VISUALIZING NEW YORK CITY IN DATA

ADLEY KIM

ABOUT ME

جامعة نيويورك أبوظبي
NYU | ABU DHABI

SOCIAL RESEARCH &
PUBLIC POLICY

NYU | Center for Urban Science + Progress

APPLIED URBAN SCIENCE &
INFORMATICS

DATA ENGINEER

AGENDA

I. Set up Jupyter Notebooks

II. Lecture: Data Science Workflow

III. Break (5-10 minutes)

IV. Lab: Visualizing Commercial Noise Complaints

JUPYTER NOTEBOOKS

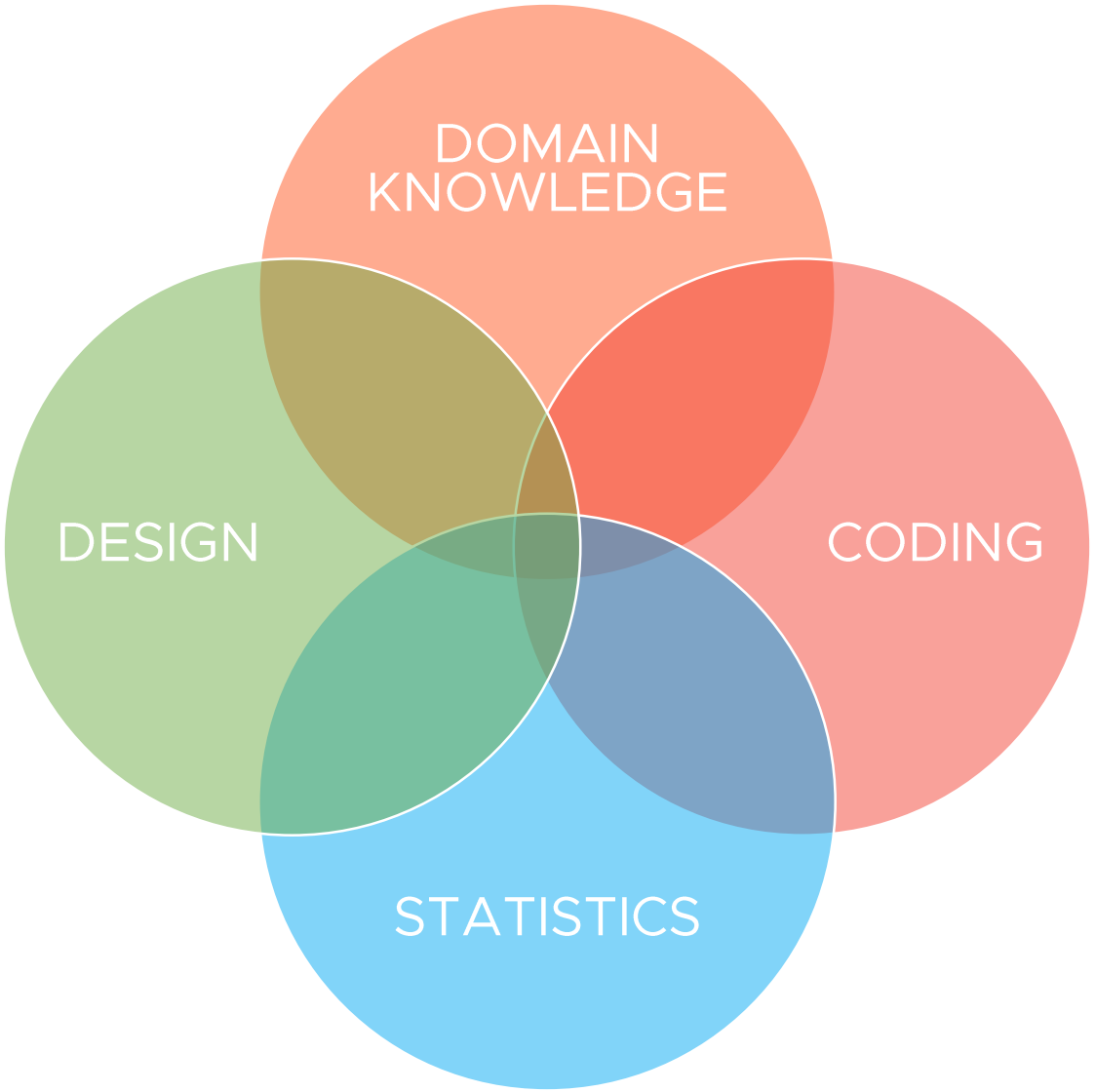
I. Go to: https://www.anaconda.com/distribution/

II. Install the Python 3.7 version of Anaconda

III. Install geopandas (`conda install geopandas`)

- » Windows: Anaconda Prompt
- » Mac: Terminal

WHAT IS DATA SCIENCE?


DOMAIN
KNOWLEDGE
DESIGN
CODING
STATISTICS

YOU DON'T NEED TO
BE A DATA SCIENTIST
TO BE DATA LITERATE.

