A Very Brief Introduction to Data Science with a Real-World Example and Data Visualisation

6G7V0026 Principles of Data Science

Luciano Gerber

Senior Lecturer
Department of Computing and Mathematics
Faculty of Science and Engineering
Manchester Metropolitan University

Block 1, Week 1 (20 to 24-Sep-21)

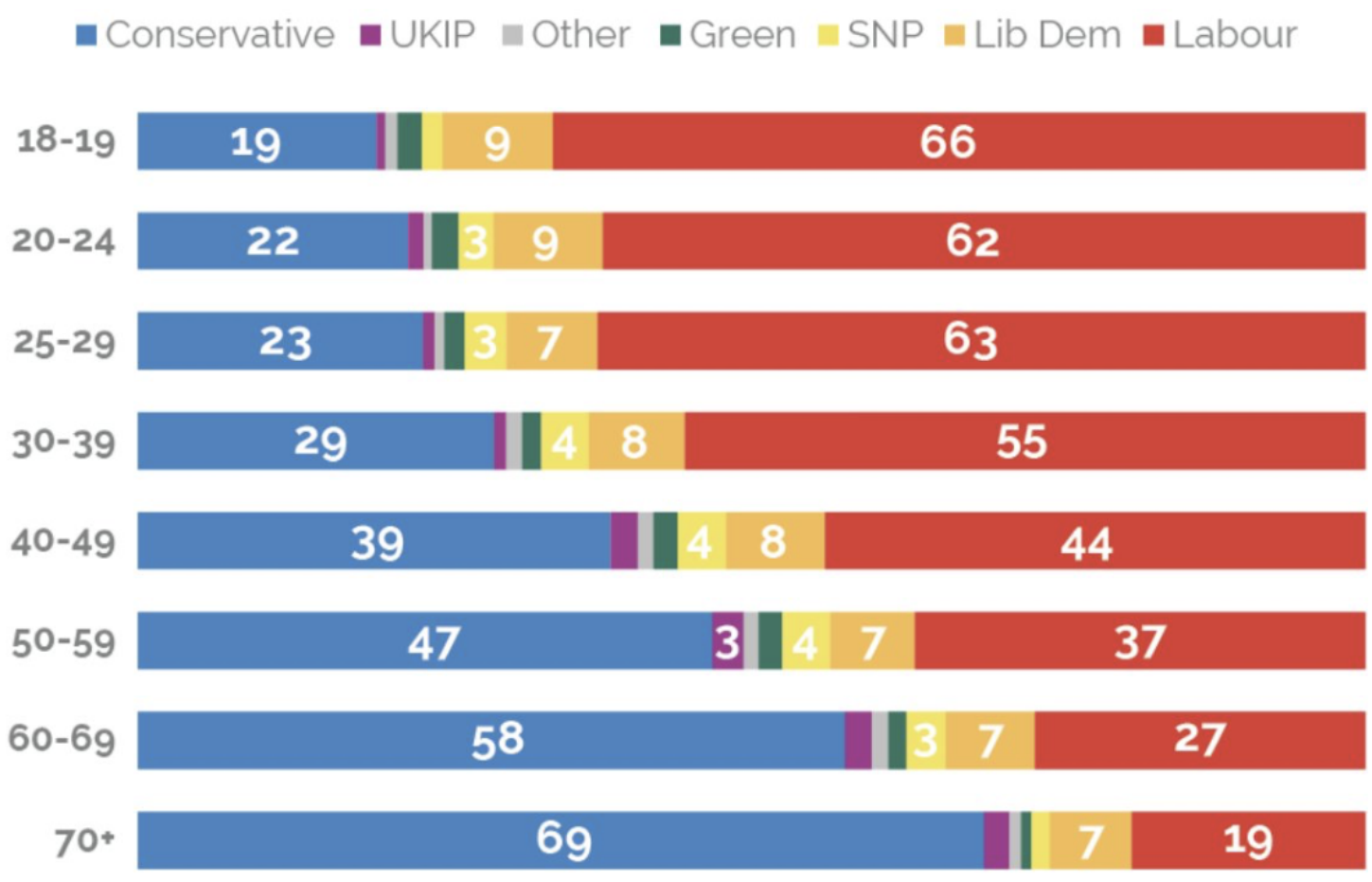
A First Quick Look at Data Science

Some Key Points

- Confluence of **computational** and **statistical** sciences (and thinking).
- In short, we want to **make sense of** (potentially complex and large) **data**. How? For example,
  - Analyse **data distributions** with data visualisation and summarisation (e.g., what is the mean and the standard deviation of patients' ages?).
  - Identify **associations** (e.g., how correlated are income and level of education?).
  - **Compare** groups (e.g., is the mortality rate of one hospital significantly different than another? How does the Brexit vote change according to different age groups?).
  - **Make predictions** (e.g., given socio-economic factors, can we predict whether someone is to be granted a loan? Can we predict the sale value of a residential property, given features such as neighbourhood and number of rooms?)

