Introduction to Data Visualization

Vetria L. Byrd, PhD
Assistant Professor
Computer Graphics Technology
Purdue University

Summer Undergraduate Research Fellowship Program (SURF Program)
Purdue University

June 13, 2017
ABOUT THIS TALK

ASSUMPTIONS

Target Audience: Beginners

Pre-requisites: No prior knowledge of visualization
ABOUT THIS TALK

GOALS

1. Provide 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
By the end of this presentation, you will
1. Understand the purpose of visualization
2. Be able to identify your data visualization needs
3. Identify some visualization tools available to assist with visualizing your data
AGENDA
INTRODUCTION TO DATA VISUALIZATION

- Brief Introduction
- Purpose Of Visualization
- High Level Overview
- Visualization Applications
- You’ve Got Data, Now What?
- Q&A
What majors are represented today?
About Me

Vetria L. Byrd, PhD

Academic Preparation

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

What I’ve Done

Visualization Initiatives

- Research Experience for Undergraduates in Collaborative Data Visualization Applications (2014/2015)

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
Research Interests

• Data Visualization
• Big Data
• Visualization of Heterogeneous Data
• Uncertainty Visualization
• High Performance Computing (HPC) for Data Visualization
• Broadening Participation in Visualization

Faculty Liaison for Purdue University’s Women in HPC Group
FOUR TYPES OF VISUALIZATIONS

GEORGES GRINSTEIN (KEYNOTE PRESENTATION, VINCI 2016)

- **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
DATA VISUALIZATIONS
WE’VE ALL SEEN THEM

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

Engineering Intelligence Through Data Visualization at Uber
https://eng.uber.com/data-viz-intel/

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.

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

Internet Users in the World (per 100 people)

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

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.

Last accessed 02/27/17
What is the purpose of Visualization?
"The purpose of visualization is "insight", not pictures."

~Ben Shneiderman

Advancing Beyond Data to True Insight

Data → Relevance

Relationship

Data becomes information when it has meaning and we understand context and relationship—the who, what, where, and when.

FROM DATA TO INSIGHT

Advancing Beyond Data to True Insight

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.

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.

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. Information becomes data when it has meaning and we understand context and relationship – the who, what, where, and when.

FROM DATA TO INSIGHT

Advancing Beyond Data to True Insight

Wisdom

Understanding

Knowledge

Information

Data

What does insight lead to?
Discovery

- Visualizing Patterns over time
- Spotting Differences

Decision Making
Analysis of Data
Explanation
Storytelling

INSIGHT LEADS TO

Discovery

Spotting Differences

- Visualizing Patterns
- Spotting Differences

How many 7’s do you see?
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
Discovery
- Visualizing Patterns over time
- Spotting Differences

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.
INSIGHT LEADS TO

Discovery
- Visualizing Patterns over time
- Spotting Differences

Decision Making
Analysis of Data

Explanation

Visualizing Spatial Relationships

Storytelling

Watch the Growth of Walmart and Sam's Club
http://datafl.ws/197

Watch the Growth of Target Stores
http://datafl.ws/198

Story Telling with Visualization

Napoleon’s Invasion of Russia in 1812 By Jacque Minard

Temperature

Army Size: 422,000

Army Size: 10,000

Army Size: 100,000

Path of retreat

Moscow

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

Temperatures:
-26
-30
-11
-21
-9
Hans Rosling’s 200 Countries, 200 Years, 4 minutes

The Joy of Stats - BBC Four
Visualization Applications

Why is visualization important?
Biovisualization (BioVis)

The visualization of biological data;
Often grouped with computer animation
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.cernea.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
Data Visualization Process
High Level Overview

There are great subtleties involved in the collection of data and its conversion into information.

Data Visualization Process

An iterative process

- obtain the data
- acquire

- parse
- filter
- mine

- represent
- refine
- interact

provide structure

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

- 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.

This is where the magic happens!

Adopted from The ParaView Tutorial, The Basics of Visualization, version 3.98
Why do we care?
I just want a pretty picture!
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

Produce Input Data

Input Data

Analyze, Filter, Reformat

Prepared Data

Apply Sci Vis Techniques

Sci Vis Model Data

Map to Geometry

Computer Graphics Data

Render, Postprocess

Image data

View Results

---

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

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

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
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/
Why Should I 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
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?
REMEMBER THE FOUR TYPES OF VISUALIZATIONS

- **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
I just need a picture for my paper or poster.
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?

Starting Point: 38 best tools for data visualization
http://www.creativebloq.com/design-tools/data-visualization-712402

More types of data: Biological, Social Media, Network Data, Survey Data, Cybersecurity, Temporal, Image data, Topical, . . . this is NOT an exhaustive list
https://www.labnol.org/software/find-right-chart-type-for-your-data/6523/
I would like to program or create my own data visualization tool
5 Libraries for Building Data Visualizations

Which one is the best programming languages for data visualization?

