Spring 3-27-2019

How Does the Visualization of Data Change how it is Interpreted?

Alexis VanderWilt
Dakota State University

Cherie Noteboom
Dakota State University

How Does the Visualization of Data Change how it is Interpreted?

Alexis VanderWilt, Undergrad Researcher
Dr Cherie Noteboom, Research Mentor
DSU DAKOTA STATE

Abstract

Big data is one of the most promising trends in technology and business today. Big data refers to extremely large data sets that may be analyzed computationally to reveal patterns, trends, and associations, especially relating to human behavior and interactions. Big data sets hold valuable information with the potential to improve efficiency in the workplace by giving us insight into the areas. How can we extract information from data? Visualizations and aggregations are frequently used tools to represent data in a manageable way. The construction of these tools requires usage of design principles to leverage human ability to translate data into knowledge that can be used to support correct decisions. Our project creates and executes a survey to discover whether participants vary in their ability to draw conclusions from open source toxicity data on Lake Michigan presented in advanced excel or tableau formats. Based on our results, we will suggest design principles for visualizations that improve the ability to comprehend data quickly.

Literature Review

As a society, we are consuming and creating more data than ever before. According to Singh, our society is creating upwards of 2.5 quintillion bytes of data each day (Sachchidanand Singh, 2012). This is making it more and more important that we have a way to analyze, comprehend, and use the data that is available (Sachchidanand Singh, 2012).

Raw data is defined as 'data that has not been processed for use' and information is defined as raw data that has been processed for use from a study done by Chaim Zins (Zims 2007). The most important thing to remember with visualization is that when designing anything it's imperative that mental road blocks are removed in order to make the data more easily interpreted (P. Keller 1993). A study by Chao Gong found that certain colors can provoke certain emotions (Chao 2009). The cool colors such as blue, green, and purple tend to cause more calm and relaxed emotions. The warmer colors like red, orange, and yellow create more of an excited feeling (Chao 2009).
We wanted to see if certain types of graphs impacted the way that users understand the data. Without proper visualization methods, important information may be disregarded. Understanding how humans translate data to knowledge and utilizing that information to design graphs that use proper colors and spacing to emphasize important information could lead to quicker analysis time, decreased human error in interpreting data, and an increase in our understanding of the immense amounts of data that are created each day.

Methodology

How Does the Visualization of Data Change how it is Interpreted?

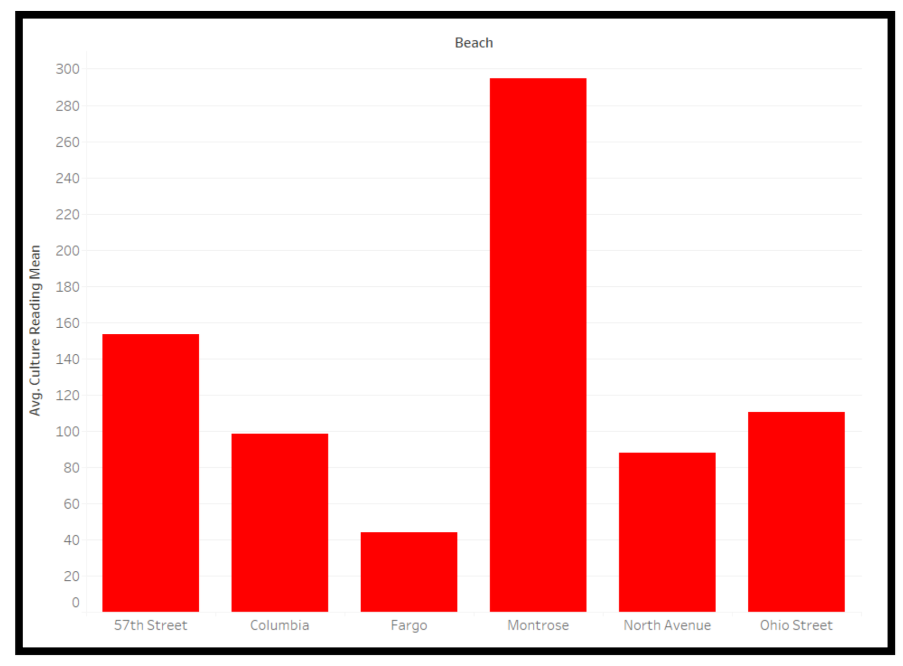
We used a survey methodology to investigate our research question "How Does the Visualization of Data Change how it is Interpreted?". Participants will receive an email with a link to a survey monkey form. They will have the option to participate or may opt out of participation. Those who do participate will answer a 5-7 minute survey with multiple choice questions evaluating the data presentations. Demographic information of college, gender and age will be collected, but none is directly personally identifiable.

