Data Visualization & Infographics

Jennifer Grek Martin
Mike Smit

School of Information Management
The Plan

1. Data Visualization: Choosing the right tools (M. Smit)
   – Hands-on: Visualizing data

2. Escaping the Default Settings: Colour and type (J. Grek Martin)
   – Hands-on: Choosing colours and typefaces/fonts in visualization

3. Context & Message (Data: M. Smit; Viz: J. Grek Martin)
   – Hands-on: Placing your visualization in context

4. The Infographic (Data: M. Smit; Viz: J. Grek Martin)
   – Hands-on: Turning a visualization into an infographic
Not The Plan

• Visualizing GIS data
• Visualizing social media data
• Working with messy / unstructured data
• (Most) statistics
• Data collection, management, policy, etc.
Resources

https://goo.gl/VFXLnJ
Segment 1: Data Visualization: Choosing the right tools

Mike Smit

School of Information Management
Topics
What Is Data?
DIKW Continuum

Marshall, Bob via Twitter @flowchainsensei.
https://twitter.com/flowchainsensei/status/408167162344648704
Data:

Information:

Knowledge:

Wisdom:
What Is Data Visualization?
What is Data Visualization?

• Communicating data using a visual medium
• Or, ascending the DIKW pyramid using visual power
Why do we visualize data?

• To explore
• To communicate
  – One message
    • e.g. “Our sales are increasing”
    • e.g. “Our product is 15% shinier”
    • e.g. “The Communist Party is gaining in the polls”
  – Multiple messages
  – A visual summary of an entire data set
Why do we visualize data?

• To explore / understand

• To communicate
  – One message
    • e.g. “Our sales are increasing”
    • e.g. “Our product is 15% shinier”
    • e.g. “The Communist Party is gaining in the polls”
  – Multiple messages
  – A visual summary of an entire data set
How do you visualize data?

• Choosing what you wish to communicate
  – Mostly out of scope
  – Except: effective visualization is often an exercise in a) summarizing data and b) deleting data

• Choosing an appropriate visualization

• Choosing the right software tool
Choosing a Visualization

• What kind of data do you have?
• What do you want to communicate?
• What is your audience / context?
What kind of data do you have?

• Specialized data (e.g. geospatial): not covered here

• General quantitative data:
  – Nominal (Discrete)
  – Ordinal (Greater or Less than)
  – Interval-Ratio (Continuous)
    • Ratio: has an absolute 0

• (In increasing order of precision)
Nominal Data

Degree: _____ Bachelor of Science
       _____ Bachelor of Arts
       _____ Bachelor of Engineering

Postal code of your home address? ______

• Obtained when names are used to (arbitrarily) label attributes
• Only indicates that there is a difference among attributes of a variable: classification
Nominal Data

• Can have artificially assigned numerical values, but numbers have no meaning
• Numbers do not express magnitude or order
• Cannot take an average
• Quantification of qualitative variables
• Example: degree program
• Graph: pie chart, bar chart
Ordinal Data

How would you rate the quality of service
Poor _ Fair _ Average _ Good _ Excellent_

• Contains information as to better or worse, or greater or less, but not about how much better or how much worse

• Classification & ranking. Indicates a difference, and attributes can be ordered or ranked
Ordinal Data

• Values themselves have no meaning, not an expression of absolute magnitude
• Ordering of numbers has a meaning

• Example: education levels
• Graph: histogram, bar chart
Interval-Ratio Data

How many points have you cumulated from our bonus point program? _____

What is your age? _____

• Classification, Ranking, and precise distance between two values can be measured
• Unlike Ordinal, which can rank from high to low but with unknown distances between categories.
Interval/Ratio Data

- Numbers and their values have meaning
- Distance between values has meaning
- Interval variables have no absolute zero point (e.g., temperature in Celsius and Fahrenheit)
- Ratio variables have a true zero point (e.g., weight, distance)
- Examples: number of people, weight
- Graph: line graph, area chart
How do I tell?

- Depends on the variable **AND** how it was collected

  - What is your annual income?

  - Do you consider yourself:
    - ___ Upper class
    - ___ Middle class
    - ___ Working class
Why does this matter?

• The type of data you have limits what you can do with it
• Graph choice depends on level of measurement.
• Reminder of examples:
  – Nominal data – bar chart, pie chart.
  – Ordinal data – histogram, bar chart.
  – Interval/Ratio data – area chart, line graph
Why does this matter?

• Bad example:
Why does this matter?

Canada Unemployment rate by Province for 2013
Why does this matter?
Choosing a Visualization

• What kind of data do you have?
• What do you want to communicate?
• What is your audience / context?
What do you want to communicate?

