Data Visualisation
Lab Experiment – 2

Name: Om Ashish Mishra
Registration Number: 16BCE0789
Slot: C2

Question 1

Problem 4

Design and implement an interactive visual idiom to help the users in choosing mobile phones with best price and features.

(What) Target - choosing mobile phone with best price and features

(Why) Tasks - Present

- Search

Set one goal for each task

i.e., present – communicate specific details precisely.

i.e., Search – browse for prices and features for a known mobile brand

Implement the goals using the following actions.

(How) Action – Encode, Navigate, Select, Arrange

CODE and OUTPUT

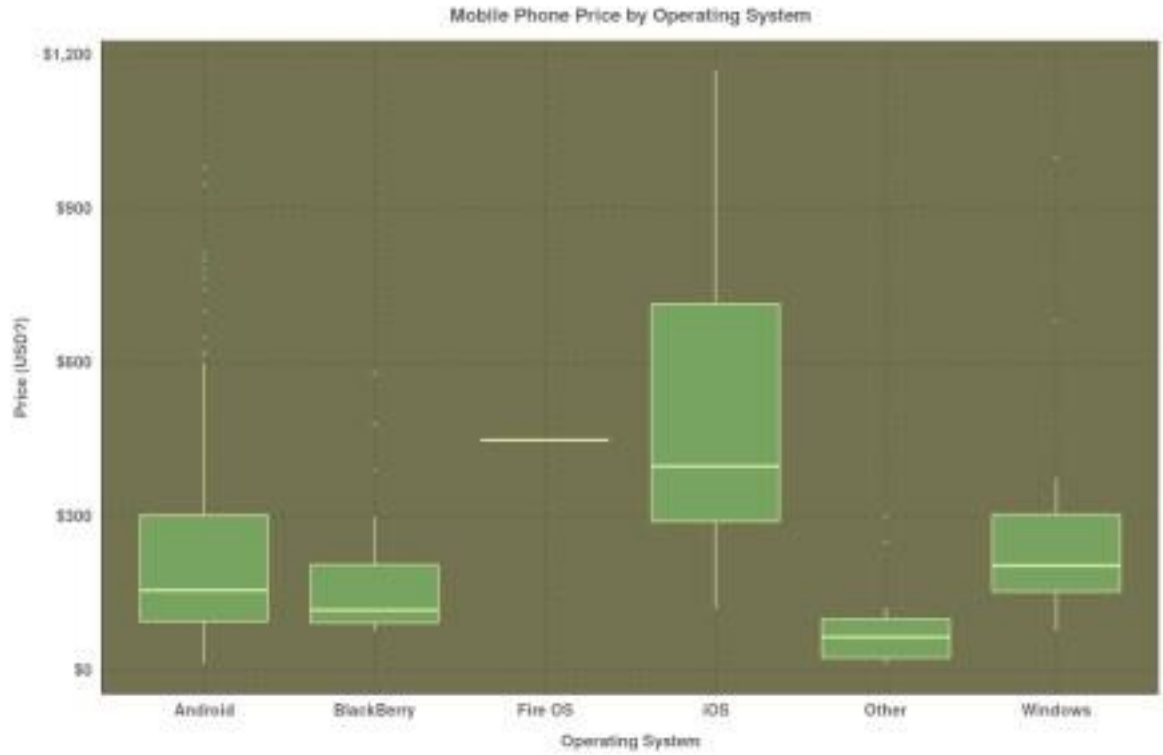
```
library(jsonlite)

cp = fromJSON(txt = "Cell Phone Data.txt", simplifyDataFrame =
TRUE) num.atts = c(4,9,11,12,13,14,15,16,18,22)

cp[,num.atts] = sapply(cp[,num.atts], function (x)
as.numeric(x)) cp$aspect.ratio = cp$att_pixels_y /
cp$att_pixels_x cp$isSmartPhone =
ifelse(grepl("smart|iphone|blackberry", cp$name,
ignore.case=TRUE) == TRUE | cp$att_screen_size >= 4,
```

