UNIVERSITY of VIRGINIA

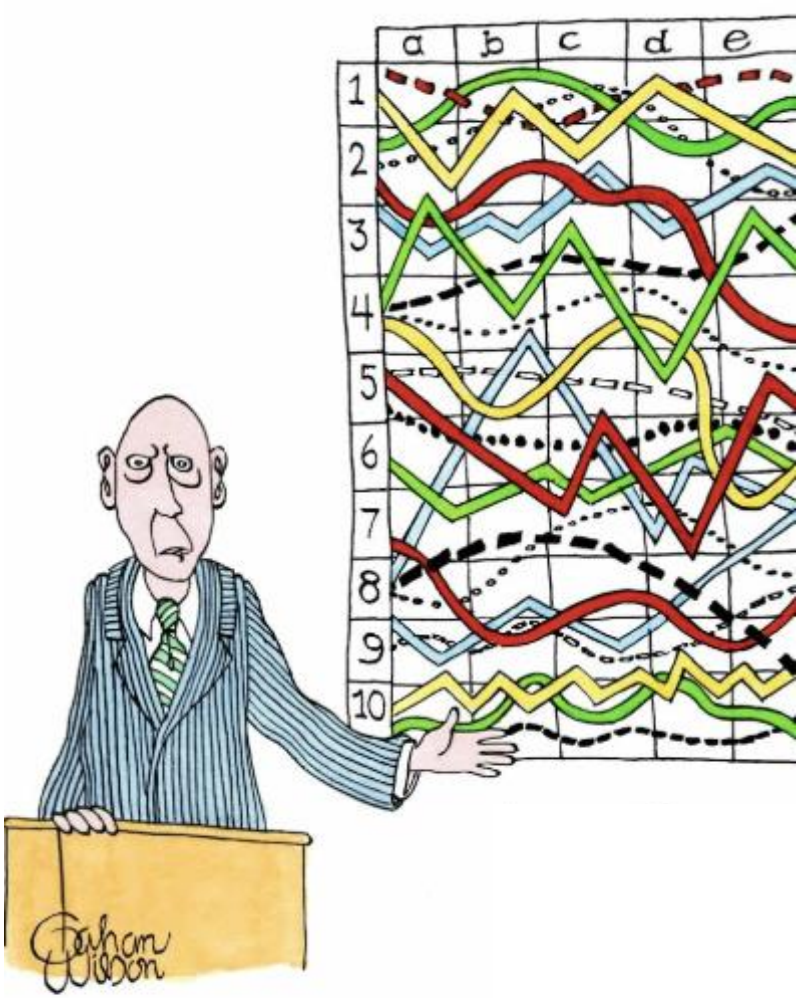
JULY 24, 2019

DATA SCIENCE FOR THE PUBLIC GOOD WORKSHOP

TIPS ON CONSTRUCING VISUALIZATIONS

VICKI LANCASTER

VAL7ZV@VIRGINIA.EDU



"I'll pause for a moment so you can let this information sink in."

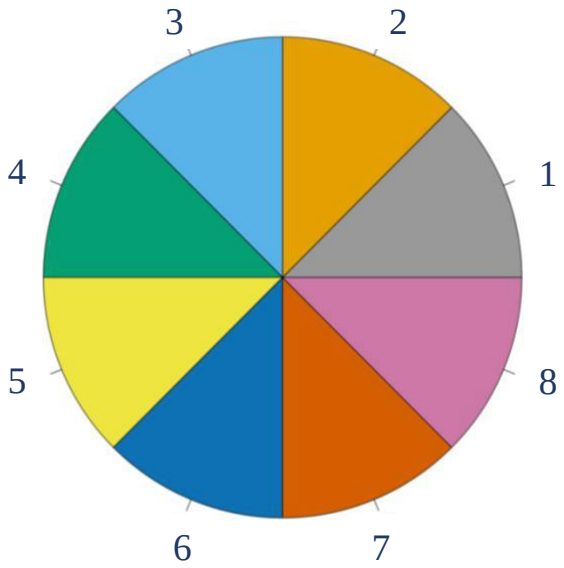
Some general non-subjective guidelines

1. Figures should stand alone and not require the reader to refer to the text.
2. When in a document, figures should be closely integrated with the description of the data in the document.
3. Avoid distorting the data.
4. Include your data source.
5. Include a title.
6. Label axes and include the units. Labels should be descriptive.
7. Use legends or annotations to explain your aesthetic mappings (how levels of a variable are mapped to visual properties for example, line type, color, pattern, etc.).

Subjective guidelines for COLOR

- Stay away from the defaults
- Use a 508 compliant colorblind palette

Colorblind Palette



1=#999999
2=#E69F00
3=#56B4E9
4=#009E73
5=#F0E442
6=#0072B2
7=#D55E00
8=#CC79A7

Very good presentation on how to make colorblind friendly figures and presentation https://jfly.uni-koeln.de/color/

Subjective guidelines for COLOR

A must read...

Escaping RGBland: Selecting Colors for Statistical Graphics

Achim Zeileis, Wirtschaftsuniversität Wien
Kurt Hornik, Wirtschaftsuniversität Wien
Paul Murrell, The University of Auckland

Article: http://epub.wu.ac.at/1692/ and Website: https://cran.r-project.org/web/packages/colorspace/vignettes/colorspace.html

Use colors palettes that are perceptually based and do not introduce optical illusions or systematic bias.

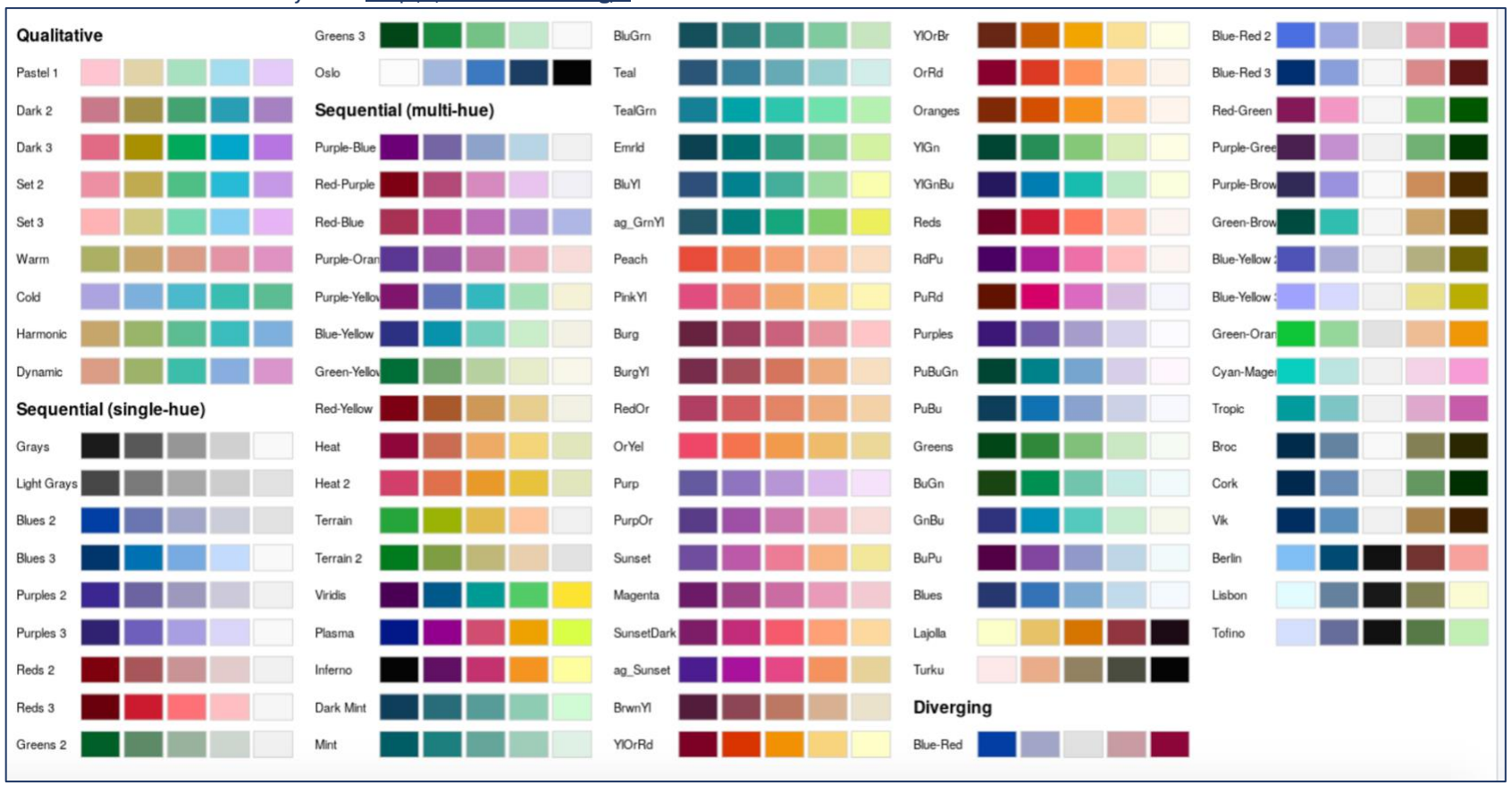
Describe colors in terms of three qualities (HCL):

- **H**ue - color or shade (five principle hues: red, yellow, green, blue purple) 
- **C**hroma - the strength or purity of the color (change by adding shades of gray) 
- **L**uminance - the light of the color 

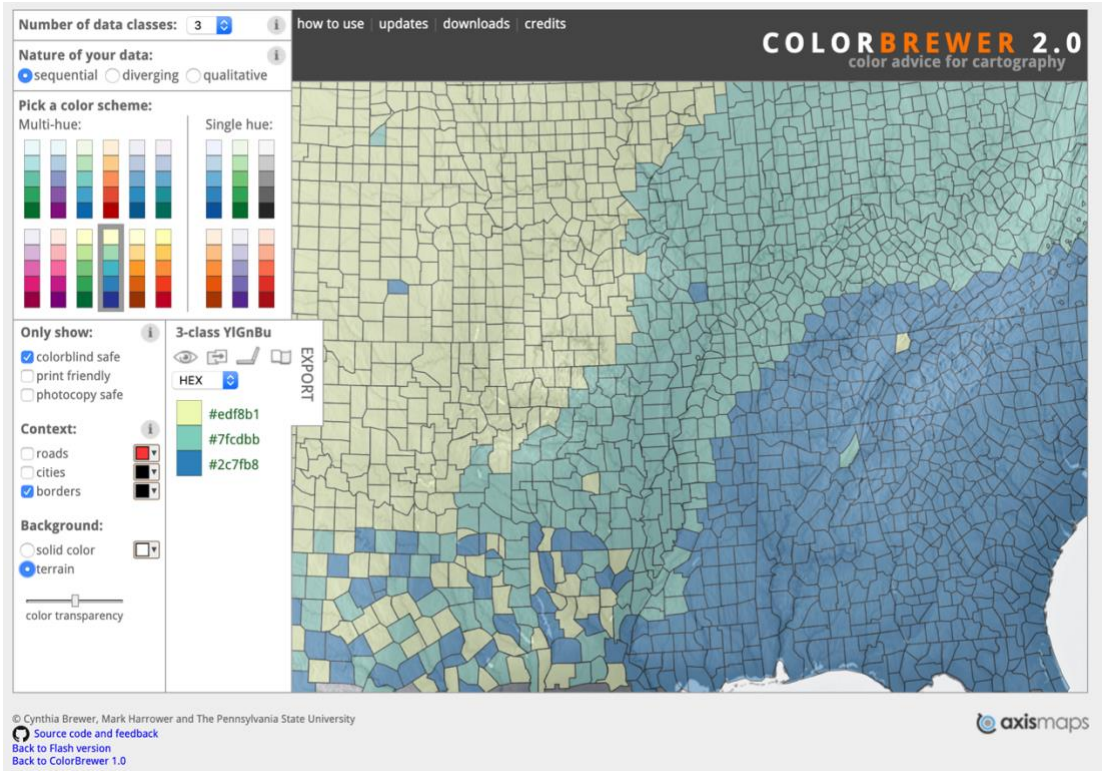
Subjective guidelines for COLOR

```
> library("colorspace")
> hcl_palettes(plot=TRUE)
```

