GENERAL ASSEMBLY

INTRO TO DATA SCIENCE

LECTURE 4: PRESENTING, SCRAPING, APIS

- **QUESTIONS ABOUT LAST CLASS** (ALREADY REVIEWED EXERCISES)
- **HEADS-UP ASSIGNMENT #1**

I. SCRAPING & APIS (OPTIONAL – SIT BACK AND RELAX)
II. HOW TO PRESENT YOUR INSIGHTS
III. VISUALIZATIONS & MATPLOTLIB (EXERCISES)

1. **INTRO, UNIX**
2. **SQL, PYTHON**
3. **PYTHON, PANDAS, MATPLOTLIB**
4. **MORE GATHERING (WEB SCRAPING, APIS) AND HOW TO PRESENT YOUR INSIGHTS**

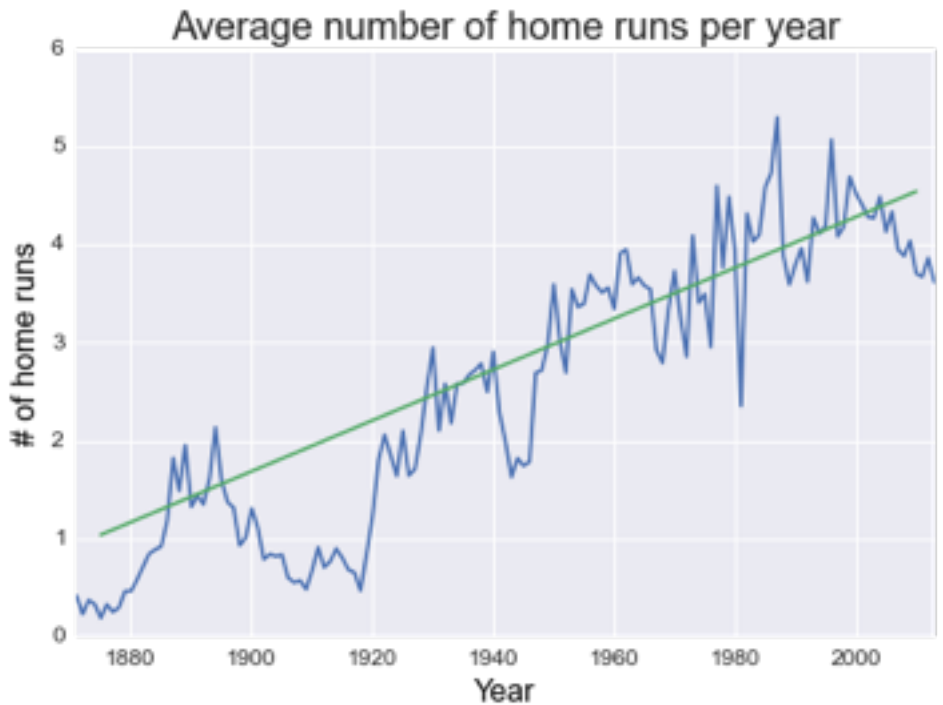
CLOSED BY ASSIGNMENT #1 (DUE SUN 7/12)

- GATHER A DATASET
- POSE A FEW QUESTIONS
- PROCESS THE DATA IN PYTHON AND PANDAS
- GENERATE DESCRIPTIVE STATISTICS
- CREATE VISUALIZATIONS
- SUBMIT YOUR WORK (DUE **SUN 7/12 MIDNIGHT**)
- PRESENT IN CLASS (**TUE 7/14**)

- GATHER A DATASET
- POSE A FEW QUESTIONS
- PROCESS THE DATA IN PYTHON AND PANDAS
- GENERATE DESCRIPTIVE STATISTICS
- CREATE VISUALIZATIONS
- SUBMIT YOUR WORK (DUE **SUN 7/12 MIDNIGHT**)
- PRESENT IN CLASS (**TUE 7/14**)

Your submission should include
- code in python (.ipynb or .py)
- results in clear plots
- *optionally,* a fancy presentation
- **not** the dataset if it's large

HOME RUNS GET MORE AND MORE COMMON



- One message per slide
- Conclusion in header (so what?)
- Clear labels in chart (title, axes)
- Trend line emphasizes message
- Source at the bottom

Think of your narrative:
your slides should form a story

Source: Lahman baseball database

I. SCRAPING & APIS
(OPTIONAL: SIT BACK & RELAX)

- Go to github.com/ga-students/DAT-23-NYC
- Scroll down to lesson #4
- Open the *Web scraping* notebook
- Have a look at the twitter_stream.py code (and run it!)

