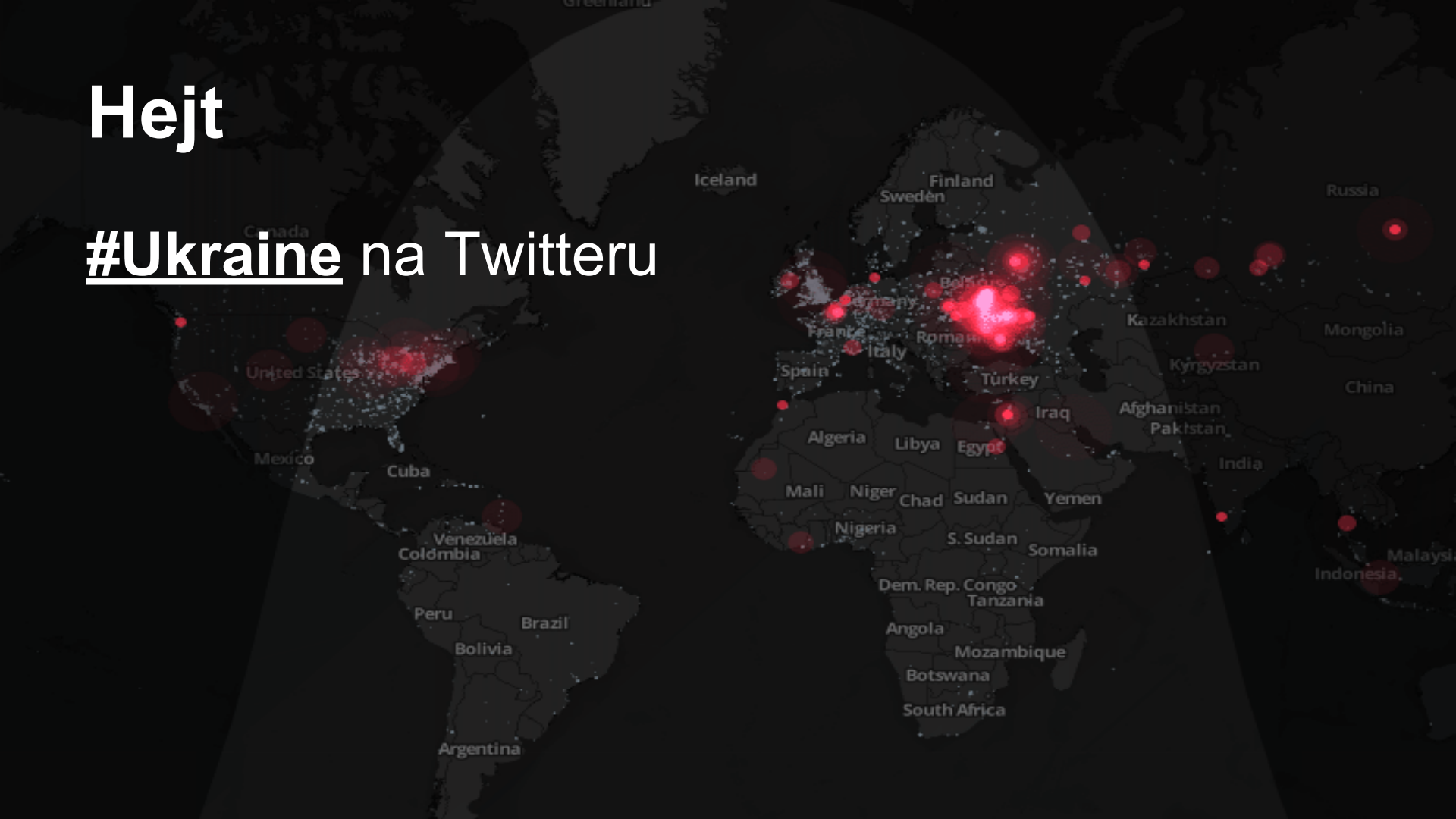
Zdroje dat

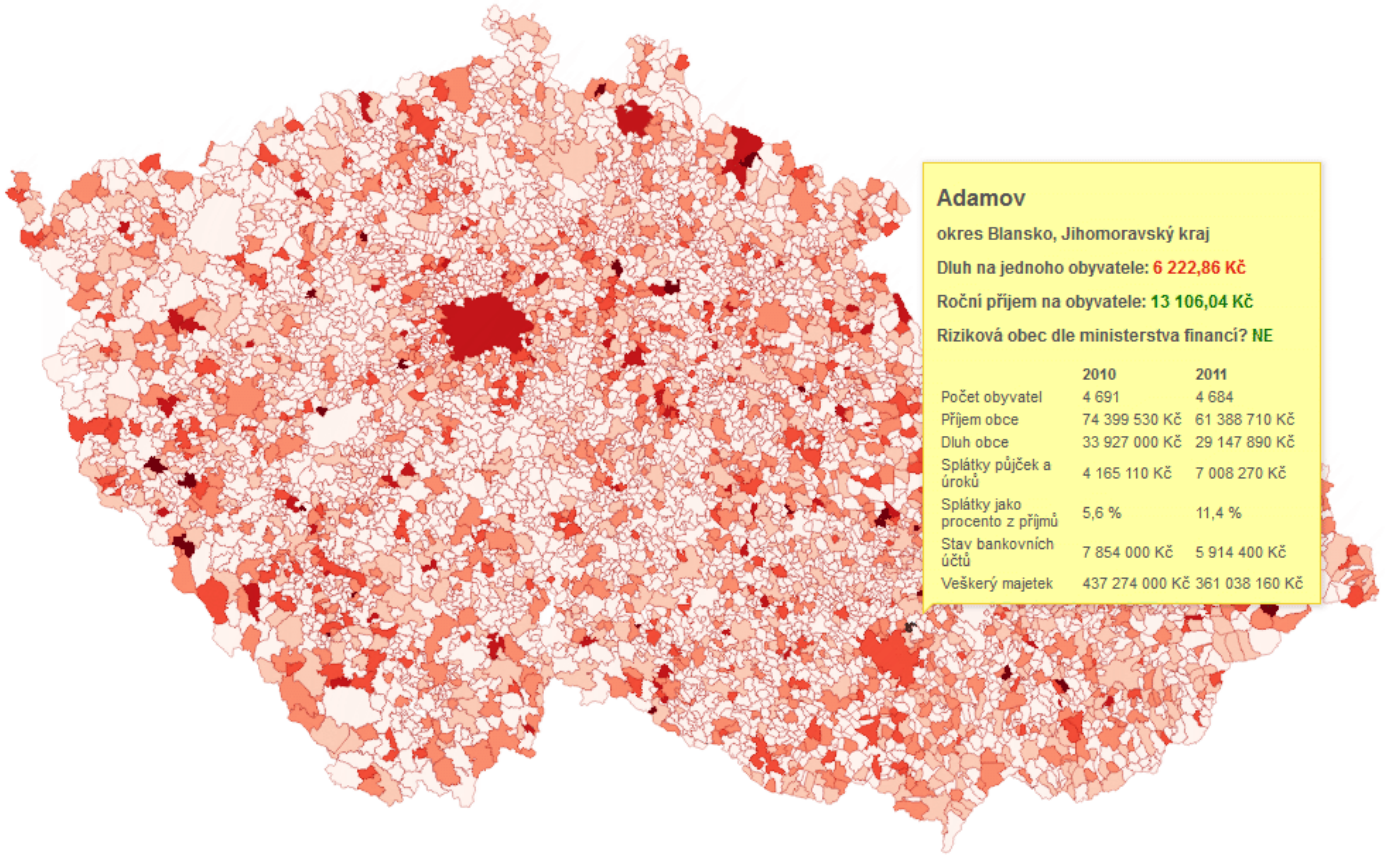
Hejt

#Ukraine na Twitteru