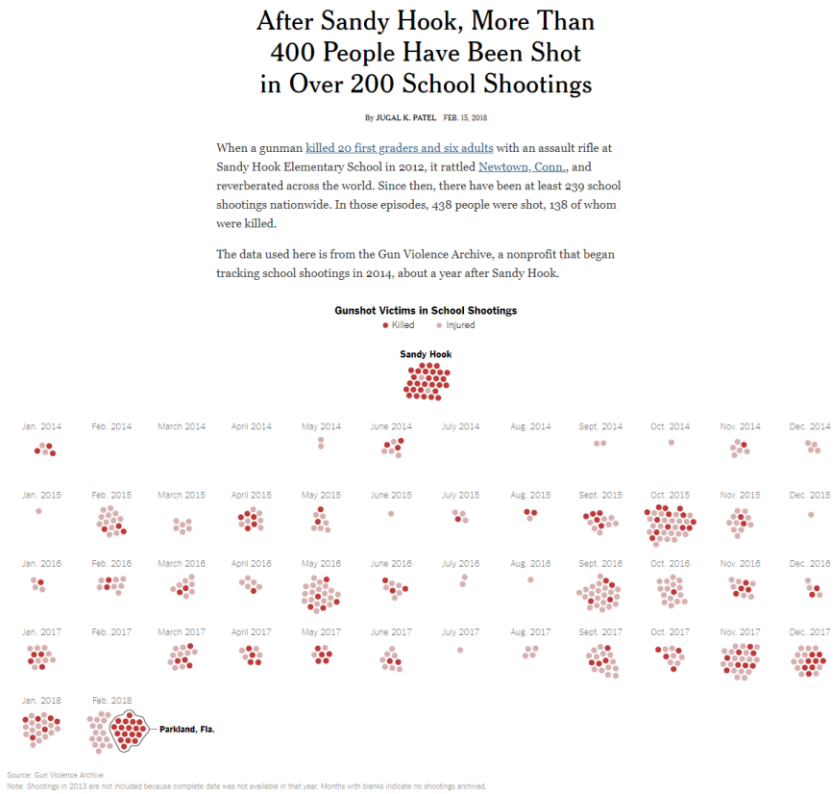
The New York Times

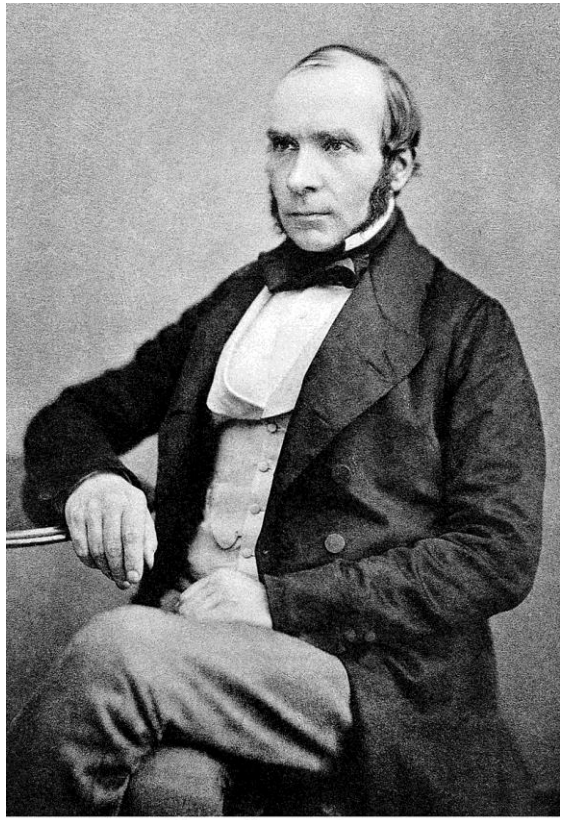
DATA JOURNALISM

“Fluency with numbers and data has become more important than ever. While journalists once were fond of joking that they got into the field because of an aversion to math, numbers now comprise the foundation for beats as wide ranging as education, the stock market, the Census and criminal justice.

More data is released than ever before—there are nearly 250,000 datasets on data.gov alone—and increasingly, government, politicians and companies try to twist those numbers to back their own agendas.”

Lindsay Rogers Cook
How We Helped Our Reporters Learn to Love Spreadsheets

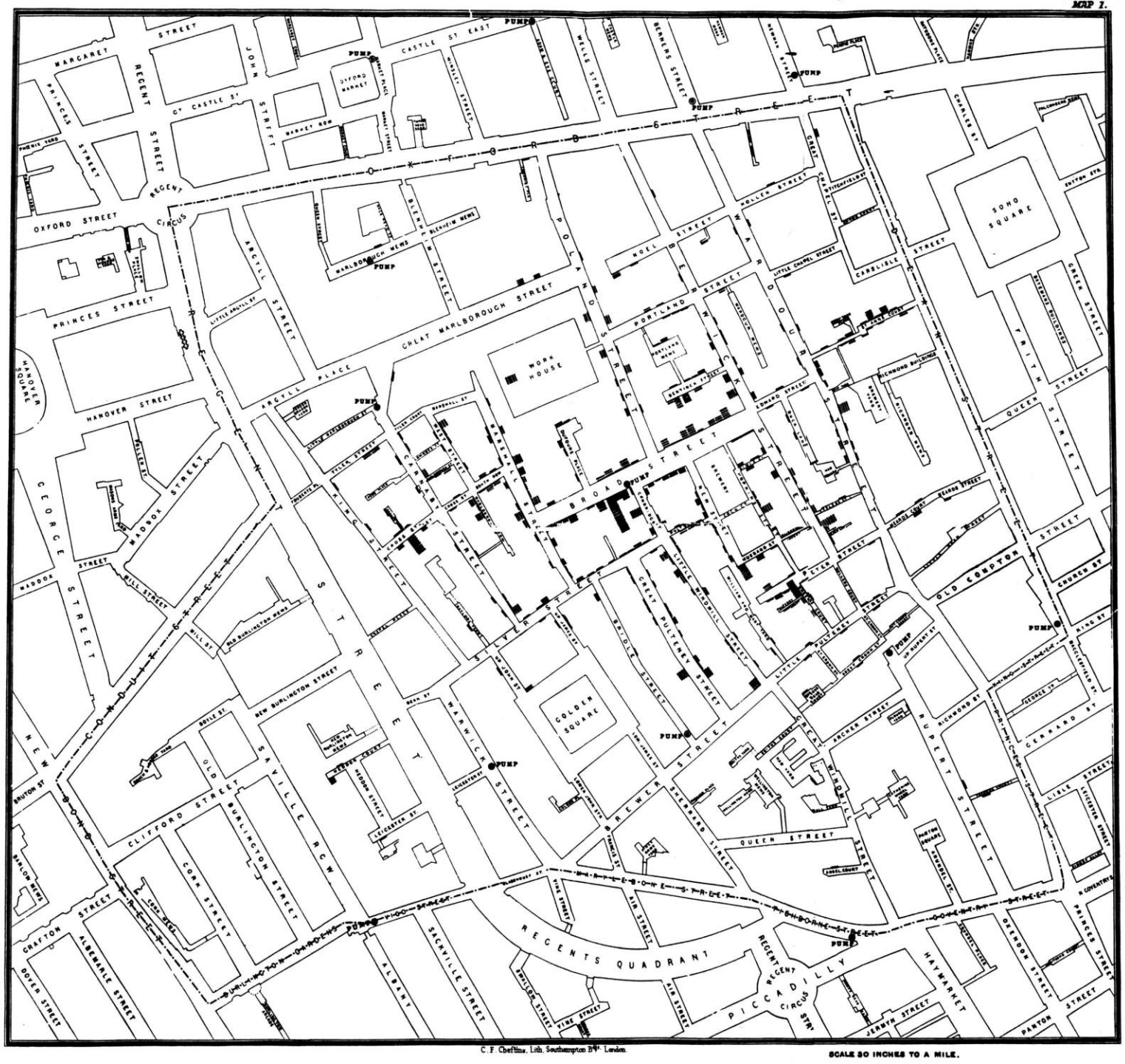




John Snow

JOHN SNOW

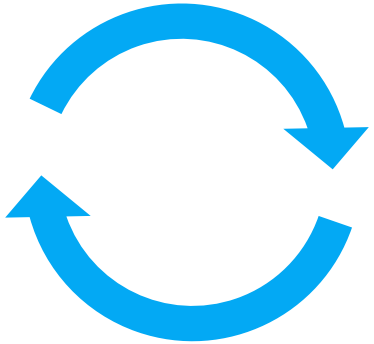
LONDON CHOLERA OUTBREAK (1854)



DATA SCIENCE WORKFLOW

THE QUESTION

WHAT DO I WANT TO KNOW?



