

Reproducible Research Project 1

Overview

This is the report associated with the first assignment in the course Reproducible Research and includes the code, questions and answers. Additional details are in the associated Readme file in the github repository.

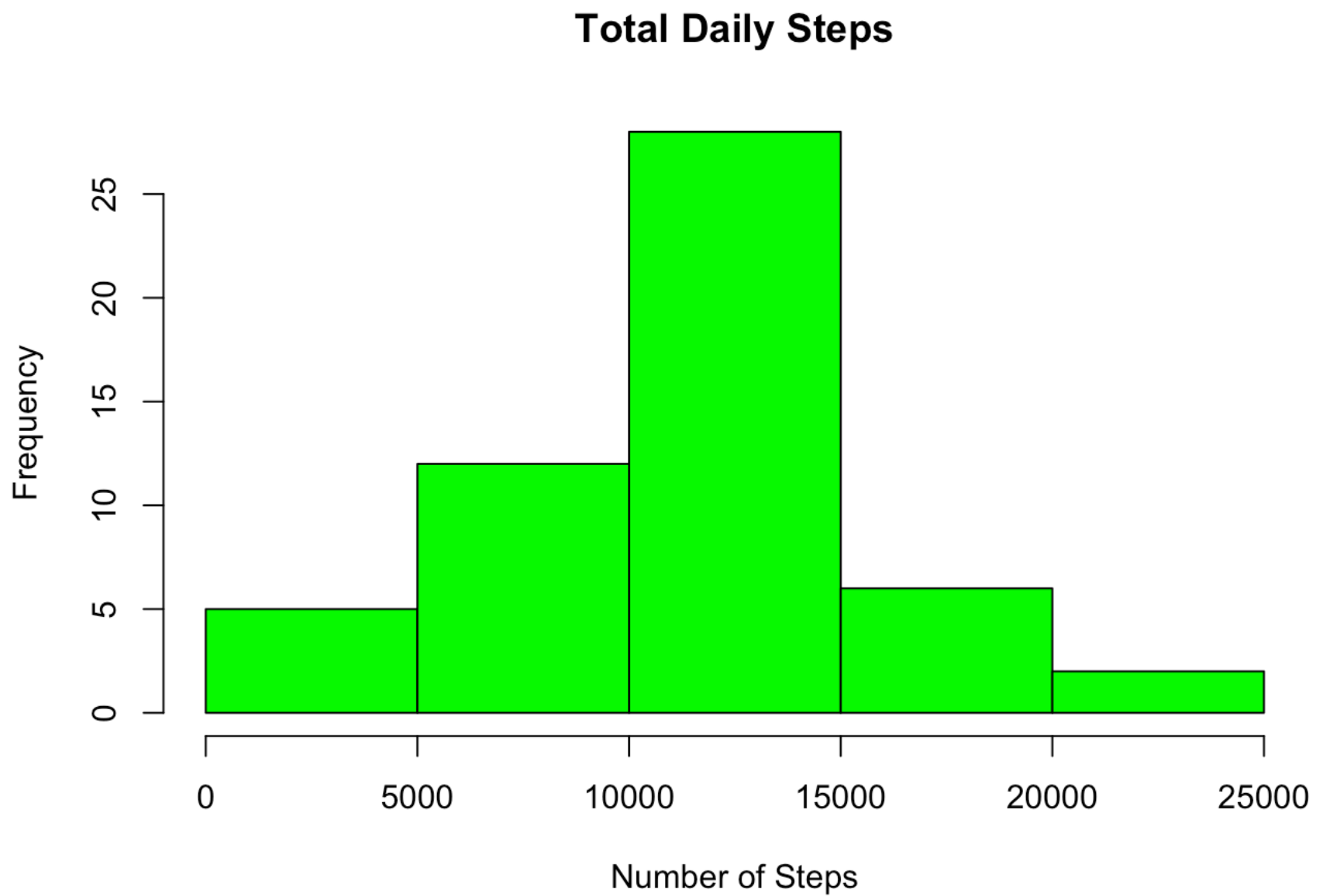
load data and graphic library

```
data <- read.csv("activity.csv")  
library(lattice)
```

What is the mean total of steps taken per day?

To answer this question it is necessary to calculate the number of steps taken per day (ignoring missing data) the output is a histogram and the calculations and outputs for rmean and rmedian.

```
dailysteps <- aggregate(steps ~ date, data, sum)  
hist(dailysteps$steps, main = paste("Total Daily Steps"), col="green", xlab="Number o  
f Steps")
```



```
rmean <- mean(dailysteps$steps)
rmedian <- median(dailysteps$steps)
print(rmean)
```

