

Untitled

Reproducible Research: Peer Assessment 1

Introduction

This project examines data gathered from monitoring devices such as the [Fitbit](#), [Nike Fuelband](#), and [Jawbone Up](#). These devices monitor a person's activity and they report measurements in order to help the user improve their health and/or understand their behavior.

Data

The personal monitoring device used gathers data at 5 minute intervals through out the day. The data consists of two months of data from an anonymous individual collected during the months of October and November, 2012 and include the number of steps taken in 5 minute intervals each day.

The variables in the raw data set are:

- **steps**: Number of steps taking in a 5-minute interval (missing values are coded as NA)
- **date**: The date on which the measurement was taken in YYYY-MM-DD format
- **interval**: Identifier for the 5-minute interval in which measurement was taken

The raw dataset is stored in a comma-separated-value (CSV) file and there are a total of 17,568 observations in this dataset.

Loading and preprocessing the data

First, we [download the zip file](#) and record the date and time of download.

```
temp <- tempfile()
download.file("http://d396qusza40orc.cloudfront.net/repdata%2Fdata%2Factivity.zip", temp)
downloadTime <- Sys.time()
print(downloadTime)
```

```
## [1] "2014-06-09 13:31:48 CDT"
```

We then unzip the downloaded zip file and read the resulting .csv file into a usable R data frame.

```
zipFileInfo <- unzip(temp, list=TRUE)
activity <- read.csv(unz(temp, as.character(zipFileInfo$Name)), header=TRUE)
unlink(temp)
```

It is worthwhile to look at the structure of the data.

```
head(activity)
```

```
##   steps      date interval
## 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
```

The data appear tidy and in a usable format.

What is mean total number of steps taken per day?

To give a general sense of activity patterns, we calculate the total number of steps per day. For this question, we exclude NA values. We perform the analysis using the dplyr package. We group the activity data by date, and then within each group, we summarize by totaling the number of steps.

```
library(dplyr)
```

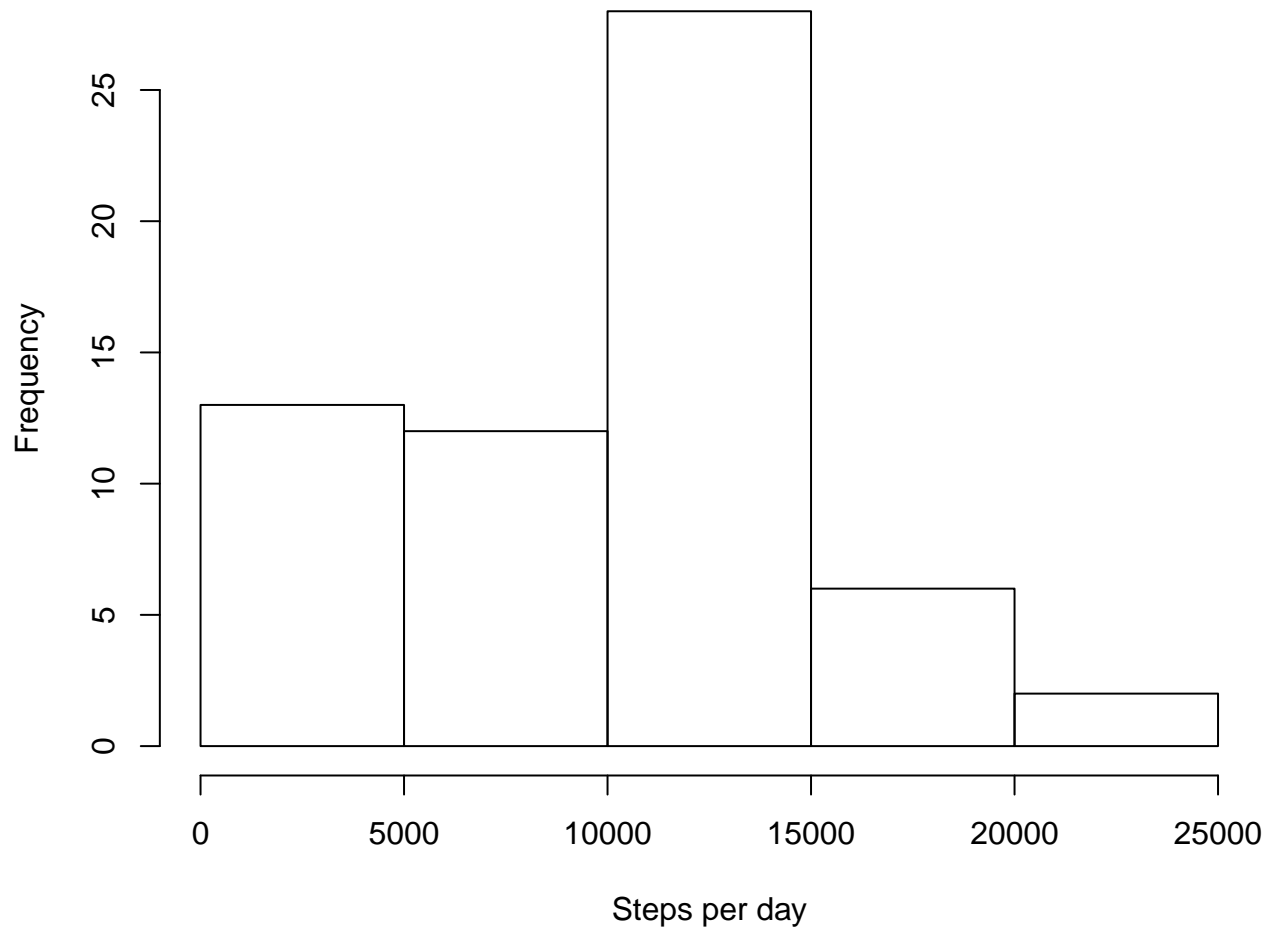
```
##
## Attaching package: 'dplyr'
##
## The following objects are masked from 'package:stats':
##
##   filter, lag
##
## The following objects are masked from 'package:base':
##
##   intersect, setdiff, setequal, union
```

```
grouped <- group_by(activity,date)
totalGrouped <- summarise(grouped,totalSteps=sum(steps,na.rm=TRUE))
```

Here is a histogram of the total number of steps per day.

```
hist(totalGrouped$totalSteps,xlab="Steps per day",main="Histogram of steps per day")
```

Histogram of steps per day



We can also calculate the mean and median values.

```
meansteps <- mean(totalGrouped$totalSteps)
mediansteps <- median(totalGrouped$totalSteps)
```

The mean value is 9354.2295 steps per day and the median value is 10395 steps per day.

What is the average daily activity pattern?

Imputing missing values

Are there differences in activity patterns between weekdays and weekends?

Assignment