



## Details

Calling all future data scientists! Join the Student Alumni Associates (SAA) for Meet the Boss: Data Science, an evening of FREE food and networking! This quarter, we're exploring the field of data science and how it has found its way into the careers of UC San Diego alumni. Our alumni panelists will represent a variety of industries and are excited to share how they found their paths within the field!

This event will feature a moderated Q&A, and after, we will break out into a night of networking. Free food and refreshments will be served.

GRAB A TICKET/RSPV HERE: [bit.ly/ucsdmtb](https://bit.ly/ucsdmtb)

APR  
29      Meet the Boss: Data Science  
Public · Hosted by UCSD Student Alumni Associates

Interested

Invite

Monday, April 29, 2019 at 6:00 PM – 8:00 PM PDT  
Next Week · 14–18°C Rain Showers

The Loft @ UCSD  
9500 Gilman Dr, La Jolla, California 92093

Show Map

Find Tickets

Free  
[bit.ly](https://bit.ly/ucsdmtb)

Our panelists:

Orysya Stus '18: Currently a data scientist at Seismic Software. She obtained both a Bachelor's in Bioengineering and a Master's in Data Science and Engineering at UC San Diego!

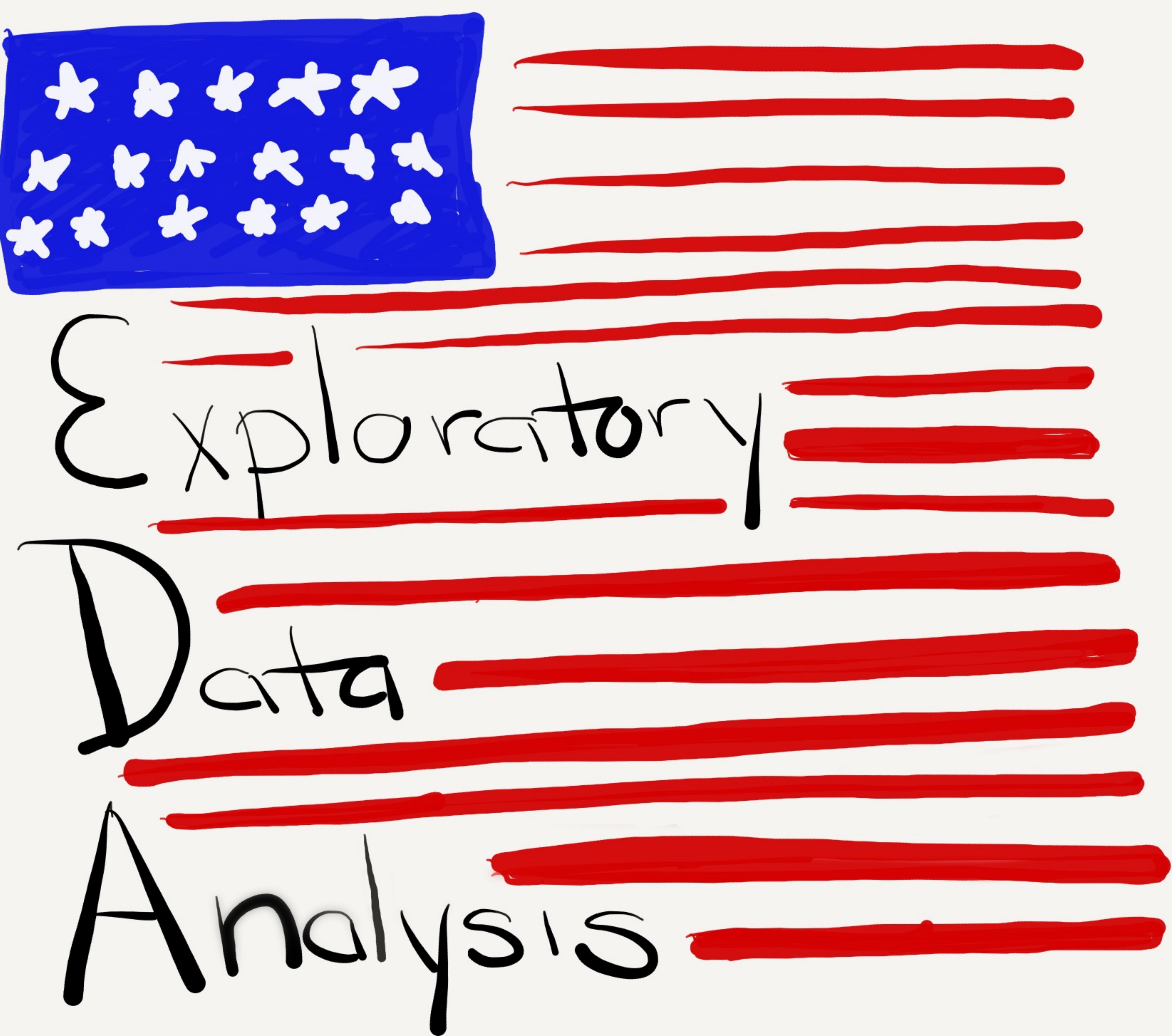
Talina Konotchick, PhD: Talina is a data analyst at Illumina with a background in Bioinformatics and Oceanography!

More speakers will be announced soon!

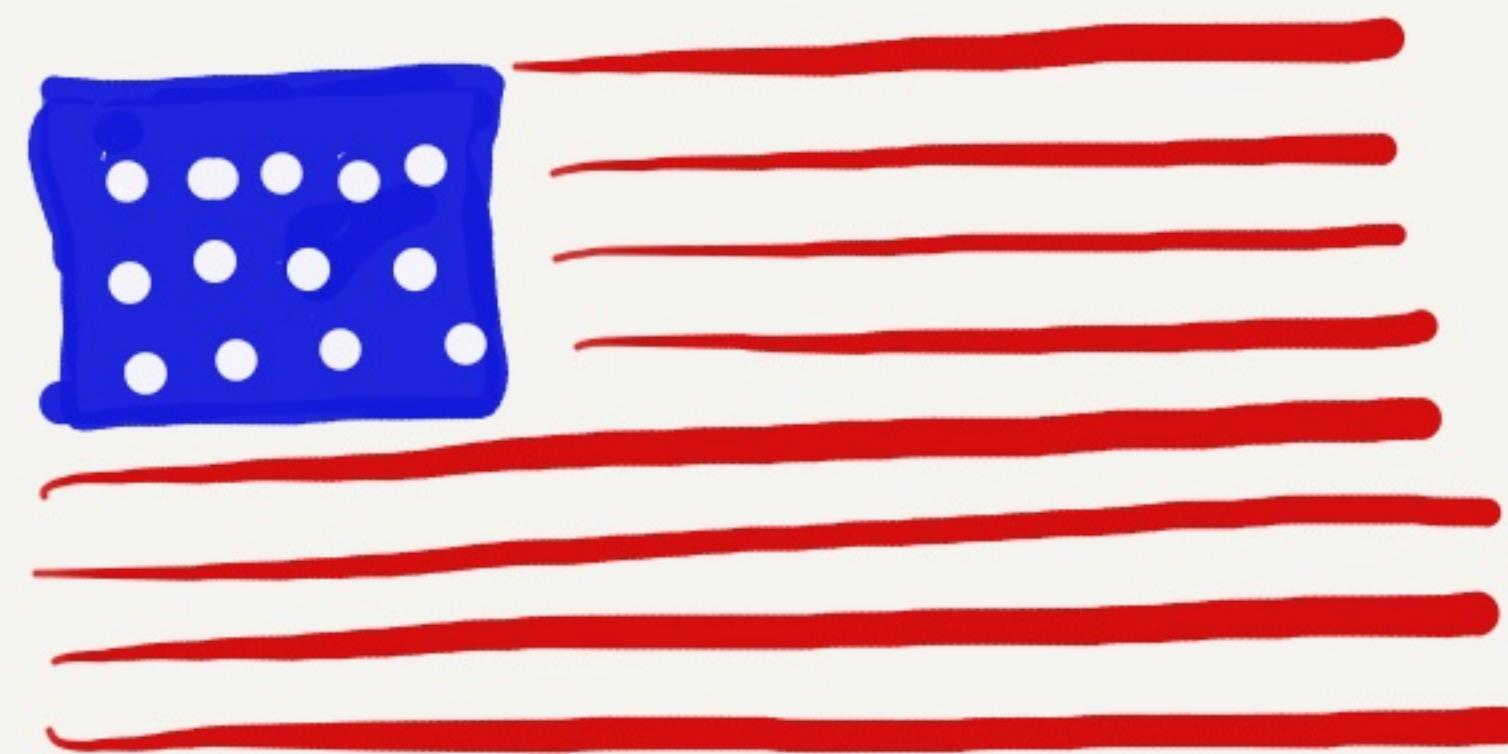
This event is for UC San Diego students only.  
Questions? Email [saa@ucsd.edu](mailto:saa@ucsd.edu)

Facebook: <https://www.facebook.com/events/652097685248663/>

Eventbrite: <https://www.eventbrite.com/e/meet-the-boss-data-science-tickets-60390412349>



