# Studying Data Science in College

Prof. Mine Dogucu
2022-04-14
Panel at Northwood High School in Irvine
bit.ly/dogucu-talks

mdogucu@uci.edu

@MineDogucu

MineDogucu

- How many of you are first years/sophomores/juniors/seniors?
- How many of you have taken a statistics class?
- How many of you have taken a computer science class?

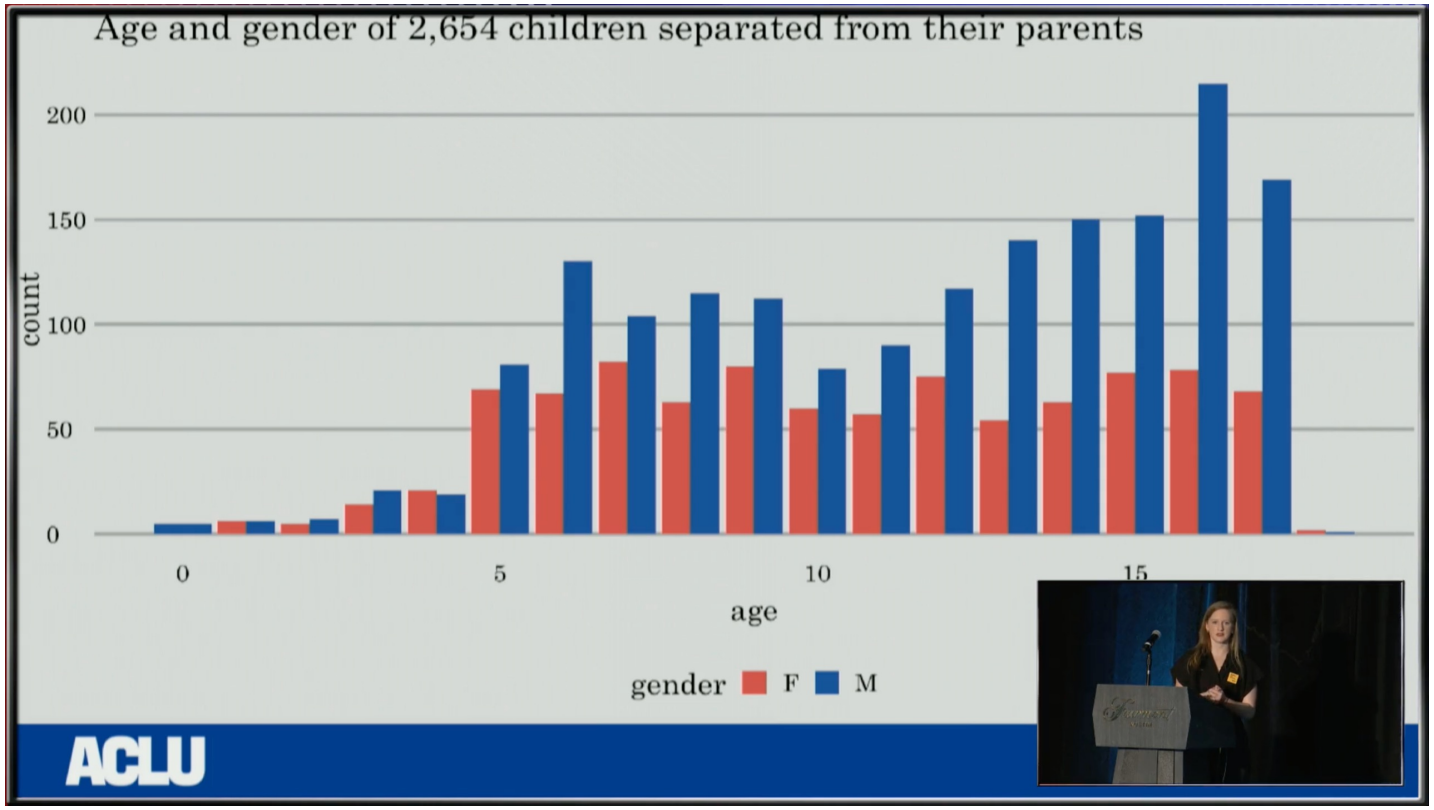
# What Do Data Scientists Do?