Vote by age

Based of a survey of 52,615 GB adults about their vote in the 2017 general election



YouGov | yougov.com

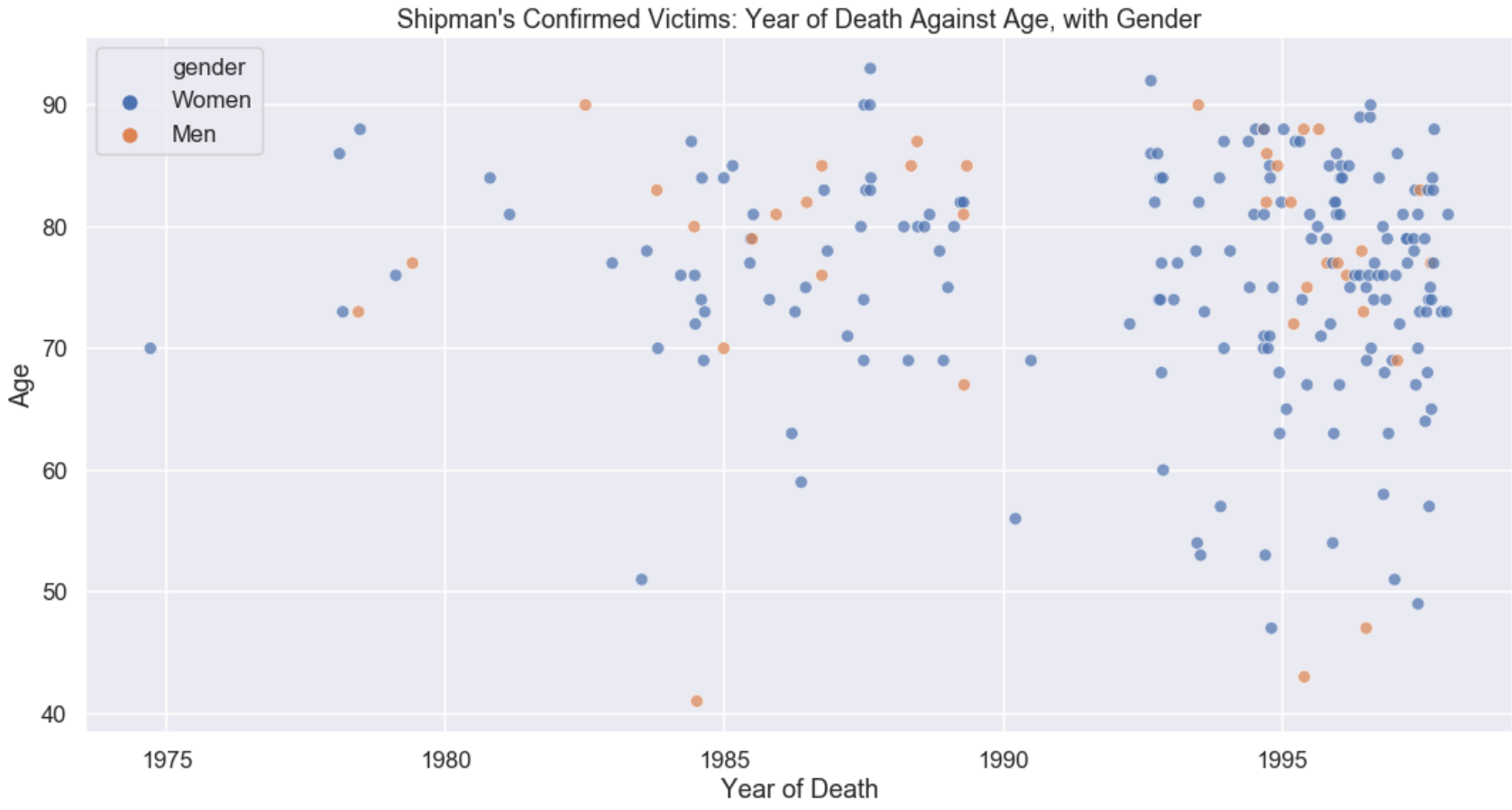
June 9-13, 2017

A Brief, Real-World Example (Comparing Groups, Data Visualisation)

