Raman Kahlon
PUID: 00291-51712

CS373 Homework 1

Part 1

A.

a. $\Omega = \{1, 2, 3, 4, 5, 6\}$
X = die turns up a 3 after i rolls
$P(X) = (\frac{5}{6})^{n-1} * \frac{1}{6}$

b. X = number of die rolls
$X = \frac{1}{6} * 1 + \frac{5}{6} * (X + 1)$
$X = \frac{1}{6} + \frac{5}{6} * X + \frac{5}{6}$
$X = 1 + \frac{5}{6} * X$
$\frac{1}{6} * X = 1$
$X = 6$
$E(X) = 6$ rolls

c. Set of outcomes = even number of die rolls = (2, 4, 6, 8, 10, …, n)
E = first time a 3 turns up is after an even number of rolls

$$P(E) = \sum_{a=1}^{\infty} (\frac{5}{6})^{(2a-1)(\frac{1}{6})}$$

$$P(E) = \frac{5}{6} * \sum_{a=1}^{\infty} (\frac{5}{6})^{(2a-1)}$$

$$P(E) = \frac{1}{6} * \sum_{a=1}^{\infty} (\frac{5}{6})^{(2a-1)}$$

$$P(E) = \frac{1}{5} * \sum_{a=0}^{\infty} (25/36)^{a}$$

$$P(E) = 5/36 * \sum_{a=0}^{\infty} (25/36)^{a} = 1/(1 - 25/36) = 5/11$$

B. $P(E) = 1 - (\frac{5}{6})^2 = 1 - 25/36 = 11/36$
$P(F) = 18/36$
$P(G) = 6/36$

a. $P(E \cap F) = P(E) - P(E \cap \neg F) = 11/36 - 6/36 = 5/36$
b. $P(E \cup F) = P(E) + P(F) - P(E \cap F) = 11/36 + 18/36 - 5/36 = 24/36 = 2/3$
c. $P(E \cup G) = P(E) + P(G) - P(E \cap G) = 11/36 + 6/36 - 2/36 = 15/36 = 5/12$
d. $P(E \cap \neg G) = P(E) - P(E \cap G) = 11/36 - 2/36 = 9/36 = 1/4$
e. $P(E \cup F \cup G) = P(E) + P(F) + P(G) - (P(E \cap F) + P(E \cap G) + P(F \cap G))$
$= 11/36 + 18/36 + 6/36 - (5/36 + 2/36 + 0/36) = 28/36 = 7/9$

C.

a. $P(C_3) = C_3/(C_1 + C_2 + C_3 + C_4)$
$P(C_3) = 3/(1+2+3+4)$
$P(C_3) = 3/10$

b. $P(C_3 \mid T) = P(C_3 \cap T)/P(T)$
$P(C_3 \mid T) = (3/10 * ½) / (1/10 * ¾ + 2/10 * ⅔ + 3/10 * ½ + 4/10 * ⅓)$
$P(C_3 \mid T) = 18/59$

D. F = student is female
C = student is majoring in computer science

a. $P(F \mid C) = P(F \cap C) / P(C) = 0.0055/0.05 = 0.11$

b. $P(C \mid F) = P(C \cap F) / P(F) = 0.0055/0.52 \approx 0.01058$

c. $P(F \mid C) = P(F \cap C) / P(C) = P(F \cap C) / 0.05 = 0.15$
$P(F \cap C) = 0.0075$
$P(C \mid F) = P(C \cap F) / P(F) = 0.0075/0.52 \approx 0.01442$

E. $P(d_1) = 0.99$
$P(d_2) = 0.97$
$P(d_3) = 0.95$

a. $P(W) = 1 - P(\neg d_1 \cap \neg d_2 \cap \neg d_3)$
$P(W) = 1 - (P(\neg d_1) * P(\neg d_2) * P(\neg d_3))$
$P(W) = 1 - (0.01 * 0.03 * 0.05)$
$P(W) = 1 - 0.000015$
$P(W) = 0.999985$

b. $P(A) = P((d_1 \cap d_2) \cup (d_3) \cup (d_1 \cap d_2 \cap d_3))$
$P(A) = P(d_1 \cap d_2) + P(d_3) + P(d_1 \cap d_2 \cap d_3)$
$P(A) = (0.99 * 0.97 * 0.05) + (0.01 * 0.03 * 0.95) + (0.99 * 0.97 * 0.95)$
$P(A) = 0.048015 + 0.000285 + 0.912285$
$P(A) \approx 0.960585$

c. $P(A \mid d_1) = 1 - P(\neg d_2 \cap \neg d_3)$
$P(A \mid d_1) = 1 - (0.03)(0.05) = 1 - 0.0015 = 0.9985$

F. X = value of sum of rolls

a. $E(X) = E(X_1) + E(X_2) + E(X_3)$
Since $E(X_1) = E(X_2) = E(X_3)$, $E(X) = 3*E(X_1)$
$E(X) = 3 * (⅙(1) + ⅙(2) + ⅙(3) + ⅙(4) + ⅙(5) + ⅙(6))$
$E(X) = 3 * (⅙ + 2/6 + 3/6 + 4/6 + ⅚ + 6/6)$
$E(X) = 3 * (21/6)$
$E(X) = 63/6 = 21/2 = 10.5$

b. $Var(X) = Var(X_1) + Var(X_2) + Var(X_3)$
Since $X_1 = X_2 = X_3$, $Var(X) = Var(3X_1) = 9 * Var(X_1)$
$Var(X_1) = E(X_1^2) - (E(X_1))^2$
$Var(X_1) = ⅙(1^2 + 2^2 + 3^2 + 4^2 + 5^2 + 6^2) - (21/6)^2$

$Var(X_1) = \frac{1}{6}(91) - (7/2)^2$

$Var(X_1) = 91/6 - 441/36 = 546/36 - 441/36 = 105/36 = 35/12$

$Var(X) = 9 * Var(X_1) = 9 * 35/12 = 105/4 = 26.25$

c. Outcomes = (A_1, A_2, A_3)

$X = \max\{A_1, A_2, A_3\} = a$

$P(X = 1) = 1/216$

$P(X = 2) = 7/216$

$P(X = 3) = 19/216$

$P(X = 4) = 37/216$

$P(X = 5) = 61/216$

$P(X = 6) = 91/216$

$E(X) = 1(1/216) + 2(7/216) + 3(19/216) + 4(37/216) + 5(61/216) + 6(91/216)$

$E(X) = 119/24$

Part 3

```r
d <- read.table(file='/Users/ramankahlon/Downloads/hw1/yelp.csv',
                sep=',',header=TRUE,quote="\"",comment.char="")
summary(d)
names(d)
```

a.

