
Data Visualization
An Introduction

Visualization assignment: Corrected due date

Table 2: Assignment Due Dates

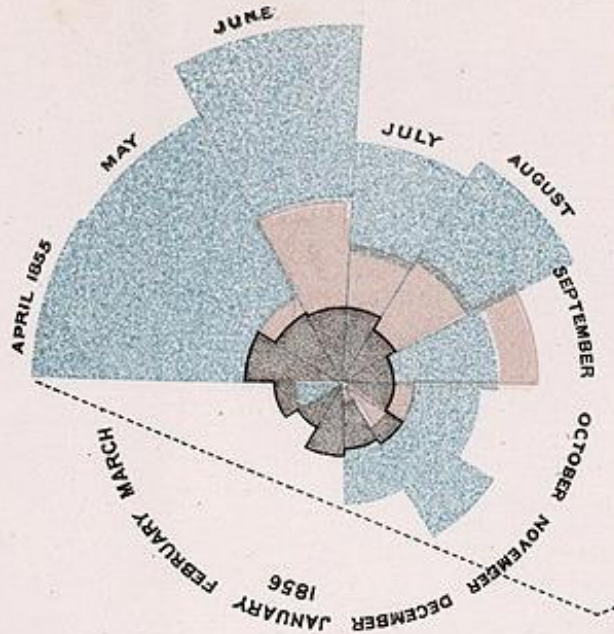
Assignment	Due Date	Points
Join Slack for Class Communication	Wednesday, December 2	1
Data Collection Assignment	Monday, December 8	5
Data Management Assignment	Monday, December 15	5
Data Visualization Assignment	Monday, February 02	5
Final Exam (Multiple Choice)	Monday, February 23 @ 10:40 AM (Building 103, Room S89 & S56) Thursday, March 23 @ 15:40 PM (Building 103, S56)	30
	Total Points	46

Recap Data Cleaning and Feature Engineering

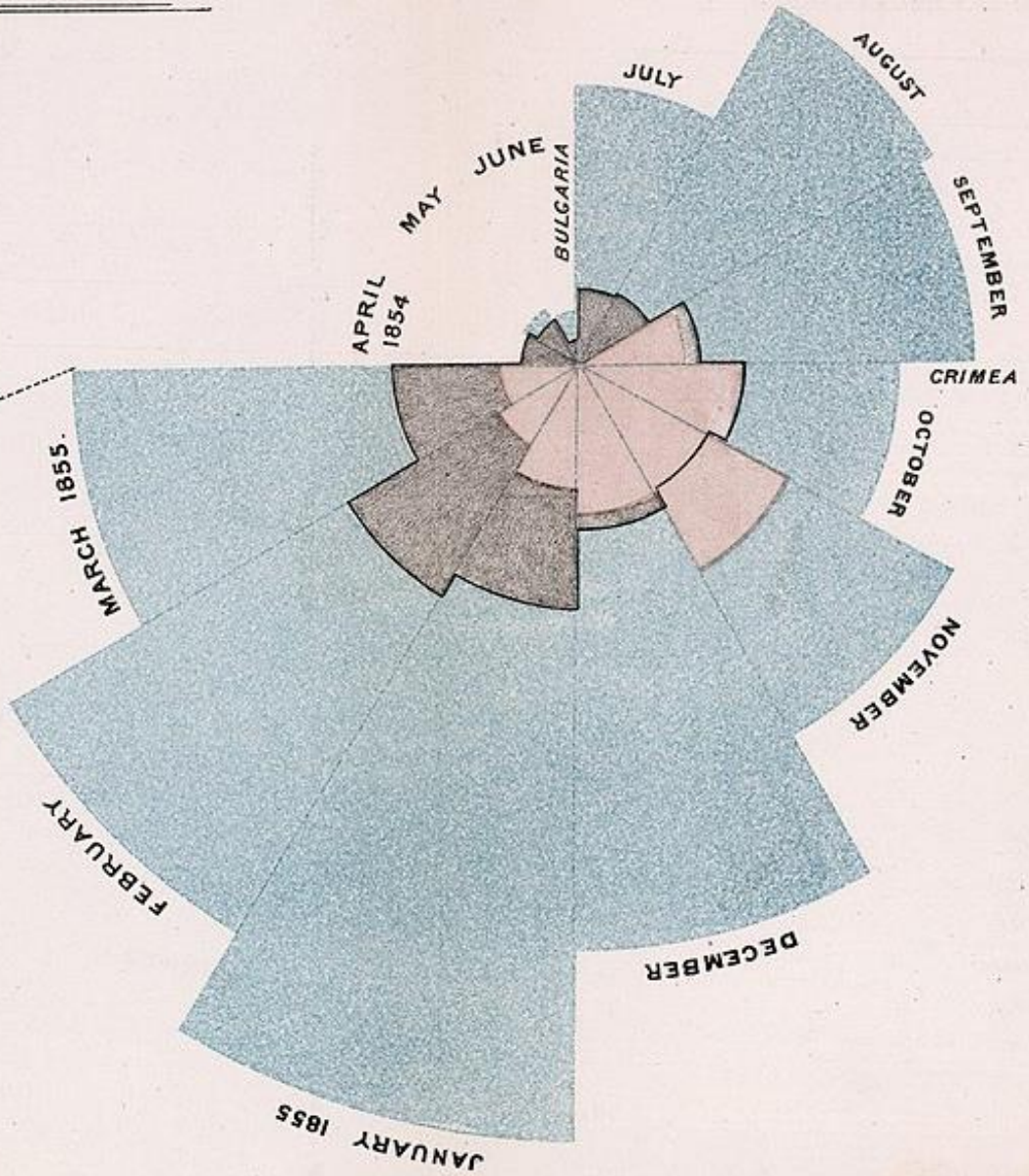

Data Visualization
An Introduction

DIAGRAM of the CAUSES of MORTALITY IN THE ARMY IN THE EAST.

2.
APRIL 1855 to MARCH 1856.



1.
APRIL 1854 to MARCH 1855.



The Areas of the blue, red, & black wedges are each measured from the centre as the common vertex.

The blue wedges measured from the centre of the circle represent area for area the deaths from Preventible or Mitigable Zymotic diseases, the red wedges measured from the centre the deaths from wounds, & the black wedges measured from the centre the deaths from all other causes.

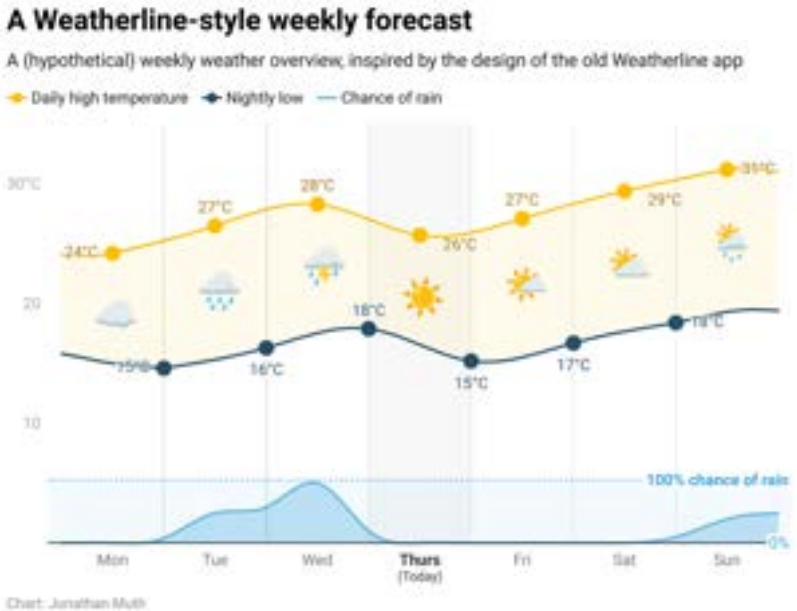
The black line across the red triangle in Novr. 1854 marks the boundary of the deaths from all other causes during the month.

In October 1854, & April 1855, the black area coincides with the red, in January & February 1856, the blue coincides with the black.

The entire areas may be compared by following the blue, the red & the black lines enclosing them.

Harrison & Sons, St. Martin's Lane.

Everyday examples of data visualizations



From overwhelming tables to understanding

	A	B	C	D	E	F
1	Order #	First Name	Last Name	Email	Country	IP address
2	1	Dalton	Kramer	dalton@email.com	France	211.91.226.1
3	2	Gita	Tetterton	gita@email.com	USA	222.153.179.
4	3	Weston	Jurgens	weston@email.com	Spain	203.123.236.
5	4	Brad	Chupp	brad@email.com	France	202.183.111.
6	5	Marybeth	Baumann	marybeth@email.com	Italy	214.132.168.
7	6	Allyson	Feder	allyson@email.com	Italy	182.108.190.
8	7	Lucile	Folks	lucile@email.com	Greece	18.64.161.62
9	8	Mickey	Rusk	mickey@email.com	Canada	40.18.115.20
10	9	Clarine	Esslinger	clarine@email.com	Greece	185.134.23.8
11	10	Kimberly	Penny	kimberly@email.com	France	34.72.165.11
12	11	Colleen	Kellough	colleen@email.com	USA	73.51.152.18
13	12	Nettie	Edmonds	nettie@email.com	Spain	94.133.138.2
14	13	Duncan	Rickenbacker	duncan@email.com	France	211.91.226.1
15	14	Marchelle	Diedrich	marchelle@email.com	Italy	222.153.179.
16	15	Mariano	Murrell	mariano@email.com	Italy	203.123.236.1

Teenage Movie Ratings

File Edit View Insert Format Data Tools Add-ons Help All changes saved in Drive

Comments Share

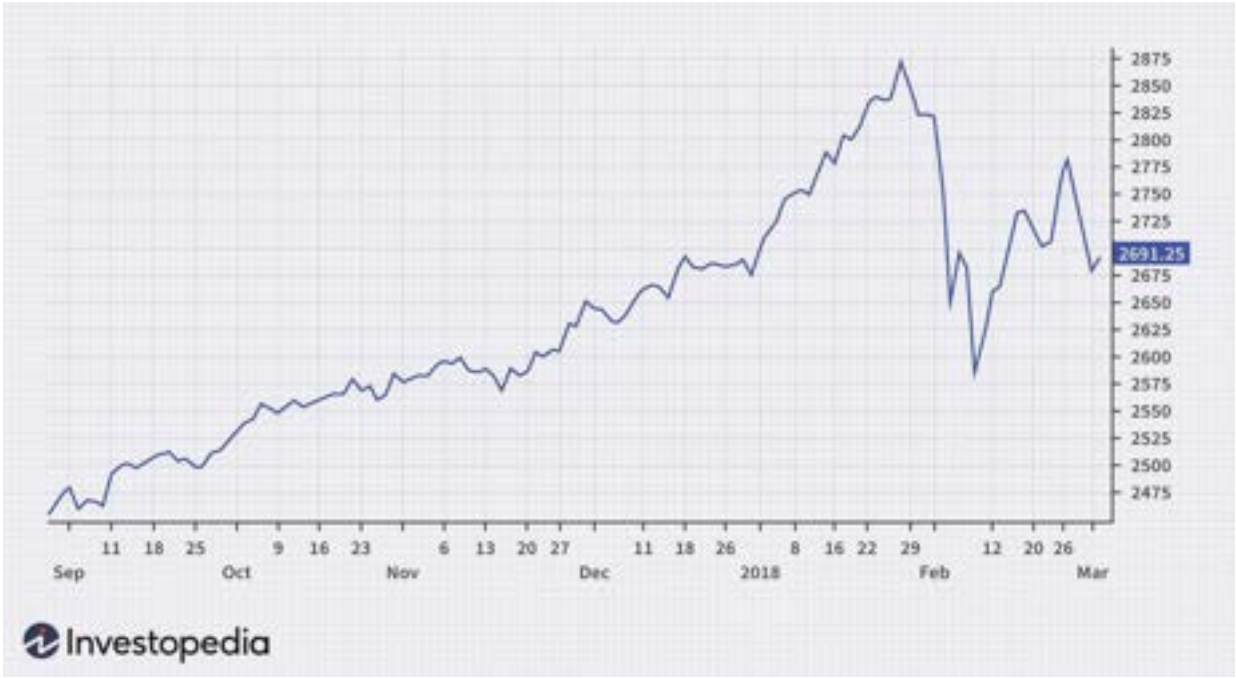
	A	B	C	D	E	F	G	H
1		F		M		Grand Total		
2		COUNT of rating	AVERAGE of rat	COUNT of rating	AVERAGE of rat	COUNT of rating	AVERAGE of rating	
3	Air Force One	5	3.2	8	3.875	13	3.615384615	
4	Chasing Amy	3	4.333333333	10	3.9	13	4	
5	Contact	7	3.428571429	13	4.384615385	20	4.05	
6	Courage Under I	5	3.8	6	3.5	11	3.636363636	
7	E.T. the Extra-Te	4	4	6	3.666666667	10	3.8	
8	Evita	6	3.5	5	3.2	11	3.363636364	
9	Fargo	2	4	10	4.5	12	4.416666667	
10	Game, The	5	3.8	7	4.714285714	12	4.333333333	
11	Independence D	7	4.571428571	7	3.285714286	14	3.928571429	
12	Liar Liar	7	3.142857143	9	3.111111111	16	3.125	
13	Mission: Impossi	3	4.333333333	9	3.777777778	12	3.916666667	
14	Phenomenon	6	3.5	7	3.571428571	13	3.538461538	
15	Return of the Jec	3	4.666666667	15	4.6	18	4.611111111	
16	Rock, The	4	4	7	4.285714286	11	4.181818182	
17	Saint, The	5	4.2	6	3	11	3.545454545	
18	Scream	10	4.4	16	4	26	4.153846154	
19	Star Wars	6	4.833333333	16	4.6875	22	4.727272727	
20	Titanic	5	4.8	7	4.571428571	12	4.666666667	
21	Toy Story	5	4.6	8	3	13	3.615384615	
22	Twelve Monkeys	2	2.5	9	4.333333333	11	4	
23	Willy Wonka and	4	4	7	3.714285714	11	3.818181818	
24	Grand Total	104	4	188	3.994680851	292	3.996575342	

