

# The Role of Visualization Capacity Building in Data Science

Vetria L. Byrd, PhD

Assistant Professor

Computer Graphics Technology

Purdue University

Liberal Arts Data Science Workshop 2018  
New College Florida

January 12-13, 2018



# About Me

Vetria L. Byrd, PhD

## Academic Preparation

- Computer Science (PhD, MS)
- Biomedical Engineering (MSMBE)



## What I Am Doing Now

### Academic Appointment

- Assistant Professor
- Purdue University
- Computer Graphics Technology
- Curriculum Development for New Major in Data Visualization
- Research Focus: Data Visualization

## What I've Done



### Visualization Initiatives

- BPViz: Broaden Participation in Visualization (2014/2016/2017)
- Research Experience for Undergraduates in Collaborative Data Visualization Applications (2014/2015)



PURDUE  
POLYTECHNIC

# About Me: Since Joining Purdue

## Assistant Professor

Purdue University/Polytechnic Institute  
Computer Graphics Technology

**Teaching:** Curriculum Development for new undergraduate major in Data Vis

**Research:** Data Visualization, Big Data Analytics, Integration and Visualization of Heterogeneous Data



ACCELERATING THE BIG DATA INNOVATION ECOSYSTEM

Steering Committee Member

## International HPC Summer School on HPC Challenges in Computational Sciences

Toronto, Canada (2015), Ljubljana, Slovenia (2016), Boulder, CO, US (2017), Prague (2018)



Open Data Explore With Us Data Stories Innovation Space About

## Daternauts

Reach higher and explore deeper. Whether you're a software engineer or coding newbie, join our NASA Datonaut community to engage with each other and subject matter experts to solve data challenges. It really IS Rocket [Data] Science.



Water Cooler Chat



Visualization Webinars



Teaching & Learning Technology Conference 2016



*Agent for “Insight”*

**PURDUE**  
POLYTECHNIC

**Q**

# How would you define data visualization?



# What is Data Visualization?

<http://mindymcadams.com/tojou/2011/10-useful-resources-about-data-visualization/>

Last accessed 02/27/17

- Representing large amounts of disparate information in a visual form often allows you to see patterns that would otherwise be buried in vast, unconnected data sets. ...
- Visualizations allow you to understand and process enormous amounts of information quickly because it is all represented in a single image or animation.

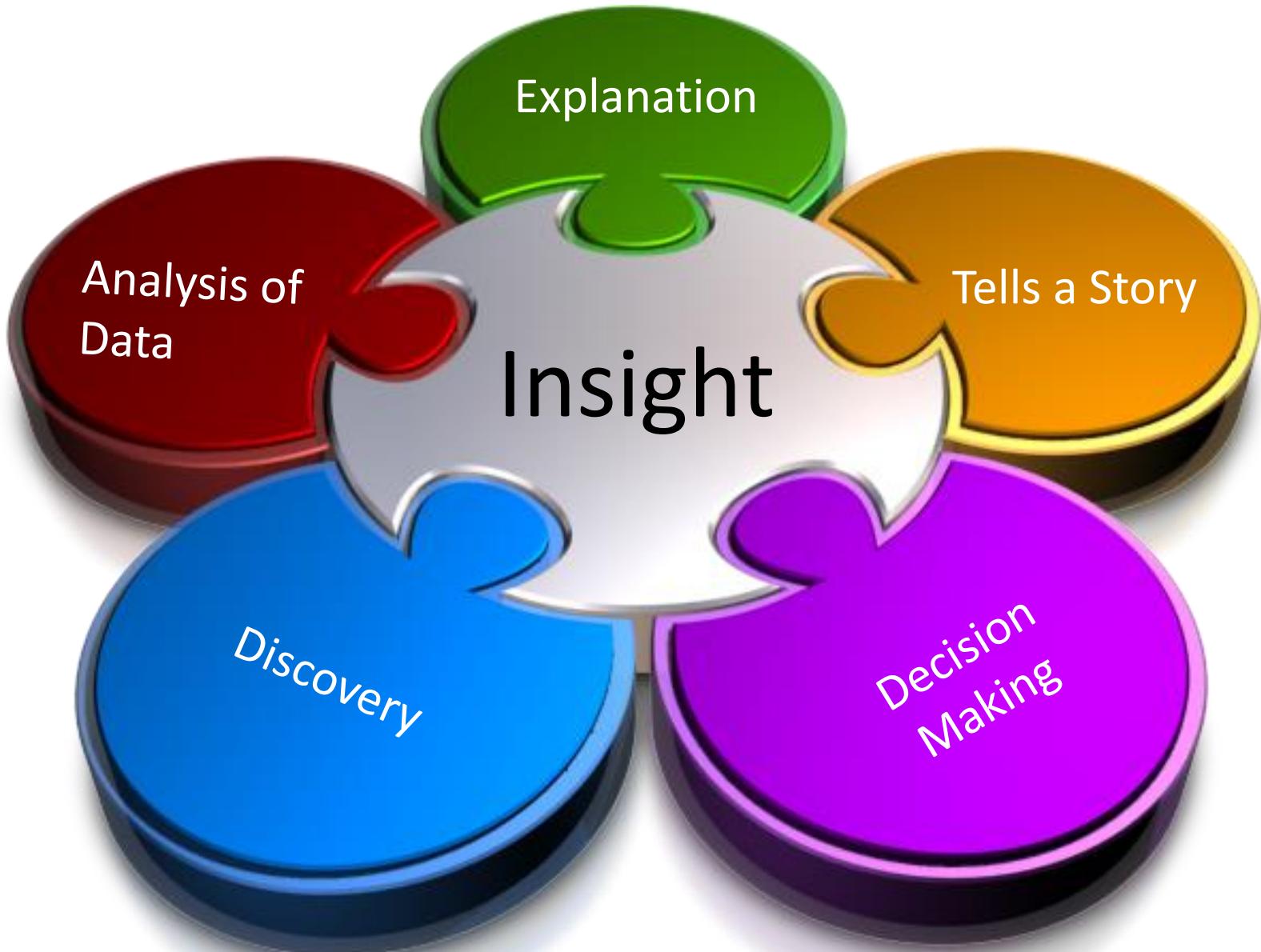
# Data Visualization

A **process** of transforming raw, complex data into a visual representation that does not overwhelm the viewer.



“The purpose of  
visualization  
is “*insight*”,  
not pictures.”

~*Ben Shneiderman*



# FROM DATA TO INSIGHT

Advancing Beyond Data to True Insight

Relationship



Data

Relevance

Source: Ackoff, Russell L., "From Data to Wisdom", Journal of Applied Systems Analysis, Volume 16, 1989 p 3-9.

