

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)	30
	Thursday, March 23 @ 15:40 PM (Building 103, S56)	
	Total Points	46

Recap Data Cleaning and Feature Engineering



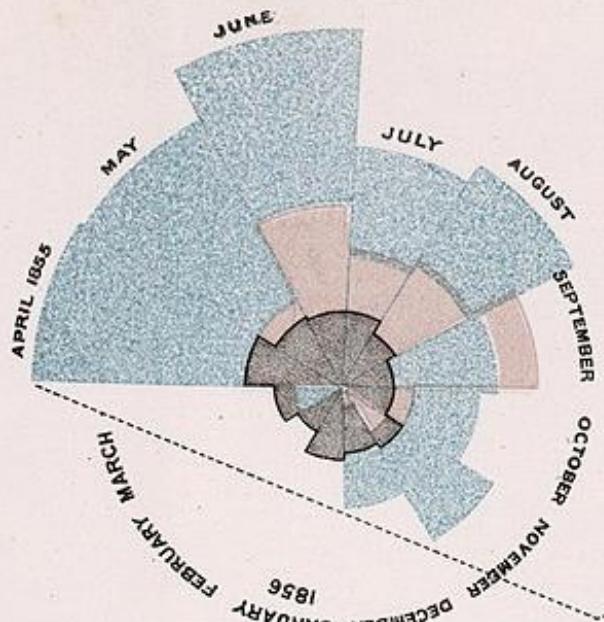
Data Visualization

An Introduction

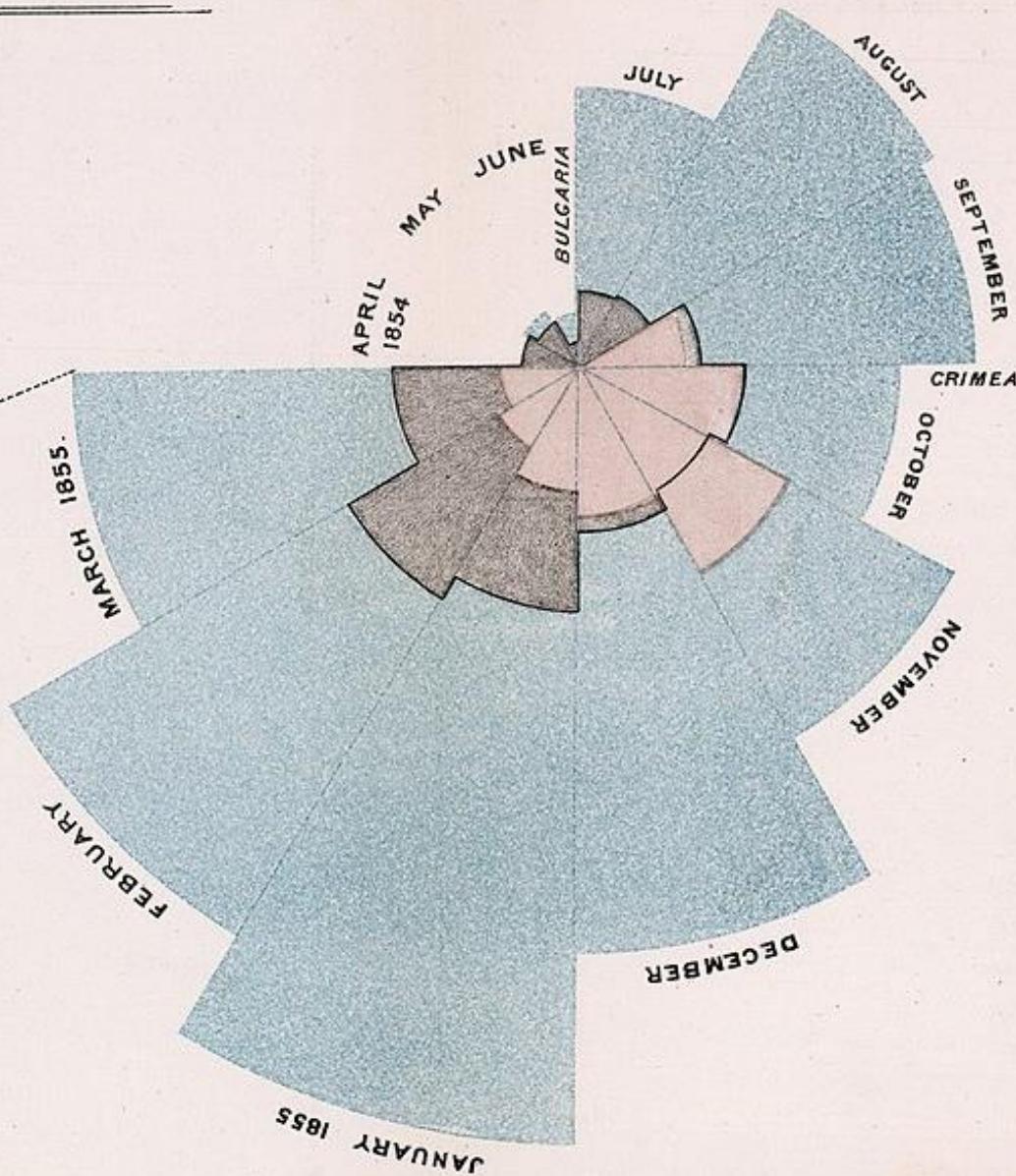


DIAGRAM OF THE CAUSES OF MORTALITY
IN THE ARMY IN THE EAST.

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 Nov^r 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.

Everyday examples of data visualizations

A Weatherline-style weekly forecast

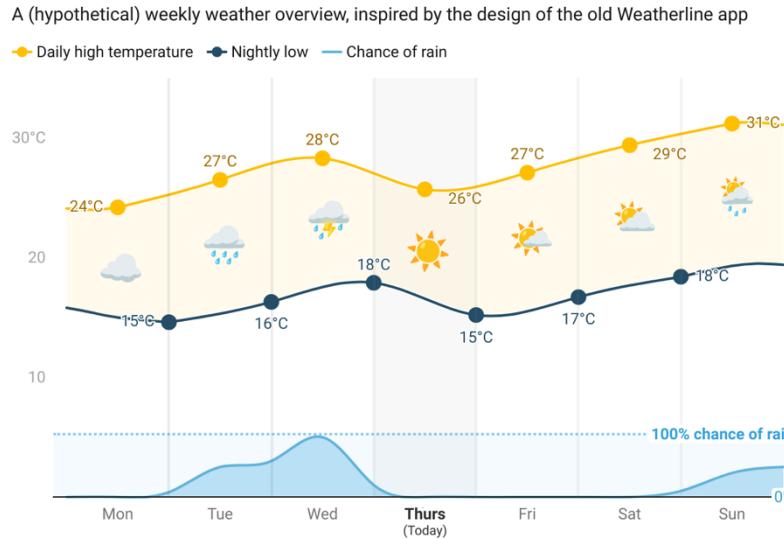
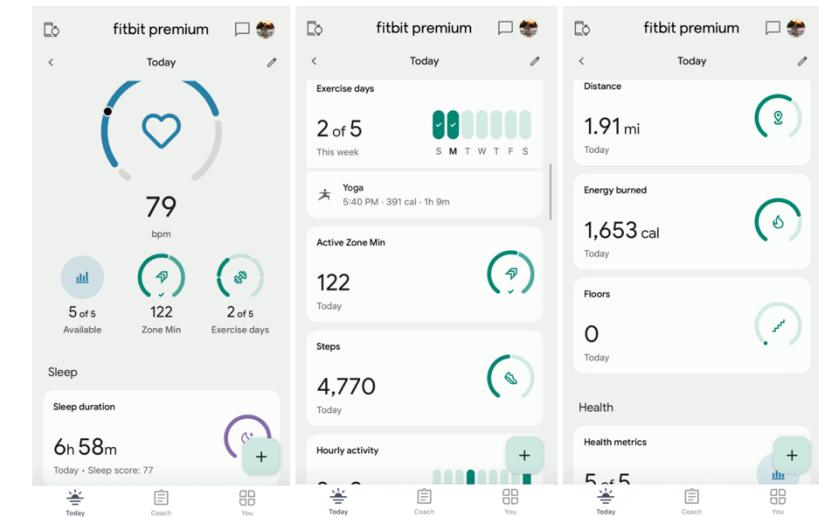


