Homework 1

AUTHOR
Mani Jose 24222863

Introduction

The ***airquality*** dataset is built into R and contains daily air quality measurements in New York City from May to September 1973.The goal of collecting this data was to monitor environmental conditions and understand how factors like temperature, wind, and solar radiation impact air pollution, particularly ozone levels, which are harmful to health.

The airquality dataset contains 153 observations, each representing a single day, it includes 6 variables :

Variable	Description
Ozone	Ozone concentration (ppb)
Solar.R	Solar radiation (Langley)
Wind	Wind speed (mph)
Temp	Maximum daily temperature (°F)
Month	Month of the observation
Day	Day of the month

The objective of this analysis is to explore the airquality dataset by visualizing key weather and pollution variables including:

1. Use boxplots to compare ozone level across different months, revealing seasonal patterns and variability in air quality and weather.
2. Analyze daily trends in ozone concentration over time through a line plot, providing insight into how air pollution fluctuates across the months.

Basic summary statistics

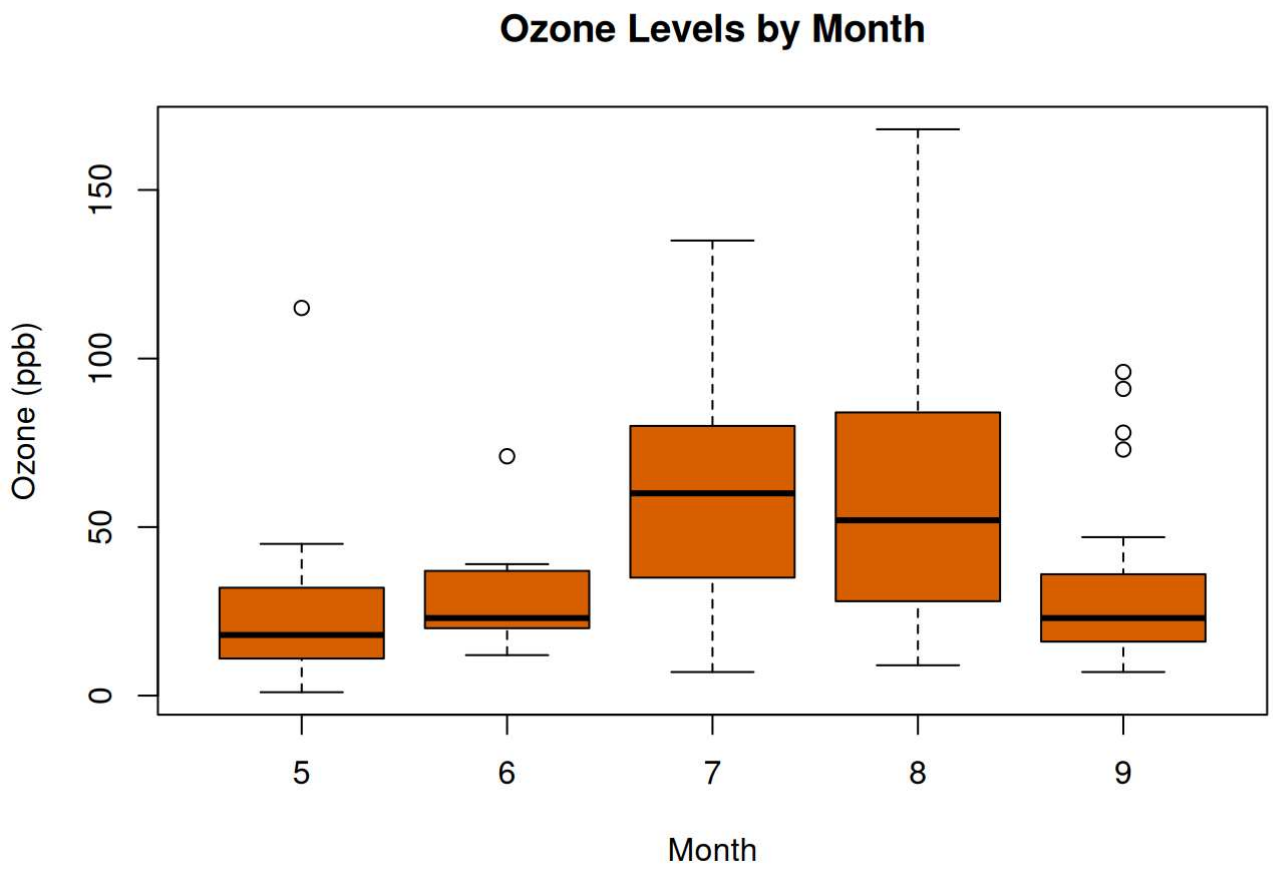
```
     Ozone           Solar.R          Wind             Temp      
 Min.   :  1.00   Min.   :  7.0   Min.   : 1.700   Min.   :56.00  
 1st Qu.: 18.00   1st Qu.:115.8   1st Qu.: 7.400   1st Qu.:72.00  
 Median : 31.50   Median :205.0   Median : 9.700   Median :79.00  
 Mean   : 42.13   Mean   :185.9   Mean   : 9.958   Mean   :77.88  
 3rd Qu.: 63.25   3rd Qu.:258.8   3rd Qu.:11.500   3rd Qu.:85.00  
 Max.   :168.00   Max.   :334.0   Max.   :20.700   Max.   :97.00  
 NA's   :37       NA's   :7                                       
     Month            Day       
 Min.   :5.000   Min.   : 1.0  
 1st Qu.:6.000   1st Qu.: 8.0  
 Median :7.000   Median :16.0  
 Mean   :6.993   Mean   :15.8  
 3rd Qu.:8.000   3rd Qu.:23.0  
```

```
Max.   :9.000   Max.   :31.0
```

