Data Science Course

Introduction

IFT6758, Fall 2020

Introduction: the era of big data

Introduction: the era of big data

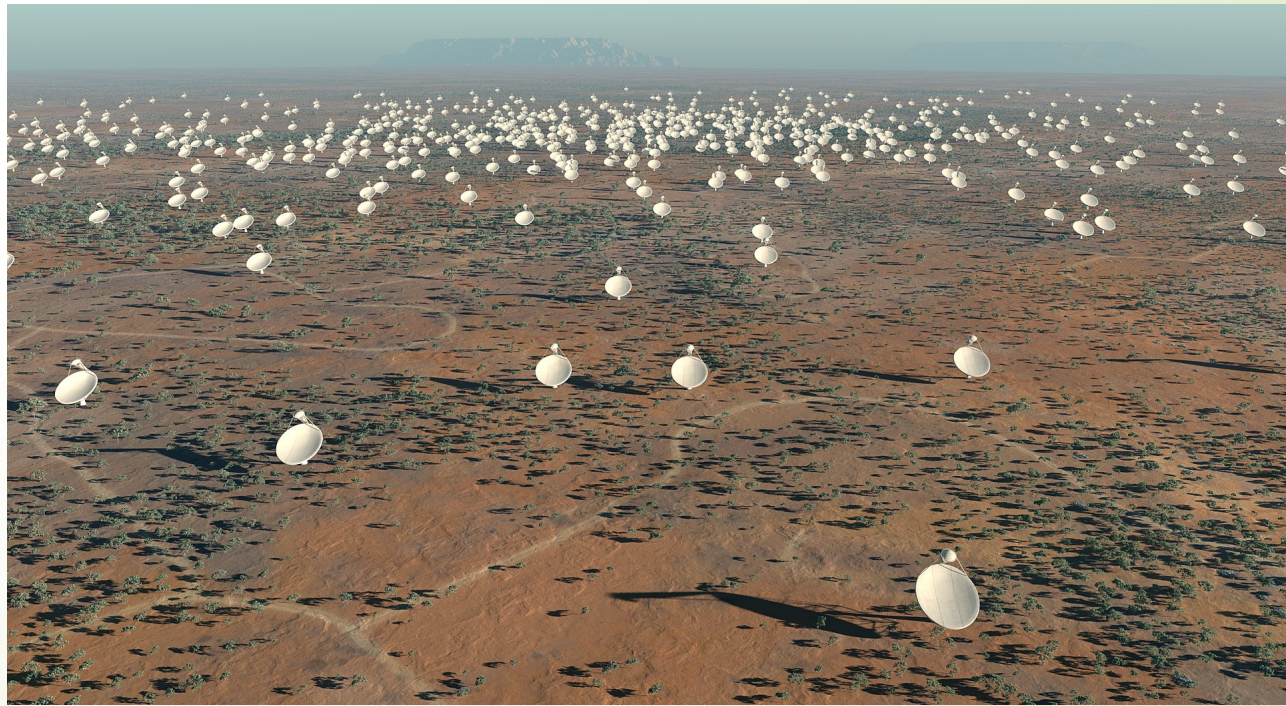
- **Billions of sensors** have been collecting data for decades

Introduction: the era of big data

- **Billions of sensors** have been collecting data for decades
- Data is routinely collected and employed for