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

```
print(rmedian)
```

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

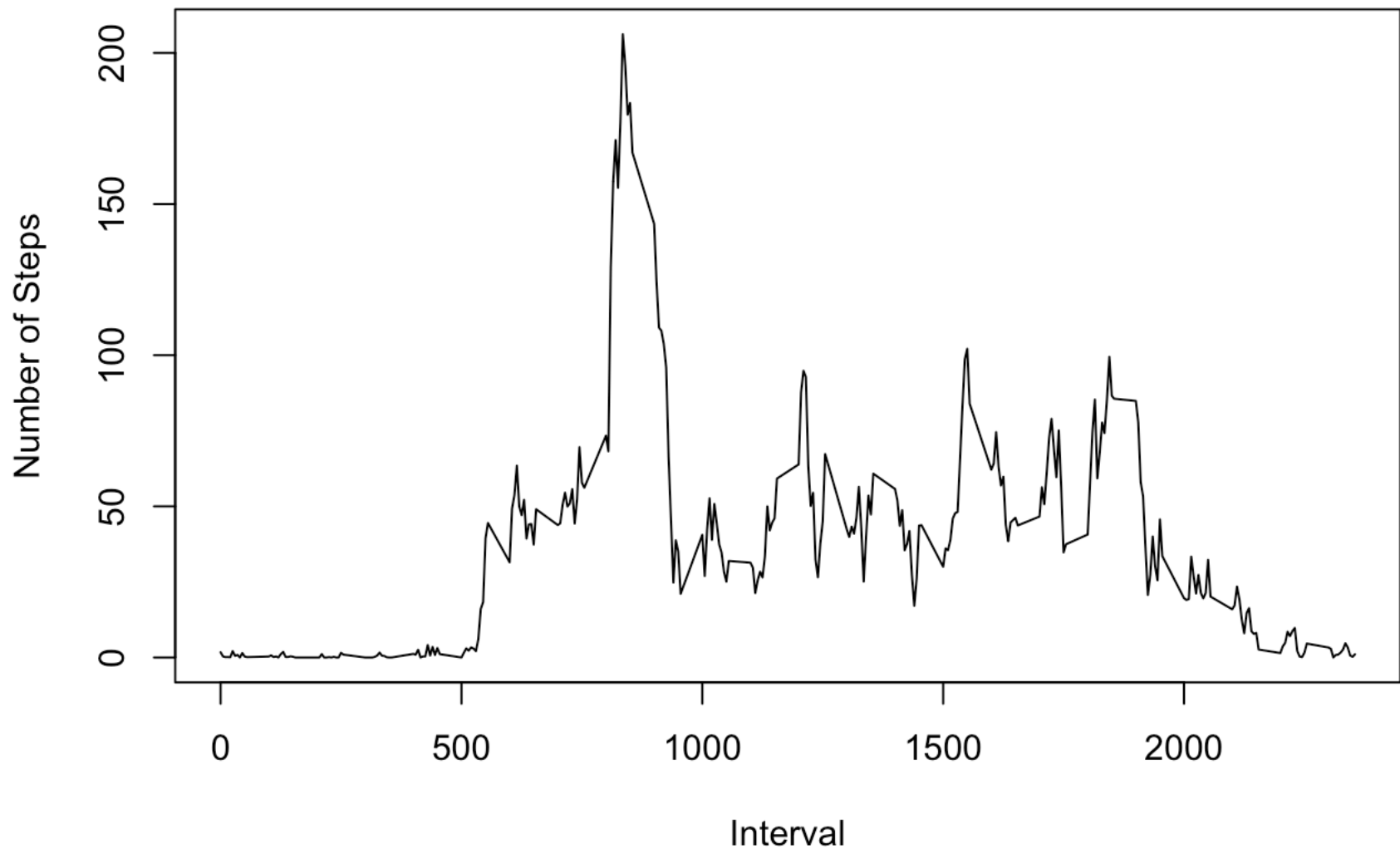
What is the daily activity pattern?

This is comprised of a time series plot (i.e. type = "1") of the 5-minute interval (x-axis) and the average number of steps taken, averaged across all days (y-axis). The 5-minute interval, on average across all the days in the data set, containing the maximum number of steps is max_interval is printed out as well.

```
stepsinterval <- aggregate(steps ~ interval, data, mean)
```

```
plot(stepsinterval$interval, stepsinterval$steps, type="l", xlab="Interval", ylab="Number of Steps", main="Average Daily Number of Steps by Interval")
```

Average Daily Number of Steps by Interval



```
max_interval <- stepsinterval[which.max(stepsinterval$steps),1]  
print(max_interval)
```

