2. Data acquisition and clening

2.1 Data sources

A dataset containing coordinates, zipcode, city name, population of the US [6]. Housing data of California containing coordinates and median house values were found from Kaggle dataset [7]. I collected 2012 IRS income by zip code of the US from Data.world [8]. To create the choropleth map of the California, I downloaded California GeoJSON file [9].

**2.2 Data cleaning and preparation**

I sliced California data from the first dataset containing coordinates, zipcode, city name, and population of the US. I merged this dataset with the dataset containing zip code and housing value. I cleaned the minor cities sharing the same coordination and left the primary cities to prevent redundant location data, and merged with the third dataset containing average income by zip code. To get the name of veterinary clinics in California, I used Foursquare API [10].

References

[1] <https://www.stevedalepetworld.com/blog/world-pet-population-data-mixed-bag/>

[2] <https://www.avma.org/KB/Resources/Statistics/Pages/Market-research-statistics-US-pet-ownership.aspx>

[3] <https://www.ftc.gov/sites/default/files/documents/public_events/pet-medications-workshop/daspros.pdf>

[4] <https://www.aavmc.org/assets/site_18/files/annual%20reports/v3_econ_2016_report3_mketvet_061416.pdf>

[5] <https://www.ibisworld.com/industry-trends/market-research-reports/professional-scientific-technical-services/professional-scientific-technical-services/veterinary-services.html>

[6] http://federalgovernmentzipcodes.us

[7] <https://www.kaggle.com/camnugent/california-housing-prices>

[8] https://data.world/jonloyens/irs-income-by-zip-code

[9] <http://boundaries.latimes.com/sets/>

[10] https://foursquare.com