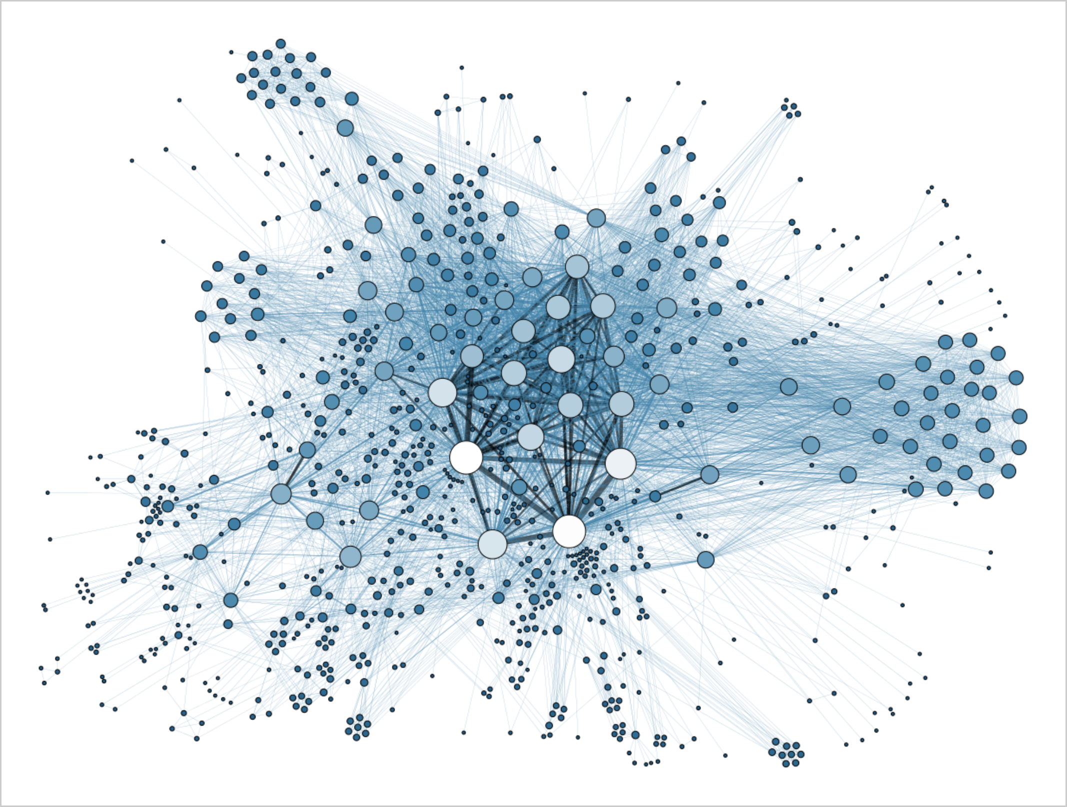
DATA VISUALIZATION

John Gentle
Data Management & Collections Group
Texas Advanced Computing Center
jgentle@tacc.utexas.edu

OVERVIEW

Visualization is the representation of an object, situation, or set of information as a chart or other image.



> *Graph representing the metadata of thousands of archive documents, documenting the social network of hundreds of League of Nations personals*

Visualization has applications in science, education, engineering, interactive multimedia, medicine, etc. (a typical application of visualization is the field of computer graphics).

THE INVENTION OF COMPUTER GRAPHICS MAY BE THE MOST IMPORTANT DEVELOPMENT IN VISUALIZATION SINCE THE INVENTION OF CENTRAL PERSPECTIVE IN THE RENAISSANCE PERIOD.

VISUALIZATION CATEGORIES

Visualizations are broadly classified into three basic categories:

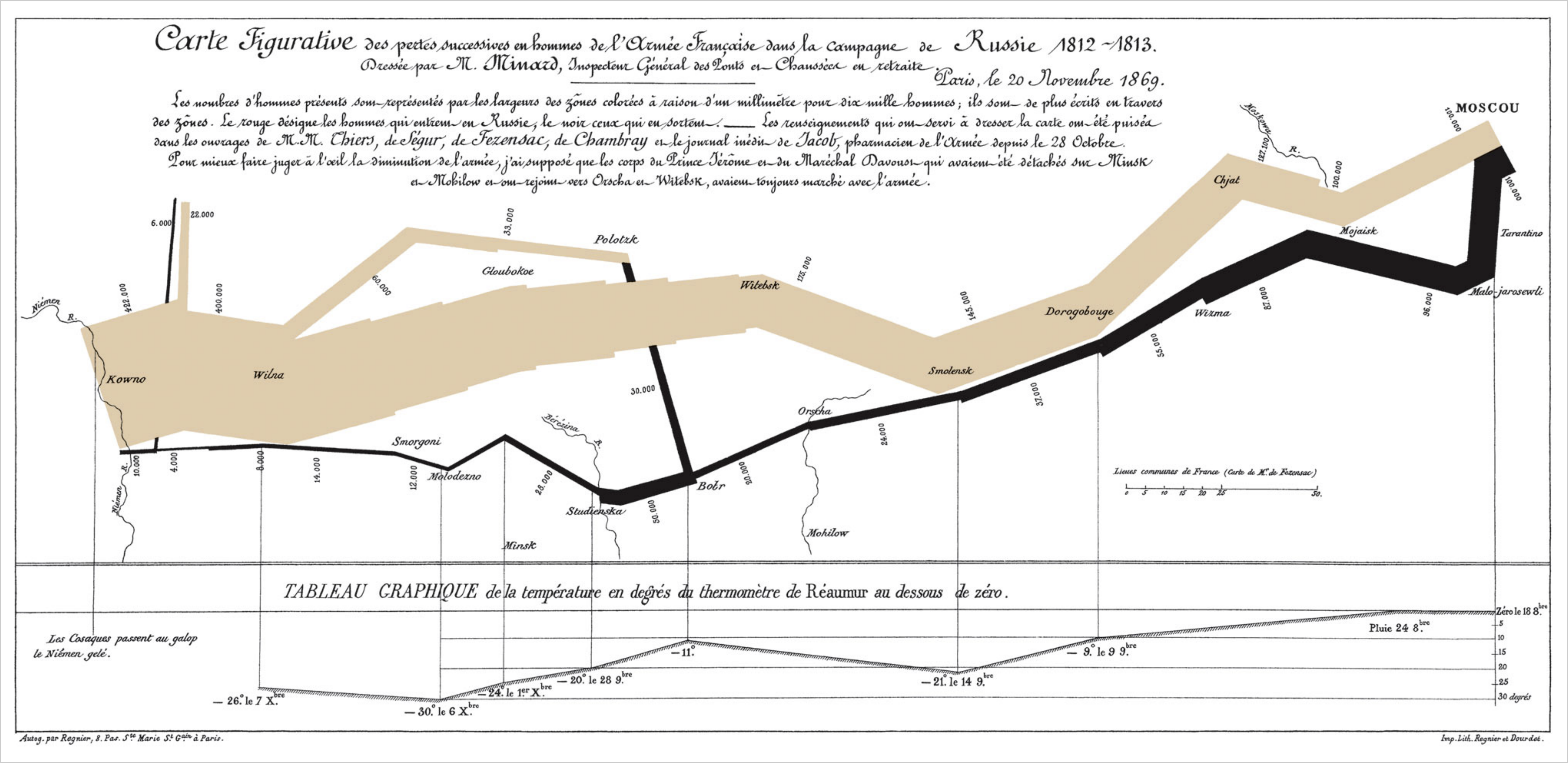
Visualization Category	*Format Changes*	*Input Required*
Static	No	No
Dynamic	Yes	No
Interactive	Either	Yes

A **static** visualization displays information in an unchanging format and requires no user input.

A **dynamic** visualization displays information in a changing state (e.g. motion graphics) but requires no user input.

