

# R Fundamentals for Public Policy

Spring 2022

Jae Yeon Kim (call me Professor Jae Yeon/Jae Yeon!)  
Assistant Professor,  
KDI School of Public Policy and Management

**Why data science?  
Data science is  
already the big thing.**



data science

San Francisco Bay Area

Search



Try  
Premium  
for free

Jobs

Experience Level 3

Full-time 1

Date Posted

Company

On-site/Remote

All filters 1

Reset

## Data science in San Francisco Bay Area

20,945 results

Set alert



1 week ago · 15 applicants



## Data Science Consultant

Accenture  
San Francisco, CA  
623 alumni work here

1 day ago · 3 applicants



## Data Science Consultant

Accenture  
San Jose, CA  
623 alumni work here

1 day ago · 0 applicants

Statistics Department: As  
Statistical Data Science

University of California, Santa C  
Santa Cruz, CA (On-site)

249 alumni work here

4 weeks ago · 1 applicant



## Lead Applied Data Science

Civis Analytics  
United States (Remote)  
Medical, Vision, Dental, 401(k), +1 be

3 alumni work here

6 days ago · 7 applicants

Leadership, Data Science Catalog,  
Product Information Extraction

Wayfair · United States (Remote) 23 minutes ago

Full-time · Mid-Senior level

10,001+ employees · Retail



data science



South Korea

Search



Try  
Premium  
for free

Jobs

Experience Level 3

Full-time 1

Date Posted

Company

On-site/Remote

All filters 1

Reset

## Data science in South Korea

249 results

Set alert

Senior Analyst - Data Science, Marketing &  
Sales

McKinsey & Company  
Seoul, Seoul, South Korea (On-site)

379 alumni work here

Promoted



UNSW

Seoul, Seoul, South Korea (On-site)

34 alumni work here

6 days ago · 3 applicants



Deep Learning Engineer

Weights & Biases  
Seoul, South Korea (Remote)

6 alumni work here

9 hours ago



[NAVER LABS] 전부문 인재영입 안내

NAVER LABS Corp.  
Gyeonggi, South Korea (Hybrid)

Actively recruiting

1 week ago

Senior Analyst - Data Science, Marketing &  
Sales

McKinsey &amp; Company · Seoul, Seoul, South Korea (On-site) 5 days ago · 45 applicants

Full-time · Associate

10,001+ employees · Management Consulting

379 school alumni

See how you compare to 45 applicants. Try Premium for free

Actively recruiting

Apply

Save

## Qualifications

- Advanced degree in computer science or quantitative discipline (statistics or mathematics)
- 4+ years of experience working with SQL and Python or R programming
- Ability to manage and query large databases including setting up and maintaining gigabyte to terabyte size databases of customer data, performing data integration, manipulation, and querying, and provide data mining and analytical support
- Strong understanding of advanced analytics or machine learning concepts for customer analytics, including segmentation, cluster analysis, consumer behavior, pricing analytics, customer insights, etc.



# Chief Data Officer

Evidence-Building Plans and Assessments

## Leadership

Department of Labor Evidence Act Officials

Chief Evaluation Officer

## Chief Data Officer

## Scott Gibbons

[Learn More about the Chief Data Office](#)

The Chief Data Officer (CDO) is responsible for data governance and lifecycle data management, supporting the Department in meeting the requirements of the Federal Data Strategy and Open Government Data Act. The CDO, located in the Office of the Assistant Secretary for Policy, serves as co-chair of the Data Governance Council and oversees the implementation of the Data Strategy across the Department.



San Francisco City Hall (Shutterstock)

The city and county of San Francisco has officially named Jason Lally as its new chief data officer.

Lally, who has been with the city for close to seven years, replaces Joy Bonaguro, the first-ever CDO, who stepped down from her position in September of 2018. Though now officially CDO, Lally has managed operations at DataSF, the city's data team, since Bonaguro's departure.

Under Bonaguro, the city saw the expansion of numerous data initiatives, including creating processes to inventory, monitor and standardize open data.



