Programming Assignment 1 - Reproducible Research

Sangeet Khullar

May 17, 2019

Read CSV Input

```
#setwd("C:\\Users\\skhullar\\OneDrive - Daiichi Sankyo\\Documents\\GitHub\\Reproducible_Research")
act <- read.csv("activity.csv")
dim(act)
```

```
## [1] 17568     3
```

Find Mean & Median Steps Per Day

1. Sum up all the steps for a given Day - (Sum(steps) - group by Date)

2. Display Histogram of Steps Per Day

```
library(dplyr)
```

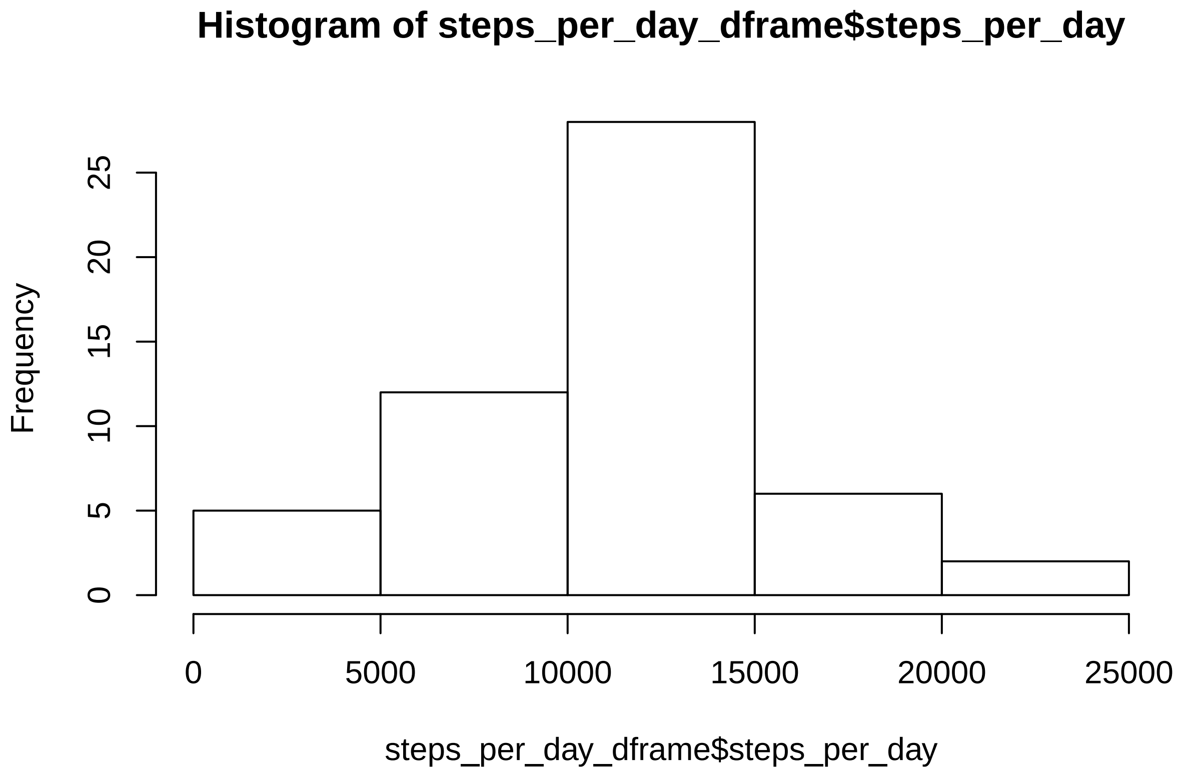
```
## Warning: package 'dplyr' was built under R version 3.5.3
```

```
##
## Attaching package: 'dplyr'
```