- **Ozone levels** vary widely from a low of 1 ppb to a high of 168 ppb, with a median value of 31.5 ppb and a mean of 42.13 ppb. *The higher mean compared to the median indicates some days with unusually high ozone concentrations* pulling the average upward. However, *37 values are missing*, which could impact analysis accuracy.
- **Solar radiation** (Solar.R) ranges between 7.0 and 334.0 units, with a median of 205.0 and an average of 185.9. *There are 7 missing values*, suggesting some gaps in solar radiation measurement.
- **Wind speeds** *are fairly consistent*, spanning from 1.7 mph to 20.7 mph, with a median of 9.7 mph and a mean of 9.96 mph, *indicating generally moderate wind conditions.*
- **Temperatures** during the recorded period range from 56°F to 97°F, averaging 77.88°F with a median of 79°F, which *aligns with typical late spring and summer weather*.

Visualizations Of Ozone Levels

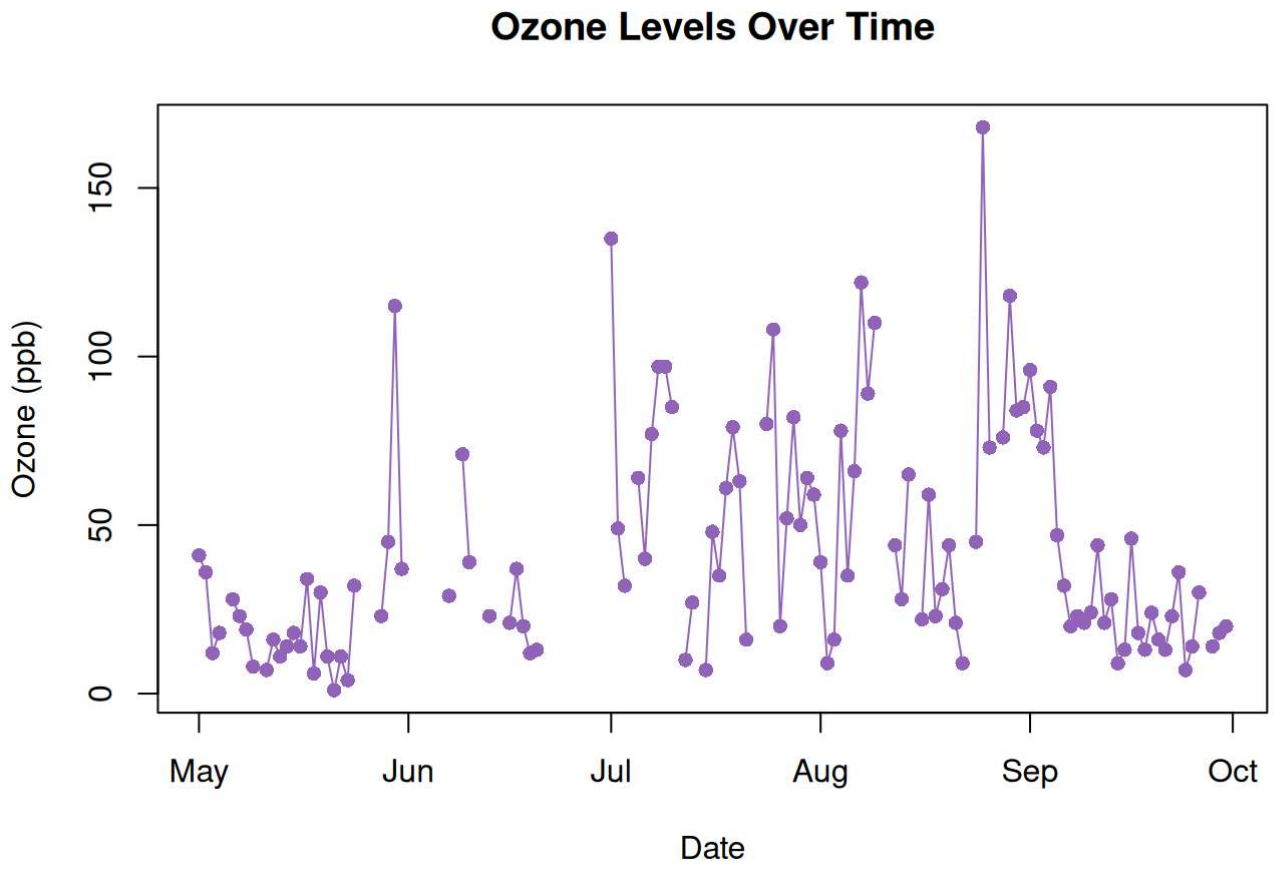
This boxplot shows how ozone concentration (ppb) varies across the months from May (5) to September (9):



- *Ozone levels are generally **lowest in May and June**, with relatively compact boxplots and fewer high values.*
- ***July and August show the highest median ozone** levels and the widest spread, indicating more variability and higher pollution during peak summer.*

- *September shows a decline in ozone levels again, similar to early summer months.*
- *Several **outliers** (circles) are present, especially **in August and September**, representing days with unusually high ozone levels.*

Ozone levels tend to peak in mid-summer, possibly due to stronger sunlight and higher temperatures which accelerate ozone formation.



This R code used generates a time series plot of daily ozone levels from the `airquality` dataset. It first creates a new `Date` variable by combining the year 1973 with the existing `Month` and `Day` columns, converting them into proper date objects. The `plot()` function then draws a line graph of ozone levels over time, the `points()` function adds solid dots at each valid observation.

Ozone levels tend to **peak during the warmest months (July–August)**, likely due to increased sunlight and temperature, which are known to enhance ozone formation. The drop-off in September suggests **seasonal effects** strongly influence ozone concentration.