supermarket sales

Mingwei Wu

Data Description

Id: computer generated sales slip invoice indentification number

Branch: branch of supercenter (3 branches are avaliable identified by A, B and C)

City: Location of supercenters

Type: type of customers, recorded by members for customers using member card and normal for without member card

Gender: gender type of customer

Product: general item categorization groups - electronic accessories, fashion accessories, food and beverages, health and beauty, home and lifestyle, sports and travel

Price: price of each product in dollar

Quantity: numbers of products purchased by customer

Tax: 5% tax fee for customer buying

Total: total price including tax

Date: date for purchase (record available from Jan 2019 to March 2019)

Time: purchase time (10am to 8pm)

Payment: payment used by customer for purchase(3 methods are available - Cash, Credit card and Ewallet)

COGS: cost of goods sold

Gross income: gross income

Rating: customer stratification rating on their overall shopping experience (on a scale of 1 to 10)

Import data

```
sales<-read.csv("supermarket.csv",header = TRUE)
head(sales)
```

```
##    Invoice.ID Branch      City Customer.type Gender           Product.line
## 1 750-67-8428      A    Yangon        Member Female      Health and beauty
## 2 226-31-3081      C Naypyitaw        Normal Female Electronic accessories
```

```
## 3 631-41-3108      A    Yangon        Normal   Male     Home and lifestyle
## 4 123-19-1176      A    Yangon        Member   Male      Health and beauty
## 5 373-73-7910      A    Yangon        Normal   Male      Sports and travel
## 6 699-14-3026      C Naypyitaw        Normal   Male Electronic accessories
##   Unit.price Quantity  Tax.5.    Total      Date  Time     Payment   cogs
## 1      74.69        7 26.1415 548.9715  1/5/2019 13:08     Ewallet 522.83
## 2      15.28        5  3.8200  80.2200  3/8/2019 10:29        Cash  76.40
## 3      46.33        7 16.2155 340.5255  3/3/2019 13:23 Credit card 324.31
## 4      58.22        8 23.2880 489.0480 1/27/2019 20:33     Ewallet 465.76
## 5      86.31        7 30.2085 634.3785  2/8/2019 10:37     Ewallet 604.17
## 6      85.39        7 29.8865 627.6165 3/25/2019 18:30     Ewallet 597.73
##   gross.margin.percentage gross.income Rating
## 1                4.761905      26.1415    9.1
## 2                4.761905       3.8200    9.6
## 3                4.761905      16.2155    7.4
## 4                4.761905      23.2880    8.4
## 5                4.761905      30.2085    5.3
## 6                4.761905      29.8865    4.1
```

##rename data frame

```
library(tidyverse)
```

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

## v ggplot2 3.3.2     v purrr   0.3.4
## v tibble  3.0.4     v dplyr   1.0.2
## v tidyr   1.1.2     v stringr 1.4.0
## v readr   1.4.0     v forcats 0.5.0

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

```
library(dplyr)
names(sales)
```

```
##  [1] "Invoice.ID"             "Branch"
##  [3] "City"                   "Customer.type"
##  [5] "Gender"                 "Product.line"
##  [7] "Unit.price"             "Quantity"
##  [9] "Tax.5."                 "Total"
## [11] "Date"                   "Time"
## [13] "Payment"                "cogs"
## [15] "gross.margin.percentage" "gross.income"
## [17] "Rating"
```

```
data<-sales%>%
  rename(Id=Invoice.ID,
         Type=Customer.type,
         Product=Product.line,
```

```
        Price=Unit.price,
        Tax=Tax.5.,
        gross_percent=gross.margin.percentage,
        gross_income=gross.income)
```