HCL Wizard for R and Python http://hclwizard.org/



Subjective guidelines for COLOR



http://colorbrewer2.org/#type=sequential&scheme=YlGnBu&n=3

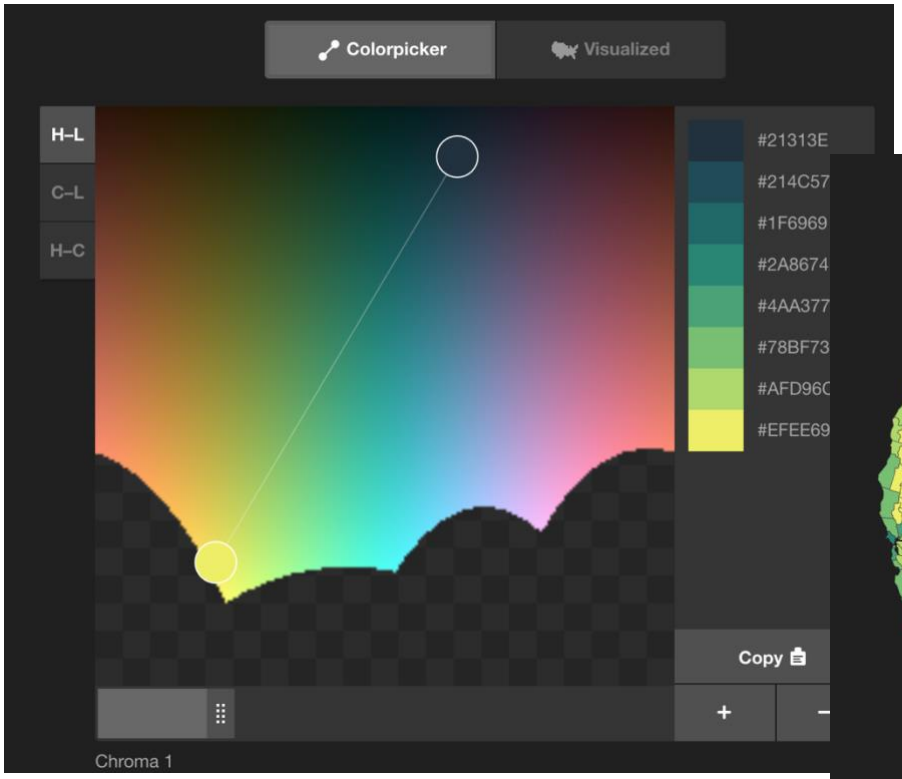
An R cheatsheet for the packages **colorspace** & **RcolorBrewer**
https://www.nceas.ucsb.edu/~frazier/RSpatialGuides/colorPaletteCheatsheet.pdf

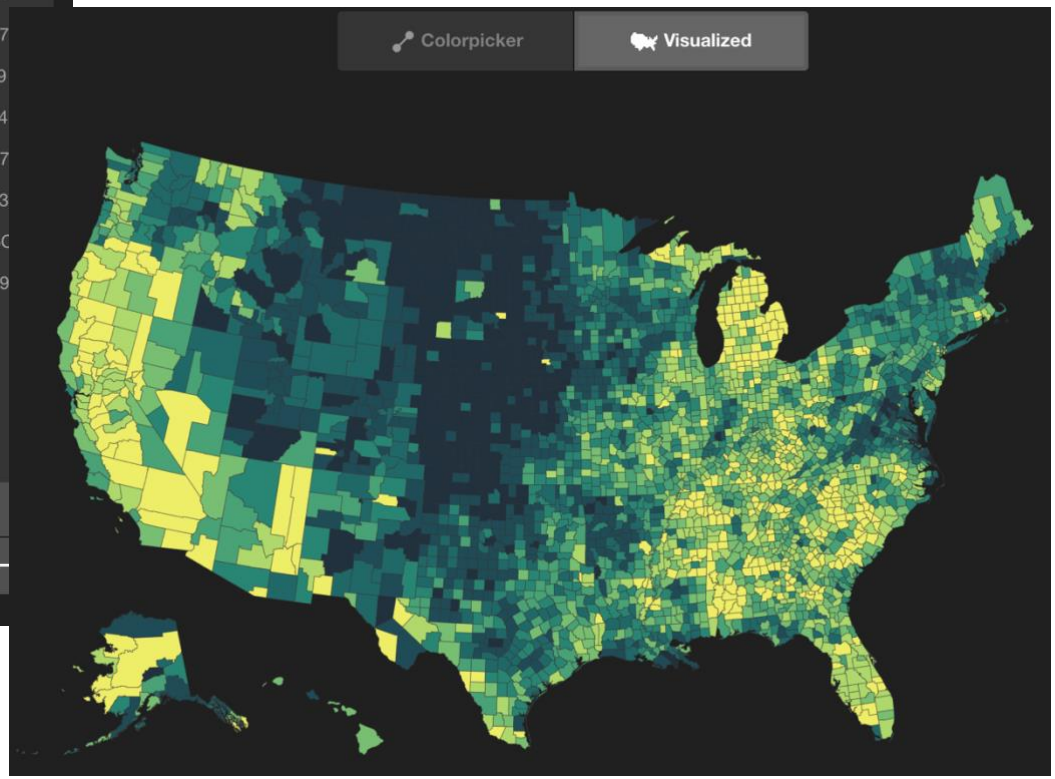
R Cookbook on color http://www.cookbook-r.com/Graphs/Colors_(ggplot2)/#a-colorblind-friendly-palette

Subjective guidelines for COLOR

Colorpicker for data

Built off Gregor Aisch's article and color conversion library chroma.js. Fork it on GitHub.

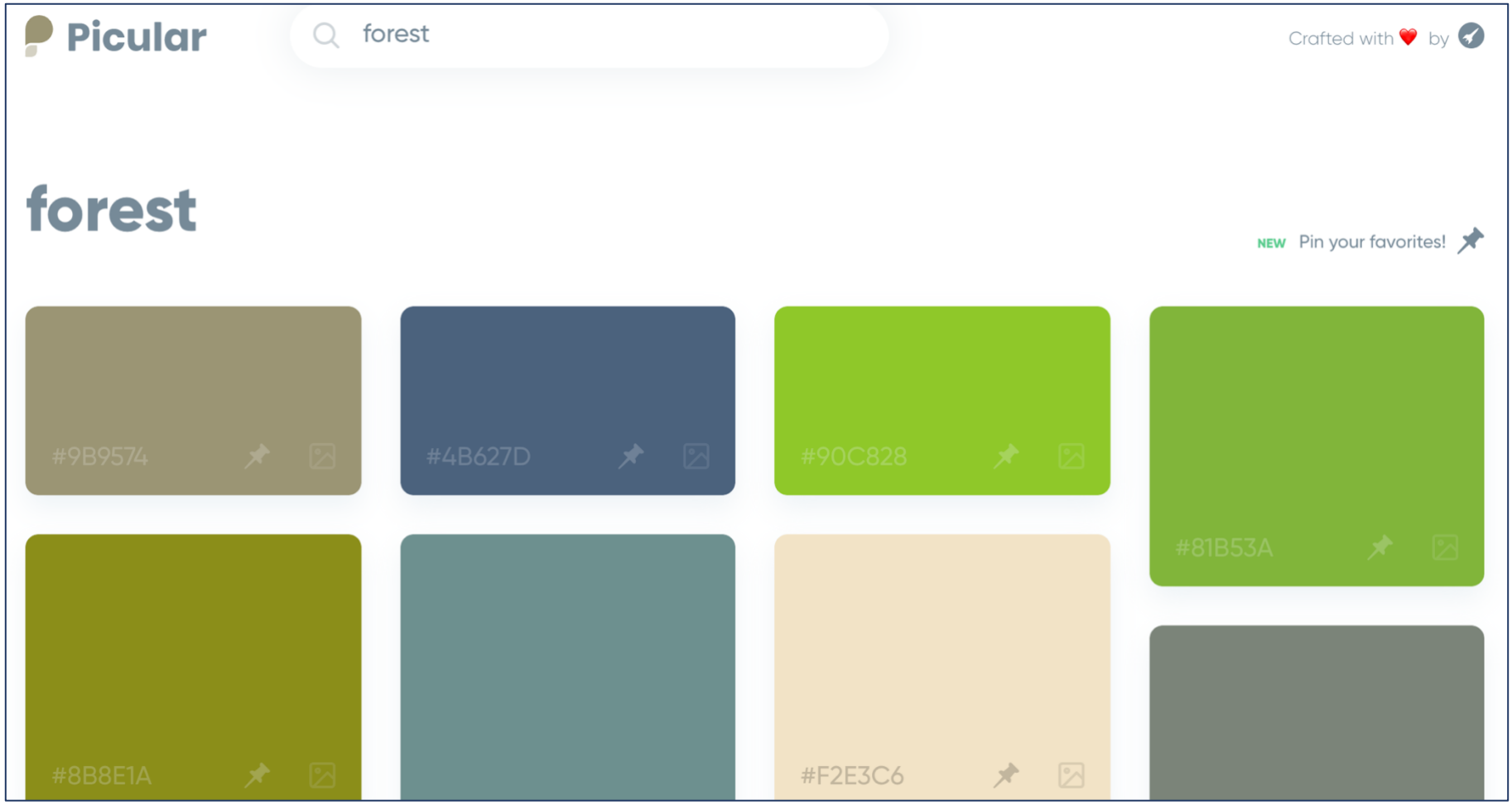




http://tristen.ca/hcl-picker/#/hlc/6/1/15534C/E2E062

Subjective guidelines for COLOR

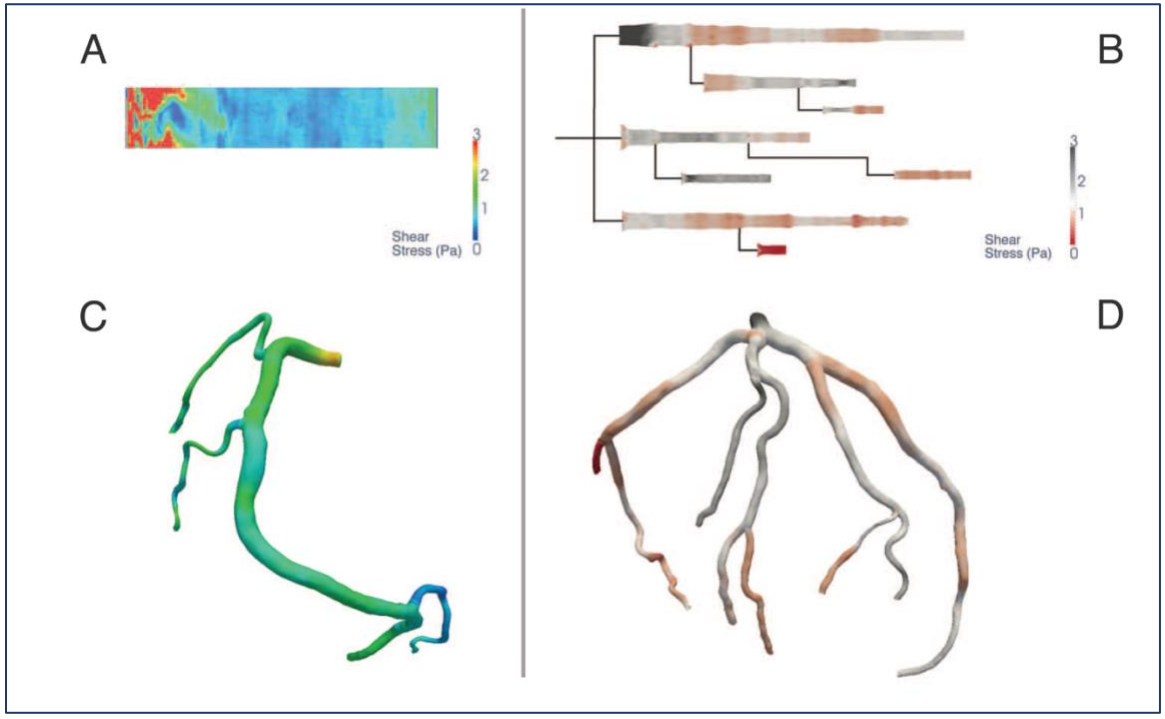
A search engine for color palettes...



https://picular.co/

Color Choice Is Important

In Borkin et al. (2011) "*Heart disease is the number one killer in the United States, finding indicators of the disease at an early stage is critical for treatment and prevention. In this paper we evaluate visualization techniques that enable the diagnosis of coronary artery disease. A key physical quantity of medical interest is endothelial shear stress (ESS).*"

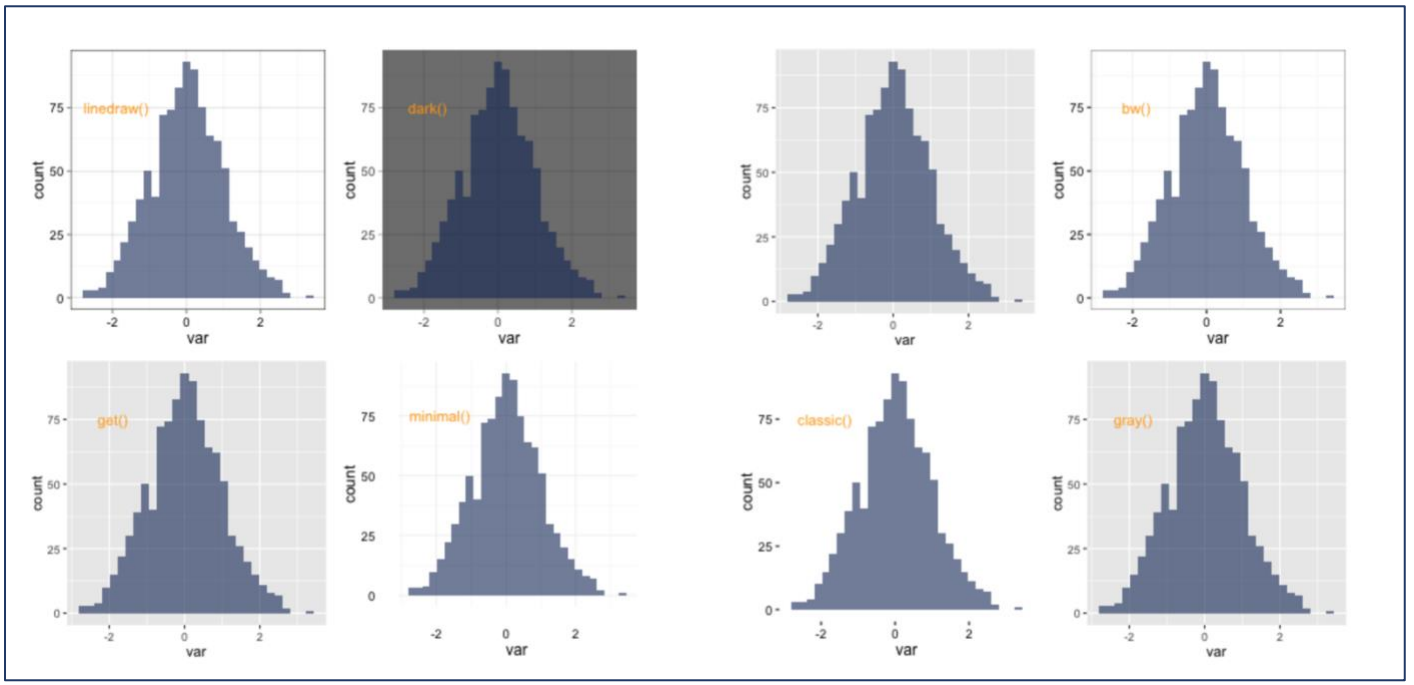


A perceptually appropriate 2D-divergent color map leads to fewer diagnostic mistakes than a 2D or 3D rainbow color map.

