Open Data Lab

Workshop #2:

Študijný materiál

*Toto podujatie bolo podporené z dotačnej schémy Ministerstva školstva, vedy, výskumu a športu SR Programy pre mládež na roky 2014 – 2021, ktorú administruje IUVENTA – Slovenský inštitút mládeže.*



Autor: Lukáš Jankovič  
[lukas.jankovic@alvaria.sk](mailto:lukas.jankovic@alvaria.sk)

Obsah

[Vizualizácia dát v Tableau Public 3](#_Toc89433065)

[Používateľské prostredie 3](#_Toc89433066)

[Načítanie datasetu 4](#_Toc89433067)

[Práca so stĺpcami a mierkami 5](#_Toc89433068)

[Vlastné stĺpce 5](#_Toc89433069)

[Vytvorenie grafu 6](#_Toc89433070)

[Vytvorenie mapy 9](#_Toc89433071)

[Práca s georeferenčnými (GIS) údajmi 10](#_Toc89433072)

[Formátovanie 11](#_Toc89433073)

[Filtre 13](#_Toc89433074)

[Spájanie datasetov 14](#_Toc89433075)

[Publikovanie 16](#_Toc89433076)

**Poznámka:**

Ak je v texte spomenuté stlačenie tlačidla myši, myslí sa tým ľavé tlačidlo. V prípade ak ide o pravé tlačidlo, je to v texte uvedené.

**Pojmy**

Prvok – graf, tabuľka, mapa

Vizualizácia – celý súbor (môže obsahovať viac prvkov a viac strán)

# Vizualizácia dát v Tableau Public

**Tableau** je nástroj na **spracovanie, editovanie a vizualizáciu dát**.

Nástroj je vhodný na spracovanie otvorených dát, rovnako ako iných druhov dát a datasetov.

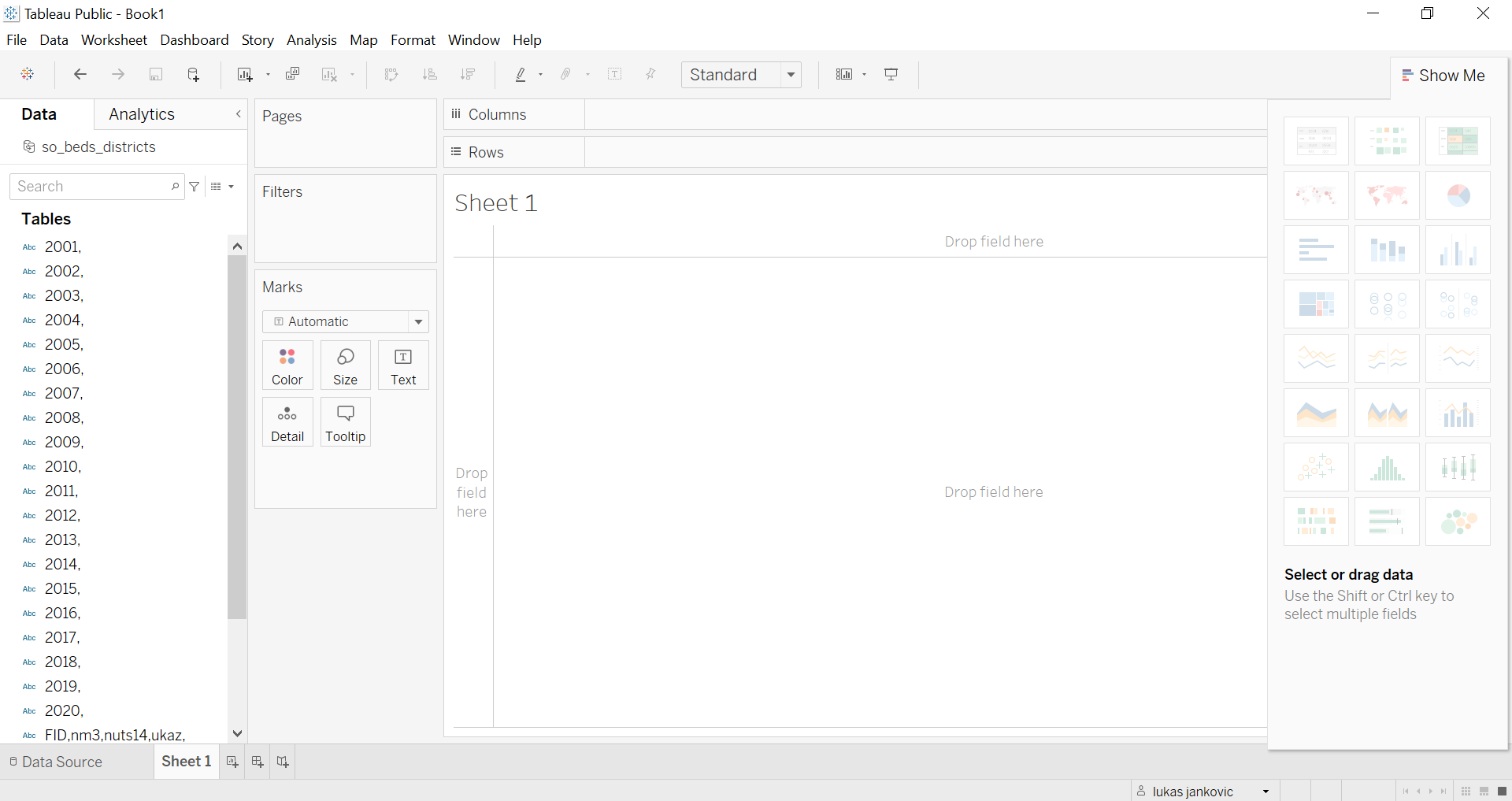
Tableau obsahuje viac produktov, **Tableau Public** je dostupný bezplatne pre MS Windows a Mac OS.

* Link na stiahnutie: <https://public.tableau.com/en-us/s/>

Podporuje viacero jazykov, napr. angličtinu a nemčinu (nie však slovenčinu a češtinu). Pre plnohodnotné použitie je **potrebné sa zaregistrovať** a vytvoriť si profil na platforme Tableau.

* Link na registráciu: <https://public.tableau.com/en-us/s/>

# Používateľské prostredie



Používateľské rozhranie používa klasické menu (File, Data, ...) okrem toho tieto panely:

* panel nástrojov,
* panel **Data**,
* panel **Pages**,
* panel **Filters**,
* panel **Marks**,
* polia na zadanie stĺpcov datasetu – **columns** a **rows** a
* panel **Show Me**, ktorý obsahuje výber prvkov (graf, mapa, tabuľka, atď.)

# Načítanie datasetu

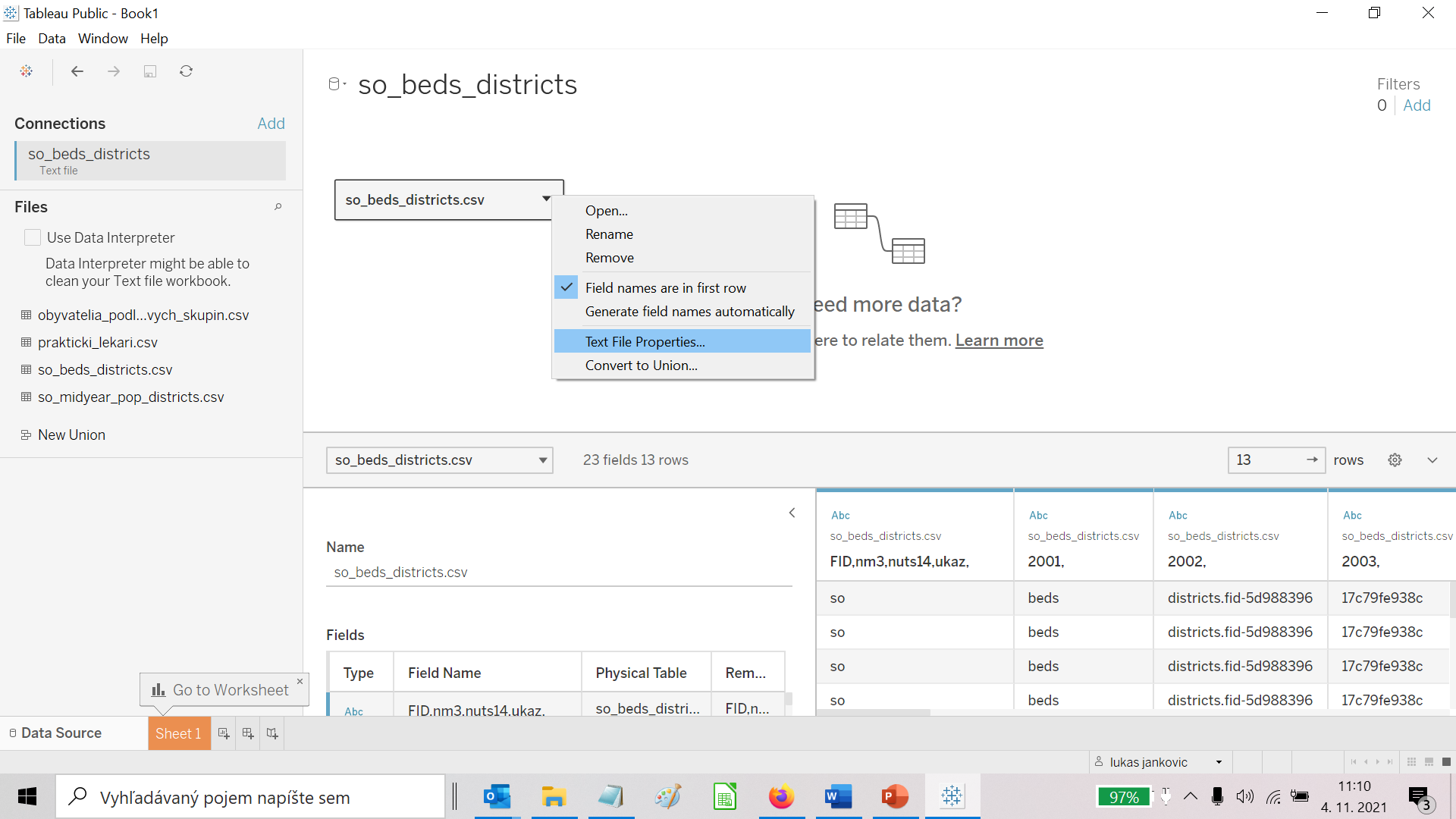
Základné formáty, ktoré vieme spracovať v Tableau sú:

