Nathan J. Chan
November 2, 2018

UC Davis, STA 141A
Professor Ulle

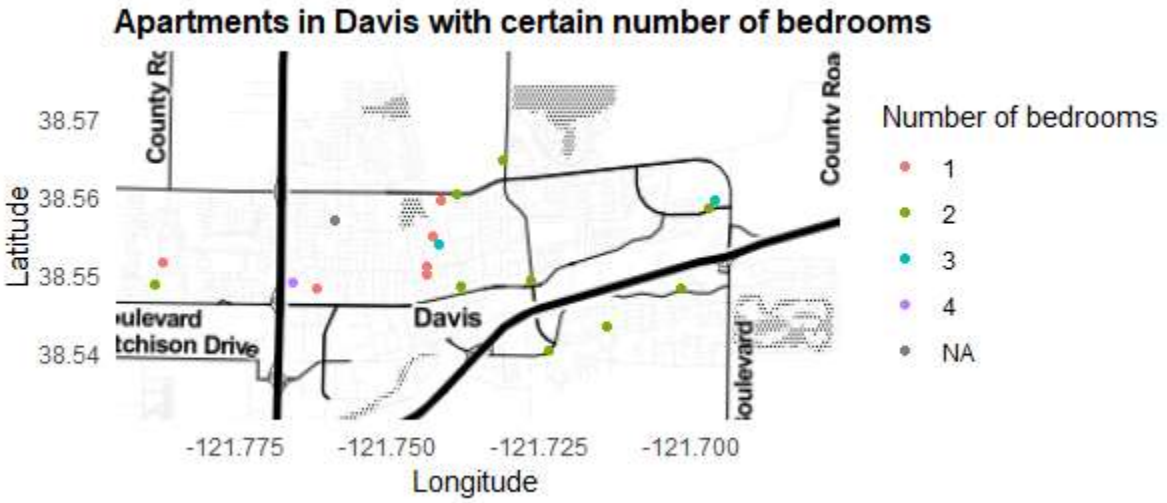
ASSIGNMENT FOUR

1. Analyzing apartments in Davis on a map

In this section, we will explore apartments in Davis, California using a map. The data is from about 50 apartment listings on Craigslist. Note that some of the data points include duplicates, and will not appear on the map.

A. In which areas of Davis are there apartment sizes of different types?

There may be certain areas of Davis that developed and grew at different times, thereby creating areas with apartments of different sizes. Here is a map of Davis, with the apartments on the map and markers to indicate how big each apartment is.



Since the data set is so small, as there were not many Davis apartment listings on Craigslist, it is hard to make any solid conclusions, if any. Without much data, it wouldn't be terrible to simply say that we cannot find any patterns at all!

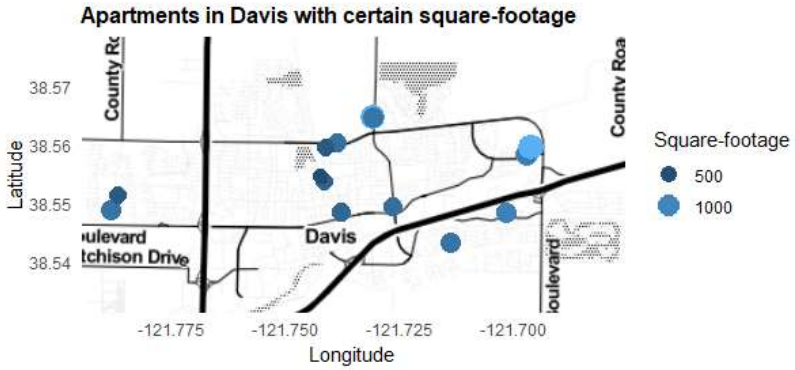
Because of the ever-growing number of students who attend UC Davis, the Davis apartment vacancy rate was 0.4% in January 2018.[1] In the entire United States, the apartment vacancy rate was 4.7% in the first quarter of 2018 (which is the same time as January 2018).[2] There will be far fewer listings in Davis than in other parts of California, and it is a consequence of circumstance.

What I can say is that many of the apartments are along main roads, because most of the areas in Davis are residential neighborhoods, with houses. There is an even spread of apartments with different numbers of bedrooms: no size of apartment clearly dominates.

---

[1] https://www.ucdavis.edu/news/uc-davis-study-city-apartment-vacancy-rate-essentially-unchanged-rental-rates/

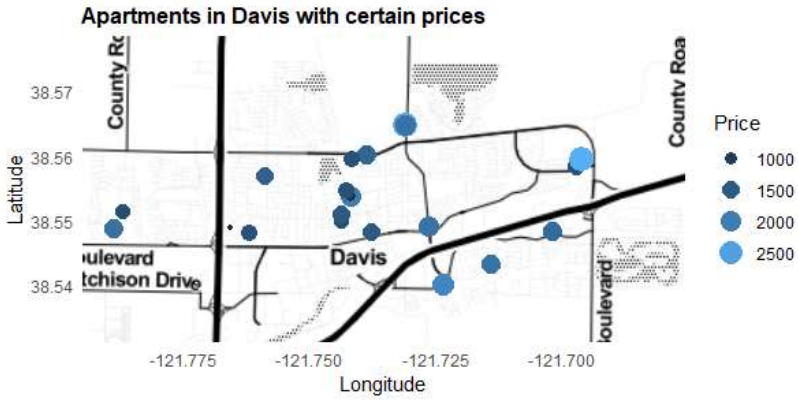
[2] https://www.reuters.com/article/property-usa-apartment/u-s-apartment-vacancy-rate-up-slightly-in-first-quarter-reis-idUSL3N1R95E5



