Week 05

Histograms & Perception

INFO 3402: Information Exposition

Brian Keegan, Ph.D.
Assistant Professor, Information Science
brian.keegan@colorado.edu

University of Colorado Boulder

Course Overview

Module	Week	Dates	Computational skill	Communication skill
Shaping	1	Jan 11, Jan 13	Loading	Documentation
	2	Jan 18, Jan 20	Aggregation	Summarization
	3	Jan 25, Jan 27	Joining	Validation
	4	Feb 1, Feb 3	Tidying	Tables
Distribution	5	Feb 8, Feb 10	Histograms	Perception
	6	Feb 15, Feb 17	Box plots	Audience
Comparison	7	Feb 22, Feb 24	Cat plots	Objectives
	8	Mar 1, Mar 3	Faceted plots	Simplicity
Trend	9	Mar 8, Mar 10	Line plots	Trust
	10	Mar 15, Mar 17	Stacked plots	Annotation
	11	Mar 22, Mar 24	Spring Break	
Relationship	12	Mar 29, Mar 31	Scatter plots	Fallacies
	13	Apr 5, Apr 7	Heatmaps	Persuasion
Spatial	14	Apr 12, Apr 14	Choropleths	Conventions
	15	Apr 19, Apr 21	Point plots	Design
Projects	16	Apr 26, Apr 28	Projects	

Recap – Module 01

Week	Skills	Datasets
01	Data science mindset; loading data; documentation	Census; Boulder+Broomfield weather
02	Types of data; groupby-aggregation; pivot tables	CDC deaths; time use
03	Concatenation; joining/merging	Baseball; county cannabis, crime, population, & COVID
04	Tidying data; melting/unmalting; table formatting	DIA and CDOT traffic

Weekly overview

- Notebook exercises (ungraded) – Assigned Tuesdays and reviewed Thursdays
- Add a Visualization Critique (ungraded) – Discussed in class on Thursdays
- Complete "Thursday Questions" form! (ungraded/optional) – Review in class on Thursdays
- Weekly Assignment (graded, 2%) – Assigned Tuesdays and due Sundays by midnight
- Weekly Quiz (graded, 1%) – End of class on Thursdays (open 12:00pm – 12:30pm)
- Module Assignment 02 (graded, 5%) – Due Tuesday, February 22 by 11am*

* We'll discuss due dates in a few slides

Readings

- Questions for Friday's Weekly Quiz 05 will be drawn from these readings

- matplotlib Tutorial – The Lifecycle of a Plot (for WA05)
- Matplotlib Tutorial – Basic Usage (for WA05)
- Berinato (2016). Chapter 2. "When A Chart Hits Our Eyes: Some Science of How We See".
- Schwabish (2021). Chapter 1. "Visual Processing and Perceptual Rankings"

Weekly Assignment 05

- **Skills**: Object-oriented matplotlib style, making histograms, multiple subplots
- **Data**: CU Boulder FCQ data 2010 – 2019

- New Weekly Assignment deadlines: Sundays at 11:59pm

Module Assignment 01

Module Assignment 01

- Tell a story using any of the datasets from Weeks 01 – 04
  - Census; Weather; CDC deaths; Time use; Baseball; County cannabis/crime/COVID; DIA/CDOT traffic
- 700 – 1000 words with at least one table and one visualization
- Use Peng & Matsui (2018) and Abela (2013) readings on forming a question, structuring a story
- Module Assignment 01 will be due Tuesday, February 08 by 11am
  - Submit URL of your Medium post to Canvas or save and submit as an HTML file
  - Tag your post on Medium with “INFO3402S22MA1” and whatever other tags you’d like

