Assignment-3—Part-II.R

frank

2023-11-11

```
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
```

```
library(ggplot2)
library(gridExtra)
```

```
##
## Attaching package: 'gridExtra'
```

```
## The following object is masked from 'package:dplyr':
##
##     combine
```

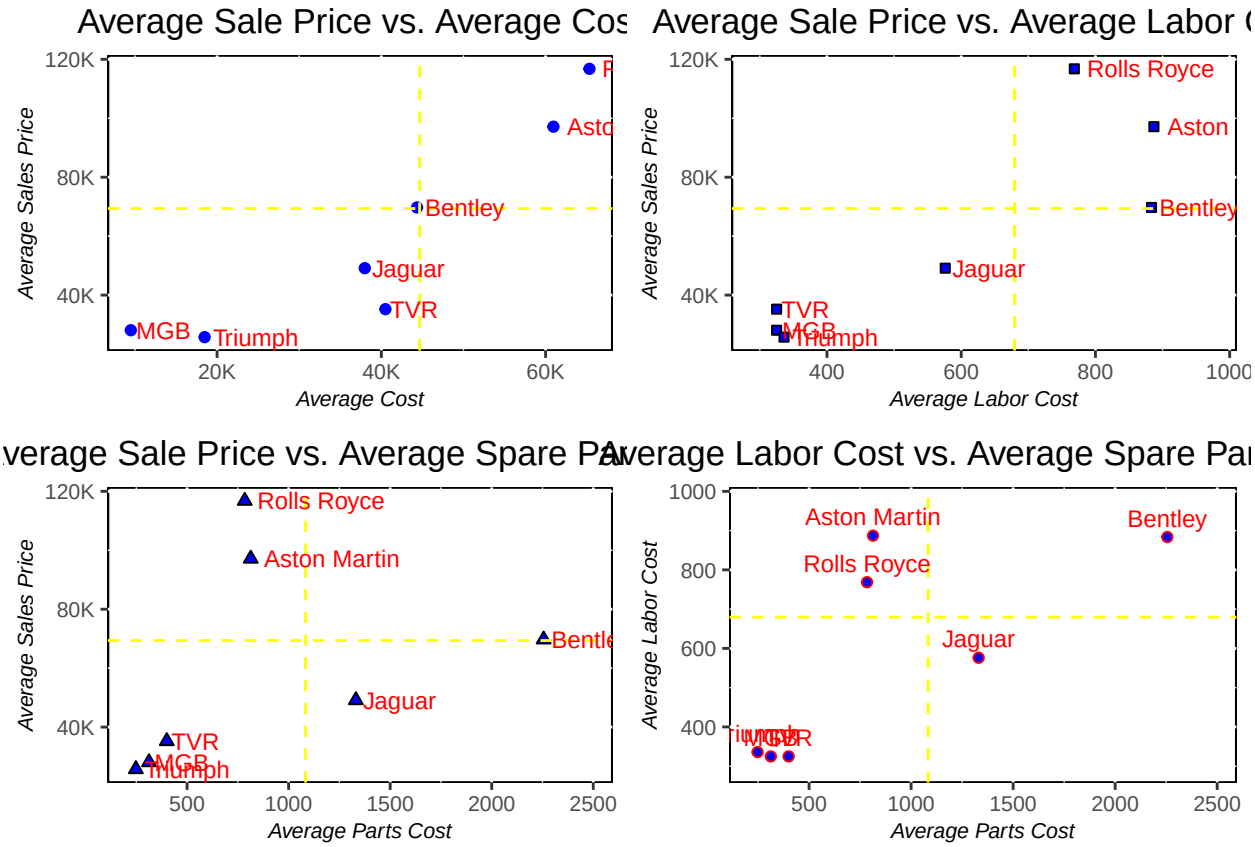
```
#########################################################
#Question 1 - Read in Data
#########################################################
car_data <- read.csv("/Users/frank/Desktop/Classes/QMB 6304 Data Visualization/Assignment 3/CarSales.csv")



#########################################################
#Question 2 - Scatter Plots
#########################################################
average_data <- car_data %>%
  group_by(Make) %>%
  summarize(
    avg_sale_price = mean(SalePrice),
    avg_cost = mean(CostPrice),
    avg_labor_cost = mean(LaborCost),
    avg_spare_parts_cost = mean(SpareParts)
  )

# Calculate average lines vert/horizontal
overall_avg_cost <- mean(car_data$CostPrice)
overall_avg_sale_price <- mean(car_data$SalePrice)
overall_avg_labor_cost <- mean(car_data$LaborCost)
```

```r
overall_avg_spare_parts_cost <- mean(car_data$SpareParts)



plot1 <- ggplot(average_data, aes(x = avg_cost, y = avg_sale_price)) +
  geom_point(shape = 19, color = "blue") +
  geom_hline(yintercept = overall_avg_sale_price, linetype =
               "dashed", size = 0.5, color = "yellow") +
  geom_vline(xintercept = overall_avg_cost, linetype = "dashed",
             size = 0.5, color = "yellow") +
  geom_text(aes(label = Make), color = "red", size = 3, hjust = -0.1) +
  labs(title = "Average Sale Price vs. Average Cost",
       x = "Average Cost", y = "Average Sales Price") +
  theme(plot.title = element_text(hjust = 0.5),axis.text =
          element_text(size = 8),axis.title = element_text(size = 8,
                                                           face= "italic"))+
  scale_x_continuous(breaks = seq(0, 80000, by = 20000),
                     labels = scales::comma_format(scale = 1e-3,
                                                   suffix = "K")) +
  scale_y_continuous(breaks = seq(0, 120000, by = 40000),
                     labels = scales::comma_format(scale = 1e-3,
                                                   suffix = "K")) +
  theme(panel.background = element_rect(fill = "white", color = "black"))
```

