

**EKONOMICKÁ UNIVERZITA V BRATISLAVE**  
**FAKULTA HOSPODÁRSKEJ INFORMATIKY**

Evidenčné číslo: 103004/B/2021/36122163742067460

**VIZUALIZÁCIA DÁT A JEJ NÁSTROJE**

**Bakalárska práca**

**2021**

**Katarína Cagáňová**

**EKONOMICKÁ UNIVERZITA V BRATISLAVE**  
**FAKULTA HOSPODÁRSKEJ INFORMATIKY**

**VIZUALIZÁCIA DÁT A JEJ NÁSTROJE**

**Bakalárska práca**

**Študijný program:** hospodárska informatika

**Študijný odbor:** bHI\_15

**Školiace pracovisko:** Katedra aplikovanej informatiky

**Vedúci záverečnej práce:** RNDr. Eva Rakovská, PhD.

**Bratislava 2021**

**Katarína Cagáňová**

## Čestné vyhlásenie

Čestne vyhlasujem, že bakalársku prácu som vypracovala samostatne a že som uviedla všetku použitú literatúru v Zozname použitej literatúry.

**Dátum:**

.....

(Katarína Cagánová)

## **POĎAKOVANIE**

Chcela by som sa poďakovať svojej vedúcej práce, RNDr. Eve Rakovskej, PhD., za vytvorenie priestoru na spracovanie tejto témy. Zároveň ďakujem za cenné rady a usmerňovanie pri písaní mojej bakalárskej práce.

## ABSTRAKT

CAGÁŇOVÁ, Katarína: *Vizualizácia dát a jej nástroje*. – Ekonomická univerzita v Bratislave. Fakulta hospodárskej informatiky. Katedra aplikovanej informatiky. – RNDr. Eva Rakovská, PhD. – Bratislava: FHI, 2021

Cieľom záverečnej práce bolo analyzovať ako efektívne vizualizovať dáta podľa charakteristík a porovnať IT nástroje na vizualizáciu dát.

V súčasnosti dáta predstavujú základ pre rozhodovanie v praxi. Ich vizualizácia značne uľahčuje rozhodovanie. Práca je zameraná na objasnenie charakteristík dát, podľa ktorých je vhodné hľadať vhodný typ vizualizácie a následne porovnáva vybrané IT nástroje na vizualizáciu dát.

Bakalárska práca sa skladá z teoretickej a praktickej časti. V teoretickej časti sa nachádzajú základné pojmy – čo je to vizualizácia, dáta a informácie. Ďalej je popísaný proces tvorby vizualizácie, aké typy grafov a dát poznáme a popísané najčastejšie používané nástroje na vizualizáciu dát. V praktickej časti boli porovnané dva vybrané nástroje. Vybrané nástroje na vizualizáciu dát boli Tableau [6] a D3.js [9].

Zistili sme, že Tableau a D3.js sú dva podobné nástroje. Tableau sa ľahko používa, je to nástroj, ktorý nám vie poslúžiť na vytváranie výkonných, pútavých vizualizácií. Tableau je veľmi dobrý na prieskum dát a nevyžaduje žiadny zásah do databázy. D3.js vyžaduje podstatne väčšie investície do času a vývojových hodín ako v prípade Tableau. D3.js prináša výhody ako je napríklad bezchybná integrácia do webových stránok. D3.js má väčšiu knižnicu vizualizácií a je flexibilnejší. Vyžaduje veľmi časovo náročné programovanie (JavaScript) a údržbu. Na prieskum informácií údajov nie je veľmi dobrý, musíme už mať na pamäti plán a návrh.

## KLÚČOVÉ SLOVÁ

vizualizácia dát, nástroje na vizualizáciu, Tableau, D3.js

## **ABSTRACT**

CAGÁŇOVÁ, Katarína: *Data visualization and its tools*. – University of Economics in Bratislava. Faculty of Economic Informatics. Department of Applied Informatics. – RNDr. Eva Rakovská, PhD. – Bratislava: FHI, 2021

The aim of the work was to analyse how to effectively visualize data according to characteristics and compare IT tools for data visualization.

At present, data are the basis for decision-making in practice. Their visualization greatly simplifies decision-making. The work is focused on clarifying the characteristics of data, according to which it is appropriate to search for a suitable type of visualization and then compares selected IT tools for data visualization.

The bachelor thesis consists of a theoretical and practical part. The theoretical part contains the basic concepts - what is visualization, data and information. Next, the process of creating a visualization is described, what types of graphs and data we know, and the most used tools for data visualization are described. In the practical part we compared two selected tools. Selected tools for data visualization were Tableau [6] and D3.js [9].

We found that Tableau and D3.js are two similar tools. Tableau is easy to use, it is a tool that can be used to create powerful, engaging visualizations. Tableau is very good at exploring data and does not require any intervention in the database. D3.js requires significantly more investment in time and development hours than in the case of Tableau. D3.js brings benefits such as flawless integration into websites. D3.js has a larger library of visualizations and is more flexible. It requires very time-consuming programming (JavaScript) and maintenance. It is not very good to research information data, we must already keep in mind the plan and proposal.

## **KEYWORDS**

data visualization, visualization tools, Tableau, D3.js

# Obsah

Úvod .....	8
1. Súčasný stav riešenej problematiky .....	9
1.1. Vizualizácia dát .....	9
1.2. Proces vytvárania vizualizácie .....	11
1.2.1. Získavanie dát .....	11
1.2.2. Triedenie dát .....	11
1.2.3. Filtrovanie dát .....	11
1.2.4. Dolovanie dát .....	11
1.2.5. Reprezentácia dát .....	11
1.2.6. Zjemňovanie dát .....	12
1.2.7. Interakcia dát .....	12
1.3. Typy dát .....	12
1.3.1. Štruktúrované a neštruktúrované dáta .....	12
1.3.2. Kvantitatívne a kvalitatívne dáta .....	13
1.4. Vstupné dátové formáty .....	15
1.5. Typy grafov .....	16
1.5.1. Stĺpcový graf .....	18
1.5.2. Čiarový graf .....	18
1.5.3. Koláčový graf .....	18
1.5.4. Prstencový graf .....	19
1.5.5. Plošný graf .....	19
1.5.6. Graf XY (závislosť) .....	20
1.5.7. Bublinový graf .....	20
1.5.8. Radarový graf .....	20
1.6. Nástroje na vizualizáciu dát .....	20
1.6.1. Tableau Public .....	21
1.6.2. Google Charts API .....	23
1.6.3. D3.js .....	24
1.6.4. RAWGraphs .....	25
1.6.5. Charts.js .....	26
1.6.6. Infogram .....	26
1.6.7. Leaflet .....	26

1.6.8. Porovnanie nástrojov na vizualizáciu dát.....	27
2. Cieľ práce.....	28
3. Metodika práce a metódy skúmania .....	30
3.1. Vizualizácia dát v Tableau .....	30
3.2. Vizualizácia dát v D3.js .....	35
4. Výsledky a diskusia .....	44
4.1. Výsledky .....	44
4.2. Diskusia.....	46
Záver .....	50
Zoznam použitej literatúry .....	52



# Úvod

Vďaka technologickému pokroku, informácie vieme získavať v akejkoľvek oblasti záujmu. V súčasnosti už neriešime problém, ako získavať dáta. V bakalárskej práci sme sa rozhodli venovať sa téme vizualizácia dát a jej nástroje. Táto práca má za úlohu rozšíriť obzory o jednotlivých dátových typoch, potrebných pri vizualizácii, typoch grafov a vizualizačných metódach. Z vizualizačných nástrojov sme použili nástroje Tableau [6] a D3.js [9].

V súčasnosti sa údaje generujú napríklad na YouTube, Tumblr, Reddit, Facebook, WhatsApp, Twitter, Instagram Gmail a LinkedIn. Pochopenie týchto údajov je veľmi dôležité, pretože je to zásadný a veľmi dôležitý subjekt organizácie, národov a inštitúcií. Veľké dáta sú zbierkou veľkých a komplexných údajov, s ktorými sa ťažko zaobchádza pomocou tradičného aplikačného softvéru na spracovanie údajov. Analýza a vizualizácia súborov dát môžu poskytnúť nové obchodné trendy, predchádzať chorobám a modelovať prognózy budúcich paradigiem a bojovať proti kriminalite a podobne. V súčasnosti najpoužívanejší nástroj na analýzu a vizualizáciu údajov, zisťovanie údajov je Tableau. Tableau je jedným z najrýchlejšie pripravovaných nástrojov business intelligence (BI). Ľahko sa učí a je pre používateľa veľmi užitočný. Tableau je softvér, ktorý pomáha používateľom skúmať a porozumieť ich údajom vytváraním interaktívnych vizualizácií. Tento softvér má také výhody, že ho možno použiť v spojení s ľubovoľnou databázou, a jeho použitie je ľahké pomocou interaktívnej vizualizácie, ktorá vyžaduje požadovaný formát. COVID-19 (koronavírus) sa ukázal ako skutočne globálna pandémia, ktorá ovplyvňuje ľudí takmer v každom kúte sveta. Užitočnosť Tableau v analytike dát COVID-19 (koronavírus) možno merať podľa jeho výkonu a rýchlosti.

# 1. Súčasný stav riešenej problematiky

## 1.1. Vizualizácia dát

Je dôležité vymedzenie základných pojmov ako sú dáta a informácie. Mládková [12] ich definovala takto:

- Dáta – „dáta sú všetko, čo môžeme monitorovať našimi zmyslami: to je cítiť, chutnať, vidieť a počuť.“
- Informácie – „informácie sú dáta, ktorým používateľ pri interpretácii priraduje dôležitosť a význam.“

Vizualizáciu dát môžeme chápať ako proces alebo techniku. Napríklad Langekamp [13] uvádza, že vizualizácia dát je technika využívajúca schopnosti ľudského mozgu vnímať dáta a rozoznávať v nich informácie vo forme vzorov a trendov. Vizualizácia dát je grafickou interpretáciou dát v podobe, ktorá uľahčuje vnímanie, získavanie relatívnych informácií a pochopenie komplexných vzťahov v dátach.

Vizualizácia sa zaoberá prezentovaním a spracovaním dát. Pre vytvorenie vizualizácie je zdrojom množstvo štatistických dát, meraných kvantitatívnych dát, alebo iných faktov o svete. Cieľom je prezentovať záznamy vo vizuálnej podobe tak, aby bolo možné získať z dát relevantné informácie, odhaľovať vzory, trendy a ďalšie správy, ktoré sa v dátach ukrývajú [1]. Vizualizácia údajov sa snaží podať informácie v stručnej, ale zrozumiteľnej podobe. Vizualizácie môžu zdôrazniť dôležité body a poskytnúť používateľovi užitočné veci, napríklad súhrny údajov [5]. Informačnú vizualizáciu je možné kategorizovať podľa dátových typov. V taxonómii od Shneidermana [3] sú informačné vizualizácie rozdelené do siedmich rôznych typov údajov: 1D, 2D, 3D, časové, stromové, sieťové a viacrozmerné údaje. Pre každý dátový typ existuje niekoľko druhov vhodných vizualizácií.

Vizuálne znázornenia poskytujú dáta v kontextoch, ktoré stimulujú mozog diváka, a odhaľujú informácie, ktoré zvyčajne nie sú zrejmé z tabuľkových dát. Spravidla je oveľa prirodzenejšie zachytiť podstatný obsah údajov pomocou vizuálnych artefaktov. Mapy sú výkonným, jasným a efektívnym prostriedkom na rozprávanie údajov prostredníctvom dát. Môžu zabaliť obrovské množstvo informácií do malých priestorov a uľahčiť tak porovnávanie údajov, poskytovanie poznatkov a odhalenie trendov, vzťahov, príčinných súvislostí a iných vzorov skrytých v číslach. Vizualizácia údajov je umením aj vedou. Graf nemusí vysvetľovať

všetko. Nie vždy to musí byť presné. Môže byť zameraný na konkrétne publikum, ktoré by malo poskytnúť kontext potrebný na jeho pochopenie a dekodovanie [16].

Vizualizácia je použitie počítačom podporovanej vizuálnej reprezentácie údajov. Na rozdiel od vizualizácie statických údajov umožňuje interaktívna vizualizácia údajov používateľom určiť formát používaný pri zobrazovaní údajov [19].

Bežné vizualizačné techniky zahŕňajú [19]:

- Čiarový graf: Ukazuje vzťah medzi položkami. Môže sa použiť na porovnanie zmien za určité časové obdobie.
- Stĺpcový graf: Používa sa na porovnanie množstiev rôznych kategórií.
- Bodový graf: Toto je dvojrozmerný graf zobrazujúci variáciu dvoch položiek.
- Koláčový graf: Používa sa na porovnanie častí celku.

Formát grafov môže mať teda podobu stĺpcového grafu, koláčového grafu, spojnicového grafu atď. Je dôležité pochopiť, ktorý graf je potrebné na vybrané dáta použiť. Vizualizácia údajov využíva počítačovú grafiku na zobrazovanie vzorov, trendov a vzťahov medzi prvkami údajov. Môže generovať koláčové grafy, stĺpcové grafy, bodové grafy a ďalšie typy údajových grafov pomocou jednoduchých rozbaľovacích ponúk a kliknutí myšou. Farby sú starostlivo vybrané pre určité typy vizualizácie. Keď sa na reprezentáciu údajov používa farba, musíme na rozlíšenie medzi dátovými prvkami zvoliť efektívne farby. Pri vizualizácii údajov sa údaje abstrahujú a sumarizujú. Priestorové premenné, ako je poloha, veľkosť a tvar, predstavujú kľúčové prvky údajov. Vizualizačný systém by mal vykonávať redukciu dát, transformáciu a premietanie pôvodného súboru údajov na obrazovku. Mal by vizualizovať výsledky vo forme tabuliek a grafov a prezentovať výsledky užívateľsky prívetivým spôsobom [19].

## **1.2. Proces vytvárania vizualizácie**

Proces vytvárania vizualizácie je tvorený z týchto fáz – získavanie, spracovanie dát a tvorby. Postup rozdelil Fry [1] do týchto krokov (obrázok 1):

### *1.2.1. Získavanie dát*

Znamená aj preberanie dát zo zdroja. Tento krok zahŕňa získavanie údajov, či už zo súboru na disku alebo zo zdroja na sieti. Rovnako ako mnoho ďalších krokov, aj tento môže byť buď mimoriadne komplikovaný (to znamená, pokúšať sa zhromaždiť užitočné údaje z veľkého systému), alebo veľmi jednoduchý (čítanie ľahko dostupného textového súboru).