Data scientists build recommender systems

amazon

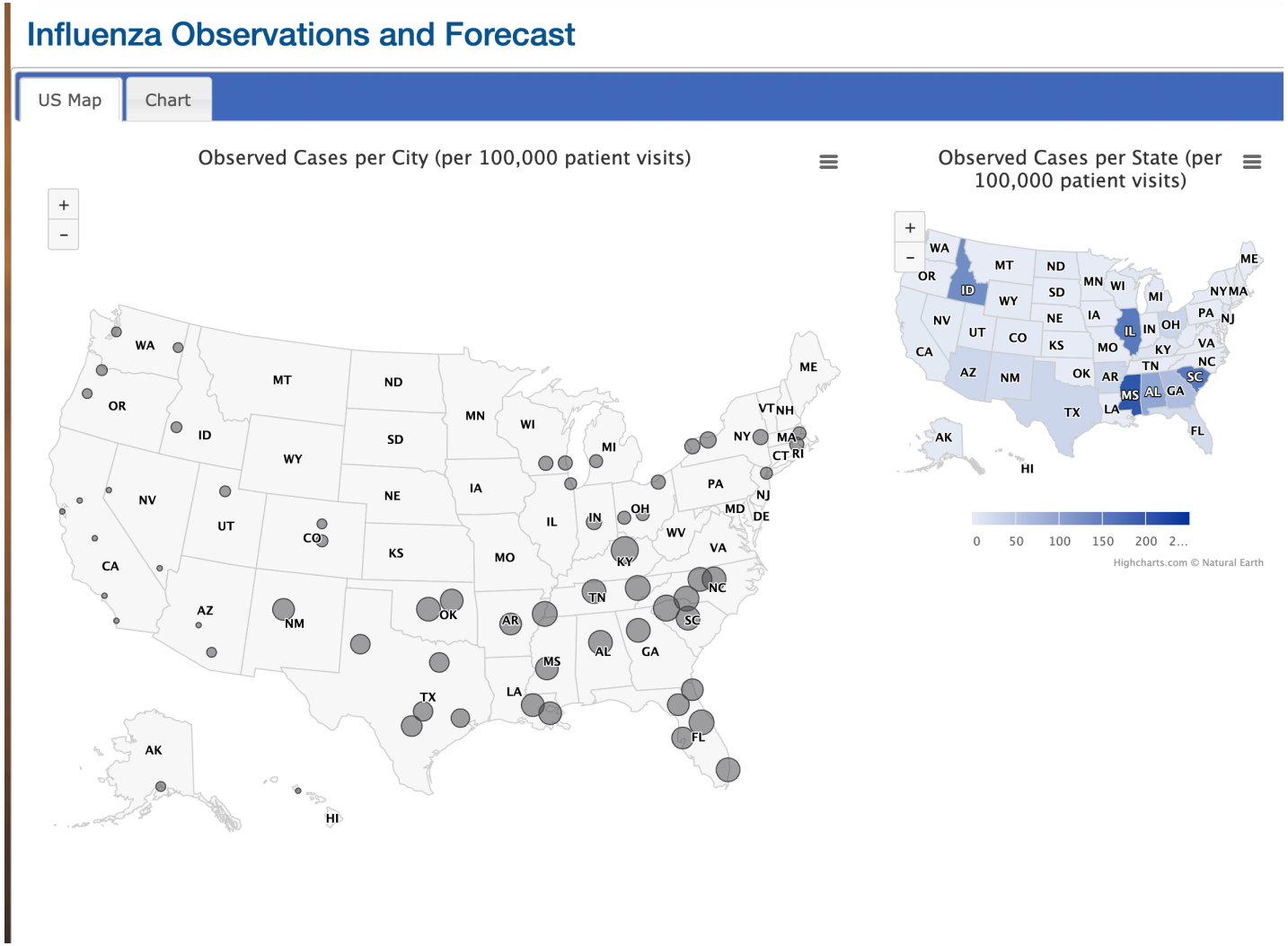
NETFLIX

Data scientists help reunite immigrant children with their families



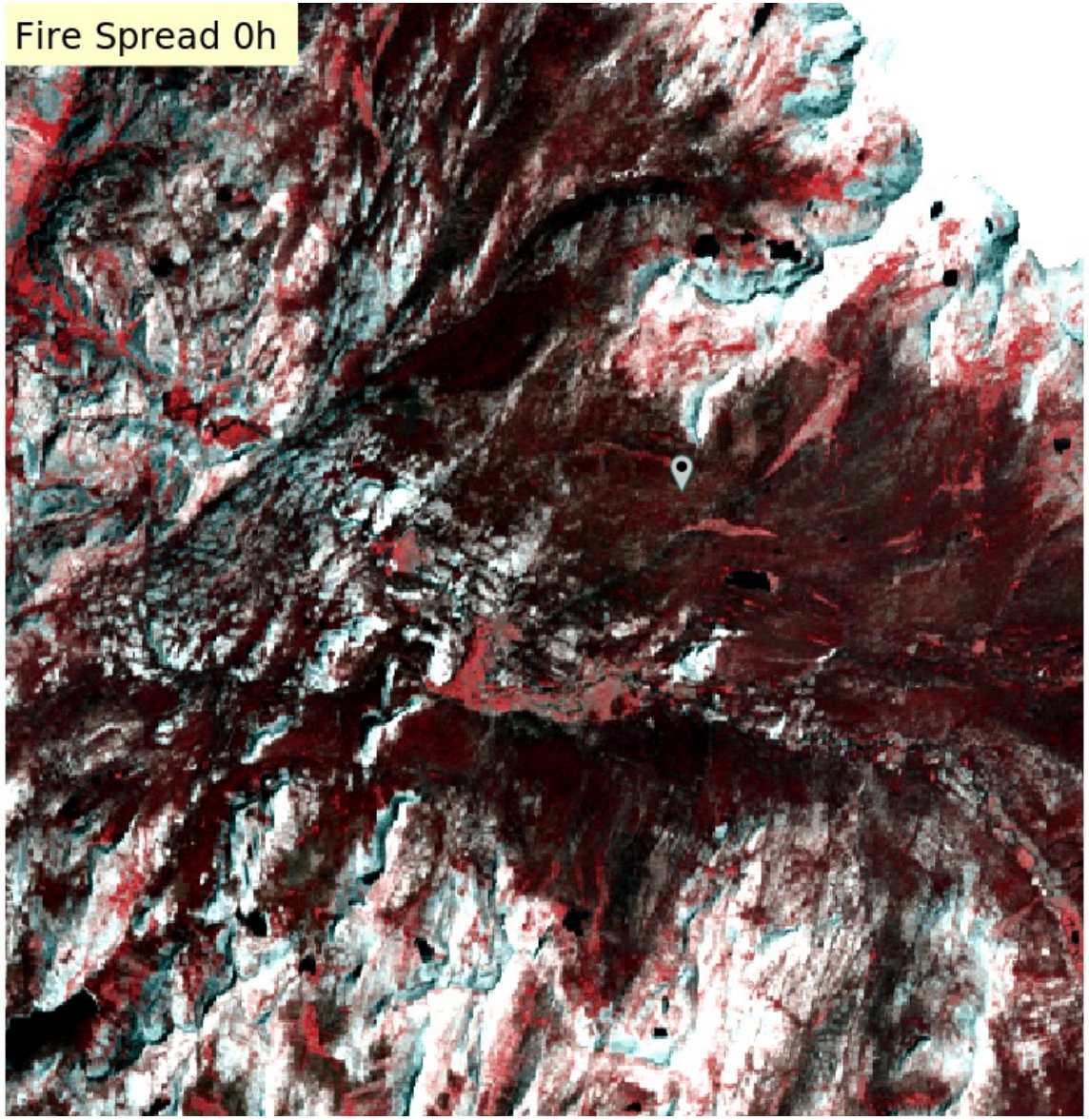

https://www.rstudio.com/resources/rstudioconf-2019/r-at-the-aclu-joining-tables-to-to-reunite-families/

Data scientists predict infectious disease outbreaks (with public health workers)




https://cpid.iri.columbia.edu/

