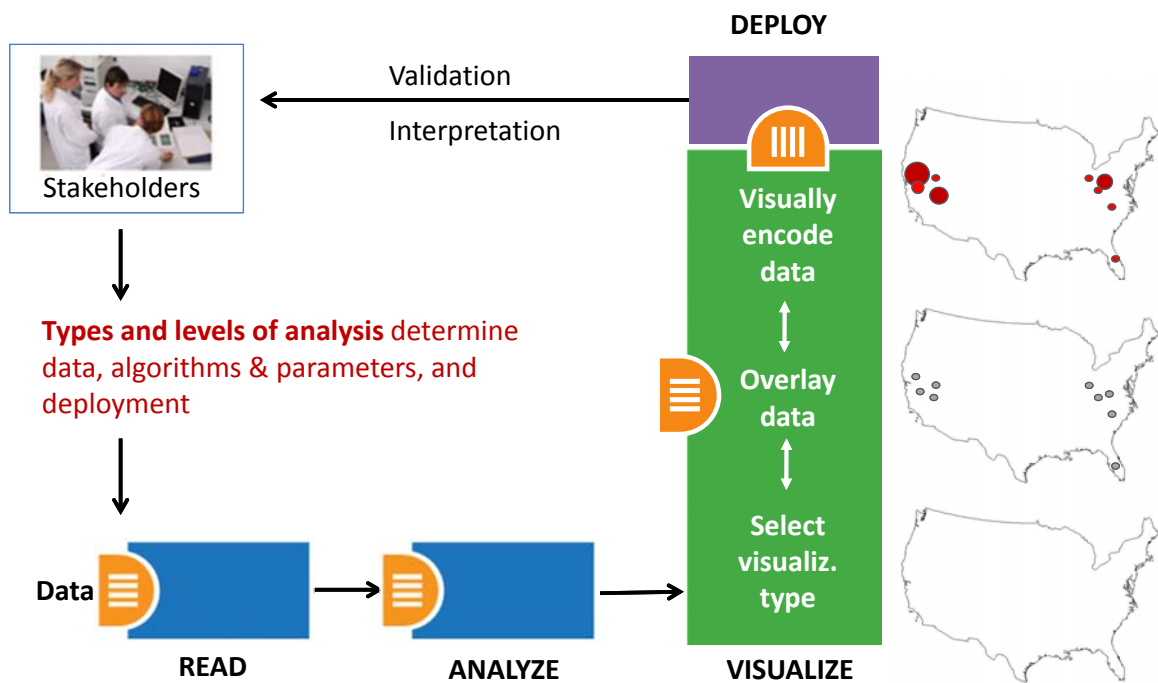


Needs-Driven Workflow Design



26

Information Visualization MOOC

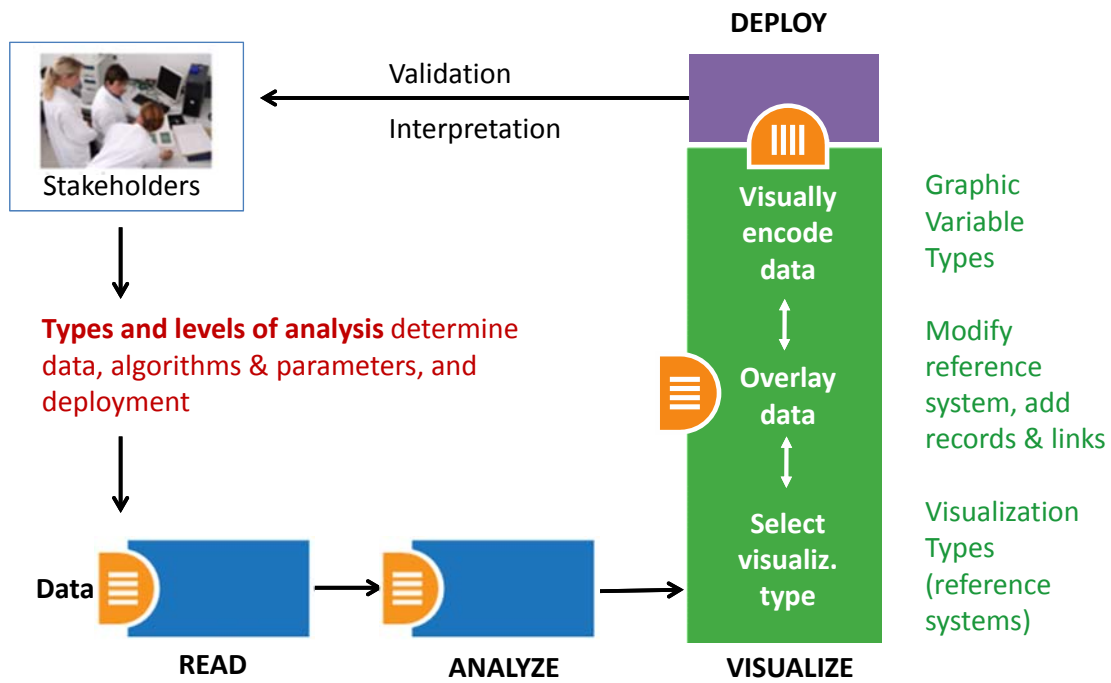
Unit 3 – “Where”: Geospatial Data

Workflow Design

Relevant Research Disciplines:

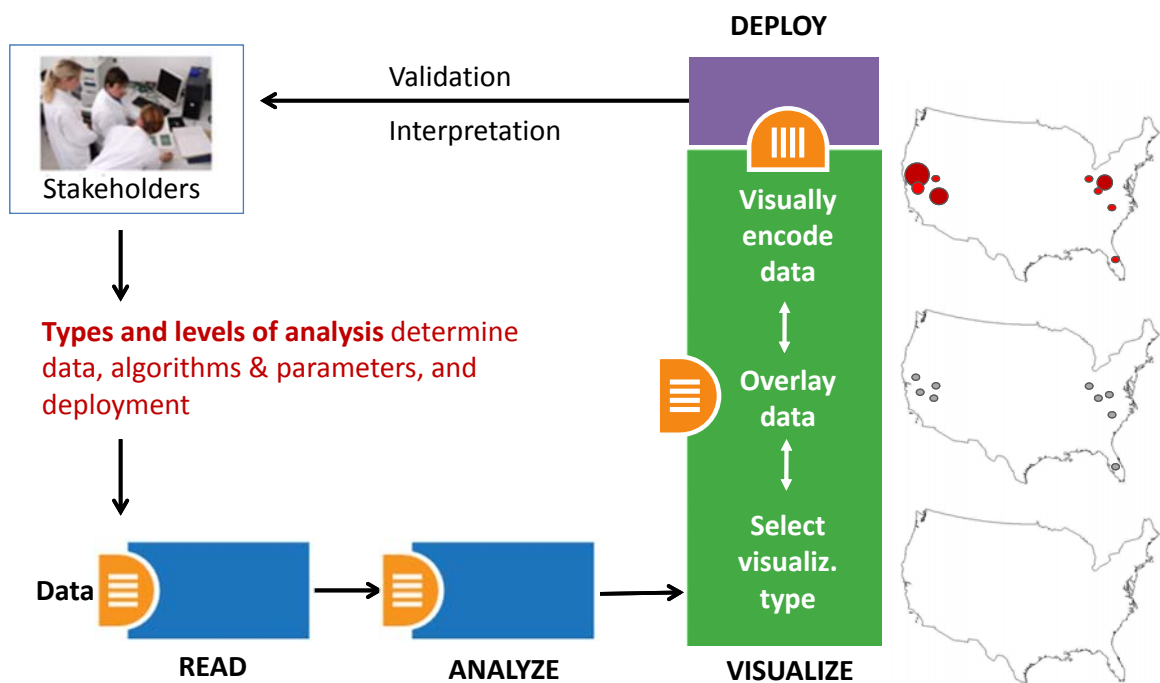
Geography, Cartography, Statistics, Information Visualization

Needs-Driven Workflow Design



28

Needs-Driven Workflow Design



29

Read Data

Data Repositories:

- Free GIS Data, <http://freegisdata.rtwilson.com/>
- KDD Datasets, <http://www.kdnuggets.com/datasets/>
- Digging into Data list of repositories, <http://www.diggingintodata.org/Repositories/tabid/167/Default.aspx>
- Scholarly Database, <http://sdb.cns.iu.edu>