### *1.2.2. Triedenie dát*

Prispôsobenie štruktúry dát. Po získaní údajov je potrebné ich roztriediť.

### *1.2.3. Filtrovanie dát*

Ďalším krokom je filtrovanie údajov a odstránenie častí, ktoré nie sú relevantné pre naše použitie. Ide o odstraňovanie všetkých údajov, ktoré sú nadbytočné a normalizácia dát.

### *1.2.4. Dolovanie dát*

Tento krok zahŕňa matematiku, štatistiku a dolovanie dát.

### *1.2.5. Reprezentácia dát*

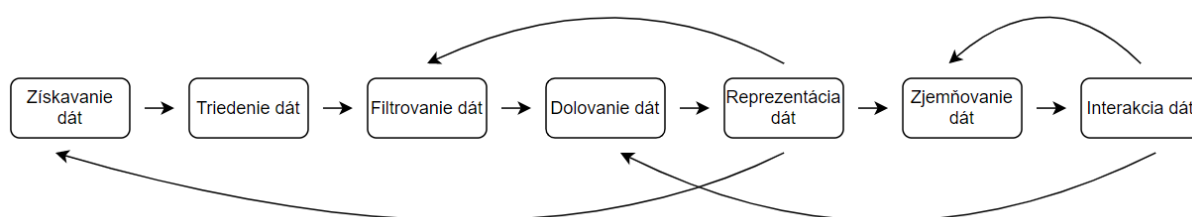
Určenie jednoduchého znázornenia, či údaje majú jeden z mnohých tvarov, napríklad či majú stĺpcový graf, zoznam alebo strom. V tomto kroku sa vyberá vhodná vizuálna reprezentácia dát.

### 1.2.6. Zjemňovanie dát

Využitie poznatkov vizuálnej komunikácie pri úpravách grafického výstupu, pridávanie detailov.

### 1.2.7. Interakcia dát

Doplnenie metód pre manipuláciu s dátami pri vizualizácii. [1]



Obrázok 1 Interakcia medzi siedmimi krokmi [1].

## 1.3. Typy dát

### 1.3.1. Štruktúrované a neštruktúrované dáta

Dáta môžeme deliť na štruktúrované a neštruktúrované dáta. Niektoré sú teda štruktúrované (napríklad v relačných databázach), ale väčšina dát je neštruktúrovaná.

Štruktúrované dáta sú organizované dáta, ktoré možno považovať za pozorovania a charakteristiky. Spravidla sa organizuje tabuľkovou metódou (riadky a stĺpce). O štruktúrovaných dátach sa všeobecne predpokladá, že je s nimi oveľa jednoduchšie pracovať a analyzovať ich. Väčšina štatistických modelov a modelov strojového učenia bola zostavená s prihliadnutím na štruktúrované dáta a nemôže pracovať na voľnej interpretácii neštruktúrovaných dátach. Prirodzená štruktúra riadkov a stĺpcov je ľahko stráviteľná pre ľudské i strojové oči [18]. Nasledujúca tabuľka 1 znázorňuje príklad na štruktúrované dáta.

ID	Meno	Vek	Rok_narodenia	Výška	Oddelenie_id
1	Alexandra	40	1981	160	6
2	Marek	33	1988	179	3
3	Erika	21	2000	171	2
4	Alena	19	2002	168	2
5	Ľudovít	35	1986	185	1
6	Marcel	26	1995	190	5
7	Peter	22	1999	176	4

Tabuľka 1 – Príklad na štruktúrované dáta (fiktívny príklad vytvorený autorkou práce).  
Oddelenie\_id = Identifikačné číslo pracovného oddelenia.

Neštruktúrované dáta sú neorganizované dáta, existujú ako bezplatná entita a neriadia sa nijakou štandardnou organizačnou hierarchiou. Neštruktúrované dáta sú napríklad: aktivita na sociálnych sieťach, snímky zo satelitov, textové súbory, video, email a podobne. Tieto dáta nevieme umiestniť do tabuľky, preto sa nedajú vizualizovať – nemáme vyjadrené závislosti, ktoré vizualizujeme [18].

### 1.3.2. Kvantitatívne a kvalitatívne dáta

Kvantitatívne dáta – je možné opísať pomocou čísel a na množine sú možné základné matematické postupy vrátane sčítania. Predstavujú sumy, počty a výsledok merania. Kvantitatívne dáta vieme zoradiť a vieme s nimi vykonávať matematické operácie.

Kvantitatívne dáta je možné rozdeliť na diskkrétne a kontinuálne (spojité) dáta.

Diskkrétne dáta sú dáta, ktoré sa počítajú. Môže nadobúdať iba určité hodnoty. Medzi príklady diskrétnych kvantitatívnych dát patrí napríklad hod kockou, pretože môže mať iba šesť hodnôt.

Kontinuálne (spojité) dáta opisujú namerané údaje. Existuje v nekonečnom rozsahu hodnôt. Dobrým príkladom nepretržitých údajov by bola váha človeka. Výška osoby alebo budovy je spojité číslo, pretože je možná nekonečná škála desatinných miest. Ďalšími príkladmi nepretržitých údajov by mohli byť čas a teplota.

Kvalitatívne dáta – nemožno opísať pomocou čísel a základnej matematiky. Tieto dáta sa všeobecne považujú za opísané pomocou prirodzených kategórií a jazyka [18].

V roku 1946 uverejnil americký psychológ Stanley Smith Stevens v časopise Science slávny príspevok s názvom „O teórii váh a merania“. V tomto príspevku tvrdil, že všetko vedecké meranie sa uskutočňovalo pomocou 4 rôznych typov váh, ktoré nazval:

- Nominálne dáta
- Ordinálne dáta
- Intervalové dáta
- Pomerové dáta

Tento dokument v podstate zjednotil kvalitatívne dáta (nominálne dáta a ordinálne dáta) a kvantitatívne dáta (dáta o intervaloch a dáta o pomere) [18].

### *Nominálna úroveň*

Nominálna úroveň pozostáva z dát, ktoré sú opísané čisto názvom alebo kategóriou. Medzi základné príklady patrí pohlavie, národnosť alebo druh. Nie sú opísané číslami, a preto sú kvalitatívne [18].

Ovocie možno rozdeliť na jablká, pomaranče, banány atď. Nemá zmysel, aby bolo možné ovocie umiestniť v usporiadanom poradí. niekedy sa čísla používajú týmto spôsobom; napríklad číslo na prednej časti autobusu má všeobecne čisto nominálnu hodnotu. identifikuje trasu, po ktorej autobus ide [18].

### *Ordinálna úroveň*

Dáta na ordinálnej úrovni nám poskytujú poradie alebo prostriedky na umiestnenie jedného pozorovania pred druhé; neposkytuje nám však relatívne rozdiely medzi pozorovaniami, čo znamená, že aj keď si môžeme zoradiť pozorovania od prvého po posledné, nemôžeme ich pridať ani odčítať, aby sme získali skutočný význam. Radové stupnice sa často používajú na meranie spokojnosti, šťastia a podobne [18].

## *Intervalová úroveň*

Pri intervalovej úrovni merania, je možné odvodiť medzeru medzi hodnotami údajov. Čas odletu a čas priletu lietadla sú definované na intervalovej úrovni. Intervalové dáta sú zábavné (a užitočné), pretože sa týkajú poradia aj rozdielov medzi premennými. Takto môžete merať štandardnú odchýlku a centrálnu tendenciu.

Príkladom intervalových dát sú teploty v stupňoch Celzia. 20 stupňov Celzia je vyššia teplota ako 10 stupňov Celzia a rozdiel medzi 20 stupňami a 10 stupňami je 10 stupňov. Rozdiel medzi 10 a 0 je tiež 10 stupňov [18].

## *Pomerová (ratio) úroveň*

S pomerovou úrovňou máme plnú výpovednú silu reálneho čísla. Môžeme vytvoriť výroky ako „Objekt A je dvakrát väčší ako objekt B.“ Hmotnosť objektu je definovaná na pomerovej úrovni. Peniaze sú definované na pomerovej škále. Použitie pomerovej úrovni znamená nulovú hodnotu použitú ako referenciu. Pomerové údaje nám hovoria o poradí premenných, rozdieloch medzi nimi a tie majú túto absolútnu nulu. Čo umožňuje vykonávať a vykresľovať najrôznejšie výpočty a závery. Pomerové údaje sú veľmi podobné intervalovým údajom, okrem toho, že nula znamená „žiadne“. Pre pomerové údaje nie je možné mať záporné hodnoty.

Napríklad výška sú pomerové dáta. Nie je možné mať negatívnu výšku. Ak je výška objektu nula, potom neexistuje žiadny objekt [18].

## **1.4. Vstupné dátové formáty**

Pre lepšiu logistiku dát je potrebné transformovať dáta do určitej štruktúry. Ak sú dáta prístupné digitálne, je možné, že bude potrebné ich transformovať, teda ich súčasnú štruktúru na štruktúru kompatibilnú s aplikáciou na zobrazenie alebo vizualizáciu [14].

Medzi štandardné dátové formáty patria [14]:

- JSON – (Java Script Object Notation)
- CSV – (Comma-Separated values)



- XML – (eXtensible Markup Language)

CSV je veľmi jednoduchý formát obyčajného textu, to znamená, že súbor je sekvencia znakov, ktorý je oddelený špeciálnymi znakmi. JSON je používaný formát na výmenu dát na webe a XML je flexibilný, jednoduchý textový formát odvodený od SGML [14].

## 1.5. Typy grafov

Vizualizácia údajov sa zvyčajne týka vizuálneho zobrazenia kvantitatívnych informácií, ako napríklad v štatistických a informačných mapách, grafoch, údajových mapách atď. ale môže obsahovať akýkoľvek druh vizuálneho znázornenia údajov, ako napríklad matematická grafika, cestné siete (systémy metra, cesty, schémy elektronických obvodov), oblaky slov, zvukové a zvukové reprezentácie, časové osi, geografické informačné systémy, chemické a atómové diagramy alebo akýkoľvek iný spôsob kódovania údajov pomocou vizuálnych artefaktov [16].

Výber grafu vyžaduje znalosť vašich údajov. Grafy sú komunikačným prostriedkom zameraným na odhaľovanie informácií, takže hlavná otázka znie [16]:

- Čo chcete ukázať? Po zodpovedaní tejto otázky by ste mali analyzovať svoje údaje a zistiť, aký druh údajov máte. Dátové hodnoty použité vo vizualizáciách možno zvyčajne klasifikovať ako jeden z týchto troch typov:**

- Kvantitatívna: Hodnota, ktorá sa dá zmerať alebo spočítať (číslo, dĺžka, plocha, uhol)

- Poradová: Hodnotu je možné zoradiť alebo porovnať (sýtosť farieb, plocha, uhol, dĺžka, slová)

- Nominálna: kategória (meno)

- Aký je účel vašej tabuľky? Prajete si odhaliť vzťahy, trendy alebo kauzalitu? Aký vzťah chcete zdôrazniť? Súvisia vaše premenné s časom alebo priestorom?**

Vizualizácie je možné usporiadať do kategórií, ktoré uľahčujú výber typu grafu, ktorý potrebujete. Väčšinu máp je možné umiestniť do jednej z týchto kategórií:

- Časové rady (vykreslenie jednej premennej za určité časové obdobie). Napríklad čiarový graf, ktorý demonštruje trend.

- Časové / lineárne (kategórie umiestnené v časovej línii). Napríklad séria udalostí.

- Priestorové / rovinné / objemové (kategórie distribuované v priestorovej mape). Napríklad kartogram alebo choropleth s dátami distribuovanými na geografickej mape.

- Porovnanie (kategórie spojené s veličinami sa porovnávajú a zoradia počas jedného obdobia.) Napríklad stĺpcový graf, ktorý porovnáva hodnoty.

- Časť k celku (kategorické členenie ako pomer k celku.) Napríklad koláčový graf s plátkami ako percentami.

- Korelácia (porovnanie dvoch alebo viacerých premenných). Napríklad scatterplot porovnávajúci dve premenné alebo bublinový graf porovnávajúci tri.

Aspekty, ktoré možno použiť na meranie kvality a integrity vizualizácií. Sú to tieto:

- Pomer dát a atramentu - Množstvo atramentu (alebo pixelov) vyhradeného na zobrazené údaje

- Chartjunk - Vizualne odpadky, ktoré nie sú relevantné pre zobrazené údaje (a ktoré často bránia).

- Lie Factor - číslo, ktoré meria integritu vizualizácie; napríklad grafy, ktoré ležia tak, že dostatočne presne neprezentujú proporcie a dĺžky.

Pomer dátového atramentu je možné zlepšiť odstránením nepotrebných čiar a štítkov z grafov. Niekedy sú riadky dôležité pre kontext, ale v interaktívnych webových vizualizáciách môžete byť veľmi minimalistickí. Podrobnosti o dopyte môžete kedykoľvek poskytnúť pomocou tipov alebo iných interaktívnych zdrojov. Komunikácia je hlboko ovplyvnená tým, ako ľudia vnímajú grafiku, a môže byť vylepšená alebo skreslená optickými ilúziami. Neexistujú grafy s klamným faktorom, ktoré by mohli divákov priviesť k nesprávnemu vnímaniu. Pozície a dĺžka sú najlepšie na vyjadrenie kvantitatívnych informácií. Ďalej nasledujú smer a uhly, potom plocha, objem, zakrivenie a nakoniec tieň, sýtosť a farby [16].

Obrazovou prezentáciou informácií je teda graf. Na zobrazenie grafov sa používajú symboly, ako sú kruh, štvorec, elipsa, obdĺžnik a iné. V podkapitolách je popis najčastejšie používaných typov grafov [4].

