



Data Visualization presented to the NCSES Analytics Working Group by UVA/BII/SDAD

March 2, 2021

AGENDA

1. Taxonomy of a Visualization
2. 508 Color Blindness Compliance
3. Examples of Branding
4. Bar Chart Alternatives
5. Response to NCSES Comments & Questions
6. Next Steps

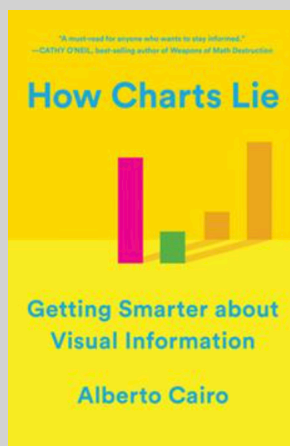


**BIOCOMPLEXITY
INSTITUTE**

Visualization = Scaffolding + Content

Scaffolding consists of...

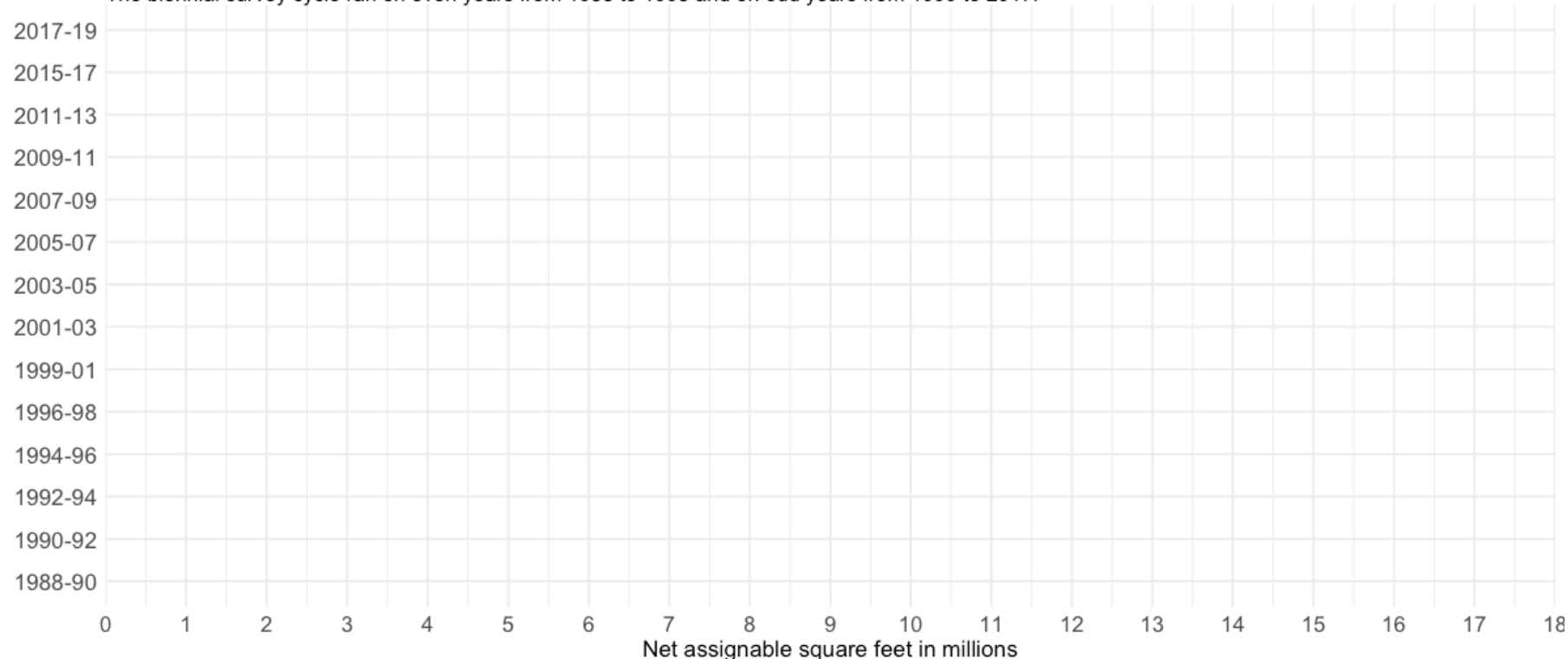
- Title
- Sub-title
- Axis Titles
- Axis Labels
- Legends
- Measurement
- Measurement Unit
- Scale
- Data Sources



Adapted from

Science and engineering research space in academic institution, change over 2-year periods: FYs 1988–2017

The biennial survey cycle ran on even years from 1988 to 1998 and on odd years from 1999 to 2017.



SOURCE: National Science Foundation, National Center for Science and Engineering Statistics, Survey of Science and Engineering Research Facilities

Visualization = Scaffolding + Content

Scaffolding consists of...

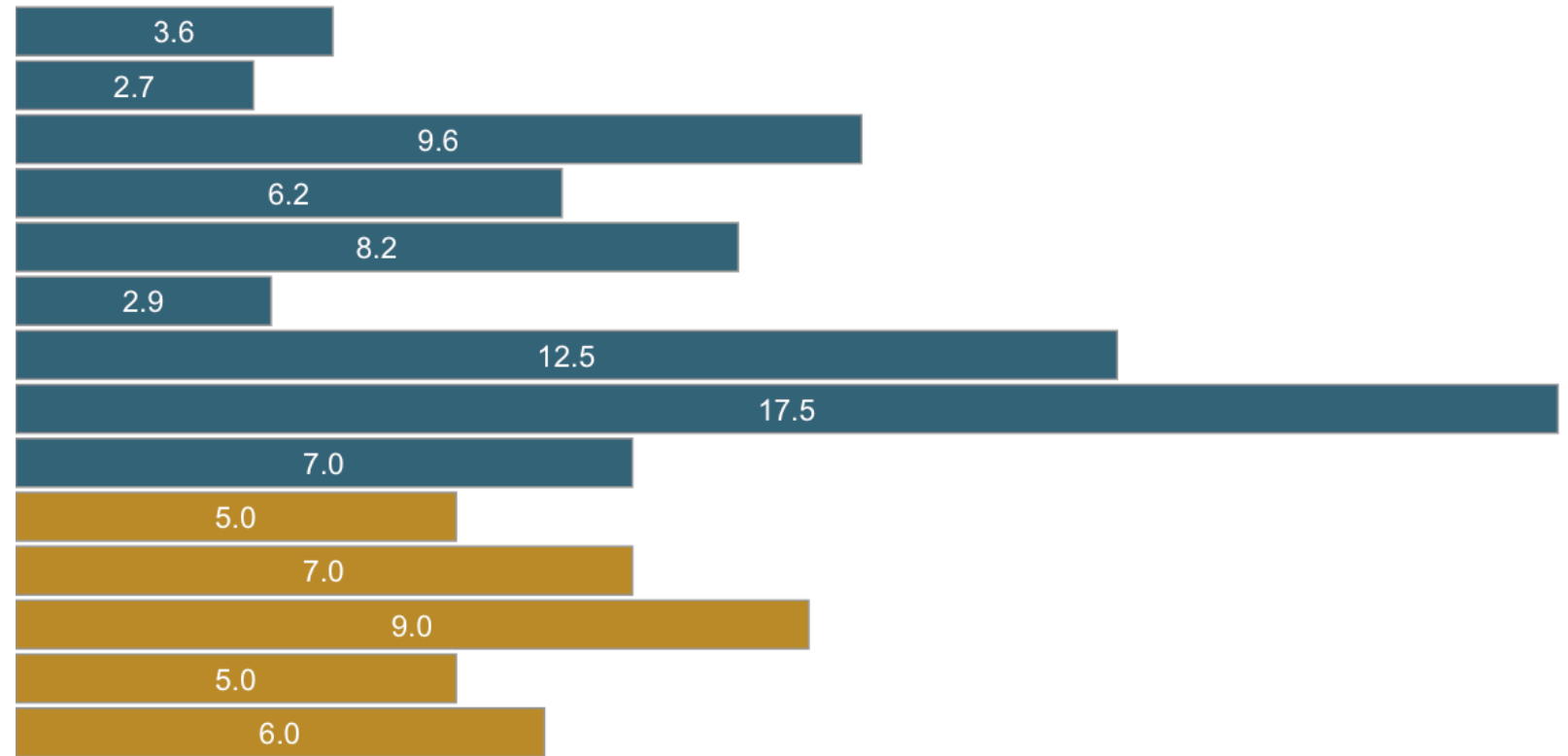
- Title
- Sub-title
- Axis Titles
- Axis Labels
- Legends
- Measurement
- Measurement Unit
- Scale
- Data Sources

When you think about branding regarding the scaffolding you can manipulate the:

- Font
- Font sizes
- Font colors
- Grid lines
- Background color
- Whether to include the NCSES logo and branding shapes

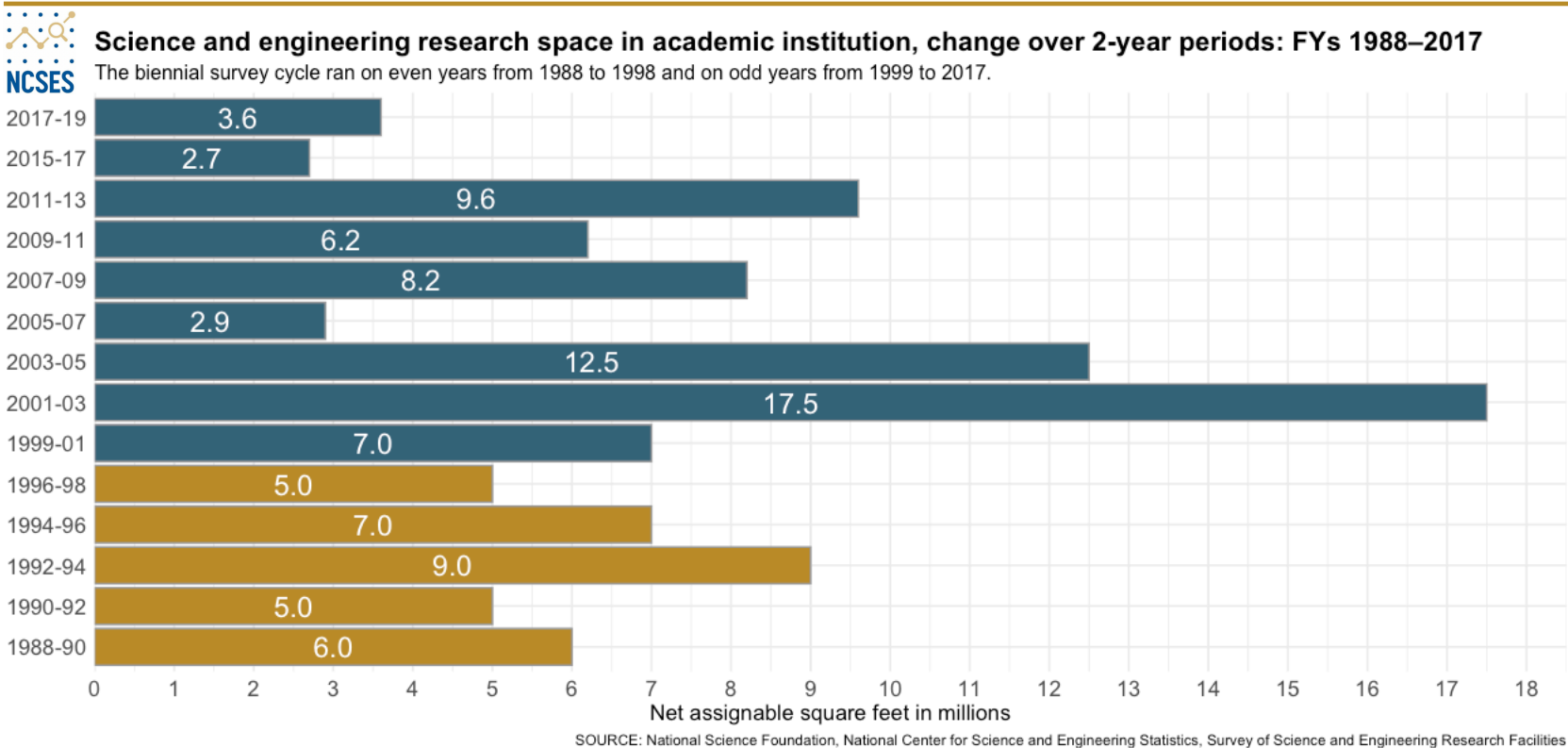
The **content** consists of...