Subjective guidelines for ggplot2 **theme_xx()**

- Stay away from the defaults
- Think twice about using background color, it is an unnecessary use of ink (most of the time)

Theme Options in ggplot2



The R Graph Gallery https://www.r-graph-gallery.com/192-ggplot-themes/

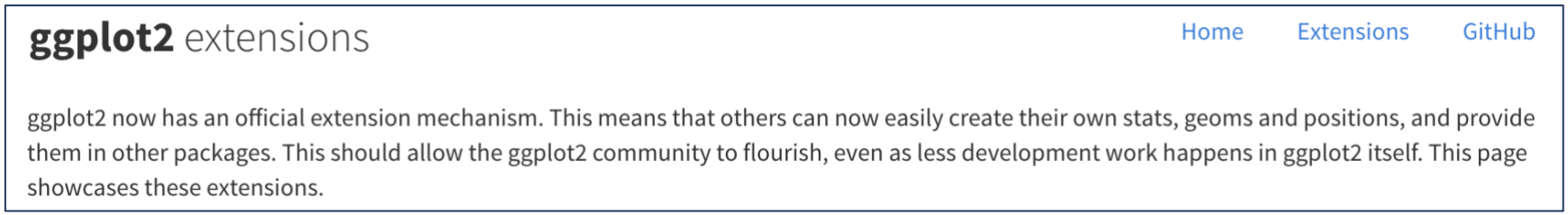
Subjective guidelines for ggplot2 **theme_xx()**

Exactly what do the themes control?

- **background color**
- **panel background color** and
- **grid lines**

There are extensions to theme_xx() in the
library("ggthemes").

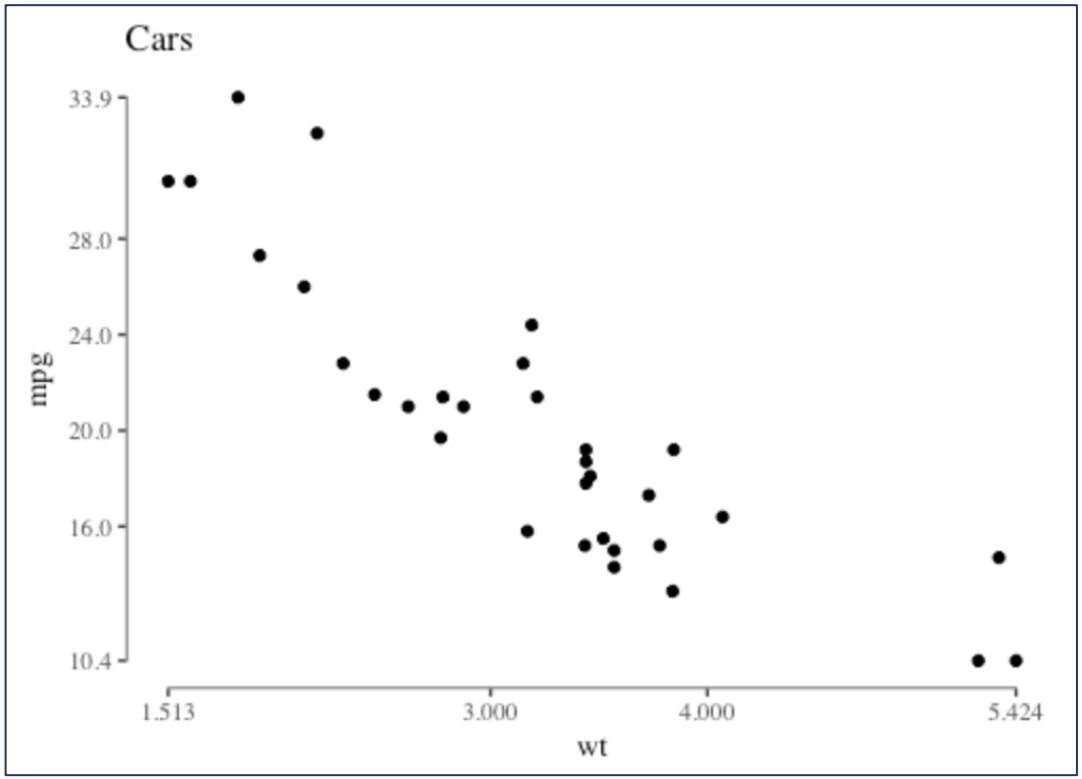
A very good reference for ggthemes...



https://www.ggplot2-exts.org/ggthemes.html

Subjective guidelines for ggplot2 **theme_xx()**

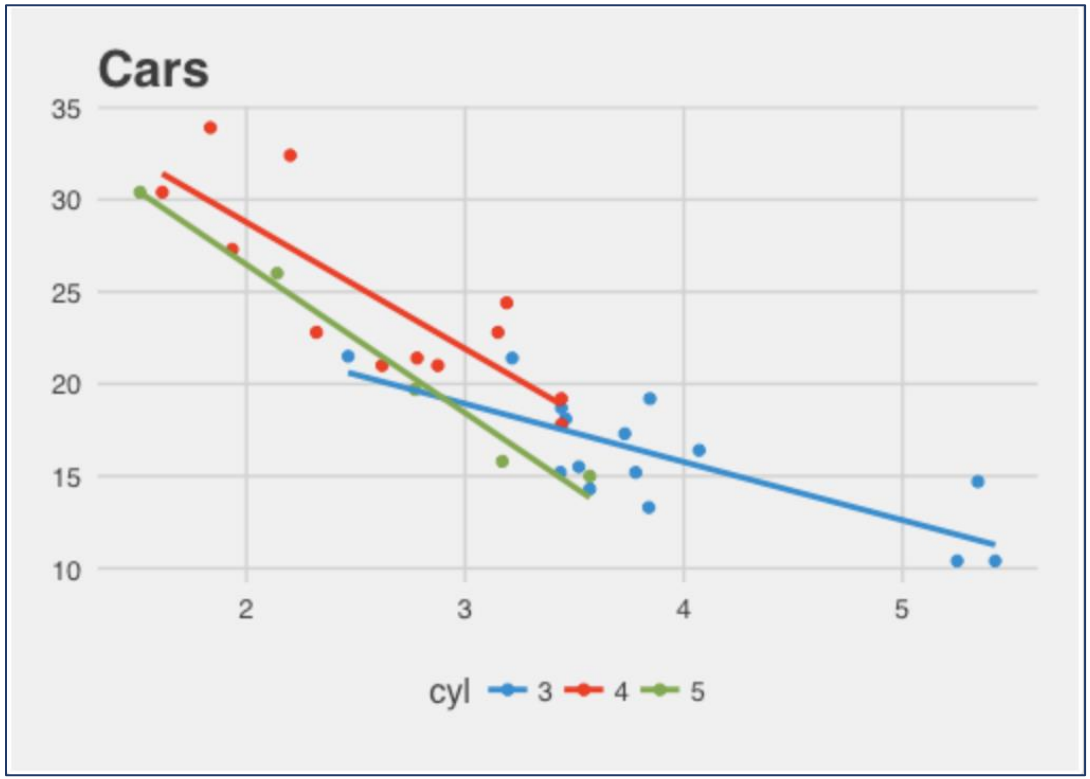
theme_tufte: a minimal ink theme based on Tufte's *The Visual Display of Quantitative Information*



https://mran.microsoft.com/snapshot/2016-12-03/web/packages/ggthemes/vignettes/ggthemes.html

Subjective guidelines for ggplot2 **theme_xx()**

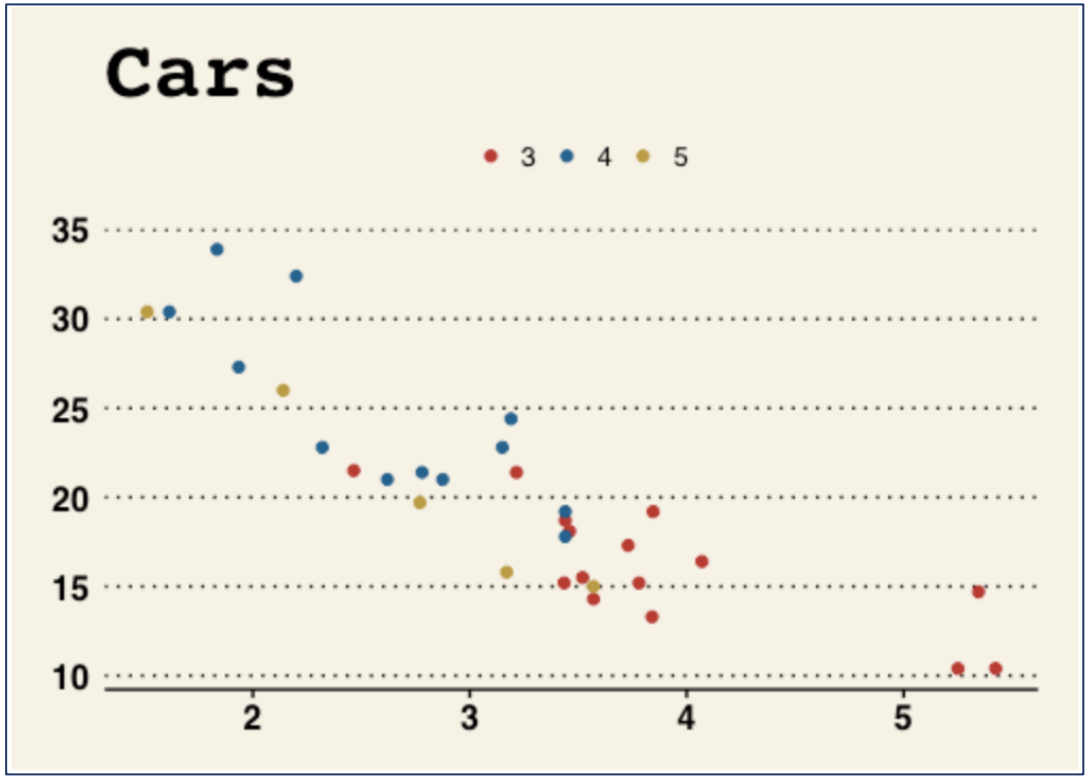
theme_fivethirtyeight: a theme based on the plots at fivethirtyeight.com



https://mran.microsoft.com/snapshot/2016-12-03/web/packages/ggthemes/vignettes/ggthemes.html

