



# BUFFALO STATE

The State University of New York

[www.buffalostate.edu](http://www.buffalostate.edu)

# Origins, Recent Developments, and Impact of Data Science

*-In God we trust. All others must bring data*

W.E. Deming-

**SUNY Buffalo State**  
**CS4ALL**  
**October 13, 2017**

# Outline

- \* Factual (i.e. natural; collected by our senses) Data.
- \* Digital Data and how it is changing the way we learn about the world.
- \* Origin of Data Science and Data Analytics
- \* The components of Data Science
- \* Impact of Data Science
- \* Curriculum and Degrees in Data Science

# Data

- \* Merriam-Webster's definition of DATA: "*factual* information (such as measurements or statistics) used as a basis for reasoning, discussion, or calculation"  
-- (I colored and italicized "factual" 😊) –
- \* What does "factual" mean?

# Data

\* I quote from Wikipedia:

“Solipsism (/ˈsɒlɪpsɪzəm/ ); from Latin solus, meaning “alone”, and ipse, meaning “self”) is the philosophical idea that only one's own mind is sure to exist. As an epistemological position, solipsism holds that knowledge of anything outside one's own mind is unsure; the external world and other minds cannot be known and might not exist outside the mind. As a metaphysical position, solipsism goes further to the conclusion that the world and other minds do not exist.”

# Factual Data

## (is it really factual?)

- \* We obtain information (i.e. factual data) about the world with our senses, but how reliable are our senses?
- \* Consider the fact that:
  - 1) We constantly confuse facts with “automatic conclusions” our mind creates.
  - 2) We constantly misuse abstraction and generalization.



**GAME #9**

# Factual Data

## (is it really factual?)

- \* Fact?: Buffalo is one of the 10 food capitals of the world!!!  
<http://www.nationalgeographic.com/travel/top-10/food-cities/>
- \* More “facts”:  
<https://www.amazon.com/Everydata-Misinformation-Hidden-Little-Consume/dp/B01DV2EJI2>



# Computing & Computer Technology

- \* Humans have done computations for thousands of years. For example the Euclidean Algorithm was first recorded 300 BC (over 2300 years ago!!!)

Step k	Equation	Quotient and remainder
0	$1071 = q_0 462 + r_0$	$q_0 = 2$ and $r_0 = 147$
1	$462 = q_1 147 + r_1$	$q_1 = 3$ and $r_1 = 21$
2	$147 = q_2 21 + r_2$	$q_2 = 7$ and $r_2 = 0$ ; algorithm ends

