12.1-FinalProject-Pt4

Brandon Sams

2/27/2020

Being a Dog Owner NYC

Overview

There are few things that can be as emotionally rewarding as having a pet in your home. The sense of companionship and pride that one gets from having a pet is incredibly well recognized. According to an survey from the American Pet Products Association, 67% of US Households own a pet. They also estimate that of the roughly 84.9 million households that own a pet, 63.4 million own a dog. This is by far the most common pet to have in a household, but this should not be news to most.

This research attempts to look at a case study when so many of those households are so very close together. Living in a big city like New York City brings additional advantages and drawbacks to everyday life. It makes sense that living with a dog in NYC would be much the same.

Thus, the problem is that living with a dog in a big city such as NYC may be different from living with a dog in the rest of the country. This changes the calculus for whether or not a person who lives in NYC should own a dog or not.

Data Sources

This research project draws on two primary data sources. First, is a list of Dog License Registrations in NYC. This helps to establish a baseline for what the dog population looks like.

```
RegisteredDogs <- read.csv("/Users/brandon.sams/Documents/RStudio/520 Week 12/DogRegistration.csv")
head(RegisteredDogs)
```

```
##   RowNumber AnimalName AnimalGender AnimalBirthMonth
## 1         1      PAIGE            F             2014
## 2         2       YOGI            M             2010
## 3         3        ALI            M             2014
## 4         4      QUEEN            F             2013
## 5         5       LOLA            F             2009
## 6         6        IAN            M             2006
##                             BreedName ZipCode LicenseIssuedDate
## 1 American Pit Bull Mix / Pit Bull Mix   10035        2014-09-12
## 2                              Boxer   10465        2014-09-12
## 3                            Basenji   10013        2014-09-12
## 4                  Akita Crossbreed   10013        2014-09-12
## 5                            Maltese   10028        2014-09-12
## 6                            Unknown   10013        2014-09-12
##   LicenseExpiredDate Extract.Year
## 1         2017-09-12         2016
## 2         2017-10-02         2016
## 3         2019-09-12         2016
```

```
## 4         2017-09-12         2016
## 5         2017-10-09         2016
## 6         2019-10-30         2016
```

The second data source is a list of Dog Bite incidents in NYC. I like this data source becuase it

```
DogBites <- read.csv("/Users/brandon.sams/Documents/RStudio/520 Week 12/DogBites.csv")
head(DogBites)
```

```
##   UniqueID      DateOfBite   Month Day Year Species
## 1        1 January 02 2015 January   2 2015     DOG
## 2        2 January 02 2015 January   2 2015     DOG
## 3        3 January 02 2015 January   2 2015     DOG
## 4        4 January 01 2015 January   1 2015     DOG
## 5        5 January 03 2015 January   3 2015     DOG
## 6        6 January 05 2015 January   5 2015     DOG
##                               Breed Age Gender SpayNeuter  Borough ZipCode
## 1                   Poodle, Standard   3      M       TRUE Brooklyn   11238
## 2                             HUSKY          U      FALSE Brooklyn   11249
## 3                           UNKNOWN          U      FALSE Brooklyn    NULL
## 4 American Pit Bull Terrier/Pit Bull   6      M      FALSE Brooklyn   11221
## 5 American Pit Bull Terrier/Pit Bull   1      M      FALSE Brooklyn   11207
## 6 American Pit Bull Terrier/Pit Bull   1      F      FALSE Brooklyn   11209
```

A nonzero amount of cleaning has taken place in both data sets. There is additional cleaning that can be done, of course, but these data sets will give us a decent look into what it is like to own a dog in NYC.

