Data Visualization | Homework 3 | David Aslanyan | R

In [29]:
```
install.packages("dplyr")

library(tidyverse)
library(ggplot2)

mobiles <- read.csv("mobiles_dataset.csv")
```

In [30]:
```
colnames(mobiles)
```

'Company.Name' · 'Model.Name' · 'Mobile.Weight' · 'RAM' · 'Front.Camera' · 'Back.Camera' · 'Processor' · 'Battery.Capacity.mAh' · 'Screen.Size.inches' · 'Launched.Price.Pakistan.PKR' · 'Launched.Price.India.INR' · 'Launched.Price.China.CNY' · 'Launched.Price.USA.USD' · 'Launched.Price.Dubai.AED' · 'Launched.Year'

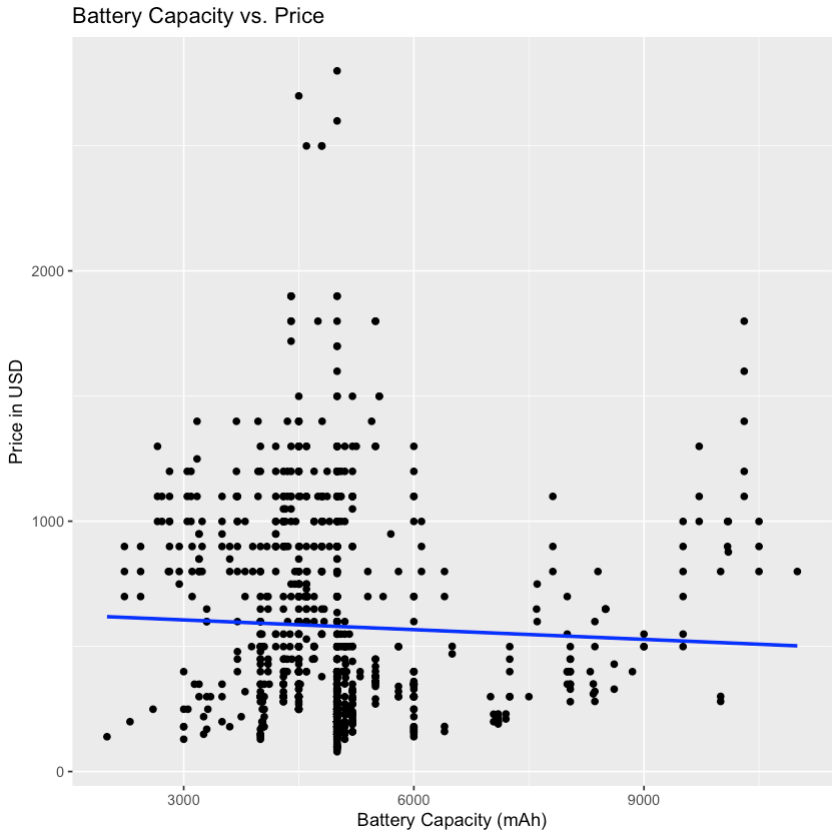
Firstly we are converting all the currencies into USD

In [31]:
```
mobiles <- mobiles %>%
  mutate(Price_USD_PKR = Launched.Price.Pakistan.PKR * 0.0036,
         Price_USD_INR = Launched.Price.India.INR * 0.011,
         Price_USD_CNY = Launched.Price.China.CNY * 0.14,
         Price_USD_AED = Launched.Price.Dubai.AED * 0.27,
         Price_USD_USA = Launched.Price.USA.USD)
```

Nextly what we want to do is to see the correlation between battery capacity and prices of the devices

In [32]:
```
ggplot(mobiles, aes(x = Battery.Capacity.mAh, y = Price_USD_USA)) +
  geom_point() +
  geom_smooth(method = "lm", se = FALSE, color = "blue") +
  labs(title = "Battery Capacity vs. Price",
       x = "Battery Capacity (mAh)", y = "Price in USD")
```

```
`geom_smooth()` using formula = 'y ~ x'
```