- Brainstorm + Question
- Discuss Analytical Approach
- Communicate



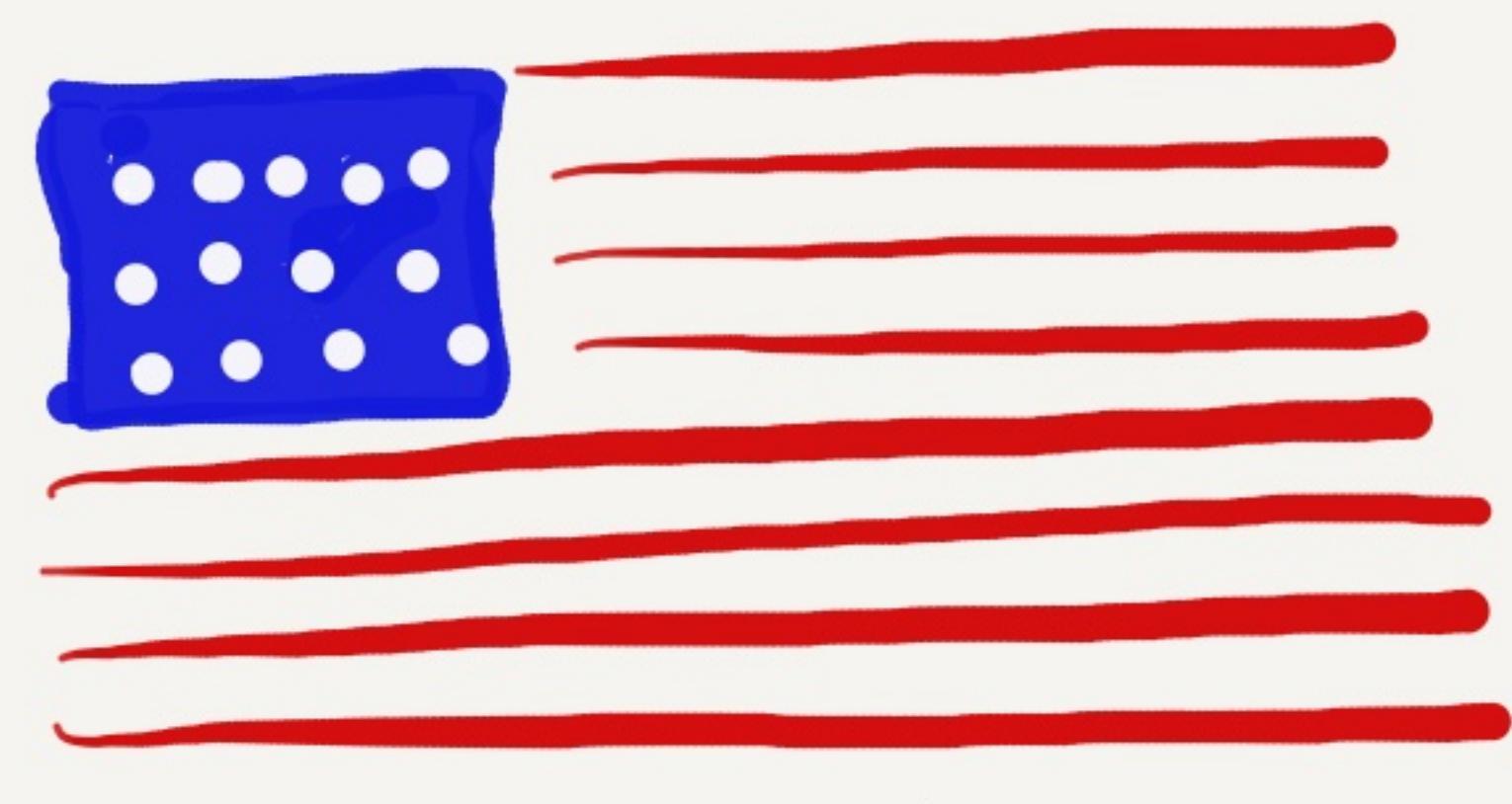
## Question

Data Science  
Question?

- actively reflect constituents?
- mean age
- differences in policy by age
- freshman congress person has less experience
  - take learning curve into account?
  - navigate bureaucracy better w/ experience?
    - how much experience?

# Does Congress Have An Age Problem?

- "they're too old"
  - avg age of congress members
    - this related to negative in society (i.e. crime rate)
      - over time
  - age related to competency?
    - older not a problem in itself
      - i.e. technology



## Question

Data science  
question:  
what to  
measure?

# Data Science Question

Age + Policy

- # drafted legisl. pieces
- votes

Age + Pass/Fail

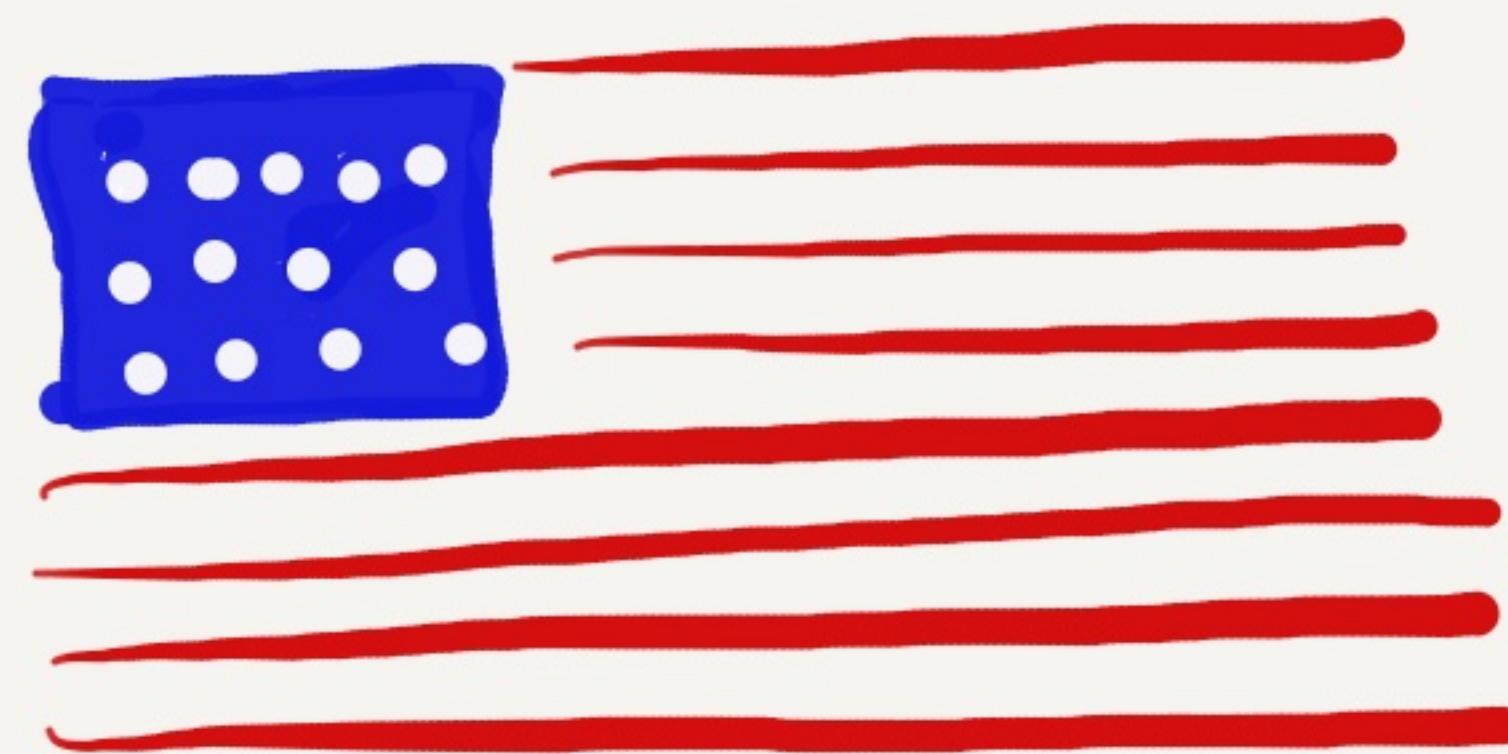
→ actually accomplish

Age + approval rating

Age + time in office

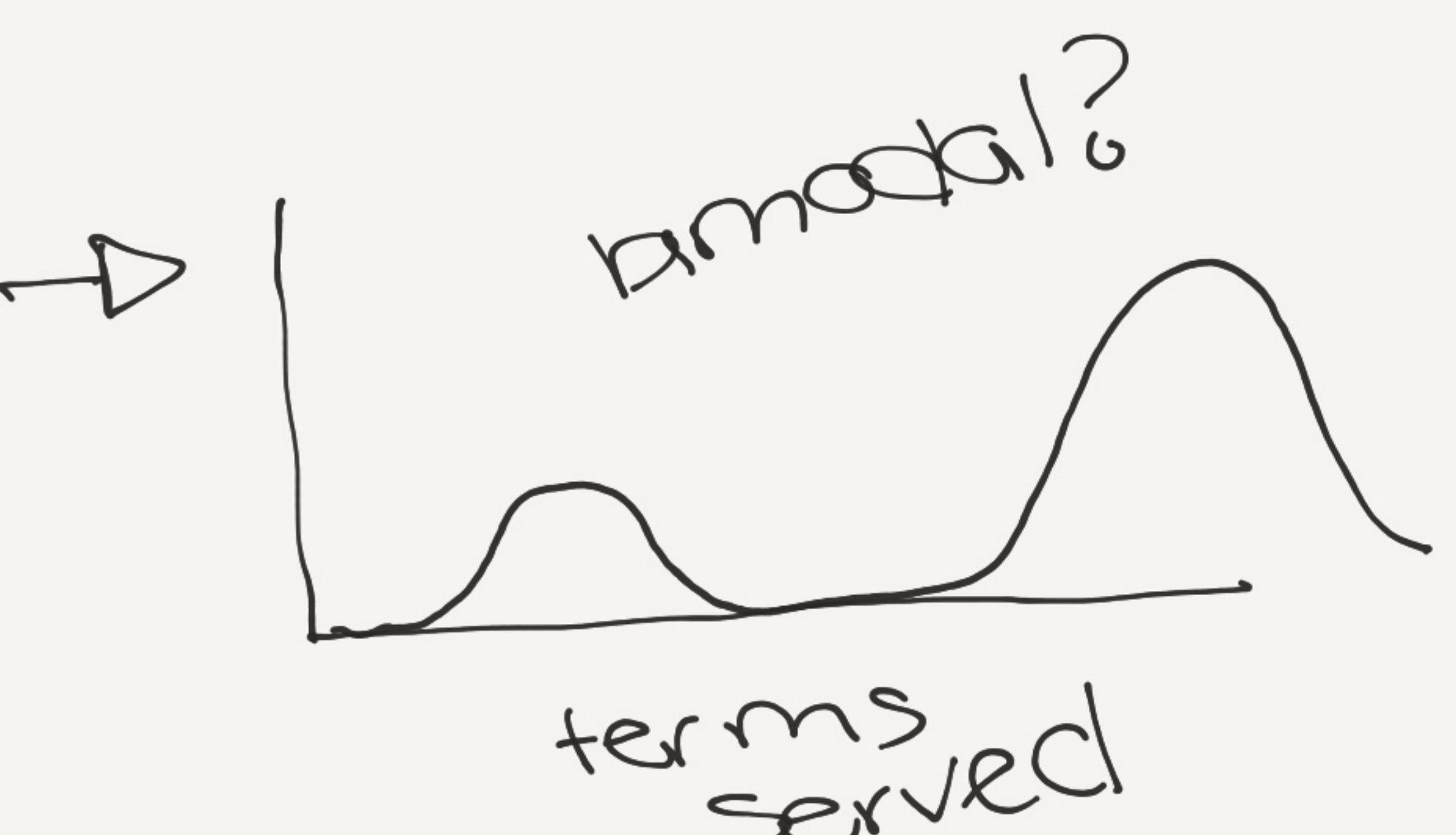
- ↑ older = more productive due to  
time in job

Age of Congress people  
over time.



# What do we already know about this topic?

## Background

- "People in Congress are old + they don't know what the Internet is"
  - Age of Congress is higher than min age requirement
    - 25 for House
    - 30 for Senate
  - Party affiliation
    - 2 party system
      - o D
      - o R
      - o others ( )
- 2-years in term limit
- { House : 435 (population)  
Senate : 100  
→ (2 per state)
- 6 years; 535 members  
no term limit
- 



What would you  
think the answer  
to these questions is?

overall?

