HM3_r

2025-03-12

Homework N2

```
library(ggplot2)
library(dplyr)
```

```
## Warning: package 'dplyr' was built under R version 4.4.2
```

```
##
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':
##
##     filter, lag
```

```
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
library(tidyr) #For prepping data for plots
```

```
## Warning: package 'tidyr' was built under R version 4.4.2
```

```
library(reshape2) #For prepping data for correlation heat maps
```

```
## Warning: package 'reshape2' was built under R version 4.4.2
```

```
##
## Attaching package: 'reshape2'
```

```
## The following object is masked from 'package:tidyr':
##
##     smiths
```

```
df_phones <- read.csv(".\\mobiles_dataset.csv")
head(df_phones)
```

```
##   Company.Name           Model.Name Mobile.Weight RAM Front.Camera Back.Camera
## 1        Apple      iPhone 16 128GB          174g 6GB         12MP        48MP
## 2        Apple      iPhone 16 256GB          174g 6GB         12MP        48MP
## 3        Apple      iPhone 16 512GB          174g 6GB         12MP        48MP
## 4        Apple iPhone 16 Plus 128GB          203g 6GB         12MP        48MP
## 5        Apple iPhone 16 Plus 256GB          203g 6GB         12MP        48MP
## 6        Apple iPhone 16 Plus 512GB          203g 6GB         12MP        48MP
##    Processor Battery.Capacity.mAh Screen.Size.inches
## 1 A17 Bionic                3600                6.1
## 2 A17 Bionic                3600                6.1
## 3 A17 Bionic                3600                6.1
## 4 A17 Bionic                4200                6.7
## 5 A17 Bionic                4200                6.7
```

```
## 6 A17 Bionic                 4200                6.7
##   Launched.Price.Pakistan.PKR Launched.Price.India.INR Launched.Price.China.CNY
## 1                      224999                   79999                    5799
## 2                      234999                   84999                    6099
## 3                      244999                   89999                    6499
## 4                      249999                   89999                    6199
## 5                      259999                   94999                    6499
## 6                      274999                  104999                    6999
##   Launched.Price.USA.USD Launched.Price.Dubai.AED Launched.Year
## 1                    799                    2799          2024
## 2                    849                    2999          2024
## 3                    899                    3199          2024
## 4                    899                    3199          2024
## 5                    949                    3399          2024
## 6                    999                    3599          2024
```

Part 1: Analytical Questions (Python & R) (I am just doing the graphs for this part explanations are in Python)

Data Preperations

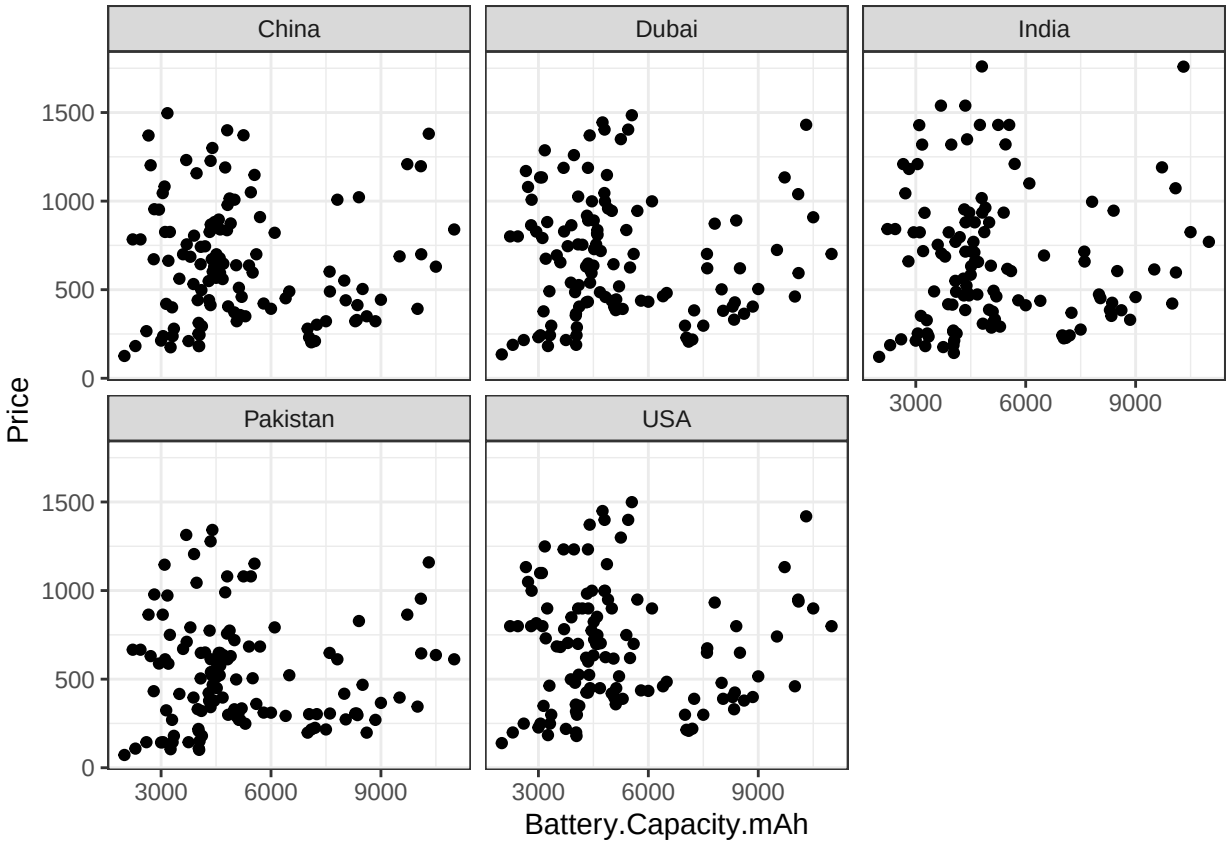
```
df_preped <- df_phones %>% mutate(Launched.Price.Pakistan.PKR = Launched.Price.Pakistan.PKR * 0.0036,
                    Launched.Price.India.INR = Launched.Price.India.INR *  0.011,
                    Launched.Price.China.CNY = Launched.Price.China.CNY * 0.14,
                    Launched.Price.Dubai.AED = Launched.Price.Dubai.AED * 0.27)
#Fixing the POCO issue
df_preped <- df_preped %>% mutate(Company.Name = ifelse(Company.Name == 'Poco', 'POCO', Company.Name))
df_preped <- df_preped %>% filter(RAM != '8GB / 12GB') #Removing the two annoying observations
df_preped$RAM = substr(df_preped$RAM,1,nchar(df_preped$RAM)-2) #Removing the GB part from the RAM
df_preped <- df_preped %>% mutate(RAM = as.numeric(RAM)) #Converting RAM to int
head(df_preped)
```

```
##   Company.Name           Model.Name Mobile.Weight RAM Front.Camera Back.Camera
## 1        Apple      iPhone 16 128GB          174g   6         12MP        48MP
## 2        Apple      iPhone 16 256GB          174g   6         12MP        48MP
## 3        Apple      iPhone 16 512GB          174g   6         12MP        48MP
## 4        Apple iPhone 16 Plus 128GB          203g   6         12MP        48MP
## 5        Apple iPhone 16 Plus 256GB          203g   6         12MP        48MP
## 6        Apple iPhone 16 Plus 512GB          203g   6         12MP        48MP
##    Processor Battery.Capacity.mAh Screen.Size.inches
## 1 A17 Bionic                3600                6.1
## 2 A17 Bionic                3600                6.1
## 3 A17 Bionic                3600                6.1
## 4 A17 Bionic                4200                6.7
## 5 A17 Bionic                4200                6.7
## 6 A17 Bionic                4200                6.7
##   Launched.Price.Pakistan.PKR Launched.Price.India.INR Launched.Price.China.CNY
## 1                   809.9964                  879.989                   811.86
## 2                   845.9964                  934.989                   853.86
## 3                   881.9964                  989.989                   909.86
## 4                   899.9964                  989.989                   867.86
## 5                   935.9964                 1044.989                   909.86
## 6                   989.9964                 1154.989                   979.86
##   Launched.Price.USA.USD Launched.Price.Dubai.AED Launched.Year
```

```
## 1                 799                755.73          2024
## 2                 849                809.73          2024
## 3                 899                863.73          2024
## 4                 899                863.73          2024
## 5                 949                917.73          2024
## 6                 999                971.73          2024
```

