STAT 33B Lab 2

Ziang Gao (3033669851)

Edit this file, knit to PDF, and:

- Submit the Rmd file on bCourses.
- Submit the PDF file on Gradescope.

If you think you'll need help with submission, please ask during the lab.

Answer all questions with complete sentences, and put code in code chunks. You can make as many new code chunks as you like. Please do not delete the exercises already in this notebook, because it may interfere with our grading tools.

As you work, you may find it helpful to be able to run your code. You can run a single line of code by pressing Ctrl + Enter. You can run an entire code chunk by clicking on the green arrow in the upper right corner of the code chunk.

Knit the document from time to time to make sure that your code runs without errors from top to bottom in a fresh R environment.

The code below controls the number of significant digits shown for the return values in your knitted document.

```
options(digits = 3)
```

Bay Area Apartments Data Set

The Bay Area Apartments Data Set is a collection of 5852 advertisements for apartments for rent in the San Francisco Bay Area. The data set was collected from Craigslist on Feb 13, 2020.

In this lab, you'll use subsetting, statistical functions, and `ggplot2` to analyze the data set.

Exercise 1

The data set is available on the bCourse as an RDS file.

Read the data set into R, then use R functions to display the following:

- The dimensions of the data set.
- The names of the columns in the data set.
- A **str**uctural summary of the data set.

```
# Your code goes here
file = readRDS("cl_apartments.rds")
dim = dim(file)
cnames = colnames(file)
summary = str(file)
```

```
## 'data.frame':    5852 obs. of  21 variables:
##  $ title       : chr  "$1695 / 1br - 700ft2 - 1 Bed 1 Bath Apartment Downtown Livermore $1695 (dubli
##  $ text        : chr  "QR Code Link to This Post\n            \n        \n1 bed 1 bath apartment for
##  $ latitude    : num  NA 37.9 37.6 37.9 37.9 ...
```

```
##  $ longitude   : num  NA -122 -122 -122 -122 ...
##  $ city_text   : Factor w/ 277 levels "  Limited Time Only ) (vallejo / benicia",..: 64 51 73 229 37
##  $ date_posted : POSIXlt, format: "2020-01-31 13:56:37" "2020-02-13 10:35:39" ...
##  $ date_updated: POSIXlt, format: "2020-02-13 09:58:45" NA ...
##  $ price       : num  1695 2120 2463 1875 2900 ...
##  $ deleted     : logi  FALSE FALSE FALSE FALSE FALSE ...
##  $ sqft        : num  700 590 777 650 600 ...
##  $ bedrooms    : num  1 1 1 1 2 2 2 2 4 1 ...
##  $ bathrooms   : num  1 1 1 1 1 2 2 2 2.5 1 ...
##  $ pets        : Factor w/ 5 levels "both","cats",..: NA 1 1 1 5 1 1 1 1 1 ...
##  $ laundry     : Factor w/ 4 levels "hookup","in-unit",..: 4 2 2 4 3 4 4 4 2 4 ...
##  $ parking     : Factor w/ 6 levels "covered","garage",..: 1 1 1 1 3 2 4 2 6 2 ...
##  $ fname       : chr  "data//sfbay/_eby_apa_d_1-bed-1-bath-apartment-downtown_7066767007.html" "data/
##  $ craigslist  : Factor w/ 2 levels "sfbay","sfbay_eby": 1 1 1 1 1 1 1 1 1 1 ...
##  $ place       : Factor w/ 129 levels "Alameda","Alamo",..: NA 88 124 128 11 1 1 1 1 1 ...
##  $ city        : Factor w/ 95 levels "Alameda","Albany",..: NA 62 91 94 7 1 1 1 1 1 ...
##  $ state       : chr  NA "CA" "CA" "CA" ...
##  $ county      : Factor w/ 18 levels "Alameda","Clark",..: NA 3 1 3 1 1 1 1 1 1 ...
```

Exercise 2

The city each apartment is in is listed in the `city` column. Note that some entries in the `city` column are missing.

