Turning Data Into Information

Sponsored by:
The Illinois Department of Public Health
and
Illinois Public Health Institute
Center for Community Capacity Development
Webinar Objectives

• Describe the different levels of measurement
• Describe the best type of visual displays for each type of data
• Describe how various visual elements such as points, bars, lines, color, shape, shading can be used to effectively encode data in visual displays
• Describe how principles from the science of cognition and perception can be utilized to effectively present data
Presenter

Mark Edgar, PhD, MPH
Illinois Public Health Institute
University of Illinois at Springfield (UIS)
Introduction to Research and Statistical Terms

• What is research?
  – “Systematic investigation with analysis of data in order to discern what is actually the case” (Burns, p. 4) as opposed to the method normally used by lay people i.e. rules of thumb, commons sense, etc.
Statistics

• When you can measure what you are speaking about and express it in numbers, you know something about it; but when you cannot measure it, when you cannot express it in numbers, your knowledge is of the meager and unsatisfactory kind.

- Lord Kelvin (British physicist)
Statistics

• A knowledge of statistics is like a knowledge of foreign languages or of algebra; it may prove of use at any time under any circumstances.

- A.L. Bowley
Others have said…

- Statistics means never having to say you're certain.
- Figures don't lie, but liars figure.
  - *Samuel Clemens*
- There are lies, damned lies, and statistics!
  - *B. Disraeli*
However...

- It’s easy to lie with statistics, but it is easier to lie without them.

-Fredereick Mosteller
THE GOVERNMENT DOESN'T WANT TO HIDE BEHIND STATISTICS... BUT IT IS NEVERTHELESS TRUE THAT 72% OF THE 13.8% OF GPs WHO RESPONDED WERE 14-25% HAPPy WITH 45% OF OUR CHANGES!
WIND YOU LIKE NUMBERS TO PROVE THAT WELFARE IS GOOD OR A GENOCIDAL PLAGUE ON THE NATIONAL SOUL?

By Signe Wilkinson, Philadelphia Daily News, Cartoonists & Writers Syndicate
Gapminder

• An example of data visualization from the Kaolinska Institute

• http://www.ted.com/index.php/talks/hans_rosling_shows_the_best_stats_you_ve_ever_seen.html
• Communication

• Variation

• Summarization
Data-Information-Knowledge Continuum

Ackoff, 1989
Data-Information-Knowledge-Understanding-Wisdom

• Data: symbols

• Information: data that are processed to be useful; provides answers to "who", "what", "where", and "when" questions

• Knowledge: application of data and information; answers "how" questions

• Understanding: appreciation of "why"

• Wisdom: evaluated understanding

(Ackoff 1989)
Introduction to Research and Statistical Terms

- What is research?
- Ways of knowing
  - Method of Tenacity
  - Method of Authority
  - Method of Intuition
  - Method of Science
Introduction to Research and Statistical Terms

• The Scientific Method
  – Control
  – Operational definition
  – Replication
  – Hypothesis testing
Introduction to Research and Statistical Terms

- Quantitative vs. Qualitative methods
  - Quantitative-numbers, measures, reliability, validity, objectivity, generalizability
  - Data are numbers
  - Qualitative- Context, uniqueness vs. generalizability, interviews, observation
  - Data are words, observations, pictures
Introduction to Research and Statistical Terms

• Explaining Variation
  – In a study of hypertension can we explain the variation in BP in relation to a drug intervention
  – In a study of teen pregnancy can we explain the variation in rates in relation to family variables (education, income, etc.)
Introduction to Research and Statistical Terms

– Ask a question
– Design a study and specify its measures
– Collect data and describe them
– Simplify the data
– Interpret the data
– Generalize the finding

(Tal, 2001)
Variables

• A person place or thing that can take on more than 1 value
  – Sex
  – Systolic BP
  – Score on a depression scale
  – Time
  – Hair color
  – Height
  – Weight
  – Cholesterol level
Introduction to Statistics

• Levels of Measurement
  – Nominal
  – Ordinal
  – Interval
  – Ratio
Levels of Measurement

– Nominal
  • Race
  • Sex
  • Yes/No
  • Pregnant/Not pregnant
Levels of Measurement