WHAT DATA DO I HAVE ACCESS TO?

DATA COLLECTION

OPEN DATA

API

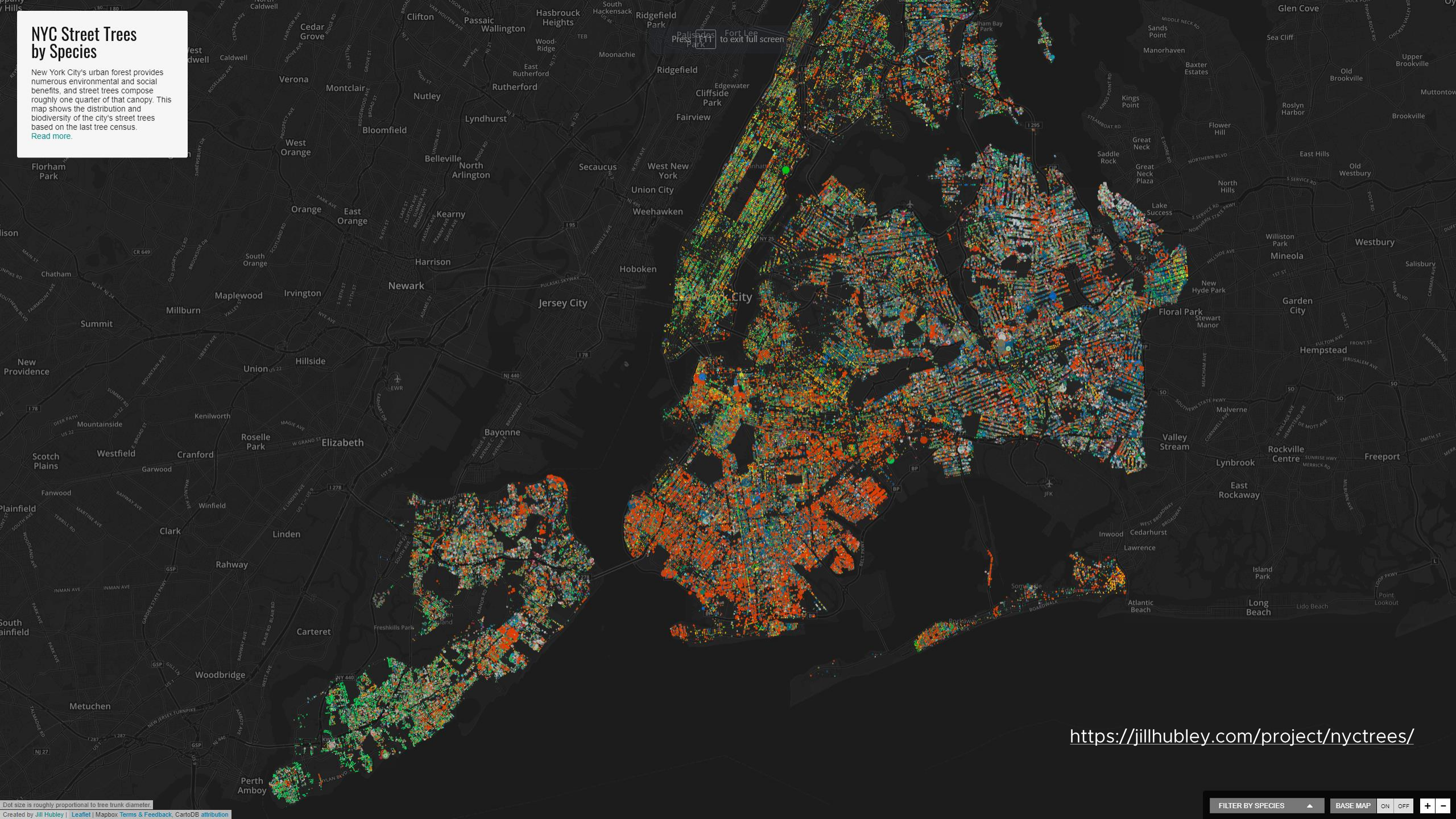
WEB SCRAPING

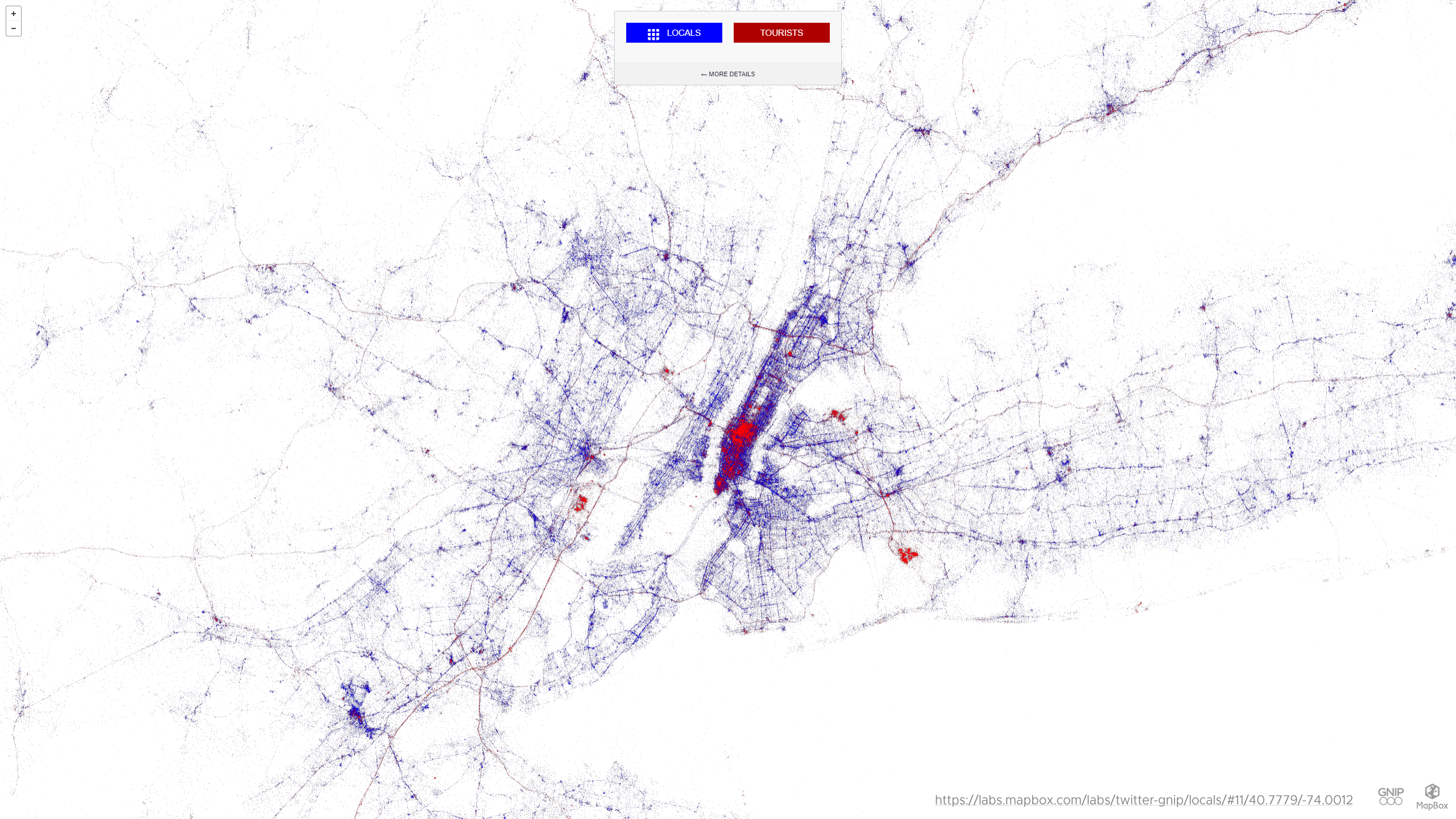
United States™ Census Bureau

NYC OpenData

yelp

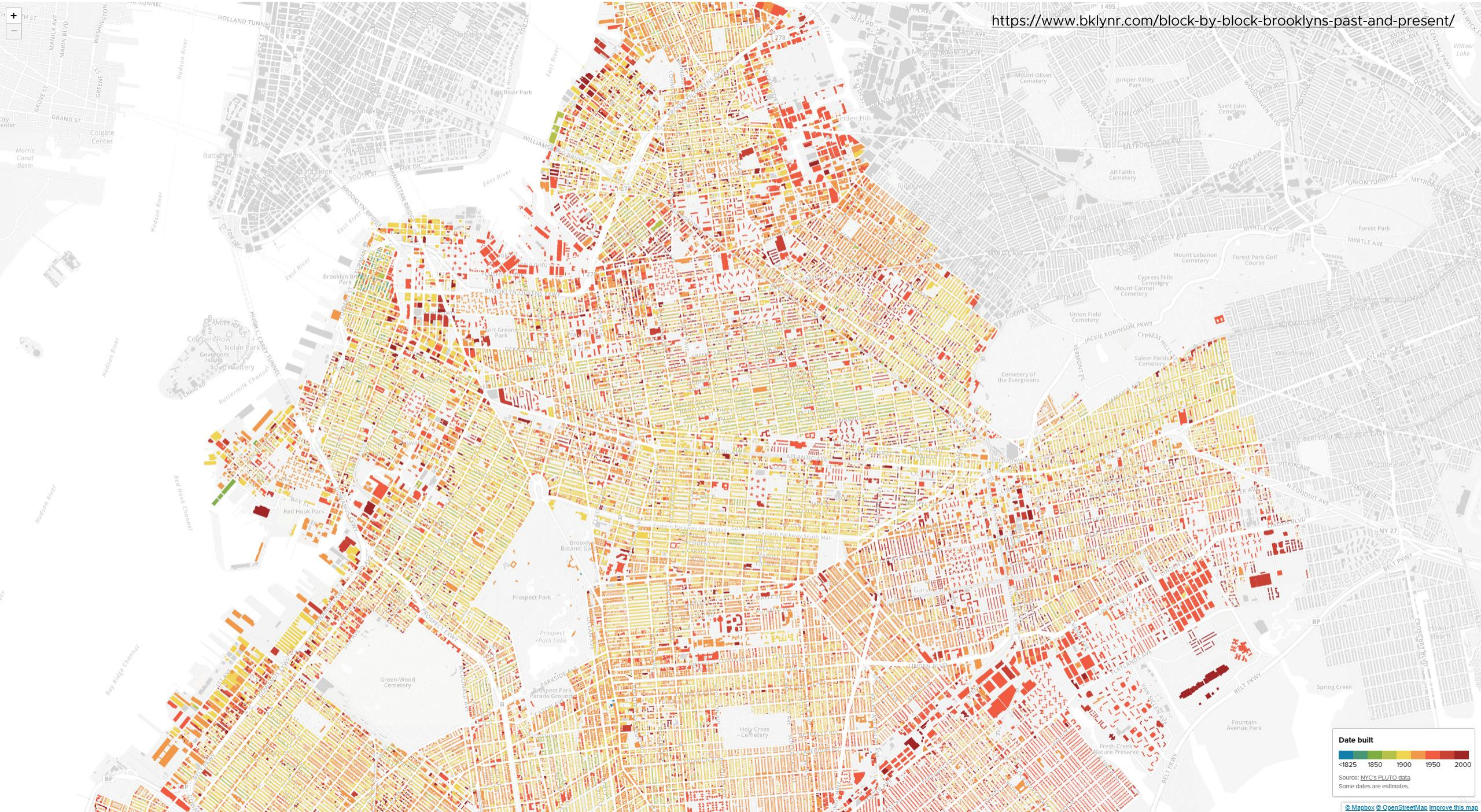
BEAUTIFULSOUP


NYC Street Trees by Species
New York City's urban forest provides numerous environmental and social benefits, and street trees compose roughly one quarter of that canopy. This map shows the distribution and biodiversity of the city's street trees based on the last tree census.
Read more.
Press F11 to exit full screen
https://jillhubley.com/project/nyctrees/
Dot size is roughly proportional to tree trunk diameter.
Created by Jill Hubley | Leaflet | Mapbox Terms & Feedback, CartoDB attribution
FILTER BY SPECIES
BASE MAP
ON
OFF


+
−
LOCALS
TOURISTS
← MORE DETAILS
https://labs.mapbox.com/labs/twitter-gnip/locals/#11/40.7779/-74.0012
GNIP
MapBox

Block by Block, Brooklyn's Past and Present

https://www.bklynr.com/block-by-block-brooklyns-past-and-present/



DATA CLEANING IS 80%* OF THE JOB

*76.3% of all statistics are made up.