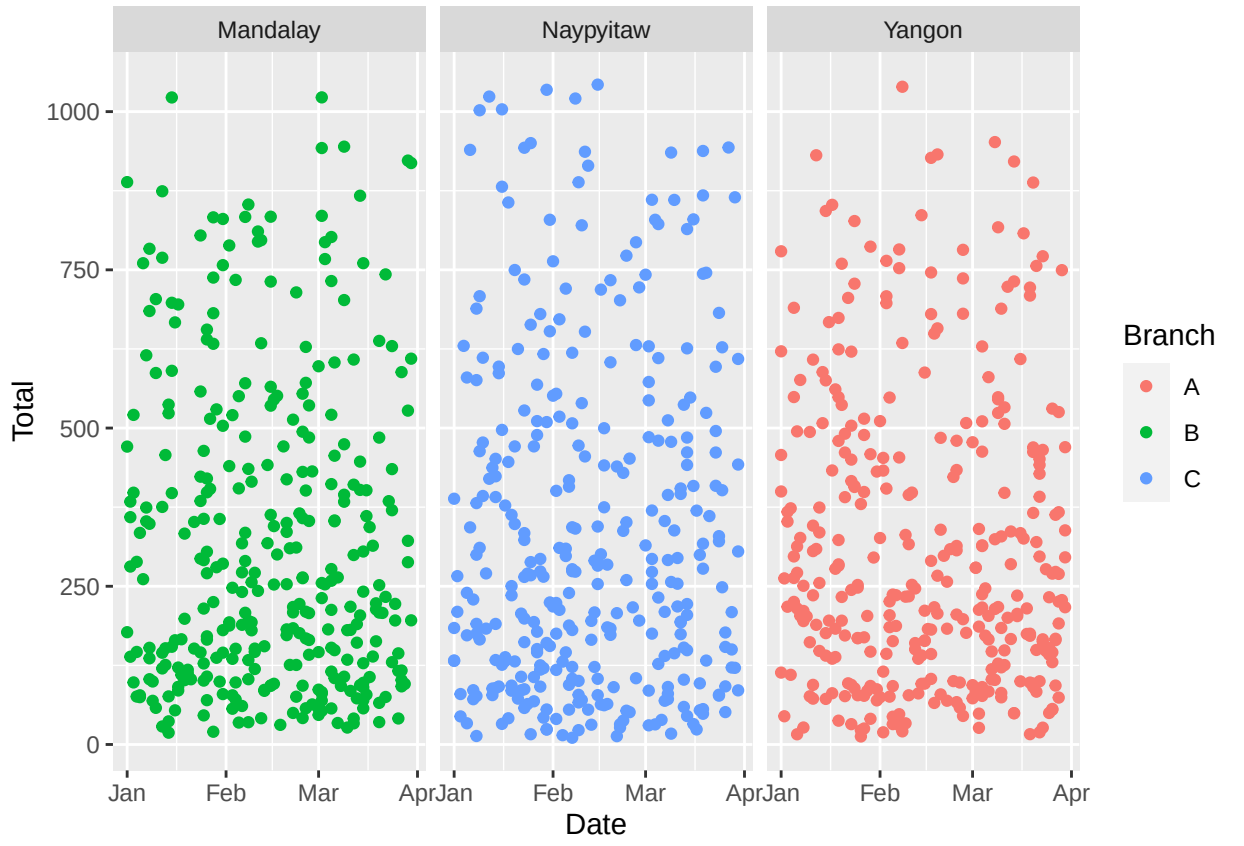
```
data$Date<-as.Date(data$Date,"%m/%d/%y")
```

```
summary(data)
```

```
##       Id                Branch              City                Type
##  Length:1000        Length:1000        Length:1000        Length:1000
##  Class :character   Class :character   Class :character   Class :character
##  Mode  :character   Mode  :character   Mode  :character   Mode  :character
##
##
##
##     Gender             Product              Price          Quantity
##  Length:1000        Length:1000        Min.   :10.08   Min.   : 1.00
##  Class :character   Class :character   1st Qu.:32.88   1st Qu.: 3.00
##  Mode  :character   Mode  :character   Median :55.23   Median : 5.00
##                                        Mean   :55.67   Mean   : 5.51
##                                        3rd Qu.:77.94   3rd Qu.: 8.00
##                                        Max.   :99.96   Max.   :10.00
##       Tax              Total              Date                Time
##  Min.   : 0.5085   Min.   :  10.68   Min.   :2020-01-01   Length:1000
##  1st Qu.: 5.9249   1st Qu.: 124.42   1st Qu.:2020-01-24   Class :character
##  Median :12.0880   Median : 253.85   Median :2020-02-13   Mode  :character
##  Mean   :15.3794   Mean   : 322.97   Mean   :2020-02-14
##  3rd Qu.:22.4453   3rd Qu.: 471.35   3rd Qu.:2020-03-08
##  Max.   :49.6500   Max.   :1042.65   Max.   :2020-03-30
##    Payment               cogs        gross_percent    gross_income
##  Length:1000        Min.   : 10.17   Min.   :4.762   Min.   : 0.5085
##  Class :character   1st Qu.:118.50   1st Qu.:4.762   1st Qu.: 5.9249
##  Mode  :character   Median :241.76   Median :4.762   Median :12.0880
##                     Mean   :307.59   Mean   :4.762   Mean   :15.3794
##                     3rd Qu.:448.90   3rd Qu.:4.762   3rd Qu.:22.4453
##                     Max.   :993.00   Max.   :4.762   Max.   :49.6500
##      Rating
##  Min.   : 4.000
##  1st Qu.: 5.500
##  Median : 7.000
##  Mean   : 6.973
##  3rd Qu.: 8.500
##  Max.   :10.000
```