* CSV,
* XML,
* XLX/XLSX,
* súbor geografických dát,
* Google Drive.

**Načítanie (import) datasetu** je možné cez voľbu **Connect** na úvodnej strane alebo **Add** v paneli **Data Source**.

Pri importe datasetu vo formáte CSV (v Tableau označené ako **Text file**) je dôležité nastaviť **správne kódovanie**. Mnoho súborov používa kódovanie **Unicode UTF-8**. Takisto je nutné správne nastaviť oddeľovač, pri CSV súboroch sa štandardne používa čiarka, niekedy sa vyskytuje aj bodkočiarka ;.

Na nastavenie sa používa voľba **Text File Properties**, kde sa dajú vykonať nastavenia kódovania a oddeľovača.



Načítanie nepotvrdzujeme žiadnym tlačidlom, jednoducho sa môžeme prepnúť do prvého hárku voľbou **Sheet 1** (dole vľavo od panelu Data Source).

Ak potrebujeme použiť viac dátových zdrojov, vytvoríme nový pomocou kliknutia na tlačidlo pri názve dátového zdroja (hore).

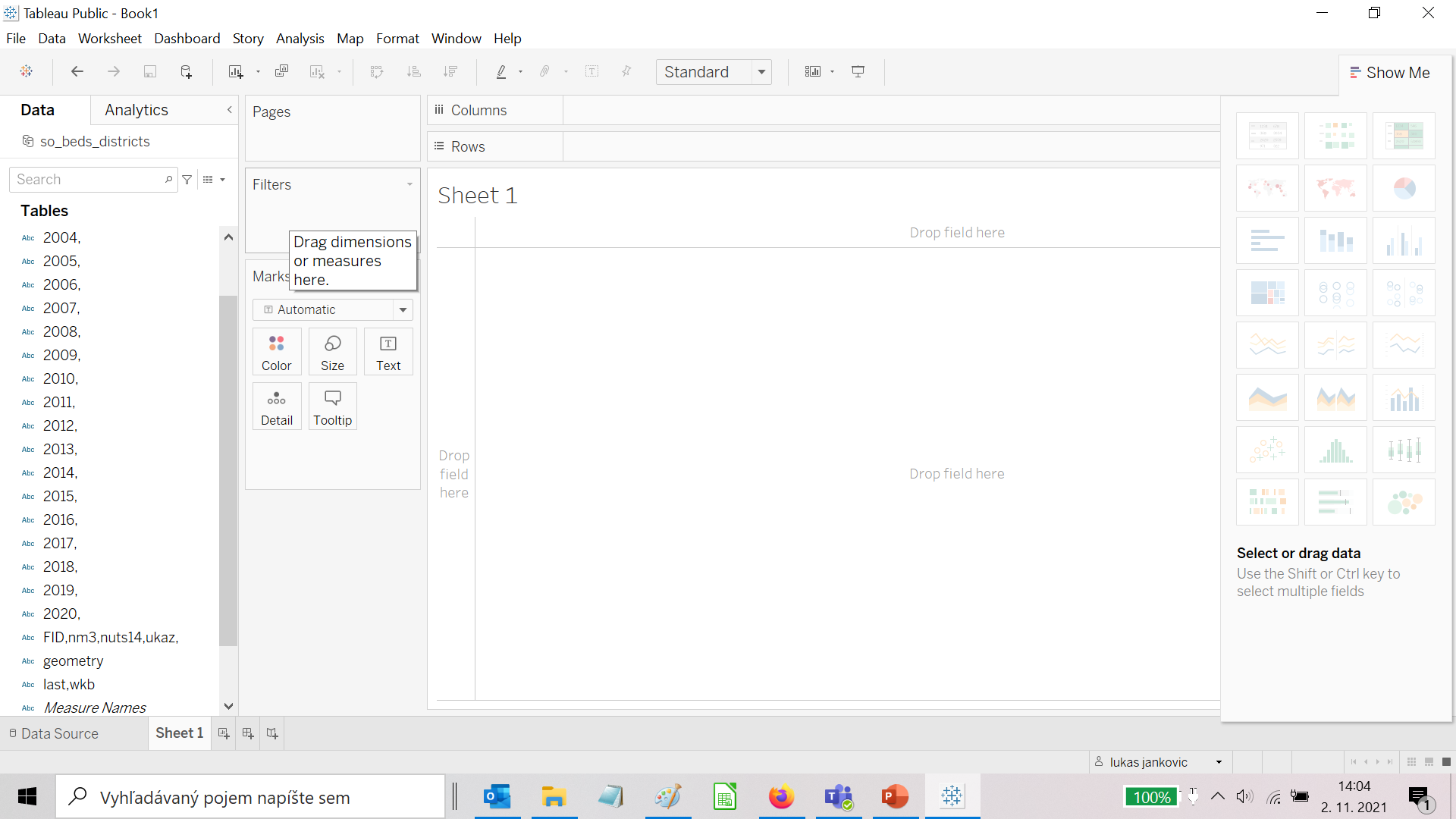
Obrázok, na ktorom je text

Automaticky generovaný popis

# Práca so stĺpcami a mierkami

**Náhľad údajov** vidíme vľavo na paneli Data. Zobrazené sú názvy stĺpcov datasetu.

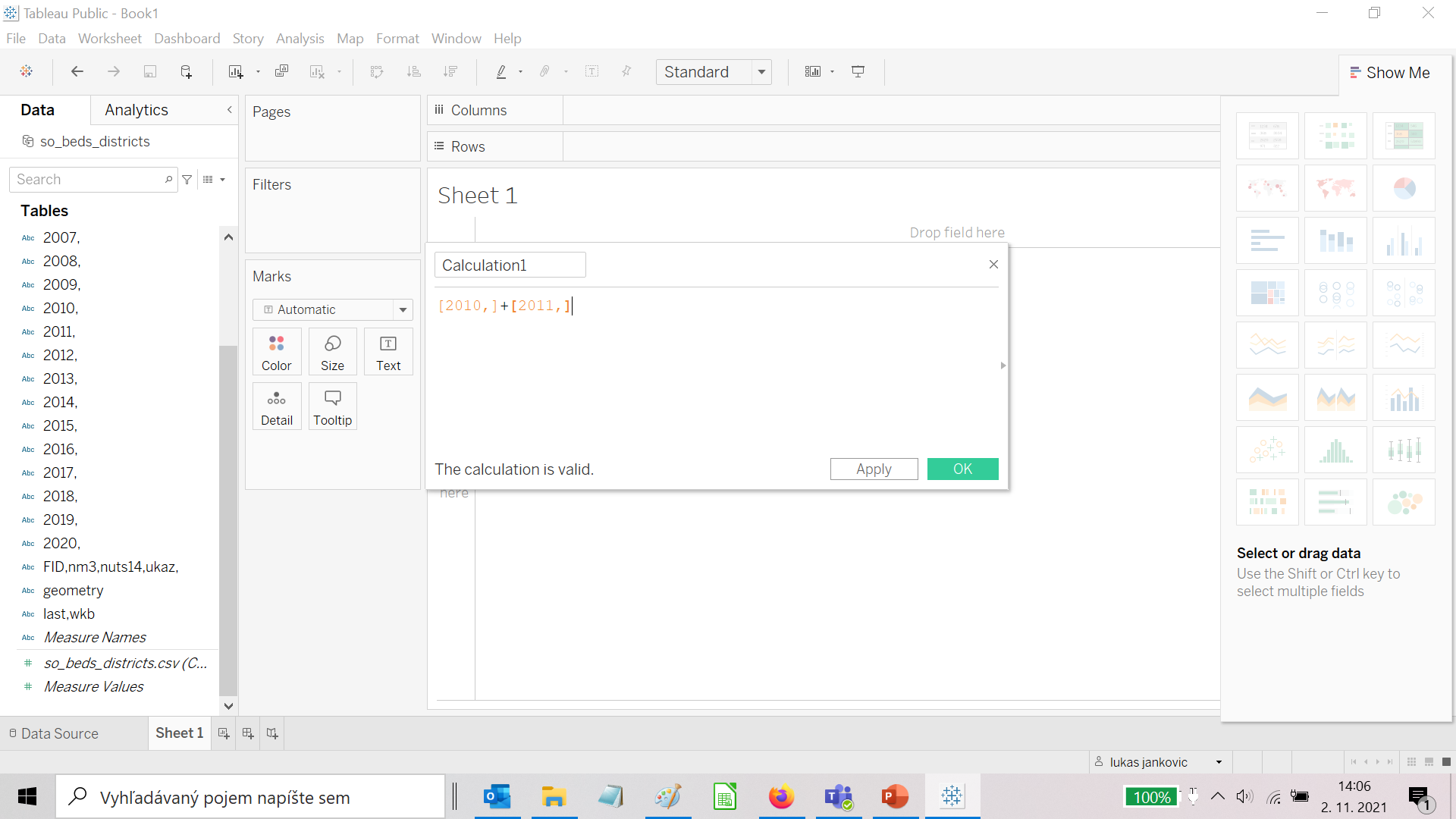
**Kliknutím pravým tlačidlom myši na názov stĺpca môžeme stĺpec premenovať alebo skryť**.



Taktiež pravým tlačidlom myši a voľbou **Change Data Type** môžeme nastaviť **typ údajov stĺpca (číslo, text, dátum,...)**. Voľbou **Geographic Role** môžeme nastaviť, ak stĺpec obsahuje GPS súradnice alebo geografický názov.

