

Lab 1

Descriptive Network Analysis – Collecting and Visualizing Data

Agenda

- Assignment context
- Collecting network data
- Visualizing network data
- Q&A

Collective Intelligence Is About To Disrupt Your Strategy: Are You Ready?

Ganes Kesari Contributor @

Data science advisor, fractional CDAO, TEDx speaker, corporate trainer

Follow

Apr 28, 2021, 06:50am EDT

VS.

Harvard
Business
Review

Subscribe

Sign In

AI And Machine Learning

ChatGPT and How AI Disrupts Industries

by Ajay Agrawal, Joshua Gans, and Avi Goldfarb

December 12, 2022



is the knowledge that arises when **people join forces, collaborate, and work together** to make decisions. It is shared or group knowledge.

vs.



is the **simulation of human intelligence by machines or softwares**. Basically, machines learn how to learn, recognise speech patterns, process information etc. in the same way as the human brain.

Part I: Collecting Network Data



Quora

Stack**Exchange**

 **stack overflow**

vs.



ChatGPT

Step 1: Choose a topic for your text

You can decide on the topic based on personal interests, research interests, or popular topical areas, among others.

Example: Medici Family in Florence during the Renaissance



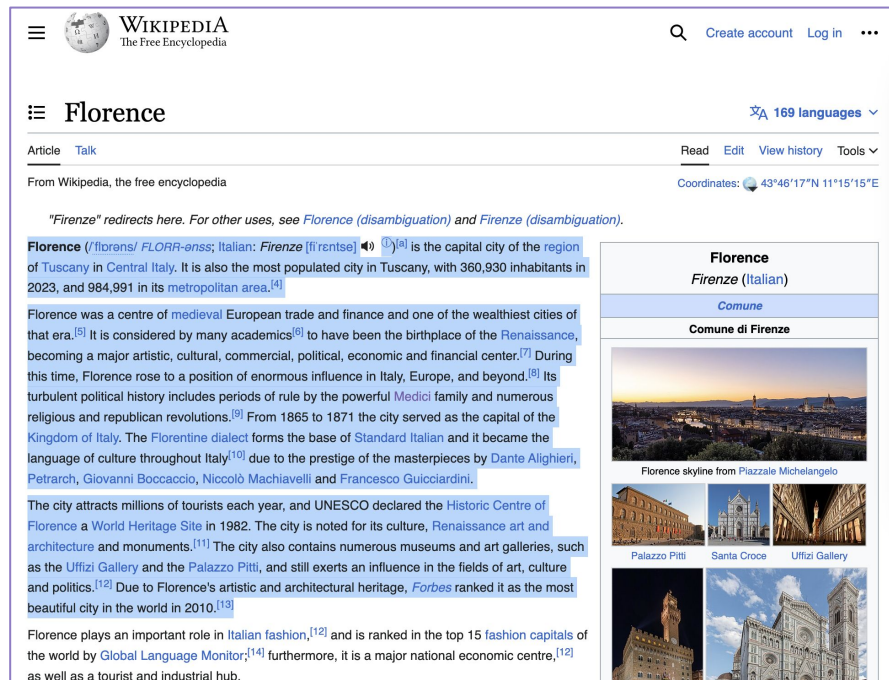
Step 2: Collect text from a collective intelligence source

For the collective intelligence sources (e.g., Wikipedia), you want to find a single webpage that contains an overview of your topic.

Example collective intelligence source:

Wikipedia → Florence (sections)

- Rise of the Medici
- Savonarola, Machiavelli, and the Medici popes



The screenshot shows the Wikipedia page for Florence. The article text is highlighted in blue, showing the city's history, its role as a center of Renaissance art and architecture, and its status as a World Heritage Site. The sidebar on the right contains a table with the title 'Florence' and 'Firenze (Italian)', followed by 'Comune' and 'Comune di Firenze'. Below this are several images: a wide view of the Florence skyline from Piazzale Michelangelo, and smaller images of the Palazzo Pitti, Santa Croce, and the Uffizi Gallery.

