Successful, Effective Data Visualizations

Cameron Riddell

DO(N'T) USE THIS C<>DE

Overview

- ➔ **What's important to think about when creating a visualization?**
- ➔ What are the components of a message?
- ➔ Why would I want to communicate visually?
- ➔ What is the dichotomy of goals?
- ➔ What are the different ways to represent data visually?
- ➔ What are some basic charts and when should I use them?
- ➔ Why is the use of color important?
- ➔ **How can I communicate more clearly?**

What should I consider?

- the message
- your audience
- layout
- metrics
- charts & marks

What is a message?

signal

---

noise

Communicating Visually

Why should I bother?

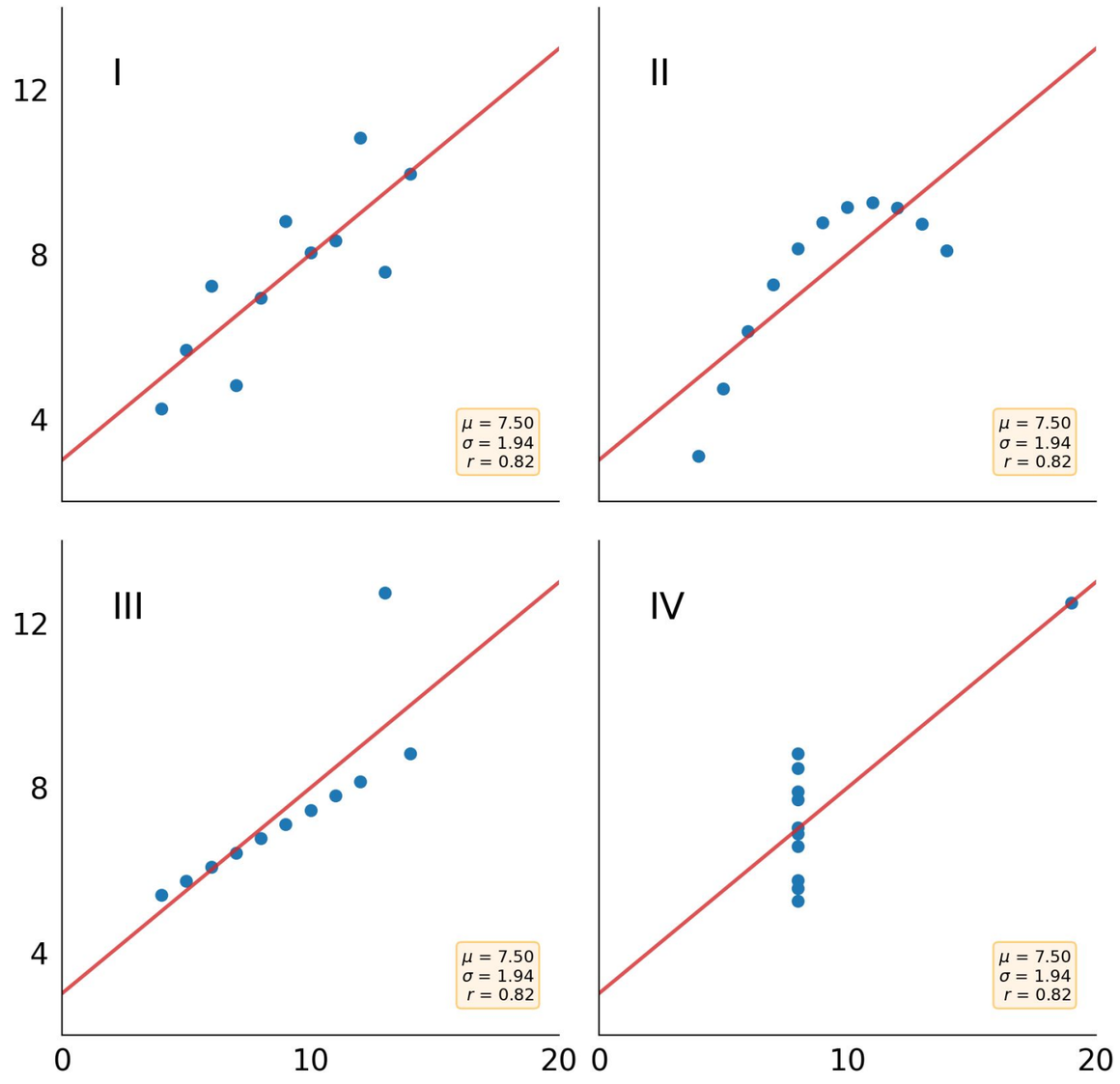
perception vs cognition: offload cognitive computation

y_1	y_2	y_3	y_4
8.04	9.14	7.46	6.58
6.95	8.14	6.77	5.76
7.58	8.74	12.74	7.71
8.81	8.77	7.11	8.84
8.33	9.26	7.81	8.47
9.96	8.10	8.84	7.04
7.24	6.13	6.08	5.25
4.26	3.10	5.39	12.50

Why should I bother?

represent <u>all</u> of the data: aggregation encompasses assumption

"Anscombe's Quartet"



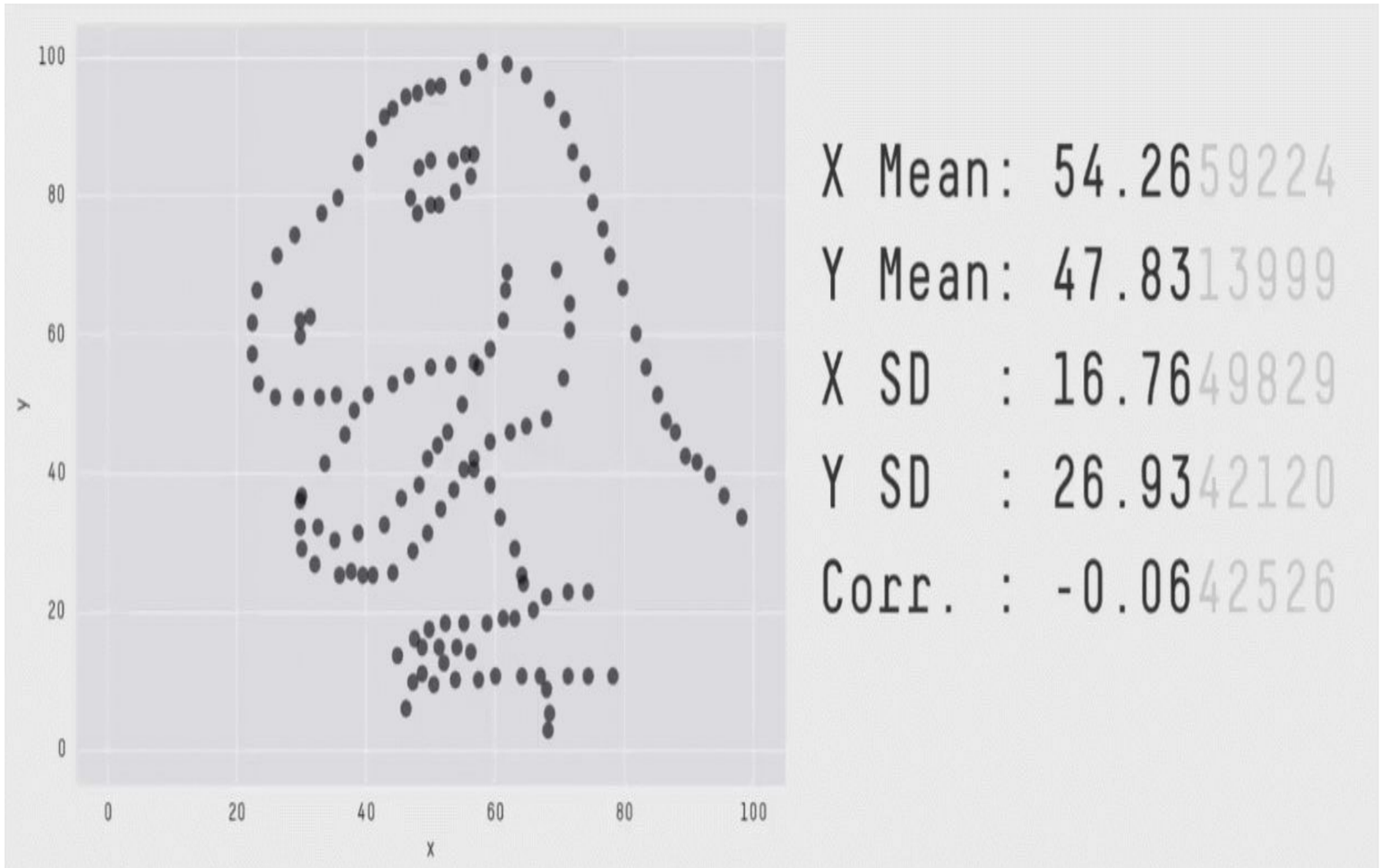
Visualization by **Cameron Riddell (Don't Use This Code)**
Adapted from John Hunter, Darren Dale, Eric Firing, Michael Droettboom and the Matplotlib development team; 2012–2023 The Matplotlib development team

DO(N'T) USE THIS C<>DE

Why should I bother?

represent <u>all</u> of the data: aggregation encompasses assumption

“Dinosaurus Dozen”



Visualization by
Matejka, Justin & Fitzmaurice, George; *Same stats, different graphs: Generating datasets with varied appearance and identical statistics through simulated annealing.*

When should I bother?