## Federal Data Strategy — Data, accountability, and transparency: creating a data strategy and infrastructure for the future

The Federal Data Strategy (FDS) encompasses a 10-year vision for how the Federal Government will

M-19-23

### MEMORANDUM FOR HEADS OF EXECUTIVE DEPARTMENTS AND AGENCIES

FROM: Russell T. Vought  
Acting Director 

SUBJECT: Phase 1 Implementation of the Foundations for Evidence-Based Policymaking Act of 2018: Learning Agendas, Personnel, and Planning Guidance

### Background

The *Foundations for Evidence-Based Policymaking Act of 2018* (“Evidence Act”),<sup>1</sup> signed into law on January 14, 2019, emphasizes collaboration and coordination to advance data and evidence-building functions in the Federal Government by statutorily mandating Federal evidence-building activities, open government data, and confidential information protection and statistical efficiency. Evidence is broadly defined and includes foundational fact finding, performance measurement, policy analysis, and program evaluation (see Appendix A).<sup>2</sup> In addition, the Evidence Act mandates a systematic rethinking of government data management to better facilitate access for evidence-building activities and public consumption.

Gov data are the  
original big data; But  
the public sector has  
been left behind in  
the data revolution.

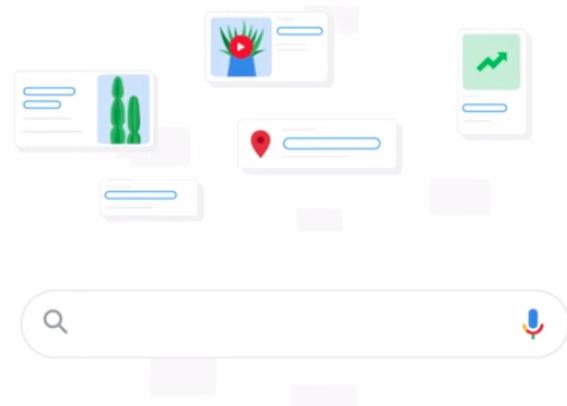
Missed  
opportunitiy!

**Quantitative methods**  
= Counting well with  
structured data  
(spreadsheets)

**Data Science =**  
**Counting well with**  
**structured/unstruct**  
**ured small/large**  
**data using computer**  
**(algorithms)**

## Maximize access to information

Google's mission is to organize the world's information and make it universally accessible and useful. That's why Search makes it easy to discover a broad range of information from a wide variety of sources. Some information is simple, like the height of the Eiffel Tower. For more complex topics, Search is a tool to explore many angles so you can form your own understanding of the world.



**Data science =  
turning data into  
actionable insights**

- A. Data for action
- B. Data for evaluation

Only on Netflix



Trending Now



big data + ML =  
recommendation  
system (relatively  
easy)

Only on Netflix



Trending Now



big data + ML =  
creating next big hits  
(really hard)

ML predicts the past not  
the future

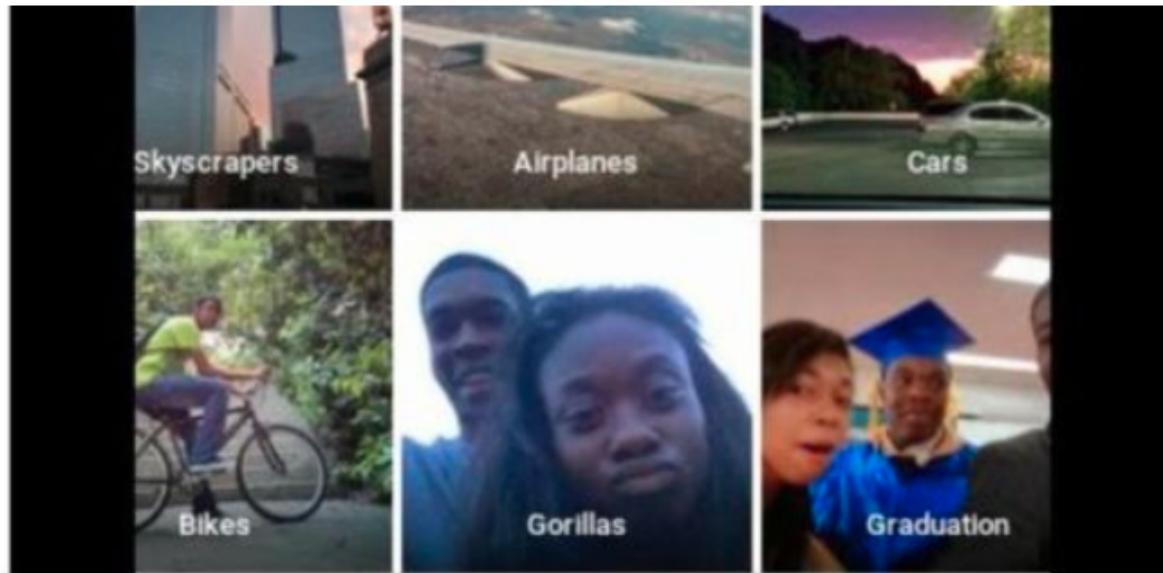
Big data != Good data

These problems are more  
serious in the public  
sector.