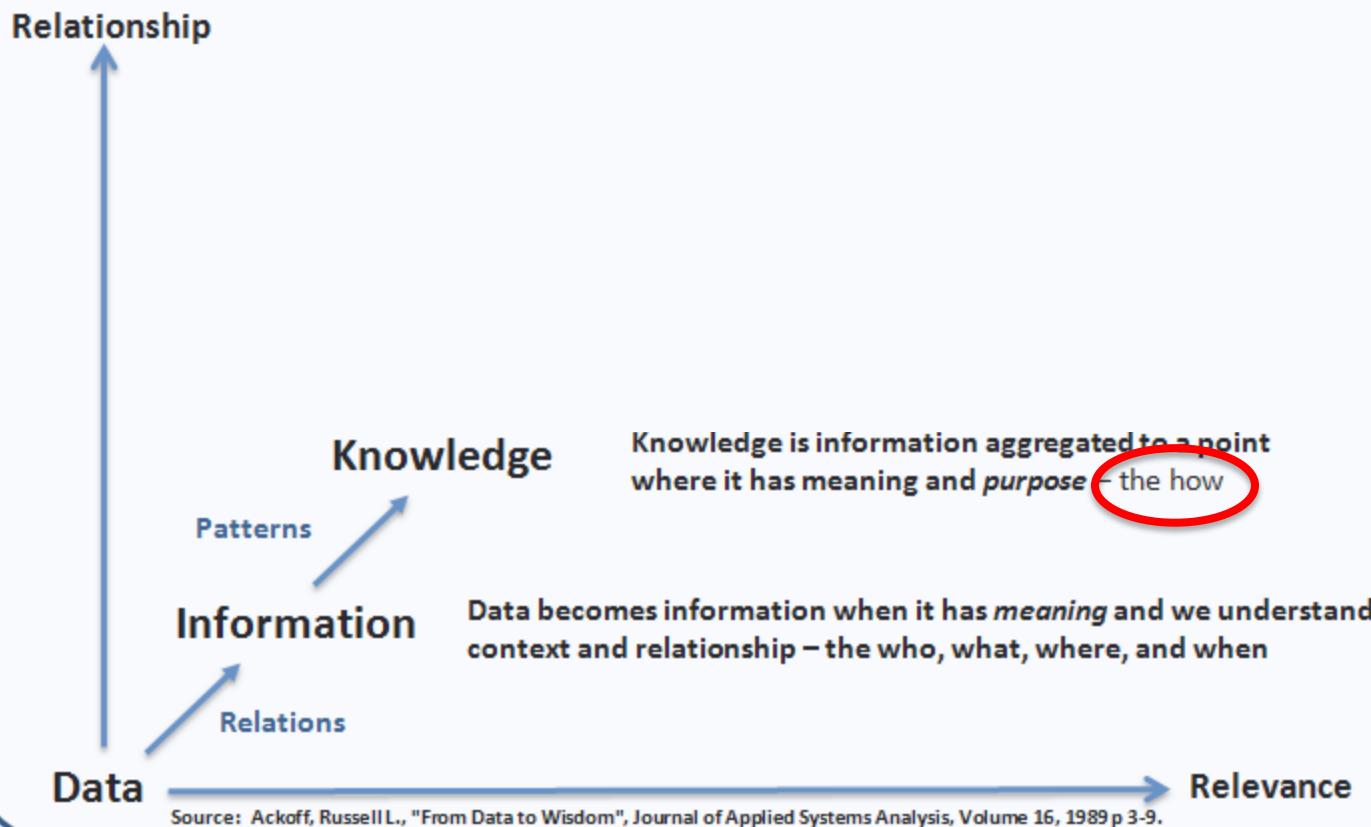
# FROM DATA TO INSIGHT

## Advancing Beyond Data to True Insight



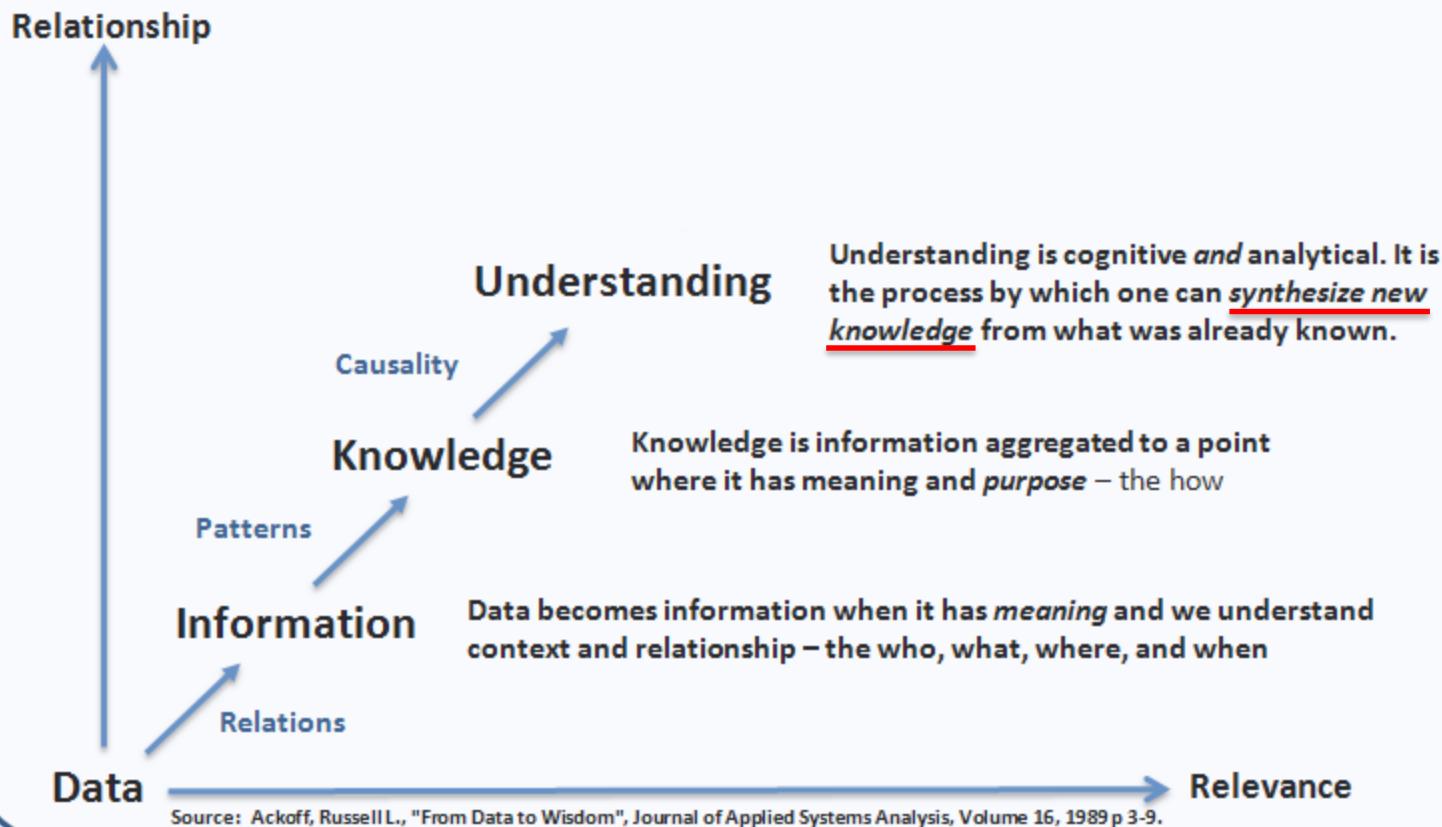
# FROM DATA TO INSIGHT

## Advancing Beyond Data to True Insight



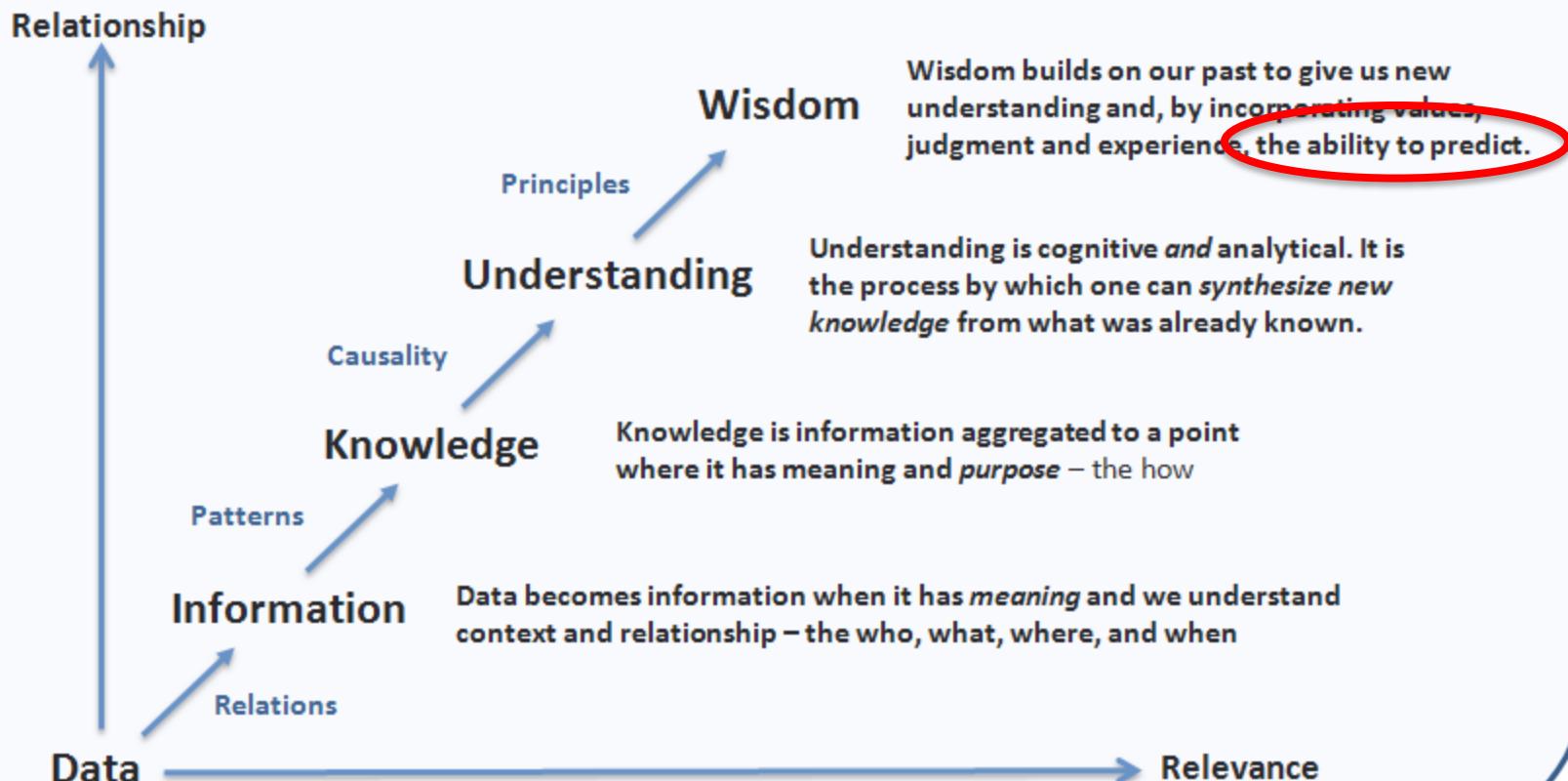
# FROM DATA TO INSIGHT

## Advancing Beyond Data to True Insight



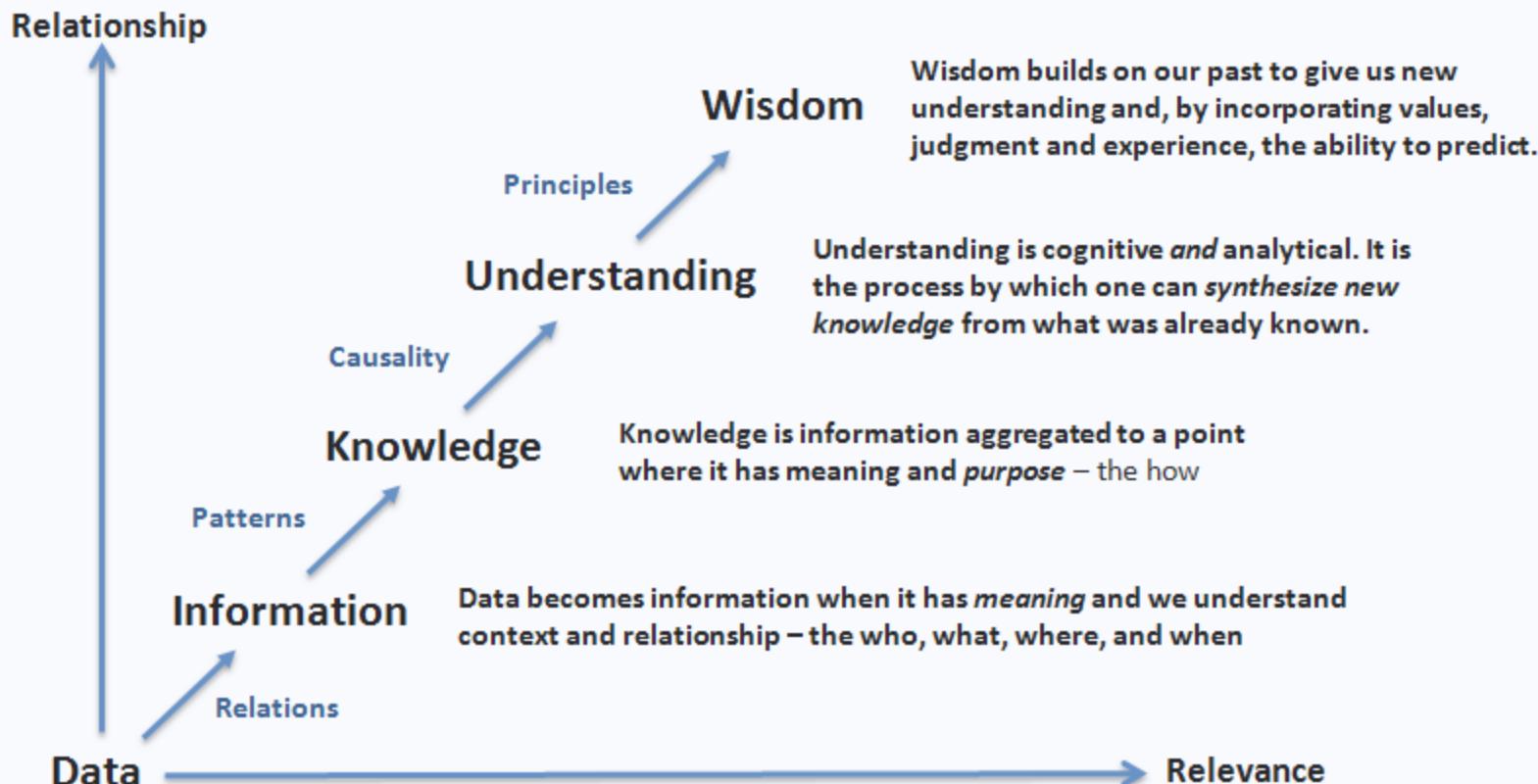
# FROM DATA TO INSIGHT

## Advancing Beyond Data to True Insight



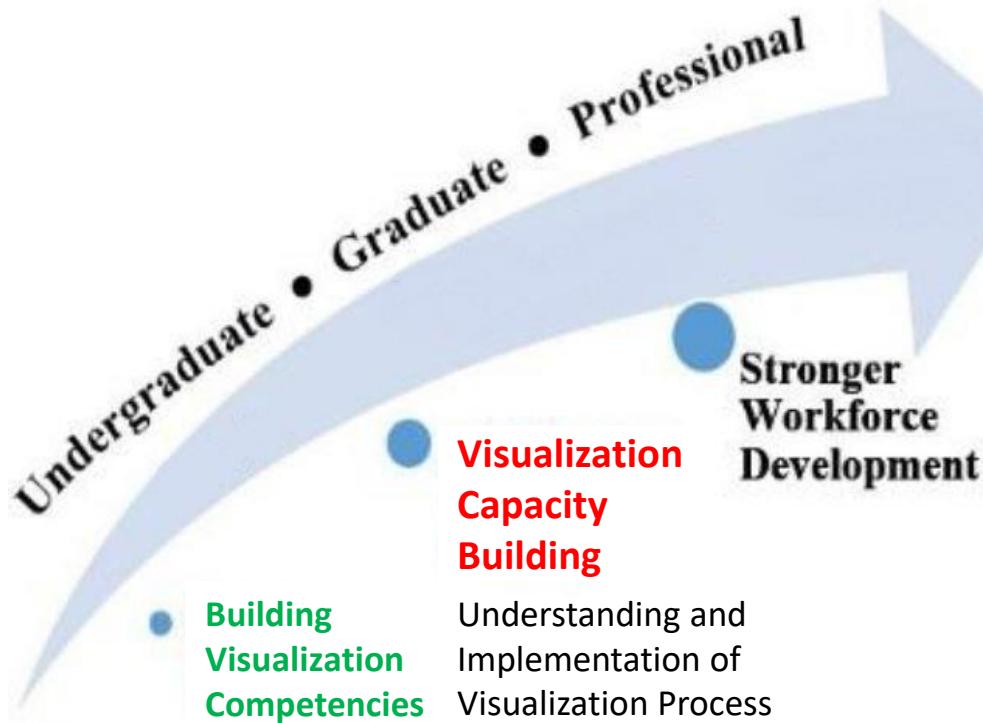
# FROM DATA TO INSIGHT

## Advancing Beyond Data to True Insight



Source: Ackoff, Russell L., "From Data to Wisdom", Journal of Applied Systems Analysis, Volume 16, 1989 p 3-9.

# Visualization Capacity Building



Visual Literacy

# Why Should We Care about Data Visualization?

Regardless of major, research interest, or academic background, at some point you will visualize some type of data.