# Vlastné stĺpce

Vlastné stĺpce sú užitočné v prípade, ak potrebujeme vytvoriť pomocné stĺpce alebo potrebujeme údaje v stĺpcoch upraviť. Vlastné stĺpce môžeme tvoriť v paneli **Data,** kliknutím na malú šípku vpravo a zvolením **Create Calculated Field**. Otvorí riadok, do ktorého **môžeme vpísať vzorec** pre nový stĺpec a takisto aj upraviť jeho názov.



Pri vytváraní nového stĺpca môžeme pracovať s existujúcimi stĺpcami ale **môžeme používať aj tradičné matematické operátory ako sú zátvorky alebo + - / \*.** Ak chceme do výpočtu **zahrnúť existujúci stĺpec**, **potiahneme stĺpec z panelu Data** do vzorca alebo ho napíšeme v hranatých zátvorkách.

Ak sa pri názve stĺpcu zobrazuje symbol „=“, ide o číselný údaj, ak trojica písmen „abc“, ide o textový stĺpec, ak sa zobrazí symbol „-=“, ide o vlastný stĺpec.

*Príklady stĺpcov:*

Mojstlpec=so\_beds\_districts[\_2020]+1000

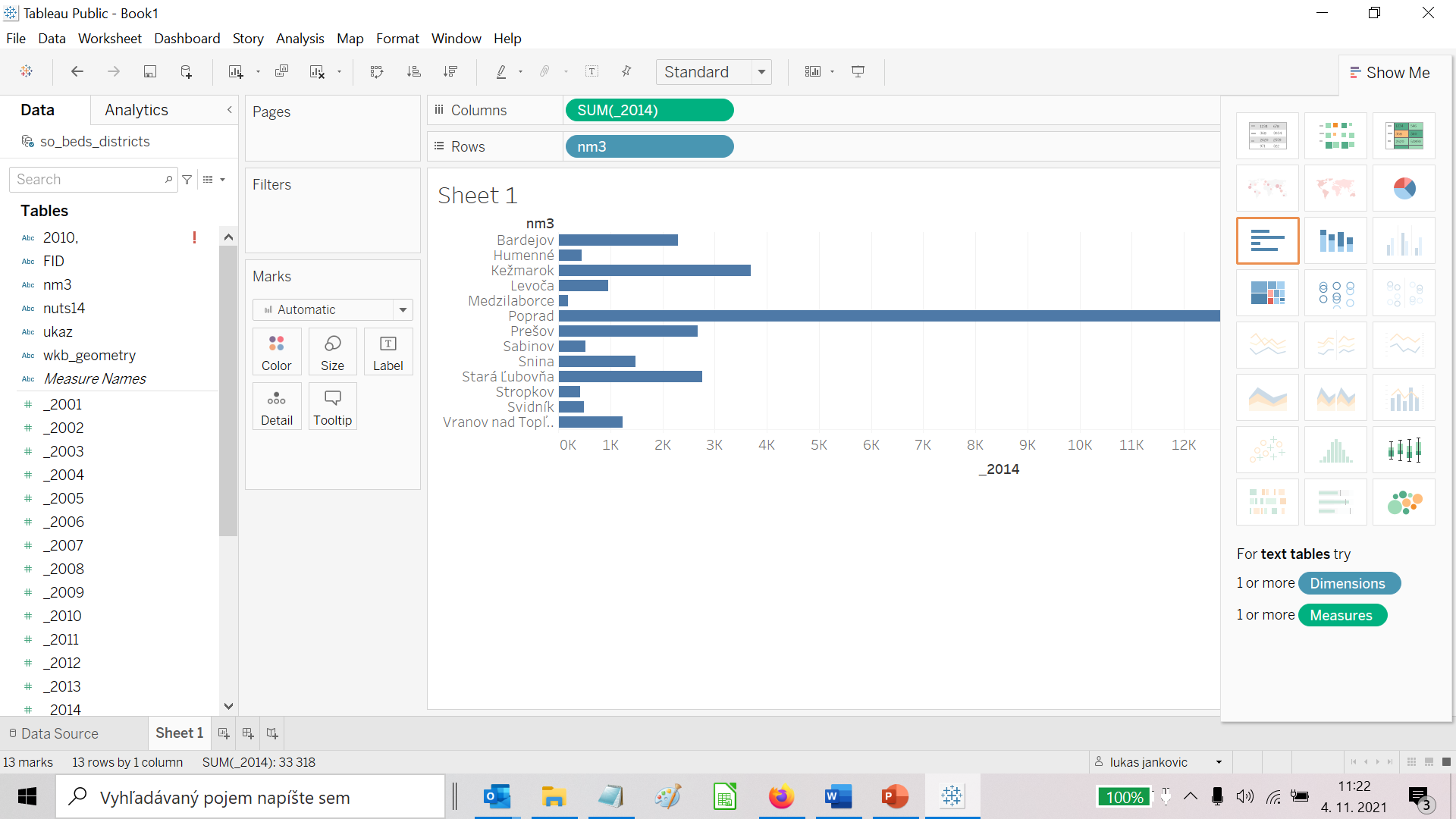
Mojstlpec=so\_beds\_districts[\_2020]+so\_beds\_districts[\_2019]

Mojstlpec= so\_beds\_districts[\_2020]/100

Mojstlpec= (so\_beds\_districts[\_2020]-20)/100

# Vytvorenie grafu

Rôzne typy grafov je možné vytvoriť cez **tlačidlá v paneli Show Me** (vpravo hore).



