Course assignment

R Markdown

An assignment to understand R markdown and reproducible research. Libraries that will be used.

```
library(dplyr)

## Warning: package 'dplyr' was built under R version 4.1.1

##
## 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

library(ggplot2)

## Warning: package 'ggplot2' was built under R version 4.1.1
```

Assignment

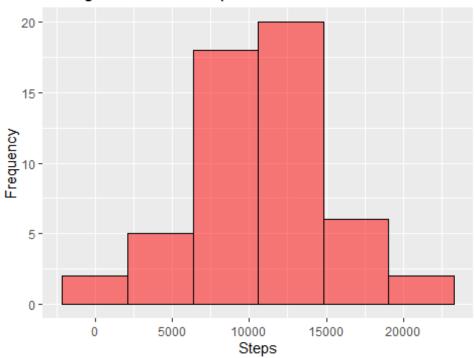The first step is to load the data set that we will be working with.

```
activity <- read.csv("data/activity.csv")
activity <- tibble::as_tibble(activity)
head(activity, n = 10)

## # A tibble: 10 x 3
##    steps date       interval
##    <int> <chr>         <int>
##  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
##  7    NA 2012-10-01       30
##  8    NA 2012-10-01       35
##  9    NA 2012-10-01       40
## 10    NA 2012-10-01       45
```

Histogram of the total steps in a day.

```
by_sum <- activity %>% group_by(date) %>% summarise(sum_Steps = sum(steps),
na.rm = TRUE)
by_sum <- tibble::as_tibble(by_sum)
```

```
ggplot(data = by_sum, aes(x = sum_Steps)) + geom_histogram (bins = 6, fill =
"red", col = "black", alpha = .5) + labs (x = "Steps", y = "Frequency", title
= "Histogram of Total steps")
```

```
## Warning: Removed 8 rows containing non-finite values (stat_bin).
```


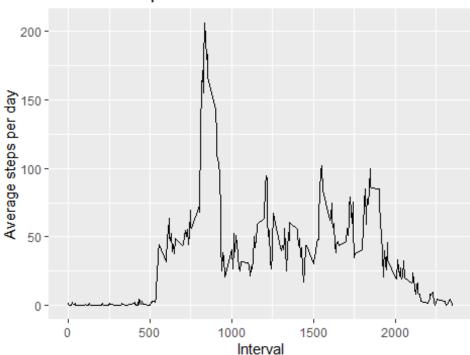Histogram of Total steps

Mean and Median for the steps taken each day.

```
summary(by_sum)
```

```
##      date              sum_Steps        na.rm
##   Length:61          Min.   :    41   Mode:logical
##   Class :character   1st Qu.: 8841    TRUE:61
##   Mode  :character   Median :10765
##                      Mean   :10766
##                      3rd Qu.:13294
##                      Max.   :21194
##                      NA's   :8
```

The median is presented as 10,765 as for the mean is 10,766.

Time series plot of the 5-minute interval (x-axis) and the average steps per day (y-axis).

```
avg_interval <- activity %>% group_by(interval) %>% summarize(avg_steps =
mean(steps, na.rm = TRUE))
qplot(interval, avg_steps, data = avg_interval, geom=c("line"), xlab =
"Interval", ylab = "Average steps per day", main = "Time Series plot")
```


Finding the number of missing values in the data set.

```
sum(is.na(activity))
```

```
## [1] 2304
```

Removing the missing values from the data using the mean average steps per day.
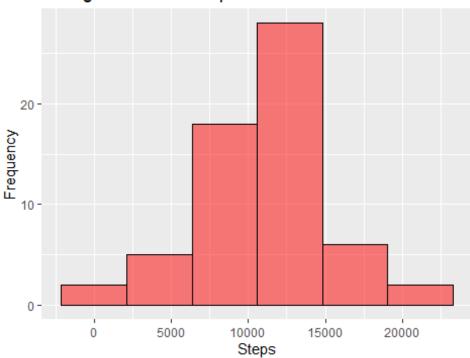
```
for (i in 1:nrow(activity)) {

    if(is.na(activity$steps[i])) {

        val <- avg_interval$avg_steps[which(avg_interval$interval ==
activity$interval[i])]
        activity$steps[i] <- val
    }
}
```

```
sum(is.na(activity))
```

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

Histogram with the missing value replaced.

```r
by_sum <- activity %>% group_by(date) %>% summarise(sum_Steps = sum(steps))
by_sum <- tibble::as_tibble(by_sum)
```

```r
ggplot(data = by_sum, aes(x = sum_Steps)) + geom_histogram (bins = 6, fill =
"red", col = "black", alpha = .5) + labs (x = "Steps", y = "Frequency", title
= "Histogram of Total steps")
```


Lets see if the mean and median have changed.

```r
summary(by_sum)
```

```
##      date               sum_Steps
##   Length:61           Min.   :   41
##   Class :character    1st Qu.: 9819
##   Mode  :character    Median :10766
##                       Mean   :10766
##                       3rd Qu.:12811
##                       Max.   :21194
```

Lastly, we will try to see if there is a pattern between steps in the weekday and steps in the weekend.

```r
activity['type_of_day'] <- weekdays(as.Date(activity$date))

activity$type_of_day[activity$type_of_day  %in% c('Saturday','Sunday') ] <-
"weekend"
```

```
activity$type_of_day[activity$type_of_day != "weekend"] <- "weekday"

activity$type_of_day <- as.factor(activity$type_of_day)
interval_sum <- aggregate(steps ~ interval + type_of_day, activity, mean)

qplot(interval, steps, data = interval_sum, geom=c("line"), xlab =
"Interval", ylab = "Number of steps", main = "") + facet_wrap(~ type_of_day,
ncol = 1)
```