



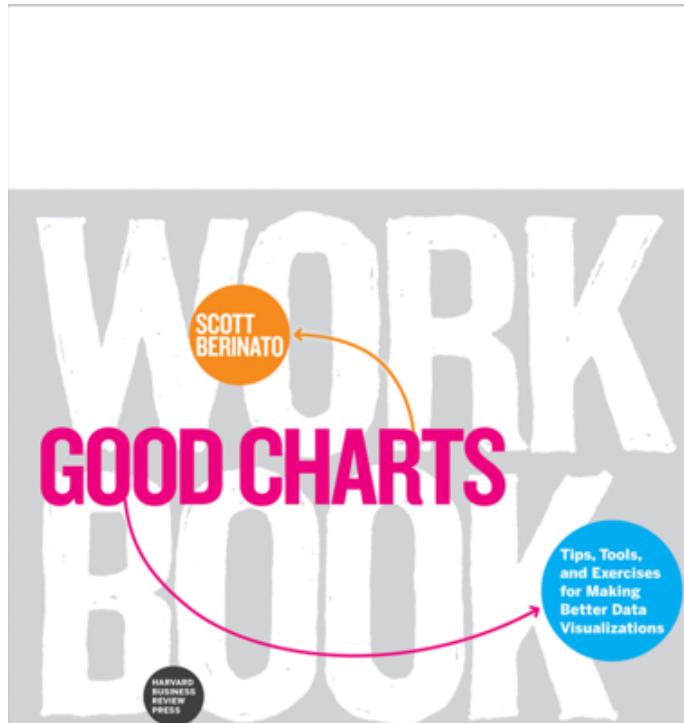
# Introduction to Data Science

Center for Data Science  
Iddo Drori, Spring 2019



# Data Visualization

# Data Visualization

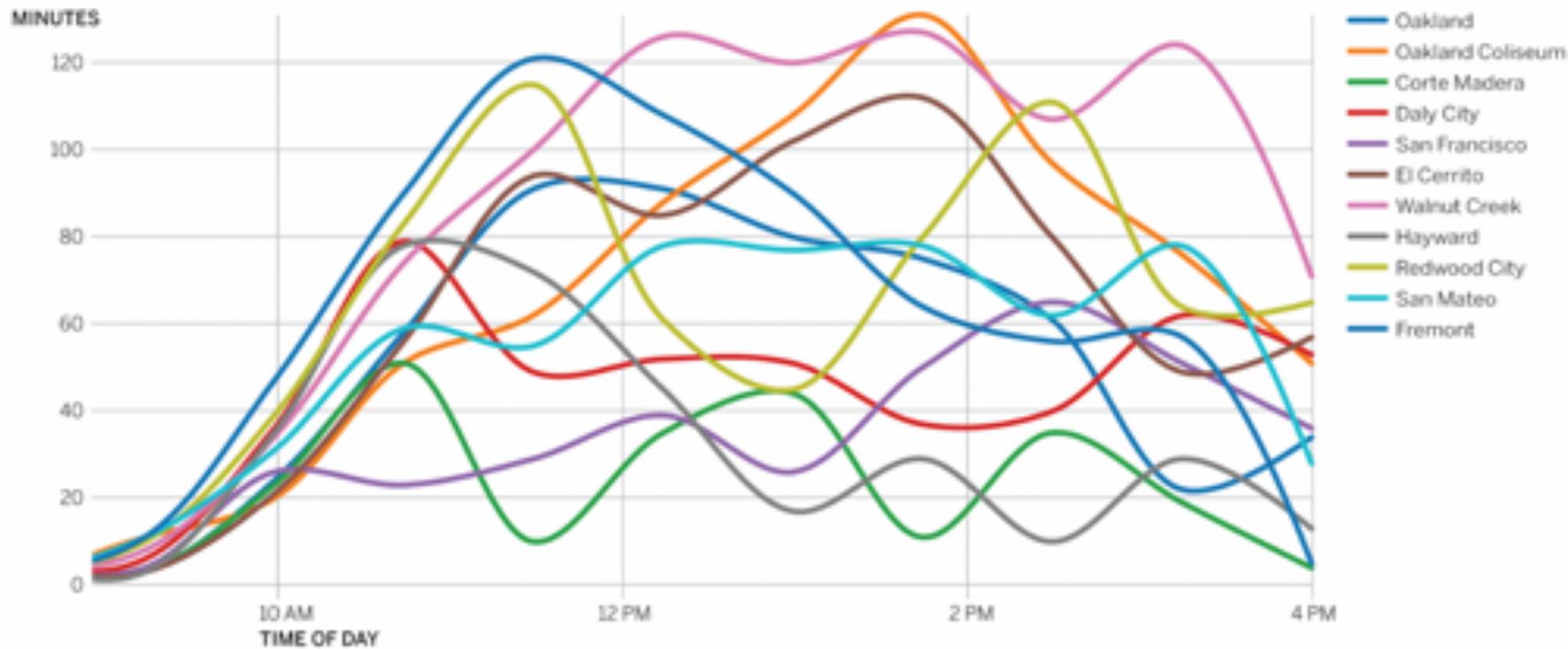


# Data Visualization

- Color
- Clarity
- Chart type
- Persuasion
- Concepts

## Color

## DMV NON-APPOINTMENT WAIT TIMES: SAN FRANCISCO VERSUS OAKLAND AREA

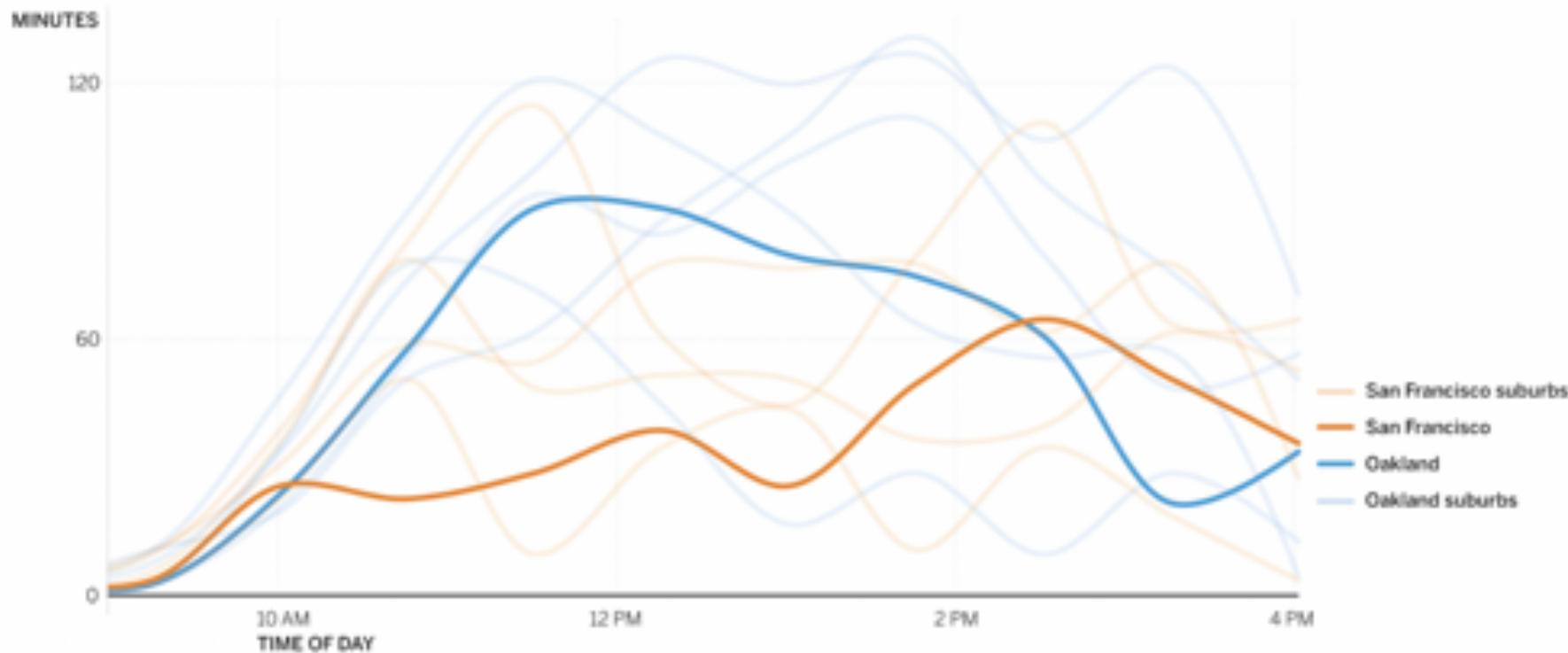


# Color

- Use less
- Use gray
- Complement or contrast
- Stick to the variables
- Think how, not which

## Color

## DMV NON-APPOINTMENT WAIT TIMES: TRY SAN FRANCISCO AT LUNCH, OAKLAND LATER



# Color

In a bar chart, you want to show comparisons between older and younger men and between older and younger women. Which color scheme do you use?

- A — Men under 20  
— Men 20–40  
— Men 40–60  
— Men over 60

- B — Men under 40  
— Men over 40

- Women under 20  
— Women 20–40  
— Women 40–60  
— Women over 60

- C — Men under 20  
— Men 20–40  
— Men 40–60  
— Men over 60

- Women under 40  
— Women over 40  
— Women under 20  
— Women 20–40  
— Women 40–60  
— Women over 60

# Color

Answer: B. The context focuses on a binary comparison—young versus old—so we shrink each gender into only two groups, under 40 and over 40. We give the two groups of men similar hues, and likewise for the women. Eight variables become four—fewer bars—and only two colors are in play. A obviously uses too much color, which would overwhelm the bar chart. The light-to-dark scheme of C maintains four distinctions per gender when we need only two. The gradient saturation also suggests different degrees of something. That could work for age, with younger people being less saturated, but it's not entirely intuitive.

# Color

In a scatter plot, you want to show the distribution of performance for four sales teams, but your goal is to highlight the performance of the European sales force against all others. Which color scheme do you use?

- A   ● Europe  
     ● North America  
     ● Asia  
     ● Africa

- B   ● Europe  
     ● North America  
     ● Asia  
     ● Africa

- C   ● Europe  
     ● North America  
     ● Asia  
     ● Africa

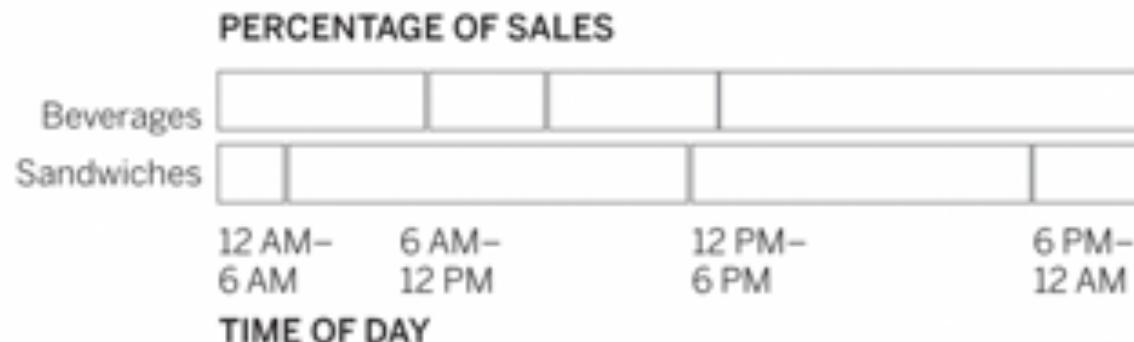
- D   ● Europe  
     ● North America  
     ● Asia  
     ● Africa

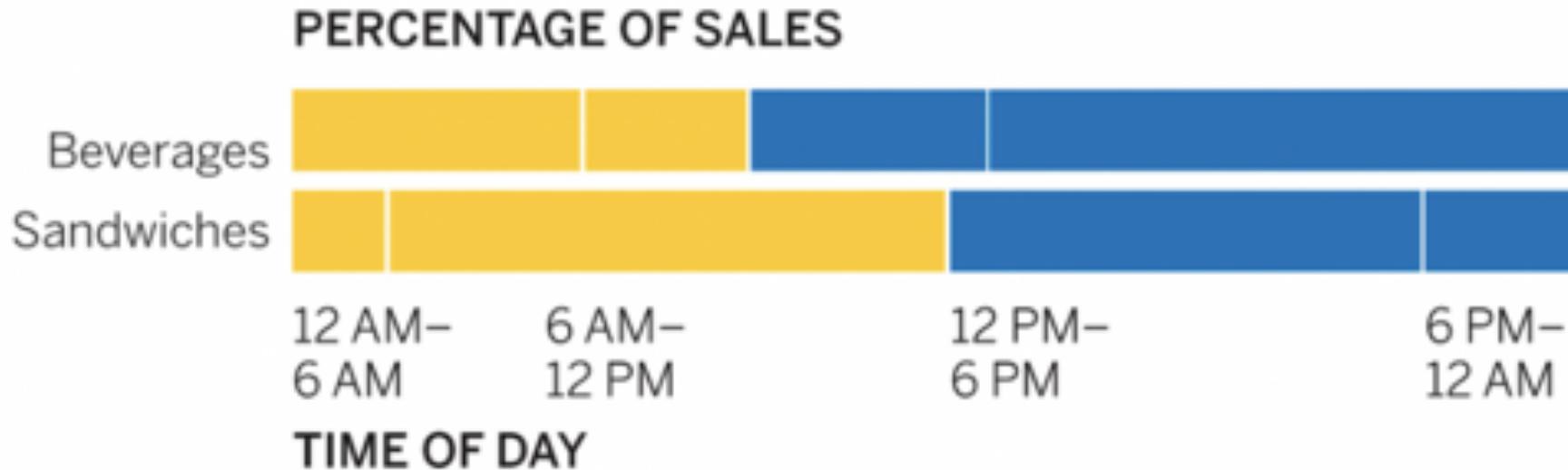
# Color

Answer: C. Europe versus all others means we want the eye to go right to Europe. The other variables exist to be compared with Europe, and distinctions among them don't matter. Giving any other region a dominant color overemphasizes it, eliminating D (four distinct colors fighting for attention) and A (two distinct color groups, though the groupings don't mean much). B wouldn't be a bad choice, but making three variables yellow would draw attention to that cluster, which would presumably have more marks on the page than Europe does, since it's three variables combined. By making the "other" group gray—and possibly not even labeling them separately but only as one variable called "other regions"—we leave no doubt: Look at Europe.

# Color

You want to compare before-noon sales to after-noon sales. Create a color scheme for the stacked bars.





# Color

You want to show answers from a Likert-type scale that ranges from “strongly agree” to “strongly disagree.” Match each survey question below to the color scheme that would be best used to describe the results of that question.

1. Please rate your feeling about the following statement: *I'm ready for the challenge of transforming this company.*
2. Please rate your feeling about the following statement: *Our leaders are ready for the challenge of transforming this company.*
3. Please rate your feeling about the following statement: *I believe in the company's strategy.*

A Strongly disagree  Strongly agree

B Strongly disagree  Strongly agree

C Strongly disagree  Strongly agree

# Color

A: 3. If you want to show strong feelings in two directions, opposing colors on the extremes and lighter shades heading into the middle works well. Here lightness reflects ambivalence, and the hues reflect positive versus negative.

B: 1. If you want to show degrees of a positive feeling (readiness), try a desaturated-to-saturated scale using a single color. Here the darkness of the pink reflects respondents' degree of readiness.

C: 2. If you want to show degrees of a negative feeling (skepticism), simply flip the previous convention and move from saturated to desaturated. Here a darker blue reflects a deeper skepticism.

In a line chart, you're comparing four price trends with an average trend. You want your audience to see the two lines that show below-average price trends. What's a good color to make the average trend line?

- A** A color similar to the ones used for the below-average trend lines, to show that they are what we want to compare with the average
- B** A color that contrasts with the below-average trend lines so that those two lines stand out
- C** Black so that it's neutral compared with the four trend lines
- D** Gray so that it's visible enough to use as a comparison but not dominant

---

Answer: C. Or possibly D. The below-average trend lines should get dominant colors, because that's where we want people to focus. At the same time, what matters is those lines' performance against average, so we don't want to overwhelm the average line. Gray might be too faint; a dark gray would probably work. If we chose A or B, we'd confuse the audience. In either case, it would appear to be another variable in the set, not an average line describing something about that set.

---

# Color

In a chart about car makers, you have many variables. Group them to reduce the number of colors used and assign a color scheme. Find a grouping that requires only two colors.

AMC	FIAT	PLYMOUTH
AUDI	FORD	PONTIAC
BMW	HONDA	RENAULT
BUICK	MAZDA	SAAB
CADILLAC	MERCEDES	SUBARU
CHEVROLET	MERCURY	TRIUMPH
CHRYSLER	NISSAN	VOLKSWAGEN
CITROËN	OLDSMOBILE	VOLVO
DATSON	OPEL	
DODGE	PEUGEOT	

# Color

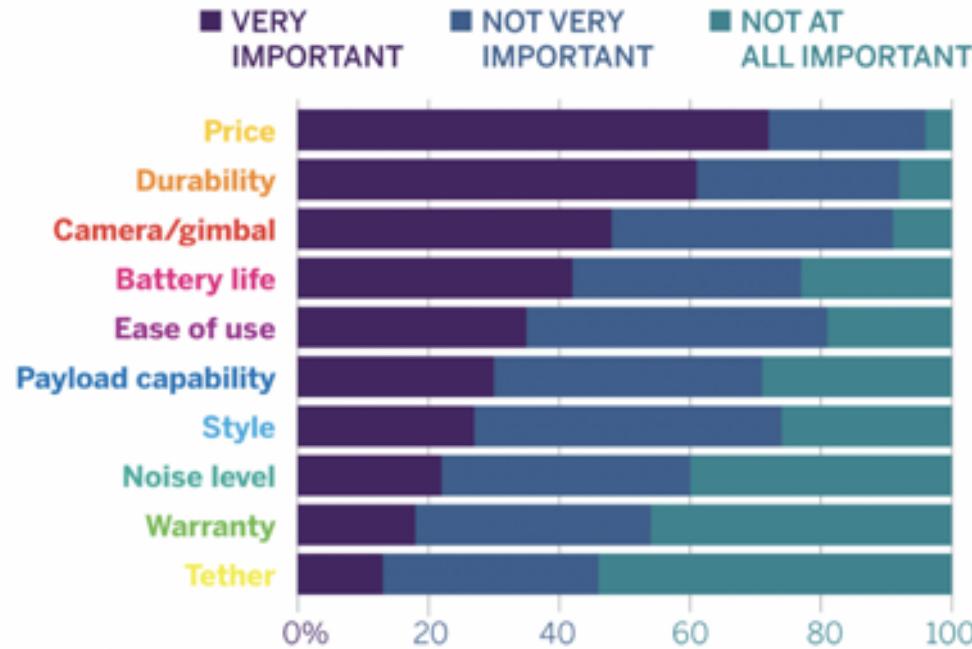
- European
- American
- Asian

- Cars in production
- Cars no longer in production

# Color

Find four places to remove color from this plot.

## WHAT'S IMPORTANT WHEN BUYING A DRONE?

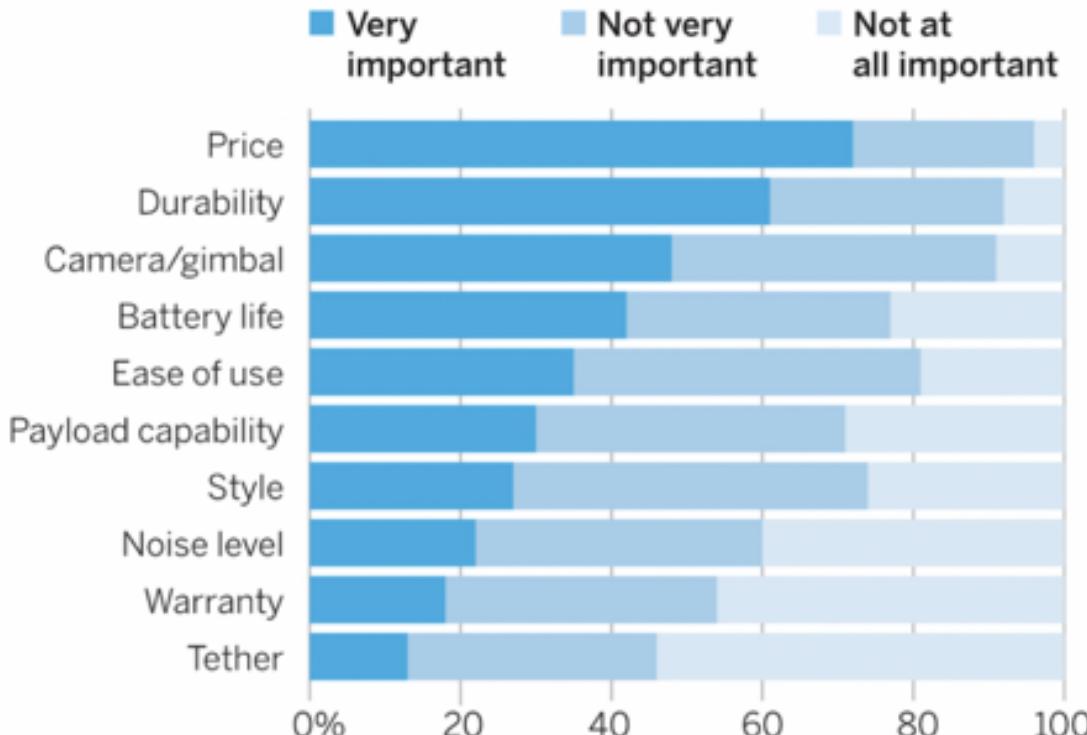


# Color

1. *The headline.* Using color here doesn't help increase the focus on the crucial idea of what's important when buying a drone. Also, using the same color for the headline and the "not at all important" value seems confusing and contradictory.
2. *The category labels.* This rainbow is unnecessary decoration, and the colors don't connect to anything else in the visualization.
3. *The key.* Sometimes making the words in a key the same color as what they represent works. Here we're already using color blocks in the key. If we keep the blocks, the words can be black.
4. *The x-axis labels.* Associating these percentages with the variables' colors is confusing. After all, 80% of people would *not* vote "not at all important." In general, labels don't need color—especially color that's been assigned to something else.

