HARIKA YENUGA

114967-MIS-658-NW2 Data Mgt, Orgztn & Visualizatn Graduate Midland Fall 2024-2025

TUFTE CASE STUDY REPORT

Dr. Dana McCann

11/17/2024

Introduction:

Data visualization is all about presenting information in a way that's easy to understand through graphs and charts. However, sometimes these visualizations can become confusing or misleading due to poor design choices. Edward Tufte, an expert in the field, created five key principles to help guide how we should design graphs to make them clear, honest, and easy to interpret.

In this report, I will be discussing the four different types of graphs and evaluating them based on Tufte's principles. These principles will help us understand how simple, clear designs make it easier to understand data, while complex or unnecessary decorations can make it harder. By analyzing these graphs, I'll explain how we can improve them to follow Tufte's principles, ensuring that the data is shown in the best way possible.

Clarity and Simplicity
He stresses the importance of clear and simple designs that allow the data to speak for itself.

Integration of Art and Data
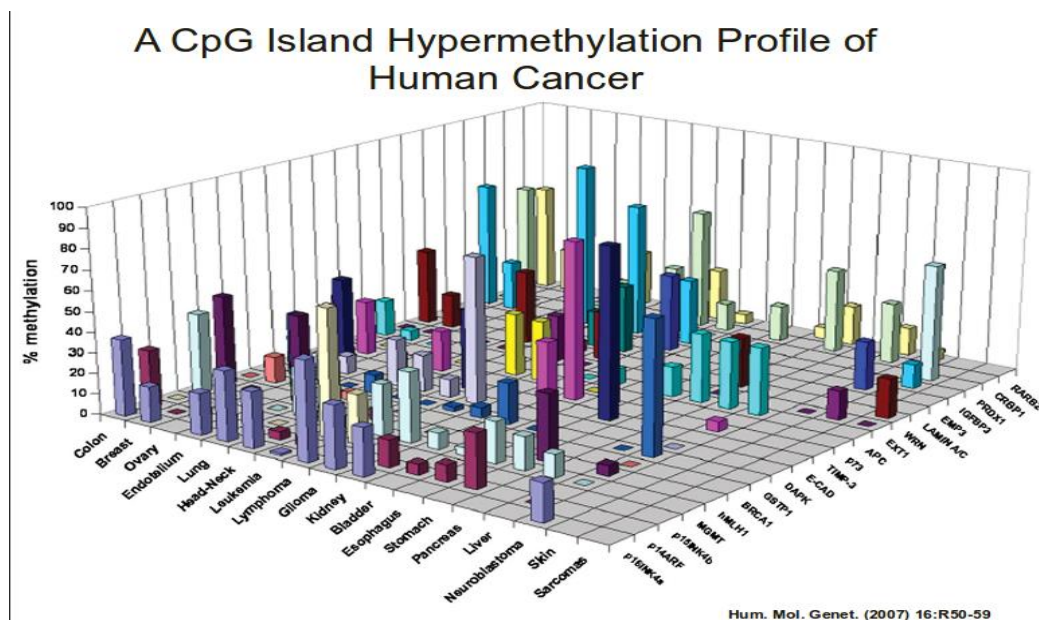Tufte views data visualization as a form of art.

Influence on Modern Data Visualization
His principles have been integrated into modern data visualization tools and practices.

Shashika. (2024, June 16) Source

TUFTE PRINCIPLES:

<u>Graphic 1: CpG Island Hypermethylation Profile in Human Cancer</u>

Dunford, C. (2023, April 13) Source

This chart shows how a chemical process called "methylation" affects different genes in various types of cancer. It uses a fancy 3D bar chart to present this information. While the chart looks attractive, it's hard to read and understand, making it unsuitable for scientific or serious presentations.

