Data Storytelling

Data storytelling is the process of communicating insights about data using narratives and visualizations.

Data is where it all starts. You may have data that has already been collected and curated or you may have to collect and clean data yourself. Check the modules on Data Collection and Data Cleaning for more information on those processes.

Next comes the narrative, or story. As with any story, the story you tell about data should be compelling and entertaining. Your audience is far more likely to embrace your story when they can identify with it.

Finally, your story should have a strong visual component. Visualizations help bring your story to life.

Let's look at each step in-depth and then some examples of good storytelling.

Data

As detailed in the modules on Data Collection and Data Cleaning, there's more to data than just 1s and 0s. Data must serve a useful purpose. Ensuring that starts with a proper data curation process. If you do not already have data, then collecting and cleaning it are your first steps. As detailed in the Data Collection and Data Cleaning modules, these processes entail several sub-steps. Create a plan to ensure that the data you collect suits your data analysis goals. This plan should be part of a larger data curation process.

There are associated modules for data. It is recommended that you look at those modules for more information if you are going to create your own datasets. If you are starting with datasets from Kaggle or other repositories, then data cleaning is the appropriate next step for you.

If you are collecting your own data, remember to keep it simple and straightforward. The key is to be consistent so that no matter who is collecting the data, you are getting the same variables in the same way. For this route, the data collection module should be your next step.

The Story (Narrative)

Stories revolve around the characters in them. For data storytelling, the data serves the same purpose. When we look at what makes a great story: compelling characters, interesting events, and a "sticky" plot that keeps the audience's attention. Your data story will have the same elements. Your data should be compelling: a vetted source for your data; data that completely encompasses the topic; interesting visuals and a "sticky" dashboard or presentation that keeps the audience's attention.

Characters: the data. When engaging in storytelling, your data is the character that goes on a journey and/or undergoes a transformation. Be sure to explain where your data starts and the setting for your data

storytelling. For example, if your data concerns employment trends, what data are you examining and why are you doing it now? Is there a recession? Does the data cover a specific segment of the working population? Next you'll want to describe the journey (give the Reader's Digest version) and transformations that take place. Did the data get transformed? What was the reason? Was there a catalyst involved? Finally, show the results of the data's journey/transformation. What was the moral of the story? What are the lessons learned?

Just as a novel or other story has a sequence of events, your story should, too. It starts with data and builds to the insights that your data analysis has revealed. Your story should follow that pattern and, if you can, include classic elements of storytelling such as foreshadowing, setting, conflict, and resolution.

Visualizations

Whether you choose to use a dashboard, static presentation, or slideshow your visuals will be the "pictures that are worth a thousand words". They should be clear, concise, and address a KPI or key finding of your analysis. Remember that our data analysis has answered questions such as "what" and "why". Your job is to make that story interesting.

When designing your presentation, keep these maxims in mind:

- Good graphics make for a good story.
- Use software designed for presentations.
- Good graphics take as much screen space as bad graphics.
- Keep your design clean and simple. White space improves displays!

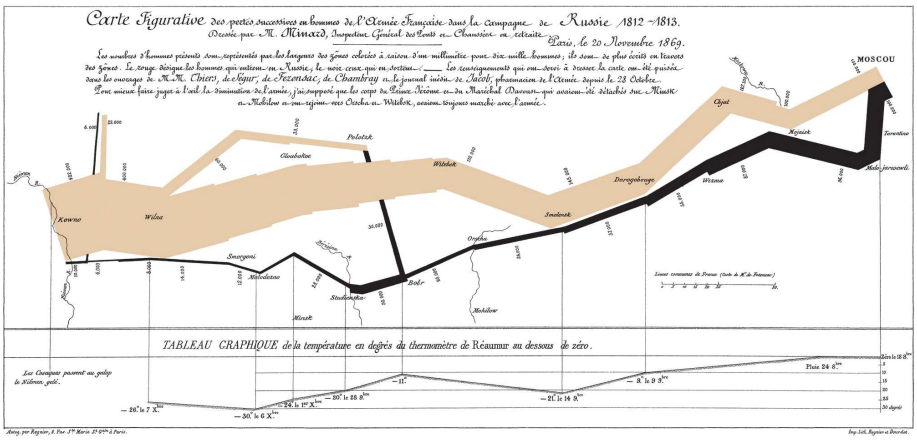
There are several ways to create visualizations to go with your analysis. A common way is to divide the screen into four areas that follow the natural line of sight for the viewer. In our society, we tend to read from left to right and in a clockwise direction. Designing your visualization with that in mind can help the viewers naturally progress through charts and graphs that flow from one to the other.

Another consideration is to remember that the "what", "why", etc. should stand out in the graphics, either by color, shape, animation, etc. It should be easy for the viewer to pick out the most important part of your graphs and charts.

Your story should be the truth, the whole truth, and nothing but the truth. There's no need to embellish or remove parts of your analysis that you don't like. Indeed, those points may be the start of another series of analysis! It's okay to put those parts in the "deemphasized" corners of the screen or a drill-down page, but do not omit them entirely!

