The Distribution of Dogs & Cats in Seattle | Project 2

Rui Ma

December 12, 2017

Introduction

Unbeknownst to a lot of people, Seattle is one of the most pet friendly cities in the United States. Since there are so many trails and parks in the city, it is extremely common for people to encounter pets on the street at any given time of the day.

This project picks three zones (zip codes) in the Greater Seattle area that have the best chance for one to bump into dogs and cats on the street, respectively, based on pet licenses issued between 2013 and 2016. We would be mainly using the "Seattle Pet Licenses" (https://catalog.data.gov/dataset/seattle-pet-licenses) and the Seattle "Zip Codes" (https://catalog.data.gov/dataset/seattle-zip-codes-ebab5) data sets. The variables that we would focus on include the license issue dates, species of the pets, and license ZIP codes from the license data set. The entries from the zip code data set that we are interested in are the zip codes and the areas of the zip code territories. We would also make use of the center coordinates of each zone from the "zipcode" data set contained in the R "zipcode" package.

Analysis and Results

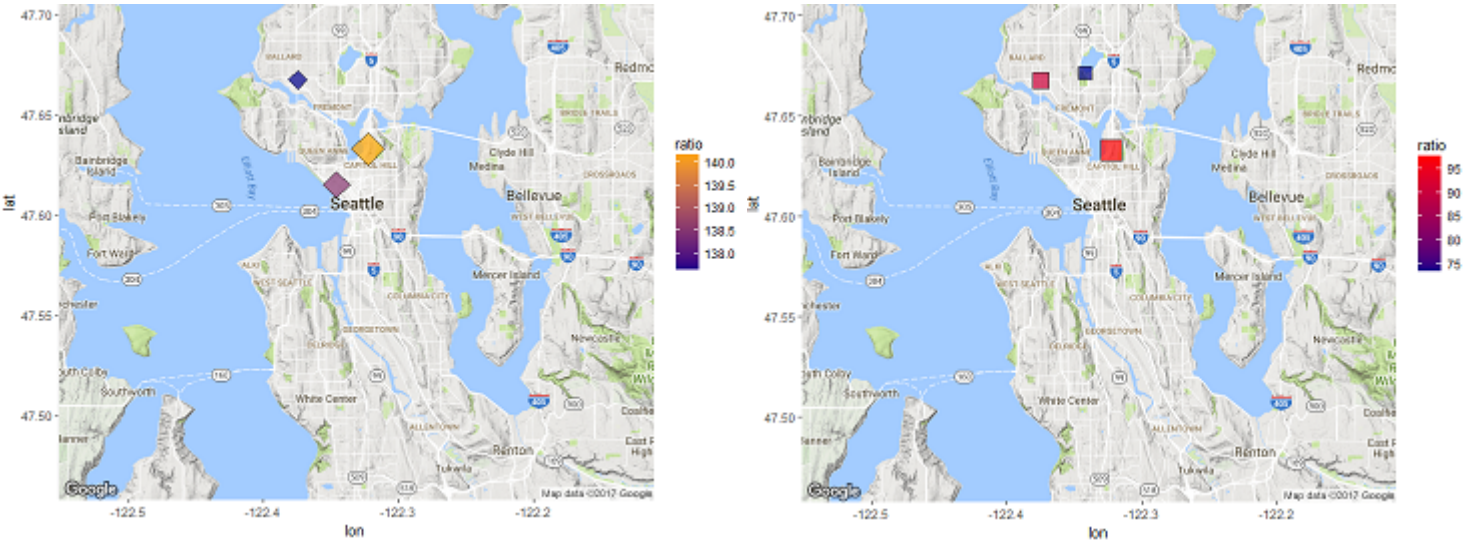
The main idea of our method is to calculate the numbers of dogs and cats per square kilometers for each zip code in the Greater Seattle area and pick the three zip codes that have the highest such ratios. First, we would have to parse the date entries in the license data set and grab all the desired ones (between year 2013 and 2016). Then, we would clean up the zip code entries from both data sets and merge the two by zip codes. After that, we would compute the ratios and sort them in descending order.

The following tables display the results of our calulation for the distribution of dogs (left) and cats (right):

Table 1: Distribution of Dogs vs. Distribution of Cats

zip	ratio	count	area	zip	ratio	count	area
98102	140.1785	1239	8838728	98102	97.07279	858	8838728
98121	138.5254	787	5681267	98107	90.53015	1244	13741278
98107	137.6146	1891	13741278	98103	73.54797	2169	29490956

In order to graph the geographic locations of the above zones (zip codes), we can utilize the "ggmap" package in R. The graphs/maps are as follows (dogs left, cats right):



We notice that **zip code 98102** has both the highest dogs/KM and cats/KM, and the ratios indicate that this region has more dogs than cats. 98121 and 98107 are the 2^{nd} and 3^{rd} respectively on the dog list, whereas 98107 and 98103 are the 2^{nd} and 3^{rd} respectively on the cat list. Coming from Seattle, I actually don't find these results surprising since all these territories, especially the 98102 area, are very densely populated and surrounded by parks.