Qualitative Analysis in R

To analyse open ended responses using R there is the RQDA and Text Mining (TM) packages. This guide is not intended to be an exhaustive resource for conducting qualitative analyses in R, it is an introduction to these packages. There are more advanced functions that are covered in the full documentation available here: http://cran.r-project.org/web/packages/RQDA/RQDA.pdf

Data Cleaning

Before examining open ended responses, it's easier to clean up the field and create a new file with just those responses. Save the.csv file to import into the RQDA GUI.

```
      ## Import data file
      df <- read.csv ("C:/…/mydata.csv", stringsAsFactors=FALSE)

      ##Creates a list of variable names
            # OR1 – Open Response Question 1, category – demographics
      vars <- c("ResponseID", "category", "OR1")

      ##Create new data frame (OR1) with variables listed above
      df <- data.frame[vars]

      ## Clean up the text for analysis
      # Function to remove leading and trailing whitespace
      trim <- function (x) gsub("^\\s+|\\s+$", "", x)
      trim (df$OR1)

      # Replace carriage returns with space
      df$OR1 <- gsub("[\r\n]", " ", df$OR1)

      # Replaces commas  with space
      df$OR1<- gsub("[,]", " ", df$OR1)

      # Replaces dashes  with space
      df$OR1<- gsub("[-]", " ", df$OR1)

      # Convert all upper case to lower case
      df$OR1 <- tolower (df$OR1)

##Save the new data frame in the project folder
write.csv (OR1, file="OpenResponse1.csv")
```

Converting all of the text to lower case is important as R is case sensitive. For example, if you want to search and code for the term "bully" you would have to search for both "Bully" and "bully". This can become onerous when there are several different iterations of a word (i.e. bully, bullied, bullying).

Text Mining

Before conducting thematic analysis you can explore the open responses using the Text Mining package (http://cran.r-project.org/web/packages/tm/tm.pdf) and look for frequently occurring words, associations between words as well as other text mining functions.

```
library (tm)
library(reshape)

## Import data file
df <- read.csv ("C:/…/mydata.csv", stringsAsFactors=FALSE)
      # OR1 – Open Response Question 1

# Create a corpus for text mining
OR1.corpus <- Corpus(VectorSource (df$OR1))
#makes all lower case
OR1.corpus <- tm_map (OR1.corpus, tolower)
#Removes Punctuation
OR1.corpus <- tm_map (OR1s.corpus, removePunctuation)

# build a term-document matrix
OR1.dtm <- TermDocumentMatrix(OR1.corpus, control =
list(stopwords=TRUE,wordLengths = c(1,30)))

# Shows cases where "assess" appears.
# If "Subscript out of bounds" error, word does not appear
melt (inspect (OR1.dtm ["assess",1:500]))

# inspect most popular words
findFreqTerms(OR1.dtm, lowfreq=10) ##Terms that appear 10 times

# Counts for top words
freqwrds <- sort (rowSums (as.matrix(OR1.dtm)),decreasing=TRUE)

# Returns top 100 words
melt(freqwrds [1:100])

# Associations of word "climate" with other terms
findAssocs(OR1.dtm, ' climate, 0.20)
```

