

# Introduction to Data Visualization

**Vetria L. Byrd, PhD**

Assistant Professor

Computer Graphics Technology

Purdue University

**International HPC Training Consortium  
Visualization Seminar Series**

**March 01, 2017**



# **ABOUT THIS WEBINAR**

## **ASSUMPTIONS**

**Target Audience:** Beginners

**Pre-requisites:** No prior knowledge of visualization required

**Software Requirements:** None

# ABOUT THIS WEBINAR

## GOALS

1. Provide viewers with an introduction to data visualization
2. Provide a summary of visualization capabilities
3. Identify first steps towards visualizing different types of data

# ABOUT THIS WEBINAR

## OBJECTIVES

1. Explore the underlying principles of data visualization,
2. Explore the visualization process
3. Explore some visualization applications
4. Explore different types of visualization tools for different types of data

# **ABOUT THIS WEBINAR**

## **EXPECTED OUTCOMES**

By the end of this webinar, viewers will

1. Understand the purpose of visualization
2. Be able to identify their data visualization needs
3. Identify the visualization tools available to assist with visualizing their data

# AGENDA

## INTRODUCTION TO DATA VISUALIZATION

- Brief Introduction
- Purpose Of Visualization
- High Level Overview
- Visualization Applications
- You've Got Data, Now What?
- Q&A

# Introductions



# About Me

Vetria L. Byrd, PhD

## Academic Preparation

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



Knowledge that will change your world

## 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)



*Agent for “Insight”*

**PURDUE**  
POLYTECHNIC

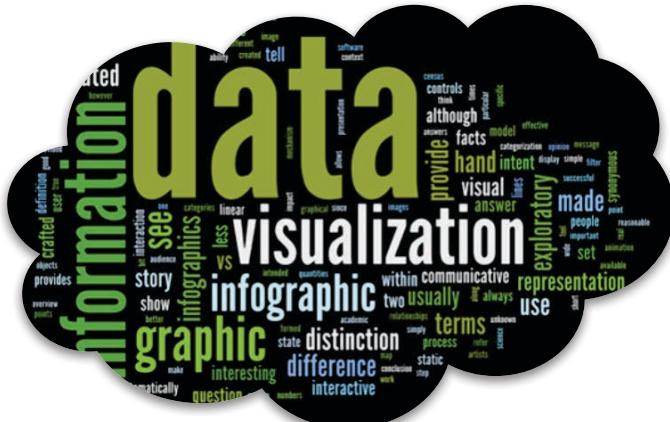
**What do you think of when  
you hear**

**Data Visualization?**



# DATA VISUALIZATIONS

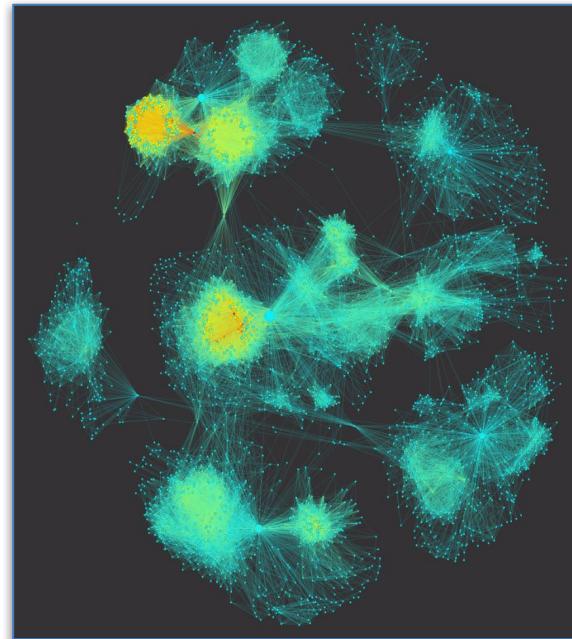
# WE'VE ALL SEEN THEM



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

## Wind Map

# Prime Number Patterns

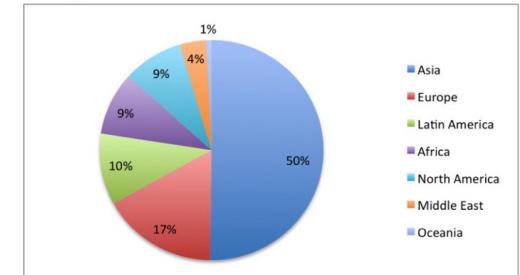


Facebook Network Visualization  
Anonymous friend networks  
Created by Christine Mintert & Fisher Adelakin  
CGT 270 Class Assignment  
Data Source:  
<http://snap.stanford.edu/data/index.html#soc>



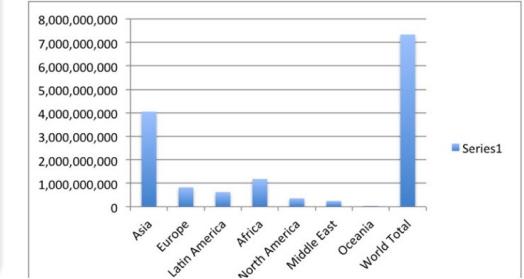
## **Internet Users in the World (Based on Region)**

## Percentage of Internet Users in Regions Around the World in 2016



## Why those percentages?

It's based on the relative population of the region to the rest of the world!



## Internet Users in the World (per 100 people)



# What is 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.

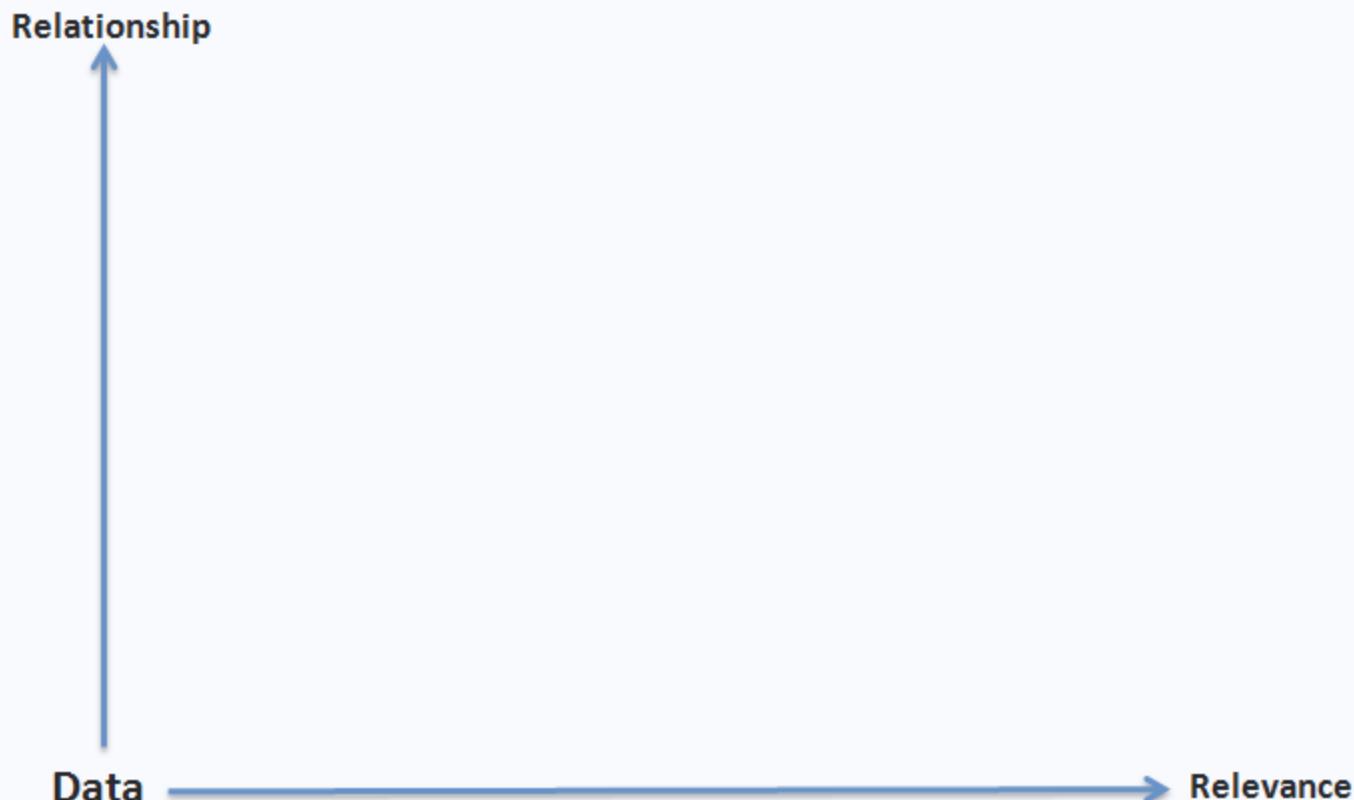
# What is the purpose of visualization?