*~ Vetria Byrd, PhD*



# Data is Everywhere!

Image Source: <http://www.centrodeinnovacionbbva.com/en/news/practical-examples-big-data-use>



# The Growth of Scientific Data

- Global growth trend of data volume, 2006–2020
- Based on “The digital universe in 2020: big data, bigger digital shadows, and biggest growth in the far east”)

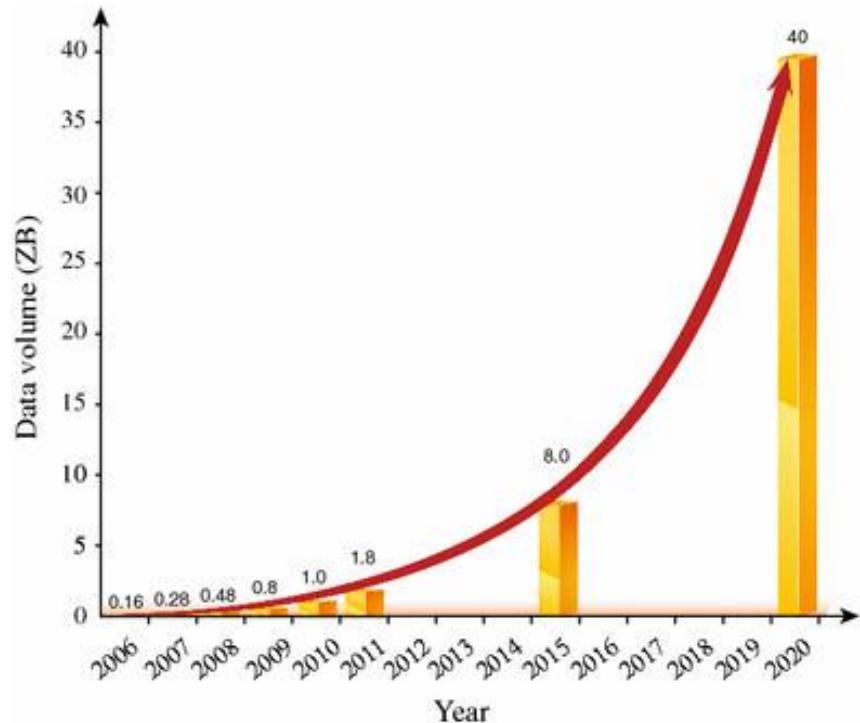
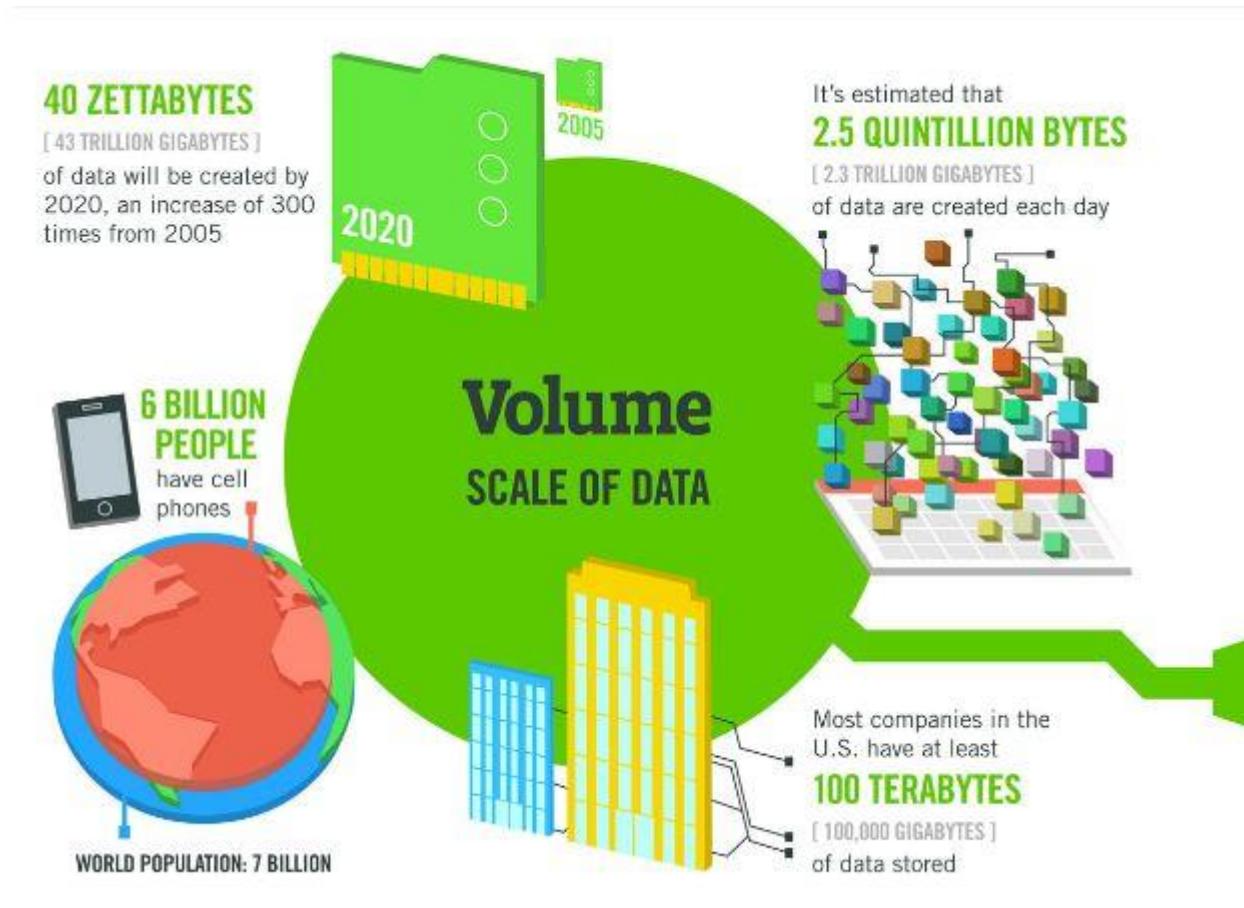


Image Source: [https://www.researchgate.net/publication/274233315\\_Scientific\\_big\\_data\\_and\\_Digital\\_Earth](https://www.researchgate.net/publication/274233315_Scientific_big_data_and_Digital_Earth)

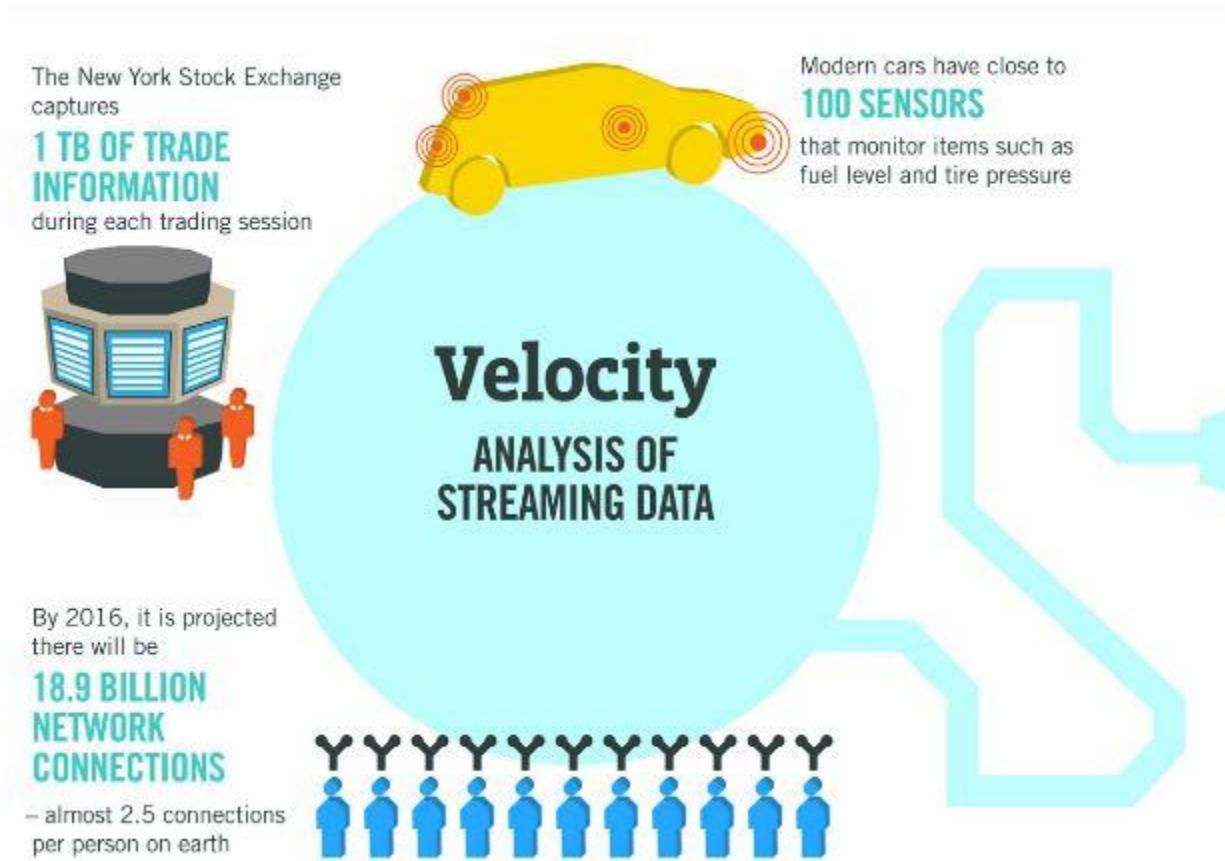
ZB: a unit of information equal to one sextillion (10<sup>21</sup>) or, strictly, 270 bytes