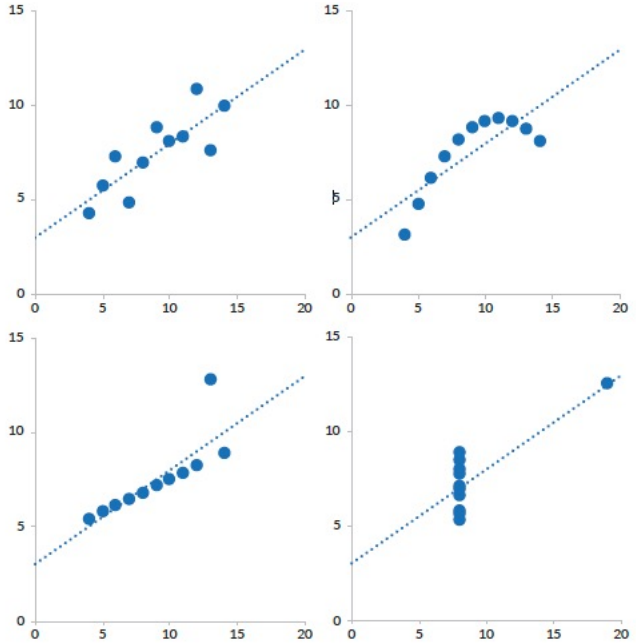
Questions

- What was hard or confusing about the assignment? What should we do differently?
- What was hard or confusing about Medium?
  - Use the “INFO3402S22A1” tag (despite what I said elsewhere)
- What can we debug so you can submit a strong assignment and build a great portfolio?

Visual perception

Motivation

- Visualization plays an important role in data analysis
  - Anscombe's Quartet famously demonstrates need
- Also about communication and storytelling, with few words
- How do you encode meaning into a picture so that your reader decodes the same meaning?
  - Alternatively: How is meaning being encoded that leads the reader to the wrong conclusions?



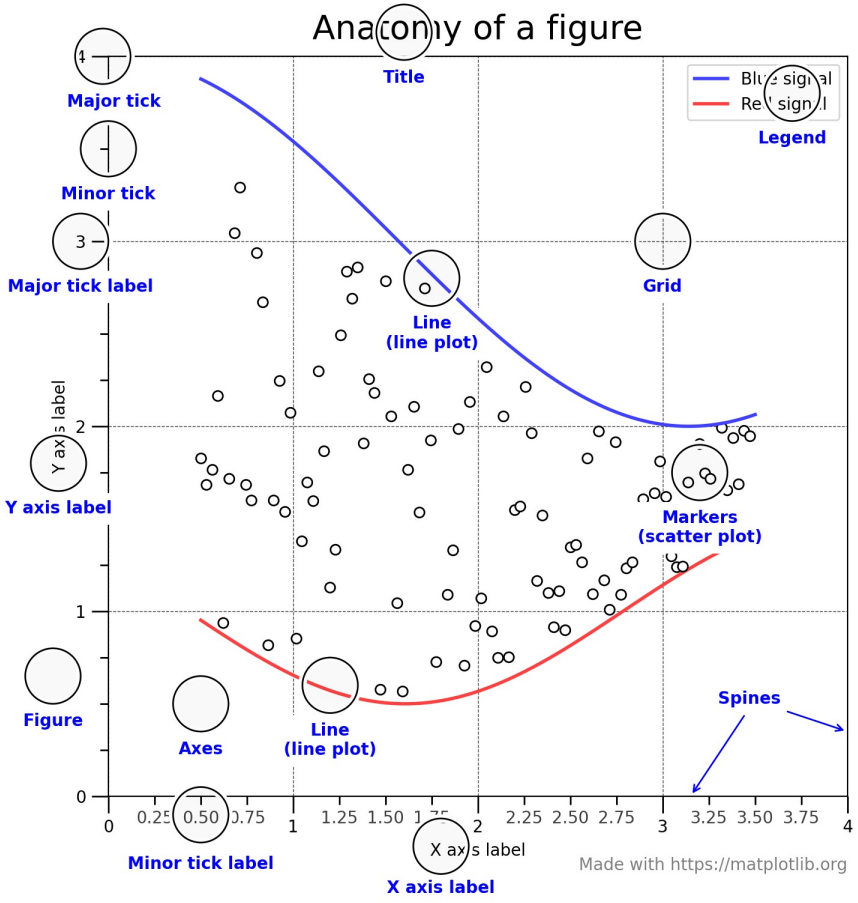
Berinato's Five Ideas

1. **We don't go in order** → Visualizations are read and interpreted differently by everyone
2. **We see first what stands out** → We focus on outliers, dominant colors, trends
3. **We see only a few things at once** → Meaning fades as more elements are present
4. **We seek meaning and make connections** → We seek stories to try to make sense
5. **We rely on conventions and metaphors** → We have expectations about how data should look