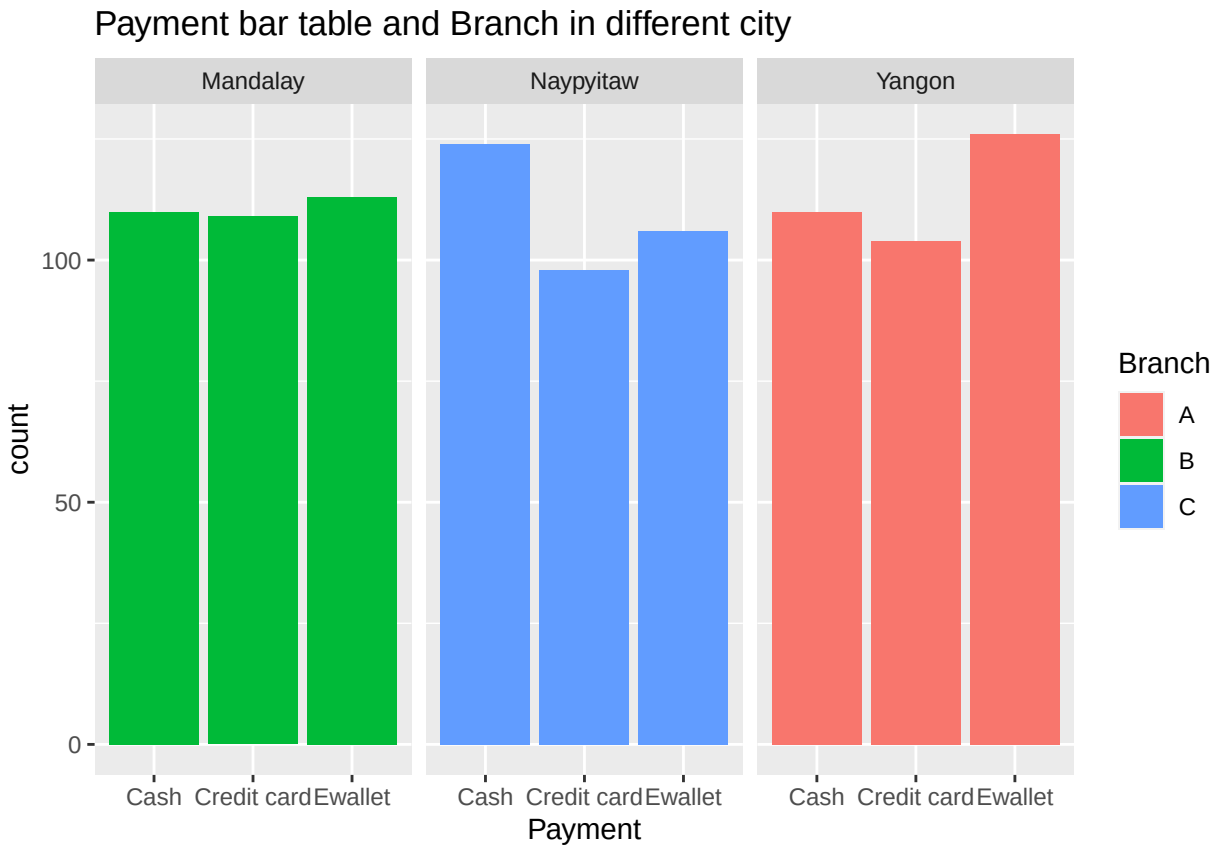
The attachment below obviously display the relationship between the city and branch. we can see the Branch A only shows in City, Yangon, Branch B in City, Mandalay, and the rest in Naypyitaw.

```
library(ggplot2)
data%>%
  ggplot(aes(Date,Total,color=Branch))+geom_point()+facet_wrap(~City)
```