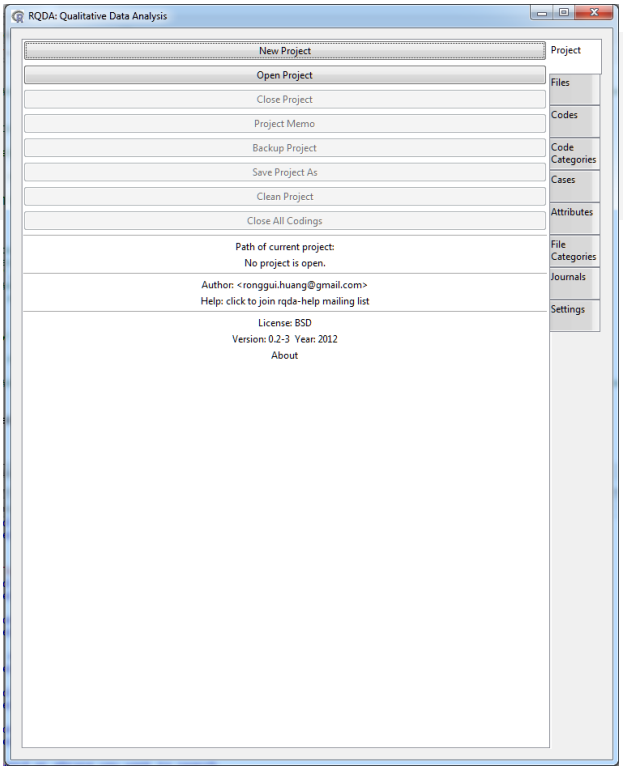
Coding the Data File

Once the open responses have been saved as a .csv file, load the RQDA package.

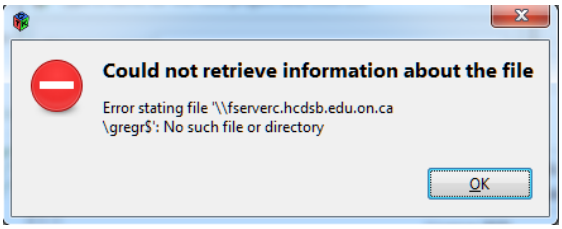
```
install.packages(RQDA)
library(RQDA)

# Launch the RQDA GUI
RQDA()
```

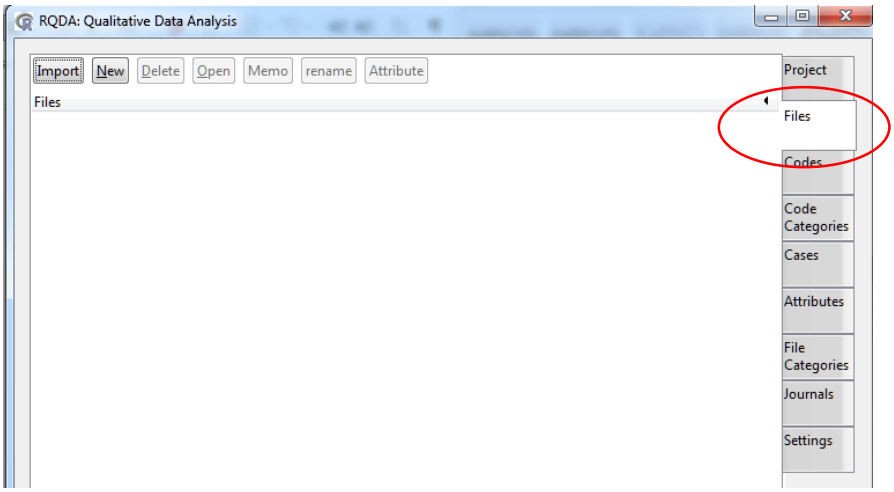
The Graphical User Interface opens in a new window