# Data scientists model wildfires




https://towardsdatascience.com/wildfire-modeling-in-yosemite-national-park-666f45c9a0a0

No matter what the data context is (influenza outbreak , wildfires or something else), data scientists try to answer (or help others answer) scientific and business questions using data.

# Key Aspects of Data Science

Domain knowledge and problem definition

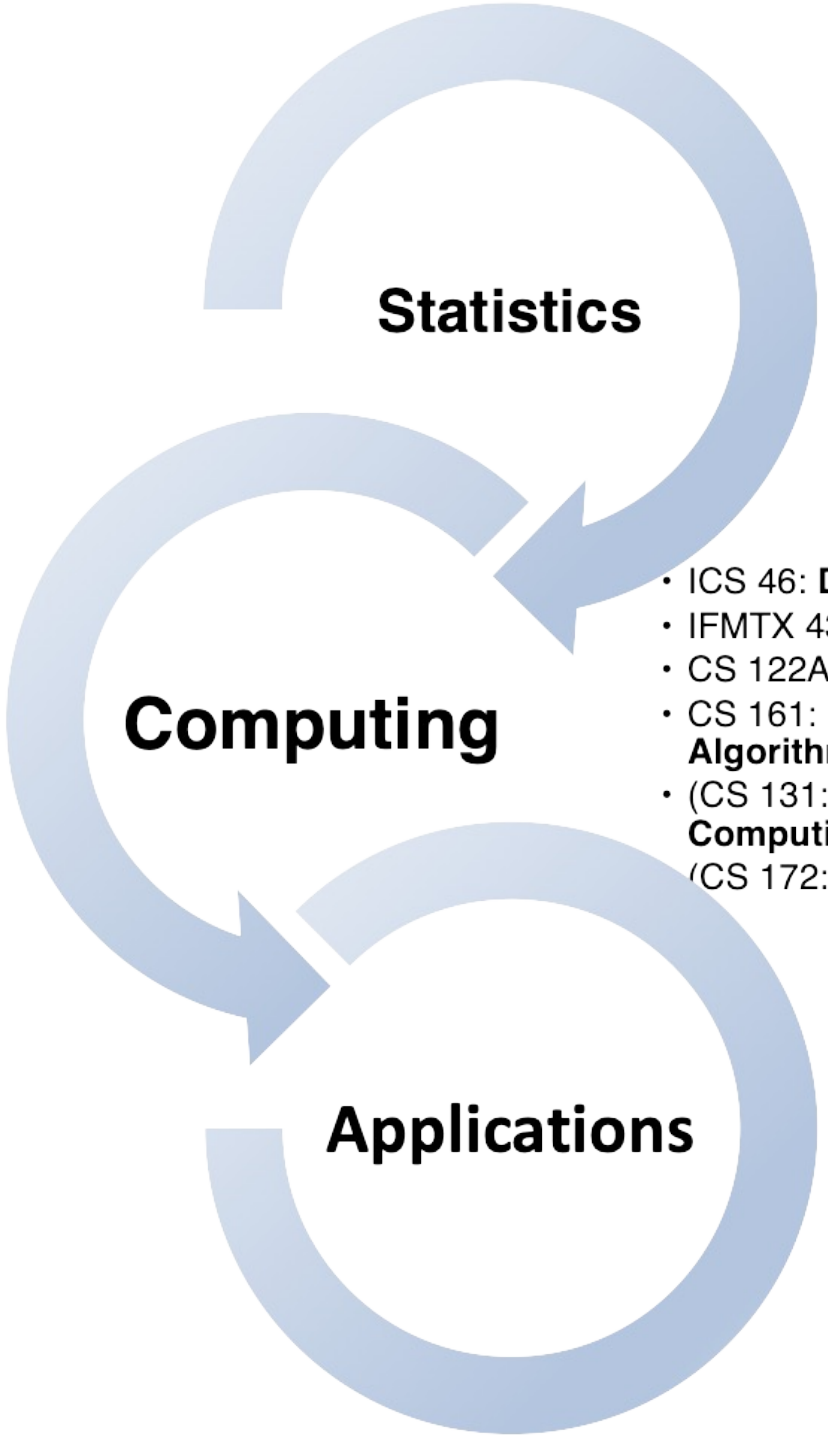
Data preparation, organization, and management