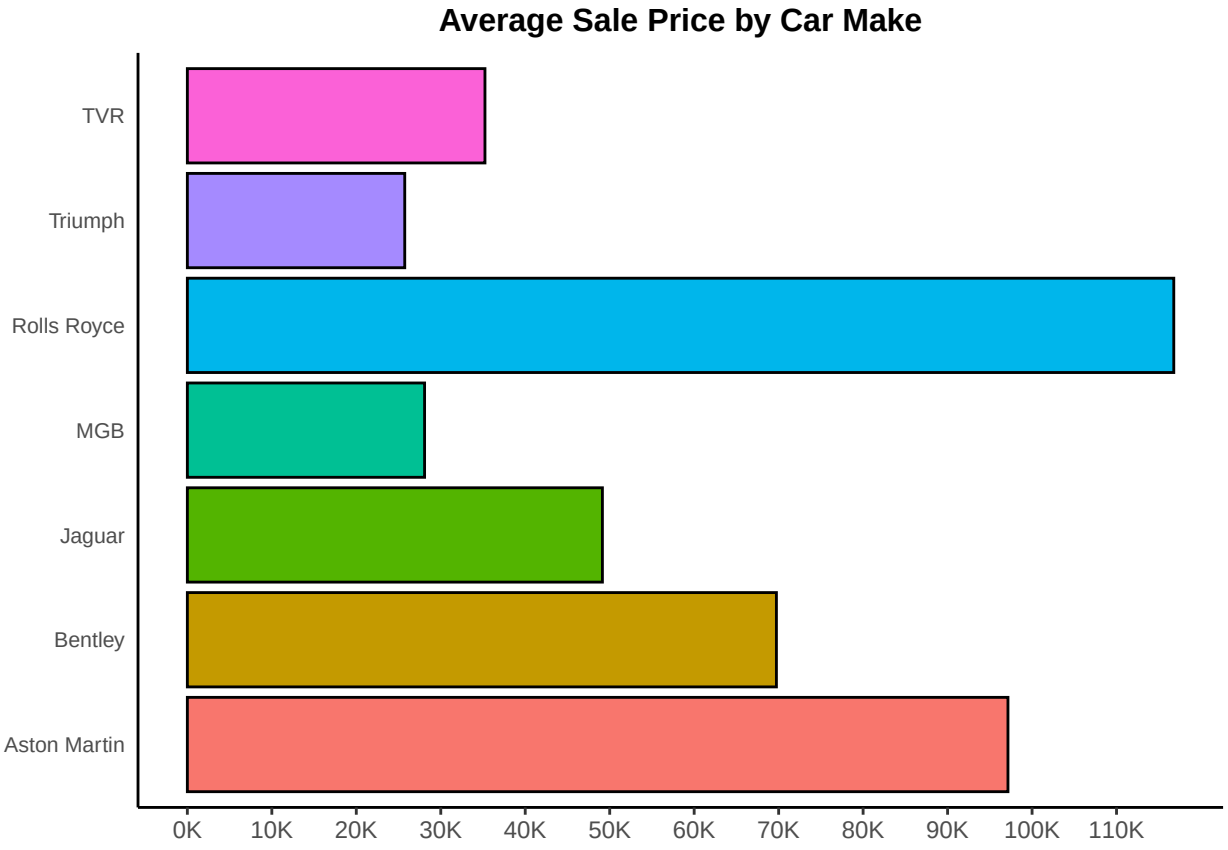
```
## Warning: Using `size` aesthetic for lines was deprecated in ggplot2 3.4.0.
## i Please use `linewidth` instead.
## This warning is displayed once every 8 hours.
## Call `lifecycle::last_lifecycle_warnings()` to see where this warning was
## generated.
```

```r
plot2 <- ggplot(average_data, aes(x = avg_labor_cost, y = avg_sale_price)) +
  geom_point(shape = 22, fill = "blue", color = "black") +
  geom_hline(yintercept = overall_avg_sale_price, linetype = "dashed",
             size = 0.5, color = "yellow") +
  geom_vline(xintercept = overall_avg_labor_cost, linetype = "dashed",
             size = 0.5, color = "yellow") +
  geom_text(aes(label = Make), color = "red", size = 3, hjust = -0.1) +
  labs(title = "Average Sale Price vs. Average Labor Cost",
       x = "Average Labor Cost", y = "Average Sales Price") +
  theme(plot.title = element_text(hjust = 0.5),axis.text =
          element_text(size = 8),axis.title = element_text(size = 8,
                                                           face= "italic"))+
  scale_x_continuous(limits = c(0.9 * min(average_data$avg_labor_cost),
                                1.1 * max(average_data$avg_labor_cost)))+
  scale_y_continuous(breaks = seq(0, 120000, by = 40000), labels = scales::
                       comma_format(scale = 1e-3, suffix = "K")) +
  theme(panel.background = element_rect(fill = "white", color = "black"))

plot3 <- ggplot(average_data, aes(x = avg_spare_parts_cost,
                                  y = avg_sale_price)) +
  geom_point(shape = 24, fill = "blue") +
