

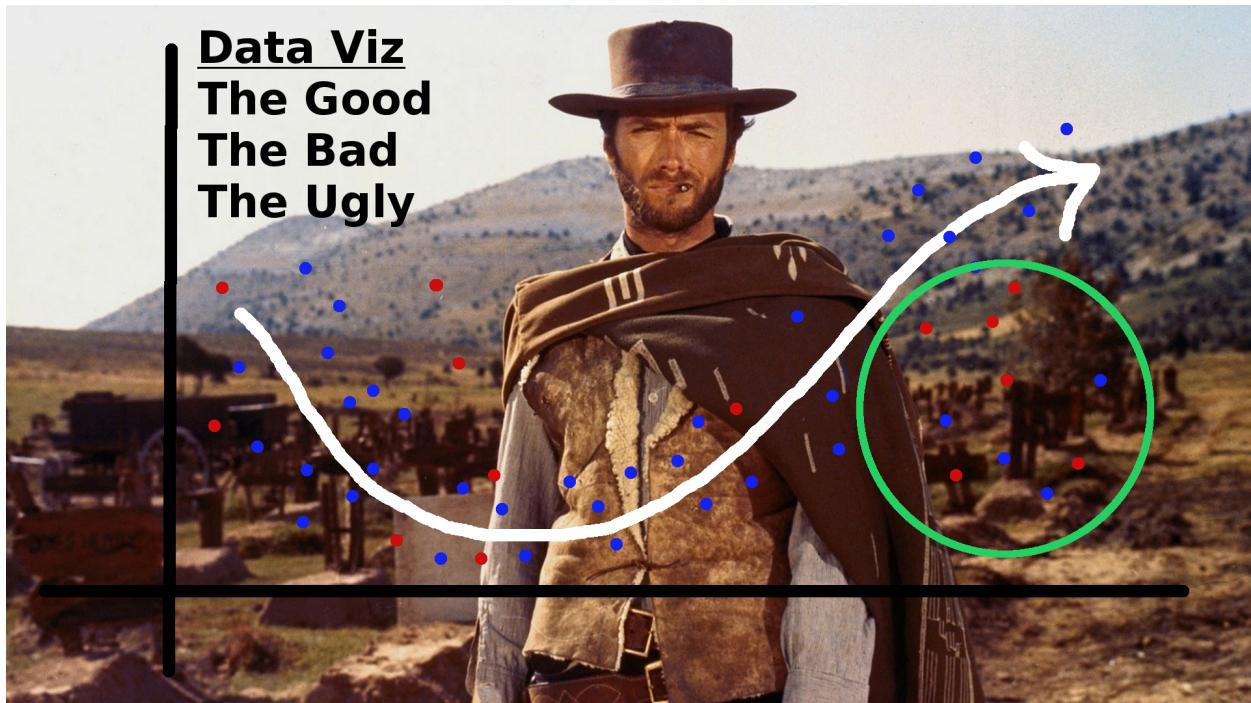
Data Visualization - Part 1

The Good, The Bad and The Ugly

Scott Stoltzman

March 14, 2017

Introduction to Data Visualization

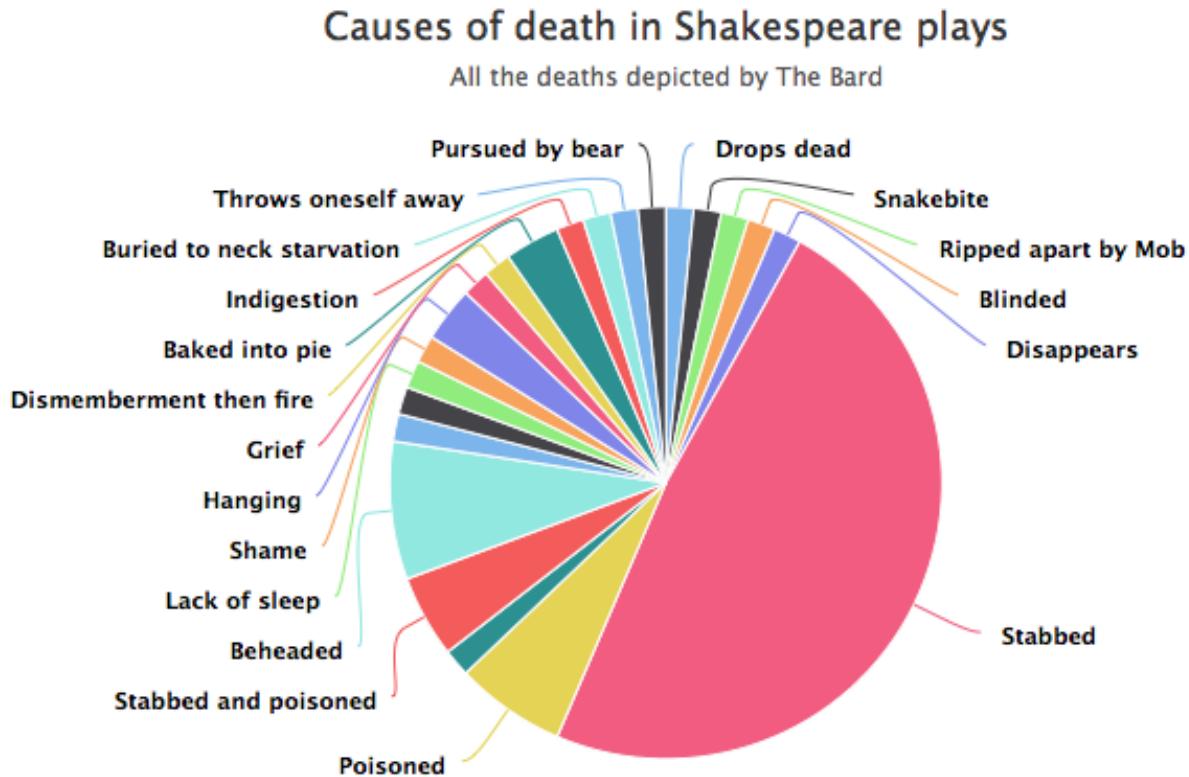


The topic of data visualization is very popular in the data science community. The market size for visualization products is valued at \$4 Billion and is projected to reach \$7 Billion by the end of 2022 according to Mordor Intelligence. While we have seen amazing advances in the technology to display information, the understanding of how, why, and when to use visualization techniques has not kept up. Unfortunately, people are often taught how to make a chart before even thinking about whether or not it's appropriate.

In short, are you adding value to your work or are you simply adding this to make it seem *less boring?* Let's take a look at some examples before going through the Stoltzmanniac Data Visualization Philosophy.

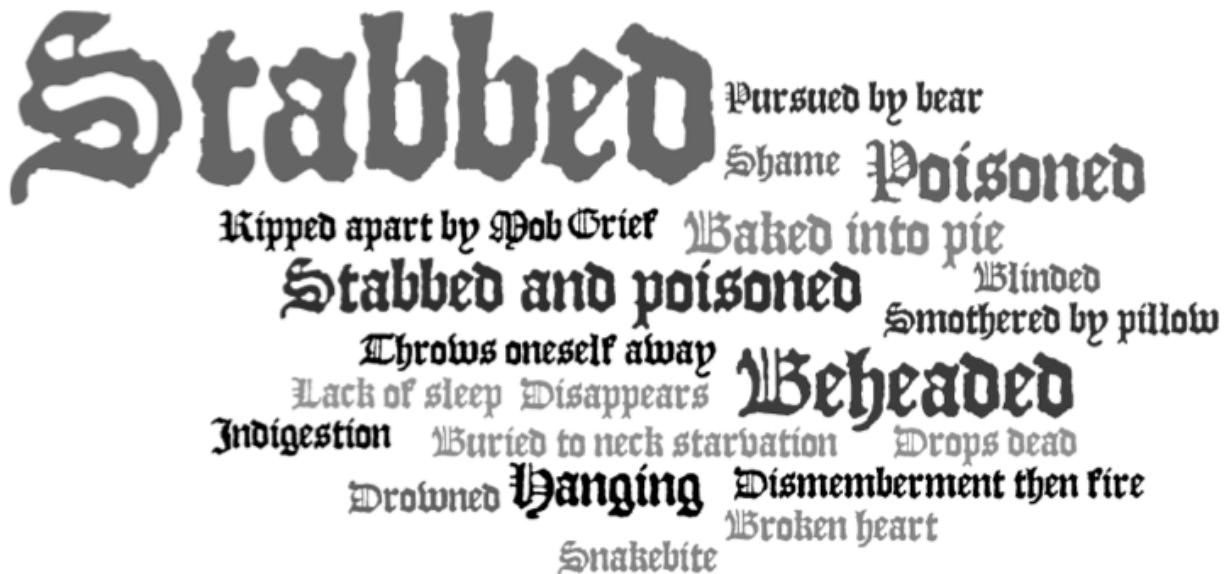
I have to give credit to Junk Charts - it inspired a lot of this post.

One author at Vox wanted to show the cause of death in all of Shakespeare



Is this not insane!?!?!

Using a legend instead of data callouts is the only thing that could have made this worse. The author could easily have used a number of other tools to get the point across. While wordles are not ideal for any work requiring exact proportions, it does make for a great visual in this article. Junk Charts Article.



Junkcharts