The bar table shows the different payment method in secret cities. In Mandalay, the bar table shows Uniform. and Naypyitaw skewed right, and Yangon seems almost symmetric shapes

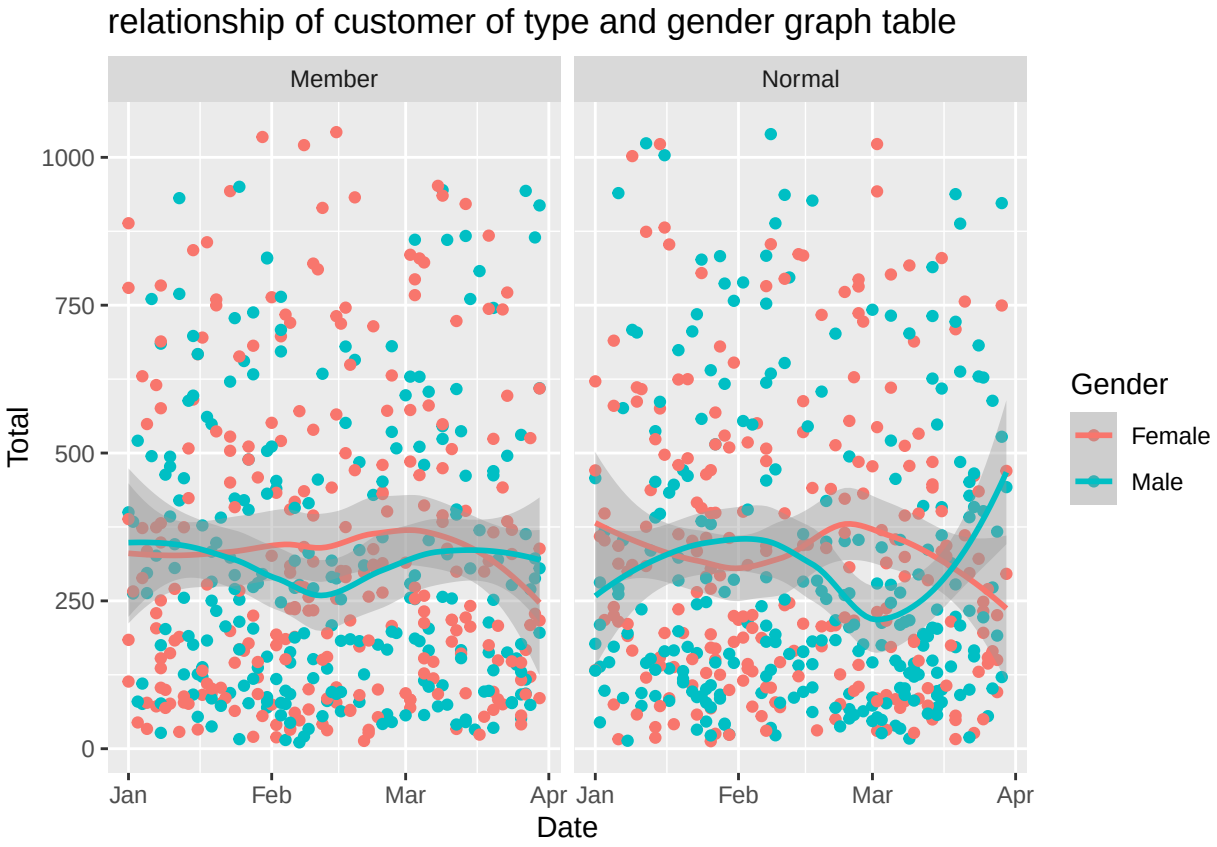
```
data%>%
  ggplot(aes(Payment,fill=Branch))+geom_bar()+ggtitle("Payment bar table and Branch in different city")+
```



The attachment is linear regression between the Gender and type of customers where they spend total with date.

