Environmental Health Big Data Analysis

Il-Youp Kwak

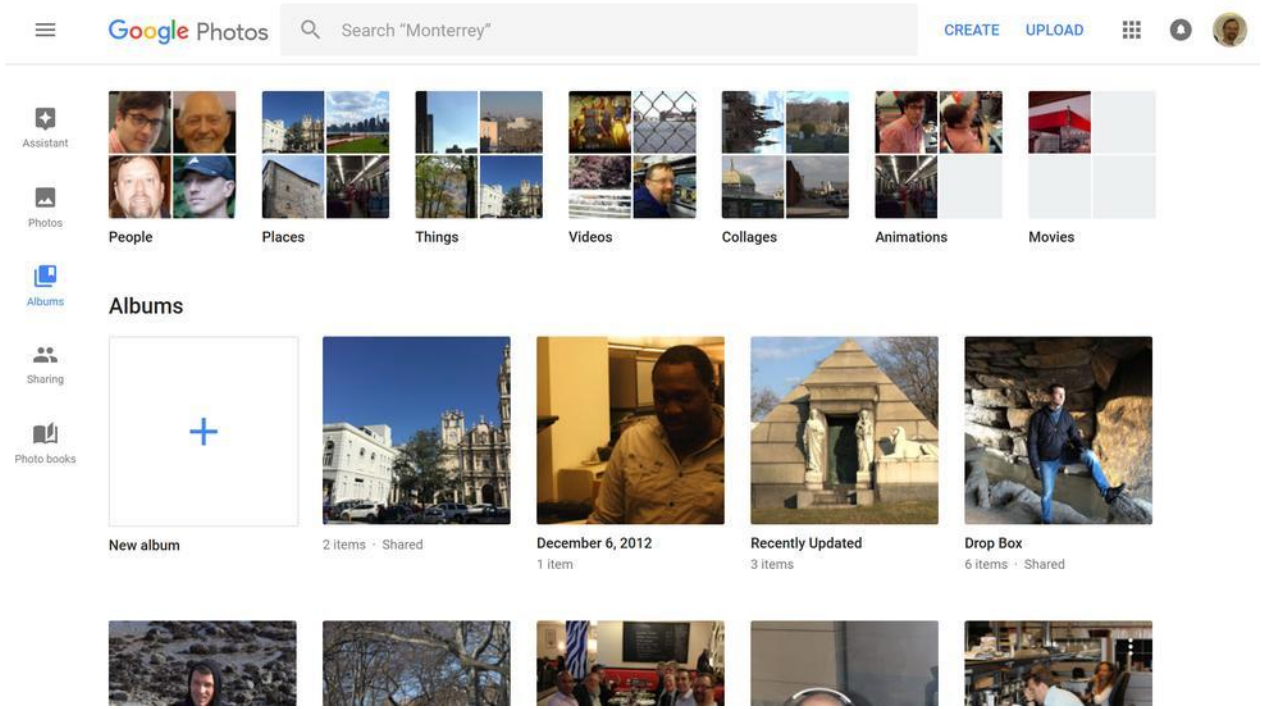
Chung-Ang University

About Me

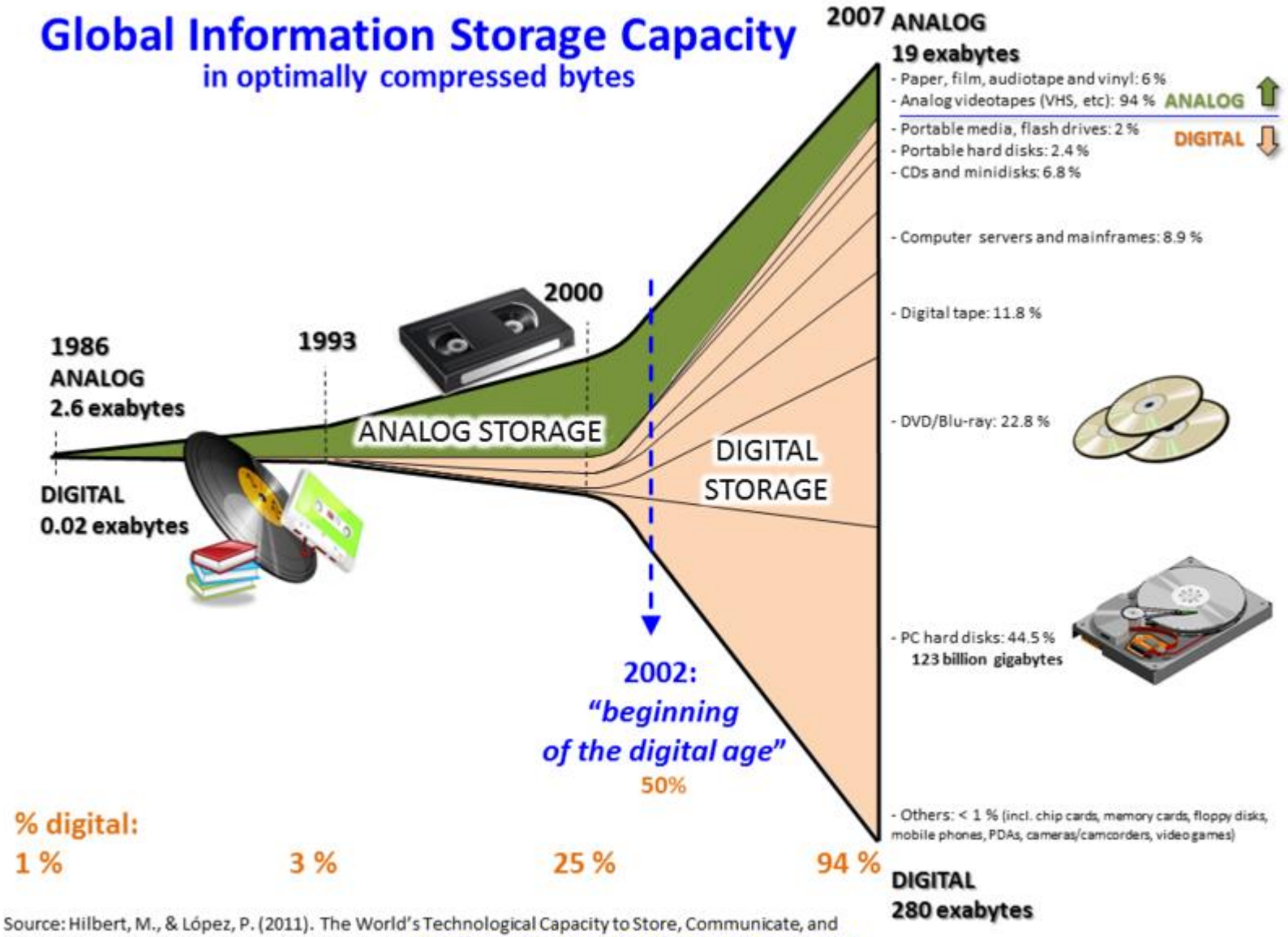
I am an Assistant Professor in the Department of Applied Statistics at Chung-Ang University; research in Statistical Genetics and Deep Learning.

I am eager to design new methods for big and complicated data. I am also excited to develop new useable tools for the method so that those methods can be widely adopted and used by other researchers

