[**Module 01Lesson – Creating Data Sets**](https://learning.rasmussen.edu/webapps/blackboard/content/listContent.jsp?course_id=_67422_1&content_id=_6076794_1&mode=reset)

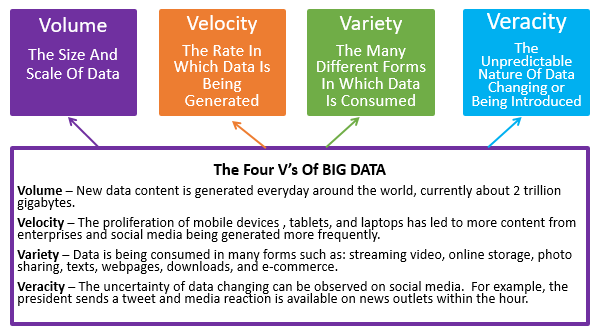
**Reading -**

* *Data Visualization: A Successful Design Process*
  + Chapter 1. The Context of Data Visualization
    - Permalink: <http://go.oreilly.com/rasmussen-college/library/view/data-visualization-a/9781849693462/ch01.html>
  + Chapter 2. Setting the Purpose and Identifying Key Factors
    - Permalink: <http://go.oreilly.com/rasmussen-college/library/view/data-visualization-a/9781849693462/ch02.html>
* *Storytelling with Data: A Data Visualization Guide for Business Professionals*
  + Introduction
    - Permalink: <http://go.oreilly.com/rasmussen-college/library/view/storytelling-with-data/9781119002253/f_07.xhtml>
  + Chapter 1 The importance of context
    - Permalink: <http://go.oreilly.com/rasmussen-college/library/view/storytelling-with-data/9781119002253/c01.xhtml>

Visualization Basics

The world that we live in today is full of data. It seems as though our lives are constantly being recorded in social media; creating a digital impression of who we are, what we like, and parts of our lives that we wish to share. We even have data to describe other data, which we call metadata. If you stop for a moment to think about exactly how much data is being creating today, it can be overwhelming. The infographic below from IBM provides a perspective on Big Data.

*Figure 1.1 The Four Vs of BIG DATA*

**

Big Data is being described here using four characteristics: volume, variety, velocity, and veracity. **Volume** describes the tremendous amount of data that is being generated in our digital world. **Variety** explains the different forms of data such as video, social media, and stored personal data. **Veracity** gives us a view of the uncertainty of data. Like breaking news, an important event can generate an enormous amount of internet traffic on thousands of websites at any given time. And finally, if we consider how quickly breaking news can happen, it speaks to the **Velocity** in which data is constantly changing. It is important to have the above understanding about data, so that when data sets are being created, consideration can be given to the four V’s. Even though there is a significant amount of data being created constantly, all of that data is not useful. Most of it is never used to generate insight, or at best it is underutilized. All data in a given data set does not have to be utilized in order to produce useful insight, but relevant data is necessary.

With so much data being generated from so many different sources, how can it be determined which of those sources contain valid data? All data should be vetted, regardless of the source. The validity of a data source begins with the credibility or reputation of the entity that produced the data. For example, it is generally accepted that data from U.S. government websites is considered credible. A central repository of data produced by the U.S. government is [www.data.gov](http://www.data.gov/). The U.S. Census Bureau also has a long history data collected. But just as the data is considered to be credible, it must also be accepted that there are certain limitations to such government data. For instance, the timing of U.S. Census Bureau reports can often lag data collection periods and therefore, is only useful for certain applications. U.S. Census reports are produced every 10 years and as a result, will miss new construction or new development areas.

Another source of credible data comes from research companies. Although operating as a for profit companies, their business is built upon their reputation for producing high quality data and insights. Some well-known research companies are: Gartner, Forrester, Consumer Reports, and J.D. Power & Associates to name a few. Research companies promote ethics in data analysis and readily shares their research methodologies with the general public. It is this transparency that aids research companies in gaining support as a credible source of data and insights. Below is one example of a research process map:

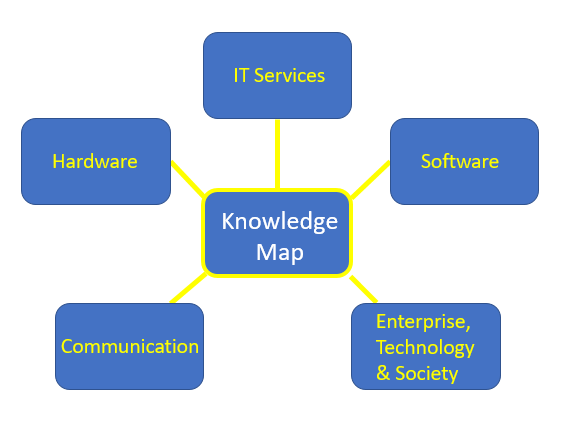
*Figure 1.2 Research Process*

**

Research companies employ thousands of analysts who are trained conduct industry research which is technology independent. Research companies ensure that their research is not influenced by specific vendors seeking drive their own products or services. This is an important distinction in gathering data from an online source which may be influenced by business agendas.

