Data Visualization Principles for Dashboard Design

Olin College, Data Dashboard Design

Jerzy Wieczorek

11/10/15
Follow along

These slides and a summary checklist are at
http://www.stat.cmu.edu/~jwieczor/
Overview

There is a spectrum between exploratory graphics (for your own understanding as you analyze the data) and explanatory graphics (for communicating results to an audience).

Today’s advice applies to both, but we’ll focus on explanatory graphs.
Overview

- What should I graph?
  (let function constrain form, show both raw data and summaries, show precision/uncertainty)
- Can people read my graphs?
  (guides and captions, image format and quality, text and color)
- Can people understand my graphs?
  (comparisons, grouping and search, cognition, consistency)
- Where can I learn more about dataviz?
  (books, people to follow)
Principles

In general:

- Design each graph with a **clear task** in mind
- Use **common sense** to help audience read the graph
- Use knowledge of **visual perception** to help audience perform the task efficiently

For dashboard design:

- Design a single-screen view, for monitoring at a glance
- Give context: compare today’s summaries and exceptions vs. the past, competitors, or thresholds
- Use tables to show exact values; use graphs to show patterns
Design for a clear task

General principles:

▶ Design each graph with a clear task in mind
▶ Use common sense to help audience read the graph
▶ Use knowledge of visual perception to help audience perform the task efficiently
A visualization is a tool for showing order and patterns in data. As the creator, you can “generate order before people’s brains try to do it on their own.”

Cairo, *The Functional Art*: Cairo read a claim that *higher education is related to lower obesity*, but no evidence was shown. He found a US state-level dataset for rates of obesity and of college education. What graphical form can help show evidence for or against this claim?
Choice of graphical form

Bubbles on a map?
(larger = higher percentage)

Percentage of people with a BA degree or higher

Percentage of obese people
Choice of graphical form

Choropleth (colored thematic map)?
(darker = higher percentage)

Percentage of people with a BA degree or higher

Percentage of obese people
Choice of graphical form

Scatterplot shows the relationship directly. Doesn’t show spatial patterns, but that wasn’t our goal, so OK.
Choice of graphical form: skeumorphs

Imitation gauges ("skeumorphs") take more space, and give less context, than simple bars

Few, “Dashboard Design for Real-Time Situation Awareness”
Choice of graphical form

Lessons

- Decide on visual task, and choose a form that supports it.
- Just because you have certain variables (geography, time, social ties) doesn't mean it's helpful to show them (map, time series, network diagram).
- Just because car dashboards have gauges doesn't mean it's a helpful form for data dashboards.
Most useful graphic forms and visual variables

Data forms: point, line, bar
(including sparklines and small multiples)

Visual variables: color, point shape, point size, line width, line type
Which age group weighs the least?

Consistency
Consistency

Give all small multiples the same structure, usually including axis limits, to make comparisons easier and reduce cognitive load.
Consistency

Ensure design changes are meaningful (tied to data changes)
Consistency

More consistent redesign, Stephen Few
Consistency

Avoid meaningless visual variables like shadow or 3D
Consistency

Lessons

- Use consistent mappings (colors and shapes, axis limits) across graphs.
- Don’t reuse the same mappings across different data variables.
- Avoid meaningless variety in design.
- Avoid shadow, 3D, and other variables not mapped to data.
Semantic associations

Orange vs blue crab species: I’ve actually seen this in a talk (crabs dataset)
Semantic associations

Lessons

- Use meaningful mappings:
  orange vs blue crab species = orange and blue symbols.
- Use conventional mappings: blue = cold, red = hot.
- “More = more”:
  deeper saturation or larger size = higher value of variable.
Legibility

General principles:

- Design each graph with a clear task in mind
- **Use common sense to help audience read the graph**
- Use knowledge of visual perception to help audience perform the task efficiently
Legibility

What is done well here? What could be improved?

![Graph showing time (sec) vs. condition (ambiguous, high, low)]
**Title or caption:** explain anything not already on axes or legend. What are the individual plotted points (raw data by subject, vs. averages by group)? Is this the full dataset or a subset?

**Axis labels:** give units (“number of events”, “$”, “%”, “m/s”). Give readable labels, not default variable names (Pretest Score, not PRE_TEST_SCORE).

**Tick marks:** use round, readable numbers and avoid scientific notation. If “Dollars” axis has ticks like “1e6”, plot “Dollars in millions” instead.

