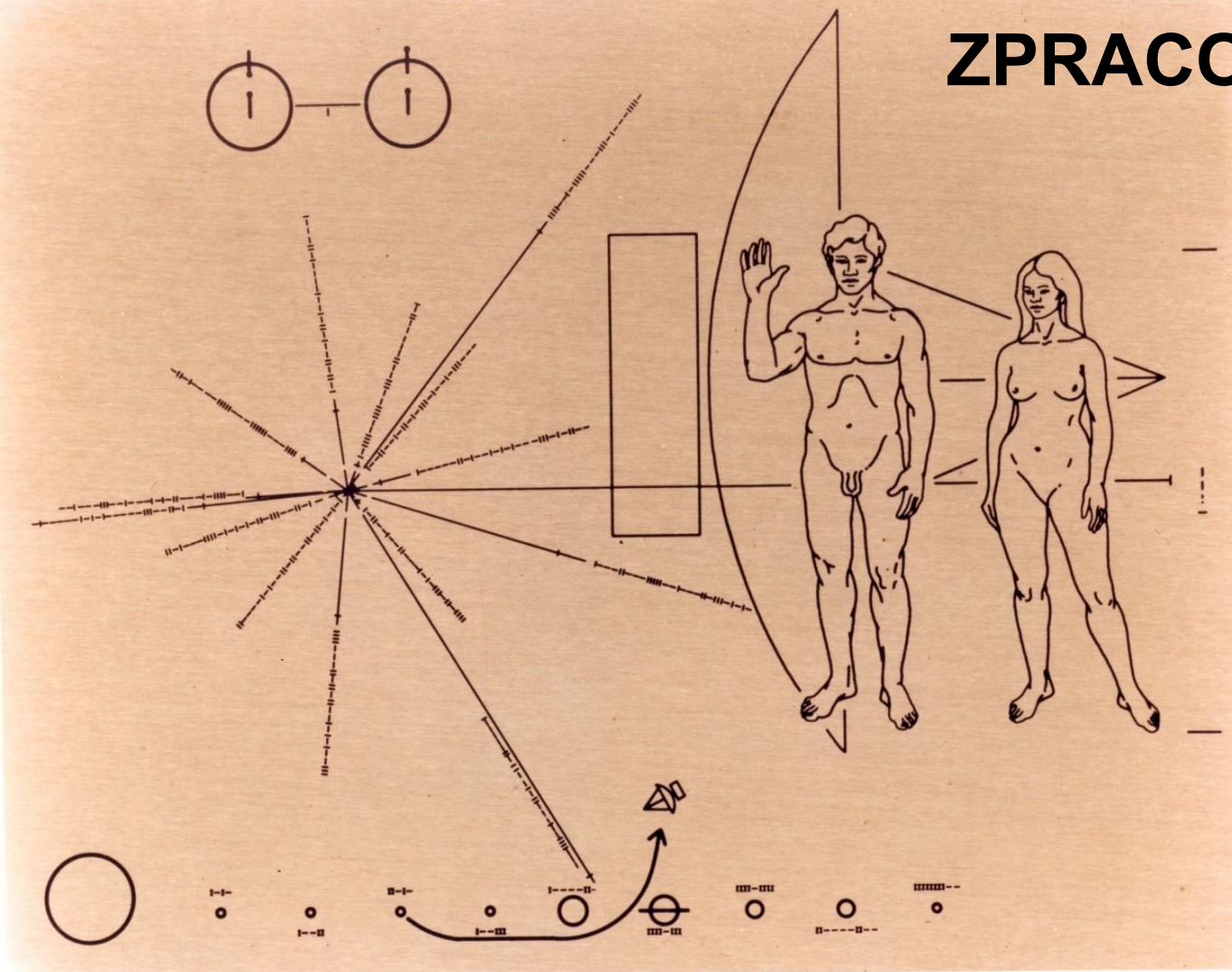


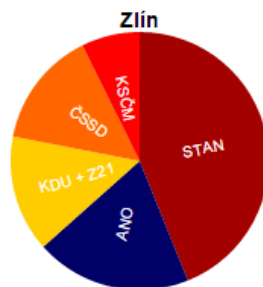
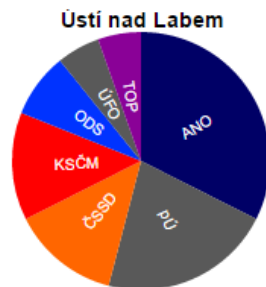
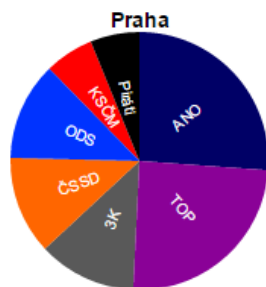
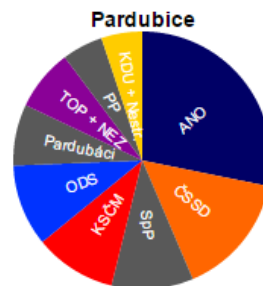
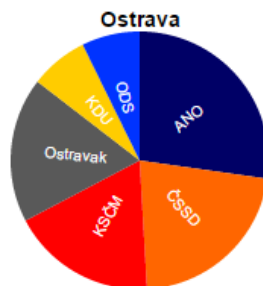
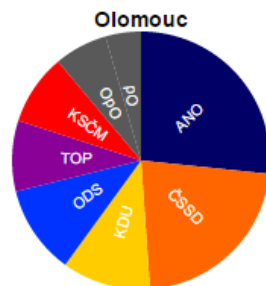
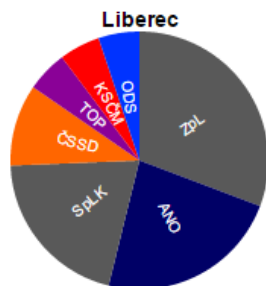
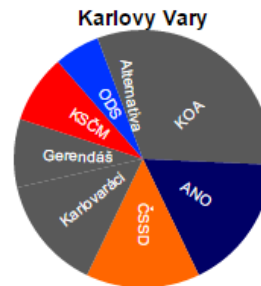