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

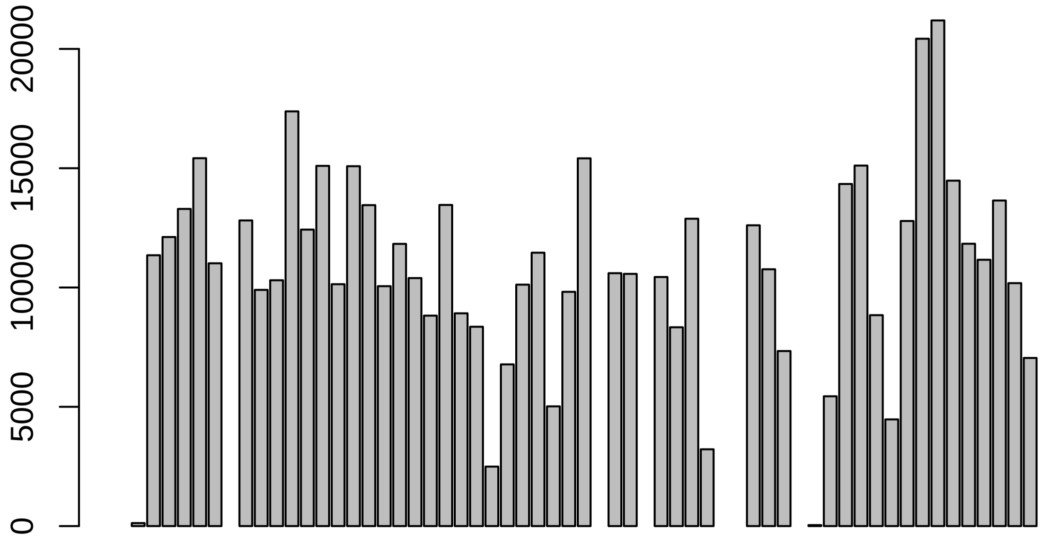
```
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
steps_per_day_dframe <- act    %>% group_by(date) %>% summarize(steps_per_day=sum(steps))
hist(steps_per_day_dframe$steps_per_day)
```



3. Display BarPlot of Steps Per Day

```
barplot(steps_per_day_dframe$steps_per_day)
```



4. Find Mean & Median value per day from the above aggregated dataset

```
mean_steps_per_day <- steps_per_day_dframe %>% summarize(mean_steps_per_day=mean(steps_per_day,na.rm = 
print(paste("Mean_Steps_Per_Day=",mean_steps_per_day))
```

```
## [1] "Mean_Steps_Per_Day= 10766.1886792453"
```

```
median_steps_per_day <- steps_per_day_dframe %>% summarize(median_steps_per_day=median(steps_per_day,na
print(paste("Median_Steps_Per_Day=",median_steps_per_day))
```

```
## [1] "Median_Steps_Per_Day= 10765"
```