Chart: Jonathan Muth



From overwhelming tables to understanding

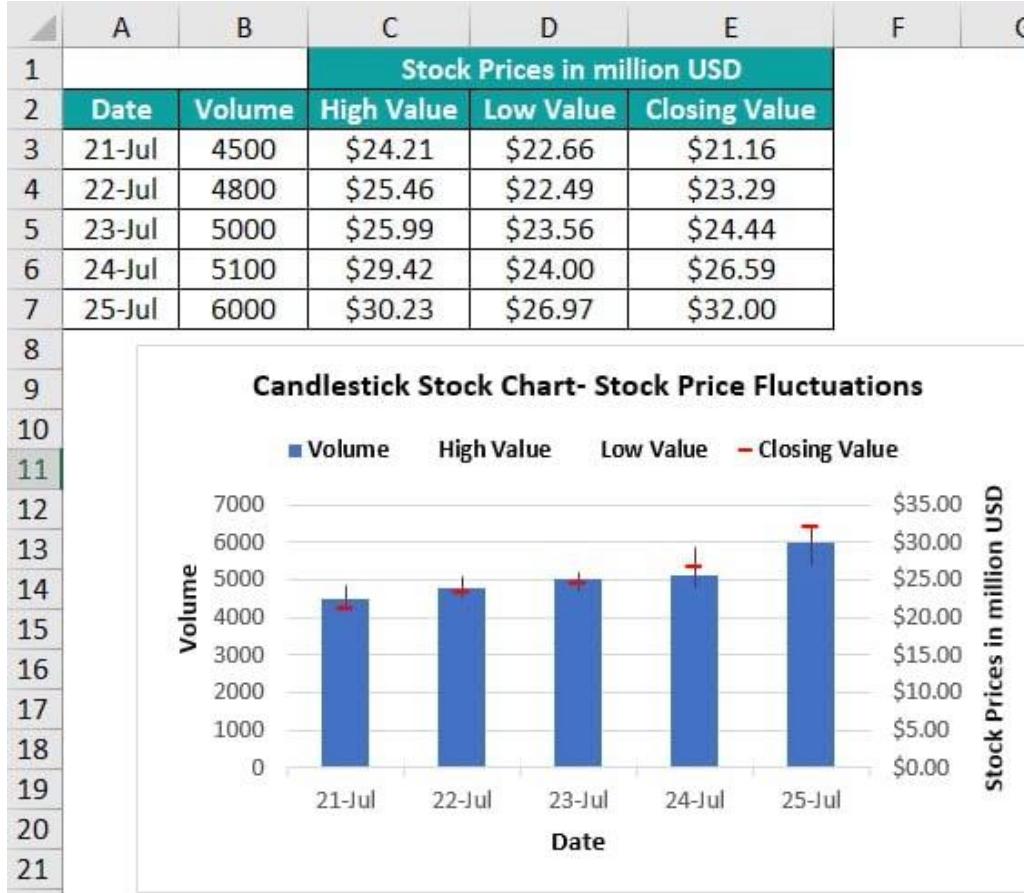
The screenshot shows a Google Sheets document titled "Teenage Movie Ratings". The interface includes a toolbar at the top with File, Edit, View, Insert, Format, Data, Tools, Add-ons, Help, and a status bar at the bottom indicating "99" rows, "4 / Card" view, "Fast" refresh, and "Delivered" status.

The document contains two main tables:

- Pivot Table (Top):** A summary table with columns for movie titles and their average ratings. The columns are labeled A through H. Row 1 contains headers for movie titles and counts. Rows 2-7 show data for "Air Force One", "Chasing Army", "Contact", "Courage Under Fire", and "E.T. the Extra-Terrestrial". Row 8 is a blank header row, and rows 9-16 show data for "Evita", "Fargo", "Game, The", "Independence Day", "Liar Liar", "Mission: Impossible", "Phenomenon", "Return of the Jedi", "Rock, The", "Saint, The", "Scream", "Star Wars", "Titanic", "Toy Story", "Twelve Monkeys", "Willy Wonka and the Chocolate Factory", "Grand Total", and "104".
- Raw Data Table (Bottom):** A detailed table with columns A through F. Column A is labeled "Order #", and column B is labeled "First Name". The table lists 16 entries, each with a unique Order #, First Name, Last Name, Email, Country, and IP address.

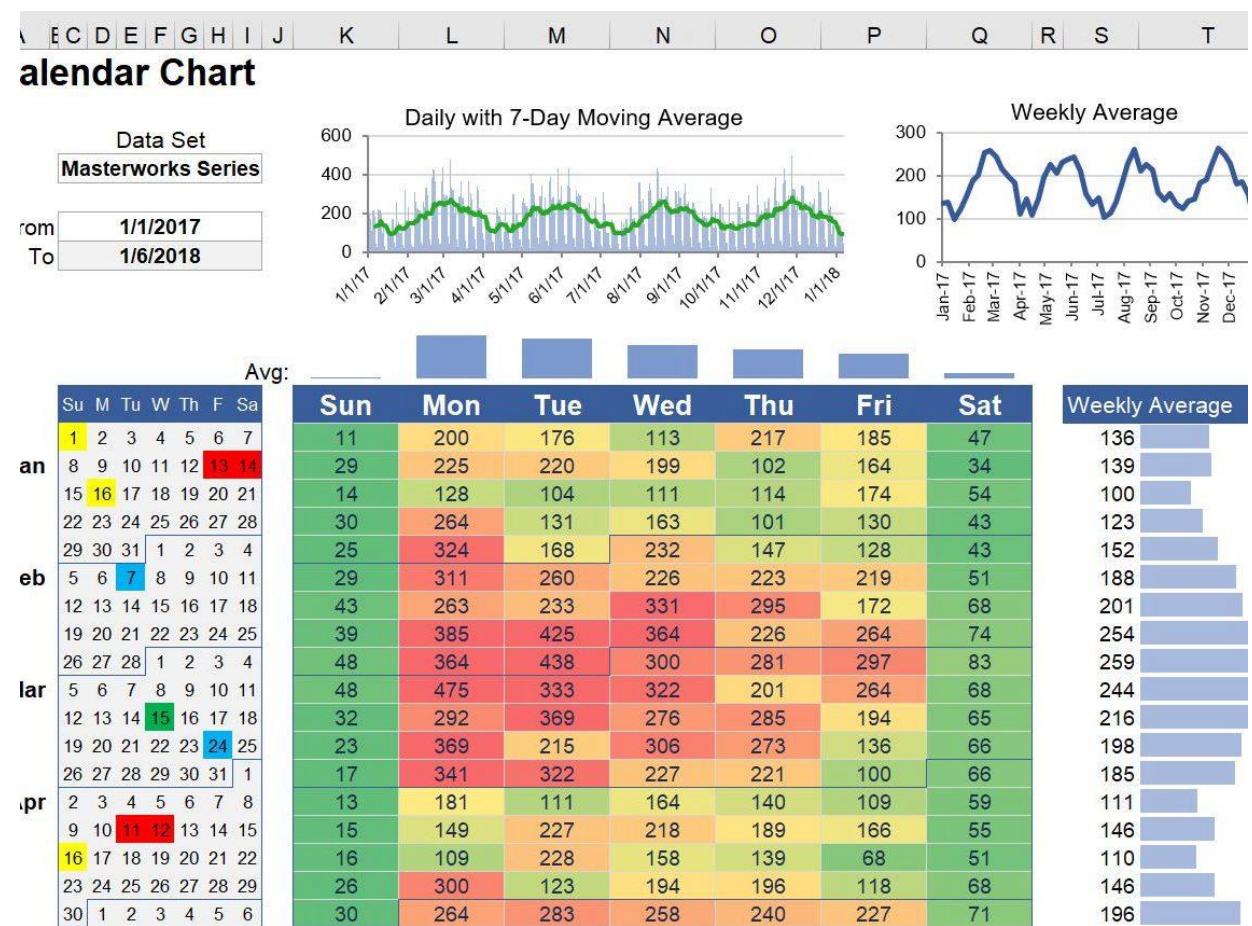
	A	B	C	D	E	F
1	Order #	First Name	Last Name	Email	Country	IP address
2	1	Dalton	Kramer	dalton@email.com	France	211.91.226.1
3	2	Gita	Tetterton	gita@email.com	USA	222.153.179.
4	3	Weston	Jurgens	weston@email.com	Spain	203.123.236.
5	4	Brad	Chupp	brad@email.com	France	202.183.111.
6	5	Marybeth	Baumann	marybeth@email.com	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

Communicate data clearly & effectively



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	



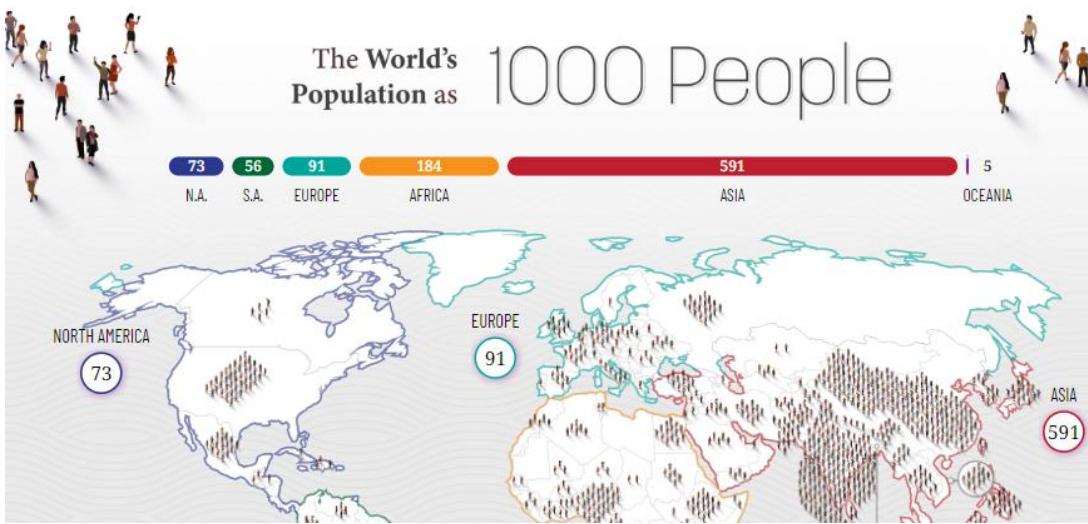
Why use data visualization?

