Test 1 Memo

Dr. Niladri Chakraborty

2025-03-28

Q1. Use the dataset 'Occupancy_data.csv' from the 'Extra Resources' section in BlackBoard. Then answer the following question. Use the 'classic' theme as background for all the plots. (i) Create a boxplot to compare the distribution of CO2 levels between occupied and unoccupied rooms. (ii) Plot a histogram of the Humidity values. Use 20 bins and fill the bars in blue with white borders. (iii) Create a line chart of Temperature over time. (iv) Multiply 100 with all 'Temperature' values and save these values with the same variable name as 'Tamperature'. Plot both Temperature and CO2 over time on the same chart using different colors. (v) Create a boxplot showing the distribution of Light levels for different occupancy statuses.

Ans.

```
# Load libraries
library(tidyverse)
```

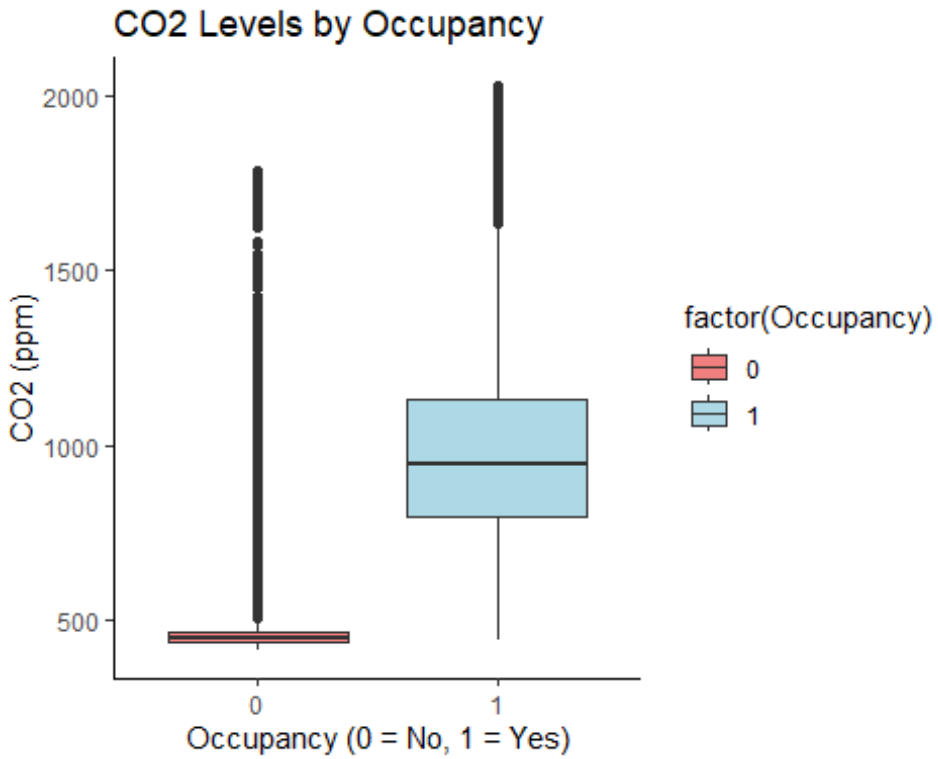
```
## ── Attaching core tidyverse packages ──────────────────────── tidyverse
2.0.0 ──
## ✓ dplyr     1.1.2     ✓ readr     2.1.4
## ✓ forcats   1.0.0     ✓ stringr   1.5.0
## ✓ ggplot2   3.4.2     ✓ tibble    3.2.1
## ✓ lubridate 1.9.2     ✓ tidyr     1.3.0
## ✓ purrr     1.0.1
## ── Conflicts ────────────────────────────────────────
tidyverse_conflicts() ──
## ✖ dplyr::filter() masks stats::filter()
## ✖ dplyr::lag()    masks stats::lag()
## ℹ Use the conflicted package (<http://conflicted.r-lib.org/>) to force all
conflicts to become errors
```

```
#library(readr)

# Read the data
data = read.csv("C:\\Users\\ChakrabortyN\\OneDrive - University of the Free
State\\Documents\\Occupancy_data.csv")
```

(i)