Principle 1 – Graphical Integrity:

So, looking at this 3D bar chart, I've noticed that it's hard to trust the data whereas, This chart uses a 3D design, but this effect distorts the data. It's really harder for me to compare the bars accurately because the 3D effect make some bars look bigger or smaller than they really are. For example, in some marketing ads, how they use 3D charts to make something looks more important than it actually is.

Principle 2 – Data Ink:

This 3D effect in this chart adds extra design elements like shadows, which took more space that could be used for the data itself. That space made the chart look more complex than it needs to be. Think of it like a piece of art where the decoration takes attention away from the main subject. A simpler design like a clean 2D bar chart would help the data stand out more clearly.

Principle 3 – Chartjunk:

The extra 3D effects, shadows, and gradients are examples of "chartjunk," which is just clutter that doesn't help you understand the data. If you look at graphs in sources like the Cancer Genome Atlas, they don't have these unnecessary features. Instead, they focus on presenting the data in a simple and clear way.

Principle 4 – Data Density:

In this chart, the 3D design takes up a lot of space where that could have been used to show more accurate data in the chart. It's more like trying to fit a lot of information into a Small room crowded with people everywhere where you can't see anything clearly. A better design would be a 2D bar chart or a heatmap, which would pack in the information in a more compact and readable way.

Principle 5 – Small Multiples:

If this chart were divided into smaller, simpler charts for each cancer type, it would make comparing the data much easier. Think about how in a report like the *Human Epigenome Atlas*, each tissue type has its own simple chart, so you can compare them one by one. That would work better here to make the data clearer and easier to understand.

Better Alternatives:

Venngage. (n.d.). *Vertical bar chart marketing stats template*. Venngage. Source

As per my Understanding, A better approach would be to use a simple 2D bar chart, like the ones you see in most scientific studies, where the above chart helps present marketing performance data clearly, such as revenue, leads, conversions, and costs, allowing users to visualize the information effectively in one place. By automatically generating charts, the template simplifies data representation without unnecessary clutter, adhering to the principle of making the data easily accessible and interpretable.

Graphic 2: Pie Chart With Overlapping Data

This chart below, tries to show how different groups overlap but makes a big mistake by summing the percentages incorrectly (178% instead of 100%). Pie charts are meant to show parts of a whole, but this one misleads the viewer.

u/jan_fon (2019, March 7) Source

Principle 1 – Graphical Integrity:

This pie chart isn't accurate because the percentages add up to 178%. In real life, percentages should never go over 100% in a single chart. You might see something like this in an ad trying to make a product look more popular by showing overlapping numbers. To fix this, we could use a Venn diagram or a bar chart where overlapping data is shown in a way that doesn't confuse the viewer.

Principle 2 – Data Ink:

The chart uses bright colors and 3D effects that don't really help you understand the data better. It's like decorating a room with unnecessary items that distract you from the important

things. By simplifying the chart using flat colors and removing the 3D effects it would focus more on the data, making it easier to read.

Principle 3 – Chartjunk:

All the extra effects, like 3D design and bright colors, are "chartjunk." These things distract from the information and make the chart harder to understand. Just like how in some reports, you only see basic pie charts without unnecessary effects, this chart would be much clearer without the extra decoration.

Principle 4 – Data Density:

The overlapping sections in the pie chart make it harder to understand the data clearly. Instead of using a pie chart, we could use a stacked bar chart or Venn diagram to show the relationships between the categories. These alternatives would show the data more clearly without confusion.

Principle 5 – Small Multiples:

This chart could be improved by splitting the data into smaller, separate charts for each category. By doing this, we would have clearer insights into each group without the overlap confusing the viewer. Tufte's principle of small multiples would help here, as each category could be shown individually to improve understanding.

Why It's Problematic ?

Misleading Percentages:

The overlapping sections confuse the rest of the crowd because the percentages don't add up correctly. For example, during the COVID-19 vaccine rollout, overlapping categories (like healthcare workers and elderly) sometimes led to inflated statistics. If data overlaps, a Venn diagram is better suited than a pie chart.