– Ordinal
  • Order means something but not equal intervals
  • Example:
    Gold, Silver, Bronze medals in marathon
Levels of Measurement

– Interval

• Equal intervals (distance) between points
• Example: Bodily temperature difference between 98.0 and 99.0 same as difference between 100.0 and 101.0
• However no absolute zero, i.e. complete absence of something
Levels of Measurement

- Ratio
  - Ordered, equal distance and an absolute zero
  - Time
  - Weight
  - T Cells
  - Money

So,
- Ratios mean something
- We can perform all arithmetic operations on the data allowing the highest level of statistical analysis
Measures of Central Tendency

– Mean
– Median
– Mode
Measures of Central Tendency

– Mean

• What we think of as average, the arithmetic mean

\[ \frac{\sum x}{N} \]

or the sum of all x elements divided by the number of elements

Mean of \((3, 6, 4, 5, 2)\) = \(\frac{3+6+4+5+2}{5} = 4\)

Mean also signified by \(M\) or by \(\mu\) (pronounced mew) for a population

\(\bar{X}\) for a sample (pronounced X-bar)
Turning Data into Information
M = 4.4
Measures of Central Tendency

– Median

– The middle value of a distribution, arranged in rank order, where equal numbers of elements are on either side

<p>| | | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>16</td>
<td>6</td>
<td>11</td>
<td>24</td>
<td>17</td>
<td>4</td>
<td>19</td>
</tr>
<tr>
<td>4</td>
<td>6</td>
<td>9</td>
<td>11</td>
<td>16</td>
<td>17</td>
<td>19</td>
</tr>
</tbody>
</table>

Median
Measures of Central Tendency

– Median
– When there are an equal number of data elements

\[
16 + 20 / 2 = 18 = \text{Median}
\]
Measures of Central Tendency

- Mode
- Most common occurring value

<table>
<thead>
<tr>
<th>23</th>
<th>28</th>
<th>24</th>
<th>9</th>
<th>24</th>
<th>21</th>
<th>18</th>
<th>24</th>
</tr>
</thead>
</table>

Mode = 24

<table>
<thead>
<tr>
<th>23</th>
<th>24</th>
<th>9</th>
<th>24</th>
<th>7</th>
<th>9</th>
<th>9</th>
<th>24</th>
</tr>
</thead>
</table>

Bimodal (9, 24)
M=5.5
Measures of Dispersion

- Measures of Dispersion
  - Range
  - Variance
  - Standard Deviation
Measures of Dispersion

- Range
  - Highest value – lowest value

6  8  12  14  20  35  42

Range = 42 - 6 = 36
Measures of Dispersion

• Variance
  – Average difference between the scores and the mean but,
    • Unfortunately this always sums to 0
Variance

M = 4

Deviation from mean

Turning Data into Information

September 10, 2010
Measures of Dispersion

• Variance

So,

– We square the difference score to make them all positive, sum them (Sum of Squares), then divide by N to arrive at the variance
Measures of Dispersion

• Standard Deviation
  – The variance is expressed in units squared
  So,
  – We take the square root of the variance to return to the original unit of measure and we have.
  – The standard deviation signified by $\sigma$ for population, SD for a sample
Distributions

• In a frequency distribution the X axis is the value of the scores
• The Y axis represents the frequency of each score
Distributions

• Skewed
• Normal (Gaussian)
  – A mathematical abstraction, but very useful
  – We know what % of all scores lie under certain portions of the curve
Normal or Gaussian Distribution

~68%

~95%

~2.5%
Normal Distribution

• We can tell where any standardized or z-score lies along the distribution and therefore what % of scores lie below or above it

• Or what % of cases or scores lie under a portion of the curve
Correlation

- The relationship between two variables or how two variables are associated
- Correlation does not imply causation
- One variable could cause another but the observed correlation does not, by itself, show causation
Correlation

- Simplest method, scatter plot
- Correlation coefficient (r)
- Ranges from
  +1.00 perfect positive
To
  - 1.00 perfect negative
  0 = no correlation
Correlation

• Size of correlation -
  – the closer to -/+1.0 the greater the correlation

• Negative does not mean no correlation
Table and Graphs

– Definitions
– Use
– Strengths
– Examples
– Summary
Tables

• Definition
  – Data is arrayed in rows and columns
  – Data is encoded as text
How are Tables Used?