### *1.5.1. Stĺpcový graf*

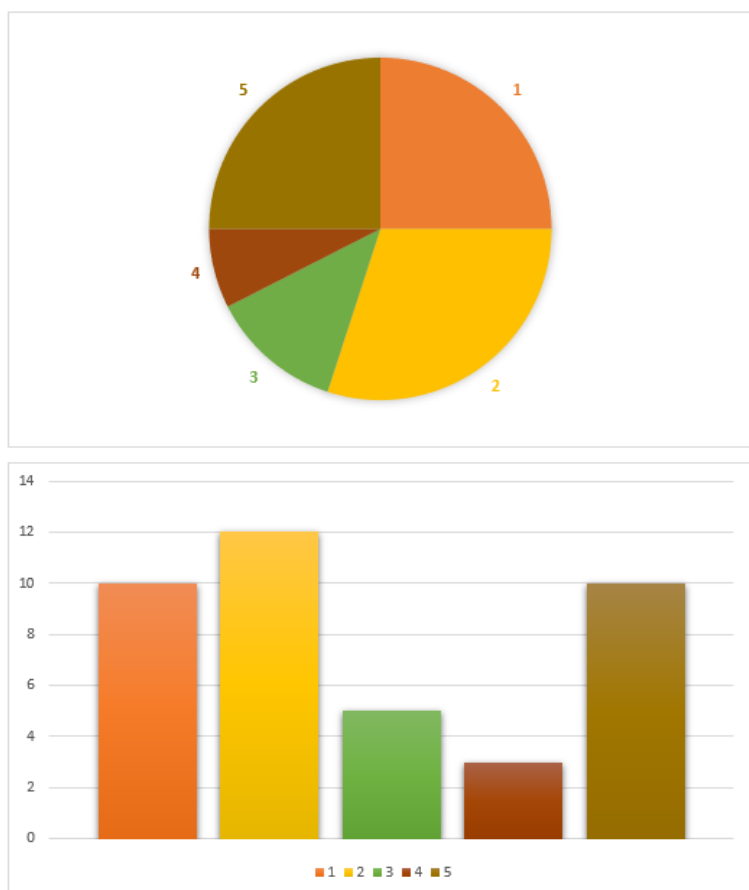
V stĺpcovom grafe možno zobraziť dáta, ktoré sú usporiadané do stĺpcov alebo riadkov v hárku. Stĺpcový graf zobrazuje kategórie pozdĺž vodorovnej osi (kategória) a hodnoty pozdĺž zvislej osi (hodnota). Skupinový stĺpcový graf zobrazuje hodnoty v dvojrozmerných stĺpcoch [4].

### *1.5.2. Čiarový graf*

V čiarovom grafe možno zobraziť dáta, ktoré sú v hárku usporiadané do stĺpcov alebo riadkov. V čiarovom grafe sú údaje kategórií rovnomerne distribuované pozdĺž vodorovnej osi a všetky údaje hodnôt sú rovnomerne distribuované pozdĺž zvislej osi. Čiarové grafy môžu zobrazovať kontinuálne údaje v čase na osi s rovnomernou mierkou a sú preto ideálne na zobrazenie trendov vývoja údajov v rovnakých intervaloch. Napríklad v intervaloch ako sú mesiac, štvrťrok alebo rok. Čiarové grafy sú jedny z najužitočnejších v prípade viacerých radov údajov na grafe. Ak máme len jeden rad údajov, odporúča sa namiesto toho použiť XY (závislosť) graf [4].

### *1.5.3. Koláčový graf*

Koláčové grafy zobrazujú veľkosť položiek v jednom rade dát proporcionálne k súčtu položiek. Koláčové grafy majú tvar kruhu alebo koláča. Kruh je rozrezaný na kúsky rôznych veľkostí. Každý kus predstavuje percento z celku. Po spočítaní sa celková suma rovná 100 percentám [22]. Koláčový graf sa odporúča použiť najmä ak máme len jeden rad údajov, žiadna z hodnôt v údajoch nie je záporná, takmer žiadna z hodnôt v údajoch nie je nulová a ak nemáme viac než sedem kategórií, pričom všetky z nich predstavujú časti celého koláča [4].



Obrázok 2 Porovnávanie hodnôt v koláčovom a stĺpcovom grafe (vlastné spracovanie)

#### 1.5.4. Prstencový graf

Prstencový graf je odvodený od koláčového grafu. Prstencové grafy zobrazujú vzťahy častí z celku. Ale môžu obsahovať viac než jeden rad dát. Každý prstenec predstavuje jeden rad údajov. V prstencových grafoch sa ťažšie orientuje [4].

#### 1.5.5. Plošný graf

Plošný graf možno použiť na zobrazenie zmeny za určité obdobie a zdôraznenie celkovej hodnoty v rámci trendu. Tieto grafy zobrazujú súčet zobrazených hodnôt a znázorňujú tak aj vzťah jednotlivých častí k celku [4].

### *1.5.6. Graf XY (závislosť)*

Graf XY obsahuje dve osi hodnôt: vodorovná os (x) a zvislá os (y) hodnôt. Spája hodnoty x a y na jednotlivé údajové body a zobrazuje ich v nepravidelných intervaloch alebo v klastroch. Grafy XY sa zvyčajne používajú na zobrazenie a porovnávanie číselných hodnôt, ako sú napríklad vedecké údaje, štatistické údaje alebo inžinierske údaje.

Graf XY možno použiť ak chceme zmeniť rozsah vodorovnej osi, chceme urobiť takúto os logaritmickou stupnicou, hodnoty pre vodorovnú os nie sú rovnomerne rozmiestnené, na vodorovnej osi existuje viac údajových bodov, chceme upraviť nezávislé stupnice osí bodového grafu tak, aby zobrazovali viac informácií o údajoch zahŕňajúcich páry alebo zoskupené množiny hodnôt alebo chceme porovnávať množstvo údajov bez ohľadu na čas – čím viac údajov zahrnieme do bodového grafu, tým lepšie porovnanie môžeme vytvoriť [4].

### *1.5.7. Bublínový graf*

V bublínovom grafe je podobne ako v bodovom grafe pridaný tretí stĺpec, ktorý určuje veľkosť v ňom zobrazených bublín, tieto bubliny predstavujú údajové body v rade údajov. Bublínový graf teda zobrazuje trojdimenzionálne dáta do dvojdimenzionálneho priestoru [4].

### *1.5.8. Radarový graf*

Radarový graf porovnáva agregované hodnoty, ktoré predstavujú viaceré rady údajov. Radarový graf je dvojdimenzionálny graf. V Radarovom grafe hodnoty udávajú, akým podielom patria do danej kategórie [4].

## **1.6. Nástroje na vizualizáciu dát**

Nástroje a techniky na vizualizáciu dát pomôžu vytvoriť dvoj- alebo troj- dimenzionálne obrázky dát, ktoré možno ľahko interpretovať a získať tak vedomosti a prehľady o týchto dátach. Vizualnou kontrolou a interakciou s dvoj- alebo trojdimenzionálnou vizualizáciou môžeme identifikovať zaujímavé, implicitné, možno predtým neznáme a potenciálne užitočné informácie alebo vzory v súbore dát. Nástroje na vizualizáciu dát poskytujú ľahší spôsob vytvárania vizuálnych reprezentácií veľkých súborov dát. Pri práci s dátami, ktoré zahŕňajú

státisíce alebo milióny dátových bodov, automatizácia procesu vytvárania vizualizácie výrazne uľahčuje prácu dizajnéra [2].

K dispozícii je obrovské množstvo nástrojov na vizualizáciu dát, ktoré sú obohatené o širokú škálu funkcií. V nasledujúcich podkapitolách sú príklady na nástroje pre vizualizáciu dát.

### *1.6.1. Tableau Public*

Tableau Public je bezplatná platforma na verejné zdieľanie a skúmanie vizualizácií dát online. Vizualizácie môžu byť vložené do webových stránok a blogov, môžu byť zdieľané prostredníctvom sociálnych sietí alebo e-mailu a môžu byť sprístupnené na stiahnutie ďalším používateľom. Softvér sa stal čoraz populárnejším medzi novinármi, spisovateľmi, blogermi, študentmi, občanmi a ďalšími [6].

Ponúka širokú škálu možností vrátane koláčových, čiarových a bublinových grafov, máp, tepelných máp, bodových grafov, ktoré využívajú informačné panely, ktoré je možné okamžite vytvoriť z rôznych súborov dát. Vykonáva agregácie, zvýraznenie alebo rozčlenenie do grafov s takou ľahkosťou, že aj začínajúci používatelia môžu vytvárať vizualizácie na osvetlenie faktov v obrovskom súbore dát. Tableau sa môže bez námahy pripojiť k údajom uloženým v súboroch Excel, CSV a textových súboroch a rozpoznať polia a formáty. Na prístup k údajom uloženým v mnohých databázach však používatelia Tableau stále potrebujú príslušné databázové konektory a odborné znalosti. Tabuľku je možné použiť na definovanie a výpočet nových premenných a na vykonávanie jednoduchých manipulácií s dátami pomocou matematických vzorcov ako Excel, je však oveľa výkonnejšia a dokáže spracovať milióny záznamov, pri ktorých program Excel zlyhá [7].

Obmedzenú analytickú schopnosť Tableau možno prekonať jej schopnosťou integrácie s niekoľkými platformami vrátane platformy Big Data Hadoop, štatistického analytického nástroja R a jeho podpory rozhrania Google Big Query API. Tableau napriek tomu poskytuje natívnu podporu proprietárnym hráčom Big Data, ako sú Amazon Red Shift a TeraData, stále mu chýba priama podpora mnohých prevládajúcich open source databáz NoSQL, ako je MongoDB [7].

## *Verzie Tableau*

Na uspokojenie rôznych potrieb rôznych používateľov ponúka Tableau Software výber z možností: Tableau Desktop, Tableau Server, Tableau Mobile alebo Tableau Public. Tableau Desktop je prítlačlivý pre jednotlivcov i malé organizácie, zatiaľ čo Tableau Server je vhodný pre organizácie s veľkým počtom používateľov [7].

### *Tableau Public*

Tableau Public je bezplatná verzia softvéru Tableau, ktorú je možné použiť na vytváranie vizualizácií. Je to ekonomická verzia, ktorá je veľmi dobrá pre ľudí, ktorí sa chcú učiť a zdieľať svoje údaje s ľuďmi. Súbor sú zverejnené na Tableau Public a nie je možné ich uložiť v systéme. Tableau Public nie je zabezpečený a ktokoľvek má prístup k údajom a môže ich sťahovať [7].

### *Tableau Desktop*

Tableau Desktop je návrhový nástroj na vytváranie vizuálnych analýz a dashboardov. Existujú dve verzie: edícia Personal a edícia Professional. Profesionálny je populárnejší, pretože sa pripája k širšej palete zdrojov údajov ako osobný. Ak neexistuje konektor pre konkrétne údaje, ku ktorým potrebujete získať prístup, môžete sa k nim pripojiť prostredníctvom štandardu Open Database (ODBC). Tableau tiež poskytuje bezplatný nástroj na spotrebu s názvom Tableau Reader. Je to licencovaná verzia s dvojtýždňovou stopou. Dáta sa ľahko importujú do dátového stroja spoločnosti Tableau z viacerých zdrojov a importujú ich kombináciou viacerých zobrazení v interaktívnom paneli. Tableau Desktop produkuje súbory s príponami twb a twbx. Používatelia počítačov Tableau môžu využívať server a online na výkonné úložisko údajov [7].

Tableau Mobile podporuje spotrebiteľov tabletov a mobilných telefónov, ktorí pridali softvéru veľkú hodnotu v dôsledku rozšírenia používania smartfónov medzi jednotlivcami. Aj keď je tablo pre verejnosť bezplatné a je ho možné použiť na vizualizácie v malom meradle, má v porovnaní s tablom obmedzené funkcie [7].

So serverom Tableau je možné zdieľať vizualizácie medzi rôznymi používateľmi a súčasne umožniť obmedzený prístup k rôznym zobrazeniam a podkladovým údajom použitím jeho používateľského filtra alebo bezpečnostnej funkcie na úrovni riadku [7].

## Parametre Tableau

Parametre nástroja	Popis
Operačný systém	Windows a Mac OS
Prístup k softwaru	Desktop a Cloud
Vstupné formáty	Databáza: Teradata, MS SQL Server, MySQL, PostgreSQL, MS Access, Oracle, Oracle Essbase, IBM DB2, Netezza, MS SQL Analysis Services, Sybase, SAP HANA, Cognos, IBM DB2, Tableau Server, Amazon Redshift, Progress, Firebird, Exacol, Apache Drill, Aster Database.  Textové súbory: Text, Excel, CSV, HTML, MS PowerPoint
Vstupné dáta	PDF, MS Excel, CSV, HTML, MS PowerPoint, JSON
Podporované mobilné zariadenie	iOS a Android
Ostatné konektory	ODBC, OLE DB, Salesforce, Google Analytics, Google BigQuery, OneDrive, Microsoft Analysis Services, Dropbox, Web Data Connector

Tabuľka 2 Parametre Tableau [7].

### 1.6.2. Google Charts API

Google Charts API je sada tried v JavaScript, vytvorené programátormi v Google. Google Charts poskytuje jednoduché možnosti na vizualizáciu dát online, priamo na webovej stránke. Zahŕňa množstvo typov grafov od čiarových grafov po zložitejšie hierarchické stromové mapy [8].

Google Charts poskytuje mnoho typov grafov, ktoré môžete použiť. Predvolený vzhľad bude zvyčajne všetko, čo potrebujete, a vždy môžete prispôsobiť graf tak, aby vyhovoval



vzhľad a štýlu vášho webu. Grafy sú vysoko interaktívne a zobrazujú udalosti, vďaka ktorým ich môžete prepojiť a vytvárať zložité informačné panely alebo iné prostredia integrované s vašou webovou stránkou. Grafy sa vykresľujú pomocou technológie HTML5 / SVG, aby sa zaistila kompatibilita medzi rôznymi prehliadačmi (vrátane VML pre staršie verzie prehliadača IE) a prenosnosť na rôzne platformy pre telefóny iPhone, iPad a Android [8].

### 1.6.3. D3.js

D3.js - Data-Driven Documents - je knižnica JavaScript na vytváranie vizualizácie dát. D3 nám vie pomôcť oživiť dáta pomocou HTML, SVG a CSS. Autorom D3 je Mike Bostock, aj keď existuje mnoho ďalších prispievateľov. Je voľne dostupný na GitHub [9].

D3 je vydávaný na základe licencie BSD, takže ho môžeme používať, upravovať a prispôbovať na nekomerčné a komerčné použitie bez akýchkoľvek nákladov. D3 uľahčuje generovanie a manipuláciu s webovými dokumentami pomocou dát. Robí to pomocou [9]:

- Načítania dát do pamäte prehliadača.
- Viazanie údajov na prvky v rámci dokumentu, vytváranie nových prvkov podľa potreby.
- Transformácie prvkov interpretáciou viazaného dátumu každého prvku a zodpovedajúcim spôsobom nastavenie jeho vizuálnych vlastností.
- Prechodové prvky medzi stavmi v reakcii na vstup používateľa.