Understanding of uncertainty (statistics)

Computing, algorithms, fitting models, machine learning

Iterative exploration and experimentation

Human judgement and interpretation




- Stats 120 ABC: **Intro to Prob and Stats**
- Stats 68: **Exploratory Data Analysis**
- Stats 110-112: **Statistical Methods**
- Stats 115: **Intro Bayesian Data Analysis**
- CS 178: **Machine Learning**
- (Stats 140: **Multivariate Statistics**)

- ICS 46: **Data Structures**
- IFMTX 43: **Intro to Software Engineering**
- CS 122A: **Intro to Data Management**
- CS 161: **Design and Analysis of Algorithms**
- (CS 131: **Parallel and Distributed Computing**)
- (CS 172: **Neural Networks/Deep Learning**)

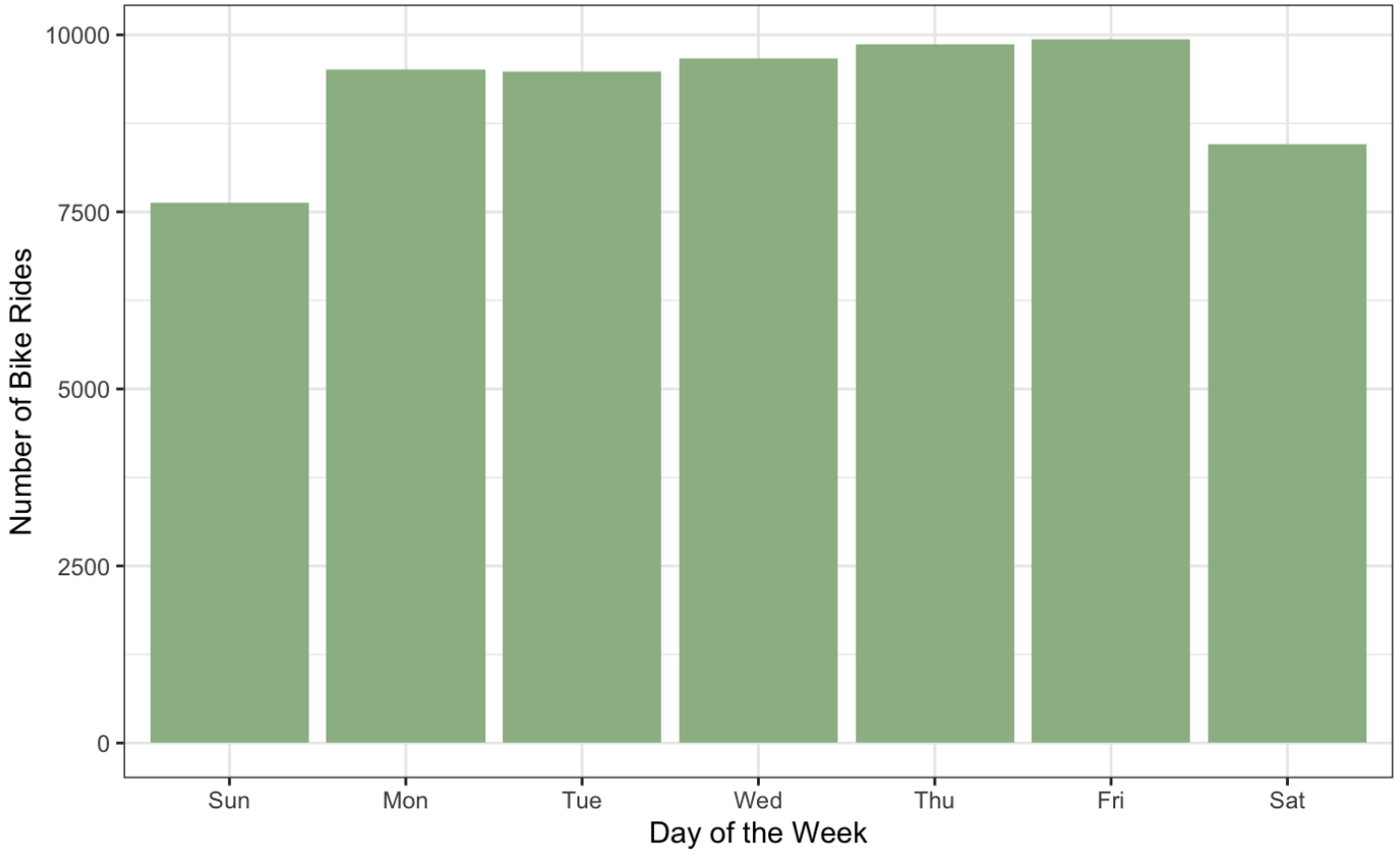
- Stats 170AB: **Data Science Capstone Project**
- INF 143: **Information Visualization**
- (INF 131: **Human Computer Interaction**)
- (CS 121: **Information Retrieval**)
- (CS 122B: **Project in Databases/Web Applications**)
- (Summer internships, e.g., junior year)