Old Photo Books to Google Photos

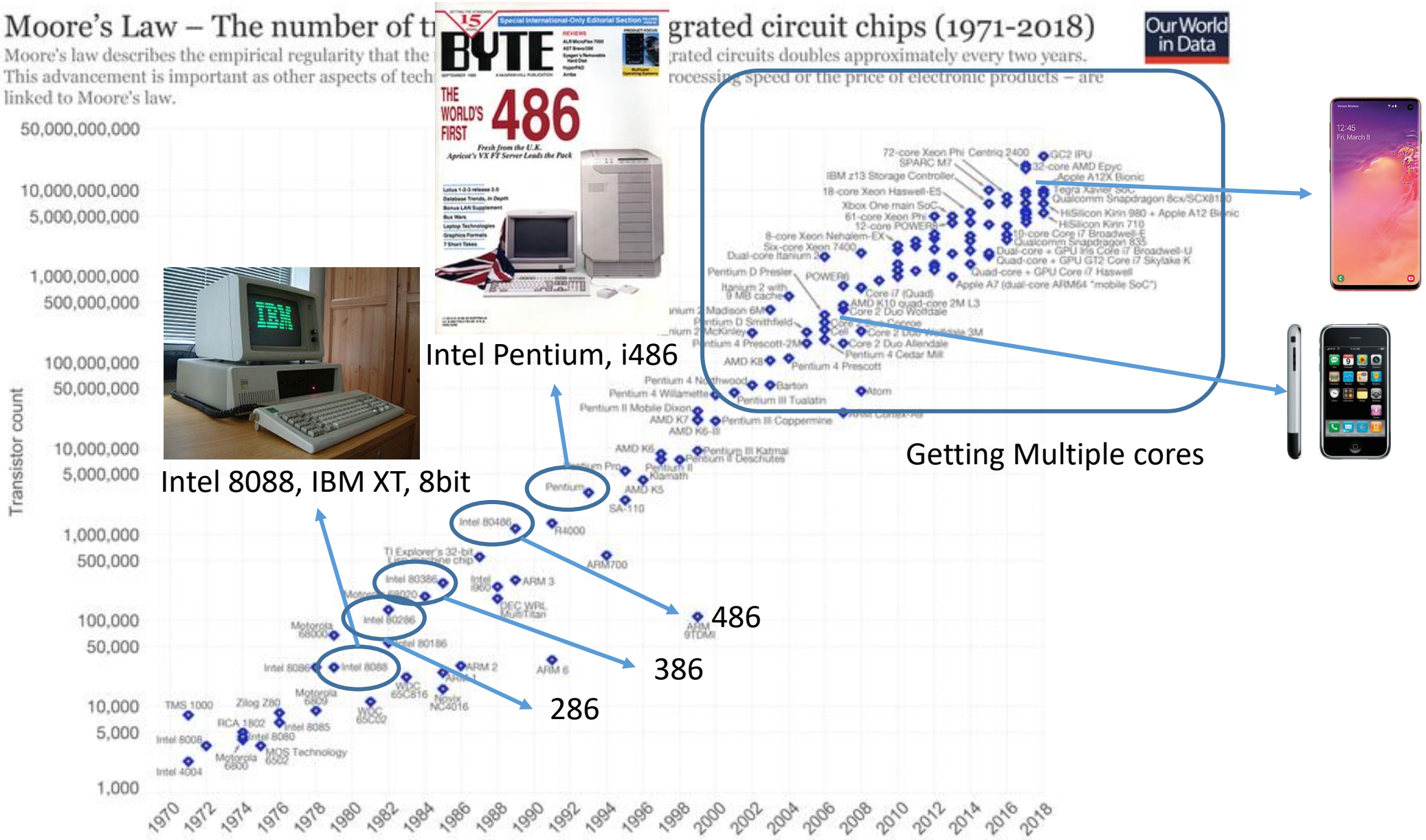


Global Information Storage Capacity is Growing



Source: Hilbert, M., & López, P. (2011). The World's Technological Capacity to Store, Communicate, and Compute Information. *Science*, 332(6025), 60 –65. http://www.martinhilbert.net/WorldInfoCapacity.html

