

PA1_template.Rmd

Jacob Asariah

August 23, 2018

Reproducible Research Course Project 1

Load the data and Process/transform the data (if necessary) into a format suitable for your analysis.

```
setwd("0://Jacob//Coursera//Course5-Reproducible Research//Project")

activitydf <- read.csv(file="activity.csv", header=TRUE, sep=",")

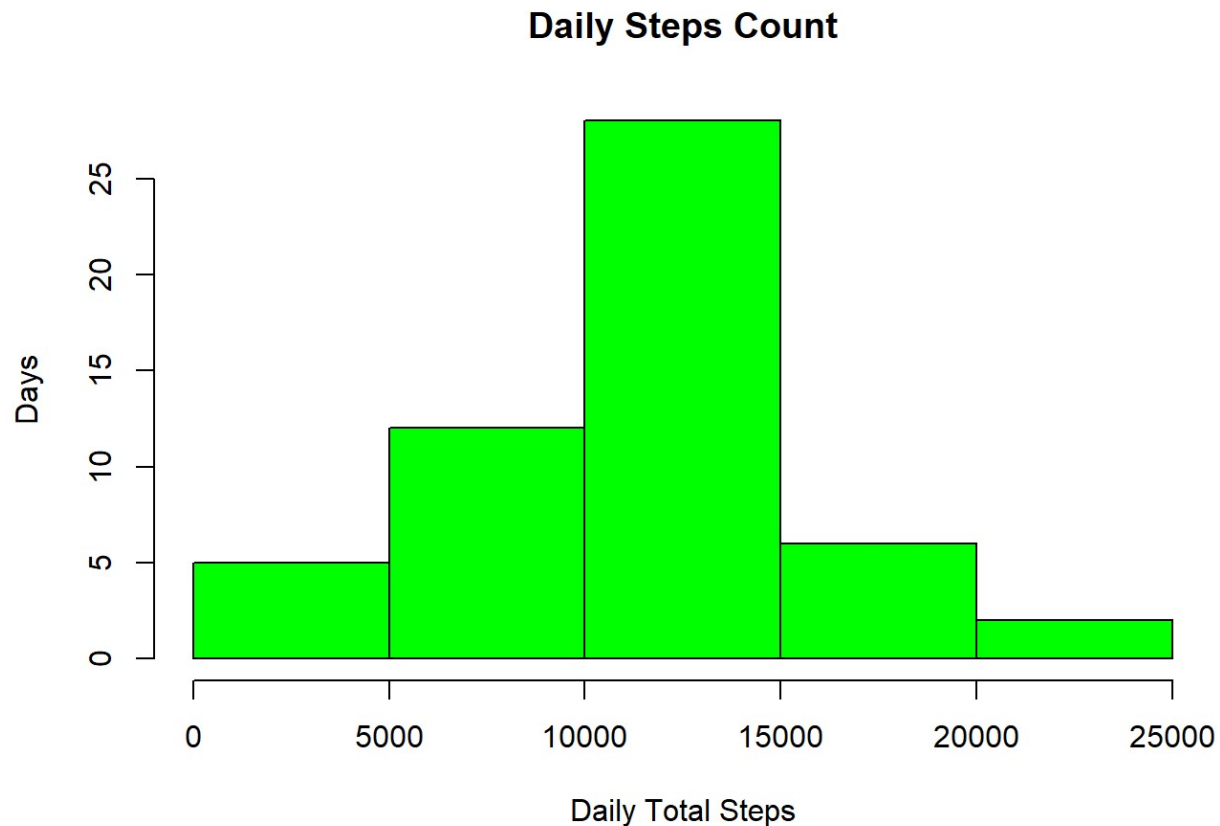
summary(activitydf)
```

```
##      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
```

What is mean total number of steps taken per day?

```
DailyStepsCount <- aggregate(steps ~ date, subset(activitydf,!is.na(activitydf$steps)), sum)

hist(DailyStepsCount$steps,col="green",xlab="Daily Total Steps",ylab="Days",main="Daily Steps Count")
```



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

```
mean(DailyStepsCount$steps)
```