  geom_hline(yintercept = overall_avg_sale_price, linetype = "dashed",
             size = 0.5, color = "yellow") +
  geom_vline(xintercept = overall_avg_spare_parts_cost, linetype = "dashed",
```

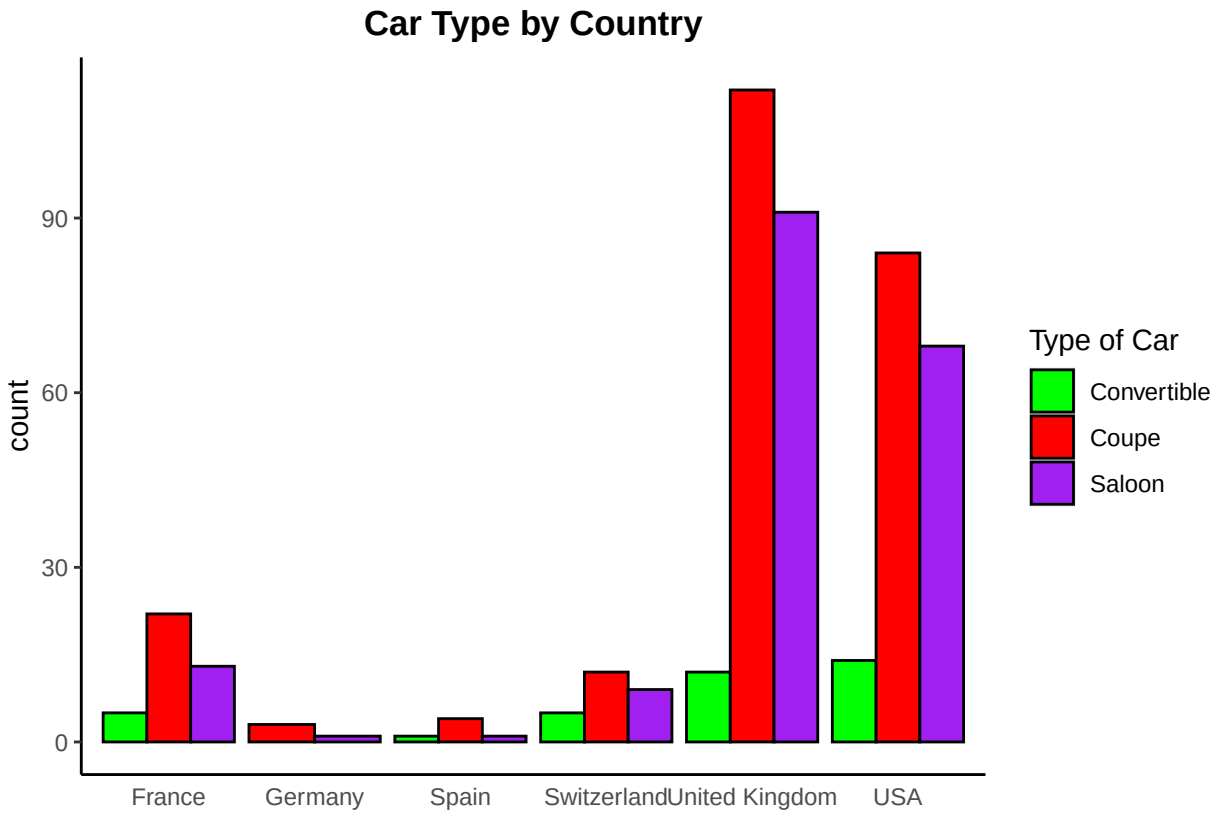
```
            size = 0.5, color = "yellow") +
  geom_text(aes(label = Make), color = "red", size = 3, hjust = -0.1) +
  labs(title = "Average Sale Price vs. Average Spare Parts Cost",
       x = "Average Parts Cost", y = "Average Sales Price") +
  theme(plot.title = element_text(hjust = 0.5),axis.text =
          element_text(size = 8),axis.title = element_text(size = 8,
                                                           face= "italic"))+
  scale_x_continuous(limits = c(0.9 * min(average_data$avg_spare_parts_cost),
                                1.1 * max(average_data$avg_spare_parts_cost)))+
  scale_y_continuous(breaks = seq(0, 120000, by = 40000), labels = scales::
                       comma_format(scale = 1e-3, suffix = "K")) +
  theme(panel.background = element_rect(fill = "white", color = "black"))

plot4 <- ggplot(average_data, aes(x = avg_spare_parts_cost, y =
                                    avg_labor_cost)) +
  geom_point(shape = 21, fill = "blue", color = "red") +
  geom_hline(yintercept = overall_avg_labor_cost, linetype = "dashed",
             size = 0.5, color = "yellow") +
  geom_vline(xintercept = overall_avg_spare_parts_cost, linetype = "dashed",
             size = 0.5, color = "yellow") +
  geom_text(aes(label = Make), color = "red", size = 3, vjust = -0.6) +
  labs(title = "Average Labor Cost vs. Average Spare Parts Cost", x =
         "Average Parts Cost", y = "Average Labor Cost") +