Subjective guidelines for ggplot2 **theme_xx()**

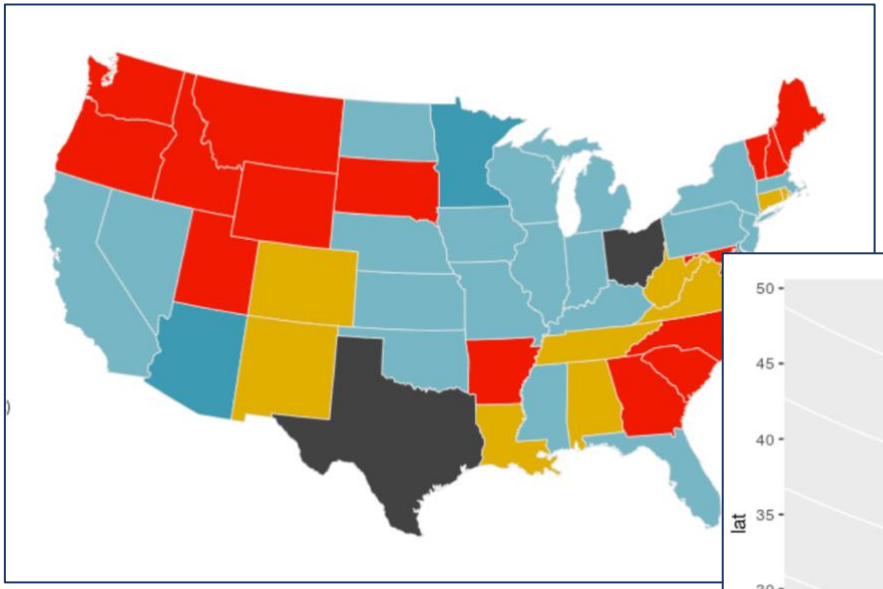
theme_wsj: a theme based on the plots in the The Wall Street Journal



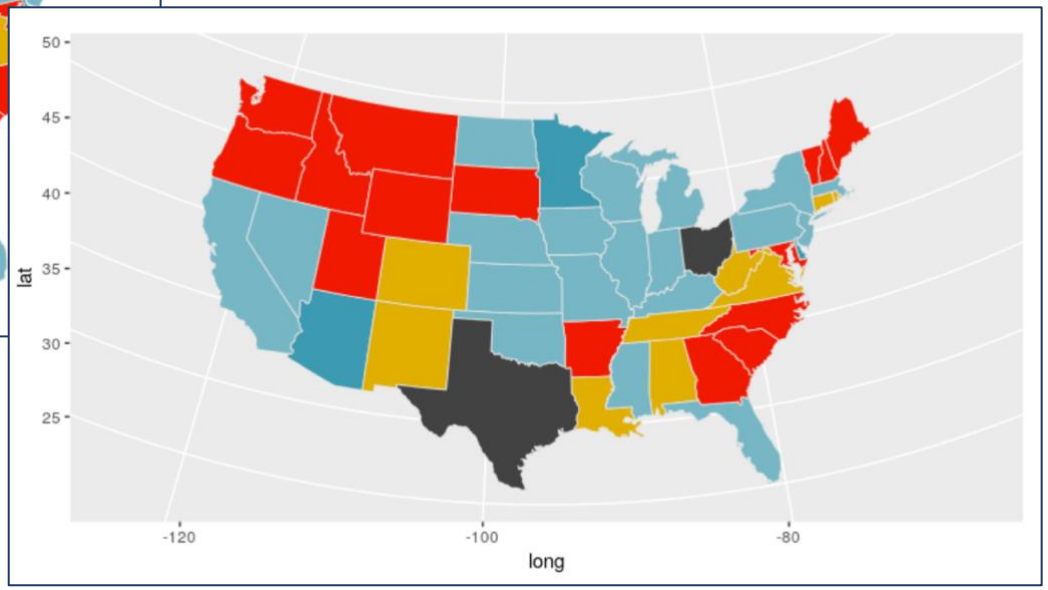
https://mran.microsoft.com/snapshot/2016-12-03/web/packages/ggthemes/vignettes/ggthemes.html

Subjective guidelines for ggplot2 **theme_xx()**

I **highly recommend** the ggthemes function for maps, theme_map(). There is **no reason** to include the axes for longitude and latitude, including them is a waste of space.



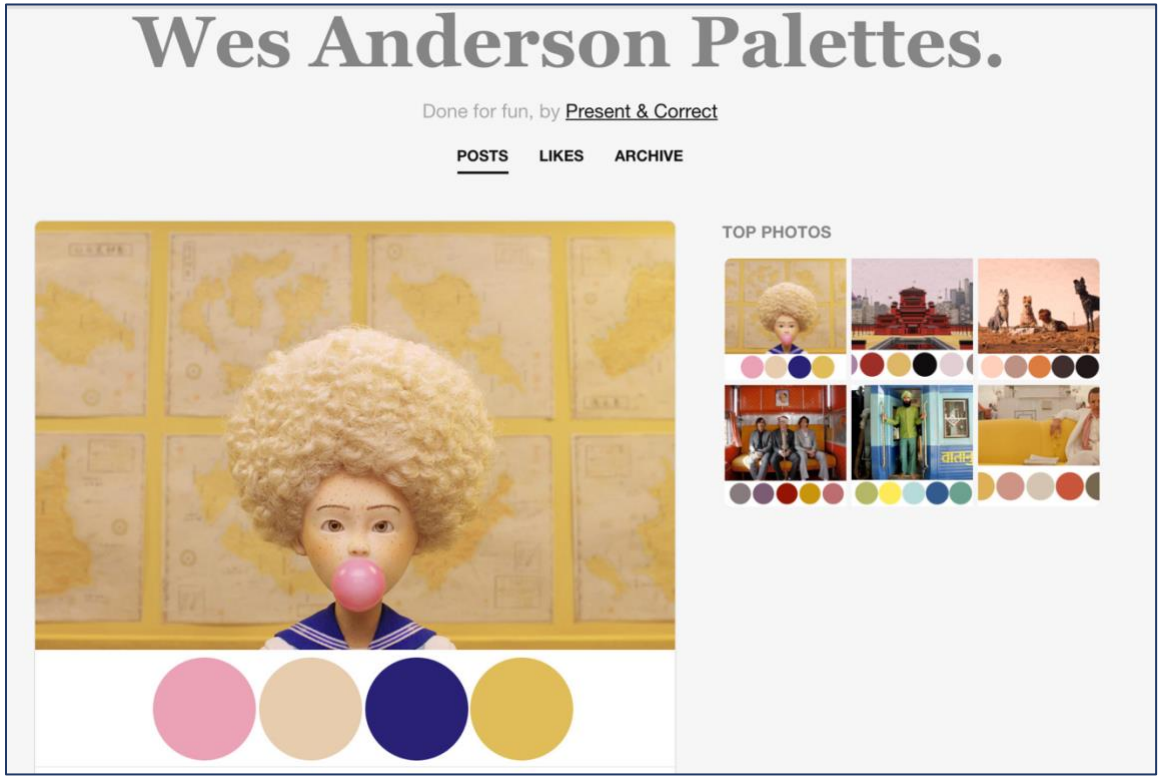
with theme_map()



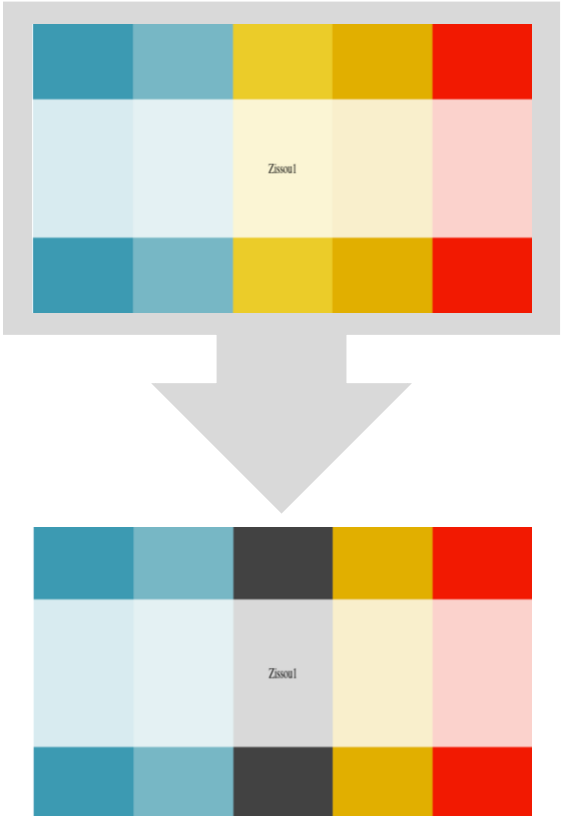
without theme_map()

Subjective guidelines for ggplot2 **theme_xx()**

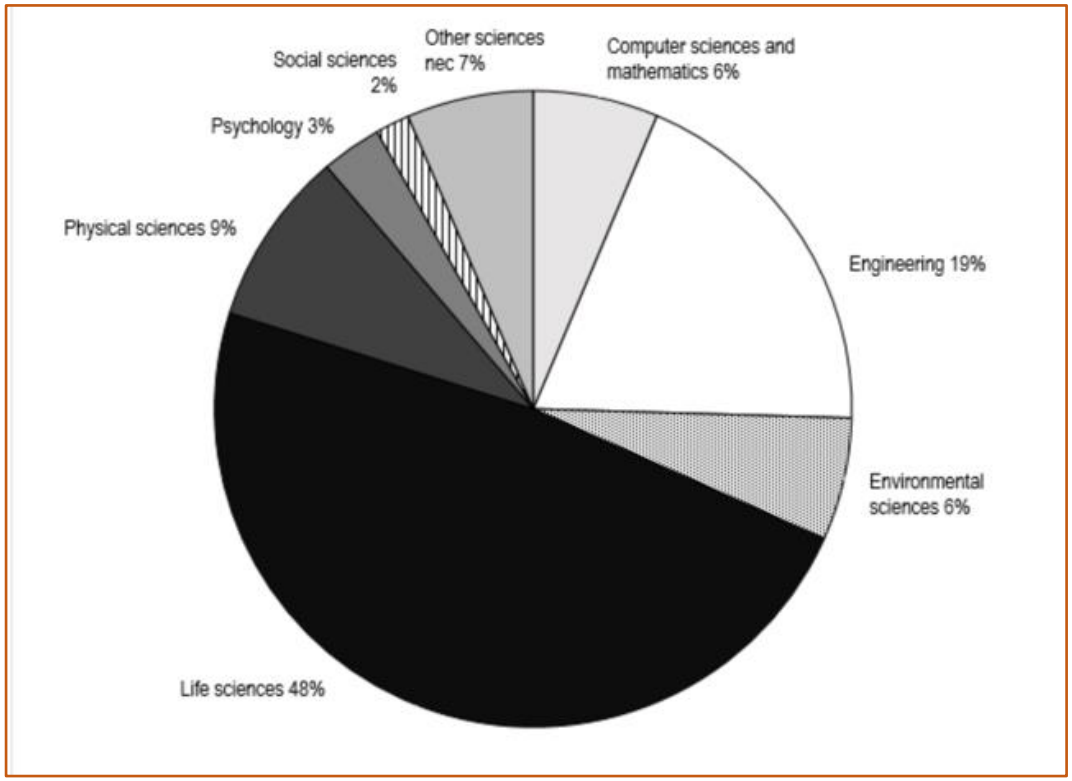
The U.S. maps used the wesanderson R package palette, Zizzous1, to create a divergent palette by substituting a dark grey for the middle color.



https://wesandersonpalettes.tumblr.com/



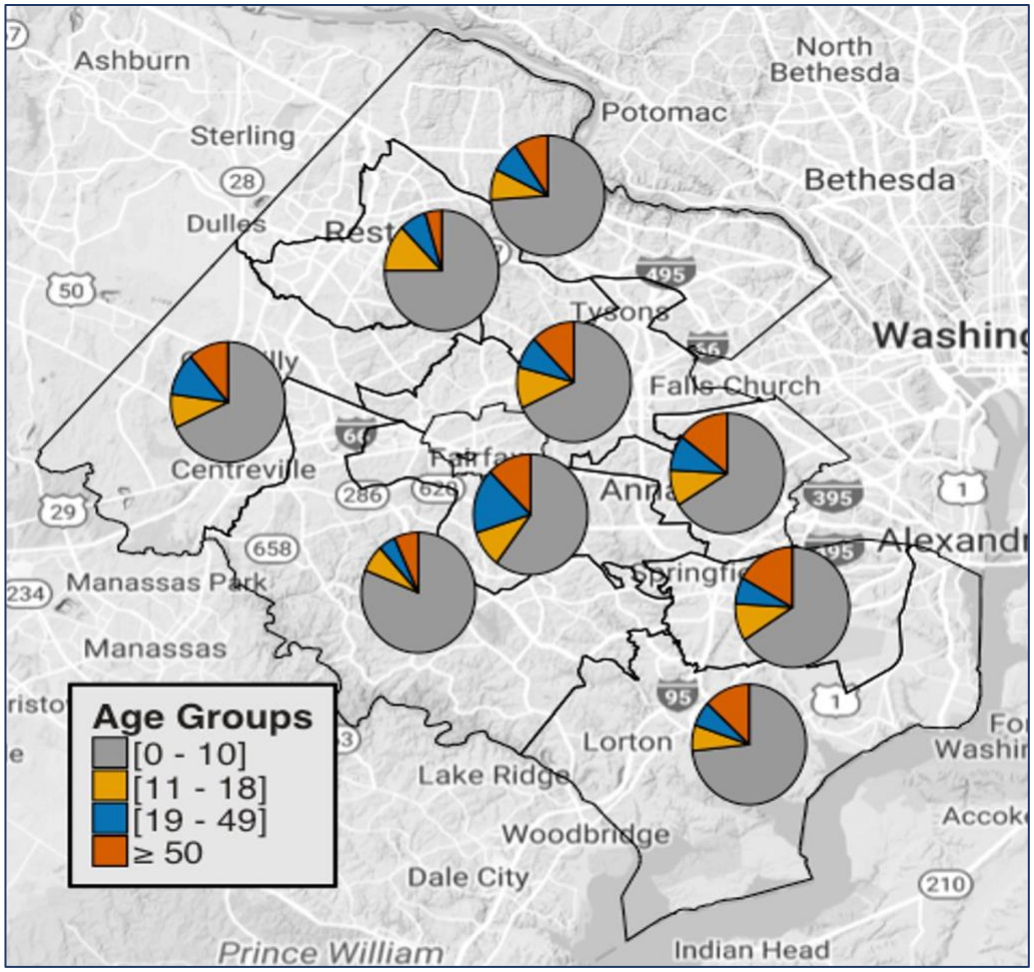
Before & After: Alternative to a Pie Chart



Pie charts only make it easy to judge the magnitude of a slice when it is close to 0%, 25%, 50%, 75%, or 100%,"[1] 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.