# The Four V's of Big Data



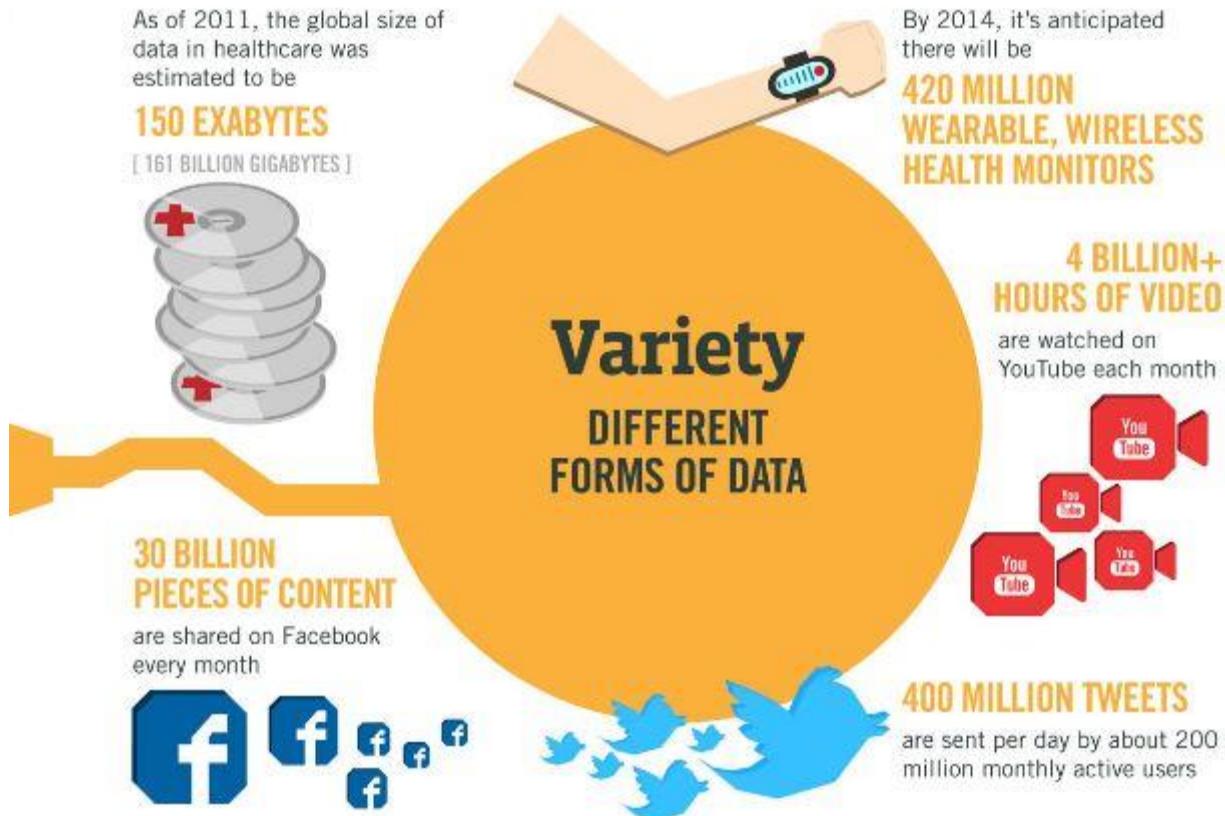
<https://www-01.ibm.com/software/data/bigdata/images/4-Vs-of-big-data.jpg>

# The Four V's of Big Data



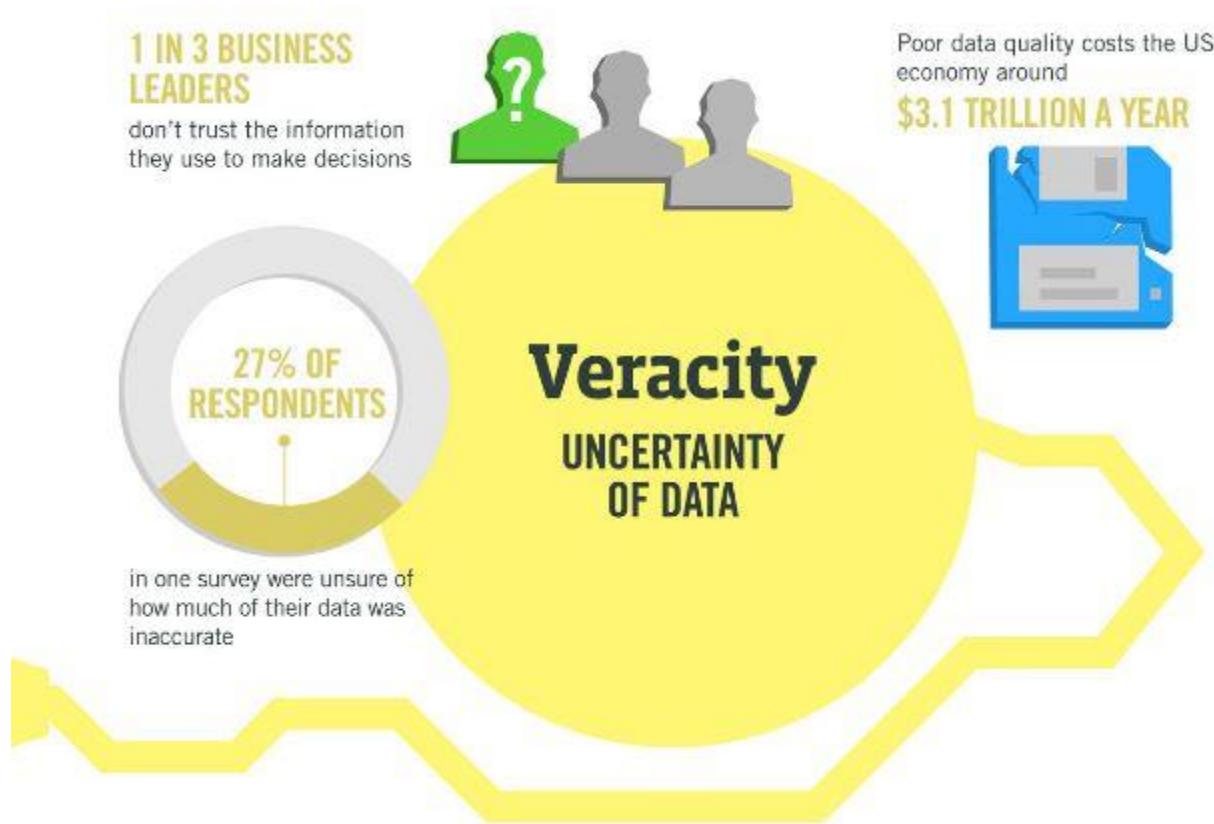
<https://www-01.ibm.com/software/data/bigdata/images/4-Vs-of-big-data.jpg>

# The Four V's of Big Data



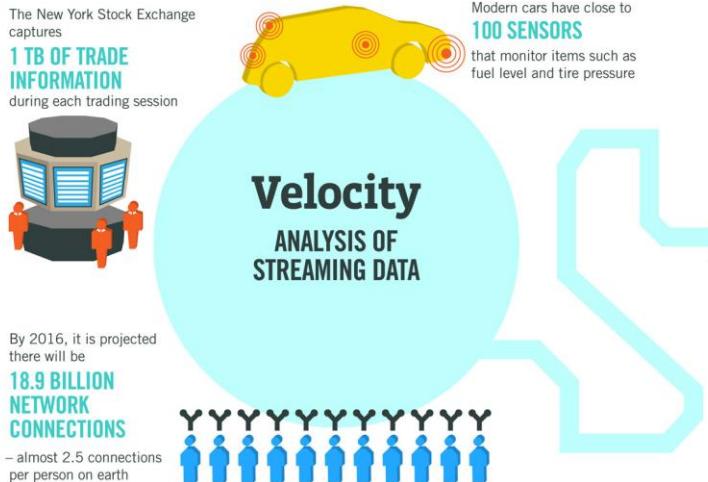
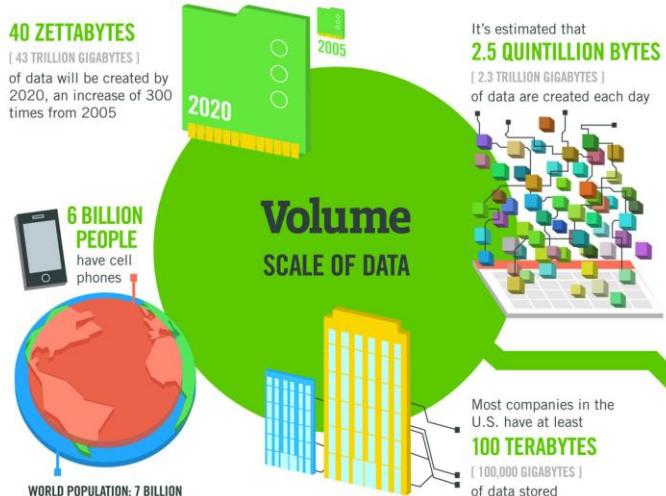
<https://www-01.ibm.com/software/data/bigdata/images/4-Vs-of-big-data.jpg>

# The Four V's of Big Data



<https://www-01.ibm.com/software/data/bigdata/images/4-Vs-of-big-data.jpg>

# IBM Big Data Platform



## The FOUR V's of Big Data

From traffic patterns and music downloads to web history and medical records, data is recorded, stored, and analyzed to enable the technology and services that the world relies on every day. But what exactly is big data, and how can these massive amounts of data be used?

As a leader in the sector, IBM data scientists break big data into four dimensions: **Volume, Velocity, Variety and Veracity**.

Depending on the industry and organization, big data encompasses information from multiple internal and external sources such as transactions, social media, enterprise content, sensors and mobile devices. Companies can leverage data to adapt their products and services to better meet customer needs, optimize operations and infrastructure, and find new sources of revenue.

By 2015  
**4.4 MILLION IT JOBS** will be created globally to support big data, with 1.9 million in the United States

As of 2011, the global size of data in healthcare was estimated to be

**150 EXABYTES**

[161 BILLION GIGABYTES]

**30 BILLION PIECES OF CONTENT**

are shared on Facebook every month



### Variety DIFFERENT FORMS OF DATA

By 2014, it's anticipated there will be **420 MILLION WEARABLE, WIRELESS HEALTH MONITORS**

**4 BILLION+ HOURS OF VIDEO** are watched on YouTube each month



**400 MILLION TWEETS** are sent per day by about 200 million monthly active users

### 1 IN 3 BUSINESS LEADERS

don't trust the information they use to make decisions



**27% OF RESPONDENTS**

in one survey were unsure of how much of their data was inaccurate

### Veracity UNCERTAINTY OF DATA

Poor data quality costs the US economy around **\$3.1 TRILLION A YEAR**



# WHY SHOULD WE CARE?

There is a demand for persons

- who understand the visualization process
- is able to transform raw complex data into a visual representation that does not overwhelm.
- who can present data in such a way that anyone can understand the implications of the data.