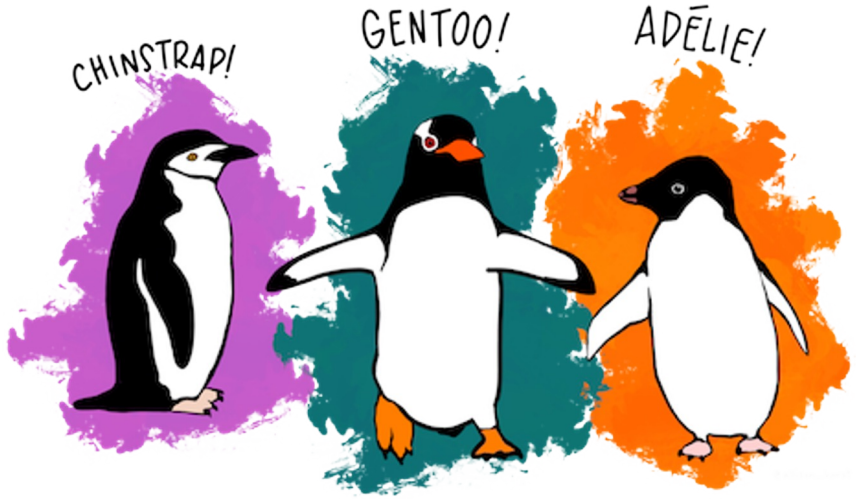
# Introduction to Data Science

# Course Goals:

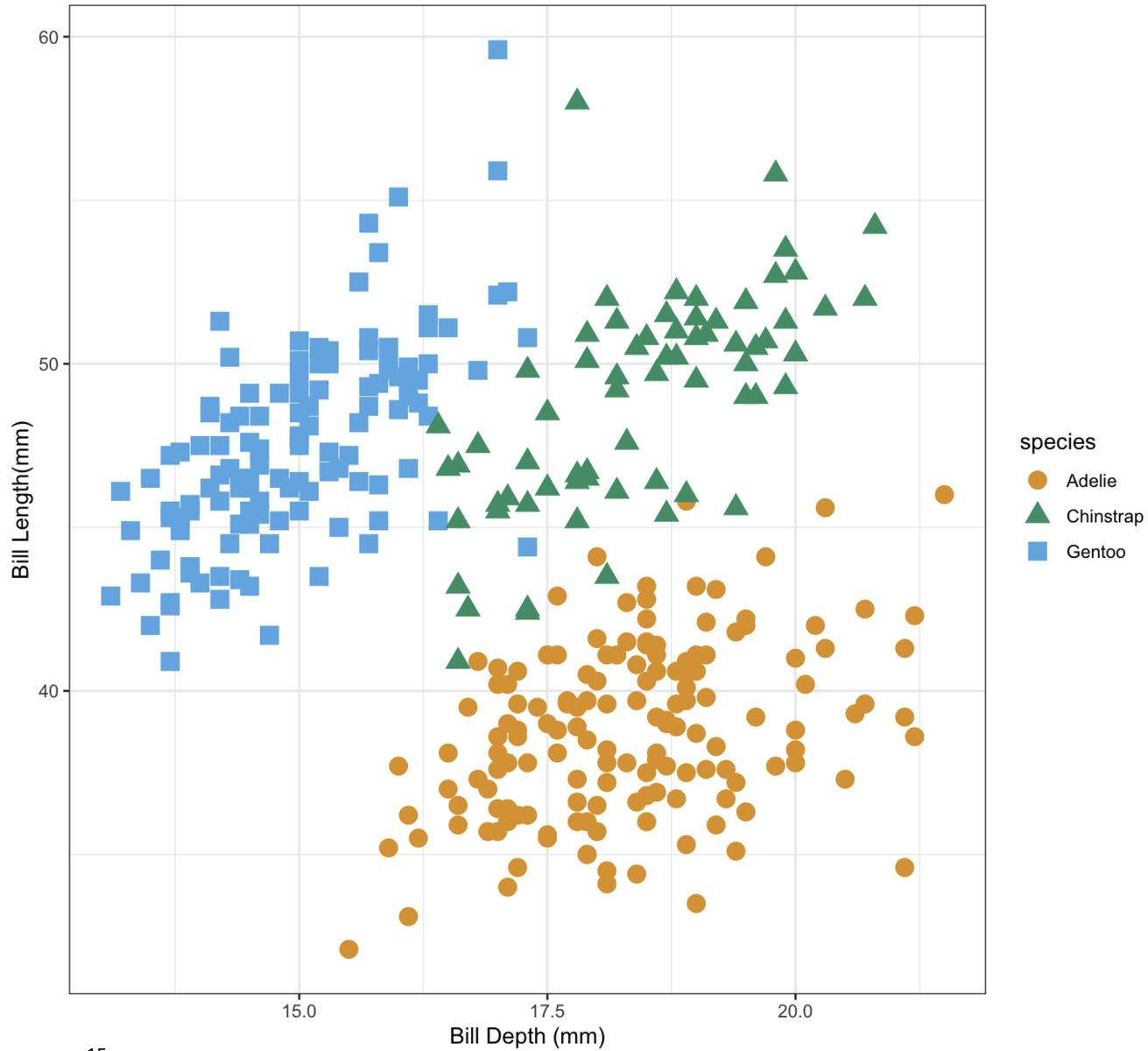
- explore data using descriptive statistics and visualizations;
- read, write, and tidy up datasets;
- make predictions and conclusions using models;
- consider impact of decisions related to data on humans, other livings, and the planet;
- write human- and machine-readable code using R.

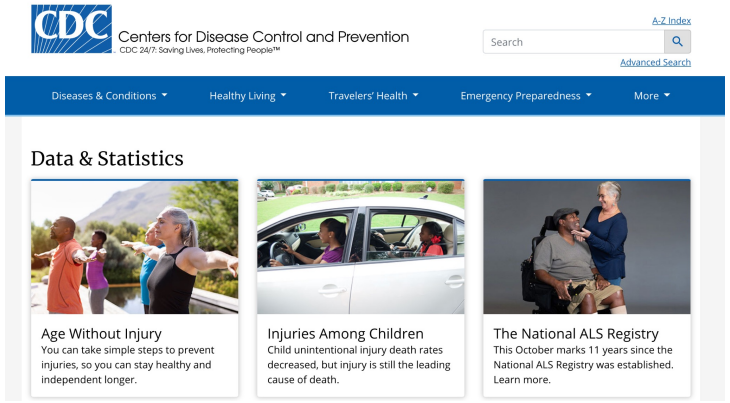
