Reproducible Research-Week 2

7/13/2019

Overview of tasks and files

Task: Our objective will be to analyzie the total, mean and median number of steps taken per day from October-November 2012. We will be plotting the average number of steps taken per 5 minute interval and evaluating the max. We will be imputing missing step data based on the average step data for the intervals and evaluating the average number of steps taken between weekdays and weekends.

files needed:

- zip file titled: repdata_data_activity.zip
- R markdown file: reprodresearch_wk2.Rmd

Unzip raw data for analysis

- Expect 17568 observations with 3 variables (steps, date, interval)

```r
getwd()
```

```
## [1] "C:/Users/rnowak/Documents/Coursera/reproducible research/week2"
```

```r
actz<-unzip("repdata_data_activity.zip")
act<-read.csv(actz)
rm(actz)

head(act)
```

```
##   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
```

```r
act1<-na.omit(act)
act1$date<- as.Date(act1$date)
```

Calculate the total number of steps taken per day, visualize and summarize

- generate data
- histogram
- mean median summary statistics

```
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
```

```
#generate data
totstep<-act1 %>%
    group_by(date) %>%
    summarize(tot=sum(steps))

#Histogram
hist(totstep$tot, xlab="Total number of steps", ylab="count",
        main="Total steps taken per day", col="blue")
```
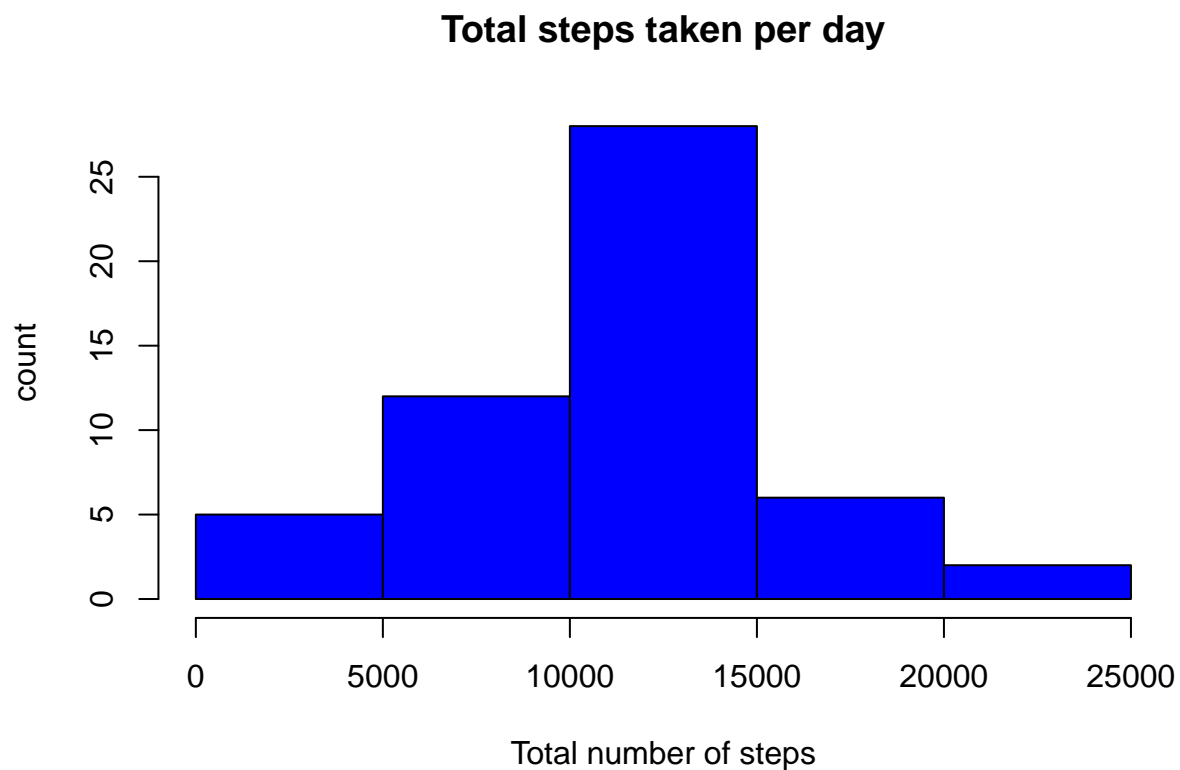

