Course 5\_Assignment1

birasafab

1/12/2021

##### Reading the required dataset for tidying and analysis

library(readr)

## Warning: package 'readr' was built under R version 3.6.3

Activity<-read\_csv("activity.zip")

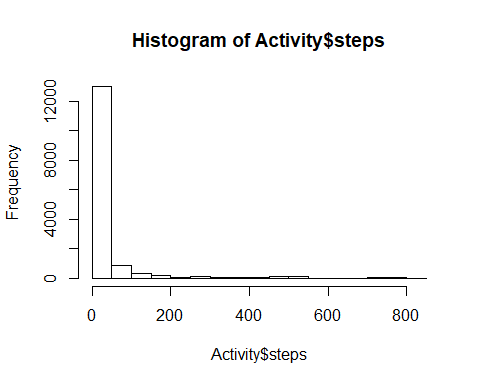
## Parsed with column specification:  
## cols(  
## steps = col\_double(),  
## date = col\_date(format = ""),  
## interval = col\_double()  
## )

head(Activity)

## # A tibble: 6 x 3  
## steps date interval  
## <dbl> <date> <dbl>  
## 1 NA 2012-10-01 0  
## 2 NA 2012-10-01 5  
## 3 NA 2012-10-01 10  
## 4 NA 2012-10-01 15  
## 5 NA 2012-10-01 20  
## 6 NA 2012-10-01 25

##### Ploting the Histogram of the steps taken per day

hist(Activity$steps)

 ##### Calculating the meaning and Median of the number of steps taken per day

library(dplyr)

## Warning: package 'dplyr' was built under R version 3.6.3

##   
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':  
##   
## filter, lag

## The following objects are masked from 'package:base':  
##   
## intersect, setdiff, setequal, union

mean\_day<-Activity %>%  
group\_by(date) %>%  
summarize(Mean = mean(steps, na.rm=TRUE))  
tail(mean\_day)

## # A tibble: 6 x 2  
## date Mean  
## <date> <dbl>  
## 1 2012-11-25 41.1  
## 2 2012-11-26 38.8  
## 3 2012-11-27 47.4  
## 4 2012-11-28 35.4  
## 5 2012-11-29 24.5  
## 6 2012-11-30 NaN

### Calculating the Median of the number of steps taken per day

median\_day<-Activity %>%  
group\_by(date) %>%  
summarize(median = median(steps, na.rm=TRUE))

# Calculating the Mean and Median of the number of Steps taken per day

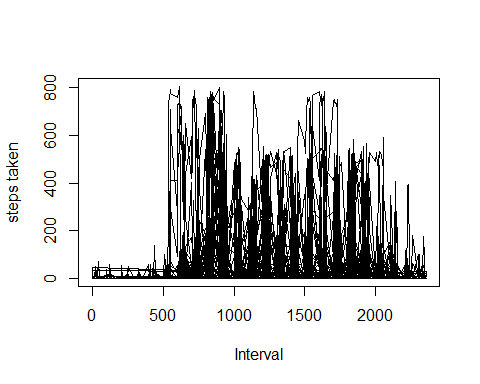
mean(Activity$steps,na.rm = TRUE)

## [1] 37.3826

median(Activity$steps,na.rm = TRUE)

## [1] 0

plot(Activity$interval, Activity$steps, type = "l", xlab = "Interval",ylab = "steps taken")



#Finding the date on which maximum number of steps were recorded

arrange(mean\_day,-Mean,date)[1,1]

## # A tibble: 1 x 1  
## date   
## <date>   
## 1 2012-11-23

### Calculating the total number of missing values

## the one way to calculate it is like the following   
sum(is.na(Activity$steps))

## [1] 2304

## or the following function can work too  
max(row(Activity))-sum(complete.cases(Activity))

## [1] 2304

## calculating the mean for missing values

## the following function serves to find the name of columns which have any missing value  
Col\_missing<-colnames(Activity)[apply(Activity,2,anyNA)]  
## the following function serves to calculates the mean value to replace all the missing values existent  
apply(Activity[,colnames(Activity) %in% Col\_missing], 2, mean, na.rm = TRUE)

## steps   
## 37.3826

## Imputing the missing values with value of the mean  
Activity\_imputes\_mean<-sapply(Activity,function(steps) ifelse(is.na(steps),mean(steps,na.rm = TRUE),steps))

Final\_impute\_mean<-sapply(Activity, function(steps) ifelse(is.na(steps),mean\_day,steps))  
  
## calculating the mean for that day to replace the missing values  
head(mean\_day)

## # A tibble: 6 x 2  
## date Mean  
## <date> <dbl>  
## 1 2012-10-01 NaN   
## 2 2012-10-02 0.438  
## 3 2012-10-03 39.4   
## 4 2012-10-04 42.1   
## 5 2012-10-05 46.2   
## 6 2012-10-06 53.5

what<-Activity %>%  
 group\_by(date) %>%  
 summarise(mean(steps,na.rm = TRUE))  
head(what)

## # A tibble: 6 x 2  
## date `mean(steps, na.rm = TRUE)`  
## <date> <dbl>  
## 1 2012-10-01 NaN   
## 2 2012-10-02 0.438  
## 3 2012-10-03 39.4   
## 4 2012-10-04 42.1   
## 5 2012-10-05 46.2   
## 6 2012-10-06 53.5

head(mean\_day)

## # A tibble: 6 x 2  
## date Mean  
## <date> <dbl>  
## 1 2012-10-01 NaN   
## 2 2012-10-02 0.438  
## 3 2012-10-03 39.4   
## 4 2012-10-04 42.1   
## 5 2012-10-05 46.2   
## 6 2012-10-06 53.5