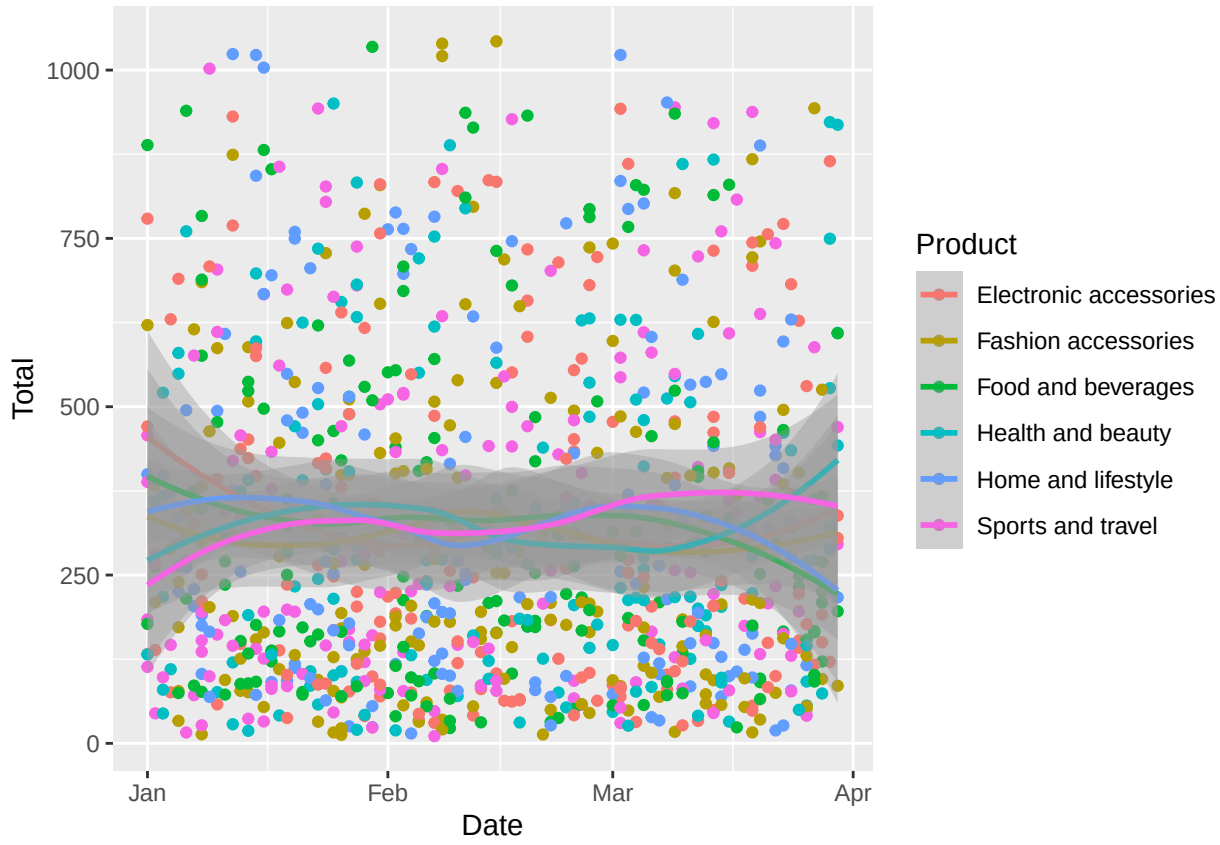
```
data%>%
  ggplot(aes(Date,Total, color=Gender))+geom_point()+geom_smooth()+facet_wrap(~Type)+ggtitle("relationsl
```

```
## `geom_smooth()` using method = 'loess' and formula 'y ~ x'
```