Sheet1 Pivot Table 6 Sheet5 Explore

99 477 Card Fast Delivered

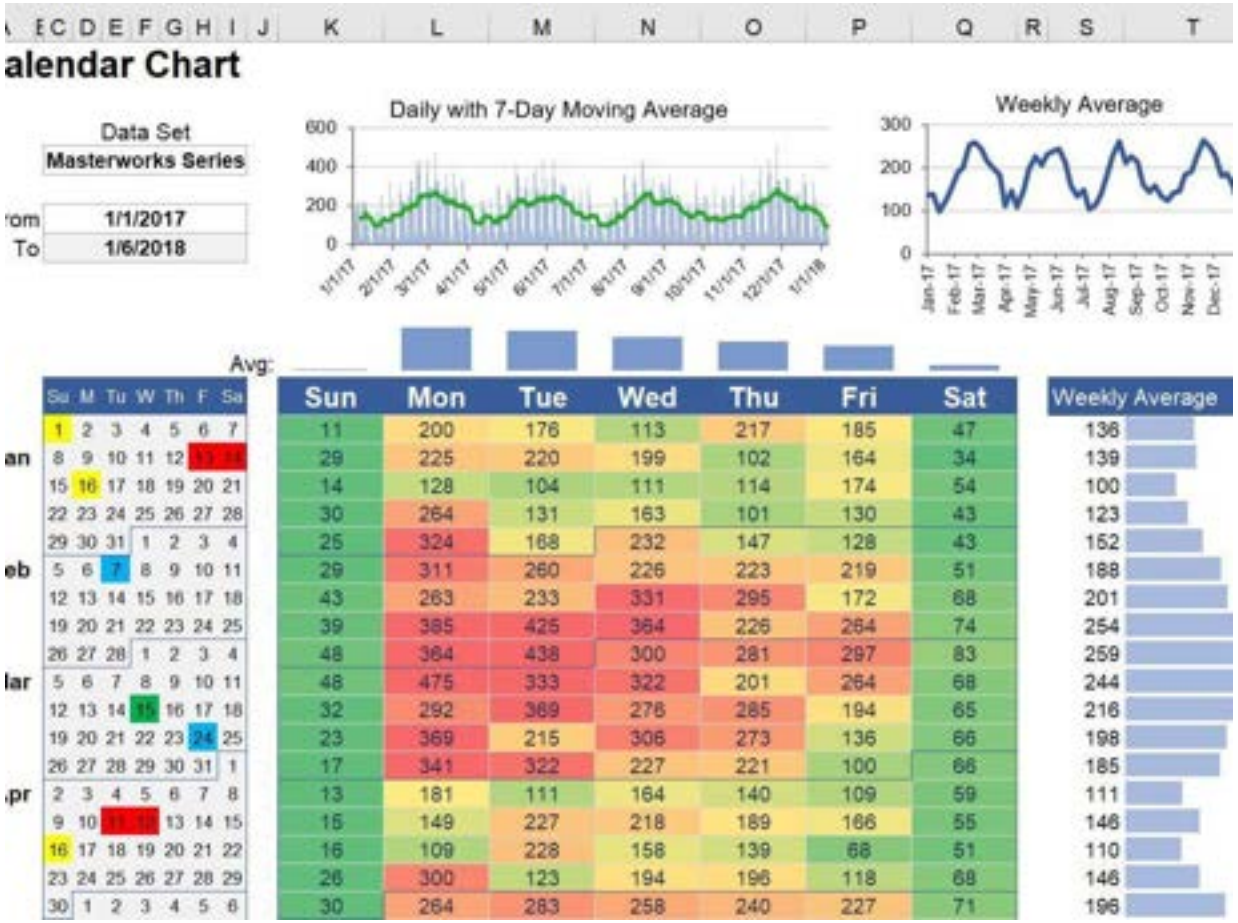
Communicate data clearly & effectively

	A	B	C	D	E	F	G
1			Stock Prices in million USD				
2	Date	Volume	High Value	Low Value	Closing Value		
3	21-Jul	4500	$24.21	$22.66	$21.16		
4	22-Jul	4800	$25.46	$22.49	$23.29		
5	23-Jul	5000	$25.99	$23.56	$24.44		
6	24-Jul	5100	$29.42	$24.00	$26.59		
7	25-Jul	6000	$30.23	$26.97	$32.00		
8							
9							
10							
11							
12							
13							
14							
15							
16							
17							
18							
19							
20							
21							



Identify trends and patterns

Sales Table							
No.	Date	Client	PID	U.P.	Qty.	Sales	Add.
0001	17/11/1	A Company	A-156	120	2	240	
0002	17/11/2	B Company	C-001	80	4	320	
0003	17/11/3	C Company	S-453	150	2	300	
0001	17/11/6	A Company	A-301	90	5	450	
0003	17/11/7	C Company	S-125	130	2	260	
0005	17/11/8	E Company	Z-120	560	1	560	
0006	17/11/9	F Company	F-021	320	5	1600	
0003	17/11/10	C Company	S-136	75	10	750	
0007	17/11/11	G Company	G-980	50	15	750	
0001	17/11/12	A Company	A-157	60	9	540	
0007	17/11/13	G Company	G-910	120	2	240	
0008	17/11/14	H Company	E-365	90	13	1170	
0003	17/11/7	C Company	S-125	130	2	260	
0005	17/11/8	E Company	Z-120	560	1	560	
0001	17/11/4	A Company	A-201	650	1	650	
0004	17/11/5	D Company	B-150	200	3	600	
0006	17/11/15	F Company	F-027	35	15	525	

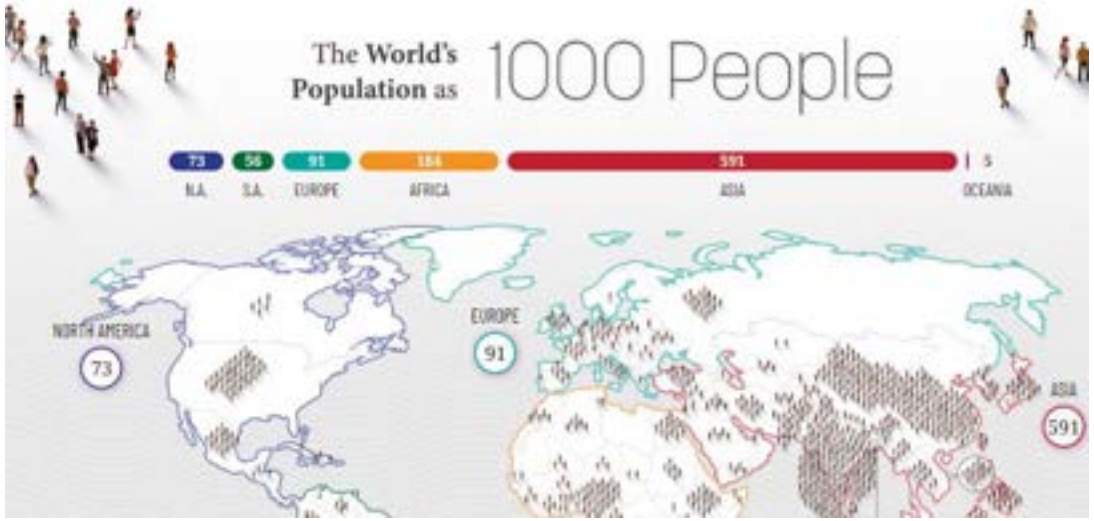


Why use data visualization?

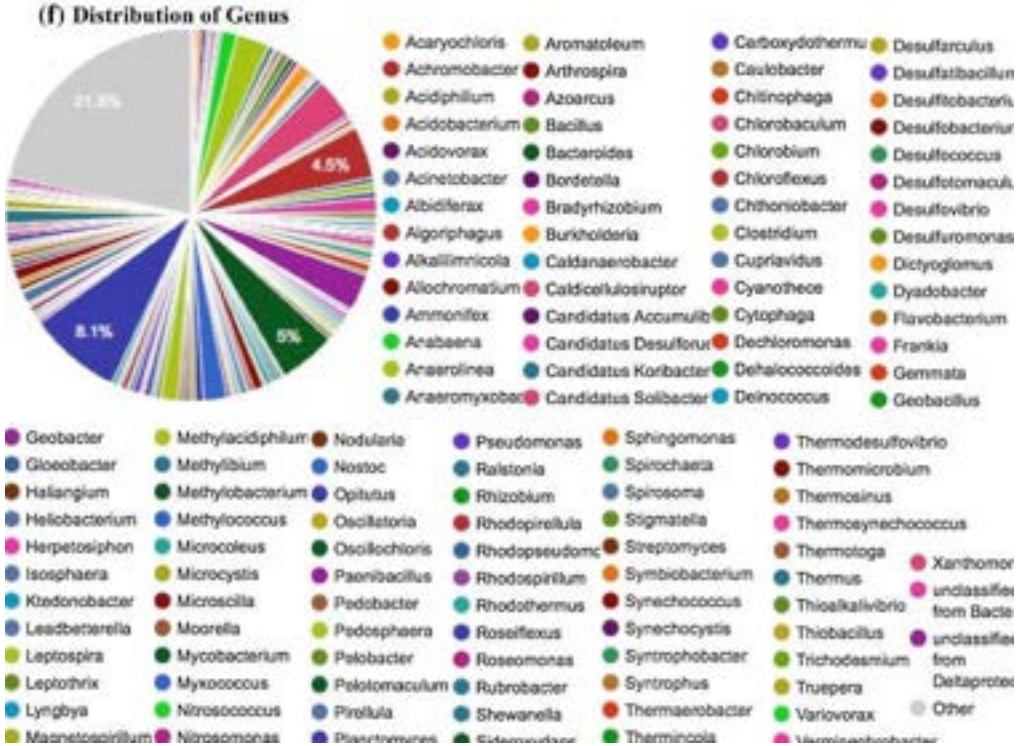
1. Make data easier to understand and remember
2. Discover unknown facts, outliers, and trends
3. Visualize relationships and patterns quickly
4. Ask better questions and make better decisions

Explore and Report (on Slack)