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

~Ben Shneiderman

## Advancing Beyond Data to True Insight

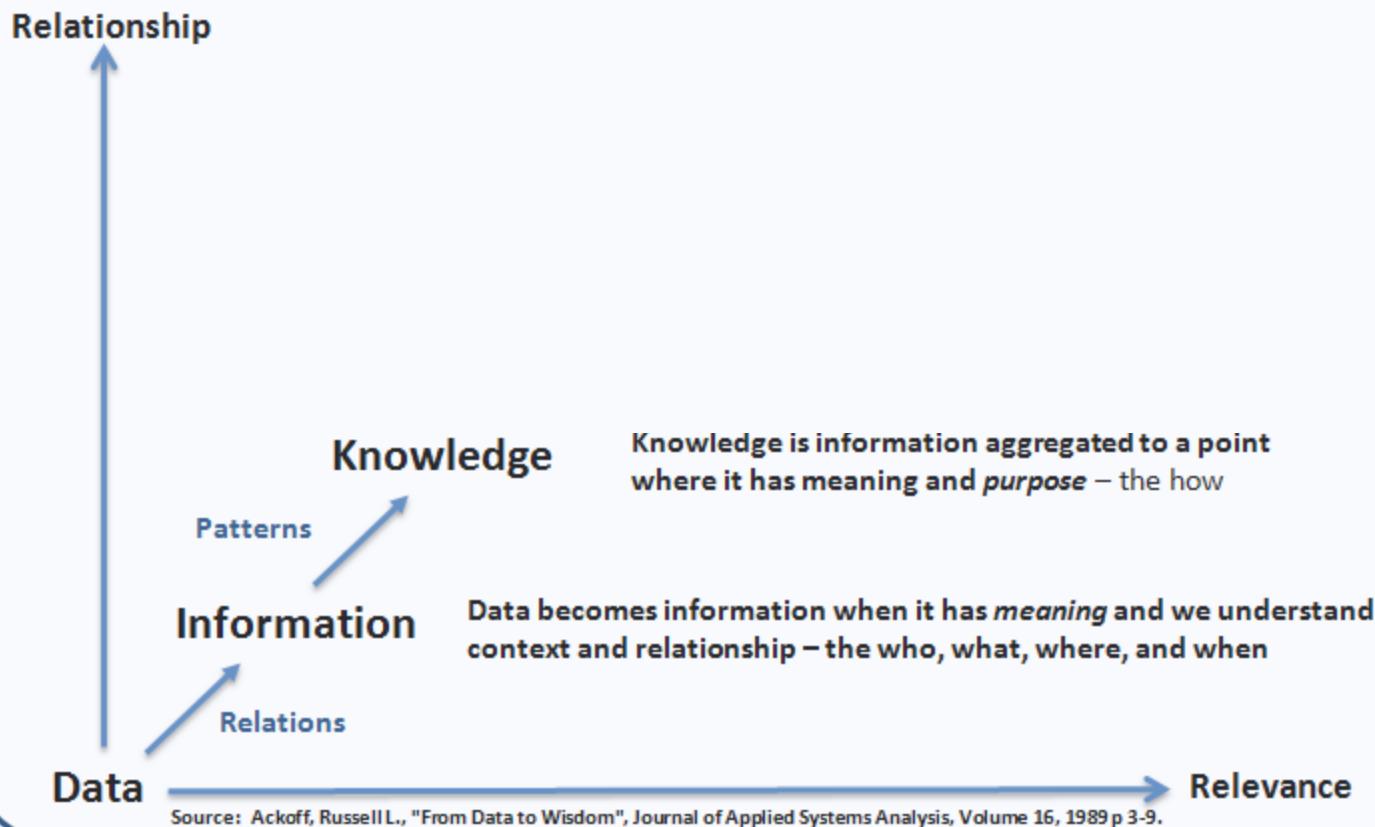


Source: Ackoff, Russell L., "From Data to Wisdom", Journal of Applied Systems Analysis, Volume 16, 1989 p 3-9.

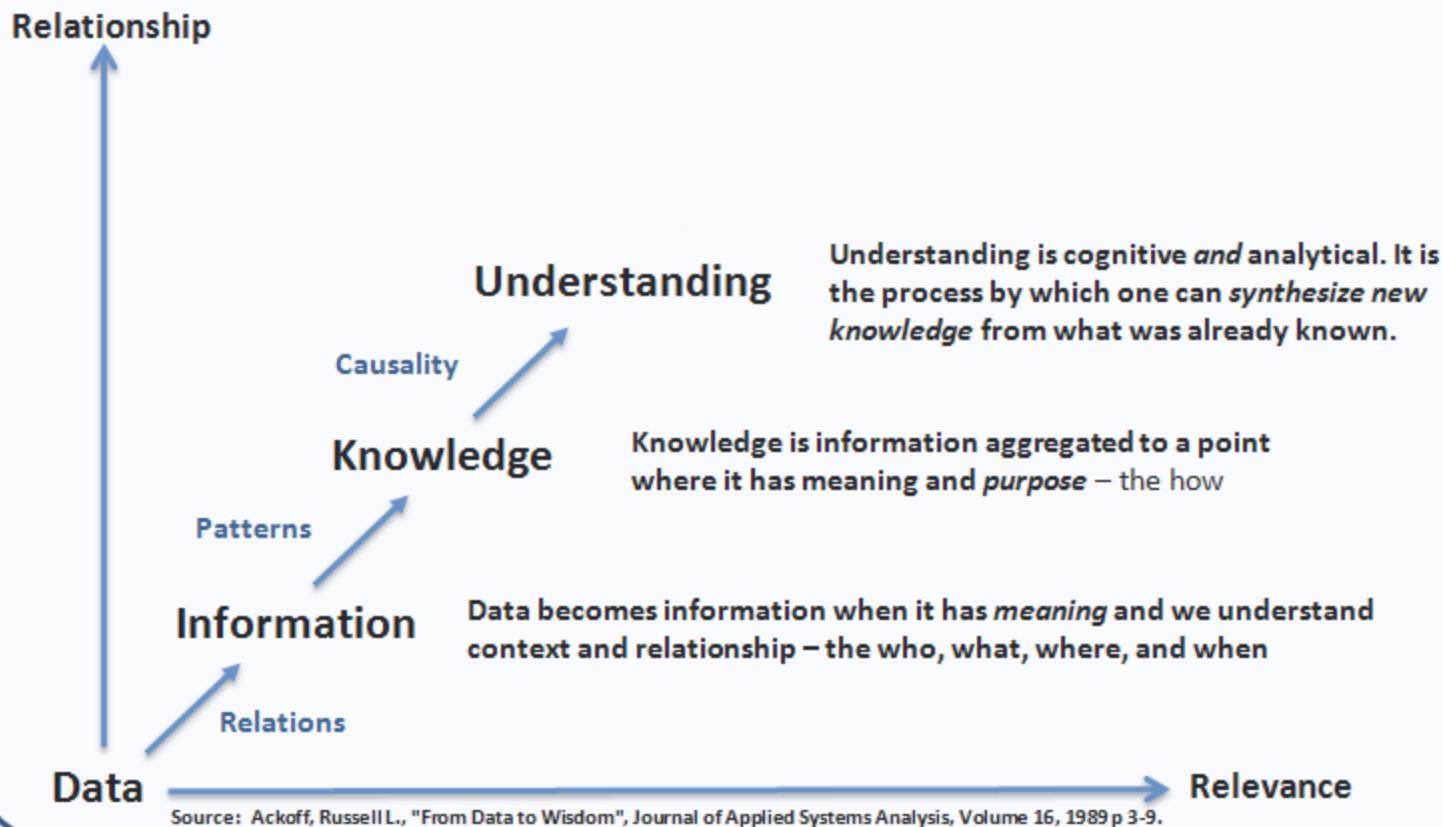
## Advancing Beyond Data to True Insight



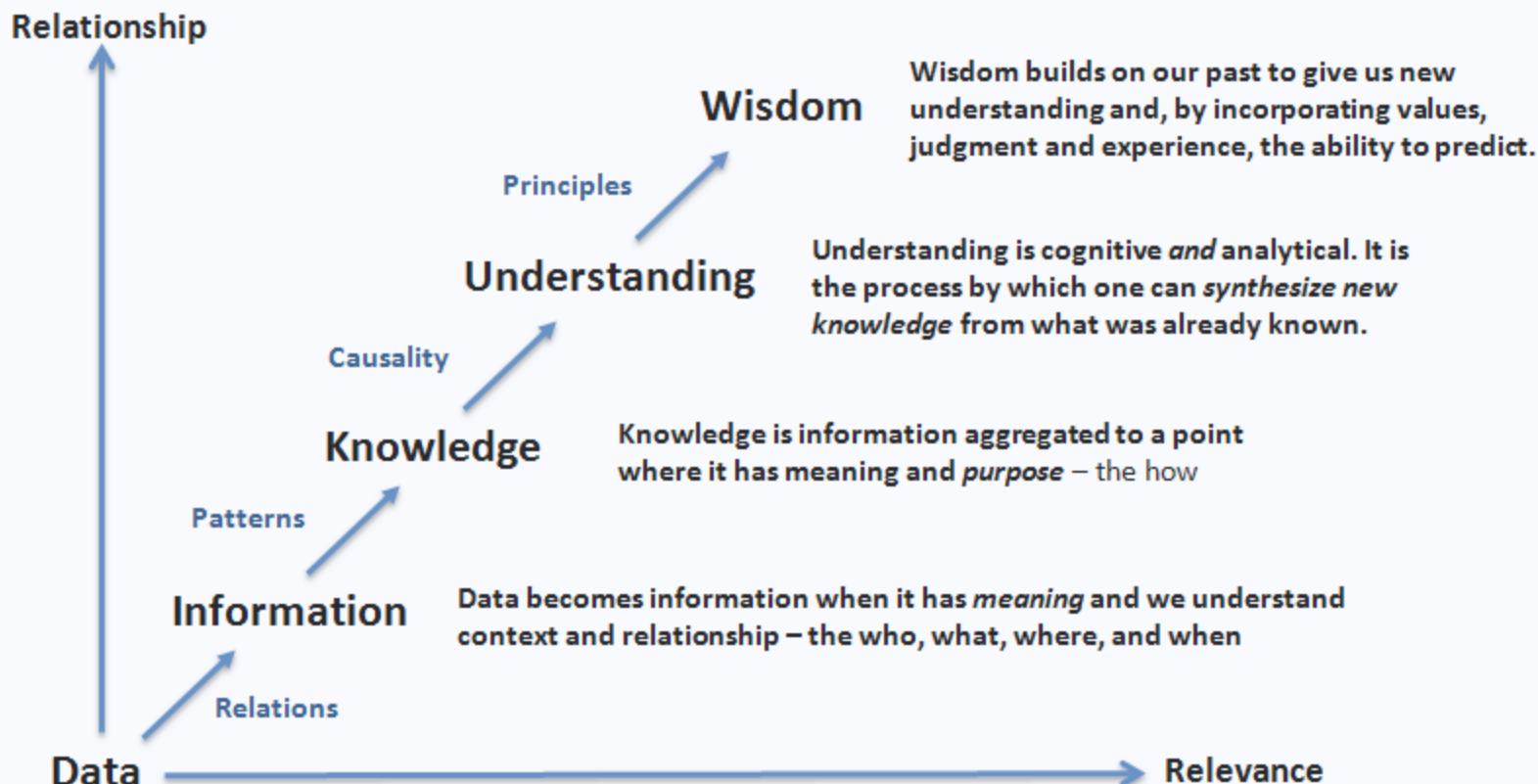
## Advancing Beyond Data to True Insight



