ASSIGNMENT 3

Nathan J. Chan
October 18, 2018

UC Davis, STA 141A
Professor Nick Ulle

1. Introduction to the Data Set

We will investigate a data set comprising of 21,948 Craigslist apartment listings from all around California. On October 15. 2018, the information in the data set was scraped from the Craigslist website that hosts all these listings. The posting dates of the listings range from September 8, 2018 to October 15, 2018. Each data point includes information of the listing, including the title, description, and date of the listing, and the price, location, number of bedrooms and bathrooms, square footage, and other features of the apartment. There is some data missing: mainly from the date updated feature, likely because not everyone updated their listing, and the square footage feature, likely because not everyone knows their square footage of their apartment.

2. Family-friendliness, Bedrooms and Bathrooms, and Comparison

A. Family-friendliness of Major and Non-major City Apartments

The family-friendliness of an apartment can be defined many ways; I will define it as the following: the apartment has 2 or more bedrooms, 1.5 or more bathrooms, allows pets, both cats and dogs, has some sort of washing machine available, and has some sort of parking available.

We will look at the number of family-friendly apartments in the city and the suburbs. I will consider the major cities to be cities with more than 900 apartment listings, which are Los Angeles, Sacramento, San Diego, San Francisco, and San Jose. Here are the results:

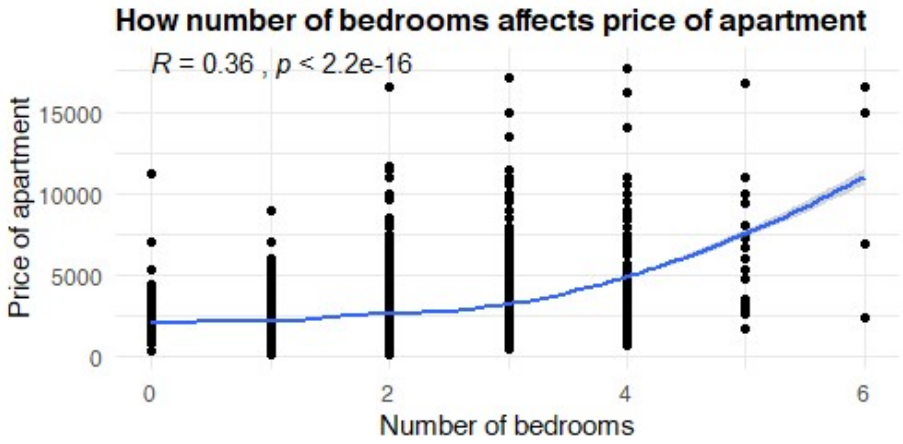
Percentage of family-friendly apartments in the major cities	Percentage of family-friendly apartments in non-major cities
1373 / 8254 = 16.63%	2569 / 11838 = 21.70%

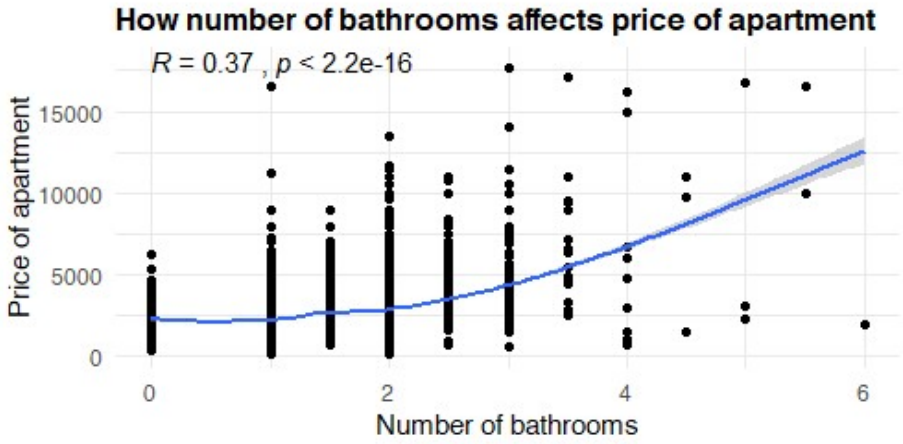
There is a much higher percentage of family-friendly apartments in non-major cities than there are in major cities. Assuming all the assumptions are met, doing a 2-proportion z-test, we find that the difference in percentages are statistically significant: there is a ~0% that these percentages are a product of change. Therefore, we can conclude that apartments in non-major cities are more likely to be family-friendly than apartments in major cities.

This may be due to the way apartments are designed in large cities compared to in non-major cities. In large major cities, apartment buildings are tall skyscrapers designed to pack in as many apartments as possible, while in small cities, apartment buildings may be a couple stories tall with more spacious apartments. Apartments in large cities are not going to focus on space, because real estate in the city is more valuable and scarce, so there will be fewer bigger, family-friendly apartments.