- How the data are represented or encoded
- An **annotation layer** that can be used to clarify a point of interest
- In this case the data are represented by the length of a bar and annotation is used to provide the encoded value of the length

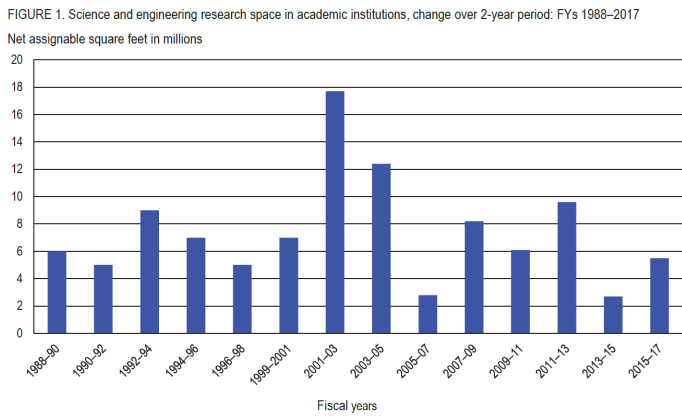


Complete Visualization

A **bar chart** is a visualization of two variables. In this case one with an ordinal level of measurement, year, and the other with a ratio level of measurement, square feet.



“The bar chart that always accompanies the Research Facilities IB (attached). What would be more dynamic alternatives?”



NOTE: The biennial survey cycle ran on even years from 1988 to 1998 and on odd years from 1999 to 2017.
SOURCE: National Science Foundation, National Center for Science and Engineering Statistics, Survey of Science and Engineering Research Facilities.

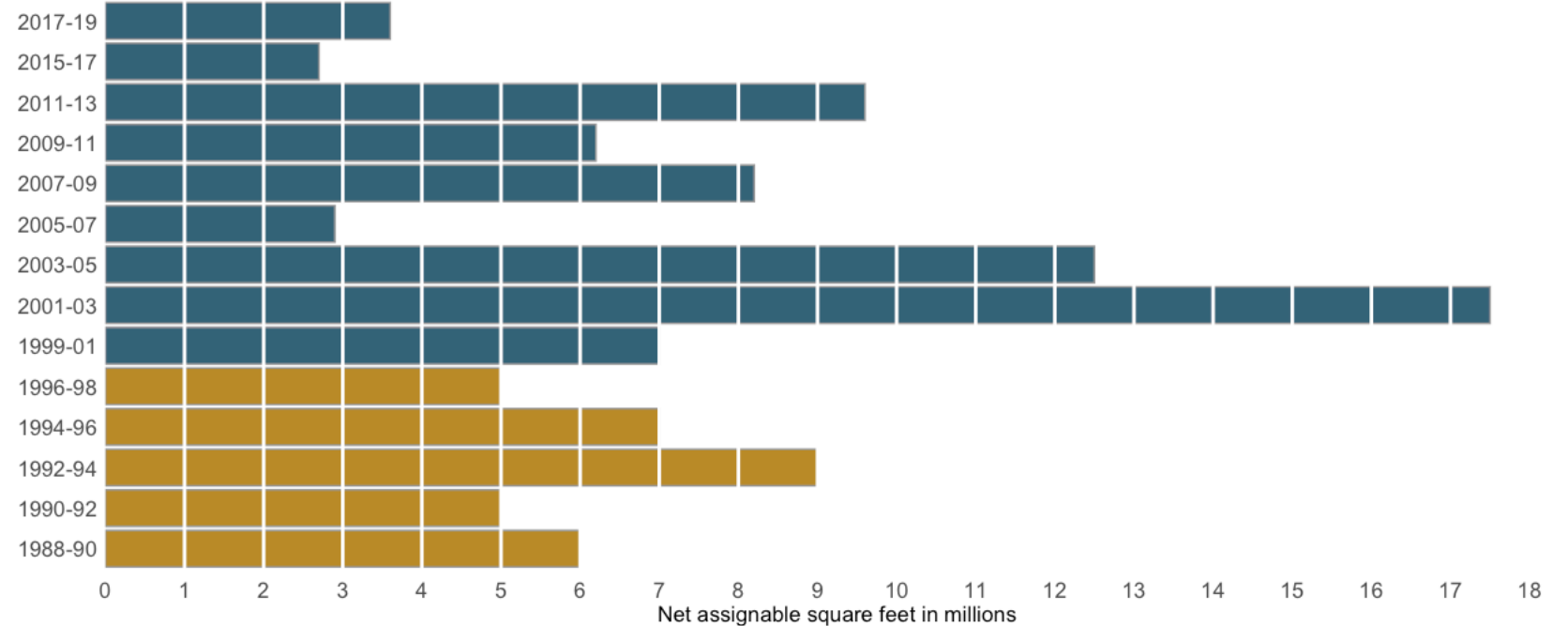
Complete Visualization

In the plot the grid lines are white and go over the bars.



Science and engineering research space in academic institution, change over 2-year periods: FYs 1988–2017

The biennial survey cycle ran on even years from 1988 to 1998 and on odd years from 1999 to 2017.

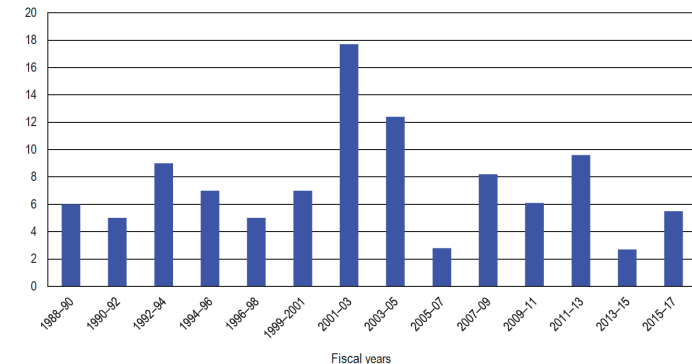


SOURCE: National Science Foundation, National Center for Science and Engineering Statistics, Survey of Science and Engineering Research Facilities

“The bar chart that always accompanies the Research Facilities IB (attached). What would be more dynamic alternatives?”

FIGURE 1. Science and engineering research space in academic institutions, change over 2-year period: FYs 1988–2017

Net assignable square feet in millions



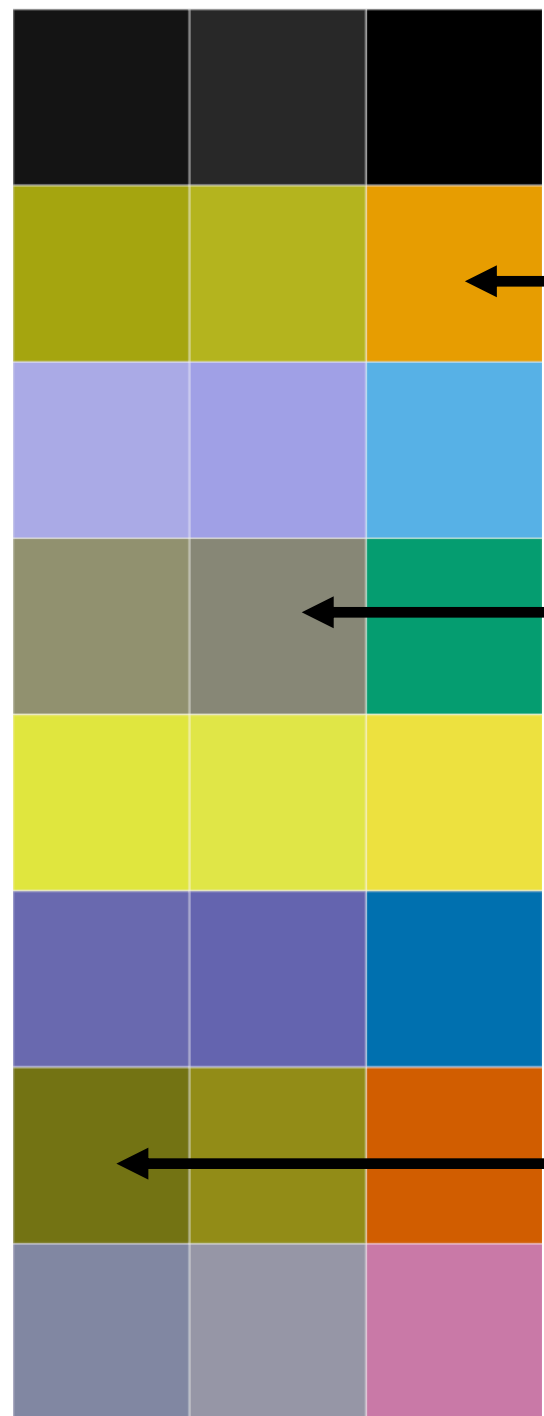
NOTE: The biennial survey cycle ran on even years from 1988 to 1998 and on odd years from 1999 to 2017.

SOURCE: National Science Foundation, National Center for Science and Engineering Statistics, Survey of Science and Engineering Research Facilities.

508 Color Compliance

Colorblindness:

- is not a total loss of color vision;
- the range of colors that are discernable is a function of the photoreceptor cells in the retinae which are sensitive to the spectrum red, green and blue.



Colorblind Palette (3rd column):
What a non colorblind person sees.

Deuteranopia (middle column):
Red-green color blindness; less sensitive to green light.

Protanopia (first column):
Red-green color blindness; less sensitive to red light.

508 Color Compliance

Resource: **Colblinder**
Provides a colorblindness simulator. Drag your visualization into the simulator and you can see the various forms of color anomalies.

