

‘Reproducible Research’

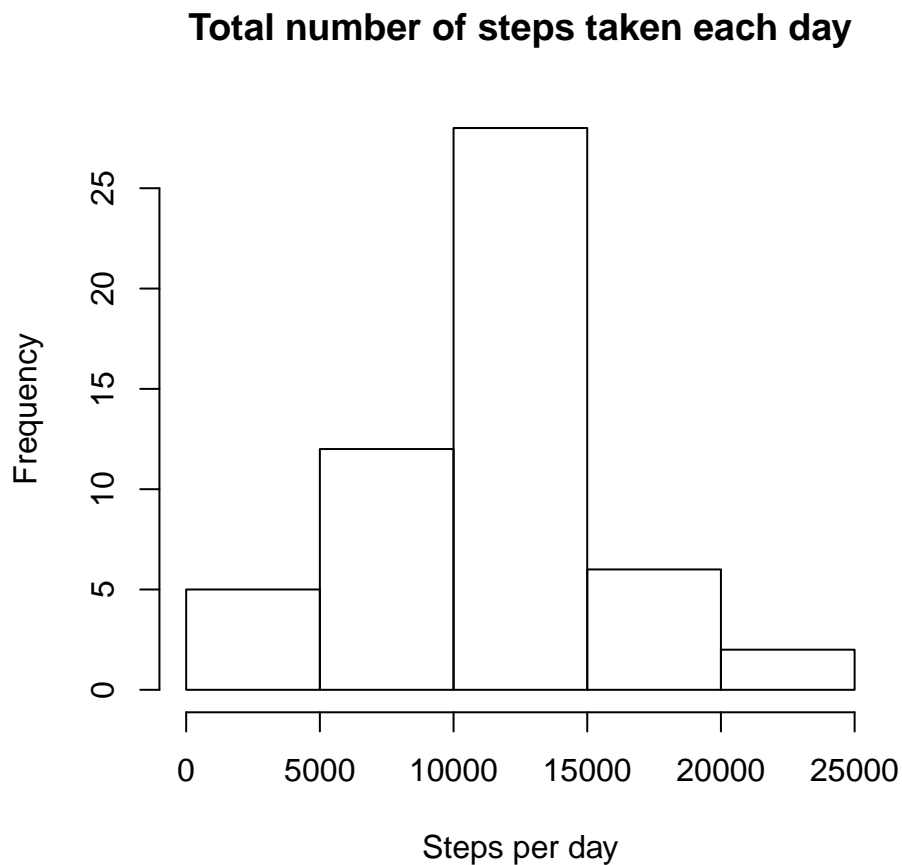
Peer Assessment No. 1 by Claudio Cer

Read data

```
file<-'D:\\users\\contre06\\Desktop\\C_reproducible_research-master\\peer_assessment_no1\\activity.csv'  
dataOrig<-read.csv(file=file,header=TRUE)  
data<-dataOrig
```

What is the mean total number of steps taken per day?

```
stepsDate<-aggregate(steps~date, data, FUN=sum)  
hist(stepsDate$steps,xlab='Steps per day',main='Total number of steps taken each day')
```



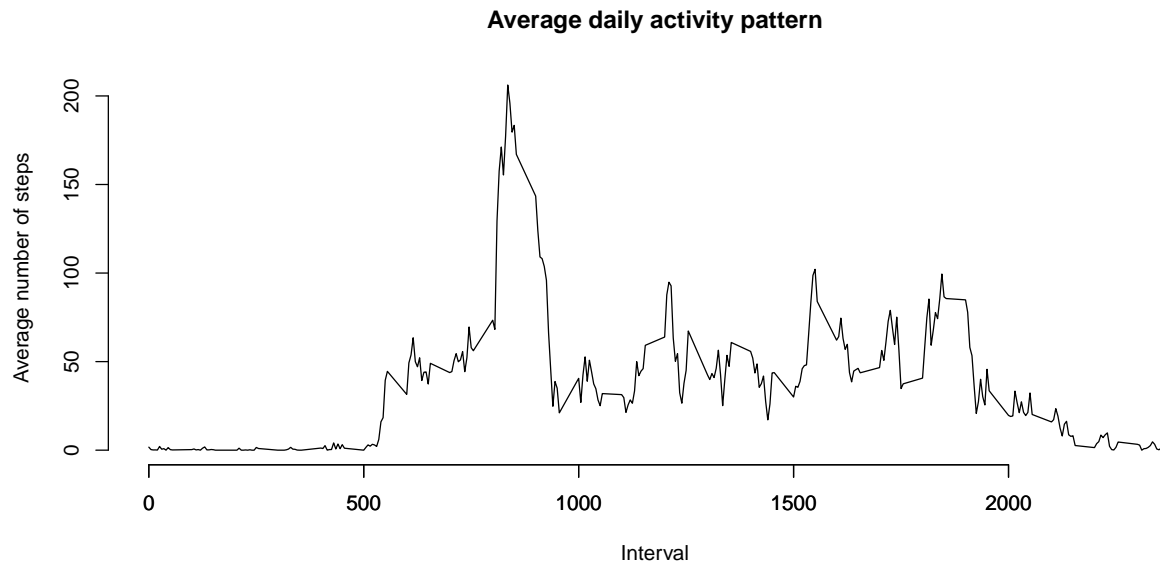
```
stepsMean<-as.integer(mean(stepsDate$steps))  
stepsMedian<-as.integer(median(stepsDate$steps))
```