B. Number of Bedrooms and Bathrooms, and Effect on Rent

We will look at whether more bedrooms or more bathrooms has a heavier impact on rent. From the data set, I removed two outrageous, and likely mistaken, prices ($9,951,095 and $34,083,742) and any prices lower than $100 to avoid skewing the data. Those two weird prices could have been a result of a typo. The lower prices are from listing that weren't even about apartments, or from people posting $0 or $1 to hint that their price was negotiable. Since it is impossible to get an actual price for these apartments whose prices are negotiable, there are two options: I can assign it a price based on apartments from similar features (finding the average of all prices with same bedroom and bathroom count and giving that average to the data point) or I can remove it entirely. I chose to remove the listing entirely.





As evidenced in the two graphs, the slope of the line of the second graph is steeper, indicating that an increase in number of bathrooms will lead to a larger increase in price than an increase in number of bedrooms. Why is this the case? I can only infer that a high number of bathrooms is rarer than a high number of bedrooms, so a higher number of bathrooms in an apartment would carry with it a higher premium.

Note that there are a few points that seem out of place, like apartments with a high number of bedrooms or bathrooms at a somewhat low price. This is due to the wide range of prices of apartments in California. Prices depend heavily on location, and an apartment in the Bay Area will cost many times more than a similar apartment in the Central Valley.

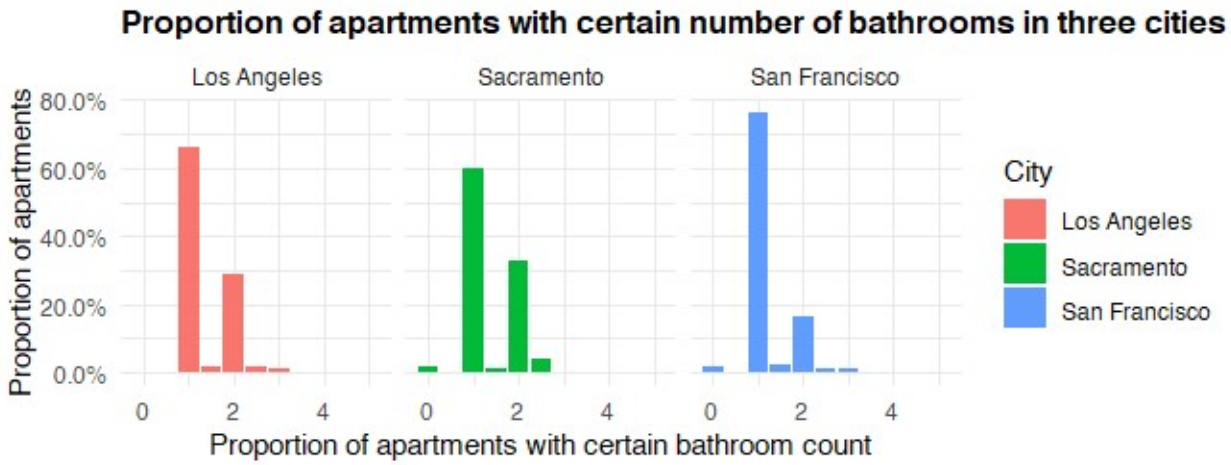
C. Similarities and Differences of Apartments in Three Cities

We will look at apartments in the three different cities, San Francisco, Los Angeles, and Sacramento, and determine how similar the apartments are in these cities. Taking a look at the figures, there are both similarities and differences between the three cities:

Figure C1:

Note: 0 bedrooms indicates a studio apartment.

