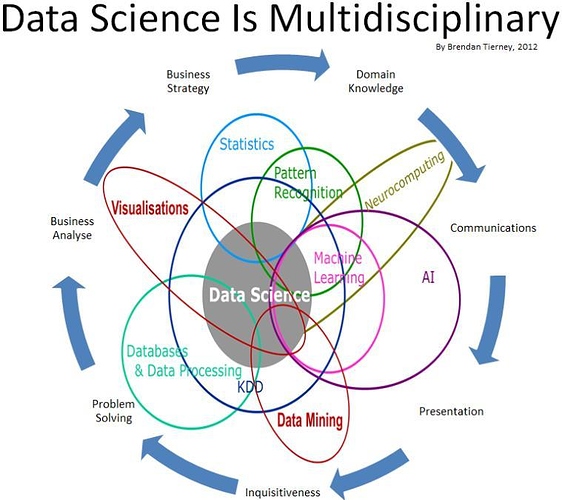
**Data Science and Understanding Analytics**

1. **Data Science:** Refers to the umbrella of techniques where you are trying to extract information and insights from data. This includes MIS reporting on the lowest level to building predictive models on the higher level.
2. **Data Mining:** refers to the science of collecting all the past data and then searching for patterns in this data. You look for consistent patterns and / or relationships between variables. Once you find these insights, you validate the findings by applying the detected patterns to new subsets of data. The ultimate goal of data mining is prediction - and predictive data mining is the most common type of data mining and one that has the most direct business applications.
3. **Data Analysis:** This is a loosely used term. People running reporting also say that they are analyzing data and so do predictive modelers. I would just take this as any attempt to make sense of data can be called as data analysis.
4. **Machine learning** - is the science of creating algorithms and program which learn on their own. Once designed, they do not need a human to become better. Some of the common applications of machine learning include following: Web Search, spam filters, recommender systems, ad placement, credit scoring, fraud detection, stock trading, computer vision and drug design. An easy way to understand is this - it is humanly impossible to create models for every possible search or spam, so you make the machine intelligent enough to learn by itself. When you automate the later part of data mining - it is known as machine learning.



“Everyone talks about it. No one really knows how to do it. It’s been the source of many rumors. Everyone thinks that everyone else is going it… So everyone claims that they’re doing it.”

But what is Big Data, by definition? Big Data typically refers to any large quantity of raw data that can be collected, stored, and analyzed through various means that reveals patterns or trends relating to behaviors – particularly that of consumers – which can be used to maximize a business’ potential. That was the point that Robert Abate made clear at the beginning of his presentation; his focus was more on utilizing Big Data effectively through the use of Data Visualizations.

**Why is Big Data important?**

What makes Big Data so useful to many companies is the fact that it provides answers to many questions that they didn’t even know they had in the first place, he said. In other words, it provides a point of reference. With such a massive amount of information, the data is able to be shaped or tested in any way that the company sees fit. In doing so, organizations are able to pinpoint issues in a more comprehensible form. Collecting masses of data and finding a trend within the data allows the businesses to move much more quickly, smoothly, and efficiently. It also allows them to eliminate problem areas before those previously elusive issues pull their profits or reputation through the proverbial mud.

Without the aid of data, innumerable operations, in all levels of the enterprise, could be threatened, impaired, or destroyed completely:

“Data is like the air that a company breathes. Just as humans cannot survive without oxygen, a company cannot survive without the data [air] necessary to its life. Without that air, the company would theoretically choke.”

Every company uses data; the more efficiently that a company uses its data, the more potential it has to flourish. Such truisms are not unknown, but according to Mr. Abate, many organizations have yet to really grasp such simple ideas:

“Part of the challenges we have are expressing to the business [what data can do], is how the business is really the theme. We have to figure out how to communicate and to create a direct link to them.”

Big Data, especially when used in accordance with infographics and other visual elements, also allows a company to generate those answers faster. This is ideal, because the faster that a company can generate information from their sources, the faster that they can generate their answers. Speed cannot be generated with more information alone, he said. Without someone to review that data, and make sense of it, that massive hump of data is still just raw data that holds no potential or value. Just the same, adding more eyes does not necessarily equal more spotted trends – the very same way that adding more employees does not increase productivity, but instead begins to hinder it.

So what should an enterprise do?

**Increase Productivity Using Big Data and Data Visualization**

If an organization can generate more beneficial visual matter, then they can increase cognition. According to an analogy employed by Mr. Abate, while many individuals work, they typically only use one computer screen. But one computer screen can only do so much: it has limitations. And with limitations, the individual cannot hone their potential. If one were to have two, or even three computer screens, their ability to multitask and come to conclusions would be increased. If their resources were increased, they could make better use of their time. When it comes to Big Data, the saying “a picture is worth a thousand words” is an understatement. It is much easier to see data in a visual/physical form than it is to comprehend it as a simple spreadsheet or in other quantitative formats.