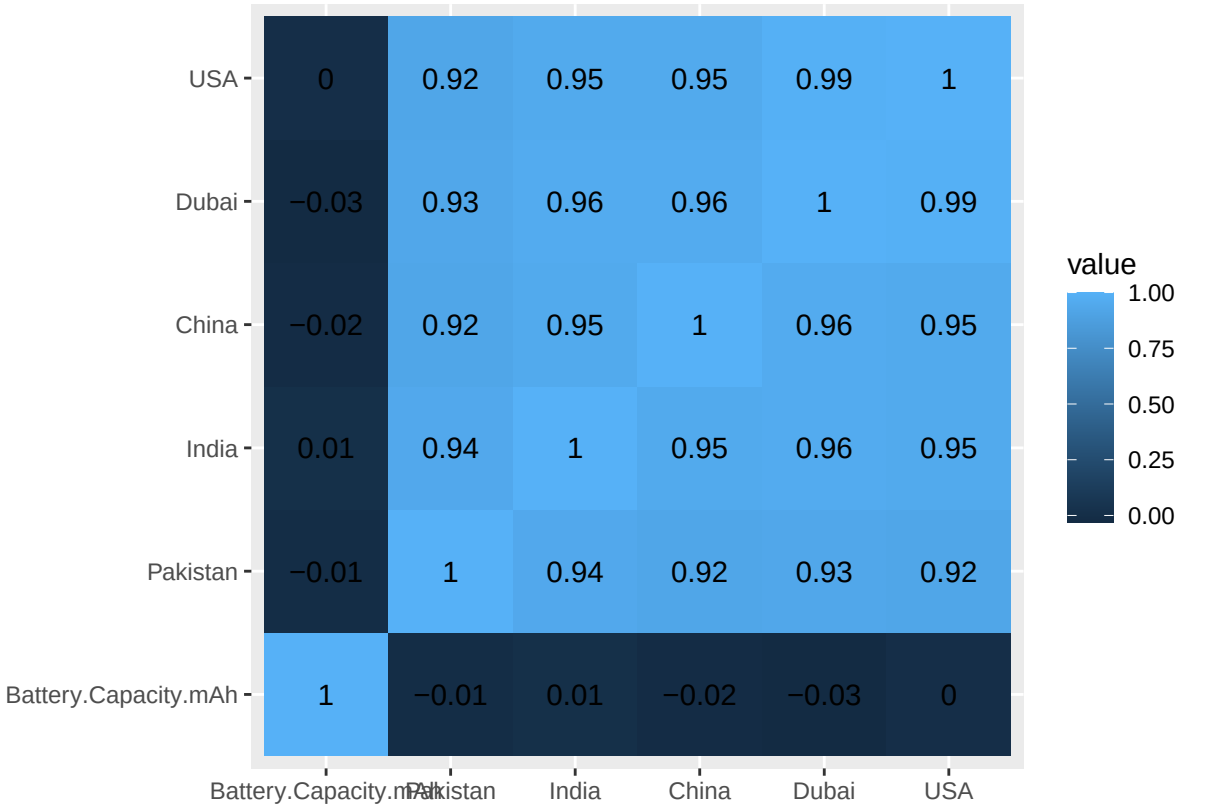
1. Does battery capacity influence the launched price of a smartphone? Check this variability across all currencies. Is there any type of difference between behaviors?

```
grouped_by_battery <- df_preped %>% group_by(Battery.Capacity.mAh) %>%
  summarise(Pakistan = mean(Launched.Price.Pakistan.PKR), India = mean(Launched.Price.India.INR),
            China = mean(Launched.Price.China.CNY), Dubai = mean(Launched.Price.Dubai.AED),
            USA = mean(Launched.Price.USA.USD))

grouped_by_battery_longwise <- grouped_by_battery %>%
  pivot_longer(!Battery.Capacity.mAh, names_to = "Country", values_to = "Price")
ggplot(grouped_by_battery_longwise, aes(x = Battery.Capacity.mAh, y = Price)) +
  geom_point() +
  facet_wrap(~Country) +
  theme_bw()
```