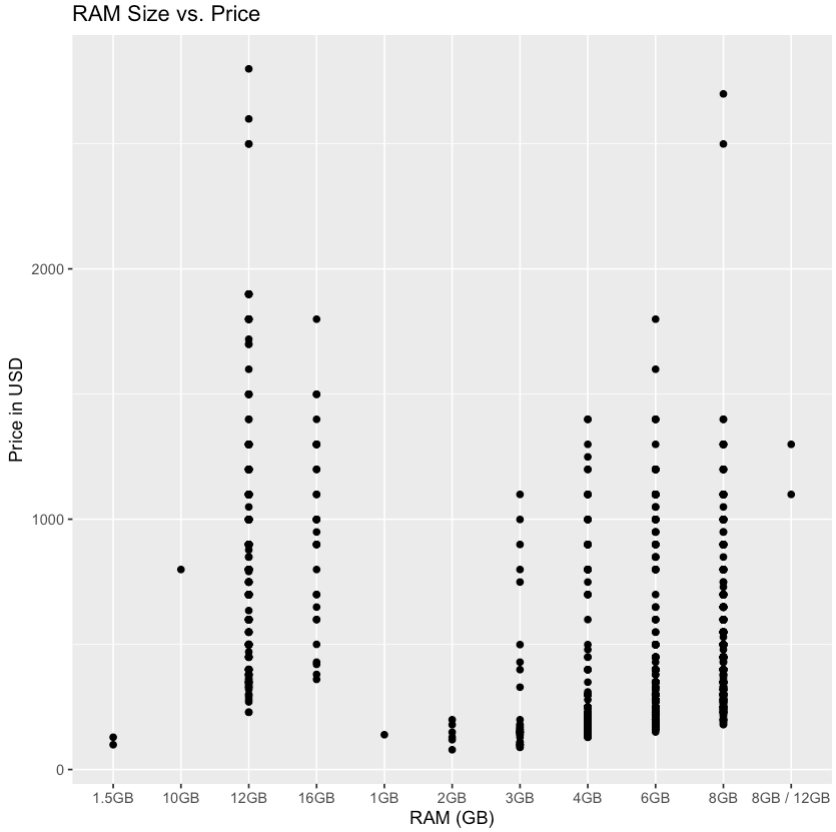


We see a weak, almost 0-ish correlation => the battery capacity not connected with the prices.

Next up we seek a correlation between RAM and Prices

In [33]:
```
ggplot(mobiles, aes(x = RAM, y = Price_USD_USA)) +
  geom_point() +
  geom_smooth(method = "lm", se = FALSE, color = "red") +
  labs(title = "RAM Size vs. Price",
       x = "RAM (GB)", y = "Price in USD")
```

```
`geom_smooth()` using formula = 'y ~ x'
```



Here we see more or less correlation => the more the RAM the higher the price.

Next up we want to see the price variation of all models in different regions.

In [34]:
```
brand_price_variation <- mobiles %>%
  group_by(Company.Name) %>%
  summarise(Price_SD = sd(c(Price_USD_PKR, Price_USD_INR, Price_USD_CNY, Pri
            .groups = "drop") %>%
  arrange(desc(Price_SD))

print(brand_price_variation)
```

```
# A tibble: 19 × 2
   Company.Name Price_SD
   <chr>           <dbl>
 1 Huawei          638.
 2 Samsung         515.
 3 Sony            401.
 4 Honor           397.
 5 Tecno           371.
 6 Google          355.
 7 Oppo            280.
 8 Apple           279.
 9 Xiaomi          275.
10 Vivo            273.
11 Motorola        267.
12 OnePlus         237.
13 Lenovo          173.
14 POCO            111.
15 Infinix         100.
16 Realme           99.8
17 iQOO             88.4
18 Nokia            45.7
19 Poco             24.0
```

Apple is not leading the list, however it is in Top 10, which means that the price pretty much varies in different regions.