explore data using descriptive statistics and visualizations;






Artwork by @allison_horst






# read, write, and tidy up datasets;

Comments

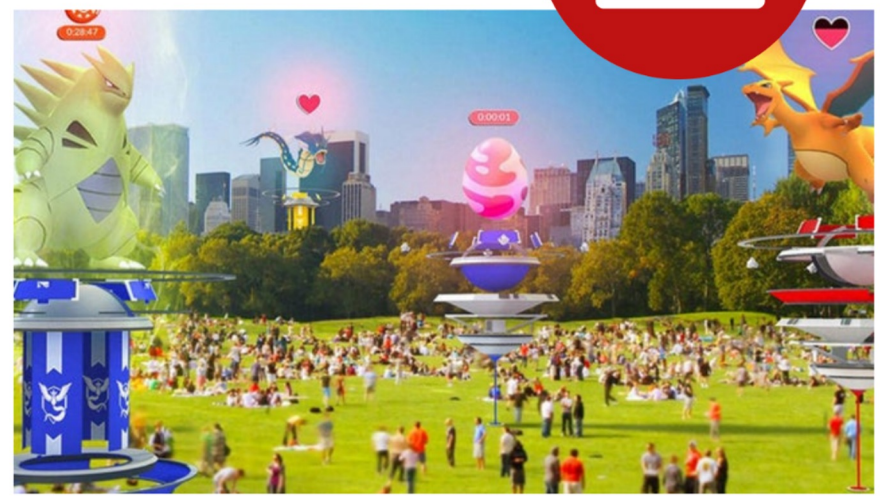
100 Pokémon Go
99 Borderlands 2
98 Divinity: Original Sin 2
97 Dishonored
96 Final Fantasy VII
95 Assassin's Creed IV: Black Flag
94 Monkey Island 2: LeChuck's Revenge
93 Burnout 3: Takedown

If you'd like to know more, check out this *list of changes to the Best 100 Games of All Time list -- and why we made those changes*.