Since 1987 - Covering the Fastest Computers In the World and the People Who Run Them



**Byrd Emphasizes Value of Visualization at XSEDE14**

By Trish Barker

ATLANTA, MARQUES

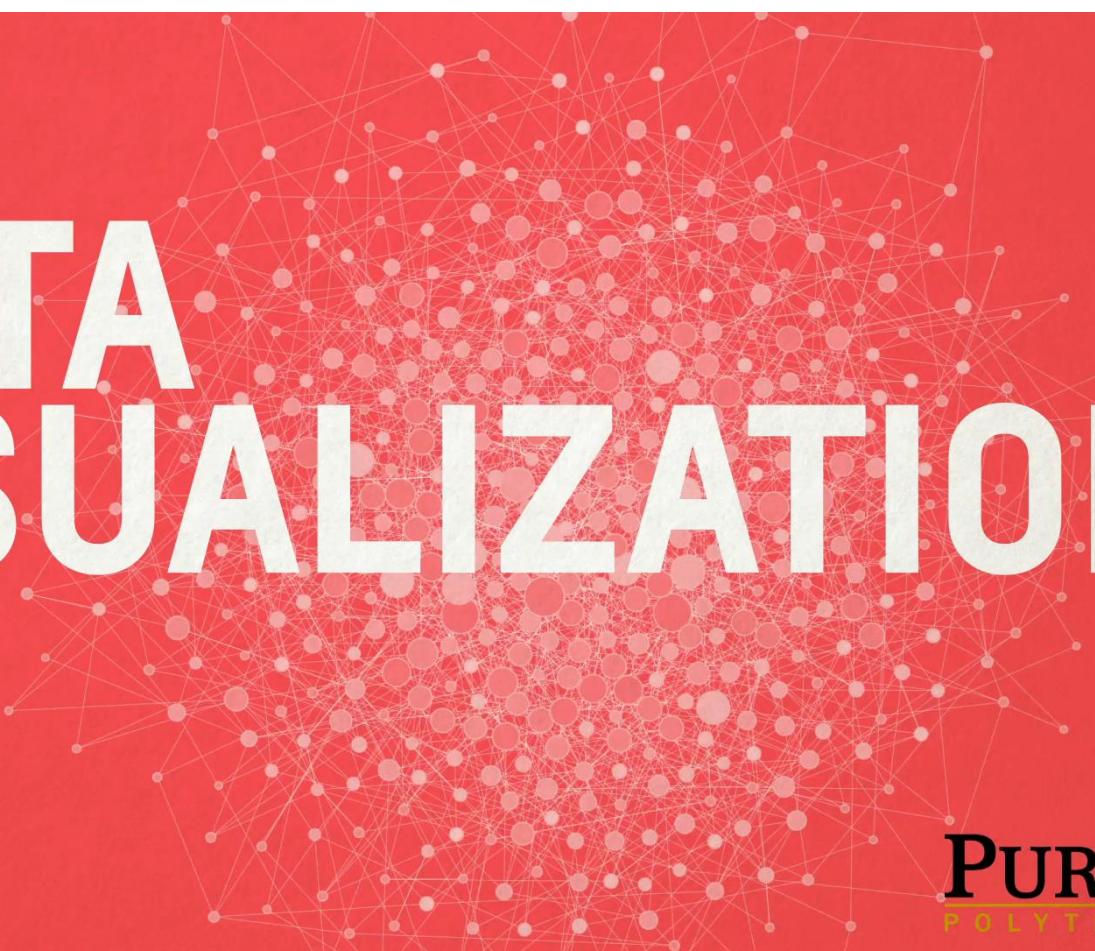
<https://www.hpcwire.com/2014/07/31/byrd-emphasizes-value-visualization-xsede14/>

July 31, 2014

# Data Visualization + Data Science = Insight

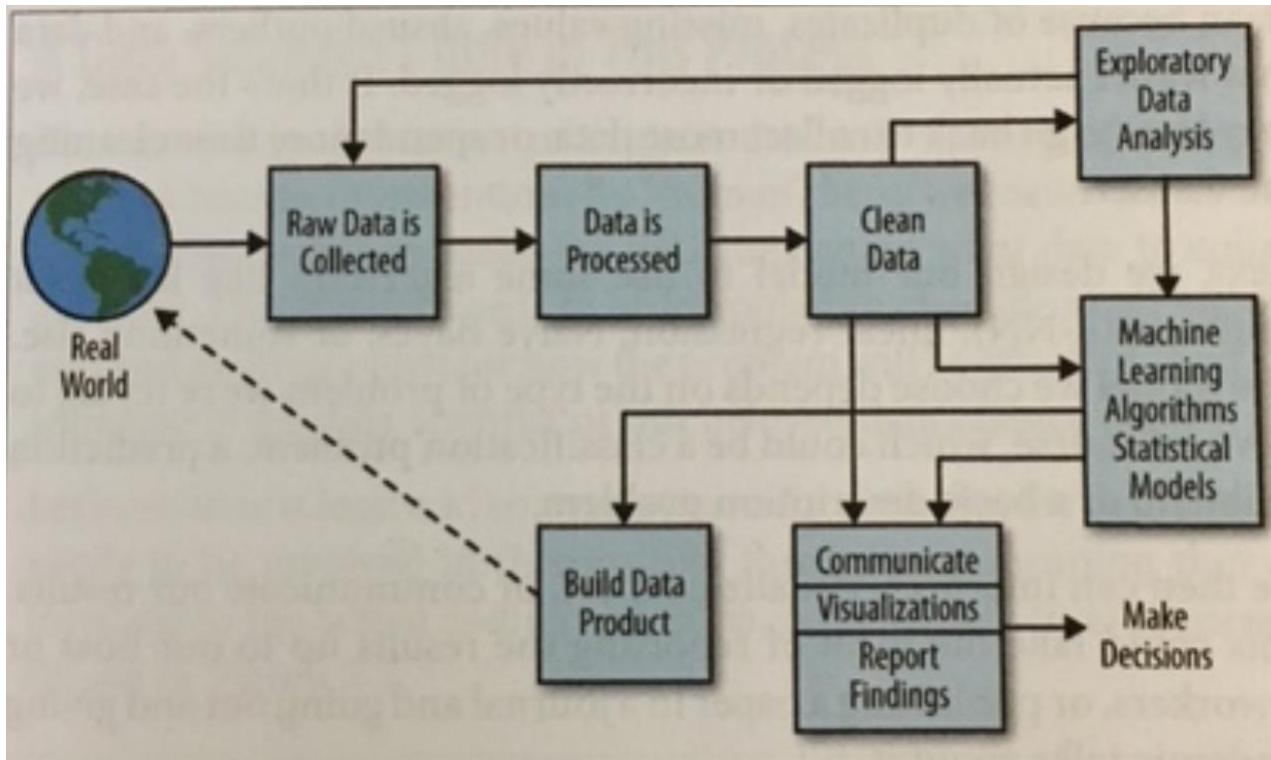


# DATA VISUALIZATION

A large, abstract network graph is positioned in the center-right area of the slide. It consists of numerous small white dots connected by thin gray lines, forming a complex web-like structure. In the lower half of the graph, there are several larger, semi-transparent red circles of varying sizes, creating a sense of depth and highlighting specific data points or clusters.

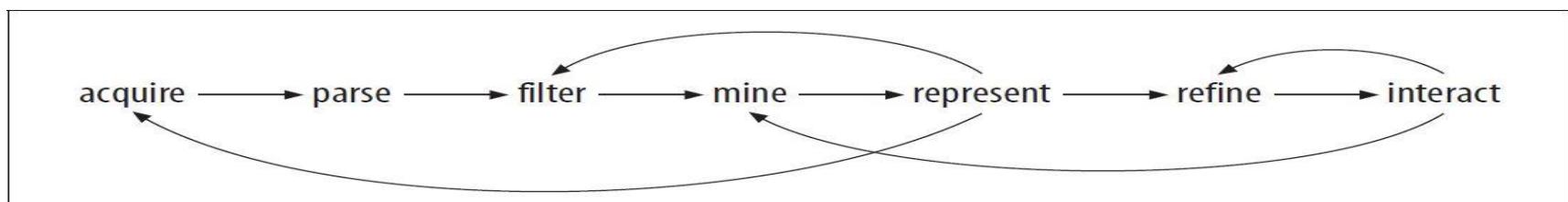
**PURDUE**  
POLYTECHNIC

# The Data Science Process



Schutt, R., & O'Neil, C. (2013). *Doing data science: Straight talk from the frontline.* " O'Reilly Media, Inc.". Figure 2-2, page 41.

Fry, B. (2007). *Visualizing data: Exploring and explaining data with the processing environment.* " O'Reilly Media, Inc.".



# Visualization Initiatives

Viz REU Site  
**BPViz**



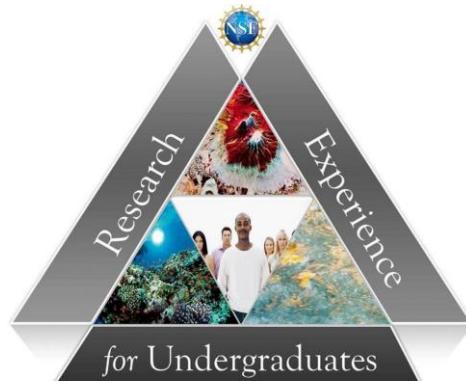
# Viz REU Site

## Research Experience for Undergraduates In Collaborative Data Visualization Applications

Host Cite: Clemson University, Clemson SC

Vetria L. Byrd, PI

Funded by NSF ACI Award 1359223

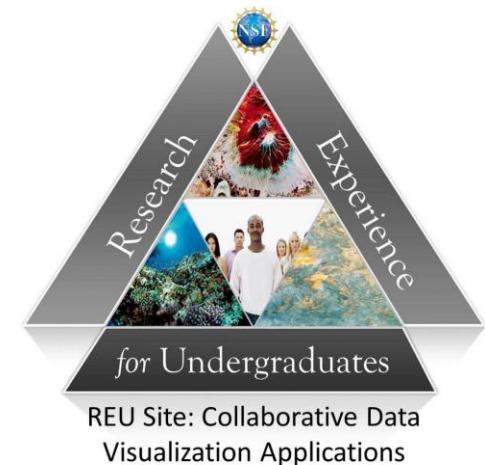


REU Site: Collaborative Data  
Visualization Applications



# Viz REU Site

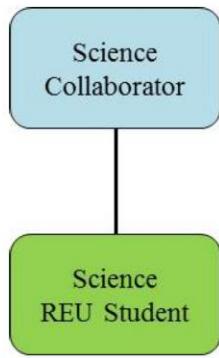
- 8 – weeks on site
- 10 – 12 Students from US Universities
- Paid Internship
- Interdisciplinary Research
- Work closely with domain expert and visualization experts
- Workshops on preparing for graduate school, funding opportunities
- Undergraduates
  - Sophomores and Juniors
  - Very academically mature freshmen



# What was Unique about the Viz REU Site

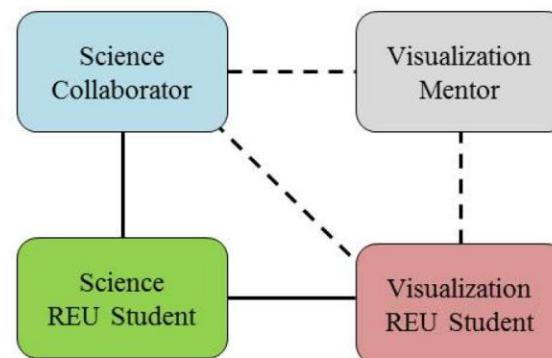
- Primary Focus: Data Visualization

Typical REU Program



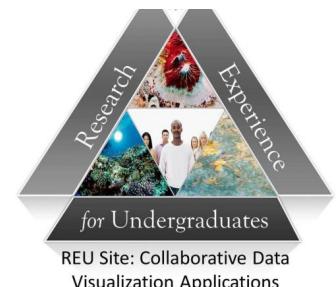
(A)

Viz REU Site



(B)

Figure 3. (A) Independent science collaborator with science REU student, (B) The REU Team. Dashed lines indicate the REU mentoring detailed in this REU. Solid lines indicate collaboration with a larger research group.