## Advancing Beyond Data to True 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.

# **Questions?**

## **Next: What does Insight Lead to**



# What does Insight lead to?



# INSIGHT LEADS TO

## Discovery

- Visualizing Patterns over time
- Spotting Differences

Decision Making

Analysis of Data

Explanation

Storytelling



# INSIGHT LEADS TO

Discovery

- Visualizing Patterns over time
- Spotting Differences

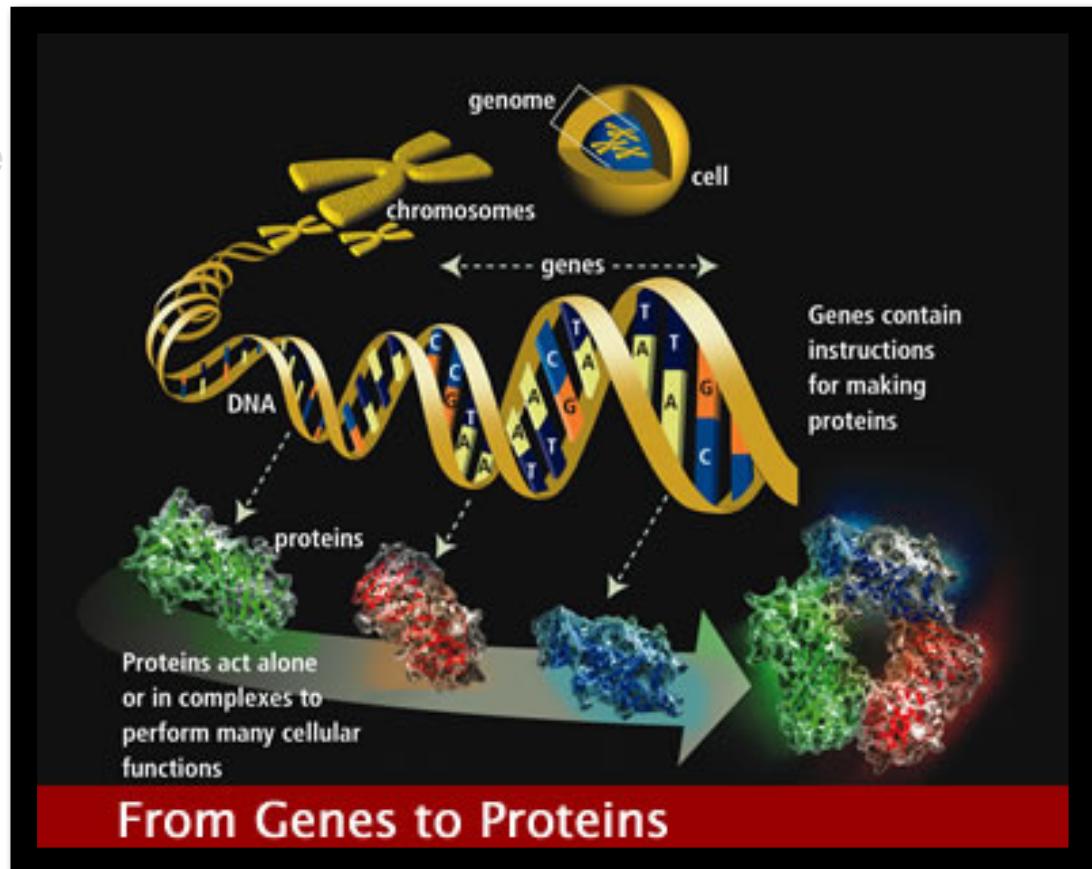
Decision Making

Analysis of Data

Explanation

Storytelling

**Allows users to answer  
questions they didn't know  
they had**



Human Genome Project

<https://pradipjntu.files.wordpress.com/2011/05/molecularmachine.jpg>

# INSIGHT LEADS TO

Discovery

Decision Making  
Analysis of Data  
Explanation  
Storytelling



*Katherine Johnson (played by Taraji P. Henson) calculates orbital insertion trajectories for the Mercury program using Euler's method in this scene from the movie Hidden Figures. Credit: TM and © 2017 Twentieth Century Fox Film Corporation. All rights reserved.*

# INSIGHT LEADS TO

## Discovery

- Visualizing Patterns over time
- Spotting Differences

## Decision Making

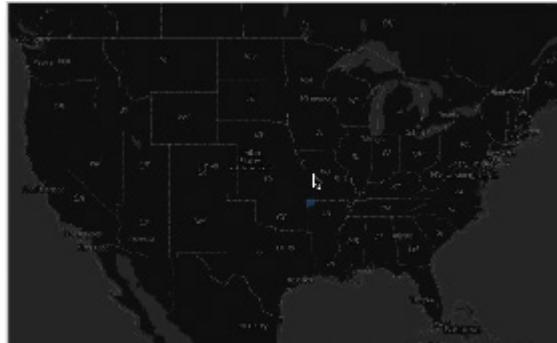
## Analysis of Data

## Explanation

### Visualizing Spatial Relationships

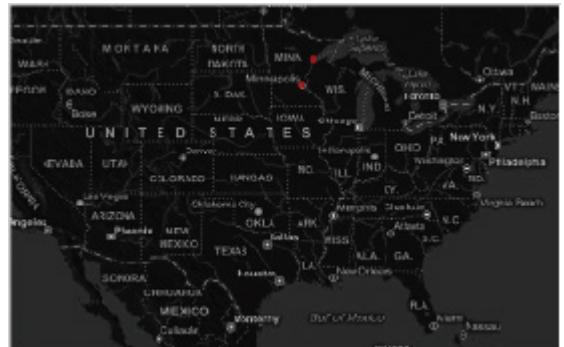
## Storytelling

Muehlenhaus, I. (2012). **Chapter 8, Visualizing Spatial Relationships**, Visualize This: The Flowing Data Guide to Design, Visualization, and Statistics, pp 271-326.



Watch the Growth of Walmart and Sam's Club

<http://datafl.ws/197>



Watch the Growth of Target Stores

<http://datafl.ws/198>

# INSIGHT LEADS TO

Discovery

- Visualizing Patterns over time
- Spotting Differences

Decision Making

Analysis of Data

Explanation

Storytelling

COVER FEATURE



**Storytelling:  
The Next  
Step for  
Visualization**

Robert Kosara and Jock Mackinlay, *Tableau Software, Seattle*

R. Kosara and J. Mackinlay, "Storytelling: The Next Step for Visualization," in *Computer*, vol. 46, no. 5, pp. 44-50, May 2013.

# Story Telling with Visualization

## Napoleon's Invasion of Russia in 1812 By Jacque Minard

Army Size:  
100,000

*Figurative Map of the successive losses in men of the French Army in the Russian campaign 1812-1813.  
Drawn up by J. Minard, Inspector General of Bridges and Roads in retirement.*

The number of men present are represented by the width of the colored zones at a rate of one millimeter for every ten thousand men; they are further written across the zones. The red designates the men who enter into Russia, the black those who leave it. — The information which has served to draw up the map has been extracted from the report of M.R. Thiers, of L'Isle, of Recouvrance, of Chambry and the unpublished diary of Jacob, pharmacist of the Army since October 21st. — In drawing the map with the zone, the diminution of the army, I have assumed that the corps of Lefèvre, those of Marshal Davout, who had been detached at Wiazma and Borodino and had rejoined around Orelia and Vitebsk, had always marched with the army.

Army Size:  
422,000

Width of band indicates the size of the army at each position

Army Size:  
10,000

-26

-30

-20 November 28

-16

-11

-21

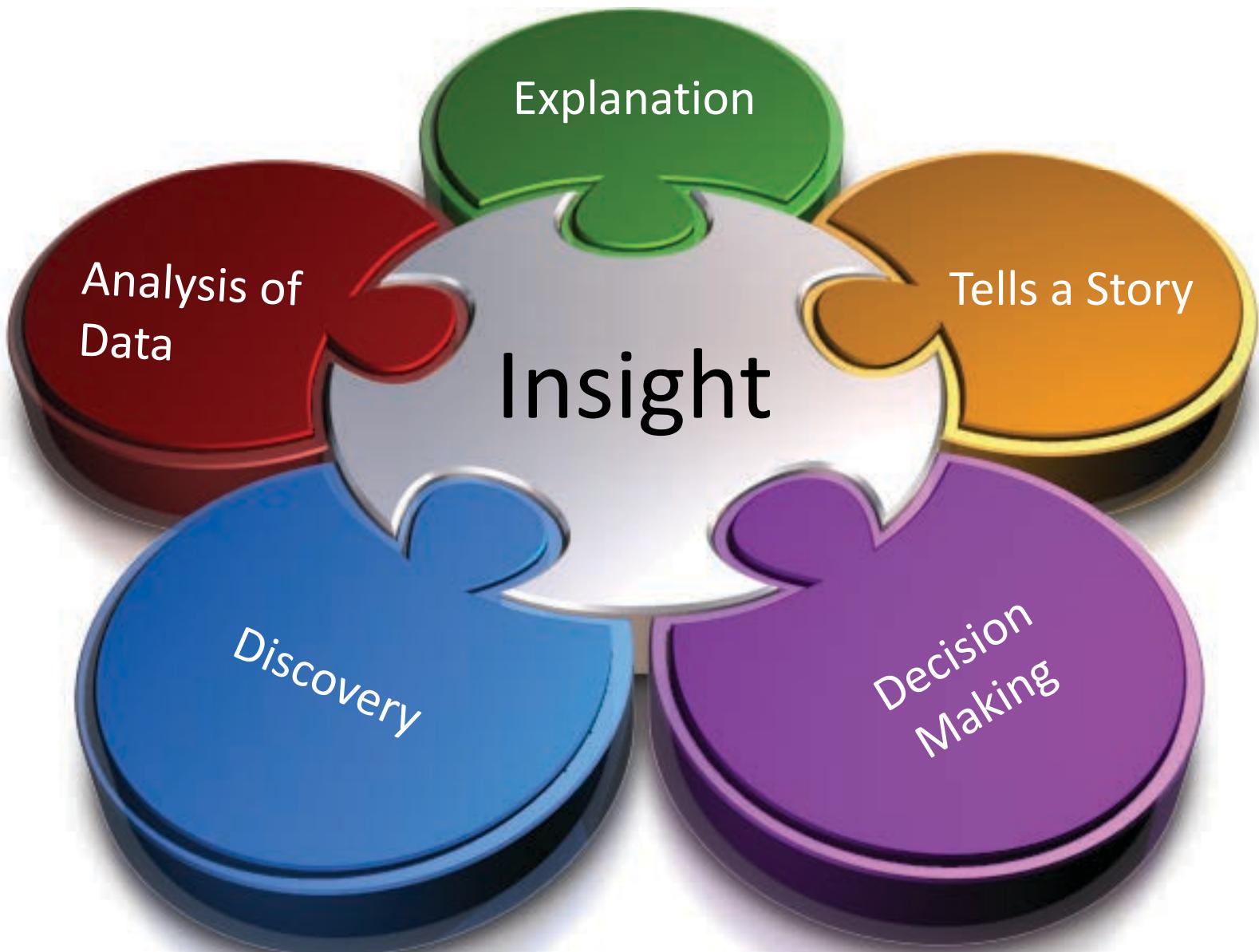
-9

Path of retreat

Graphic Table of Press (Map of Russia)