# Pokémon Go

100

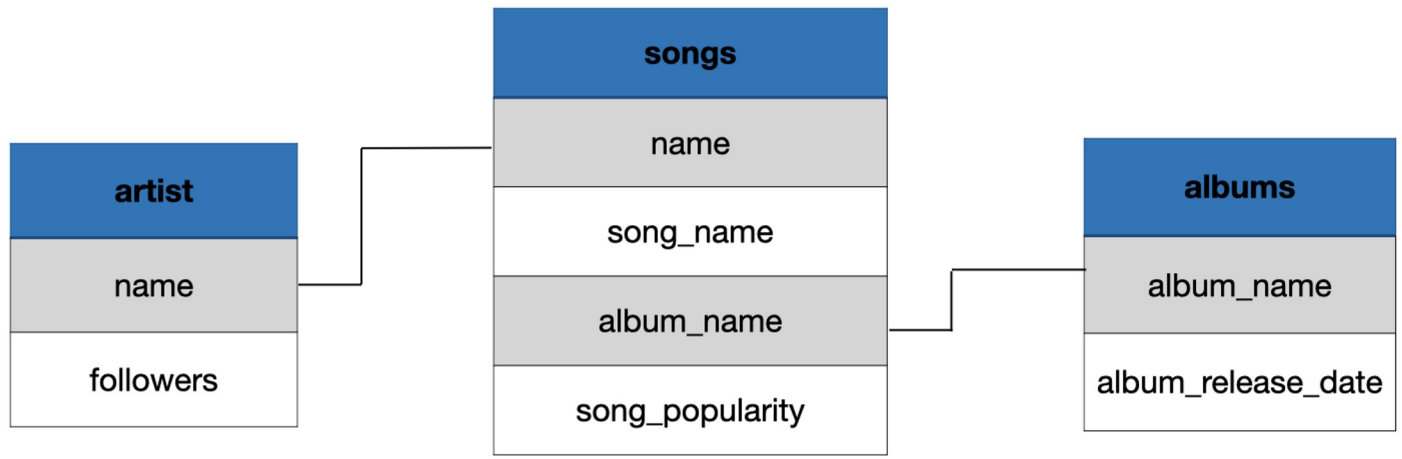
**RELEASED**
2016



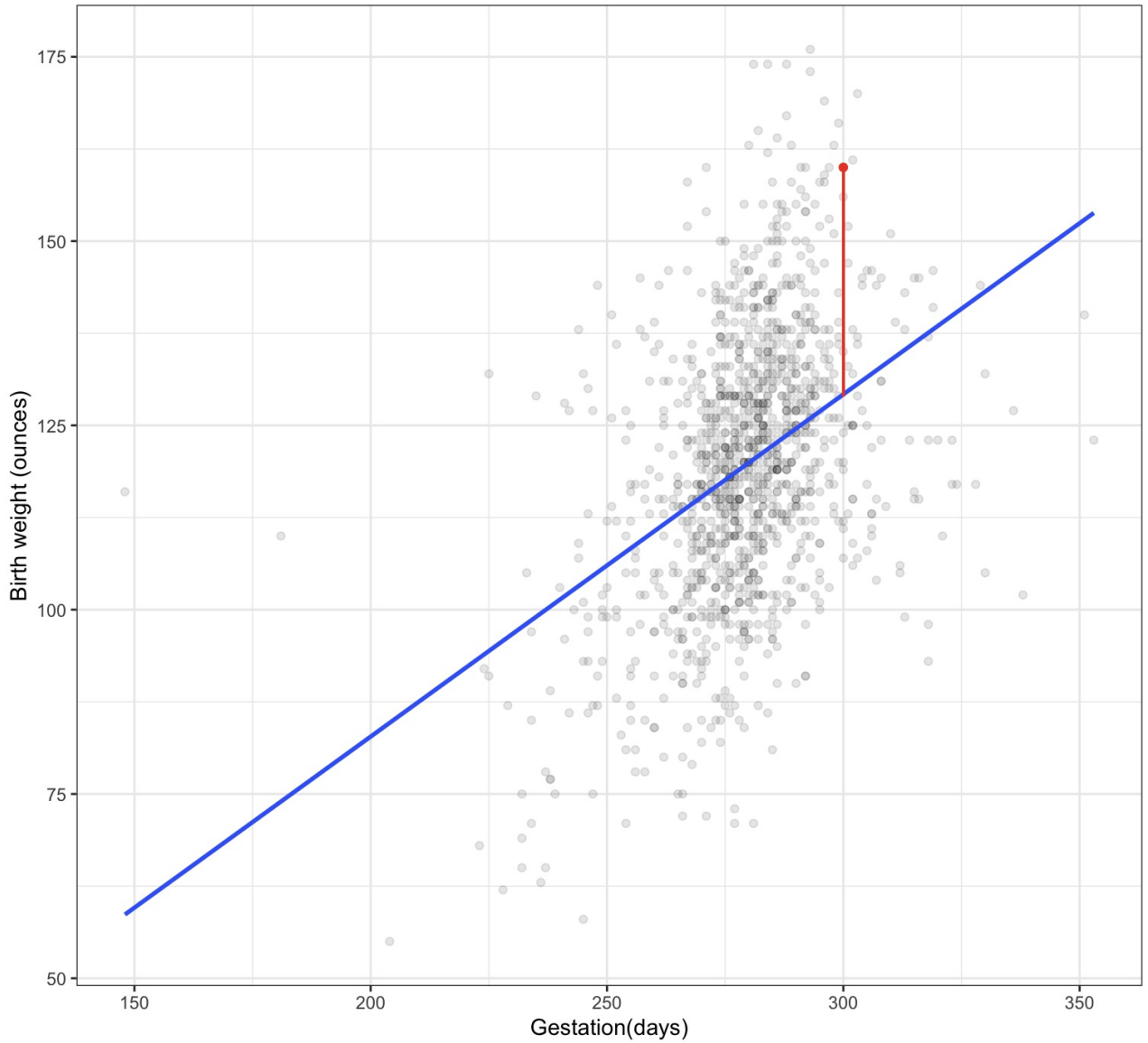
**JUSTIN VACHON →** Pokemon GO in 2019 is a game I shouldn't care about. When it launched in 2016 it was in a lot of ways a mediocre experience. Outside of catching the original 151 Pokemon the game itself relied heavily on the nostalgia of the Pokemon franchise and augmented reality gimmick of having them show up in the real world. If you didn't care about the IP, the game itself was very lacking. In 2019, the game is flooded with a multitude of tasks, activities, and events that can involve anyone from yourself to a large group of people. These additions create an experience that incentivizes users to be more dedicated to daily play without feeling like a grind. Friendship has been introduced and allows users to now exchange gifts, trade or even battle each other. Quests (research tasks as they are referred to in-