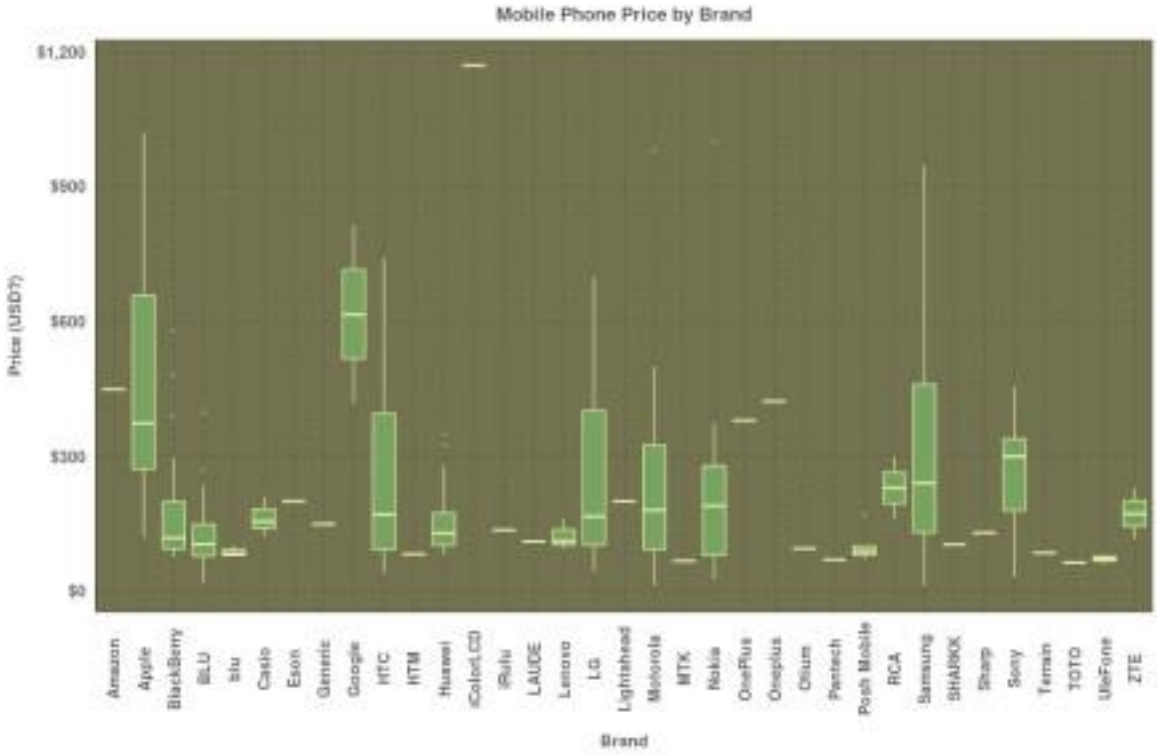
```
"Yes", "No")

library(ggplot2)

library(ggthemr)

library(scales)

ggthemr("camoflauge")


ggplot(cp, aes(x=att_brand, y=price)) + geom_boxplot() +
ggtitle("Mobile Phone Price by Brand") +
theme(axis.text.x=element_text(angle=90, size=14, vjust=0.5),
axis.text.y=element_text(size=14),
axis.title.x=element_text(size=15),
axis.title.y=element_text(size=15),
plot.title=element_text(size=17)) +
scale_y_continuous(labels=dollar, name="Price (USD?)") +
scale_x_discrete("Brand")

ggplot(cp, aes(x=att_weight, y=price)) + geom_point(size=3) +
ggtitle("Mobile Phone Price by Weight") +
theme(axis.text.x=element_text(size=14, vjust=0.5),
axis.text.y=element_text(size=14),
axis.title.x=element_text(size=15),
axis.title.y=element_text(size=15),
plot.title=element_text(size=17)) +
scale_y_continuous(labels=dollar, name="Price (USD?)") +
scale_x_continuous("Weight (oz)") + stat_smooth(se=FALSE)

ggplot(cp, aes(x=att_screen_size, y=price)) +
geom_point(size=3) + ggtitle("Mobile Phone Price by Screen
Size") + theme(axis.text.x=element_text(size=14, vjust=0.5),
axis.text.y=element_text(size=14),
axis.title.x=element_text(size=15),
axis.title.y=element_text(size=15),
plot.title=element_text(size=17)) +
scale_y_continuous(labels=dollar, name="Price (USD?)") +
scale_x_continuous("Screen Size (in)") + stat_smooth(se=FALSE)

ggplot(cp, aes(x=att_ram, y=price)) + geom_point(size=3) +
ggtitle("Mobile Phone Price by Amount of RAM") +
theme(axis.text.x=element_text(size=14, vjust=0.5),
```

