BBB

Yuhan Xu

1/19/2020

```
library(tidyverse)
```

```
## -- Attaching packages ---------------------------------------------------
-------------- tidyverse 1.3.0 --

## v ggplot2 3.2.1     v purrr   0.3.3
## v tibble  2.1.3     v dplyr   0.8.3
## v tidyr   1.0.0     v stringr 1.4.0
## v readr   1.3.1     v forcats 0.4.0

## -- Conflicts ------------------------------------------------------------
-------- tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
library(gmodels)
BookBinders = read.csv("BBB.csv")
attach(BookBinders)
```

1

Q: What percent of BookBinders customers are female?

```
CrossTable(gender)
```

```
##
##
##    Cell Contents
## |-------------------------|
## |                       N |
## |         N / Table Total |
## |-------------------------|
##
##
## Total Observations in Table:  50000
##
##
##            |         F |         M |
##            |-----------|-----------|
##            |     33302 |     16698 |
##            |     0.666 |     0.334 |
##            |-----------|-----------|
##
```

```
##
##
##
```

A: 66.6% of BookBinders customers are female.

2

Q: Which three states account for the largest percentage of BookBinders's customers?

```
CrossTable(state)
```

```
##
##
##    Cell Contents
## |-------------------------|
## |                       N |
## |         N / Table Total |
## |-------------------------|
##
##
## Total Observations in Table:  50000
##
##
##            |        AE |        CT |        DC |        DE |        MA |
##            |-----------|-----------|-----------|-----------|-----------|
##            |         5 |      2512 |       339 |       711 |      4252 |
##            |     0.000 |     0.050 |     0.007 |     0.014 |     0.085 |
##            |-----------|-----------|-----------|-----------|-----------|
##
##
##            |        MD |        ME |        NH |        NJ |        NY |
##            |-----------|-----------|-----------|-----------|-----------|
##            |      4172 |       343 |       665 |     11068 |     16530 |
##            |     0.083 |     0.007 |     0.013 |     0.221 |     0.331 |
##            |-----------|-----------|-----------|-----------|-----------|
##
##
##            |        PA |        RI |        VA |        VI |        VT |
##            |-----------|-----------|-----------|-----------|-----------|
##            |      8718 |       402 |        27 |        45 |       211 |
##            |     0.174 |     0.008 |     0.001 |     0.001 |     0.004 |
##            |-----------|-----------|-----------|-----------|-----------|
##
##
##
##
```

```
sort(table(state), decreasing = TRUE)
```

```
## state
##    NY    NJ    PA    MA    MD    CT    DE    NH    RI    ME    DC    VT
VI
## 16530 11068  8718  4252  4172  2512   711   665   402   343   339   211
45
##    VA    AE
##    27     5
```

A: NY, NJ and PA account for the largest percentage of BookBinders's customers.

3

Q: What is the average Total $ spent, the average Total # of book purchases, and the average number of months since last purchase?

```
summarise(BookBinders, avg_spent = mean(total_), avg_book = mean(purch),
avg_months = mean(last))
```

```
##   avg_spent avg_book avg_months
## 1  208.3183  3.89022   12.35816
```

A: Average Total $ spent is $208.3, the average Total # of book purchases is 3.9, and the average number of months since last purchase is 12.4.

4

Q: Calculate the correlation between customers' total spending on books and their total spending on non-book products.

```
cor(book_, nonbook_)
```

```
## [1] 0.1574359
```

```
cor.test(book_, nonbook_)
```

```
##
##  Pearson's product-moment correlation
##
## data:  book_ and nonbook_
## t = 35.648, df = 49998, p-value < 2.2e-16
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  0.1488761 0.1659721
## sample estimates:
##       cor
## 0.1574359
```

A: The correlation is 0.16.

5

Q: Which book categories have sold the most books? Which have sold the least?

```
sort(
  summarise(
    BookBinders,
    tot_child = sum(child),
    tot_youth = sum(youth),
    tot_cook = sum(cook),
    tot_do_it = sum(do_it),
    tot_refernce = sum(refernce),
    tot_art = sum(art),
    tot_geog = sum(geog)
  ),
  decreasing = TRUE
)
```