<https://www.color-blindness.com/coblis-color-blindness-simulator/>

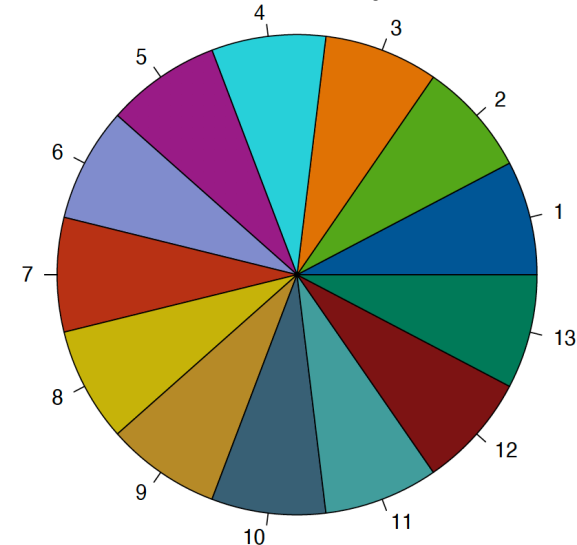
R Package colorBlindness



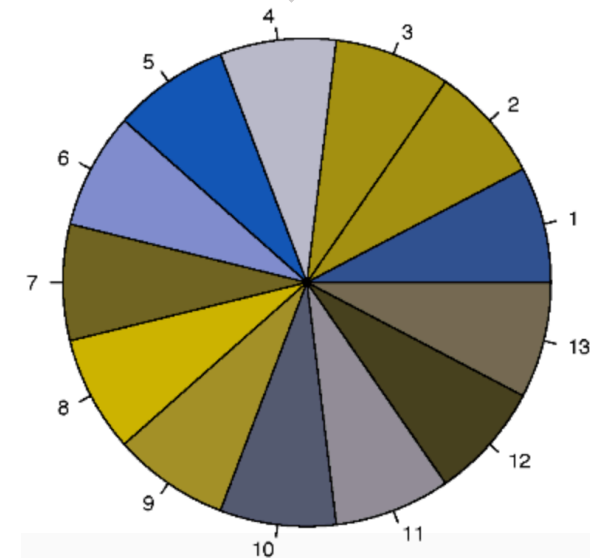
Red-Blind/Protanopia



NCSES Primary Colors



Red-Blind/Protanopia



- 1 = #005799
- 2 = #54A719
- 3 = #E57503
- 4 = #2ED1D9
- 5 = #9D1E87
- 6 = #8190CE
- 7 = #BE3213
- 8 = #CBB608
- 9 = #BA8D29
- 10 = #386379
- 11 = #469FA0
- 12 = #7F1416
- 13 = #007D59

508 Color Compliance

Resources:

Color Universal Design:
How to make figures and
presentations that are
friendly to Colorblind
people.

<https://jfly.uni-koeln.de/color/>

Some ideas on how to produce visualizations that are color deficient friendly:

1. Use thicker lines and bigger symbols make it easier to distinguish colors.
2. Try to keep the number of colors to a minimum.
3. Use redundant coding. Do not convey information with just colors, for example, in line graphs use colors and symbols (circles, triangles, or different line types (dashed, solid, dotted, etc.)).
4. Do not use legends, provide labels within the figure.

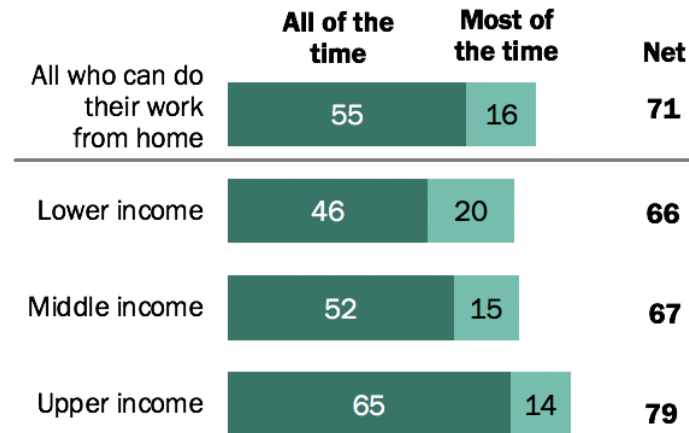
Pew: very novel scaffold (legend, axes titles & labels)

Branding consists of...

- Title: black bolded
- Sub-title: grey italic different font than the title
- Axis Titles: grey same font as the title or minimal
- Axis Labels: grey same font as the title: minimum and maximum or none
- Legends: annotation
- Measurement: identified in sub-title
- Data Sources: identified at the bottom

Most who can do their job from home say they are currently doing so all or most of the time

Among employed adults who say that, for the most part, the responsibilities of their job can be done from home, % saying they are currently working from home ...

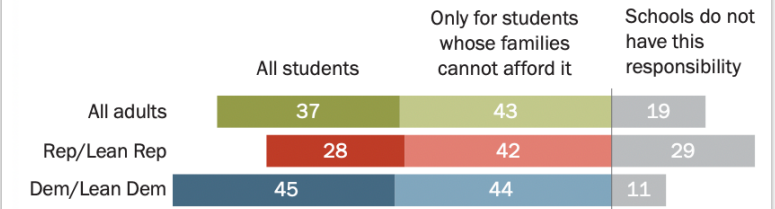


Note: Figures may not add to subtotals due to rounding. Family income tiers are based on adjusted 2019 earnings.
Source: Survey of U.S. adults conducted Oct. 13-19, 2020.
"How the Coronavirus Outbreak Has – and Hasn't – Changed the Way Americans Work"

PEW RESEARCH CENTER

While most Americans say schools should provide computers to at least some students during outbreak, parties divided on if this should be done for all

% of U.S. adults who say that K-12 schools have a responsibility to provide laptops or tablet computers to ___ in order to help them complete schoolwork at home during the coronavirus outbreak



Note: Those who did not give an answer are not shown.

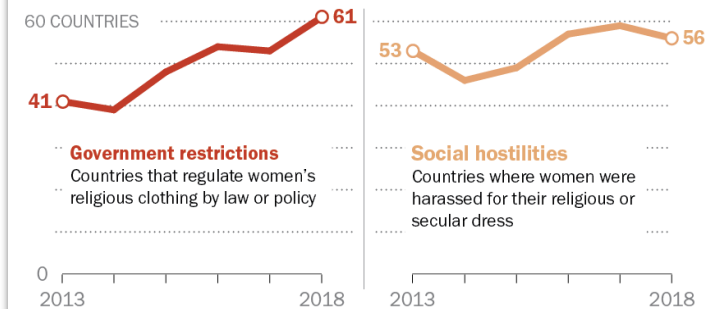
Source: Survey of U.S. adults conducted April 7-12, 2020.

"53% of Americans Say the Internet Has Been Essential During the COVID-19 Outbreak"

PEW RESEARCH CENTER

Restrictions on women's dress have risen around the world in the past five years

Number of countries where women experienced ...



Note: Government restrictions include both rules that prohibit religious garb and rules that require it. Social hostilities data for each year also takes into account incidents from the two previous years to capture ongoing hostilities.
Source: Pew Research Center analysis of external data.

"In 2018, Government Restrictions on Religion Reach Highest Level Globally in More Than a Decade"

PEW RESEARCH CENTER

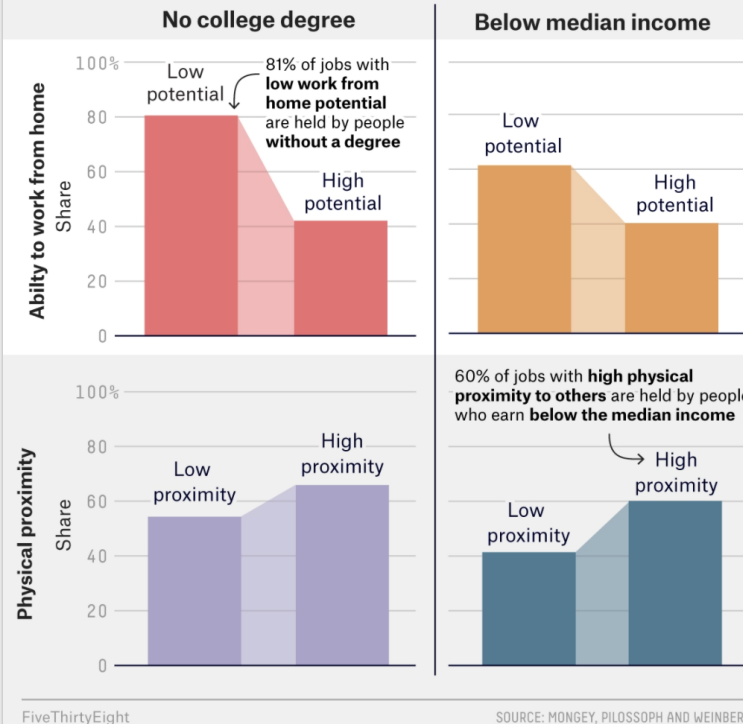
FiveThirtyEight: grey background & annotated legend

Branding consists of...

- Title: black bolded
- Sub-title: grey italic
- Axis Titles: same color as title
- Axis Labels: grey same font as the title
- Legends: annotation
- Measurement: identified in sub-title
- Data Sources: identified at the bottom

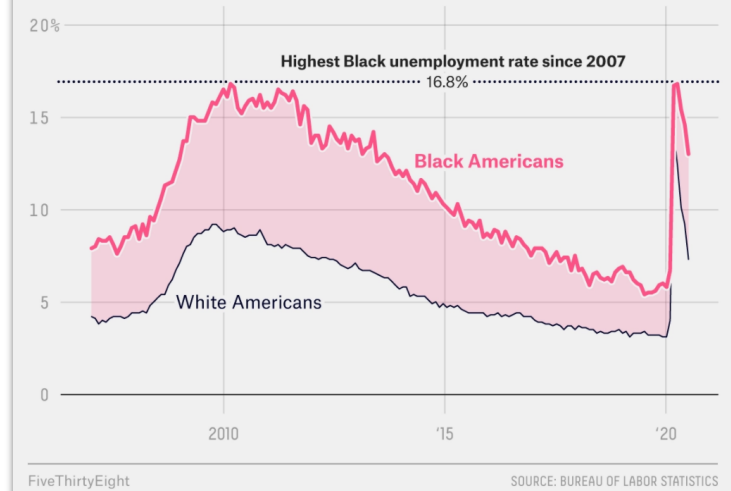
Vulnerable populations tend to have the high-risk jobs

Share of workers among high and low likelihoods of being able to work from home and high and low physical proximity to others at work



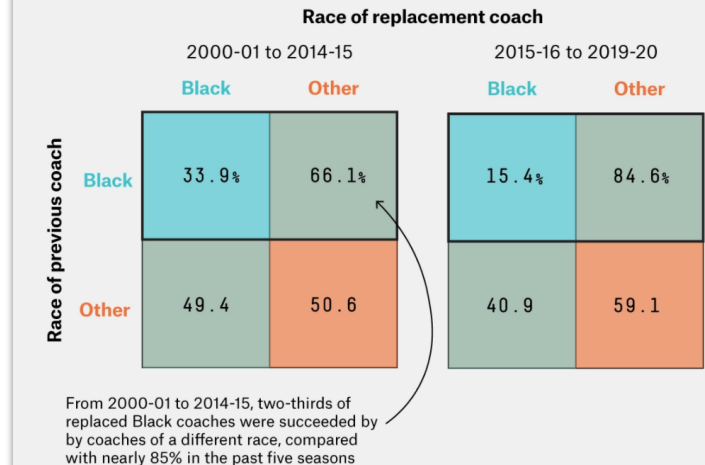
The Black unemployment rate remains very high

Unemployment rate by race, Jan. 2007 through Aug. 2020



Black coaches aren't being replaced by Black coaches

NBA head coaches who left a team, broken down by race and by the race of the coaches replacing them at the start of the next season



Interim coaches were not included in these calculations.