  theme(plot.title = element_text(hjust = 0.5),axis.text =
          element_text(size = 8),axis.title =
          element_text(size = 8, face= "italic"))+
  scale_x_continuous(limits = c(0.9 * min(average_data$avg_spare_parts_cost),
                                1.1 * max(average_data$avg_spare_parts_cost)))+
  scale_y_continuous(limits = c(0.9 * min(average_data$avg_labor_cost), 1.1 *
                                  max(average_data$avg_labor_cost)))+
  theme(panel.background = element_rect(fill = "white", color = "black"))

# 2x2 matrix
grid.arrange(plot1, plot2, plot3, plot4, ncol = 2)
```




```
#######################################################
#Question 3 - Bar Plots
#######################################################


horizontal_bar_plot <- ggplot(average_data,
                              aes(x = Make, y = avg_sale_price, fill = Make)) +
  geom_bar(stat = "identity", color = "black") +
  labs(title = "Average Sale Price by Car Make", x = NULL, y = NULL) +
  theme(
    plot.title = element_text(hjust = 0.5, size = 12, face = "bold"),
    axis.text.y = element_text(size = 8),
    axis.ticks.y = element_blank(),
    panel.background = element_rect(fill = "white", color = "white"),
    panel.grid.major = element_blank(),
    panel.grid.minor = element_blank(),
    axis.line = element_line(color = "black"),
    legend.position = "none"
  ) +
  scale_y_continuous(breaks = seq(0, max(average_data$avg_sale_price),
                                  by = 10000), labels = scales::comma_format
                     (scale = 1e-3, suffix = "K")) +
  coord_flip()