# Color

## WHAT'S IMPORTANT WHEN BUYING A DRONE?



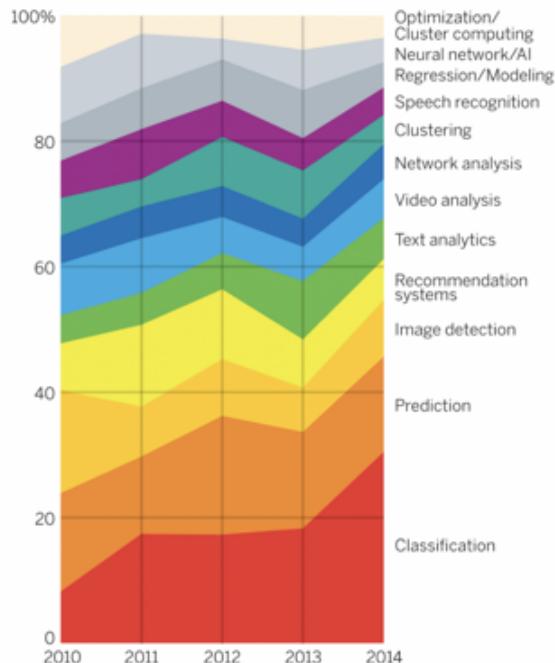
# Color

Find a logical way to reduce the amount of color in this stacked area graph.

## 12 COMMON MACHINE LEARNING TECHNIQUES

These approaches were identified through an analysis of more than 1,150 research papers over a four-year period.

PERCENTAGE OF TOTAL RESEARCH PAPERS



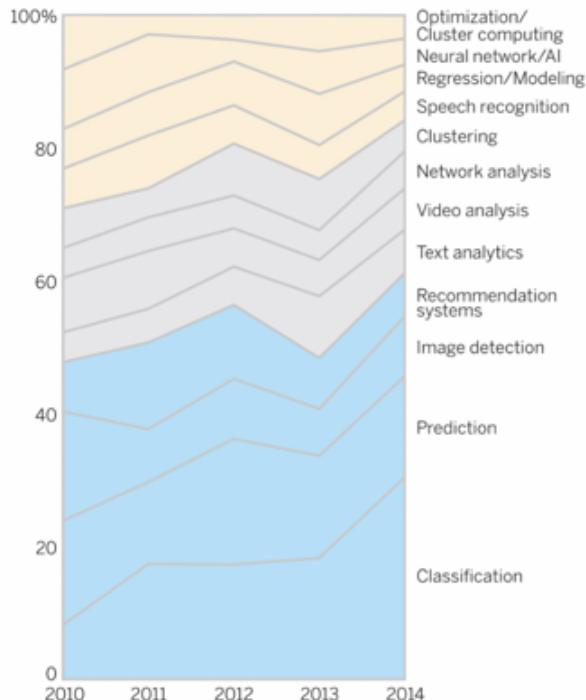
SOURCE: ACCENTURE INSTITUTE FOR HIGH PERFORMANCE, ANALYSIS  
OF A STANFORD MACHINE LEARNING RESEARCH PAPER DATABASE

# Color

## 12 COMMON MACHINE LEARNING TECHNIQUES

These approaches were identified through an analysis of more than 1,150 research papers over a four-year period.

PERCENTAGE OF TOTAL RESEARCH PAPERS



SOURCE: ACCENTURE INSTITUTE FOR HIGH PERFORMANCE ANALYSIS  
OF A STANFORD MACHINE LEARNING RESEARCH PAPER DATABASE

# Color

## HOW WE SPEND OUR TIME



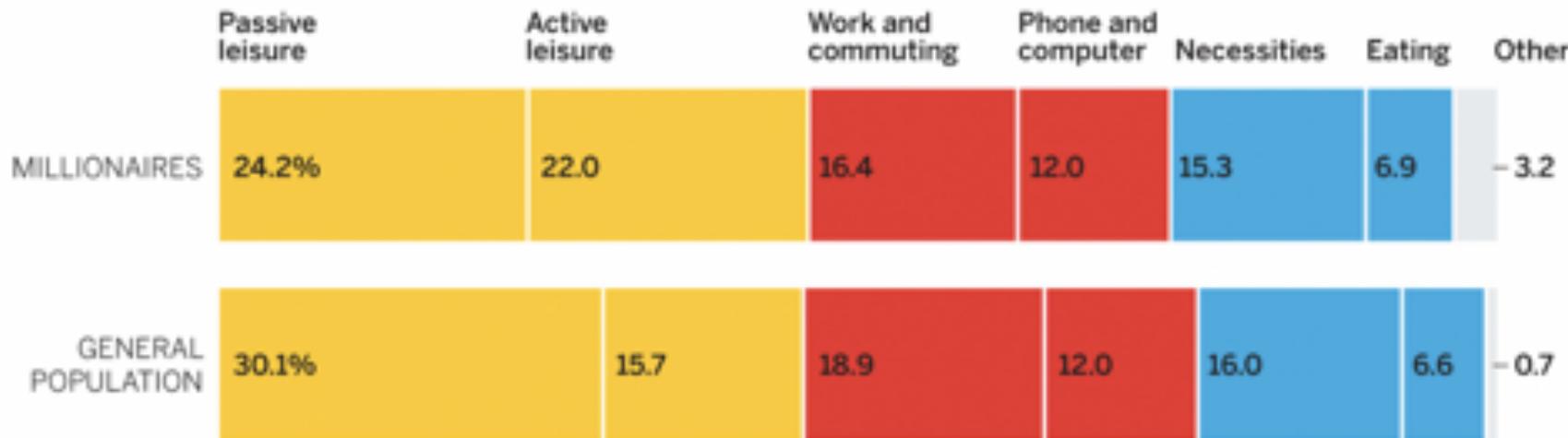
- 
1. Find up to three places you can remove color regardless of context.
  2. Find a way to group the variables using fewer colors but maintaining useful distinctions.
  3. You want to discuss leisure time with your audience. Create a color scheme for this context.
  4. Find a way to maintain seven colors without creating an overwhelming rainbow effect.
-

## Color

**HOW WE SPEND OUR TIME**

# Color

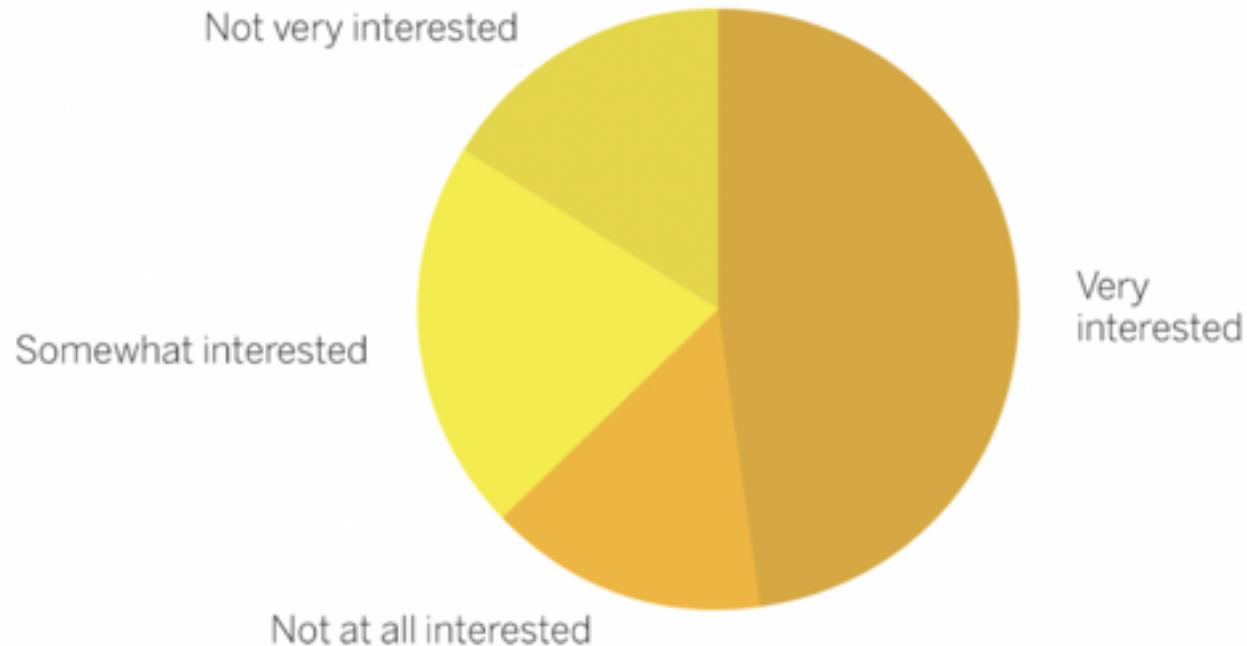
## HOW WE SPEND OUR TIME



## HOW WE SPEND OUR TIME



## Color

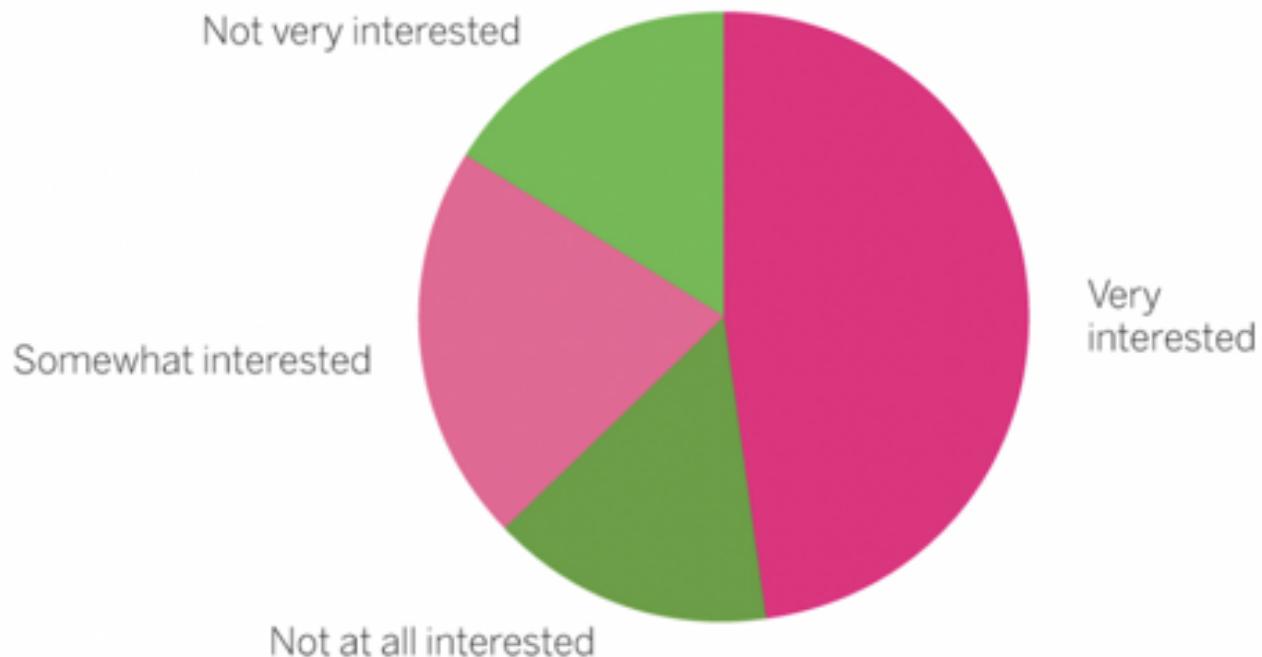
**HOW INTERESTED ARE YOU IN BUYING A DRONE?**

1. Identify two problems with the colors used in this pie chart.
2. Remake the chart with a new color scheme.
3. You want your audience to see the proportion of people who have a positive interest in drones. Create a color scheme for this new goal.
4. You want to discuss only the interest in drones of respondents under 30. Remake the chart with a new color scheme to focus on that group, using the following data:

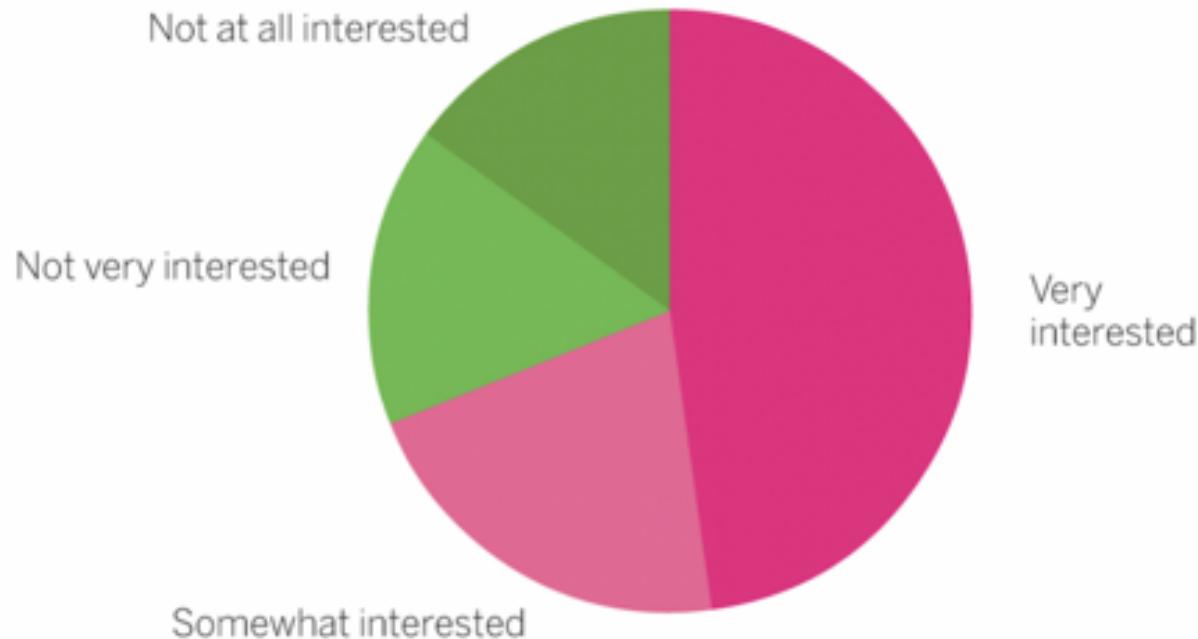
	UNDER 30	OVER 30
Very interested	34%	14%
Somewhat interested	14	7
Not very interested	5	11
Not at all interested	2	13

- 
1. *The colors are not grouped logically.* I count two groups of variables: interest and lack of interest. But here darker yellows are assigned to “very interested” and “not very interested” and brighter yellows to “not at all interested” and “somewhat interested.” Lightness and darkness aren’t used logically.
-

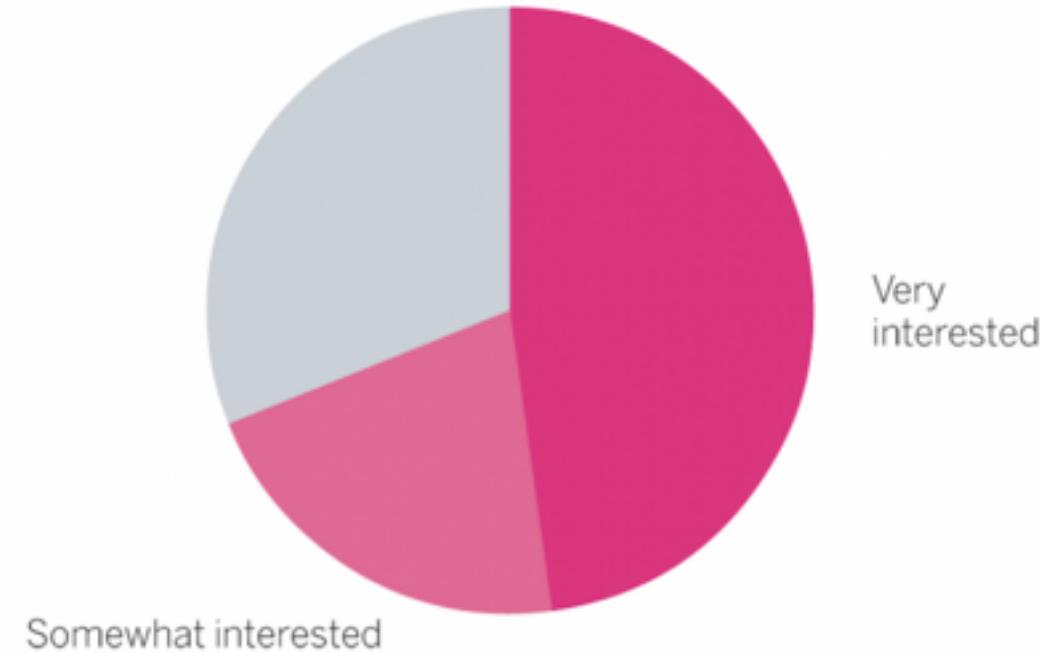
## HOW INTERESTED ARE YOU IN BUYING A DRONE?