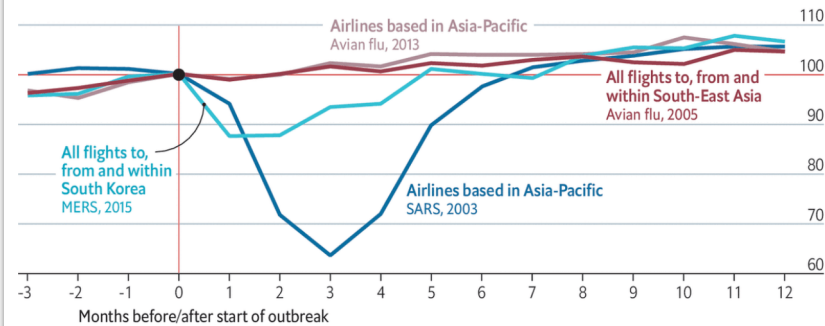
Economist: scaffold with axes labels on the right and top

Branding consists of...

- Title: black bolded
- Sub-title: grey
- Axis Titles: none or minimal
- Axis Labels: grey same font as the title on the right
- Legends: a single use of legends otherwise annotation is used
- Measurement: identified in sub-title
- Data Sources: identified at the bottom
- Red line at the top

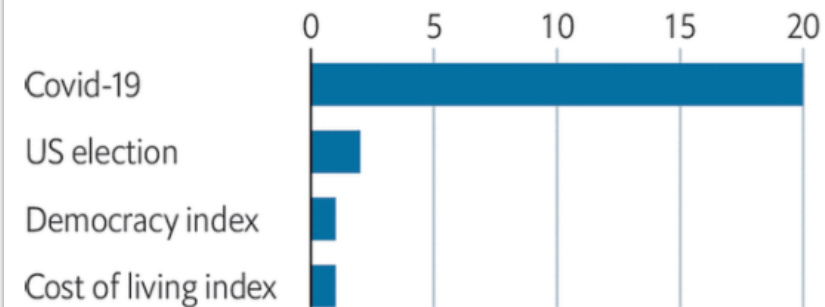
Taking flight

Impact of virus outbreaks on aviation, revenue-passenger km*
Start of outbreak=100



Viral content

Subject matter of our 24 most popular articles*, 2020



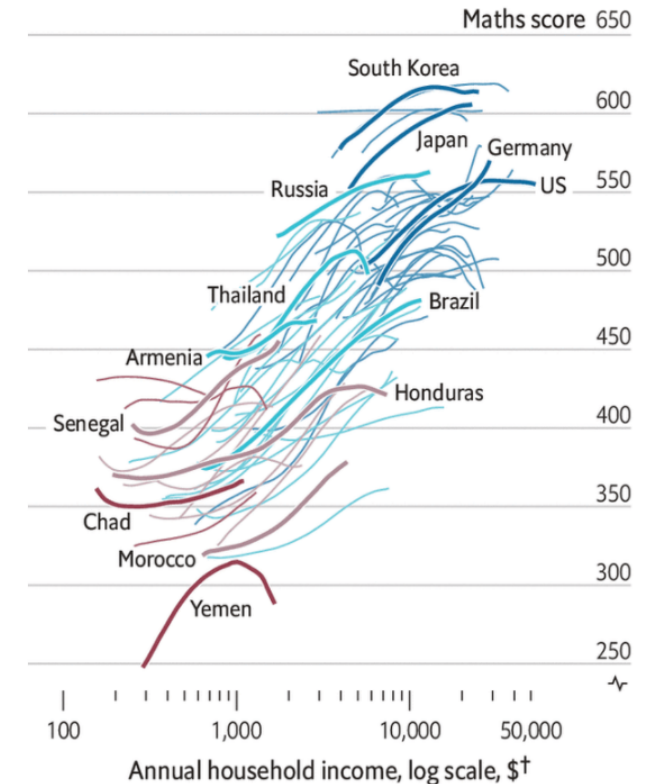
The Economist

It all adds up

Household income and maths test scores*
Between the 10th and 90th percentile of income in each country

Country income group

— High — Upper middle — Lower middle — Low



*Estimated TIMSS score equivalent, out of 1,000

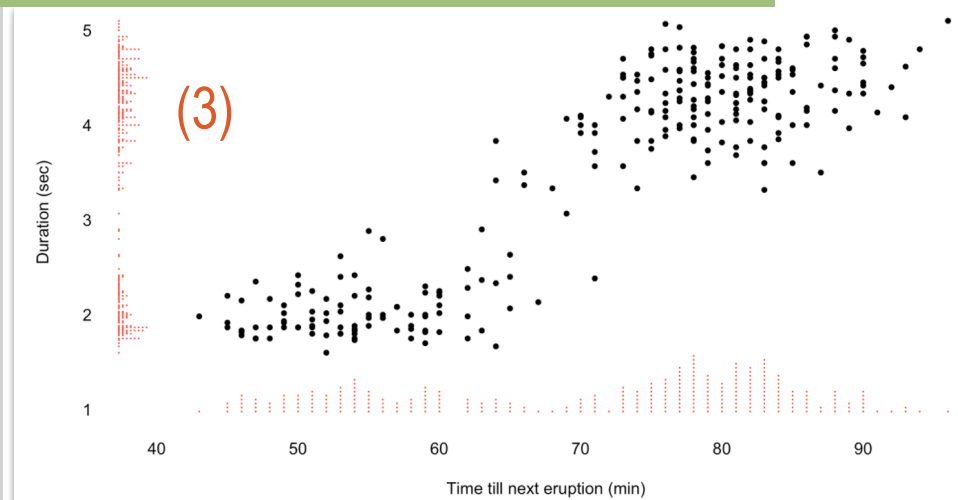
†In 2005 dollars, at purchasing-power parity (PPP)

Sources: "A Rosetta Stone for Human Capital", by D. Patel and J. Sandefur; World Bank; The Economist
The Economist

Tufte: adding additional information to the axes

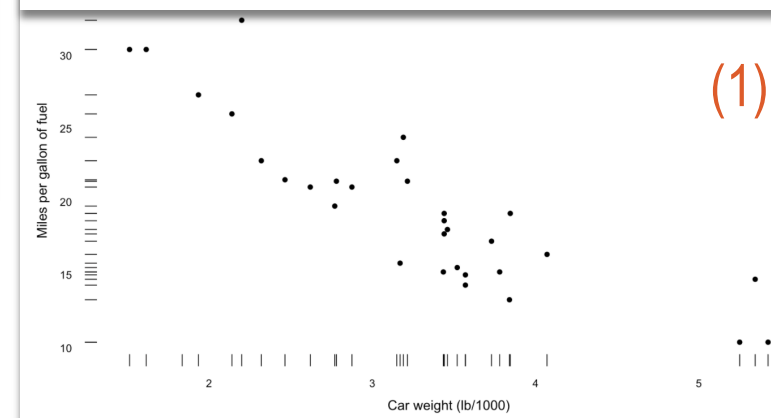
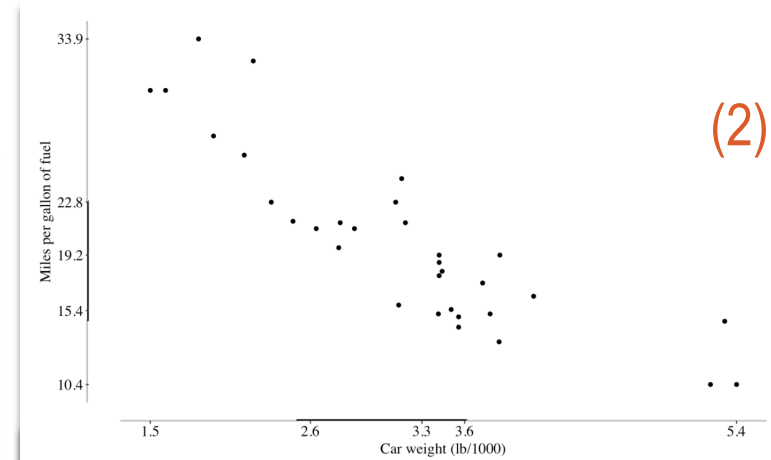
- Branding** consists of...
- packing a lot of information either using the axes
 - (2) labels - the five-number summary
 - (3) tick marks – marginal histograms