In [35]:
```
mobiles <- mobiles %>%
  mutate(Company.Name = toupper(Company.Name))

unique(mobiles$Company.Name)
```

'APPLE' · 'SAMSUNG' · 'ONEPLUS' · 'VIVO' · 'IQOO' · 'OPPO' · 'REALME' · 'XIAOMI' · 'LENOVO' · 'MOTOROLA' · 'HUAWEI' · 'NOKIA' · 'SONY' · 'GOOGLE' · 'TECNO' · 'INFINIX' · 'HONOR' · 'POCO'

We ware making a list to see the brands that are more buget-option oriented or vica versa

In [36]:
```
mobiles <- mobiles %>%
  mutate(Price_Category = case_when(
    Price_USD_USA < 300 ~ "Budget",
    Price_USD_USA >= 300 & Price_USD_USA <= 700 ~ "Mid-Range",
    TRUE ~ "Premium"
  ))

brand_segments <- mobiles %>%
  group_by(Company.Name, Price_Category) %>%
  summarise(Count = n()) %>%
  spread(key = Price_Category, value = Count, fill = 0)

print(brand_segments)
```

```
`summarise()` has grouped output by 'Company.Name'. You can override using the
`.groups` argument.
# A tibble: 18 × 4
# Groups:   Company.Name [18]
   Company.Name Budget `Mid-Range` Premium
   <chr>         <dbl>       <dbl>   <dbl>
 1 APPLE             0           8      89
 2 GOOGLE            0          12       9
 3 HONOR            29          37      25
 4 HUAWEI            0          15      27
 5 INFINIX          41          15       0
 6 IQOO              0           3       0
 7 LENOVO           10           5       0
 8 MOTOROLA         22          33       7
 9 NOKIA            10           0       0
10 ONEPLUS          10          23      20
11 OPPO             46          59      24
12 POCO             17          15       0
13 REALME           43          26       0
14 SAMSUNG          26          19      39
15 SONY              0           3       6
16 TECNO            18          12       9
17 VIVO             33          37      16
18 XIAOMI            6          12       9
```

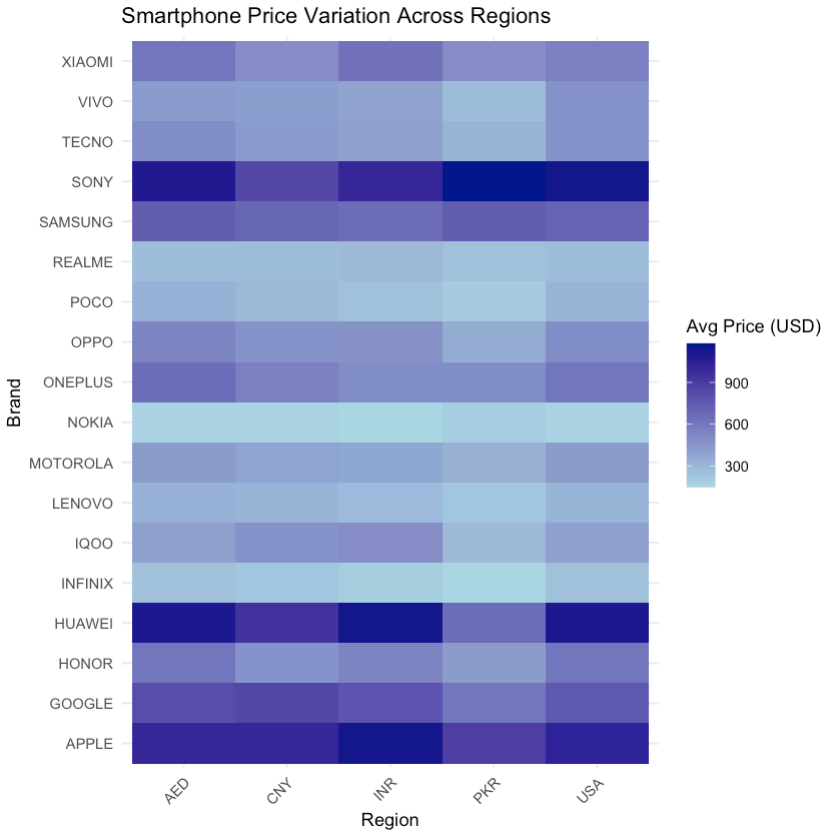
The analysis reveals that while most brands offer a mix of budget, mid-range, and premium models, some brands focus exclusively on premium devices, like Apple. Budget-friendly brands like Realme and Xiaomi lead the lower and mid-range categories, while brands like Samsung and OnePlus are balanced in terms of covering all segments.