**DID YOU KNOW?**

There are currently five (of eight) generations of Pokemon in the game, with more being added each year.


songs
name
song_name
album_name
song_popularity
artist
name
followers
albums
album_name
album_release_date

make predictions and conclusions using models;



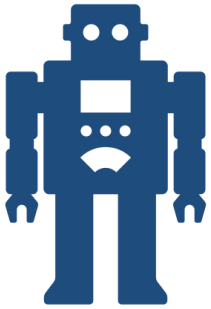
logistic regression example:

**Are Emily and Greg More Employable than Lakisha and Jamal? A Field Experiment on Labor Market Discrimination**

consider impact of decisions related to data on humans, other livings, and the planet;

Finding data online does not grant permission to use

Human subjects

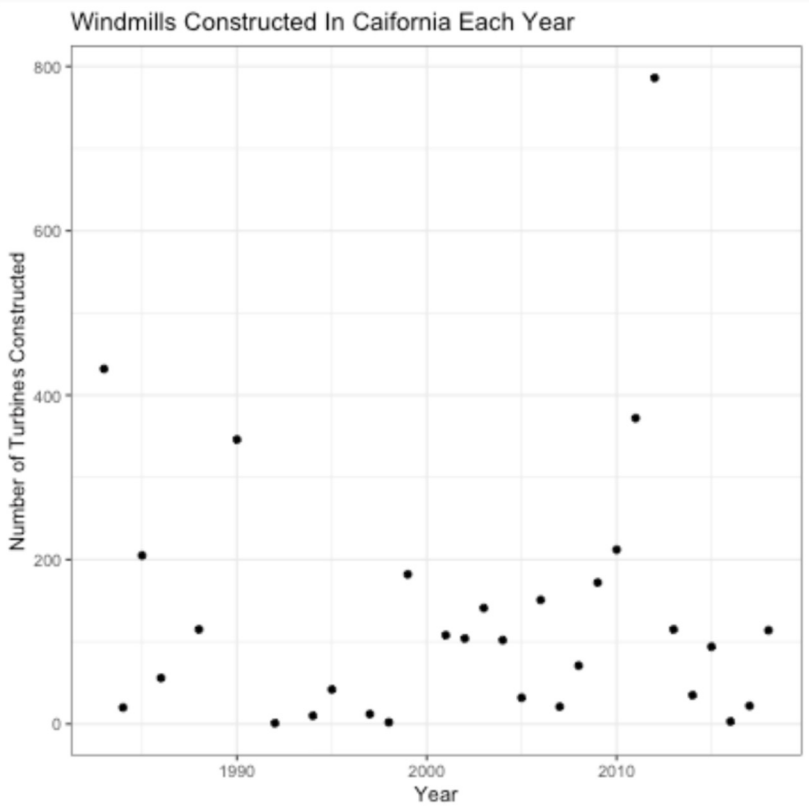


Bot access

Commercial use

write human- and machine-readable code using R.

# Windmills in California Over Time



4078 turbines in California.