Introduction to Data Visualization

Vetria L. Byrd, PhD
Assistant Professor
Computer Graphics Technology

Summer Research Opportunity Program (SROP)
Purdue University

June 7, 2017
ABOUT THIS TALK

ASSUMPTIONS

Target Audience: Beginners

Pre-requisites: No prior knowledge of visualization

Software Requirements: None
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
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 TALK

EXPECTED OUTCOMES

By the end of this presentation, 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
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)

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

- Research Experience for Undergraduates in Collaborative Data Visualization Applications (2014/2015)
What do you think of when you hear Data Visualization?
DATA VISUALIZATIONS
WE’VE ALL SEEN THEM

Wind Map
http://hint.fm/wind/

Prime Number Patterns
https://www.jasondavies.com/primos/

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#socnets

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)

Created by Mrdhula Venkataramani, CGT 270 class assignment
What is Data Visualization?
What is 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

FROM DATA TO WISDOM


**Advancing Beyond Data to True Insight**

- **Data**
  - **Information**
    - **Knowledge**
      - **Understanding**
        - **Wisdom**
          - Wisdom builds on our past to give us new understanding and, by incorporating values, judgment and experience, the ability to predict.
          - Understanding is cognitive and analytical. It is the process by which one can **synthesize new knowledge** from what was already known.
          - Knowledge is information aggregated to a point where it has meaning and **purpose** – the how.
          - Data becomes information when it has **meaning** and we understand context and relationship – the who, what, where, and when.

What is the purpose of Visualization?
What does insight lead to?
Discovery
- Visualizing Patterns over time
- Spotting Differences

Decision Making
Analysis of Data
Explanation
Storytelling

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: ™ and © 2017 Twentieth Century Fox Film Corporation. All rights reserved.
Discovery
• Visualizing Patterns over time
• Spotting Differences
Decision Making
Analysis of Data
Explanation
   **Visualizing Spatial Relationships**
Storytelling

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

Story Telling with Visualization

Napoleon’s Invasion of Russia in 1812 By Jacque Minard

![Figurative Map of the movement from Moscow to St. Petersburg in 1812-1813](map.png)

- **Moscow**
- **Path of retreat**

- **Temperature**
  - -26
  - -30
  - -11
  - -21
  - -9

- **Army Size**
  - 422,000
  - 10,000
  - 100,000

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

- Explanation
- Tells a Story
- Analysis of Data
- Discovery
- Decision Making
FOUR TYPES OF VISUALIZATIONS

GEORGES GRINSTEIN (KEYNOTE PRESENTATION)

- **Exploratory**
  Have no hypotheses about the data
  Explore data interactively as undirected searches

- **Confirmatory**
  Have specific hypotheses about the data
  Goal-oriented examination of the hypotheses

- **Presentation**
  Facts to be presented are fixed a priori
  Select appropriate presentation techniques

- **Interactive**
  Interactions with a pre-defined animation
Visualization Applications

Why is visualization important?
Visualizația ațăului (BioVis)

The visualization of biological data;
Often grouped with computer animation

March 2010 | volume 7 | number 3
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.cerne.net/wp-content/uploads/2013/03/internet.gif
Geographic Visualization

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

Scientific Visualization (SciVis)

Primarily concerned with the visualization of three-dimensional phenomena
Emphases 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
Broadening Participation in Visualization
Vetria Byrd, Organizer
BPViz’17 in conjunction with IEEE VIS, October 2017, Phoenix, AZ
Data Visualization Process
High Level Overview
Data
Visualization Process

An iterative process

1. Obtain the data
2. Remove all but the data of interest
3. Choose a basic visual model, such as a bar graph, list or tree
4. Add methods for manipulating the data or controlling what features are visible

- Acquire
- Parse
- Filter
- Mine
- Represent
- Refine
- Interact

Provide structure

Apply methods from statistics or data mining to discern patterns or place the data in mathematical context

Improve the basic representation to make it clearer and more visually engaging

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.