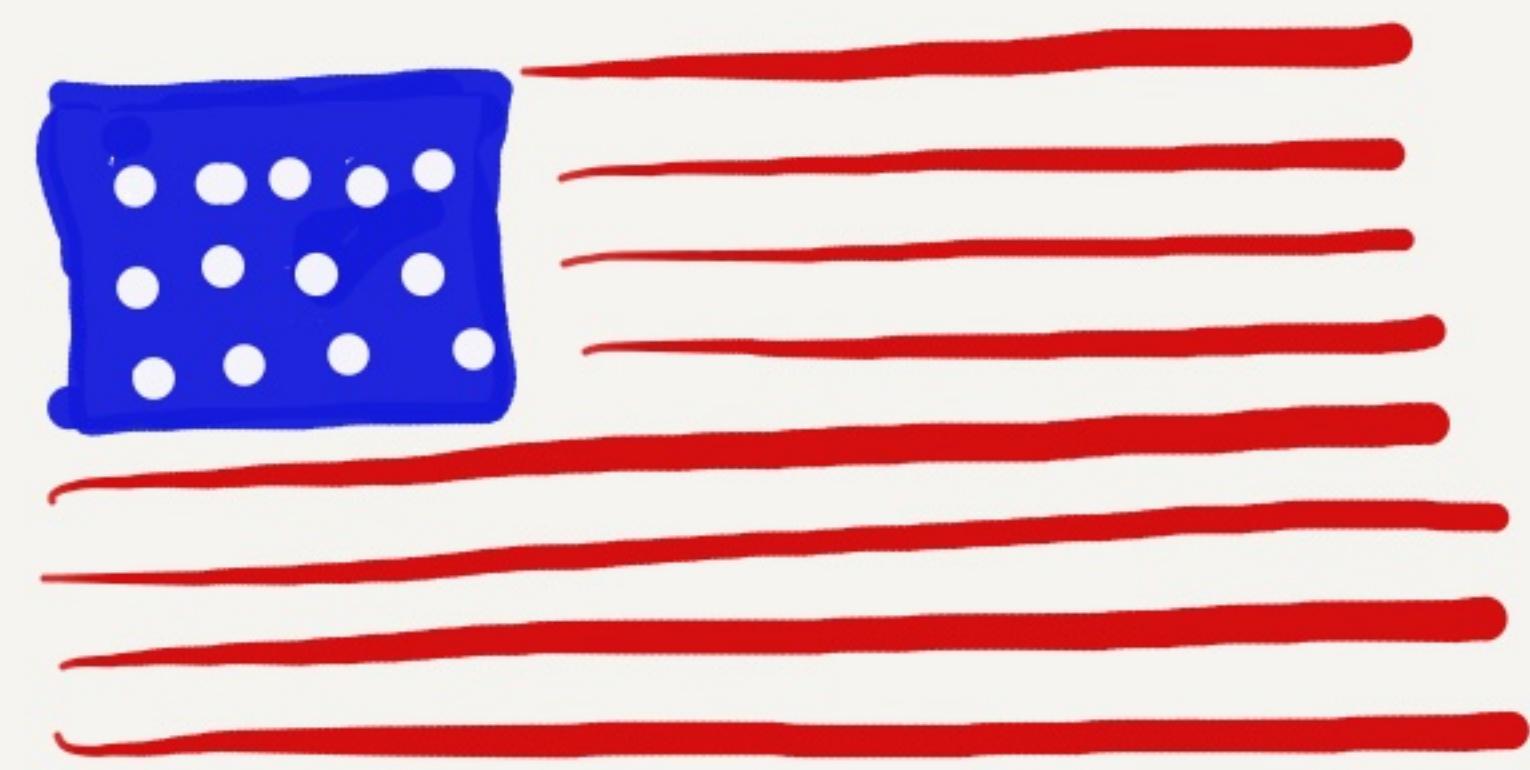
- decrease; younger now than historically (~50y)
- increase over long term
  - big age of life expectancy
  - + ↑ retirement age
- by party
  - ↓ mean age
  - democratic lower than Repub

trend over time?

House v.s. Senate?  
- older + longer term

↑ Senate > House  
age

- Gap in competency by age
- increase ineffectiveness over time; level off



# What data/information would we need to answer this question?

## Data

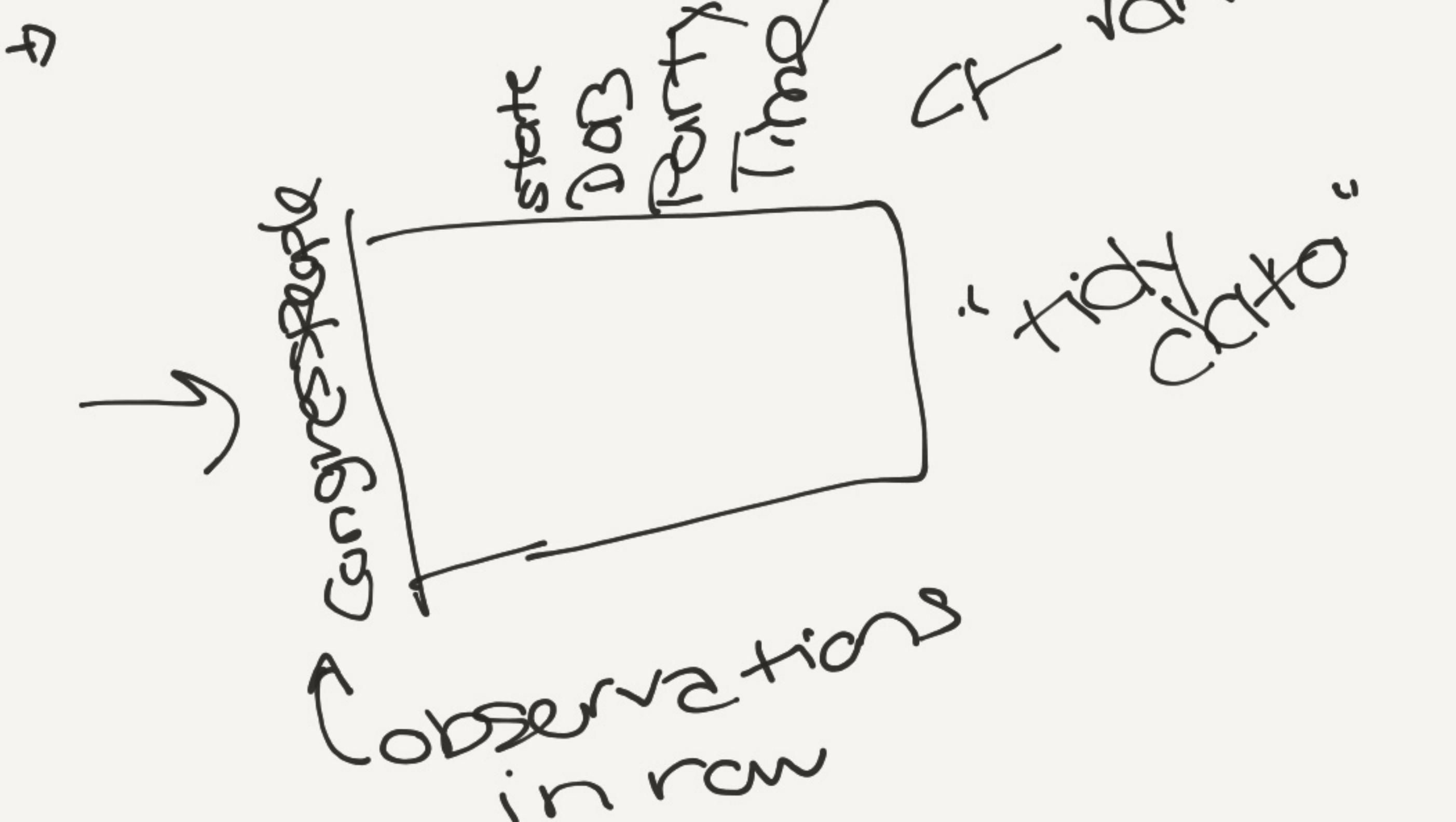
- Polling data
  - in touch w/base?
  - approval rating by age
  - issues important by age
- Items of legislation
  - drafted
  - discussed in committees
  - voted on
  - passed?

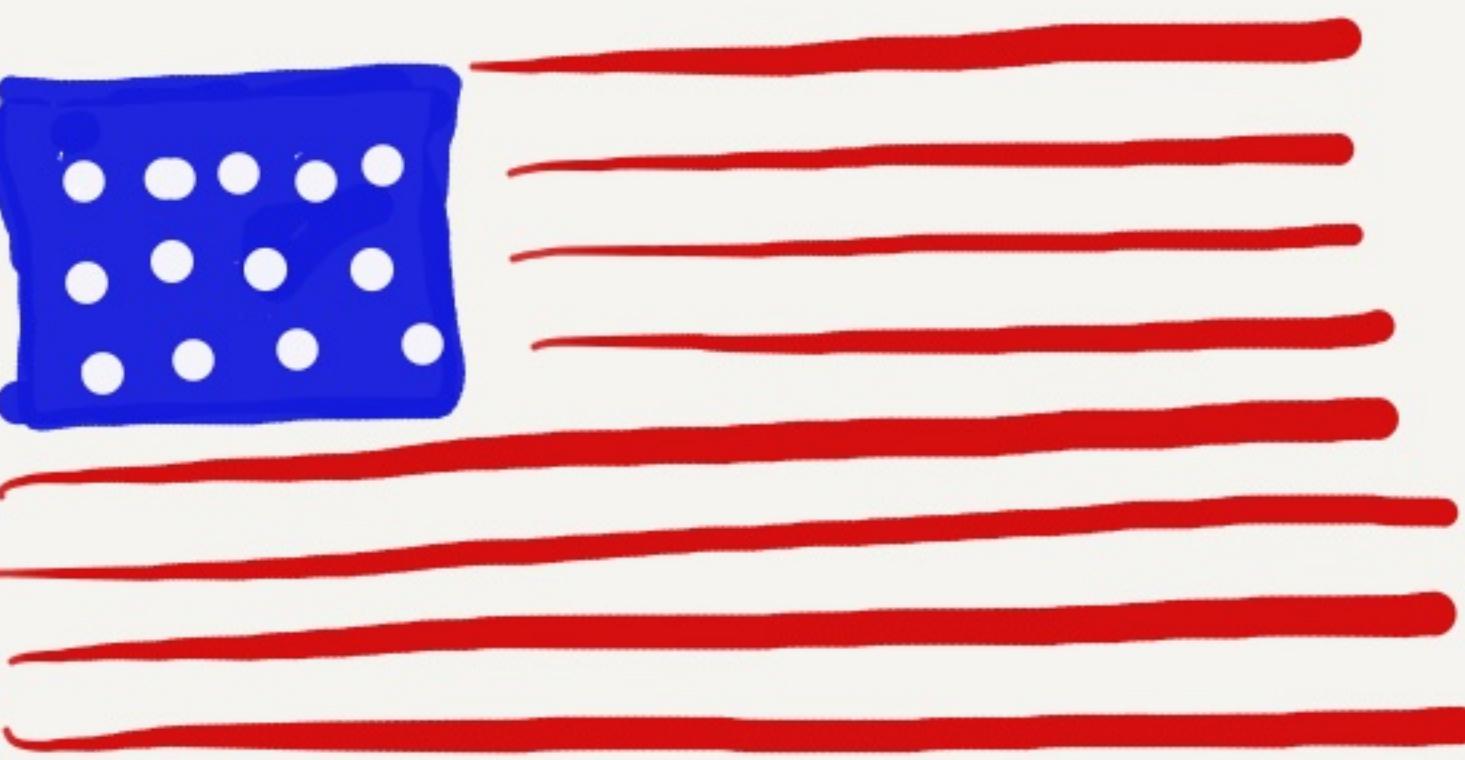
multiple state  
states

## Congress people over time

- Age (DOB)
- Party
- Time (when? how long?)
- ~~for~~ House/Senate?
- Educational/Career Background
- SES/wealth
- State/Region?