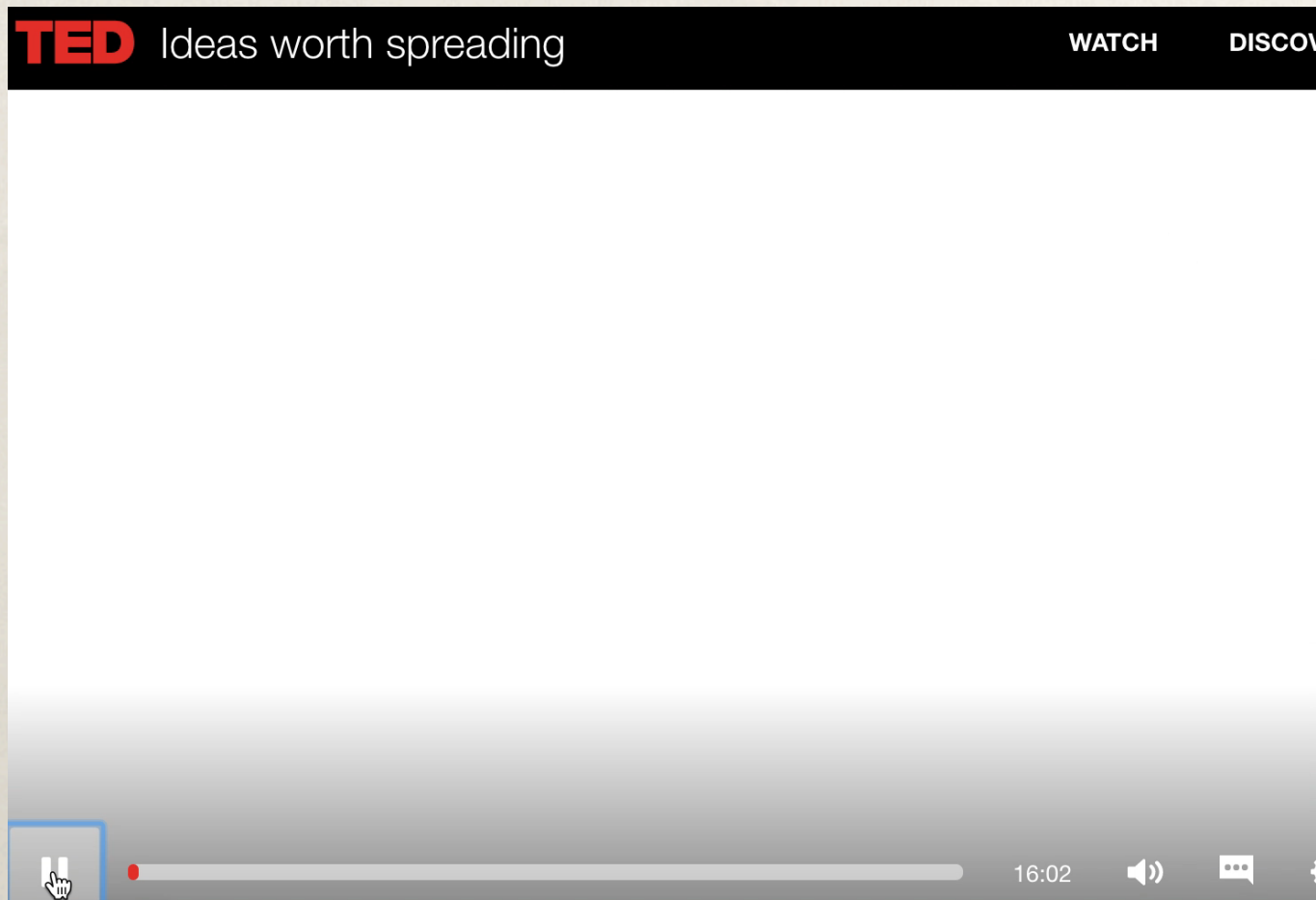
# Computing & Computer Technology

- \* Humans have created technology for thousands of years



# Computing & Computer Technology

- \* Tim Berners-Lee invented the World Wide Web

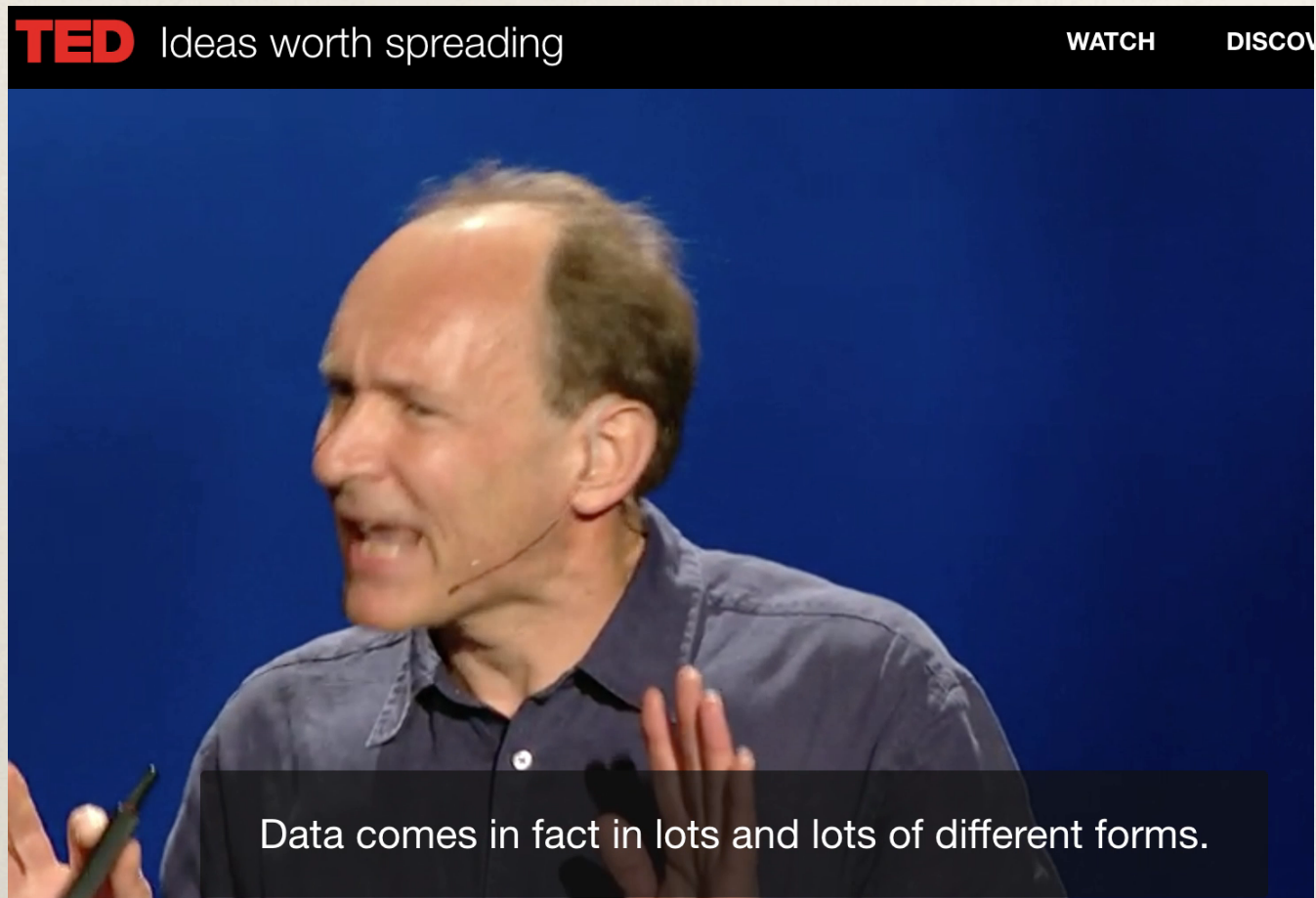


# Computing & Computer Technology

- \* Tim Berners-Lee invented the World Wide Web in 1989.  
Why? What is the WWW?
- \* The WWW allows information in different places and with different formats to be displayed in one place:  
The browser in your computer.
- \* It took less than 20 years for the WWW to be useful to all of us. (Fact: facebook had 100 million users in 2008.)

# Computing & Computer Technology

- \* Why is Raw Data + Internet a killer combo



# Fast forward to Digital Data

- \* Artificial Intelligence Data Analytics (Machine Learning, to be specific) was used to individual's reaction to sugar intake:

[http://www.cell.com/abstract/S0092-8674\(15\)01481-6](http://www.cell.com/abstract/S0092-8674(15)01481-6)

- \* This is about two years of sleep brain waves I measured on myself:

[Sleep data 2011-2012-2013](#)



# Fast forward to Digital Data

- \* It turns out we can record most (maybe all) of our sensorial perceptions using 0's and 1's.  
(IS THIS AMAZING TO YOU? IT IS TO ME ☺ )
- \* Some amazing facts about data:  
[file:///Users/carbonjo/Documents/Documents/MacBook-BackUp/Documents/DATA-Analytics-PSM\\_BS/CS4ALL-Oct13-2017/Big%20Data%20Mind-Boggling%20Facts%20Everyone%20Must%20Read.html](file:///Users/carbonjo/Documents/Documents/MacBook-BackUp/Documents/DATA-Analytics-PSM_BS/CS4ALL-Oct13-2017/Big%20Data%20Mind-Boggling%20Facts%20Everyone%20Must%20Read.html)

# Big Data humor

Moore's Law of Big Data:

"The Amount of Nonsense Packed  
Into the Term "BIG DATA" Doubles  
Approximately Every Two Years."

