

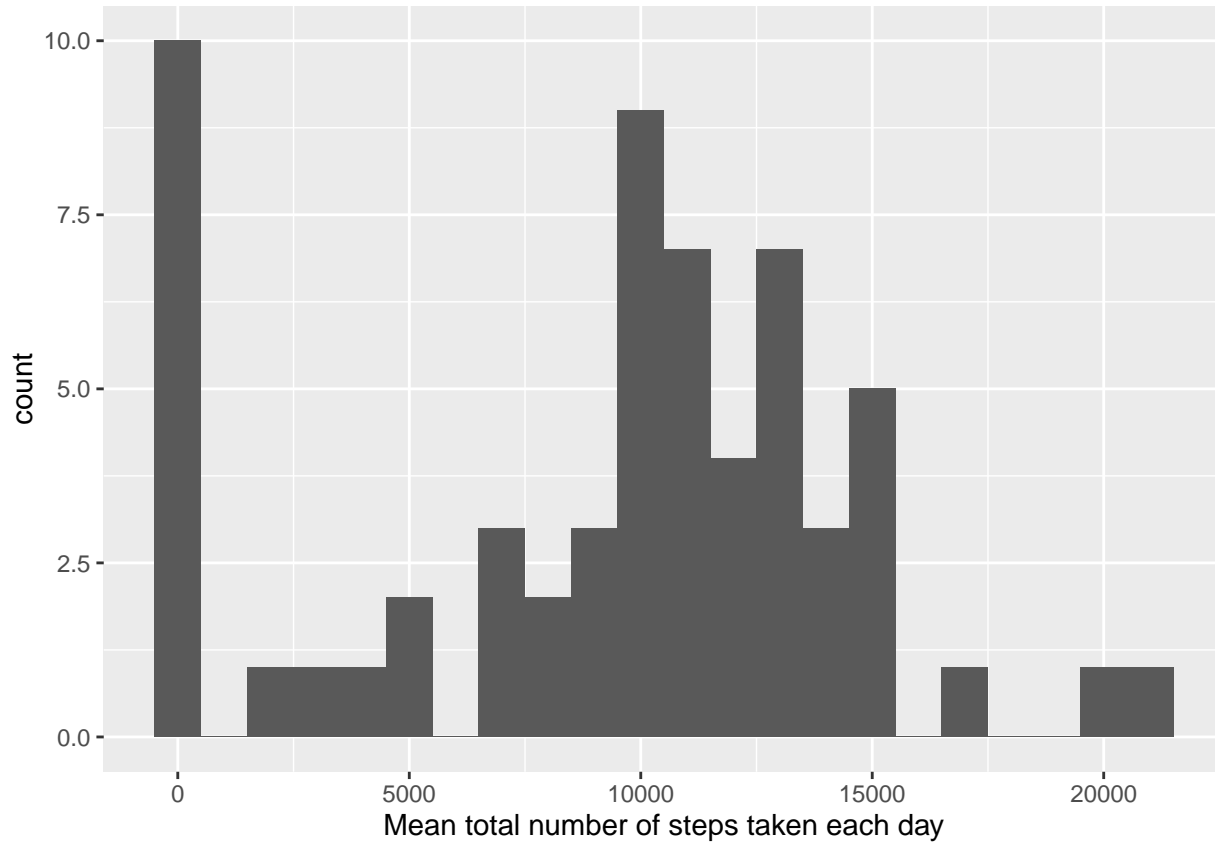
Reproducible Research: Peer Assessment 1

Code for reading in the dataset and/or processing the data

```
unzip(zipfile="activity.zip")
data <- read.csv("activity.csv")
```

Histogram of the total number of steps taken each day

```
library(ggplot2)
total.steps <- tapply(data$steps, data$date, FUN=sum, na.rm=TRUE)
qplot(total.steps, xlab="Mean total number of steps taken each day", binwidth=1000)
```



```
# Mean number of steps taken each day
mean(total.steps, na.rm=TRUE)
```

