From CSV to Stratigraphic Plot

The `read.csv` command is used to read a .csv file. Unfortunately, the .csv files downloaded from Neotoma are not easy to use. I downloaded pollen data from Devil's Lake. Pollen counts start in line 10. Lines 7 - 9 contain entries that are not numeric. Therefore R assumes that the entire file contains factors and not numbers. This is a major problem when trying to make a stratigraphic plot. There are several solutions to this problem. I will present one below.

The `read.csv` command allows us to skip a certain number of lines when importing a data file into R. Pollen counts start in line 10. We will therefore skip the first 9 lines. We also need to tell R that there is no header in this data set (i.e. the columns do not have names)

```
devils_lake <- read.csv('dataset684.csv',skip=9,header = FALSE)
```

Now R understands that this file actually contains numbers. The next challenge is to rearrange the data. In our current data set the first row contains counts of *Abies*. To use `strat.plot` *Abies* percentages have to be in the first column of our data set. We therefore need to transpose the data. The first five columns do not contain count data. I therefore omit these columns before transposing. I do also assign column names and replace *NA* with *0*.

```
pollen_counts <- t(devils_lake[,-c(1:5)])
colnames(pollen_counts) <- devils_lake[,1]
pollen_counts[is.na(pollen_counts)] <- 0
counts_final <- pollen_counts
```

We want to plot pollen percentages. We therefore have to transform pollen counts into pollen percentages. The Devil's Lake stratigraphy contains 63 taxa. I therefore decided to exclude a few taxa. I only include taxa that cross the 5% threshold twice.

```
percentages <- 100*counts_final/rowSums(counts_final)
inclusion.criterion <- apply(percentages,2,function(x) (sum(x>5))>1)
percentages_clean <- percentages[,inclusion.criterion]
```

It is now possible to make a stratigraphic plot (I deliberately omit the code making the diagram from this file).

```
## This is rioja 0.9-15
```
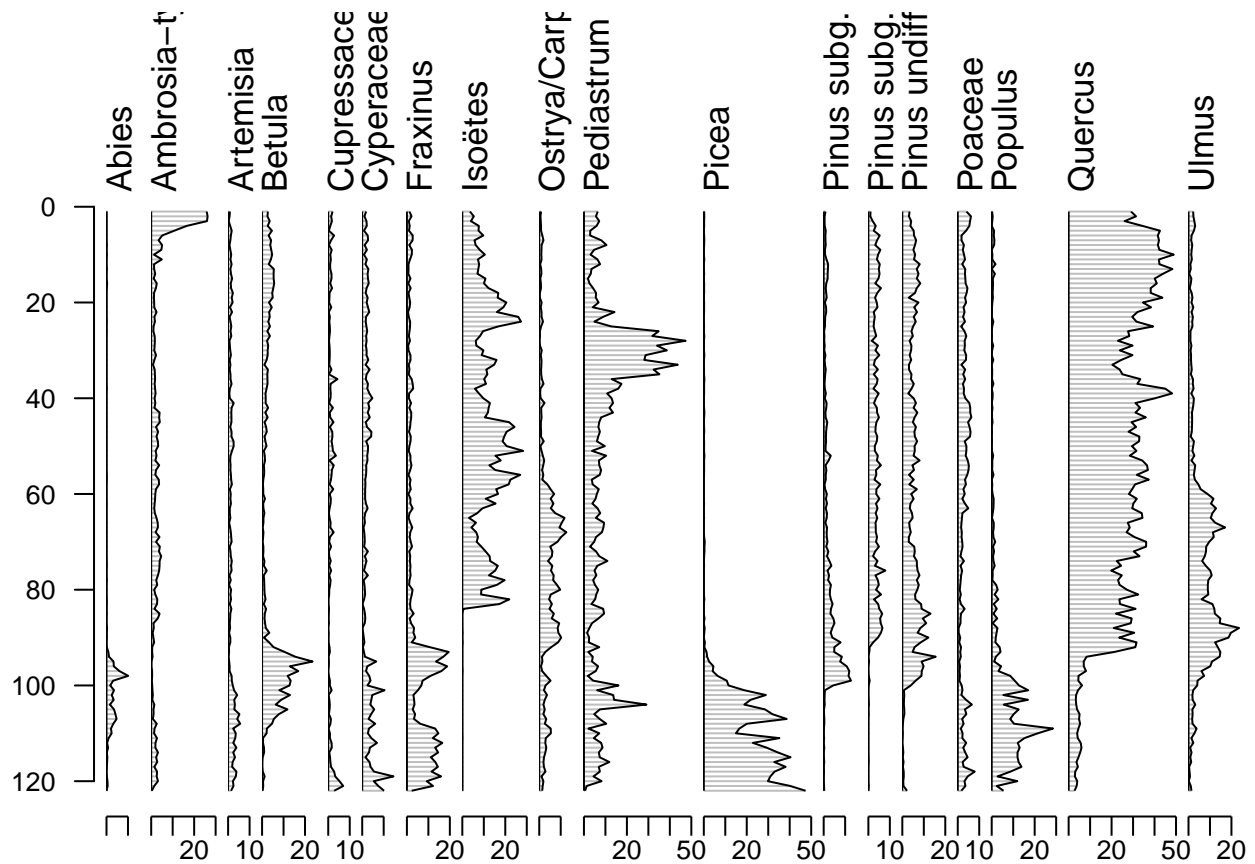
Unfortunately, there is no simple solution to get depths. I will again show one solution to obtain depths.

```
devils_lake_total <- read.csv('dataset684.csv')
depths.factor <- devils_lake_total[devils_lake_total$name=='Depth',6:ncol(devils_lake_total)]
depths.character <- as.character(unlist(depths.factor))
depths <- as.numeric(depths.character)
```