Florence
Firenze (Italian)

Comune

Comune di Firenze

Florence skyline from Piazzale Michelangelo

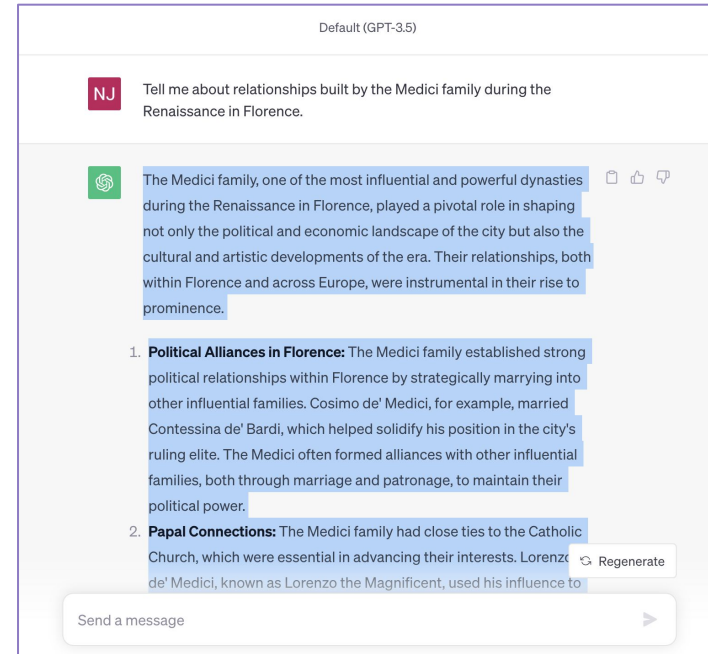
Palazzo Pitti Santa Croce Uffizi Gallery

Step 3: Collect text from ChatGPT using similar prompts

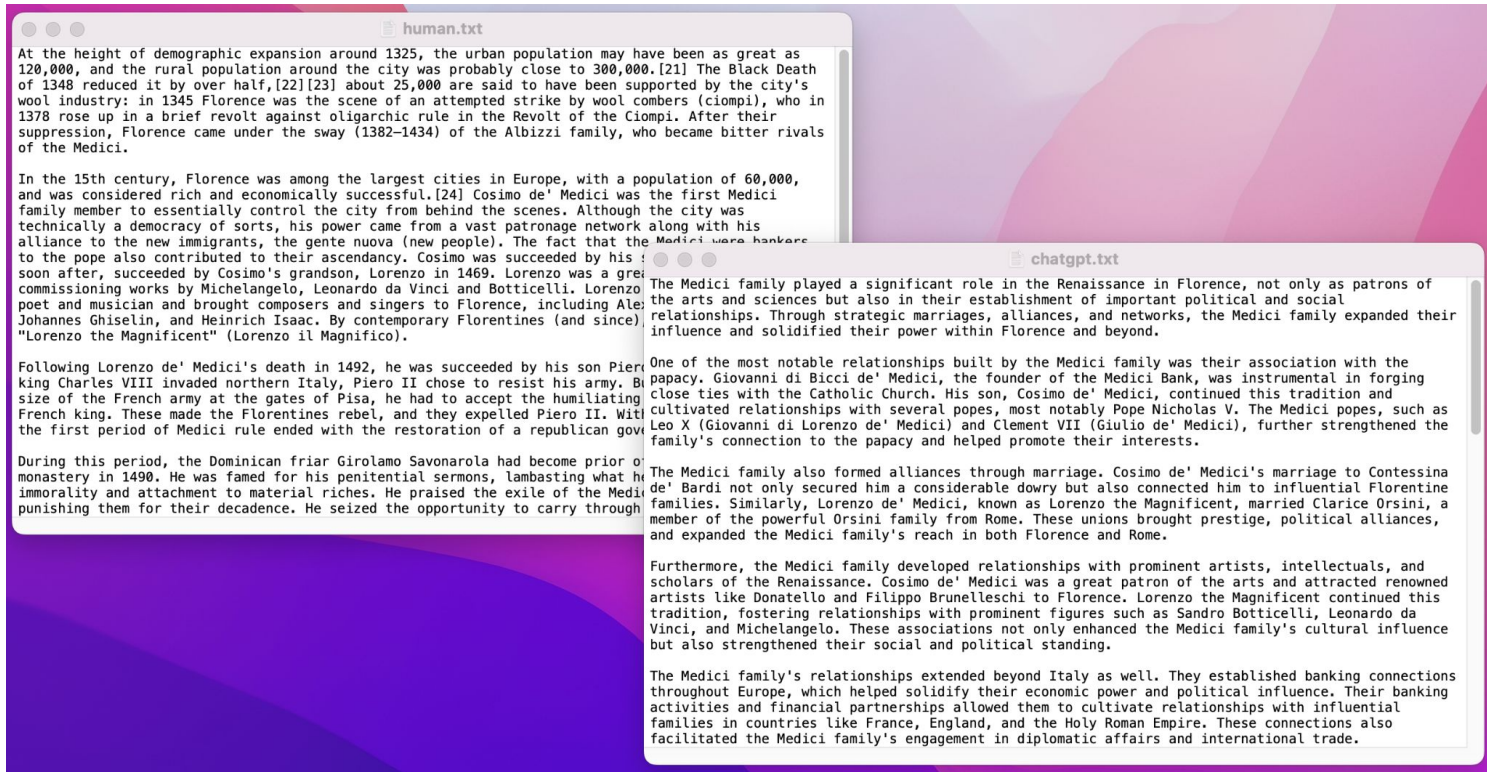
For ChatGPT, you should query it to collect similar information.