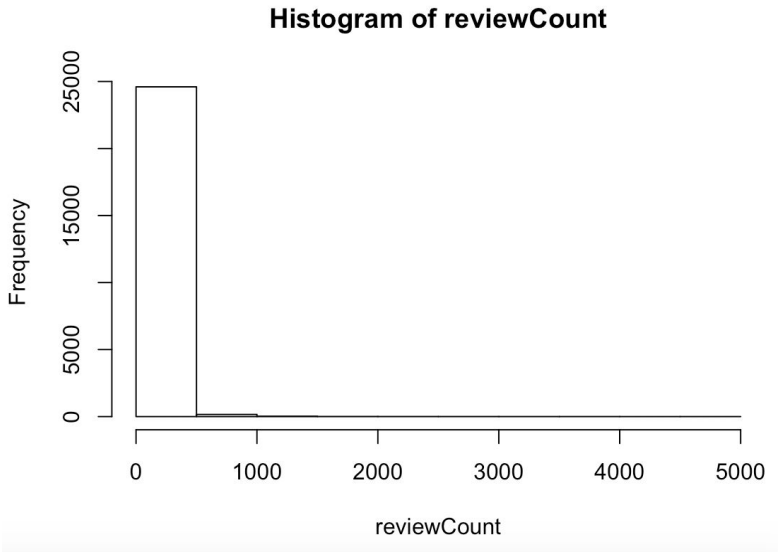
```
> summary(d)
                 business_id              name                                                                              fullAddress             city            state
 __etvGuL2dh_a1LOT0gNYQ:    1   Starbucks :  407   Bellagio Las Vegas\n3600 S Las Vegas Blvd\nThe Strip\nLas Vegas, NV 89109               :   21   Las Vegas : 5256   AZ     :9301
 __kNfrrGoUXoF-BYciMU_Q:    1   McDonald's:  275   Las Vegas, NV                                                                          :   17   Phoenix   : 3072   NV     :6296
 __Y2jjdCFHvq3rzSbpDBlw:    1   Subway    :  256   5000 S Arizona Mills Cir\nTempe, AZ 85282                                              :   14   Charlotte : 1993   QC     :2389
 _-1EgXrkOlKajCsmasuEgg:    1   Walgreens :  158   3131 Las Vegas Blvd. South\nThe Strip\nLas Vegas, NV 89109                              :   13   Pittsburgh: 1467   NC     :2370
 _-6I6VXjr-NiwIBa_1uI4A:    1   Taco Bell :  148   Monte Carlo Hotel and Casino\n3770 Las Vegas Blvd S\nThe Strip\nLas Vegas, NV 89109:   13   Scottsdale: 1296   PA     :1613
 _-9pMxBWtG_x8l4rHWBasg:    1   Wendy's   :  113   2000 E Rio Salado Pkwy\nTempe, AZ 85281                                                :   12   Montral   : 1267   WI     :1089
 (Other)               :24807   (Other)   :23456   (Other)                                                                                :24723   (Other)   :10462   (Other):1755
   latitude       longitude            stars          reviewCount          checkins          open                neighborhoods                                    categories
 Min.   :32.88   Min.   :-115.370   Min.   :1.000   Min.   :   3.00   Min.   :    3   Mode :logical   []             :15727   ['Mexican', 'Restaurants']              : 1331
 1st Qu.:33.54   1st Qu.:-114.977   1st Qu.:3.000   1st Qu.:   8.00   1st Qu.:   16   FALSE:3580      ['The Strip']:  816   ['Food', 'Coffee & Tea']                :  844
 Median :36.03   Median :-111.924   Median :3.500   Median :  18.00   Median :   48   TRUE :21233     ['Southeast']:  639   ['Pizza', 'Restaurants']                :  831
 Mean   :37.53   Mean   : -97.298   Mean   :3.544   Mean   :  49.03   Mean   :  166                   ['Downtown'] :  533   ['Chinese', 'Restaurants']              :  776
 3rd Qu.:40.41   3rd Qu.: -80.807   3rd Qu.:4.000   3rd Qu.:  48.00   3rd Qu.:  155                   ['Westside'] :  526   ['Burgers', 'Fast Food', 'Restaurants']:  549
 Max.   :55.99   Max.   :   8.549   Max.   :5.000   Max.   :4578.00   Max.   :14203                   ['Eastside'] :  447   ['Restaurants', 'Italian']              :  509
                                                                                                      (Other)      : 6125   (Other)                                 :19973
          alcohol           noiseLevel        attire         priceRange       delivery              ambience                 parking                      dietaryRestrictions waiterService
              :    3              : 7947          : 7005   Min.   :1.000   Mode :logical   ['casual']:7878   ['lot']           :10348                                :24696   Mode :logical
 beer_and_wine: 2497   average   :10957   casual:17129   1st Qu.:1.000   FALSE:14471               :7875   []                : 6675   ['vegan']                      :   45   FALSE:6208
 full_bar     : 7565   loud      : 1622   dressy:  640   Median :2.000   TRUE :3093      []        :6348   ['street']        : 3046   ['vegetarian']                 :   23   TRUE :10351
 none         :14748   quiet     : 3562   formal:   39   Mean   :1.631   NA's :7249      ['divey'] : 716                     : 2456   []                             :   20   NA's :8254
                       very_loud:   725                  3rd Qu.:2.000                   ['trendy']: 567   ['garage']        :  907   ['dairy-free', 'vegetarian']:    7
                                                         Max.   :4.000                   ['classy']: 320   ['street', 'lot']:  364   ['vegan', 'vegetarian']       :    5
                                                         NA's   :903                     (Other)   :1109   (Other)           : 1017   (Other)                        :   17
     smoking        outdoorSeating     caters                        recommendedFor goodForGroups    goodForKids
        :21862   Mode :logical   Mode :logical                               :7859   Mode :logical   Mode :logical
 no     :  904   FALSE:10989     FALSE:6503      []                        :4932   FALSE:2054      FALSE:506
 outdoor: 1415   TRUE :8698      TRUE :5932      ['lunch']                 :4324   TRUE :17078     TRUE :1283
 yes    :  632   NA's :5126      NA's :12378     ['dinner']                :2553   NA's :5681      NA's :23024
                                                 ['lunch', 'dinner']:1966
                                                 ['breakfast']             :1004
                                                 (Other)                   :2175
```

b.