Average daily activity pattern

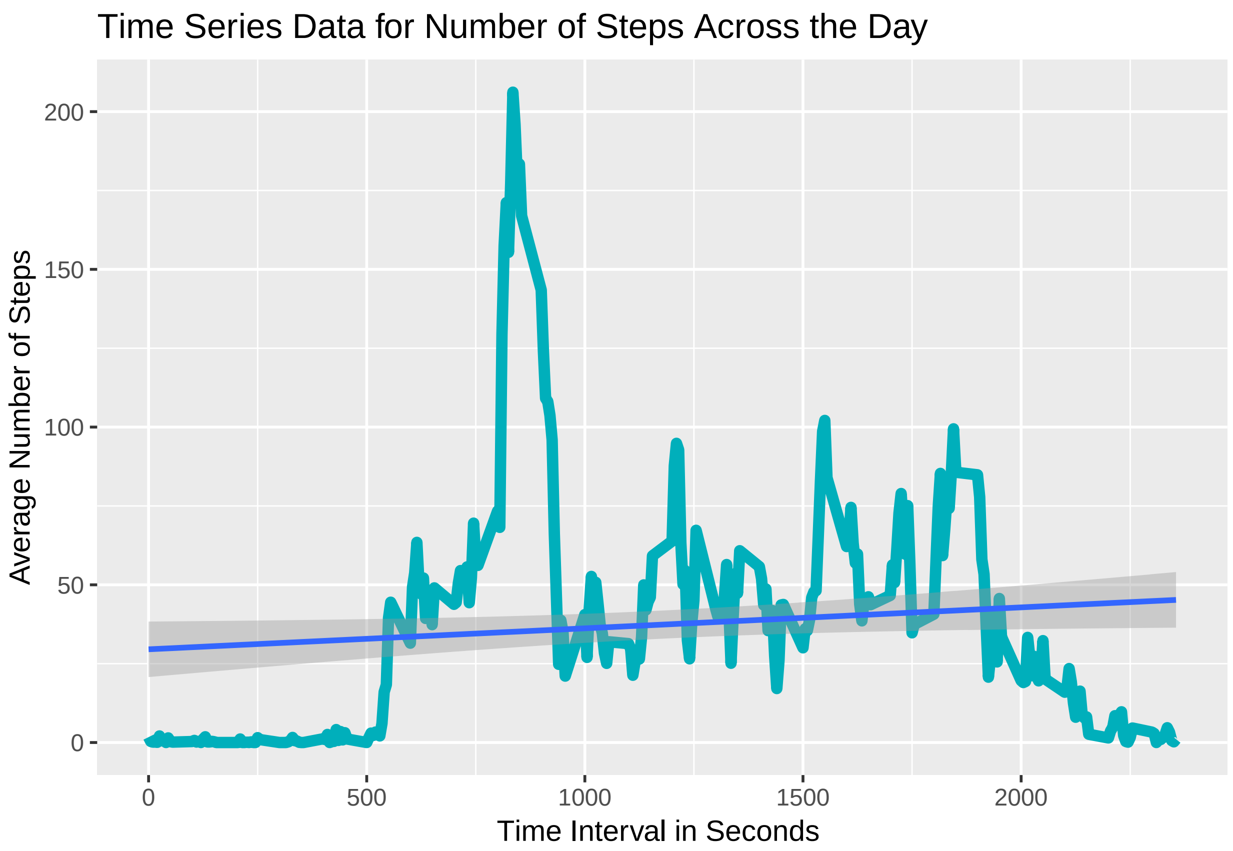
1. Time series plot of the 5-minute interval & the average number of steps taken, averaged across all days

```
library(ggplot2)
```

```
## Warning: package 'ggplot2' was built under R version 3.5.3
```

```
avg_steps_per_5min_interval <- act    %>% group_by(interval) %>% summarize(avg_steps_for_interval= mean

ggplot(data = avg_steps_per_5min_interval, aes(x = interval, y = avg_steps_for_interval))+
  geom_line(color = "#00AFBB", size = 2) +
   geom_smooth(method = "lm") +
     labs(x = "Time Interval in Seconds ") +
          labs(y = "Average Number of Steps") +
            labs(title = "Time Series Data for Number of Steps Across the Day")
```



2. Which 5-minute interval, on average across all the days in the dataset, contains the maximum number of steps?

```
max_step_average <- max(avg_steps_per_5min_interval$avg_steps_for_interval, na.rm = TRUE)

max_step_average_interval <- avg_steps_per_5min_interval[which(avg_steps_per_5min_interval$avg_steps_fo
                                    max_step_average), ]
print(paste("Max_Step_Average_Interval=",max_step_average_interval$interval))
```

```
## [1] "Max_Step_Average_Interval= 835"
```

Imputing missing values

Report the total number of missing values in the dataset

```
summary(act)
```

```
##      steps                  date          interval     
##  Min.   :  0.00   2012-10-01:  288   Min.   :   0.0  
##  1st Qu.:  0.00   2012-10-02:  288   1st Qu.: 588.8  
##  Median :  0.00   2012-10-03:  288   Median :1177.5  
##  Mean   : 37.38   2012-10-04:  288   Mean   :1177.5  
##  3rd Qu.: 12.00   2012-10-05:  288   3rd Qu.:1766.2  
##  Max.   :806.00   2012-10-06:  288   Max.   :2355.0  
##  NA's   :2304     (Other)   :15840
```

```
count_na_values <- sum(is.na(act$steps))
print(paste("Count_Of_Records_With_Missing_Step_Values=",count_na_values))
```

```
## [1] "Count_Of_Records_With_Missing_Step_Values= 2304"
```