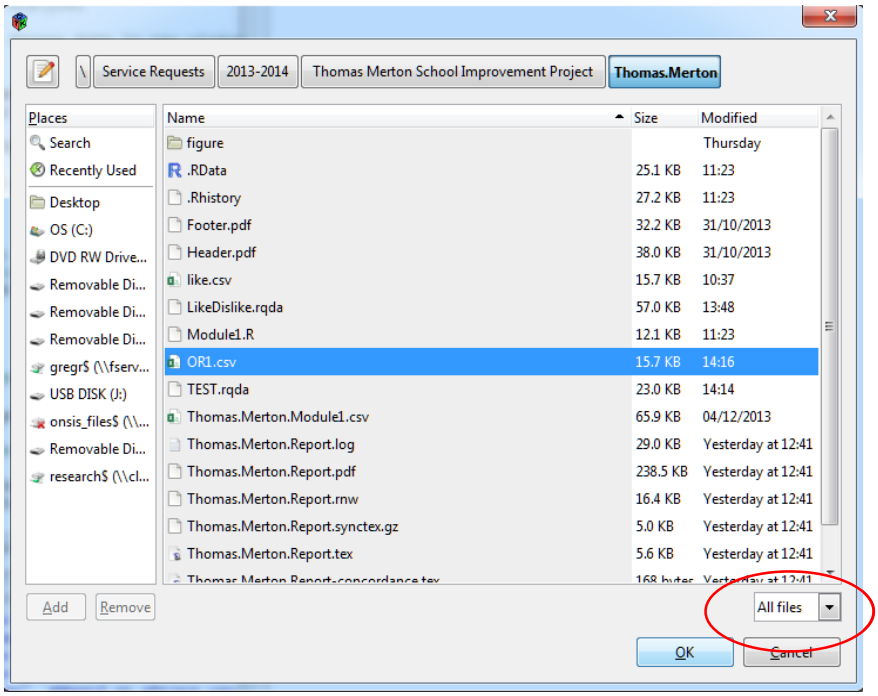
Click on “New Project”. You will likely get an error message. Just click “OK” and name your project.



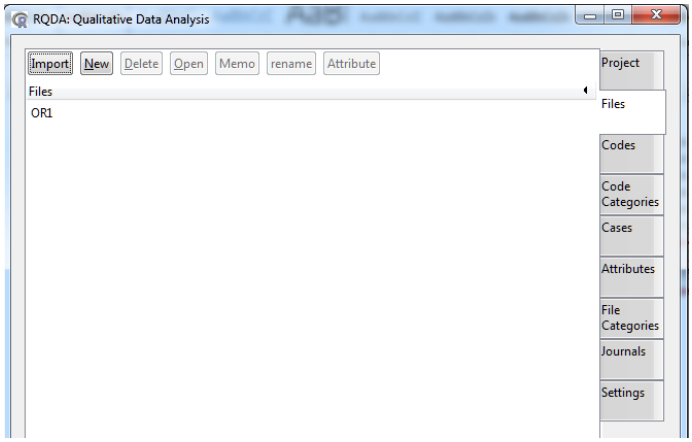
Click on the “File” tab (along the right hand side), and “Import”.

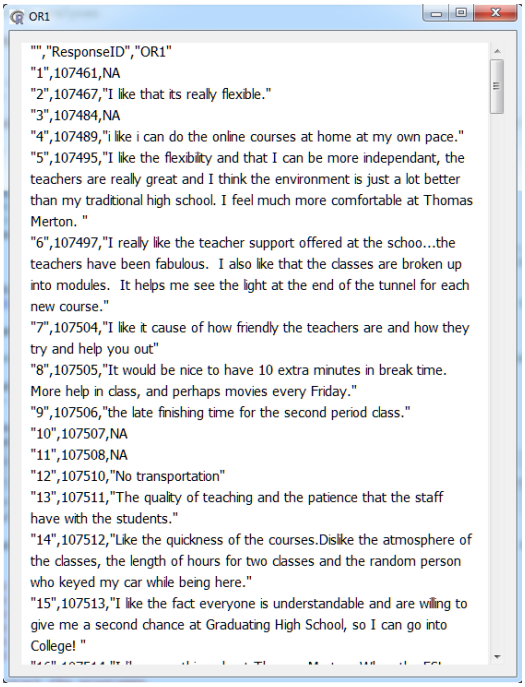


You may see the same error message, just click “OK” and then navigate to where the Open Response .csv file is saved. You may have to change the file type (bottom right) from text files to All files.



