



## **Review of Visualizations in NCSES Publications based on Population Surveys**

**Vicki Lancaster, Ph.D.**  
**Principal Scientist**  
**Val7zv@virginialedu**

**University of Virginia**  
**Biocomplexity Institute and Initiative**  
**Social and Decision Analytics Division**

**April 10, 2019**

# Introduction

This is the first of a series of three papers, the first two provide comments and suggestions on the visualizations appearing in NCSES publications and provide alternative ways to display the data; the third is a survey of the visualizations in other federal government statistical agency publications.

This first paper reviews the visualizations in NCSES publications that report on surveys administered to an entire population. In these publications there is no need to display the variability of the data summaries displayed in the figures (i.e., means). Population surveys include: Earned Doctorates (SED), Graduate Students and Postdoctorates in Science and Engineering (GSS), Higher Education Research and Development (HERD), State Government Research and Development, FFRDC Research and Development, Science and Engineering Research Facilities, Federal Science and Engineering Support to Universities, and Federal Funds for Research and Development. Publications based on these surveys include numerous InfoBriefs and the Special Reports, Doctorate Recipients from U.S. Universities and Women, Minorities, and Persons with Disabilities in Science and Engineering.

All three papers are organized by topic defined by the number of variables being displayed and the measurement level of the variables, or whether the visualization is geospatial, displays a difference between two variables, or has time on the horizontal axis. Within a topic area, the figures outlined in orange are from NCSES publications and the figures outlined in grey provide an alternative way to display the same data. The alternative visualizations presented in the paper have not been endorsed or approved by NCSES for use in NCSES publications.

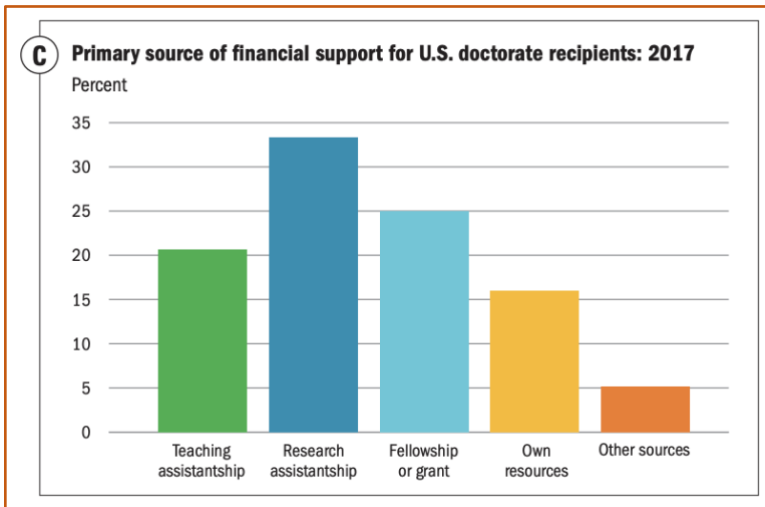
## Topics

1. Graphs with 1-categorical variable
2. Graphs with 2-categorical variables
3. Graphs with  $\geq 3$ -categorical variables
4. Graphs with Time on the Horizontal Axis
5. Graphs that Display a Difference
6. Geospatial Graphs

*This project has been funded, either wholly or in part, with federal funds from the National Center for Science and Engineering Statistics (NCSES) within the National Science Foundation (NSF) under cooperative support agreement number NCSE-1641251. The contents of this publication do not necessarily reflect the views or policies of NCSES or NSF, nor does mention of trade names, commercial products, or organizations imply endorsement of same by the U.S. Government.*

# 1. Graphs with 1-categorical variable

## Special Report: 2017 Doctorate Recipients from U.S. Universities<sup>1</sup>



### Bar Chart

One Variable:

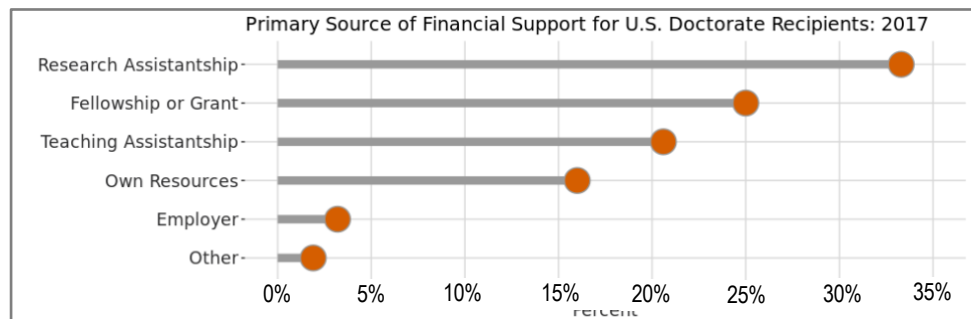
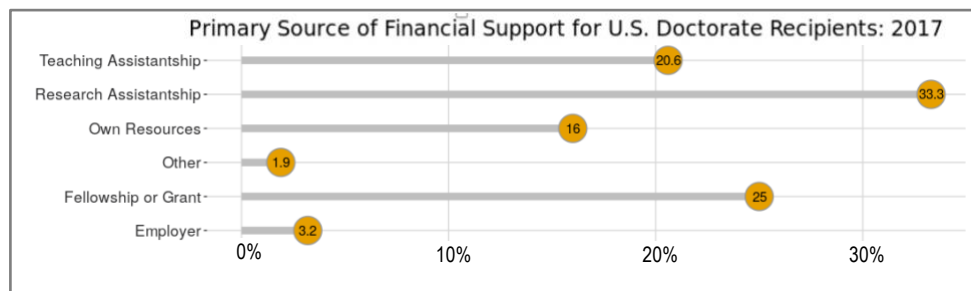
> Nominal: Doctorate financial support (5 levels)

### COMMENTS:

> This a lot of ink for just five numbers, which could easily be displayed in a table. Suggestion: since the axis labels are long and one category is missing (Employer) flip the coordinates and display the data using a [lollipop chart](#).<sup>2</sup>

The data are displayed below using lollipop charts.

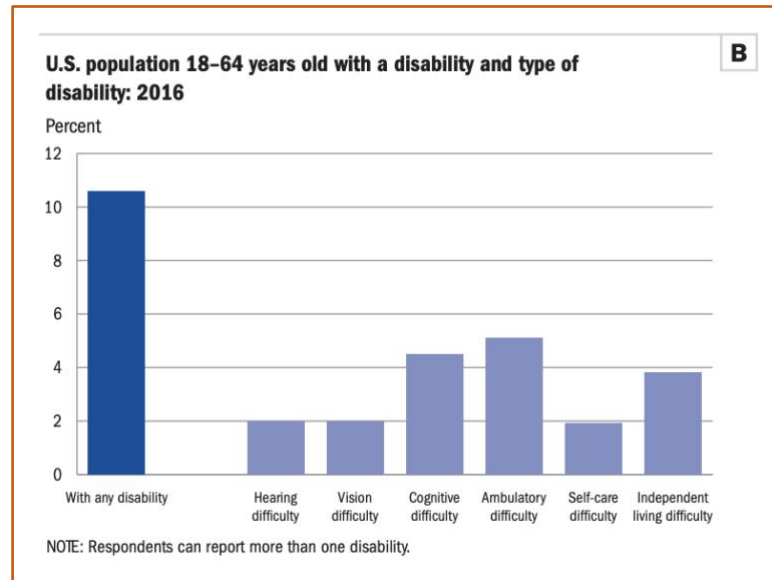
There are a number of options on how the levels of the variable can be arranged. In the bottom figure the levels are ordered by the percentage, largest to smallest. They can also be arranged by a second variable that is not visualized. The actual percentages can be included in the display (top figure) or not (bottom figure).



<sup>1</sup> 2017 Doctorate Recipients from U.S Universities <https://nces.nsf.gov/pubs/nsf19301/> (Accessed on May 5, 2019)

<sup>2</sup> ferdio <https://datavizproject.com/data-type/lollipop-chart/> (Accessed on May 5, 2019)

## Special Report: 2019 Women, Minorities, and Persons with Disabilities in Science and Engineering



### Bar Chart

One Variable:

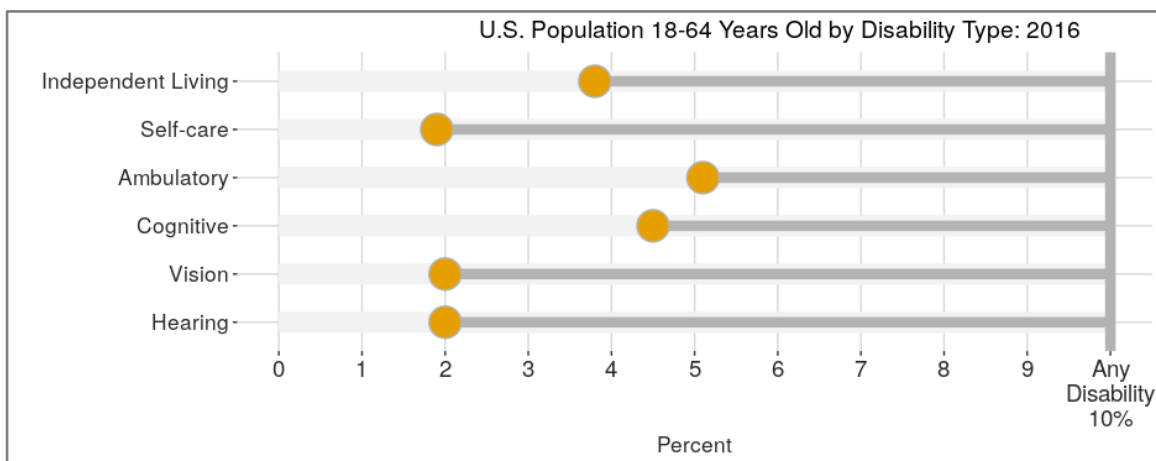
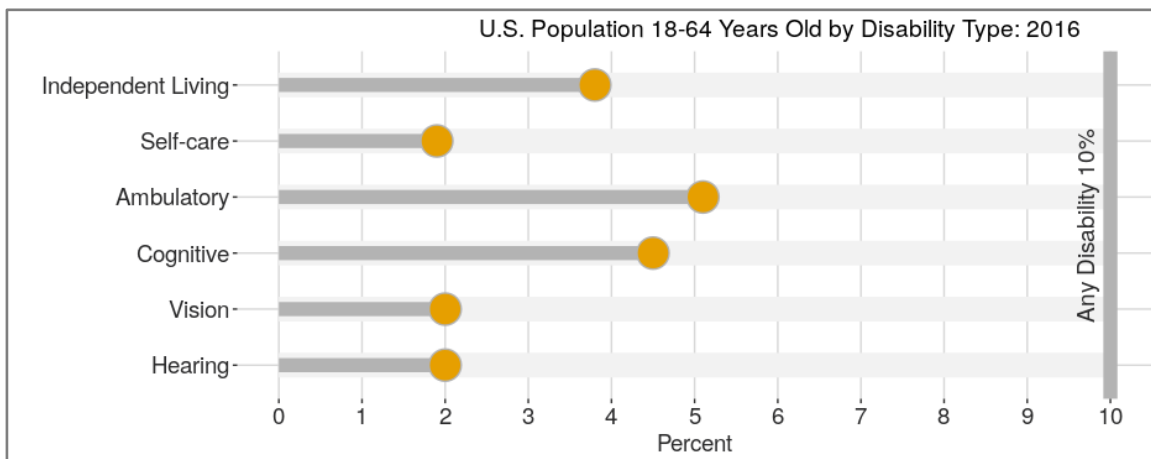
> Nominal: Type of disability (6 levels)

