

**Univerzita Karlova v Praze**

**Filozofická fakulta**

**Ústav informačních studií a knihovnictví**

# **DIPLOMOVÁ PRÁCE**

**Simona Beránková**

**Vizuální reprezentace dat jako nástroj pro prezentování veřejně dostupných dat  
statistických úřadů na webových stránkách**

Visual data representation as a tool for presenting public data on statistical offices  
websites

**2015**

Konzultant práce: Mgr. Josef Šlerka

Vedoucí práce: doc. RNDr. Jiří Souček, DrSc.

*Prohlašuji, že jsem diplomovou práci vypracovala samostatně, že jsem řádně citovala všechny použité prameny a literaturu a že práce nebyla využita v rámci jiného vysokoškolského studia či k získání jiného nebo stejného titulu.*

*V Praze dne .....*

*.....podpis*

## *Abstrakt*

Cílem práce bylo zmapovat stav webových stránek statistických úřadů a podat zprávu o míře poskytování veřejně dostupných dat a využívání různých typů vizuální reprezentace dat na webových stránkách statistických úřadů, jakožto jednoho z nástrojů naplňování poslání a komunikování statistik těchto úřadů směrem k veřejnosti.

Klíčová slova: vizualizace, vizuální reprezentace, data, statistické úřady, poslání, mise, veřejnost, infografika, český statistický úřad

## *Abstrakt anglicky*

*The main task of this diploma thesis was to map conditions of statistical offices websites and review if they fulfill their mission toward public by providing public accessible data and how and if they use the visual representation as a tool for helping people to understand statistics.*

*Klíčová slova – visualization, data, statistic office, statistics, milion, public,*

1.	Úvod.....	8
2.	Základní pojmy a jejich definice .....	9
2.1.	Data a jejich význam pro statistiku a vizuální reprezentaci.....	9
2.1.1.	Co jsou data .....	9
2.1.2.	Vlastnosti dat .....	11
2.1.3.	Typy dat .....	11
2.1.3.1.	Raw data .....	11
2.1.3.2.	Agregovaná data.....	12
2.1.3.3.	Metadata .....	12
2.1.4.	Základní kroky zpracování dat (data processing).....	13
2.1.5.	Jak data úřady poskytují.....	14
2.1.6.	Požadavky na data z pohledu statistiky a vizuální reprezentace dat .....	15
2.1.7.	Závěr: .....	15
2.2.	Informace jako produkt vizuálního zobrazení .....	16
2.2.1.	Teorie informace .....	16
2.2.2.	Matematická teorie komunikace v kontextu vizualizace .....	16
2.2.3.	Vizualizovaná informace jako věc, proces i znalost .....	17
2.2.4.	Filozofický koncept informace jako nejvhodnější přístup k informacím produkovaným vizualizací.....	18
2.2.5.	Informace jako znak .....	19
2.2.6.	Zařazení informace .....	19
2.2.7.	Závěr .....	20
2.3.	Vizuální zobrazení dat .....	21
2.3.1.	Teoretické vymezení vizuální reprezentace dat.....	21
2.3.2.	Charakteristika vizuální reprezentace dat .....	22
2.3.3.	Sémiotická podstata grafického zobrazení .....	22
2.3.4.	Vytváření vizuální reprezentace .....	23
2.3.5.	Funkce a cíle vizualizace.....	24
2.3.6.	Vizuální reprezentace dat jako nástroj k tvoření informací .....	25

2.3.7.	Vizuální zobrazení jako rozhraní pro vnímání informace .....	25
2.3.8.	Vizuální zobrazení dat jako řešení information overload .....	26
2.3.9.	Ztráta informací .....	27
2.3.10.	Interaktivita a narativita .....	27
2.3.11.	Specifické typy vizualizací užívané statistickými úřady k prezentaci dat....	28
2.3.11.1.	Infografika .....	28
2.3.11.2.	Prostředí na explorační analýzu .....	29
2.3.11.3.	Vizualizace .....	30
2.3.11.4.	Kartografické zobrazení .....	31
3.	Vizualizace v kontextu statistických úřadů .....	32
3.1.	Poslání statistických úřadů jako objasnění důvodu užívání vizualizace.....	32
3.1.1.	Vytvářet statistiky .....	33
3.1.2.	Integrita vs. Včasnost.....	34
3.1.3.	Snižování zátěže.....	34
3.1.4.	Rozvíjet obor a být autoritou .....	35
3.1.4.1.	Služba pro veřejnost a služba státní správě.....	35
3.1.4.1.1.	Služba státní správě .....	36
3.1.4.1.2.	Služba veřejnosti .....	37
3.1.5.	Vzdělávání.....	38
3.1.6.	Závěr .....	39
3.2.	Historie využívání vizuální reprezentace dat pro potřeby a prezentaci statistických úřadů.....	39
3.3.	Rozvoj principů statistického myšlení.....	40
3.3.1.	Obnovení zájmu statistiků o vizuální reprezentaci dat .....	40
3.4.	Co se snaží statistické úřady dokázat publikováním vizualizací .....	42
3.4.1.	Cílová skupina uživatelů.....	43
3.5.	Závěr: .....	45
4.	Metodologická část.....	46
4.1.	Výzkumný problém a kontext.....	46

4.2.	Cíl výzkumu a výzkumné otázky .....	48
4.3.	Stručný popis metody výzkumu.....	49
4.4.	Sledované kategorie .....	49
4.5.	Zápis údajů .....	54
4.6.	Čištění dat .....	55
4.7.	Zaznamenané a nezaznamenané weby statistických úřadů .....	55
4.4.	Kvantitativní zhodnocení dat .....	55
5.	Výzkumná část .....	57
5.1.	Přehled o datech .....	57
5.1.1.	Datové formáty .....	57
5.1.2.	Sociální média.....	59
5.1.3.	Interaktivní a neinteraktivní grafický obsah.....	60
5.1.4.	Ověření hypotéz.....	60
5.1.4.1.	Ověření hypotézy 1.....	60
5.1.5.	Geografická poloha ve vztahu k naplňování poslání statistických úřadů....	62
5.1.5.1.	První skupina.....	62
5.1.5.2.	Druhá skupina .....	64
5.1.5.3.	Třetí skupina.....	66
5.1.6.	Závěr: .....	67
5.2.	Ověření hypotézy 2.....	67
5.2.4.	Náplň poslání statistických úřadů ve vztahu k HDP dle jednotlivých kontinentů.....	69
5.2.4.1.	Africké státy.....	69
5.2.4.2.	Amerika.....	71
5.2.4.3.	Evropa .....	72
5.2.4.4.	Asie .....	73
5.2.4.5.	Oceánie.....	74
5.2.5.	Revize hypotézy .....	75
5.3.	Ověření hypotézy 3.....	75

5.3.4.	Distribuce jednotlivých typů grafické reprezentace dat.....	77
5.3.5.	Porovnání se statistickými úřady, které disponují méně než polovinou různých strojově čitelných datových formátů .....	79
5.3.6.	Závěr: .....	80
6.	Závěr: .....	82
7.	Seznam použité literatury	84
8.	Seznam obrázků:.....	101

## 1. Úvod

Během relativně krátké doby prošly statistické úřady v přístupu k veřejnosti poměrně velkou proměnou. Z uzavřeného světa se stal obor, který kromě vytváření, archivování a analyzování dat čím dál více přijímá svou odpovědnost vůči veřejnosti, tedy lidem, poskytují statistická data. Jako součást poslání této instituce se tak kromě produkování statistik stala transparentnost, otevřenost a vstřícnost. Statistické úřady jsou v současné době autoritou, jejímž úkolem je komunikovat statistiky, vzdělávat veřejnost a vést společnost k porozumění datům o sobě samé.

Tento proces otevírání se byl v současné podobě započatý zhruba před 25 lety v Británii a k jeho rychlému postupu přispělo pozdější rozšíření a zpřístupnění internetu. Velký zlomem bylo první přesunutí části datových archivů do kyberprostoru a znamenalo poskytnutí dat na jediné kliknutí komukoli na kterémkoli místě na světě. Přesto tento krok, který dnes vnímáme jako samozřejmý stav, nemohl znamenat konec procesu otevírání se statistických úřadů novým uživatelům dat. Během velmi krátké doby došlo k renesanci a znovuobjevení spojení dat a jejich grafického zobrazení. Tentokrát ovšem nikoli pro potřeby samotných statistiků, kteří tuto metodu využívali v období, kdy ještě neexistovaly dnes používané metody aplikované statistiky, ale pro představení složitých problémů a vztahů v datech ve snadno srozumitelné podobě.

Základy renesance tohoto propojení postavili už v 60. letech reformátoři statistiky J. W. Tukey a Edward Tufte, kteří vrátili grafickému zobrazení dat jeho funkčnost, smysl a modernizovali ho pro využití v současném světě, který je daty nasycen. K plnému využití potenciálu pro komunikování statistik touto formou ale mohlo dojít až s možnostmi webu 2.0, který umožnil vyvinutí řadu nástrojů na zpracování a vizualizování velkých objemů dat. A právě tato cesta propojení stále se rozvíjejícího oboru počítačové vizualizace dat využívaného jako prostředek komunikace významu velkých množství dat, kterými statistické úřady disponují, veřejnosti, je předmětem této práce.

Na téma propojení grafické reprezentace dat a využití statistickými úřady byla vytvořena řada studií, část z nich se soustředila na zkoumání vhodnosti takového propojení (Friendly, 1999), další se zabývaly vytvořením nových metod a přístupů, které by v tomto ohledu mohly být uplatněny (Prodromou, 2013), nebo popisovaly strategii jednotlivých úřadů (australský). Přesto je kompletní zhodnocení celkového fenoménu, ve kterém statistické úřady působí jako autorita, která poskytuje veřejnosti nejen data, ale i vzhled ve formě, která je srozumitelná a snadno pochopitelná, málo zmapované.



Cílem práce je proto podat zprávu o míře a využívání vizuální reprezentace dat na webových stránkách statistických úřadů jakožto jednoho z nástrojů naplňování poslání těchto úřadů směrem k veřejnosti.

Tato práce se drží tradiční struktury diplomové práce a je rozdělena do tří základních částí – teoretickou, metodologickou a analytickou. V první části se věnuji vymezení pojmu grafického zobrazení dat a jeho různých forem, které využívají statistické úřady, dále popisují poslání statistických úřadů a předkládám průřez současným využívání vizualizací v rámci tohoto poslání. V metodologické části vytyčuji hypotézy, definuji cíle a popisují způsob sběru dat a jejich analýzy. Stěžejní částí mé práce je předložení stavu zkoumaného fenoménu a ověření stanovených hypotéz. V závěru shrnuji nové poznatky, které tato práce přináší.

## **2. Základní pojmy a jejich definice**

### **2.1. Data a jejich význam pro statistiku a vizuální reprezentaci dat**

Data jsou pro statistiku velmi cenná nejen jako zdroj, ale také pro svou hodnotu vyplývající z rozsahu sběru v rámci poslání statistických úřadů. Obsahují záznamy o značném počtu respondentů, stavu společnosti a takové údaje mohou být při neopatrnosti snadno vystaveny zneužití. I proto je zacházení s nimi ošetřeno v řadě zákonů, právních předpisů a směrnicích.<sup>1</sup>

Právě o data a často dokonce big data v syrové nebo zpracované podobě, se v posledních letech, kromě odborníků, začala zajímat také veřejnost. Tomuto zájmu pomohlo dostatečné rozšíření výkonných počítačů, schopných zaznamenat datový proud prakticky v reálném čase, zobrazit velké množství dat a následně jej zpracovat, popřípadě tato data extrahovat dříve nedostupnými technikami jako je scrapping. Rostoucí zájem ilustruje například publikování článku Data Is the New Oil z roku 2006 (Palmer, 2006) a řada dalších s podobným titulkem, které představují data jako bohatství čekající na zpeněžení podobně, jako je tomu právě u ropy.

#### **2.1.1. Co jsou data**

---

<sup>1</sup> V České republice se jedná o zákony: Zákon o ochraně osobních údajů 101/2000 Sb., Zákon o svobodném přístupu k informacím 106/1999 Sb., Zákon o statistické službě 89/1995 Sb. Dále jde o Kodex evropské statistiky, Nařízení Evropského parlamentu a Rady (ES) č. 223/2009, Nařízení Evropského Parlamentu a Rady (EU) 2015/759, Nařízení Komise (EU) č. 557/2013, Rozhodnutí Komise 2012/504/EU, Nařízení Komise (EU) č. 611/2013. V Estonsku jde například zákon Official Statistical Act – Riigi Teataja, Ve Velké Británii o Official Statistics Act a další.

Data existují v různých podobách a typech (více v 2.1.3) a k jejich popisu lze přistupovat z různých pohledů. Samotných definic dat je značné množství (Zins, 2007, s. 479-493) a konkrétní znění závisí na kontextu nebo na zvoleném oboru.<sup>2</sup>

Většina definic pojmu „data“ vychází z latinského významu slova, kdy jde o plurál slova „datum“, znamenající věc, která je dána. Například Buckland (1991, s. 351-352) popisuje datum jako všechno, co může rozšířit znalosti a lze zaznamenat nebo předat ústně. Další velmi široké pojetí představuje Floridi (2010, s. 23), který vychází ze slova „dedemona“, řeckého označení pro data a tvrdí, že data jsou v reálném světě cokoli, co postrádá uniformitu. Na jazyku, formátu a médiu dat přitom nezáleží.

Obecně lze k většině teorií říci, že vědci k této entitě přistupují jako ke zdroji, stavebnímu kamenu, prvnímu článku informačního řetězce či podstavě informační pyramidy, které se skládají z informace, znalostí a moudrosti (Shedroff, c1999, s. 267-293). Je tak nejmenší rozeznatelnou jednotkou věci, kterou lze měřit (Zins, 2007, s. 479-493). Stonier tvrdí, že data jsou nepropojená, nezprocesovaná fakta a pozorování (Stonier in Zins, 2007, s. 479-493). Z různých definic je nicméně patrné, že vědci nejsou jednotní v tom, zda jsou data přímo fakta nebo jde o jejich reprezentaci (Hey, 2004, s. 6;). V kontextu sémiotiky je možné data také považovat za znaky, které označují entity nebo vlastnosti v blízkém kontextu (Zins, 2007, s. 284).

Buckland ve svém systému nepoužívá dnes tradiční hierarchii a považuje data za jeden z typů informací, přičemž rozlišuje informaci jako proces, věc a vědomost. Data spadají do stejného typu jako texty, dokumenty, události, objekty a v podstatě jakýkoli jiný systém skladující a vydávající informace. Data jsou tak hmatatelný typ informace. (1991, s. 351-360)

Jiní data považují za entity nižšího významu a hodnoty. Pro Woodwarda jsou důkazem existence jevů, tedy tím, co je zachyceno, a v podstatě představují pouhý šum, z něhož se vědci, včetně statistiků, snaží extrahovat signál, který je právě onen jev, jenž je následně podroben zkoumání (1989, s. 393-421). Pro tento typ sebraných dat byl definován také pojem *capta*, (Lanigan, 1994, s. 285) který označuje „něco sebraného, aktivně vzatého“, na rozdíl od dat, která představují něco daného. Pro Checklanda a Hollwela (2006, s. 47-55) jde v případě dat a *capt* o sousledný vztah. Data jsou všechny masy faktů, pozorování a konceptů, které existují. *Capta* jsou potom mezistupněm mezi daty a informacemi, tedy takovými daty, která jsou zaznamenána a vybrána na základě kritérií. Přímou říkají, že „*využitím tohoto konceptu můžeme definovat data pocházející z výzkumné aktivity jako capta, tj. data sebraná od širokého spektra zdrojů, které reprezentují a popisují jen výzkumnou aktivitu.*“ (Checkland & Hollwel, 2006, s. 52). Data

---

<sup>2</sup> Definice tohoto pojmu se vzájemně liší ve statistice, matematice, informační vědě či IT

vstupující do procesu tvorby statistik a vizualizací jsou tedy podle tohoto přístupu ve skutečnosti *capta*.

### **2.1.2. Vlastnosti dat**

Z jednotlivých teorií v předešlé podkapitole lze odvodit základní požadavky a obecné vlastnosti dat. Konkrétní požadavky na data z pohledu statistiky a vizualizace jsou rozebrány v další podkapitole. Checkland a Holwell (2006, s. 50-55) na základě ontologických přístupů definovali 3 pohledy na data - a to objektivní, subjektivní a intersubjektivní. Z objektivního pohledu jsou podle nich data faktickým výsledkem měření, jsou naprosto neutrální, bez kontextu, a jsou nestructurovaná. Ze subjektivního pohledu potom lze o datech říci, že nejsou nutně pravdivá, mohou obsahovat chyby a lze pochybovat, zda jsou pravdivou reprezentací faktu. Data reprezentují informace a jsou jedinou možností, jak je vyjevit, neboť jsou hmatatelný typ informace (Buckland, 1991, 351-354). Pouze data jsou tím, co lze komunikovat, přestože sama o sobě nemají také žádný význam. Pokud reprezentují názor, nemusí být nutně objektivní.

Z intersubjektivního pohledu potom platí, že smyslem dat je připustit komunikaci. K tomu je nutné, aby byla data kódována v ustálené formě a byla alespoň částečně strukturovatelná. Data obsahují také kontext, který ozřejmuje konkrétní důvod jejich sběru. (Checkland & Holwell , 2006, s. 50-55)

### **2.1.3. Typy dat**

#### **2.1.3.1. Raw data**

Takzvaně syrová data, jsou data, která obsahují šum a nejsou interpretovaná (Harris et al., 2003, s. 1509). Podle Hershe jde o záznam „syrového pozorování světa, s minimální kontextovou interpretací.“ (Zins, 2007, s. 484) Jde o data v původní podobě tak, jak byla sesbírána. Jsou tedy ve stavu, v němž není možné je použít a musí dojít k jejich dalšímu zpracování, tj. odstranění šumu, strukturování, v některých případech i k redukování některých záznamů ať už z důvodu chybovosti nebo přílišného vybočování.

V kontextu statistických úřadů se raw data mohou vyskytovat jako data z administrativních zdrojů nebo nezpracované vyplněné dotazníky získané od respondentů v rámci šetření, která jsou předmětem dalšího zpracování jako anonymizace nebo agregování do takové formy, v níž není možné identifikovat jednotlivé uživatele. Tento proces se děje v souladu s příslušnými zákony. Pro celou Evropskou unii platí Nařízení Evropského Parlamentu a Rady č. 223/2009 ze

dne 11. 3. 2009 o evropské statistice, které vymezuje problematiku statistické důvěrnosti. V české legislativě tomuto nařízení odpovídá zákon č. 89/1995 Sb., o státní statistické službě. Ochrana osobních údajů respondentů je zaručena zákonem č. 101/2000 Sb., O ochraně osobních údajů. Legislativní prostor je dále vymezen zákonem č. 106/1999 Sb., O svobodném přístupu k informacím a zákonem č. 412/2005 Sb., O ochraně utajovaných informací a o bezpečnostní způsobilosti (Holý, 2011, s. 28).

### **2.1.3.2. Agregovaná data**

Agregovaná data jsou data strukturovaná a vznikají částečným zpracováním mikrodat.<sup>3</sup> Děje se tak obvykle sumarizací nebo průměrováním. Agregovaná data obsahuje například Veřejná databáze ČSÚ i databáze dalších statistických úřadů, do kterých se ukládají statistické údaje vznikající při zpracování statistických šetření. Metoda agregace dat je v případě databází volena z důvodu zajištění anonymity, ochrání citlivých dat a nepřehlcení uživatelů. V průběhu procesu dochází zcela logicky ke ztrátě informací (Clark et al., 1976, s. 429-430), což je v případě vytváření vizuálních reprezentací dat spíše kontraproduktivní, neboť může vést ke zkreslení výsledků, tj. zobrazení.

Floridi<sup>4</sup> nazývá jak mikrodata, tak agregovaná data, daty primárními. Tedy takovými, která jsou v databázi a byla sesbírána ke specifickému účelu pro konkrétní systém. (2010, 29-32)

### **2.1.3.3. Metadata**

Metadata jsou definována jako data o datech (Hey & Trefethen, 2003, s. 2) nebo jako informace o datech (Edwards et al., 2011, s. 667-669). Popisují věci nebo informační objekty a jsou nepostradatelná při archivování dat a vytváření databází, kdy jsou součástí tzv. archivačního procesu (Albright & Lyle, 2010, s. 17-21), kterému předchází procesování dat a následuje vytvoření databáze.

Metadata obsahují řadu informací o obsahu i kontextu vzniku dat, jejich umístění, formátu a významu či názvu proměnných, aj. Umožňují lepší orientaci v souboru při hledání dat a setů. Jejich účelem je snižovat tzv. data friction, tj. eliminovat neshody údajů, nepřesnosti, chyby vznikající v případě vzájemného kontaktu dvou systémů. Nejsou tedy výstupem nebo produktem, ale jejich role je poskytovat informace, prakticky slouží jako náhrada komunikace mezi autorem datasetu a jeho pozdějším potenciálními uživateli z různých institucí, zemí, v různé době (Edwards et al., 2011, s. 672).

---

<sup>3</sup> Konkrétní údaje o obyvatelích či domácnostech neboli jakýchkoli jednotkách

<sup>4</sup> Floridi dále klasifikuje ještě operational data – informující o dynamice a fungování systému a “odvoditelných datech”, která vznikají analýzou nebo srovnáváním dat s podobným kontextem. A sekundární data, která definuje jako opak dat primárních, tedy záznamy o tom, co chybí.

Právě pro usnadnění používání metadat došlo v průběhu posledních 15 let k vývoji a ustálení standardů s cílem vytvořit běžně užívaný a pro lidi i stroje srozumitelný jazyk k archivaci digitálních dokumentů (Hey & Trefethen, 2005 in Edwards et al., 2011, s. 672; Jones et al., Edwards et al., 2011, s. 672; Lawrence et al., 2009 in Edwards et al., 2011, s. 672; Michener, 2006 Edwards et al., 2011, s. 672; Signell et al., 2008 in Edwards et al., 2011, s. 672), jež umožňují prodloužit životnost dokumentů. V současnosti se pro účely strukturování databází obvykle používá značkovací jazyk XML<sup>5</sup>, který k definování struktury a významu libovolných dat používá tagy.<sup>6</sup> Data v tomto strojově čitelném formátu poskytují i některé z databází dostupné na webech statistických úřadů.<sup>7</sup>

#### **2.1.4. Základní kroky zpracování dat (data processing)**

Data před jakýmkoli dalším použitím vyžadují alespoň základní zpracování. Jde o několik ustálených kroků, které zahrnují sběr, zachycení, editování, opravy a odhad procedur (Chambers, 1993, s. 182-184). V případě statistických dat je vhodné zahrnout ještě postupy související s ochranou dat a zabránění identifikace respondentů.

Sběr a zachycení dat probíhá, v případě statistických úřadů, zpravidla prostřednictvím elektronických či korespondenčních dotazníků<sup>8</sup>. Ve specifických případech navštěvují respondenty komisaři. Takto získaná data je nutné podrobit procesu čištění, při kterém může dojít ke zmenšení datasetu až o 10 procent záznamů (De Veaux, Hand, 2005, s. 231-238), které se mohou ukázat jako nevyhovující. Při čištění dochází k harmonizování, opravám dat a k agregaci.

Podstatnou součástí processingu je anonymizace dat. Cílem toho procesu je upravit, pozměnit, datové záznamy před publikováním tak, aby zůstalo co nejvíce informací co nejkonkrétnějších a zároveň nedošlo k identifikování jednotlivých respondentů (De Veaux, Hand, 2005, s. 231-238). Statistické úřady jsou za takovou úpravu dat odpovědné. Jednou z metod anonymizace je například data swapping, při kterém dochází k záměně citlivých proměnných (Holý, 2011, s. 27).

Při anonymizaci je problémem nejen odkrytí identity konkrétního člověka, které se eliminuje odstraněním osobních údajů jako je jméno či adresa, ale i rozpoznání jedince na základě kombinace kategorií a údajů, kdy určité údaje mohou například prozradit lokalizaci

---

<sup>5</sup> Pokusů definovat standard pro metadat bylo samozřejmě více, například OAIS, CF, net cdf. Všechny přistupují k metadatům jako explicitním informacím kompilovaným ve standardizovaných kategoriích. Pro formalizaci metadat se používají i další jazyky jako DAMIL, OIL, EML.

<sup>6</sup> Extensible Markup Language (XML). [online]. Available from: <http://www.w3.org/XML/> [Accessed 2015-12-25]

<sup>7</sup> Například Veřejná databáze ČSÚ, Statistical Database estonského statistického úřadu, aj

<sup>8</sup> Přes rozhraní aplikace dante/Web v případě ČSÚ

respondenta.<sup>9</sup> Zpravidla se na tyto úkony používá zvláštní software.<sup>10</sup> Během zpracování může dojít také k řadě chyb, které vedou k záměrnému nebo nezáměrnému zkreslení a následně i k chybným závěrům. Tento problém vzniká zejména v případě analýzy sekundárních<sup>11</sup> dat, sebraných původně k jinému účelu. Pro nový účel však mohou být nedostatečná i v případě, že pro původní účel mohla být perfektní (De Veaux, Hand, 2005, 231-238).

### 2.1.5. Jak data úřady poskytují

Jedním ze základních poslání statistických úřadů je služba veřejnosti (více v kapitole 3.1), která zahrnuje zejména poskytování dat,<sup>12</sup> kdy jde nejen o poskytnutí dat k dalšímu použití a naplnění služby veřejnosti, ale také o formu kontroly využití veřejných peněz (European Statistical system Committee. 2011) a demonstraci transparentnosti (Albright & Lyle, 2010). Poskytovaná data musí splňovat několik vlastností – především být anonymizovaná, pravdivá, přesná, v běžně čitelných formátech, aktuální, aktualizovaná a nezkrácená. Ve většině případů jsou data rovněž agregovaná.

Data jsou veřejnosti dostupná obvykle na webových stránkách příslušného statistického úřadu. Díky rozšíření a zrychlení internetového připojení je možné zpřístupnit data komukoli, kdo na web získá přístup. Děje se tak obvykle zdarma a běžných strojově čitelných či nečitelných datových formátech – nejčastěji XLS, PDF nebo CSV (více ve výzkumné části 55.3).

S prezentováním dat jinde než přímo na webu příslušného úřadu souvisí také využívání API, setu protokolů a postupů sloužících k vytvoření prostředí, které zpřístupní příslušné datové sady a záznamy. Tuto formu poskytování dat zatím podporuje pouze několik statistických úřadů jako například ONS, estonský statistický úřad, katalánský statistický úřad nebo US Census. Dalším způsobem prezentace jsou například rychlé informace, které značně proměnily způsob zveřejňování dat – z periodického cyklu vydávání přehledů a analýz se stal prakticky kontinuální stream jednoduchých textových přehledů o změnách a aktuálních statistikách. Řada úřadů také poskytuje údaje prostřednictvím mobilní aplikace a publikací.

Zřejmě nejviditelnější<sup>13</sup> platforma, kterou pro prezentování dat používá ČSÚ, se jmenuje Veřejná databáze<sup>14</sup>. Zpuštěna byla v roce 2006 a skládá se z úložiště dat opatřených metadaty a předdefinovaných tabulek, grafů, map. Tabulky jsou navrženy dynamicky, grafický výstup lze zobrazit prostřednictvím nástroje Google Motion Chart. Podporované vkládání grafů na vlastní

<sup>9</sup> Pro tyto účely se používá data swapping, tedy záměna citlivých proměnných či mikroagregace, tj. kombinace podobných záznamů a určení průměru v rámci skupinového skóre u každého případu

<sup>10</sup> V ČSÚ se pro tyto účely používá programy  $\mu$ -ARGUS pro oblast ochrany mikrodat, a  $\tau$ -ARGUS produkuje zabezpečené tabulky. Ačkoli je účelem systémů ARGUS je minimalizovat riziko prozrazení za současné maximalizace vypovídací schopnosti výsledného zpracování, dochází i při jejich použití ke ztrátě dat.

<sup>11</sup> Zde: data využita k jiným účelům než byla zaznamenána. Např. administrativní data

<sup>12</sup> Pro český statistický úřad je závazně popsáno v zákoně 89/1995 Sb., a Kodexu evropské statistiky Statistický,

<sup>13</sup> Veřejná databáze ČSÚ má desetitisíce přístupů ročně

<sup>14</sup> <http://vdb.czso.cz>

webové stránky je součástí strategie, jak šířit výsledky statistických šetření, neboť v současné době se statistické úřady nesoustřeďují pouze na přivádění uživatelů na web, ale naopak vytváří nástroje pro automatické sdílení dat, díky nimž mohou koncoví uživatelé statistické údaje využívat ve vlastních aplikacích či zařízeních (Roszka, 2015, s. 60-76). Podobné nástroje na exploraci dat poskytují na svých webových stránkách také další statistické úřady a mezinárodní organizace jako Eurostat, OSN či African Development Bank. (podrobněji 2.3.11.2)

### **2.1.6.Požadavky na data z pohledu statistiky a vizuální reprezentace dat**

Data jsou pro vizuální reprezentaci dat, stejně jako ve statistice, primárním zdrojem (Mazza, c2009; Tufte, c2001), pro něž platí, že musí být především pravdivá, neboť zkreslená data vedou k chybným závěrům i v případě, že byl dodržen správný postup zpracování (De Veaux & Hand, 2005, s. 232-234). Je také důležité, aby byla čitelná a dostupná nejlépe ve strojově čitelných formátech.

Stejně jako pro účely statistiky i pro vytváření vizuálních reprezentací je podstatné, aby používaná data byla aktuální nebo alespoň platná. Včasnost ale není pro tvorbu vizuálních reprezentací dat vždy nejdůležitějším kritériem. Zde mu předchází například požadavek na co největší bohatost a rozsah datasetu, jež je pro vyjevení informací prostřednictvím grafického zobrazení stěžejní (Tufte, c2001, s. 13). Taková situace nastává v případě zobrazování časové změny, kdy je naopak žádoucí shromáždění i dat, která již platná nejsou. Přesto jejich zobrazení poskytne cenné informace. Pro oba obory je ale podstatná tzn. relevance dat, tedy výběr vhodných dat.<sup>15</sup> Pokud data nejsou relevantní pro daný případ, pak ani odpovídající metoda zpracování a správný postup nezajistí očekávaný výstup nebo výsledek.

### **2.1.7.Závěr**

Data jsou entity, které samy o sobě nenesou žádný význam, reprezentují fakta a popisují jevy. Pro statistiku jsou proto předmětem sběru stejně tak, jako jsou pro vizualizaci předmětem zájmu a dalšího zpracování (Tufte, c2001; Mazza, c2009; Chen, c2006). Vyskytují se v řadě typů, forem a formátů, z nichž některé slouží přímo jako předmět zájmu, jiné například k jejich popsání. Zároveň musí splňovat určité požadavky nutné k jejich zpracování. V případě statistických úřadů vzniká vůči veřejnosti přesně definovaná odpovědnost, na jejímž základě se vymezují konkrétní požadavky na vlastnosti dat. Pokud vizuální reprezentace dat vzniká v rámci této odpovědnosti, měla by splňovat stejné požadavky jako samotné statistiky. Jedním z nich je například respektování ochrany osobních dat a anonymizace, aktuálnost, nezkrivenost a pravdivost.

---

<sup>15</sup> V kontextu vizualizace může jít o výběr odpovídajícího datasetu k zobrazení, v případě statistik může k otázce relevance dojít například při využívání administrativních dat

## **2.2. Informace jako produkt vizuálního zobrazení**

Jak je zjevné z řady konceptů Data-Information-knowledge-Wisdom DIKW modelů, včetně Shedroffova kontinua porozumění (c1999, s. 267-293), základním stavebním kamenem pro tvorbu informací jsou data, jejichž koncept byl popsán v předchozí kapitole. Vizuální reprezentace dat organizuje a reprezentuje data tak, aby jejím konečným produktem byly informace (Mazza, c2009, s. 18-23). Jsou-li tedy informace produktem, potom se musí v informačním řetězci vizuální zobrazení nacházet právě mezi daty a informacemi (Mazza, c2009, s. 8-10).

### **2.2.1. Teorie informace**

Informace je základní oběživo, cíl a výsledek dobře postavené vizuální reprezentace dat. Vědci přistoupili na shodu, že jde o polysémantický koncept, jehož vysvětlení závisí na způsobu užití. Proto je běžné, že se v souvislosti s informací hovoří o řadě konceptů, například Fischerově informaci (Kullback, 1978, s. 26-28), Shannonově informaci (Floridi, 2010, s.24-30), aj. Nejčastěji zmiňované koncepty pohlížejí na informaci ve dvou základních přístupech – kvalitativním a kvantitativním - které následně formují dvě základní teorie přístupu k informaci. Starší, kvantitativní, přístup je reprezentován matematickou teorií komunikace, v níž se Shannon zabývá syntaktickou informací. Kvalitativní Floridiho pojetí se také označuje jako filozofie informace, jež se zabývá informací sémantickou. Ta je pro popsání informace, která vzniká pomocí vizuální reprezentace dat vhodnější, a proto je v této souvislosti obecně platně používanou teorií (Floridi, 2005, s. 351-370).