## HOW INTERESTED ARE YOU IN BUYING A DRONE?

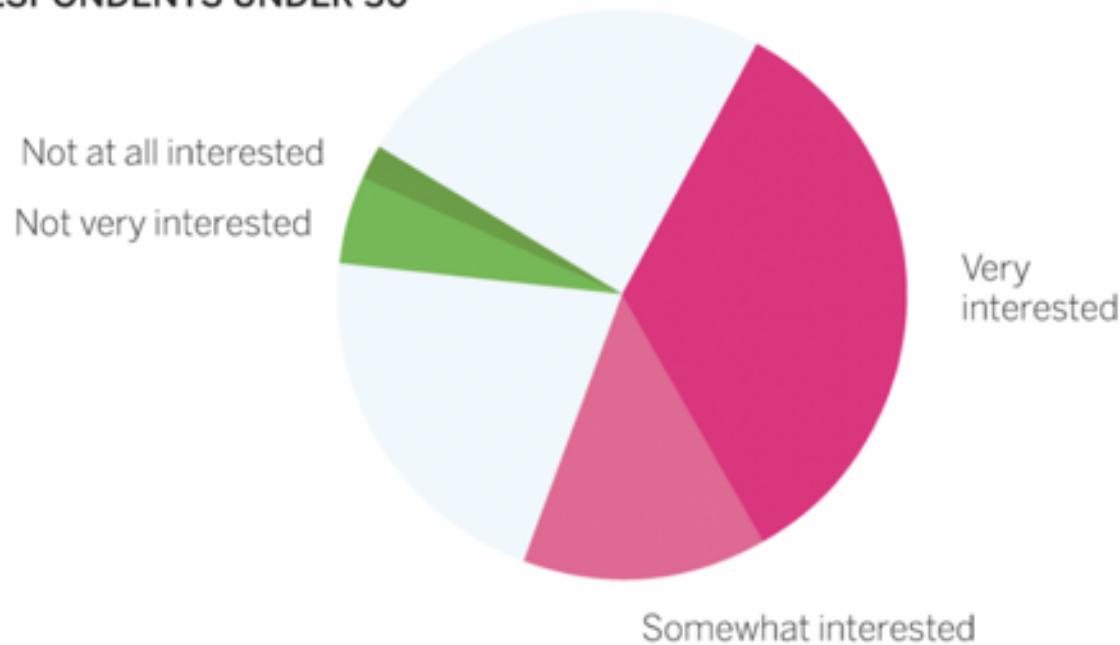


## HOW INTERESTED ARE YOU IN BUYING A DRONE?



## HOW INTERESTED ARE YOU IN BUYING A DRONE?

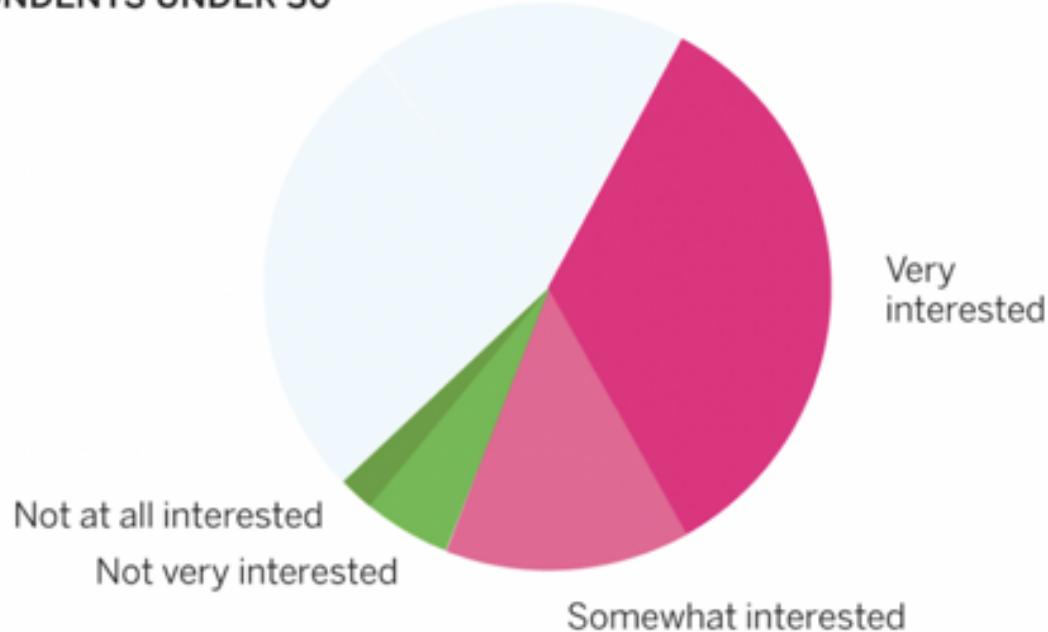
RESPONDENTS UNDER 30



## Color

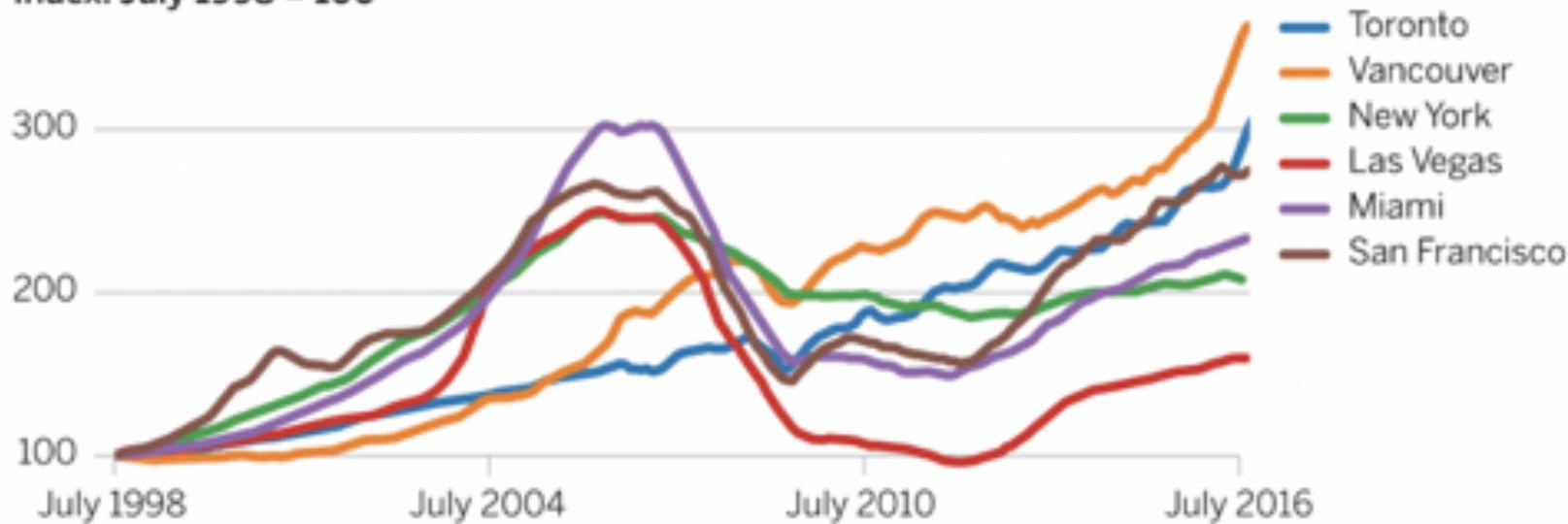
**HOW INTERESTED ARE YOU IN BUYING A DRONE?**

RESPONDENTS UNDER 30



## HOUSE PRICE INDEX FOR SELECTED CITIES

Index: July 1998 = 100

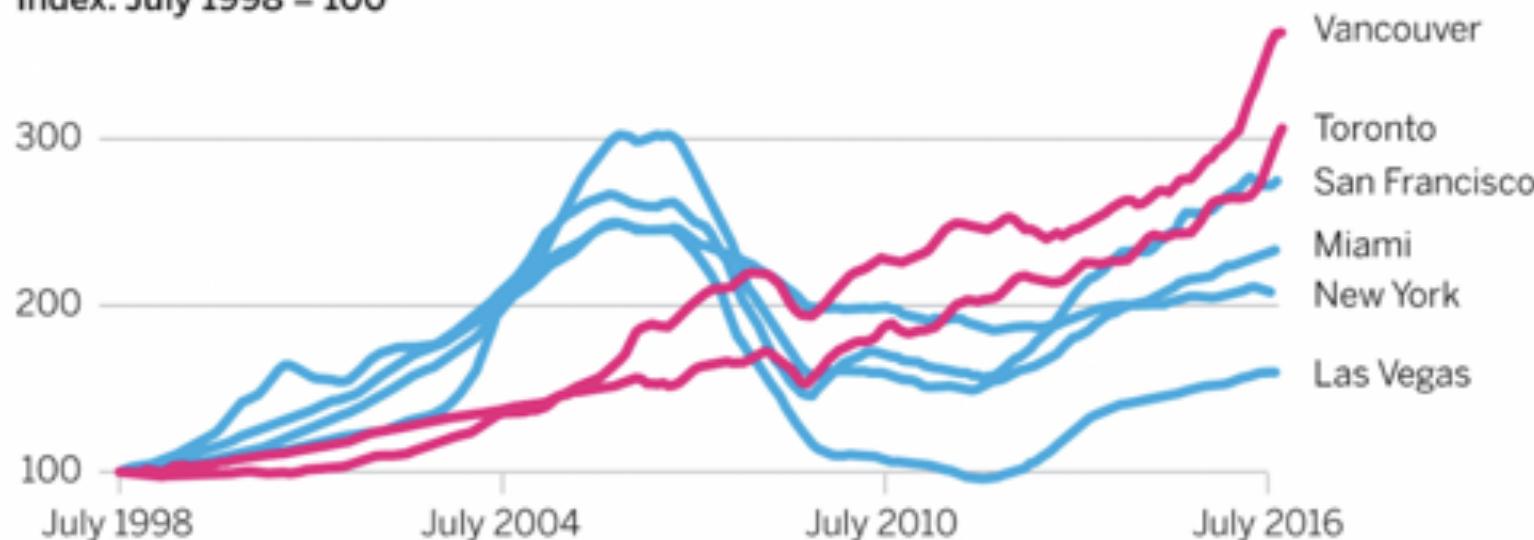


SOURCE: S&P/CASE SHILLER; TERANET HPI; MACLEAN'S

1. Create a logical grouping of variables and a color scheme for the grouping.
2. Create a color scheme that helps your audience focus on Canadian housing prices.
3. Create two more versions of this chart that use color to focus on trends in the data.

## CANADIAN HOUSE PRICES RISE PAST U.S. PRICES

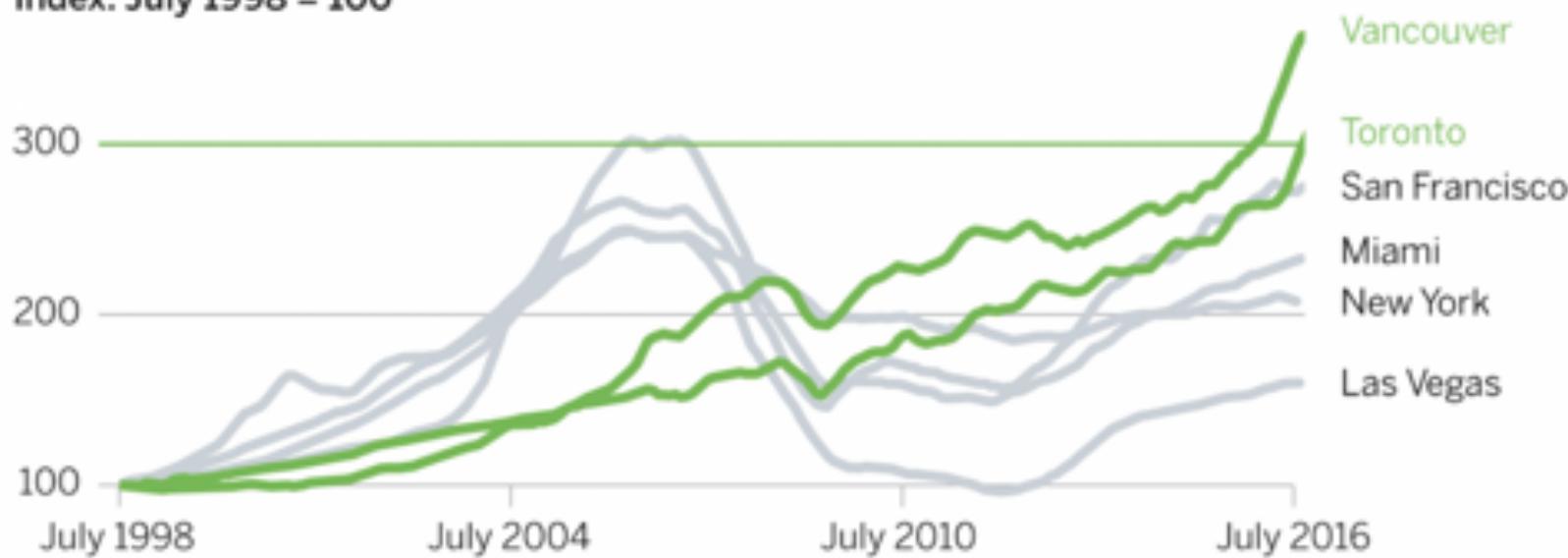
Index: July 1998 = 100



SOURCE: S&P/CASE SHILLER; TERANET HPI; MACLEAN'S

## CANADIAN HOUSE PRICES: STEADY GROWTH OR NEW BUBBLE?

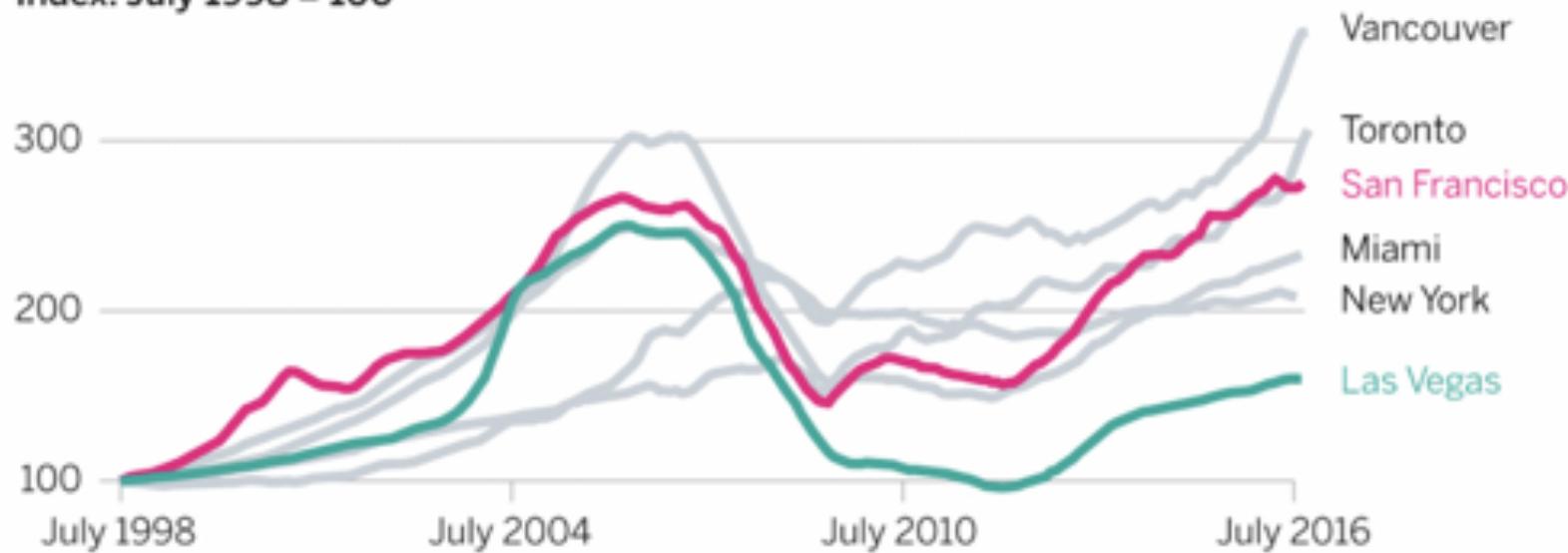
Index: July 1998 = 100



SOURCE: S&P/CASE SHILLER; TERANET HPI; MACLEAN'S

## SIMILAR BUBBLES, DIFFERENT RECOVERIES

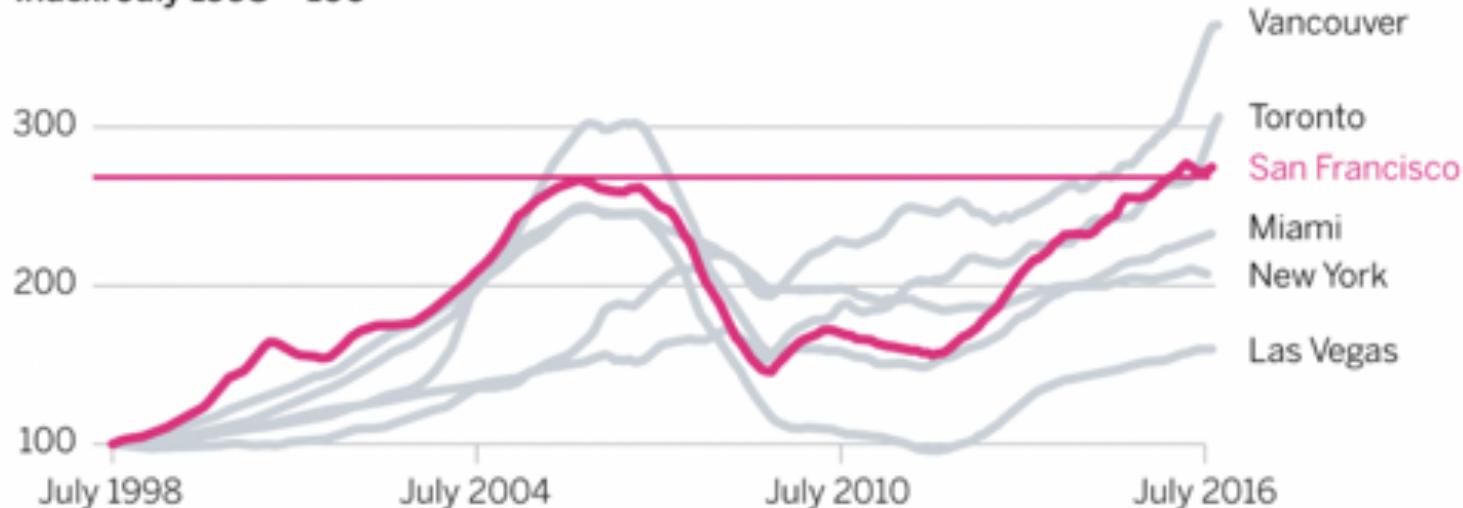
Index: July 1998 = 100



SOURCE: S&P/CASE SHILLER; TERANET HPI; MACLEAN'S

## SAN FRANCISCO BUBBLES UP AGAIN

Index: July 1998 = 100



SOURCE: S&P/CASE SHILLER; TERANET HPI; MACLEAN'S

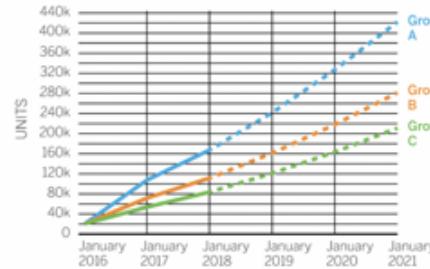
# Clarity

- Take stuff away
- Remove redundancy
- Limit color and eye travel
- Know how people think
- Describe ideas, not structure
- Align everything

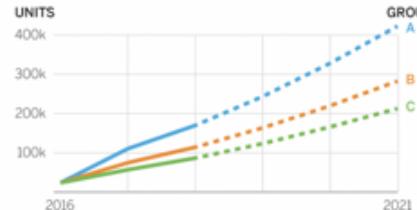
# Clarity

You want to use a line chart to show the year-to-year trend in three forecasts. Which grid creates the clearest experience for that context?

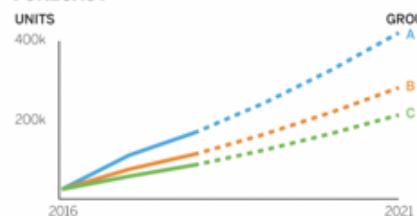
A FORECAST



B FORECAST



C FORECAST



# Clarity

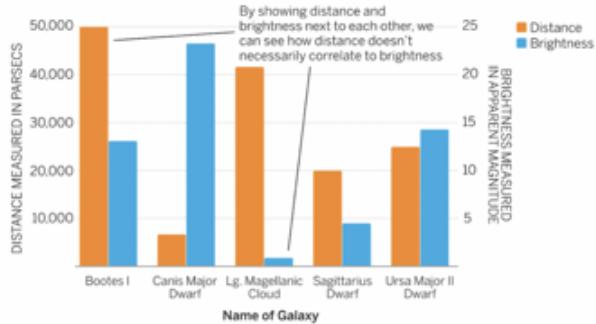
Answer: B. The key is the context of needing to talk about “year-to-year trends.” That disqualifies C, which—in an effort to be as simple as possible—has literally removed any sign of those trends. In a presentation in which the five-year trend matters, it might be a fine chart, but it’s not enough for our context. The first chart, A, also provides year-to-year trends, but it’s probably not simple enough. The number and heaviness of the grid lines on the y-axis make it hard to read. If those grid lines were lighter, as in B, one could argue that they’re useful. They give us more-exact values at each year’s intersecting grid line. But even if the lines weren’t so dominant, it would be hard to connect them to their value, because values are scrunched together. You could make the chart higher to create more space between values, but even then the visual busyness would fight the trend lines.

# Clarity

Find a common element in the following bar charts that makes both unclear.  
Find one element in each that makes it uniquely unclear.

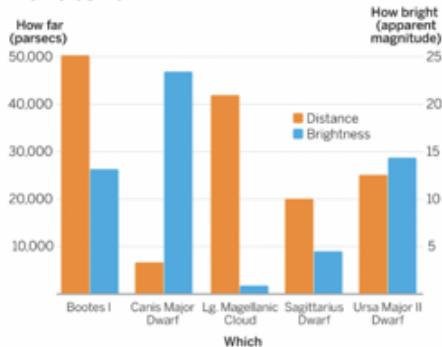
## MEASURING THE DISTANCE AND BRIGHTNESS OF NEARBY GALAXIES

Distance in parsecs is shown on the left side, and brightness in apparent magnitude is shown on the right.



## DISTANCE AND BRIGHTNESS

They vary by galaxy.



# Clarity

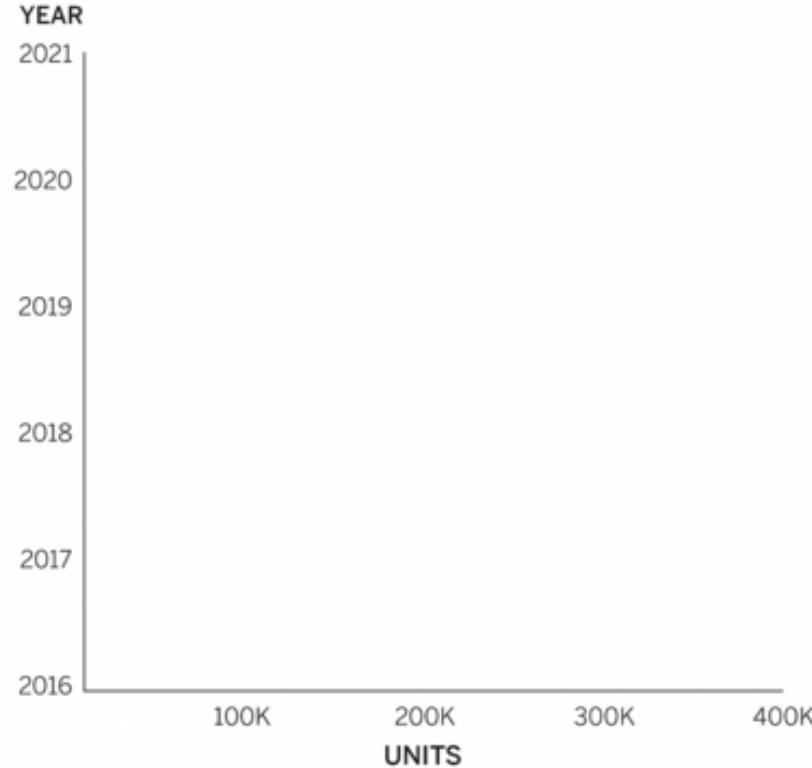
*Common element:* The categories on the x-axis are organized alphabetically. Although that's a valid way to organize information, it's not clarifying here—it just creates a haphazard collection of comparisons. And galaxies are not arranged alphabetically in the universe. We're comparing distances and brightness. Arranging the galaxies on one of those progressions—nearest to farthest, or brightest to dimmest—would provide an anchor point for seeing what one variable does to the other. To me, distance is the more accessible variable, so I would arrange the data to progress from near to far. Far to near would challenge a perceptual convention that we start on the left with the nearest thing. It would feel weird to start with the farthest away and progress to the nearest.