No grid lines



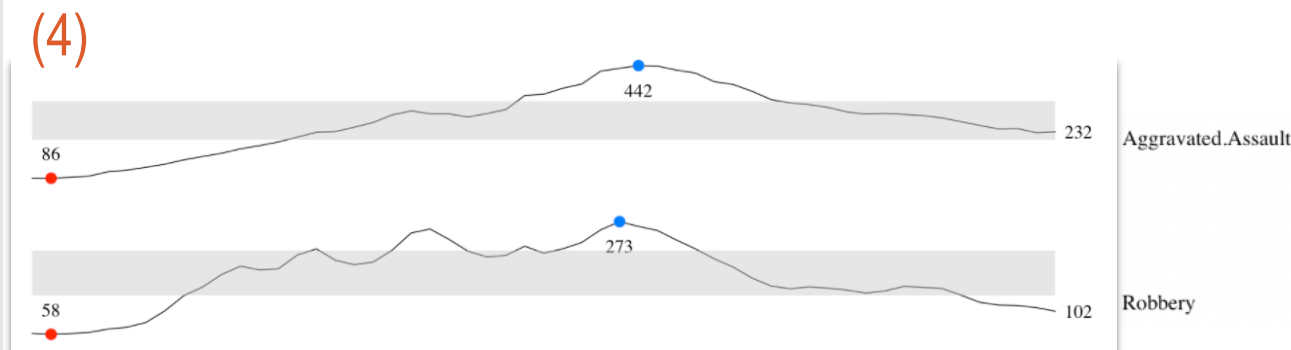
Dot Plots

(1) Dot-Dash, (2) Dot-Range frame, (3) Dot-Marginal histogram



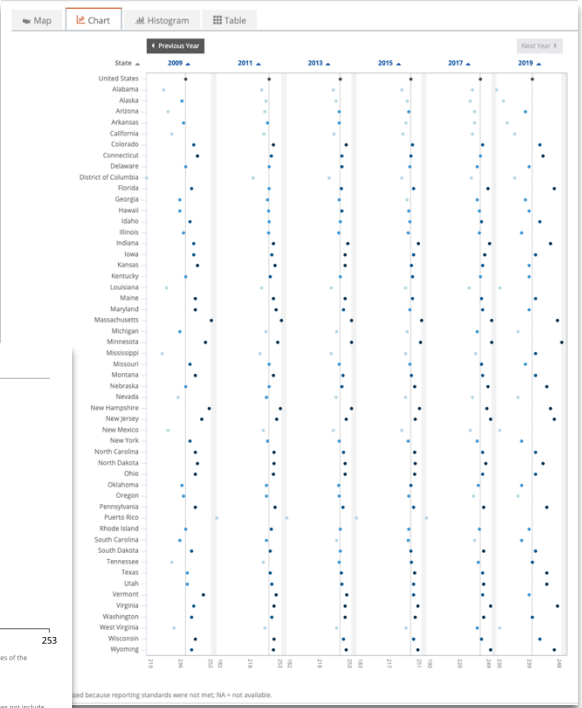
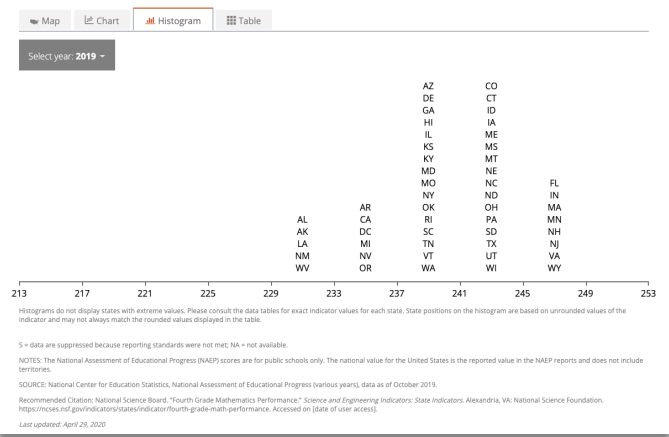
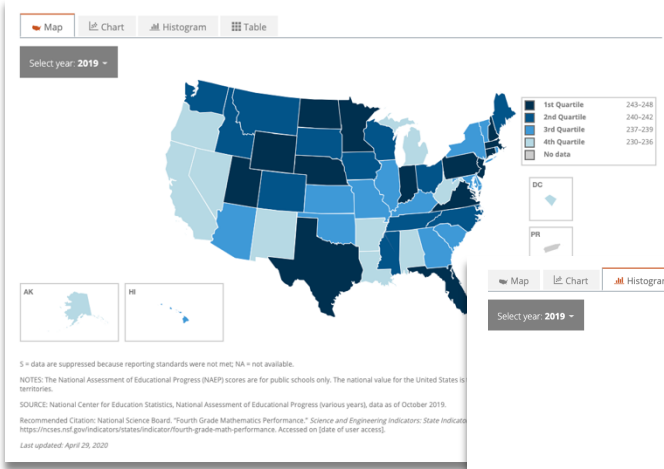
Sparklines

Blue Dot = Maximum
Red Dot = Minimum
Grey Rectangle = area between the 1st and 3rd quartiles



State Indicators

Q1 “The state indicators are one of NCSES’s more popular products so we should try to create engaging and effective visualizations.”



| State ^ | 2000 ^ | 2003 ^ | 2005 ^ | 2007 ^ | 2009 ^ | 2011 ^ | 2013 ^ | 2015 ^ | 2017 ^ | 2019 ^ | Trend |
|----------------------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|-------|
| United States | 224 | 234 | 237 | 239 | 239 | 240 | 241 | 240 | 239 | 240 | |
| Alabama | 217 | 223 | 225 | 229 | 228 | 231 | 233 | 231 | 232 | 230 | |
| Alaska | NA | 233 | 236 | 237 | 237 | 236 | 236 | 236 | 230 | 232 | |
| Arizona | 219 | 229 | 230 | 232 | 230 | 235 | 240 | 238 | 234 | 238 | |
| Arkansas | 216 | 229 | 236 | 238 | 238 | 238 | 240 | 235 | 234 | 233 | |
| California | 213 | 227 | 230 | 230 | 232 | 234 | 234 | 232 | 232 | 235 | |
| Colorado | NA | 235 | 239 | 240 | 243 | 244 | 247 | 242 | 241 | 242 | |
| Connecticut | 234 | 241 | 242 | 243 | 245 | 242 | 243 | 240 | 239 | 243 | |
| Delaware | NA | 236 | 240 | 242 | 239 | 240 | 243 | 239 | 236 | 239 | |
| District of Columbia | 192 | 205 | 211 | 214 | 219 | 222 | 229 | 231 | 231 | 235 | |

State Indicators

Q1 *“The state indicators are one of NCSES’s more popular products so we should try to create engaging and effective visualizations.”*

Comments: Provide sufficient information about the data being displayed so the viewer will know how it can be used and interpreted.

“This indicator represents the average score in mathematics of each state’s fourth grade public school students on the National Assessment of Educational Progress (NAEP). ... Not all students participate in NAEP assessments; a sample of schools and students is selected to represent each participating state.”

- How is the sample selected? What percentage of 4th grade students in public schools participate in each state? Is the percentage the same for each state? What percentage of schools are represented?

State Indicators

Q1 *“The state indicators are one of NCSES’s more popular products so we should try to create engaging and effective visualizations.”*

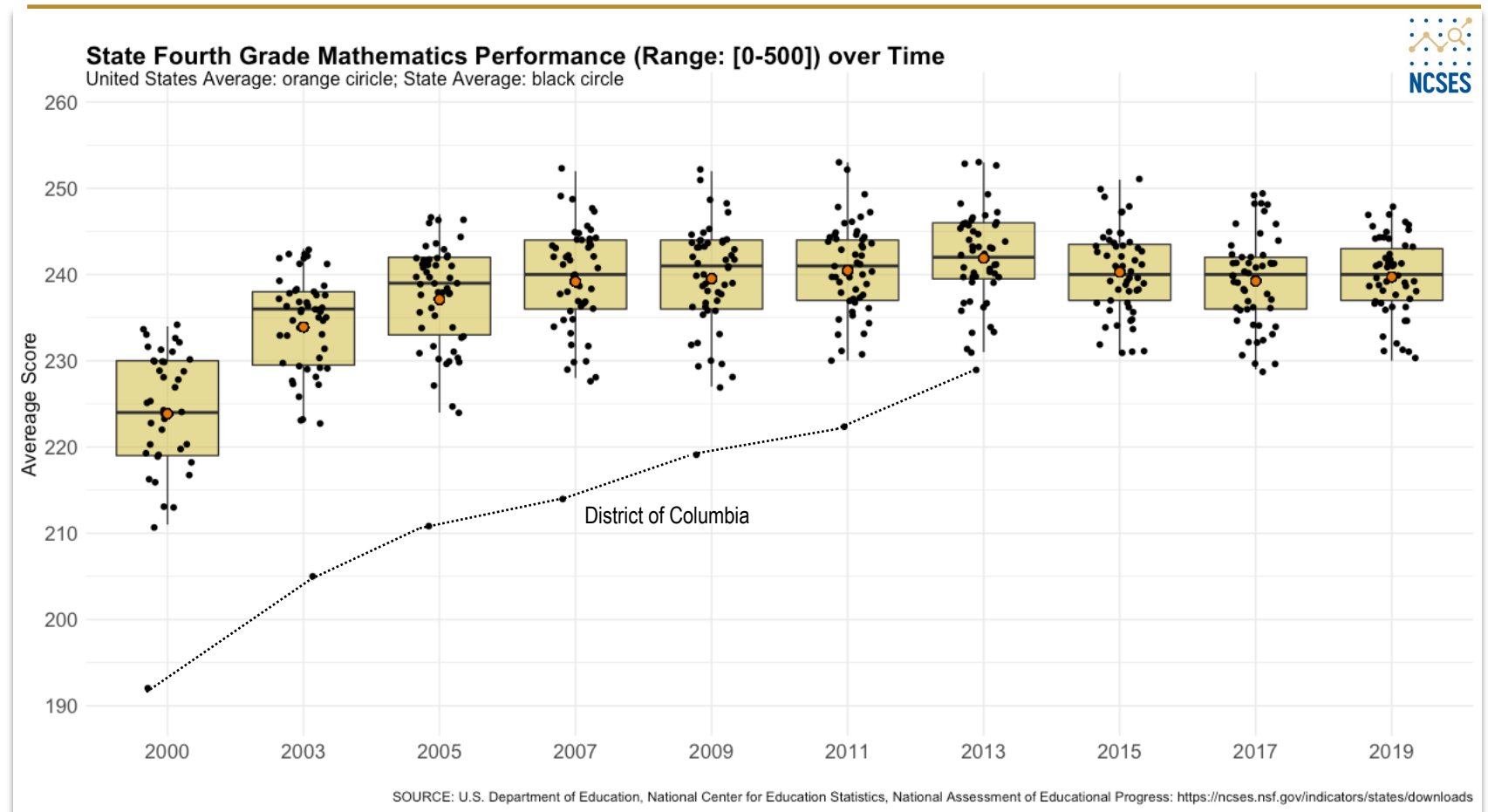
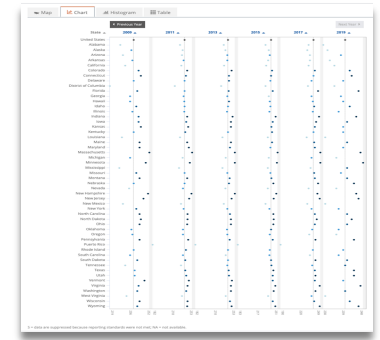
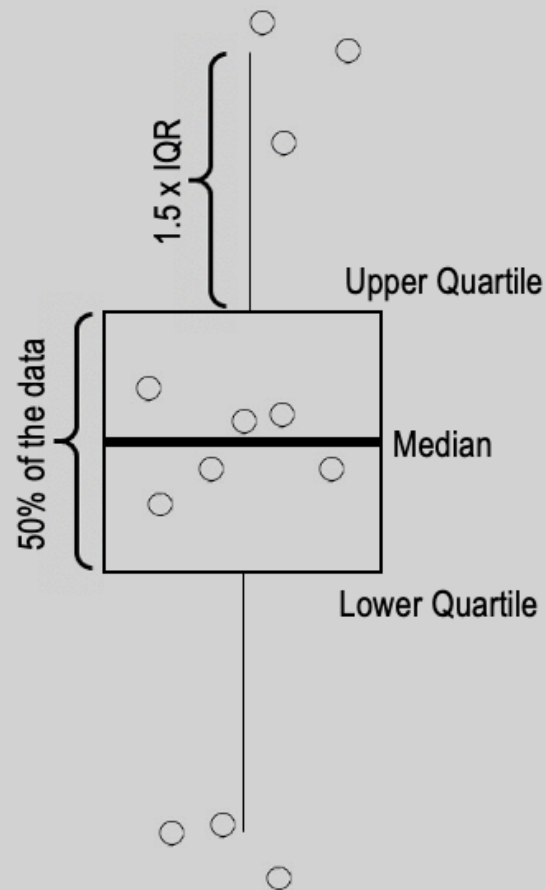
Comments: Instead of thinking of various ways to display the data think what questions users might have about the data.

1. What are the state and national trends over time?
2. How do states differ from the national average over time (above/below the average)?
3. What percentage of students are represented in the estimates over time?
4. What percentage of schools are represented in the estimates over time?
5. Would it make sense to provide data summaries by school regions?