Cool Visualizations



Ugly (or not useful) Visualizations



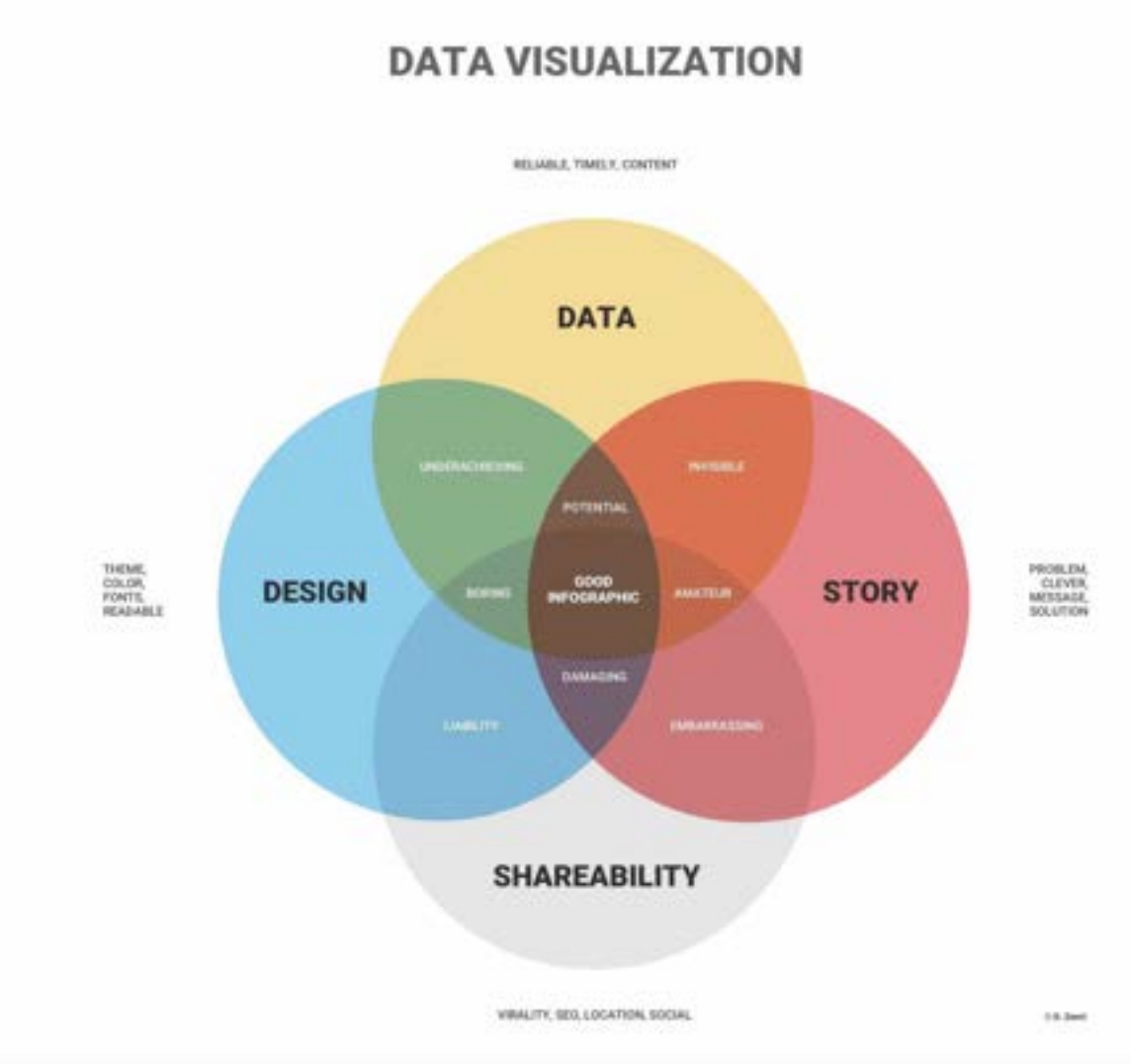
What makes a visualization cool or ugly? Useful?

“Maybe stories are just data with a soul.”

— Brené Brown

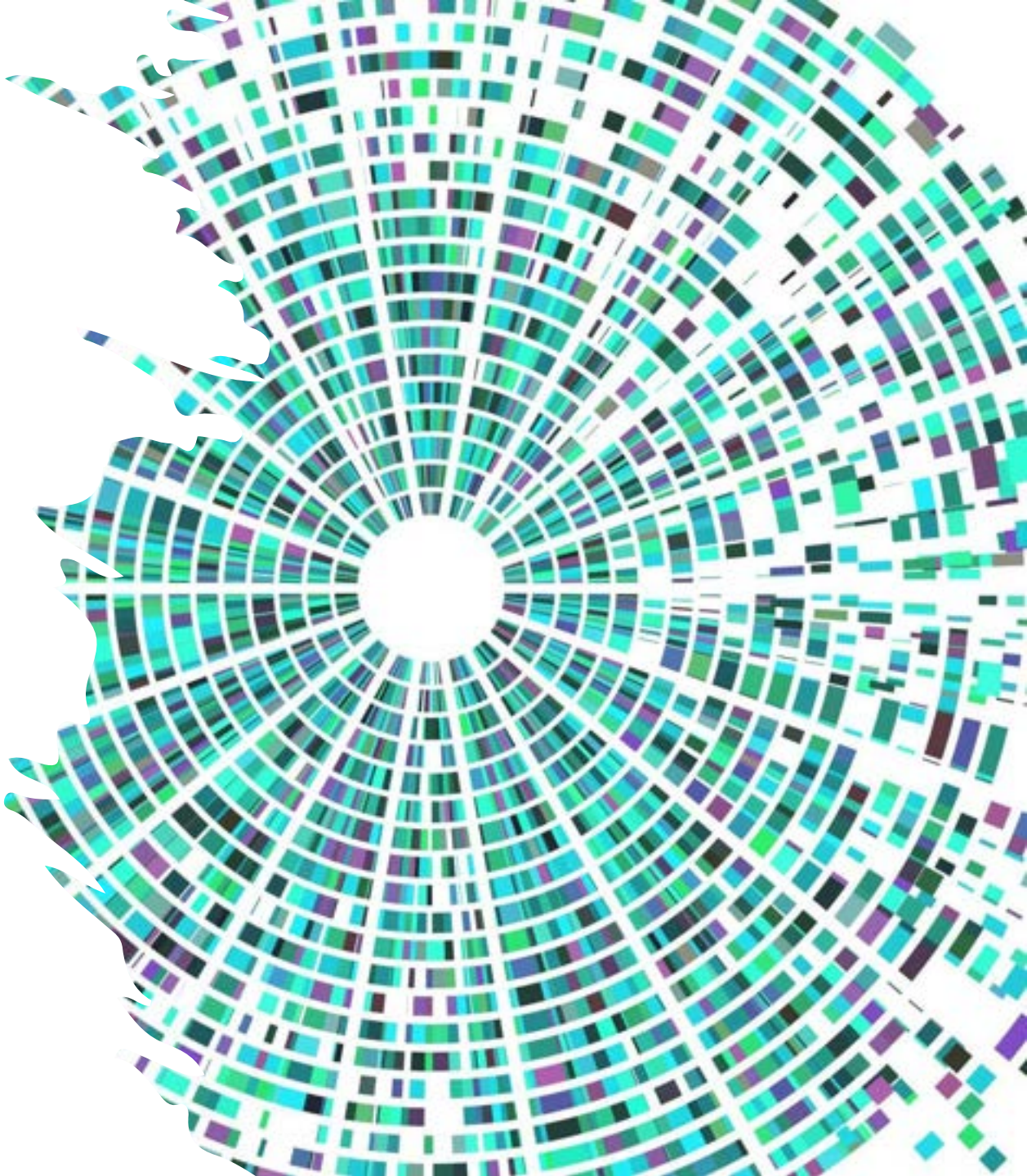
What Makes a Good Data Visualization?

- Good data visualizations are created when communication, data science, and design collide.
- Data visualizations done right offer key insights into complicated datasets in ways that are meaningful and intuitive.



Data Visualization

‘Complex ideas communicated with clarity, precision, and efficiency.’ E. Tufte (Yale Professor)



Data Visualization Principles:

1. Reduce clutter
2. Create order
3. Give focus

1. Reduce 'clutter' - visual noise

- Remove all unnecessary elements.
- Empty spaces in visualizations are as important as pauses in speaking.

Group	Metric A	Metric B	Metric C
Group 1	$X.X	Y%	Z,ZZZ
Group 2	$X.X	Y%	Z,ZZZ
Group 3	$X.X	Y%	Z,ZZZ
Group 4	$X.X	Y%	Z,ZZZ
Group 5	$X.X	Y%	Z,ZZZ

1. Reduce 'clutter' - visual noise

- Remove all unnecessary elements.
- Empty spaces in visualizations are as important as pauses in speaking.

Group	Metric A	Metric B	Metric C
Group 1	$X.X	Y%	Z,ZZZ
Group 2	$X.X	Y%	Z,ZZZ
Group 3	$X.X	Y%	Z,ZZZ
Group 4	$X.X	Y%	Z,ZZZ
Group 5	$X.X	Y%	Z,ZZZ

“Whitespace isn’t just empty—it gives the eye a break and creates balance in the design.”

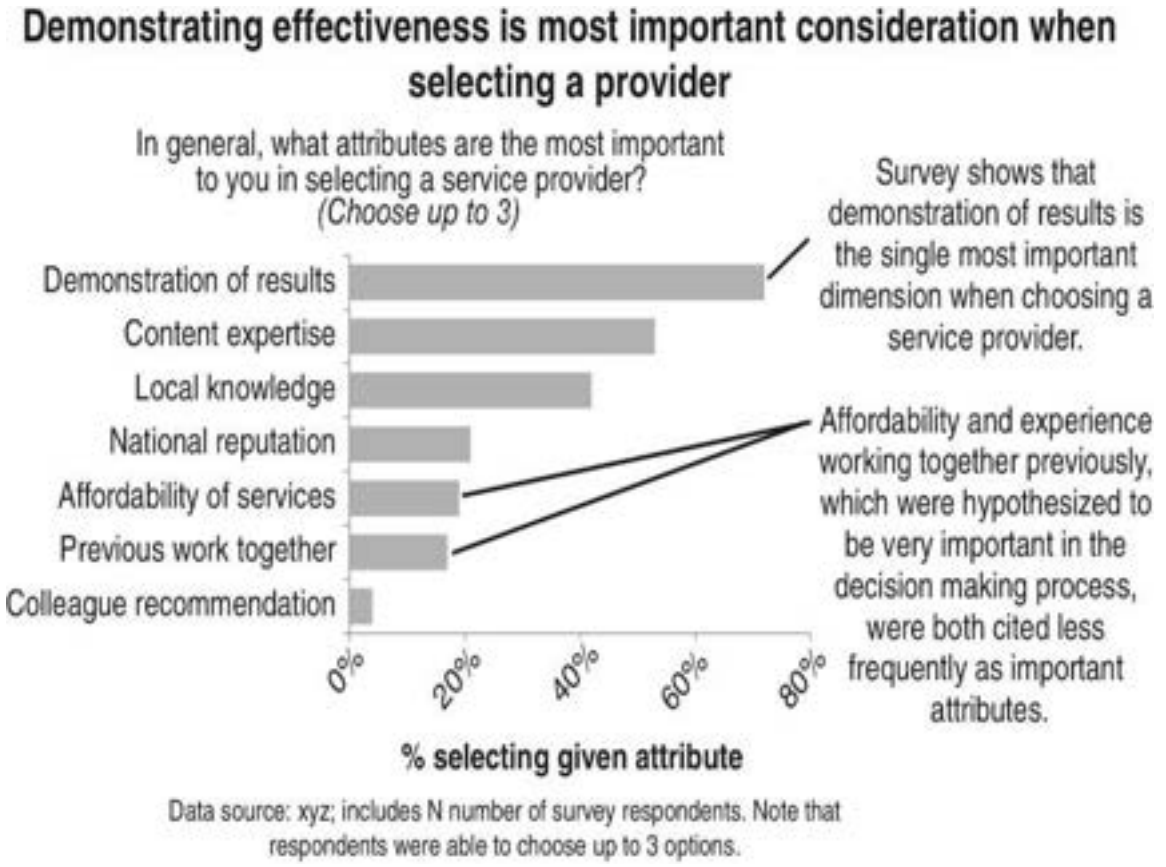
Cluttered Example: Monthly Sales by Region

Month	North	South	East	West	Online	Retail	Total
Jan	1137	1274	1411	1548	1685	1822	1959
Feb	1274	1548	1822	2096	2370	2644	2918
Mar	1411	1822	2233	2644	3055	3466	3877
Apr	1548	2096	2644	3192	3740	4288	4836
May	1685	2370	3055	3740	4425	5110	5795
				4288	5110	5932	6754
				4836	5795	6754	7713

Download **declutter.pptx** and create order. Post your solution on our Slack channel

2. Create order

- Align elements along implicit lines.
- Again: Empty spaces in visualizations are as important as pauses in speaking

