

```

> # Q2. Using function data() to load R data set air quality
> data(airquality)           # Load the built-in dataset
> # Q2A - How many variables are there? How many observations are there?
> ncol(airquality)           # Number of variables
[1] 6
> nrow(airquality)           # Number of observations
[1] 153
> print(paste("The airquality dataset contains", ncol(airquality), "variables
and", nrow(airquality), "observations."))
[1] "The airquality dataset contains 6 variables and 153 observations."
> # Q2B - Are there missing values in the data set? If so, clean up the data
by removing those observations with missing values
> sum(is.na(airquality))     # Total number of missing values
[1] 44
> airquality_no_na <- na.omit(airquality) # Remove rows with any NA values
> nrow(airquality_no_na)     # Number of rows after removing NAs
[1] 111
> print(paste("The cleaned airquality dataset contains",
nrow(airquality_no_na), "observations after removing rows with missing
values."))
[1] "The cleaned airquality dataset contains 111 observations after removing
rows with missing values."
> # Q2C - Show the mean, median, mode, and 1st and 3rd quartiles of variables
Temp and Wind in the cleaned data set.
> # ----- TEMP -----
> mean_temp <- mean(airquality_no_na$Temp)
> median_temp <- median(airquality_no_na$Temp)
> temp_freq <- table(airquality_no_na$Temp)
> mode_temp <- as.numeric(names(temp_freq[temp_freq == max(temp_freq)]))
> q_temp <- quantile(airquality_no_na$Temp, c(0.25, 0.75))
> print(paste("Temp - Mean:", mean_temp))
[1] "Temp - Mean: 77.7927927927928"
> print(paste("Temp - Median:", median_temp))
[1] "Temp - Median: 79"
> print(paste("Temp - Mode:", paste(mode_temp, collapse = ", ")))
[1] "Temp - Mode: 81"
> print(paste("Temp - 1st Quartile:", q_temp[1]))
[1] "Temp - 1st Quartile: 71"
> print(paste("Temp - 3rd Quartile:", q_temp[2]))
[1] "Temp - 3rd Quartile: 84.5"
> # ----- WIND -----
> mean_wind <- mean(airquality_no_na$Wind)
> median_wind <- median(airquality_no_na$Wind)
> wind_freq <- table(airquality_no_na$Wind)
> mode_wind <- as.numeric(names(wind_freq[wind_freq == max(wind_freq)]))
> q_wind <- quantile(airquality_no_na$Wind, c(0.25, 0.75))
> print(paste("Wind - Mean:", mean_wind))
[1] "Wind - Mean: 9.93963963963964"
> print(paste("Wind - Median:", median_wind))

```

```

[1] "Wind - Median: 9.7"
> print(paste("Wind - Mode:", paste(mode_wind, collapse = ", ")))
[1] "Wind - Mode: 10.3, 11.5"
> print(paste("Wind - 1st Quartile:", q_wind[1]))
[1] "Wind - 1st Quartile: 7.4"
> print(paste("Wind - 3rd Quartile:", q_wind[2]))
[1] "Wind - 3rd Quartile: 11.5"
> # Q2D - In the cleaned data set, how many days with a Temp higher than 78,
and how many days with a Wind lower than 9?
> temp_above_78 <- sum(airquality_no_na$Temp > 78)
> wind_below_9 <- sum(airquality_no_na$Wind < 9)
> print(paste("Number of days with Temp > 78:", temp_above_78))
[1] "Number of days with Temp > 78: 57"
> print(paste("Number of days with Wind < 9:", wind_below_9))
[1] "Number of days with Wind < 9: 43"
> # Q2E - Add a new column Celsius to the cleaned data set, showing
temperature in Celsius.
> airquality_no_na$Celsius <- (airquality_no_na$Temp - 32) * 5 / 9
> # Show the first few rows to confirm the new column
> head(airquality_no_na)
  Ozone Solar.R Wind Temp Month Day  Celsius
1    41     190  7.4   67     5   1  19.44444
2    36     118  8.0   72     5   2  22.22222
3    12     149 12.6   74     5   3  23.33333
4    18     313 11.5   62     5   4  16.66667
7    23     299  8.6   65     5   7  18.33333
8    19      99 13.8   59     5   8  15.00000
> print("The cleaned dataset now includes a new column Celsius (temperature in
Celsius).")
[1] "The cleaned dataset now includes a new column Celsius (temperature in
Celsius)."
```