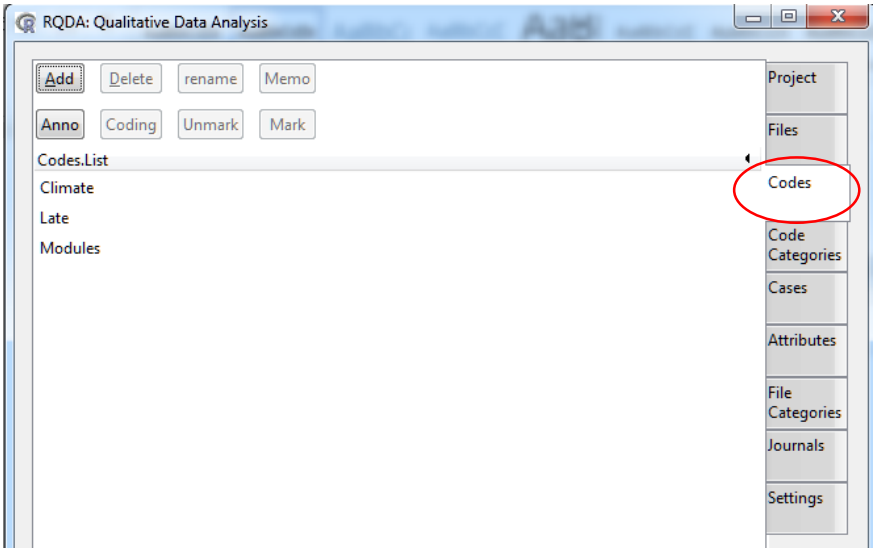
You should now see your file under the File tab. If you double-click on the file name a new window will open with the text.





You can now set up your codes under the Codes tabs. In this example, the question was “What do you like about your school?”.

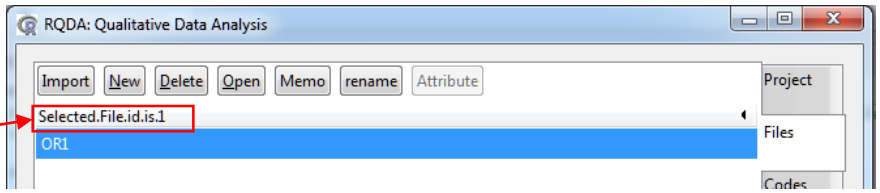
The codes are the themes that emerge from the text mining (i.e. Climate, Modules and Late Policy).



Auto-Coding Text

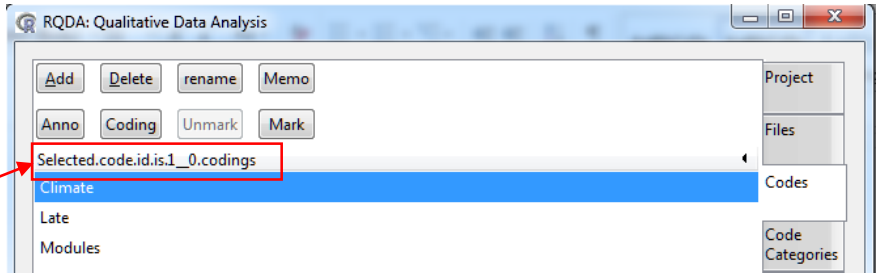
Coding can also be done using the R console using the `codingBySearch` function. In order to use this function, you have to identify the File ID and the Code ID in the GUI.

To get the FileID, click on the File tab in the GUI, and highlight the file. Above the file name will be the File ID number.

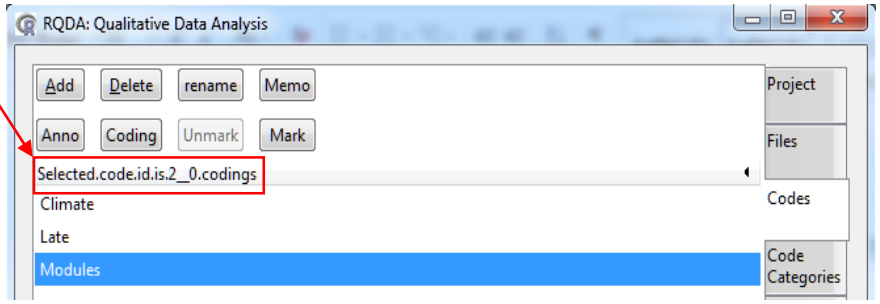


Follow the same procedure for the Code ID.