• Three main categories for most communication tasks, the Three C’s
  – Comparison: best/worst, high/low
    • Column / bar chart
  – Change: usually change in values over time: trends, transitions
    • Line graph
    • Area chart
  – Composition: breaking a total up into its constituent parts
    • Pie chart
Aside re: Pie Charts

• Pie charts show composition, but should only be used to show large differences in relative size
A YEAR IN CANADA:

- January
- February
- March
- April
- May
- June
- July
- August
- September
- October
- November
- December
Three C’s: in Combination

• Comparison and Composition
  – Stacked bar chart

Category 1

Category 2

Category 3

Category 4

Series 1

Series 2

Series 3
Three C’s: in Combination

- Change and Composition
  - Stacked area chart
Three C’s: in Combination

- Comparison and Change
  - Line chart
Three C’s: in Combination

- Comparison, Change, and composition
- Incomprehensible abominations
Three C’s: in Combination

- Comparison, Change, and composition
  - Incomprehensible abominations
Choosing a Visualization

• What kind of data do you have?
• What do you want to communicate?
• What is your audience / context?
### Audience & Presentation Context

#### US iPad Users and Penetration, 2010-2014

<table>
<thead>
<tr>
<th></th>
<th>2010</th>
<th>2011</th>
<th>2012</th>
<th>2013</th>
<th>2014</th>
</tr>
</thead>
<tbody>
<tr>
<td>iPad users (millions)</td>
<td>11.5</td>
<td>28.0</td>
<td>41.9</td>
<td>53.9</td>
<td>60.8</td>
</tr>
<tr>
<td>% change</td>
<td>-</td>
<td>143.9%</td>
<td>49.8%</td>
<td>28.5%</td>
<td>12.9%</td>
</tr>
<tr>
<td>% of total population</td>
<td>3.7%</td>
<td>8.9%</td>
<td>13.2%</td>
<td>16.9%</td>
<td>18.9%</td>
</tr>
<tr>
<td>% of internet users</td>
<td>5.1%</td>
<td>12.1%</td>
<td>17.5%</td>
<td>22.0%</td>
<td>24.2%</td>
</tr>
<tr>
<td>% of tablet users</td>
<td>88.0%</td>
<td>83.0%</td>
<td>76.4%</td>
<td>71.2%</td>
<td>68.0%</td>
</tr>
</tbody>
</table>

*Note: individuals of any age who use an iPad at least once per month*

*Source: eMarketer, Nov 2011*
iPad users in the United States (millions)

iPad users in US, as % of all tablet users

- 2010
- 2011
- 2012
- 2013
- 2014

0%
100%
0%
20%
40%
60%
80%
100%
61 million Americans use iPads
SO....

• Know your data
• Know your three C’s
• Know your audience
Finally...

• Be honest with your data

• Common lies:
  – Using 3d is a form of (usually) unintentional dishonesty
  – Truncating axes
  – Confusing data representations
  – Not following the usual norms
In 2014
New Democrats raised more than $9.5 million to kick off this election year stronger than ever.

$9,522,705.74

46,523
Canadians chipped in.

More than 10,000 supporters made their first-ever donation.

Thanks for an incredible year!
Is truncating the Y-axis misleading?

Percent of the time

NO

YES
53% Workforce planning
50% Standardizing HR metrics
43% Standardizing HR dashboards/reports
38% Delivering workforce analytics to executives
37% Delivering workforce analytics to Line Managers/business leaders
26% Predictive analysis
24% More powerful tools for HR analysts
15% Unifying different workforce analytics solutions
13% Replacing/upgrading existing analytics systems
UNEMPLOYMENT RATE UNDER PRESIDENT OBAMA

2011
SOURCE: BUREAU OF LABOR STATISTICS
Gun deaths in Florida

Number of murders committed using firearms

Source: Florida Department of Law Enforcement

C. Chan 16/02/2014

2005
Florida enacted its ‘Stand Your Ground’ law

873 721
Choosing the right software tool
Data Visualization Tools

- Do you want charts, graphs, or plots?
  - SAP's Lumira Cloud
  - Tableau Public
  - Plot.ly
  - IBM Watson Analytics
  - Microsoft Excel
Remove to improve (the data-ink ratio)
Context & Message

Mike Smit
Context of Your Data

• Context examples:
  – Measurement units
  – Where is my data from
  – How was my data collected
  – What are caveats / limitations of my data
  – Metadata in general

• Not necessarily in the visualization
Referencing Figures

• Things to avoid:
  – “The results are shown in Figure 2.”
  – ... or just inserting a figure and waiting for the reader to figure it out!

• Reference the figure when you talk about that data

• Describe non-obvious aspects of the figure
### Audience & Presentation Context

**US iPad Users and Penetration, 2010-2014**

<table>
<thead>
<tr>
<th></th>
<th>2010</th>
<th>2011</th>
<th>2012</th>
<th>2013</th>
<th>2014</th>
</tr>
</thead>
<tbody>
<tr>
<td>iPad users (millions)</td>
<td>11.5</td>
<td>28.0</td>
<td>41.9</td>
<td>53.9</td>
<td>60.8</td>
</tr>
<tr>
<td>% change</td>
<td>-</td>
<td>143.9%</td>
<td>49.8%</td>
<td>28.5%</td>
<td>12.9%</td>
</tr>
<tr>
<td>% of total population</td>
<td>3.7%</td>
<td>8.9%</td>
<td>13.2%</td>
<td>16.9%</td>
<td>18.9%</td>
</tr>
<tr>
<td>% of internet users</td>
<td>5.1%</td>
<td>12.1%</td>
<td>17.5%</td>
<td>22.0%</td>
<td>24.2%</td>
</tr>
<tr>
<td>% of tablet users</td>
<td>88.0%</td>
<td>83.0%</td>
<td>76.4%</td>
<td>71.2%</td>
<td>68.0%</td>
</tr>
</tbody>
</table>

