**Option 2: Is it better to use a bar chart or a line chart when you have a time dimension? What are the pros/cons of each?**

One of the first things that I learned in this program was the general concept of data interpretation. It is clearly oversimplified that data is only valuable if you know how to visualize it and give context. That means whenever you have data you want to visualize, make sure you use the right charts. While your data might work with multiple chart types, it’s up to you to select the one that ensures your message is clear and accurate. Choosing the right chart is not easy. Depending on the analysis of the data, the charts that you choose have to make sense and relate to that analysis. There are four main types of charts:

* Comparison charts are used to compare one or more datasets. They can compare items or show differences over time.
* Relationship charts are used to show a connection or correlation between two or more variables.
* Composition charts are used to display parts of a whole and change over time.
* Distribution charts are used to show how variables are distributed over time, helping identify outliers and trends.

For this discussion, I will focus on comparison charts because I will define and differentiate bar charts and line charts. I will also present the advantages and disadvantages of each.

1. Bar graph may be the easiest type to graphs to create and understand. They are primarily used to compare different items and to help avoiding clutter when you have several items to compare. It’s also a good choice when you want to compare things between different groups. You could use a bar graph if you want to track change over time as long as the changes are significant. Although the graphs can be plotted vertically or horizontally, the most usual type of bar graph is vertical.

* Advantages of bar graphs:
  + Just like I mention above, bar graphs are very easy to create, read, and interpret.
  + Bar graphs are used more than other graphs because they are easy to be adapted to many different functions. They are very flexible; meaning that they can be used in several concepts.
  + Another aspect of the flexibility of bar graphs is that they are very easy to be edited. That means they give the author the opportunity to add data labels at ends of bars
  + Finally, there is a lot of room for longer text labels when using bar graphs.
* Disadvantages of bar graphs:
  + One thing that I don’t like about bar graph is that they are not too friendly when it comes to big data. They become cluttered with too many categories.
  + Clustered bar charts are harder to read as data series are added.

1. Just like bar graphs, line charts can be used to show relationships within a continuous data set, and can be applied to a wide variety of categories, including daily number of visitors to a site or variations in stock prices. They are also a good way to compare the trend against previous periods to add further context particularly when reviewing seasonal trends. A line graph showing two trend lines with revenue data from the current and preceding years allows a very simple way to compare performance against previous years. They can also be used to display several dependent variables against one independent variable.

* Advantages of line graphs:
  + Line chart is that it is very easy to understand and make.
  + They are used to handle large amount of data.
* Disadvantages of line graphs:
  + One of the disadvantages include not being able to use too many lines as it makes the line chart and the information on it cluttered and hard to understand.
  + When comparing data sets, line graphs are only useful if the axes follow the same scales. Some experts recommend no more than 4 lines on a single graph; any more than that and it becomes difficult to interpret.

References:

* <https://sciencing.com/advantages-disadvantages-using-graphs-math-8315872.html>
* <https://www.betterevaluation.org/en/evaluation-options/LineGraph>
* <http://cahare.sh-original-media.com/disadvantages-of-bar-charts/>

**Option 4: Pick one of the model visuals in Chapter 6 of Storytelling with Data and provide an analysis of what you see first. Provide your interpretation of the visual, what stands out to you, what you would do differently, what you think worked well, etc.**

It is very clear to everyone by now that I always go back to basic concepts to analyze or to define my answers. I am doing the same thing for this paper. In statistics, standardization is the process of putting different variables on the same scale. This process allows you to compare scores between different types of variables. Typically, to standardize variables, you calculate the mean and standard deviation for a variable. Then, for each observed value of the variable, you subtract the mean and divide by the standard deviation. This process produces standard scores that represent the number of standard deviations above or below the mean that a specific observation falls. For instance, a standardized value of 2 indicates that the observation falls 2 standard deviations above the mean. This interpretation is true regardless of the type of variable that you standardize.

The main goals and objectives of data standardization are to:

* Implement data administration in ways that provide clear, consistent, unambiguous, and easily accessible data throughout an organization.
* Standardize and register data elements to meet the requirements for data sharing and interoperability among information systems throughout an organization.
* Promote standardization in the organization, consistent with requirements for sharing data among business owners, with other Federal agencies, and with private and other Governmental communities.
* Minimize or eliminate the cost and time required to transform, translate, or research redundant data elements across various systems.
* Incorporate applicable standards related to the industry of that organization before creating standards or using common commercial best practices.
* Provide steps for the development and/or modification of newly-identified data and, provide towards the required level for correctness, consistency, and completeness of these data elements as represented in their respective data models.
* Provide workflows and instructions for the implementation of data standards.
* Reuse of data found in the organization data dictionaries and data models.