Here you can see that “Climate” has code ID 1 and “Modules” has code ID 2.



The _0 shows the number of times the code has been applied.



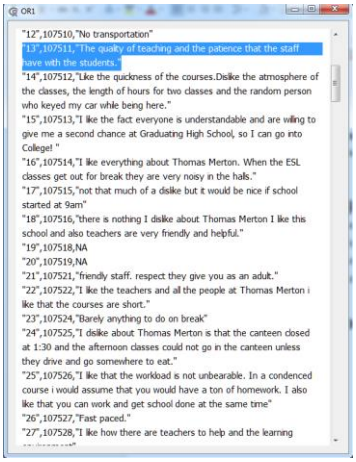
With the RQDA windows open, return to the R console and set up the `codingBySearch` script.

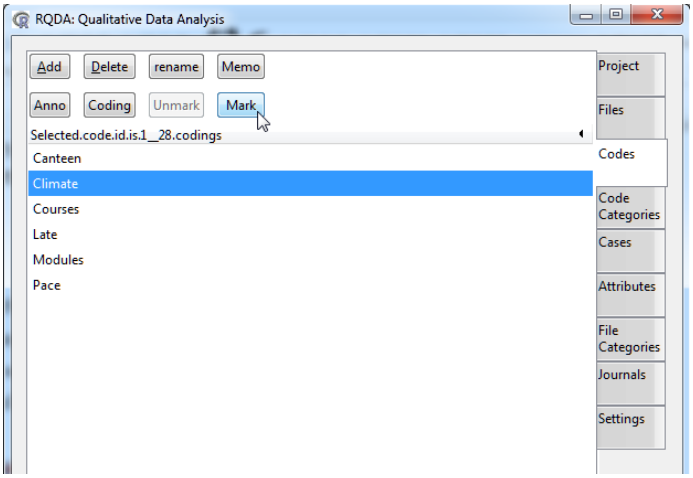
```
codingBySearch("climate", ##word or phrase you want to search
               fid="1",##FileID, from GUI
               cid="1") ##CodeID, from GUI
```

This code passes over the file and every entry that contains the word “climate” will be coded as such. Some categories may have multiple entries.

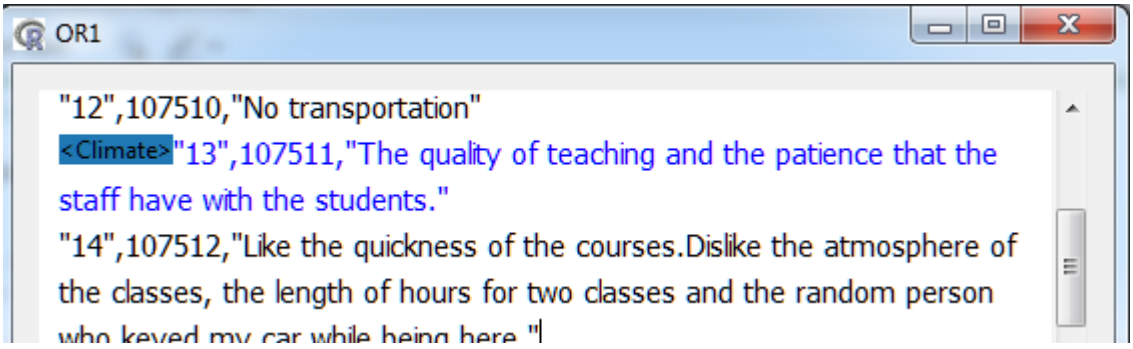
Manually Coding Text

Automatically coding text works great for themes that have already been established, however manual coding of the file is still a necessity. Some themes may not be evident from the text mining, or there may be misspellings. Have both the RQDA GUI and the window with the text open. Highlight the entry in the text box that you wish to code, ensure the proper code is highlighted in the GUI and click “Mark”,





The text will appear highlighted in the text window with the code label beside it.