> One reference value: Any disability

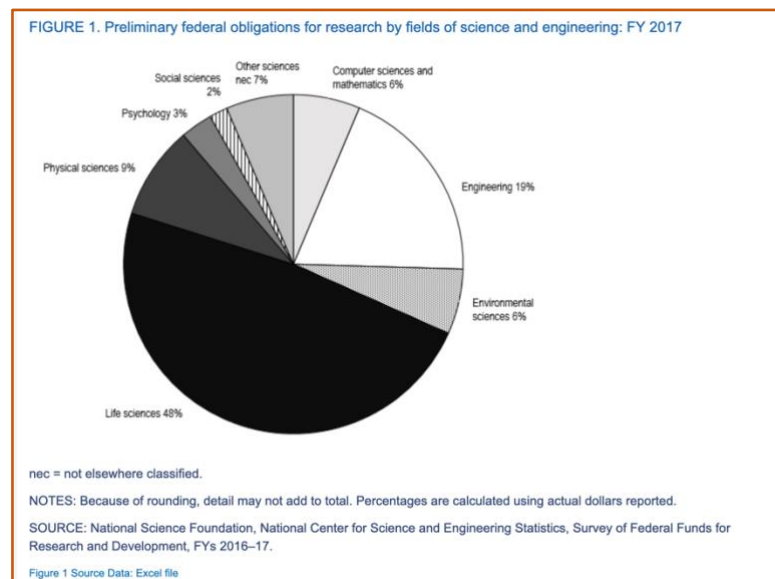
### COMMENTS:

> A bar chart that includes a reference value is common among NCSES figures. A suggestion for an alternative visualization is to flip the coordinates and display the data using a [lollipop chart](#) and highlight the reference value using an axis grid line.

The data are displayed below using lollipop charts with a reference line.



## Federal R&D Obligations Increase 3% in FY2017: Research Obligations Decrease Slightly While Those for Development Increase<sup>3</sup>



### Pie Chart

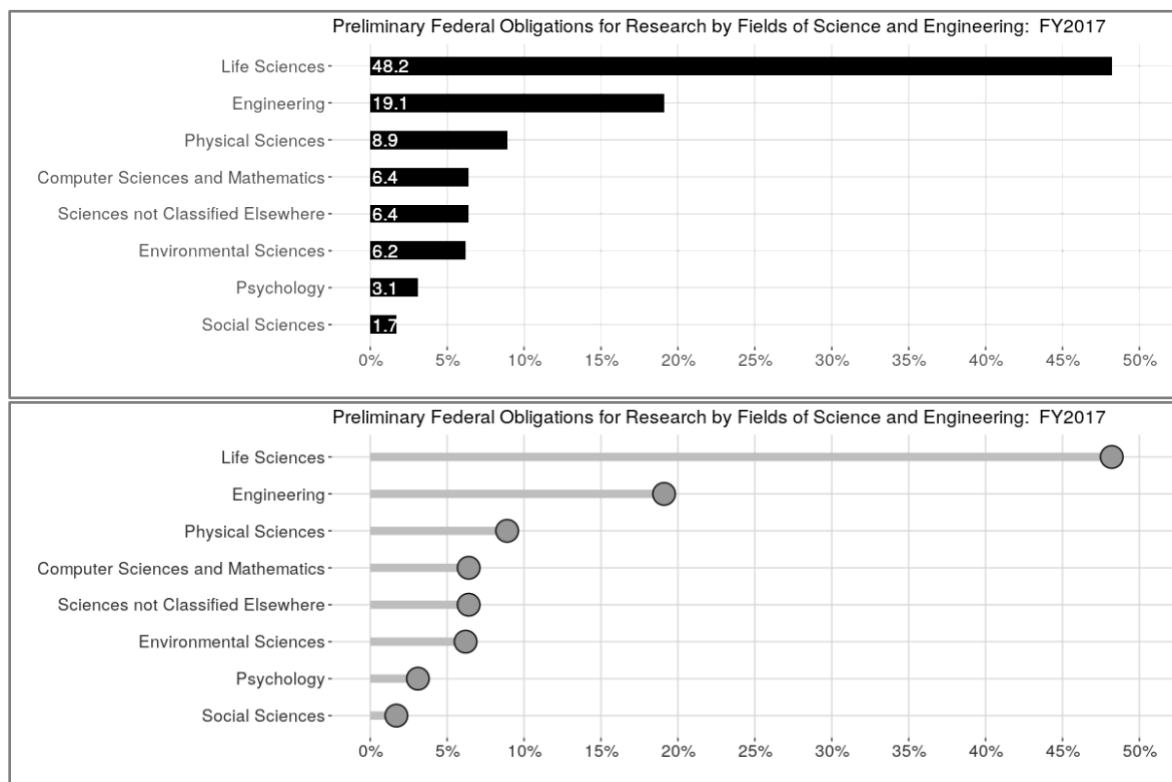
One Variable

> Nominal: Federal obligations to science and engineering fields (8 levels)

### COMMENTS:

> “Pie charts only make it easy to judge the magnitude of a slice when it is close to 0%, 25%, 50%, 75%, or 100%,”<sup>4</sup> but that is not the case in the figure on the left. The slices range from 2 to 48% and the lack of color makes distinguishing between the slices more difficult. Shades of grey are used to differentiate the slices with the exception of a single pattern for the social sciences; normally employing a different aesthetic would signal a different

variable. If color is not an option in InfoBriefs, then one suggestion is to use a bar chart with the bars ordered from largest to the smallest percentage and the value inserted within the bar (see the figure below).

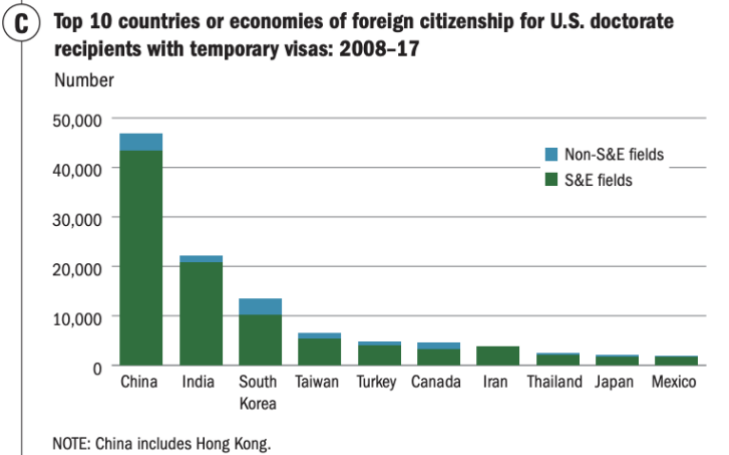


<sup>3</sup> InfoBrief July 30, 2018 <https://www.nsf.gov/statistics/2018/nsf18311> (Accessed on May 5, 2019)

<sup>4</sup> Stephen Few (2007) Save the Pies for Desert [https://www.perceptualedge.com/articles/visual\\_business\\_intelligence/save\\_the\\_pies\\_for\\_dessert.pdf](https://www.perceptualedge.com/articles/visual_business_intelligence/save_the_pies_for_dessert.pdf) (Accessed on May 5, 2019)

## 2. Graphs with 2-categorical variables

### Special Report: 2017 Doctorate Recipients from U.S. Universities<sup>5</sup>



#### Simple Stacked Bar Graph

Two Variables:

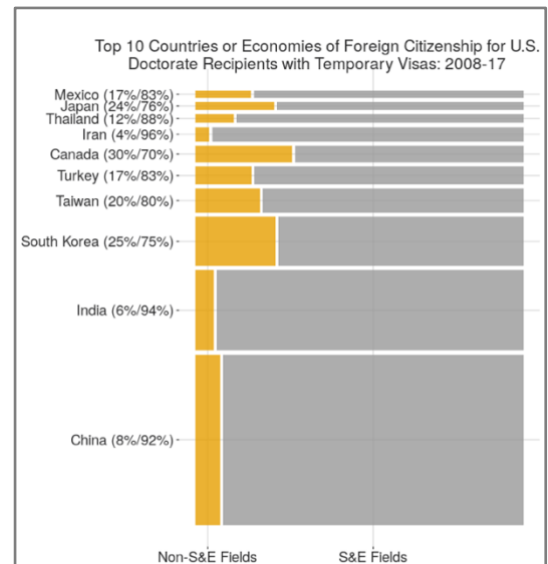
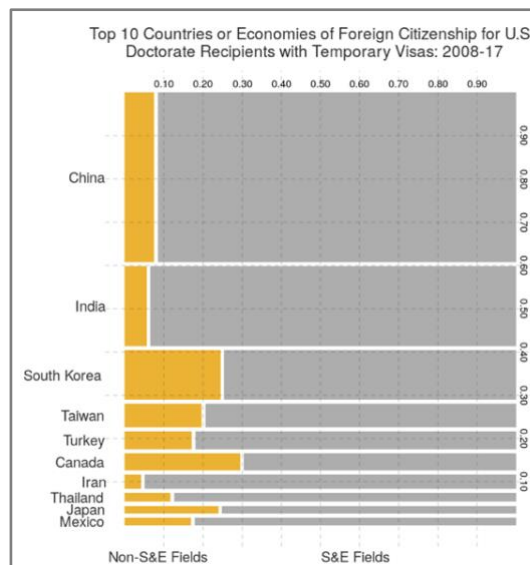
- > Nominal: S&E fields (2 levels)
- > Nominal: Country/Economy (10 levels)

#### COMMENTS:

> The large difference in the number of doctorates between China and Mexico (China has over 23x more doctorates than Mexico) make the graph hard to interpret for the smaller level, Non-S&E fields. A suggestion is to display the data using a [mosaic plot](#)<sup>6</sup> (see the figure below).