Temperature

Moscow



# Questions?

## Next: Visualization Applications



# Visualization Applications

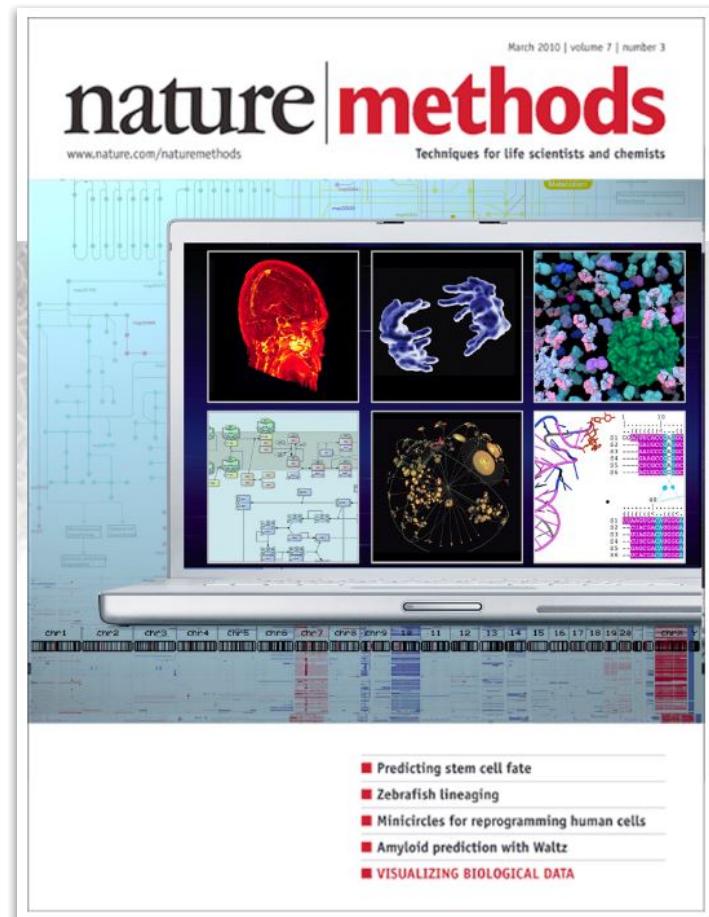
## Why is visualization important?



# Visualization Applications

## Biovisualization (BioVis)

The visualization of biological data;  
Often grouped with computer animation



March 2010 | volume 7 | number 3

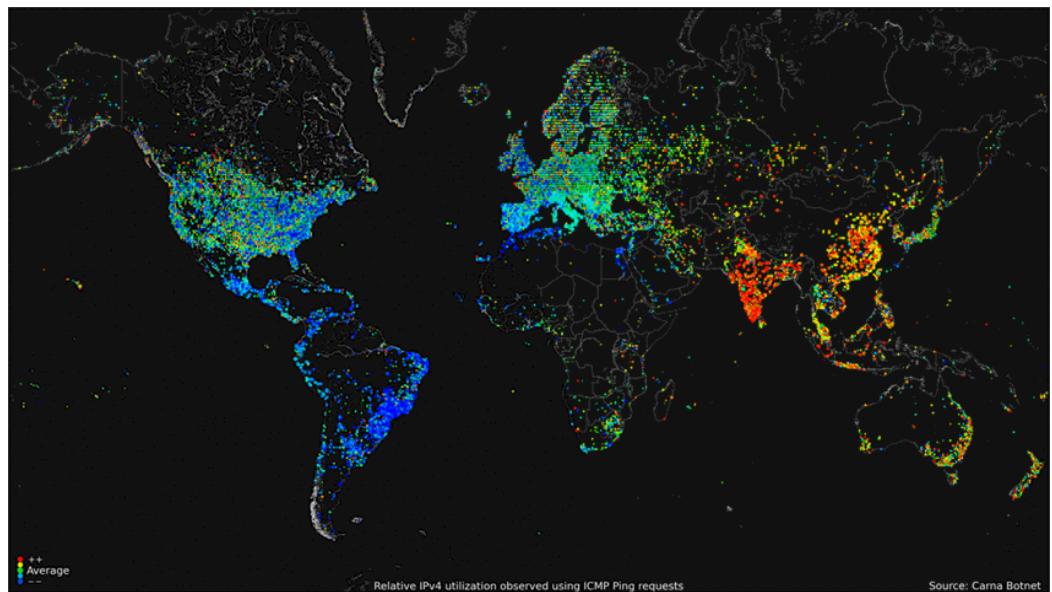
# Visualization Applications

## Information Visualization (InfoVis)

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



Social Media Data  
Survey Data  
Observed Data



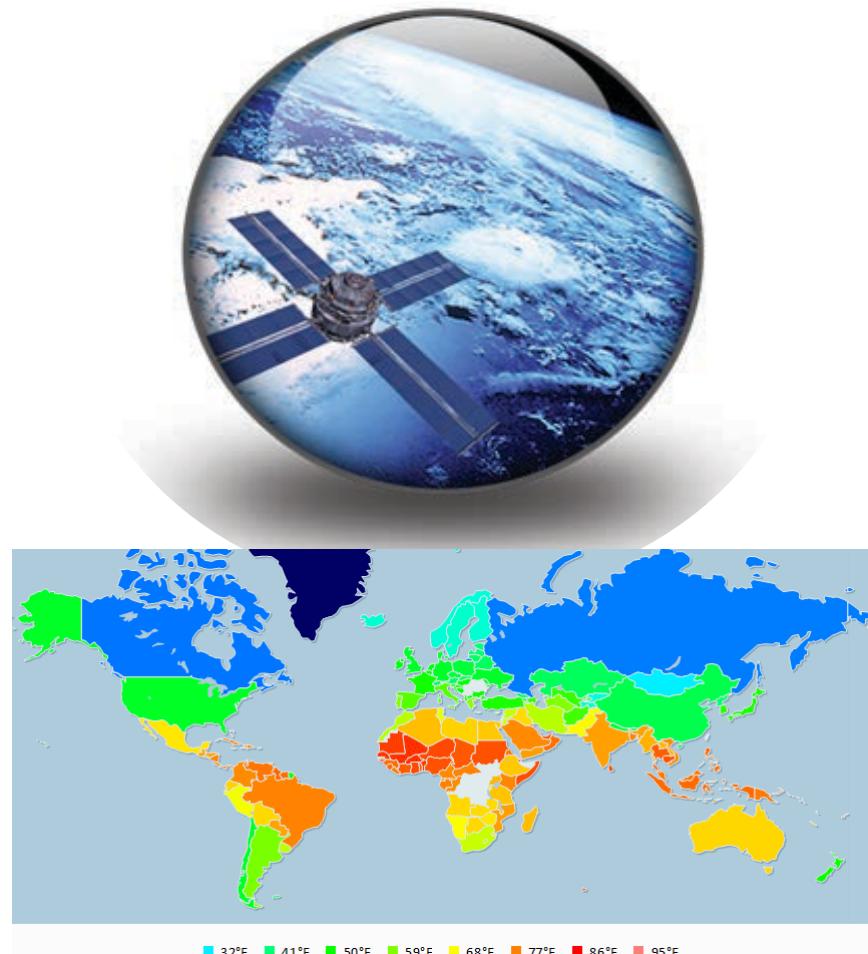
Internet Usage

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>

# Questions?

## Next: High Level Overview



# Data Visualization Process

## High Level Overview



# Data Visualization Process

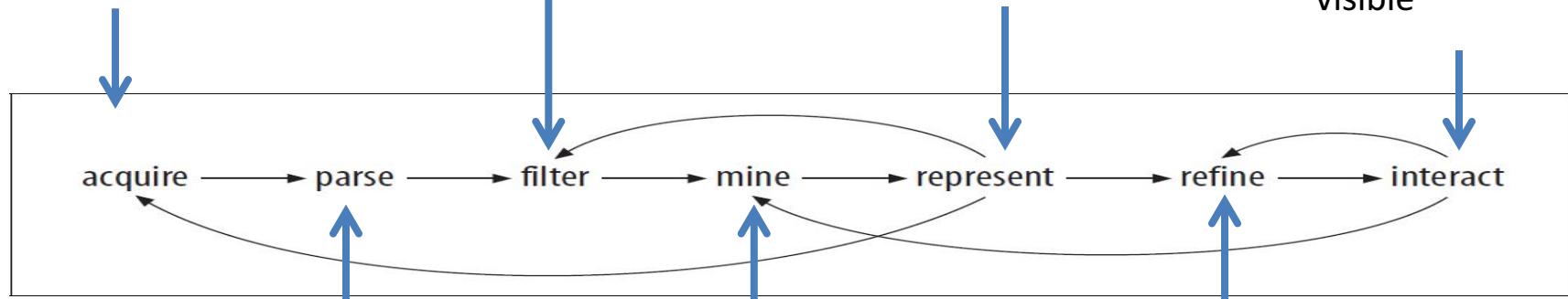
*An iterative process*

obtain the data

remove all but  
the data of  
interest

choose a basic  
visual model,  
such as a bar  
graph, list or tree

Add methods for  
manipulating the  
data or controlling  
what features are  
visible



Adopted from Visualizing Data: Exploring and Explaining Data with the Processing Environment  
by Ben Fry, O'Reilly (p 15)

# Visualization Process

Taking raw data and converting it to a form that is viewable and understandable to humans.

