

# Econ 106

Lecture 18

slides derived from:

<https://r4ds.had.co.nz/graphics-for-communication.html>

# Reminders

- Lab 5 is due Sunday, 11:59pm (best 4 of 5 labs count)
- Final Project is due Sunday, 11:59pm

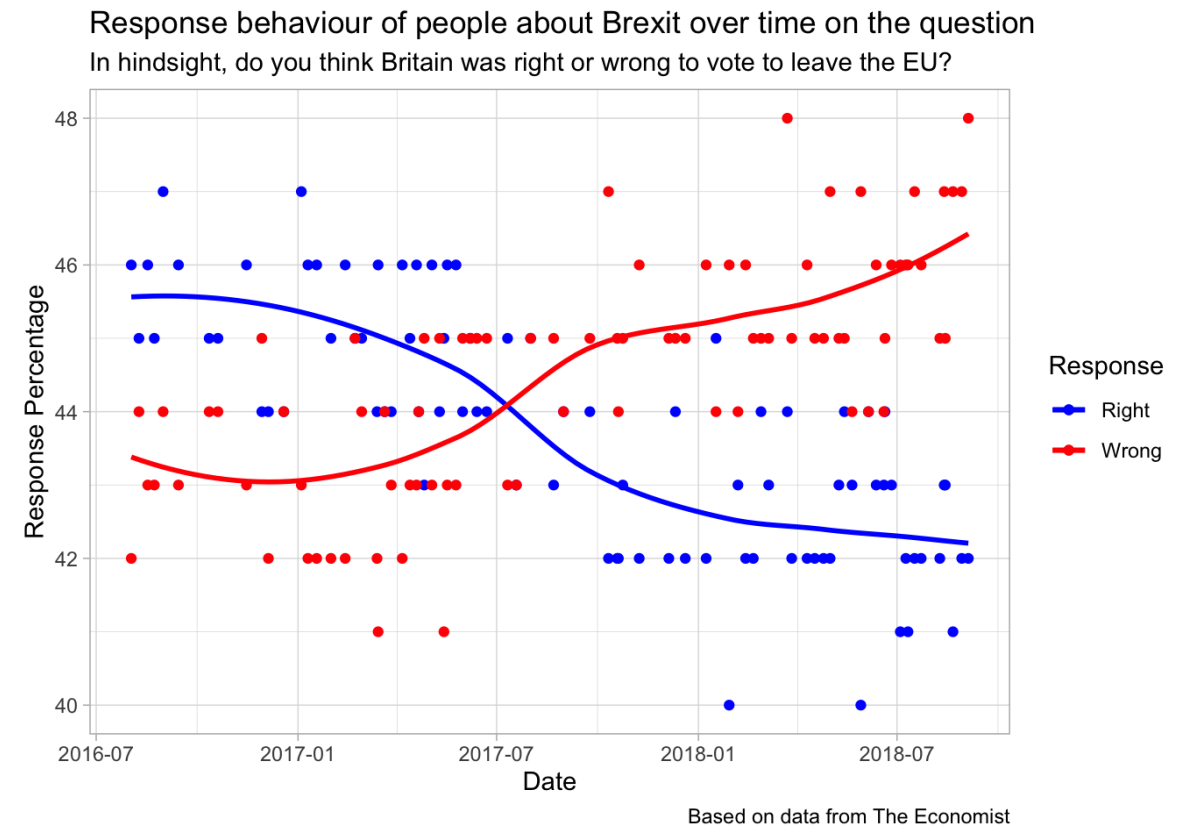
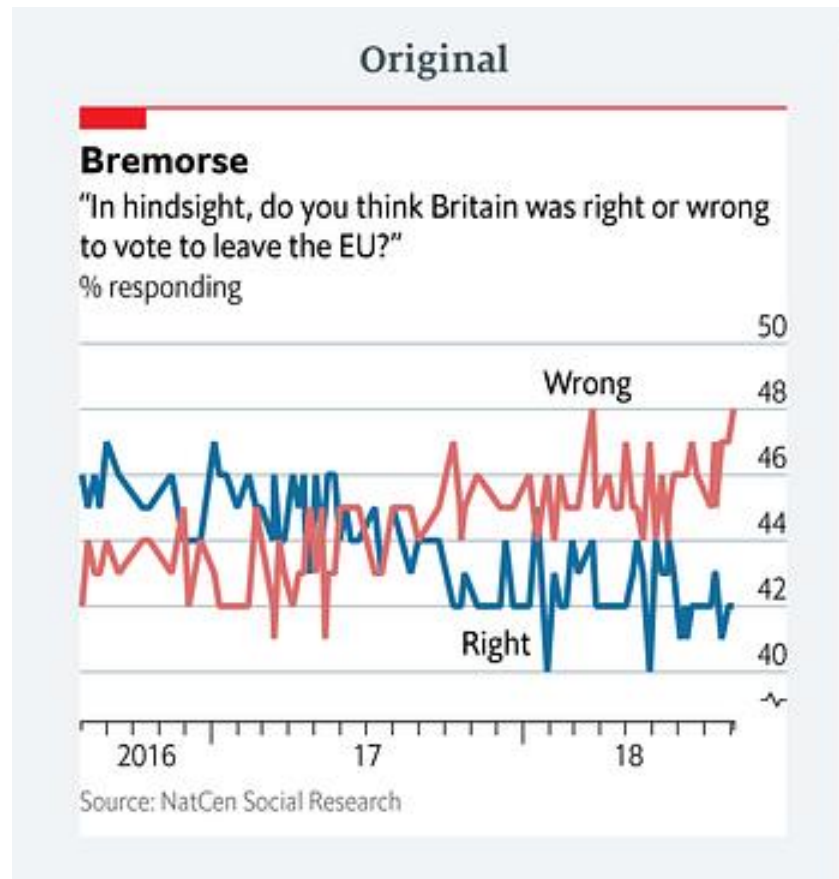
<https://pollev.com/vsovero>

# Extra Credit Opportunity

- Please fill out a teaching evaluation: <https://ieval.ucr.edu/ieval/login>
- Extra credit will be based on the class response rate:
  - 25% response rate → add .25% to your overall grade
  - 50% response rate → add .50% to your overall grade
  - 75% response rate → add .75% to your overall grade
  - 100% response rate → add 1% to your overall grade
- Deadline: Friday 11:59pm