Example ChatGPT prompts:

- Tell me about relationships built by the Medici family during the Renaissance in Florence.
- Tell me more about the links between the Medici and artists.



Step 4: Save your text in .txt or .rtf files



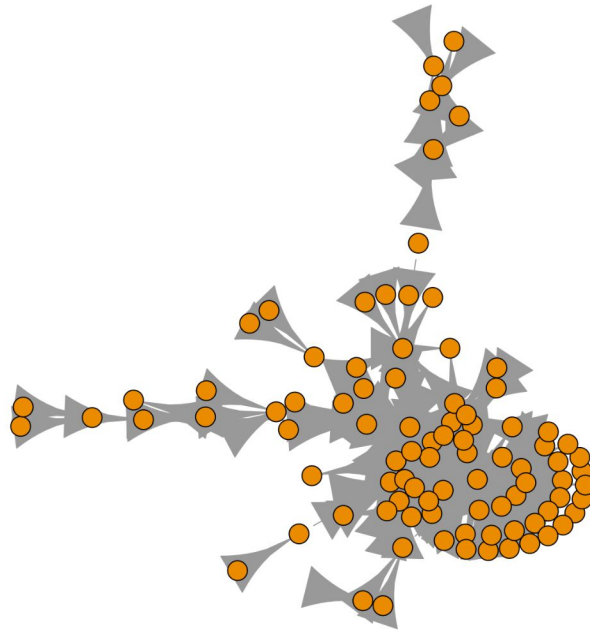
Step 5: Generate Semantic Networks

Some Details to Note:

- The raw text is initially “cleaned” by:
 - Removing **stop words** (words that are frequent but provide little information – e.g., “I”, “the”, “we’ll”, “it’s”, etc.)
 - **Lemmatization** (grouping together different inflected forms of the same word – e.g., the lemma of “ran”, “runs”, and “running” is simply “run”)
- The network ties represent words that co-occur in the text **within 10 words** of each other, excluding **rare word co-occurrences** (those that occur less than 3 times in the full body of text)
- **LDA** is used to group the words into 3 topics

