Some warming-up exercises on the use of R for massaging and visualization
SPR4104, autumn 2014

Data:
simple numbers about word occurrences in corpora per week
http://folk.uio.no/dssantos/cursoR/veraoPortugal.txt
http://folk.uio.no/dssantos/cursoR/veraoBrasil.txt

read and plot, compare the two

```
summerPT<-read.table("veraoPortugal.txt")
summerBR<-read.table("veraoBrasil.txt")
colnames(summerPT)<-c("week","PTsummer")
colnames(summerBR)<-c("week","BRsummer")
plot(summerPT)
```

to plot them together

```
plot(summerPT,type="n")
points(summerPT, pch=1)
points(summerBR, pch=2)
legend(1,240,c("Portugal","Brazil"),pch=c(1,2))
```

Try the opposite way (plot(summerBR,type="n"), and see why it would not be a good idea.

To get one data frame:
```
newsummer<-merge(summerPT, summerBR, by.x=c("week"),by.y=c("week"))
```

add a new term
http://folk.uio.no/dssantos/cursoR/NatalCHAVE.txt
Redraw. Remerge...

Tired of getting one column per file?
Read http://folk.uio.no/dssantos/cursoR/fourseasons.txt

```
seasons<-read.table("fourseasons.txt")
colnames(seasons)<-c("season","total","variety","varietysize","perc")
summary(seasons)
```

What went wrong? Try instead

```
seasons<-read.table("fourseasons.txt", dec=",")
colnames(seasons)<-c("season","total","variety","varietysize","perc")
```

Plot
```
barplot(seasons[seasons$variety=="BR",]$perc,
names=seasons$season[c(1,3,5,7)])
```

Create a matrix with the two varieties separate

```
matrix<-rbind(seasons[seasons$variety=="BR",]$perc,
seasons[seasons$variety=="PT",]$perc)
```

Try different barplot options:

```
barplot(matrix)
barplot(matrix,beside=TRUE)
barplot(matrix,beside=TRUE,names.arg=seasons$season[c(1,3,5,7)],legen
d=c("BR","PT"))
```

Maybe this one is better?

```
legend(1,70,c"BR","PT"))
```

Now try a totally different dataset

```
library(languageR)
names(dative)
table(dative$SemanticClass)
xtabs(~RealizationOfRecipient + SemanticClass, data=dative)
mosaicplot(xtabs(~RealizationOfRecipient + SemanticClass,
data=dative))
mosaicplot(xtabs(~RealizationOfRecipient + SemanticClass + Modality,
data=dative))
```

And still another one

```
emos<-read.table("http://folk.uio.no/dssantos/cursoR/emotions.txt",
header=TRUE)
names(emos)
```

Get a colum for love and one for hate, and compare the two.

```
emos$love<-emos$amor.min+emos$amor
emos$hate<-emos$odio.min+emos$odio
```

Better to compare the relative frequencies

```
emos$loverel<-emos$love/emos$Tamanho
emos$haterel<-emos$hate/emos$Tamanho
emos
boxplot(emos$loverel,emos$haterel)
plot(loverel,  xlab="",ylab="",type="n",las=2,xaxt="n")
axis(1,at=1:length(loverel),labels=emos$Corpus,las=2)
points(loverel~Corpus,pch=1)
points(haterel~Corpus,pch=2)
legend(1, .014, c("love", "hate"), pch=c(1,2), bty="n")
```