### **2.2.2. Matematická teorie komunikace v kontextu vizualizace**

Přestože se Shannonova matematická teorie komunikace <sup>16</sup>zabývá ryze technickým vysvětlením pojmu informace a týká se přenosu, datové komprese a kódování dat, kde v žádné fázi přenosu není předmětem porozumění obsahu, vychází z této teorie i model vzniku sémantické informace (Floridi, 2009, s. 13). Navzdory tomu, že jsou oba koncepty informace prezentovány jako

---

<sup>16</sup> V něm jde o přenos informace od informujícího k informovanému obsahující kódér vysílající signál a dekodér, který signál přímá. Doručení zprávy v původní podobě potom ohrožuje šum



opačné póly, jsou ve Floridiho pojetí informace zobrazeny jako součásti jediného konceptuálního diagramu, tzn. Informační mapy, kde je filozofie informace de facto dalším stupněm Shannonovy teorie (Floridi, 2009, s. 13-53). Tohoto přístupu Floridi dosáhl vytvořením obecné definice informace, kde Shannonovu informaci považuje za tzv. „dobře zformovaná data.“ (Floridi, 2009, s. 16). Blízkost obou systémů<sup>17</sup> je zřejmá také z podobnosti mezi úkolem vizuálního zobrazení dat a komunikačního modelu – tedy doručení informace od zdroje k příjemci a dále adresátovi, a to rychle a přesně. Podobně jako u Shannonova kódéru je navíc do vizualizace „zakódována“ informace a adresát ji dekóduje (Chen et al., 2014, s. 7). V případě sémantické informace a jejího přenosu pomocí vizualizace platí, že přílišná orientace na porozumění obsahu přenášené informace není žádoucí – například může dojít k porušení soukromí nebo k identifikaci respondenta. Tomu se v případě statistických úřadů předchází anonymizací a agregací dat/publikováním agregovaných datasetů. (této části jsem se věnovala v kapitole 2.1

### **2.2.3. Vizualizovaná informace jako věc, proces i znalost**

Ačkoli je obecně přijímanou teorií v souvislosti s tvorbou vizuálního zobrazení dat koncept filozofie informace (Ware, 2013, s. 25-30; Card et. Al. 1999, s. 1-35)<sup>18</sup>, je nutné v této souvislosti prezentovat také Bucklandovu teorii informace (1991, s. 351-355), která nabízí základní klasifikaci pro rozdílné, na informaci závislé, aktivity. Pro Bucklanda je informací už samotný proces informování (1991, s. 351-352), stávání se informovaným, tedy měnění dosavadní uživatelské znalosti. Do informace-jako-procesu je zahrnut také akt sdělování, tedy komunikace znalosti nebo zprávy.

Vyvození závěru, výsledek z toho, co je přijato, tedy z procesu informování, je pojmenováno jako informace-jako-žnalost<sup>19</sup>. Oba tyto typy jsou nehmotné a nehmatatelné, žnalost je sama o sobě nepřenositelný typ informace, který ale lze zhmotnit a to prostřednictvím posledního typu – informace-jako-věci<sup>20</sup>, což je pojem atributivně použitý pro objekty, dokumenty<sup>21</sup>, data a datasety<sup>22</sup>. Jejich kvalitou je schopnost komunikovat informaci a propůjčovat žnalosti, samy

---

<sup>17</sup> Ačkoli existuje mezi oběma systémy řada podobností na to, aby bylo podle Floridiho a Chena ospravedlnitelné uvážit informační teorii jako teoretický rámec k „k podepření“ vizualizace, jsou si vědomi velkých rozdílů například v zapojení lidského faktoru, transformování dat během vizualizačního procesu navzdory doručení zprávy ve stejném typu u Shannonova komunikačního modelu a důrazu na kompresi zprávy.

<sup>18</sup> Teorii publikoval Luciano Floridi, vlastněm a povaze vizualizace ovšem odpovídá také koncept 3 typů informace (Buckland, 1991)

<sup>19</sup> Information-as- knowledge

<sup>20</sup> Information-as-thing

<sup>21</sup> Model je informativní reprezentace originálu Dokument je zde generický termín označující jakoukoli fyzickou informaci, definici odpovídá i diagram či vizualizace

<sup>22</sup> Datasety jsou označovány jako „informace“ protože jsou informativní, tedy nesou informaci

jsou ale pouhou reprezentací nehmotné znalosti, s níž na rozdíl od znalosti lze interagovat přímo. Právě takovou reprezentací je vizuální zobrazení<sup>23</sup>.

#### **2.2.4. Filozofický koncept informace jako nejvhodnější přístup k informacím produkovaným vizualizací**

Jak už bylo zmíněno, filozofie informace je komplexní a nejčastěji obecně přijímanou teorií informace, která je aplikovaná na proces vizuální reprezentace dat a grafické zobrazení (Chen et al., 2013, s. 3424). Na informaci nahlíží ze sémantické perspektivy, tedy se zajímá především o obsah sdělení, které nese a způsob jeho vysílání a přijímání.

V tomto kontextu je informace definována ve své nezákladnější podobě jako spojení dat a významu (Floridi, 2009, s. 16). V komplexnějším pojetí se označuje za konceptuální labyrint (Floridi, 2009, s. 15), ve kterém se rozlišuje sémantická informace na instrukce a fakta, a dále na nepravdivá a pravdivá fakta, z nichž mohou dále vzniknout znalosti. V rámci filozofického konceptu informace se považuje za danost, že informace se sestává z dat<sup>24</sup>. Tento fakt je zároveň základem definice pro General Definition of Information (GDI), obecnou definici informace (Floridi, 2009, s. 16), která zní:

- 1) Informace se sestává z dat, v počtu větším než 1,
- 2) Tato data jsou dobře uspořádaná<sup>25</sup>,
- 3) Tato dobře uspořádaná data mají význam, jsou smysluplná – projdou-li předchozím procesem zformování, mají význam a dávají smysl.

Ačkoli v definici GDI<sup>26</sup> není explicitně uvedena podmínka pravdivosti, teorie předpokládá presumpci informativnosti, což znamená, že data jsou uvažována dobře zformovaná, smysluplná a pravdivá, pokud není dokázán opak. Chybná/lživá informace je zároveň stále považována za informaci<sup>27</sup>.

---

<sup>23</sup> Do tohoto konceptu patří ještě 4. element, takzvané zpracování, procesování informací, který je odlišný od informace-jako-procesu. Společně s vlastnostmi jako je (ne)hmotnost dokáže popsat 4 aspekty informace.

<sup>24</sup> Což dostatečně ověřily jiné obory jako Informační věda in Information Systems Theory, Methodology, Analysis and Design, in Information (Systems) Management, in Database Design and in Decision Theory)

<sup>25</sup> Well-formed information (např. jsou sdružena dle pravidel syntaxu, tedy formy, kompozicem, koncepce, struktury či jazyka, či analyzována podle určitého klíče.

<sup>26</sup> Navzdory podmíněnosti (Hjorland, 2007, 1449) informace, může i v rámci GDI existovat nezávisle na informovaném. Tento princip general neutrality, který připouští existenci informace bez informovaného subjektu, tedy to, že se význam nenachází v mysli uživatele (Floridi, 2005, s. 358), posouvá vizualizaci do pozice nositele, nikoli zprostředkovatele informace, jakéhosi agenta, který indikuje změnu, neboť informace je právě informací o změně (Floridi, 2005, s. 359).

<sup>27</sup> Toto dělení například u Bucklanda (1991, s. 351-360) není, ten nepravdivou informaci nepovažuje za informaci (mis a disinformace)

Floridi zároveň ale pravdivou informací povyšuje na vyšší sémantický koncept, tzn. Silně sémantickou informací<sup>28</sup>, jejíž podmínkou je právě její pravdivost, nikoli ale správnost. Pravdivost je definována jako vzdálenost sdělované informace od faktu, zdroje či situace, o kterém má informace referovat a označuje se jako míra informativnosti.(2009, s. 16) Výsledek procesu data modelingu, tedy formování dat, nemusí přesně reprezentovat skutečnou povahu analyzovaného, podobně jako fotografie, statistická data dostupná pro veřejnost, která jsou agregovaná a anonymizovaná nebo vizuální zobrazení dat, jež ze své podstaty trpí ztrátou informací (Chen et al., 2013, s. 3429).

V souvislosti se statistickými daty a informacemi je hledání pravdivosti částečně miskoncepcí, neboť současná statistika se zakládá na pravděpodobnostech. V tomto kontextu je spíše vhodné zmínit podmínku spolehlivosti a důvěryhodnosti dobře zformovaných a smysluplných dat. Taková data by měla splňovat 22 vlastností od přesnosti až po platnost (Mai, 2013, s. 682). Na důvěryhodnost informace má vliv také reputace jejího autora. V tomto kontextu je vhodné zmínit statistické úřady, jež by měly být v ideálním případě chápány jako spolehlivé, objektivní a důvěryhodné instituce.

### **2.2.5. Informace jako znak**

Popsání informace, kterou produkuje vizualizační proces, jako informací sémantickou, vede k zájmu o konvenčnost informace v podmínkách obecné definice informace (GDI). Explicitně informační podmíněnost zmiňuje také Hjørland, který tvrdí, že „informace má subjektivní a situační porozumění, tedy, že tvrzení a myšlenky budou znamenat něco jiného v různých společnostech.“(2007, s. 1449). Tato konvenčnost informací propojuje s Peircovým konceptem označujícího a označovaného<sup>29</sup>.

Peirce informací označuje přímo za znak, neboť jde o jev, který je reprezentací objektu, jež musí být interpretován (Vickers, 2013, s. 1-2). Je mu tedy – konvenčně, na základě dohody, závisí tedy na kontextu - přisouzen význam, což je jedna z podmínek pro to, aby se data mohla stát informacemi. Hjørland (2007, s. 1449) zároveň tvrdí, že proces získávání významu je komunikací a odtud je tedy možné také odvodit, že výstupem vizualizace jsou informace, neboť řada teorií ji rovněž považuje za komunikaci (Spence, 2001 a Cleveland 1994 in Friendly, M. 1999; Shedroff, 1999, s. 267-270).

### **2.2.6. Zařazení informace**

---

<sup>28</sup> Strongly Semantical Information, SSI

<sup>29</sup> Tedy se sémiotikou

Z dosavadní části této kapitoly je zřejmé, že informace je složitějším konceptem krystalizujícím z dat. Přesto ani tato forma nestojí na pomyslném vrcholu pyramidy, ale je součástí komplexnější struktury nazývané Data-Information-Knowledge-Wisdom (DIKW), která se v informační vědě obvykle prezentuje jako pyramida s podstavou z dat a vrcholem v moudrosti. Pro potřeby vizualizace je ale vhodnější spíše zpodobnění řetězce, které poprvé popsal Nathan Shedroff (c1999, s. 267-293), který neevokuje hierarchii, ale ukazuje přerod jednoho jevu do druhého s naznačením vztahů, popsáním vlastností a definováním cíle. Tento koncept se nazývá kontinuum porozumění<sup>30</sup>.

Jeho součástí je informační design, který formuje data tak, aby byla snadno pochopitelná bez důrazu na design, který je předmětem interakčního designu zaměřeného na interaktivitu a storytelling (kapitola 2.3.10) Důraz je kladen na pochopení transformovaných dat<sup>31</sup>, která se v ideálním případě u příjemce přetvářejí na znalosti. Znalost vzniká osvojením informace, v prezentaci a myslí uživatele<sup>32</sup>. O její co nejefektivnější sdělení se snaží princip kontinua interaktivity. Znalost je nejhodnotnějším typem zpráv, který můžeme sdělit (Shedroff, 1999, s. 267-293). Znalosti mají společné malé množství sdílené zkušenosti, je proto obtížné je komunikovat a najít vhodnou formu jejich prezentace pro široké publikum.

Moudrost je ryze interní a osobní úroveň chápání. Jedná se v podstatě o metaznalosti, které jsou výsledkem vyhodnocování, retrospekce, uvažování a interpretace znalostí. Moudrost zároveň není možné vytvořit tak jako informace, ani ji nelze sdělovat jako znalost.

Aby bylo možné tyto koncepty naplnit, musí být především zřejmý důvod komunikování zprávy. Celý proces porozumění tak začíná vytvořením smyslu a následně vyvinutím vhodného způsobu prezentace na straně autora prezentace. Jsou-li tyto podmínky naplněny, je výsledkem kontinua porozumění zkušenost, kterou uživatel na základě interakce a pochopení smyslu, nabývá.

### **2.2.7.Závěr**

Informace je koncept, který je v ideálním případě produktem vizualizace. Vzniká transformací dat a je zároveň předmětem interakce při tvoření znalosti na základě zkušeností v rámci kontinua porozumění. Jde o jev, který může být asociován s několika vysvětleními. Pro potřeby vizualizace se ale nejlépe uplatní filozofický přístup k sémantické Informaci, který se zabývá jejím významem, proměnou, relevancí, spolehlivostí, užitečností, čitelností a interpretací. Hledá odpověď na to, kdy se entita může nazývat informací. I proto autor tohoto konceptu navrhl tzv.

---

<sup>30</sup> Continuum of Understanding

<sup>31</sup> Ta sice sama o sobě nenesou význam, ale při nalezení vzorců a vzájemných vztahů je možné jim význam přiřadit

<sup>32</sup> Nejlepším způsobem pro získání znalosti je podle Shedroffa (1999, 267-293) interakce, snahou je tedy vytvářet co nejvíce interaktivní prezentace

obecnou definici informace, která se uplatní nejen při tvorbě vizualizace, ale i v souvislosti s publikováním dat a vytvářením statistických analýz v rámci statistických úřadů, což jsou jedny ze základních poslání těchto institucí.

### **2.3. Vizualní zobrazení dat**

Vizualní zobrazení dat je se statistickou praxí a analýzou sebraných dat spojeno již od počátku pokusů o státem organizovaný systematický sběr dat (Friendly, 2008, s. 502). Před vynalezením moderních statistických metod bylo právě grafické<sup>33</sup> zobrazení stěžejním nástrojem pro porozumění jejich významu (více vysvětlují v kapitole 3.3.1) Současný zájem o grafické zobrazení je tedy renesancí přístupu, který těží především ze své jednoduchosti, snadné pochopitelnosti a explicitního předvedení výsledku. Výhodou současné vlny zájmu o vizualní zobrazení dat jsou především vyspělé počítačové technologie, software, požadavky na otevřenost úřadů a okamžitý, bezplatný přístup ke statistickým datům ve zpracovatelné podobě a proměna vnímání poslání statistických úřadů.

#### **2.3.1. Teoretické vymezení vizualní reprezentace dat**

Vizualní reprezentace dat je takový typ zobrazení, jenž kromě samotného odvození dat, splňuje také podmínku prostého mapování, kdy zástupný symbol v obraze odpovídá právě jednomu datovému prvku (Vickers, et al., 2013, s. 1050). Systém by měl tedy obsahovat co nejmenší množství redundantních prvků nereprezentujících žádnou informaci, což usnadňuje interakci uživatelů se zobrazením, která by měla probíhat netriviální formou umožňující jeho manipulaci. (Ziemkiewicz & Kosara, 2007, s. 2)<sup>34</sup>. Poslední vlastností, která je nutná pro označení zobrazení jako vizualní reprezentace dat je syntaktická notationalita. Tato podmínka odvozená z Goodmanovy definice notational systému znamená, že v systému skládajícím se ze symbolů a značek, žádný není zaměnitelný s některým z jiných. (Vickers et. al. 2013, s. 1060).

Vizualní reprezentace dat obsahuje také aspekty, které potenciálně pomáhají uživatelům k intuitivnímu nalezení a efektivnímu zobrazení informace (Chen, 2006, s. 22). Strukturální modeling (Shedroff, 1999, s. 267-270) umožňuje detekovat, extrahovat a zjednodušovat skryté vztahy, jež formují strukturu charakterizovanou kolekcí dokumentů nebo jiných datasetů. Díky

---

<sup>33</sup> Označení grafická reprezentace dat i vizualní reprezentace dat jsou pojmy zaměnitelné a běžně využívané v odborné literatuře.

<sup>34</sup> Netriviální interaktivita se vyskytuje ve 2 variantách. Manipulativní umožňuje zasahovat do procesu, který generuje, reprezentaci. Nástroje jsou například posouvání, přibližování nebo otočení. Transformativní umožňuje ovládat procesní fázi a uživatel má vliv na data vstupující do vizualizačního procesu. Příkladem je filtrování podle vlastností.

použití těchto principů je uživatel na základě zobrazení schopen například zodpovědět, jaká je základní struktura komplexní sítě nebo kolekce dokumentů a jaké jsou mentální modely v myslích lidí, kteří vizualizaci vytvářeli. Strukturální modeling je doplněn aspektem grafické prezentace, která převádí výsledek předchozího procesu do vizuálně vnímatelné podoby. Struktura tak může být podrobena vizuálnímu zkoumání integrovaná do systému znalostí uživatele (Ware, 2011, s. 22).

### **2.3.2. Charakteristika vizuální reprezentace dat**

Většinu definic vizuální reprezentace dat lze rozdělit do několika přístupů, které se průběžně vyvíjejí a vzájemně doplňují. Za dnes již poněkud archaický přístup lze považovat charakterizování vizuální reprezentace dat jako reprezentace lidského mentálního procesu. V současnosti je tento model zpravidla nahrazován dvěma paralelními přístupy. V prvním z nich je vizuální reprezentace dat považována za nástroj k zesílení kognitivních procesů v lidském mozku pomocí počítače (Zhu & Chen, 2005, s. 140; Card, 1999, s. 6), tzn. computer-aided seeing (Chen & Floridi, 2013, s. 3422). V druhém pohledu se na vizualizaci nahlíží spíše jako na proces než na pomocnou entitu (Ziemkiewicz & Kosara, 2007, s. 1; Chen & Floridi, 2013, s. 3425-3429; Few, 2007, s. 1). Tento přístup je v současnosti také bližší přístupu statistických úřadů v kontextu současného využívání tohoto způsobu prezentování dat (Smith, 2013, s. 180; Forbes, 2010, s. 1). Používané formy vizuální reprezentace dat jsou zde považovány za proces komunikace, který je založen na kvalitativních či kvantitativních datech. Jejím výsledkem je obraz reprezentující a právě komunikující tato data v čitelné podobě, jenž umožňuje prozkoumání a vyzkoušení (Ziemkiewicz & Kosara, 2007, s. 1; Azam et al., 2013, s. 9). Úspěšné provedení aktu komunikace, se potom hodnotí podle množství informací extrahovaných ze zobrazení (Friendly, 1999, s. 321-322), kdy ideálem je uživateli nabídnout co nejintuitivnější prostředí, ve kterém může extrahovat maximum hodnotných informací za minimálního úsilí a časové investice (Chen, 2006, s. 20)<sup>35</sup>.

### **2.3.3. Sémiotická podstata grafického zobrazení**

Uplatněním přístupů sémiotiky na teorii vizuální zobrazení nebo obraz se zabývala řada filozofů<sup>36</sup>. V moderní sémiotice potom Peirce považoval tuto formu za ikonický typ znaku, tento přístup ale odmítl Nelson Goodman (Schneiderman, 1996, s. 336), který zobrazení

<sup>35</sup> Míru časové investice ovlivňují 3 typy proměnných. - data centrické atributy jako je velikost a uspořádanosti datasetu, human centrické atributy jako je rychlost pochopení a míra interakce nebo podrobnost typ média a prezentace. (Chen et al., 2013, s. 6-7)

<sup>36</sup> Jedním z nich byl například Wilhelm Leibnitz, který navrhl strukturu pro reprezentování významu pomocí znaků, která se sestávala z charakteristiky elementu, kombinace, aplikace a kreativního použití znakových kombinací (Hardt, 2004, s. 19).

představuje jako symbolický systém, systém symbolů, tedy něco co představuje reprezentaci něčeho jiného vyjádřenou nikoli verbálním jazykem, ale vlastnostmi symbolů<sup>37</sup>. Podle Goodmana je si je graf a obraz velmi podobný<sup>38</sup>, obsahuje ale méně vlastností a jevů, které nesou syntaktický význam. Na základě tohoto vysvětlení je tedy možné pracovat s vizuální reprezentací dat, za niž lze graf považovat (Tufte, 1983), jako se symbolickým typem znaku, který reprezentuje, zastupuje něco, co se nachází mimo obraz.

Pro pochopení vztahů v rámci vizuální reprezentace dat a mapování z jedné domény na druhou lze použít sumarizaci pomocí Peircovy sémiotické triády. Tento koncept je založen na triadickém vztahu, který zahrnuje objekt, z pohledu tvůrce set dat, který bude reprezentován, reprezentamen neboli symbol či formu, již znak nabývá, tedy jednu z forem vizuální reprezentace dat (více vysvětluji v kapitole 2.3) a takzvaný interpretant, jímž je v případě vizuální reprezentace dat lidský rozum. (Mai, 2013, s. 679). Vickers (et al. 2013, s. 5-8) potom z pohledu příjemce vizuální reprezentace dat vidí jako objekt set dat, jež je transformováním převeden do vizuálního zobrazení, tedy do své obrazové reprezentace. Interpretantem je rovněž lidský rozum, mysl, v níž se zformuje koncept, jak daný symbol - tedy reprezentaci - interpretovat a jak jednotlivé symboly číst. Jde tedy o zpětné vytvoření si obrazu referenčního objektu – zde datasetu - v mysl. Díky transformování do některé z forem vizuálního zobrazení už ale se zřetelnějším povědomím o vztazích mezi daty. Jak je z obou pohledů zřejmé, je vyžadováno, aby producent a adresát sdíleli kontextuální znalost důležitou pro úspěšné pochopení smyslu, neboť výsledkem je porozumění a získání znalostí nikoli o datech, která ve výsledné rozšíření podobě nejsou ani předmětem zkoumání, ale o systému, o němž se data sbírají a zpracovávají.

### **2.3.4. Vytváření vizuální reprezentace**

Přijmeme-li, že vizuální reprezentace dat je kromě podpůrného nástroje také komunikací, je možné její tvorbu popisovat jako odvození od původního Shannonova komunikačního modelu v jeho matematické teorii komunikace (Chen & Floridi, 2013, s. 3430), jehož úkolem je dopravit co nejrychleji a nejpresněji zprávu od odesilatele k adresátovi. Sestává se ze zdroje informace kódujícího zprávu, jež vysílá signál k příjemci, jež ji dekoduje a doručí na místo určení. Model počítá také s existencí rušivého prvku, kanálu, který je zdrojem informačního šumu (Floridi, 2010, s. 37-48). V případě vizuální reprezentace abstraktních dat a komunikování sémantické informace ve formě zobrazení je proces podobný. Základní úkoly, které je během procesu nutné splnit, jsou rozděleny do 3 hlavních kategorií – získávání informací, analýza informací a

---

<sup>37</sup> Uvádí 2 typy obrazů, ty co referují k něčemu mimo obraz a těm, které jsou tím, k čemu je referováno a nemají žádný předobraz, žádnou referenci v reálném světě

<sup>38</sup> Graf považuje v podstatě také za obraz lišící se jen menším množstvím „nasycenosti“ tedy méně jeho částí hraje sémantickou roli v rozhodování, co znamenají (Ziemkiewicz, Kosara, 2007, 2,3)

rozšiřování informací<sup>39</sup>. Celý proces může být chápán také na základě identifikování potřeb potenciálních uživatelů – do mentálního procesu tak vstupuje rozhodování o formě. Celý proces je potom rozčleněn do 4 kroků tak, že zobrazení předchází samotné analýze, která se děje až u příjemce, přičemž v procesu ji nahrazuje „dávání smyslu datům, tedy vytváření sémantické informace“ (Chen & Floridi, 2013, s. 3424). Proces potom odkazuje k dvěma různým aspektům vizualizačního procesu ve 2 sémantických kontextech. „Prohlížení“ – (Chen & Floridi, 2013, s. 3425) neboli specifikování zaznamenáníhodné informace, vytvoření vhodné vizuální reprezentace a předání, sdělení a představení vizuální reprezentace uživatelům. Tato fáze se soustředí na část vizualizačního procesu zprostředkovaného počítačem, což zahrnuje algoritmy pro filtrování, vizuální mapování a renderování. „Vidění“ (Chen & Floridi, 2013, s. 3422) potom zahrnuje uživatelův myšlenkový proces a kognitivní zkušenost interpretování přijaté informace a konvertování informace do mentálních reprezentací toho, co má informace v úmyslu sdělit. Tento stav se také nazývá „making visible to ones mind“ (Chen & Floridi, 2013, s. 3422). Tyto dva procesy dohromady vytváří tzv. vizualizační pipeline, v níž „cirkuluje informace,“ (Chen & Floridi, 2013, s. 3422-3428).

### 2.3.5. Funkce a cíle vizualizace

Vizuální reprezentace dat se klasifikují podle svého účelu. Friendly (1999, s. 323) rozlišuje především, zda reprezentace vzniká za účelem prezentování nebo analýzy, nicméně je možné, aby naplňovalo obě potřeby. Je-li jeho úkolem prezentování dat, potom je cílem informovat, přesvědčit a podpořit zájem. V opačném případě je zobrazení určeno k exploraci, průzkumu dat a jako podklad pro tvoření modelů.

Funkcí vizuální reprezentace dat je potom především interpretování rozsáhlých množství informací, které umožňuje řešit problémy, jež by bylo za použití jiných metod obtížné řešit. Vizuální zobrazení může fungovat také jako důkaz, poskytnout persvazivní informaci o datech, případně potvrdit či vyvrátit hypotézu, pro jejíž ověření jsou porovnávány různé datasey (Mazza, 2009, s. VIII). Mimo jiné také poskytuje také okamžitou zpětnou vazbu, která dokáže upozornit na chyby v datech i zpracování.

---

<sup>39</sup> Vyhledávání – zahrnuje operace pro objevování datového prostoru – jako vyhledávání, přehled, navigace, zooming.; Analýza - - široké množství analytických úkolů, jako nalezení extrémů, anomálií a clusterů, třídění, filtrování, kombinování a segmentování dat. porovnávání, identifikování korelací a evaluování hypotéz; Šíření - zahrnuje operace vedoucí k prezentování informace, které mají ostatním pomoci pochopit data tyto procesy jsou například sumarizace, anotace, ilustrace.



Za cíl vizuální reprezentace dat je obvykle považováno zprostředkování vhledu do dat a vyjevení datových struktur (Zhu & Chen, 2009, s. 139-140), nicméně Floridi, Chen a Borgo uvádí jako stěžejní a jediný cíl vizuálního zobrazení atraktivního pro uživatele, kteří nejsou odborníky v oboru, úsporu času a zrychlení porozumění při zkoumání problému (2014, s. 6-9). Proti tomuto pragmatickému přístupu z pohledu uživatele, se staví pojetí Stephena Fewa, jenž uvádí jako stěžejní cíl z pohledu tvůrce podnítit zaujetí, informování a vzbuzení zájmu uživatele. Neuspěje-li konkrétní vizuální reprezentace dat v tomto jediném bodě, potom je celý proces jejího vytváření považován za neúspěšný (2007, s. 1-6 ; 2009, s. 7-9).

### **2.3.6. Vizuální reprezentace dat jako nástroj k tvoření informací**

Pojem vizuální reprezentace dat nebo jinak také vizualizace<sup>40</sup> je ve svém v původním významu vzdálen od současného přeneseného pojetí. Původně se jednalo o označení ryze nehmotného konceptu vytváření vnitřních konstrukcí v mozku (Ware, 2011, s. 2), tedy procesu, který se odehrává výhradně v mysli člověka (Ware, 2011, s. 2; Mazza, 2009, s. 65; Zhu & Chen, 2005, s. 171). Tato kognitivní aktivita byla ovšem s rozvojem a zpřístupňování počítačových technologií externalizována a v současné době zastupuje pojem spíše význam grafické reprezentace dat nebo konceptů (Ware, 2011, s. 2) propojujících počítač, lidský mozek prostřednictvím oka, kterému se tak dostane pomoci při identifikování vzorců a extrahování informací na základě vhledu do velkého množství dat (Zhu & Chen, 2005, s. 140). Vizualizace<sup>41</sup> se tak stala polysémantickým konceptem, jehož princip ale zůstává stejný – transformovat data na informace (Mazza, 2009, s. 9). Z tohoto je tedy zřejmé, že se nutně nachází mezi těmito entitami (Mazza, 2009, s. 18).

V současném pojetí je vizuální reprezentování dat obor, který kombinuje postupy z počítačové vědy, designu, statistiky a dataminingu, psychologie a částečně také vizuálního umění (Ziemkiewicz & Kosara, 2011, s. 1).

### **2.3.7. Vizuální zobrazení jako rozhraní pro vnímání informace**

Grafické zobrazení informace využívá na rozdíl od textové prezentace informací nejen konvencí dané přiřazení významu prvkům, ale i intuitivní a vrozené principy odhalování vztahů, které je možné při konstruování vizualizace uplatnit. Respektováním pravidel Gestalt teorie (Mazza, 2009, s. 41-44) o kombinacích barev a tvarů (Healey, 2008, s. 1-7) je možné zabránit vytvoření

---

<sup>40</sup> Pojem vizualizace se běžně využívá jako synonymum pro vizuální reprezentaci dat, v této práci ale referuje k jednomu ze specifických typů zobrazení, jež produkují statistické úřady, proto v tomto kontextu není dále používán.

<sup>41</sup> Zde ve významu vizuální reprezentace dat

vizuální reprezentace, která bude navzdory snaze o informativnost spíše zkreslující a zavádějící, což je problematické, neboť uživatelé nejčastěji interagují právě s grafickým řešením, které by jim záměr autora mělo zprostředkovat a nikoli se záměrem samotným (Ware, 2011, s. 10-12).

Vzhledem k limitům lidského vnímání je vhodné, aby vizuální reprezentace dat představovala zhruba jeden odstavec textu (Tufte, 1983, s. 181). Rizikem jsou v tomto ohledu především rušivé prvky, které nenesou význam. Takový typ prvků se Tufte nazývá chart junk (1983, s. 107-123) a je definován jako poměr mezi množstvím funkčního „inkoustu“, který je nutný k dostatečně názornému zobrazení a množstvím „inkoustu“ který je v zobrazení navíc z estetických důvodů<sup>42</sup>

### **2.3.8. Vizuální zobrazení dat jako řešení information overload**

Všechny předchozí příklady účelů a smyslů, které může vizuální reprezentace dat mít, by se ve své podstatě ale měly potkávat. A to především ve snaze redukovat information overload (Mazza, 2009, s. 105; Shedroff, 1999, s. 271), případně informační úzkost (Marcus, 2007, s. 4), pokud připustíme její existenci. Právě nástroje a povaha vizuální reprezentace dat, jejíž datový bandwidth je větší než pro každý jiný typ médií (Schneidermann, 1996, s. 337), umožňují zvládat tento úkol díky schopnosti modifikovat vstupní data, změnit vizuální mapování<sup>43</sup> a umožnit uživatelům alespoň částečně manipulovat s výsledným vygenerovaným obrazem (Mazza, 2009, s. 8).

Podíl na redukování information overload má i samo vytváření vizualizace (Mazza, 2009, s. 9), čehož statistické úřady využily k vytvoření prostředí určenému k exploratorní analýze agregovaných dat z databank statistických úřadů<sup>44</sup> (více vysvětlují v kapitole 2.3.11.2) Zároveň je ale tento způsob v kontextu informačního přesycení značně problematický (Marcus, 2007, s. 4), především proto, že není možné odhadnout informační a statistickou gramotnost uživatele. Hrozí tedy riziko, že uživatel systému jej nepochopí, nebude schopen správně sestavit výsledný graf a následně neporozumí informaci (Weiner, 1984, s. 138-139).

Úkolem vizuálního zobrazení dat při snaze snížit information overload je vyvarovat se nebo alespoň snížit na možné minimum, takzvané spouštěče (Mai, 2013, s. 682-686) informační úzkosti, za které považuje příliš velké množství informací, jejich nerelevantnost a redundantnost

---

<sup>42</sup> Doporučení, jak se rušení vyvarovat a prezentovat data a informace jasně, precizně a efektivně, nazývá Tufte grafickou excelencí (1983,13-53)

<sup>43</sup> Vizuální mapování je definováno 3 strukturami: Prostorovým substrátem, tedy prostorem ohraničeným osami x a y, grafickými prvky, tedy čímkoli, co se objeví v prostoru a grafickými vlastnostmi vlastnosti jako je velikost, orientace, barva, textura a tvar

<sup>44</sup> Např. ve Veřejná databázi ČSÚ, Statistics Explorer švédského statistického úřadu, aj.

pro danou situaci. Rizikem je také již zmíněná omezená nebo chybějící informační a statistická gramotnost. Tento problém by měla vizualizace kompenzovat svým kvalifikovaným sestavením tak, aby se informace stala na první pohled viditelnou (Tufté, 1983, s. 13-53). Z tohoto důvodu se řada statistických úřadů, které dbají na naplňování svého poslání, uchyluje poskytování dalších typů vizuální reprezentace dat jako jsou vizualizace, infografiky nebo kartografické zobrazení (toto téma je předmětem výzkumné části kapitoly 5)

Takové kvalifikované sestavení nazývá Shedroff informačně interakční design, jehož součástí je informační design, který se zaměřuje na organizaci a prezentaci dat, jehož účelem je vytvořit jasně pochopitelnou komunikaci (1999, s. 267-269), díky které data získají hodnotu<sup>45</sup>. Druhou součástí je interakční design, který se zabývá vytvářením a prezentováním předchozí struktury, přičemž za nejúspěšnější strategii považuje interaktivitu a prezentaci informace pomocí metody storytellingu, která kopíruje osvojený lidský způsob předávání informace pomocí příběhu (Shedroff, 1999, s. 276).

### **2.3.9. Ztráta informací**

Ke ztrátě informací v příběhu vizualizačního procesu dochází téměř vždy. Lze ji rozdělit na záměrnou ztrátu informací, která představuje například agregaci dat, filtrování nebo clusterování, tedy úkony, které probíhají před zpracováním i během interakce s hotovým zobrazením. Nezáměrná ztráta informací potom vychází z limitu přístroje nebo lidského vnímání (Vickers, 2013, s. 12) a dochází k ní zpravidla při malém rozlišení nebo při snaze zabránit přesycení grafu (Few, 2008, s. 1-6), ke kterému dochází při příliš velkém počtu vnesených hodnot v prostoru.