*Unique unclear element:* For chart 1, it's redundancy. The chart is laden with repetition. The title, the subtitle, both y-axes, the key, and the caption all effectively say the same thing. If you identified the pointers, that's also a good catch. It's confusing to have the pointers refer to bars from different galaxies, and the long lines create visual distraction, too.

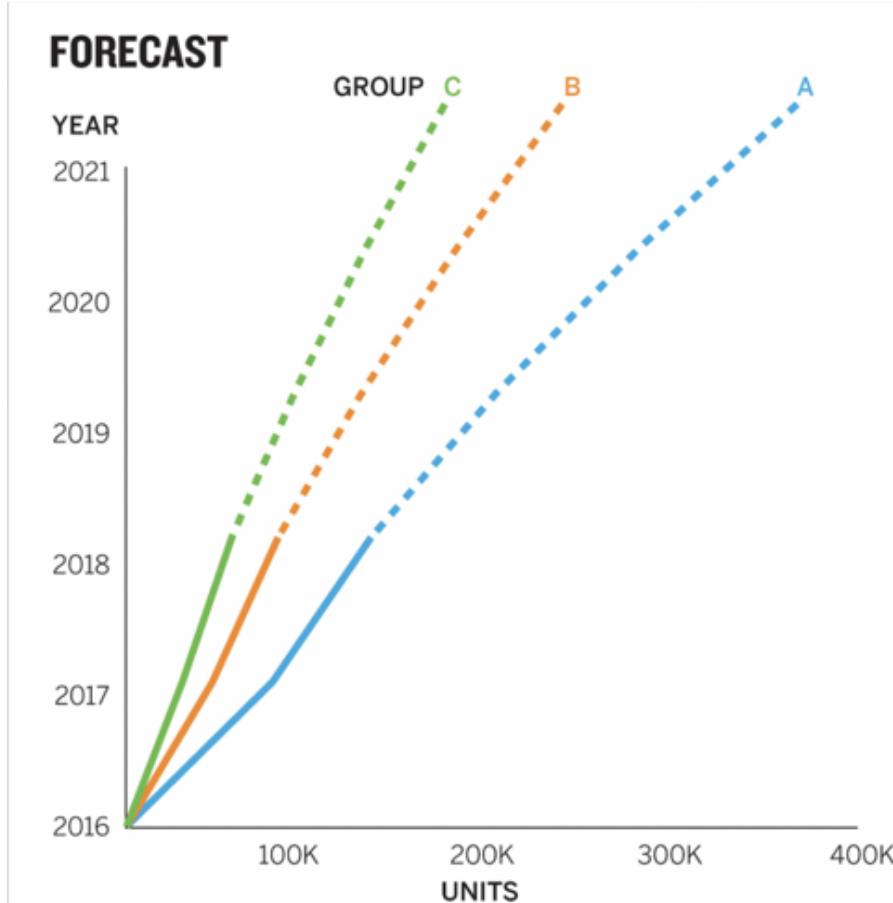
For chart 2, it's ambiguity. This chart fixes many of the previous version's problems, but it oversimplifies. The title is vague—distance and brightness of what?—and the subtitle is reflexively obvious. It doesn't help us understand why we're looking at this thing. The axis labels are nice and short, but what do they refer to? How far is what? How bright is what? The label "Which" on the x-axis is aggressively unclear. If you don't know that Bootes I et al. are galaxies, you're lost. But even if we imagine that we can figure out this presentation, the arrangement of the data will thwart us. I find myself asking, *What does this mean?* It looks more like a thoughtless display of data than a visual that conveys an idea. What's the point?

# Clarity

Will a data visualization plotted on these axes be clear? Why?



# Clarity



# Clarity

In a scatter plot you want to plot average health care spending against average life expectancy for several countries. The plot will show a positive correlation across all countries except one: the United States, which spends the most but achieves only a middle-to-low life expectancy. Which headline would you choose for clarity?

- A** Health Care Spending vs. Life Expectancy for Several Countries
- B** Investing in Health Care Works—Nearly Everywhere
- C** Countries That Spend More on Health Care Live Longer
- D** What Does Health Care Spending Tell Us About Life Expectancy?

# Clarity

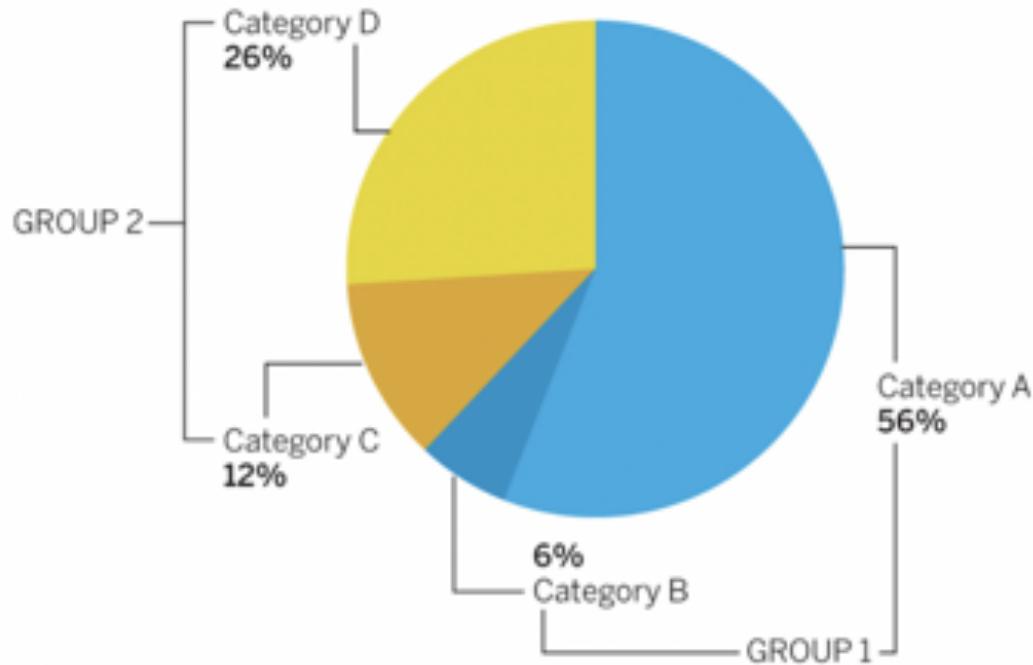
---

Answer: B. This title comes closest to conveying the idea we want people to see: The United States is an outlier. The others would not be wrong, but each suffers from some shortcoming. C comes close to revealing the idea, but it doesn't acknowledge the outlier—which is the focus not only for us but also for anyone who looks at the chart. Outliers draw the eye. This headline also contains a grammatical gaffe: countries don't live longer; people in countries do.

---

# Clarity

What makes this pie chart less clear than it could be? How would you fix it?



# Clarity

The lines make this chart much less clear than it could be. For one thing, they create too much eye travel. Connecting the visual element to its label requires some navigation, and it's not direct navigation. It's work to follow all the elbows in those pointers. Also, nothing is aligned and labels are placed at different and arbitrary distances from the chart. The Group labels are the most puzzling choice here: they create extra lines, they're far from the visual, and they're redundant. We've already created groups with color. There's no need to label them again.

# Clarity

You want to show customers' relative happiness or anger by region on a map.  
Which color scheme will be clearest?



# Clarity

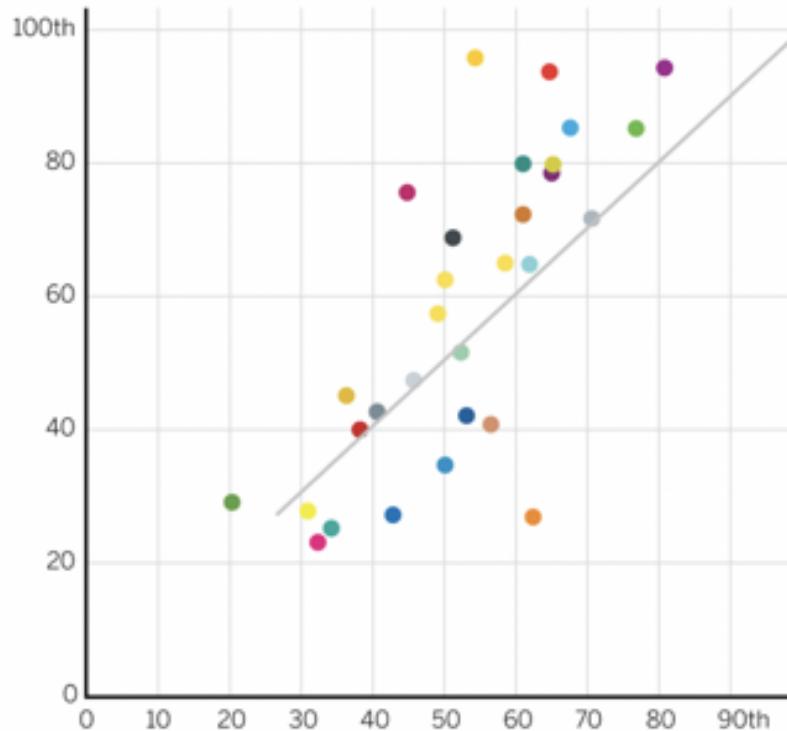
---

Answer: A. On this spectrum we want colors to fit convention. Because we associate red with heat and negativity, it makes sense to increase red's intensity as the anger level rises. This is especially true when the juxtaposed color is green, which is associated with safety and positivity. Those conventions make C a poor choice. Regions colored in deep red on a map would not immediately convey "happy customers." As for the gray scale of B, it's not bad, but it's linear. Instead of conveying a comparison of happy versus sad, it suggests a degree of one variable: anger.

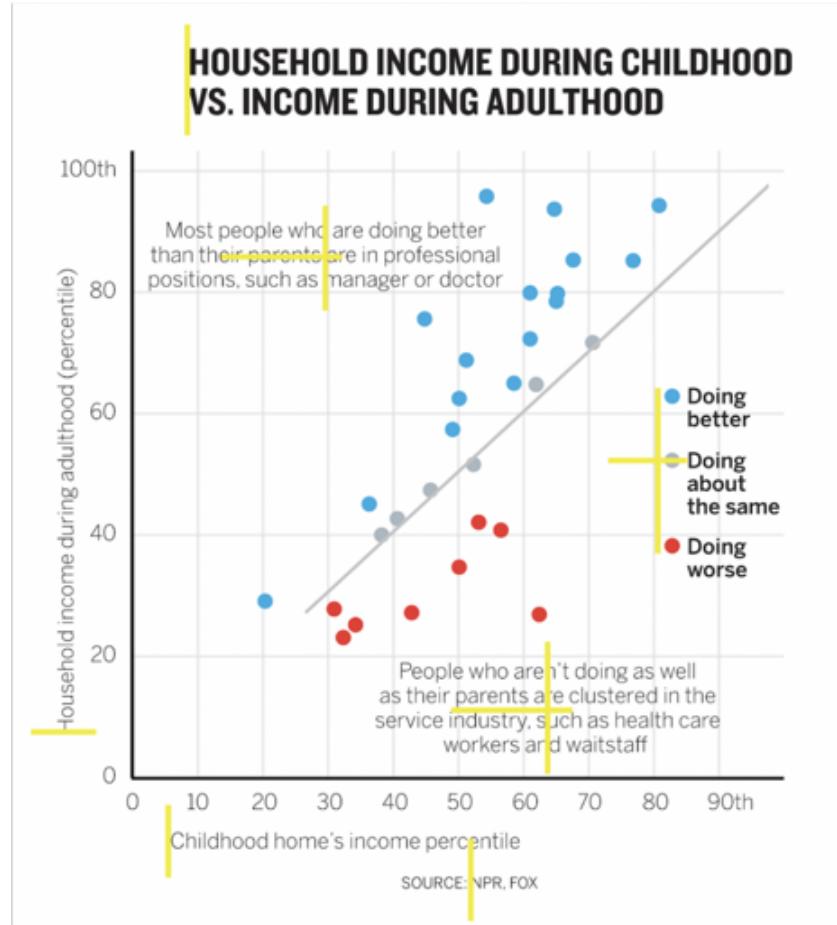
---

# Clarity

Without knowing what this scatter plot is about, what in the visual field would you attack first to improve its clarity?



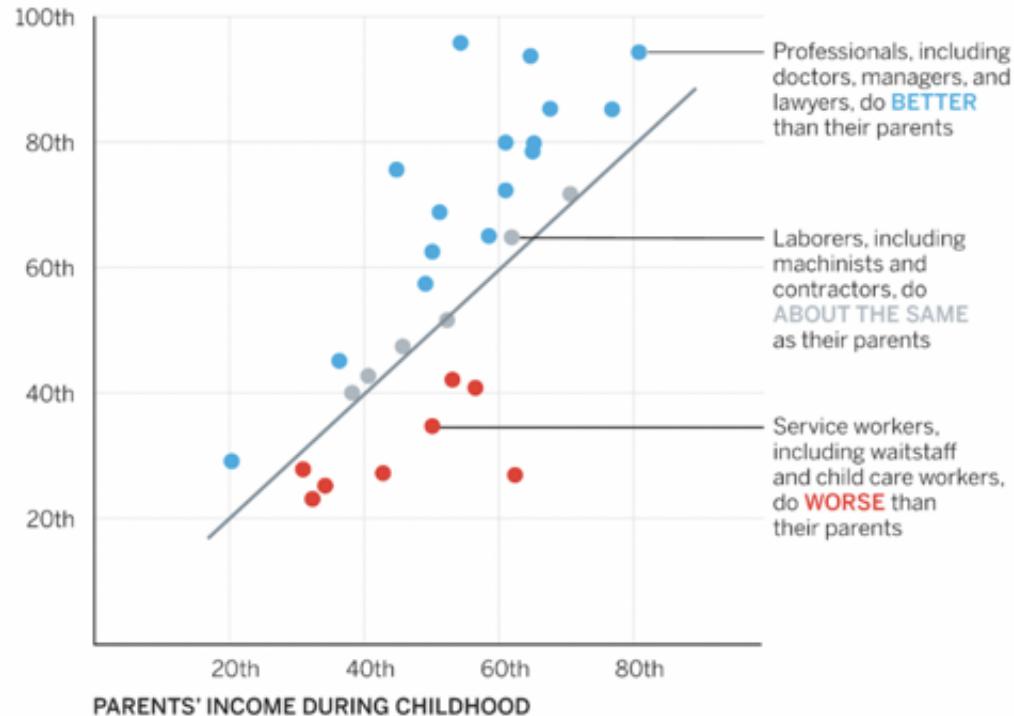
# Clarity



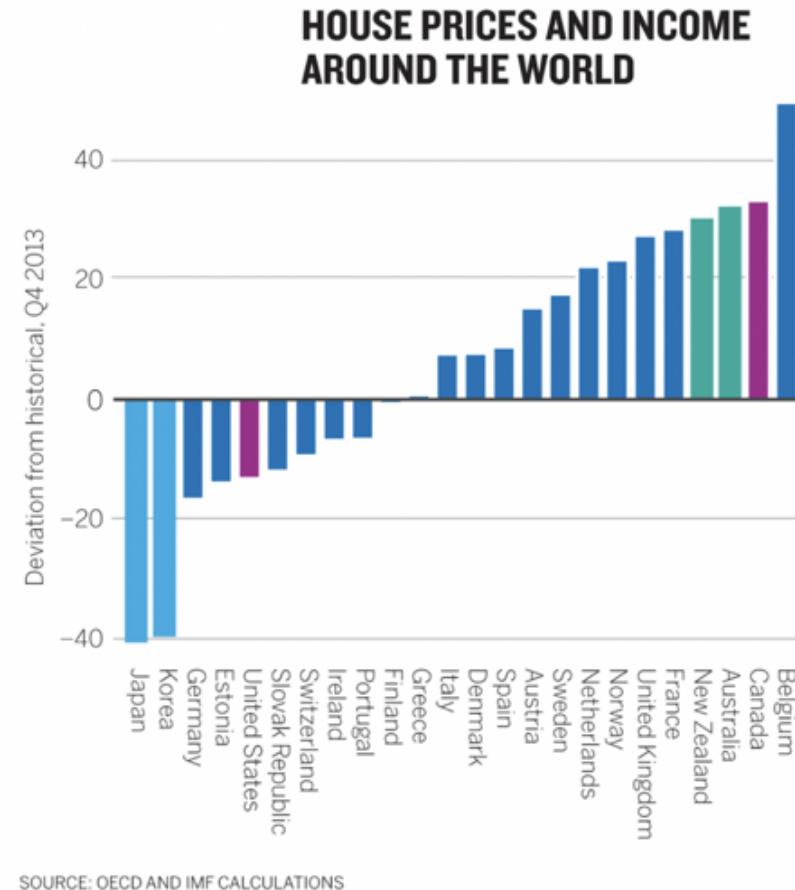
# Clarity

## WHO'S DOING BETTER THAN THEIR PARENTS? WHO'S DOING WORSE?

INCOME PERCENTILE IN ADULTHOOD



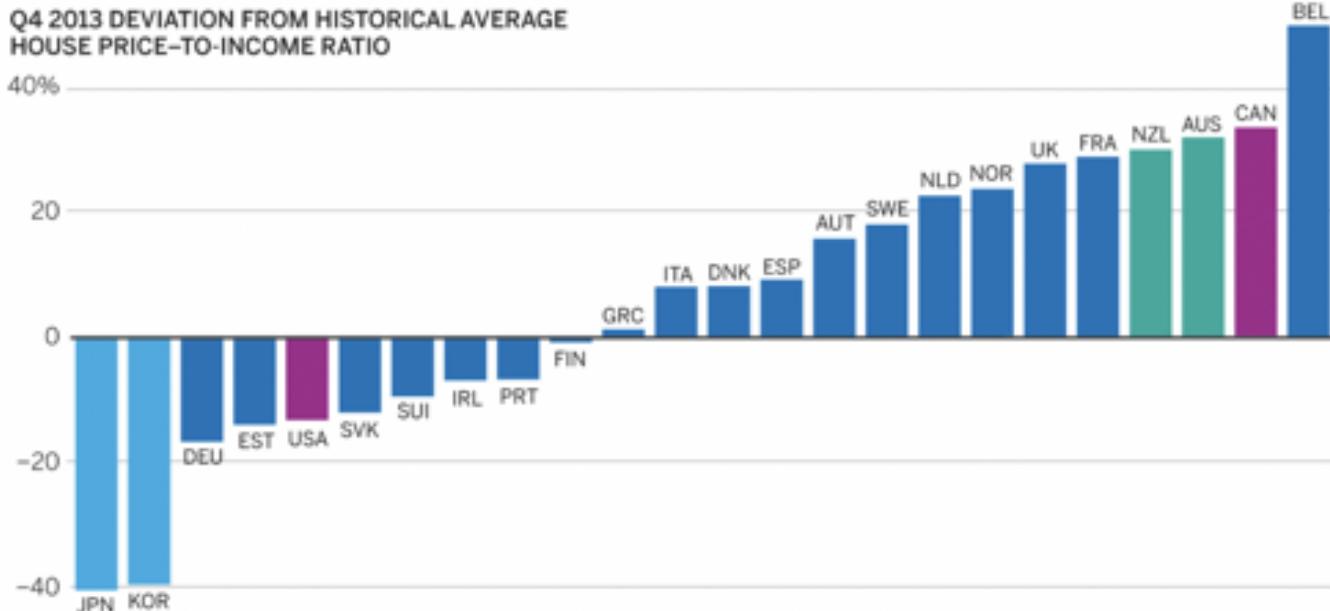
# Clarity



# Clarity

## THE HOUSE PRICE-TO-INCOME RATIO IS OUT OF WHACK, AGAIN

This key indicator dipped where the housing crisis hit hardest. In other markets, it shot up, signaling another possible bubble.



SOURCE: OECD AND IMF CALCULATIONS

# Clarity



# Clarity

1. Identify at least two instances of redundancy and say how you'd eliminate them.
2. Identify at least three more elements that make the chart less clear than it could be.
3. Sketch a clearer version that maintains focus on all three variables. Assume that the y-axis refers to dollars. Assume that the promo schedule is as follows:

Day 1: Original promo

Day 5: Follow-up email 1

Day 14: Follow-up email 2

Day 20: Last-chance email

4. Sketch a version of this chart that focuses clearly on a comparison of sneaker and jacket promotions.
5. Sketch a version that shows the “valuable” promotion period versus the “costly” promotion period (assuming that your analysis showed that promotional activities after 12 days were not cost-beneficial).