He gave a simple and easy explanation:

“To put it simply, if sales were to go down in the southwest, then one could use another computer, or phone, to bring up the weather in the southwest. Only then would one realize that there was a snowstorm in the southwest, which gives a direct correlation as to why the sales plummeted that week. Because clearly no one is going to be buying products if they can’t reach the store!”

**Why the Sudden Reliance on Big Data?**

The emergence of massive quantities of cellphones, tablets, flip computers, as well as the growth of Cloud Computing, along with other data sources such as sensors and the Internet of Things, have precipitated larger masses of data than anyone has ever seen. According to figures given in the presentation, in just the last two years alone, more data has been created than in all of history. To put that factoid into perspective, consider the fact that networking computers have been around since the 1950s, and it will become much easier to see why Big Data is such a big deal.

For even more shock value, consider these numbers that Mr. Abate presented on one of his slides:

* Every sixty seconds, 98,000 tweets or more are made on Twitter.
* Every sixty seconds, 695,000 status updates are made on Facebook.
* Every sixty seconds, 11 million instant messages are sent.
* Every sixty seconds, 698,445 Google searches are made.
* Every sixty seconds, 168 million or more emails are sent.
* Every sixty seconds, 1,820 terabytes (TB) of data is created.

It is no wonder why the last two years exceeded the rest of history in terms of data creation.

All of this information is useless; however, if there is no method implemented to organize it into a Data Lake or some other way the data assets can be leveraged effectively.

**Making Good Use of the Data**

Understanding the attributes that the data pertains to is key to using Big Data successfully. This is otherwise known as “getting to know your customer,” so that the company can market to their customers’ needs with a much higher success rate.

He gave an example of a case study where his team helped to cleanse the data of one of their clients. They removed any unrelated or outlier data from their data sets and thus narrowed it down to one key question or demographic. In doing so, they were able to identify which of their products sold more, which of their products did not sell, and thus could be eliminated. They looked at four primary data elements: revenue, frequency, value, and tenure. Mr. Abate stressed that anything more than four elements at a time, on any given visualization, become too difficult for people to follow. By eliminating products that did not sell, they were increasing their future revenue by cutting off waste. But they could not have done any of this without utilizing Data Visualization.

Data Visualization was key. By increasing their use of visualizations, the business found the value they sought. Creating more infographics, and “more screens,” allowed them to get more information, faster. It allowed them to realize they already knew much of the information that should have been obvious before. This increased the functionality of the department because they were able to ask better questions. It created linkages between data points that seemingly didn’t have any links at all. They were able to locate the good and the bad data, and in doing so they were able to maximize their productivity as well as the value of the information they collected.

Utilizing Big Data assets properly is important for any company, no matter their size. When Big Data’s potential is maximized by employing visualizations, trends that were not seen beforehand can be located with ease. As discussed by Mr. Abate, these trends can provide “information” that can be turned into valuable “insights” such as: who their customers are, how many customers they have, who the high-value and low-value customers are, and so many others.

He ended his presentation by discussing some Key Takeaways. He stated that Big Data Visualizations are the future and that it’s necessary to use more than one “screen” to gain “brain” insights. He included a few Best Practices as well: build iteratively, combine up to four “measures” per graphic (but not more), visualize movements pictorially, and make sure to plan for extreme Web scales.

**Steps Included in the process of Analysis of Data**

For most businesses and government agencies, lack of data isn’t a problem. In fact, it’s the opposite: there’s often too much information available to make a clear decision.

With so much data to sort through, you need something more from your data:

* You need to know it is the right data for answering your question;
* You need to draw accurate conclusions from that data; and
* You need data that informs your decision making process

In short, you need better data analysis. With the right data analysis process and tools, what was once an overwhelming volume of disparate information becomes a simple, clear decision point.

To improve your data analysis skills and simplify your decisions, execute these five steps in your data analysis process:

**Step 1: Define Your Questions**

In your organizational or business data analysis, you must begin with the right question(s). Questions should be measurable, clear and concise. Design your questions to either qualify or disqualify potential solutions to your specific problem or opportunity.

For example, start with a clearly defined problem: A government contractor is experiencing rising costs and is no longer able to submit competitive contract proposals. One of many questions to solve this business problem might include: Can the company reduce its staff without compromising quality?