Data Visualization Basics with the R Programming Language
I am not a programmer but
I am interested in data visualization
VISUALIZATION TOOLS

A STARTING POINT: OPEN SOURCE VISUALIZATION TOOLS

Information Visualization

Gephi (https://gephi.org/)
Tableau (not open source but free)
https://www.tableau.com/products/desktop

Scientific Visualization

ParaView (https://www.paraview.org/)
VisIt (https://wci.llnl.gov/simulation/computer-codes/visit/)

Geo Visualization

ARC GIS (https://www.arcgis.com/features/index.html)
D3.js (https://d3js.org/)
Civil Engineering Tools
http://guides.lib.purdue.edu/c.php?g=352179&p=3057468

Biological Visualization Tools
https://www.aiche.org/sbe/resources/resource-directory/software

Mathematics
http://visualizingmath.tumblr.com/

Data Visualization Basics with the R Programming Language
Interactive visualization tools:
https://constructive.co/insights/6-best-data-visualization-tools-2016-pt-1/

D3.js (https://d3js.org/)
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.
Data is Everywhere!
The Four V’s of Big Data

40 ZETTABYTES (45 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.

6 BILLION PEOPLE have cell phones.

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.
- 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 (161 BILLION GIGABYTES)**

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

**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

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.

Veracity - Uncertainty of Data

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

IBM Big Data Platform

**Volume - Scale of Data**
- 40 Zettabytes (43 trillion gigabytes) of data will be created by 2020, an increase of 300 times from 2005.
- 6 billion people have cell phones.
- World population 7 billion.

**Velocity - Analysis of Streaming Data**
- By 2015, 4.4 million IT jobs will be created globally to support big data, with 1.9 million in the United States.
- By 2016, it is projected there will be 18.9 billion network connections - almost 2.5 connections per person on earth.
- 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.

**Variety - Different Forms of Data**
- 30 billion pieces of content are shared on Facebook every month.
- 400 million tweets are sent per day by about 300 million monthly active users.
- 4 trillion hours of video are watched on YouTube each month.

**Veracity - Uncertainty of Data**
- In one survey, 27% of respondents were unsure of how much of their data was inaccurate.
- 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.

**The Four V's of Big Data**
- Volume: With 40 zettabytes of data expected by 2020, the sheer amount of data is staggering.
- Variety: With data in all forms, from structured to unstructured, big data requires unique analysis.
- Velocity: The speed at which data is generated and accessed is crucial for real-time decision-making.
- Veracity: Determining the accuracy and quality of data is essential for trust and reliability.

As of 2011, the global size of data in healthcare was estimated to be 150 exabytes (1.3 trillion gigabytes).

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

Sources: McKinsey Global Institute, Twitter, Cisco, Gartner, EMC, SAS, IBM, NPIPEC, GSA

**VISUALIZATION TOOLS FOR BIG DATA**

**THIS IS NOT AN EXHAUSTIVE LIST**

- **Google charts** ([https://developers.google.com/chart/](https://developers.google.com/chart/)) (Free) – display live data on your website
- **Tableau Desktop** ([https://www.tableau.com/products/desktop](https://www.tableau.com/products/desktop)) – 14 day free trial
- **D3** ([https://d3js.org/](https://d3js.org/)) : Data Driven Document (Free): Javascript library for visualising big data
- **Fusion Chart** ([http://www.fusioncharts.com/](http://www.fusioncharts.com/)) - Javascript charting library for the web and mobile devices, spread across 120 countries with having clients such as Google, Intel, Microsoft and many others. However, you need a bit knowledge on Javascript for implementing it.
- **High Charts** ([https://www.highcharts.com/](https://www.highcharts.com/)) Highcharts is a charting library written purely in Javascript hence, a bit knowledge of Javascript is necessary for implementing this tool. It uses HTML5, SVG and VML for displaying charts across various browsers (from IE6+) and devices like android, iPhone etc.
- **Canvas** ([http://canvasjs.com/](http://canvasjs.com/))

**A review of 20 big data visualization tools:**
Why is Visualization Important?
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 – It doesn’t hurt to have a background and hands on experience in data visualization

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

Sources:
- https://www.indoed.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=&t=1
- www.payscale.com/research/US/Job=Data_Scientist_%2F_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

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?

• Mathlete
• Strong math skills are table stakes

Do NOT let this scare you!

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&l=&&&tm=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/
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?

- **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

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=&t1=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=%25D: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=&t=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=%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?

- 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/
- www-01.ibm.com/software/data/infosphere/data-scientist/
- www.fico.com
- www.indeed.com/salary?q1=data+scientist&l=1&tm=1
- www.payscale.com/research/US/Job=Data_Scientist_%2F_Engineer/Salary
WANT MORE?

https://keshif.me/demo/VisTools?utm_content=26335725&utm_medium=social&utm_source=twitter
Visualization is a Process

- Your 1st visualization will not be your last

- The 1st visualization tool you use will not be your last
The purpose of visualization is "Insight"
Questions?
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