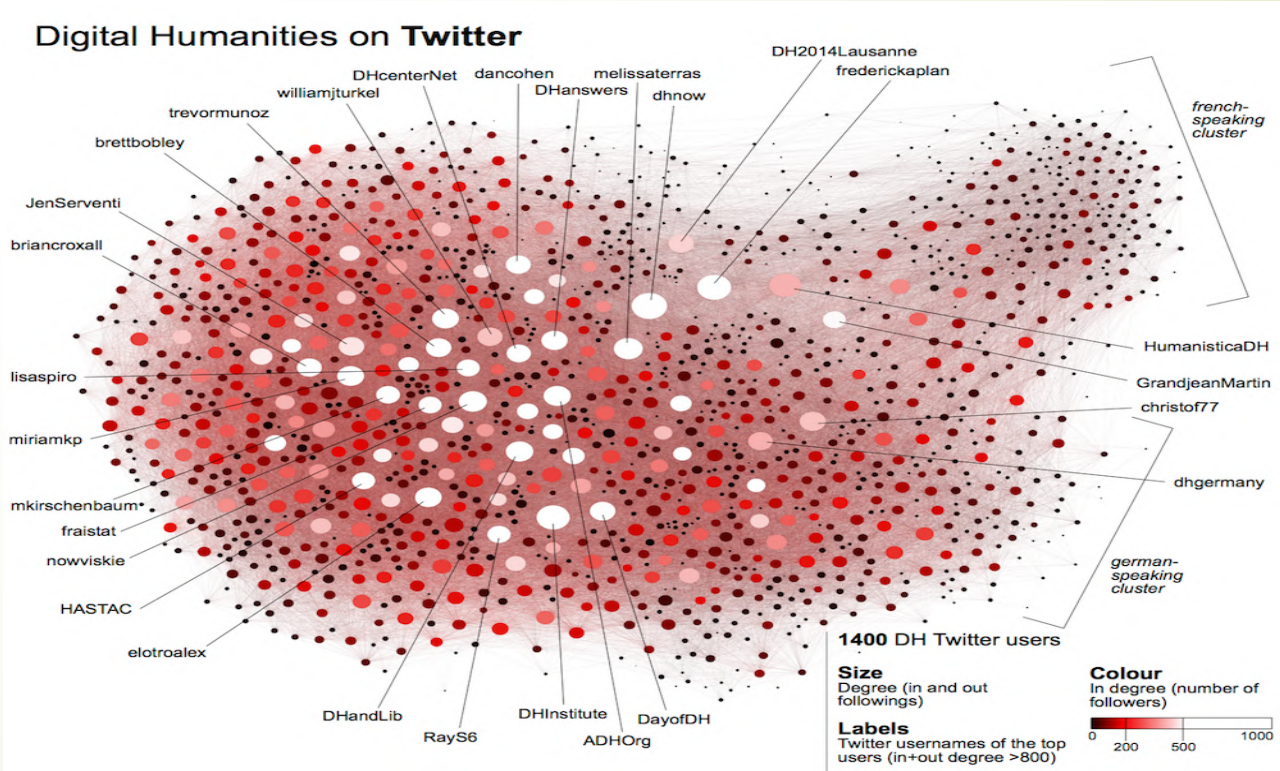
Introduction: the era of big data

- **Billions of sensors** have been collecting data for decades
- Data is routinely collected and employed for
  - Scientific discovery



Introduction: the era of big data

- **Billions of sensors** have been collecting data for decades
- Data is routinely collected and employed for
  - Scientific discovery
  - Commercial purposes



Introduction: the era of big data

- **Billions of sensors** have been collecting data for decades
- Data is routinely collected and employed for
  - Scientific discovery
  - Commercial purposes
  - Data driven society

Introduction: the era of big data

- **Billions of sensors** have been collecting data for decades
- Data is routinely collected and employed for
  - Scientific discovery
  - Commercial purposes
  - Data driven society
- Huge resource of public datasets

Introduction: the era of big data

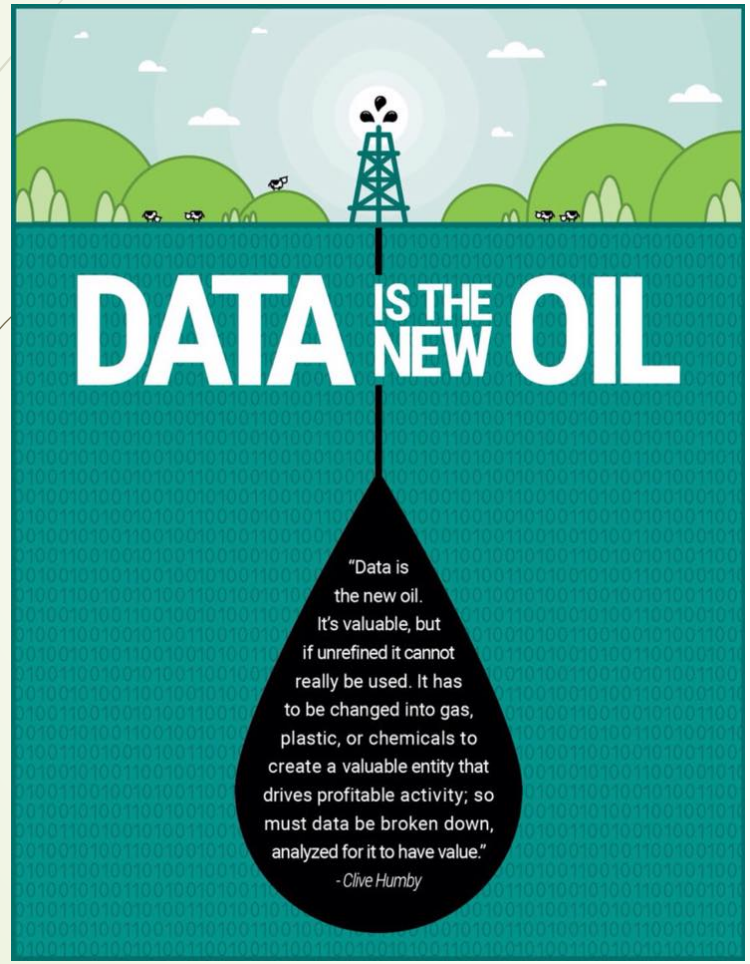
- **Billions of sensors** have been collecting data for decades
- Data is routinely collected and employed for
  - Scientific discovery
  - Commercial purposes
  - Data driven society
- Huge resource of public datasets
  - https://github.com/awesomedata/awesome-public-datasets