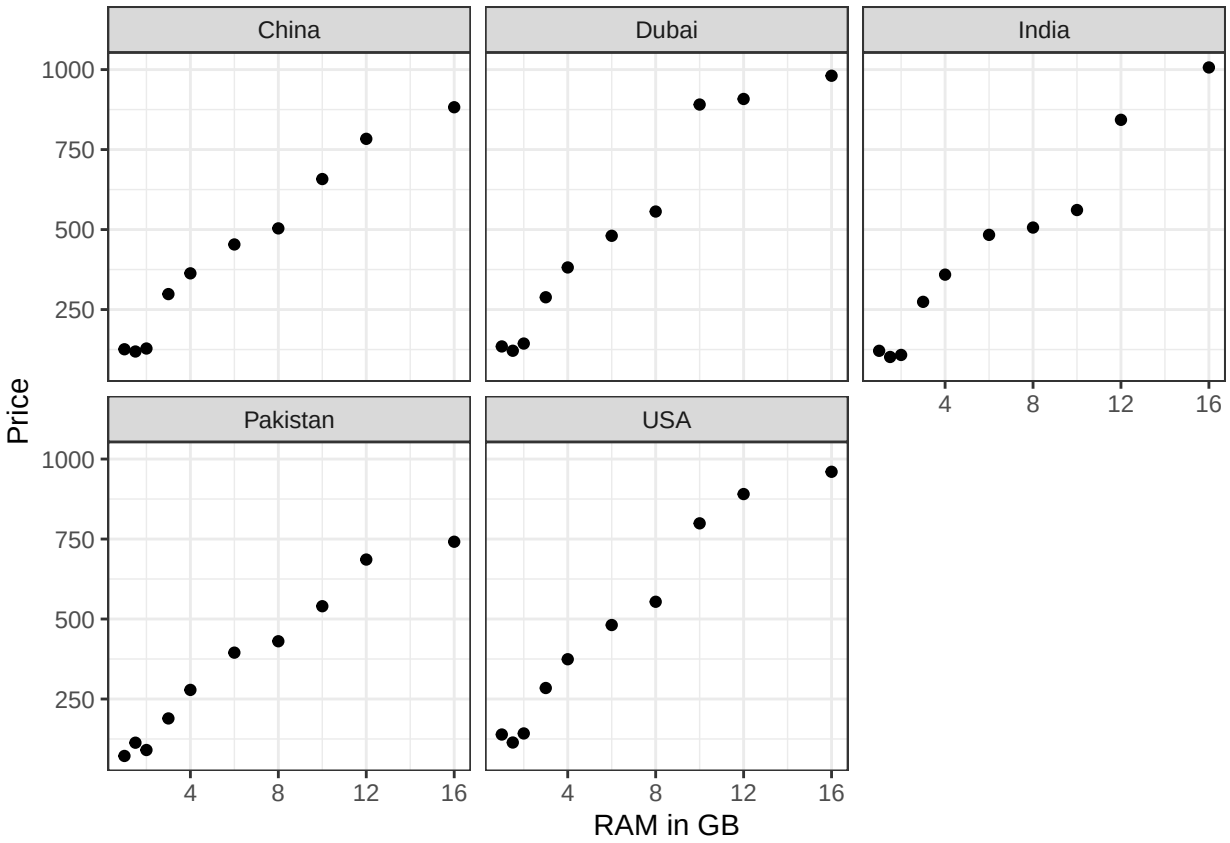

```
cormat <- round(cor(grouped_by_battery),2)
melted <- melt(cormat)
ggplot(melted, aes(x=Var1, y=Var2, fill=value)) + geom_tile() +
  geom_text(aes(Var2, Var1, label = value), color = "black", size = 4) +
```

```
  labs(x='',y='')
```

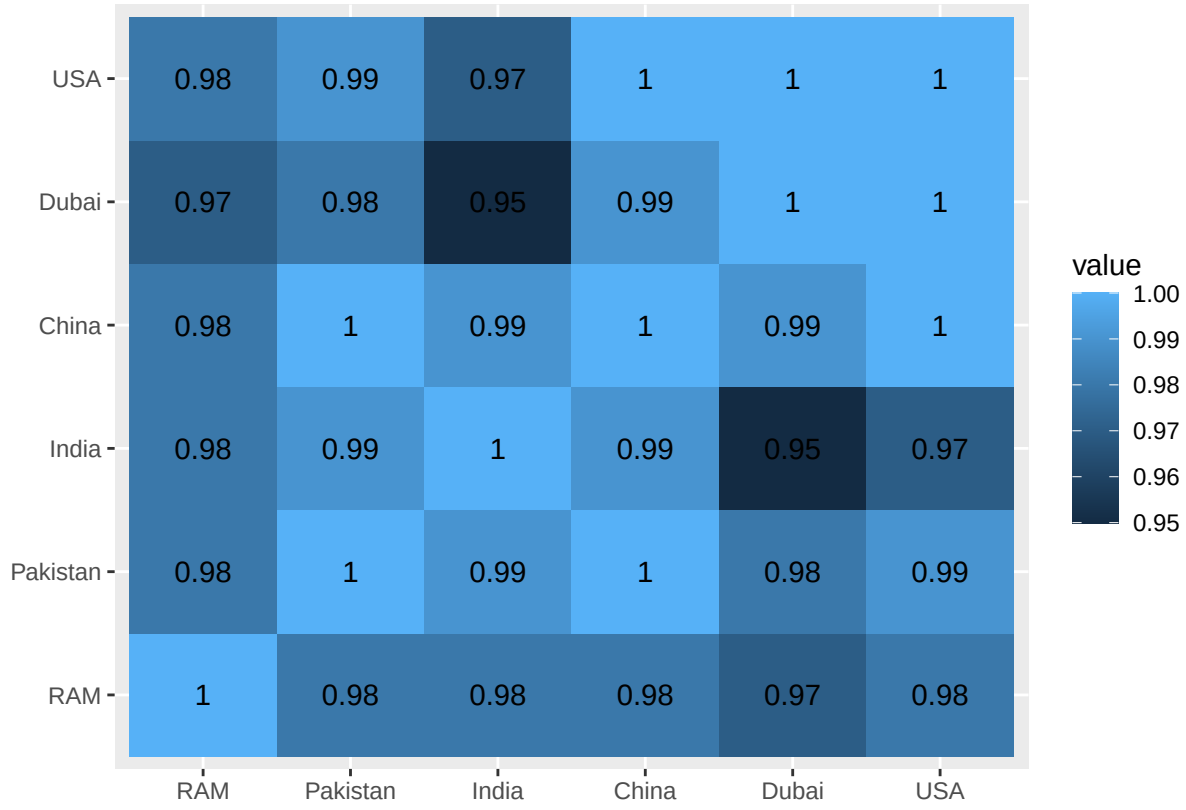


2. Does RAM size impact the price of smartphones? Check this variability across all currencies. Is there any type of difference between behaviors?

```
grouped_by_RAM <- df_preped %>% group_by(RAM) %>%
  summarise(Pakistan = mean(Launched.Price.Pakistan.PKR), India = mean(Launched.Price.India.INR),
            China = mean(Launched.Price.China.CNY), Dubai = mean(Launched.Price.Dubai.AED),
            USA = mean(Launched.Price.USA.USD))

grouped_by_RAM_longwise <- grouped_by_RAM %>%
  pivot_longer(!RAM, names_to = "Country", values_to = "Price")
ggplot(grouped_by_RAM_longwise, aes(x = RAM, y = Price)) +
  geom_point() +
  facet_wrap(~Country) +
  theme_bw() +
  labs(x = 'RAM in GB')
```

