CIA-3: Descriptive Analysis of Air Quality Dataset

Zishan Mallick

Contents

1 A. Problem Statement and Objectives

1.0.1 Problem Area:

Air pollution is a significant environmental issue. Understanding metrics like ozone concentration and its relationship with temperature, wind, and solar radiation can help in environmental policy and planning.

1.0.2 Objective:

To perform a descriptive analysis of the **airquality** dataset using R, identify patterns, and interpret results with social and ethical context.

---

2 B. Descriptive Analytics and Insights

```r
# Load the airquality dataset
data(airquality)

# Add Month as a factor for better plotting
airquality <- airquality %>%
  mutate(Month_Factor = factor(Month, labels = month.abb[5:9]))
```

2.1 Descriptive Statistics

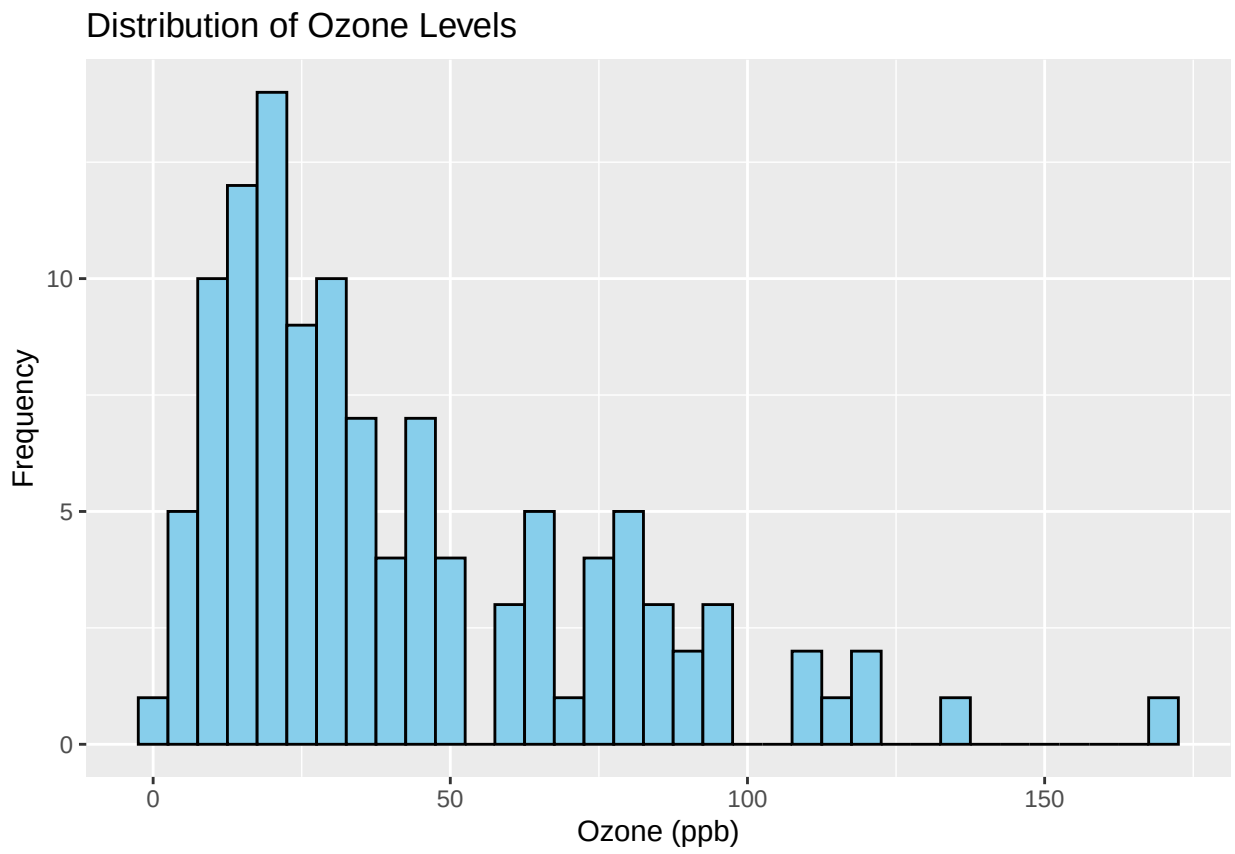
```r
# Summary stats for numeric variables
desc_stats <- airquality %>%
  summarise(across(where(is.numeric), list(mean = ~round(mean(.x, na.rm = TRUE), 2),
                                           sd = ~round(sd(.x, na.rm = TRUE), 2))))
kable(desc_stats, caption = "Descriptive Statistics of Air Quality Variables") %>%
  kable_styling()
```

