The Foundations of Data Visualization in Practice

Tuesday, June 9, 2015
8:30 AM – 4:00 PM

Training Pre-requisites
Skills: Familiarity with MS Office & Tableau Software
Tools: Laptop, wired mouse, Microsoft Office 2010 (or later versions)

Workshop Topics

1. Introductions (8:30 – 8:45)

2. Foundational Basics (8:45 – 9:45)
   a. Historical Examples:
      i. Dr. John Snow’s plot of the cholera epidemic in London
      ii. Florence Nightingale’s coxcomb charts on death toll in hospitals
      iii. Charles Joseph Minard’s thematic map of the fate of Napoleon’s army in Russia
   b. Tufte Design Principles
      i. Graphical Excellence
         1. Is that which gives the viewer the greatest number of ideas in the shortest time with the least ink in the smallest space
         2. Graphical excellence is nearly always multivariate
         3. Mobilize every graphical element, perhaps several times over, to show the data
         4. Data density and small multiples – for non-data ink, less is more. For data ink, less is a bore.
         5. And graphical excellence requires telling the truth about data.
         6. Graphical elegance is often found in simplicity of design and complexity of data.
      ii. Graphical Integrity
         1. Representation of Data should be directly proportional to the numerical quantities represented.
         2. Use clear, detailed and thorough labeling.
         3. Show data variation not design variation
         4. Graphic must not quote data out of context
            Lie Factor = (size effect show in graphic) / (size effect of data)
      iii. Principles of Data Graphics
         1. Above all else show data
         2. Maximize the data-ink ratio
         3. Erase non-data-ink.
         4. Erase redundant data-ink.
         5. Revise and edit.
   c. Stages in Data Visualization
      i. Acquire - data collection
      ii. Parse - categorize and structure data
iii. Filter - prevent data overload
iv. Mine - draw connections between data and overall goal
v. Represent - choose visual model type
vi. Refine - improve clarity and data-ink ratio
vii. Interact - allow methods of manipulation; allow users to control what/how they perceive the data

Break (9:45 – 10:00)

3. Visualization Methods (10:00 – 10:30)
   a. Relationships between data points
      i. Network diagram
      ii. Scatterplot
      iii. Matrix chart
   b. Compare values
      i. Bar/column charts
      ii. Block histograms
      iii. Bubble chart
   c. Track values over time
      i. Bar/column charts
      ii. Line graph
      iii. Stack graph
   d. Static parts of a whole
      i. Pie/ring chart
      ii. Stacked 100% bar/column charts
      iii. Treemap
   e. Brief exercise for matching data sets with visualization methods

   a. Exploration and Explanation
   b. Avoiding mistakes and pitfalls
      i. When to use tables and when to use graphs
      ii. Avoid chart junk – often computer generated.
      iii. Opt art side effects
   c. Don’t over-do it!
      i. Occam’s Razor – “What can be done with fewer is done in vain with more”
      ii. Ward Cunningham – “What is the simplest thing that could possibly work; it should be that and nothing simpler.”
      iii. Make all visual distinctions as subtle as possible, but still clear and effective. – Tufte
   d. Formatting
      i. Resolution, typography, color
      ii. Data density
      iii. Export formats for production specs

Lunch 11:15AM – 12:30PM
5. Data Sources (12:30-1:00PM)
   a. BLS / Workforce Information Database
   b. Census (American Fact Finder)
   c. BEA

6. Data cleansing, rounding, making an index, analysis, confidentiality (1:00-1:30PM)
   a. Example 1: Import WID LABFORCE table into Tableau an unemployment plot
   b. Example 2: Another example in Highcharts

Break (1:30PM-1:45PM)

7. Overview of Some Data Visualization Tools - (1:45-2:45PM)
   a. MS Office
      i. Go over examples in Excel and do a work through of an example.
   b. Tableau
      i. Go over examples (Will and John), then show how uploaded to blog.
   c. Highcharts
      i. Go over examples (John w/ JSFiddle), then show how to upload to blog.
   d. Infographics
      i. Present two examples and method of production

8. Practice Data Visualizations: Hands On (2:45-3:45PM)
   a. Practice Example 1
   b. Practice Example 2

9. Open Discussion Session (3:45-4:00PM)
   a. Final Evaluation