Exploratory

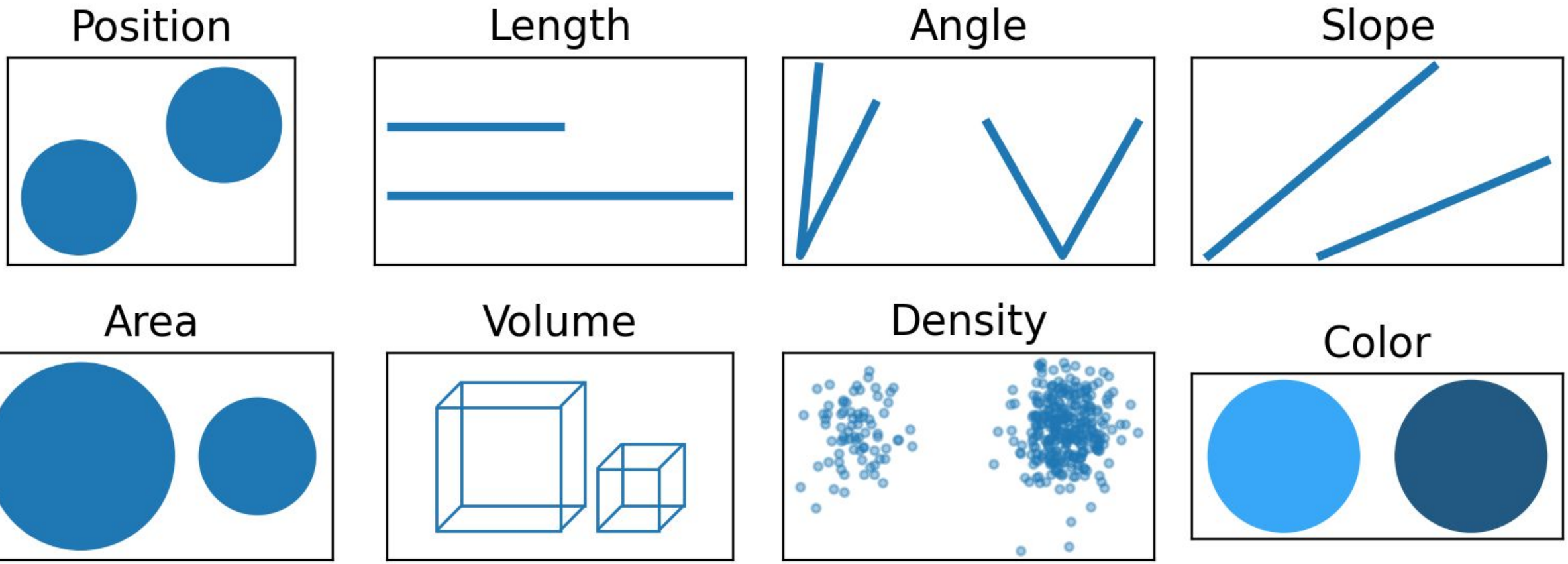
- facilitate understanding of data and relationships therein
    - data-driven explorations
- make a decision based on some data
    - has a closed-form solution been reached?
- uncover questions

Explanatory

- communicate findings of an analysis/investigation
- justify a decision or its reasoning

Visual Encodings

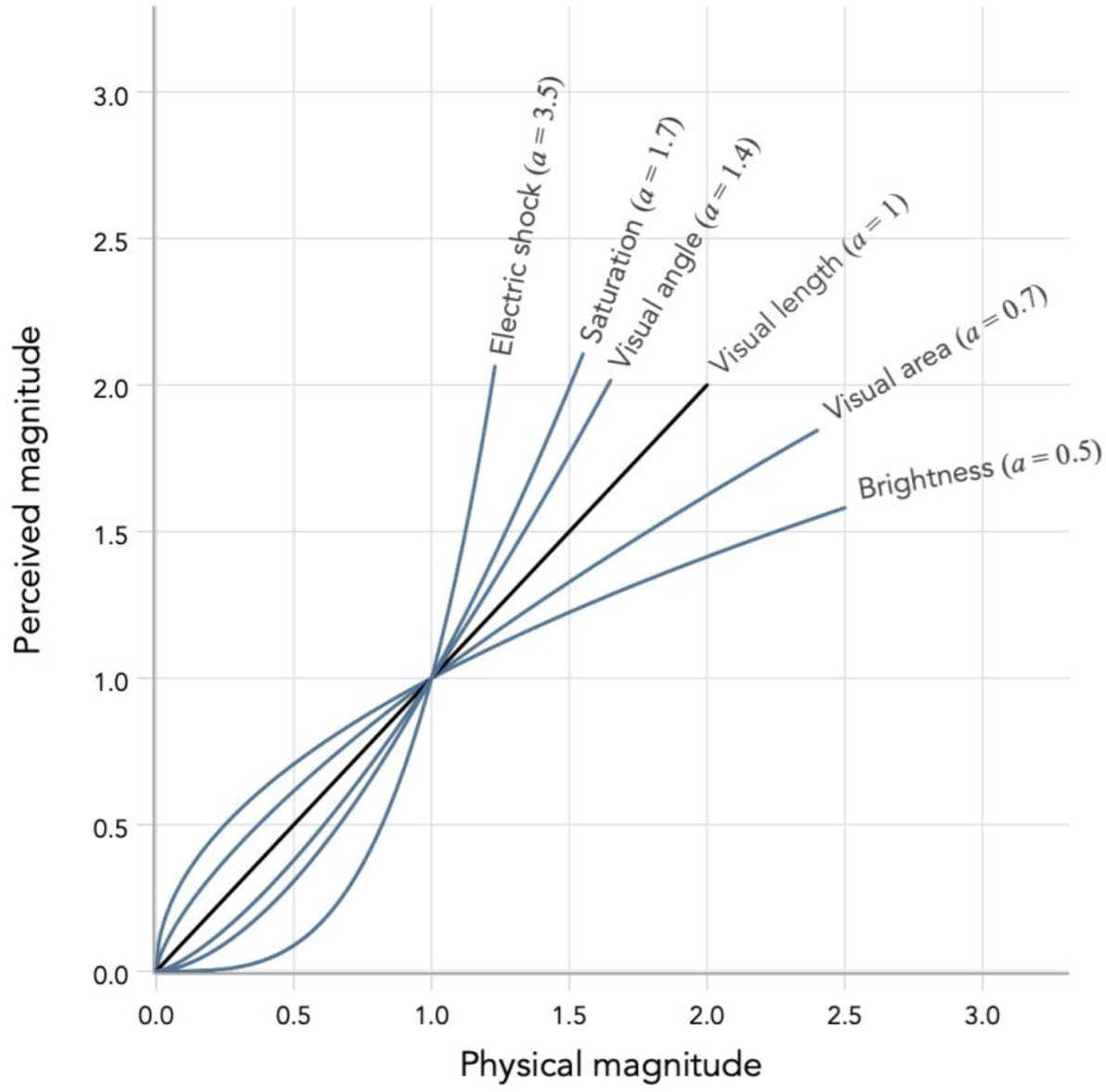
What channels do I have?



Visualization by **Cameron Riddell (Don’t Use This Code)**
Adapted from John Hunter, Darren Dale, Eric Firing, Michael Droettboom and the Matplotlib development team; 2012–2023 The Matplotlib development team

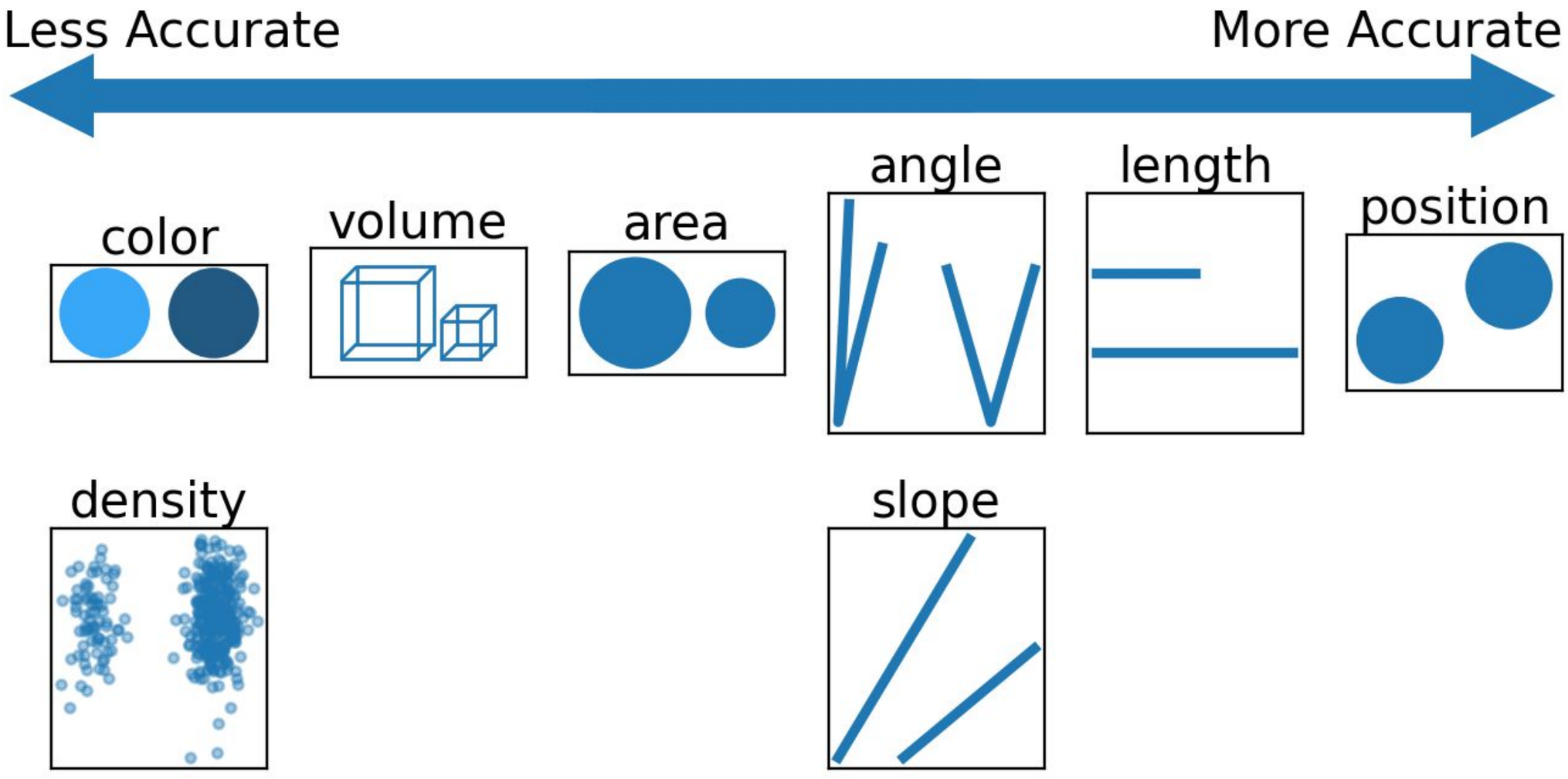
DO(N’T) USE THIS C<>DE

What is Stevens' Law?



Visualization by Demetris Christodoulou
Demetris Christodoulou, *Perceived magnitude*. Graph workflow

Which is quantitatively best?