Introduction: the era of big data

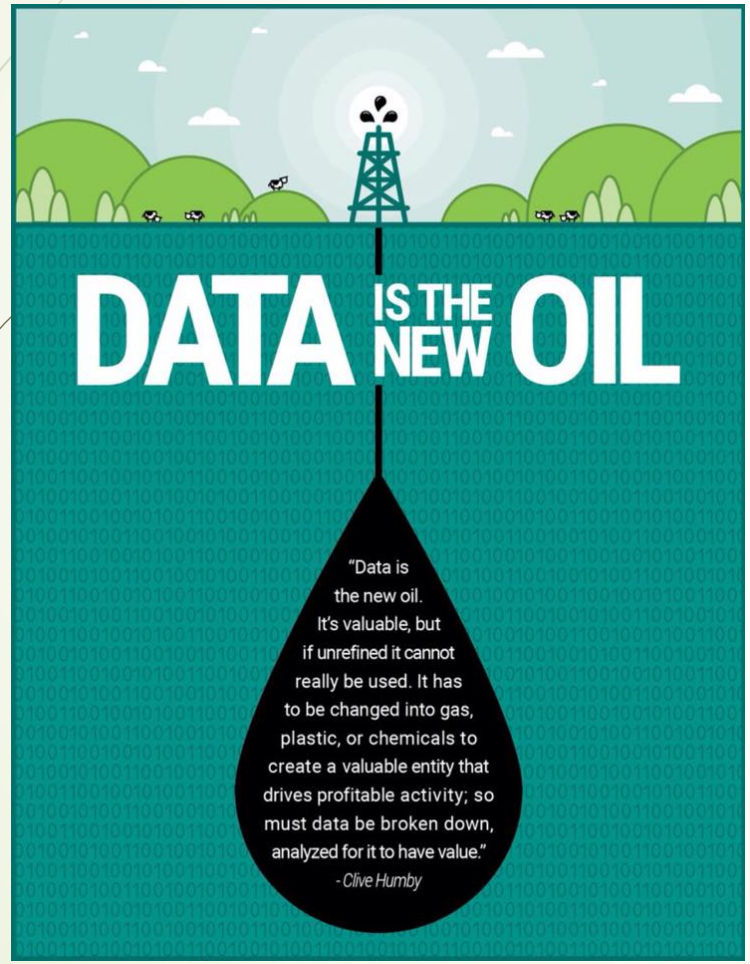
- **Billions of sensors** have been collecting data for decades
- Data is routinely collected and employed for
  - Scientific discovery
  - Commercial purposes
  - Data driven society
- Huge resource of public datasets
  - https://github.com/awesomedata/awesome-public-datasets
- It's important that data scientists work **responsibly** and for **greater good**

