Reproducible Research: Peer Assessment 1

Loading and preprocessing the data.

## Set the directory in which the activity.zip file exists as the working ## directory in the RStudio.  
## Data can now be directly read and unzipped.  
library(knitr)

## Warning: package 'knitr' was built under R version 3.1.2

library(markdown)  
  
data <- read.csv(unzip("activity.zip","activity.csv"))  
## do some exploratory analysis to understand the structure and content of data.  
nrow(data)

## [1] 17568

summary(data)

## 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

summary of data indicate there are 2304 NA's in the data set. There are three variables and 17568 records.

In the next step we compute the daily sum of total steps without and with removal of NA values and and plot the respective histograms side-by-side.

## Compute the daily sum of total steps without and with the removal of NAs. Plot the results side-by-side for comparison.  
par(mfrow=c(1,2))  
daily\_sum <- tapply(data$steps, data$date, sum)  
hist(daily\_sum, breaks=20, main = "with NAs not removed")  
daily\_sum <- tapply(data$steps, data$date, sum, na.rm=T)  
hist(daily\_sum, breaks =20,main = "with NAs removed")

The Change in the frequency of "0" counts as sum points to the difference in the computations. Further to dermine the mean, we need to have the NAs removed

What is mean total number of steps taken per day?

mean\_total = mean(daily\_sum)  
round(mean\_total)

## [1] 9354

The mean number of stpes per day is 9354.

What is the average daily activity pattern?

## mean per interval is computed with the removal of NAs.  
daily.int.mean = NULL  
daily.int.mean <- tapply(data$steps, data$interval, mean, na.rm = TRUE)  
plot(daily.int.mean ~ data$interval[1:length(daily.int.mean)], type = "l", xlab = "Interval", ylab = "daily actvity (means steps/ interval)")

Determine the 5 minute interval containing the maximum number of steps

daily.int.mean[match(max(daily.int.mean),daily.int.mean)]

## 835   
## 206.1698

The interval is 835 and the maximum number of steps is 206.

Imputing missing values

## set the daily mean values for corresponding interval is used to replace ## the NA's  
data[is.na(data)] <- tapply(data$steps, data$interval, mean, na.rm = TRUE)  
## test with summary call for removal of NAs.  
summary(data)

## 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.: 27.00 2012-10-05: 288 3rd Qu.:1766.2   
## Max. :806.00 2012-10-06: 288 Max. :2355.0   
## (Other) :15840

## Determine the daily sum with the new set of data and plot.  
daily\_sum <- tapply(data$steps, data$date, sum)  
hist(daily\_sum, breaks=20, main = "NAs replaced with interval means")

Are there differences in activity patterns between weekdays and weekends? Yes. See below.

data["day"] <- NA  
data$day <- weekdays(as.Date(data$date))  
summary(data)

## steps date interval day   
## Min. : 0.00 2012-10-01: 288 Min. : 0.0 Length:17568   
## 1st Qu.: 0.00 2012-10-02: 288 1st Qu.: 588.8 Class :character   
## Median : 0.00 2012-10-03: 288 Median :1177.5 Mode :character   
## Mean : 37.38 2012-10-04: 288 Mean :1177.5   
## 3rd Qu.: 27.00 2012-10-05: 288 3rd Qu.:1766.2   
## Max. :806.00 2012-10-06: 288 Max. :2355.0   
## (Other) :15840

head(data)

## steps date interval day  
## 1 1.7169811 2012-10-01 0 Monday  
## 2 0.3396226 2012-10-01 5 Monday  
## 3 0.1320755 2012-10-01 10 Monday  
## 4 0.1509434 2012-10-01 15 Monday  
## 5 0.0754717 2012-10-01 20 Monday  
## 6 2.0943396 2012-10-01 25 Monday

tail(data)

## steps date interval day  
## 17563 2.6037736 2012-11-30 2330 Friday  
## 17564 4.6981132 2012-11-30 2335 Friday  
## 17565 3.3018868 2012-11-30 2340 Friday  
## 17566 0.6415094 2012-11-30 2345 Friday  
## 17567 0.2264151 2012-11-30 2350 Friday  
## 17568 1.0754717 2012-11-30 2355 Friday

data.weekday <- subset(data, day == "Monday"|day =="Tuesday"| day=="Wednesday" |day =="Thursday" |day =="Friday")  
data.weekend <- subset(data, day == "Sunday"|day =="Saturday")  
data.wd = NULL  
data.wd <- tapply(data.weekday$steps, data.weekday$interval, mean, na.rm = TRUE)  
data.we = NULL  
data.we <- tapply(data.weekend$steps, data.weekend$interval, mean, na.rm = TRUE)  
par(mfrow = c(1,2))  
plot(data.wd ~ data$interval[1:length(daily.int.mean)], main = "Weekdays", type = "l", xlab = "Interval", ylab = "daily actvity (means steps/ interval)")  
plot(data.we ~ data$interval[1:length(daily.int.mean)], main = "Weekends", type = "l", xlab = "Interval", ylab = "daily actvity (means steps/ interval)")

