Power in Numbers: Making Data Work for You
Association for Rural and Small Libraries: 2015 Conference
October 2, 2015
Little Rock, AR

Deanne W. Swan, Ph.D.
Sr. Statistician
Office of Planning, Research & Evaluation
Institute of Museum & Library Services
dswan@imls.gov

Timothy Owens
Sr. Library Program Officer
Office of Library Services
Institute of Museum & Library Services
towens@imls.gov

Data analysis is a **process** of inspecting, cleaning, transforming, and modeling data with the goal of uncovering information, supporting decision making, and *telling stories*.

Human Perception as the Foundation for Effective Visualizations

- Data Visualization for Human Perception http://www.interaction-design.org/encyclopedia/data_visualization_for_human_perception.html
- Visual encoding (data types, color, shape, value) http://www.targetprocess.com/articles/visual-encoding.html
- Process for creating data visualization http://blog.visual.ly/the-process-of-creating-data-visualizations/

Principles for Effective Graphical Displays

- ACCENT
  - **A**pprehension: Can you perceive the relationship among the variables?
  - **C**larity: Can you visually distinguish all of the elements of the graph?
  - **C**onsistency: In a series of graphs, are the meanings of elements, shapes and symbols the same?
  - **E**fficiency: Do you portray a complex relationship in the simplest way possible?
  - **N**ecessity: Do you need the graph, or will another format (text, table) be more effective?
  - **T**ruthfulness: Are the elements accurately positioned and scaled?

From D. A. Burn (1993). Designing Effective Statistical Graphs, in Rao (Ed.) Handbook of Statistics.

Graphic Design Principles

- Sort data before graphing
- Make effective use of graph titles
- Maintain appropriate aspect ratios
- Limit the use of pie charts
- Avoid fancy formatting
  - Avoid use of colors in the background or plot area. Reserve color for key data points in the graph.
  - Avoid 3D effects.
  - Avoid fancy visual effects, such as gradients, shadows, patterned fill, and soft edges.
  - Avoid the tendency to enhance the graph with clip art or pictures.
- Eschew Chartjunk
  - Remove gridlines and borders.
  - Avoid overloading the graph with data labels. Focus only on key data points.
  - Use legends only when necessary. Sometimes it is apparent.
  - Remove axes that do not add information.

Information Design

- Information Design for Advocacy http://www.opensocietyfoundations.org/sites/default/files/visualizing_20080311.pdf
- Edward Tufte: http://www.edwardtufte.com/
- Michael Friendly: http://www.datavis.ca/

Inspiration

- http://flowingdata.com/
- http://eagereyes.org/
- http://thewhyaxis.info/
- http://visualisingdata.com/
- http://visual.ly/

Types of Visualizations

- Classification Charts http://charts6.excelcharts.com/blog/wp-content/uploads/2013/11/classification-chart-types.png
- Common Visualization Types http://guides.library.duke.edu/vis_types

Selecting a Graph Type

- Juice Analytics Chart Choose http://labs.juiceanalytics.com/chartchooser/index.html
- Choose a Good Chart http://extremepresentation.typepad.com/files/choosing-a-good-chart-09.pdf
- Effective Chart Design http://www.perceptualedge.com/images/Effective_Chart_Design.pdf

Popular Tools

- Cleaning data
  - Open Refine http://openrefine.org/
  - Data Wrangler http://vis.stanford.edu/wrangler/
  - Mr. Data Converter http://shancarter.github.io/mr-data-converter/
- Statistical Analysis
  - R and (RStudio) http://www.rstudio.com/
  - Python (scipy & numpy & pandas) http://datacommunitydc.org/blog/2013/07/python-for-data-analysis-the-landscape-of-tutorials/
  - StatWing https://www.statwing.com/
- Geospatial tools
  - QGIS (QuantumGIS) http://qgis.org/en/site/
  - TileMile https://www.mapbox.com/tilemill/
  - CartoDB http://cartodb.com/
- Visual analytics/Business Intelligence
  - Tableau Public http://www.tableausoftware.com/public/
  - Google Fusion Tables http://www.google.com/drive/apps.html#fusiontables
  - ManyEyes http://www-958.ibm.com/software/data/cognos/manyeyes/
- Excel
  - Jorge Camoes's Excel Charts Blog http://www.excelcharts.com/blog/
  - Learn to use:
    - Power Pivot http://www.microsoft.com/en-us/bi/powerpivot.aspx
    - Power View http://www.microsoft.com/en-us/bi/Products/PowerView.aspx
    - Analysis Toolpak http://www.add-ins.com/Analysis_ToolPak.htm
  - Python + Excel = DataNitro https://datanitro.com/

Useful Software for Visualization

- Selected Tools from Datavisualization.ch http://selection.datavisualization.ch/
- 86 Helpful Tools for the Data Professional http://infospace.ischool.syr.edu/2011/10/19/86-helpful-tools-for-the-data-professional-plus-45-bonus-tools/
- 22 free tools for data visualization and analysis http://www.computerworld.com/s/article/9215504/22_free_tools_for_data_visualization_and_analysis

Mapping, GIS & Geospatial analysis

- A Gentle Introduction to GIS http://download.osgeo.org/qgis/doc/manual/qgis-1.0.0_a-gentle-gis-introduction_en.pdf
- QGIS basics for journalists http://multimedia.journalism.berkeley.edu/tutorials/qgis-basics-journalists/

Resources

- IMLS
    - Statistical Collection Data Files https://www.imls.gov/research-tools/data-collection
    - Data Portal https://data.imls.gov/
    - Publications https://www.imls.gov/publications
- IMLS Funded Projects
    - iMap Libraries http://www.imaplibraries.org/index.html
    - Digital Inclusion Study Tool http://digitalinclusion.umd.edu/content/interactive-map

Places to get data

- Government
    - FedStats.gov http://www.fedstats.gov/
    - TIGER http://www.census.gov/geo/www/tiger/
    - American Fact Finder http://factfinder2.census.gov/faces/nav/jsf/pages/index.xhtml
    - Data portals and data catalogs
        - Data.gov & others http://www.data.gov/opendatasites
- Universities, research, and institutional repositories
    - FigShare http://figshare.com/
    - DataCite http://www.datacite.org/repolist
    - Cornell's Data Sources for Social Scientists http://ciser.cornell.edu/info/datasource.shtml
- Library specific data
    - USAC e-rate data http://www.slforms.universalservice.org/DRT/Default.aspx
    - NTIA National Broadband Data http://www.broadbandmap.gov/data-download

Recommended Books:

- Tufte, E. (2001). *The visual display of quantitative information.* Cheshire, Conn: Graphics Press.
- Cleveland, W. (1993). *Visualizing data.* Murray Hill, N.J. Summit, N.J: At & T Bell Laboratories Published by Hobart Press.
- Few, S. (2012). *Show me the numbers : designing tables and graphs to enlighten.* Burlingame, Calif: Analytics Press.
- Yau, N. (2011). *Visualize this : the FlowingData guide to design, visualization, and statistics.* Indianapolis, Ind: Wiley Pub.
- Wilkinson, L. & Wills, G. (2005). *The grammar of graphics*. New York: Springer.