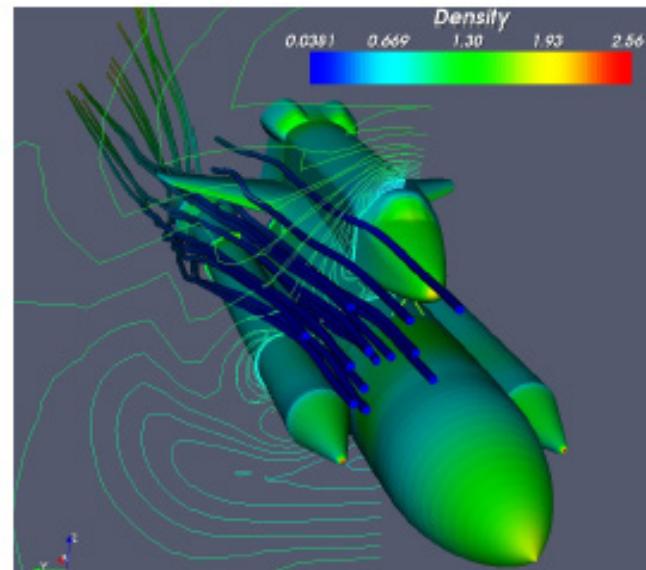
```
0265640 132304 139732 032051 037334 024721 035013 052226 001662
0265650 025587 064663 054605 043244 074076 124153 135216 126514
0265700 144210 056426 044700 042650 165230 137037 009655 006254
0265720 134453 124327 076005 027034 102634 170774 023702 067274
0265740 072451 007335 347520 063064 157435 133057 155358 114503
0265760 107204 102316 371451 046040 120223 001774 030477 046679
0266000 171317 115055 355117 334444 167220 041405 147227 050505
0266020 004137 046472 324015 134360 173550 053517 044635 021135
0266040 070126 047705 313754 175477 105532 076535 172366 056239
0266050 041023 074017 327113 003214 037025 037640 065171 123424
0266100 062701 032400 340000 165341 072410 100032 125455 056546
0266110 006715 071402 055672 332571 105645 170073 050376 072117
0266140 024451 007424 314200 077733 024434 012546 172404 102345
0266160 010223 050270 055164 1364684 047154 125525 112514 032315
0266200 0316041 175055 042765 025015 176334 017234 110050 014515
0266210 127156 030746 354234 125001 151144 163706 136237 164276
0266240 137055 052276 361755 315466 005322 132587 078216 002655
0266260 171456 126315 317155 055763 016177 014450 112755 055527
0266300 003767 175357 304754 036435 172172 150750 043643 145410
0266320 072074 000007 040627 076652 178011 002151 125132 140214
0266340 060115 014356 015164 067027 120205 070242 039065 101234
0266380 170601 170206 040437 127277 124446 186681 041462 116821
0266400 020243 005602 004145 321574 124651 005634 071331 102070
0266410 157504 160307 366330 070251 024520 134433 167273 030635
0266440 135614 105371 344160 010652 007385 026416 150716 100413
0266460 026630 002110 000630 123224 076033 140764 000737 003276
0266500 114050 042647 304475 310537 066715 104754 075547 112254
0266520 030374 144253 077734 015157 002513 173526 035531 150003
0266540 146235 015335 024486 380101 072857 040764 165518 156412
0266560 165410 067255 356160 106405 136770 030516 064740 022092
0266600 142166 129707 375121 073170 076357 037293 031396 015232
0266620 075074 016744 044055 102290 110063 033350 052765 172463
```



# Visualization Process

There are several steps between raw data and a finished visualization

```
0265640 132304 139732 032051 037334 024721 015013 052226 001662  
0265650 025587 084653 054605 043244 074078 124153 135216 126514  
0265700 144210 055426 044700 042550 165230 137037 005655 006554  
0265730 134453 124327 376005 027034 107634 170724 023702 067274  
0265740 072451 007735 347820 053064 157435 138057 155356 114503  
0265760 107204 102816 371451 046040 120223 001774 030477 046679  
0266000 171317 118055 355117 138444 167210 041405 147227 050505  
0266020 004137 046472 324015 134360 175550 053517 044635 021135  
0266040 070172 047705 313754 175477 105532 076515 177356 056393  
0266050 041023 074017 327113 003214 037025 037640 055171 123424  
0266100 062701 082406 340000 165341 072410 100032 125455 056646  
0266110 066716 071402 055672 132571 105645 170073 050376 072117  
0266140 024453 007424 014200 077238 024434 012546 172404 102345  
0266123 040223 050270 055164 136454 047154 125525 112514 082315  
0266200 0316041 176055 042765 025015 176314 017234 110050 014515  
0266220 112156 090746 354234 125001 151144 163706 196237 164276  
0266240 137055 082276 361755 118466 005312 132567 078216 002655  
0266260 171456 126351 317155 065763 016177 014460 112765 055527  
0266300 003767 175367 104754 036435 172172 150750 043643 145410  
0266320 072074 000007 040627 076652 178031 062151 125132 140214  
0266340 060115 014356 015164 067027 120205 070242 039055 131234  
0266380 170601 170206 040437 127277 124446 185681 044462 116321  
0266400 020243 005602 004145 121574 124651 005634 071331 102070  
0266420 157504 160307 266380 074251 024510 134433 167278 030535  
0266440 133614 106371 344160 010652 007385 026416 160716 100415  
0266460 025630 007210 000630 121224 076033 140764 000737 003276  
0266500 134050 042647 304475 310537 066716 104754 075447 112254  
0266520 030374 144251 077734 015157 025253 173526 035531 150003  
0266540 146207 015135 02446 138101 072857 040754 165518 156412  
0266550 165410 067251 356160 106406 136770 030516 064740 022032  
0266600 142166 123707 375121 073170 076357 037233 031136 015232  
0266620 075074 018744 044055 102290 110063 033350 052755 172463
```



# **Why do we care? I just want a pretty picture!**



# Visualization

Visualization is the tool that will take us forward from the traditional output of high performance computing (HPC) that we are used to into a visual medium that allows researchers to *collaborate* and *elaborate* on the finding's they've got.

*Tim Carroll*

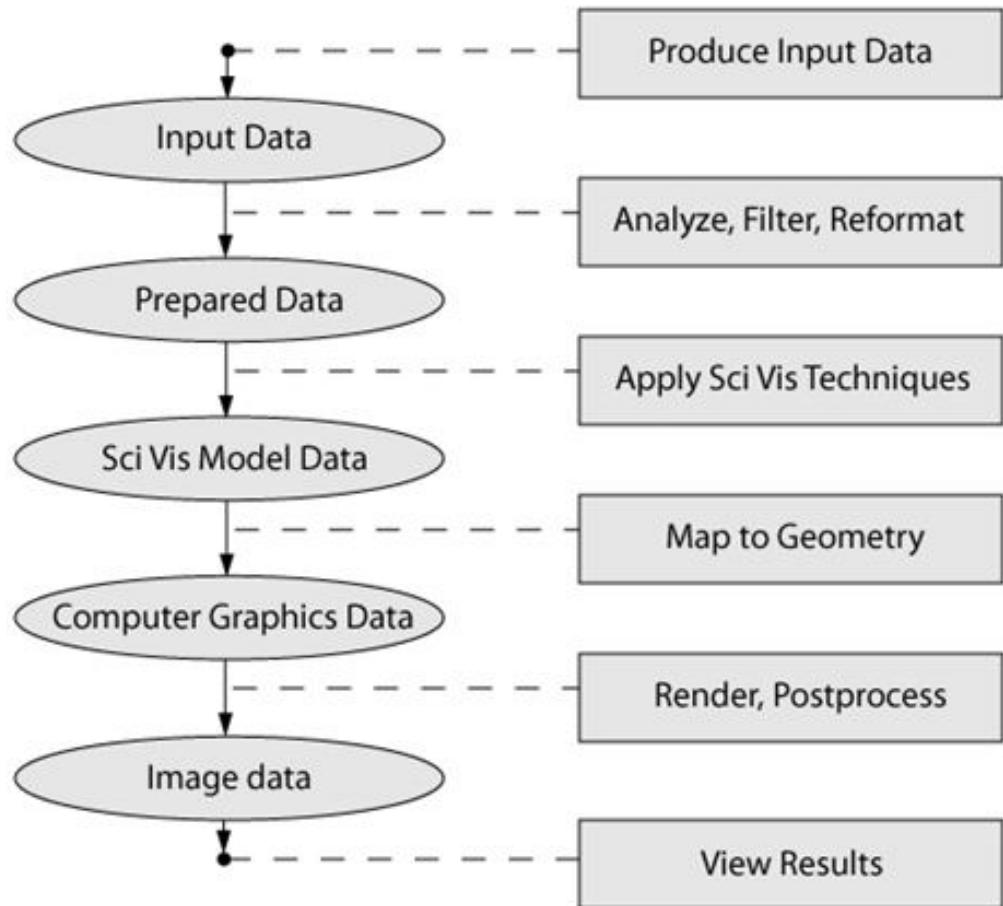
*Director and Global Lead,  
Dell Research Computing Solutions  
HPC Source (Spring 2011)*

# Scientific Visualization

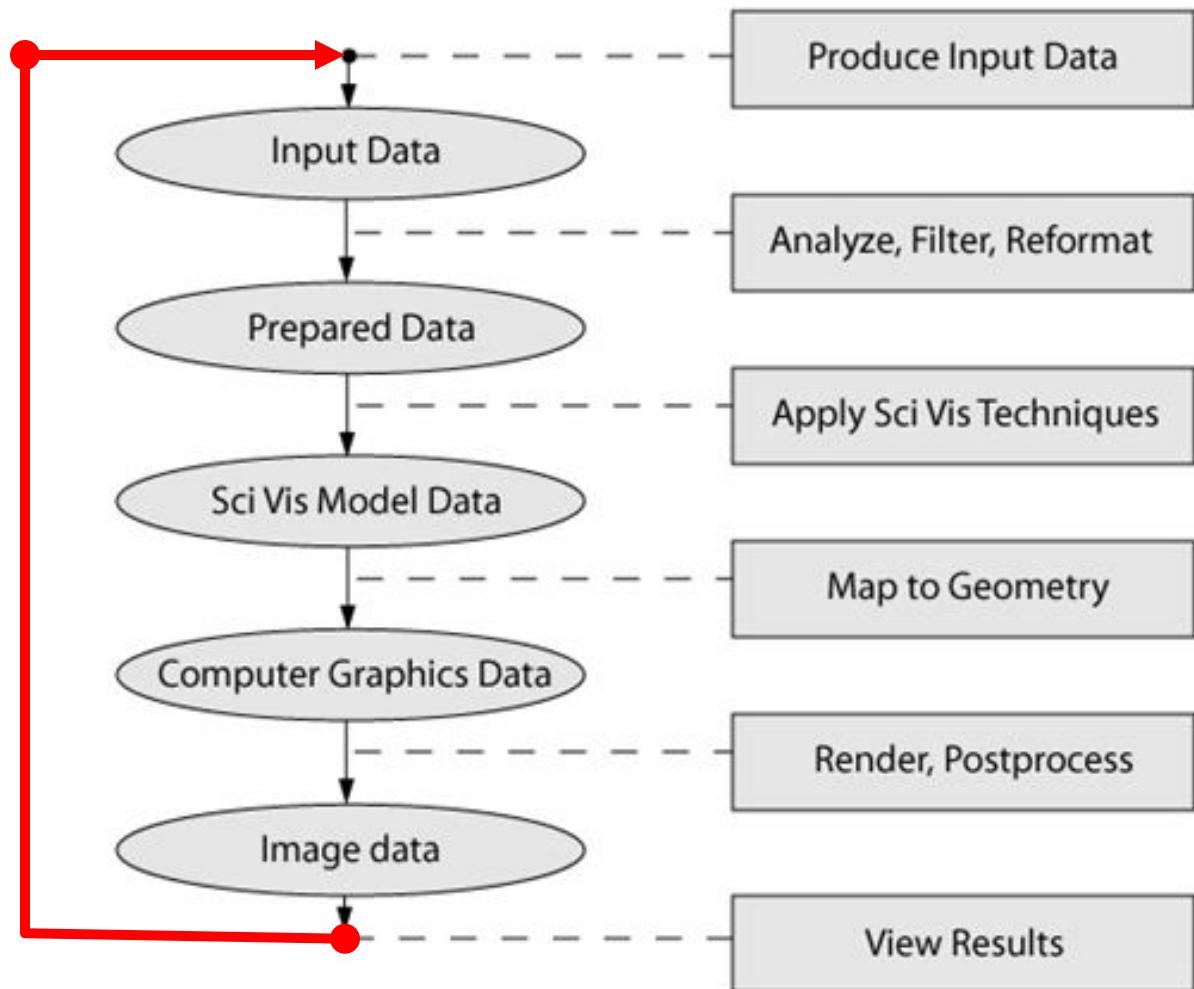
- Primarily concerned with the visualization of three-dimensional phenomena (architectural, meteorological, medical, biological, etc.),
- Where the emphasis is on realistic renderings of volumes, surfaces, illumination sources, and so forth, perhaps with a dynamic (time) component.