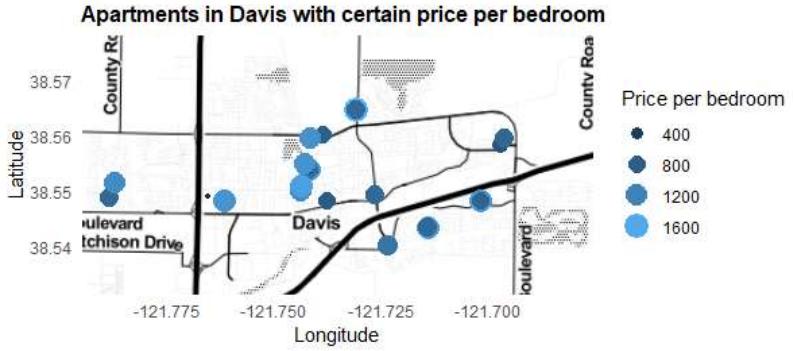
Looking at the apartments by square-footage conveys similar results. The apartments on the main roads (which are darker on the map) are larger than the other apartments near the middle of the map, on a less-traversed road. (These claims are not very sturdy as we lack evidence.)

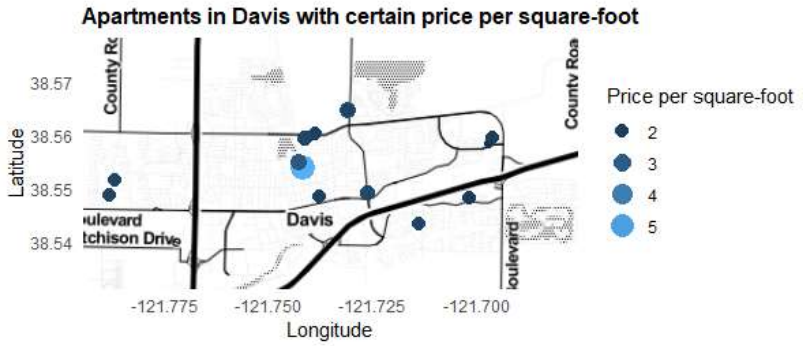
B. Are apartments closer to UC Davis more expensive than apartments farther away?

Answering this question would benefit UC Davis students looking for apartments, who must weigh the cost of being close to campus and, perhaps, the extra price of rent.



Contrary to what I expected, apartments up for rent near UC Davis appear cheaper than those farther away, especially those just north of UC Davis. (Note that UC Davis is in the bottom left area of the word “Davis” on the map.) This may be because these cheaper apartments are also smaller. Here are two maps with price per bedroom and price per square-foot:



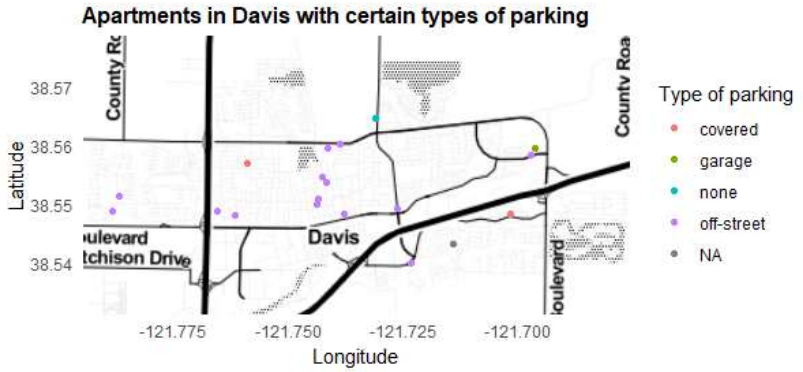


The reason the apartments closer to UC Davis seem cheaper is because they have fewer bedrooms, or are smaller. But in comparison to the other apartments, price per bedroom or price per square-foot is about the same or even more expensive. By looking at more descriptive data, we figured out our initial inclination was correct: that apartments closer (by closer, I mean those just north of UC Davis, because those are more accessible compared to apartments east of UC Davis) to UC Davis are more expensive when taking size into account.

Again, we do not have enough data to make solid conclusions. The data set is very small. There are large swaths of land on the map that don't have any apartment listings at all. It is impossible to form any sort of analysis better than a mild claim with little evidence. That is simply the nature of the data set.

C. Is there a connection between where an apartment is located and the type of parking available?

Different areas in Davis may have different types of parking because of how different communities are organized. Spacious areas may have dedicated parking, and smaller apartment areas may rely on street parking, or have no parking at all. Here is a map with each point representing the type of parking each apartment has:

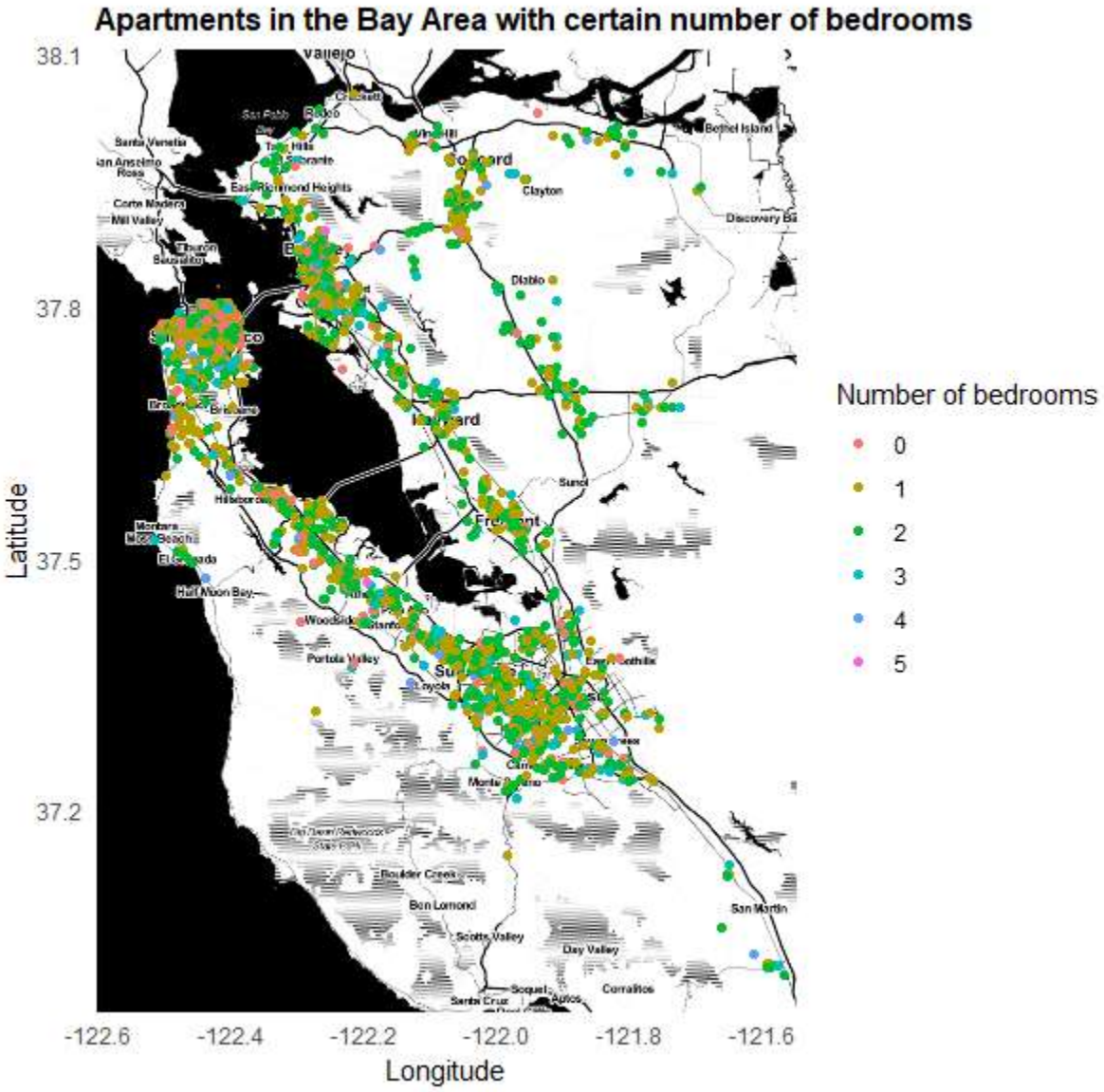


Most apartments in Davis have off-street parking, meaning these apartments all have their own dedicated parking areas. It doesn't seem that specific areas affect this trend, as most areas have apartments with dedicated parking. However, this is not the case for all apartments. There are a few apartments without parking, but there is not enough information to make claims about the entire area. The same can be said about apartments with parking: we don't have enough data to say that the area in which an apartment resides affects the type of parking available.