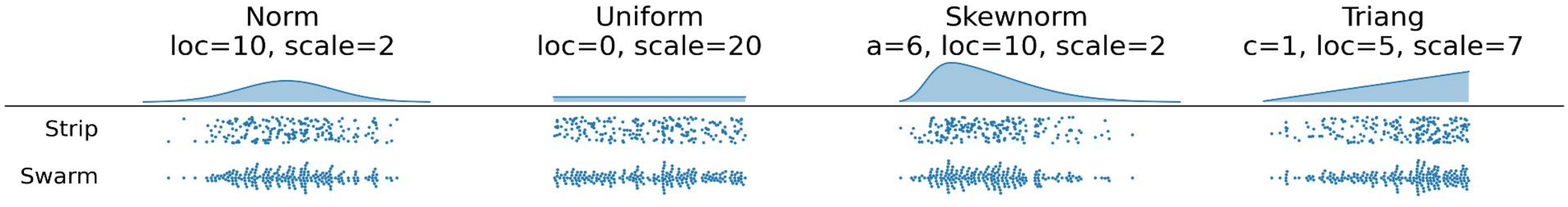
Visualization by **Cameron Riddell (Don't Use This Code)**
Adapted from *Automating the Design of Graphical Presentations Of Relational Information*, Jock Mackinlay

DO(N'T) USE THIS C<>DE

Charts

... and when to use them !

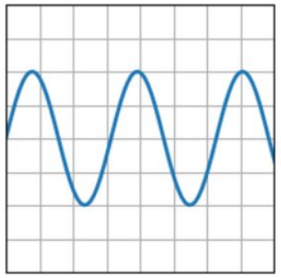
How do I show one variable?



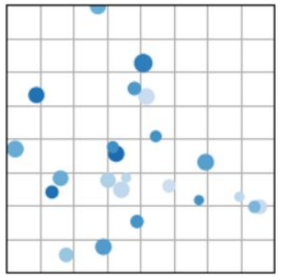
Visualization by **Cameron Riddell (Don’t Use This Code)**

DO(N’T) USE THIS C<>DE

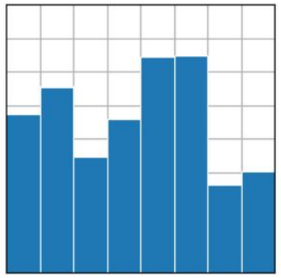
How do I show two variables?



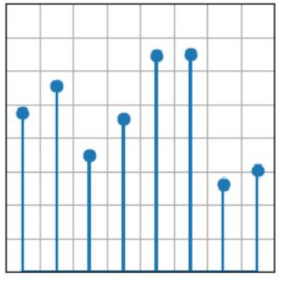
Line



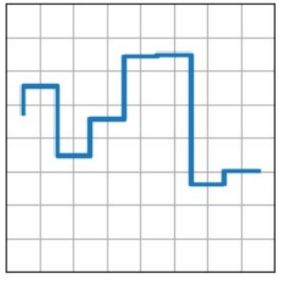
Scatter



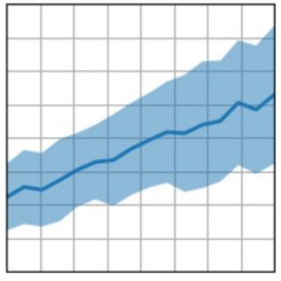
Bar



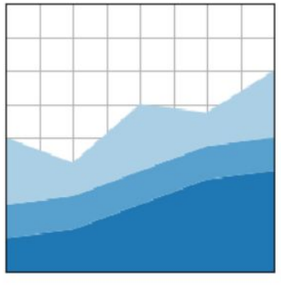
Stem



Step



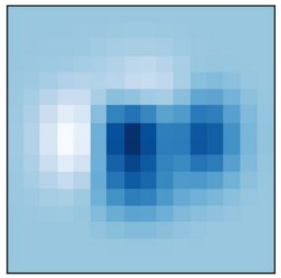
Between Area



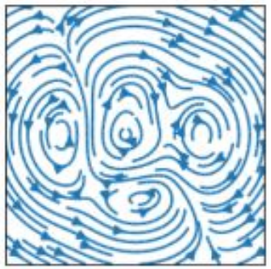
Area

Visualization by
John Hunter, Darren Dale, Eric Firing, Michael Droettboom and the Matplotlib development team; 2012–2023 The Matplotlib development team.

How do I show ≥3 variables?

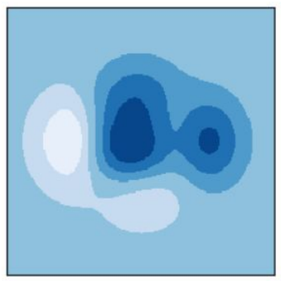


Matrix



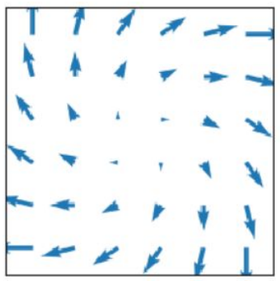
Stream

Contour



Filled Contour

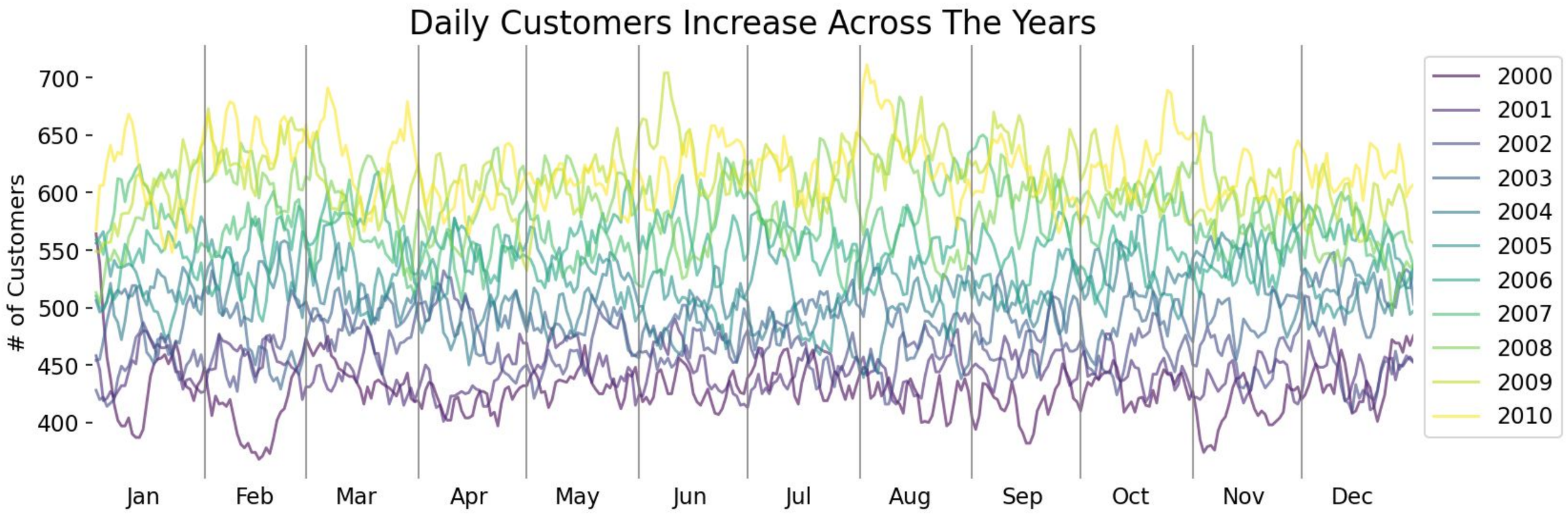
Choropleth



Quiver

Visualization by
John Hunter, Darren Dale, Eric Firing, Michael Droettboom and the Matplotlib development team; 2012–2023 The Matplotlib development team

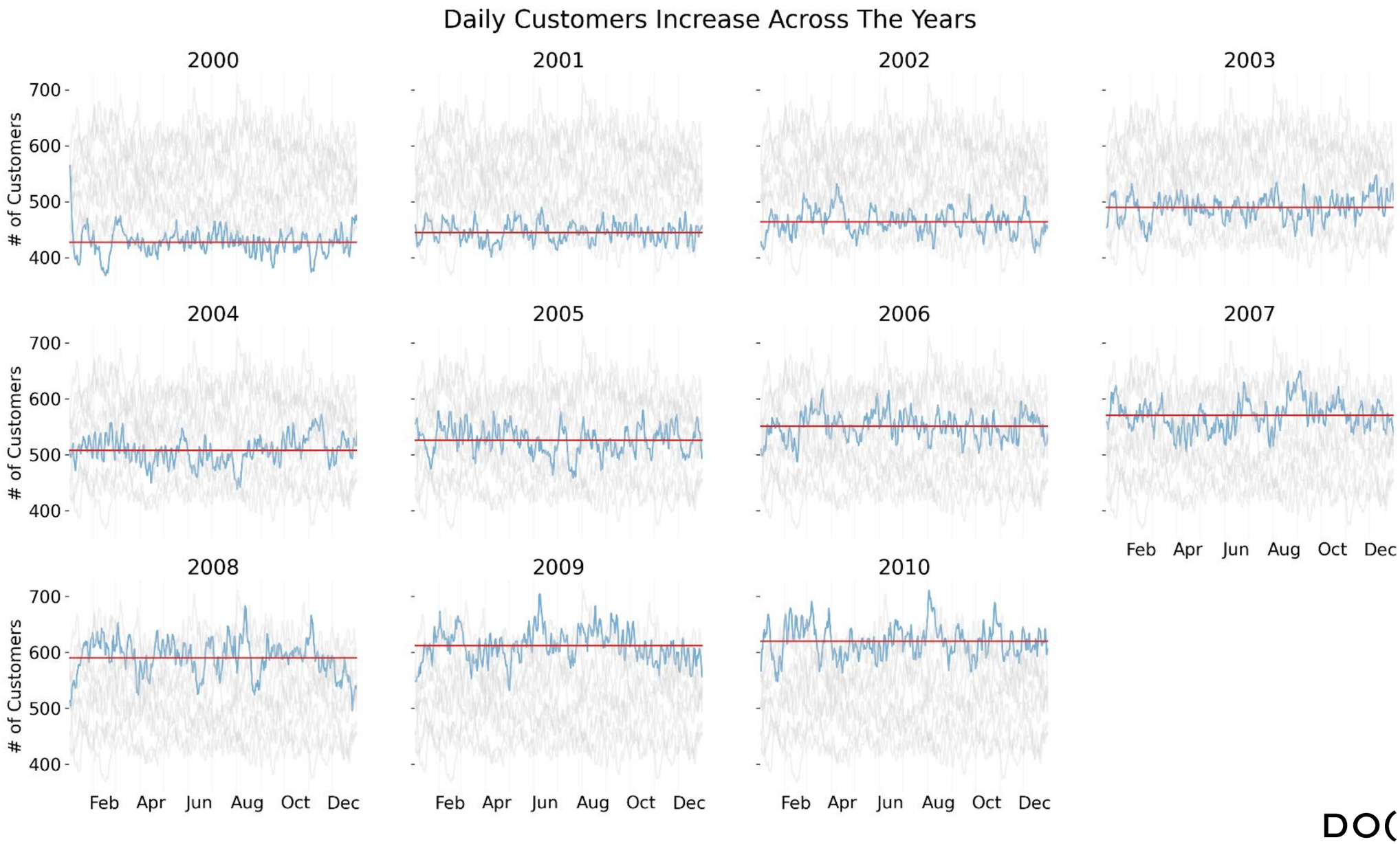
How do I show ≥3 variables?



