Assignment One–Information Visualization
"Let us do better than Professor C!"

Robert S Laramee

October 3, 2016

Number of Credits for CSCM37:
15% of module
Recommended hours: 20-25 hours
Color Printed and Blackboard Submission Deadline:
31 October by 11:00am
Please see the submission instructions on the module web page.

1 Problem Statement and Data Chaos

Professor C was approached by the Public Health Data Administration (`https://www.gov.uk/government/organisations/public-health-england`) to help them with a serious challenge. The National Health Care Service (NHS), UK, has collected a massive amount of health care data about National General Practice Profiles aka, CCGs (Clinical Commissioning Groups) in the UK. The data contains information about each person that has visited a CCG service and the diagnoses that have been given. The data that has been collected for each CCG is the following.

1. Population Age Distribution
2. Practice Summary
3. Estimated Disease Prevalence
4. CVD Coronary Heart Disease
5. CVD Stroke and TIA
6. CVD Heart Failure and atrial fibrillation
7. CVD Risk Factors for CVD
8. Diabetes
9. Mental Health
10. Respiratory Disease
11. Chronic Kidney Disease
12. Musculoskeletal Conditions
13. Other Conditions
14. Secondary Care Use–Outpatients
15. Secondary Care Use–A&E and Inpatients
16. Child Health

In addition, each of the above has been collected over a number of years from 2009-2014. It is a lot of complex data. In fact, it feels like data chaos.

Now he/she would like to know if visualization can be used to gain insight into his/her data.

As with real-world projects, acronyms and special terminology are abundant. Thus, you may need to contact do some research, like Googling NHS terms, in order to gain an understanding of the terminology. You can also ask Doris Hain, a Fingertips Technical Specialist for Public Health Data Science, Public Health England any questions. Her email address is: `profilefeedback@phe.gov.uk`.

The data can be downloaded from:
`fingertips.phe.org.uk/profile/general-practice/`

Professor C would like to obtain and convey the insight contained in the data visually. 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 average size of each CCG?
2. What does the distribution of CVD look like?
3. Which CCGs have the highest incidence of Diabetes?

Professor C would also like to find out some less obvious facts, such as:

1. Are there any interesting patterns in the data?
2. Are there trends in the data?
3. What are the outliers?
4. What insight can we gain from multi-variate visualizations of the data?
5. Are there any relationships between the attributes? For example Mental Health versus CVD.

Professor C made some attempts to visualize the data using old-fashioned pie charts, bar charts, bubble charts, and line graphs. He was disappointed by the results, which are not very insightful, not visually uninformative, nor aesthetically impressive. *Can we do better than Professor C?*

2 Part 1: Tasks for the first 60 points (maximum marks: 60)

You are required to select three or more appropriate visualization tools for visualizing the dataset concerned.

You are allowed and encouraged to further abstract the data, e.g., aggregating objects 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 or creating new spreadsheets.

Your task is to produce five *different* visualizations which can convey some meaningful and hopefully interesting insight about the data. *You are to do better than Professor C! That means you are required to use more advanced visualizations than typical line graphs, bar charts, pie charts, and bubble charts.*

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

Simply creating more standard pie charts, line charts, and bar charts is not good enough. You 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 Geng et al. [2] for an example.

Description Template: Provide the following information for each visualization you create:

- The visualization itself as an image
- Visualization Type: The name/type of the visualization
- Tool Name: The tool that was used to generate the visualization
- Visualization Mappings: Each of the visualization mappings, e.g., size is mapped to population, color is mapped to temperature, x-position is mapped to time etc.
- Observations: Things we can learn from the visualization, e.g, from this visualization we can see this pattern...

Sample Description:

- Visualization Type: Figure 1 is a treemap.
- Tool Name: ManyEyes
- Visualization Mappings:
  - each bottom level rectangle represents a car type
  - the cars are grouped by country
  - the position of each rectangle / car is determined by decreasing frequency, the more cars manufactured, the further to the left.
  - the size of each bottom-level rectangle is mapped to the difference in mpg between the city and highway MPG of each car type.
  - color is also mapped to the difference between city and highway MPG for each car
- Observations: We can clearly see an outlier, i.e., that the Toyota is the only car with greater city MPG than highway MPG. We can also see that South Korea produces cars. I did not know this previously.

Digital Maps: You are encouraged to explore the use of digital maps in your visualizations. Using digital maps will yield bonus points. 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.

A paper called, *Liquid Diagrams: Information Visualisation Gadgets* [1] discusses the use of digital maps for visualization.

3 Part 2: Tasks for the next 40 points (maximum marks: 40)

Most information visualizations are two dimensional in space. In other words, they are confined to an xy plane. For another possible 40 points, carry out a comparative study of two-dimensional (x,y) vs three-dimensional (x,y,z) information visualizations using the same data set from Part 1. The term dimensional in this context refers to spatial dimensionality.

Create *two* 3D information visualizations and compare them with their 2D counterpart, e.g., 3D scatterplots vs 2D scatterplots. For full credit, create 3D visualizations that can that can be rotated.