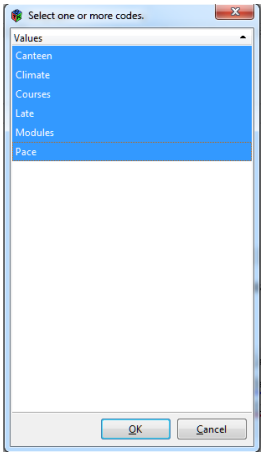


Exporting the Coded File

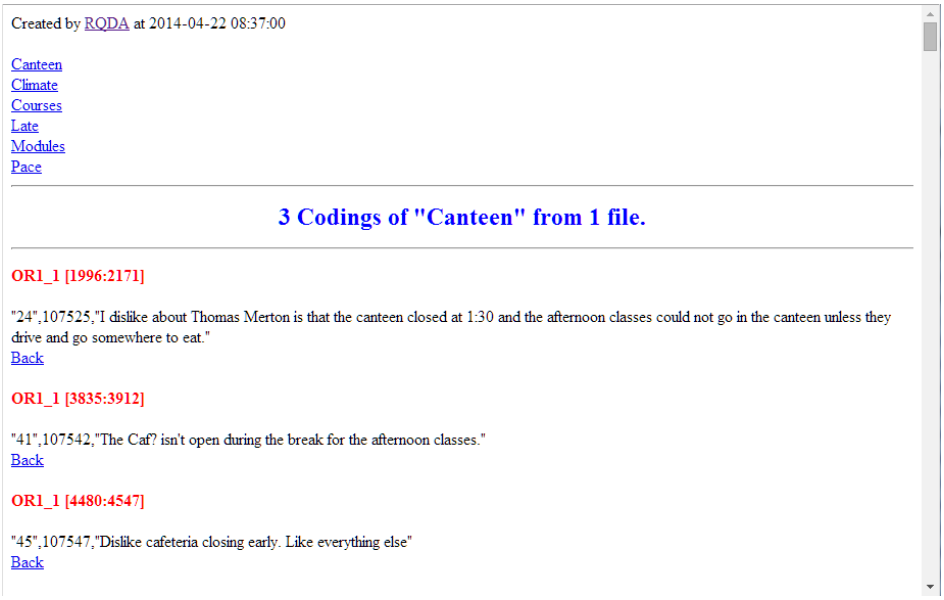
You can export the coded file as an HTML file that will show all the entries under a particular category. Going back to the R console:

```
exportCodings(file = "OpenResponse.html", Fid = 1,
              order = ("fname"), append = FALSE,
              codingTable="coding")
```

A new window will open with all of your codes listed. Highlight the codes that you are interested in (or all of them) by holding the Control key and clicking on each code.

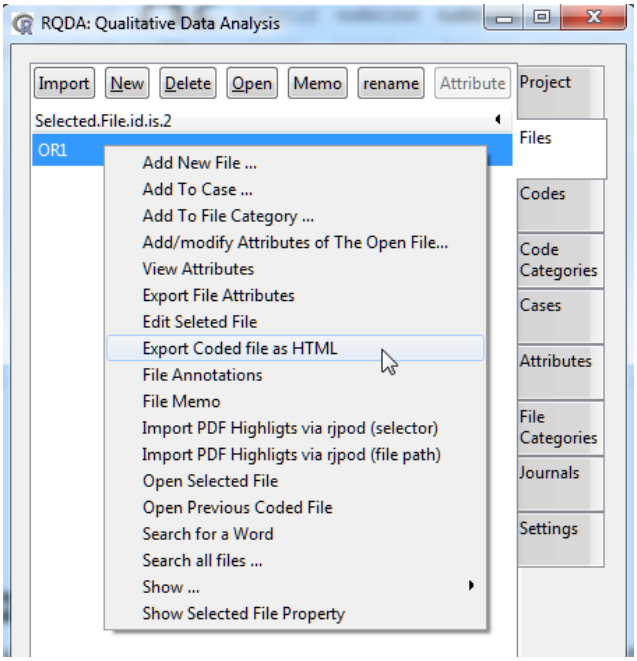


Open the HTML file and you will see each code listed at the top of the page which are hyperlinked to sections that contain all the entries associated with that code.