In a mosaic plot (also referred to as a marimekko plot), both axes are on a percentage scale that determines both the width and height of each segment. In this case the country or economy is on the vertical axis where it is easy to see that China, India, and South Korea make up more than half of the foreign doctorate recipients (as noted in the text of the report), but the reader can still see the contributions of the countries with fewer recipients. One can easily make comparison between non-S&E versus S&E fields within and between each country or economy which was not possible in the previous plot.

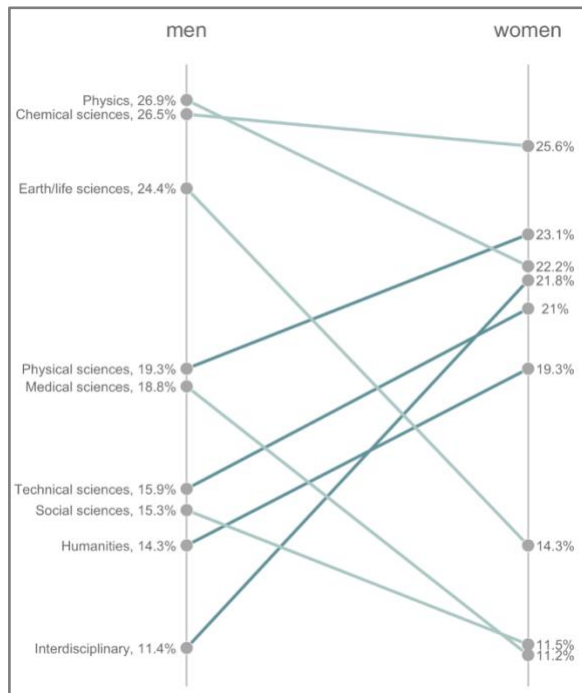
Other examples are given on page 5.



<sup>5</sup> 2017 Doctorate Recipients from U.S. Universities <https://nces.nsf.gov/pubs/nsf19301/> (Accessed on May 5, 2019)

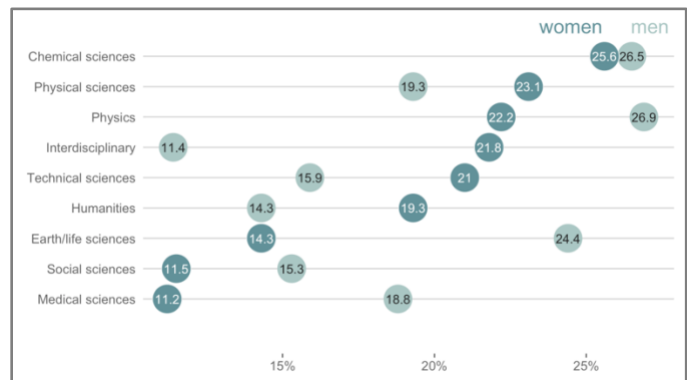
<sup>6</sup> Data Visualization Catalog [https://datavizcatalogue.com/methods/marimekko\\_chart.html](https://datavizcatalogue.com/methods/marimekko_chart.html) (Accessed on May 5, 2019)

## Alternative to the simple stacked bar graph displayed on page 4:<sup>7</sup>



SLOPE PLOT

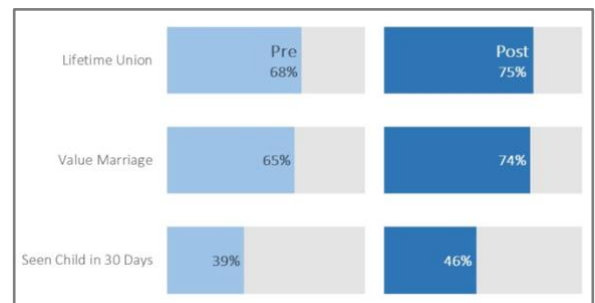
When one variable has only two levels there are more options, for example, a slope plot, horizontal dot plot, or side-by-side bar charts.



HORIZONTAL DOTS



PANEL CHART WITH CONTEXTUAL SHADING

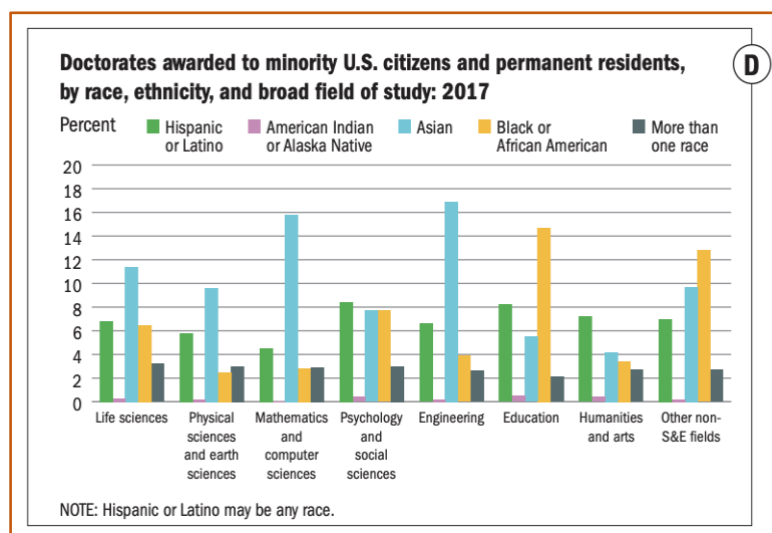


SIDE-BY-SIDE BAR CHARTS

<sup>7</sup> Depict Data Studio May 31, 2015, Alternative to the Clustered Bar Chart, <https://depictdatastudio.com/clustered-bar-chart/> (Accessed on May 5, 2019)



## Special Report: 2017 Doctorate Recipients from U.S. Universities<sup>8</sup>



### Multi-set Bar Chart

Two Variables:

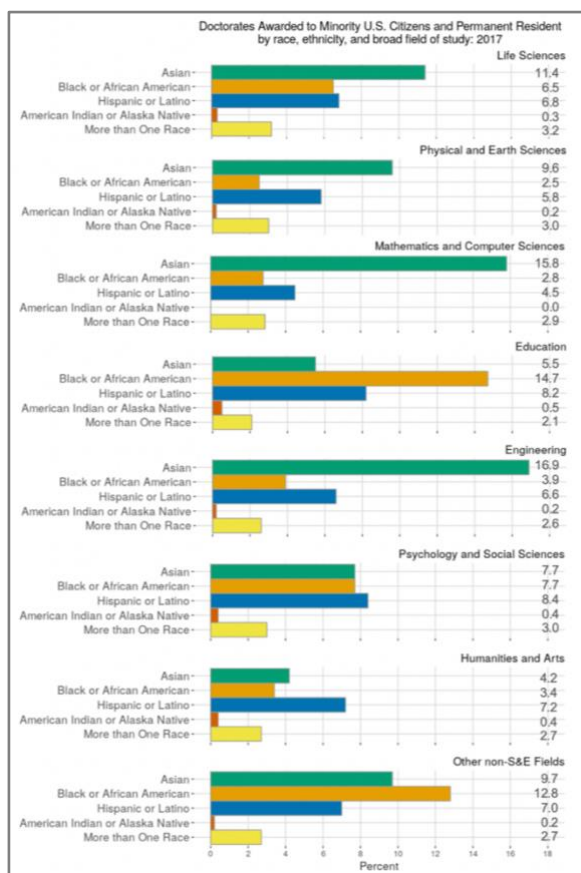
> Nominal: Broad field of study (8 levels)

> Nominal: Race/Ethnicity (5 levels)

### COMMENTS:

> Multi-set bar or grouped bar charts are difficult to interpret; it is difficult to make comparisons across the broad fields of study for a particular race/ethnicity, especially in the case of the minority group, American Indian or African Americans, where the bars are barely visible. One suggestion is to enlarge the figure, flip the coordinates, and separate the broad fields of study into stackable figures.

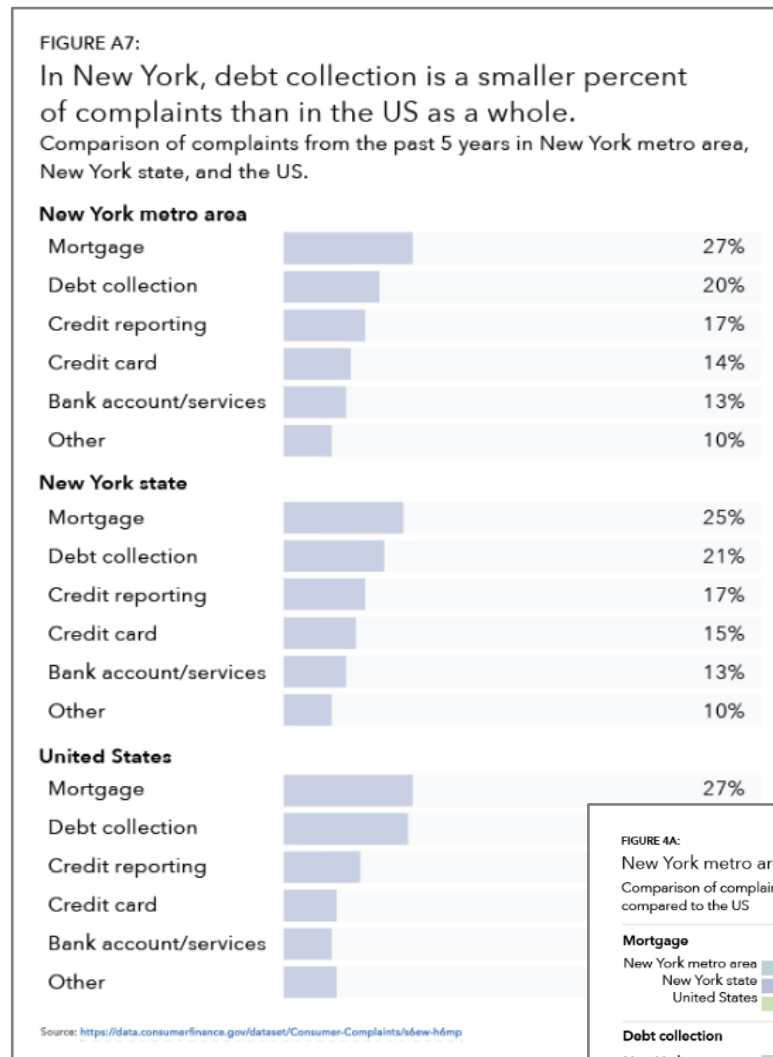
The data used in this figure are displayed on the right using stackable bar charts. Page 7 contains additional figures displaying 2-categorical variables with multiple levels.