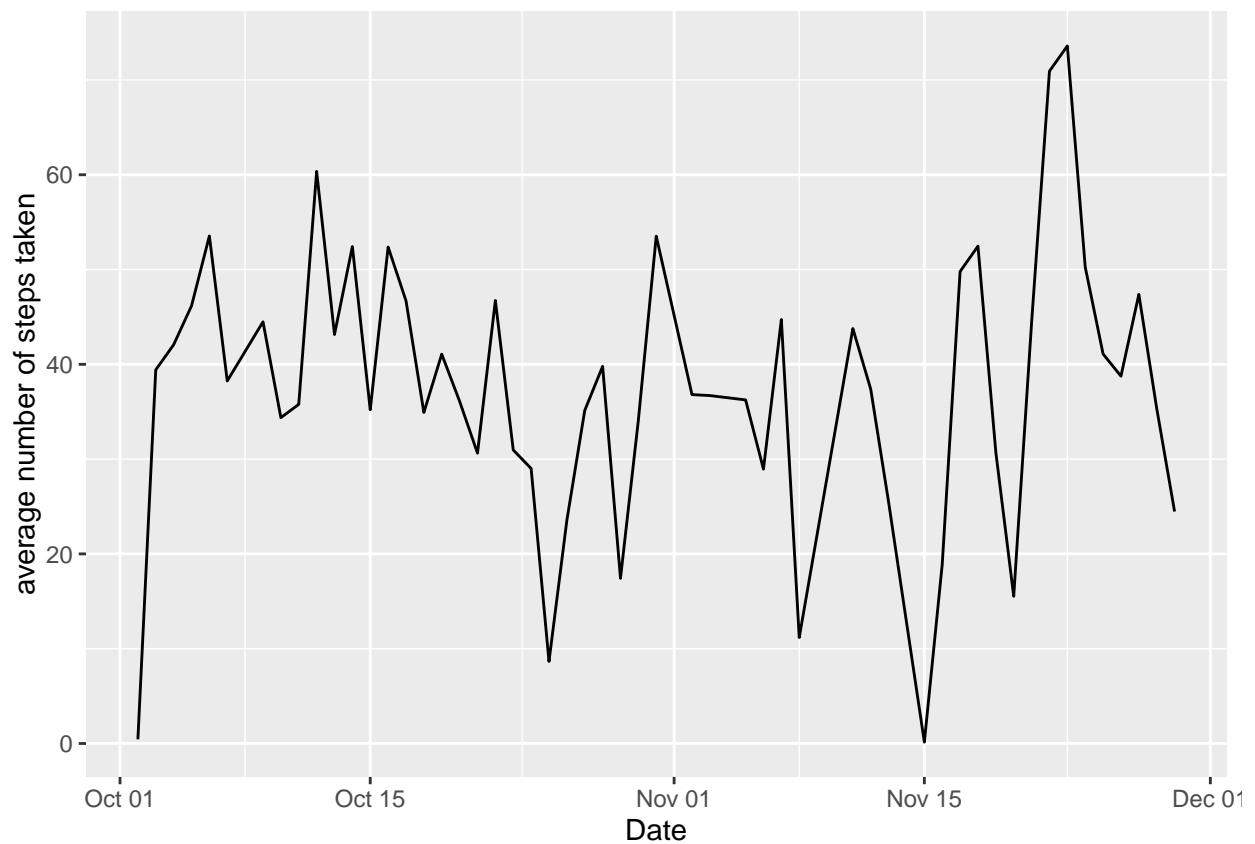
```
## [1] 9354.23
```

```
# Median and median number of steps taken each day
median(total.steps, na.rm=TRUE)
```

```
## [1] 10395
```

Time series plot of the average number of steps taken

```
library(ggplot2)
av.steps <- tapply(data$steps, data$date, FUN=mean, na.rm=TRUE)
q<-data.frame(unique(data$date),unnamed(av.steps))
colnames(q) <- c("Date", 'Steps')
q <- na.omit(q)
q$Date <- as.Date(q$Date)
ggplot(data=q, aes(x=Date, y=Steps))+geom_line()+ylab("average number of steps taken")
```



```
averages <- aggregate(x=list(steps=data$steps), by=list(interval=data$interval),
                      FUN=mean, na.rm=TRUE)
```

The 5-minute interval that, on average, contains the maximum number of steps

```
averages[which.max(averages$steps),]
```

```
##      interval      steps
## 104         835 206.1698
```