Data Formats:

- Vector—e.g., Shape files, PS
- Raster—e.g., TIFF, JPEG
- Tabular—e.g., in CSV
- Software specific

30

Preprocess Data

- Thresholding
- Unification, see Gazetters
- Aggregation, see Aggregation

Example: “Mapping the Diffusion of Information among Major U.S. Research Institutions”

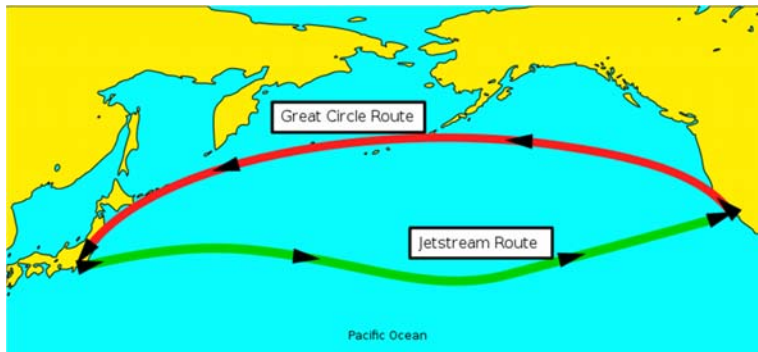
Need a compromise between maintaining geographic identity and statistical significance. For example, should IU be represented as one or eight campuses? IUB has two ZIP codes—represent IUB by one or two places?

Reference: Börner, Katy, Shashikant Penumathy, Mark Meiss, and Weimao Ke. 2006. [“Mapping the Diffusion of Information among Major U.S. Research Institutions.”](#) *Scientometrics* 68 (3): 415-426.

31

Geospatial Analysis - Linkages

- Longest link = $\frac{1}{2}$ Earth's circumference
- *Earth's Circumference* at the Equator: 24,901.55 miles (40,075.16 km).
- *Earth's Circumference* between the North and South Poles: 24,859.82 miles (40,008 km)



Airline routes between San Francisco and Tokyo: great circle (red) and following the jet stream (green) when heading eastwards.

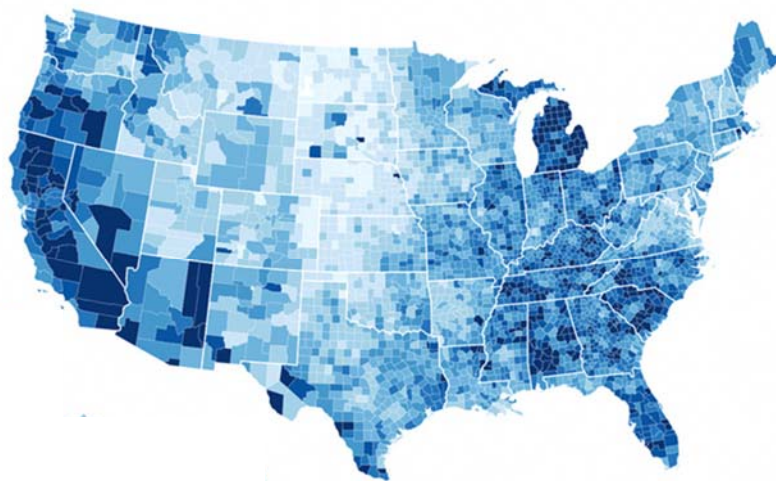
32

Visualizations

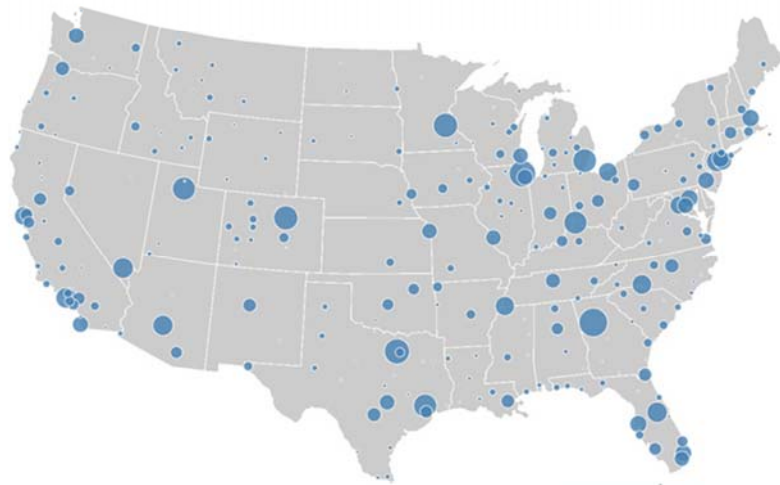
Subsequent slides show a map of the US with