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

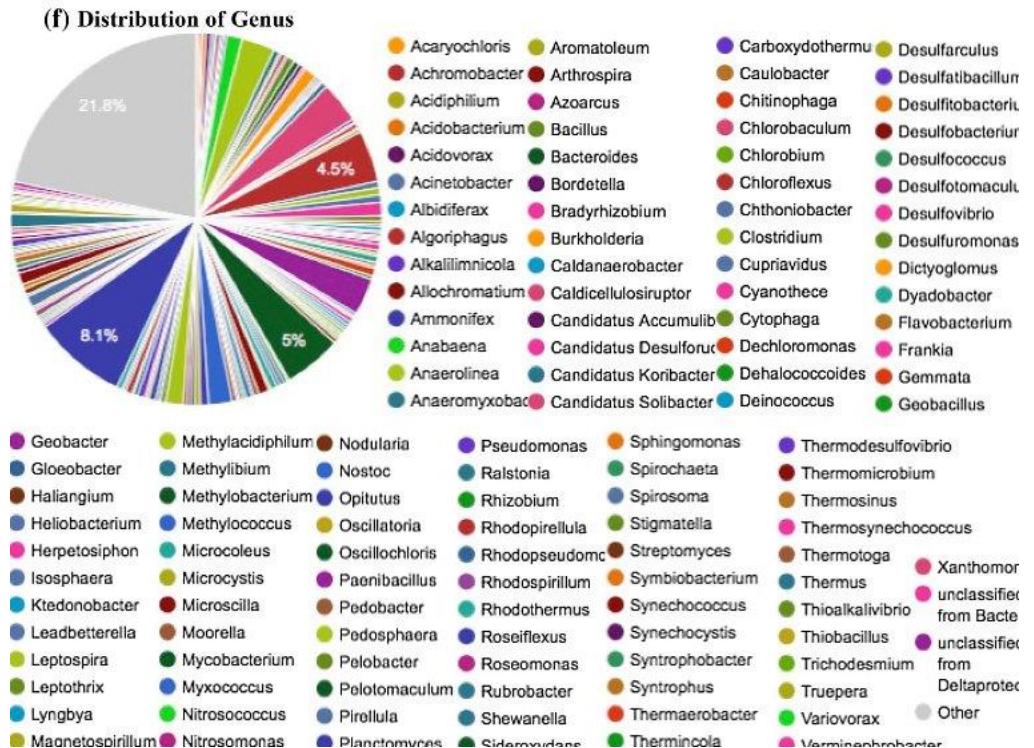


Explore and Report (on Slack)

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?

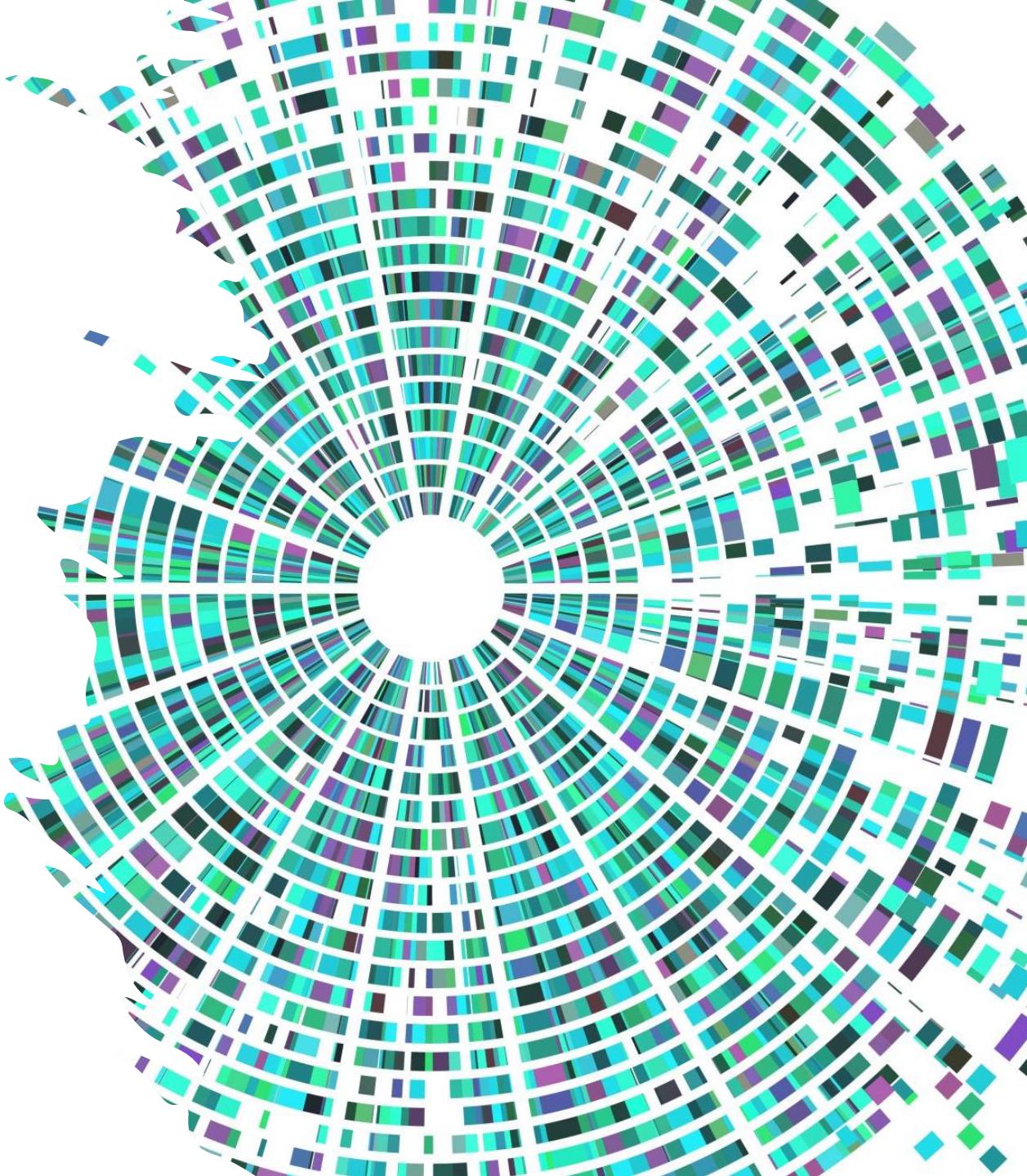
- Good data visualizations are created when communication, data science, and design collide.
- Data visualizations done right offer key insights into complicated datasets in ways that are meaningful and intuitive.

DATA VISUALIZATION



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

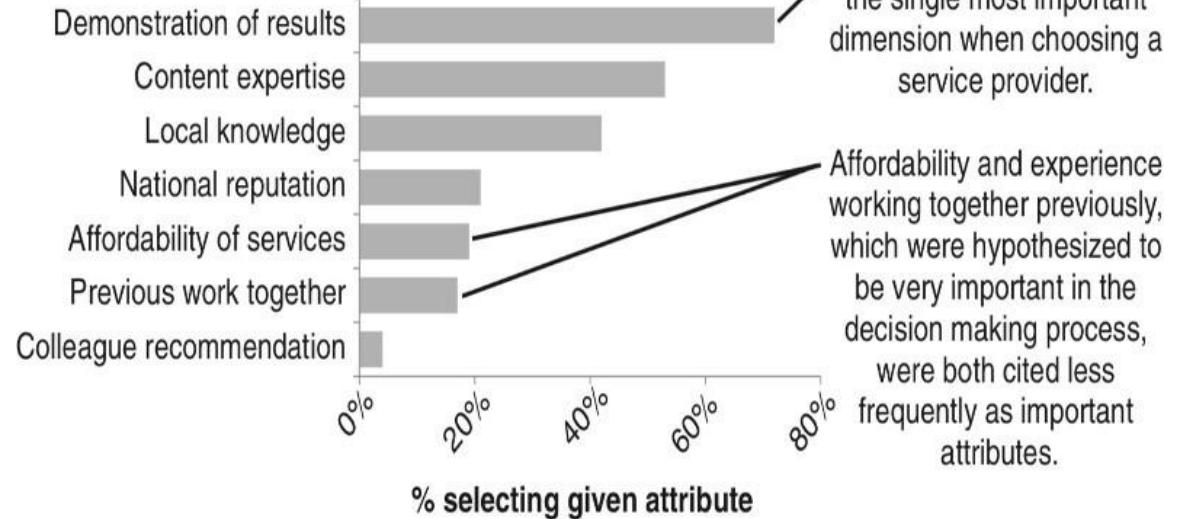
Month	North	South	East	West	Online	Retail	Total
Jan	1137	1274	1411	1548	1685	1822	1959
Feb	1274	1548	1822	2096	2370	2644	2918
Mar	1411	1822	2233	2644	3055	3466	3877
Apr	1548	2096	2644	3192	3740	4288	4836
May	1685	2370	3055	3740	4425	5110	5795
Download declutter.pptx and create order. Post your solution on our Slack channel				4288	5110	5932	6754
				4836	5795	6754	7713

2. Create order

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

Demonstrating effectiveness is most important consideration when selecting a provider

In general, what attributes are the most important to you in selecting a service provider?
(Choose up to 3)



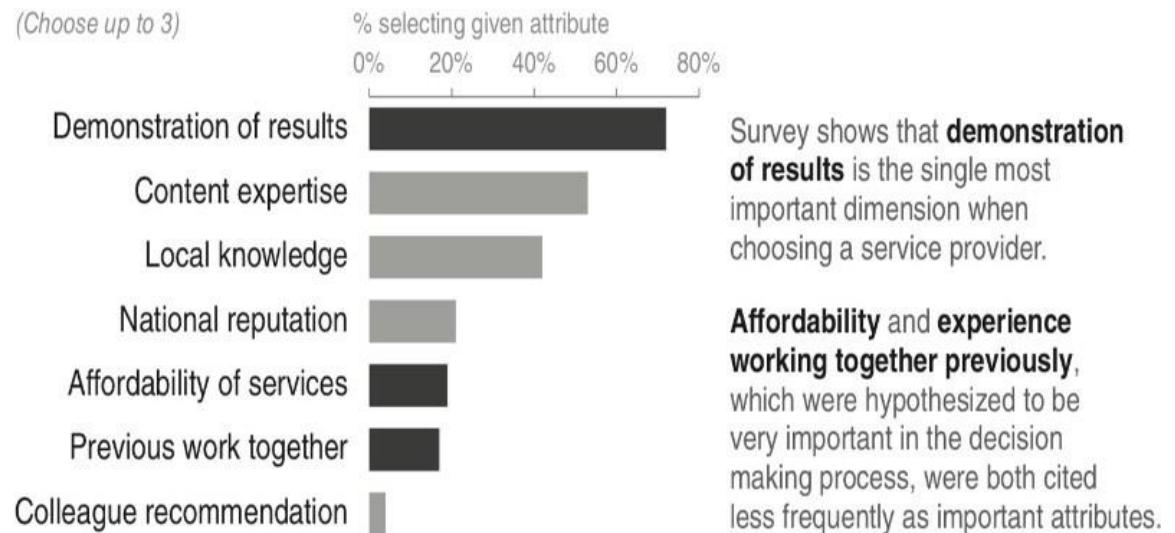
Data source: xyz; includes N number of survey respondents. Note that respondents were able to choose up to 3 options.