print(horizontal_bar_plot)
```

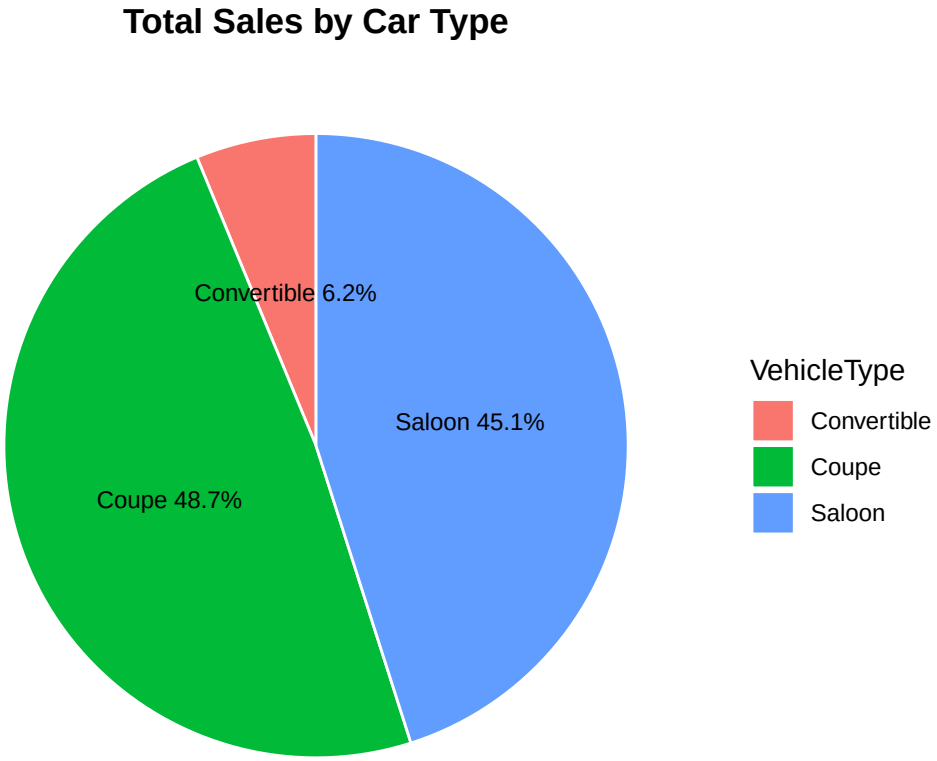



```
#########################################################
#Question 4 - Grouped Bar Plot
#########################################################


ggplot(car_data, aes(x = CountryName, fill = VehicleType)) +
  geom_bar(position = "dodge", color = "black") +
  labs(title = "Car Type by Country", x = "", fill = "Type of Car") +
  scale_fill_manual(values = c("green", "red", "purple")) +
  theme(
    plot.title = element_text(hjust = 0.5, face = "bold"),
    panel.background = element_rect(fill = "white", color = "white"),
    axis.line = element_line(color = "black"),
    axis.ticks.x = element_blank()
  )
```

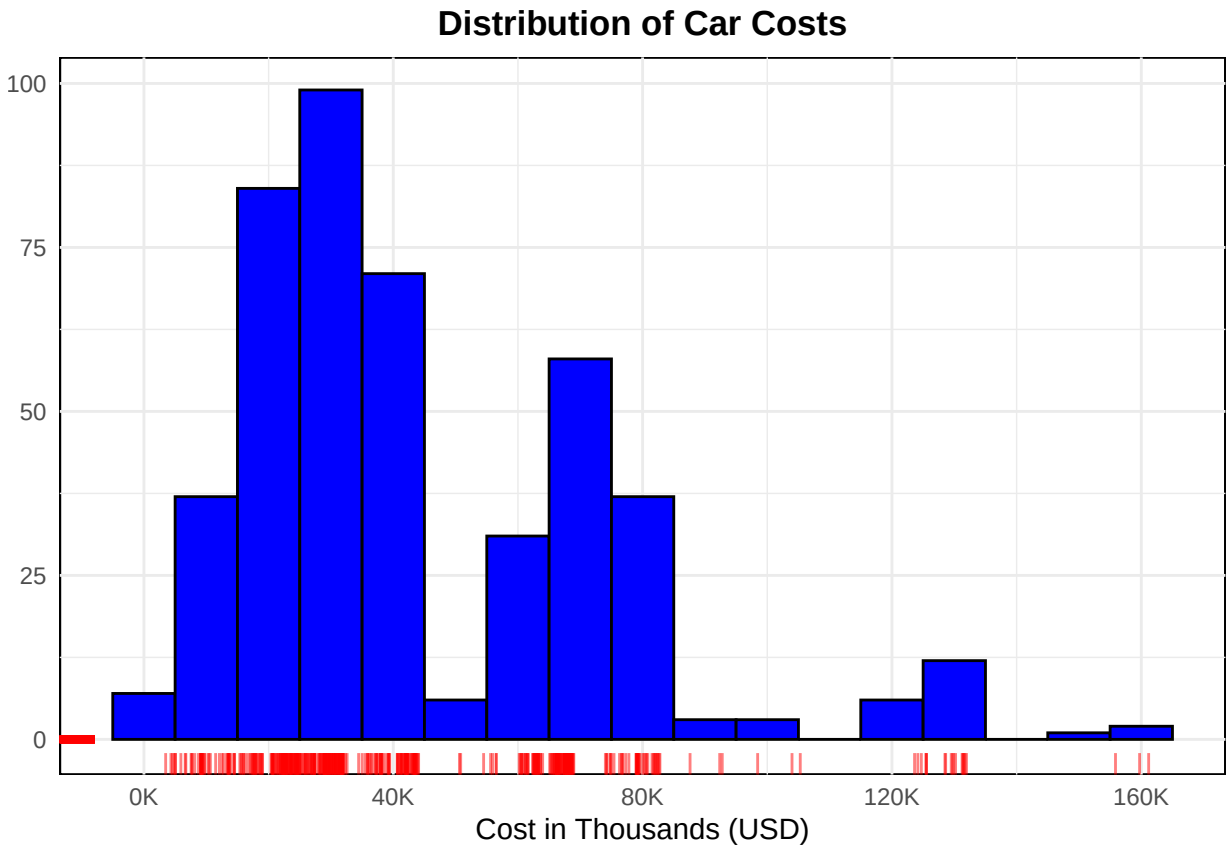



```
#########################################################
#Question 5 - Pie Charts
#########################################################

car_sales_total <- car_data %>%
  group_by(VehicleType) %>%
  summarize(total_sales = sum(SalePrice))

# Calculate percentage of total sales
car_sales_total <- car_sales_total %>%
  mutate(percentage = total_sales / sum(total_sales) * 100)

# Create a pie chart with informative slice labels
ggplot(car_sales_total, aes(x = "", y = total_sales, fill = VehicleType)) +
  geom_bar(stat = "identity", color = "white") +
  coord_polar(theta = "y") +
  labs(title = "Total Sales by Car Type") +
  theme_minimal() +
  theme(
    plot.title = element_text(hjust = 0.5, face = "bold"),
    axis.text = element_blank(),  # Remove axis text
    axis.ticks = element_blank(),  # Remove axis ticks
    axis.title = element_blank(),  # Remove axis titles
