

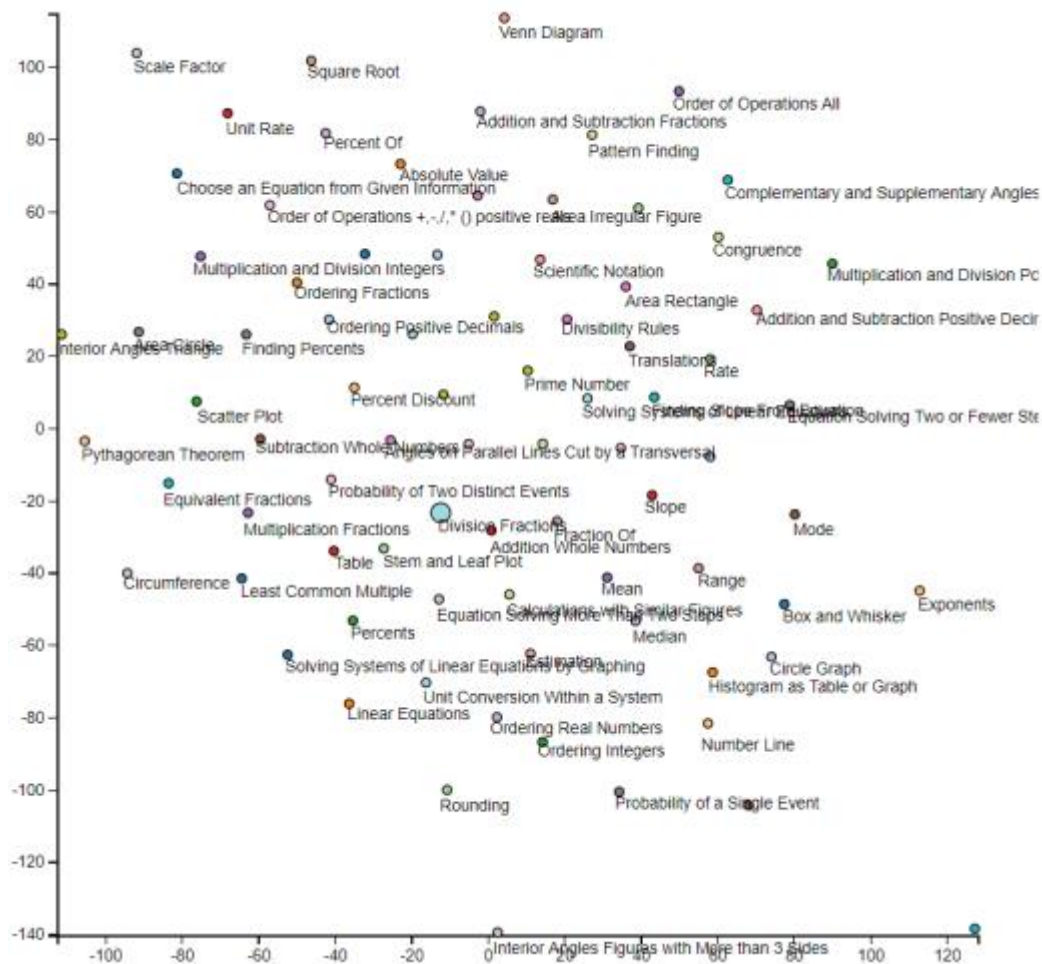
INFO C260F

Homework: Visualizing representation of problems and skills

Vinitra Swamy, Madeline Wu, Wilton Wu

Link to Github Repository: <https://github.com/vinitra/word2vecSkills>

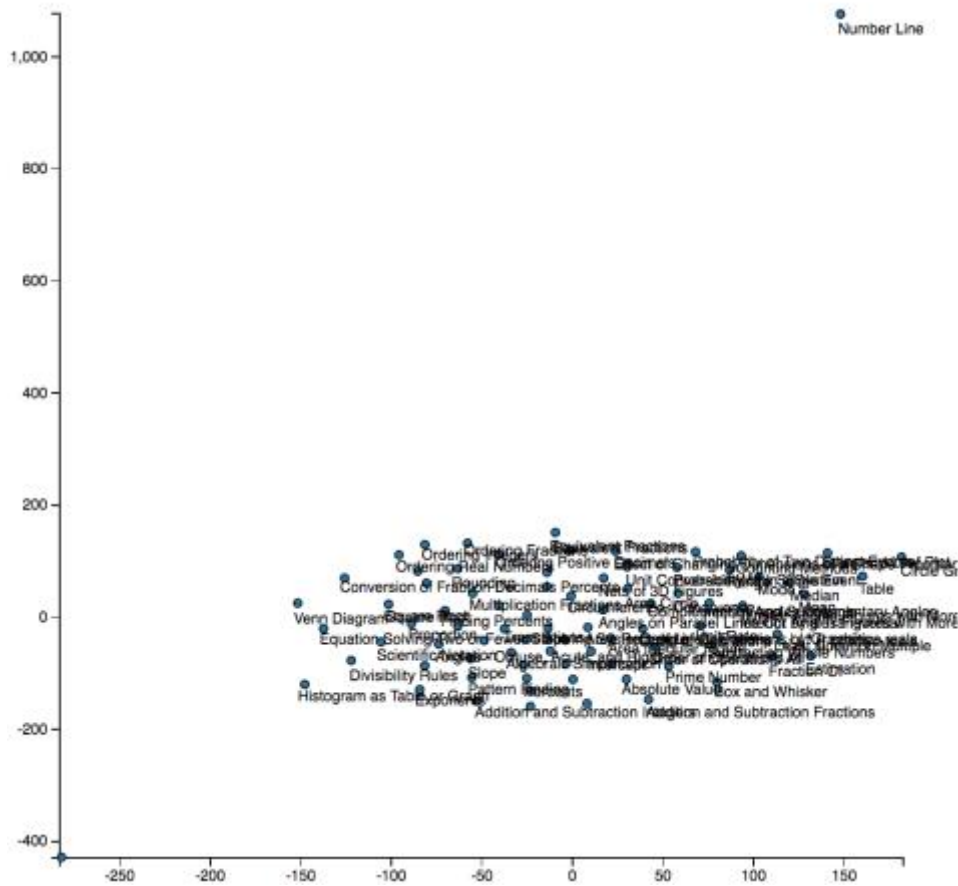
1. See skills_skipgram.py or skipgram_model.ipynb for model code
2. See skills_skipgram.py or skipgram_model.ipynb for vector extraction code.
The following scatterplot was generated using d3-scatterplot



