Computer Science

CSC380: Principles of Data Science

Wrapup

Prof. Jason Pacheco

TA: Enfa Rose George

TA: Saiful Islam Salim

Outline

- Data Science Ethics and Fairness
- Course Recap
- Additional Resources
- Final Exam Overview

Outline

- Data Science Ethics and Fairness
- Course Recap
- Additional Resources
- Final Exam Overview

Data Science Ethics

The movement to hold AI accountable gains more steam

First-in-US NYC law requires algorithms used in hiring to be "audited" for bias.

KHARI JOHNSON, WIRED.COM – 12/5/2021, 6:10 AM

[Article Link](#)

As Data Science / AI / ML become more standard, we need to address fairness and ethics…

State of Michigan's mistake led to man filing bankruptcy

Paul Egan Detroit Free Press

The secret bias hidden in mortgage-approval algorithms

By EMMANUEL MARTINEZ and LAUREN KIRCHNER/The Markup August 25, 2021

Senators Question Regulators About Tenant Screening Oversight

Facebook's race-blind
around hate speech
e expense of Black
v documents show

- Washington Post

ExamSoft's remote bar exam sparks privacy and facial recognition concerns

- Venture Beat

Data Science Ethics

- NYC adopted law requiring audits of algorithms used in hiring
- White house proposes an AI bill of rights to disclose when AI makes decisions with societal impact
- EU lawmakers require inspection of AI deemed high-risk
- Analysis of automated hiring software found to be biased to appearance, software program used to create resume, accent, or whether applicants have a bookshelf in the background
- Photo ID software works well for white men—black women, not so much

Data Science Ethics

Aspects of ethics include…

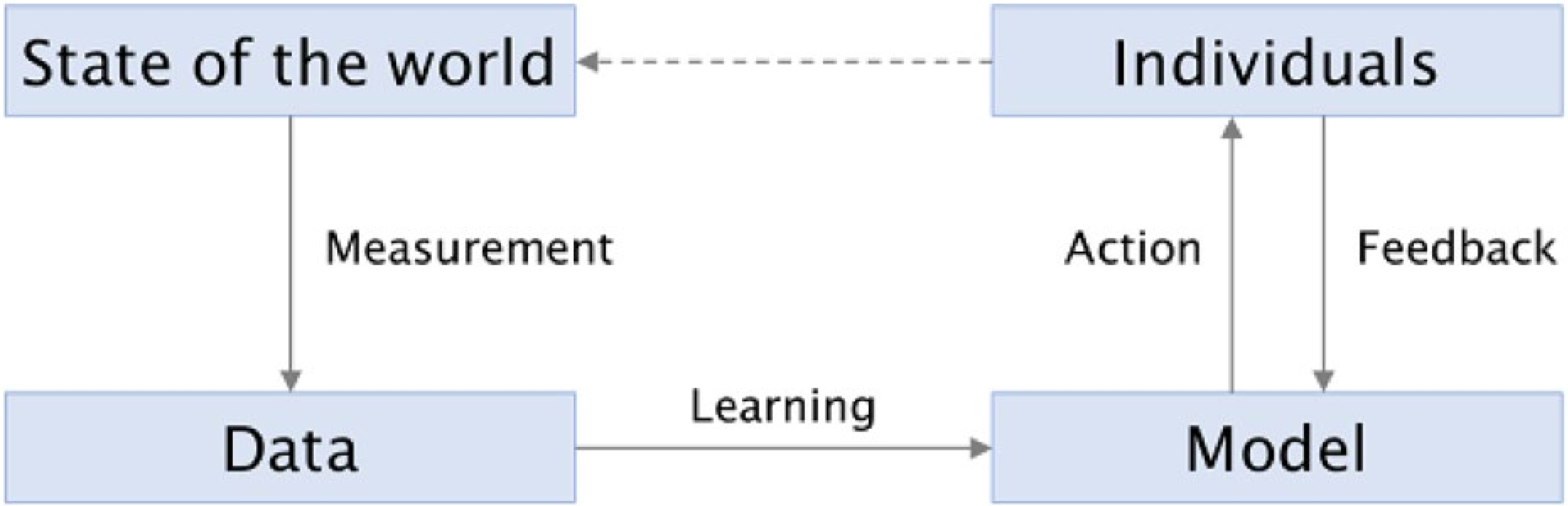
Security Who has access to the data?

Privacy Can data be used to identify individuals?

Fairness Are predictions biased across groups?

Transparency Do users know what they are consenting to? Are model decisions interpretable?

Impacts of Data Science



It is rare for data science to *not* involve people in some way

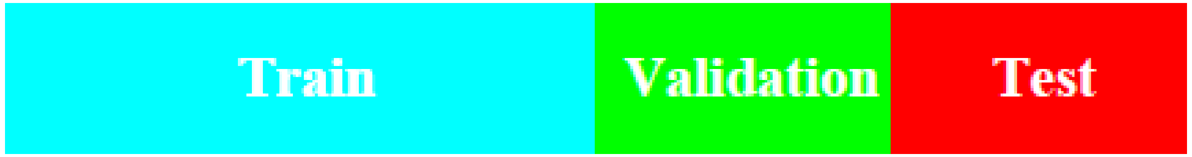
- Of the top 30 recent Kaggle competitions, 14 involve making decisions that directly affect people
- An additional 5 have obvious indirect affect on people
- Only 9 had no obvious impact on people

[Source: Brocas et al. “Fairness and ML”]

Data Science Fairness

Fairness issues can arise from biases in the data…

- Are there observable biases in the data?
- Can we correct for them?



- Differences in the distributions of training / test data?
- Can we detect these differences and avoid / correct them?

Training data reflect disparities, distortions, and biases from the real world and measurement process…

For each model a data scientist should ask… Does learning the model preserve, mitigate, or exacerbate these disparities?

Example Machine translation “She is a doctor” reverse translates to “He is a doctor” in many languages due to data biases.