D3 negeneruje preddefinované ani konzervované vizualizácie. D3 je určený predovšetkým na vysoko prispôbené vizualizačné práce, či už ide o navrhovanie jednorazových vysvetľujúcich tabuliek alebo komplexných interaktívnych prieskumných nástrojov. Je to najsilnejší nástroj na vizualizáciu na webe, pretože umožňuje od nuly vyvíjať všetko. V D3 nie sú žiadne šablóny ani sprievodca grafom. Na D3 je však postavených mnoho vynikajúcich nástrojov, ktoré poskytujú prístup k predkonfigurovaným typom grafov [9].

Hlavné funkcie D3 nespracovávajú bitmapové mapy, ako sú mapy poskytované pomocou Google Maps. D3 neskrýva pôvodné údaje, pretože je D3 kód na strane klienta, klientovi sa musia odoslať požadované vizualizované údaje. Ak údaje nemožno zdieľať,

netreba používať D3. Alternatívy zahŕňajú použitie proprietárnych doplnkov prehliadača alebo predbežné vykreslenie vizualizácií ako statických obrázkov a ich odoslanie do prehliadača [9].

S D3 je možné ľubovoľné údaje viazať na dokumentový objektový model (DOM), a preto je možné v dokumente vykonávať manipuláciu s údajmi, ako aj transformácie založené na údajoch. D3 vytvára vizualizáciu pomocou Scalable Vector Graphics (SVG), HTML a Cascaded Style Sheets (CSS) a spracováva rôzne dátové formáty vrátane CSV, JSON a GeoJSON. Viac podrobností nájdete na webovej stránke. Je to bezplatný a otvorený zdrojový nástroj, ktorý však na tvorbu grafiky vyžaduje základné znalosti html, java skriptu, CSS, SVG. Kódovanie nízkoúrovňových funkcií je potrebné dokonca aj na vytvorenie jednoduchých vizualizácií, ktoré sú pomerne náročné a značne predlžujú čas vývoja, pretože D3 nemá na výber vopred pripravenú knižnicu grafov alebo grafov. Vstavané opakovane použiteľné funkcie a ukážky bežne používanej grafiky sú však ľahko dostupné. D3 dokáže vytvoriť ľubovoľnú predstaviteľnú vizualizáciu v rámci limitov kódovania a môže ponúknuť vynikajúcu interaktivitu. Je to hlavne kvôli jeho transformačnej povahe, ktorá využíva webové štandardy na vytvorenie vizualizácie, kde návrh v ovládacích prvkoch SVG môže priniesť najlepšiu možnú vizualizáciu v porovnaní s inými nástrojmi. Umožňuje animované prechody a grafické zložitosti pri vyšších snímkových frekvenciách. D3 podporuje všetky moderné prehliadače vrátane IE 9+, Android a IOS, ale niektoré funkcie sa v starších prehliadačoch nedokážu zobraziť. V D3 je ťažké dosiahnuť utajenie údajov pre obmedzený prístup medzi rôznymi používateľmi [7].

#### *1.6.4. RAWGraphs*

RAW je bezplatná aplikácia, ktorá uľahčuje používanie knižnice D3 JavaScript. Tento nástroj má množstvo možností prispôsobenia, ktoré vám umožňujú upravovať štýl každého z vašich výtvorov. Jedinou nevýhodou RAW je obmedzená knižnica grafov. Momentálne je na výber 16 možností. Ak ste si však istí, že máte JavaScript, môžete si pridať svoj vlastný a stále sa starať o svoj styling pomocou nástroja WYSIWIG builder [15].

RAW Graphs je otvorený rámec vizualizácie údajov, ktorý bol vytvorený s cieľom uľahčiť každému vizuálnu reprezentáciu komplexných údajov. Cieľom programu RAW Graphs, ktorý bol pôvodne koncipovaný ako nástroj pre dizajnérov, je poskytnúť chýbajúce prepojenie medzi tabuľkovými aplikáciami (napr. Microsoft Excel, Apple Numbers,

OpenRefine) a editormi vektorovej grafiky (napr. Adobe Illustrator, Inkscape, Sketch). Projekt, ktorý vedie a udržiava výskumné laboratórium DensityDesign (Politecnico di Milano), bol vydaný verejne v roku 2013 a mnohí ho považujú za jeden z najdôležitejších nástrojov v oblasti vizualizácie údajov [15].

#### *1.6.5. Charts.js*

Chart.js je bezplatný, otvorený zdroj a udržiava ho aktívna komunita vývojárov na GitHub. Ako knižnica vizualizácie údajov je na GitHub na druhom mieste, čo sa týka počtu hviezdíček, ktoré má. Prvé miesto patrí D3.js, oveľa väčšej a komplexnej knižnici so strmou krivkou učenia. Na to, aby ste mohli používať Chart.js, musíte ešte vedieť JavaScript, HTML a CSS. Okrem jazyka JavaScript je tiež založený na iných webových štandardoch, ako sú DOM, CSS a Canvas. Grafy sa automaticky vykresľujú na plátne a riadia všetku konfiguráciu plátna, jeho veľkosť a pomer pixelov [16].

#### *1.6.6. Infogram*

Infogram má veľa rôznych máp / infografík a poskytuje až 35 rôznych máp a 200 rôznych máp. Poskytuje tiež funkciu súkromného zdieľania. V tomto nástroji je k dispozícii viac ako 20 hotových návrhov. Infogram umožňuje vykonať veľa úprav vizualizácie. Vizualizácie generované týmto nástrojom sú vysoko interaktívne. Hlavnou výhodou tohto nástroja je rozsah dostupných možností vstupu pre údaje. Podporuje xls,xlsx, csv, Dropbox, Onedrive, JSON feed, Google Analytics a umožňuje vyhľadávanie dát v globálnych zdrojoch. Hlavnou nevýhodou však je, že všetky tieto funkcie sú k dispozícii iba v platenej verzii. Bezplatná verzia je veľmi základná a neposkytuje veľa funkcií. Bezplatná verzia navyše obsahuje aj obrovské logo vo vygenerovanej vizualizácii [17].

#### *1.6.7. Leaflet*

Leaflet je open source knižnica JavaScriptu pre mobilné, interaktívne mapy. Bol vyvinutý tímom oddaných prispievateľov pod vedením Vladimíra. Funguje efektívne na všetkých hlavných mobilných a stolných platformách. Primárne využíva výhody HTML5 a

CSS3 v moderných prehliadačoch a je tiež spätne kompatibilný. Poskytuje rozsiahlu sadu doplnkov. Tento nástroj je prezentovaný ako ľahko použiteľné a dobre zdokumentované API.

Je to jednoduchý a čitateľný zdrojový kód, do ktorého vývojár ľahko prispieva. Jednou z hlavných atrakcií tohto nástroja je, že je veľmi ľahký (približne 33 kB) [17].

#### 1.6.8. Porovnanie nástrojov na vizualizáciu dát

Názov nástroja	Open source	Úroveň pôsobenia	Zariadenie na export obrázkov	Špecifické pre doménu	Podpora Javascript	Podporované vstupné formáty
Tableau	Nie	Vysoká	Zabudované	Nie	Áno	Tabuľky, cloudové databázy, relačné databázy
D3	Áno	Vysoká	Manuálne	Nie	Áno	JSON, CSV, geoJSON
RAW	Áno	Vysoká	Zabudované	Nie	Áno	Vstupné dáta
GoogleCharts	Áno	Limitovaná	Nie	Nie	Áno	Vstupné dáta
Infogram	Nie	Vysoká	Zabudované	Nie	Áno	Xls, xlsx, csv, Dropbox, Onedrive, JSON, Google analytics, Vyhľadávanie údajov v globálnych zdrojoch
Leaflet	Áno	Dobrá	Zabudované	Nie	Áno	Csv, kml, GeoJSON, GPX, WMX

Tabuľka 3 Porovnanie nástrojov na vizualizáciu dát [17].

## 2. Cieľ práce

Cieľom záverečnej práce je analyzovať ako efektívne vizualizovať dáta podľa charakteristík a porovnať IT nástroje na vizualizáciu dát.

V súčasnosti dáta predstavujú základ pre rozhodovanie v praxi. Ich vizualizácia značne uľahčuje rozhodovanie. Práca je zameraná na objasnenie charakteristík dát, podľa ktorých je vhodné hľadať vhodný typ vizualizácie a následne porovnáva vybrané IT nástroje na vizualizáciu dát.

Aby sme dosiahli hlavný cieľ, potrebujeme urobiť vyhľadávanie údajov na tému k publikovanej literatúre ako aj súhrn či prehľad týchto údajov a postupne charakterizovať a definovať základné pojmy týkajúce sa vizualizácie dát, procesu transformácie dát do vhodnej formy.

V prvej kapitole sa venujeme pojmom, čo je to vizualizácia dát, aký je proces vizualizácie dát a popísali sme, aké štandardné dátové formáty poznáme. Pri práci s vizualizáciou dát sa používajú hlavne dátové formáty JSON, CSV a XML [14]. Obrazovou prezentáciou informácií je graf a preto sú popísané najčastejšie používané typy grafov. Taktiež v prvej kapitole sú definované najlepšie, resp. najpoužívanejšie nástroje na vizualizáciu dát [17] a z nich vybrané do praktickej časti dva nástroje na základe zvolených kritérií.

V praktickej časti budeme pracovať s dvoma nástrojmi na vizualizáciu dát. Parciálne ciele v praktickej časti sú postupne nasledujúce:

- Zvolenie kritérií pre výber nástrojov a následne výber nástrojov.
- Naštudovanie manuálov ako pracovať s nástrojmi a ich inštalácia.
- Výber vhodného datasetu na vizualizáciu.
- Príprava dát na spracovanie pomocou nástrojov.
- Testovanie vizuálnej reprezentácie.
- Porovnanie získaných výsledkov vizualizácie dát prostredníctvom dvoch vybraných nástrojov.

V záverečnej práci používame metódy ako je analýza, predovšetkým v teoretickej časti, kde je potrebné analyzovať zdroje a urobiť z nich výber. Taktiež je potrebné preštudovať a porovnať rôzne nástroje na vizualizáciu dát a urobiť výber.

V záverečnej práci používame aj metódu selekcie a komparácie (porovnávanie). Čiastočne budeme v praktickej časti narábať s metodikou procesu pre datamining a vizualizáciu dát (CRISP metodika - Cross Industry Standard Process for Data Mining).

### 3. Metodika práce a metódy skúmania

V praktickej časti bakalárskej práce sme sa zoznámili s rôznymi druhmi vizualizácie dát. Tieto druhy vizualizácie dát sme aplikovali na vybrané dáta o COVID-19. Zdrojom dát je databáza z webovej stránky [www.kaggle.com](http://www.kaggle.com), ktorá zhromažďuje rôzne databázy [10].

#### 3.1. Vizualizácia dát v Tableau

Pandémia COVID-19 zasiahla takmer všetky aspekty života na celom svete a dodávateľia dátovej analýzy sa mobilizujú, aby pomohli jednotlivcom, vedcom, vládam a podnikom pochopiť podobu a rozsah krízy. V tomto scenári je Tableau veľmi výhodným nástrojom. Pretože Tableau je nástroj na analýzu a vizualizáciu údajov, ktorý sa dá pohodlne spojiť s mnohými zdrojmi údajov.

Pokračujúce šírenie choroby COVID-19 (koronavírus) rýchlo mení náš každodenný život, potrebujeme analýzu údajov a vizualizácie. V tomto scenári sa Tableau často používa ako nástroj, ktorý pomáha rýchlo analyzovať a vytvárať vizualizácie dát pracovných listov a dashboardov. Tableau nám umožňuje vytvárať dashboardy, ktoré poskytujú použiteľné informácie a rýchlejšie rozširujú podnikanie. Produkty Tableau sa vždy prevádzkujú vo virtualizovaných prostrediach, ak sú nakonfigurované so správnym hardvérom a operačným systémom. S Tableau sa môže pracovať s neusporiadanými dátami COVID-19 (koronavírus) [10] a pomocou Tableau vytvárať rôzne vizualizácie pomocou zabudovaných funkcií, ktoré ponúka Tableau.

Tableau pomáha organizáciám pri analýze budúcich dát bez ohľadu na ich budúce ciele. Nástroj Tableau skúma vizualizácie a pozorovanie dát o COVID-19 (koronavírus). Ukážka je uvedená na obrázku 2.