Adopted from The ParaView Tutorial, The Basics of Visualization, version 3.98
There are several steps between raw data and a finished visualization.
Why do we care?
I just want a pretty picture!
What does the pretty picture mean?

A Pretty Picture is Nothing without Meaning.
What’s Missing?

http://www.bu.edu/tech/research/training/tutorials/introduction-to-scientific-visualization-tutorial/the-scientific-visualization-pipeline/
Scientific Visualization Pipeline: Step 1 . . .

**Produce Data**

Simulated Data Images Numerical Some measured value Observed Phenomena

**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

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**

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?
Why should you care about Visualization?
Data is Everywhere!
The Four V’s of Big Data

- **Volume**: Scale of Data
  - 40 Zettabytes (43 Trillion Gigabytes) of data will be created by 2020, an increase of 300 times from 2005.
  - It’s estimated that 2.5 Quintillion Bytes (2.3 Trillion Gigabytes) of data are created each day.
  - World Population: 7 Billion
  - Most companies in the U.S. have at least 100 Terabytes (100,000 Gigabytes) of data stored.

The Four V’s of Big Data

- The New York Stock Exchange captures **1 TB OF TRADE INFORMATION** during each trading session.
- Modern cars have close to **100 SENSORS** that monitor items such as fuel level and tire pressure.

**Velocity**
**ANALYSIS OF STREAMING DATA**

By 2016, it is projected there will be **18.9 BILLION NETWORK CONNECTIONS**—almost 2.5 connections per person on earth.

The Four V’s of Big Data

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

150 EXABYTES
(1.61 BILLION GIGABYTES)

30 BILLION PIECES OF CONTENT
are shared on Facebook every month

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

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

The Four V’s of Big Data

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.

Poor data quality costs the U.S. economy around $3.1 trillion a year.

Veracity: Uncertainty of Data

IBM Big Data Platform

The FOUR V’s of Big Data

Volume
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.

Velocity
By 2015, 4.4 million IT jobs will be created globally to support big data, with 1.9 million in the United States.

Variety
As of 2013, the global size of data in healthcare was estimated to be 150 exabytes (163 trillion gigabytes).

Veracity
By 2014, it’s anticipated there will be 420 million wearable, wireless health monitors.

40 ZETTABYTES
40 trillion gigabytes of data will be created by 2020, an increase of 300 times from 2005.

6 BILLION PEOPLE have cell phones.

The New York Stock Exchange captures 1 TB of trade information during each trading session.

100 SENSORS that monitor items such as fuel level and tire pressure.

V 2.5 QUIN TILLION BYTES
2.5 quintillion bytes of data are created each day.

WORLD POPULATION 7 BILLION

1 IN 3 BUSINESS LEADERS don’t trust the information they use to make decisions.

Poor data quality costs the US economy around $3.1 trillion a year.

27% OF RESPONDENTS in one survey were unsure of how much of their data was accurate.

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.

Source: McKinsey Global Institute, Twitter, Cisco, Gartner, EMC, SAS, IBM, MEPTEC, GAS
Why Should You Care About Visualization

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

~ Vetria Byrd
WHY SHOULD YOU CARE?

There is a demand for

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

Regardless of major, research interest, or academic background, etc., at some point you will visualize some type of data. ~ Vetria Byrd, 2014

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

July 31, 2014
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?

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 student version is free)

Scientific Visualization
- ParaView
- VisIt

Geo Visualization
- ARC GIS
- D3.js

Cyber Security Visualization
- Survey of Security Visualization
- Survey of Cybersecurity 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.
What does this person look like?
THIS PERSON LOOKS LIKE YOU!

Computer Science
Biology
Engineering

Journalism
Industry
Chemistry

Physics
Non-STEM

The Arts

Research

Academia

Statistics

STEM

Social Media

Athletics
ANATOMY OF A DATA SCIENTIST

The era of Big Data has created a talent gap for people who can pull actionable insights out of raw data. The data scientist—called “the sexiest job of the 21st century” by Harvard Business Review—is in demand, with a 15,000% jump in job posts between 2011–2012. In the US, the average salary for these sought-after scientists is around $100,000.

