Tuango Analysis

true true

Preliminaries

Determine notebook defaults:

Load packages:

Read in the data:

```
# use load("filename.Rdata") for .Rdata files
# rm(list=ls())
load("/Users/dain/Programs/R_Projects/MKTG_482_HW2/Tuango_rfm.Rdata")
```

Assignment answers

Question 1

What percentage of customers responded (i.e. bought anything) after the push message?

```
tuango %>% filter(buyer==1) %>%
  summarise(perc_buyer=percent(n()/nrow(tuango), 0.01))
```

```
# A tibble: 1 x 1
  perc_buyer
  <chr>
1 3.10%
```

Question 2

Of those who bought, what was the average spending?

(Hint: use filter(buyer==1) in the dplyr package to grab the relevant rows in the data.)

```
tuango %>% filter(buyer==1) %>%
  summarise(average_spend=dollar(mean(ordersize)))
```

```
# A tibble: 1 x 1
  average_spend
  <chr>
1 $202.36
```

Question 3

Create (independent) quintile variables for recency, frequency and monetary. Display a table with the first six observations of the quintile variables only.

(Hint: review the file "RFM_BBB.Rmd" which goes through the calculations for the Bookbinders RFM analysis).

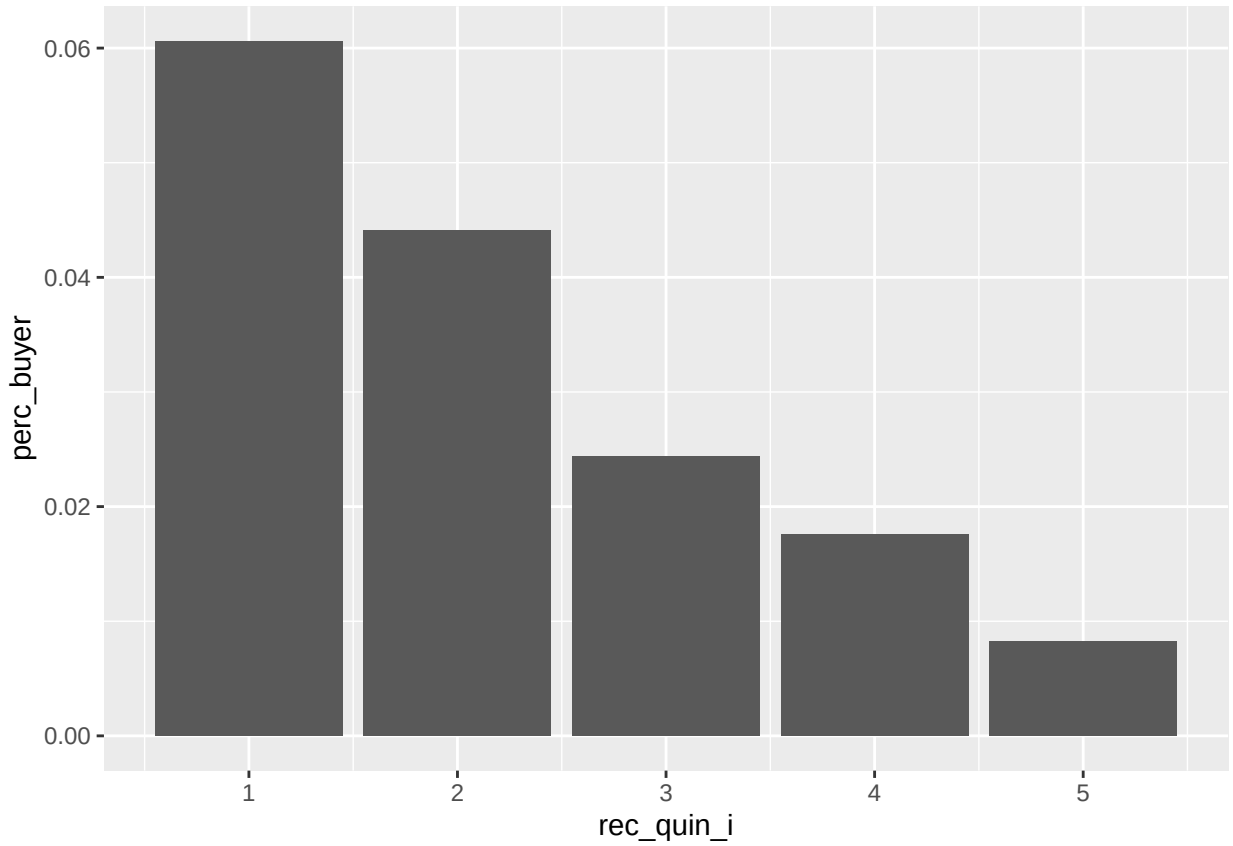
```
tuango_iq <- tuango %>%
  mutate(rec_quin_i = ntile(recency, 5),
         freq_quin_i = ntile(-frequency, 5),
         mon_quin_i = ntile(-monetary, 5))

tuango_iq %>%
  select(recency, frequency, monetary, rec_quin_i, freq_quin_i, mon_quin_i) %>%
  slice(0:6)
```

```
# A tibble: 6 x 6
  recency frequency monetary rec_quin_i freq_quin_i mon_quin_i
    <int>     <int>    <dbl>      <int>       <int>      <int>
1     309         7     39.8          5           1          3
2     297         8     39.8          5           1          3
3     295         1     72.9          5           3          2
4     277         1     40            5           3          3
5     259         1     21            5           3          4
6     243         1     19.9          5           3          4
```