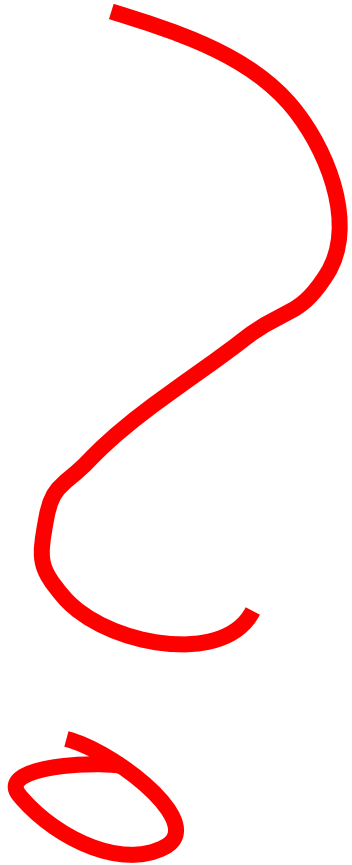
Hejt

- **Mapa dluhů**
- **Mapa Ukrajiny**
- **Bezpečnost na D1**



Plán

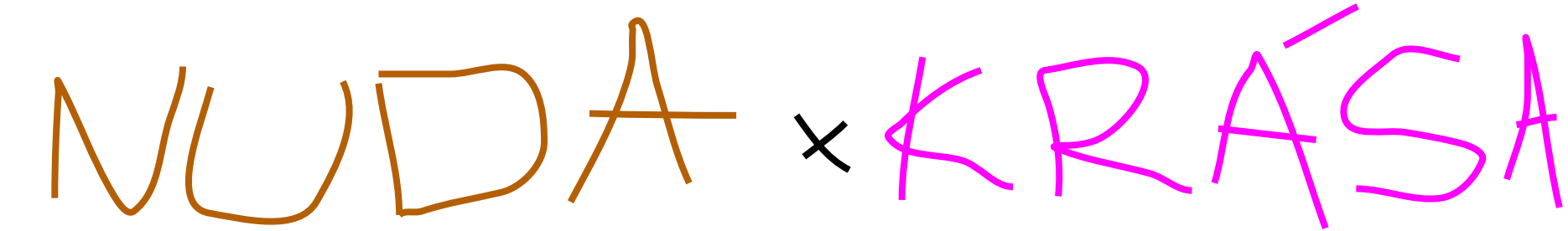