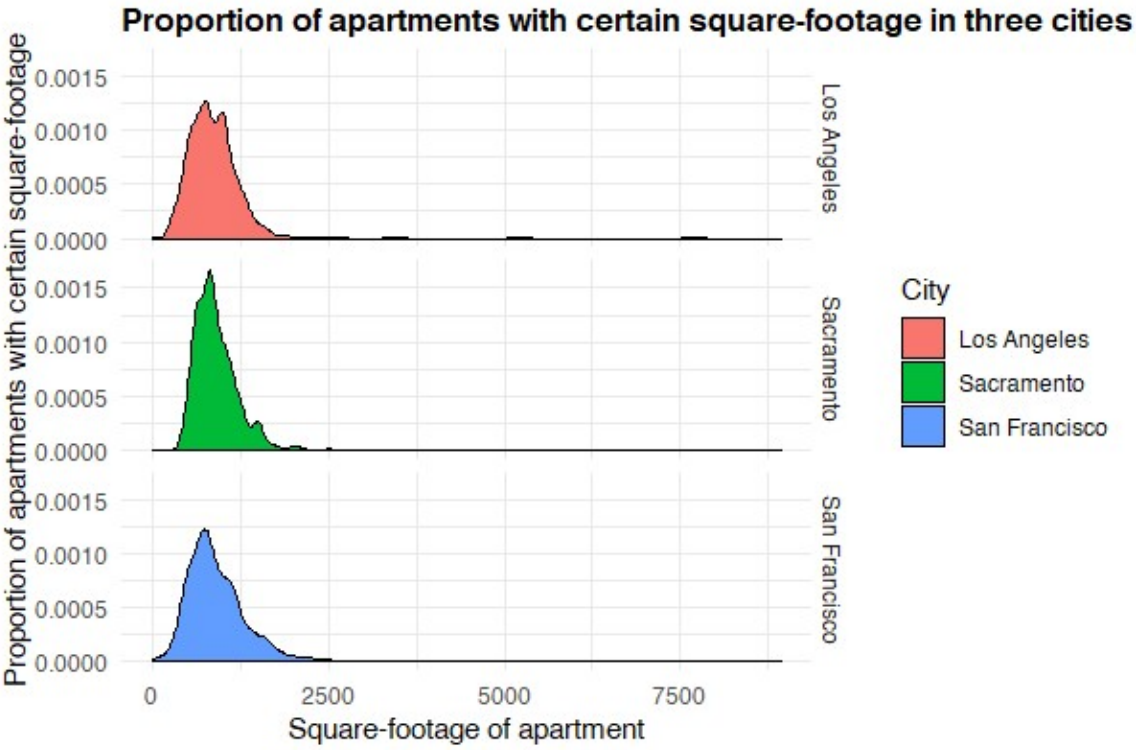
Figure C2:



Note: 0.5 bathrooms means a toilet only, while 1 bathroom means a toilet and shower.

Figure C3:

Figure C4:



In *Figure C1*, we see the proportion of apartments with a certain number of bedrooms in the three cities. Sacramento has about the same number of 1 and 2 bedroom apartments (with slightly more 2 bedroom apartments), with very few studio and 3 bedroom apartments. Los Angeles and San Francisco are more similar in terms of their distributions: there are more 1 bedroom than 2 bedroom apartments, and there are more 2 bedroom than studio apartments. San Francisco has a higher percentage of studio and 3 bedroom apartments, while Los Angeles has a higher percentage of 1 and 2 bedroom apartments.

In *Figure C2*, there overall trend is similar across the cities: there are more apartments with 1 bathroom than 2 bathrooms. However, the vast majority of the apartments in San Francisco have 1 bathroom, while there is a good percentage of apartments in Los Angeles and Sacramento that have 2 bathrooms. Sacramento has a few percentages with 3 bathrooms: more than Los Angeles and San Francisco.

In *Figure C3*, we see that the distribution of prices across the three cities are very different. The prices of apartments in Sacramento are about the same within the city. In Los Angeles, the distribution starts to spread out wider, with more apartments at a higher price, and in San Francisco, most apartments are more expensive than in both Sacramento and Los Angeles, and there is a larger range of more expensive prices.

In *Figure C4*, the square-footage of the apartments are about the same. Los Angeles has a strange spike in higher square-footage apartments, and Sacramento also has a small spike in very high square-footage apartments. San Francisco has a wider range of apartments.

Overall, the differences between the three cities are pretty consistent. San Francisco has a wider range of apartments, in terms of price and square-footage. Apartments in San Francisco are generally smaller, in terms of number of bedrooms and bathrooms. This result may be due to the technology boom in San Francisco, where new apartment buildings are being constructed and young people, who don't have families, rent smaller apartments and work in tech. Los Angeles is very similar. An apartment renter may look at Sacramento for the best prices and consistent offerings of 1 or 2 bedroom apartments. This outcome may be from Sacramento's milder economy and offset location, far from Silicon Valley.

3. Questions from the Data Set

1) *How do apartments with or without certain amenities, like washing machines and parking, differ in price?*

This question may be useful to apartment shoppers looking at how much of a premium a they might pay for special features of an apartment. Renters can weigh the extra cost of having these features with the convenience of having them, comparing that to the cost of traveling and paying for laundry or parking somewhere else.

2) *What is the relationship between square footage of an apartment and price? How might the relationships differ between different cities?*

Square footage might not even matter when searching for an apartment. As we know, an apartment in a prime location, however small, may cost more than a giant apartment somewhere else. This information may be useful to both apartment shoppers and owners wanting to know how much square footage matters when comparing prices to apartments.

3) *Will apartments with laundry also have parking? And vice versa?*

It would be interesting to see if apartments that have certain amenities are also likely to have other amenities. This is a broad, general question that might be helpful for people that own and build apartments.

