Data Visualization in R

Supreeya Srasom

Libraries

```
library(tidyverse)
library(lubridate)
library(patchwork)
library(scales)
library(glue)
library(vroom)
```

Data Overview

Use data queried from the data.world, which represents the coffee chain.

```
coffee <- vroom("coffee_chain.csv", show_col_types = FALSE)
```

```
glimpse(coffee)
```

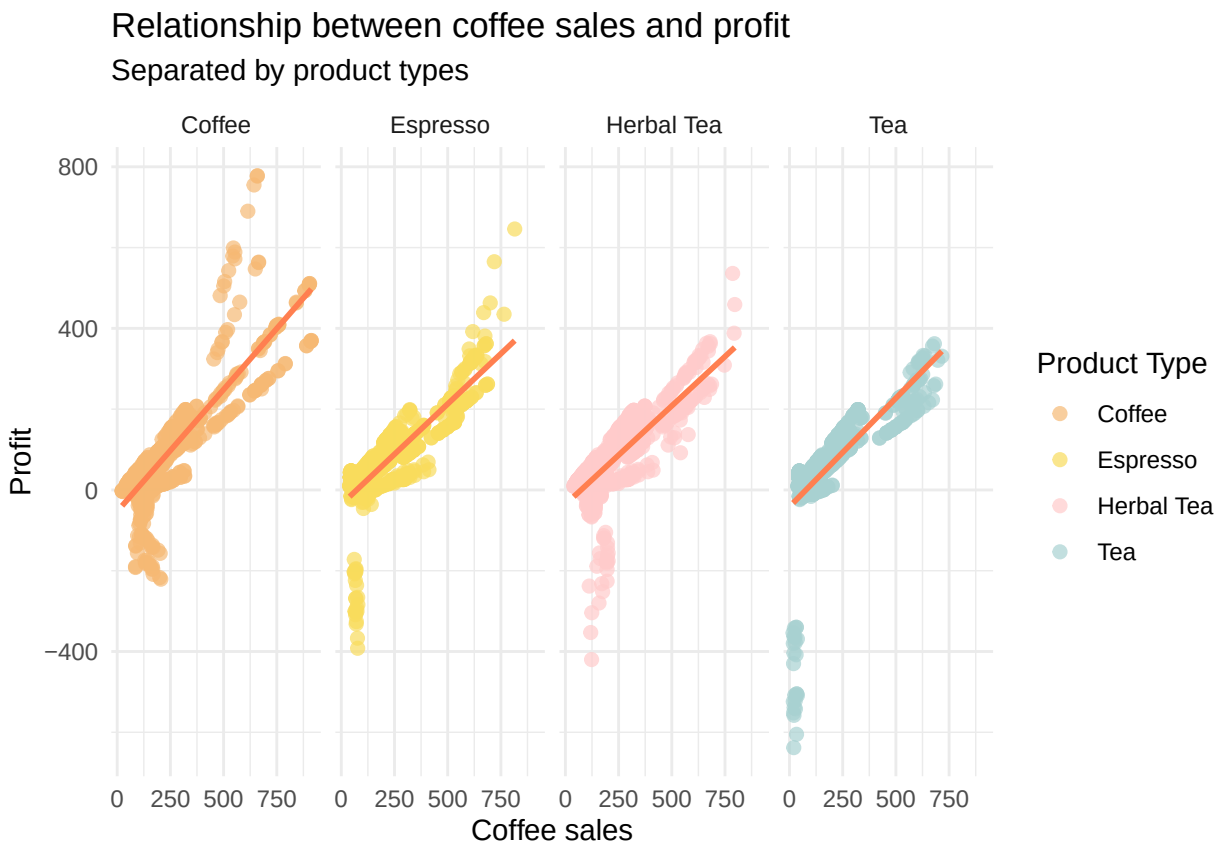
```
## Rows: 4,248
## Columns: 21
## $ `Area Code`         <dbl> 970, 719, 720, 303, 720, 719, 970, 719, 970, 719, ~
## $ Ddate               <date> 2001-01-12, 2002-01-12, 2003-01-12, 2004-01-12, 2~
## $ Market              <chr> "Central", "Central", "Central", "Central", "Centr~
## $ `Market Size`       <chr> "Major Market", "Major Market", "Major Market", "M~
## $ Product             <chr> "Decaf Irish Cream", "Decaf Irish Cream", "Decaf I~
## $ `Product Type`      <chr> "Coffee", "Coffee", "Coffee", "Coffee", "Coffee", ~
## $ State               <chr> "Colorado", "Colorado", "Colorado", "Colorado", "C~
## $ Type                <chr> "Decaf", "Decaf", "Decaf", "Decaf", "Decaf", "Deca~
## $ `Budget Cogs`       <dbl> 100, 100, 100, 100, 110, 130, 130, 130, 100, 100, ~
## $ `Budget Margin`     <dbl> 140, 140, 140, 150, 150, 180, 190, 200, 150, 160, ~
## $ `Budget Profit`     <dbl> 110, 110, 110, 120, 120, 140, 150, 160, 120, 150, ~
## $ `Budget Sales`      <dbl> 240, 240, 240, 250, 260, 310, 320, 330, 250, 260, ~
## $ `Coffee Sales`      <dbl> 234, 232, 234, 245, 256, 301, 312, 323, 245, 265, ~
## $ Cogs                <dbl> 95, 95, 95, 100, 104, 123, 127, 132, 100, 108, 81,~
## $ Inventory           <dbl> 821, 809, 799, 822, 871, 947, 1007, 994, 981, 971,~
## $ Margin              <dbl> 139, 137, 139, 145, 152, 178, 185, 191, 145, 157, ~
## $ Marketing           <dbl> 26, 26, 26, 28, 29, 34, 35, 36, 28, 30, 22, 24, 26~
## $ `Number of Records` <dbl> 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1,~
## $ `Number Of Records` <dbl> 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1,~
## $ Profit              <dbl> 101, 99, 101, 105, 112, 132, 139, 144, 106, 116, 8~
## $ `Total Expenses`    <dbl> 38, 38, 38, 40, 40, 46, 46, 47, 39, 41, 33, 36, 40~
```