# Student Demographics

ALL students, from ALL backgrounds

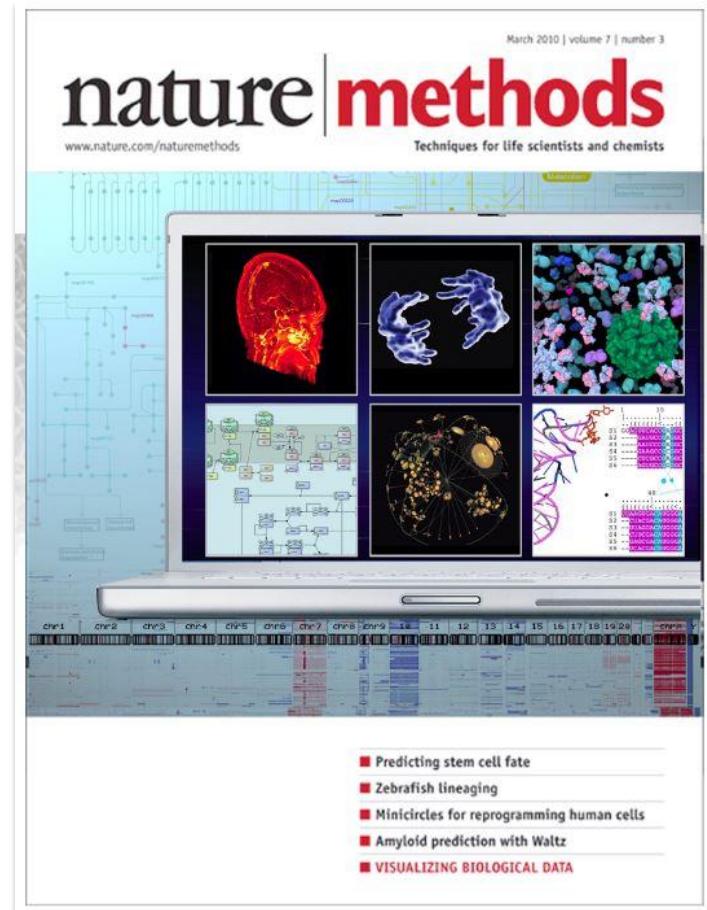
- Computer Science
- Engineering
- Digital Humanities
- Chemistry
- Statistics
- Mathematics
- Physics
- Biology
- More . . .

I specifically try to create teams of students from different backgrounds

# Visualization Applications

## Biovisualization (BioVis)

The visualization of biological data;  
Often grouped with computer animation



March 2010 | volume 7 | number 3

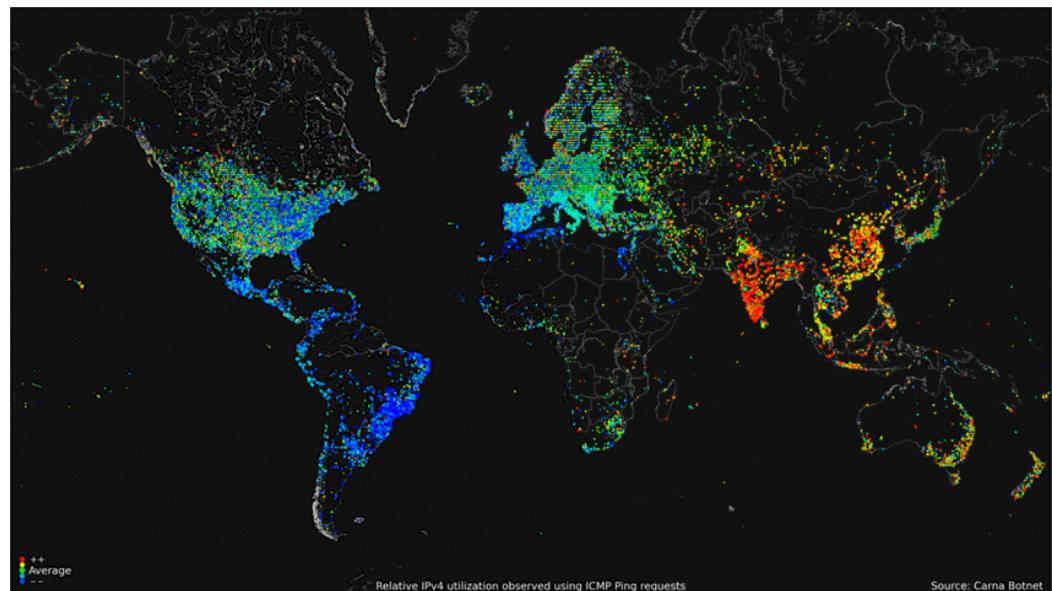
# Visualization Applications

## Information Visualization (InfoVis)

Interdisciplinary  
Study of the “visual  
representation of  
large-scale collections  
of non-numerical  
information



Social Media Data  
Survey Data  
Observed Data



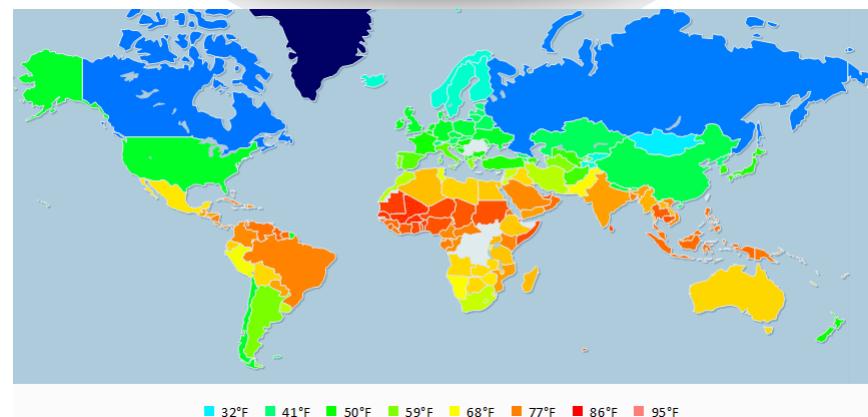
Internet Usage

Source: <http://www.cerneia.net/wp-content/uploads/2013/03/internet.gif>

# Visualization Applications

## Geographic Visualization

Communicates geospatial information in ways that, when combined with human understanding, allow for data exploration and decision-making processes.



MacEachren, A.M. and Kraak, M.J. 1997 Exploratory cartographic visualization: advancing the agenda. *Computers & Geosciences*, 23(4), pp. 335-343. Jiang, B., and Li, Z. 2005. Editorial: Geovisualization: Design, Enhanced Visual Tools and Applications. *The Cartographic Journal*, 42(1), pp. 3-4. MacEachren, A.M. 2004. Geovisualization for knowledge construction and decision support. *IEEE computer graphics and applications*, 24(1), pp.13-17

# Visualization Applications

## Scientific Visualization (SciVis)

Primarily concerned with the visualization of three-dimensional phenomena

Emphasizes on realistic renderings of volumes, surfaces, illumination sources, etc.

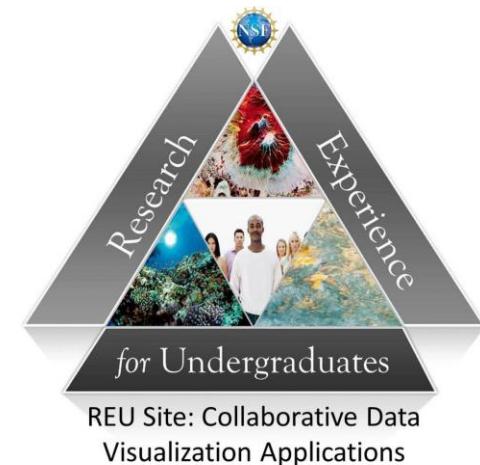


Image Source:

<http://www.sci.utah.edu/the-institute/highlights/24-research-highlights/cibc-highlights/253-top-scientific-visualization-research-problems.html>

# Participating Departments

- Genetics and Biochemistry
- Geophysics
- Sociology
- Molecular Modeling and Simulation
- Plant Population and Community Ecology
- Inorganic Chemistry
- Parks Recreation Tourism Management
- Social Media, Social Media Listening, Social Networks, Management of Information Systems
- Computer Science
- Physical Chemistry
- Digital Humanities
- Anthropology and Sociology
- Biological Sciences
- School of Education
- Electrical and Computer Engineering
- Civil Engineering
- Genetics & Biochemistry
- Sports Information
- Molecular and Cellular Biology
- And More . . .



# Viz REU Cohorts

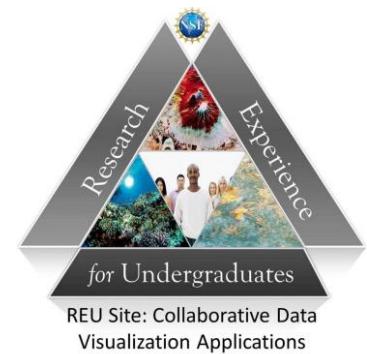
2014



2015



- 22 Students Participated
- Several have matriculated to graduate school (MIT, Georgia Tech, UC Irvine)
- One accepted to the PhD program at Purdue in the Byrd Vis Lab (Fall 2018)
- Data Visualization Scientist at Northwestern



# The Viz REU Experience

- Mentoring
- Networking
- Professional Development
- Elevator Pitch
- Collaboration
- Uncovering the magic behind the resources



# The Viz REU Experience

## It is NOT a Summer Vacation

- Introduction to Research
  - CITI Training
  - Research Methods Seminars
- Introduction to Data Visualization
  - Visualization Process
  - Visualization Principles, Methods, Tools
  - How Not to lie with visualization
  - Guest speakers
- Technical Writing
- Round Table with CIO
- Professional Presentations
  - XSEDE Student Program
  - SC: Supercomputing Conference
  - Regional Presentations (Off campus)



