

STAT 209

Data Computing and Visualization

June 3rd, 2021

Colin Reimer Dawson

Outline

Elements of data graphics
Graphical Elements
Human-Sensitive Design

The Grammar of Graphics

Elements of data graphics

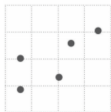
- Visual cues
 - position, size, color, etc.
- Coordinate system
 - how are data points organized?
- Scale
 - relationship between variable and distance in space
- Context
 - what in the world is the data about?
- Faceting
 - What are the sub-parts (facets) of the graph?

Visual cues

When you visualize data, you encode values to shapes, sizes, and colors.

Position

Where in space the data is

**Length**

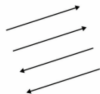
How long the shapes are

**Angle**

Rotation between vectors

**Direction**

Slope of a vector in space

**Shapes**

Symbols as categories

**Area**

How much 2-D space

**Volume**

How much 3-D space

**Color saturation**

Intensity of a color hue

**Color hue**

Usually referred to as color



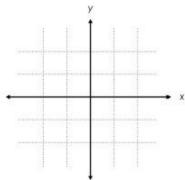
FIGURE 3-3 Visual cues

Coordinate systems

There are a variety of them, from cylindrical to spherical, but these three will cover most of your bases.

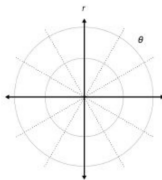
Cartesian

If you've ever made a graph, the x- and y-coordinate system will look familiar to you.



Polar

Pie charts use this system. Coordinates are placed based on radius r and angle θ .



Geographic

Latitude and longitude are used to identify locations in the world. Because the planet is round, there are multiple projections to display geographic data in two dimensions. This one is the Winkel tripel.

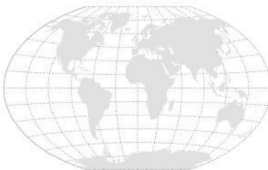


FIGURE 3-13 Commonly used coordinate systems

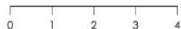
Source: Nathan Yau, *Data Points*

Scales

Along with coordinate systems, they dictate where the shapes are placed and how objects are shaded.

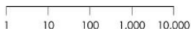
Linear

Values are evenly spaced



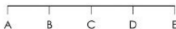
Logarithmic

Focus on percent change



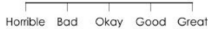
Categorical

Discrete placement in bins



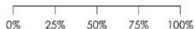
Ordinal

Categories where order matters



Percent

Representing parts of a whole



Time

Units of months, days, or hours



FIGURE 3-15 Scales

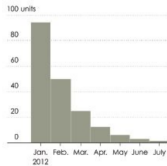
Source: Nathan Yau, *Data Points*

Working parts

Several pieces work together to make a graph. Sometimes these are explicitly shown in the visualization and other times they form a visual in the background. They all depend on the data.

Title of this Graph

A description of the data or something worth highlighting to set the stage.

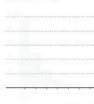


Source: Somewhere reputable

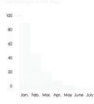
Title of this Graph
A description of the data or something worth highlighting to set the stage.



Title of this Graph
A description of the data or something worth highlighting to set the stage.



Title of this Graph
A description of the data or something worth highlighting to set the stage.



Title of this Graph
A description of the data or something worth highlighting to set the stage.

**Visual Cues**

Visualization involves encoding data with shapes, colors, and sizes. Which cues you choose depends on your data and your goals.

Coordinate System

You map data differently with a scatterplot than you do with a pie chart. It's x- and y-coordinates in one and angles with the other. It's cartesian versus polar.

Scale

Increments that make sense can increase readability, as well as shift focus.

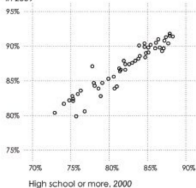
Context

If your audience is unfamiliar with the data, it's your job to clarify what values represent and explain how people should read your visualization.