Total steps taken per day

```r
#mean and median
  mnstep<-mean(totstep$tot)
  mdstep<-median(totstep$tot)
  print(mnstep)
```

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

```r
  print(mdstep)
```

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

Time series plot of average number of steps taken

- generate summary data
- plot the average septs taken per 5 minute interval
- max average steps per interval

```r
#generate summary data
  intvave<- act1 %>%
      group_by(interval) %>%
      summarize(avestep=mean(steps))

#plot
  plot(intvave$interval, intvave$avestep, type="l",
      xlab="interval",
      ylab="average",
      main="Average steps taken per 5 minute interval")
```

Average steps taken per 5 minute interval
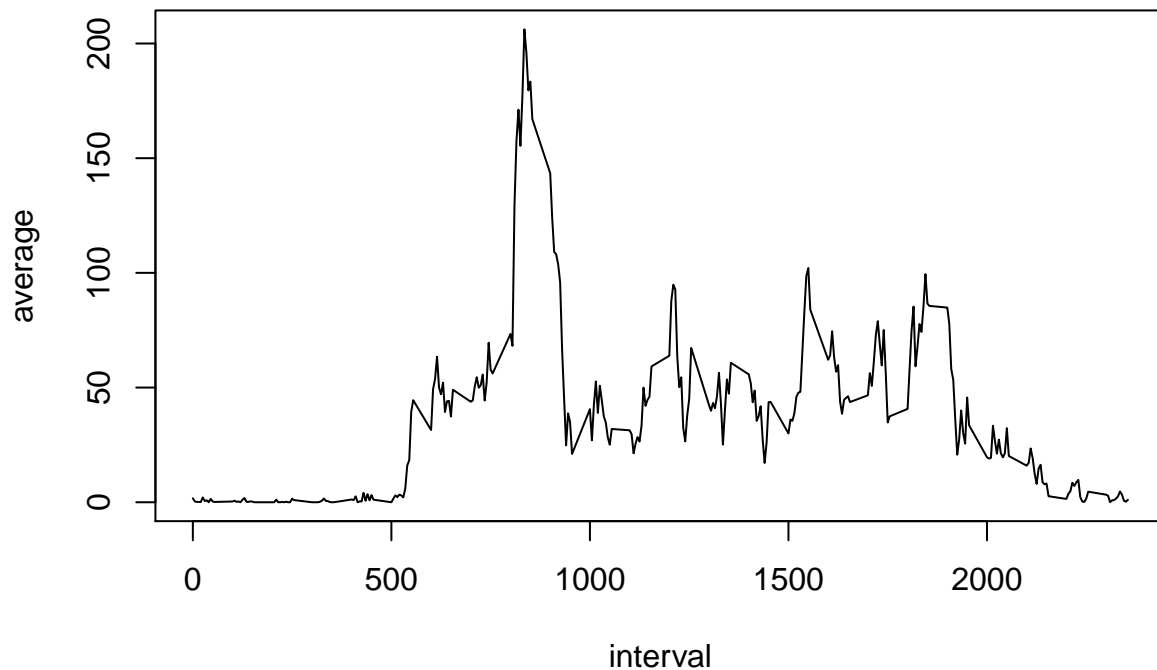

```r
#max steps
  maxstep<-intvave$interval[which.max(intvave$avestep)]
  print(maxstep)
```