4) *Which types of apartments allow pets? What kinds of pets?*

This information would be most useful to pet owners who want to rent an apartment, giving them an estimate about what kinds of apartments they can expect will and will not allow pets; for example, different cities or different apartment sizes may have different pet policies.

5) *What are the proportions of apartments with different number of bedrooms? How do the proportions differ by city?*

Apartment hunters may be interested in what their choices are when looking at apartments, so this question would give more information on the apartment market and what percentage of apartments they can expect will have the number of bedrooms they require.

6) *Do similar apartments with different pet policies, laundry arrangements, or parking designations have different prices?*

The question is asking whether different features among similar apartments cost different. For example, will a 2 bedroom and 1 bathroom apartment with covered parking cost different from a 2 bedroom and 1 bathroom apartment with garage parking? Similar to Question 1, if there is a difference, this information will help renters gauge the extra cost of different apartment amenities.

7) *Is there a difference in apartments whose listings are posted on certain days of the week?*

To clarify, this question asks about how the apartment listings differ depending on the day of the week it was posted. This question isn't necessarily helpful to anyone, but finding out the answer would still be interesting. Maybe certain people own certain apartments, and these certain people create their posts on different days of the week.

8) *How clustered or spread out are the apartment listings?*

Using the latitude and longitude information, we could see how apartment listings are bunched together. Maybe different cities have apartments more spread out than others. This information can be helpful to social scientists studying the trend of the closeness of apartments, or to a renter or owner wondering about the closeness of other apartments.

9) *How much different are the prices of apartments in San Francisco or Los Angeles?*

It's common knowledge that housing in California is very expensive, and the cost of apartments may differ in different parts of California. Someone who doesn't know much about California may worry about apartment prices in different cities. San Francisco and Los Angeles are both population and industry centers, with San Francisco focusing on technology and Los Angeles focusing on entertainment. This question is helpful to someone deciding whether to rent an apartment in San Francisco or Los Angeles.

10) For a certain number of bedrooms, how many bathrooms are expected?

There are many options for apartments. This information will be helpful to an apartment renter interested in the expected options for apartment configurations. There may be interesting trends in the number combinations of bedrooms and bathrooms, and some may be more popular than others.

4. Answering Five Questions

1) How do apartments with or without certain amenities, like washing machines and parking, differ in price?

Hypothesis: Apartments with amenities cost slightly more than apartments without amenities.

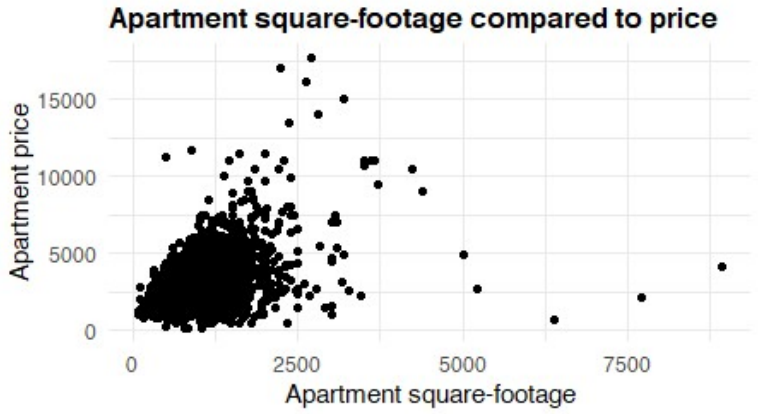
Comparing average price of apartments	Apartments with both laundry and parking	Apartments without either laundry or parking
Average price	$2464.886	$2293.006
Standard deviation	1131.438	1302.49
Total number of apartments	17824	650

I looked at two subsets of the data set: all the apartments that have some sort of laundry and parking, and the apartments that had neither laundry nor parking. I performed a 2-sample t-test, assuming all the assumptions were met, and got a p-value of 0.0004755, meaning there was about a 0.04755% probability the difference between them was a result of chance. Therefore, we have good evidence that the apartments with the amenities are on average more expensive than those without amenities.

2) What is the relationship between square footage of an apartment and price?

Hypothesis: As apartment square-footage increases, price increases. Different cities will have extreme or milder slopes, depending on the real estate market in a particular city.

First, I removed a few data points, where the square-footage was either way too small, indicating the lister probably didn't know the square-footage, and way too big, which was likely a typo. Here is the resulting scatter plot:



The relationship is not as clear as I hoped, although the positive relationship is fairly easy to see. The reason the scatter plot spreads out in a v shape is likely because of how different the apartment prices are in various parts of California. A small apartment in a high-demand area like San Francisco costs more than a giant apartment in the Central California.

3) Will apartments with laundry also have parking? And vice versa?

