DATA 22700 Spring 2023

Assignment 2: Color and cartography

Due April 21, 2023

In this assignment, students will use Altair in a computational notebook to develop a series of maps and provide detailed rationales for their design choices. The purpose of this assignment is to focus on design choices related to color scales and cartography, so special emphasis will be placed on these design choices.

Students will *work alone*.

Students should submit their assignment as a ipynb file on Gradescope. Be sure to evaluate all code blocks so the desired output you would like to submit is showing. Otherwise, your submission may appear to be incomplete when the course staff evaluates your work.

Illegal animals kept as pets data

In this assignment students will explore and answer questions about a particular dataset using Altair. Specifically, students will investigate the Illegal Animals represented in the NYC 311-calls dataset. 311 calls are a matter of public record and so are recorded and made available to the public.

This dataset can be found at NYC Open Data and filtering it (using their web interface to include Complaints of type "Illegal Animal Kept as Pet" and "Illegal Animal Sold"). If that doesn't work or for some reason then you can download an older version of the dataset here. It's a good practice to know how to download things from government websites, so the spirit of this assignment is download the above data yourself, but no penalty if you use the version of the data we prepared.

Note that students will need to supplement this data with geoJSON data for New York City (NYC) and other demographic data for NYC Boroughs, which can be found online. Finding these datasets will be good practice for the project!

Technical specification

Using Altair in an iPython notebook, answer the below questions about the dataset. Each answer should be expressed as a map or small multiples of maps rendered using Altair whose meaning is unambiguous:

1. What type of animal is most commonly sold in each NYC Borough?
2. What type of animal is most commonly kept as a pet in each NYC zip code?
3. What part of the NYC Borough of Queens is reported as having the highest number of animal reports?
4. What is the rate that each species is reported as appearing in each Borough relative

to the number of people who live in that Borough? (Hint this will require acquiring an additional dataset).

5. Within each zipcode, how does the way that complaints are filed correlate with the type of animal being reported?
6. Do illegal animal sales and illegal pet keeping tend to be co-located?
7. What is an amusing additional fact about this dataset?

Students will produce one map per question above. Whereever possible, **students should use choropleth maps** (i.e., regions on a map colored according to a scale) in order to demonstrate what you have learned about applying color scales to data. Students who forego using color in this assignment risk getting a mark of Needs improvement.

Each map should be preceded by a recap of the question it is answering and should be followed by a markdown text block containing a write-up about the design choices in the map. Write-ups should provide a rigorous rationale for students' design decisions, especially design choices pertaining to color scales, label placement, level of geospatial aggregation (e.g., zip code vs county vs state), and map projections. Document the visual encodings used and why they are appropriate for the data. How do these decisions facilitate effective communication? How to they answer the question that prompted you to create each map?