2. Analyzing apartments in the Bay Area on a map

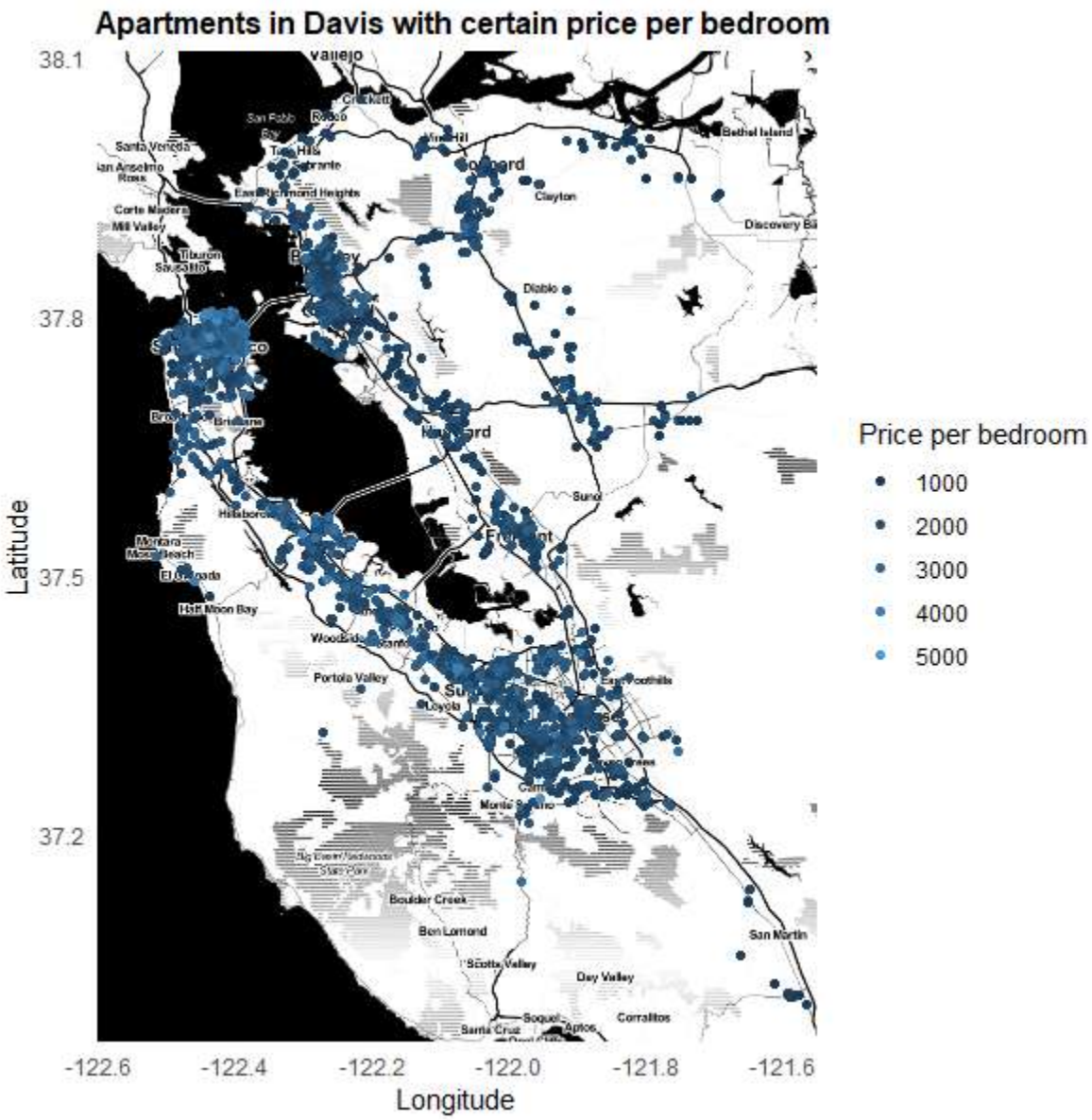
The lack of data in Davis led to frivolous conclusions. Next, we will examine data from the Bay Area where the square footage is many times larger and apartment listings are abundant. Specifically, we will look at the counties of San Francisco, San Mateo, Santa Clara, Alameda, and Contra Costa. We will use the same questions, at least, to the extent possible, as above.

A. In which areas of the Bay Area are there apartment sizes of different types?



There is a good mix of apartment sizes throughout the Bay Area. Most of them are 1 or 2 bedroom apartments. There are some studio apartments (where the number of bedrooms is 0) scattered throughout, and most of them are in San Francisco. The larger apartments are also all over the place. The only pattern is that there is no pattern; where one goes in the Bay Area, he or she can expect apartments of all sizes.

