Course Introduction

Data Science & AI

Sabine De Vreese **Stijn Lievens** **Lieven Smits** **Bert Van Vreckem**

2021–2022

HO GENT

Contents

Course guide

Introduction

HO
GENT

Course guide

HO
GENT

Learning goals

- Descriptive Statistics
  - Knows some descriptive measures for data (central tendency, dispersion).
  - Can calculate some descriptive measures for data using statistical software (Python).
- Data visualisation
  - Knows different types of plots to represent data visually.
  - Can visualize data using appropriate plots
- Probability
  - Knows the basic rules with regard to calculating with probabilities.
  - Knows the properties of some important probability distributions.

HO GENT

Learning goals

- Bivariate analysis
  - Can quantify and appropriately test the relationship between two variables.
  - Can construct a simple linear model to show the relationship between two or more variables.
- Time series analysis
  - Can discuss some common models to predict time series and/or detect anomalies.
  - Can indicate the importance of testing the accuracy of a model in a methodologically correct manner.

HO GENT

Course contents

- Introduction, sampling
- Univariate analysis
- Probability, central limit theorem, statistical testing
- Bivariate analysis: qualitative variables
- Bivariate analysis: qualitative vs. quantitative variables
- Bivariate analysis: quantitative variables
- Time series analysis

HO
GENT

Learning materials

Published on Chamilo!

- Lecture slides
- Textbook: Rajagopalan (2021) *A Python Data Analyst's Toolkit: Learn Python and Python-based Libraries with Applications in Data Analysis and Statistics.* Springer.
- Google Colab (`https://colab.research.google.com`)
- Software (optional!):
  - Python (+libraries)
  - Visual Studio Code
  - Github account, Git client
- Lab assignments with example code (Github)

HO GENT

Teaching methods

- 3 hours per week
- classroom instruction (lecture)
- exercises & lab assignments

HO
GENT

Recommendations for learning

- Attend classes!
- Take notes
- Use effective learning strategies
- Make an effort

HO
GENT

Planning

Wk	Subject
1	Course intro, sampling
2	
3	Univariate analysis
4	
5	Probability, central limit theorem
6	Statistical testing: *z*-test
7	Statistical testing: Student *t*-test
—	*Easter holiday*

HO
GENT

Planning

Wk	Subject
8	Bivariate analysis: χ^2 test, Cramér's V
9	Bivariate analysis: qual. vs. quant. variable two-sample *t*-test, effect size
10	Bivariate analysis: quantitative variables Linear regression
11	Time series analysis
12	Catch-up session (if needed)

HO
GENT

Assessment

Written open book exam, with use of computer. Allowed:

- Slides, course material on Chamilo
- Textbook
- Your notes, solutions to exercises
- Python example code
- Software for statistical analysis (Python, VS Code)

HO GENT

Introduction

HO
GENT

Lies, damn lies & statistics

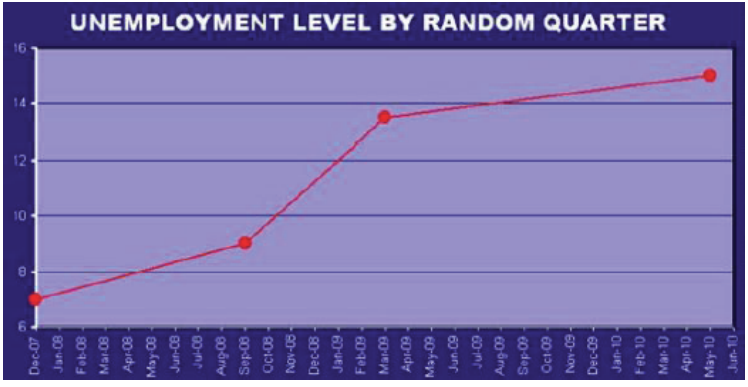
- Data visualisation mistakes and gaffes are
- Media outlets, politicians, special interest groups, shady people on Facebook, …misrepresent or misinterpret objective data to “prove” their point.

HO
GENT

Lies, damn lies & statistics

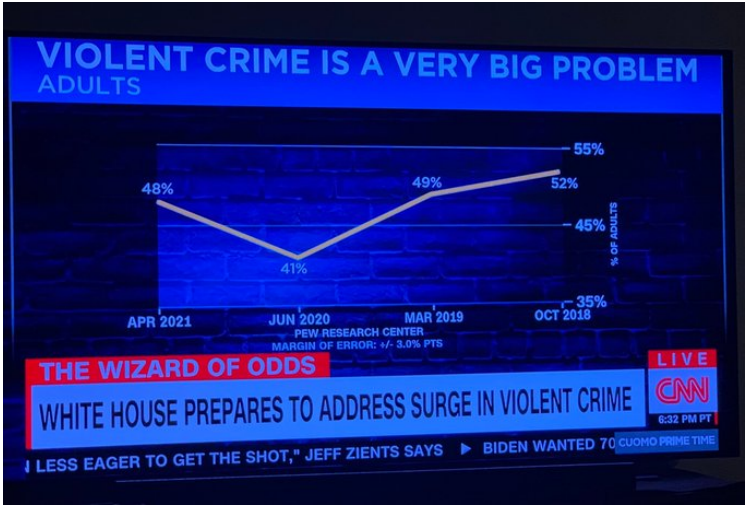
HO GENT

Lies, damn lies & statistics



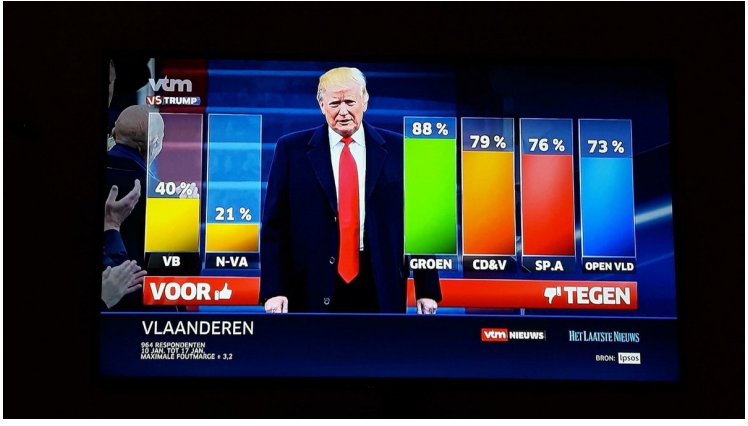
HO GENT

Lies, damn lies & statistics



HO
GENT

Lies, damn lies & statistics

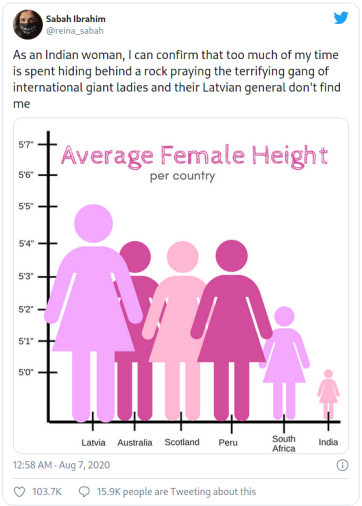


HO GENT

Lies, damn lies & statistics

HO
GENT

Lies, damn lies & statistics



HO GENT

Lies, damn lies & statistics

HO
GENT

Lies, damn lies & statistics

HO GENT

Why learn data science?

- The amount of data is exploding
- Data drives business decisions
- It's important to analyse and visualise data correctly!
- Tools and data are more accessible than ever

HO
GENT