

PRINCIPLES OF DATA VISUALIZATION I

Matt Brems

DSI+

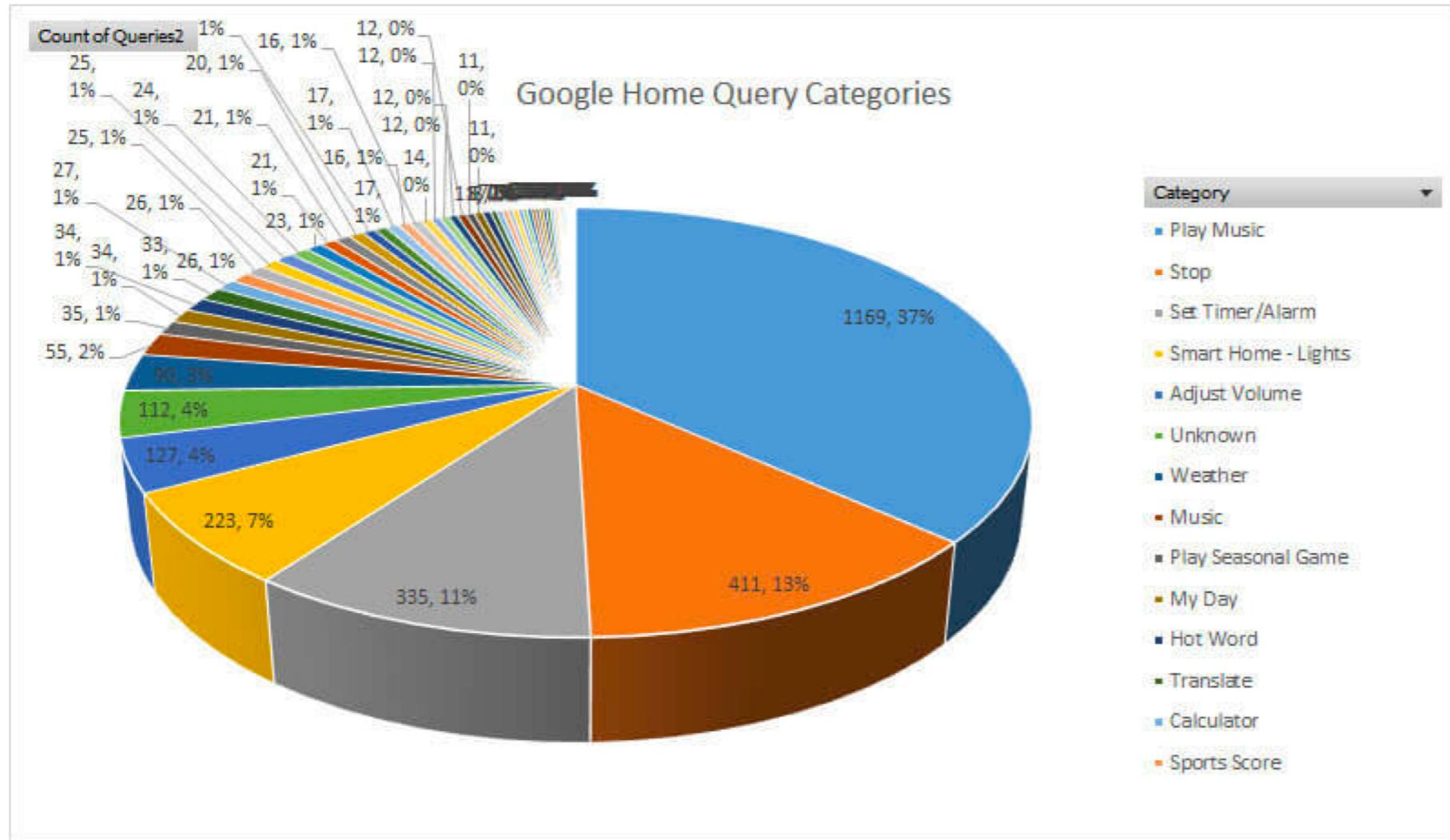
PRINCIPLES OF DATA VISUALIZATION I

LEARNING OBJECTIVES

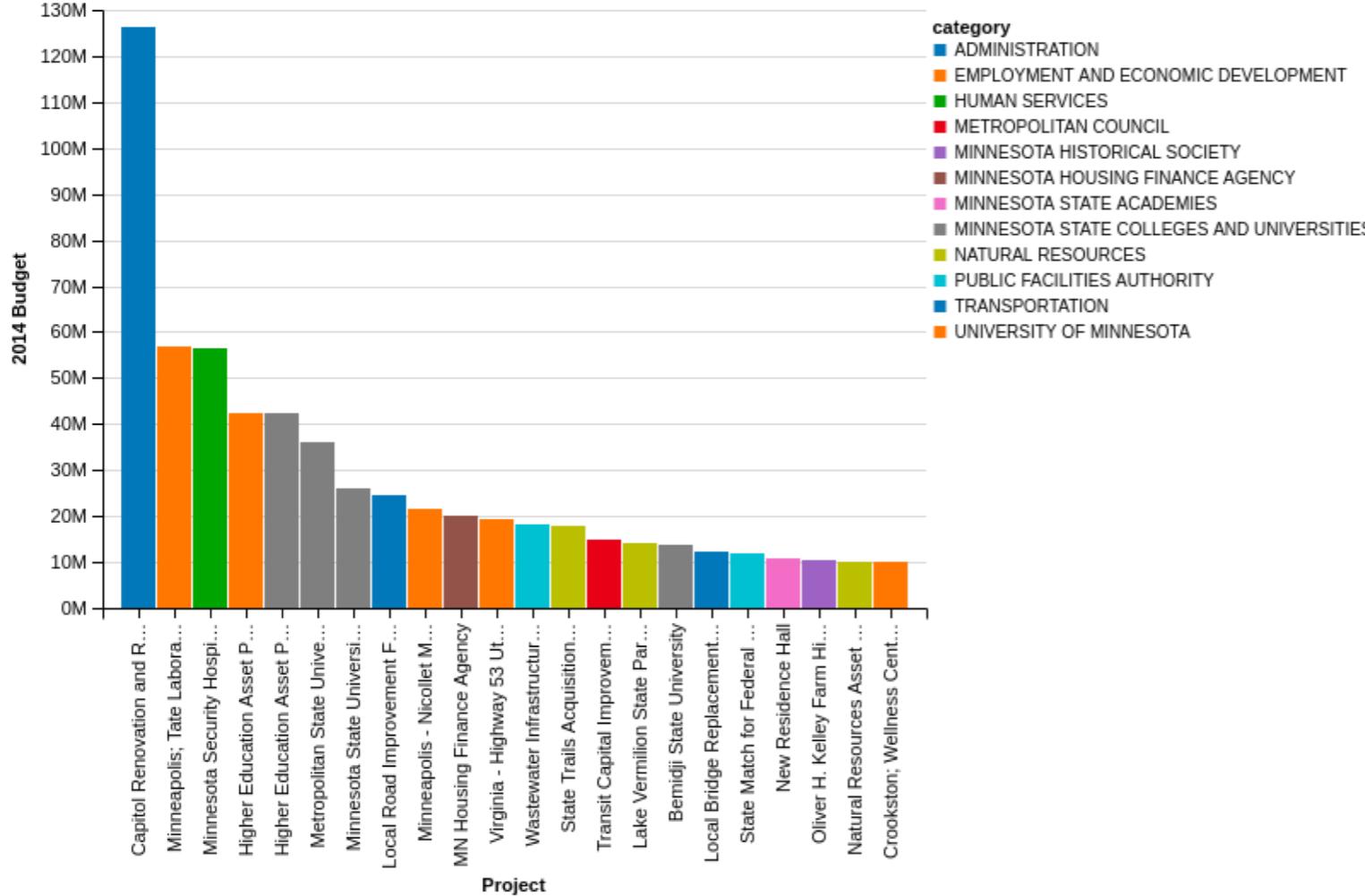
By the end of the lesson, students should be able to:

- 1. Critique data visualizations and recommend changes.
- 2. Identify what type of visualization may be appropriate for a given scenario.

GOOD OR BAD VISUALIZATION?

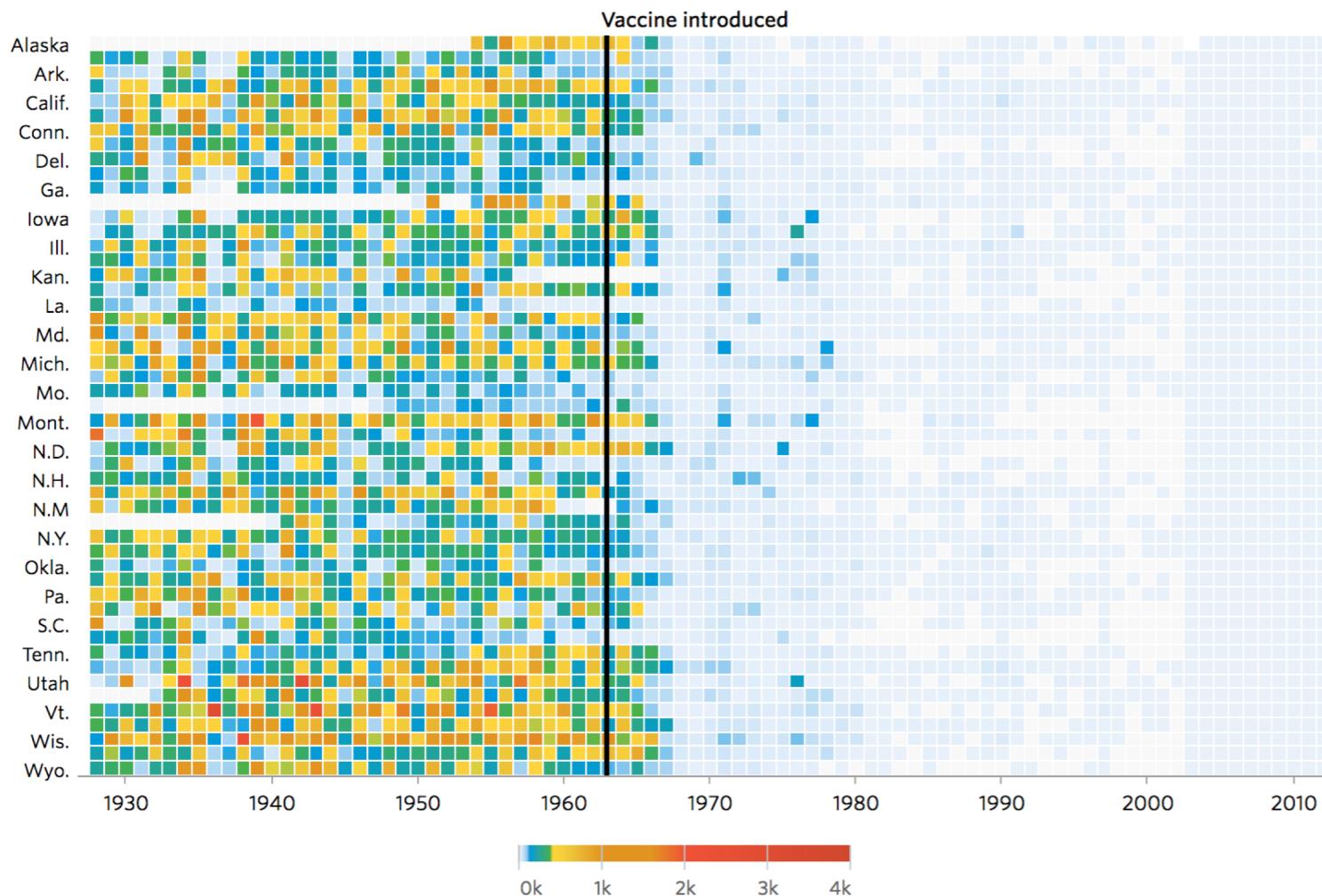


GOOD OR BAD VISUALIZATION?