We can now make a stratigraphic plot with depths.
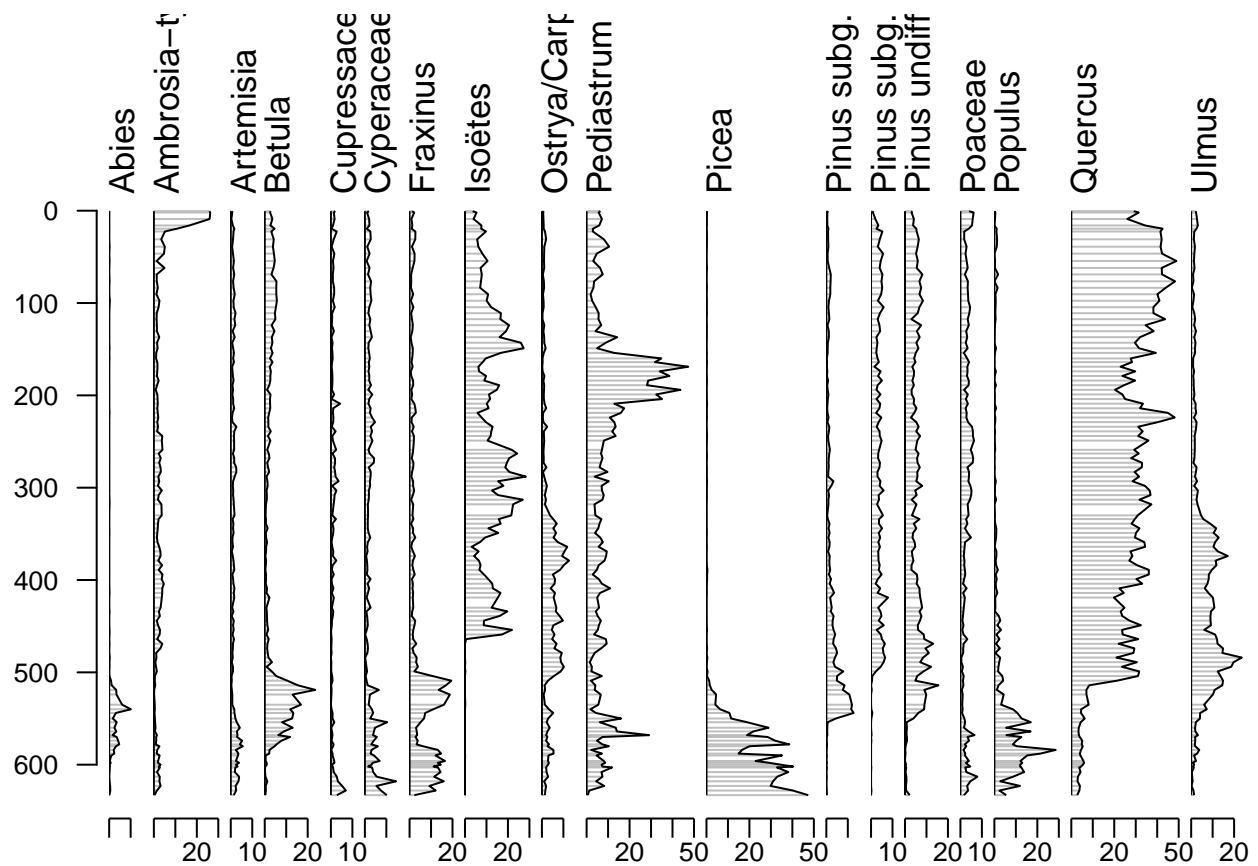
If you want to have fun making a stratigraphic plot, you can load data from Abernethy Forest, Scotland.

```r