Data is the new oil

Data is the new oil

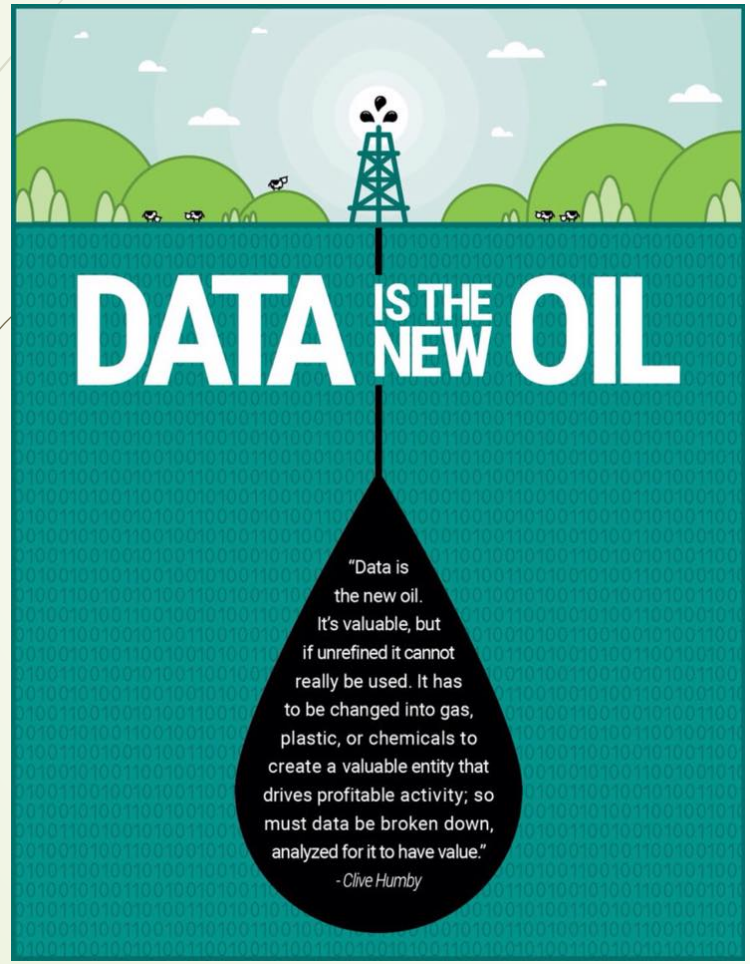


Data is the new oil



Or, is it?

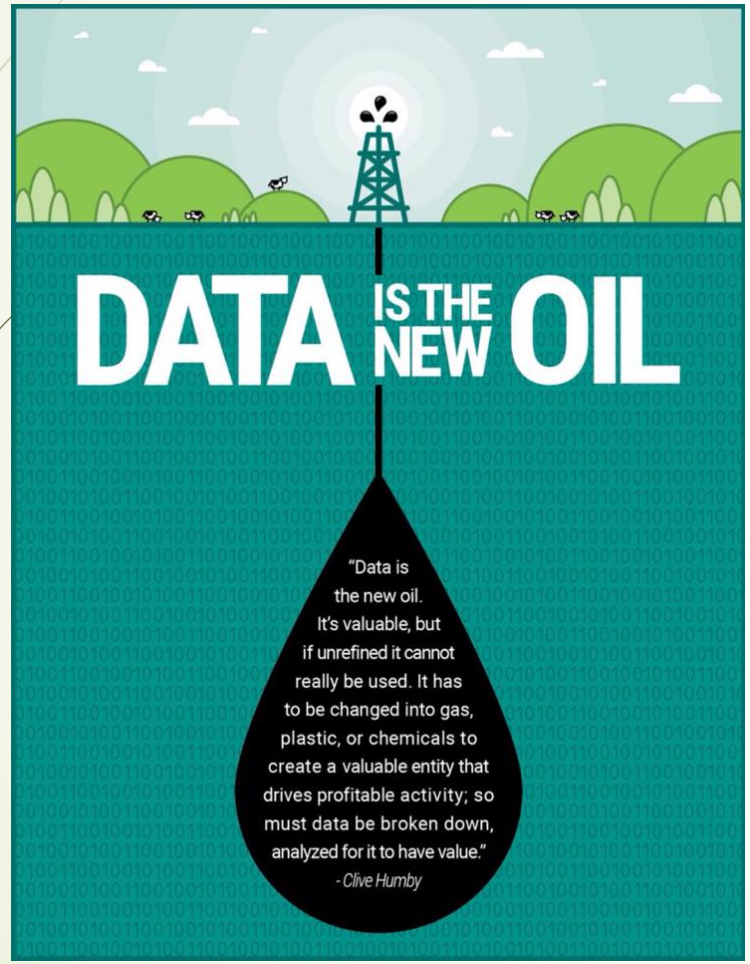
Data is the new oil



Or, is it?