• Look up individual data points

• Look for patterns in data
Strengths

• High precision/specificity

• Can handle different units of measure
Precision/specificity

• Pi to 100 places!
  – 3.1415926535 8979323846 2643383279 5028841971 6939937510 5820974944 5923078164 0628620899 8628034825 3421170679
### Precision

<table>
<thead>
<tr>
<th>Site</th>
<th>Units Sold</th>
<th>Dollar amount</th>
<th>% of Forecast</th>
</tr>
</thead>
<tbody>
<tr>
<td>Plant A</td>
<td>25</td>
<td>$1205</td>
<td>57%</td>
</tr>
<tr>
<td>Plant B</td>
<td>37</td>
<td>$1751</td>
<td>62%</td>
</tr>
<tr>
<td>Plant C</td>
<td>24</td>
<td>$1190</td>
<td>52%</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Site</th>
<th>Units Sold</th>
<th>Dollar amount</th>
<th>% of Forecast</th>
</tr>
</thead>
<tbody>
<tr>
<td>Plant A</td>
<td>25</td>
<td>$1205.67</td>
<td>57.59%</td>
</tr>
<tr>
<td>Plant B</td>
<td>37</td>
<td>$1751.99</td>
<td>62.33%</td>
</tr>
<tr>
<td>Plant C</td>
<td>24</td>
<td>$1190.90</td>
<td>52.23%</td>
</tr>
</tbody>
</table>
### Different Units of Measure

<table>
<thead>
<tr>
<th>Site</th>
<th>Units Sold</th>
<th>Dollar Amount</th>
<th>% of Forecast</th>
</tr>
</thead>
<tbody>
<tr>
<td>Plant A</td>
<td>25</td>
<td>$1205</td>
<td>57%</td>
</tr>
<tr>
<td>Plant B</td>
<td>37</td>
<td>$1751</td>
<td>62%</td>
</tr>
<tr>
<td>Plant C</td>
<td>24</td>
<td>$1190</td>
<td>52%</td>
</tr>
</tbody>
</table>

Turning Data into Information

September 10, 2010
Use tables when...

- There is a need to
  - “Look up” individual values
  - Compare individual values
  - Accommodate precise values
  - Accommodate different units of measurement
Showing Relationships in Tables

• Types of relationships
  – Quantitative to categorical
    • One set of quantitative values to one set of categorical subdivisions
    • One set of quantitative values and the intersection of multiple categories
  – Quantitative to quantitative
    • One set of quantitative values related to multiple categorical divisions
    • >1 set of quantitative values related to same categorical division
Quantitative to categorical

– One set of quantitative values to one set of categorical subdivisions

<table>
<thead>
<tr>
<th>Subdivisions</th>
<th>Dollar amount</th>
</tr>
</thead>
<tbody>
<tr>
<td>Miranda</td>
<td>$1205.67</td>
</tr>
<tr>
<td>Mallory</td>
<td>$1751.99</td>
</tr>
<tr>
<td>Elena</td>
<td>$1190.90</td>
</tr>
</tbody>
</table>

1 to 1 Quantitative to Categorical relationship
Quantitative to Categorical Relationships

– One set of quantitative values to one set of categorical subdivisions

<table>
<thead>
<tr>
<th>Income Level</th>
<th>% with Insurance</th>
</tr>
</thead>
<tbody>
<tr>
<td>&lt;$15,000</td>
<td>61%</td>
</tr>
<tr>
<td>$15,000-34,999</td>
<td>79%</td>
</tr>
<tr>
<td>$35,000-50,000</td>
<td>91%</td>
</tr>
<tr>
<td>&gt; $50,000</td>
<td>94%</td>
</tr>
</tbody>
</table>

1 to 1 Quantitative to Categorical relationship
Quantitative to Categorical Relationships