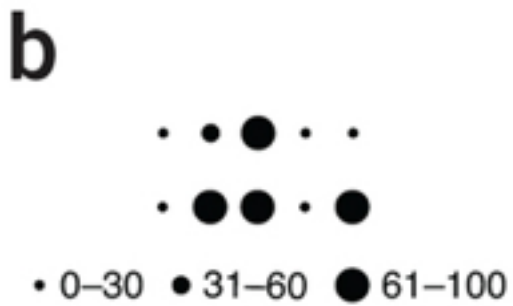
II. HOW TO PRESENT YOUR INSIGHTS

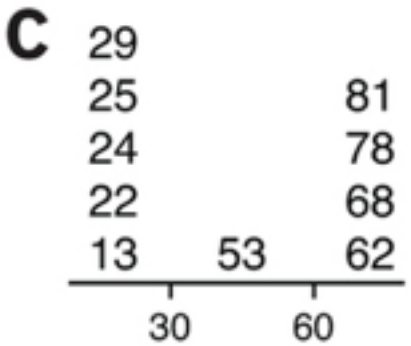
III. VISUALIZATIONS
MATPLOTLIB, SEABORN, VINCENT

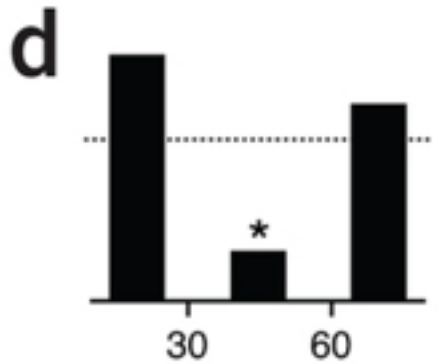
a

