

# Digital Methods: Learning Journal Template

Mette Maria Bonnema

Autumn 2019

**0.1 Thoughts / Intentions**

**0.2 Regex tasks**

**0.3 Open Refine**

# 1 29/10/2019

## 1.1 Thoughts / Intentions

**12:00 am:** I finished class at 11, cried for an hour in frustration and is now ready to understand more of the different templates we have been introduced to. **12:05 am:** Decided to work with my classmate Victor, so that the process of understanding the template goes faster. I must admit, I do not understand a lot about computers.

**12:10am:** Getting an overview of the different templates. My goal this coming week is to 1. get an overview of the different templates 2. solving the assignment for next week.

**13:00pm:** Beginning to work with Regex.

## 2 Regex exercise, Voyant -stoplist to R

### Task

Conversion of the Stopwordlist from docs into a long list of words without quotation signs.

- Opened data sets from google docs.
- Copied words into regex101.com
- Tried to mark all of the quotation signs by writing (") in the regular expression column.
- It didn't work
- I tried multiple things and found that if you wrote ", " in the regular expression, you would find all the quotation- signs and all the commas, except the first and the last one.
- I now need to substitute my regular expression with a newline.
- Since I don't know how to do this, I googled it, and found that newline expression is \n.
- By writing this in the substitution field, I completed my task. Although I couldn't "grab" the first and the last quotation- sign.
- I later talked with my friend who solved it in a different manner. In this way, it's possible to grab the first quotation-sign. First I inserted my text in Regex. Then I made my words into groups by writing (.+). I didn't get the newline error in this manner, so I also typed \n.

- I wanted to make quotation-signs around all words. I grabbed the word group in the substitution field by typing "\$" in this manner i managed to get the same result as my first try, AND grab the first and the last quotation sign.  
**Task completed**

## 2.1 Second Task: conversion of Regex stoplist to Voyant

- **Task:** To convert R -stoplist into a stopwordslist for Voyant.
- Marked the words from the Docs document and copied them into Regex
- my strategy is to exterminate all the signs, that I dont want.
- step 1. I type: " in my regular expression. In leave my substitution field blank. In that way I exterminate all the quotation-marks.
- step 2. I copy the text beneath the substitution field. This text is without "-marks. The new text can be inserted into text string. My goal is now to remove all the commas and make newlines.
- step 3. typing a comma and space in my regular expression like this: (, ) i find both commas and the following space, which I am not interested in keeping. In the substitution field, I can now write \n. The list is now perfect for voyant use.

## 2.2 Conclusion for todays work

14.30

I have solved todays tasks. Although it was difficult to understand Regex at first, it was very satisfying to finally solve the tasks and see the beautiful results.

### 3 OpenRefine tuesday d. 5.11

Tuesday the fifth of November: Today we're working with OpenRefine. I'm a bit nervous about, since I don't feel it was easy to understand last weeks class-instructions. My personal goal is therefore to listen very carefully and hopefully understand, what Adela is explaining.

- **Thoughts**

We are kindly being introduced to OpenRefine by Adela and Petra. It seems like the tempo for todays lesson is perfect.

- Creation of the Openrefine project: In OpenRefine I have opened an html-file, that Adela has given us. Afterwards I press "Create Project", and OpenRefine shows itself.
- The intention with this class, seems to be a better understanding of OpenRefine and its opportunities.
- Question to Adela: what are the advantages with OpenRefine compared to Excel? Adela tells, that the clever thing about OpenRefine is the data sorting. It allows you to sort your data in a clever way, and you can always find the original dataset, since OpenRefine save and store your changes. This allows you to regret former actions, which Excel doesn't in the same way. This seems very smart, indeed.

#### 3.1

##### Hands on OpenRefine

- Beginning to work with OpenRefine. The dataset we have been given is archeological. It looks very nicely, but there are a few mistakes, that we have to clean up.
- first task is to clean up the villages. There are several functions in OpenRefine which all seems very scary.
- Selecting "Facet" in the village box. Selecting "text facet" whereafter a box appears to the left in the picture. I now have a overview of the different villages that have been noted in the datasat.
- Merging defaults in the datasat, by pressing "Cluster" in the facet. I hereby merge the villages which have been misspelled or similar mistakes.
- Afterwards: we try several ways of reshaping the data, so it is more useable. There are a lot of steps included, and I didn't have the time to write everything down. This annoys me, since I cannot remember, what I did.

- Discovery of the Undo/Redo button in OpenRefine. This allows you to see every step in your work with the dataset. This is a really clever way to save and possibly undo your changes.
- apparently when you try to apply a commando in the middle of the list of commandos, you delete everything. I did this and can't manage to regret the command. All my work is now deleted.

## 4 Monday 11.11

Monday the 11th: lecture with Voytech and Adela

- Today we had a lecture which I didn't understand much of. I am generally very confused about what we're meant to do with the paper and the final product.
- After today's lecture I sat down with my study group to find solutions in OpenRefine and Regex. It turns out that they're just as confused as I am. I tried to explain to one of them how OpenRefine worked, but ended up deleting all my commandos. Since I didn't remember how I sorted the data in the first place, I wasn't able to reconstruct my former file. I had a minor mental breakdown and took contact to Max Odsbjerg for help.