    panel.grid = element_blank(),  # Remove grid lines
    panel.background = element_rect(fill = "white", color = "white")
  ) +
  geom_text(aes(label = paste(VehicleType, sprintf("%.1f%%", percentage)),
                y = total_sales,
```

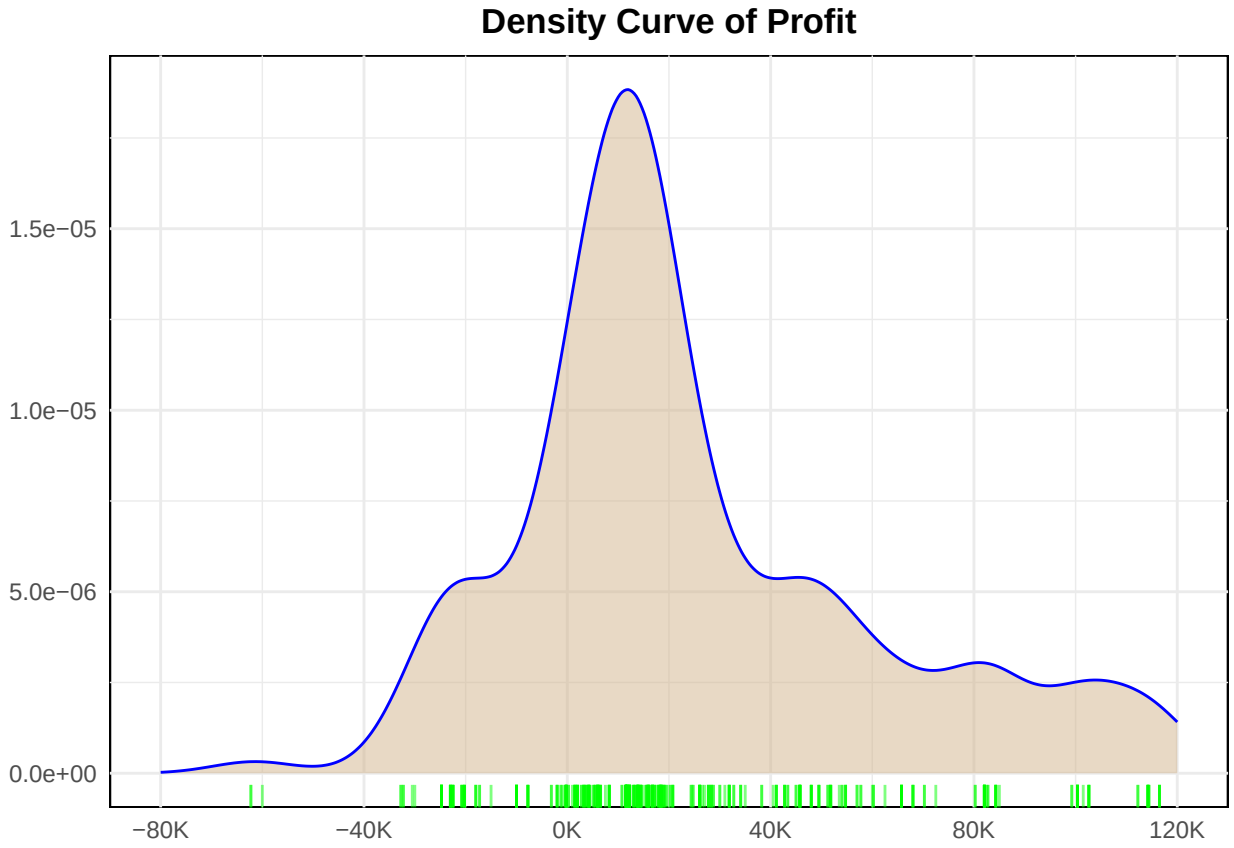
```
            group = VehicleType),
          position = position_stack(vjust = 0.5),
          color = "black", size = 3)
```



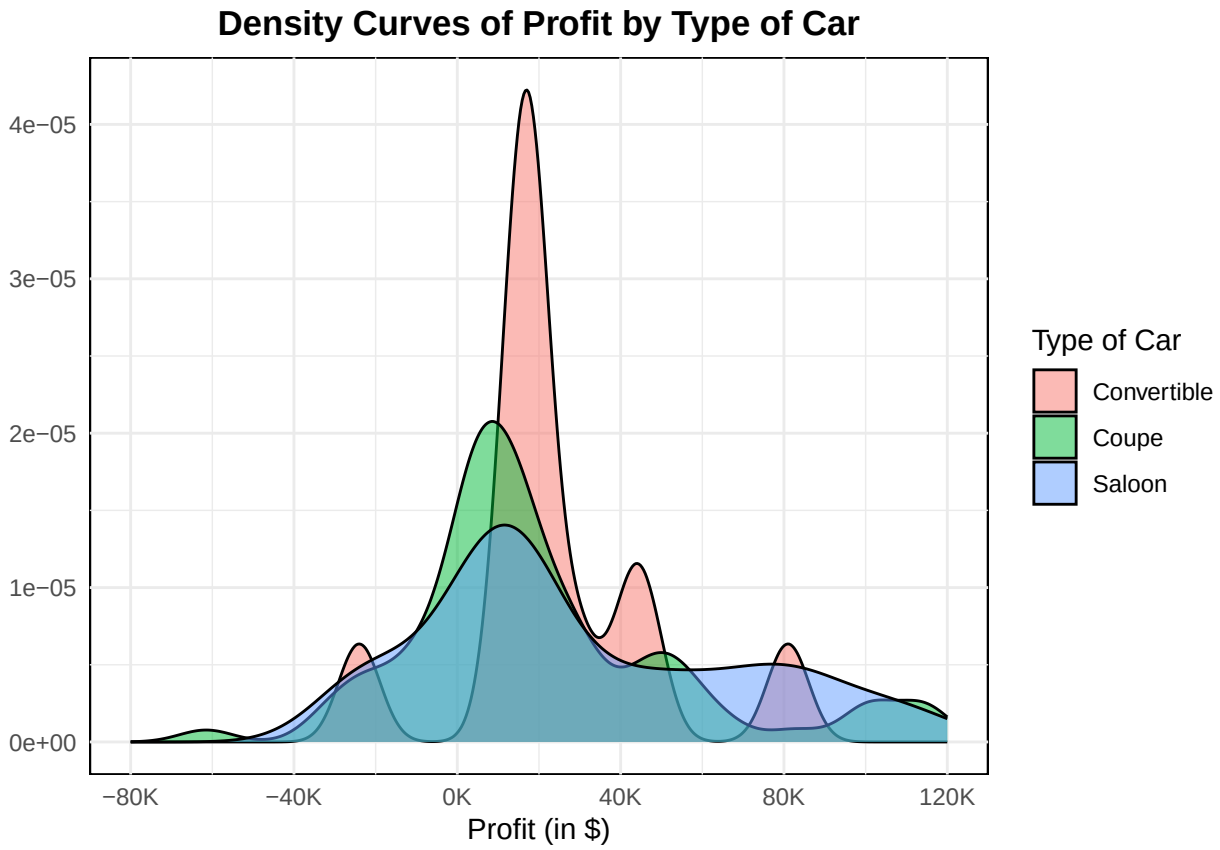

```
#####################################################
#Question 6 - Histograms
#####################################################

bin <- 10000

ggplot(car_data, aes(x = CostPrice)) +
  geom_histogram(binwidth = bin, fill = "blue", color = "black") +
  geom_rug(mapping = aes(y = 0), position = position_jitter(width = 2000),
           alpha = 0.5, color = "red") +
  labs(
    title = "Distribution of Car Costs",
    x = "Cost in Thousands (USD)",
    y = NULL) +
  scale_x_continuous(
    breaks = seq(0, max(car_data$CostPrice), by = 40000),
    labels = scales::comma_format(scale = 1e-3, suffix = "K")
  ) +
  theme_minimal() +
  theme(
    plot.title = element_text(hjust = 0.5, face = "bold"),
    panel.background = element_rect(fill = "white", color = "black")
  )
```




```
#########################################################
#Question 7 - Density Curves
#########################################################

profit_vector <- with(car_data, SalePrice - CostPrice)

car_data$Profit <- profit_vector


ggplot(car_data, aes(x = Profit)) +
  geom_density(fill = "tan", color = "blue", alpha = 0.5) +
  geom_rug(color = alpha("green", 0.5), sides = "b") +