3. The structure of this visualization seems both random/scattered, but also organized. All the math skills are almost evenly spread and similar skills, like mean, median, and range are very close in proximity to one another (near the center of the scatterplot). The skill “Box and Whisker” is just slightly off “Mean”, “Median”, and “Range” as well. This makes sense because all of these corresponding skills are more related than skills such as “Square Root”. There also aren’t any random outliers, probably since all these skills are math related.

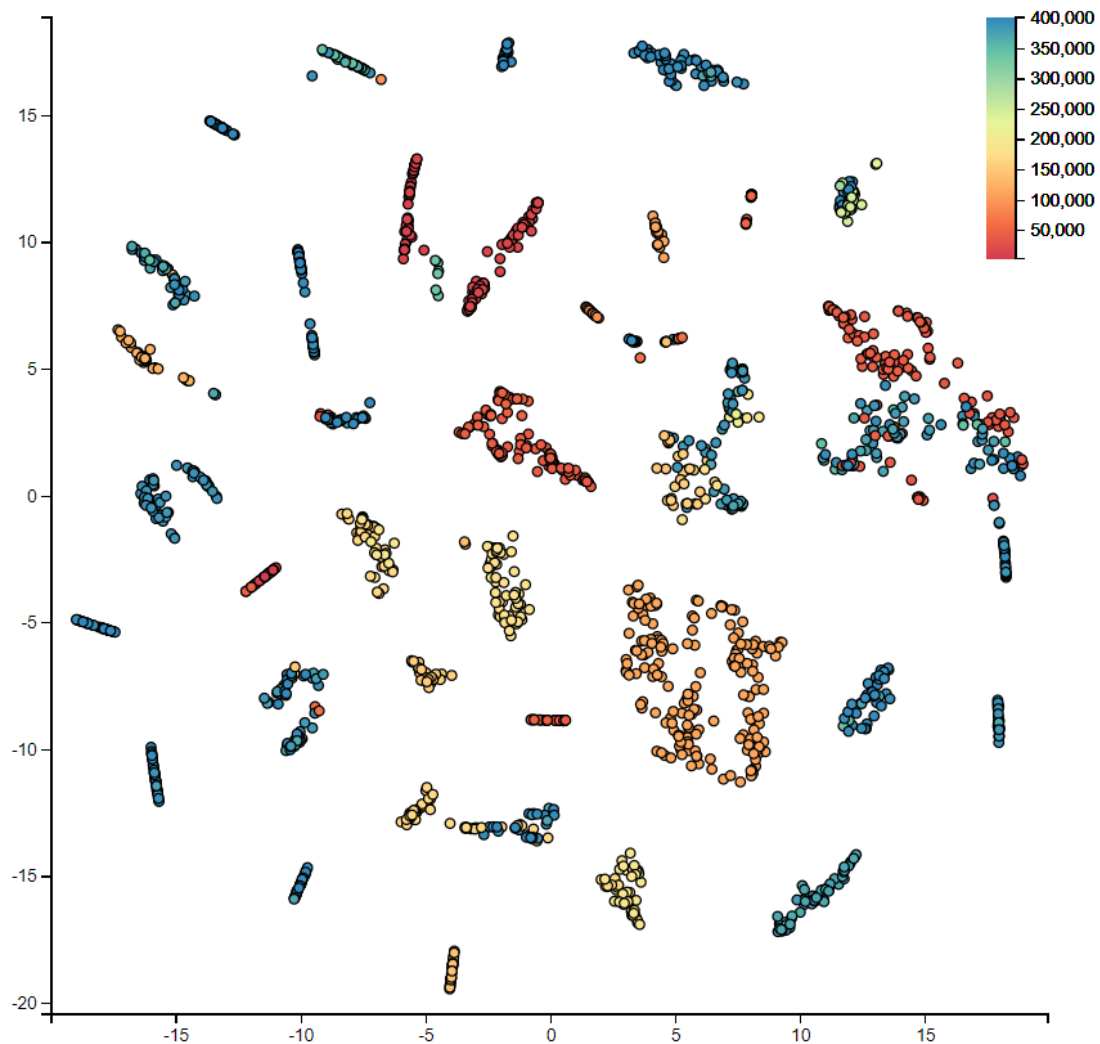
4. Changed window size to 1