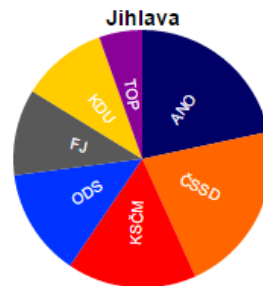
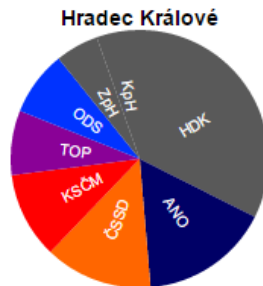
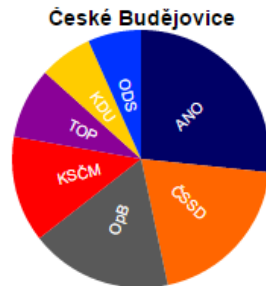
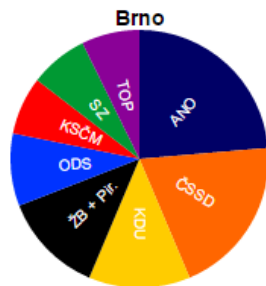
ZPRACOVÁNÍ DAT



Dnes

- Hejt
- Nástroje pro zpracování dat
- Typické postupy a problémy
- Trochu seminář!

- Zastupitelstvo
- Koalice

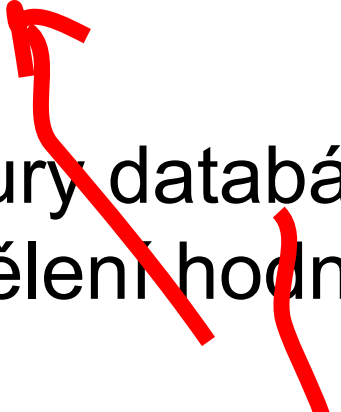


Ještě víc koláčových grafů:

106 s Cibulkou

- otázky?