Fill up the missing values in the dataset

```
library(mice)
```

```
## Warning: package 'mice' was built under R version 3.5.3

## Loading required package: lattice

##
## Attaching package: 'mice'

## The following objects are masked from 'package:base':
##
##     cbind, rbind
```

```
tempData <- mice(act,m=5,maxit=50,meth='pmm',seed=500)
```

```
##
##  iter imp variable
##   1   1  steps
##   1   2  steps
##   1   3  steps
##   1   4  steps
##   1   5  steps
##   2   1  steps
##   2   2  steps
##   2   3  steps
##   2   4  steps
##   2   5  steps
##   3   1  steps
##   3   2  steps
##   3   3  steps
##   3   4  steps
##   3   5  steps
##   4   1  steps
##   4   2  steps
##   4   3  steps
##   4   4  steps
##   4   5  steps
##   5   1  steps
##   5   2  steps
##   5   3  steps
##   5   4  steps
##   5   5  steps
##   6   1  steps
##   6   2  steps
##   6   3  steps
##   6   4  steps
##   6   5  steps
##   7   1  steps
##   7   2  steps
##   7   3  steps
##   7   4  steps
##   7   5  steps
##   8   1  steps
##   8   2  steps
##   8   3  steps
##   8   4  steps
##   8   5  steps
##   9   1  steps
##   9   2  steps
##   9   3  steps
##   9   4  steps
##   9   5  steps
##   10   1  steps
##   10   2  steps
##   10   3  steps
##   10   4  steps
##   10   5  steps
##   11   1  steps
##   11   2  steps
##   11   3  steps
##   11   4  steps
##   11   5  steps
##   12   1  steps
##   12   2  steps
##   12   3  steps
##   12   4  steps
##   12   5  steps
##   13   1  steps
##   13   2  steps
##   13   3  steps
##   13   4  steps
##   13   5  steps
##   14   1  steps
##   14   2  steps
##   14   3  steps
##   14   4  steps
##   14   5  steps
##   15   1  steps
##   15   2  steps
##   15   3  steps
##   15   4  steps
##   15   5  steps
##   16   1  steps
##   16   2  steps
##   16   3  steps
##   16   4  steps
##   16   5  steps
##   17   1  steps
##   17   2  steps
##   17   3  steps
##   17   4  steps
##   17   5  steps
##   18   1  steps
##   18   2  steps
##   18   3  steps
##   18   4  steps
##   18   5  steps
##   19   1  steps
##   19   2  steps
##   19   3  steps
##   19   4  steps
##   19   5  steps
##   20   1  steps
##   20   2  steps
##   20   3  steps
##   20   4  steps
##   20   5  steps
##   21   1  steps
##   21   2  steps
##   21   3  steps
##   21   4  steps
##   21   5  steps
##   22   1  steps
##   22   2  steps
##   22   3  steps
##   22   4  steps
##   22   5  steps
##   23   1  steps
##   23   2  steps
##   23   3  steps
##   23   4  steps
##   23   5  steps
##   24   1  steps
##   24   2  steps
##   24   3  steps
##   24   4  steps
##   24   5  steps
##   25   1  steps
##   25   2  steps
##   25   3  steps
##   25   4  steps
##   25   5  steps
##   26   1  steps
##   26   2  steps
##   26   3  steps
##   26   4  steps
##   26   5  steps
##   27   1  steps
##   27   2  steps
##   27   3  steps
##   27   4  steps
##   27   5  steps
##   28   1  steps
##   28   2  steps
##   28   3  steps
##   28   4  steps
##   28   5  steps
##   29   1  steps
##   29   2  steps
##   29   3  steps
##   29   4  steps
##   29   5  steps
##   30   1  steps
##   30   2  steps
##   30   3  steps
##   30   4  steps
##   30   5  steps
##   31   1  steps
##   31   2  steps
##   31   3  steps
##   31   4  steps
##   31   5  steps
##   32   1  steps
##   32   2  steps
##   32   3  steps
##   32   4  steps
##   32   5  steps
##   33   1  steps
##   33   2  steps
##   33   3  steps
##   33   4  steps
##   33   5  steps
##   34   1  steps
##   34   2  steps
##   34   3  steps
##   34   4  steps
##   34   5  steps
##   35   1  steps
##   35   2  steps
##   35   3  steps
##   35   4  steps
##   35   5  steps
##   36   1  steps
##   36   2  steps
##   36   3  steps
##   36   4  steps
##   36   5  steps
##   37   1  steps
##   37   2  steps
##   37   3  steps
##   37   4  steps
##   37   5  steps
##   38   1  steps
##   38   2  steps
##   38   3  steps
##   38   4  steps
##   38   5  steps
##   39   1  steps
##   39   2  steps
##   39   3  steps
##   39   4  steps
##   39   5  steps
##   40   1  steps
##   40   2  steps
##   40   3  steps
##   40   4  steps
##   40   5  steps
##   41   1  steps
##   41   2  steps
##   41   3  steps
##   41   4  steps
##   41   5  steps
##   42   1  steps
##   42   2  steps
##   42   3  steps
##   42   4  steps
##   42   5  steps
##   43   1  steps
##   43   2  steps
##   43   3  steps
##   43   4  steps
##   43   5  steps
##   44   1  steps
##   44   2  steps
##   44   3  steps
##   44   4  steps
##   44   5  steps
##   45   1  steps
##   45   2  steps
##   45   3  steps
##   45   4  steps
##   45   5  steps
##   46   1  steps
##   46   2  steps
##   46   3  steps
##   46   4  steps
##   46   5  steps
##   47   1  steps
##   47   2  steps
##   47   3  steps
##   47   4  steps
##   47   5  steps
##   48   1  steps
##   48   2  steps
##   48   3  steps
##   48   4  steps
##   48   5  steps
##   49   1  steps
##   49   2  steps
##   49   3  steps
##   49   4  steps
##   49   5  steps
##   50   1  steps
##   50   2  steps
##   50   3  steps
##   50   4  steps
##   50   5  steps
```

```
## Warning: Number of logged events: 250
```

