Storytelling with Data

Emorie D. Beck, Ph.D
University of California, Davis
@EmorieBeck

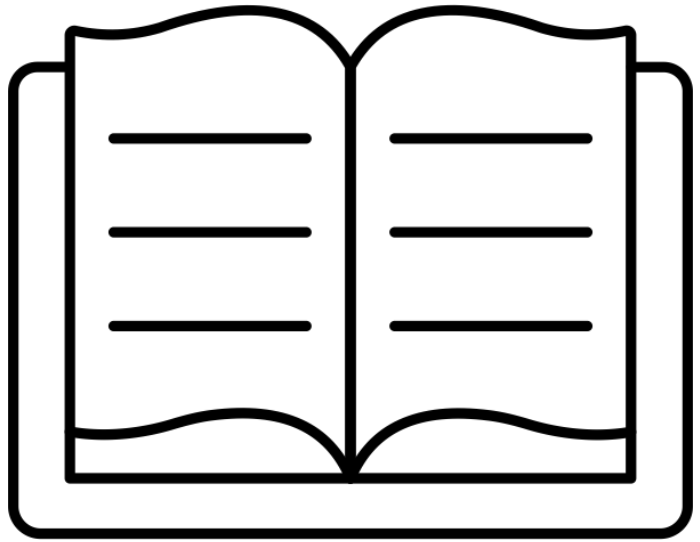
UC DAVIS
UNIVERSITY OF CALIFORNIA

What Is Data Visualization?

Data visualization is the graphical representation of data and information.

What Are the Goals of Data Visualization?

- **Summarize complex information**
- **Reveal difficult to detect trends and patterns in data**
- **Tell a story with data**

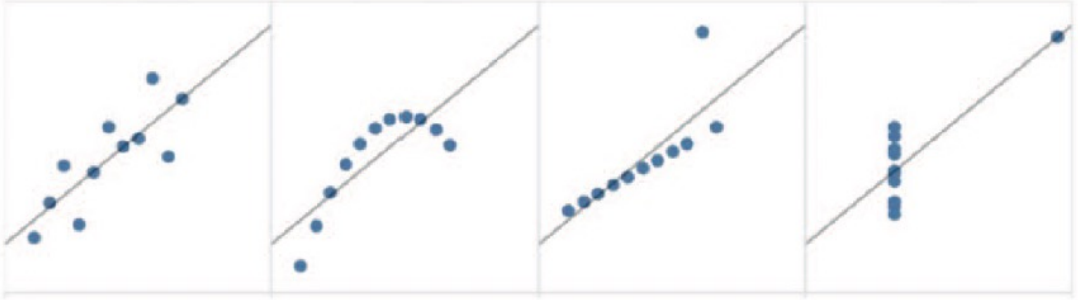


What Are the Goals of Data Visualization?

	x	y	x	y	x	y	x	y
	10	8.04	10	9.14	10	7.46	8	6.58
	8	6.95	8	8.14	8	6.77	8	5.76
	13	7.58	13	8.74	13	12.74	8	7.71
	9	8.81	9	8.77	9	7.11	8	8.84
	11	8.33	11	9.26	11	7.81	8	8.47
	14	9.96	14	8.10	14	8.84	8	7.04
	6	7.24	6	6.13	6	6.08	8	5.25
	4	4.26	4	3.10	4	5.39	19	12.50
	12	10.84	12	9.13	12	8.15	8	5.56
	7	4.82	7	7.26	7	6.42	8	7.91
	5	5.68	5	4.74	5	5.73	8	6.89
Mean	9.00	7.50	9.00	7.50	9.00	7.50	9.00	7.50
Stdev	3.32	2.03	3.32	2.03	3.32	2.03	3.32	2.03
	R	0.82	**R**	0.82	**R**	0.82	**R**	0.82

What Are the Goals of Data Visualization?

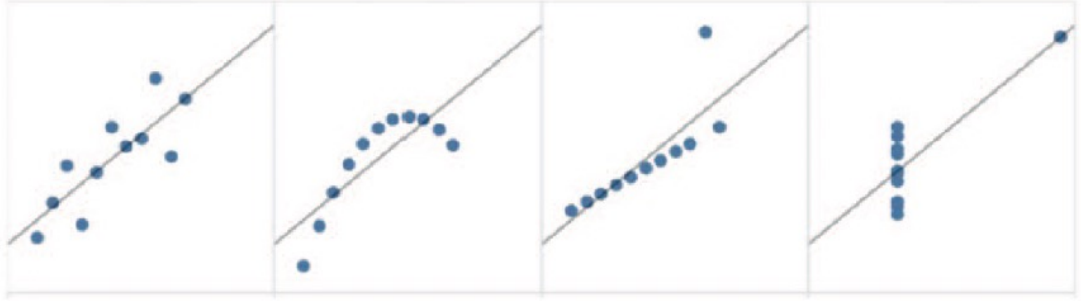
x	y	x	y	x	y	x	y
10	8.04	10	9.14	10	7.46	8	6.58
8	6.95	8	8.14	8	6.77	8	5.76
13	7.58	13	8.74	13	12.74	8	7.71
9	8.81	9	8.77	9	7.11	8	8.84
11	8.33	11	9.26	11	7.81	8	8.47
14	9.96	14	8.10	14	8.84	8	7.04
6	7.24	6	6.13	6	6.08	8	5.25
4	4.26	4	3.10	4	5.39	19	12.50
12	10.84	12	9.13	12	8.15	8	5.56
7	4.82	7	7.26	7	6.42	8	7.91
5	5.68	5	4.74	5	5.73	8	6.89
Mean	9.00	7.50	9.00	7.50	9.00	7.50	9.00	7.50
Stdev	3.32	2.03	3.32	2.03	3.32	2.03	3.32	2.03
	R	0.82	**R**	0.82	**R**	0.82	**R**	0.82



What Are the Goals of Data Visualization?

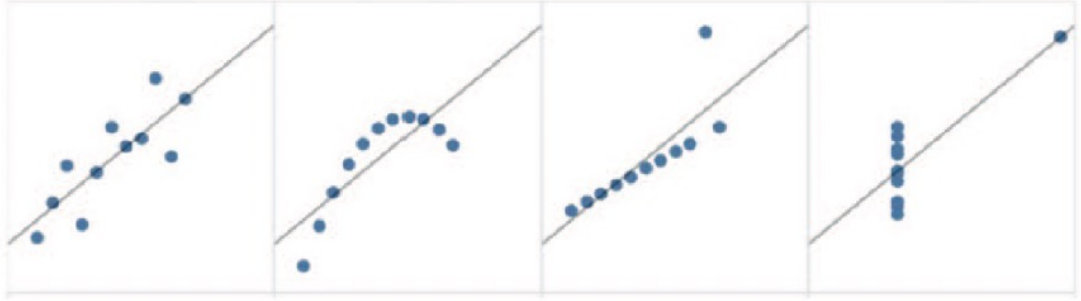
x	y	x	y	x	y	x	y
10	8.04	10	9.14	10	7.46	8	6.58
8	6.95	8	8.14	8	6.77	8	5.76
13	7.58	13	8.74	13	12.74	8	7.71
9	8.81	9	8.77	9	7.11	8	8.84
11	8.33	11	9.26	11	7.81	8	8.47
14	9.96	14	8.10	14	8.84	8	7.04
6	7.24	6	6.13	6	6.08	8	5.25
4	4.26	4	3.10	4	5.39	19	12.50
12	10.84	12	9.13	12	8.15	8	5.56
7	4.82	7	7.26	7	6.42	8	7.91
5	5.68	5	4.74	5	5.73	8	6.89
Mean	9.00	7.50	9.00	7.50	9.00	7.50	9.00	7.50
Stdev	3.32	2.03	3.32	2.03	3.32	2.03	3.32	2.03
	R	0.82	**R**	0.82	**R**	0.82	**R**	0.82

X Mean : 54.26
Y Mean : 47.83
X SD : 16.76
Y SD : 26.93
Corr. : -0.06

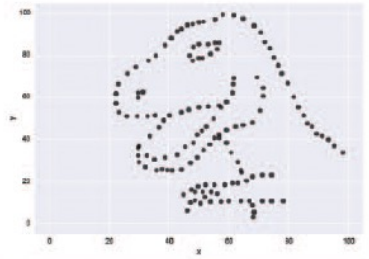


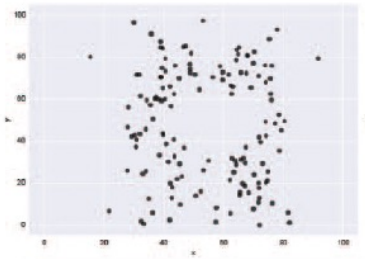
What Are the Goals of Data Visualization?

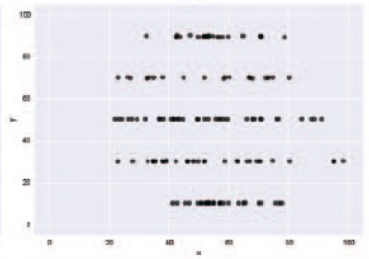
x	y	x	y	x	y	x	y
10	8.04	10	9.14	10	7.46	8	6.58
8	6.95	8	8.14	8	6.77	8	5.76
13	7.58	13	8.74	13	12.74	8	7.71
9	8.81	9	8.77	9	7.11	8	8.84
11	8.33	11	9.26	11	7.81	8	8.47
14	9.96	14	8.10	14	8.84	8	7.04
6	7.24	6	6.13	6	6.08	8	5.25
4	4.26	4	3.10	4	5.39	19	12.50
12	10.84	12	9.13	12	8.15	8	5.56
7	4.82	7	7.26	7	6.42	8	7.91
5	5.68	5	4.74	5	5.73	8	6.89
Mean	9.00	7.50	9.00	7.50	9.00	7.50	9.00	7.50
Stdev	3.32	2.03	3.32	2.03	3.32	2.03	3.32	2.03
	R	0.82	**R**	0.82	**R**	0.82	**R**	0.82

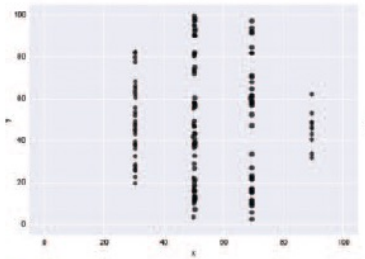


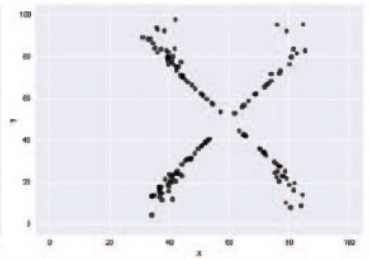
X Mean : 54.26
Y Mean : 47.83
X SD : 16.76
Y SD : 26.93
Corr. : -0.06

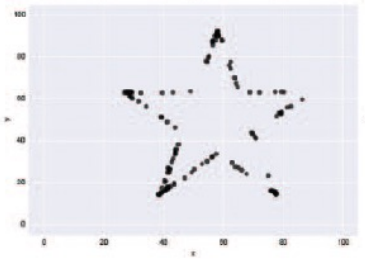


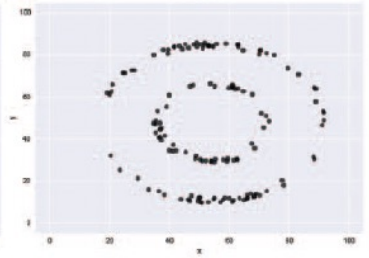




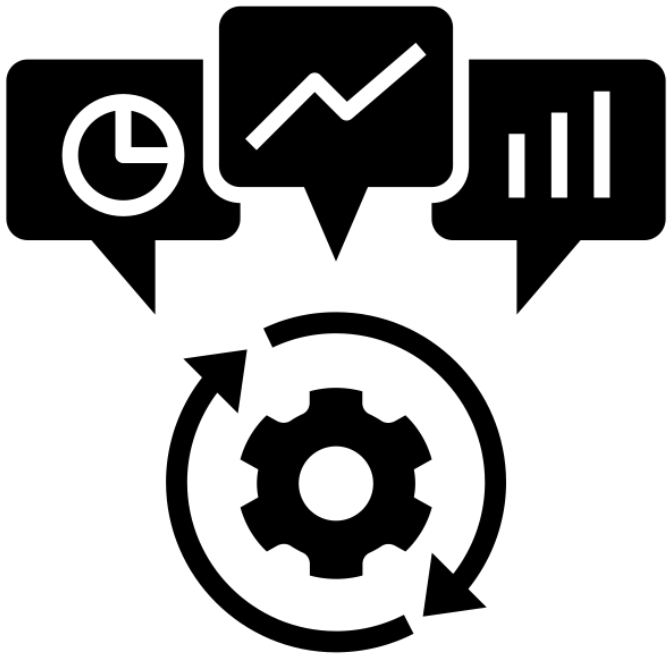








Why Should I Care About Data Visualization?



- Helps to clarify complex ideas
- Requires you think through your own ideas and data
- Data visualization is a skill and a rapidly evolving field / tool itself

Why Should I Care About Data Visualization?

https://www.youtube.com/watch?v=jbkSRLYSojo

Activity #1

Find one example of a data visualization you think is a good and another that you think is not (take your time and look for 5-10 min). Then, share your visualization with a partner and write down what makes each good and/or bad (~5-10 min).

Part 1:
Principles of Good Visualizations

Four Key Questions for Building Good Data Visualizations

What story are you / your data trying to tell?

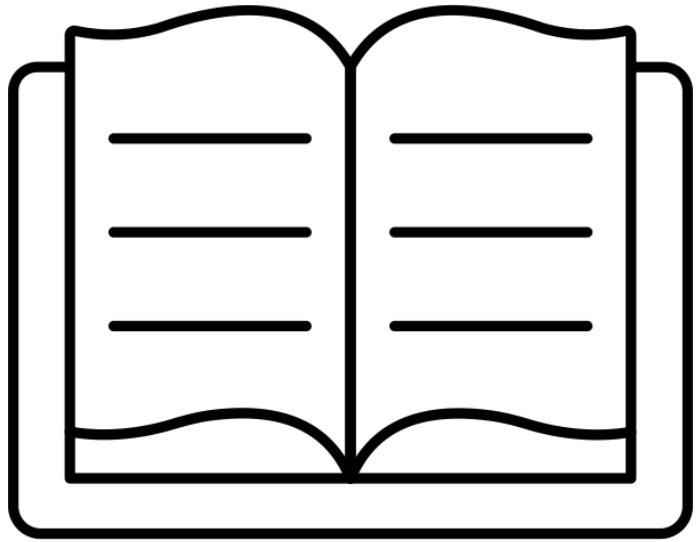
What type of data visualization will most simply communicate your story?

Who is your audience?

What type of data are you trying to summarize?

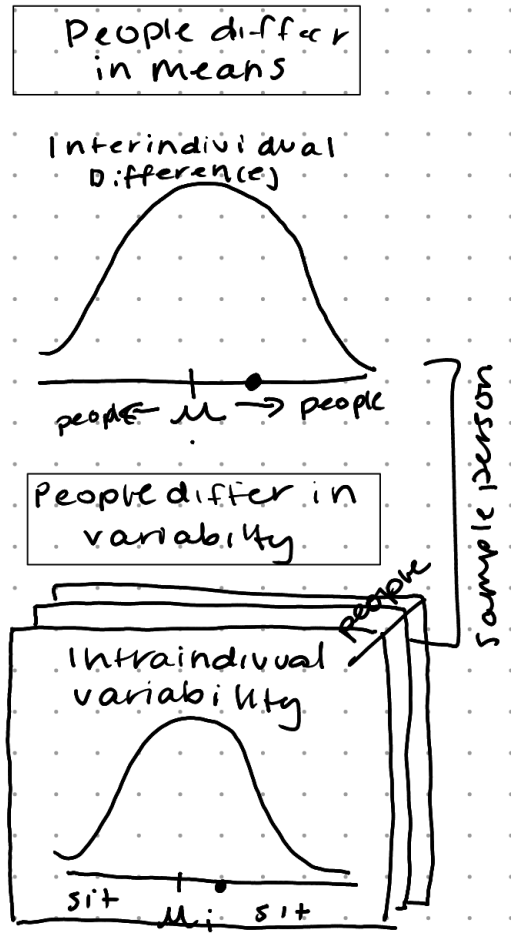
What story are you / your data trying to tell?