Hypothesis: Apartments with laundry are more likely to have parking than apartments without laundry. Apartments with parking are more likely to have laundry that apartments without it.

	Given the apartment has laundry…	Given the apartment does not have laundry…
Percentage of apartments with parking	14057 / 15340 = 91.636%	400 / 695 = 57.554%
Percentage of apartments without parking	1283 / 15340 = 8.364%	295 / 695 = 42.446%

	Given the apartment has parking…	Given the apartment does not have parking…
Percentage of apartments with laundry	14057 / 14457 = 97.233%	1283 / 1578 = 81.305%
Percentage of apartments without laundry	400 / 14457 = 2.767%	295 / 1578 = 18.695%

My hypothesis was correct: if an apartment has laundry, it is more likely to have parking. If the apartment has laundry, there is a 91.636% chance it has parking, while if it doesn't have laundry, there is a 57.554% chance it has laundry.

If an apartment has parking, it is slightly more likely to have parking. If the apartment has parking there is a 97.233% chance it has laundry, while if it doesn't have parking, there is a 81.305% chance it has parking.

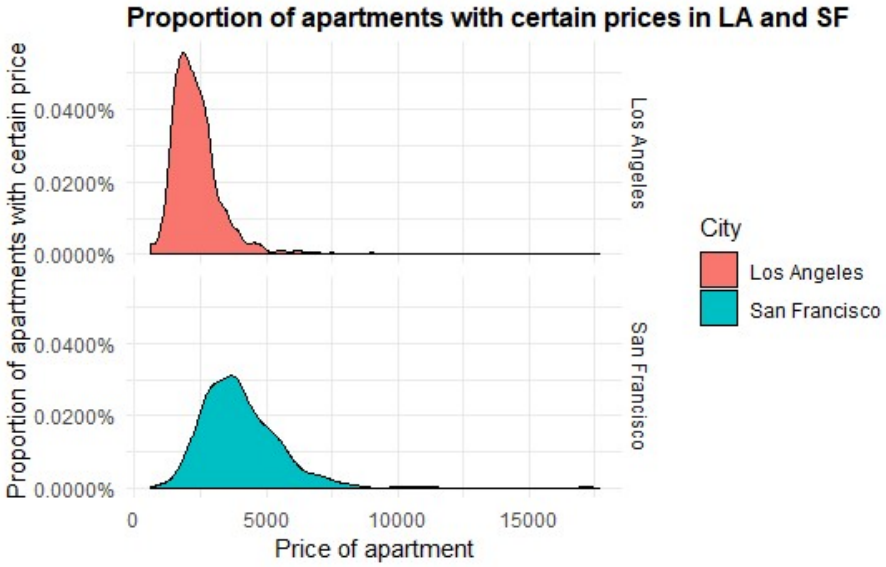
It seems that parking is a more exclusive feature than laundry. No laundry means high chance of no parking. No parking means there's probably still laundry. How does this makes sense? It makes sense because there are a very small number of apartments with laundry. If an apartment doesn't even bother to have laundry, then it won't bother to have parking. Parking is more exclusive. But, if an apartment does have laundry, there is a very high chance it also has parking. If the apartment has parking, it's very, very likely it also has laundry, since parking is already a special feature, and if it doesn't have parking, then it probably still had laundry, because parking is exclusive and not all apartments have it (but most apartments have laundry.)

4) How much different are the prices of apartments in San Francisco and Los Angeles?

Hypothesis: Apartments in Los Angeles and San Francisco will be about the same price.

As a sort of extension from *Figure C3*, I want to investigate further on the distribution of prices. Here are the distribution of prices:

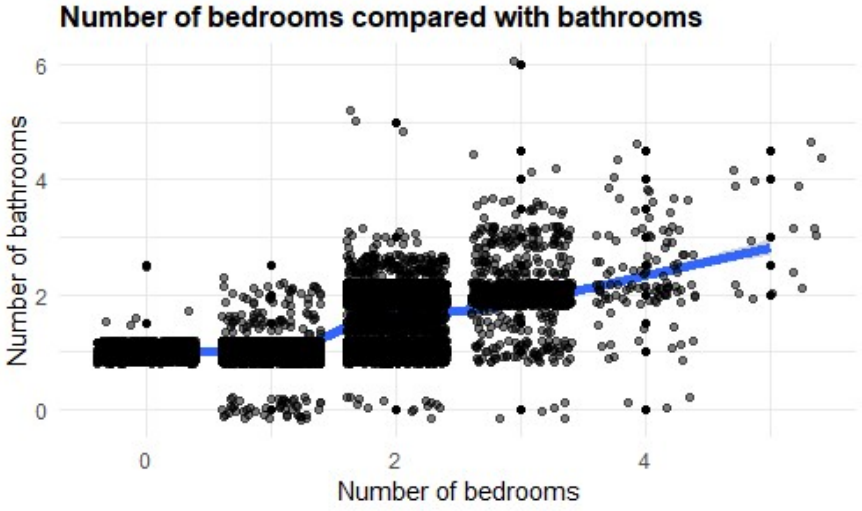
	Apartments in Los Angeles	Apartments in San Francisco
Average price	$2353.463	$4079.152
Standard deviation	1008.241	1686.741
Total number of apartments	1197	1157