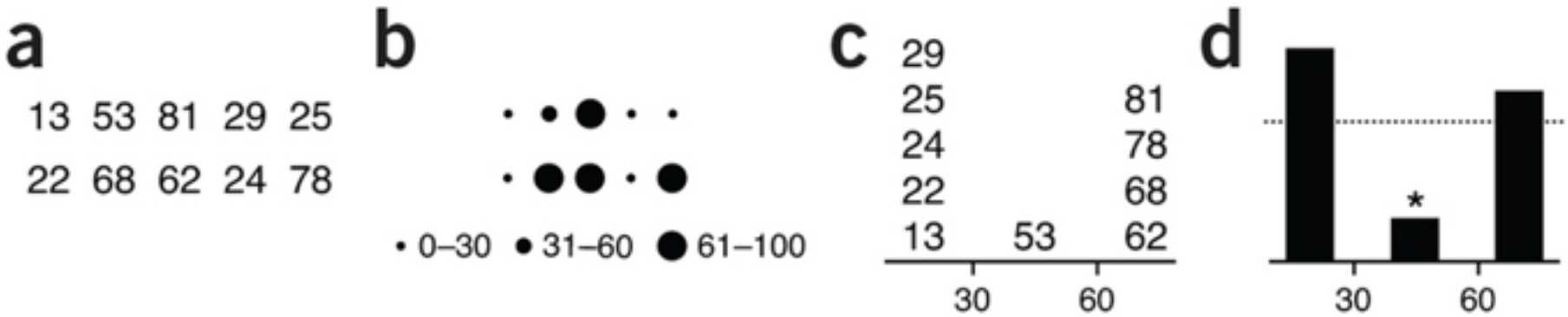
13 53 81 29 25

22 68 62 24 78

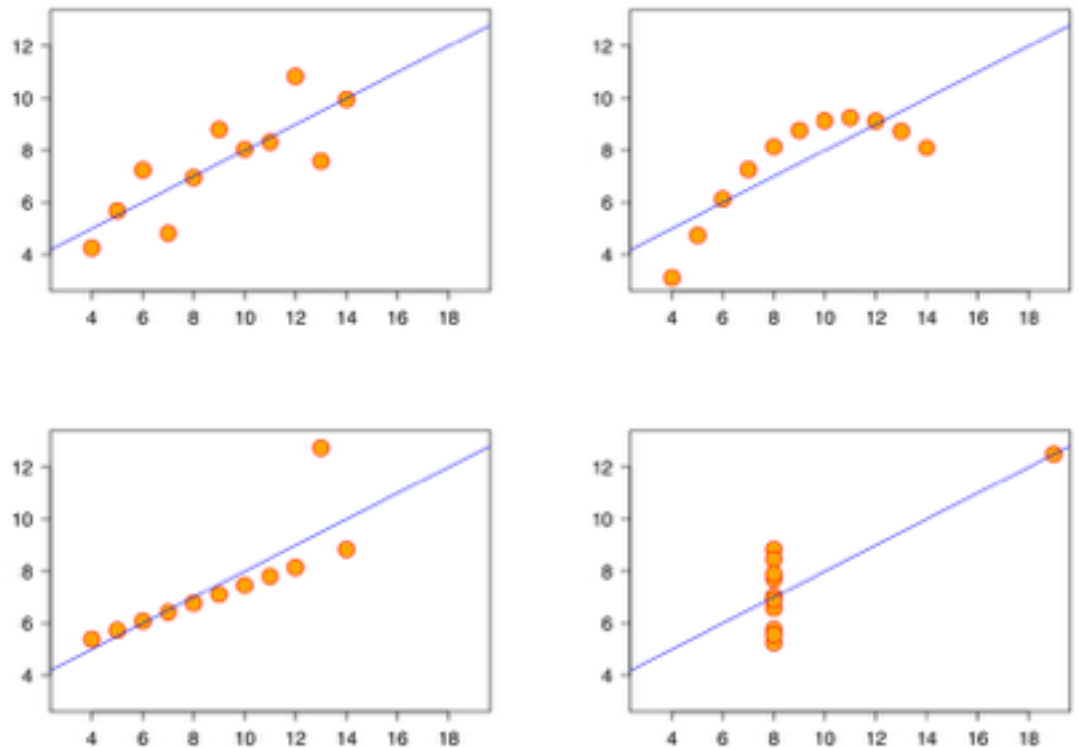

b
0–30
31–60
61–100


C
29
25
24
22
13
81
78
68
62
53
30
60

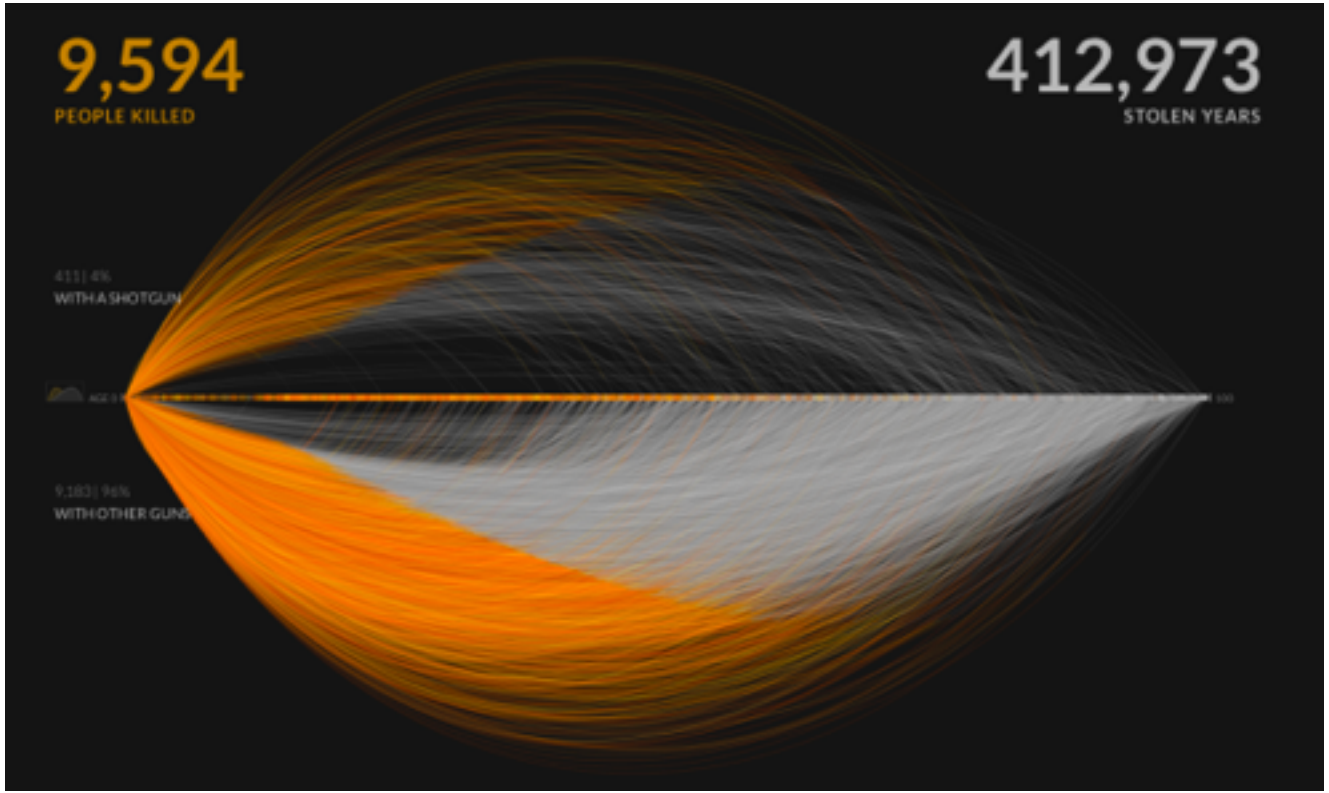

d
*
30