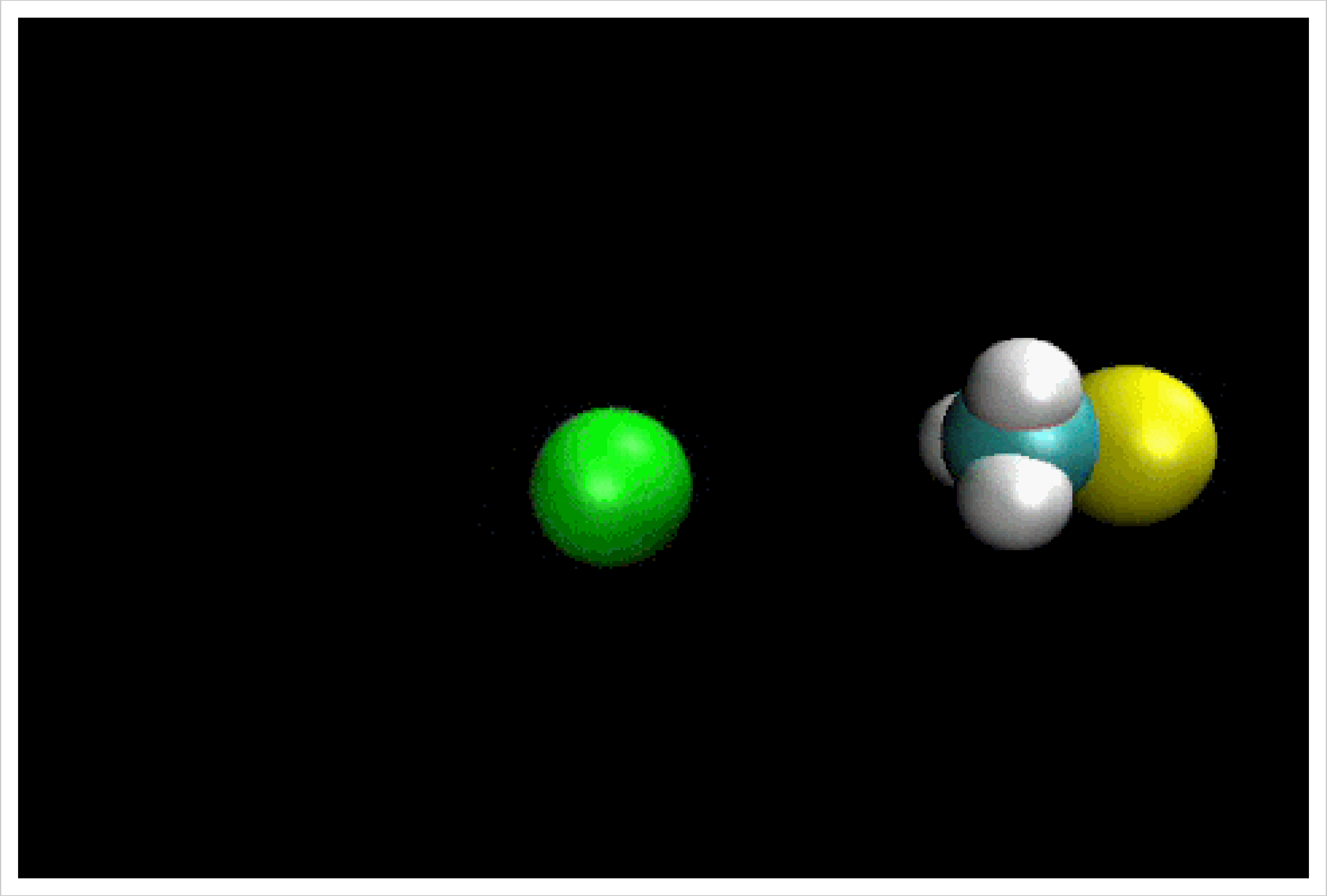
An **interactive** visualization displays information in a static *or* dynamic state but will change based on user input.

STATIC VISUALIZATION



Charles Minard's 1869 chart showing the number of men in Napoleon's 1812 Russian campaign army, their movements, as well as the temperature they encountered on the return path.

DYNAMIC VISUALIZATION



Cl- + CH3I- -> ClCH3 + I- SN2 Reaction Dynamics. Both direct [left] and indirect [right] reaction mechanisms are observed in the trajectories. The direct reaction occurs via the traditional SN2 mechanism with Cl- colliding backside with direct displacement of I- . The indirect reaction occurs via a CH3 rotation, roundabout mechanism in which Cl- strikes the CH3-moiety causing it to rotate around the I-atom and Cl- displacing I- after one CH3 revolution with the backside of the CH3-moiety positioned next to Cl-.

INTERACTIVE VISUALIZATION

Stream Graph (Mike Bostock)

Node Graph (Mike Bostock)

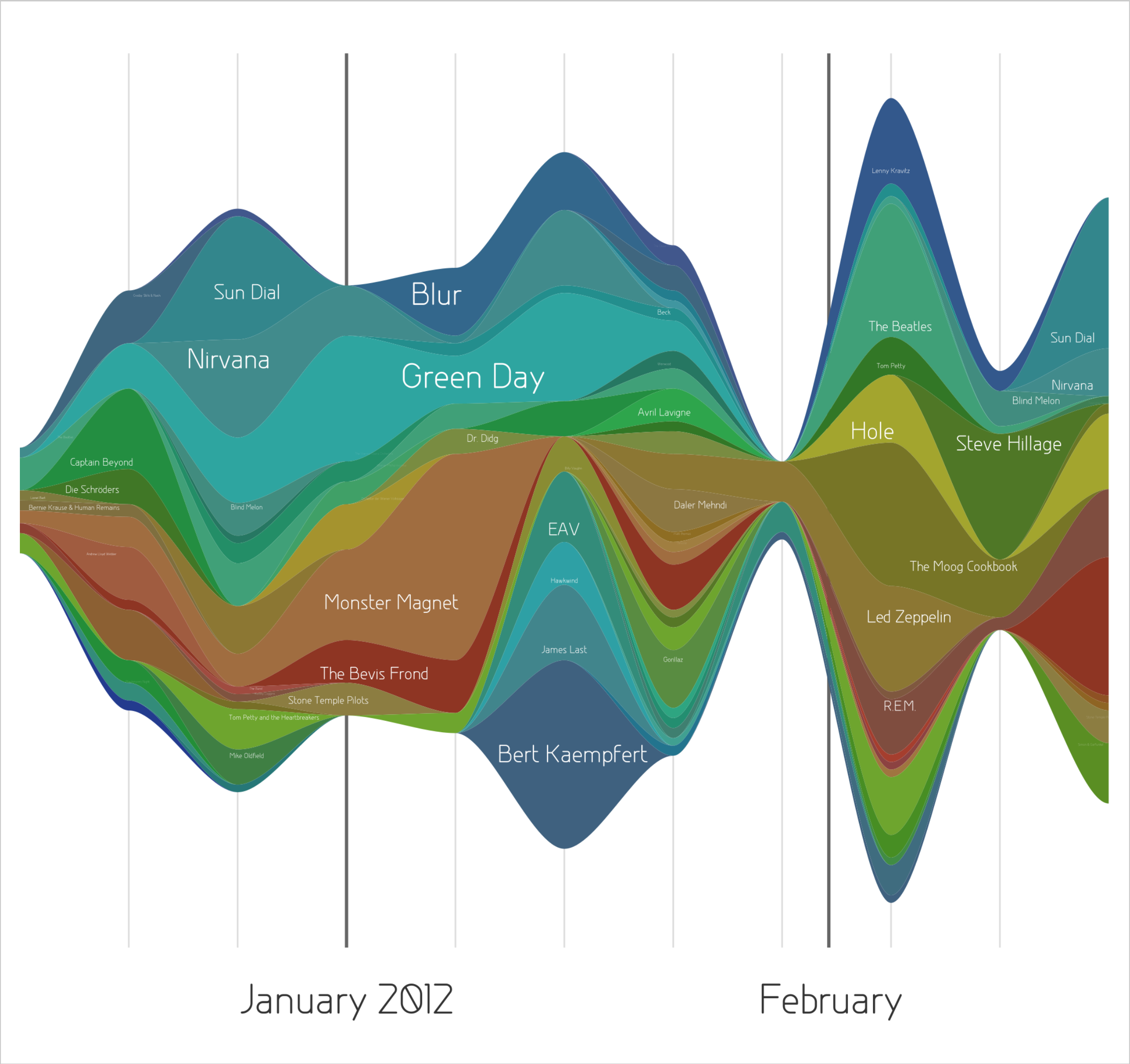
DISTINGUISHING VISUALIZATION TYPES

Apart from the categorization between static, dynamic and interactive visualizations, the next useful distinction is between the types of visualization: abstract and model-based.

Abstract visualizations show completely conceptual constructs in 2D or 3D. These generated shapes are completely arbitrary.

