

Reproducible Research: Peer Assessment 1

Loading and preprocessing the data

Show any code that is needed to.

1. Load the data.

```
## [1] "/media/caraya/DATA/CURSOS/02-Datascience/Reproducible_Research/RepData_PeerAssessment1"

## [1] "activity.csv"      "activity.zip"      "doc"
## [4] "Hist1.png"         "instructions_fig"   "PA1_template_files"
## [7] "PA1_template.html" "PA1_template.md"    "PA1_template.Rmd"
## [10] "README.html"       "README.md"
```

```
da<-read.csv("activity.csv")
```

2. Process/transform the data (if necessary) into a format suitable for your analysis

```
da$date<-as.Date(da$date)
h<-trunc(da$interval/100)
m<-da$interval-h*100
time<-paste(h,m,sep=":")
datetime<-as.POSIXlt(paste(da$date,time,sep=" "))
```

What is mean total number of steps taken per day?

1. Calculate the total number of steps taken per day.

The mean total number of steps taken per day is:

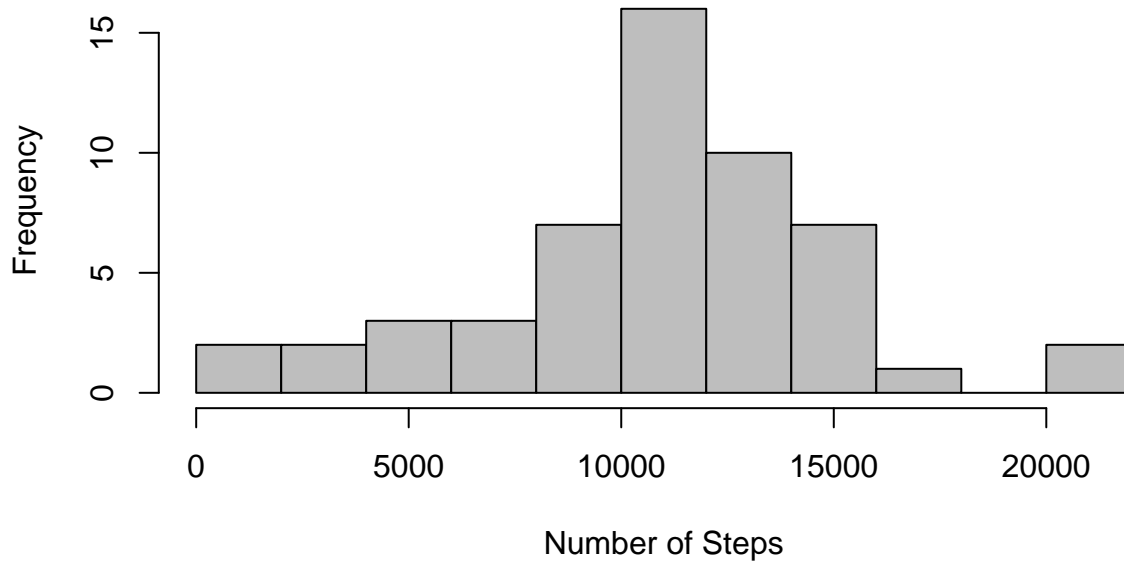
```
mean(with(da, tapply(steps, date, sum)),na.rm=T)
```

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

2. If you do not understand the difference between a histogram and a barplot, research the difference between them. Make a histogram of the total number of steps taken each day.

```
hist(with(da, tapply(steps, date, sum))
     ,main="Histogram: Total number of steps taken each day"
     ,xlab="Number of Steps"
     ,col="gray"
     ,breaks=8)
```

Histogram: Total number of steps taken each day



3. Calculate and report the mean and median of the total number of steps taken per day.

```
summary(with(da, tapply(steps, date, sum)))
```