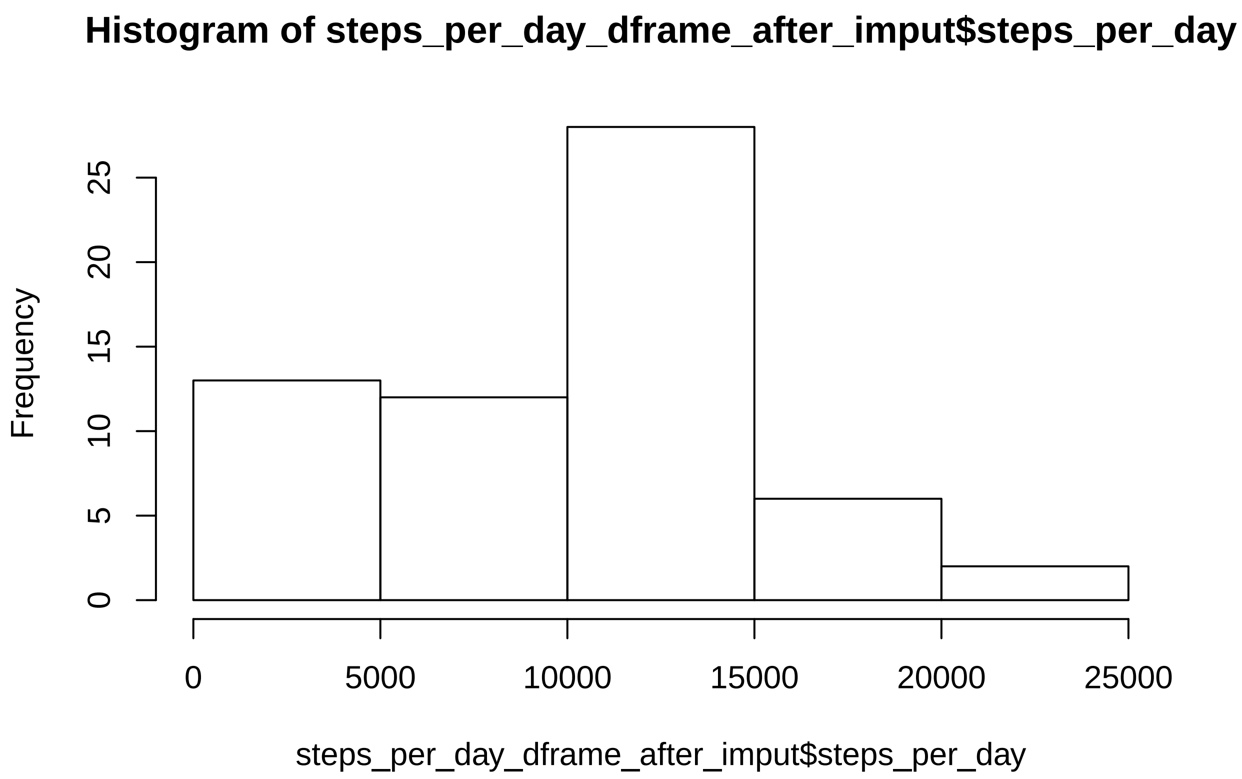
```
completedData <- complete(tempData,1)
```