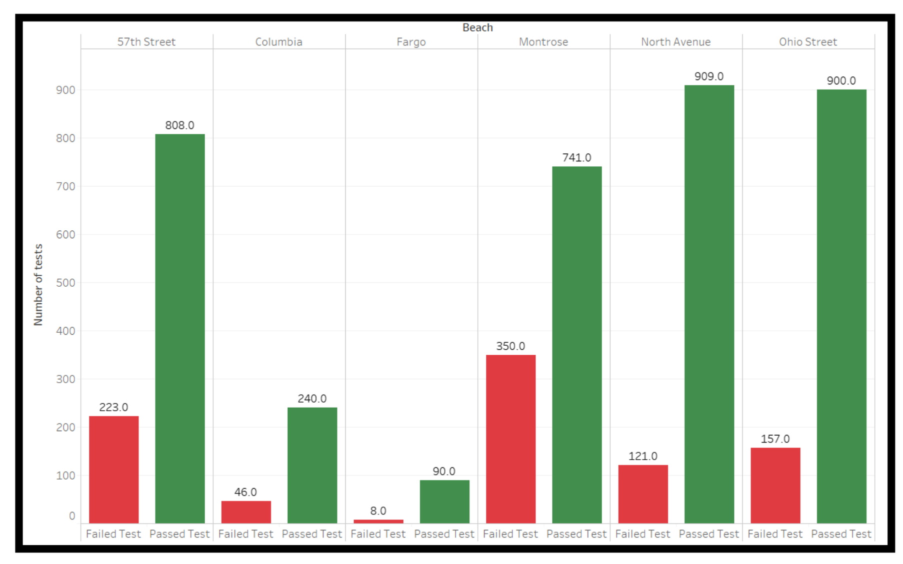
Case Scenario

The city of Chicago, Illinois rests on the edge of Lake Michigan. Any surface water, such as Lake Michigan, has the potential to collect Coliform bacteria such as Escherichia coli, also known as E. coli. Coliform bacteria are bacteria that are naturally found in the intestines of warm-blooded animals (Bacteria in Surface Waters, 2011). These bacteria can find their way into surface waters by way of wastewater treatment plants, failing septic systems, wild animal waste, and storm water runoff (Bacteria in Surface Waters, 2011). While these types of bacteria are typically not dangerous in small doses, it is recommended that swimming beaches should never have more than 88 colony forming units (CFU) per 100 mL in one sample (Bacteria in Surface Waters, 2011). The data was provided by the City of Chicago (Beach Lab Data, 2018).

Initial analysis of the data shows the E. coli levels at each beach to provide insight on safety levels related to disease-causing bacteria and viruses in the water at each beach.

Survey Visualizations

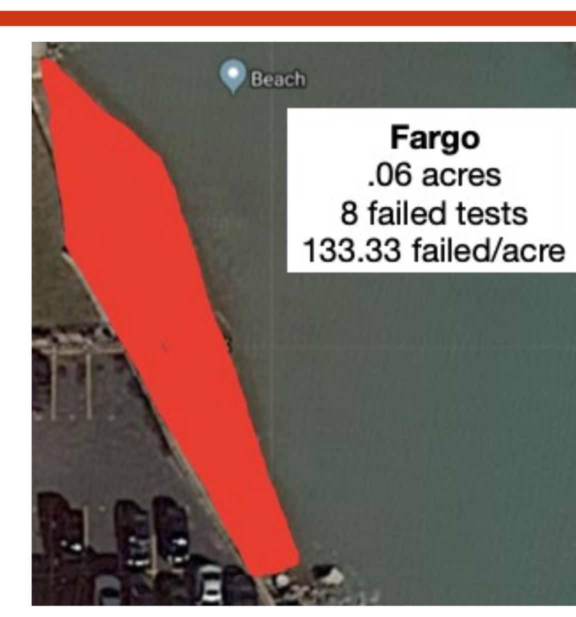




The first visualization is a simple bar graph with bars representing the average number of E.Coli per beach.

The second visualization is a stacked bar chart, with two bars per beach representing the number of E.Coli samples that exceed and did not exceed the recommended value (88/ml).

The final visualization is a set of density maps of colors based on the number of failed tests per acre of beachfront.



Results

Degree Programs		Age		Gender		Beach	Bar Chart	Stacked Bar Chart	Density Map
Undergraduate	136	<18	0	Male	88	57th St	0	0	7
		18-24	132			Columbia	0	0	0
		25-34	3	Female	45	Fargo	0	3	45
		35-44	0			Montrose	122	5	47
		45-54	1	Other	1	North Ave.	1	39	5
		55-64	0			Ohio St.	1	79	30
		65+	0			Insufficient	10	8	
Graduate	115	<18	0	Male	84	57th St	0	0	1
		18-24	9			Columbia	0	0	0
		25-34	52	Female	29	Fargo	0	1	36
		35-44	25			Montrose	96	0	32
		45-54	16	Other	0	North Ave.	0	16	4
		55-64	9			Ohio St.	0	76	28
		65+	1			Insufficient	11	11	

Conclusions

We are in the data collection phase of our project. In the future, we will continue data analysis and determine whether the data identifies significant differences in student assessment of the data visualizations by age, gender or type of student.

Expected Contributions

Our contribution to the discipline will be expanding on the ideas of data visualization and knowledge activation. This project provides a mentoring opportunity to expose me to research and further my education in this field. The project is also a great opportunity for Dr. Cherie Noteboom to provide a mentoring opportunity for an undergraduate Computer Science student. This also increases Undergraduate Student Research activity at DSU.

References

2011. Bacteria in Surface Waters. Concord: New Hampshire Department of Enviornmental Services. https://www.epa.gov/beach-tech/nationwide-bacteria-standards-protect-swimmers-beaches.
2018. Beach Lab Data. November 19. https://data.cityofchicago.org/Parks-Recreation/Beach-Lab-Data/2ivx-z93u.
Chao, Gong. 2009. "Human-Machine Interface: Design Principles of Visual Information in Human-Machine Interface Design." International Conference on Intelligent Human-Machine Systems and Cybernetics. Hangzhou, Zhejiang, China: Institute of Electrical and Electronics Engineers. 262-265.
P. Keller, M. Keller, S. Markel et al. 1993. "Visual Cues: Practical Data Visualization World Wide Web." Computers in Physics 297-298.
Sachchidanand Singh, Nirmala Singh. 2012. "Big Data Analytics." 2012 International Conference on Communication, Information & Computing Technology (ICCICT),. Mumbai, India: Institute of Electrical and Electronics Engineers. 1-4.
Zims, Chaim. 2007. "Conceptual Approaches for Defining Data, Information, and Knowledge." Journal of the American Society for Information Science and Technology 479-493.