### 4.1 Meeting with Max and working with R

- I went to Max since I was stressed and confused. I told him that I would like to have help with three things
  1. Finding a dataset.
  2. Getting an overview of the different programs of which to use.
  3. Getting some tools to work with this dataset and guidance in which programs I should use
- Max was very nice and told me not to worry. I worked with him for 1 hour and 15 minutes in which I finally found my dataset.
- Since Max is most comfortable with Rstudio, he guided me through that.
- Apparently, the commands in Regex101 is the same as in Rstudio. Regex is incorporated in Rstudio. I felt like, I had Regex under my control, so it made more sense to work with Rstudio than OpenRefine, which I found frustrating.
- Max taught me how to install usepackages for Rstudio, which would be useful, when finding a dataset. See picture below.

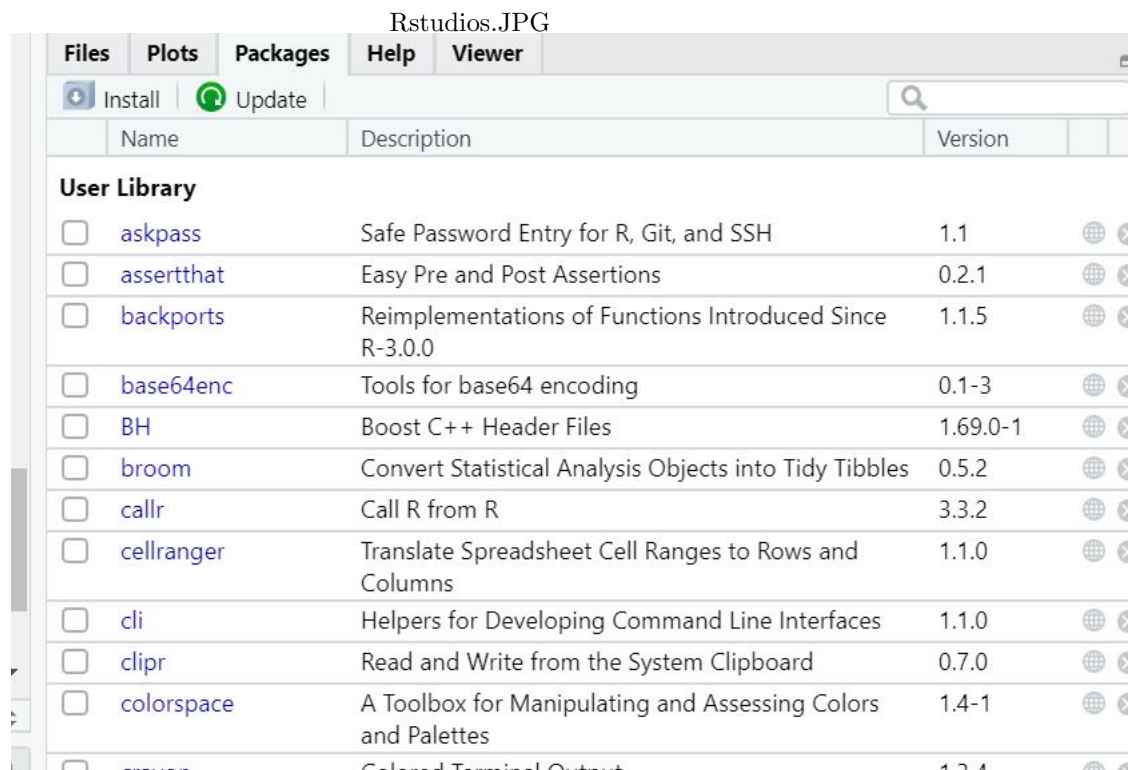


Figure 1: Caption

- In the picture, you see the icon "Packages". I chose this, and selected "install" whereafter I installed "rtweet" "tidyverse" and "tidytext".
- In the left side of the program, I can see my "storyline" of the programming. This gives me a great overview of the different commands, that I have run. See the picture below.

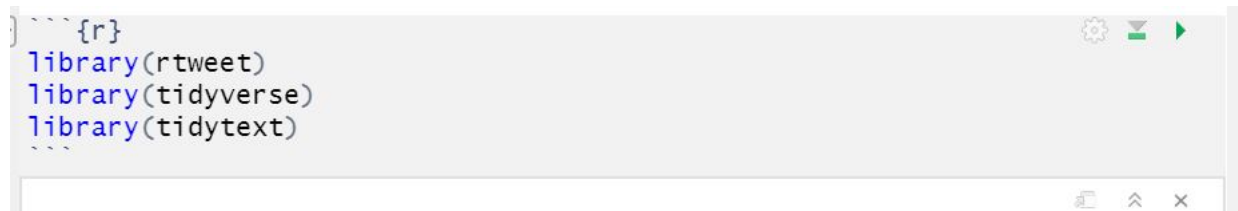


Figure 2: Usepackages R

## 4.2 Working with R

- Beginning a new R document. It has to be "R markdown".
- Struggling with where to begin. Looking through the things I did with Max and the links, he sent me. The easiest way to find answers to your questions is to google it. Everything pops up here.
- Everytime I make a command, I have to press "run" on the green button.
- I've followed an example, that Max gave me the links to. [http://hax.odsbjerg.dk/twitter\\_scrape.html4\\_visualisation](http://hax.odsbjerg.dk/twitter_scrape.html4_visualisation)
- After followings Max's example, I have a better understanding of how Rstuidos work. All the grey boxes in the program is the commands for R. If there are any minor mistakes in this, it is impossible to go through with your command.
- I have now made a graph of the most used words related to the hashtags ()Trump and ()Obama.