Best Alternative:
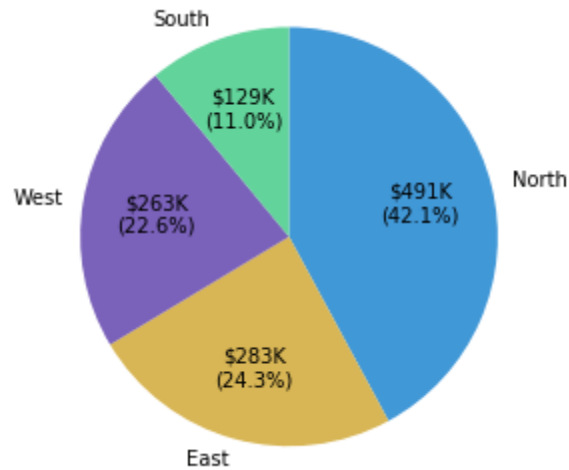
Atlassian. (n.d.).
Source

The Above mentioned Pie chart shows the clear picture of the 100% data which will helps the users to find the information easily. Whereas, Data-ink ratio = data-ink / total ink used to print the graphic

Graphic 3: COVID-19 Death Rates and Lockdown Timeline

A Rainbow-Colored Tornado

The chart below uses a spiral design to show how lockdowns affected death rates from COVID-19. While the design is visually creative, it does not shows any clarity, and making the data difficult to follow.
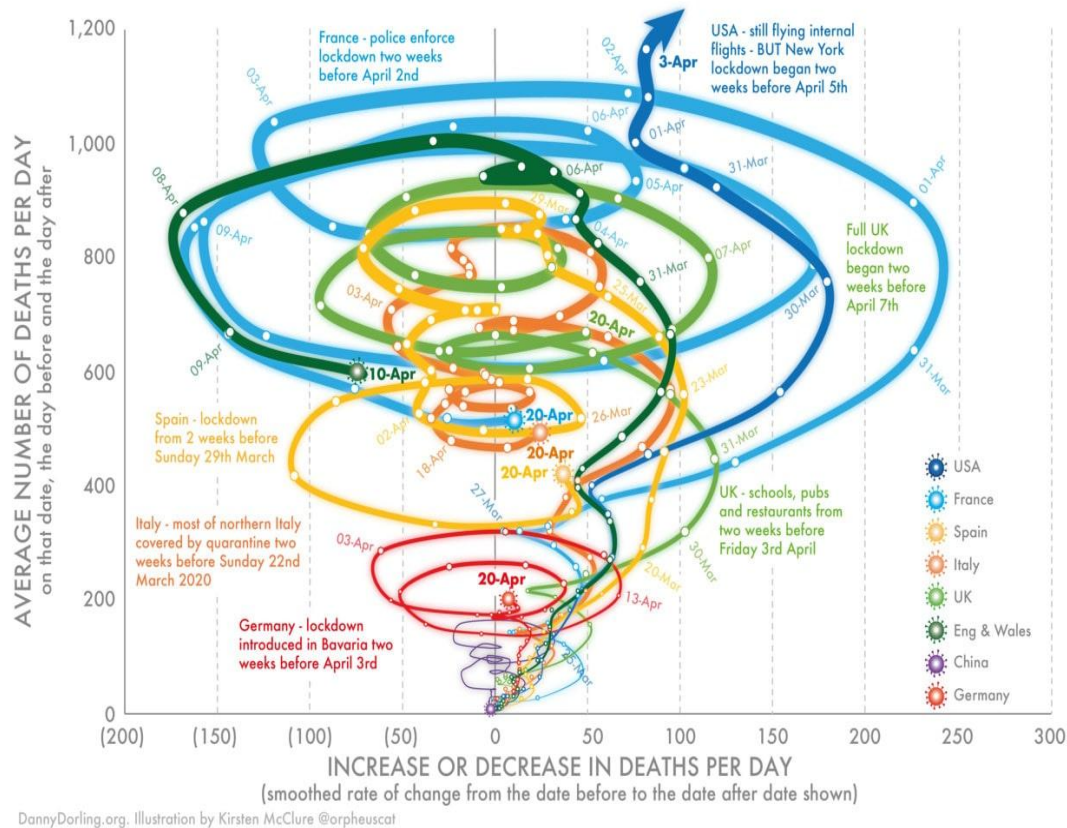
Mitsiakina, J. (2022, September 7) Source

