Mapping 311

Dr Jon Roberts

311 data is very rich

- > 5 million rows
- Most have zip codes
- All divided into categories using the descriptor column
- We can map it!

Parse the data

- Run through all the rows of the data
- Skip lines with no zip code (many of these are taxi complaints)
- Keep a running list of the different descriptions
  - https://github.com/jonroberts/nycenters

Count the CSV

Create a json structure with one entry for each zip, and a count of the 311s:

{“NY10012”:{

“Trapping Pigeon”: 1,

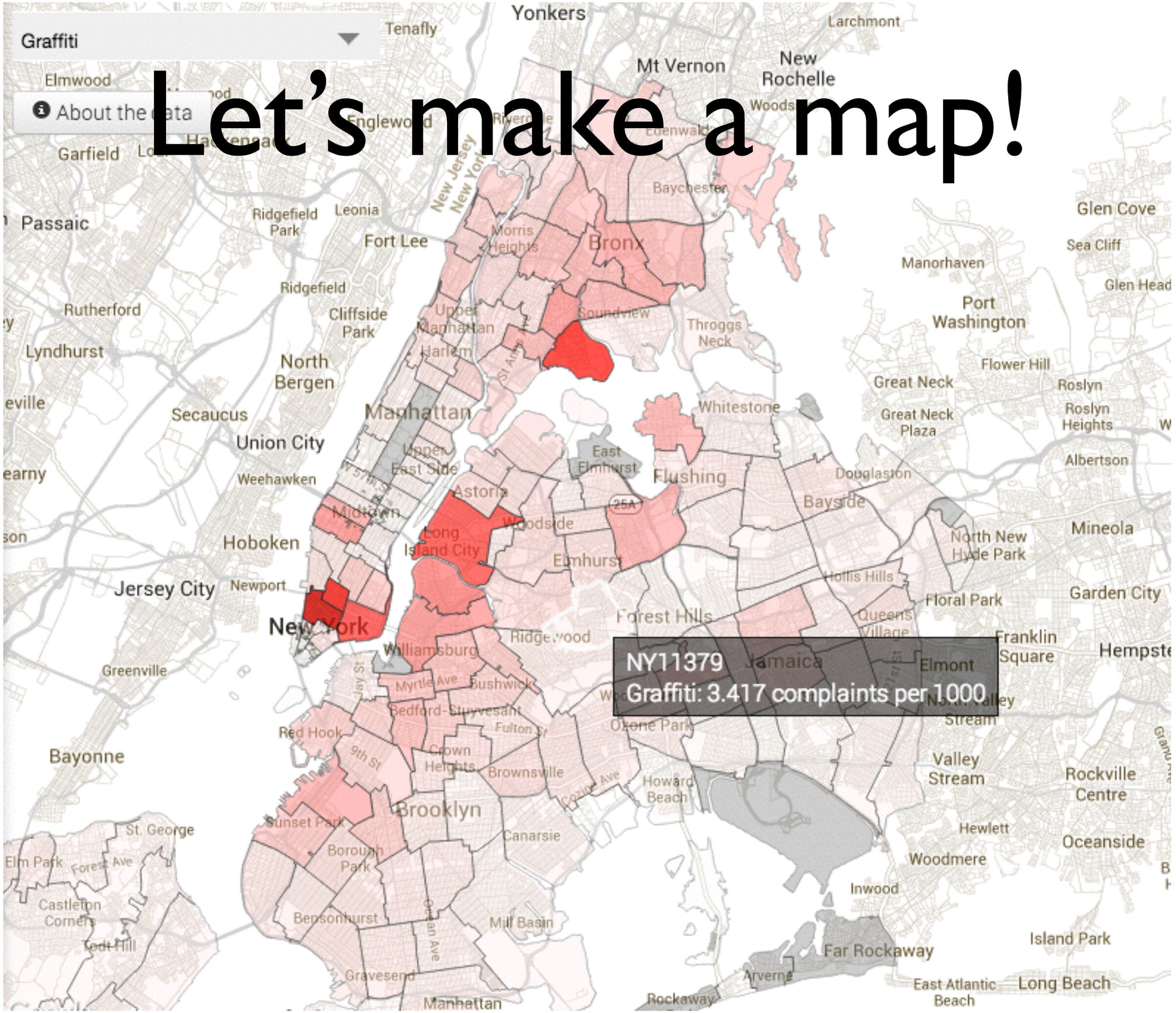
“Poison Ivy”: 0, ...

}, ... {}}

Normalise by the population of each zip

- Here I cheat and use data that I pulled for EnergyZip
- Population data comes from the American Community Survey (Census)
- Then we can get a # of complaints per 1000 population!

Let’s make a map!



http://jrsandbox.com/nycenter

Some caveats

- Those missing zip codes shouldn't all be thrown away
- Some disambiguation is needed on 311 description fields (e.g. 'noise' and 'noise commercial' might need to be combined)
    - What was the one trapping pigeon complaint?

Future directions

- There's a *lot* more data in this set to show
- There's a lot more demographic info per zipcode we could use for analysis beyond just normalising per population
- What about squeeky wheels?
    - This set isn't normalised for many complaints by one person.

Take the code!

- This is a start
- Please take it and build more awesome stuff!
- Suggest more map layers.
- Anything with zipcode info is *easy* to add.