# Scientific Visualization Pipeline

What's Missing?



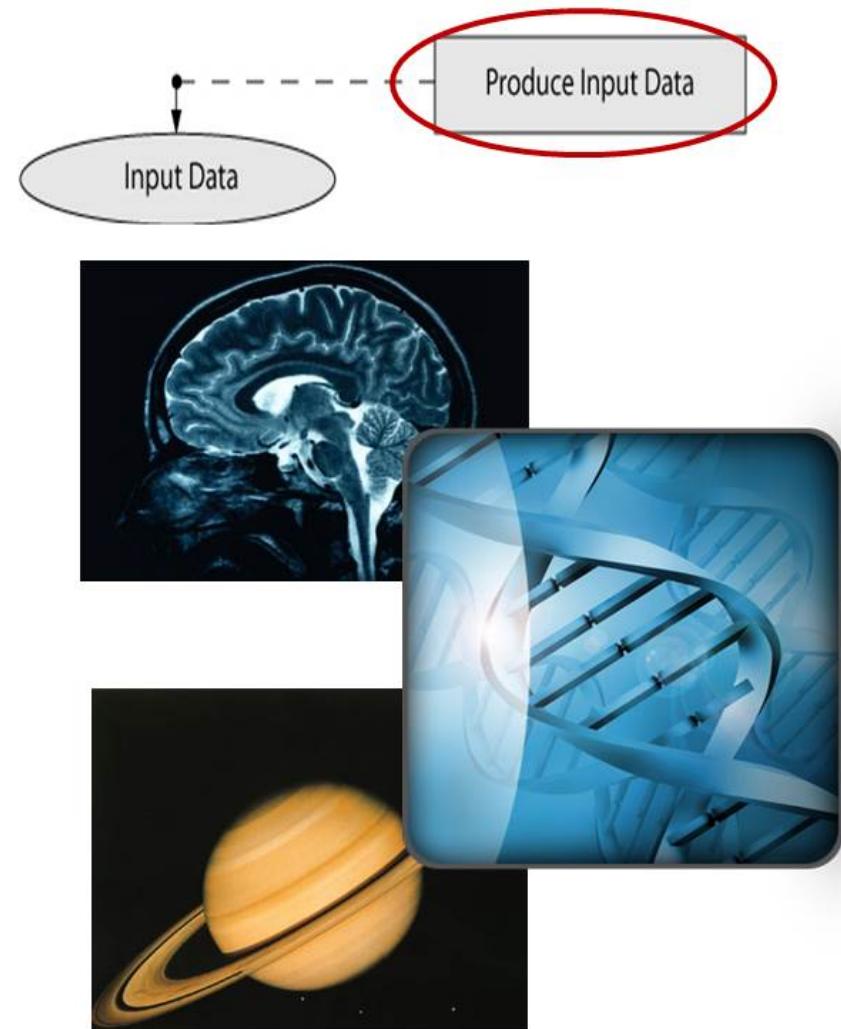
# Scientific Visualization Pipeline



# *Scientific Visualization Pipeline: Step 1 . . .*

## **Produce Data**

Simulated Data  
Images  
Numerical  
Some measured value  
Observed Phenomena



Adopted from

<http://www.bu.edu/tech/research/training/tutorials/introduction-to-scientific-visualization-tutorial/the-scientific-visualization-pipeline/>

# *Scientific Visualization Pipeline:*

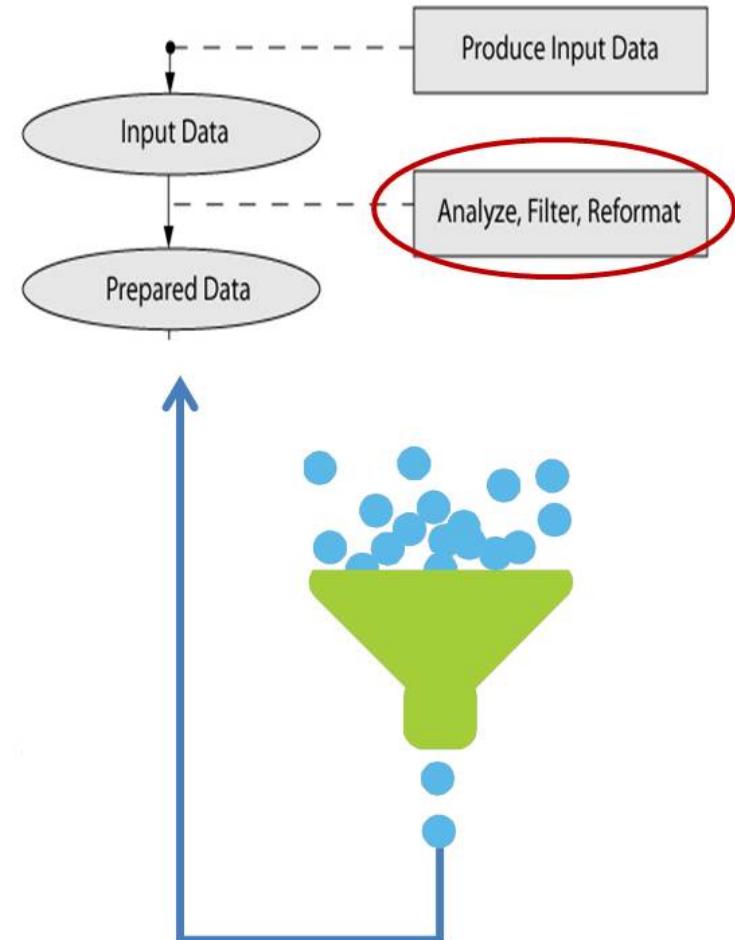
*Step 2 . . .*

## **Analyze, Filter, Reformat**

Cleaning up the data

- Removing noise
- Replacing missing values
- Clamping values to be within a specific range of interest

Performing operations to yield more useful data



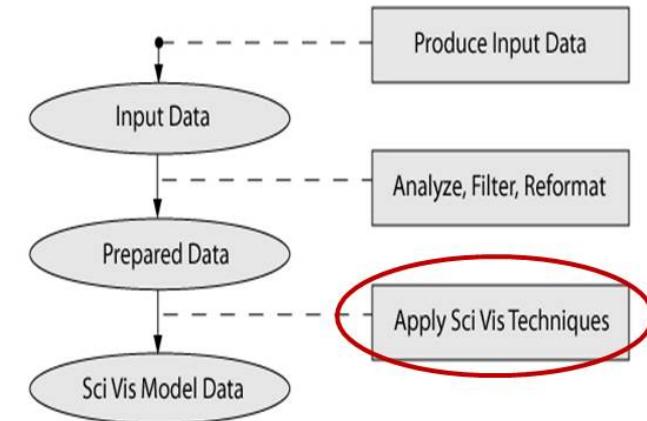
Adopted from

<http://www.bu.edu/tech/research/training/tutorials/introduction-to-scientific-visualization-tutorial/the-scientific-visualization-pipeline/>

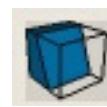
# *Scientific Visualization Pipeline: Step 3*

## **Apply SciVis Techniques**

- Converts raw information into something more understandable
- Visually extracting meaning from a scientific data set using various techniques



Contour



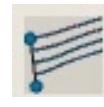
Clip



Threshold



Glyphs



Streamlines

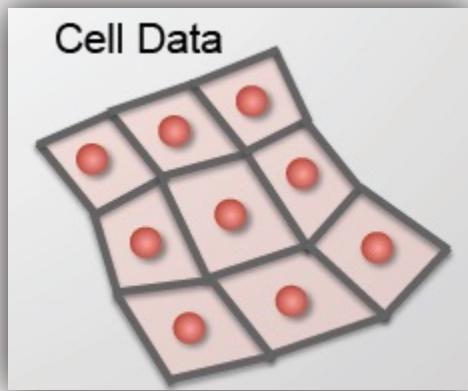
Adopted from

<http://www.bu.edu/tech/research/training/tutorials/introduction-to-scientific-visualization-tutorial/the-scientific-visualization-pipeline/>