B. Are apartments close to popular areas more expensive than apartments farther away?

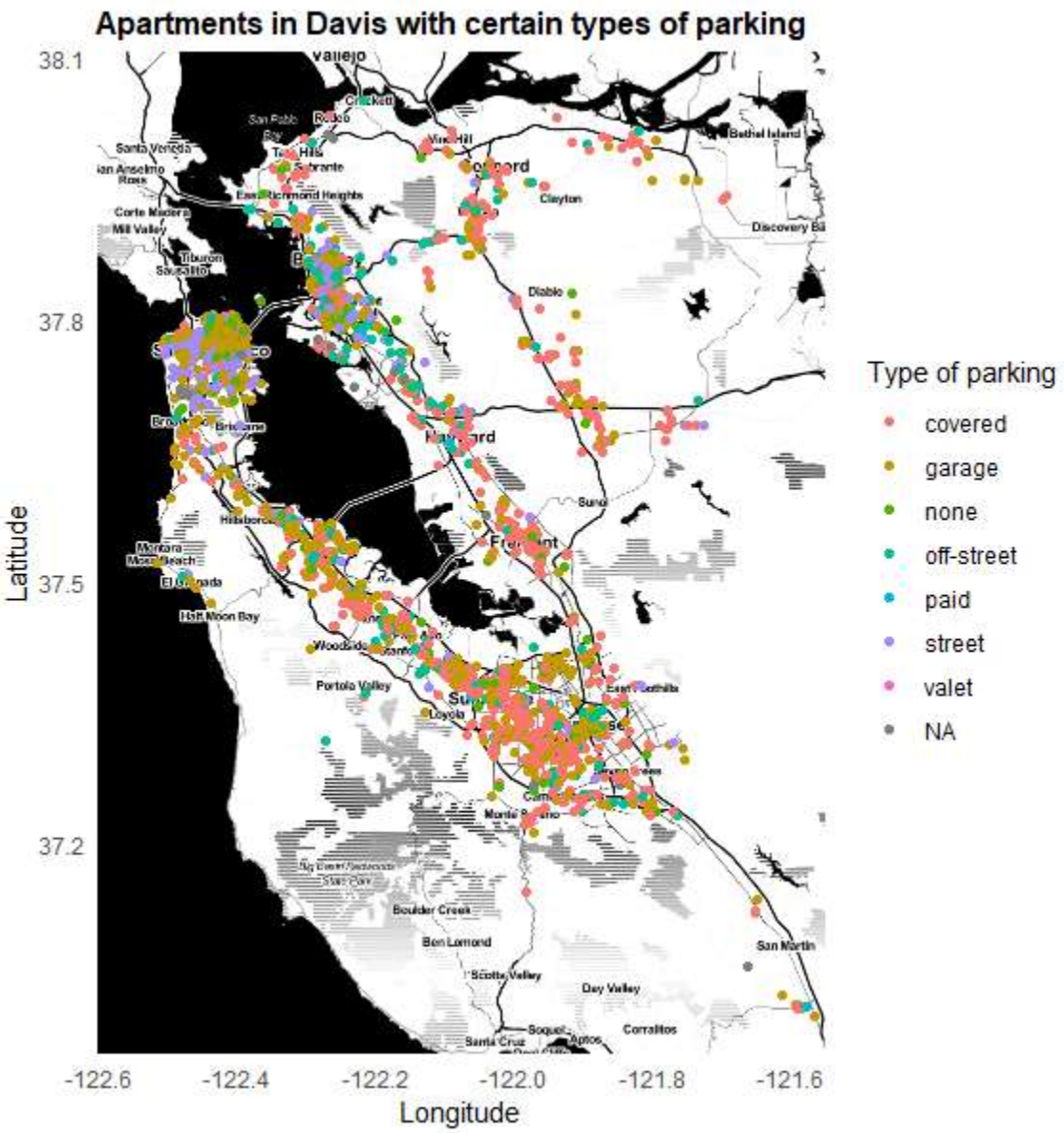


Note: studio apartments are not included on this map.

The cost per bedroom seems similar throughout the Bay Area. There are several high cost apartments scattered around, and there is a high prevalence of expensive apartments in San Francisco. As the we get farther away from the bay, there are fewer expensive apartments.

It appears that, yes, more popular area have more expensive apartments. In San Francisco, where all the technology companies are, the apartments are more expensive. On the west side of the bay, there are more high cost apartments compared to the east side of the bay, likely also because of the tech companies in Silicon Valley. There are a high number of apartments in Berkeley and San Jose from UC Berkeley and the large city of San Jose, although the apartments are not as expensive.

C. Is there a connection between where an apartment is located and the type of parking available?



There are a number of easily visible trends on this map. In San Francisco, Berkeley, and Oakland, there is a high number of apartments with street parking that anywhere else on the map. San Francisco also has a lot of garage parking. In San Francisco, space is rare, so apartment complexes must build parking garages underground or utilize whatever street parking is available, or have no parking at all. Berkeley and Oakland, even if it has more street parking than other places, also has a good mix of all sorts parking options. There seems to be more paid parking in Berkeley and Oakland than in other parts of the Bay Area.

Looking at the east side of the bay and further inland (not including Berkeley and Oakland) there is a lot of garage parking and covered parking, likely because there is more space available for dedicated parking structures compared to the city. (Apartments not in the city with garages

usually have the garage built underneath the apartment above ground.) The areas south of San Francisco, along the west side of the bay, and San Jose follow the same pattern, with most of the parking being garage or covered parking.

This finding is surprising, because San Jose and San Francisco are both large cities, but the parking available in the two cities are very different. San Jose must be less starved for space than San Francisco, since there is dedicated parking for apartment residents, compared to the reliance on street parking in San Francisco.

D. Analysis in Davis versus the Bay Area

It is much easier to make conclusions about apartments in the Bay Area than in Davis because there is more data available. With more data, we can make out a signal, rather than stumbling around the noise. More data gives us stronger evidence to back up our claims.

The downside of having so much data is that, simply, there is too much data. Because the Bay Area is so large and our map can only be so large, it's harder to see any patterns, and our analysis is dependent on the map, so it makes it harder to draw conclusions. A point on the map must be large enough to see, but small enough to not obscure the other points. There are some solutions, but they aren't very effective: changing the shape would make it even harder to see patterns, changing the transparency would cause the colors to overlap and make new colors, and changing the size will, of course, obscure other points.

Next time, we should limit our scope. Instead of looking at an entire region, we could instead focus on a city, like we did with Davis, but the city we choose would be a bit larger and also have an abundance of apartment listings, like San Francisco or San Jose.