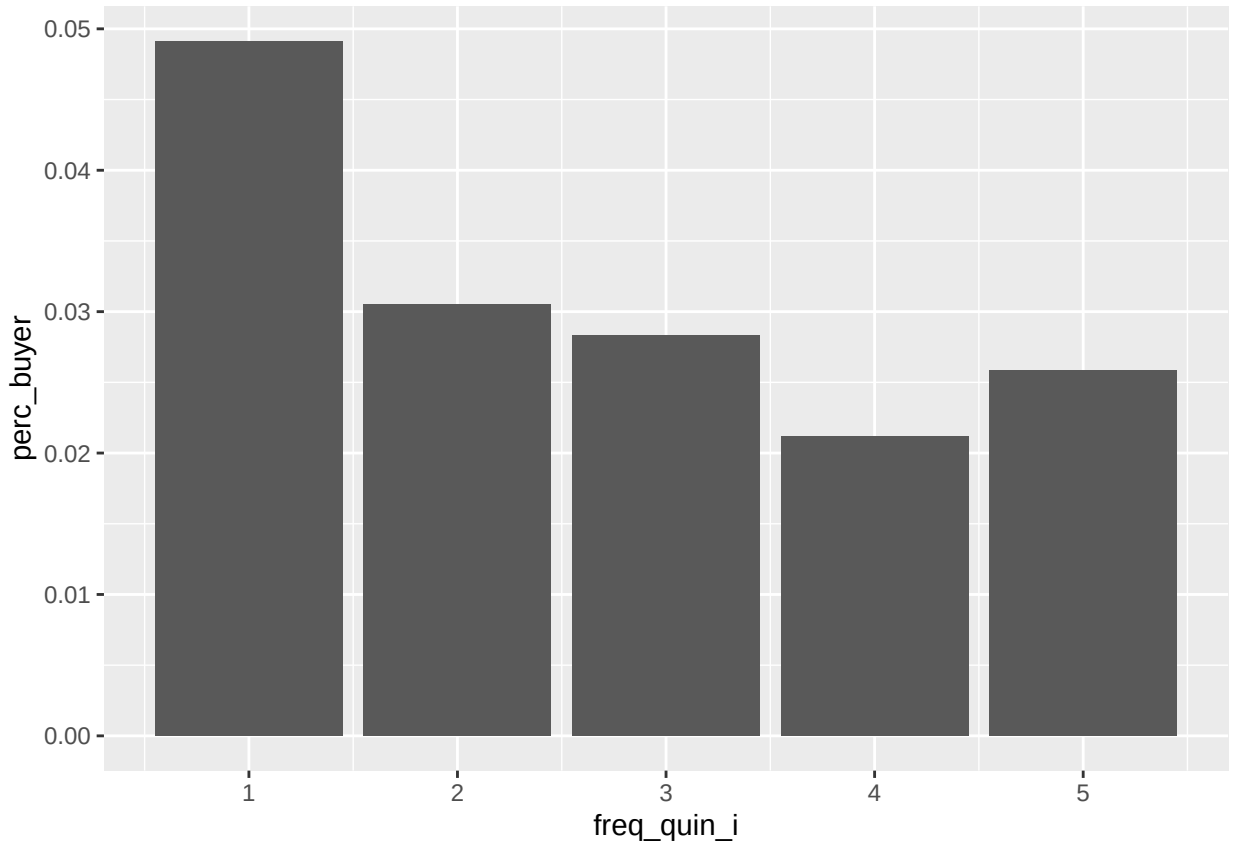
Question 4

Create bar chart showing the response rate (i.e., the proportion of customers who bought something) to this deal by • recency quintile • frequency quintile • monetary quintile