Get the subset of apartments in Berkeley. Then answer the following:

1. How many advertisements are there for Berkeley apartments?
2. What is the mean price for Berkeley apartments?
3. What is the median price for Berkeley apartments?

```r
# Your code goes here
berkeley = file[which(file$city == "Berkeley"), ]
num_ads = nrow(berkeley)
avg_price = mean(berkeley$price, na.rm = TRUE)
med_price = median(berkeley$price, na.rm = TRUE)
```

WRITE YOUR ANSWERS BELOW:

1. There are 355 advertisements for Berkeley apartments
2. The mean price for Berkeley apartments is 3242.583 dollars
3. The median price for Berkeley apartments is 2960 dollars

Exercise 3

How do the mean and median prices for Berkeley apartments compare to San Francisco? Discuss in 1-3 sentences.

```r
# Your code goes here
sf = file[which(file$city == "San Francisco"), ]
avg_sf = mean(sf$price, na.rm = TRUE)
med_sf = median(sf$price, na.rm = TRUE)
diff_avg = avg_sf - avg_price
diff_med = med_sf - med_price
```

WRITE YOUR ANSWER BELOW: On avaerage, the prices for Berkeley apartments are 304.856 dollars lower than SF apartments, and the median of the prices for Berkeley apartments is $530 lower than that of

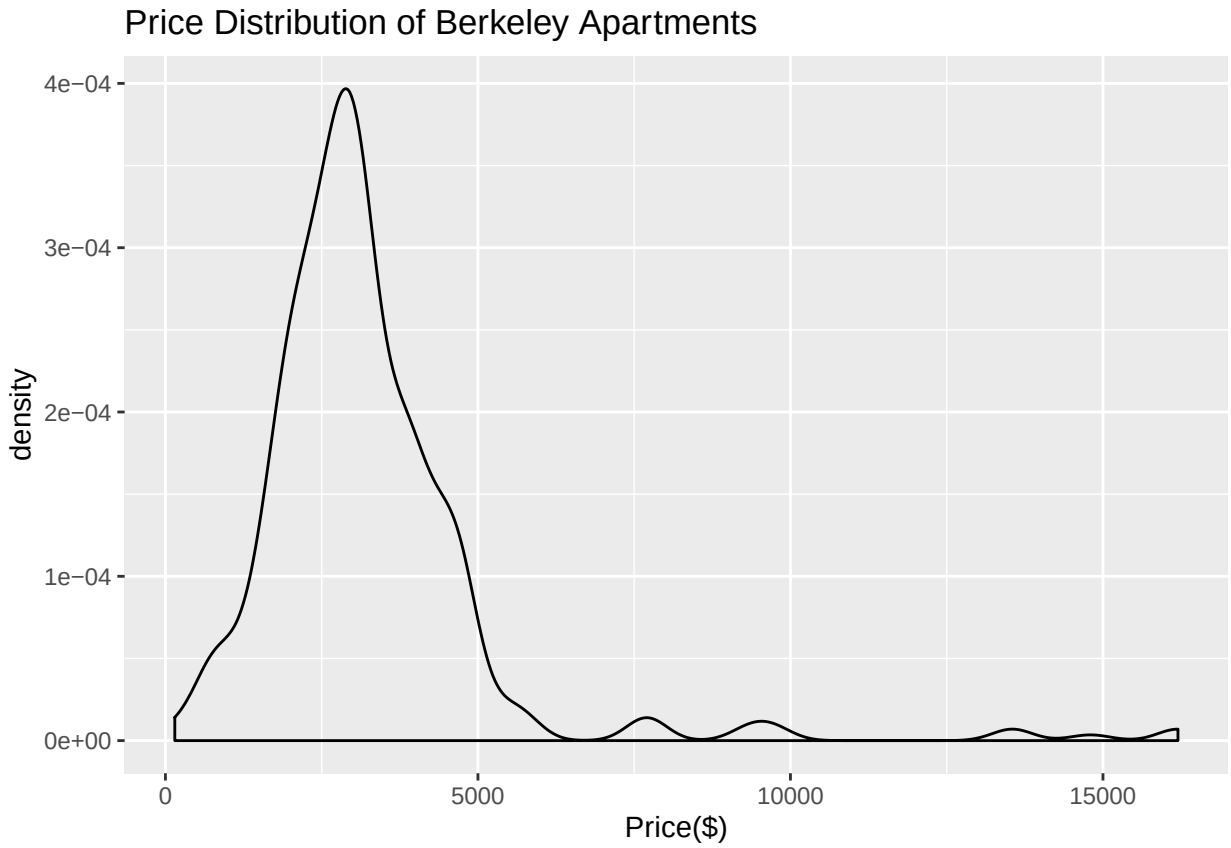
SF apartments.