Exporting the Coded File – Part 2

The RQDA package does not handle summaries or sub-groups very well. A work around is to export the coded file as HTML from the GUI and copy/pasting it into Excel.
Right click on the file and select “Export Coded file as HTML”.



The resulting file looks like this:

Created by RQDA at 2014-04-22 08:46:21

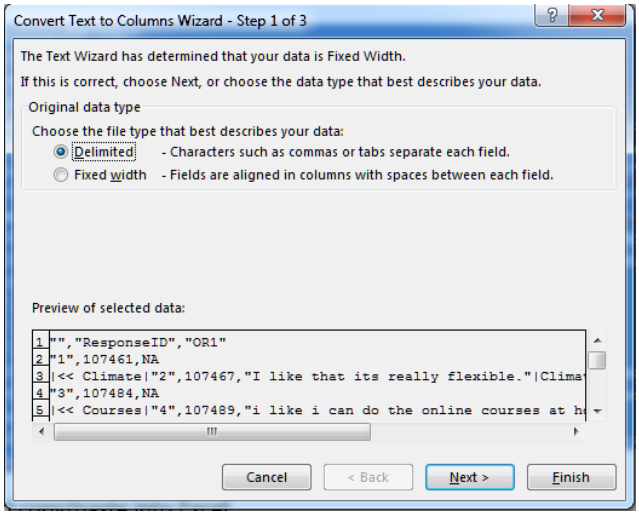
","ResponseID","OR1"
"1",107461,NA
|<< **Climate**|"2",107467,"I like that its really flexible."|*Climate >>*|
"3",107484,NA
<< **Courses**	"4",107489,"i like i can do the online courses at home at my own pace."	*Courses >>*
<< **Climate**	<< **Pace**	"5",107495,"I like the flexibility and that I can be more independant, the teachers are really great and I think the environment is just a lot better than my traditional high school. I feel much more comfortable at Thomas Merton. "	*Climate >>*	*Pace >>*
<< **Climate**	<< **Modules**	"6",107497,"I really like the teacher support offered at the schoo...the teachers have been fabulous. I also like that the classes are broken up into modules. It helps me see the light at the end of the tunnel for each new course."	*Climate >>*	*Modules >>*
<< **Climate**	"7",107504,"I like it cause of how friendly the teachers are and how they try and help you out"	*Climate >>*
"8",107505,"It would be nice to have 10 extra minutes in break time. More help in class, and perhaps movies every Friday."
"9",107506,"the late finishing time for the second period class."
"10",107507,NA
"11",107508,NA
"12",107510,"No transportation"
<< **Climate**	"13",107511,"The quality of teaching and the patience that the staff have with the students."	*Climate >>*
<< **Courses**	"14",107512,"Like the quickness of the courses.Dislike the atmosphere of the classes, the length of hours for two classes and the random person who keyed my car while being here."	*Courses >>*
<< **Climate**	"15",107513,"I like the fact everyone is understandable and are willing to give me a second chance at Graduating High School, so I can go into College! "	*Climate >>*

To clean it up, highlight the page (starting at “”, “ResponseID”, “OR1”) and copy/paste into Excel.