60


a
13 53 81 29 25
22 68 62 24 78
b
0–30
31–60
61–100
c
29
25
24
22
13
53
81
78
68
62
30
60
d
*
30
60

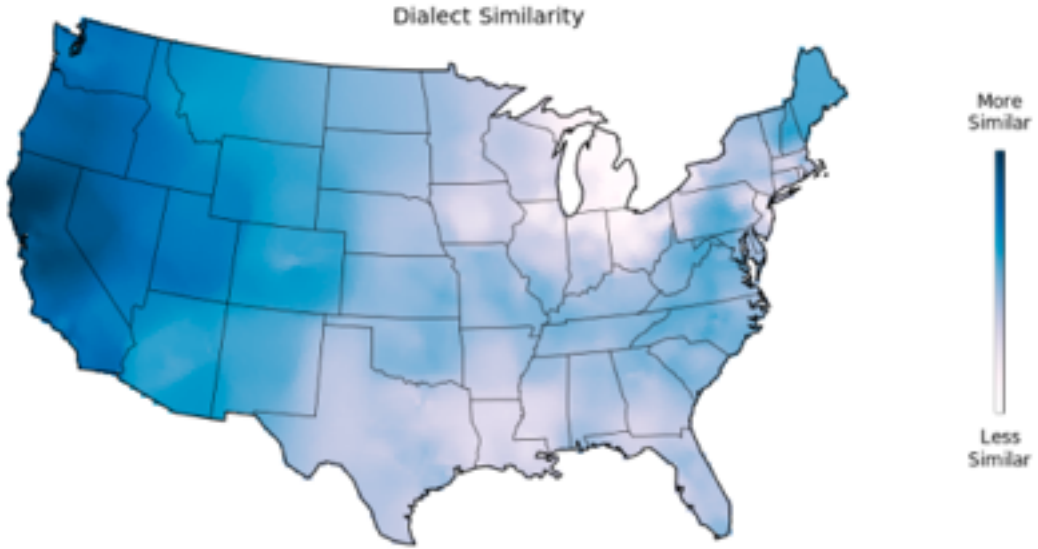
Always make some plots of your data before you apply any algorithms!