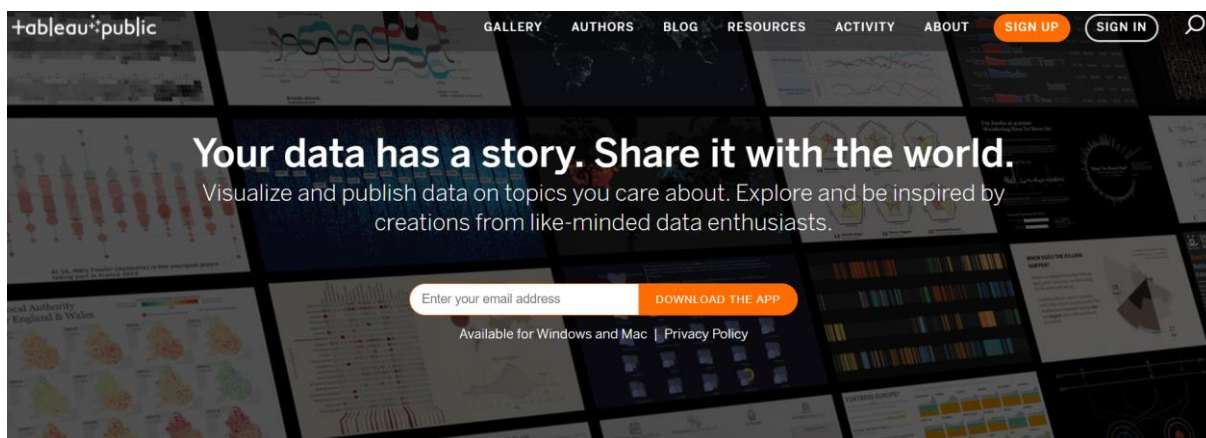
	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q
1	Country/Region	Confirmed	Deaths	Recovered	Active	New cases	New deaths	New recovery	Deaths / 1000	Recovered / 1000	Deaths / 1000	Confirmed / 1000	1 week change	% WHO Region			
2	Afghanistan	36263	1269	25198	9796	106	10	18	3,5	69,49	5,04	35526	737	2,07	Eastern Mediterranean		
3	Albania	4880	144	2745	1991	117	6	63	2,95	56,25	5,25	4171	709	17	Europe		
4	Algeria	27973	1163	18837	7973	616	8	749	4,16	67,34	6,17	23691	4282	18,07	Africa		
5	Andorra	907	52	803	52	10	0	0	5,73	88,53	6,48	884	23	2,6	Europe		
6	Angola	950	41	242	667	18	1	0	4,32	25,47	16,94	749	201	26,84	Africa		
7	Antigua and Barb	86	3	65	18	4	0	5	3,49	75,58	4,62	76	10	13,16	Americas		
8	Argentina	167416	3059	72575	91782	4890	120	2057	1,83	43,35	4,21	130774	36642	28,02	Americas		
9	Armenia	37390	711	26665	10014	73	6	187	1,9	71,32	2,67	34981	2409	6,89	Europe		
10	Australia	15303	167	9311	5825	368	6	137	1,09	60,84	1,79	12428	2875	23,13	Western Pacific		
11	Austria	20558	713	18246	1599	86	1	37	3,47	88,75	3,91	19743	815	4,13	Europe		
12	Azerbaijan	30446	423	23242	6781	396	6	558	1,39	76,34	1,82	27890	2556	9,16	Europe		
13	Bahamas	382	11	91	280	40	0	0	2,88	23,82	12,09	174	208	119,54	Americas		
14	Bahrain	39482	141	36110	3231	351	1	421	0,36	91,46	0,39	36936	2546	6,89	Eastern Mediterranean		
15	Bangladesh	226225	2965	125683	97577	2772	37	1801	1,31	55,56	2,36	207453	18772	9,05	South-East Asia		
16	Barbados	110	7	94	9	0	0	0	6,36	85,45	7,45	106	4	3,77	Americas		
17	Belarus	67251	538	60492	6221	119	4	67	0,8	89,95	0,89	66213	1038	1,57	Europe		
18	Belgium	66428	9822	17452	39154	402	1	14	14,79	26,27	56,28	64094	2334	3,64	Europe		
19	Belize	48	2	26	20	0	0	0	4,17	54,17	7,69	40	8	20	Americas		
20	Benin	1770	35	1036	699	0	0	0	1,98	58,53	3,38	1602	168	10,49	Africa		
21	Bhutan	99	0	86	13	4	0	1	0	86,87	0	90	9	10	South-East Asia		

Obrázok 2 Ukážka z dátového súboru – informácie o COVID-19 [10]. (Country/Region – krajina/oblasť, Confirmed – Potvrdené (prípady), Deaths – Úmrtia, Recovered – Zotavený (prekonali COVID-19), Active – Aktívny (majú COVID-19), New cases – Nové prípady, New deaths – nové úmrtia, New recovery – nový vyliečený)

Postupným pridávaním rozmerov do grafov sme ukázali, ako zostrojovať prehľadné vizualizácie a vyčítať z nich potrebné informácie.

### Získanie Tableau Public a prvý kontakt

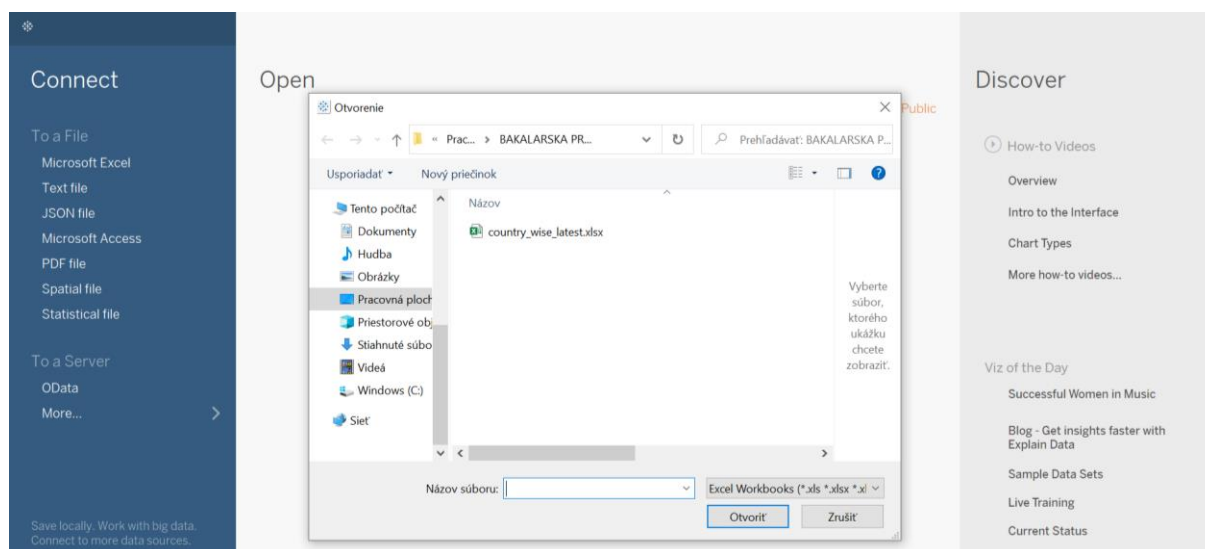
Na získanie vizualizačného nástroja Tableau Public je potrebné navštíviť webovú stránku [11], vyplniť svoju e-mailovú adresu a stlačiť tlačidlo „DOWNLOAD THE APP“. Po stlačení tlačidla sa začne automaticky sťahovať Tableau do počítača. Nasledujúci obrázok 3 znázorňuje webovú stránku po otvorení odkazu [11].



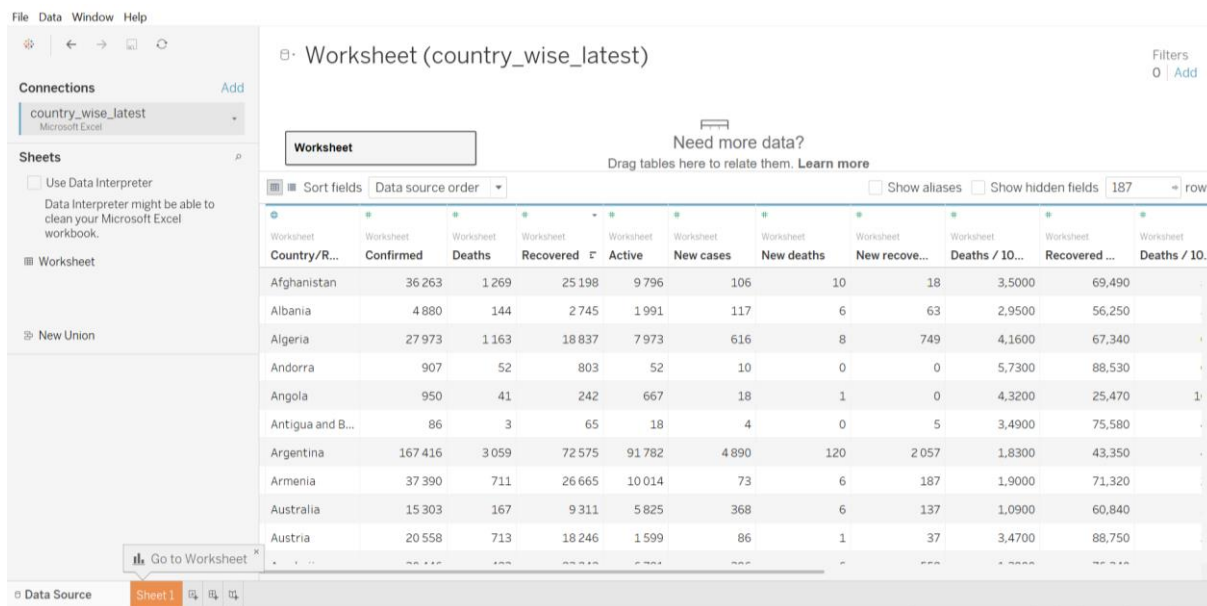


Obrázok 3 Tableau Public – sťahovanie aplikácie Tableau Public z webovej stránky [11].

Aby sme mohli začať pracovať s Tableau, musíme Tableau pripojiť k zdroju dát. Tableau je kompatibilný s mnohými zdrojmi údajov. Zdroje dát podporované Tableau sa zobrazujú na ľavej strane úvodnej obrazovky. Niektoré bežne používané zdroje dát sú program Excel, textový súbor alebo relačná databáza. Dá sa tiež pripojiť k zdroju cloudovej databázy, ako sú Google Analytics, Amazon Redshift a podobne. Po otvorení počítačového programu, používateľ si zvolí zdroj dát. Dáta, sa nachádzajú v MS Excel súbore, takže na ľavej lište „Connect“ si v časti „To a File“ sme zvolili „Microsoft Excel“ a vybrali sme dátový súbor, ktorý sme potrebovali. Na Obrázku 4 je postup pridania dátového zdroja a na obrázku 5 je v Tableau pripojený dátový zdroj.



Obrázok 4 Postup pridávania dátového zdroja na prácu v Tableau Public (dátový zdroj [10]).



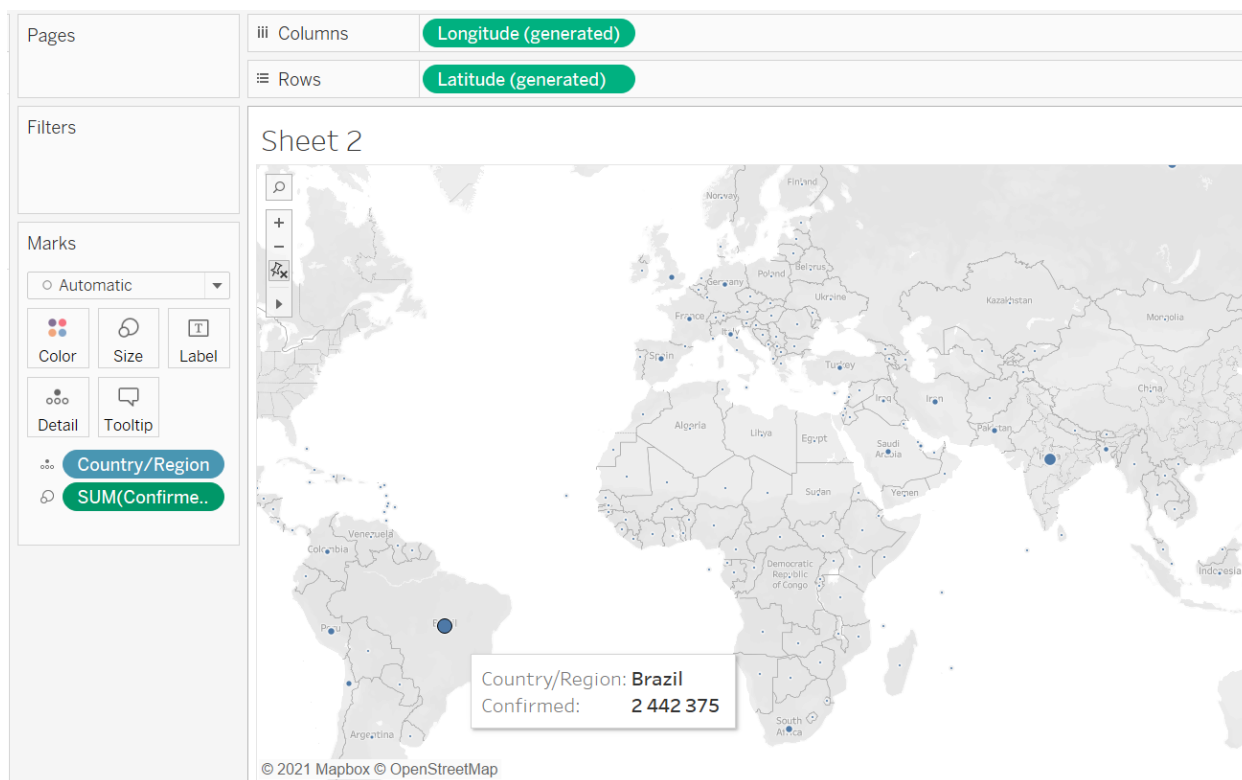
Obrázok 5 Dátový zdroj v Tableau Public (dátový zdroj [10])

Tableau poskytuje základné a pokročilé spôsoby triedenia, ku ktorým je ľahký prístup prostredníctvom ikon alebo ponúk. Najzákladnejší spôsob triedenia je pomocou ikon, ktoré sa zobrazujú v ponuke panela nástrojov. Ikony zoradenia ponuky na paneli nástrojov poskytujú vzostupné a zostupné zoradenie.

Tableau tiež poskytuje ikony triedenia blízko nadpisov a značiek. Kliknutím na ikonu triedenia, ktorá sa vznáša nad pravou stranou nadpisu podkategórie, sa poskytuje vzostupné a zostupné zoradenie pomocou textu kategórie produktu. Ikony triedenia, ktoré sa objavujú nad a pod osou značky, poskytujú vzostupné a zostupné zoradenie na základe hodnôt zobrazených značkami a tiež pridávajú zoradenie podľa poradia zdrojov dát.

Tableau ponúka zaujímavé mapovacie schopnosti na prekrytie našich údajov do relevantných máp. Tableau automaticky priradí geografickú rolu dátovým položkám, ktoré majú priradené určité názvy a dátové typy zobrazené na obrázku 6. Všetky ostatné položky majú predvolenú geografickú rolu - žiadna.

Ak Tableau správne neidentifikuje našu geografickú položku, meníme túto rolu kliknutím pravým tlačidlom myši na položku, výberom položky Geografická rola a výberom správnej geografickej roly zo zoznamu. Keď je našej dátovej položke priradená geografická rola, odlišuje sa to malou ikonou zemegule. Naš súbor dát môže obsahovať niektoré nesprávne kódované položky, ktoré Tableau nedokáže mapovať. Na obrázku 6 je Geografická rola, vytvorená pomocou Tableau, ktorá obsahuje dáta o COVID-19 [10].



Obrázok 6 Geografické roly v Tableau s dátami o COVID-19 (dátový zdroj [10]).

### 3.2. Vizualizácia dát v D3.js

O tvorbe vizualizácie dát pomocou nástroja D3.js sme sa informovali na GitHubu [20].

Knižnicu D3 sme použili jej priamym prepojením na stránku HTML zo siete Content Delivery Network (CDN). CDN je sieť serverov, na ktorých sa hostia súbory a doručujú sa používateľovi na základe ich geografického umiestnenia.