  labs(title = "Density Curve of Profit", x = NULL, y = NULL) +
  scale_x_continuous(
    limits = c(floor(min(car_data$Profit, na.rm = TRUE) / 40000) * 40000,
               ceiling(max(car_data$Profit, na.rm = TRUE) / 40000) * 40000),
    breaks = seq(floor(min(car_data$Profit, na.rm = TRUE) / 40000) * 40000,
                 ceiling(max(car_data$Profit, na.rm = TRUE) / 40000) * 40000,
                 40000),
    labels = scales::comma_format(scale = 1e-3, suffix = "K")
  ) +
  theme_minimal() +
  theme(
    plot.title = element_text(hjust = 0.5, face = "bold"),
    panel.background = element_rect(fill = "white", color = "black")
  )
```




```
#########################################################
#Question 8 - Multiple Density Curves
#########################################################


ggplot(car_data, aes(x = Profit, fill = VehicleType)) +
  geom_density(alpha = 0.5) +
  labs(title = "Density Curves of Profit by Type of Car", x = "Profit (in $)",
       y = NULL, fill = "Type of Car") +
  scale_x_continuous(
    limits = c(floor(min(car_data$Profit, na.rm = TRUE) / 40000) * 40000,
               ceiling(max(car_data$Profit, na.rm = TRUE) / 40000) * 40000),
    breaks = seq(floor(min(car_data$Profit, na.rm = TRUE) / 40000) * 40000,
                 ceiling(max(car_data$Profit, na.rm = TRUE) / 40000) * 40000,
                 40000),
    labels = scales::comma_format(scale = 1e-3, suffix = "K")
  ) +
  theme_minimal() +
  theme(
    plot.title = element_text(hjust = 0.5, face = "bold"),
    axis.title.y = element_blank(),
    panel.background = element_rect(fill = "white", color = "black")