Data Visualization

Chart 1: Relationship between coffee sales and profit

```
four_types <- coffee %>%
  count(`Product Type`) %>%
  arrange(desc(n)) %>%
  head(4)

coffee %>%
  filter(`Product Type` %in% four_types$`Product Type`) %>%
  ggplot(aes(`Coffee Sales`, Profit, color = `Product Type`)) +
  geom_point(size = 2, alpha = 0.7) +
  geom_smooth(formula = y ~ x, method = "lm", se = FALSE,
              color = "coral", linewidth = 1) +
  scale_color_manual(values = c("#F6B974",
                                "#F9DC5C",
                                "#FFCBCB",
                                "#A8D1D1")) +
  facet_wrap(~ `Product Type`, ncol = 4) +
  theme_minimal() +
  labs(title = "Relationship between coffee sales and profit",
       subtitle = "Separated by product types",
       x = "Coffee sales",
       y = "Profit")
```

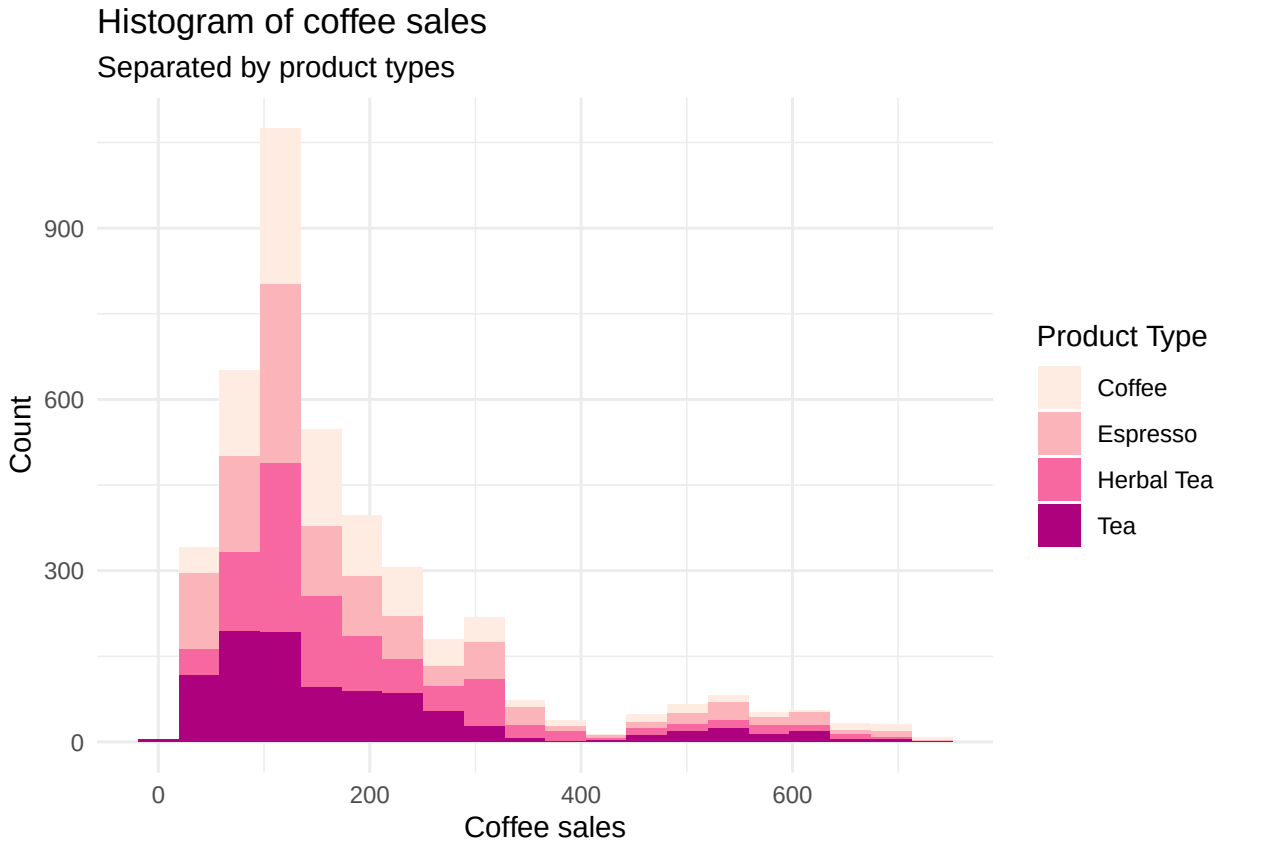


The scatter plot shows the relationship between coffee sales and profit by product types. As sales increase, the profit of coffee also increases.

Chart 2: Histogram of coffee sales

```
data_coffee <- coffee %>%
  count(`Product Type`) %>%
  arrange(desc(n)) %>%
  head(4)

coffee %>%
  filter(`Coffee Sales` < 750,
         `Product Type` %in% data_coffee$`Product Type`) %>%
  ggplot(aes(`Coffee Sales`, fill = `Product Type`)) +
  geom_histogram(bins = 20) +
  scale_fill_brewer(type = "seq", palette = "RdPu") +
  theme_minimal() +
  labs(title = "Histogram of coffee sales",
       subtitle = "Separated by product types",
       x = "Coffee sales",
       y = "Count")
```

