Assignment 1 — Information visualization

Deadline 11 a.m. on Tuesday 30/03/2021

Dr T. Torsney-Weir

Credits for CSC337/CSCM37 25% of module

Motivation

Last night Indiana Jones was on the TV and you watched it. You thought maybe being an archaeologist sounds like a fun career. Then you realized how often Indiana Jones was in a life-threatening situation and decided that maybe the leather jacket/adventuring archaeologist isn't for you. However, you realize you're taking this amazing visualization course and maybe you can leverage some of those skills to inspect certain historical places. So you decide to look at the Pleiades dataset and see if you can find some interesting relationships.

You decide to download the locations dataset from that resource at http://at lantides.org/downloads/pleiades/dumps/pleiades-locations-20210201.csv.gz. You have a look at the dataset and it is a lot of complex data. Maybe you can use visualization to help make some sense of the complex data. There is also a README file for the dataset at http://atlantides.org/downloads/pleiades/du mps/README.txt which describes the fields of the CSV file. There are other datasets available on that site which you can use.

Understanding the data

As with real-world projects, acronyms and special terminology are prevalent. Thus, you may need to conduct some research, like reading the data description online or googling terms you're unsure of. You may also ask questions on Discord or http://VisGuides.org if you need help.

Goals of the analysis

The goal is to create visualizations that maximize our understanding of the data. In addition to some obvious factual information, such as:

1. What is the oldest place in each location?
2. What is the distribution of structure types around the world?
3. Can you make a map of an ancient city?
4. How do newer settlements in a civilization absorb older ones?

There are also some less obvious questions, such as:

1. Are there any interesting patterns or trends in the data?
2. What are the outliers?
3. What insight can we gain from multi-variate visualizations of the data?
4. Are there any relationships between the different time ages?

Your boss shows you a bar chart they made in excel and asks for, "something like this." However, you know from your insightful vis lectures that single graphics of old-fashioned pie charts, bar charts, bubble charts, and line graphs might not be very insightful or aesthetically impressive. We resolve to do better (and maybe get a promotion)!

Coursework: A Real-World Challenge (maximum marks 50)

Your task is to use literate visualization to produce **five different** visual designs which can convey some meaningful and hopefully interesting insight about the data. You are required to supply five unique observations about the data (and not just repeat the same observation over and over). For each design, fill out a literate visualization schema as discussed in https://github.com/gicentre/litvis /blob/main/documents/tutorials/introduction/intro2.md.

Each visualization design you create should be in a separate markdown file. Thus you will have 5 files at the end. **You don't have to use elm-vegalite** for this coursework. You can use another tool, like Altair if you want. However, **you must include images and code in the markdown files** to create your visualizations in whatever tool you used. There is a prepared set of files at https://github.com/SwanseaU-TTW/csc337_coursework1.

You are allowed and encouraged to further abstract the data, e.g., aggregating data records into groups, combining some attributes together, or add new data, e.g., cities, geographic regions. In fact, you can do anything you like to the data and you are expected to make changes, e.g., perhaps combining some spreadsheets together, creating new spreadsheets, or gathering data from other sources.

Description Template

Provide the following information for each visual design you create. In parenthesis are the literate visualization labels to use to identify these in the source files:

Aim (`aim`): Things we can learn from the visualization, e.g., from this visualization we can see this pattern...

Visual Design Type (`vistype`): The name/type of the visual design

Image: The visualization itself as an image

Visual Mappings (`vismapping`): Each of the visual design mappings. Include the data mapping information about color, shape, size, position (x,y axes), and any other visual mappings.

Data Preparation (`dataprep`): Any modifications to the original data that had to be performed to generate your beautiful image.

Improvements (`improvements`): What are the limitations of your visualization? What improvements could you make to it?

A good observation requires interpretation of the resulting image that you generate.

What is better?