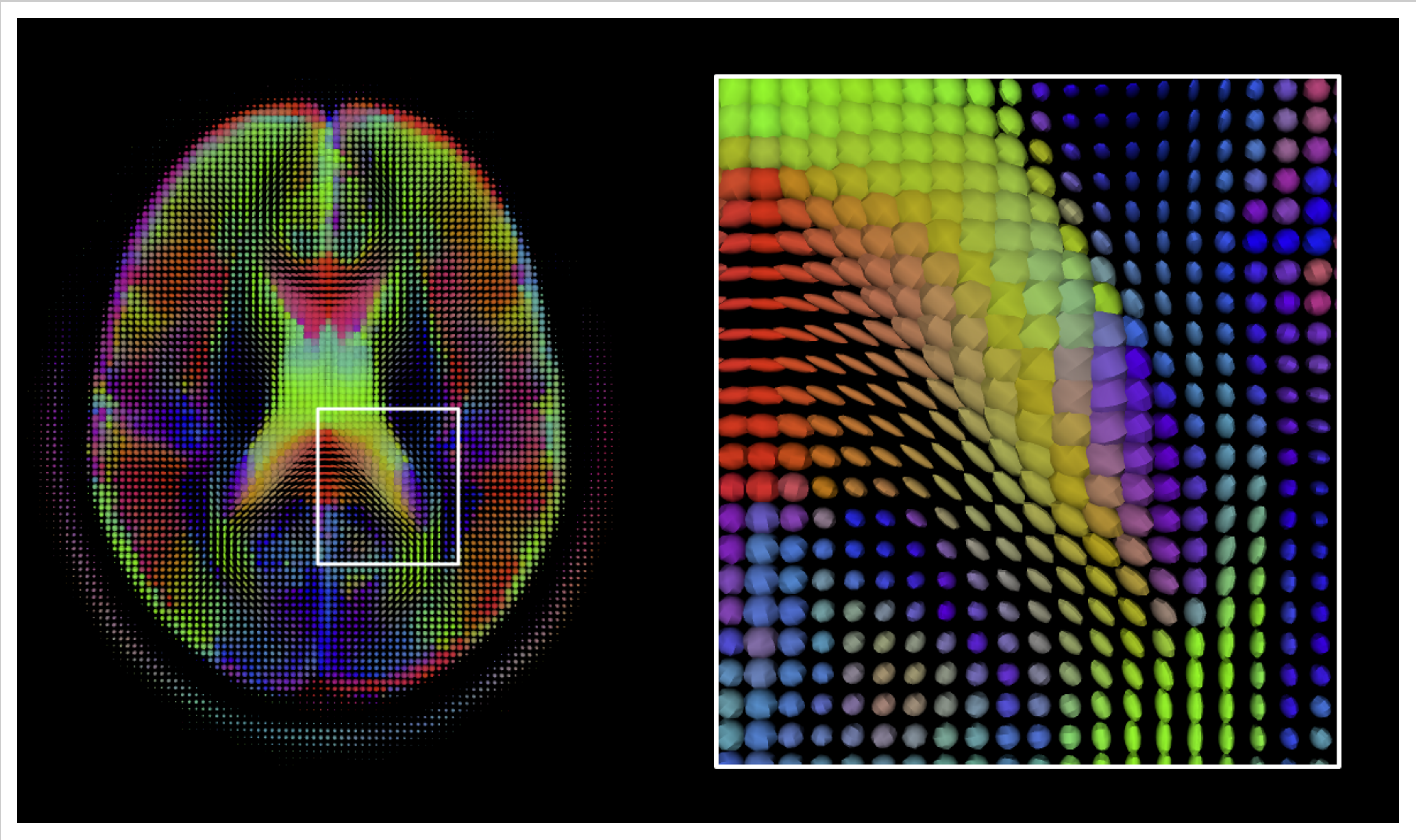
Model-based visualizations either place overlays of data on real or digitally constructed images of reality or make a digital construction of a real object directly from the scientific data.

ABSTRACT VISUALIZATION



A streamgraph showing someone's Last.fm music listening habits over time.

MODEL-BASED VISUALIZATION



The tensor field in a diffusion MRI image. A mid-axial slice of the ICBM diffusion tensor image template. Each voxel's value is a tensor represented here by an ellipsoid. Color denotes principal orientation: red = left-right, blue=inferior-superior, green = posterior-anterior.

SPATIAL REPRESENTATION

Another way to think of the two visualization types (abstract vs. model-based) is in terms of their *Spatial Representation*.

The representational constraints required by the visualization will often indicate the area of visualization that would be most appropriate to communicate the findings.

Spatial Representation Constraints	***Visualization Area***
Researchers Choice (Any)	*Data Visualization*, *Information Visualization*, *Knowledge Visualization*
Supplied (2D, 3D)	*Product Visualization*, *Educational Visualization*, *Scientific Visualization*
Supplied (GIS)	*Geospatial Visualization*

AREAS OF VISUALIZATION

Data visualization can be divided into several broad areas:

- Data Visualization
- Information Visualization
- Knowledge Visualization
- Educational Visualization
- Product Visualization
- Scientific Visualization
- Geographic Visualization

The first three areas are roughly ordered according to the *data-information-knowledge cycle* whereby data is structured and organized into information, information is structured and organized into knowledge and knowledge is used to make an informed decision (based on contextualized data).

The distinctions between the subsequent areas of educational, product, scientific and geographic visualizations are derived from their intended use and the spatial reference constraints inherent in them.

AREAS OF VISUALIZATION (TABLE)

Area	*Type (Spatial Representation)*	*Categories (Static, Dynamic)*	*Interactive*
Data Visualization	Any	Any	Optional
Information Visualization	Any	Any	Optional
Knowledge Visualization	Any	Any	Optional
Educational Visualization	Any	Any	Optional
Product Visualization	Model-based (usually)	Any	Optional
Scientific Visualization	Model-based	Any	Optional
Geographic Visualization	Model-based (GIS)	Any	Yes*

Geographic visualization is usually interactive, but not always.

DATA VISUALIZATION

Data Visualization is the presentation of data in a visualization format.

It is often considered the modern analog of *Visual Communication*.

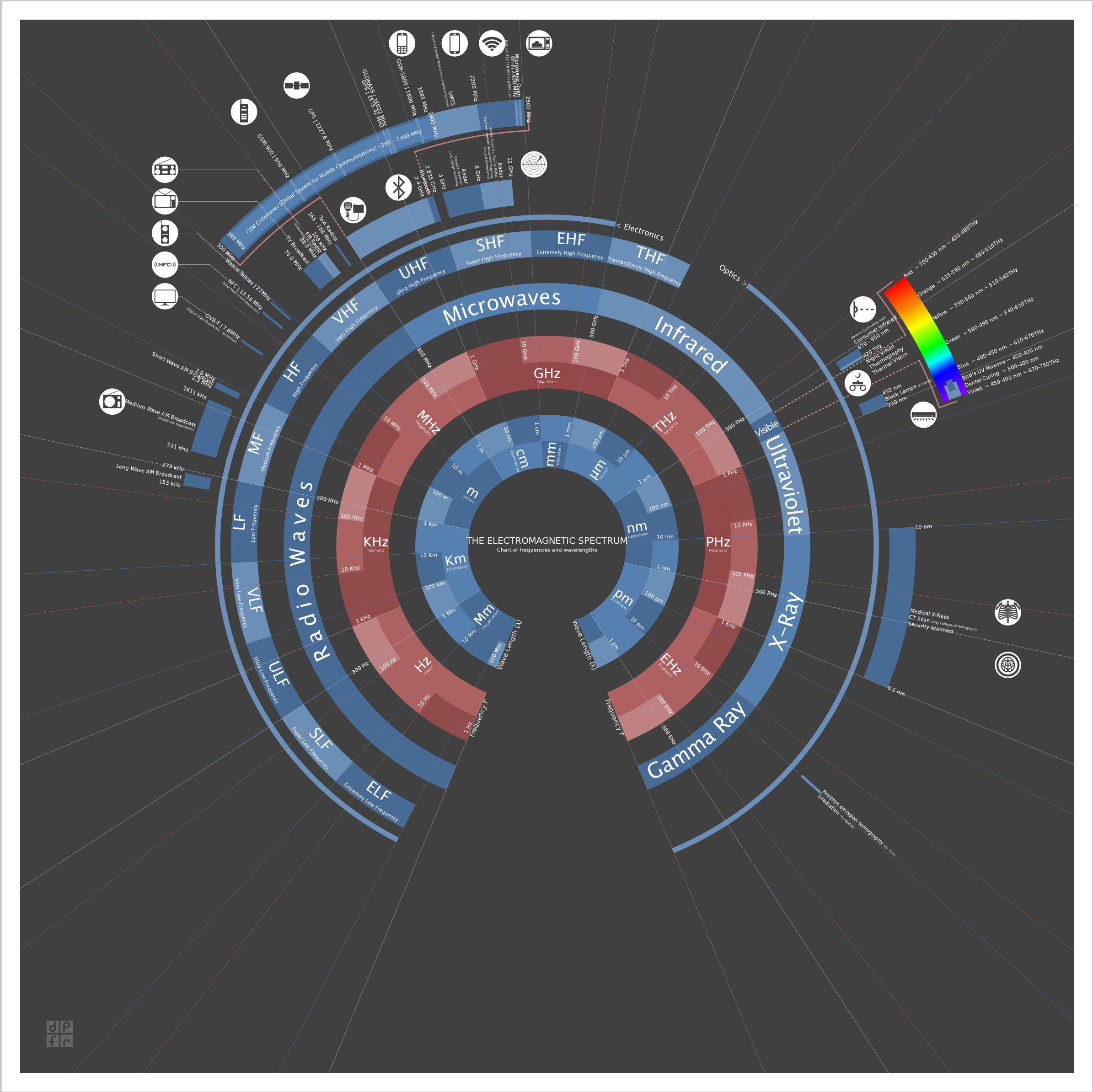
Data is the raw material of information. It is the product of research and discovery.