Table 1: Descriptive Statistics of Air Quality Variables

Ozone_mean	Ozone_sd	Solar.R_mean	Solar.R_sd	Wind_mean	Wind_sd	Temp_mean	Temp_sd	Month_
42.13	32.99	185.93	90.06	9.96	3.52	77.88	9.47	

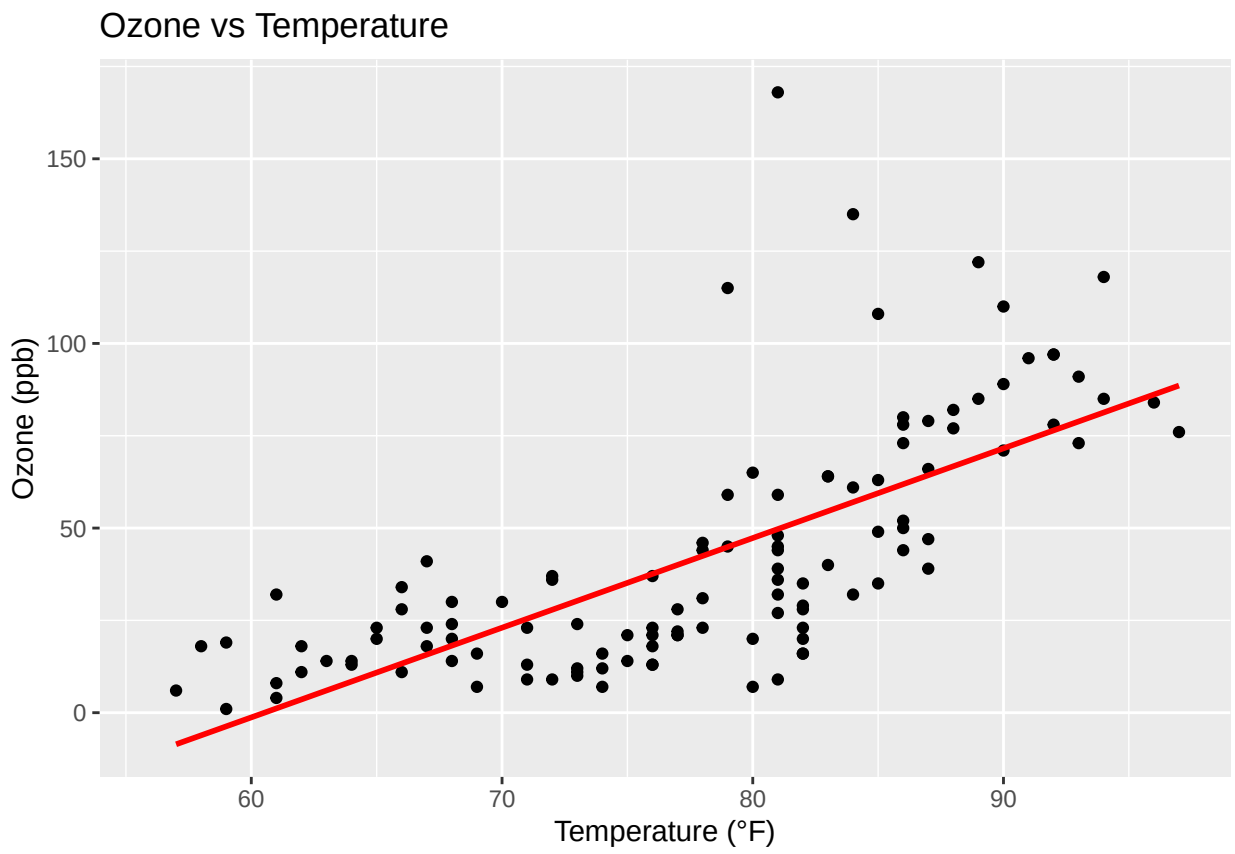
2.2 Histogram of Ozone Levels

```r
# Plot histogram of Ozone levels
ggplot(airquality, aes(x = Ozone)) +
  geom_histogram(binwidth = 5, fill = "skyblue", color = "black", na.rm = TRUE) +
  labs(title = "Distribution of Ozone Levels", x = "Ozone (ppb)", y = "Frequency")
```