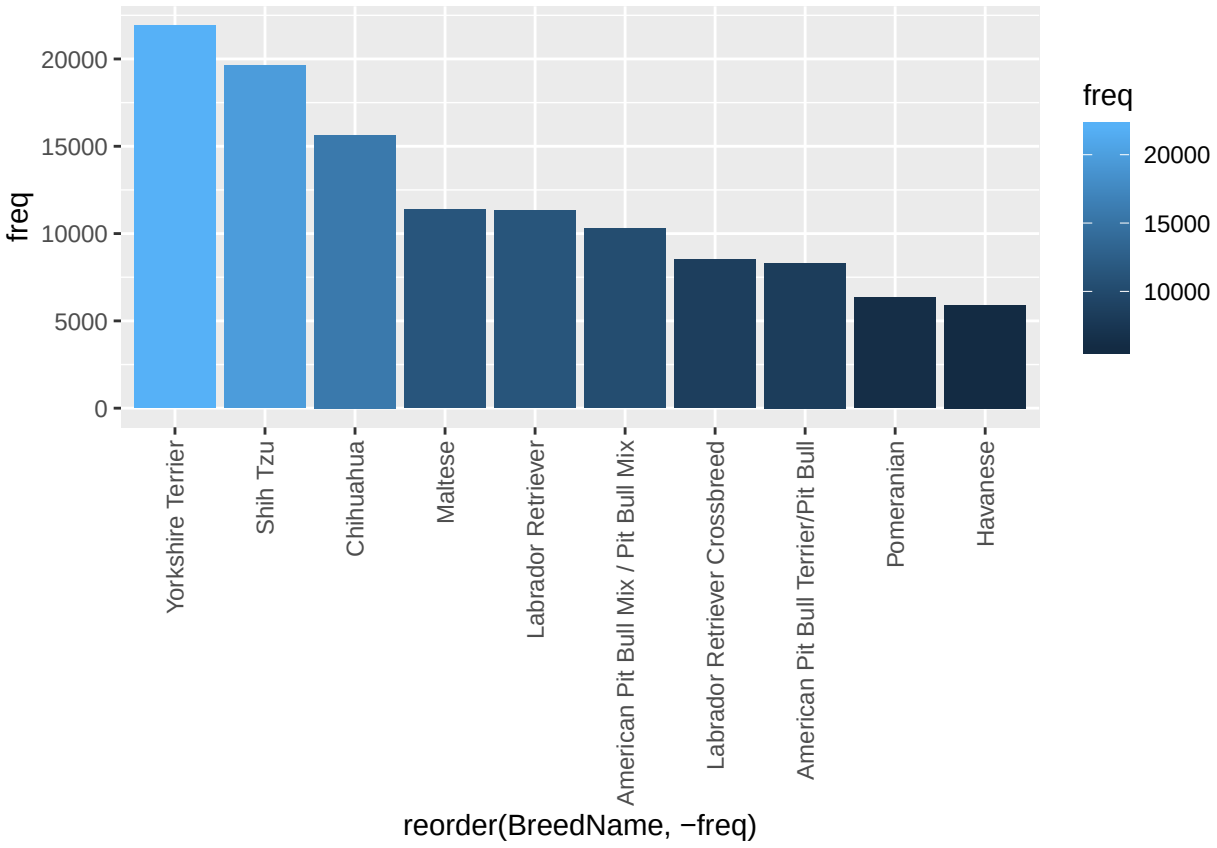
Analysis

We can look at the top ten most popular dog breeds, and one thing that I noticed was that 4 out of the top 5 were small dogs, weighing less than 15 pounds, on average. I anticipate that this has a lot to do with the size of the living conditions in NYC. Apartments are canonically quite small in the area, due to the cost per square foot of most apartments in the area. This likely has a lot to do with it.

```
BreedCount <- count(RegisteredDogs,'BreedName')
BreedCountSorted <- BreedCount[order(-BreedCount$freq),]
# skip the first row, becasue the most common breed is "Unknown"
TopTenBreeds <- head(tail(BreedCountSorted,-1),10)
TopTenBreeds
```