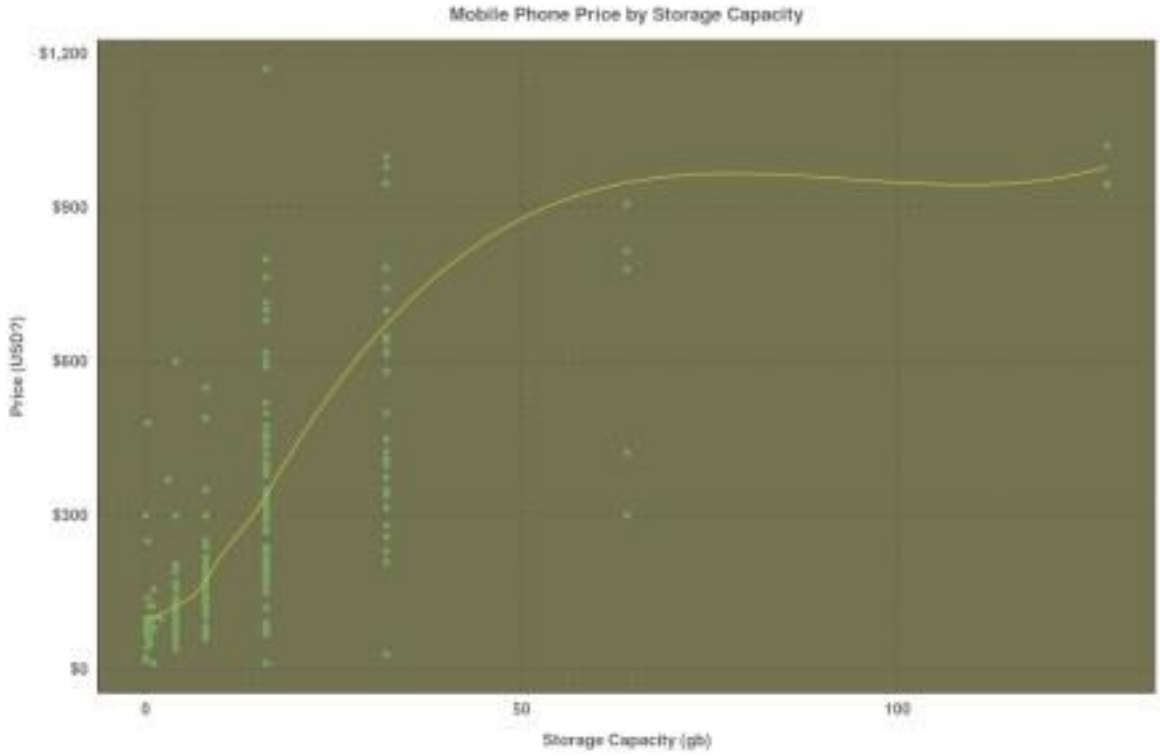
```
axis.text.y=element_text(size=14),
axis.title.x=element_text(size=15),
axis.title.y=element_text(size=15),
plot.title=element_text(size=17)) +
scale_y_continuous(labels=dollar, name="Price (USD?)") +
scale_x_continuous("RAM (gb)") + stat_smooth(se=FALSE)
ggplot(cp, aes(x=att_sd_card, y=price)) + geom_point(size=3) +
ggtitle("Mobile Phone Price by SD Card Capacity") +
theme(axis.text.x=element_text(size=14, vjust=0.5),
axis.text.y=element_text(size=14),
axis.title.x=element_text(size=15),
axis.title.y=element_text(size=15),
plot.title=element_text(size=17)) +
scale_y_continuous(labels=dollar, name="Price (USD?)") +
scale_x_continuous("SD Card Capacity (gb)") +
stat_smooth(se=FALSE)

ggplot(cp, aes(x=ifelse(cp$att_dual_sim == 1, "Yes", "No"),
y=price)) + geom_boxplot() + ggtitle("Mobile Phone Price by
Dual Sim") + theme(axis.text.x=element_text(size=14,
vjust=0.5), axis.text.y=element_text(size=14),
axis.title.x=element_text(size=15),
axis.title.y=element_text(size=15),
plot.title=element_text(size=17)) +
scale_y_continuous(labels=dollar, name="Price (USD?)") +
scale_x_discrete("Has Dual Sim Card?")

ggplot(cp, aes(x=att_storage, y=price)) + geom_point(size=3) +
ggtitle("Mobile Phone Price by Storage Capacity") +
theme(axis.text.x=element_text(size=14, vjust=0.5),
axis.text.y=element_text(size=14),
axis.title.x=element_text(size=15),
axis.title.y=element_text(size=15),
plot.title=element_text(size=17)) +
scale_y_continuous(labels=dollar, name="Price (USD?)") +
scale_x_continuous("Storage Capacity (gb)") +
stat_smooth(se=FALSE)

ggplot(cp, aes(x=att_battery_mah, y=price)) +
geom_point(size=3) + ggtitle("Mobile Phone Price by Battery
Capacity") + theme(axis.text.x=element_text(size=14,
vjust=0.5), axis.text.y=element_text(size=14),
axis.title.x=element_text(size=15),
axis.title.y=element_text(size=15),
```