Visualization by **Cameron Riddell (Don't Use This Code)**

DO(N'T) USE THIS C<>DE

How do I show ≥3 variables?



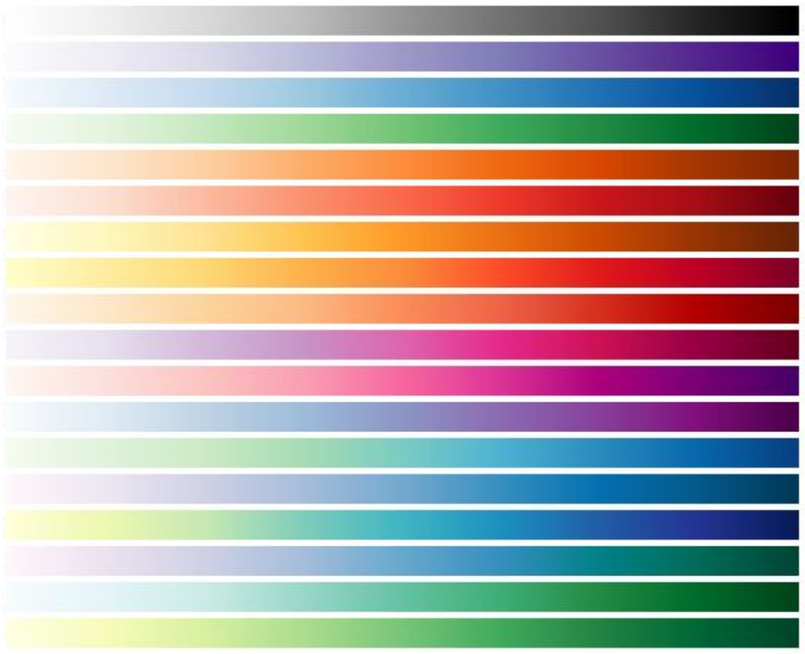
Visualization by **Cameron Riddell (Don't Use This Code)**

DO(N'T) USE THIS C<>DE

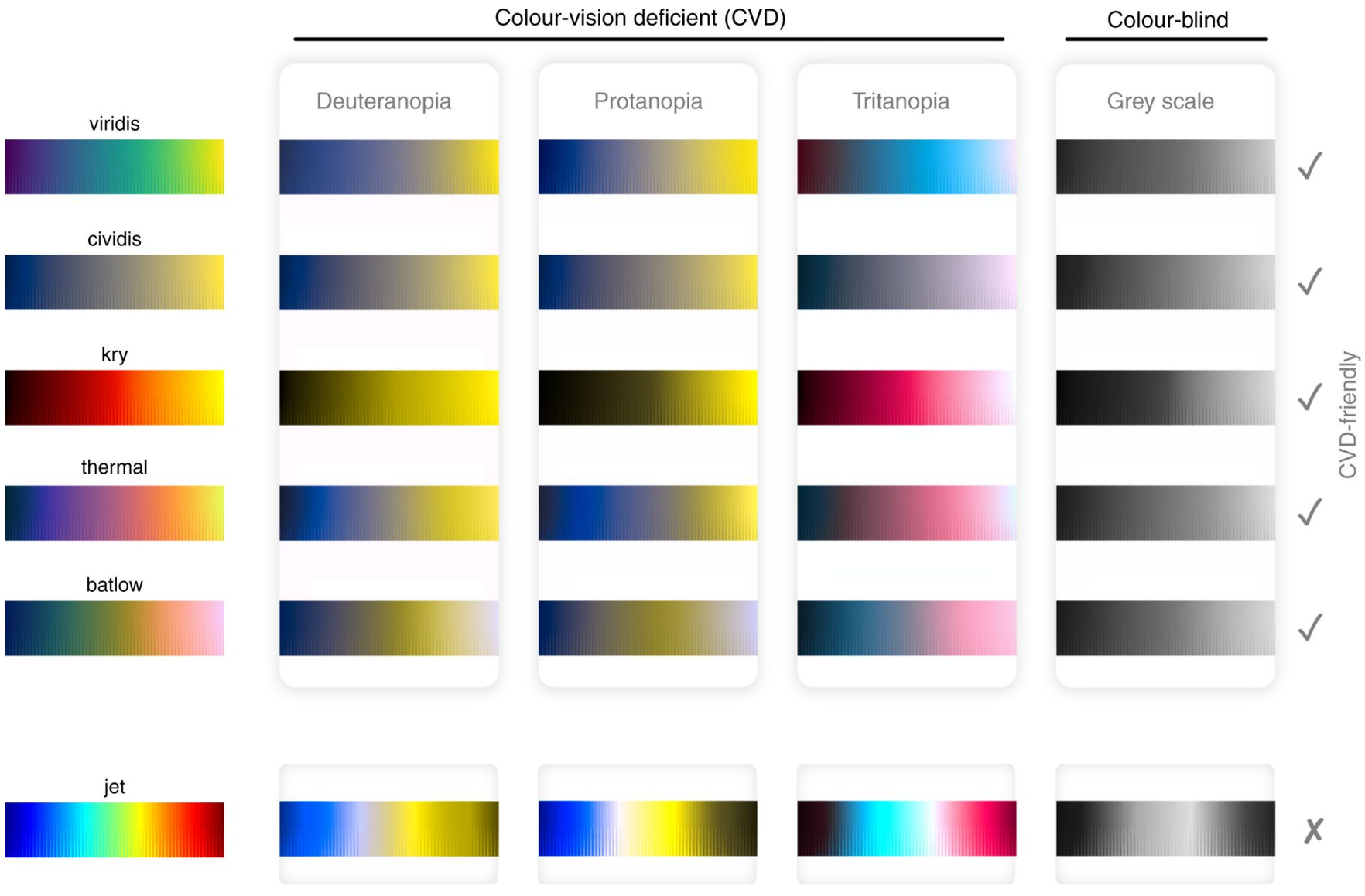
Color

What are my color choices?

Sequential

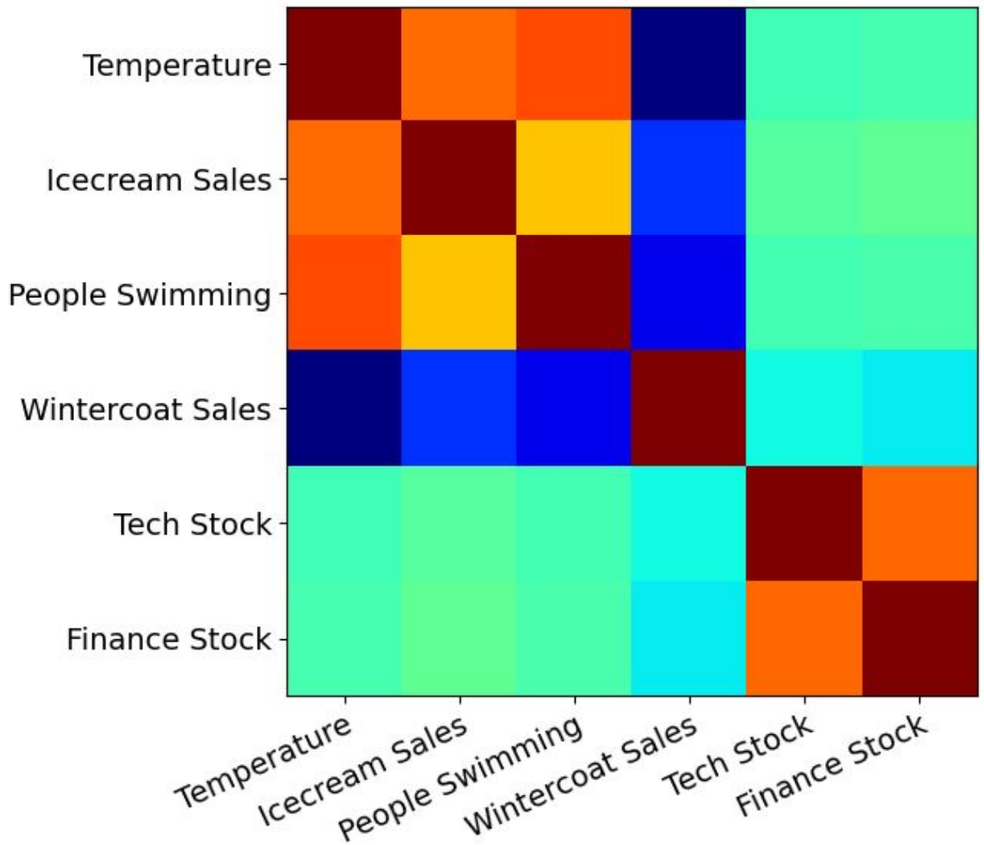


What are effective colors?