Mean total number of steps taken per day: 10766

Median of the total number of steps taken per day: 10765

What is the average daily activity pattern?

```
stepsInterval<-aggregate(steps~interval,data, FUN=mean)
plot(stepsInterval$interval,stepsInterval$steps,type='l',main='Average daily activity pattern',xlab='Interval',ylab='Average number of steps')
axis(1)
```



```
intervalMax<-stepsInterval$interval[stepsInterval$steps==max(stepsInterval$steps)]
```

Interval 835 contains the maximum number of steps.

Imputing missing values

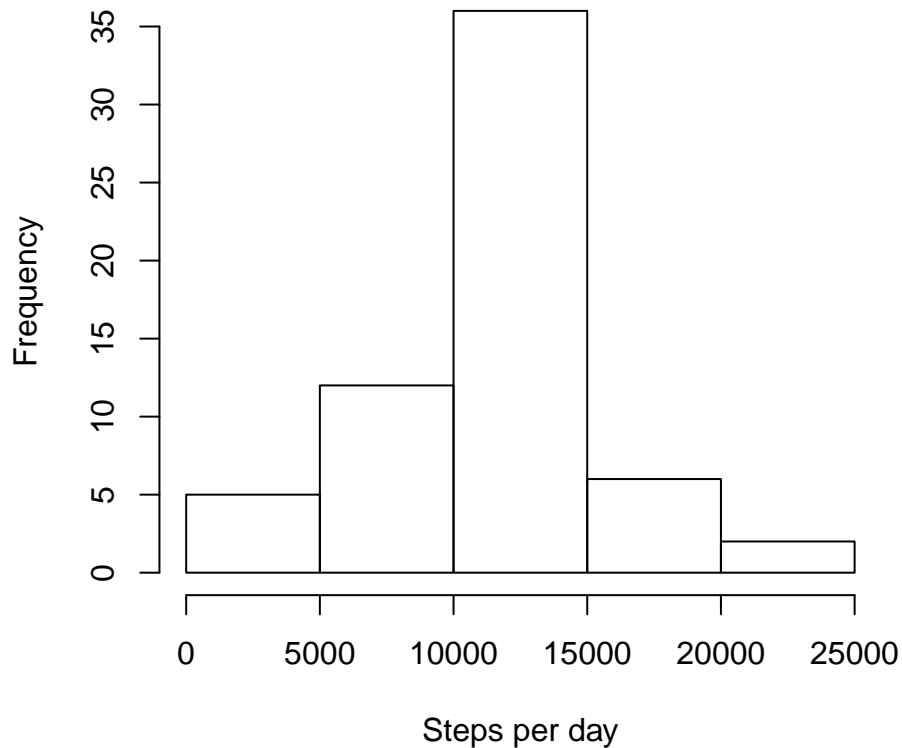
```
numberNA<-length(data[complete.cases(data)==FALSE,1])
```

The number of missing values is 2304.

```
dataFill<-dataOrig
completeCases<-complete.cases(dataFill)
for (i in 1:length(dataFill[,1])){
  if (completeCases[i]==FALSE){
    dataFill$steps[i]<-stepsInterval$steps[stepsInterval$interval==dataFill$interval[i]]
  }
}
```

```
stepsDateFill<-aggregate(steps~date, dataFill, FUN=sum)
hist(stepsDateFill$steps,xlab='Steps per day',main='Total number of steps taken each day')
```

Total number of steps taken each day



```
stepsMeanFill<-as.integer(mean(stepsDateFill$steps))
stepsMedianFill<-as.integer(median(stepsDateFill$steps))
```

Mean total number of steps taken per day: 10766

Median of the total number of steps taken per day: 10766

==> The mean and median total number of steps taken per day changed only slightly as the missing values were apparently evenly distributed filled with average steps per interval. Due to the filling with average values, the mean class of the histogram became larger.

Are there differences in activity patterns between weekdays and weekends?