Unlike oil, *when dealing with data, it is far from clear how exactly to turn that data into profits*.

Data is the new oil



Or, is it?

Unlike oil, *when dealing with data, it is far from clear how exactly to turn that data into profits*.

Why data is not new oil

Know the data

Know the data

- **Be comfortable with data**

Know the data

- **Be comfortable with data**
  - Build competence working with multimodal data sets
  - Exposure to the full data science workflow

Know the data

- **Be comfortable with data**
  - Build competence working with multimodal data sets
  - Exposure to the full data science workflow
- **Become a data detective**

Know the data

- **Be comfortable with data**
  - Build competence working with multimodal data sets
  - Exposure to the full data science workflow
- **Become a data detective**
  - Learn to ask good questions
  - Reason about uncertainty, think critically

Know the data

- **Be comfortable with data**
  - Build competence working with multimodal data sets
  - Exposure to the full data science workflow
- **Become a data detective**
  - Learn to ask good questions
  - Reason about uncertainty, think critically
- **Learn responsible data science**

Know the data

- **Be comfortable with data**
  - Build competence working with multimodal data sets
  - Exposure to the full data science workflow
- **Become a data detective**
  - Learn to ask good questions
  - Reason about uncertainty, think critically
- **Learn responsible data science**
  - Understand risks at all stages of data science workflow

