

Reproducible Research: Peer Assessment 1

Loading and preprocessing the data

```
## unzip file and read dataset into dataframe
unzip("activity.zip")
datActivity <- read.csv("activity.csv")
## convert the date column from character to date type
datActivity$date <- as.Date(datActivity$date, "%Y-%m-%d")
## display summary of dataset
str(datActivity)
```

```
## 'data.frame': 17568 obs. of 3 variables:
## $ steps : int NA NA NA NA NA NA NA NA NA ...
## $ date : Date, format: "2012-10-01" "2012-10-01" ...
## $ interval: int 0 5 10 15 20 25 30 35 40 45 ...
```

What is mean total number of steps taken per day?

```
## load libraries utilized in the analysis
library(dplyr)
```

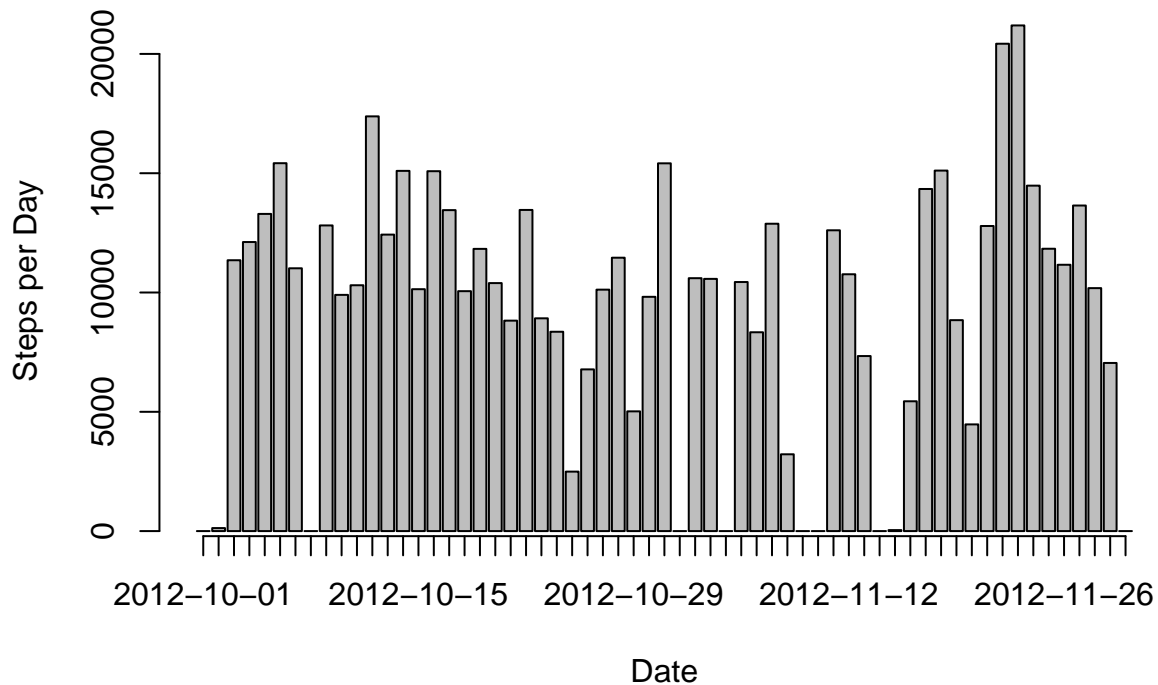
```
##
## Attaching package: 'dplyr'

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

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

```
library(ggplot2)
```

```
## summarize the dataset grouped by date and calculate the total steps per day
stepsbyday <- datActivity %>% group_by(date) %>% summarize(total=sum(steps, na.rm=TRUE))
## create a histogram of total steps per day
barplot(total~date, data=stepsbyday, axis.lty=1, xlab="Date", ylab="Steps per Day")
```



```
## show summary statistics of overall mean and median of the dataset
paste("Mean total number of steps per day:", round(mean(stepsbyday$total)))
```