– One set of quantitative values and the intersection of multiple categories

<table>
<thead>
<tr>
<th>Salesperson</th>
<th>January</th>
<th>February</th>
<th>March</th>
</tr>
</thead>
<tbody>
<tr>
<td>Miranda</td>
<td>$1122.22</td>
<td>$1205.67</td>
<td>$1113.82</td>
</tr>
<tr>
<td>Mallory</td>
<td>$2013.12</td>
<td>$1751.99</td>
<td>$1221.86</td>
</tr>
<tr>
<td>Elena</td>
<td>$1012.33</td>
<td>$1190.90</td>
<td>$1222.06</td>
</tr>
</tbody>
</table>
Quantitative to Categorical Relationships

– One set of quantitative values and the intersection of multiple categories

<table>
<thead>
<tr>
<th>Income Level</th>
<th>% with Insurance by Year</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>2003</td>
</tr>
<tr>
<td>&lt;$15,000</td>
<td>61%</td>
</tr>
<tr>
<td>$15,000-34,999</td>
<td>79%</td>
</tr>
<tr>
<td>$35,000-50,000</td>
<td>91%</td>
</tr>
<tr>
<td>&gt; $50,000</td>
<td>94%</td>
</tr>
</tbody>
</table>
Quantitative to Quantitative Relationships

– One set of quantitative values related to multiple categorical subdivisions

<table>
<thead>
<tr>
<th>Salesperson</th>
<th>Sales</th>
<th>Returns</th>
<th>Net Sales</th>
</tr>
</thead>
<tbody>
<tr>
<td>Miranda</td>
<td>$1122.22</td>
<td>$1205.67</td>
<td>$1113.82</td>
</tr>
<tr>
<td>Mallory</td>
<td>$2013.12</td>
<td>$1751.99</td>
<td>$1221.86</td>
</tr>
<tr>
<td>Elena</td>
<td>$1012.33</td>
<td>$1190.90</td>
<td>$1222.06</td>
</tr>
</tbody>
</table>
Quantitative to Quantitative Relationships

– One set of quantitative values related to multiple categorical subdivisions

<table>
<thead>
<tr>
<th>Income Level</th>
<th>% with Insurance by Year</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>2003</td>
</tr>
<tr>
<td>&lt;$15,000</td>
<td>61%</td>
</tr>
<tr>
<td>$15,000-34,999</td>
<td>79%</td>
</tr>
<tr>
<td>$35,000-50,000</td>
<td>91%</td>
</tr>
<tr>
<td>&gt; $50,000</td>
<td>94%</td>
</tr>
</tbody>
</table>
## Quantitative to Quantitative Relationships

> 1 set of quantitative values related to the same categorical subdivision

<table>
<thead>
<tr>
<th>Salesperson</th>
<th>Sales</th>
<th>Returns</th>
<th>Net Sales</th>
</tr>
</thead>
<tbody>
<tr>
<td>Miranda</td>
<td>$1122.22</td>
<td>$1205.67</td>
<td>$1113.82</td>
</tr>
<tr>
<td>Mallory</td>
<td>$2013.12</td>
<td>$1751.99</td>
<td>$1221.86</td>
</tr>
<tr>
<td>Elena</td>
<td>$1012.33</td>
<td>$1190.90</td>
<td>$1222.06</td>
</tr>
</tbody>
</table>
Quantitative to Quantitative Relationships

> 1 set of quantitative values related to the same categorical subdivision

<table>
<thead>
<tr>
<th>Income Level</th>
<th>% with Insurance by Year</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>2003</td>
</tr>
<tr>
<td>&lt;$15,000</td>
<td>61%</td>
</tr>
<tr>
<td>$15,000-34,999</td>
<td>79%</td>
</tr>
<tr>
<td>$35,000-50,000</td>
<td>91%</td>
</tr>
<tr>
<td>&gt; $50,000</td>
<td>94%</td>
</tr>
</tbody>
</table>
## Tables

<table>
<thead>
<tr>
<th>Function</th>
<th>Relationship Type</th>
<th>Relationship</th>
</tr>
</thead>
<tbody>
<tr>
<td>Look up</td>
<td>Quantitative-to- Categorical</td>
<td>Single set of quantitative values and a single set of categorical subdivisions</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Single set of quantitative values and intersection of multiple categories</td>
</tr>
<tr>
<td>Comparison</td>
<td>Quantitative-to- Quantitative</td>
<td>Single set of quantitative values and associated with multiple categorical subdivisions</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Multiple sets of quantitative values and same categorical subdivisions</td>
</tr>
</tbody>
</table>

