In Class Assignment Week 2

Part 1:

1. The data set at rnf6080.dat records hourly rainfall at a certain location in Canada, every day from 1960 to 1980. First, we need to load the data set into R using the command read.table(). Use the help function to learn what arguments this function takes. Once you have the necessary input, load the data set into R and make it a data frame called rain.df.

```
#Load data file
rain.df = as.data.frame(read.table('rnf6080.dat'))
```

2. How many rows and columns does rain.df have? (If there are not 5070 rows and 27 columns, something is wrong; check the previous part to see what might have gone wrong in the previous part.)

```
#Finds number of rows and columns
nrows = nrow(rain.df)
ncols = ncol(rain.df)
nrows
```

```
## [1] 5070
```

```
ncols
```

```
## [1] 27
```

3. What are the names of the columns of rain.df?

```
#Finds names of columns
col_names = names(rain.df)
col_names
```

```
##  [1] "V1"  "V2"  "V3"  "V4"  "V5"  "V6"  "V7"  "V8"  "V9"  "V10" "V11"
## [12] "V12" "V13" "V14" "V15" "V16" "V17" "V18" "V19" "V20" "V21" "V22"
## [23] "V23" "V24" "V25" "V26" "V27"
```

4. What is the value of row 5, column 7 of rain.df?

```
#Finds value of row 5,7
value_row5_col7 = rain.df[5,7]
value_row5_col7
```

```
## [1] 0
```

5. Display the second row of rain.df in its entirety.

```
#Prints the second row
print_second_row = rain.df[2,]
print_second_row
```

```
##   V1 V2 V3 V4 V5 V6 V7 V8 V9 V10 V11 V12 V13 V14 V15 V16 V17 V18 V19 V20
## 2 60  4  2  0  0  0  0  0  0   0   0   0   0   0   0   0   0   0   0   0
##   V21 V22 V23 V24 V25 V26 V27
## 2   0   0   0   0   0   0   0
```

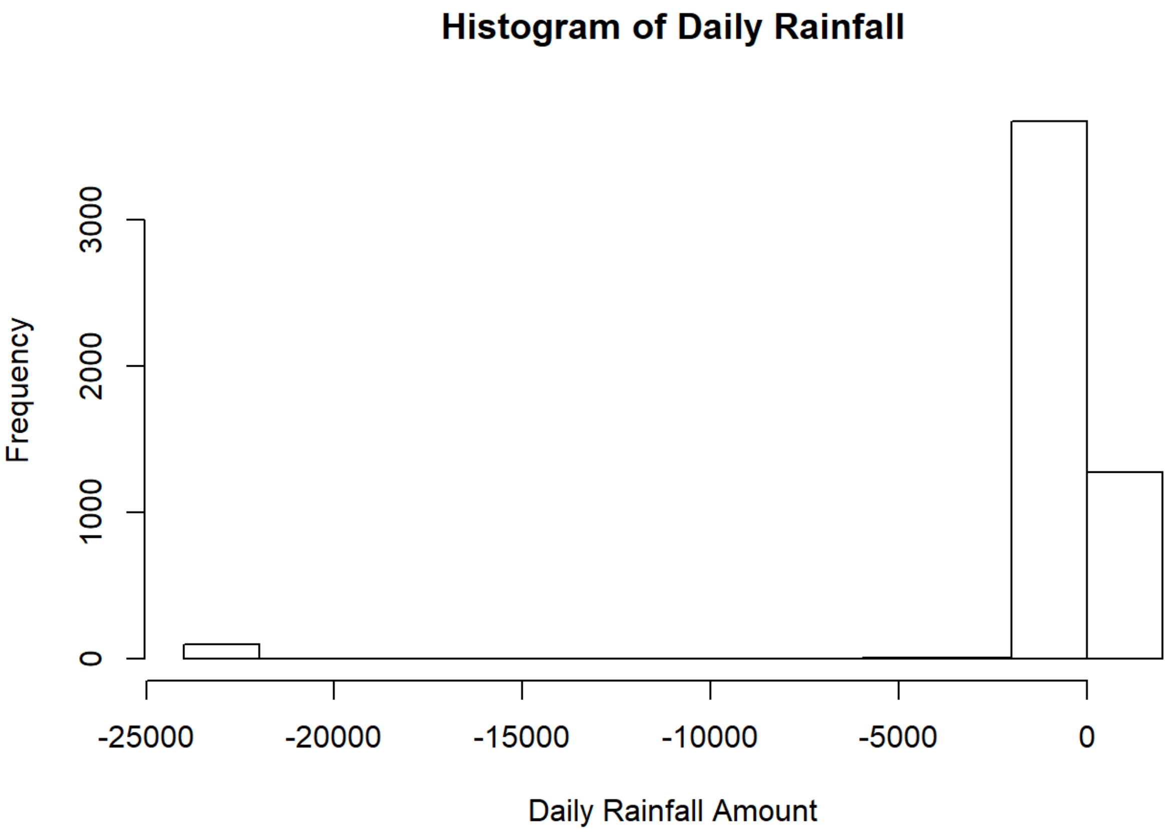
6. Explain what the command below doesby running it on your data and examining the object. (You may find the display functions head() and tail() useful here.) Is it clear now what the last 24 columns represent?

```
names(rain.df) <- c("year","month","day",seq(0,23))
```

This command changes the labels of the first three columns to “year”, “month”, and “day” respectively. The last 24 columns represent the hour of the day.