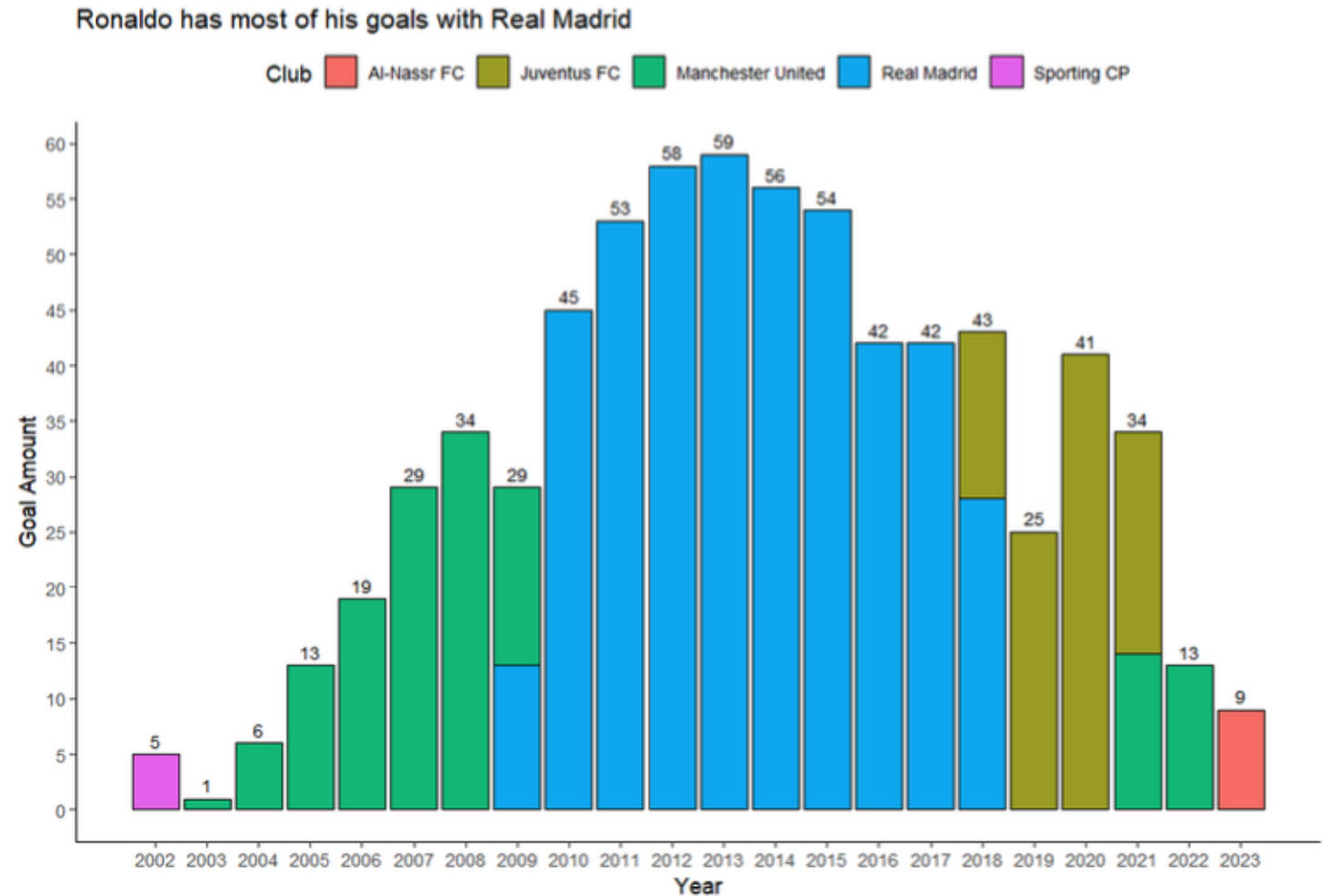
# #tidytuesday

Inspired by the Economist article [here](#):



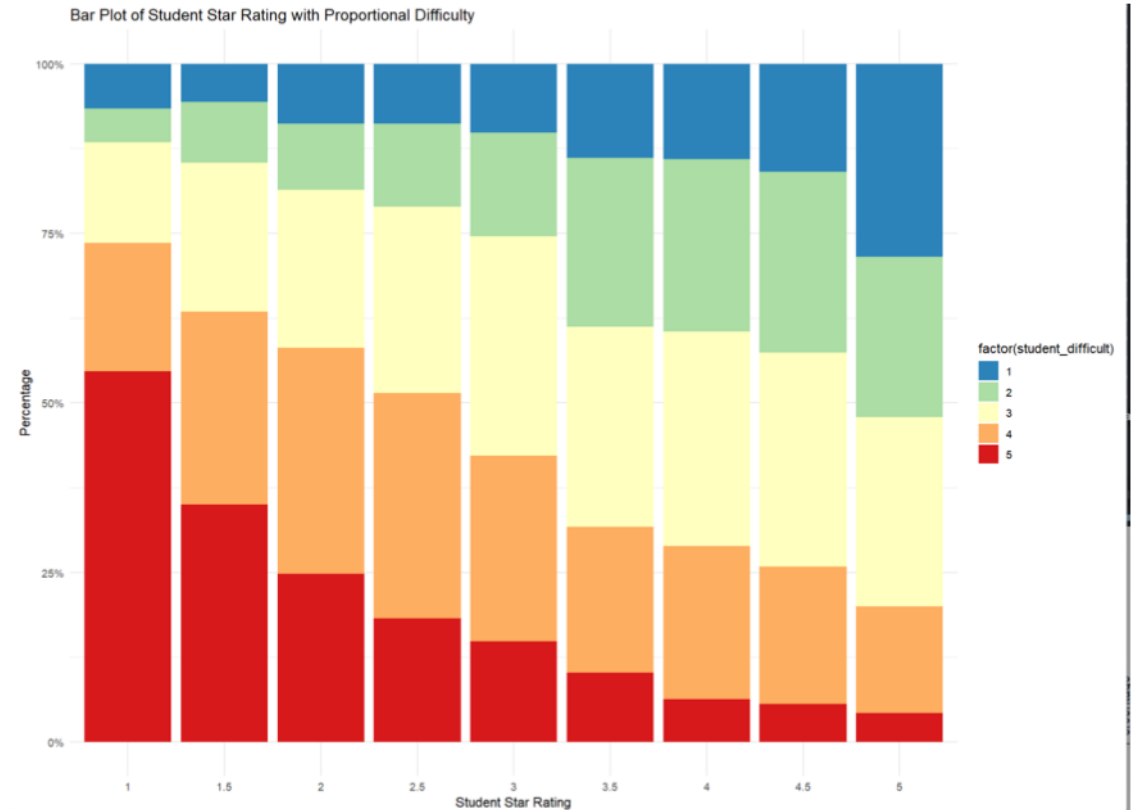
# Miscellaneous Student Examples

# Histogram with color mapping



# Bar Plot with Color Mapping

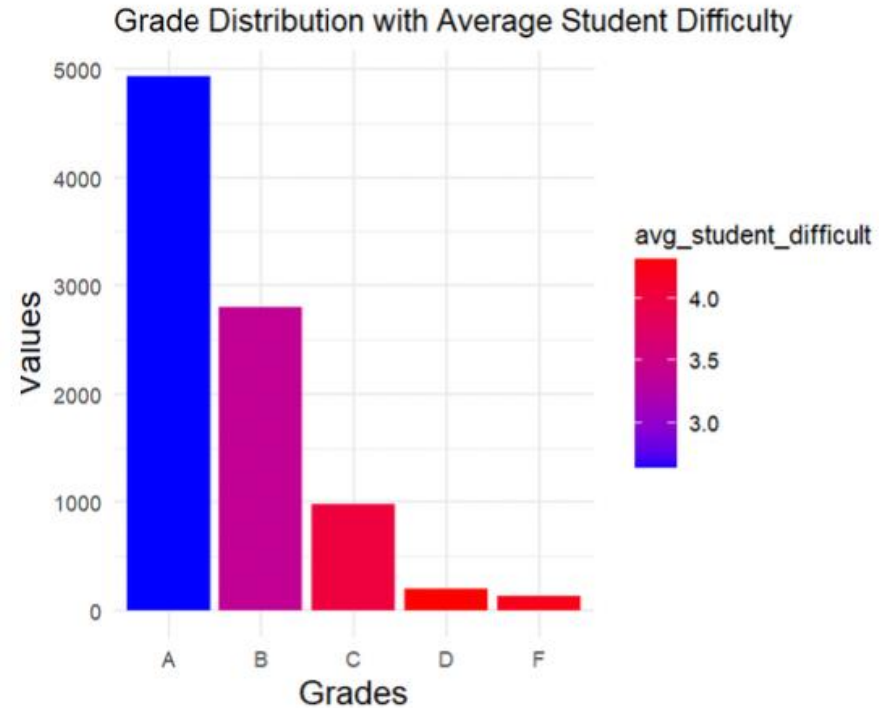
Title should describe the findings in the chart, otherwise great use of color and labeling (I like the caption with the data source)