Visualization by
Crameri, F., Shephard, G.E. & Heron, P.J. The misuse of colour in science communication. Nat Commun 11, 5444 (2020)

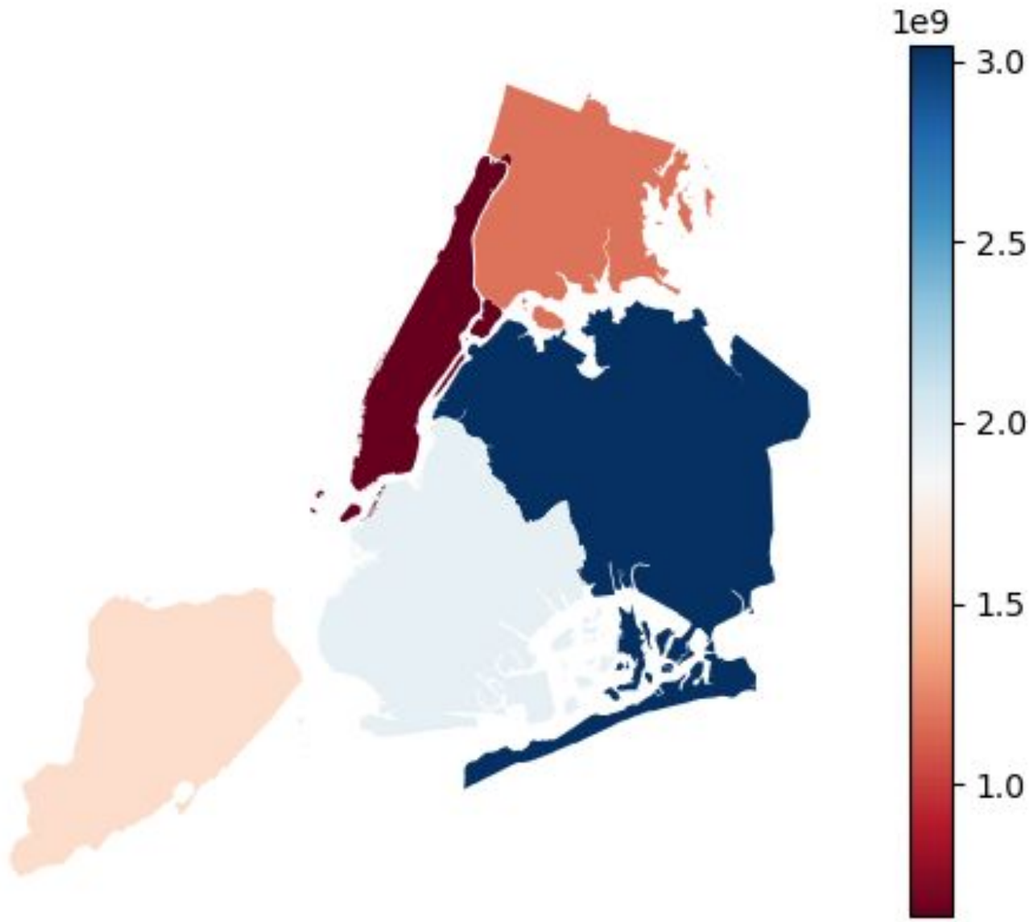
What are ineffective colors?



Visualization by **Cameron Riddell (Don't Use This Code)**

DO(N'T) USE THIS C<>DE

What are ineffective colors?



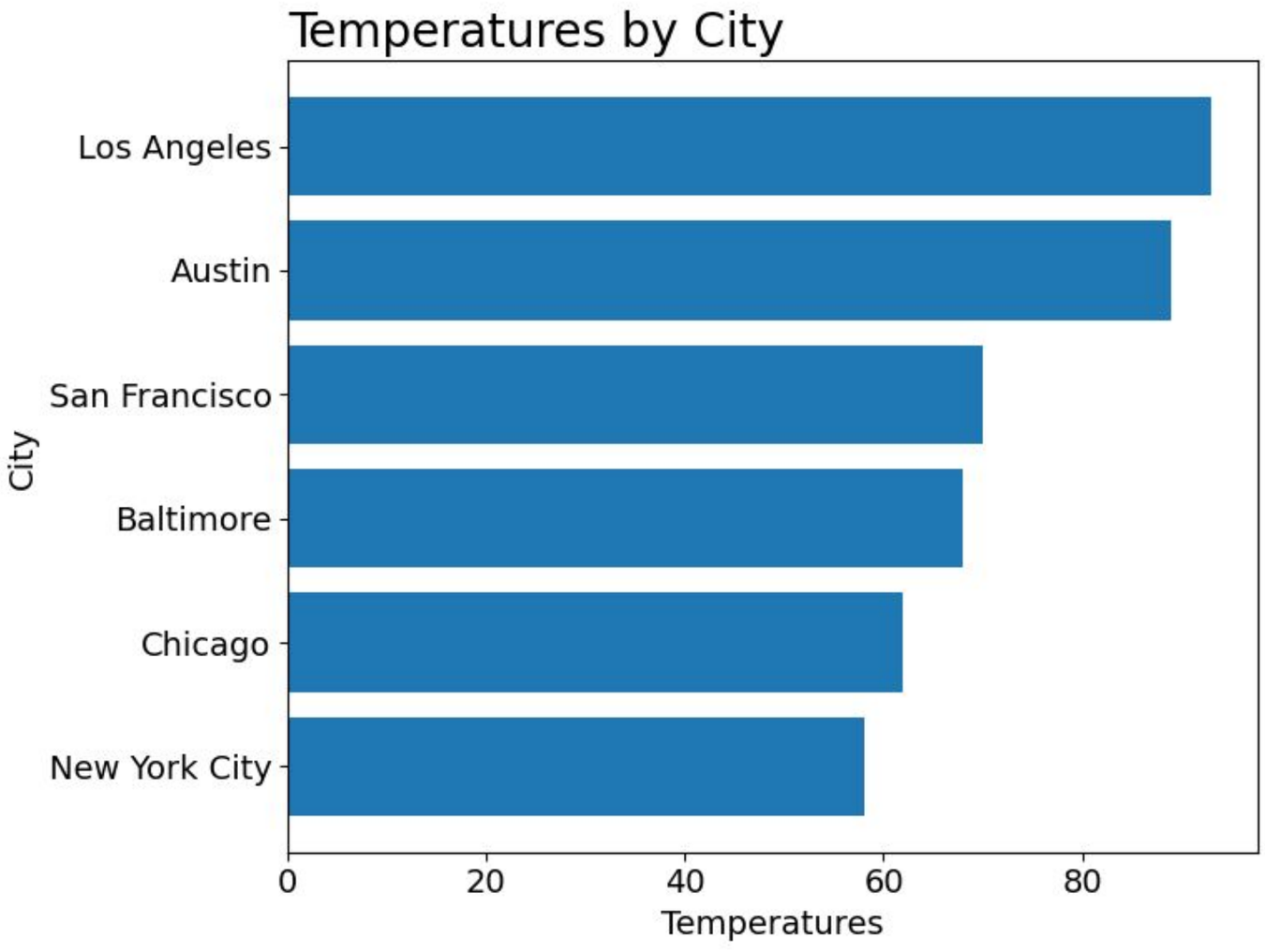
Visualization by **Cameron Riddell (Don’t Use This Code)**

DO(N’T) USE
THIS C<>DE

Visualizing the Message

... and amplifying signal!

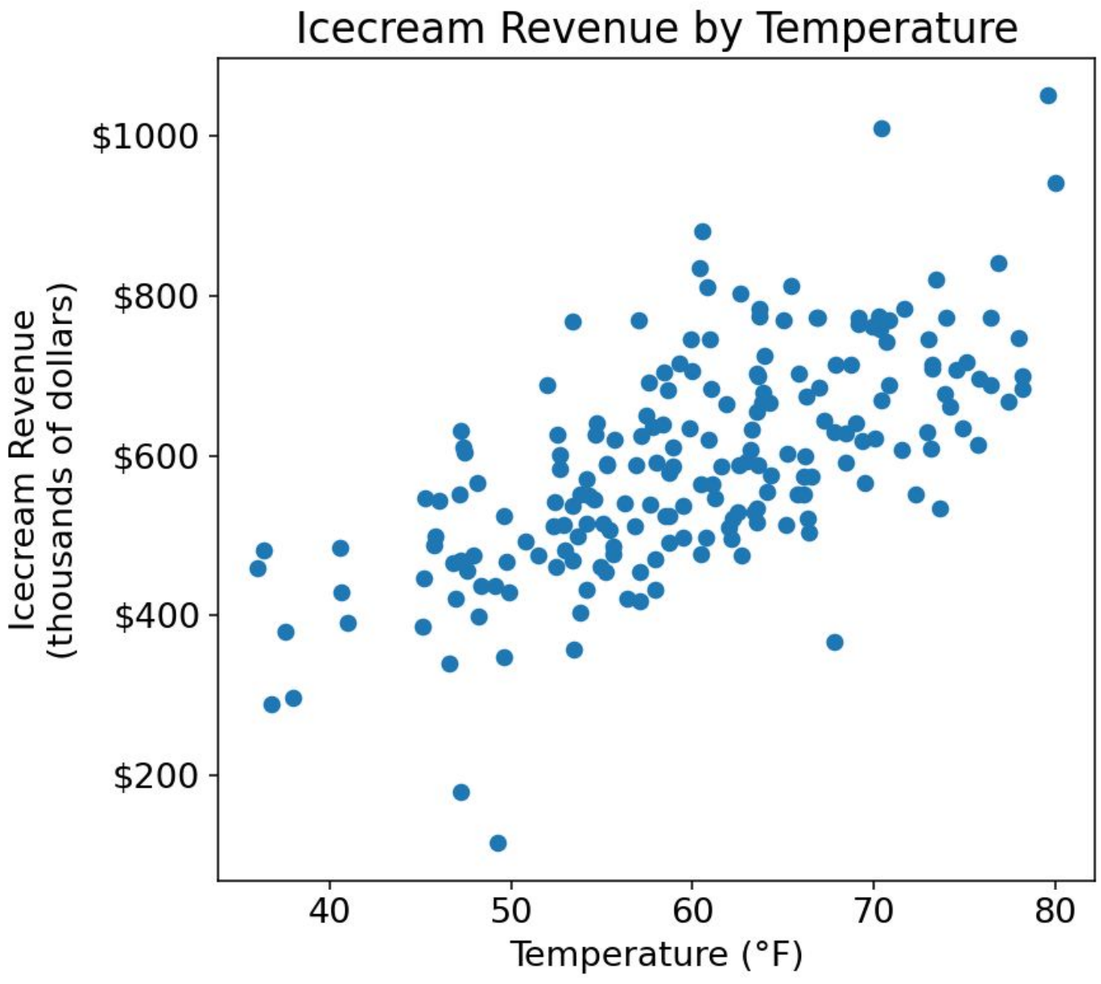
How do I put the “B.L.U.F.”?