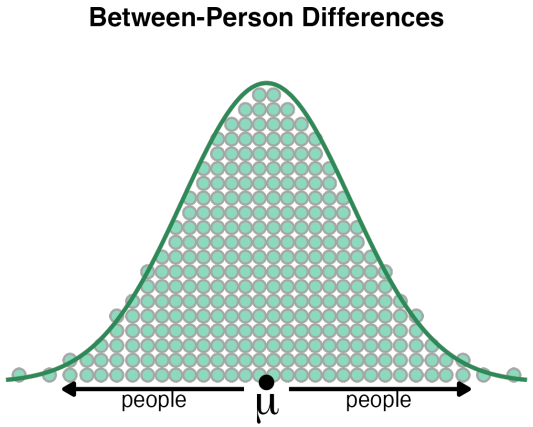
- **Start at the end: what do you want your audience to walk away knowing?**
- **Your visualization should be a journey to that**

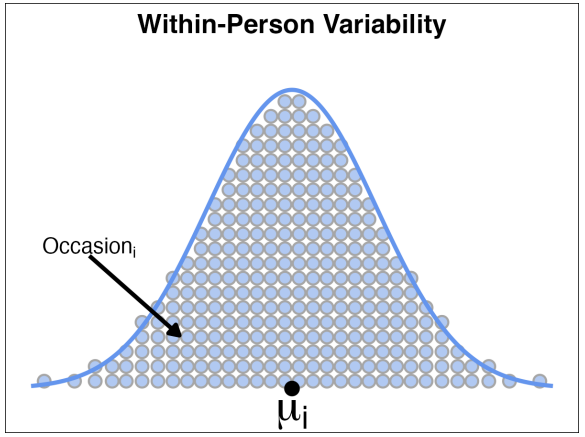


What story are you / your data trying to tell?

- **Tip: start on paper or tablet and draw your visualization**
- **The story evolves, so should your visualizations**



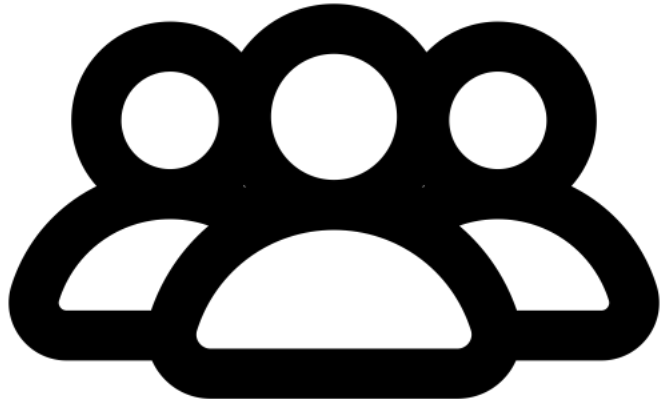




Who is your audience?

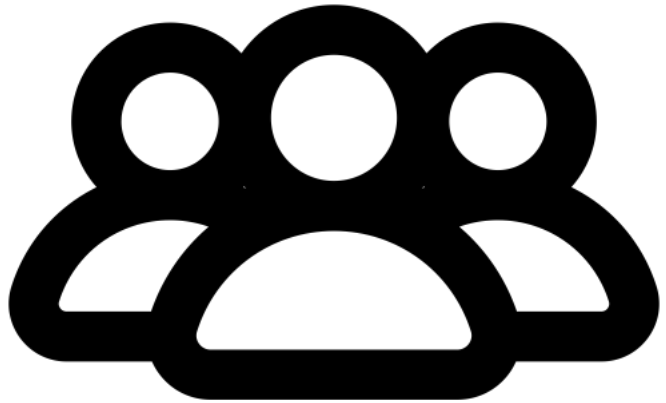
- **The story you tell should depend on your audience**

- **Ask yourself: What does your audience know? What do you want your audience to know? What steps connect their current knowledge to that?**



Who is your audience?

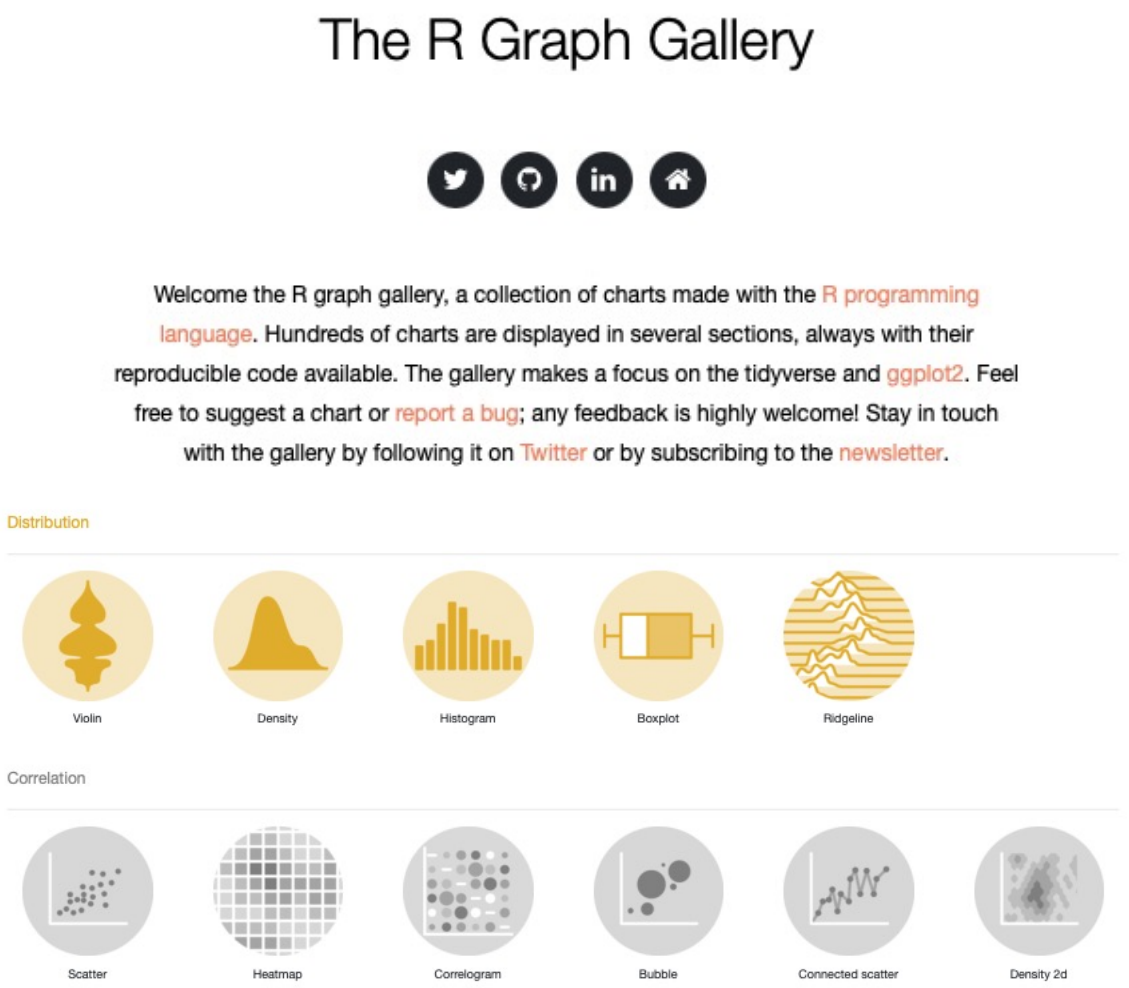
- Avoid jargon
- Use clear titles
- Avoid visual clutter
- Use color effectively, not liberally



What type of data visualization will most simply communicate your story?

- **Data visualization requires creativity, but it's also a knowledge-based skill**

- **Use online resources (e.g., https://r-graph-gallery.com/) to browse types of visualization**



What type of data are you trying to summarize?

Data Type

Visualization Type

Activity #2

Go to https://r-graph-gallery.com. Browse some of the types of data visualizations. Choose two and write down what kinds of data may be most appropriate for different visualizations (~5-10 min).

The Cognitive Psychology of Data Visualization

- **There is a whole field of researchers who study how we perceive data visualizations most efficiently and accurately!**

- **This research draws on Gestalt Principles and Cognitive Psychology to improve visualizations**

The Cognitive Psychology of Data Visualization

6 Common Types of Visual Aesthetics

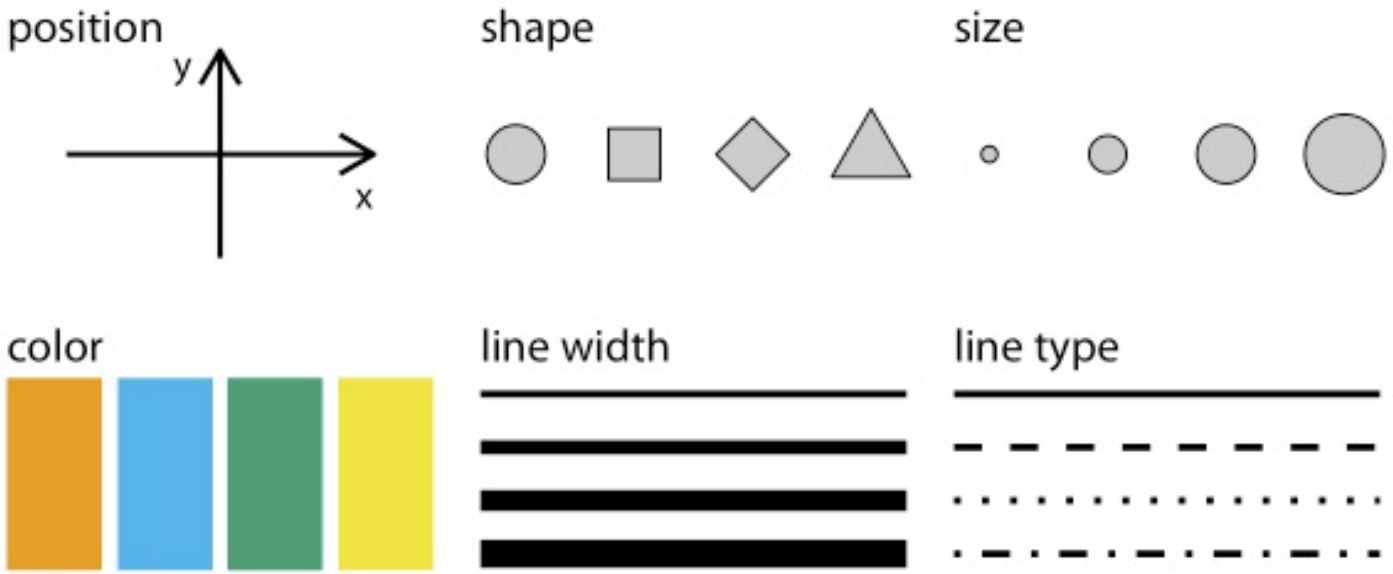
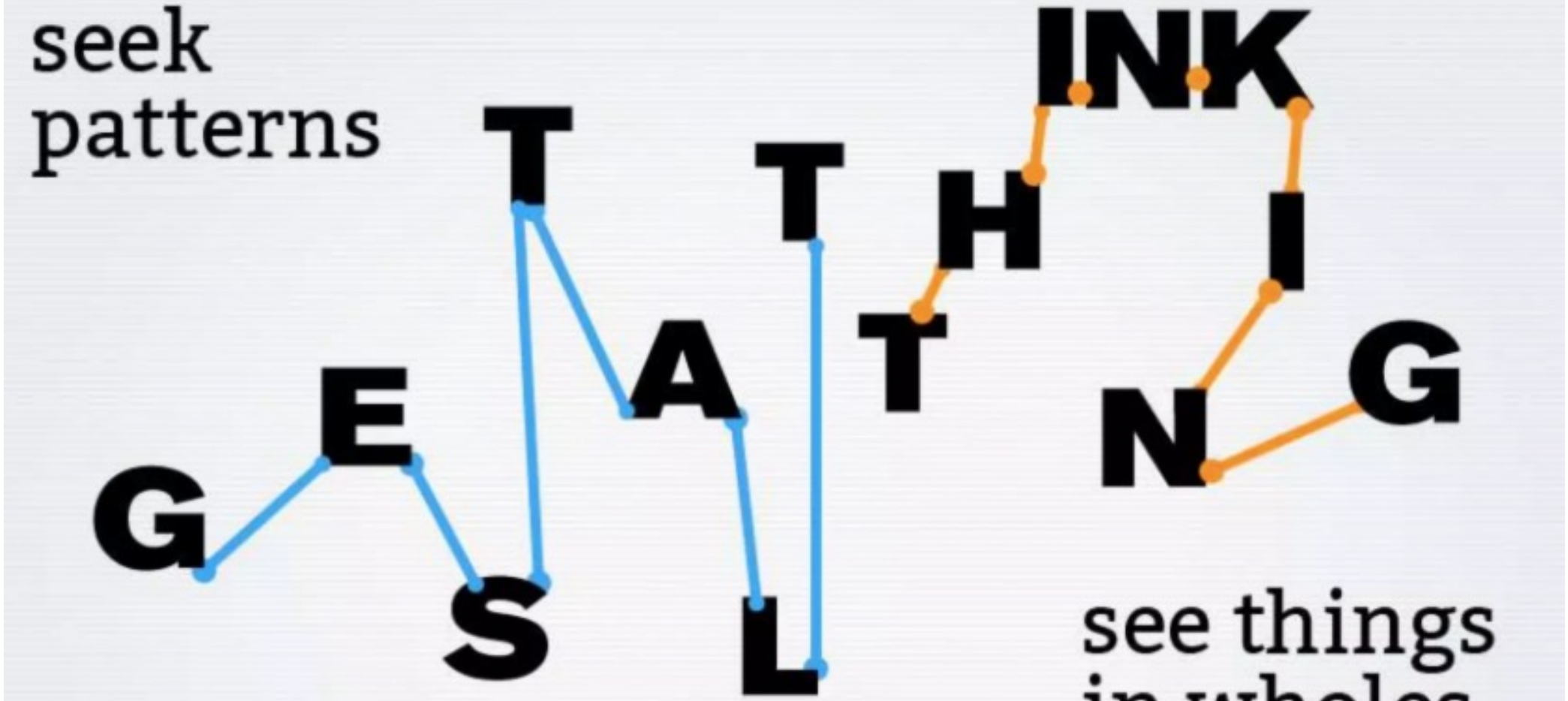


Figure 2.1: Commonly used aesthetics in data visualization: position, shape, size, color, line width, line type. Some of these aesthetics can represent both continuous and discrete data (position, size, line width, color) while others can usually only represent discrete data (shape, line type).


seek
patterns
GESTALT
THINKING
see things
in wholes,
sometimes based on
partial information
make connections,
where sometimes none exist.