Part II: Visualizing Network Data

Network Visualization



Network Visualization

Please check out this page: <https://kateto.net/network-visualization>



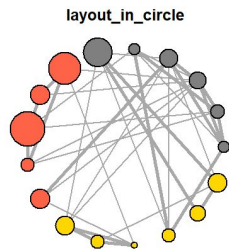
Static and dynamic network visualization with R

June 14th, 2017 | Tags: animation, centrality, D3, dataset, graph, igraph, JavaScript, map, ndtv, network, network analysis, network science, network tutorial, network visualization, networkD3, node, plot, R, reciprocity, RStudio, sna, Statnet, threejs, transitivity, visNetwork, visualization

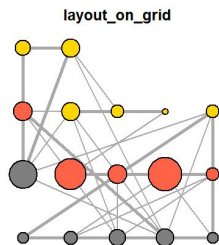
Comparing Layout Algorithms

Different stochastic algorithms attempt to arrange nodes and links in a “useful” way.

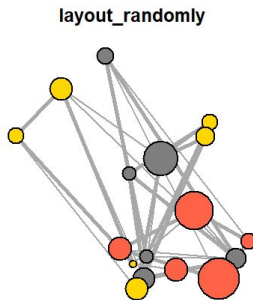
Results vary based on your network ... and how many times you run the algorithm.



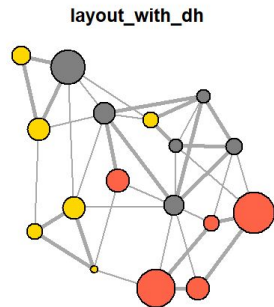
Nodes in a circle
(evenly spaced)



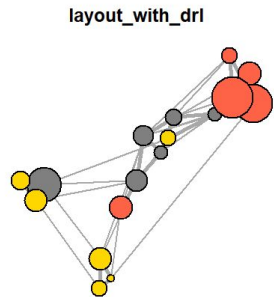
Nodes spaced out
on a uniform grid



Nodes placed
completely at
random



Davidson-Harel
Algorithm: Stochastic
optimization of edges



Distributive Recursive
Layout Algorithm: Kind of
like “springs”

