setwd("C:/Users/12555/Desktop")

#Data tyding

a=read.csv("apple.csv")

str(a)

View(a)

# select finished data

a\_sub=subset(a,a$Finished=="True")

# select useful variables

a\_sub=subset(a\_sub,select = c(18:39))

# replace "" and "undefined" with NA

a\_sub[a\_sub==""|a\_sub=="undefined"]=NA

# change structure

f=c(1,2,3,4,6,7,9,10,11,12,13,14,15,17,19,20,21,22)

a\_sub[,f]=lapply(a\_sub[,f],as.factor)

a\_sub[,-f]=lapply(a\_sub[,-f],as.numeric)

# replace NA with the median

a\_sub$Q5[is.na(a\_sub$Q5)]=median(a\_sub$Q5,na.rm=T)

a\_sub$Q8\_1[is.na(a\_sub$Q8\_1)]=median(a\_sub$Q8\_1,na.rm=T)

a\_sub$Q13\_32[is.na(a\_sub$Q13\_32)]=median(a\_sub$Q13\_32,na.rm=T)

a\_sub$Q15[is.na(a\_sub$Q15)]=median(a\_sub$Q15,na.rm=T)

# select completed data

a\_sub=a\_sub[complete.cases(a\_sub$Q4),]

View(a\_sub)

#Analysis 1-3

library(ggplot2)

# Show the distribution of Age and Income

table(a\_sub$Q2, a\_sub$Q3)

# Analysis 1: No. of Apple鈥檚 products (Q4) and Age (Q2)

t1 <- table(a\_sub$Q2, a\_sub$Q4)

t1\_df <- as.data.frame.matrix(t1)

t1\_df["Sum",] <- colSums(t1)

t1\_df

## Null hypothesis (H0): Age and the number of Apple鈥檚 product are unrelated.

## Alternative hypothesis (H1): Age and the number of Apple鈥檚 product are related.

chisq.test(table(a\_sub$Q4, a\_sub$Q2))

## P-value > 0.05

## Not to reject H0.

## So we conclude age and the number of Apple鈥檚 product do not have a significant relationship.

## Visualize the response

ggplot(a\_sub, aes(Q4, fill=factor(Q2))) +

geom\_bar(position = "dodge", colour = "black") +

labs(x = "No. of Apple's product", y = "Count"

, title = "No. of Apple鈥檚 products and Age"

, fill = "Age Range")

## Correlation test

Age <- as.numeric(a\_sub$Q2)

ProductNo <- as.numeric(a\_sub$Q4)

## H0: True correlation is not equal to 0

cor.test(Age, ProductNo)

## P-value > 0.05

## Not to reject H0.

## So we conclude age and the number of Apple鈥檚 product do not have a significant relationship.

# Analysis 2: No. of Apple product (Q4) & Income (Q3)

unique(a\_sub$Q3)

a\_sub$Q3 <- factor(a\_sub$Q3

, levels = c("Less than 5,000HKD"

, "5,000HKD - 10,000HKD"

, "10,000HKD - 50,000HKD"))

## Visualize the response

ggplot(a\_sub, aes(Q4, fill=factor(Q3))) +

geom\_bar(position = "dodge", colour = "black") +

labs(x = "No. of Apple's product", y = "Count"

, title = "No. of Apple鈥檚 products and Income"

, fill = "Income Range")

table(a\_sub$Q3, a\_sub$Q4)

## Correlation test

Income <- as.numeric(a\_sub$Q3)

ProductNo <- as.numeric(a\_sub$Q4)

## H0: True correlation is not equal to 0

cor.test(ProductNo, Income)

## P-value > 0.05

## Not to reject H0.

## So there is no relationship between income and the no. of Apple鈥檚 product.

# Analysis 3: No. of Apple鈥檚 products (Q4) and Comments on Price (Q7)

a\_sub$Q7 <- factor(a\_sub$Q7

, levels = c("Too cheap", "Cheap"

, "Affordable", "Expensive"

, "Too expensive"))

ggplot(a\_sub, aes(Q4, fill=factor(Q7))) +

geom\_bar(position = "dodge", colour = "black") +

labs(x = "No. of Apple's product", y = "Count"

, title = "No. of Apple鈥檚 products and Price"

, fill = "Price")

## H0: Comments on Price and the number of Apple鈥檚 product are unrelated.

## H1: Comments on Price and the number of Apple鈥檚 product are related.

chisq.test(table(a\_sub$Q7, a\_sub$Q4))

## P-value > 0.05

## Not to reject H0.

## So we conclude comments on price and the number of Apple鈥檚 product do not have a significant relationship.

## Correlation test

Price <- as.numeric(a\_sub$Q7)

cor.test(ProductNo, Price)