TIDY DATA

Y	SPATIAL UNIT	TEMPORAL UNIT	X_1	X_2	X_n

1 row = 1 observation for 1 space at 1 time

VARIABLE TYPES

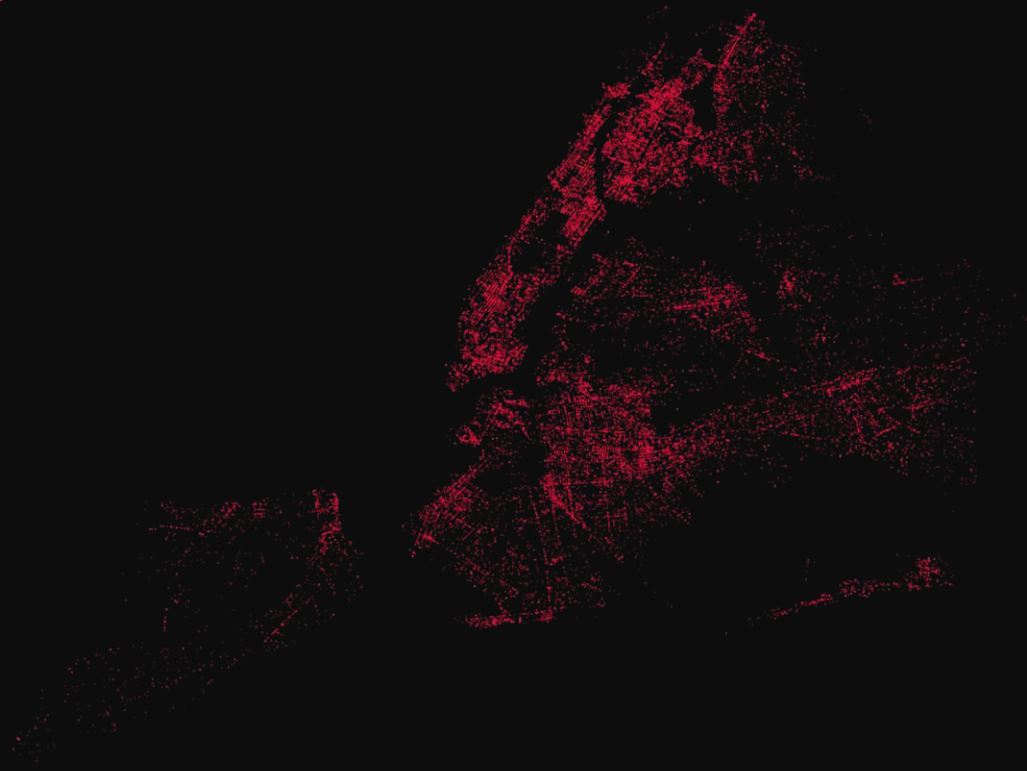
- → Continuous (numbers, percentages)
- → Discrete (classes, categories)
- → Ordinal (rankings)

LEVELS OF SPATIAL AGGREGATION

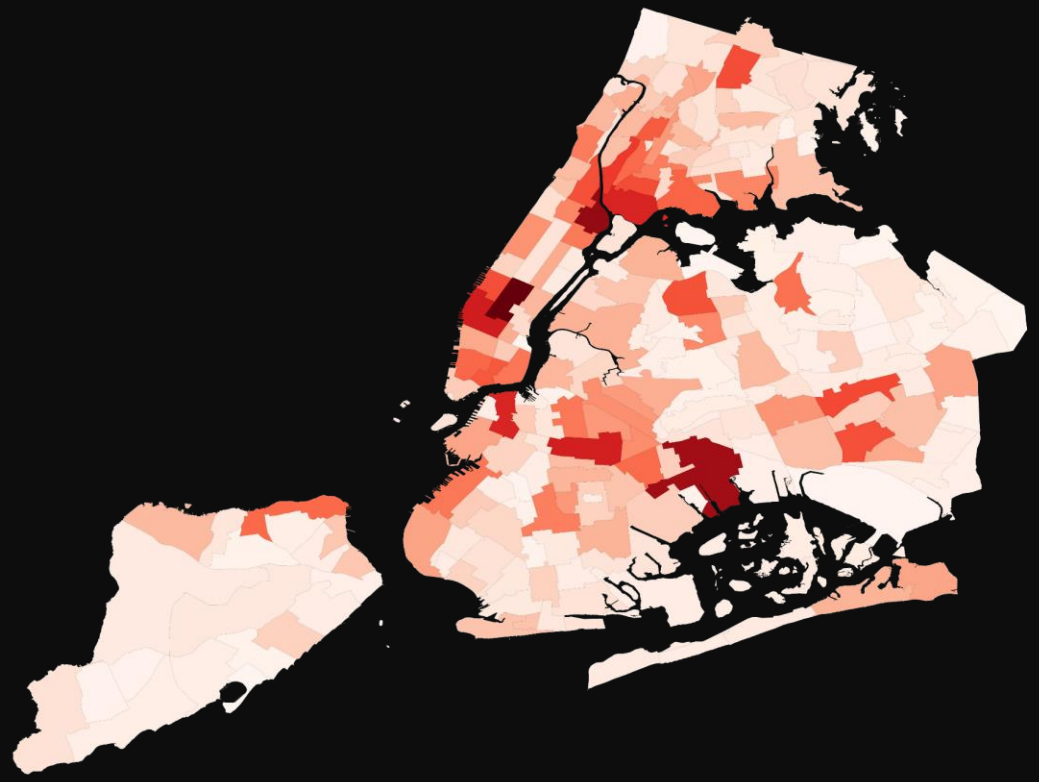
NYPD ARRESTS 2019

https://data.cityofnewyork.us/Public-Safety/NYPD-Arrest-Data-Year-to-Date-/uip8-fykc

All arrests made by the NYPD in the year 2019.



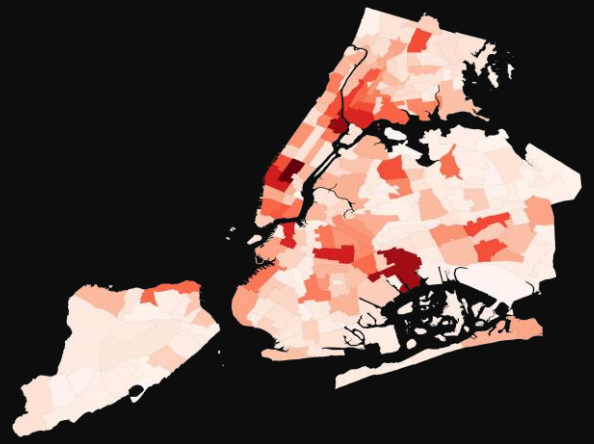
BOROUGH



NEIGHBORHOOD

CENSUS TRACT

BOROUGH



NEIGHBORHOOD

CENSUS TRACT

WHAT’S MISSING?

