

Práce s daty

ANALÝZA DAT

25. ÚNORA 2021

Práce s daty v médiích



David Merhaut, Letovice

Dnes 2:50

 +1 |  -33

2800 nakažených a 81 mrtvých? Tak to je úmrtnost cca 35% !!!!! Kde je teda 2% procent co všude píšou? Nebo neumím matematiku? Takže jestli nekecaj s množstvím nakažených tak je to teda Průser !!!

novinky.cz

Typy proměnných

- nominální (nominal)
 - o dvou hodnotách lze říci pouze to, zda jsou stejné či různé
 - např. pohlaví, TV stanice, typ formy vysílání apod.

kvalitativní data

- ordinální (ordinal)
 - u hodnot je navíc možné určit jejich pořadí („co je více a co méně“)
 - např. úroveň dosaženého vzdělání, míra spokojenosti s obsahem pořadu apod.

lze užít jako oba typy dat

- intervalová (interval)
 - lze vypočítat, o kolik je jedna hodnota větší (menší) než druhá
 - např. míra růstu HDP, výše inflace apod.

kvantitativní data

- poměrová (ratio)
 - lze vypočítat, kolikrát je jedna hodnota větší (menší) než druhá – existuje absolutní nula, pouze kladné hodnoty
 - např. počet hlasů získaných ve volbách, počet poslanců hlasujících pro radní ČRo apod.

Kvantitativní data

- informace, které na sebe berou numerickou podobu
- je možné je statisticky zpracovávat
- jejich hlavní výhodou je schopnost efektivně popsat velké množství případů
- jejich hlavní nevýhodou je neschopnost hlubšího porozumění jednotlivým případům
- význam kvantitativních dat v dnešním světě roste, a to včetně oblasti médií
- je důležité být si vědom nedostatků a vyvažovat je využitím kvalitativních dat