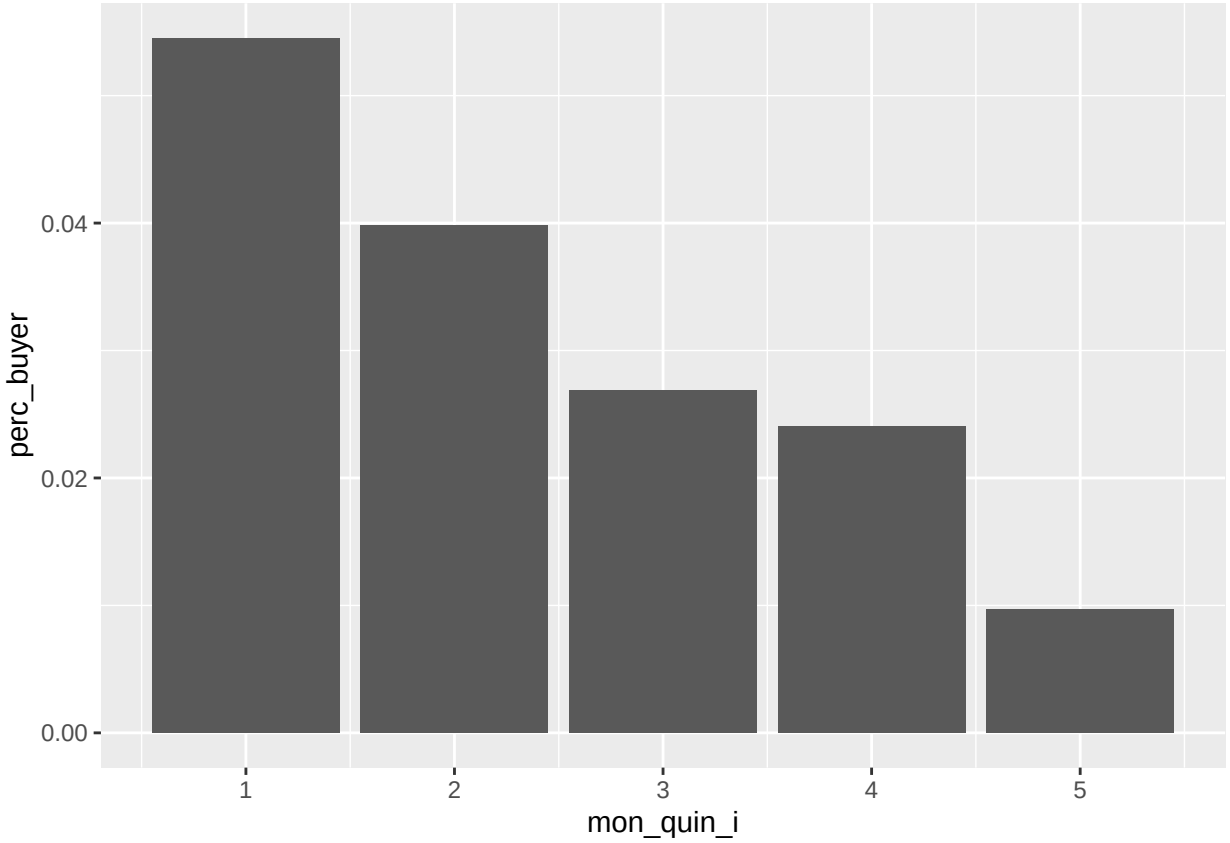
```
tuango_iq %>%
  group_by(rec_quin_i) %>%
  summarize(perc_buyer = mean(buyer)) %>%
  ggplot() + geom_col(aes(x = rec_quin_i, y = perc_buyer))
```




```
tuango_iq %>%
  group_by(freq_quin_i) %>%
  summarize(perc_buyer = mean(buyer)) %>%
  ggplot() + geom_col(aes(x = freq_quin_i, y = perc_buyer))
```




```
tuango_iq %>%
  group_by(mon_quin_i) %>%
  summarize(perc_buyer = mean(buyer)) %>%
  ggplot() + geom_col(aes(x = mon_quin_i, y = perc_buyer))
```