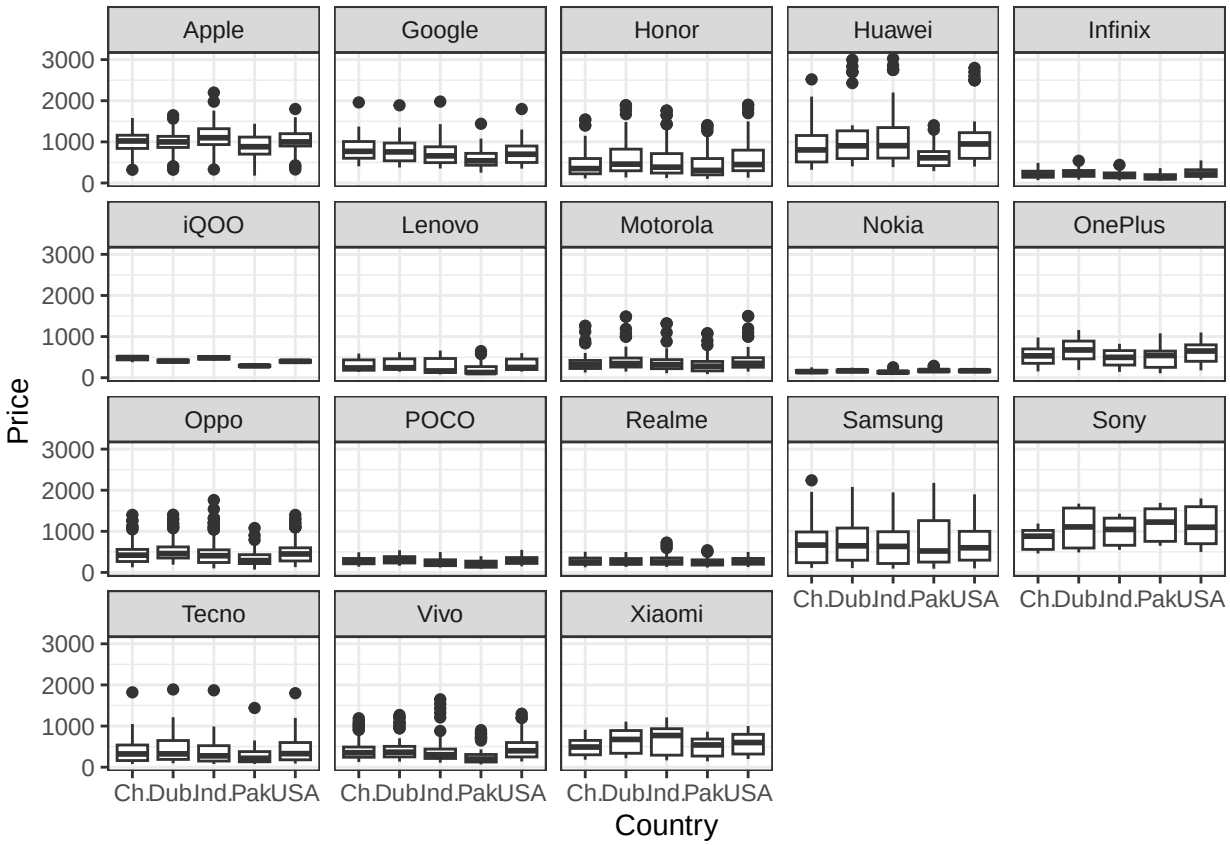



```
cormat <- round(cor(grouped_by_RAM),2)
melted <- melt(cormat)
ggplot(melted, aes(x=Var1, y=Var2, fill=value)) + geom_tile() +
  geom_text(aes(Var2, Var1, label = value), color = "black", size = 4) +
  labs(x='',y='')
```



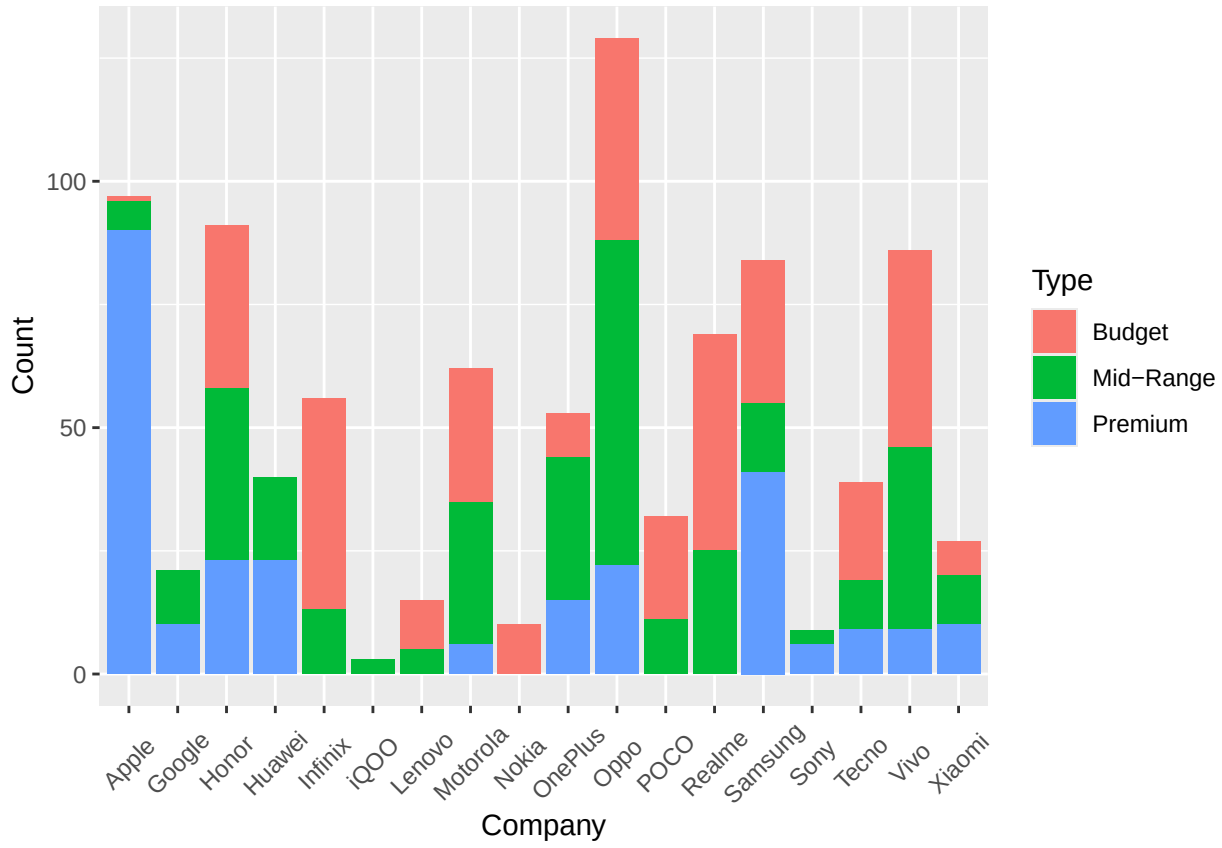
3. Do Apple devices have a higher price variation across different regions compared to other brands? In which country do Apple devices have the highest markup? Are there brands with more stable pricing across regions?

```
pain <- df_preped %>% select(Company.Name, Launched.Price.Pakistan.PKR, Launched.Price.India.INR,
                             Launched.Price.China.CNY, Launched.Price.USA.USD, Launched.Price.Dubai.AED)
pain <- pain %>% rename(Pak. = Launched.Price.Pakistan.PKR, Ind. = Launched.Price.India.INR,
                        Ch. = Launched.Price.China.CNY, USA = Launched.Price.USA.USD,
                        Dub. = Launched.Price.Dubai.AED)
pain_longwise <- pain %>%
  pivot_longer(!Company.Name, names_to = "Country", values_to = "Price")

ggplot(pain_longwise, aes(x=Country, y=Price)) + geom_boxplot() + facet_wrap(~Company.Name) + theme_bw()
```