The Lifecycle of a Plot

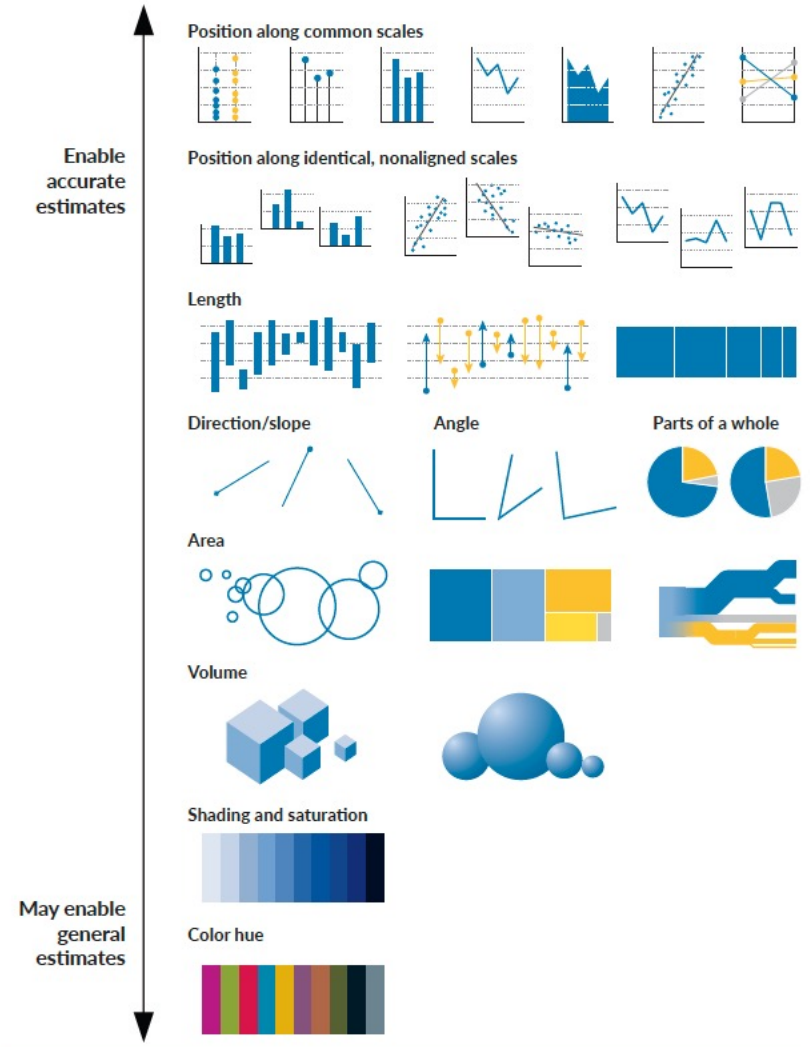
- The default design decisions in a visualization are a starting point, not a finish line
- Literally ***any*** part of a matplotlib figure can be customized in some way
  - But some are easier to change than others!
- Adding labels & annotations and changing colors & styles are all part of the design process
- What look like big scary code blocks were developed by a process of accumulation
  - Literally **no one** expects you to be able to write all this code out from memory on the first time.
  - Use the documentation, Q&A sites, tutorials, other code – but make sure to cite it!
  - Try something. → Does it make it better or worse? → Keep the better things. → Try something else.