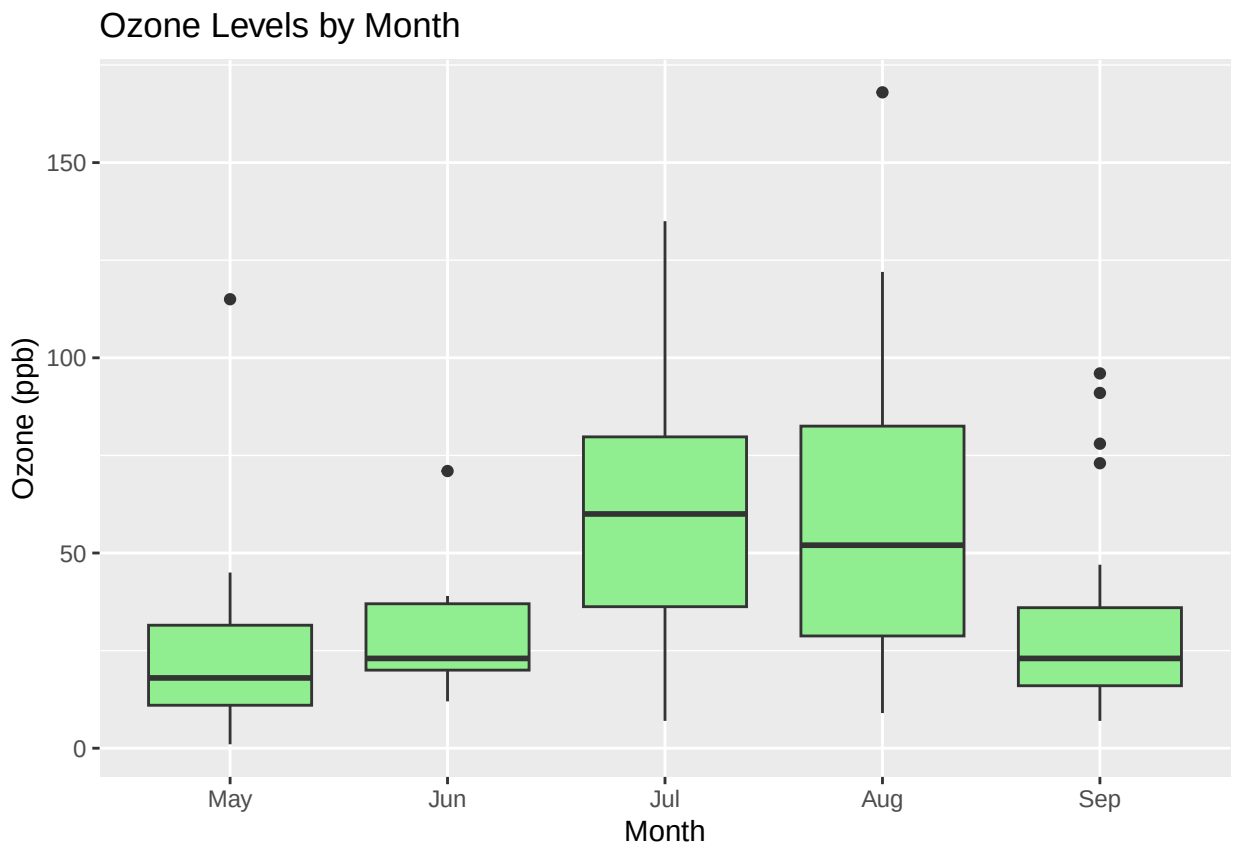
2.3 Scatter Plot: Ozone vs Temperature

```r
# Plot ozone vs temperature
ggplot(airquality, aes(x = Temp, y = Ozone)) +
  geom_point(na.rm = TRUE) +
  geom_smooth(method = "lm", se = FALSE, color = "red", na.rm = TRUE) +
  labs(title = "Ozone vs Temperature", x = "Temperature (°F)", y = "Ozone (ppb)")
```



