Project 2 Proposal

The George Washington University Data Analytics and Data Visualization Boot Camp

02 January 2019

Patricia Escalante, Melinda Wiley, Yanjun Zhou

Project 2 Reported Parasitic Infections in the United States

Drafted Proposal

Project 2 will utilize Python Flask powered RESTful API, HTML/CSS, JS, one (1) database (MySQL, MongoDB, SQLite) with a minimum of 100 records with at least three (3) user-driver interactions. Data munging and data visualization utilizing CDC provided datasets and parasite information from https://www.cdc.gov/parasites/az/index.html#s and <https://data.cdc.gov>.

Web scrape https://www.cdc.gov/parasites/az/index.html#s to pull the list of parasites. Working this aspect through jupyter notebook with BeautifulSoup.

Utilize JSON APIs provided by CDC from https://data.cdc.gov focusing NNDSS – Table II using the list from the Alphabetical Index of Parasitic Diseases - 91 listed.

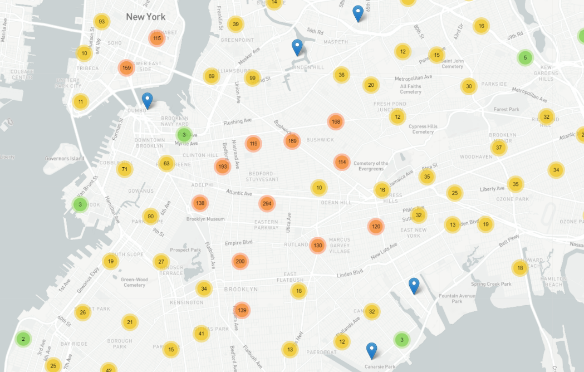
Initial layout idea will be US Mapping to define infections through the years utilizing Leaflet or Plotly. May chart infection cases over the years in a bar or line chart. Due to the number of parasites listed and the amount of data provided by the CDC, the dataset may be cut down from 91 over 3 to 5 years to a smaller dataset. This is due to the NNDSS Table II broken out into several APIs. These JSONs will be read into a database, probably MySQL or SQLite.

The 1 JS library not previously covered will be TBD.

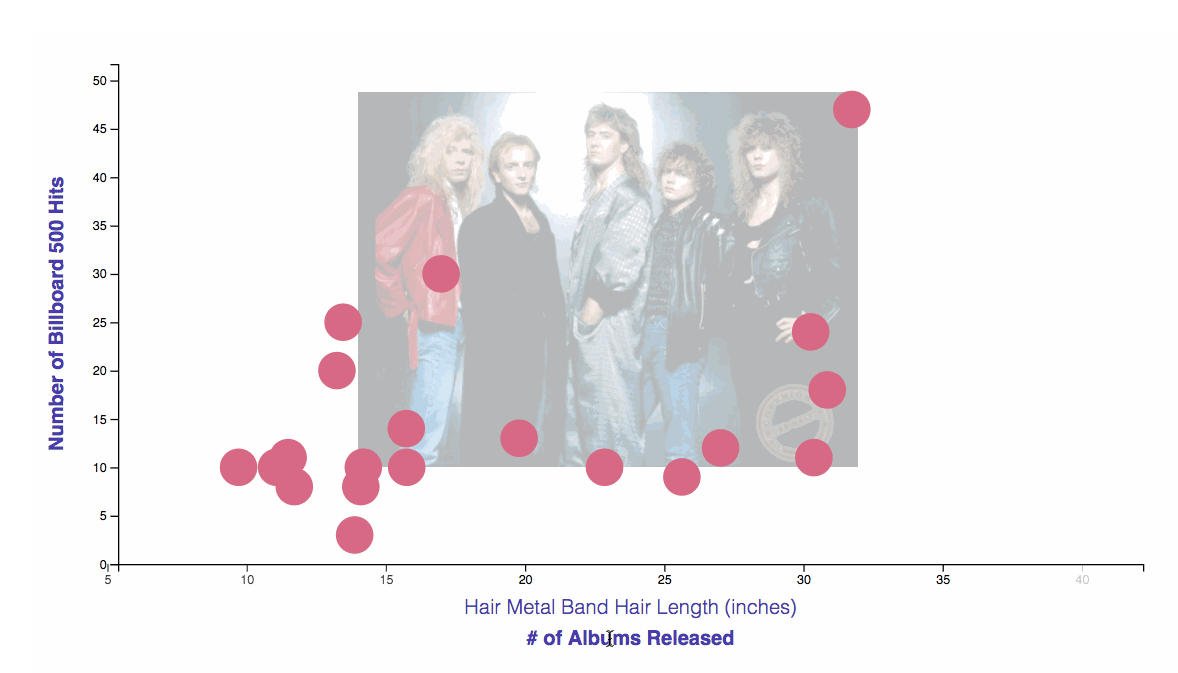
Resources, so far: <https://www.cdc.gov/parasites/az/index.html#s>, <https://data.cdc.gov>