Parts of a matplotlib Figure



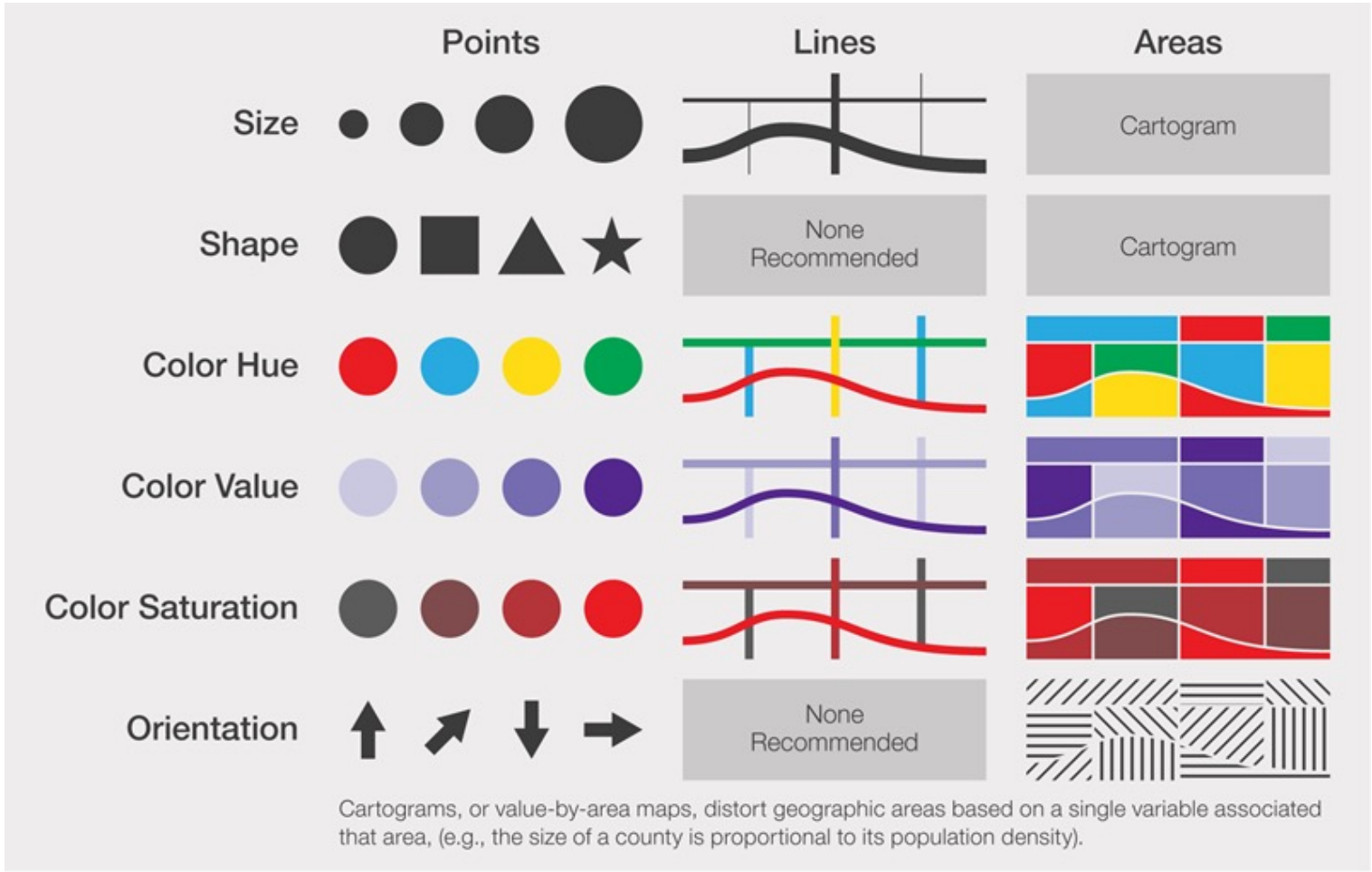
Perceptual ranking

- Visualizations can be designed for different purposes
- Some visualizations let readers make accurate estimates and comparisons (top) but can be harder to interpret
- Other visualizations can be easier to interpret (bottom) but readers cannot make accurate estimates or comparisons



Schwabish, J. (2021). Chapter 1.

Visual variables



Cartograms, or value-by-area maps, distort geographic areas based on a single variable associated that area, (e.g., the size of a county is proportional to its population density).

	Qualitative	Quantitative	
	Nominal	Ordinal	Numerical
Size	P	G	G
Shape	G	P	P
Color Hue	G	M[a]	M[a]
Color Value	P	G	M
Color Saturation	P	G	M
Orientation	G	M	M
Arrangement	M	P	P
Texture	G	M	M
Transparency	M	G	P
Crispness	P	G	P
Resolution	P	G	P

G = good; M = marginally effective; P = poor

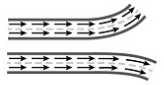
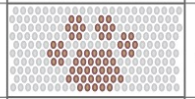
[a] The particular hues selected must be logically ordered.

UCGIS. Symbolization and Visual Variables.

Visual hierarchy and layout

Type	Stronger		vs		Weaker
Size*	Large		vs		Small
Color: Hue	Warm		vs		Cool
Color: Saturation	Intense		vs		Pale
Color: Value‡	Dark/Light		vs		Light/Dark
Focus	Sharp		vs		Blurry
Position	Center		vs		Periphery
Continuity	Edged		vs		Continuous
Grouping*	Isolated		vs		Grouped
Arrangement	Ordered		vs		Random
Distribution	Dense		vs		Sparse
Cropping	Entire		vs		Cropped
Detail*	Intricate		vs		Generalized