Create a new Data Set with missing values populated

```
completedData <- complete(tempData,1)
```

Make a histogram of the total number of steps taken each day

```
steps_per_day_dframe_after_imput <- completedData    %>% group_by(date) %>% summarize(steps_per_day=sum
hist(steps_per_day_dframe_after_imput$steps_per_day)
```



Calculate the mean and median total number of steps taken per day AFTER Imput

```
mean_steps_per_day_after_imput  <- steps_per_day_dframe_after_imput %>% summarize(mean_steps_per_day=mea
print(paste("Mean_Steps_Per_Day_After_Imput=",mean_steps_per_day_after_imput))
```

```
## [1] "Mean_Steps_Per_Day_After_Imput= 9354.22950819672"
```

```
median_steps_per_day_after_imput <- steps_per_day_dframe_after_imput %>% summarize(median_steps_per_day
print(paste("Median_Steps_Per_Day_After_Imput=",median_steps_per_day_after_imput))
```

```
## [1] "Median_Steps_Per_Day_After_Imput= 10395"
```

Mean & Median Before Imput

```
print(paste("Mean_Steps_Per_Day_Before_Imput=",mean_steps_per_day))
```

```
## [1] "Mean_Steps_Per_Day_Before_Imput= 10766.1886792453"
```

```
print(paste("Median_Steps_Per_Day_Before_Imput=",median_steps_per_day))
```

```
## [1] "Median_Steps_Per_Day_Before_Imput= 10765"
```

Impact of imputing missing data on the estimates of the total daily number of steps

```
diff_mean <-  mean_steps_per_day_after_imput -mean_steps_per_day
print(paste("Change_in_Mean_Steps_Per_day=",diff_mean))
```

```
## [1] "Change_in_Mean_Steps_Per_day= -1411.95917104856"
```

```
diff_median <- median_steps_per_day_after_imput - median_steps_per_day
print(paste("Change_in_Median_Steps_Per_day=",diff_median))
```

```
## [1] "Change_in_Median_Steps_Per_day= -370"
```