Source: [Big Data Set from RateMyProfessor.com for Professors' Teaching Evaluation - Mendeley Data](#)

# Bar Plot with Color Mapping

Interesting use of color  
(color is a gradient based on  
the difficulty rating)



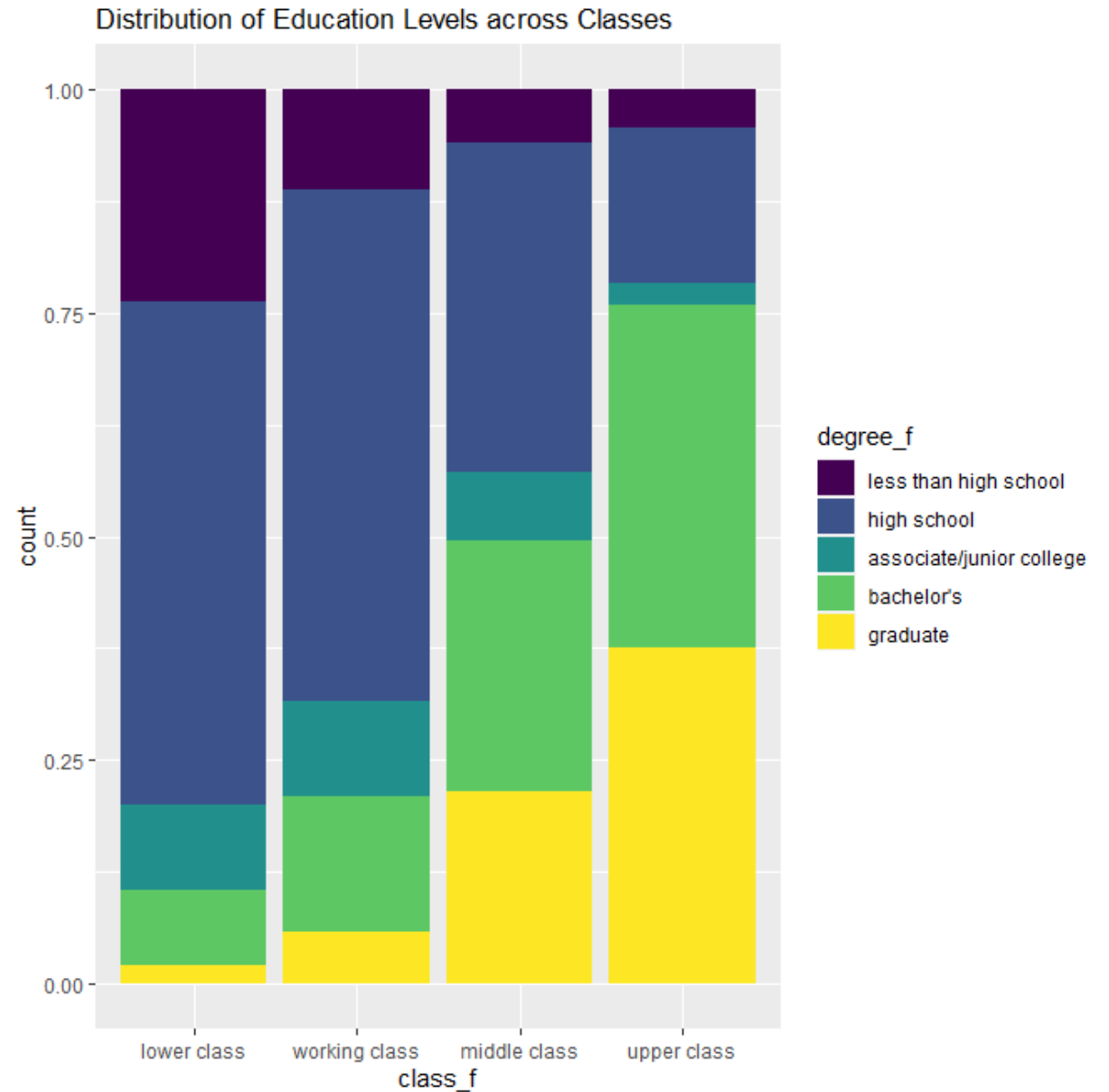
Source: [Big Data Set from RateMyProfessor.com for Professors' Teaching Evaluation - Mendeley Data](#)



- More education often leads to higher income
- Education can also show people ways of moving up in social class
- Higher education is expensive

Thoughtful interpretation of the findings  
(upper class has higher levels of educational attainment)

Good use of color gradient (viridis)



# Providing Context

Student provides additional context for the findings in the chart (uses knowledge of sneaker related events)

## 04 Visualization 1: Nike & Adidas Sales Over Time

Stockx Data

### Off White

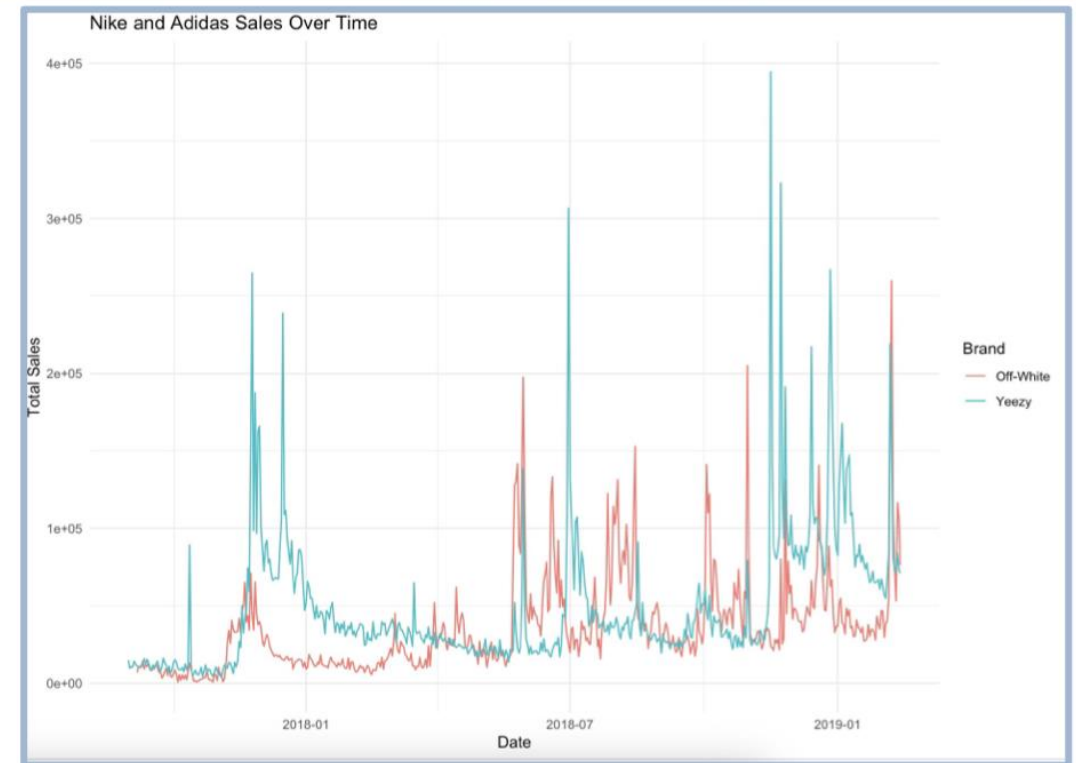
Has some spikes late 2018, but its highest is right after JAN 2019

