

Homework 3 Part 1-4

Mariam Mkrtichyan

2025-03-19

R Markdown

This is an R Markdown document. Markdown is a simple formatting syntax for authoring HTML, PDF, and MS Word documents. For more details on using R Markdown see <http://rmarkdown.rstudio.com>.

When you click the **Knit** button a document will be generated that includes both content as well as the output of any embedded R code chunks within the document. You can embed an R code chunk like this:

```
library(ggplot2)
library(dplyr)

##
## 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

# Load dataset
df <- read.csv("mobiles_dataset.csv")

# Rename columns to match R-friendly names
df <- df %>%
  rename(company_name = Company.Name, launched_price_usa_usd = Launched.Price.USA.USD)

# Define custom order for companies
brand_order <- c("Apple", "Google", "Honor", "Huawei", "Infinix", "Lenovo", "Motorola",
                 "Nokia", "OnePlus", "Oppo", "Poco", "Realme", "Samsung", "Sony", "Tecno",
                 "Vivo", "Xiaomi")

# Remove missing values and set factor order
df_usa <- df %>%
  filter(company_name != 'NA') %>%
  filter(!is.na(launched_price_usa_usd)) %>%
  mutate(company_name = factor(company_name, levels = brand_order))

# Updated color mapping based on the provided image
```

```

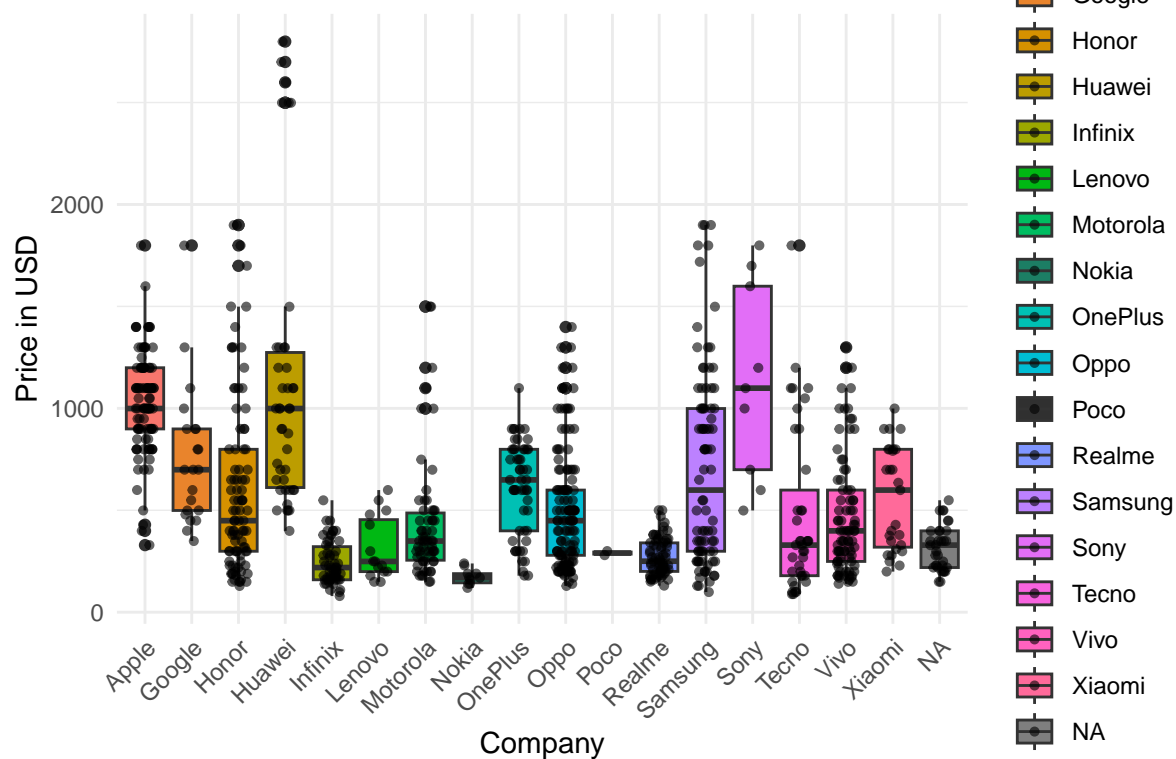
brand_colors <- c(
  "Apple" = "#F8766D", "Google" = "#E9842C", "Honor" = "#D69100",
  "Huawei" = "#BC9D00", "Infinix" = "#9CA700", "iQOO" = "#598907",
  "Lenovo" = "#00B813", "Motorola" = "#00BD61", "Nokia" = "#1C7E64",
  "OnePlus" = "#00C0B4", "Oppo" = "#00BDD4", "Poco" = "#303030",
  "POCO" = "#00A7FF", "Realme" = "#7F96FF", "Samsung" = "#BC81FF",
  "Sony" = "#E26EF7", "Tecno" = "#F863DF", "Vivo" = "#FF62BF",
  "Xiaomi" = "#FF6A9A"
)

# Create boxplot
ggplot(df_usa, aes(x = company_name, y = launched_price_usa_usd, fill = company_name)) +
  geom_boxplot(width = 0.8, alpha = 1, position = position_dodge(width = 1.5)) +
  geom_jitter(width = 0.2, size = 1.1, alpha = 0.6) +
  labs(
    title = "Price Distribution by Company in USA",
    subtitle = "A boxplot showing how the price varies by company, with individual data points overlaid",
    x = "Company",
    y = "Price in USD"
  ) +
  theme_minimal() +
  scale_fill_manual(values = brand_colors) +
  guides(color = guide_legend(title = "Company Name")) +
  theme(
    axis.text.x = element_text(angle = 45, hjust = 1),
    legend.position = "right",
    plot.title = element_text(face = "bold", size = 14),
    plot.subtitle = element_text(face = "italic", size = 10)
  )