**Step 2: Set Clear Measurement Priorities**

This step breaks down into two sub-steps: A) Decide what to measure, and B) Decide how to measure it.

**A) Decide What To Measure**

Using the government contractor example, consider what kind of data you’d need to answer your key question. In this case, you’d need to know the number and cost of current staff and the percentage of time they spend on necessary business functions. In answering this question, you likely need to answer many sub-questions (e.g., Are staff currently under-utilized? If so, what process improvements would help?). Finally, in your decision on what to measure, be sure to include any reasonable objections any stakeholders might have (e.g., If staff are reduced, how would the company respond to surges in demand?).

**B) Decide How To Measure It**

Thinking about how you measure your data is just as important, especially before the data collection phase, because your measuring process either backs up or discredits your analysis later on. Key questions to ask for this step include:

* What is your time frame? (e.g., annual versus quarterly costs)
* What is your unit of measure? (e.g., USD versus Euro)
* What factors should be included? (e.g., just annual salary versus annual salary plus cost of staff benefits)

**Step 3: Collect Data**

With your question clearly defined and your measurement priorities set, now it’s time to collect your data. As you collect and organize your data, remember to keep these important points in mind:

* Before you collect new data, determine what information could be collected from existing databases or sources on hand. Collect this data first.
* Determine a file storing and naming system ahead of time to help all tasked team members collaborate. This process saves time and prevents team members from collecting the same information twice.
* If you need to gather data via observation or interviews, then develop an interview template ahead of time to ensure consistency and save time.
* Keep your collected data organized in a log with collection dates and add any source notes as you go (including any data normalization performed). This practice validates your conclusions down the road.

**Step 4: Analyze Data**

After you’ve collected the right data to answer your question from Step 1, it’s time for deeper data analysis. Begin by manipulating your data in a number of different ways, such as plotting it out and finding correlations or by creating a pivot table in Excel. A pivot table lets you sort and filter data by different variables and lets you calculate the mean, maximum, minimum and standard deviation of your data – [JUST BE SURE TO AVOID THESE FIVE PITFALLS OF STATISTICAL DATA ANALYSIS](http://www.bigskyassociates.com/blog/bid/356764/5-Most-Important-Methods-For-Statistical-Data-Analysis).

As you manipulate data, you may find you have the exact data you need, but more likely, you might need to revise your original question or collect more data. Either way, this initial analysis of trends, correlations, variations and outliers helps you [FOCUS YOUR DATA ANALYSIS ON BETTER ANSWERING YOUR QUESTION](http://www.bigskyassociates.com/blog/bid/339496/Are-You-Using-The-Right-Data-Analysis-Techniques) and any objections others might have.

During this step, data analysis tools and software are extremely helpful. Visio, Minitab and Stata are all good software packages for advanced statistical data analysis. However, in most cases, nothing quite compares to Microsoft Excel in terms of decision-making tools. If you need a review or a primer on all the functions Excel accomplishes for your data analysis, we recommend this [HARVARD BUSINESS REVIEW CLASS](http://hbr.org/product/baynote/an/2001HF-HTM-ENG?referral=00506).

**Step 5: Interpret Results**

After analyzing your data and possibly conducting further research, it’s finally time to interpret your results. As you interpret your analysis, keep in mind that you cannot ever prove a hypothesis true: rather, you can only fail to reject the hypothesis. Meaning that no matter how much data you collect, chance could always interfere with your results.

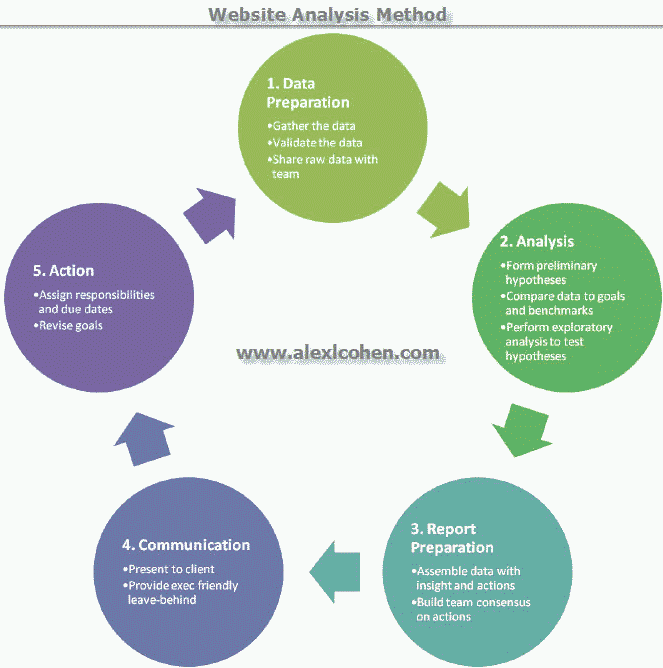
As you interpret the results of your data, ask yourself these key questions:

* Does the data answer your original question? How?
* Does the data help you defend against any objections? How?
* Are there any limitation on your conclusions, any angles you haven’t considered?