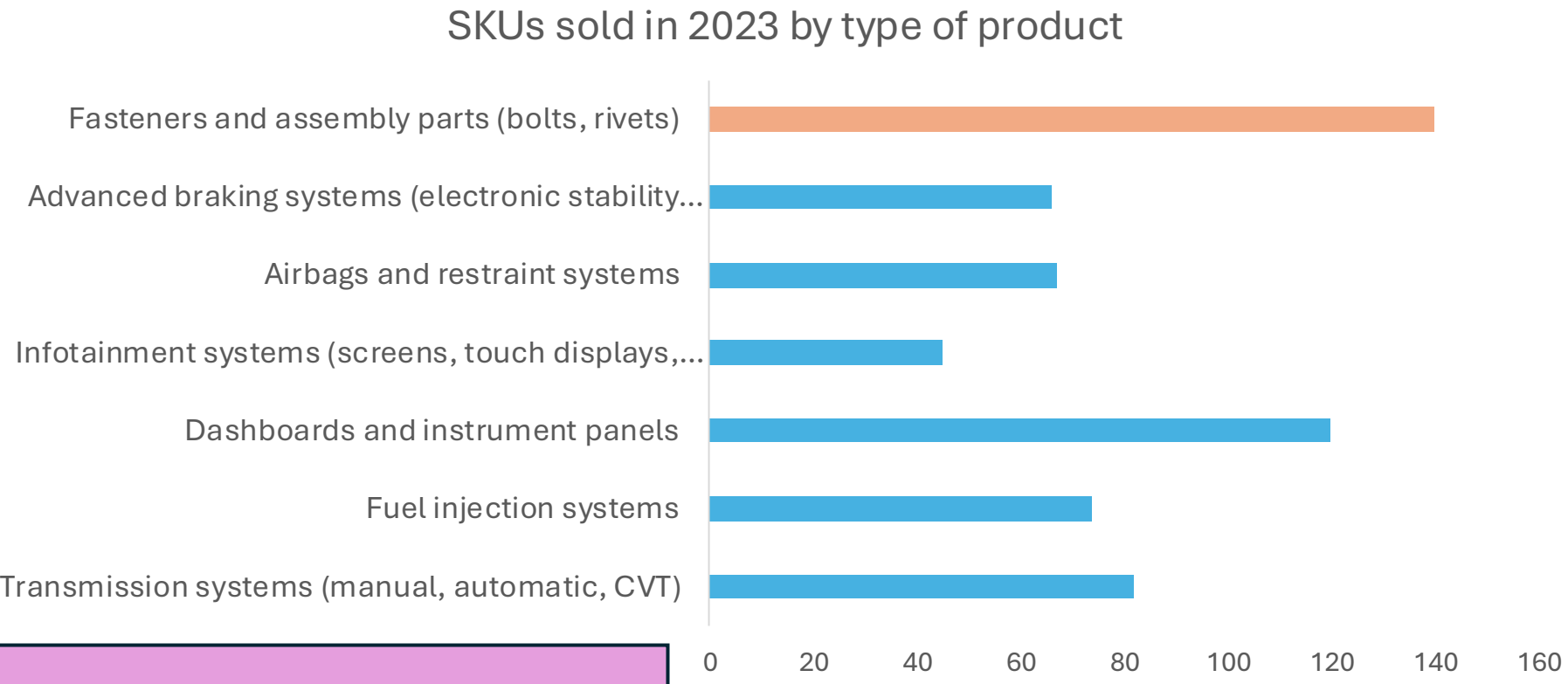


2. Create order

- Align elements along implicit lines.
- Again: Empty spaces in visualizations are as important as pauses in speaking

Fasteners and assembly parts were our best selling products in 2023

Infotainment systems were the least sold products

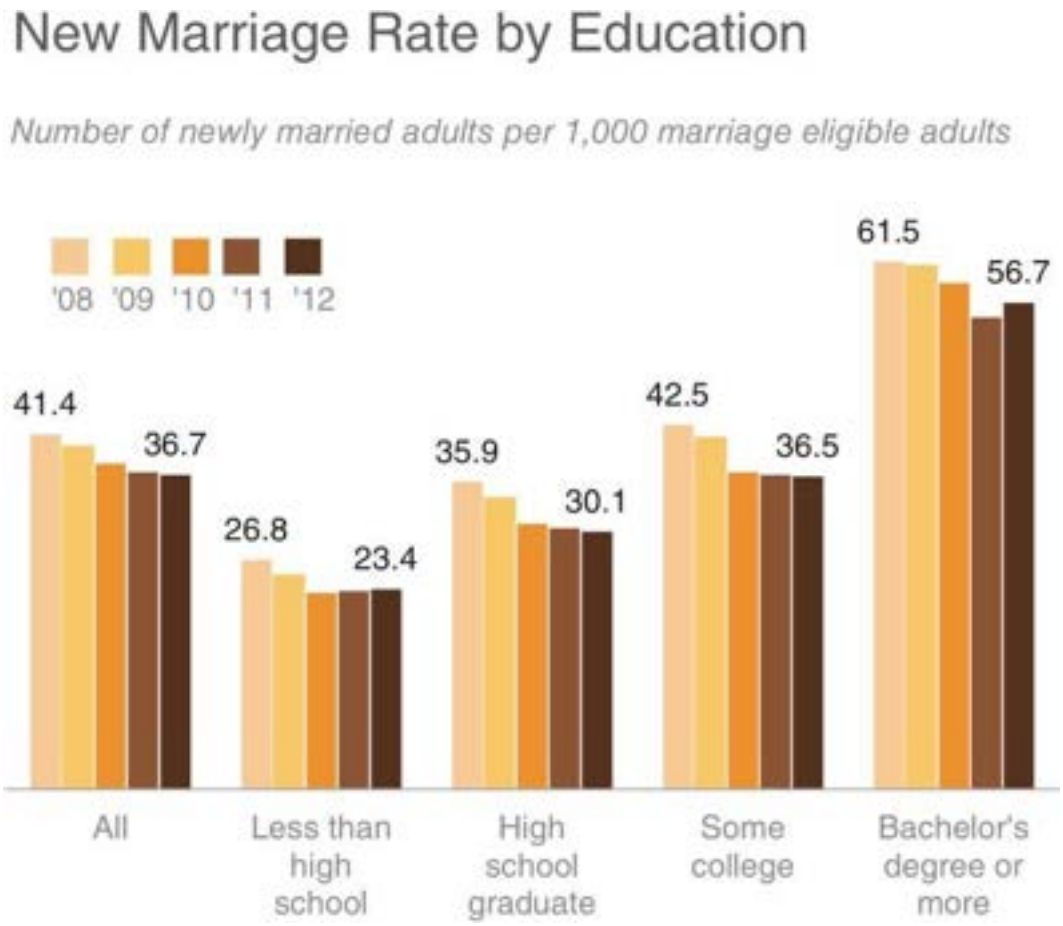


Download **order.pptx** and create order. Post your solution on our Slack channel

Best and least selling products in 2023

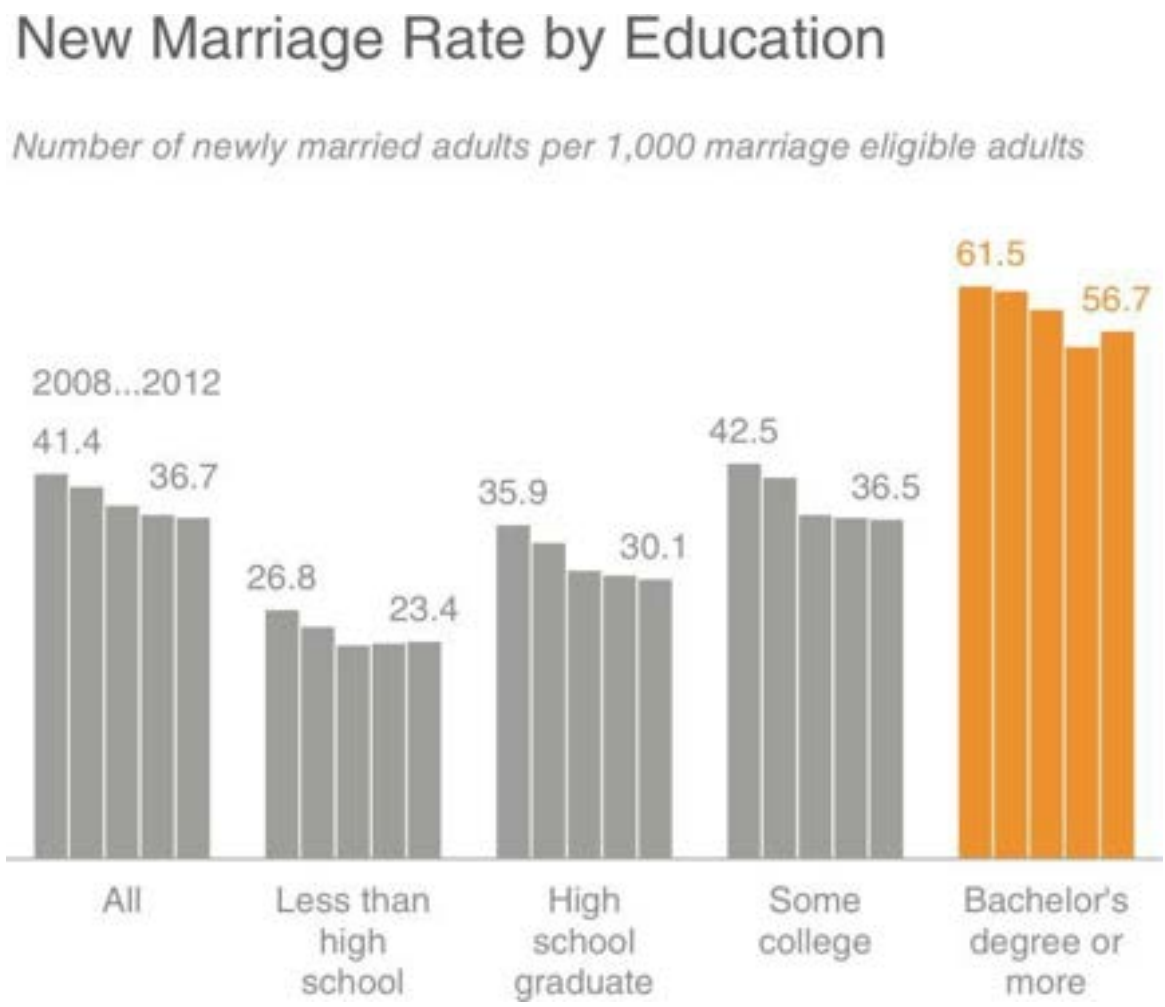
3. Give focus

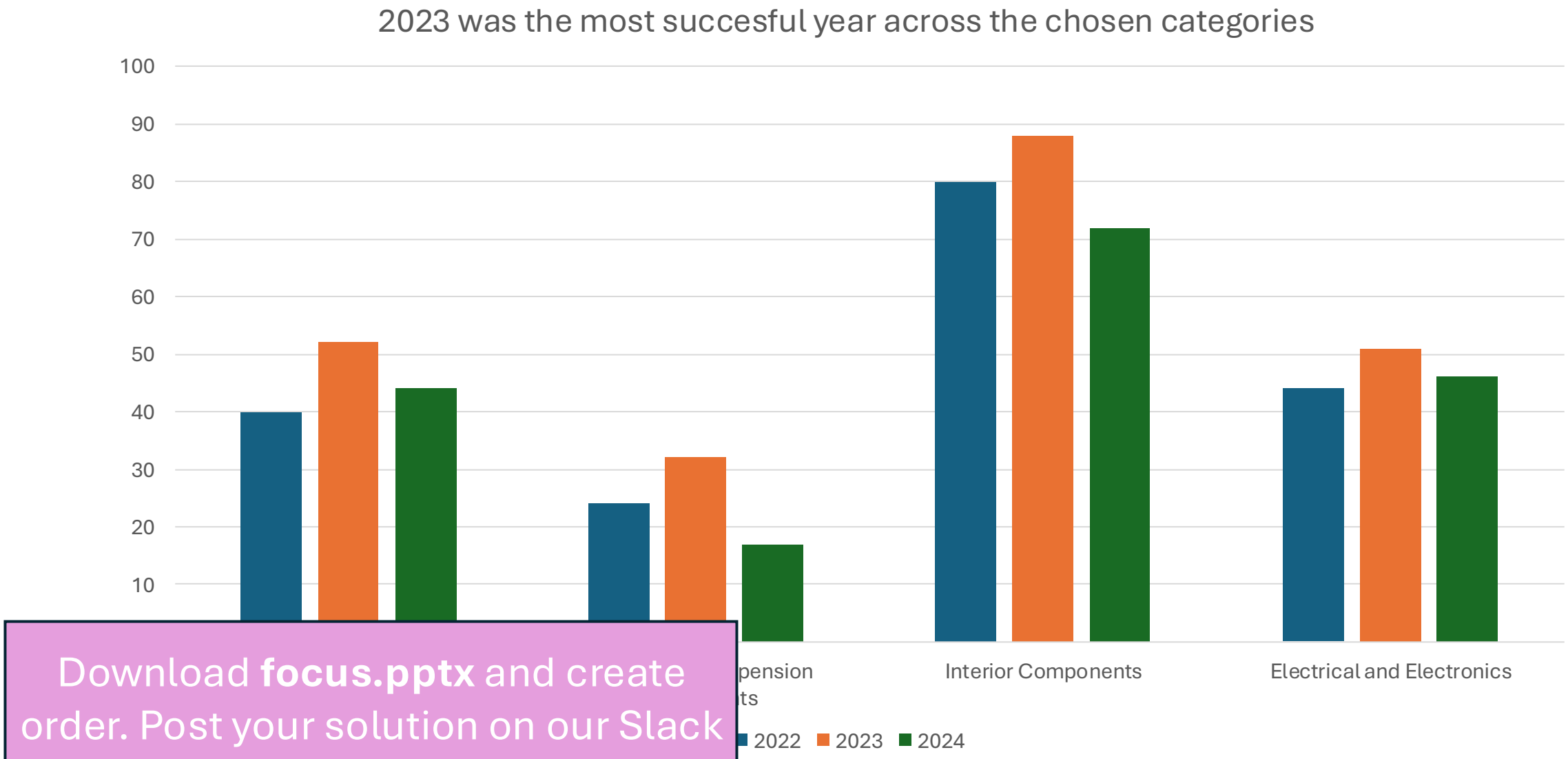
- Use few colors and use them strategically (applies also to contrast).
- Choose colors that are appropriate and pleasing.
- Be consistent.