Visualization by **Cameron Riddell (Don’t Use This Code)**

DO(N’T) USE THIS C<>DE

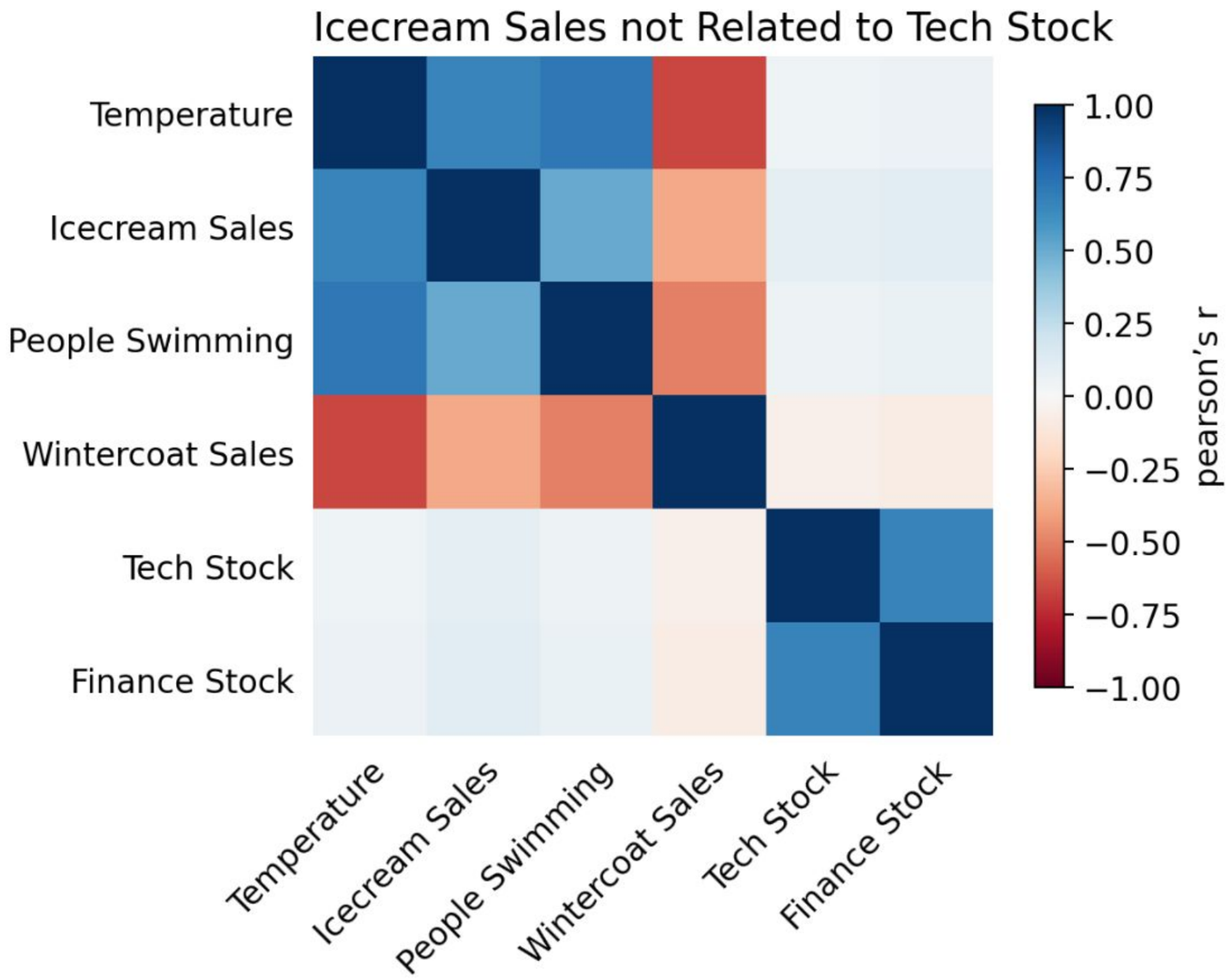
“Bottom Line Up Front”



Visualization by **Cameron Riddell (Don’t Use This Code)**

DO(N’T) USE
THIS C<>DE

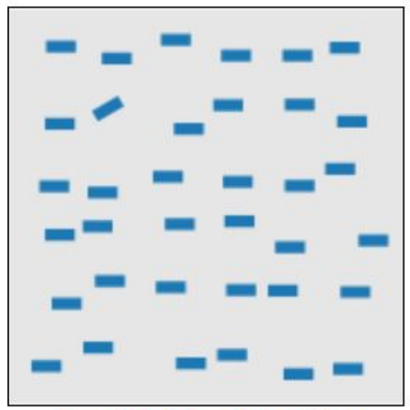
What is redundant encoding?



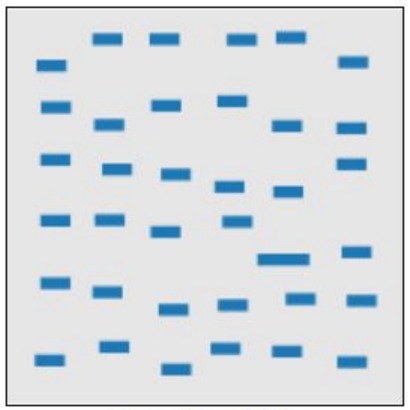
Visualization by **Cameron Riddell (Don't Use This Code)**

DO(N'T) USE THIS C<>DE

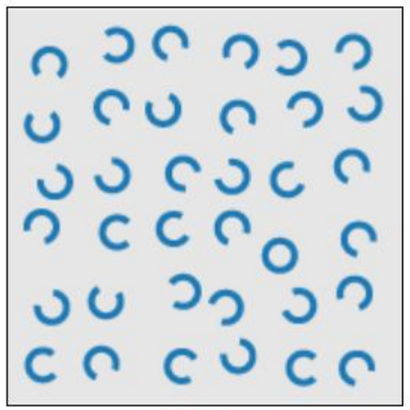
How do I draw attention?



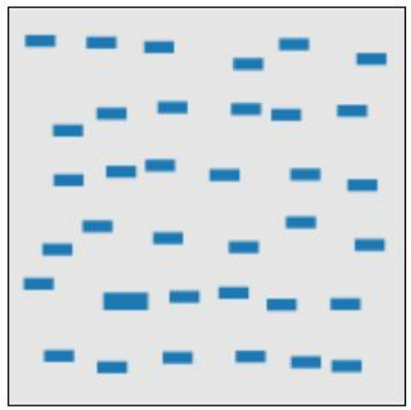
line (blob) orientation
Julész & Bergen 83; Sagi & Julész 85a, Wolfe et al. 92; Weigle et al. 2000



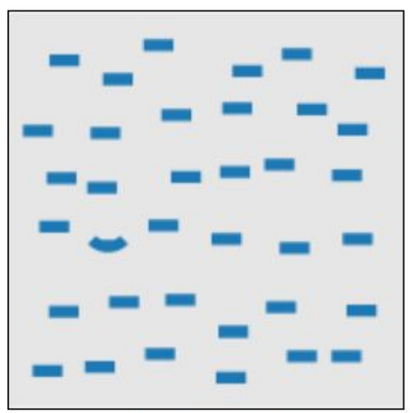
length, width
Sagi & Julész 85b; Treisman & Gormican 88



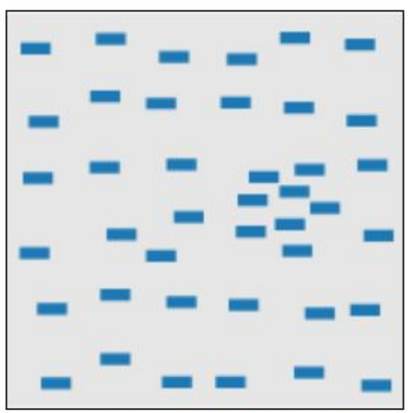
closure
Julész & Bergen 83



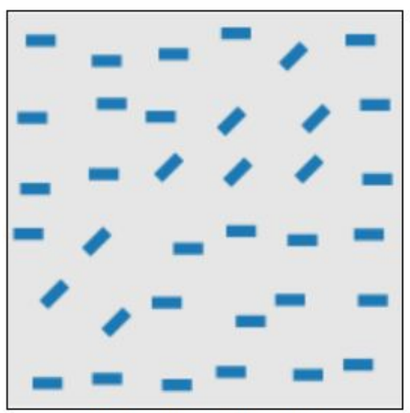
size
Treisman & Gelade 80; Healey & Enns 98; Healey & Enns 99