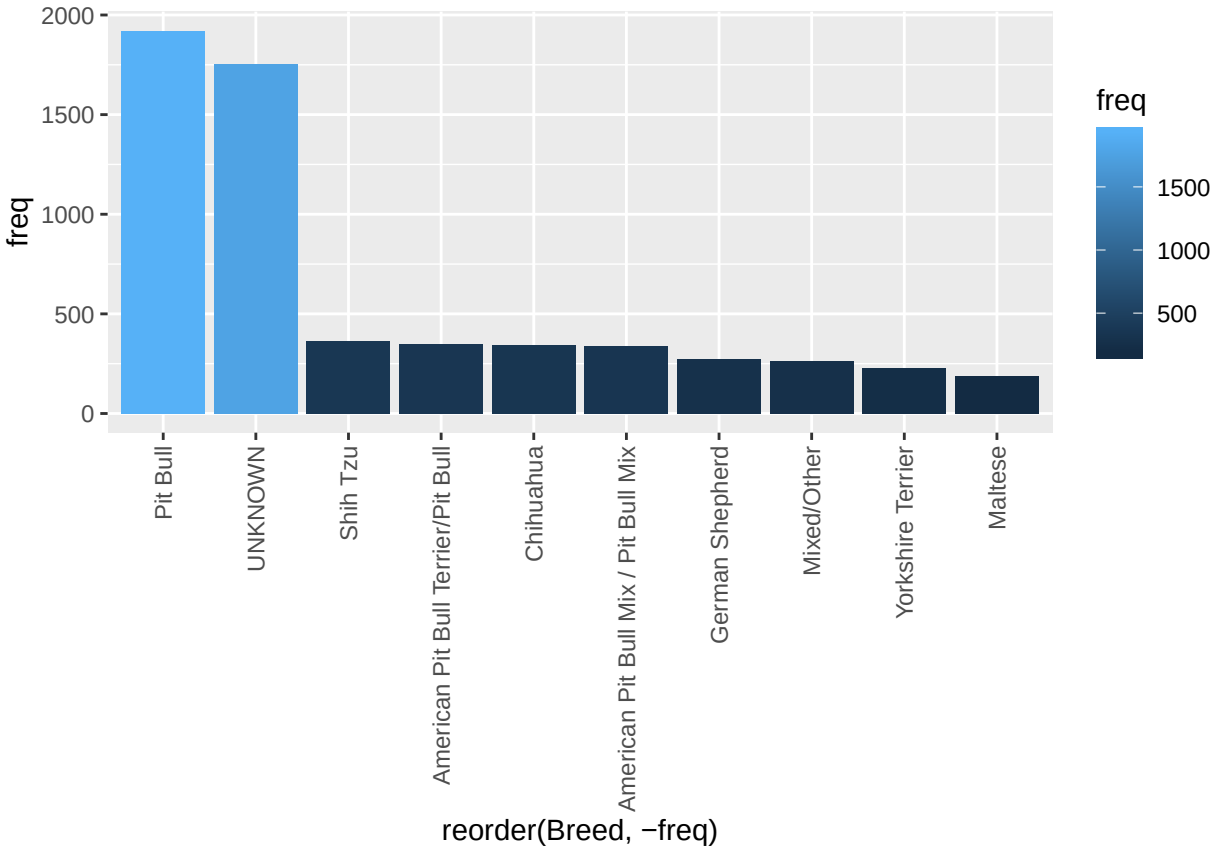
```
##                               BreedName  freq
## 318                   Yorkshire Terrier 21922
## 276                            Shih Tzu 19631
## 83                             Chihuahua 15647
## 188                              Maltese 11391
## 180                   Labrador Retriever 11327
## 14  American Pit Bull Mix / Pit Bull Mix 10304
## 181        Labrador Retriever Crossbreed  8511
## 15    American Pit Bull Terrier/Pit Bull  8319
## 232                           Pomeranian  6345
## 155                             Havanese  5910
```

```
ggplot(TopTenBreeds,aes(x = reorder(BreedName,-freq), y = freq, fill = freq)) + geom_bar(stat = "identi
```



We can also take a look at the dog bite data set, and see which dog breeds appear frequently in that list. I included the "UNKNOWN" this time, because I think it helps to show how frequently the number one breed appears in this list.

```
BiteBreedCount <- count(DogBites,'Breed')
BiteBreedCountSorted <- BiteBreedCount[order(-BiteBreedCount$freq),]
BiteTopTenBreeds <- head(BiteBreedCountSorted,10)
ggplot(BiteTopTenBreeds,aes(x = reorder(Breed,-freq), y = freq, fill = freq)) + geom_bar(stat = "identi
```