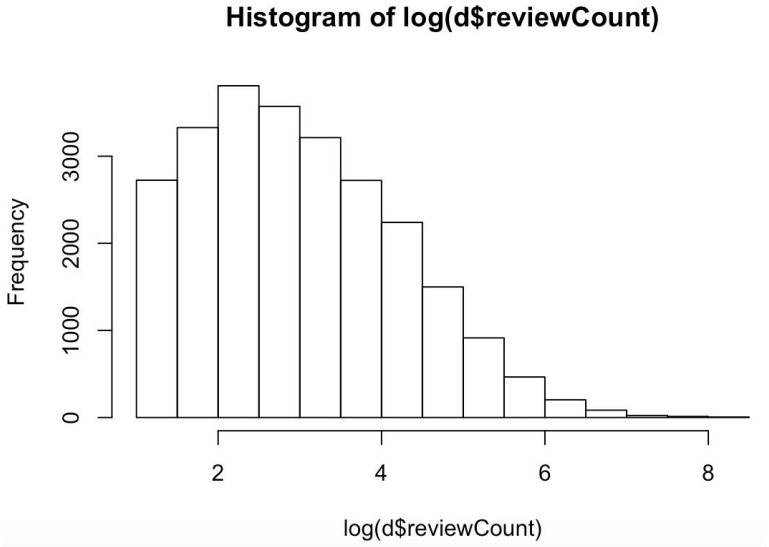
```
> names(d)
 [1] "business_id"     "name"             "fullAddress"     "city"            "state"                "latitude"        "longitude"    "stars"
 [9] "reviewCount"     "checkins"         "open"            "neighborhoods"   "categories"           "alcohol"         "noiseLevel"   "attire"
[17] "priceRange"      "delivery"         "ambience"        "parking"         "dietaryRestrictions" "waiterService"   "smoking"      "outdoorSeating"
[25] "caters"          "recommendedFor"   "goodForGroups"   "goodForKids"
```

Part 4

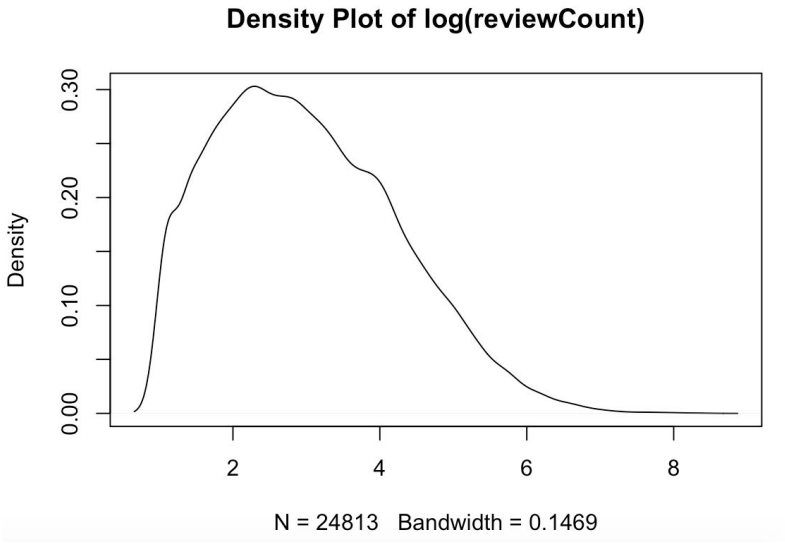
A. a.
```
hist(d[,'reviewCount'],xlab="reviewCount",main="Histogram of reviewCount")
```



b.
```
hist(log(d$'reviewCount'))
```



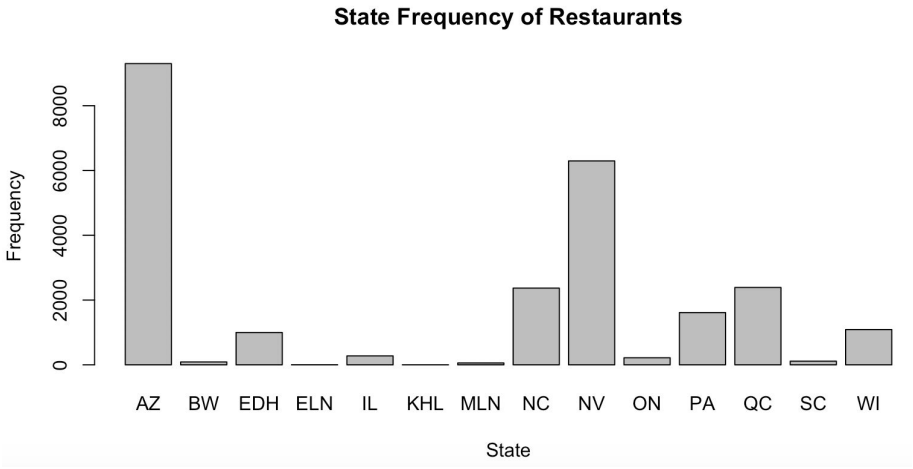
c.
```
plot(density(log(d$'reviewCount')),main="Density Plot of log(reviewCount)")
```



d. All three plots are skewed right, showing that most of the reviewCount values are towards the lower end of the range from 0 to 5000. This indicates that there are also a few upper outliers towards the reviewCount=5000 end that cause the plots to be skewed right. However, they also have differences, since the histogram of reviewCount has different units than the histogram and density plots of log(reviewCount). The log graphs make it much easier to see these upper outliers and the gradual decrease in frequency/density as the values along the x-axis increase. The log graphs also make the data less skewed.

B.

```
d['state']
counts <- table(d$'state')
names <- names(counts)
barplot(counts, main="State Frequency of Restaurants",
        xlab="State", ylab="Frequency",names.arg=names)
```



Part 5

A.

```
#5
servesPizza <- vector()
for(i in 1:length(d$'categories')){
  if(grepl('Pizza',d$'categories')[i]==TRUE)
    servesPizza <- append(servesPizza, TRUE)
  else
    servesPizza <- append(servesPizza, FALSE)
}
goodForBreakfast <- vector()
for(i in 1:length(d$'categories')){
  if(grepl('breakfast',d$'recommendedFor')[i]==TRUE)
    goodForBreakfast <- append(goodForBreakfast, TRUE)
  else
    goodForBreakfast <- append(goodForBreakfast, FALSE)
}
data <- cbind(servesPizza, goodForBreakfast)
d <- cbind(d, data)
```