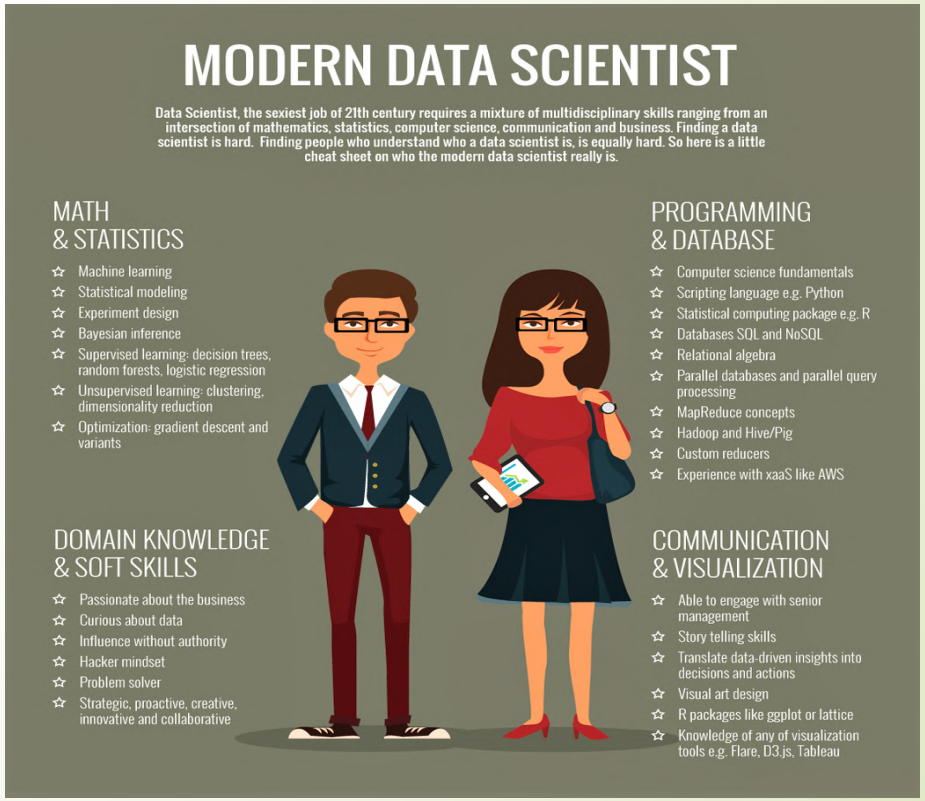
Takeaway from this course

Takeaway from this course

Prepare you to be responsible and competent data scientists

Takeaway from this course

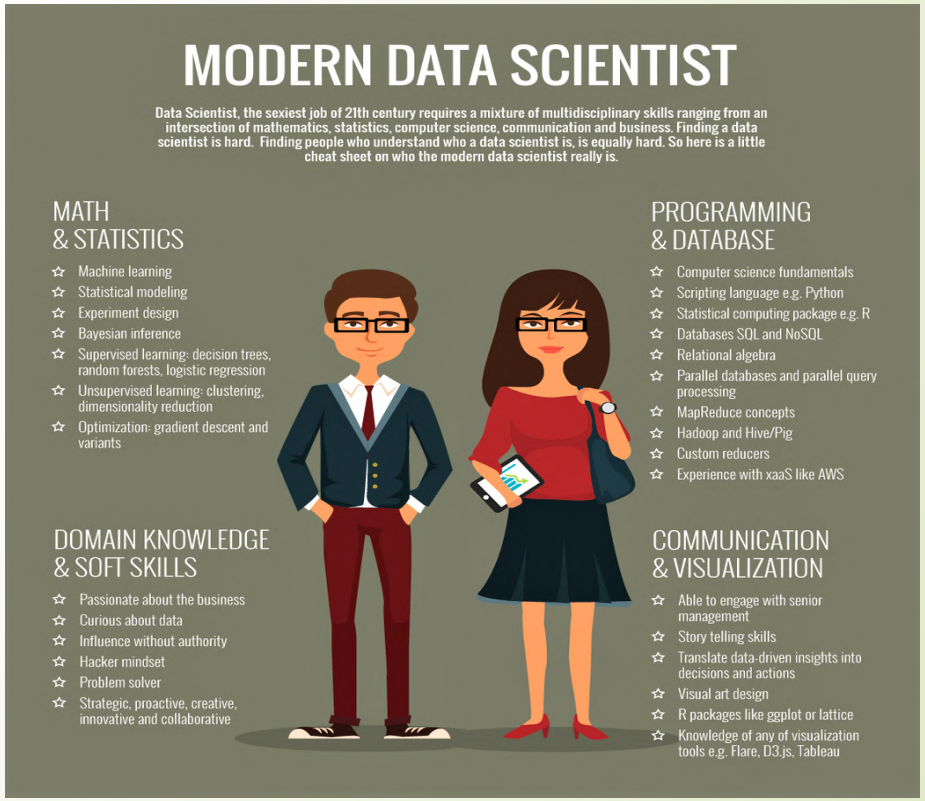
Prepare you to be responsible and competent data scientists



Takeaway from this course

Prepare you to be responsible and competent data scientists

- Learn how to…
  - Apply data science in your own field
  - Work in industry or research
  - Understand data's role in society



Course outline (subject to minor changes)

Course outline (subject to minor changes)

- **Part 1: Summaries and Inferences**
  - Data visualization and transformation
  - Supervised and unsupervised learning from data
  - Inference and model comparison

Course outline (subject to minor changes)

- **Part 1: Summaries and Inferences**
  - Data visualization and transformation
  - Supervised and unsupervised learning from data
  - Inference and model comparison
- **Part 2: Nontabular data**
  - Text and Image
  - Graph mining
  - Special topics: Reinforcement Learning and Deep Learning

Course outline (subject to minor changes)

- **Part 1: Summaries and Inferences**
  - Data visualization and transformation
  - Supervised and unsupervised learning from data
  - Inference and model comparison
- **Part 2: Nontabular data**
  - Text and Image
  - Graph mining
  - Special topics: Reinforcement Learning and Deep Learning
- **Part 3: Frontiers**
  - Advanced inference
  - Ensembling
  - Privacy and explainability

Logistics

- Course **website**: https://jhelum-ch.github.io/DataScience_IFT6758/
- Fill Survey for access to discussion forum
- Grading: **25%** Final project, **30%** HW, **30%** Midterm, **15%** Kaggle competition
- Instructor & TA Office Hours: **TBA**

- Use **StudiUM Forum** to reach out to the instructor
- In case of emergency:
  - **Instructor: Jhelum – chakravj@mila.quebec**
  - **Head TA: Pravish – pravishsainath@gmail.com**
  - **TAs: Yutong Yan, Alexander Peplowski, Harmanpreet Singh, Akshay Singh Rana**

Resources

- Books: Statistical Learning, Python Handbook, Introduction to Data Science
- Online courses: freeCodeCamp , Harvard CS109
- Harvard's Data Science Review