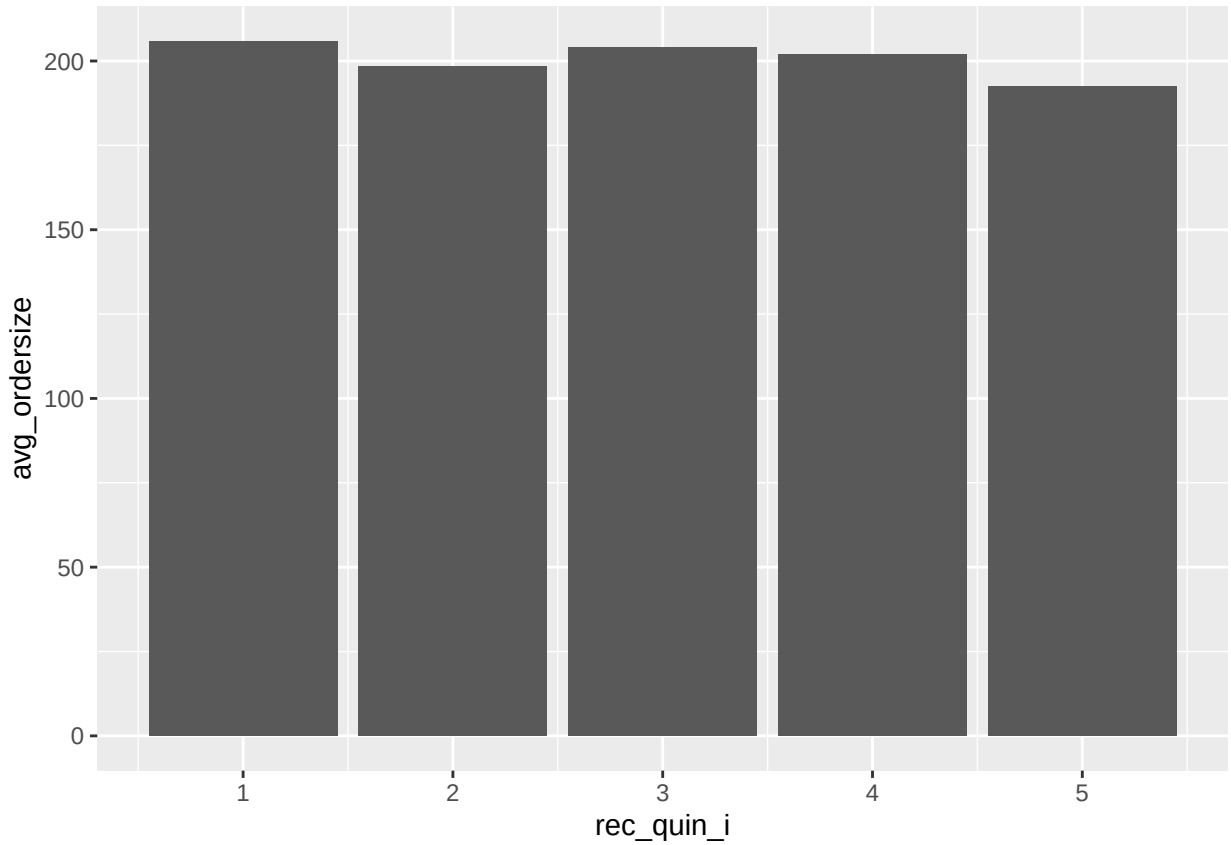


Question 5

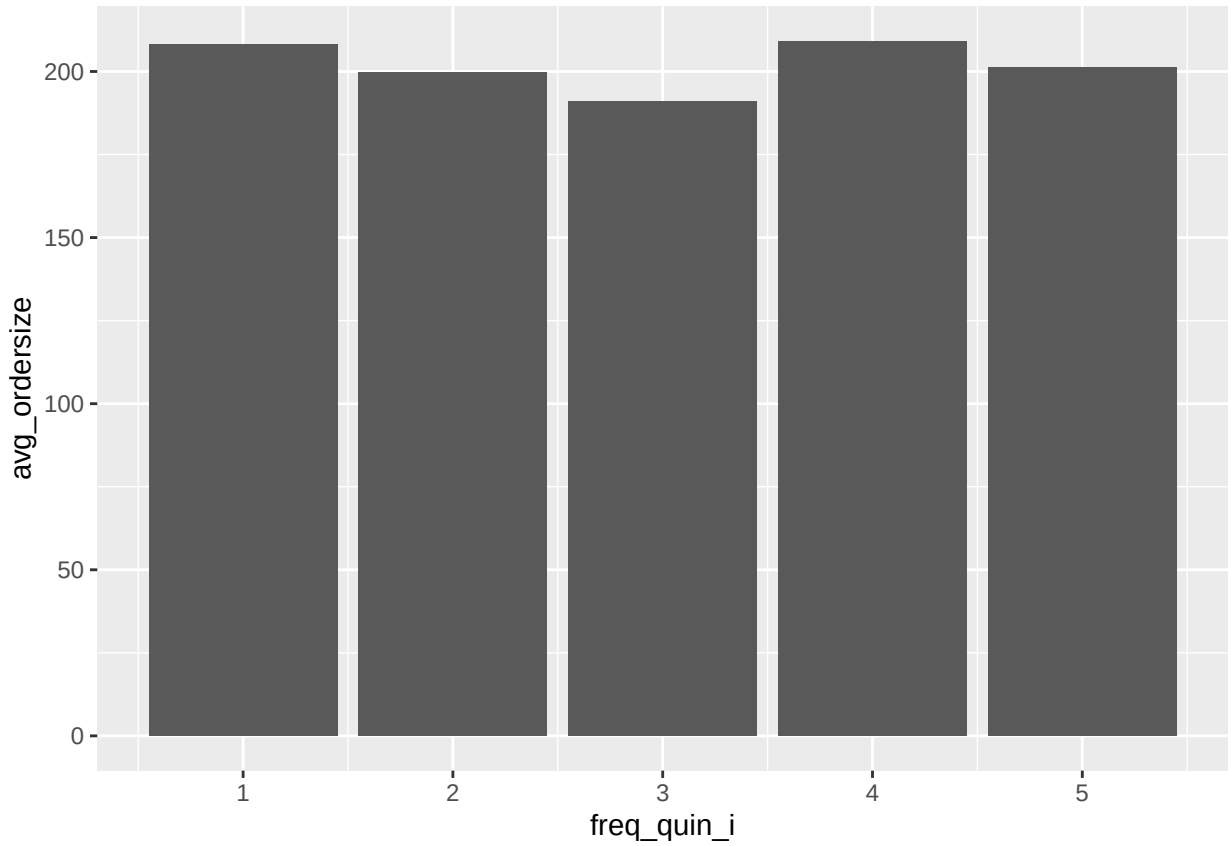
Repeat questions 4 using only those customers who placed an order after the push message, i.e. create bar charts showing the average spending (in RMB) by recency, frequency and monetary quintile.