-Mike Pluta, 2014-08-10



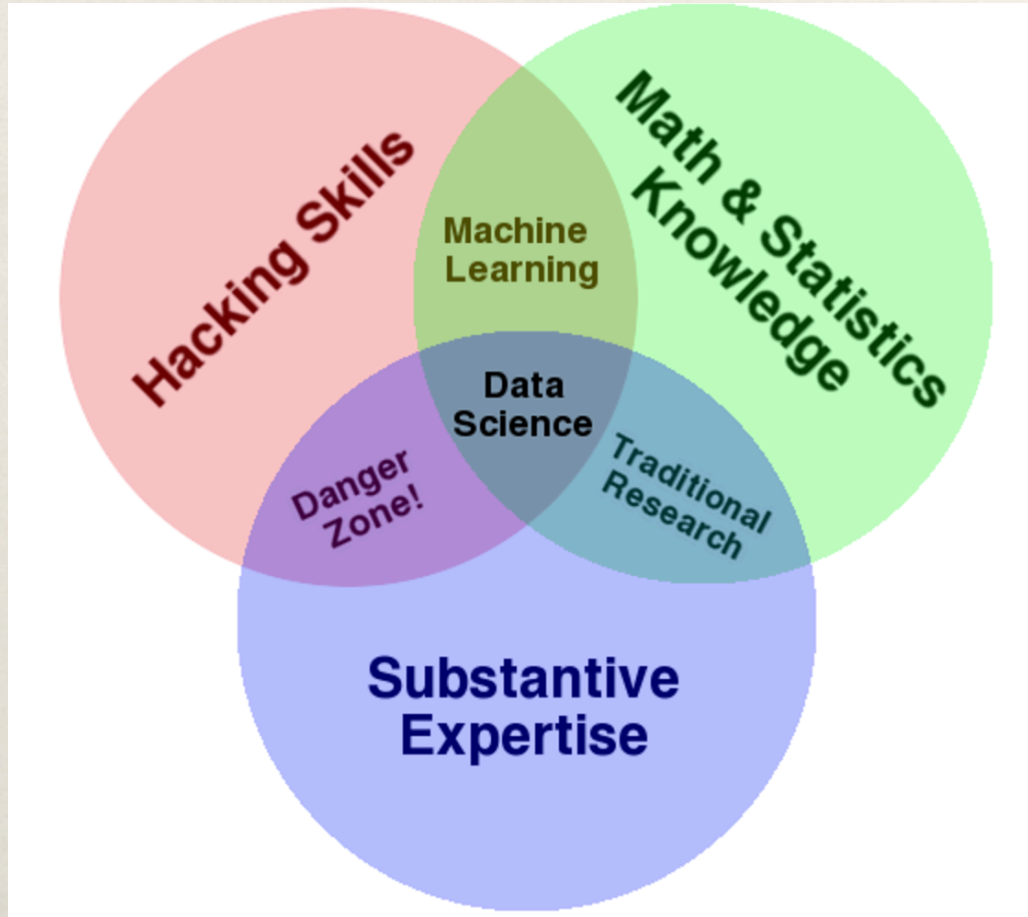
# Data Science: An Introduction/ A History of Data Science

- \* Quote from WikiBooks “Data Science: An Introduction/ A History of Data Science”:
- \* “The term "Data Science" was coined at the beginning of the 21st Century. It is attributed to William S. Cleveland[1] who, in 2001, wrote "Data Science: An Action Plan for Expanding the Technical Areas of the Field of Statistics." [2] About a year later, the International Council for Science: Committee on Data for Science and Technology[3] started publishing the CODATA Data Science Journal beginning April 2002.[4] Shortly thereafter, in January 2003, Columbia University began publishing The Journal of Data Science .[5]”