## Other factors?





# What Python tools would we use?

## Setup

package)

- numpy
- scipy (analysis)
- matplotlib (other)
- pandas (dataframes)  
→ wrangling

- get the data
- web scrape (BeautifulSoup?)
- census data? (Age of Pop<sup>n</sup> ....)
- Congress website? (congresspeople)
- API (newsources?)

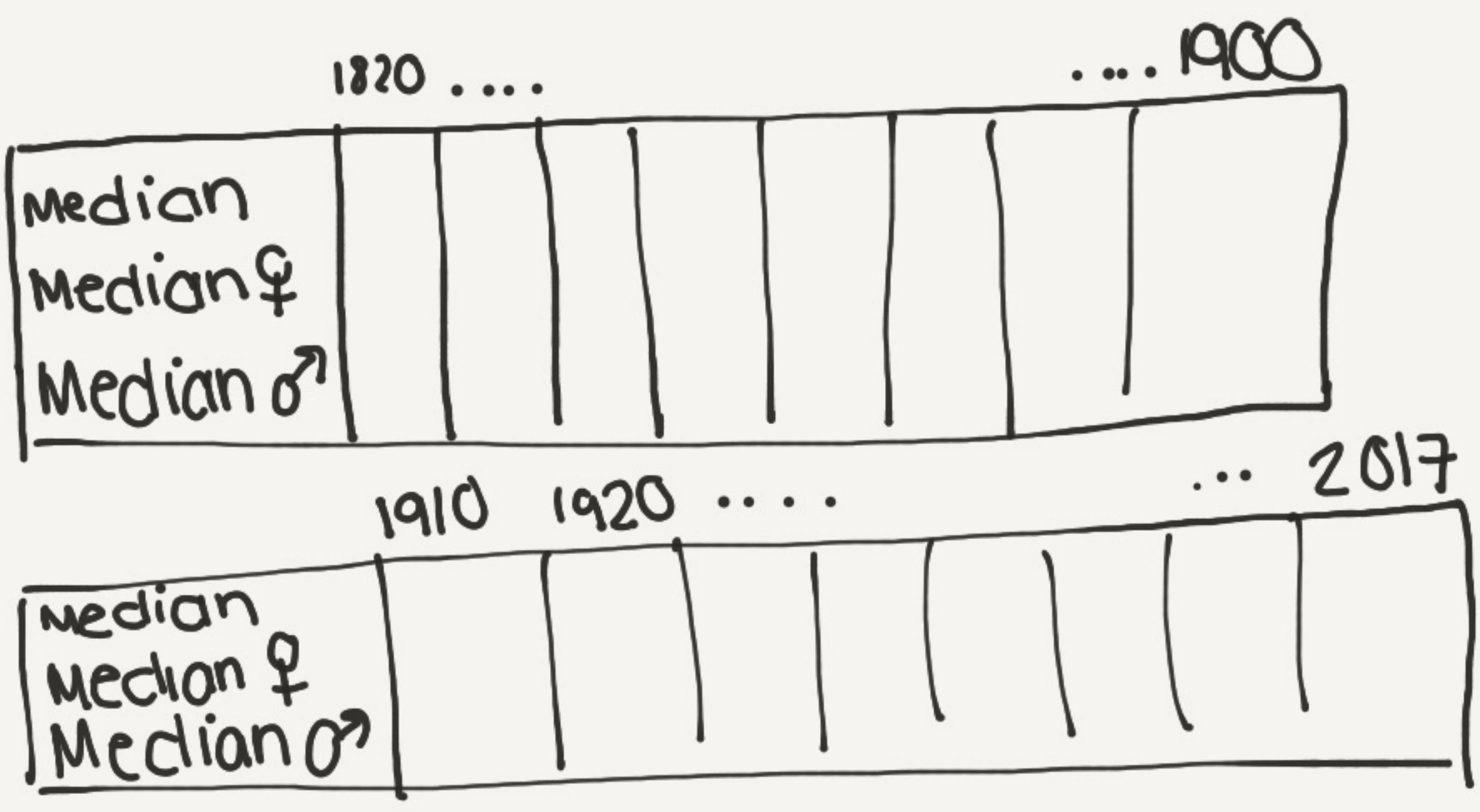


## REVIEW: DAY2

### Tools:

matplotlib/seaborn  
pandas,  
sklearn?  
beautifulsoup?  
numpy

### Median US Age (from Wikipedia)



General Question: Does Congress have an age problem?

Data Science : (1) Does the average age of Congress reflect the average age of the US ?

- (2) What has this trend looked like historically?
  - differ by chamber? (House v. senate)
  - differ by party? (D vs R)

### Congress Data Set

Congress [80 - 113]	chamber (house ; senate)	first name	last name	birthday (YYYY-MM-DD)	state (two letter)	Party (D; R; AL; L; I; ID)	incumbent (Yes; No)	term start (YYYY-MM-DD)	age (years); i.e. 85.9)
---------------------	--------------------------	------------	-----------	-----------------------	--------------------	----------------------------	---------------------	-------------------------	-------------------------



## Descriptive Analysis

trend

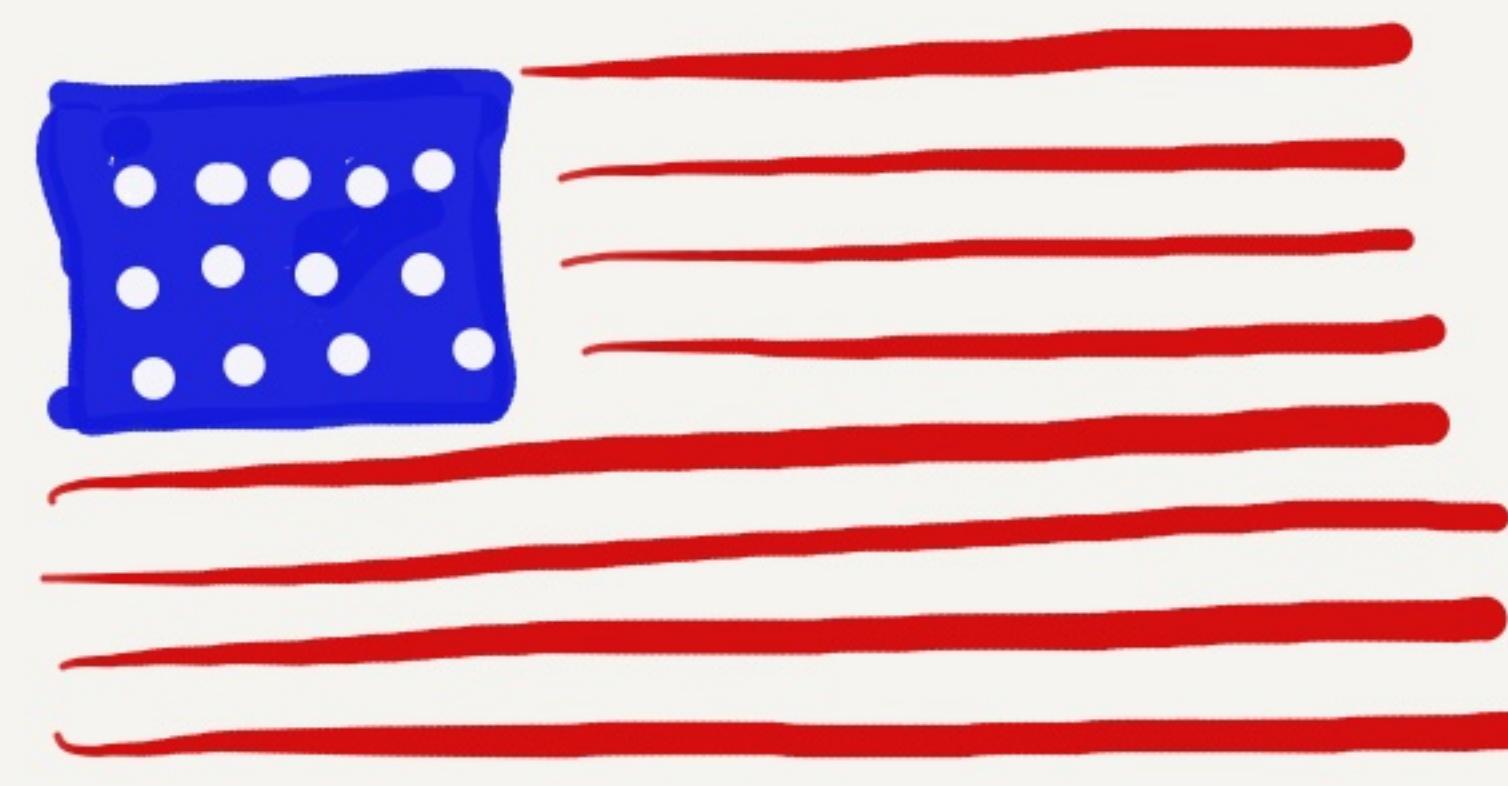
- each year, avg age of congress  $\pm$  sd

wish list

voting history

What would you do to summarize these data?

- size of dataset:  
18,635; 13 columns
- center + spread of all quant
  - \* age (mean: 53y)
  - \* length of term
- avg starting age (age+incumbent)
- proportion of freshman
  - by year
  - # of terms served by individual
- missingness
  - by party
  - by chamber
  - over time
- issues: diff years
  - medians over time  
(granularity: summarized w/n year)

