Reproducible Research - Project 1

Eric Glass

Saturday, January 17, 2015

code for reading in the dataset and/or processing the data:

```
library(plyr)
```

```
## Warning: package 'plyr' was built under R version 3.1.1
```

```
url<-"http://d396qusza40orc.cloudfront.net/repdata%2Fdata%2Factivity.zip"
download.file(url, "activity.zip")
unzip("activity.zip", exdir = getwd())
data<-read.csv("activity.csv")
```

histogram of the total number of steps taken each day:

```
S=with(data,tapply(steps,date,sum,na.rm=T))
hist(S,main="",xlab="total number of steps",ylab="")
```
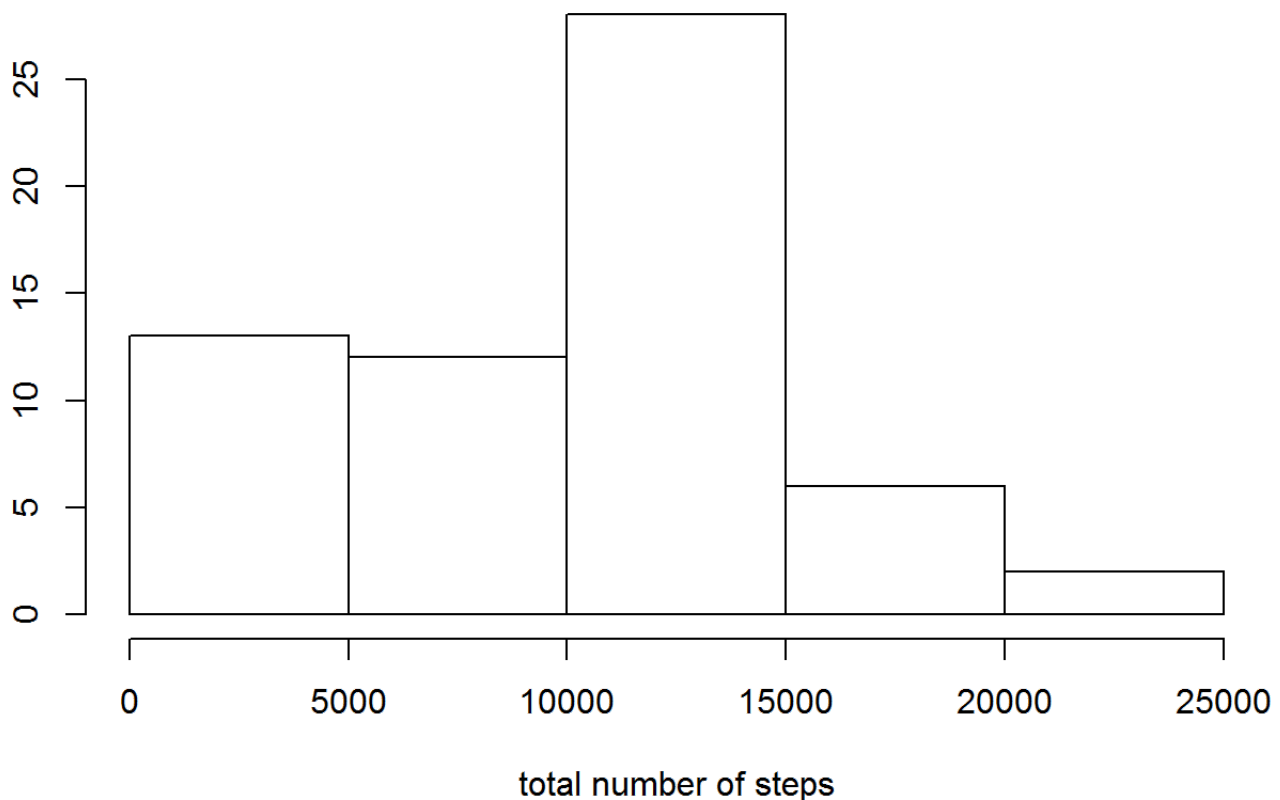
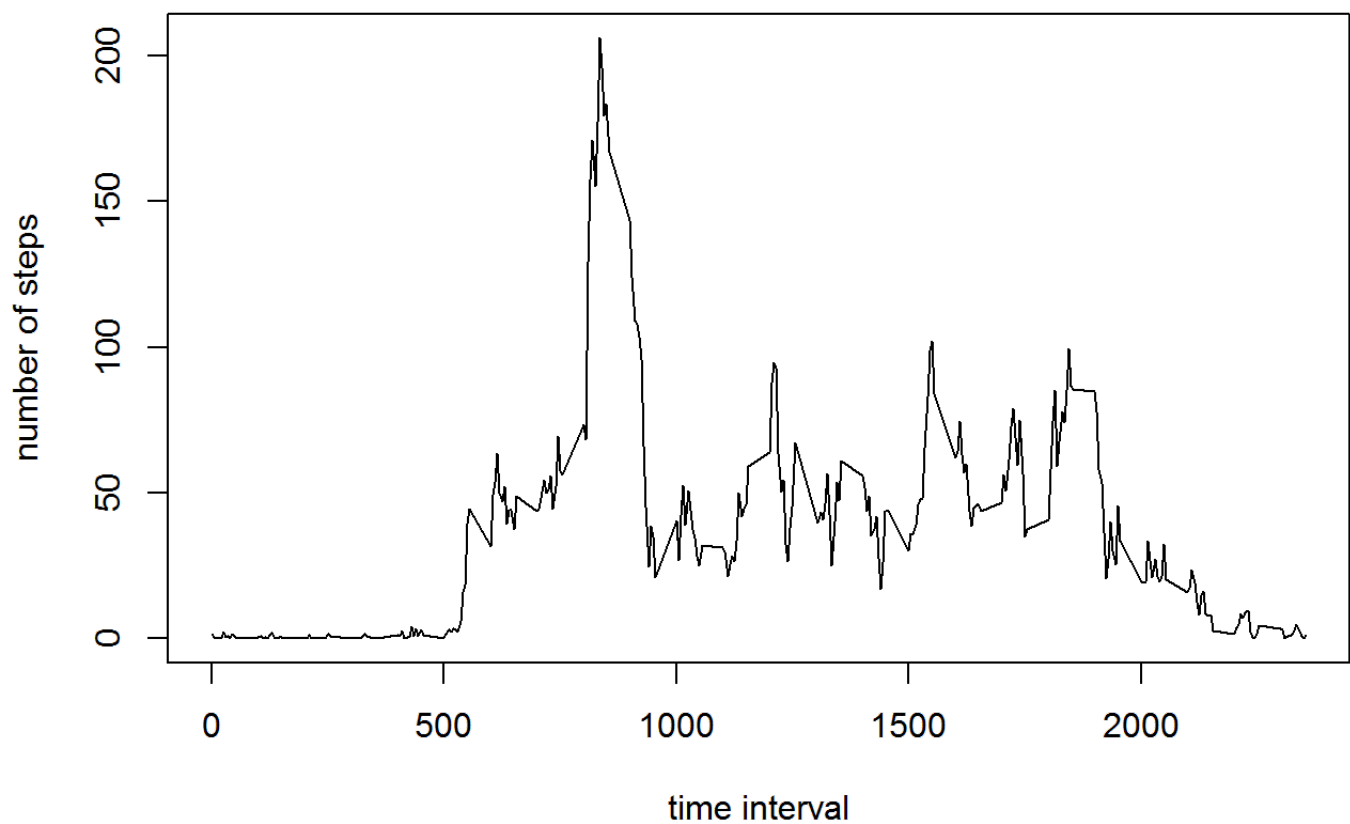mean and median number of steps taken each day:

```
mean(S)
```