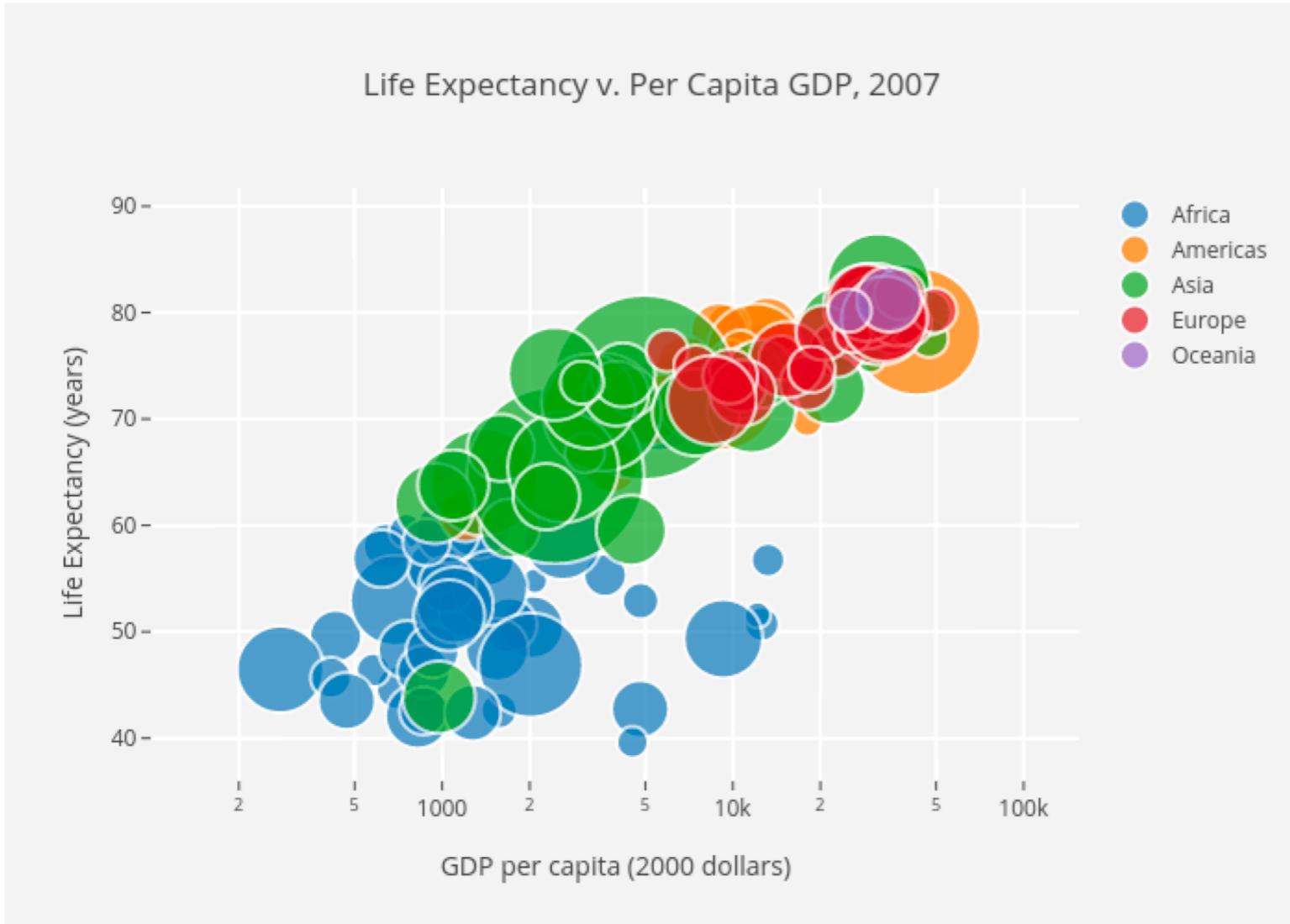
GOOD OR BAD VISUALIZATION?

Measles



Note: CDC data from 2003-2012 comes from its Summary of Notifiable Diseases, which publishes yearly rather than weekly and counts confirmed cases as opposed to provisional ones.

GOOD OR BAD VISUALIZATION?



LESSON 1: UNDERSTANDING THE CONTEXT

- When visualizing data, there are usually two types of visualizations:
 - Exploratory visualization.
 - Explanatory visualization.
- What is appropriate for presentations?

LESSON 1: UNDERSTANDING THE CONTEXT

- When visualizing data for the audience, it's important to keep in mind the “Who, What, and How?”
 - Who is your audience?
 - Who are you?

LESSON 1: UNDERSTANDING THE CONTEXT

- When visualizing data for the audience, it's important to keep in mind the "Who, What, and How?"
 - What do you need your audience to know or do?
 - If you aren't able to briefly summarize what the audience should learn or what action they should take after you communicate this information to them... is there a need to communicate this information at all?
 - How can you use data to make your point?
 - It's helpful to know this before collecting data, if possible!

LESSON 1: UNDERSTANDING THE CONTEXT

- If you're attempting to understand the context for a situation, here is a list of questions that might be helpful when consulting with stakeholders.
1. What background information is relevant/essential?
 2. Who is the audience or decision-maker? What do we know about them?
 3. What biases does our audience have that might make them supportive of or resistant to our message?
 4. What data is available that would strengthen our case? Is our audience familiar with this?
 5. What factors might weaken our case? Do we need to proactively address them?
 6. What would a successful outcome look like?
 7. If you only had a single sentence to talk to your audience, what would you say?

LESSON 1: UNDERSTANDING THE CONTEXT

- A pro-tip: We often have attachment to materials we've created on a computer. As you begin your work, consider starting “low-tech,” like on a pad of paper, before moving to a PowerPoint or Jupyter notebook.

LESSON 2: CHOOSING AN EFFECTIVE VISUAL

- While there are many, many visuals one can create, usually a handful will suffice for the vast majority of situations!
- Here, we go through twelve visuals you'll commonly encounter (and may want to consider using).

LESSON 2: CHOOSING AN EFFECTIVE VISUAL

91%

Simple text

	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%

Table



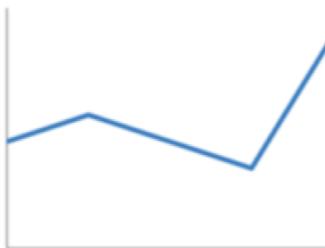
Scatterplot



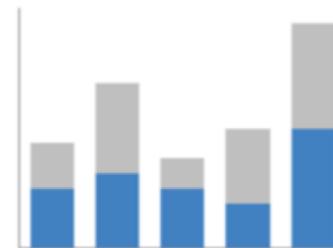
Vertical bar



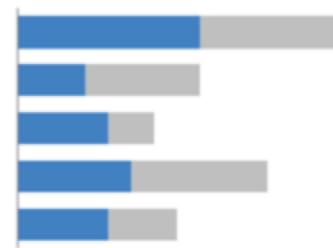
Horizontal bar



Line



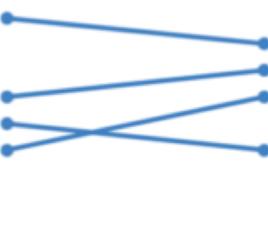
Stacked vertical bar



Stacked horizontal bar

	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%

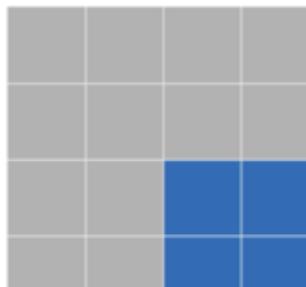
Heatmap



Slopegraph



Waterfall



Square area

LESSON 2: CHOOSING AN EFFECTIVE VISUAL – SIMPLE TEXT

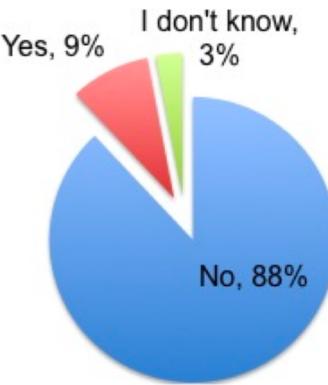
- If you just have one or two numbers to communicate, simple text might do the job.
- Be careful about any context you leave out by only using one number.

In a recent survey, we asked our users if there is anything they would like to see us change

9% said yes.

88% responded no, 3% responded / don't know.

Would you change anything?



LESSON 2: CHOOSING AN EFFECTIVE VISUAL – TABLE

- Tables are good when communicating to a mixed audience where each member can find their row/column of interest.
- Tables are usually a bad idea in a live presentation.
- Don't let borders or other design elements compete for attention!

	Q1	Q2	Q3
Actual	15	20	22
Projected	17	21	20
Delta	-2	-1	+2

	Q1	Q2	Q3
Actual	15	20	22
Projected	17	21	20
Delta	-2	-1	+2

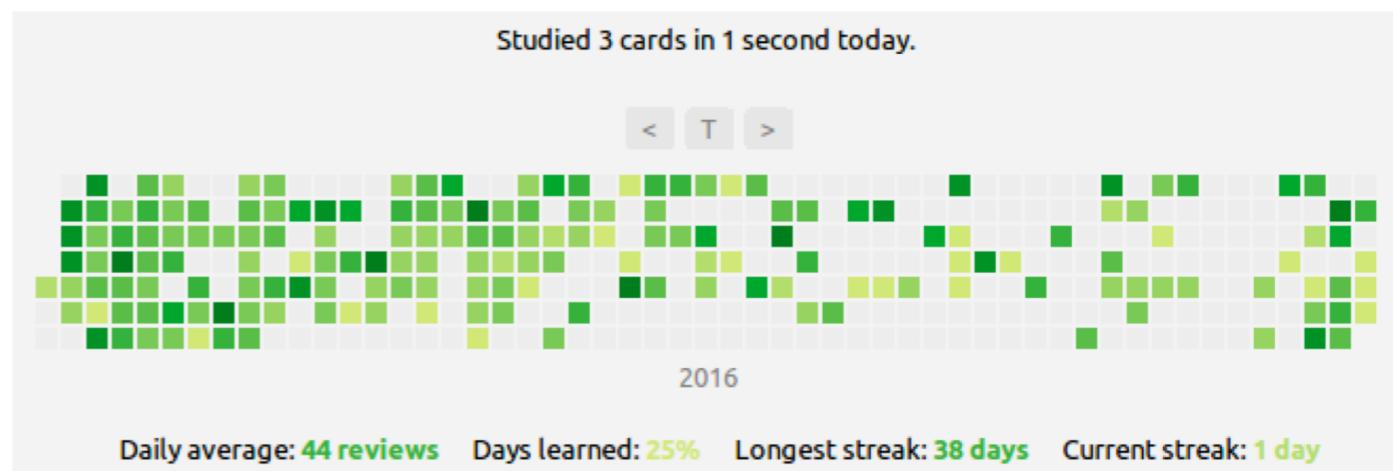
	Q1	Q2	Q3
Actual	15	20	22
Projected	17	21	20
Delta	-2	-1	+2

LESSON 2: CHOOSING AN EFFECTIVE VISUAL – HEATMAP

- Heatmaps are good in a situation where we may want to have a larger table in a live presentation, but want to guide the audience's eye to particular areas of the table.
- CNK recommends using one color along a gradient instead of multiple colors.