B. a.

```
quantile(d$'checkins')
  0%   25%   50%   75%  100%
   3    16    48   155 14203
```

b.

```
checkins <- subset(d, checkins <= 16)
summary(checkins)
```

```
> summary(checkins)
              business_id                 name                                                fullAddress          city          state
 __Y2jjdCFHvq3rzSbpDBlw:   1   Subway           :  73   Las Vegas, NV                                    :   9   Las Vegas : 776   AZ     :1752
 _-6I6VXjr-NiwIBa_1uI4A:   1   Pizza Hut        :  70   5000 S Arizona Mills Cir\nTempe, AZ 85282        :   7   Montral   : 645   QC     :1387
 _-EB8tQzBlM_jLkgtRw4Rg:   1   Domino's Pizza   :  49   160 University Avenue W\nWaterloo, ON N2L 3E9:   6   Phoenix   : 584   NV     : 927
 _04PNAespgMZVXBJrkmbNA:   1   McDonald's       :  49   Montral, QC                                      :   6   Edinburgh : 540   NC     : 528
 _0DI4UXAaFC6hOYpBadtIw:   1   Burger King      :  39   138, avenue Atwater\nMontreal, QC H4C 2G3        :   5   Pittsburgh: 451   EDH    : 521
 _0ZajBG5CSBSyxeeZV276g:   1   Papa John's Pizza:  33   224 E 7th St\nFirst Ward\nCharlotte, NC 28202:   5   Montreal  : 436   PA     : 514
 (Other)               :6385   (Other)          :6078   (Other)                                          :6353   (Other)   :2959   (Other): 762
    latitude       longitude          stars        reviewCount        checkins        open             neighborhoods
 Min.   :32.88   Min.   :-115.370   Min.   :1.000   Min.   :  3.000   Min.   : 3.000   Mode :logical   []           :4347
 1st Qu.:33.65   1st Qu.:-112.073   1st Qu.:3.000   1st Qu.:  4.000   1st Qu.: 5.000   FALSE:1378      ['New Town'] : 123
 Median :36.24   Median : -80.953   Median :3.500   Median :  7.000   Median : 8.000   TRUE :5013      ['Downtown'] : 118
 Mean   :40.22   Mean   : -85.417   Mean   :3.484   Mean   :  9.004   Mean   : 8.739                   ['The Strip']: 102
 3rd Qu.:45.50   3rd Qu.: -73.608   3rd Qu.:4.000   3rd Qu.: 11.000   3rd Qu.:12.000                   ['Southeast']:  98
 Max.   :55.99   Max.   :   8.549   Max.   :5.000   Max.   :230.000   Max.   :16.000                   ['Westside'] :  91
                                                                                                      (Other)      :1512
                    categories            alcohol         noiseLevel      attire       priceRange      delivery                ambience
 ['Pizza', 'Restaurants']  : 371              :   0           :3072         :2237   Min.   :1.000   Mode :logical                :3145
 ['Mexican', 'Restaurants']: 311   beer_and_wine: 477   average  :1708   casual:3954   1st Qu.:1.000   FALSE:2771      []           :2599
 ['Chinese', 'Restaurants']: 291   full_bar     :1403   loud     : 324   dressy: 181   Median :2.000   TRUE :1112      ['casual']   : 481
 ['Restaurants', 'Italian']: 181   none         :4511   quiet    :1118   formal:  19   Mean   :1.674   NA's :2508      ['divey']    :  54
 ['Food', 'Coffee & Tea']  : 125                        very_loud: 169                 3rd Qu.:2.000                   ['trendy']   :  24
 ['Food', 'Grocery']       : 120                                                       Max.   :4.000                   ['romantic']:  15
 (Other)                   :4992                                                       NA's   :663                     (Other)      :  73
            parking                                dietaryRestrictions waiterService      smoking     outdoorSeating    caters
 []             :2983                                     :6348     Mode :logical          :5938   Mode :logical   Mode :logical
                :1430   ['vegan']                         :  30     FALSE:1422      no     : 140   FALSE:2836      FALSE:972
 ['lot']        :1074   []                                :  10     TRUE :2180      outdoor: 244   TRUE :1524      TRUE :773
 ['street']     : 713   ['dairy-free', 'vegan', 'vegetarian']:   1   NA's :2789      yes    :  69   NA's :2031      NA's :4646
 ['street', 'lot']:  58   ['dairy-free', 'vegetarian']      :   1
 ['garage']     :  57   ['vegetarian']                    :   1
 (Other)        :  76   (Other)                           :   0
           recommendedFor goodForGroups   goodForKids     servesPizza     goodForBreakfast
                 :2776    Mode :logical   Mode :logical   Mode :logical   Mode :logical
 []              :2332    FALSE:756       FALSE:54        FALSE:5738      FALSE:6279
 ['lunch', 'dinner']: 353   TRUE :3734      TRUE :141       TRUE :653       TRUE :112
 ['lunch']       : 345    NA's :1901      NA's :6196
 ['dinner']      : 321
 ['breakfast']   :  64
 (Other)         : 200
```

c.

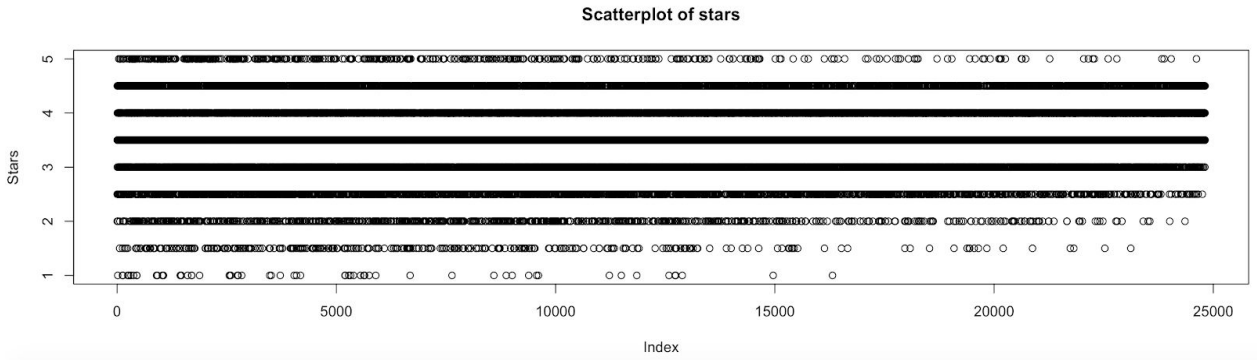
```
> summary(checkins$'checkins')
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  3.000   5.000   8.000   8.739  12.000  16.000
> summary(checkins$'stars')
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  1.000   3.000   3.500   3.484   4.000   5.000
> summary(checkins$'noiseLevel')
            average      loud     quiet very_loud
     3072      1708       324      1118       169
> summary(checkins$'priceRange')
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's
  1.000   1.000   2.000   1.674   2.000   4.000     663
> summary(checkins$'reviewCount')
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  3.000   4.000   7.000   9.004  11.000 230.000
> summary(checkins$'goodForGroups')
   Mode    FALSE     TRUE     NA's
logical      756     3734     1901
```