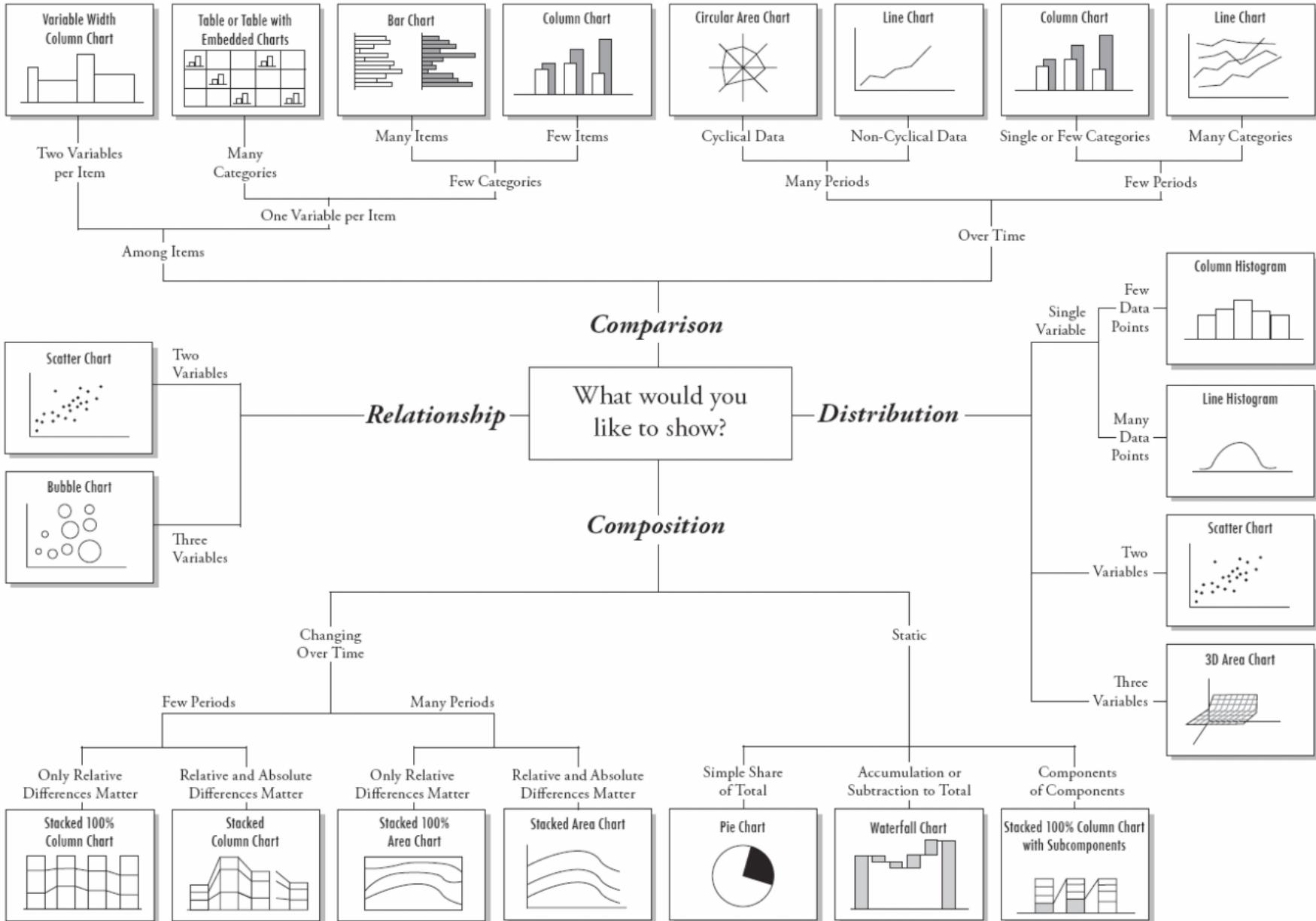
# The Data Dictates Which Tool is Used

Each research project is assessed and visualization tools determined as the visualization needs are determined.

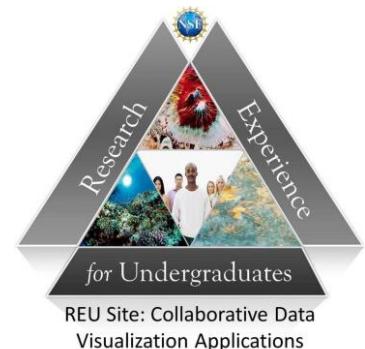
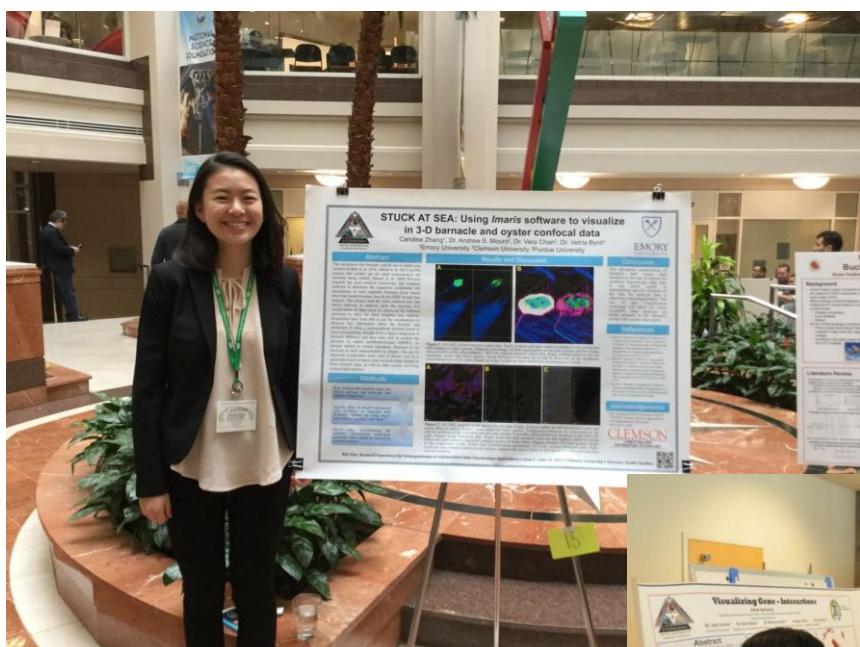
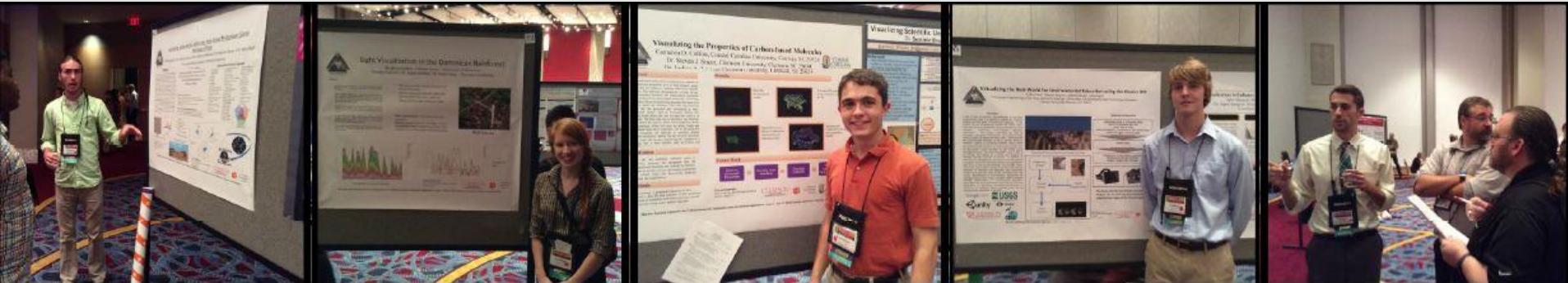
- Open source tools (mostly)
- Information Visualization (Tableau)
- Scientific Visualization (ParaView, VisIt)
- Others as needed

# Programming Tools for Visualization

- Processing (recommended for students with no programming experience)
- Data Driven Documents (D3.js)
- Python (interfaces with ParaView and VisIt)
- Any others that students are comfortable with



# 2014 Viz REU Cohort



REU Site: Collaborative Data  
Visualization Applications

# 2014 Student Conference Poster Presentations

**Burns, C.**, DeWalt, S., Ickes, K. Light visualization in the Dominican Rainforest. Submitted to 2015 Emerging Researchers National (ERN) Conference in STEM. The ERN Conference is cosponsored by the American Association for the Advancement of Science (AAAS), Education and Human Resources Programs (EHR) and the NSF Division of Human Resource Development (HRD), within the Directorate for Education and Human Resources (EHR). Washington, D.C., on February 19-21, 2015.

**Sampong, A.**, Levine, J., Feltus, A., Smith, M., Joshi, A., Sapra K. Visualizing Gene - Interactions within the Rice and Maize Network. 2014 AGU Fall Meeting, December 15-19, 2014, San Francisco, CA.

**Sampong, A.**, Feltus, A., Levine, J., Byrd, V. Using JavaScript to View the Gene Interactions in the Rice and Maize Networks. 2014 National Association of Black Geoscientists (NABG) Meeting, September 17-20, 2014, Richland, WA.

**Bosch, M.**, Moysey, S., Mobley, C., Boyer, M., Byrd, V. (VisMentor). Expressing Sustainability within the Web-Based Multiplayer Game Naranpur Online. CUR REU Symposium, October 26-27, 2014, Arlington, VA

**Bosch, M.**, Moysey, S., Mobley, C., Boyer, M., Byrd, V. (VisMentor). Expressing Sustainability within the Web-Based Multiplayer Game Naranpur Online. XSEDE14 Student Poster Presentation, July 2014, Atlanta, GA.

**Slonecki, T.**, Sarupria, S., Levine, J. (VisMentor). Visualization to Enhance Rare Event Simulations of Ice Nucleation. XSEDE14 Student Poster Presentation, July 2014, Atlanta GA.

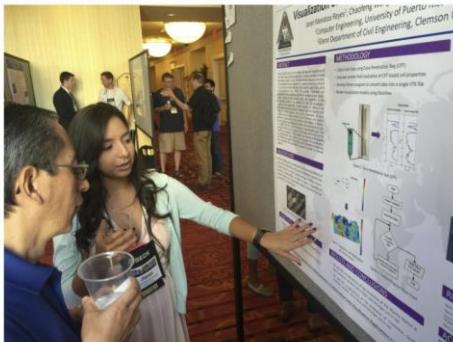
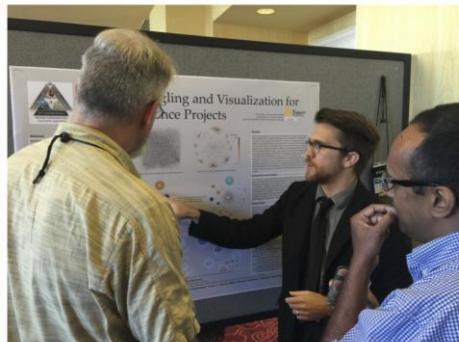
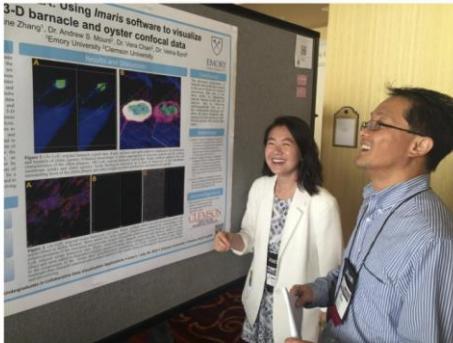
**Burns, C.**, DeWalt, S., Byrd, V. (VisMentor). Light Visualization in the Dominican Rainforest. XSEDE14 Student Poster Presentation, July 2014, Atlanta GA.

**Collins, C.**, Stuart, S., Levine, J. (VisMentor). Visualizing the Properties of Carbon-based Molecules. XSEDE14 Student Poster Presentation, July 2014, Atlanta, GA.

**Fant, L.**, Moysey, S., Boyer, M., Byrd, V. (VisMentor). Virtualizing the Real-World for Environmental Education using the Oculus Rift. XSEDE14 Student Poster Presentation, July 2014, Atlanta, GA.