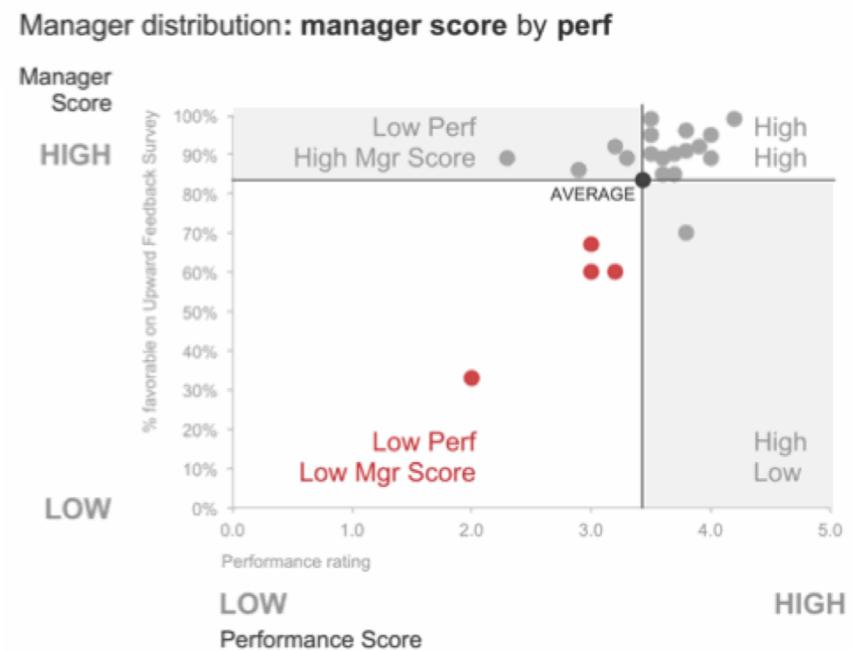
Question	Movie1	Movie2	Movie3	Movie4	Movie5	Movie6	Movie7	Movie8	Movie9	Movie10	Average by Movie
Question1	10.00	8.76	4.33	5.00	7.40	4.00	7.00	4.00	7.20	9.20	6.69
Question2	8.00	4.00	7.00	7.90	6.50	3.50	5.00	5.50	7.80	6.80	6.20
Question3	7.00	9.00	3.67	0.00	1.50	8.30	1.80	2.20	8.40	0.00	4.19
Question4	6.00	7.00	8.00	4.67	4.00	3.67	4.00	6.50	1.50	4.00	4.93
Question5	3.00	4.00	8.50	6.90	3.00	7.80	6.00	7.90	3.60	6.60	5.73
Question6	4.80	4.20	4.20	6.00	4.20	2.80	3.80	4.20	4.20	5.60	4.40
Average by Question	6.47	6.16	5.95	5.08	4.43	5.01	4.60	5.05	5.45	5.37	5.36

RANKINGS 10.0 8.0 5.0 3.0 0.0



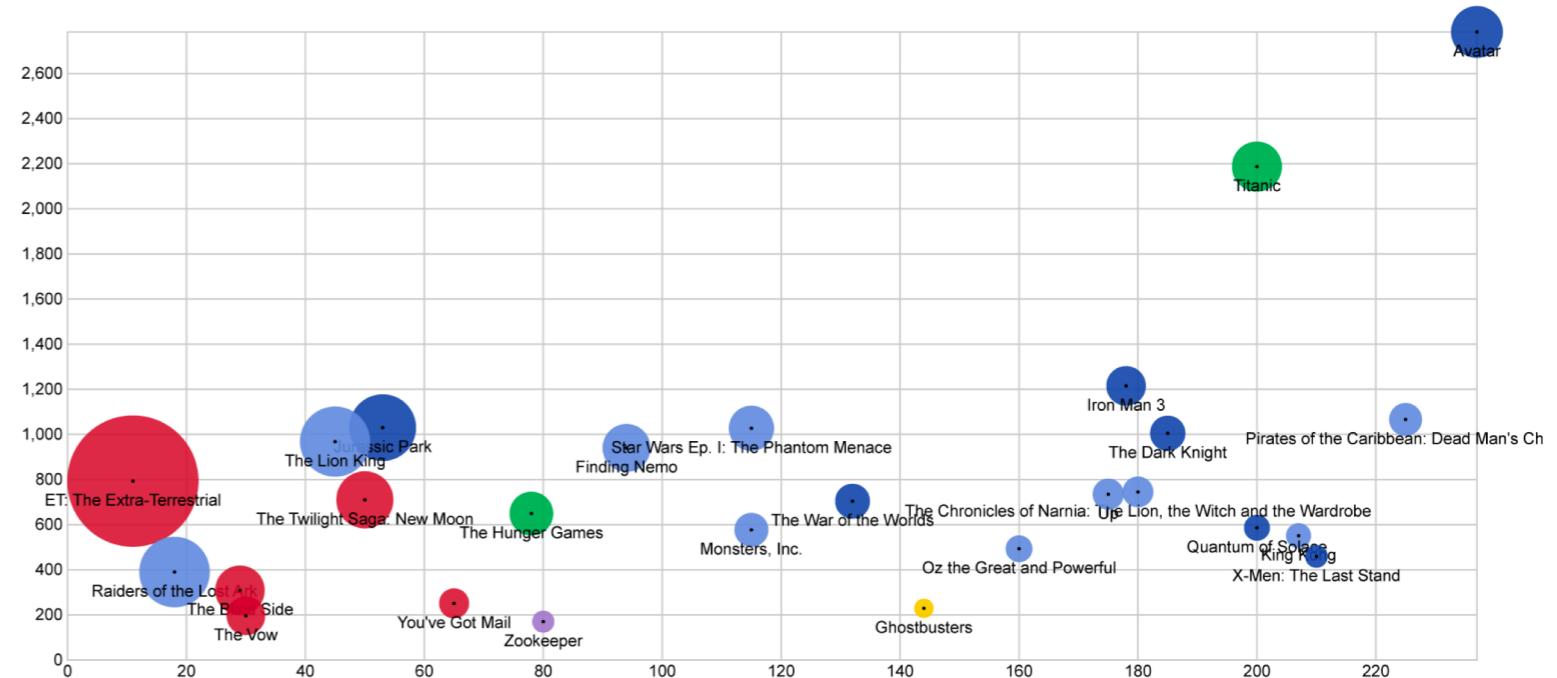
LESSON 2: CHOOSING AN EFFECTIVE VISUAL – SCATTERPLOT

- Scatterplots are great for comparing two quantitative variables.
 - We can easily visualize relationships.
- These can be easily adapted to provide context.
 - Be careful to not add too many dimensions to your data without clearly explaining them!



LESSON 2: CHOOSING AN EFFECTIVE VISUAL – SCATTERPLOT

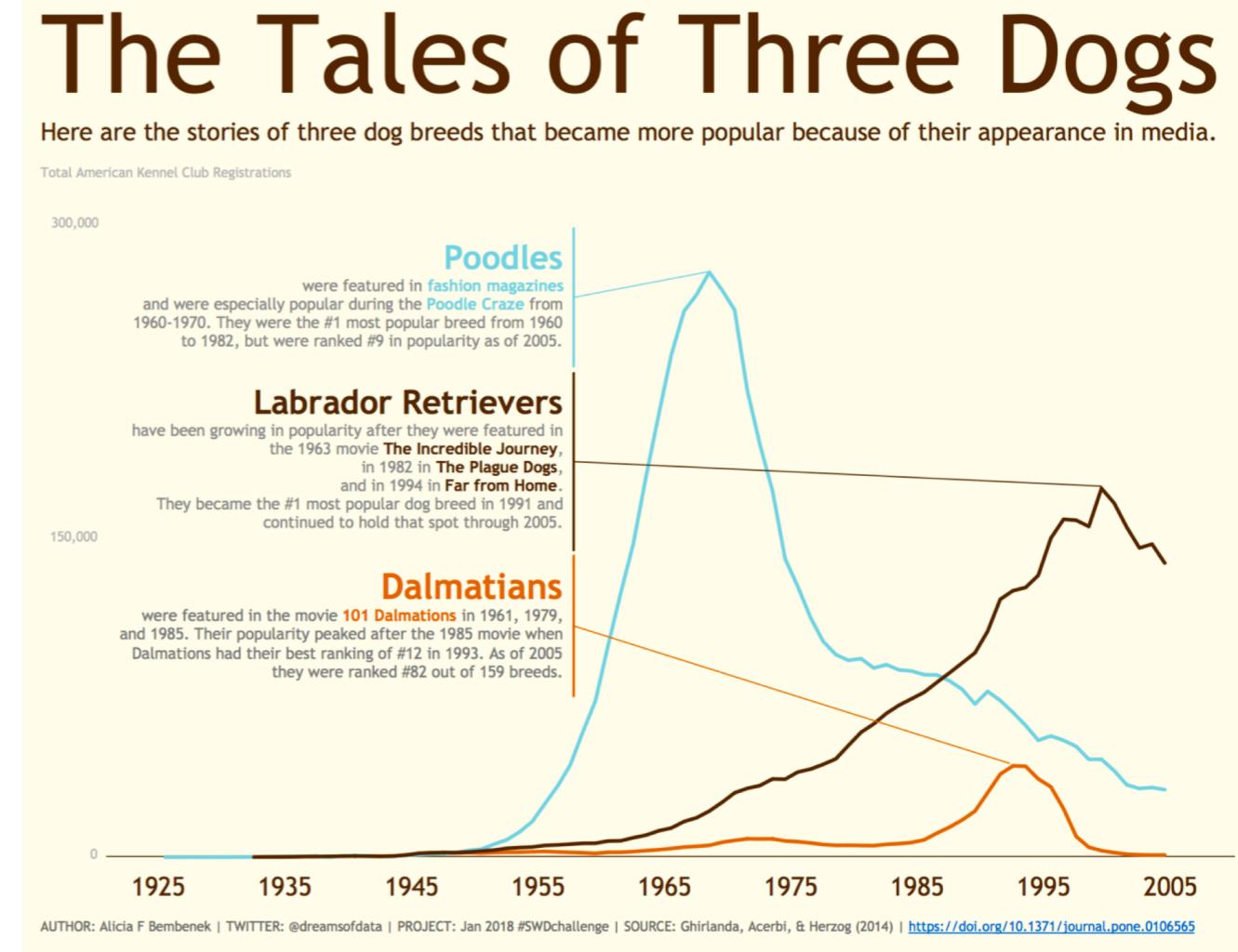
- Be careful to not add too many dimensions to your data without clearly explaining them!