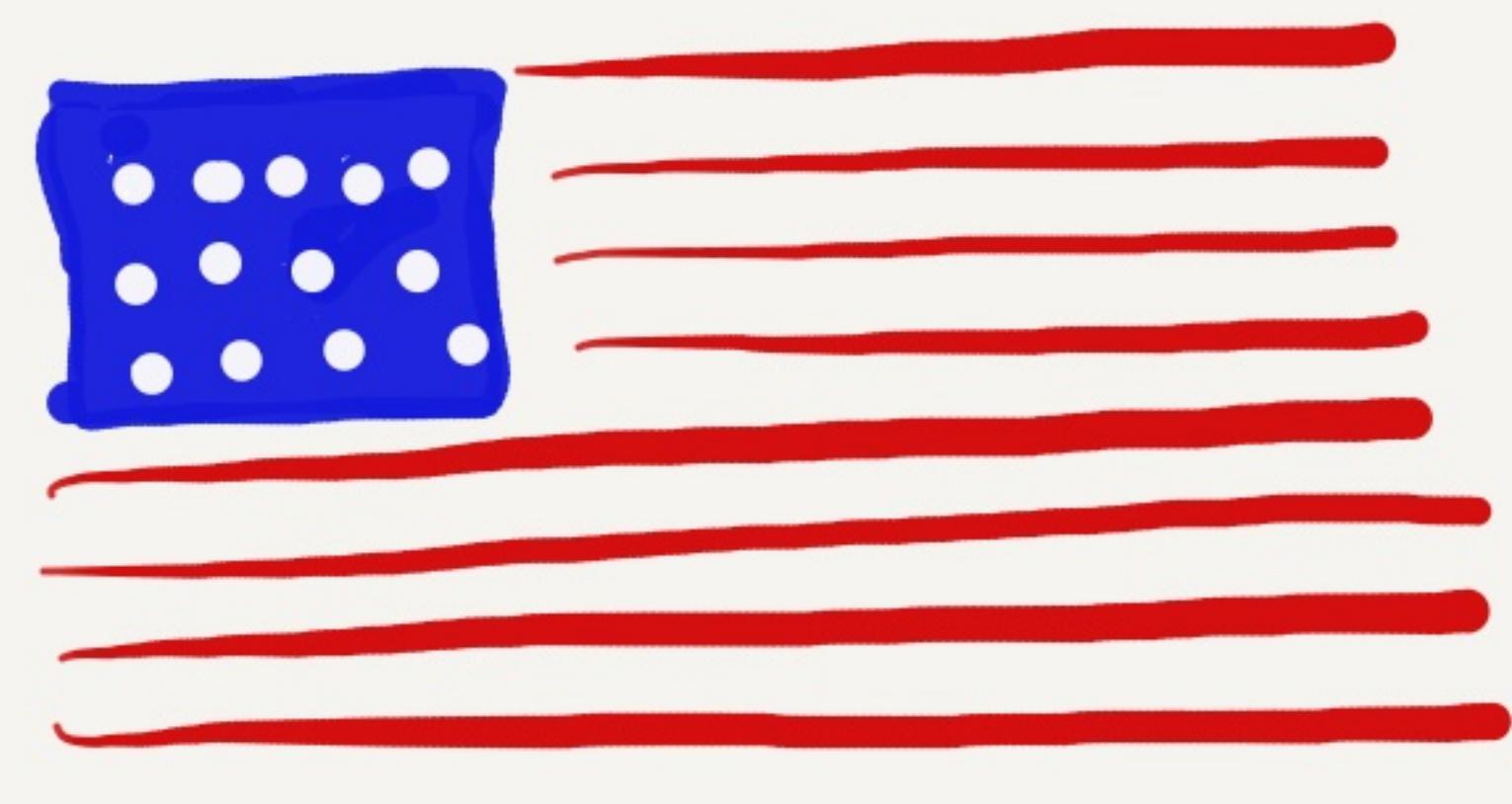


EDA

brainstorm  
a list...  
etc.

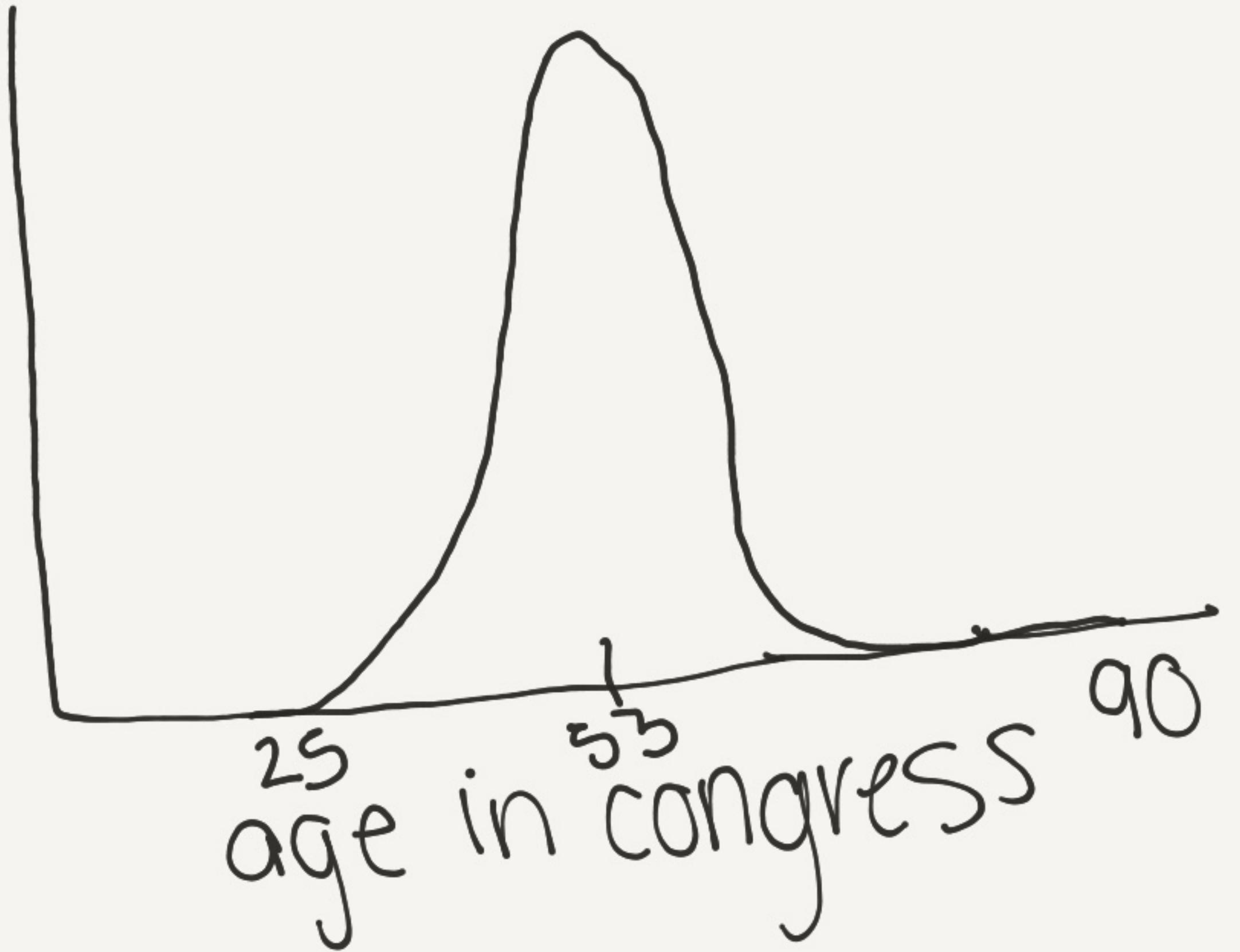
How would you explore  
the data?

- individual dist' quant
  - histogram
  - density plot
- age vs. years
  - overall
  - D vs R



EDA

How would you explore  
the data?



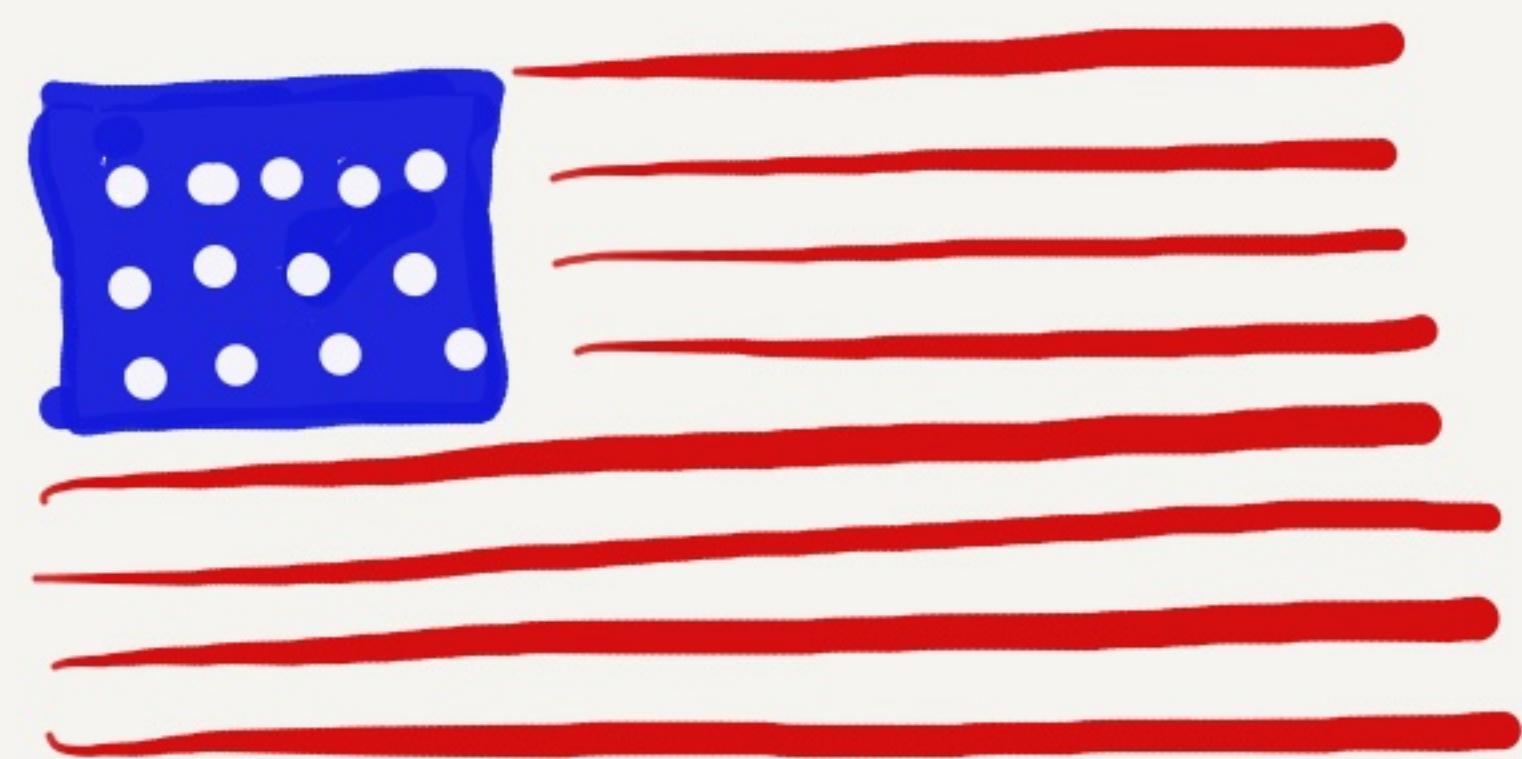
US Population

Median: 33y  
2017

limit / caveat:

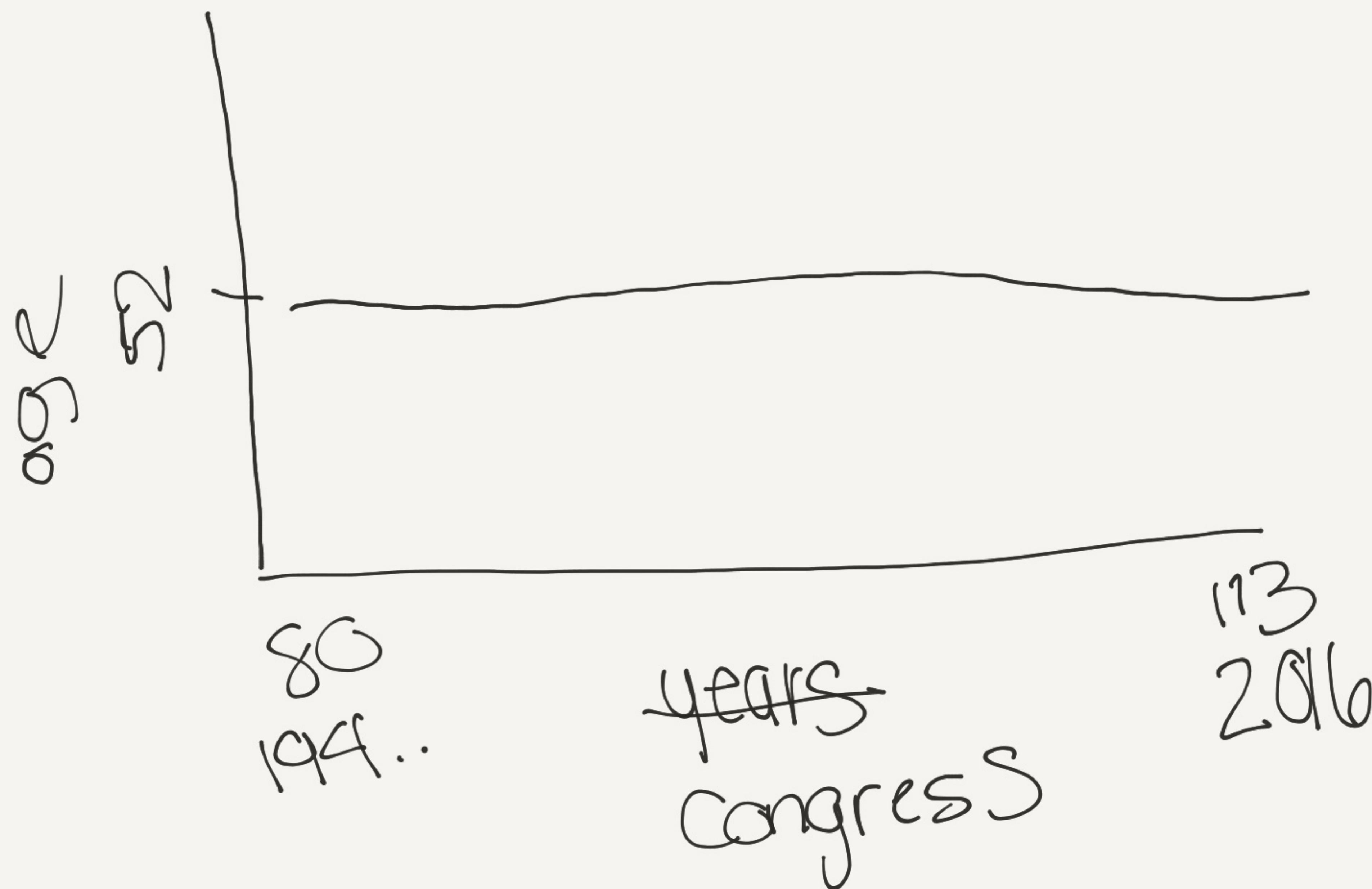
→ didn't the  
parties  
flip flop?

→ only consider  
adults  $\geq 25$   
or  $\geq 30$   
(senate)

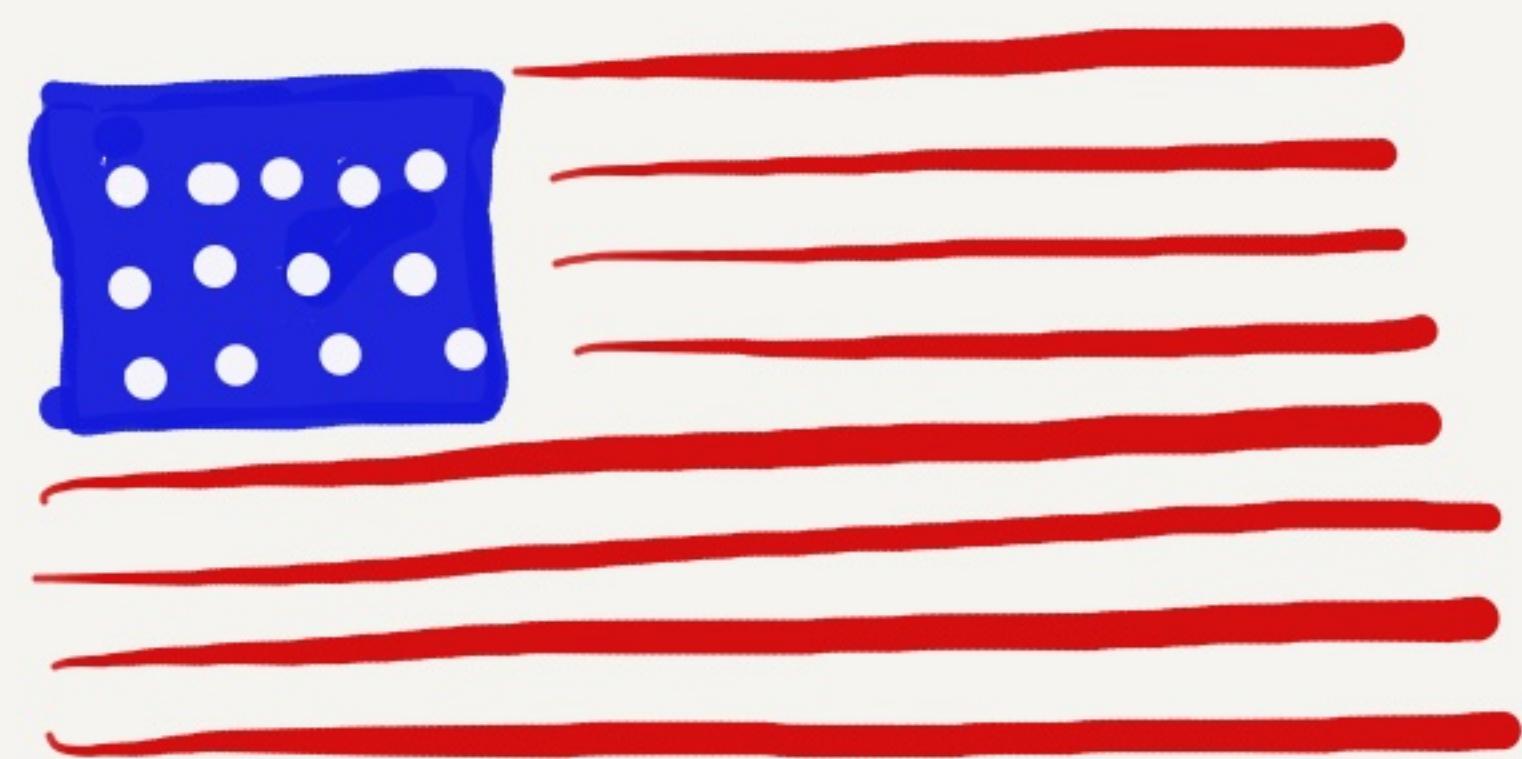


EDA

What exploratory visualizations would you generate?



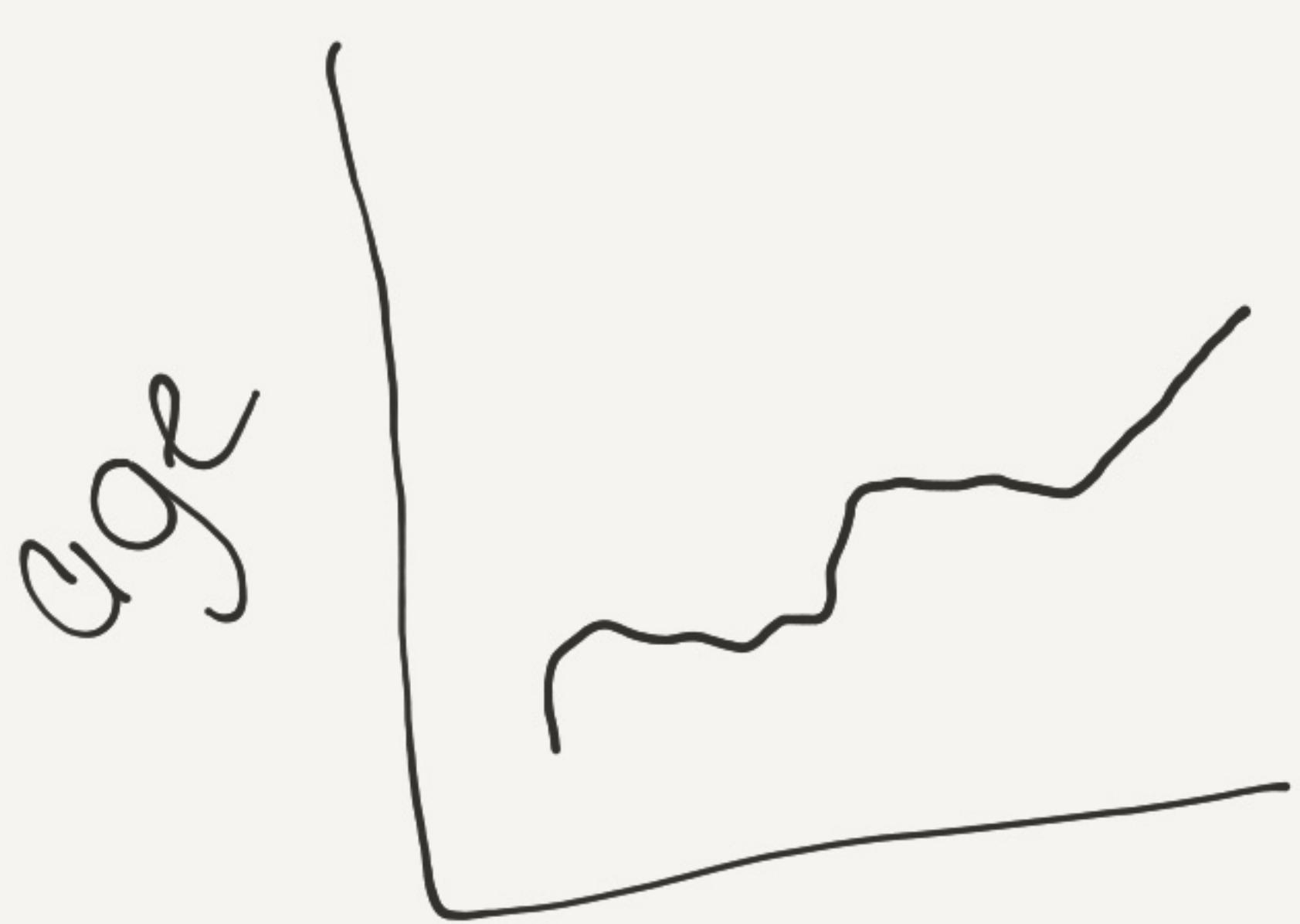
- unlikely
- incumbency / time ↑
- life expectancy



What exploratory visualizations would you generate?