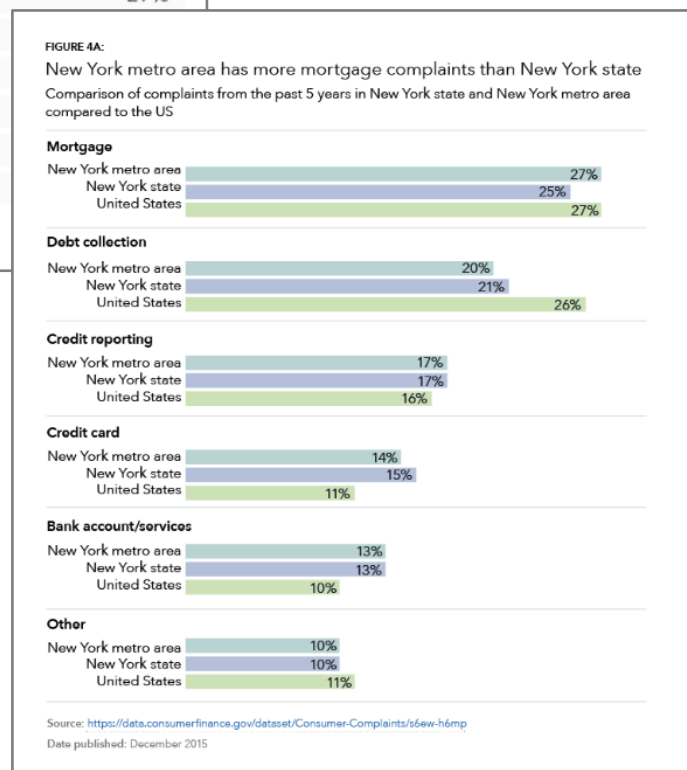
<sup>8</sup> 2017 Doctorate Recipients from U.S. Universities <https://nces.nsf.gov/pubs/nsf19301/> (Accessed on May 5, 2019)



## Alternative to the multi-set bar chart data displayed on page 6:<sup>9</sup>



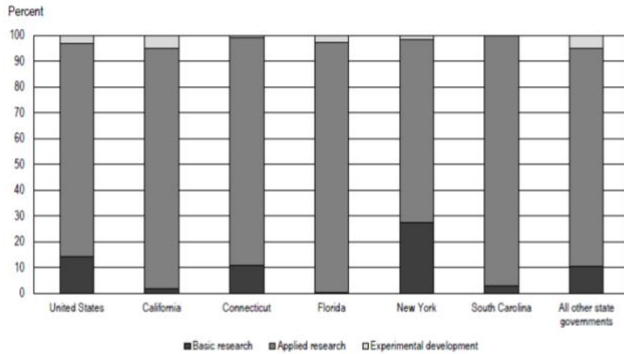
When there are more than two levels for each variable, flip the axis and use the levels of one variable to label the individual bar charts that are plotted one on top of each other. Adding the value to the bar chart makes for less mental energy to interpret the display. The figures on this page display the same data using two different configurations.



<sup>9</sup> [DATA VIZ] Bar Chart Bonanza!, March 3, 2016, <https://github.com/cfpb/design-manual/issues/401> (Accessed on May 5, 2019)

## InfoBriefs: State Government R&D Expenditures Increase 7% in FY 2017; Health-Related R&D Up 13%<sup>10</sup>

FIGURE 1. State agency intramural R&D, by type, for the five states with the highest level of intramural R&D: FY 2017



NOTES: U.S. total reflects all 50 states and the District of Columbia. Detail may not add to total because of rounding.

SOURCE: National Science Foundation, National Center for Science and Engineering Statistics, Survey of State Government Research and Development, FY 2017.

Figure 1 Source Data: Excel file

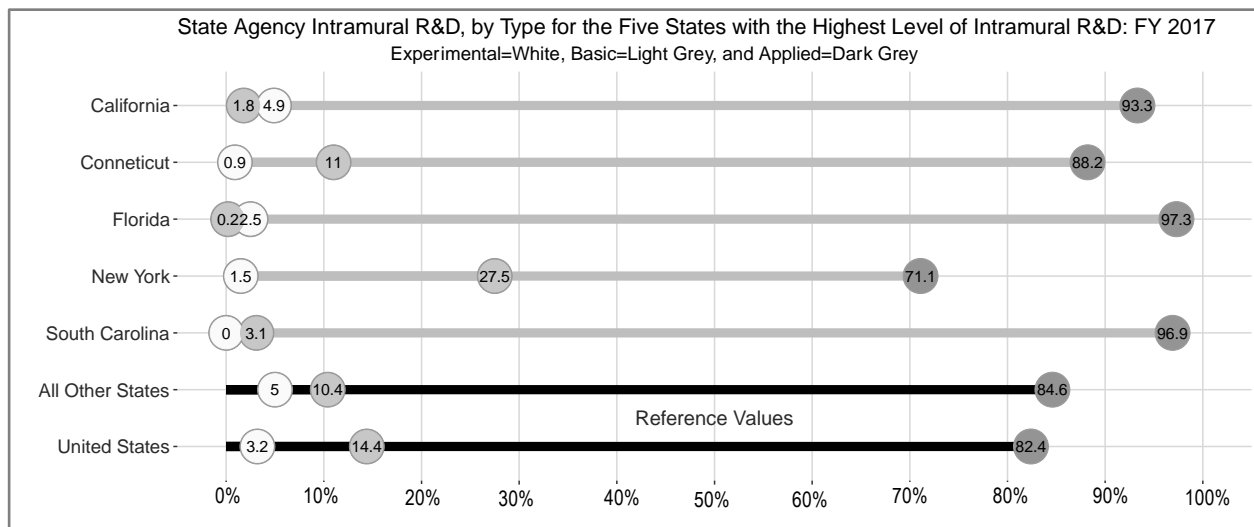
### Stacked Bar Chart

Two Variables:

- > Nominal: Types of research (3 levels)
- > Nominal: States (5 levels)
- > Two reference values: All other states and United States

### COMMENTS

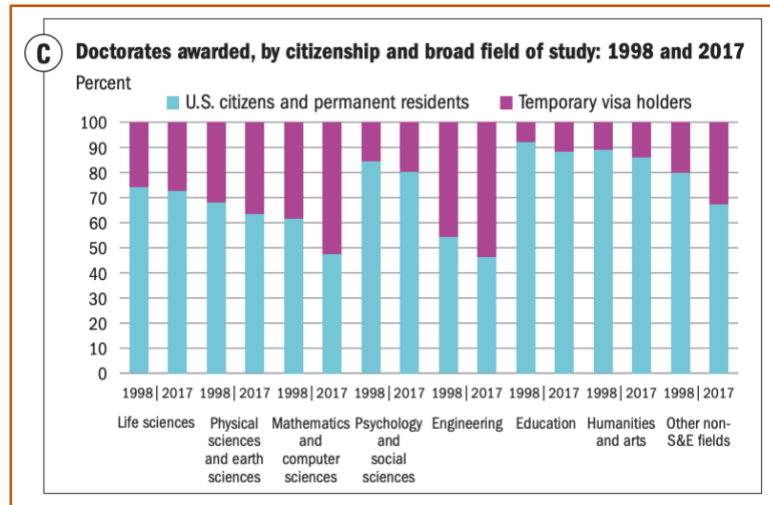
> It is difficult to see all the levels of the types of research. One suggestion is to use a horizontal dot plot and insert the values in the dots. The reference values are identified. It makes it easier to see that the five largest states all have basic research percentages smaller than the U.S. average and with the exception of New York they have larger applied research percentages.



<sup>10</sup> InfoBriefs December 19, 2019, <https://nsf.gov/statistics/2019/nsf19305/> (Accessed on May 5, 2019)

### 3. Graphs with $\geq 3$ -categorical variables

#### Special Report: 2017 Doctorate Recipients from U.S. Universities<sup>11</sup>



#### Multi-set Stacked Bar Graph

Three Variables:

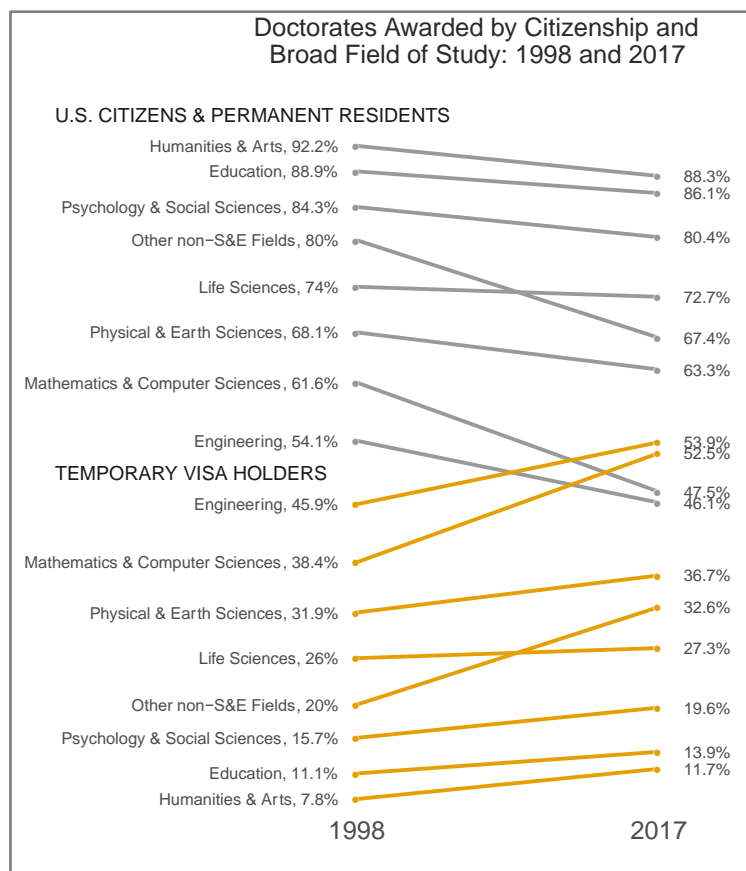
- > Nominal: Broad field of study (8 levels)
- > Nominal: Citizenship (2 levels)
- > Interval: Time (2 years)

#### COMMENTS:

> Two colors are being used to display three variables. The figure does not tell the story of the data without exerting a lot of mental energy. A suggestion is to display the data using a [slope plot](#)<sup>12</sup> (see below and on page 10).

The slope plot contains two axes, one for each level of time (1998 and 2017). The broad fields of study are each represented by a line connecting the percentages of doctorates awarded in 1998 and 2017.