A single piece of data has no meaning unless the context is understood.

Data needs to be *transformed* into information.

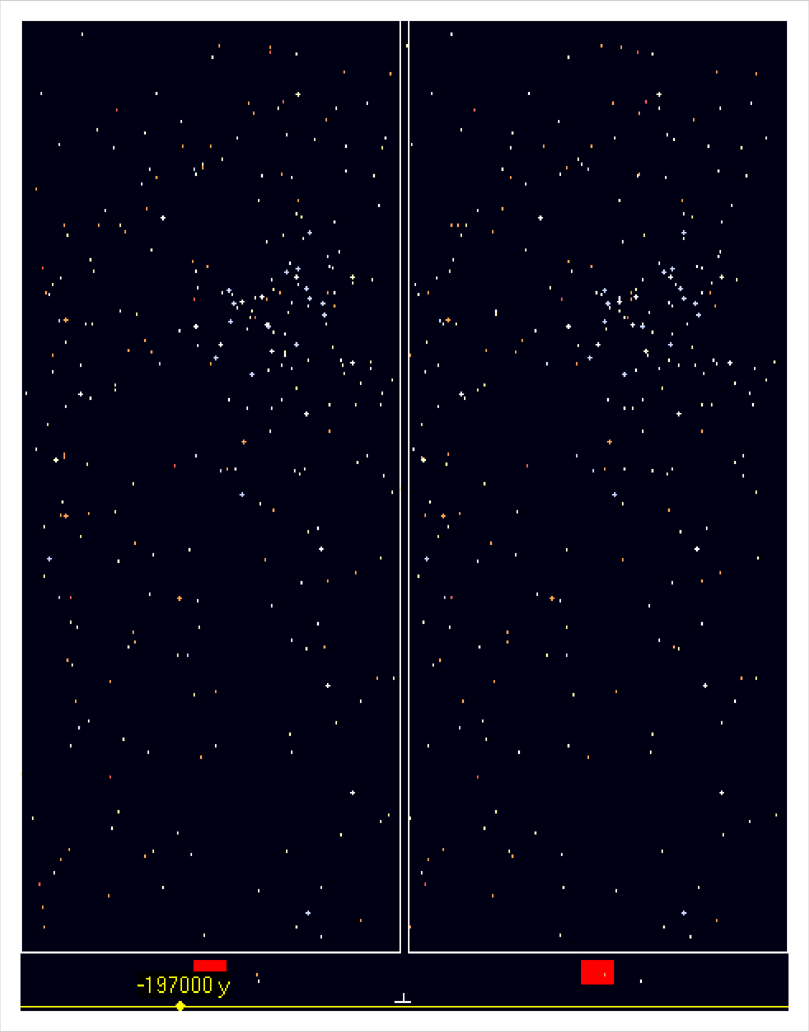
> *Individual bits or "bytes" of "raw" biological data (e.g. the number of individual plants of a given species at a given location) do not by themselves inform the human mind. However, drawing various data together within an appropriate context yields information that may be useful (e.g. the distribution and abundance of the plant species at various points in space and time). In turn, this information helps foster the quality of knowing (e.g. whether the plant species is increasing or decreasing in distribution and abundance over space and time). Knowledge and experience blend to become wisdom--the power of applying these attributes critically or practically to make decisions.*
>
> *~ The Atlantic Canada Conservation Data Centre*

DATA VISUALIZATION (STATIC)



The Electromagnetic Radiation Spectrum Infographic with labels and commonly used or known frequencies and wavelengths.

DATA VISUALIZATION (DYNAMIC)



Pleiades - the movement of stars in -/+ 200 000 years. To view this picture you need cross-eyed viewing.

DATA VISUALIZATION (INTERACTIVE)

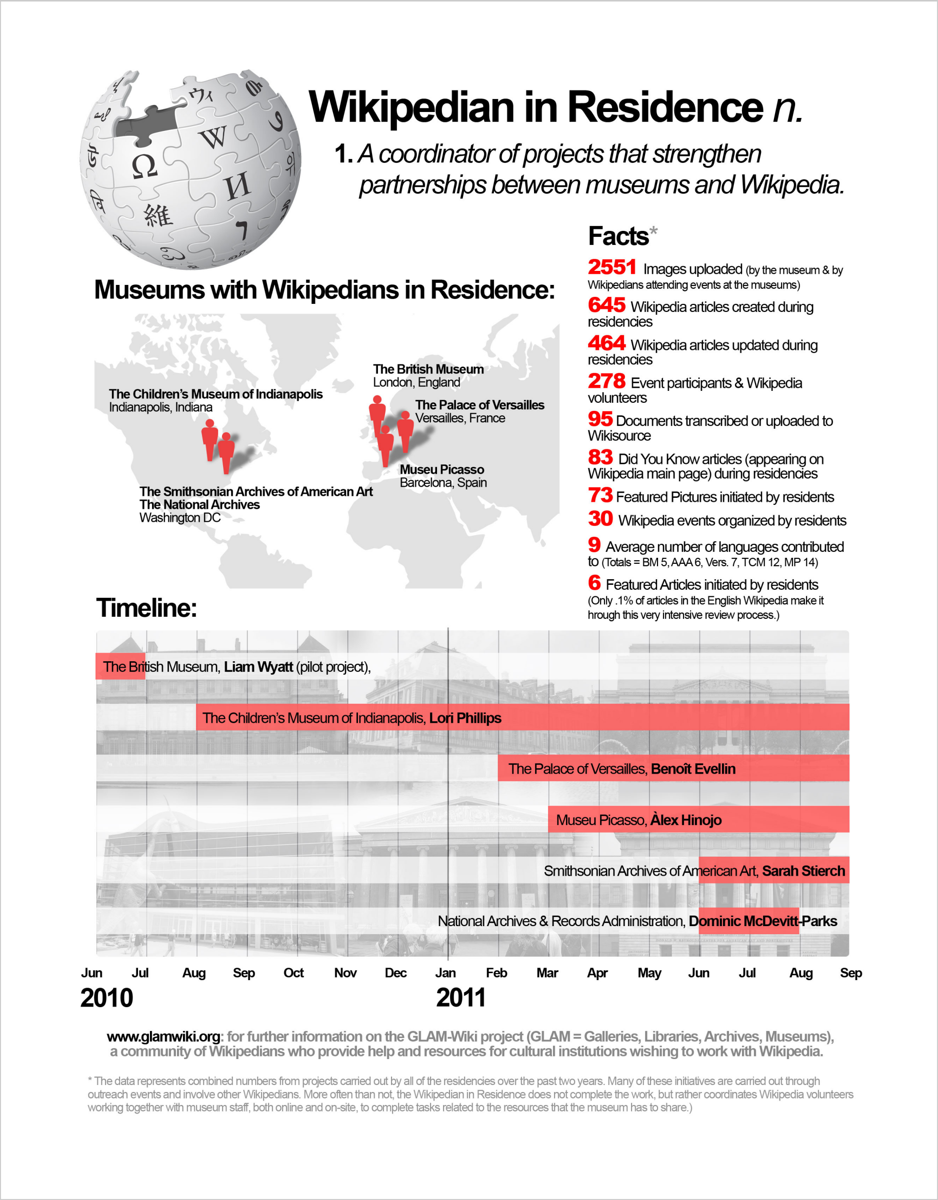
Zoomable Tree Map (Mike Bostock)

Sunburst Graph (Mike Bostock)

INFORMATION VISUALIZATION

Information Visualization is the study of visual representations of abstract data to reinforce human cognition and derive new insights.