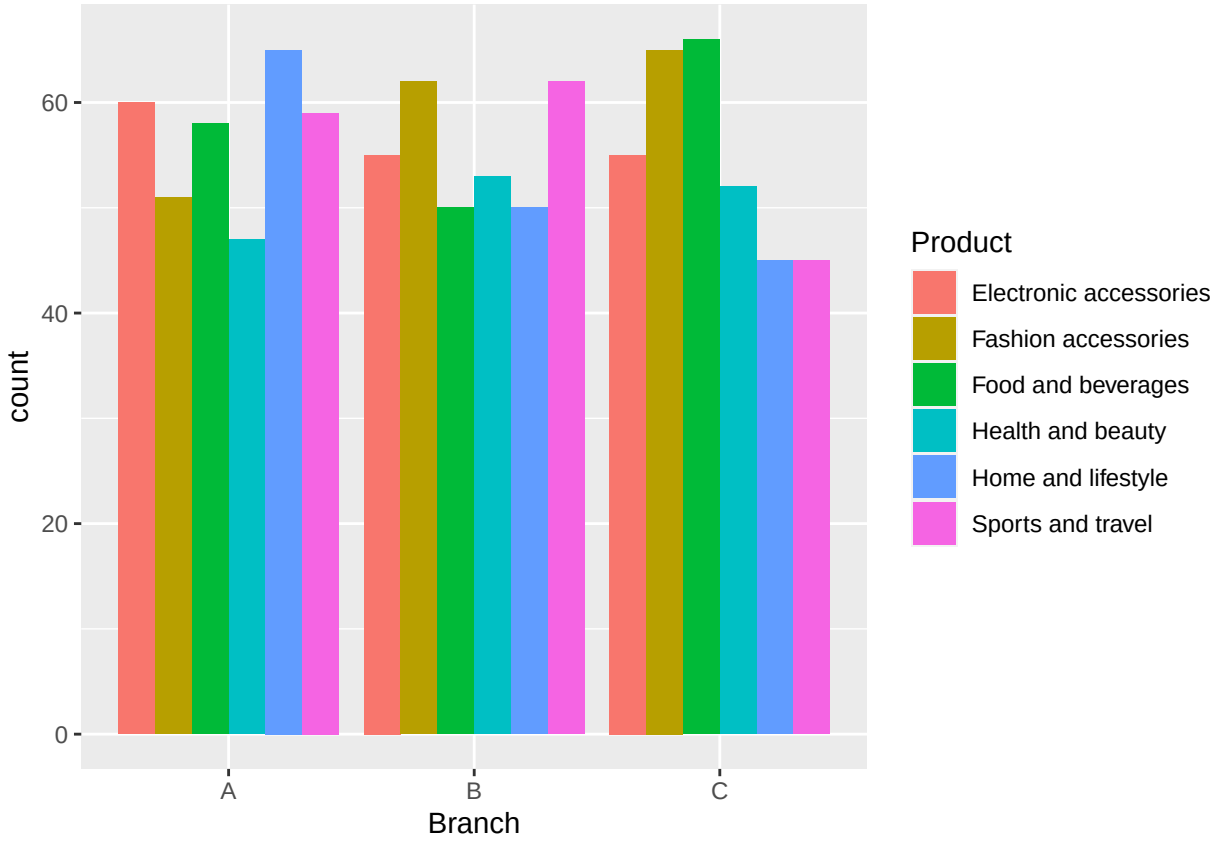



```
data%>%
  ggplot(aes(Date,Total, color=Product))+geom_point()+geom_smooth()
```

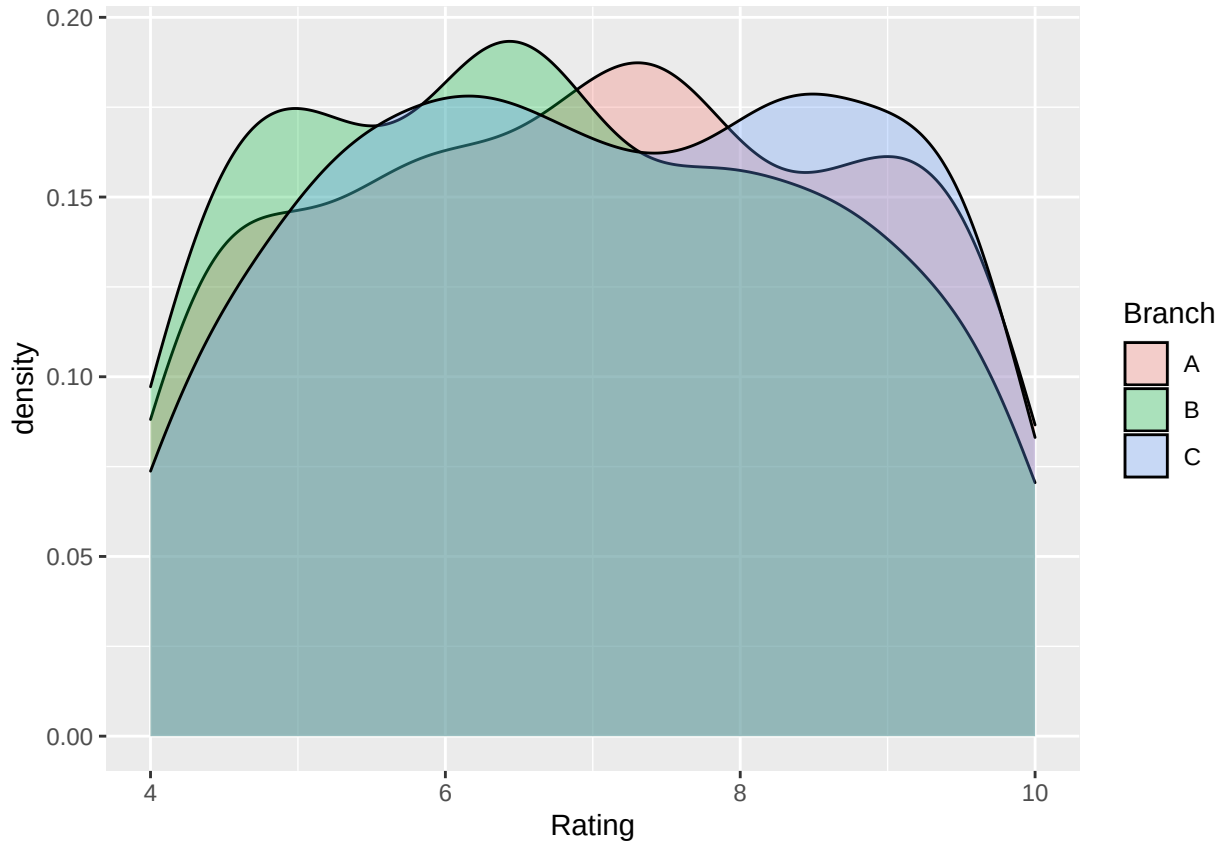
```
## `geom_smooth()` using method = 'loess' and formula 'y ~ x'
```