Ideal Visuals:

Different markers with layers



D3 charting with transitions and interactions



<https://github.com/Yanjun7/Parasites>

https://github.com/minwiley/Parasites.git

Requirements

1.Your visualization must include a Python Flask powered RESTful API, HTML/CSS, JavaScript, and at least one database (MySQL, MongoDB, SQLite, etc.)

2.Your project should fall into one of the below four tracks:

• A custom "creative" D3.js project (i.e. non-standard graph or chart)

• A combination of Web Scraping and Leaflet or Plotly

• A dashboard page with multiple charts all updating from the same data

• A "thick" server that performs multiple manipulations on data in a database prior to visualization (must be approved)

3.Your project should include at least one JS library that we did not cover.

4.Your project must be powered by a dataset with at least 100 records.

5.Your project must include some level of user-driven interaction (e.g. menus, dropdowns, textboxes, etc.)

6.Your final visualization should ideally include at least three views

Schedule

1.Wed:

Between now and Saturday, you will need to start brainstorming topics with your group and researching potential datasets. Your focus should specifically center around:

• Selecting a Topic

• Finding a Dataset

• Finding Inspiration

• "Sketching" your ideal visuals

• Creating a 1-Page Proposal

2.End-of-Class Saturday:

You will need to create a 1 page proposal. The proposal should include:

• A brief articulation of your chosen topic and rationale

• A link to your dataset(s) and a screenshot of the metadata if it exists.

• 3-4 screenshots of relevant "inspiring" visualizations that frame your creative fodder (or just links)

• A sketch of the final design

• A link to the primary GitHub repository you'll be housing your work in

3.Mon, Wed; Project Work

4.Saturday Jan 12: Presentation

Datasets:

https://www.data.gov/

https://opendata.cityofnewyork.us/

https://www.capitalbikeshare.com/system-data

https://www.census.gov/data.html

https://dc.gov/page/open-data

https://www.eia.gov/

https://www.bjs.gov/

https://tinyletter.com/data-is-plural/archive

https://www.kaggle.com/datasets

https://www.nass.usda.gov/Publications/Ag\_Statistics/

https://www.bls.gov/

Places for viz inspiration:

https://flowingdata.com/

https://github.com/d3/d3/wiki/Gallery

https://www.mapbox.com/gallery/

https://bl.ocks.org/mbostock

https://beta.observablehq.com/collection/@observablehq/visualization

https://www.mapbox.com/mapbox-gl-js/example/simple-map/

Some commonly used javascript libraries (see more here, but not exhaustive https://en.wikipedia.org/wiki/List\_of\_JavaScript\_libraries):

The Big Boys (too steep of learning curve probably for all except perhaps Jquery):

• Jquery

• Underscore

• Node (for servers)

• AngularJS

• ReactJS

Visualization

• Mapbox.js (built on top of leaflet) https://www.mapbox.com/mapbox-gl-js/example/simple-map/

• Highcharts

• See others above

Jquery related

• Jquery UI

• Jquery Autocomplete https://github.com/devbridge/jQuery-Autocomplete

• Jquery tagit (https://github.com/aehlke/tag-it)

• Lots and lots of others

Ideas:

How has Cap Bike Share changed in ridership since the introduction of other bikes/scooters?

How often does your area have a “white Christmas” (looking at historical weather data)?

How does pollution affect public health in a particular area?

How president’s clemency and pardons have changed over the years.

Investigating food deserts

updated linked for review

https://www.cdc.gov/parasites/az/index.html#s

https://data.cdc.gov

https://www.newscientist.com/gallery/mg20327161300-enemy-within-human-parasites/

https://www.youtube.com/watch?v=-ttmPv-VNaY

https://reference.medscape.com/slideshow/intestinal-parasites-6010996

https://beta.observablehq.com/@mbostock/tadpoles

https://docs.google.com/document/d/1zJzYYBlsRm\_a\_iG-RKERvhnKN7YddV1GI1iPcSSBZQE/edit

https://leafletjs.com/examples/choropleth/

https://www.cbsnews.com/news/parasites-causing-infections-in-the-us-cdc-says/

https://wonder.cdc.gov/nndss/static/2018/51/2018-51-table1.html