Exercise 4

A density plot represents a distribution of values as a smooth curve. The height of the curve at a given point corresponds to how likely values are to fall near that point.

Use `ggplot2`'s `geom_density()` geometry to make a density plot of the price of apartments in Berkeley. Add an appropriate title to the plot with the `labs()` function. *Hint: see the online documentation for examples.*

```
# Your code goes here
library(ggplot2)
ggplot(berkeley, aes(berkeley$price)) + geom_density() + labs(title = "Price Distribution of Berkeley Ap
```



Exercise 5

Some of the apartments advertised in Berkeley list prices above 10,000 USD. Investigate these apartments:

1. How many of these apartments are there?
2. Do the listed prices appear to be correct? Use information from other columns as evidence.

   Hint: you can pretty-print the text in the `text` *and* `title` *columns with the* `cat()` *function. If you do, please DO NOT include the pretty-printed text in your submitted PDF file. Instead, tell us what you discovered.*

```
# Your code goes here
exp = berkeley[which(berkeley$price > 10000),]
exp
```

```
##
## 122         $13550 / 6br - Fully Furnished 6+ Bedroom Home Avail for Rent near UCB (berkeley north /
## 162                      $16200 / 7br - Rent the Fully Furnished AHR House on Dwight Way near UCB (be
## 3194 $14800 / 7br - 2017ft2 - Beautiful Luxury 7-Bedroom 3-Bathroom - 3 blocks from UC Berkeley! (be
## 3219        $13550 / 6br - Fully Furnished 6+ Bedroom Home Avail for Rent near UCB (berkeley north /
## 3318                     $16200 / 7br - Rent the Fully Furnished AHR House on Dwight Way near UCB (be
##
## 122
## 162  QR Code Link to This Post\n            \n        \nDwight House is a 7-bedroom home with 17 beds
## 3194
## 3219
## 3318 QR Code Link to This Post\n            \n        \nDwight House is a 7-bedroom home with 17 beds
##      latitude longitude              city_text         date_posted
## 122      37.9      -122 berkeley north / hills 2020-02-07 14:44:46
## 162      37.9      -122               berkeley 2020-02-07 16:11:35
## 3194     37.9      -122               berkeley 2020-02-12 16:54:22
## 3219     37.9      -122 berkeley north / hills 2020-02-07 14:44:46
## 3318     37.9      -122               berkeley 2020-02-07 16:11:35
##             date_updated price deleted sqft bedrooms bathrooms pets laundry
## 122  2020-02-13 11:21:40 13550   FALSE   NA        6         3 none    none
## 162  2020-02-13 11:21:27 16200   FALSE   NA        7         3 none in-unit
## 3194 2020-02-12 16:56:05 14800   FALSE 2017        7         3 none in-unit
## 3219 2020-02-13 11:21:40 13550   FALSE   NA        6         3 none    none
## 3318 2020-02-13 11:21:27 16200   FALSE   NA        7         3 none in-unit
##      parking
## 122   street
## 162     none
## 3194    none
## 3219  street
## 3318    none
##                                                                                fname
## 122      data//sfbay/_eby_apa_d_berkeley-fully-furnished-6-bedroom-home_7071138684.html
## 162        data//sfbay/_eby_apa_d_berkeley-rent-the-fully-furnished-ahr_7071192117.html
## 3194   data//sfbay_eby/_eby_apa_d_berkeley-beautiful-luxury-7-bedroom-3_7074303979.html
## 3219 data//sfbay_eby/_eby_apa_d_berkeley-fully-furnished-6-bedroom-home_7071138684.html
## 3318   data//sfbay_eby/_eby_apa_d_berkeley-rent-the-fully-furnished-ahr_7071192117.html
##      craigslist    place     city state  county
## 122       sfbay Berkeley Berkeley    CA Alameda
## 162       sfbay Berkeley Berkeley    CA Alameda
## 3194  sfbay_eby Berkeley Berkeley    CA Alameda
## 3219  sfbay_eby Berkeley Berkeley    CA Alameda
## 3318  sfbay_eby Berkeley Berkeley    CA Alameda
```

```
# Separate chunk to show code for pretty-printing, but hide output.
cat(exp$text, exp$title)
```

WRITE YOUR ANSWERS BELOW:

1. There are 5 apartmenets
2. The prices are correct