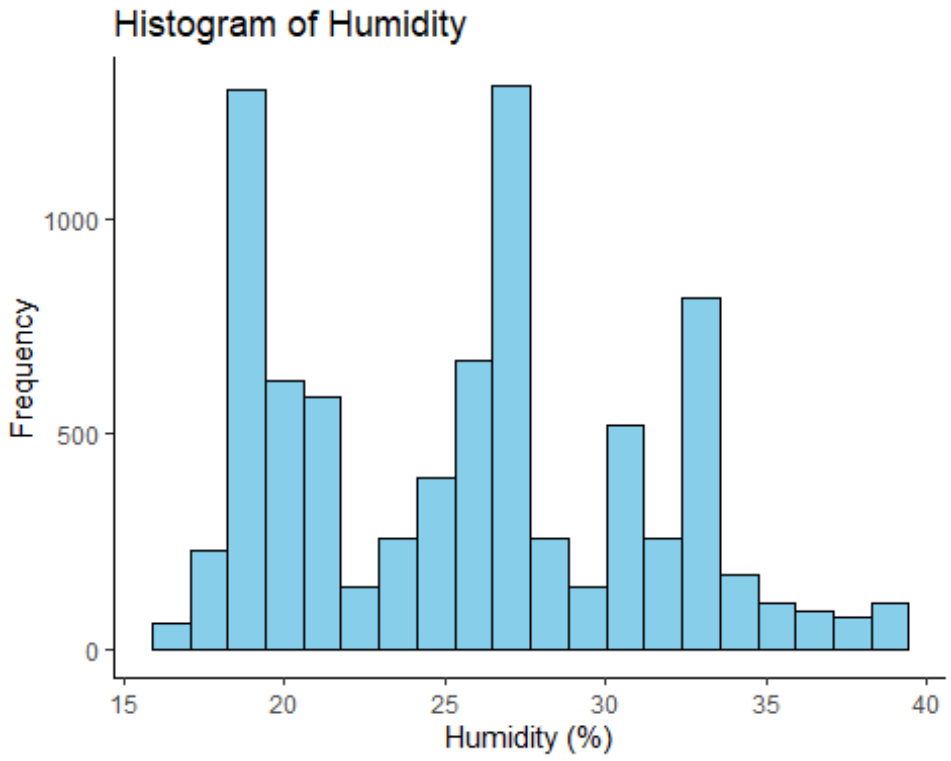
```
ggplot(data, aes(x = factor(Occupancy), y = CO2, fill = factor(Occupancy))) +
  geom_boxplot() +
  labs(title = "CO2 Levels by Occupancy",
       x = "Occupancy (0 = No, 1 = Yes)", y = "CO2 (ppm)") +
```

```
  scale_fill_manual(values = c("lightcoral", "lightblue")) +
  theme_classic()
```



(ii)

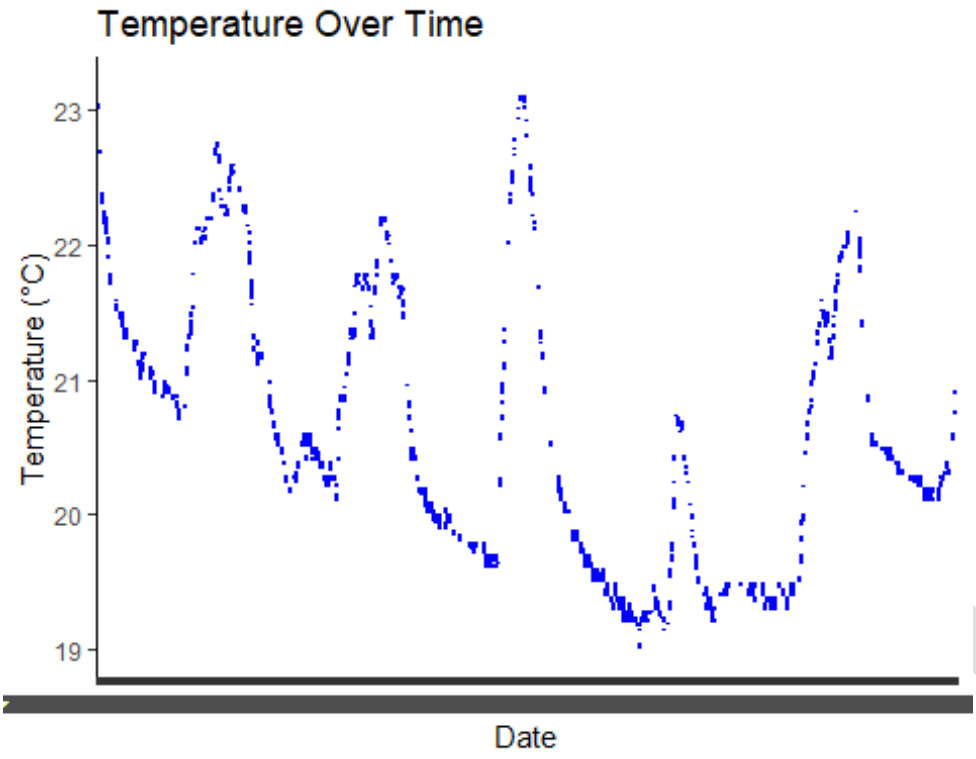
```
ggplot(data, aes(x = Humidity)) +
  geom_histogram(bins = 20, fill = "skyblue", color = "black") +
  labs(title = "Histogram of Humidity",
       x = "Humidity (%)", y = "Frequency") +
  theme_classic()
```



