Coffee

Prashant

17 March 2019

# install.packages("reshape")  
library(reshape)

## Warning: package 'reshape' was built under R version 3.5.3

library(ggplot2)

## Warning: package 'ggplot2' was built under R version 3.5.3

library(plyr)

## Warning: package 'plyr' was built under R version 3.5.3

##   
## Attaching package: 'plyr'

## The following objects are masked from 'package:reshape':  
##   
## rename, round\_any

library(grid)  
library(gridExtra)

## Warning: package 'gridExtra' was built under R version 3.5.3

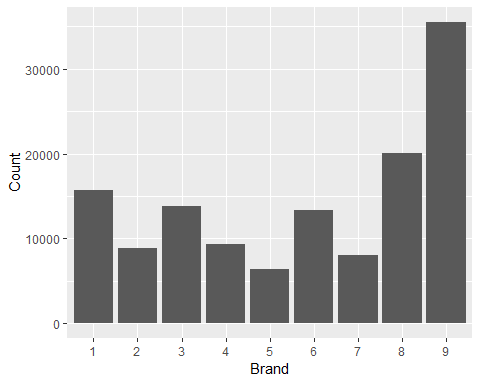
library(readxl)

## Warning: package 'readxl' was built under R version 3.5.3

coffee <- read\_excel("E:/Assignment 2/coffee.xlsx")  
View(coffee)

### Which brands of coffee are more popular? Given a brand, are all variants equally preferred?

coffee$Brand <- as.factor(coffee$Brand)  
count\_brand <- count(coffee$Brand)  
names(count\_brand)<- c("Brand", "Count")  
ggplot(count\_brand,aes(Brand,Count))+geom\_bar(stat = "Identity")



### What is the price of different coffee brands?

cost<-coffee[,c("Brand","Price\_per\_Packet")]  
cost<-count(cost, c("Brand","Price\_per\_Packet"))  
cost<-cast(cost, Brand ~ Price\_per\_Packet)

## Using freq as value column. Use the value argument to cast to override this choice

cost

## Brand 1 2 3  
## 1 1 345 12030 3287  
## 2 2 1163 5843 1866  
## 3 3 4915 8917 NA  
## 4 4 7359 2011 NA  
## 5 5 912 2627 2823  
## 6 6 1922 4302 7050  
## 7 7 1393 3496 3122  
## 8 8 2553 8568 8934  
## 9 9 9075 19580 6893

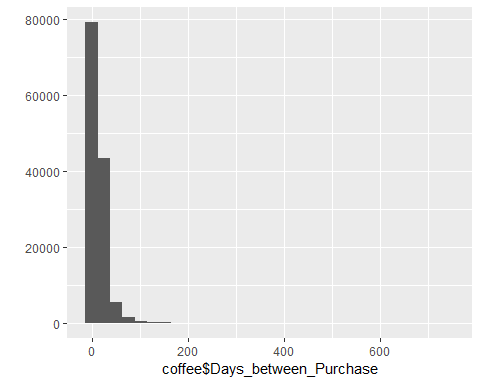
### How frequently does a household buy coffee? How many packets of coffee are bought at a time?

summary(coffee$Days\_between\_Purchase)

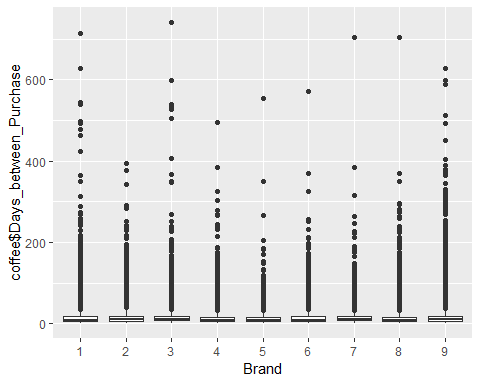
## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 1.00 6.00 9.00 15.14 17.00 741.00

qplot(coffee$Days\_between\_Purchase,data=coffee)

## `stat\_bin()` using `bins = 30`. Pick better value with `binwidth`.



ggplot(coffee,aes(Brand,coffee$Days\_between\_Purchase))+geom\_boxplot()



quantile(coffee$Days\_between\_Purchase)

## 0% 25% 50% 75% 100%   
## 1 6 9 17 741

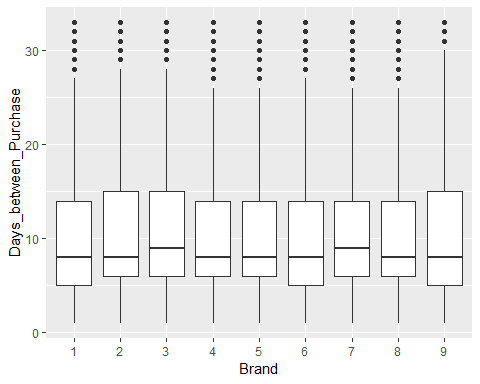
q1<-6  
q3<-17  
iqr<-q3-q1  
upper <- q3+(iqr\*1.5)  
upper

## [1] 33.5

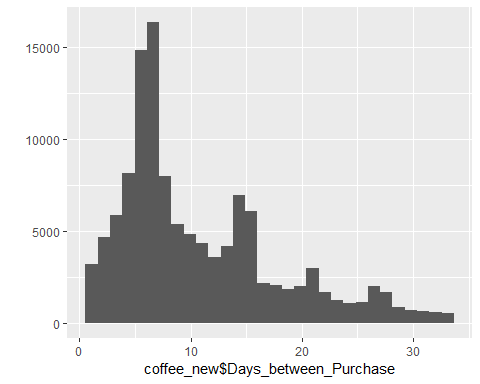
coffee\_new <-subset(coffee, Days\_between\_Purchase<upper)  
summary(coffee\_new$Days\_between\_Purchase)

## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 1.00 6.00 8.00 10.79 14.00 33.00

coffee\_new$Brand <- as.factor(coffee\_new$Brand)  
ggplot(coffee\_new, aes(Brand,Days\_between\_Purchase))+geom\_boxplot()



qplot(coffee\_new$Days\_between\_Purchase,bins=30)



# On an average most households buy coffee in 9-12 days

count(coffee$No\_of\_Packet)$freq/sum(count(coffee$No\_of\_Packet)$freq)

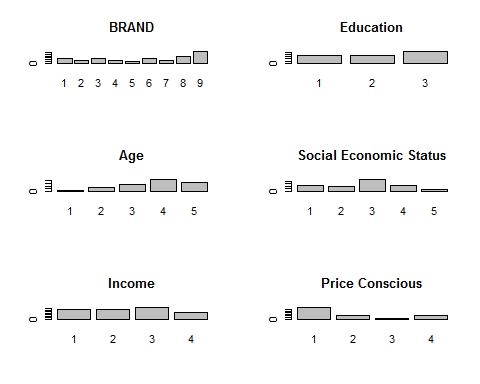
## [1] 0.80178034 0.14726001 0.05095964

#80% bought one packet at a time, 15% bought two packets while 5% bought three or more packets

### What factors have an impact on household’s coffee purchase?

brand<-(count(coffee\_new$Brand))  
edu<-(count(coffee\_new$Education))  
age<-(count(coffee\_new$Age))  
sec<-(count(coffee\_new$SEC))  
income<-(count(coffee\_new$Income))  
price\_cons<-(count(coffee\_new$Price\_Conscious))

par(mfrow=c(3,2))  
barplot(brand$freq,names.arg=brand$x,main="BRAND")  
barplot(edu$freq,names.arg=edu$x,main="Education")  
   
  
barplot(age$freq,names.arg=age$x,main="Age")  
barplot(sec$freq,names.arg=sec$x,main="Social Economic Status")  
   
  
barplot(income$freq,names.arg=income$x,main="Income")  
barplot(price\_cons$freq,names.arg=price\_cons$x,main="Price Conscious")



# Hence the factors that impact Coffee sales are Brand, Age, Socio Economic Status and Price Conscious where as   
# Education and Income do not have much impact on Coffee sales

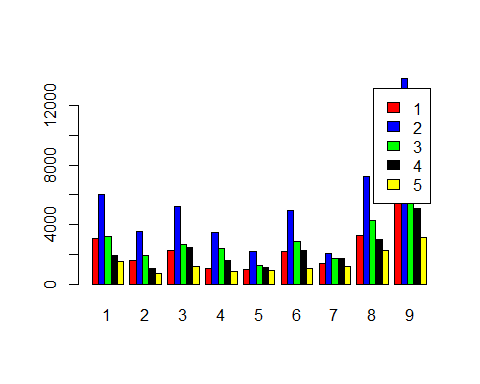
## What are the factors that have an impact on a household’s coffee purchase pattern? Does brand preference depend on household size? Does purchase depend on a person’s income or education level?

house <- cast(count(coffee[,c("Brand","Household\_Sz")], c("Brand", "Household\_Sz"))  
 ,Household\_Sz ~ Brand)

## Using freq as value column. Use the value argument to cast to override this choice

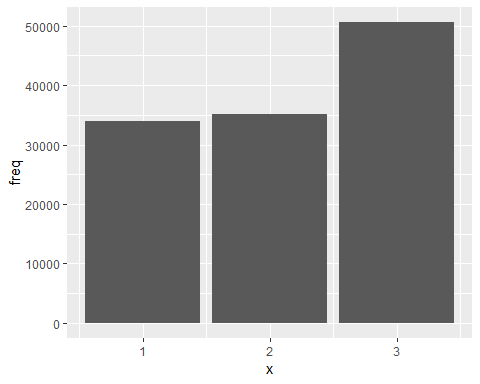
# Using freq as value column. Use the value argument to cast to override this choice

barplot((as.matrix(house)),beside = TRUE,legend.text = TRUE,col = c("red","blue","green","black","yellow"))



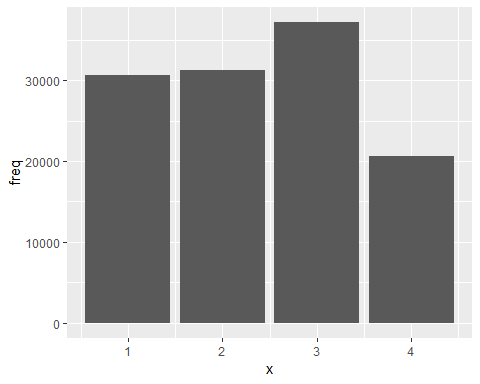
#Brand preference does not depend on house hold size

ggplot(edu, aes(x=x,y = freq)) + geom\_bar(stat = "identity")



#The purchase pattern does not depend on educaction

ggplot(income, aes(x=x,y = freq)) + geom\_bar(stat = "identity")



##### The purchase patter does not depend on income