Data Science Ethics

A real-live example of dataset bias…

https://translate.google.com/

Exhibits gender bias in many languages…

…largely the result of using highly-parameterized neural networks with inadequate training data

Assessing Gender Bias in Machine Translation – A Case Study with Google Translate

Marcelo Prates MORPRATES@INF.UFRGS.BR
Pedro Avelar PEDRO.AVELAR@INF.UFRGS.BR
Luis C. Lamb LAMB@INF.UFRGS.BR
Federal University of Rio Grande do Sul

Abstract

Recently there has been a growing concern in academia, industrial research labs and the mainstream commercial media about the phenomenon dubbed as *machine bias*, where trained statistical models – unbeknownst to their creators – grow to reflect controversial societal asymmetries, such as gender or racial bias. A significant number of Artificial Intelligence tools have recently been suggested to be harmfully biased towards some minority, with reports of racist criminal behavior predictors, Apple's Iphone X failing to differentiate between two distinct Asian people and the now infamous case of Google photos' mistakenly classifying black people as gorillas. Although a systematic study of such biases can be difficult, we believe that automated translation tools can be exploited through gender neutral languages to yield a window into the phenomenon of gender bias in AI.

In this paper, we start with a comprehensive list of job positions from the U.S. Bureau of Labor Statistics (BLS) and used it in order to build sentences in constructions like "He/She is an Engineer" (where "Engineer" is replaced by the job position of interest) in 12 different gender neutral languages such as Hungarian, Chinese, Yoruba, and several others. We translate these sentences into English using the Google Translate API, and collect statistics about the frequency of female, male and gender-neutral pronouns in the

Data Science Fairness

Example We are building a system to screen mortgage applications. Suppose we collect training data from two demographic groups: 85% White and 15% Black

- Predictive accuracy on the held-out validation set is 95%
- Only 5% error
- Should we sign off on the system as good?

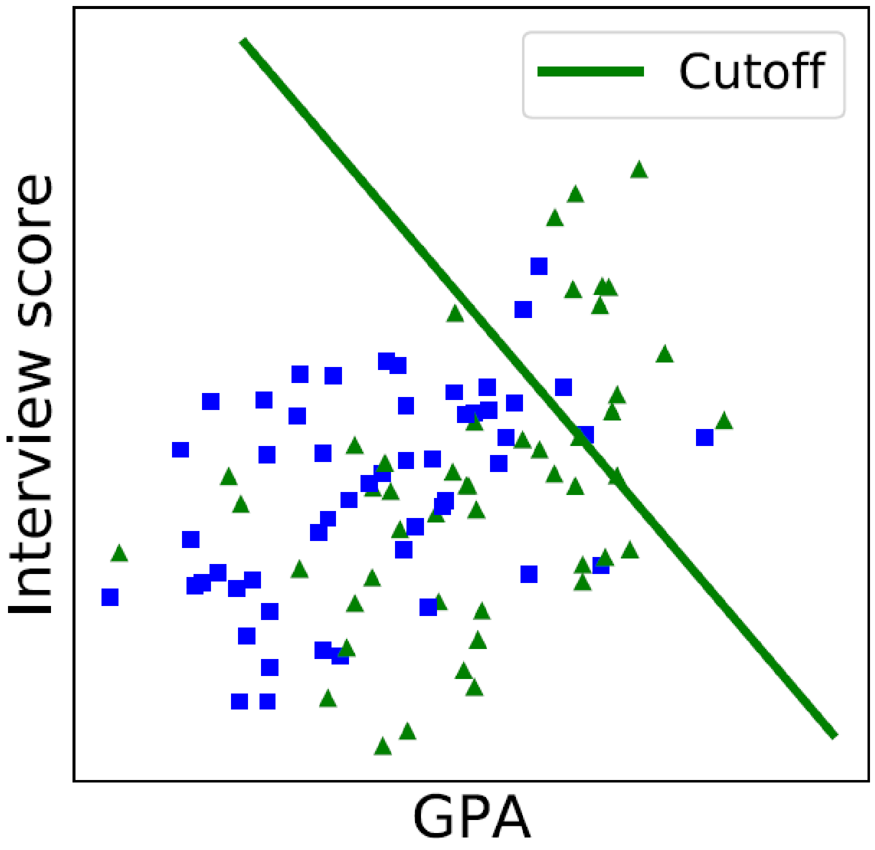
With 5% total error can have up to 66% error on the underrepresented group. Need to report error by each group and aim for 95% accuracy in any group.

Data Science Fairness

Example You are building a system for college admissions based on GPA and interview score (obviously a toy example)

- Fit a least squares regression model
- Model does not account for two demographic groups (blue / green)
- Does this make it fair? (fairness-as-blindness)

Admission rate much lower for blue cohort

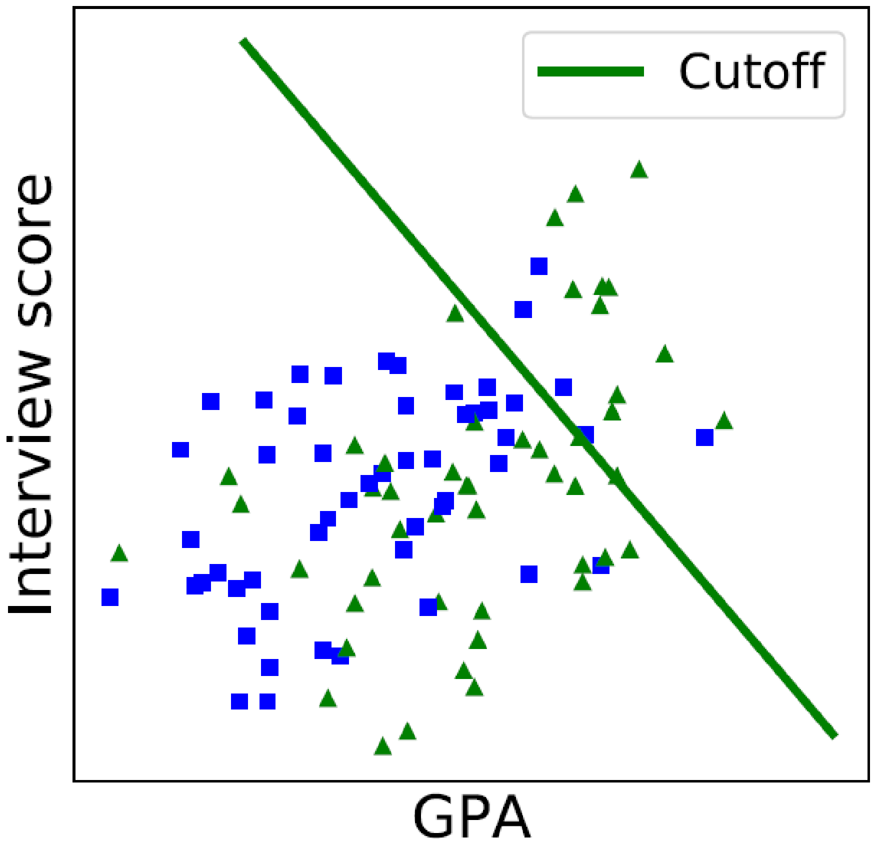


[Source: Brocas et al. “Fairness and ML”]

Data Science Fairness

How to address this behavior?