INFORMATION VISUALIZATION (STATIC)



INFORMATION VISUALIZATION (DYNAMIC)

Wind Map

INFORMATION VISUALIZATION (INTERACTIVE)

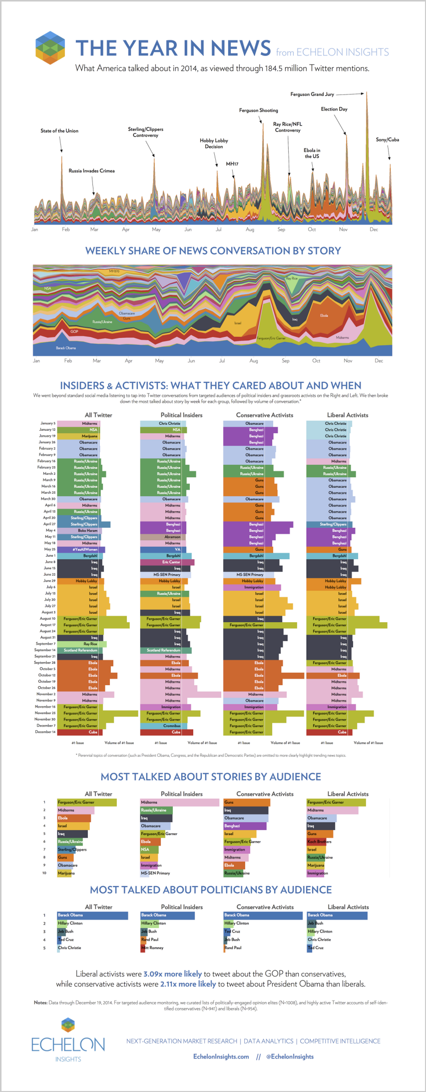
The Depth of the Problem

MVP Sports Franchises

KNOWLEDGE VISUALIZATION

Knowledge Visualization is the use of visual representations to improve the transfer of knowledge between two or more individuals by using computer and non-computer-based visualization methods complementarily, focusing on transferring insights and creating new knowledge in groups.

KNOWLEDGE VISUALIZATION (STATIC)



link

KNOWLEDGE VISUALIZATION (DYNAMIC)

A day in the life: US Thanksgiving on Google Flights

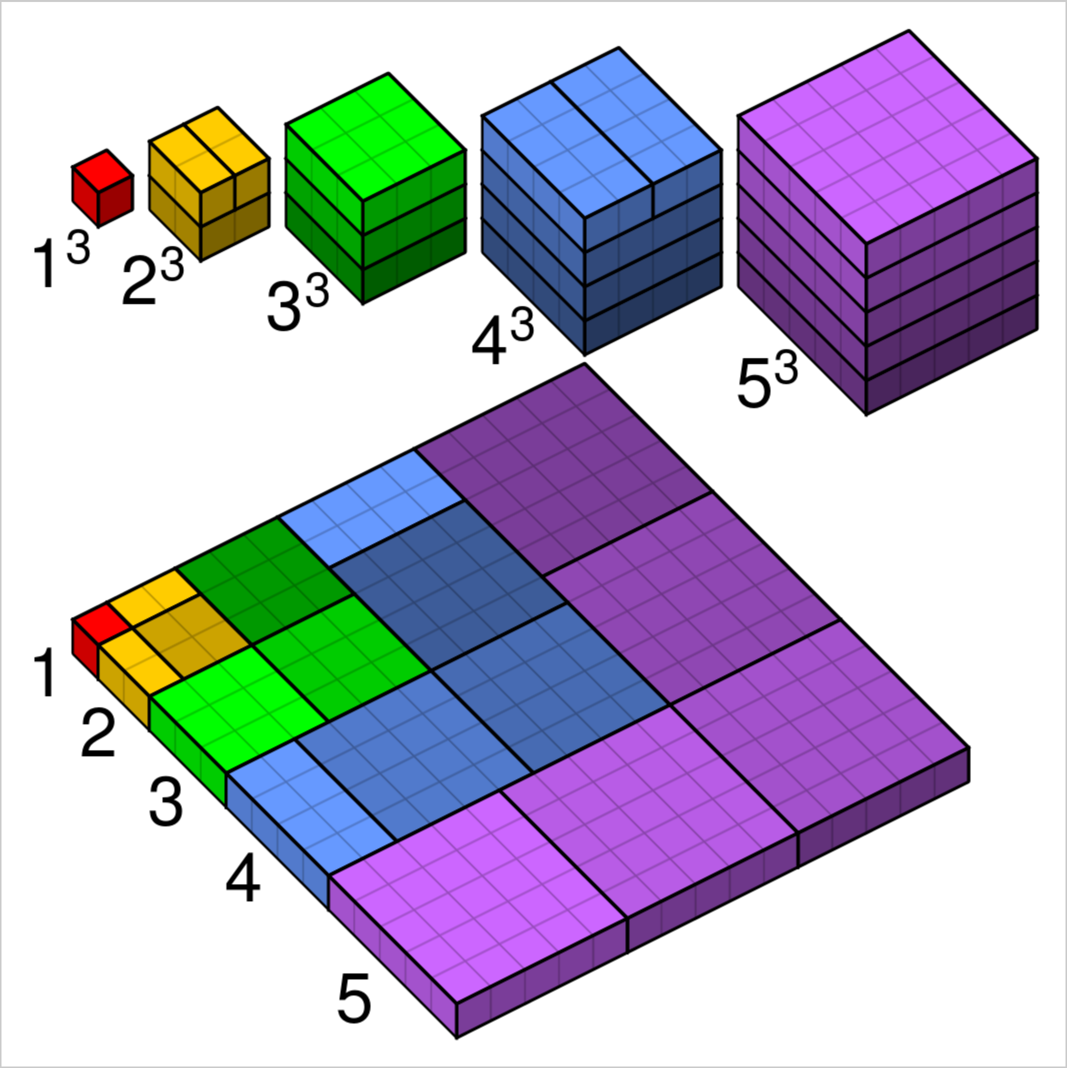
KNOWLEDGE VISUALIZATION (INTERACTIVE)

Crossfilter Demo (ES2015 + D3v4)

EDUCATIONAL VISUALIZATION

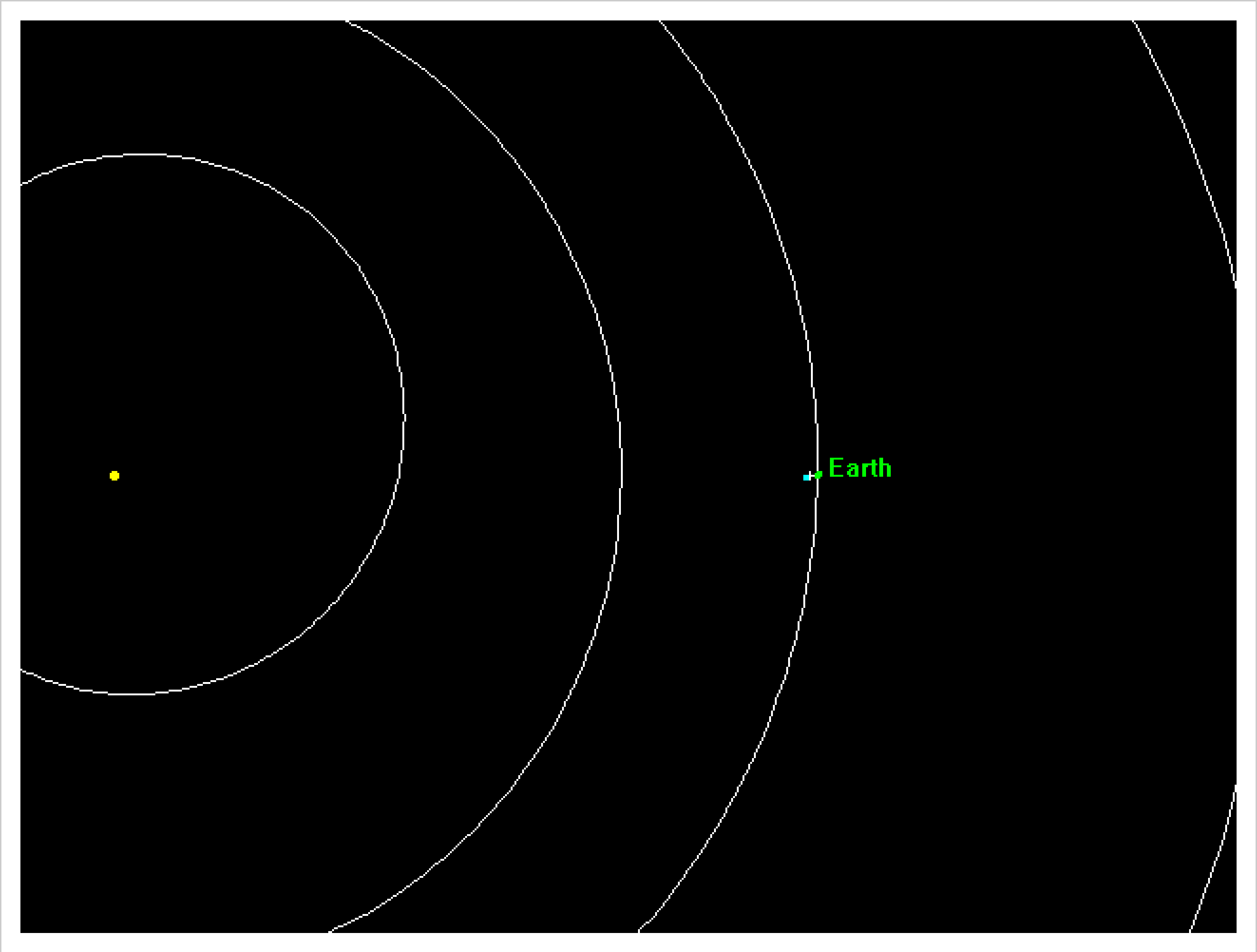
Educational Visualization is using a simulation, usually created on a computer, to create an image of something so it can be taught about (ex. atomic structure).