2. Create order

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

Demonstrating effectiveness is most important consideration when selecting a provider

In general, **what attributes are the most important** to you in selecting a service provider?

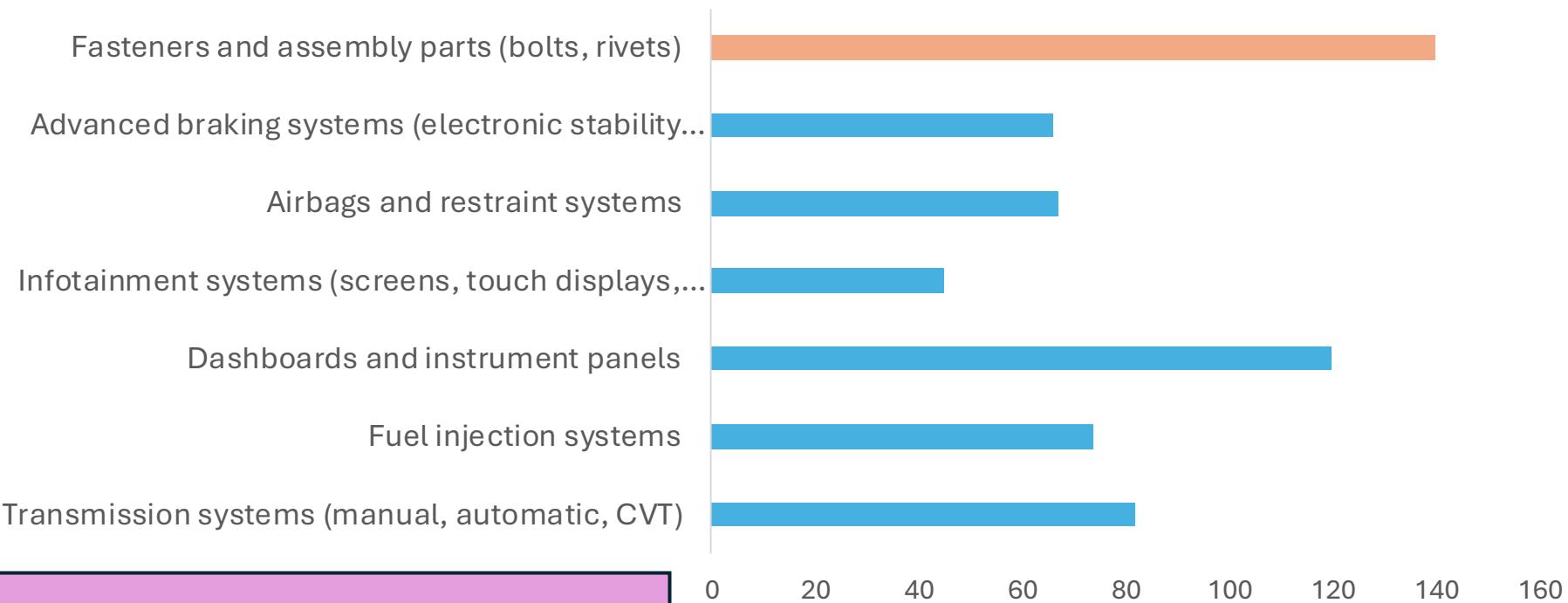


Data source: xyz; includes N number of survey respondents.
Note that respondents were able to choose up to 3 options.

