Guidelines on the Use of Graphs

Christopher G. Healey
Institute for Advanced Analytics
Department of Computer Science
North Carolina State University
healey@ncsu.edu
http://www.csc.ncsu.edu/faculty/healey
Charts and Graphs

• Well known, well recognized presentation technique

• Basic types of charts and graphs are recognized by most viewers
  ✦ line graph, bar graph usually recognized
  ✦ pie chart, histogram, scatterplot, Venn less common

• Different graphs are designed for different types of data and analysis tasks
  ✦ import to choose the “correct” graph for the data and task

• Graphs are meant to provide insight to your viewers

• Understand your audience and their needs
Graph Types

• Different graphs are used for different tasks:
  ✦ tracking changes (line graph)
  ✦ correlations, relationships (X-Y plot)
  ✦ correlations, clustering (scatterplot)
  ✦ category comparison (bar chart)
  ✦ distributions (pie chart)
  ✦ uniqueness and overlap (Venn diagram)

• Graphs are not meant for absolute value determination
  ✦ tables should be used if determining exact values is important

• Graph elements can represent multiple attribute values
Graph Selection

Independent variable quantitative?  
- Quantitative
  - One independent variable?
    - Yes  
      - Scalar
        - Independent variable ordinal?  
          - Ordinal
            - See local trends?
              - Yes  
                - One independent variable?
                  - No  
                    - multi-line graph
                  - Yes  
                    - See individual data points?
                      - No  
                        - grouped bar graph
                      - Yes  
                        - line graph, bar graph
            - No  
              - multi-line graph
          - No  
            - See individual data points?
              - Yes  
                - line graph, bar graph
              - No  
                - grouped bar graph
    - No  
      - See local trends?
        - Yes  
          - See individual data points?
            - No  
              - multi-line graph
            - Yes  
              - line graph, bar graph
      - No  
        - See individual data points?
          - No  
            - multi-line graph
          - Yes  
            - line graph, bar graph
    - Qualitative
      - One independent variable?
        - Yes  
          - See local trends?
            - Yes  
              - See individual data points?
                - No  
                  - line graph, bar graph
                - Yes  
                  - line graph, bar graph
            - No  
              - See individual data points?
                - No  
                  - multi-line graph
                - Yes  
                  - line graph, bar graph
        - No  
          - See local trends?
            - Yes  
              - See individual data points?
                - No  
                  - grouped bar graph
                - Yes  
                  - grouped bar graph
Define Target Audience

- Who is your target audience?
  - Defines the presentation type(s) they are familiar with

- What do they know about the data?
  - May want to include known data to define context
  - May want to leave out known data if its presentation is redundant

- What do they want to know about the data?
  - Defines which data to show, which presentation type(s) to use

- What do they expect to see?
  - Defines “uninteresting” results

- What will they do with the information?
  - e.g., exploring the data, validating known results, presenting results?
  - Defines which presentation type(s) to use
Define Message

- What do the data show?
  - Depends on analysis task(s) and result(s) found
  - Exploration, show new results
  - Validation, show that expected results are correct
  - Presentation, show important results in an easy-to-comprehend manner

- Is there more than one main message?
  - Ensure individual messages are clear and memorable
  - Do not use a single presentation for multiple messages unless necessary

- What aspects of the message should be highlighted?
  - Ensure main point(s) are obvious point in the presentation
Tufte’s PowerPoint Folio

• Basic guidelines on computer presentation

Edward R. Tufte

The Cognitive Style of PowerPoint: Pitching Out Corrupts Within

Military parade, Stalin Square, Budapest, April 4, 1936.
Graph Recommendations

- Pie chart
  - comparison of relative amounts, description of components
- Bar graph
  - comparison of items, relationships between items
  - time series data
- Line graph
  - time series and frequency distribution
- Scatterplot
  - analysis of relation
- Venn diagram
  - analysis of uniqueness and commonality (overlap)
Bar Graph

- Composed of discrete bars
- Represents categorical data for comparison across categories

Bar graph of ice, sleet, and snow in Knoxville, categorized by month
Bar Graph

- Composed of discrete bars
- Represents categorical data for comparison across categories

Grouped bar graph of US patent applications, patents issued, licenses signed, categorized by year
Bar Graph

- Composed of discrete bars
- Represents categorical data for comparison across categories

Stacked or composite bar graph of production on two lines, categorized by day of week
Bar Graph

- Composed of discrete bars
- Represents categorical data for comparison across categories

Topographical bar chart of New York City population during the day, and at night
Pie Chart

- Circular chart
- Relative comparison of amounts of parts of a whole

Pie chart of advertising dollars categorized by media type
**Pie Chart**

- Circular chart
- Relative comparison of amounts of parts of a whole

3D pie chart of expenditure costs, categorized by expenditure type
Pie Chart

- Circular chart
- Relative comparison of amounts of parts of a whole

Exploded 3D pie chart of transportation use categorized by transportation type
Pie Chart

- Circular chart
- Relative comparison of amounts of parts of a whole

Pie chart of population by US state
Line Graph

- Displays the relationship between two types of information
- Also useful for time-series trends

Line graph of California’s population by year
**Line Graph**

- Displays the relationship between two types of information
- Also useful for time-series trends

Time series line graphs of S&P 500 earnings versus price index from 1870-2010
**Line Graph**

- Displays the relationship between two types of information
- Also useful for time-series trends

Stacked time series line graph of baby name popularity from 1900–2003
- Bars that display frequency of continuous data divided into interval “bins”

Histogram of population age binned by age 20 year age intervals
**Histogram**

- Bars that display frequency of continuous data divided into interval “bins”

Interval histogram (with overlay for comparison) of pixel brightnesses binned by greyscale
Histogram

- Bars that display frequency of continuous data divided into interval “bins”
Scatterplot

- Explore relationship between two variables

Scatterplot of county’s average income versus vote percentage for George Bush
Scatterplot

- Explore relationship between two variables

Scatterplot of Yellowstone's Old Faithful wait time versus eruption duration
Scatterplot

• Explore relationship between two variables

3D scatterplot (3 axes) with a fourth variable visualized using colour
Venn Diagram

- Show commonality and overlap

Mathematical description of Venn diagram’s areas of intersection, union, and single item
Venn Diagram

- Show commonality and overlap

Venn diagram for \( n=5 \) sets with five-fold rotational symmetry
Venn Diagram

- Show commonality and overlap

British Isles Venn diagram
Contact Information

Christopher G. Healey

healey@csc.ncsu.edu

http://www.csc.ncsu.edu/faculty/healey