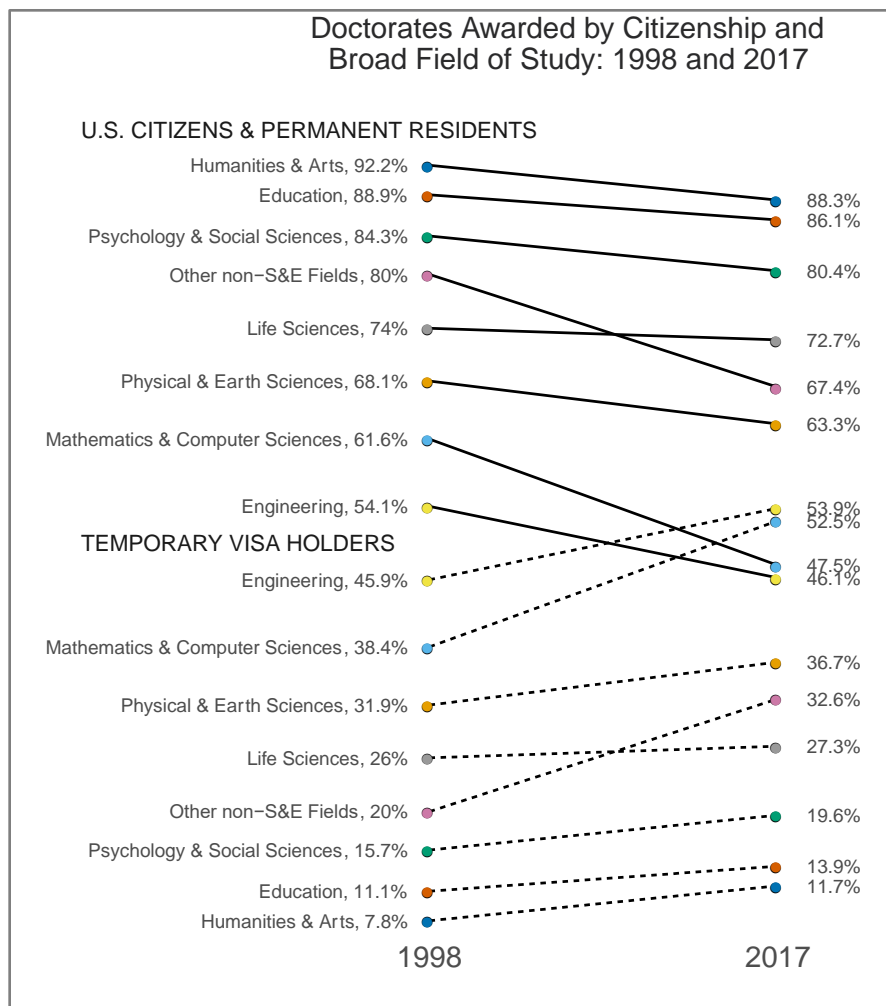
This plot makes it very clear that the percentage of doctorates for temporary visa holders is increasing (upward slope) while it is decreasing for U.S. citizens and permanent residents (downward slope). In the display the colored lines identify the two categories of citizenship and also upward and downward slopes.



<sup>11</sup> 2017 Doctorate Recipients from U.S. Universities <https://nces.nsf.gov/pubs/nfs19301/> (Accessed on May 5, 2019)

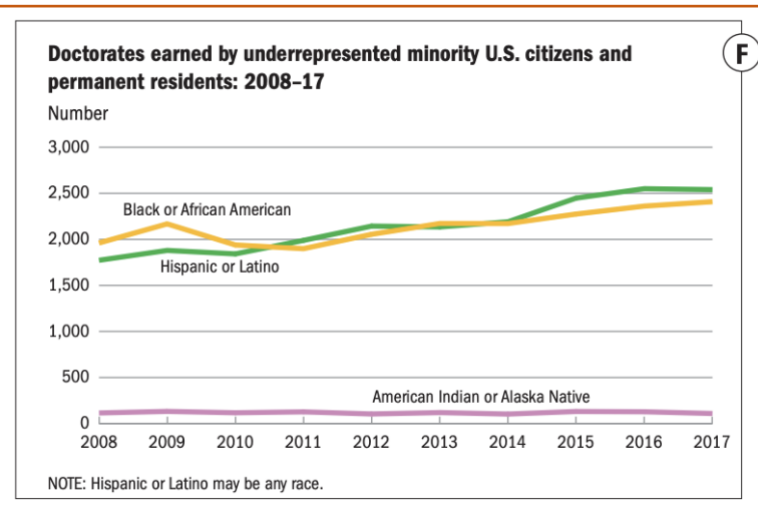
<sup>12</sup> Seeing Data <http://seeingdata.org/taketime/inside-the-chart-slope-graph/> (Accessed May 5, 2019)

In this slope plot color is used to identify the broad fields of study.



## 4. Graphs with Time on the Horizontal Axis

### Special Report: 2017 Doctorate Recipients from U.S. Universities<sup>13</sup>



#### Longitudinal (2008-2017) Line Graph

Two Variables:

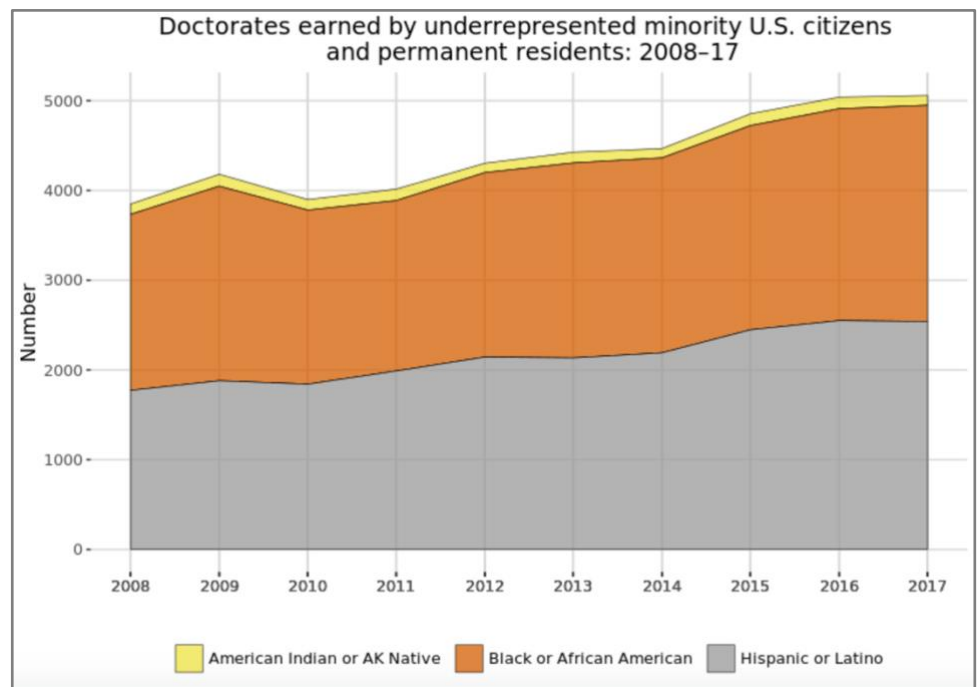
- > Nominal: Minority U.S. citizens (3 levels)
- > Interval: Time (10 years)

#### COMMENTS:

> It is easy to look at this figure and miss the American Indian or Alaska Native minority group, the numbers are so small compared to the other two minority groups. One way to give each of the three minority groups equal prominence is to use a [stacked area graph](#).<sup>14</sup>

The data used in this figure are displayed below using a stacked area graph.

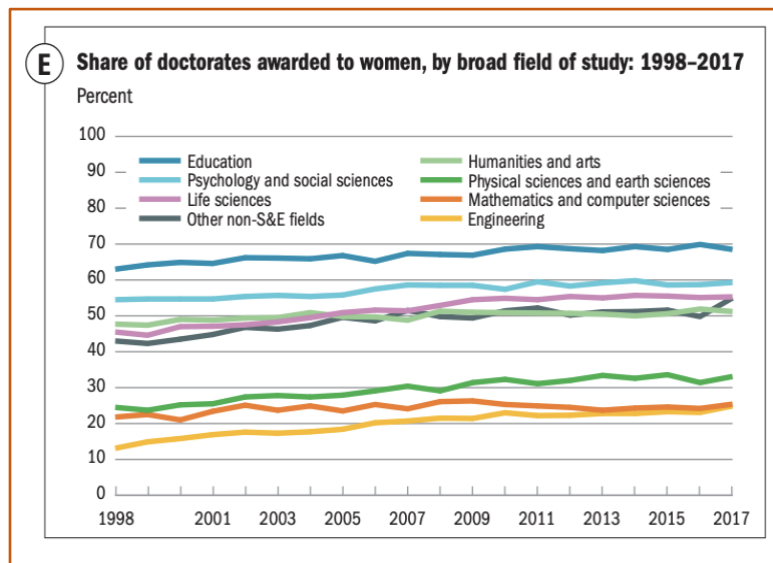
The graph displays the total number of U.S. doctorates over the years 2008 to 2017 as well as the relative proportions of each of the three minority groups. The graph displays the evolution of the whole, doctorate recipients for underrepresented minorities is increasing, and the relative proportions of the three minority groups, Black or African American and Hispanic or Latino groups are increasing whereas American Indian or Alaska Natives has remained the same. Stacked area graphs are not without their [critics](#)<sup>15</sup> but in this case it works and provides a clear picture of the differences in the three groups over time.



