# Report 4

Jie Gu
SID: 913953707

In this report, I will continue to examine the Craigslist data set to make some further inferences on spatial characteristics of the rental market and demography.

Part I: spatial characteristics of the rental market in Davis

In this part, I will investigate some features of rental market in Davis with a map.

Firstly, Let's see how the price of renting an apartment varies in Davis. The following is a dot plot in the map of Davis
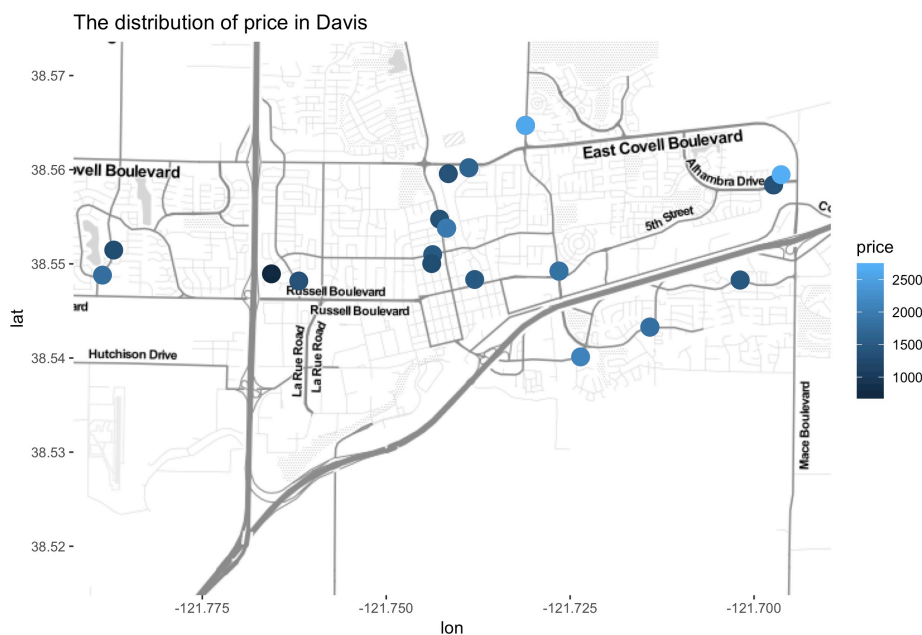
In this plot, we can see that the price of renting an apartment is relatively low in the west of Davis; the price is relatively high in the east of Davis. The price of renting an apartment in east of Davis is even higher than that near downtown, which is approximately in the middle part of the map. The reason for this may be that the apartments in the east of Davis is relatively new compared with other parts of Davis. Also, many students of UC Davis tend to live in the east part of Davis. Therefore, the price in east is relatively higher.

After examine the price, we can also study the pets' policy in Davis. More precisely, we can check whether there exist some areas in which most apartments allow pets. That is to say, if the pets' policy is related to location, where should we go to rent an apartment assuming that we have pets. Now, let's take a lot at a dot plot of whether pets are allowed in the map of Davis.
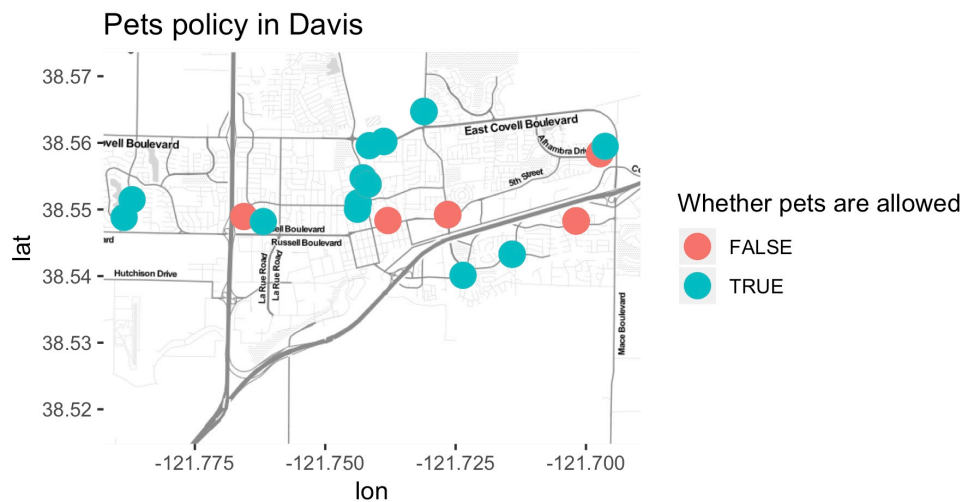
In this plot, 'FALSE' represents pets are not allowed; 'TRUE' represents pets are allowed. We can see that most of apartments allow pets in Davis. Also, along the high way in the east of Davis, the number of apartments that do not allow pets is similar to the number of apartments that allow pets. In addition, most apartments in middle part of Davis allow pets. The situation in the west of Davis is similar to the middle part. Therefore, if we have a pet, we can go to the west of Davis or middle part of Davis to rent an apartment.