# Clarity

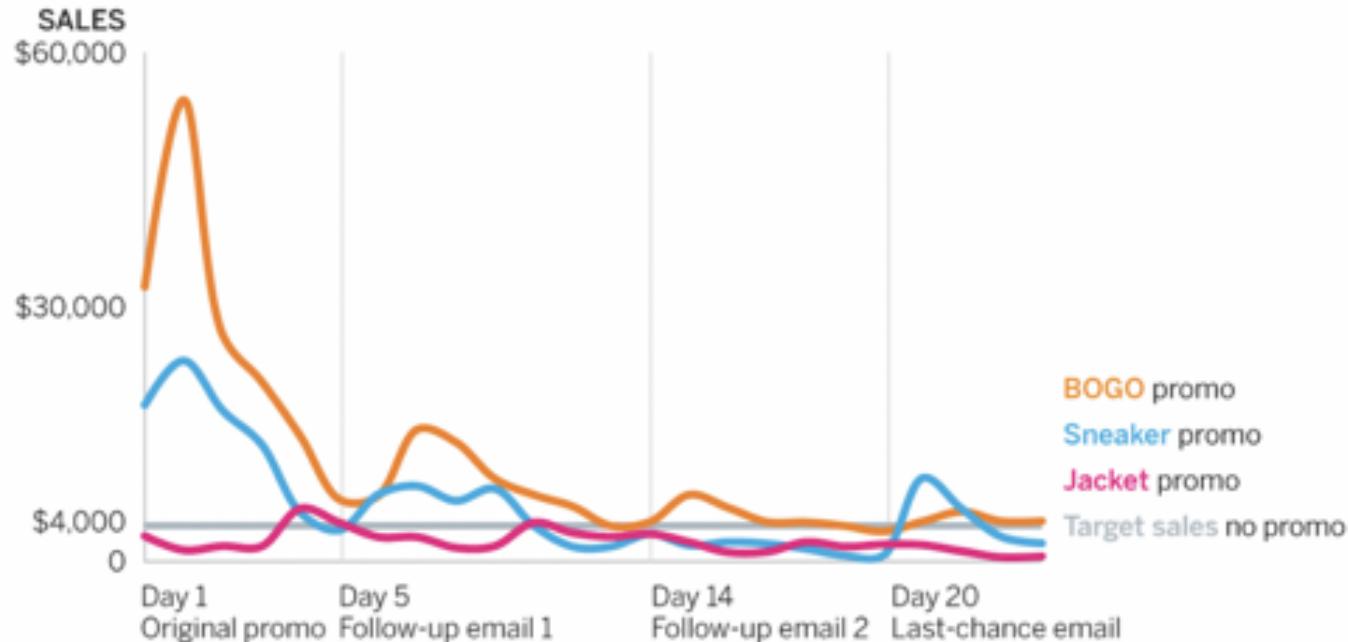
- 1.** *1. The x-axis labels.* Repeating “day” 24 times clutters up the page and makes it hard to see which day connects to which vertical grid line once the numbers are in double digits. Furthermore, the axis itself does the job of telling us what the values 1 through 24 are. You could also consider not putting every day on the x-axis. Ask yourself whether it’s important that the audience be able to access every day’s value. If it’s not—if, say, only the days that email promotions went out are important—you could eliminate most of these labels altogether.
- 2. Promotion “bumps” captions.* The words in the caption boxes describe what we see in the line chart and also point to what they describe, where there are other labels. So we’ve shown a bump, talked about a bump, pointed to the bump, and labeled the bump. The bump seems important, but so much information around it makes it harder to use. Should I look at the chart? Read the caption? Start with the labels? Which pointers should I start with? There’s no clear way through this chart because of the redundancy.
- 3. The headline.* This headline describes the structure of the chart: store promotions (points along the trend line) and sales (y-axis) over two-plus weeks (x-axis). Remember that users usually don’t read from the headline down—they start with the visual field and use the headline as a “confirming cue” to check whether what they think they see is in fact what they’re being shown. Use your headline to describe the main idea of a chart, not its structure.

# Clarity

- 2.* 1. *The y-axis.* It needs more information. Is this dollars? Units? What?
- 2. The grid lines.* Their heaviness and number distract from the trend lines. Grid lines are most valuable when you want the audience to be able to map values anywhere along the x-axis to the y-axis. In other words, if specific points on the trend line are important, grid lines will help. Ask how many x-axis points need to be accessed on the y-axis: One every week? Every day? (If individual points are more important than the general trend, perhaps a trend line isn't the right form. Maybe a dot plot or a table would serve you better.) If the trend line matters most, the grid lines can go, along with the visual noise they create.
- 3. Color.* The color scheme of the trend lines is perfectly fine, but the additional colors on the page create confusion. The thick teal of the target-sales line gives it too much prominence, and the dashes make an already busy grid busier. This is secondary information, for reference. A thin gray would work better. The dark-gray caption boxes draw our eyes, so we fight between looking at the chart and reading the captions. They also add to an overall heaviness.
- 4. Pointers.* Although they're effective in connecting captions to the bumps described, they create confusion by crossing over lines and using color. The various shapes and angles lessen clarity and create unnecessary eye travel.
- 5. Alignment.* There is none. Labels float; captions are offset; the key is on its own plane. So many unique entry points erode clarity.

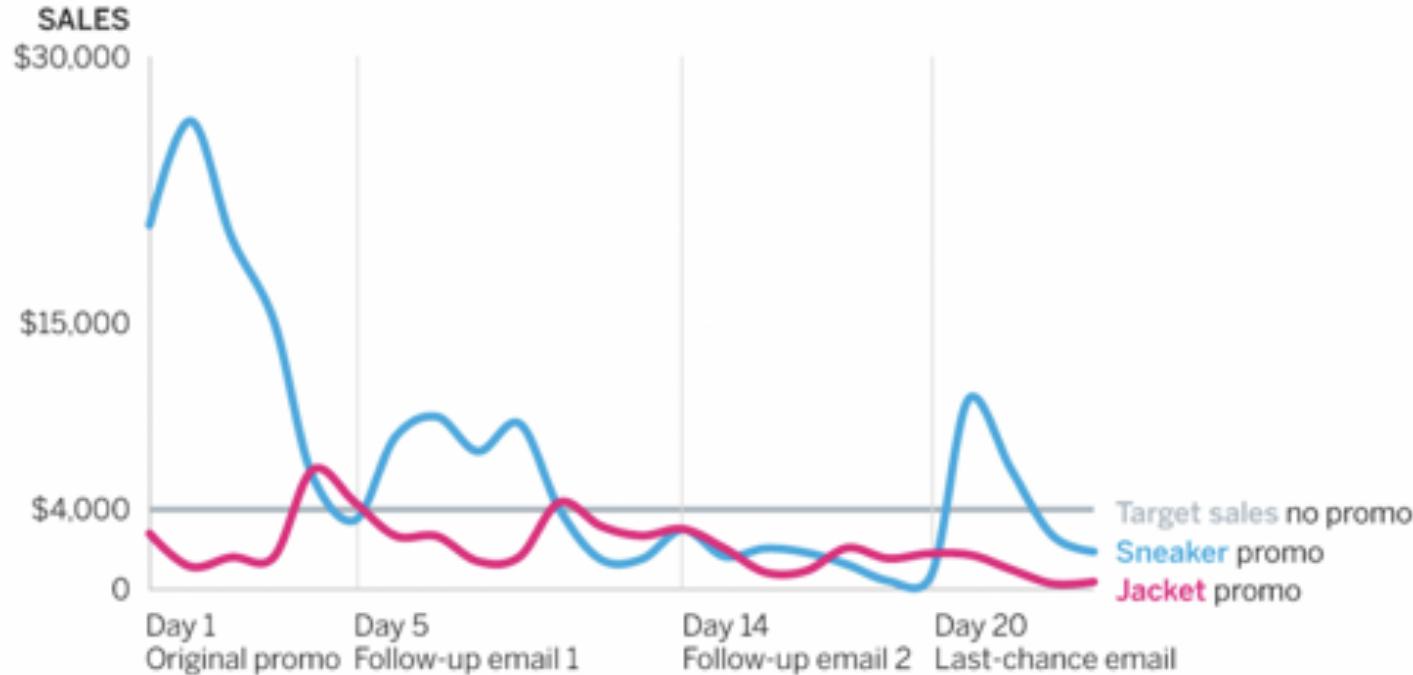
## THE LONG-TAIL LIFE OF STORE PROMOTIONS

Emails bumped sales, but less each time



# Clarity

## COMPARING PROMOS: SNEAKERS DID BETTER THAN JACKETS



# Clarity

## STORE PROMOTIONS ARE WORTH IT—FOR 12 DAYS



# Clarity

## IN-DEMAND SKILLS IN 15 BUSINESS FUNCTIONS

Some skills matter much more than others based on nearly 25 million job postings collected between September 2014 and August 2015.



SOURCE: "THE HUMAN FACTOR" BY BURNING GLASS TECHNOLOGIES, NOVEMBER 2015

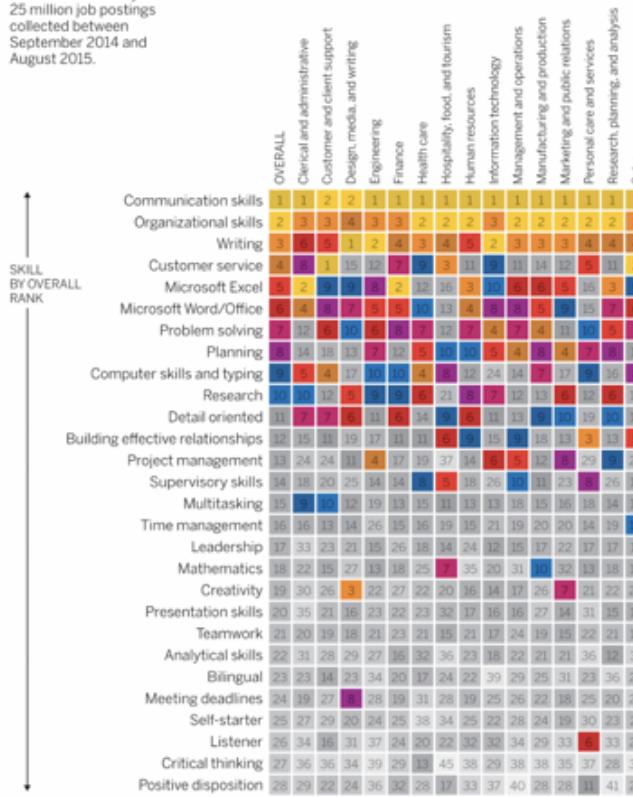
# Clarity

- 
1. Describe how this visualization's axes are organized and identify at least two alternative ways to organize them.
  2. Sketch a new version of it using a different organizing principle.
  3. Describe the general color scheme of this heat map.
  4. Sketch a new version using a different color scheme to improve clarity.
-

# Clarity

## THE MOST IN-DEMAND SKILLS IN 15 BUSINESS FUNCTIONS

Derived from nearly  
25 million job postings  
collected between  
September 2014 and  
August 2015.



SOURCE: "THE HUMAN FACTOR," BY BURNING GLASS TECHNOLOGIES, NOVEMBER 2015

# Clarity

## TOP SKILLS REQUIRED IN 15 BUSINESS FUNCTIONS

Not surprising: Communication is crucial. Surprising: Teamwork doesn't crack the top 20.

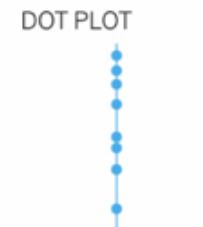
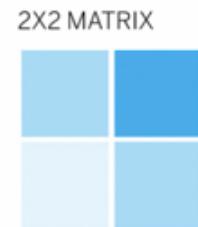
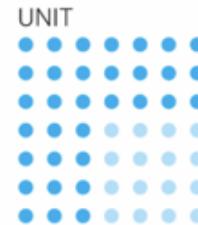
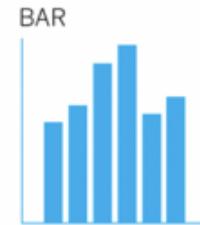
	OVERALL	Clerical and administrative	Customer and client support	Design, media, and writing	Engineering	Finance	Healthcare	Hospitality, food, and tourism	Human resources	Information technology	Management and operations	Manufacturing and production	Marketing and public relations	Personal care and services	Research, planning, and analysis	Sales	
Communication skills	1	1	2	2	1	1	1	1	1	1	1	1	1	1	1	1	
Organizational skills	2	2	3	4	3	3	3	3	3	3	3	3	3	3	3	3	
Writing	3	6	5	3	2	4	3	4	5	2	3	3	3	3	4	4	
Customer service	4	8	11	15	12	7	9	3	11	9	11	14	12	5	11	2	
Microsoft Excel	5	2	9	9	8	8	12	16	3	10	6	6	5	16	3	9	
Microsoft Word/Office	6	4	8	7	5	5	10	13	4	8	8	5	9	15	7	6	
Problem solving	7	12	6	10	6	8	7	12	7	4	7	4	11	10	5	7	
Planning	8	14	18	13	7	12	5	10	10	5	4	8	4	7	8	11	
Computer skills and typing	9	5	4	17	10	10	4	8	12	24	14	7	17	9	16	8	
Research	10	10	12	5	9	9	6	21	8	7	12	13	6	12	6	18	
Detail oriented	11	7	7	6	11	6	14	9	6	11	13	9	10	19	10	12	
Building effective relationships	12	15	11	19	17	11	11	6	9	15	9	18	13	3	13	5	
Project management	13	24	24	11	4	17	19	37	14	6	5	12	8	29	9	22	
Supervisory skills	14	18	20	25	14	14	8	5	18	26	10	11	23	8	26	14	
Multitasking	15	9	10	12	19	13	15	11	13	13	18	15	16	18	14	13	
Time management	16	16	13	14	26	15	16	19	15	21	19	20	20	14	19	10	
Leadership	17	33	23	21	15	26	18	14	24	12	15	17	22	17	17	17	
Mathematics	18	22	15	27	13	18	25	7	35	20	31	10	32	13	18	16	
Creativity	19	30	26	3	22	27	22	20	16	14	17	26	7	21	22	20	
Presentation skills	20	35	21	16	23	22	23	32	17	16	16	27	14	31	15	15	
Teamwork	21	20	19	18	21	23	21	15	21	17	24	19	15	22	21	19	
Analytical skills	22	31	28	29	27	16	32	36	23	18	22	21	21	36	12	31	
Bilingual	23	23	14	23	34	20	17	24	22	39	29	25	31	23	36	24	
Meeting deadlines	24	19	27	8	28	19	31	28	19	25	26	22	18	25	20	29	
Self-starter	25	27	29	20	24	25	38	34	25	22	28	24	19	30	23	21	
Listener	26	34	16	31	37	24	20	22	32	32	34	29	33	6	33	23	
Critical thinking	27	36	36	34	39	29	13	46	38	29	38	36	35	37	28	39	
Positive disposition	28	28	29	22	24	26	32	28	17	39	37	30	28	28	11	21	28

SOURCE: "THE HUMAN FACTOR," BY BURNING GLASS TECHNOLOGIES, NOVEMBER 2015.  
RANKINGS BASED ON 25 MILLION JOB POSTINGS COLLECTED BETWEEN SEPTEMBER 2014 AND AUGUST 2015.

# Chart Types

- Know the basic categories
- Listen to how you describe things
- Rely on basic charts
- Don't overlook tables
- Consider using one axis

# Chart Types

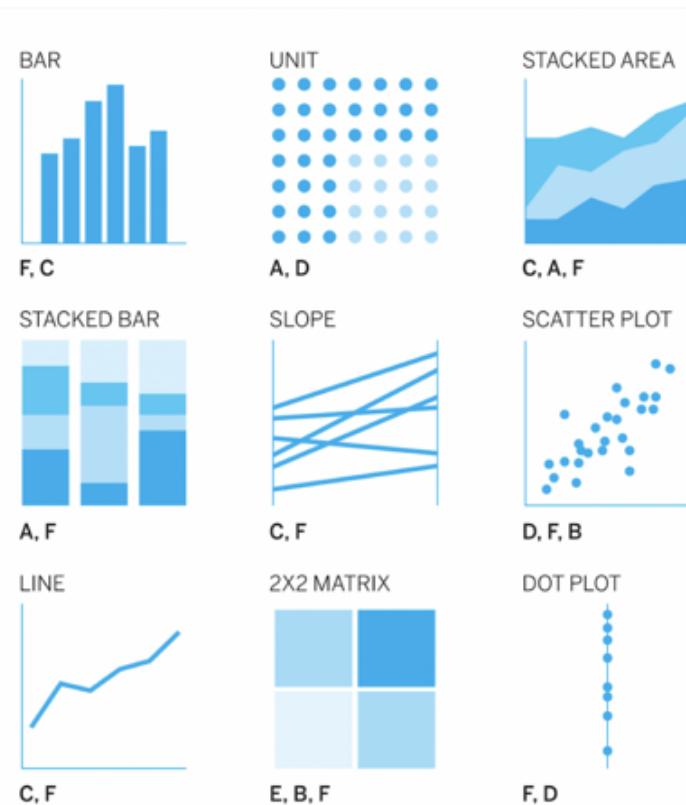


A. Show proportions  
D. Show distributions

B. Map something  
E. Explain concepts

C. Show trends  
F. Make comparisons

# Chart Types



A. Show proportions  
D. Show distributions

B. Map something  
E. Explain concepts

C. Show trends  
F. Make comparisons

# Chart Types

---

A simple data set shows the average number of hours spent in meetings per employee last year and this year at headquarters and in two satellite offices. How might you display this?

---

# Chart Types

	LAST YEAR	THIS YEAR
HQ	510	570
Satellite Office A	325	295
Satellite Office B	300	210

# Chart Types

---

You want to categorize football players on two dimensions: speed and strength. Each player gets a score for each dimension. What would be a good chart type to map how the players compare with one another?

---

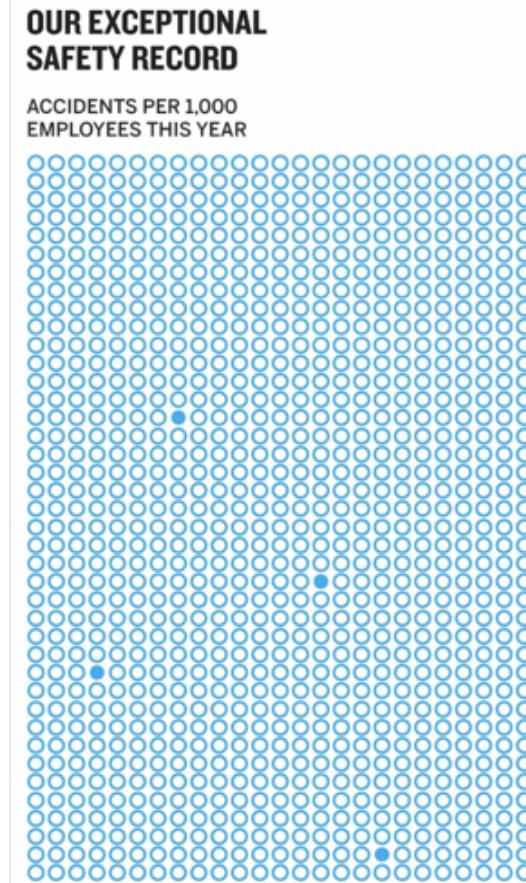
# Chart Types



# Chart Types

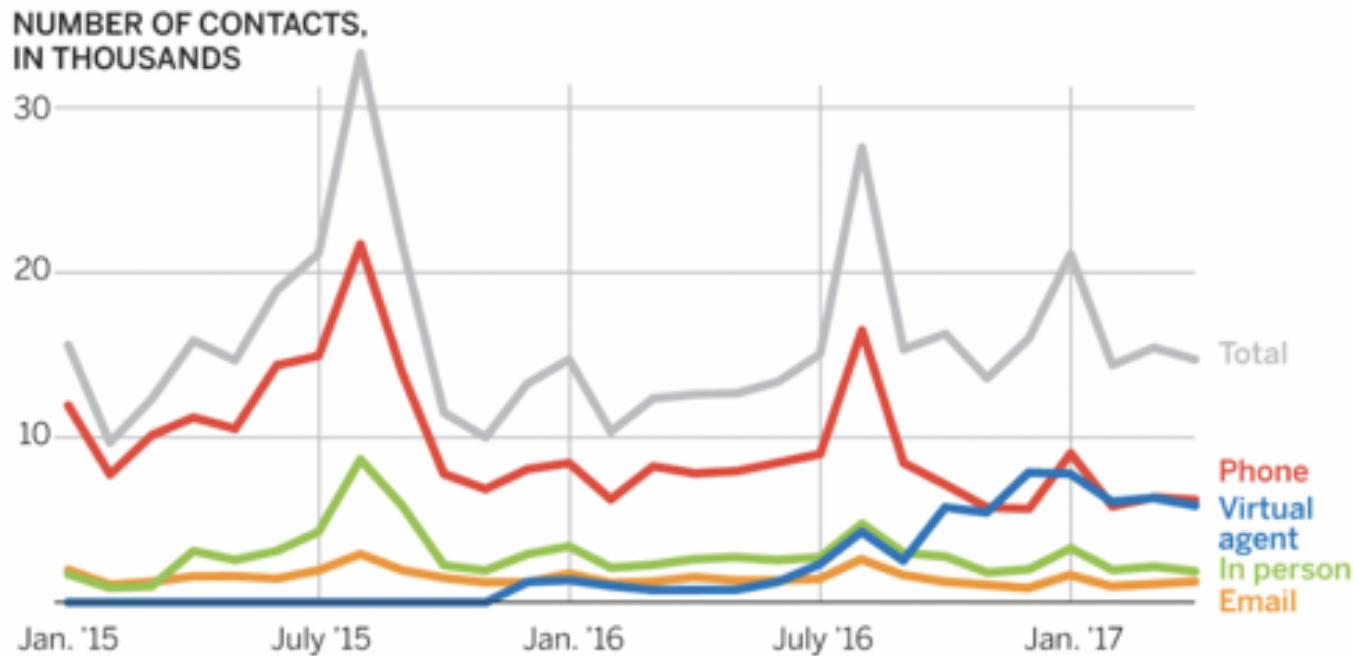
You want to convey how rare on-the-job accidents are in your factories—only 4 employees out of 1,000 have been injured in the past year. What type of visualization might powerfully convey this?