*Note: individuals of any age who use an iPad at least once per month*

*Source: eMarketer, Nov 2011*
iPad users in the United States (millions)

iPad users in US, as % of all tablet users

2010 2011 2012 2013 2014

2010 2011 2012 2013 2014
Visual Context

• How does the human mind perceive the world around them?
Human Vision is Optimized to See Structure

Gestalt Principles of Visual Perception

- Proximity
- Similarity
- Continuity
- Closure
- Symmetry
- Figure/ground
- Common fate

Adapted from slides by Jeff Johnson illustrating his book *Designing with the Mind in Mind*, 2014.
Gestalt Principle: Proximity

• Items that are closer appear grouped

Left: rows

Right: columns
Gestalt Principle: Proximity

- Google search results

  Orchidaceae - Wikipedia, the free encyclopedia
  en.wikipedia.org/wiki/Orchidaceae
  Orchids are easily distinguished from other plants, as they share some very evident apomorphies. Among these are: bilateral symmetry (zygomorphism), many ...
  Etymology - Distribution - Taxonomy - Characteristics

  orchids, gift orchids, hobby growing
  www.orchids.com/
  Orchids.com Logo, Login | Home | My Account | Customer Service | Sign In or Register | Shopping Cart. 1-888-4ORCHID (1-888-467-2443) - 9 am-5 pm (Pacific) ...

  Pacific Orchid Exposition - San Francisco Orchid Society
  www.orchidsanfrancisco.org/poe.html
  The San Francisco Orchid Society would like to express its thanks to the Taiwan Tourist Bureau CBS and KCBS for their continued support in production and ...

  Orchids in the Park - San Francisco Orchid Society
  www.orchidsanfrancisco.org/orchidsinthepark.html
  Thanks to all volunteers and attendees for a successful "Orchids In The Park" event. -- Images by Jeff Harris. -- Images by Jeff Harris. <back to top> ...
Gestalt Principle: Proximity

• Google search results without blank space
Gestalt Principle: Similarity
Gestalt Principle: Continuity

- We tend to see continuous forms
- Blue line and orange line; not blue-and-orange > and <
- Even if cross-point is covered
Gestalt Principle: Continuity
Gestalt Principle: Continuity

- MacOS slider

Turn off when computer is not used for:

- 5 secs
- 10 secs
- 30 secs
- 1 min
- 5 mins
- Never
Gestalt Principle: Closure

- We tend to see whole, closed objects, not collections of fragments

- Overlapping circles & triangles, not odd fragments
Gestalt Principle: Closure

• We tend to see whole, closed objects, not collections of fragments

• Stacked pages, not odd bits of images
Gestalt Principle: Symmetry

• We tend to see simple figures rather than complex ones

• E.g., two overlapping diamonds; not other, more complex combinations
Gestalt Principle: Symmetry

• Thagard: *Coherence in Thought & Action*
Gestalt Principle: Figure/Ground
Gestalt Principle: Figure/Ground
Gestalt Principle: Common Fate

• Items that move together appear grouped
<table>
<thead>
<tr>
<th>VERY PRECISE QUANTITATIVE PERCEPTION</th>
<th>NOT VERY PRECISE QUANTITATIVE PERCEPTION</th>
<th>NO QUANTITATIVE PERCEPTION</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Length</strong></td>
<td><strong>Width</strong></td>
<td><strong>Orientation</strong></td>
</tr>
<tr>
<td><img src="image1" alt="Graph" /></td>
<td><img src="image2" alt="Graph" /></td>
<td><img src="image3" alt="Graph" /></td>
</tr>
<tr>
<td><strong>2D position</strong></td>
<td><strong>Size</strong></td>
<td><strong>Shape</strong></td>
</tr>
<tr>
<td><img src="image4" alt="Graph" /></td>
<td><img src="image5" alt="Graph" /></td>
<td><img src="image6" alt="Graph" /></td>
</tr>
<tr>
<td><strong>Intensity</strong></td>
<td><strong>Enclosure</strong></td>
<td></td>
</tr>
<tr>
<td><img src="image7" alt="Graph" /></td>
<td><img src="image8" alt="Graph" /></td>
<td></td>
</tr>
</tbody>
</table>
What is an infographic?

Infographics are a fun and quick way to learn about a topic without a ton of heavy reading. There are many different styles of infographics and data visualizations, but the ultimate goal for all infographics is to be shared. Learn what makes a great infographic, why they are useful for everyone, some tips to create a viral infographic and the numbers to back it all up.

An infographic is:

- A data-rich visualization of a story or thesis
- A tool to educate and inform
- A way to build brand awareness and inbound links at half the cost of standard online marketing campaigns