```
## [1] "Mean total number of steps per day: 9354"
```

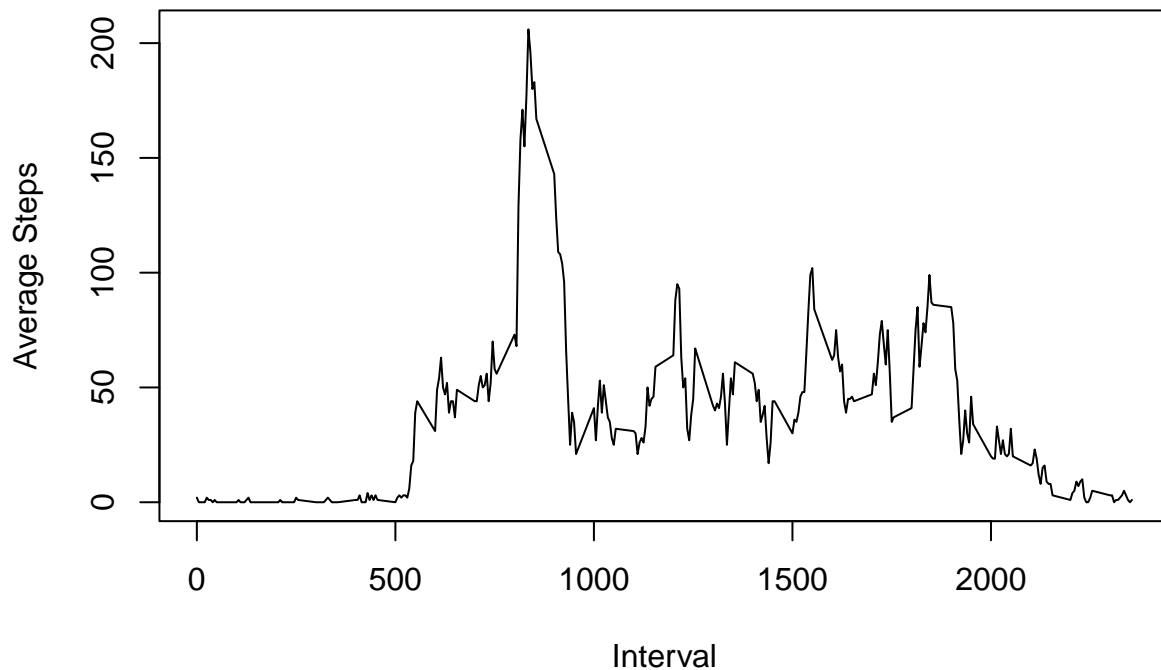
```
paste("Median total number of steps per day:", median(stepsbyday$total))
```

```
## [1] "Median total number of steps per day: 10395"
```

What is the average daily activity pattern?

Step activity predominates in the morning time frame.

```
## summarize the dataset grouped by time interval and calculate the average steps per interval
stepsbytime <- datActivity %>% group_by(interval) %>% summarize(avg=mean(steps, na.rm=TRUE))
stepsbytime <- transform(stepsbytime, avg=round(avg))
## create time-series plot of the average steps per time interval
plot(stepsbytime$interval, stepsbytime$avg, type="l", xlab="Interval", ylab="Average Steps")
```



```
## determine the time interval with the maximum average steps and display the interval number and steps
maxVal <- stepsbytime[which.max(stepsbytime$avg),]
paste("The interval with the maximum average number of steps:", maxVal[1])
```