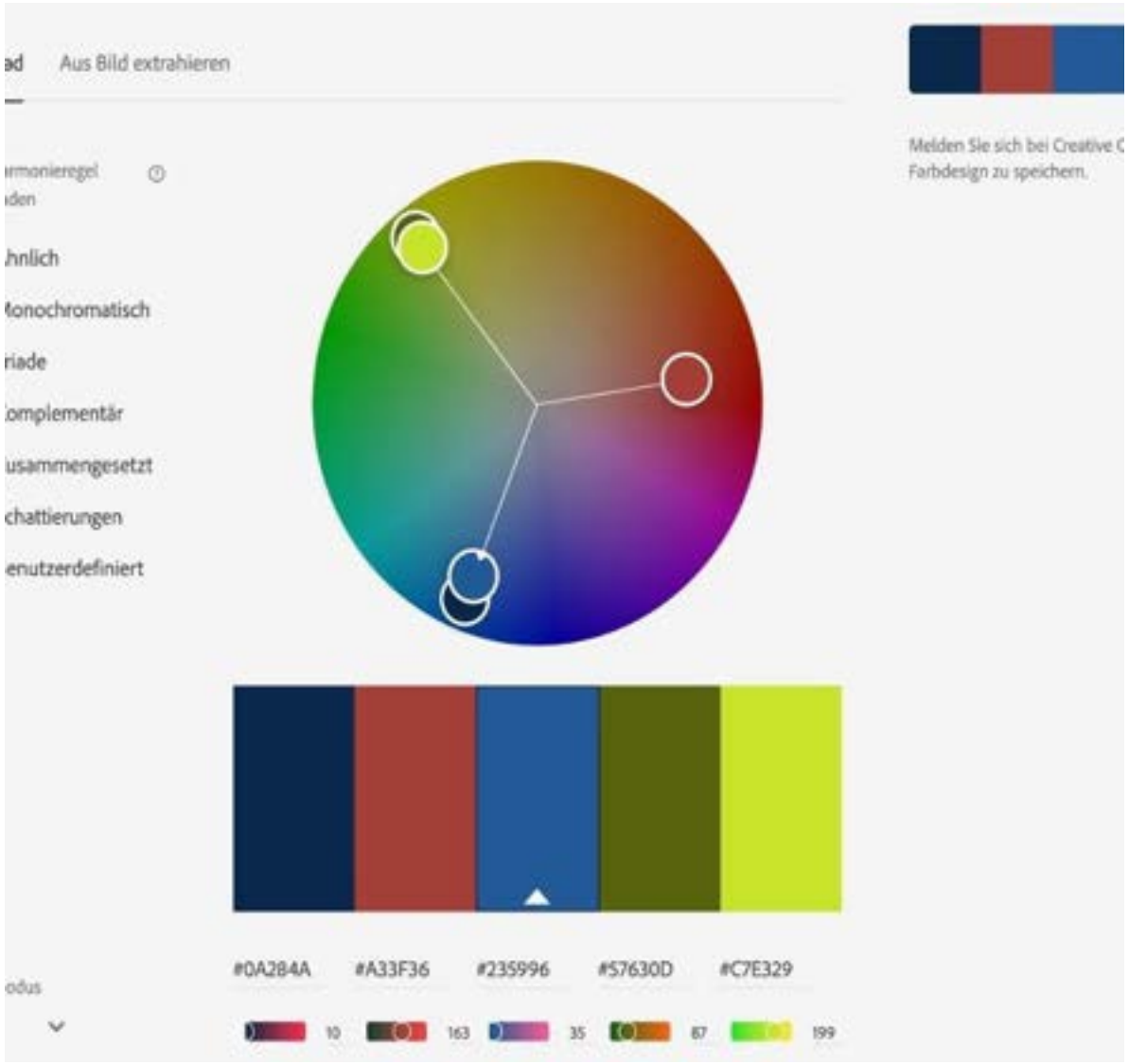
3. Give focus

- Use few colors and use them strategically (applies also to contrast).
- Choose colors that are appropriate and pleasing.
- Be consistent.





Download **focus.pptx** and create order. Post your solution on our Slack channel



3. Give focus

- Use few colors and use them strategically (applies also to contrast).
- Choose colors that are appropriate and pleasing.
- Tools:
  - **color.adobe.com**
  - coolors.co
  - install.packages("viridis")
- Be consistent.

3. Give focus

- Use few colors and use them strategically (applies also to contrast).
- Choose colors that are appropriate and pleasing.
  - color.adobe.com
  - **coolors.co**
  - install.packages("viridis")
- Be consistent.

Who?

- Think about your Audience
  - Prior knowledge Expectations?
  - What do you want your audience to know or do?
  - What tone do you want your communication to set?
  - What biases does your audience have?
  - Is your audience familiar with this data?

Memory and attention span of humans are limited.

- **Visual attention** is the cognitive process that mediates the **selection** of **important information** from the environment.

- This selection is usually **controlled** by **bottom-up and top-down** attentional biasing.

Information processing is strongly influenced by:

- Top-down bias: expectations and prior knowledge - voluntary guidance of attention by internal goals
- Bottom-up bias: visual characteristica - involuntarily capture of attention by salient events in the environment.

And you will read this last

You will read this first

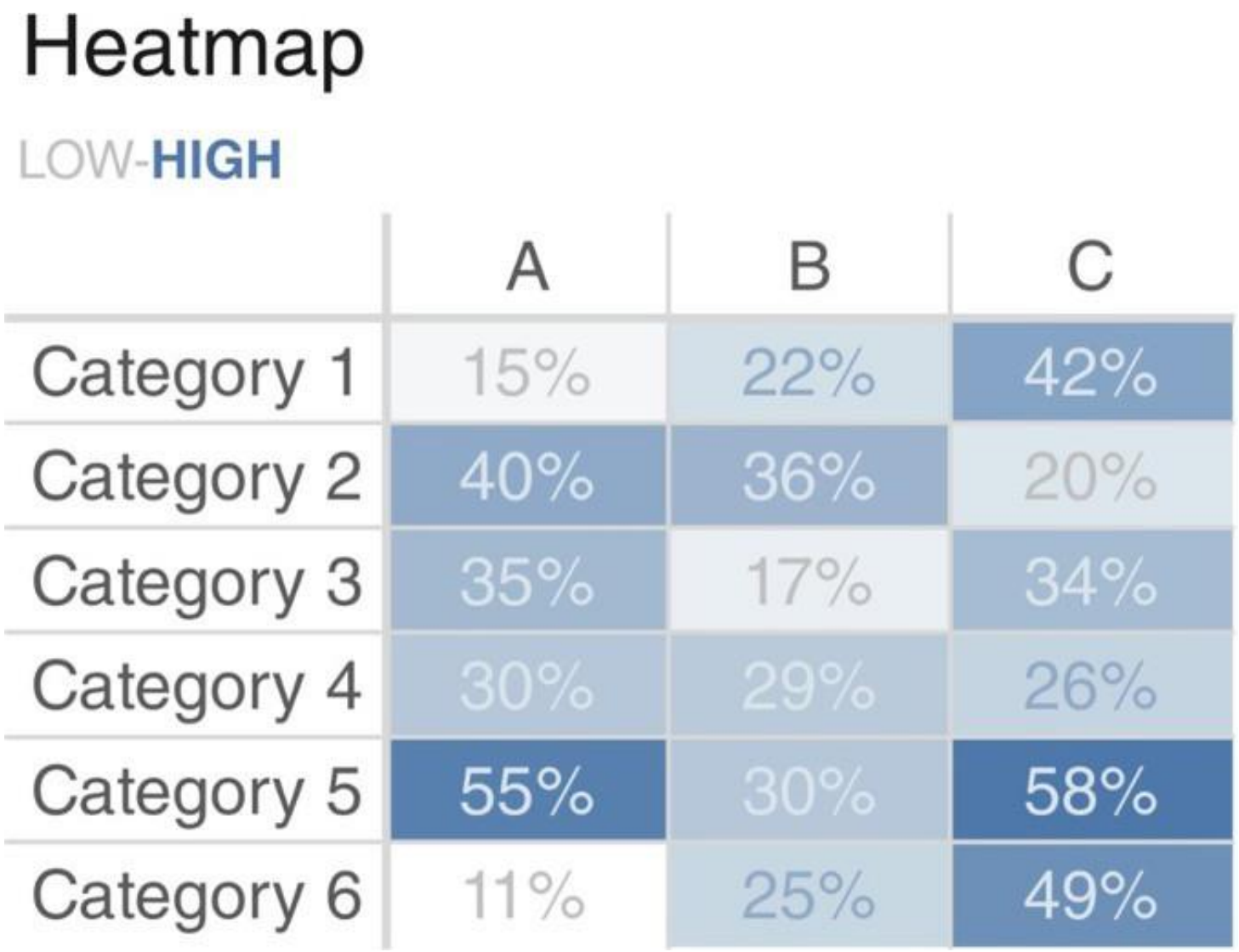
And then you will read this

Then this one

Visual attention is by definition a selective process

Tables

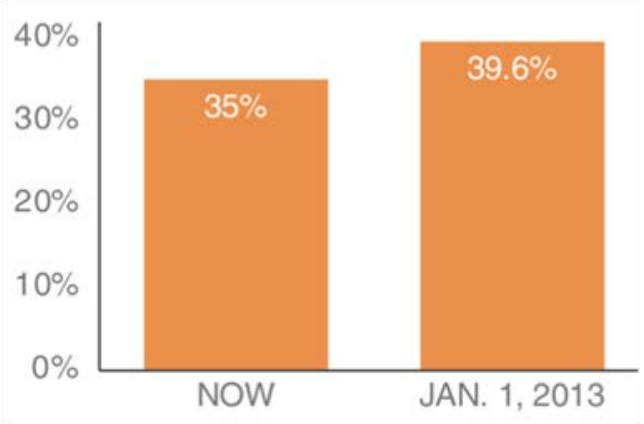
- Are precise and complete
- Are processed serially and need time.
- Decluttered tables are easier to read.
- Colors can help with giving focus.



Charts

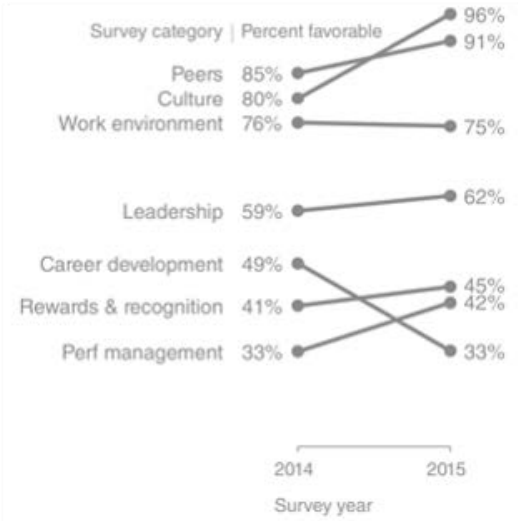
- Are often easier to understand than tables.
- Are often more difficult to create than tables.