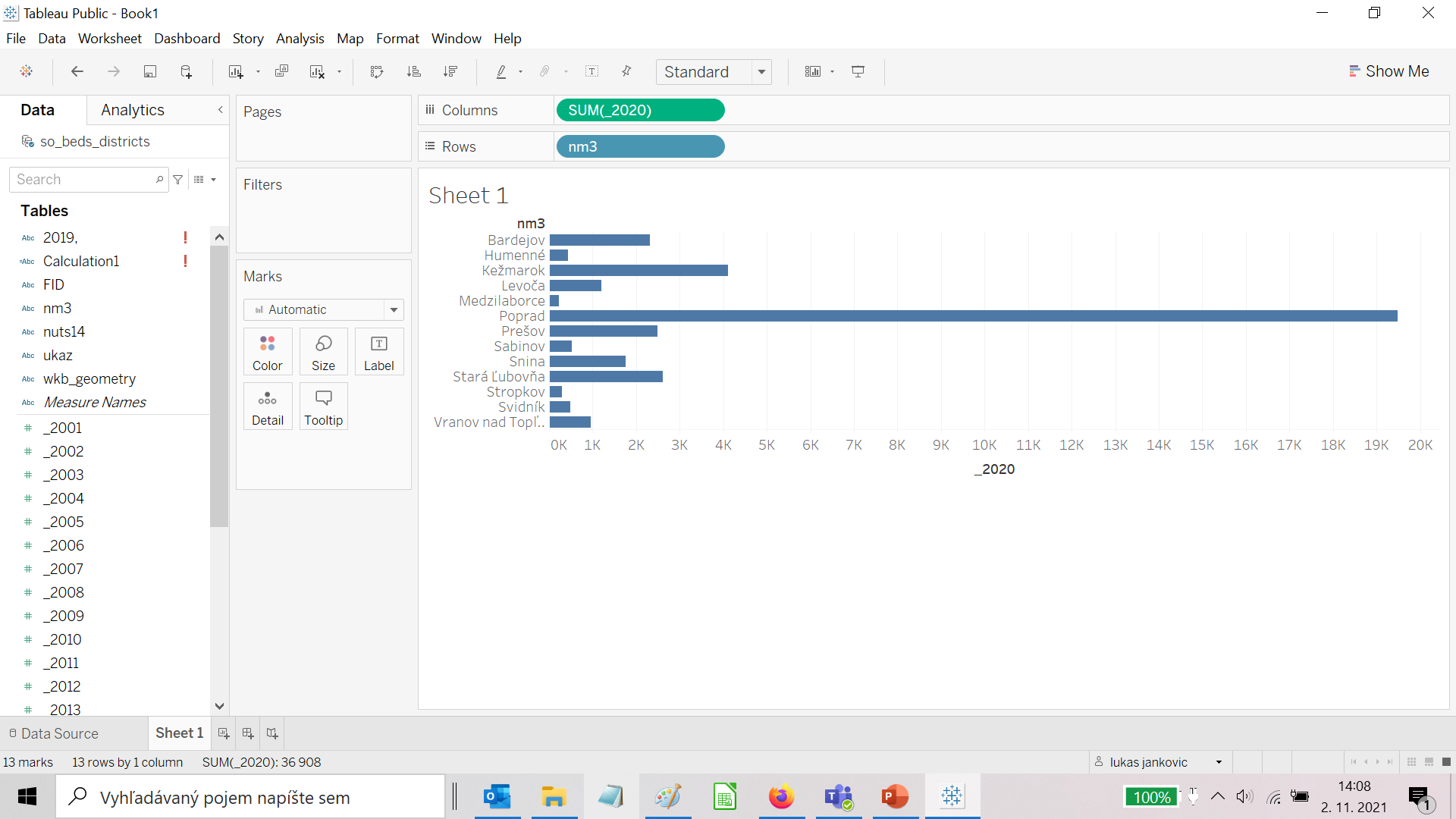
Najskôr však musíme názvy stĺpcov potiahnuť do polí **Columns** a**Rows**.

Pole columns zobrazuje stĺpce a pole rows riadky.

Ak napríklad chceme zobraziť stĺpcový graf, potiahneme do poľa Columns dimenziu (napríklad názvy okresov) a do poľa Rows hodnoty (napríklad počet izieb v okresoch).

Údaje môžeme vymeniť kliknutím na ikonu **Swap Rows and Columns **.

Do **Rows** a **Columns** môžeme potiahnuť aj **viac stĺpcov**.



Ak klikneme pravým tlačidlom na názov stĺpca a zvolíme z **Measure**, **môžeme zmeniť výpočet** na Priemer, Medián, Minimum alebo Maximum alebo iný výpočet.

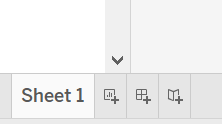
Taktiež môžeme v paneli **Marks** meniť aj farbu (**Color**), veľkosť (**Size**), popis (**Label**), detaily (**Detail**) alebo bublinu (**Tooltip**) pre graf tým, že potiahneme stĺpec nad jedno z tlačidiel a pustíme. Ak klikneme na tlačidlo, môžeme ďalej nastaviť spôsob zobrazenia.

Ak si chceme **pridať viacero hárkov, dashboardov alebo príbehov** do vizualizácie, môžeme tak urobiť dole, pomocou **tlačidla plus**. **Môžeme ich premenovať alebo vymazať ak klikneme na ich názov pravým tlačidlom myši**. Takisto ich môžeme aj **skryť** (nezobrazia sa používateľovi).