Toto rozdělení lze také použít k výběru vhodných zobrazení. K nezáměrné ztrátě dochází například při zobrazení dat v kartogramu. Záměrná ztráta je potom jednou z charakteristik infografiky.

### **2.3.10. Interaktivita a narativita**

Smyslem interaktivity je zviditelnit veškerou zkušenost, interakci mezi myslí a přístrojem, nosičem, zařízením (Shedroff, 1999, s. 276). Míra použité interaktivity v zobrazení znamená zároveň větší možnost kontroly pohybu a sofistikované navigace uživatelů. Zároveň jde o nástroj umožňující lépe budovat zkušenost a lépe a rychleji si tak osvojit informaci vyplývající

---

<sup>45</sup> Jiní autoři nazývají tuto fázi modelování či informační processing

ze zobrazení. Neznamena to ale zároveň, že je zobrazení více nasyceno informacemi. Princip využití interaktivity prezentuje vizualizační mantra (Schneidermann, 1996, s. 336-337), jejíž následování umožňuje získat rychle a přehledně znalost o konkrétní reprezentaci.

S interaktivitou úzce souvisí metoda storytellingu (Few, 2009, s. 1-10), která vychází z teorie o vizuálním zobrazení jako komunikaci, při níž lze uplatnit stejná pravidla, jako platí při sdělování příběhů jinými formami komunikace v podobě, v jaké dává logický smysl. Pro úspěšné sdělení a pochopení informace je tedy vhodné dodržovat například Griceovy konverzační maximy (Mai, 2013, s. 676-685), tedy principy slušné a smysluplné konverzace. Aby komunikace fungovala, je nutné sdělit zprávu jasně a stručně a ve správném, nemanipulujícím znění a v kontextu, který ukáže případné kauzální vztahy. Zároveň musí obsahovat prvek překvapení neočekávanou konkrétní pointu či nějakou novou informaci, která vyvolá emoce.

### **2.3.11. Specifické typy vizualizací užívané statistickými úřady k prezentaci dat**

#### **2.3.11.1. Infografika**

Jako infografiku lze označit statické grafické zobrazení, obraz, které zobrazuje množství informací vycházející z určitého setu dat. Kombinuje prvky grafického designu, grafů, textu a čísel s cílem komunikovat informace přitažlivou v jednoduše interpretovatelné formě pro publikum, které disponuje pouze nízkým statistickým vzděláním a předpokládá se u něj také malá statistická gramotnost (Few, 2007, s. 2). Infografika umožňuje zjednodušit komplexní informace, ale tato transformace má za následek vynechání některých vztahů, což následně redukuje statistický problém a může být i limitující pro pochopení konkrétní situace. Taková situace nastává v případě čísel, která jsou uváděna jen jako jednotlivá fakta postrádající dodatečný text a kontext (Albers, 2015, s. 268-270).

Infografiky lze v závislosti na složitosti a způsobu, jakým jsou informace prezentované, rozdělit do 4 kategorií (Albers, 2015, s. 268-269). Nejjednodušší formou je graficky zpracovaný ekvivalent zarážkového listu, složitější formou je potom fotografie nebo centrální obrázek, k němuž jsou vedené příslušné textové poznámky často obsahující číselné údaje. Tyto dva typy se v prezentacích statistických úřadů nacházejí pouze zřídka (Albers, 2015, s. 268-269).

Častější jsou v případě užívání statistickými úřady obě komplexnější formy – tzv. „plochá informace“, tedy sled grafů, obrazů či čísel souvisejících s jedním tématem, které ukazují danou tematiku z různých perspektiv. Tomuto typu infografiky chybí předem dané pořadí čtení

jednotlivých komponent, ale podporuje srovnávání různých vztahů a dat (Albers, 2015, s. 268-270). Poslední typ se nazývá informační proud (Albers, 2015, s. 268-270) a naznačuje určitý proces, vývoj, který je nutný pro správné pochopení informace. Je zde předem daný sled prvků, který je potřebné následovat. Oba tyto typy grafik produkují například statistické úřady v Německu (Destatis.de, 2013), Itálii (Istat.it, 2015) či USA (US Census, 2012)

Ačkoli je infografika vnímaná jako vhodný nástroj pro pochopení komplexních dat, nejde vždy o jednoznačný fakt. Na základě uživatelského testování zadané Department of Commerce při US Census Bureau vyšlo najevo (Gareau et al., 2013, s. 161-172), že současné komplexní infografiky jsou sice lépe srozumitelné než výhradně textový popis ekvivalentního problému, ale spíše pro uživatele s vyšším než středoškolským vzděláním, kteří uvedli, že neměli problémy s orientací v dokumentu. Jako jeden z problémů se v tomto případě ukázal i příliš odborný slovník.

Uživatelé od infografiky očekávají informace vztažené k reálnému světu (Gareau et al., 2013, s. 171). Vzhledem k značně šumnému prostředí, které z podstaty infografiky obsahuje značné množství redundantních grafických prvků, je důležité dostatečně zdůrazňovat stěžejní informace, neboť pro uživatele je snadné důležitou informaci přehlédnout.

#### **2.3.11.2.        Prostředí na explorační analýzu**

Prostředí na explorační analýzu umožňuje uživatelům zkoumat data pomocí předdefinovaných grafických nástrojů a vytvářet tak vlastní grafické zobrazení z agregovaných dat dostupných databázi, která je v případě statistických úřadů vždy aplikace. Prostředí je navrženo tak, aby bylo možné jednoduchými kroky vykreslit grafy či kartografické zobrazení. Účelem tohoto prostředí je umožnit prozkoumání povahy a vztahů dat v rámci vybraného datasetu či korelací mezi větším množstvím datových souborů. Zobrazení dat, které je výstupním produktem tohoto nástroje ale nemá prezentační, nýbrž analytickou funkci (Leinhardt & Wassermann, 1979, s. 311). V postupu konstruování grafu jde pro uživatele o opačný princip vnímání než u ostatních forem vizuálního zobrazení, neboť uživatel začíná interakci s daty a reprezentací a pokračuje odhalováním vzorců a informací ukrytých v datech (Good, 1983, s. 291). Konečným cílem není jako v případě vizualizace ohromit, přesvědčit, ale porozumět datům, která porovnává a nahlíží na ně z více perspektiv. Estetická stránka v tomto případě není prioritou a stěžejní je přehlednost a přístup k potřebným statistickým datům a statistickým funkcím a nástrojům (Good, 1983, s. 291).

Prostředí aplikací na explorační analýzu dat využívaných statistickými úřady vychází z principu exploratory data analysis představené J. W. Tukeym v roce 1977 (Tukey, 1977). Jedná se o techniku bez přesných pravidel, již představuje pouze set nástrojů jejichž použitím uživatel zkoumá nejpravděpodobnější významy a vzorce v datech (Good, 1983, s. 290). Základní aktivitou je opakovaná interakce s daty a zkoušení různých modelů (Beniger, 1978, s. 1-2). Cílem je potom vyjevení významných rysů, struktur, vzorců a outlierů. Tato metoda umožňuje organizovat a sumarizovat data, a to s absencí přímé teorie a tradičních statistických metod. Uživatel do prostředí a k datům může přicházet s určitou představou o formulování hypotézy (Good, 1983, s. 291), výsledek je ale nepředpokládatelný a absence takové předpřípravy není překážkou (Leinhardt & Wasserman, 1979, s.310-314). I v tomto prostředí interagují statistická data a grafické zobrazení, je proto možné uplatnit principy Visual seeking Mantra.

Implementování aplikací na explorační analýzu statistických dat je značně rozšířené. Na svých webových stránkách jej poskytuje řada statistických úřadů po celém světě (více vysvětlují ve výkumné části 5.2) Přestože je již nyní toto prostředí pro uživatele velmi intuitivní – zejména v produktech Development Bank of Africa (Dataportal DAB, 2013), Statistisches Bundesbüro (Genesis,2015) nebo Statistika Centalbyrån (Statistics Explorer,2015), obsahuje pro uživatele bez statistického vzdělání či s omezenou statistickou gramotností stále řadu limitů a rizik. Jedná se především o velké množství dostupných možností explorační, jejichž vhodnost nedokáže nedostatečně vzdělaný uživatel posoudit (Beniger, 1978, s. 1-2), dalším rizikem je například výběr nesrovnatelných datasetů či takových dat, u nichž korelace nedává smysl.

### **2.3.11.3. Vizualizace**

Vizuální zobrazení dat je možné kategorizovat a třídit na základě řady faktorů. Například na základě vstupních dat, informativnosti výstupního obrazu či úkolu (Chen, Floridi, 2013, s. 3425). Právě na základě vstupních dat (Card, 1999, s. 19) rozlišuje Card vizualizaci na 3 základní typy – vědeckou vizualizaci, softwarovou vizualizaci a informační, potažmo datovou, vizualizaci.

Vědecká vizualizace je vizualizací vlastností prostoty fyzických objektů a umožňuje pochopit fyzické jevy uložené v datech velkých datasetů. Data přichází z různých senzorů, satelitů, medicínských skenů nebo teleskopů. Analyzovaná data se týkají výhradně přírodních věd, lidského těla, buněk, DNA, molekuly, tras letadel aj. Datasetsy obvykle obsahují velké množství proměnných, k jejichž zobrazení se používají grafy pro multivariantní datovou analýzu, jako jsou glyphs, 3D zobrazení nebo stromy (Card et al., 1999, s. 15).

Informační a softwarová vizualizace jsou potom určeny k vizuálnímu znázornění abstraktních dat, tedy dat nefyzické povahy, často s původem ve fyzickém světě. Softwarová vizualizace umožňuje pochopení a efektivního používání počítačového softwaru, využívá se především pro vzdělávání (Zhu, Chen, 2005, s. 140).

Informační, datová, vizualizace umožňuje identifikování vzorců, korelací nebo vzorců nalézajících se v datasetech. Vstupní data mohou být jak ve strukturované, tak i nestrukturované formě. Vizualizace tohoto typu se využívá ke grafické reprezentaci, vyjevení vzorců ukrytých v datech a k obohacení o poznání o uložených datech. Vizualizace významně usnadňuje porozumění vlastnostem nebo strukturám v rozsáhlých množstvích dat. Současným trendem je například interakce mezi tímto typem vizualizací a dataminingovými algoritmy (Zhu, Chen, 2005, s. 145-146).

Vizualizace se pro účely zobrazování statistických dat používají například v případě zobrazení rozsáhlých datových setů referujících o změnách v čase. Pomocí vizualizace je možné zobrazit rozsáhlé sety dat přehlcení uživatele, neboť dochází k silné informační redukci. V současné době se dává přednost především interaktivní a animované vizualizaci časových řad před dříve statickým zobrazením tzv. small multiples (Tufté, 1983, s. 161-177).

Jedním z tradičně využívaného typu vizualizace je prostředí nástroje Gapminder. Tento nástroj umožňuje animovat časový vývoj rozsáhlých setů dat týkajících se zejména socioekonomických ukazatelů v jednotlivých zemích či v rámci jedné země. Dalším, pro statistické úřady typickým příkladem, je dynamická populační pyramida animující nárůst či propad populace, a to včetně trendů v určité generaci či predikce do budoucnosti (Destatis, 2015; *National Population Projection, 2014*; *Population projection, 2015*).

#### **2.3.11.4. Kartografické zobrazení**

Geografické zobrazení tvoří zvláštní případ vizualizace, která využívá kartografy vytvořených map a standardně se nachází v kartesiánské soustavě. Zobrazení geografické pozice je proto pro laiky intuitivní a přirozeně vychází ze získaných znalostí zobrazení světa a určitých území (Hallisey, 2005, s. 359). Kartografická zobrazení jsou využívána zejména k redukování informace a díky čemuž umožňuje pochopit a prezentovat rozsáhlé a komplexní datasety.

Uživatelům umožňuje rychlou orientaci<sup>46</sup>, neboť zobrazované informace jsou organizované a pasují na geografickou strukturu. Psychologickou výhodou při vnímání kartografické vizualizace je značná důvěřivost a nekritický přístup k takovým typům zobrazení (Hallisey, 2005, s. 358-361). Pro dobrou orientaci v tomto typu zobrazení je nezbytná znalost příslušné konvence<sup>47</sup> a funkčně vytvořená legenda (Hallisey, 2005, s. 358-361).

Geovizualizační komunikační model je v porovnání s ostatními typy vizuálních zobrazení dat odlišný a místo kódéru a dekodéru obsahuje kartografovu koncepci, která je transformována do mapy (Hallisey, 2005, s. 352). Cílem tohoto typu modelu je identifikovat kauzální mechanismy a struktury, které vedou k jevům ve specifických podmínkách a porozumět prostorovým informacím skrz interaktivní nebo neinteraktivní vizuální zobrazení (Dykes, 1998, s. 485-486). Použitím geografické vizualizace je možné dospět k informacím o umístění prvku, o jeho okolí, určit lokální anomálie a změnu. Kartografická vizualizace slouží také jako iterační proces umožňující pozorování a srovnání se znalostmi uživatele (Hallisey, 2005, s. 353).

Statistické úřady využívají kartografické zobrazení ve 3 formách. Především jde o vytvoření nástrojů pro interakci uživatelů s daty v intuitivním prostředí (Forbes, 2010, s. 1), například v analytickém nástroji Statistics of Sweden (Statistics Explorer, 2015) či interaktivní mapový nástroj Commuterview Statistic Office of New Zealand (Forbes, 2010, s. 4). Oba tyto nástroje využívají geovizualizace pro redukování dat. Druhým směrem je potom vizualizování prostorových dat při využívání externích nástrojů, jako je statistický balík R umožňující mapování a geografické analýzy, GoogleMaps a GoogleEarth, které obsahují několik volně dostupných otevřených GIS systémů, na jejichž bázi úřady vytvářejí vlastní aplikace. Poslední, nejméně interaktivní formou kartografických zobrazení jsou vlastní statistické atlasy.

### **3. Vizualizace v kontextu statistických úřadů**

#### **3.1. Poslání statistických úřadů jako objasnění důvodu užívání vizualizace**

Poslání statistických úřadů prošlo během posledních 50 let výrazným vývojem a postupně bylo rozšiřováno podle potřeb, které se odvíjely od nastavení společnosti i politicko-ekonomické situace. Koncem 90. let kromě národních zákonů vznikly i nadnárodní dokumenty, ustanovení a dohody, jejichž úkolem je hlídat nikoli obsahovou náplň, ale především procedurální postupy a etické zásady (European Statistical System Committee, 2011, s. 1-8). Posledním stupněm

---

<sup>46</sup> Rizikem této přístupnosti je například fakt, že kartografické zobrazení ale může způsobit i zkratkovité vnímání, například iluze

<sup>47</sup> Zobrazování v kartézské soustavě, měřítko, význam barev a šíře linií, aj.



definování poslání je vytváření unikátních strategií a vlastních misí. Ty se navzdory respektování stejných mezinárodních dohod a kodexů v jednotlivých případech liší a odrážejí to, co je prioritou konkrétního úřadu. Plnění základní požadavků přitom zůstává zachováno.

### **3.1.1. Vytvářet statistiky**

Mise jednotlivých statistických úřadů jsou rozdílné a jejich konkrétní podobu služby si určuje konkrétně každý úřad a stát sám, ač se i díky existenci mezinárodních dokumentů jako je Kodex evropské statistiky (European Statistical system Committee, 2011) nebo Základní principy oficiální statistiky (United Nations, 2014, s. 1-2) značně podobají. Obecně je ale možné říci, že základním a prvním posláním všech statistických úřadů je a byl sběr dat a vytváření statistik (Leonard, 1952, s. 14).

Zákony, samoregulační dokumenty, kodexy a diskuse odborníků přesně definovaly podmínky, postupy a zásady, jichž musí sběr a zpracování dat dostát (více vysvětlují v kapitole 2.1.3.2) Statistické úřady mají za povinnost v rámci vytváření statistik informovat o demografickém, socioekonomickém a ekologickém vývoji země, což je zároveň jeden z požadavků v dokumentu zpracovaném UN, který definuje základní principy statistické služby (United Nations, 2014, s. 1-2).

Českému statistickému úřadu ukládá stát za povinnost vytvářet statistické informace (Česká republika, 1995, s. 1-26). Statistickou informací se rozumí agregovaný set individuálních údajů týkajících se jednotlivé právnické nebo fyzické osoby, získaný pro statistické účely. Posláním Estonského statistického úřadu je vytváření „oficiálních statistik“ popsanych jako kvalitativní a kvantitativní agregované a reprezentativní informace o společenských jevech, které jsou výsledkem statistického procesování dat (Estonsko, 2011). Singapurský úřad se zajímá výhradně o ekonomickou prosperitu a změny v populaci. Bez bližšího upřesnění má produkovat co nejkvalitnější statistiky o těchto segmentech (Singapur, 2015, s. 1-2). S požadavkem na maximální kvalitu produkovaných statistik souvisí také zajištění nezávislosti úřadu. Je například nepřijatelné, aby statistický úřad přijímal data od politických institucí (Beranová, B. 2011, s. 4-7). Jen tak mohou statistické úřady plně přijmout odpovědnost za poskytované statistiky a splnit poslání, které žádá, aby produkovaly včasné, objektivní a důvěryhodné statistiky (Český statistický úřad. 2011, s. 1). Jen tak mohou vytvářet důvěryhodný obraz o stavu a vývoji společnosti (Český statistický úřad. 2011, s. 1; Statistics Estonia, 2015).

Uplatňování těchto principů bylo přínosné zejména během finanční krize. Například v případě Řecka jsou důvěryhodná primární data pro Eurostat zásadní a Evropská komise se zaměřuje na

jejich hlídání i vzhledem k poskytování úvěrů této zemi (Gébllová, 2012, s. 5). K částečnému selhání naplňování požadavku důvěryhodnosti v průběhu 80. let v Británii (McLennan, 1995, s. 470) a v první polovině 90. let v České republice. V době privatizace, liberalizace ekonomiky a přerodu na jiný typ obchodování bylo nutné změnit také metodologii statistických zjišťování a v případě obchodních statistik se v letech 1990-1994 nepodařilo vytvořit žádná data. Došlo ke ztrátě zjišťovaných subjektů a periodické šetření se vrátilo až v 2. polovině 90. let (Mašátová, 2012, s. 10).

### **3.1.2. Integrita vs. Včasnost**

Součástí poslání vytváření výše uvedených statistik je také balancovat prakticky neustálé dilema mezi integritou a včasností publikovaných statistik. Neboť jejich úkolem je vytvářet komplexní set sjednocených statistik (definičně, koncepčně, s harmonizovaným názvoslovím a příslušným označením). Pro uživatele mají zásadní význam oba faktory. Je důležité mít jak dostatečně čerstvé údaje, tak i konzistentní statistiky (McLennan, 2013, s. 1-28).

Integrita ve statistickém pojetí znamená pochopení skutečnosti, tedy publikování přesných a objektivních statistik ve stanoveném termínu za použití standardizované metodologie. Integritu úřady ukazují přijetím odpovědnosti za obsah publikovaných statistických výstupů (McLennan, 1995, s. 467) například publikováním na webu nebo vytvořením softwaru unifikující data podle standardů ESU a EU (Estonian Statistics, 2008). K odpovědnějšímu naplňování požadavku integrity a včasnosti statistik přispěla finanční krize v roce 2008, kdy v rámci stabilizačních procesů statistické úřady publikovaly v krátkých intervalech údaje o změnách zásadních ekonomických indikátorů. (Bohatá, 2011, s. 5-10).

### **3.1.3. Snižování zátěže**

Jako významnou součást strategie, vize a mise následujících let je pro statistické úřady snižování zátěže respondentů. Úřady se samy motivují k opětovnému využívání již jednou publikovaných dat a snaží se využívat administrativní data pořízená ostatními úřady za účelem evidence. Není tak například již v zájmu ČSÚ opakovat znovu obchodní census uskutečněný jako první pokus o získání dat o počtu provozoven v roce 1999 (Mašátová, 2012, s. 10), neboť tyto údaje je možné získat z jiných databází. ČSÚ se na tento úkol zaměřuje od roku 2008 (Trendová, 2011, s. 25-27) a pro respondenty připravil webový průvodce vyplňováním výkazů a základní FAQ na [www.vykazy.cz](http://www.vykazy.cz). Základem naplnění této části mise je zabránění duplikování informací a využívání existujících zdrojů dat (European Statistical system Committee, 2011). ČSÚ konkrétní podobu zefektivnění služby, které vedlo ke snížení administrativní zátěže u

podnikatelů, prezentoval v návaznosti na požadavky vlády.<sup>48</sup> Se snižováním zátěže souvisí také snižování finanční náročnosti statistik a tedy zlevnění fungování úřadu.

### **3.1.4. Rozvíjet obor a být autoritou**

Důležitou součástí mise statistických úřadů je fungování jako odborná autorita. Naplňovat toto poslání znamená konstantní snahu o zlepšování a zrychlení výsledků, metod a měření (Moser, 1976, s. 60) I z tohoto důvodu existuje úřad jako samostatná entita, aby mohl dosáhnout odborného respektu. Nezávislosti instituce je zásadní nejen pro fungování v rámci státu, ale také v mezinárodním prostředí. S ohrožením tohoto poslání se v minulém roce potýkal ČSÚ, kdy statističtí odborníci apelovali na politiky s vyjmutím ČSÚ z působnosti služebního zákona (Ritschelová, 2014, s. 15-18)

Počátkem 90. let měl na redefinování této části mise statistického úřadu v Británii vliv thatcherismus poté, co po velkém personálním zúžení a omezení rozpočtu v 80. letech nebyl schopen Central Statistical Office (CSO) naplňovat své poslání (Moore, 1992, s. 23). Byla vyhrazena exkluzivita poskytování některých služeb výhradně pro veřejný sektor, a to mimo jiné znamenalo i deklarování monopolu pro Central Statistical Office (CSO) v poskytování oficiálních statistik. Tento monopol na produkování dat mají v současnosti zajištěn statistické úřady i v ostatních zemích, neboť obvykle není žádná alternativa, na kterou by se veřejnost mohla s důvěrou obrátit (Bohatá, M. 2011, s. 5-10).

Naplňování pozice autority v oblasti rozvoje oboru a metodologie zavazuje to k publikování nových metod, které akademická obec hodnotí a podrobuje kritice a využívá je pro vyučování oboru. Smyslem těchto studií je zlepšit stávající metody a upravit studijní náplň podle nejmodernějších přístupů v praxi (Smadar, 2011, s. 130) ČSÚ tuto podmínku naplňuje jak zavedením takzvaných Peer Reviews, tak dlouhodobým publikováním recenzovaného a neimpaktovaného časopisu Statistika<sup>49</sup> a spoluprací se Sociologickým ústavem (Mana, 2014, s. 7).<sup>50</sup>

#### **3.1.4.1. Služba pro veřejnost a služba státní správě**

---

<sup>48</sup> Dne 29. srpna 2012 k Záměru zefektivnění výkonu státní statistické služby a Návrhu úsporných opatření

<sup>49</sup> Určeného na rozdíl od časopisu Statistika a my pro odbornou obec

<sup>50</sup> Souží k archivaci a zpřístupnění dat, vznikl za účelem podpory sekundární analýzy dat (informace, technologie, metody, výuka, výzkumná činnost – metodologie, kvalita a harmonizace dat, přehledové studie, obsahuje anonymizovaná data. Spolupráce také ve výuce s UTB ve Zlíně.

Jak už z názvů zákonů a kodexů týkajících se fungování a mise statistických úřadů vyplývá, statistické úřady provozují službu. Její příjemci jsou zpravidla dvě skupiny - státní správa a ostatní veřejnost. Orientace na veřejnost je základním cílem pouze posledních 25 let. Na přelomu 80. a 90. let proběhla v Británii rozsáhlá diskuse, která měla za následek kompletní změnu tehdy platného paradigmatu Reynerovy doktríny (Laux, 2013, s. 2). Svůj vliv na změny sehrál také pád železné opony (United Nations, 2014, s. 1-2). V současné době je kladen velký tlak na maximální rozšiřování služeb směrem k veřejnosti, demokratizaci statistik a zvyšování uživatelského komfortu.

#### **3.1.4.1.1. Služba státní správě**

Do konce 80. let se statistické úřady věnovaly téměř výhradně poslání sloužit státu. Je to zjevné například z předrevolučních zákonů o statistice<sup>51</sup>, kde úkolem bylo především hlídat naplňování plánu řízeného hospodářství. Až do roku 1992 nijak nezmiňují jakoukoli možnost veřejnosti žádat o přístup ke statistickým datům. Teprve dva porevoluční zákony<sup>52</sup> upravily službu úřadu nejen vůči státní správě, ale i veřejnosti. V současnosti je Český statistický úřad odpovědný za poskytování dat vládě, státní správě a resortům, kde musí být imunní vůči politickým tlakům. Pro české prostředí platí, že také koordinuje a vede přehled statistické služby, kterou vykonávají ministerstva.

Ve Velké Británii má zajišťování statistik velkou tradici v tzv. decentralizovaném systému působnosti v rámci ministerstev. Teprve v průběhu 50. let došlo k alespoň formálnímu sjednocení vzájemně nezávislých statistických oddělení pod instituci nazvanou Government Statistical System (GSS). Tato oddělení měla za úkol poskytovat statistiky a analýzy dat. Toto přidružení k resortům vzniklo z důvodu asistence správě a vládě při zjišťování dat a interpretacích výsledků i nastavování „správných otázek“. Statistická zjišťování tak mohla vzniknout na politickou objednávku (Moore, 1992, s. 15-18).

Snížení stavů i rozpočtu Centrálního statistického úřadu i GSS v 70. letech vedlo k omezení aktivity a zpracovávání statistik a dalo počátkem 80. let vzniknout dokumentu, který se nazývá Reynerova doktrína. Ta tvrdila, že je nutné rozšířit podporu pro GSS na úkor druhé větve statistického úřadu sloužící k sběru ostatních statistik, které nebyly v působnosti ministerstev a omezit přístup veřejnosti k datům, analýzám a výsledkům statistických zjišťování (Alldritt, Young, Laux, 2008, s. 1-14). Tento stav postupně vedl k nedůvěře v úřad a tvorbu nekvalitních statistik. Nakonec se počátkem 90. let přistoupilo ke zrušení dokumentu. V roce 1999 jej nahradil White Paper Building Trust in Statistics definující požadavky na nezávislost instituce a

---

<sup>51</sup> 99/1961 Sb, 1/1967, Zákon 1/1969 Sb.

<sup>52</sup> 287/1992 sb., 89/1995 Sb.

v roce 2002 Etický kodex statistického úřadu, kterým byl dokončen návrat k tezi, že národní zájem není data tajit, ale naopak zveřejnit (Moser, 1976, s. 59-63).

#### **3.1.4.1.2. Služba veřejnosti**

Služba veřejnosti je prakticky nejmladší součástí poslání statistických úřadů. O poskytování dat veřejnosti se poprvé začalo uvažovat v 70. letech, tyto snahy ale ukončilo zavedení Reynerovy doktríny (více vysvětlují v kapitole 3.1.4.1.1). Podruhé se do iniciování většího otevření úřadu vložila také OSN a společně s britským statistickým úřadem v roce 1992 sestavili „Základní principy oficiální statistiky“. Deset základních požadavků, které by měl naplňovat statistický úřad demokratické země, vznikl v souvislosti s rozpadem východního bloku a potřebou redefinice úkolů statistických úřadů v postsocialistických zemích (United Nations. 2014, s. 1-2). Kromě jiných požadavků žádá otevřenost úřadu a důvěryhodnost dat.

*V zákoně o státní statistické službě je o službě k veřejnosti uvedena jediná věta: „(ČSÚ) vydává, zveřejňuje a poskytuje statistické informace, včetně metodiky, informuje veřejnost o sociálním, ekonomickém, demografickém a ekologickém vývoji České republiky a jejich jednotlivých částí“ (Česká republika. 1995, s. 4). Ve strategii k naplňování mise vytvořenou přímo úřadem, se uvádí, že ČSÚ se snaží „maximalizovat užitek a komfort uživatelů. Principy jsou dostupnost, standardizace, operativnost, dialog, důvěrnost, přidaná hodnota, cílená komunikace, personifikace.“ (Český statistický úřad. 2011, s. 1)*

Se strohými a krátkými konstatováním souzní také mnohem rozsáhlejší a podrobnější strategie naplňování mise služby veřejnosti statistického úřadu v Estonsku (Estonian Statistics, 2015). Veřejnosti mimo výše uvedeného garantuje spolehlivost, objektivnost a transparentnost nebo všeobecnou dostupnost statistik. Vedle závazků jako vysvětlování metodiky si klade cíle jako je reakce na připomínky, rozvoj moderních způsobů šíření statistik a dat, srozumitelné analýzy a rozvoj nových systémů. Zavazuje se i k pořádání otevřeného panelu s veřejností, který má zajistit maximální respektování uživatelských potřeb.

Důležitým krokem k větší odpovědnosti vůči veřejnosti a k demokratizaci přístupu ke statistikám a analýzám dat měl a má rozvoj ICT, internetu a osobních počítačů. Ty umožňují jednodušší dostupnost dat a zjednodušují jejich nabídku, také ale zvyšují očekávání uživatelů směrem k rychlosti a velikosti nabídky dat. V důsledku těchto požadavků dal Eurostat k dispozici svůj současný internetový servis v únoru 2000. Tzn. Green Paper o informacích ve veřejném sektoru v informační společnosti (Alldritt, Young, Laux, 2008, s. 1-14) potvrdil již nastolený trend otevřenosti úřadů a mimo jiné ukládá například úřadům závazek poskytovat svá data na stránkách zdarma. K tomuto dokumentu se ČR připojila po vstupu do EU.

S demokratizačním procesem a zlepšováním naplňování poslání vůči veřejnosti souvisí také důvěra v tuto službu, zejména v případě, pokud jde o monopol, u něž není dostupná alternativa. Pro minimalizování rizika nedůvěry ve veřejný sektor byla po letech platnosti Reynerovy doktríny v roce 1991 zavedena ve Velké Británii tzv. Citizen's Charter (Laux, 2013, s. 5). Konkrétní závazky vedly ke zlepšení fungování veřejného sektoru a staly se později základem otevřeného přístupu statistické služby i v dalších zemích Evropy (Laux, 2013, s. 3-5). Pro posílení důvěry ve služby ONS byl roku 2002 navíc zaveden National Statistic Code of Practise (UK Statistics Authority, 2009).

V kontextu služby veřejnosti je také důležité zavedení tzn. information managementu, jehož principem je využití webu jako platformy pro poskytování statistik uživatelům. Na stránkách se následně odehrávají veškeré služby od výběru dat po jejich analýzu a poradenství odborníků. V současnosti je trendem navíc poskytování dat ke stažení a metadat (Cook, 2003, s. 1-4).<sup>53</sup>

Odpovědnost k veřejnosti, kterou má statistický úřad, lze rozdělit na různé "menší" odpovědnosti k různým typům veřejnosti, které se vzájemně prolínají. Veřejnost, která vyplňuje formuláře má právo na jejich nenáročnost, ochranu osobních údajů a anonymizaci. Médii by měla být zajištěna podpora při interpretaci dat, akademická obec má právo na mikrodata a recenzování vydaných nových metodologií a technik. Celá veřejnost má potom právo vědět o využití veřejných financí (McLennan, B. 2013). Laická veřejnost dále očekává jednoduchý, ale vysvětlující komentář, analýzu nebo interpretaci, která je nutná k pochopení publikovaných dat.

Tyto požadavky naplňuje v případě ČSÚ například Veřejná databáze, Informační systém nebo prezentování dat pomocí geostatistik (Klauda, P. 2011, s. 107-111). V případě jiných úřadů jde zpravidla o (ne)interaktivní vizuální reprezentace dat a webové nástroje na jednoduché i náročné zpracování a vizualizování dat, jejichž přiblížení je smyslem této práce.

### **3.1.5. Vzdělávání**

Součástí služby veřejnosti je také vzdělávání a pozvedávání statistické gramotnosti obyvatel. To je závaznou součástí Britského etického kodexu a centrální statistický úřad má za povinnost publikovat statistické materiály určené pro vzdělávání na středních školách (Cook, L. 2003, s. 1-4). Tato část mise je taktéž vedena v závazných mezinárodních dokumentech<sup>54</sup> a součástí mise ČSÚ, zákonně závazné pro statistický úřad ČR tato služba veřejnosti ale není, ačkoli si sami statistici stěžují na nízkou úroveň statistické gramotnosti obyvatelstva (Víchová, J. 2011, s. 24-28), přesto, že úřad aktivně vytváří popularizační projekty jako je Minisčítání či soutěž o nejlepší statistický plakát.

<sup>53</sup> Například formou poskytování různých forem výstupů nebo API (Koláčková, 2011, s. 44-45)

<sup>54</sup> European Statistic Code of Practice (European Statistical system Committee, 2011) a Fundamental Principles of Official Statistics (United Nations, 2014)

### 3.1.6. Závěr

Poslání statistických úřadů je obecně služba. Nejčastěji se rozděluje na službu veřejnosti a službu státní správě. Vzniká tak neustálý konflikt mezi požadavky, které statistický úřad musí naplňovat (Moser, C. 1976, s. 59-76).