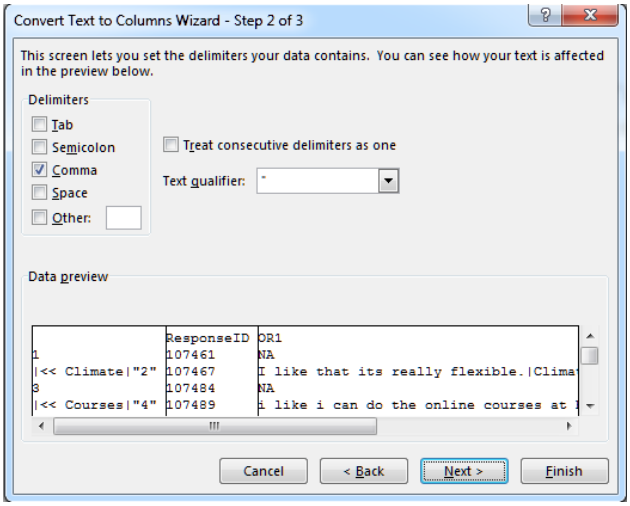
Created by RQDA at 2014-04-22 08:46:21

```
"","ResponseID","OR1"
"1",107461,NA
|<< Climate|"2",107467,"I like that its really flexible."|Climate >>|
"3",107484,NA
|<< Courses|"4",107489,"i like i can do the online courses at home at my own pace."|Courses >>|
|<< Climate|<< Pace|"5",107495,"I like the flexibility and that I can be more independant, the teachers are really great and I think the environment is just a lot better than my traditional high school. I feel much more comfortable at Thomas Merton. "|Climate >>|Pace >>|
|<< Climate|<< Modules|"6",107497,"I really like the teacher support offered at the schoo...the teachers have been fabulous. I also like that the classes are broken up into modules. It helps me see the light at the end of the tunnel for each new course."|Climate >>|Modules >>|
|<< Climate|"7",107504,"I like it cause of how friendly the teachers are and how they try and help you out"|Climate >>|
"8",107505,"It would be nice to have 10 extra minutes in break time. More help in class, and perhaps movies every Friday."
"9",107506,"the late finishing time for the second period class."
"10",107507,NA
"11",107508,NA
"12",107510,"No transportation"
|<< Climate|"13",107511,"The quality of teaching and the patience that the staff have with the students."|Climate >>|
|<< Courses|"14",107512,"Like the quickness of the courses.Dislike the atmosphere of the classes, the length of hours for two classes and the random person who keyed my car while being here."|Courses >>|
|<< Climate|"15",107513,"I like the fact everyone is understandable and are willing to give me a second chance at Graduating High School, so I can go into College!
"|Climate >>|
```

Once pasted, select “Text to Columns” and ensure “Delimited” is selected. (This is where removing punctuation in the file becomes important) and click Next.



Ensure that “Comma” is selected and click “Finish:.



You will now have each entry with the codings listed in the first cell. You can insert columns beside Column A and use the Text to Columns function again with "|" as the delimiter instead of a comma. This is also handy when there are subgroups in the file, such as grade or gender.

	A	B	C	D
1		ResponseID	OR1	
2	1	107461	NA	
3	|<< Climate|"2"	107467	I like that its really flexible.|Climate	
4	3	107484	NA	
5	|<< Courses|"4"	107489	i like i can do the online courses at h	
6	|<< Climate|<< Pace|"5"	107495	I like the flexibility and that I can be	
7	|<< Climate|<< Modules|"6"	107497	I really like the teacher support offer	
8	|<< Climate|"7"	107504	I like it cause of how friendly the tea	
9	8	107505	It would be nice to have 10 extra min	
10	9	107506	the late finishing time for the second	
11	10	107507	NA	
12	11	107508	NA	
13	12	107510	No transportation	
14	|<< Climate|"13"	107511	The quality of teaching and the patie	
15	|<< Courses|"14"	107512	Like the quickness of the courses.Dis	
16	|<< Climate|"15"	107513	I like the fact everyone is understanc	

Once this file has been cleaned up, you can import back into R to create summaries of the data.