Background

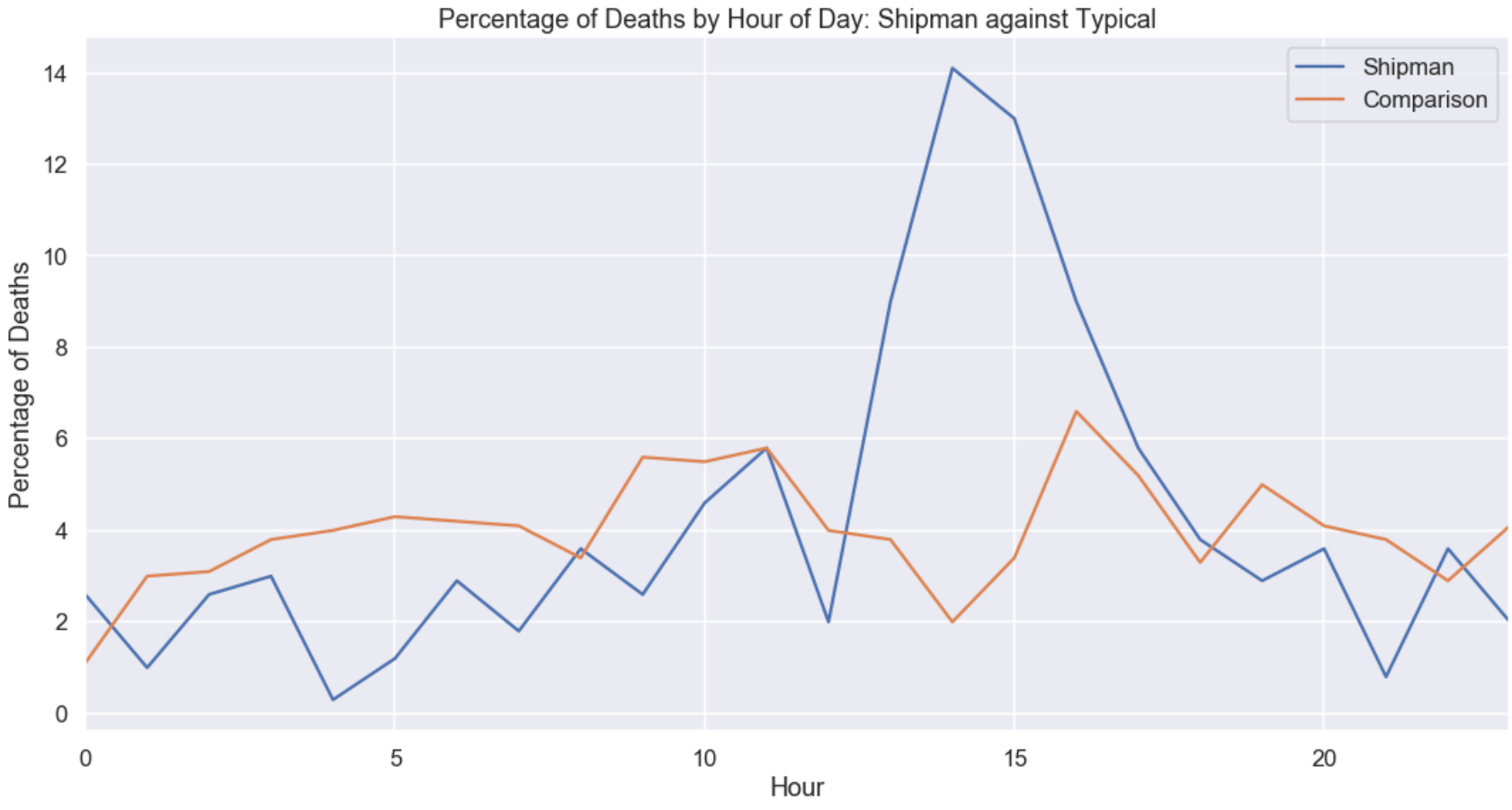
- A taster of Forensic Data Science. **Warning: this example is a little grim**
- It looks at aggregate data on gender, age, and time of death of patients of a convicted serial-killer.
- This is taken from The Art of Statistics. The author, Sir David Spiegelhalter, as a statistician, was asked to give evidence at a public inquiry.
- Harold Shipman was a family doctor working in Bury (Greater Manchester) that, between 1975 and 1998, killed at least 215 of his patients with a opiate overdose.
- Some of the (broad) questions of a data scientist could ask, based on the data, are:
  - What could one learn from Shipman's killings? Any patterns?
  - Could they have been detected/prevented?

Victims by Year of Death, Age, and Gender



- Please use this Padlet to comment on what we can learn from this graph.

Victims by Time of Day, Compared to a Typical Physician



- Please use this Padlet to comment on what we can learn from this graph.