Exponential Laws of Computing Growth



Computers are getting faster,

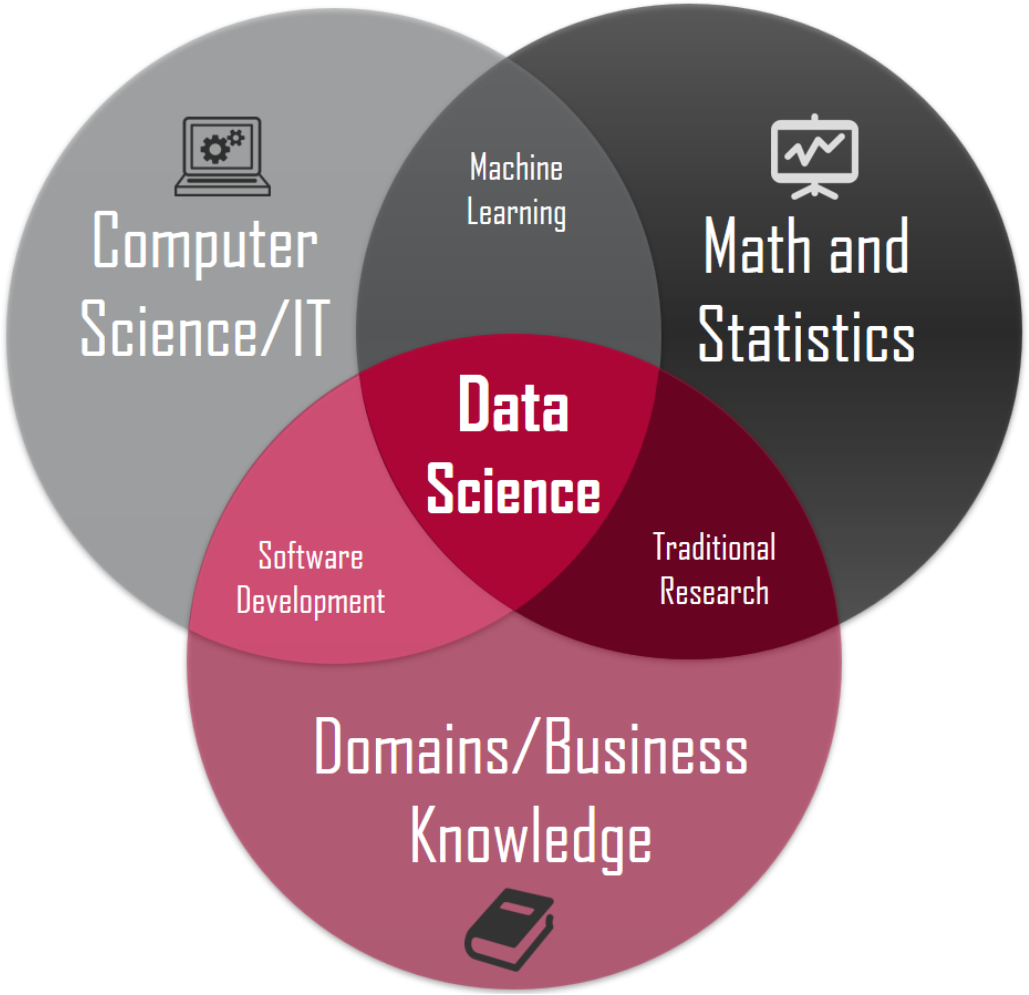
Information storage growing bigger

Opens Big Data Era

A good eco for AI

Growing need for Data Scientists

Core Skills for Data Scientist



Intellectual curiosity

Openness to learning new things

Ability to solve problems in unique way

Passion for innovation

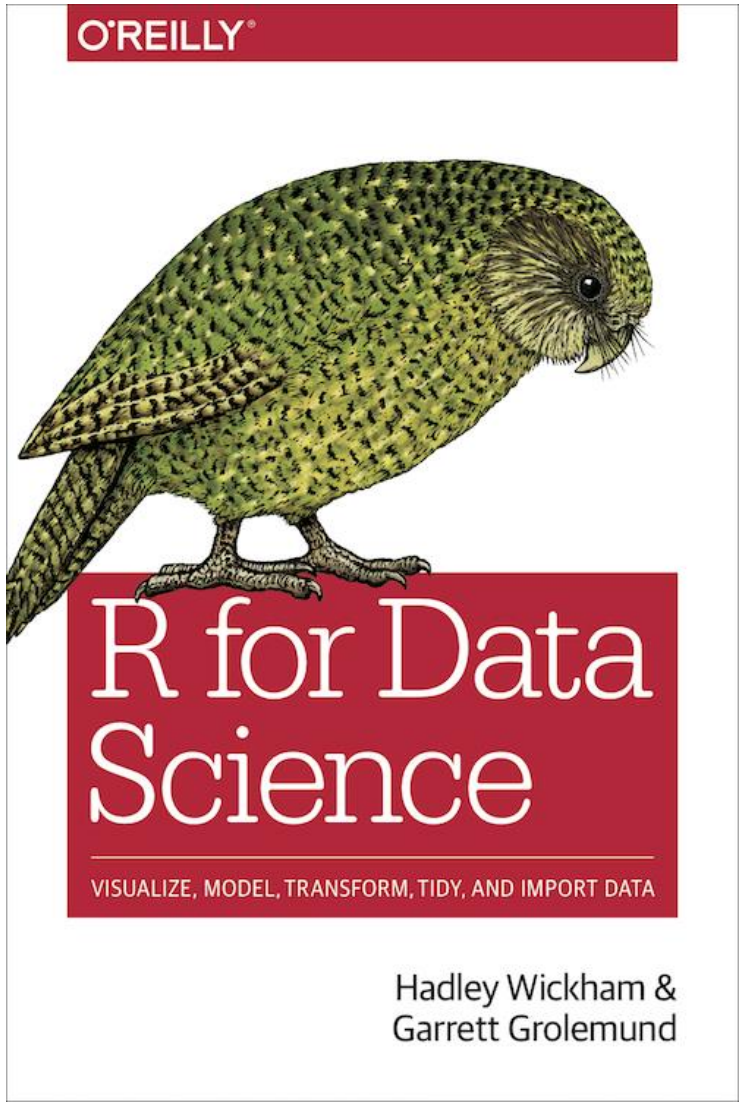
About this course?

