Activity_monitor

Ying Zhu

January 31, 2017

Introduction

It is now possible to collect a large amount of data about personal movement using activity monitoring devices such as a Fitbit, Nike Fuelband, or Jawbone Up. These type of devices are part of the "quantified self" movement – a group of enthusiasts who take measurements about themselves regularly to improve their health, to find patterns in their behavior, or because they are tech geeks. But these data remain under-utilized both because the raw data are hard to obtain and there is a lack of statistical methods and software for processing and interpreting the data.

This assignment makes use of data from a personal activity monitoring device. This device collects 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.

### read data

```
data1<-read.csv("activity.csv")
head(data1)
```

```
##   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
```

compute total number of steps taken each day

```
attach(data1)
aggdata<-aggregate(steps, by=list(date), FUN=sum, na.rm=TRUE)
colnames(aggdata)<-c("date","sum_steps")
hist(aggdata$sum_steps,breaks=10,xlab="total number of steps",main="Histogram of total number of steps
```

Histogram of total number of steps taken each day
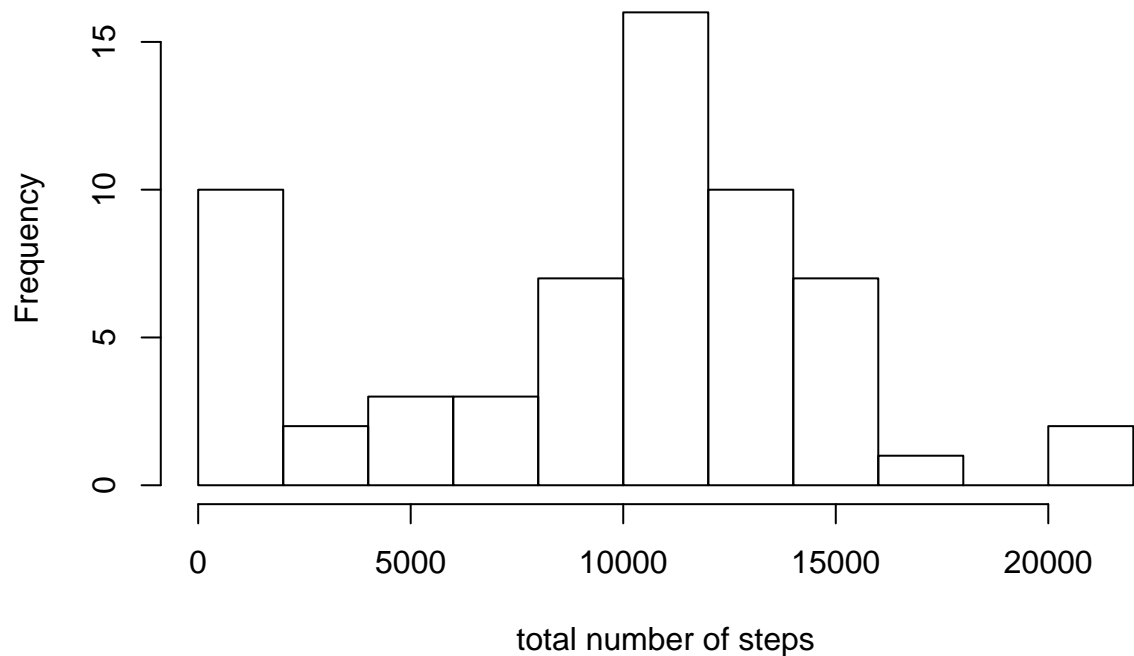
com-

pute mean and median of total steps taken each day

```r
mean(aggdata$sum_steps)
```