EDUCATIONAL VISUALIZATION (STATIC)



Proof without words of the Nicomachus theorem (squared triangular numbers).

EDUCATIONAL VISUALIZATION (DYNAMIC)



The motion of J002E3, showing how the object was captured into its chaotic orbit around the Earth by passing near the L1 point, looping around the Earth for 6 orbits, and then leaving Earth's orbit. The Sun is to the left in these animations.

EDUCATIONAL VISUALIZATION (INTERACTIVE)

Over the Decades, How States Have Shifted

Four Ways to Slice Obama's 2013 Budget Proposal

Political Influence

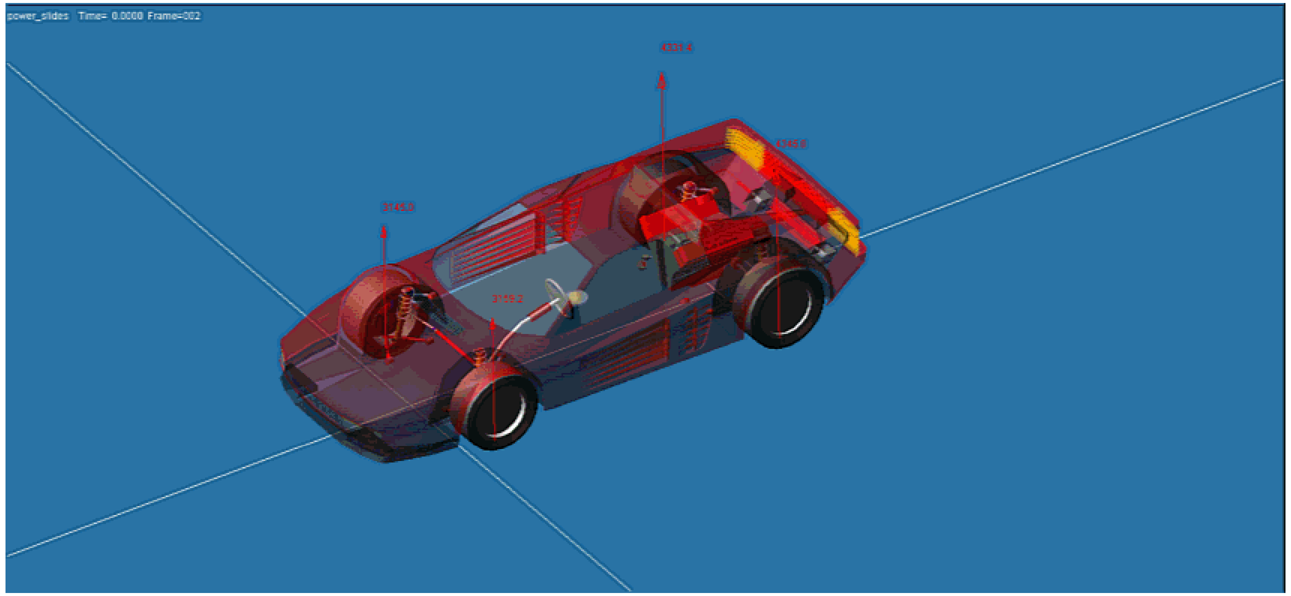
PRODUCT VISUALIZATION

Product Visualization is the viewing and manipulation of 3D models, technical drawing and other related documentation of manufactured components and large assemblies of products.

PRODUCT VISUALIZATION (STATIC)

2D rendering of a 3D CAD model of a car design in Sketchup.

PRODUCT VISUALIZATION (DYNAMIC)



3D electric motor.

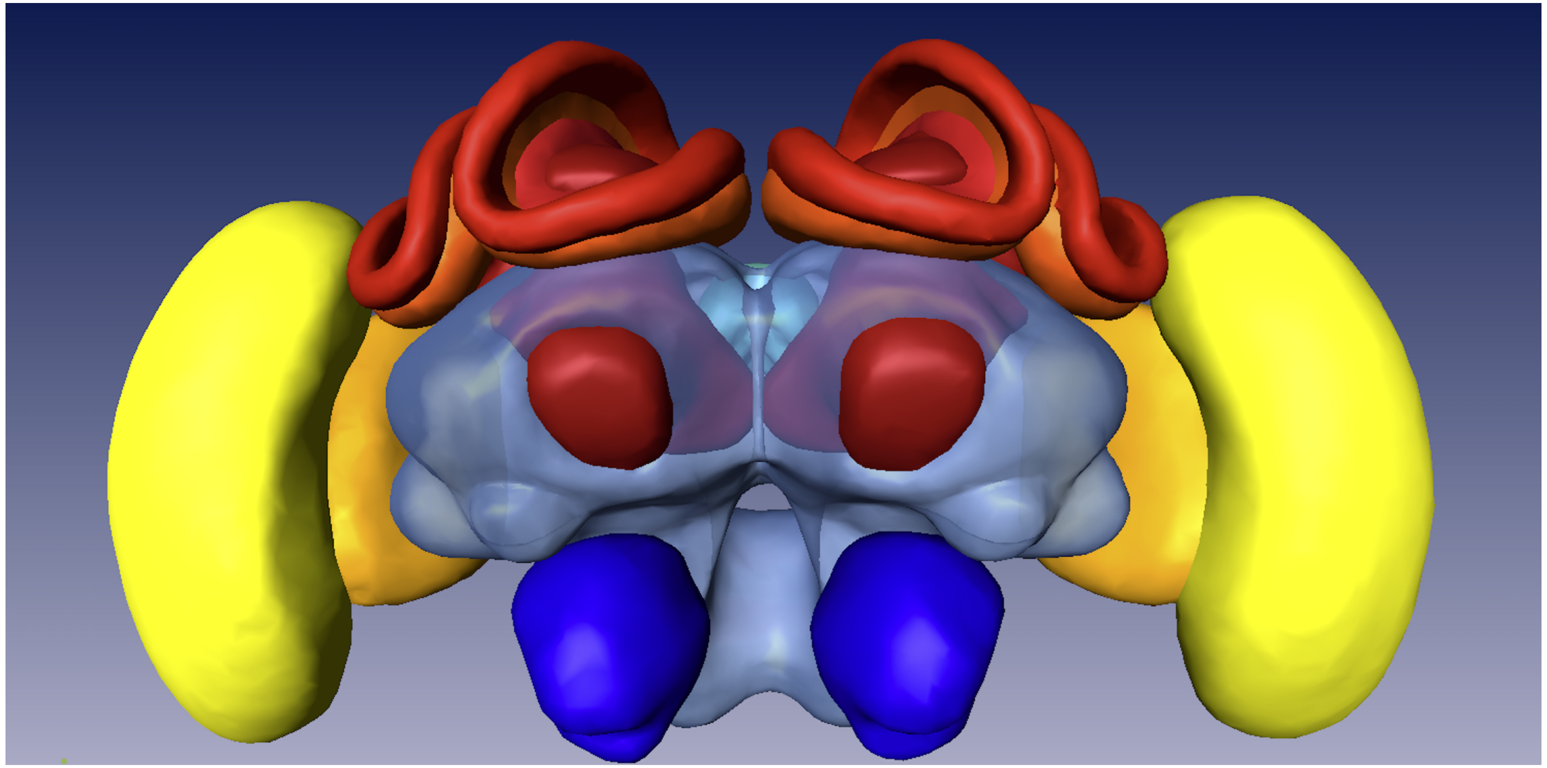
PRODUCT VISUALIZATION (INTERACTIVE)