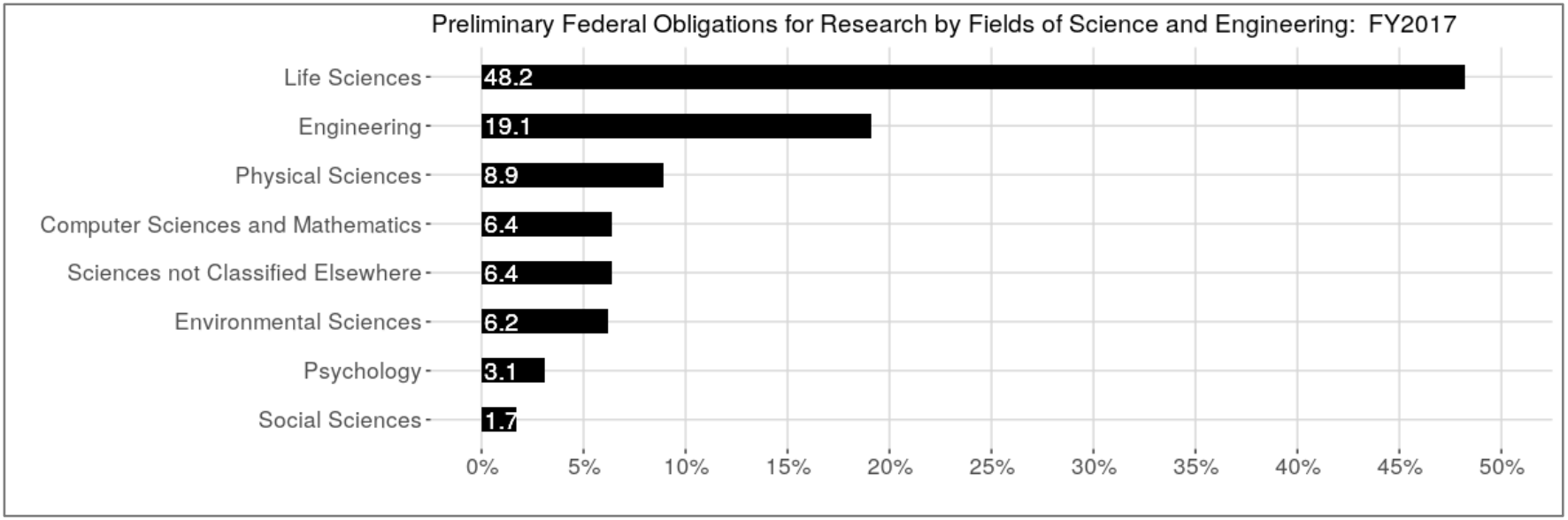
[1] Stephen Few (2007) Save the Pies for Desert
https://www.perceptualedge.com/articles/visual_business_intelligence/save_the_pies_for_dessert.pdf

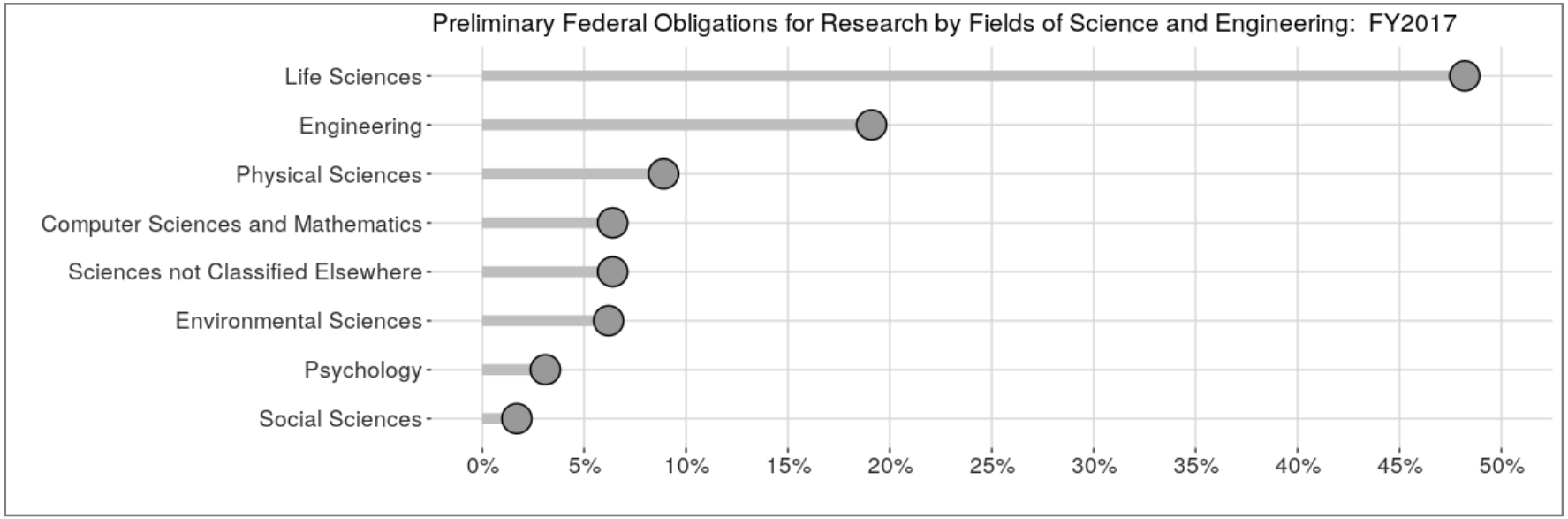
Before & After: Alternative to a Pie Chart



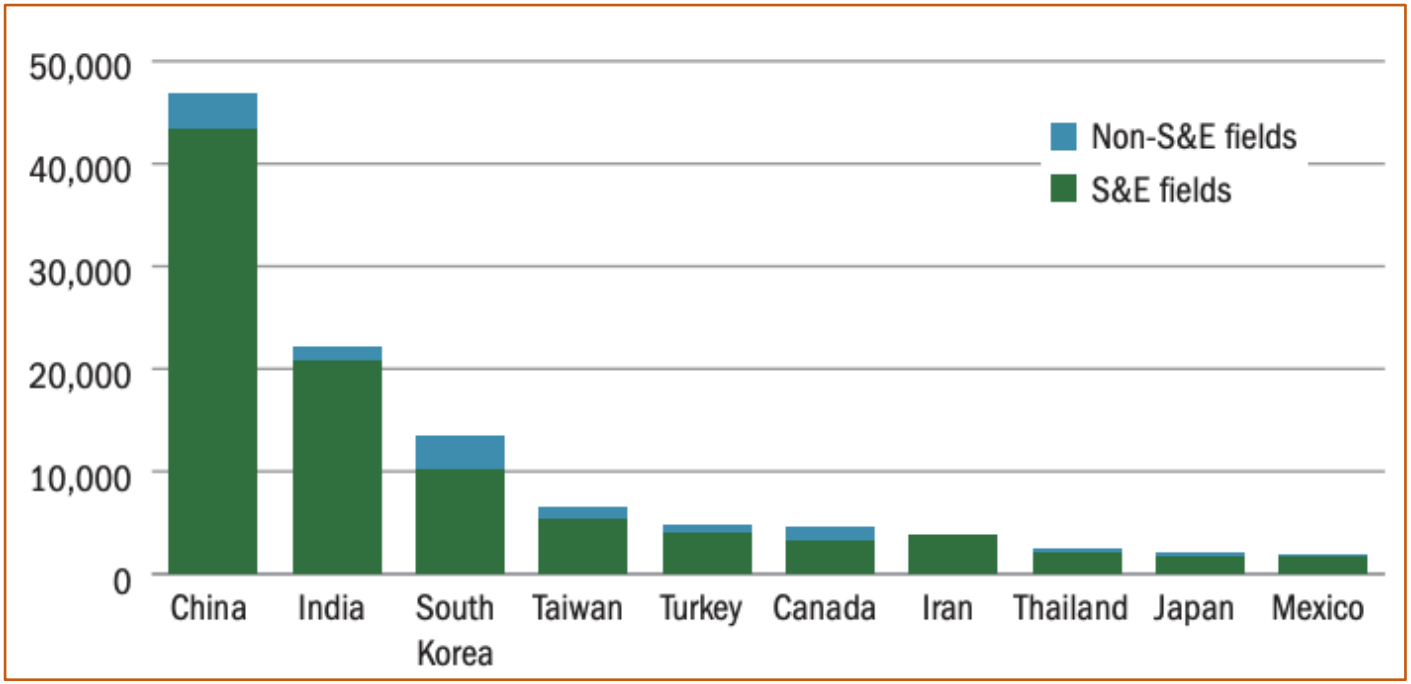
A more appropriate use of pie charts? Is there another way to display this a categorical variable on a map?

Before & After: Alternative to a Pie Chart



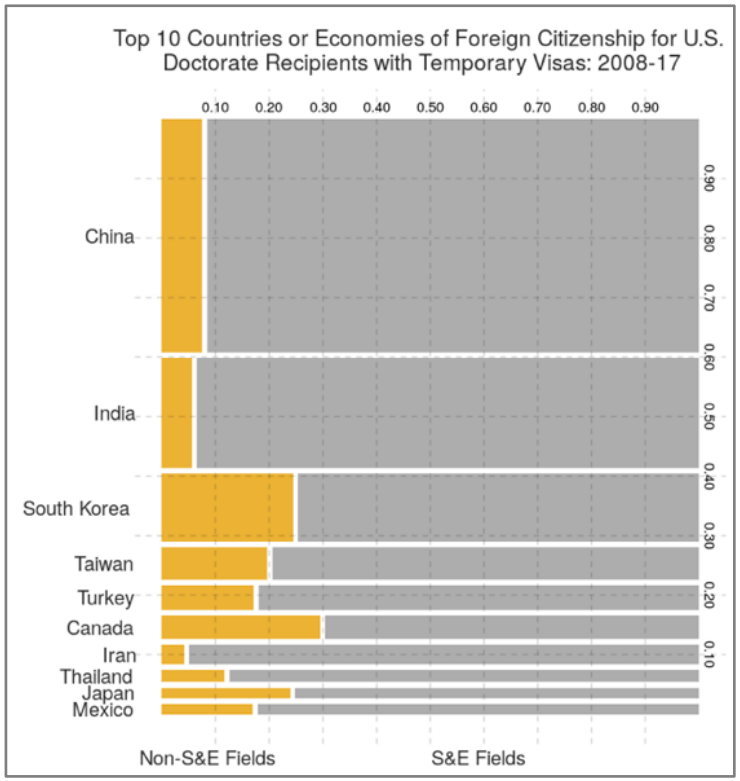


Before & After: Alternative to a Bar Chart



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.