Knižnicu D3 sme zahrnuli pomocou adresy URL - CDN „<https://d3js.org/d3.v4.min.js>“ na svoju stránku.

```
<!doctype html>
<html lang="en">
  <head>
    <meta charset="utf-8">
    <meta name="viewport" content="width=device-width, initial-scale=1, shrink-to-fit=no">
    <script src="https://d3js.org/d3.v4.min.js"></script>
    <style>
      svg {
        border: 1px solid #ddd;
      }
    </style>
  </head>
```

Ďalej máme vykreslené dáta.

```
<body>
  <div id="chart" style="width: 600px"></div>
  <script>
    const data = [1269, 144, 1163, 52, 41, 3, 3059, 711, 167, 423, 11, 141, 2965, 7, 538, 9822, 2, 35];
    const labels = ['Afghanistan','Albania','Algeria','Andorra','Angola','Antigua and Barbuda','Argentina','Armenia','Australia',
      'Austria','Azerbaijan','Bahamas','Bahrain','Bangladesh','Barbados','Belarus','Belgium','Belize','Benin'];
```

Dáta sa dajú vykresliť pomocou pripojenia .csv súboru tak, ako je nižšie uvedené.

```
d3.csv("abcd.csv", function(error, data) {
  if (error) {
    throw error;
  }
```

```
});
```

V tomto kroku sa načíta súbor abcd.csv pomocou metódy d3.csv(). Pridali sme spracovanie chýb v prípade, že sa súbor nepodarí načítať.

Maximálna šírka jedného stĺpca, takže stĺpec nevyzerá ako box, pomocou:

```
const max_bar_width = 100;
```

Maximálna výška prvku SVG (obsahuje posunutie zhora a zdola):

```
const svg_height = 600;
```

Farba stĺpcov je vytvorená pomocou:

```
const bar_color = "#1CF0C1";
```

Horný a dolný okraj grafu:

```
const top_offset = 50;  
const bottom_offset = 50;
```

Stmavenie alebo zosvetlenie farby, pri posunutí kurzoru na stĺpec graf stmavne. (Táto časť kódu je použitá z <https://stackoverflow.com/a/13532993/10468888>)

```
function shadeColor(color, percent) {  
  
    var R = parseInt(color.substring(1,3),16);  
    var G = parseInt(color.substring(3,5),16);  
    var B = parseInt(color.substring(5,7),16);  
  
    R = parseInt(R * (100 + percent) / 100);  
    G = parseInt(G * (100 + percent) / 100);  
    B = parseInt(B * (100 + percent) / 100);  
  
    R = (R<255)?R:255;  
    G = (G<255)?G:255;  
    B = (B<255)?B:255;  
  
    var RR = ((R.toString(16).length==1)?"0"+R.toString(16):R.toString(16));  
    var GG = ((G.toString(16).length==1)?"0"+G.toString(16):G.toString(16));
```

```

var BB = ((B.toString(16).length==1)?'0'+B.toString(16):B.toString(16));

return '#' + RR + GG + BB;
}

```

Pripojenie SVG (Scalable Vector Graphics):

```

const svg = d3.select("#chart")
  .append("svg")
  .attr("fill", "red")
  .attr("width", '100%')
  .attr("height", svg_height);

```

Aby graf reagoval na výpočet, nastavíme šírku 100% - šírku vypočítame v pixeloch:

```

const svg_width = svg.node().getBoundingClientRect().width;

```

Rozhodnutie o šírke stĺpca v závislosti od dostupného priestoru a počtu stĺpcov, ktoré sa majú vykresliť:

```

let bar_width = Math.round((svg_width - 100) / data.length);
if (bar_width > max_bar_width) {
  bar_width = max_bar_width;
}

```

Nastavenie medzery medzi dvoma stĺpcami - namiesto pevnej hodnoty ju nastavujeme dynamicky:

```

const spacing = 0.5 * bar_width;

```

Do stredu zarovnáme graf, ktorý posunieme zľava doprava:

```

let left_offset = Math.round((svg_width - bar_width * data.length) / 2);
if (left_offset < 0) {
  left_offset = 0;
}

```

Pomocou tejto časti sme vytvorili mierku pre os X:

```
const scale = d3.scaleLinear()
  .domain([0, Math.max(...data)])
  .range([0, svg_height - top_offset - bottom_offset]);
```

Vytvorenie mierky pre os Y - mohli sme použiť vyššie uvedenú mierku, ale pre os Y potrebujeme obrátenú doménu.

.domain() - toto je napísané na osi: maximálna hodnota dát o úmrtiach až 0 hodnoty.

.range() - toto je miesto, kde je umiestnená os.

```
const scale_y_axis = d3.scaleLinear()
  .domain([Math.max(...data), 0])
  .range([0, svg_height - top_offset - bottom_offset]);
```

Vytvorenie prvku popisu:

```
const tooltip = d3.select("body")
  .append("div")
  .attr("class", "d3-tooltip")
  .style("position", "absolute")
  .style("z-index", "10")
  .style("visibility", "hidden")
  .style("padding", "15px")
  .style("background", "rgba(0,0,0,0.6)")
  .style("border-radius", "5px")
  .style("color", "#fff")
  .text("Bar Chart COVID-19");
```

Pridanie obdĺžnika:

```
const rect = svg.selectAll("g")
  .data(data)
  .enter()
  .append("rect")
  .attr("fill", bar_color)
  .attr("x", (d, i) => left_offset + bar_width * i)
  .attr("y", d => svg_height - bottom_offset)
  .attr("width", bar_width - spacing)
  .on("mouseover", function(d, i) {
    tooltip.html(`Data: ${d}`).style("visibility", "visible");
    d3.select(this)
      .attr("fill", shadeColor(bar_color, -15));
  })
  .on("mousemove", function(){
    tooltip
```

```

        .style("top", (event.pageY-10)+"px")
        .style("left",(event.pageX+10)+"px");
    })
    .on("mouseout", function() {
        tooltip.html(`).style("visibility", "hidden");
        d3.select(this).attr("fill", bar_color);
    });

```

Pridanie textu:

```

svg.selectAll("g")
  .data(data)
  .enter()
  .append("text")
  .attr("dominant-baseline", "text-before-edge")
  .attr("text-anchor", "middle")
  .attr("fill", "#000000")
  .attr("x", (d, i) => left_offset + bar_width * i + bar_width/2 - spacing/2)
  .attr("y", svg_height - bottom_offset + 5)
  .attr("style", "font-family:Verdana")
  .text((d, i) => labels[i]);

```

Pridanie osi x:

```

svg.append("line")
  .attr("stroke", "#000000")
  .attr("stroke-width", 2)
  .attr("x1", left_offset)
  .attr("y1", svg_height - bottom_offset)
  .attr("x2", bar_width * data.length + left_offset - spacing)
  .attr("y2", svg_height - bottom_offset);

```

Pridanie osi y:

.attr() - Riadi polohu osi.

```

svg.append("g")
  .attr("transform", "translate(0," + top_offset + ")")
  .call(d3.axisRight(scale_y_axis));

window.onload = () => {

```

Nastavenie animácie (nastavenie animácie pri zobrazení grafu):



```

rect.transition()
  .ease(d3.easeLinear)
  .duration(1000)
  .attr('y', d => svg_height - bottom_offset - scale(d))
  .attr("height", d => scale(d));
};
</script>
</body>
</html>

```

Celý kód, vytvorený pomocou nástroja D3.js v HTML prostredí. Pomocou tohto kódu bol vytvorený stĺpcový graf, ktorý je na obrázku 8.

```

<!DOCTYPE html>

<html lang="en">
  <head>
    <meta charset="utf-8">
    <meta name="viewport" content="width=device-width, initial-scale=1, shrink-to-fit=no">
    <script src="https://d3js.org/d3.v4.min.js"></script>
    <style>
      svg {
        border: 1px solid #ddd;
      }
    </style>
  </head>
  <body>
    <div id="chart" style="width: 600px"></div>
    <script>

      const data = [1269, 144, 1163, 52, 41, 3, 3059, 711, 167, 423, 11, 141, 2965, 7, 538, 9822, 2, 35];
      const labels = ['Afghanistan', 'Albania', 'Algeria', 'Andorra', 'Angola', 'Antigua and Barbuda', 'Argentina', 'Armenia', 'Australia', 'Austria', 'Azerbaijan', 'Bahamas', 'Bahrain', 'Bangladesh', 'Barbados', 'Belarus', 'Belgium', 'Belize', 'Benin'];

      const max_bar_width = 100;
      const svg_height = 600;
      const bar_color = "#1CF0C1";
      const top_offset = 50;
      const bottom_offset = 50;

      function shadeColor(color, percent) {

```

```

var R = parseInt(color.substring(1,3),16);
var G = parseInt(color.substring(3,5),16);
var B = parseInt(color.substring(5,7),16);

R = parseInt(R * (100 + percent) / 100);
G = parseInt(G * (100 + percent) / 100);
B = parseInt(B * (100 + percent) / 100);

R = (R<255)?R:255;
G = (G<255)?G:255;
B = (B<255)?B:255;

var RR = ((R.toString(16).length==1)?"0"+R.toString(16):R.toString(16));
var GG = ((G.toString(16).length==1)?"0"+G.toString(16):G.toString(16));
var BB = ((B.toString(16).length==1)?"0"+B.toString(16):B.toString(16));

return "#" + RR + GG + BB;
}

const svg = d3.select("#chart")
  .append("svg")
  .attr("fill", "red")
  .attr("width", '100%')
  .attr("height", svg_height);

const svg_width = svg.node().getBoundingClientRect().width;

let bar_width = Math.round((svg_width - 100) / data.length);
if (bar_width > max_bar_width) {
  bar_width = max_bar_width;
}

const spacing = 0.5 * bar_width;

let left_offset = Math.round((svg_width - bar_width * data.length) / 2);
if (left_offset < 0) {
  left_offset = 0;
}

const scale = d3.scaleLinear()
  .domain([0, Math.max(...data)])
  .range([0, svg_height - top_offset - bottom_offset]);

const scale_y_axis = d3.scaleLinear()
  .domain([Math.max(...data), 0])
  .range([0, svg_height - top_offset - bottom_offset]);

```

```

const tooltip = d3.select("body")
  .append("div")
  .attr("class", "d3-tooltip")
  .style("position", "absolute")
  .style("z-index", "10")
  .style("visibility", "hidden")
  .style("padding", "15px")
  .style("background", "rgba(0,0,0,0.6)")
  .style("border-radius", "5px")
  .style("color", "#fff")
  .text("Bar Chart COVID-19");

const rect = svg.selectAll("g")
  .data(data)
  .enter()
  .append("rect")
  .attr("fill", bar_color)
  .attr("x", (d, i) => left_offset + bar_width * i)
  .attr("y", d => svg_height - bottom_offset)
  .attr("width", bar_width - spacing)
  .on("mouseover", function(d, i) {
    tooltip.html(`Data: ${d}`).style("visibility", "visible");
    d3.select(this)
      .attr("fill", shadeColor(bar_color, -15));
  })
  .on("mousemove", function(){
    tooltip
      .style("top", (event.pageY-10)+"px")
      .style("left", (event.pageX+10)+"px");
  })
  .on("mouseout", function() {
    tooltip.html("").style("visibility", "hidden");
    d3.select(this).attr("fill", bar_color);
  });

svg.selectAll("g")
  .data(data)
  .enter()
  .append("text")
  .attr("dominant-baseline", "text-before-edge")
  .attr("text-anchor", "middle")
  .attr("fill", "#000000")
  .attr("x", (d, i) => left_offset + bar_width * i + bar_width/2 - spacing/2)
  .attr("y", svg_height - bottom_offset + 5)
  .attr("style", "font-family: Verdana")
  .text((d, i) => labels[i]);

```

```

svg.append('line')
  .attr('stroke', '#000000')
  .attr('stroke-width', 2)
  .attr('x1', left_offset)
  .attr('y1', svg_height - bottom_offset)
  .attr('x2', bar_width * data.length + left_offset - spacing)
  .attr('y2', svg_height - bottom_offset);

svg.append('g')
  .attr('transform', "translate(0," + top_offset + ")")
  .call(d3.axisRight(scale_y_axis));

window.onload = () => {

  rect.transition()
    .ease(d3.easeLinear)
    .duration(1000)
    .attr('y', d => svg_height - bottom_offset - scale(d))
    .attr('height', d => scale(d));
};
</script>
</body>
</html>

```

## 4. Výsledky a diskusia

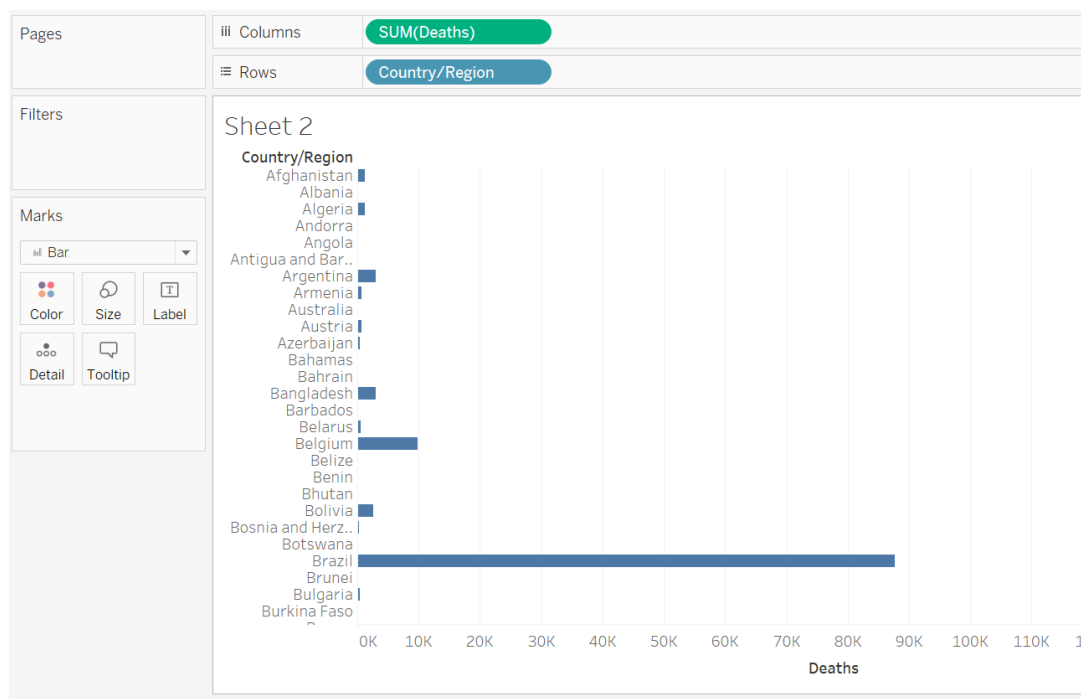
### 4.1. Výsledky

V bakalárskej práci sme pracovali s databázou, ktorá obsahovala dáta o prípadoch nakazení a úmrtia na COVID-19. Databázu sme použili zo zdroja [10].