Virgil joined Louis Vuitton

### Yeezy

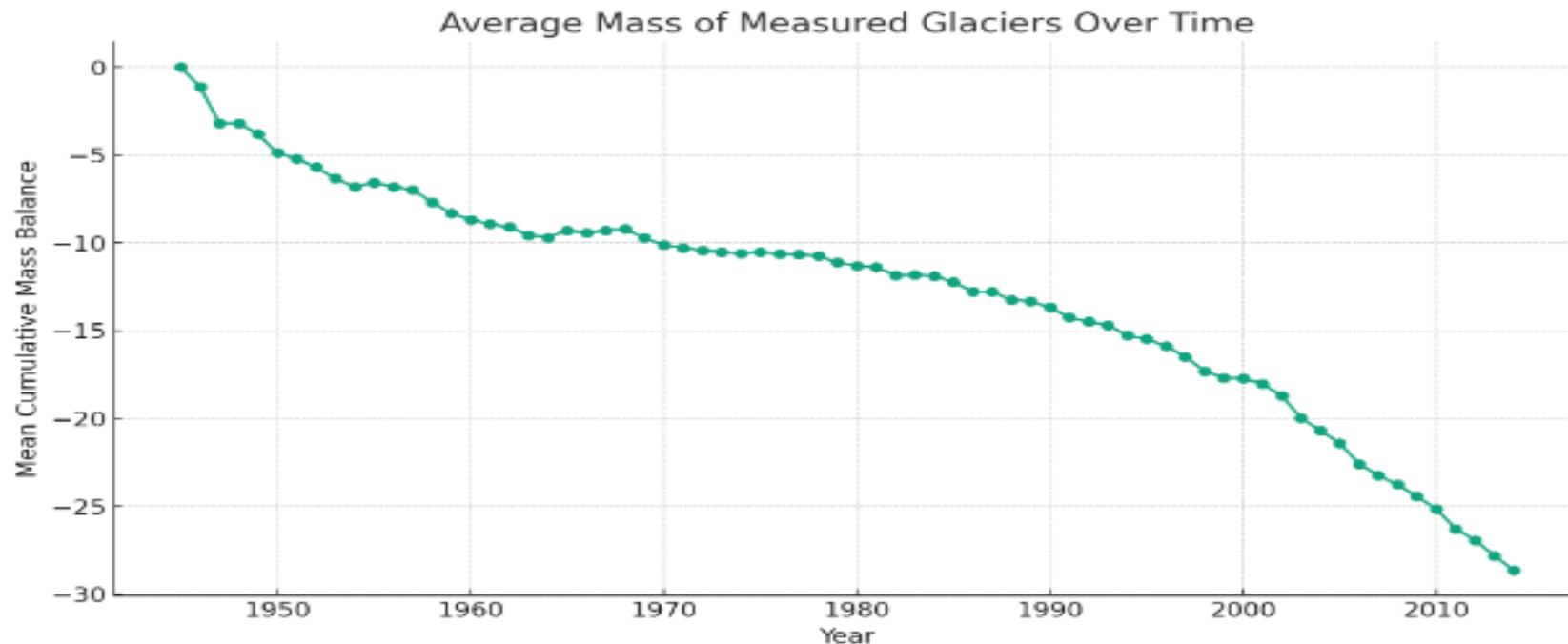
Sales overpower Off white, almost double, biggest spike, late 2018

Kanye dropped 2 albums : Kids See Ghost and Ye



# Line Graph (Trends over Time)

It's not that impressive to see a single line graph- should try to group by an additional variable



# Don't do this

The plot showing the average mass of measured glaciers over time, based on the dataset you provided. The plot illustrates the trend in the mean cumulative mass balance of glaciers from 1945 onwards.

One of the most direct effects of melting glaciers is the rise in sea levels. Glaciers store a large amount of the Earth's freshwater, and when they melt, this water flows into the oceans. This contributes to the overall increase in sea levels, which can lead to the flooding of coastal areas, affecting millions of people worldwide.

Glaciers are important habitats for unique ecosystems. As they retreat, these ecosystems can be lost.

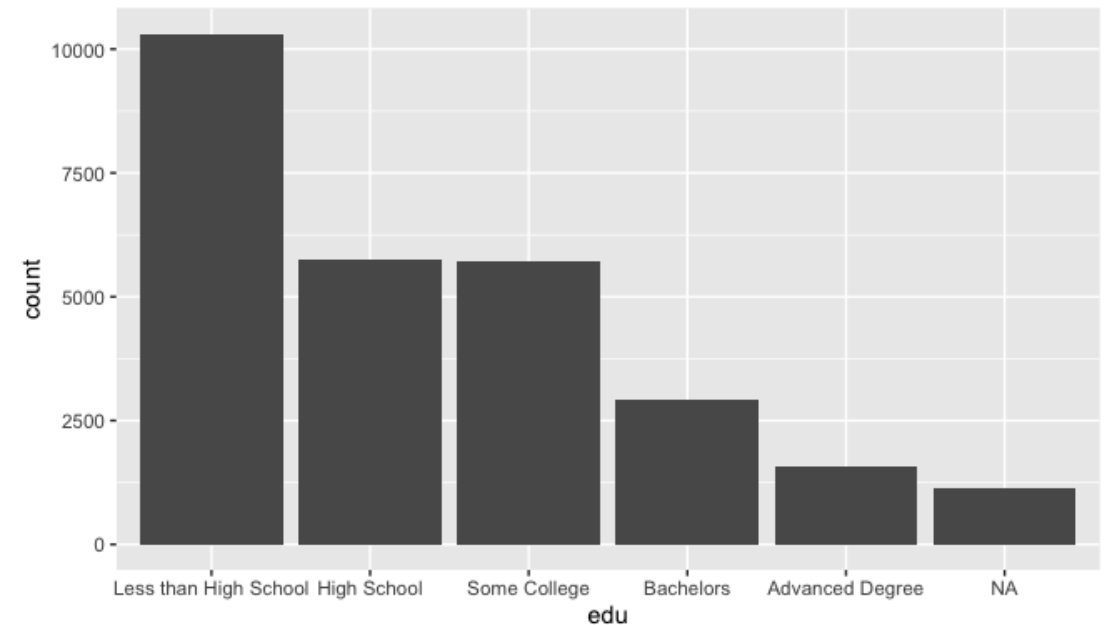
# Outline

- Visualization Best Practices
- Customizing your plots

# Bar Plot Customizations

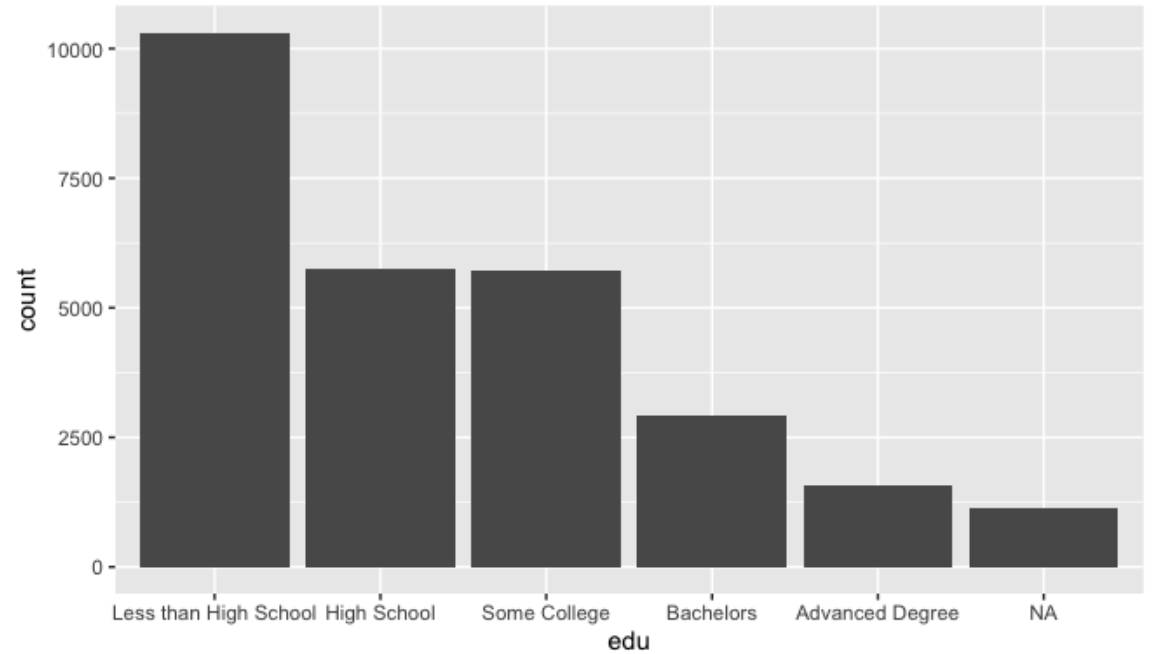
- What can we do to improve this bar plot?

```
ggplot(data=acs,  
mapping=aes(x=edu))+  
geom_bar() +
```



