

HW 1, STAT 450

Due: Friday, September 6

Directions: The assignment should be completed using Quarto and submitted to Canvas as a self-contained HTML or PDF file.

Exercise 1

- (a) What are the four common data types of vectors in R?
- (b) Determine the data type of each vector below. Print out the value of each vector and use the `class()` function.

```
a <- c(1, 2, 3, 4, 5)
b <- 1:5
c <- c("blue", "orange", "red")
d <- c(T, T, T, T)
e <- c("1", 2, 3)
f <- c(7, NA, NA, 5, 3)
g <- c(7, "NA", "NA", 5, 3)
h <- c()
```

Exercise 2

Use R to calculate the sum of the squares of all integers from 1 to 100:

$$1^2 + 2^2 + 3^2 + \cdots + 100^2$$

Exercise 3

R has a built-in vector called `rivers` which contains the lengths of 141 major North American rivers. Type `help(rivers)` in the console to learn about the data set.

```
# print out first 6 river lengths
head(rivers)
```

```
## [1] 735 320 325 392 524 450
```

- (a) Find the mean and standard deviation of the river lengths.
- (b) What is the longest river length? What is the shortest river length?
- (c) How many rivers in the data set are longer than 1000 miles?
- (d) What percent of rivers in the data set are longer than 1000 miles?

Exercises 4 and 5 will use the `airquality` data frame, which is already loaded into R.

```
head(airquality)
```

```
##   Ozone Solar.R Wind Temp Month Day
## 1    41     190  7.4   67     5   1
## 2    36     118  8.0   72     5   2
## 3    12     149 12.6   74     5   3
## 4    18     313 11.5   62     5   4
## 5    NA       NA 14.3   56     5   5
## 6    28       NA 14.9   66     5   6
```

The data frame contains daily air quality measurements in New York from May to September 1973. Type `help(airquality)` in the console to read about this data in the help menu.

Exercise 4

Run the following code to subset the `Ozone` column and assign it to a variable called `Ozone1`.

```
Ozone1 <- airquality$Ozone
```

- (a) Use `is.na()` to remove the missing data (NA values) from the vector `Ozone1`. Assign the vector with the missing values removed to a variable called `Ozone2`. How many NA values were removed?
- (b) Compute the min, median, mean, max, and standard deviation of the numeric vector `Ozone2`.
- (c) Run the following commands, and explain how each command handles missing data.

```
summary(airquality$Ozone)
sd(airquality$Ozone)
sd(airquality$Ozone, na.rm = TRUE)
```

Exercise 5

Make a scatter plot with `Temp` on *x*-axis and `Ozone` on the *y*-axis. Label the *x*-axis “Temperature (degrees F)” and the *y*-axis “Ozone (ppb)”. Describe the association between the two variables.