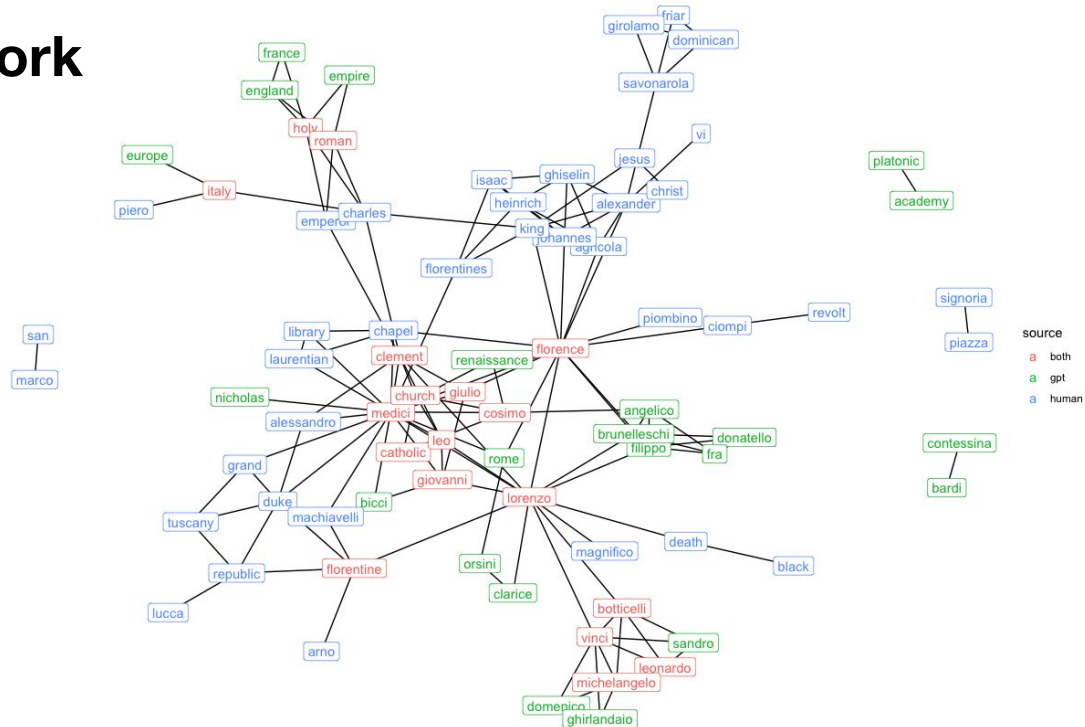
Layouts using igraph

Tutorial by Katherine Ognyanova

<https://kateto.net/network-visualization>

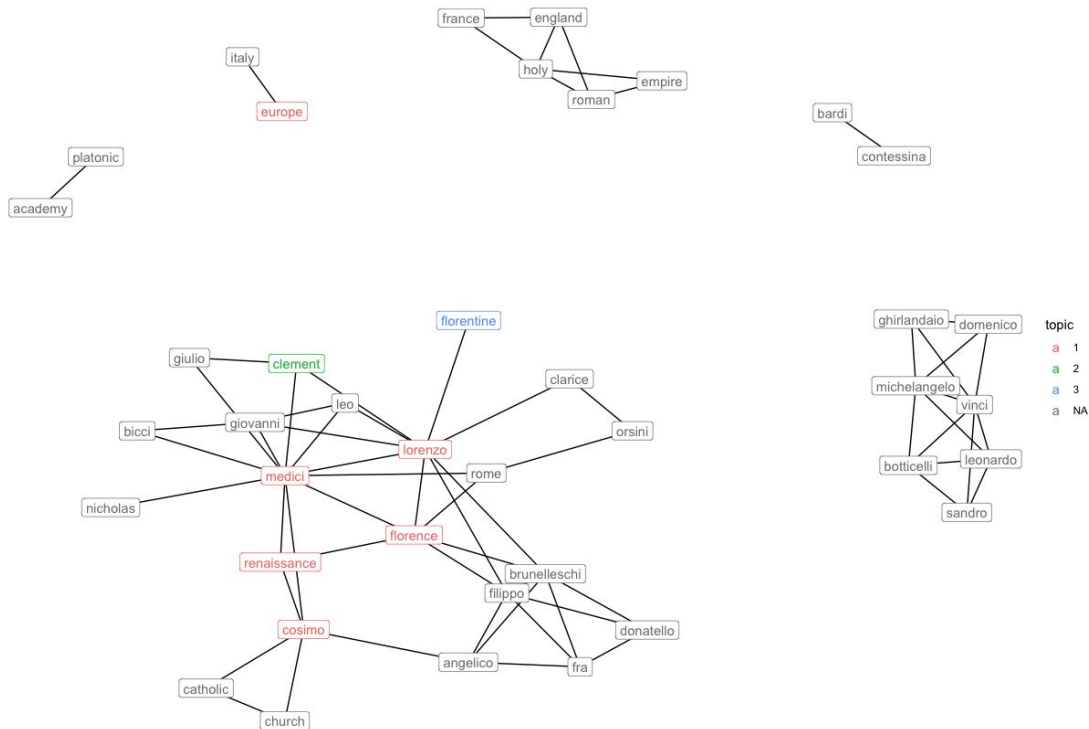
Interpreting the Network Visualization

Combined Network



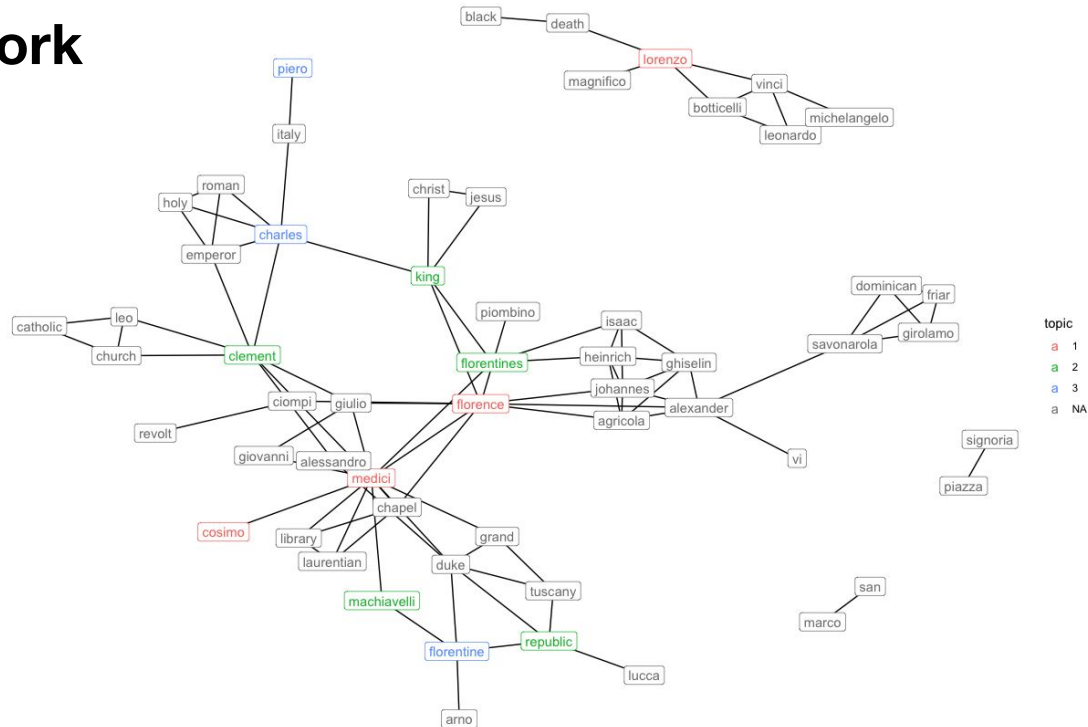
Interpreting the Network Visualization

GPT Network



Interpreting the Network Visualization

Wikipedia Network



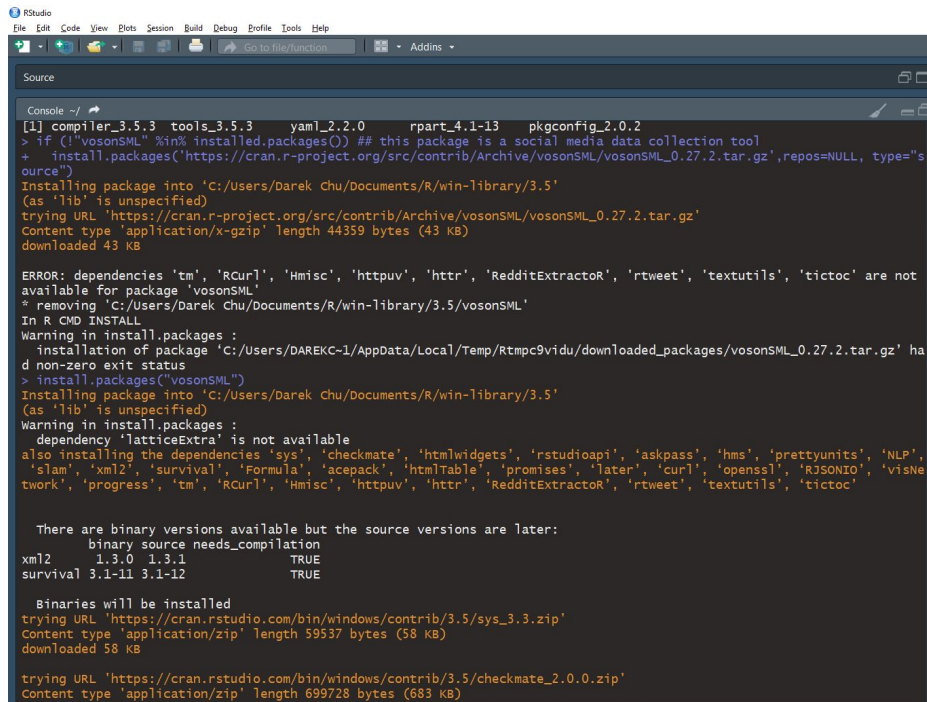
Some Tips

General Tips

- Collect data as soon as possible (**0 points for not collecting data**)
- Make sure to run the code and confirm that it works
 - Execute the code chunk by chunk - Don't execute the entire script!
 - For Windows users, Ctrl+Enter
 - For Mac users, Cmd+Return
- Once you collect data, save it as an RData file
- Focus on interpreting results

When Asking Questions

- Please include screenshots:



```
RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
Source
Console ~/
[1] compiler_3.5.3 tools_3.5.3 yaml_2.2.0 rpart_4.1-13 pkgconfig_2.0.2
> if (!"vosonSML" %in% installed.packages()) ## this package is a social media data collection tool