# Class Exercise

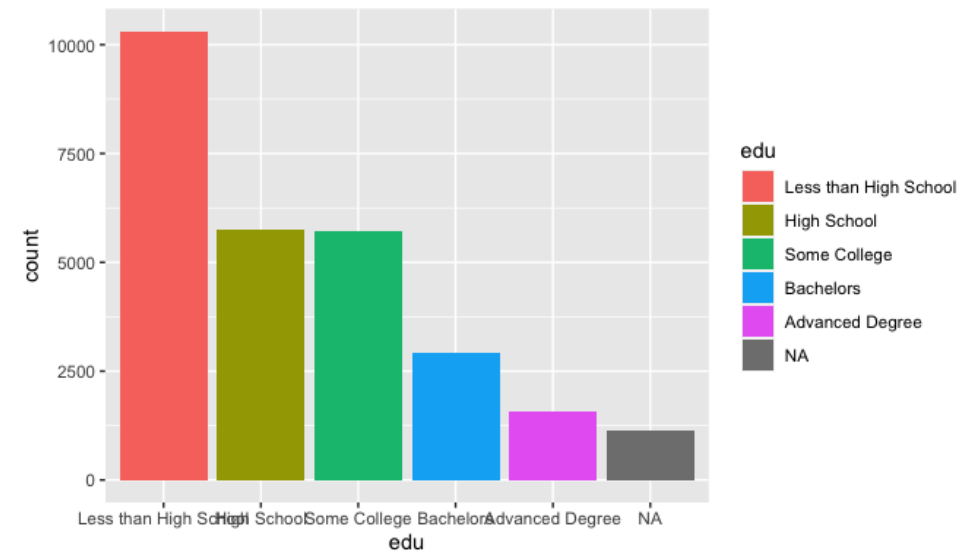
- improve on this bar plot by adjusting:
  - labels
  - scales
  - theme



# Bar Plot Customizations: add color

- We use the fill argument to add color to the bar chart
- ggplot will choose the default color palette

```
ggplot(data=acs,  
mapping= aes(x=edu))+  
geom_bar ( aes(fill=edu))
```

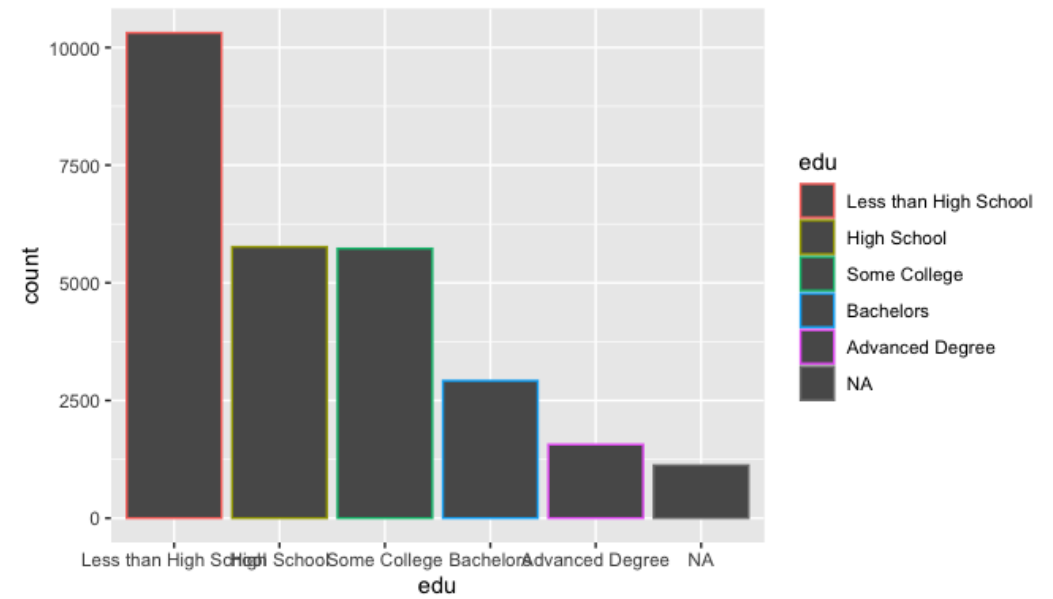




# Bar Plot Customizations: fill vs color

- Remember to use **fill** instead of **color** for 2d shapes
- otherwise only the outline gets colored

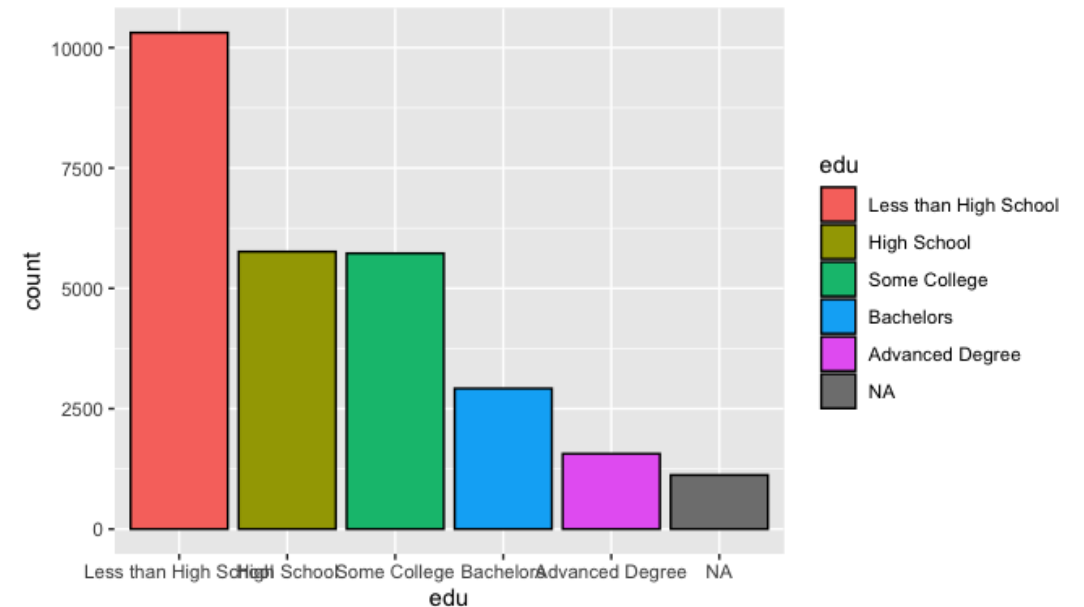
```
ggplot(data=acs,  
mapping= aes(x=edu))+  
geom_bar ( aes(color=edu))
```



# Bar Plot Customizations: fill vs color

- can we use **fill** and **color**?
- Sure!
- I added a black border in this example