# Google ‘appalled’ as Photos app labels black people ‘gorillas’

Google removes ‘gorilla’ tag from app, after it misidentified images of black people

🕒 Fri, Jul 3, 2015, 07:21 Updated: Fri, Jul 3, 2015, 08:45



diri noir avec banan @jackyalcine · Jun 29

Google Photos, y'all [REDACTED] My friend's not a gorilla.

A New York-based software developer tweeted Google about the fact its app had mistakenly labelled a black couple as being “gorillas”. Image via Twitter.

Civic data science =  
data science for  
public policy, civil  
society, and  
democracy

**Data supply chain:**

**Quality data ->**

**Sound evidence ->**

**Effective policy**

# THE DATA SCIENCE HIERARCHY OF NEEDS

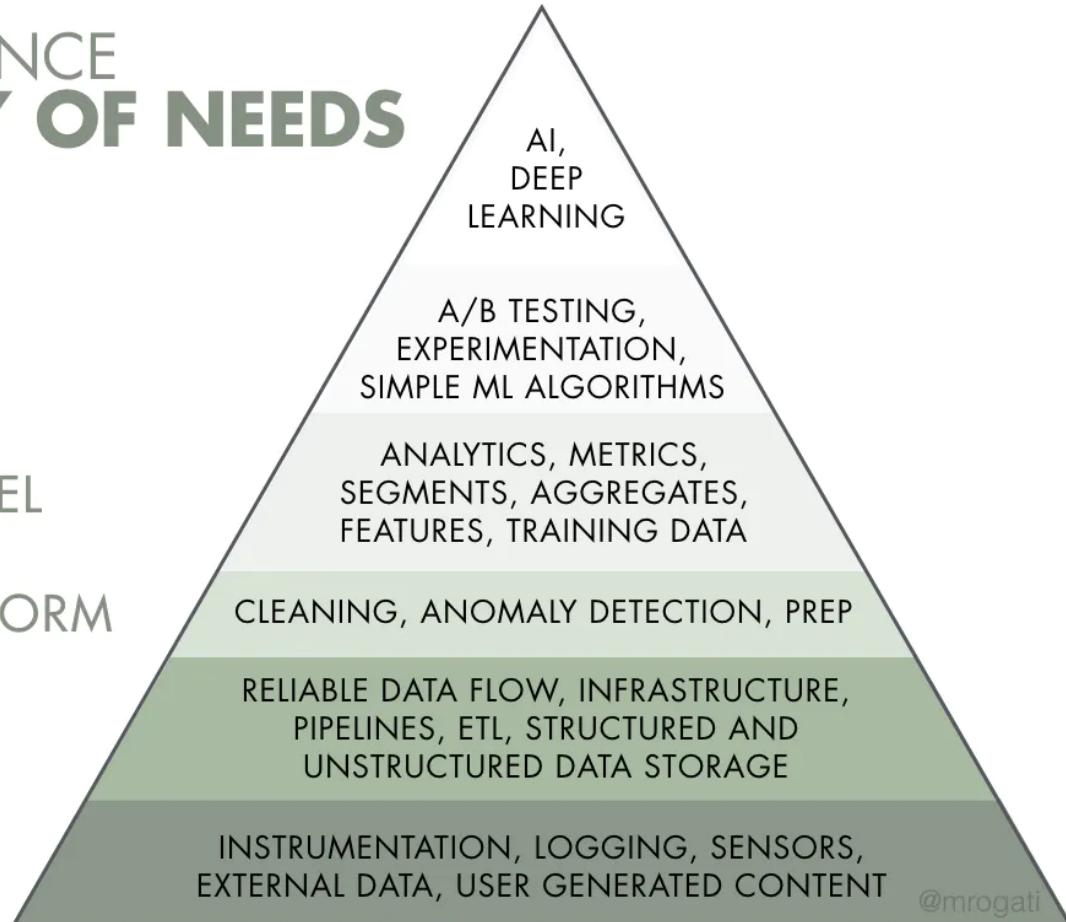
LEARN/OPTIMIZE

AGGREGATE/LABEL

EXPLORE/TRANSFORM

MOVE/STORE

COLLECT

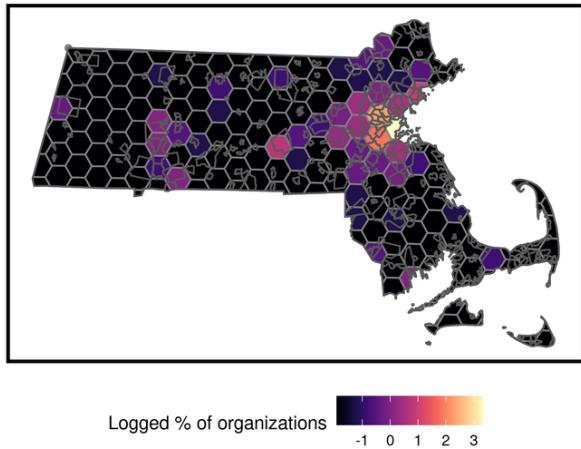