3. Oldest populations in the southern Bay Area and rental market effects

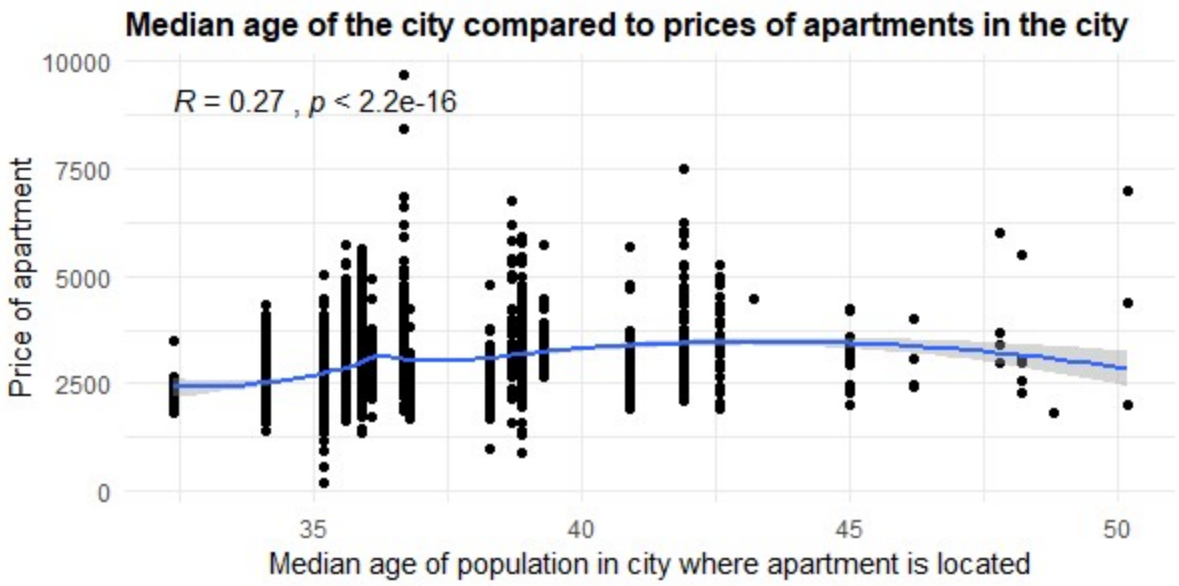
We will use data from the 2010 US Census to identify the oldest populations in the southern Bay Area whether they have any effect on the rental market.

I will define the southern Bay Area as the combination of the following cities: Fremont, Redwood City, Mountain View, San Mateo, Menlo Park, Palo Alto, San Carlos, Atherton, Portola Valley, Foster City, Belmont, Los Altos Hills, Half Moon Bay, Los Altos, Woodside, San Jose, Campbell, Gilroy, Los Gatos, Santa Clara, Sunnyvale, Milpitas, Cupertino, Morgan Hill, and Saratoga.

I will define "oldest populations" as places which have the highest median age from the total population. The top five cities with the oldest populations are the following:

City	Median age
Los Altos Hills	50.2
Woodside	48.8
Atherton	48.2
Saratoga	47.8
Los Altos	46.2

I took the median ages of each city and plotted it against the prices of the apartments in each city. Here is the plot:



The relationship between the two variables is fairly weak. The regression line is almost completely straight. This means that as the median age of the population in the city increases, there is no change in the price of apartment.

However, there do appear to be fewer apartments in cities where the population is older. The reasoning for this could be that old people live closer to old people, and older populations have already settled in and a high turnover rate is unlikely.

So older populations don't affect the rental market through apartment price, but they may affect the market through apartment availability.

4. Citation

D. Kahle and H. Wickham. ggmap: Spatial Visualization with ggplot2. The R Journal, 5(1), 144-161. URL http://journal.r-project.org/archive/2013-1/kahle-wickham.pdf