Díky rozvoji výpočetní techniky a internetu v současné době dochází k vylepšování komfortu uživatelů v dostupnosti a interpretaci dat. Úřady se snaží nabídnout maximální otevřenost služeb, důvěryhodnost a spolehlivost dat. Úkolem úřadu není jen produkovat data, ale také působit jako autorita. To znamená nezávislost, působení jako autorita a usilování o zvyšování statistické gramotnosti ve společnosti. K naplňování uvedených poslání může pozitivně přispět jako nástroj také vizuální reprezentace dat.

### 3.2. Historie využívání vizuální reprezentace dat pro potřeby a prezentaci statistických úřadů

Využívání vizuální prezentace dat statistickými úřady není novým fenoménem. Toto spojení bylo velmi intenzivní již v 19. století, zejména pak v jeho druhé polovině<sup>55</sup>, kdy došlo k dostatečnému rozvoji grafických metod<sup>56</sup>, způsobů prezentace<sup>57</sup> a k rozvoji institucionalizovaného systému sběru statistických dat. V tomto období, které se nazývá „zlatý věk statistické grafiky“ (Friendly, 2008, s. 503) vznikaly rozsáhlé vizualizace a soubory, pro které bylo typické velké množství neustálých inovací, jež později podpořilo i rozvoj reprodukčních technologií jako (barevná) litografie a offsetový tisk. Tehdejšímu rozsahu použití vizuálního zobrazení nedokáže současné využití zdaleka konkurovat (Friendly, 2008, s. 503), i pro to, že v tomto období plnila jinou funkci než v současném pojetí (více vysvětlují v kapitole 3.4) a představovala stěžejní způsob hloubkové analýzy dat, neboť ještě nebyly vynalezeny moderní statistické metody používané ve 20. století.

Statistická grafika byla doménou zejména francouzských statistiků, kteří vytvořili například korelaci sledující míru vzdělanosti obyvatel, bydliště a míru kriminality, díky kartografickému zobrazení dokázali odhalit velké rozdíly ve školní docházce žáků na severu a jihu země. V Británii byl přesto tento způsob analýzy oficiálně schválen Royal Statistical Society pro využití UK Statistical Office až roku 1885<sup>58</sup> (Friendly, 2008, s. 521-523).

<sup>55</sup> V tomto období vznikla například Minardova grafika tažení Napoleonovy armády do Ruska (1869)

<sup>56</sup> Jako je vynález koláčového a sloupcového grafu, scatterplotu, aj. Williamem Playferem

<sup>57</sup> Například kartografické zobrazení

<sup>58</sup> Již předtím ale došlo k úspěšnému využití například při mapování zdroje cholery v Londýně lékařem Johnem Snowem v roce 1854.

### 3.3. Rozvoj principů statistického myšlení

V průběhu 2. poloviny 19. století došlo také k rozvoji statistického myšlení, které umožnilo statistiku dále inovovat. Jednou z největších změn, které měly ve výsledku vliv i na způsoby zobrazování, byla transformace „zákonu frekvence chyb na Supreme law of unreason,“ (Friendly, 2008, s. 506) díky čemuž mohl vzniknout koncept „průměrného člověka“ (Friendly, 2008, s. 506). Tato centrální hodnota, průměr, se následně stala předmětem primárního zájmu. V roce 1878 vznikla také první publikace, která se zabývala vytvářením grafů, La Methode Graphique. Autor Etienne-Jules Marey v ní popsal dosud známé teorie, objevy a příklady tehdy známých analytiků Minarda, Playfaira nebo Nightingale. Vznik některých metod ovlivnily dokonce vynálezy, které se statistikou neměly příliš společného<sup>59</sup>.

Okolo roku 1870 byly grafické reprezentace běžně rozšířené mezi řadou statistických úřadů v Evropě i USA (Friendly, 2008, s. 517). Vznikaly statistická alba a atlasy, které obsahovaly řadu zobrazení, jež statistici od 70. let 20. století často objevovali znovu (Friendly, 2008, s. 502). Jedním z nejrozšířenějších typů zobrazení tohoto období byly například statistické atlasy, kdy mapa samotná sloužila spíše jako orientační podklad a hlavní zájem se upínal na čísla a grafy.

Persvazivní funkci vizuální reprezentace dat poprvé demonstrovala Florence Nightingale, která pomocí ledovcového grafu demonstrovala špatné hygienické podmínky vojáků v bojích. Tuto vlastnost prezentace zdůrazňuje většina novodobých teoretiků od Jacquese Bertina (Bertin, 1967), přes Edwarda Tufteho, který ji nazývá též jako interokularitu (2001, s. 13-53), až po Stephena Fewa.

#### 3.3.1. Obnovení zájmu statistiků o vizuální reprezentaci dat

Po útlumu v první polovině 20. Století došlo v 70. letech k obnovení zájmu o tuto formu prezentace dat. Vzniklo také několik stěžejních teoretických publikací, které ovlivňují odborníky i v současnosti. Bertinova *Semiology of Graphics* (Bertin, 1967) je studií struktury organizování vizuálních prvků grafiky směrem k vyjevování vztahů v datech vedoucí k posílení lidského vnímání těchto vztahů. J.W. Tukey popsal zcela novou metodu vizuálního přístupu k analyzování dat (Tukey, 1977), která klade důraz na důležitost výběru konkrétní metody zobrazování pro porozumění datům.

---

<sup>59</sup> Jedním z nich je třeba fotografie – experiment Edwarda Myubridge později inspiroval ke vzniku tzn. Small multiples (Friendly, 2008, s. 508; Tufte, 1983a)



Americká vláda v polovině 70. let hledala způsob, jak informovat o současných socioekonomických trendech vedení země, vládní činitele a zároveň také veřejnost. Jedním z příkladů publikací určených právě pro veřejnost je pravidelně vydávaná brožura STATUS publikovaná Department of Commerce, která obsahovala řadu rozsáhlých, barevných a funkčně zpracovaných grafů týkajících se ekonomických a sociálních trendů. Navzdory zájmu a snahy American Statistical Association pokračovat ve vydávání i po změně vlády v roce 1976 bylo ale její vydávání po několika měsících zrušeno. Později v témže oddělení vznikla ještě publikace Graphic Presentation of Statistical Information, která zahrnovala inovativní postupy při prezentování multidimenzionálních dat ve 2D grafech, mapy nebo diagramy. (Marcus, 2007, s. 2)

Podobně důležitou publikaci jako Bertin a Tukey, vydal v roce 1983 Edward Tufte. V knize Visual Display of Quantitative Data (1983) publikoval teorii týkající se maximalizace nasycení užitečné, účelné informace. Zaměřil se také na praktické nástrahy při prezentování dat vedoucí k preciznímu předání informací, integritě zobrazení co nejjednoduššímu vnímání.

Dříve než se vizuální reprezentace dat etablovala v praxi národních i mezinárodních statistických úřadů a agentur, rozšířila se tato metoda v rámci přírodních věd a v roce 1986 vznikla metoda vědecké vizualizace a později datová vizualizace určená pro zpracování abstraktních dat<sup>60</sup>. V průběhu zejména 90. let a následně počátku 21. století vznikaly v tomto oboru stěžejní publikace, které ovlivnily jeho současnou podobu Ware (2001), Card (1999), Schneidermann (1992) Robertson (1991) (Chen, 2006, s. 3-8).

Podobně jako v 2. polovině 19. století, tak i na konci století 20 a počátku 21., měly na další používání a rozšíření vizualizací značný vliv nové technologie. Nyní zvyšující se výkon počítačů, dostupnost internetu a zdokonalující se grafické uživatelské prostředí. Právě díky tomu mohly vzniknout online i offline nástroje na tvorbu vizualizací jako je Many Eyes, Google vizualizace, D3, nástroj Trendylazer či specializované statistické nástroje a softwarové balíky jako SAS, SPSS, S-PLUS, a R, které navíc umožňují vytvořené analýzy graficky zobrazit pomocí statických vizualizací.

Tento trend zájmu o vizuální zobrazení dat zachytily a využívají v posledních 10 letech i statistické úřady. Pro řadu z nich se webové stránky staly hlavním publikačním místem, kam soustřeďují veškeré publikační aktivity, včetně vizuálních zobrazení a nástrojů na explorační

---

<sup>60</sup> V roce 1990 se potom konala první vědecká konference na téma vizualizace od ilee.org. (Institute of Electrical and Electronic Engineers (IEEE)).

analýzu dat. Poprvé takovou službu na webu nabídl uživatelům UK Office for National Statistics v roce 2005, když včlenil aplikaci umožňující vytvářet grafy z dat na lokální úrovni (Konference Otevřená data, 2014). Předznamenal tak současný trend, kdy se očekává využití poskytnutých dat k vytváření vlastních analýz přímo uživateli a tedy i generování znalostí. V roce 2007 ONS sestavila Data vizualizační centrum, kde jsou prozkoumávány možnosti interaktivního, animovaného či grafického prezentování dat. Tento přístup je v souladu jak s evropským statistickým kodexem, tak i s vlastním kodexem ONS v němž jsou statistici po velké krizi důvěry v 80. letech vedeni k transparentnosti a aktivnímu vytvoření bohatší zkušenosti pro uživatele (Smith, 2008, s. 120).

V roce 2009 se k tomuto trendu připojil také Eurostat, který publikoval platformu Statistics Explained, sloužící jako průvodce evropskými statistikami snadno srozumitelnou grafickou formou. Pokračování ve využívání vizuální reprezentace dat jako nástroje pro prezentaci oficiálních statistik potvrdilo i vydání European Statistics Code of Practice (European Statistical system Committee. 2011).

Práci s vizualizacemi a vyváření nástrojů na exploratorní analýzu dat si postupně osvojily i další statistické úřady – zejména Statistisches Bundesamt (Interaktiv & Anschaulich, 2013), Statistics of South Africa, severoamerický US Census (Quick Stats, 2015), italský Istat (Istat.it. 2013) a mezinárodní organizace jako je OECD, OSN, UN nebo African Development Bank (*Dataportal African Development Bank Group*). Český statistický úřad se k tomuto trendu připojil vytvořením Veřejné databáze.

### **3.4. Co se snaží statistické úřady dokázat publikováním vizualizací**

Jednou z částí mise statistických úřadů je komunikace informací a statistik veřejnosti. Pro naplňování tohoto poslání statistické úřady v současnosti zpravidla využívány vlastní webové stránky a prezentace (Rafalowska, 2006, s. 147). V souladu s přijetím Evropského kodexu statistiky (European Statistical system Committee. 2011, s. 9, 11-15) a zpravidla také deklarovaním ve vlastních stanovách a kodexech by se tyto instituce měly snažit o sdělování informací co nejsrozumitelnější formou (European Statistical system Committee. 2011, s. 1). Naplňování těchto požadavků vyžaduje respektování jednotlivých cílových skupin<sup>61</sup>, prostředků a způsobů, jakými statistické úřady mohou komunikovat svá sdělení.

---

<sup>61</sup> Statistický úřad v Nizozemsku definoval na 15 různých skupin, které si vždy žádají specifický způsob oslovení a komunikace. Ať už se jedná o přizpůsobení a zvolení odpovídající formy sdělení nebo zvoleného média. (Statistics Netherlands. 2014)

Jednou z možností považované za nástroj efektivní komunikace informací může být právě vizualizace. Tak už se děje například v případě italského statistického úřadu Istat. V nově deklarované strategii „Communication-4-Inovation“ (Work Session on the Communication of Statistics, 2014) je definováno použití vizualizace jako jeden ze stěžejních způsobů komunikace statistických analýz a informací rozmanitému publiku, přičemž respektování různých stupňů statistické gramotnosti publika je deklarováno jako výzva, nikoli jako překážka.

Při prezentaci své činnosti a sebraných dat se přitom soustředí zejména na publikování obsahu na webových stránkách, které jsou dnes úřady chápány jako nejsilnější a neúčinnější prostředek k zasažení veřejnosti, neboť tímto způsobem mohou nemediovaně oslovit až statisíce lidí<sup>62</sup>. Veřejnost na stránkách očekává transparentní a jasnou komunikaci, tedy vlastnosti, které správně vytvořená vizuální reprezentace dat uživatelům také může nabídnout. V britském Kodexu statistika je dokonce konkrétně uvedeno, že „lidé mají právo znát statistikovo porozumění zprávě od statistika tak, jako mají právo na data samotná“ (UK Statistics Authority, 2009, s. 13)

Vizuální reprezentaci dat v různých formách je možné už z podstaty své definice a existence považovat za optimální nástroj k naplnění výše uvedených úkolů a zájmů statistických úřadů. (Konference Otevřená data, 2014, video) Tento předpoklad je možné podložit daty o návštěvnosti z německého statistického úřadu, jehož webová sekce sdružující odkazy na interaktivní obsah čítá měsíčně 17 tisíc přístupů, což je považováno za uspokojující (Kiekenbeck, J. 2015, emailová komunikace)

Řada statistických úřadů si vhodnost vizuální reprezentace dat pro zprostředkování porozumění komplexním datům veřejnosti dlouhodobě uvědomuje a využívá její potenciál. Jako příklady je možné uvést US Census, který krátce již v polovině 70. let vydával veřejnosti dostupnou brožuru obsahující socioekonomické trendy zpracované ve formě snadno pochopitelných a atraktivních statických grafů a vizualizací (Marcus, 2007, s. 3). Dalším příkladem je produkce britského statistického úřadu Office for National Statistic (ONS), který v roce 2007 zřídil vlastní datavizualizační centrum (DVC), vyvíjející nové vizualizace i prostředí pro exploraci dat. Zároveň vyrábí řadu interaktivních a animovaných grafů, které ukazují strukturální změny v populaci a podporují nástroje k přizpůsobení pro každého uživatele (Smith, 2013, s. 175-180). Řadu prezentací šíří mezi veřejnost ve spolupráci s největšími médii v zemi.

#### **3.4.1. Cílová skupina uživatelů**

---

<sup>62</sup> ONS má měsíčně 700 tisíc unikátních přístupů na web, ČSÚ 50K denně 2015, (Konference Otevřená data, 14.3.2014)

Pro dosažení odpovídajícího účinku vizuální reprezentace dat je pro statistické úřady důležité respektovat cílovou skupinu, jíž jsou prezentace určeny. Cílová skupina příjemců statistik se s přechodem z papírového na webové publikování analýz a dat a postupnou demokratizací přístupu k datům značně rozšířila (Konference Otevřená data, 2014, video; Kiekenbeck, J., 2015, emailová komunikace). Je daleko více zřejmé, že uživatelům může chybět statistická gramotnost a v tomto kontextu je nezbytné respektovat limity uživatelů a nevytvářet taková vizuální zobrazení, která jsou příliš složitá a tedy určená spíše pro tvůrce samotné nebo jiné odborníky. Tento problém odhalil například US Census při testování užitečnosti a přívětivosti svých infografik, které byly v konečném důsledku lépe čitelné spíše pro vzdělanější než pro méně vzdělané uživatele, pro něž jsou vytvářeny (Gareau et al., 2015, s. 171). V datovém centru ONS se snaží riziko neinformovanosti vizuálních reprezentací minimalizovat vytvářením kolaborativních týmů skládajících se z datových analytiků, designérů a vývojářů, jejichž úkolem je navrhnout prezentaci založenou na běžných zkušenostech a potřebách občana a uživatele v reálném světě (Smith, 2013, s.173-185).

Statistickí si také uvědomují, že pomocí vizuálního zobrazení dat mohou rozvíjet statistické myšlení a zvyšovat statistickou gramotnost uživatelů (Novotný, 2014) a zároveň využívat persvazivní funkci vizuální reprezentace dat. Český statistický úřad (Černý, 2011, s. 16-18) v tomto kontextu uvádí jako příklad zobrazování dlouhých časových řad. ČSÚ pro tyto účely používal Google Motion Chart, tedy převzatý nástroj Trendylazer od společnosti Gapminder, jenž je dostupný ve Veřejné databázi. V témže nástroji se používá také prezentace statistických dat v mapách.

Vizuální reprezentace dat má jako nástroj ke komunikování statistických dat a analýz pro ČSÚ značný potenciál. Vyplývá to například ze zájmu o informační služby ČSÚ. Zde je patrné, že více než polovinu dotazů na ústředí vznáší veřejnost či firmy, tedy uživatelé, u nichž nelze s jistotou předpokládat odborné statistické vzdělání (Géblová, 2013). Dotazů je přitom několik desítek tisíc a ukázalo se, že jednoduše pochopitelné informace, které je možné rychle najít na webových stránkách, se uživatelé naučili vyhledávat sami a již o ně odborníky nežadají (Géblová, 2013), přitom dříve právě nejjednodušší dotazy převažovaly. Má tedy zřejmě smysl uvažovat vytváření takových způsobů prezentování složitějších informací, které by bylo snadno pochopitelné a mohlo i v tomto případě zvýšit počet dotazů, jejichž odpovědi by byli uživatelé schopni najít sami.

Tomuto řešení přitom nahrává i tematické zaměření poptávaných témat. Na základě informací využívání informačních služeb se ukazuje, že jde o oblasti, které už jiné statistické úřady úspěšně zpracovaly do datových vizualizací. Zájem ze strany tazatelů (Géblová, 2013) je především o porovnání spotřebitelských cen, jež zpracoval německý Statistisches Bundesamt

(Destatis.de, 2013), demografickou statistiku již ONS zpracoval jako interaktivní populační pyramidu obsahující predikci vývoje složení obyvatel až do roku 2083 (ONS, 2013). Komplexním příkladem řešení dotazů na zjišťování povšechného stavu a proměn socioekonomických ukazatelů o stavu země, je ročně aktualizovaný katalog Noi Italia pro rok 2015 (Istat.it, 2015)

### **3.5. Závěr:**

Je zřejmé, že statistické úřady si začínají uvědomovat svou roli komunikování statistik, neboť lidem dokáže pomoci v chápání současného světa v duchu slov bývalého britského a australského statistika Bill McLennan, tedy, že aby byly statistiky užitečné, je nutné je používat<sup>63</sup>. Vzhledem k tomu, že publikem a uživatelem dat produkovaných statistickými úřady jsou lidé s různou specializací a statistickou gramotností, řada úřadů pochopila, že klíčem ke komunikování jejich výstupů je nabídnutí co nejjednodušší a nejjasnější formy. A právě vizuální reprezentace dat, jak už se ukázalo v historii, by mohla být funkčním nástrojem pro účinné porozumění statistickým datům, neboť umožňuje přístup, interpretování, pochopení a použití statistických dat ve formě, která při správném zpracování a respektování limitů vizuálního zobrazení dat a lidského vnímání navazuje na obvyklé lidské zkušenosti a znalosti a rozšiřuje je prostřednictvím zviditelňování trendů, vztahů v datech a podporuje další uvažování.

---

<sup>63</sup> "to be useful official statistics need to be used".

## 4. Metodologická část

Tato část práce je věnována zázemí následujícího kvantitativního výzkumu. V rámci podkapitol vysvětlují jednotlivé zkoumané položky, věnují se způsobu získávání dat ze stránek statistických úřadů, procesu jejich zpracování a fázím analýzy. V této části také definují cíle a výzkumné otázky, které vycházejí jak z prvotního pohledu na sbíraná data, tak z dokumentů jednotlivých úřadů a mezinárodních organizací, v nichž jsou deklarovány povinnosti a poslání statistických úřadů (UK Statistics Authority, 2009; Český statistický úřad, 2011; European Statistical system Committee, 2011).

### 4.1. Výzkumný problém a kontext

Je zřejmé, že statistické úřady si i při působení ve virtuálním prostoru uvědomují svoji odpovědnost vůči veřejnosti, která vyplývá z dokumentů definujících poslání těchto institucí (UK Statistics Authority, 2009, s. 13; European Statistical System Committee, 2011; Český statistický úřad, 2011), s čímž souvisí i komunikování statistik prostřednictvím vizuálních zobrazení dat na vlastních webových stránkách, neboť uživatelům může pomoci v chápání současného světa v duchu slov bývalého britského a australského statistika Billa McLennana, tedy, aby byly statistiky užitečné, je nutné je používat (McLennan, 1995, s. 478)

Vzhledem k různým stupňům statistické gramotnosti a rozličné specializaci uživatelů statistických dat, vzaly statistické úřady na vědomí, že klíčem ke komunikování jejich výstupů je nabídnutí co nejvíce možností prezentování a poskytování dat. A právě vizuální reprezentace dat, jak už se ukázalo v historii<sup>64</sup>, by mohla být funkčním nástrojem pro účinné porozumění statistickým datům jednoduchou a jasnou formou, neboť umožňuje přístup, interpretování, pochopení a použití statistických dat ve formě, která při správném zpracování a respektování limitů vizuálního zobrazení dat a lidského vnímání navazuje na obvyklé lidské zkušenosti a znalosti a rozšiřuje je prostřednictvím zviditelňování trendů, vztahů v datech a podporuje další uvažování.

Soudím, že právě tento přístup, kdy jsou data a informace prezentovány nejen datovou, ale i vizuální formou, lze považovat za úspěšné pochopení a realizování plnění části poslání statistických úřadů směrem k veřejnosti, a to jak poskytováním veřejně dostupných dat, tak odbornou asistencí jakožto statistické autority při snaze o jejich chápání, vzdělávání a dbání o zvyšování statistické gramotnosti. Předpokládám tedy, že statistické úřady jsou znalé svojí

funkce a zvolené formy vizuálního prezentování dat pro používají pro její naplňování. Tato hypotéza bude později hlouběji rozvedena.

Vizualizace a grafické zobrazení statistických dat je známá metoda používaná v období raného vývoje statistických úřadů pro analýzu dat z průzkumů. Počátkem 20. století tuto zobrazovací metodu nahradily moderní, dosud využívané analytické metody jako je modelování a také metody aplikované statistiky, která umožňuje podrobnější a náročnější analýzy s větší podrobností a detailností (Friendly, Morgan 1997, 76). Renesanci využití vizuální reprezentace dat přinesly publikace Edwarda Tufteho (Tufte, 1983) a J. W. Tukeyho (Tukey, 1977), z jejichž teorií a následného uvedení do praxe těží vizualizovaní statistických dat dodnes. Jde především o moderní pravidla zobrazování grafu a vizualizace a teoretické základy pro vytvoření prostředí aplikací na explorační analýzu veřejných dat nabízených statistickými úřady zpravidla ve strukturovaných databázích. Tyto nástroje a obsahy se, po internetové revoluci a rozvoji možností webu 2.0, staly základními součástmi obsahu webových prezentací řady statistických úřadů.

Výzkumy a analýzy využití vizuální reprezentace dat při prezentování veřejně dostupných statistických dat jsou v posledních letech časté. Michael Friendly ve své prezentaci pro zástupce statistických úřadů na univerzitě v Yorku v roce 2007 uvedl (2003, s. 21), že konzumenti od statistických grafů očekávají jednoduchost, doslova „data dovnitř, obraz ven“ a pouze malá část uživatelů si žádá kompletní kontrolu nad vytvářením zobrazení a grafickými detaily. M. Pfankuch (Pfankuch & Wild, 2000, s. 134) potom v kvalitativní analýze statistického myšlení uvádí technologický progres a využívání grafických nástrojů pro prezentaci dat jako přelomový počín. Umožnilo to jak ztraktivnit a zvýšit zájem a přístup k datům i lidem bez technického vzdělání, tak možnost poskytnout uživatelům větší diverzitu při prezentování, neboť do té doby vypadaly vizuální prezentace velmi podobně, a to tak, jak to zavedlo první prezentování určitého typu dat (tzn. First-in-the-field). Evaluační výzkum provedený počátkem 70. let (Fienberg, 1979, s. 165-178) vyhodnocuje úspěšnost zavedení novinek, které se v průběhu 60. až konce 70. let ustálily jako způsob grafického prezentování dat. Vyzdvihuje zejména o použití kartografického zobrazení a grafů zobrazujících multidimenzionální data ve dvou dimenzích za použití small multiples. Podromou (2013, s. 4-5) ve své metaanalýze, hovoří o budoucnosti statistické grafiky a to především v kontextu vyučování studentů, neboť právě reprezentování a interpretace dat, společně s větším využíváním explorační analýzy dat a modelovacími aplikacemi, jež na základě dostupných dat dokáží modelovat budoucí situaci, má být novou budoucností pro vyučování a chápání statistiky. Greenfield (1993, s. 292) potom analyzuje změnu kultury v přístupu veřejnosti ke statistickým datům, kterou prochází i samotní statistici, kteří si teprve na počátku 90. let zvykali na otevřenost veřejných dat a limity statisticky málo

gramotné veřejnosti. Jako jeden z potenciálně funkčních nástrojů, který by měl zacelit mezeru mezi těmito dvěma skupinami, je uváděna právě vizuální reprezentace dat. V článku z roku 2010 (Wannabe, 2010) je evaluováno využívání několika grafických zobrazení dat, jehož výsledkem je potvrzení vhodnosti užití takového způsobu prezentace dat i ve statistických článcích, Friendly (1991) potom už počátkem 90. let experimentálně využil multivariantních zobrazení pro statistická data.

Přestože k tématu vizuálního reprezentování statistických dat vznikla řada studií a analýz, žádná z nich se nezabývá poskytováním veřejně dostupných dat a vzhledu do nich touto formou, jako způsobem naplňování poslání statistických úřadů. Vzhledem k rozšíření technologických a technických možností, zvýšení rychlosti a dostupnosti internetového připojení a vyspělosti nástrojů pro tvorbu vizuálních zobrazení, nabírá tato vizuální reprezentace dat statistickými úřady na aktuálnosti, neboť se tato forma stane zřejmě i v budoucnosti rovnocennou součástí veřejné služby těchto institucí. Je zde proto prostor pro vytvoření datově podložené kvantitativní analýzy zkoumající míru naplňování poslání statistických úřadů pomocí těchto nástrojů a veřejně dostupných dat, určených k dalšímu použití.

#### **4.2. Cíl výzkumu a výzkumné otázky**

Cílem tohoto výzkumu je zjistit, jak plní statistické úřady směrem k veřejnosti své deklarované poslání, a to prostřednictvím poskytování veřejně dostupných dat v různých formátech a nástrojích a zjistit jakou roli v této souvislosti hraje vizuální reprezentování dat jako prostředek k zjednodušení chápání statistik.

Jako poslání statistických úřadů směrem k veřejnosti je chápáno poskytování statistik a dat, a to nejlépe v takové formě, která je pro uživatele komfortní při dalším zpracování, jejich komunikování směrem k veřejnosti ve snadno pochopitelné formě, působení jako statistická autorita, vzdělávání a zvyšování statistické gramotnosti uživatelů.

Výzkumný problém jsem rozpracovala do několika hypotéz, které byly formulovány před započítím výzkumného šetření.

**Hypotéza H1: Statistické úřady z podstaty svého poslání poskytují veřejnosti data a tuto povinnost realizují v nejvyšší možné míře i prostřednictvím svých oficiálních webových stránek.**



**Hypotéza H2: Statistické úřady zemí, které jsou hodnocené jako otevřené a výkonné ekonomiky a státy tak mají dostatek financí na zabezpečení naplňování poslání statistických úřadů, jsou ty, které poskytují otevřená data a nástroje v nejvyšší možné míře, a to nejméně v 35 procentech.**

**Hypotéza H3: Vizuální formy prezentace dat produkovaných statistickými úřady jako způsob zvyšování statistické gramotnosti, koreluje s širší nabídkou strojově čitelných dat, než je obvyklé.**

### **4.3. Stručný popis metody výzkumu**

Následující výzkum je veden kvantitativní metodou. Probíhal formou rešerše jednotlivých oficiálních webových stránek statistických úřadů dle seznamu OSN <sup>65</sup> a to jak jejich mezinárodní, tak národní verze. Mezinárodní, obvykle méně obsáhlá verze, je vždy přeložena do anglického či francouzského jazyka, národní verze poskytují informace v jednom nebo více úředních jazycích. Existují také webové stránky statistických úřadů, jejichž mezinárodní a národní verze splývají. Jedná se zejména o webové stránky úřadů zemí s angličtinou jako úředním jazykem, výjimku tvoří bývalé francouzské kolonie v Africe, jejichž webové prezentace jsou pouze ve francouzštině.

O každé zobrazené webové stránce byl proveden záznam do sestaveného archu. Zkoumána byla přítomnost jednotlivých zvolených prvků, přičemž do tabulky byly vpisovány hodnoty 1 v případě, že daný web příslušnou položku obsahoval, či nic v případě, že ji neobsahoval. V rámci kategorie „interaktivní grafický obsah“ byla výjimečně využita i hodnota 2, jež rozlišovala, zda je daná položka vytvořena přímo statistickým úřadem, nebo se jedná o využití národní mutace aplikace, kterou vytvořila některá z mezinárodních organizací Development Bank Africa, World Bank nebo OSN. V dodatečném řádku označeném jako „typ vizualizace“ bylo v případě zápisu alespoň jedné číslice 2 uvedeno, které konkrétní nástroje statistický úřad poskytuje.

### **4.4. Sledované kategorie**

V příloženém archu (elektronická příloha) je uvedeno 22 hlavních položek, které kategorizují formy poskytování veřejně dostupných dat prostřednictvím webových stránek statistických úřadů. Vybrané kategorie jsou dále rozděleny na podkategorie.

První skupinou jsou položky základních produktů, které statistické úřady na svých webových stránkách nabízejí. Společným znakem je datový set a alespoň elementární uvedení do kontextu.

---

<sup>65</sup> [http://unstats.un.org/unsd/methods/inter-natlinks/sd\\_natstat.asp](http://unstats.un.org/unsd/methods/inter-natlinks/sd_natstat.asp)

**Rychlé informace** – číslo 1 je do archu zaznamenáno v případě, že webové stránky nabízí pravidelně aktualizované údaje o nejnovějším stavu indexů a statistik týkajících se konkrétní země. Jejich aktuální hodnoty se přepočítávají dle nejnovějších dat pocházejících od statistických subjektů, jako jsou instituce, firmy nebo jednotlivci. Periodicita jednotlivých revizí se liší a může dosahovat jednoho týdne až několika měsíců. Rychlé informace kvantitativně popisují socioekonomický stav dané země a zpravidla informují o stavu ekonomických ukazatelů jako HDP, stav dovozu a vývozu nebo míry nezaměstnanosti. Tato položka byla vybrána, neboť tento typ informace je jeden ze základních ukazatelů stavu země, pro obyvatele i zahraniční návštěvníky webu.

**Newsletter/tiskové zprávy** – číslo 1 je do archu zaznamenáno v případě, že statistický úřad umožňuje zaregistrovat email pro rozesílání newsletterů a aktuálních informací, či je možné tyto aktuální informace najít na webu ve formě tiskových zpráv. Tato forma poskytuje základní povědomí o dalším novém obsahu, který statistický úřad vytvořil.

**Publikace na webu (ročenka, analýzy, komentáře, bulletin aj.)** – číslo 1 je do archu zaznamenáno v případě, že jsou na webu publikovány rozsáhlejší monotematické dokumenty označované jako komentář, analýza nebo ročenka, které nabízejí souborná data za sledované období nebo jsou ve sledovaném odvětví vydávané pravidelně s určitou periodicitou. Tyto publikace se skládají z textové, tabulkové a grafické části a obvykle jsou poskytovány ve formátu PDF nebo jiném strojově nečitelném formátu. Hlavní funkcí těchto dokumentů je poskytnutí analytického pohledu statistického úřadu na daný jev reprezentovaný daty. Tento typ dokumentů je velmi důležitý, neboť předkládá zanalyzování jevu a jeho vysvětlení a význam v širším kontextu. Zároveň tyto publikace poskytují veřejnosti strukturovaná, harmonizovaná a agregovaná data o konkrétním jevu za delší časové období.

Druhou skupinou jsou položky produktů na webech statistických úřadů, jejichž součástí jsou data ve strukturované formě v jednom z níže uvedených formátů.

**Datové kolekce/průřezové statistiky** – číslo 1 je do archu zaznamenáno v případě, že statistický úřad disponuje datovými sety publikovanými přímo na svých webových stránkách. Liší se od datových archivů a databází především tím, že nejsou lokalizované v databázi se samostatnou webovou adresou, ale nachází se přímo na stránkách. Zpravidla jsou takové statistiky k dispozici dle témat nebo jsou uváděny jako zdroj dat pro analýzy. Vyskytují se v různých datových formátech, které jsou popsány níže. Poskytováním dat touto formou dochází k naplňování poslání statistických úřadů.

**Datové archivy s databáze** – číslo 1 je do archu zaznamenáno v případě, že webové stránky statistického úřadu obsahují stránku nebo odkaz na stránku, na níž úřady poskytují datové sety ve formátech určených k dalšímu zpracování a neobsahují žádný další doplňující text. V tomto nástroji jsou publikovány výhradně kategorizované datové soubory. V některých případech je databáze zároveň doplněna aplikací na exploratorní analýzu dat, tato možnost je ale zaznamenána zvlášť v kategorii níže.

V případě této položky bylo v archu rozlišeno, zda jde o databázi vytvořenou samotným úřadem, pak byla příslušná buňka vyplněna číslem 1. Pokud bylo vyplněno číslo 2, jedná se o databázi vytvořenou jednou z mezinárodních organizací, jejíž národní mutace byly poskytnuty více statistickým úřadům. Takovou databázi lze identifikovat jednak jednotným názvem (například \*NADA nebo REDATAM), jednotným vizuálním standardem a především identifikací organizace, která ji vytvořila. Tato identifikace se nachází v patičce či v horní části stránky stránky.

Uváděné externí databáze:

**NADA** – microdatový katalogový nástroj vyvinutý Světovou bankou, který slouží k procházení, vyhledávání, přístupu a stahování veřejně dostupných dat. Využívá DDI (Data Documentation Initiative), tedy mezinárodně uznávaný metadatový standard založený na xml.<sup>66</sup>

**REDATAM** – program vyvinutý mezinárodní organizací CEPAL. Jedná se o databázi mikrodat vytvořenou speciálně pro malé karibské státy, které do prostředí mohou nahrát zvolená data. V současnosti je k dispozici offline i online verze tohoto programu, který umožňuje prohlížet a transformovat zvolené datové sety.

