What is WTFcsv?

WTFcsv gives you a quick overview of what is in your .csv file. It helps you quickly move to asking your spreadsheet some questions that can help you find stories to tell. This hands-on activity helps participants build their data literacy while looking at some fun datasets and brainstorming questions you can ask about them.

Learning Goals

• Understanding that data has many types, including numbers, text, and dates.

• Increased ability to form hypotheses about and find stories from metadata.

• Understanding that one can find many stories in one dataset.

• Understanding that datasets can be combined to answer more complicated questions.

Run the Activity

Solving a Problem

Open up a spreadsheet in Excel or something and poll the room for how many participants spend time working with data like this. Introduce the CSV file as simply one standard way to save spreadsheet data. Finding stories in a CSV can be hard, especially when there are lots of columns and rows. It's helpful to start by thinking about what questions you can even ask your data. WTFcsv helps you quickly analyze a CSV to understand what is in it so you can start asking it some questions.

Share Inspirational Examples

Show an open data portal from a place near the workshop location (for instance, *http://data.gov* in the US). Talk about the wide variety of datasets that are available online in portals like the one you are showing. Download our UFO sample data (*https://databasic.io/wtfcsv/ufo.csv*) and open it in Excel. Show that each column holds different types of information about each row.

Introduce the Tool

Open WTFcsv (*https://databasic.io/wtfcsv*) with the "UFO Sightings" sample data. Explain that each card on the results page is a summary of the information in all the rows for one column.

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 sttick paper to walls so participants can draw.

Supplies

• Computers
  1 for every 3 participants

• Copies of the WTFcsv question handout
  https://databasic.io/wtfcsv/handout.pdf

• Pens

Run the Activity (continued)

Demonstrate flipping a card over to flip between the visual and text summaries. Give examples of questions to ask, other data you'd need to answer that question, and how you'd get it. For instance, one question you could ask the UFO data is "Do towns with more people have more sightings?" To answer this, in addition to the UFO data you'd also need to get population data for all the towns. You could get this information from US Census records.

Launch the Activity

1. Participants have 15 minutes.

2. Participants work in teams of three.

3. Each team gets one WTFcsv Question Handout.
   https://databasic.io/wtfcsv/handout.pdf

4. Each team uses WTFcsv to analyze one of the sample CSVs and brainstorm questions to ask it.

5. Each team writes down the most interesting question they came up with on the handout.

Share-Back

As you walk around keep an eye out for different types of questions to highlight. Ask for groups that want to share back their top question. 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 questions. Focus on these questions and topics:

- Are all of the answers to your questions contained in the dataset.

  Probably not, sometimes you need to find another dataset or call someone to get the information you need

- Where is the data from? If the source of the data isn't revealed, you should be skeptical.

  We downloaded the UFO data from the National UFO Reporting Center (http://www.nuforc.org).

- Do you see places where values are missing? Missing values are one way data can be "messy"!

  Ask participants for some ways data can be "messy"; and what "clean" data might mean.

- WTFcsv shows you aggregated summary data, but sometimes you can ask interesting questions about just one row in a dataset, or look for an "outlier". Did participants find any stories about outliers?

Reminders

- We aren't trying to answer the questions, just make a list of ones to ask.

- There are no stupid questions.

- Some data needs to be "normalized" for comparison. For instance, to compare the number of sightings by town you should normalize it by population so you compare sightings per person.

Terms to Introduce

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

Data Types
Data is often classified by type. Common types of data include numbers, dates and text.

Metadata
Metadata tells you about your data so you can understand what it is.

Normalization
A process to make data points comparable in the same way.

Outlier
One piece of your dataset that is very different from the rest of it.