+ install.packages('https://cran.r-project.org/src/contrib/Archive/vosonSML/vosonSML_0.27.2.tar.gz', repos=NULL, type="s
source")
Installing package into 'C:/Users/Darek Chu/Documents/R/win-library/3.5'
(as 'lib' is unspecified)
trying URL 'https://cran.r-project.org/src/contrib/Archive/vosonSML/vosonSML_0.27.2.tar.gz'
Content type 'application/x-gzip' length 44359 bytes (43 KB)
downloaded 43 KB

ERROR: dependencies 'tm', 'RCurl', 'Hmisc', 'httpuv', 'httr', 'RedditExtractor', 'rtweet', 'textutils', 'tictoc' are not
available for package 'vosonSML'
* removing 'C:/Users/Darek Chu/Documents/R/win-library/3.5/vosonSML'
In R CMD INSTALL
Warning in install.packages :
  installation of package 'C:/Users/DAREKC-1/AppData/Local/Temp/Rtmpc9vidu/downloaded_packages/vosonSML_0.27.2.tar.gz' ha
d non-zero exit status
> install.packages("vosonSML")
Installing package into 'C:/Users/Darek Chu/Documents/R/win-library/3.5'
(as 'lib' is unspecified)
warning in install.packages :
  dependency 'latticeExtra' is not available
also installing the dependencies 'sys', 'checkmate', 'htmlwidgets', 'rstudioapi', 'askpass', 'hms', 'prettyunits', 'NLP',