Source: Anscombe's quartet *en.wikipedia.org/wiki/Anscombe%27s_quartet*

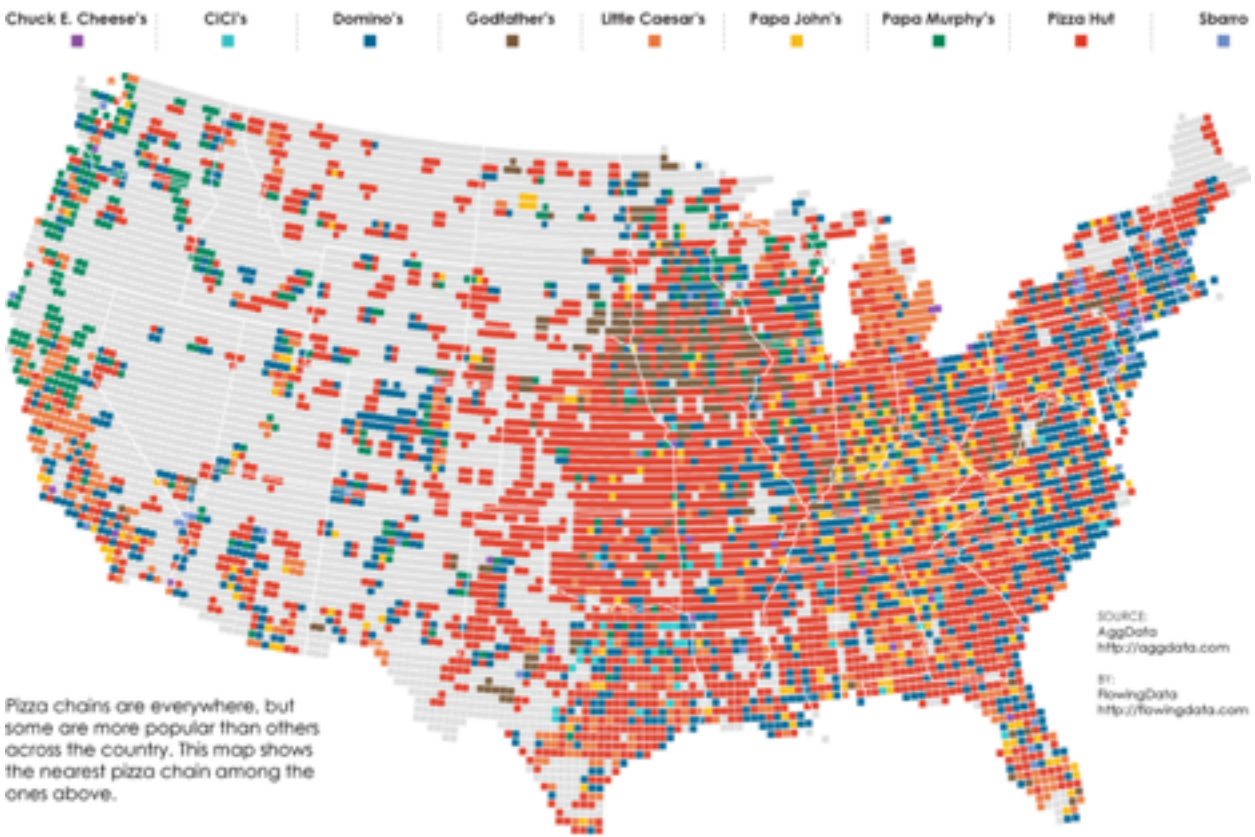

9,594
PEOPLE KILLED
412,973
STOLEN YEARS
411 | 4%
WITH A SHOTGUN
9,183 | 96%
WITH OTHER GUNS


Trott George
Peck Samuel
Revere Paul



	Most Similar Cities	
1	Santa Rosa CA	59.5
2	Vallejo CA	59.2
3	Fremont CA	58.9
4	San Francisco CA	58.8
5	Modesto CA	58.0

	Least Similar Cities	
1	Detroit MI	33.7
2	Warren MI	34.1
3	Toledo OH	35.0
4	Flint MI	35.1
5	Lansing MI	35.1


Chuck E. Cheese's
CiCi's
Domino's
Godfather's
Little Caesar's
Papa John's
Papa Murphy's
Pizza Hut
Sbarro
SOURCE:
AggData
http://aggdata.com
BY:
FlowingData
http://flowingdata.com
Pizza chains are everywhere, but some are more popular than others across the country. This map shows the nearest pizza chain among the ones above.

- Go to github.com/ga-students/DAT-23-NYC
- Scroll down to lesson #4
- Open the *Visualizations* notebook