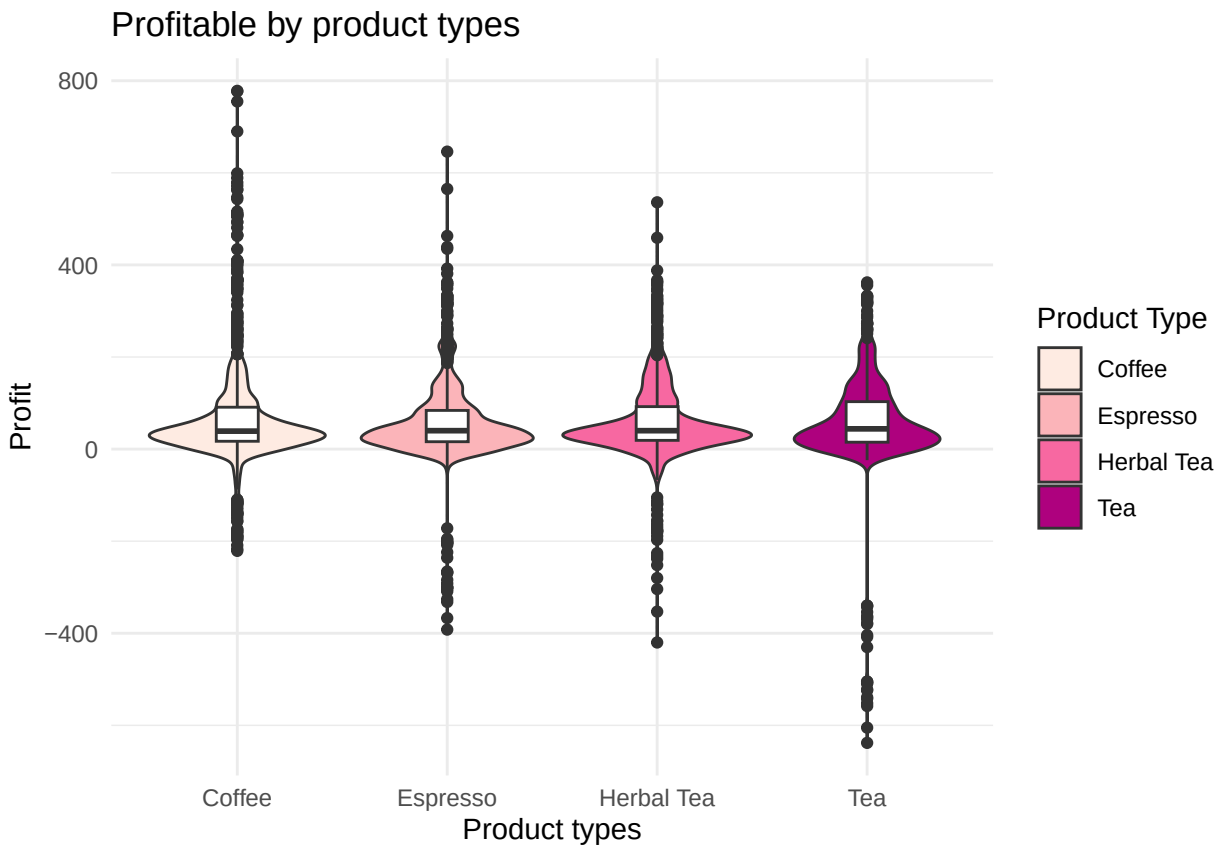


This histogram shows the count of the coffee sales by product types.

Chart 3: Profitable by product types

```
four_types <- coffee %>%
  count(`Product Type`) %>%
  arrange(desc(n)) %>%
  head(4)

coffee %>%
  filter(`Product Type` %in% four_types$`Product Type`) %>%
  ggplot(aes(`Product Type`, Profit, fill = `Product Type`)) +
  geom_violin() +
  geom_boxplot(width = 0.2, fill = "white") +
  scale_fill_brewer(palette = "RdPu") +
  theme_minimal() +
  labs(title = "Profitable by product types",
       x = "Product types",
       y = "Profit")
```