The full benefits of data standardization will only be achieved if organizations use the same data element definitions and if those definitions are available for all business partners to search, retrieve, and use for file transfer specification development. At global or domestic organization, data standardization is supported through the Enterprise Data Dictionary. These tools support seamless integration of data between the organization systems. These tools provide benefits and support such as, but not limited to:

* Data Sharing: Facilitates data sharing within and between the organization and business owners and its partners where allowed, and where mutually agreed upon.
* Reliable Data: Enterprise Data Standardization will allow the organization personnel to be more productive and self-sufficient in their jobs by knowing where information is located, what it looks like, and its meaning. This will facilitate the reduction in occurrences of multiple systems capturing the same data, and displaying different answers to the same question.
* Commonly Defined Data: Definitions from individual systems will be replaced with Enterprise wide data definitions. Application interoperability depends on standardization, and standardization of data is only possible when data is defined and understood.
* Integration of Operations: The support of integrated operations among lines of business, communities of practice, and the facilitation of decision-making using standard data.

Data standardization plays a significant role, especially in the healthcare industry, in creating a common format of data to allow different institutions to collaborate in an efficient way while using various information processing tools. This helps clinics, hospitals, and other medical organizations perform large-scale analytics, mutual research, and share effective methodologies. Here is a clear example that I can share. Advent Health just bought Florida Hospital, which had been in Central Florida since 1908. Advent Health clearly understand that creating a good impression in the Central Florida community is very important. They understand that healthcare data standards are the commonly adopted rules regarding definitions and formats of health information, and the methods of recording, storing, and sharing data within medical organizations. At a basic level, data standardization is aimed at defining what information can and should be collected, determining a way for this information to be represented, and deciding how to encode it for further transmission.

The standards involve the following types of data:

* Medications;
* Medical devices;
* Medical records;
* Radiological images;
* Medical monitoring systems;
* Payment information; and
* Reimbursement information.

Healthcare data standards apply to both computer-based and paper-based healthcare data systems. They involve the following components:

* Data elements;
* Interchange formats;
* Terminologies; and
* Representation standards.

References:

Wu, Z., Irizarry, R., Gentleman, R., Martinez Murillo, F. Spencer, F. A Model Based Background Adjustment for Oligonucleotide Expression Arrays. Journal of American Statistical Association 99, 909-917 (2004)

Bolstad, B. M., Irizarry, R. A., Astrand, M., and Speed, T. P., A comparison of normalization methods for high density oligonucleotide array data based on variance and bias, Bioinformatics, 19, 185 (2003).

Yang, Y. H., Buckley, M. J., Dudoit, S., and Speed, T. P. (2002). Comparison of methods for image analysis on cDNA microarray data. Journal of Computational and Graphical Statistics, 11 (1), 108-136.

Hello David,

The second part of the term project give me the opportunity to see how important it is to use the appropriate charts to catch an audience’s attention. Every single business in the world depends on clear and easy ways to comprehend data analysis. Based on chart data graphs, business leaders can demonstrate how to play or expect production teams and buy deep into data to determine sales patterns. Data may indicate that animal care products are falling, but high-income consumers expect most sales. These statistics can be used to improve support for this portion of the subscriber to increase conversion rates and revenue growth for this category. Another important advantage of data quality is that we do not only pay data graphical databases, but also allow you to modify the form, remove unnecessary and continue with the details. This is a great eye and attracts attention to business operations better and provides better communication. It also offers many multidisciplinary methods for presenting data.

I think you did a great job with your post. Researchers agree that vision is our dominant sense: 80–85% of information we perceive, learn or process is mediated through vision. It is even more so when we are trying to understand and interpret data or when we are looking for relationships among hundreds or thousands of variables to determine their relative importance. One of the most effective ways to discern important relationships is through advanced analysis and easy-to-understand visualizations. Data visualization is applied in practically every field of knowledge. Scientists in various disciplines use computer techniques to model complex events and visualize phenomena that cannot be observed directly, such as weather patterns, medical conditions or mathematical relationships. Data visualization provides an important suite of tools and techniques for gaining a qualitative understanding. The two basic techniques that I believe are still the best way to illustrate data are:

* Line Plot: the simplest technique, a line plot is used to plot the relationship or dependence of one variable on another. To plot the relationship between the two variables, we can simply call the plot function.
* Bar Chart: Bar charts are used for comparing the quantities of different categories or groups. Values of a category are represented with the help of bars and they can be configured with vertical or horizontal bars, with the length or height of each bar representing the value.