**Grid lines:** omit or make very light. Do not overwhelm the data.

**Legend:** use same ordering as on plot. Or, if possible, omit and use direct labels instead.
Color

Ensure your colors can be distinguished from each other:

- Around 10% of men and 1% of women are colorblind.
- Some color palettes do not photocopy well.

Use Color Brewer to choose a well-tested palette.
Image quality

Understand when to use bitmap vs. vector formats.

Ensure graph’s text size is similar to surrounding body text.

Don’t stretch graphs! Create & save them at the right size in the first place.
Vector vs bitmap explained

Bitmap: common formats

- **jpg/jpeg** is lossy, designed for photos but not text/charts
- **png** is lossless, good for text/charts, common on web

Vector: common formats

- **svg** can display in browser, common on web
- **pdf** is for standalone doc (or to put inside another pdf)
### Recommended formats and resolutions

<table>
<thead>
<tr>
<th>Software</th>
<th>Recommended graphics device</th>
</tr>
</thead>
<tbody>
<tr>
<td>Illustrator</td>
<td>svg</td>
</tr>
<tr>
<td>pdflatex</td>
<td>pdf, png (600 ppi)</td>
</tr>
<tr>
<td>Office</td>
<td>png (600 ppi)</td>
</tr>
<tr>
<td>web</td>
<td>png (72 ppi)</td>
</tr>
</tbody>
</table>

(ppi = Pixels Per Inch)

From Wickham, *ggplot2*, Table 8.3
Save images at intended final size

Decide on target size and create the graph that way from the start. Gives better quality than changing size after saving, and avoids stretching/distorting the graph.

For example:

- Default textwidth in a LaTeX article is around 5.4 inches
- My Wordpress blog has a default width of 500 pixels
Visual Perception

In general:

- Design each graph with a clear task in mind
- Use common sense to help audience read the graph
- Use knowledge of visual perception to help audience perform the task efficiently
Quantitative comparisons

We’ll try an experiment on next few slides:

<table>
<thead>
<tr>
<th></th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
</tr>
</thead>
<tbody>
<tr>
<td>Positions</td>
<td>1</td>
<td>?</td>
<td>?</td>
<td>?</td>
</tr>
<tr>
<td>Lengths</td>
<td>1</td>
<td>?</td>
<td>?</td>
<td>?</td>
</tr>
<tr>
<td>Angles</td>
<td>1</td>
<td>?</td>
<td>?</td>
<td>?</td>
</tr>
<tr>
<td>Areas</td>
<td>1</td>
<td>?</td>
<td>?</td>
<td>?</td>
</tr>
</tbody>
</table>
Quantitative perceptual tasks: position, aligned

A B C D

32 / 70
Quantitative perceptual tasks: length
Quantitative perceptual tasks: angle
Quantitative perceptual tasks: area
Quantitative perceptual tasks: answers

<table>
<thead>
<tr>
<th></th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
</tr>
</thead>
<tbody>
<tr>
<td>Positions</td>
<td>1</td>
<td>3/4</td>
<td>1/4</td>
<td>2/4</td>
</tr>
<tr>
<td>Lengths</td>
<td>1</td>
<td>2/4</td>
<td>3/4</td>
<td>1/4</td>
</tr>
<tr>
<td>Angles</td>
<td>1</td>
<td>2/3</td>
<td>1/3</td>
<td>4/3</td>
</tr>
<tr>
<td>Areas</td>
<td>1</td>
<td>2/4</td>
<td>1/4</td>
<td>3/4</td>
</tr>
</tbody>
</table>

Cleveland and McGill (1984)

Cleveland, *The Elements of Graphing Data*
Quantitative perceptual tasks: effect of angle orientation

Same angle looks wider when bisector is horizontal.
Ordering of perceptual tasks

Cleveland and McGill’s ordering (split over 2 slides)

- 2D position along common, aligned scale
- 2D position along common, but unaligned scales
- Length

Allows more accurate comparisons
Ordering of perceptual tasks

- Slope
- Area
- Volume
- Colour hue
- Colour intensity
- Angle

Allows more generic comparisons
Distance

Cleveland and McGill (1984)

Figure 4. Graphs from position-length experiment.
Quantitative perceptual tasks

Lessons:

- Best to show quantitative variables with position or length. Use points, lines, or bars.
- Bars encode length, so start bars at 0. If you must zoom in, use dotplots (encoding position) instead.
- Avoid stacked bars (not aligned). Use dots or lines (aligned baselines) instead.
- Avoid pies, area, and volume entirely.
- Choose and order hues sensibly. Use Color Brewer.
- Place things-to-be-compared near each other.
Pre-attentive processing: example task