1. STANDARDIZE UNITS

2. ACCOUNT FOR TIME

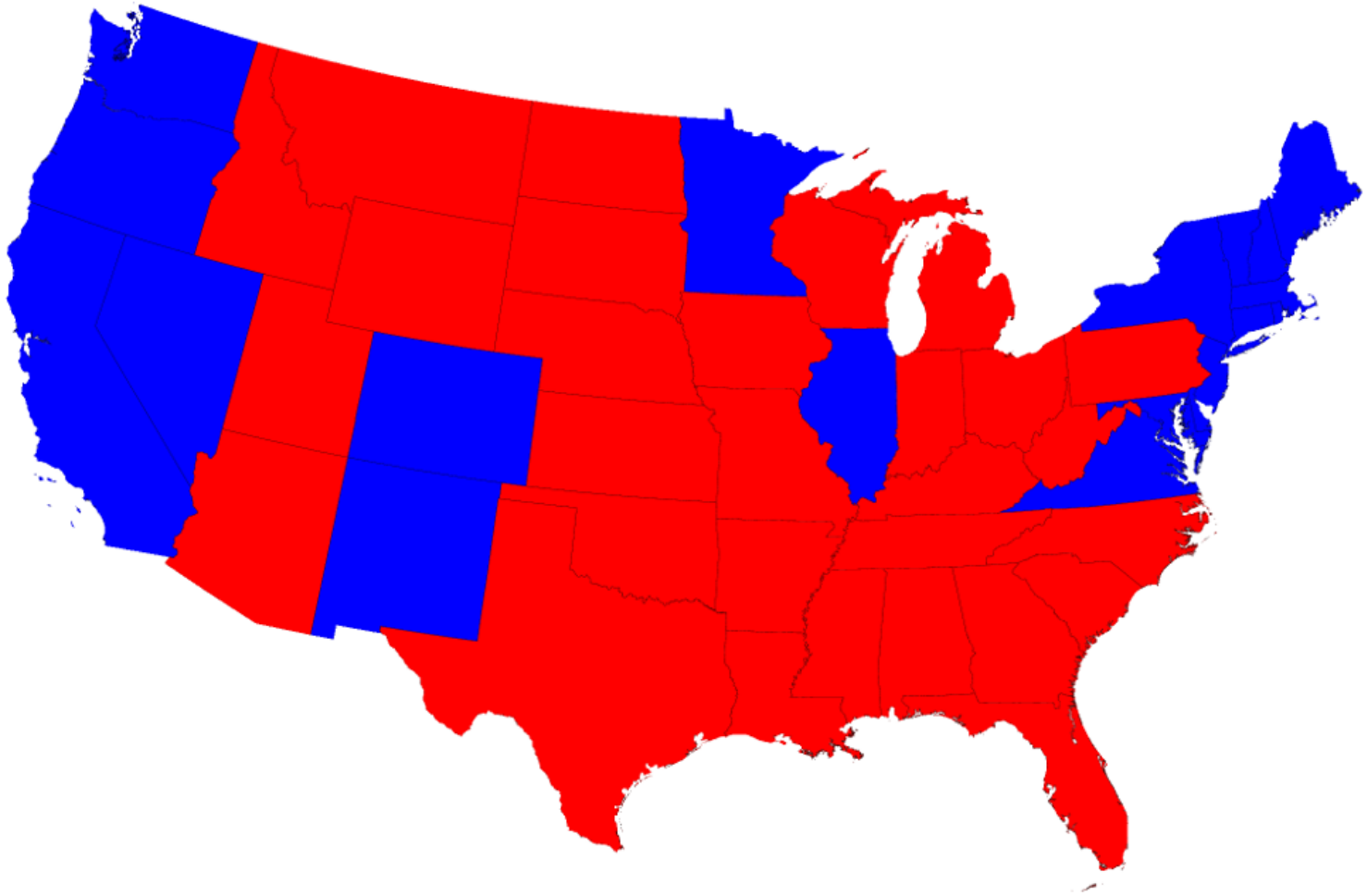
“YOUR DEFAULT POSITION SHOULD BE SKEPTICISM”