(iii)

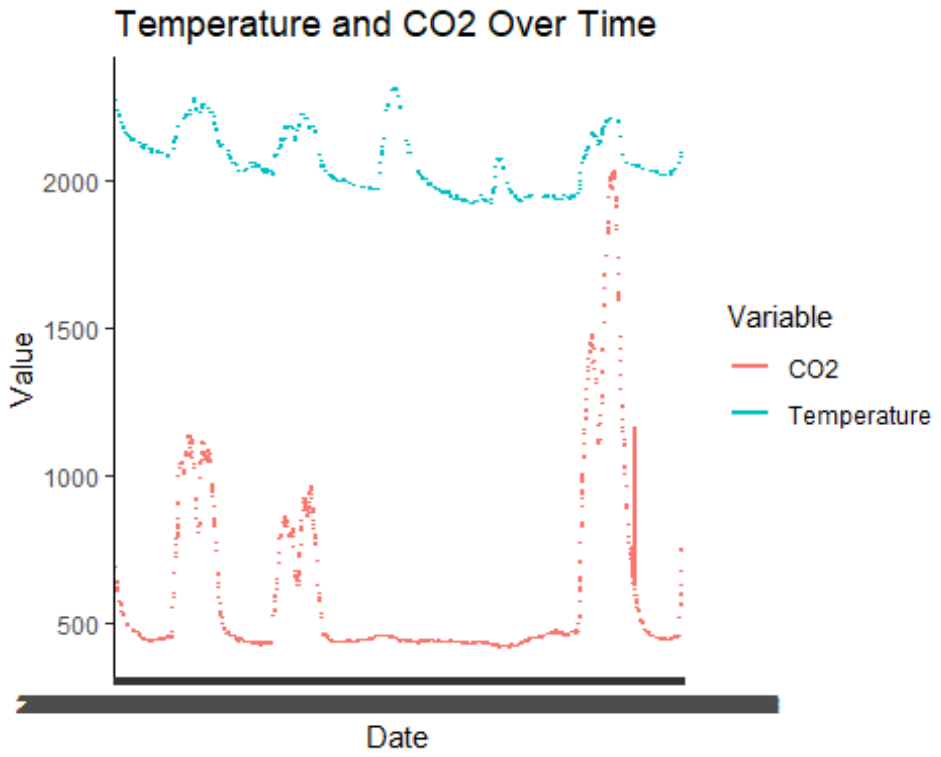
```
ggplot(data, aes(x = date, y = Temperature)) +
  geom_line(color = "blue", size = 1) +
  labs(title = "Temperature Over Time",
       x = "Date", y = "Temperature (°C)") +
  theme_classic()
```

```
## 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.
```