Zpracování dat

- surová data -> data pro analýzu/vizualizaci
 - první krok, kdy si osaháte data
 - **čištění dat:** oprava chyb
 - **strukturace dat:** změna struktury databáze (vypuštění sloupců/řádků, rozdělení hodnot, spojování databází...)
- 

Typické problémy

- kdo nejvíc/nejméně (filtrování dat)
- kdo nejvíc/nejméně dohromady (agregace dat, pivot tables, R)
- kolik je to na 1 obyvatele (spojování dat, Google Fusion Tables, R)
- jak to bude vypadat v grafu (náhledová vizualizace)
- je to v PDF, kua! (OCR, Tabula)

Typické problémy

- datum a čas zvlášť (spojování/rozdělování hodnot, formátování buněk)
- rozsypala se čeština (kódování jazyka, Sublime Text)
- PROČ ZASE! (Ctrl+C, Ctrl+V) (formátování buněk)
- ALE TO JE ČÍSLO!!! (. a ,) (nastavení jazyka, hromadné nahrazení)

Typické problémy

- nefungují funkce (volba nástroje)
- chci odmazat každý druhý řádek (R, DataWrangler)
- jak to vypadá na mapě (Google Fusion Tables, QGis)
- chci pěknou a rychlou časovou osu (Timeline JS, R, D3)
- jak to vypadá v čase na mapě (?)

Typické problémy

- jak vyčistit tenhle bordel (Open Refine)
- chci hromadně nahradit text, co začíná tímhle a končí tadytím (regulární výrazy)
- moc dat, Excel to neskousne (R)
- chci statistické funkce (R)
- nutně musím stáhnout ČSFD (scrapování) (R, ScraperWiki)
- ČSFD mě zablokovalo :((((TunnelBear :)))

Typické (hardcore) problémy

- Šlerka mluvil o grafu vztahů (Gephi)
- našel jsem API, co s tím (R)
- co dělá tenhle zdrojový kód (JS Fiddle)
- upravím si cizí kód (Sublime Text, D3)
- potřebuju spolupracovat na kódu (Git)
- chci se pochlubit, co jsem naprogramoval (GitHub)

Základní nástroje

- MS Office 2013
 - analýza dat v pivot tables
 - limity Excelu 2003, 2007, 2010, 2013
- Libre Office
 - ideální na CSV
- Google Docs
 - spolehlivější funkce, online funkce, sdílení
 - pomalejší
 - Google Fusion Tables
- texták (Sublime Text)
 - znaková sada, náhled CSV, editor kódu ...

Další nástroje

- R
 - free statistický soft
 - obrovská komunita, milion funkcí
 - kurz na Courseře
- Open Refine
 - čištění dat - včetně lingvistických
 - dřív Google Refine, špatná podpora
- QGis
 - všechno nad mapami, kromě zveřejňování online
 - lacinější varianta ArcGisu

Další nástroje

- Gephi
 - SNA
 - parádní statické obrázky, obtížné interaktivní zveřejnění
- DataWrangler
 - úprava špatně formátovaných tabulek
- TunnelBear
 - změna IP adresy
- Tabula
 - dolování dat z PDF
- Mr. Data Converter
 - převody formátů dat

Další nástroje

- Data Wrapper
 - rychlá vizualizace
- Google Charts
 - vlastní vizualizace, hotové kusy kódu
- D3.js
 - kompletně vlastní kód
- Git, GitHub
 - sdílení kódu

A teď prakticky!

- distribuce léčiv: SÚKL
- 106?
- výzkumné otázky?
- www.sukl.cz/2014

Odtahy v Brně

- DO ÚTERÝ TAJNÝ !!

Odtahy v Brně

- dataviz.cz/brnoodtahy.csv
- na hraní

Odtahy v Brně

- teď na mapu
- Google Fusion Tables

DÚ

- vybrat si finální příběh
- odevzdat surová data
- odevzdat upravená a vyčištěná data