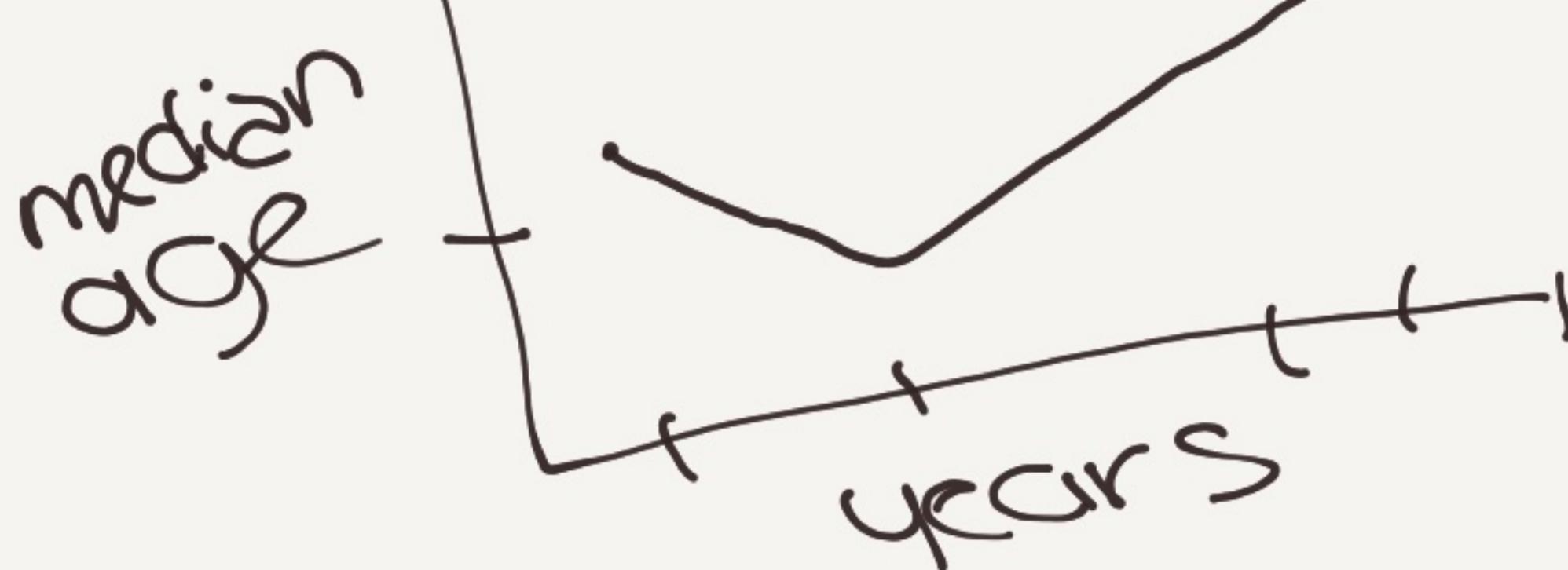
EDA

Congress



gears

us population



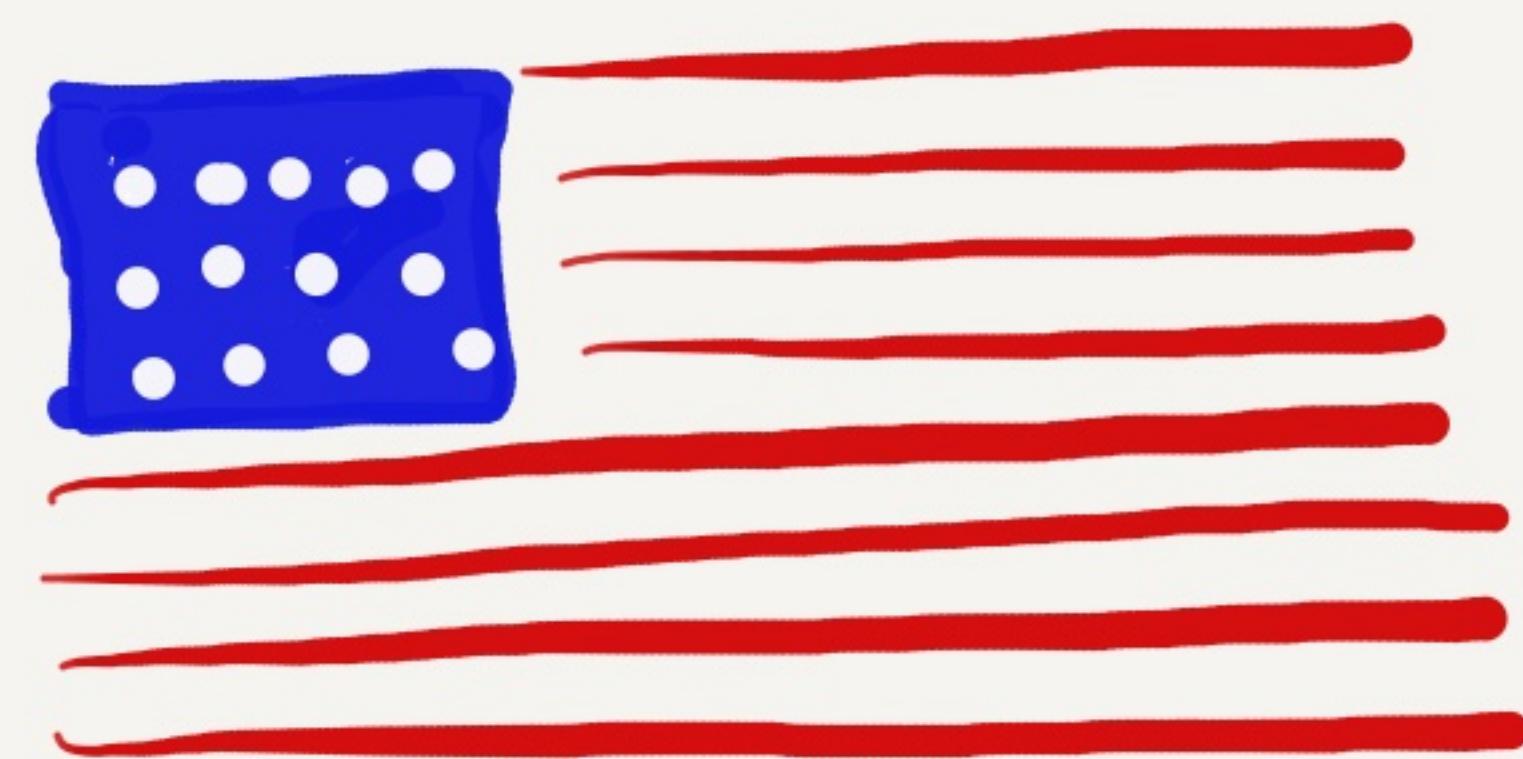
years

age

age

years

— R  
— D

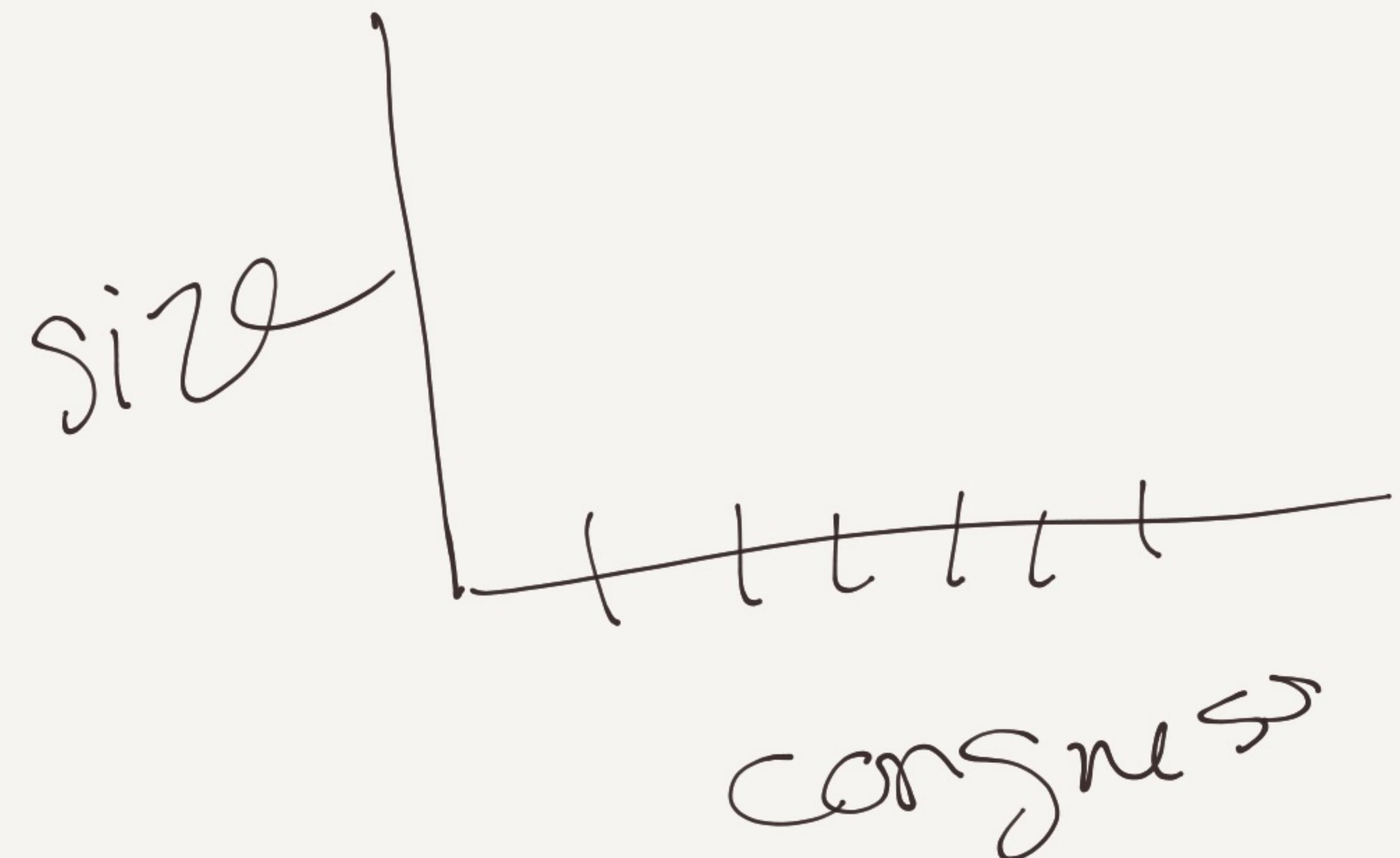


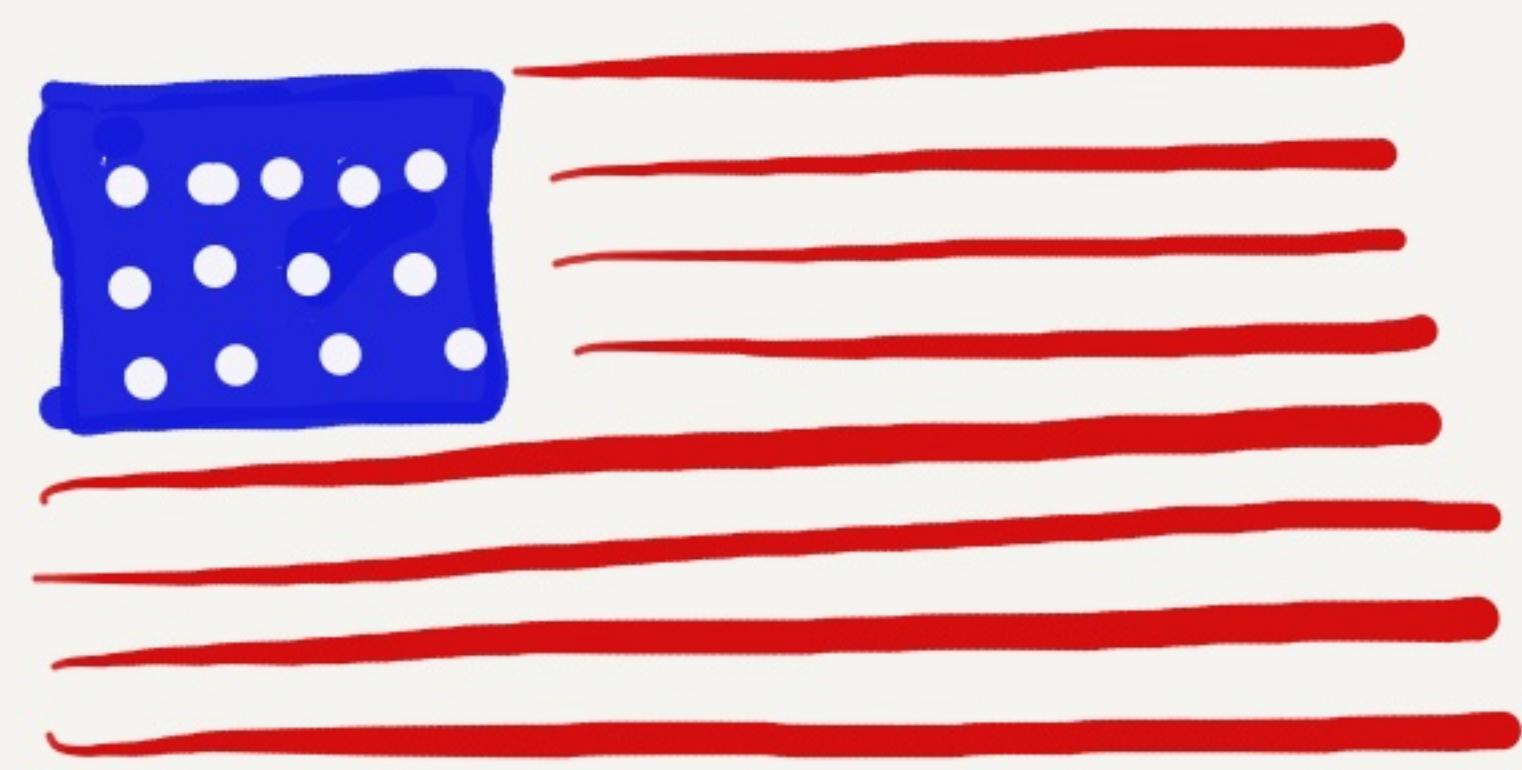
EDA

What exploratory visualizations would you generate?

→ size over time

→ ↑ House





EDA

typical  
term per  
Congress?

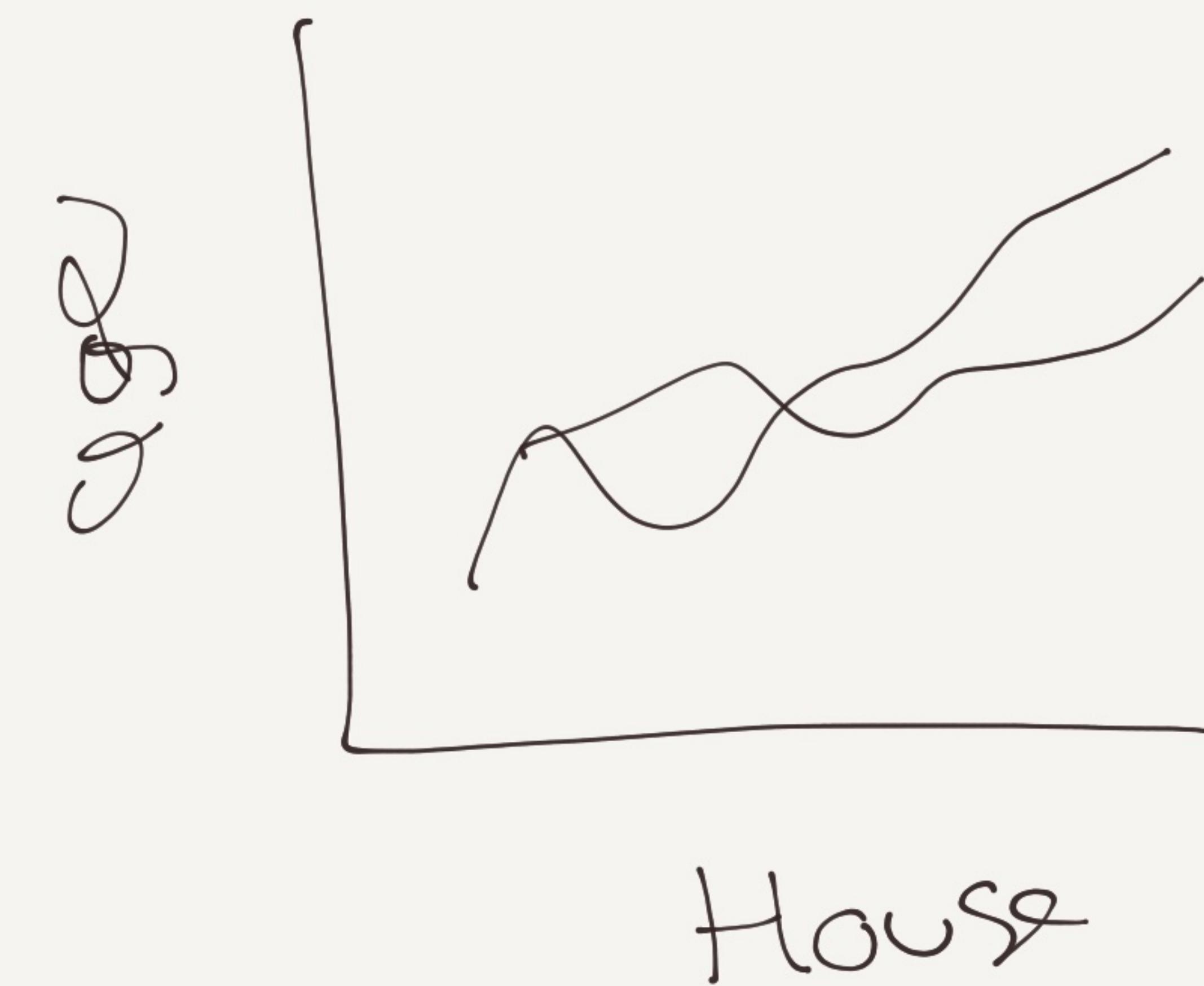
What exploratory  