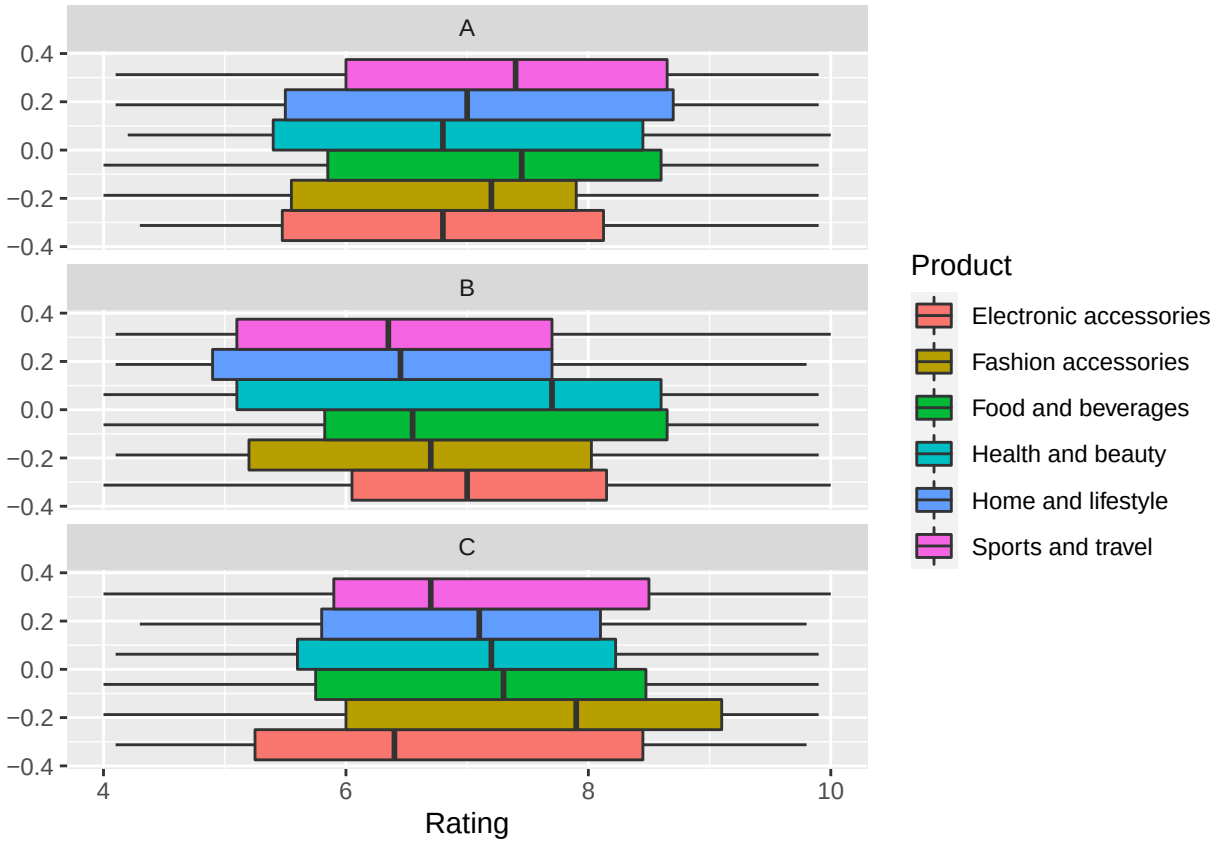

```
ggplot(data=data)+geom_bar(mapping= aes(x=Branch, fill=Product), position="dodge")
```




```
ggplot(data=data)+geom_density(mapping= aes(x=Rating,fill=Branch),alpha=0.3)
```




```
data%>%
  ggplot(aes(Rating,fill=Product))+geom_boxplot(position="dodge")+facet_wrap(~Branch,nrow = 3)
```