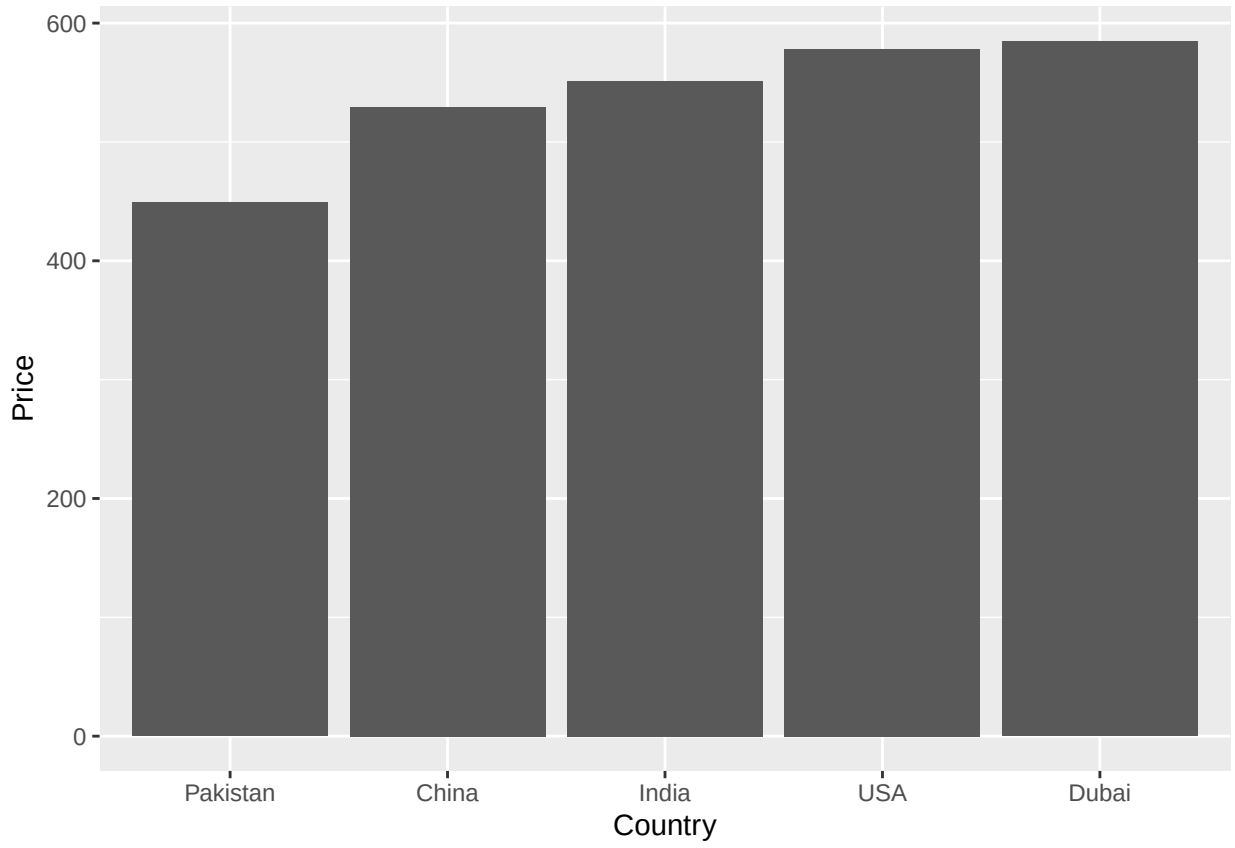
4. Do all smartphone brands have flagship and budget-friendly models, or do some brands only focus on premium devices? Check how many models each brand has in each segment. Determine whether a brand covers all three segments or focuses only on premium/mid-range.

```
df_preped <- df_preped %>% mutate(Avg.Price = (Launched.Price.Pakistan.PKR + Launched.Price.India.INR +
                                  + Launched.Price.USA.USD + Launched.Price.Dubai.AED) / 5)
df_preped <- df_preped %>% mutate(Device.Type = ifelse(Avg.Price < 300, 'Budget', ifelse(Avg.Price < 70
                                                                                        'Mid-Range', '
grouped_by_Comp_and_Device_Type <- df_preped %>% count(Company.Name,Device.Type,.drop=FALSE)

ggplot(grouped_by_Comp_and_Device_Type, aes(fill=Device.Type, y=n, x=Company.Name)) +
  geom_bar(position="stack", stat="identity") +
  labs(fill = 'Type', x = 'Company', y='Count') +
  theme(
    axis.text.x = element_text(angle = 45, vjust = 0.5)
  )
```



5. Which region offers the most affordable smartphone prices on average? Are there any brands that price their phones significantly lower in one region compared to others?