State Indicators

Median = 50% of the values lie above and below the median

Inter-Quartile Range (IQR) =
Upper Quartile – Lower Quartile

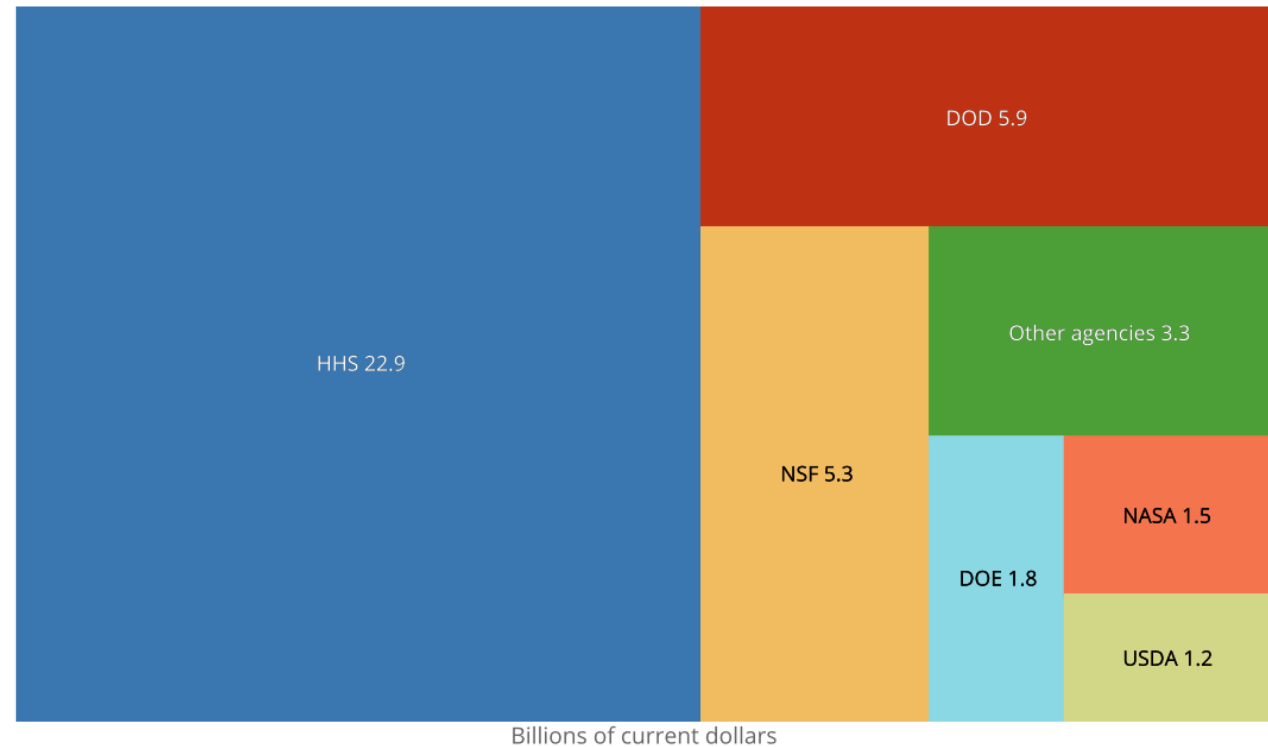


Tree Maps ?

Q3 “We used tree maps in this report, partly to break up the monotony of bar chart after bar chart (if you look at the report’s other figures). In the forthcoming report, we’re planning to use bar charts instead anyway. Which is better? Does it matter? Are there different ways to display data like these?”

FIGURE 5B-7

Federally financed academic R&D expenditures, by agency: FY 2018



DOD = Department of Defense; DOE = Department of Energy; HHS = Department of Health and Human Services; NASA = National Aeronautics and Space Administration; NSF = National Science Foundation; USDA = Department of Agriculture.

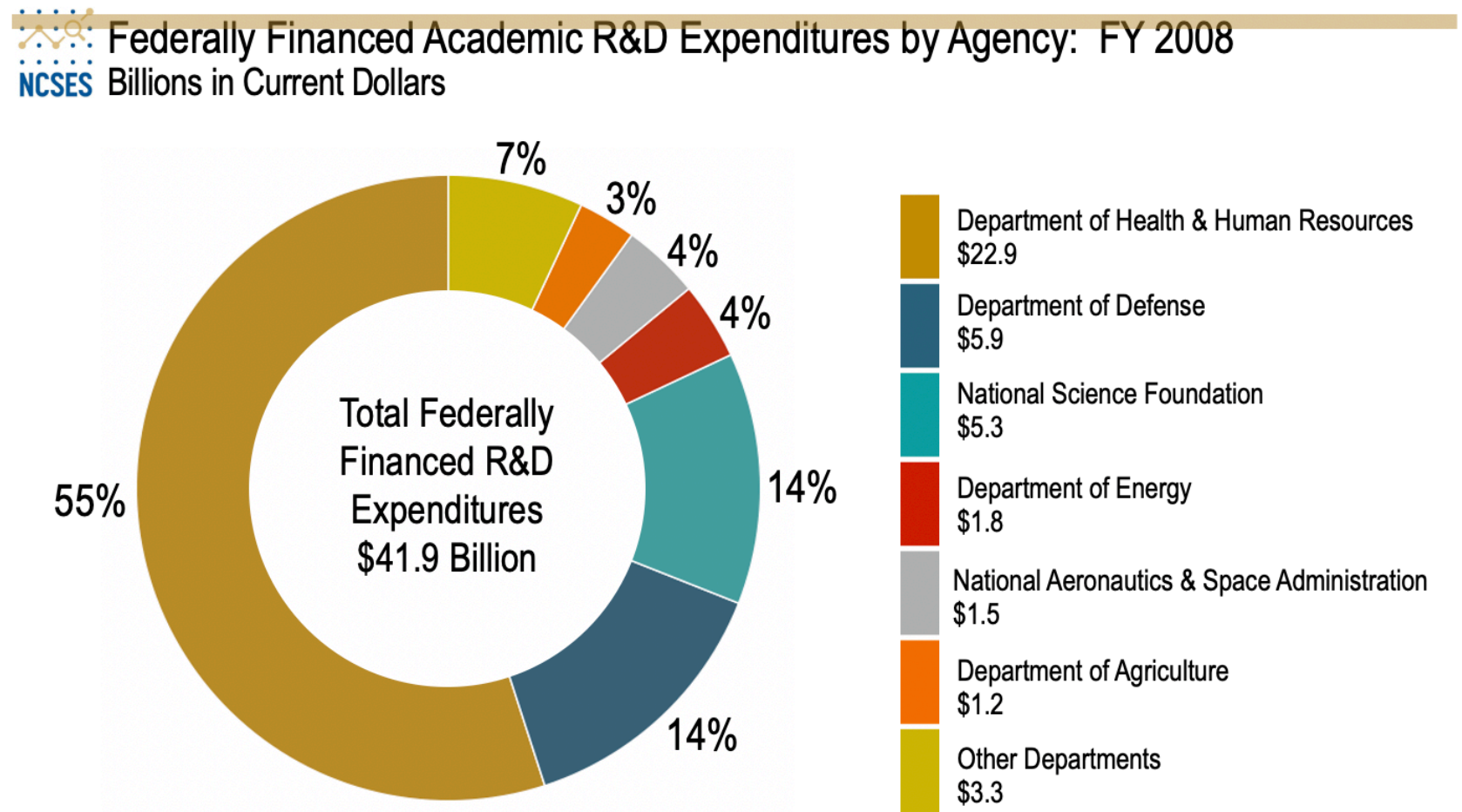
Source(s)

National Center for Science and Engineering Statistics, National Science Foundation, Higher Education Research and Development Survey (HERD), FY 2018.

Science and Engineering Indicators

Tree Maps ?

Q3 “We used tree maps in this report, partly to break up the monotony of bar chart after bar chart (if you look at the report’s other figures). In the forthcoming report, we’re planning to use bar charts instead anyway. Which is better? Does it matter? Are there different ways to display data like these?”



“One Pagers”

Q4 “NCSES has started to release “one-pagers.” The two so far can be found here:

<https://www.nsf.gov/statistics/2020/nsf20315/> and <https://www.nsf.gov/statistics/2020/nsf20304/>.

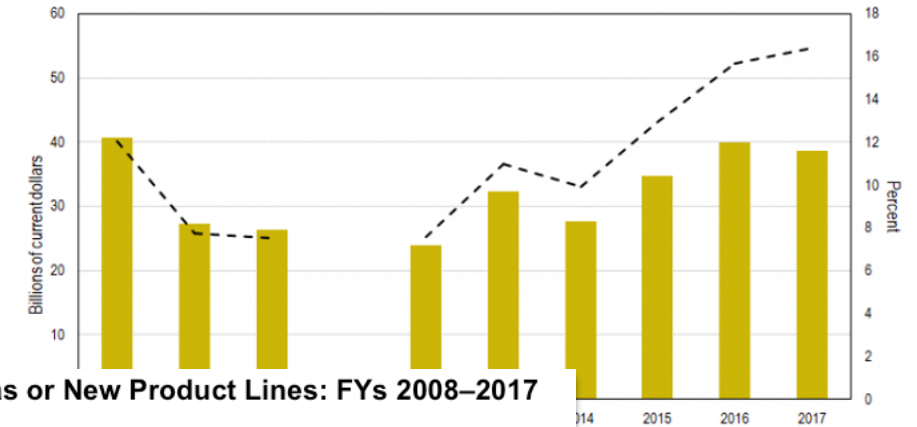
They have one graphic and a short amount of text. Any thoughts?”

Put annotation inside the figure to make it easier and quicker to interpret.

U.S. Businesses Invest \$55 Billion in R&D Directed at New Business Areas and New Product Lines

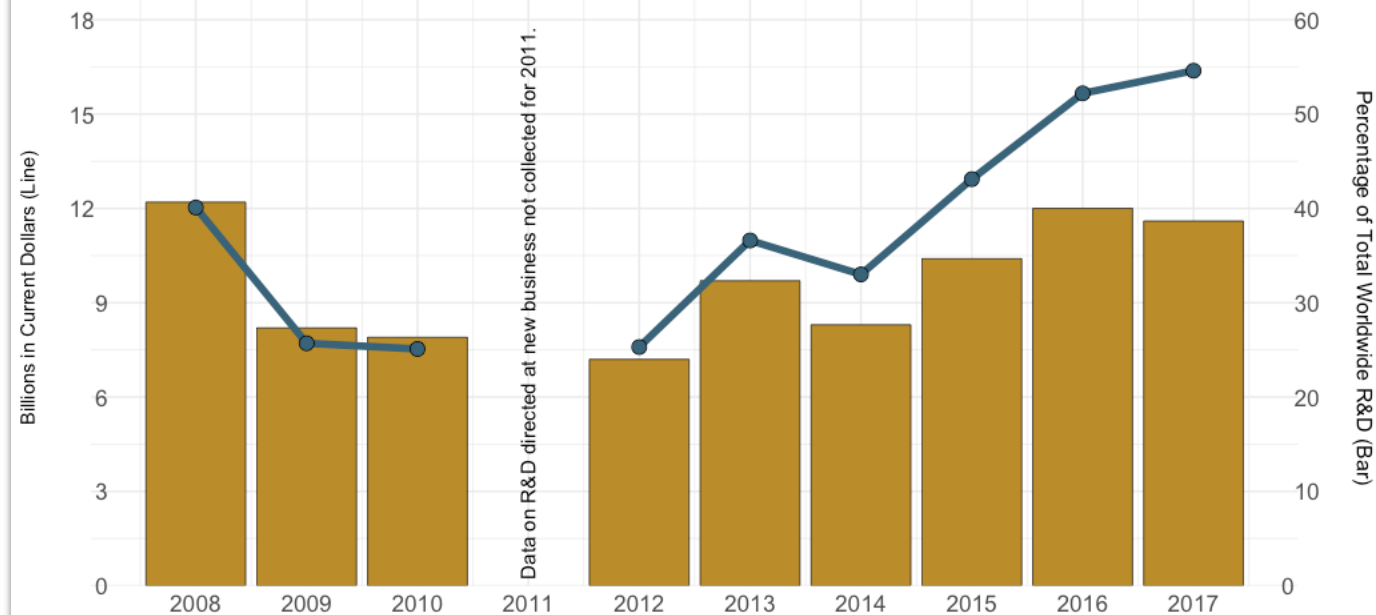
FIGURE 1

R&D directed at new business areas: 2008–17



Worldwide R&D Directed Toward New Business Areas or New Product Lines: FYs 2008–2017

NCSES: Business Research and Development Survey



SOURCE: 2017 <https://nces.nsf.gov/pubs/nsf20311/>; 2010, 2015–16 <https://www.nsf.gov/statistics/industry/>; 2008–09, 2012, data are available in the survey's Methodology Report, available from the author

“One Pagers”

Q4 “NCSES has started to release “one-pagers.”