juffrouj.wordpress.com

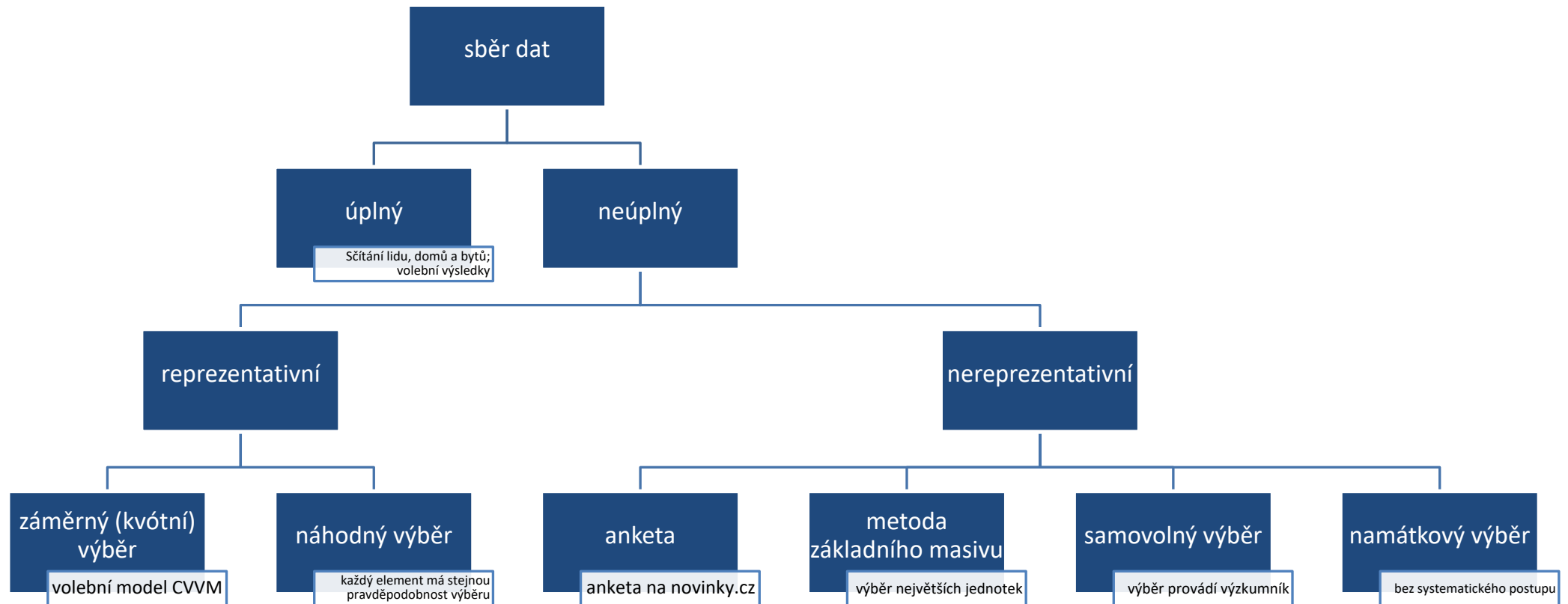
Cvičení

- jakého typu jsou následující proměnné (nominální, ordinální, intervalové, nebo poměrové?):
 - jméno hosta
 - počet odvysílaných pořadů
 - pořadí pořadů v bodovém scénáři
 - místo narození respondenta
 - vzdálenost redakce média od Prahy
 - věk voliče
 - rozdíl v počtu odvysílaných pořadů z roku na rok
 - délka pořadu



integrativeosteopathy.com.au

Sběr (kvantitativních) dat



Práce s datovým setem

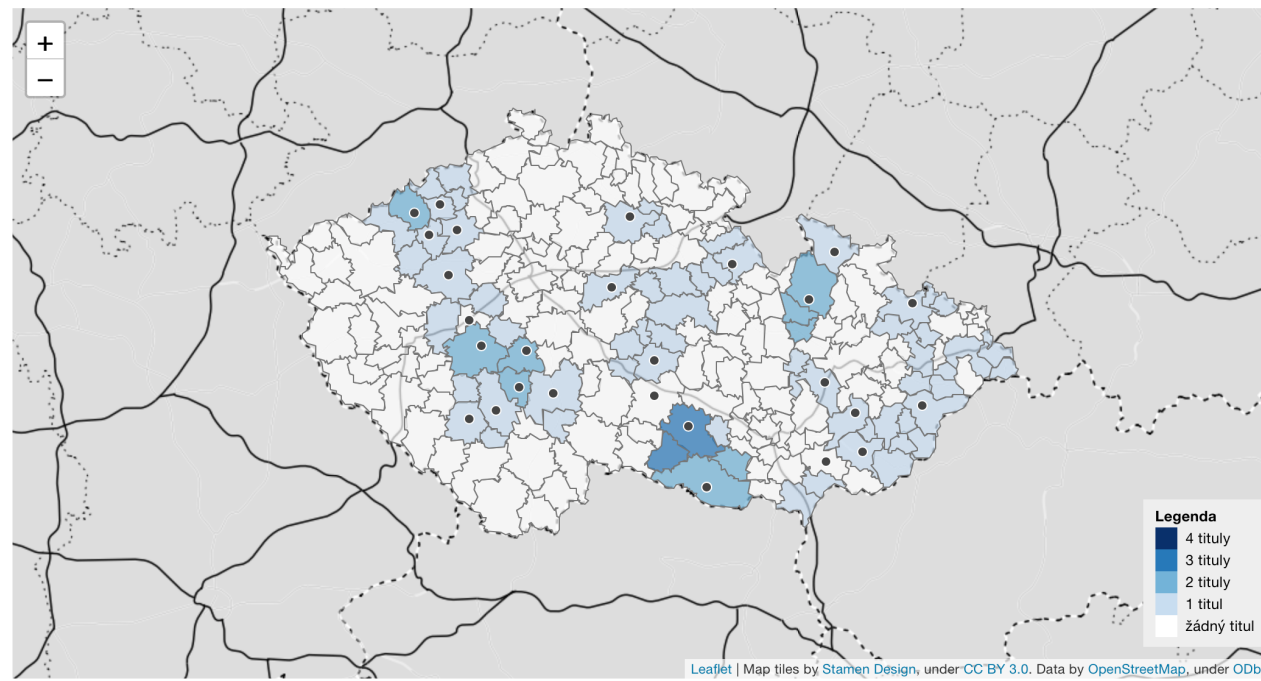
- zajímavá data poskytuje například [Český sociálněvědní datový archiv](#)
 - volební, mediální, genderové a mnohé další výzkumy
- každý datový set je možné stáhnout včetně potřebné dokumentace
- v nabídce digitálního archivu je i možnost základních analýz
- pro stažení dat je třeba být registrován (zdarma)
- podívejme se detailněji například na pravidelný průzkum Naše společnost od CVVM