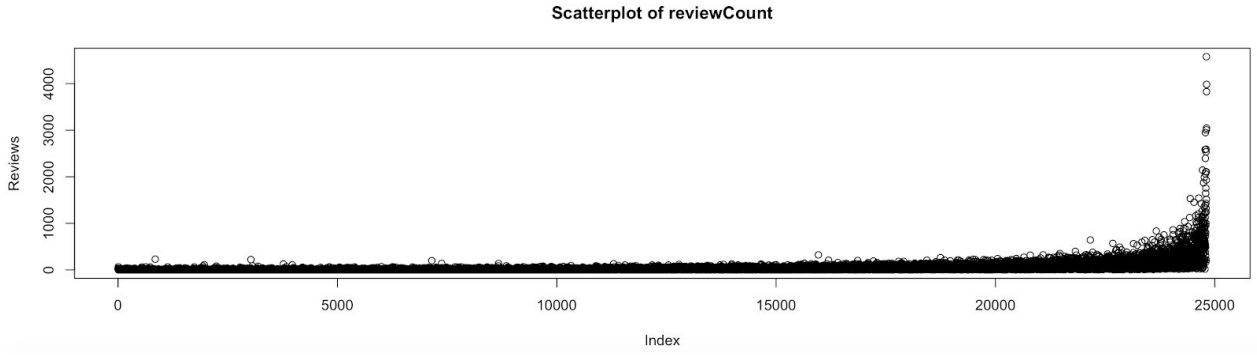
The distribution for the attributes in each dataset is almost the same or at least very similar. The only apparent difference between the datasets is the total number of fields for the attributes.

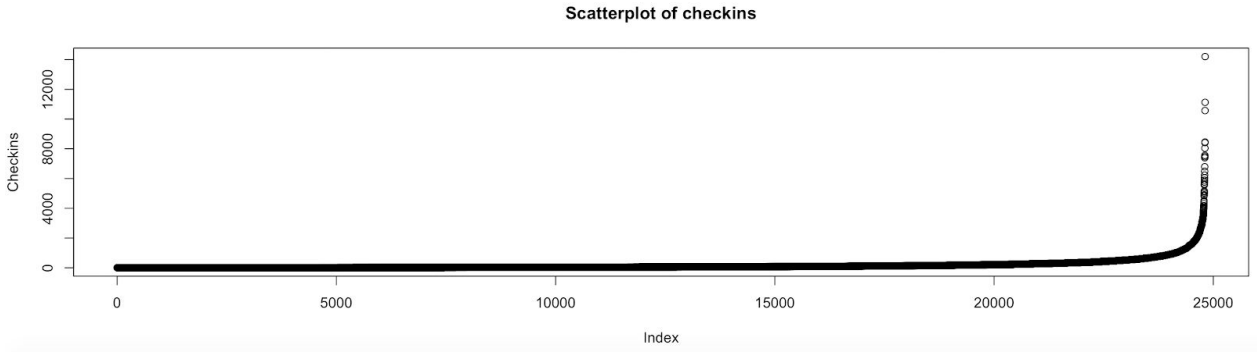
Part 6

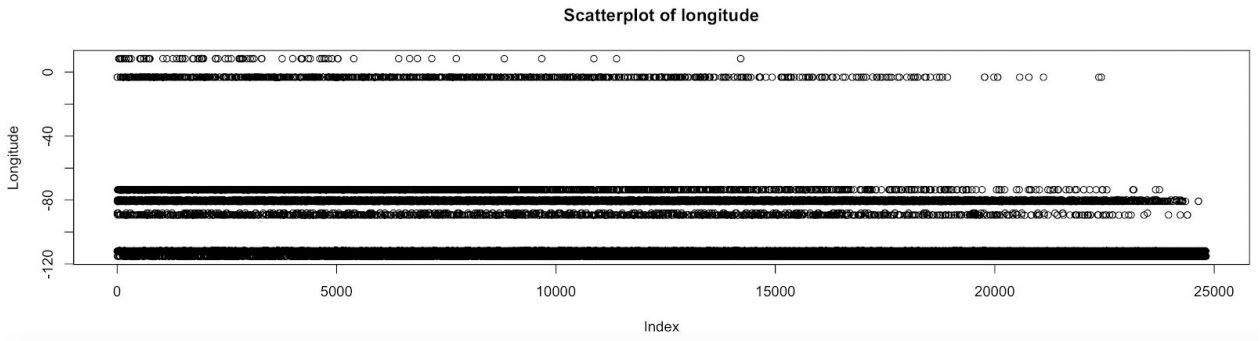
A.

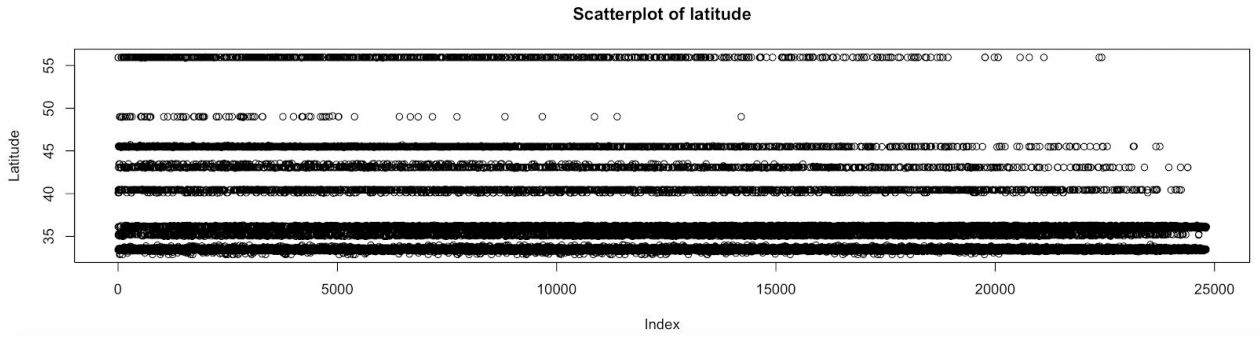
```
#6
plot(d$'stars', main='Scatterplot of stars', ylab='Stars')
plot(d$'reviewCount', main='Scatterplot of reviewCount', ylab='Reviews')
plot(d$'checkins', main='Scatterplot of checkins', ylab='Checkins')
plot(d$'longitude', main='Scatterplot of longitude', ylab='Longitude')
plot(d$'latitude', main='Scatterplot of latitude', ylab='Latitude')
```











The attributes ‘reviewCount’ and ‘checkins’ exhibit the most association, since it seems that the number of reviews and checkins drastically increase towards the upper end of the index axis (bottom end of the dataframe). This positive association makes sense since, for a greater number of visitors (checkins), we would expect the restaurant to have a larger number of reviews. Also, the more that the visitors enjoy a specific restaurant, it is more likely for them to leave a review than an average/mediocre restaurant.

