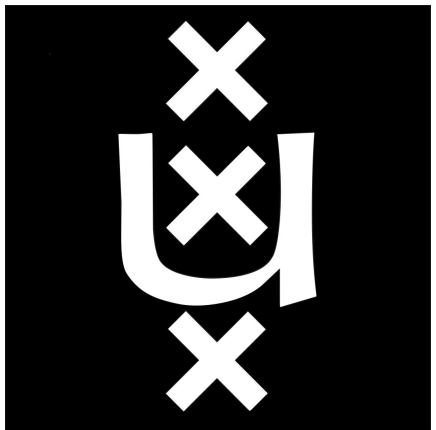
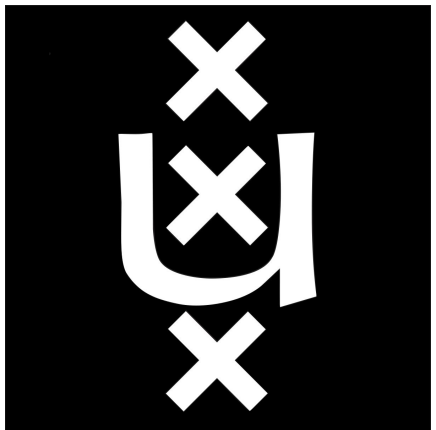
AI for Society

BSc AI 2020/21



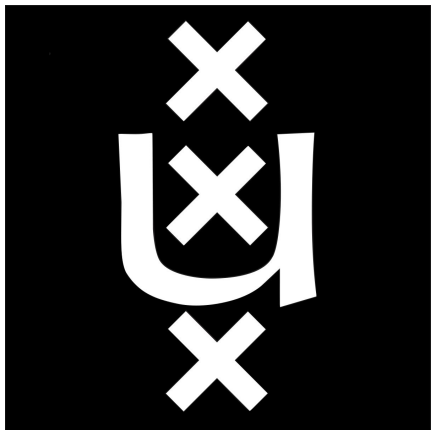
Week 3: Analysing Society and Culture

Tobias Blanke



PART 1: Flipped Classroom

Research-led Teaching



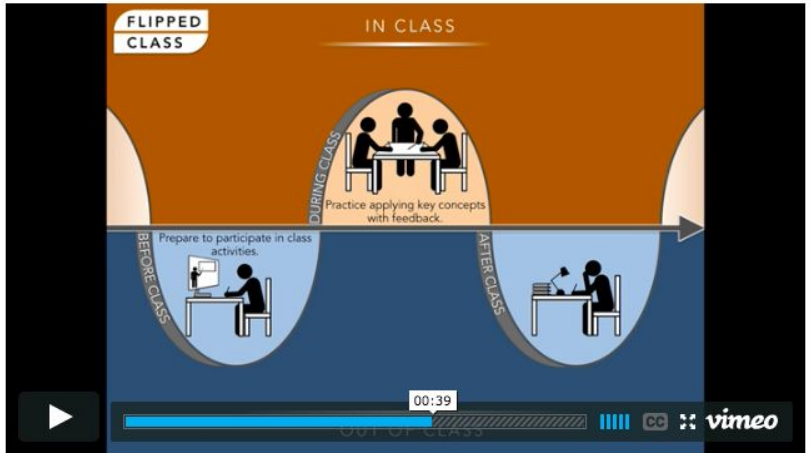
Flipped Classroom

The flipped classroom is a pedagogical model in which the typical lecture and homework elements of a course are reversed.

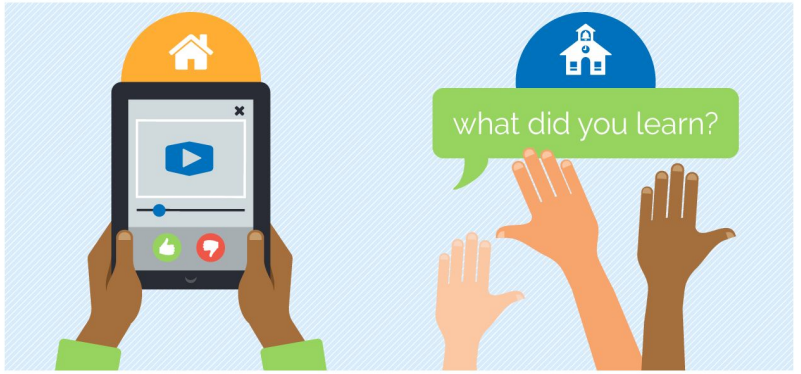
Devoting class time to concepts might give instructors a better opportunity to detect errors in thinking.