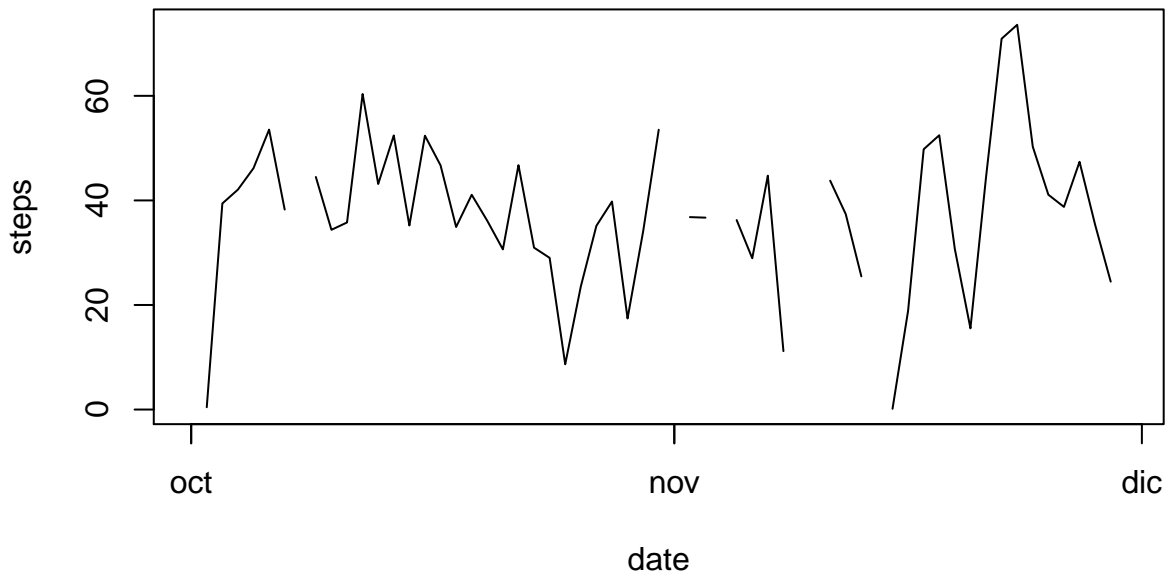
##	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.	NA's
##	41	8841	10760	10770	13290	21190	8

What is the average daily activity pattern?

1. Make a time series plot (i.e. type = "l") of the 5-minute interval (x-axis) and the average number of steps taken, averaged across all days (y-axis)

```
par(mfrow=c(1,1))
y<-with(da, tapply(steps, date, mean))
x<-with(da, tapply(date, date, max))
x<-as.POSIXlt(names(x))
plot(x,y,main="Daily Activity Pattern",type="l",xlab="date",ylab="steps")
```

Daily Activity Pattern



2. Which 5-minute interval, on average across all the days in the dataset, contains the maximum number of steps?

```
m<-max(da$steps,na.rm=T)
da[da$steps==m && !is.na(da$steps),]
```

```
## [1] steps    date      interval
## <0 rows> (or 0-length row.names)
```

Imputing missing values

Note that there are a number of days/intervals where there are missing values (coded as NA). The presence of missing days may introduce bias into some calculations or summaries of the data.

1. Calculate and report the total number of missing values in the dataset (i.e. the total number of rows with NAs)

The total number of missing values in the dataset is:

```
length(da[is.na(da$steps),1])+
length(da[is.na(da$date),1])+
length(da[is.na(da$interval),1])
```

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

2. Devise a strategy for filling in all of the missing values in the dataset. The strategy does not need to be sophisticated. For example, you could use the mean/median for that day, or the mean for that 5-minute interval, etc.
3. Create a new dataset that is equal to the original dataset but with the missing data filled in.

4. Make a histogram of the total number of steps taken each day and Calculate and report the mean and median total number of steps taken per day. Do these values differ from the estimates from the first part of the assignment? What is the impact of imputing missing data on the estimates of the total daily number of steps?

```
mean(with(da, tapply(steps, date, sum)),na.rm=F)
```

```
## [1] NA
```