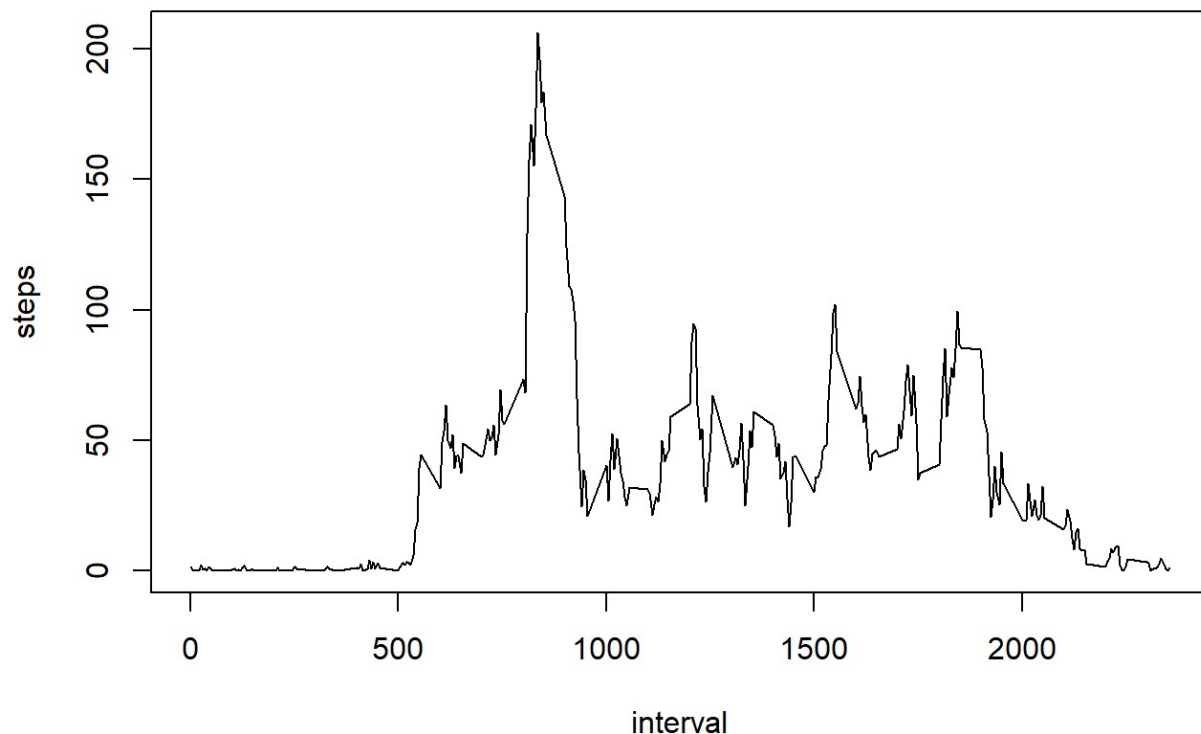
```
## [1] 10766.19
```

```
median(DailyStepsCount$steps)
```

```
## [1] 10765
```

What is the average daily activity pattern?

```
IntervalCount <- aggregate(steps~interval, data=activitydf, mean, na.rm=TRUE)
plot(steps~interval, data=IntervalCount, type="l")
```



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

```
IntervalCount[which.max(IntervalCount$steps),]$interval
```

```
## [1] 835
```

Imputing missing values

Calculate and report the total number of missing values in the dataset.

```
sum(is.na(activitydf$steps))
```

```
## [1] 2304
```