- GPA correlates with group—omit it as a predictor?
  - Would dramatically impact accuracy
- Pick separate cutoffs (fit separate model) for each group
  - No longer blind to demographics
  - What is the goal for picking cutoffs? Same admission rates?
- Could optimize for diversity among selected candidates
  - Measuring similarity is non-trivial



[Source: Brocas et al. “Fairness and ML”]

Classification Fairness Criteria

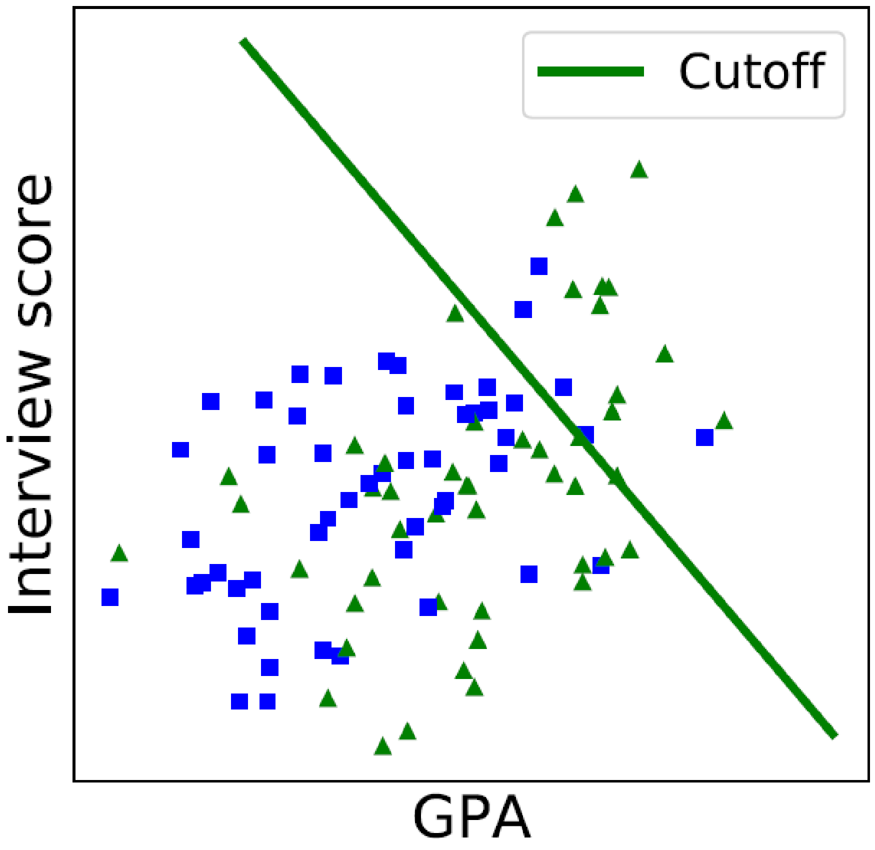
Let *A* be a sensitive attribute, target variable *Y*, and classifier prediction *R*.

Example In our admissions case,

A : Demographic group

R : Prediction of admission

Y : Actual acceptance outcome



Classification Fairness Criteria

Independence	Separation	Sufficiency
$R \perp A$	$R \perp A \mid Y$	$Y \perp A \mid R$

Independence The prediction and attribute are independent

Example The probability of predicting admission doesn't differ across demographic groups,

$$P(R \mid A = a) = P(R \mid A = b)$$

Demographic parity, statistical parity, group fairness, disparate impact

Classification Fairness Criteria

Independence	Separation	Sufficiency
$R \perp A$	$R \perp A \mid Y$	$Y \perp A \mid R$

Separation Score and attribute are conditionally independent, given the classifier decision

Example There is no relationship between prediction and attribute within accepted / non-accepted groups,

$$P(R \mid Y = 1, A = a) = P(R \mid Y = 1, A = b)$$

$$P(R \mid Y = 0, A = a) = P(R \mid Y = 0, A = b)$$

Classification Fairness Criteria

Independence	Separation	Sufficiency
$R \perp A$	$R \perp A \mid Y$	$Y \perp A \mid R$

Sufficiency Outcome and attribute are independent given the model prediction

Example There is no relationship between whether someone is admitted and their demographic group within predictions

$$P(Y \mid R = 1, A = a) = P(Y \mid R = 1, A = b)$$

$$P(Y \mid R = 0, A = a) = P(Y \mid R = 0, A = b)$$

Data Science Fairness

In short… there is a lot to say on ethics and fairness… and much can be quantified rigorously…