**Hárok** – jeden prvok (graf, mapa, tabuľka)

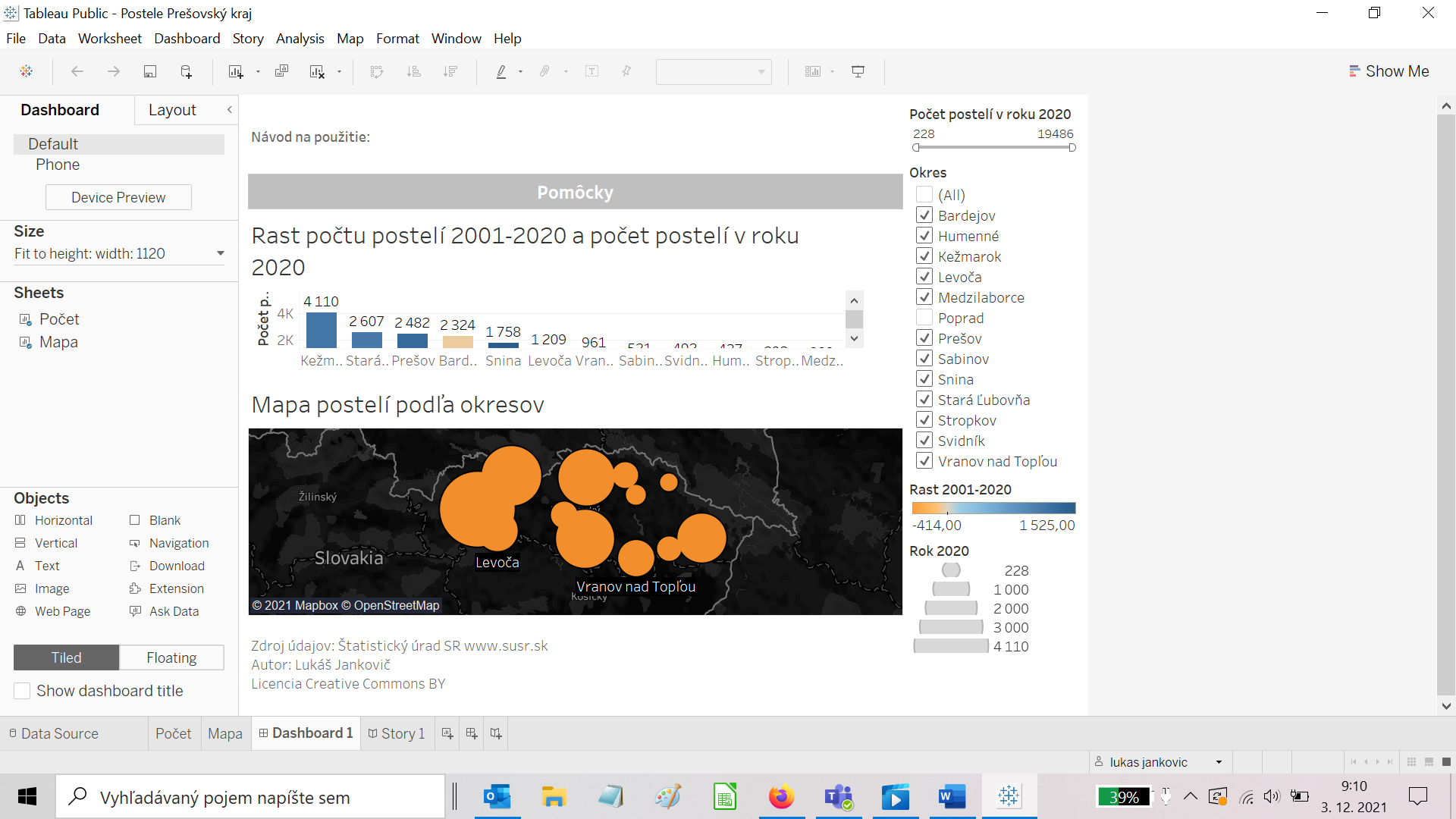
**Dashboard** – viacero hárkov

**Story** – viacero dashboardov

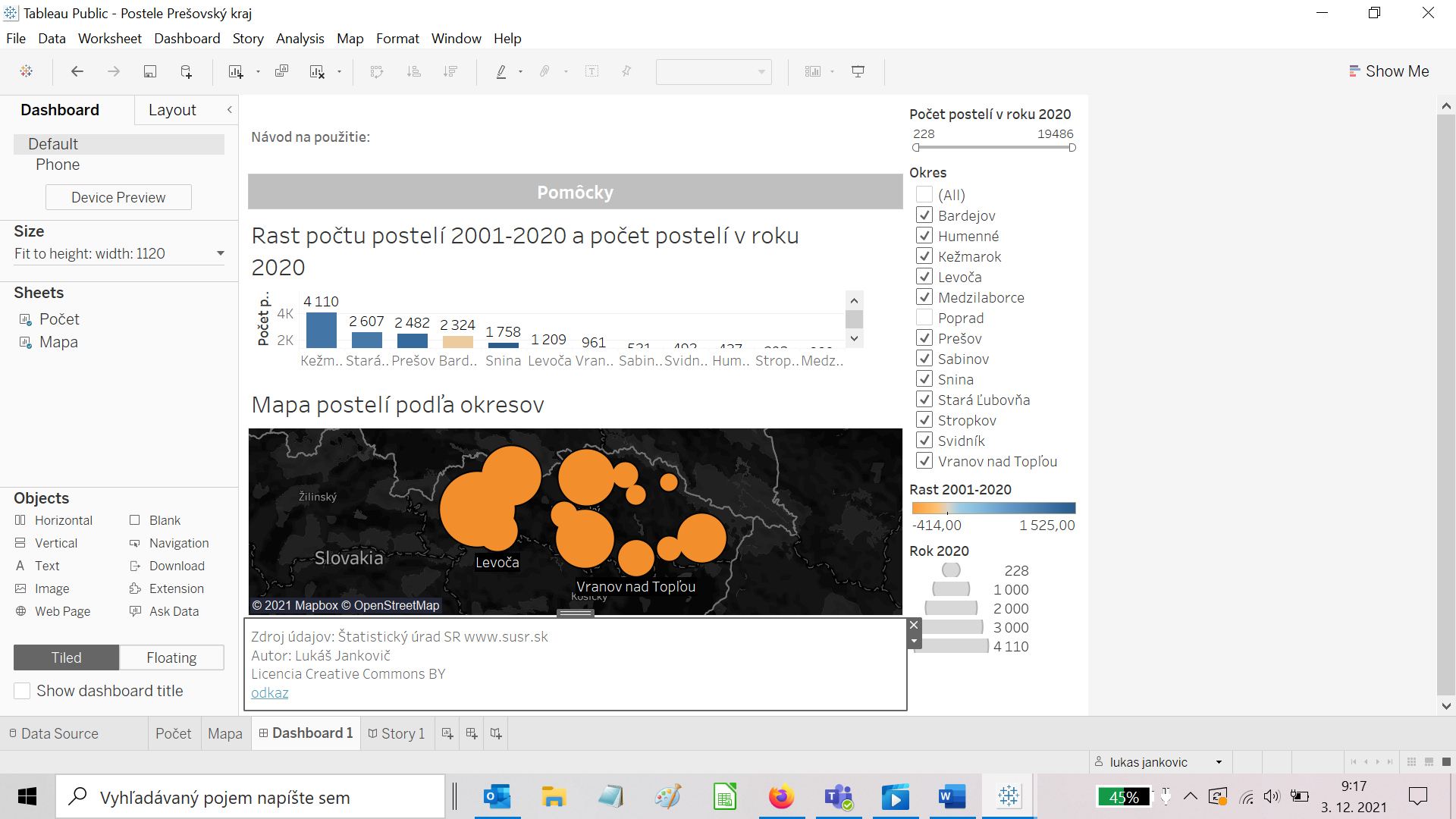


**Pridanie textového poľa do Dashboardu**

Textové pole môžeme natiahnuť do Dasboardu pomocou potiahnutia myšou **Text.** Do poľa môžeme písať a môžeme meniť jeho formátovanie (farba, veľkosť, štýl a zarovnanie písma).



Odkaz môžeme do poľa pridať skopírovaním napríklad z Internetu alebo wordového dokumentu. Označíme celý [odkaz](http://www.alvaria.sk/) (napr. vo Worde) a skopírujeme do poľa.



# Vytvorenie mapy

Mapu vytvoríme buď pomocou GPS súradníc alebo názvov – odporúčam používať názvy miest alebo krajín.

V prípade ak chceme použiť názvy:

1. Nastavíme stĺpcu s názvami geografickú rolu (**Geograhpical role**)
2. Potiahneme stĺpec do poľa **Column**
3. Nastavíme typ prvku na **symbol** **maps** **(bod na mape)** alebo **maps (plochy)** vpravo v paneli **Show me**.

V prípade ak chceme použiť GPS súradnice musíme mať súbor s GPS súradnicami. Tie si vieme zadarmo geokódovať z adries alebo názvov obcí napríklad cez službu <https://geocode.localfocus.nl/>

Postup:

1. Nastavíme stĺpcom geografickú rolu **Latitude** a **Longitude**
2. Potiahneme stĺpec **Latitude** do poľa **Columns** a stĺpec **Longitude** do poľa **Rows**
3. Nastavíme stĺpce ako dimenzie - **Dimension** (pravým tlačidlom myši)
4. Nastavíme typ prvku na **symbol** **maps** (bod na mape).

Obrázok, na ktorom je mapa

Automaticky generovaný popis

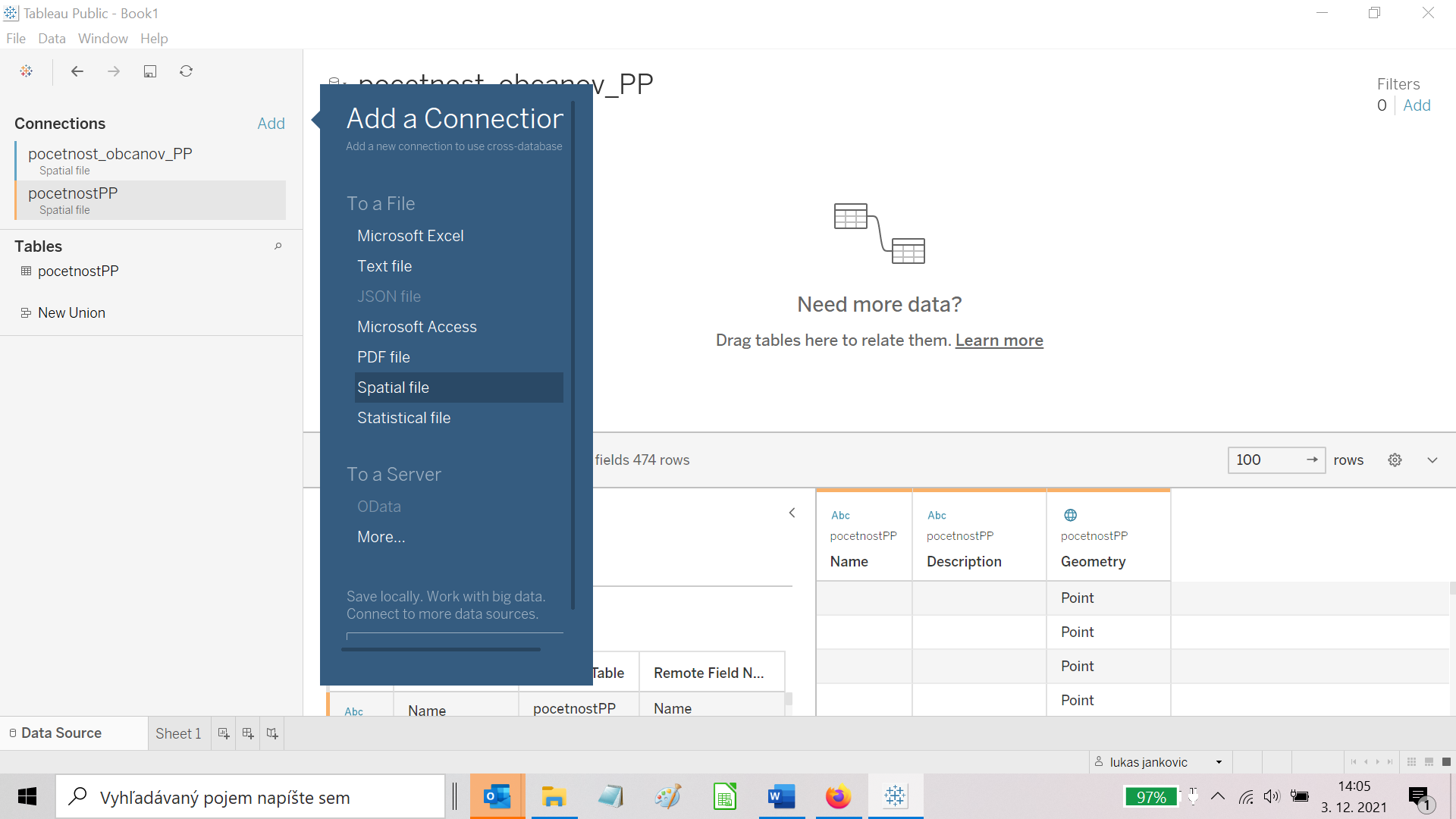
Ak chceme formátovať mapu, môžeme pomocou položky hlavného menu **Map** nastaviť štýl mapy, zobrazenie mapy a ďalšie parametre.