Doing a 2-sample t-test, we find that the p-value is 0%, meaning there is no chance that the average prices are the same. This is evident in the two charts, where San Francisco clearly has higher prices than Los Angeles; San Francisco also has a wider range of prices. The housing markets in these two cities seem very different, despite them both being in California, and the difference between them is enormous. An apartment shopper deciding between the two cities based on price should definitely choose Los Angeles.

5) For a certain number of bedrooms, how many bathrooms are expected?

Hypothesis: The number of bathrooms will always be less than or equal to the number of bedrooms. For every 2 bedrooms, there will be about 1 to 2 bathrooms.



Note: I added some “jitter” to the points so all the scatter points won’t overlap with each other.

Studio and 1 bedroom apartments will nearly always have 1 bathroom, as would be expected. There are some apartments with no bathrooms, because the bathrooms are shared. Many of the 1 bedrooms apartments with no bathrooms are from multiple listings from a senior home, where the bathrooms are shared. 2 bedroom apartments have a lot of options in terms of the bathroom count: there are many apartments with 1, 1.5, 2, and 2.5 bathrooms.

As the number of bedrooms increases, the number of bathrooms increases. It is a milder increase; there will usually be fewer bathrooms than bedrooms in any given apartment, and, on average, an increase in 1 bedroom is approximately an increase in 0.5 bathrooms.

5. Limitations of the Data Set and Effect on Conclusions

The data set was generated by scraping the web postings from Craigslist, so whatever information was on the listing ended up in the dataset. Since this was an automated process, there is likely no bias in how the data set was created.

As identified throughout the analysis, there were several weird data points that were removed, including square-footage and prices that were too small or too big. In these cases, the lister either made a typo or purposely put an inaccurate price indicate that the price was negotiable. (It is common on Craigslist to list items as $1.) While it is possible to simply give a price to the apartment (the given price being the average price of similar apartments), I chose to remove them instead.

There is a lot of missing data, either because the program didn’t read all the information or the lister didn’t including the information. Whenever there was analysis that required data from the features with missing data, I removed the listings with missing data.

It is important to note that while my analysis makes claims about the entire apartment market in certain places, these apartments listings are only a small sample from all the apartments. Almost all apartments anywhere are occupied, so they are not listed, and many apartment owners won’t use Craigslist to rent off their apartment, as there are other tools like Zillow and Apartments.com. Some places are in such high demand that they don’t need to have a listing to attract attentions from renters.

The apartments listed on Craigslist are definitely not a wholly accurate representation of all apartments in California. For example, a high end apartment complex won’t waste their time posting something on a website like Craigslist (as I believe Craigslist seems more to be a place where normal people can publicize their listings, not large companies). Nonetheless, the data from Craigslist is a fine source to get a general idea of what apartments look like around California.

My conclusions, therefore, must be taken with the knowledge of the limitations. In my analysis, I assume that the sample is representative of both the renting market for apartments and for all apartments in general. However, this make not be the case, and a better sample may be required before any sort of confirmation of my claims.

The housing market in California is wild and treacherous, and we must be wary of making sweeping generalizations. Information about apartments is very important to both owners and renters, and economists, social scientists, and lawmakers, so we must tread lightly.