Car Visualizer

SCIENTIFIC VISUALIZATION

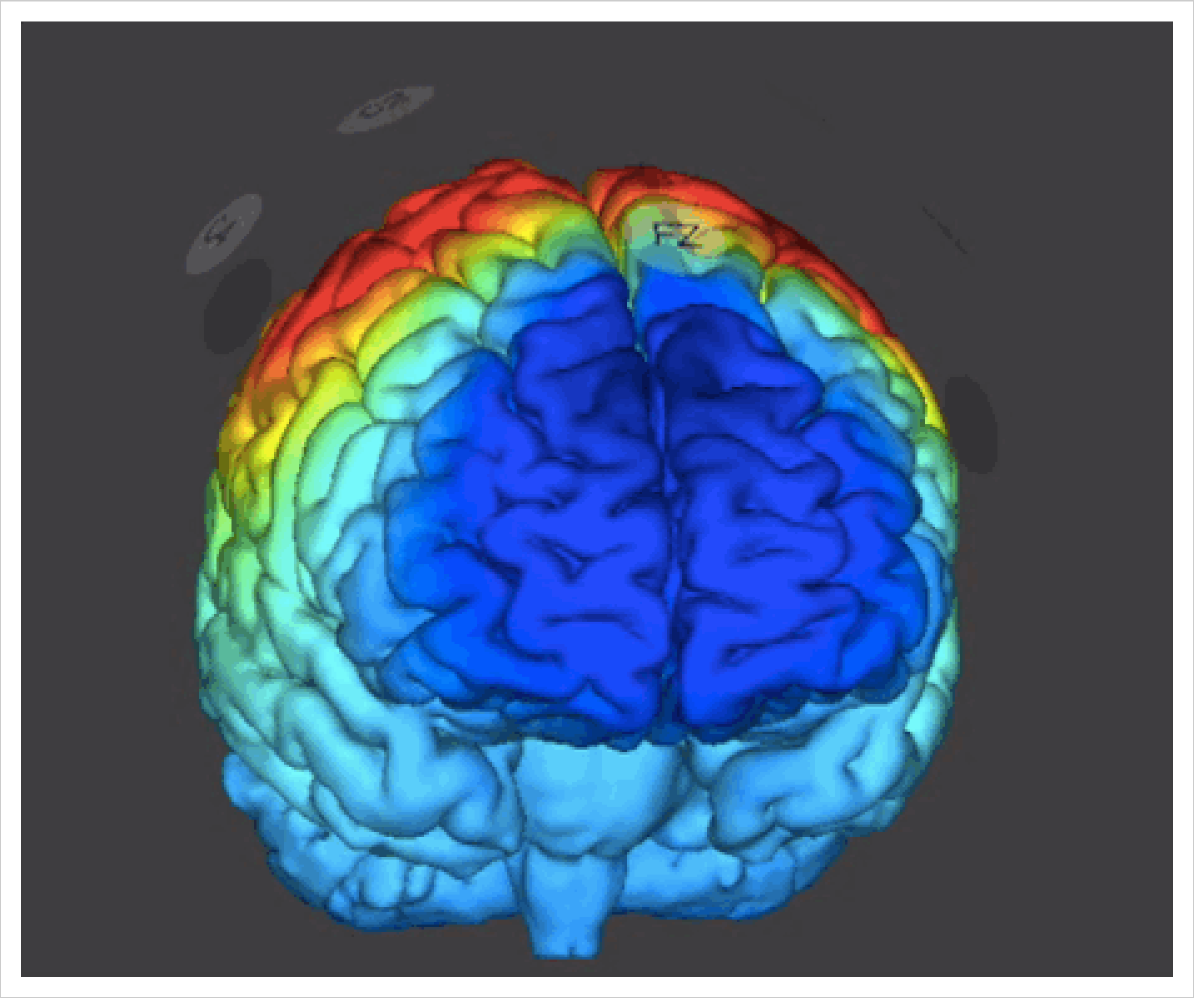
Scientific Visualization is primarily concerned with the visualization of three-dimensional phenomena (architectural, meteorological, medical, biological, etc.), where the emphasis is on realistic renderings of volumes, surfaces, illumination sources, and so forth, perhaps with a dynamic (time) component. The purpose of scientific visualization is to graphically illustrate scientific data to enable scientists to understand, illustrate, and glean insight from their data.

SCIENTIFIC VISUALIZATION (STATIC)



3D model of a honey bee's brain.

SCIENTIFIC VISUALIZATION (DYNAMIC)



3D model of a human brain.

SCIENTIFIC VISUALIZATION (INTERACIVE)

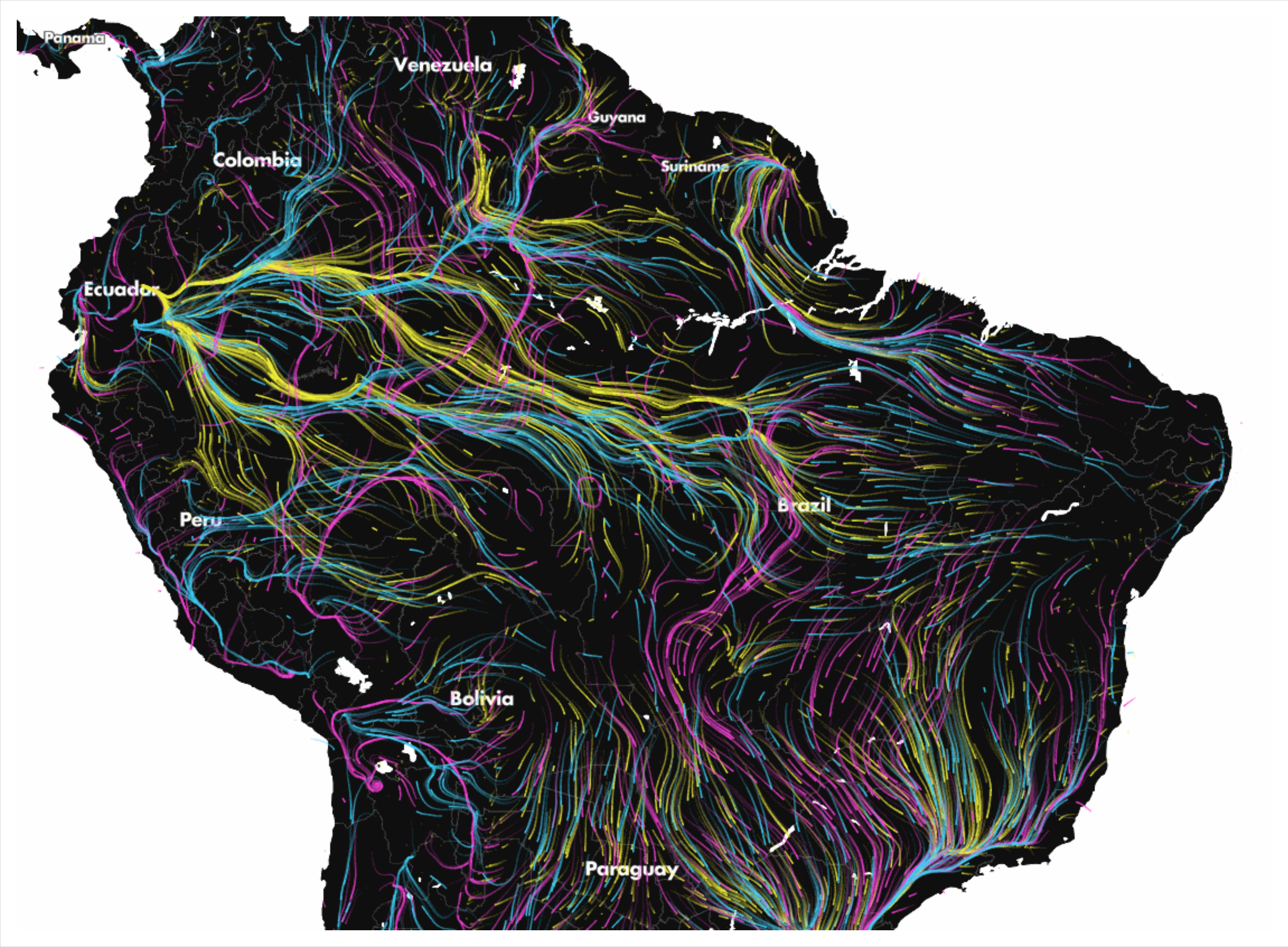
GEOGRAPHIC VISUALIZATION

Geographic Visualization (a.k.a. *Geovisualization*) refers to a set of tools and techniques supporting the analysis of geospatial data through the use of interactive visualization. Geovisualization emphasizes knowledge construction over knowledge storage or information transmission by communicating geospatial information in ways that (when combined with human understanding) allow for data exploration and decision-making processes.

GEOGRAPHIC VISUALIZATION (STATIC)

A map of U.S. river flowlines based on the NHDPlusV2 geo-spatial, hydrologic framework dataset built by the US EPA Office of Water, assisted by the US Geological Survey.

GEOGRAPHIC VISUALIZATION (DYNAMIC)



GEOGRAPHIC VISUALIZATION (INTERACTIVE)

US Wind Patterns

Plotting Earthquakes with D3.js + Leaflet

Small Arms & Ammunition - Imports & Exports

GEOGRAPHIC VISUALIZATION CHALLENGES

The proper conformance between various projections in geospatial data is an endless dance of transformation.