Moreover, we can solve the problem that which part of Davis should we go if we want to rent an apartment with more bedrooms and relatively large size with this data set. To solve this problem, we can use a dot plot with color representing square feet of apartments and size representing number of bedrooms. The following is the plot.

Situation about size and bedrooms of apartments

In this plot, we can know that the size of apartments in the middle part is relatively small. Even though some apartments have 3 or 4 bedrooms in the middle part, their size is still small. The reason may be that these apartments are near downtown, and it is really normal that the apartments near downtown are relatively small since most areas of downtown are used for commerce instead of apartments. Additionally, many of the apartments in the north-east of Davis have more than 2 bedrooms and relatively large size compared with other parts in Davis. Therefore, if we want to rent an apartment with more bedrooms and relatively large size, we can go to the north-east of Davis.

Although we have many findings on the price, pets' policy, size and the number of bedrooms of apartments in Davis, these findings all have limitations. The limitation is that the number of posts of apartments in Davis is too small. That is to say, the sample size is too small. So our findings may not be precise since maybe these posts cannot represent a general situation in Davis. Hence, we need more observations to investigate these features.

Part II: spatial characteristics of the rental market in southern half of the San Francisco Bay Area

In this part, I will investigate some features of rental market in southern half of the San Francisco Bay Area like what I did for Davis.