library(rioja)
data(aber)
pollen.percentages <- aber$spec
depths.aber <- as.numeric(rownames(pollen.percentages))
```
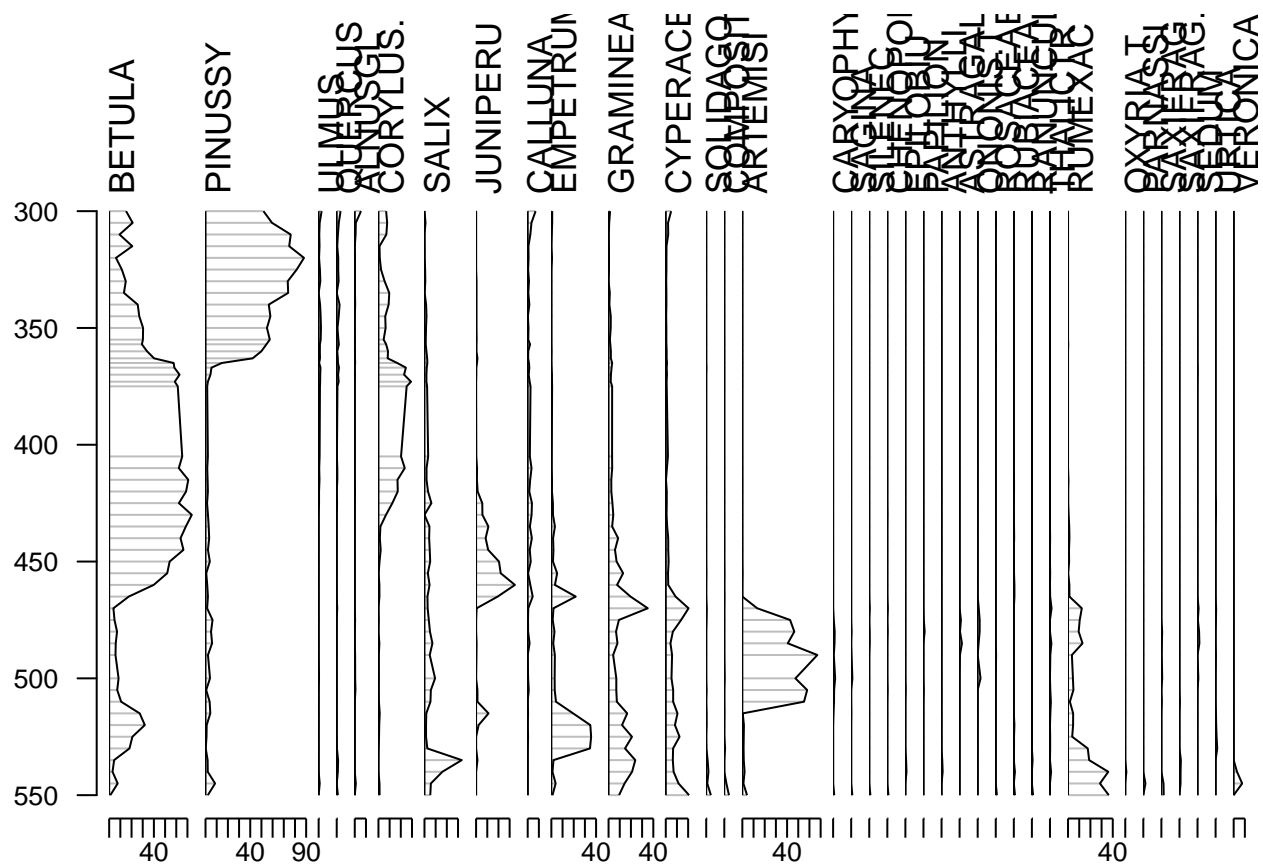
You could try to exclude unimportant taxa.