Projections

EXPLORE GEOGRAPHIC VISUALIZATION

These links are to live CodePen samples that you can experiment with.

D3

- Simple Map
- Bar Graph
- Chord Diagram

LEAFLET

- Flight Path
- Leaflet + Google Maps
- Satellite Imagery
- Geometric Functions
- Custom Scale
- Dynamic Markers with Clustering

GEOGRAPHIC VISUALIZATION EXERCISES

LET'S MAKE A MAP! (2012)

This tutorial will cover how to make a modest map from scratch using D3 and TopoJSON. It will also show you a few places where you can find free geographic data online, and how to convert it into a format that is both efficient and convenient for display.

Article | Demo (CodePen)

GEOSPATIAL CODING TUTORIALS (2016)

These tutorials will get you started with Geospatial concepts and Python or R.

Raster Processing with Python Tools | Intro to PyQGis (Python + QGis) | Intro to GIS and Spatial Analysis (R)

COMMAND LINE CARTOGRAPHY (2017)

This multipart tutorial will teach you to make a thematic map from the command line using d3-geo, TopoJSON and ndjson-cli—free, open-source tools written in JavaScript. You will make a choropleth of California's population density. (For added challenge, substitute your state of choice!)

Part 1 | Part 2 | Part 3

GENERAL WORKFLOWS

1. Plan your visualization
2. Acquire the data from the sources
3. Validate/Cleanse/Transform the data as is required
4. Implement the visualization design
5. Load the data into the visualization
6. Adjust the visualization until it meets your criteria
7. Render and publish your visualization (or deploy if interactive)

GENERAL BEST PRACTICES

1. Know your Purpose
    - Be clear on your objective. Consider whether you are you trying to educate your audience, influence the audience or encourage your audience to take some action.
2. Know your Data
    - Start with basic visualizations and explore the dimensions of the data.
3. Data Ethics
    - Ensure you are using valid, current data with the proper permissions.
4. Visual Design
    - Choose the correct kind of visual to communicate your intent with the data.
5. Pay Close Attention to Details
    - Ensure that there are no mislabeled or misspelled items, that everything behaves correctly, etc. Attention to detail matters when decisions are being made based on the date presentation.

VISUALIZATION LIBRARIES & PACKAGES

PYTHON

- matplotlib
- seaborn
- ggplot
- bokeh
- plotly
- vispy

R PACKAGES

- dplyr (Rdocs)
- foreign (Rdocs)
- cluster (Rdocs)
- ggplot2 (Rdocs)
- ggplot2

VISUALIZATION APPLICATIONS & WEB LIBRARIES

FRAMEWORKS, APPLICATIONS & PLATFORMS

- Processing
- Gephi
- Trifecta Wrangler
- Tableau

JAVASCRIPT LIBRARIES

- D3
- Vega (Visualization Grammar)
- DyGraphs
- Raphael
- Google Charts
- ChartsJS
- RawGraphs
- ThreeJS
- WebVR
- Timeline

GEOSPATIAL APPLICATIONS & FRAMEWORKS

GEOSPATIAL SUITES & APPLICATIONS

- OpenGeo Suite
- QGis
- GDAL
- PDAL
- GeoKettle
- TileMill

MAP SERVERS & TILE SERVERS

- GeoServer
- Mapserver
- TileServerGL (FOSS)

GEOSPATIAL WEB RESOURCES

WEB MAPPING LIBRARIES

- Leaflet
- OpenLayers
- Polymaps
- DataMaps
- amCharts
- WebGL Globe

WEB MAPPING APIS

- Google Map API
- Esri Web API

WEB MAPPING PLATFORMS & TILE SERVICES

- Mapbox
- Arcgis.com
- OpenStreetMap
- Google Maps
- TileServer (Commercial)

DATA SOURCES

Government

Data.gov

US Census Bereau

US FDA

NCES

FBI: Uniform Crime Reporting

World

CIA World Factbook

International Monetary Fund

Unicef

United Nations ODC

WHO

Drug War Facts

MORE DATA SOURCES

Academic & Scientific

PEW Iinternet Project

Open Science Data Cloud

Healthdata.gov

NOAA National Centers for Environmental Information

National Weather Service

NASA Planetary Data Service

NASA Earth Data

Social Media

Google Trends

Facebook Graph API

Twitter API

Instagram API

Foursquare API

EVEN MORE DATA SOURCES

Journalism & Media

NY Times API

Associated Press API

AWS Public Datasets

Wikipedia Database

five-thirty-eight (github)

Other

Subreddit Datasets

AWS Million Song Dataset

IMDB

Generatedata.com

Gapminder

RESOURCES

Print vs. Web, Static vs. Interactive

Minard's graphic of Napolean in Russia

Analyzing Minard's Visualization Of Napoleon's 1812 March

Data, Information and Knowledge in Visualization

Stacked to Grouped Bars Chart (M. Bostock)

MERRA Image (NASA)

jdem846 Utility

Kevin Gill - Geovisualizations

Maptools.org

RESOURCES (CONTINUED)

Information is Beautiful

Learning from Architects: The Difference between Knowledge Visualization and Information Visualization

JPL Infographics (NASA)

Stream Graph (Mike Bostock)

Zoomable Tree Map (Mike Bostock)

Node Graph (Mike Bostock)

Sunburst Graph (Mike Bostock)

Projections

D3 Blocks (Mike Bostock)

US Rivers (Mike Bostock)

Map of American Rivers: A vector tile demonstration and tutorial (Nelson Minar)

NHDPlus Version 2 Data

Plotting Earthquakes with D3.js + Leaflet

US Wind Patterns

ATTRIBUTIONS

File: Social Network Analysis Visualization.png | **Source:** Own work | **Author:** Martin Grandjean | **Date:** 2 November 2013, 22:02:28 | **Copyright:** Creative Commons Attribution-Share Alike 3.0 Unported | link

File: Minard.png | **Source:** see upload log | **Author:** Charles Minard (1781-1870) | **Date:** November 20, 1869 | **Copyright:** Creative Commons Attribution-Share Alike 3.0 Unported | link

File: chemical-reaction.gif | **Source:** TACC | **Author:** TACC | **Date:** 09/29/2016 | **Copyright:** 3-Clause BSD License | link

File: LastGraph example.svg | **Source:** Own work | **Author:** Psychonaut | **Date:** 24 July 2012 | **Copyright:** Creative Commons CC0 1.0 Universal Public Domain Dedication | link

File: DiffusionMRI glyphs.png | **Source:** Medical imaging software | **Author:** Tucanis | **Date:** 2012-08-22 | **Copyright:** Creative Commons Attribution-ShareAlike 3.0 License | link

File: Electromagnetic Radiation Spectrum Infographic.svg | **Source:** Own work | **Author:** Duarte Farrajota Ramos | **Date:** 9 February 2014 | **Copyright:** Creative Commons Attribution-Share Alike 3.0 Unported | link

File: Astro 4D m45 cr anim.gif | **Source:** Own work | **Author:** Alexander Meleg | **Date:** 13 March 2010 | **Copyright:** Creative Commons Attribution-Share Alike 4.0 International | link

File: Wikipedian in Residence Infographic.jpg | **Source:** The Children's Museum of Indianapolis | **Author:** The Children's Museum of Indianapolis | **Date:** July 2011 | **Copyright:** Creative Commons Attribution-Share Alike 3.0 Unported | link

File: U.S. Rivers | **Source:** Own work | **Author:** Mike Bostock | **Date:** May 17, 2013 | **Copyright:** Copyright (c) 2013-2017, Michael Bostock All rights reserved. | link

File: Future ozone layer concentrations.gif | **Source:** NASA | **Author:** NASA/Goddard Space Flight Center Scientific Visualization Studio | **Date:** 21 May 2009 | **Copyright:** Public Domain | link

THANK YOU

PRESENTATION SOURCE

The presentation requires NodeJS and Bower be installed on your system.

NodeJS: https://nodejs.org/en/

Bower.io (requires node):

```
npm install -g bower
```

Once these are installed...

Download the presentation here: https://github.com/jgentle/data-visualization

Or clone it locally:

```
git clone https://github.com/jgentle/data-visualization.git
```

Then...

```
cd data-visualization
npm install
bower install
grunt serve
```

Enjoy the presentation!

PRINTING THIS PRESENTATION

Printing functionality requires the presentation to be running on a web server (either locally or online).

You can then print this presentation by appending `?print-pdf` to the end of the URL:

```
# localhost.
http://localhost:9000/?pdf-print/

# live online.
http://mydomain.com/my-presentation/?pdf-print/
```

Follow it then enter the standard CTRL/CMD+P (or use browser menu > Print...).

The content will look weird until you print preview... but it will print OKAY!

For convenience I have included the localhost link below.

Print Ready Version