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

Impute missing step data then visualize and run crude summaries

- impute first
- histogram of total number of steps per day after imputation
- mean and median number of steps taken each day after imputation

```r
#impute
  actimp<-act
  for (i in 1:nrow(actimp)){
      if (is.na(actimp$steps[i])) {
        int<-which(actimp$interval[i] == intvave$interval)
        actimp$steps[i]<-intvave[int,]$avestep
      }
  }
#format date
  actimp$date <- as.Date(actimp$date)

#summarize data
```
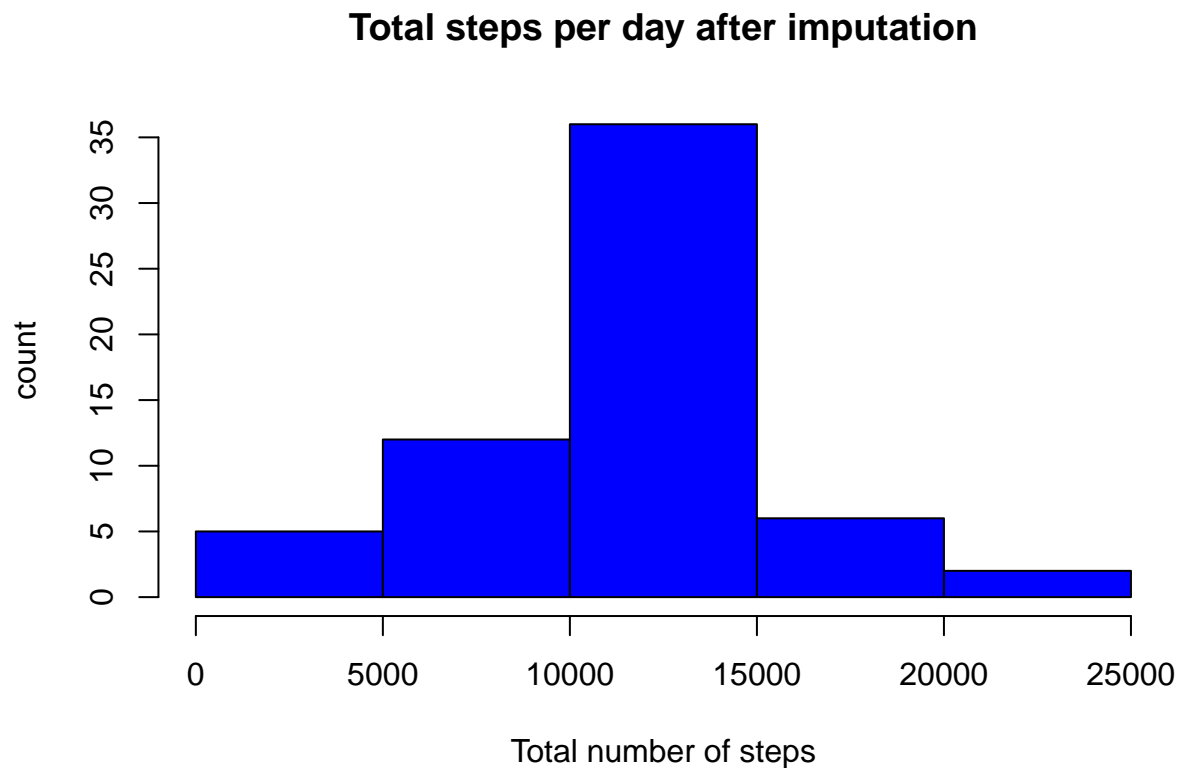
```
  totstepi<-actimp %>%
    group_by(date) %>%
    summarize(tot=sum(steps))

#histogram
  hist(totstepi$tot, xlab="Total number of steps", ylab="count",
      main="Total steps per day after imputation", col="blue")
```

Total steps per day after imputation


```
#mean and median number of steps
  mnstepi<-mean(totstepi$tot)
  mdstepi<-median(totstepi$tot)
  print(mnstepi)
```

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

```
  print(mdstepi)
```

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

Panel plot comparing average # of steps taken per 5 minute interval across weekdays and weekend