Turning Data into Information
September 10, 2010
Table Variations

• Unidirectional

<table>
<thead>
<tr>
<th>Race</th>
<th>% Seen a Dr. in Past 12 Months</th>
</tr>
</thead>
<tbody>
<tr>
<td>Black</td>
<td>79%</td>
</tr>
<tr>
<td>Asian/Pacific Islander</td>
<td>94%</td>
</tr>
<tr>
<td>American Indian</td>
<td>81%</td>
</tr>
<tr>
<td>White</td>
<td>99%</td>
</tr>
</tbody>
</table>

Categories Arranged in 1 direction
Table Variations

• Bidirectional

<table>
<thead>
<tr>
<th>Race</th>
<th>% Seen a Dr in Past 12 Months</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Rural IL</td>
</tr>
<tr>
<td>Black</td>
<td>79%</td>
</tr>
<tr>
<td>Asian/Pacific Islander</td>
<td>94%</td>
</tr>
<tr>
<td>American Indian</td>
<td>81%</td>
</tr>
<tr>
<td>White</td>
<td>99%</td>
</tr>
</tbody>
</table>
### Summary

- **Quantitative to Categorical Relationships**

<table>
<thead>
<tr>
<th>Relationship</th>
<th>Structural Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>Single set of quantitative values and a single set of categorical subdivisions</td>
<td><strong>Unidirectional</strong> Yes</td>
</tr>
<tr>
<td></td>
<td><strong>Bidirectional</strong> Not applicable- single set of categorical subdivisions</td>
</tr>
<tr>
<td>Single set of quantitative values and intersection of multiple categorical subdivisions</td>
<td><strong>Unidirectional</strong> Yes. Sometimes preferable due to convention</td>
</tr>
<tr>
<td></td>
<td><strong>Bidirectional</strong> Yes. Saves space</td>
</tr>
</tbody>
</table>

Turning Data into Information
September 10, 2010

69
### Summary

- **Quantitative to Quantitative Relationships**

<table>
<thead>
<tr>
<th>Relationship</th>
<th>Structural Type</th>
</tr>
</thead>
</table>
| Single set of quantitative values associated with multiple categorical subdivisions | Unidirectional: Yes  
Bidirectional: Yes. Quantitative values close together |
| Multiple sets of quantitative values associated with the same set of categorical subdivisions | Unidirectional: Yes.  
Bidirectional: Yes, but tends to get messy |
Graphs

- Values displayed in area demarcated by axes
- Values encoded as visual objects
- Axes provide scales (both quantitative and categorical) to assign values and labels to visual objects
Use Graphs When…

- Graphs are used to display relationships among multiple variables by giving them shape
- Allow for identification of:
  1. Trends
  2. Comparisons
  3. Exceptions/anomalies
  4. Similarities/differences
  5. Variance
Use Graphs When...

<table>
<thead>
<tr>
<th>Use Tables When</th>
<th>Use Graphs When</th>
</tr>
</thead>
<tbody>
<tr>
<td>Need to look up values</td>
<td>Message is contained in shape of the values</td>
</tr>
<tr>
<td>Need to compare individual values</td>
<td>To reveal relationships among multiple values</td>
</tr>
<tr>
<td>Precise values required</td>
<td></td>
</tr>
<tr>
<td>Quantitative values involve more than 1 unit of measurement</td>
<td></td>
</tr>
</tbody>
</table>
Encoding Data in Graphs

• Elements Used to Encode Data
  – Points
  – Lines
  – Bars
  – Shapes with 2-D area
Points

- Simple Scatter plot

Values on X and Y

X Axis

Y Axis
Lines

- Connect individual data points
- Show trend of series of data points
Lines

- Trend line/fit line/"line of best fit"

[Graph showing a scatter plot with a trend line]
Bars

- Really a “thick” line
- Thickness should be equal or eye “sees” greater value
- Horizontal or vertical works
2-D Areas