Now we create a Heatmap to see how the prices of the same brand differ depending on the region.

In [46]:
```
install.packages("ggplot2")
install.packages("reshape2")
```

```
The downloaded binary packages are in
        /var/folders/cz/ysxc02l92kv04yfdk1rqmzx80000gn/T//RtmpM2fmz2/downlo
aded_packages

The downloaded binary packages are in
        /var/folders/cz/ysxc02l92kv04yfdk1rqmzx80000gn/T//RtmpM2fmz2/downlo
aded_packages
```

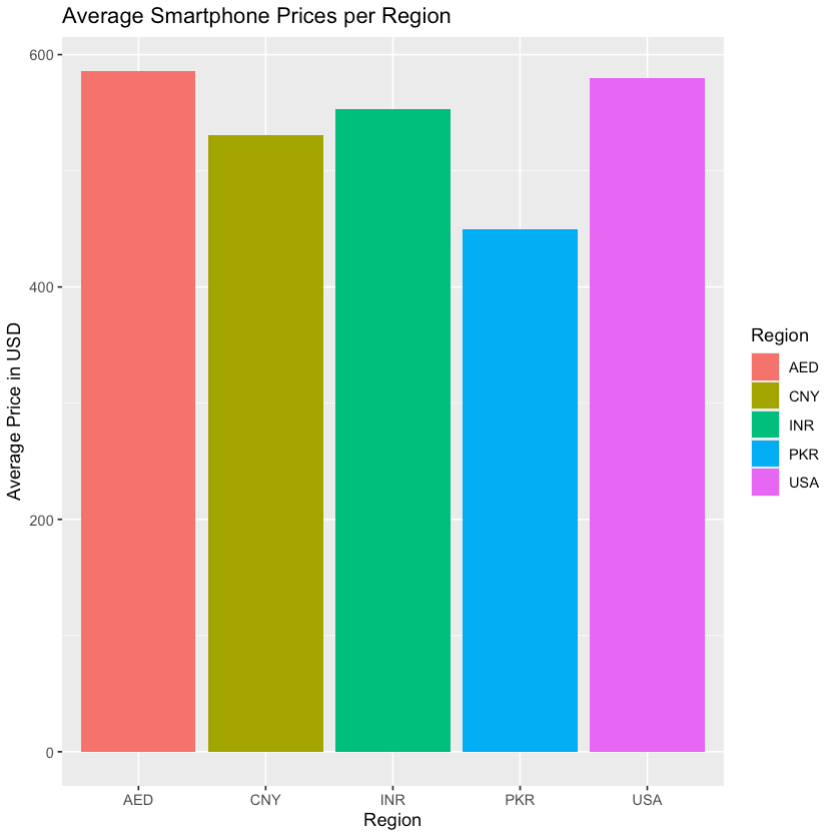
In [47]:
```
library(ggplot2)
library(reshape2)

brand_region_prices <- mobiles %>%
  group_by(Company.Name) %>%
  summarise(Avg_Price_PKR = mean(Price_USD_PKR, na.rm = TRUE),
            Avg_Price_INR = mean(Price_USD_INR, na.rm = TRUE),
            Avg_Price_CNY = mean(Price_USD_CNY, na.rm = TRUE),
            Avg_Price_AED = mean(Price_USD_AED, na.rm = TRUE),
            Avg_Price_USA = mean(Price_USD_USA, na.rm = TRUE),
            .groups = "drop")

brand_region_long <- melt(brand_region_prices, id.vars = "Company.Name",
                          variable.name = "Region", value.name = "Avg_Price"

brand_region_long$Region <- gsub("Avg_Price_", "", brand_region_long$Region)

ggplot(brand_region_long, aes(x = Region, y = Company.Name, fill = Avg_Price
  geom_tile() +
  scale_fill_gradient(low = "lightblue", high = "darkblue") +
  labs(title = "Smartphone Price Variation Across Regions",
       x = "Region", y = "Brand",
       fill = "Avg Price (USD)") +
  theme_minimal() +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))
```