Find and count the 6s

```
5 9 3 2 4 6 5 1 7 9 3 9 8 5 2 1 2 7 2 5
9 2 8 5 9 4 3 5 7 4 5 4 6 9 4 1 0 8 4 5
9 8 2 0 3 4 7 4 7 6 1 8 7 5 9 1 7 4 4 5
5 0 5 6 2 5 8 1 7 0 3 8 7 5 8 3 6 7 2 6
9 3 5 8 7 8 3 4 0 2 9 3 9 2 1 2 9 4 4 2
```
Preattentive processing: example task

Find and count the 6s now

```
5 9 3 2 4 6 5 1 7 9 3 9 8 5 2 1 2 7 2 5
9 2 8 5 9 4 3 5 7 4 5 4 6 9 4 1 0 8 4 5
9 8 2 0 3 4 7 4 7 6 1 8 7 5 9 1 7 4 4 5
5 0 5 6 2 5 8 1 7 0 3 8 7 5 8 3 6 7 2 6
9 3 5 8 7 8 3 4 0 2 9 3 9 2 1 2 9 4 4 2
```
Preattentive processing

We automatically process and notice certain features, while others require conscious thought to find.

We process faster when there are few categories to distinguish.
Preattentive processing: features

Colin Ware, *Information Visualization*
Preattentive processing: features

- Gray/value
- Enclosure
- Convexity/concavity
- Addition
- Juncture
- Parallelism
Preattentive processing: color and icons

With toned-down colors, red flags stand out more effectively

Few, “Dashboard Design for Real-Time Situation Awareness”
Preattentive processing

Lessons

▶ Distinguish categorical groups by features like hue & shape.
▶ Hue also lets you use direct labels instead of a legend.
▶ Don’t try to show too many groups on one plot. Use small multiples to show more sub-groups.
▶ If highlighting one group/element, use a preattentive attribute. Highlight only that element, not all elements.
Separable dimensions

Some examples from Colin Ware, *Information Visualization*

<- More integral . . . More separable ->

- red-green
- yellow-blue
- x-size
- y-size
- size orientation
- color shape
- color motion
- color location
Integral dimensions example

US Census Bureau map using hue and saturation
Lessons

- Use color and another variable (shape, size, orientation, motion).
- Use small multiples rather than different plotting symbols.
- Avoid mixing 2 aspects of color, or 2 aspects of size.
- Don’t show too many grouping variables at once.
Dashboard design principles:

- **Design a single-screen view, for monitoring at a glance**
- Give context: compare today’s summaries and exceptions vs. the past, competitors, or thresholds
- Use tables to show exact values; use graphs to show patterns
Single-screen view

**Overall Performance**
- Hold Time
- Call Duration
- Abandonments

**Utilization**
- Reps Today: 23
- Reps Online: 20

**Volume**
- This Hour: 373
- Today: 1,322
- This Month: 25,934
- Per Hour Today: 17,834

**Rep Performance**

<table>
<thead>
<tr>
<th>Name</th>
<th>Orders Per Hr</th>
<th>Calls Per Hr</th>
<th>Call Duration (minutes)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Jacobs, S</td>
<td>5</td>
<td>10</td>
<td></td>
</tr>
<tr>
<td>McKinsey, J</td>
<td>5</td>
<td>13</td>
<td></td>
</tr>
<tr>
<td>Smith, V</td>
<td>6</td>
<td>12</td>
<td></td>
</tr>
<tr>
<td>Wilcox, R</td>
<td>9</td>
<td>14</td>
<td></td>
</tr>
<tr>
<td>Clark, P</td>
<td>10</td>
<td>14</td>
<td></td>
</tr>
<tr>
<td>Simons, B</td>
<td>10</td>
<td>16</td>
<td></td>
</tr>
<tr>
<td>Newman, A</td>
<td>11</td>
<td>15</td>
<td></td>
</tr>
<tr>
<td>Bailey, S</td>
<td>11</td>
<td>16</td>
<td></td>
</tr>
<tr>
<td>Barclay, T</td>
<td>11</td>
<td>17</td>
<td></td>
</tr>
<tr>
<td>Jimenez, J</td>
<td>12</td>
<td>16</td>
<td></td>
</tr>
<tr>
<td>X Chou, A</td>
<td>12</td>
<td>17</td>
<td></td>
</tr>
<tr>
<td>Kata, H</td>
<td>12</td>
<td>17</td>
<td></td>
</tr>
<tr>
<td>Silverstein, C</td>
<td>13</td>
<td>18</td>
<td></td>
</tr>
<tr>
<td>Schuster, P</td>
<td>13</td>
<td>18</td>
<td></td>
</tr>
<tr>
<td>Truman, M</td>
<td>13</td>
<td>19</td>
<td></td>
</tr>
<tr>
<td>X Pierce, B</td>
<td>14</td>
<td>19</td>
<td></td>
</tr>
<tr>
<td>Fisher, J</td>
<td>14</td>
<td>20</td>
<td></td>
</tr>
<tr>
<td>Jung, T</td>
<td>14</td>
<td>20</td>
<td></td>
</tr>
<tr>
<td>English, S</td>
<td>15</td>
<td>21</td>
<td></td>
</tr>
<tr>
<td>Wiley, P</td>
<td>15</td>
<td>21</td>
<td></td>
</tr>
<tr>
<td>Johnson, N</td>
<td>16</td>
<td>21</td>
<td></td>
</tr>
<tr>
<td>X Lucas, J</td>
<td>16</td>
<td>22</td>
<td></td>
</tr>
<tr>
<td>Forester, R</td>
<td>17</td>
<td>23</td>
<td></td>
</tr>
</tbody>
</table>

