

# Independent Study Proposal

Noah Dunn

February 2020

## 1 Data Visualization with Processing

**Data Visualization** is a catch-all term for using data in any capacity that is represented by some form of a picture, video, or another type of media. It takes hold in the form of graphs or charts in some publications, in others, it may be screenshots or complex diagrams. Data visualization is used to communicate to the reader an idea that may be complicated to understand in plain text. It is also used to give the reader a distinct mental model as they process a body of text, providing a nice break to a long stream of text.

**Processing** is a programming language and IDE wrapped up in one convenient package. Processing was created as a way to do Java graphics without needing as much overhead or user-written code to achieve very simple visuals at a quick rate. Some modern projects have adopted Processing as a means of generating advanced data visualization models, particularly for use in presentations, videos, and slideshows to gain a better understanding of large data sets. As a result, data visualization employing Processing has become a powerful tool for the modern researcher to present information they have discovered.

## 2 Topic Scope

Processing has a very large toolbox for a language that focuses on simple implementation. In my independent study, I intend to avoid all forms of 3D visualization in pursuit of 2D visualization. My focus in the study is not the language of Processing, but understanding the models that have been built with it already. However, given how Processing streamlines a lot of common data visualization operations in its language, I think it will be beneficial to point these out as they arise. In particular, I wish to continue work on visualization for my current taxi-cab research project, showing taxi-cab routes over time, as well as path lengths and areas of highest taxi usage.

One of the potential projects I will be working on for my thesis is the idea of using causal inference to predict medical related trends. Causal inference is a very new computer science and statistics field that is heavily dependent on mathematical graphs. Given that these graphs are very data driven and visual in nature. I believe that, should I take on this path as my primary research, there

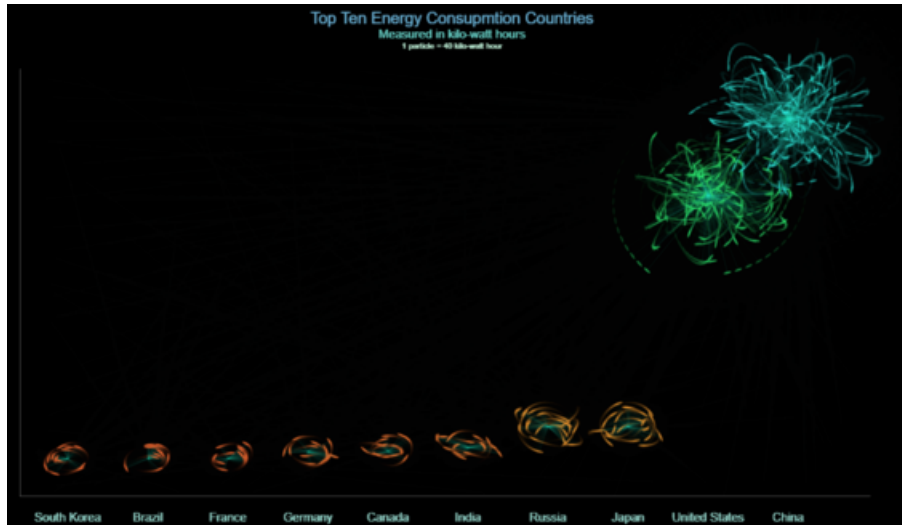


Figure 1: Sample Data Visualization in Processing [2]

may be some value in looking into how to show causal inference relationships on the data visualization front, again in 2D

### 3 A List of Resources

Processing was built in 2001 by Casey Reas and Ben Fry. Later in 2012, Daniel Shiffman joined as a lead developer.[3] The following list provides some of the resources I have already explored and intend to explore in dealing with processing and data visualization.

- Daniel Shiffman's Youtube Channel: *The Coding Train* [8]
- The Processing Foundation's Online Tutorials on Data Processing [1]
- An Online Course from "Lynda.com by LinkedIn" [6]
- Ben Fry's book on Data Visualization [4]
- Data Visualization Projects provided by Open Processing [7]
- Google's current Visualization research projects [5]

### 4 Justification of Appropriateness

Although my end research topic is still up in the air at this point, I think Data Visualization is essential in some form in the majority of research projects in Computer Science. Data is likely going to be used or presented somewhere

along the line, and the reader must have a firm understanding of what that data represents.

Processing is a relatively new technology, but thanks to the success of a Daniel Shiffman's YouTube channel, the community is very friendly to learners of all education backgrounds. Processing offers developers a very unique experience in that they can create large, complex animations and graphical compositions with very little actual code. Processing takes Java from a very complicated graphics engine to a simplified powerhouse. With all of this given, it makes sense to learn Processing for quick and easy data visualization that will be useful in any data driven research.

## 5 Potential Outcomes

The end goal is to become as close to an expert in processing as I can in my independent study. I am familiar with Processing on a surface level, but I would like to attain mastery of its graphics capabilities to produce stellar data visualization models and images. The obvious take away from gaining knowledge in this field is to be able to spin up temporary data visualization models as quickly as possible to understand more about a data-set before I begin working with it.

My more advanced goals revolve around being able to understand what kinds of data visualization are preferred and to produce one that is both accurate and visually appealing. I also believe that gaining proficiency in this technology would enable me to teach future supervisors and colleagues some techniques in order to help out data visualization in their research and papers. I also am taking a job in a field that is heavily dependent on having management make important decisions based on the available data. I believe this skill will offer a great benefit to my future work life in addition to academia.

## References

- [1] Processing: Data, 2020.
- [2] Jem Brown. Top ten energy consuming countries, 2017.
- [3] The Processing Foundation. Processing: People, 2020.
- [4] Ben Fry. *Visualizing Data*. 1st edition, 2008.
- [5] Google. Big picture group, 2020.
- [6] Lynda.com From LinkedIn. Processing: Interactive data visualization, 2020.
- [7] Open Processing. Sketches received hearts during anytime, 2020.
- [8] Daniel Shiffman. The Coding Train, 2020.