Principle 1 – Graphical Integrity:

The spiral design in this chart makes it hard to compare the death rates between countries because the lines overlap. I feel It's completely like trying to compare two people standing too close to each other where you can't even really see who's taller. A better approach would be to use simple line graphs, like the ones on the Johns Hopkins University's COVID-19 dashboard, where each country's data is displayed clearly and separately.

Principle 2 – Data Ink:

This spiral chart is full of unnecessary effects, like thick lines and multiple colors, which make it harder to focus on the actual data. If you look at Our World in Data COVID-19 visualizations, they use clean, simple line graphs to show trends in a clear way without extra effects.

Principle 3 – Chartjunk:

The spiral and overlapping lines are best examples of chartjunk. These visual effects don't make the data clearer they just confuse the viewer. If we use simple line graphs instead of this spiral chart, we can see the trends more clearly, just like how *The New York Times* COVID tracker does it.

Principle 4 – Data Density:

This chart spreads out the data in a way that reduces its density. It tries to fit a lot of information into a spiral, but it gets confusing. A clearer way to show this data would be with separate line charts for each country, allowing us to compare them side by side without clutter.

**Principle 5 – Small Multiples:**

Small multiples would work great here. By showing each country's data in its own chart, we could easily compare trends across countries. Just like the European Centre for Disease Prevention and Control ECDC often does with COVID-19 data, breaking it down into small charts would make the trends much clearer.

Why It's Problematic?

Too Decorative:

The thick lines and bright colors take attention away from the actual data. A cleaner format, such as those used by Our World in Data, focuses on presenting the data without unnecessary distractions.

Better Alternatives:
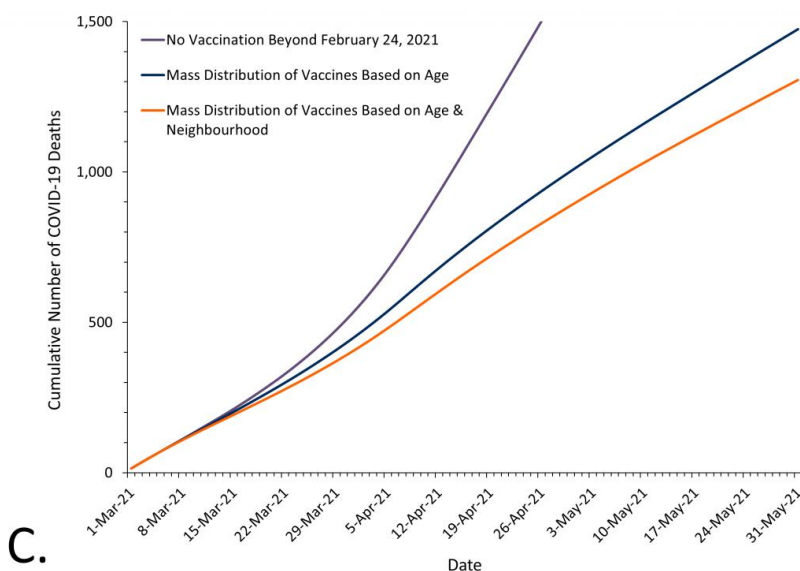


Brown, K. A., Stall, N. M., Joh, E., et al. (2021). Source

A line graph with separate lines for each country would allow people to see clear trends. Breaking the data into small multiples (one graph for each country) would also make it easier to compare trends side by side.