(Hint: use filter(buyer==1) in the dplyr package to first create a dataframe containing only consumers who placed an order. Use that dataframe as an input to ggplot.)

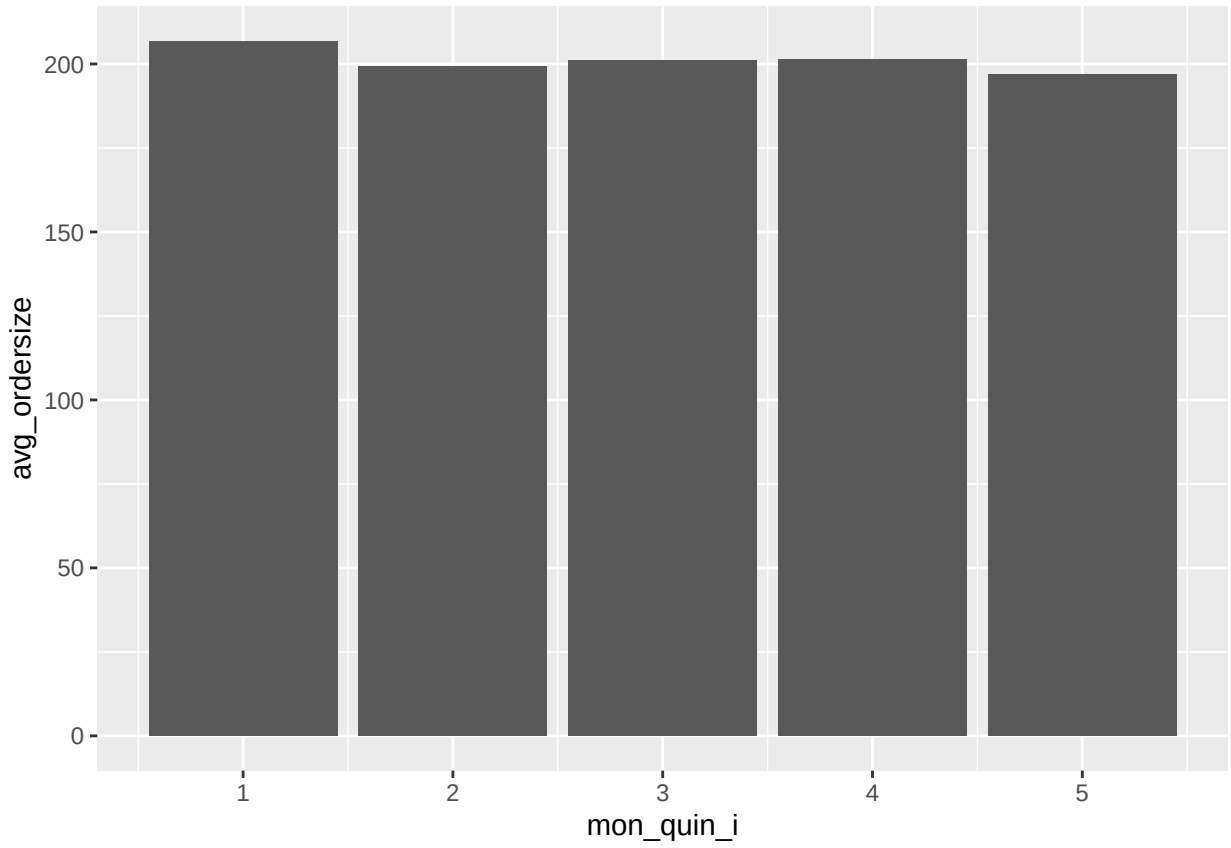
```
tuango_iq %>%
  filter(buyer==1) %>%
  group_by(rec_quin_i) %>%
  summarize(avg_ordersize = mean(ordersize)) %>%
  ggplot() + geom_col(aes(x = rec_quin_i, y = avg_ordersize))
```




```
tuango_iq %>%
  filter(buyer==1) %>%
  group_by(freq_quin_i) %>%
  summarize(avg_ordersize = mean(ordersize)) %>%
  ggplot() + geom_col(aes(x = freq_quin_i, y = avg_ordersize))
```




```
tuango_iq %>%
  filter(buyer==1) %>%
  group_by(mon_quin_i) %>%
  summarize(avg_ordersize = mean(ordersize)) %>%
  ggplot() + geom_col(aes(x = mon_quin_i, y = avg_ordersize))
```