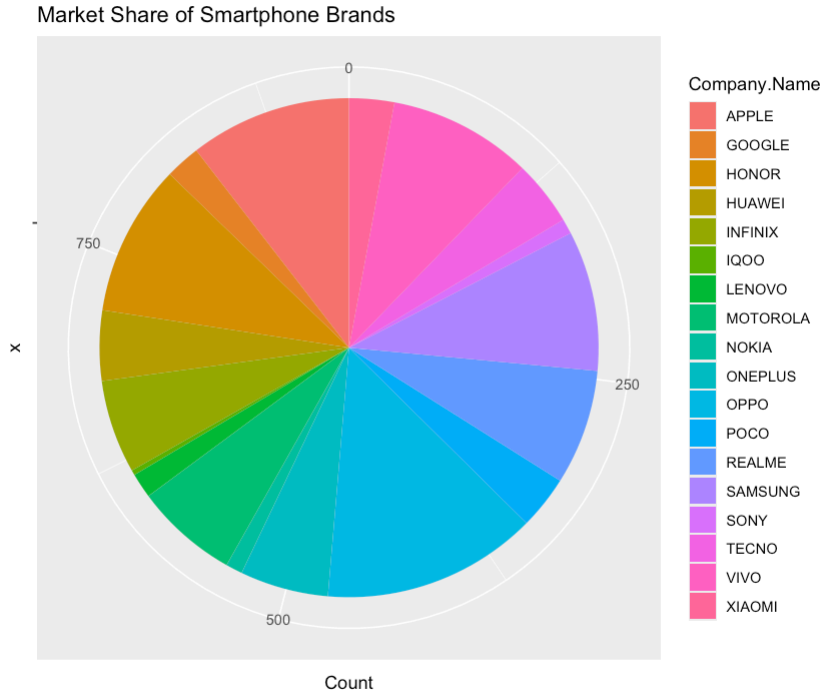
From the average price comparison across regions we notice that China and India generally have the most affordable smartphone prices, while USA and Dubai have higher prices, especially for brands like Huawei and Sony

Part 2 | Visualization

In [48]:
```
avg_prices <- mobiles %>%
  summarise(PKR = mean(Price_USD_PKR, na.rm = TRUE),
            INR = mean(Price_USD_INR, na.rm = TRUE),
            CNY = mean(Price_USD_CNY, na.rm = TRUE),
            AED = mean(Price_USD_AED, na.rm = TRUE),
            USA = mean(Price_USD_USA, na.rm = TRUE))

avg_prices_long <- pivot_longer(avg_prices, cols = everything(), names_to =

ggplot(avg_prices_long, aes(x = Region, y = Avg_Price, fill = Region)) +
  geom_bar(stat = "identity") +
  labs(title = "Average Smartphone Prices per Region", x = "Region", y = "Av
```



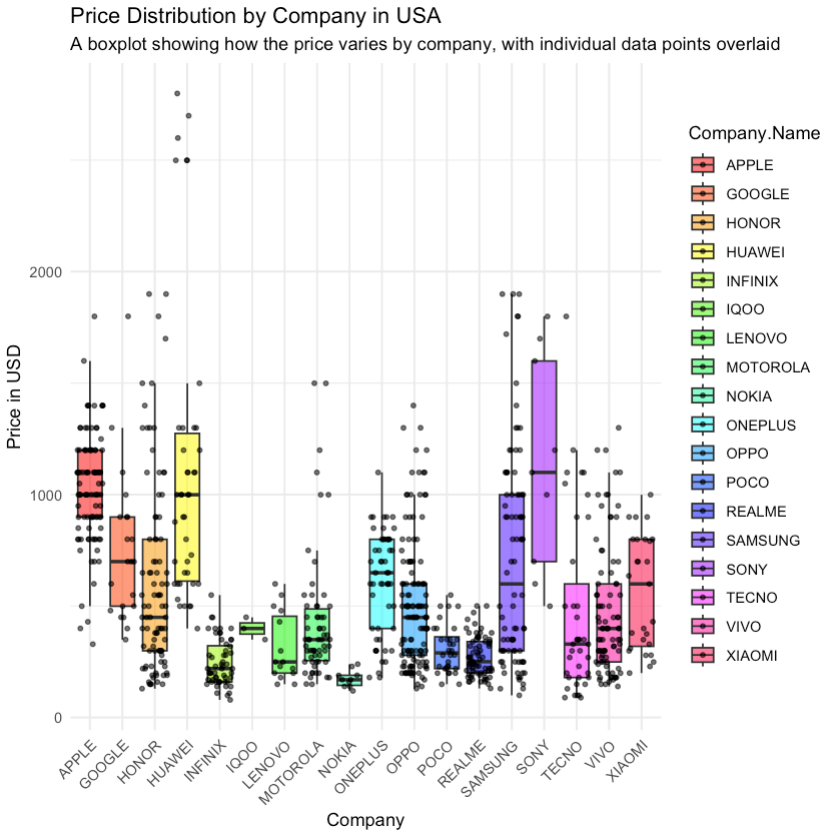
In [49]:
```
brand_market_share <- mobiles %>%
  group_by(Company.Name) %>%
  summarise(Count = n())

ggplot(brand_market_share, aes(x = "", y = Count, fill = Company.Name)) +
  geom_bar(stat = "identity", width = 1) +
  coord_polar("y", start = 0) +
  labs(title = "Market Share of Smartphone Brands")
```