Devise a strategy for filling in all of the missing values in the

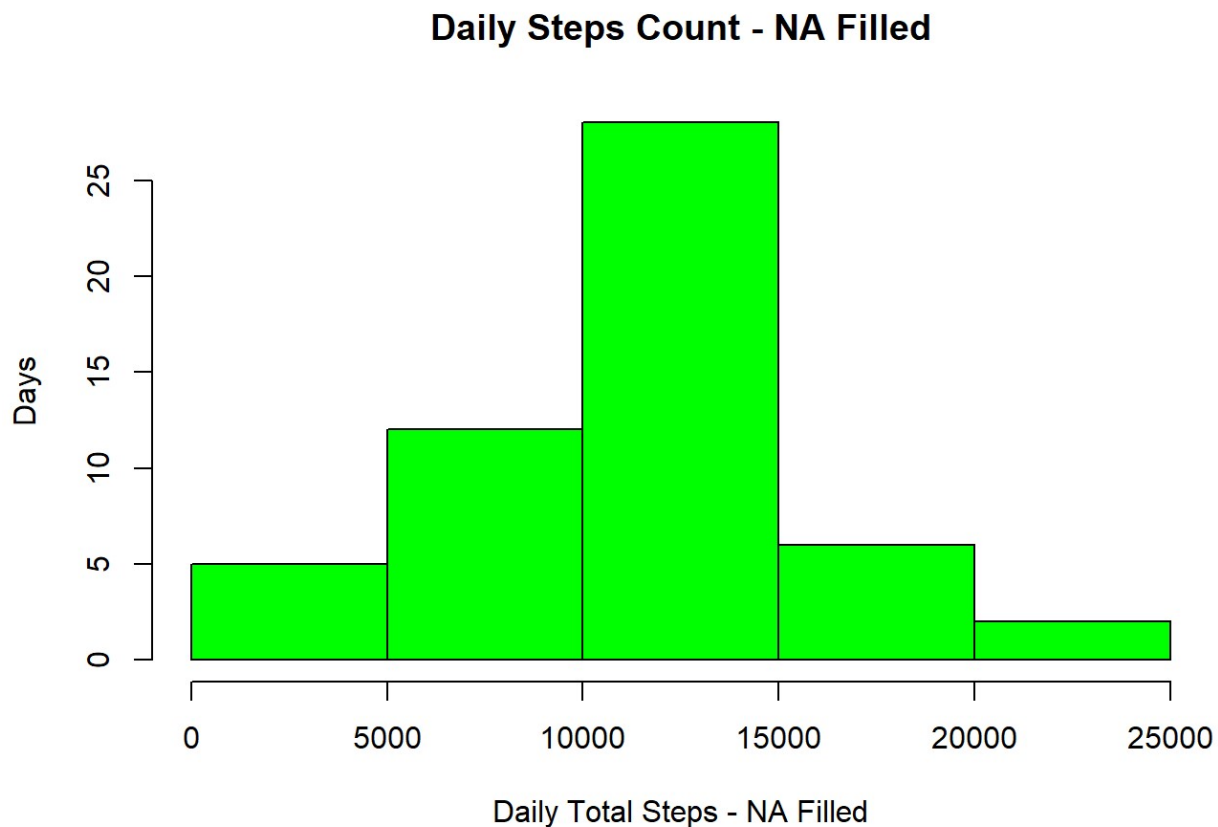
dataset.

Create a new dataset that is equal to the original dataset but with the missing data filled in.

```
activitydfNA <- activitydf
activitydfNA[["steps"]][is.na(activitydfNA[["steps"]])] <- 0
DailyStepsCount <- aggregate(steps ~ date, subset(activitydfNA,!is.na(activitydf$steps)), sum)
```

Make a histogram of the total number of steps taken each day and Calculate and report the mean and median total number of steps taken per day.

```
hist(DailyStepsCount$steps,col="green",xlab="Daily Total Steps - NA Filled",ylab="Days",main="Daily Steps Count - NA Filled")
```



Calculate and report the mean and median total number of

steps taken per day.

```
mean(DailyStepsCount$steps)
```

```
## [1] 10766.19
```

```
median(DailyStepsCount$steps)
```

```
## [1] 10765
```

Are there differences in activity patterns between weekdays and weekends?

Create a new factor variable in the dataset with two levels - “weekday” and “weekend” indicating whether a given date is a weekday or weekend day.

```
activitydfNA$date <- as.Date(strptime(activitydfNA$date, format="%Y-%m-%d"))

weekdaysdf <- transform(activitydfNA, wday = ifelse(weekdays(activitydfNA$date) == "Saturday" | weekdays(activitydfNA$date) == "Sunday", "weekend", "weekday"))
```

Make a panel plot containing a time series plot type=“l” of the 5-minute interval (x-axis) and the average number of steps taken, averaged across all weekday days or weekend days (y-axis).

```
library(lattice)

xyplot(steps ~ interval | wday, weekdaysdf, type = "l", layout = c(1, 2), xlab = "Interval", ylab = "Number of steps")
```