Question 6

What do the above bar charts reveal about the likelihood of response and the size of the order across the different recency, frequency, and monetary quintiles?

```
"
The response rates of the first and second quintile of Recency & Monetary are fairly
higher than the last 3 quintiles. The first Frequency quintile is higher, but the
other 4 quintiles are relatively the same. It appears there is little difference
amongst the avg ordersize quintiles, so orderise of all purchases is relatively uniform.
"
```

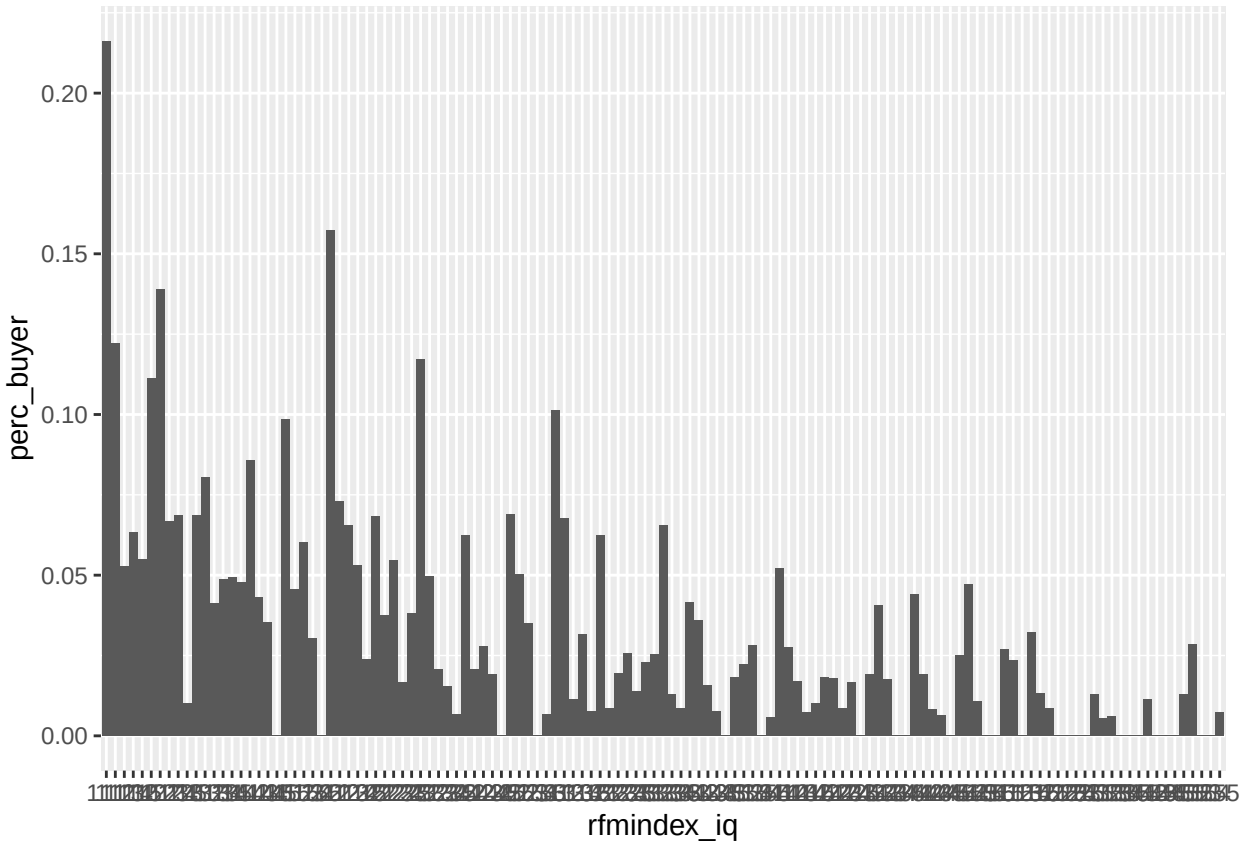
```
[1] "\nThe response rates of the first and second quintile of Recency & Monetary are fairly \nhigher tha
```

Part 2

Question 7

Create the independent RFM index in R.