At first, similar to the exploration for Davis, let's see how the price of renting an apartment varies in southern half of the San Francisco Bay Area. The following is a dot plot of price in the map of southern half of the San Francisco Bay Area.
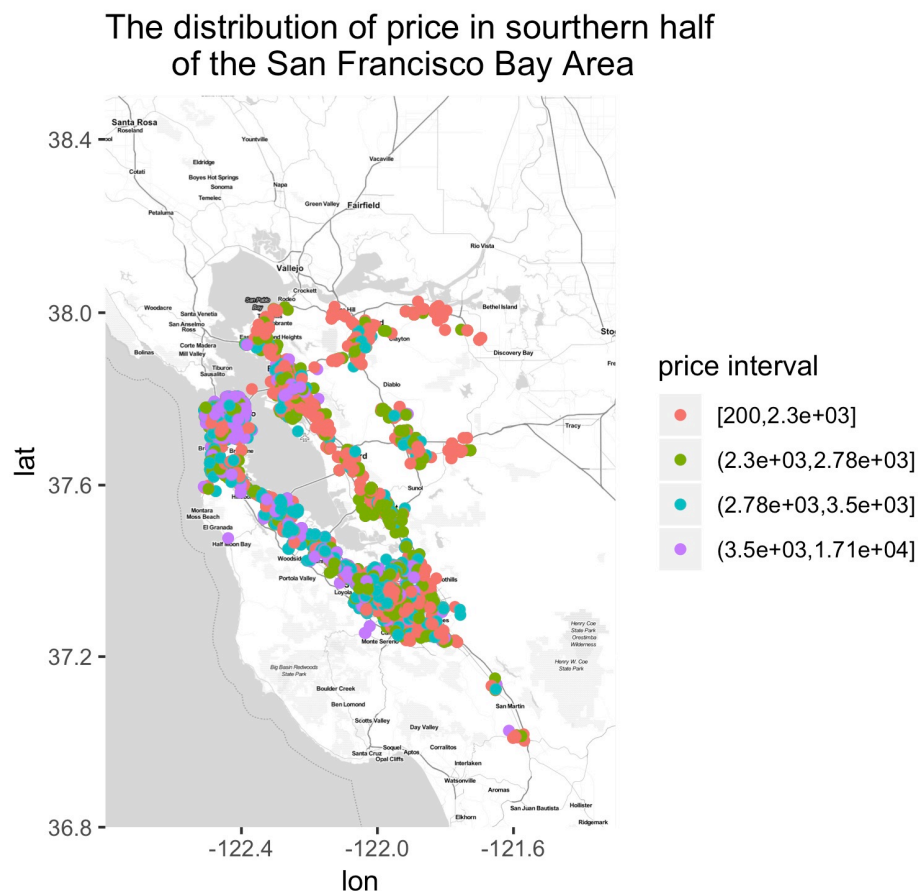
In this plot, we can see that the north-west part of this area, which is San Francisco, has many apartments with price over $3500. Also, the price of renting an apartment in San Francisco is relatively higher compared with other parts. The price of renting an apartment varies a lot in the South part of this area, which is San Jose. Moreover, the price of renting an apartment in north-east part of this area, which is Contra Costa county, is relatively lower than other parts. These findings can be explained by the economics and job opportunities in each area. For instance, the economic situation in San Francisco is good and there are many job opportunities there, so the price of

renting an apartment is higher. For people who want to rent an apartment with a low price in southern half of the San Francisco Bay Area, they can go to Contra Costa county.

Now, since many people have cars, let's pay attention to the parking situation in southern half of the San Francisco Bay Area. More precisely, we can solve the problem that where people should go to rent an apartment if they want garages. Here is a dot plot of parking situation in southern half of the San Francisco Bay Area.

Parking Situation in southern half of the San Francisco Bay Area
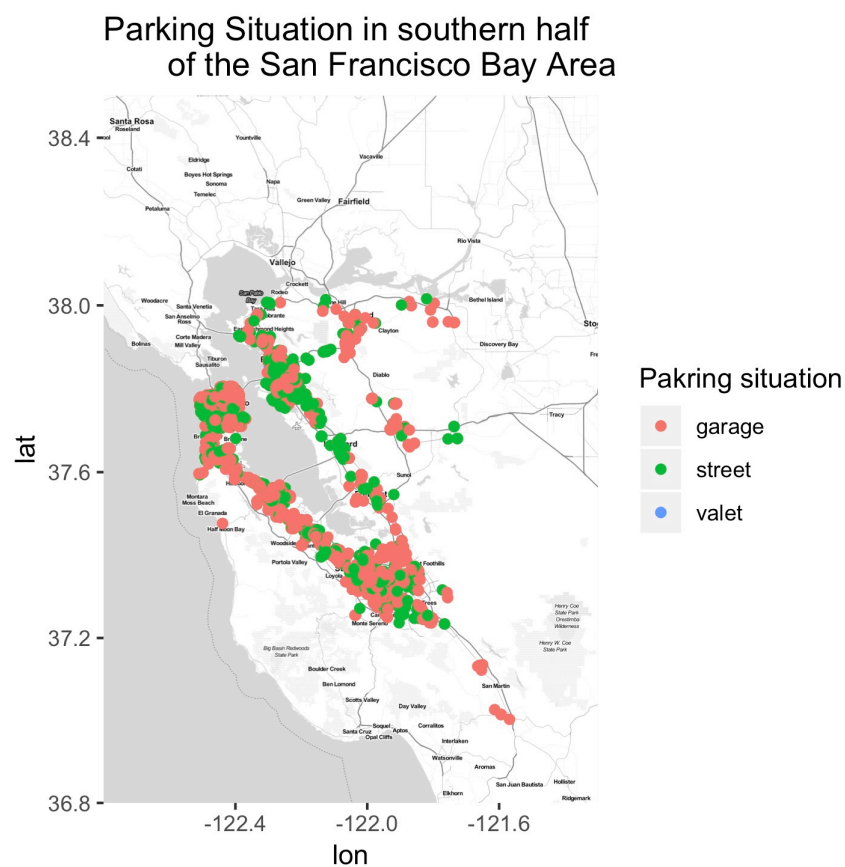
From this graph, we can know that the proportion of apartments with garage is higher in San Jose compared with that in San Francisco. Therefore, for a person who want to work in major cities and have a requirement for garage, he can choose San Jose. Besides, along the coastline from San Jose to San Francisco, the proportion of apartments with a garage is high. This proportion is also high in the north-east part of southern half of the San Francisco Bay Area, which is Contra Costa. While between San Francisco and Contra Costa, which is Berkeley, the proportion of apartments with