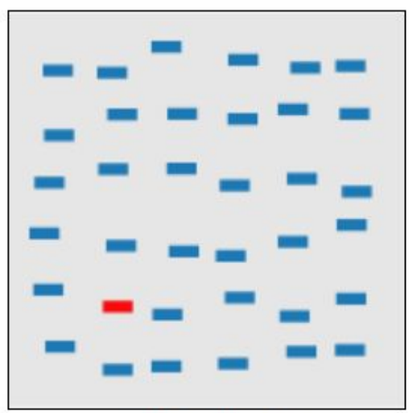
curvature
Treisman & Gormican 88



density, contrast
Healey & Enns 98; Healey & Enns 99



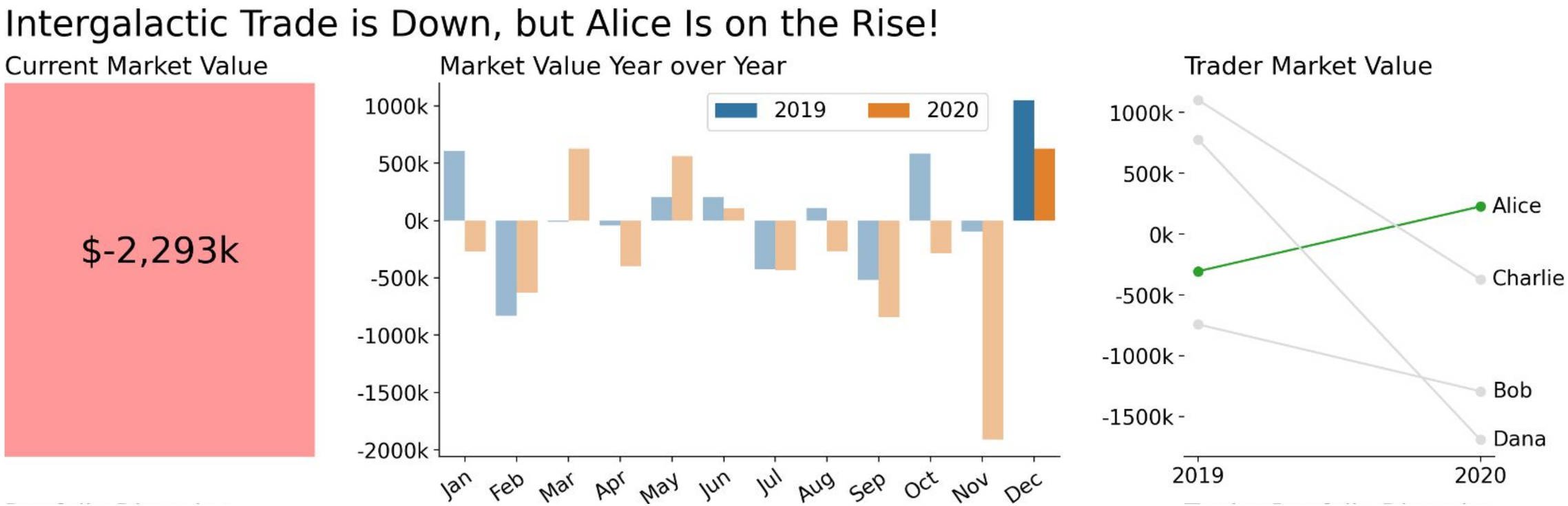
number, estimation
Sagi & Julész 85b; Healey et al. 93; Trick & Pylyshyn 94



colour (hue)
Nagy & Sanchez 90; Nagy et al. 90; D'Zmura 91; Kawai et al. 95; Bauer et al. 96; Healey 96; Bauer et al. 98; Healey & Enns 99

Visualization by
Healey, Christopher G. & Enns, James T. *Attention and Visual Memory in Visualization and Computer Graphics*. IEEE TRANSACTIONS ON VISUALIZATION AND COMPUTER GRAPHICS, VOL. 18, NO. 7, JULY 2012

How do I draw attention?



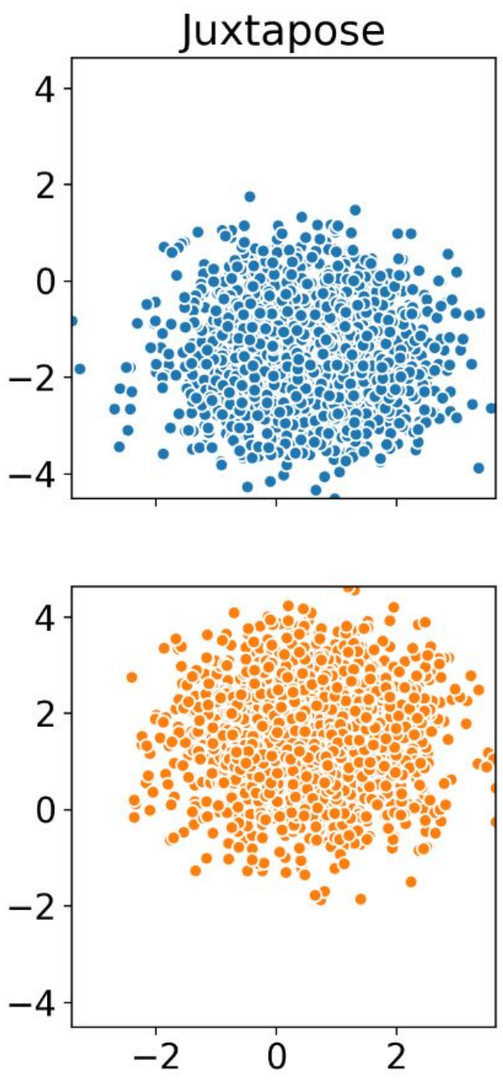
Visualization by **Cameron Riddell (Don't Use This Code)**

DO(N'T) USE THIS C<>DE

Visualizing the Message

... and reducing noise!

What is overplotting?

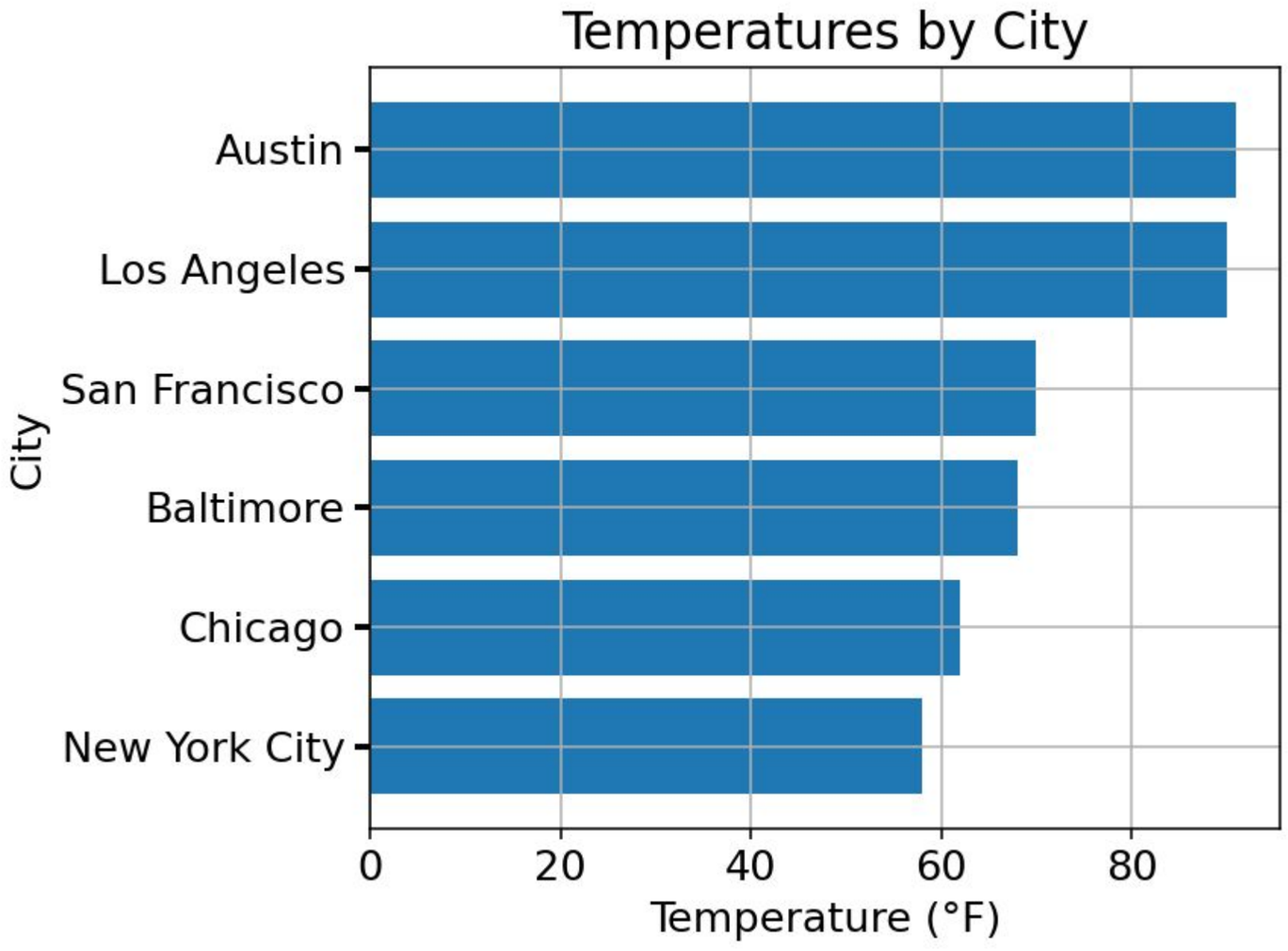


Visualization by **Cameron Riddell (Don't Use This Code)**

DO(N'T) USE
THIS C<>DE

How do I eliminate “chart junk”?

“data-to-ink ratio”