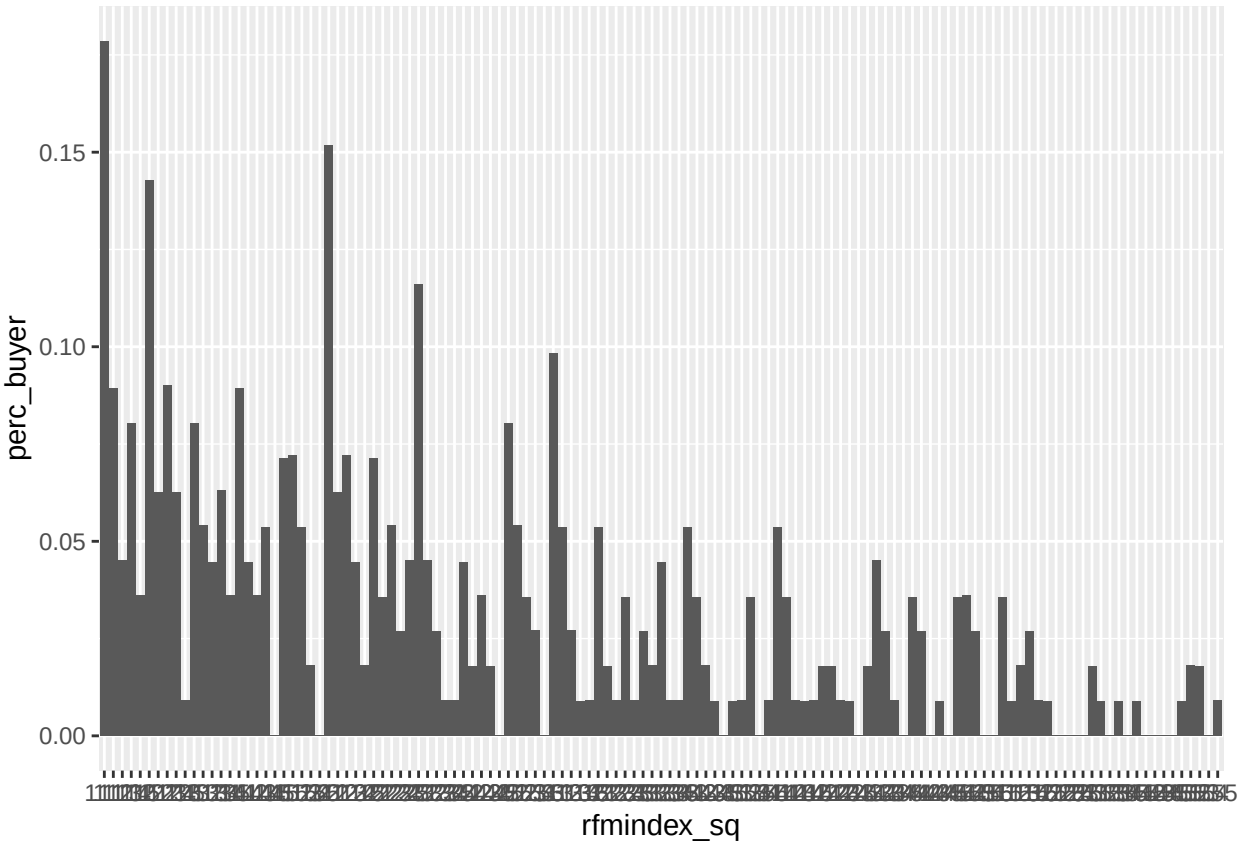
```
tuango_iq <- tuango_iq %>%
  mutate(rfmindex_iq = paste(rec_quin_i, freq_quin_i, mon_quin_i, sep = ''))

tuango_iq %>%
  group_by(rfmindex_iq) %>%
  summarise(perc_buyer=mean(buyer)) %>%
  ggplot() + geom_col(aes(x = rfmindex_iq, y = perc_buyer))
```