- Great example of 5-D graph can be found here:
[https://www.gapminder.org/tools/#\\$chart-type=bubbles](https://www.gapminder.org/tools/#$chart-type=bubbles)

LESSON 2: CHOOSING AN EFFECTIVE VISUAL – LINE GRAPH

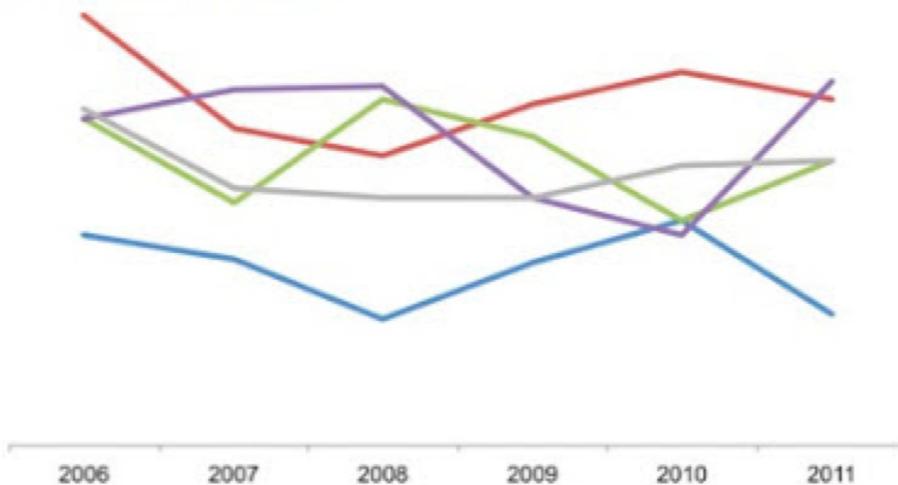
- Line plots are helpful for showing one or more series of data, usually over time.
- Make sure that the intervals are plotted in consistent intervals! (i.e. there isn't a gap on the x -axis.)



LESSON 2: CHOOSING AN EFFECTIVE VISUAL – LINE GRAPH

The Spaghetti Graph

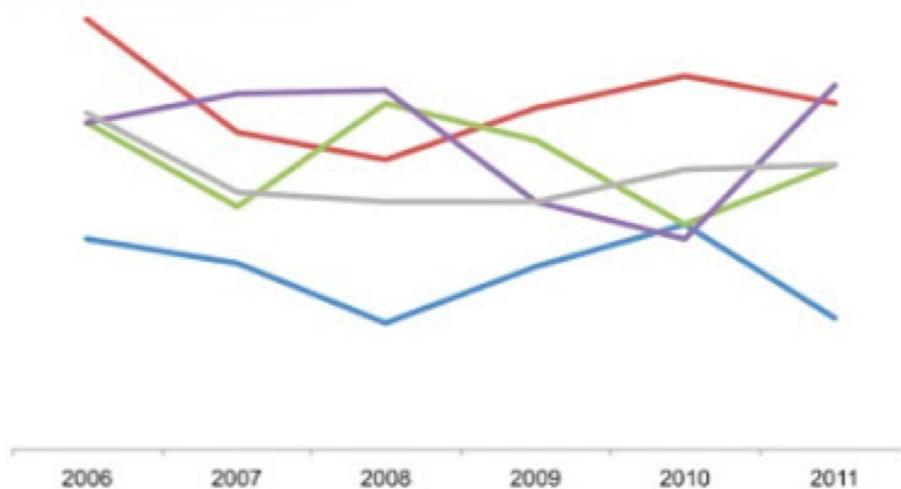
It doesn't even matter what the details are because it is nearly impossible to pull any insight from this visual



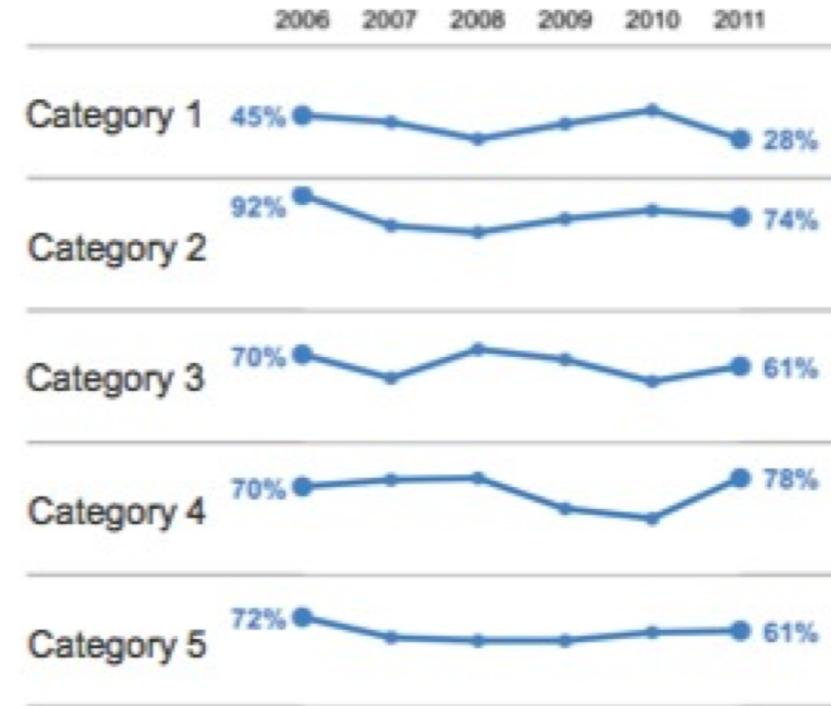
LESSON 2: CHOOSING AN EFFECTIVE VISUAL – LINE GRAPH

The Spaghetti Graph

It doesn't even matter what the details are because it is nearly impossible to pull any insight from this visual



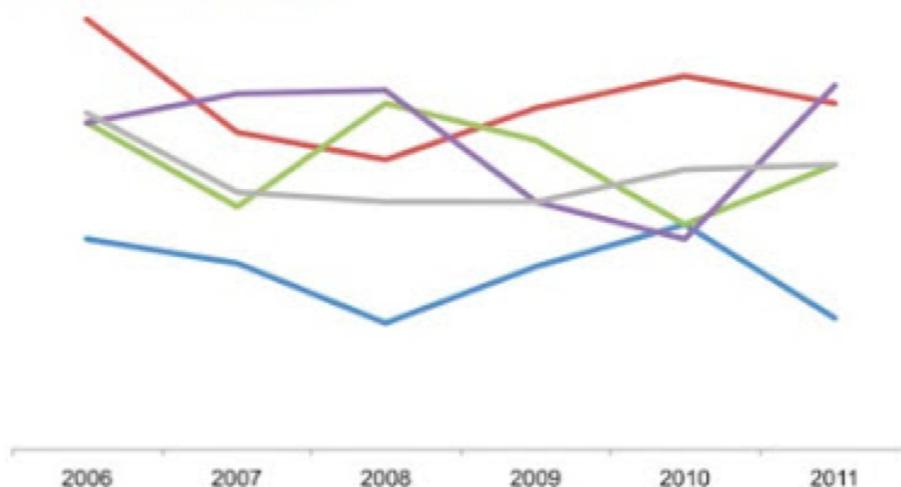
The Spaghetti Graph: Untangled



LESSON 2: CHOOSING AN EFFECTIVE VISUAL – LINE GRAPH

The Spaghetti Graph

It doesn't even matter what the details are because it is nearly impossible to pull any insight from this visual



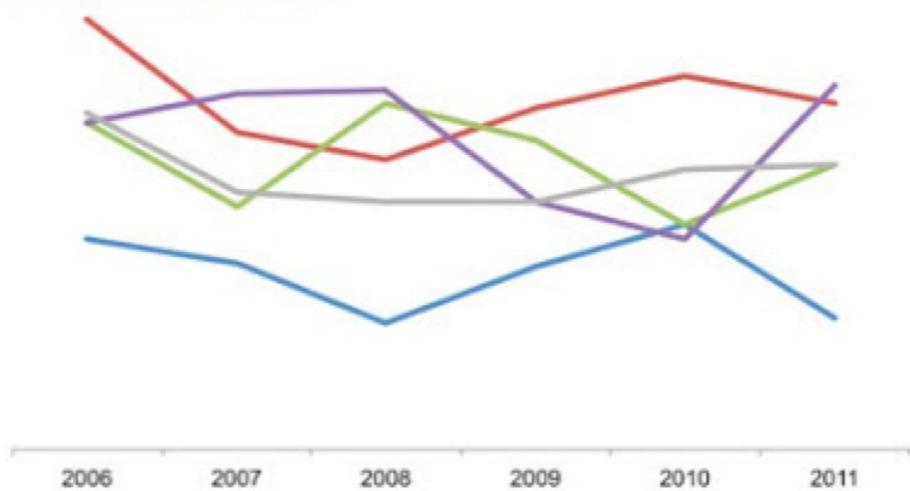
Budget over time by category



LESSON 2: CHOOSING AN EFFECTIVE VISUAL – LINE GRAPH

The Spaghetti Graph

It doesn't even matter what the details are because it is nearly impossible to pull any insight from this visual