An effective flip requires careful preparation.

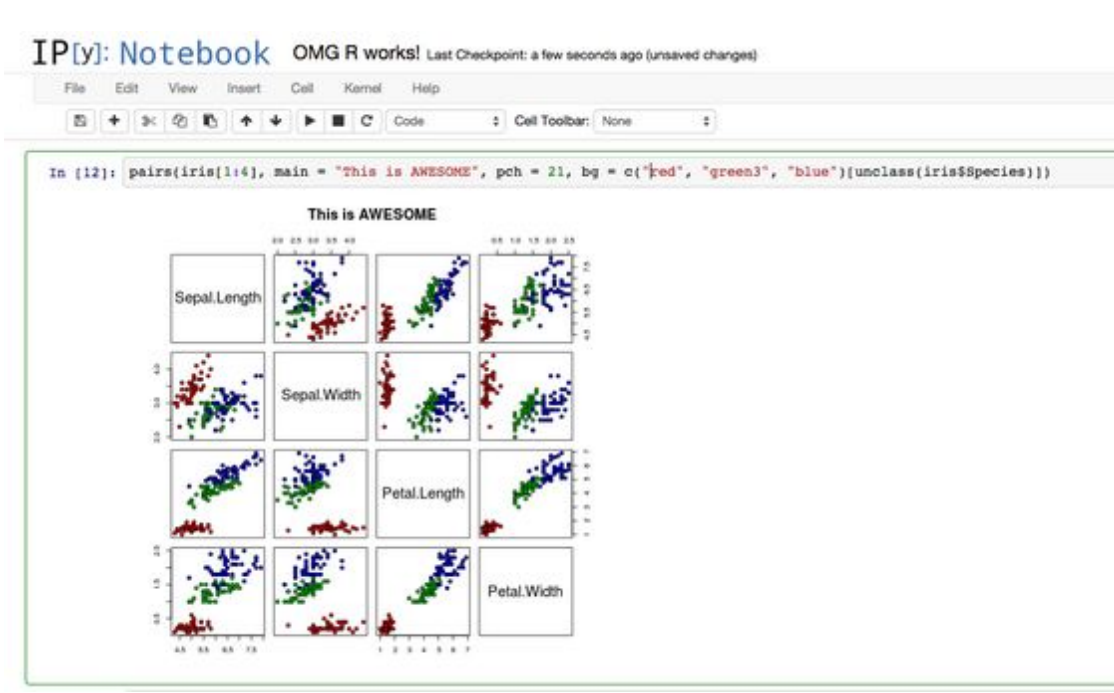
New tools such as notebooks may emerge to support the out-of-class portion of the curriculum



http://facultyinnovate.utexas.edu/teaching/flipping-a-class



Learning by doing: Jupyter notebooks



Notebooks are documents produced by the Jupyter Notebook App which contain both computer code and rich text elements (paragraph, equations, figures, links, etc...).

Notebook documents are both human-readable documents containing the analysis description and the results (figures, tables, etc..) as well as executable documents which can be run to perform data analysis.

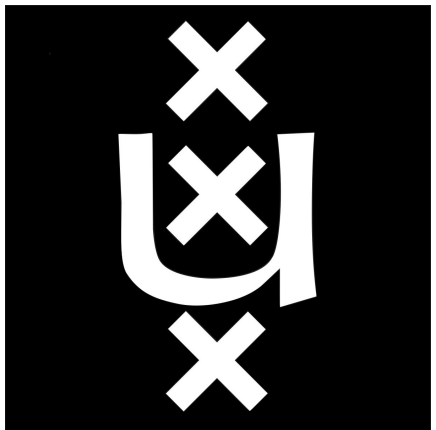
Example: Communities in Digital Culture and Societies

- Research-led teaching:
  - Understanding political communities of practice (CoP) in parliaments
  - Made possible by extensive online publications of parliamentary records (US Congress)
- Methodological perspective
  - Unsupervised Learning: Data exploration through clustering

Unsupervised learning: Who belongs to which group?