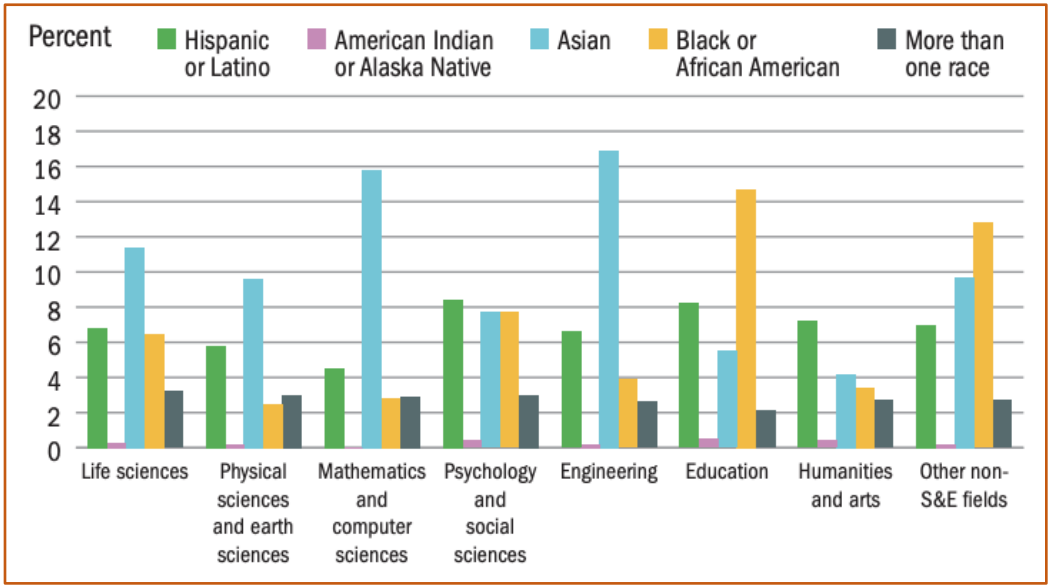
Before & After: Alternative to a Bar Chart



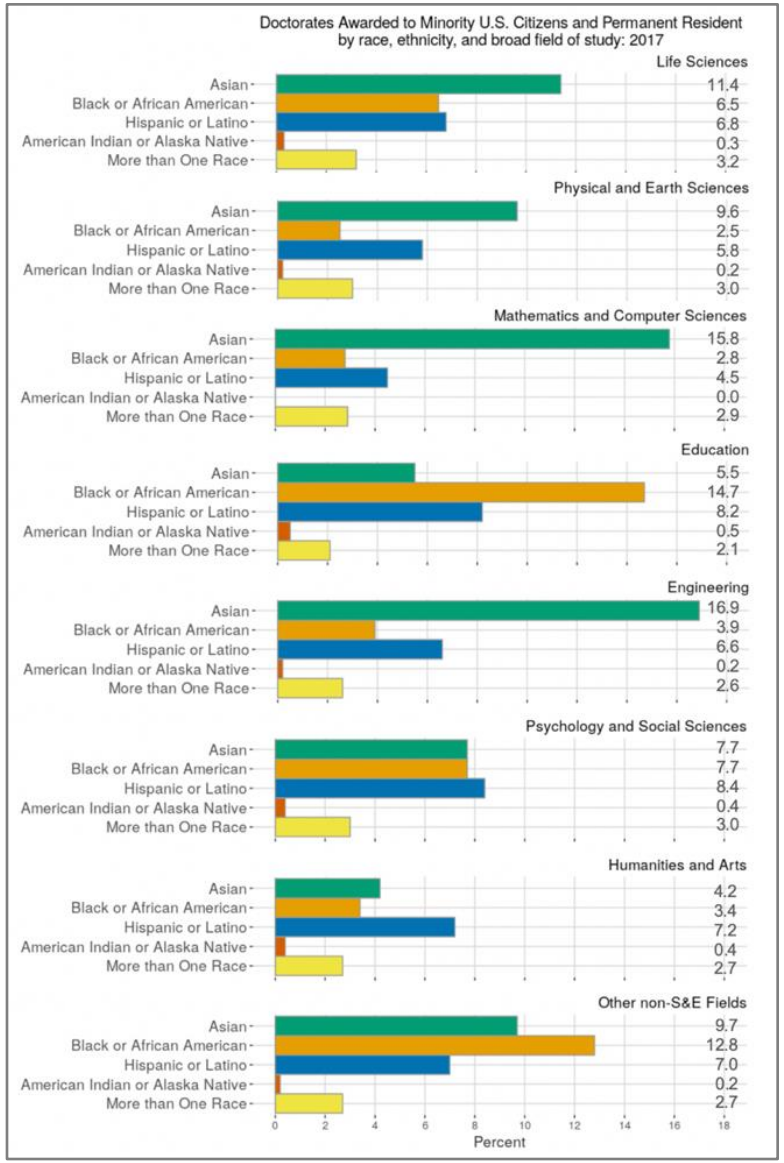
The data displayed as a mosaic plot[2]. In a mosaic plot (also referred to as a marimekko plot), both axes are on a percentage scale that determines both the width and height of each segment. In this case the country or economy is on the vertical axis where it is easy to see that China, India, and South 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 makes comparison between non-S&E versus S&E fields within and between each country or economy.

[2] Data Visualization Catalog https://datavizcatalogue.com/methods/marimekko_chart.html (Accessed on May 5, 2019)

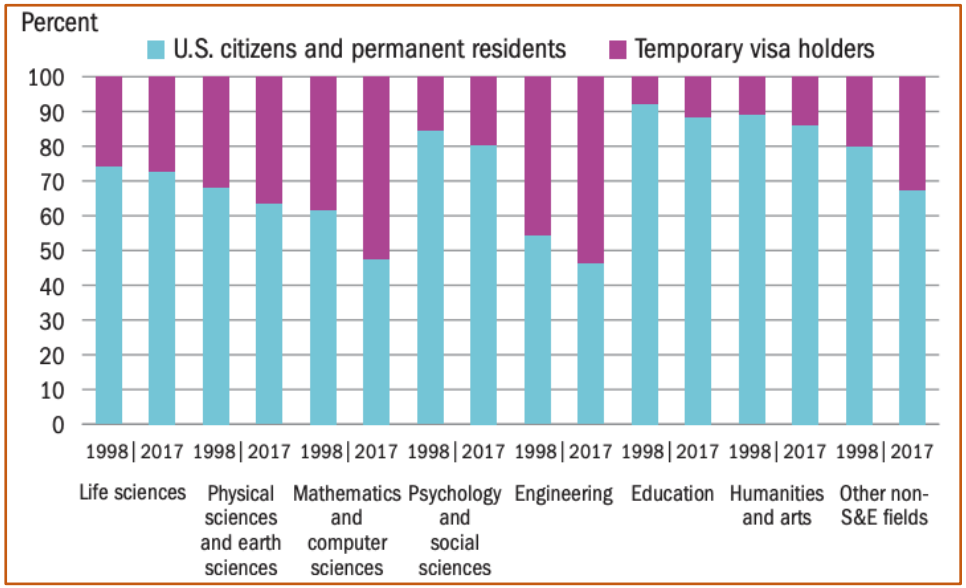
Before & After: Alternative to a Bar Chart



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.



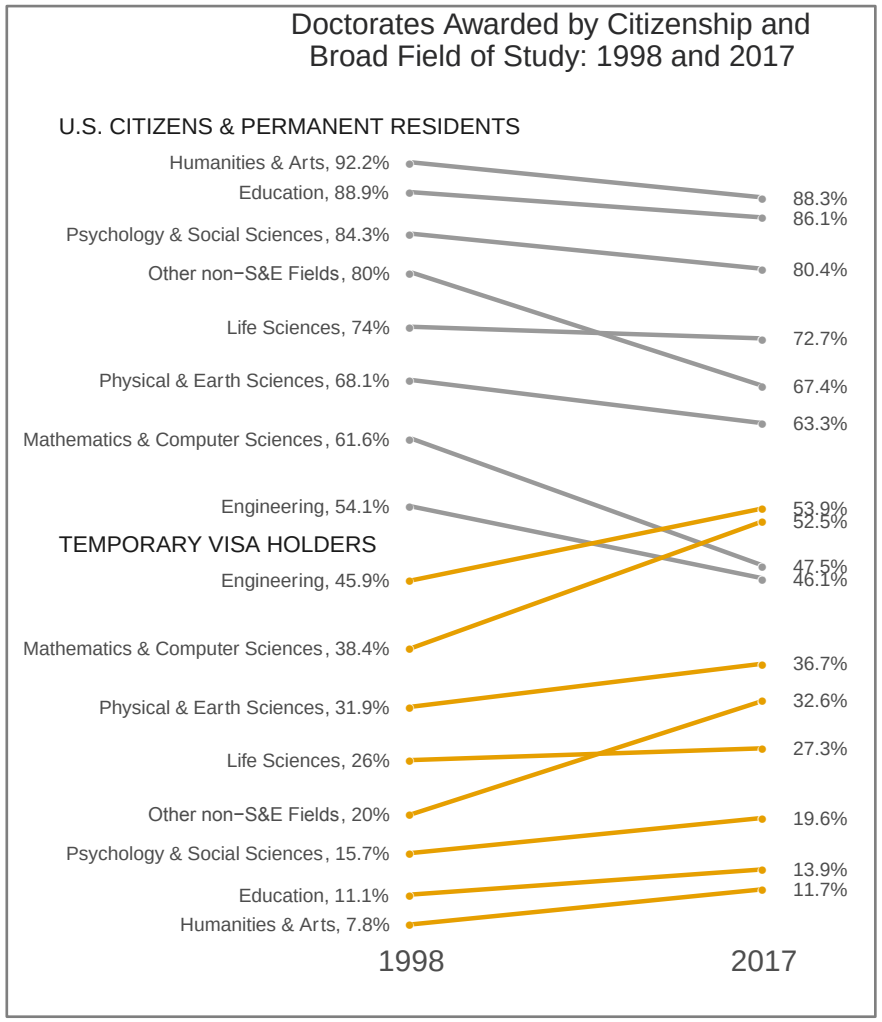
Before & After: Alternative to a Bar Chart

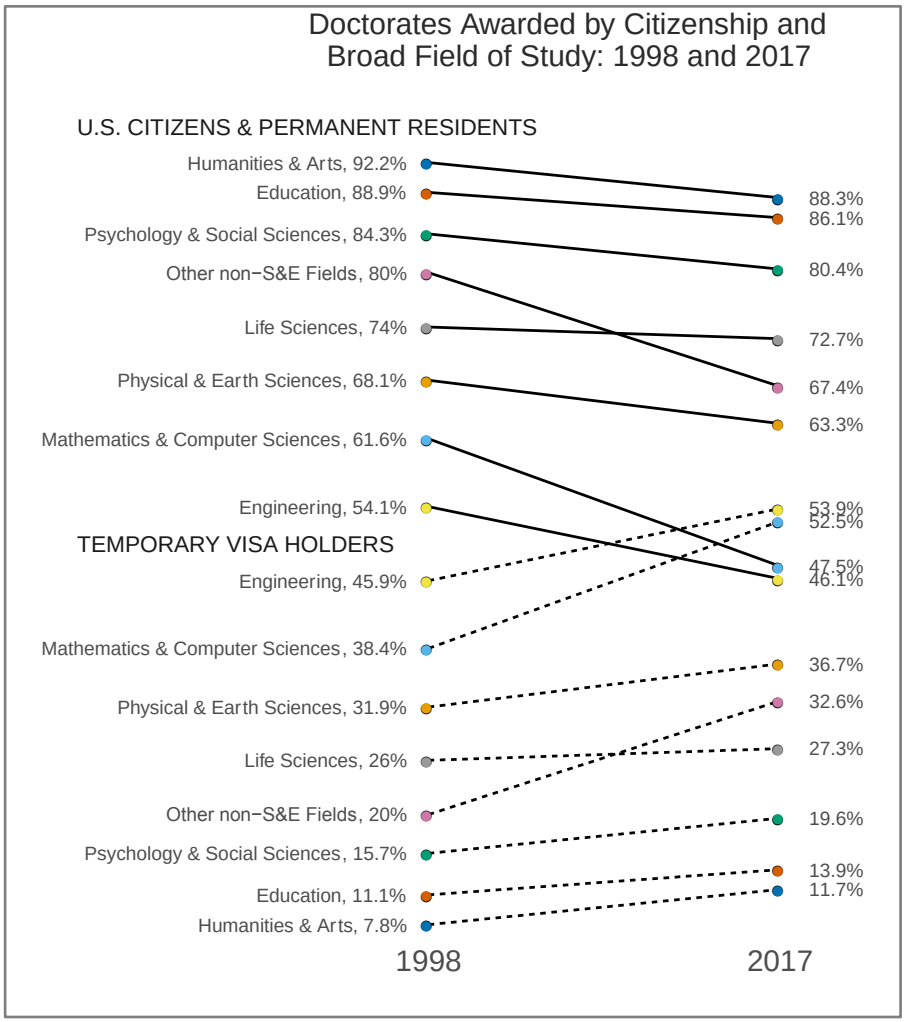


Two colors are being used to display three variables. The figure does not tell the story of the data without exerting a lot of mental energy. A suggestion is to display the data using a slope plot[3]. The slope plot contains two axes, one for each level of time (1998 and 2017). The broad fields of study 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

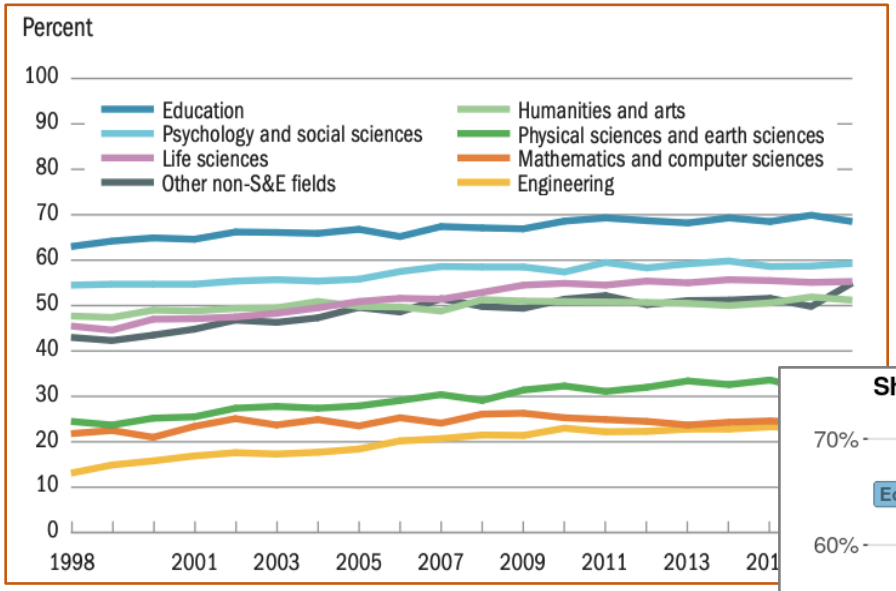
[3] Seeing Data http://seeingdata.org/taketime/inside-the-chart-slope-graph/ (Accessed May 5, 2019)

