Cheatography

The Five C's of Data Visualization Cheat Sheet

by [deleted] via cheatography.com/2754/cs/7518/

Introduction

Organizations pull data from their ERP, CRM, and financial systems to use for reporting. The data users understand the data, and ideas begin rolling, but this often leads them to ask for more and more data, and increasingly complex scenarios evolve as they figure how they want to operate. It's up to the data stewards of the organization to be able to say "yes, we can deliver it."

But in order for all this to be possible, there must be control of this data and the relationships between entities so that the data is manageable. There are numerous industry statistics which cite that the majority, even as much as 80% to 90%, of all data collected, is not used

It does not matter how many, how big, how sophisticated, how thorough, or how detailed all the data you collect is. If you can't turn that data into actionable information for the people who need it, it's worthless. By following the five C's of data visualization you have a system in place to be successfu

Credit: Heine Krog Iversen is CEO of TimeXtender.
http://www.dbta.com/Editorial/Trends-and-Applications/The-Five-Cs--of-Data-Visualization-106821.aspx

1. Capture

We can Google and create data source connections to most any data source we have within 15 minutes. However, we must be able to replicate that connection to data across to all the individuals who need this data. We've got to be able to scale. This may be harder and more costly to do when the data idea we had is a complex one. We do not want to be the bottleneck for others accessing production and live transactional systems. We want to have a limited set of individuals perform the capture and acquisition and then point people to it. This apparent extra layer ups the ability to insulate all downstream reporting from migrations of source systems in the future and prevents transactional systems from a barrage of pings..

2. Clean

We've spoken about the volumes of data available. It's overwhelming. The increased amount of data introduces an overall degradation in the quality of the data coming in. So now it's up to the systems, in this case the data repository system, to be the authority on it. More data generated with limited numbers of rules put in place on data entry leads to incomplete items that need filling in and cleansing. Web traffic data comes in volumes. Many times it is irrelevant information to you, and you must seamlessly and continuously filter the data. Before data visualization happens, having a clean set of data is imperative. One effort should move on cleaning the captured repository. Downstream, if this simple rule is followed, questions and reduced confidence in the final results drop dramatically. The data users can then make educated decisions on what changes need to be made. This is the core reason why people look at data.

The 5 Cs of Data Visualization

3. Combine

Experience shows that great data visualizations come when pairing data from varied sources. However, the hardest way to combine data is to do it cross-platform. Once you've captured data in a common repository, you have removed completely any barriers to combining sources. In fact, you've generated a single new source and cut your technical challenges exponentially for every source you add beyond that.

4. Calculate

Calculations seem to gain a mind of their own in every reporting system. And similar to the clean data conversation, calculations done once and used in many places must be pivotal to an end solution. One would not want every report in the organization calculating the margin percent on a sale after removing sales tax. It's prone to simple order of operations errors, localization errors, and more. Centralized calculations done prior to data visualization allow for the ad hoc analysis to render real results that can be acted on quickly. Central calculations lead to an increased confidence in the data system and in the results and findings..

5. Control

Control is what you gain by putting together the C's into data before visualization. Control can sometimes be a word that gets misconstrued; however, having control over your data allows you to be more agile in your decision-making process—and ultimately more reactive to a data-hungry business. That kind of control leads to good data visualization.