PART 2: Exploratory Data Analysis

Because it is important

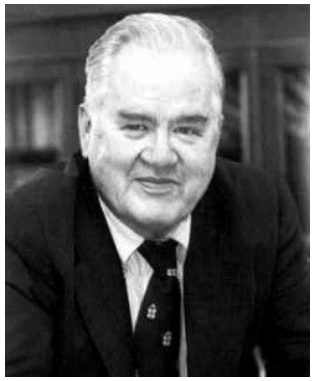


Exploratory Data Analysis

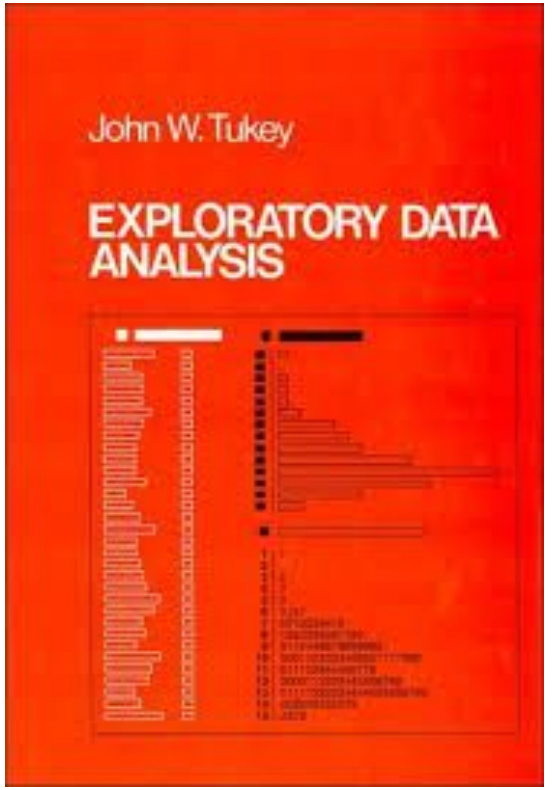
- One of the most important and overlooked part of statistics/data science is Exploratory Data Analysis or EDA
- Allows you to generate hypotheses as well as get a feel for you data

- Descriptive vs Inferential
  - take data from samples and make generalizations about a population

Exploratory Data Analysis
1977 - John Tukey



- Based on insights developed at Bell Labs in the 1960s
- Techniques for visualizing and summarizing data
- What can the data tell us? (in contrast to "confirmatory" data analysis)
- Introduced many basic techniques:
  - 5-number summary, box plots,…



The Trouble with Summary Stats

Set A		Set B		Set C		Set D	
X	Y	X	Y	X	Y	X	Y
10	8.04	10	9.14	10	7.46	8	6.58
8	6.95	8	8.14	8	6.77	8	5.76
13	7.58	13	8.74	13	12.74	8	7.71
9	8.81	9	8.77	9	7.11	8	8.84
11	8.33	11	9.26	11	7.81	8	8.47
14	9.96	14	8.1	14	8.84	8	7.04
6	7.24	6	6.13	6	6.08	8	5.25
4	4.26	4	3.1	4	5.39	19	12.5
12	10.84	12	9.11	12	8.15	8	5.56
7	4.82	7	7.26	7	6.42	8	7.91
5	5.68	5	4.74	5	5.73	8	6.89