a garage is low. Finally, we almost cannot see apartments with valet parking in the graph since this situation is rare. Therefore, people can go to the areas along the coastline from San Jose to San Francisco and Contra Costa if they have a requirement for garage.

Finally, we can check pets' policy to solve that if the pets' policy is related to location, where should we go to rent an apartment assuming that we have pets, just like I did for Davis.
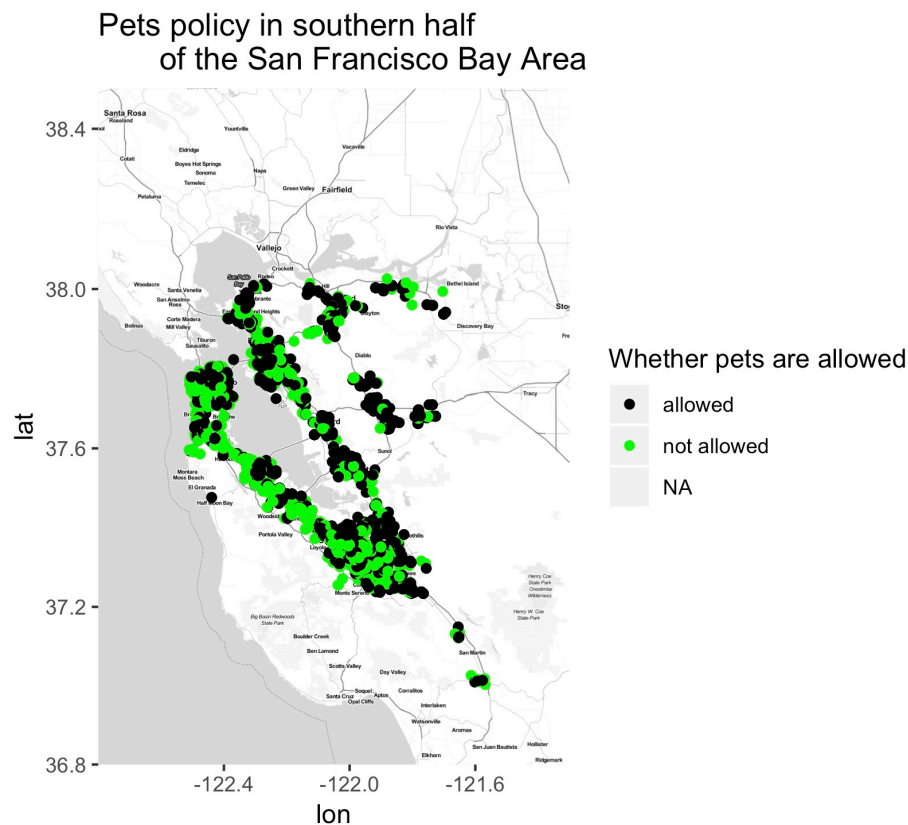