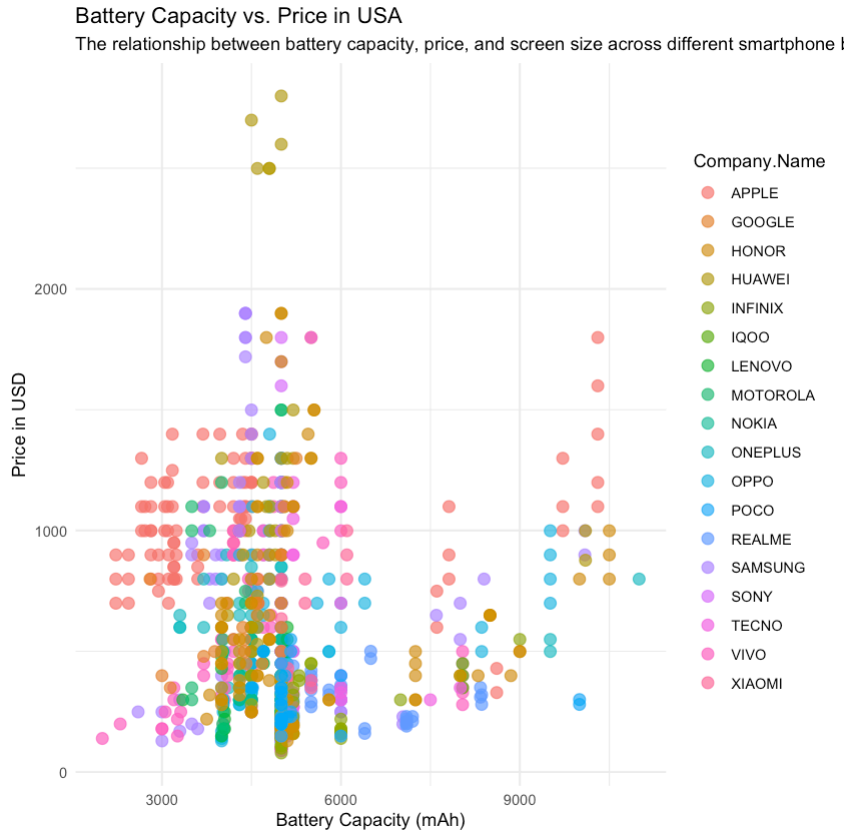
Part 3 | Recreation of Graph

In [50]:
```
library(ggplot2)
library(dplyr)

mobiles$Company.Name <- factor(mobiles$Company.Name)

ggplot(mobiles, aes(x = Company.Name, y = Price_USD_USA, fill = Company.Name
  geom_boxplot(outlier.shape = NA, alpha = 0.6) +
  geom_jitter(color = "black", size = 1, alpha = 0.6) +
  theme_minimal() +
  labs(title = "Price Distribution by Company in USA",
       subtitle = "A boxplot showing how the price varies by company, with i
       x = "Company", y = "Price in USD") +
  theme(axis.text.x = element_text(angle = 45, hjust = 1),
        legend.position = "right") +
  scale_fill_manual(values = rainbow(length(unique(mobiles$Company.Name))))
```

