Reproducible Research Assignment 1

Scott Hansen

Sunday February 14, 2016

Load necessary libararies

library(dplyr)

## Warning: package 'dplyr' was built under R version 3.1.3

##   
## Attaching package: 'dplyr'  
##   
## The following object is masked from 'package:stats':  
##   
## filter  
##   
## The following objects are masked from 'package:base':  
##   
## intersect, setdiff, setequal, union

library(stringr)  
library(ggplot2)

## Warning: package 'ggplot2' was built under R version 3.1.3

library(Hmisc)

## Loading required package: grid  
## Loading required package: lattice  
## Loading required package: survival  
## Loading required package: splines  
## Loading required package: Formula  
##   
## Attaching package: 'Hmisc'  
##   
## The following objects are masked from 'package:dplyr':  
##   
## combine, src, summarize  
##   
## The following objects are masked from 'package:base':  
##   
## format.pval, round.POSIXt, trunc.POSIXt, units

library(timeDate)

## Warning: package 'timeDate' was built under R version 3.1.3

1. Code for Reading in the data set and/or processing the data

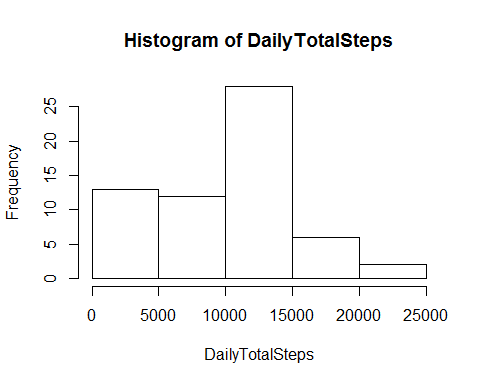
setwd("E:\\Data Scientist\\Data Science Track\\Reproducible Reseearch\\Project 1\\Reproducible-Research-Project-1")  
IntervalData<- data.frame(read.csv("activity.csv", stringsAsFactors = FALSE, header = TRUE))

* Preprocess Data Convert interval to time

IntervalData$DateTime<-paste(IntervalData$date, str\_pad(IntervalData$interval, 4,side="left", pad="0"))

1. Histogram of the total number of steps taken each day

DailyTotalSteps<-tapply(IntervalData$steps, IntervalData$date, sum, na.rm=TRUE)  
  
hist(DailyTotalSteps, breaks=7)



1. Mean and Median of steps taken each day

MeanSteps<-mean(DailyTotalSteps)  
MedianSteps<-median(DailyTotalSteps)

-Mean

MeanSteps

## [1] 9354.23

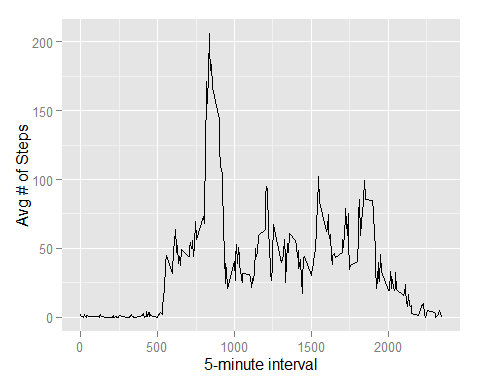
-Median

MedianSteps

## [1] 10395

1. Time series plot of the average steps taken

AvgStepsPerInterfal<- IntervalData[!is.na(IntervalData$steps),]%>%  
group\_by(interval)%>%   
summarise(Avg=mean(steps))  
  
qplot(interval, Avg, data=AvgStepsPerInterfal, geom="line", xlab="5-minute interval", ylab="Avg # of Steps")



1. 5 minute interval that on average contains the maximum number of steps

* The maximum number of steps occurs at interval

mostSteps <- which.max(AvgStepsPerInterfal$Avg)  
  
MaxInterval <-AvgStepsPerInterfal[mostSteps,]  
  
MaxInterval$interval

## [1] 835

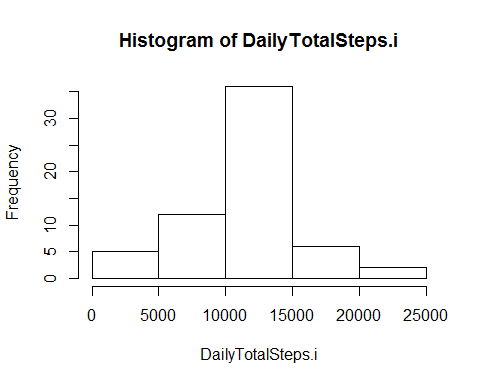
1. Code to determine how to impute missing data

* I selected the mean of all steps to replace missing values

IntervalDataClean<-IntervalData  
IntervalDataClean$steps<-impute(IntervalDataClean$steps, fun=mean)

1. Histogrram of the total number of steps taken after missing values are imputed

DailyTotalSteps.i<-tapply(IntervalDataClean$steps, IntervalData$date, sum)  
hist(DailyTotalSteps.i, breaks=7)



MeanSteps.i<-mean(DailyTotalSteps.i)  
MedianSteps.i<-median(DailyTotalSteps.i)

1. Panel plot comparing the average number of steps taken per 5-minute Interval across weekdays and week

IntervalDataClean$DayType<- ifelse(isWeekday(IntervalDataClean$date, wday = 1:5), 'WeekDay', 'WeekEnd')  
  
IntervalDataClean.S<-aggregate(steps~interval+DayType, data=IntervalDataClean, mean)  
  
  
qplot(interval, steps, data=IntervalDataClean.S, geom="line", facets=.~DayType, color=DayType, xlab="5-minute interval", ylab="Avg # of Steps")