5. R code appendix

```
# Nathan J. Chan
# STA 141A
# Professor Ulle
# November 2, 2018
# Assignment 4

cl = readRDS("cl_apartments.rds")

library(ggplot2)
library(ggmap)
```

```
library(stringr)
library(sf)
library(ggpubr)
library(ggpmisc)

# remove two weird points and all NA,
# and remove weird sqft and super low prices
cl2 = subset(cl, cl$price != 9951095 &
               cl$price != 34083742 &
               cl$price > 100 &
               cl$sqft != 200000 &
               cl$sqft > 10
             )

# 1. Analyzing apartments in Davis on a map

davis = subset(cl2, cl2$city == "Davis" | cl2$place == "Davis")

davis_coords = c(-121.795024, 38.531215, -121.675951, 38.578497)
davis_map = get_stamenmap(davis_coords, zoom = 12, maptype = "toner")

# A. In which areas of Davis are there apartment sizes of different types?

# change bedrooms to categorical variable
# https://stackoverflow.com/questions/16639484/how-to-convert-integer-into-
categorical-data-in-r
davis$bedrooms = as.factor(davis$bedrooms)

# On map by number of bedrooms
ggmap(davis_map) +
  geom_point(aes(longitude, latitude, color = bedrooms), davis) +
  labs(x = "Longitude", y = "Latitude", title = "Apartments in Davis with certain
number of bedrooms") +
  guides(size = guide_legend(title = "Number of bedrooms"), color =
guide_legend(title = "Number of bedrooms")) +
  theme_minimal() +
  theme(plot.title = element_text(size = 12, face = "bold"), text =
element_text(family = "Helvetica"))

# On map by square-footage
ggmap(davis_map) +
  geom_point(aes(longitude, latitude, color = sqft, size = sqft), davis) +
  labs(x = "Longitude", y = "Latitude", title = "Apartments in Davis with certain
square-footage") +
  guides(size = guide_legend(title = "Square-footage"), color = guide_legend(title =
"Square-footage")) +
  theme_minimal() +
  theme(plot.title = element_text(size = 12, face = "bold"), text =
element_text(family = "Helvetica"))

# Note: there are some duplicates in the data set.
# I will ignore them, as they do not appear on the map
# (they are stacked on top of each other and appear as one point)

# B. Are apartments closer to UC Davis more expensive than apartments farther away?
```

```
# On map by price
ggmap(davis_map) +
  geom_point(aes(longitude, latitude, color = price, size = price), davis) +
  labs(x = "Longitude", y = "Latitude", title = "Apartments in Davis with certain
prices") +
  guides(size = guide_legend(title = "Price"), color = guide_legend(title = "Price"))
+
  theme_minimal() +
  theme(plot.title = element_text(size = 12, face = "bold"), text =
element_text(family = "Helvetica"))

# On map by price per bedroom
ggmap(davis_map) +
  geom_point(aes(longitude, latitude, color = price/bedrooms, size = price/bedrooms),
davis) +
  labs(x = "Longitude", y = "Latitude", title = "Apartments in Davis with certain
price per bedroom") +
  guides(size = guide_legend(title = "Price per bedroom"), color = guide_legend(title
= "Price per bedroom")) +
  theme_minimal() +
  theme(plot.title = element_text(size = 12, face = "bold"), text =
element_text(family = "Helvetica"))

# On map by price per square-foot
ggmap(davis_map) +
  geom_point(aes(longitude, latitude, color = price/sqft, size = price/sqft), davis)
+
  labs(x = "Longitude", y = "Latitude", title = "Apartments in Davis with certain
price per square-foot") +
  guides(size = guide_legend(title = "Price per square-foot"), color =
guide_legend(title = "Price per square-foot")) +
  theme_minimal() +
  theme(plot.title = element_text(size = 12, face = "bold"), text =
element_text(family = "Helvetica"))

# C. Is there a connection between where an apartment is located and the type of
parking available?

# On map by type of parking
ggmap(davis_map) +
  geom_point(aes(longitude, latitude, color = parking), davis) +
  labs(x = "Longitude", y = "Latitude", title = "Apartments in Davis with certain
types of parking") +
  guides(color = guide_legend(title = "Type of parking")) +
  theme_minimal() +
  theme(plot.title = element_text(size = 12, face = "bold"), text =
element_text(family = "Helvetica"))

# 2. Analyzing apartments in the Bay Area on a map

bay_area = subset(cl2, cl2$county == "San Francisco" |
                    cl2$county == "San Mateo" |
                    cl2$county == "Santa Clara" |
                    cl2$county == "Alameda" |
```

```
                  cl2$county == "Contra Costa"
                  )

# bay_area = subset(bay_area, is.na(bay_area$bedrooms) == FALSE)

bay_area_coords = c(-122.604945, 36.957870, -121.551308, 38.105303)
bay_area_map = get_stamenmap(bay_area_coords, zoom = 10, maptype = "toner")

# A. In which areas of the Bay Area are there apartment sizes of different types?

bay_area2 = bay_area
bay_area2$bedrooms = as.factor(bay_area2$bedrooms)

# On map by number of bedrooms
ggmap(bay_area_map) +
  geom_point(aes(longitude, latitude, color = bedrooms),
bay_area2[is.na(bay_area2$bedrooms) == FALSE, ]) +
  labs(x = "Longitude", y = "Latitude", title = "Apartments in the Bay Area with
certain number of bedrooms") +
  guides(size = guide_legend(title = "Number of bedrooms"), color =
guide_legend(title = "Number of bedrooms")) +
  theme_minimal() +
  theme(plot.title = element_text(size = 12, face = "bold"), text =
element_text(family = "Helvetica"))

# On map by square-footage
ggmap(bay_area_map) +
  geom_point(aes(longitude, latitude, color = sqft), bay_area2[is.na(bay_area2$sqft)
== FALSE, ]) +
  labs(x = "Longitude", y = "Latitude", title = "Apartments in Bay Area with certain
square-footage") +
  guides(size = guide_legend(title = "Square-footage"), color = guide_legend(title =
"Square-footage")) +
  theme_minimal() +
  theme(plot.title = element_text(size = 12, face = "bold"), text =
element_text(family = "Helvetica"))
# (will not include in report because it doesn't show anything useful)

# B. Are apartments close to popular areas more expensive than apartments farther
away?

# On map by price
ggmap(bay_area_map) +
  geom_point(aes(longitude, latitude, color = price), bay_area2) +
  labs(x = "Longitude", y = "Latitude", title = "Apartments in the Bay Area with
certain prices") +
  guides(size = guide_legend(title = "Price"), color = guide_legend(title = "Price"))
+
  theme_minimal() +
  theme(plot.title = element_text(size = 12, face = "bold"), text =
element_text(family = "Helvetica"))
# (will not include; not very helpful)

# On map by price per bedroom
ggmap(bay_area_map) +
```

```
  geom_point(aes(longitude, latitude, color = price/bedrooms),
bay_area[bay_area$bedrooms > 0, ]) +
  labs(x = "Longitude", y = "Latitude", title = "Apartments in Davis with certain
price per bedroom") +
  guides(size = guide_legend(title = "Price per bedroom"), color = guide_legend(title