For this option, you are required to write a short report (recommended 500-1000 words), provide your analysis of the

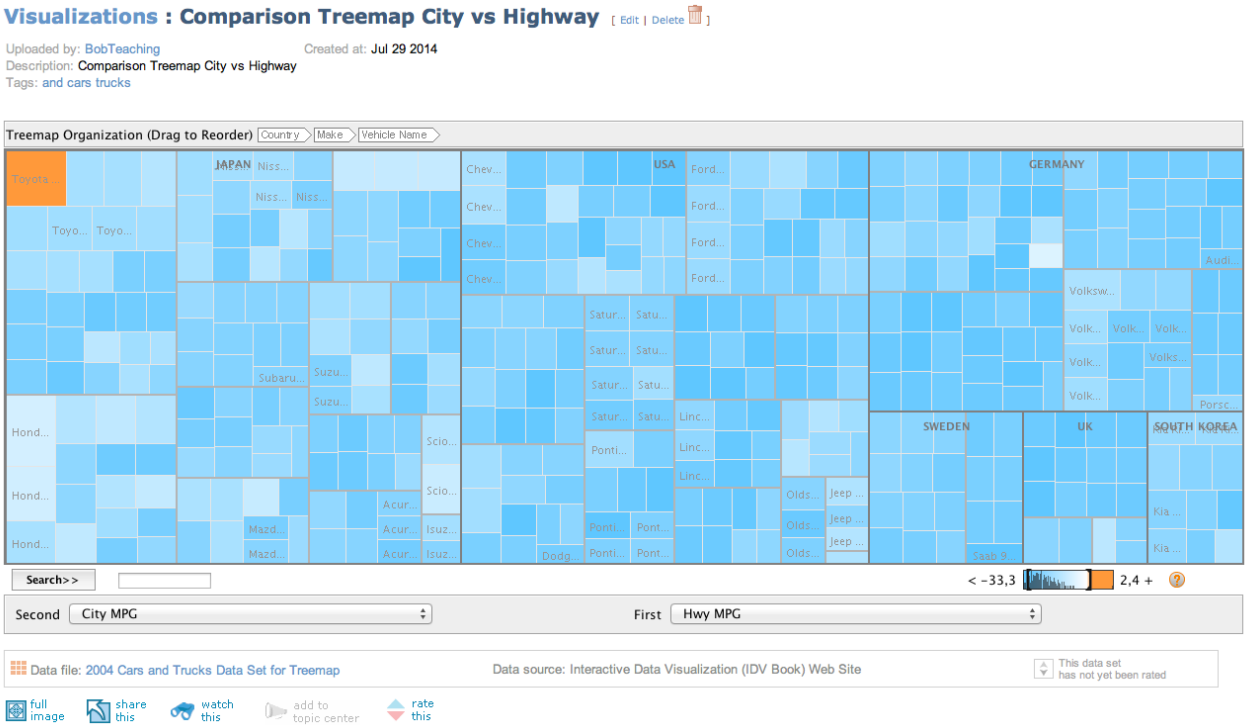

Figure 1: A comparison treemap.

relative merits of the 2D vs 3D visualizations. Aid the discussions with visualization results using both techniques.

You may consider the following aspects of visualization in your analysis (but not limited to):

1. Which tools did you use to create the 3D visualizations?
2. Are there any advantages to using 3D visualization?
3. Are there any disadvantages to using 3D visualization (as opposed to 2D) ?
4. Which type of visualization do you think is more intuitive?

4 Useful Links to Visualization Tools

You may consider using the following visualization tools. Links to these tools (and more) can be found on the module web page.

1. Data Driven Documents:
   `http://d3js.org/`
2. WEAVE:
   `https://oicweave.org/`
3. Scaffold Hunter:
   `http://scaffoldhunter.sourceforge.net/`
4. IBM Many Eyes:
   `http://www-01.ibm.com/software/analytics/many-eyes/`
5. Mondrian:
   `http://www.theusrus.de/Mondrian/`
6. XMDV:
   `http://davis.wpi.edu/xmdv/`
7. Topcat:
   `http://www.star.bris.ac.uk/~mbt/topcat/`
8. Tableau:
   `http://www.tableausoftware.com/`
   (free trial available)

Look carefully at the selection of information visualization tools provided on the module web page for tools with the capability to create 3D visualizations. Any of the tools listed on the module web page may be used. Tools not listed there may also be used. A paper called, *Visual Analysis of Document Triage Data* [2] 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 a module teaching assistant.

5 Submission

You are required to submit a document, which contains:

Your name, a summary (0.5-1 A4) about your approach, and tools that you have used and the insight your visualization provide. If the data has been modified in order to create your visualizations, please describe the changes that were

made. Please also indicate the number of hours spent on this part of the assignment for help us to calibrate the difficulty levels in future assignments. Show five visualizations, each of which is accompanied by a caption with fewer than 250 words. You may use the caption to indicate what type of visualization (e.g., parallel coordinates, Treemap), and highlight the important insight depicted in the visualization. *Make sure each of your visualization types are distinct. In other words, two Treemap visualizations are two instances of one type of visualization.*

You may submit more than five visualizations, but you must select five to be marked as part of this required level.

You are required to submit a *printed copy* of your report by the deadline and upload your report via blackboard and submit PDF format is strongly encouraged. Open Office Document Format (.odt) is also good. Word document format (.doc) is not acceptable.

References

[1] K. Andrews and M. Lessacher. Liquid Diagrams: Information Visualisation Gadgets. In *Proceedings of the 14th International Conference on Information Visualisation (IV'10)*, pages 104–109. IEEE Computer Society Press, July 2010.

[2] Z. Geng, R.S. Laramee, F. Loizides, and G. Buchanan. Visual Analysis of Document Triage Data. In *International Conference on Information Visualization Theory and Applications (IVAPP)*, pages 151–163, 2011. (available online).