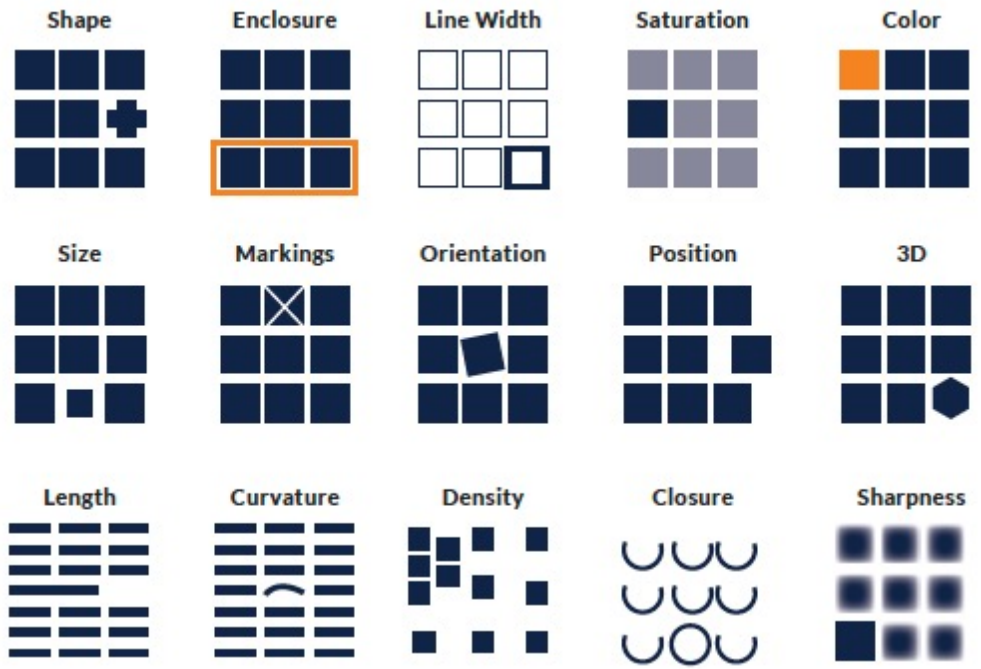
*Cartography texts disagree on which is stronger and which weaker.
‡Value's strength depends on the background, dark is stronger on a light background and light on dark one.

Name		Type	Description
Proximity		Grouping	Objects close to each other group into clusters or other forms
Similarity		Grouping	Objects with similar visual variable characteristics (other than location) form groups.
Direction		Grouping	Objects that flow in similar direction form groups.
Common Fate		Grouping	Objects that share the same movement form groups.
Experience or Familiarity		Grouping	Familiar shapes or arrangements form groups.
Symmetry		Grouping	Objects that are symmetrical about a line or point form a group
Stages		Grouping	Complex series of objects group into steps or stages (primarily for animations).
Simplicity		Interpretation	A group of complex objects will form a single, simpler shape
Figure-Ground		Interpretation	Some objects will group and be promoted as figure above a ground (this is the most important gestalt principal for cartography!).
Closure		Interpretation	The eye will complete an object's form to create a whole.
Good Continuation		Interpretation	The eye will continue the form of an object that is partially hidden or that is interrupted.

UCGIS. Visual Hierarchy and Layout.

Gestalt principles of perception

- **Proximity** → close objects are similar
- **Similarity** → objects sharing style are similar
- **Enclosure** → objects within a boundary are similar
- **Closure** → missing data is imputed by proximity
- **Continuity** → aligned objects are related
- **Connection** → connected objects are similar



Schwabish, J. (2021). Chapter 1.

Where to go next?

- Visual and interaction design have both scientific and gestalt approaches
- Color theory: https://gistbok.ucgis.org/bok-topics/color-theory
- *Semiology of Graphics* → theoretical foundations by Bertin & Berg in 1967
- *Grammar of Graphics* → 2005 framework adopted by many visualization packages

- Do you have to do *everything* in matplotlib? Of course not! But a great place to start!

Kesari, G. (2018). My Talk on Grammar of Graphics: The Secret Sauce of Powerful Data Stories.

Notebook time!

- Download “Week 05 – Lecture.ipynb”
  - We’ll be using data from previous classes

Next class

Next Class

- Review concepts and exercises from last class
  - Complete “Thursday Questions” form! https://forms.gle/1tA27125rjFkH2R46 (ungraded/optional)
- Critique a data narrative or visualization
  - Post a link and a few sentences about a data visualization on Canvas discussion (ungraded/optional)
- Time to work on Weekly Assignment
  - Weekly Assignment due on Sunday by midnight by submitting HTML notebook to Canvas
- Weekly quiz at the end of class (12:00–12:30)

Module Assignment 02

- Module Assignment 02 will be due Tuesday, February 08 by 11am
  - Submit URL of your Medium post to Canvas or save and submit as an HTML file
  - Tag your post on Medium with “INFO3402S22A2” and whatever other tags you’d like
- Discuss details on Thursday