Title of the project

Pets in Seattle

Names of participants

Cynthia Wang: yw3393@columbia.edu

Olivia Huang: sh3963@columbia.edu

Yihui Wang: yw3361@columbia.edu

Brief description of the proposed visualizations / analyses

As pets lovers, we are interested in finding and visualizing the relationship between pet ownership (e.g. breeds, name) and household basic information (e.g. income, education level) by each neighborhood/zip code in Seattle.

Hypothesis: People with higher income level will raise more pets

Since we are using the Seattle pet dataset, we will focus more on region-level data analysis. In the pet dataset, we have zip code and year column. It would be a good idea to do a time series analysis to see the trend of the most popular pets and their breeds for each year in different areas.

Links to data sources

SOI Tax Stats - Individual Income Tax Statistics - ZIP Code Data (SOI) :

https://www.irs.gov/statistics/soi-tax-stats-individual-income-tax-statistics-zip-code-data-soi

This dataset has individuals' income information by zip codes from 2011 - 2017.

Seattle Pet Licenses:

https://data.seattle.gov/Community/Seattle-Pet-Licenses/jguv-t9rb

This dataset contains a list of active/current pet licenses in Seattle. It has data from 2000 to 2000 about license number, animal names, species, primary breed, secondary breed, and zip codes.

We are still looking for more datasets that can provide other demographic information by zip codes (e.g. number of people in household, race, education level). We'd really appreciate it if you can share some datasets you know about!

Types of visualizations

We intend to use geospatial density maps to demonstrate the distribution of pet ownership in Seattle; line charts and scatter plots to illustrate the relationship between pet ownership and social economic factors, including income and other possible factors; pie charts, bar charts, and word clouds to show comparison among different social categories and highlight categories with the largest frequencies.

Are there any significant hurdles that you have doubts about? Would not solving them render the project incomplete?

We are having a hard time finding datasets with demographic information with specific zip codes due to privacy issues. If we can't find enough variables about household information then we need to shift our focus from visualizing the relationship between pets and demographic characteristics to the characteristics of pet ownership.