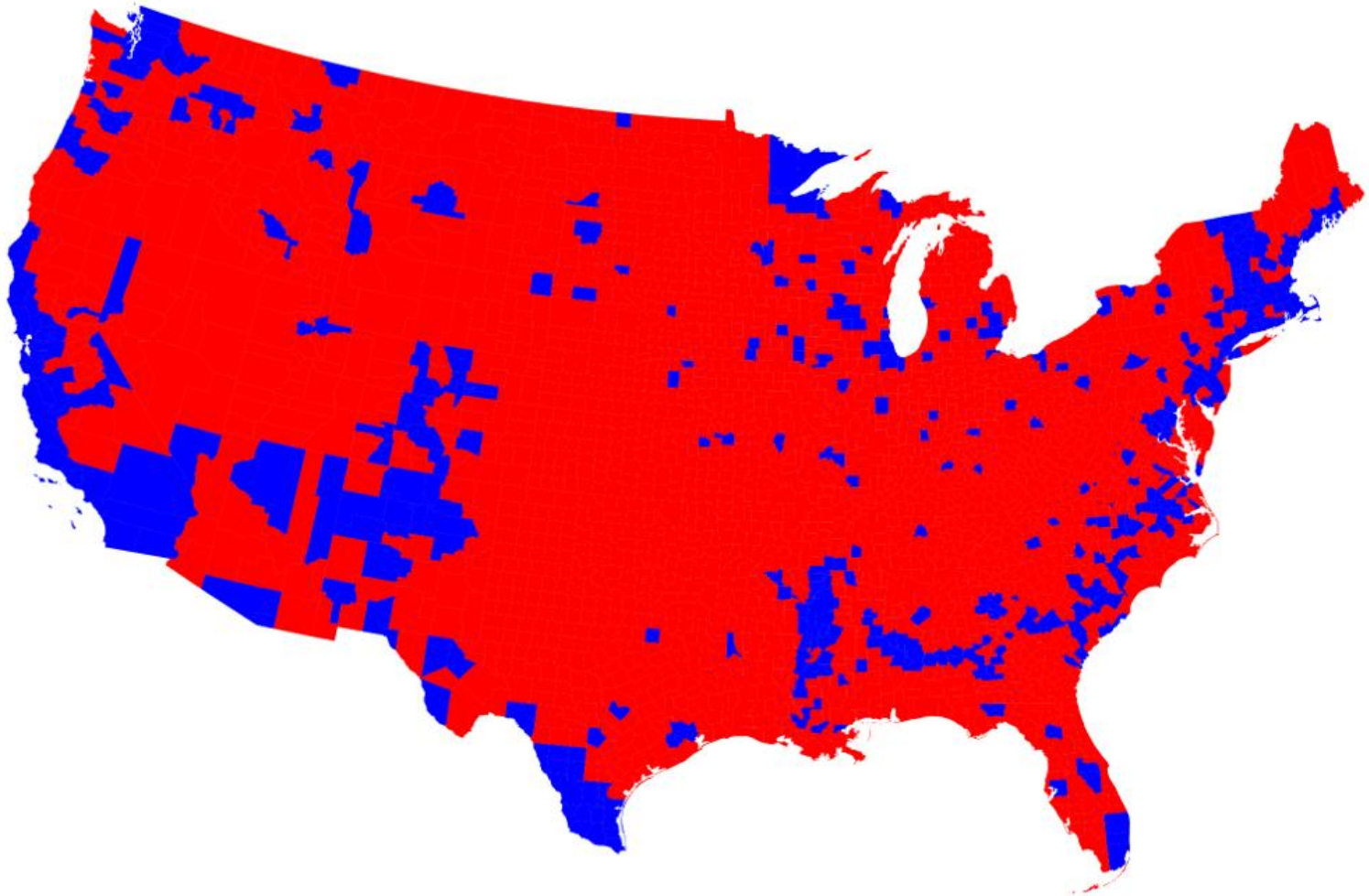
– HADLEY WICKHAM

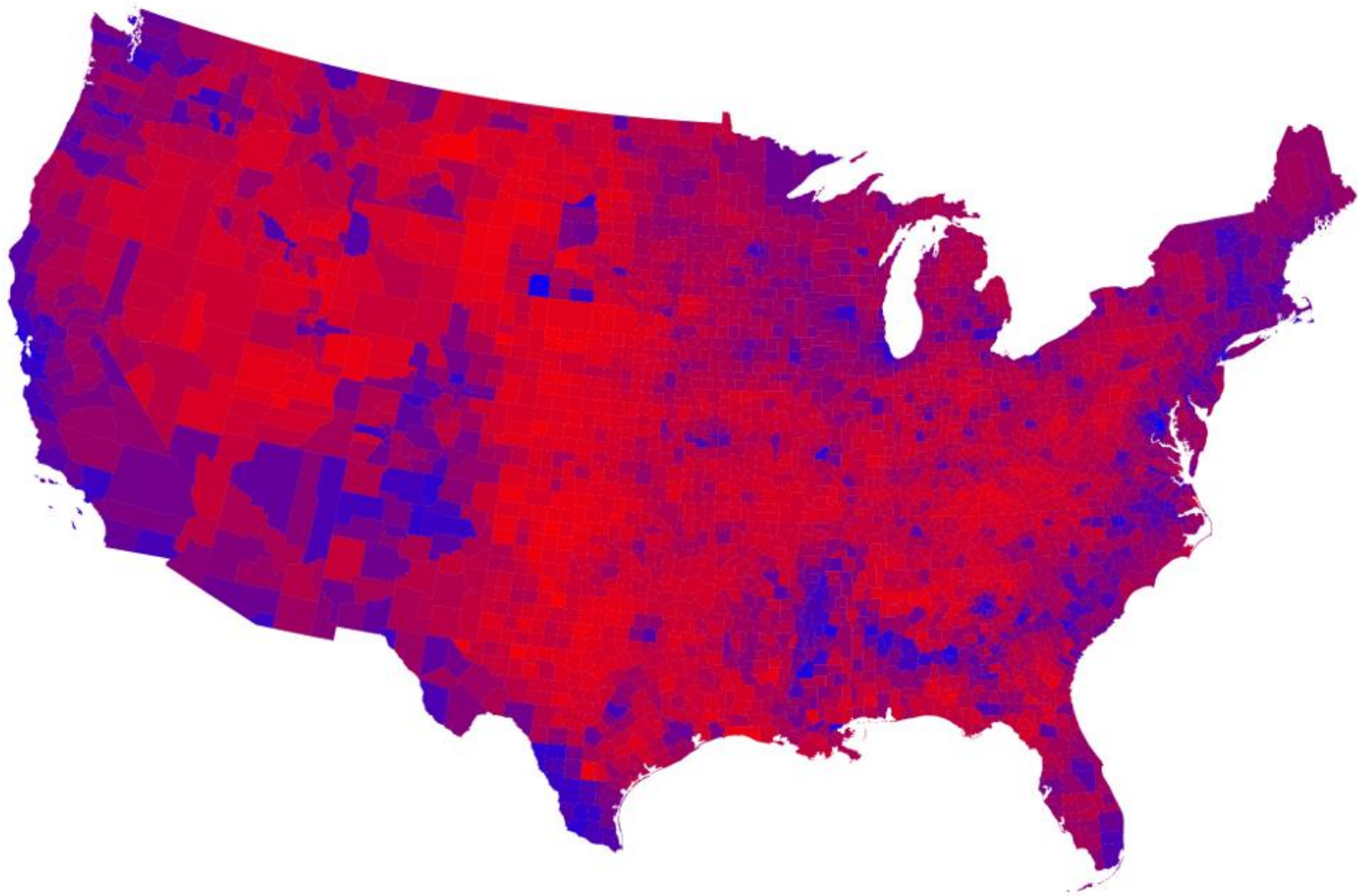
https://www.propublica.org/nerds/hadley-wickham-your-default-position-should-be-skepticism-and-other-advice-for-data-journalists