# Práca s georeferenčnými (GIS) údajmi

V Tableau môžeme pracovať aj s **georeferenčnými údajmi** (formát SHP, KML, geoJSON).

Tableau podporuje body rovnako aj krivky a plochy.

Súbor načítame pomocou voľby **Spatial file**.



V súbore nájdeme automaticky vygenerovaný stĺpec **Geometry**. Tableau vytvorí body na mape, ktoré sa správajú zatiaľ ako jeden bod. Ak chceme body rozlíšiť, natiahneme do poľa Detail ešte iný jedinečný stĺpec (napr. názov)

Obrázok, na ktorom je text

Automaticky generovaný popis

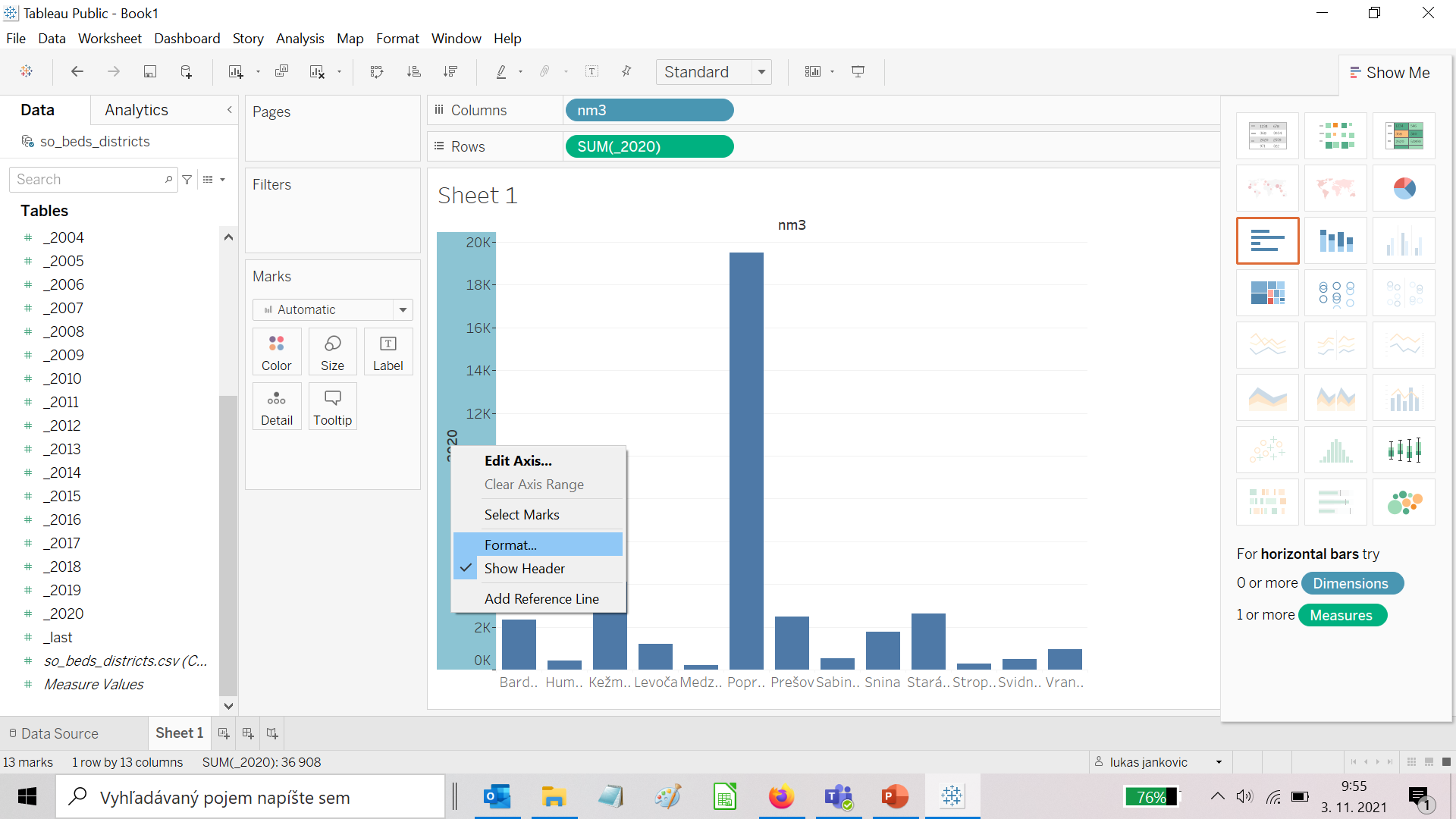
Tento stĺpec potiahnutím myšou prenesieme do poľa **Marks (detail)**.

# Formátovanie

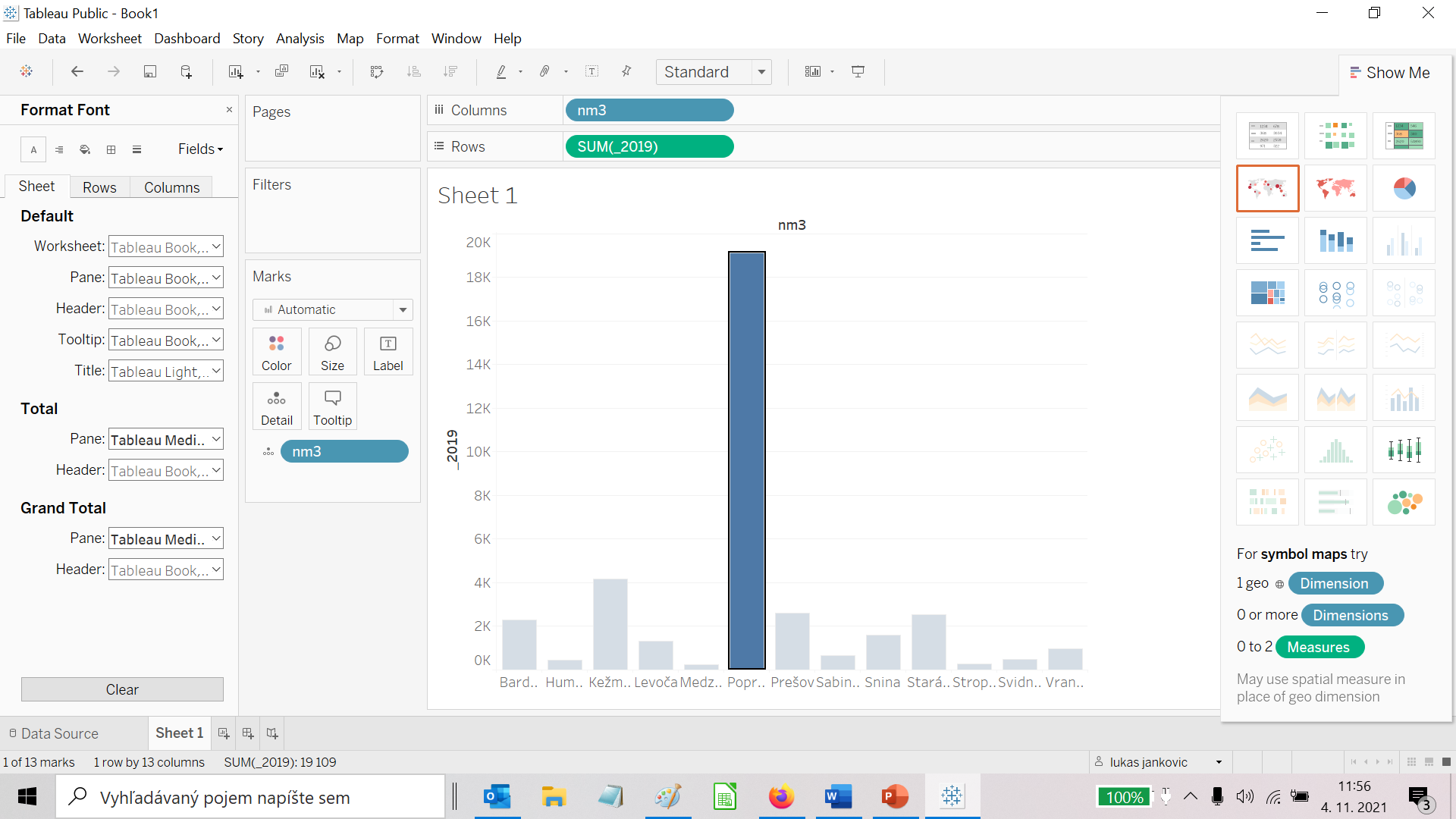
Nastavenie formátovania umožňuje meniť vzhľad prvkov – grafov, máp, atď. (farby, veľkosť a typ písma, pozadie atď.)