- Pie
- Stacked bar
Encoding Categorical Data

- Position
- Color
- Point shape
- Line style
Position

- Clear which bar is related to which category
Color

- Color (shading) encodes different values- Legend identifies

Percentage with Insurance

- White
- Non-white

Rural IL | Urban IL | IL | US

Turning Data into Information
September 10, 2010
Point Shape

- Point shape differentiates categories

### Percentage with Insurance

<table>
<thead>
<tr>
<th></th>
<th>White</th>
<th>Non-white</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rural IL</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Urban IL</td>
<td></td>
<td></td>
</tr>
<tr>
<td>IL</td>
<td></td>
<td></td>
</tr>
<tr>
<td>US</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Turning Data into Information
September 10, 2010
Point Shape and Color

- Using shape and color to encode data

Percentage with Insurance

- White
- Non-white

Rural IL Urban IL IL US

Turning Data into Information
September 10, 2010
Point Shape and Color
Fill Patterns

• **A Bad Example** of encoding with fill patterns

Percentage with Insurance

- **White**
- **Non-white**

<table>
<thead>
<tr>
<th></th>
<th>Rural IL</th>
<th>Urban IL</th>
<th>IL</th>
<th>US</th>
</tr>
</thead>
<tbody>
<tr>
<td>0%</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>20%</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>40%</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>60%</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>80%</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>100%</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Fill Patterns

• A little better but, **don’t do it!**
Line Style

- Using line style to differentiate data series

Percentage with Insurance

- White
- Non-white
- Other
Relationships in Graphs

- Nominal
- Time-series
- Ranking
- Part-to-Whole
- Frequency distribution
- Correlation
Nominal

- Categorical subdivisions have no order they are just different
- Goal is to show the differences in the quantitative values related to each category

<table>
<thead>
<tr>
<th>Percentage with Insurance</th>
</tr>
</thead>
<tbody>
<tr>
<td>US</td>
</tr>
<tr>
<td>IL</td>
</tr>
<tr>
<td>Urban IL</td>
</tr>
<tr>
<td>Rural IL</td>
</tr>
</tbody>
</table>

Turning Data into Information
September 10, 2010
Time-series

- Shows quantitative values associated with time (years, months, decades)
Ranking

- Ordinal relationships

Percentage with Insurance

- US
- Urban IL
- Rural IL
- IL

Turning Data into Information
September 10, 2010
Part-to-Whole

- Ratio relationships
- Comparing individual quantitative values to the total of the quantitative values
Part-to-Whole

- Better
- Even Better

![Bar chart showing various categories like Protection, Promotion, Admin, Supplies, Misc, Other with percentages ranging from 0% to 40%.]
Part-to-Whole

- Best!

Bar chart showing the distribution of categories:
- Admin
- Promotion
- Protection
- Misc
- Supplies
- Other
Deviation

- Show deviation of quantitative values from some set of primary (reference) values

Expenses to Budget

-20% -10% 0% 10% 20% 30% 40%

Admin
Promotion
Protection
Misc
Supplies
Other
Deviation

Immunization Rates - IL

Turning Data into Information
September 10, 2010
Frequency Distribution

- Histogram shows frequency - Vertical by convention

Cases By Age Group

- 0 to 5: 10 cases
- 6 to 11: 5 cases
- 12 to 16: 15 cases
- 16 and over: 20 cases
Frequency Distribution