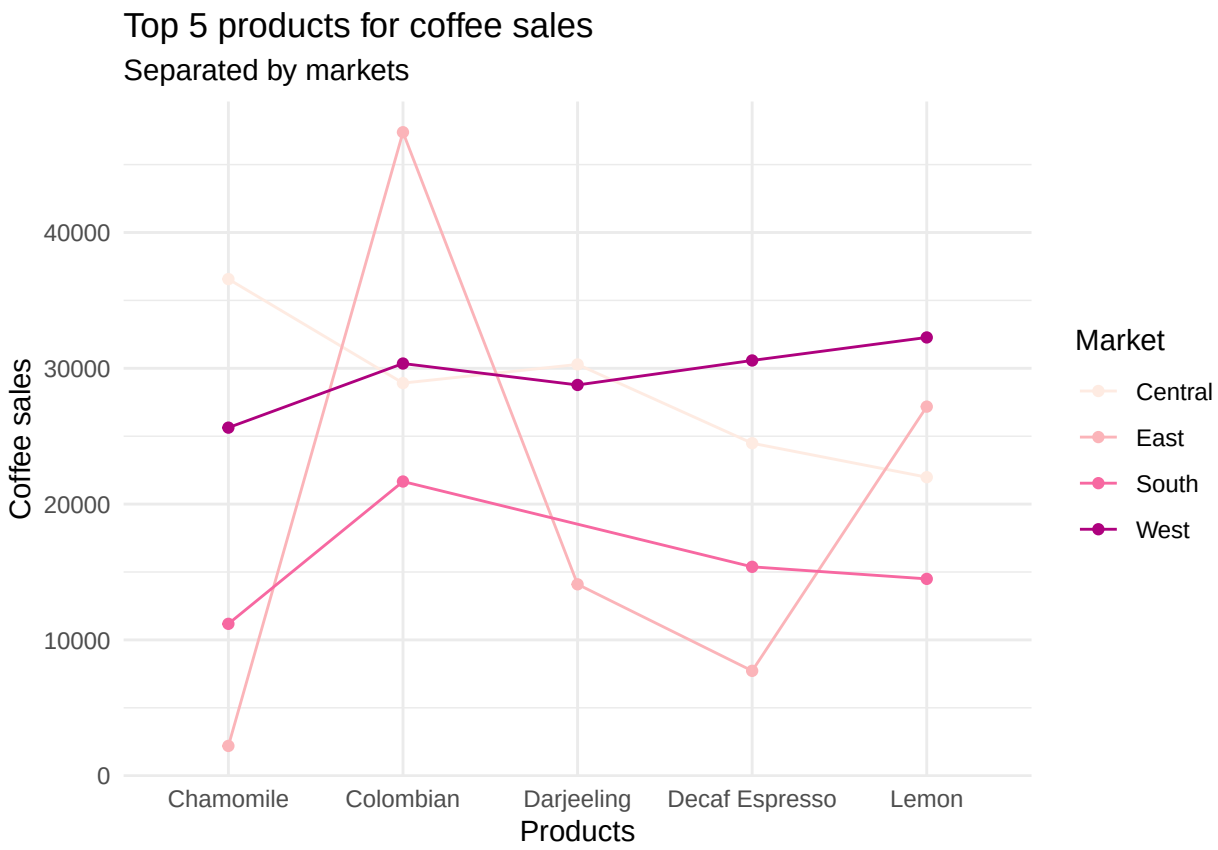
The violin plot and box plot shows the profitable by product types. There are too many outliers in tea.

Chart 4: Top 5 products for coffee sales

```
marketing <- coffee %>%
  count(Market) %>%
  arrange(desc(n)) %>%
  head(4)
```

```r
products <- coffee %>%
  group_by(five_products = Product) %>%
  summarise(sum_profit = sum(Profit),
            sum_sales = sum(`Coffee Sales`), .groups = "drop") %>%
  arrange(desc(sum_profit)) %>%
  head(5)

coffee %>%
  filter(Market %in% marketing$Market,
         Product %in% products$five_products) %>%
  group_by(five_products = Product, Market) %>%
  summarise(total = sum(`Coffee Sales`), .groups = "drop") %>%
  ggplot(aes(five_products, total, group = Market,
             color = Market)) +
  geom_point() +
  geom_line() +
  scale_color_brewer(palette = "RdPu") +
  theme_minimal() +
  labs(title = "Top 5 products for coffee sales",
       subtitle = "Separated by markets",
       x = "Products",
       y = "Coffee sales")
```



It can be seen that the top 5 products that continue to increase in coffee sales are west markets.