

Princípy dátovej vedy - prednáška 1

Samuel Rosa

Fakulta matematiky, fyziky a informatiky
Univerzity Komenského v Bratislave

Vyučujúci:

Vladimír Boža, M-25, vladimir.boza@fmph.uniba.sk – cvičenia

Samuel Rosa, M-240, samuel.rosa@fmph.uniba.sk – prednášky

Web:

<http://www.iam.fmph.uniba.sk/ospm/Rosa/teaching.htm>

- všetky informácie

Vyučujúci:

Vladimír Boža, M-25, vladimir.boza@fmph.uniba.sk – cvičenia
Samuel Rosa, M-240, samuel.rosa@fmph.uniba.sk – prednášky

Web:

<http://www.iam.fmph.uniba.sk/ospm/Rosa/teaching.htm>

- všetky informácie
- **Google classroom:** kód nwsxwbj
- Konzultácie: dohodnúť
- Otázky vítané!
- Anonymná spätná väzba:
<https://forms.gle/jDZq4WrV5XUBDcWz9>
- Tykanie

Literatúra:

- James, Witten, Hastie, Tibshirani: An Introduction to Statistical Learning (2nd ed.), Springer, 2021
- Skiena: The Data Science Design Manual, Springer, 2017
- Grus: Data Science from Scratch, O'Reilly Media, 2019

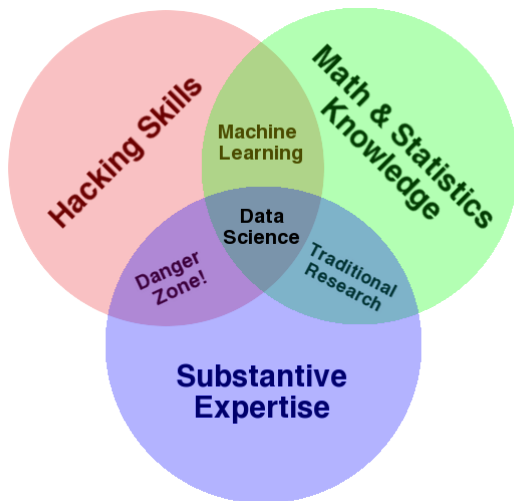
Hodnotenie: http://www.iam.fmph.uniba.sk/ospm/Rosa/PDV/PDV_hodnotenie.pdf

- zo skúšky aspoň polovica bodov, inak Fx

Opisovanie:

- neopisujte!
- -100% príslušnej časti bodového hodnotenia
- diskusia ohľadom domácich úloh je ok, ALE:
 - nerobte si poznámky
 - počkajte niekoľko hodín, kým začnete realizovať vlastné riešenie

Dátová věda



zdroj: Drew Conway



Josh Wills ✓

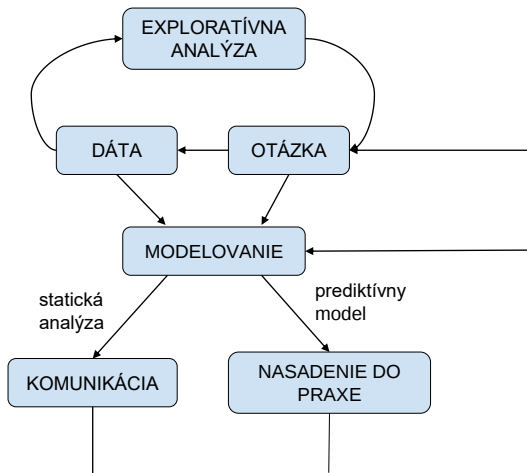
@josh_wills

...

Data Scientist (n.): Person who is better at statistics than any software engineer and better at software engineering than any statistician.

6:55 PM · May 3, 2012

1,752 Reposts **96** Quotes **2,140** Likes **88** Bookmarks



Voľná osnova:

1. Exploratívna analýza
2. **Predikčné modely**
3. Princípy štatistiky
4. Zhukovanie
5. Pravdepodobnostné modelovanie
6. Štatistické testy
7. Veľké dáta
8. Etika

Ciele:

- zoznámenie sa s princípmi
- zoznámenie sa s metódami
- prehľad, akými smermi sa dá ísť ďalej

Exploratívna analýza

National Center for Health Statistics

CDC > NCHS

National Health and Nutrition Examination Survey

About NHANES



What's New



Webinar

Questionnaires, Datasets, and
Related Documentation



Survey Participants



Biospecimen Program



New Content and Proposal
Guidelines



[Print](#)

National Health and Nutrition Examination Survey

Survey Participants



If you were selected,
learn more about
participating

Survey Data and Documentation



Access data,
documentation &
response rates

Publications and Products



View health and
nutrition reports & CDC
Growth Charts

Data Analysis Tutorials

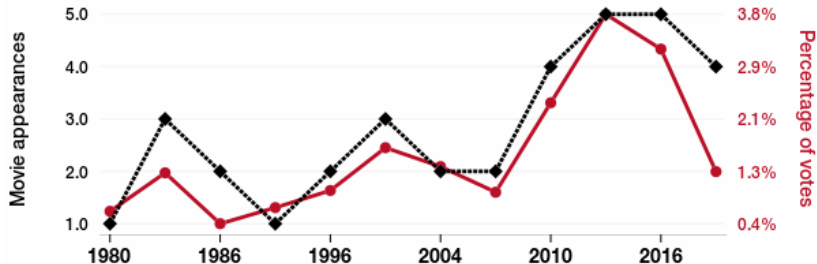


Review step-by-step
guidance on using
NHANES data

The number of movies Nicolas Cage appeared in

correlates with

Votes for Libertarian Senators in Illinois

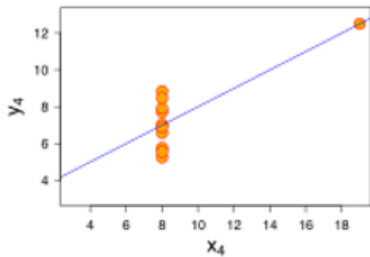
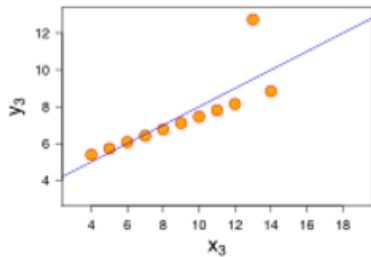
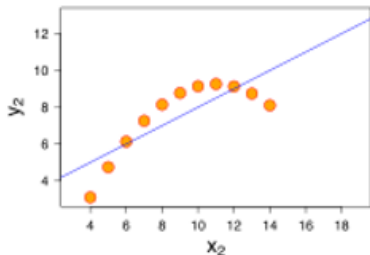
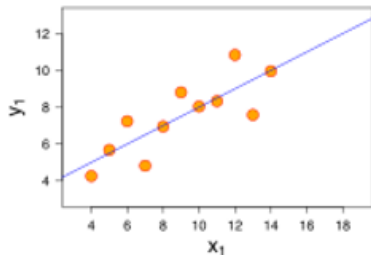


◆ The number of movies Nicolas Cage appeared in · Source: The Movie DB

● Percentage of votes cast for Federal Libertarian Senate candidates in Illinois
· Source: MIT Election Data and Science Lab, Harvard Dataverse

1980-2020, $r=0.885$, $r^2=0.784$, $p<0.01$ · tylervigen.com/spurious/correlation/5847

Anscombe's quartet



Datasaurus dozen