Fasteners and assembly parts were our best selling products in 2023

Infotainment systems were the least sold products

SKUs sold in 2023 by type of product



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

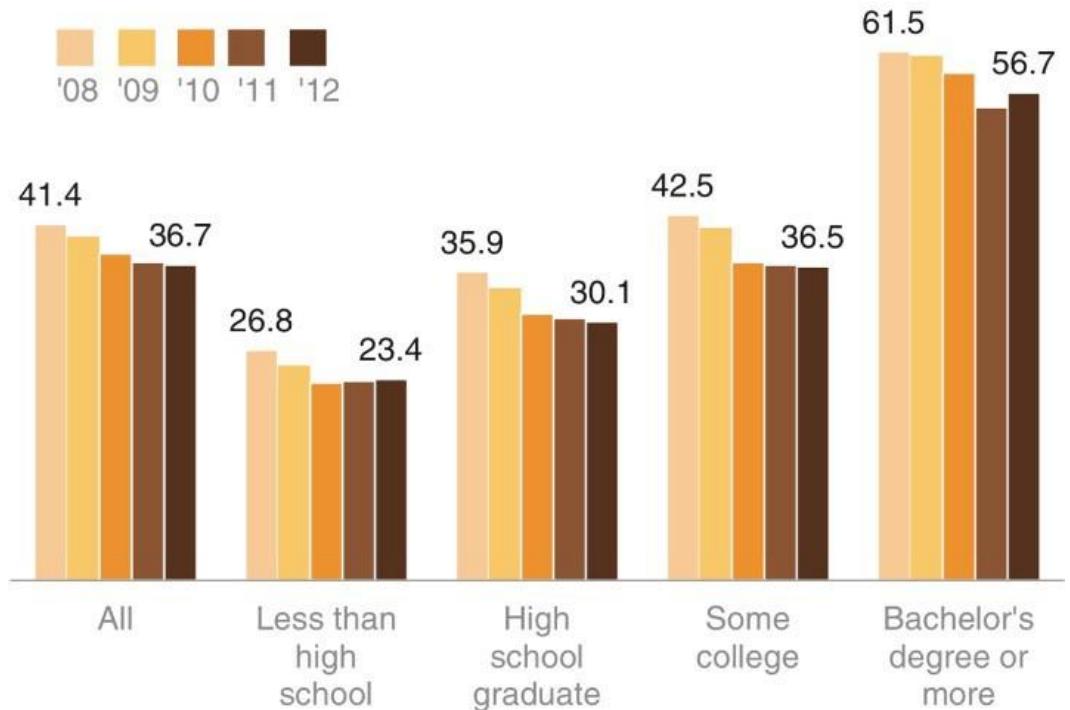
Best and least selling products in 2023

3. Give focus

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

New Marriage Rate by Education

Number of newly married adults per 1,000 marriage eligible adults

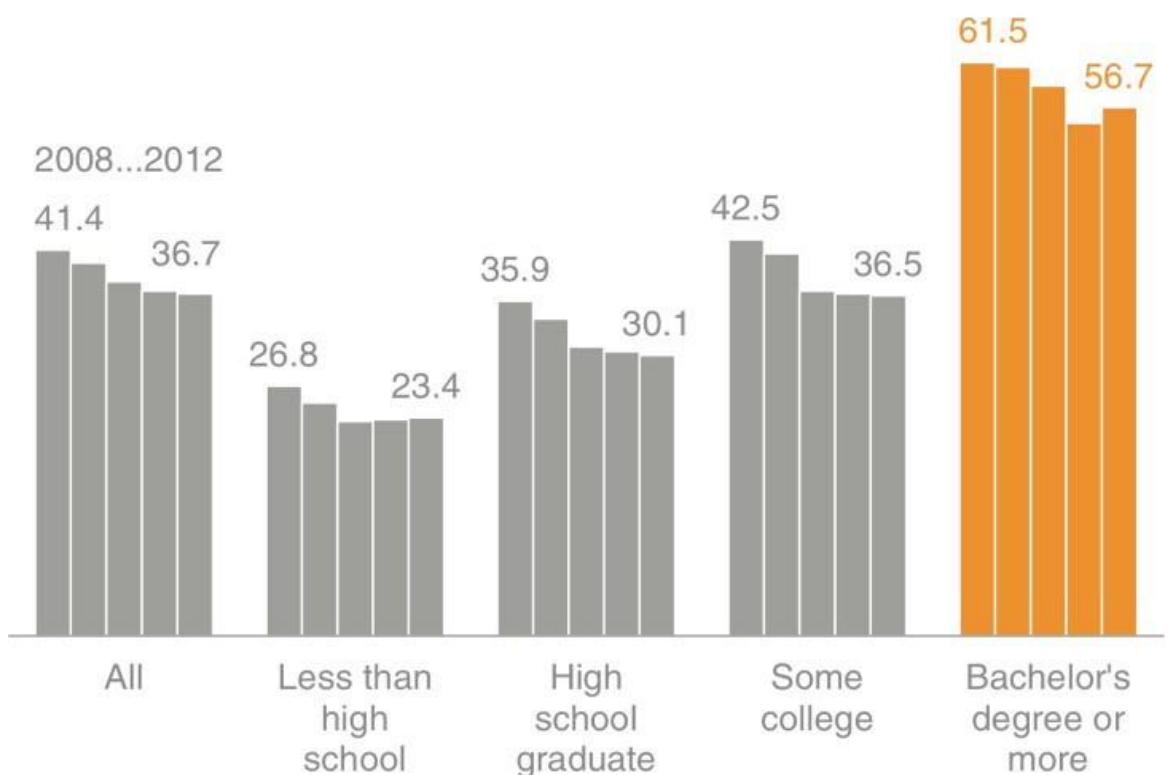


3. Give focus

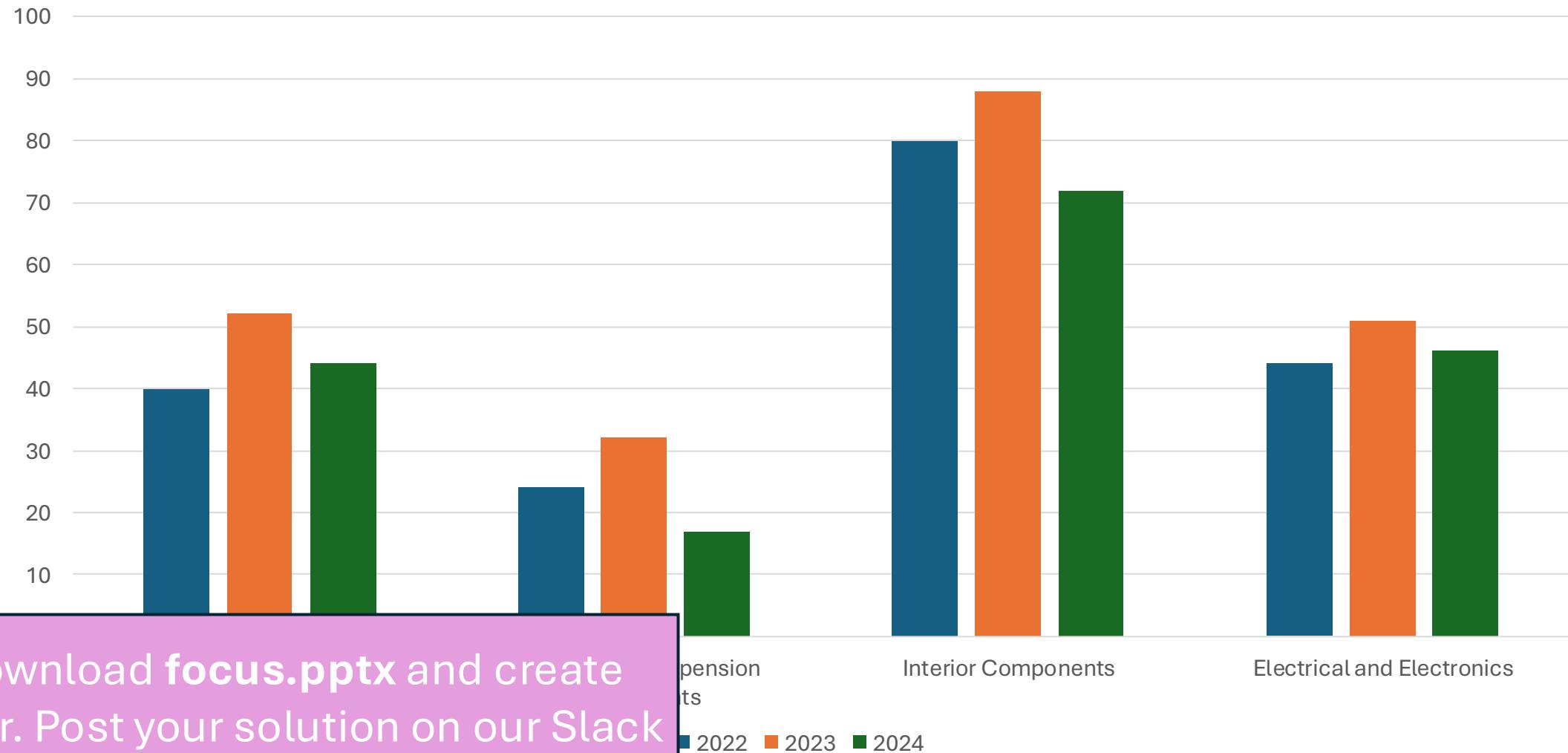
- Use few colors and use them strategically (applies also to contrast).
- Choose colors that are appropriate and pleasing.
- Be consistent.

New Marriage Rate by Education

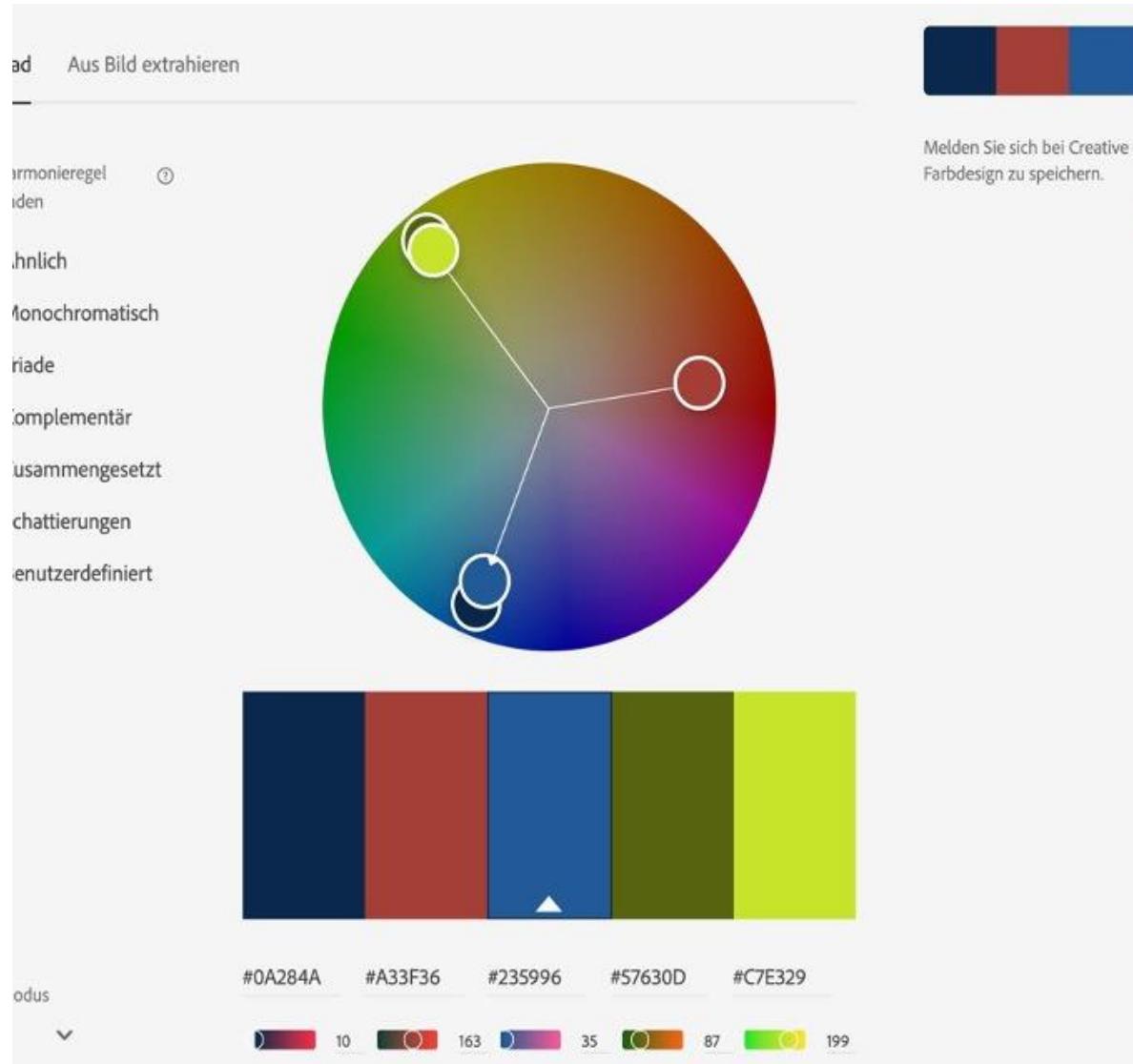
Number of newly married adults per 1,000 marriage eligible adults



2023 was the most successful year across the chosen categories



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**

Visual attention is
by definition a
selective process

And then you will read this

Then this one

Tables

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

Heatmap

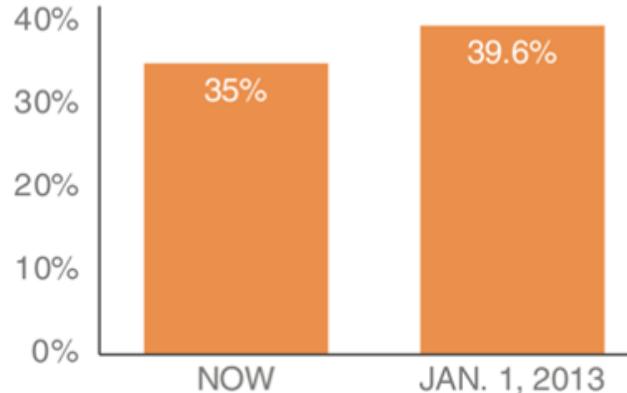
LOW-HIGH

	A	B	C
Category 1	15%	22%	42%
Category 2	40%	36%	20%
Category 3	35%	17%	34%
Category 4	30%	29%	26%
Category 5	55%	30%	58%
Category 6	11%	25%	49%