<sup>13</sup> 2017 Doctorate Recipients from U.S. Universities <https://nces.nsf.gov/pubs/nsf19301/> (Accessed on May 5, 2019)

<sup>14</sup> Data visualization Catalog [https://datavizcatalogue.com/methods/stacked\\_area\\_graph.html](https://datavizcatalogue.com/methods/stacked_area_graph.html) (Accessed on May 5, 2019)

<sup>15</sup> from Data to Viz, The issue with Stacking, <https://www.data-to-viz.com/caveat/stacking.html> (Accessed on May 5, 2019)



### Longitudinal (1998–2017) Line Graph

Two Variables:

> Nominal: Broad field of study (8 levels)

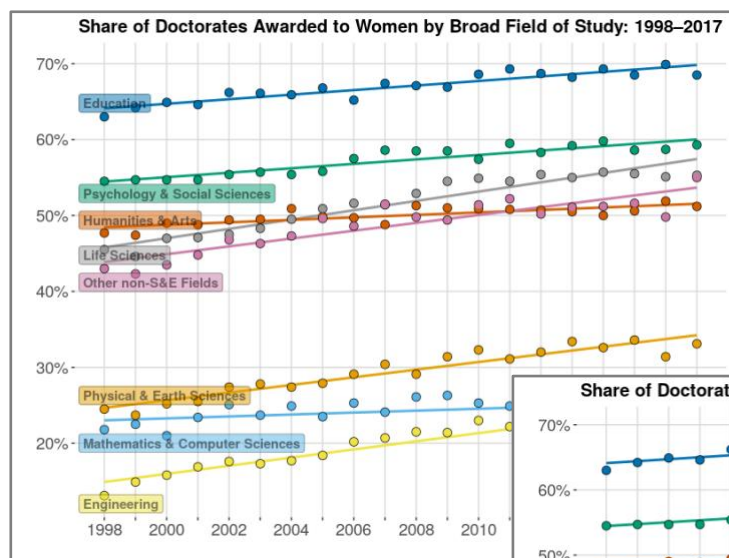
> Interval: Time (19 years)

### COMMENTS:

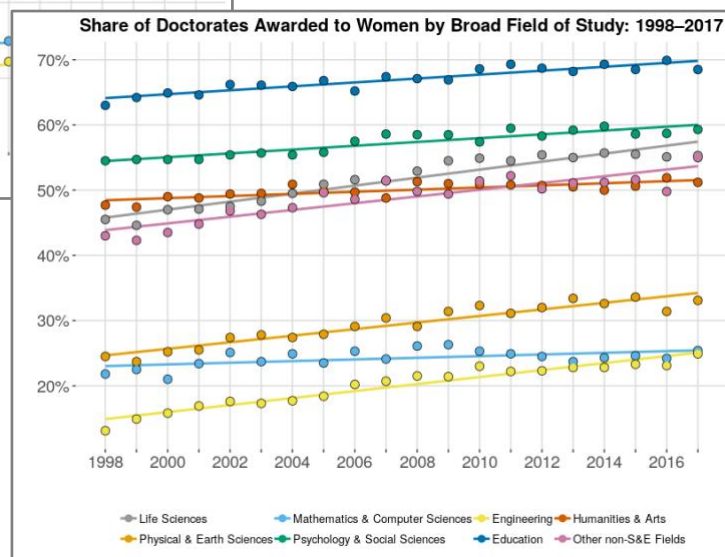
> In order to be 508 compliant use colorblind colors.

> Reducing the white space would make it easier to make comparisons between the fields of study over time. A suggestion is to change the vertical axis from [0–100] to [10–70], to reduce the non-information white space.

The data used in this figure are displayed below and on the next page.

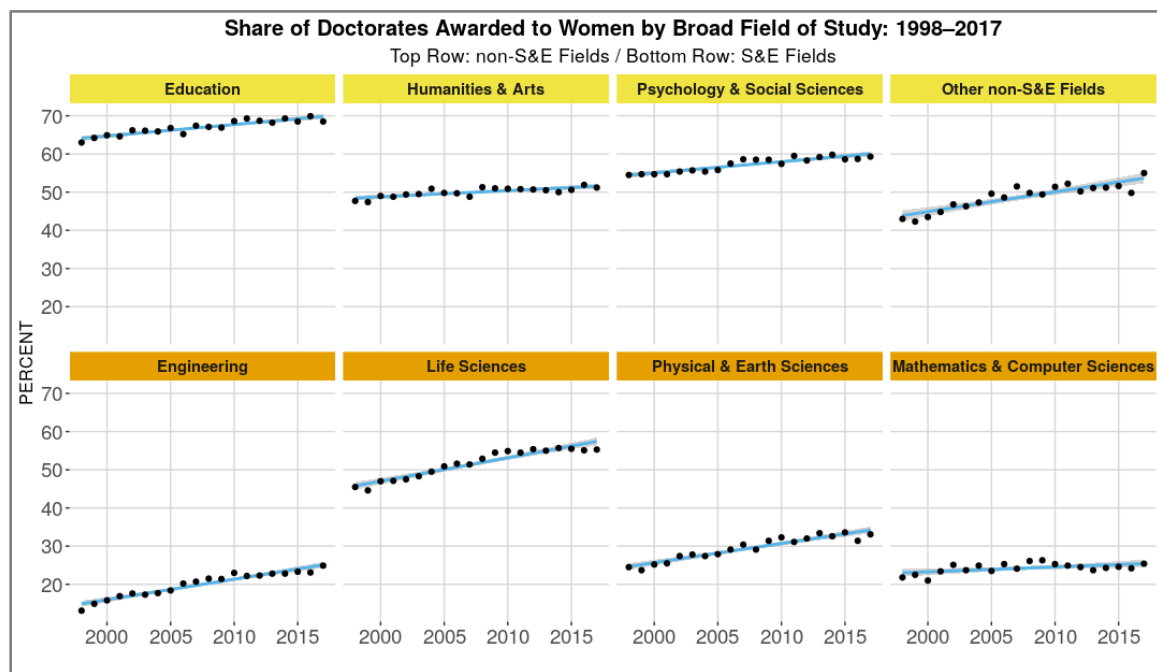
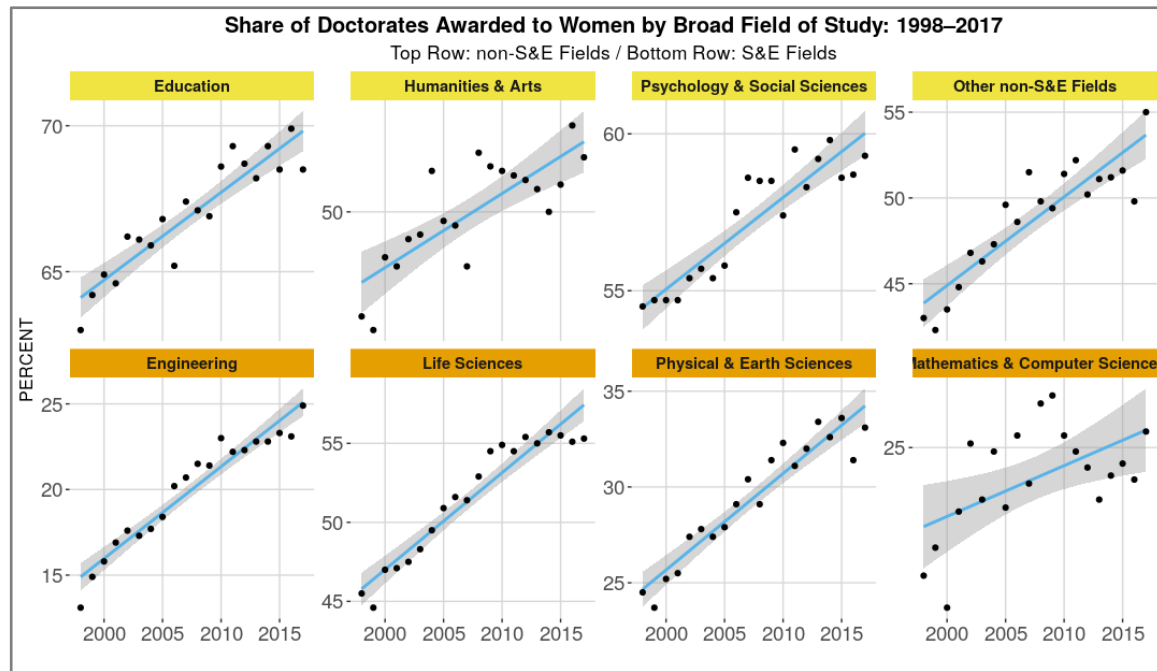


In both plots the range of the vertical axis is reduced to the range of the data. The lines connecting the points are replaced with the predicted lines from a simple linear regression model. The predicted line makes it easier to see the trends and variation over time. Two different legends are displayed, one that labels the lines (left) and has a legend at the bottom of the figure (below).



<sup>16</sup> 2017 Doctorate Recipients from U.S. Universities <https://nces.gov/pubs/nfs19301/> (Accessed on May 5, 2019)

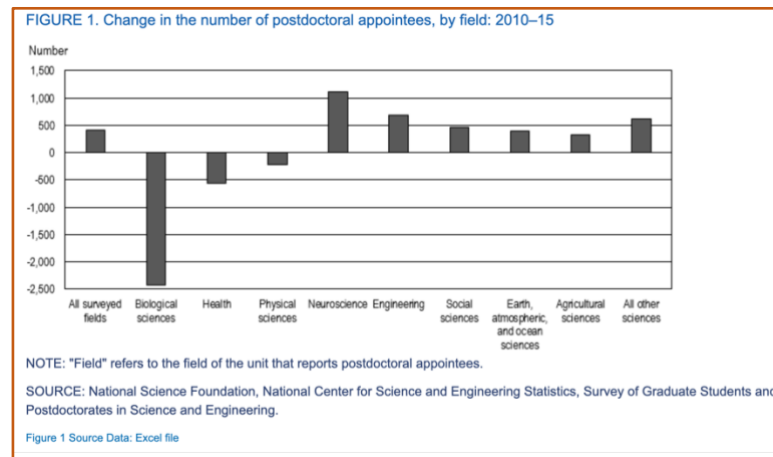
Other suggestions for longitudinal data are to plot the groups separately, two examples using lattice plots are provided below. In both cases the figures are arranged by broad field of study categories, non-S&E (top row) and S&E (bottom row); simple linear regression lines are displayed along with 95% confidence bands. In the top plot the vertical axes are allowed to vary with the range of the broad fields of study. In the top plot one is able to see the 95% confidence bands around the simple linear regression lines which are hidden in the bottom plot where the vertical axes are fixed. The bottom plot it is easy to see that, with the exception of the Life Sciences, there are more doctorates awarded to women in non-S&E fields.