- Zdroje a formáty dat - přednáška
- Čištění dat - seminář
- Scrapování webu - seminář



Zdroje dat

Připravená data
Standardní formáty
Málo práce
Weby institucí

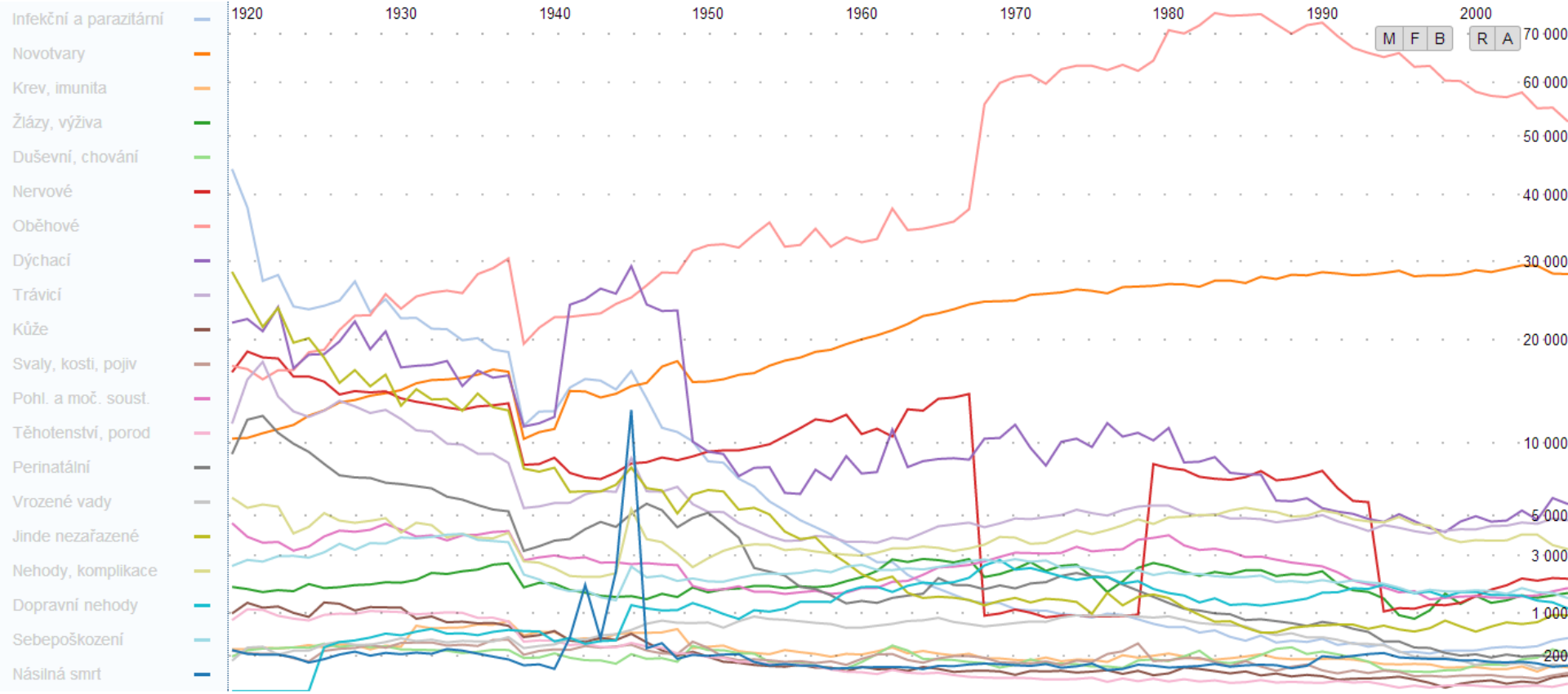
Data neexistují
Existují, ale jsou tajná
Spousta práce
(I programování!)