Možnosť formátovania máme pri každom prvku (mapa, graf, tabuľka, iné).

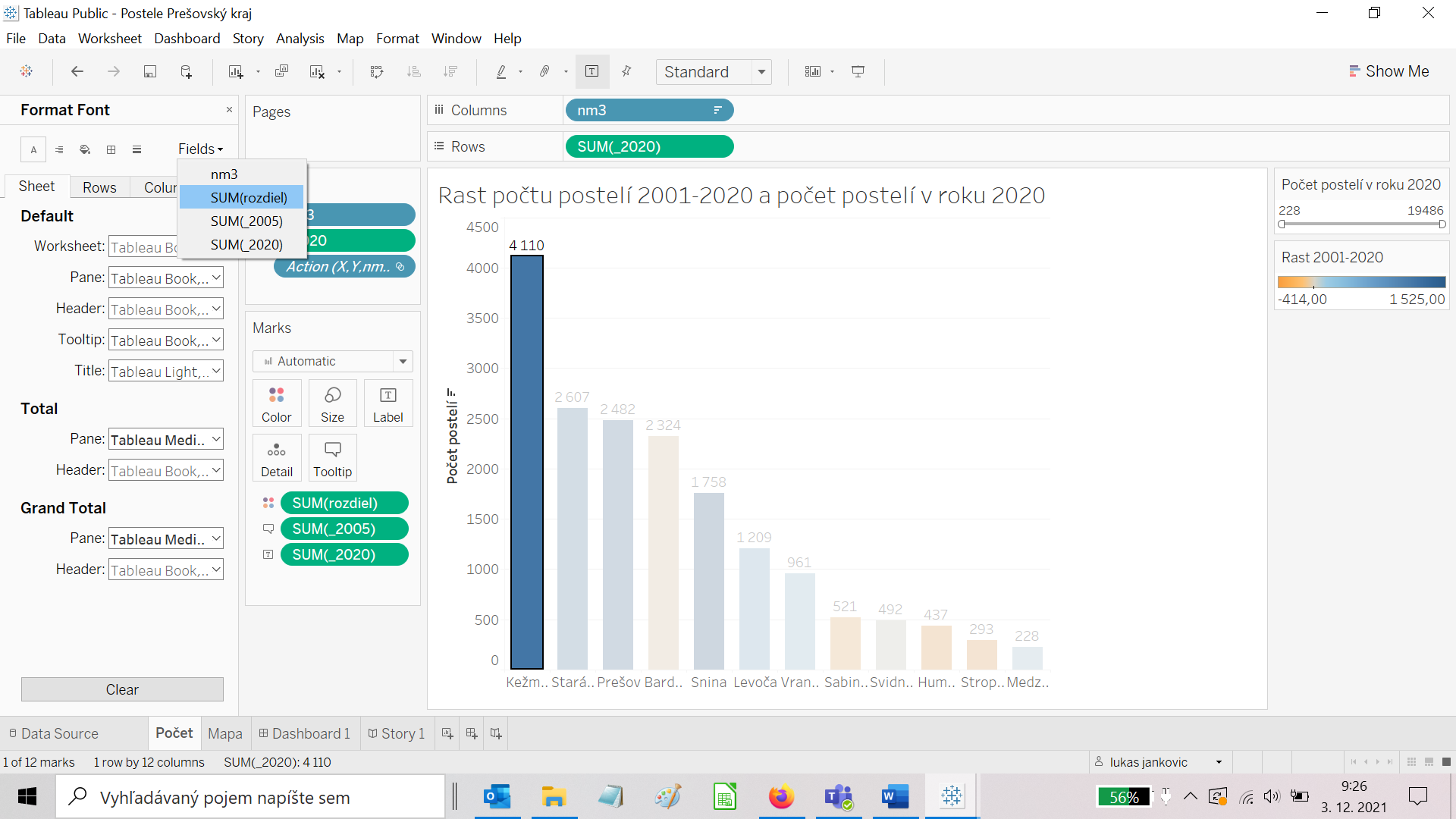
Možnosti formátovania môžeme použiť tak, že klikneme pravým tlačidlom myši na prvok, ktorý chceme formátovať a vyberieme **Format**.

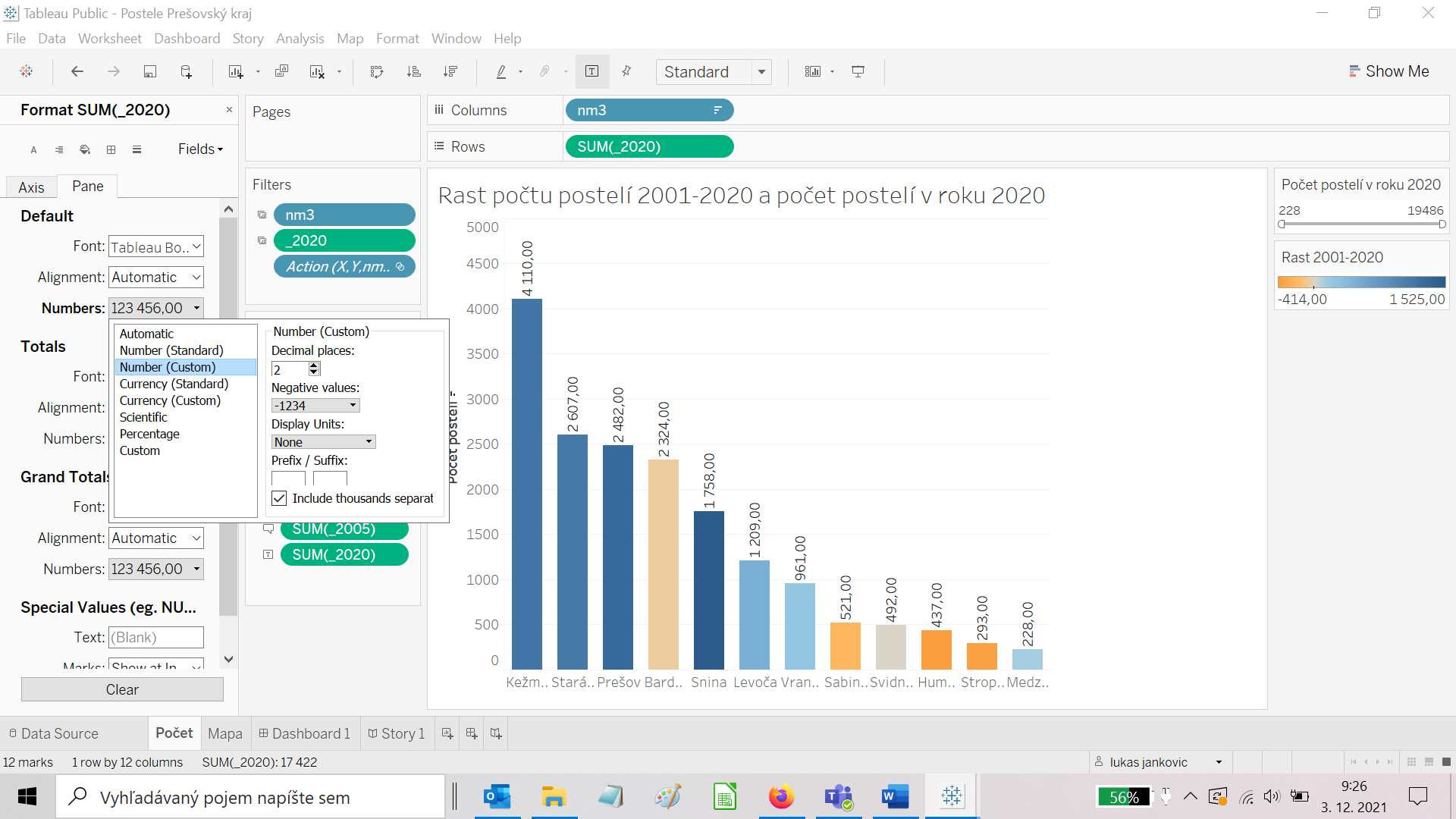


Následne sa nám zobrazí **Format Font** panel, pomocou ktorého môžeme nastaviť farby, písma ale aj počet desatinných miest pri číslach a ďalšie nastavenia.



