GEOG 594: R & R-Studio Exercise

Submitted by: Harshal Sanap
RedID: 822044776
Dated: 9/20/2018

1. Attach the two Big Data Definition Word Cloud Images.

Image1: WordCloud with words Big and Data

Image 2: WordCloud without Words Big and Data

2. Go to the Live Well Data Portal (https://data.livewellsd.org/) Pick up another data types (other cancer data or other injuries data). Import the data into R and conduct some basic statistical analysis and draw some new visualization graphics (different from the previous example). Attached your R-script for creating the analysis and visual graphics in the report.

- Answer to this question is attached in a separate HTML file.

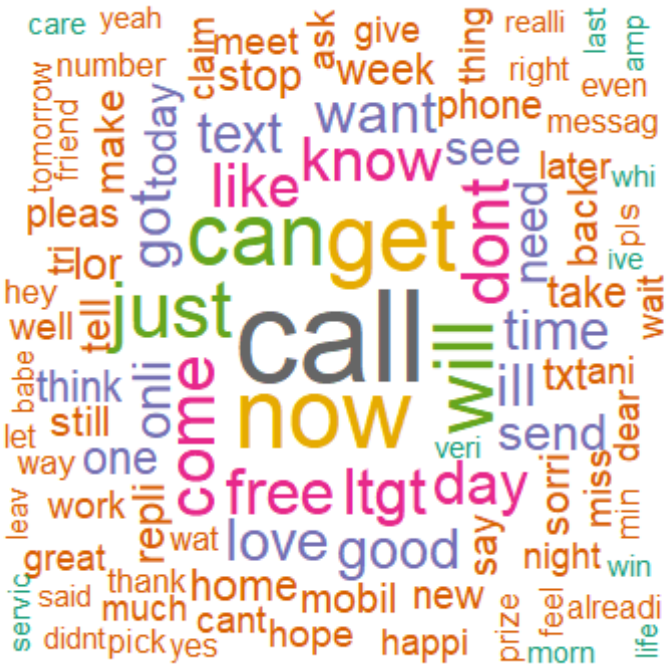
GEOG 594: R & R-Studio Exercise

Submitted by: Harshal Sanap
RedID: 822044776
Dated: 9/20/2018

3. Select a Webpage or a group of text files, create a word cloud map. In the report, indicate the text sources and the output WordCloud image.

Text Source: Twitter user comments for a product for the purpose of sentiment analysis. Downloaded from Kaggle.

Image3: WordCloud with twitter dataseet



4. In the Word Cloud exercise, we need to manually copy each student's definition in Blackboard into a single text file. Can you think about any better data collection methods or procedure? Which tool or software can be used for this task?

- According to my experience we can perform web scraping / crawling to access the required information from Blackboard using the below two approaches:
    i. Using Python: We can use pythons BeautifulSoup package to scrape the required data from blackboard. We would first need to create or establish a session with blackboard by giving our username and password credentials. Once a session is established then we can redirect the program to go to the discussions board and write a crawler to request individual student's definition. The request can be made only by giving specific element found using inspect element as the input parameter in the program or by simply reading in the whole page. The output of the request would generally be a JSON file from which we would read the elements that we need, in this case that would be the definitions element posted by individual students.

GEOG 594: R & R-Studio Exercise

Submitted by: Harshal Sanap
RedID: 822044776
Dated: 9/20/2018

(The technical name of this definition field can be found using inspect element on chrome browser)

ii. Using Data Miner: I had used this incredible tool for scraping information off website. We do not need to write any code like in the above approach and this is a tool which can be installed for free as part of Chrome or any other browser extension. On installing this extension, you go to the website that you want to scrape and define a recipe (formula on what fields you want to read from the website) stating how you would like to crawl and scrape the given website. Once you define the recipe, you can run it and the output can be downloaded in csv, excel or copy to clipboard format.