6. R Code Appendix

```
# STA 141A Assignment 3
# Nathan Chan

cl = readRDS("cl_apartments.rds")

# 1

find_max = function(data) {
  # Reads date, prints the latest date
  max = 0
  for (i in 1:nrow(data)) {
    temp = as.Date(data[i, 6])
    if (is.na(temp) | is.na(max)) {
      next
    }
    if (temp > max) {
      max = temp
      print(max)
    } else {
      next
    }
  }
}

find_min = function(data) {
  # Reads date, prints the earliest date
  min = as.Date("2018-10-15")
  for (i in 1:nrow(data)) {
    temp = as.Date(data[i, 6])
    if (is.na(temp) | is.na(min)) {
      next
    }
    if (temp < min) {
      min = temp
      print(min)
    } else {
      next
    }
  }
}

find_max(cl)
find_min(cl)

dates = sort(cl$date_posted)

# 2
```

```
unique(cl$laundry)
table(cl$laundry)
table(cl$parking)

ff = subset(cl, cl$bedrooms >= 2 & cl$bathrooms >= 1.5 & cl$pets == "both" &
cl$laundry != "none" &
              cl$parking != "none" & cl$city != "NA")

library("plyr")
cities_count = count(cl, "city")
cities_count
sort(cities_count$freq)

ff_cities = subset(ff, ff$city == "Los Angeles" | ff$city == "Sacramento" |
ff$city == "San Diego" |
                     ff$city == "San Francisco" | ff$city == "San Jose")

cities = subset(cl, (cl$city == "Los Angeles" | cl$city == "Sacramento" |
cl$city == "San Diego" |
                     cl$city == "San Francisco" | cl$city == "San Jose") &
cl$city != "NA")

ff_not_cities = subset(ff, ff$city != "Los Angeles" & ff$city != "Sacramento"
& ff$city != "San Diego" &
                         ff$city != "San Francisco" & ff$city != "San Jose")

not_cities = subset(cl, (cl$city != "Los Angeles" & cl$city != "Sacramento" &
cl$city != "San Diego" &
                       cl$city != "San Francisco" & cl$city != "San Jose") &
cl$city != "NA")

prop_city = nrow(ff_cities) / nrow(cities)
prop_not_cities = nrow(ff_not_cities) / nrow(not_cities)

nrow(ff_cities)
nrow(cities)
nrow(ff_not_cities)
nrow(not_cities)

library(ggplot2)
library(ggpubr)
library(ggpmisc)

# http://www.sthda.com/english/articles/32-r-graphics-essentials/131-plot-
two-continuous-variables-scatter-graph-and-alternatives/

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

```
g1 = ggplot(cl2, aes(x = bedrooms, y = price)) + geom_point() +
geom_smooth(method = 'loess') +
  labs(x = "Number of bedrooms", y = "Price of apartment",
       title = "How number of bedrooms affects price of apartment") +
theme_minimal() +
  theme(plot.title = element_text(size = 12, face = "bold"), text =
element_text(family = "Helvetica")) +
  stat_cor(method = "pearson", label.y = 18000)

g2 = ggplot(cl2, aes(x = bathrooms, y = price)) + geom_point() +
geom_smooth(method = 'loess') +
  labs(x = "Number of bathrooms", y = "Price of apartment",
       title = "How number of bathrooms affects price of apartment") +
theme_minimal() +
  theme(plot.title = element_text(size = 12, face = "bold"), text =
element_text(family = "Helvetica")) +
  stat_cor(method = "pearson", label.y = 18000)

library(gridExtra)
grid.arrange(g1, g2)

test = unique(cl$title)

# 3

SF = subset(cl, cl$city == "San Francisco")
LA = subset(cl, cl$city == "Los Angeles")
SM = subset(cl, cl$city == "Sacramento")

three_cities = subset(cl, (cl$city == "Los Angeles" | cl$city == "Sacramento"
| cl$city == "San Francisco")
                      & cl$city != "NA")
three_cities = subset(three_cities, three_cities$sqft > 10)

library("ggridges")
library("scales")

# https://stackoverflow.com/questions/36604127/creating-a-bar-plot-with-
proportions-on-ggplot

ggplot(three_cities, aes(bedrooms, y = ..prop.., fill = city)) + geom_bar() +
facet_wrap(~ city) +
  labs(x = "Proportion of apartments with certain bedroom count", y =
"Proportion of apartments",
       title = "Proportion of apartments with certain number of bedrooms in
three cities") +
  guides(fill = guide_legend(title = "City")) + theme_minimal() +
  theme(plot.title = element_text(size = 12, face = "bold"), text =
element_text(family = "Helvetica")) +
```

```
  scale_y_continuous(labels = percent_format())

ggplot(three_cities, aes(bathrooms, y = ..prop.., fill = city)) + geom_bar()
+ facet_wrap(~ city) +
  labs(x = "Proportion of apartments with certain bathroom count", y =
"Proportion of apartments",
       title = "Proportion of apartments with certain number of bathrooms in
three cities") +
  guides(fill = guide_legend(title = "City")) + theme_minimal() +
  theme(plot.title = element_text(size = 12, face = "bold"), text =
element_text(family = "Helvetica")) +
  scale_y_continuous(labels = percent_format())

ggplot(three_cities, aes(price, fill = city)) + geom_density() +
facet_grid(city ~ .) +
  labs(x = "Price of apartment", y = "Proportion of apartments with certain
price",
       title = "Proportion of apartments with certain prices in three
cities") +
  guides(fill = guide_legend(title = "City")) + theme_minimal() +
  theme(plot.title = element_text(size = 12, face = "bold"), text =
element_text(family = "Helvetica")) +
  scale_y_continuous(labels = percent_format())

ggplot(three_cities, aes(sqft, fill = city)) + geom_density() +
facet_grid(city ~ .) +
  labs(x = "Square-footage of apartment", y = "Proportion of apartments with
certain square-footage",
       title = "Proportion of apartments with certain square-footage in three
cities") +
  guides(fill = guide_legend(title = "City")) + theme_minimal() +
  theme(plot.title = element_text(size = 12, face = "bold"), text =
element_text(family = "Helvetica"))