Code to describe and show a strategy for imputing missing data

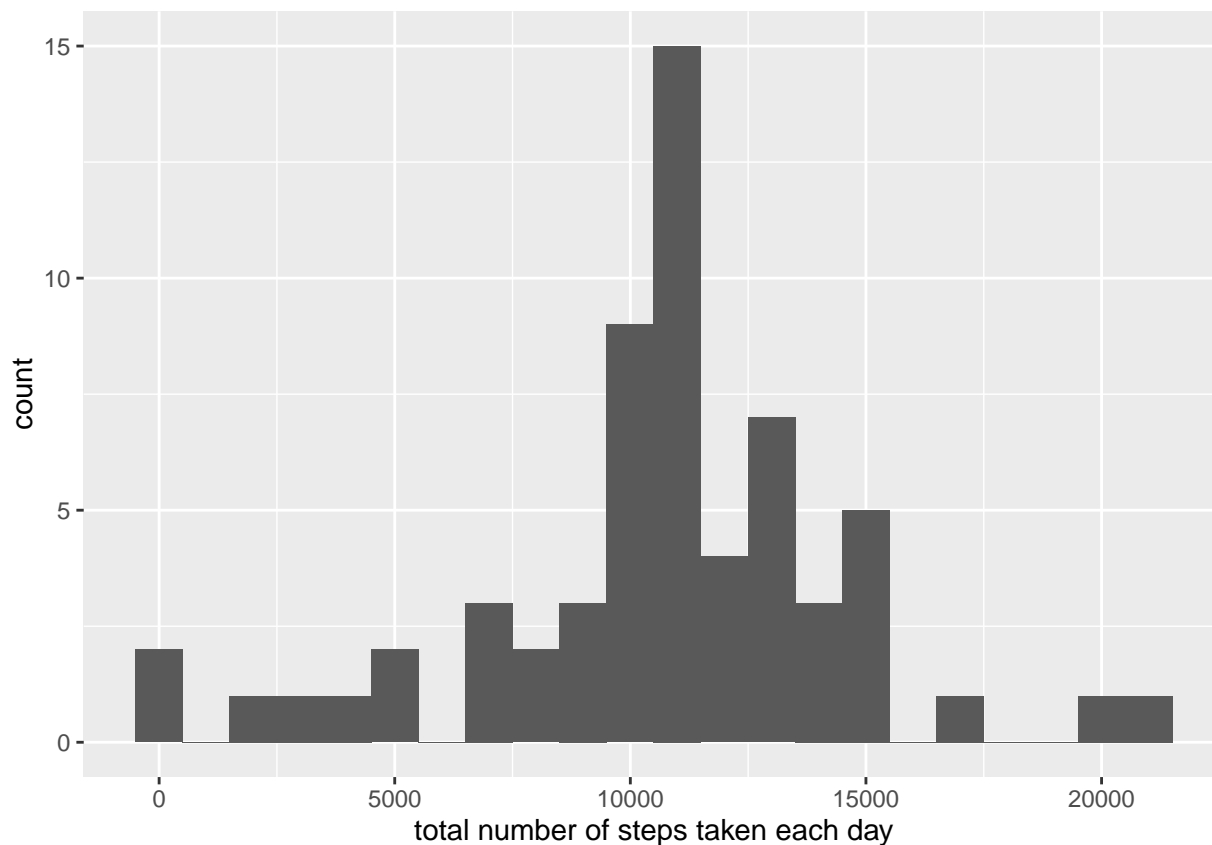
```
missing <- is.na(data$steps)
# How many missing
table(missing)

## missing
## FALSE TRUE
## 15264 2304

# Replace each missing value with the mean value of its 5-minute interval
fill.value <- function(steps, interval) {
  filled <- NA
  if (!is.na(steps))
    filled <- c(steps)
  else
    filled <- (averages[averages$interval==interval, "steps"])
  return(filled)
}
filled.data <- data
filled.data$steps <- mapply(fill.value, filled.data$steps, filled.data$interval)
```

Histogram of the total number of steps taken each day after missing values are imputed

```
total.steps = tapply(filled.data$steps, filled.data$date, FUN=sum)
qplot(total.steps, binwidth=1000, xlab="total number of steps taken each day")
```



```
mean(total.steps)
```

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

```
median(total.steps)
```

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

Panel plot comparing the average number of steps taken per 5-minute interval across weekdays and weekends

```
weekday.or.weekend <- function(date) {  
  day <- weekdays(date)  
  if (day %in% c("Monday", "Tuesday", "Wednesday", "Thursday", "Friday"))  
    return("weekday")  
  else if (day %in% c("Saturday", "Sunday"))  
    return("weekend")  
  else  
    stop("invalid date")  
}  
filled.data$date <- as.Date(filled.data$date)  
filled.data$day <- sapply(filled.data$date, FUN=weekday.or.weekend)  
averages <- aggregate(steps ~ interval + day, data=filled.data, mean)  
ggplot(averages, aes(interval, steps)) + geom_line() + facet_grid(day ~ .) +  
  xlab("5-minute interval") + ylab("Number of steps")
```