6 Useful Gestalt Principles

Proximity

Figure / Ground

Closure

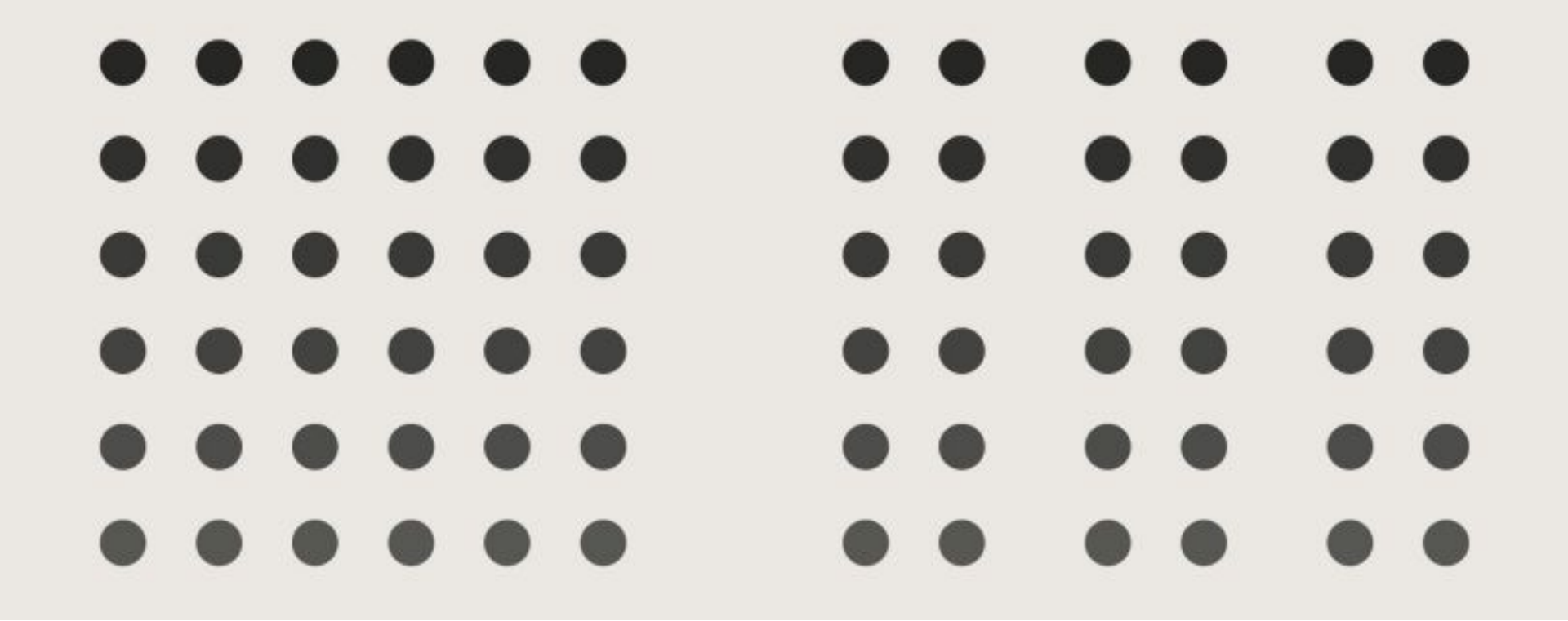
Symmetry

Similarity

Continuity

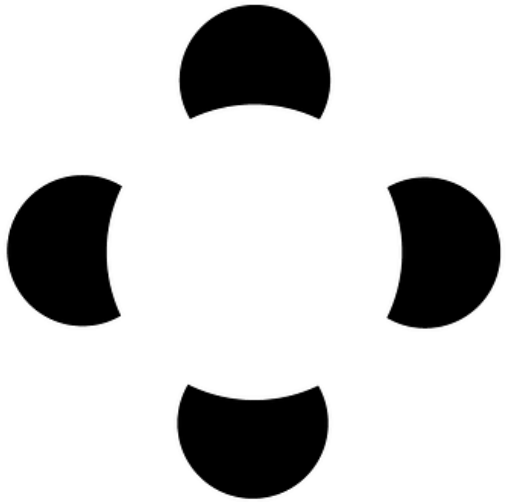
Proximity

<u>The Law of Proximity</u>: People perceive visual elements related to how closely they are positioned to one another

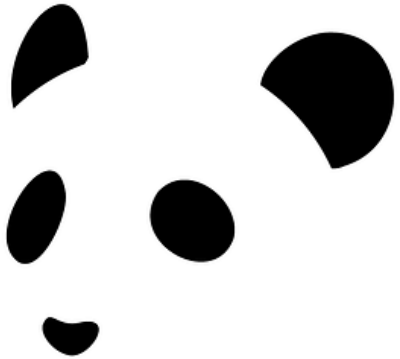


Closure

The Law of Closure: Our tendency to perceive segmented visual elements as complete or whole objects, even when we're missing



4 SMALL CIRCLES OR
1 BIG ONE?



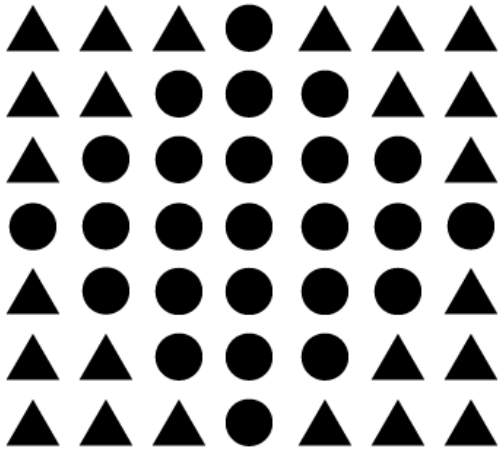
5 BLACK SHAPES OR
1 PANDA BEAR?

3 CURVES OR
1 FULL BLACK CIRCLE?

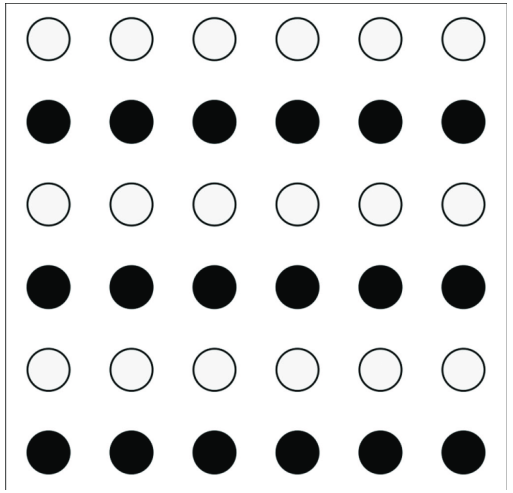
Similarity

<u>The Law of Similarity</u>: the human brain will group together things that appear similar (it also tends to assign them the similar function)

SHAPE



COLOR



SIZE

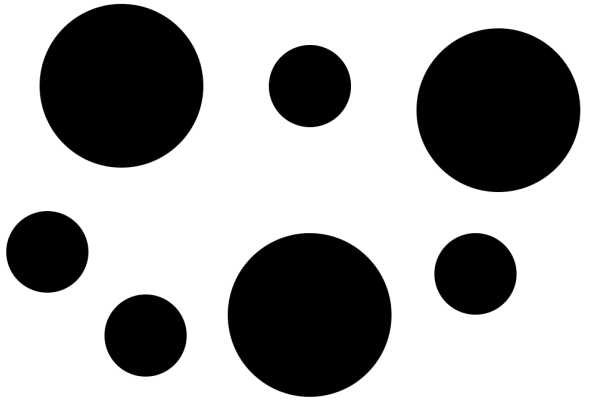
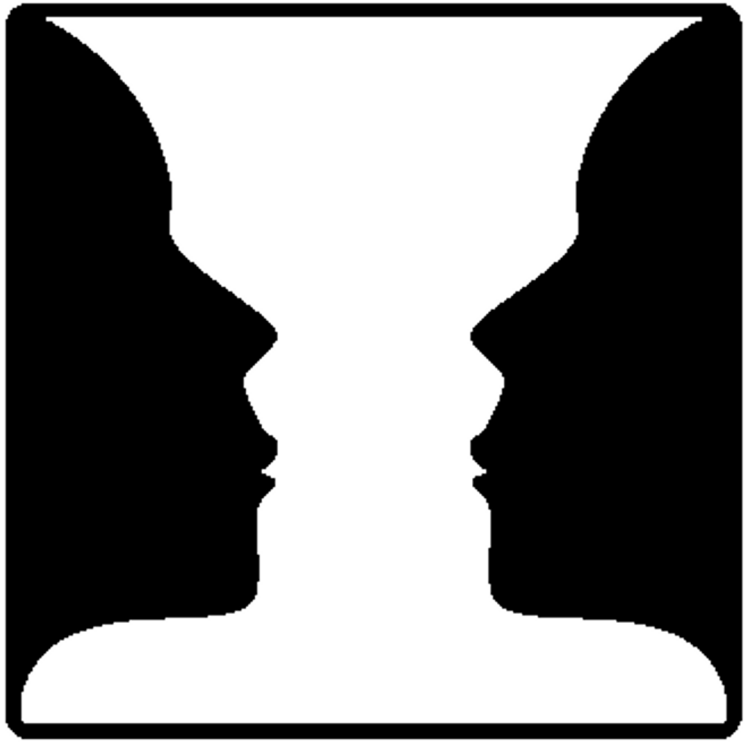


Figure / Ground

<u>The Law of Figure / Ground</u>: the brain will unconsciously place objects either in the foreground or the background

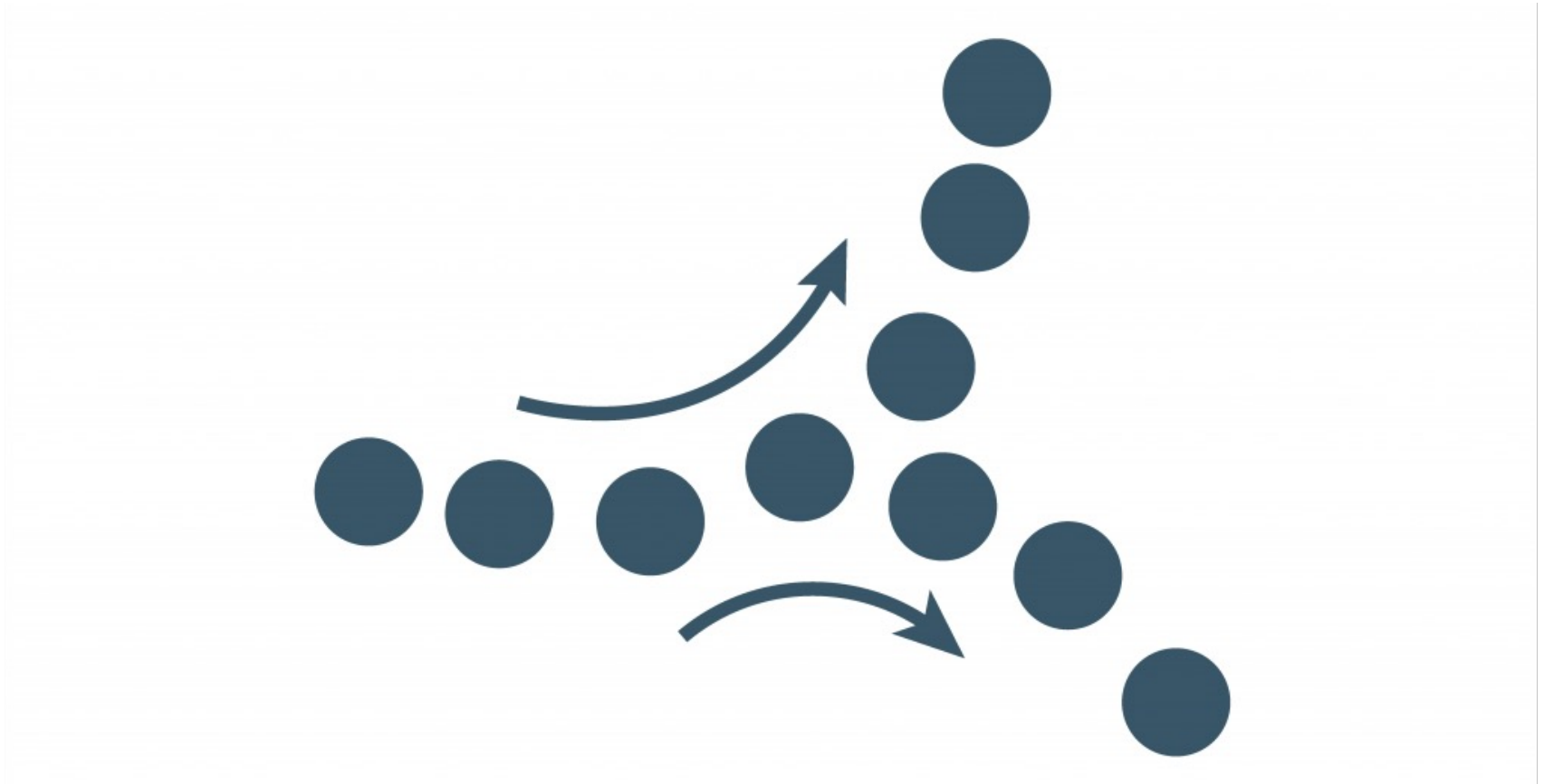


Symmetry

The Law of Symmetry: Visual elements that are symmetrical to each other tend to be perceived as a unified group

Continuity

The Law of Continuity: human brains tend to perceive any line as continuing its established direction



Activity #3

Choose graphs from the previous activities and identity which Gestalt Principles they rely on. Are there other Gestalt Principles that could have improve the visualization?

The Cognitive Psychology of Data Visualization

- **There is a whole field of researchers who study how we perceive data visualizations most efficiently and accurately!**

- **For a review see Franconeri et al. (2021)**

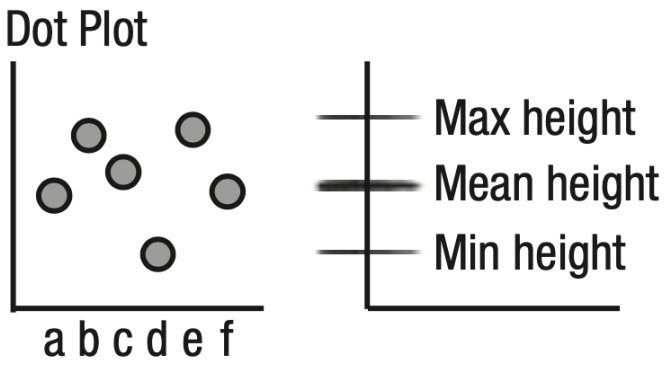
The Cognitive Psychology of Data Visualization

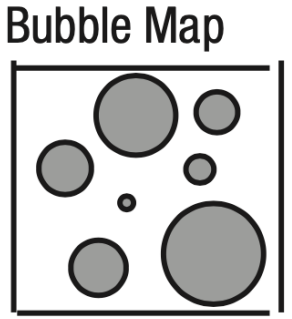
Principles for Efficient Visualization

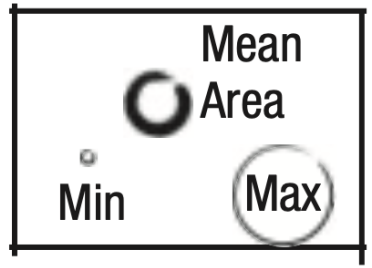
1. Use visualizations to allow viewers to powerfully compute statistics
2. Avoid visual processing limits: making comparisons
3. Control comparison with visual grouping cues
4. Guide viewer to the most important comparison
5. Avoid taxing limited working memory

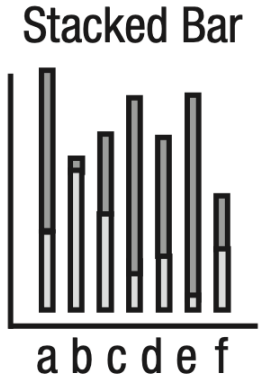
The Cognitive Psychology of Data Visualization