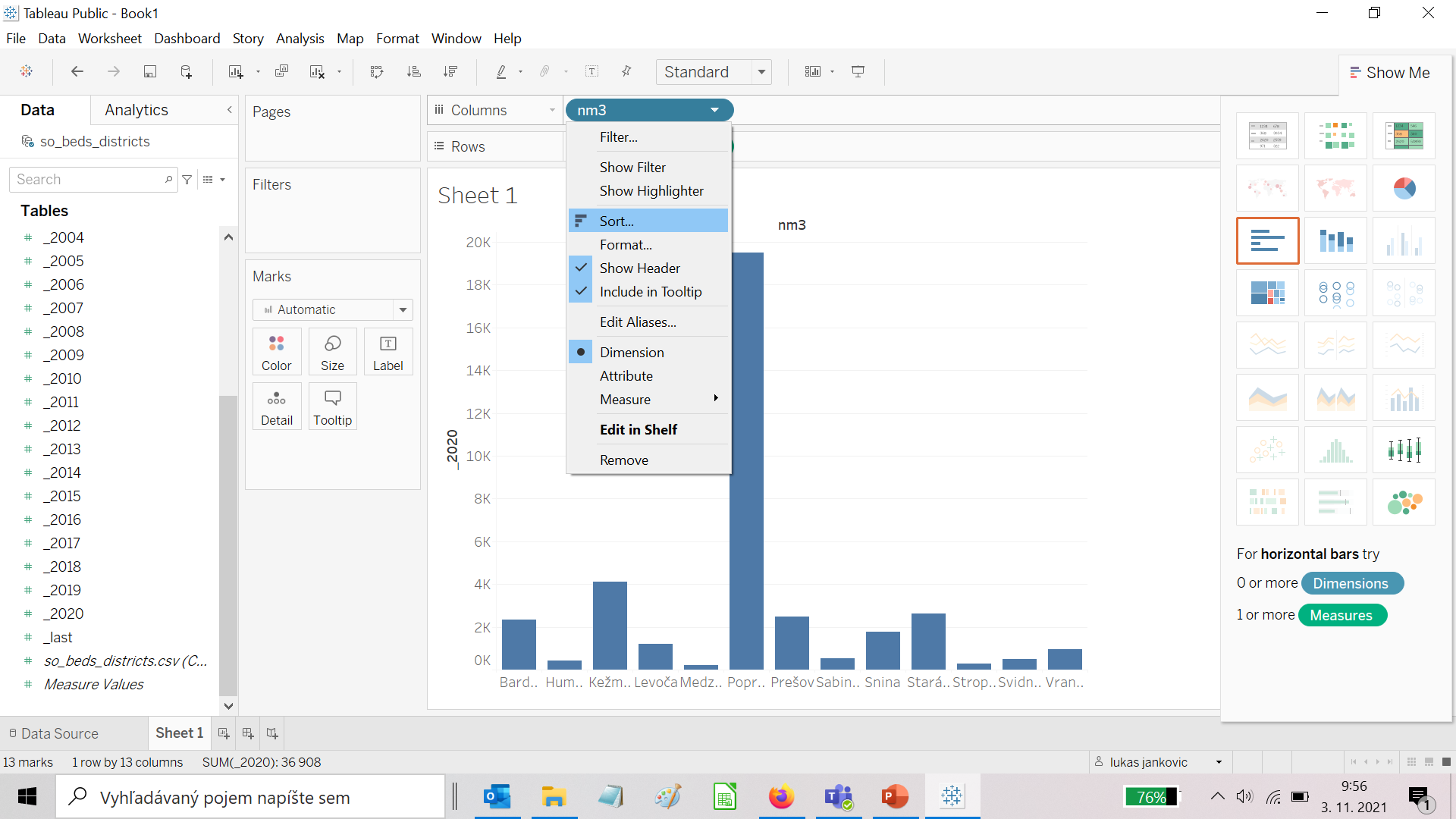
Ak potrebujeme zmeniť formátovanie čísla, vyberieme z ľavého panelu **Format font** požadované pole (field) a pomocou voľby **Numbers** vyberieme požadované formátovanie – počet desatinných miest, oddeľovače, zobrazenie tisícov, miliónov prefix alebo sufix (napr. mena). Takisto môžeme zvoliť aj zobrazenie percent.





Ak chceme **zmeniť text názvu osi**, klikneme pravým tlačidlom myši na názov osi a vyberieme **Edit Axis**. V tomto nastavení môžeme takisto zmeniť mierku osi.

Ak chceme **zmeniť zoradenie údajov**, klikneme pravým tlačidlom myši na názov stĺpca v poli **Columns** alebo **Rows** a klikneme na **Sort**.



# Filtre

**Filtrovanie umožňuje zobraziť iba vybranú časť údajov**, ktoré sa majú zobraziť vo vizualizácii. To môže byť užitočné v prípade ak potrebujeme z vizualizácie **niektoré dáta vylúčiť** alebo aj v prípade, ak chceme používateľovi našej vizualizácie ponechať **možnosť vybrať si z položiek alebo si vyhľadať položky**.

Filtrovať údaje môžeme tak, že potiahneme stĺpec do panela **Filters.** Zobrazí sa nám okno, v ktorom môžeme hodnoty, ktoré chceme odfiltrovať a ktoré majú v dátach zostať.

Obrázok, na ktorom je text, vnútri, snímka obrazovky

Automaticky generovaný popis

Ak chceme aby mal **možnosť filtrovania k dispozícii aj používateľ vizualizácie**, klikneme pravým tlačidlom myši na názov stĺpca a vyberieme **Show Filter**. Filter sa nám zobrazí vpravo, môžeme ho formátovať pomocou malej šípky vpravo.

# Spájanie datasetov

**Spájanie datasetov je veľmi užitočné, pretože pridávame do dát pridanú hodnotu.** Môžeme vizualizovať aj jeden dataset, pridaním iných údajov však **vytvárame unikátny produkt.** Veľmi často sa spájanie datasetov používa ak potrebujeme **spojiť podobné dáta z rôznych zdrojov** (napríklad dotácie z rôznych miest).

V Tableau poznáme **2 základné typy** spájania údajov – **zlúčenie (Join) a pripojenie (Union)**. Rozdiel medzi nimi je ilustrovaný obrázkom nižšie.

Obrázok, na ktorom je stôl

Automaticky generovaný popis

Zlúčenie

Pripojenie

Pri **zlúčení** ide o dva datasety, ktoré majú **spoločný unikátny stĺpec**, na základe ktorého sa údaje spoja. Príkladom je ak máme dataset s prevádzkami a ich adresami a dataset s adresou a rokom výstavby budovy. Na základe spoločného unikátneho stĺpca, adresy, vieme do datasetu s prevádzkami doplniť rok výstavby budovy.

**Pri pripojení ide o dva datasety**, ktoré obsahujú v **zásade podobné údaje, ktoré potrebujeme spojiť do jedného datasetu s ktorým budeme ďalej pracovať.** Príkladom sú napríklad dotácie rôznych miest, kde pri každom meste budeme mať napríklad sumu, žiadateľa, názov projektu.

**Pri spájaní datasetov musíme dbať na to, aby spájané stĺpce mali rovnaký typ (napr. spájame text s textom, čísla s číslami, atď.)**

**Pri zlučovaní datasetov používame na zlúčenie pole, ktoré má nastavený typ na text (aj ak obsahuje čísla).**

Pri **zlúčení datasetov** **potiahneme druhý dataset k prvému**, pričom sa nám medzi nimi vytvorí oranžová čiara – spojenie. Následne nastavíme dole pole, ktoré je v oboch datasetoch spoločné.

**Obrázok, na ktorom je text

Automaticky generovaný popis**

Pri **pripojení datasetov** pridáme prvý dataset, **klikneme na šípku vpravo a vyberieme Convert to Union**. Následne **potiahneme datasety, ktoré sa majú pripojiť**.

Obrázok, na ktorom je text

Automaticky generovaný popis

# Publikovanie

Vizualizáciu publikujeme vybratím **File > Save To Tableau Public**.

Aby sme vizualizáciu mohli zverejniť, prihlásime sa na server. Tu použijeme meno a heslo, ktoré sme získali pri registrácii.

Obrázok, na ktorom je text

Automaticky generovaný popis

Následne sa vizualizácia nahrá na server, kde môžeme získať kód na jej zdieľanie kliknutím na ikonku vpravo dole s 3 krúžkami .

Vizualizáciu môžeme následne zdieľať na sociálnych sieťach, posielať e-mailom alebo zabudovať do svojej web stránky (embed kód)