If your interpretation of the data holds up under all of these questions and considerations, then you likely have come to a productive conclusion. The only remaining step is to use the results of your data analysis process to decide your best course of action.

By following these five steps in your data analysis process, you make better decisions for your business or government agency because your choices are backed by data that has been robustly collected and analyzed. With practice, your data analysis gets faster and more accurate – meaning you make better, more informed decisions to run your organization most effectively.

Want to draw the most accurate conclusions from your data? Click below to download a free guide from Big Sky Associates and discover how the right data analysis drives success for your organization.



**Big data**

Big data is one of the misunderstood (and misused) terms in today’s market. But with a clearer understanding the how to apply big data to business intelligence (BI), you can help customers navigate the ins and outs of big data, including how to get the most from their big data analytics.

**What Do You Mean By “Big Data”?**

Big data can be applied to real-time fraud detection, complex competitive analysis, call center optimization, [consumer sentiment analysis](http://briefingsdirect.blogspot.com/2013/10/democratic-national-committee-leverages.html), intelligent traffic management, and to manage smart power grids, to name only a few applications. Big data is characterized by [three primary factors](http://www.forbes.com/sites/oreillymedia/2012/01/19/volume-velocity-variety-what-you-need-to-know-about-big-data/): volume (too much data to handle easily); velocity (the speed of data flowing in and out makes it difficult to analyze); and variety (the range and type of data sources are too great to assimilate). In short, big data simply means more than an organizations can manage effectively with their current BI program.  But with the right analytics, big data can deliver richer insight since it draws from multiple sources and transactions to uncover hidden patterns and relationships.

There are four types of big data BI that really aid business:

1. Prescriptive – This type of analysis reveals what actions should be taken. This is the most valuable kind of analysis and usually results in rules and recommendations for next steps.
2. Predictive – An analysis of likely scenarios of what might happen. The deliverables are usually a predictive forecast.
3. Diagnostic – A look at past performance to determine what happened and why. The result of the analysis is often an analytic dashboard.
4. Descriptive – What is happening now based on incoming data. To mine the analytics, you typically use a real-time dashboard and/or email reports.

**Big Data Analytics in Action**

**Prescriptive analytics** is really valuable, but largely not used. [According to Gartner](http://www.enterpriseappstoday.com/business-intelligence/gartner-taps-predictive-analytics-as-next-big-business-intelligence-trend.html), 13 percent of organizations are using predictive but only 3 percent are using prescriptive analytics. Where big data analytics in general sheds light on a subject, prescriptive analytics gives you a laser-like focus to answer specific questions. For example, in [the health care industry](http://medcitynews.com/2013/03/four-types-of-healthcare-analytics-that-providers-are-using-to-improve-population-health/#ixzz2tEc2sgSv), you can better manage the patient population by using prescriptive analytics to measure the number of patients who are clinically obese, then add filters for factors like diabetes and LDL cholesterol levels to determine where to focus treatment. The same prescriptive model can be applied to almost any industry target group or problem.

**Predictive analytics** use big data to identify past patterns to predict the future. For example, some companies are using predictive analytics for sales lead scoring. Some companies have gone one step further use predictive analytics for the entire sales process, analyzing lead source, number of communications, types of communications, social media, documents, CRM data, etc. Properly tuned predictive analytics can be used to support sales, marketing, or for other types of complex forecasts.

**Diagnostic analytics** are used for discovery or to determine why something happened. For example, for a social media marketing campaign, you can use descriptive analytics to assess the number of posts, mentions, followers, fans, page views, reviews, pins, etc. There can be thousands of online mentions that can be distilled into a single view to see what worked in your past campaigns and what didn’t.

**Descriptive analytics** or data mining are at the bottom of the big data value chain, but they can be valuable for uncovering patterns that offer insight. A simple example of descriptive analytics would be assessing credit risk; using past financial performance to predict a customer’s likely financial performance. Descriptive analytics can be useful in the sales cycle, for example, to categorize customers by their likely product preferences and sales cycle.

As you can see, harnessing big data analytics can deliver big value to business, adding context to data that tells a more complete story. By reducing complex data sets to actionable intelligence you can make more accurate business decisions. If you understand how to demystify big data for your customers, then your value has just gone up tenfold.