Externí databáze dat spojené s nástrojem na explorační analýzu dat budou uváděny v příslušné položce.

**Datové formáty** – Následující formáty byly zvoleny neboť jsou obecně rozšířené či poskytují komfort při užívání a další práci s daty. Přestože byly vybrány běžně využívané datové formáty, je zřejmé, že některé se pro další využití a zpracování nehodí. Tento problém se týká zejména skupiny strojově nečitelných dat.

Formáty ve skupině strojově čitelných dat byly zvoleny s ohledem na kompatibilitu s běžně používanými programy ke zpracování dat.

Tímto způsobem jsem se rozhodla revidovat, jak statistické úřady plní své poslání formou nabízení dat, neboť pokud by poskytovaly data, jež nejsou strojově čitelná v běžně

---

<sup>66</sup> Software je volně dostupný zde: <http://www.ihsn.org/home/software/nada>

používaných programech, jednalo by se spíše o vytvoření bariéry než naplňování deklarované mise.

**Strojově čitelná data** – položka Strojově čitelná data označuje soubory dat, které jsou již vyčištěné, harmonizované, strukturované a agregované a k dispozici v níže uvedených formátech. Úřady musí zajistit, aby touto formou nedošlo k identifikování respondentů. Tyto formáty byly zvoleny, protože jsou obecně rozšířené.

CSV - číslo 1 je do archu zapsáno v případě, že statistický úřad nabízí soubory dat v tomto formátu, který je určený pro výměnu tabulkových dat.

XLS – číslo 1 je do archu zapsáno v případě, že statistický úřad nabízí soubory dat v tomto formátu pro výměnu dat, který je čitelný programem MS Excel, jedním z nejrozšířenějších tabulkových editorů

JSON - číslo 1 je do archu zapsáno v případě, že statistický úřad nabízí soubory dat ve formátu JavaScript Object Notation určeného pro výměnu dat. Je jednoduše čitelný i zapisovatelný člověkem a snadno analyzovatelný i generovatelný strojově. Jedná se o textový, na jazyce nezávislý, formát.

API - číslo 1 je do archu zapsáno v případě, že statistický úřad nabízí možnost používat rozhraní Application Programming Interface, určené pro volání dat z databáze.

ostatní – číslo 1 je do archu zapsáno v případě, že statistický úřad nabízí soubory dat v ostatních formátech strojově čitelných dat, které nejsou mezi výše uvedenými - např. spss,

XML – číslo 1 je do archu zapsáno v případě, že statistický úřad nabízí soubory dat v tomto značkovacím jazyce, který se používá se **pro snadnou výměnu informací a komunikaci** nezávislou na konkrétní aplikaci či platformě. Hlavními výhodami je jeho nezávislost, podpora národních kódování a jednoduchý převod na jiné formáty

**Strojově nečitelná data** – položka strojově nečitelná data zahrnuje soubory dat ve formátu, jež není možné ihned zpracovávat, ale je nutné data převést do jiného formátu či scrapovat. Následující formáty byly zvoleny, neboť jsou obecně rozšířené a často používané statistickými úřady.

PDF - číslo 1 je do archu zapsáno v případě, že statistický úřad nabízí soubory dat v tomto souborovém formátu vyvinutém pro ukládání dokumentů nezávisle na [softwaru](#) i [hardwaru](#), na kterém byly pořízeny

HTML – číslo 1 je do archu zapsáno v případě, že statistický úřad nabízí soubory dat, které se nachází přímo na dané webové stránce

ZIP – číslo 1 je do archu zapsáno v případě, že statistický úřad nabízí soubory dat v tomto kompresním formátu

ostatní - číslo 1 je do archu zapsáno v případě, že statistický úřad nabízí soubory dat v ostatních formátech strojově nečitelných dat, které nejsou vyjmenované

DOC. – číslo 1 je do archu zapsáno v případě, že statistický úřad nabízí soubory dat v tomto textovém souboru

Třetí skupinou položek, které statistické úřady na svých webech nabízejí, jsou nástroje a produkty poskytující vizuální reprezentaci dat. Byly vybrány, neboť jsou prostředkem naplňování poslání statistických úřadů a uživatelům pomáhají pochopit složité statistické jevy. Součástí této skupiny jsou následující položky:

**Vizualizace** – číslo 1 je zapsáno v případě, že webová stránka obsahuje tento interaktivní způsob grafické prezentace dat. Vizualizace je předmětem zkoumání této práce a je jedním z prostředků naplňování poslání statistických úřadů

**interaktivní projekty** – číslo 1 je zapsáno v případě, že webová stránka obsahuje vizuální či audiovizuální obsah na webových stránkách, který nesplňuje kritéria pro vizuální zobrazení, ale je vytvořen za účelem interakce uživatele s prostředím a získáváním informací

**mobilní aplikace** – číslo 1 je zapsáno v případě, že web obsahuje funkční odkaz na aplikace vytvořené pro mobilní telefony

**aplikace na EDA** – číslo 1 je zapsáno v případě, že webová stránka obsahuje odkaz na prostředí určené ke grafickému způsobu zkoumání patternů v datech. Je obvyklé, že tento nástroj je doplněn databází agregovaných, strukturovaných a vyčištěných dat, které je možné vyvolat a v různé formě analyzovat. Tato položka je předmětem zkoumání této práce a je jedním z prostředků naplňování poslání statistických úřadů.

Tato položka je rozlišena na aplikace, které vytvářejí statistické úřady interně, ty byly v archu označeny 1 a aplikace, které jsou vytvářeny mezinárodními organizacemi a dále poskytovány k použití statistickým úřadům včetně databází agregovaných, vyčištěných a strukturovaných dat. Tyto aplikace jsou označeny 2 a v poznámkách konkrétně vyjmenovány. Jedná se o:

DevInfo – databázový systém vyvinutý pod hlavičkou OSN. Jedná se o nástroj pro organizaci ukládání a prezentování dat v jednotné formě ke sdílení dat na národní úrovni. Je k dispozici všem členským státům OSN.

Countrystat – je webový systém pro uchovávání statistik o zemědělství, cenách obchodu a spotřebě produktů. Kromě databáze poskytuje grafický nástroj na analýzu dat na národní a lokální úrovni. Je k dispozici více než 20 státům Latinské Ameriky, subsaharské Afriky a Asie Afristat/OpendedataAfrica.org - explorační aplikace vytvořená African Development Bank, kombinující národní a mezinárodní statistiky. Nástroj umožňuje vizualizovat data 52 afrických států.

**infografika** – číslo 1 je zapsáno v případě, že webová stránka obsahuje tento statický způsob grafické prezentace dat. Tato položka je předmětem zkoumání této práce a je jedním z prostředků naplňování poslání statistických úřadů (viz výše)

**GIS systémy a atlasy** – mapová interaktivní a statická forma grafické reprezentace dat, která je předmětem zkoumání této práce a je jedním z prostředků naplňování poslání statistických úřadů

Poslední skupinou položky produktů, které statistické úřady na svých webech nabízejí, jsou ostatní způsoby, jak statistické úřady zpřístupňují data. Byly vybrány, neboť jsou také nástrojem, jehož prostřednictvím statistické úřady naplňují své poslání. Součástí této skupiny jsou následující položky:

**Videa** – číslo 1 je do archu zapsáno v případě, že webová stránka obsahuje audiovizuální obsah, který je součástí naplňování mise statistických úřadů.

**Učební materiály** - číslo 1 je do archu zapsáno v případě, že webová stránka obsahuje sekci s výukovým materiálem

**Účty na SM/ publikování na SM** - číslo 1 je do archu zapsáno v případě, že je na webových stránkách uveden odkaz na některou z následujících sociálních sítí:

Twitter

Facebook

Youtube

Pinterest

Instagram,

SlideShare

LinkedIn

Blog

Storify

**časopis** - číslo 1 je do archu zapsáno v případě, že je na webu možné najít úřadem vydávaný časopis či magazín

**volební data** - číslo 1 je do archu zapsáno v případě, že web obsahuje odkaz na výsledky voleb

#### **4.5. Zápis údajů**

Základním dotazem, který rozhodoval o tom, zda bude proveden zápis číslice 1 či vynechání, respektive 2, byl:

„Obsahuje webová stránka tohoto statistického úřadu tento konkrétní prvek?“

Zápis čísla 1 následoval v případě, že odpověď byla: *ANO, webová stránka statistického úřadu příslušný prvek obsahuje*, rozhodnutí o ponechání prázdného pole následovalo v případě, že odpověď na otázku byla: *NE, prvek pro příslušnou položku nebyl nalezen*. Číslo 2 je speciální rozlišovací možnost čísla 1. Tato možnost nastává pouze ve výše uvedených specifických případech.

#### **4.6. Čištění dat**

Ve vyplněném archu nebylo nutné provádět následné čištění dat, neboť zaznamenaná data nebyla kontaminována údaji nebo záznamy, které by jej mohly znehodnotit.

#### **4.7. Zaznamenané a nezaznamenané weby statistických úřadů**

Ačkoli bylo zobrazeno celkem 185 webových stránek statistických úřadů (Information on National Statistical Systems, 2015), zaznamenaný byly údaje o 159 z nich. Důvodem pro vyřazení byly nefunkční webové stránky ve zkoumaném období, ačkoli není známo, co bylo příčinou a zda se jednalo o krátkodobý výpadek či setrvalý stav.

Konkrétně byly pro stav mimo provoz vyřazeny stránky statistických úřadů Thajska, Sýrie, Iráku, Comorosu, Eritrey, Libérie, Bureau of Justice Statistics, Barmy, Kazachstánu, Severní Korey, Řecka, Burundi, Gambie, Mozambiku, Chile, Curacao, Saint Kitts and Nevis, Aruby, Bermud, Kajmanských ostrovů, Mexika, Jordánska, Guam, Tokelau, Ukrajiny a Srbska.

#### **4.4. Kvantitativní zhodnocení dat**

Kompletní datový soubor tvoří 159 záznamů o oficiálních webových stránkách statistických úřadů (seznam je součástí přílohy), přičemž bylo ještě před započnutím vyhodnocování bylo výše uvedených 26 vyřazeno pro nemožnost zobrazit příslušný web (viz příloha), ačkoli bylo cílem zmapovat stav všech oficiálních webových stránek statistických úřadů.

Pro ověřování hypotéz byl nejprve pomocí součtu hodnot v jednotlivých sloupcích příslušejících statistickým úřadům zjištěn stav položek na každé webové stránce a následně porovnán se 100procentním stavem. Výsledkem byl procentuální údaj, který byl následně porovnán s hodnotami ostatních statistických úřadů. Zároveň bylo vytvořeno pořadí jednotlivých statistických úřadů dle míry naplnění poslání u zveřejňování veřejně dostupných dat na webových stránkách.

V druhém kroku byl k výsledkům předchozího kroku přiřazen základní údaj o ekonomickém stavu země, tj. nejčerstvější údaj o výši HDP na obyvatele<sup>67</sup>. Následně byl celý set rozdělen na 3 podobně velké skupiny vytvořené na základě pořadí v předchozím kroku.

---

<sup>67</sup> <http://data.worldbank.org/indicator/NY.GDP.PCAP.PP.CD>



## 5. Výzkumná část

Poskytování veřejně dostupných dat bylo ověřováno v celkem 22 položkách (popsané v kapitole 4.4). Po vyplnění záznamů o přítomnosti jednotlivých položek, jejichž prostřednictvím statistické úřady na svých webových stránkách poskytují veřejně dostupná data, došlo k porovnání se stavem, který představuje přítomnost 100 procent zvolených položek. Poté byly provedeny součty všech hodnot v položkách listu, které obsahovaly nenulový údaj.

Vyhodnocením dat v tabulce bylo zjištěno, že žádný ze statistických úřadů nenaplnuje své poslání v poskytování veřejně dostupných dat na svých webových stránkách na 100 procentech. Tento stav je kromě aktivity samotných úřadů ovlivněn částečně také zákony dané země,<sup>68</sup> nebo nedostupností některých způsobů poskytování veřejně dostupných dat<sup>69</sup>.

### 5.1. Přehled o datech

Nejčastěji statistické úřady publikují na svých webových stránkách veřejně dostupná data prostřednictvím datových kolekcí a průřezových statistik, tedy základních informací o dané zemi, a to ve 135 případech ze 159. Téměř stejný počet byl zjištěn u publikací, jako jsou ročenky či další rozsáhlé publikace, vydávané s pravidelnou periodicitou. Datové archivy a databáze<sup>70</sup> a analýzy nabízí více než 100 webových stránek úřadů. Rychlé informace o stavu národních účtů a aktuálních nejdůležitějších údajích o zemi poskytuje na svých webech 99 úřadů.

#### 5.1.1. Datové formáty

Poskytování dat statistickými úřady v různých formátech na jednotlivých webech je jednou ze zásadních položek. Data v alespoň jednom formátu poskytuje celkem 131 statistických úřadů a 28 dalších data samostatně v žádném z formátů neposkytuje.

Alespoň jeden strojově čitelný datový formát poskytuje 107 statistických úřadů a alespoň jeden strojově nečitelný formát nabízí celkem 80 statistických úřadů. Nejvíce rozšířeným strojově čitelným datovým formátem je XLS, data v tomto formátu poskytuje k dalšímu zpracování 98 statistických úřadů. Druhým nejčastěji poskytovaným strojově čitelným formátem je formát CSV, který však na svých webových stránkách poskytuje méně než polovina předchozího počtu, jedná se o 43 webových prezentací statistických úřadů. Oba tyto nejrozšířenější strojově čitelné formáty dohromady lze najít na 39 webových stránkách statistických úřadů. Nejvíce využívaný strojově nečitelný formát je formát PDF, v němž nabízí data 60 úřadů.

---

<sup>68</sup> Například v případě zveřejňování volebních dat

<sup>69</sup> Případ některých sociálních médií – například v Číně či v některých zemích na Středním východě, je Facebook či Twitter oficiálně nedostupný.

<sup>70</sup> S blíže nespecifikovanými typy formátů dat

Většina úřadů, které umožňují stažení dat v nejrozšířenějším strojově čitelném formátu XLS, nabízí také aplikace pro explorační analýzu dat, jedná se o 74 z 98 webů statistických úřadů. Pokud se z tohoto počtu vytknou prezentace statistických úřadů Evropy, jichž tuto kombinaci nabízí 27, jde kromě Arménie o aplikace, které si statistické úřady vytvořily pro vlastní potřeby a bez využití platforem poskytovaných mezinárodními organizacemi. Úřady asijských zemí, které data v XLS a aplikaci na explorační analýzu dat poskytují, odkazují v 5 případech vlastní vytvořené aplikace<sup>71</sup> a v 5 případech na národní mutace aplikací vytvořených některou z mezinárodních organizací<sup>72</sup>, statistický úřad Spojených Arabských emirátů poskytuje obě varianty. Podobný poměr mezi vlastními a poskytnutými aplikacemi je i v případě webových stránek úřadů států v Jižní, Střední a Severní Americe. Úřady afrických států využívají ve velké většině poskytnutá prostředí vytvořená Africkou rozvojovou bankou nebo OSN<sup>73</sup>.

Formát CSV je svázán s poskytováním nástroje na explorační analýzu dat dokonce ve 37 ze 43 případů, kdy úřady data v tomto formátu na webové stránce poskytují. Pouze kanadský a malawský statistický úřad neumožňuje zároveň exportovat data i ve formátu XLS. S výjimkou Zambie, Zimbabwe, Ghany, Guiney Bissau a Malawi tento formát využívají pouze ty statistické úřady, které na své webové stránky umístily vlastní aplikaci na explorační analýzu dat.

V případě ostatních strojově čitelných datových formátů JSON, XML a prostředí API, lze říci, že byly vždy přítomny společně s nejméně jedním z dvojice výše uvedených typů.

V případě strojově čitelných dat lze u ostatních typů grafického zobrazení dat, tedy infografiky, vizualizace, aplikací využívajících GIS a kartografů, říci, že je vytvářejí zejména ty statistické úřady, které na svých webových stránkách poskytují nejméně jeden strojově čitelný formát dat. Výjimkou jsou v případě infografik Siera Leone, Honduras, Jamaika, Nepál a Papua Nová Guinea<sup>74</sup>. V případě vizualizací, data v alespoň jednom strojově čitelném datovém formátu nenabízí z 22 statistických úřadů<sup>75</sup> pouze Moldávie, Etiopie a Malajsie. Aplikace využívající GIS a atlasy poskytují s výjimkou Mikronéské republiky, Fiji a Moldávie všechny statistické úřady, které zároveň poskytují data ke stažení v alespoň jednom ze strojově čitelných formátů.

Obecně vzato lze říci, že pravděpodobnost shody poskytování strojově čitelných dat a alespoň jednoho způsobu grafické reprezentace dat, je u formátu XLS, CSV a API 86 a 88 procent. Pokud se jedná o formát JSON a XML, je v jejich případě shoda stoprocentní.

---

<sup>71</sup> Čína, Afghánistán, Vietnam, Spojené arabské emiráty, Turecko

<sup>72</sup> Jde o aplikaci DevInfo vyvinutou OSN; na webu ji z asijských států nabízí Indie, Jemen, Tádžikistán, Omán

<sup>73</sup> Jedná se o 27 úřadů vůči 3

<sup>74</sup> Jedná se o úřady států Nigerie, JAR, Uruguay, Trinidadu a Tobago, US Cesus, Hongkongu, Nového Zélandu, Chorvatska, České republiky, Litvy, Lotyšska, Nizozemska, Portugalska, Polska, Itálie a Ománu

<sup>75</sup> Singapur, Pobřeží slonoviny, Kongo, Etiopie, JAR US Census, Hongkong, Malajsie, Austrálie, Nový Zéland, Francie, Maďarsko, Německo, Litva, Irsko, Moldávie, Nizozemsko, Portugalsko, Itálie, Švýcarsko, Británie, Tunisko

V případě strojově nečitelných dat je pravděpodobnost, že na webové stránce bude zároveň některý z typů grafické reprezentace dat, nižší. Pro formát PDF platí 73procentní pravděpodobnost, pro HTML a ZIP se jedná o 70, respektive 60procentní pravděpodobnost. Vzhledem k nízké vzájemné korelaci výskytu těchto dvou forem, se tato analýza bude dále zabývat pouze souvislostí mezi poskytováním vizuálních reprezentací s formáty strojově čitelných dat.

### 5.1.2. Sociální média

Sociální média jsou v současné době jedním ze zdrojů informací a statistické úřady je chápou jako další nástroj k naplňování svého poslání<sup>76</sup>. Statistické úřady nejčastěji pro šíření zvoleného obsahu využívají sociální médium Twitter, řada účtů má i více než 100 tisíc followerů<sup>77</sup>. K účtu na Facebooku a Youtube odkazuje na svých webových stránkách 40, respektive 34 statistických úřadů. K účtům na ostatních sociálních médiích jako LinkedIn, Storify, SlideShare, Pinterest a Instagram odkazují statistické úřady na webech pouze jednotlivě<sup>78</sup>.

Sociální média jsou jako součást naplňování statistického poslání směrem k veřejnosti využívána nejvíce úřady v Evropě. Účtem alespoň na jednom sociálním médiu disponuje 76 procent úřadů. Nejméně na aktivní jsou na sociálních médiích úřady v Oceánii, s výjimkou Austrálie a Nového Zélandu. Úřady zemí v Asii, Africe a Americe<sup>79</sup> jsou na nejméně jednom sociálním médiu zastoupeny ve zhruba 30 procentech. Úřady v Evropě jsou nejčastěji aktivní na Twitteru, úřady asijských států používají spíše YouTube. Státy ve Střední, Jižní a Severní Americe využívají všechna 3 nejpoužívanější média ve stejném poměru, v případě Afriky mírně převažuje používání Facebooku a Twitteru nad YouTube.

Blogy využívá pouze několik statistických úřadů – statistické kanceláře USA, finský úřad a lotyšský úřad, tedy země, v nichž je obecně kladen velký důraz na otevřenost a transparentnost. Dále je jde o keňský a ománský úřad, také tyto statistické úřady kladou velký důraz na otevřenost, poskytování široké škály typů vizuálních reprezentací dat a působení na sociálních médiích.

Ostatní sociální média využívá například britský ONS (Storify, LinkedIn), ománský statistický úřad (Pinterest, Instagram), portugalský úřad a italský Istat (Pinterest, Slideshare) nebo Český statistický úřad (SlideShare).

---

<sup>76</sup> Dle výsledků prezentace na jednotlivých sociálních médiích jde například US Census, ONS, Estonský statistický úřad, Omán či sdělení zástupců ČSÚ v soukromém rozhovoru

<sup>77</sup> @ONS, @USCensus, @Statistiekbs.

<sup>78</sup> Slideshare – například ONS, Instagram – Omán, aj.

<sup>79</sup> Termín dále používaný pro Jižní, střední a Severní Ameriku

### **5.1.3. Interaktivní a neinteraktivní grafický obsah**

Alespoň jednu z položek Interaktivního a neinteraktivního vizuálního obsahu na svých webových stránkách publikovalo 115 statistických úřadů. Nejčastěji jde o aplikace pro explorační analýzu dat, aplikaci využívající GIS, obecně interaktivní projekty, kterými se ale tato práce dále nezabývá, atlasy a následně vizualizace a infografiky, které využívá 22, respektive 21 statistických úřadů. Nejméně jsou v této kategorii položek využívány mobilní aplikace a videa umístěná přímo na webových stránkách.

Interaktivní projekty a mobilní aplikace vytvářejí zejména úřady evropských států a států, na jejichž formování měla vliv britská monarchie a v současnosti jsou považovány za součást západního světa. Jedná se zejména o úřady Austrálie, Nového Zélandu a Spojených států amerických. Kromě předešlých tento obsah nabízí také statistické úřady Ománu, Číny, Mongolska, Spojených Arabských emirátů a Maroka. Také v případě interaktivních projektů jsou hlavními producenty statistické úřady evropských zemí, USA, Austrálie, JAR a Nový Zéland. Výjimky v podobě úřadů afrických zemí Keni, Namibie a Egypta jsou ale rovněž bývalou součástí britského impéria. Podrobnější analýza využívání ostatních typů Interaktivní a neinteraktivního grafického obsahu bude provedena dále.

### **5.1.4. Ověření hypotéz**

#### **5.1.4.1. Ověření hypotézy 1**

**Naplňování poslání statistických úřadů z pohledu komunikace a poskytování statistických dat probíhá v nejvyšší možné míře i prostřednictvím webových stránek.**

Předmětem této hypotézy bylo zjištění, v jaké míře dle z vybraných kritérií naplňují statistické úřady své poslání z hlediska komunikace statistik veřejnosti. V této fázi byla vyhodnocována data ze 159 záznamů o webových stránkách.

Zhodnocení bylo provedeno pomocí výpočtu<sup>80</sup>, který identifikoval, na kolik procent se danému úřadu daří na svých webových stránkách naplňovat poslání v komunikování statistik a poskytování dat veřejnosti. Výsledkem je procentuální údaj, který byl v samostatné tabulce následně porovnán s hodnotami, kterých dosáhly webové prezentace ostatních statistických úřadů, díky čemuž bylo možné zjistit pořadí jednotlivých statistických úřadů dle míry naplňování poslání vůči veřejnosti na základě stanovených kritérií. Po seřazení v pořadí od

---

<sup>80</sup> Součet přítomných položek/maximální počet položek

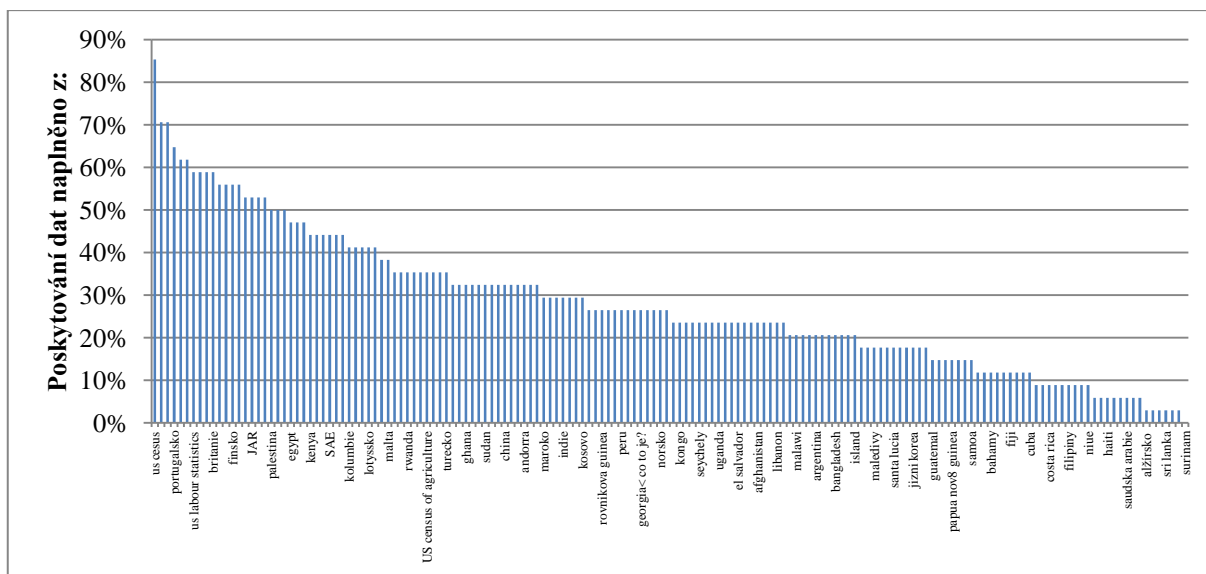
nejnižšího k nejvyššímu, byla jednotlivým úřadům přiřazena identifikace dle kontinentů a poznámka, zda je součástí daného webu také národní mutace databáze nebo aplikace na explorační analýzu dat, které byly vytvořeny mezinárodními organizacemi (jejich popis v 4.4)

Jako nejvíce odpovědný v této části posláni statistických úřadů byl vyhodnocen web US Census s mírou naplnění 85 procent, přičemž žádný z úřadů nedosahoval 100procentního naplnění. V případě webové stránky této statistické agentury nebylo naplněno poskytování dat o výsledcích voleb<sup>81</sup>, poskytování dat ve formátu XML a působení na sociálních médiích Storify a Instagram. Nejméně své posláni naopak naplňuje prostřednictvím webových stránek statistický úřad středoamerického státu Surinam. Na jeho webových stránkách nebyla nalezena žádná v této práci stanovená položka, jež by dokazovala naplnění mise tohoto statistického úřadu vůči veřejnosti.

Následující graf ukazuje distribuci míry naplnění u jednotlivých úřadů.(detail grafu v příloze 1). V trendu je možné identifikovat, že medián míry u zkoumaných webových stránek statistických úřadů je méně než 25 procent. Znamená to tedy, že polovina z nich poskytuje uživatelům pouze méně než jednu čtvrtinu položek, které byly identifikovány jako zásadní pro naplnění služby veřejnosti v rámci poskytování a komunikování statistických dat. Alespoň polovinu, tedy 50 procent zjišťovaných položek, nabízí jen 20 ze 159 zkoumaných webových stránek statistických úřadů, v relativním vyjádření se tedy jedná o 12 procent. Většina z těchto úřadů se nachází v Evropě.

---

<sup>81</sup> Tento nedostatek je dán faktem, že zákonnou povinností US Census Bureau není poskytování volebních dat na webových stránkách. Tato data jsou poskytována na webu jiné státní instituce.



Graf 1

### 5.1.5. Geografická poloha ve vztahu k naplňování poslání statistických úřadů

Jednotlivé weby statistických úřadů byly rovnoměrně rozděleny do 3 skupin dle míry naplňování vytyčených kritérií. Každou ze skupin tvoří záznamy o podobném počtu webových stránek statistických úřadů.

Skupinu, v níž se nacházejí statistické úřady, kterým se daří prostřednictvím webových stránek naplňovat poslání statistických úřadů z pohledu poskytování veřejně dostupných dat a jejich komunikace, tvoří 55 záznamů. Byly do ní zařazeny úřady, které vytyčená kritéria splňují na nejméně 35 procent. Druhou skupinu tvoří záznamy o 60 webových stránkách statistických úřadů, které naplňují poslání<sup>82</sup> z 32 procent až 21 procent. Poslední skupina sdružuje úřady, které poskytují méně než 18 procent zkoumaných položek.

#### 5.1.5.1. První skupina

První skupina je tvořena téměř z poloviny<sup>83</sup> záznamy o webových stránkách statistických úřadů v Evropě, což představuje více než 50 procent celkového počtu úřadů zemí na tomto kontinentu, zbývající část skupiny je tvořena rovnoměrným zastoupením záznamů o webových stránkách statistických úřadů zemí Ameriky, Asie a Afriky<sup>84</sup>. Nejméně početné zastoupení má Oceánie, a to prezentací úřadů Austrálie a nového Zélandu, v kontextu celého kontinentu jde ale o čtvrtinu zemí, tedy větší procentuální zastoupení, než mají úřady zemí Afriky i Asie. Tuto skupinu

<sup>82</sup> Naplňování poslání z hlediska poskytování veřejně dostupných dat a komunikování statistik

<sup>83</sup> Ze 43 procent

<sup>84</sup> Afrika 8 webových prezentací (18 procent), Asie 7 webových prezentací (19 procent), Amerika 6 záznamů (24 procent), včetně 4 statistických agentur Spojených států.

zároveň z 60 procent tvoří úřady, jejichž webové stránky, splňují alespoň polovinu vytyčených kritérií.

Úřady neevropských zemí v této skupině je možné rozdělit na dvě části. První byly v minulosti kolonizovány nebo byly po určitou dobu ve správě Velké Británie a nyní jsou zpravidla členy Commonwealthu<sup>85</sup>. Velký počet úřadů ze zemí sdružených v Commonwealthu v této skupině lze vysvětlit jako pozitivní důsledek vlivu procesu obnovování důvěry ve statistický systém, kterým prošla a stále prochází Velká Británie od počátku 90. let po krizi v 80. letech (tomuto problému se věnuji v kapitole 3.1.4.1). Tento proces, v jehož druhé fázi vznikl Kodex statistika (UK Statistic Authority, 2009), závazně ukládá statistikům z ONS nad rámec mezinárodních dohod naplňovat potřeby uživatelů a výsledky šetření komunikovat jasnou a srozumitelnou formou, která jejich respektuje limity (Laux, 2013, s. 4). Druhou část tvoří úřady států, které nebyly kolonií žádné evropské země jako Japonsko, Ázerbajdžán nebo Omán, nebo vznikly vyhlášením nezávislosti na jiných kolonizátorech z Evropy, než byla Velká Británie. Jedná se o Kolumbii, Rwandu, Senegal a Benin.

Úřady evropských států jsou zde zároveň s výjimkou Švýcarska členskými státy Evropské unie, ačkoli některé tradičně otevřené země Skandinávie jako Švédsko nebo Norsko v této skupině chybí<sup>86</sup>. Opačným případem je například Německo, zejména díky působení na několika sociálních médiích, vzdělávacím materiálům a široké nabídce produktů využívající grafickou prezentaci dat. Tato aktivita se německému statistickému úřadu zúročuje, stránka, která sdružuje odkazy na vizualizace a další grafické reprezentace dat (Destatis.de Interaktiv, 2015) dosahuje měsíčně až 17 tisíc přístupů a souží také jako pomůcka ke vzdělávání,<sup>87</sup> což úřad považuje za uspokojivé. Mezi úřady států s webovými stránkami, které komunikují a poskytují statistiky pomocí alespoň poloviny vybraných položek, se řadí také Český statistický úřad, a to navzdory méně rozmanitému množství strojově čitelných a nečitelných formátů dat. Tento stav vynahrazuje širokou škálou jiných způsobů šíření vlastního obsahu jako je mobilní aplikace, vzdělávací a interaktivní projekt Minisčítání, infografiky publikované prostřednictvím měsíčně zveřejňovaného časopisu a účtu na Twitteru<sup>88</sup>. Na svých webových stránkách úřad disponuje také vlastní aplikací na explorační analýzu dat (Český statistický úřad, 2011).

Z úřadů afrických států jsou v této skupině zastoupeny jak ty, jejichž ekonomiky lze v kontextu Afriky označit výkonnou (jde o úřady Keni, Egyptu, JAR, Nigérie, Tuniska), tak rozvojové státy Senegal a Rwanda, jimž k naplňování poslání výrazně pomáhají nástroje vytvořené

---

<sup>85</sup> USA; Singapur, Nový Zéland, Austrálie, JAR, Palestina (do roku 1948 část oblasti pod britskou správou), Egypt, Keňa, Kanada, Namibie, Spojené arabské emiráty, Nigérie.

<sup>86</sup> Tyto země nabízí pouze omezené množství typů vizuální reprezentace dat.

<sup>87</sup> Na základě emailových dotazů, celé znění emailu v příloze

<sup>88</sup> [Twitter.com/statistickyurad](https://twitter.com/statistickyurad)

mezinárodními organizacemi jako je OSN či African Development Bank<sup>89</sup>. Statistické úřady těchto prvních pěti států disponují širokým spektrem forem, které pomáhají uživatelům lépe a jednodušeji vnímat data. V případě JAR se jedná dokonce o 2 nástroje na explorační analýzu dat (pro laickou veřejnost SuperWeb<sup>90</sup>, pro odbornou o nástroj Nesstar<sup>91</sup>) rozsáhlý prostor věnovaný studijním materiálům, řadu kalkulátorů a příběhy připravené pro publikování v médiích. Keňský statistický úřad využívá širokou škálu poskytnutých nástrojů (NADA (Kenada, 2013), Africadata.org (KenInfo.org, 2015) a se věnuje osvětě prostřednictvím portálu otevřených dat (Kenya Open data, 2015), na kterém úřad vzdělává, poskytuje podporu vývojářům, publikuje analýzy a odkazuje na mapovou aplikaci.