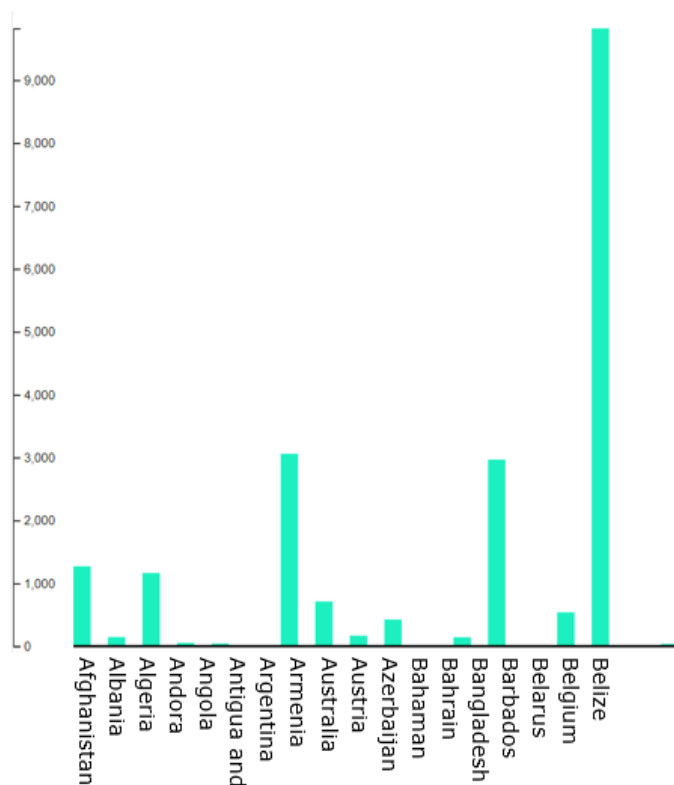
Pomocou databázy o COVID-19 sme vytvorili pomocou dvoch vizualizačných nástrojov, ktoré sú Tableau a D3.js ich vizualizáciu.

Výsledkom sú na obrázku 7 a obrázku 8 dva stĺpcové grafy, ktoré zahrňujú dáta o úmrtiach na COVID-19 v rôznych štátoch a regiónoch. Stĺpcový graf na obrázku 7 bol získaný pomocou nástroja Tableau a na obrázku 8 bol získaný pomocou nástroja D3.js. Napriek tomu, že grafy majú mierne odlišný vzhľad, tak ich hodnoty sú rovnaké. Na obrázku 8 má stĺpcový graf, vytvorený pomocou vizualizačného nástroja D3.js na osi y hodnoty o úmrtiach ľudí na COVID-19 a na osi x názvy postihnutých štátov a regiónov. Stĺpcový graf na obrázku 7, vytvorený pomocou vizualizačného nástroja Tableau obsahuje názvy postihnutých štátov a regiónov na osi y a na osi x obsahuje hodnoty o úmrtiach na COVID-19.

Stĺpcový graf vytvorený pomocou nástroja Tableau bol menej časovo náročný ako v prípade tvorby stĺpcového grafu pomocou nástroja D3.js. Používatelia pri vizualizácii pomocou nástroja D3.js majú kontrolu nad tým, ako chcú formovať a vytvárať interaktívne vizuálne dáta. Pomocou nástroja Tableau sme vedeli urobiť rôzne vizualizácie dát a nemal problém vytvoriť vizualizáciu aj pomocou geografického (mapového) grafu (obrázok 7), s dátami z databázy o COVID-19 o úmrtiach na COVID-19 v postihnutých štátoch a regiónoch.



Obrázok 7 Vizualizácia dátového súboru COVID-19 [10] pomocou nástroja Tableau (Na osi y sú informácie o štátoch a regiónoch a na osi x sú znázornené dáta o úmrtiach na COVID-19).

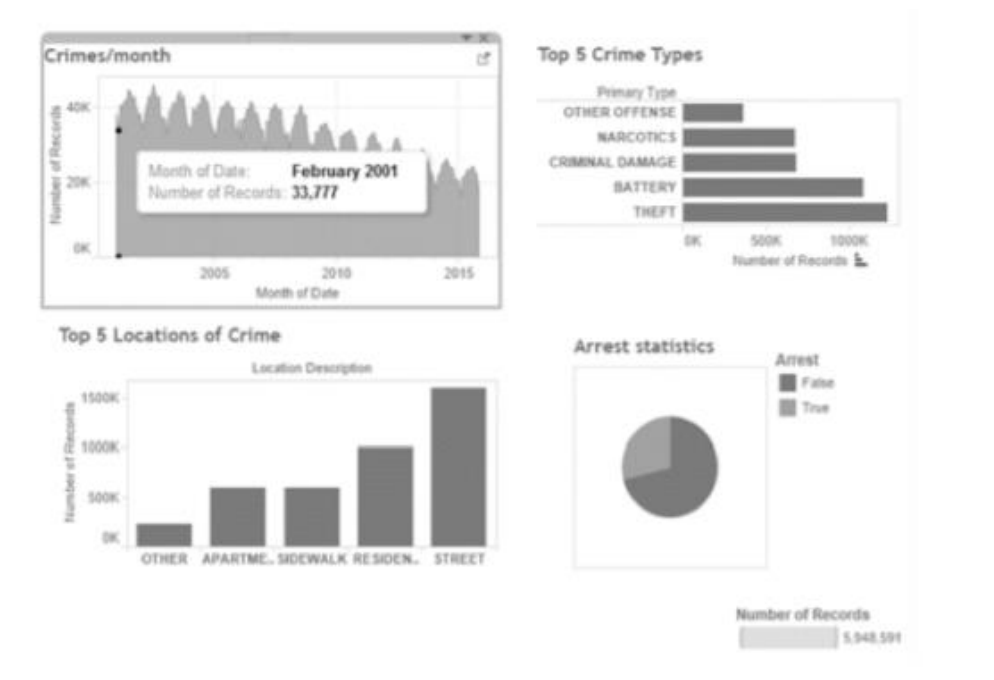


Obrázok 8 Vizualizácia dátového súboru COVID-19 [10] pomocou nástroja D3.js (Na osi y sú dáta o úmrtiach na COVID-19 a na osi x sú znázornené informácie o štátoch a regiónoch.).

## 4.2. Diskusia

Lekha et al. [7] v článku implementovali vizualizáciu veľkého súboru údajov o kriminalite mesta Chicago, kde boli odhalené fakty a trendy týkajúce sa trestných činov v priebehu rokov, a to využitím dvoch vizualizačných nástrojov, Tableau a D3.js. Vykonalí komparatívnu analýzu uvedených nástrojov.

Dátovým súborom použitým pre ich vizualizáciu pomocou nástroja Tableau a nástroja D3.js, bol súbor CSV „zločin 2001 až súčasnosť“ z dátového portálu v Chicagu, ktorý bol sprístupnený 7. decembra 2015 a ktorý obsahuje 5 948 951 záznamov s 23 atribútmi, pričom každý záznam odráža iné trestné činy hlásené v meste Chicago, prevzaté zo systému chicagského policajného oddelenia CLEAR (Citizen Law Enforcement Analysis and Reporting).



Obrázok 9 Vizualizácia dátového súboru kriminality v Chicagu pomocou nástroja Tableau [7].



Obrázok 10 Vizualizácia dátového súboru kriminality v Chicagu pomocou nástroja D3.js v rokoch 2012 – 2015 [7].

Na obrázku 9 je znázornená vizualizácia pomocou nástroja Tableau a na obrázku 10 je znázornená vizualizácia pomocou nástroja D3.js. Pri vizualizácii pomocou nástroja D3.js tvrdili, že okrem zobrazenia presných mier nad oblasťami grafu poskytovala vizualizácia vynikajúcu interaktivitu, ako napríklad výber oblasti (časového obdobia) v grafe zločinov za mesiac, aby bolo možné sledovať zmeny odrážané aj v iných grafoch. Ale vizualizácia nedokázala zobraziť grafy, keď išlo o kompletný súbor údajov asi 6 miliónov záznamov.

Lekha et al. [7] zistili, že Tableau dokáže identifikovať rozmery a miery celého súboru dát a vytvoriť vizualizáciu za niekoľko minút, bez toho, aby museli napísať jeden riadok kódu. Ďalej zistili, že D3.js nedokázal vytvoriť vizualizáciu ak zahrnuli celý súbor dát s približne 6 miliónmi záznamov. Pre rýchlu a ľahkú vizualizáciu zahrňajúcu bežne používané mapy alebo grafy je teda Tableau vhodný a súbor D3.js je možné použiť, ak existujú mimoriadne požiadavky na mapovanie alebo vysoké požiadavky na interaktivitu.

Ich záverom bolo, že vizuálna analýza súboru dát odhalila zaujímavé vzory, ktoré by si inak mohli nevšimnúť. Porovnávacia analýza dvoch nástrojov odhalila, že vizualizačné nástroje otvoreného zdroja vyžadujú veľa vylepšení, aby uspokojili potreby vznikajúcich problémov s vizualizáciou veľkých dát. Žiadny nástroj nemohol byť dokonalým riešením pre všetky problémy s vizualizáciou veľkých dát, kde ho bolo treba zvoliť na základe mnohých faktorov vrátane nákladov a času na vývoj, dostupných odborných znalostí, objemu a funkcií dátovej



sady, požiadaviek na vizualizáciu vrátane vizuálnej príťažlivosti, znalostného obsahu, úroveň interaktivity a možností zdieľania.

Bendremer et al. [21] vo svojej práci mali kľúčovú výzvu vyvinúť vizualizácie, ktoré by boli užitočné a implementovateľné. Cieľom práce bolo vytvoriť vizualizačný nástroj, ktorý by bol užitočný pre napríklad akademických poradcov alebo Centrum kariérneho rozvoja.

Medzi kľúčové aspekty vývoja vizualizácií patrili rozhodnutia o tom, ako usporiadať dashboard, výkon filtrovania a interaktivity a zabezpečenie toho, aby vizualizácie boli skutočne zmysluplné.

Softvér na vizualizáciu dát Tableau sa spočiatku zdal pre nich byť najlepším riešením pre projekt. WPI (Worcester Polytechnic Institute - súkromná výskumná univerzita vo Worcester v štáte Massachusetts) už Tableau používa, takže malo pre nich zmysel investovať čas do programu, ktorý by pracovníci WPI mohli v budúcnosti ľahko aktualizovať. Tableau vyniká rýchlou vizualizáciou dát pre získanie štatistík. Avšak čokoľvek, čo prekonáva základnú vizualizáciu, si vyžaduje hlbokú znalosť Tableau, pretože to v podstate vyžaduje napísanie vizualizácie úplne od začiatku.

D3.js je jednou z najpopulárnejších knižníc webovej vizualizácie a prešli na D3.js predovšetkým z týchto dôvodov:

1. Výkon a nástroje - D3.js je navrhnutý na vytváranie interaktívnych vizualizácií. D3 navyše ponúka veľa „pomocných“ funkcií, aby sa zabránilo opakovanému výpočtu.

2. Väčšia podpora + Komunita - D3.js beží na JavaScripte, ktorý je od roku 2021 asi najbežnejším programovacím jazykom. Vďaka obľúbenosti tohto jazyka, komunita podpory jazyka JavaScript (a D3.js) je podstatne väčšia ako Tableau, čo umožňuje rýchlejšie ladenie a tvorbu prototypov.

Vizualizácia dát pomocou nástroja Tableau pri používaní databázy s informáciami o úmrtiach na COVID-19 bola jednoduchšia ako pri používaní nástroja D3.js, nemuseli sme zasahovať do databázy. Nástroj Tableau vytvoril vizualizácie také, aké sme potrebovali bez toho, aby sme museli napísať jeden alebo viac riadkov kódu. Nástroj Tableau bez žiadnych problémov vytvoril stĺpcový graf o úmrtiach na COVID-19 v daných krajinách a regiónoch. Načítanie dát trvalo pár minút. Podľa Lekha et al. [7] vývoj vizualizácie pomocou D3.js zahŕňal niekoľko úloh vrátane vytvorenia správnej štruktúry a napísania niekoľko riadkov kódu.

Nástroj D3.js zlyhal pri tvorbe vizualizácií, keď bol zahrnutý celý súbor dát. D3.js síce nedokázal pracovať s databázou s viac ako 6 miliónmi záznamov, ale má výhody, ktoré sú lepšie ako pri vizualizácií pomocou Tableau. D3.js má širokú komunitu podpory, je open source a nepotrebujeme na vizualizáciu v D3.js žiadnu inú technológiu okrem prehliadača.

Pri vytváraní vizualizácie pomocou nástroja Tableau a D3.js sme tiež zistili, že vizualizačný nástroj Tableau je rýchlejší, vytvorí vizualizácie za pár minút, pričom v porovnaní s vizualizačným nástrojom D3.js potrebujeme viac znalostí na tvorbu vizualizácie a viac času na vizualizáciu.

Tableau má svoje vlastné obmedzené prostredie, čo znamená, že pre množinu dát je možné použiť iba určité typy grafov. Vizualizácie musia byť zdieľané prostredníctvom servera Tableau Server a sú obmedzené na možnosti formátovania obsiahnuté v Tableau. Na prístup k softvéru Tableau Desktop je potrebné mať platenú licenciu, čo je z finančného hľadiska problém pre programátorov. Študentská verzia je rok bezplatná.

Bendremer et al. [21] pri používaní nástroja Tableau prešli na D3.js, s ich dôvodmi súhlasím, D3.js má väčšiu knižnicu vizualizácií a je flexibilnejší. Jedným z hlavných bodov v prospech nástroja D3.js je, že si môžeme vizualizácie vytvoriť tak, ako chceme. Vizualizačný nástroj D3.js, ako už bolo spomenuté, je otvorený zdroj tzv. open source, s ktorým je kedykoľvek možné zdieľať vizualizáciu prostredníctvom webových stránok. To dáva D3.js väčšiu výhodu oproti Tableau. D3.js pracuje s webovými štandardmi, takže na využitie nástroja D3.js nepotrebujeme žiadnu inú technológiu alebo doplnok okrem prehliadača. D3 pracuje s webovými štandardmi ako HTML, CSS a SVG, na prácu s D3.js nie je potrebný žiadny nový vzdelávací nástroj.