  )
```




```
#######################################################
#Question 9 - Box Plots
#######################################################

set.seed(123)
car_data <- data.frame(
  CarType = rep(c("Sedan", "SUV", "Truck"), each = 50),
  SalePrice = c(rnorm(50, mean = 30000, sd = 5000),
                rnorm(50, mean = 35000, sd = 6000),
                rnorm(50, mean = 40000, sd = 7000))
)

# Create the box plot
p <- ggplot(car_data, aes(x = CarType, y = SalePrice, fill = CarType)) +
  geom_boxplot(color = "gray", outlier.shape = 1, outlier.fill = "white") +
  stat_boxplot(geom = "errorbar", width = 0.5)

# Customize the plot
p + theme_minimal() +
  theme(
    panel.background = element_rect(fill = "white", color = "black"),
    axis.text.y = element_text(size = 10),
    axis.ticks.y = element_line(color = "black", size = 0.5),
    axis.title.y = element_text(size = 12),
    axis.title.x = element_blank(),
    plot.title = element_text(face = "bold", size = 14, hjust = 0.5),
    legend.position = "none"
```

```
  ) +
  scale_y_continuous(labels = scales::dollar_format(scale = 1e-3, suffix = "K"),
                     breaks = seq(0, max(car_data$SalePrice), by = 40000)) +
  labs(
    title = "Comparison of Sale Prices for Different Car Types",
    y = "Sale Price (in Thousands)"
  )
```

```
## Warning: The `size` argument of `element_line()` is deprecated as of ggplot2 3.4.0.
## i Please use the `linewidth` argument instead.
## This warning is displayed once every 8 hours.
## Call `lifecycle::last_lifecycle_warnings()` to see where this warning was
## generated.
```