Blog for 10-275 Sweave Article

We used the R script files we worked on in Wednesdays 10/25 class to write a paper. The subject of the paper is using tidytext & dplyr with janeauster to make a wordcloud. In addition we learned about the gutenbergr package which contains many books.

First you need to install packages in R Studio (install all the packages first) then library (some installations require a “re-start.” If you already have these installed all you need to do is library:

First, we'll need to install \& library the following packages:

tidytext

tm (don't need to library) -tm stands for text mining

wordcloud

stringr

dplyr

janeaustenr

gutenbergr

Knitr

First, we did some preliminary exploration on the gutenbergr package:

We opened an R script file.

In order to find a book in the gutenbergr package, we need to know the ID number.

We ran the code: gutenberg\_works(title=='Dracula') in our rscript and got the following in the console:

> gutenberg\_works(title=='Dracula')

# A tibble: 1 x 8

gutenberg\_id title author gutenberg\_author\_id language gutenberg\_bookshelf

<int> <chr> <chr> <int> <chr> <chr>

1 345 Dracula Stoker, Bram 190 en Gothic Fiction/Movie Books/Horror/Mystery Fiction

# ... with 2 more variables: rights <chr>, has\_text <lgl>

>

So the ID for Dracula is 345.

Next we tried to find Frankenstein. But the title of Frankenstein isn’t Frankenstein.

The following did not work:

gutenberg\_works(title==str\_detect(title'Frankenstein'))

gutenberg\_works(author==str\_extract(author,'Shelley'))

gutenberg\_authors(author=='Shelley,Mary')

After class note: what did work:

gutenberg\_works(str\_detect(author, "Shelley"))

gutenberg\_works(str\_detect(title, 'Frankenstein'))

84 Frankenstein; Or, The Modern Prometheus Shelley, Mary Wollstonecraft

Back to Sweave:

First under Tools -> Global Options choose Sweve on the left & at the top change Sweave to Kntr

Now open a new Sweave file.

It opens up showing:

\documentclass{article}

\begin{document}

\end{document}

These control the Sweave document. If the Begin or end document aren’t paired, it will not compile into a pdf. MAKE SURE that the \end document is always the last thing. Put several returns before it to make sure you have “room” to work & keep this at the bottom of the page.

After the begin document type in the following:

\title{sense and sensibility wordcoud}

\author{Madam President Cake AKA Heidi L Beezub}

\maketitle

Then we need to add

\begin{abstract}

\end{abstract}

You need to thave the corresponding “end” Type the summary intro for our paper between the beginning & end.

Take note. Sweave uses special characters. For example if you type an ampersand & you have to use a backslash \ in front of it, or Sweave wants to read it as a function.

The backslash \ is the escape character (double backslash can be used to create a new line).\\

A backslash needs to go before the underscore \\_ or ampersand \& as these have special meaning in R for formatting text.\\

The > I gather is for a new paragraph but you can’t use the escape backslash to escape out of it.

Use the tic mark in upper left key (under the tilde) for the first quote and the last quote is the regular quote mark. If using double quotes you need 2 tics. Note that for a contraction such as can't you use the regular quote mark.

As we put in each section of the Sweave, be sure to hit the ‘Compile Pdf’ button. That way you will know your document is compiling.

If you get an error message it will have a line number (that doesn’t correspond to your Sweave document line numbers). Go to the ‘view log’ button (just above the error messages). Scroll through until you see a (what looks like a number 1 but is really an l) 1. Then some numbers so for example:

l.61 In our October 25 &

27 classes we used tidytext to construct a wordclou...

I can't figure out why you would want to use a tab mark

here. If you just want an ampersand, the remedy is

simple: Just type `I\&' now. But if some right brace

up above has ended a previous alignment prematurely,

you're probably due for more error messages, and you

might try typing `S' now just to see what is salvageable.

It gives you the “text” where you made the mistake you just need to find it & fix it.

Now we are ready to add a section. We type

\section{Section Title}

With the section title in between the curley brackets {}

Sweave keeps track of the section numbers & if we add a section in the middle it automatically re-numbers.

We can type text (keeping in mind those special characters)

And insert r code using the green ‘+c’ icon at the top right of the working sweave file.

When we compile the pdf, it shows not only the code, but the code output.

We added text & code based on our ‘cleaned” data from the Jane Austen wordcloud.

Monday 10/30 we will work on building a bibliography.

We will have a Halloween project of doing an Sweave on a Halloween themed novel.

Fin

Bibliography

For our bibliography we have to do several things.

At the top of our document right under the document class we need to add

Backslashusepackage\{natbib\}

Directly under that we need to add the following code:

Left carrot left carrotecho=FALSEright carrot right carrot=

Packages left carrot dash c('dplyr','stringr','tidytext','wordcloud','tm','janeaustenr')

knitr::write\_bib(packages,file='packages.bib')

@

Then just before the end document put in:

Backslash bibliographystyle\{apa\}

Backslash bibliography\{Wordcloud\_WutheringHeights,packages\}

Backslash nocite\{\\*\}