Pomocou nástroja Tableau a D3.js sme vytvorili stĺpcové grafy znázornené na obrázku 7 v Tableau a na obrázku 8 v D3.js. Napriek tomu, že grafy majú mierne odlišný vzhľad, tak ich hodnoty sú rovnaké. Napriek tomu, že vizualizáciu pomocou nástroja v Tableau vieme vytvoriť bez napísania jediného riadka kódu ako tvrdí Lekha et al. [7], pomocou nástroja D3.js sa napriek časovej náročnosti získavania informácií o danom nástroji a ďalšiemu písaniu kódu, tvorila vizualizácia lepšie. Pomocou nástroja D3.js sme vedeli vytvoriť vizualizáciu takú, ako sme potrebovali. Vedeli sme pridať rôzne animácie ako napríklad pohyb stĺpcov pri otvorení webového prehliadača.

## Záver

Grafická vizualizácia dát na internetových stránkach pomáha ich návštevníkom nielen k lepšiemu vnímaniu hodnotených javov, ale celkový dizajn stránky je veľmi dôležitým prvkom aj z hľadiska opakovanej návštevnosti a v súčasnosti patrí medzi významné marketingové nástroje vývoja internetu ako takého. Jednotlivé API od spoločnosti Google sú navrhované tak, aby pomohli tvorcom internetových stránok čo najjednoduchším spôsobom implementovať do stránok viaceré funkčne vysoko kvalitné prvky.

Vizualizácia informácií je viackrokový proces. Na začiatku potrebujeme vedieť, akým spôsobom budeme vizualizovať dostupné dáta. Výber správnej techniky závisí od viacerých faktorov, napríklad od typu dát, ktorými disponujeme, od cieľového publika, ktorému informácie prezentujeme, od spôsobu, akým chceme vysvetliť dáta a súvislosti medzi nimi. Po zvolení vhodnej vizualizačnej metódy môžeme pristúpiť k samotnej vizualizácii dát. Avšak, vytvorením výsledného grafu proces vizualizácie nekončí. Veľmi dôležitým krokom v celej postupnosti je interpretácia vytvoreného obrazu.

Tableau používame na vytváranie vysoko efektívnych vizualizácií bežných analýz údajov, ktoré nám pomáhajú vidieť a porozumieť našim údajom. Pomocou Tableau môžu ľudia, vedci a výskumní pracovníci rýchlo získať spoľahlivé informácie a dokonca údaje sami analyzovať. Pomocou Tableau sa dajú definovať diskrétne a spojité dáta a preskúmať, kedy sa dajú použiť ktorékoľvek z nich na vysvetlenie údajov, a tiež zistiť, ako môže Tableau používať rôzne typy geografických údajov, ako sa pripojiť k viacerým zdrojom údajov a ako vytvárať vlastné mapy.

Tableau sa používa na vizualizácie dát, ktoré môžeme použiť na spojenie s takmer ľubovoľnou databázou, vykreslenie našich dát a získanie prehľadov o ich správnom pochopení. Pomocou Tableau vieme vytvárať dashboardy, modely, interaktívne vizualizácie, ako aj dôkladne analyzujeme naše dáta. Tableau poskytuje príležitosti pre vás alebo kľúčové subjekty s rozhodovacou právomocou objavovať dátové vzory COVID-19, ako napríklad pandemické správanie COVID-19.

S Tableau sa ľahko pracovalo, veľmi dobre poslúžil pri vytváraní pútavých vizualizácií pri vybraných dátach s obsahom informácií ohľadom COVID-19. Tableau nevyžadoval žiadny zásah do databázy. Tableau má pre používateľa výborné prostredie, ktoré umožňuje aj netechnickým používateľom rýchlo a ľahko vytvárať napríklad dashboardy.

Nástroj na vizualizáciu D3.js je časovo náročný na programovanie. Poskytuje niekoľko spôsobov ľahkej vizualizácie údajov. Je nenáročný na systémové prostriedky. D3.js je zostavený pomocou JavaScriptu a vizualizácie je možné ľahko hostiť na webe v rôznych prehliadačoch. Používa otvorené webové technológie HTML, SVG a CSS. Pri práci s veľkými súborami dát je príliš pomalý a nie je moc vhodný na vytváranie máp. JSON a CSV formáty sú integrovateľné a konvertované na flexibilitu dát v D3.js.

## Zoznam použitej literatúry

[1] FRY, B. Computational Information Design (Disertation), MIT, Massachusetts, USA, 2004, 175 s. Dostupné na internete: <https://benfry.com/phd/>

[2] SOUKUP, T. DAVIDSON, I. Visual Data Mining: Techniques and Tools for Data Visualization and Mining [elektronický zdroj], Wiley, 2002, 416 s., ISBN: 978-0-471. Dostupné na internete: [https://books.google.sk/books?hl=sk&lr=&id=9HsVknVqbo4C&oi=fnd&pg=PR7&dq=visualization+data+tools&ots=anmX-dm4ko&sig=QxpFmBNWpQ17OkIeBQ7FCy-SXl4&redir\\_esc=y#v=onepage&q=visualization%20data%20tools&f=false](https://books.google.sk/books?hl=sk&lr=&id=9HsVknVqbo4C&oi=fnd&pg=PR7&dq=visualization+data+tools&ots=anmX-dm4ko&sig=QxpFmBNWpQ17OkIeBQ7FCy-SXl4&redir_esc=y#v=onepage&q=visualization%20data%20tools&f=false)

[3] SHNEIDERMAN, B. The eyes have it: A task by data type taxonomy for information visualizations, in: Visual Languages, 1996. Proceedings., IEEE Symposium on. IEEE, 336–343 s.

[4] Dostupné typy grafov v Office [online]. Dostupné na internete: <https://support.microsoft.com/sk-sk/office/dostupn%c3%a9-typy-grafov-v-office-a6187218-807e-4103-9e0a-27cdb19afb90?ocmsassetid=ha001233737&correlationid=81547bad-8852-4547-890d-c762de6395eb&ui=sk-sk&rs=sk-sk&ad=sk>

[5] Data visualization – Java Programming [online]. Dostupné na internete: <https://java-programming.mooc.fi/part-14/1-data-visualization>

[6] GANCHEV D., Application of Tableau Public Free Data Visualization Software in the Pesticide Science and Ecotoxicology Education, 2017. Dostupné na internete: [https://www.researchgate.net/publication/319454203\\_Application\\_of\\_Tableau\\_Public\\_Free\\_Data\\_Visualization\\_Software\\_in\\_the\\_Pesticide\\_Science\\_and\\_Ecotoxicology\\_Education/citations](https://www.researchgate.net/publication/319454203_Application_of_Tableau_Public_Free_Data_Visualization_Software_in_the_Pesticide_Science_and_Ecotoxicology_Education/citations)

[7] NAIR, L., SHETTY, S., SHETTY, S "Interactive visual analytics on Big Data: Tableau vs D3. js." *Journal of e-Learning and Knowledge Society* 12.4 (2016). Dostupné na internete: [http://je-lks.org/ojs/index.php/Je-LKS\\_EN/article/view/1128/611](http://je-lks.org/ojs/index.php/Je-LKS_EN/article/view/1128/611)

[8] Google: Introduction to Using Chart Tools, 2020. Dostupné na internete: <https://developers.google.com/chart/interactive/docs/index>

[9] MURRAY, S. Interactive Data Visualization for the Web: An Introduction to Designing with D3. Dostupné na internete: [https://books.google.sk/books?hl=sk&lr=&id=NGwvDwAAQBAJ&oi=fnd&pg=PR2&dq=Interactive+Data+Visualization+for+the+Web&ots=rvTCooftOp&sig=D-rkD1BEh3fgXAYG72S7KPikdVM&redir\\_esc=y#v=onepage&q=Interactive%20Data%20Visualization%20for%20the%20Web&f=false](https://books.google.sk/books?hl=sk&lr=&id=NGwvDwAAQBAJ&oi=fnd&pg=PR2&dq=Interactive+Data+Visualization+for+the+Web&ots=rvTCooftOp&sig=D-rkD1BEh3fgXAYG72S7KPikdVM&redir_esc=y#v=onepage&q=Interactive%20Data%20Visualization%20for%20the%20Web&f=false)

[10] COVID-19 Dataset, kaggle.com Databáza dostupná na internete: [https://www.kaggle.com/imdevskp/corona-virus-report?select=country\\_wise\\_latest.csv](https://www.kaggle.com/imdevskp/corona-virus-report?select=country_wise_latest.csv)

[11] Tableau Public, 2021. Dostupné na internete: <https://public.tableau.com/en-us/s/>

[12] MLÁDKOVÁ, L. 2005. Moderní přístupy k managementu. Tacitní znalost a jak ji řídit. 1. vydání. Praha : C. H. Beck, 2005. 80-7179-310-8.

[13] LANGENKAMP, J., 2010. Glossary. Information Management. Information Management and SourceMedia, Inc., 2010

[14] CUESTA, H., KUMAR, S. 2016. Practical Data Analysis-Second Edition. Birmingham: Packt Publishing , 2016. 338 s. ISBN 978-1-78528-971-2 .

[15] RAWGraphs About. Dostupné na internete: <https://rawgraphs.io/about>

[16] ROCHA, D., H. 2019. Learn Chart.js: Create interactive visualizations for the Web with Chart.js 2, 2019, ISBN 978-1-78934-248-2, Dostupné na internete: [https://books.google.sk/books?hl=sk&lr=&id=NHCLDwAAQBAJ&oi=fnd&pg=PP1&dq=Chart.js+description+&ots=OhZubIVedZ&sig=-xg5RMrFaw8f-6aePwyqEX2ux\\_E&redir\\_esc=y#v=onepage&q=Chart.js%20description&f=false](https://books.google.sk/books?hl=sk&lr=&id=NHCLDwAAQBAJ&oi=fnd&pg=PP1&dq=Chart.js+description+&ots=OhZubIVedZ&sig=-xg5RMrFaw8f-6aePwyqEX2ux_E&redir_esc=y#v=onepage&q=Chart.js%20description&f=false)

[17] Data Visualization tools – A case study. Dostupné na internete: [https://d1wqtxts1xzle7.cloudfront.net/51130753/88\\_Paper\\_310816159\\_IJCSIS\\_Camera\\_Ready\\_pp\\_834-838.pdf?1483187232=&response-content-disposition=inline%3B+filename%3DData\\_Visualisation\\_Tools\\_A\\_Case\\_Study.pdf&Expires=1625931825&Signature=F7UyvWiID-bAZgQxuWynRtL2Bl4jMvSazqm6R79b9h6LeGxWQqubCTxOWFvzXspdp02hkvGrGcbtYLhDsoCKG09vE~z4PhF3ch~MduF5szA82wThE20OhukaOJmWszE2WA94DnWRZU4AxaZKIVDh3kjFBxGFnyN0o3SPsPKFC-WA1hcECLLMRfyZ59kQyekE2M3gGjDVof5Q~mJoIbtAIui9FDPr95szWiB0hrqfbD~6ptuS](https://d1wqtxts1xzle7.cloudfront.net/51130753/88_Paper_310816159_IJCSIS_Camera_Ready_pp_834-838.pdf?1483187232=&response-content-disposition=inline%3B+filename%3DData_Visualisation_Tools_A_Case_Study.pdf&Expires=1625931825&Signature=F7UyvWiID-bAZgQxuWynRtL2Bl4jMvSazqm6R79b9h6LeGxWQqubCTxOWFvzXspdp02hkvGrGcbtYLhDsoCKG09vE~z4PhF3ch~MduF5szA82wThE20OhukaOJmWszE2WA94DnWRZU4AxaZKIVDh3kjFBxGFnyN0o3SPsPKFC-WA1hcECLLMRfyZ59kQyekE2M3gGjDVof5Q~mJoIbtAIui9FDPr95szWiB0hrqfbD~6ptuS)

mdAoijaU~SugQ2U8Dc7Zz8JFC7c4sekFuv0agisDz7cHrGPthfzvdfPvKI0Da4HnKjyJAGeR  
bAibTqthrNucQBOMl6E1hIUkl83xrw\_\_&Key-Pair-Id=APKAJLOHF5GGSLRBV4ZA

[18] OZDEMIR S., 2016, Principles of Data Science, 2016, ISBN 9781785888922

[19] SADIKU, M. N. O., SHADARE, A. E., MUSA, S. M., AKUJOBI, C. M., Data visualization, Dostupné na interete: [https://www.researchgate.net/profile/Adebawale-Shadare/publication/311597028\\_DATA\\_VISUALIZATION/links/5851945608aef7d0309f20a7/DATA-VISUALIZATION.pdf](https://www.researchgate.net/profile/Adebawale-Shadare/publication/311597028_DATA_VISUALIZATION/links/5851945608aef7d0309f20a7/DATA-VISUALIZATION.pdf)

[20] Data Driven Documents. Dostupné na internete: <https://github.com/d3/d3/wiki>

[21] BENDREMER, E., FABRY, K., GUPTA, A., 2021, Mapping WPI Alumni Career Trajectories Through Data Visualization. Worcester Polytechnic Institute. 2021. 24391 E-project-051321-141105, Dostupné na internete: [https://digital.wpi.edu/concern/student\\_works/mc87pt00k](https://digital.wpi.edu/concern/student_works/mc87pt00k)

[22] HUDAK, H. C. *Charts and Graphs*. Weigl Publishers, 2019. Dostupné na internete: [https://books.google.sk/books?hl=sk&lr=&id=68K2DwAAQBAJ&oi=fnd&pg=PA4&dq=types+of+graphs+and+charts&ots=rvKf9rSy5M&sig=Eg7GAun6RupekpIN4EvNVLtx5yU&redir\\_esc=y#v=onepage&q=types%20of%20graphs%20and%20charts&f=false](https://books.google.sk/books?hl=sk&lr=&id=68K2DwAAQBAJ&oi=fnd&pg=PA4&dq=types+of+graphs+and+charts&ots=rvKf9rSy5M&sig=Eg7GAun6RupekpIN4EvNVLtx5yU&redir_esc=y#v=onepage&q=types%20of%20graphs%20and%20charts&f=false)