Barplot

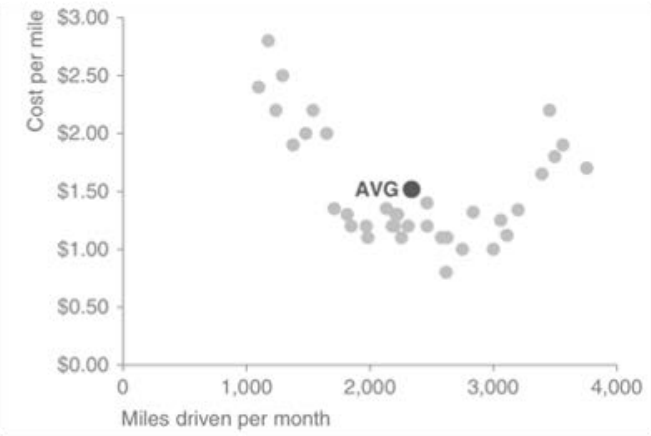


Lineplot

Slope plot

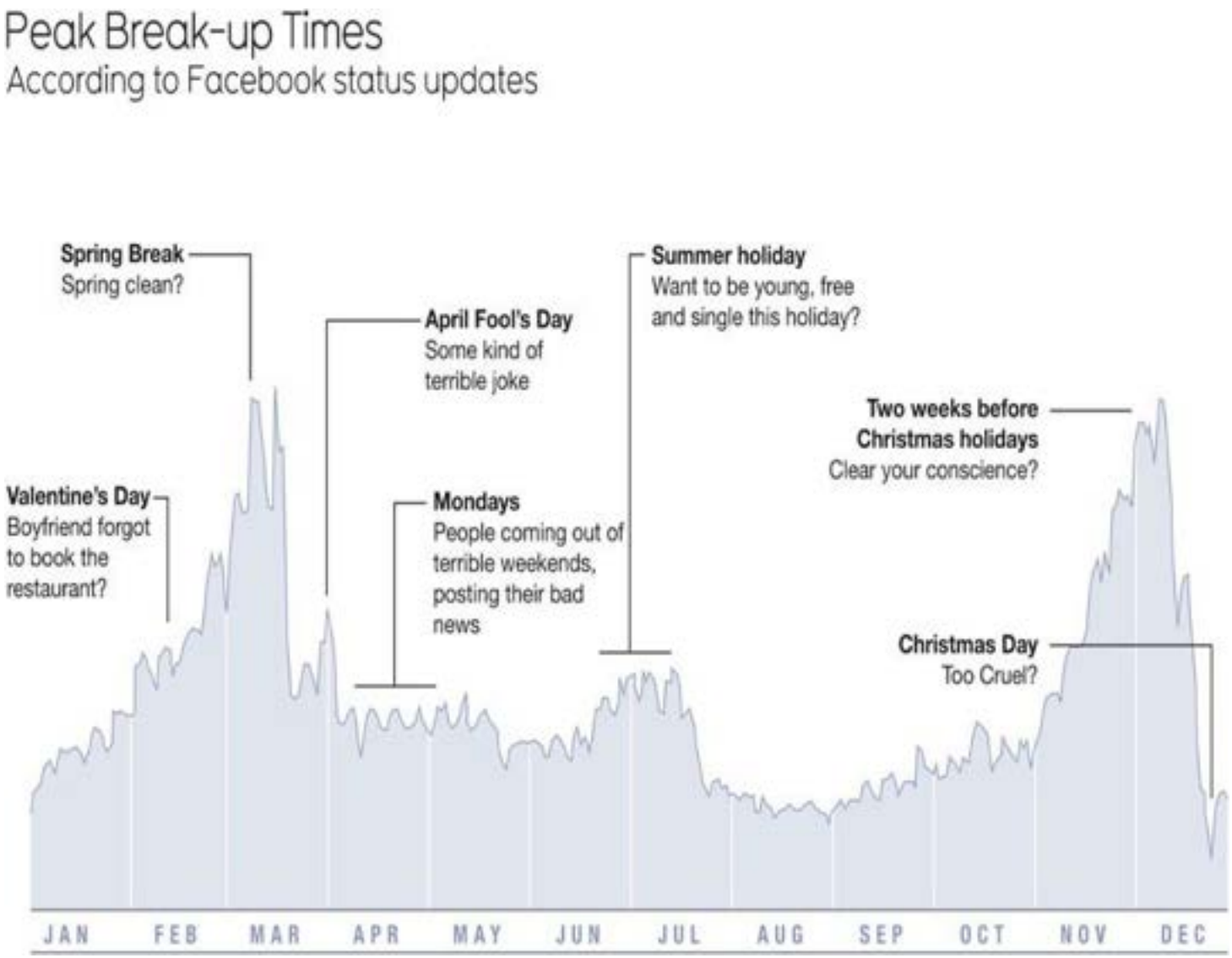


Scatterplot



Annotation

- Annotation is very important
- Axes titles
  Plot title
  Subtitle
  Legend
- Annotations in the plot



Colors

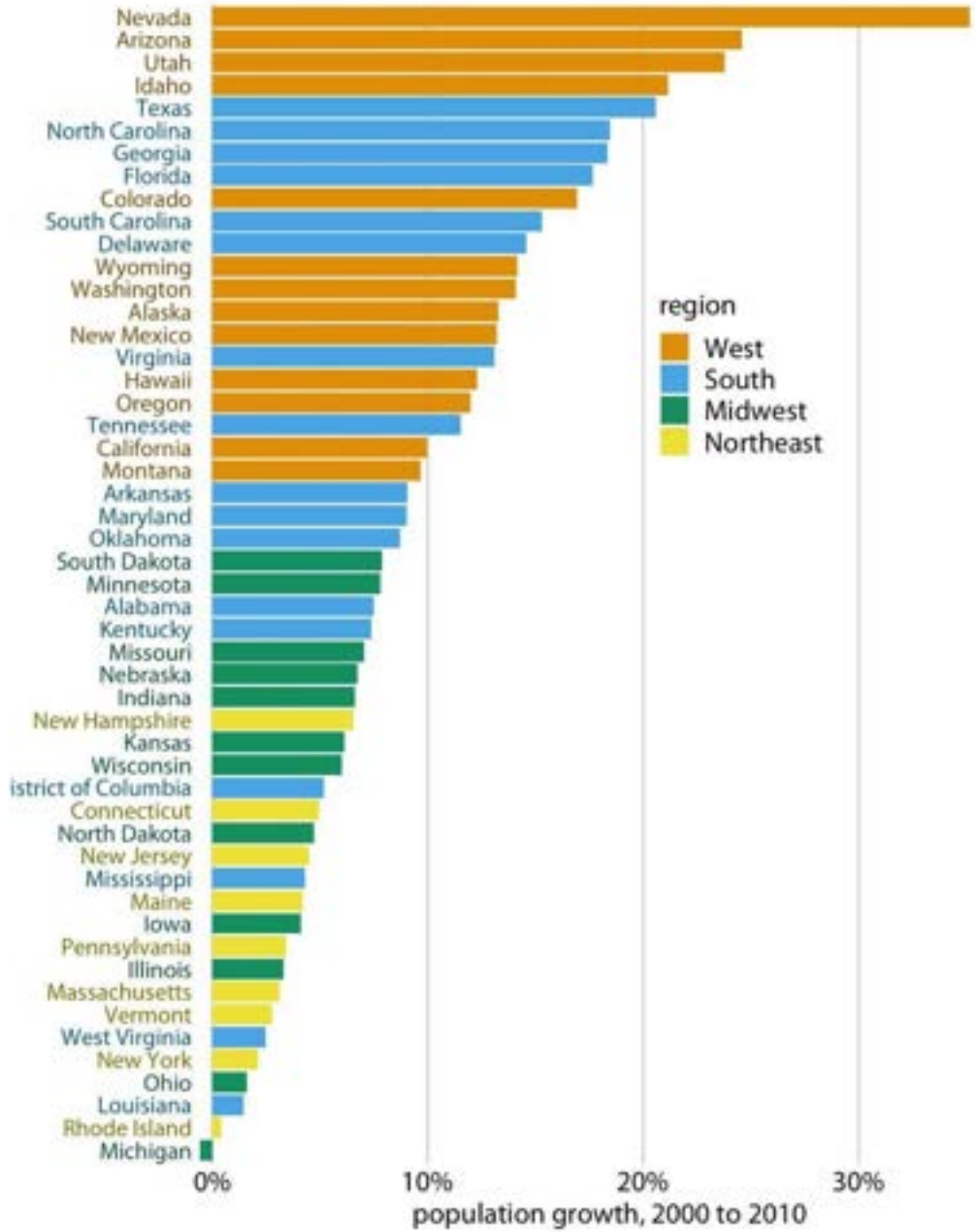
- Colors as a Tool to Distinguish: Qualitative color scale. Chosen to look clearly distinct from each other

- Color to Represent Data Values: Quantitative data values, such as income and temperature. A sequence of colors that indicate which values are larger or smaller. Sometimes we need to visualize the deviation of data values in one of two directions. We may want to show those different colors, so that it is clear whether a value is positive or negative

- Color as a Tool to Highlight: elements in a color or set of colors that stand out against the rest of the figure— achieved with accent color scales.

Source : Fundamentals of Data Visualization by Claus O. Wilke

Colors as a Tool to Distinguish: Qualitative color scale. Chosen to look clearly distinct from each other


Nevada
Arizona
Utah
Idaho
Texas
North Carolina
Georgia
Florida
Colorado
South Carolina
Delaware
Wyoming
Washington
Alaska
New Mexico
Virginia
Hawaii
Oregon
Tennessee
California
Montana
Arkansas
Maryland
Oklahoma
South Dakota
Minnesota
Alabama
Kentucky
Missouri
Nebraska
Indiana
New Hampshire
Kansas
Wisconsin
istrict of Columbia
Connecticut
North Dakota
New Jersey
Mississippi
Maine
Iowa
Pennsylvania
Illinois
Massachusetts
Vermont
West Virginia
New York
Ohio
Louisiana
Rhode Island
Michigan
region
West
South
Midwest
Northeast
0%
10%
20%
30%
population growth, 2000 to 2010

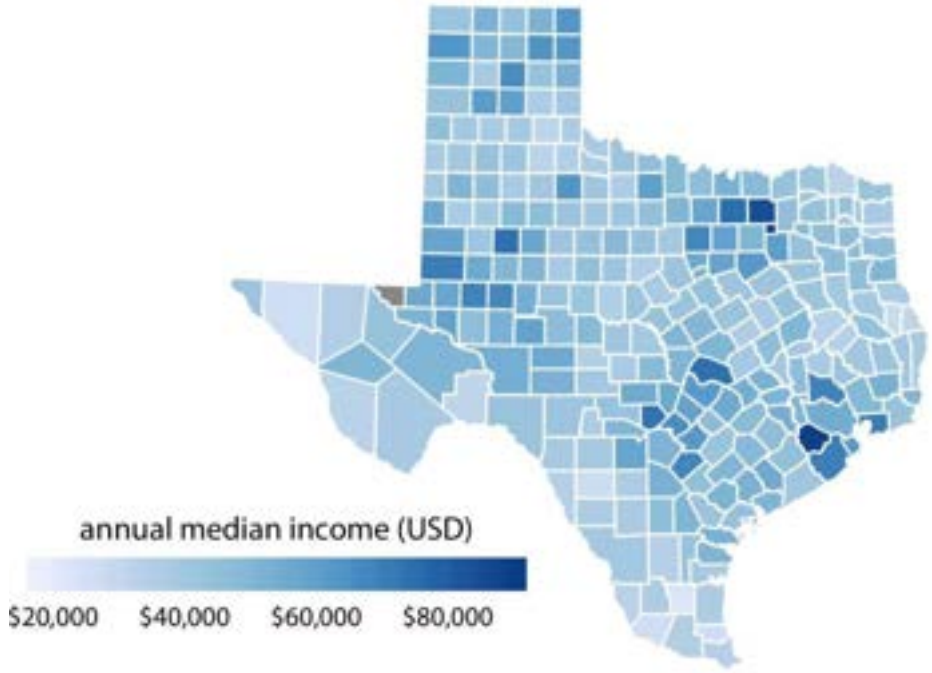
Color to Represent Data Values: Quantitative data values, such as income and temperature. Sequential colors can indicate which values are larger or smaller.