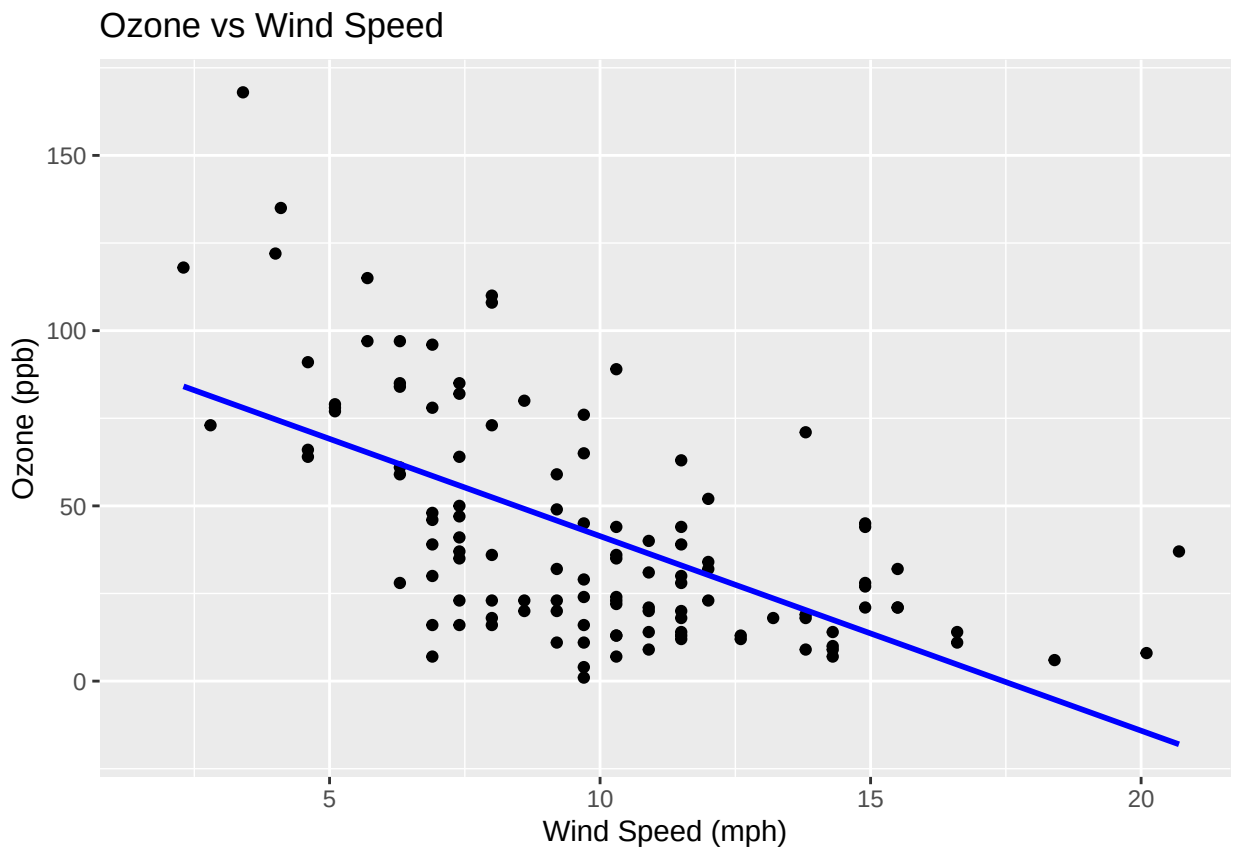
2.4 Boxplot: Ozone by Month

```r
# Boxplot of Ozone levels by Month
ggplot(airquality, aes(x = Month_Factor, y = Ozone)) +
  geom_boxplot(fill = "lightgreen", na.rm = TRUE) +
  labs(title = "Ozone Levels by Month", x = "Month", y = "Ozone (ppb)")
```



2.5 Scatter Plot: Ozone vs Wind

```r
# Ozone vs Wind plot
ggplot(airquality, aes(x = Wind, y = Ozone)) +
  geom_point(na.rm = TRUE) +
  geom_smooth(method = "lm", se = FALSE, color = "blue", na.rm = TRUE) +
  labs(title = "Ozone vs Wind Speed", x = "Wind Speed (mph)", y = "Ozone (ppb)")
```



2.6 Correlation Matrix

```
# Correlation between numeric variables
cor_matrix <- cor(airquality[, c("Ozone", "Solar.R", "Wind", "Temp")], use = "pairwise.complete.obs")
kable(round(cor_matrix, 2), caption = "Correlation Matrix of Air Quality Variables") %>%
  kable_styling()
```

Table 2: Correlation Matrix of Air Quality Variables

	Ozone	Solar.R	Wind	Temp
Ozone	1.00	0.35	-0.60	0.70
Solar.R	0.35	1.00	-0.06	0.28
Wind	-0.60	-0.06	1.00	-0.46
Temp	0.70	0.28	-0.46	1.00

---

3 C. Key Insights

- Ozone increases with temperature and decreases with wind speed.

- Seasonal variation shows July–August have higher ozone levels.
- Wind may help disperse ozone and reduce pollution levels.

---

4 D. Ethical and Social Considerations

- Urban areas with higher heat and stagnant wind conditions may face increased health risks.
- Making air quality data public can promote environmental justice and help inform vulnerable communities.
- Visualizations like these can be used in policy decisions to trigger alerts and mitigation efforts.

---

5 E. Conclusion and Recommendations

- Monitor ozone levels during warmer months.
- Improve green infrastructure to counter heat-induced ozone rise.
- Promote wind corridor planning to disperse pollution.

---

6 Appendix: Wireframe Summary

- **Data**: `airquality`
- **Tools**: R, ggplot2, dplyr, RMarkdown
- **Outputs**: Summary stats, plots, interpretation, ethical discussion