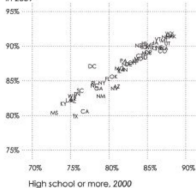
Source: Nathan Yau, *Data Points*

Position

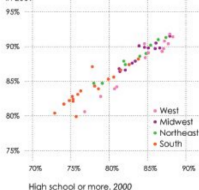
High school
or more,
in 2009

**Position + Symbols**

High school
or more,
in 2009

**Position + Color**

High school
or more,
in 2009

**Position + Symbols + Color**

High school
or more,
in 2009

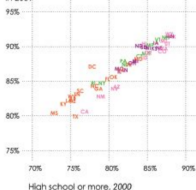
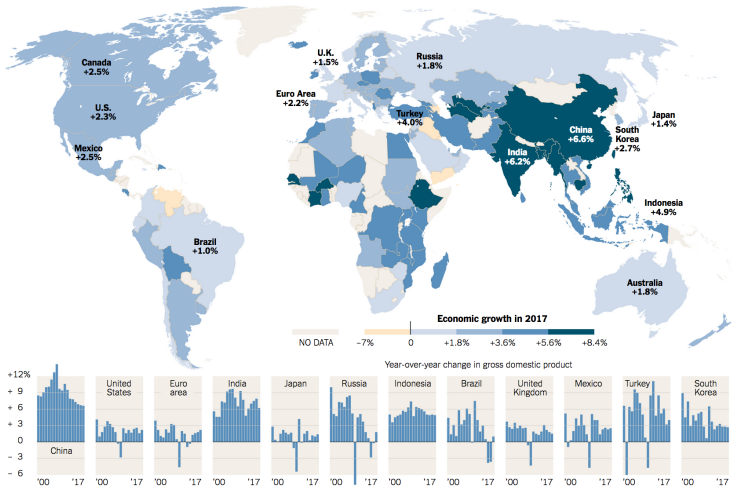


FIGURE 3-39 Position, symbols, and colors in scatterplots

Growth Across the Globe

For the first time since the financial crisis a decade ago, all of the world's major economies are growing.



Some figures are estimates

Source: The Conference Board; Bureau of Labor Statistics | By Karl Russell

A Perceptual Hierarchy

Cleveland and McGill (1985): people better at judging:

- position than size
- length than angle
- 1D differences than 2D differences
- 2D differences than 3D differences
- size than color

Dos and Don'ts

- Prefer length scale to color scale
- Never use pie charts
- ABC: Always Bring Context
- Above all else, *show the data* (Edward Tufte)

The Grammar of Graphics

- *The Grammar of Graphics* by Leland Wilkinson (1999, 2005) set out to define "parts of speech" and "grammar rules" of data visualization
- An "ontology of graphs"
- Implemented in R in ggplot2 package (Hadley Wickham, 2010)
 - ggplot (or ggplot1) was sort of a beta version; not really used today

Graphical elements in ggplot2

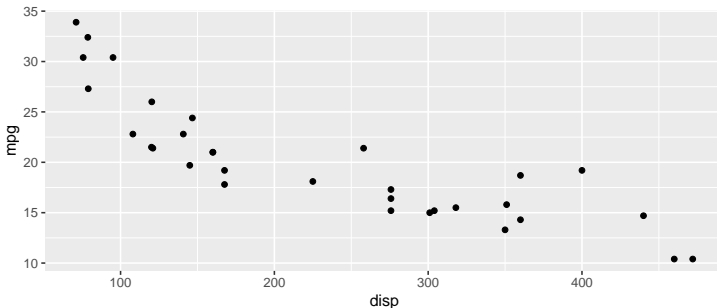
Graphical element	ggplot2 object(s)
The dataset	<code>data=</code> argument
Geometric objects	<code>geom_*()</code> functions
Mappings of variables to cues	<code>aes()</code> function
Scales	<code>scale_*()</code> functions
Faceting	<code>facet_wrap()</code> , <code>facet_grid()</code>

Combination elements via the '+' operator to produce a plot as an R *object* (can be assigned to things, operated on later, etc.)

The minimal template

Plots must at a minimum have data (`data=`), a mapping (`aes()`), and at least one geometry element (`geom_*()`)

```
library(tidyverse)
ggplot(data = mtcars, aes(x = disp, y = mpg)) +
  geom_point()
```



Something more complex

```
ggplot(data = mtcars, aes(x = disp, y = mpg, color = factor(cyl))) +  
  geom_point() +  
  geom_line() +  
  facet_wrap(~am) +  
  scale_color_brewer(palette = "Set1")
```