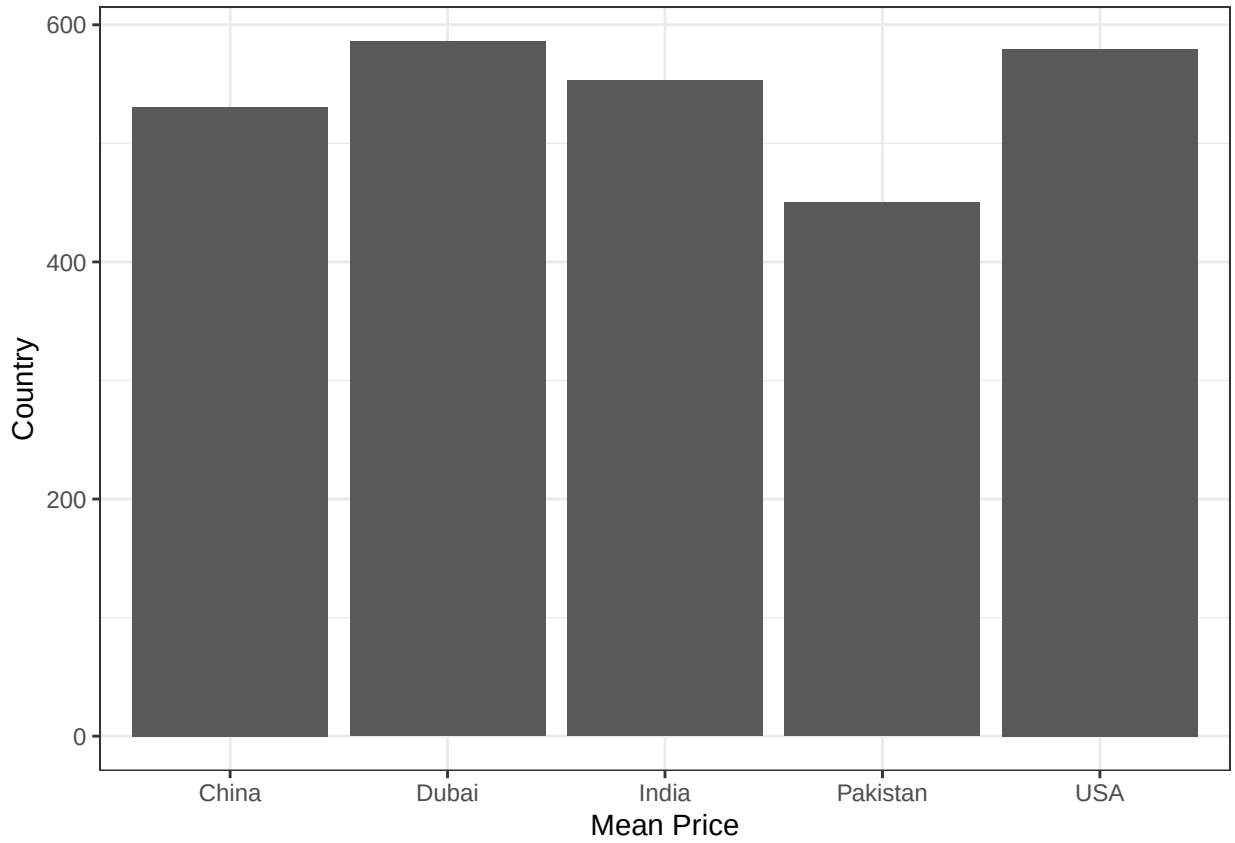
```
df_last <- df_preped %>% summarise(Pakistan = mean(Launched.Price.Pakistan.PKR), India = mean(Launched.
            China = mean(Launched.Price.China.CNY), Dubai = mean(Launched.Price.Dubai.AED),
            USA = mean(Launched.Price.USA.USD))
df_last_longwise <- df_last %>%
  pivot_longer(c(Pakistan, India, China, Dubai, USA),names_to = "Country", values_to = "Price") %>% arra

ggplot(df_last_longwise, aes(x=reorder(Country, Price), y=Price)) + geom_bar(stat='identity') +
  labs(x = 'Country')
```



Part 2: Visualization (Python & R)

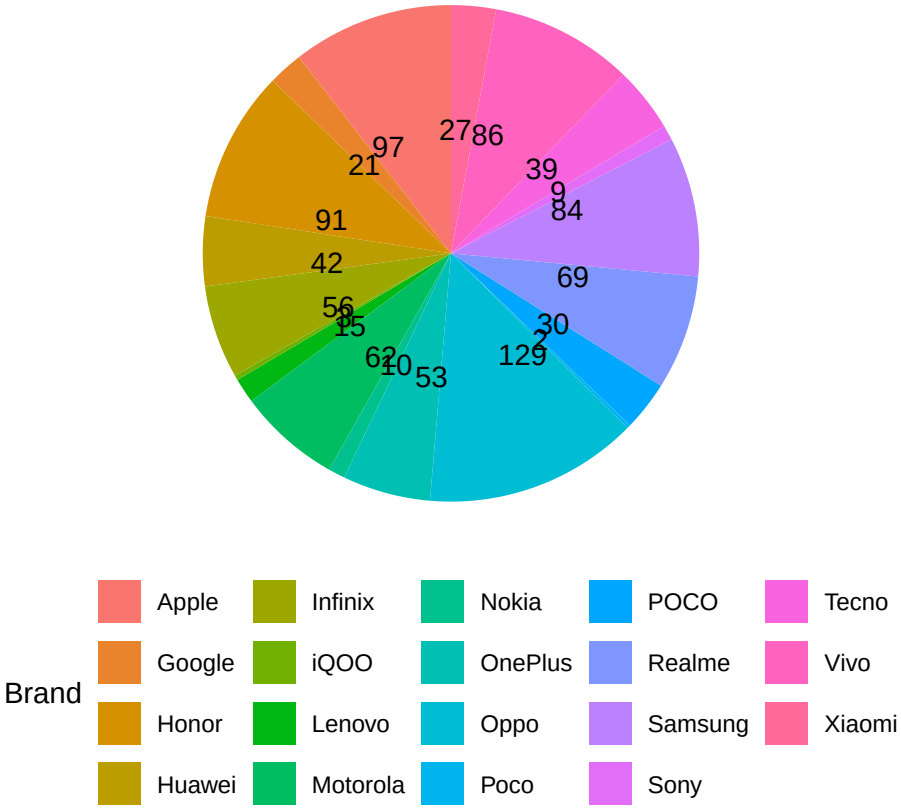
1. Plot a bar chart for average price per region in USD.

```
bar_df <- df_phones %>% select(c(Launched.Price.Pakistan.PKR,Launched.Price.India.INR,
                      Launched.Price.China.CNY, Launched.Price.USA.USD, Launched.Price.Dubai.AED)) %>%
                    Launched.Price.India.INR = Launched.Price.India.INR *  0.011,
                    Launched.Price.China.CNY = Launched.Price.China.CNY * 0.14,
                    Launched.Price.Dubai.AED = Launched.Price.Dubai.AED * 0.27) %>% summarise_all(mean)
bar_df <- data.frame(t(bar_df)) %>% `colnames<-`(c("Mean_Price"))
bar_df$Country <- rownames(bar_df)
ggplot(bar_df, aes(x=Country, y=Mean_Price)) + geom_bar(stat='identity') +
  scale_x_discrete(labels=c("China", "Dubai", "India", "Pakistan", "USA")) +
  labs(x = 'Mean Price', y= 'Country') +
  theme_bw()
```