FAIRNESS AND MACHINE LEARNING

Limitations and Opportunities

Solon Barocas, Moritz Hardt, Arvind Narayanan

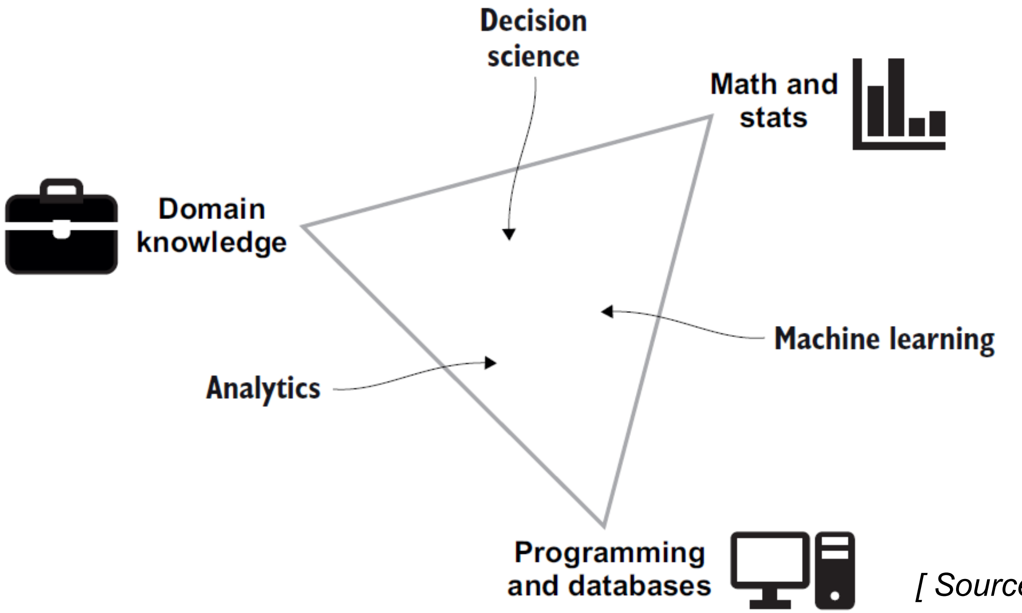
https://fairmlbook.org/

Outline

- Data Science Ethics and Fairness
- Course Recap
- Additional Resources
- Final Exam Overview

What is “Data Science”?

My Definition: *The process of using data to answer questions, extract knowledge, and predict future outcomes.*



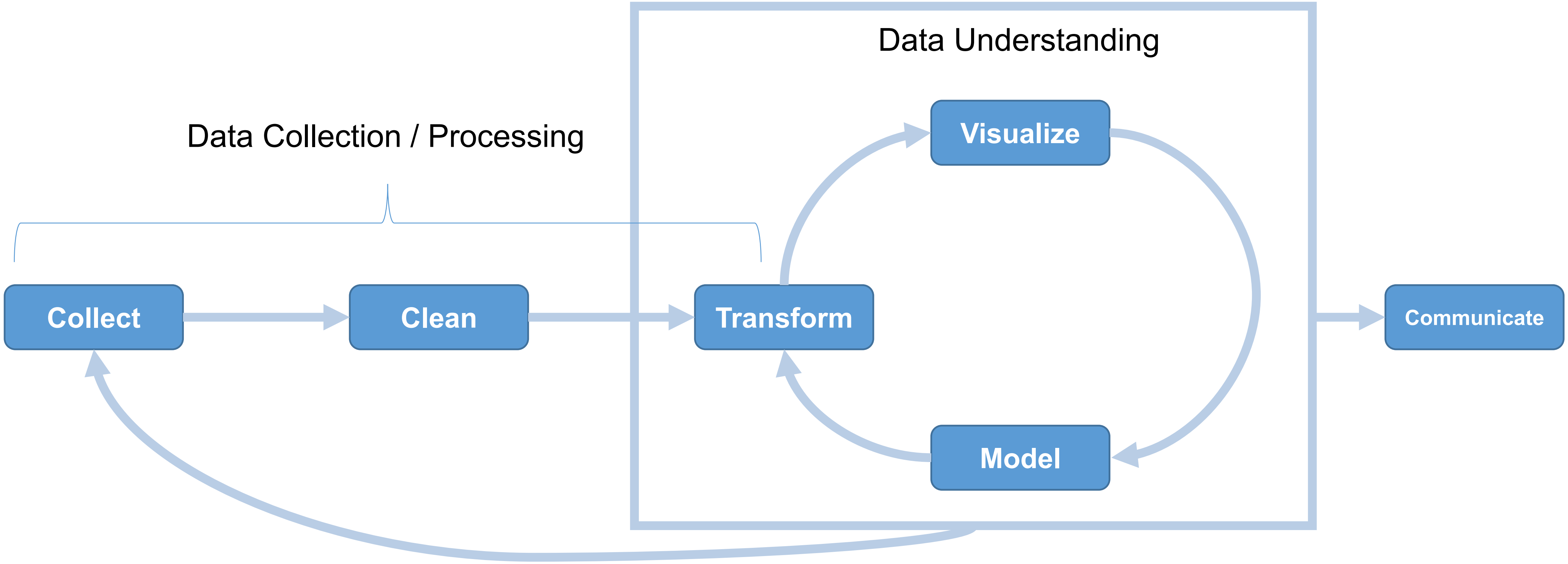
Caveat: I don’t *love* this figure since it leaves out **visualization.**

[Source: Robinson, E. and Nolis, J.]

Data Science Is:

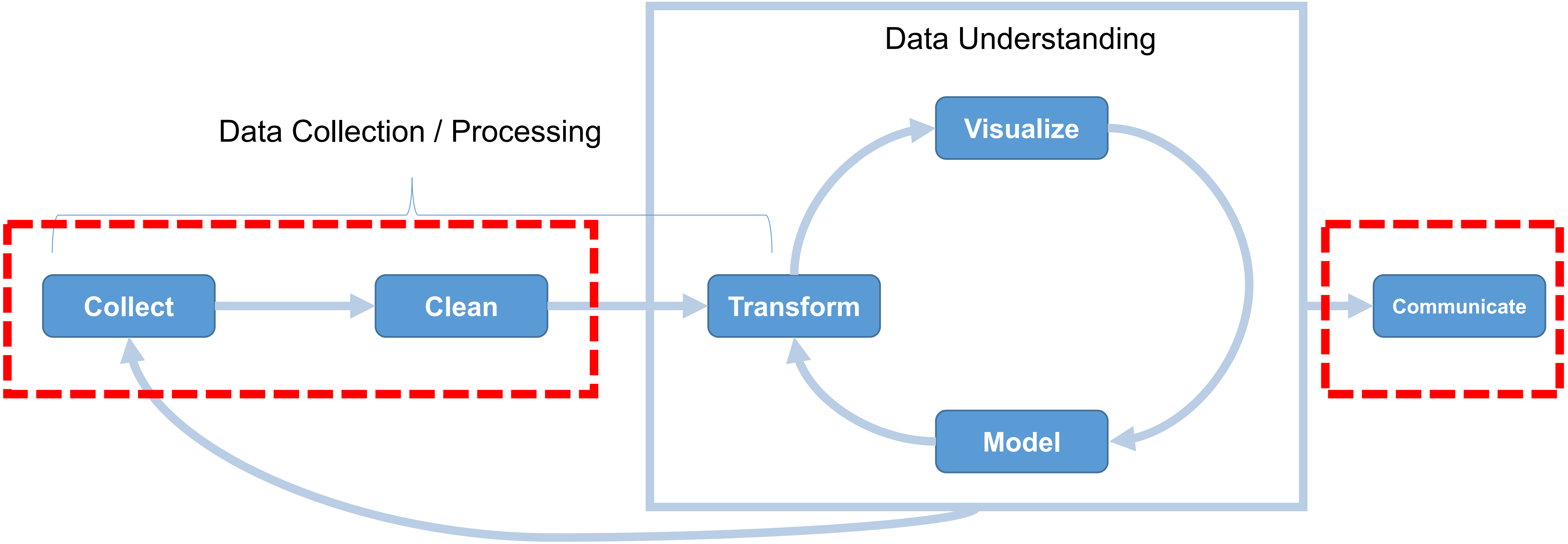
- **Interdisciplinary**: Combines tools and techniques from Math / Statistics / CS
- **Exploratory**: Understanding data requires creative exploration and visualization
- **Applied Statistics & Probability** + extra stuff to handle, process, and visualize data