**Telesales Dashboard**

Few, *Information Dashboard Design*
Give context

Dashboard design principles:

- Design a single-screen view, for monitoring at a glance
- **Give context:** compare today’s summaries and exceptions vs. the past, competitors, or thresholds
- Use tables to show exact values; use graphs to show patterns
Sparklines give historical context.
Bullet graphs compare to performance thresholds.
Use tables for lookup

Dashboard design principles:

- Design a single-screen view, for monitoring at a glance
- Give context: compare today’s summaries and exceptions vs. the past, competitors, or thresholds
- **Use tables to show exact values; use graphs to show patterns**
Use tables for lookup

Graphs are better for making comparisons and finding patterns. Tables are better for highlighting individuals or showing precise values, if audience needs them:

▶ Exceptional performers: most and least effective sales reps; best and worst selling product lines
▶ Precise numeric values: dollar amounts down to the penny for Accounting purposes
Principles

In general:

- Design each graph with a clear task in mind
- Use common sense to help audience read the graph
- Use knowledge of visual perception to help audience perform the task efficiently

For dashboard design:

- Design a single-screen view, for monitoring at a glance
- Give context: compare today’s summaries and exceptions vs. the past, competitors, or thresholds
- Use tables to show exact values; use graphs to show patterns
Further resources

- **My checklist** for the material covered today
- Few, *Information Dashboard Design*: short but solid coverage of dashboard design
- Few, *Show Me The Numbers*: general graph and table design in a business context
- Cairo, *The Functional Art*: great overview of infographics from a data journalism perspective
- Robbins, *Creating More Effective Graphs*: quick, accessible summary of classic advice by Tufte and Cleveland; includes a graph design checklist
- Ware, *Information Visualization*: thorough textbook on insights from perception research
- Donahue, *Fundamental Statistical Concepts in Presenting Data*: real case studies from a statistical consultant
Thanks!

PS: Don’t forget your UOCD and HFID skills: do some usability testing on your dashboard!
A few more principles, if there’s time

- Align small multiples for the task
- Rank/order informatively
- Show derived variables directly
- Use Gestalt principles
Alignment

Among male newborns, compare by race
Among male newborns, compare by race:
easier search now, though harder comparison
Western state areas (1000s of sq miles)
Western state areas (1000s of sq miles)
Derived variables

William Playfair, one of the earliest line charts
What does the difference look like?
Derived variables

Differences shown directly, by Cleveland and McGill

CHART OF BALANCE AGAINST ENGLAND

Figure 28. Playfair data.
Alignment, ranking, and derived variables

Lessons

- Decide on visual task, and helpfully align elements to be compared.
- As you explore data, try several arrangements.
- Order your dots/bars meaningfully: rank by a variable, not alphabetically.
- If differences or ratios are interesting, compute and plot them directly.
Gestalt

Gestalt = “pattern” in German

We automatically structure data into patterns / groups using certain features
Lessons

- Distinguish categorical groups by similarity, proximity, or enclosure.
- Use proximity to structure your layout (arrange small multiples).
- Use connection to show groups on line chart, parallel coordinates chart, or network graph.
- To highlight one group, use enclosure or similarity.