# 2015 Cohort Conference Presentations

XSEDE 2015



2015 VisREU Cohort at CSS/REU Mini-Conference  
College of Charleston  
July 11, 2015

# 2015 Conference Talks

## REUNS/IEEE MASS 2015 Conference

**O'Brien, S.**, Song, B. *Colorful Math: Using LIDAR Data to Three-Dimensionally Visualize and Analyze the Dynamics of a Deciduous Tree*. REUNS/IEEE MASS 2015 Conference, Dallas, TX, October 19-22, 2015.

**Salazar, C.**, Giacalone, K., and Mobley, C. *Is the Glass Half Full or Half Empty? Is the Water Murky or Clear?* REUNS/IEEE MASS 2015 Conference, Dallas, TX, October 19-22, 2015.

**James, J.**, Kallin, J., and Gantt, J. *The Virtual Athletic Event*. Clemson Booth Presentation, Super Computing 2015, Austin, TX, November 15-20, 2015.

## NC/SC REU Site Conference

**Cox, S.**, Kallin, J., and Gantt, J. (Research Mentors), Byrd, V.L. (Vis Mentor). *How to improve representation of sports statistics?* NC/SC REU Site Conference, College of Charleston, Charleston, South Carolina, July 11, 2015.

**Elavsky, F.**, Hankins, G. (Research Mentor), Gemmill, J. (Vis Mentor). *Digital Wrangling and Visualization for Correspondence Projects*. NC/SC REU Site Conference, College of Charleston, Charleston, South Carolina, July 11, 2015.

**Hutula, K.**, Riggs, S. (Research Mentor), Gemmill, J. (Vis Mentor). *Using Visualization to Effectively Capture Multimodal Data Over Time*. Clemson University, Clemson, South Carolina.

**James, J.**, Kallin, J., and Gantt, J. (Research Mentors), Byrd, V.L. (Vis Mentor). *The Virtual Athletic Event*. NC/SC REU Site Conference, College of Charleston, Charleston, South Carolina, July 11, 2015.

**Kyle, K.**, Feltus, A., and Smith, M. (Research Mentors), Gemmill, J. (Vis Mentor). *Detangling genetic "hairballs": implementing an abstracted, gene cluster view for the gPICTviz visualization tool*. NC/SC REU Site Conference, College of Charleston, Charleston, South Carolina, July 11, 2015.

**Mendoza, J.**, Chen, Q. (Research Mentor), Byrd, VL (Vis Mentor). *3D visualization of soil data of seismic hazard analysis*. NC/SC REU Site Conference, College of Charleston, Charleston, South Carolina, July 11, 2015.

**O'Brien, S.**, Song, B. (Research Mentor), Byrd, V.L. (Vis Mentor). *Colorful Math: Using LIDAR Data to Three-Dimensionally Visualize and Analyze the Dynamics of a Deciduous Tree*. NC/SC REU Site Conference, College of Charleston, Charleston, South Carolina, July 11, 2015.

**Salazar, C.**, Giacalone, K., and Mobley, C. (Research Mentors), Byrd, V.L. (Vis Mentor). *Is the Glass Half Full or Half Empty? Is the Water Murky or Clear?* NC/SC REU Site Conference, College of Charleston, Charleston, South Carolina, July 11, 2015.

**Smith, P.**, Mount, A. (Research Mentor), Byrd, V.L. (Vis Mentor). *Visualization of Jupiter's Radio Storms Using Worldwide Telescope*. NC/SC REU Site Conference, College of Charleston, Charleston, South Carolina, July 11, 2015.

**Warner, B.**, White, DL (Research Mentor), Byrd, VL (Vis Mentor). *Developing Analytic Web Tools for Environmental Data Using R Shiny*. NC/SC REU Site Conference, College of Charleston, Charleston, South Carolina, July 11, 2015.

**Zhang, C.**, Mount, A. (Research Mentor), Byrd, V.L. (Vis Mentor). *Using Imaris software to visualize in 3-D the mechanism of barnacle attachment*. NC/SC REU Site Conference, College of Charleston, Charleston, South Carolina, July 11, 2015.

# Broadening Participation in Visualization



# BPViz Workshop

Biennial Workshop

**Goals:** (1) Broaden Participation of women and members of underrepresented groups in data visualization

Underrepresentation refers to gender, ethnicity, academic background, limited access to resources.

(2) Encourage participants to consider STEM fields and professions.

# BPViz Workshop

Biennial Workshop

Diverse in participants and in speakers

**Speakers:** Industry, Academia, Research

**Participants:** undergraduates, graduate students, junior/senior faculty, post-docs, research scientists, lecturers, administrative, K-12 (first year).

## Biennial Workshops

### Day 1:

Hands-on Workshops  
Visualization Panelists  
Meet the Panelists Sessions  
Poster Session/ Vis Showcase

### Day 2:

Polar Vortex!



1<sup>st</sup> CRA-W/CDC/NSF Broadening  
Participation in Visualization (BPViz)  
Workshop, Clemson University,  
Clemson, South Carolina (2014)

## Biennial Workshops

### Day 1:

- Hands-on Workshops
- Visualization Panelists
- Meet the Panelists Sessions
- Poster Session/ Vis Showcase

### Day 2:

- Travel to National Center for Supercomputing Applications (NCSA)
- Hands-on Workshops



BPViz'16 Co-located at Purdue University and University of Illinois Urbana Champaign (NCSA) (2016)

Sponsored by Computing Research Association (CRA) – Women

# BPViz'18

- Broadening Participation in Visualization Workshop
- Two-day workshop that introduces data visualization to all persons with an interest in data visualization
- Sponsored by Computing Research Association (CRA) – Women and NSF
- 2018 Co-located at Purdue University and University of Illinois at Urbana Champaign, NCSA
  - June 13-14, 2018
  - **Now Accepting Applications**
  - Application Process Closes January 19, 2018



The screenshot shows the homepage of the BPViz website. At the top, there's a navigation bar with links for HOME, ABOUT, WORKSHOP DETAILS, and CONTACT. The main title "Broadening Participation in Visualization" is displayed above a large photograph of a group of people at a workshop. Below the photo, there's a section for "Workshop Dates" (June 13-14, 2018), a "Location" section (Co-Located at Purdue University and University of Illinois, Urbana Champaign National Center for Supercomputing Applications), and a detailed "Broadening Participation in Visualization" section.

Broadening Participation in Visualization

Workshop Dates:  
June 13-14, 2018

Location  
Co-Located at Purdue University and University of Illinois, Urbana Champaign National Center for Supercomputing Applications

Broadening Participation in Visualization

Visualization plays an important role in all levels of scholarship, and across interdisciplinary, research and social landscapes. It is imperative that the importance and benefits of a field with such far reaching impact embrace the diverse demographic of the future workforce for whom data visualization skills will be needed if not required. In order to meet the growing demand for persons with capacity to utilize data visualization to solve complex problems in research and in the workforce exposure to data visualization at all levels is essential. This workshop aims to be a catalyst for broadening participation of women, members of underrepresented groups and underrepresented disciplines in visualization, fostering a community of current and future scholars with interests in visualization through mentoring and encouraging participants to consider visualization as a career path. The workshop is designed to inform, inspire and encourage participants, specifically participants historically underrepresented at the conference.

<http://bpviz.org>

# Byrd Data Visualization Lab



**PURDUE**  
POLYTECHNIC



# Byrd Vis Lab: Research Interests

- Data Visualization
- Big Data
- Visualization of Heterogeneous Data
- Uncertainty Visualization
- High Performance Computing for Data Visualization
- Broadening Participation in Visualization

# Byrd Vis Lab

## Students

- Graduate Students
  - PhD (1), MS (2)
- Undergraduates (4)
  - Office of Undergraduate Research (OUR) Scholars (2)
  - Statistics Learning Community Scholars (4)



Looking for more  
Students!

# Byrd Vis Lab: Research Interests

## Visualizing Heterogeneous Data

Data Integration

Uncertainty  
Visualization

Machine  
Learning

High  
Performance  
Computing

Data Bases

Data Source: Lpus Data (Inherently Heterogeneous in nature)

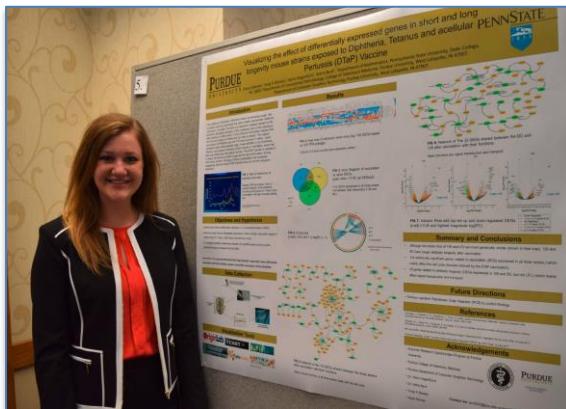


Lpus  
Framework

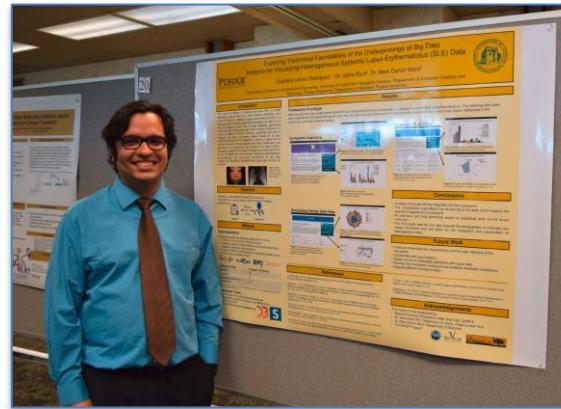
# Byrd Vis Lab: Undergraduate Research

## Purdue Summer Research Opportunity Program

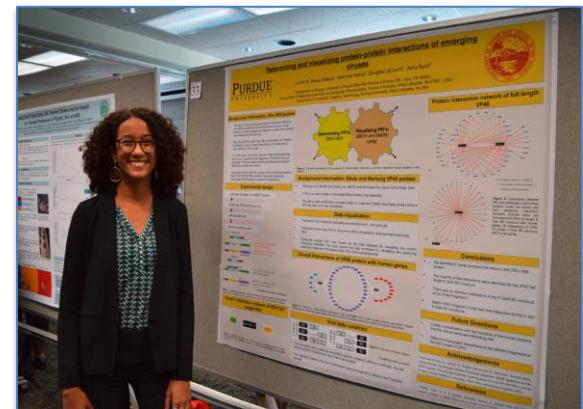
- Collaborative, Interdisciplinary Research
- 8 to 10 – Weeks, Paid Experience



Research Focus: Whooping Cough



Research Focus: Lupus



Research Focus: Zika Virus

# **Byrd Data Visualization Lab**

**I am looking for graduate students with a  
data science background!**



*When you're*  
**CURIOS**  
*you find lots  
of interesting  
things to do.*

*- Walt Disney*

[LegacyTravel.com/travelquotes](http://LegacyTravel.com/travelquotes)

# Questions?

Vetria L. Byrd

Assistant Professor

Computer Graphics Technology

[vlbyrd@purdue.edu](mailto:vlbyrd@purdue.edu)

<https://polytechnic.purdue.edu/profile/vbyrd>

<http://web.ics.purdue.edu/~vbyrd/>



@VByrdPhD, @BPViz

Purdue Polytechnic Institute



**PURDUE**  
POLYTECHNIC

Thank You Image Source:

<http://careerconfidential.com/category/thank-you-notes/>

<http://careerconfidential.com/wp-content/uploads/2015/02/ThankYou2.jpg>