Z úřadů asijských států jsou v této skupině zastoupeny státy s vysokým HDP na obyvatele - arabské státy těžící především z prodeje ropy a dále jde o Japonsko a Singapur. Výjimkou je úřad území Palestiny, jehož HDP je výrazně nižší<sup>92</sup>. Všechny statistické úřady těchto států poskytují širokou škálu typů grafických reprezentací dat a interaktivního obsahu obecně. Ománský statistický úřad odkazuje na webu k mobilní aplikaci, s výjimkou vizualizace komunikuje statistická data pomocí všech sledovaných typů grafické reprezentace dat, nabízí využití API a aktivně působí na šesti z osmi sledovaných sociálních médií. V podobném stavu je z hlediska naplňování také stránka singapurského statistického úřadu, který rovněž poskytuje všechny sledované typy grafické reprezentace dat, s výjimkou infografiky.

Statistické úřady zemí z Jižní a Severní Ameriky zastoupené této skupině jsou s výjimkou Kolumbie a Brazílie jednotlivé statistické kanceláře Spojených států, které mají obecně velmi vysoký standard v naplňování poslání statistických úřadů z hlediska komunikování statistik a poskytování veřejných dat prostřednictvím různých forem. Zcela suverénní postavení má v tomto ohledu US Census Bureau, který kromě poskytování volebních dat<sup>93</sup> a dat ve formátu XML poskytuje veškeré sledované nástroje, formáty a typy. Podobně jako v případě Velké Británie, je tento stav dán postupným historickým vývojem, kdy tyto statistické kanceláře implementovaly způsoby vizuálního prezentování dat již v průběhu 70. let (Marcus, 2007, s. 4).

#### **5.1.5.2. Druhá skupina**

Skupinu, která sdružuje státy, jejichž statistické úřady na webových stránkách naplňují poslání v poskytování veřejně dostupných dat z jedné pětiny až jedné třetiny, tvoří 62 států. Převažují v ní úřady afrických států a nejsou v ní zastoupeny úřady států z oblasti Oceánie.

---

<sup>89</sup> kromě nástroje na explorační analýzu dat Afristat, poskytují lokální mutaci datového archivu NADA

<sup>90</sup> (Superweb, 2015)

<sup>91</sup> (Nesstar, 2015)

<sup>92</sup> 143. místo

<sup>93</sup> Dle zákona je to záležitostí jiného úřadu



Nejpočetnější skupina úřadů afrických zemí tvoří více než polovinu z celkového počtu zkoumaných úřadů z tohoto kontinentu. Charakteristickým znakem jejich webových stránek je především odkaz na lokální mutaci aplikace k explorační analýze dat vytvořenou některou z mezinárodních organizací<sup>94</sup>, jež využívá 20 z celkem 25 statistických úřadů zejména ze střední a jižní Afriky. Tato strategie úřadům umožňuje naplňovat alespoň v základní formě statistické poslání i v případě, že samy nejsou schopné vytvořit vlastní nástroje. Země jako Etiopie, Guinea Bissau, Ghana a další obvykle poskytují základní informace o stavu národních účtů, některou z publikací a základní informace o zemi v podobě statistik o stavu obyvatelstva, aj. Tento obsah lze zpravidla zobrazit ve formátu PDF. Webové prezentace těchto zemí jsou jednoduché a žádný další obsah či typy grafické reprezentace dat nejsou poskytnuty<sup>95</sup>.

Statistické úřady zemí v Asii<sup>96</sup> až na výjimky, odkazující na mutaci aplikace DevInfo (Indie, Jemen, Kyrgystán a Tadžikistán), výše zmíněných aplikací nevyužívají. Navzdory tomu nabízejí alespoň jeden formát strojově čitelných dat<sup>97</sup> a kromě Číny a Hongkongu také alespoň jeden formát strojově nečitelných dat. Vizuální zobrazení dat jako způsob komunikování statistik se nicméně využívají jen málo. Exploraci dat pomocí grafů umožňuje pouze čínský, malajský, afghánský a vietnamský statistický úřad. Bangladéšský úřad používá pro prezentování statistik aplikaci využívající geografický informační systém<sup>98</sup>.

V Jižní a Střední Americe je situace v souvislosti s využíváním aplikací na explorační analýzu dat rozdílná a statistickým úřadům těchto států je poskytován datový archiv REDATAM, který pomáhá širší nabídce strojově čitelných dat na jednotlivých webech. Tyto úřady obvykle produkují data ve dvou strojově čitelných formátech dat<sup>99</sup>. V porovnání s úřady asijských států v této skupině poskytují tyto úřady také častěji i některé typy grafické reprezentace dat. Jedná se zejména infografiky v případě úřadů Trinidadu a Tobaga a Uruguaye. Zobrazení statistik v mapách umožňuje webová stránka statistického úřadu Argentiny a Paraguaye.

V této skupině se nacházejí také úřady evropských států. Jde především o státy, které nejsou součástí Evropské unie a nejsou tedy vázány Kodexem evropské statistiky. Výjimku v tomto případě tvoří statistický úřad Slovenska, Švédska a Belgie. Navzdory předpokladům, míře rozvinutosti zemí a platným závazkům vůči výše uvedenému dokumentu, na svých webových

---

<sup>94</sup> Countrystat, DevInfo a Africadata

<sup>95</sup> Jedná se o země Maroko, Burkina Faso, Pobřeží Slonoviny, Rovnikova Guinea, Tanzanie, Kamerun, Kongo, Mauritánie, Mali, Seychely, Togo a Uganda

<sup>96</sup> Jedná se o úřady Afghánistánu, Hongkongu, Číny, Indie, Izraele, Mongolska, Nepálu, Jemenu, Tádžikistánu, Pákistánu, Vietnamu Pákistánu, Vietnamu, Bangladéše a Libanonu.

<sup>97</sup> S výjimkou Nepálu

<sup>98</sup> GIS

<sup>99</sup> Výjimkou jsou úřady států Maledivy, Jamaica, Argentina a Nicaragua

stránkách kromě základních způsobů prezentace veřejně dostupných dat<sup>100</sup> poskytují ve srovnání s ostatními členy EU pouze málo dalších nástrojů. Úřady Švédska a Slovenska umožňují alespoň exploraci dat ve vlastní aplikaci a využívání jednoho strojově čitelného formátu. V případě všech zmíněných států je společná také malá aktivita na sociálních médiích. Podobnou úroveň komunikování statistik a poskytování statistických dat mají i další úřady evropských států v této skupině<sup>101</sup>. Úřady Andorry a Kosova umožňují exploraci dat kromě EDA aplikace také pomocí map, moldavský statistický úřad vytváří vizualizace a na webové stránce odkazuje ke stažení mobilní aplikace.

### 5.1.5.3. Třetí skupina

Skupina tvoří úřady, jejichž webové stránky naplňují poslání v poskytování veřejně dostupných dat a komunikování statistik, pouze v malé míře<sup>102</sup>. Nejvíce statistických úřadů v této skupině<sup>103</sup> se nachází ve státech Jižní a Střední Ameriky. V celkovém poměru tvoří jejich počet v této skupině téměř polovinu počtu všech zkoumaných úřadů z tohoto kontinentu. V případě webových stránek úřadů z asijských a afrických zemí jde pouze o třetinu a čtvrtinu celkového počtu všech sledovaných prezentací statistických úřadů<sup>104</sup>. V případě statistických úřadů států v Ocenánii je rozdíl nejvíce kontrastující. Ačkoli je počet záznamů o úřadech v této oblasti téměř nejnižší, jejich počet představuje více než 80 procent celkového počtu webových prezentací úřadů z této oblasti. Webové prezentace evropských úřadů v této skupině naopak představují pouze 16 procent a jedná se o úřady velmi malých zemí a městských států jako je Černá Hora nebo Monako. Jedinou výjimku v tomto trendu představuje Rusko, které na svých webových stránkách neposkytuje prakticky žádnou formu veřejně dostupných dat, ani nepůsobí na žádném ze sociálních médií.

Aplikace k exploraci dat pomocí grafů či datové archivy využívá pouze desetina těchto úřadů zejména ze středoamerických ostrovních států Kostariky, Barbadosu a Svaté Lucie a úřadů rozvojových afrických států jako Burundi a Rovnickové Guiney.

S výjimkou statistického úřadu Venezuely, Jižní Korey, Fayerských ostrovů, Bosny a Hercegoviny a výše zmíněných států, neposkytují úřady data ve strojově čitelných formátech<sup>105</sup>. Ze sledovaných typů vizuálních reprezentací dat je na webových stránkách úřadů v této skupině možné nalézt pouze infografiky<sup>106</sup> a aplikaci zobrazující data v mapách<sup>107</sup>. Statistické úřady

---

<sup>100</sup> Rychlé informace, publikace, analýzy

<sup>101</sup> Rumunský statistický úřad jako jediný z úřadů evropských zemí v této skupině využívá nástroj Devinfo.

<sup>102</sup> Čítá 50 záznamů.

<sup>103</sup> Celkem 14

<sup>104</sup> V této části se nachází 10 respektive 11 úřadů,

<sup>105</sup> Ale i přesto poskytují statistiky v jednom ze strojově nečitelných formátů PDF a HTML

<sup>106</sup> Papua Nová Guinea, Honduras, Jižní Korea

<sup>107</sup> Fiji, Mikronésie Jižní Korea

zařazené v této skupině také s výjimkou Hondurasu a Guatemaly neodkazují na svých webových stránkách na vlastní účty na sociálních médiích.

#### **5.1.6. Závěr:**

Rozdělení webových prezentací statistických úřadů do 3 skupin dle míry naplňování poslání v podobě komunikace statistik a poskytování veřejně dostupných dat ukázalo v tomto ohledu základní trendy. Prvním z nich nenaplnění předpokladu vysoké míry statistické mise statistických úřadů vůči veřejnosti prostřednictvím webových stránek. Pouze 15 procent států poskytuje službu prostřednictvím více než 50 procent sledovaných položek. Zároveň je v tomto ohledu patrná vysoká míra poskytování dat a komunikování statistik u většiny států EU a mimoevropských států, které byly ovlivněny koloniální politikou Velké Británie. Prezentace úřadů některých států, které jsou označovány za otevřené a dbající na transparentnost ale dosahuje úrovně, která je srovnatelná s úřady rozvojových zemí Afriky.

Výraznou pomocí v naplňování tohoto poslání jsou nástroje a datové archivy poskytované mezinárodními organizacemi, což je zřejmé zejména z výsledků statistických úřadů států střední a jižní Afriky.

## **5.2. Ověření hypotézy 2**

**Míra plnění poslání statistických úřadů je přímo úměrná ekonomické situaci země a otevřenosti její ekonomiky.**

Tato hypotéza předpokládá, že statistické úřady zemí, které jsou hodnocené jako otevřené a výkonné ekonomiky a státy tak mají dostatek financí na zabezpečení naplňování poslání statistických úřadů, jsou ty, které poskytují otevřená data a nástroje na jejich zpracování ve více než 35 procentech.

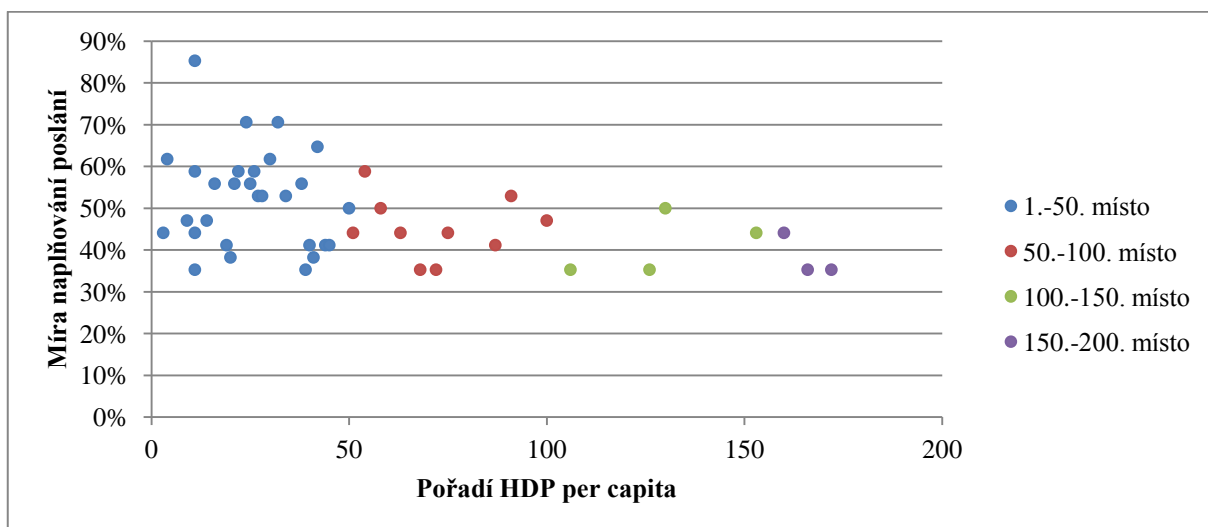
Postup vyhodnocování této hypotézy navazuje na předchozí kapitolu. Úřadům seřazeným dle míry naplňování vytyčené mise bylo přiřazeno pořadí dle aktuální výše HDP na obyvatele<sup>108109</sup>. Vyhodnocení hypotézy probíhalo pomocí bodového grafu, jenž usnadnil identifikování korelací.

Následující graf ukazuje distribuci celkem 55 statistických úřadů, které na svých webových stránkách naplňují poslání definované zvolenými kategoriemi na nejméně 35 procent a korelaci tohoto stavu s pozicí v žebříčku podle výkonnosti ekonomiky.

---

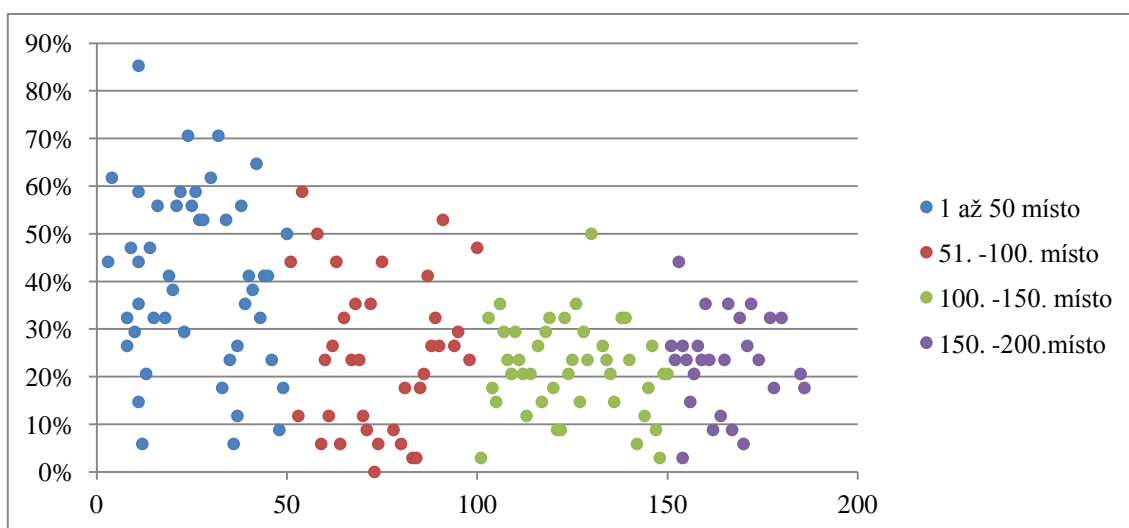
<sup>108</sup> Seznam států dle aktuálního HDP se nachází v elektronické příloze

<sup>109</sup> Jsem si vědoma limitů, které tento údaj má, zvolen byl zejména z důvodu, že jím disponuje většina států na světě. Přesnější by byl zřejmě index otevřenosti ekonomiky, který se sice počítá, přesto jej má pouze 75 zemí



Graf 2

Pokud nejsou data uvedena do kontextu, zdá se pravděpodobné, že daná hypotéza je alespoň částečně platná, neboť mezi úřady, které plní své poslání na nejméně 35 procent jsou skutečně z většiny úřady států, které se v žebříčku HDP na obyvatele umístily na čelních místech. V následujícím porovnání s ostatními skupinami je ale zřejmé, že ekonomická situace země nemusí mít primárně souvislost s mírou naplňování poslání statistických úřadů. Tento trend je patrný, jak je viditelné výše, u úřadů, které poslání naplňují z více než 50 procent. Tento stav ale není možné aplikovat na celou škálu, což je zřejmé z grafu, který zobrazuje údaje o všech zkoumaných webových stránkách statistických úřadů.



Graf 3

Je patrné, že země, které se v celosvětovém žebříčku umístily na 100. a zejména 150. a nižším místě, nemusí zároveň také nejméně naplňovat své poslání. Jak je z grafu zřejmé často jejich webové prezentace dosahují i na úroveň států s ve skupině s nejvyšším HDP na obyvatele.

Přestože se nejvyšší míra poskytování veřejně dostupných dat a komunikování statistik koncentruje ve státech s nejvyšším HDP, skupina, jež naplňuje poslání nejméně z jedné čtvrtiny až jedné třetiny je nejsilněji zastoupena státy na 80. až 160. místě. Její součástí jsou i ty nejchudší státy světa, jež se nachází v Africe, jako Burundi, Malawi nebo Kongo. Tato specifika budou dále osvětlena.

Skupinu, která poslání statistických úřadů naplňuje nejméně, tvoří dvě skupiny států. Jednou z nich jsou málo rozvinuté asijské státy jako Tádžikistán, malé státy v Oceánii jako Nauru nebo Papua Nová Guinea a malé evropské země jako je Monako či San Marino. Na druhé straně jde o úřady výrazně bohatších zemí jako je Rusko, státy v Jižní a Střední Americe a některé asijské státy, které jsou v podobné situaci z hlediska přepočtu HDP na hlavu. Tyto vzorce budou rozebrány dále.

#### **5.2.4. Náplň poslání statistických úřadů ve vztahu k HDP dle jednotlivých kontinentů**

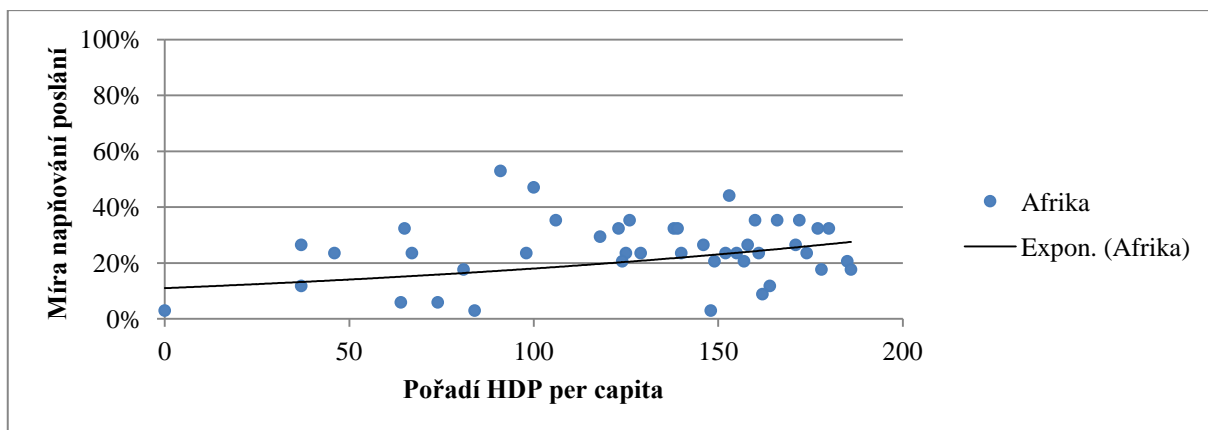
##### **5.2.4.1. Africké státy**

Africké státy se až na výjimky jako Seychely, Rovnicková Guinea, JAR a Gabon řadí mezi rozvojové státy, které mají velmi nízké HDP na obyvatele. Přesto jsou to právě úřady těchto států mezi 120. a 160. místem v žebříčku pořadí HDP na obyvatele, které naplňují poslání statistických úřadů ve větší míře, než je tomu u bohatších států v tomto regionu<sup>110</sup>. Navzdory vyššímu HDP<sup>111</sup>, naplňují statistické úřady států Gabon, Alžírsko, Botswana či Rovnicková Guinea, poslání statistických úřadů dle daných kategorií pouze z méně než 10 procent. Zároveň neposkytují na svých webových stránkách žádný typ grafické reprezentace dat ani neumožňují využívání dat ve strojově čitelném formátu.

---

<sup>110</sup> , jako jsou Gabon, Alžírsko, Botswana či Rovnicková Guinea

<sup>111</sup> Mezi 37 a 84. místem



Graf 4

Druhou skupinou států jsou již zmíněné málo rozvinuté země s nízkým HDP na obyvatele, které navzdory předpokladům naplňují posláním statistických úřadů lépe, než úřady rozvinutějších zemí či úřady zemí s podobnou silou ekonomiky v jiné části světa<sup>112</sup>. Tento jev byl již zmíněn v souvislosti s rozdělením jednotlivých skupin, podrobně v 5.1.5. Úřady států jako jsou Pobřeží Slonoviny, Kamerun, Timor Leste, Lesotho, Jižní Súdán a další ve středo a jihoafrickém regionu<sup>113</sup> využívají nástrojů mezinárodních organizací v národních mutacích, které vedou ke zvýšení míry naplňování posláním statistických úřadů prostřednictvím webových stránek. Datovým archivem NADA a několika vzájemně podobnými aplikacemi, které jsou kombinací databáze a nástroje na explorační analýzu dat<sup>114</sup> disponuje například webová stránka úřadu Burkiny Faso, Burundi nebo Kamerunu a dalším<sup>115</sup>. Je paradoxní, že jde často o statistické úřady v zemích, v nichž je omezena svoboda obyvatel<sup>116</sup>. Například statistický úřad ve Rwandě poskytuje nejen více typů nástrojů umožňujících grafickou exploraci, ale i data ve více typech strojově čitelných formátů, než statistický úřad ve Švédsku či v Belgii.

Výjimky z obou výše zmíněných skupin jsou statistické úřady států, jež byly v minulosti součástí britského impéria a v kontextu Afriky mají vysoké HDP, ačkoli jde ve světovém kontextu o země, které se nachází na 91. – 153. místě žebříčku<sup>117118</sup>. Na webových stránkách tyto úřady nabízí 2 nástroje na explorační analýzu dat, z čehož je vždy alespoň vytvořen přímo úřadem, nejméně dva strojově čitelné formáty dat, API, aplikaci k prohlížení statistik založenou na GIS a dále v případě JAR mobilní aplikaci a infografiku. Všechny zmíněné úřady také působí nejméně na dvou sociálních médiích.

<sup>112</sup> Například Oceánie

<sup>113</sup> Příloha 2

<sup>114</sup> DevInfo, Countrystat a Africadata

<sup>115</sup> Džibuti, Etiopie či Guinea Bissau

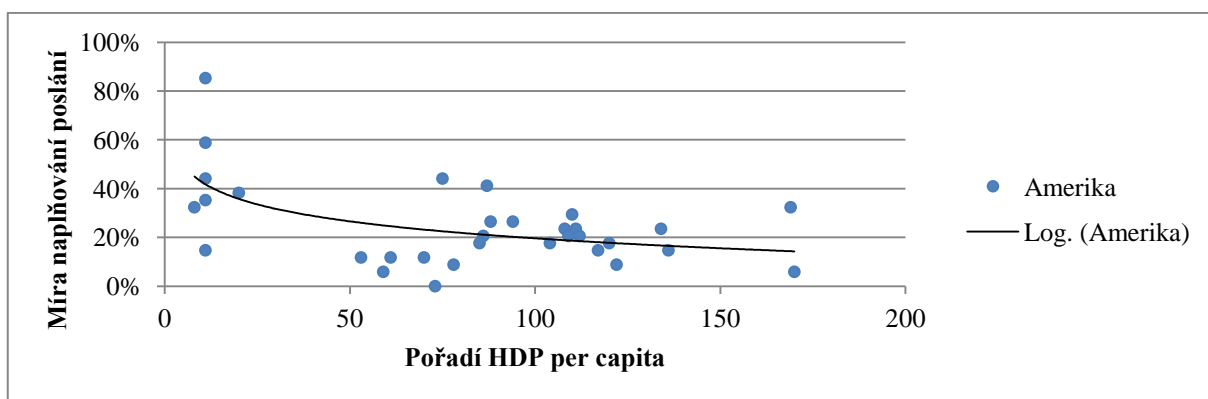
<sup>116</sup> Vlády v zemích jako je Súdán, Etiopie, Rwanda nebo Zimbabwe jsou označovány jako diktátorské

<sup>117</sup> Keni, JAR a Egypta

<sup>118</sup> JAR 91., Keňa 153., Egypt 100.

### 5.2.4.2. Amerika

Pod jednotné označení Amerika byly zahrnuty webové prezentace statistických úřadů států v Jižní, Střední a Severní Americe. Z ekonomického hlediska i zkoumání naplňování statistické mise jsou mezi těmito zeměmi velké rozdíly. Americký US Census je úřadem, který naplňuje své poslání nejlépe ze všech hodnocených webových prezentací a na celém kontinentu je jediným úřadem, který naplňuje vytyčená kritéria z více než 50 procent. Jak již bylo uvedeno výše, disponuje všemi formami grafického interaktivního obsahu a všemi typy grafické reprezentace dat. Kanadský statistický úřad navzdory příslušnosti země ke Commonwealthu a 20. nejvýkonnější ekonomice světa, nedisponuje, s výjimkou aplikace na explorační analýzu dat, žádným dalším typem grafické reprezentace dat. Naplňování poslání tohoto statistického úřadu a komunikování statistik je tak méně důsledné než u úřadů států, které se v žebříčku HDP na obyvatele nachází na hranici 100. místa<sup>119</sup>.



Graf 5

Jak je zřejmé z grafu, úřady skupiny států kumulujících se mezi 80. a 120. místem žebříčku HDP na obyvatele naplňují 16 až 30 procent vytyčených kritérií. V tomto rozmezí se nachází téměř jedna pětina všech zkoumaných statistických úřadů. Dvě třetiny těchto úřadů<sup>120</sup> jsou, podobně jako země subsaharské Afriky, v poskytování služeb veřejnosti prostřednictvím webu závislé na nástrojích, které byly vyvinuty mezinárodními organizacemi<sup>121</sup>. Úřady ekonomicky vyspělejších států, jejichž HDP na osobu je mezi 50. a 80. místem žebříčku, nenabízí žádné nástroje na exploraci a grafické zobrazení dat<sup>122</sup>.

<sup>119</sup> JAR, Egypt, Keňa

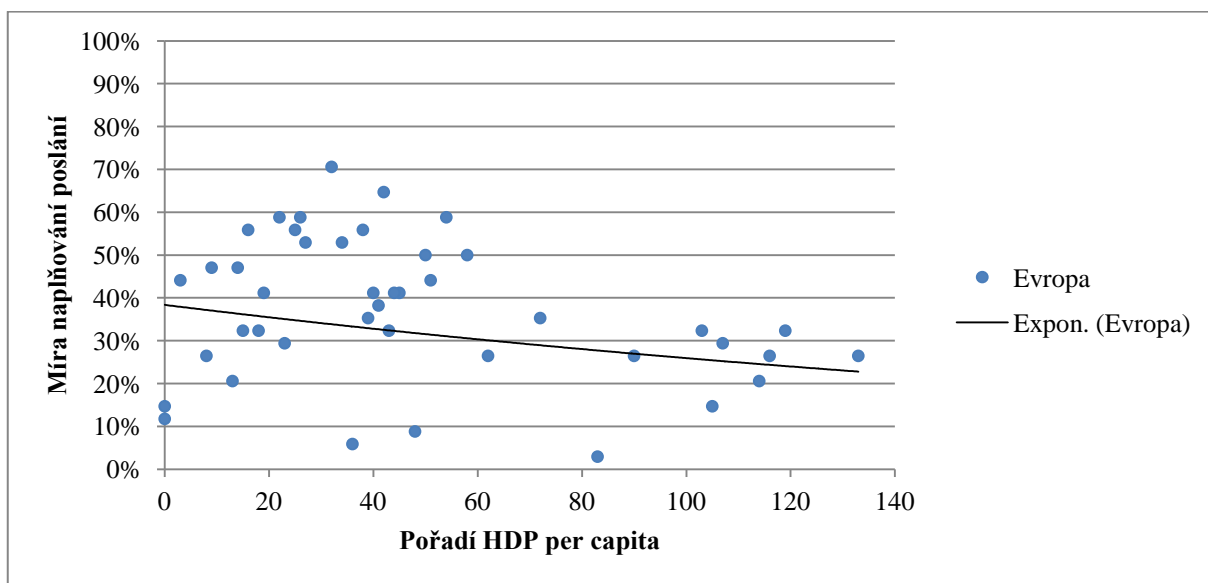
<sup>120</sup> Vyvinuté externě – Dominikánská republika, El Salvador, Svatá Lucie, Guyana, Paraguay, Nicaragua. Vlastní – Peru, Belize,

<sup>121</sup> REDATAM

<sup>122</sup> ani v jednom z případů tyto nástroje k dispozici neměli, i z tohoto důvodu úřady těchto malých ostrovní zemí v Karibském moři jako Panama, Surinam Bahamy nebo Kuba

### 5.2.4.3. Evropa

Statistické úřady evropských států společně s Austrálií, US Census a Novým Zélandem nejlépe naplňují požadavky této hypotézy o souvislosti ekonomické vyspělosti a aktivně projevované zodpovědnosti vůči mezinárodně schváleným povinnostem, deklarovaným misím a požadavkům na naplňování poslání statistických úřadů.



Graf 6

Obecně se většina států evropských zemí nachází do 80. místa žebříčku HDP na obyvatele a jejich statistické úřady se obvykle prezentují takovými webovými stránkami, které splňují vytyčená kritéria nejméně z jedné třetiny. Nejlépe se z evropských statistických úřadů daří naplňovat službu veřejnosti prostřednictvím webu těm statistickým úřadům, které jsou zároveň členy Evropské unie. Tyto úřady ještě společně se švýcarským statistickým úřadem naplňují sledované položky většinou z více než 50 procent<sup>123</sup>. Je pravděpodobné, že na tento stav má vliv závaznost naplňování Kodexu evropské statistiky, který mimo jiné ukládá za povinnost zvyšovat statistickou gramotnost veřejnosti, vzdělávat a komunikovat statistiky ve formě, která je snadno pochopitelná. Tyto státy také nabízejí nejméně jeden formát strojově čitelných dat, vlastní aplikaci na explorační analýzu, řadu vizualizací<sup>124</sup> a infografik<sup>125</sup>. Kromě britského statistického úřadu v tomto ohledu vynikají úřady Itálie, Německa, Švýcarska, Portugalska, Litvy a České republiky.

Výjimkami z popsaného trendu jsou webové stránky statistických úřadů malých městských států jako Monako, San Marino, či Andorra, ale také států, u nichž byla předpokládána větší

<sup>123</sup> Výjimka Rumunsko Švédsko

<sup>124</sup> Například Rakousko, Nizozemsko

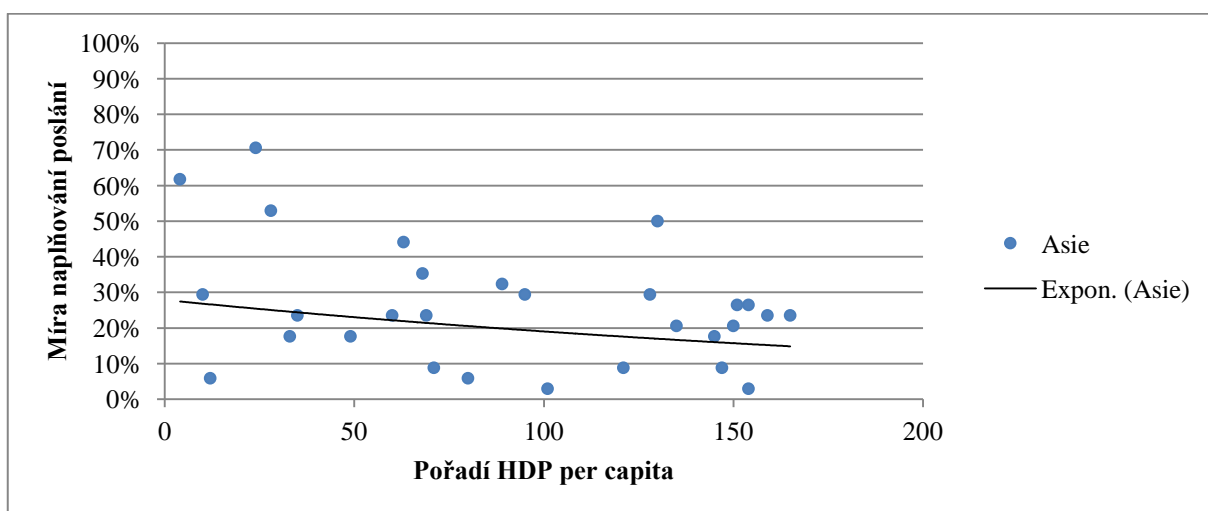
<sup>125</sup> ČR



otevřenost a vstřícnost k respektování dohod nastavených v rámci EU a k využívání interaktivních a grafických nástrojů k exploraci dat. Konkrétně se jedná o Švédsko, Belgie a Norsko, tedy státy s HDP mezi TOP 30 a s výjimkou Norska dlouholeté členy EU<sup>126</sup>.