Data dostupná všem

- ČSÚ (**nuda**, **krása**)
- Ministerstva, ŘSD, ÚZIS ...
- Globálně: Eurostat, OSN, WB, WHO ...
- Lokálně: Open Data (**USA**, **jinde**)
- Tipy: **Datablog**

Živě


Infekční a parazitární
Novotvary
Krev, imunita
Žlázy, výživa
Duševní, chování
Nervové
Oběhové
Dýchací
Trávicí
Kůže
Svaly, kosti, pojiv
Pohl. a moč. soust.
Těhotenství, porod
Perinatální
Vrozené vady
Jinde nezařazené
Nehody, komplikace
Dopravní nehody
Sebepoškození
Násilná smrt
1920
1930
1940
1950
1960
1970
1980
1990
2000
M F B
R A
70 000
60 000
50 000
40 000
30 000
20 000
10 000
5 000
3 000
1 000
200

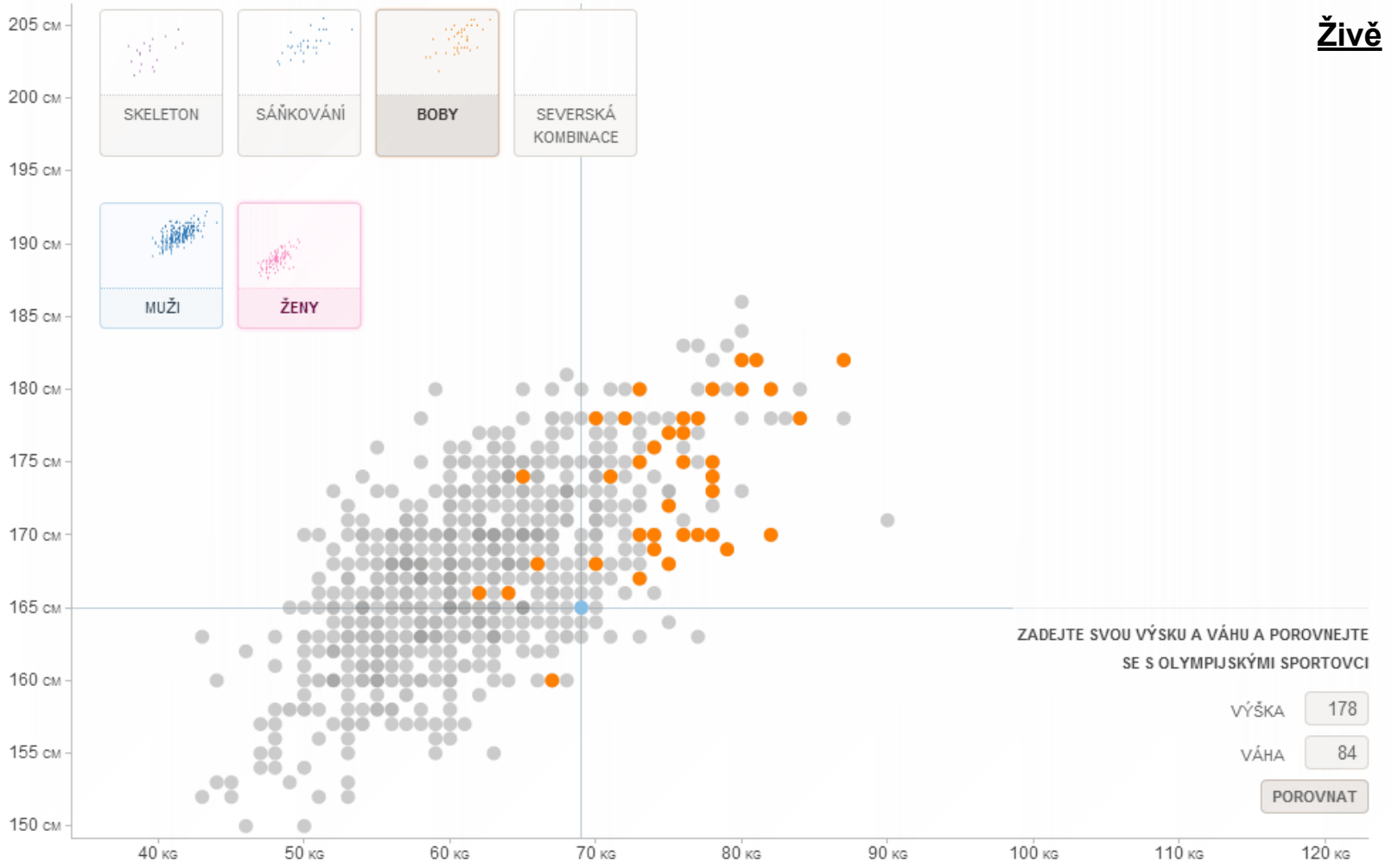
Data dostupná všem, ale

- Jak vypadají data pro volební prognózy ČT?
- Kolik platí VZP za konkrétní diagnózy?
- Kdo cestuje s prezidentskými delegacemi?
- **Odpověď: <u>106</u>!**
- Pozor na zapadlé kouty webu, PDF

Data, co zatím neexistují

- Diskuze o mazání na Wikipedii (**interaktivně**)
- Barvy v různých kulturách (**interaktivně**)
- Vzorce brněnských semaforů
- **FF odpověď:** papír, nůžky, pastelky a ∞ času
- **FI odpověď:** scrapování