- generate data

- classify steps as weekday or weekend
- panel plot comparing average steps between weekdays and weekends

```
#generate data
  actimp$day<-weekdays(actimp$date)
  actimp$daytype<-"weekday"
  actimp$daytype[actimp$day %in% c("Saturday","Sunday")]<-"weekend"

#classify steps
  stepsd<-actimp %>%
    group_by(daytype, interval) %>%
    summarize(avestep=mean(steps))

#panel plot
  library(ggplot2)
  qplot(interval,avestep, data=stepsd,
        type="l",
        geom="line",
        xlab="interval",
        ylab="Average No. Steps",
        main="Average steps per interval for weekdays and weekend",
        facets=daytype~.)
```

```
## Warning: Ignoring unknown parameters: type
```

Average steps per interval for weekdays and weekend
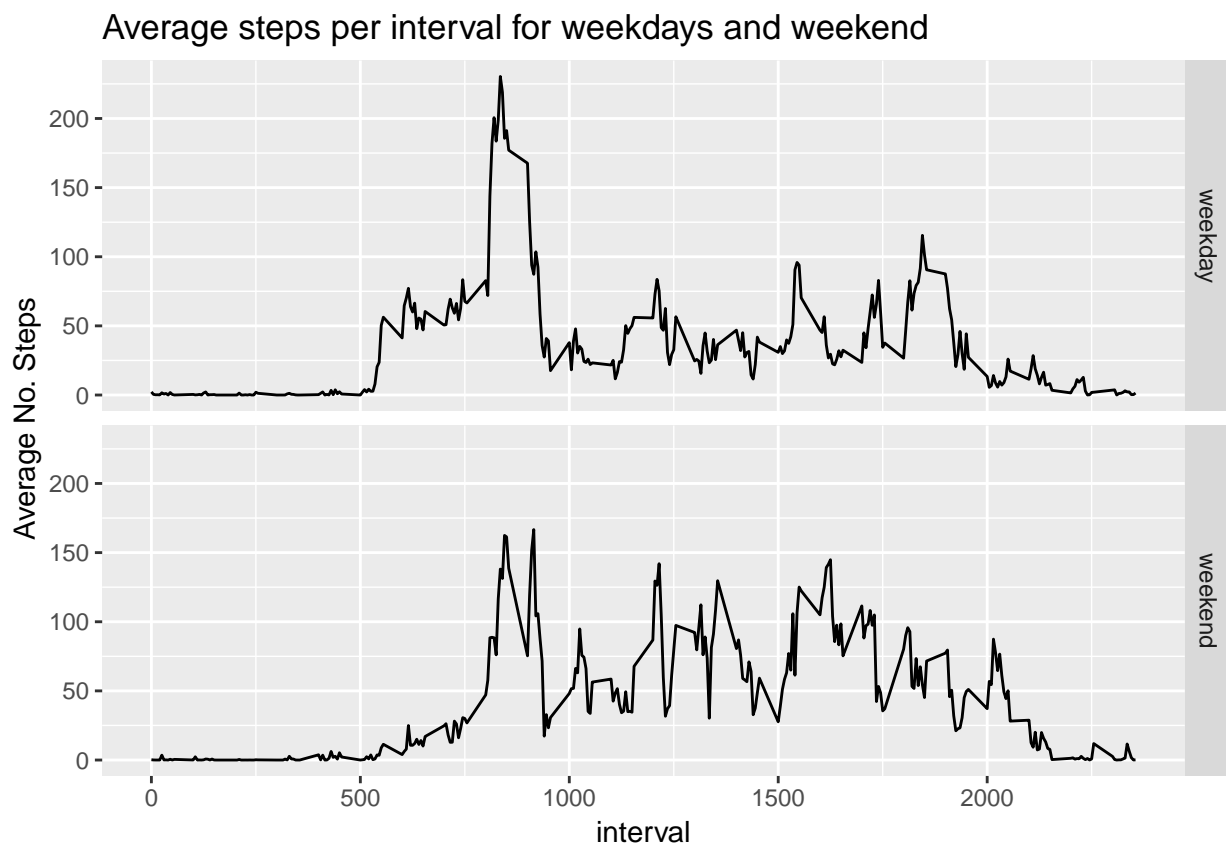
Note: more steps taken in earlier part of weekday as compared to weekend