Data Science Workflow



[Adapted from: Grolemund and Wickham, 2018]

Data Science Workflow



Only touched on these briefly…

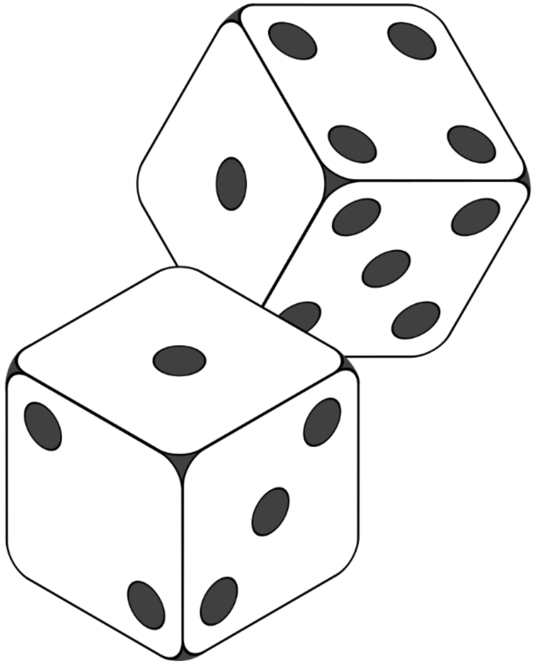
[Adapted from: Grolemund and Wickham, 2018]

Course Overview

Course Objective *Introduction to basic concepts in data science and machine learning.*

Probability and Statistics	Data Handling and Visualization	Machine Learning	Ethics and Fairness
Random events / variables, distributions / densities, moments, descriptive stats, estimation	Reading & cleaning, transformation & preprocessing, visualization	Predictive models, supervised learning, unsupervised learning, model checking	Data privacy, ethics, fairness

Probability and Statistics



Suppose we roll <u>two fair dice</u>…

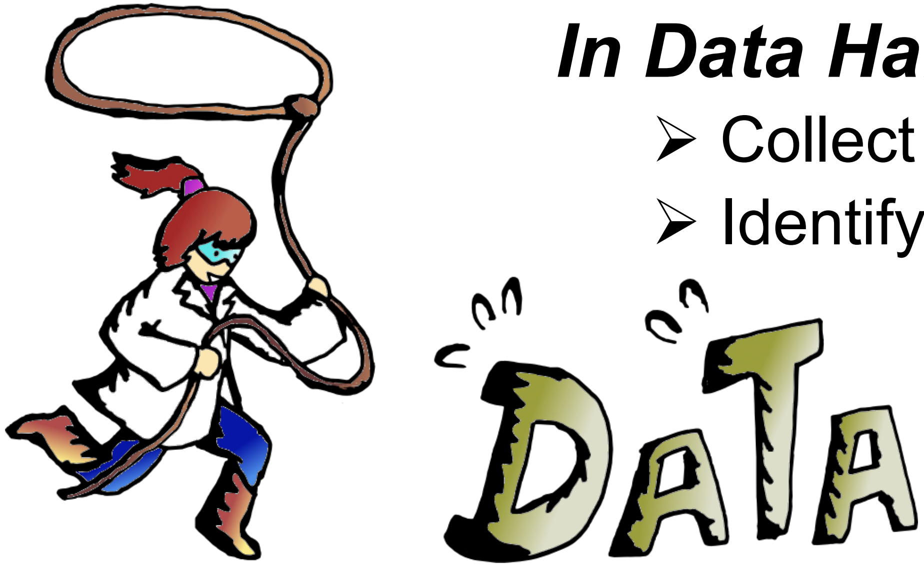
- What are the possible outcomes?
- What is the *probability* of rolling **even** numbers?

*… this is an **experiment** or **random process**.*

We learned how to…

- Mathematically formulate outcomes and their probabilities?
- Describe characteristics of random processes
- Estimate unknown quantities (e.g. are the dice actually fair?)
- Characterize the uncertainty in random outcomes
- Identify and measure dependence among random quantities