```

Price Distribution by Company in USA

A boxplot showing how the price varies by company, with individual data points overlaid

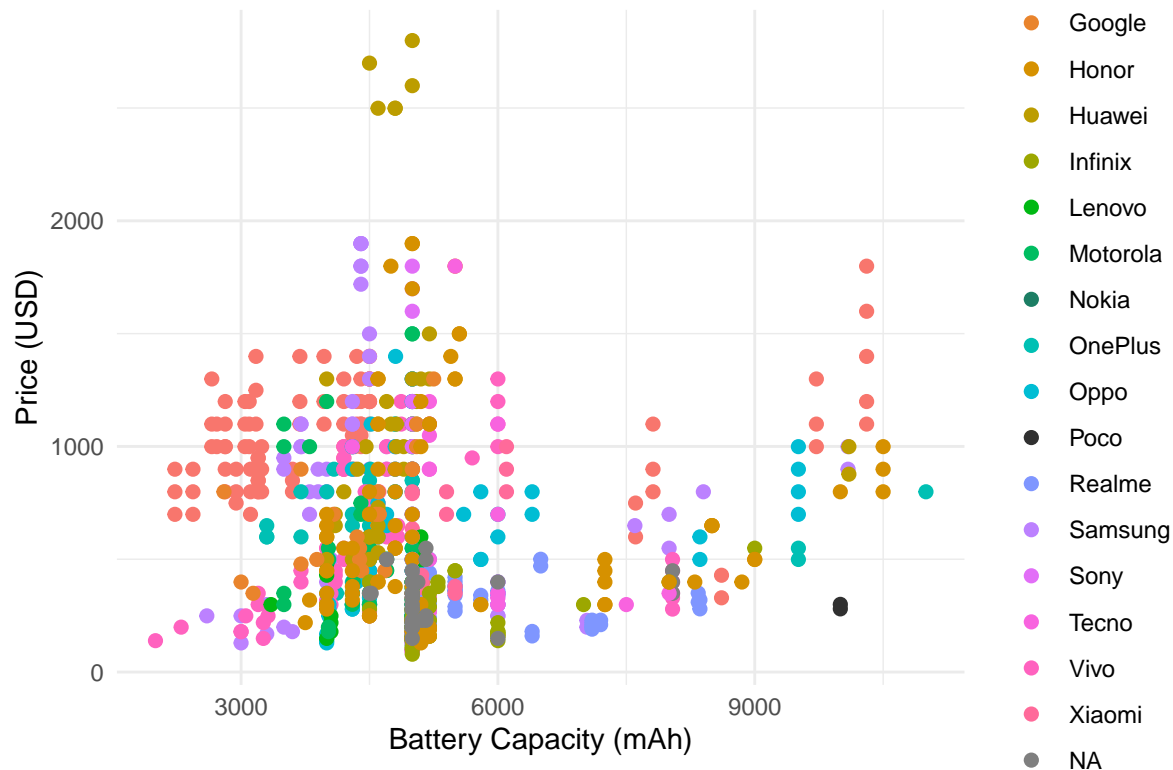


Plot 2: Battery Capacity vs Price Scatter Plot

```
ggplot(df_usa, aes(x = Battery.Capacity.mAh, y = launched_price_usa_usd, color = company_name)) +
  geom_point(size = 2, alpha = 1) +
  labs(
    title = "Battery Capacity vs. Price in USA",
    subtitle = "The relationship between battery capacity, price, and screen size across different smart phones",
    x = "Battery Capacity (mAh)",
    y = "Price (USD)"
  ) +
  theme_minimal() +
  scale_color_manual(values = brand_colors) +
  guides(color = guide_legend(title = "Company Name")) +
  theme(
    plot.title = element_text(face = "bold", size = 14),
    plot.subtitle = element_text(face = "italic", size = 10),
    legend.position = "right"
  )
```

Battery Capacity vs. Price in USA

The relationship between battery capacity, price, and screen size across different smartphone brands



```
# Filter for top 5 brands
top_brands <- c("Apple", "Honor", "Oppo", "Samsung", "Vivo")
df_top <- df_usa %>%
  filter(company_name %in% top_brands) %>%
  mutate(company_name = factor(company_name, levels = top_brands))

# Create the plot
ggplot(df_top, aes(x = Battery.Capacity.mAh,
  y = launched_price_usa_usd,
  shape = company_name,
  color = Screen.Size.inches)) +
  geom_point(size = 2.5, alpha = 0.6) +
  scale_shape_manual(values = c(16, 17, 18, 15, 19)) +
  scale_y_continuous(breaks = seq(500, 2000, by = 500)) +
  scale_x_continuous(breaks = seq(2000, 10000, by = 2000)) +
  labs(title = "Battery Capacity vs. Price",
    subtitle = "Different Shapes for Each Brand, Color by Screen Size (USA)",
    x = "Battery Capacity (mAh)",
    y = "Price (USD)") +
  theme_minimal() +
  theme(
    plot.title = element_text(face = "bold", size = 14),
    plot.subtitle = element_text(face = "italic", size = 10),
    legend.position = "right",
    legend.box = "vertical"
  ) +
  guides(
```

```

shape = guide_legend(title = "Brand"),
color = "none"
)

```

Battery Capacity vs. Price for Top 5 Brands

Different Shapes for Each Brand, Color by Screen Size (USA)