# Chart Types



# Chart Types

## HOW STUDENTS CONTACT THE REGISTRAR'S OFFICE

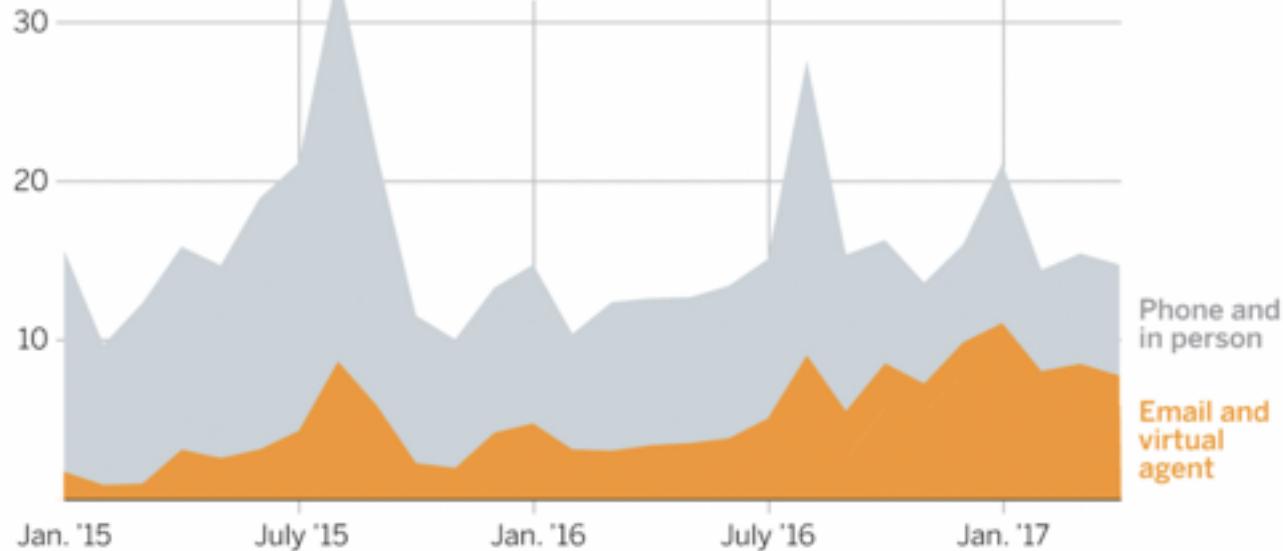


SOURCE: MARK MCCONAHAY

# Chart Types

## STUDENTS MOVE TO EMAIL AND VIRTUAL AGENTS

NUMBER OF CONTACTS  
IN THOUSANDS



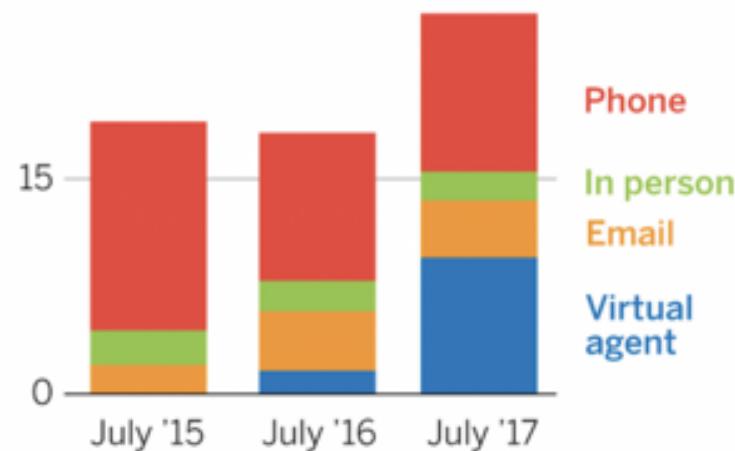
SOURCE: MARK MCCONAHEY

# Chart Types

## HOW STUDENTS CONTACT THE REGISTRAR'S OFFICE

NUMBER OF CONTACTS  
IN THOUSANDS

30



SOURCE: MARK MCCONAHEY

# Chart Types

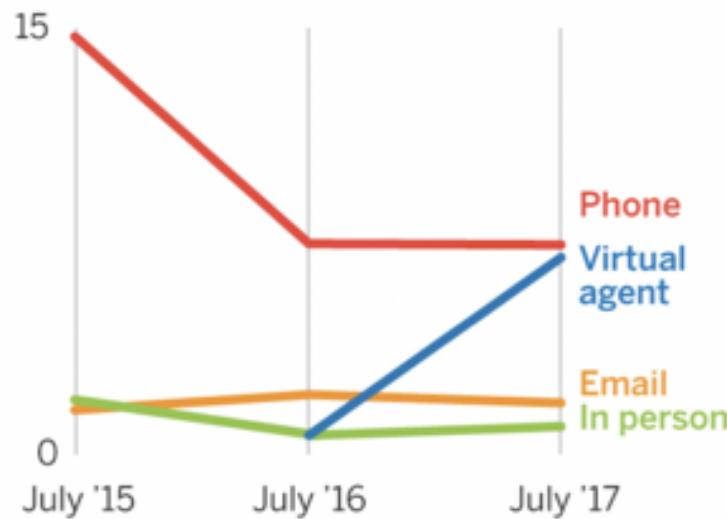


SOURCE: MARK MCCONAHEY

# Chart Types

## HOW STUDENTS CONTACT THE REGISTRAR'S OFFICE

NUMBER OF CONTACTS  
IN THOUSANDS



SOURCE: MARK MCCONAHAY

# Chart Types

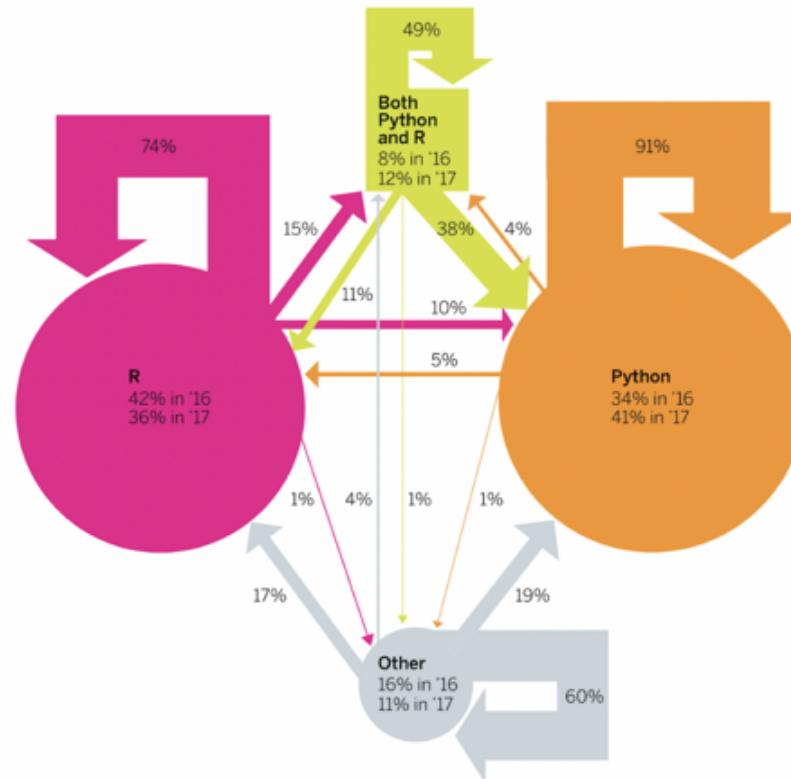


SOURCE: MARK MCCONAHAY

# Chart Types

## ANALYTICS, DATA SCIENCE, AND MACHINE LEARNING PLATFORMS

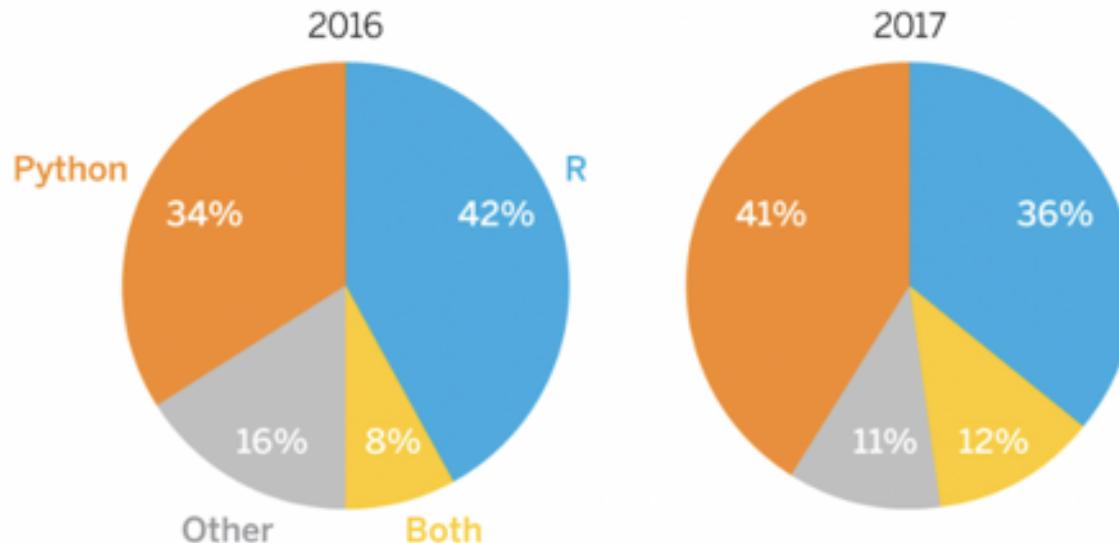
Change from 2016 to 2017



# Chart Types

## PYTHON GAINS, R WANES

Data scientists are flocking to Python



SOURCE: DATA FROM KDNUGETS.COM

# Chart Types

## PYTHON GAINS, R WANES

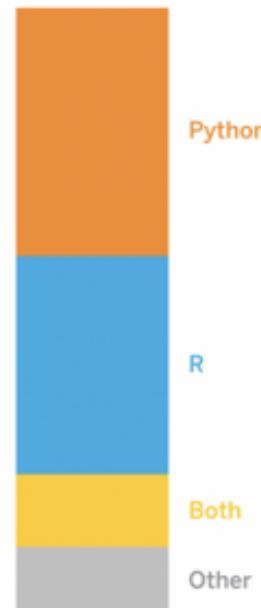
Data scientists are flocking to Python

### USAGE SHARE

2016



2017



## PYTHON GAINS, R WANES

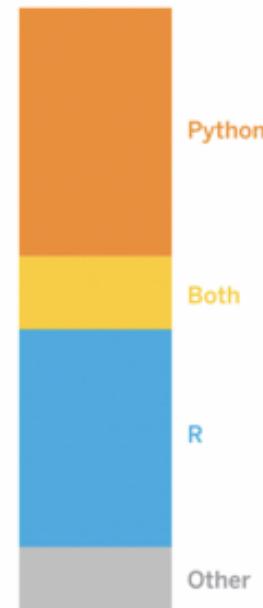
Data scientists are flocking to Python

### USAGE SHARE

2016



2017

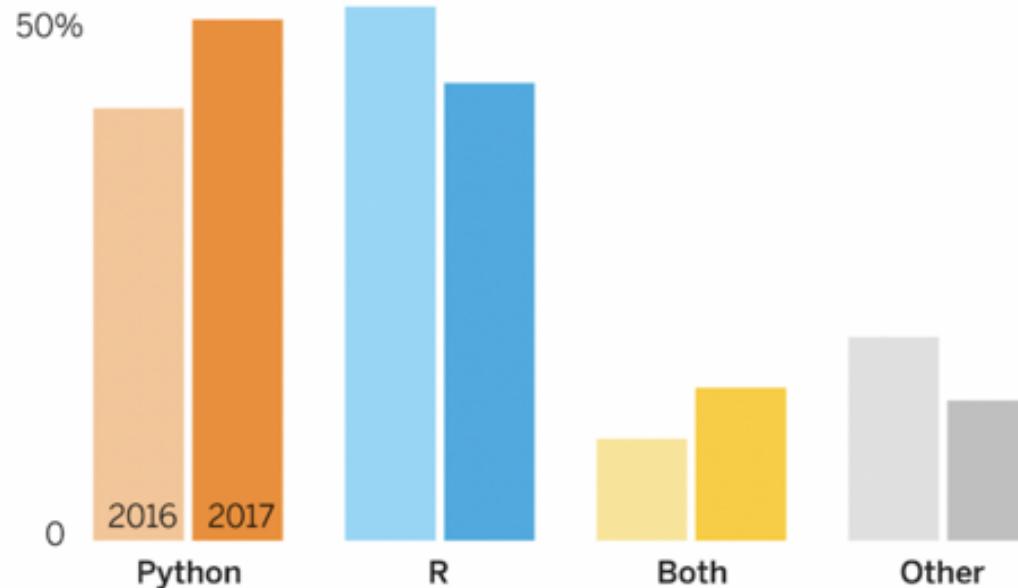


# Chart Types

## PYTHON GAINS, R WANES

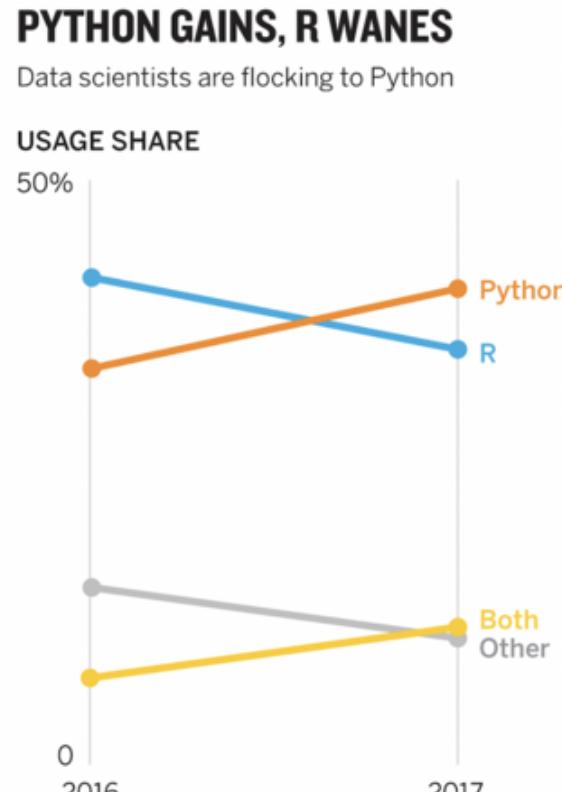
Data scientists are flocking to Python

### USAGE SHARE



SOURCE: DATA FROM KDNUGETS.COM

# Chart Types



SOURCE: DATA FROM KDNUGETS.COM

# Chart Types

## THE CHANGING LANDSCAPE OF ANALYTICS PLATFORMS

Data scientists are flocking to Python

### % SHARE

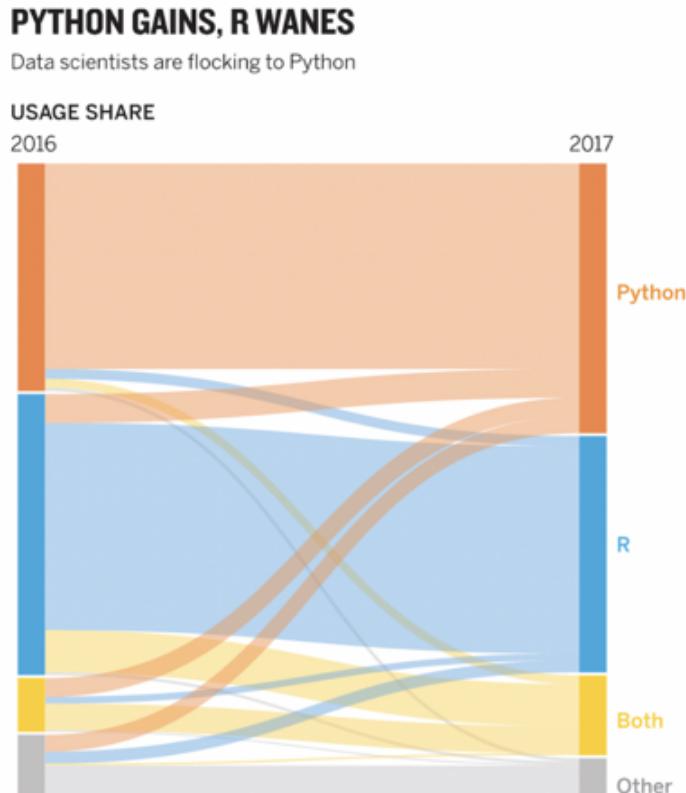
	2016	2017
Python	34	41
R	42	36
Both	8	12
Other	16	11

### % TRANSFER OF SHARE

	To Python	From Python	From R	From Both	From Other
To Python		10		38	19
To R	5			11	17
To Both	4		15		4
To Other	1		1	1	

SOURCE: DATA FROM KDNUGETS.COM

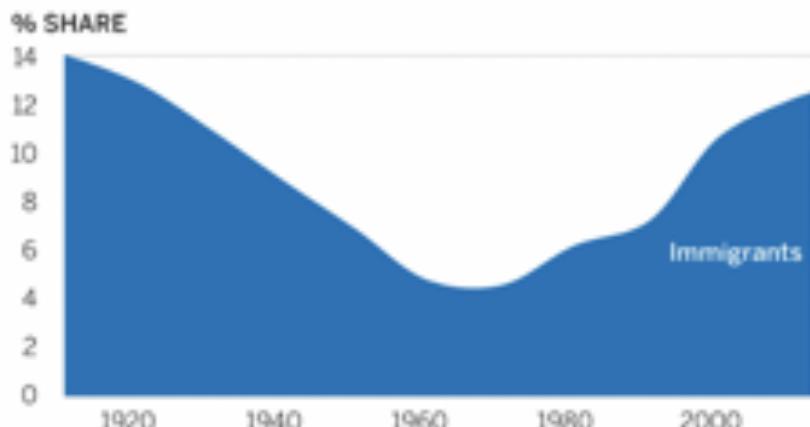
# Chart Types



- Shift your context question
- Emphasize and isolate
- Consider reference points
- Point things out
- Use narrative structure

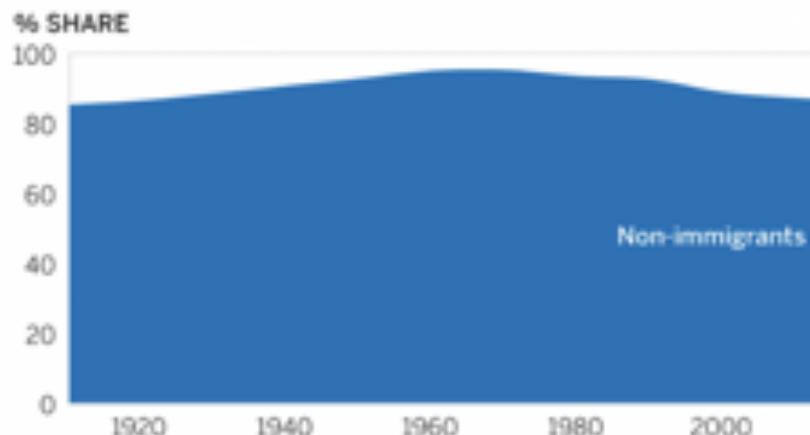
# Persuasion

IMMIGRANT SHARE OF THE POPULATION APPROACHES 1910 LEVELS



SOURCE: PEW RESEARCH

NON-IMMIGRANT SHARE OF THE POPULATION, 1910–2013



# Persuasion

---

Which of the following statements is a good contextual starting point for developing a persuasive chart?

- A** "I need to show them our declining overseas revenue."
- B** "I want them to see the two-year revenue trend in every market so that they can see that revenues are declining more in most of the 13 overseas markets than they are in our home markets—and by percentages higher than seasonal or historical averages."
- C** "I need to prove to them that the revenue trend overseas is troubling. Especially in the past six months."

# Persuasion

This chart convincingly shows a precipitous drop in employee vacation days taken. Identify two ways in which it may be unfairly manipulative.