## 5. Graphs that Display a Difference

### InfoBriefs: Field Composition of Postdocs Shifts as Numbers Decline in Biological Sciences and in Clinical Medicine<sup>17</sup>



#### Bar Chart

Two Variables:

> Nominal: Broad field of study (8 levels)

> Interval: Time (2 years)

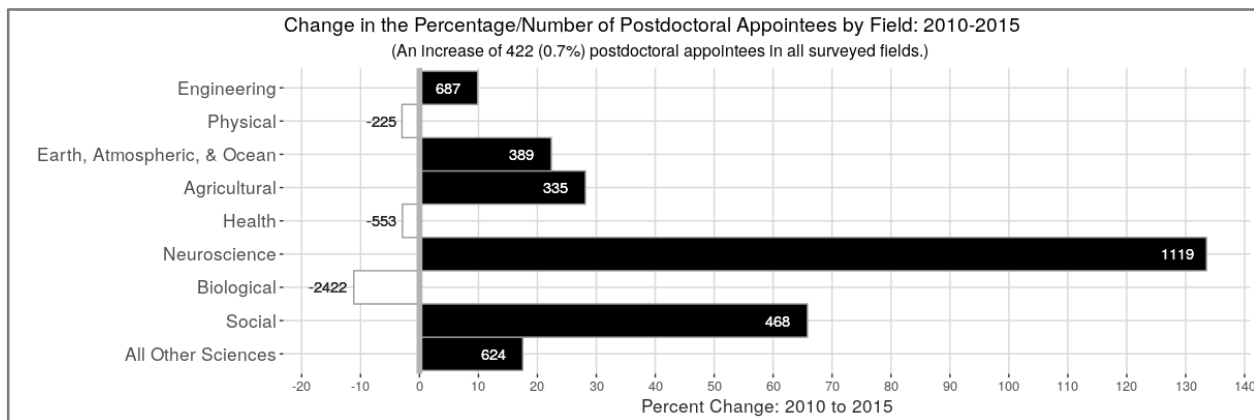
> One reference value: All surveyed fields

#### COMMENTS

> There is a lot of white space in this figure. By using the change in the number of appointees from 2010 to 2015 rather than the relative percent differences, you miss the dramatic increase in neuroscience postdoctoral appointees.

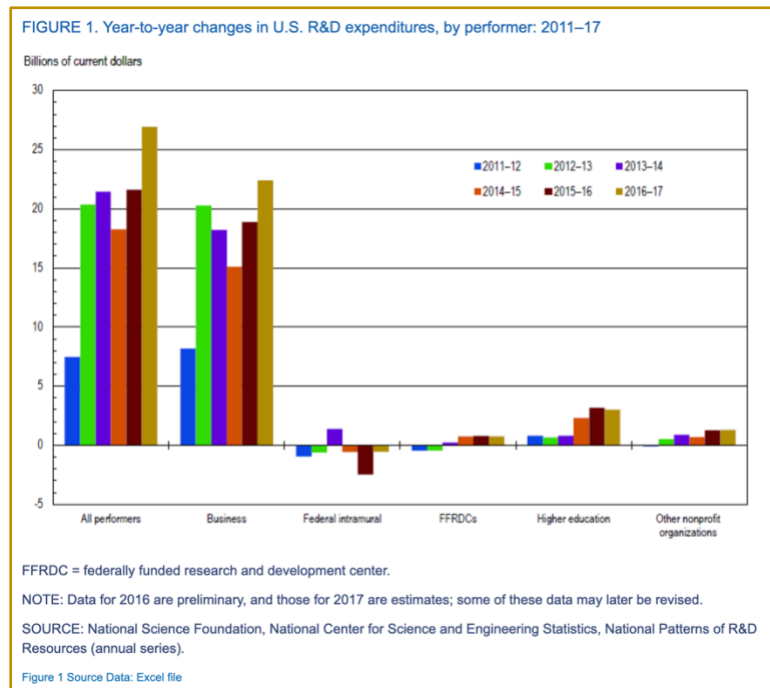
One suggestion is to use black and white to denote the difference between

increases and decreases, plot the relative percent difference, and insert the actual numerical change within the bar. The reference value, the relative percent different for all surveyed fields is displayed in the subtitle.



<sup>17</sup> InfoBrief, February 27, 2019, <https://www.nsf.gov/statistics/2017/nsf17309/> (Accessed May 5, 2019)

InfoBriefs: U.S. R&D Increased by \$22 Billion in 2016, to \$515 Billion; Estimates for 2017 Indicate a Rise to \$542 Billion<sup>18</sup>



### Multi-set Bar Chart

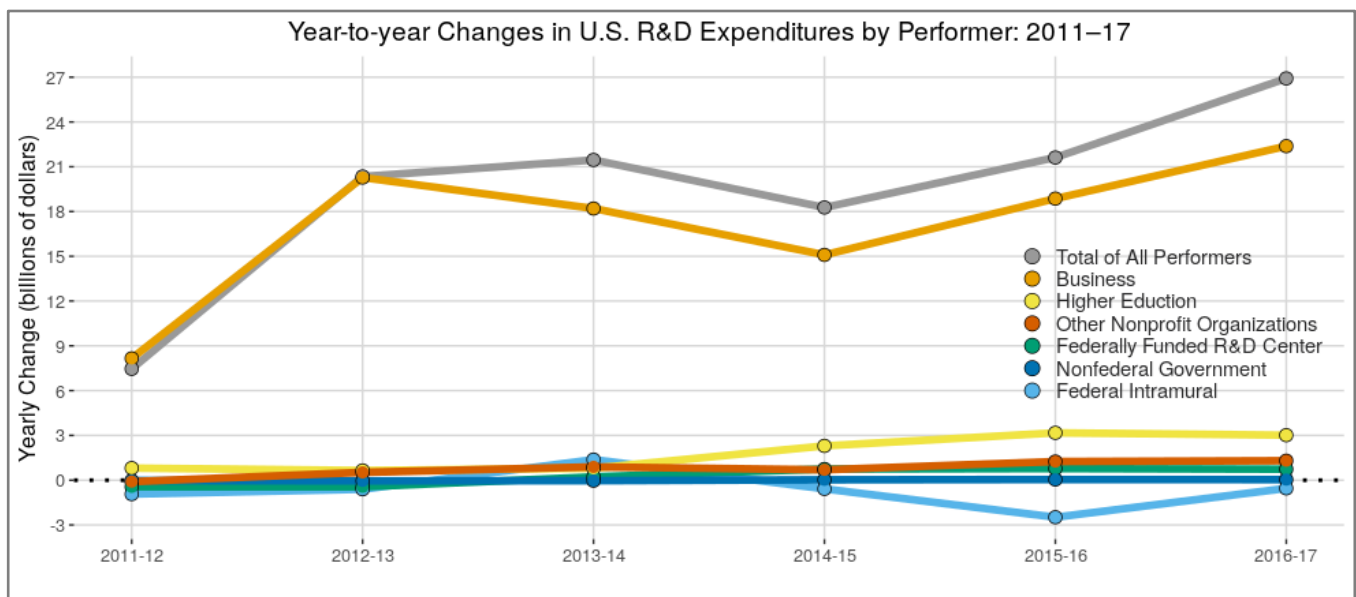
Two Variables:

- > Nominal: Performer (5 levels)
- > Interval: Time (6 1-year intervals)
- > One reference value: All performers

### COMMENTS

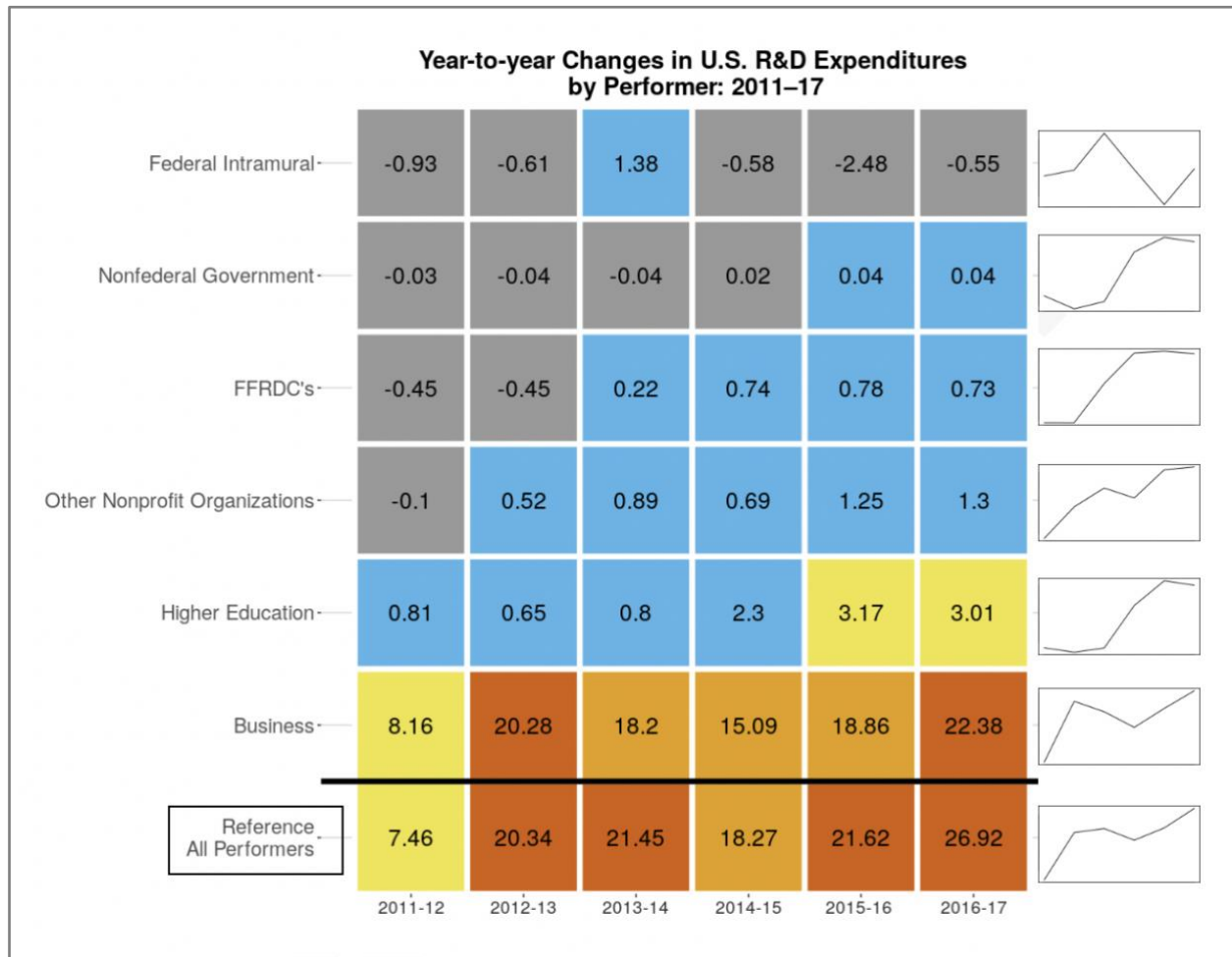
- > To be 508 compliant use colorblind colors.
- > Multi-set bar or grouped bar charts are difficult to interpret. In this case since time intervals are one of the plotting variables one suggestion is to replot the data as a line graph.

The data used in the figure on the left figure are displayed below using a line graph.