Originally published by Monica Rogati on

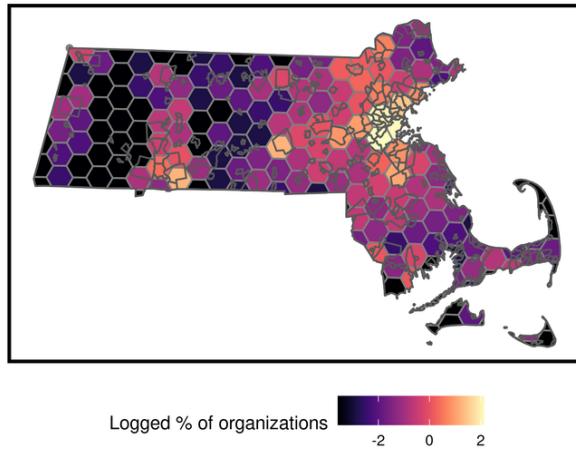
June 13th 2017

★ 173,979 reads

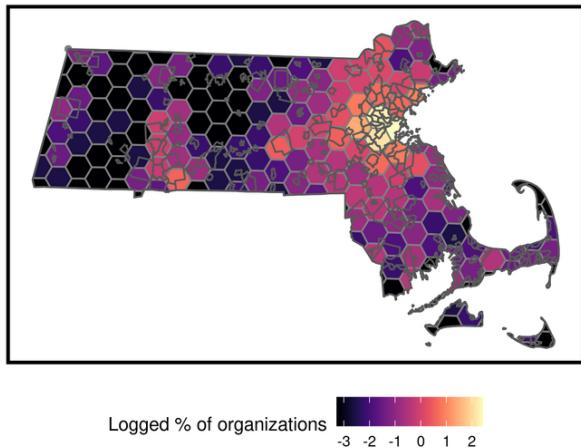
Civic organizations in MA  
*Excluding PO Box and local chapters*



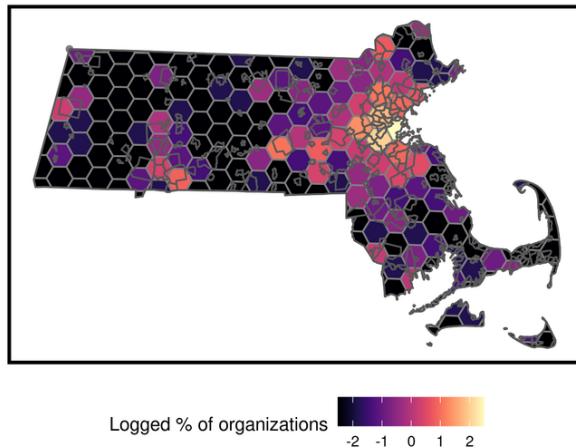
Religious organizations in MA  
*Excluding PO Box and local chapters*



Education organizations in MA  
*Excluding PO Box and local chapters*



Health organizations in MA  
*Excluding PO Box and local chapters*



My own project  
with Hahrie  
Han & Milan de  
Vries (Johns  
Hopkins  
University)

Spent 1+ year  
just collecting,  
cleaning,  
parsing, and  
joining data

# A Day In the Life of a Line Cook at One of NYC's Fanciest Restaurants

Nana Araba Wilmot makes the food almost every night at Le Coucou, but you'd never know her name.