Import sql library

```
library(sqldf)
```

```
## Loading required package: gsubfn
```

```
## Loading required package: proto
```

```
## Warning in doTryCatch(return(expr), name, parentenv, handler): unable to load shared object '/Library
##   dlopen(/Library/Frameworks/R.framework/Resources/modules//R_X11.so, 6): Library not loaded: /opt/X
##   Referenced from: /Library/Frameworks/R.framework/Resources/modules//R_X11.so
##   Reason: image not found
```

```
## Could not load tcltk.  Will use slower R code instead.
```

```
## Loading required package: RSQLite
```

```
sqldf("select *
      from data
      group by Id
      order by Rating desc, Total desc
      limit 10")
```

```
##             Id Branch      City   Type Gender                Product Price
## 1  423-57-2993      B  Mandalay Normal   Male      Sports and travel 93.39
## 2  866-70-2814      B  Mandalay Normal Female Electronic accessories 52.79
## 3  347-34-2234      B  Mandalay Member Female      Sports and travel 55.07
## 4  725-56-0833      A    Yangon Normal Female      Health and beauty 32.32
## 5  285-68-5083      C Naypyitaw Member Female      Sports and travel 24.74
## 6  641-51-2661      C Naypyitaw Member Female     Food and beverages 87.10
## 7  109-28-2512      B  Mandalay Member Female    Fashion accessories 97.61
## 8  166-19-2553      A    Yangon Member   Male      Sports and travel 89.06
## 9  410-67-1709      A    Yangon Member Female    Fashion accessories 63.88
## 10 868-52-7573      B  Mandalay Normal Female     Food and beverages 99.69
##    Quantity     Tax    Total       Date  Time     Payment   cogs gross_percent
## 1         6 28.0170 588.3570 2020-03-27 19:18     Ewallet 560.34      4.761905
## 2        10 26.3950 554.2950 2020-02-25 11:58     Ewallet 527.90      4.761905
## 3         9 24.7815 520.4115 2020-02-03 13:40     Ewallet 495.63      4.761905
## 4        10 16.1600 339.3600 2020-02-20 16:49 Credit card 323.20      4.761905
## 5         3  3.7110  77.9310 2020-02-15 17:47 Credit card  74.22      4.761905
## 6        10 43.5500 914.5500 2020-02-12 14:45 Credit card 871.00      4.761905
## 7         6 29.2830 614.9430 2020-01-07 15:01     Ewallet 585.66      4.761905
## 8         6 26.7180 561.0780 2020-01-18 17:26        Cash 534.36      4.761905
## 9         8 25.5520 536.5920 2020-01-20 17:48     Ewallet 511.04      4.761905
## 10        5 24.9225 523.3725 2020-01-14 12:09        Cash 498.45      4.761905
##    gross_income Rating
## 1       28.0170   10.0
## 2       26.3950   10.0
## 3       24.7815   10.0
## 4       16.1600   10.0
## 5        3.7110   10.0
## 6       43.5500    9.9
## 7       29.2830    9.9
## 8       26.7180    9.9
## 9       25.5520    9.9
## 10      24.9225    9.9
```

```
sqldf(" select count(*)
      from data
      where Type == 'Member'
      ")
```

```
##   count(*)
## 1      501
```