1. Choropleth map of unemployment data in 2008
2. Proportional symbol overlay of airport traffic data
3. Linkage overlays of airport traffic data

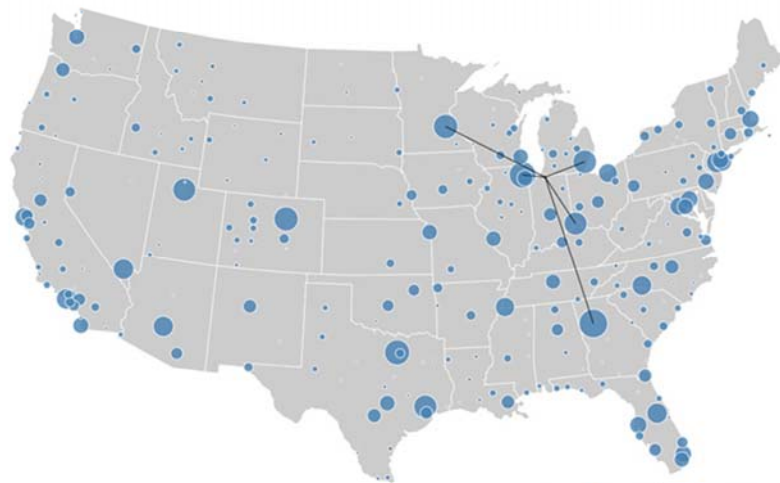
33



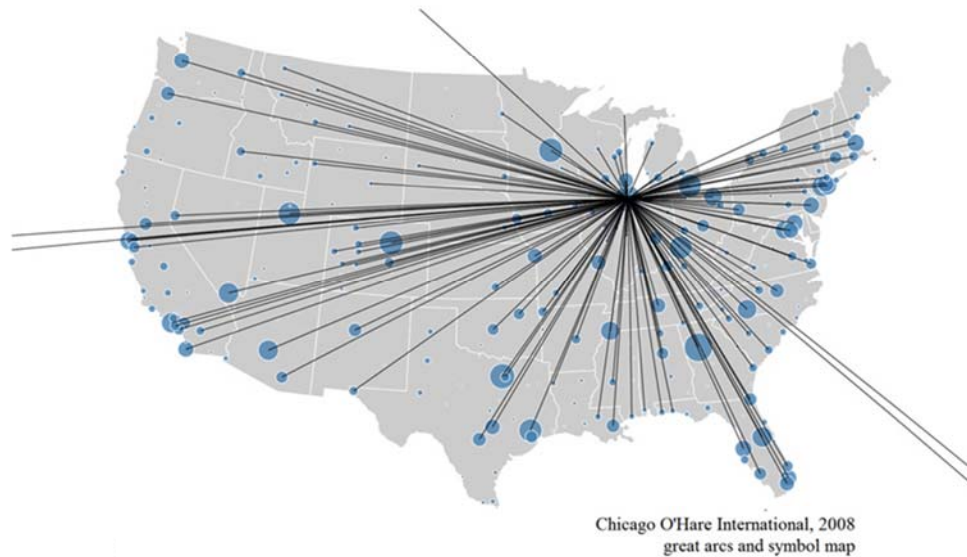
U.S. Unemployment, 2008



Tucson International, 2008
great arcs and symbol map



South Bend Regional, 2008
great arcs and symbol map



Relevant Tools

- List of GISsoftware, also open software, http://en.wikipedia.org/wiki/List_of_geographic_information_systems_software
- Data-Driven Documents (D3), <http://d3js.org>
- Tableau, <http://tableausoftware.com>
- IBM's Many Eyes, <http://www-958.ibm.com/software/data/cognos/manyeyes>



Please post your favorite to Twitter, Flickr using tags “ivmooc” and “#geotools.”