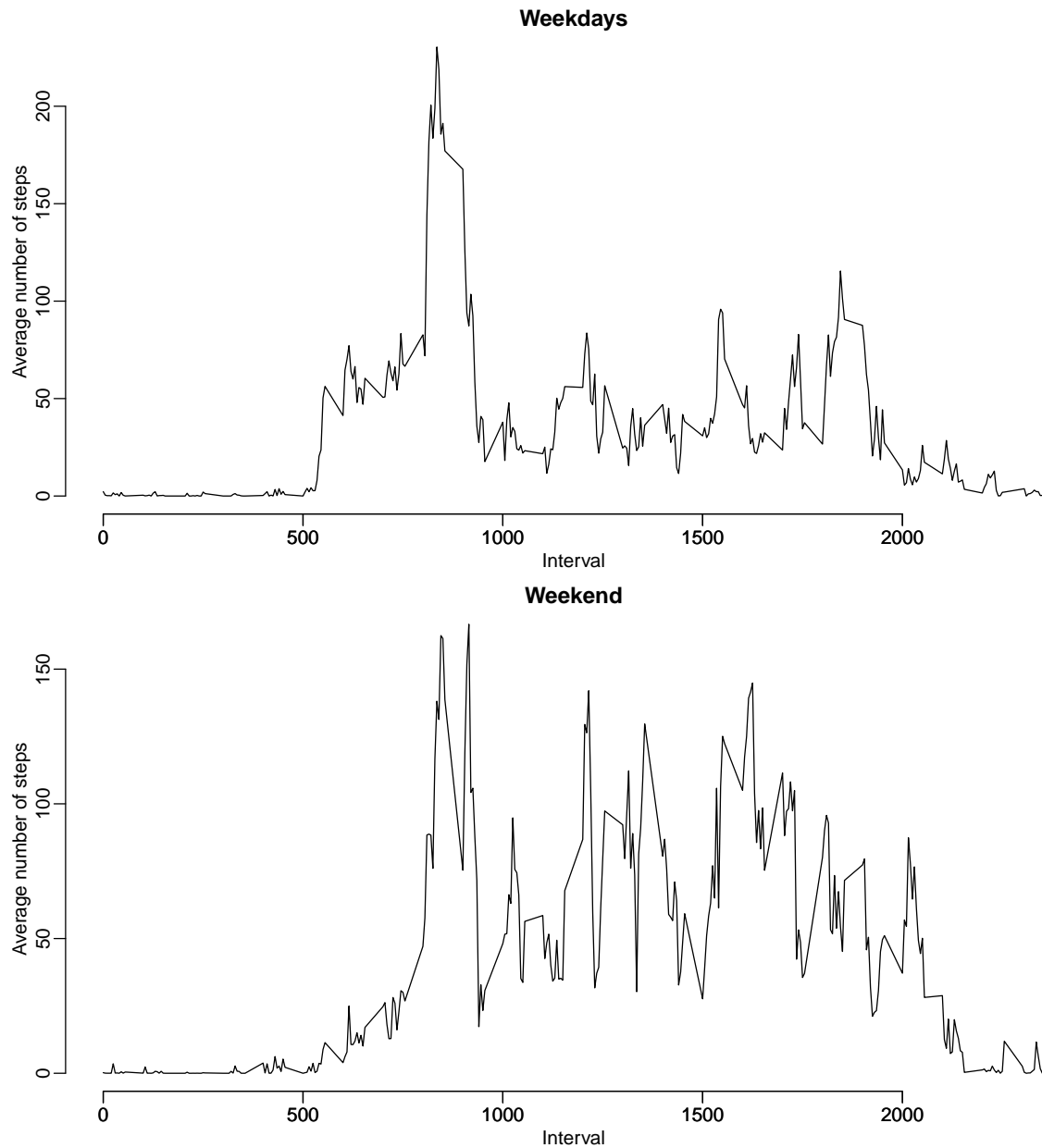
```
weekday<-weekdays(as.Date(dataFill$date))
wd<-c("lunedì", "martedì", "mercoledì", "giovedì", "venerdì")
we<-c("sabato", "domenica")
for (w in wd){weekday[weekday==w]<- 'weekday'}
for (w in we){weekday[weekday==w]<- 'weekend'}
dataFill$weekday<-factor(weekday)

stepsIntervalWD<-aggregate(steps~interval, dataFill[dataFill$weekday=='weekday',], FUN=mean)
```

```

stepsIntervalWE<-aggregate(steps~interval,dataFill[dataFill$weekday=='weekend'], FUN=mean)
par(mfrow=c(2,1),mar=c(3,3,1,3),mgp=c(1.5, 0.5, 0))
plot(stepsIntervalWD$interval,stepsIntervalWD$steps,type='l',main='Weekdays', xlab='Interval',ylab='Average number of steps',axis(1))
plot(stepsIntervalWE$interval,stepsIntervalWE$steps,type='l',main='Weekend', xlab='Interval',ylab='Average number of steps',axis(1))

```



==> Activity on the weekend is more evenly distributed throughout the day.