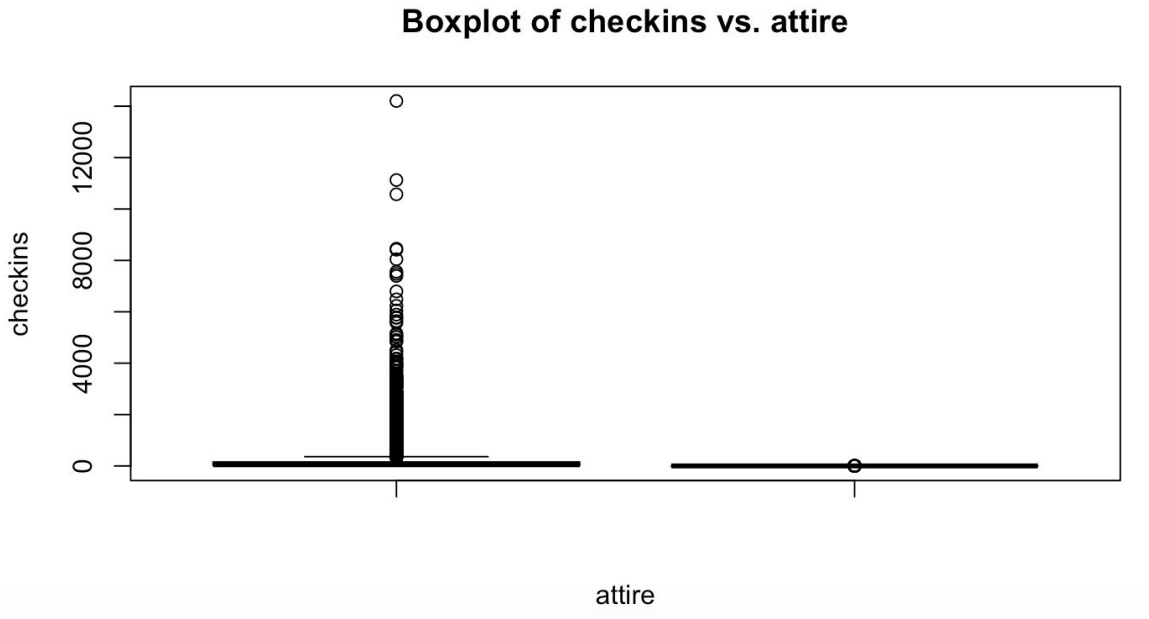
B.

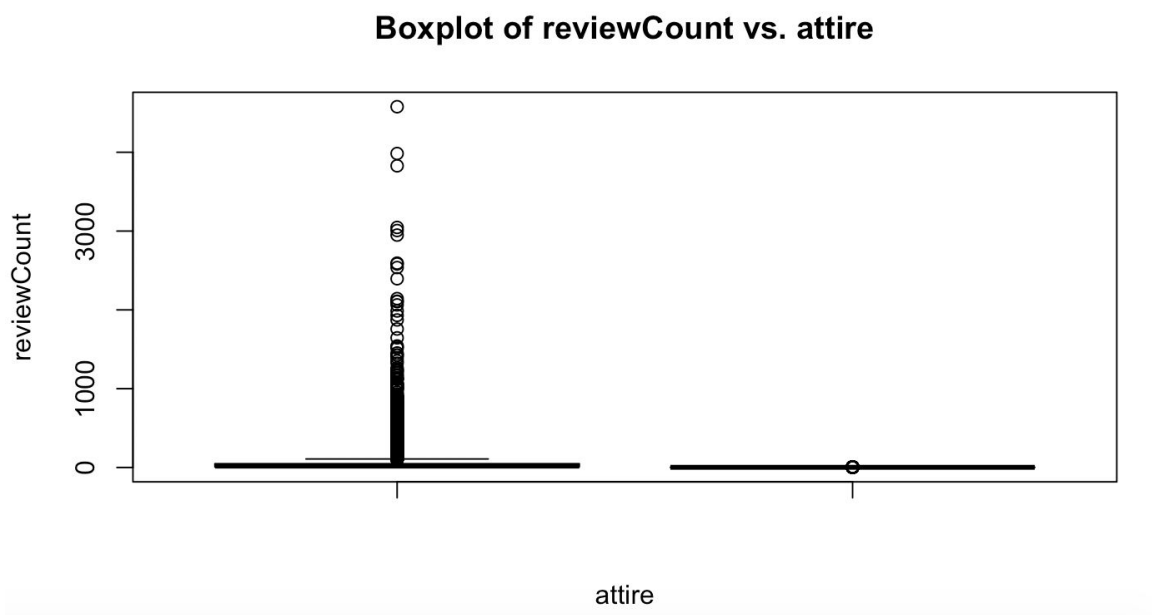
```
> cor(d$'stars',d$'reviewCount')
[1] 0.1070506
> cor(d$'stars',d$'checkins')
[1] 0.09440071
> cor(d$'stars',d$'longitude')
[1] 0.1174446
> cor(d$'stars',d$'latitude')
[1] 0.1211631
> cor(d$'reviewCount',d$'checkins')
[1] 0.8274936
> cor(d$'reviewCount',d$'longitude')
[1] -0.1294142
> cor(d$'reviewCount',d$'latitude')
[1] -0.09850936
> cor(d$'checkins',d$'longitude')
[1] -0.1789531
> cor(d$'checkins',d$'latitude')
[1] -0.1526046
> cor(d$'longitude',d$'latitude')
[1] 0.8811018
```