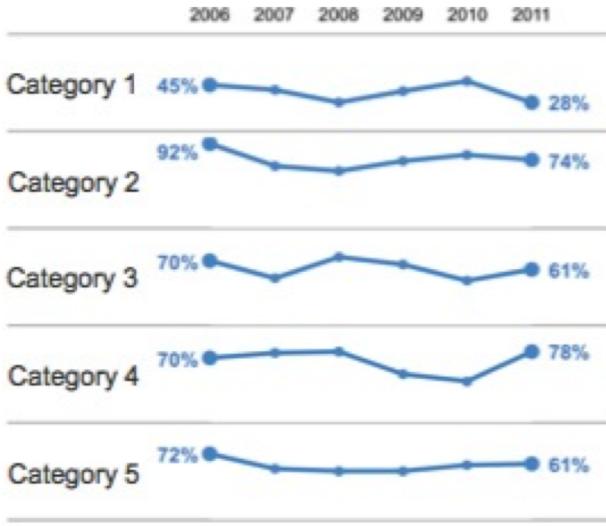


Budget over time by category



LESSON 2: CHOOSING AN EFFECTIVE VISUAL – LINE GRAPH

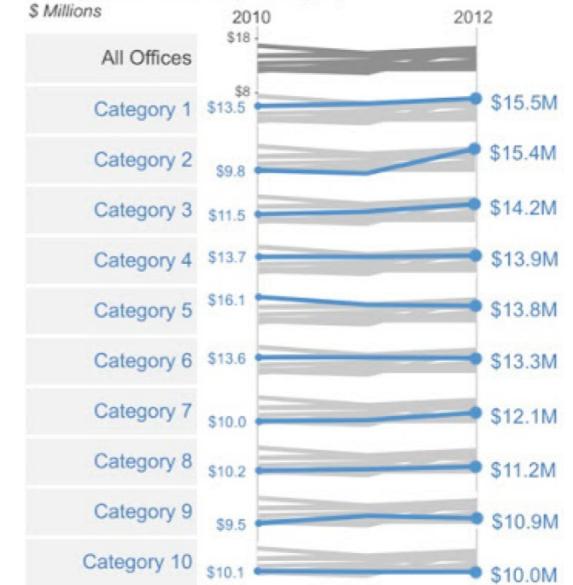
The Spaghetti Graph: Untangled



Budget over time by category



Budget over time by category



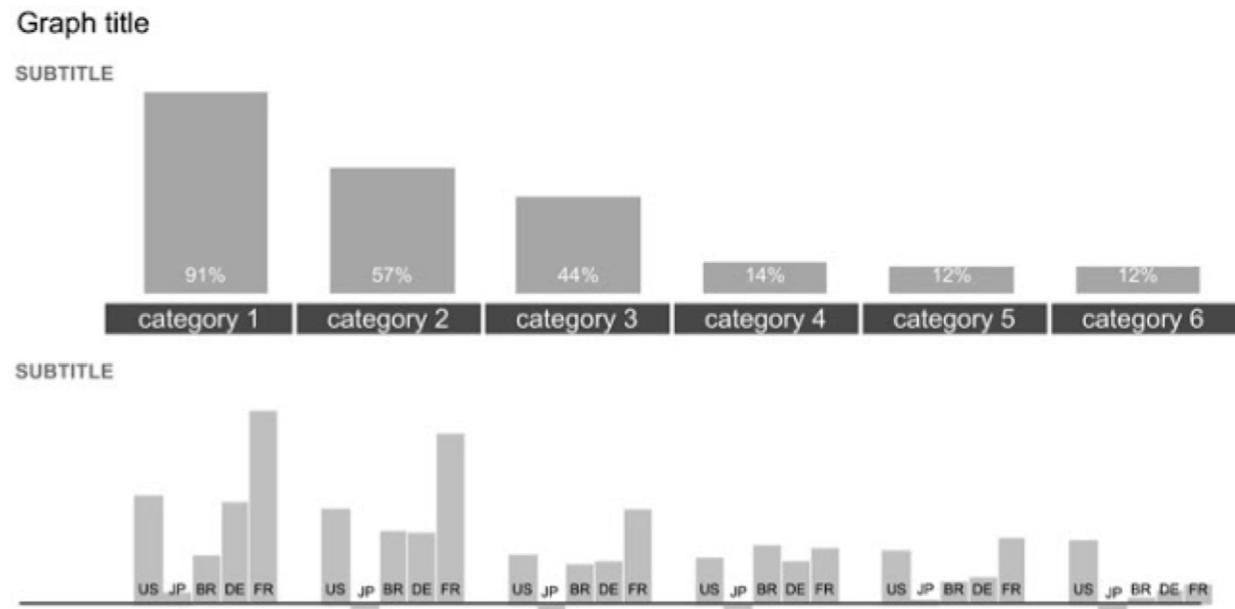
LESSON 2: CHOOSING AN EFFECTIVE VISUAL – SLOPEGRAPH

- Slopegraphs are useful for comparing categories over two points of comparison (usually two time periods).
 - You can see the absolute change and the visual increase/decrease.
- Use color to draw attention to a small group of categories.
- May work poorly if there are many overlapping categories.



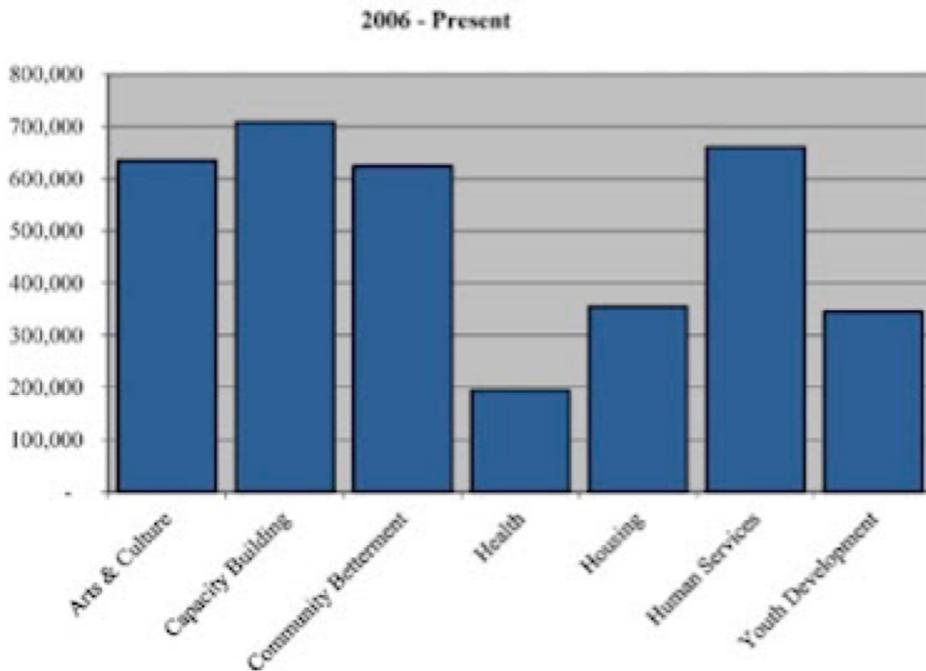
LESSON 2: CHOOSING AN EFFECTIVE VISUAL – VERTICAL BAR CHART

- Vertical bar charts can visually display one or more series of data broken out by categories.
- You will want to pay special attention to spacing to ensure that we don't confuse which category is which.
- If you have too many series, the graph may become unreadable.



LESSON 2: CHOOSING AN EFFECTIVE VISUAL – HORIZ. BAR CHART

Investment by area of impact



We invest primarily in four areas

Since we began investing in 2006, **four areas have received more than \$600K each, accounting for 75% of total grantmaking activity**

Investment by Area of Impact
2006 - Present

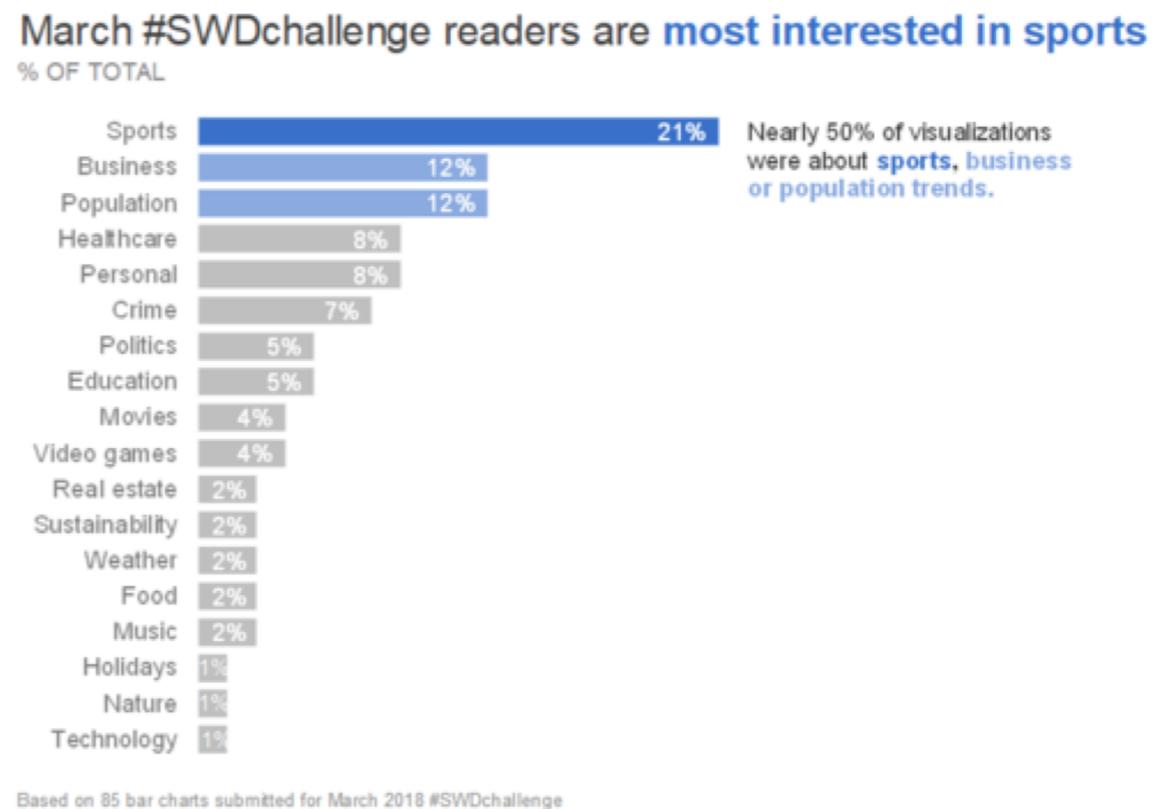
	Dollars in '000s
Capacity Building	\$710
Human Services	\$670
Arts & Culture	\$630
Community Betterment	\$620
Housing	\$360
Youth Development	\$340
Health	\$190

LESSON 2: CHOOSING AN EFFECTIVE VISUAL – HORIZ. BAR CHART

- Horizontal bar charts can visually display one or more series of data broken out by categories.