#### 5.2.4.4. Asie

Navzdory velkému rozpětí pořadí asijských zemí v žebříčku HDP, je zřejmé, že s výjimkou tří outlierů, nedosahuje žádná webová prezentace statistických úřadů vyšší míry naplňování poslání než 50 procent. V případě asijských statistických úřadů je souvislost mezi ekonomickou vyspělostí a otevřeností a velkou rozmanitostí nabídky nástrojů a datových formátů na webových stránkách nejednoznačná.



Graf 7

Z grafu je zřejmé, že jak v ekonomicky vyspělých státech jako je Hongkong, Saudská Arábie, Jižní Korea nebo Izrael, tak i v zemích s nižším HPD na osobu jako Kyrgystán či Indie je možné nalézt úřady, které poskytují více než třetinové naplňování poslání i takové, které na svých webových prezentacích nenabízí téměř žádné strojově čitelné datové formáty či infografiky nebo vizualizace.

V prezentacích statistických úřadů těchto zemí se s výjimkou úřadů SAE, Jemenu, Tádžikistánu Malajsie a Indie nevyskytují národní mutace nástrojů, které by byly poskytnuty mezinárodními organizacemi<sup>127</sup>.

Outliery v Asii představují státy s předním postavením ve světové ekonomice<sup>128</sup>, jejichž statistické úřady zároveň provádí prostřednictvím webových stránek kvalitní službu veřejnosti.

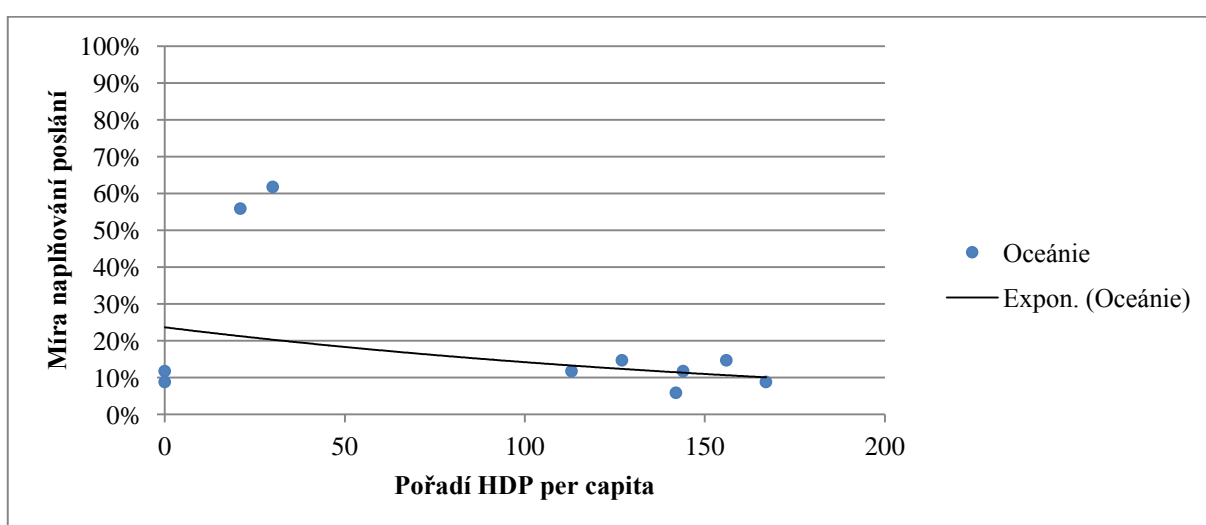
<sup>126</sup> Úřady těchto zemí nenabízí tyto země téměř žádné nástroje. Kromě Norska, které Umožňuje využití API a dalších 2 formátů nejsou ani příliš vstřícné k poskytování dat.

<sup>127</sup> Výše zmíněné země využívají aplikace DevInfo, Malajsie poskytuje statistiky v aplikaci vyvinutou Tableau software.

S výjimkou Singapuru se navíc nejedná o státy, které byly v minulosti součástí britského impéria. Tyto státy nabízí vždy alespoň dva způsoby grafické reprezentace dat. Jedná se vždy o nástroj na explorační analýzu dat spojený s datovým archivem, dále o infografiky a aplikace GIS, které umožňují prohlížení statistik v mapách.

#### 5.2.4.5. Oceánie

Země, které byly zařazeny pod označení Oceánie, potvrzují danou hypotézu. Statistické úřady malých ostrovních států v Tichém oceánu s nízkým HDP naplňují své poslání prostřednictvím webové prezentace pouze velmi omezeně a míra naplnění zvolených kritérií nepřesahuje jednu pětinu.



Graf 8

V kontrastu s tímto stavem jsou státy s výkonnými ekonomikami – Austrálie a Nový Zéland. Statistické úřady těchto států naplňují předpoklad souvislosti otevřenosti statistických úřadů a výkonné ekonomiky. Prostřednictvím webových stránek naplňují z více než 50 procent poslání statistických úřadů v podobě služby veřejnosti. Zejména statistický úřad Nového Zélandu publikuje řadu vizualizací a infografik na témata, která jsou pro tamní společnost důležitá – jako je stárnutí obyvatelstva, inflace a soužití s původním obyvatelstvem. Austrálie se zaměřuje na vytváření interaktivního obsahu a vzdělávacích mobilních aplikací založených na datech ze sčítání lidu<sup>129</sup>. Obě tyto země jsou podobně jako většina dalších v tomto regionu součástí Commonwealthu.

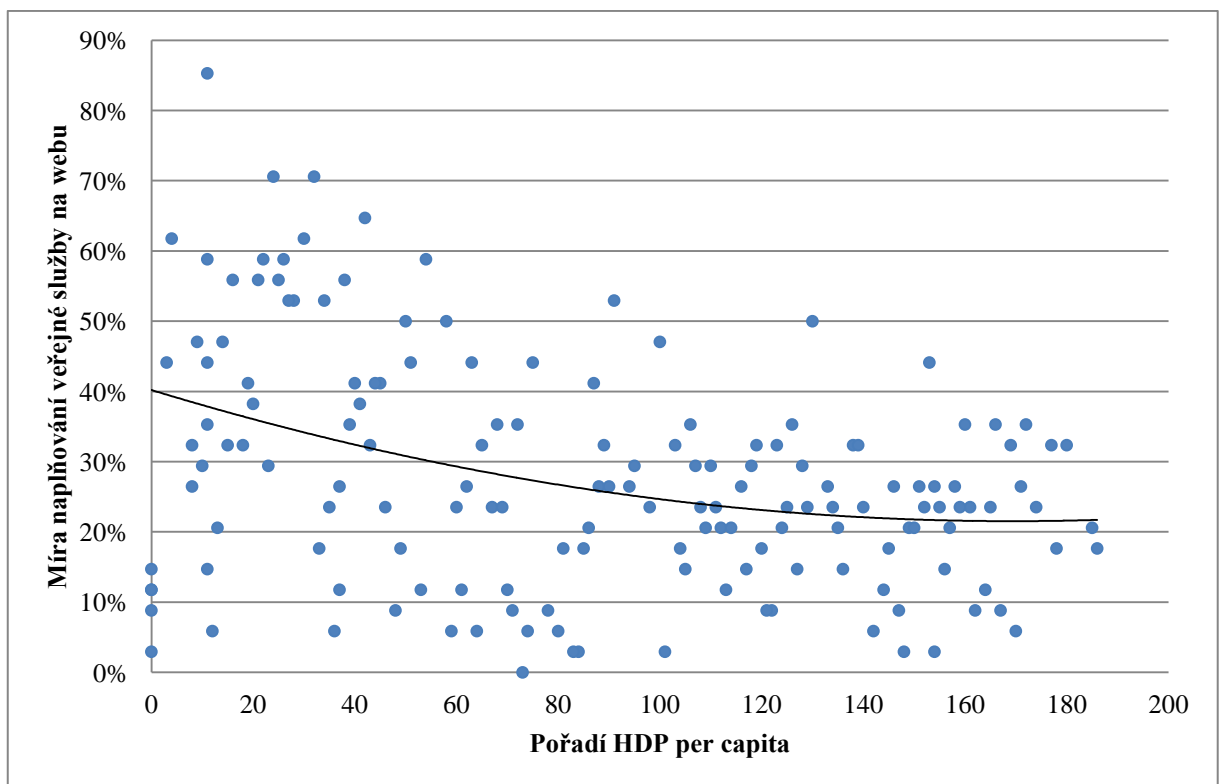
<sup>128</sup> Omán, Japonsko, Singapur

<sup>129</sup> Run that town

### 5.2.5. Revize hypotézy

Hypotézu, tak jak byla v původním znění nastavena, se nepodařilo průkazně ověřit. Neplatí tedy, že státy s výkonnou ekonomikou zároveň disponují statistickými úřady, které poskytují veřejnosti co nejvíce možností nahlížet do statistik, a to jak prostřednictvím typických produktů, datových kolekcí v dále zpracovatelných formátech dat, tak i prostřednictvím vizualizací.

Přesto je možné na základě předchozí analýzy a následujícího souhrnného grafu, jímž byla proložena spojnice trendu, říci, že platí tvrzení, které zní: Lze potvrdit existenci trendu, který ukazuje, že společně se snižující se výší HDP bude zároveň klesat míra naplňování veřejné služby statistických úřadů prostřednictvím jejich webových prezentací. Tento stav ale ovlivňuje poskytování nástrojů, které byly vytvořeny mezinárodními organizacemi, jak je zřejmé v případě statistických úřadů zemí Afriky.



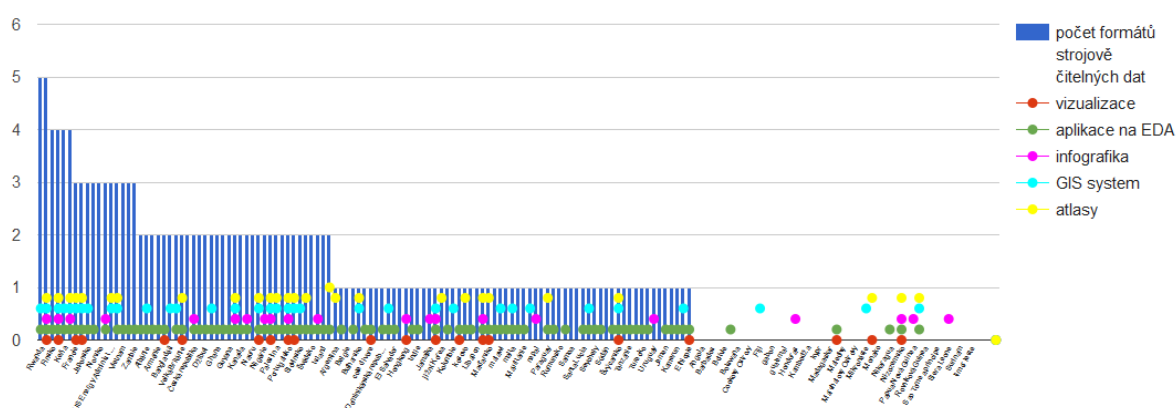
Graf 9

### 5.3. Ověření hypotézy 3

#### vztah počtu typů vizuálních reprezentací dat a počtu strojově čitelných datových formátů

Podle této hypotézy koreluje vyšší počet vizuálních forem prezentace dat produkovaných statistickými úřady s širší nabídkou strojově čitelných dat, tedy dat, které se používají mimo jiné k vytvoření vizuální reprezentace, než je obvyklé.

Pro její ověření byl vytvořen multidimenzionální graf, který ukazuje vztah počtu strojově čitelných datových formátů, jež jednotlivé statistické úřady na svých webových stránkách nabízejí, reprezentovaný sloupci, a barevně vyznačených sledovaných typů vizuálních reprezentací dat. Tento podklad (plná velikost v příloze 3) slouží jako hlavní zdroj pro ověření dané hypotézy. Prostřednictvím tohoto grafu bylo ověřeno, zda úřady, které nabízejí větší, než průměrné množství strojově čitelných datových formátů také zároveň nabízejí větší než průměrné množství různých typů sledovaných vizuálních reprezentací dat. Za sledované vizuální reprezentace dat jsou považovány vizualizace, infografiky, aplikace na explorační analýzu dat, kartografické zobrazení neboli atlasy a aplikace využívající na zobrazení dat geografický informační systém (GIS).



Graf 10

Předmětem zkoumání jsou webové stránky statistických úřadů, na nichž je možné najít data v nejméně polovině ze zvolených strojově čitelných formátů – tedy alespoň tři formátů z XLS, CSV, XML, JSON a prostředí API. Celkem se jedná o 17 webových stránek statistických úřadů představujících jednu desetinu celkového počtu zkoumaných prezentací. Lze o nich říci, že jde téměř z poloviny o statistické úřady z Evropy a s výjimkou Norska také o členské státy EU<sup>130</sup>. HDP na obyvatele těchto zemí je na vyšší než 50. pozici v pořadí států světa<sup>131</sup>.

Další čtvrtina webových prezentací patří úřadům afrických států. U všech těchto prezentací lze mimo další formy vizuální prezentace dat najít odkaz na národní mutaci datového archivu

<sup>130</sup> Norsko, Lucembursko Norsko Polsko Francie Irsko Finsko Itálie

<sup>131</sup> Viz elektronická příloha

NADA a některého z nástrojů na explorační analýzu dat, jež byly vytvořeny mezinárodními organizacemi<sup>132</sup>.

V rámci Severní, Jižní a Střední Ameriky jsou jediným státem, jehož statistické úřady poskytují na webových stránkách tři a více strojově čitelných datových formátů, Spojené státy americké. V segmentu stránek, jež bude dále zkoumán, se nachází Úřad pro Census, Úřad pro pracovní statistiky a Úřad pro energetickou administrativu. Z asijských států se jedná o webové stránky úřadů ve Vietnamu, Ománu a Japonsku, z nichž pouze Omán využívá datový archiv vytvořený na opensource platformě.

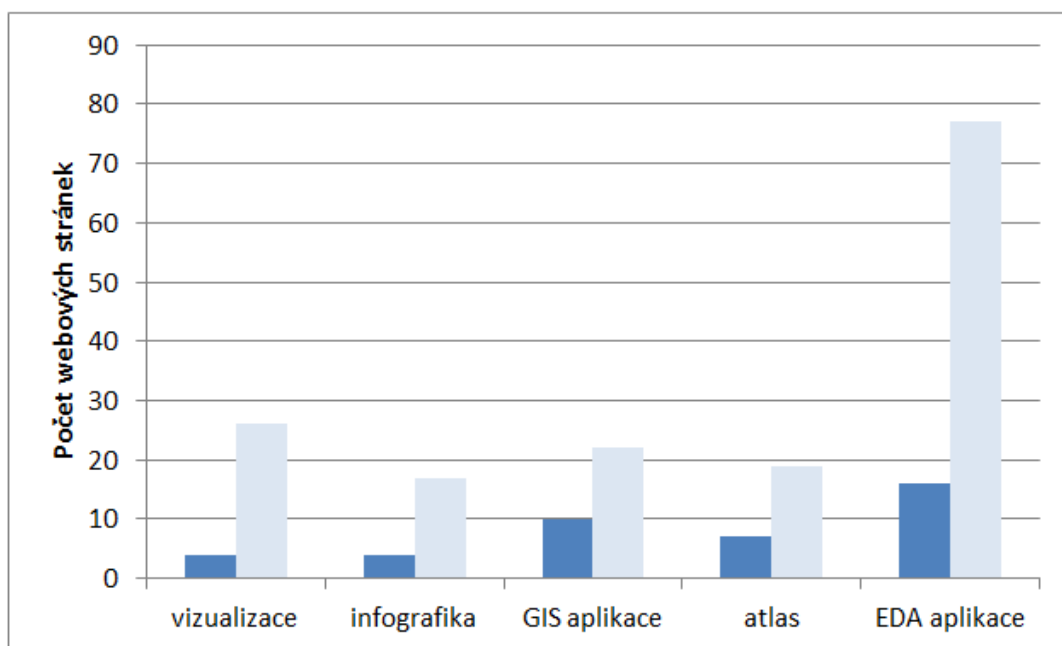
Dle předchozí charakteristiky je zjevné, že skupina států s velkým množstvím datových formátů se překrývá pouze částečně se skupinou statistických úřadů, které naplňují poslání v oblasti komunikování statistik v nejvyšší možné míře.

#### **5.3.4. Distribuce jednotlivých typů grafické reprezentace dat**

V této skupině zkoumaných webových stránek se koncentruje nejvyšší nasycenost typů grafických reprezentací dat. S výjimkou Norska disponují všechny webové prezentace těchto statistických úřadů alespoň jedním ze zkoumaných typů vizuální reprezentace dat. Více než polovina států potom disponuje aplikací založenou na geografickém informačním systému a atlasech. Nejméně jsou mezi těmito státy využívány vizualizace a infografiky. Následující graf ukazuje počet jednotlivých typů grafických zobrazení v poměru k celkovému počtu všech webových stránek, na nichž je možné daný typ najít.

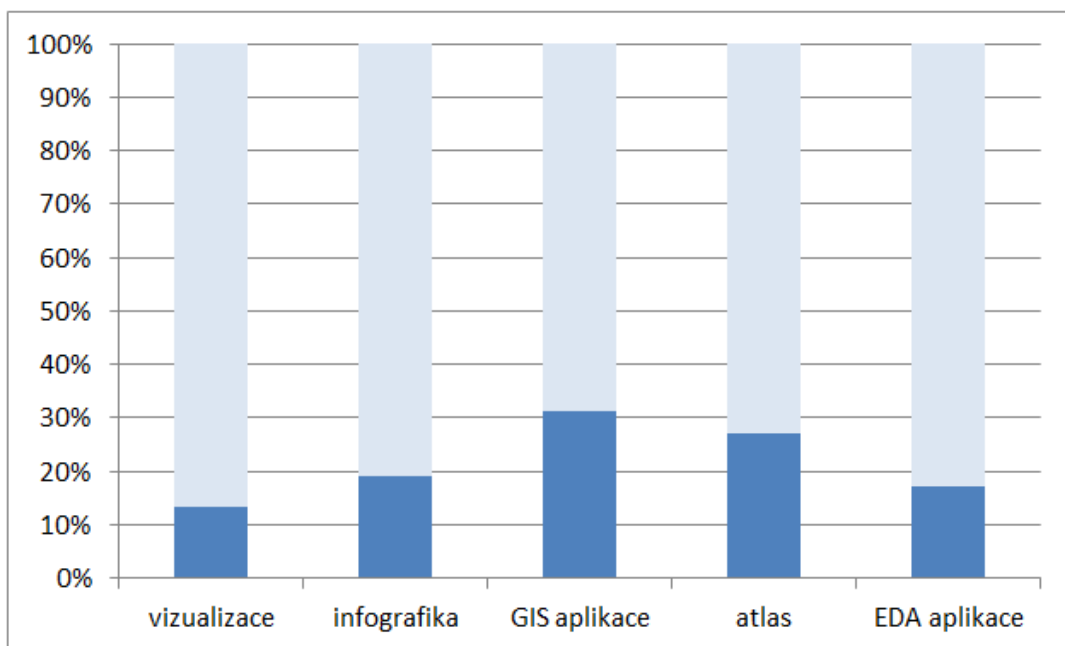
---

<sup>132</sup> Africadata, Countrystat, Devinfo



Graf 11

Na dalším grafu jsou zřetelné procentuální poměry, které ukazují, že 10 procent zkoumaných webů statistických úřadů na svých webových stránkách koncentruje třetinu všech aplikací využívajících GIS, 25 procent kartografických zobrazení, pětinu infografik, šestinu vizualizací a 18 procent aplikací na explorační analýzu dat. Ač má tento typ největší absolutní zastoupení, v celkovém kontextu to představuje pouze necelou pětinu, je tedy hojně využíván i statistickými úřady, jenž nejsou součástí této skupiny.



Graf 12

Celkem 17 států, které tvoří 10 procent celkového počtu zkoumaných webů splňujících podmínku poskytování alespoň 3 strojově čitelných datových typů, koncentruje dohromady čtvrtinu využití všech zkoumaných datových typů.

Při exploraci distribuce jednotlivých typů vizuální reprezentace dat je zřejmé, že 88 procent webových stránek v této skupině disponuje aplikací na explorační analýzu dat. Více než 60 procent poskytuje aplikace fungující na bázi GIS, více než 40 procent poskytuje kartografické zobrazení. Vizualizace a infografiky potom využívá 25 procent webů.

### 5.3.5. Porovnání se statistickými úřady, které disponují méně než polovinou různých strojově čitelných datových formátů

Ve skupině webových prezentací, v nichž statistické úřady nabízejí méně než polovinu statistickými úřady běžně používaných strojově čitelných dat, se předpokládá, že poskytují méně typů grafických reprezentací dat než skupina, která stanovený počet formátů poskytuje. Z následujících tabulek je zřejmé, že tento předpoklad je platný.

počet typů	počet webových stránek, které podmínku splňují	procentuální vyjádření
alespoň 1	16	94,12%
alespoň 2	11	64,71%
alespoň 3	6	35,29%
více	5	29,41%

Tabulka 1

počet typů	počet webových stránek, které podmínku splňují	procentuální vyjádření
alespoň 1	92	64,34%
alespoň 2	38	26,57%
alespoň 3	13	9,09%
více	10	6,29%

Tabulka 2

Ačkoli je platnost hypotézy zřejmá, existuje v této skupině celkem 13 outlierů, kteří navzdory hypotéze disponují větším počtem typů grafických reprezentací, než bylo očekáváno. Tyto statistické úřady jsou součástí skupiny úřadů, které na webových stránkách naplňují své poslání z více než 50 procent, přestože nedisponují vyšším než průměrným počtem typů běžně nabízených strojově čitelných formátů dat. Tyto státy jsou uvedeny v následující tabulce:

Úřad státu	počet typů graf. reprezentací dat	počet typů strojově čitelných dat
Chorvatsko	2	4
Německo	2	4
Nový Zéland	2	5
Portugalsko	2	5
Singapur	2	4
Palestina	2	3
Velká Británie	2	3
JAR	1	4
Maďarsko	1	4
Litva	1	3
Švýcarsko	1	4
Rakousko	1	4

Tabulka 3

### 5.3.6.Závěr:

Přestože se ukázala platnost hypotézy, která tvrdí, že úřady poskytující více než průměrný počet typů strojově čitelných typů dat zároveň nabízí také více typů grafické reprezentace dat, není možné zároveň říci, že platnost hypotézy funguje i opačným směrem. Tedy neplatí, že úřady, které nabízejí vyšší, než průměrný počet typů grafických zobrazení dat, zároveň nabízí i nadprůměrný počet strojově čitelných datových formátů.

Statistické úřady, jejichž webové prezentace nabízejí nadprůměrný počet vizuálních reprezentací dat a naplňují službu veřejnosti ve stanovených kritériích z více než 50 procent, nekladou dle zjištění vždy takový důraz na rozmanitost datových formátů, ale spíše volí



formáty, které jsou běžné a tedy i pro uživatele s omezenou informační gramotností dostupné, a nevytvářejí zbytečnou bariéru. Tyto úřady se soustředí spíše na rozmanitost nástrojů, jimiž je možné data získat vhléd do dat a porozumět statistikám, což je také jedna ze zásadních funkcí statistických úřadů.

## 6. Závěr:

V první části této práce vysvětluji specifika vizuální reprezentace dat směrem k lidskému vnímání, definuji poslání statistických úřadů, uvádím historický kontext a následně vysvětluji důvody, proč se statistické úřady rozhodly pro využívání vizualizací jako nástroje pro komunikování statistik a poskytování vhledu do veřejně dostupných dat.

Prezentuji vizualizace jako nástroj k tvoření informací a vnímání informací, čímž jej považuji za vhodný způsob zprostředkování veřejně dostupných dat veřejnosti. V poslední části popisuji, cílovou skupinu příjemců tohoto způsobu prezentování veřejně dostupných informací a shrnuji dosavadní zkušenosti statistických úřadů s poskytováním a komunikováním veřejně dostupných dat a statistik vizuální formou. A to jak prostřednictvím vizualizací, tak i ostatních nástrojů, které statistické úřady pro tyto účely využívají.

V druhé části práce věnuji kvantitativní analýze webových stránek jednotlivých statistických úřadů, který byl realizován za účelem zjištění stavu naplňování poslání jednotlivých úřadů směrem k veřejnosti a ověření, v jaké míře jednotlivé statistické úřady na svých webových stránkách využívají možnosti vizuální reprezentace dat. Prostřednictvím jednotlivých hypotéz potom byly ověřeny vybrané souvislosti.

Výsledky, ke kterým jsem došla, ukázaly, že míra naplňování poslání statistických úřadů v poskytování veřejně dostupných dat prostřednictvím webových stránek je v celosvětovém kontextu poměrně malá a u úřadů většiny států nedosahuje ani poloviny ze stanovených kritérií. Zjistila jsem také, že úřady, které poskytují dané položky v alespoň 50 procentech, zároveň nabízejí širší škálu typů nástrojů vizuální reprezentace dat než ty úřady, které tuto podmínku nesplňují.

Při ověřování druhé hypotézy jsem vyvrátila předpoklad přímé souvislosti mezi otevřeností statistických úřadů, a tedy mimo jiné poskytování veřejných dat v různých formách, a otevřenosti, potažmo výkonnosti ekonomiky dané země. Neplatí tak, že státy, které si mohou dovolit více investovat do statistické služby, mají zároveň tuto službu více otevřenou veřejnosti. mezinárodní dohody, vůči kterým se státy zavázaly (European Statistical System Committee, 2011), pozůstatek vlivu z historie (Velká Británie a Commonwealth) a aktivita mezinárodních organizací poskytující vybraným statistickým úřadům již hotové nástroje.

Podářilo se nicméně ověřit hypotézu, která dávala do souvislosti vyšší než průměrný počet poskytovaných formátů strojově čitelných dat, jakožto typu dat, z něž mohou být tvořeny právě vizuální reprezentace dat a zvýšenou pravděpodobnost používání právě zkoumaných typů vizuální reprezentace jako je kartografické zobrazení, vizualizace, infografika, aplikace na exploratorní analýzu dat a aplikace využívající geografický informační systém.

Pomocí obou částí se tak, navzdory limitům například v podobě nedostupnosti některých webových prezentací statistických úřadů, podařilo zmapovat současný stav fenoménu využívání vizuální reprezentace dat statistickými úřady ke komunikování statistik a jako prostředek nabízení veřejně dostupných informací a myslím si, že si našel své opodstatnění jako analýza fenoménu, jehož význam a oblíba mezi statistickými úřady stoupá<sup>133</sup>. Další výzkumy na toto téma by se mohly zaměřit například na tematickou analýzu různých typů vizuálních reprezentací dat nebo na souvislost prezentování dat touto formou a míry gramotnosti obyvatelstva daného státu.

---

<sup>133</sup> Z dat, která se nepoužila, vyplývá, že roste počet těch vizualizací, které se dělají.

## 7. Seznam použité literatury

2006. Data Communication - Emerging International Trends and Practices of the Australian Bureau of Statistics. Australian Bureau of Statistics [online]. Available from: <http://www.abs.gov.au/ausstats/abs@.nsf/2f762f95845417aeca25706c00834efa/13154B524A63CF45CA25711000213162?opendocument> [Accessed 2015-12-16]

2012. Visualizations [online]. Available from: <http://www.census.gov/data/visualizations.html> [Accessed 2015-12-16]

2013. England and Wales, 1961-2087. Www.ons.gov.uk [online]. Available from: <http://www.ons.gov.uk/ons/interactive/england-and-wales-population-pyramid---dvc2/index.html> [Accessed 2015-12-16]

2013. Interactive Content. Istat [online]. Available from: <http://www.istat.it/en/products/interactive-contents> [Accessed 2015-12-16]

2013. Noi Italia. Www.istat.it [online]. Available from: <http://noi-italia.istat.it/> [Accessed 2015-12-16]

2013. Preis Kaleidoskop. Www.destatis.de [online]. Available from: <https://www.destatis.de/Voronoi/PreisKaleidoskop.svg> [Accessed 2015-12-16]

2013. Preis Kaleidoskop. Www.destatis.de [online]. Available from: <https://www.destatis.de/DE/ZahlenFakten/Indikatoren/Konjunkturindikatoren/Preismonitor/Preismonitor.html> [Accessed 2015-12-16]

2013. Quickstats. US Census [online]. Available from: <http://www.census.gov/quickfacts/table/PST045215/00> [Accessed 2015-12-16]

2015. About Statistics Estonia [online]. Available from: <http://www.stat.ee/about> [Accessed 2015-12-16]

2015. Data vizualizace. Www.csb.gov.lv/ [online]. Available from: [www.csb.gov.lv/en/dati/data-visualisation-42751.html](http://www.csb.gov.lv/en/dati/data-visualisation-42751.html) [Accessed 2015-12-16]

2015. Veřejná database [online] Available from: <http://vdb.czso.cz/> [Accessed 2015-12-16]

Albers, M. Infographics and Communicating Complex Information. 2015 IN: [online]. pp. 267-280. Available from: [http://link.springer.com/10.1007/978-3-319-20898-5\\_26](http://link.springer.com/10.1007/978-3-319-20898-5_26) [Accessed 2015-12-16]

Albright, J. J., & Lyle, J. A.. 2010. Data Preservation through Data Archives. PS: Political Science and Politics, 43(1), 17–21. Retrieved from <http://www.jstor.org/stable/25699287> [cit. 2015-12-16]

Alexa, Chapter 3: Data, Information and Meaning [online]. Disertační práce. Available from: <http://repository.up.ac.za/bitstream/handle/2263/27367/03chapter3.pdf> [Accessed 2015-12-16]

Alldritt,, R., Young, R. & Laux,, R. 2008. Independence for UK official statistics: the new UK Statistics Authority.

Azzam, T., Evergreen, S., Germuth, A. & Kistler, S.. 2013. Data Visualization and Evaluation. New Directions for Evaluation [online], vol. 2013, issue 139, pp. 7-32. Available from: <http://doi.wiley.com/10.1002/ev.20065> [cit. 2015-12-16]

Beniger, J. 1978. Review. Contemporary Sociology [online], vol. 7, no. 1, pp. 64-65. Available from: <http://www.jstor.org/stable/2065930> [cit. 2015-12-16]

Beranová, B. 2011. Statistika musí být objektivní a nezávislá, jinak postrádá smysl. Statistika a my [online], vol. 2011, 11-12, pp. 4-7. Available from: [https://www.czso.cz/documents/10180/20541267/180411111204\\_06.pdf/51cd8da2-eed9-4f26-8344-a77a537be1ea?version=1.0](https://www.czso.cz/documents/10180/20541267/180411111204_06.pdf/51cd8da2-eed9-4f26-8344-a77a537be1ea?version=1.0) [cit. 2015-12-16]

Bertin, J. 1967. Semiology of graphics: diagrams, networks, maps, Překlad William J Berg. 1st ed. Redlands, Calif.: ESRI Press.

Bohatá, M. 2011. The Role of the Official Statistics in the Context of the Global Crisis. Statistika, vol. 48, no. 1, pp. 5-10.

Buckland, M. 1991. Information as Thing. Journal of the American Society for Information Science [online], vol. 42, issue 5, pp. 351-360. Available from:

<http://skat.ihmc.us/rid=1KR7VC4CQ-SLX5RG-5T39/BUCKLAND%281991%29-informationasthing.pdf> [cit. 2015-12-16]

Card, S., Mackinlay, J. & Shneiderman, B. c1999. Readings in information visualization: using vision to think. Third edition. San Francisco, Calif.: Morgan Kaufmann Publishers.

Cieslar, J. Když nejste nezávislý, pak nejste důvěryhodný. Statistika a my [online]. Available from: <https://www.czso.cz/csu/czso/510046c991> [Accessed 2015-12-16]

Clark, W., Avery, K., Batcheller, A., Bowker, G. & Borgman, C.. 1976. The Effects of Data Aggregation in Statistical Analysis. Geographical Analysis [online], vol. 8, issue 4, pp. 428-438. Available from: <http://doi.wiley.com/10.1111/j.1538-4632.1976.tb00549.x> [cit. 2015-12-16]

Cook, L.. 2003. Editorial: The National Statistician: 2 Years On. Journal of the Royal Statistical Society. Series A (statistics in Society), 166(1), 1–4. Retrieved from <http://www.jstor.org/stable/3559824>

CountryStat 2013. [online]. Available from: <http://www.countrystat.org/> [Accessed 2015-12-16]

Cover, T. & Thomas, J. c2006. Elements of information theory: making sense of the field [online]. 2nd ed. Hoboken, N.J.: Wiley-Interscience. [Accessed 2015-12-16]

Černý, M. 2011. Vizualizace – cesta k vyšší atraktivitě veřejné databáze. Statistika a my [online], vol. 2011, no. 5, pp. 16-18. Available from: [https://www.czso.cz/documents/10180/20541251/1804110516\\_17.pdf/0755030b-8c7c-4fb1-9000-cfd5017771a2?version=1.0](https://www.czso.cz/documents/10180/20541251/1804110516_17.pdf/0755030b-8c7c-4fb1-9000-cfd5017771a2?version=1.0) [cit. 2015-12-16]

Česká republika. 1995. Zákon o státní statistické službě 89/1995 Sb. [online], pp. 1-26. Available from: [https://www.czso.cz/documents/10180/25163201/zakon\\_c\\_89\\_1995\\_sb\\_20150101.pdf/dadc8367-72bc-463f-872e-d1acf31baec1?version=1.0](https://www.czso.cz/documents/10180/25163201/zakon_c_89_1995_sb_20150101.pdf/dadc8367-72bc-463f-872e-d1acf31baec1?version=1.0) [cit. 2015-12-16]

Český statistický úřad. 2011. Mise, vize s priority ČSÚ. [online], pp. 1. Available from: <http://csugeo.i->

server.cz/csu/redakce.nsf/i/mise\_vize\_a\_priority\_csu/\$File/Mise\_2012.pdf [cit. 2015-12-16]

Dataportal African Development Bank Group [online]. Available from: <http://dataportal.afdb.org/DataAnalysis.aspx>, [Accessed 2015-12-16]

Dataportal Genesis Statistisches Bundesamt [online]. Available from: <https://www-genesis.destatis.de/genesis/online/logon> [Accessed 2015-12-16]

De Veaux, R., Hand. 2005. How to Lie with Bad Data. *Statistical Science* [online], vol. 20, issue 3, pp. 231-238. Available from: <http://projecteuclid.org/Dienst/getRecord?id=euclid.ss/1124891289/> [cit. 2015-12-16]

DevInfo. 2015. [online]. <http://www.devinform.org/libraries.aspx/Home.aspx> [Accessed 2015-12-16]

Duspivová, K. 2012. Zlepšujeme mzdové statistiky. *Statistika a my* [online], vol. 2, issue 6, pp. 38. Available from: <https://www.czso.cz/documents/10180/20555411/1804120638.pdf/d1078951-199b-40b5-b02f-45acde2c447d?version=1.0> [cit. 2015-12-16]

Dykes, J. 1998. Cartographic visualization: exploratory spatial data analysis with local indicators of spatial association using Tcl/Tk and cd. *The Statistician*, vol. 47, no. 3, pp. 485-497.