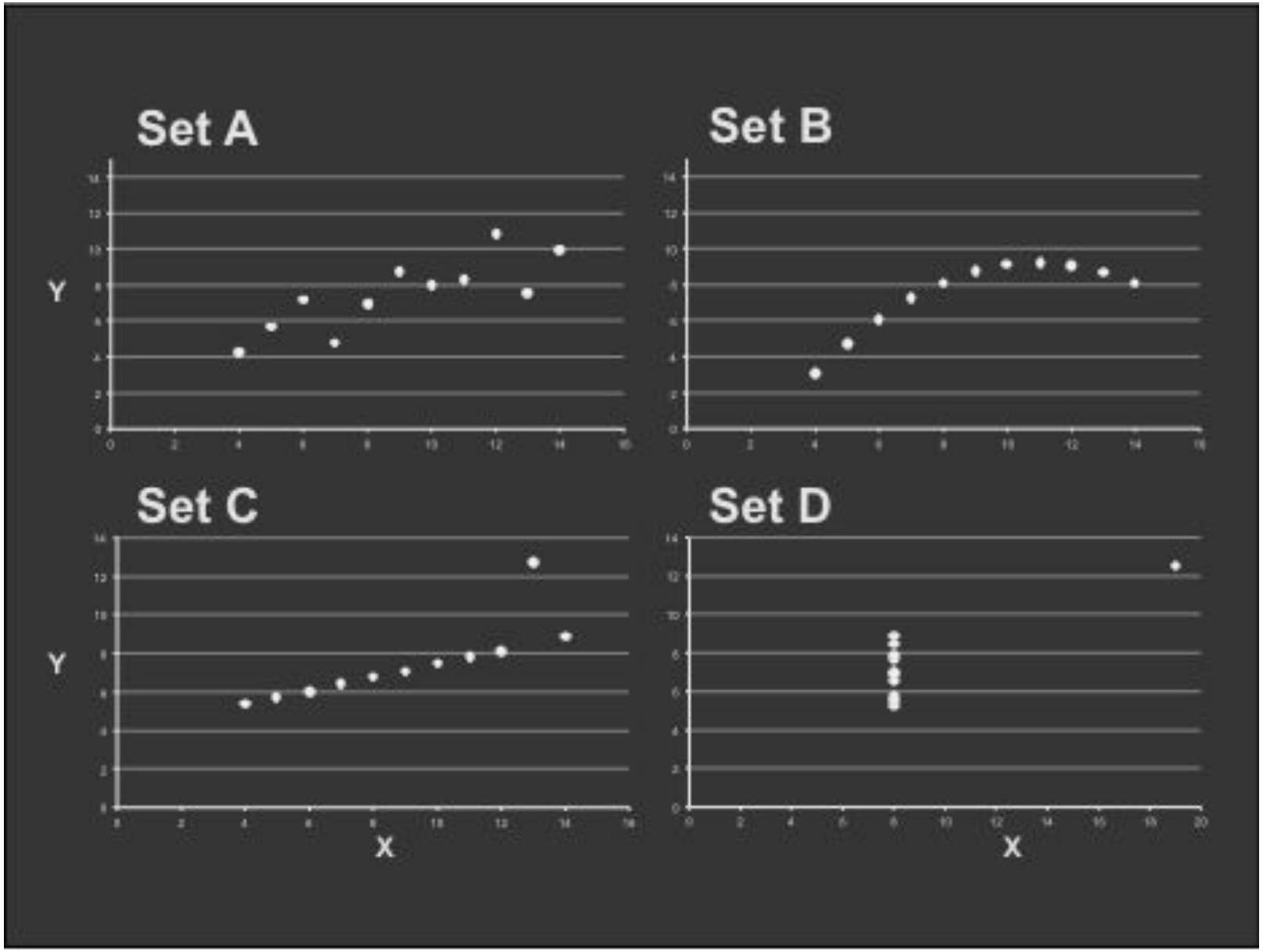
Summary Statistics Linear Regression

$u_X = 9.0$ $\sigma_X = 3.317$ $Y = 3 + 0.5\ X$

$u_Y = 7.5$ $\sigma_Y = 2.03$ $R^2 = 0.67$

[Anscombe 73]