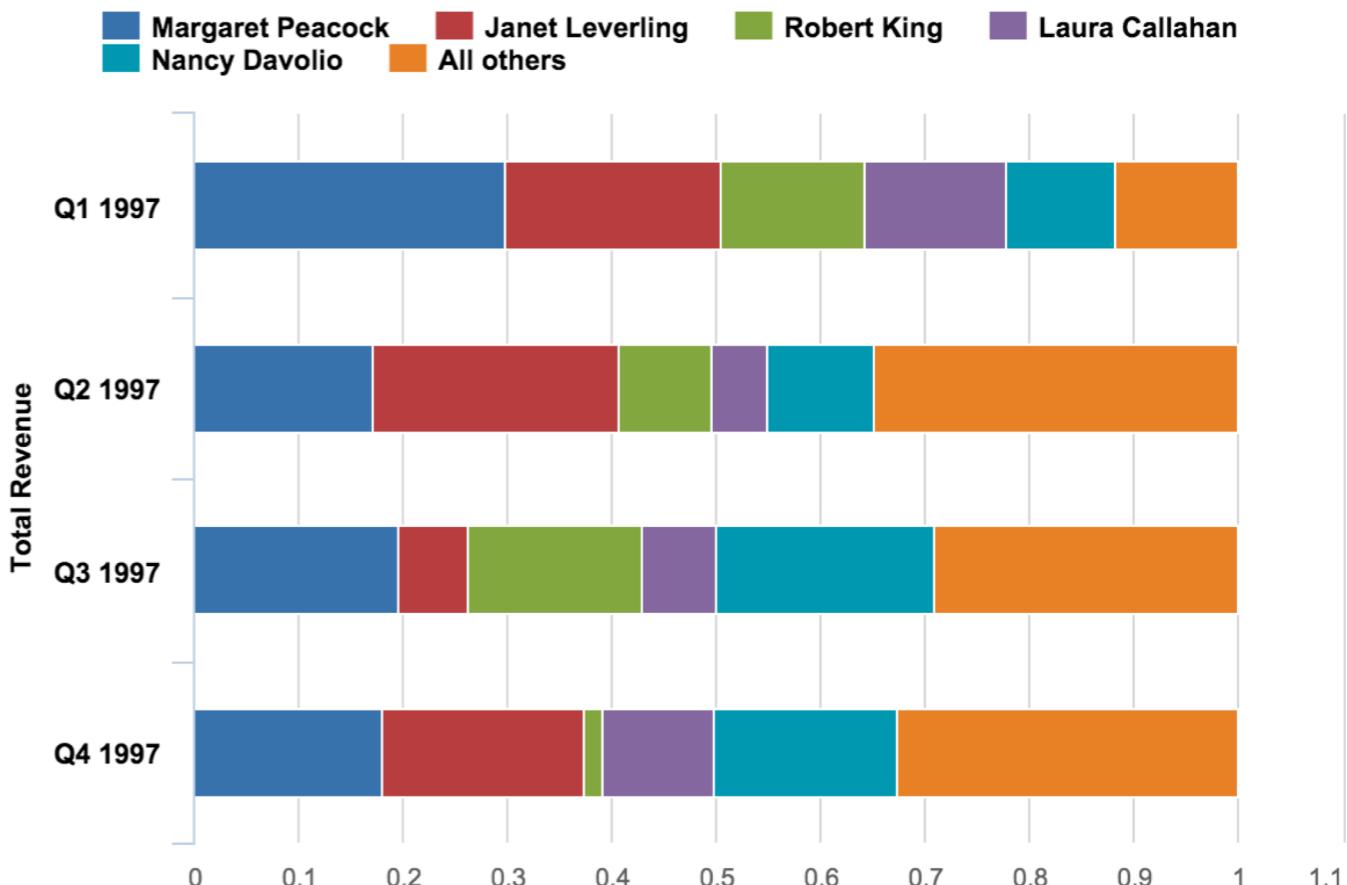
- Given the “Z” motion most eyes make when viewing charts, horizontal bar charts are usually more effective than vertical bar charts.

- We can have longer category names!



LESSON 2: CHOOSING AN EFFECTIVE VISUAL – STACKED BAR CHART

- Stacked bar charts can visually display one or more series of data broken out by categories.
- We can easily compare the value closest to the baseline(s) across categories, but it becomes more difficult for us to compare other values.

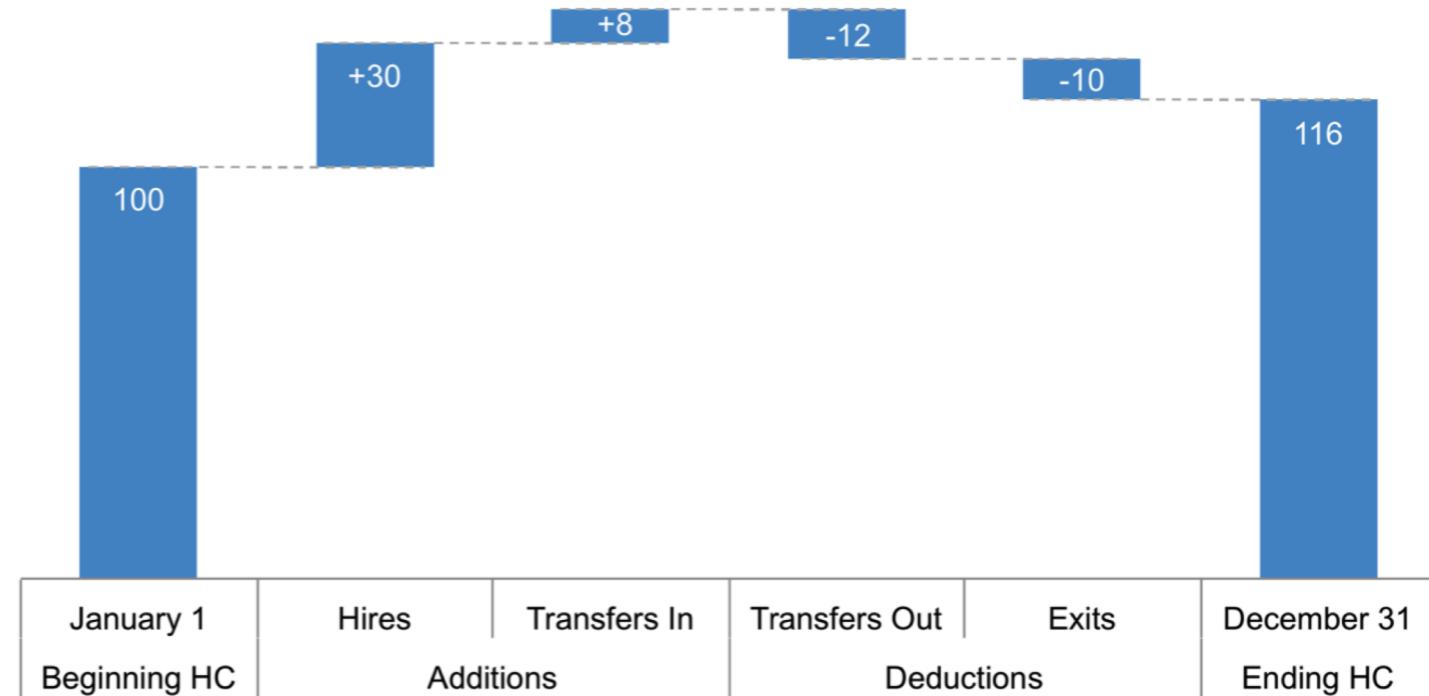


LESSON 2: CHOOSING AN EFFECTIVE VISUAL – WATERFALL CHART

- Waterfall charts are like stacked vertical bar charts for one series where we can look at changes over time.
- Want to make one of these? Use a stacked bar chart but set the color of the other bars to the same color as the background.

Headcount math

Though more employees transferred out of the team than transferred in, aggressive hiring means overall headcount (HC) increased 16% over the course of the year.



LESSON 2: CHOOSING AN EFFECTIVE VISUAL – AREA CHART

- Area charts are generally terrible - we aren't very good at visually comparing areas.
- However, a square area chart may be useful when comparing numbers of vastly different magnitude.

▪ \$1,000 invested with Warren Buffet in 1966 would be worth \$6,500,000 today.



LESSON 2: CHOOSING AN EFFECTIVE VISUAL

91%

Simple text

	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%

Table



Scatterplot



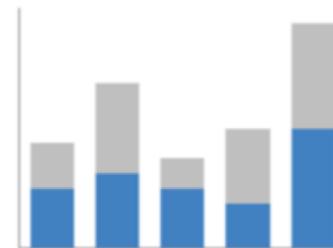
Vertical bar



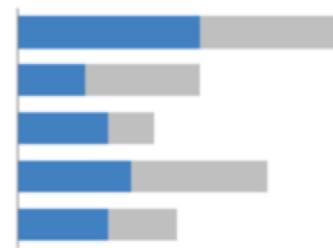
Horizontal bar



Line



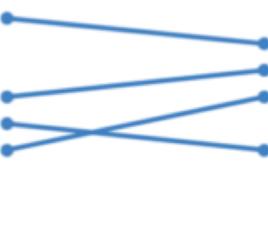
Stacked vertical bar



Stacked horizontal bar

	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%

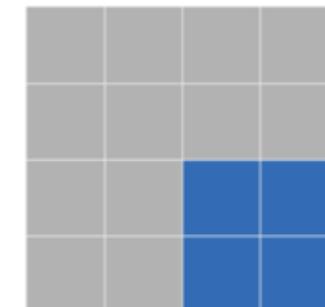
Heatmap



Slopegraph



Waterfall



Square area

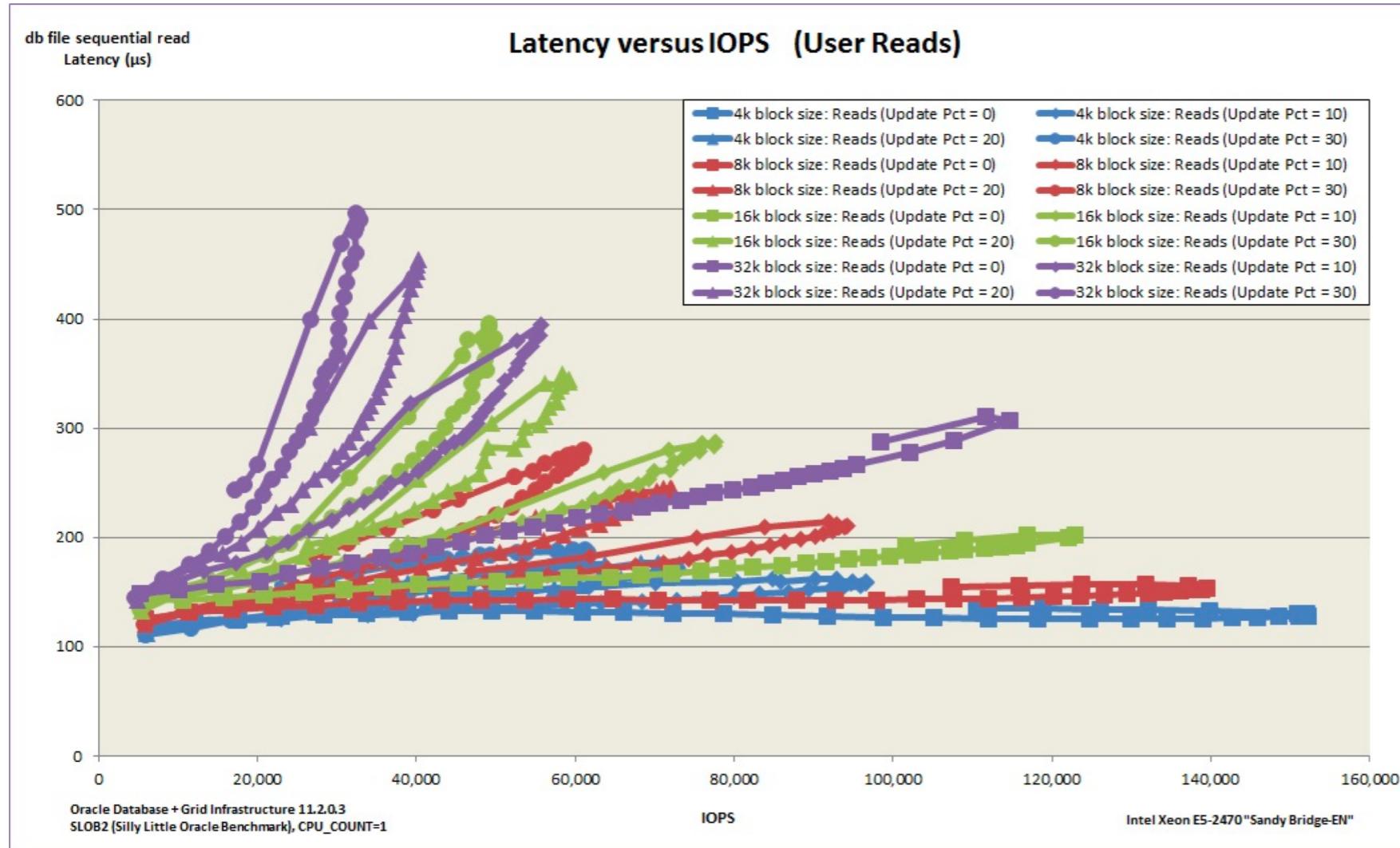
LESSON 2: CHOOSING AN EFFECTIVE VISUAL – WHAT TO AVOID?