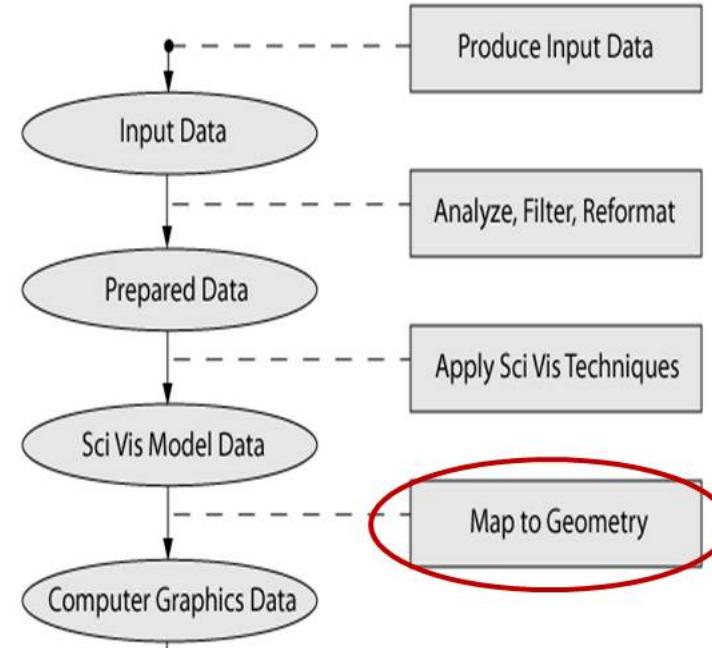
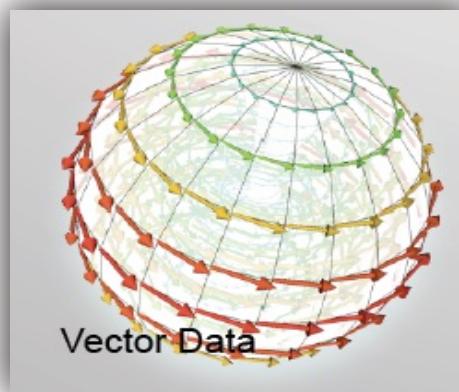
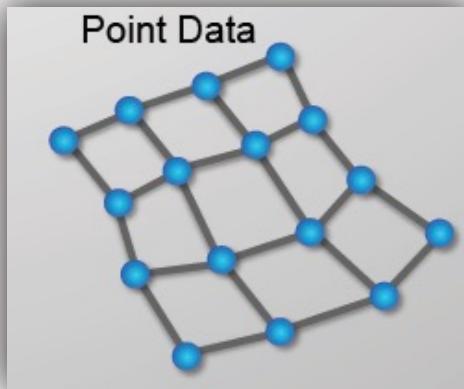
# *Scientific Visualization Pipeline*

Step 4 . . .

## Map to Geometry



Scalars, vectors,  
tensors  
1D, 2D, 3D  
Mesh



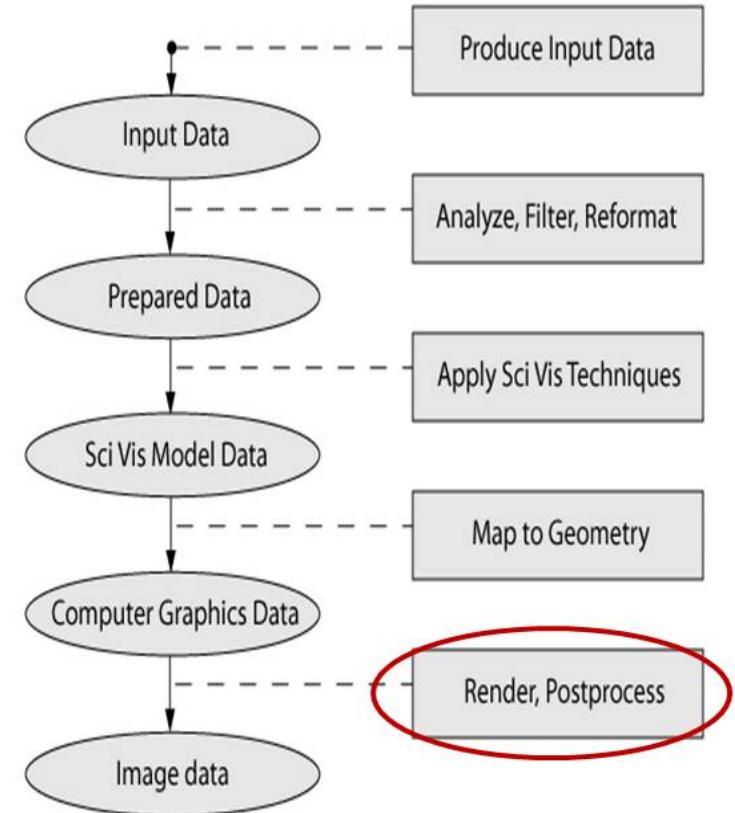
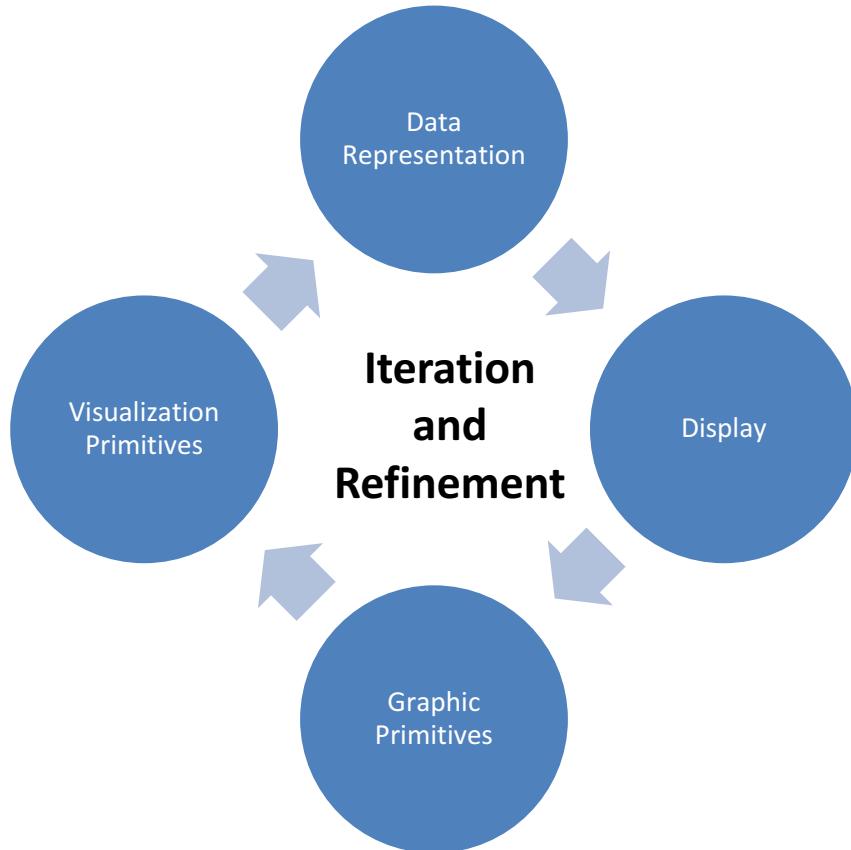
Adopted from

<http://www.bu.edu/tech/research/training/tutorials/introduction-to-scientific-visualization-tutorial/the-scientific-visualization-pipeline/>

# *Scientific Visualization Pipeline:*

*Step 5 . . .*

## **Render, Post Process**



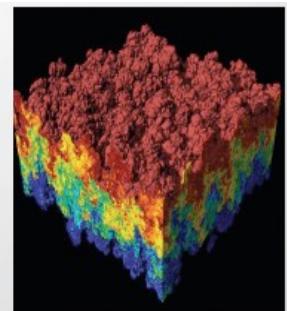
Adopted from

<http://www.bu.edu/tech/research/training/tutorials/introduction-to-scientific-visualization-tutorial/the-scientific-visualization-pipeline/>

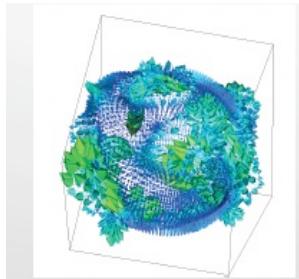
# *Scientific Visualization Pipeline:*

*Step 6 . . .*

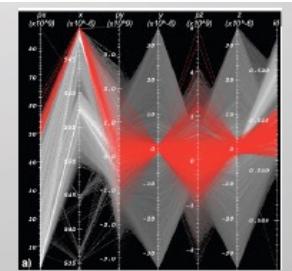
## View Results



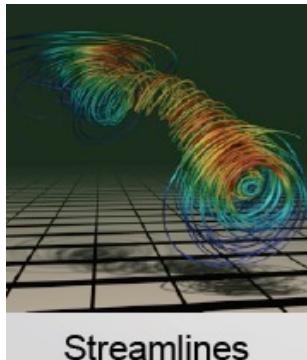
Pseudocolor Rendering



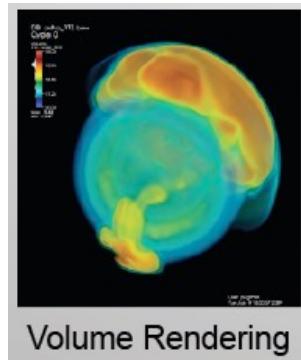
Vector / Tensor Glyphs



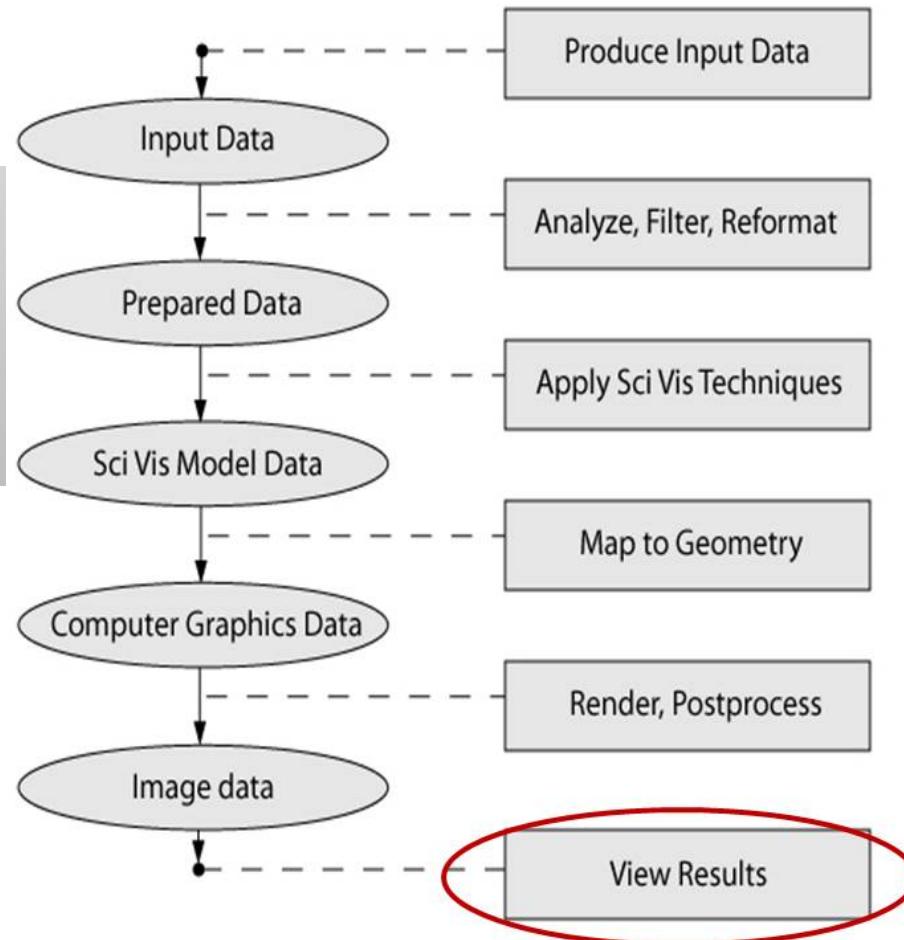
Parallel Coordinates



Streamlines



Volume Rendering



Adopted from

<http://www.bu.edu/tech/research/training/tutorials/introduction-to-scientific-visualization-tutorial/the-scientific-visualization-pipeline/>