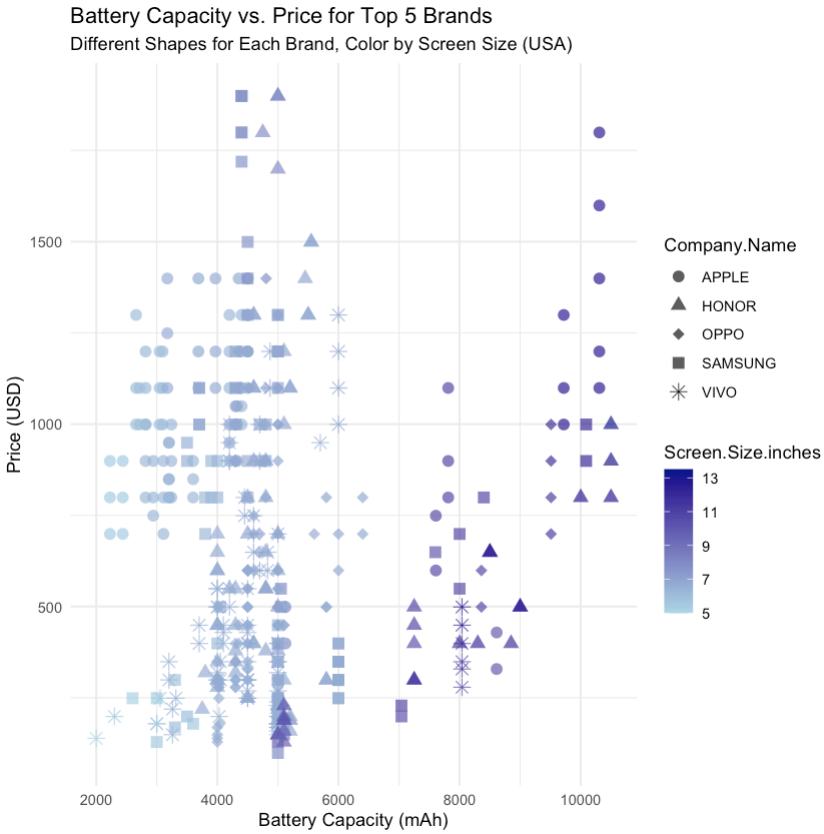
```
The downloaded binary packages are in
        /var/folders/cz/ysxc02l92kv04yfdk1rqmzx80000gn/T//RtmpM2fmz2/downlo
aded_packages

The downloaded binary packages are in
        /var/folders/cz/ysxc02l92kv04yfdk1rqmzx80000gn/T//RtmpM2fmz2/downlo
aded_packages
```



In [52]:
```
library(ggplot2)

ggplot(mobiles, aes(x = Battery.Capacity.mAh, y = Price_USD_USA, color = Cor
  geom_point(size = 3, alpha = 0.7) +
  theme_minimal() +
  labs(title = "Battery Capacity vs. Price in USA",
       subtitle = "The relationship between battery capacity, price, and sc
       x = "Battery Capacity (mAh)", y = "Price in USD") +
  theme(legend.position = "right")
```



In [57]:
```
library(ggplot2)
library(dplyr)

top_5_brands <- mobiles %>%
  count(Company.Name) %>%
  arrange(desc(n)) %>%
  slice(1:5) %>%
  pull(Company.Name)

mobiles_top5 <- mobiles %>% filter(Company.Name %in% top_5_brands)

brand_shapes <- c(16, 17, 18, 15, 8)

ggplot(mobiles_top5, aes(x = Battery.Capacity.mAh, y = Price_USD_USA, shape
  geom_point(size = 3, alpha = 0.7) +
  scale_color_gradient(low = "lightblue", high = "darkblue") +
  theme_minimal() +
  labs(title = "Battery Capacity vs. Price for Top 5 Brands",
       subtitle = "Different Shapes for Each Brand, Color by Screen Size (US
       x = "Battery Capacity (mAh)", y = "Price (USD)") +
  scale_shape_manual(values = brand_shapes) +
  theme(legend.position = "right")
```



For Part 4, the findings are the same as mentioned in the Python version.