Trouble with summary statistics

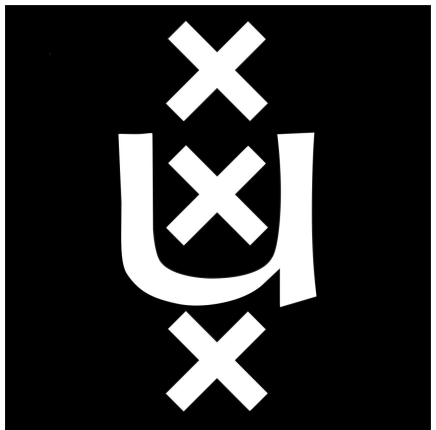


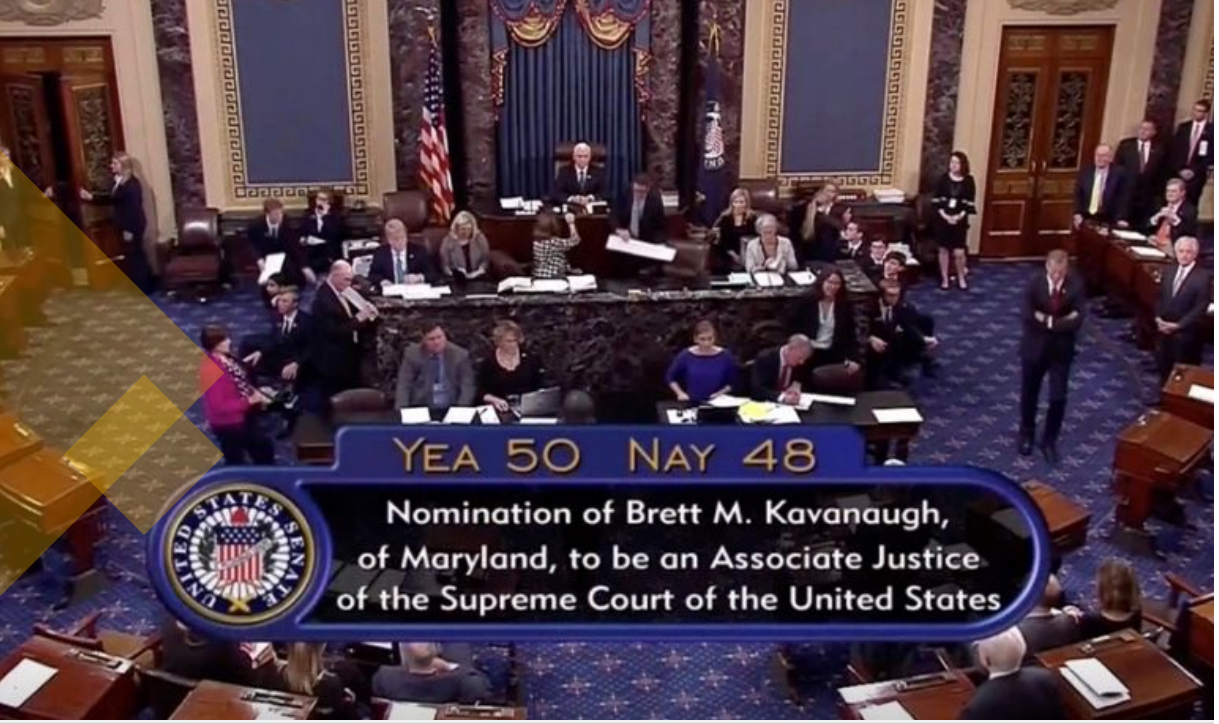
Examples of Research Questions: Descriptive and Inferential

- Descriptive Stats
  - “Who are the most common group of parliamentarians working together?”
- Hypothesis Testing
  - “Is there a difference in number of collaborations across two periods of the Senate?”
- Segmentation/Classification
  - What are the common characteristics of these senators working together?
- Prediction
  - Will new senators become a part of cross-party collaborations?

adapted from Provost and Fawcett, “Data Science for Business”

PART 3: Today and Lab





Cluster 1 (N = 3,376)	Cluster 2 (N = 601)	Cluster 3 (N = 1,036)	Cluster 4 (N = 3,279)	Clust (N = 21
swimming cheerleading cute sexy hot dance dress hair mall hollister abercrombie shopping clothes	band marching music rock	sports sex sexy hot kissed dance music band die death drunk drugs	basketball football soccer softball volleyball baseball sports god church Jesus bible	???
Princesses	Brains	Criminals	Athletes	Basket

Today

- Jupyter notebook – partly complete
- Video recordings that explain the background

- Voting behaviour in the 114th congress
  - Cleaning data
  - Clustering to test party coherence
  - Clustering/visualisation to find out how many parties are there in reality?

- Consumer Choices of Teens
  - Social Network Analysis
  - Clustering result analysis

- Getting data: Web Scraping

Lab

- Catch-up on experiences with Lecture material
- 1st hour Formative Assessment: EDA of voting behaviour in UN Assembly

- Group Assessment: Collaboration Networks in the US Congress
  - How sponsors bills together in the US Congress
  - Association Mining: Apriori Algorithm and Networks

  - Main challenge: Extract the data from US Congress website + other sources

    Always design for error. While scraping drop data that does not compute

    Only download a small proof-of-concept of bills

    Be gentle on their servers! This is only a training exercise

    Split up into smaller teams per group: Data Scraping, Algorithms, etc.

Report

As it will be difficult to download enough bills in order to find out about real relationships, please answer the following question:

“

How do you think association analysis can be applied to other areas in digital society? Name two new case studies, the research question you would like to ask, the kind of data you would like to collect, your methodology and what results you would expect?

“