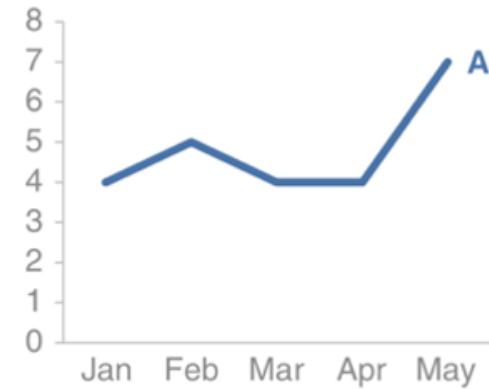
Charts

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

Barplot



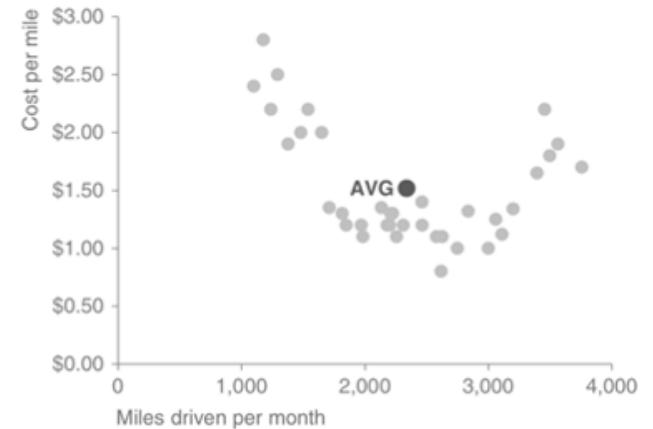
Lineplot



Slope plot



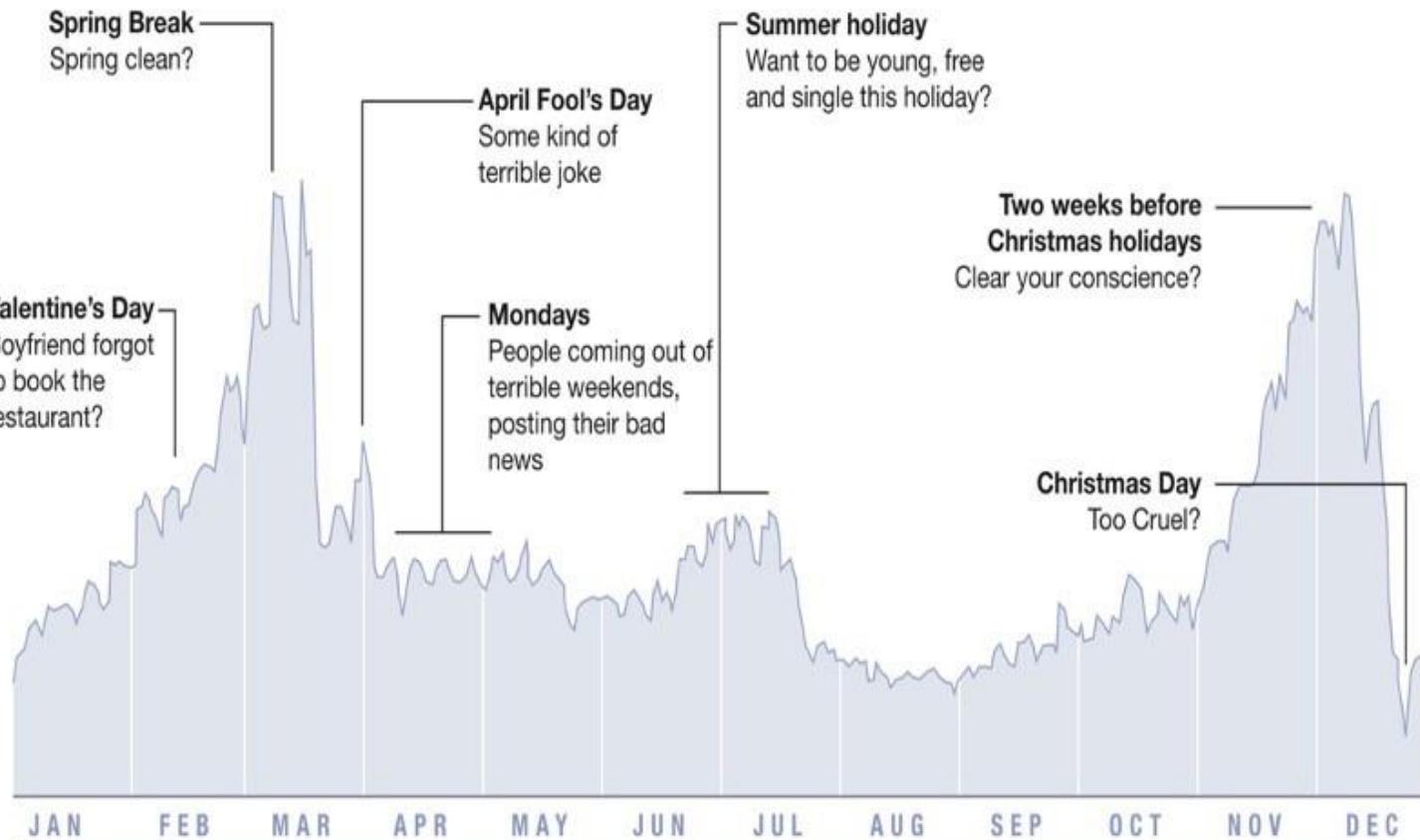
Scatterplot



Annotation

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

Peak Break-up Times
According to Facebook status updates



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.

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

Okabe Ito

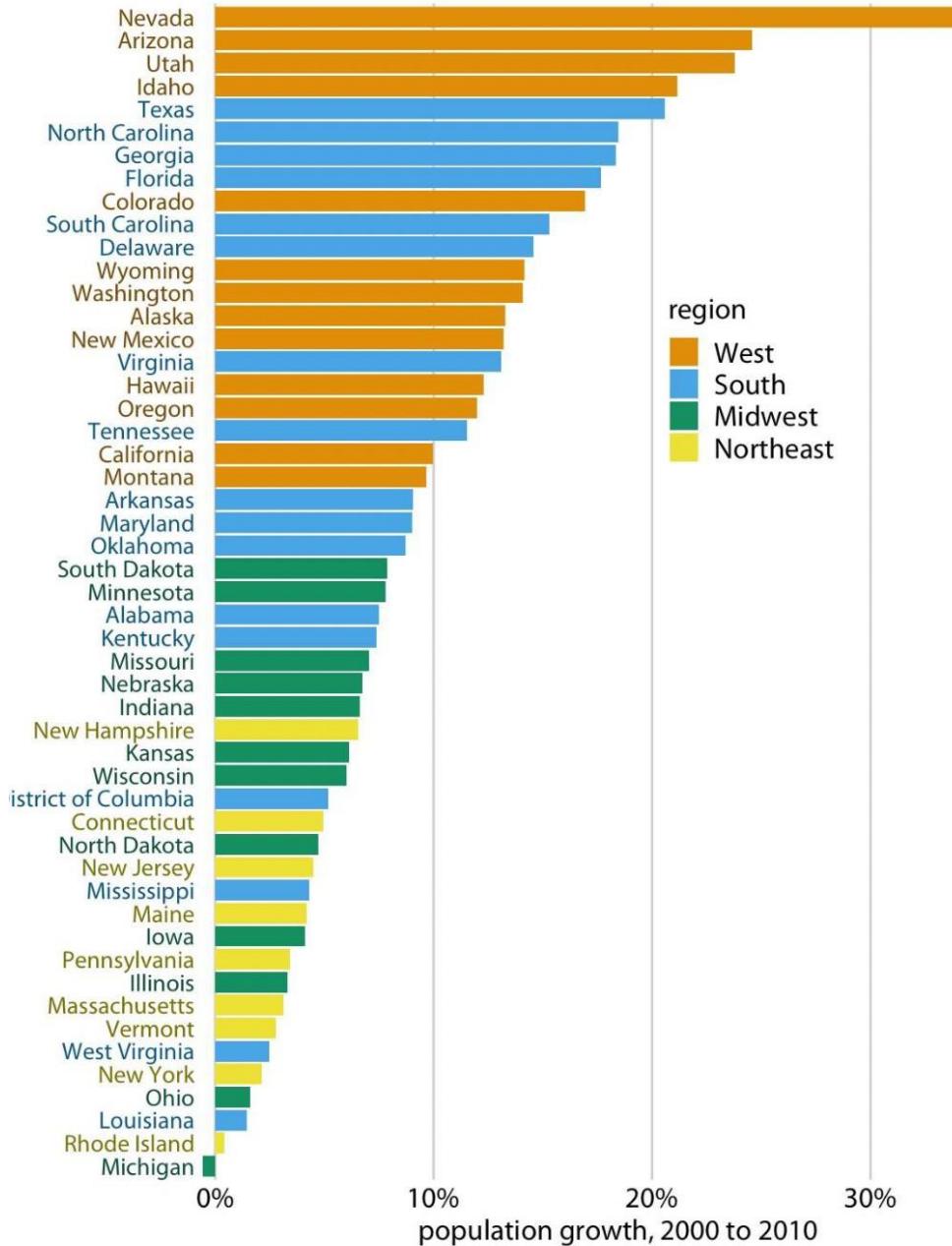


ColorBrewer Dark2



ggplot2 hue



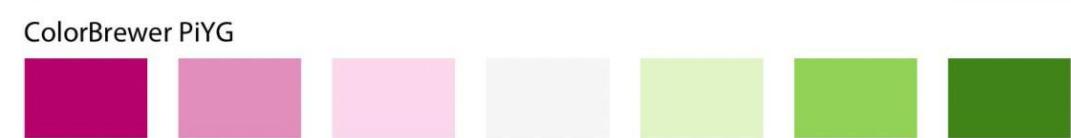


Color to Represent Data Values: Quantitative data values, such as income and temperature. Sequential colors can indicate which values are larger or smaller.