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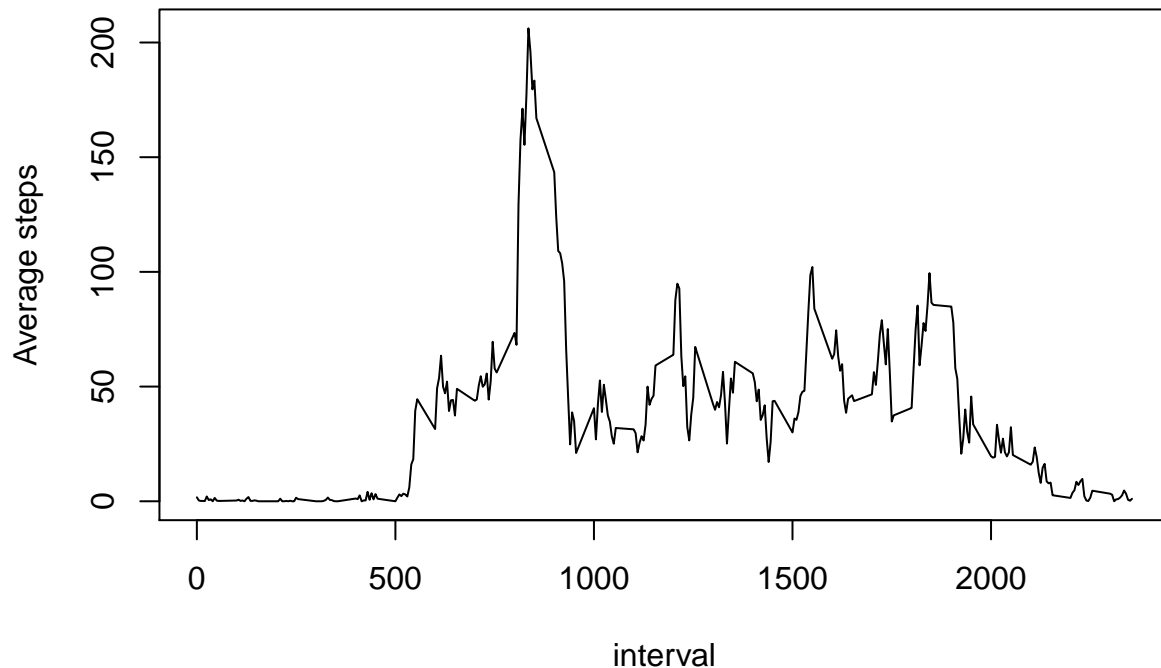```

```r
median(aggdata$sum_steps)
```

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

Time series plot of the average number of steps taken

```r
avg_steps<-aggregate(steps, by=list(interval), FUN=mean, na.rm=TRUE)
colnames(avg_steps)<-c("interval","avgsteps")
plot(avg_steps$interval,avg_steps$avgsteps,type="l",xlab="interval",ylab="Average steps",main="average
```

average number of steps taken, averaged across all days

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

```
newdata<-avg_steps[order(-avg_steps$avgsteps),]
max_step_interval<-newdata$interval[1]
max_step_interval
```

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

Code to describe and show a strategy for imputing missing data

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

```
sum(is.na(data1$steps))
```

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

Devise a strategy for filling in all of the missing values in the dataset. The strategy does not need to be sophisticated. For example, I could use the mean for that 5-minute interval across all days. Create a new dataset that is equal to the original dataset but with the missing data filled in.

```
data2<-merge(data1,avg_steps,by="interval")
NAset<-which(is.na(data2$steps)==TRUE)
data2$steps[NAset]=data2$avgsteps[NAset]
data3<-data.frame(data2$steps,data2$date,data2$interval)
colnames(data3)<-c("steps","date","interval")
head(data3)
```

```
##      steps         date interval
## 1 1.716981 2012-10-01        0
## 2 0.000000 2012-11-23        0
## 3 0.000000 2012-10-28        0
## 4 0.000000 2012-11-06        0
## 5 0.000000 2012-11-24        0
## 6 0.000000 2012-11-15        0
```