- Histogram over time
- Frequency polygon emphasizes shape of distribution

Cases by Day

Days

Cases by Day

Days
Deviation

Cases by Day

Days

Cases
0
5
10

1 2 3 4 5 6 7 8 9 10
Correlation

- Simple Scatter plot

![Graph showing correlation with values on the x and y axes.](image)
Correlation

• Simple Scatter plot

Values on X and Y

X Axis

Y Axis
## Relationships Summary

<table>
<thead>
<tr>
<th>Relationship</th>
<th>Points</th>
<th>Lines</th>
<th>Points &amp; Lines</th>
<th>Bars</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Nominal</strong></td>
<td>Avoid</td>
<td>Avoid</td>
<td>Avoid</td>
<td>Horizontal or vertical</td>
</tr>
<tr>
<td><strong>Time-Series</strong></td>
<td>Avoid</td>
<td>Categories on X axis;</td>
<td>Categories on X axis;</td>
<td>Categories on X axis;</td>
</tr>
<tr>
<td></td>
<td></td>
<td>quantitative on Y axis.</td>
<td>quantitative on Y axis.</td>
<td>quantitative on Y axis.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Emphasis on pattern</td>
<td>Emphasis on individual values and</td>
<td>Emphasis on individual values</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>pattern</td>
<td></td>
</tr>
<tr>
<td><strong>Ranking</strong></td>
<td>Avoid</td>
<td>Avoid</td>
<td>Avoid</td>
<td>Horizontal preferable</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>with values sorted</td>
</tr>
<tr>
<td><strong>Part-to-Whole</strong></td>
<td>Avoid</td>
<td>Avoid</td>
<td>Avoid</td>
<td>Horizontal or vertical</td>
</tr>
<tr>
<td><strong>Deviation</strong></td>
<td>Avoid</td>
<td>Good when combined with time</td>
<td>Good when combined with time series</td>
<td>Horizontal or vertical</td>
</tr>
<tr>
<td></td>
<td></td>
<td>series</td>
<td>and when you need some emphasis on</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>individual values</td>
<td></td>
</tr>
<tr>
<td><strong>Frequency</strong></td>
<td>Avoid</td>
<td>Frequency polygon</td>
<td>Avoid</td>
<td>Histogram</td>
</tr>
<tr>
<td><strong>Distribution</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Correlation</strong></td>
<td>Scatter</td>
<td>Avoid</td>
<td>Trend line</td>
<td>Horizontal or vertical</td>
</tr>
<tr>
<td></td>
<td>plot</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

(Adapted from Few 2004, p. 83)
Visual Perception, Cognition and Communication

• Iconic memory
  – Pre-attentive processing

• Short term memory (working memory)
  – Temporary, limited storage capacity
  – 3-9 objects “in the queue”

• Long term memory
Visual Perception, Cognition and Communication

- Attributes of form in pre-attentive processing
  - Orientation
  - Line length
  - Line width
  - Size
  - Shape
  - Curvature
  - Added marks
  - Enclosure

(Ware 2000)
Visual Perception, Cognition and Communication

• Attributes of Color
  – Hue (what we usually think of as color)
  – Intensity (lightness/darkness)
  – Saturation (purity of the color)
Visual Perception, Cognition and Communication

- Context
Gestalt Principles

• Principle of Proximity
• Principle of Similarity
• Principle of Enclosure
• Principle of Closure
• Principle of Continuity
• Principle of Connection
### Principle of Proximity

<p>| | | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
</tbody>
</table>

<p>| | | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
</tbody>
</table>

Turning Data into Information

September 10, 2010
Principle of Similarity

- Notice how quickly attention is drawn to the similar objects
Principle of Similarity

- Subtle differences can lead the eye to read down columns

<table>
<thead>
<tr>
<th>XXX</th>
<th>XXX</th>
<th>XXX</th>
<th>XXX</th>
<th>XXX</th>
<th>XXX</th>
<th>XXX</th>
<th>XXX</th>
</tr>
</thead>
<tbody>
<tr>
<td>XXX</td>
<td>XXX</td>
<td>XXX</td>
<td>XXX</td>
<td>XXX</td>
<td>XXX</td>
<td>XXX</td>
<td>XXX</td>
</tr>
<tr>
<td>XXX</td>
<td>XXX</td>
<td>XXX</td>
<td>XXX</td>
<td>XXX</td>
<td>XXX</td>
<td>XXX</td>
<td>XXX</td>
</tr>
<tr>
<td>XXX</td>
<td>XXX</td>
<td>XXX</td>
<td>XXX</td>
<td>XXX</td>
<td>XXX</td>
<td>XXX</td>
<td>XXX</td>
</tr>
</tbody>
</table>