Data Handling and Visualization



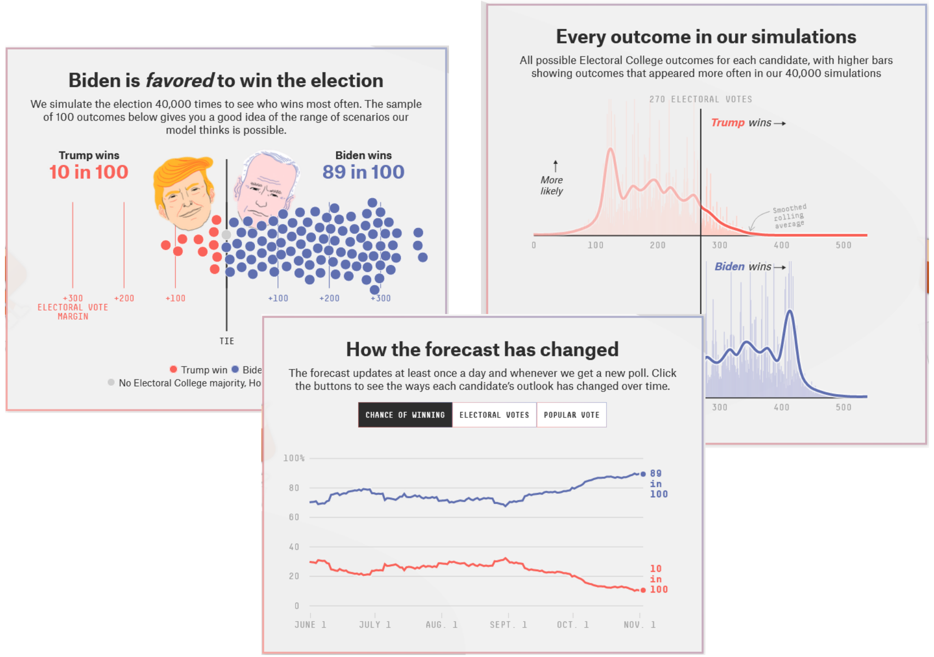
[Image Source: Code A Star]

In Data Handling learned to…

- Collect data through population sampling
- Identify and avoid biased population samples
- Clean data and correct errors
- Transform and preprocess data (*wrangling*)

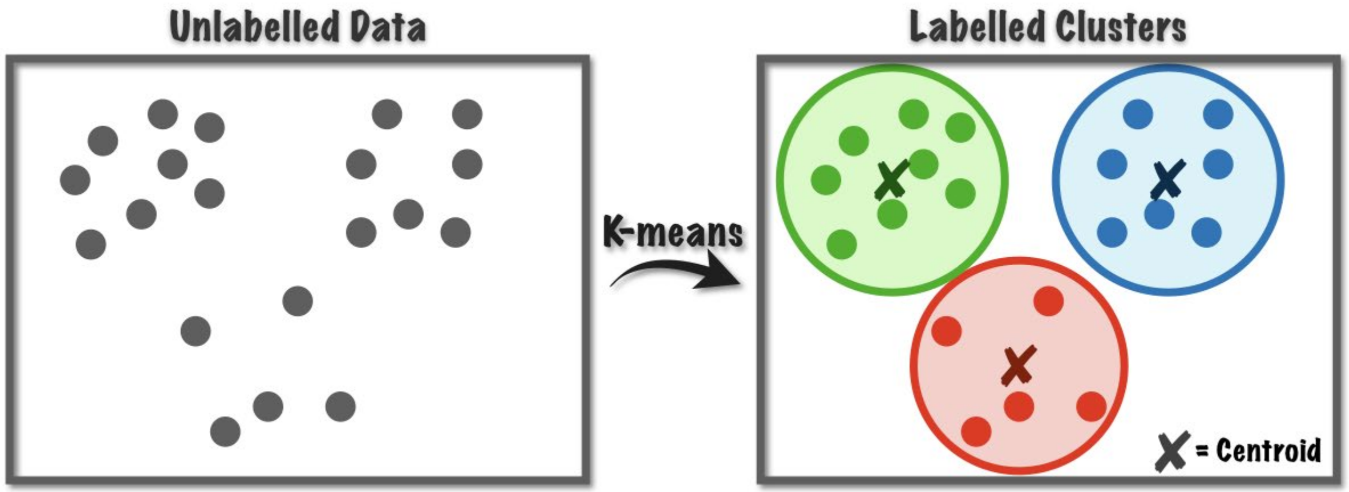
In Data Visualization we learned…

- Why visualization is important
- Exploratory data analysis
- Common forms of visualization
- Pitfalls and gotchas



Machine Learning

How do use data to learn underlying patterns and predict unknowns?



In Machine Learning we learned…

- Principles of prediction
- Use of training / validation / test data
- Unsupervised vs. supervised learning
- Linear and nonlinear models

[Image Source: Towards Data Science]

Outline

- Data Science Ethics and Fairness
- Course Recap
- Additional Resources
- Final Exam Overview