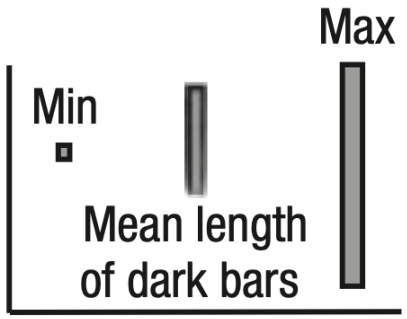
1. **Use visualizations to allow viewers to powerfully compute statistics**

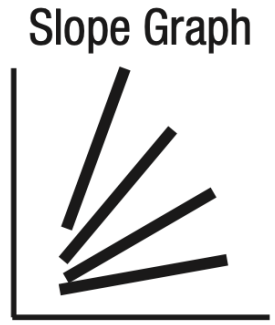


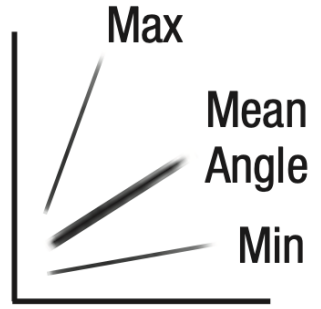












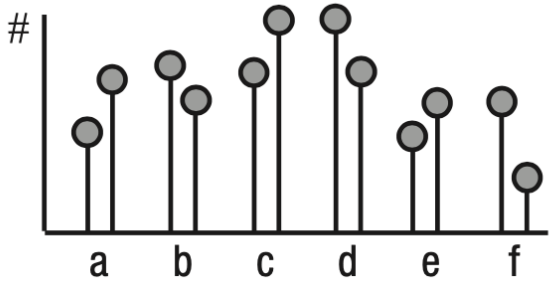
Franconeri et al., 2021

The Cognitive Psychology of Data Visualization

2. Avoid visual processing limits: making comparisons

Vision Is Sluggish for Comparisons

Isolating pairs with “larger second values” is tough…



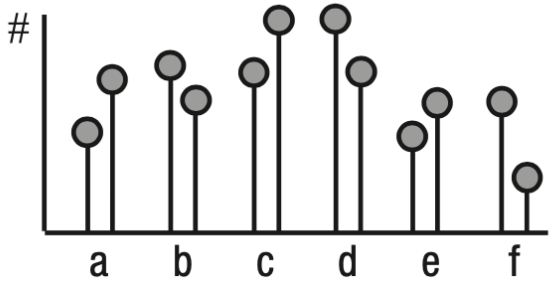
Franconeri et al., 2021

The Cognitive Psychology of Data Visualization

2. Avoid visual processing limits: making comparisons

Vision Is Sluggish for Comparisons

Isolating pairs with “larger second values” is tough…



So guide viewers to the right comparisons

Tool: Shortcut comparisons by adding direct depictions of the deltas, as below

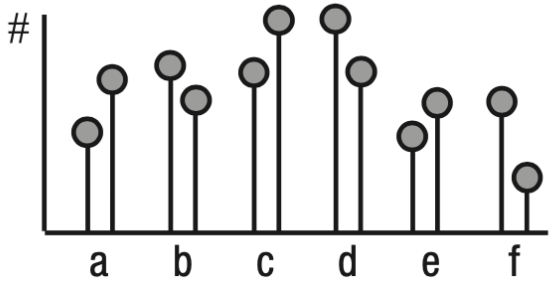
Franconeri et al., 2021

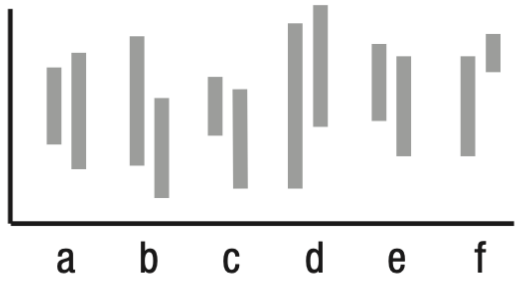
The Cognitive Psychology of Data Visualization

2. Avoid visual processing limits: making comparisons

Vision Is Sluggish for Comparisons

Isolating pairs with “larger second values” is tough…





So guide viewers to the right comparisons

Tool: Shortcut comparisons by adding direct depictions of the deltas, as below

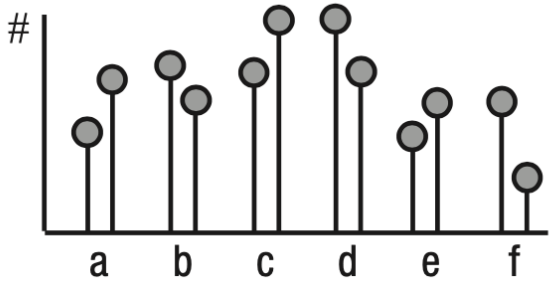
Franconeri et al., 2021

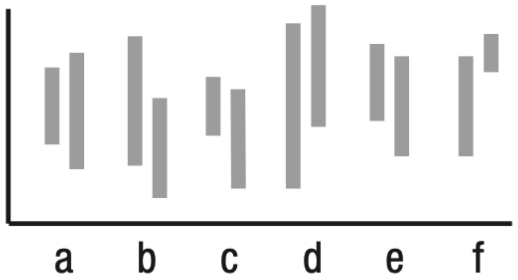
The Cognitive Psychology of Data Visualization

2. Avoid visual processing limits: making comparisons

Vision Is Sluggish for Comparisons

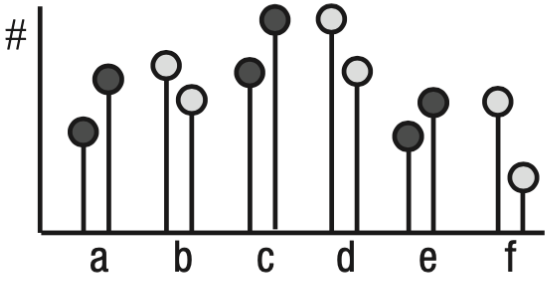
Isolating pairs with “larger second values” is tough…





So guide viewers to the right comparisons

Tool: Shortcut comparisons by adding direct depictions of the deltas, as below

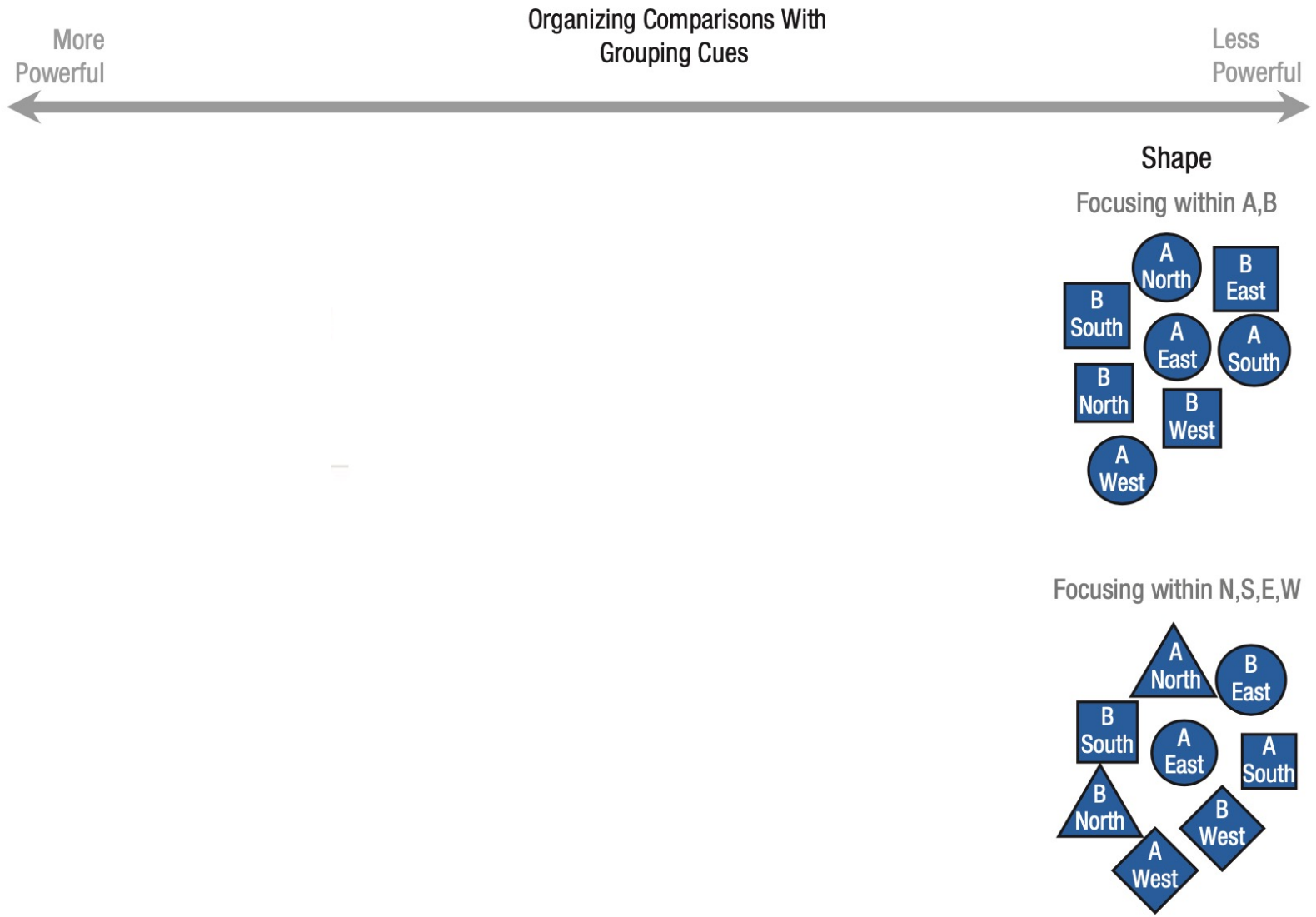


“a, c, & e have increased”

Tool: Highlight and annotate the right comparisons for your viewers, as above.

The Cognitive Psychology of Data Visualization

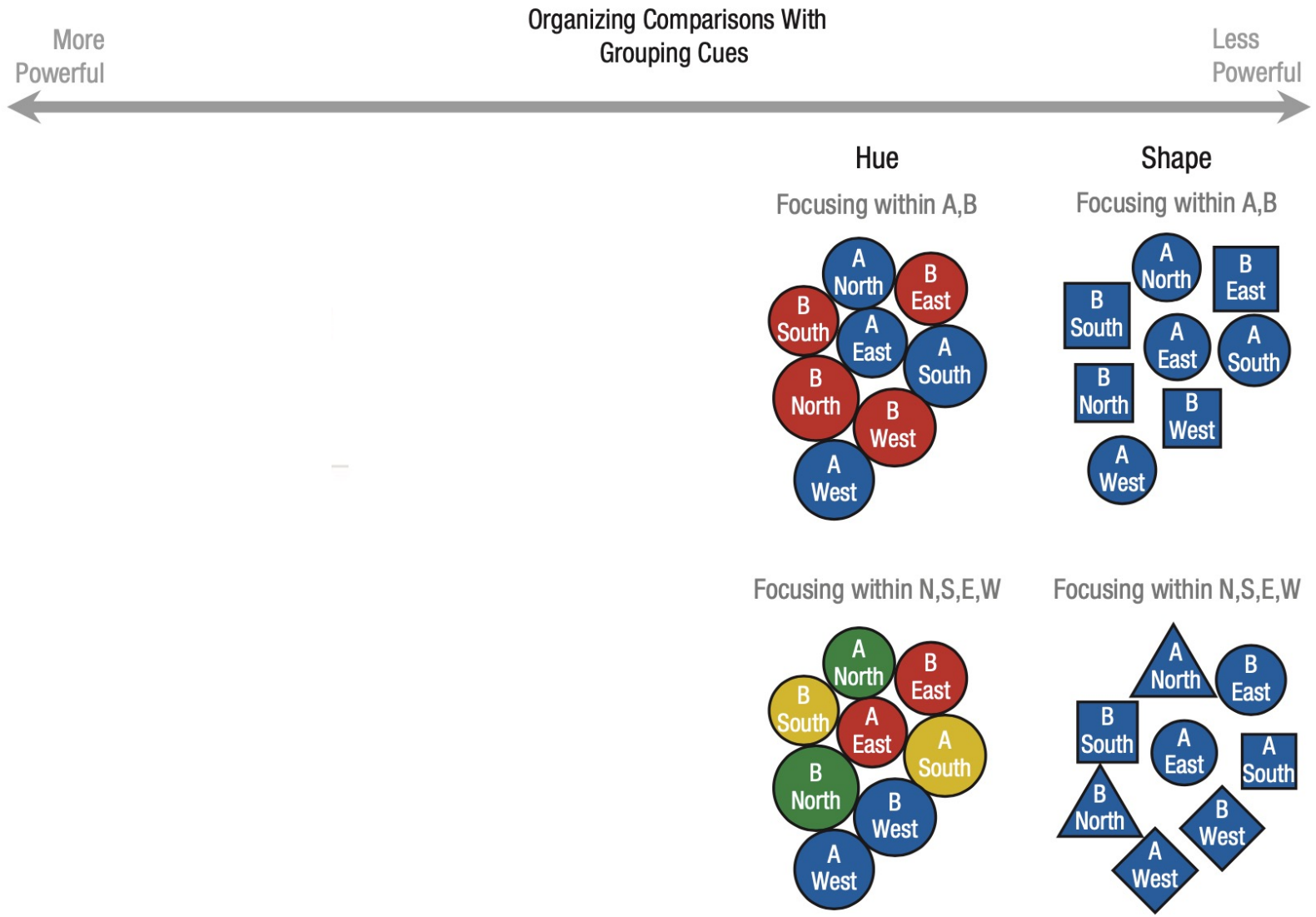
3. Control comparison with visual grouping cue



Franconeri et al., 2021

The Cognitive Psychology of Data Visualization

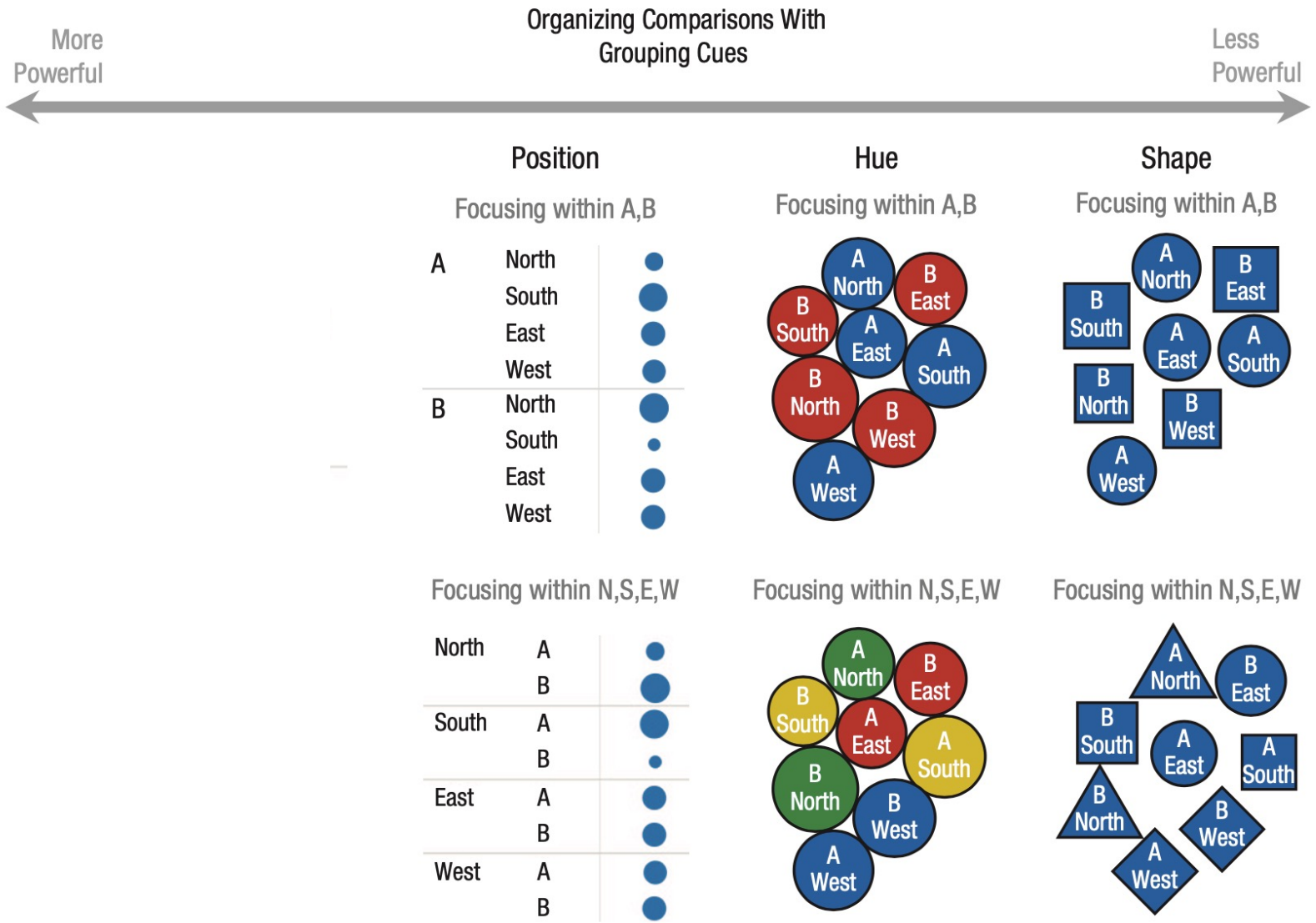
3. Control comparison with visual grouping cue