- Or across rows

<table>
<thead>
<tr>
<th>XXX</th>
<th>XXX</th>
<th>XXX</th>
<th>XXX</th>
<th>XXX</th>
<th>XXX</th>
<th>XXX</th>
<th>XXX</th>
</tr>
</thead>
<tbody>
<tr>
<td>XXX</td>
<td>XXX</td>
<td>XXX</td>
<td>XXX</td>
<td>XXX</td>
<td>XXX</td>
<td>XXX</td>
<td>XXX</td>
</tr>
<tr>
<td>XXX</td>
<td>XXX</td>
<td>XXX</td>
<td>XXX</td>
<td>XXX</td>
<td>XXX</td>
<td>XXX</td>
<td>XXX</td>
</tr>
<tr>
<td>XXX</td>
<td>XXX</td>
<td>XXX</td>
<td>XXX</td>
<td>XXX</td>
<td>XXX</td>
<td>XXX</td>
<td>XXX</td>
</tr>
<tr>
<td>XXX</td>
<td>XXX</td>
<td>XXX</td>
<td>XXX</td>
<td>XXX</td>
<td>XXX</td>
<td>XXX</td>
<td>XXX</td>
</tr>
<tr>
<td>XXX</td>
<td>XXX</td>
<td>XXX</td>
<td>XXX</td>
<td>XXX</td>
<td>XXX</td>
<td>XXX</td>
<td>XXX</td>
</tr>
</tbody>
</table>

Turning Data into Information
September 10, 2010
Principle of Enclosure

• We see objects enclosed together as belonging together
Principle of Enclosure

- Drawing attention to “outliers”
Principle of Closure

• We have a perceptual tendency to complete objects
Principle of Closure

• So, we don’t really need a full enclosure, 2 axes will do
Principle of Continuity

- We perceive aligned objects as “complete”
Principle of Connection

- Notice how difficult it is to see a pattern here
Principle of Connection

- Lines “connect the dots” and communicate the shape of the data distribution
## Summary of Gestalt Principles

<table>
<thead>
<tr>
<th>Principle</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Proximity</td>
<td>Objects close together perceived as a group</td>
</tr>
<tr>
<td>Similarity</td>
<td>Objects that share attributes are perceived as a group</td>
</tr>
<tr>
<td>Enclosure</td>
<td>Objects that appear to have a boundary (are enclosed) are perceived as a group</td>
</tr>
<tr>
<td>Closure</td>
<td>Open structure are perceived as closed or complete</td>
</tr>
<tr>
<td>Continuity</td>
<td>Objects that are aligned or appear to be continuous are perceived as a group</td>
</tr>
<tr>
<td>Connection</td>
<td>Objects that are connected are perceived as a group</td>
</tr>
</tbody>
</table>
Tufte

- Data-ink ratio = \frac{\text{Data-ink}}{\text{Total Ink in graph}}

- Chartjunk
  - Unintentional optical art (illusions)
  - The Grid
  - Self-promoting graphics
Data-Ink Ratio

• Notice how the gridlines overwhelm the data
• They add lots of ink but not much data-ink
Data-Ink Ratio

• What could we do here to maximize the data-ink ratio?
Data-Ink Ratio

• Take out borders, gridlines and re-format Y axis scale

Percentage with Insurance

- White
- Non-white

0% 20% 40% 60% 80% 100%

Rural IL Urban IL IL US
Data-Ink Ratio

• Improved data-ink ratio on the right

Percentage with Insurance

White □ Non-white □
Chartjunk

• We’ve seen an example of unintentional optical art i.e. chartjunk
Chartjunk

• This is so much better

Percentage with Insurance

<table>
<thead>
<tr>
<th></th>
<th>White</th>
<th>Non-white</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rural IL</td>
<td>80%</td>
<td>70%</td>
</tr>
<tr>
<td>Urban IL</td>
<td>80%</td>
<td>70%</td>
</tr>
<tr>
<td>IL</td>
<td>80%</td>
<td>70%</td>
</tr>
<tr>
<td>US</td>
<td>80%</td>
<td>70%</td>
</tr>
</tbody>
</table>
Data-Ink Ratio

- Grids should be subtle, if used at all
Tufte’s Principles

- Above all, show the data
- Maximize data-ink ratio
- Erase non-data-ink
- Erase redundant data ink
- Forgo chartjunk
- Revise and edit

(Tufte 2001)
References


Feedback

• Please complete the online feedback survey following this session
• Your input is used to plan future offerings.
• Check your email for a link to the survey or you will find it on the IPLAN website where you downloaded this Webinar.
Join us for a live Q and A session

Dial – 1 (877) 411-9748
Enter Passcode – 3467868#
If you have training or technical assistance follow-up needs, contact:

Laurie Call, Director
Center for Community Capacity Development, IPHI
Laurie.Call@iphionline.org