Data Science Competitions

Competitions can be a great way to hone your skills…

CodaLab

MACHINE HACK

topcoder™

kaggle

Analytics Vidhya

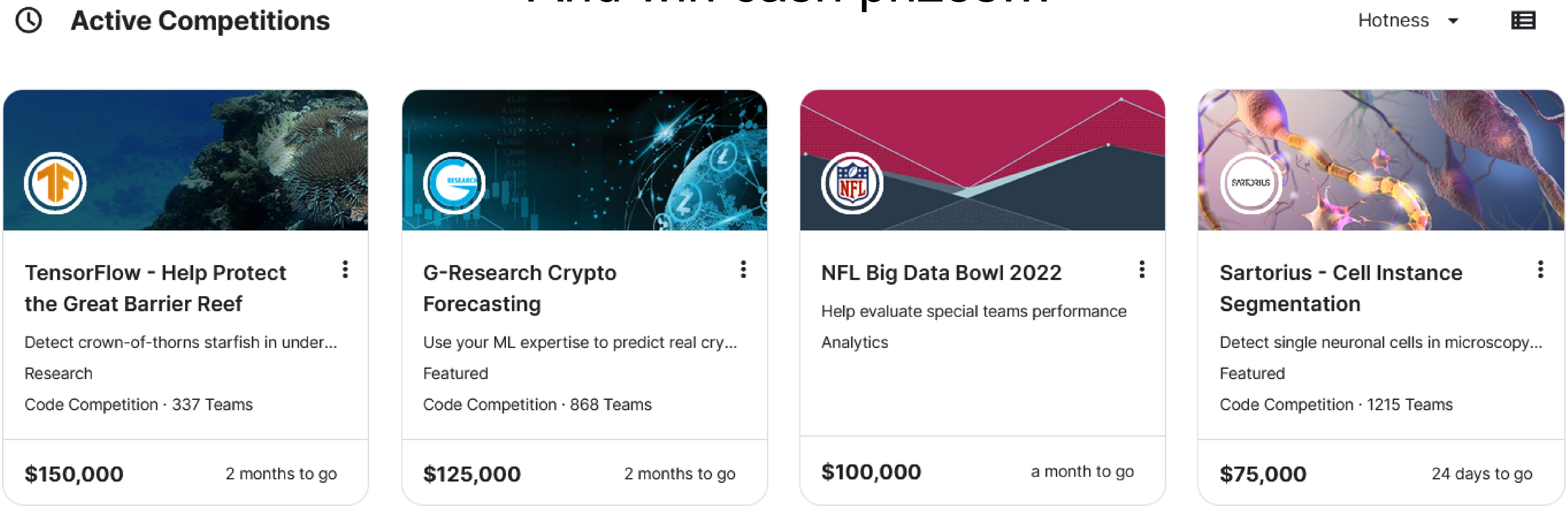
IDA

INTERNATIONAL
DATA ANALYSIS OLYMPIAD

DRIVENDATA

Data Science Competitions

And win cash prizes…



Can also be a great source for datasets to practice

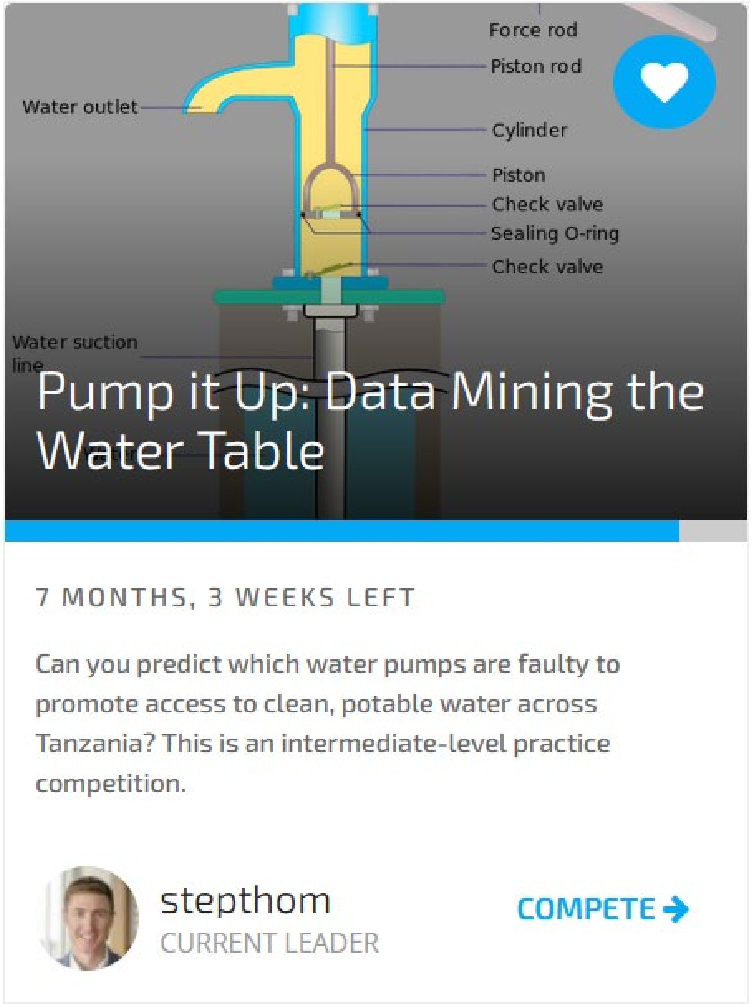
www.Kaggle.com

Data Science Competitions

Cash prizes aren't the only goal…

DRIVENDATA

- Focuses on social impact
- Challenges last 2-3 months
- Real-world predictive problems
  - Detecting hateful content online
  - Predicting disease spread
  - Predicting damage from earthquakes
  - …
- Submissions are released as open source



Additional Relevant Courses

- CSC 480 : Principles of Machine Learning
- CSC 444 : Introduction to Data Visualization
- ISTA 457 : Neural Networks
- ESOC 330 : Digital Dilemmas : Privacy, Property, and Access
- MATH 574M : Statistical Machine Learning

Videos

3Blue1Brown

- Accessible videos on a variety of math topics
- Nicely produced, engaging graphics
- A number of ML / Data Science / Statistics topics covered

Steve Brunton – YouTube Channel

- More detailed videos on math / engineering topics
- Good linear algebra and machine learning videos
- Associated book,

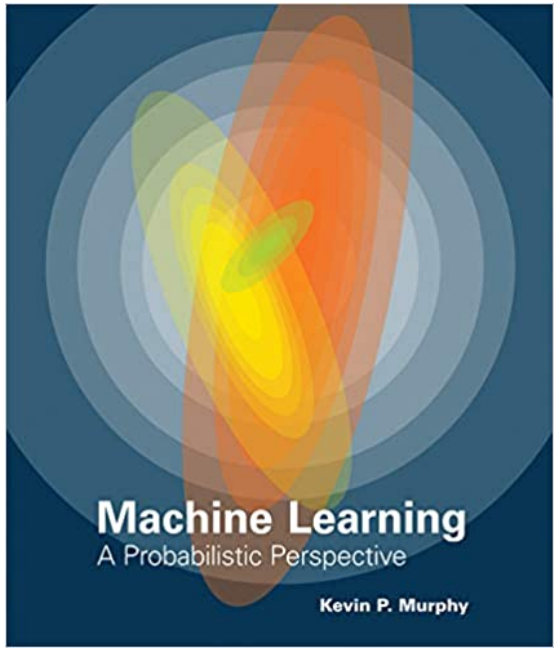
Data-Driven Science and Engineering : ML, Dynamical Systems, and control