7. Create a new column in the data frame called daily, which is the sum of the rightmost 24 columns. With this column, create a histogram of the values in this column, which are supposed to be daily rainfall values. What is wrong with this picture?

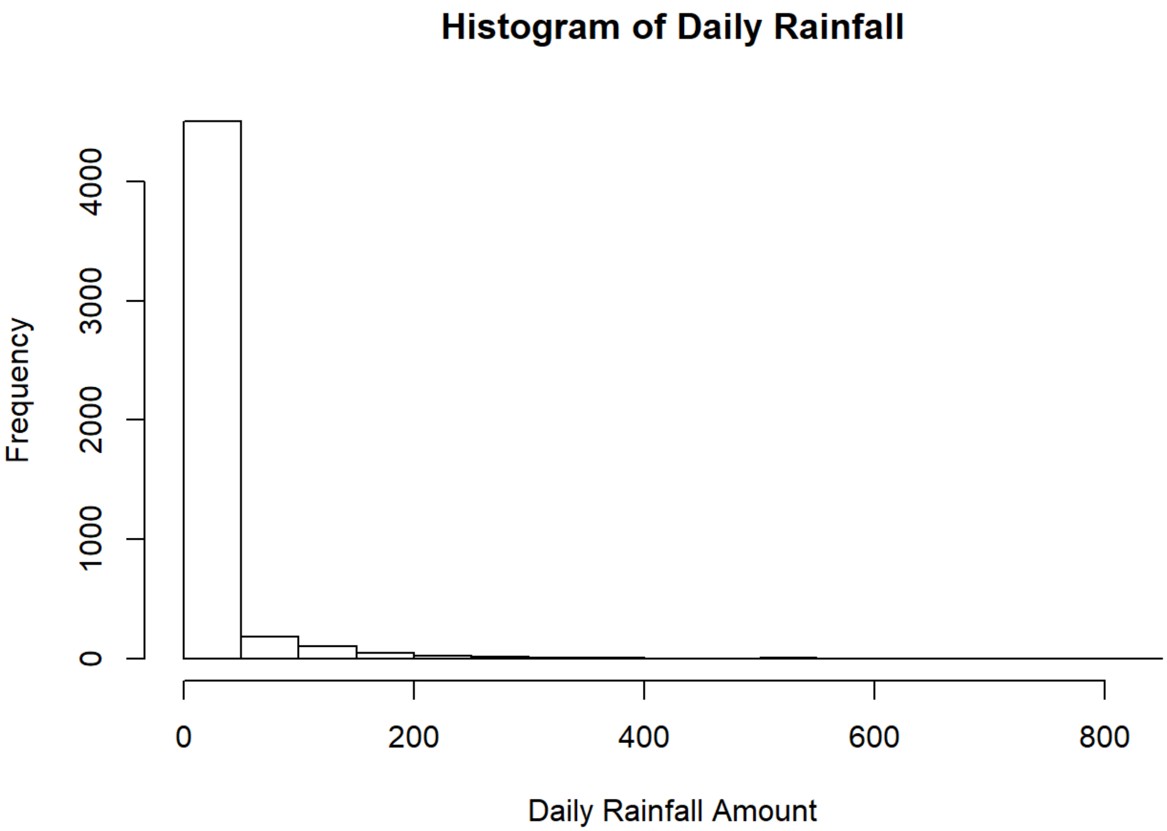
```
#Adds column
rain.df =
rain.df %>%
  mutate(daily = rowSums(rain.df[, 4:27]))
#histogram of daily rainfall
hist(rain.df$daily, main = "Histogram of Daily Rainfall", xlab = "Daily Rainfall Amoun
t", ylab = "Frequency")
```



This histogram has negative values. There can be no negative rainfall amounts.

8. Create a new data frame rain.df.fixed that takes the original and fixes it for the apparent flaw you have discovered. Having done this, produce a new histogram with the corrected data and explain why this is more reasonable.

```
#Changed all values <0 to NA
rain.df[rain.df < 0] = NA
hist(rain.df$daily, main = "Histogram of Daily Rainfall", xlab = "Daily Rainfall Amoun
t", ylab = "Frequency")
```



This graph is much more reasonable because there are no negative values.