R Week 1 Quiz

Lilian

7/30/2020

Solutions to problem set using hw1_data

First, I used RStudio to import the .csv data (I had to upload this from my local directory to kgen using scp). This required installation of the readr package, followed by read_csv()

```
library(readr)
hw1_data <- read.csv("/gfs/work/llam/R-lessons/datasciencecoursera/hw1_data.csv")
```

11) In the dataset provided for this Quiz, what are the column names of the dataset?

```
colnames(hw1_data)
```

```
## [1] "Ozone"   "Solar.R" "Wind"    "Temp"    "Month"   "Day"
```

12) Extract the first 2 rows of the data frame and print them to the console.

```
hw1_data[1:2,]
```

```
##   Ozone Solar.R Wind Temp Month Day
## 1    41     190  7.4   67     5   1
## 2    36     118  8.0   72     5   2
```

13) How many observations (i.e. rows) are in this data frame?

```
nrow(hw1_data)
```

```
## [1] 153
```

14) Extract the last 2 rows of the data frame and print them to the console. What does the output look like?

```
hw1_data[152:153, ]
```

```
##     Ozone Solar.R Wind Temp Month Day
## 152    18     131  8.0   76     9  29
## 153    20     223 11.5   68     9  30
```

```
tail(hw1_data, n=2)
```

```
##     Ozone Solar.R Wind Temp Month Day
## 152    18     131  8.0   76     9  29
## 153    20     223 11.5   68     9  30
```

15) What is the value of Ozone in the 47th row?

```
hw1_data$Ozone[47]
```

```
## [1] 21
```

16) How many missing values are in the Ozone column of this data frame?

```r
is.na(hw1_data$Ozone)
```

```
##   [1] FALSE FALSE FALSE FALSE  TRUE FALSE FALSE FALSE FALSE  TRUE FALSE FALSE
##  [13] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
##  [25]  TRUE  TRUE  TRUE FALSE FALSE FALSE FALSE  TRUE  TRUE  TRUE  TRUE  TRUE
##  [37]  TRUE FALSE  TRUE FALSE FALSE  TRUE  TRUE FALSE  TRUE  TRUE FALSE FALSE
##  [49] FALSE FALSE FALSE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE
##  [61]  TRUE FALSE FALSE FALSE  TRUE FALSE FALSE FALSE FALSE FALSE FALSE  TRUE
##  [73] FALSE FALSE  TRUE FALSE FALSE FALSE FALSE FALSE FALSE FALSE  TRUE  TRUE
##  [85] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
##  [97] FALSE FALSE FALSE FALSE FALSE  TRUE  TRUE FALSE FALSE FALSE  TRUE FALSE
## [109] FALSE FALSE FALSE FALSE FALSE FALSE  TRUE FALSE FALSE FALSE  TRUE FALSE
## [121] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## [133] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## [145] FALSE FALSE FALSE FALSE FALSE  TRUE FALSE FALSE FALSE
```

then sum these to get the count of NA = TRUE

```r
sum(is.na(hw1_data$Ozone))
```

```
## [1] 37
```

17) What is the mean of the Ozone column in this dataset? Exclude missing values (coded as NA) from this calculation.

We need to remove NA by na.rm = TRUE

```r
mean(hw1_data$Ozone, na.rm = TRUE)
```

```
## [1] 42.12931
```

18) Extract the subset of rows of the data frame where Ozone values are above 31 and Temp values are above 90. What is the mean of Solar.R in this subset?

```r
mean(hw1_data[which(hw1_data$Ozone > 31 & hw1_data$Temp > 90),]$Solar.R)
```

```
## [1] 212.8
```

19) What is the mean of "Temp" when "Month" is equal to 6?

```r
mean(hw1_data[which(hw1_data$Month == 6),]$Temp)
```

```
## [1] 79.1
```

20) What was the maximum ozone value in the month of May (i.e. Month is equal to 5)?

```r
max(hw1_data[which(hw1_data$Month == 5),]$Ozone, na.rm = TRUE)
```

```
## [1] 115
```