BY AMANDA SHAPIRO

March 1, 2018



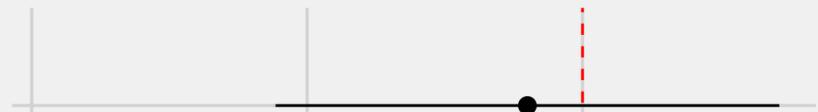
Photo By Jimmy Fontaine

Learn to be a line cook first

From Excel/SPSS/Stata  
to R (general purpose  
programming  
language): Make your  
data work efficient,  
scalable, and  
reproducible

## Background survey, N: 10

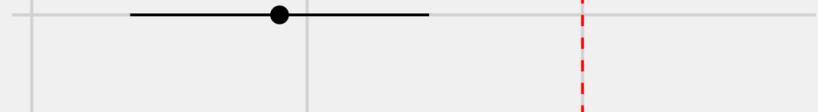
How much do you have experience in Excel (5-very much, 3-moderate, 1-very little)?



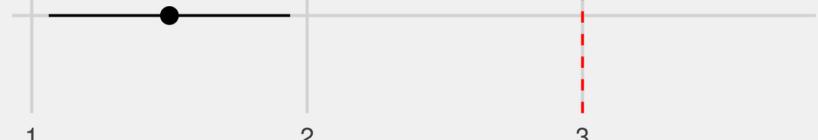
How much do you have experience in Stata/SPSS (5-very much, 3-moderate, 1-very little)?



How much do you have experience in R (5-very much, 3-moderate, 1-very little)?



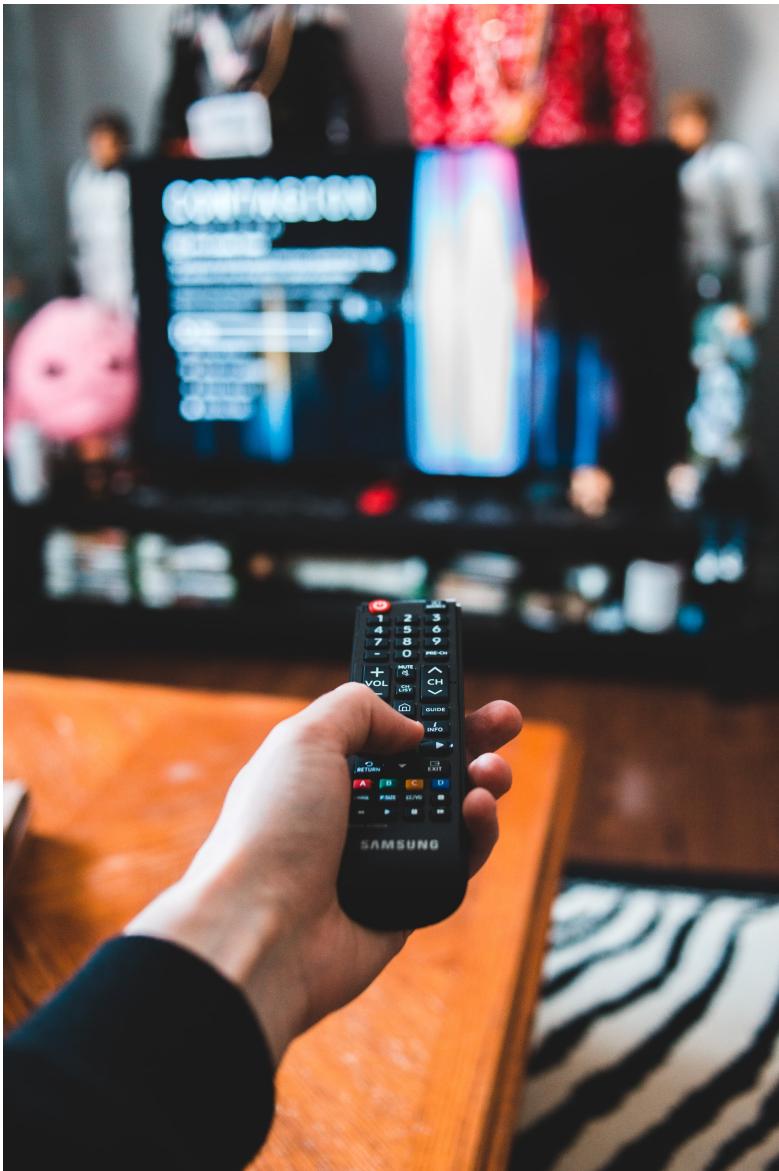
How much do you have experience in Python (5-very much, 3-moderate, 1-very little)?