Videos

MIT Open Courseware

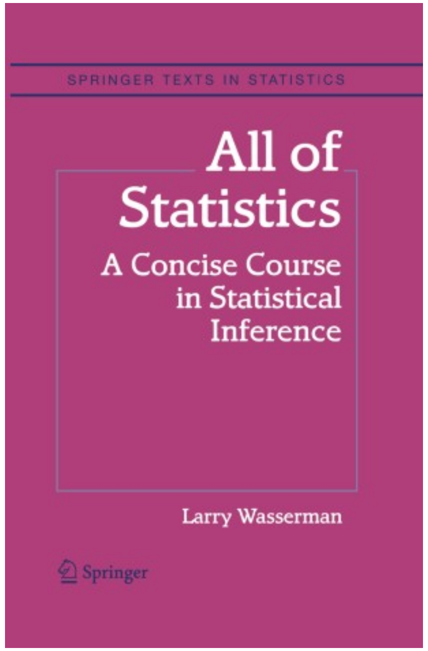
- Lots of topics freely available
- Excellent Linear Algebra course by Prof. Gilbert Strang (YouTube lectures)
- All assignments and exams available online

Textbooks



Murphy, K. "Machine Learning: A Probabilistic Perspective." MIT press, 2012

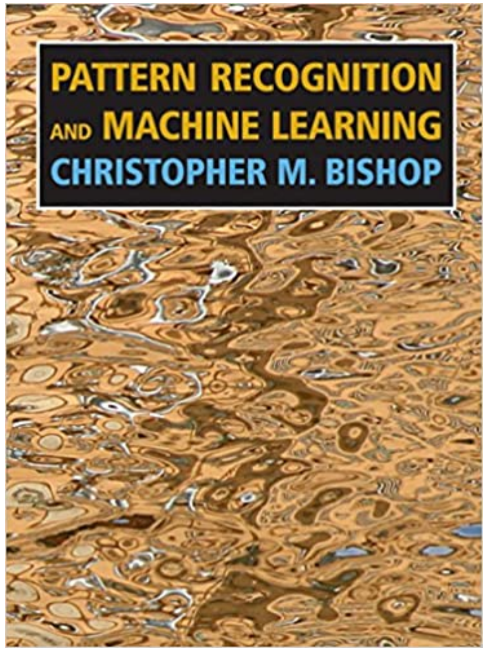
(UA Library)



Wasserman, L. “All of Statistics.” Springer, 2004

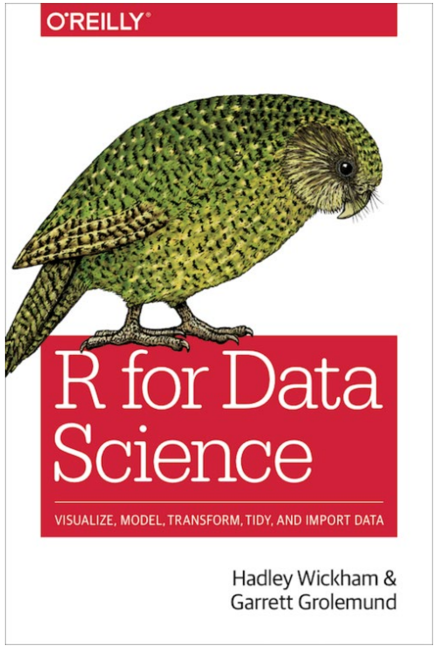
(Springer)

Textbooks



Bishop, C. "Pattern Recognition and Machine Learning." Springer, 2006

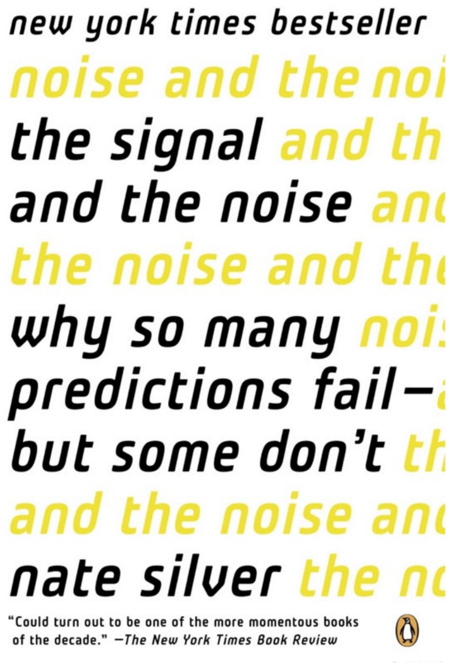
[(Microsoft)]



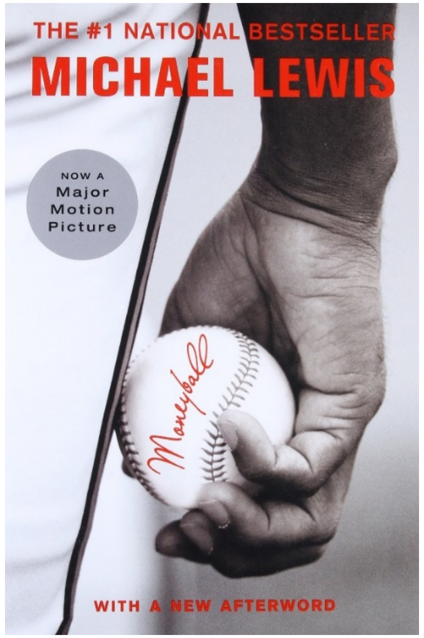
Wickham and Grolemund. “R for Data Science.” O’Reilly, 2016

[(O’Reilly)]

Non-Textbooks



Silver, N. "The Signal and The Noise." Penguin, 2015



Lewis, M. "Moneyball." W. W. Norton, 2011

Outline

- Data Science Ethics and Fairness
- Course Recap
- Additional Resources
- Final Exam Overview

Final Exam

- Out Thursday 12 / 9, Due Wed 12 / 15 @ 11:59
- Worth 20 points total
- 3 Required questions (Jupyter notebooks)
- 1 Extra credit (written)

Final Exam

Question 1 : Data analysis and visualization

- Review Homework 5

- Load data / do some cleaning and preprocessing / display visualizations

- Review boxplots, scatterplots, pie charts

Question 2 : Linear / Nonlinear Regression

- Review Homework 7 and part of Homework 8 (PolynomialFeatures)
- Fit linear regression model
- Compute polynomial features
- Fit nonlinear regression

Question 3 : PCA + High-Dimensional Clustering

- Review PCA lecture and HW9
- Fit K-Means to high-dimensional dataset
- Fit PCA model to reduce dimensions
- Fit K-Means to reduced dimension model

Questions?