Graphic 4: Job Gains and Losses Over Time (Sectors)

This chart uses colored dots to represent changes in employment across different job sectors. While the idea is good, the execution makes it harder to understand the data.
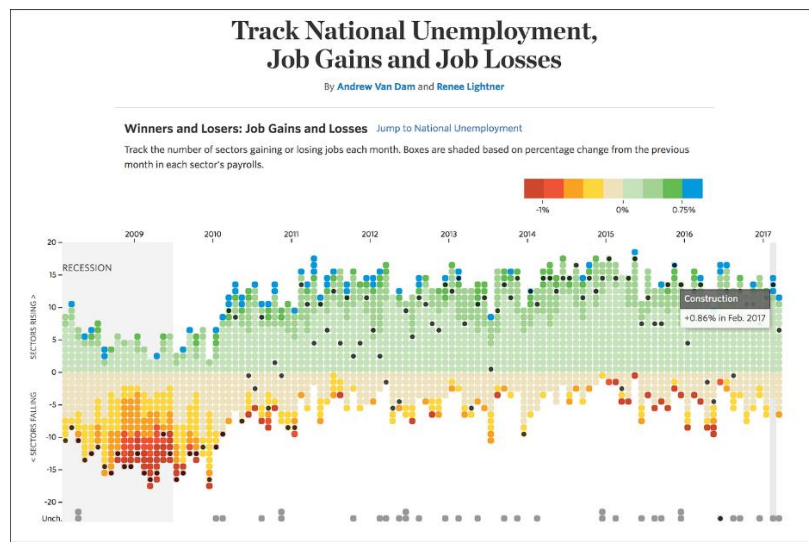
Jovanovic, M. (2024, July 24). Source

Principle 1 – Graphical Integrity:

This chart uses dots with different sizes and gradients, which makes it hard to tell the exact number of job gains or losses. It's like looking at a picture where the details are blurry. A better design would use simple lines or consistent markers to show job changes clearly, like the reports on gender wage gaps from the Organization for Economic Cooperation and Development (OECD), which use clear line graphs to show trends.

Principle 2 – Data Ink:

The gradients and shaded areas in this chart use up space that could be used for actual data. Instead of fancy effects, we could use clean lines or simple markers to represent job changes. This would give us a much clearer picture of the trends, without wasting space on unnecessary design features.

Principle 3 – Chartjunk:

The gradient shading and extra visual effects are chartjunk because they don't help us understand the data any better. By removing these and keeping only the basic information like

simple trend lines we could focus on the job trends without the distraction of extra decoration. This is how organizations like the *World Economic Forum* create clean, effective visuals.
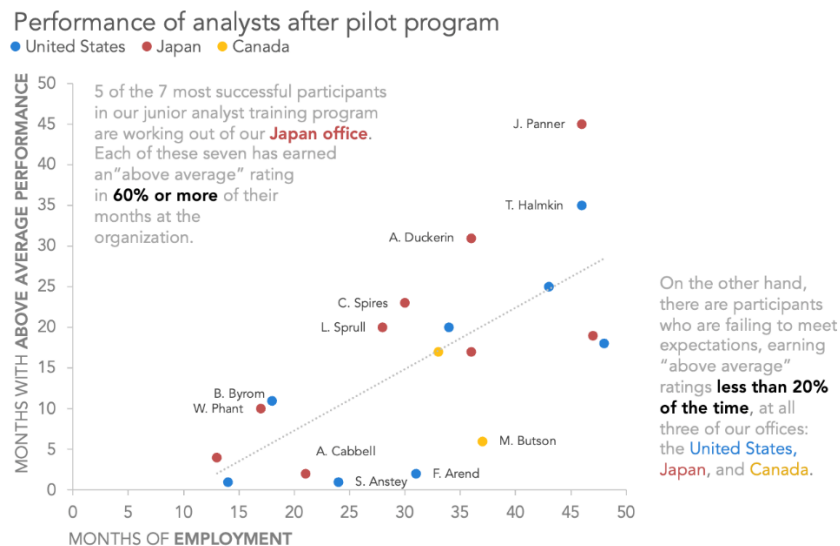
Principle 4 – Data Density:

The chart could show more information in less space if we got rid of the unnecessary gradients and shading. A simpler chart, using lines or consistent markers, would allow us to see job trends more clearly and fit more data into the same amount of space. This would be like how Federal Reserve Economic Data FRED shows economic data in compact, easy-to-read charts.

