

Introduction

GIS works best with clean, consistent, quantifiable data. Humanists often work with messy, inconsistent, qualitative and subjective data. We can sometimes force humanistic data to conform to the standards of GIS, but just as often we might prefer to preserve and visualize the messiness, ambiguity, or uncertainty of our materials. This recipe suggests ways of preserving uncertainty and ambiguity in our data. (If you are interested in visualizing uncertainty and ambiguity, see [this recipe](#).)

Certainty refers to how confident we are that a date, location, or other piece of data is correct. If your great-great grandmother was born either in Minsk or Warsaw, you can be only 50% certain of each location.

Ambiguity refers to being unable to quantify a date, location, or other piece of data precisely. If your data is imprecise - not machine-readable without translation (or perhaps transformation) into a more precise value - it qualifies as ambiguous. If your great-great grandmother lived in the Pale of Settlement but you don't know where, you only have an ambiguous location.

Ingredients

A spreadsheet for your data (Google Sheets or AirTable recommended; Excel may cause importing issues)

How to do it

1. Determine what kind of ambiguity or uncertainty you have in your data. Common kinds include:
 - a. Imprecise dates, locations, or qualitative data
 - b. Imaginary locations
 - c. Subjective, experiential knowledge
 - d. Unlikely or impossible times and geographies
2. Determine whether that ambiguity or uncertainty is significant and needs to be preserved/visualized. For instance, if change over time is not important to your research/teaching question, then being uncertain about dates has no bearing on your work. You can just ignore or cut that data and move on.
3. If uncertainty/ambiguity is significant, in your spreadsheet, create a new column, ideally close to the column whose uncertainty/ambiguity the new column will record. Be sure to label it in a way that will help you remember what other column it is associated with.
4. In the new column, record the uncertainty or ambiguity in your data using discrete values, e.g., a range of 1 through 5 for certainty; "True" and "False" or "1" and "0" for ambiguity. For example:

Lat	Long	Location	Ambiguous Location
50.075539	14.4378	Prague	TRUE
51.507351	-0.127758	London	TRUE
40.668333	-74.045556	Port of New York	TRUE
40.737153	-73.98857	Clarendon Hotel	FALSE
40.733946	-73.982325	Dvorak House in New York	FALSE
40.735732	-73.987261	National Conservatory of Music	FALSE
40.765	-73.98	Music Hall (Carnegie), New York	FALSE
41.79	-87.58	Chicago World Exposition, 1893	TRUE
43.204972	-91.950997	Spillville, Iowa	TRUE

Note that GIS requires precise locations, whether you know them or not. In this example, specific latitudes and longitudes have been chosen so that each location can appear on a map, but some of the entries are precise, accurate locations, and some are provisional approximations that nevertheless need to be included in the data set. The approximations can be chosen so that they appear in obviously incorrect places to avoid confusion (like a park or an ocean), or so that they are near the likely correct, but ambiguous, location.

Ambiguous dates -- relating to events that happened within a range, or that happened in a given month or year but on an unknown day -- are best entered as ranges in your spreadsheet, allowing you to indicate temporal (un)certainty in maps that animate points chronologically. Four columns are recommended (earliest begin date, latest begin date, earliest end date, latest end date), as in the following example.

1	Latitude	Longitude	Place	Location	Earliest Begin Date	Latest Begin Date	Earliest End Date	Latest End Date
2	42.125	-80.0806	Erie High School	Erie PA	1884-01-01	1884-12-31	1885-01-01	1885-12-31
3	42.1249	-80.0888	Clark Commercial	Erie PA	1885-01-01	1885-12-31	1885-01-01	1885-12-31
4	42.1249	-80.0888	Clark Commercial	Erie PA	1885-01-01	1885-12-31	1885-01-01	1885-12-31
5	42.128	-80.0876	Episcopal Cather	Erie PA	1885-01-01	1885-12-31	1885-01-01	1885-12-31
6	42.125	-80.0806	Erie High School	Erie PA	1885-02-22	1885-02-22	1885-02-22	1885-02-22
7	42.1249	-80.0888	Clark Commercial	Erie PA	1885-02-27	1885-02-27	1885-02-27	1885-02-27
8	42.1248	-80.0835	YMCA	Southeast Corne	1885-04-01	1885-04-30	1885-04-01	1885-04-01
9	42.1302	-80.0872	First Baptist Chu	5th St & Peach S	1886-04-01	1886-04-30	1886-04-01	1886-04-30
10	42.1283697	-80.0703123	Cross and Crown	E 12th St & Ash	1886-08-01	1886-08-31	1886-08-01	1886-08-31
11	42.1267992	-80.0900361	First Methodist C	707 Sassafras St	1887-01-01	1887-12-31	1887-01-01	1887-12-31
12	42.126	-80.0829	Loyal Ladies Lea	Erie PA	1887-01-01	1887-12-31	1887-01-01	1887-12-31

5. In addition, a “Notes” or “Details” column can afford opportunities to explain the nature of the uncertainty of ambiguity. If two sources disagree, for instance, you can indicate as much - whether as a reminder for you or for users of your data. For instance, “We know Burleigh performed in the city in March, but not exactly when or where.”
6. At the end of your process, you could separate out data into distinct spreadsheets according to its uncertainty or ambiguity. For instance, if a number of records include confirmed information and a number of records include unconfirmed information whose certainty or ambiguity is equivalent, then separating those two or more subsets into distinct spreadsheets will allow a map user to toggle various layers on and off according to uncertainty/ambiguity. For an example, see <https://stolaf.maps.arcgis.com/apps/webappviewer/index.html?id=abcf07d9a224a6886355e77c3a800>

[0a&mobileBreakPoint=1](#), where if you click on the “Layers” widget (which looks like three pieces of paper stacked) you can select/unselect “Marian Anderson Unconfirmed.”

How it works

While it is possible to enter ranges for spatial information, as recommended above for dates, it is not practicable for most projects. Expression of that type of ambiguity is best handled in the visualization process. In such instances, it may be advisable to record ambiguity (“True” or “1” in the appropriate column) and uncertainty (1 through 5), as the second value can be used to express the geographic extent of ambiguity graphically.

Further Resources

“Formulating Ambiguity in a Database”

(<https://nodegoat.net/blog.p/82.m/21/formulating-ambiguity-in-a-database>)

Johanna Drucker, “Humanities Approaches to Graphical Display”

(<http://www.digitalhumanities.org/dhq/vol/5/1/000091/000091.html>)