# Data Science or Data Analytics?

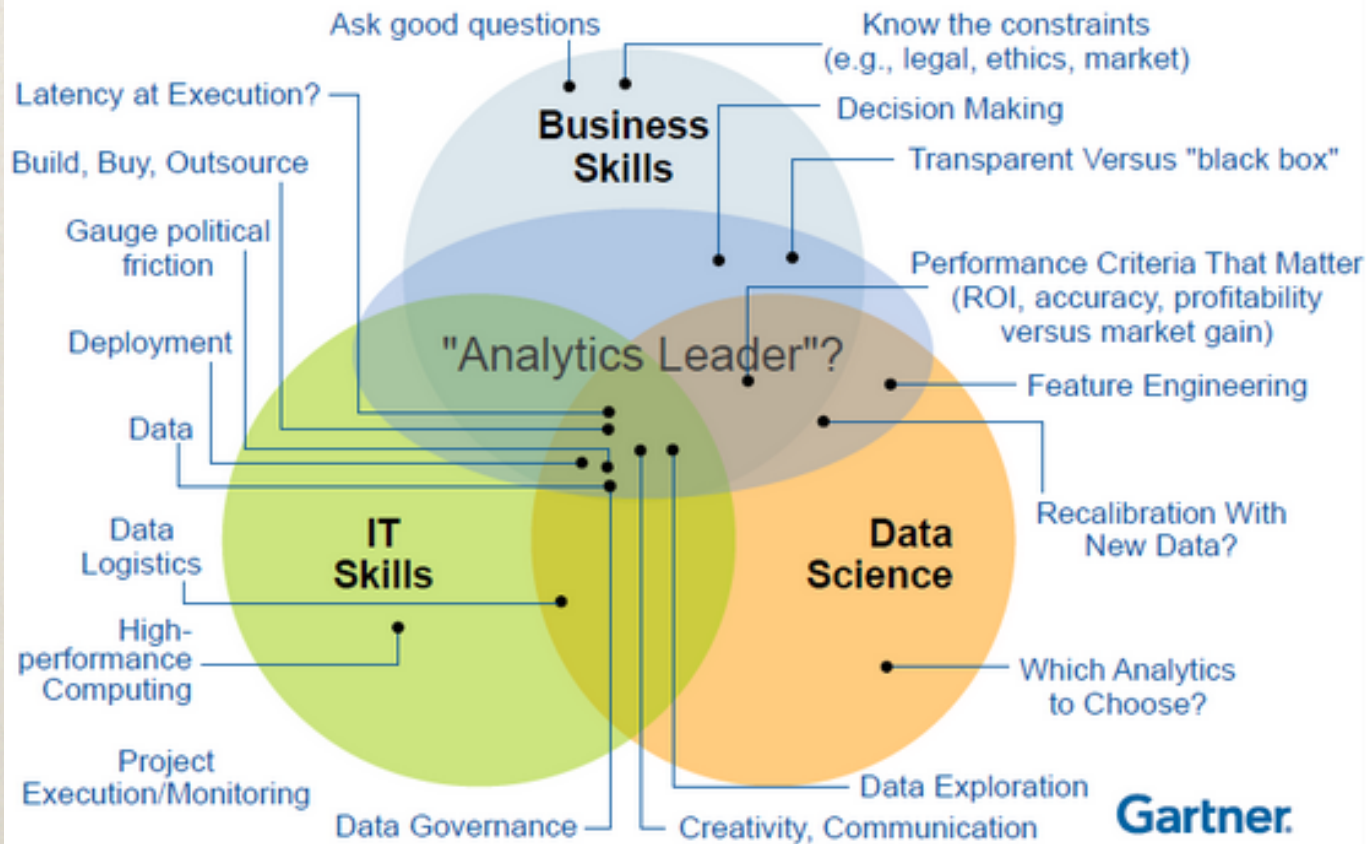
- \* **Data Science** is the growing body of knowledge used to collect, organize and extract information from data. Data Science is growing into a stand-alone field, with undergraduate and graduate degrees being offered in it.
- \* **Data Analytics** tends to be a name for the skill practiced by Data Scientists. It is different than, and a superset of, Data Analysis and Statistical Analysis.