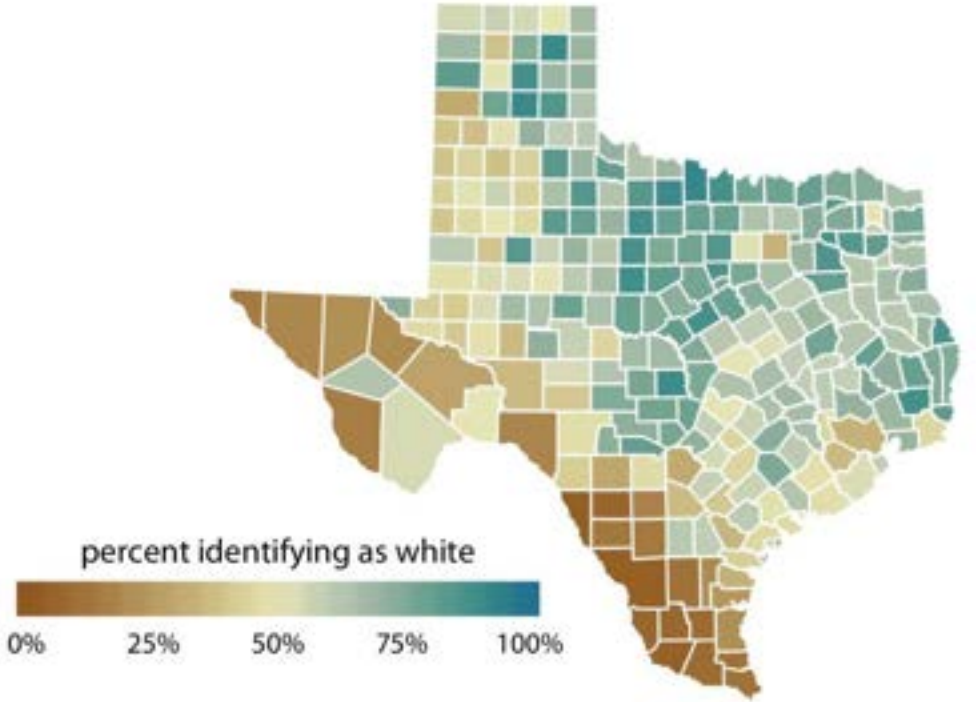
Sometimes we need to visualize the deviation of data values in one of two directions - use **Divergent colors.** We may want to show those different colors, so that it is clear whether a value is positive or negative

- **Sequential**

- **Divergent**

Colors

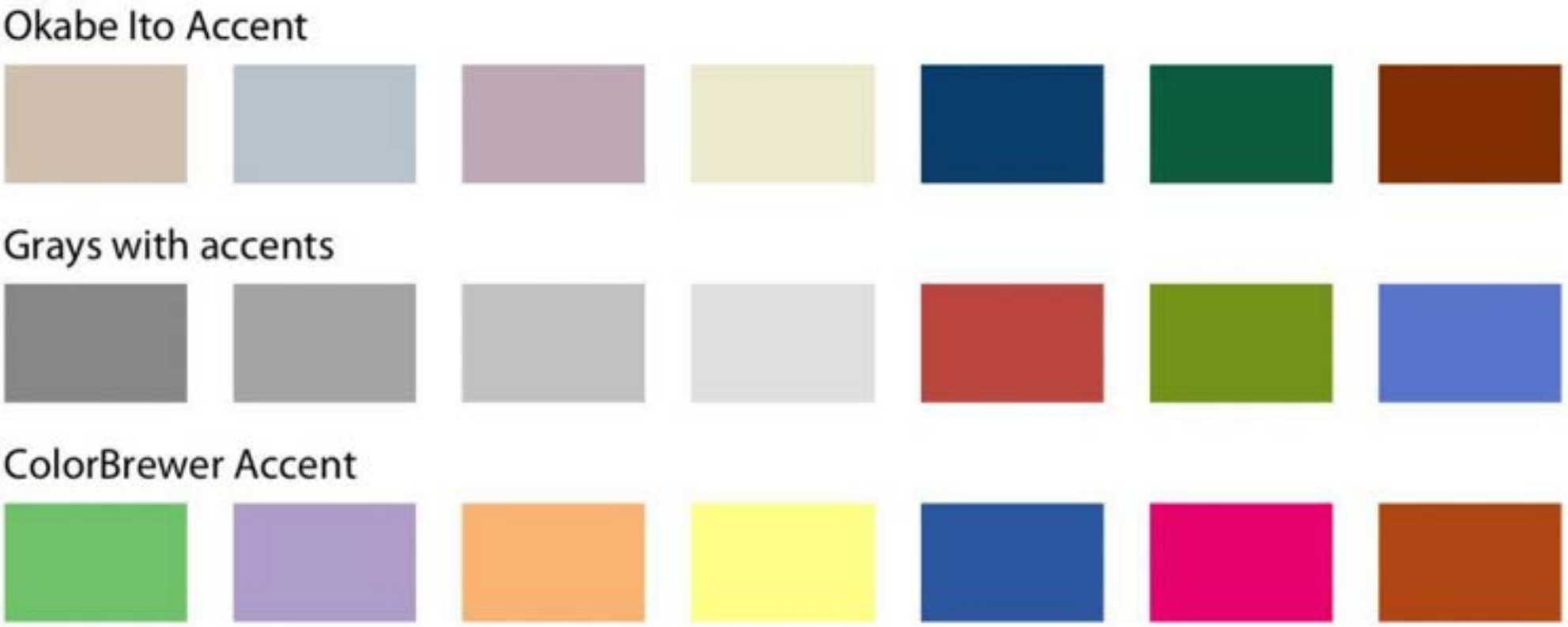




Sequential

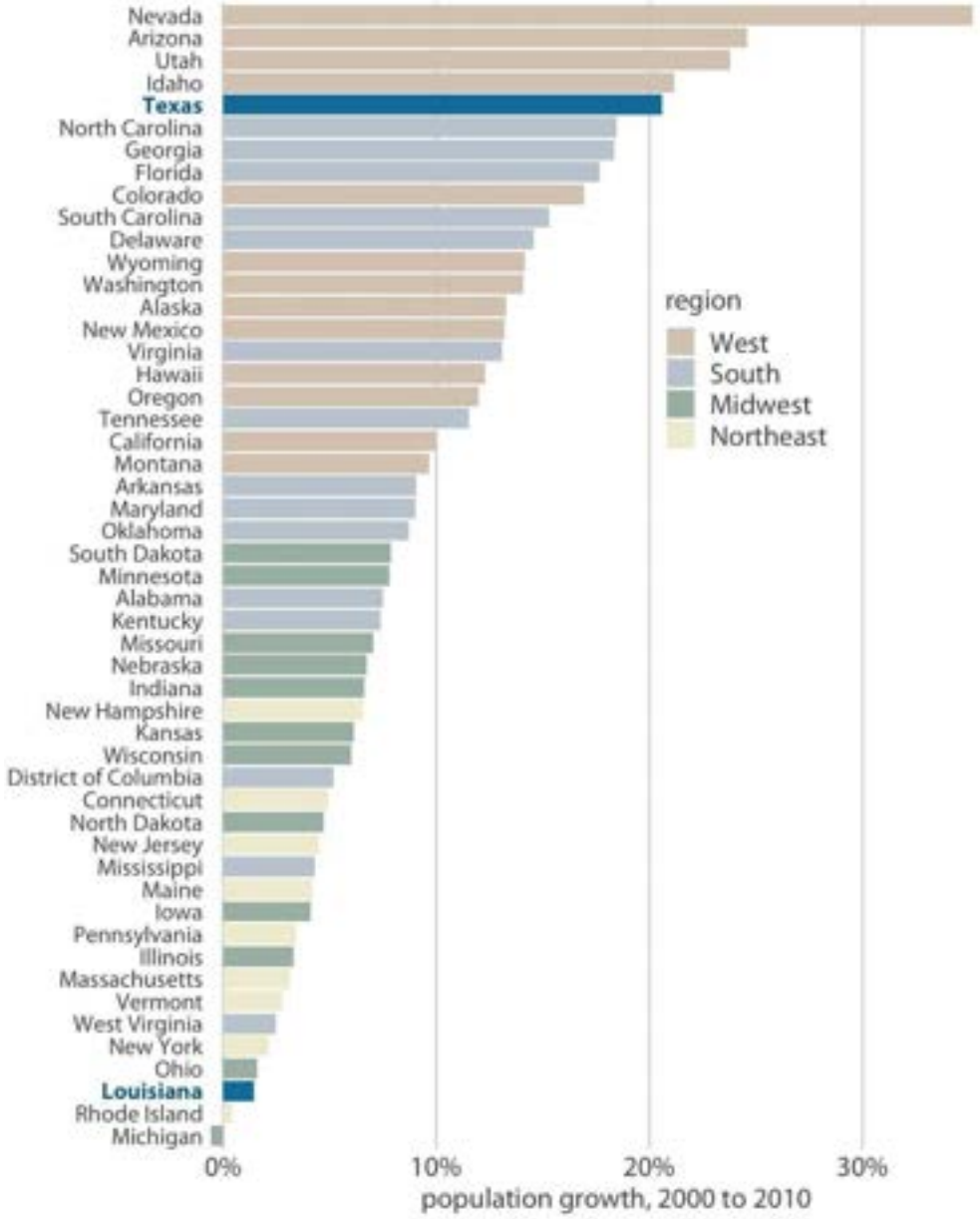
Divergent

Color as a Tool to Highlight: elements in a color or set of colors that stand out against the rest of the figure— achieved with accent color scales.



Source : Fundamentals of Data Visualization by Claus O. Wilke

Colors

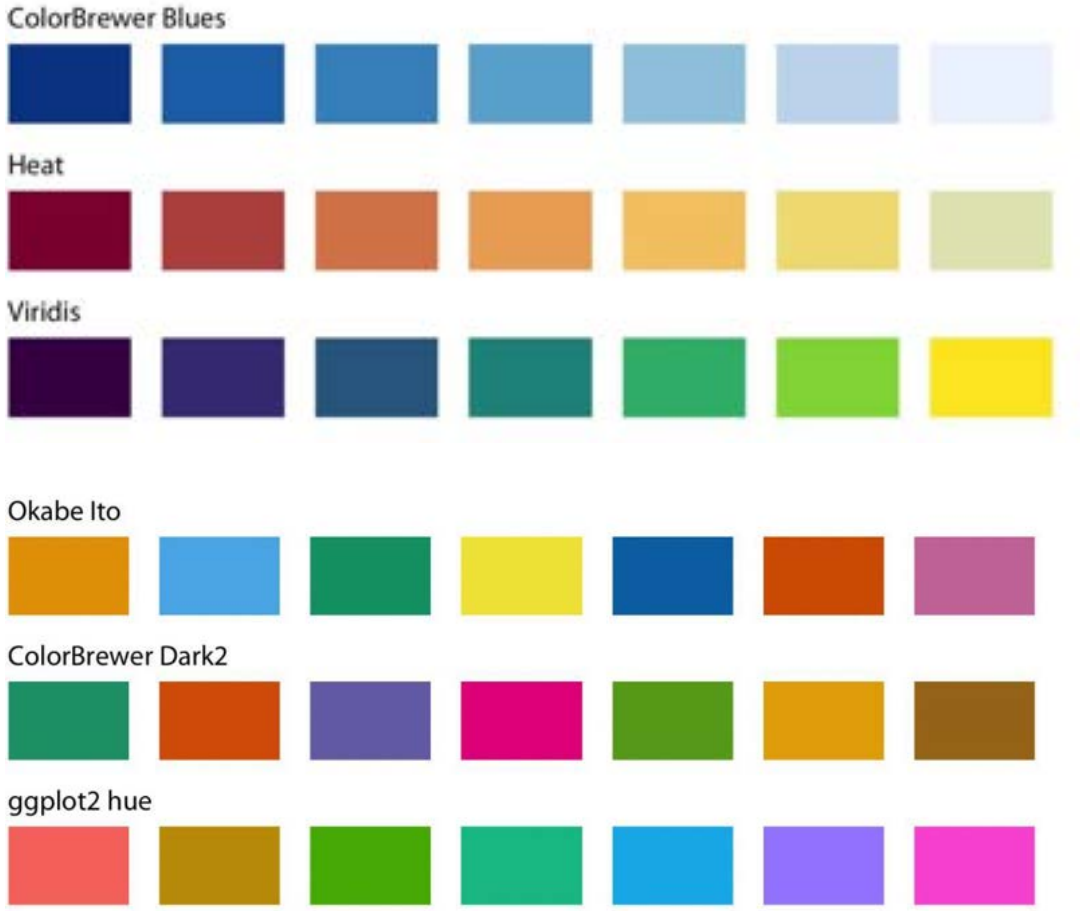


Source : Fundamentals of Data Visualization by Claus O. Wilke

Consider a visualization problem that requires **a) a qualitative** color palette and **b) a quantitative** color palette.