Part of what makes finding research easier is indexing content into categories. This makes searching for the right content more readily available to clients. This is done in the form of an indexing map. This indexing map is what adds structure to the data, in a relational format similar to what you may find in a relational database. The indexing map contains scores of research content into specific research categories making research data more searchable and cross-referenced. Below is a general representation of an indexing map.

*Figure 1.3 Index Map*

**

Another vital source of data comes from non-profit organizations whose purpose is to serve the general public. Examples of such organizations include IEEE who developed standards for such things as the internet and WiFi, government consortiums such as the World Economic Forum. Non-profit organizations have produced valuable data for research purposes and is often funded by government grants. An example of this comes from universities whose research may be to find the cure for certain diseases, or spark a change in technology with innovation labs.

Regardless of the source mentioned above, it should be understood that good data can come from many sources. The independence of the data and transparency of the research is what provides a foundation as a reliable data source. Data collection methodologies should not contain any biases which may be injected into research by the observer. Whenever, considering a data source, be sure to gather background on how the data was gathered, content which may have been excluded, and the methodology behind the research.

Download Tableau

In order to get started with Tableau, you will first need to download the student version of the software program. Rasmussen students are eligible for a free one year subscription of the software program. You will need to provide proof of enrollment at Rasmussen College (Such as a copy of unofficial transcripts). These instructions are available at: [https://www.tableau.com/academic/students#form](https://www.tableau.com/academic/students).

When you reach the “Tableau For Students” section, select the “Get Tableau For Free” button in the middle of the page. DO NOT select “Try Now” at the top right of the page. Try Now will route to a 14-day trial of Tableau and will not be sufficient for the entire term. Selecting the correct download is an important initial step, so **be sure to download the free one year version**.

**Note:** To register for Tableau, you will be asked to provide proof of enrollment at Rasmussen College by submitting an unofficial transcript. To obtain your unofficial transcript follow these steps:

* Log in to their Rasmussen Student Portal.
* Then go to the Student Account Center and click on My Program Progress to find the unofficial transcript. **Note**: There is no charge for the unofficial transcript (only an official one).
* Submit the unofficial transcript to the Tableau student signup site when you download their software.

You should receive confirmation of your Tableau registration within 2-3 days of submitting your unofficial transcript. **Contact your instructor if you do not hear back from Tableau and/or run into any complications completing the setup of the software.**

Course Project Introduction

**Project Overview - Consumption Awareness Program**

For your course project, you will apply the knowledge and skills gained throughout this course to the case scenario below.

**Case Scenario:**

The Department of Energy Efficiency has commissioned you to develop a study on energy consumption and would like to increase public awareness of the cost of electricity throughout the United States. Your study will make use of relevant data and visualization tools to answer the following questions:

* Which metrics are best for understanding the nature of the problem?
* How does geography and seasonality affect energy consumption
* Have advancements in technology reduced energy consumption over time?

In addition, your study will:

* Make use of actual data sets to provide answers to the key questions.
* Make use of data visualization software to present your summary of findings and key insights.
* Provide effective data visualization that is appropriate for the intended audience.
* Provide a visualization map or infographic that threads together relevant data sets into a single view.
* Compare quantitative and qualitative data sets.
* Include an executive summary that explains the credibility of data sources, summarizes key insights, and describes how energy consumption has changed over time. Your executive summary should be presented in an objective manner using editorial thinking.

Your overall final study should be 8 - 10 pages, including a title page, executive summary, content, and references.

[Review the Course Project Instructional Supplement](https://content.learntoday.info/Learn/QMB3300fw_Summer_17/qmb3300fw-summer-17/media/Course%20Project%20Instructional%20Supplement.docx) get started.

**Requirements**

Apply the course project case scenario for each of the submissions below.

**Due Date**

**The final submission for your course project is due in Module 06.** There will be individual assignments along the way. The module during which they are due is noted in the time line below.

**Time Line**

|  |  |
| --- | --- |
| **Module** | **Assignment** |
| 01 | Course Project Overview |
| 01 | **Evaluating Data Sets** |
| 02 | **Visualization Tool Comparison** |
| 03 | **Comparing Data Sets** |
| 04 | **Audience Perceptions** |
| 05 | **Create a Visualization Map** |
| 06 | **Final Submission** |