**MODULE 6 Analytics Visualization**

*Storytelling with Data: A Data Visualization Guide for Business Professionals*

Chapter 10 Final thoughts

Permalink: <http://go.oreilly.com/rasmussen-college/library/view/storytelling-with-data/9781119002253/c10.xhtml>

**Removing Bias**

We live in a world where access to information has been made relatively easy with the internet. We are also constantly being influenced by companies who wish to sell their products, political organizations pushing their agendas, religious organizations seeking to convert you, and people we are connected with sharing their thoughts on social media. All of these are examples of people trying to influence us and their communication is filled with bias.

Now imagine yourself being called in for jury duty. You are sworn in for a case and asked to remove your personal bias and only consider the facts of a case. You are called upon to set aside any preconceived notions about social-economic class, race, age, religion, or sexual orientation. You are supposed to consider the facts of the case at hand and not introduce bias into your judgement about the case if you say you had a similar situation in life as perhaps the defendant.

As data analysts, you are called upon to present data fairly and accurately. But there are ways in which data can be presented which may influence the impact on the audience and the story you are trying to tell with data. Let’s consider a few examples, so that as data analysts, you can be informed and make the proper choices.

**The Three Averages**

Averages are calculated in three different ways. First the mean is calculated by adding all the values in a data set, then dividing the total by the number of unique values. The median is the number that falls exactly midway between a set of numbers. For example, if there are 44 values in a data set. The value at 22 will represent the median value. The mode is the value which occurs most often. If a real estate company was promoting home prices in a given area as part of a cost of living calculation, there can be a big difference between the median home price and average home price.

**Price vs. Price per Quantity**

Let’s say that you are presenting data that shows the difference in price between several brands of laundry detergent. Which product costs less? A consumer may not look at the number of loads that each bottle will wash. One bottle may wash 40 loads of clothes, and another 55 loads. Will the price then be a fair reflection of the cost to the consumer? Most stores display a cost per ounce number on the price sticker on the shelf, but many people do not take this into account.

**Choosing Scales on Graphs**

Another way that business professionals can intentionally or unintentionally skew data is to manipulate the scale in which data is presented. If you were graphing the temperature in several cities around the country during the summer months, your graph will have temperatures which are similar. We all know it is hot during the summer. If my graph is set to start at 0 degrees Fahrenheit and go up to 110 degrees Fahrenheit, then we may see trend lines which are closely grouped together. Now consider the same data set, with a scale which goes from 80 degrees to 110 degrees. The trend lines would show a larger gap between the temperatures. Both graphs would be factually correct, but would have completely different visuals if presented on a screen. Changing the scale of a graph can give a false impression and therefore, you should always orient your audience to the scale you are using in your data in order to remove bias from your data presentation.

**Ethics in Data Presentation**

Considering just the couple examples above, you can see how easy it would be to manipulate outcomes and influence an audience to think a certain way. While this may be the norm when promoting products for profit, other organizations must operate with integrity of their data. The finance industry came under fire for misrepresenting financial results and today we have the Sarbanes-Oxley regulations. There are so many cases where individuals have too much power to manipulate data or simply make a mistake and there is no verification. Having data controls, data governance, and data audits in place at your organization is always a good practice. Whenever data is presented to you, observe how the data is represented. Ask about the source, and ask how the data was prepared. You will become more knowledgeable about the data you are analyzing.