1. **Pie charts.** We aren't good at visually comparing areas. Our interpretation of size might depend on the orientation of the pie chart. If there are many categories, some categories may not be visible. (Use a bar chart instead!)
2. **Donut charts.** They're similar to pie charts, but worse. (They're hollow and we're supposed to identify area by the length of the curve of each piece.)
3. **3-D charts.** This is confusing and makes interpretation harder, not easier.
4. **Secondary y-axes.** Rather than comparing two series on the same set of axes (which is confusing), just have two separate visuals!

LESSON 3: CLUTTER IS YOUR ENEMY!

- Clutter, or anything on a graph that doesn't contribute to a better understanding of the story you want to tell, is bad.
- We want our “data to ink” ratio to be high. (Edward Tufte.)
- Think about a blank page. Every single thing we add to that blank page takes more of our mental energy to process.

LESSON 3: CLUTTER IS YOUR ENEMY!



LESSON 3: CLUTTER IS YOUR ENEMY!

- Want to get rid of clutter? Try the following things:
 - Putting axes/grid as part of the background by making grey instead of black. (Or delete them entirely!)
 - Instead of showing an association using lines, show association using the other Gestalt principles. (i.e. same color, same shape, shading)
 - Get rid of labels that aren't important (i.e. instead of labeling all points on a graph, only label the points to which you want to draw attention!)
 - Change axes to be “in thousands” and drop the extra zeroes.
 - Use multiple graphs instead of piling everything on one graph.
- Cole Nussbaumer Knaflic presentation:
<http://www.storytellingwithdata.com/blog/2016/3/1/declutter-your-data-visualizations>

LESSON 4: FOCUS AUDIENCE ATTENTION

- In presenting visualizations, you want your audience to focus on what you want to focus on.
- Having too much on one graph can be distracting or disorienting.
- Use visual hierarchy to guide your audience!

LESSON 4: FOCUS AUDIENCE ATTENTION

- People's eyes generally move in a "Z" motion, so we should leverage that where possible.

We invest primarily in four areas

Since we began investing in 2006, **four areas have received more than \$600K each, accounting for 75% of total grantmaking activity**

Investment by Area of Impact

2006 - Present

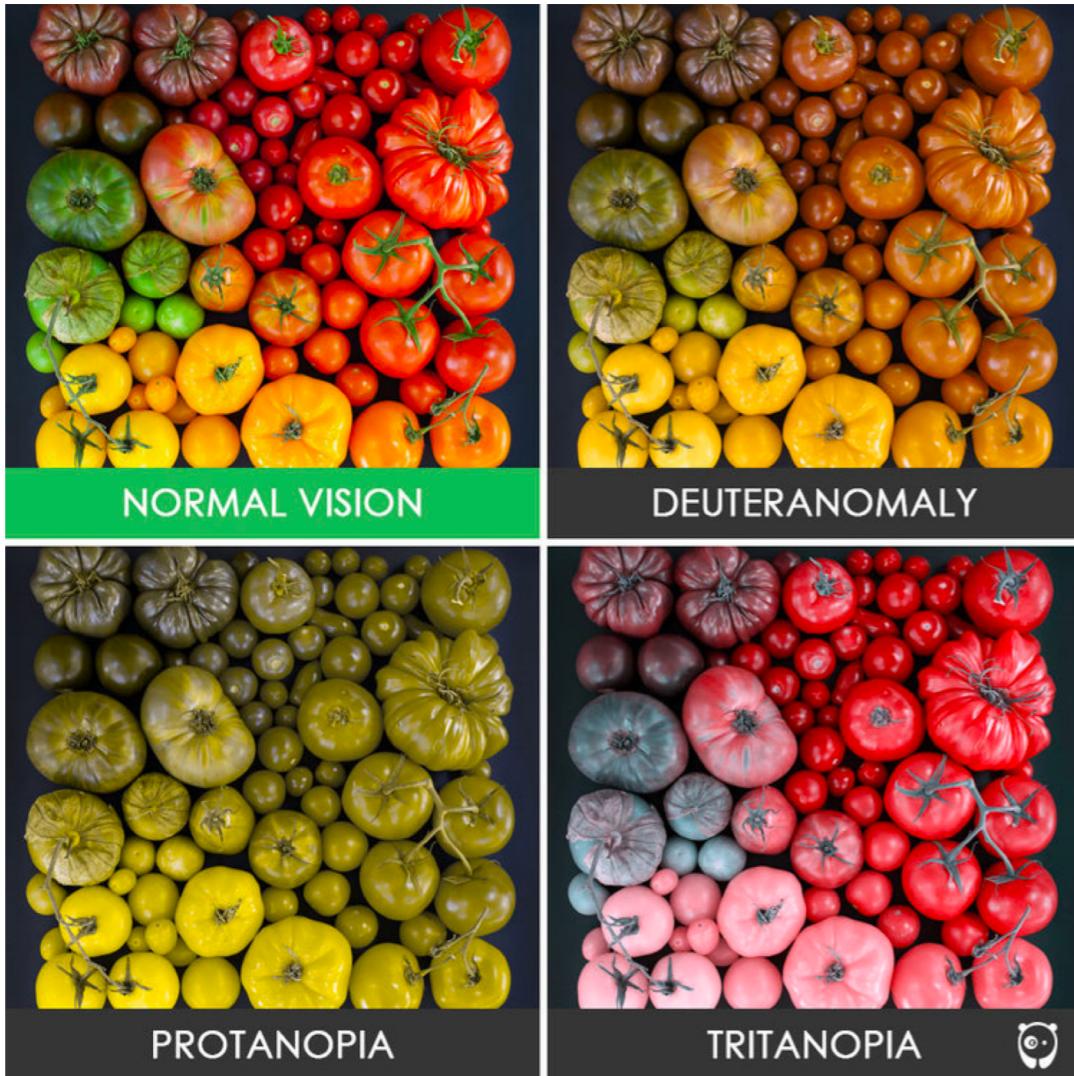
Dollars in '000s



LESSON 4: FOCUS AUDIENCE ATTENTION

- Use color to draw focus, but only sparingly. (Use gradients of one color if possible.)
- Use bold to emphasize text. Avoid italics if possible.
- When labeling your graph, don't feel obligated to write "X vs. Y" at the top. Feel free to make the label the intended takeaway of your graph!
- Anything that isn't needed – delete it! (Note that there are no axes on the previous slide.)

SIDE BAR: INCLUSIVE DESIGN



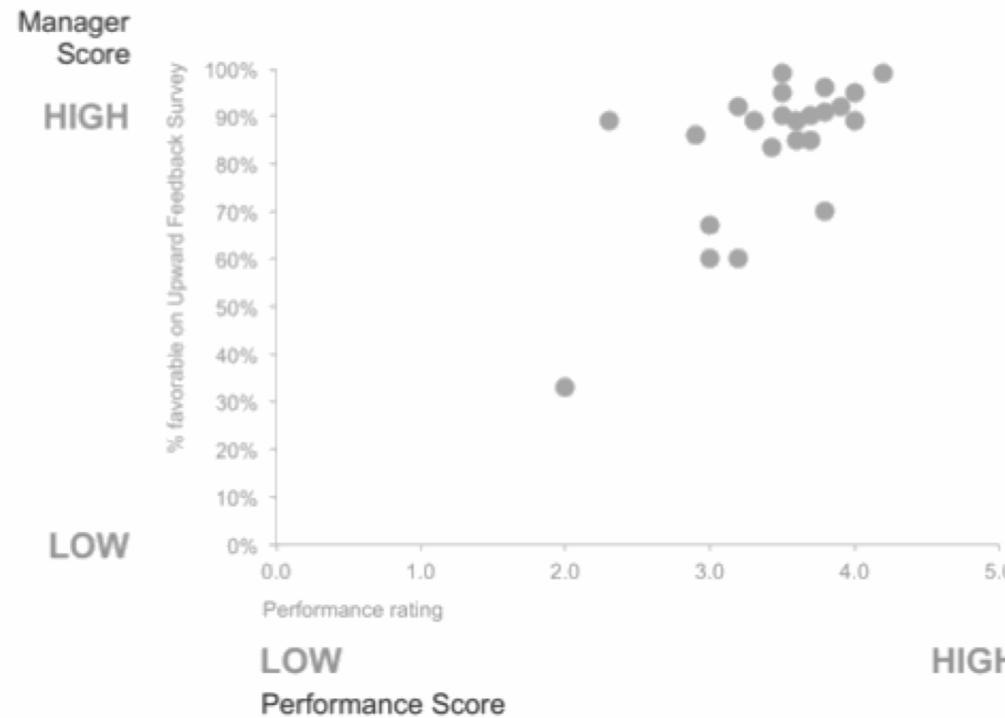
- 253,000,000 individuals across the world have some type of visual impairment.
- Red-green colorblindness (protanopia) affects one in twelve men of European descent and one in two hundred women of European descent.
- Want to see if your visual has enough contrast?
<http://www.checkmycolours.com/>