Create visualizations that:

- convey information and knowledge
- aid discovery of: patterns, deviation, hierarchy, relationships and association
- identify relationships between data attributes
- depict data at different scales
- separate noise from the signal
- are intuitive to laypersons and easy to learn
- are aesthetically pleasing

Examples: Simply creating more standard pie charts, line charts, and bar charts is not good enough. Some examples of more advanced visual designs are treemaps, parallel coordinates, matrix charts, scatter plots etc. You will see many examples in the lectures. If you'd like to get a head start, you may watch the course lectures from last year on YouTube (https://www.youtube.com/playlist?list=PLZo40sVmw_4No1_K-Xllf04O977PYngq1), look at the slides on Blackboard, or read the recommended text book[@ward:2015]. You are also encouraged to ask questions in lectures.

Describe the insight that your visualizations provide. What can we learn from your visualizations? How are they better than a standard line, pie, or bar chart? Also, a common mistake that students make is using a treemap to show non-hierarchical data. Treemaps are a hierarchical data visualization technique. Without a hierarchy, they may be even less effective than a bar chart. However, you can *create a hierarchy* from your data to visualize. See the example submission and Geng et al.[@geng:2011a] for examples.

Digital Maps: You are encouraged to explore the use of digital maps in your visualizations. For a good source of digital maps, please visit the following URL: http://commons.wikimedia.org/. You can then do a search for the relevant maps.

Hands-On Tutorials: For extra help, there are extensive tutorials at https://github.com/gicentre/litvis/tree/main/documents/tutorials. There's also a very nice tutorial on youtube: https://www.youtube.com/watch?v=pMHmQX3TZ8A.

You may certainly use the digital, geo-spatial maps from these tutorials in your submission. However, you will not receive credit for the bar charts and line charts demonstrated in the tutorial. Also, a paper called, *Liquid Diagrams: Information Visualisation Gadgets*[@andrews:2010a] discusses the use of digital maps for visualization.

Submission

Below is a list of possible tools that you can use. If you want to use a tool or technology not on this list then please clear it with me first. **If not, you will receive a 0 for the assignment.**

You are required to submit a document, which contains your name, student number, and **five narrative examples** with complete template information as in the example submission. *Make sure each of your visual design types is distinct.* In other words, two Treemap layouts are two instances of one visual design.

Create a directory with your student number. Then put the final pdf, all markdown, and all data files used. Make sure it's laid out as:

```
<student number>
|- coursework.md
|- followup1.md
|- followup2.md
|- followup3.md
|- final.md
|- <studentnumber>.pdf
|- <datafile 1>
|- <datafile 2>
|- ...
```

Then, zip this directory. Please create a **.zip file** and not a tar, gzip, or any other format. Make sure extracting this zip file properly recreates this directory.

You are required to upload the zip file to Canvas. **If you only submit the code or the pdf then you will receive a mark of 0**

Creating a pdf from VS Code preview

The easiest way I've found to create a pdf from the preview is:

1) Right click on the preview window
2) Select "Open in Browser" from the menu. Your web browser should open
3) Use the print to pdf functionality of your browser
4) Once you have all 5 of these, combine the pdfs together.

Useful Links to Visualization Tools

Here is a list of visualization tools that you may use for this assignment. You can potentially use something else but you **must** clear it with me first.

- elm-vegalite (Elm): https://github.com/gicentre/elm-vegalite
- Data Driven Documents (JS): http://d3js.org/
- Vega (JS): https://vega.github.io/vega/
- Altair (Python): https://altair-viz.github.io/
- VegaLite.jl (Julia): https://github.com/queryverse/VegaLite.jl
- vegalite (R): https://cran.r-project.org/web/packages/vegalite/index.html

I've deliberately not included packages like ggplot and matplotlib in this list because it is not easy to produce interactive tools using them.

Tableau and Lyra are not included as I want you to have more experience with using code to generate visualizations. These are excellent tools for prototyping though!

To help you get started with litvis, there are a number of tutorials at https://github.com/gicentre/litvis/tree/main/documents/tutorials.

Both carefully prepared tutorials start from the very beginning with a description of a sample data set and proceed step-by-step creating a selection of visual designs.

A paper called, *Visual Analysis of Document Triage Data*[@geng:2011a], provides a nice overview of free information visualization tools and their features. If you have any questions or concerns, please do not hesitate to contact the module lecturer or the module teaching assistant.

References