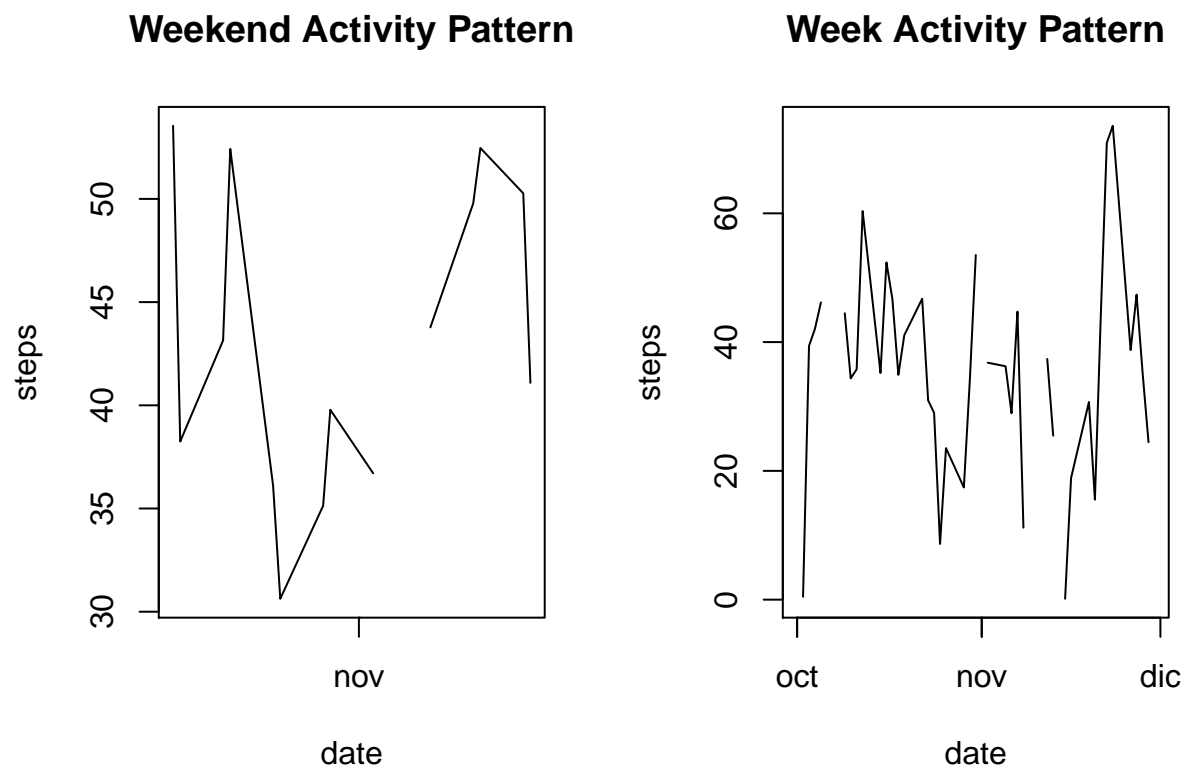
Are there differences in activity patterns between weekdays and weekends?

For that, we do: * Extract the week's days.

```
da<-cbind(da,weekdays(da$date))
colnames(da)<-c("steps","date","interval","weekday")
y1<-with(da, tapply(steps[da$weekday %in% c("sábado","domingo")], date[da$weekday %in% c("sábado","domingo")], date[da$weekday %in% c("sábado","domingo")], FUN=sum))
x1<-with(da, tapply(date[da$weekday %in% c("sábado","domingo")], date[da$weekday %in% c("sábado","domingo")], FUN=as.POSIXlt))
x1<-as.POSIXlt(names(x1))

y2<-with(da, tapply(steps[!da$weekday %in% c("sábado","domingo")], date[!da$weekday %in% c("sábado","domingo")], date[!da$weekday %in% c("sábado","domingo")], FUN=sum))
x2<-with(da, tapply(date[!da$weekday %in% c("sábado","domingo")], date[!da$weekday %in% c("sábado","domingo")], FUN=as.POSIXlt))
x2<-as.POSIXlt(names(x2))

par(mfrow=c(1,2))
plot(x1,y1,main="Weekend Activity Pattern",type="l",xlab="date",ylab="steps")
plot(x2,y2,main="Week Activity Pattern",type="l",xlab="date",ylab="steps")
```



There is no difference between Weekend and day of week Activity Pattern.