# Data Science (D. Conway, 2013)



# Data Science Venn Diagram (revisited)

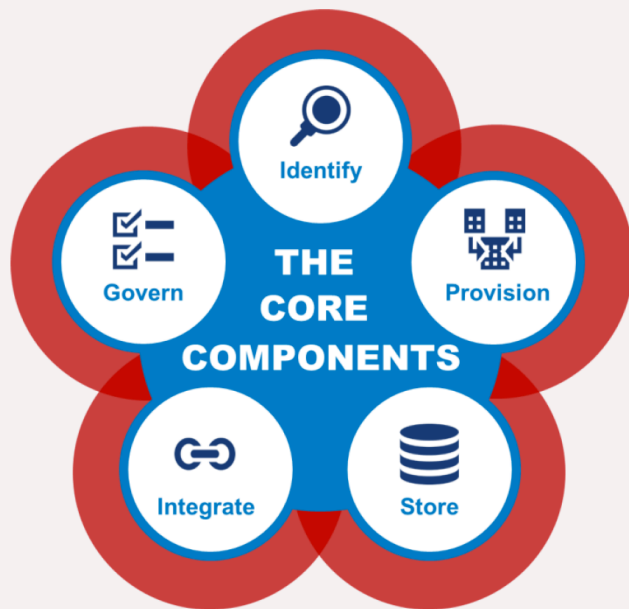
## Driving the Success of Data Science Solutions: Skills, Roles and Responsibilities ...



# Data strategy skills (E. Levy, VP-Data mangmt, SAS)

Data Strategy

## The Core Components



- Identify** The ability to identify data and understand its meaning regardless of its structure, origin, or location
- Provision** Enabling data to be packaged and made available while respecting all rules and access guidelines
- Store** Persisting data in a structure and location that supports access and processing across the enterprise
- Integrate** Moving and combining data residing in multiple locations and providing a unified view of the data
- Govern** Establishing and communicating information policies and mechanisms to ensure effective data usage.

# Who need this skills?

Why should you care? The analytics skills are needed now



Open jobs asking for analytics skills in 2015

**2.3M**

Forecast of population with analytics skills by 2018

**2.9M**

Notes: US data only.

Source: Burning Glass Technologies analysis of 26.9 million US job postings from 2015. McKinsey Global Institute, Big Data: The next frontier for innovation, competition, and productivity (June 2011).