Question 8

Create the sequential RFM index in R.

```
tuango_iq_sq <- tuango_iq %>%
  mutate(rec_quin_s = ntile(recency, 5)) %>%
  group_by(rec_quin_s) %>%
  mutate(freq_quin_s = ntile(-frequency, 5)) %>%
  group_by(rec_quin_s, freq_quin_s) %>%
  mutate(mon_quin_s = ntile(-monetary, 5)) %>%
  ungroup()

tuango_iq_sq <- tuango_iq_sq %>%
  mutate(rfmindex_sq = paste(rec_quin_s, freq_quin_s, mon_quin_s, sep = ''))

tuango_iq_sq %>%
  group_by(rfmindex_sq) %>%
  summarise(perc_buyer=mean(buyer)) %>%
  ggplot() + geom_col(aes(x = rfmindex_sq, y = perc_buyer))
```



Question 9

Visually examine the observations in the database. What do you notice about the independent and sequential index values? That is – do the two approaches generally yield the same RFM index for any given customer? What do you see as the pros and cons of the two approaches (from a statistical as well as logical perspective) and why?

(Hint: Calculate and compare how many consumers are in each "cell" for the independent and sequential index. You can count the number of observations using the dplyr summarise(n()) function.)

```
tuango_iq %>%
   group_by(rfmindex_iq) %>% summarise(count=n()) %>% arrange(desc(count)) %>%
   arrange(desc(rfmindex_iq)) %>%
   slice(0:5)
```

```
# A tibble: 5 x 2
  rfmindex_iq count
  <chr>       <int>
1 555           138
2 554            52
3 553            71
4 552           105
5 551           155
```

```
tuango_iq_sq %>%
   group_by(rfmindex_sq) %>% summarise(count=n()) %>% arrange(desc(count)) %>%
   arrange(desc(rfmindex_sq)) %>%
   slice(0:5)
```

```
# A tibble: 5 x 2
  rfmindex_sq count
  <chr>       <int>
1 555           111
2 554           111
3 553           112
4 552           111
5 551           112
```

```
"
The results are predominantly the same (the graphs are similar). Sequential method
results are more evenly distributed customer count within each cell.~

Sequential (Nested):
Pros:
* Even distribution of customers within each RFM cell / group
Cons:
* Assignment of customers and the results become harder to intrepet.

Independent:
Pros:
* Unambiguous assigment for 3 RFM categories.
Cons:
* For smaller samples, may result in less even distribution of combined
RFM scores,resulting in more empty cells.
"
```

```
[1] "\nThe results are predominantly the same (the graphs are similar). Sequential method \nresults are
```

Part 3

Question 10

What is the breakeven response rate?

```
# Marginal Cost: 2.5RMB
# Fee per sale: 50%
# Average Order Size: ~$202

marginal_cost <- 2.5
gross_rev <- tuango %>% filter(buyer==1) %>%
  summarise(average_spend=mean(ordersize))
rev_split <- 0.5
net_rev <- gross_rev$average_spend * rev_split
response_rate <- marginal_cost / net_rev
percent(response_rate, 0.01)
```

```
[1] "2.47%"
```

Question 11

What is the projected profit in RMB if you offer the deal to all remaining 264,841 customers.

```
ncustomers <- 264841
marginal_cost <- 2.5
total_response_rate <- tuango_iq_sq %>% summarise(perc_buyer=mean(buyer))
```

```
dollar(ncustomers * total_response_rate$perc_buyer * net_rev - (ncustomers * marginal_cost))
```

```
[1] "$168,368"
```

Question 12

Consider offering the deal only to those of the 264,841 customers in RFM cells (using the sequential n-tiles approach) with a response rate that is equal to or greater than the breakeven response rate. * Calculate the response rate for the sequential quintiles approach. * Merge the response rate back into main data. * Calculate who you should have mailed to in your data. * Calculate the projected profit in RMB. What is it?

(Hint: review the file "RFM_BBB.Rmd" which goes through the calculations in the Bookbinders RFM analysis).

```
resp_rate_sq_df <- tuango_iq_sq %>%
  group_by(rfmindex_sq) %>%
  summarise(resp_rate_sq=mean(buyer)) %>%
  ungroup()

tuango_final <- tuango_iq_sq %>%
  left_join(resp_rate_sq_df)

tuango_final <- tuango_final %>%
  mutate(mailto_sq=1*(resp_rate_sq>response_rate))

targeted_customers_sq <- tuango_final %>%
  filter(mailto_sq==1) %>%
  summarise(frac_mailed=n()/nrow(tuango_final), resp_rate=mean(buyer))

# targeted_customers_sq

ntargeted_customers <- 264841 * targeted_customers_sq$frac_mailed
targeted_resp_rate <- targeted_customers_sq$resp_rate

dollar(ntargeted_customers * targeted_resp_rate * net_rev - (ntargeted_customers * marginal_cost))
```

```
[1] "$404,472"
```