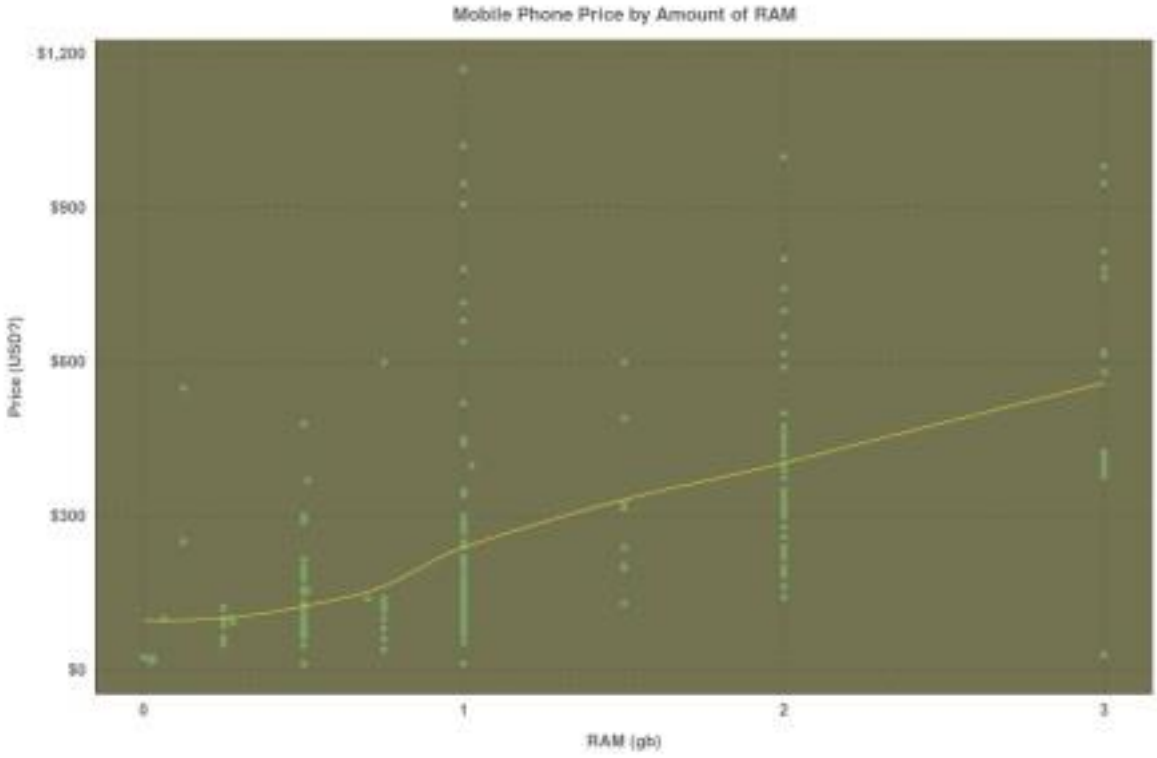
```
plot.title=element_text(size=17)) +
scale_y_continuous(labels=dollar, name="Price (USD?)") +
scale_x_continuous("Battery Capacity (mAh)") +
stat_smooth(se=FALSE)

ggplot(cp, aes(x=aspect.ratio, y=price)) + geom_point(size=3)
+ ggtitle("Mobile Phone Price by Aspect Ratio") +
theme(axis.text.x=element_text(size=14, vjust=0.5),
axis.text.y=element_text(size=14),
axis.title.x=element_text(size=15),
axis.title.y=element_text(size=15),
plot.title=element_text(size=17)) +
scale_y_continuous(labels=dollar, name="Price (USD?)") +
scale_x_continuous("Aspect Ratio (Y Pixels / X Pixels)") +
stat_smooth(se=FALSE)

ggplot(cp, aes(x=isSmartPhone, y=price)) + geom_boxplot() +
ggtitle("Mobile Phone Price by Smart Phone Status") +
theme(axis.text.x=element_text(size=14, vjust=0.5),
axis.text.y=element_text(size=14),
axis.title.x=element_text(size=15),
axis.title.y=element_text(size=15),
plot.title=element_text(size=17)) +
scale_y_continuous(labels=dollar, name="Price (USD?)") +
scale_x_discrete("Is it a Smart Phone?")

ggplot(cp, aes(x=att_os, y=price)) + geom_boxplot() +
ggtitle("Mobile Phone Price by Operating System") +
theme(axis.text.x=element_text(size=14, vjust=0.5),
axis.text.y=element_text(size=14),
axis.title.x=element_text(size=15),
axis.title.y=element_text(size=15),
plot.title=element_text(size=17)) +
scale_y_continuous(labels=dollar, name="Price (USD?)") +
scale_x_discrete("Operating System") library(caret)
control = trainControl(method="cv")
in_train = createDataPartition(cp$price, p=.8, list=FALSE)
model.gbm = train(price ~ att_brand + att_weight +
att_screen_size + att_ram + att_sd_card
+ att_dual_sim + att_storage +
att_battery_mah + att_os, data=cp,
method="gbm", trControl=control, verbose=FALSE,
subset=in_train) cp$att_brand = factor(cp$)
```