Hello Lenin,

I am happy that you have chosen student loan as your story telling. I actually wish that I chose that topic. I just want to mention the reality of what is really behind the student loan crisis. There are several factors that have contributed to the student loan debt crisis in the U.S., beginning with rising tuition prices.For the 2017-18 academic year, the average cost of tuition, fees, and room and board at a public four-year university totaled $25,620 for out-of-state students. The cost climbed to $34,740 for students at private four-year universities. 529 college savings plans can help with paying college expenses but only 13% of families use them, according to Sallie Mae's 2017 How America Pays for College Report. Instead, 42% of families borrow to pay for college, including loans taken out by both students and parents.

The lure of loan forgiveness may also be seen as a contributing factor. The federal Public Service Loan Forgiveness program offers student loan forgiveness for grads who pursue a career in public service. That's a tempting prospect, which may lead students to lean on loans more heavily, with the expectation that they'll be forgiven later. But, the program isn't permanent and could be altered or canceled, making it risky for borrowers to count on loan forgiveness. The student loan debt crisis is compounded by the number of borrowers falling delinquent on their loans. As of the first quarter of 2018, 11% of student loan borrowers were 90 days or more delinquent or in default on their loans. While that's a slight decline from the previous quarter, it suggests that a substantial number of borrowers are struggling to keep up with their loan payments. I have one question for everyone who actually ready my reply. Would you rather pay more taxes in order to have free education?

Hello Paul,

Your discussion about the differences between the donut and pie chart is very impressive. I want to comment a little more on the visualization analysis instead of writing most of the things that you already mention in your post. Now whenever we talk about graph analysis, we also refer to graph algorithms or graph analytics. Graph analytics are analytic tools used to determine strength and direction of relationships between objects in a graph. The focus of graph analytics is on pairwise relationship between two objects at a time and structural characteristics of the graph as a whole.

Graph Algorithms or Graph Analytics are used in a number of applications.

* + Clustering - the grouping of objects based on their characteristics such that there is high intra- cluster similarity and low inter-cluster similarity. Applications include machine learning, data mining, statistics, image processing, and numerous physical and social science applications.
  + Cutting or Partitioning – To find the cut with the fewest number of crossing edges. Applications include finding weak spots in data and communications networks, and community detection in social networks.
  + Search – Breadth first search and Depth first search.
  + Shortest path – To find the shortest path between two nodes of interest. Applications include social network analysis, transportation logistics and many other optimization problems.
  + Widest path – To find a path between two designated vertices in a weighted graph, maximizing the weight of the minimum-weight edge in the path. Applications include IP traffic routing and traffic-sensitive path planning.
  + Connected components – A strongly connected graph is one where you can get to every node in the graph from any starting node. The strongly connected components are the maximal sub- regions of a graph for which each sub-region is strongly connected. Applications include social network analysis.

References:

https://www.pyramidanalytics.com/data-visualization-examples-in-pyramid?subid=0

<https://www.sas.com/en_us/insights/big-data/data-visualization.html>

Hello William,

My high school adopted an international baccalaureate program that allow us to attend any university around the world upon completing the program. I was in 10th grade when I was introduced to a concept called “Map Mind” that help us studying for the baccalaureate exams. Map mind was just a normally outline that helped the student connecting any missing subtopics that can be easy to forget. After reading and understand tree graphs, I notice that it is the same general concepts that I had learned in 10th grade. Normal graphs evolved from the field of mathematics. They are primarily used to describe a model that shows the route from one location to another location. A graph consists of a set of nodes and a set of edges. An edge is a pair of nodes that are connected. A path is the term used to describe traveling between nodes that share an edge.

A tree data structure, like a graph, is a collection of nodes. There is a root node. The node can then have children nodes. The children nodes can have their own children nodes called grandchildren nodes. This repeats until all data is represented in the tree data structure. A tree is a graph that has no cycles. That means it is a graph that starts and ends at the same vertex. A child node can only have one parent. For this reason trees are not a recursive data structure. In computer programming, trees are used all the time to define data structures. They are also used as the basis for algorithms to solve problems. The most common implementations of a graph are finding a path between two nodes, finding the shortest path from one node to another and finding the shortest path that visits all nodes.