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

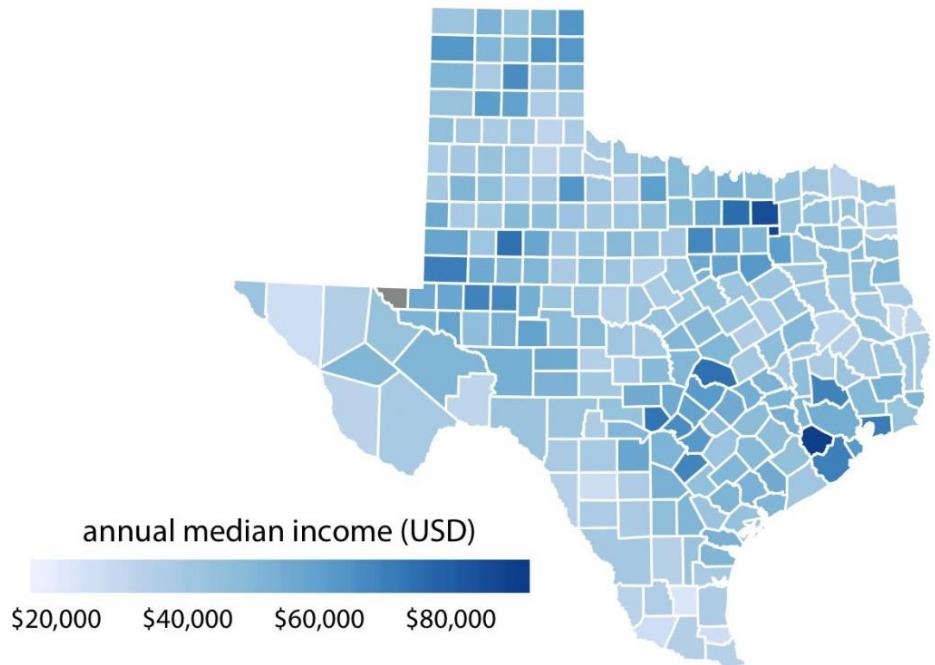


- **Sequential**
|

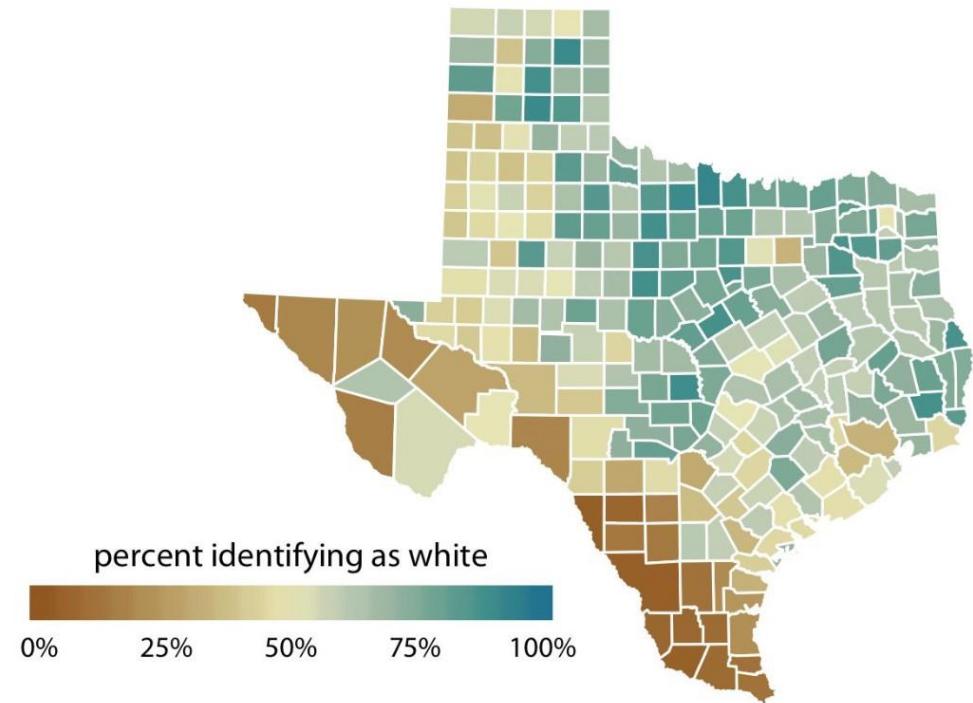


- **Divergent**

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.