So what makes a good data scientist?

http://www.houghtoncdsa.org/liberal-arts-data-science-seriously/
WHAT MAKES A GOOD DATA SCIENTIST?

- Degreed In Geek
- Problem Solving Prowess
- Mathlete
- Suit-Able
- Insight Whisperer
- Quantastic
- Curiouser and Curiouser
- Agile and Adaptive

http://www.houghtoncdsa.org/liberal-arts-data-science-seriously/

Sources:
www.indeed.com/jobtrends?q=%22Data+Scientist%22&l=%5D&relative=1
http://management.fortune.cnn.com/2013/05/10/big-data-jobs/
http://hbr.org/2012/10/data-scientist-the-sexiest-job-of-the-21st-century/ar/1
www.fico.com
www.indeed.com/salary?q1=data+scientist&l1=&t=1
www.payscale.com/research/US/Job=Data_Scientist_%2F_Engineer/Salary
WHAT MAKES A GOOD DATA SCIENTIST?

- Degreed In Geek – It doesn’t hurt to have a background and hands on experience in data visualization

Image Source: https://www.pinterest.com/pin/213076626089856136/

http://www.houghtoncdsa.org/liberal-arts-data-science-seriously/

Sources:
www.indeed.com/jobtrends?q=%22Data+Scientist%22&t=%5D&relative=1
http://management.fortune.cnn.com/2013/05/10/big-data-jobs/
http://hbr.org/2012/10/data-scientist-the-sexiest-job-of-the-21st-century/ar/1
www.fico.com
www.indeed.com/salary?q1=data+scientist&t1=&tm=1
www.payscale.com/research/US/Job=Data_Scientist_%2E_F_Engineer/Salary
WHAT MAKES A GOOD DATA SCIENTIST?

- **Problem Solving Prowess** – A problem solver at heart who’s able to devise creative solutions to real-world problems.
- Knows how to define those problems precisely, spot elusive patterns and connect the dots

http://www.houghtoncdsa.org/liberal-arts-data-science-seriously/

Sources:
- www.indiegogo.com/jobtrends?q=%22Data+Scientist%22&l=%5D&relative=1
- http://management.fortune.cnn.com/2013/05/10/big-data-jobs/
- www-01.ibm.com/software/data/infosphere/data-scientist/
- www.fico.com
- www.indiegogo.com/salary?q1=data+scientist&l1=&tm=1
- www.payscale.com/research/US/Job=Data_Scientist_%2F_Engineer/Salary
WHAT MAKES A GOOD DATA SCIENTIST?

- Mathlete
- Strong math skills are table stakes

Do NOT let this scare you!

There is a visualization component.

http://www.houghtoncdsa.org/liberal-arts-data-science-seriously/

Sources:
www.indeed.com/jobtrends?q=%22Data+Scientist%22&l=%5D&relative=1
http://management.fortune.cnn.com/2013/05/10/big-data-jobs/
http://hbr.org/2012/10/data-scientist-the-sexiest-job-of-the-21st-century/ar/1
www-01.ibm.com/software/data/infosphere/data-scientist/
www.fico.com
www.indeed.com/salary?q1=data+scientist&l1=&t1=1
www.payscale.com/research/US/Job=Data_Scientist_%2F_Engineer/Salary
WHAT MAKES A GOOD DATA SCIENTIST?

- **Suit-Able** – Has the know-how and finesse to be a business leader.
- Today, data scientists can lead from the backroom to the boardroom

http://www.houghtoncdsa.org/liberal-arts-data-science-seriously/

**Sources:**
- www.indeed.com/jobtrends?q=%22Data+Scientist%22&l=%5D&relative=1
- http://management.fortune.cnn.com/2013/05/10/big-data-jobs/
- www-01.ibm.com/software/data/infosphere/data-scientist/
- www.fico.com
- www.indeed.com/salary?q1=data+scientist&l1=&tm=1
- www.payscale.com/research/US/Job=Data_Scientist_%2F_Engineer/Salary
WHAT MAKES A GOOD DATA SCIENTIST?