## Polychoric Correlation

library(psych)

a\_sub$ProductNo <- ProductNo

a\_sub$Price <- Price

polychoric(a\_sub[,c("ProductNo","Price")])

## The polychoric correlation coefficient is -0.31.

#Analysis 4 and 5

library(ggplot2)

with(a\_sub,prop.table(table(ProductNo,Airdrop),margin=1))

table <- as.data.frame(table(a\_sub$Q4,a\_sub$Q12))

ggplot(table,aes(x=Var1, y= Freq, fill=Var2))+

geom\_col(position='dodge')+

labs(x="Number of Products",y="Count",fill="Airdrop")+

ggtitle("Numbers of Porducts and Views about Airdrop")

ggplot(a\_sub,aes(Q16,fill=Q4))+

geom\_bar(position='fill')+

labs(x="Attributes",fill="Number of Products")+

ggtitle("Attracting Attributes")

ggplot(a\_sub,aes(Q16))+

geom\_bar(fill="orange",colour="black")+

labs(x="Attributes")

unique(a\_sub$Q12)

Airdrop <- rep(NA, length(a\_sub$Q12))

Airdrop[a\_sub$Q12=="Definitely not"] <- 1

Airdrop[a\_sub$Q12=="Probably not"] <- 2

Airdrop[a\_sub$Q12=="Might or might not"] <- 3

Airdrop[a\_sub$Q12=="Probably yes"] <- 4

Airdrop[a\_sub$Q12=="Definitely yes"] <- 5

a\_sub$Airdrop <- Airdrop

unique(a\_sub$Q4)

ProductNo <- rep(NA, length(a\_sub$Q4))

ProductNo[a\_sub$Q4=="1"] <- 1

ProductNo[a\_sub$Q4=="2"] <- 2

ProductNo[a\_sub$Q4=="3"] <- 3

ProductNo[a\_sub$Q4=="4"] <- 4

ProductNo[a\_sub$Q4=="More than4"] <- 5

a\_sub$ProductNo <- ProductNo

df <- subset(a\_sub, as.numeric(ProductNo) & as.numeric(Airdrop))

library(psych)

polychoric(df[,c('Airdrop','ProductNo')] )

#Analysis 6

setwd("C:/Users/12555/Desktop")

library(dplyr)

library(coefplot)

data1=read.csv("apple.csv")

col1=colnames(data1)

data2=read.csv("apple.csv",header = F,skip = 3)

colnames(data2)=col1

data=data2[,c("Q4","Q8\_1","Q9\_3","Q10","Q15")]

colnames(data)=c("Number.of.apple.products","Sat\_quality","Sat\_after.sales.service",

"Sat\_design","Sat\_environment.friendly")

unique(data$Sat\_after.sales.service)

data =data %>%

dplyr::mutate(Sat\_after.sales.service = as.character(Sat\_after.sales.service)) %>%

dplyr::mutate(Sat\_after.sales.service=ifelse(Sat\_after.sales.service=="Extremely Satisfied",5,Sat\_after.sales.service)) %>%

dplyr::mutate(Sat\_after.sales.service=ifelse(Sat\_after.sales.service=="Satisfied",4,Sat\_after.sales.service)) %>%

dplyr::mutate(Sat\_after.sales.service=ifelse(Sat\_after.sales.service=="Common",3,Sat\_after.sales.service)) %>%

dplyr::mutate(Sat\_after.sales.service=ifelse(Sat\_after.sales.service=="Dissatisfied",2,Sat\_after.sales.service)) %>%

dplyr::mutate(Sat\_after.sales.service=ifelse(Sat\_after.sales.service=="Extremely Dissatisfied",1,Sat\_after.sales.service)) %>%

dplyr::mutate(Sat\_after.sales.service = as.numeric(Sat\_after.sales.service))

summary(data$Sat\_design)

data$Sat\_design=as.factor(data$Sat\_design)

data$Number.of.apple.products=as.factor(data$Number.of.apple.products)

data[data == 'Quite appealling' ] ='4'

data[data == 'Appealling' ] ='3'

data[data == 'Common' ] ='2'

data[data == 'Ugly' ]= '1'

data[data == 'More than4' ] ='5'

data$Sat\_design=as.numeric(data$Sat\_design)

data$Number.of.apple.products=as.numeric(data$Number.of.apple.products)

data=na.omit(data) # no missing values

model1=lm(Number.of.apple.products~

Sat\_quality+Sat\_after.sales.service+

Sat\_design+Sat\_environment.friendly,data=data)

summary(model1)

confint(model1)

coefplot(model1,

intercept=F,

outerCI=1.96,

lwdOuter=1.5,

ylab= "Satisfied rating of Apple",

xlab="Impact on number of owned Apple products",

title="Coefficient Plot")