Improve Practical Computational Skills (R and Python)

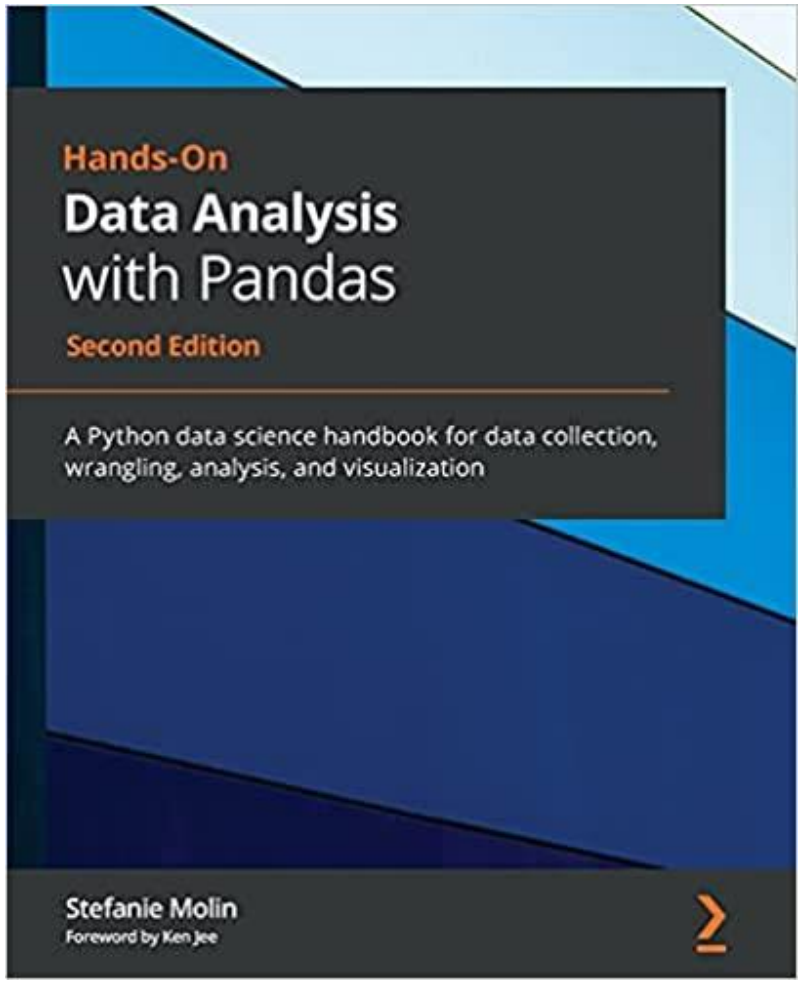
Learn reproducible research techniques (Github, Rmd, Jupyter Notebook)

Team project using big data (Final Project)

Textbooks



https://r4ds.had.co.nz/



https://github.com/stefmolin/Hands-On-Data-Analysis-with-Pandas-2nd-edition

Plans

1	곽일엽	Course Introduction. Colab 실습, Jupyter notebook 실습, Learn Github	
2	곽일엽	Markdown, Rmarkdown, Jupyter notebook	
3	곽일엽	R for Data Science 1 (dplyr)	HW1
4	곽일엽	R for Data Science 2 (tidyr)	
5	곽일엽	R for Data Science 3 (ggplot)	
6	곽일엽	Working with Pandas dataframe	HW2
7	곽일엽	Data Wrangling with Pandas	
8	곽일엽	중간고사	

Plans

9	곽일엽	Aggregating Pandas DataFrames	HW3		
10	곽일엽	Visualizing Data with Pandas and Mat plotlib, Visualizing multi-dimensional data (t-SNE, UMAP)			
11	곽일엽	Plotting with Seaborn and Customizati on Techniques			
12	곽일엽	Financial Analysis – Bitcoin and the St ock Market ;		Zoom, 5 min plan presentation	
13	곽일엽	Rule-Based Anomaly Detection			
14	곽일엽	Getting Started with Machine Learning in Python			
15	곽일엽	Talk from the industry (talk from data scientist?)			
16	곽일엽	팀별 데이터 분석 결과 발표	기말프로젝트 레포트 완성본, 기말 ppt, 개인 제출자료 제출	기말고사는 팀별 발표로 대체. 각 팀별로 데이터 분석 및 결과를 약 15분간 발표 진행	

Discussion (Team Project)

- Each team will have 3~4 members (4~5?)
- Pick any data
  - https://datahub.io/search
  - https://datasetsearch.research.google.com/
  - https://registry.opendata.aws/
  - Etc
- Zoom session for team selection (when?)

Thank you! ☺