Franconeri et al., 2021

The Cognitive Psychology of Data Visualization

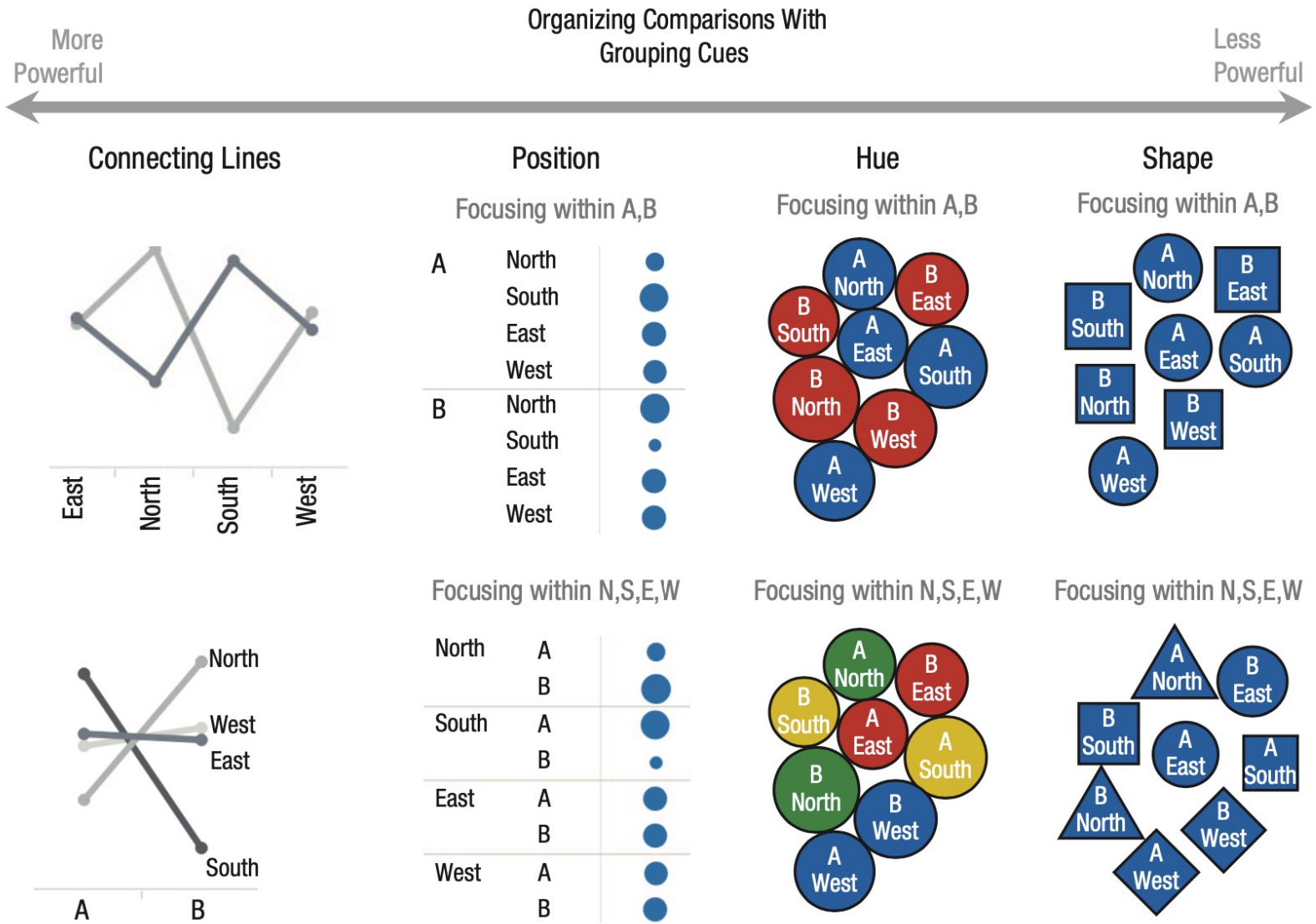
3. Control comparison with visual grouping cue



Franconeri et al., 2021

The Cognitive Psychology of Data Visualization

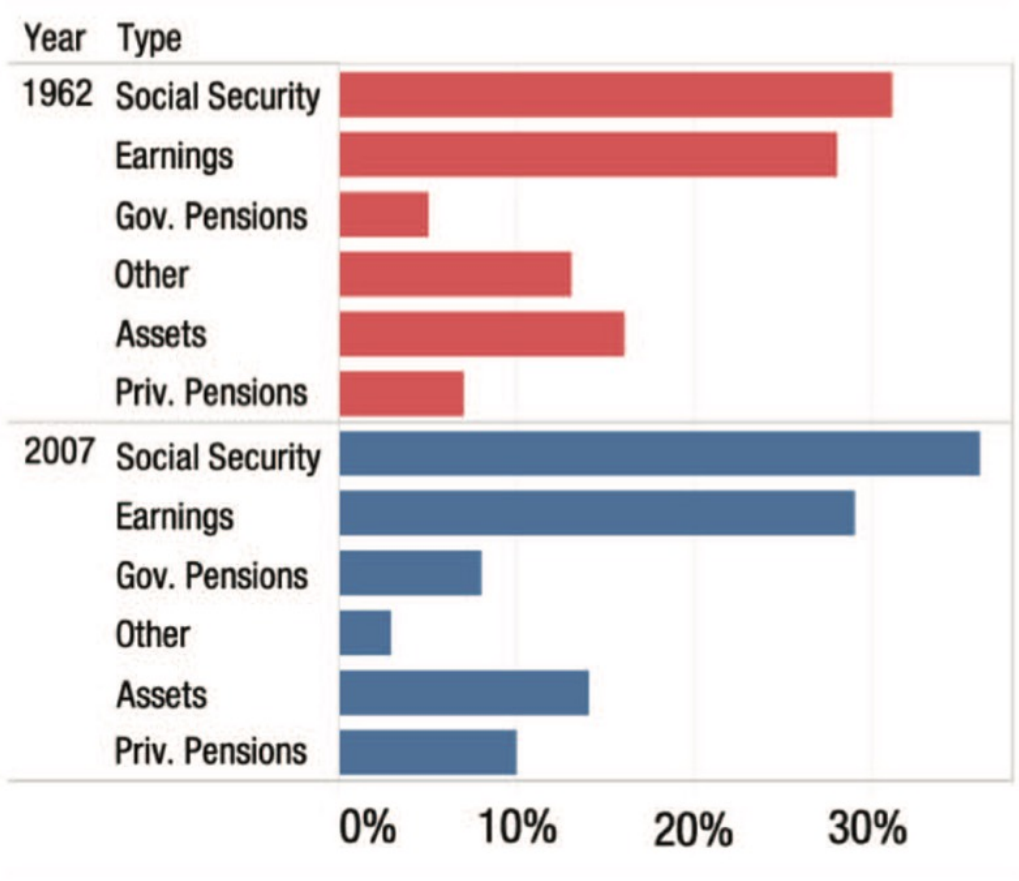
3. Control comparison with visual grouping cue



Franconeri et al., 2021

The Cognitive Psychology of Data Visualization

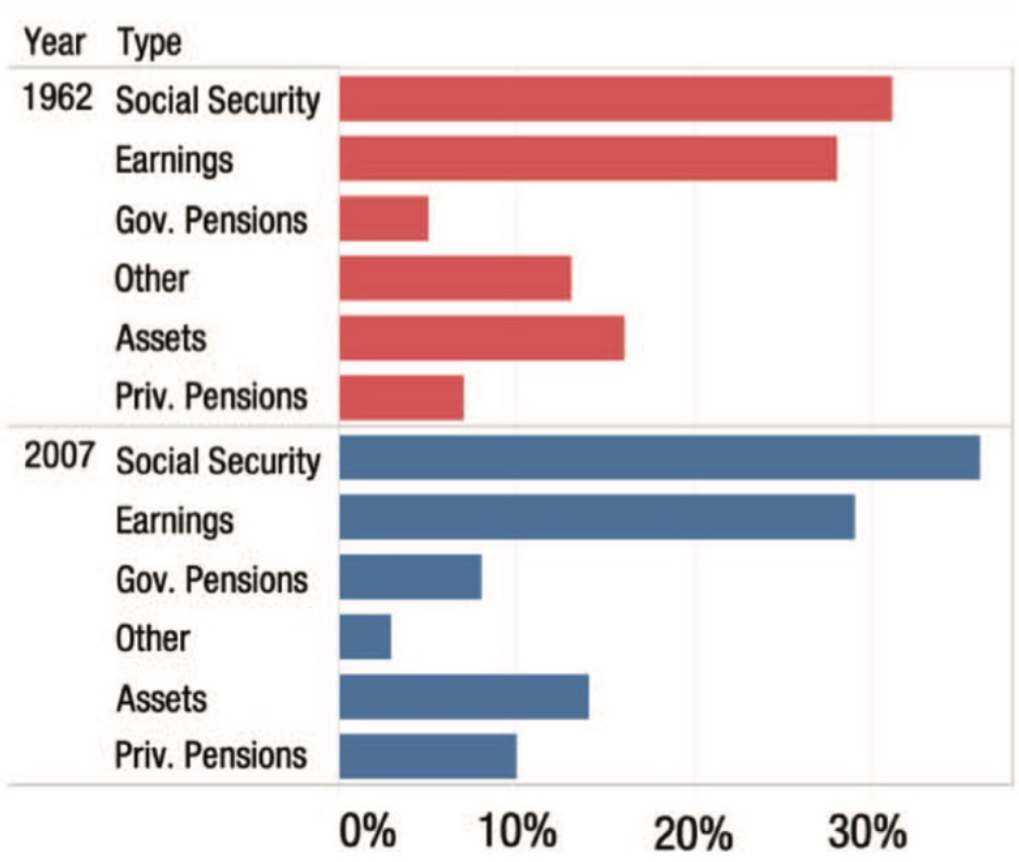
4. Guide viewer to the most important comparison