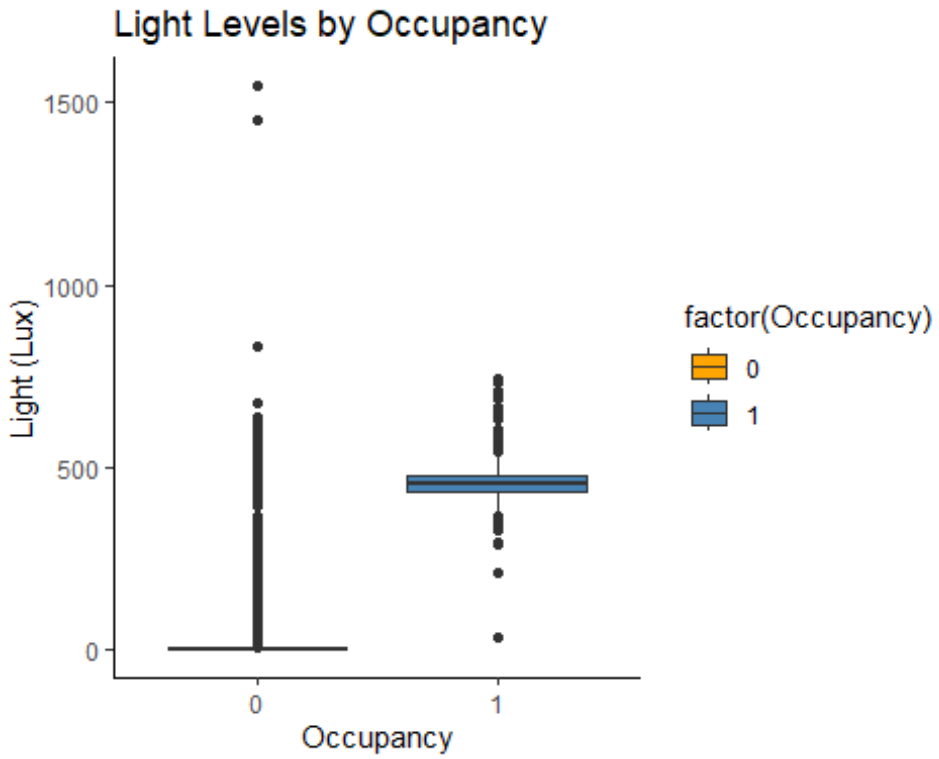
(iv)

```
library(tidyverse)

# Select and scale data
data_long = data %>%
  select(date, Temperature, CO2)

data_long$Temperature = data_long$Temperature * 100  # Scale temperature

# Reshape using gather()
data_long = gather(data_long, key = "Variable", value = "Value", Temperature,
CO2)


ggplot(data_long, aes(x = date, y = Value, color = Variable)) +
  geom_line(size = 1) +
  labs(title = "Temperature and CO2 Over Time",
       x = "Date", y = "Value") +
  theme_classic()
```



(v)

```
ggplot(data, aes(x = factor(Occupancy), y = Light, fill = factor(Occupancy)))
+
  geom_boxplot() +
  labs(title = "Light Levels by Occupancy",
       x = "Occupancy", y = "Light (Lux)") +
  scale_fill_manual(values = c("orange", "steelblue")) +
  theme_classic()
```



Q2. Use the 'mpg' dataset. (i) Now create a histogram of the 'displ' variable which is the engine displacement in liters. Add labels for the X and Y axis as 'Displacement' and 'Frequency', respectively. Also add a title for the histogram plot as 'Histogram of the engine displacement in liters'. The histogram bars should be filled with 'blue' color and the borders should be in 'black'. Use the 'classic' theme for the background.