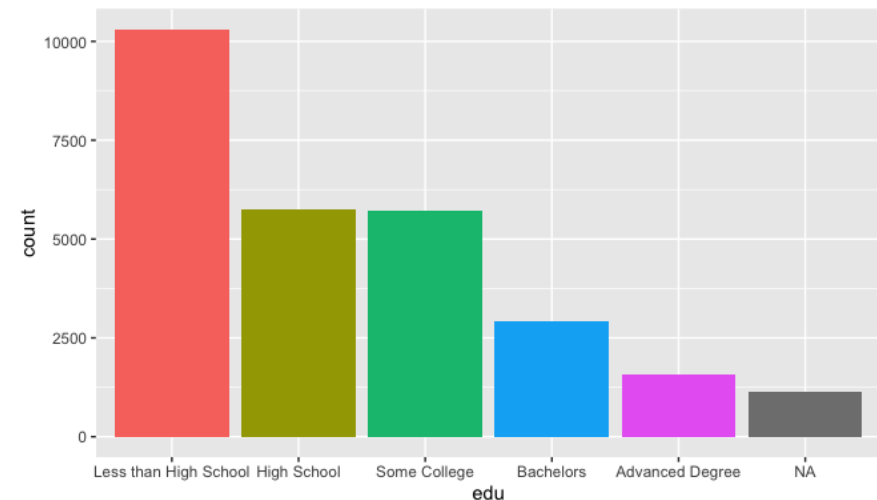
```
ggplot(data=acs,  
mapping= aes(x=edu))+  
geom_bar ( aes(fill=edu), color = "black" )
```



# Bar Plot Customizations: Remove legend

- We can remove the legend because it doesn't provide any additional information
- **legend.position** controls where the legend is drawn:
  - right (the default)
  - left
  - top
  - bottom
  - none (suppress legend)

```
ggplot(data=acs,  
mapping= aes(x=edu))+  
geom_bar ( aes(fill=edu)) +  
theme(legend.position = 'none')
```



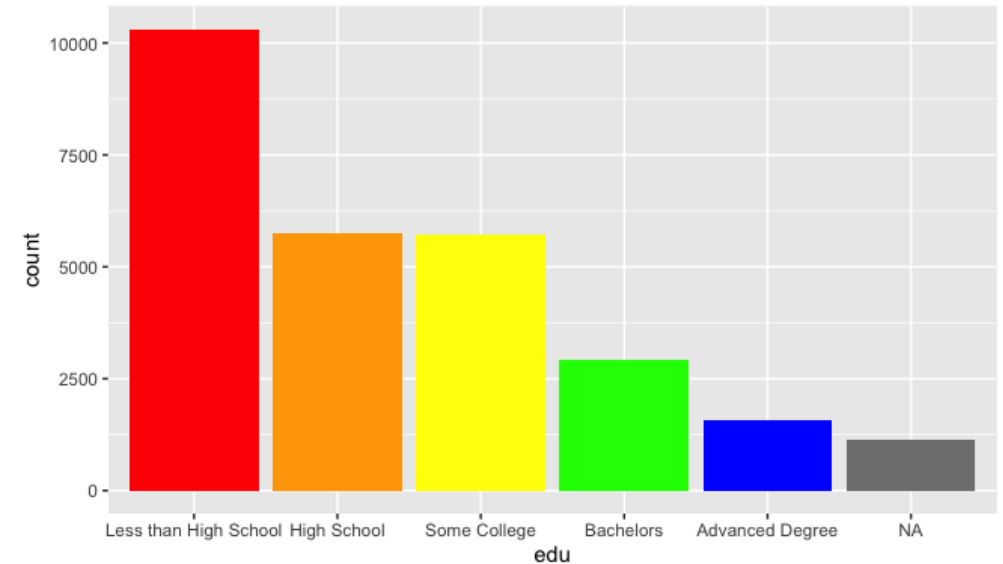
# Color Scales for Categorical Variables

- To adjust color on a 1-d shape (point), we can use:
  - `scale_color_manual()`: manually create color scale
  - `scale_color_brewer()`: use a ColorBrewer palette
- To adjust color on a 2-d shape (bar), we can use:
  - `scale_fill_manual()`: manually create color scale
  - `scale_fill_brewer()`: use a ColorBrewer palette

# Manually Select Color Scale

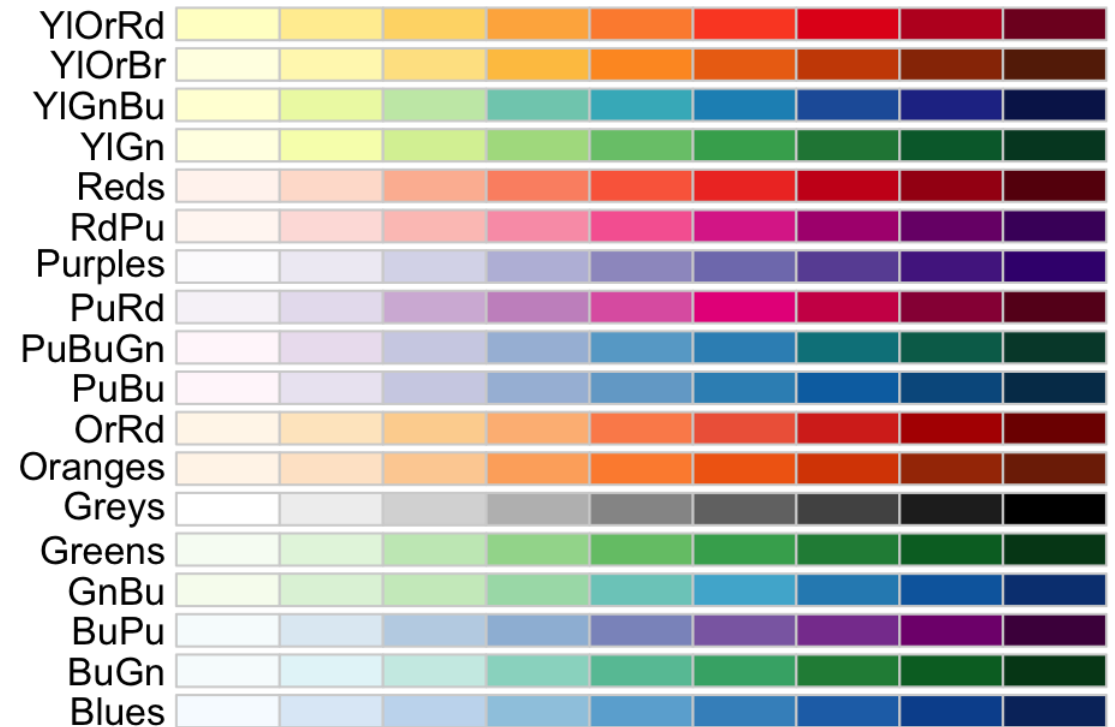
- Note: the color for NA can't be adjusted manually

```
ggplot(data=acs,  
  mapping= aes(x=edu))+  
  geom_bar(aes(fill=edu))+  
  theme(legend.position = 'none') +  
  scale_fill_manual(values= c("red", "orange", "yellow",  
    "green", "blue"))
```