Differences in activity patterns between weekdays and weekends

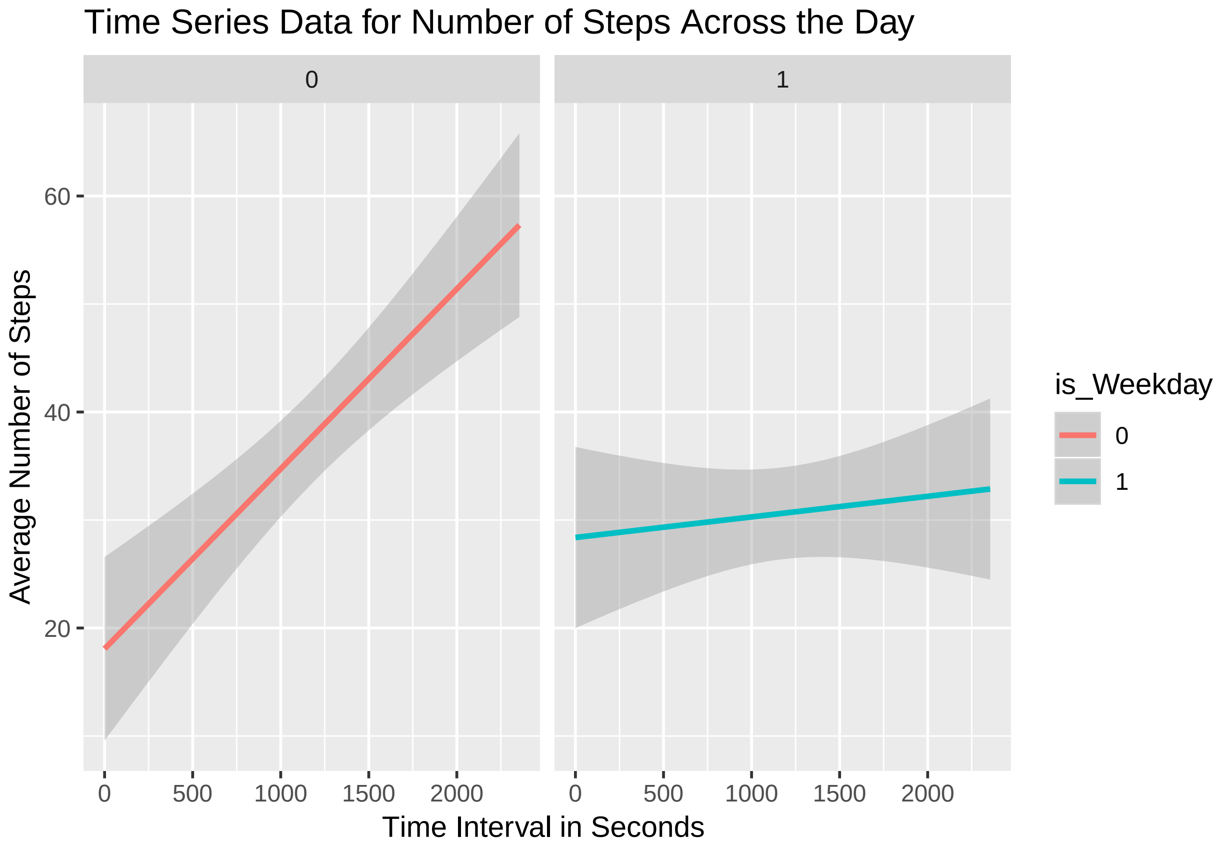
New factor variable weekday and weekend

```
for (row in 1:nrow(completedData)) {
    day  <- weekdays(as.Date(completedData[row, "date"]))
       if(day %in% c("Monday","Tuesday","Wednesday","Thursday","Friday"))
         {
        completedData[row,"is_Weekday"] <- "1"
         }
        else
          {
        completedData[row,"is_Weekday"] <- "0"
          }
}
```

Panel Time Series Plot time series plot

```
avg_steps_per_5min_interval_imp <- completedData  %>% group_by(interval,is_Weekday) %>%
      summarize(avg_steps_for_interval= mean(steps,na.rm = TRUE))

ggplot(data = avg_steps_per_5min_interval_imp, aes(x = interval, y = avg_steps_for_interval ,color=is_We
   facet_grid( .~ is_Weekday) +
    geom_smooth(method = "lm") +
     labs(x = "Time Interval in Seconds ") +
          labs(y = "Average Number of Steps") +
            labs(title = "Time Series Data for Number of Steps Across the Day")
```


Time Series Data for Number of Steps Across the Day
0
1
Average Number of Steps
60
40
20
0
500
1000
1500
2000
Time Interval in Seconds
is_Weekday
0
1

The End