Edward R. Tufte. 1990. *Envisioning information*. 11th printing, Nov. 2006. Cheshire, Conn: Graphics Press.

Edwards, P. N., Mayernik, M. S., Batcheller, A. L., Bowker, G. C., & Borgman, C. L.. 2011. Science friction: Data, metadata, and collaboration. *Social Studies of Science*, 41(5), 667–690. Available from: <http://www.jstor.org/stable/41301955> [cit. 2015-12-16]

Edwards, P., Mayernik, M., Batcheller, A., Bowker, G. & Borgman, C.. 2011. Science friction. *Social Studies of Science* [online], vol. 41, issue 5, pp. 667-690. Available from: <http://sss.sagepub.com/cgi/doi/10.1177/0306312711413314> [cit. 2015-12-16]

Estonian Statistics. 2015. *Strategy of Statistics Estonia 2015–2020* [online]. Estonsko, 2015 [cit. 2015-12-16]. Dostupné z: <http://www.stat.ee/strategy>

Estonko. 2011. Official Statistics Act. Statistika a my [online]. Available from:  
<https://www.riigiteataja.ee/en/eli/506012015002/consolide> [Accessed 2015-12-16]

European Statistical system Committee. 2011. European Statistic code of Practice [online]. Brusel: EU. [Accessed 2015-12-16]

Extensible Markup Language (XML). [online]. Available from:  
<http://www.w3.org/XML/> [Accessed 2015-12-16]

Few, S. 2004. Show Me the Numbers. Designing Tables and Graphs to Enlighten. Analytics Press, Oakland, CA, 2004

Few, S. 2007 Sticky stories told with numbers, Infographics and Communicating Complex Information. Perceptual Edge [online], vol. 2007, February, pp. 1-6. Available from:  
[https://www.perceptualedge.com/articles/visual\\_business\\_intelligence/sticky\\_stories\\_with\\_nums.pdf](https://www.perceptualedge.com/articles/visual_business_intelligence/sticky_stories_with_nums.pdf) [cit. 2015-12-16]

Few, S. 2008. What Ordinary People Need Most from Information Visualization Today. Perceptual Edge [online], vol. 2008, August, pp. 1-7. Available from:  
[http://www.perceptualedge.com/articles/visual\\_business\\_intelligence/what\\_people\\_need\\_from\\_infovis.pdf](http://www.perceptualedge.com/articles/visual_business_intelligence/what_people_need_from_infovis.pdf) [cit. 2015-12-16]

Few, S. 2008a. Solutions to the Problem of Over-Plotting in Graphs. Perceptual Edge [online], vol. 2008, September/October, pp. 1-6. Available from:  
[https://www.perceptualedge.com/articles/visual\\_business\\_intelligence/overplotting\\_in\\_graphs.pdf](https://www.perceptualedge.com/articles/visual_business_intelligence/overplotting_in_graphs.pdf) [cit. 2015-12-16]

Few, S. 2009. Statistical Narrative Telling Compelling Stories with Number. Perceptual Edge [online], vol. 2009, July/August, pp. 1-10. Available from:  
[https://www.perceptualedge.com/articles/visual\\_business\\_intelligence/statistical\\_narrative.pdf](https://www.perceptualedge.com/articles/visual_business_intelligence/statistical_narrative.pdf) [cit. 2015-12-16]

Few, S. 2011. The Chart Junk Debate. Perceptual Edge [online], vol. 2011, no. 1, pp. 1-11. Available from:  
[https://www.perceptualedge.com/articles/visual\\_business\\_intelligence/the\\_chartjunk\\_debate.pdf](https://www.perceptualedge.com/articles/visual_business_intelligence/the_chartjunk_debate.pdf) [cit. 2015-12-16]



- Fienberg, S. 1979. Graphical Methods in Statistics. *The American Statistician* [online], vol. 33, issue 4, pp. 165-. Available from:  
<http://www.jstor.org/stable/2683729?origin=crossref> [cit. 2015-12-16]
- Floridi, L. 2005. Is Semantic Information Meaningful Data?. *Philosophy and Phenomenological Research* [online], vol. 70, issue 2, pp. 351-370. Available from:  
<http://doi.wiley.com/10.1111/j.1933-1592.2005.tb00531.x> [cit. 2015-12-16]
- Floridi, L. 2009. Philosophical Conceptions of Information. IN: [online]. pp. 13-53. Available from: [http://link.springer.com/10.1007/978-3-642-00659-3\\_2](http://link.springer.com/10.1007/978-3-642-00659-3_2) [Accessed 2015-12-16]
- Floridi, L. 2010. Information: a very short introduction [online]. 1st pub. Oxford: Oxford University Press. [Accessed 2015-12-16]
- Forbes, S. 2010. Data visualization and official statistics: Providing new evidence and enhancing understanding. [online], pp. 1-6. Available from:  
[https://www.bis.org/ifc/events/2011\\_dublin\\_48\\_04\\_forbes.pdf](https://www.bis.org/ifc/events/2011_dublin_48_04_forbes.pdf) [cit. 2015-12-16]
- Friendly, M. & Kwan, E.. 2003. Effect ordering for data displays. *Computational Statistics & Data Analysis* [online], vol. 43, issue 4, pp. 509-539. Available from:  
<http://linkinghub.elsevier.com/retrieve/pii/S0167947302002906> [cit. 2015-12-16]
- Friendly, M. 1999. Visualizing Categorical Data. In: Sirken, M., Mackinlay, J. & Shneiderman, B. c1999. *Cognition and survey research: using vision to think*. Third edition. New York: Wiley.
- Friendly, M. 2005. Milestones in the History of Data Visualization. IN: *Classification — the Ubiquitous Challenge* [online]. Berlin/Heidelberg: Springer-Verlag, pp. 34. Available from: [http://link.springer.com/10.1007/3-540-28084-7\\_4](http://link.springer.com/10.1007/3-540-28084-7_4) [Accessed 2015-12-16]
- Friendly, M. 2008. The Golden Age of Statistical Graphics. *Statistical Science* [online], vol. 23, issue 4, pp. 502-535. Available from:  
<http://projecteuclid.org/euclid.ss/1242049392> [cit. 2015-12-16]

- Friendly, M. c1999. Visualizing Categorical Data. IN: Cognition and survey research [online]. New York: Wiley, pp. 319-348. Available from:  
<http://datavis.ca/papers/casm/casm.pdf> [Accessed 2015-12-16]
- Gareau, M., Keegan, R. & Wang, L.. 2013. An Exploration of the Effectiveness of Infographics in Contrast to Text Documents for Visualizing Census Data. Statistical Journal of the IAOS [online], vol. 29, no. 3, pp. 161-172. Available from:  
[http://link.springer.com/10.1007/978-3-319-20612-7\\_16](http://link.springer.com/10.1007/978-3-319-20612-7_16) [cit. 2015-12-16]
- Gébllová, A. 2012. Zásadní je důvěryhodnost statistik. Statistika a my [online], vol. 2012, no. 3, pp. 5. Available from:  
<https://www.czso.cz/documents/10180/20555413/1804120305.pdf/775f32b4-6a46-4380-8189-5dbfb1038af7?version=1.0> [cit. 2015-12-16]
- Gébllová, A. 2013. Informační služby v roce 2012. Statistika a my [online], vol. 2013, no. 1, pp. 16-18. Available from: <https://www.czso.cz/csu/czso/2d0035b1ff> [cit. 2015-12-16]
- Girard, J., Allison, M. 2008. Information Anxiety: Fact, Fable or Fallacy. The Electronic Journal of Knowledge Management Volume 6 Issue 2, pp. 111 - 124, Available from: [www.ejkm.com](http://www.ejkm.com) [cit. 2015-12-16]
- Good, I. J.. 1983. The Philosophy of Exploratory Data Analysis. Philosophy of Science 50 (2). [University of Chicago Press, Philosophy of Science Association]: 283–295. Available from: <http://www.jstor.org/stable/188015> [Accessed 2015-12-16]
- Greenfield, T. Communicating Statistics. [online]. Available from:  
<http://www.jstor.org/stable/2982733?origin=crossref> [cit. 2015-12-16]
- Hallisey, E. 2005. Cartographic Visualization. The Professional Geographer [online], vol. 57, issue 3, pp. 350-364. Available from:  
<http://www.tandfonline.com/doi/abs/10.1111/j.0033-0124.2005.00483.x> [cit. 2015-12-16]
- Hardt, M. 2004. Aesthetics, Semiotics & Design. Bergen National Academy of the Arts, pp. 1-31.

- Harris, T., Mayernik, M., Batcheller, A., Bowker, G. & Borgman, C.. 2003. Data Models and the Acquisition and Manipulation of Data. *Philosophy of Science* [online], vol. 70, issue 5, pp. 1508-1517. Available from: <http://www.jstor.org/stable/10.1086/377426> [cit. 2015-12-16]
- Healey, C. G., Enns, J. T. 2012. Attention and Visual Memory in Visualization and Computer Graphics. [online] *IEEE Transactions on Visualization and Computer Graphics* **18**, 7, 1170–1188 [cit. 2015-12-16]
- Healey, Ch. G., et al. 2008. Perception in visualization. Retrieved February, 2007, 10, pp. 1-21. Available from: [http://portal.sbgrid.org/devel/viz/CS171/www-csc-ncsu-edu\\_faculty\\_healey\\_PP\\_index-html\\_0zufsm22.pdf](http://portal.sbgrid.org/devel/viz/CS171/www-csc-ncsu-edu_faculty_healey_PP_index-html_0zufsm22.pdf) [cit. 2015-12-16]
- Hey, J. 2004.,. The Data, Information, Knowledge, Wisdom Chain:The Metaphoricallink
- Hey, T. & Trefethen, A.. 2003. The Data Deluge. *Grid Computing* [online], pp. 809. Available from: <http://doi.wiley.com/10.1002/0470867167.ch36> [cit. 2015-12-16]
- Hjørland, B. 2007. Information. *Journal of the American Society for Information Science and Technology* [online], vol. 58, issue 10, pp. 1448-1456. Available from: <http://doi.wiley.com/10.1002/asi.20620> [cit. 2015-12-16]
- Holý, V. 2011. Jak a proč chráníme důvěrné statistické údaje. *Statistika a my* [online], vol. 2011, 11/12, pp. 26-28. Available from: [https://www.czso.cz/documents/10180/20541267/18041111226\\_27.pdf/de5198cc-b1dd-4b71-a95c-62fa3140b335?version=1.0](https://www.czso.cz/documents/10180/20541267/18041111226_27.pdf/de5198cc-b1dd-4b71-a95c-62fa3140b335?version=1.0) [cit. 2015-12-16]
- Chambers, J. 1993. Greater or lesser statistics. *Statistics and Computing* [online], vol. 3, issue 4, pp. 182-184. Available from: <http://link.springer.com/10.1007/BF00141776> [cit. 2015-12-16]
- [Checkland, P](#) & Holwell. 2006. Data, capta, information and knowledge. in *Introducing Information Management: the business approach*. Elsevier, London, New York and Amsterdam, pp. 47-55
- Chen, M. & Floridi, L.. 2013. An analysis of information visualisation. *Synthese* [online], vol. 190, issue 16, pp. 3421-3438. Available from: <http://link.springer.com/10.1007/s11229-012-0183-y> [cit. 2015-12-16]

Chen, M., Floridi, L. & Borgo, R. 2014. What Is Visualization Really For?. IN: The Philosophy of Information Quality [online]. Cham: Springer International Publishing, pp. 75-93. Available from: [http://link.springer.com/10.1007/978-3-319-07121-3\\_5](http://link.springer.com/10.1007/978-3-319-07121-3_5) [Accessed 2015-12-16]

Information on National Statistical Systems [online]. 2015. Available from: [http://unstats.un.org/unsd/methods/inter-natlinks/sd\\_natstat.asp](http://unstats.un.org/unsd/methods/inter-natlinks/sd_natstat.asp) [Accessed 2015-12-16]

Interaktiv & Anschaulich. Destatis - Statistisches Bundesamt [online]. Available from: <https://www.destatis.de/DE/Service/InteraktivAnschaulich/InteraktivAnschaulich.html;jsessionid=F85F0CBB48B771BF4CA463F780A85A8D.cae2> [Accessed 2015-12-16]

Kiekenbeck, J. 2015 Emailová komunikace s Destatis.de z 18.9. 2015. [online], pp. 1. Available from: [https://mail.volny.cz/?utm\\_source=volnyHP&utm\\_medium=mailbox](https://mail.volny.cz/?utm_source=volnyHP&utm_medium=mailbox) [cit. 2015-12-16]

Klauda, P. 2011. Site-Oriented Statistics and its Geoinformatic Potential. Statistika [online], vol. 48, no. 2, pp. 107-111. Available from: <https://www.czso.cz/documents/10180/20542065/180211q2107-110.pdf/2105001d-53c1-4f7e-8ac1-ba4a2fd04c86?version=1.0> [cit. 2015-12-16]

Klauda, P. 2011. Site-Oriented Statistics and its Geoinformatic Potential. Statistika [online], vol. 48, no. 2, pp. 107-111. Available from: <https://www.czso.cz/documents/10180/20542065/180211q2107-110.pdf/2105001d-53c1-4f7e-8ac1-ba4a2fd04c86?version=1.0> [cit. 2015-12-16]

Koláčková, H. 2011. Web je hlavně pro studenty zdrojem cenných údajů. Statistika a my [online], vol. 2011, 11-12, pp. 44-45. Available from: [https://www.czso.cz/documents/10180/20541267/180411111244\\_45.pdf/c9f01376-d9c5-4bd3-98bb-aa7137589b71?version=1.0](https://www.czso.cz/documents/10180/20541267/180411111244_45.pdf/c9f01376-d9c5-4bd3-98bb-aa7137589b71?version=1.0) [cit. 2015-12-16]

Konference <http://www.otevrenadata.cz>. 14.4.2014. Otevřená data: Chytřejší stát a lepší služby - YouTube. Youtube.com [online]. Available from: <https://www.youtube.com/watch?v=y0aqObi5AOA> [Accessed 2015-12-16]

Kullback., S. 1978. Information theory and statistics. [2. printing]. Gloucester, Mass: Peter Smith. Chen, C. c2006. Information visualization: beyond the horizon [online]. 2nd ed. London: Springer. [Accessed 2015-12-16]

- LANIGAN, R. 1994. *Capta versus data: Method and evidence in communicology*. *Human Studies* [online]. 1994, **17**(2): 285-285 [cit. 2015-12-16]. DOI: 10.1007/BF01323607. ISSN 0163-8548. Dostupné z: <http://link.springer.com/10.1007/BF01323607>
- Larkin, J. & Simon, H.. 1987-01-03. *Why a Diagram is (Sometimes) Worth Ten Thousand Words*. *Cognitive Science* [online], vol. 11, issue 1, pp. 65-100. Available from: <http://doi.wiley.com/10.1111/j.1551-6708.1987.tb00863.x> [cit. 2015-12-16]
- Laux, R. 2013. *UK Official Statistics - from New Public Management to Public Value*. [online]. Available from: [www.statisticsauthority.gov.uk%2Freports---correspondence%2Freports%2Fconference-papers%2Fuk-official-statistics---from-new-public-management-to-public-value.pdf](http://www.statisticsauthority.gov.uk%2Freports---correspondence%2Freports%2Fconference-papers%2Fuk-official-statistics---from-new-public-management-to-public-value.pdf) [cit. 2015-12-16]
- Leinhardt, S. & Wasserman, S.. 1979. *Exploratory Data Analysis*. *Sociological Methodology* [online], vol. 10, pp. 311-. Available from: <http://www.jstor.org/stable/270776?origin=crossref> [cit. 2015-12-16]
- Leonard, W. 1952. *Statistical Activities of the United Nations*. *The American Statistician* [online], vol. 6, issue 2, pp. 14-. Available from: <http://www.jstor.org/stable/2681712?origin=crossref> [cit. 2015-12-16]
- Mai, J. 2013. *The quality and qualities of information*. *Journal of the American Society for Information Science and Technology* [online], vol. 64, issue 4, pp. 675-688. Available from: <http://doi.wiley.com/10.1002/asi.22783> [cit. 2015-12-16]
- Makalouš, I. 2011. *Reducing Respondents' Burden in the Czech Statistical Office from Respondents' Point of View*. *Statistika* [online], vol. 48, issue 3, pp. 71-78. [cit. 2015-12-16]
- Mana, M. 2014. *ČSÚ = brána k vědeckému zkoumání*. [online], vol. 2012, no. 6, pp. 7. Available from: <https://www.czso.cz/documents/10180/20555411/1804120607.pdf/bb8f6e82-4e26-4dc9-809d-57f12030a8ae?version=1.0> [cit. 2015-12-16]
- Marcus, A. 2007 *Information Graphics: A Celebration and Recollection*. *Perceptual Edge* [online], vol. 2007, February, pp. 1-11. Available from:

<https://www.perceptualedge.com/articles/guests/Information%20Graphics.pdf> [cit. 2015-12-16]

Mašátová, M. 2012. Historie změn ve statistice. *Statistika a my* [online], vol. 2012, no. 2, pp. 10. Available from:

<https://www.czso.cz/documents/10180/20555419/1804120210.pdf/2ff9d4bd-fb6e-48ec-9a0a-a61c277d811c?version=1.0> [cit. 2015-12-16]

Mazza, R. c2009. *Introduction to information visualization* [online]. London: Springer. [Accessed 2015-12-16]

McLennan, B. 1995. You Can Count on Us--With Confidence. *Journal of the Royal Statistical Society. Series A (Statistics in Society)* [online], vol. 158, issue 3, pp. 467-. Available from: <http://www.jstor.org/stable/10.2307/2983441?origin=crossref> [cit. 2015-12-16]

McLennan, B. 1995. You Can Count on Us--With Confidence. *Journal of the Royal Statistical Society. Series A (Statistics in Society)* [online], vol. 158, issue 3, pp. 467-. Available from: <http://www.jstor.org/stable/10.2307/2983441?origin=crossref> [cit. 2015-12-16]

MCLENNAN. B. 2013 The role and position of the national statistical office in the overall government structure [online]. : 1-28 [cit. 2015-12-16]. Dostupné z: [http://unstats.un.org/unsd/methods/statorg/Workshops/Yangon/Session3\\_StatisticalSystems\\_McLennan.pdf](http://unstats.un.org/unsd/methods/statorg/Workshops/Yangon/Session3_StatisticalSystems_McLennan.pdf)

Moore, P. 1992. A National Statistical Commission. *Journal of the Royal Statistical Society. Series A (Statistics in Society)* [online], vol. 155, issue 1, pp. 5-. Available from: <http://www.jstor.org/stable/2982666?origin=crossref> [cit. 2015-12-16]

Moore, P. 1992. A National Statistical Commission. *Journal of the Royal Statistical Society. Series A (Statistics in Society)* [online], vol. 155, issue 1, pp. 5-. Available from: <http://www.jstor.org/stable/2982666?origin=crossref> [cit. 2015-12-16]

Moser, C. 1976. The Role of the Central Statistical Office in Assisting Public Policy Makers. *The American Statistician* [online], vol. 30, issue 2, pp. 59-. Available from: <http://www.jstor.org/stable/2683794?origin=crossref> [cit. 2015-12-16]

- Moser, C. 1976. The Role of the Central Statistical Office in Assisting Public Policy Makers. *The American Statistician* [online], vol. 30, issue 2, pp. 59-. Available from: <http://www.jstor.org/stable/2683794?origin=crossref> [cit. 2015-12-16]
- National Population Projection [online]. 2014. Available from: <http://www.ons.gov.uk/ons/interactive/uk-national-population-projections---dvc3/index.html> [Accessed 2015-12-16]
- Nelder, J. 1986. Statistics, Science and Technology. *Journal of the Royal Statistical Society. Series A (General)* [online], vol. 149, issue 2, pp. 109-. Available from: <http://www.jstor.org/stable/10.2307/2981525?origin=crossref> [cit. 2015-12-16]
- Novotný, M. 2014. Nebojte se čísel, nekoušou. *Statistika a my* [online], vol. 2014, no. 10. Available from: <http://www.statistikaamy.cz/2014/12/nebojte-se-cisel-nekousou/> [cit. 2015-12-16]
- Opendata Africa 2013. [online]. Available from: <http://opendataforafrica.org/> [Accessed 2015-12-16]
- Or, O. Data Visualisation and Its Application in Official Statistics. Census and Statistics Department, Hong Kong, China [online], pp. 1-6. Available from: <http://www.statistics.gov.hk/wsc/CPS109-P4-S.pdf> [cit. 2015-12-16]
- Palmer, M. 2006. Data is the New Oil. [Ana.blogs.com/maestros/2006/11/data\\_is\\_the\\_new.html](http://ana.blogs.com/maestros/2006/11/data_is_the_new.html) [online]. Available from: [http://ana.blogs.com/maestros/2006/11/data\\_is\\_the\\_new.html](http://ana.blogs.com/maestros/2006/11/data_is_the_new.html) [Accessed 2015-12-16]
- Parlament ČR. 2015. ZÁKON č . 166/1993 Sb., O NEJVYŠŠÍM KONTROLNÍM ÚŘADU. Praha.
- Pfannkuch, M., & Wild, J. 2000. Statistical Thinking and Statistical Practice: Themes Gleaned from Professional Statisticians. *Statistical Science* vol. 15, no. 2. Institute of Mathematical Statistics: pp. 132–152. Available from: <http://www.jstor.org/stable/2676728> [cit. 2015-12-16]
- Prodromou, A.. 2013. Data Visualisation and Statistics from the Future. IN: 59th International Statistical Institute (ISI) World Statistics Congress, Hong Kong, China,

25th - 30th August, 2013 [online] pp. 1-3. Available from: <http://e-publications.une.edu.au/1959.11/13463> [cit. 2015-12-16]

Rafalowska, H. 2006. Building the reputation of a statistical office through effective communication. *Statistical Journal of the United Nations Economic Commission for Europe*, [online], vol. 22, no. 2, pp. 147-156. Available from: <http://content.iospress.com/articles/statistical-journal-of-the-united-nations-economic-commission-for-europe/sju00613> [cit. 2015-12-16]

Ritschelová, I. 2014. 20 let Základních principů oficiální statistiky v OSN. *Statistika a my* [online]. Available from: <http://www.statistikaamy.cz/2014/03/20-let-zakladnich-principu-oficialni-statistiky-v-osn/> [Accessed 2015-12-16]

Ritschelová, I. 2014. Návrh služebního zákona ohrožuje odbornou úroveň statistiky. [online], vol. 2014, no. 2, pp. 15-18. Available from: <http://www.statistikaamy.cz/2014/02/navrh-sluzebniho-zakona-ohrozuje-odbornou-uroven-statistiky/> [cit. 2015-12-16]

Ritschelová, I. 2014. Prioritní úkoly ČSÚ na rok 2014. *Statistika a my* [online], vol. 2014, issue 11-12, Available from: <http://www.statistikaamy.cz/2014/01/prioritni-ukoly-csu-na-rok-2014/> [cit. 2015-12-16]

Roszka, W. 2015. Some Practical Issues Related to the Integration of Data from Sample Surveys. *Statistika*, vol. 49, no. 1, pp. 60-76

#### Seznam citací

Shedroff, N. c1999. *Information Interaction Design: A Unified Field Theory of Design*. In: *Information design*. 1st MIT Press pbk. ed. Cambridge, Mass.: MIT Press, pp. 267-293.

Shneiderman, B. 1996. The eyes have it: a task by data type taxonomy for information visualizations. IN: *Proceedings 1996 IEEE Symposium on Visual Languages* [online]. IEEE Comput. Soc. Press, pp. 336-343. Available from: <http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=545307> [Accessed 2015-12-16]



Shneiderman, B. 1996. The eyes have it: a task by data type taxonomy for information visualizations. IN: Proceedings 1996 IEEE Symposium on Visual Languages [online]. IEEE Comput. Soc. Press, pp. 336-343. Available from: <http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=545307> [Accessed 2015-12-16]

Singapore Department of Statistics. Singapore Department of Statistics. [online]. Available from: <http://www.singstat.gov.sg/about-us/history> [cit. 2015-12-16]

Smadar G. 2011. Responsibility Toward the Self and Responsibility Toward the Other and the Gap Between. *Daat: A Journal of Jewish Philosophy & Kabbalah* no. 70. Bar Ilan University Press: 97–131. Available from: <http://www.jstor.org/stable/24232411>[cit. 2015-12-16]

Smith, A. 2008. Web 2.0 and official statistics: The case for a multi-disciplinary approach. *Statistical Journal of the IAOS* [online], vol. 25, no. 1, pp. 117-124. Available from: [https://www.virtualstatisticalsystem.org/vss\\_uploads/Web\\_2.0\\_and\\_official\\_statistics\\_Smith\\_and\\_Rogers.pdf](https://www.virtualstatisticalsystem.org/vss_uploads/Web_2.0_and_official_statistics_Smith_and_Rogers.pdf) [cit. 2015-12-16]

Smith, A. 2013. Data Visualisation and Beyond: A multi-disciplinary approach to promote user engagement with Official Statistics. *Statistical Journal of the IAOS* [online], vol. 29, no. 3, pp. 173-185. Available from: <http://content.iospress.com/articles/statistical-journal-of-the-iaos/sji00783> [cit. 2015-12-16]

Spence, R. 2014. *Information Visualization: An Introduction* [online]. Available from: <http://www.springer.com/br/book/9783319073408> [Accessed 2015-12-16]

Statistics Explained [online]. Available from: [http://ec.europa.eu/eurostat/statistics-explained/index.php/Main\\_Page](http://ec.europa.eu/eurostat/statistics-explained/index.php/Main_Page) [Accessed 2015-12-16]

Statistics Explorer [online]. Available from: [http://www.scb.se/Kartor/Statistikatlas\\_42\\_KN\\_201205\\_English/index.html](http://www.scb.se/Kartor/Statistikatlas_42_KN_201205_English/index.html) [Accessed 2015-12-16]

Statistics Netherlands. 2014. What is the value of official statistics and how do we communicate that value?. IN: *Conference of European Statistics*, pp. 2-8.

Stephen Few. 2012. Show me the numbers: designing tables and graphs to enlighten. 2nd ed. Burlingame, Calif: Analytics Press.

Strategy of Statistics Estonia 2010-2015 [online]. Estonsko, 2015 [cit. 2015-12-16]. Dostupné z: <http://www.stat.ee/strategy> 2008

Trendová, P. 2011. Státní statistická služba mezi mlýnskými kameny. [online], vol. 2011, 11-12, pp. 25-27. Available from: [https://www.czso.cz/documents/10180/20541267/180411111224\\_25.pdf/40cf8163-431c-4dee-9258-6e4003286d8b?version=1.0](https://www.czso.cz/documents/10180/20541267/180411111224_25.pdf/40cf8163-431c-4dee-9258-6e4003286d8b?version=1.0) [cit. 2015-12-16]

Tufte, E. c2001. The visual display of quantitative information [online]. 2nd ed. Cheshire, Conn.: Graphics Press. [Accessed 2015-12-16]

Tukey, J. c1977. Exploratory data analysis. Reading, Mass.: Addison-Wesley Pub. Co.

Tukey, J. c1977. Exploratory data analysis. Reading, Mass.: Addison-Wesley Pub. Co.

Tukey, J. W. 1972. Some Graphic and Semigraphic Displays. Statistical Papers in Honor of George W. Snedecor [online], pp. 293-316. Available from: [http://www.edwardtufte.com/bboard/q-and-a-fetch-msg?msg\\_id=0003v7](http://www.edwardtufte.com/bboard/q-and-a-fetch-msg?msg_id=0003v7) [cit. 2015-12-16]

UK Statistics Authority. 2009 Code of Practice for Official Statistics., pp. 1-19.

United Nations. 2014. Fundamental principles of statistics. [online], pp. 2. Available from: <http://unstats.un.org/unsd/dnss/gp/FP-New-E.pdf> [cit. 2015-12-16]

Van den Elzen, S. & Van Wijk, J.. 2013. Small Multiples, Large Singles. Computer Graphics Forum [online], vol. 32, 3pt2, pp. 191-200. Available from: <http://doi.wiley.com/10.1111/cgf.12106> [cit. 2015-12-16]

Veřejná databáze [online]. Available from: <http://vdb.czso.cz/> [Accessed 2015-12-16]

Veselá, M. 2014. Tvorba statistik přísluší statistikům. Statistika a my [online]. Available from: <http://www.statistikaamy.cz/2014/02/tvorba-statistik-prislusi-statistikum/> [Accessed 2015-12-16]

Vickers, P., Faith, J. & Rossiter, N.. 2013. Understanding Visualization. IEEE Transactions on Visualization and Computer Graphics [online], vol. 19, issue 6, pp.

1048-1061. Available from:

<http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=6311404> [cit. 2015-12-16]

Víchová, J. 2011. Ještě se nenarodil popularizátor statistiky,. Statistika a my [online], vol. 2011, no. 2, pp. 24-28. Available from:

[https://www.czso.cz/documents/10180/20541259/1804110224\\_27.pdf/8acaf64a-73e5-4acf-a6dc-d4d2f27cbd2e?version=1.0](https://www.czso.cz/documents/10180/20541259/1804110224_27.pdf/8acaf64a-73e5-4acf-a6dc-d4d2f27cbd2e?version=1.0) [cit. 2015-12-16]

Wainer, H. 1984. How to Display Data Badly. The American Statistician [online], vol. 38, issue 2, pp. 137-147. Available from:

<http://www.jstor.org/stable/2683253?origin=crossref> [cit. 2015-12-16]

Ware, C. 2013. Information visualization: perception for design. Third edition. Waltham, MA: Morgan Kaufmann.

Watanabe, T. 2010. Let's Practice What We Preach. Teaching Children Mathematics [online] vol. 16, no. 7. National Council of Teachers of Mathematics: 380–81.

<http://www.jstor.org/stable/41199497>.

Woodward, J. 1989. Data and phenomena. Synthese [online], vol. 79, issue 3, pp. 393-472. Available from: <http://link.springer.com/10.1007/BF00869282> [cit. 2015-12-16]

Work Session on the Communication of Statistics. 2014. Communicating the NSO value proposition – ensuring that official statistics is a well recognized brand. IN: UNITED NATIONS ECONOMIC COMMISSION FOR EUROPE (UNECE) CONFERENCE OF EUROPEAN STATISTICIANS, pp. 1-9.

Wurman, R. 1989. Information anxiety. 1st ed. New York: Doubleday

x State & society - Population projection 2015. [online]. Available from:

<https://www.destatis.de/EN/FactsFigures/SocietyState/Population/PopulationProjection/PopulationProjection.html> [Accessed 2015-12-16]

Zhu, B. & Chen, H.. 2005. Information visualization. Annual Review of Information Science and Technology [online], vol. 39, issue 1, pp. 139-177. Available from:

<http://doi.wiley.com/10.1002/aris.1440390111> [cit. 2015-12-16]

Ziemkiewicz, C. & Kosara, R.. 2007. Understanding Information Visualization in the Context of Visual Communication. [online], pp. 1-8. Available from: <https://www.purdue.edu/discoverypark/vaccine/assets/pdfs/publications/pdf/Understanding%20Information.pdf> [cit. 2015-12-16]

Zins, C. 2007. Conceptual approaches for defining data, information, and knowledge. *Journal of the American Society for Information Science and Technology* [online], vol. 58, issue 4, pp. 479-493. Available from: <http://doi.wiley.com/10.1002/asi.20508> [cit. 2015-12-16]

## 8. Seznam obrázků:

Tabulka 1 .....	80
Tabulka 2 .....	80
Tabulka 3 .....	80
Graf 1 .....	62
Graf 2 .....	68
Graf 3 .....	68
Graf 4 .....	70
Graf 5 .....	71
Graf 6 .....	72
Graf 7 .....	73
Graf 8 .....	74
Graf 9 .....	75
Graf 10 .....	76
Graf 11 .....	78
Graf 12 .....	79