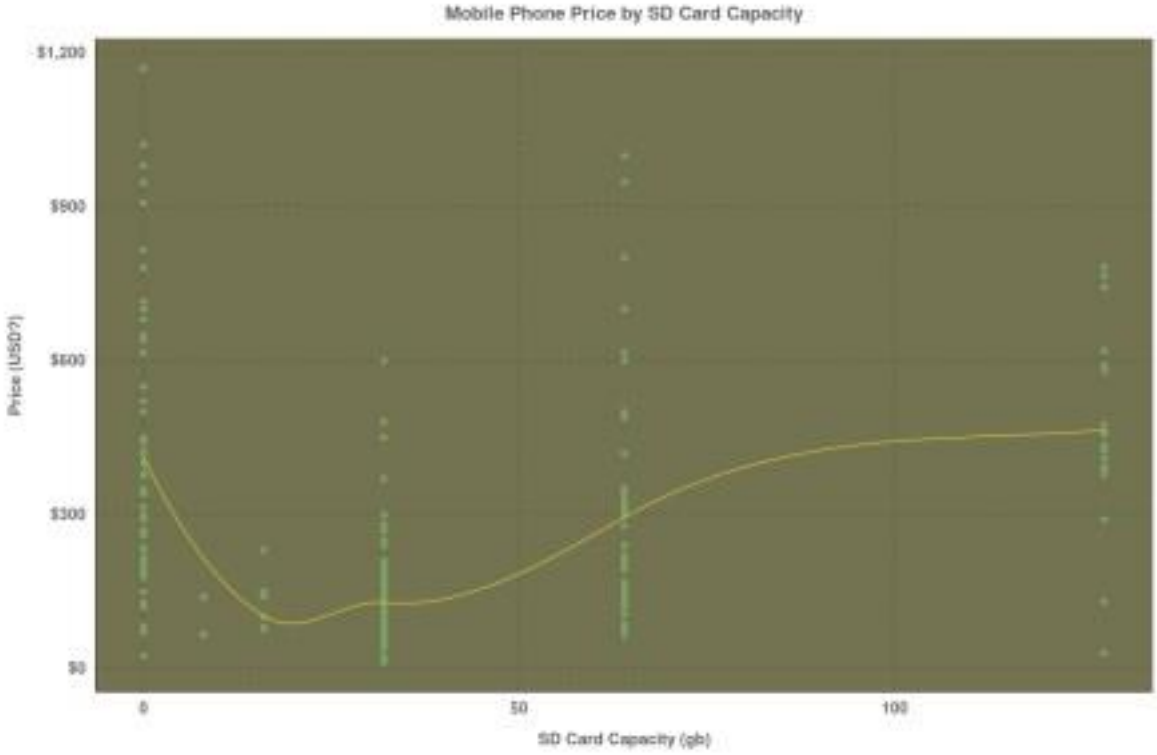
```
cp.test = cp[-in_train,] cp.test =
subset(cp.test, att_brand != 'TOTO') cp.test =
na.omit(cp.test)
cp.test$pred.price = predict(model.gbm, cp.test)

ggplot(cp.test, aes(x=pred.price, y=price)) +
geom_point(size=3) + ggtitle("Mobile Phone Price by Predicted
Price") + theme(axis.text.x=element_text(size=14, vjust=0.5),
axis.text.y=element_text(size=14),
axis.title.x=element_text(size=15),
axis.title.y=element_text(size=15),
plot.title=element_text(size=17)) +
scale_y_continuous(labels=dollar, name="Price (USD?)") +
scale_x_continuous("Predicted Price", labels=dollar) +
geom_abline(intercept=0, slope=1, colour="yellow") +
stat_smooth(se=FALSE)
```