Got to **color.adobe.com** or **coolors.co** and create the palettes.

Post a screenshot of you palette together with a brief description of the visualization problem on Slack.



Sketching your ideas!

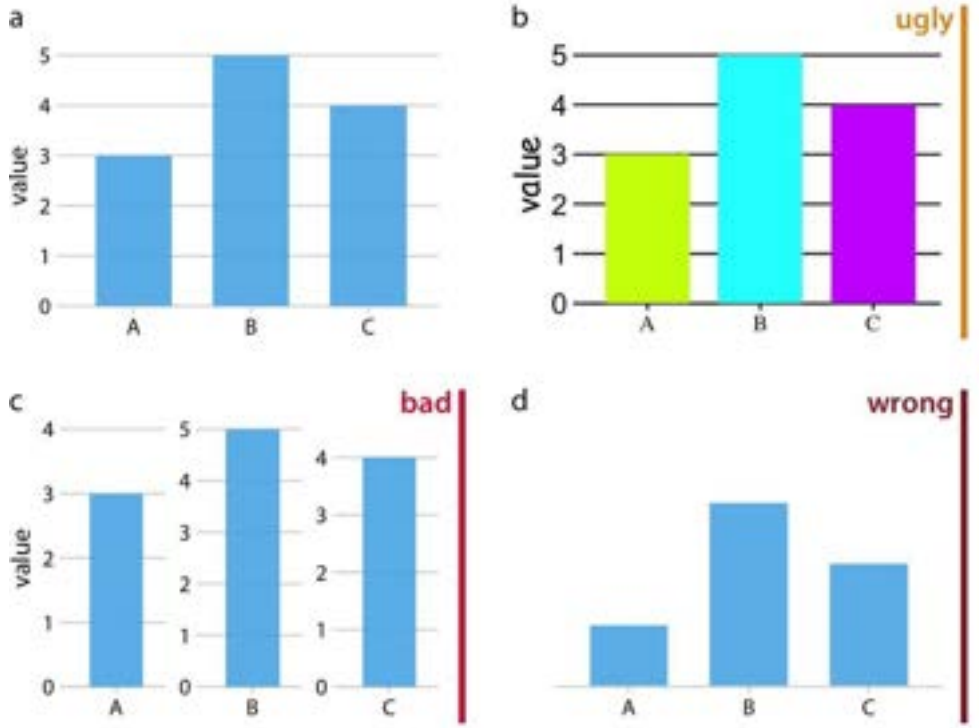
Sketching helps us integrate different kinds of knowledge

- Drawing an idea makes explicit the metaphors you're using to think about the idea;
- Physical representations allow you to see characteristics and relationships of concepts more easily;
- Our brains constantly translate the visual and the verbal, so externalizing this process helps us communicate and process more effectively.

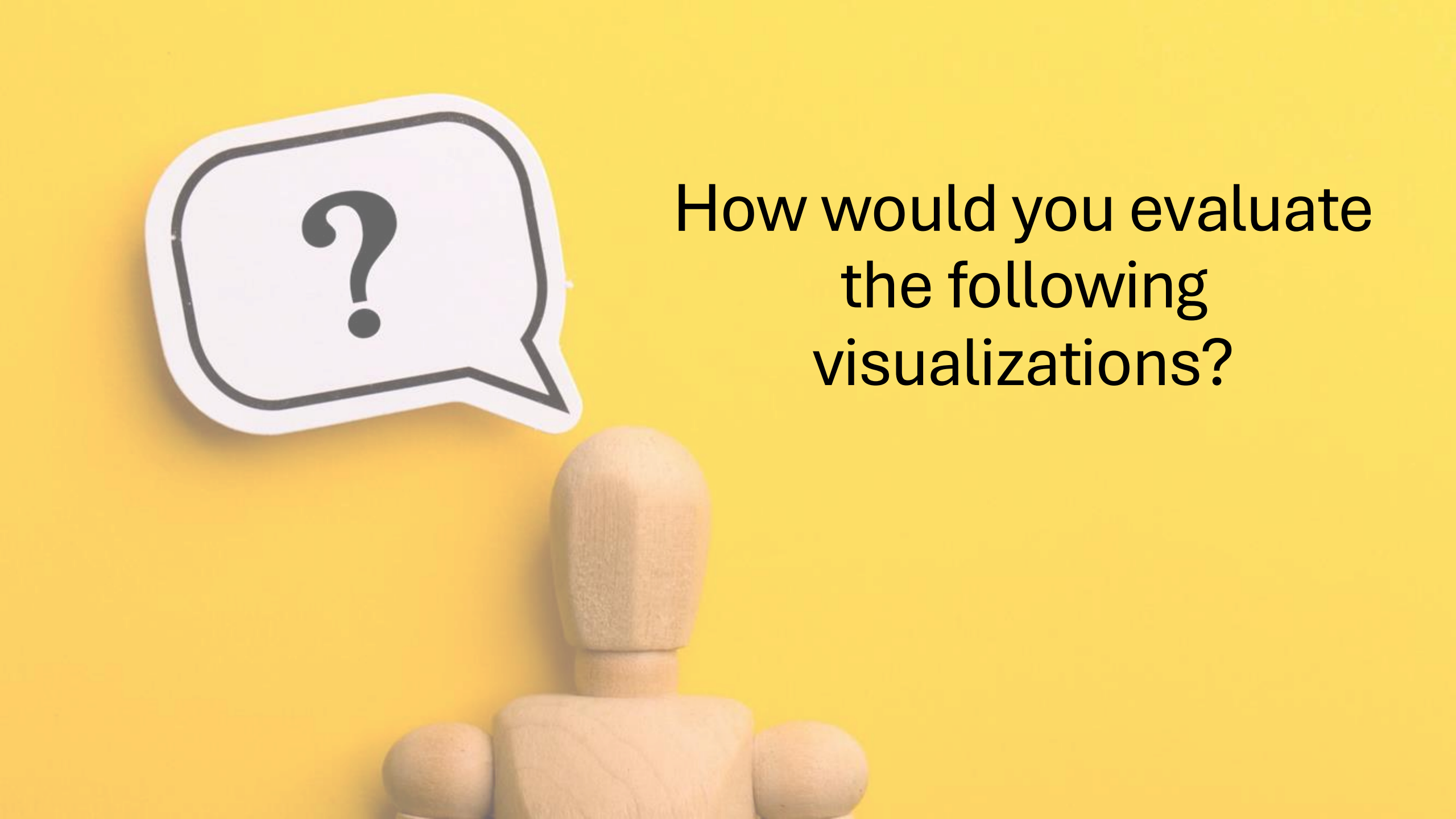
Bad - a figure that has problems related to perception; it may be unclear, confusing, overly complicated.

Ugly - a figure that has aesthetic problems but otherwise is clear and informative

Wrong - a figure that has problems related to mathematics: it is objectively incorrect

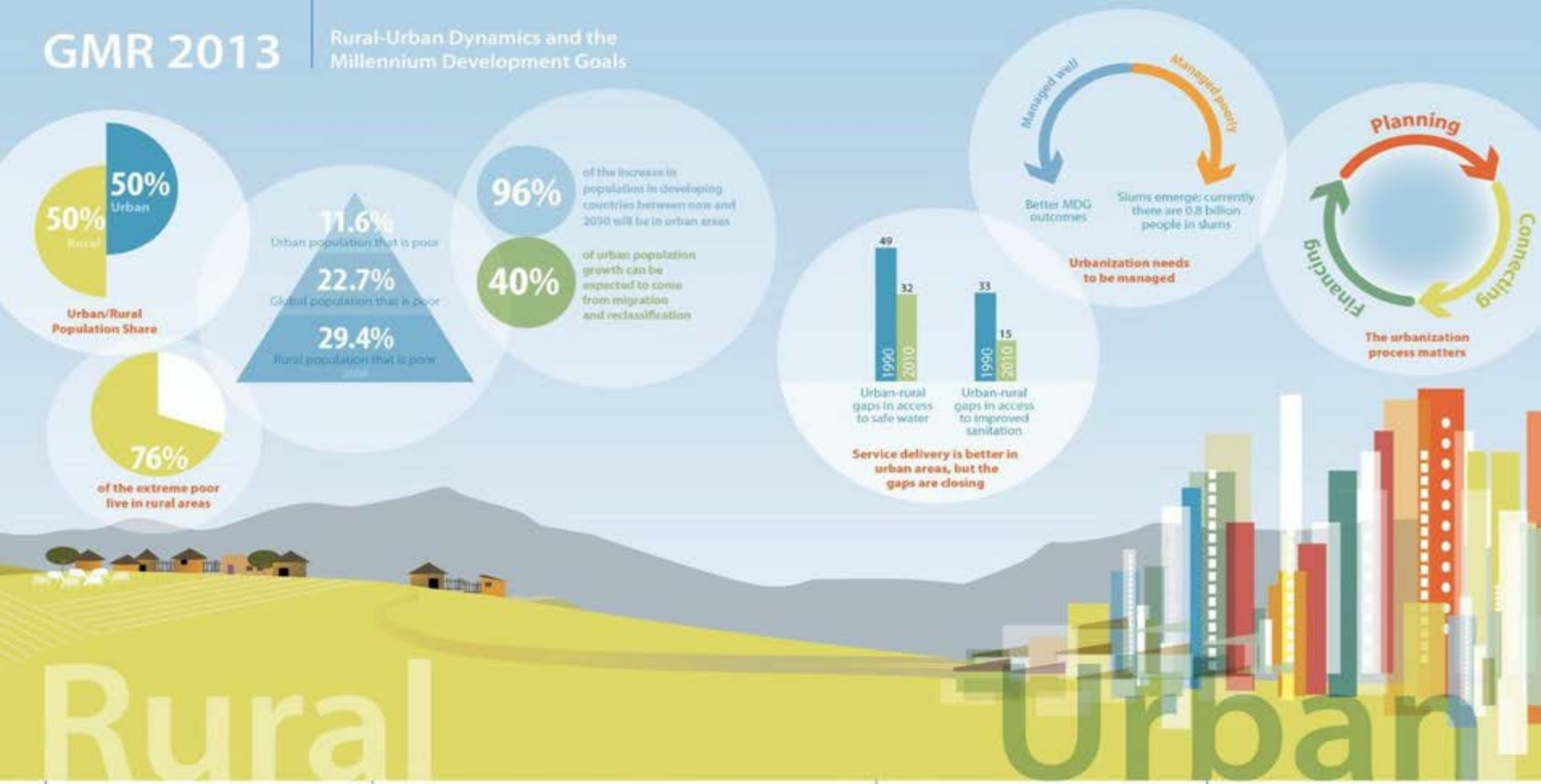


Source : Fundamentals of Data Visualization by Claus O. Wilke


How would you evaluate the following visualizations?

GMR 2013

Rural-Urban Dynamics and the Millennium Development Goals



Four MDG targets have been met; MDG 1.a (halving extreme poverty), two parts of MDG 7 (access to safe water and improved lives of slum dwellers, and part of MDG 3.a (gender parity in primary education). Progress on the remaining MDGs is limited, except for MDG 3.a (gender parity in primary and secondary education), which is close to being on target.

Populations are typically seen as being spatially bipolar. In reality, people and poverty are located along a spectrum from rural to urban, with many types of settlements from small to large towns. The experience is that the smaller the town, the higher the poverty rate, with less access to MDG-related services.

The MDGs reflect the basic needs of all citizens, and governments should aim to meet them fully in both urban and rural areas. But resources are scarce, and priorities must be set. Much of the sequencing will depend on local conditions regarding degree of urbanization and rural-urban differences in MDG outcomes.

Urbanization by itself is no guarantee for success. If unregulated and poorly planned, urbanization can lead to disproportionate increases in slums. GMR 2013 calls for an integrated strategy to better manage the planning-connecting-financing formula of urbanization.