With window=1, the visualization looked very “squashed”. It seems as though the visualization is more limited and thus more clustered with a smaller window size (except for the skill number line for some reason).



5. Code in `assistent_skipgram.py`

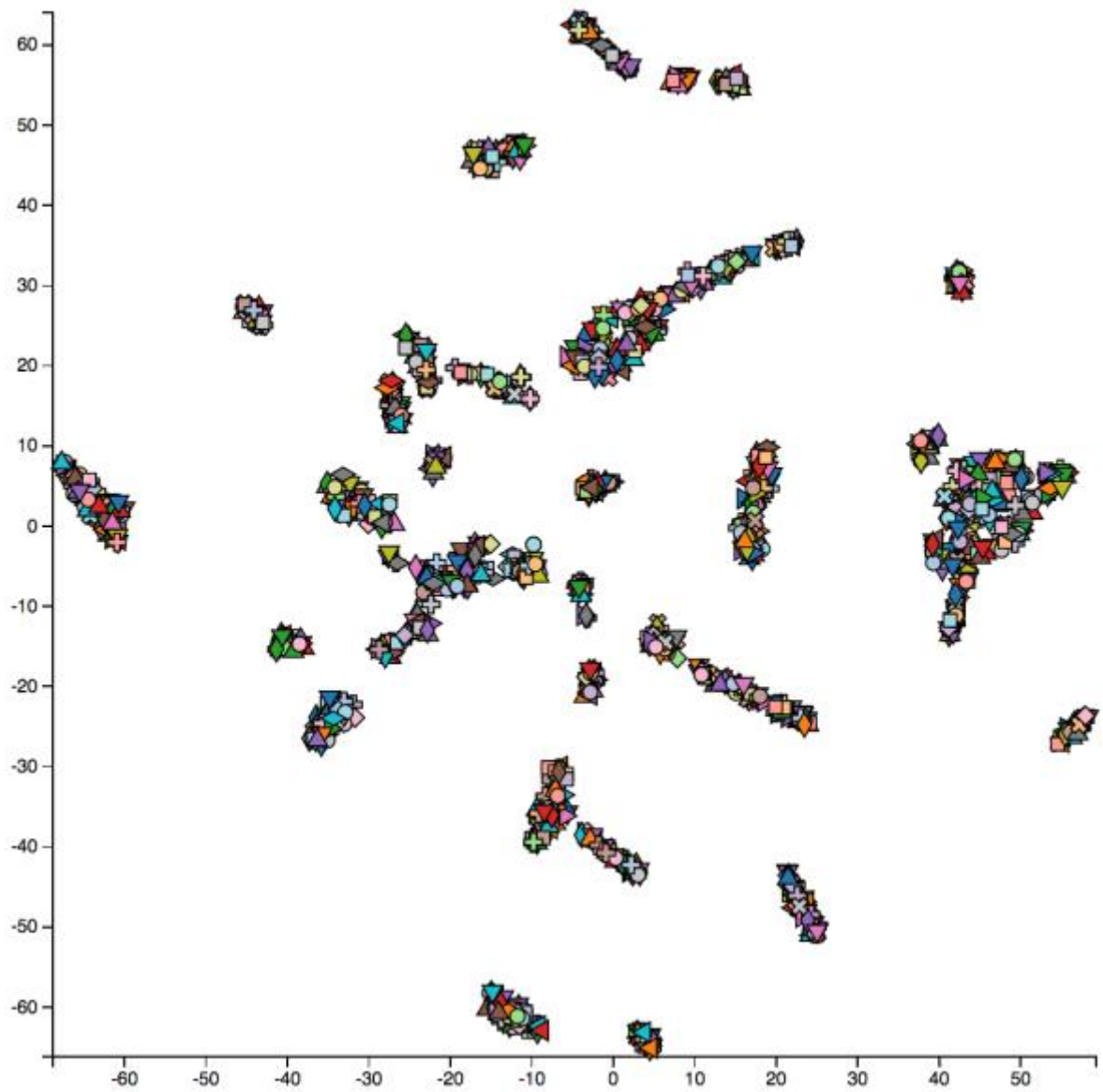
The following scatterplot was generated using `d3-scatterplot`



The structure of the visualization for `assistent.tsv` is also scatter but organized. There are obvious clusters for some skills, but many other skills interweave with one another.

(answer continued on next page)

D3 scatterplot for window=1



Does not cluster by skill anymore.