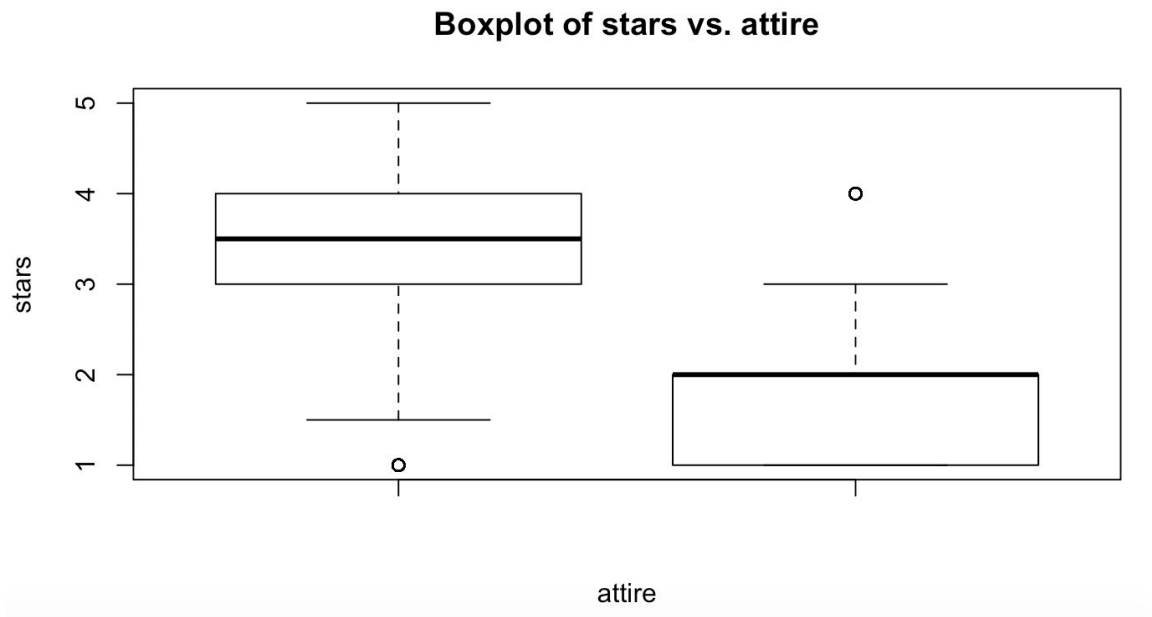
The pair of attributes with the largest positive correlation is ‘longitude’ and ‘latitude’ (correlation = 0.8811018) and the second largest positive correlation was between ‘checkins’ and ‘reviewCount’, which exhibited the most association in part a. Meanwhile, the pair of attributes with the largest negative correlation is ‘checkins’ and ‘longitude’ (correlation = -0.1789531). Negative longitude is the western hemisphere, so this negative correlation would make sense because we would expect that most Yelp users are located in the U.S., so most of the checkins would occur in the U.S.

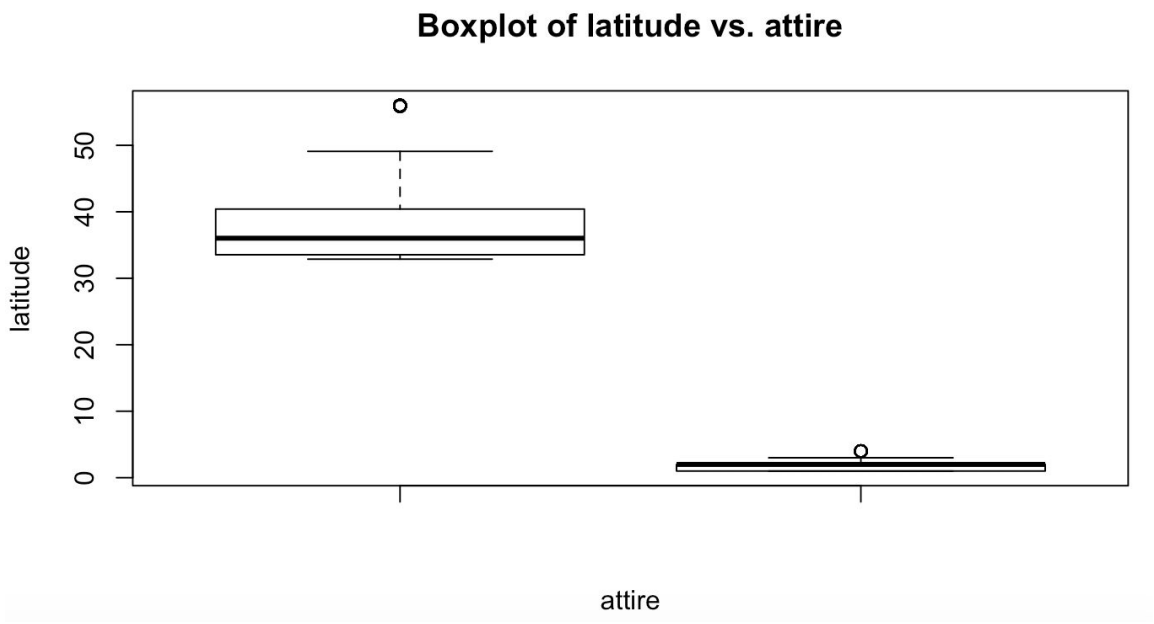
C.

```
boxplot(d$'checkins', d$'attire', main="Boxplot of checkins vs. attire",
        xlab="attire", ylab="checkins")
boxplot(d$'reviewCount', d$'attire', main="Boxplot of reviewCount vs. attire",
        xlab="attire", ylab="reviewCount")
boxplot(d$'stars', d$'attire', main="Boxplot of stars vs. attire",
        xlab="attire", ylab="stars")
boxplot(d$'latitude', d$'attire', main="Boxplot of latitude vs. attire",
        xlab="attire", ylab="latitude")
```




Boxplot of reviewCount vs. attire
reviewCount
0
1000
3000
attire


Boxplot of stars vs. attire
stars
1
2
3
4
5
attire



a. The ‘stars’ attribute exhibits the most association with ‘attire’. The attributes ‘stars’ and ‘attire’ are both discrete variables having similar boxplots. This is expected since a high ‘stars’ rating for a restaurant might indicate that it is a classy/formal establishment, which requires classy or formal attire.

b.

```
> casual <- subset(d, attire=="casual")
> dressy <- subset(d, attire=="dressy")
> formal <- subset(d, attire=="formal")
> none <- subset(d, attire=="")
> quantile(casual$stars)
  0%  25%  50%  75% 100%
 1.0  3.0  3.5  4.0  5.0
> quantile(dressy$stars)
  0%  25%  50%  75% 100%
 1.5  3.5  4.0  4.0  5.0
> quantile(formal$stars)
  0%  25%  50%  75% 100%
 1.0  2.5  3.5  4.0  5.0
> quantile(none$stars)
  0%  25%  50%  75% 100%
 1.0  3.0  3.5  4.0  5.0
```