Make histogram of total steps taken each day,compute the mean, median steps taken each day. Are they different from the original?

```
attach(data3)
```

```
## The following objects are masked from data1:
##
##     date, interval, steps
```
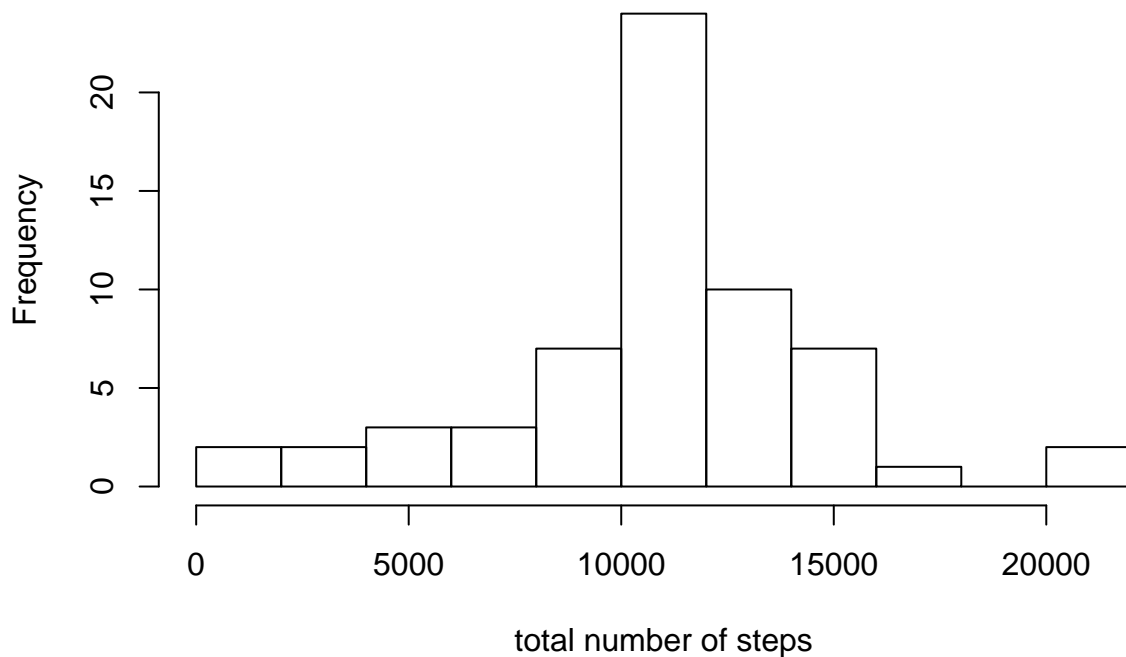
```
aggdata2<-aggregate(steps,by=list(date),FUN=sum,na.rm=TRUE)
colnames(aggdata2)<-c("date","sum_steps")
hist(aggdata2$sum_steps,breaks=10,xlab="total number of steps",main="Histogram of total number of steps
```

Histogram of total number of steps taken each day


```
mean(aggdata2$sum_steps)
```

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

```
median(aggdata2$sum_steps)
```

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

Are there differences in activity patterns between weekdays and weekends?

Create a new factor variable in the dataset with two levels – "weekday" and "weekend"

```
d1<-as.Date(data3$date)
data3$day<-weekdays(d1)
r1<-which(data3$day==c("Saturday","Sunday"))
data3$weekd<-"weekday"
data3$weekd[r1]<-"weekend"
head(data3)
```

```
##      steps       date interval     day   weekd
## 1 1.716981 2012-10-01        0  Monday weekday
## 2 0.000000 2012-11-23        0  Friday weekday
## 3 0.000000 2012-10-28        0  Sunday weekday
## 4 0.000000 2012-11-06        0 Tuesday weekday
## 5 0.000000 2012-11-24        0 Saturday weekend
## 6 0.000000 2012-11-15        0 Thursday weekday
```

make panel plots

Time series plot 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)
attach(data3)
```

```
## The following objects are masked from data3 (pos = 4):
##
##     date, interval, steps
```

```
## The following objects are masked from data1:
##
##     date, interval, steps
```

```
aggdata3<-aggregate(steps, by=list(interval,weekd), FUN=mean, na.rm=TRUE)
colnames(aggdata3)<-c("interval","weekd","steps")
xyplot(steps~interval|weekd,type="l",data=aggdata3,layout=c(1,2))
```