Mobile Phone Price by Brand
$1,200
$900
$600
$300
$0
Price (USD?)
Amazon
Apple
BlackBerry
BLU
Ivlu
Casio
Eson
Generic
Google
HTC
HTM
Huawei
iColorLCD
iRulu
LAUDE
Lenovo
LG
Lightshead
Motorola
MTK
Nokia
OnePlus
Oneplus
Olium
Pantech
Posh Mobile
RCA
Samsung
SHARKK
Sharp
Sony
Terrain
TOTO
UleFone
ZTE
Brand

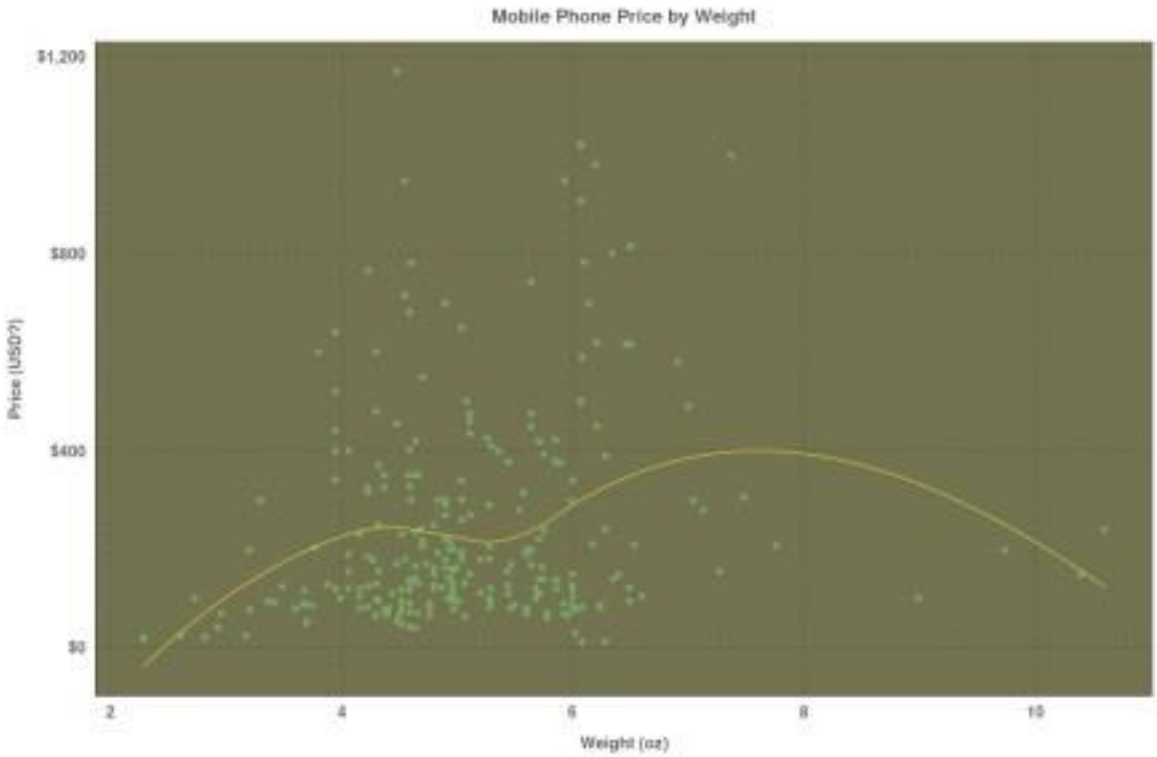

Mobile Phone Price by Storage Capacity
$1,200
$900
$600
$300
$0
Price (USD?)
0
50
100
Storage Capacity (gb)


Mobile Phone Price by Amount of RAM
$1,200
$900
$600
$300
$0
Price (USD?)
0
1
2
3
RAM (gb)


Mobile Phone Price by SD Card Capacity
$1,200
$900
$600
$300
$0
Price (USD?)
0
50
100
SD Card Capacity (gb)


Mobile Phone Price by Screen Size
$1,200
$900
$600
$300
$0
Price (USD)
2
3
4
5
6
7
Screen Size (in)


Mobile Phone Price by Battery Capacity
$1,200
$900
$600
$300
$0
Price (USD)
1000
2000
3000
4000
5000
Battery Capacity (mAh)



RESULT

Importance of each feature in the dataset

gbm variable importance

only 20 most important variables shown (out of 41)

	Overall
att_storage	100.0000
att_battery_mah	59.7597
att_weight	46.5410
att_ram	27.5871
att_osiOS	26.9977
att_screen_size	21.1106
att_sd_card	20.1130
att_brandSamsung	9.1220

Question 2

Problem 6