2. Create a pie chart of the market share of smartphone brands.

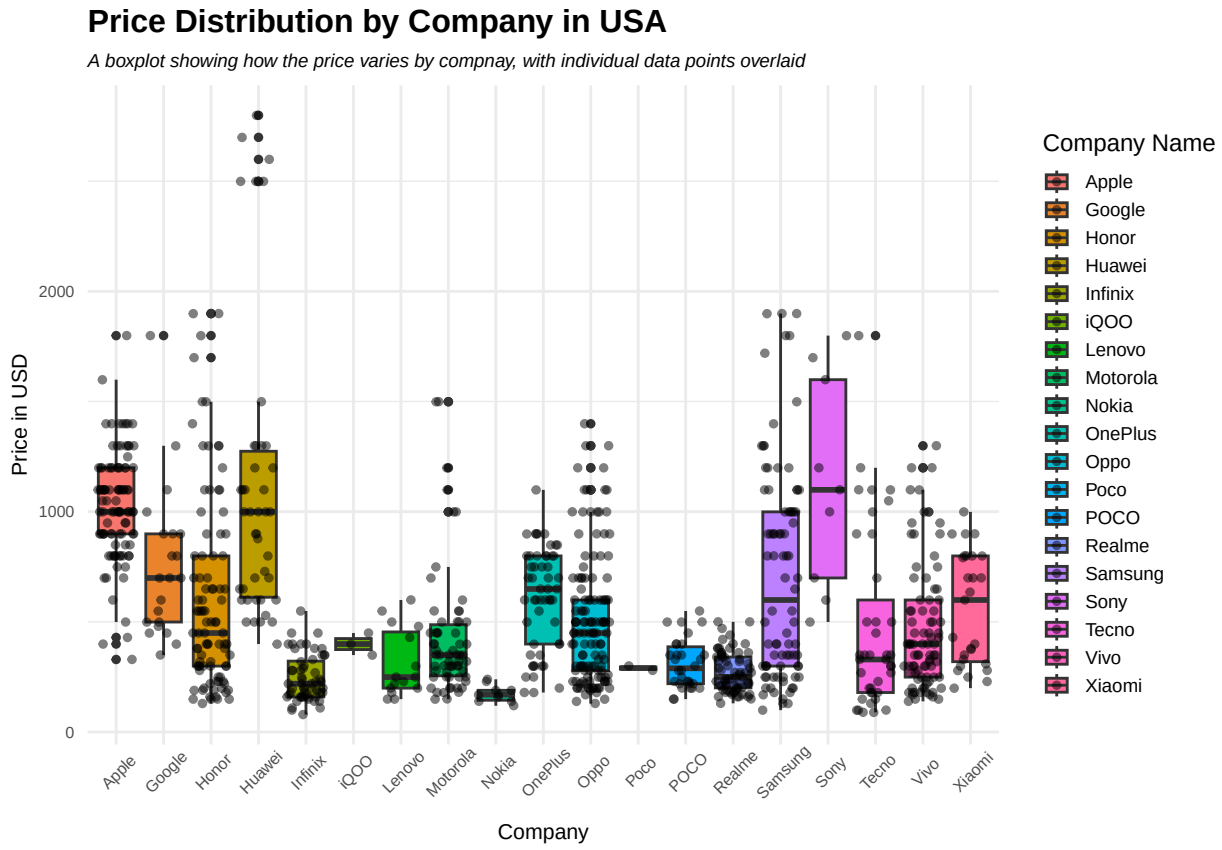
```
#Way to many categories for a pie chart nothing can save this.
pie_df <- df_phones %>% group_by(Company.Name) %>% count()
ggplot(pie_df, aes(x = "", y = n, fill = Company.Name)) +
  geom_col() +
  geom_text(aes(label = n),
            position = position_stack(vjust = 0.2)) +
  coord_polar(theta = "y") +
  labs(fill="Brand") +
  theme_void() +
  theme(legend.position = "bottom")
```



Part 3: Recreate (R only)

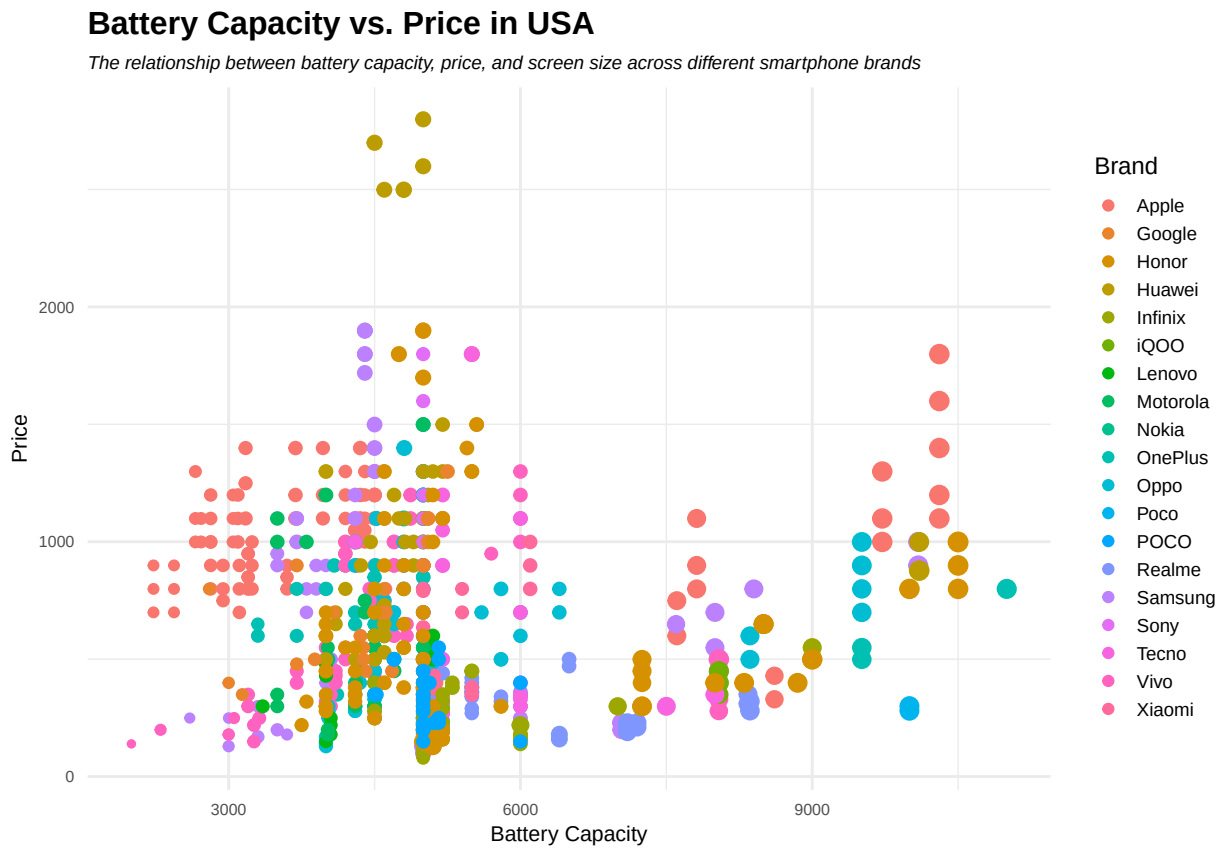
1)

```
ggplot(df_phones, aes(x=Company.Name, y=Launched.Price.USA.USD, fill=Company.Name)) +
  geom_boxplot(outlier.size = 1) +
  geom_jitter(color="black", size=1, alpha=0.5) +
  labs(title='Price Distribution by Company in USA', x='Company', y='Price in USD',
       subtitle='A boxplot showing how the price varies by compnay, with individual data points overlai
       fill = 'Company Name') +
  theme_minimal() +
    theme(
      plot.title = element_text(face="bold", size=12),
      plot.subtitle = element_text(face="italic", size=7),
      axis.text.x = element_text(angle = 45, vjust = 0.8, size = 6),
      axis.text.y = element_text(size = 6),
      legend.text = element_text(size=7),
      legend.title = element_text(size=9),
      legend.key.size = unit(0.37, 'cm'),
      axis.title.x = element_text(size = 8),
      axis.title.y = element_text(size = 8)
    )
```