## THE ABRUPT FALL OF THE AMERICAN VACATION



SOURCE: OXFORD ECONOMICS ANALYSIS, BASED ON A GFK PUBLIC AFFAIRS AND COMMUNICATION SURVEY AND DATA BY U.S. BUREAU OF LABOR STATISTICS, ON BEHALF OF "PROJECT: TIME OFF" (2014)

# Persuasion

## THE ABRUPT FALL OF THE AMERICAN VACATION

AVERAGE ANNUAL  
VACATION DAYS USED

22

20

18

16

1978

1996

2014

% FULL-WEEK  
VACATIONS

2.8

2.6

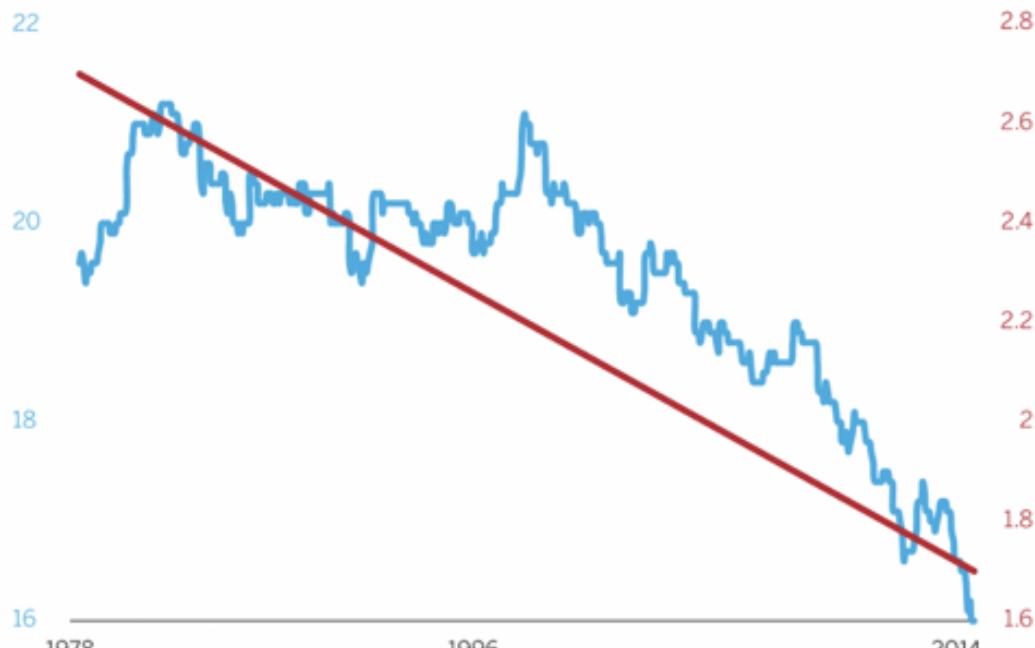
2.4

2.2

2

1.8

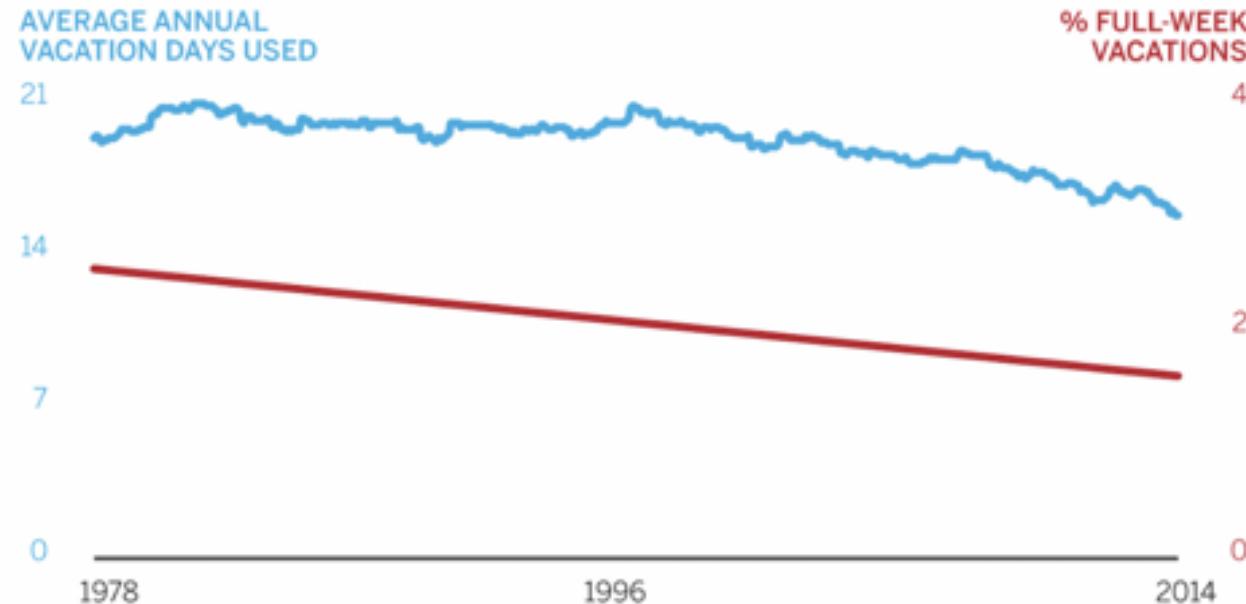
1.6



SOURCE: OXFORD ECONOMICS ANALYSIS, BASED ON A GFK PUBLIC AFFAIRS AND COMMUNICATION SURVEY AND DATA BY U.S. BUREAU OF LABOR STATISTICS, ON BEHALF OF "PROJECT: TIME OFF" (2014)

# Persuasion

## THE ABRUPT FALL OF THE AMERICAN VACATION



SOURCE: OXFORD ECONOMICS ANALYSIS, BASED ON A GFK PUBLIC AFFAIRS AND COMMUNICATION SURVEY AND DATA BY U.S. BUREAU OF LABOR STATISTICS, ON BEHALF OF "PROJECT: TIME OFF" (2014)

# Persuasion

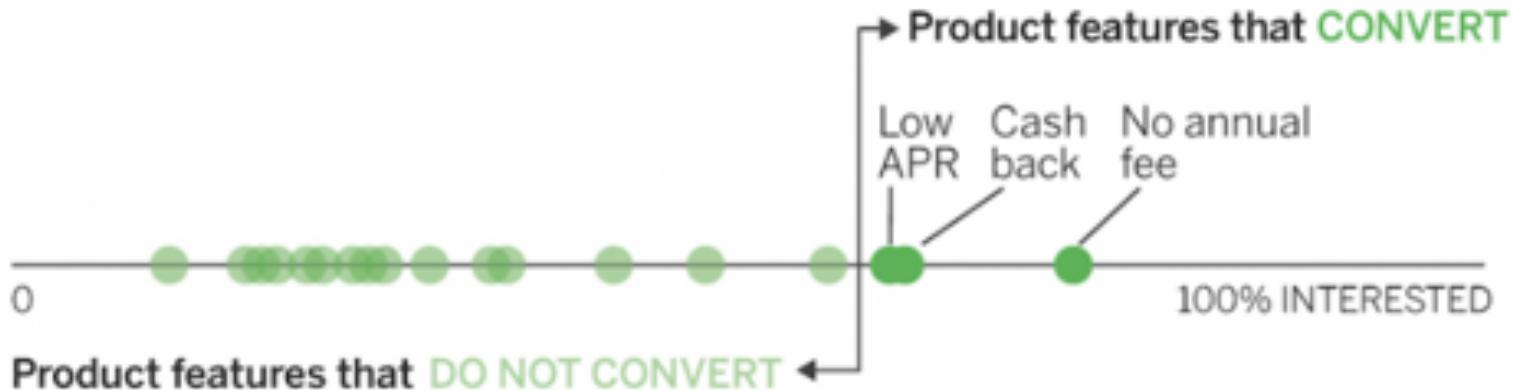
You want to convince your boss that marketing should focus on fewer product features when offering a new credit card. Your analysis tells you that only the top three features offered drive meaningful conversion. Sketch a way to make a more persuasive case.

## WHAT DO CUSTOMERS WANT FROM THEIR CREDIT CARDS?



- Retailer rewards
- Entry period bonus incentive
- Transferable points
- Restaurant rewards
- No foreign transaction fees
- Airline miles/travel rewards
- Hotel rewards
- Comprehensive mobile app
- Online chat support
- Contactless payment
- Access to events
- Concierge service

## FEW FEATURES DRIVE ACCOUNT CREATION

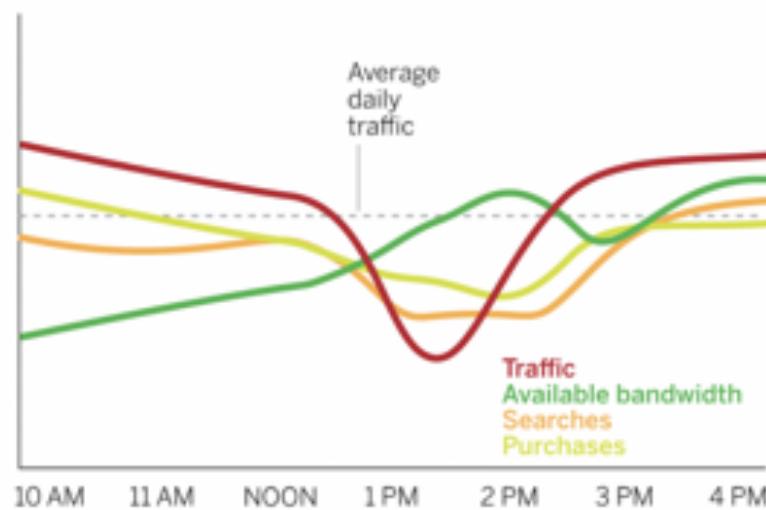


SOURCE: COMPANY RESEARCH

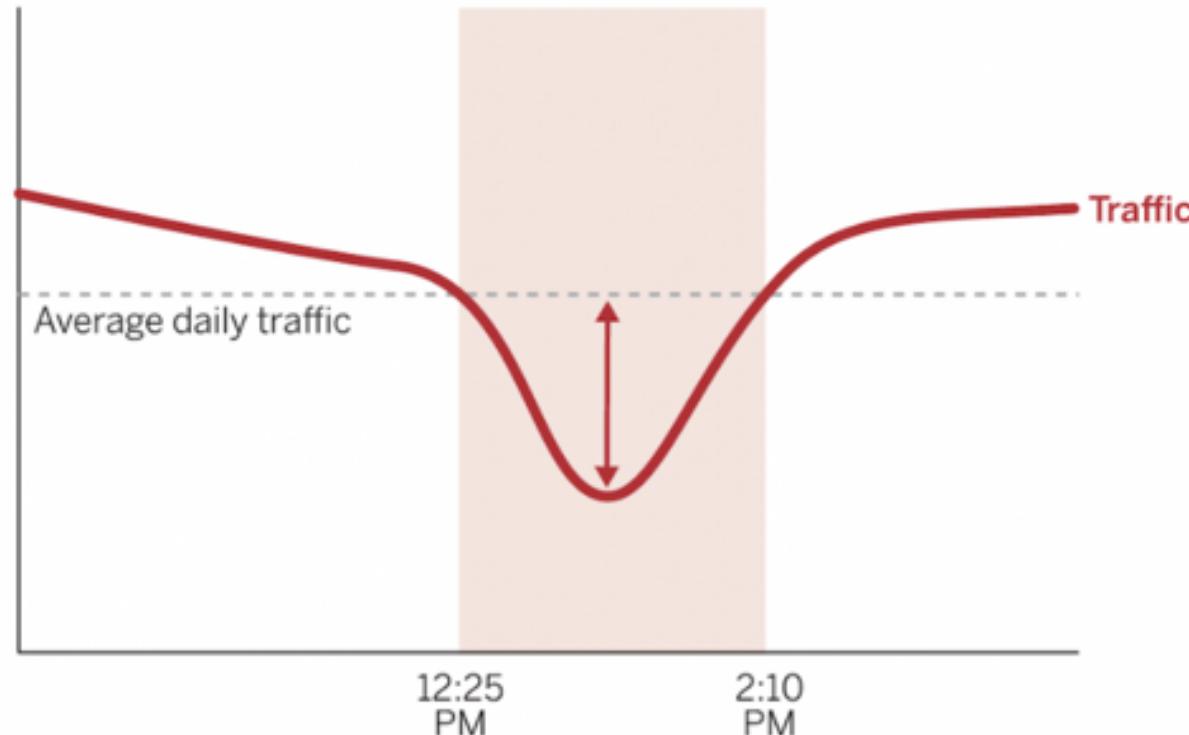
# Persuasion

You want to show your team that there's an opportunity to increase traffic to your website between 12:25 p.m. and 2:10 p.m. Adjust this chart to make it more persuasive on that point.

THE LUNCHTIME TRAFFIC DIP



## THE LUNCHTIME TRAFFIC OPPORTUNITY



# Persuasion

Which of the following statements will most likely lead to a chart that veers into unfair manipulation?

- A** "I need to convince them that they've set the year-end goals way too high, that there's no historical precedent for such aggressive targets, and that we're being set up for failure."
- B** "I have to show them how ridiculous they're being and that they're doing this just to make us look bad. The year-end goal is so off the charts that we'll basically fail by the end of the first quarter. There's no chance for success, and they need to stop being idiots."
- C** "I want to show them how unlikely it is that we can attain the goals they've set for the year."

# Persuasion

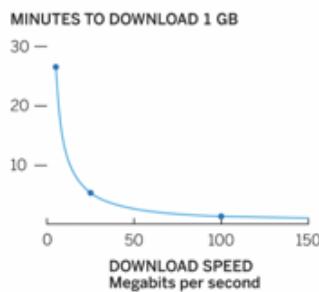
The scatter plot here shows a reasonably linear correlation between price and download speed. But the line chart shows that download speed is not a linear function. How might you adjust the scatter plot to amplify the point that our sense of the value of the bandwidth we pay for is skewed?

## THE PRICE OF SPEED



SOURCE: STEFANO PUNTINI

## THE TIME BENEFIT OF SPEED

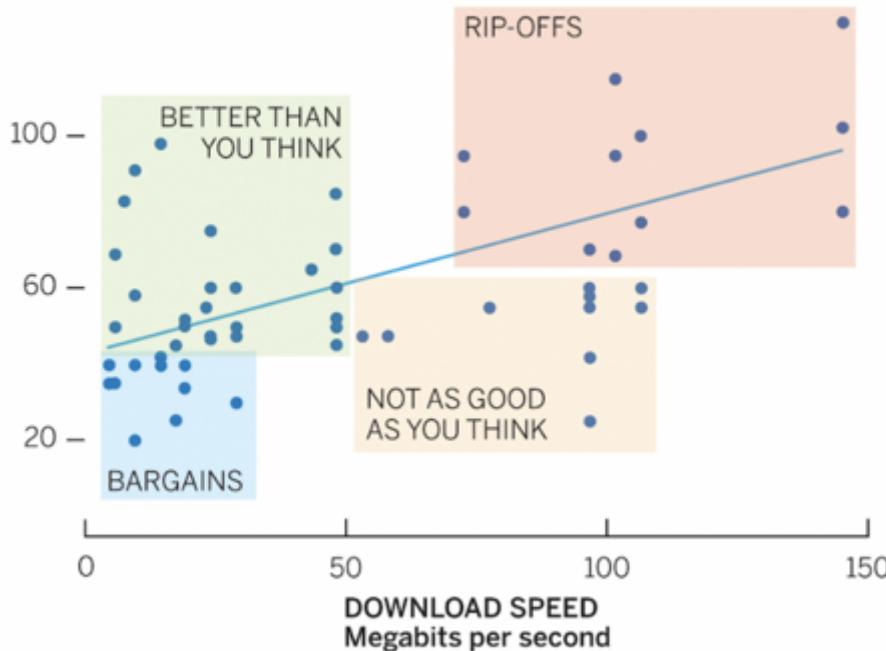


# Persuasion

## THE PRICE OF SPEED

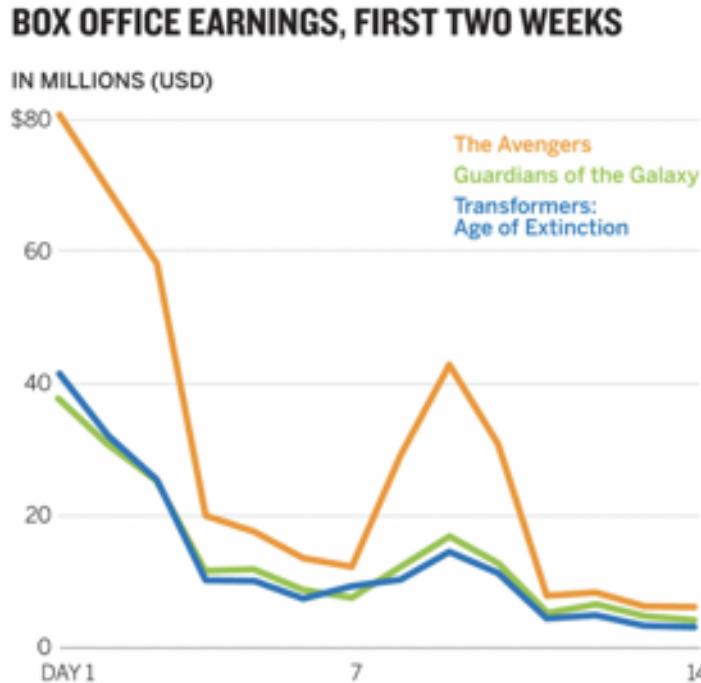
MONTHLY PRICE

\$140 —

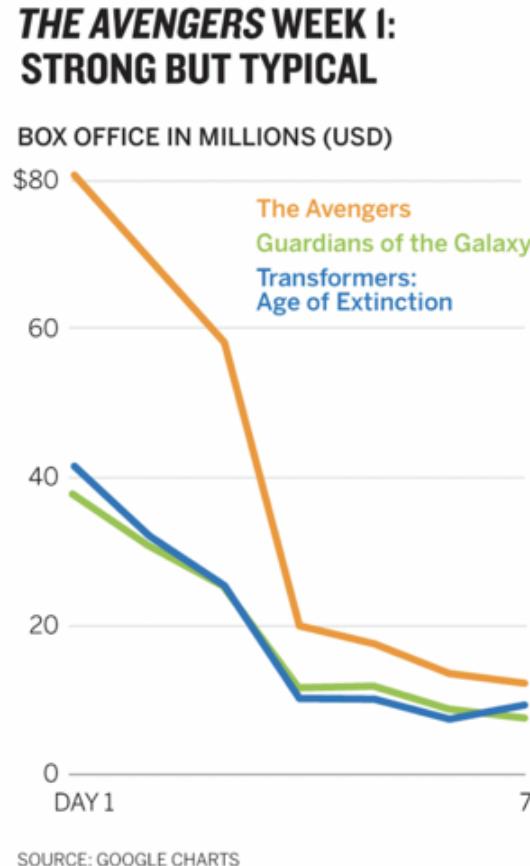


# Persuasion

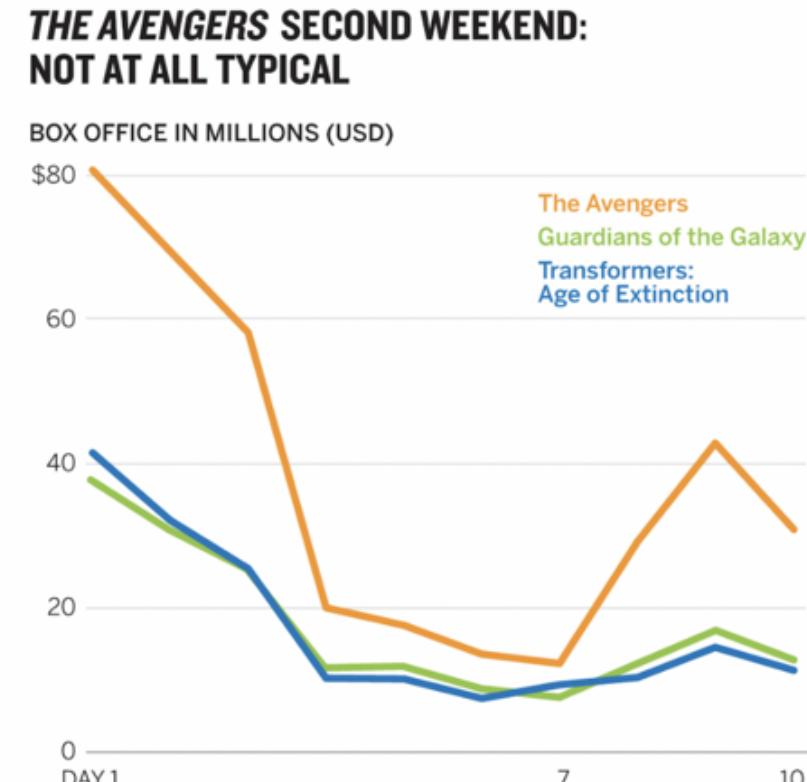
Tell a story with this chart. Write out a setup, a conflict, and a resolution and then sketch how you could present each part of the story.