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

Input missing data

There are a number of days and intervals where there are missing values. The presence of missing days may introduce bias into some calculations or summaries of the data. The following code calculates and reports the total number of missing values by inserting the average for each interval. NAs were determined to be zeros. the new data set, `incdailystep`, includes the updated data. The output is a histogram and new mean and median data respectively for the new dataset.

```

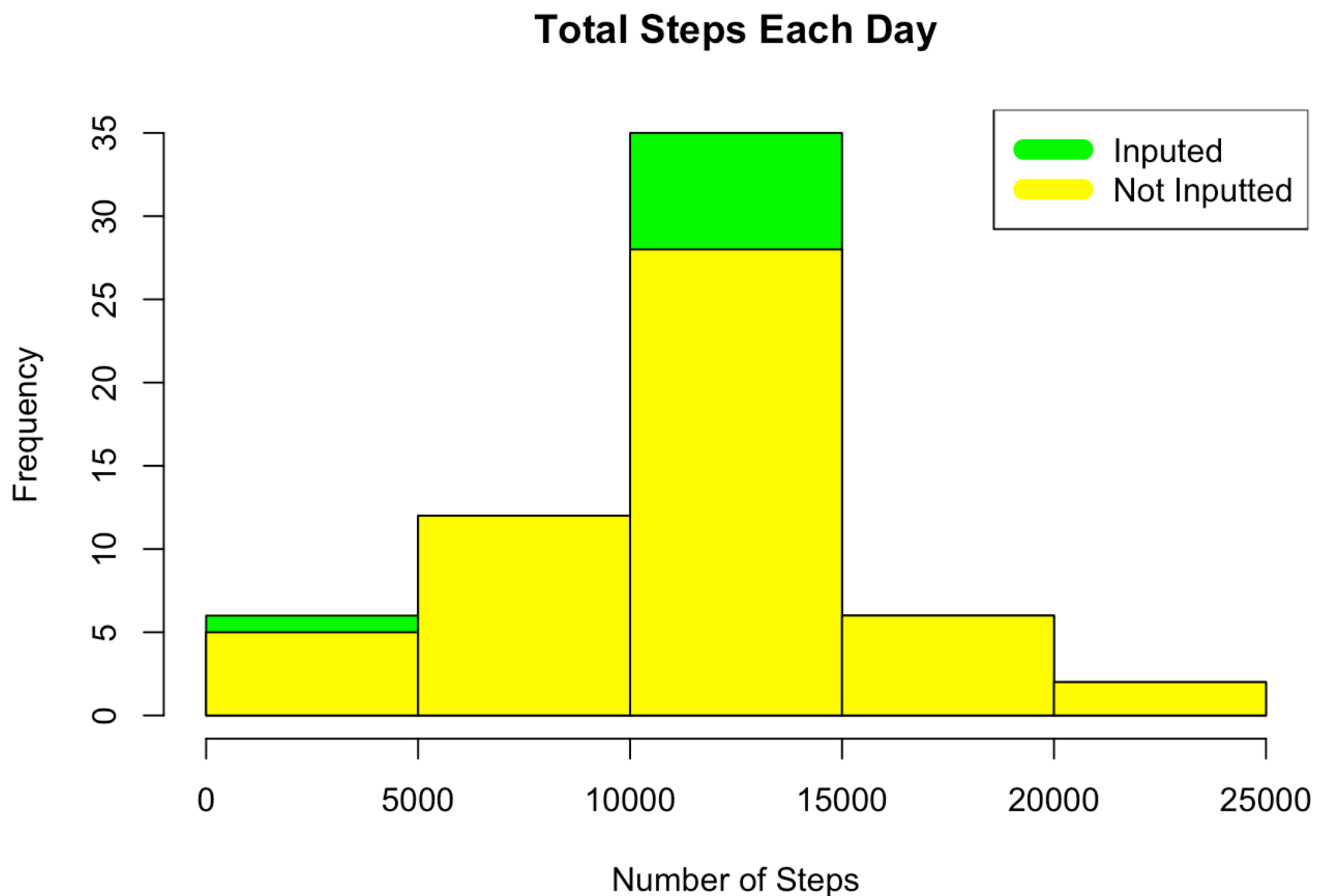
incomplete <- sum(!complete.cases(data))

datainputed <- transform(data, steps =
                           ifelse(is.na(data$steps), stepsinterval$steps
[match(data$interval, stepsinterval$interval)], data$steps)
                           )
datainputed[as.character(datainputed$date) == "2012-10-01", 1] <- 0

incdailysteps <- aggregate(steps ~ date, datainputed, sum)
hist(incdailysteps$steps, main = paste("Total Steps Each Day"), col="green", xlab="Number of Steps")

hist(dailysteps$steps, main = paste("Total Steps Each Day"), col="yellow", xlab="Number of Steps", add=T)
legend("topright", c("Inputed", "Not Inputted"), col=c("green", "yellow"), lwd=10)

```



```
#step 4
```

```
rmeaninc <- mean(incdailysteps$steps)
rmedianinc <- median(incdailysteps$steps)
```

```
meanDif <- rmeaninc - rmean
medDif <- rmedianinc - rmedian
```

```
TotalDif <- sum(incdailysteps$steps) - sum(dailysteps$steps)
print(rmeaninc)
```

```
## [1] 10589.69
```

```
print(rmeaninc)
```

```
## [1] 10589.69
```

Are there differences between activity pattern on weekdays and weekends?

Answering this question entails using the weekdays function and the output is a panel plot containing a time series plot to show the differences.

```
weekdays <- c("Monday", "Tuesday", "Wednesday", "Thursday",
              "Friday")
datainput$dow =
  as.factor(ifelse(is.element(weekdays(as.Date(datainput$date))), weekdays), "Weekday", "Weekend")
)

incstepsinterval <- aggregate(steps ~ interval + dow, datainput, mean)

xyplot(incstepsinterval$steps ~ incstepsinterval$interval | incstepsinterval$dow, main=
"Average Daily Steps by Interval", xlab="Interval", ylab="Steps", layout=c(1,2), type="l")
```

Average Daily Steps by Interval