Okabe Ito Accent



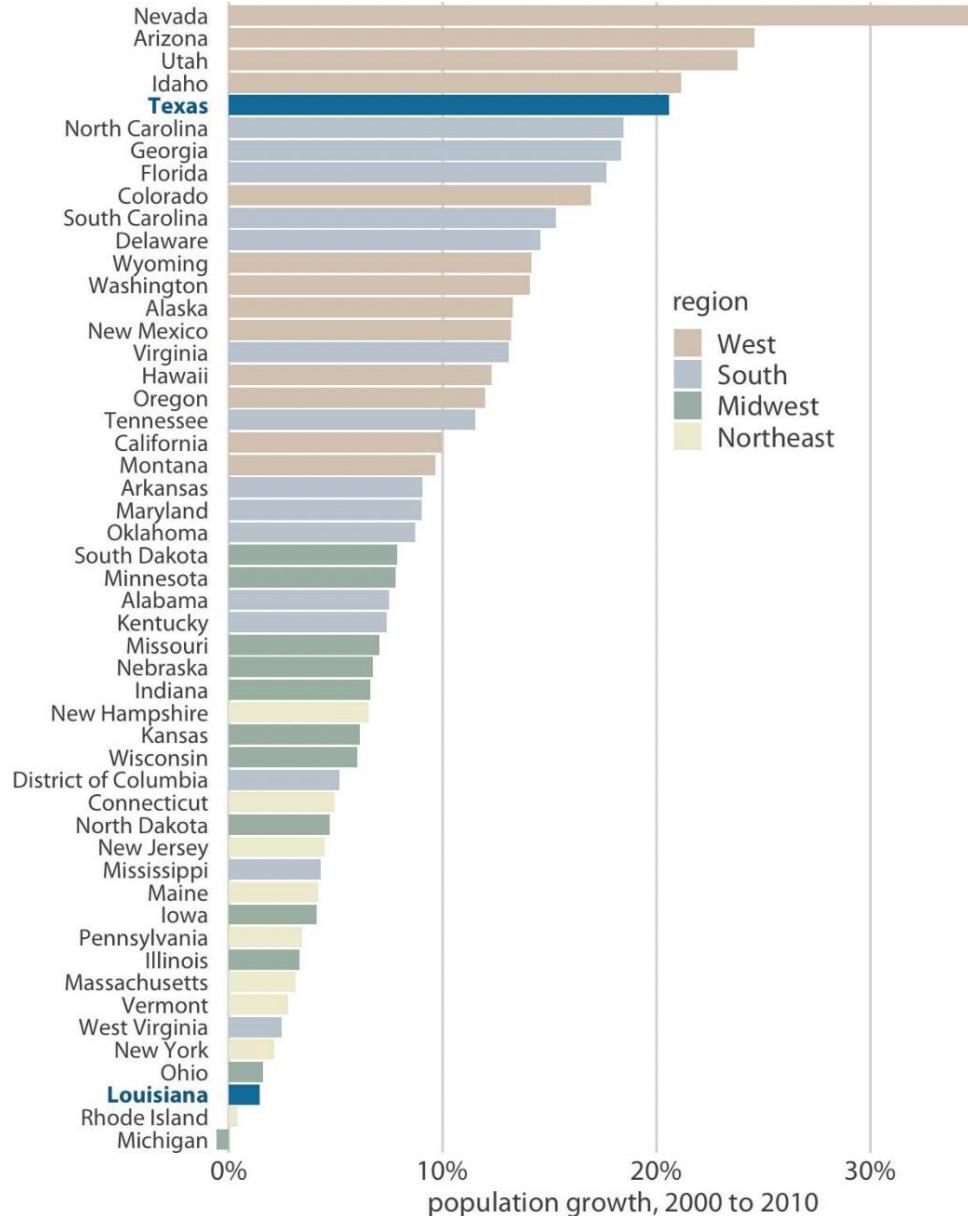
Grays with accents



ColorBrewer Accent



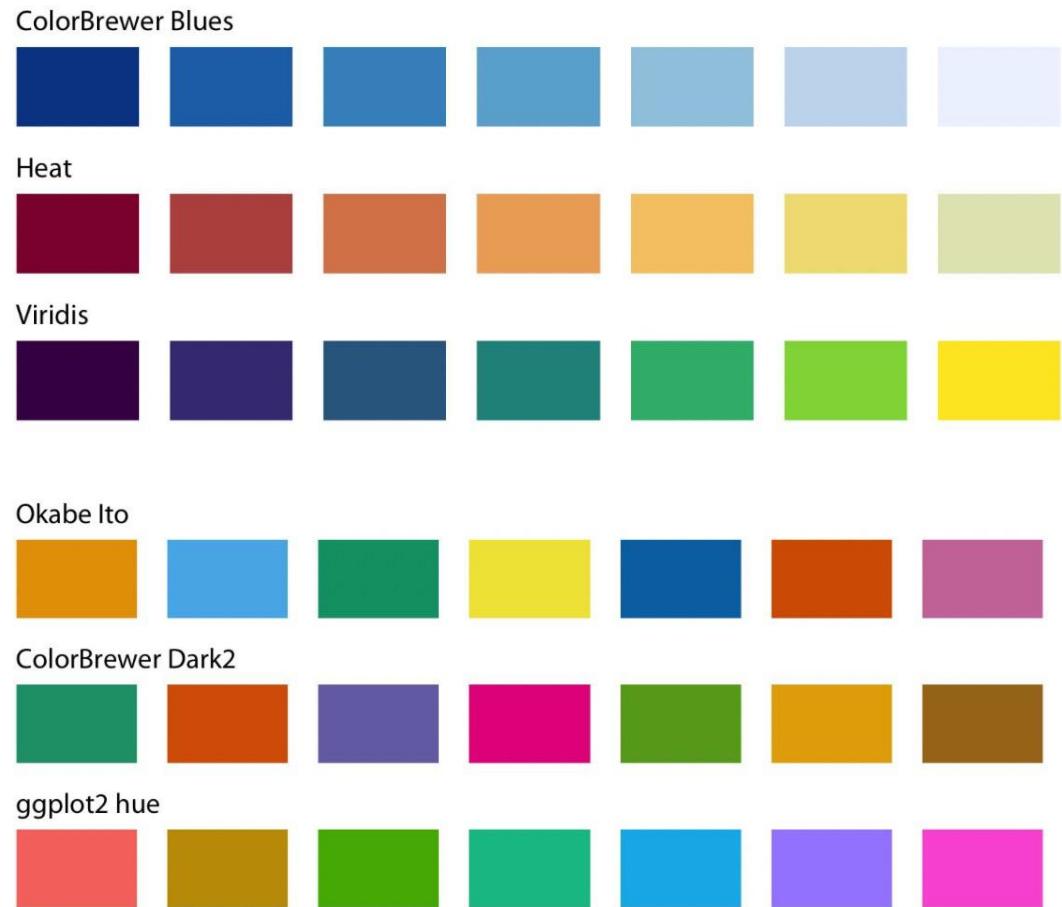
Colors



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.



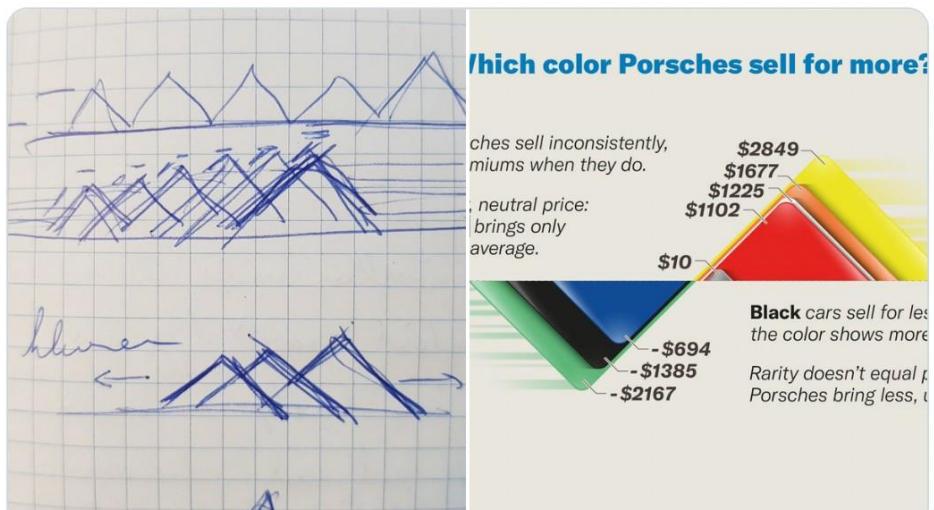


@Sonja_Kuijpers@vis.social
@SonjaKuijpers · Follow



The initial sketch 🖍, and the outcome 💻
Would love to see others as well!
#sketchvsoutcome
Put yours in reply 🤘😊

#dataviz #datavisualisation #illustration #GraphicDesign
#illustrator #sketch



11:37 AM · Jan 18, 2021

Sketching your ideas!

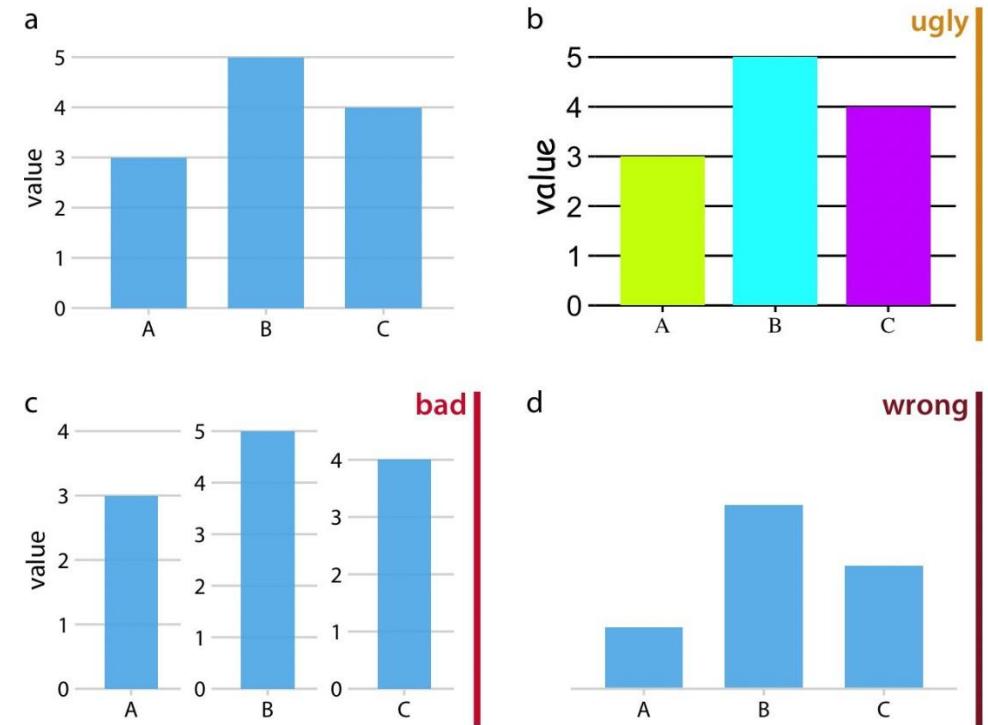
Sketching helps us integrate different kinds of knowledge

- 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.

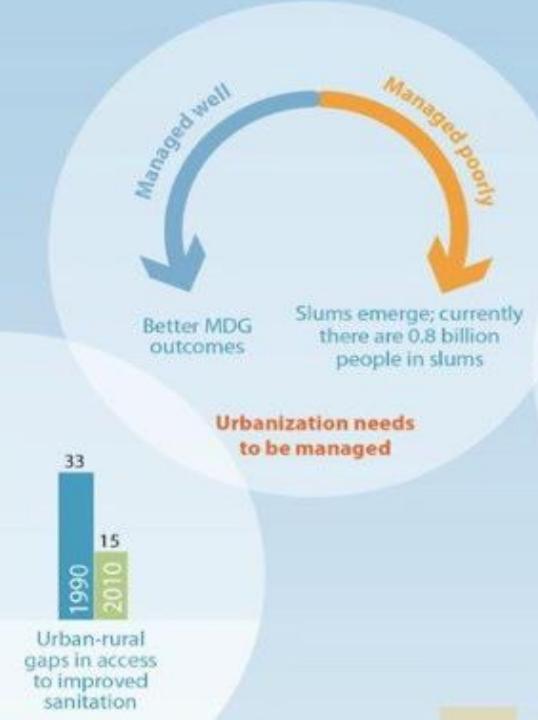
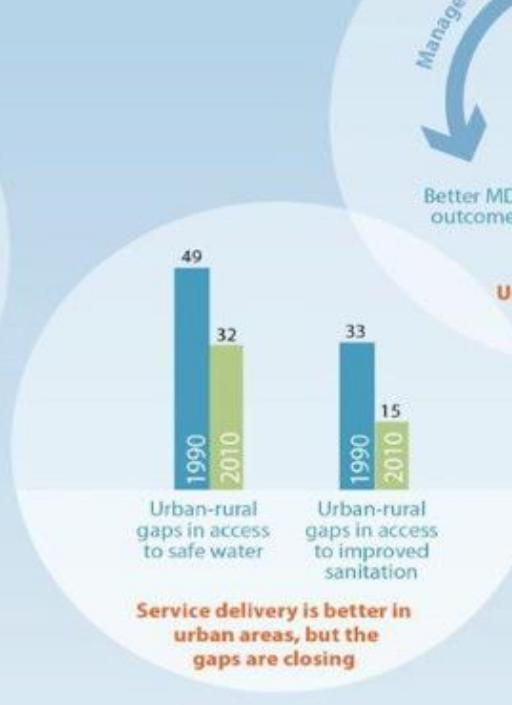
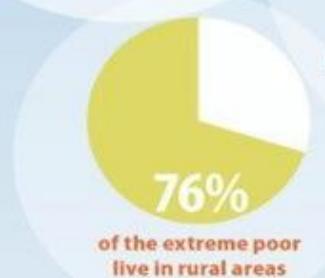
Ugly - a figure that has aesthetic problems but otherwise is clear and informative.

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





How would you evaluate
the following
visualizations?



Rural

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.