# Questions?

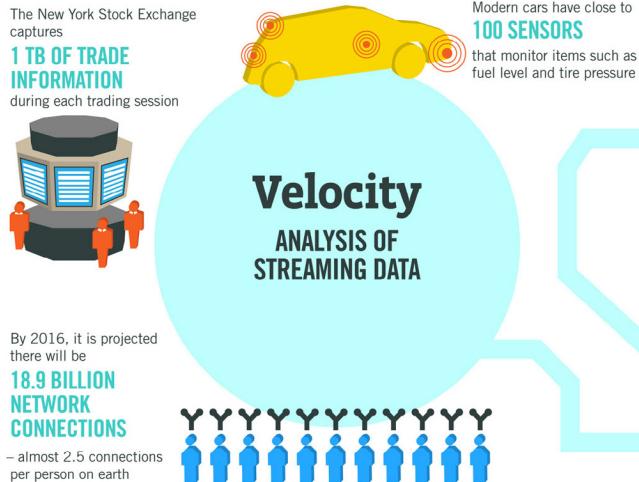
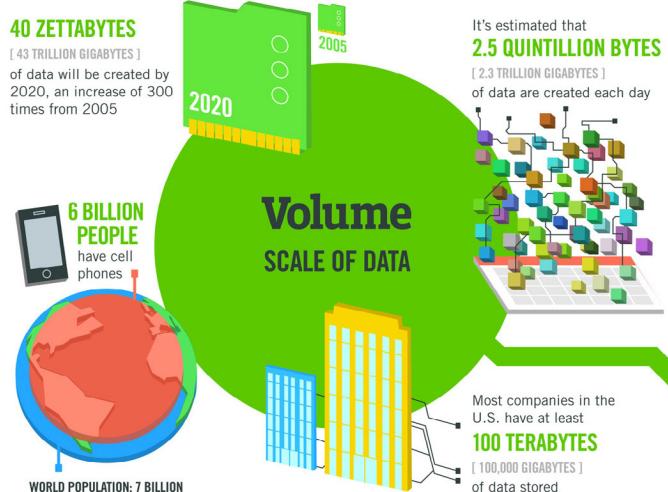
## Next: The Importance of Data Visualization



# Why is Data Visualization Important?



# 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]



### Variety DIFFERENT FORMS OF DATA

**30 BILLION PIECES OF CONTENT**

are shared on Facebook every month



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

Poor data quality costs the US economy around

**\$3.1 TRILLION A YEAR**



### 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

# Why should you care about Visualization?



# Why Should You Care About Visualization

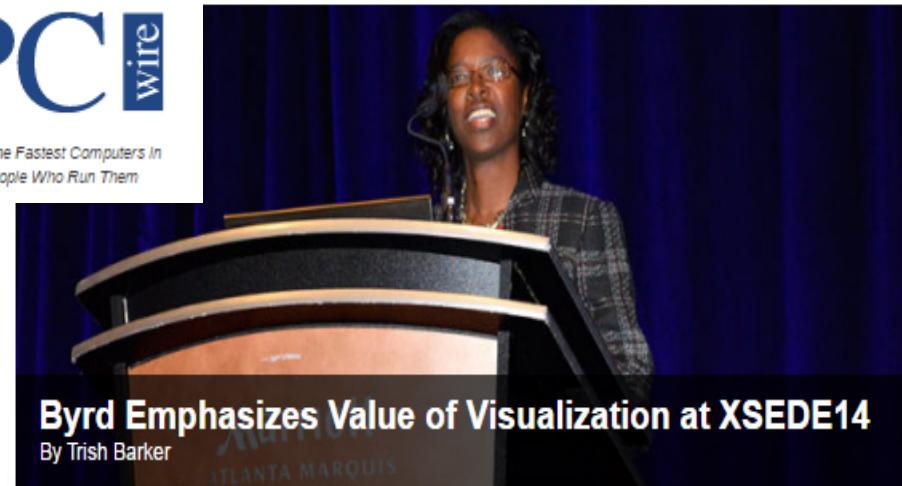
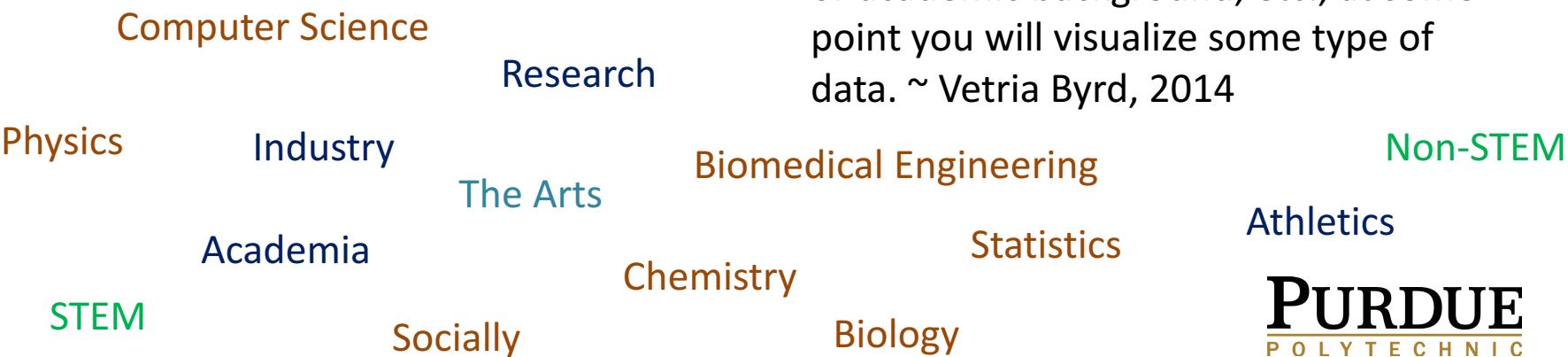
- There is a demand for people who understand the visualization process and is able to transform raw complex data into a visual representation that does not overwhelm.
- Regardless of major, research interest, or academic background, at some point you will visualize some type of data. ~ Vetria Byrd, 2017

# WHY SHOULD YOU CARE?



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

- There is a demand for people who understand the visualization process and is able to transform raw complex data into a visual representation that does not overwhelm.



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

July 31, 2014

# You've Got Data

## Now What?



# YOU'VE GOT DATA

NOW WHAT?

## Questions

1. What does the data look like?
2. What needs to be communicated?
3. What are you interested in utilizing the resulting visualization(s) for
  - ✓ Analysis of data
  - ✓ Explanation
  - ✓ Communication  
(Storytelling)
  - ✓ Discovery
  - ✓ Decision Making
4. What has been done before?
5. [Where do I start?](#)



Image Source: Types of data. Translation of document hosted by João Batista Neto at

[https://commons.wikimedia.org/wiki/File:Data\\_types\\_-\\_pt\\_br.svg](https://commons.wikimedia.org/wiki/File:Data_types_-_pt_br.svg)

More types of data: Biological, Social Media, Network Data, Survey Data, Cybersecurity, Temporal, Image data, Topical, . . . this is NOT an exhaustive list

# Visualization is a Process



# RESOURCES

## A STARTING POINT: OPEN SOURCE VISUALIZATION TOOLS

- Information Visualization
  - Gephi
  - Tableau (not open source but free)
- Scientific Visualization
  - ParaView
  - VisIt
- Geo Visualization
  - ARC GIS
  - D3.js
- Cyber Security Visualization
  - Survey of Security Visualization
  - Survey of Cybersecurity Visualization

# WANT MORE?

[https://keshif.me/demo/VisTools?utm\\_content=26335725&utm\\_medium=social&utm\\_source=twitter](https://keshif.me/demo/VisTools?utm_content=26335725&utm_medium=social&utm_source=twitter)

Created with **Keshif**

## 430 Data-Visualization Tools

**Highlights**

Rating	Count
★★★	8
★★	37
★	90

**Features**

Feature	Count
Charting	256
Web-based	242
Design focused	61
Spreadsheet	55
Infographics	33
Statistical	32
Color	31
Scientific	24

**Data Types**

Type	Count
Multivariate	150
Time	136
Geographical	118
Network	52
Text	21

**Data Processing**

Process	Count
Search	5
Aggregate	14
Filter	12
Advanced	12
Formula	8
Convert	8
Programming	7
Clean	6
Yes	5
Extract	5
Scrape	4
Transform	3

**Cost**

Cost Type	Count
Free	241
Paid	71
Free & Paid	33

**Data-Visualization Tools**

- Keshif**           
- D3.js**     
- Tableau Public**   
- ColorBrewer 2.0**    
- NodeXL**   
- Leaflet.js**   
- RStudio**  
- Crossfilter** 
- dc.js** 

**Main**

Category	Count
Categorized	342
Active	291
Open-Source	184

**Tags**

Tag	Count
Search	5
Platform	23
d3	23
Excel	10
Business Intelligence	10
Storytelling	8
PDF	5
VIS2015	4
Vega	4
Technology	4
jQuery	4
SVG	3
Mockups	3
Journalism	3
iOS	3
Faceted Browsing	3

**Programming**

Language	Count
Graphical Interface	120
JavaScript	89
R	20
Python	20
Scripting	12
HTML	10
Java	8
RESTful API	6
Ruby	5
SQl	4
JSON	4
PHP	3
Scala	2
Processing	2
Other	2

**Author**

Author	Count
39 Rows + 23 More	100

Data is public at [Google Sheets](#). Contribute to this open resource! For info, contact [@adilyalcin](#). All rights for the images belong to their respective owners. Acknowledgements

# 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



Thank You Image Source:

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

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