# ColorBrewer Sequential Color Scales

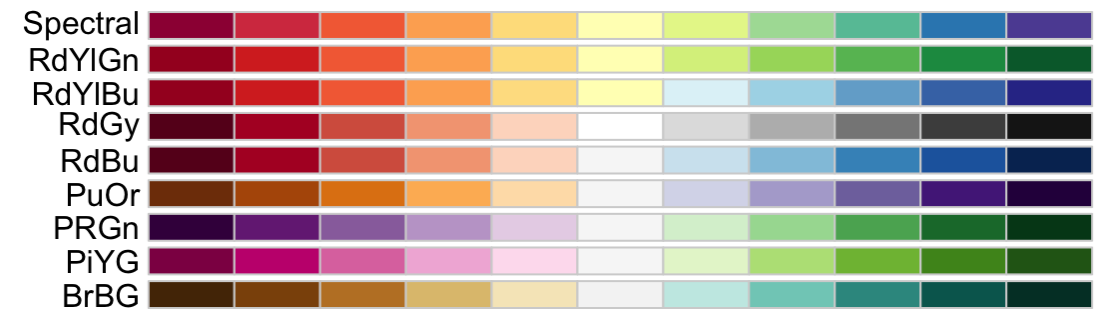
- ColorBrewer provides sets of colors (palettes)
- Sequential palettes are good for ordinal categorical variables
- Educational levels:
  - high school
  - college
  - graduate school



# ColorBrewer Diverging Color Scales

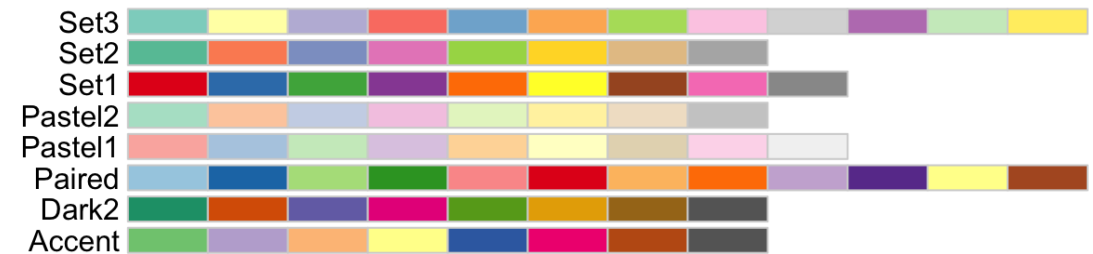
Diverging palettes are good for ordinal categorical variables

- Use this when your values are ordered in two directions relative to a center.
- political affiliation:
  - liberal
  - centrist
  - conservative



# ColorBrewer Qualitative Color Scales

- Qualitative (nominal) palettes are good for categorical Variables whose values have no ordering.
- Major:
  - Economics
  - Business
  - Statistics

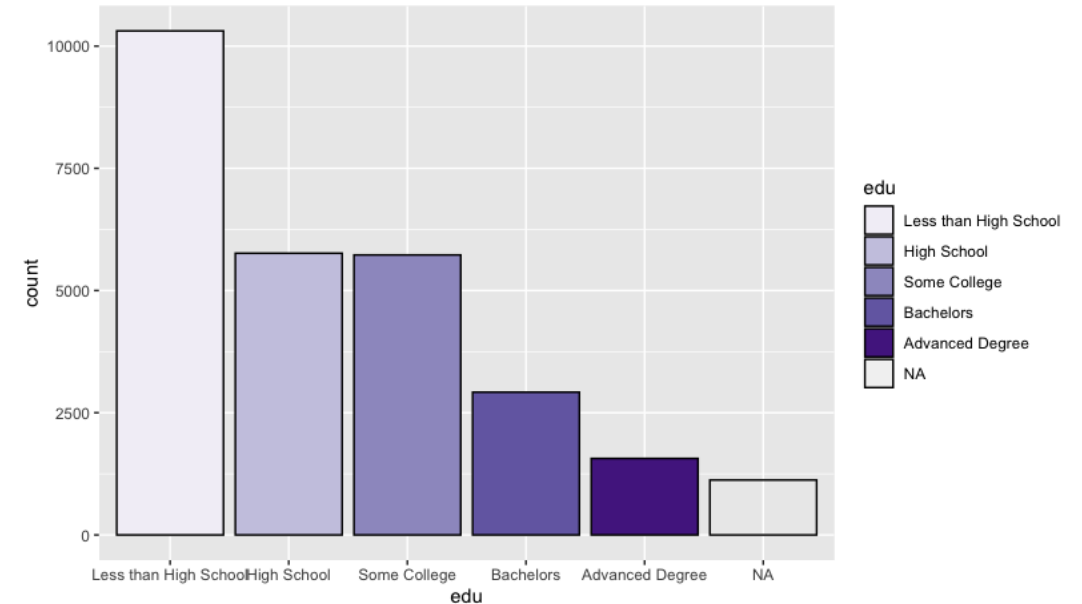




# Change the Palette

- Education is ordinal, so I will select a sequential color scale

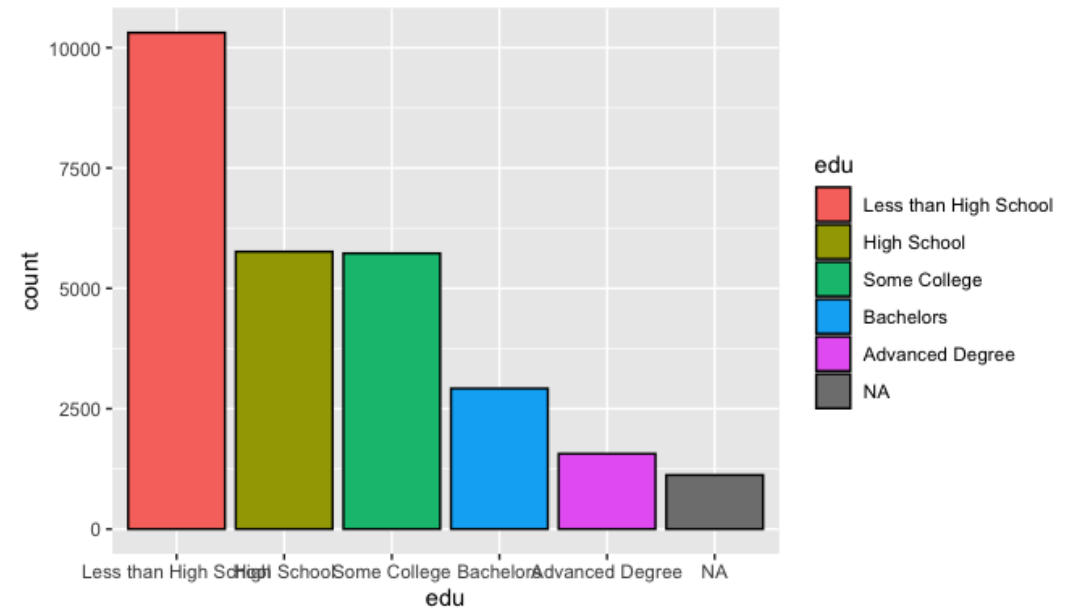
```
ggplot(data=acs,  
mapping= aes(x=edu))+  
geom_bar(aes(fill=edu), color = "black" )+  
scale_fill_brewer(palette= "Purples")
```



# What happened?

- I changed the palette for 1d objects (lines and points), not 2d objects

```
ggplot(data=acs,  
mapping= aes(x=edu))+  
geom_bar(aes(fill=edu), color = "black" )+  
scale_color_brewer(palette= "Purples")
```