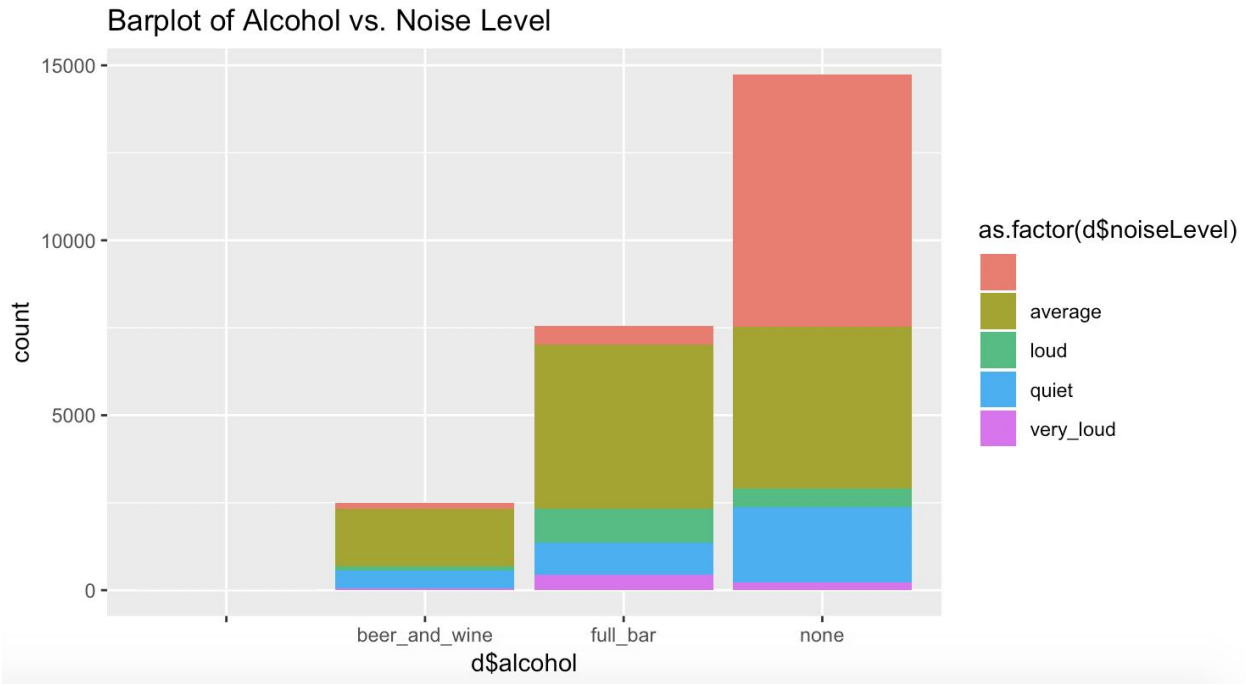
From the R output above, we can observe that the none and casual subsets have the same 5 summaries. The only difference is that the dress attire data has a high median and minimum. This supports the claim that dressy attire is associated with high ratings (‘stars’).

Part 7

1.

```
library(ggplot2)
ggplot(data = d) + geom_bar(aes(x=d$alcohol, fill = as.factor(d$noiseLevel))) +
  ggtitle("Barplot of Alcohol vs. Noise Level")
```

a.



b. Both of the variables (alcohol and noise level) are discrete. Due to this, a bar plot would be sufficient to compare them.

c. Assume that X is alcohol and Y is noise level. Then, X is associated with Y.

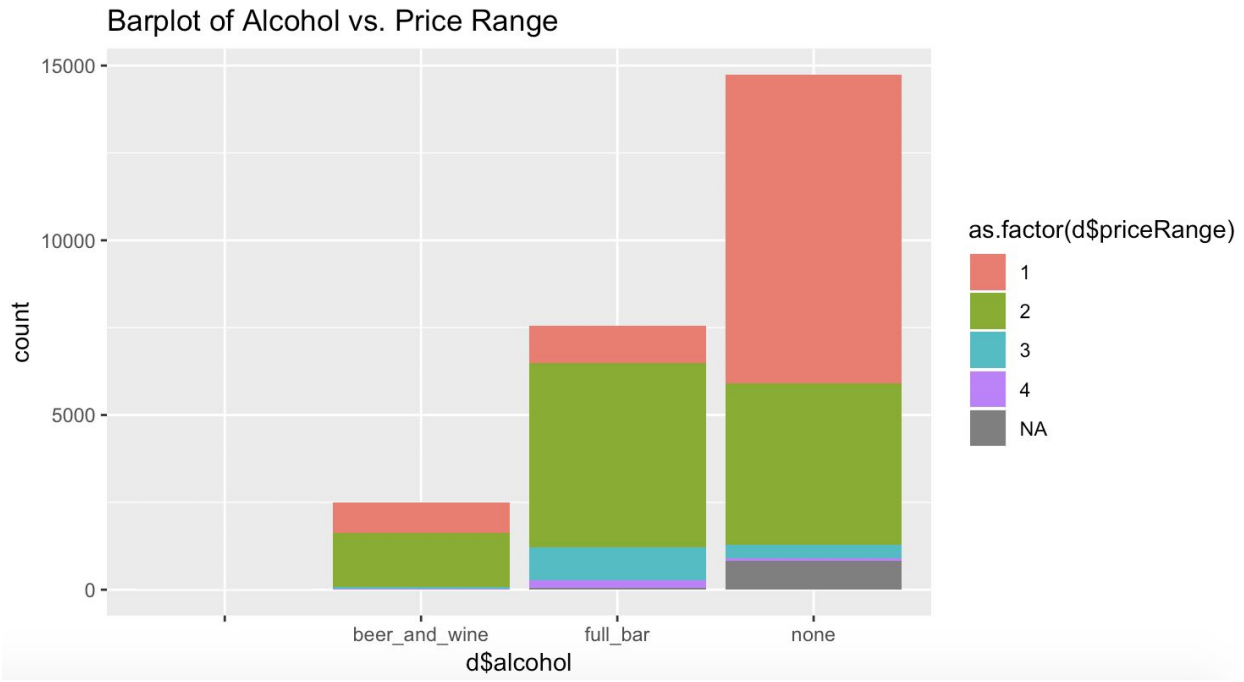
d. From this barplot, we can see that the more alcohol that a restaurant offers, the higher the average noise level of that restaurant will be.

e. This is a directional causal hypothesis.

2.

```
ggplot(data = d) + geom_bar(aes(x=d$alcohol, fill = as.factor(d$priceRange))) +
  ggtitle("Barplot of Alcohol vs. Price Range")
```

a.



b. Both of the variables (alcohol and price range) are discrete. Due to this, a bar plot would be sufficient to compare them.

c. Assume that X is alcohol and Y is price range. Then, X is associated with Y.

d. From this barplot, we can see that the more alcohol that a restaurant offers, the higher the average price range will be.

e. This is a directional causal hypothesis.