Before & After: Alternative to a Bar Chart

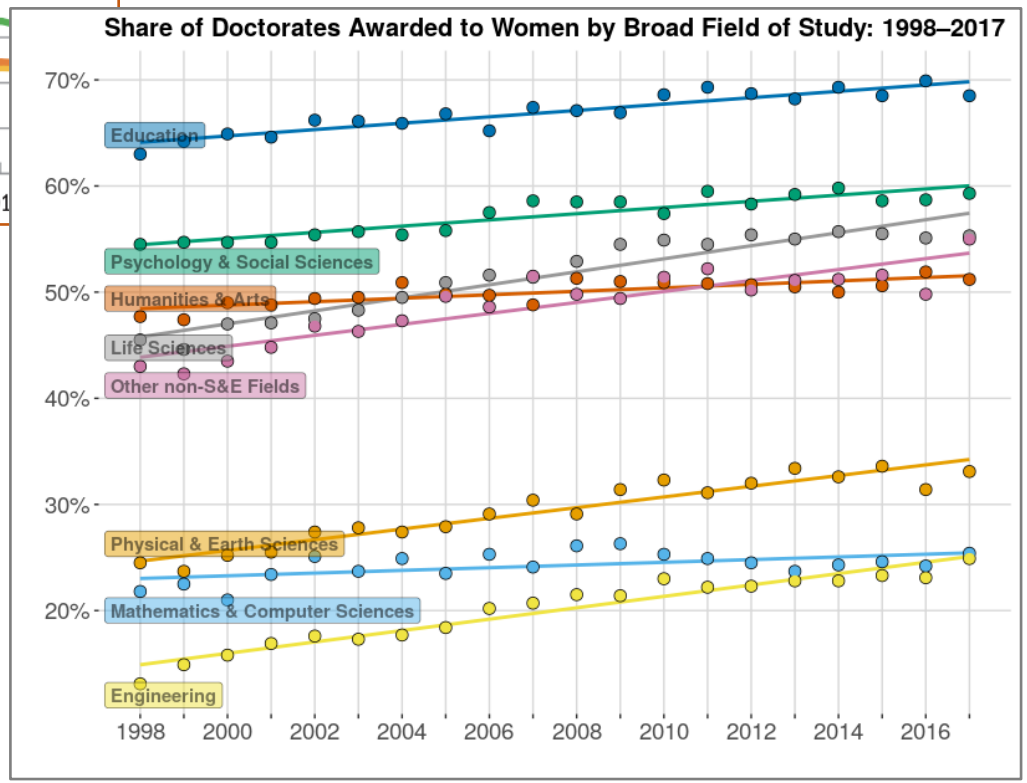




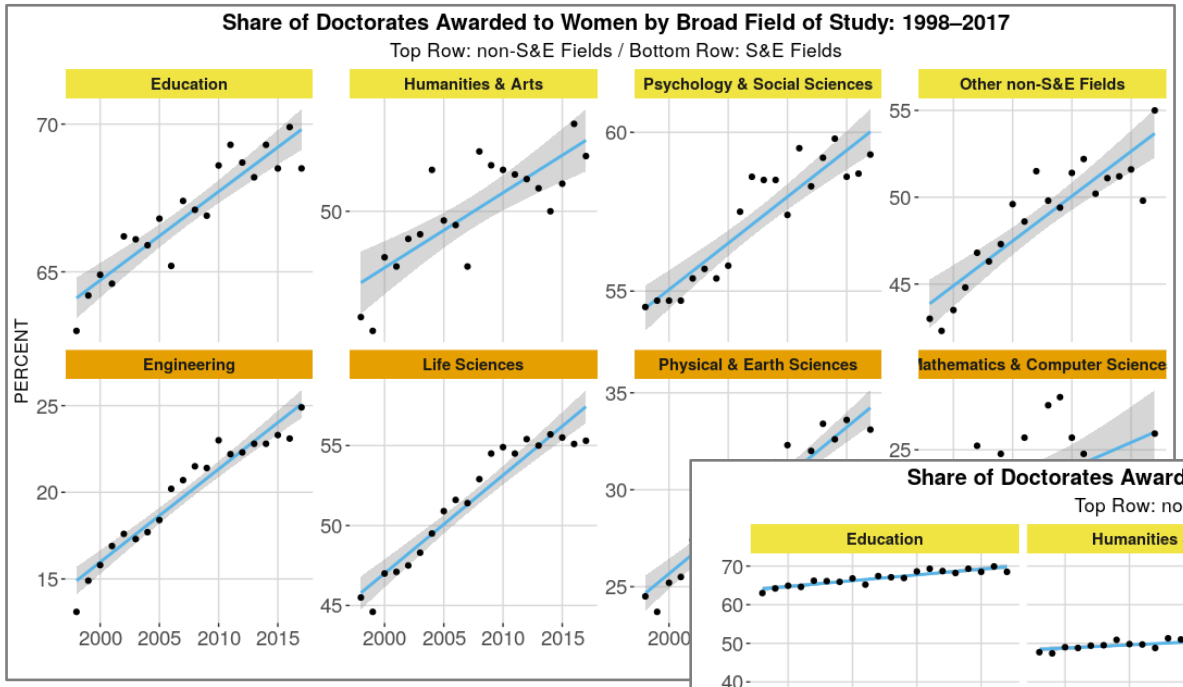
Before & After: Alternative to a Line Charts

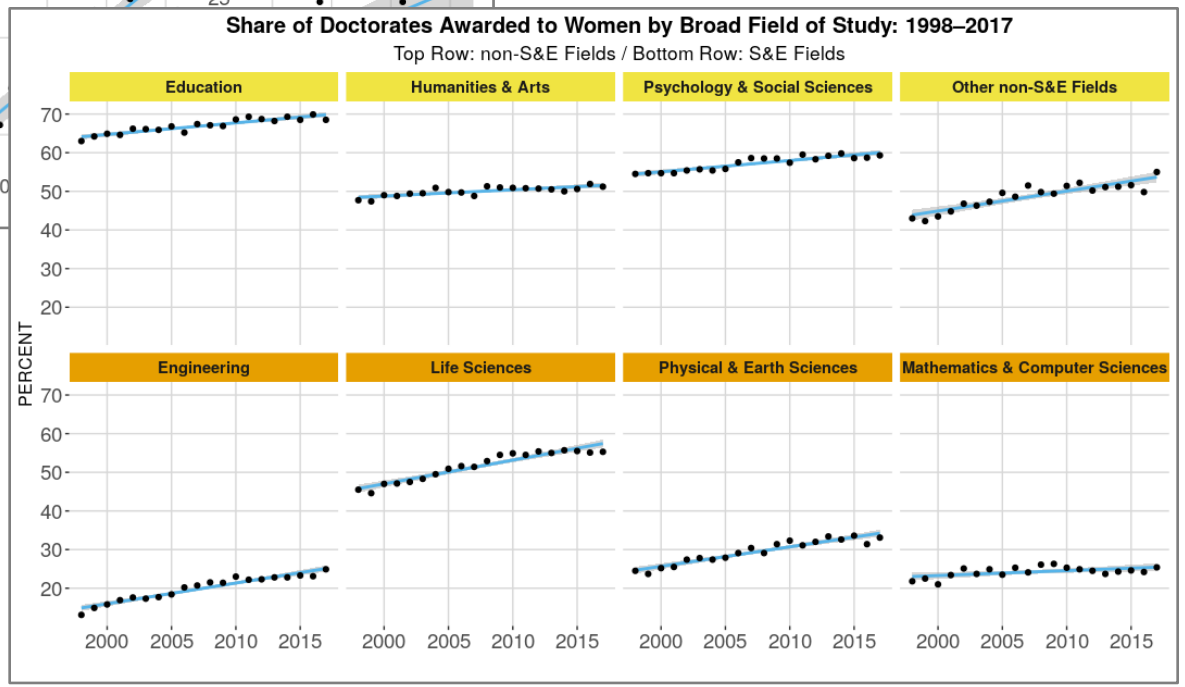


Reducing the white space would make it easier to make comparisons between the fields of study over time. A suggestion is to change the vertical axis from [0-100] to [10-70], to reduce the non-information white space.