From this graph, the proportion of apartments which allow pets is lower than that of apartments which do not allow pets in San Francisco and the areas along the coastline from San Jose to San Francisco. Also, the number of apartments which allow pets seems to be similar to the number of apartments which do not allow pets in San Jose. As for Contra Costa county, the proportion of apartments which allow pets is relatively higher compared with other parts of southern half of the San Francisco Bay Area. Hence, for people who have a pet, they can choose Contra Costa county to rent

an apartment. If they also want to live in big city, San Jose is better than San Francisco from the aspect of pets.

In general, I can make stronger conclusion about the San Francisco Bay Area than I could about Davis since the number of posts about rental information is much larger in this area than in Davis. In other words, since the sample size in this area is much bigger than that in Davis, the plot for this area can be more representative than the plot for Davis. Therefore, I can make a stronger conclusion.

Part III: Which places in the southern San Francisco Bay Area have the oldest populations? How does this relate to the rental market, if at all?

Firstly, the places having the oldest populations means having the highest percent of 65 years or older. Thus, the places in the southern San Francisco Bay Area with the oldest populations is Walnut Creek. To solve how this relates to the rental market, I analyze the percent of 65 years or older with price. Let's look at a dot plot about the relation between percent of 65 years or older and mean price.

The relation between percent of 65 years or older and mean price
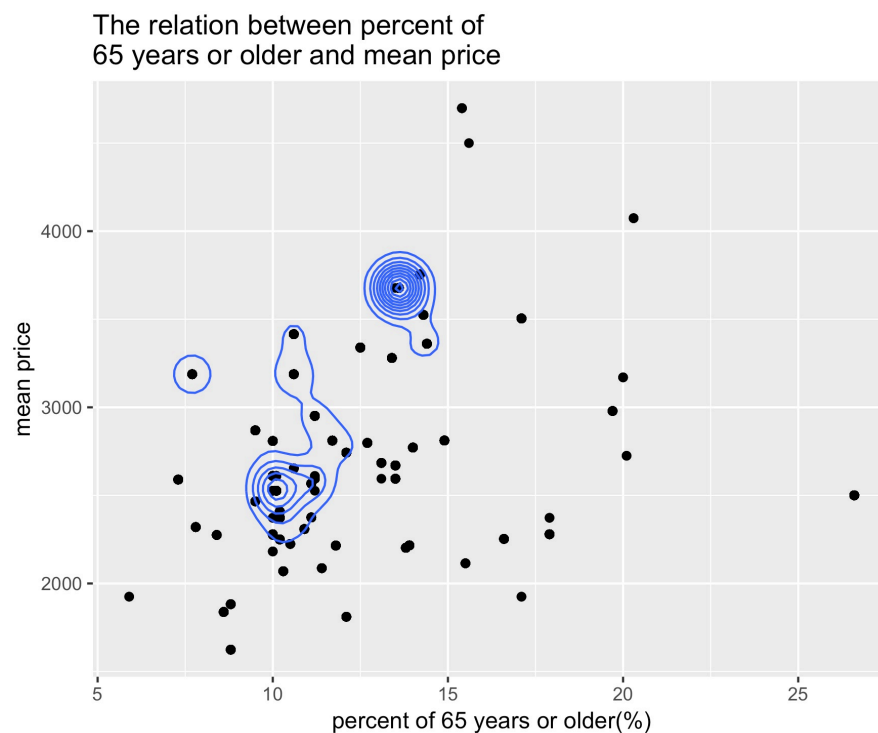
From this graph, we can see that there may exists a linear relationship between percent of 65 years or older and mean price. More precisely, with the increase in percent of 65 years or older, the mean price of renting an apartment also increase. However, this relation is not true for all observations. For instance, Walnut Creek, the city with the oldest population, seems to be an outlier in the graph. The mean price of renting an apartment in Walnut Creek is lower than that in almost half of the places in the southern San Francisco Bay Area. Therefore, this relationship cannot apply to all of the places in the southern San Francisco Bay Area.