Živě

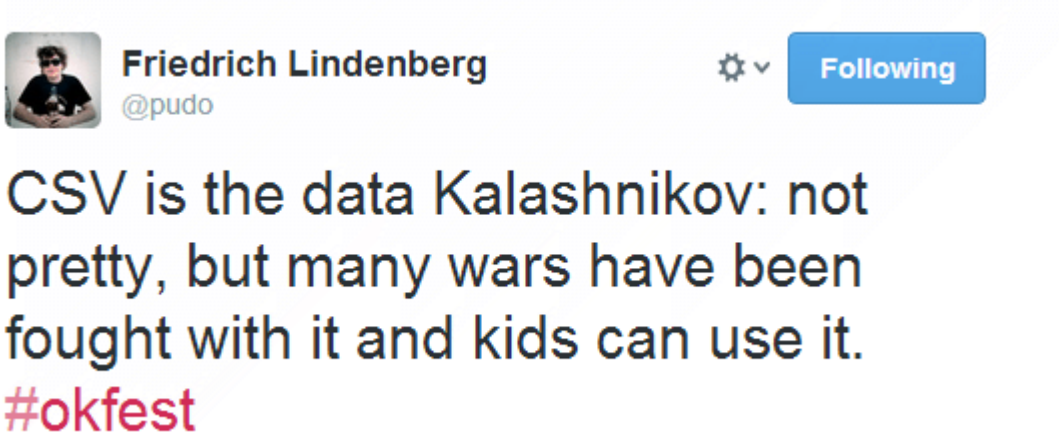

SKELETON
SÁŇKOVÁNÍ
BOBY
SEVERSKÁ KOMBINACE
MUŽI
ŽENY
205 CM
200 CM
195 CM
190 CM
185 CM
180 CM
175 CM
170 CM
165 CM
160 CM
155 CM
150 CM
40 KG
50 KG
60 KG
70 KG
80 KG
90 KG
100 KG
110 KG
120 KG
ZADEJTE SVOU VÝSKU A VÁHU A POROVNEJTE SE S OLYMPIJSKÝMI SPORTOVCI
VÝŠKA 178
VÁHA 84
POROVNAT

Formáty souborů

- XLS

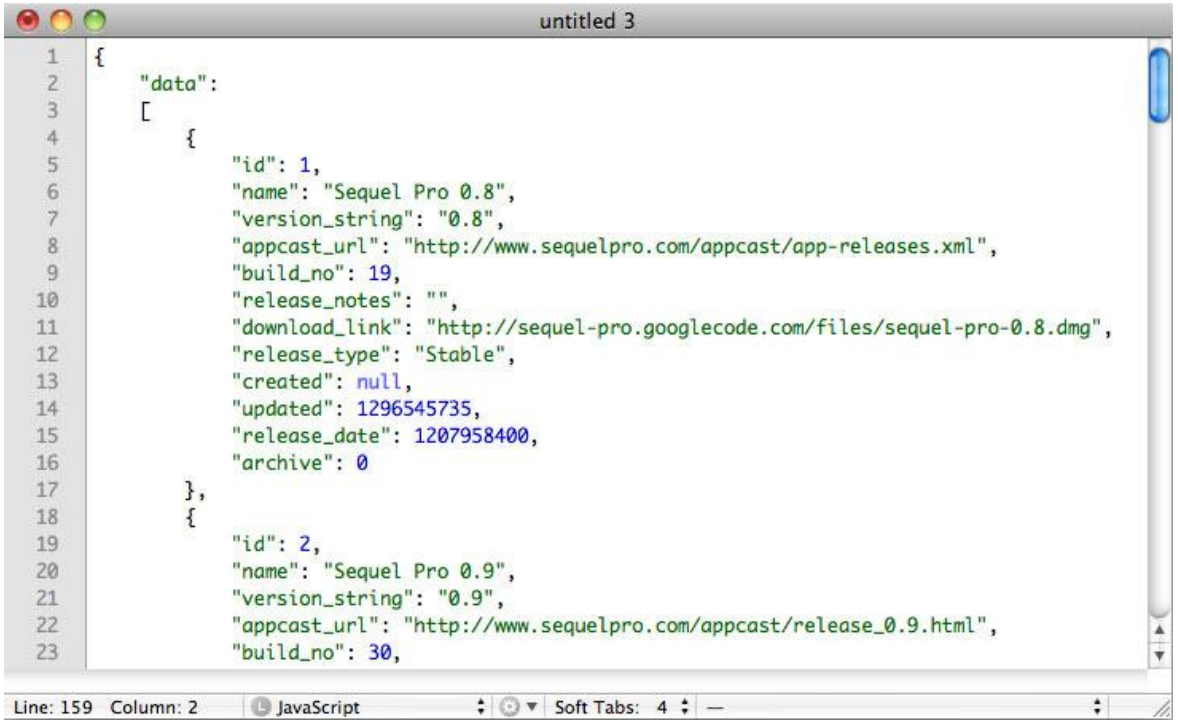
Formáty souborů

- XLS
- CSV



Formáty souborů

- XLS
- CSV
- JSON



Formáty souborů

- XLS
- CSV
- JSON
- **Mr. Data Converter**

Čištění dat

Čištění dat

- Vyčistit bordel (typicky: veřejné zakázky)
- Učesat klíčové proměnné
- Smazat ostatní
- Nejistota!

Google Docs

- Zkopírujte si data (**sportovci v Soči/sochi.jdem.cz**)
- Disciplína, stát, pohlaví: přeložit
- Jméno: rozdělit na křestní a příjmení
- Výška a váha: rozhodnout, kdo má zhubnout
- Výška a váha: orientační graf pro sporty