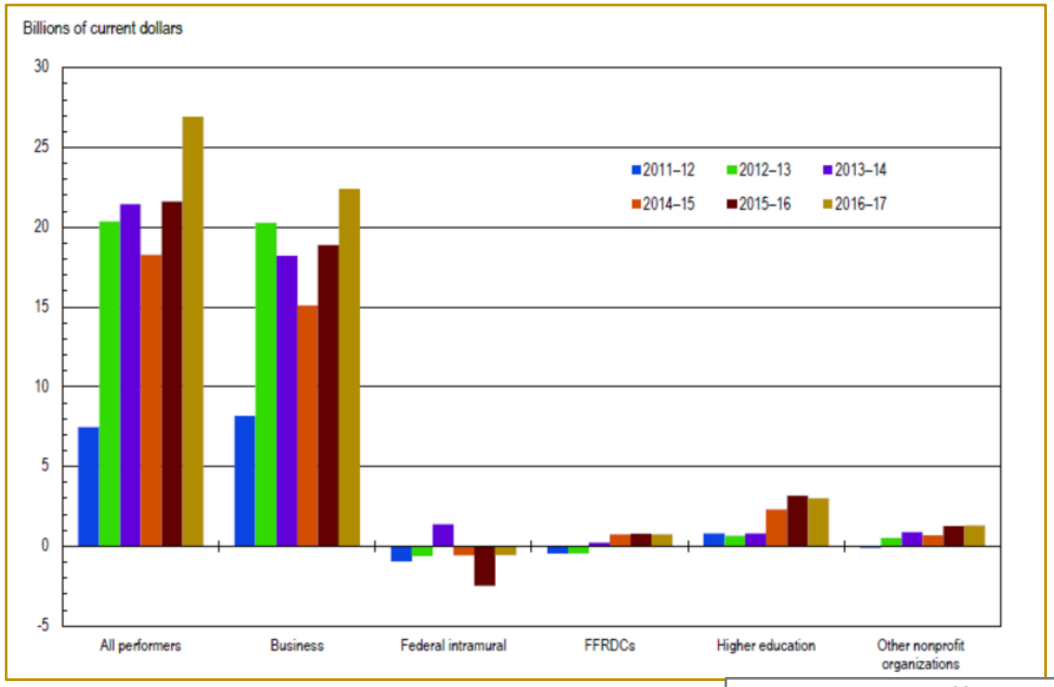


Before & After: Alternative to a Line Charts

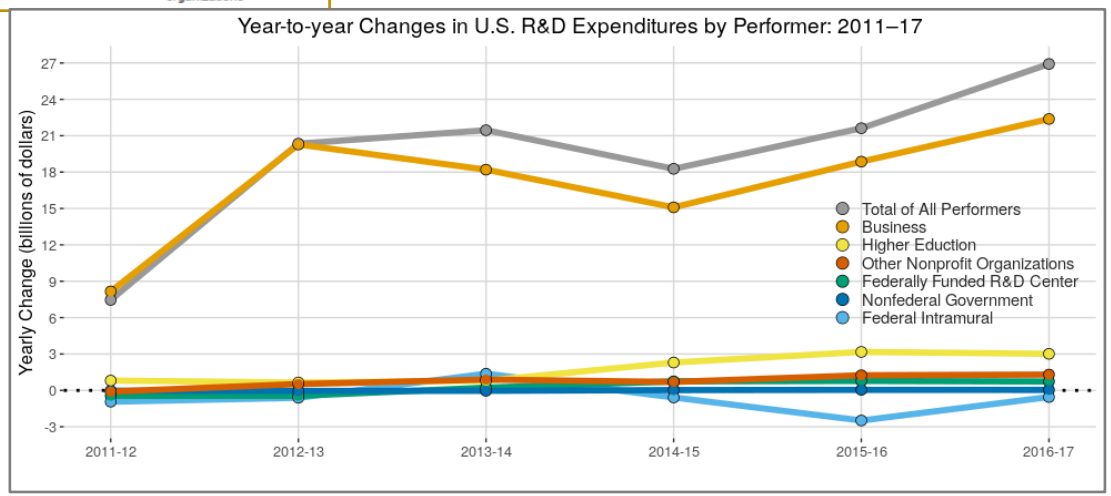




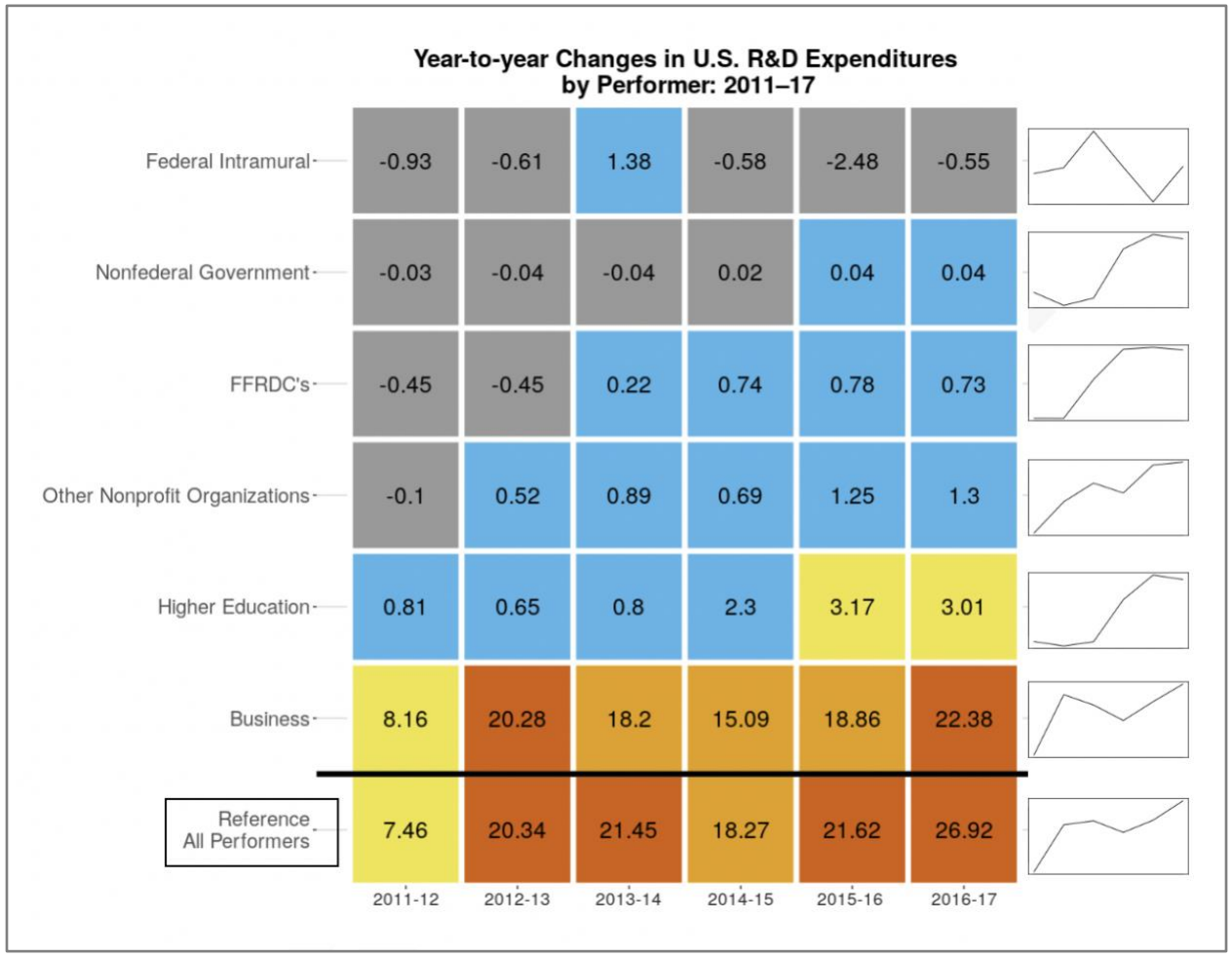
Before & After: Alternative to a Bar Chart



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.



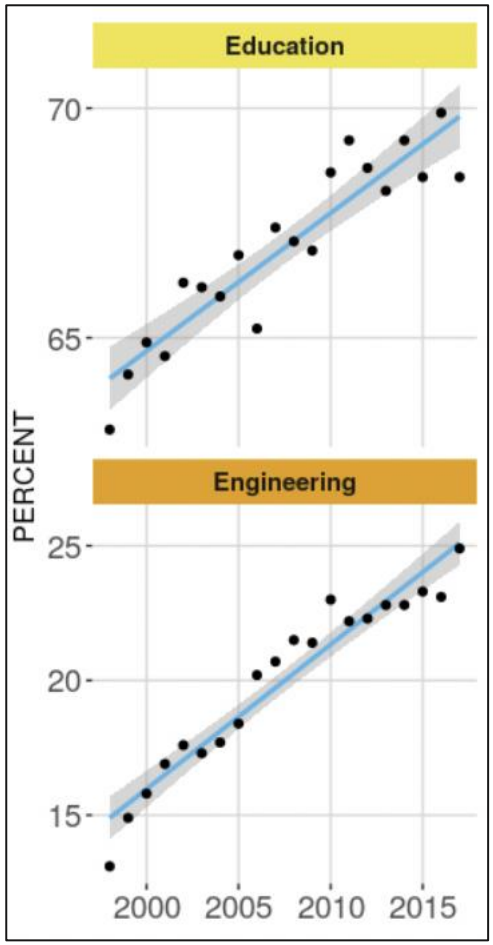
Before & After: Alternative to a Bar Chart



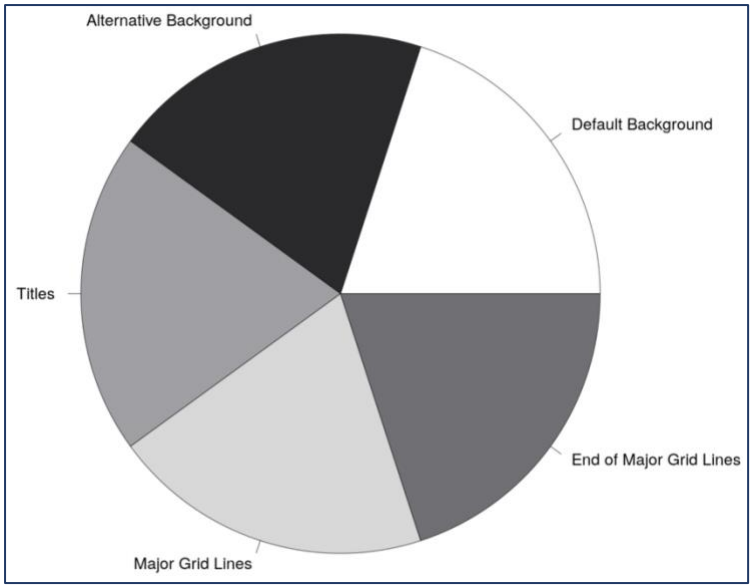
In cases where the values differ by an order of magnitude, one suggestion is to dichotomize the values and use a heat map which will display the categories using colors and can include the actual values. In the plot below trend lines are included to aid in interpretation.

theme_SDAD

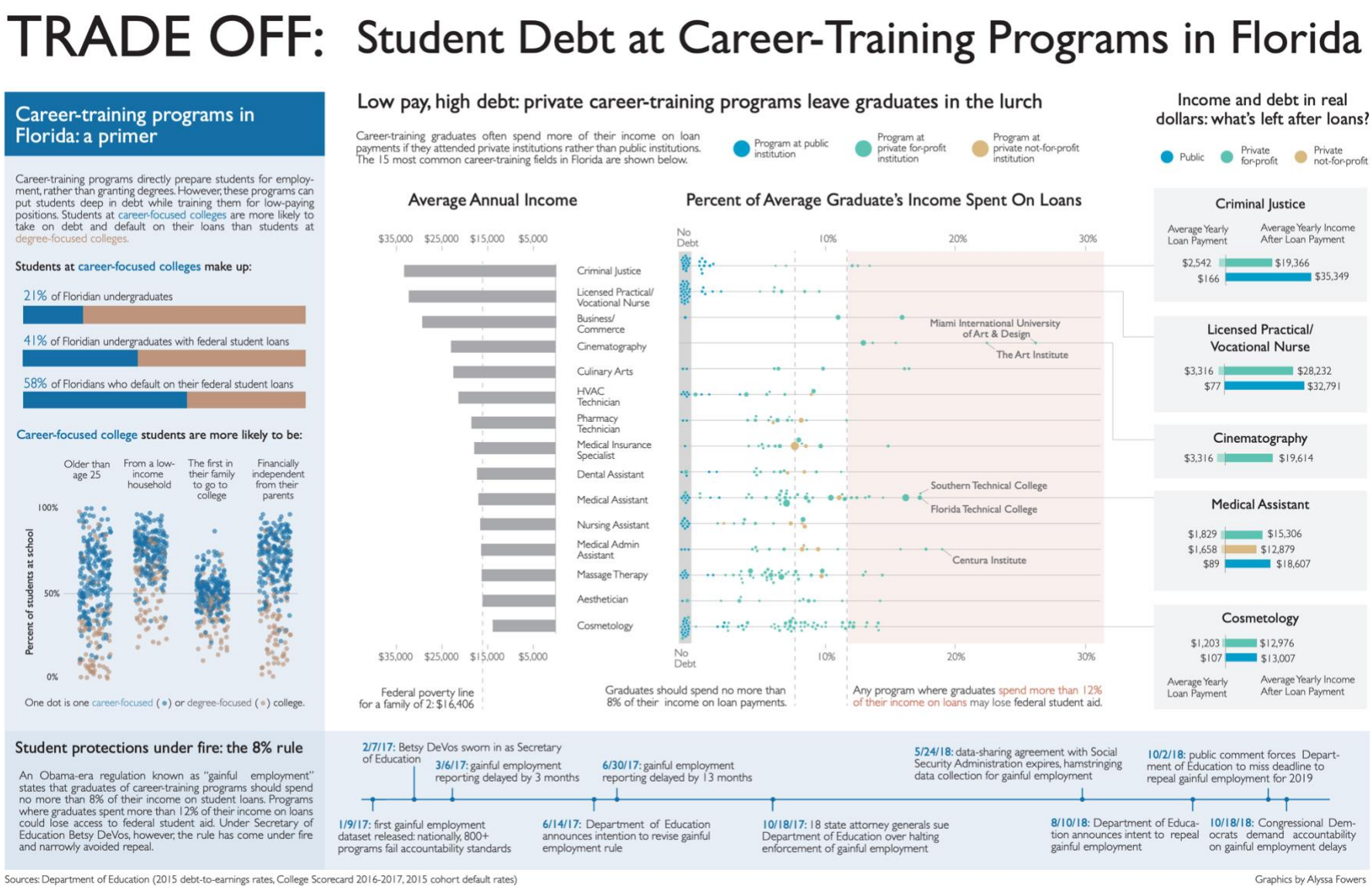
theme_SDAD was used to construct the previous graphs. It is very minimal, no background color, no axis lines, no box around the figure, ...



Colors Used in theme_SDAD



Good Example of a Poster



Online References for Color

Stephen Few's Practical Rules for Using Color in Charts

http://www.perceptualedge.com/articles/visual_business_intelligence/rules_for_using_color.pdf

Escaping RGBland: Selecting Colors for Statistical Graphics

https://eeecon.uibk.ac.at/~zeileis/papers/Zeileis+Hornik+Murrell-2009.pdf

R Colorspace Package

https://cran.r-project.org/web/packages/colorspace/vignettes/colorspace.html

R Dichromat Package for Color Blindness

http://australianantarcticdatacentre.github.io/GentleR/2015-07-23/dichromat.pdf

R Viridis Package

https://cran.r-project.org/web/packages/viridis/vignettes/intro-to-viridis.html
https://timogrossenbacher.ch/2016/12/beautiful-thematic-maps-with-ggplot2-only/

Online References for ggplot2 Themes

Themes included in ggplot2
https://www.r-graph-gallery.com/192-ggplot-themes/

The ggthemr package – Theme and colour your ggplot figures
https://www.shanelynn.ie/themes-and-colours-for-r-ggplots-with-ggthemr/

ggplot2 Themes Gallery
https://www.datanovia.com/en/blog/ggplot-themes-gallery/

Online References for Visualization Ideas

R you ready to make charts? (very clever and funny)
https://www.williamrchase.com/slides/ggplot_intro.html#27

Flowing Data (memberships available)
https://flowingdata.com/

Consumer Financial Protection Design Manual
https://cfpb.github.io/design-manual/data-visualization/data-visualization.html

Data Visualization by Kiernan Healy
https://socviz.co/index.html#preface

Data Visualization Catalog
https://datavizcatalogue.com/

Dataviz Project by ferdio
https://datavizproject.com/

Online References for 508 Compliance

508 Compliant Visualizations

https://www.hhs.gov/web/section-508/making-files-accessible/checklist/word/index.html

Making Your Charts and Graphs 508 Compliant - MSKTC

https://msktc.org/lib/docs/KT_Toolkit/Charts_and_Graphs/Charts_and_Graphics_508c.pdf