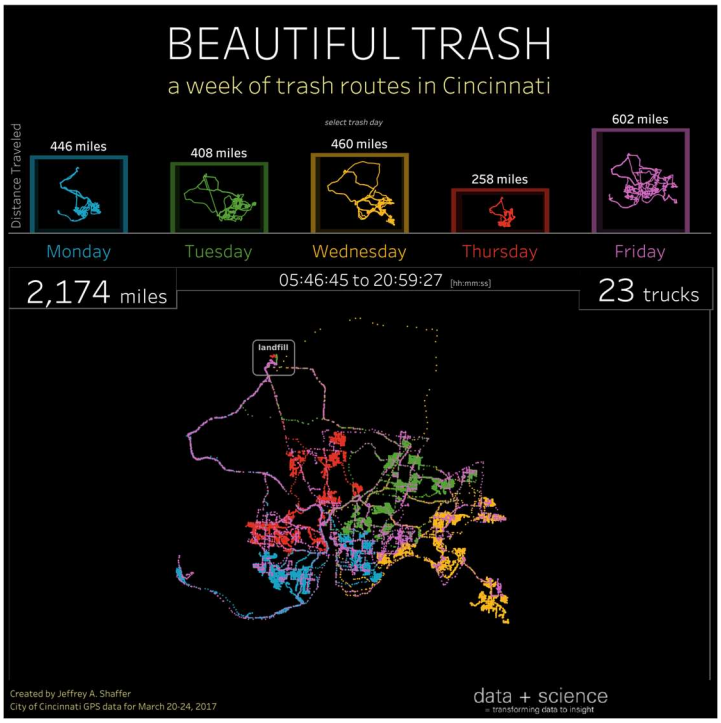
While Tableau and Power BI are two industry powerhouses in the display of data and data analysis, you can use Microsoft Excel to create some compelling visualizations, too. Both Tableau and Power BI have learning curves associated with them, so if you already know Excel, that's a great place to start.

Classic examples of good storytelling



One of the best examples of good storytelling comes long before the invention of computers and graphic analysis software. **Charles Joseph Mindard** made a storytelling graphic that covers **Napoleon's campaign in Russia**. This one single graphic encapsulates the dramatic reduction in forces that happened during the campaign as well as several correlated events such as temperature and location of forces.

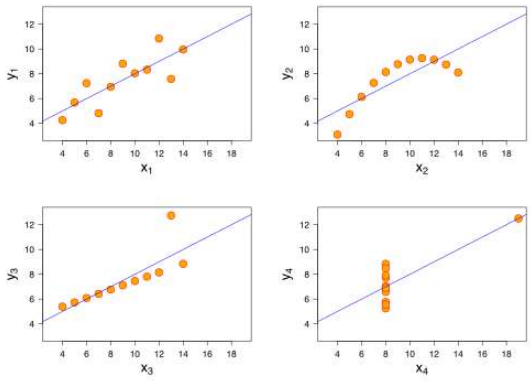
Jeffery Shaffer, Tableau Zen Master and Data Visualization expert, made trash delivery routes into a compelling story with his **"Beautiful Trash"** story. By using colors, a dark theme, and a map of Cincinnati, it is easy for viewers to see how a large city coordinates trash collection and how many miles are driven each day to do it.



On a single display users can see KPIs such as total miles, daily miles, the number of trucks, areas of the city covered (total and by day), etc.

Shaffer combines maps with a column chart to show which days have the most/least mileage and at the same time, which parts of the city are covered. That is maximizing screen real estate!

Anscombe's Quartet



Anscombe's Quartet is a graphic that is famous for showing how a visual analysis of data is necessary. The datasets represented by the associated image all have the same statistical measures for center, and produce the same line of best fit, but are clearly different visually.

Outliers and non-linear trends are just two reasons why graphic analysis of data is vital. Note that the further out an outlier is, the more "leverage" it has on the line of best fit.

Resources

1. Data storytelling with Microsoft Power BI. https://powerbi.microsoft.com/en-us/data-storytelling/
2. Data Storytelling: How to Effectively Tell a Story with Data (Harvard). https://online.hbs.edu/blog/post/data-storytelling
3. Data-Driven Storytelling (Standord, webinar). https://online.hbs.edu/blog/post/data-storytelling
4. Stories in Tableau. https://help.tableau.com/current/pro/desktop/en-us/stories.htm
5. Data Storytelling, Business Intelligence Trends (Tableau, PDF). https://www.tableau.com/reports/business-intelligence-trends/data-storytelling
6. Storytelling with Data (Tableau). https://www.tableau.com/solutions/customer/storytelling-data-0
7. Data storytelling (Columbia, grades 9-12 audience). https://www.earth.columbia.edu/videos/view/data-storytelling

This material is based upon work supported by the National Science Foundation under Grant DUE 2055411.Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.