Graphing Ideas

**Purpose:**

Create a graph (or graphs) that is not only representative of the data, but also scalable with the data. This means that if the data available grows very large, the usefulness of the graph will not diminish when it tries to represent the data.

**Things to Consider:**

* The words that we are comparing will only increase, and become more niche
* The starting words that we have now are very broad.
* Some words may be specific to only one group (Bohemian, corporate, traditional)
* Is there a way to seperate words into categories by machine learning?
  + Doing it by hand may be tedious.
  + Also, there may be other groups of words that are significant
* We can use 2 or more graphs jointly
* The graph should be as simple as possible, but also has to convey as much meaning as possible. The reader should not have to strain their brain to understand a graph because the whole purpose of a graph is to easily convey information to the audience.

**Example of category and words:**

Social Interaction

* Meeting
* Group study
* Talking
* Friends
* Chatting

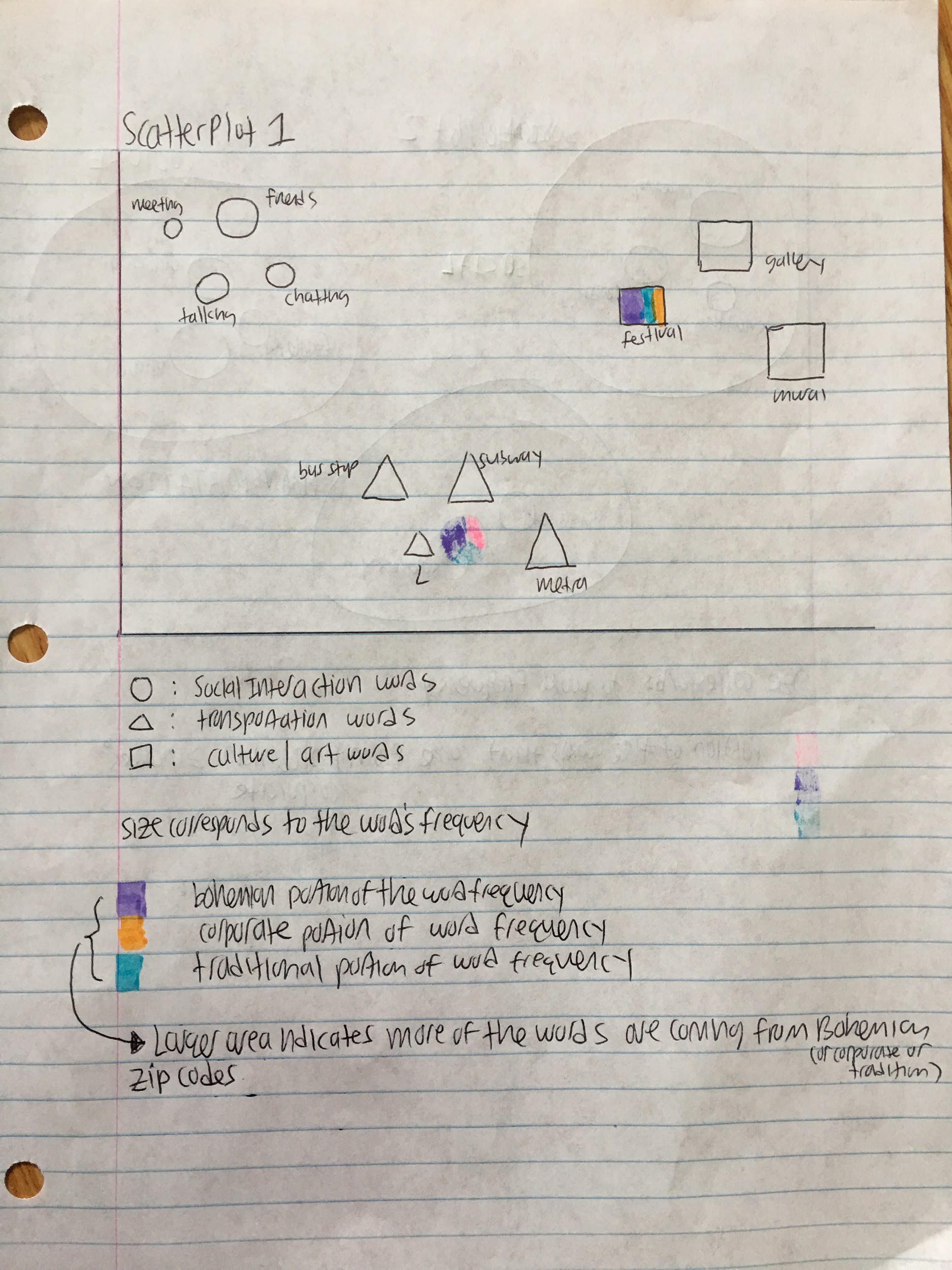
Transportation

* Bus stop
* L
* Subway
* Metra

Culture

* Graffitti
* Street art
* Mural
* Gallery
* Festival

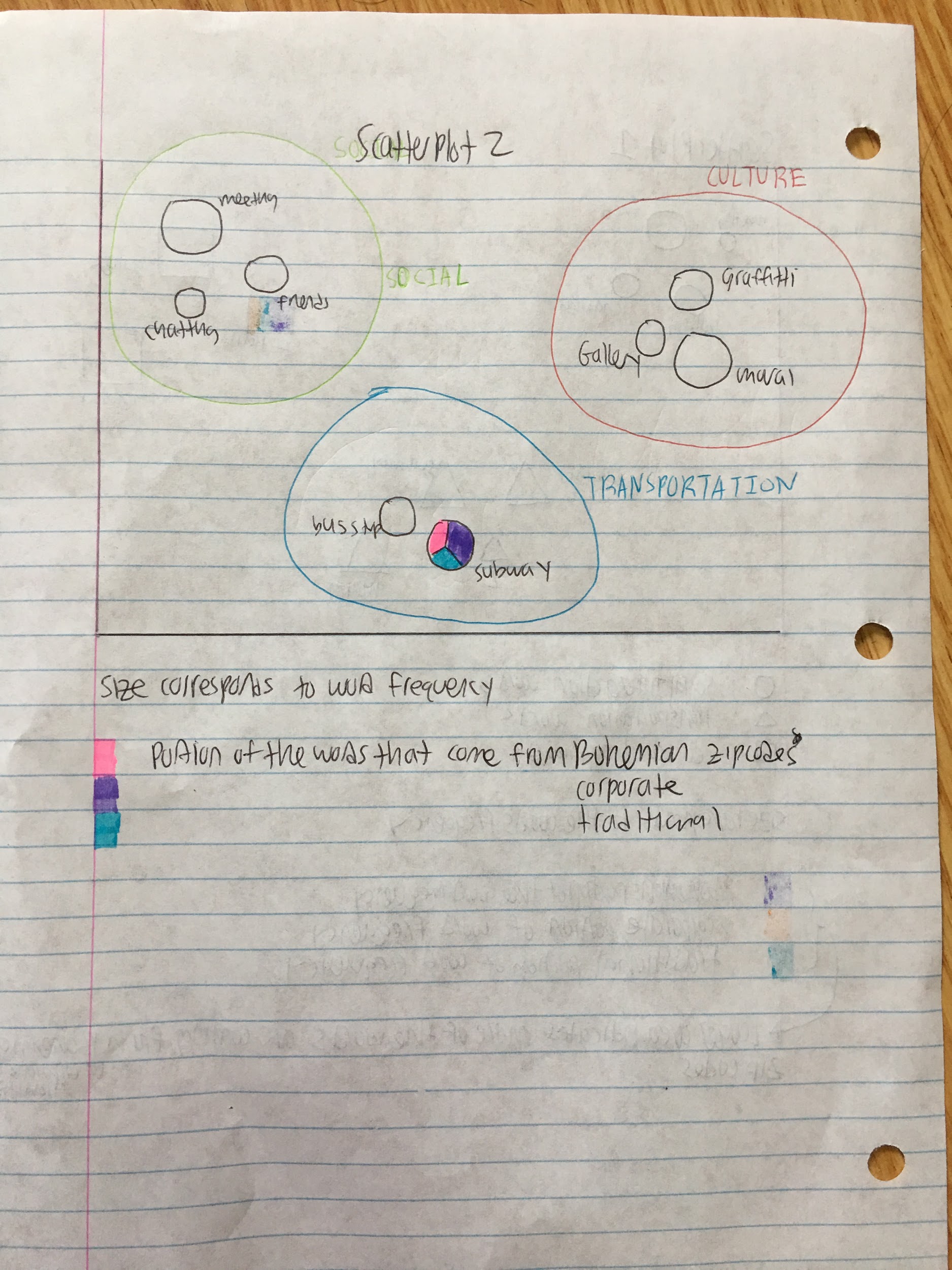
**Plots:**

**Scatterplot 1:  
**

Cons:

* the different shapes confuses the sizes (how do we compare the size of the triangle to the size of the circle)?
* X, y axis don’t have any significant meaning

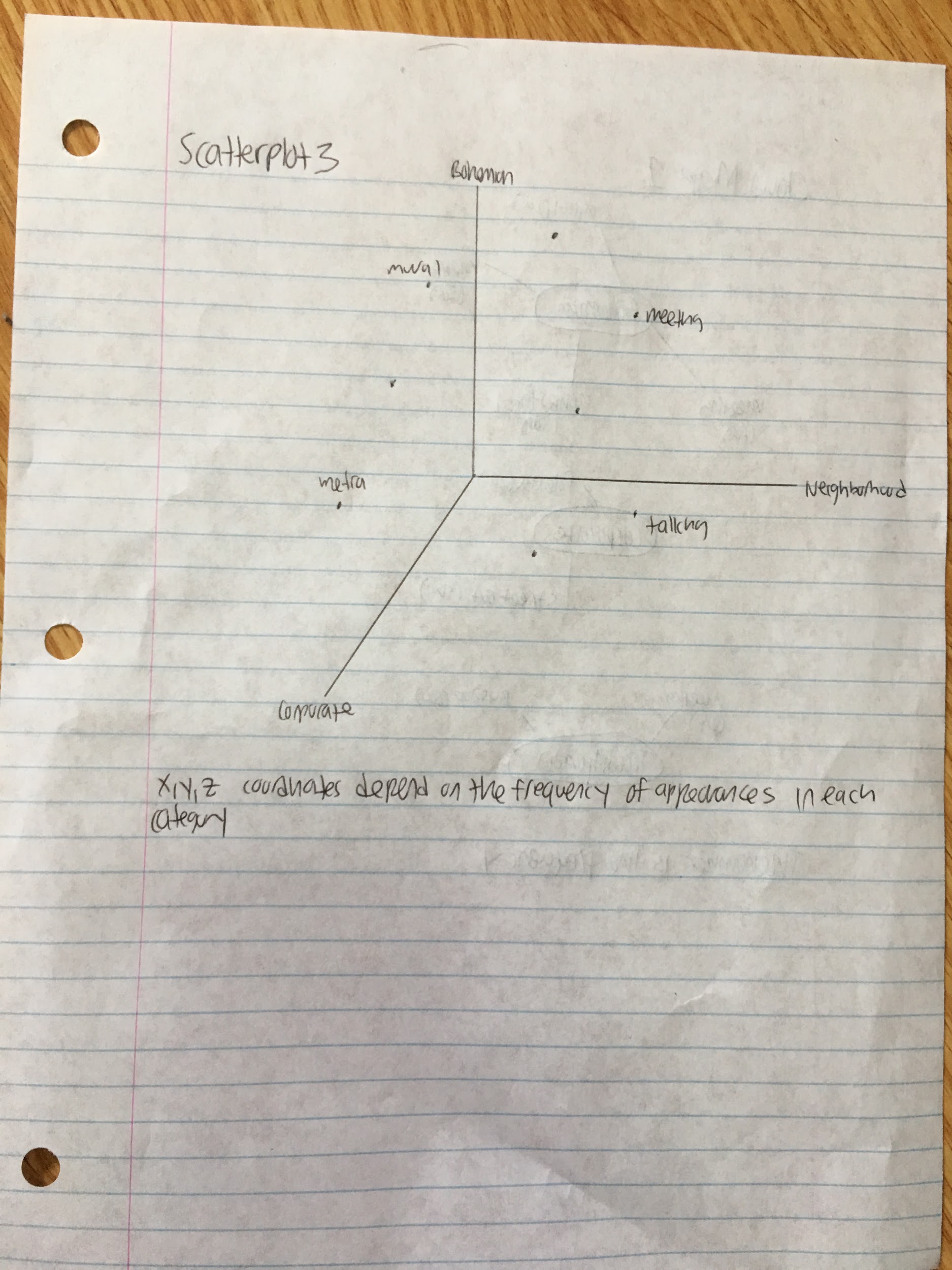
**Scatterplot 2:**

****

Cons:

* Larger circles may overlap
* X,y axis don’t have significant meaning

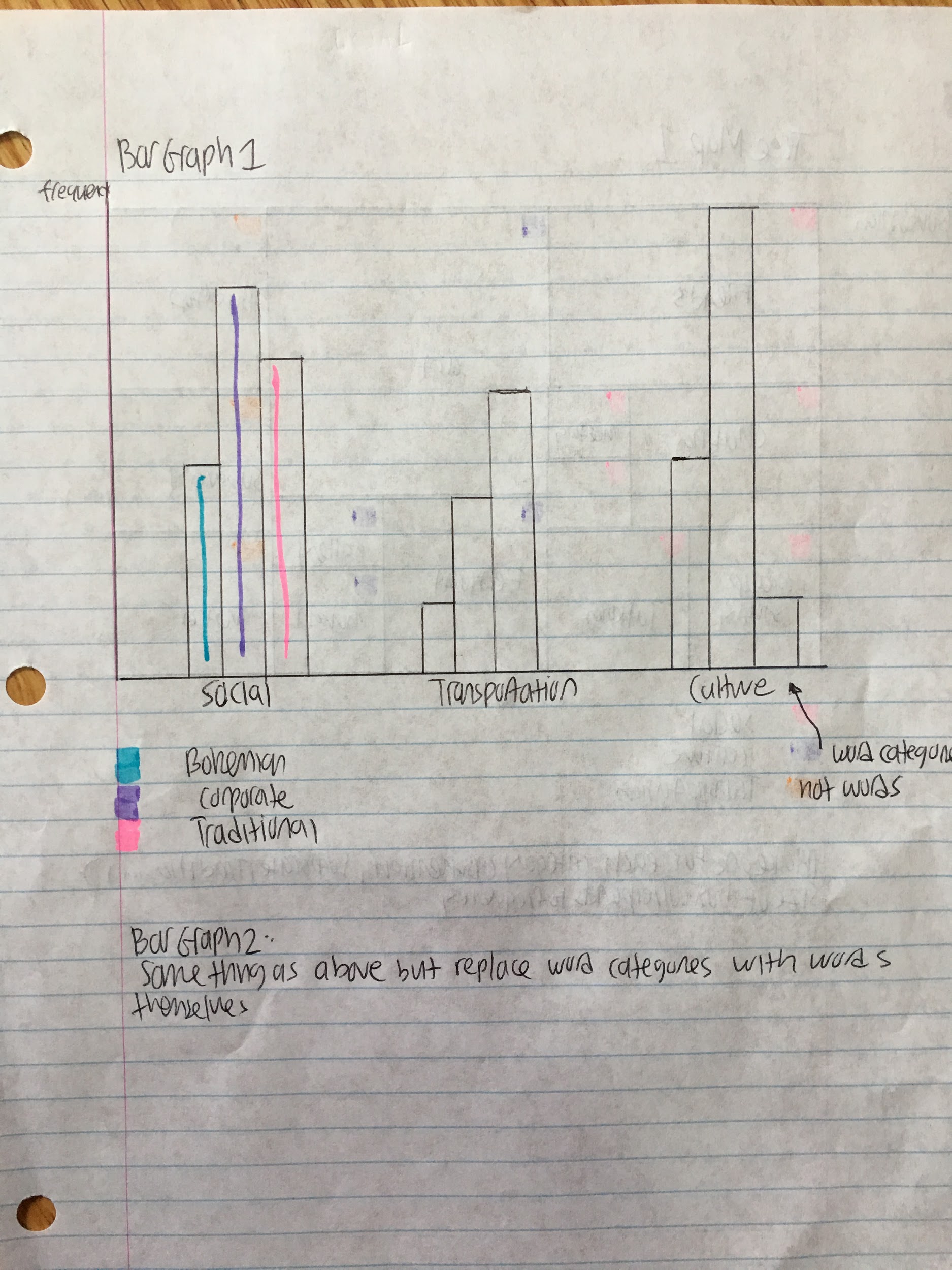
**Scatterplot 3:**



Cons:

* 3-dim on 2-dim paper is problematic (hard to see where the points really are)
* No logical grouping of words

**Bar graph 1:**

****

Cons:

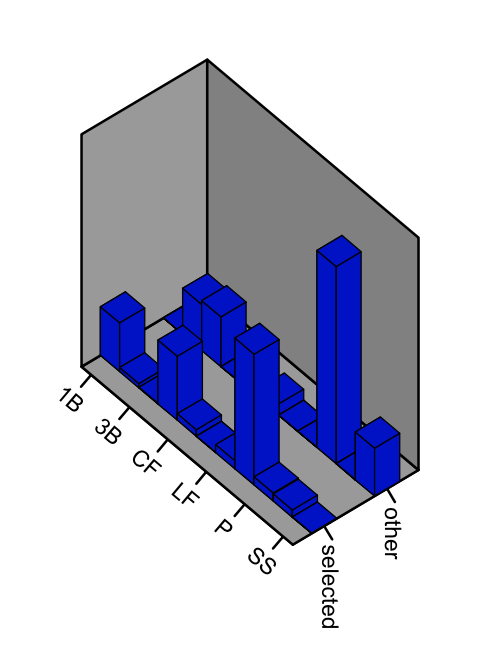
* Doesn’t say much about individual words

**Bar graph 2:**

Cons:

* Unscalable with many words
* Ordering of x axis may be a problem

**Bar graph 3:**



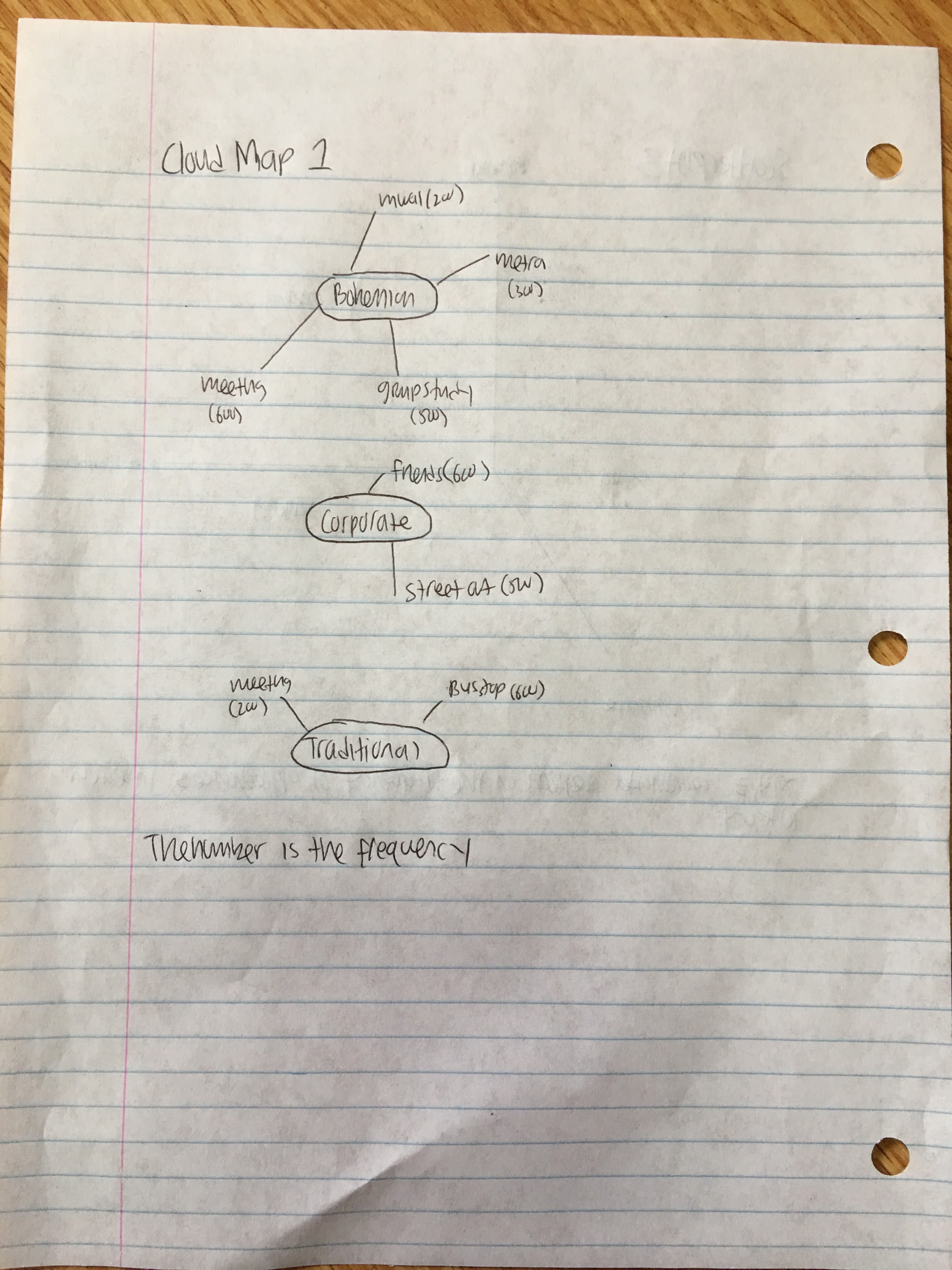
Source:

* page 223 from the textbook

Something like this, but where it says “selected”, “other”, replace it by “Bohemian”, “Corporate”, “Traditional”.

Replace the other axis with words.

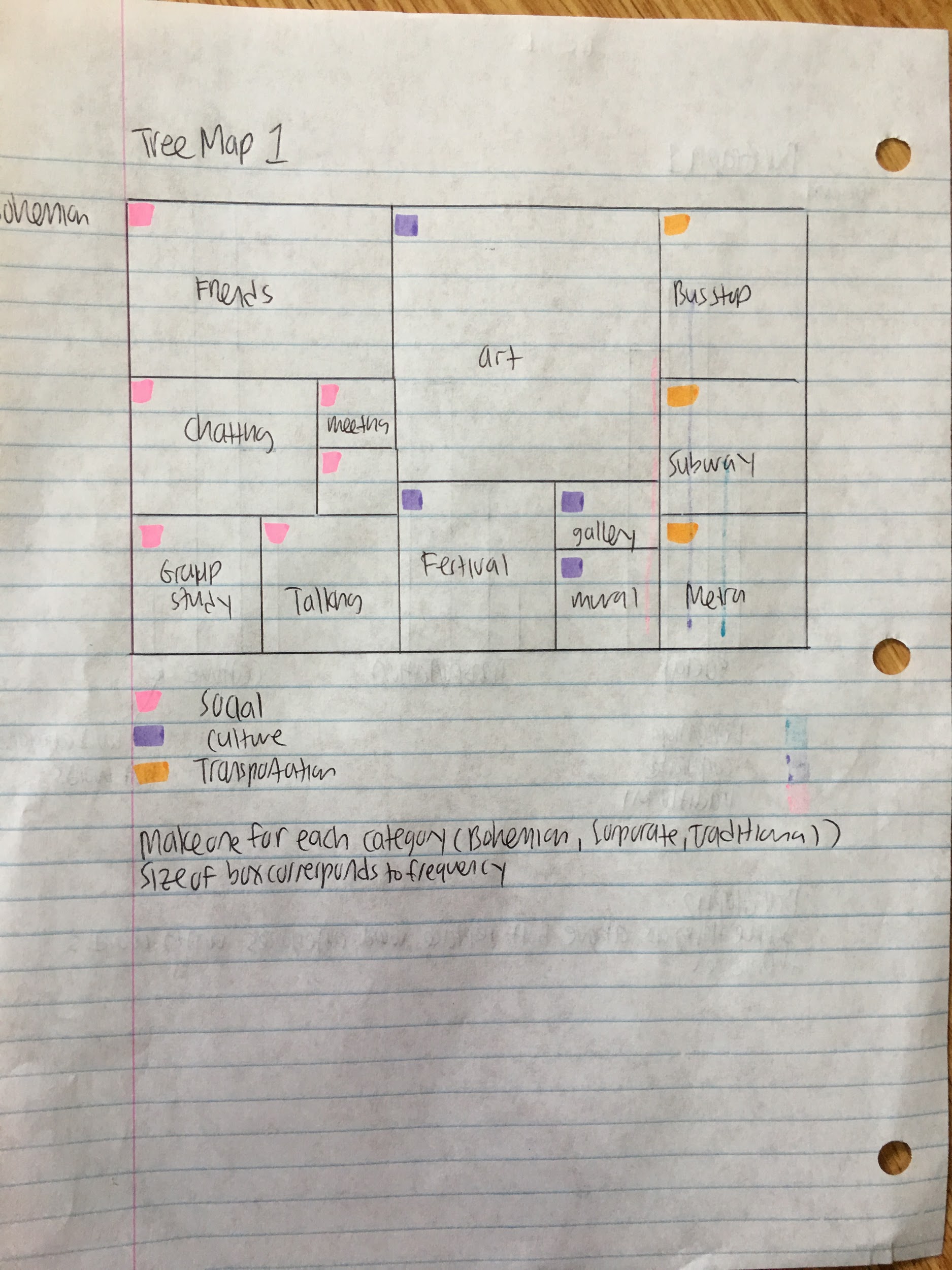
**Cloud Map 1:**



Cons:

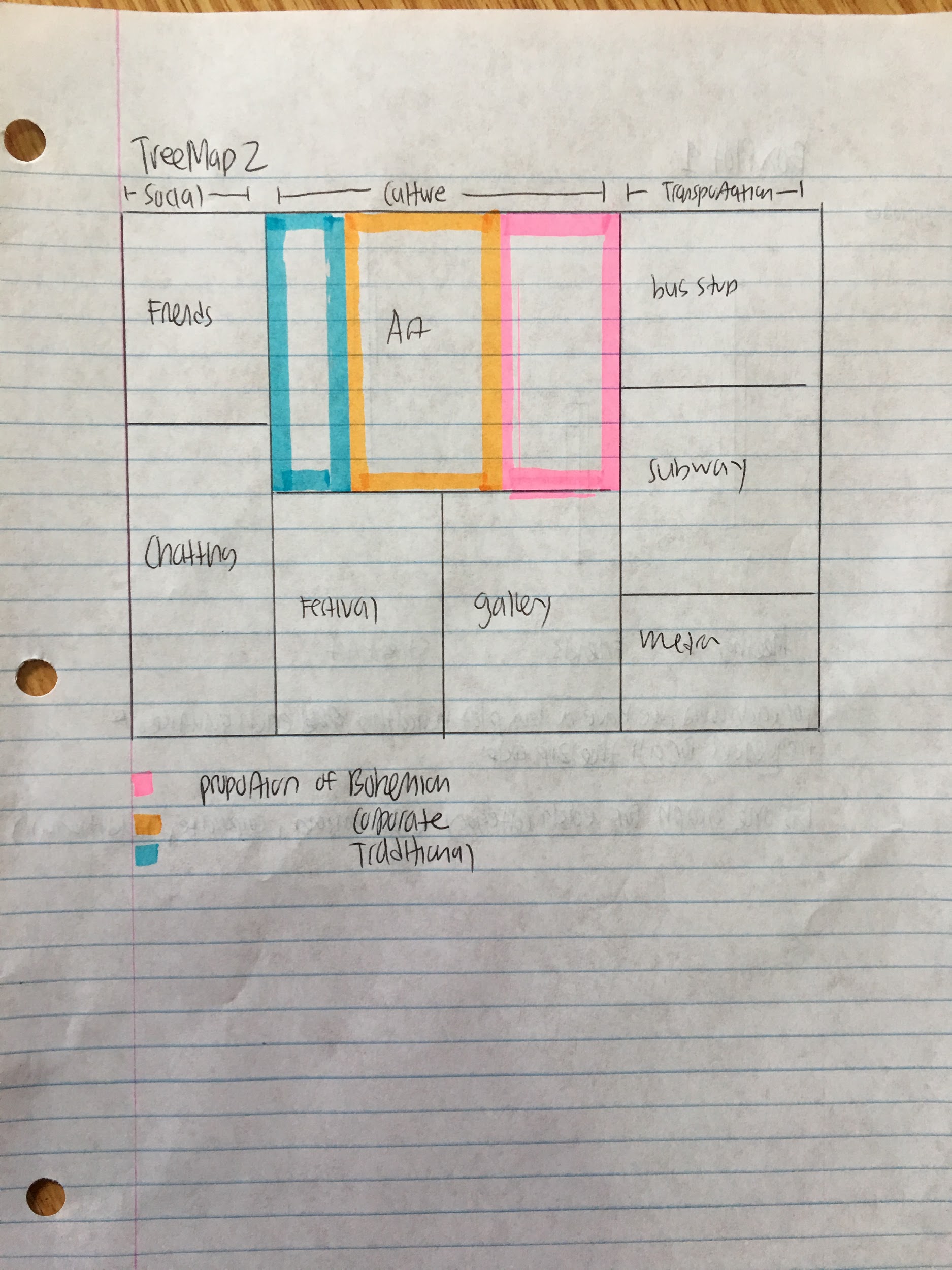
* very difficult to explicitly compare words from one category to another

**Tree Map 1:**

****

Cons:

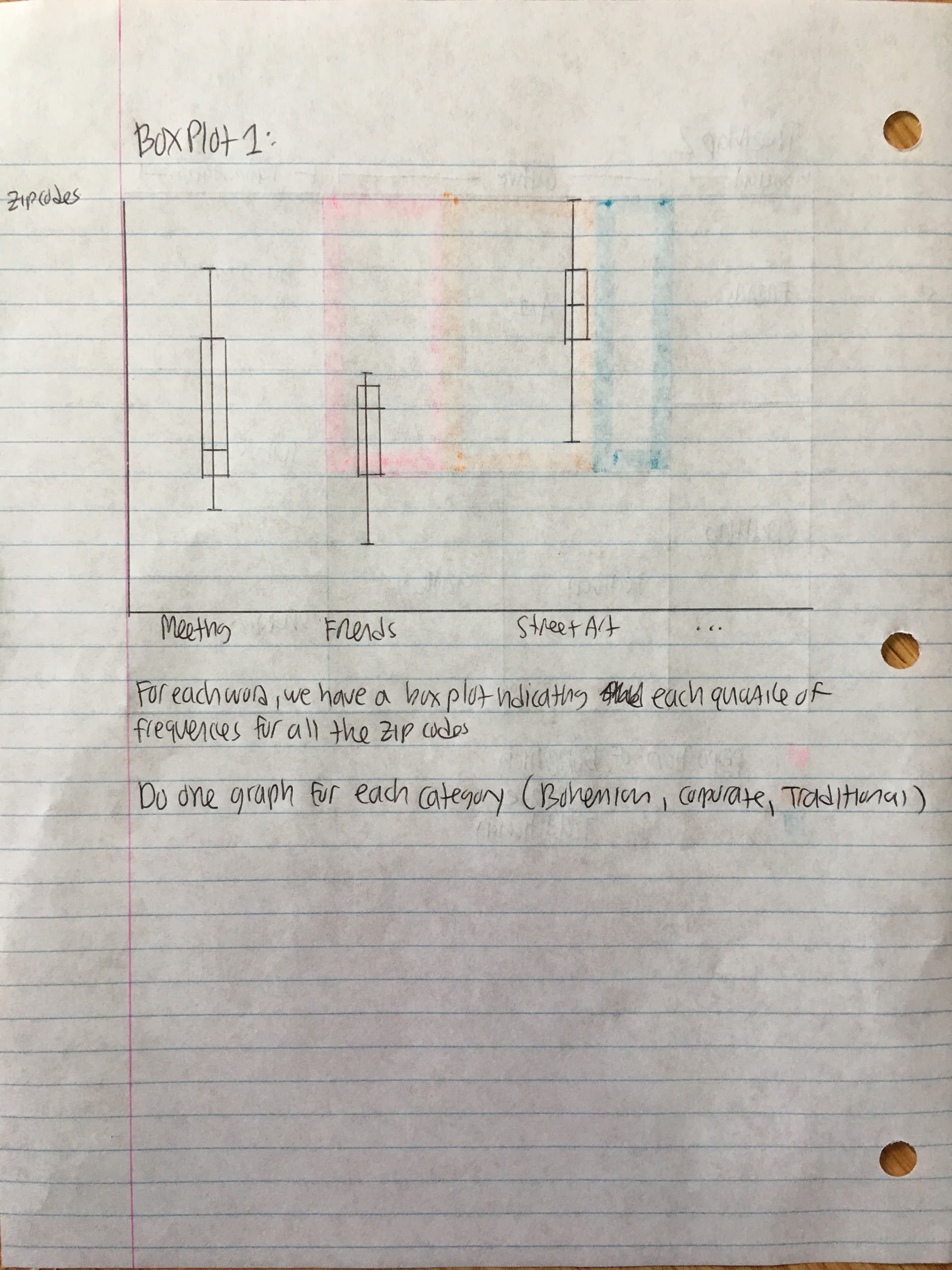
* Only really useful for proportions and not absolute amounts. The different categories may have different numbers of words, making comparisons more difficult.

**Tree Map 2:**

Cons:

* same as tree map 1

**Box Plot 1:**

****

Cons:

* not very scalable

**Plots that may be useful to consider:**

* Hierchal trees (see page 129 of the book)
* Box plots
* Panel scatterplots
* Linked plots

**Things to check out again:**

* Page 325 of the textbook