Appendix

```r
library(ggplot2)
library(ggmap)
library(Lahman)
library(stringr)

#Before solving questions, read data and adjust data
#================================================
cl = readRDS('cl_apartments.rds')



#adjust the data like I did in assignment3
cl_ca <- subset(cl,state == 'CA')
cl_no_duplicate = cl_ca[!duplicated(cl$title),]
cl_no_duplicate =
cl_no_duplicate[!duplicated(cl_no_duplicate$text),]
cl_no_duplicate$price[cl_no_duplicate$price>=30000000] <-
3408
cl_no_duplicate$price[cl_no_duplicate$price==9951095] <-
995
cl_small_price <- subset(cl_no_duplicate,price <= 100)
cl_no_duplicate$price[cl_no_duplicate$price <= 100] <- NA
cl_no_duplicate$sqft[cl_no_duplicate$sqft >= 10000] <- NA
head(sort(cl_no_duplicate$sqft))
cl_no_duplicate$sqft[cl_no_duplicate$sqft <= 3] <- NA
cl = cl_no_duplicate

#Read Census data
ann = read.csv('DEC_10_SF1_SF1DP1_with_ann.csv',header =
TRUE)

#see the feature of data
```

```r
names(ann)
head(ann$GEO.display.label)

#remove extra information in GEO
ann$GEO.display.label = str_remove_all(ann[,3],'
(city|CDP),| California')

#Check difference in two variables
setdiff(cl$city,ann$GEO.display.label)

#merge the data
cl_ann = merge(cl,ann,by.x = 'city',by.y =
'GEO.display.label')




#Q1:Use maps to explore at least 3 features of the
Craigslist posts in Davis.
#==========================================================
=====

#get the data for davis
Davis <- subset(cl_ann,city == 'Davis')


#get the map for daivs
bbox = c(
  -121.792517,38.514802,
  -121.689546,38.573602
)#From piazza
m = get_stamenmap(bbox, zoom = 13, maptype = "toner-
lite")
```

```r
#Q2:question1: How does the price of aparments vary in
Davis
ggmap(m) + geom_point(aes(x = longitude,y =
latitude,color = price),Davis,size = 5) +
  labs(title = 'The distribution of price in Davis')

#Q2:question2: if the pets??? policy is related to
location,
#where should we go to rent an apartment assuming that we
have pets.

table(Davis$pets)
#add new variables to Davis to record whether a pet is
allowed.
pets_policy = (Davis$pets == 'both' | Davis$pets ==
'cats' |
              Davis$pets == 'dogs' | Davis$pets ==
'negotiable')
Davis$pets_policy= pets_policy

#remove na in pets
Davis_no_na_pet <- subset(Davis,is.na(pets_policy) !=
TRUE)

#a dotplot of pets_policy in Davis
ggmap(m) + geom_point(aes(x = longitude,y =
latitude,color = pets_policy),Davis_no_na_pet,size = 5) +
  labs(color = 'Whether pets are allowed',title = 'Pets
policy in Davis')


#Q2:question3: which part of Davis should we go
```

```r
#if we want to rent an apartment with more bedrooms and
relatively large size
ggmap(m) + geom_point(aes(x = longitude,y = latitude,size
= as.character(bedrooms),color = sqft),Davis) +
  labs(size = 'number of bedrooms',title = 'Situation
about size and bedrooms of apartments')




#Q2:Repeat question 1 for the southern half of the San
Francisco Bay Area
#(counties San Francisco, San Mateo, Santa Clara,
Alameda, and Contra Costa).
#============================================================
=====

#get the data for San Francisco Bay Area
county_name = c('San Francisco','San Mateo','Santa
Clara','Alameda','Contra Costa')
SF_BayArea = cl_ann[cl_ann$county%in%county_name,]

