

New Course: Statistical Data Science and Visualisation

MSc Mathematical Statistics, Institute of Mathematical Statistics, Faculty of Mathematics and Computer Science, University of Tartu

Kaspar Märtens, Tanel Pärnamaa

<http://andmeteadus.github.io/>

MOTIVATION

Study IT in .ee   UNIVERSITY OF TARTU
Institute of Computer Science

Problem

- How to create attractive graphs for visual storytelling?
- How to make your data analysis reproducible?
- How to increase interdisciplinarity between computer science and statistics?

Solution

There was no such course at the University of Tartu, so we have created and organised one. Moreover, our course „Statistical Data Science and Visualisation“ is a small step towards bridging the gap between computer science and statistics.

OUTLINE OF TOPICS

Week 1

- How to create an automatic stock market report?
- Tools for reproducible research in R.

Week 2

- How to detect fraud?
- Reproducing and analysing research paper „Growth in a Time of Debt“ by Harvard economists.

Week 3

- How to create an interactive data product?
- Answering questions from the course feedback data:
 - Is it easier to get better grades in some faculties?
 - Do courses with smaller workload get better feedback?

Week 4

- Techniques for creating visualisations like these:

Week 5

- How to scrape data from the web?
- Voting patterns of Estonian Parliament members.

Week 6

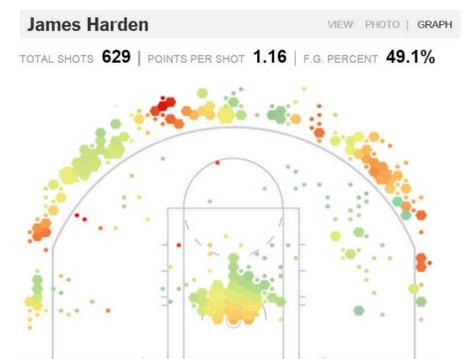
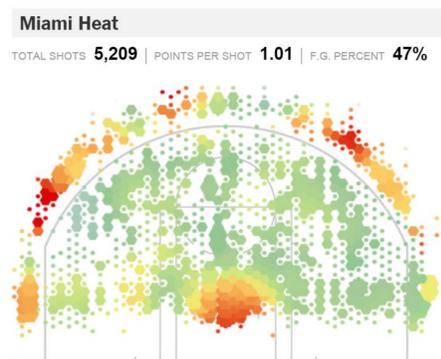
- How to visualize complex highdimensional data?
- Dimensionality reduction.

Week 7

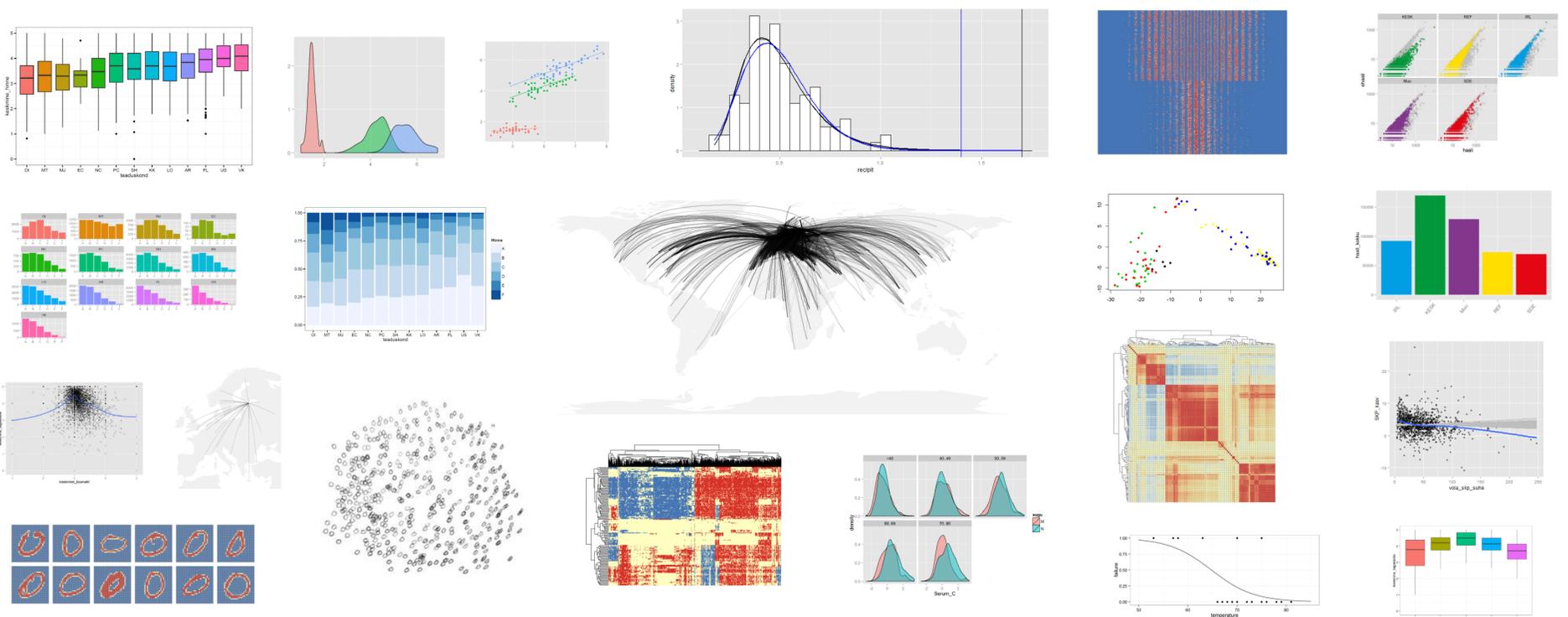
- How to predict death?
- Reproducing the analysis in „Biomarker Profiling by Nuclear Magnetic Resonance Spectroscopy for the Prediction of All-Cause Mortality: An Observational Study of 17,345 Persons“ (K. Fischer *et al*, 2014)

Week 8

- What's wrong with doping tests?
- Analysing the Andrus Veerpalu case.



In 2015 spring semester, we had Bsc, Msc and PhD students from Stats and CS. During the course, they created tons of graphs, for example:



COURSE MATERIALS

All materials are publicly available on the course webpage <http://andmeteadus.github.io/> The focus is on exploring recent case-studies from different areas. These are accompanied by introductory minimal examples in R. To cover new concepts, we have provided links to selected video lectures from the world top universities.