```
## [1] "The interval with the maximum average number of steps: 835"
```

```
paste("The maximum average number of steps in an interval:", maxVal[2])
```

```
## [1] "The maximum average number of steps in an interval: 206"
```

Imputing missing values

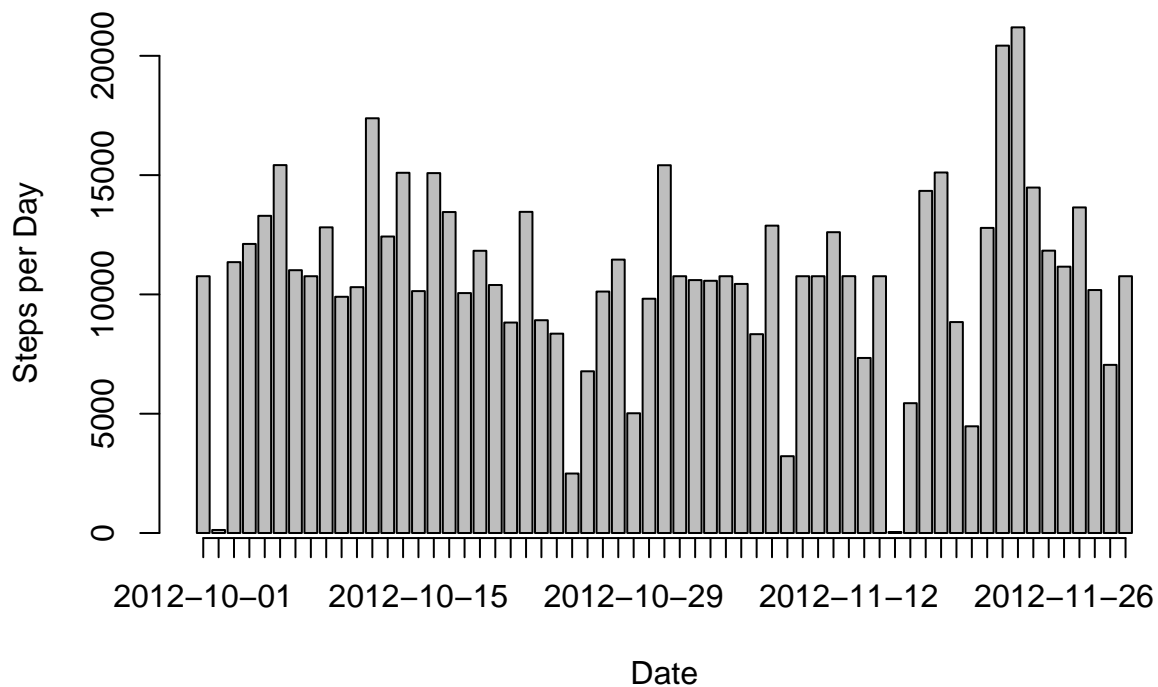
```
## calculate and display the number of rows with missing step data
paste("The number of rows with missing values for steps:", sum(is.na(datActivity$steps)))
```

```
## [1] "The number of rows with missing values for steps: 2304"
```

```
## impute NA values by copying the dataframe, iterating through each row and
## if the value is missing insert the corresponding average steps for that time interval
imputedActivity <- datActivity
for (i in 1:nrow(imputedActivity)) {
  if (is.na(imputedActivity$steps[i]))
    imputedActivity$steps[i] <- stepsbytime[stepsbytime$interval==imputedActivity$interval[i],2]
}
```

The overall shape of the histogram is not different except for fewer days with no steps recorded. The mean value increased significantly and is closer to the median.

```
## summarize the imputed dataset grouped by date and calculate the total steps per day
imputedstepsbyday <- imputedActivity %>% group_by(date) %>% summarize(total=sum(steps, na.rm=TRUE))
## create a histogram of total steps per day
barplot(total~date, data=imputedstepsbyday, axis.lty=1, xlab="Date", ylab="Steps per Day")
```



```
## show summary statistics of overall mean and median of the imputed dataset
paste("Mean total number of steps per day:", round(mean(imputedstepsbyday$total)))
```

```
## [1] "Mean total number of steps per day: 10766"
```

```
paste("Median total number of steps per day:", median(imputedstepsbyday$total))
```

```
## [1] "Median total number of steps per day: 10762"
```

Are there differences in activity patterns between weekdays and weekends?

On weekdays the step activity is predominantly in the morning, whereas on weekends it is more uniform throughout the day.

```
## add weekday/weekend factor variable in imputed dataset
imputedActivity$day <- as.factor(ifelse(weekdays(imputedActivity$date, abbreviate=TRUE) %in% c("Sat", "Sun"), "weekend", "weekday"))

## summarize the imputed dataset grouped by whether weekday or weekend and time interval
## calculate the average steps in each time interval
wdaystepsbytime <- imputedActivity %>% group_by(day, interval) %>% summarize(avg=mean(steps, na.rm=TRUE))
wdaystepsbytime <- transform(wdaystepsbytime, avg=round(avg))
## create time-series plots of the average steps per time interval for the weekdays and weekends
ggplot(wdaystepsbytime, aes(interval, avg)) + geom_line() + facet_grid(day~.)
```