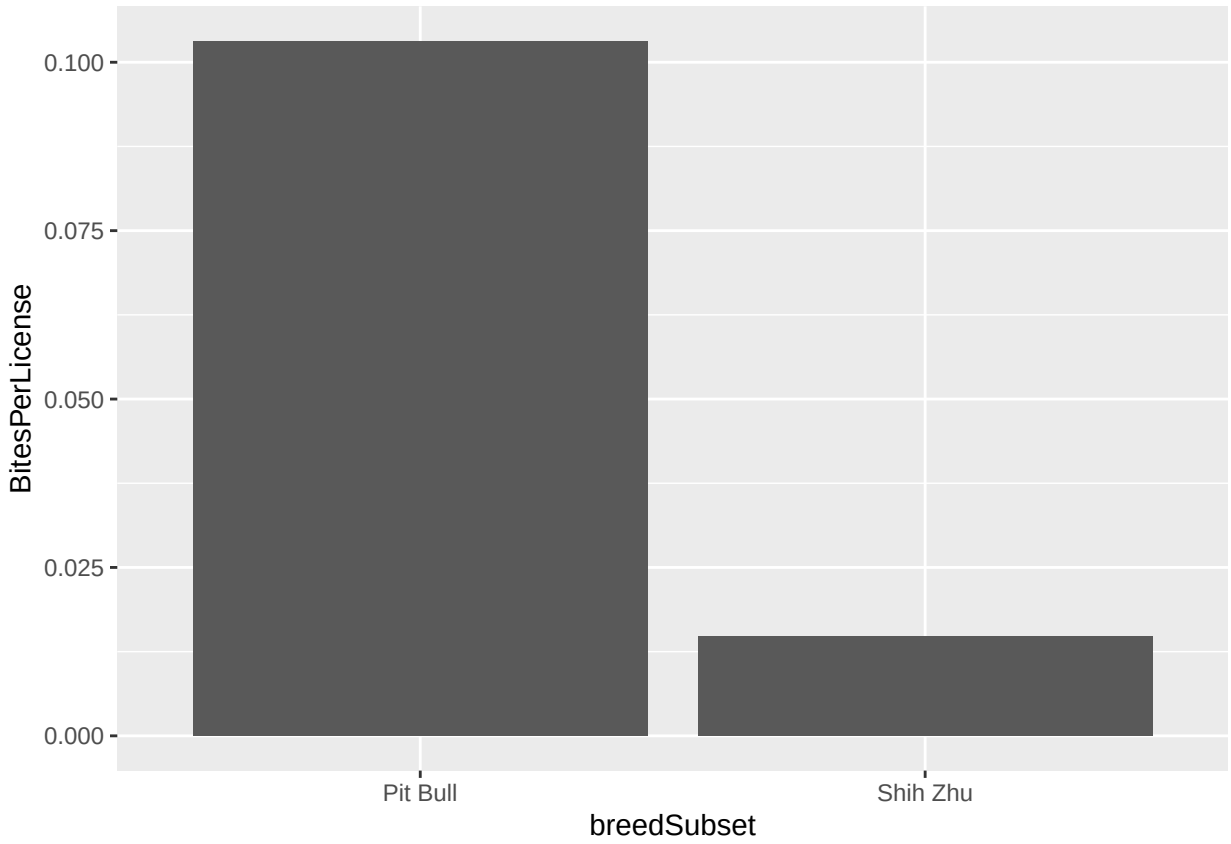


The largest bar, by far, is the "Pit Bull" breed. Of course, a breed will appear more frequently in this list if it is common to own in general. We can adjust for this by finding bites per number of times that breed has been registered. We will take a look at the top two most common breeds by bite, ignoring the "UNKNOWN" breed. That is, Pit Bull and Shih Tzu.

```
breedSubset <- c("Pit Bull","Shih Zhu")
breedLicensed <- c(18623,24661)
breedBites <- c(1921,364)
breedSubsetFrame <- data.frame(breedSubset,breedLicensed,breedBites)
breedSubsetFrame$BitesPerLicense <-with(breedSubsetFrame, breedBites / breedLicensed)
breedSubsetFrame
```

```
##   breedSubset breedLicensed breedBites BitesPerLicense
## 1    Pit Bull         18623       1921      0.10315202
## 2    Shih Zhu         24661        364      0.01476015
```

```
ggplot(breedSubsetFrame,aes(x = breedSubset,y = BitesPerLicense)) + geom_bar(stat = "identity")
```



Even then, the Pit Bull breed still drastically outnumbers the next most common, the Shih Tzu, by an order of magnitude.

I'm not sure what to do with this information exactly, but it is clear that this is statistically significant. This may be a part of the reason why some leases put Pit Bulls on the restricted breeds list, but I would hesitate to recommend that a person strays away from owning a Pit Bull in NYC, because it still may be the right option for some people.

Addressing the Problem

I attempted to find some of the challenges that a dog owner living in New York City could be facing. Due to the data sets that I had available, I looked into dog bites, and cross referenced this with records of what dogs exist in the city. This determined that one of the major problems that a person may encounter is breed restrictions on living spaces.

Implications

This analysis implies that being a dog owner is still a very popular thing to do, even in a crowded city like New York. There were more than 300,000 dog registrations available in my data set, which indicates that there is still an opportunity for a good life to be had, even if you are a dog owner. Just keep in mind breed restrictions when finding a place to live.

Limitations

There were several limitations in the analysis I performed. When being a dog owner in NYC, there are loads more things to consider than just the probability that a dog will bite. Some may even argue that that is such

a rare event, that it may not even be worth considering. I would be inclined to agree, but I was limited by my skill set in R and the data sets I had available. I wanted to cross reference two data sets, and use them for comparative analysis, which I got the chance to do. But yes, more diversity of data sets would be beneficial.

Another limitation is that I only cross compared two breeds, rather than the full list. A breed in one list is not necessarily the same as in the other, and further data cleaning would be required to ensure these two would play well together.