

Tutorial on Basic Data Analysis of Texts
Eduardo Navas (ean13@psu.edu)

Introduction to Principles of Data Visualization

Exercise: Data Mine two texts to discover patterns for comparative analysis.
See details for assignment below.

Introduction:

Go over the following online resources:

Top 20 Data Visualization Tools

<http://www.creativebloq.com/web-design/top-data-visualisation-tools-9122832>

Software Studies Initiative:

<http://lab.softwarestudies.com/>

Infographics:

<http://infoqr.am/>

<http://visual.ly/>

<http://dailyinfographic.com/>

Discuss Principles of Text Analysis

Create an account on Many Eyes, and run some basic tests:

<http://www-969.ibm.com/software/analytics/manyeyes/#/>

(Note: You will need to run it on Safari or a browser which can have its Java Applet extension enabled)

Create an account with Infoqr.am:

<http://infoqr.am/>

You can experiment with the tools available on the above sites once you have gathered some data as discussed below.

For an inclass exercise use Voyant for data-mining a large text. This could be a website or a PDF of a book:

<http://voyeurtools.org/>

Look for patterns and share what you discover in class. Output the images for you to study later. Make sure to keep notes of the different techniques that will be covered during class lecture.

Once you have a sense of how a basic analysis functions with the tool above, you can also develop visualizations using the following tools:

<http://docs.voyant-tools.org/tools/>

Consider how the visualization with each tool affects how the data is understood. You may choose one of these visualizations to include as part of your written argument/assessment on the

data. You should spend some time experimenting with the different visualizations, and read their respective documentation to then decide if, how, why it may be useful for your own analysis.

Exercise: Think of two books or texts that you want to data mine for similarities and differences. *You should consider books or texts that are relevant or related to the research you are performing in class.* For your mining of text, make sure you are able to access PDF versions online, or html versions. Run them through Voyant first and perform the methods discussed during class.

Once you have gathered data, you can then decide how best to visualize it. You may use one of the tools above to develop infographics, or you may use the data to develop your argument as part of a paper, or both. This is up to your professor.

You should have a question that you want to explore such as, but not limited to:

- 1) What are some of the common words used in the books? How does this affect the way the books are structured and aesthetically experienced?
- 2) If it is said that one author is influenced by the other, how is this taking place directly in material production vs. intertextuality?
- 3) What are the formal structures of the books?
- 4) What political and cultural implications can be furthered discussed based on your analysis.

It takes a long time to fully mine a text, so try to look for something very specific that you can manage for the amount of three weeks. You should focus on one basic question and data mine for it as much as possible then think of how to design an infographic that will show the results of your analysis. You are encouraged to include graphs and charts that you develop during your research as part of the infographic. You should choose material that you have read and are familiar with. This will enable you to develop a qualitative analysis while using quantitative information.

Think of a design that will best represent the areas from your research you want to feature. You should spend some time evaluating the design options available at infogr.am or another site you prefer. You can use one of these designs to present your materials or you can design one of your own. If you develop your own design, you must make sure it is mathematically accurate.