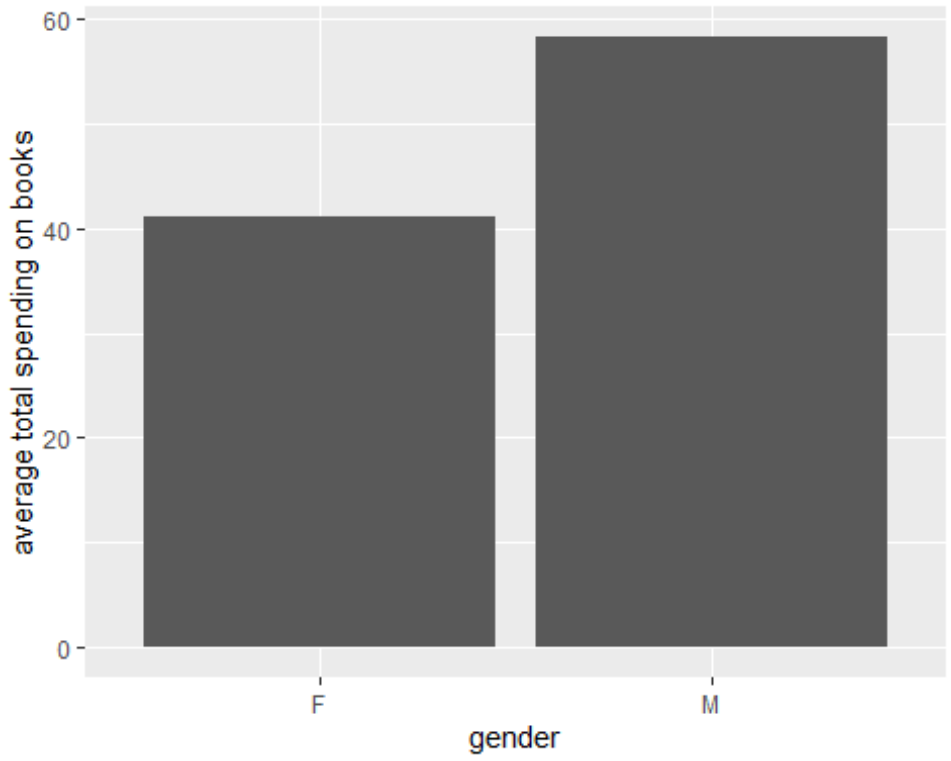
```
##   tot_cook tot_child tot_geog tot_do_it tot_youth tot_art tot_refernce
## 1    46830     42723    27348     23153     19549   19296        15612
```

A: Cook books have been sold the most. Reference books have been sold the least.

6

Q: Create a bar chart showing the average total spending on books for males and females.
A:

```
BookBinders1 = BookBinders %>% group_by(gender) %>% summarise(avg_spending =
mean(book_))
ggplot(data = BookBinders1, aes(x = gender, y = avg_spending)) +
  labs(x = "gender", y = "average total spending on books") +
  geom_bar(stat = "identity")
```



7

Q: For both males and females, find their respective total number and also the percent who bought "The Art History of Florence."

```
BookBinders %>% group_by(gender, buyer) %>% summarise(total = n()) %>%
  mutate(percent = total/sum(total)*100)
```

```
## # A tibble: 4 x 4
## # Groups:   gender [2]
##   gender buyer total percent
##   <fct>  <int> <int>   <dbl>
## 1 F          0 30913   92.8
## 2 F          1  2389    7.17
## 3 M          0 14565   87.2
## 4 M          1  2133   12.8
```

A: The total number of males who bought "The Art History of Florence." is 2389, which accounts for 7.17%. The total number of females who bought "The Art History of Florence." is 2133, which accounts for 12.77%.

8

Q: For both males and females, determine the total number of purchases and the average number of purchases by males vs. females.

```
BookBinders %>% group_by(gender) %>%
  summarise(total_purchase = sum(purch), avg_purchase = mean(purch))
```

```
## # A tibble: 2 x 3
##   gender total_purchase avg_purchase
##   <fct>           <int>        <dbl>
## 1 F              111968         3.36
## 2 M               82543         4.94
```

A: The total number of purchases by males is 111,968. The average number of purchases by males is 3.36. The total number of purchases by females is 82,543. The average number of purchases by females is 4.94.

9

Q: Determine the minimum, the maximum, and the average number of months between customers' first purchase and their most recent purchase.

```
summarise(
  BookBinders,
  mini_month = min(first - last),
  max_month = max(first - last),
  avg_month = mean(first - last)
)
```

```
##   mini_month max_month avg_month
## 1          0        72  13.31076
```

A: The minimum number of months is 0. The maximum number of months is 72. The average number of months is 13.31.

10

Q: What percent of repeat customers (those with two or more total purchases) bought "The Art History of Florence?"

```
BookBinders %>% filter(purch >= 2) %>% group_by(buyer) %>% summarise(total =
n()) %>%
  mutate(percent = total/sum(total)*100)
```

```
## # A tibble: 2 x 3
##   buyer total percent
##   <int> <int>   <dbl>
## 1     0 31282    89.7
## 2     1  3598    10.3
```

A: The percent of repeat customers is 10.31%.