Generate a Node Edge Diagram (directed) for the data given in the table below. The attributes : Relationship, Age and Lives in should be displayed when there is a mouse click on each node. Use different shapes to differentiate gender. Also use similar node colors for the same places in which persons are living.

Source	Destination	Relationship	Age	Lives in
A	B	Daughter	75	Coimbatore
A	C	Son	78	Tiruppur
B	D	Son	55	Madurai
B	E	Son	52	Madurai
C	C1	Daughter	56	Shankarankovil
C	C2	Daughter	50	Coimbatore
D	D1	Daughter	25	Coimbatore
D	D2	Daughter	18	Madurai
E	E1	Daughter	20	Tiruppur
E	E2	Son	23	Nagarkovil
D1	F	Son	2	Thiruvarur

CODE and OUTPUT

```
data = read.csv("data.csv",header = TRUE)

Adata = data.frame(Node = 'A', Age = 100, Lives.in =
 Coimbatore ')
 vertices = data.frame(Node = data$Destination, Age = data$Age,
'


Lives.in = data$Lives.in)
vertices = rbind(Adata,vertices)

library(visNetwork) nodes = data.frame(id = vertices$Node,
label= vertices$Node) nodes$title = vertices$Age
nodes$color.background = c("slategrey", "tomato",
"gold","orange","red","green")[as.factor(vertices$Lives.in)]
nodes$title = paste0("<p>Age:", vertices$Age,"<br>Lives in :",
vertices$Lives.in ,"</p>")
```

```
links = data.frame(from = data$Source, to = data$Destination,
label = data$Relationship) visNetwork(nodes, links) %>%
visEdges(arrows = 'to')
```