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

```
median(S)
```

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

time series plot of the average number of steps taken (averaged across all days) versus the 5-minute intervals:

```
T=with(data,tapply(steps,interval,mean,na.rm=T))
plot(names(T),T,type="l",xlab="time interval",ylab ="number of steps")
```


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

```
which(T==max(T))
```

```
## 835
## 104
```

```
T[which(T==max(T))]
```

```
##      835
## 206.1698
```

describe and show with code a strategy for imputing missing data: first, identify n/a in summary table. then if n/a, impute missing data by replacement w/ actuals.

```
summary(data)
```

```
##      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
```
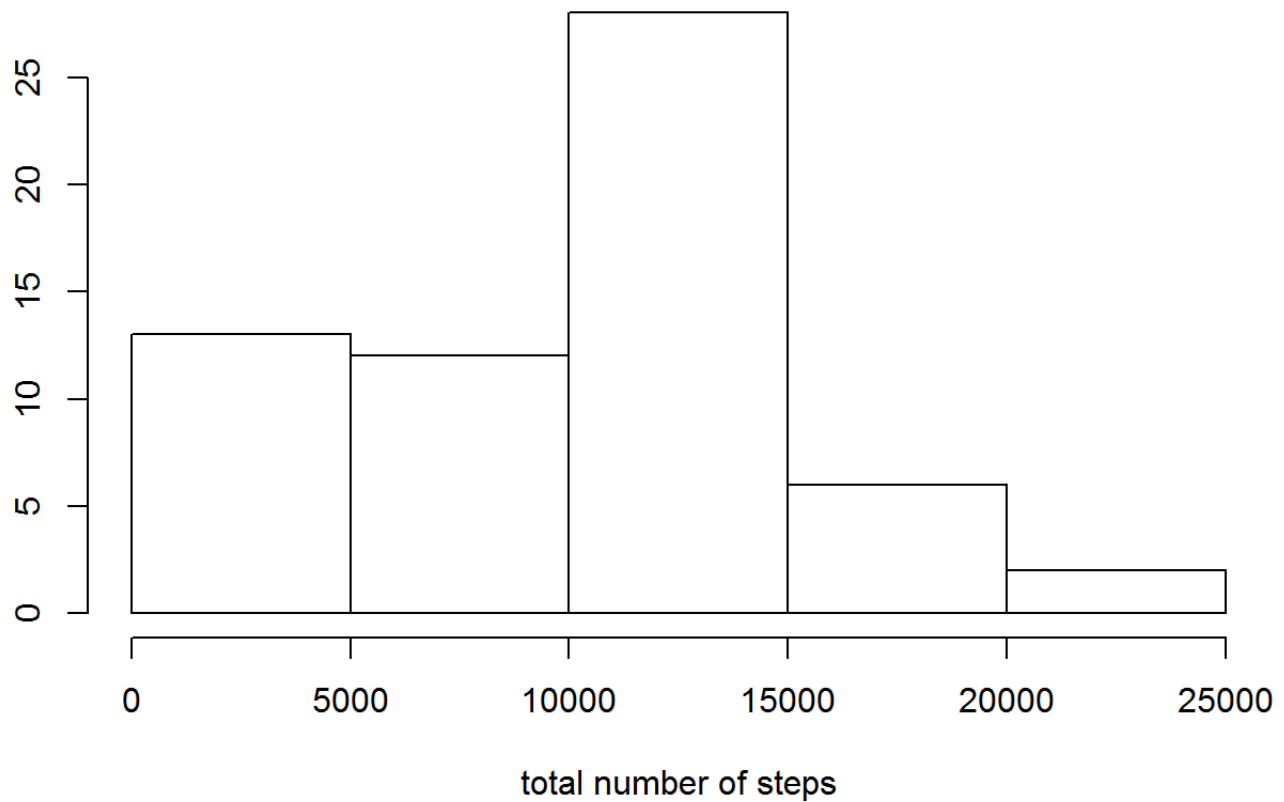
```
data2 = data
for (k in 1:17568)
{if (is.na(data[k,"steps"])) data2[k,"steps"]= T[[as.character(data[k,"interval"])]]
}
```

histogram of the total number of steps taken each day after missing values were imputed:

```
S2=with(data2,tapply(steps,date,sum,na.rm=T))
hist(S,main="",xlab="total number of steps",ylab="")
```

mean and median number of steps taken each day:

```
mean(S2)
```

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

```
median(S2)
```

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

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

```
data2$day=weekdays(as.Date(data2$date))
data2wday =subset(data2, ! day %in% c("Saturday","Sunday"))
data2wkend =subset(data2,  day %in% c("Saturday","Sunday"))

par(mfrow=c(2,1))
Tday=with(data2wday,tapply(steps,interval,mean))
plot(names(Tday),Tday,type="l",xlab="Weekday Activity",ylab ="number of steps")
Tend=with(data2wkend,tapply(steps,interval,mean))
plot(names(Tend),Tend,type="l",xlab="Weekend Activity",ylab ="number of steps")
```