# Persuasion



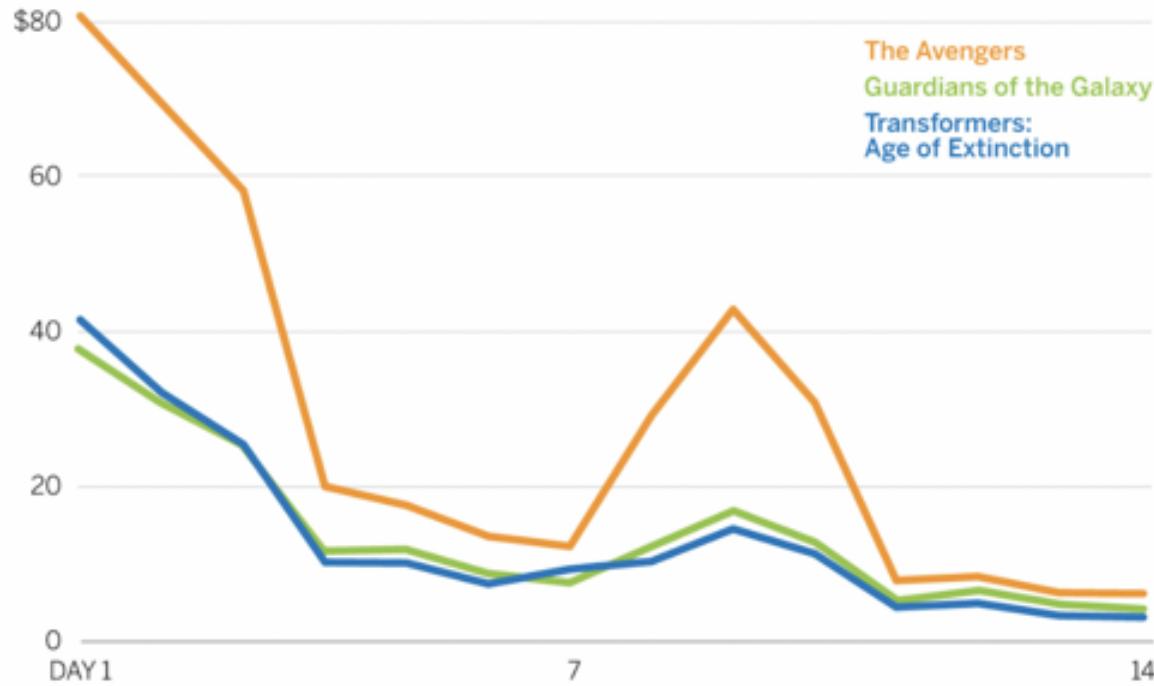
# Persuasion



# Persuasion

## THE AVENGERS WEEK 2: BACK TO NORMAL

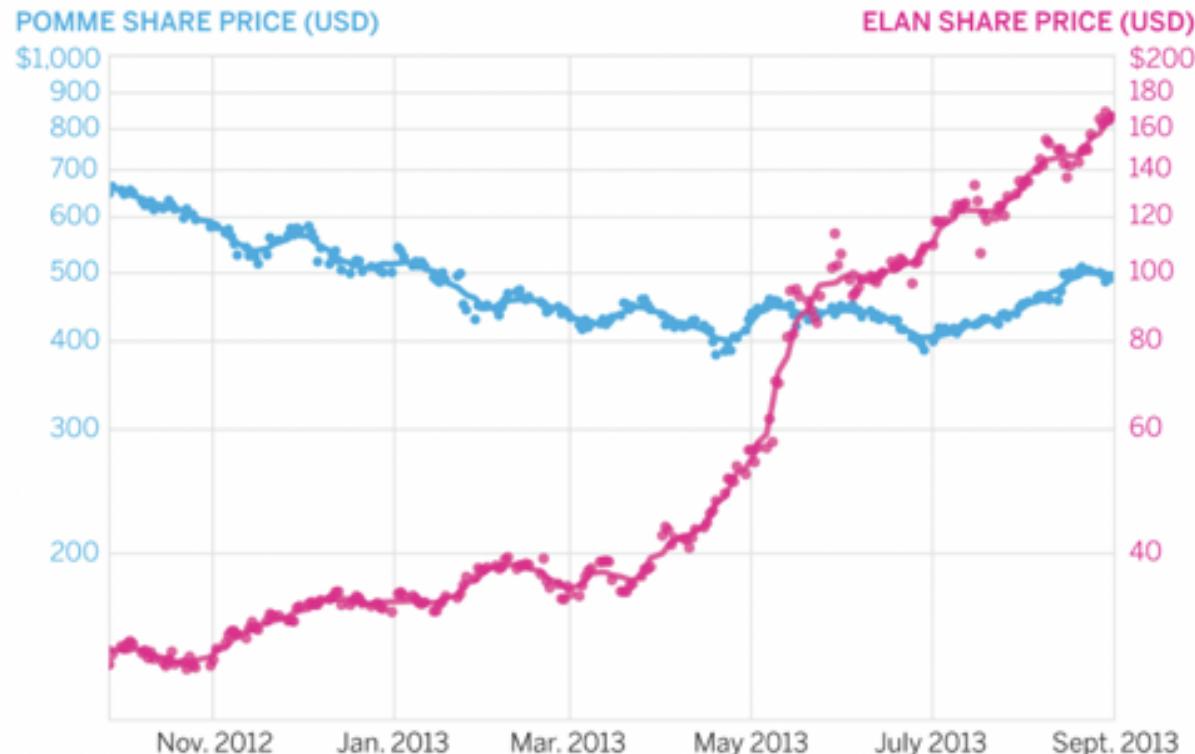
BOX OFFICE IN MILLIONS (USD)



SOURCE: GOOGLE CHARTS

# Persuasion

## ELAN INC. OVERTAKES POMME CO.



# Persuasion

1. Find and explain three ways this chart unfairly manipulates the user.
2. Sketch a version of it using share price that better reflects reality.
3. Sketch an alternative chart that supports the idea that Elan Inc. is the stronger investment.

# Persuasion

## ELAN INC. STRONGER THAN POMME CO.

SHARE PRICE (USD)

\$800

600

400

200

0

Elan

Pomme



# Persuasion

## ELAN INC. RISING, POMME CO. EBBING

% CHANGE IN STOCK PRICE SINCE SEPT. 17, 2012

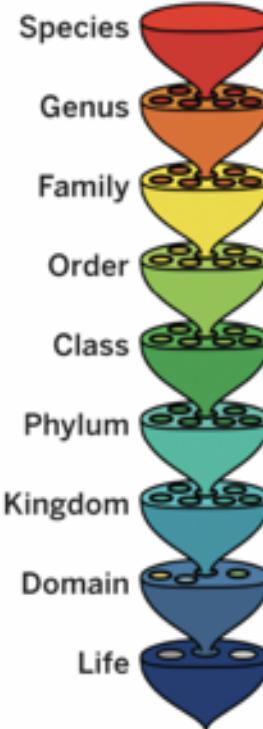


# Concepts

- Avoid mixed metaphors
- Be less literal
- Stick to conventions
- Edit

# Concepts

Find three elements that make this chart difficult to use and sketch an alternative approach to showing the biological classification system.



# Concepts

## BIOLOGICAL CLASSIFICATION WITH EXAMPLE

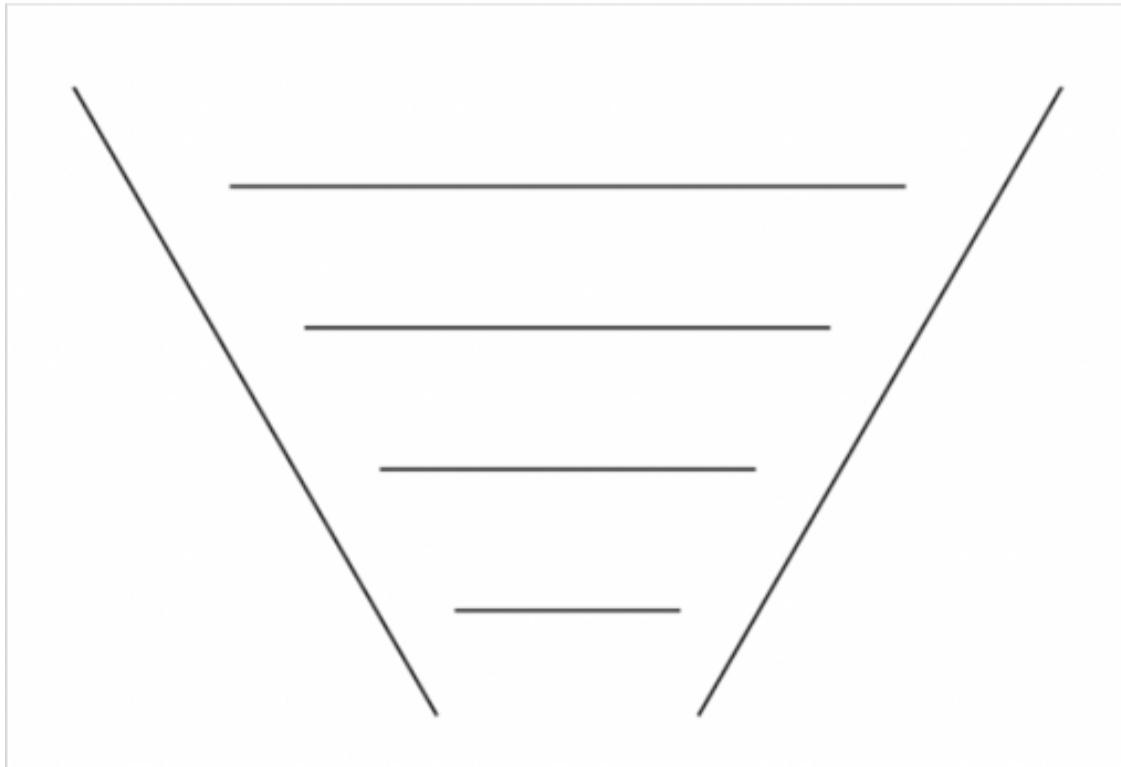
<u>LIFE</u>	
DOMAIN	Eukarya
KINGDOM	Animalia
PHYLUM	Chordata
CLASS	Mammalia
ORDER	Carnivora
FAMILY	Mustelidae
GENUS	<i>Enhydra</i>
SPECIES	<i>Enhydra Lutris</i> Sea otter

# Concepts

Sketch a new version of the classic purchase funnel below, eliminating all elements you think are extraneous.



# Concepts



# Concepts

---

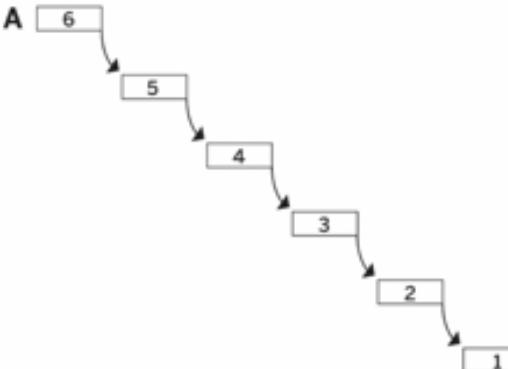
Maslow's hierarchy of needs is a framework for understanding human motivation. It describes human needs from the largest, most basic ones to more-specific and higher-level ones, which can be met only after the basic needs are achieved. The six levels of his hierarchy, starting with the most basic, are:

1. Physiological
2. Safety
3. Belonging and love
4. Esteem
5. Self-actualization
6. Self-transcendence

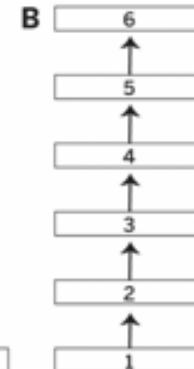
# Concepts

Which of the following would be a good conceptual form for showing Maslow's hierarchy?

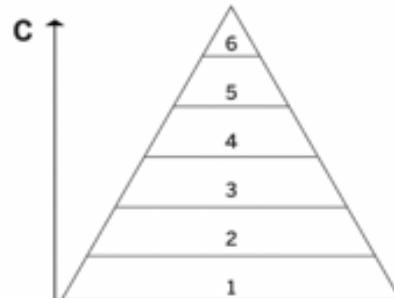
A



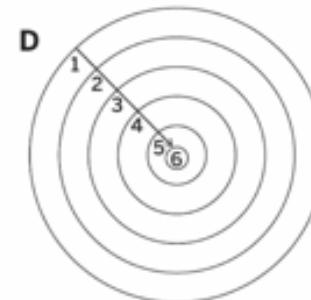
B



C

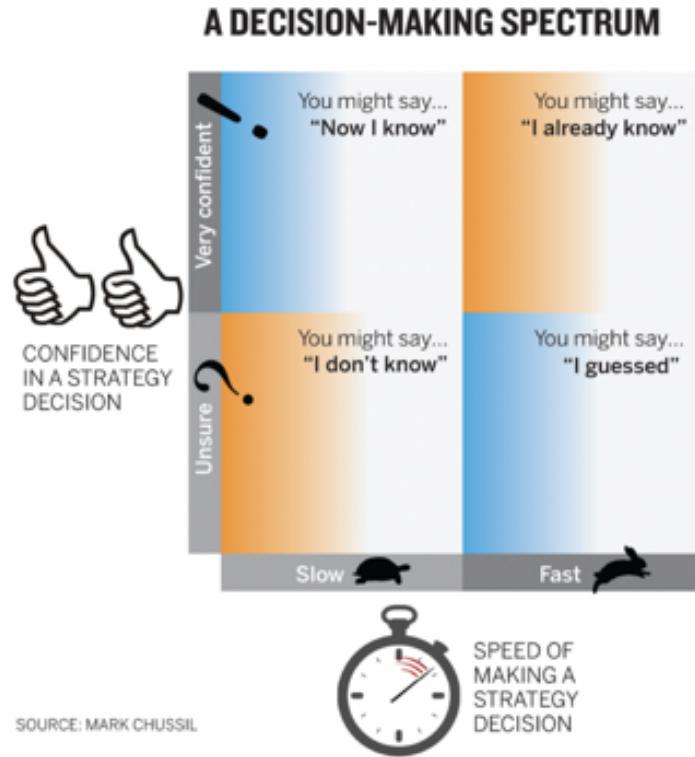


D



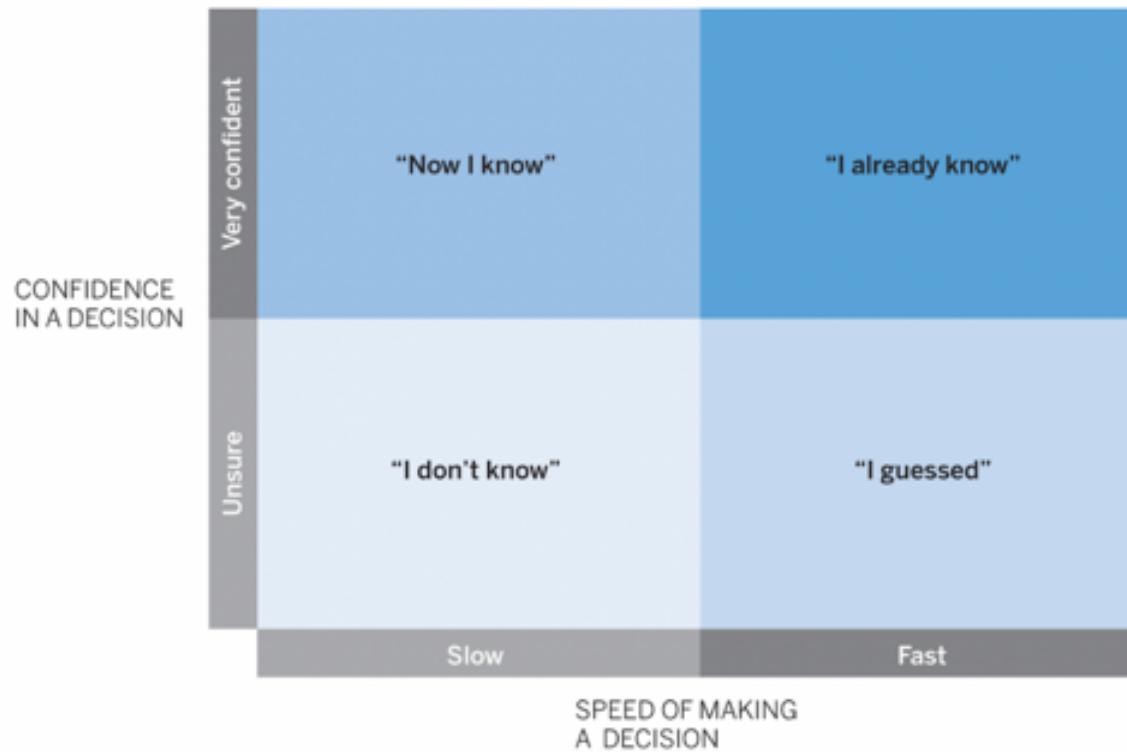
# Concepts

Identify elements that reduce the clarity of this 2x2 matrix and then sketch an improved version.



# Concepts

## 4 STYLES OF STRATEGY DECISION MAKING



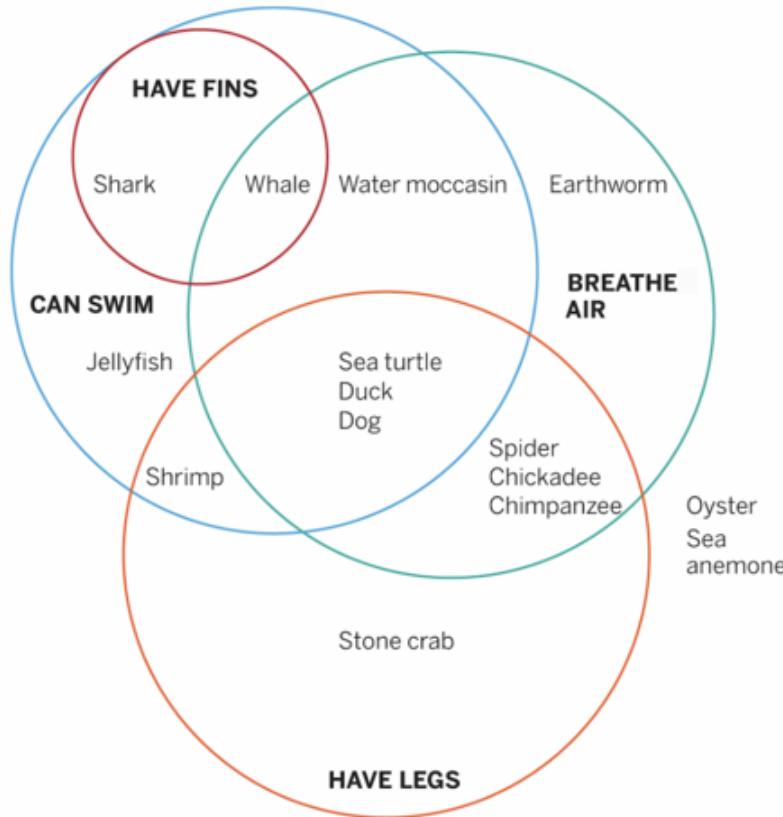
SOURCE: MARK CHUSSIL

# Concepts

How might you transform this table into a conceptual visualization?

	BREATHES AIR	CAN SWIM	HAS FINS	HAS LEGS
Chickadee	X			X
Chimpanzee	X			X
Dog	X	X		X
Duck	X	X		X
Earthworm	X			
Jellyfish		X		
Oyster				
Sea anemone				
Sea turtle	X	X		X
Shark		X	X	
Shrimp		X		
Spider	X			X
Stone crab				X
Water moccasin	X	X		
Whale	X	X	X	

# Concepts

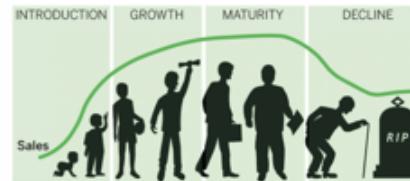


SOURCE: ADAPTED FROM A CHART BY DAVID WALBERT: LEARNNC.ORG/LP/PAGES/2646

# Concepts

You want to create a conceptual representation of a product's life cycle showing low sales at introduction, increasing sales through a growth phase, peaking sales at maturity, and declining sales during the last phase. Which representation would you choose?

## A THE PRODUCT LIFE CYCLE



## B THE PRODUCT LIFE CYCLE

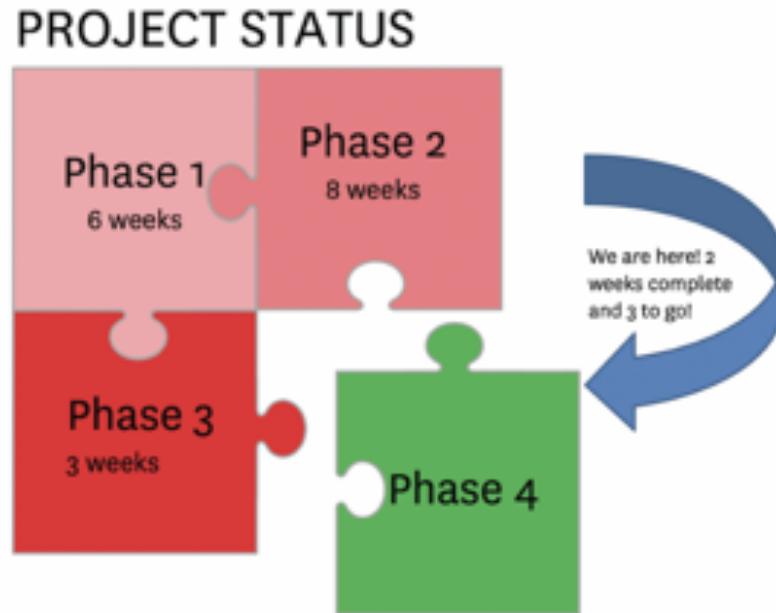


## C THE PRODUCT LIFE CYCLE



# Concepts

Find three elements that make this conceptual chart more difficult to use than it needs to be, and then reinvent it for improved clarity.



## PROJECT TIMELINE



# Example

<http://www.fallen.io/ww2/>