Zásady práce s datovým setem

- je třeba být seznámen a neustále pracovat s kódovací knihou
- při analýze je nutné vyřadit informace, které jsou mimo pracovní škálu (bez odpovědi, nevím, chybějící informace apod.) – typicky jsou označeny např. jako „9“ nebo „99“
- při práci s daty je třeba si vždy uvědomit, zda jde o proměnné nominální, ordinální, intervalové nebo poměrové
- nepřeházet řazením dat pouze v jednom sloupci všechna pozorování (při seřazení hodnot v jednom sloupci se přirozeně musí změnit pořadí celých řádků, nikoliv jen polí v předmětném sloupci)
- pozor na překlepy – mohou se objevit i v profesionálních datasetech
- uvědomte si, jak příslušný software pracuje s prázdnými poli (chybějící informace)
- při vlastním kódování vhodně nazývejte proměnné a maximálně využívejte kvantitativní podobu informace (ideálně si ihned запиšte způsob operacionalizace)

Analýza médií

- už s jednoduchými nástroji lze vytvářet zajímavé analýzy
- podívejme se na [článek](#) na webu iRozhlas.cz o regionálních novinách



Data: Lenka Waschková Císařová

irozhlas.cz

Replikace – wordcloud

- podívejme se na [článek](#) serveru iRozhlas.cz o jednání v Poslanecké sněmovně
- wordcloud je oblíbeným nástrojem zobrazení textových dat
- nástroj se pohybuje na pomezí práce s kvalitativními a kvantitativními daty
- pro výrobu wordcloudu je možné využít některý z webových nástrojů
 - např. <https://www.wordclouds.com>
 - nebo <https://wordart.com/create>

Replikace – wordcloud

- [článek](#) serveru iRozhlas.cz využíval data z [digitální knihovny](#) českého parlamentu
- čím více aktuální je volební období, tím více dat knihovna z dané doby obsahuje



psp.cz

Replikace – wordcloud

- pro vytvoření unikátního obsahu se podívejme na novoroční projevy a vánoční poselství českých prezidentů
- zdrojem může být např. unikátní [databáze](#) Českého rozhlasu
- obsahuje všechny projevy a poselství od roku 1935

**Od Společnosti národů k
,lepšolidem`. Poslechněte si
prezidentské vánoční projevy
od roku 1935 dodnes**

irozhlas.cz

Uhádnete prezidenta?



Uhádnete prezidenta?



Václav Havel, 1. ledna 1990

Replikace – wordcloud

- vytvořme sami analýzu některého projevu
- zdrojem je tedy [databáze](#) Českého rozhlasu
- využijme <https://wordart.com/create>
 - nejprve je třeba importovat příslušný text
 - následně odstranit spojky, ukazovací zájmena či předložky
 - pro finální zobrazení je možné využít celou řadu parametrů



Replikace – tweety

- podívejme se na [článek](#) z ČT24 o tweetování o brexitu
- společně si ukažme postup analýzy

O brexitu ve 400 tisících tweetech: dominuje důvěra i očekávání

5. 3. 2019

Datum odchodu Velké Británie z Evropské unie se nezadržitelně blíží. Stále přitom není jasné, k jaké podobě brexitu nakonec dojde. Ze všech stran jsou proto analyzovány názory britských poslanců a unijních vyjednávačů. Jak ale brexit v poslední fázi jeho vývoje vnímají sami obyvatelé Spojeného království a dalších zemí? Web ČT24 přináší unikátní analýzu 400 tisíc tweetů o brexitu.

ct24.ceskatelevize.cz

Shrnutí

- existují různé typy proměnných – podle typů dat volíme nástroje analýzy
- při práci s daty je třeba dodržovat řadu pravidel
- už základní deskriptivní datové analýzy pak umožňují vytvářet velmi zajímavé výstupy
- je na nich založena velká část datové žurnalistiky, ale i analýzy médií
- důležité je především přijít s dobrým nápadem a aplikovat vhodné analytické nástroje