### 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)

### **ASSIGNMENT #1: DATA EXPLORATION**

- 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)

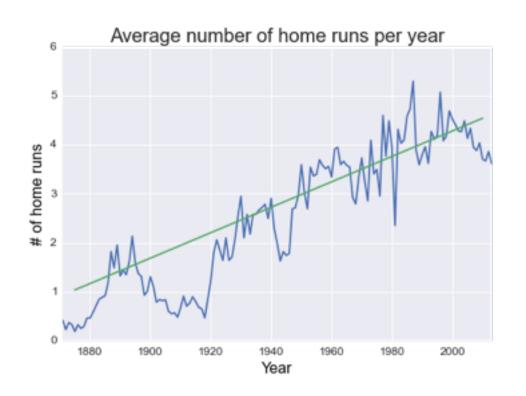
### **ASSIGNMENT #1: DATA EXPLORATION**

- 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

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

#### **EXERCISES**

- 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

DATA VIZ

a

13 53 81 29 25

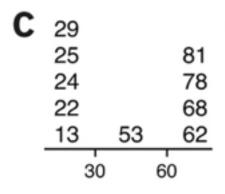
22 68 62 24 78

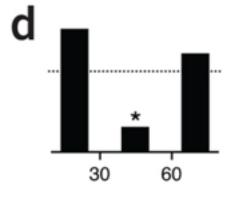
b

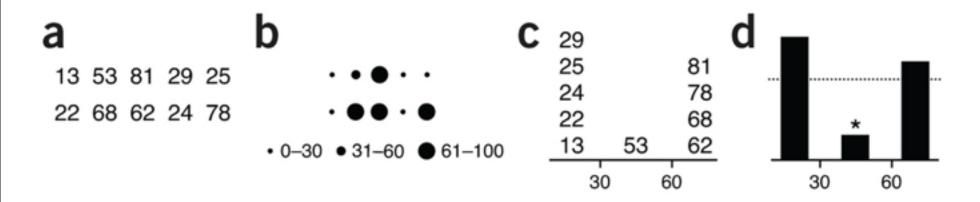
• • • •

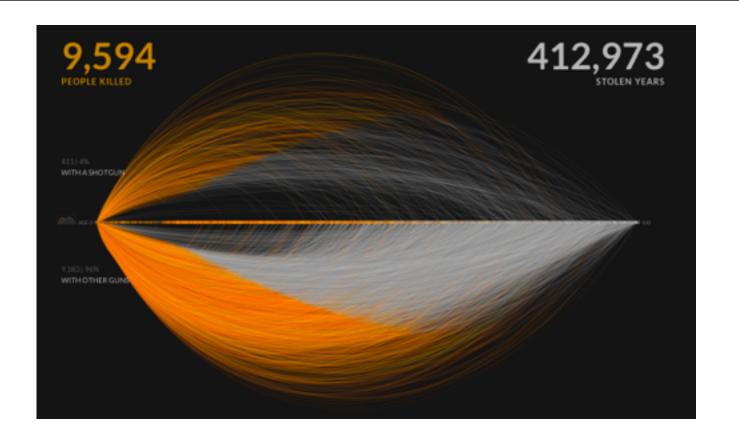
· • • · •

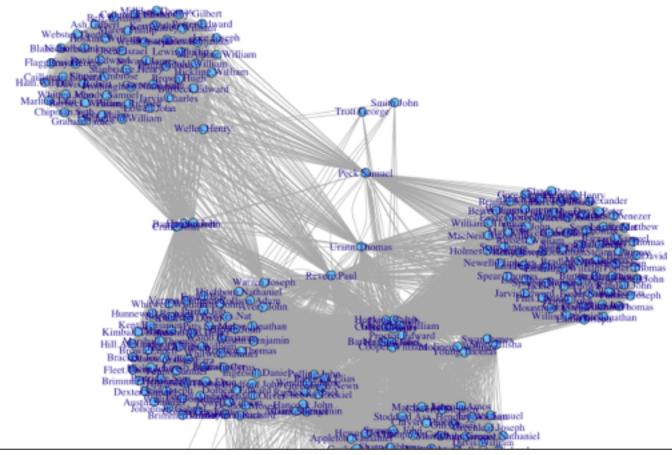
• 0-30 • 31-60 ● 61-100

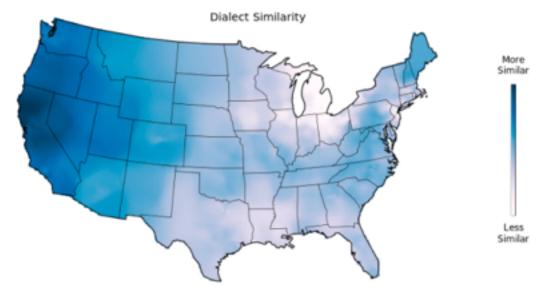








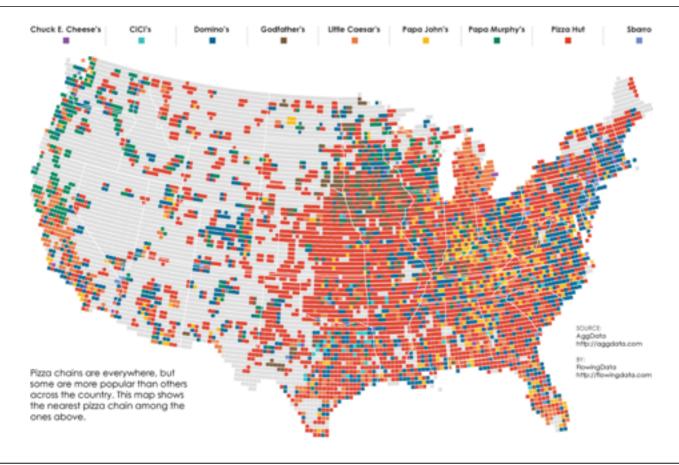




Joshua Katz, Dept of Statistics, NC State University http://spark.rstudio.com/(katz/DialectQuitz/

	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



EXERCISES 11

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