= "Price per bedroom")) +
  theme_minimal() +
  theme(plot.title = element_text(size = 12, face = "bold"), text =
element_text(family = "Helvetica"))
# (not including studio apartments)

# On map by price per square-foot
ggmap(bay_area_map) +
  geom_point(aes(longitude, latitude, color = price/sqft), bay_area) +
  labs(x = "Longitude", y = "Latitude", title = "Apartments in Davis with certain
price per square-foot") +
  guides(size = guide_legend(title = "Price per square-foot"), color =
guide_legend(title = "Price per square-foot")) +
  theme_minimal() +
  theme(plot.title = element_text(size = 12, face = "bold"), text =
element_text(family = "Helvetica"))
# (will not include; not very helpful)

# C. Is there a connection between where an apartment is located and the type of
parking available?

# On map by type of parking
ggmap(bay_area_map) +
  geom_point(aes(longitude, latitude, color = parking), bay_area) +
  labs(x = "Longitude", y = "Latitude", title = "Apartments in Davis with certain
types of parking") +
  guides(color = guide_legend(title = "Type of parking")) +
  theme_minimal() +
  theme(plot.title = element_text(size = 12, face = "bold"), text =
element_text(family = "Helvetica"))

# 3. Oldest populations in the southern Bay Area and rental market effects

# read census data
# change column names
# https://stackoverflow.com/questions/20956119/assign-headers-based-on-existing-row-
in-dataframe-in-r
census = read.csv("2010_census_data/DEC_10_SF1_SF1DP1_with_ann.csv")
colnames(census) = as.character(unlist(census[1, ]))
census = census[-1, ]

# get only southern Bay Area
south_bay_area = subset(bay_area, bay_area$place == "Fremont" |
                          bay_area$place == "Redwood City" |
                          bay_area$place == "Mountain View" |
                          bay_area$place == "San Mateo" |
                          bay_area$place == "Menlo Park" |
                          bay_area$place == "Palo Alto" |
                          bay_area$place == "San Carlos" |
                          bay_area$place == "Atherton" |
```

```
                            bay_area$place == "Protola Valley" |
                            bay_area$place == "Foster City" |
                            bay_area$place == "Belmont" |
                            bay_area$place == "Los Altos Hills" |
                            bay_area$place == "Half Moon Bay" |
                            bay_area$place == "Los Altos" |
                            bay_area$place == "Woodside" |
                            bay_area$place == "San Jose" |
                            bay_area$place == "Campbell" |
                            bay_area$place == "Gilroy" |
                            bay_area$place == "Los Gatos" |
                            bay_area$place == "Santa Clara" |
                            bay_area$place == "Sunnyvale" |
                            bay_area$place == "Milpitas" |
                            bay_area$place == "Cuptertino" |
                            bay_area$place == "Morgan Hill" |
                            bay_area$place == "Saratoga"
                            )

# get data only related to southern Bay Area
south_census = subset(census, str_detect(census$Geography, "Fremont") |
                        str_detect(census$Geography, "Redwood City") |
                        str_detect(census$Geography, "Mountain View") |
                        str_detect(census$Geography, "San Mateo") |
                        str_detect(census$Geography, "Menlo Park") |
                        str_detect(census$Geography, "Palo Alto") |
                        str_detect(census$Geography, "San Carlos") |
                        str_detect(census$Geography, "Atherton") |
                        str_detect(census$Geography, "Protola Valley") |
                        str_detect(census$Geography, "Foster City") |
                        str_detect(census$Geography, "Belmont") |
                        str_detect(census$Geography, "Los Altos Hills") |
                        str_detect(census$Geography, "Half Moon Bay") |
                        str_detect(census$Geography, "Los Altos") |
                        str_detect(census$Geography, "Woodside") |
                        str_detect(census$Geography, "San Jose") |
                        str_detect(census$Geography, "Campbell") |
                        str_detect(census$Geography, "Gilroy") |
                        str_detect(census$Geography, "Los Gatos") |
                        str_detect(census$Geography, "Santa Clara") |
                        str_detect(census$Geography, "Sunnyvale") |
                        str_detect(census$Geography, "Milpitas") |
                        str_detect(census$Geography, "Cuptertino") |
                        str_detect(census$Geography, "Morgan Hill") |
                        str_detect(census$Geography, "Saratoga")
                      )

# ignore this function; doesn't work
# for some reason cannot compare the strings properly
merge_data = function(south_bay_area, south_census) {
  # Create a vector with the median ages in the same order as the south_bay_area
  # Will add median age vector to south_bay_area
  median_age = c()
  for (i in 1:nrow(south_bay_area)) {
    for (j in 1:nrow(south_census)) {
```

```
      if (str_detect(as.character(south_census[j, ]$Geography),
as.character(south_bay_area[i, ]$place)) == TRUE) {
        median_age[i] = as.character(south_census[j,
]$Percent..SEX.AND.AGE...Total.population...Median.age..years.)
        break
      }
    }
  }
  return(median_age)
}

write.csv(south_census, "south_census.csv")
south_census = read.csv("south_census.csv")

remove_excess = function(data) {
  # remove parts of the string that shouldn't be there
  # allow us to merge
  for (i in 1:nrow(data)) {
    if (str_detect(data[i, 3], "city")) {
      data[i, 3] = str_remove(data[i, 3], " city, California")
    }
    else if (str_detect(data[i, 3], "town")) {
      data[i, 3] = str_remove(data[i, 3], " town, California")
    }
  }
  return(data)
}

south_census = remove_excess(south_census)

combined_bay_area = merge(south_bay_area, south_census, by.x = "city", by.y =
"Geography")

# Plot median age against price of apartment
ggplot(combined_bay_area, aes(x =
Number..SEX.AND.AGE...Total.population...Median.age..years., y = price)) +
  geom_point() +
  labs(x = "Median age of population in city where apartment is located", y = "Price
of apartment",
       title = "Median age of the city compared to prices of apartments in the city")
+
  geom_smooth(method = 'loess') +
  theme_minimal() +
  theme(plot.title = element_text(size = 12, face = "bold"), text =
element_text(family = "Helvetica")) +
  stat_cor(method = "pearson")
```