- Insight Whisperer – Can develop unique insights, apply them to solve problems and explain them to people without overwhelming them.

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

~Ben Shneiderman

Sources:
www.indeed.com/jobtrends?q=%22Data+Scientist%22&l=%5D&relative=1
http://management.fortune.cnn.com/2013/05/10/big-data-jobs/
http://hbr.org/2012/10/data-scientist-the-sexiest-job-of-the-21st-century/ar/1
www-01.ibm.com/software/data/infosphere/data-scientist/
www.fico.com
www.indeed.com/salary?q1=data+scientist&l1=&tm=1
www.payscale.com/research/US/Job=Data_Scientist_%2F_Engineer/Salary
WHAT MAKES A GOOD DATA SCIENTIST?

- **Quantastic** – successful data scientists come not only from math backgrounds, but also from many other fields.
- They have programming skills or . . .
- The ability to learn programming languages and represent concepts via computer code

http://www.houghtoncdsa.org/liberal-arts-data-science-seriously/

**Sources:**
- www.indeed.com/jobtrends?q=%22Data+Scientist%22&l=%5D&relative=1
- http://management.fortune.cnn.com/2013/05/10/big-data-jobs/
- www-01.ibm.com/software/data/infosphere/data-scientist/
- www.fico.com
- www.indeed.com/salary?q1=data+scientist&l1=&tm=1
- www.payscale.com/research/US/Job=Data_Scientist_%2F_Engineer/Salary
WHAT MAKES A GOOD DATA SCIENTIST?

- Curiouser and Curiouser – Critical Thinking and a relentlessly inquisitive nature are at the center of an analytic mindset

When you’re CURIOUS you find lots of interesting things to do.
- Walt Disney

http://www.houghtoncdsa.org/liberal-arts-data-science-seriously/

Sources:
- www.indeed.com/jobtrends?q=%22Data+Scientist%22&l=%26D&relative=1
- http://management.fortune.cnn.com/2013/05/10/big-data-jobs/
- www-01.ibm.com/software/data/infosphere/data-scientist/
- www.fico.com
- www.indeed.com/salary?q1=data+scientist&l1=&tm=1
- www.payscale.com/research/US/Job=Data_Scientist_%2F_Engineer/Salary
WHAT MAKES A GOOD DATA SCIENTIST?

- Agile and Adaptive – versatile enough to apply their expertise to multiple industries, from retail to banking, insurance to government, healthcare to airlines

The possibilities are endless!

http://www.houghtoncdsa.org/liberal-arts-data-science-seriously/

Sources:
www.indeed.com/jobtrends?q=%22Data+Scientist%22&l=%5D&relative=1
http://management.fortune.cnn.com/2013/05/10/big-data-jobs/
http://hbr.org/2012/10/data-scientist-the-sexiest-job-of-the-21st-century/ar/1
www-01.ibm.com/software/data/infosphere/data-scientist/
www.fico.com
www.indeed.com/salary?q1=data+scientist&l1=&tn=1
www.payscale.com/research/US/Job=Data_Scientist_%2FSalary
Data Scientists

Places for Employment

- Government agencies
- Science Institutes
- Retail Companies (that analyze large amounts of data to drive their business)
  - Amazon
  - Target
  - Netflix

Data Scientists

Additional Fields of Employment

- Healthcare
- Journalism
- Biotech
- Finance
- Insurance
- Hospitality
- Manufacturing
- Transportation

https://keshif.me/demo/VisTools?utm_content=26335725&utm_medium=social&utm_source=twitter
Questions?
Vetria L. Byrd
Assistant Professor
Computer Graphics Technology
vlbyrd@purdue.edu
https://polytechnic.purdue.edu/profile/vbyrd
http://web.ics.purdue.edu/~vbyrd/
@VByrdPhD, @BPViz

Purdue Polytechnic Institute