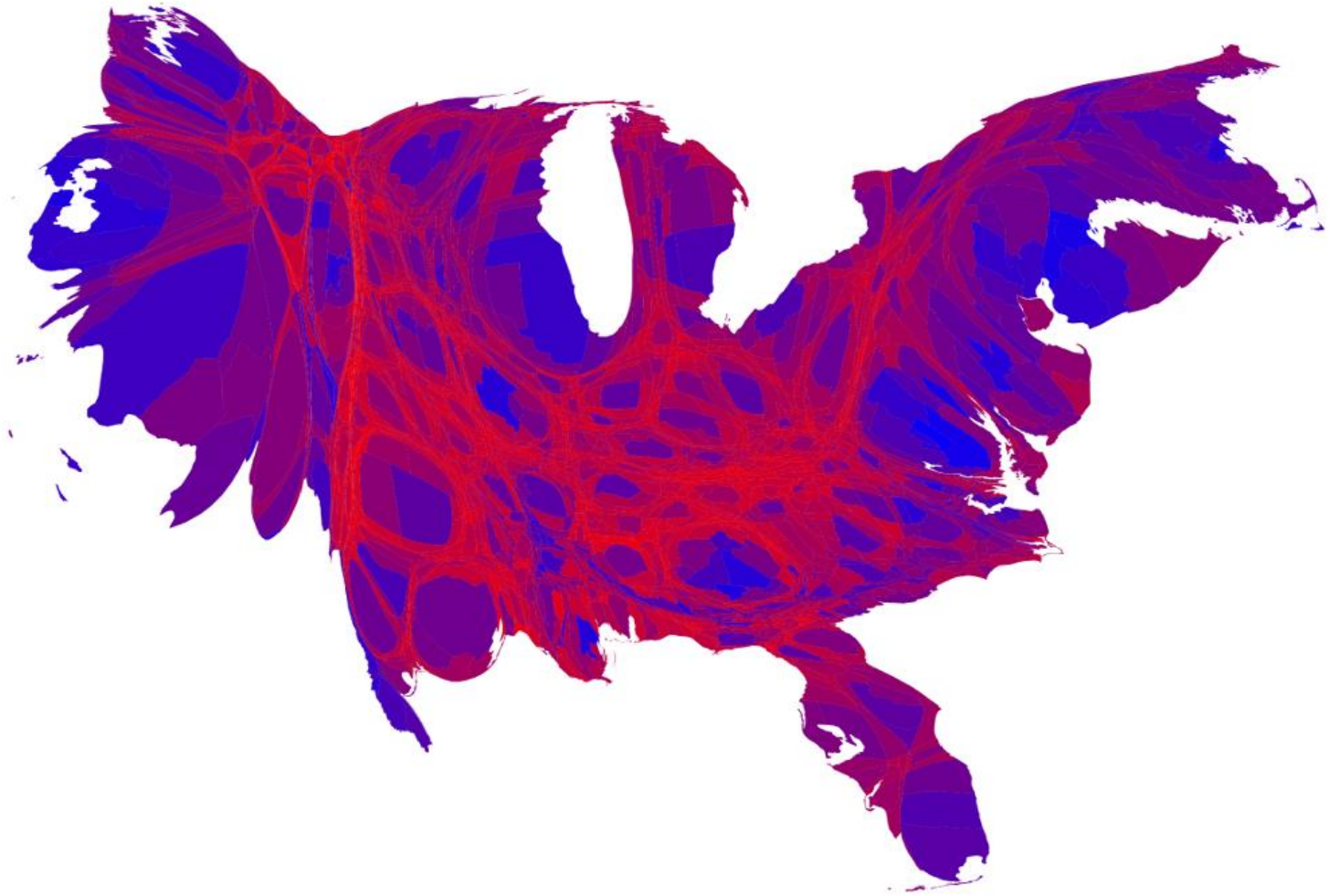
2016 PRESIDENTIAL ELECTION OUTCOMES

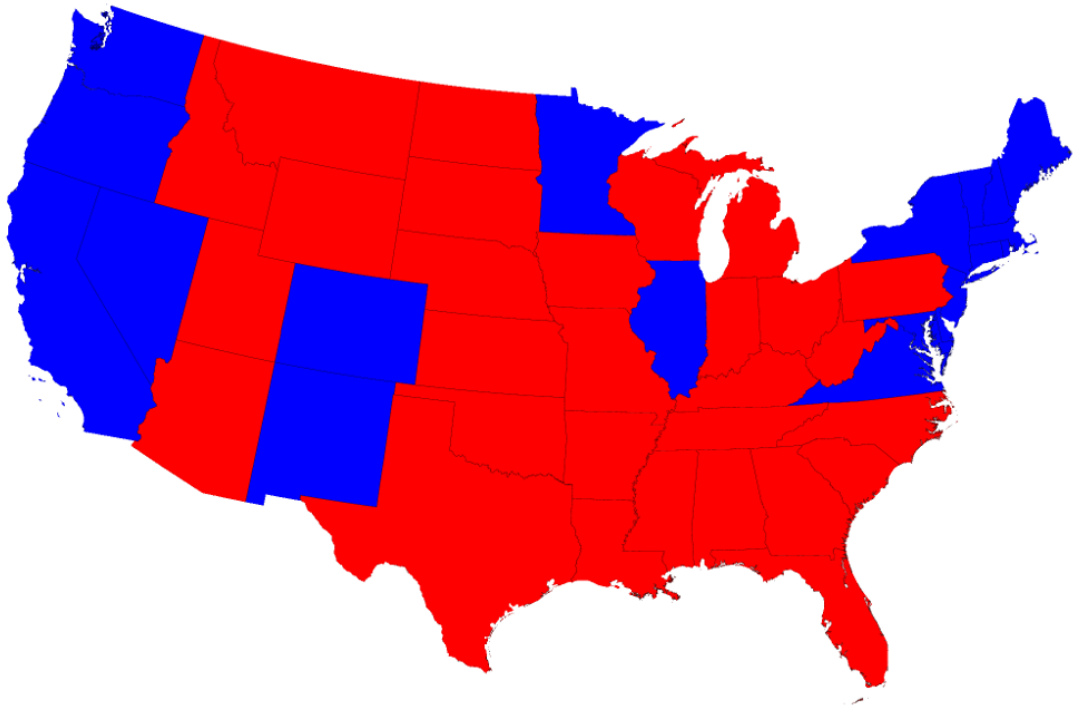
Mark Newman, Department of Physics and Center for
the Study of Complex Systems, University of Michigan

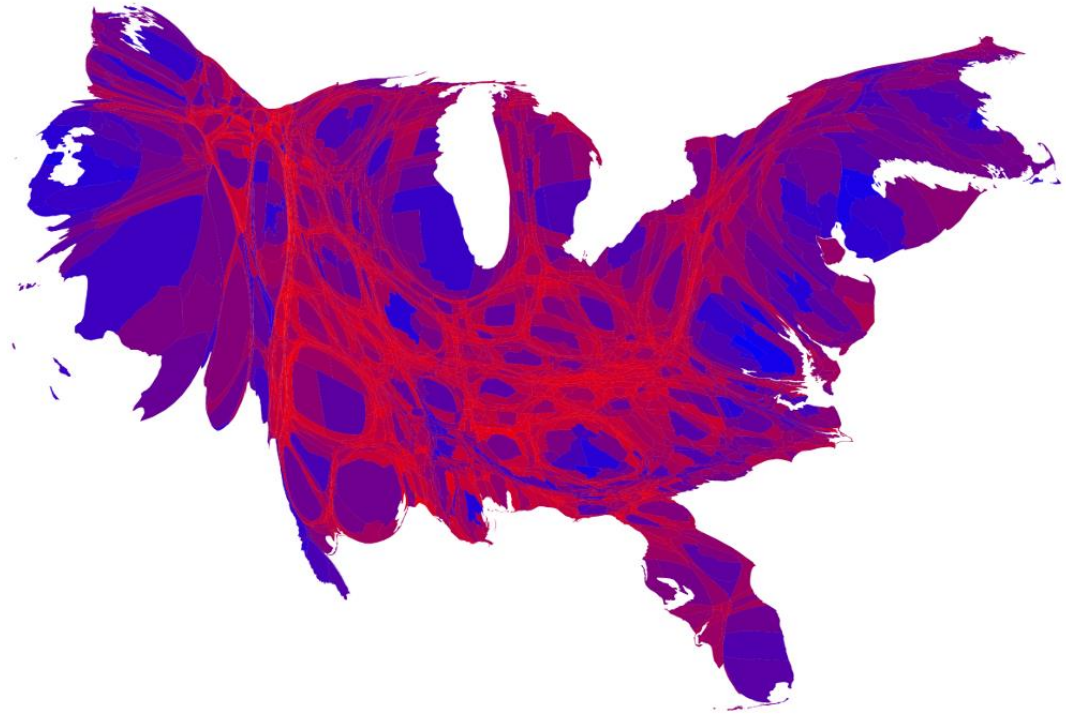












IT’S REALLY EASY TO LIE WITH DATA

RECAP: DATA SCIENCE WORKFLOW

QUESTIONS?