'slam', 'xml2', 'survival', 'Formula', 'acepack', 'htmlTable', 'promises', 'later', 'curl', 'openssl', 'RJSONIO', 'visNe
twork', 'progress', 'tm', 'RCurl', 'Hmisc', 'httpuv', 'httr', 'RedditExtractor', 'rtweet', 'textutils', 'tictoc'

There are binary versions available but the source versions are later:
      binary source needs_compilation
xml2    1.3.0  1.3.1                TRUE
survival 3.1-11 3.1-12                TRUE

Binaries will be installed
trying URL 'https://cran.rstudio.com/bin/windows/contrib/3.5/sys_3.3.zip'
Content type 'application/zip' length 59537 bytes (58 KB)
downloaded 58 KB

trying URL 'https://cran.rstudio.com/bin/windows/contrib/3.5/checkmate_2.0.0.zip'
Content type 'application/zip' length 699728 bytes (683 KB)
```

Common Error

- If you're running into this error:

```
Error in separate_wider_delim(unnest_tokens(pivot_longer(text, cols = c(gpt, :  
could not find function "separate_wider_delim"
```

- Run **packageVersion("tidyverse")**
 - If you have anything less than **version 2.0.0**, please run **install.packages("tidyverse")**, then restart R, and run the code again from the beginning

Thank You!