LESSON 5: THINK LIKE A DESIGNER

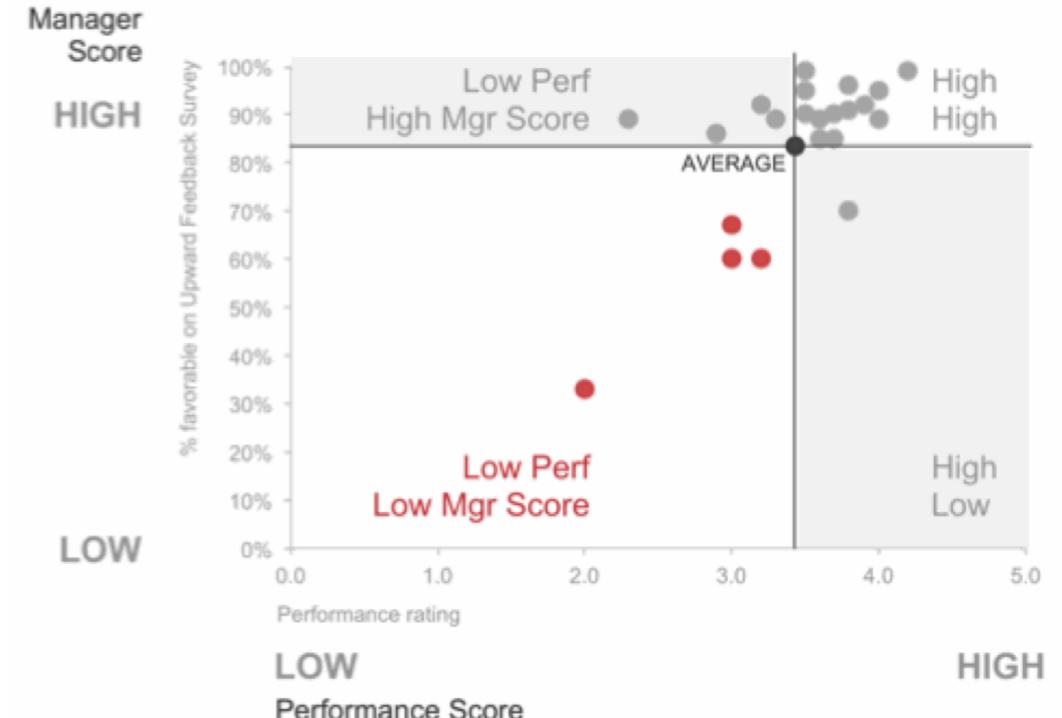
- Highlight the important stuff. (Bold, colored text, etc.)
 - At most 10% of your visual should be highlighted. (Lidwell, Holden, and Butler, 2003)
- Eliminate distractions.
 - Not all data are equally important (for visualization purposes).
 - Consider combining small groups into a larger “Other” category.
 - “Would eliminating this change anything?”
 - Push necessary but non-message-impacting items to the background.
 - Light grey works well for this.

LESSON 5: THINK LIKE A DESIGNER

Manager distribution: manager score by perf

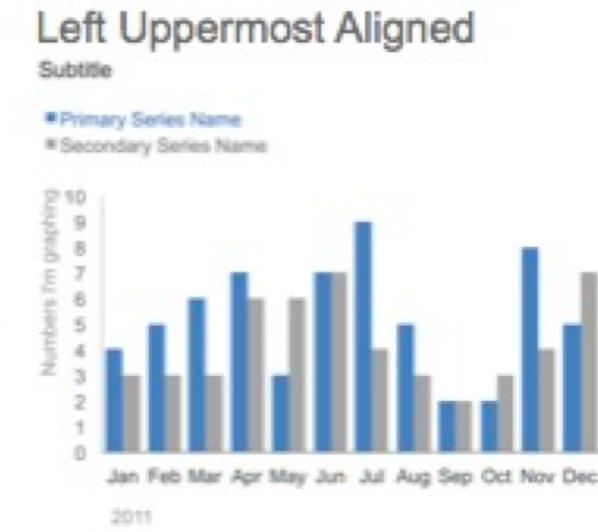
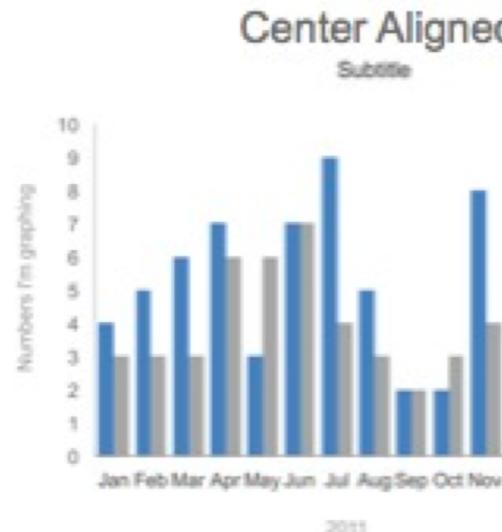


Manager distribution: manager score by perf

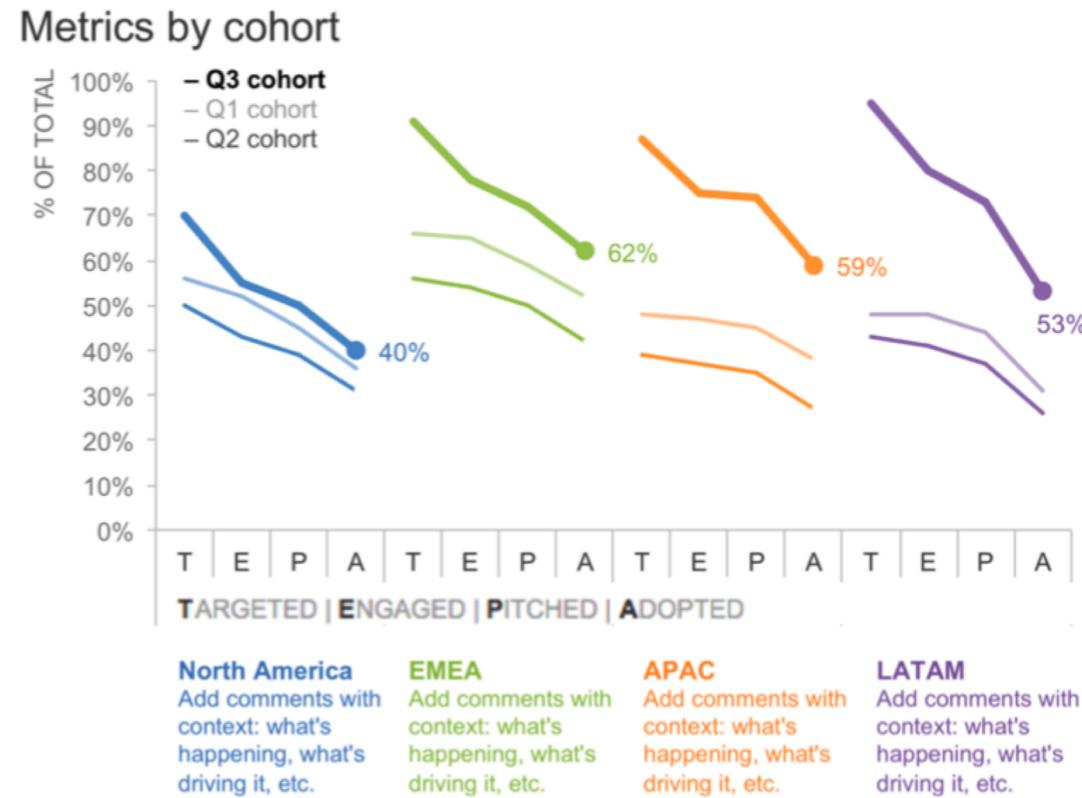
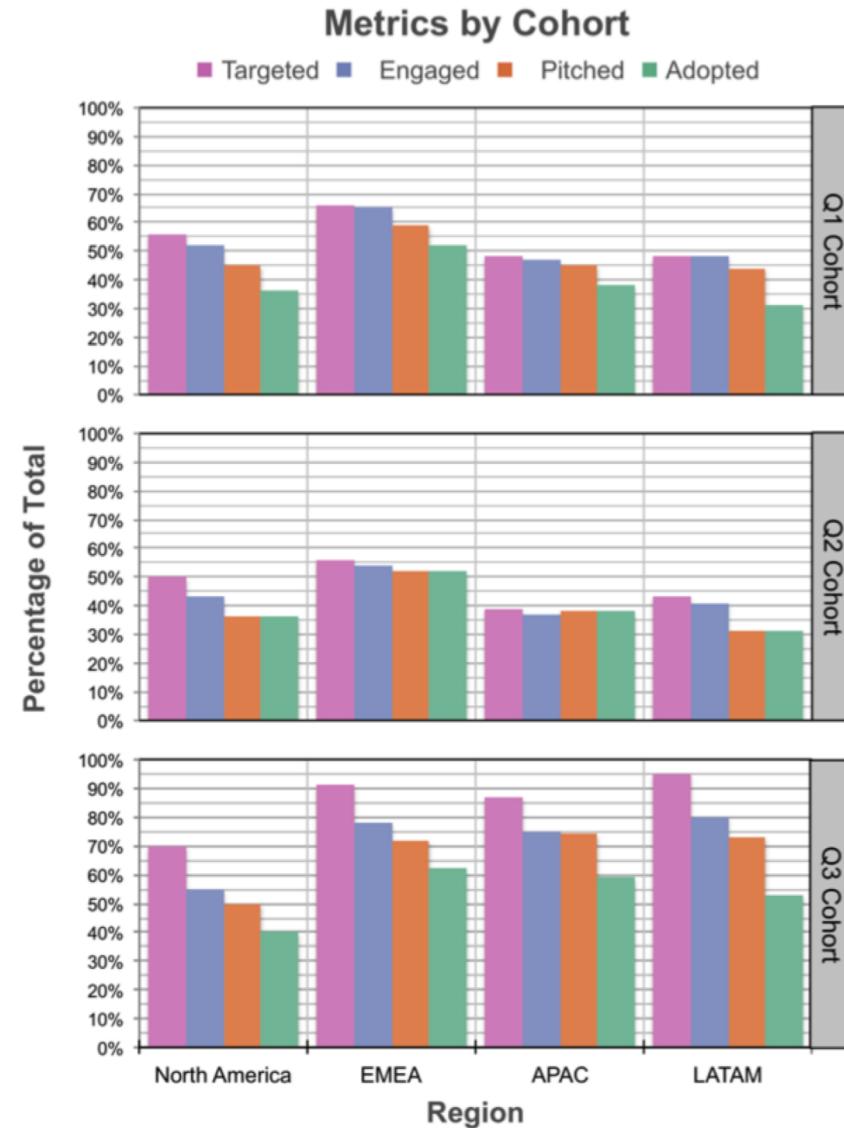


LESSON 5: THINK LIKE A DESIGNER

- Align your text!
 - Text that is centered requires mental lift on the audience.



LESSON 5: THINK LIKE A DESIGNER



LESSON 6: TELL A STORY

- We can be persuaded to act in two ways, says Robert McKee in a *Harvard Business Review* interview: through conventional rhetoric or through story.
 - Story unites people through common emotion and attention.
 - Order your story chronologically or by leading with the ending.
 - Chronologically: “So, we’re faced with this problem. How do we tackle it?”
 - Leading with the Ending: “Today, I’m going to ask you to support me in this. Let me talk about why this is so important.”

LESSONS REVISITED

- Lesson 1: Understand the context. (Who, What, How)
- Lesson 2: Choose an appropriate visual display. (What type of graph?)
- Lesson 3: Eliminate clutter. (What can I delete or make more subtle?)
- Lesson 4: Focus attention where you want it. (How do I emphasize important things?)
- Lesson 5: Think like a designer. (How do I organize my visualization?)
- Lesson 6: Tell a story! (How do I communicate my visualization?)

CREDIT

- The majority of the examples and recommendations from this lesson come from Cole Nussbaumer Knaflic's book "Storytelling with Data."

“HOMEWORK”

- As you work on your SAT project, think about a visual you have made. What changes can you make to make your visual more impactful, less cluttered, and a more coherent part of the story you want to tell?