Sketch a Data Story

Use WordCounter to analyze a musician's lyrics and sketch a story with their words

v1.0.0

What is WordCounter?

WordCounter analyzes text and tells you the most common words and phrases. It helps you quantitatively summarize your text so you can investigate how words are used. Sketching a data story helps participants build their data literacy by finding and drawing a story about what words show up in famous musicians' lyrics.

Learning Goals

- Increased ability to analyze and reason about text data.
- Understanding that one can find many stories in one dataset.
- Ability to quickly communicate a data-driven story to others.

Run the Activity

Solving a Problem

Text is data too! Finding patterns in large sets of text can be hard to do by hand, but computers can help. WordCounter helps you start to analyze text quantitatively but showing you the most commonly used words and phrases.

Share Inspirational Examples

Visualizing music lyrics is a popular thing to do. Show examples like Tahir Hemphill's "Rap Research Lab" (http://rapresearchlab.com) or Javier Arce's "Spotimap" (http://javier.is/spotimap). Ask participants what stories they see in these examples.

Introduce the Tool

Open up WordCounter (*https://databasic.io/wordcounter*) and choose "Elvis Presley" from the sample data drop-down menu. Press "Count!" to show an analysis of all of Elvis' lyrics. Show your group how the word cloud of his lyrics shows that he used the word "love" a lot, but only by looking at the bigrams and trigrams can you start to understand how and why. The fact that Elvis said "love" so times but only said "attorney" once could be a fun and revealing story to sketch. Mention that you can download CSVs of the results for further analysis.

Total Time

30 to 45 Minutes

Audience

3 - 100 people. Ages 12+.
Designed for grades 6 - 12, Higher Ed classrooms, News Organizations, Non-profits, and Community Workshops. No prior experience with data is required.

Space

- A projector and computer.
- Ability to break out into small groups of 3 clustered around a computer.
- Large tables or floor, or tape to stick paper to walls so participants can draw.

Supplies

- Computers
  1 for every 3 participants
- Large pieces of paper
  roughly 2 feet x 3 feet
- Thick crayons or markers

Run the Activity (continued)

Launch the Activity

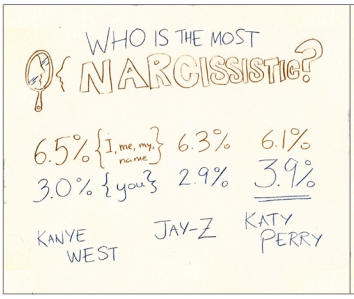
1. Participants have 15 minutes.
2. Participants work in teams of three.
3. Each team uses WordCounter to analyze the lyrics of a musician and find a story to tell about their lyrics.
   https://databasic.io/wordcounter
4. Each team sketches a visual way to tell that story on a big piece of paper with crayons.

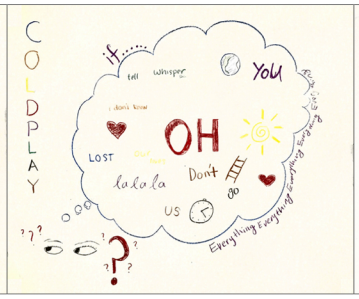
Share-Back

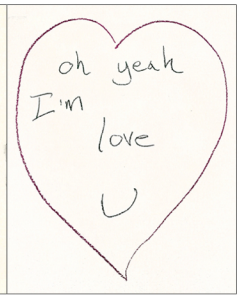
As you walk around keep an eye out for different types of stories to highlight. If you see any particularly interesting visuals, or novel stories, keep them in mind to call on during the share-back time at the end of the workshop. With your time constraints in mind, ask for groups that want to share back their story. Limit them to 1 minute, so you don't get lost in one group's example. As folks share, make sure to call on any groups you noticed earlier that had novel or different stories they created.
Focus on these themes:

- There are many types of stories you can tell with the same data.
- Sketching is a helpful way to communicate your ideas to those around you.
- Working with data can be fun.
- Demonstrate how you can download the CSVs to continue your analysis in Excel.

Sample Sketches







Reminders

- Stick figures are fine.
- Data stories can be as simple as "Elvis said 'love' way more than 'clambake'".
- The "stopwords" checkbox lets you ignore or include common words like "the" or "and".
- Don't worry about being perfect, just sketch something.

Terms to Introduce

Bigrams
A bigram is a two word phrase.

csv
A "comma-separated values" file; a standard way to save tables of information in a plain text file.

Qualitative data
Data that describes something (like a picture or quote).

Quantitative Data
Data that counts or measures something (like a number).

Stopwords
Stopwords are words in that are helpful to ignore for text analysis, such as "this", "that", "and" and "but".

Trigrams
A trigram is a three word phrase.

Word Cloud
A picture of all the words in a document, where each word is bigger if it is used more, and smaller if it is used less.

DataBasic is a suite of easy-to-use web tools for beginners that introduce concepts of working with data. DataBasic is a project of the Engagement Lab at Emerson College and the MIT Center for Civic Media. Supported by the Knight Foundation.