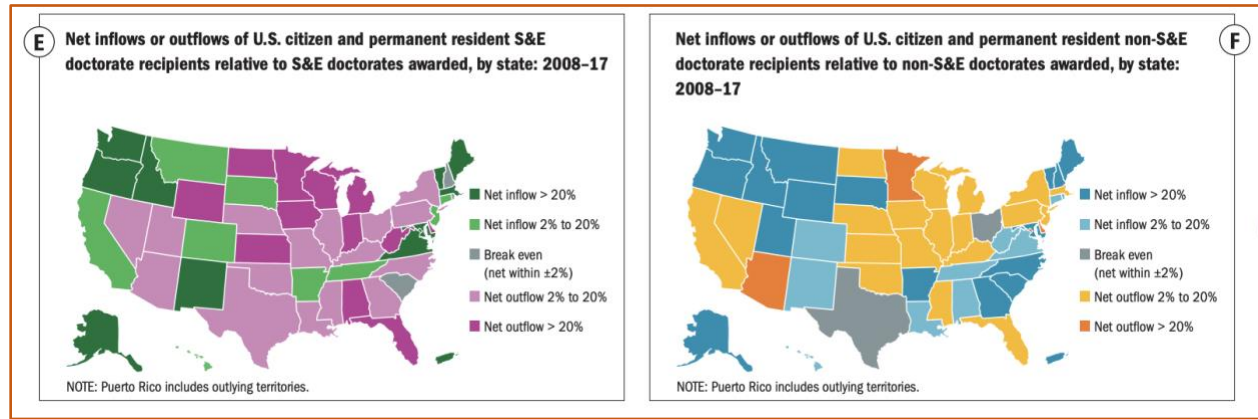
<sup>18</sup> InfoBrief, February 27, 2019, <https://www.nsf.gov/statistics/2019/nsf19308/> (Accessed May 5, 2019)

In cases where the values differ by an order of magnitude, one suggestion is to dichotomize the values and use a heat map which will display the categories using colors and can include the actual values. In the plot below trend lines are included to aid in interpretation.



## 6. Geospatial Graphs

Patterns of interstate mobility: What are the employment and geographic trends? Net inflows and outflows by state<sup>19</sup>



### Chloropleth Map (1998-2017)

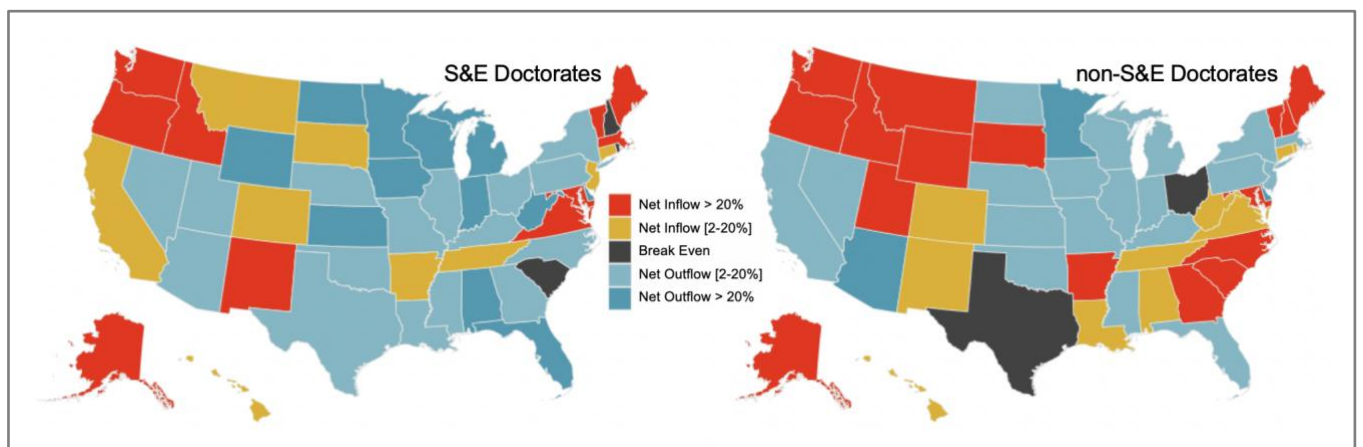
Two Variables:

> Nominal: States + Puerto Rico (51 levels)

> Interval: Inflow/outflow of S&E and non-S&E doctorate recipients between 2008 and 2017

COMMENTS:

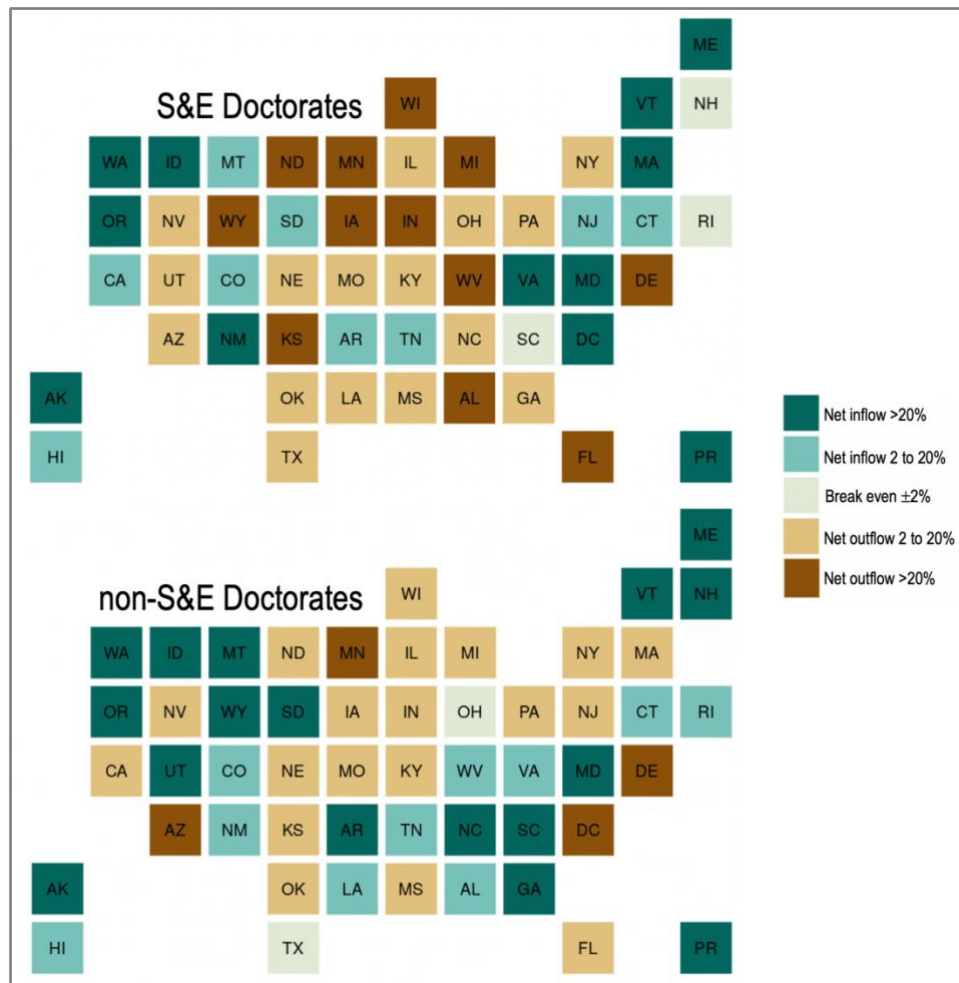
> If the same color palette was used for both maps it would make it easier to make comparisons between S&E and non-S&E doctorate recipients (see below and on the next page).



<sup>19</sup> 2017 Doctorate Recipients from U.S. Universities <https://nces.gov/pubs/nsf19301/> (Accessed on May 5, 2019)

Cartogram heatmaps are an alternative to choropleth maps on the previous page, that are based on work by the Washington Post graphics department in their report on "[The states most threatened by trade](http://www.washingtonpost.com/wp-srv/special/business/states-most-threatened-by-trade/)".<sup>20</sup> "State bins" preserve as much of the geographic placement of the states as possible but has the look and feel of a traditional heatmap (see page 16).

Net Inflows and Outflows of U.S. Citizen and Permanent Resident S&E and non-S&E Doctorate Recipients by State: 2008-2017

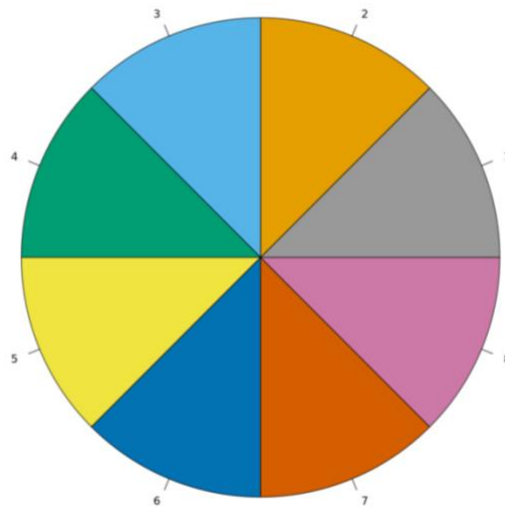


<sup>20</sup> Washington Post, August 22, 2014, <http://www.washingtonpost.com/wp-srv/special/business/states-most-threatened-by-trade/>

**General Comments:**

1. Reduce the noninformation white space by reducing the axis to cover the range of the data. In the case of line graphs where the horizontal axis is time, this will make it easier to detect trends and differences in trends over time.
2. Use a colorblind palette to be 508 compliant.

Color blind palette with grey  
(black can be substituted for grey).



3. When possible use the same color for each level of a variable throughout the publication this will make comparisons between the figures easier.
4. NCSES figures often include reference values - a level of a variable that is a summary of all or a portion of the other levels. These values should be treated separately in the figure.

**Some tips:**

1. Avoid distorting the data.
2. Make large data sets coherent.
3. Encourage the eye to make comparisons.
4. Closely integrate the statistical displays with the description of the data in the document.

Adapted from E. R. Tufte (2001) *The Visual Display of Quantitative Information*, 2<sup>nd</sup> Edition, Graphics Press, Cheshire, Connecticut

## References:

- Data Visualization Catalog  
<https://datavizcatalogue.com/>
- Dataviz Project by ferdio  
<https://datavizproject.com/>
- 508 Compliant Visualizations  
<https://www.hhs.gov/web/section-508/making-files-accessible/checklist/word/index.html>
- Escaping RGBland: Selecting Colors for Statistical Graphics  
<https://eeecon.uibk.ac.at/~zeileis/papers/Zeileis+Hornik+Murrell-2009.pdf>
- Consumer Financial Protection Design Manual  
<https://cfpb.github.io/design-manual/data-visualization/data-visualization.html>