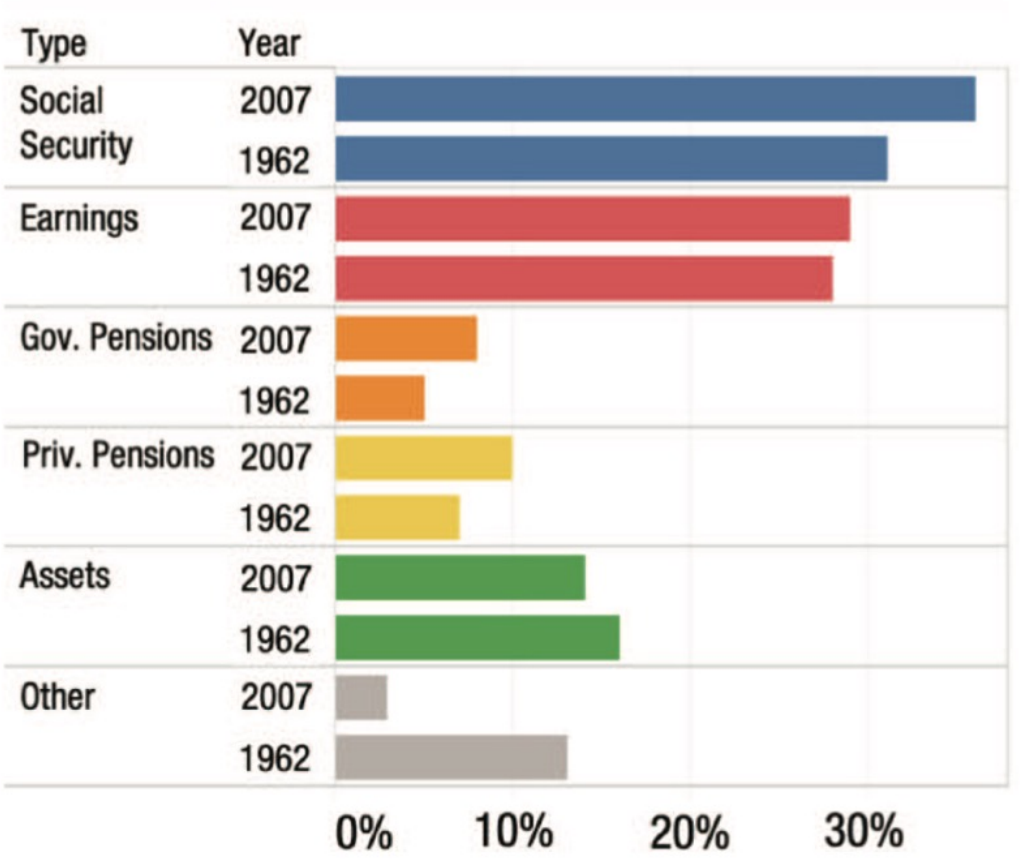


Franconeri et al., 2021

The Cognitive Psychology of Data Visualization

4. Guide viewer to the most important comparison

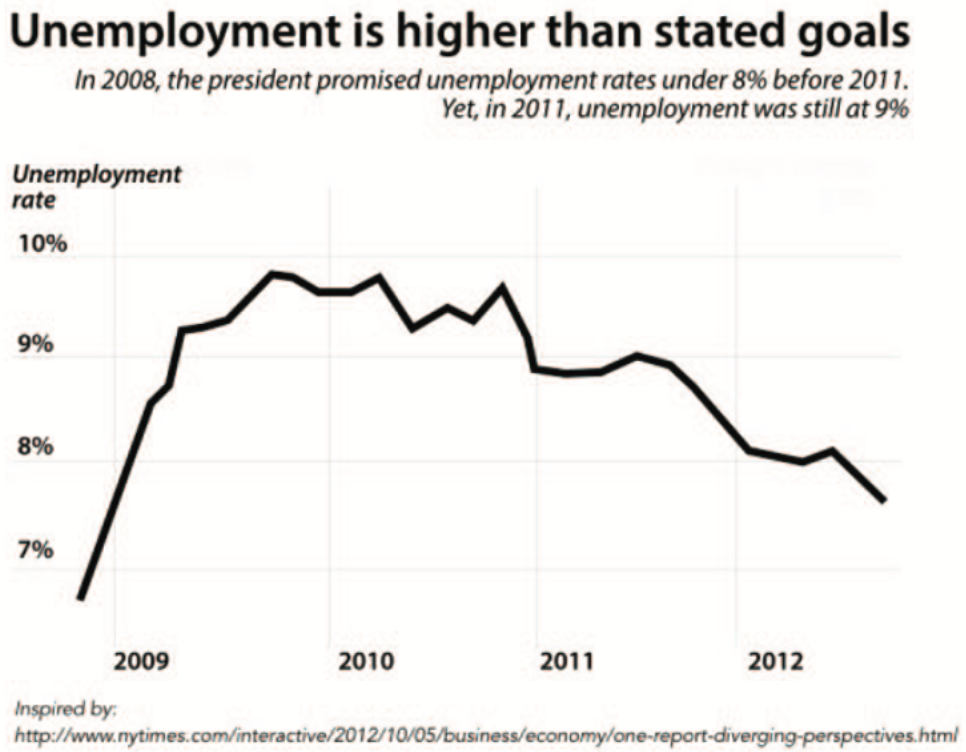




Franconeri et al., 2021

The Cognitive Psychology of Data Visualization

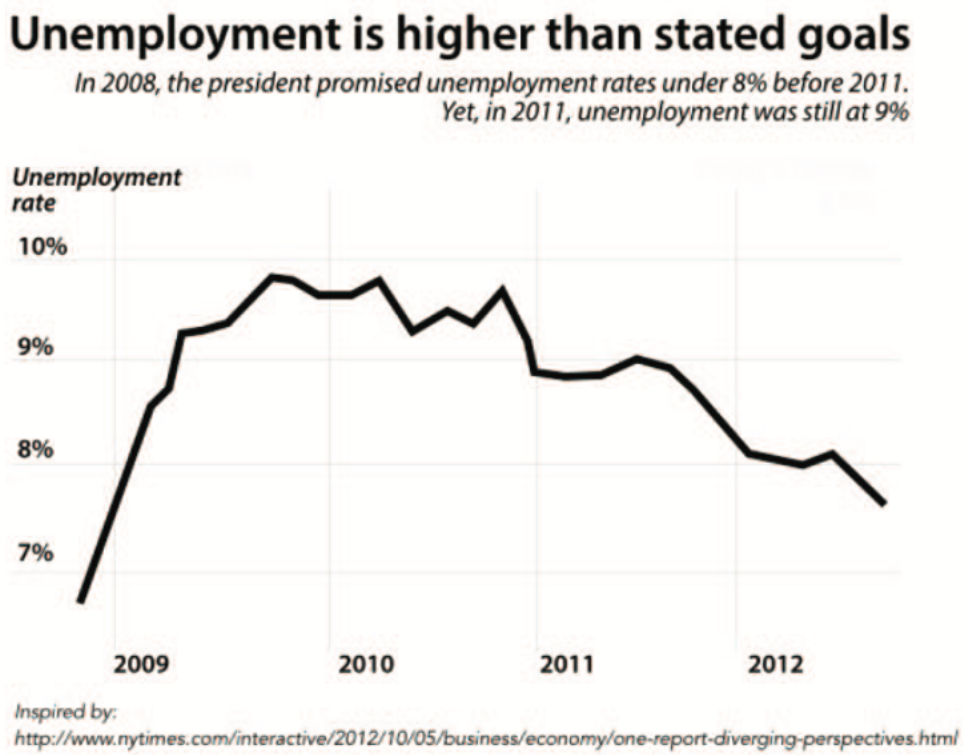
4. Guide viewer to the most important comparison

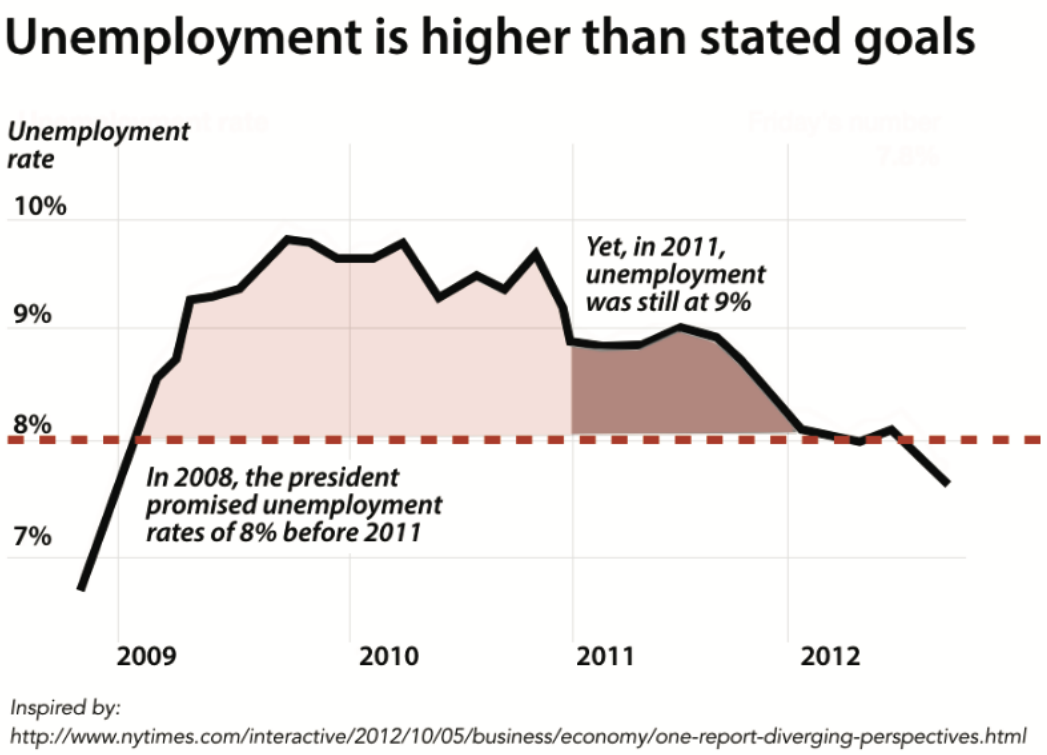


Franconeri et al., 2021

The Cognitive Psychology of Data Visualization

4. Guide viewer to the most important comparison

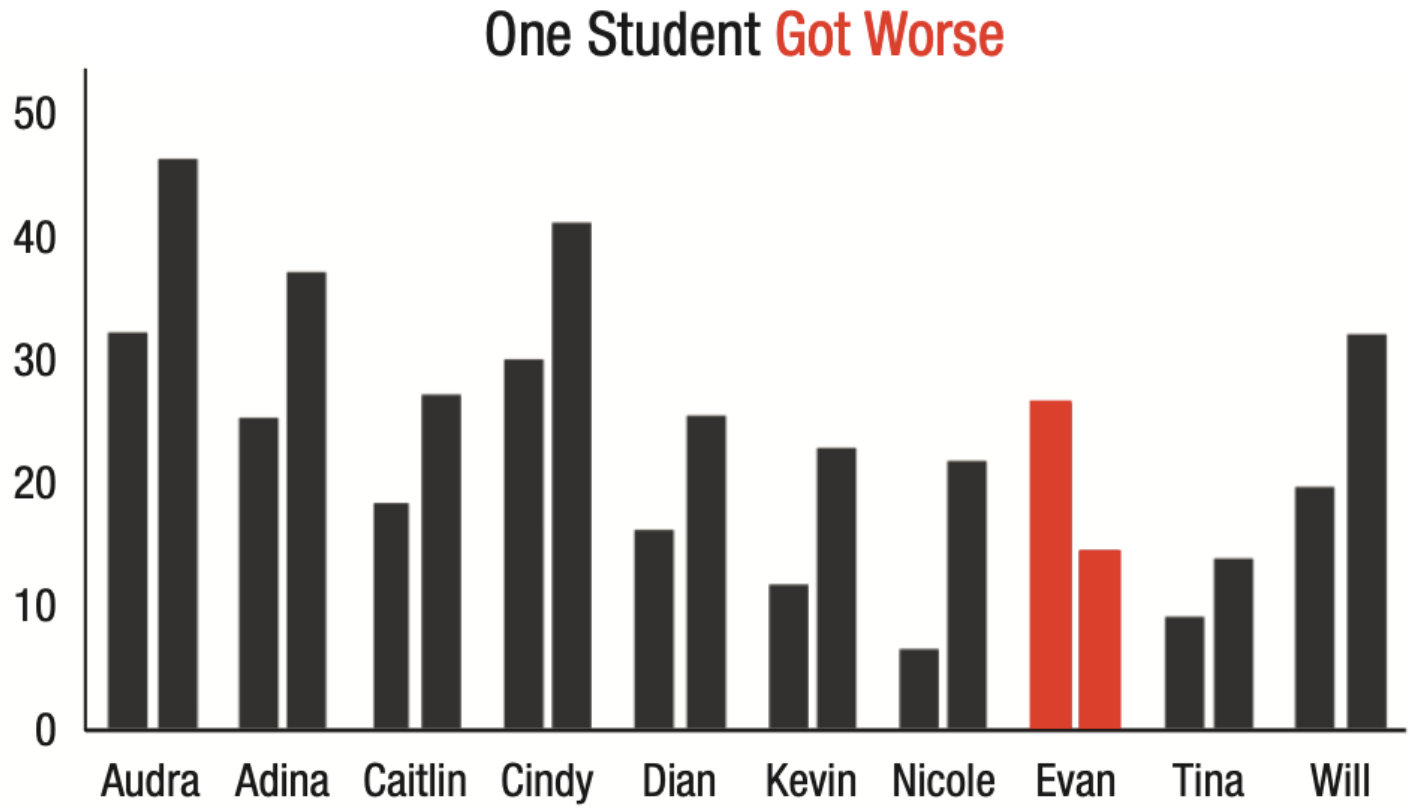




Franconeri et al., 2021

The Cognitive Psychology of Data Visualization

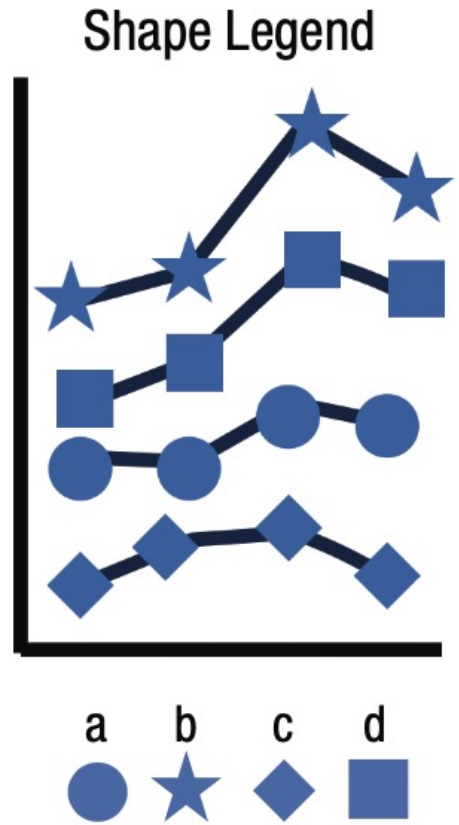
4. Guide viewer to the most important comparison