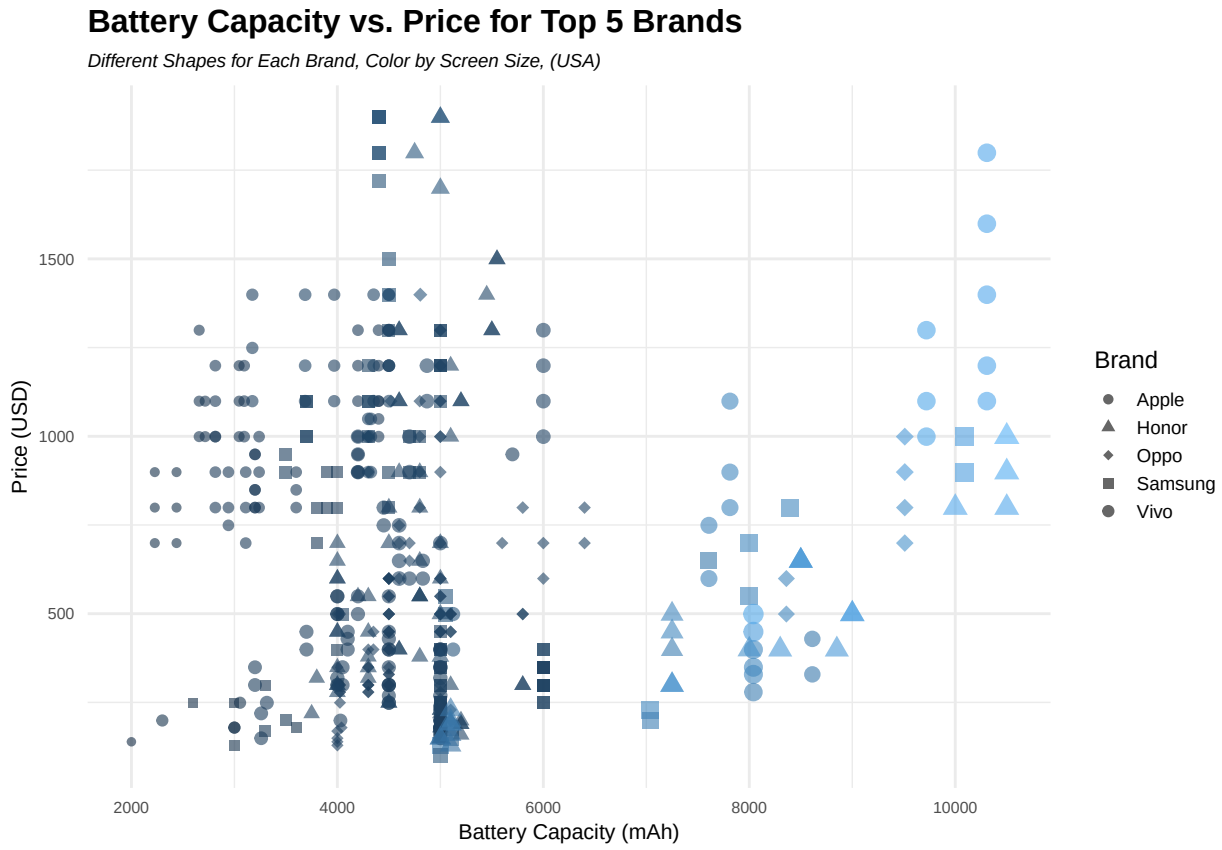
2)

```
ggplot(df_phones, aes(x=Battery.Capacity.mAh, y=Launched.Price.USA.USD, color=Company.Name, size=Screen
  geom_point() +
  scale_size(range = c(1,3)) +
  theme_minimal() +
  labs(title='Battery Capacity vs. Price in USA', x='Battery Capacity', y='Price',
       subtitle='The relationship between battery capacity, price, and screen size across different sma
       color = 'Brand') +
  guides(size="none") +
  theme(
    plot.title = element_text(face="bold", size=12),
    plot.subtitle = element_text(face="italic", size=7),
    axis.text.x = element_text(size = 6),
    axis.text.y = element_text(size = 6),
    legend.text = element_text(size=7),
    legend.title = element_text(size=9),
    legend.key.size = unit(0.37, 'cm'),
    axis.title.x = element_text(size = 8),
    axis.title.y = element_text(size = 8)
  )
```



3)

```
top5_df <- df_phones %>% filter(Company.Name %in% c('Apple', 'Honor', 'Oppo', 'Samsung', 'Vivo'))
ggplot(top5_df, aes(x=Battery.Capacity.mAh, y=Launched.Price.USA.USD, shape=Company.Name,
                    color=Screen.Size.inches, size=Screen.Size.inches)) +
  geom_point(alpha=0.6) +
  scale_size(range = c(1,3)) +
  scale_shape_manual(values = c(16, 17, 18, 15, 19)) +
  labs(title='Battery Capacity vs. Price for Top 5 Brands', x='Battery Capacity (mAh)', y='Price (USD)'
       subtitle='Different Shapes for Each Brand, Color by Screen Size, (USA)',
       shape = 'Brand') +
  guides(size='none', color='none') +
  theme_minimal() +
  theme(
    plot.title = element_text(face="bold", size=12),
    plot.subtitle = element_text(face="italic", size=7),
    axis.text.x = element_text(size = 6),
    axis.text.y = element_text(size = 6),
    legend.text = element_text(size=7),
    legend.title = element_text(size=9),
    legend.key.size = unit(0.37, 'cm'),
    axis.title.x = element_text(size = 8),
    axis.title.y = element_text(size = 8)
  )
```


Battery Capacity vs. Price for Top 5 Brands
Different Shapes for Each Brand, Color by Screen Size, (USA)
Price (USD)
1500
1000
500
2000
4000
6000
8000
10000
Battery Capacity (mAh)
Brand
Apple
Honor
Oppo
Samsung
Vivo