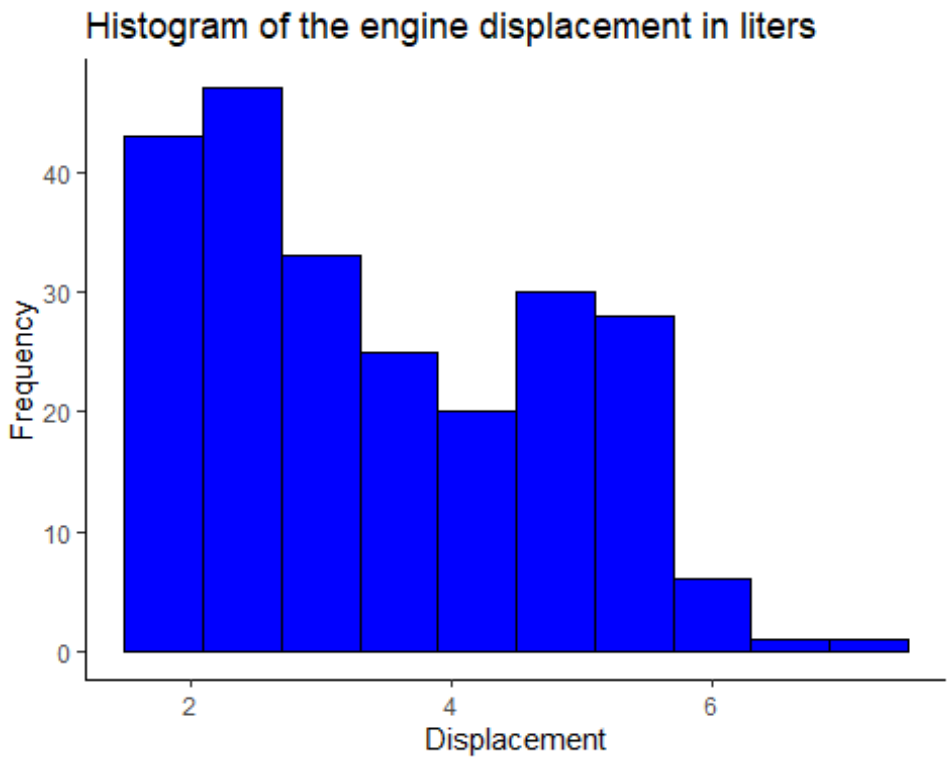
(ii) Create a subset of the 'mpg' data for the following manufacturers: corolla", "altima", "jetta", "maxima", "pathfinder 4wd", "grand prix". Now use the same subset of the 'mpg' data, then create boxplots for the 'displ' variable (Engine displacement in liters) for different manufacturers. All the boxplots should be in the same plot. Use different colors for different manufacturers. Provide a title of the boxplot as 'Engine displacement for different manufacturers'. Provide X and Y axis labels as 'Manufacturer' and 'Engine displacement in liters', respectively. Use the 'minimal' theme for the plot.

Ans.