Franconeri et al., 2021

The Cognitive Psychology of Data Visualization

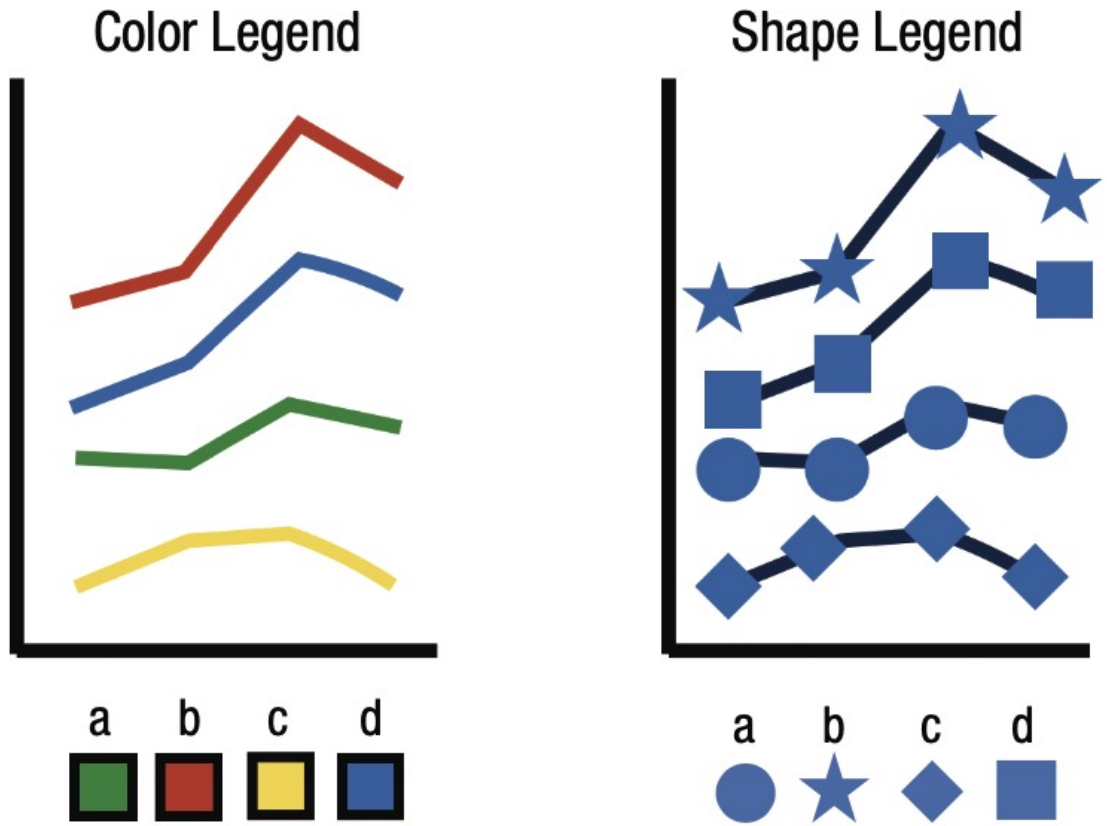
5. Avoid taxing limited working memory



Franconeri et al., 2021

The Cognitive Psychology of Data Visualization

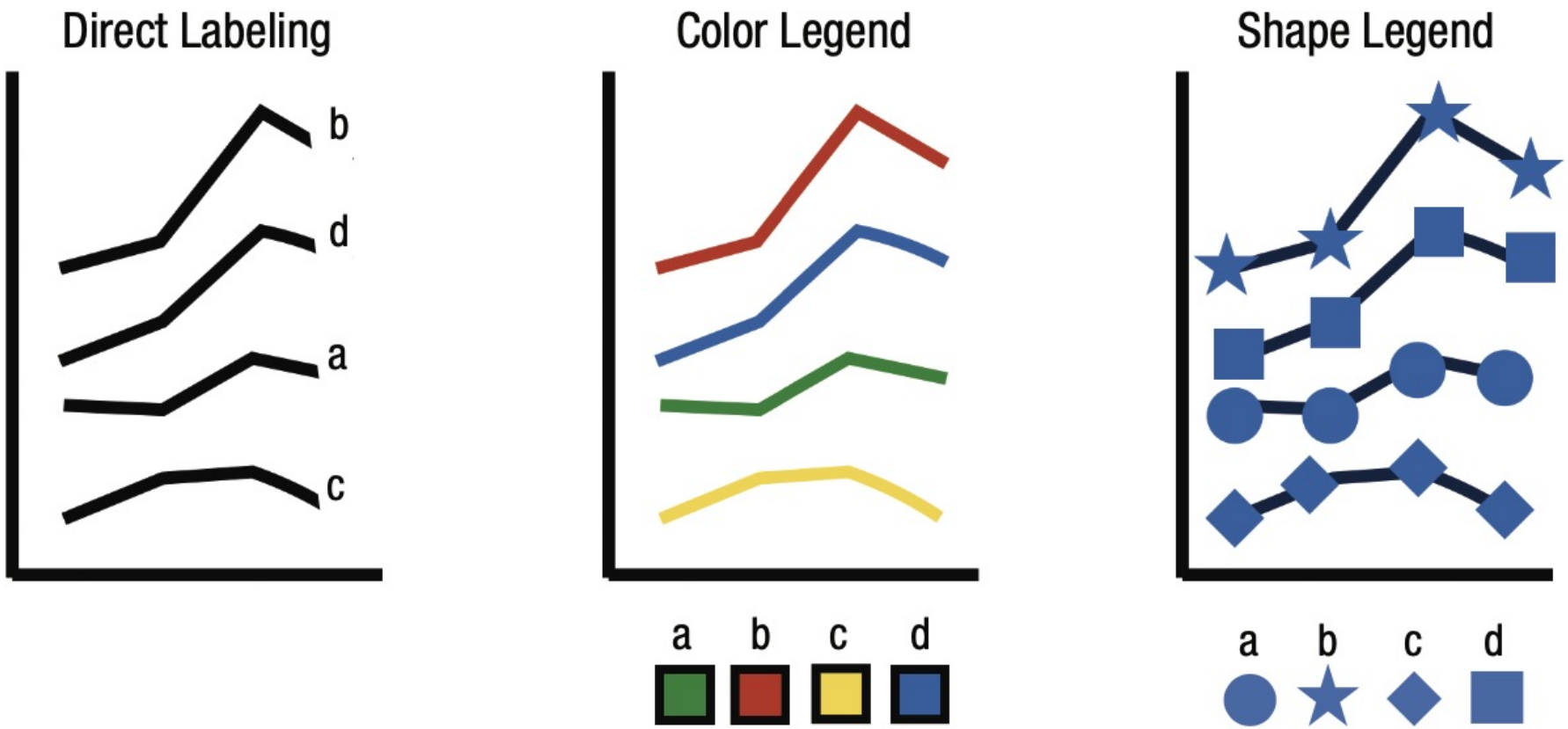
5. Avoid taxing limited working memory



Franconeri et al., 2021

The Cognitive Psychology of Data Visualization

5. Avoid taxing limited working memory



Franconeri et al., 2021

Some Final Notes

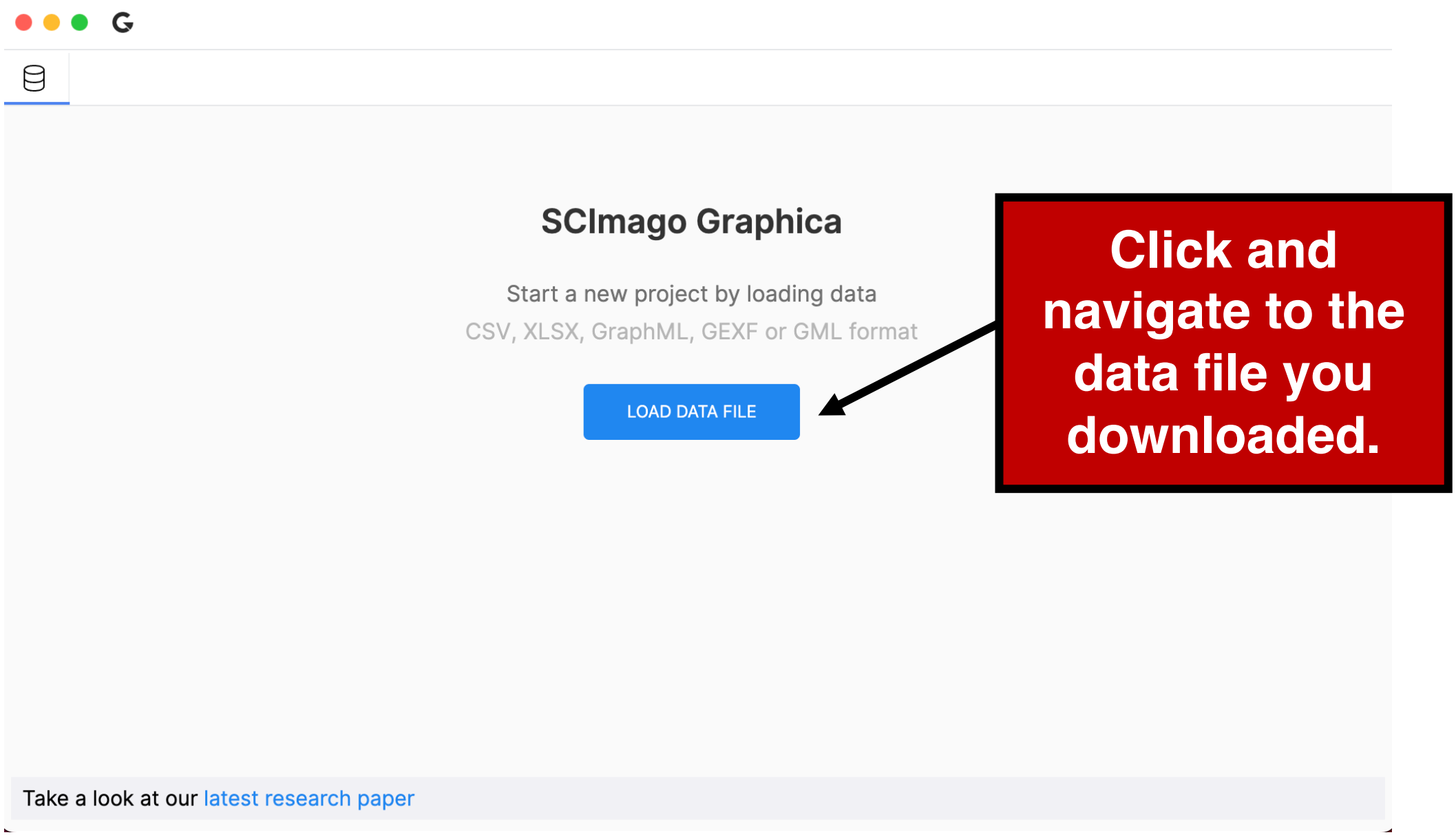
1. Increase your font size on all labels and titles
2. Use colorblind friendly palettes
3. Note all visualizations reflect data. Adobe, powerpoint, etc. are great tools for this.
4. Good visualizations can take time, planning, and feedback
5. Avoid 3D visualizations in most cases
6. Use animations and interactive graphics sparingly / when appropriate
7. Have fun!

Part 2:

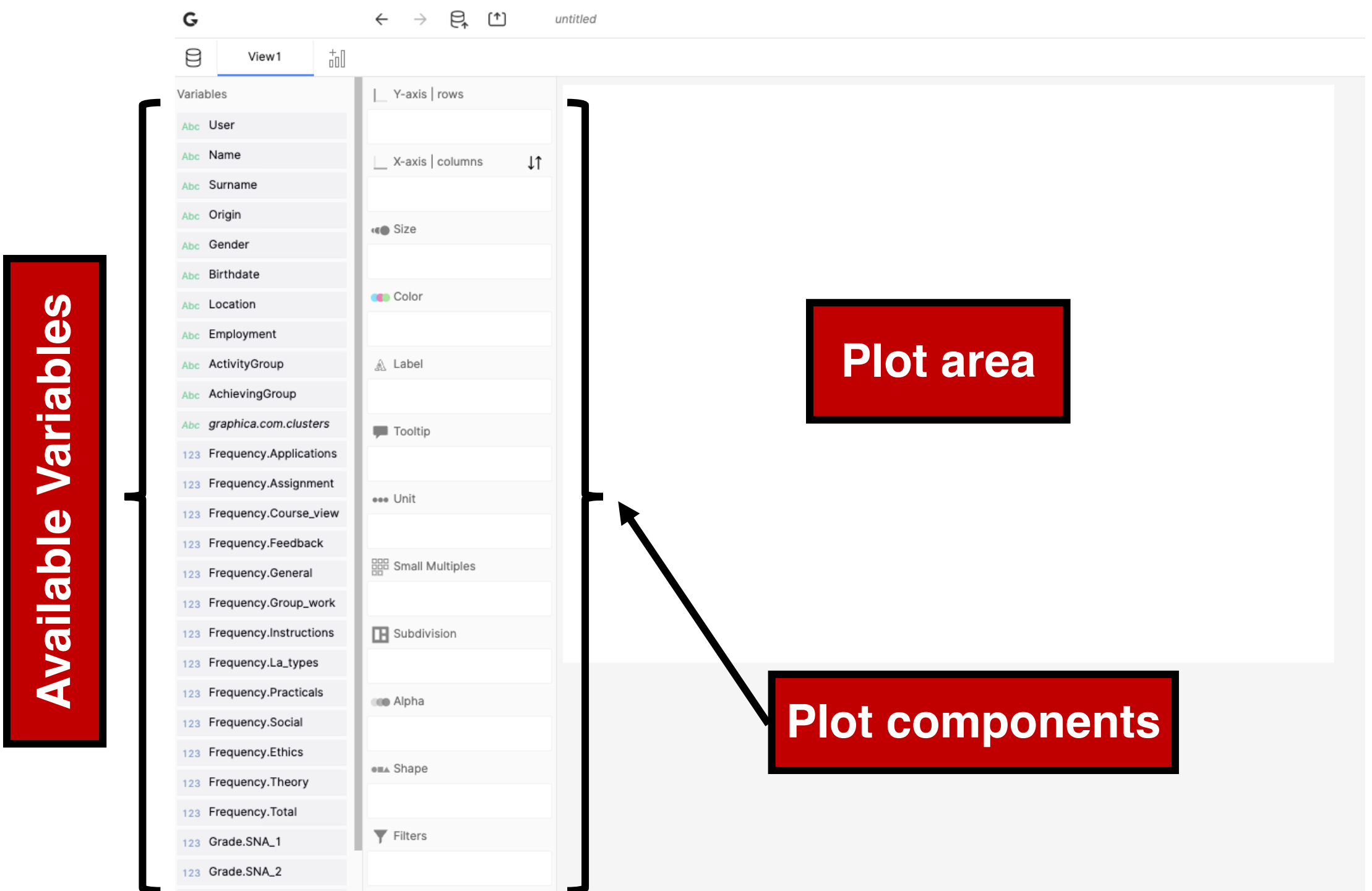
Building Data Visualizations

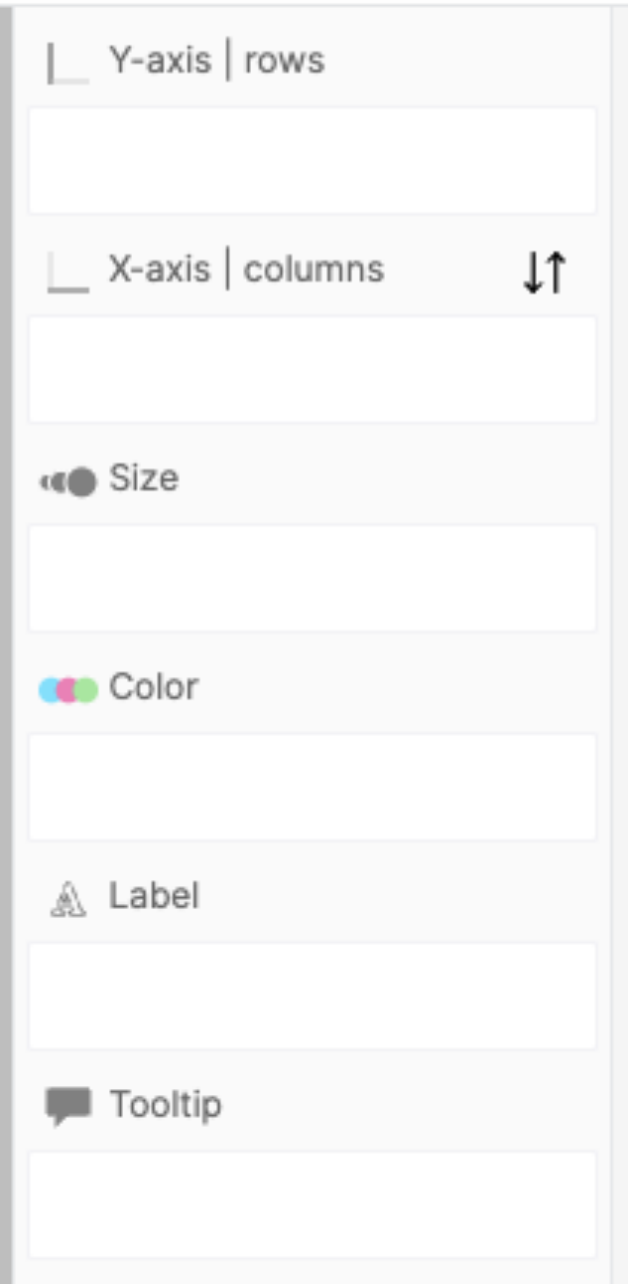
Tool: https://www.graphica.app/

Data: https://tinyurl.com/mvvys6ne


SCImago Graphica
Start a new project by loading data
CSV, XLSX, GraphML, GEXF or GML format
LOAD DATA FILE
Take a look at our latest research paper
Click and navigate to the data file you downloaded.

Remote	N	2	121	103	7
Remote	N	0	62	294	22
On campus	N	0	41	53	0
On campus	N	0	44	49	0
On campus	None	0	9	119	9
On campus	Part-time	21	46	262	50
On campus	None	0	33	189	22
On campus	None	5	53	226	26
On campus	None	0	26	58	13
On campus	None	0	52	199	8
Remote	None	9	52	153	14
On campus	None	0	25	111	24
Remote	Part-time	9	84	246	42
Remote	Full-time	3	23	176	25
On campus	None	2	115	107	8
On campus	None	0	22	63	10
On campus	Full-time	0	50	88	18
On campus	None	0	42	50	0


untitled
View1
Variables
Abc User
Abc Name
Abc Surname
Abc Origin
Abc Gender
Abc Birthdate
Abc Location
Abc Employment
Abc ActivityGroup
Abc AchievingGroup
Abc graphica.com.clusters
123 Frequency.Applications
123 Frequency.Assignment
123 Frequency.Course_view
123 Frequency.Feedback
123 Frequency.General
123 Frequency.Group_work
123 Frequency.Instructions
123 Frequency.La_types
123 Frequency.Practicals
123 Frequency.Social
123 Frequency.Ethics
123 Frequency.Theory
123 Frequency.Total
123 Grade.SNA_1
123 Grade.SNA_2
Y-axis | rows
X-axis | columns
Size
Color
Label
Tooltip
Unit
Small Multiples
Subdivision
Alpha
Shape
Filters
Available Variables
Plot area
Plot components



Y-axis (vertical)