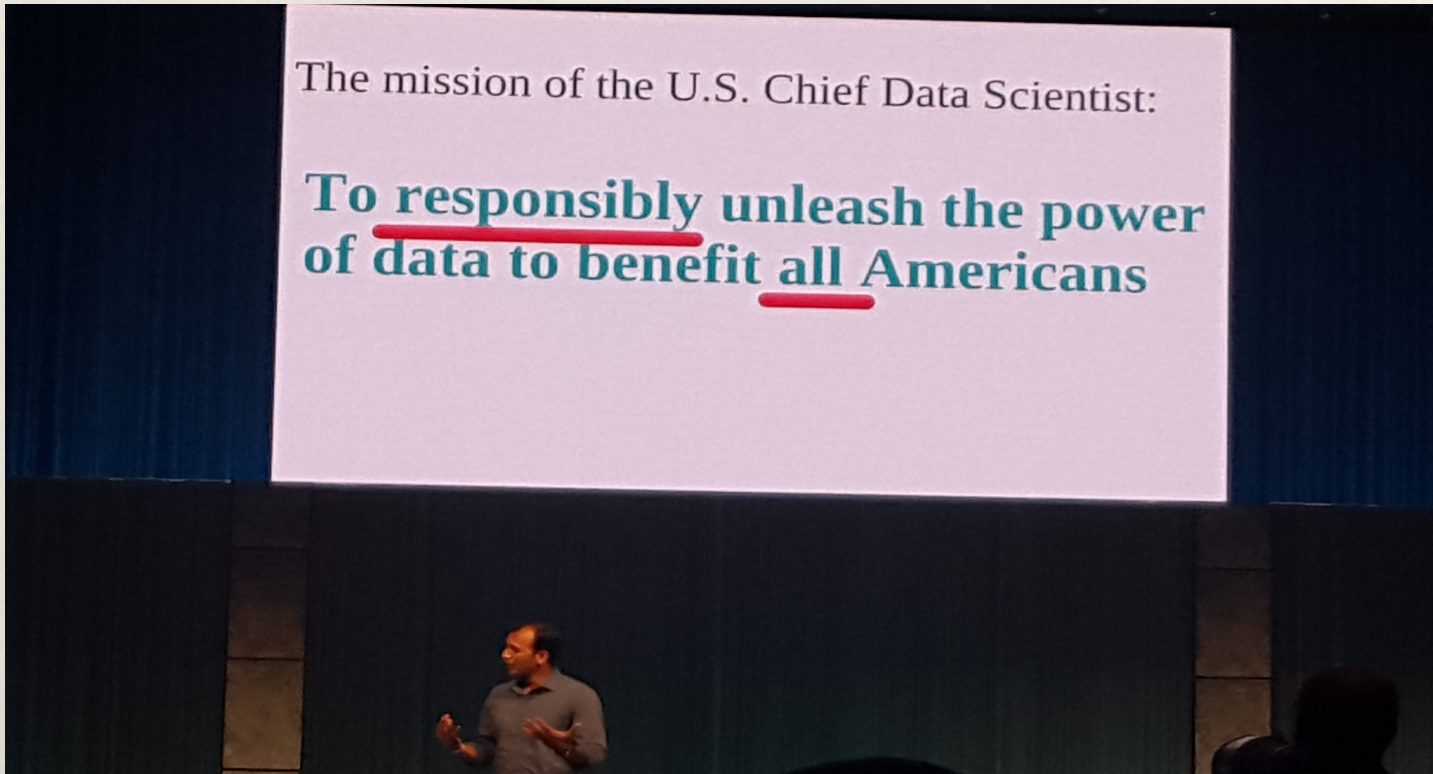
# D.J. Patil

(Former Chief Data Scientist of the United States  
Office of Science and Technology Policy)

## Government role

The mission of the U.S. Chief Data Scientist:

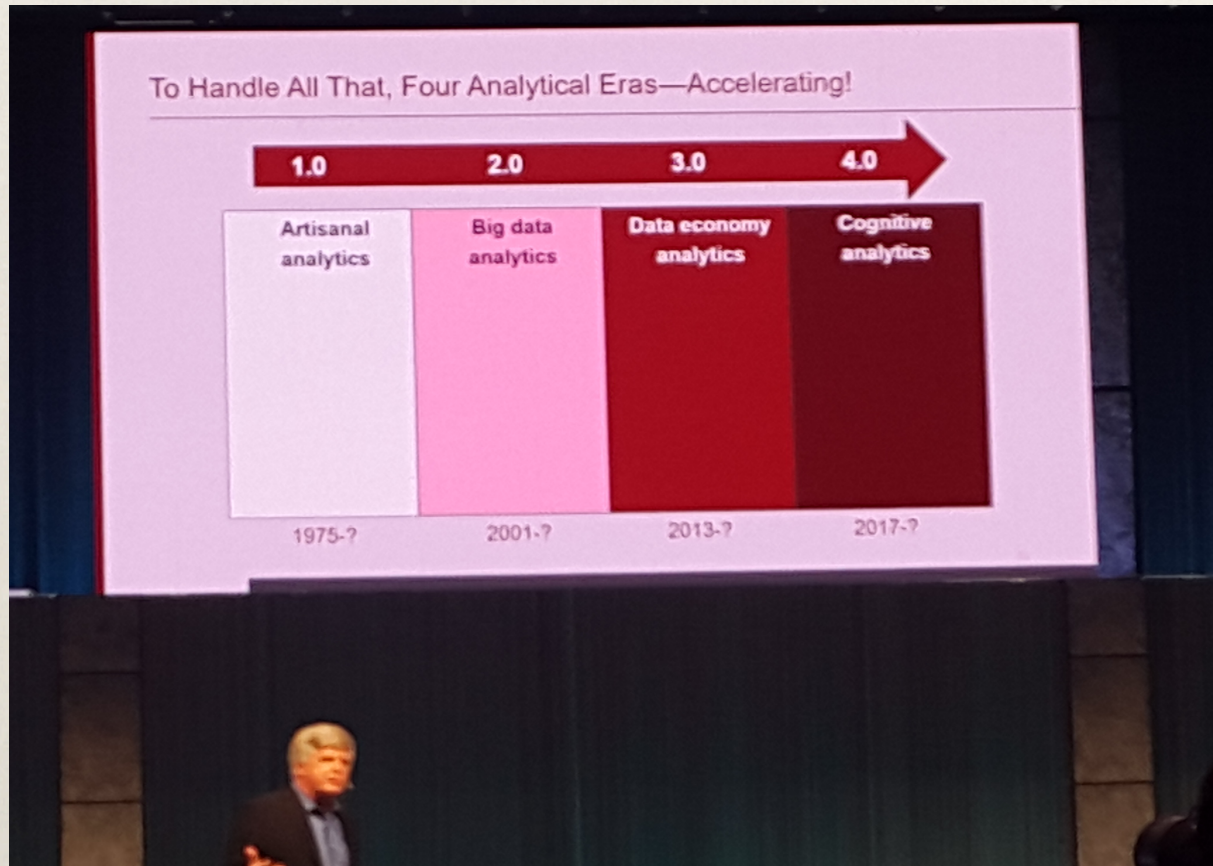
**To responsibly unleash the power  
of data to benefit all Americans**



# Tom Davenport

(<http://www.tomdavenport.com/>)

## Evolution of Data Science





# Where is all this leading to?

- \* Data Science deals with a much broader aspect of data than any Statistics, Mathematics, Computer Science or any other academic field ever dealt with before.
- \* The theoretical and practical questions in Data Science come from many fields.
- \* Data Science is becoming part of the bag of tools of all academic disciplines.
- \* **The price of light is less than the cost of darkness.** *Arthur C. Nielsen, Market Researcher & Founder of ACNielsen.*

# Conferences and previous successful programs at Buffalo State

- \* DAAD conference (2015), DAAD conference (2016) and several others.
- \* PSM in Professional Applied and Computational Mathematics (2010).
- \* Some Data Analytics curriculum at Buffalo State
  1. Introduction to Data Science
  2. Statistics for Data Science
  3. Mathematics for Data Science
  4. Geospatial Programming

# Professional Science Master's Data Analytics Across Disciplines (General structure)

- \* Math/Statistics
- \* Computer Science
- \* Business skills. Experiential components. Internships
- \* -----
- \* Strong ties between academic, government, non-profit and professional sectors permeates in all the curriculum
- \* Strong ties between different disciplines
- \* Entrepreneurial mind set

# References

1. William S. Cleveland. "Faculty Page". Retrieved 6 July 2012.
2. Cleveland, W. S. (2001). "Data Science: An Action Plan for Expanding the Technical Areas of the Field of Statistics". *International Statistical Review / Revue Internationale de Statistique* 69 (1).
3. "International Council for Science : Committee on Data for Science and Technology". Retrieved 6 July 2012.
4. "CODATA Data Science Journal". Volume 1, Issue 1. Retrieved from Japan Science and Technology Information Aggregator. April 2002. Retrieved 6 July 2012.
5. "The Journal of Data Science". Volume 1, Issue 1. Columbia University. January 2003. Retrieved 6 July 2012.

# Thanks for being here!

- \* Open discussion