To be clear, I'm not close to being perfect when it comes to visualizations in my blog. The sizes, shapes, font colors, etc. tend to get out of control and I don't take the time in R to tinker with all of the details. However, when it comes to displaying things professionally, it has to be spot on! So, I'll walk through my theory and not worry too much about aesthetics (save that for a time when you're getting paid).

The Good, The Bad, The Ugly

“The Good” visualizations:

- Clearly illustrate a point
- Are tailored to the appropriate audience
 - Analysts may want detail
 - Executives may want a high-level view
- Are tailored to the presentation medium
 - A piece in an academic journal can be analyzed slowly and carefully
 - A slide in front of 5,000 people in a conference will be glanced at quickly
- Are memorable to those who care about the material
- Make an impact which increases the understanding of the subject matter

“The Bad” visualizations:

- Are difficult to interpret
- Are unintentionally misleading
- Contain redundant and boring information

“The Ugly” visualizations:

- Are almost impossible to interpret
- Are filled with completely worthless information
- Are intentionally created to mislead the audience
- Are inaccurate

Coming soon:

- Determining whether or not you need a visualization
- Choosing the type of plot to use depending on the use case
- Introduction to the ggplot2 in R and how it works
- Visualization beyond the standard charts and graphs

As always, the code used in this post is on my GitHub