Principle 5 – Small Multiples:

Using small multiples for each sector would help compare job trends more easily. Instead of one cluttered chart, we could have smaller charts for each sector, making it simpler to understand the changes in each one. This is similar to how The Economist compares data across different regions or sectors using small, clear charts.

Best Alternatives:

Japan has had the most successful participants in our junior analyst training pilot program

Performance of analysts after pilot program
● United States ● Japan ● Canada

5 of the 7 most successful participants in our junior analyst training program are working out of our **Japan office**. Each of these seven has earned an"above average" rating in **60% or more** of their months at the organization.

On the other hand, there are participants who are failing to meet expectations, earning "above average" ratings **less than 20% of the time**, at all three of our offices: the United States, Japan, and Canada.

These Scatter Plot like above, are best alternatives to show the numerical data graphs in terms of analysis. And this scatter plot shows how accurately data can be represented which is a good data for clear understanding.

Conclusion of these principles

In conclusion, Tufte's five principles suggest us that the key to effective data visualization is to make the information understandable, accurate, and clear. When creating charts, it's important to stay clear of complicated designs or superfluous decorations that could detract from the main information. For example, while extra hues or 3D effects may give the chart a more professional look, they often make it more difficult to read the actual data. Instead, using more straightforward forms, such as line graphs or 2D bar charts, guarantees that the attention remains on the data, making it simpler for users to understand the relationships and patterns being displayed.

Any data visualization's ultimate objective is to convey information in the most open and honest manner that is possible. We can ensure that the intention is conveyed properly by eliminating unnecessary aspects, focusing on the important information, and utilizing designs that are straightforward and easy to understand. By following with these principles, a chart's appearance is improved, and its capacity to educate and assist individuals in making better judgments based on the data will also be increased.

References

Tufte, E. R. (2001). The visual display of quantitative information (2nd ed.). Graphics Press.

Mitsiakina, J. (2022, September 7). Misleading data visualization: This is what you should avoid. XB Software. https://xbsoftware.https://xbsoftware.com/blog/misleading-data-visualization/

Dunford, C. (2023, April 13). When data visualization really isn't useful (and when it is). Old Street Solutions. https://www.oldstreetsolutions.com/good-and-bad-data-visualization

u/jan_fon. (2019, March 7). A pie chart out of 178. Reddit: CrappyDesign. https://www.reddit.com/r/CrappyDesign/comments/fpdv53/a_pie_chart_out_of_https://www.reddit.com/r/CrappyDesign/comments/fpdv53/a_pie_chart_out_of_178/?rdt=49964*178/?rdt=4 9964*

Jovanovic, M. (2024, July 24). What is financial data visualization. wpDataTables. https://wpdatatables.com/financial-data-visualization/

The Cancer Genome Atlas (TCGA):

National Cancer Institute. (n.d.). The Cancer Genome Atlas (TCGA). Retrieved November 17, 2024, from https://datacommons.cancer.gov

World Bank Data:

World Bank. (n.d.). World Bank open data. Retrieved November 17, 2024, from https://data.worldbank.org

Our World in Data (COVID-19 visualizations):

Ritchie, H., & Roser, M. (2020). Coronavirus pandemic (COVID-19). Our World in Data. Retrieved November 17, 2024, from https://ourworldindata.org/covid-cases

Federal Reserve Economic Data (FRED):

*Federal Reserve Bank of St. Louis. (n.d.). FRED Economic Data. Retrieved November 17, 2024,
from https://fred.stlouisfed.org*