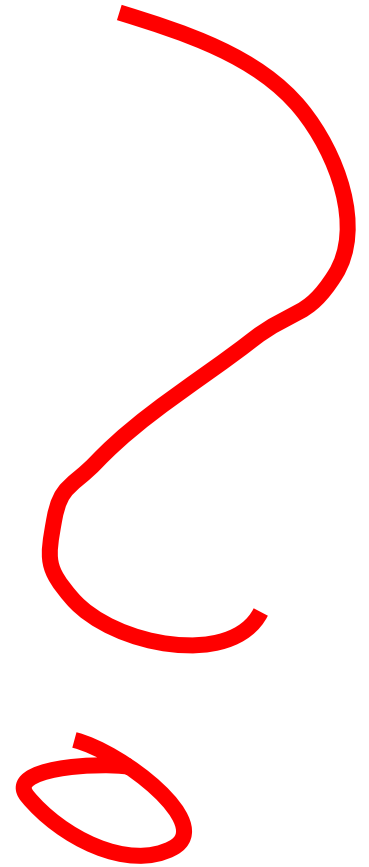


# **Analýza a vizualizace**

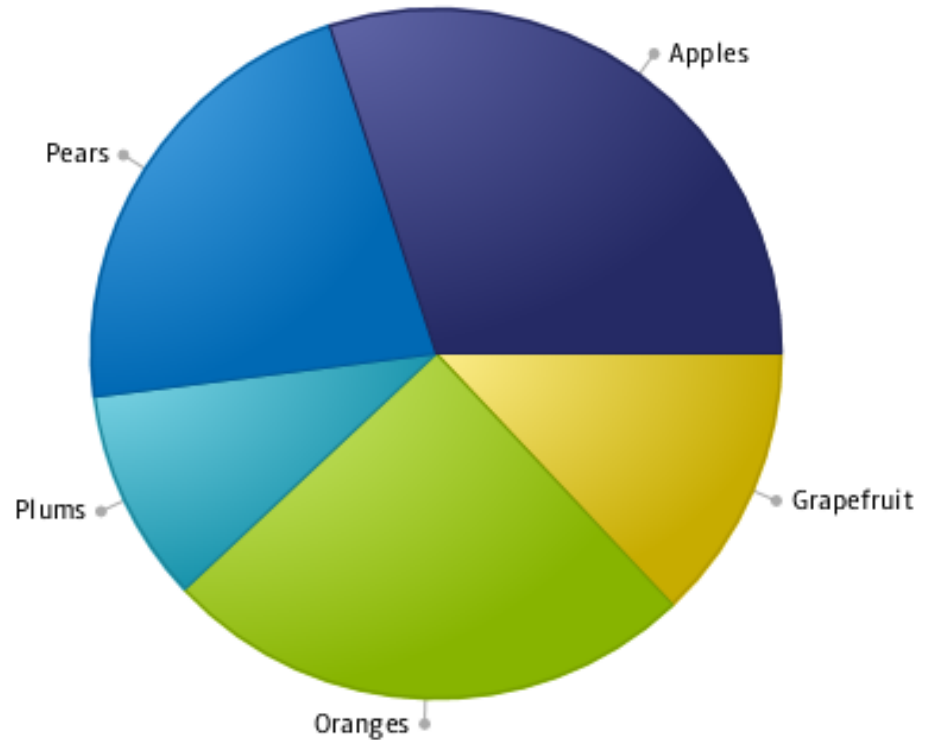
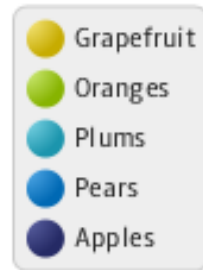
# Plán