The two so far can be found here:

<https://www.nsf.gov/statistics/2020/nsf20315/> and <https://www.nsf.gov/statistics/2020/nsf20304/>.

They have one graphic and a short amount of text. Any thoughts?”

Put annotation inside the figure to make it easier and quicker to interpret.

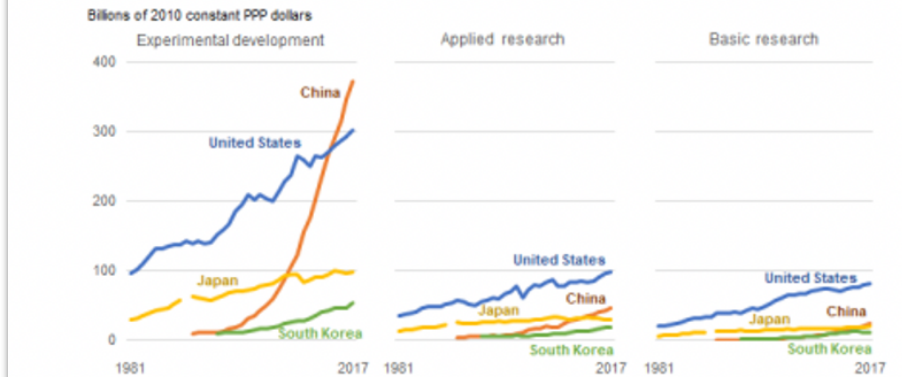
For example,

1. take the definitions and place them under the plot titles;
2. express the definition of total R&D spending as an equation and place it within the plot;
3. define PPP that is in the vertical axis title and place the axis title on the left side of the plot.

The United States Invests More in Applied and Basic Research than Any Other Country but Invests Less in Experimental Development than China

FIGURE 1

China leads the world in experimental development spending



PPP = purchasing power parity.

Note(s)

Data not available for China before 1991, South Korea before 1995, and Japan in 1990.

Source(s)

OECD Research and Development Statistics, Gross domestic expenditure on R&D by sector of performance and source of funds (https://stats.oecd.org/Index.aspx?DataSetCode=GERD_TORD, accessed 23 August 2019) and NCSES, National Patterns of R&D Resources (annual series).

Although total R&D spending in 2017 by the United States (\$483 billion, enumerated in constant 2010 purchasing power parity dollars) exceeds China's R&D expenditures (\$443 billion), China's annual investment in experimental development surpasses that of the United States. China's spending on experimental development has grown rapidly in recent years to over \$370 billion in 2017, nearly \$70 billion greater than the United States.

U.S. spending on applied research has grown steadily since the early 1980s, rising to nearly \$100 billion in 2017. U.S. expenditures on basic research have shown a similar trend and were just below \$85 billion in 2017.

Experimental development is systematic work that draws on knowledge gained from research and practical experience and is directed toward producing new or improving existing products or processes. Applied research is original investigation undertaken to acquire new knowledge that is directed primarily toward a specific, practical aim or objective. Basic research is experimental or theoretical work undertaken primarily to acquire new knowledge of the underlying foundations of phenomena and observable facts without any particular application or use in view. Like basic and applied research, experimental development generates additional knowledge. Total R&D expenditures are the sum of spending on experimental development, applied research, and basic research.

Data are from the National Center for Science and Engineering Statistics (NCSES), National Patterns of R&D Resources, and from the Organization for Economic Cooperation and Development (OECD).

Vipin Arora, NCSES.

December 2019

Visualization Dictionary

Q5 Would it be helpful to be able to see visualization examples based on the variables you are displaying? For example, to address questions like:

“I want to show share of master’s degrees awarded to students on temporary visas by field over time, so that fits Category X in the report.”

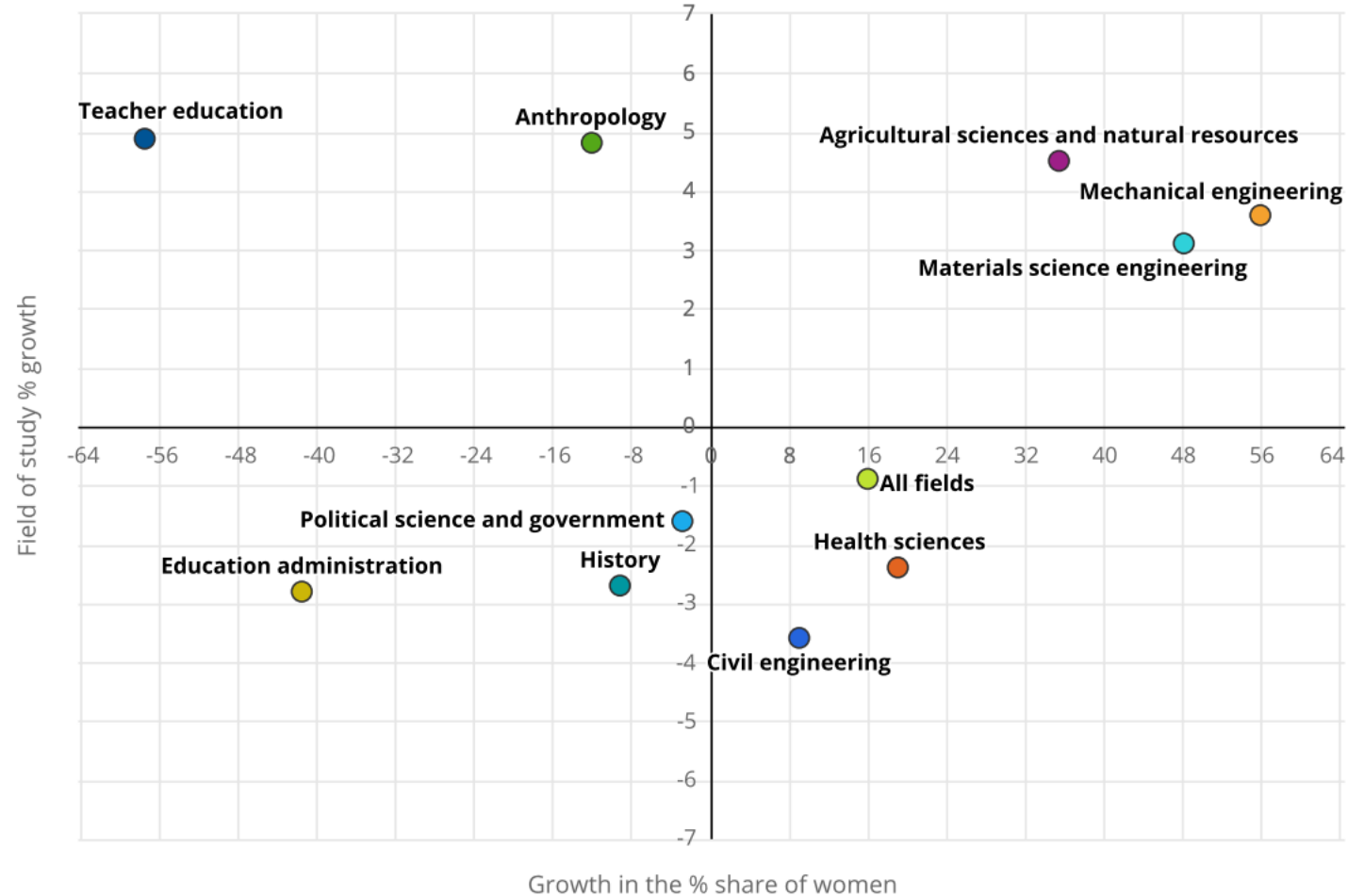
| Discrete | Continuous | Spatial |
|------------------|-------------|------------------------|
| Gender | Time | State |
| Race/Ethnicity | Square Feet | County |
| Citizenship | Dollars | Congressional District |
| Occupation | Percentage | Country |
| Federal Agency | | |
| Industry | | |
| Institution Type | | |
| | | |
| | | |
| | | |

SED Digest Report

Q5 “It would be great if they can look at the SED Digest report. The 2018 is on the website and the 2019 Digest will be released in 3 weeks Figure 12 in the upcoming Digest is new and it would be very helpful to get feedback on it from them because this is a topic that we cover every year.”

Figure 12

Fastest changing fields of study for female doctorate recipients and rates of change: 2010–19