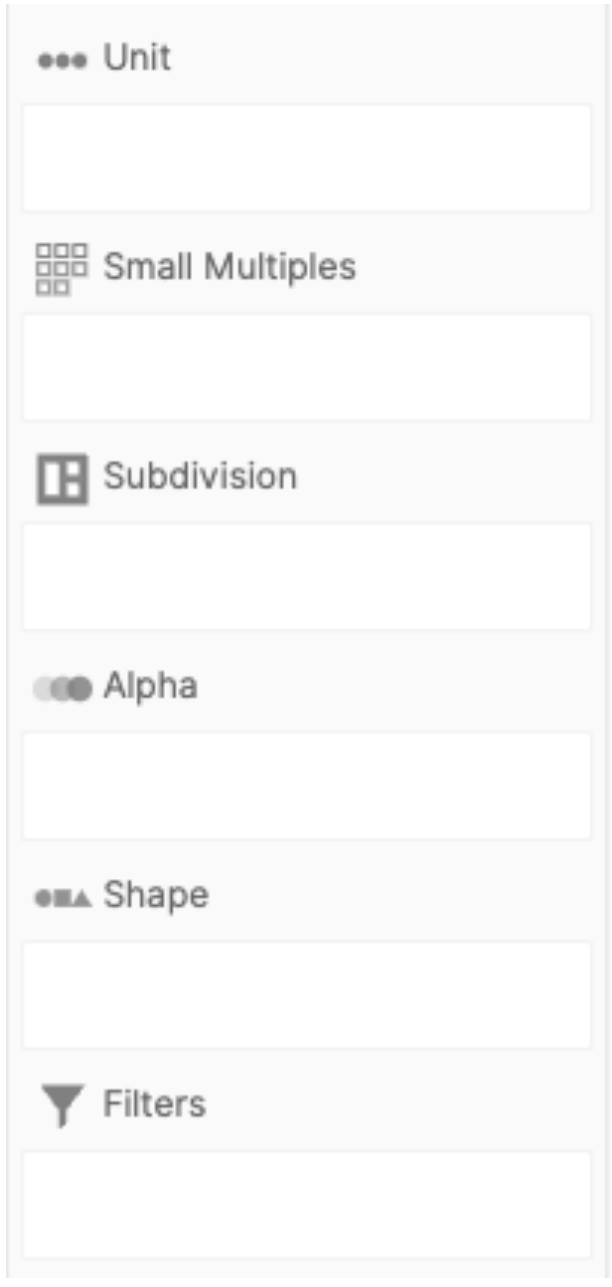
X-axis (horizontal)

Size components based on a variable

Color components based on a variable

Label components based on a variable

Pop-up text when hover cursor

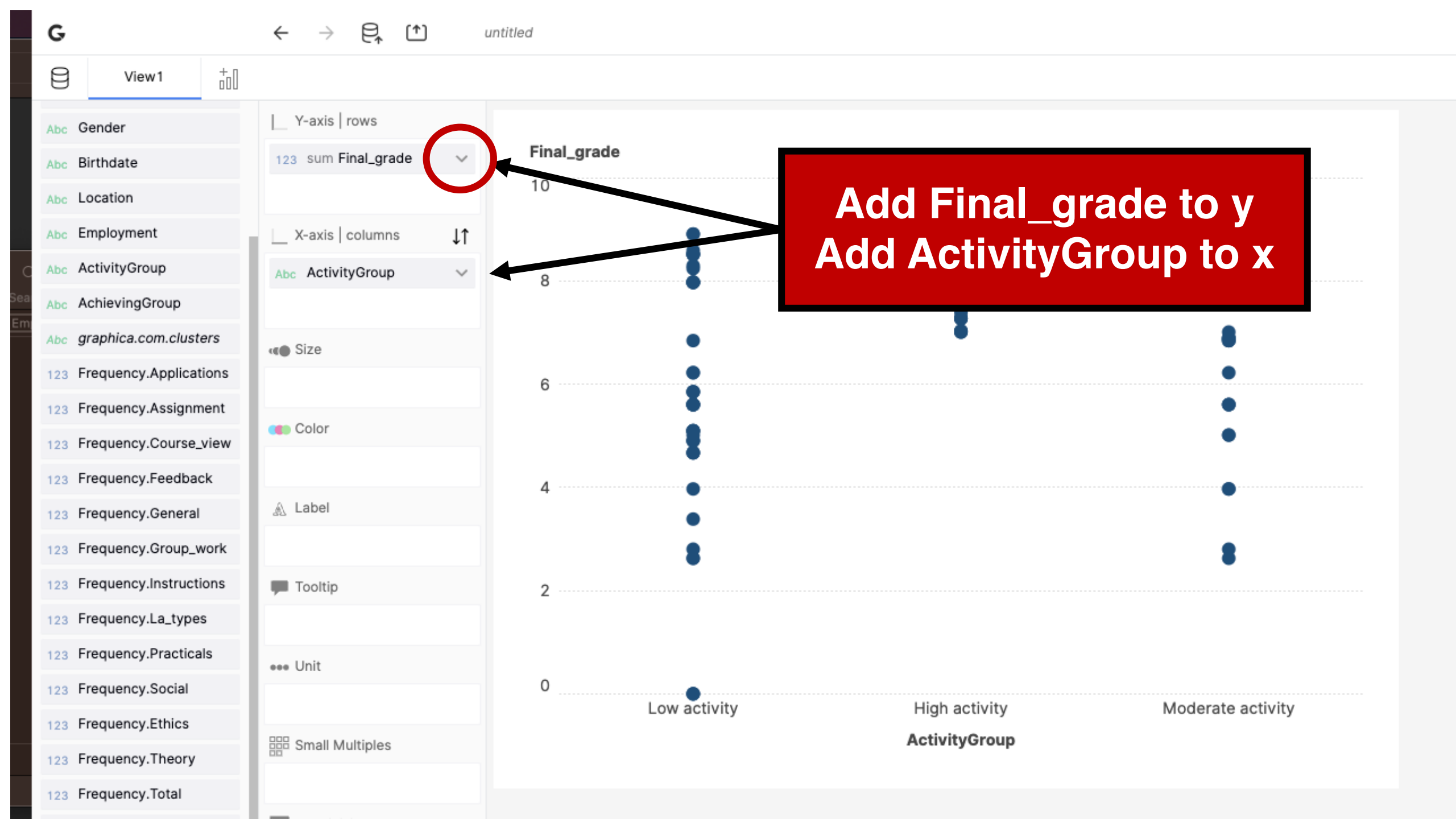


Facet data by variables

Change opacity based on a variable

Change shape based on a variable

Change shape based on a variable


untitled
View1
Abc Gender
Abc Birthdate
Abc Location
Abc Employment
Abc ActivityGroup
Abc AchievingGroup
Abc graphica.com.clusters
123 Frequency.Applications
123 Frequency.Assignment
123 Frequency.Course_view
123 Frequency.Feedback
123 Frequency.General
123 Frequency.Group_work
123 Frequency.Instructions
123 Frequency.La_types
123 Frequency.Practicals
123 Frequency.Social
123 Frequency.Ethics
123 Frequency.Theory
123 Frequency.Total
Y-axis | rows
123 sum Final_grade
X-axis | columns
Abc ActivityGroup
Size
Color
Label
Tooltip
Unit
Small Multiples
Final_grade
10
8
6
4
2
0
Low activity
High activity
Moderate activity
ActivityGroup
Add Final_grade to y
Add ActivityGroup to x

untitled

View1

Abc Gender
Abc Birthdate
Abc Location
Abc Employment
Abc ActivityGroup
Abc AchievingGroup
Abc *graphica.com.clusters*
123 Frequency.Applications
123 Frequency.Assignment
123 Frequency.Course_view
123 Frequency.Feedback
123 Frequency.General
123 Frequency.Group_work
123 Frequency.Instructions
123 Frequency.La_types
123 Frequency.Practicals
123 Frequency.Social
123 Frequency.Ethics
123 Frequency.Theory
123 Frequency.Total

Y-axis | rows

Final_grade

Scale

Aggregation function

- ✓ Sum
- Average
- Median
- Min
- Max

Min value: automatic
Max value: automatic

Reversed scale

Title

Labels

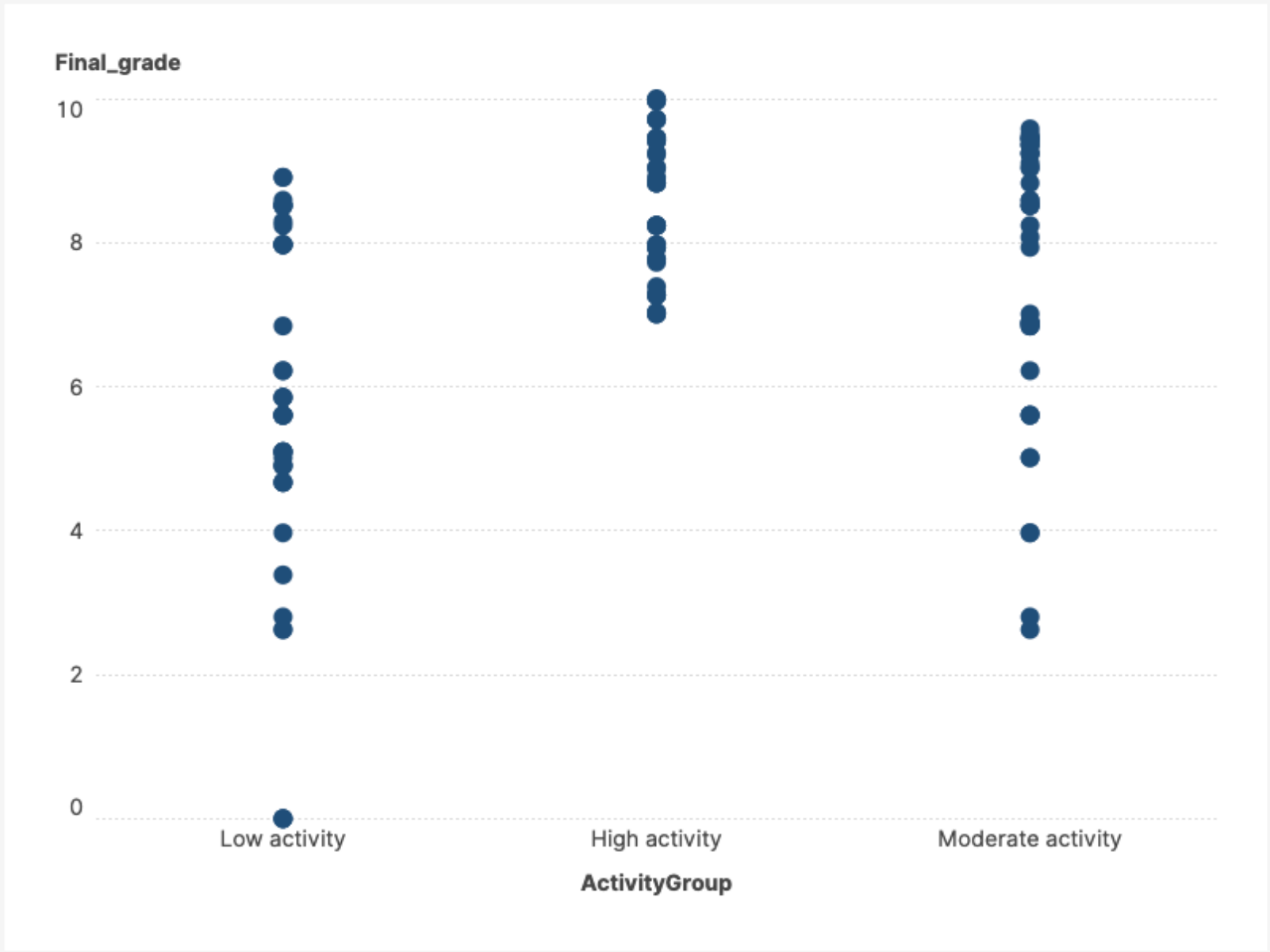
Grid lines

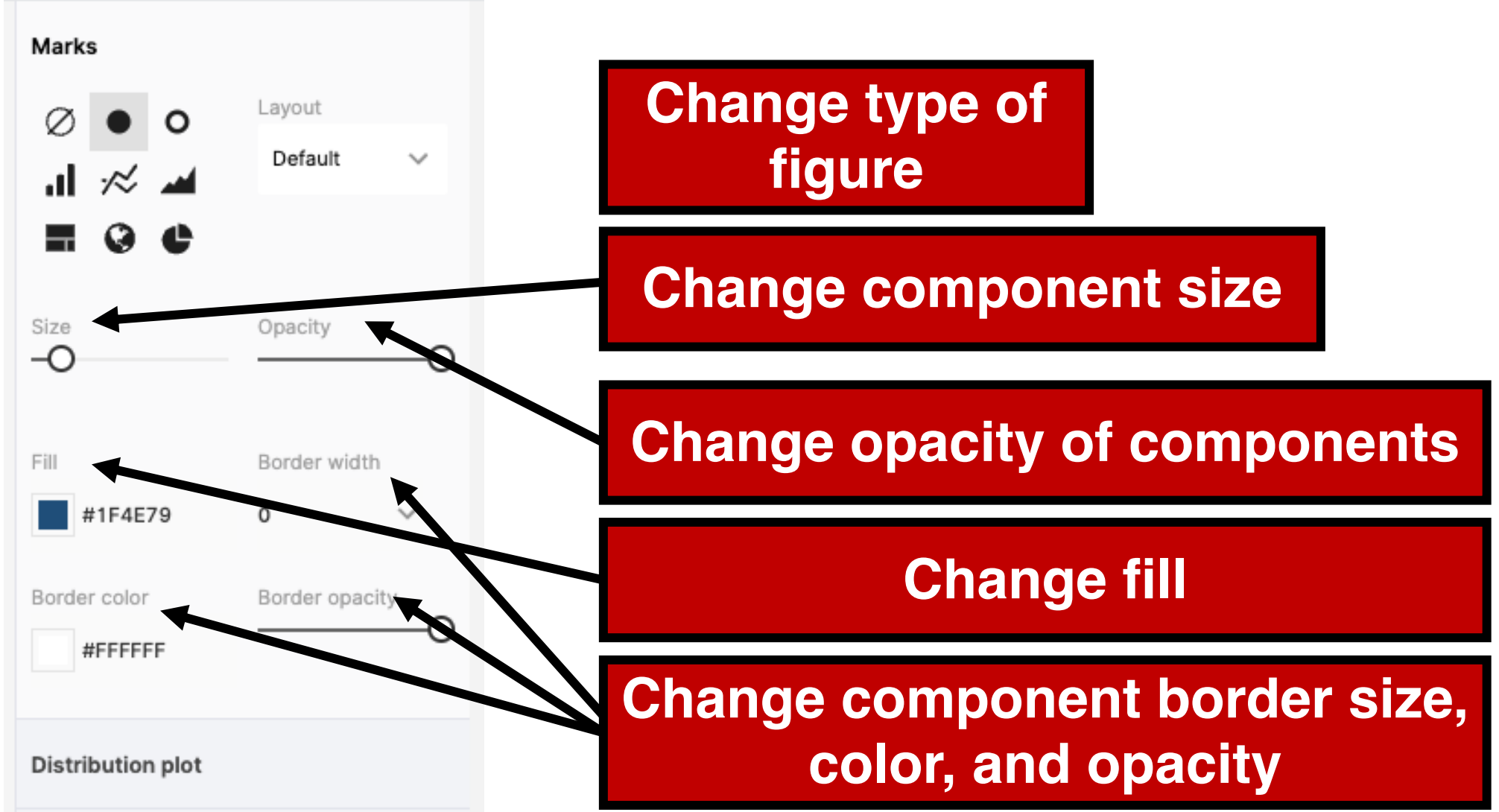
Position

Interactive filter

Number format

Small Multiples




Marks
Layout
Default
Size
Opacity
Fill
#1F4E79
Border width
0
Border color
#FFFFFF
Border opacity
Distribution plot
Reference lines
Page setup
Interactions
Change type of figure
Change component size
Change opacity of components
Change fill
Change component border size, color, and opacity

Activity #4

1. Choose two continuous variables and create a figure with a trend line, with color based on a categorical variable and size based on a frequency variable.

2. Choose one continuous and one categorical variable and create a figure with color based on a second categorical variable, and small multiples based on a third categorical variable.

3. Choose two categorical variables and create a figure capturing the proportion of people in different groups.

Resources & Learning More

- Wilke's *Fundamentals of Data Visualization*
- Healy's *Perception in Visualization*
- Healy's course site
- Kazakova (2021)
- Franconeri (2021)
- Jessica Hullman's website
- Matthew Kay's work on visualizing uncertainty
- My (2022) course GitHub