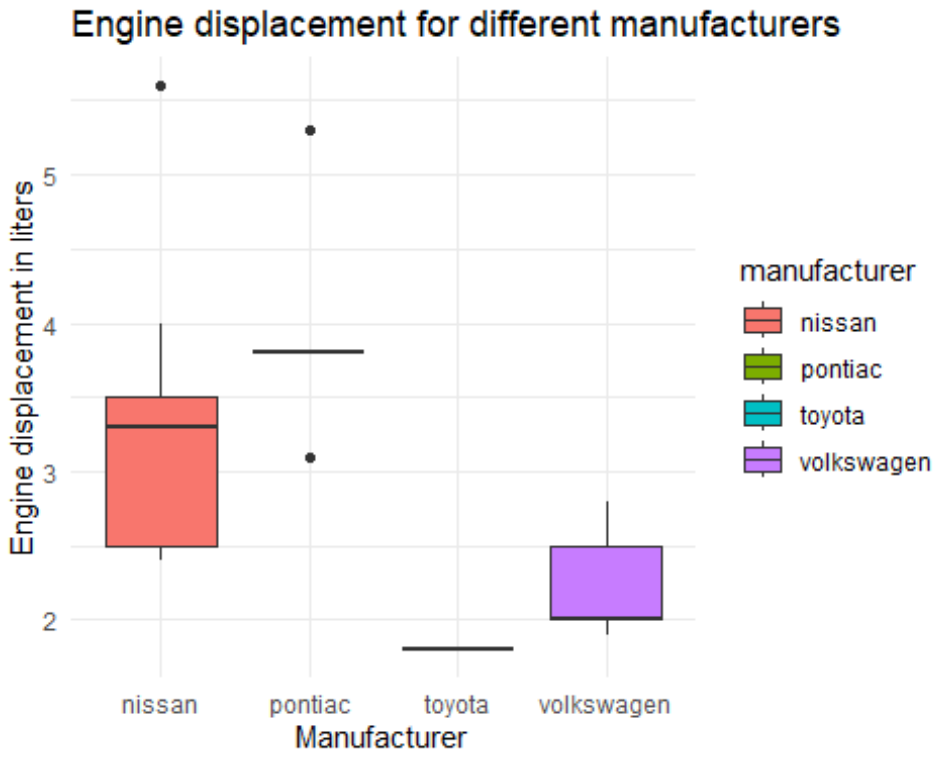
```
library(tidyverse)
data = mpg
str(data)
```

```
## tibble [234 × 11] (S3: tbl_df/tbl/data.frame)
##  $ manufacturer: chr [1:234] "audi" "audi" "audi" "audi" ...
##  $ model       : chr [1:234] "a4" "a4" "a4" "a4" ...
##  $ displ       : num [1:234] 1.8 1.8 2 2 2.8 2.8 3.1 1.8 1.8 2 ...
##  $ year        : int [1:234] 1999 1999 2008 2008 1999 1999 2008 1999 1999
```

```
2008 ...
##  $ cyl          : int [1:234] 4 4 4 4 6 6 6 4 4 4 ...
##  $ trans        : chr [1:234] "auto(l5)" "manual(m5)" "manual(m6)"
"auto(av)" ...
##  $ drv          : chr [1:234] "f" "f" "f" "f" ...
##  $ cty          : int [1:234] 18 21 20 21 16 18 18 18 16 20 ...
##  $ hwy          : int [1:234] 29 29 31 30 26 26 27 26 25 28 ...
##  $ fl           : chr [1:234] "p" "p" "p" "p" ...
##  $ class        : chr [1:234] "compact" "compact" "compact" "compact" ...

head(data)

## # A tibble: 6 × 11
##   manufacturer model displ  year   cyl trans      drv     cty   hwy fl
class
##   <chr>        <chr> <dbl> <int> <int> <chr>      <chr> <int> <int> <chr>
<chr>
## 1 audi         a4      1.8  1999     4 auto(l5)   f        18    29 p
compa…
## 2 audi         a4      1.8  1999     4 manual(m5) f        21    29 p
compa…
## 3 audi         a4      2    2008     4 manual(m6) f        20    31 p
compa…
## 4 audi         a4      2    2008     4 auto(av)   f        21    30 p
compa…
## 5 audi         a4      2.8  1999     6 auto(l5)   f        16    26 p
compa…
## 6 audi         a4      2.8  1999     6 manual(m5) f        18    26 p
compa…

disp <- data$displ
# Plot histogram
ggplot(data, aes(x = disp)) +
  geom_histogram(bins = 10, fill = "blue", color = "black") +
  labs(title = "Histogram of the engine displacement in liters", x =
"Displacement", y = "Frequency")+ theme_classic()
```




```r
library(dplyr)

# Create a subset of the dataset for specified models
selected_models <- c("corolla", "altima", "jetta", "maxima", "pathfinder
4wd", "grand prix")
subset_data <- mpg %>%
  filter(model %in% selected_models)

# Plot boxplot
ggplot(subset_data, aes(x = manufacturer, y = displ, fill = manufacturer)) +
  geom_boxplot() +
  labs(title = "Engine displacement for different manufacturers", x =
"Manufacturer", y = "Engine displacement in liters") +
  theme_minimal()
```



Q3. Use the 'airquality' dataset. Create a dataframe with the variables 'Ozone', 'Solar.R', 'Temp' from the dataset. Now create a correlation plot for these variables using the GGally package. Interpret the correlation plot.

```
[8+2 = 10]
```

```
library(GGally)
```

```
## Warning: package 'GGally' was built under R version 4.3.1
```

```
## Registered S3 method overwritten by 'GGally':
##   method from
##   +.gg   ggplot2
```