Visualization by **Cameron Riddell (Don’t Use This Code)**

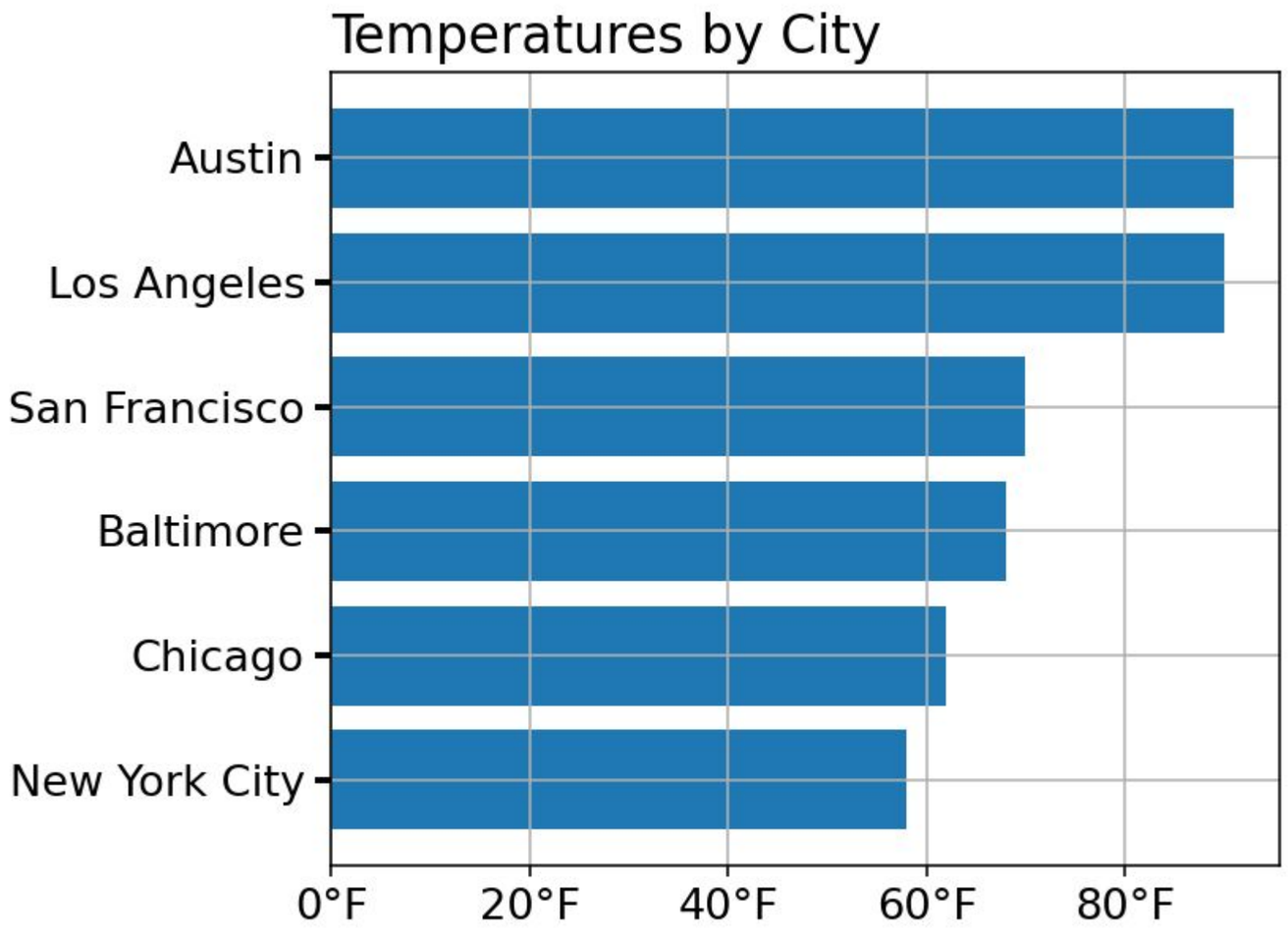
DO(N’T) USE THIS C<>DE

How do I eliminate “chart junk”?

“data-to-ink ratio”

➔ remove redundant labels

➔ left align title



Visualization by **Cameron Riddell (Don’t Use This Code)**

DO(N’T) USE THIS C<>DE

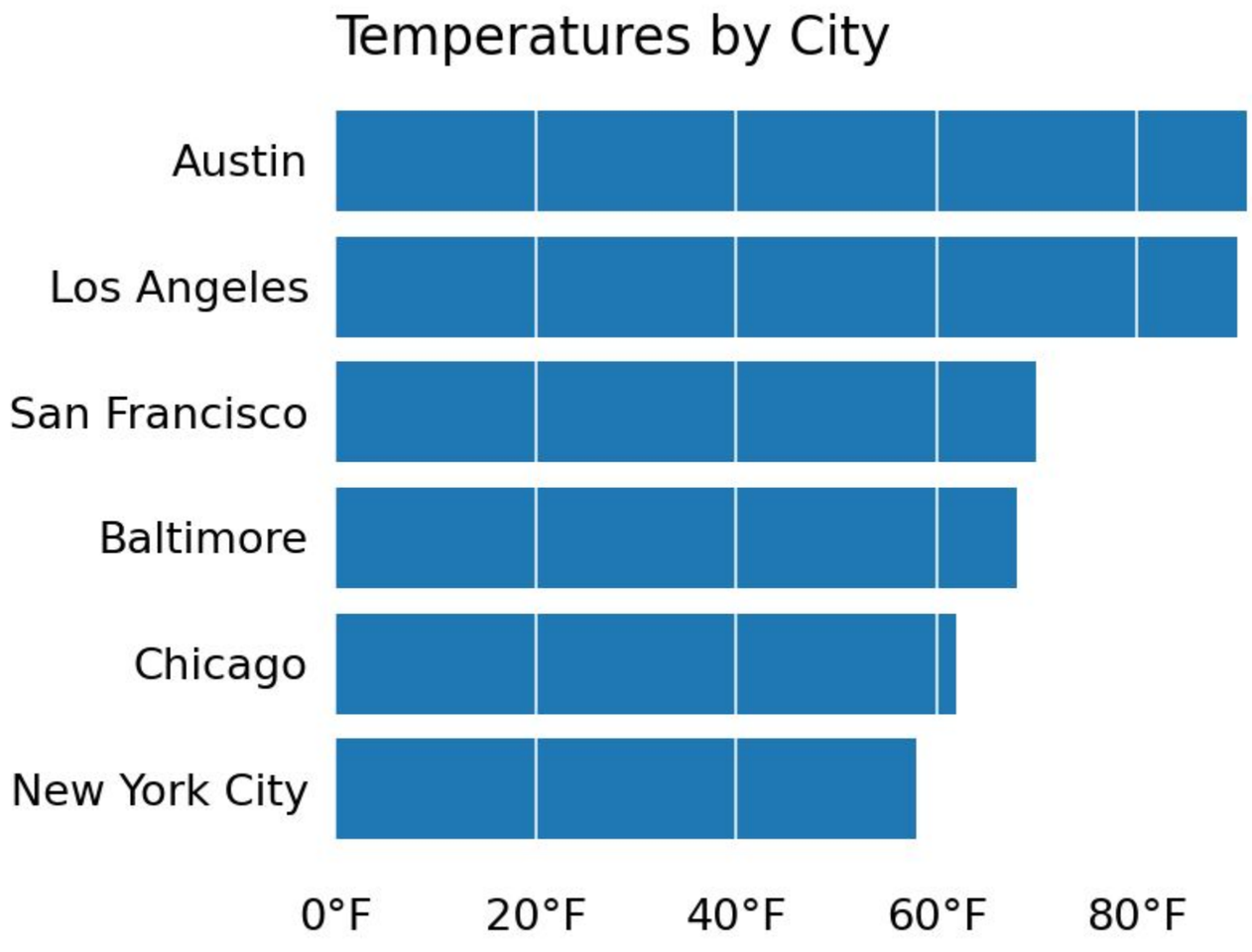
How do I eliminate “chart junk”?

“data-to-ink ratio”

➔ remove redundant labels

➔ left align title

➔ remove excess lines



Visualization by **Cameron Riddell (Don’t Use This Code)**

DO(N’T) USE THIS C<>DE

How do I eliminate “chart junk”?

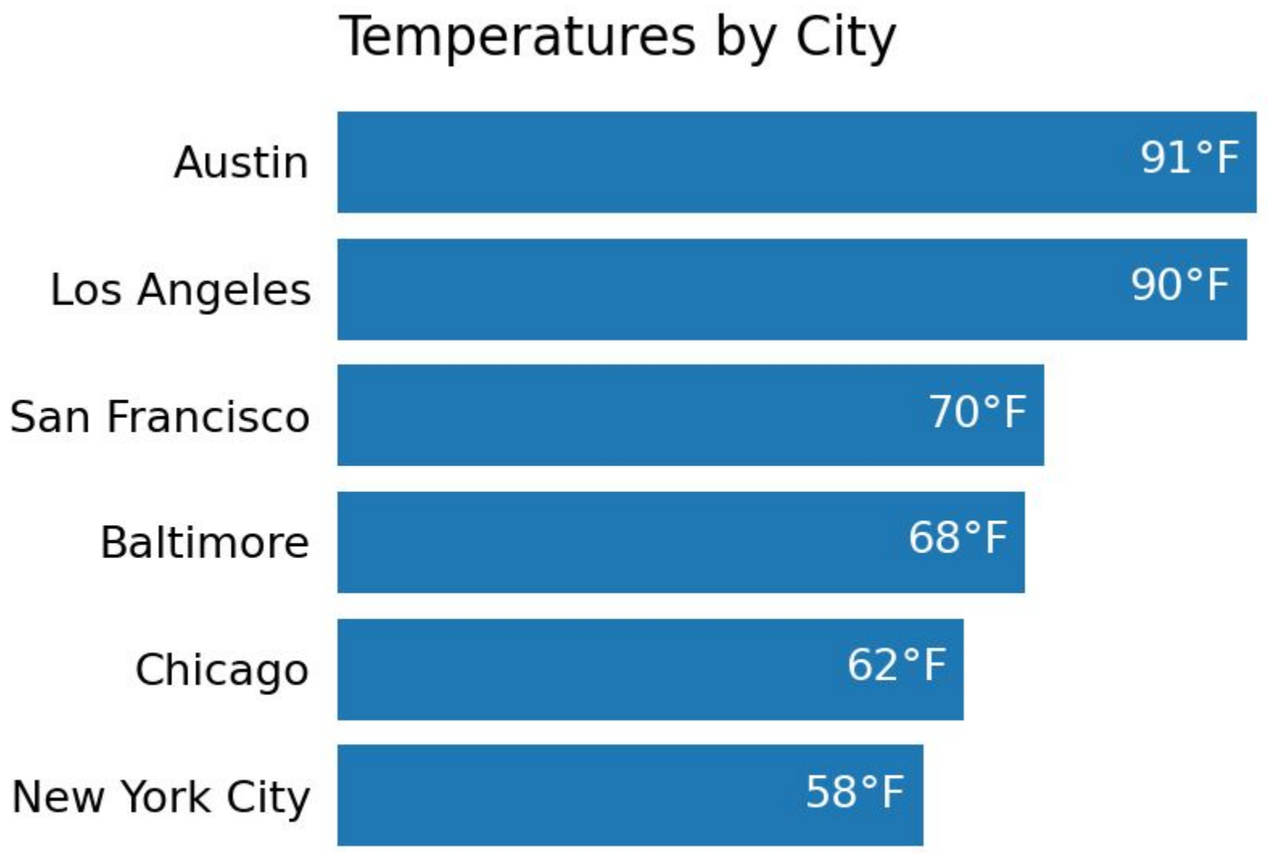
“data-to-ink ratio”

➔ remove redundant labels

➔ left align title

➔ remove excess lines

➔ inline limited data



Visualization by **Cameron Riddell (Don’t Use This Code)**

DO(N’T) USE THIS C<>DE

How do I *explore* better?

Exploratory

- ❖ conveys descriptive information, while faithfully representing the data
- ❖ facilitates the exploration of relationships
  - ➢ meaningful interactions
- ❖ don’t overdo dimensionality
  - ➢ superimposition vs juxtaposition

How do I *explain* better?

Explanatory (“signal-to-noise ratio”)

message ⇌ audience ⇒ layout ⇒ metric ⇒ encodings

amplify signal

- bottom line up front
- redundant encoding
- guide attention
- appropriate metric & encoding

minimize noise

- avoid overplotting
- appropriate filtering/aggregation
- remove chart junk