# Changing formating:
# https://stackoverflow.com/questions/28243514/ggplot2-change-title-size
# https://stackoverflow.com/questions/34522732/changing-fonts-in-ggplot2
# http://www.sthda.com/english/wiki/ggplot2-title-main-axis-and-legend-titles

# 4

# 1)

with_amenities = subset(cl2, cl2$laundry != "none" & cl2$parking != "none")
without_amenities = subset(cl2, cl2$laundry == "none" & cl2$parking ==
"none")

mean(with_amenities$price, na.rm = TRUE)
mean(without_amenities$price, na.rm = TRUE)
```

```
sd(with_amenities$price, na.rm = TRUE)
sd(without_amenities$price, na.rm = TRUE)

nrow(with_amenities)
nrow(without_amenities)

# 2)

ggplot(cl2, aes(x = sqft, y = price)) + geom_point() +
  labs(x = "Apartment square-footage", y = "Apartment price", title =
"Apartment square-footage compared to price") +
  theme_minimal() + theme(plot.title = element_text(size = 12, face =
"bold"), text = element_text(family = "Helvetica"))

ggplot(three_cities, aes(x = sqft, y = price)) + geom_point() +
facet_grid(city ~ .) +
  labs(x = "Apartment square-footage", y = "Apartment price",
       title = "Apartment square-footage compared to price") +
  theme_minimal() + geom_smooth(method = 'lm') +
  theme(plot.title = element_text(size = 12, face = "bold"), text =
element_text(family = "Helvetica"))

# 3)

cl3 = subset(cl2, is.na(cl2$laundry) != TRUE & is.na(cl2$parking) != TRUE)

yes_laundry = subset(cl3, cl3$laundry != "none")

no_laundry = subset(cl3, cl3$laundry == "none")

yes_parking = subset(cl3, cl3$parking != "none")

no_parking = subset(cl3, cl3$parking == "none")

nrow(subset(yes_laundry, yes_laundry$parking != "none"))
nrow(subset(yes_laundry, yes_laundry$parking == "none"))

nrow(yes_laundry)

nrow(subset(no_laundry, no_laundry$parking != "none"))
nrow(subset(no_laundry, no_laundry$parking == "none"))

nrow(no_laundry)

nrow(subset(yes_parking, yes_parking$laundry != "none"))
nrow(subset(yes_parking, yes_parking$laundry == "none"))

nrow(yes_parking)

nrow(subset(no_parking, no_parking$laundry != "none"))
```

```
nrow(subset(no_parking, no_parking$laundry == "none"))

nrow(no_parking)

# 4)

lasf = subset(three_cities, three_cities$city != "Sacramento")

ggplot(lasf, aes(price, fill = city)) + geom_density() + facet_grid(city ~ .)
+
  labs(x = "Price of apartment", y = "Proportion of apartments with certain
price",
       title = "Proportion of apartments with certain prices in LA and SF") +
  guides(fill = guide_legend(title = "City")) + theme_minimal() +
  theme(plot.title = element_text(size = 12, face = "bold"), text =
element_text(family = "Helvetica")) +
  scale_y_continuous(labels = percent_format())

mean(lasf$price[lasf$city == "Los Angeles"], na.rm = TRUE)
mean(lasf$price[lasf$city == "San Francisco"], na.rm = TRUE)

sd(lasf$price[lasf$city == "Los Angeles"], na.rm = TRUE)
sd(lasf$price[lasf$city == "San Francisco"], na.rm = TRUE)

lasf = subset(lasf, is.na(lasf$price) == FALSE)
nrow(subset(lasf, lasf$city == "Los Angeles"))
nrow(subset(lasf, lasf$city == "San Francisco"))

# 5)

bed_bath = subset(cl2, is.na(cl2$bedrooms) == FALSE & is.na(cl2$bathrooms) ==
FALSE)

# https://stackoverflow.com/questions/46845342/plot-data-with-duplicate-
points

ggplot(bed_bath, aes(x = bedrooms, y = bathrooms)) + geom_point() +
geom_smooth(method = "loess", size = 2) +
  theme_minimal() + geom_jitter(alpha = 0.5) +
  theme(plot.title = element_text(size = 12, face = "bold"), text =
element_text(family = "Helvetica")) +
  labs(x = "Number of bedrooms", y = "Number of bathrooms", title = "Number
of bedrooms compared with bathrooms")
```