Lecture 4: Data Structures in R

In-Class Activities

1. Answer the following questions using the built-in R dataset `airquality`:

   (a) Create a new column called `date` that combines `Month` and `Day` from `airquality`, assuming the year is 1973 (use `as.Date()` and `paste0()` functions).

   (b) Create a new logical column called `windy` that is `TRUE` when `Wind > 10` and `FALSE` otherwise.

   (c) Display the observations where `Temp > 85` or `windy` is `TRUE`.

   (d) Display the observations where `Temp > 85` and `windy` is `TRUE`.

   (e) Create a second data frame called `late_summer_df` by filtering `airquality` to only months August or September (`Month` in {8,9}).

   (f) Print out the occurrences in `airquality` when the month was either May or August and the ozone level was less than 50. (Write your condition so it means "(May or August) and Ozone < 50".)

   (g) Print out the occurrences in `airquality` when the month was May, or the month was August with an ozone level less than 50. (Write your condition so it means "May, or (August and Ozone < 50)".)

   (h) Find the number of observations in `airquality` where `windy` is `TRUE`.

   (i) Find the mean temperature (`Temp`) for `windy` observations in `airquality`.

2. Answer the following questions using the built-in R dataset `mtcars`:

   (a) Create a matrix called `car_mat` that contains only the columns `mpg`, `hp`, and `wt`. Make sure it is stored as a matrix (not a dataframe).

   (b) Extract the submatrix consisting of rows 2 through 6 and columns 1 and 3.

   (c) Display the rows of `car_mat` where `hp > 150` and `wt < 3`. Use `which()` to get the row indices first.

   (d) Replace every value of `hp` that is greater than 200 with 200.

   (e) Create a list called `car_list` with the following components:

   - `raw` = the full `mtcars` dataset
   - `mat` = your updated `car_mat`
   - `heavy` = the row names of the cars with `wt > 3`
   - `averages` = a vector containing the mean of `mpg` and the mean of `wt`.

   (f) Use `str()` to inspect `car_list`

   (g) Access and print (a) the `averages` component, (b) the mean `mpg` value inside `averages`, and (c) the second entry of the `heavy` component