visualizations would you  
generate?

wish list

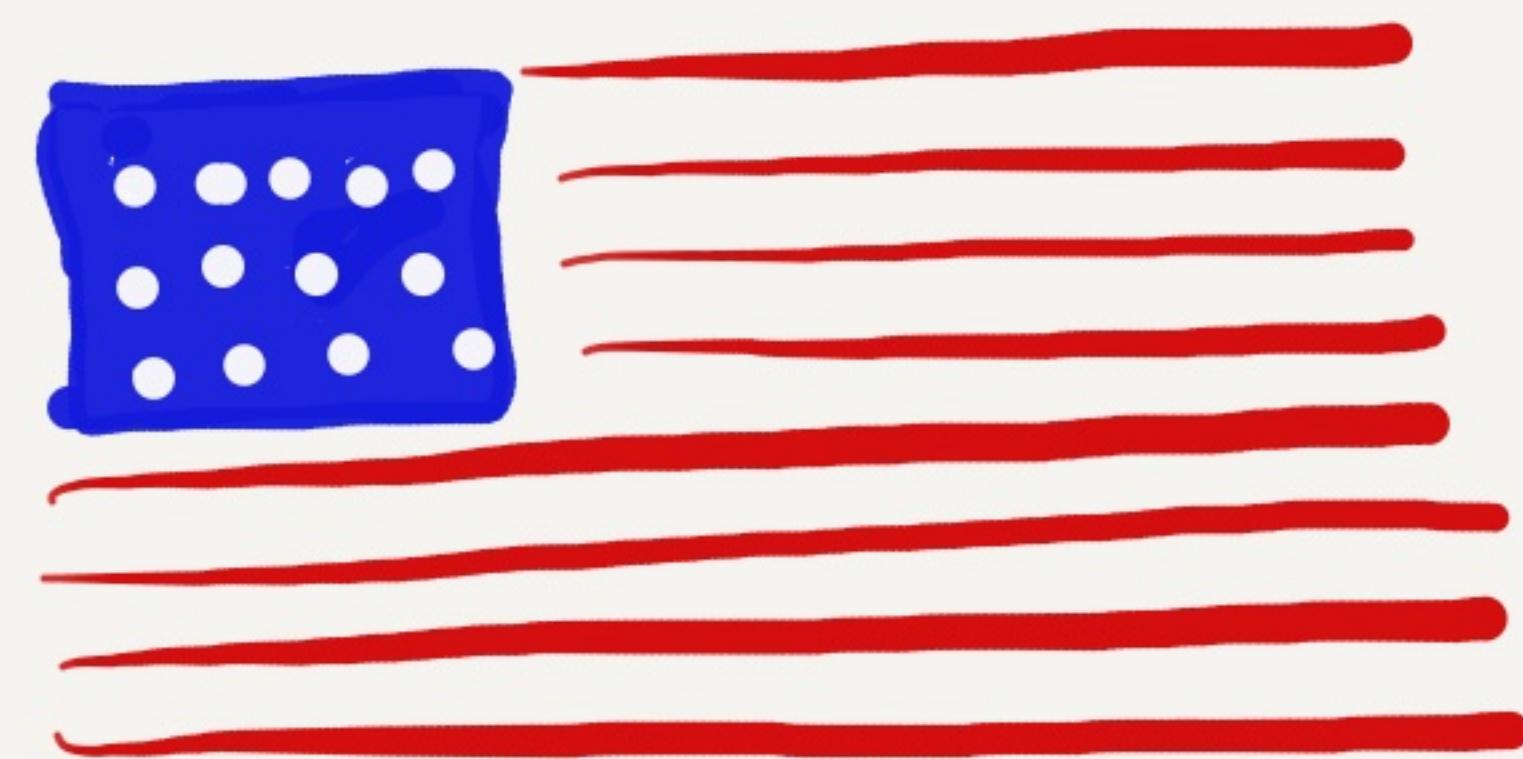
# of pieces  
of legislation  
each year



Senate



House



Communication

- publications  
→ detailed report

## How would | could you share your results?

- wikipedia / newspaper
- advocacy group
  - Twitter / Medium <sup>(blog post)</sup>  
(big reach)
- job : whoever paid
  - executive report
  - detailed

Know your audience