

# **Course Introduction**

## **Data Science**

**Sabine De Vreese   Lieven Smits   Pieter Van Der Helst   Bert Van Vreckem**  
**2024–2025**

**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.

# 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

# Learning materials

Published on Chamilo!

- Course material: Jupyter Notebooks with example code and lab assignments: <https://github.com/HoGentTIN/dsai-labs>
- Lecture slides

Software: choose one of the options below:

- Online: Google Colab (<https://colab.research.google.com>)
- Local:
  - Python (+libraries)
  - Visual Studio Code + Python/Jupyter extensions
  - Github account, Git client

# Teaching methods

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



# Recommendations for learning

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

# 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

# Planning

Wk	Subject
8	Bivariate analysis: $\chi^2$ test, Cramér's V
—	<i>Easter holiday</i>
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)

# Assessment

- Written open book exam
- On your computer
- Assignment = Jupyter Notebook
- Fill out responses in the notebook & hand it in

# Assessment

Open book: what is allowed?

- Course material and slides, learning path on Chamilo
- Textbook (PDF, printed)
- Your personal notes, solutions to exercises
- Software for statistical analysis (Google Colab, Python, VS Code)
- Internet search

# Assessment

Open book: what is not allowed?

- Communication with others (Discord, Messenger, ...)
- Forums, Q&A sites, social media, ...
- Use of LLMs: ChatGPT, Github Copilot, Bing Chat, Gemini, ...
- Earplugs with electronics (e.g. Bluetooth)

**HO  
GENT**

# Introduction

**HO  
GENT**

# Lies, damn lies & statistics

- Data visualisation mistakes and gaffes are common
- 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



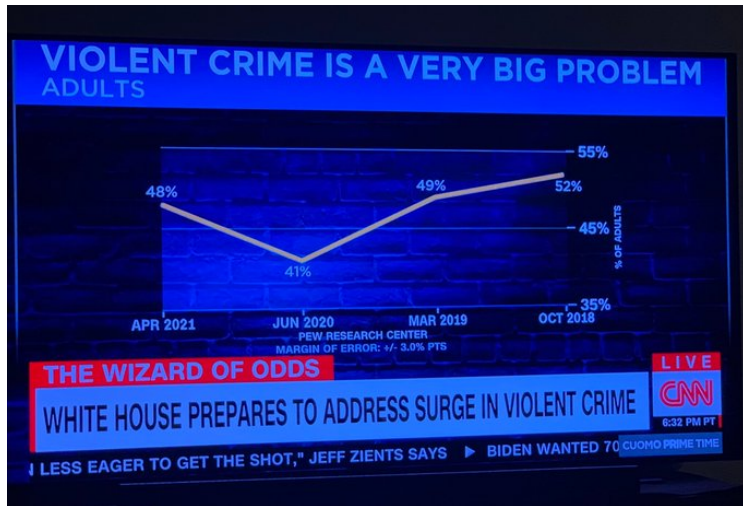
**HO  
GENT**

# Lies, damn lies & statistics



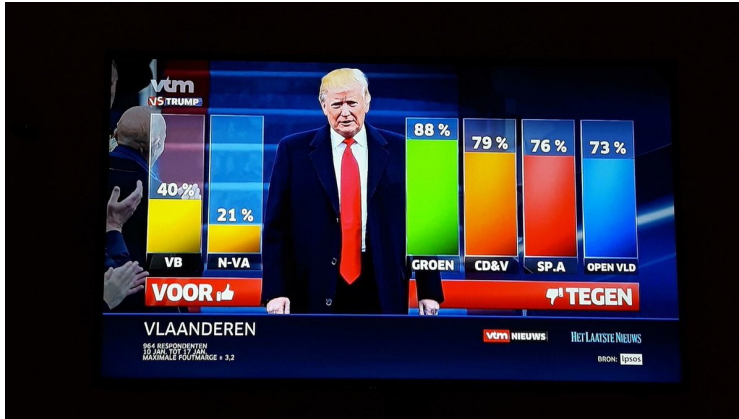
**HO  
GENT**

# Lies, damn lies & statistics



**HO  
GENT**

# Lies, damn lies & statistics



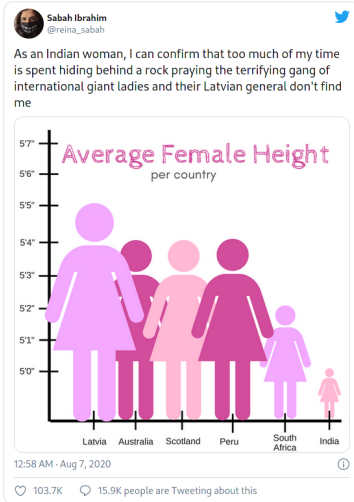
**HO  
GENT**

# Lies, damn lies & statistics



**HO  
GENT**

# Lies, damn lies & statistics



**HO  
GENT**

# Lies, damn lies & statistics



**HO  
GENT**

# Lies, damn lies & statistics

DE  
REDACTIE.BE

21 februari 2014 16:28 DE EN FR

9 °C 139 KM FILE

Het journaal 1 - 21/02/14

Het weer 13.30u

100" Journaal 15u

HOME VIDEOZONE LIVE CENTER

PROGRAMMA'S NIEUWS WAUTERS VS. WAES OOK DAT NOG 60 SECONDEN VK14 OP TV LIVESTREAM MEEST BEKEKEN

**VASTGOED**  
**WOONHUIS € 196 000**  
**APPARTEMENT € 207 500**

**Eigen dak boven hoofd wordt steeds duurder**

Andere afleveringen

Het journaal 1 - 21/02/14

Het journaal L - 20/02/14

Het journaal 1 - 20/02/14

Het journaal L - 19/02/14

**HO  
GENT**



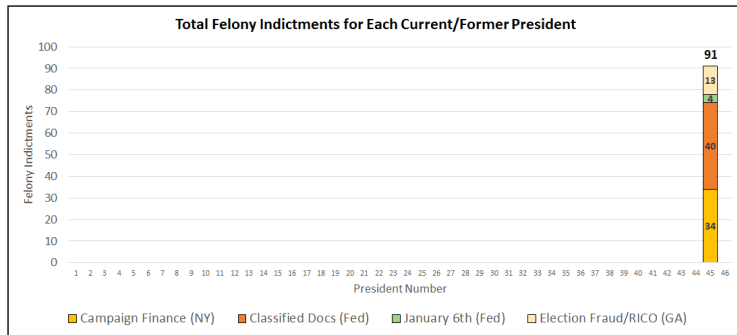
# Lies, damn lies & statistics

On average, US presidents have been charged with  $\approx 1.978$  felony charges.

**HO  
GENT**

# Lies, damn lies & statistics

On average, US presidents have been charged with  $\approx 1.978$  felony charges.



**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