By the end of the semester, you will understand the code below; You will acquire code literacy (=programming fundamentals)

```
1 # an example CSV data
2 df <- read.csv(file path)
3
4 # calculate the mean GPA among the students
5 # entered in 2020 by master program type
6 df %>%
7   filter(year == 2020) %>%
8   group_by(KDI master program) %>%
9   summarize(mean = mean(GPA))
```

# Hands-on Follow-along



[Code](#)[Issues](#)[Pull requests](#)[Actions](#)[Projects](#)[Wiki](#)[Security](#)[Insights](#)[Settings](#)[main](#) [Branches](#)[1 branch](#) [0 tags](#)[Go to file](#)[Add file](#)[Code](#)**Jae Yeon Kim** Merge branch 'main' of github.com:KDSS-DSPPM/r-fundamentals0f354d6 13 days ago [30 commits](#)[README.md](#)

Merge branch 'main' of github.com:KDSS-DSPPM/r-fundamentals

13 days ago

[software\\_setup.md](#)

add software install guideline

13 days ago

[README.md](#)

## R Fundamentals for Public Policy

Spring 2022

This course introduces and demonstrates how fun and exciting data analysis with R could be. Mastering computational tools and techniques not only enable social scientists to collect, wrangle, analyze, and interpret data with less pain and more fun, but it also let them work on research projects that would previously seem impossible. Due to the time constraint, the course focuses on data wrangling: the most fundamental and time-consuming component of the data analysis workflow. After completing this course, students will know how to turn messy raw data into structured datasets using the R programming language. In addition, the course will help students develop a framework for learning computational methods and using data to make the government work for everyone and empower citizens to negotiate power.

<https://github.com/KDIS-DSPPM/r-fundamentals>

Lecture notes (including in-class code examples) + the syllabus + the sign-up link for office hour



This course is  
highly useful  
but also hard;

Learn how to  
push yourself  
to the limit

# Community-building

R Fundamentals fo... ▾ 

- Threads
- More
- Channels
  - # general
  - # introduce-yourself**
  - # random
- Add channels
- Direct messages
  - Slackbot
  - Jae Yeon Kim you
  - Naveen Kori
- Add teammates

## # introduce-yourself ▾

+ Add a bookmark

with your team. [Edit description](#)

Today ▾



**Jae Yeon Kim** 12:44 PM

joined #introduce-yourself.



**Jae Yeon Kim** 12:56 PM

Please introduce yourself here. There's no fixed format but at least mention your name and program. I will go first. I'm Jae Yeon Kim. I am an Assistant Professor at the KDI School of Public Policy and Management, and I teach computational social science and civic data science. I received my PhD in political science from UC Berkeley, where I was Senior Data Science Fellow at D-Lab, and was a pre-and post-doc at the SNF Agora Institute at Johns Hopkins University, where I'm still affiliated as a researcher. I love using data to improve public policy, empower citizens, and strengthen democracy.



**Naveen Kori** 5:41 PM

joined #introduce-yourself.



**Naveen Kori** 5:44 PM

I am Naveen Kori from India pursuing MDP (2021-22). Looking ahead for a good learning experience! Thank you.



Unless absolutely necessary, I will use this Slack workspace for the Q&A. Check the announcement for the invite link.



**Do Googling (Stack  
Overflow),  
Ask questions,  
Help Each other**

# Syllabus review

# What to expect:

- 1. Lectures:  
cloud-based**
- 2. Assignments:  
R markdown**
- 3. Lots of surveys (KSS)**

# Get to know each other:

1. Preferred name,
2. Program & Year,
3. Something interesting about  
yourself
4. Motivation for taking the class

To do:

1. Fill out the background survey
2. Introduce yourself on Slack