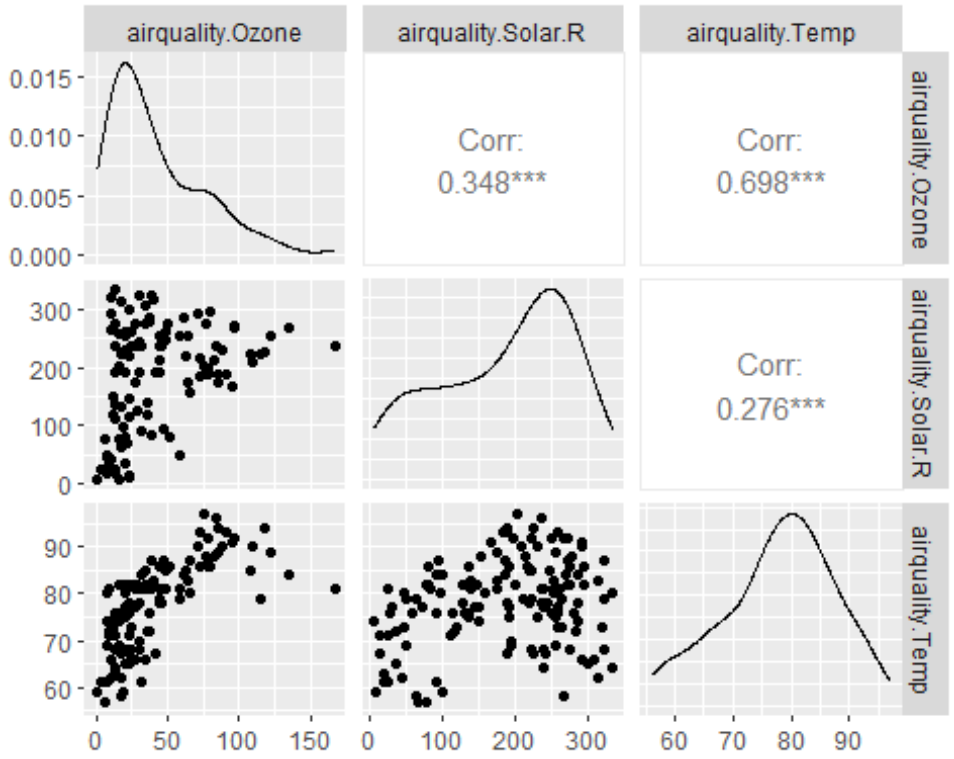
```
data("airquality")
data = data.frame(airquality$Ozone,airquality$Solar.R,airquality$Temp)
ggpairs(data)
```

```
## Warning: Removed 37 rows containing non-finite values (`stat_density()`).
```

```
## Warning in ggally_statistic(data = data, mapping = mapping, na.rm = na.rm,
:
## Removed 42 rows containing missing values
```

```
## Warning in ggally_statistic(data = data, mapping = mapping, na.rm = na.rm,
:
## Removed 37 rows containing missing values
```

```
## Warning: Removed 42 rows containing missing values (`geom_point()`).

## Warning: Removed 7 rows containing non-finite values (`stat_density()`).

## Warning in ggally_statistic(data = data, mapping = mapping, na.rm = na.rm,
:
## Removed 7 rows containing missing values

## Warning: Removed 37 rows containing missing values (`geom_point()`).

## Warning: Removed 7 rows containing missing values (`geom_point()`).
```



Strong positive correlations are evident among all the variables, suggesting that an increase in one variable tends to be associated with an increase in the others. The distribution of the Ozone levels shows a positive skew, while the distributions of the other two variables exhibit negative skews.

Q4. Write a code using a 'for' loop to compute the sum of all even numbers from 1 to 100. [10]

```r
# Initialize the sum variable
sum_evens <- 0

# For loop to sum even numbers from 1 to 100
for (i in 1:100) {
  if (i %% 2 == 0) {  # Check if the number is even
    sum_evens <- sum_evens + i
```

```
  }
}

sum_evens

## [1] 2550
```

Q5. Write a code using a ‘while’ loop that finds the smallest integer such that its square is greater than 1000. The program should print the integer and its square. [10]

```
x <- 1

while (x^2 <= 1000) {
  x <- x + 1
}

cat("The smallest integer whose square is greater than 1000 is:", x, "\n")

## The smallest integer whose square is greater than 1000 is: 32

cat("Its square is:", x^2, "\n")

## Its square is: 1024
```