#get the map for San Francisco Bay Area
bbox2 = c(-122.80,36.8,
        -121.3,38.5)
m2 = get_stamenmap(bbox2, zoom = 10, maptype = "toner-
lite")



#Q2:question1:how the price of renting an apartment
varies
#in southern half of the San Francisco Bay Area.
```

```r
#cut the price into four interval
SF_BayArea$price_interval = cut_number(SF_BayArea$price,
4)
#get the dotplot of price in map
SF_BayArea_noNA_price <-
subset(SF_BayArea,price_interval !=
is.na(price_interval))
ggmap(m2) + geom_point(aes(x = longitude,y =
latitude,color = price_interval),SF_BayArea_noNA_price) +
  labs(title = 'The distribution of price in sourthern
half
     of the San Francisco Bay Area',color = 'price
interval')



#Q2:question2:parking situation in southern half of the
San Francisco Bay Area.

#Define different parking type
SF_BayArea$Parking_situation[SF_BayArea$parking%in%c('str
eet','off-street')] = 'street'
SF_BayArea$Parking_situation[SF_BayArea$parking%in%c('gar
age')] = 'garage'
SF_BayArea$Parking_situation[SF_BayArea$parking%in%c('val
et')] = 'valet'
#remove NA in parking_situation
SF_BayArea_parking_noNA <-
subset(SF_BayArea,Parking_situation !=
is.na(Parking_situation))
#get a dot plot of parking in the map
```

```r
ggmap(m2) + geom_point(aes(x = longitude,y =
latitude,color =
Parking_situation),SF_BayArea_parking_noNA) +
  labs(title = 'Parking Situation in southern half
      of the San Francisco Bay Area',
      color = 'Pakring situation')



#Q2:question3:check pets??? policy to solve that if the
pets??? policy is related to location,
#where should we go to rent an apartment assuming that we
have pets
#add new varibles to record hwther pets are allowed
SF_BayArea$pets_policy[SF_BayArea$pets%in%c('both','cats'
,'dogs','negotiable')] = 'allowed'
SF_BayArea$pets_policy[SF_BayArea$pets%in%c('none')] =
'not allowed'

#get a dot plot of pet policy in map
ggmap(m2) + geom_point(aes(x = longitude,y =
latitude,color = pets_policy),SF_BayArea) +
scale_color_manual(values = c('allowed' = 'black','not
allowed' = 'green')) +
  labs(title = 'Pets policy in southern half
      of the San Francisco Bay Area',
      color = 'Whether pets are allowed')



#Q3:Which places in the southern San Francisco Bay Area
have the oldest populations?
#How does this relate to the rental market, if at all?
```

```r
#I have merged the data at the begining of the code.
#tansfer discrete data to continuous data
SF_BayArea$HD02_S025 = as.character(SF_BayArea$HD02_S025)
SF_BayArea$HD02_S025 = as.double(SF_BayArea$HD02_S025)

#To check which place have the oldest population
max(SF_BayArea$HD02_S025)
SF_BayArea[SF_BayArea$HD02_S025 ==
max(SF_BayArea$HD02_S025),]$place



#To get the mean price in each palce
mean_price = aggregate(price~place,SF_BayArea,mean)
names(mean_price) = c('place','mean_price')

#to get a dot plot of the percent over 65 years vs mean
price
SF_BayArea_meanprice = merge(SF_BayArea,mean_price,by =
'place')
ggplot(SF_BayArea_meanprice,aes(x = HD02_S025,y =
mean_price)) + geom_point() +
  geom_density2d() +
  labs(x = 'percent of 65 years or older(%)',
  y = 'mean price',
  title = 'The relation between percent of
65 years or older and mean price')
```