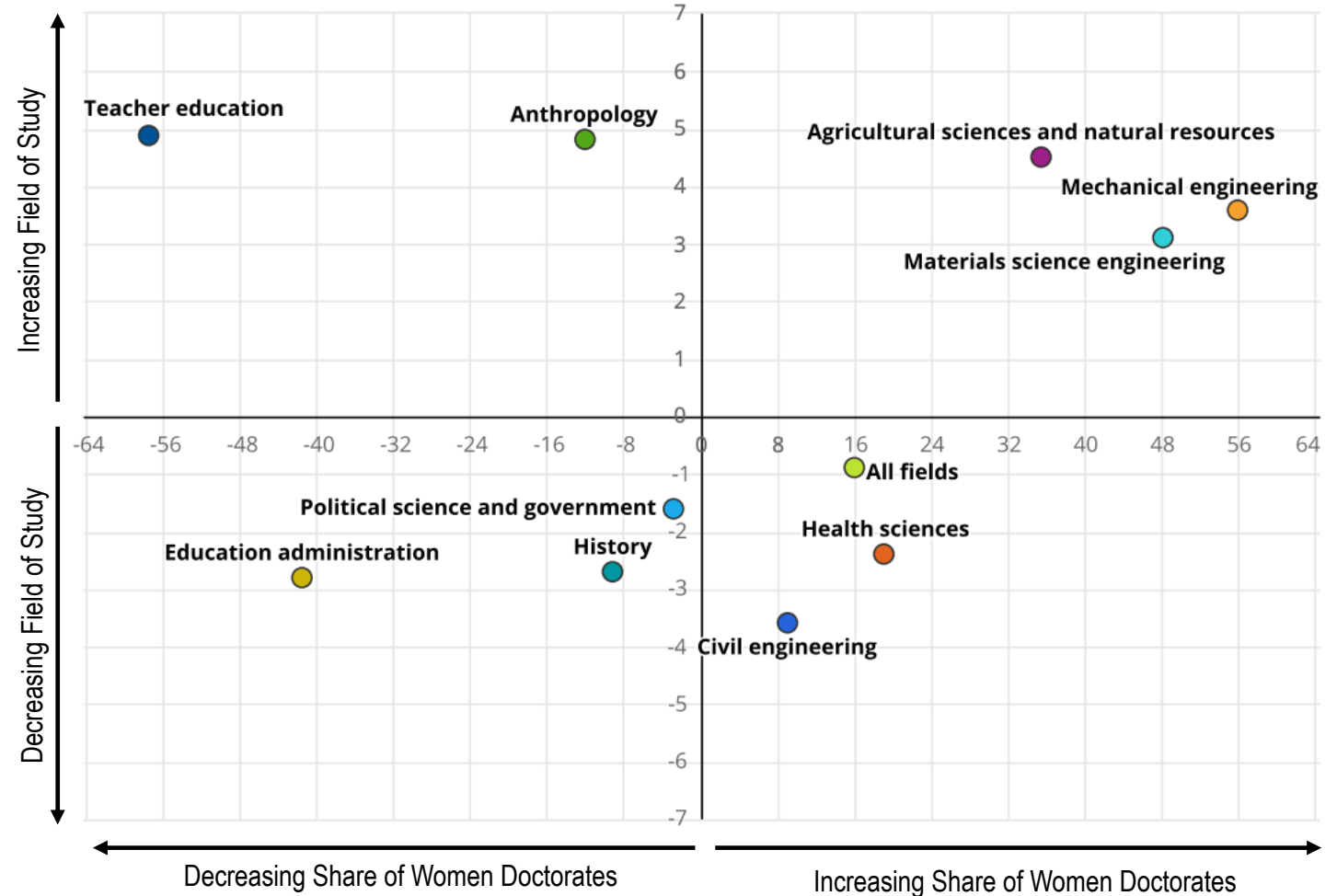


SED Digest Report

Q5 “It would be great if they can look at the SED Digest report. The 2018 is on the website and the 2019 Digest will be released in 3 weeks Figure 12 in the upcoming Digest is new and it would be very helpful to get feedback on it from them because this is a topic that we cover every year.”

Figure 12.

Percentage change in fields of study and female doctorates recipients between 2010 and 2019

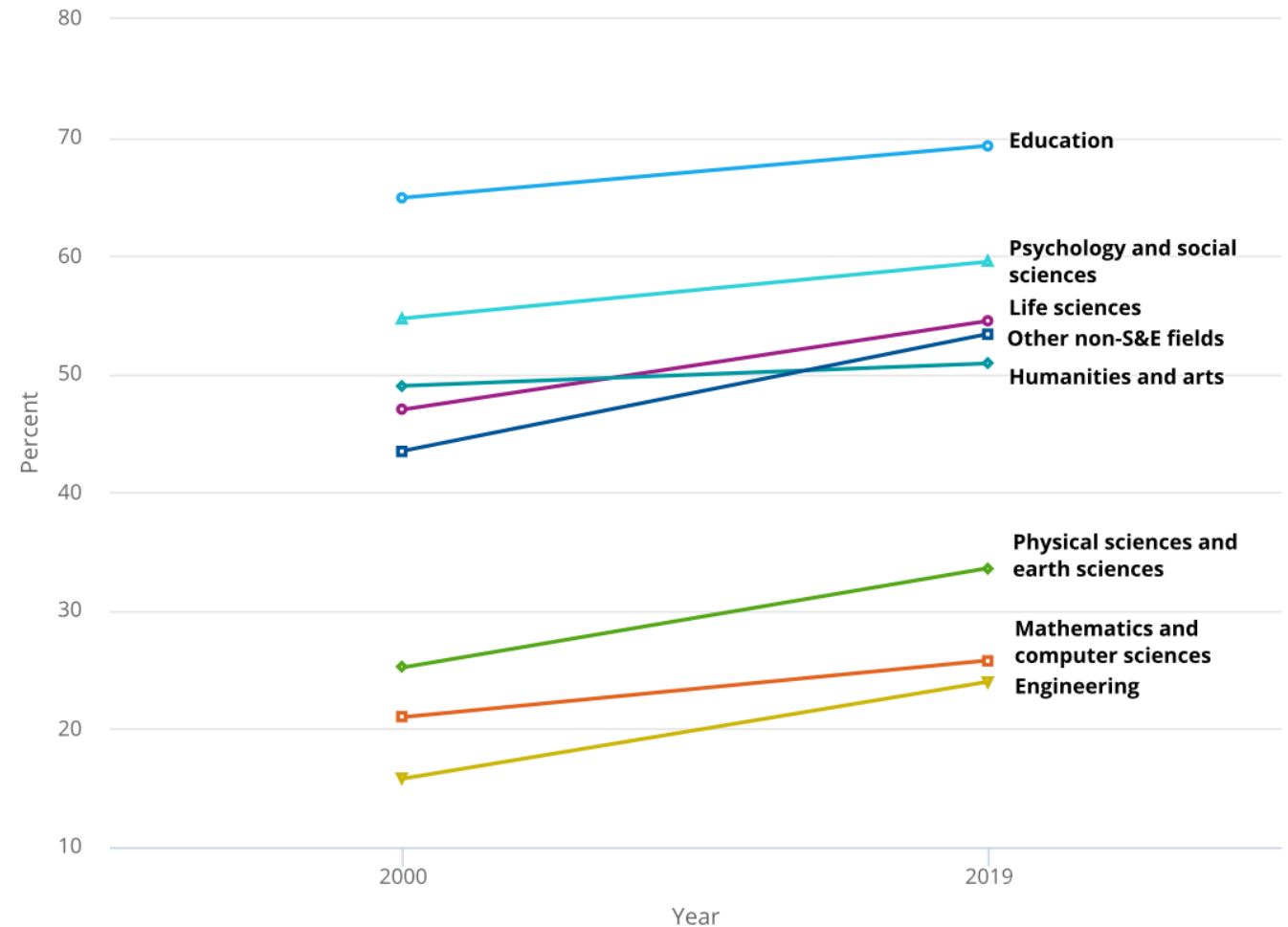


SED Digest Report

Q5 *“It would be great if they can look at the SED Digest report. The 2018 is on the website and the 2019 Digest will be released in 3 weeks Figure 12 in the upcoming Digest is new and it would be very helpful to get feedback on it from them because this is a topic that we cover every year.”*

Figure 11

Doctorates awarded to women, by broad field of study: 2000 and 2019

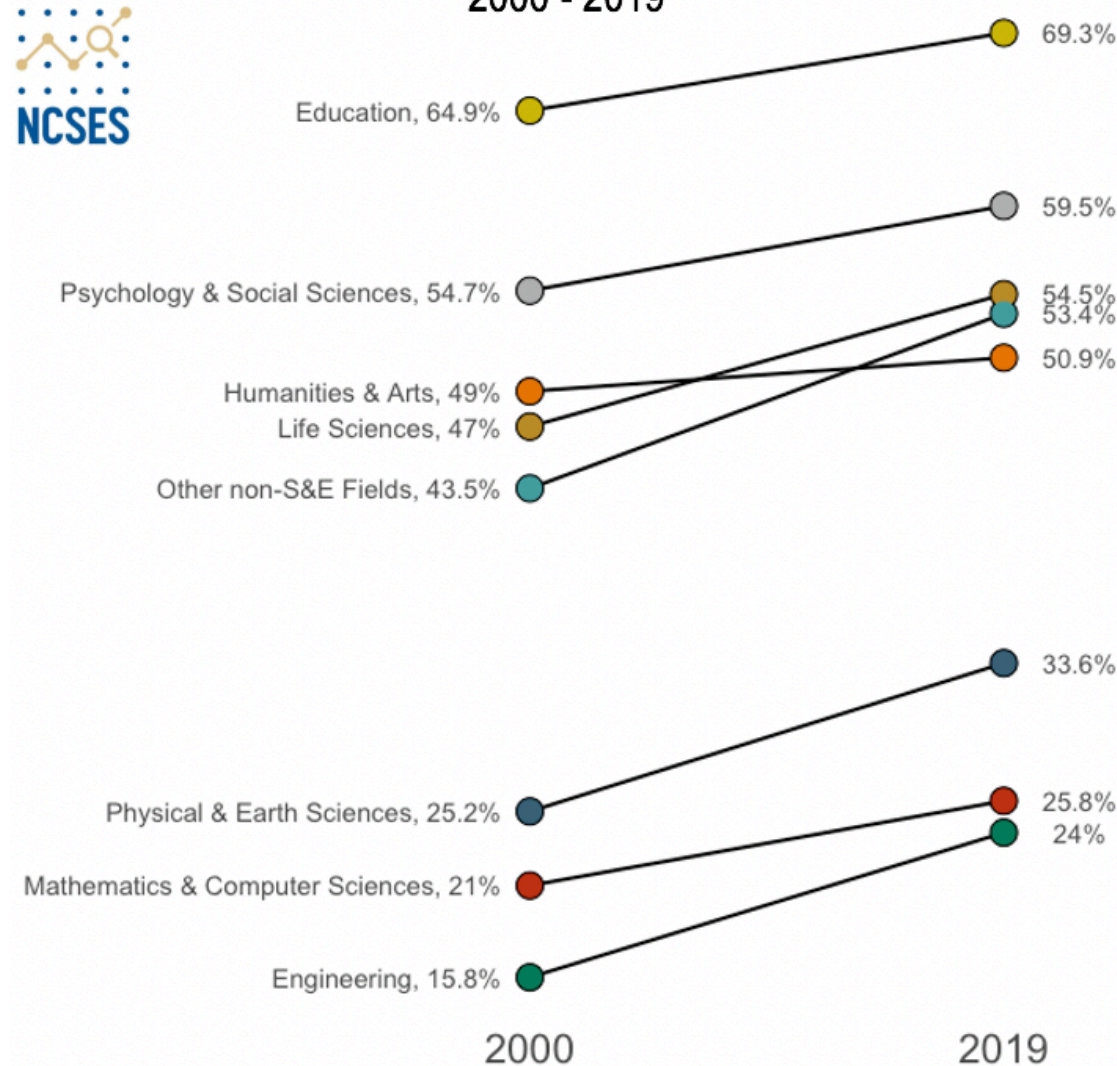


<https://nces.nsf.gov/pubs/nsf21308/report/fields-of-study#women>

SED Digest Report

Q5 “It would be great if they can look at the SED Digest report. The 2018 is on the website and the 2019 Digest will be released in 3 weeks Figure 12 in the upcoming Digest is new and it would be very helpful to get feedback on it from them because this is a topic that we cover every year.”

Doctorates Awarded to Women by Broad Field of Study
2000 - 2019



Getting rid of some of the white space and including the percentages for both years.



CONTACT:

VICKI LANCASTER

Social & Decision Analytics Division
Biocomplexity Institute & Initiative | University of Virginia
Suite 2910 | 1100 Wilson Blvd | Arlington, VA 22209
val7zv@virginia.edu