- Hejty
- Rest - CSV
- Analýza dat - mapy
- Scrapování webu
- DÚ

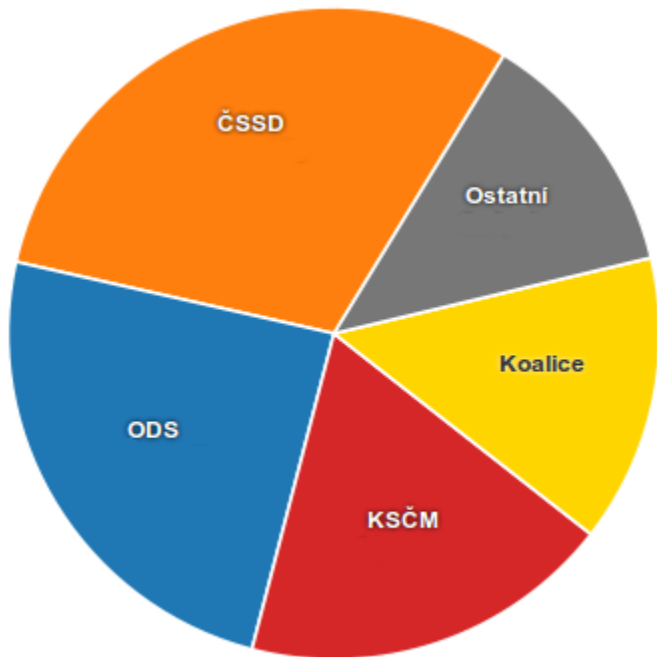


# Hejt: piecharty

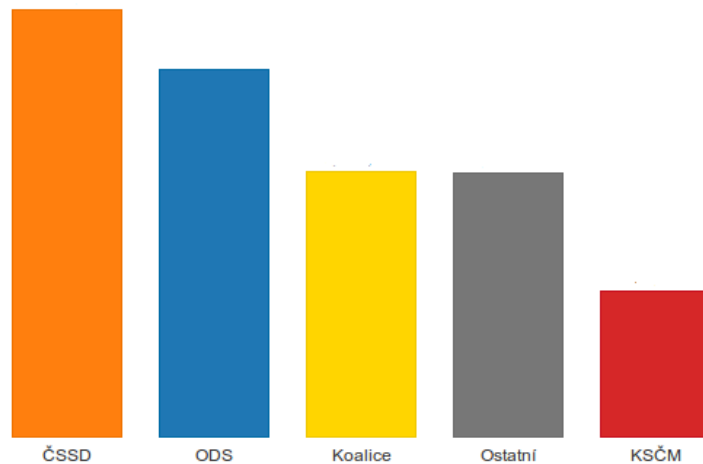
Imported Fruits



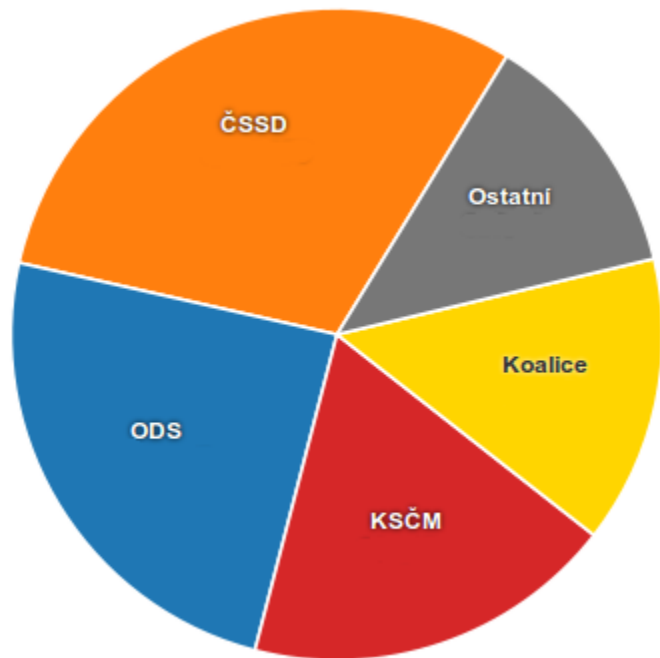
# Hejt: piecharty



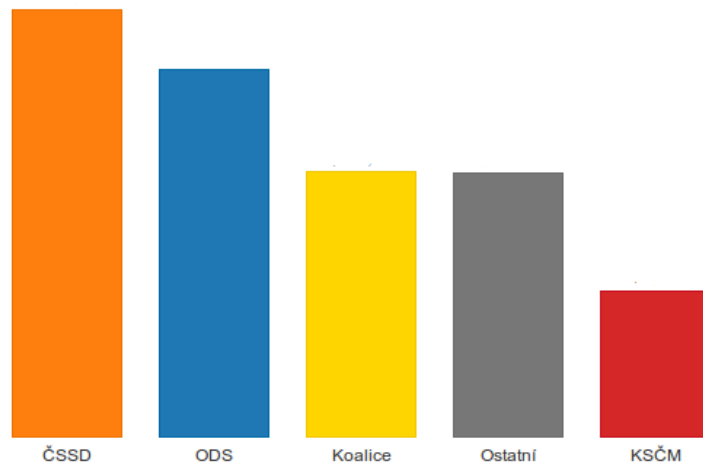
×



# Hejt: piecharty



×







■ FLIGHT 370 | 4:57 PM | MAR 17, 2014

## How Statisticians Could Help Find That Missing Plane

By CARL BIALIK

What happened to Malaysia Airlines Flight 370, and where is it now?

Statistical tools can't answer those questions any more definitively than Malaysian officials have. Yet they can help refine and focus the hunt for the

### TOP STORIES



# Láv

- Vizualizace dat na FB





# Formáty souborů

- CSV



Friedrich Lindenberg

@pudo



Following

CSV is the data Kalashnikov: not pretty, but many wars have been fought with it and kids can use it.

#okfest

# CSV do Excelu, semafony na mapu

- data: [semafony.jdem.cz](http://semafony.jdem.cz)
- Google Fusion Tables

# Google Maps

- Kde jsou průměrně nejstarší sportovci?
- Data: [sochivek.jdem.cz](http://sochivek.jdem.cz)
- Vytvořit z dat Fusion table
- Vyhledat KML s hranicemi: [research.google.com/tables](http://research.google.com/tables)
- Spojit obě tabulky

# OpenRefine

- Čištění překlepů, nekonzistencí, bordelů
- ... a všechno ostatní
- Ale vlastní skript, blbě se gúglí
- Offline (stáhněte si)
- Data z registru (stáhněte si)

# Registr dopravních vozidel

- “ŠTĚPÁN INDIVIDUÁL”
- “PEPA 1”
- “ŠOA SPORT”

# Scrapování 01/2014 leden

alba, stránky, přání, zápisníky



Vintage

Knoflíčky

Pro švadlenky

Inchies

Výběr alba

Projekt Life



# Scrapování

- “Olizování webu”
- Když existují data a neexistuje dataset
- Naprosto v pořádku u státní správy
- Pozor na soukromé weby

# Soči

- Strojové procházení webu a vyzobávání dat
- Pochopit strukturu stránky
- Pochopit strukturu webu



# Jak se to dělá?

- R/Python/ScraperWiki
- Bacha na odpojování

# DÚ

- Zkusit si scrapovací doplněk pro Chrome
- Číst 538!