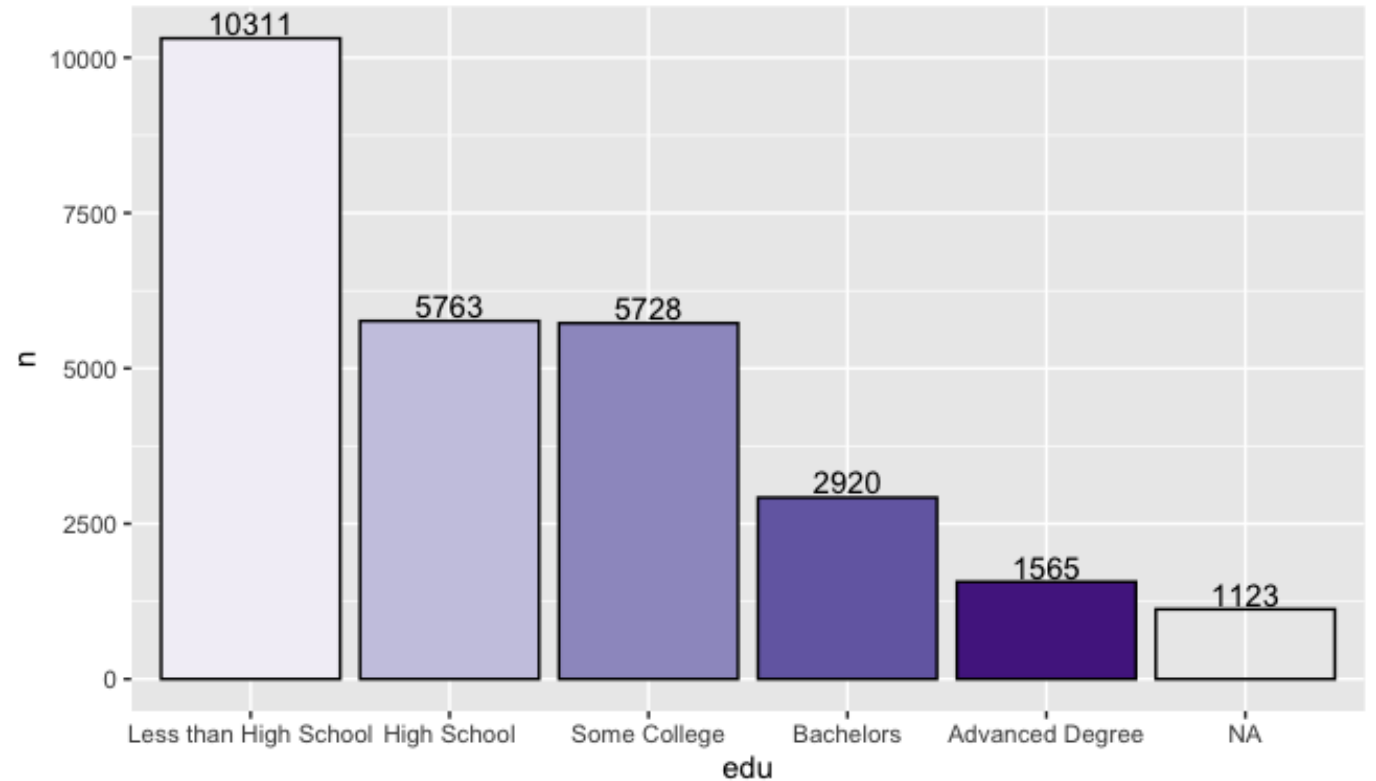
# Class Exercise

- Make a bar plot of race
- use the Set1 color palette

<https://pollev.com/vsovero>

# Adding text labels

- Adding the values of the counts provides more detail to our bar plots



# Step 1: Make a Frequency Table

- This gives us the values to add to the bar plot

```
degree_frequency<-acs%>%  
count(edu)
```

	edu	n
1	Less than High School	10311
2	High School	5763
3	Some College	5728
4	Bachelors	2920
5	Advanced Degree	1565
6	NA	1123

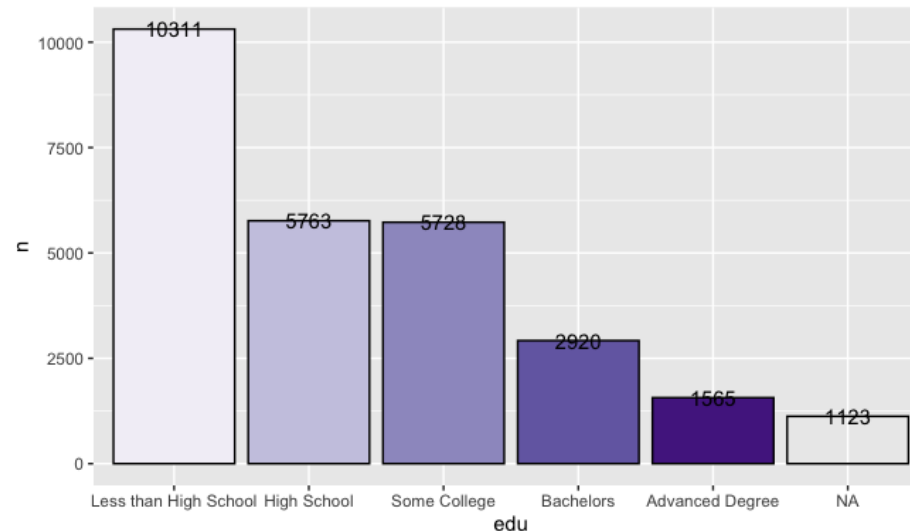
## Step 2: Make a Bar Plot with `geom_col()`

- We use `geom_col()` instead of `geom_bar()`
- This means we use the frequency table as our data
- `geom_text()` adds a label at the y-values

	edu	n
1	Less than High School	10311
2	High School	5763
3	Some College	5728
4	Bachelors	2920
5	Advanced Degree	1565
6	NA	1123



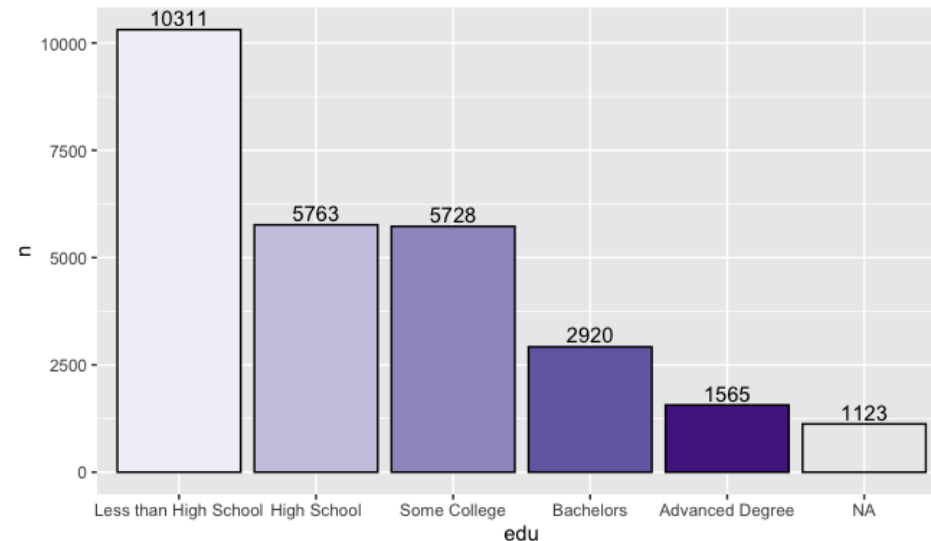
```
ggplot(data=degree_frequency,  
mapping=aes(x=edu, y=n))+  
geom_col(aes(fill=edu), color="black")+  
geom_text(aes(label=n))+  
scale_color_brewer(palette="Purples")+  
theme(legend.position='none') +
```



# Adjust the position of the text

- We can adjust the position of the text using `vjust` and `hjust`

```
ggplot(data=degree_frequency,  
mapping= aes(x=edu, y=n))+  
geom_col(aes(fill=edu), color = "black" )+  
geom_text(aes(label=n, vjust=-.3 )+  
scale_color_brewer(palette= "Purples" )+  
theme(legend.position = 'none') +
```



# Class Exercise

- Make a bar plot of race
- use the Set1 color palette
- add the counts to the top of each bar

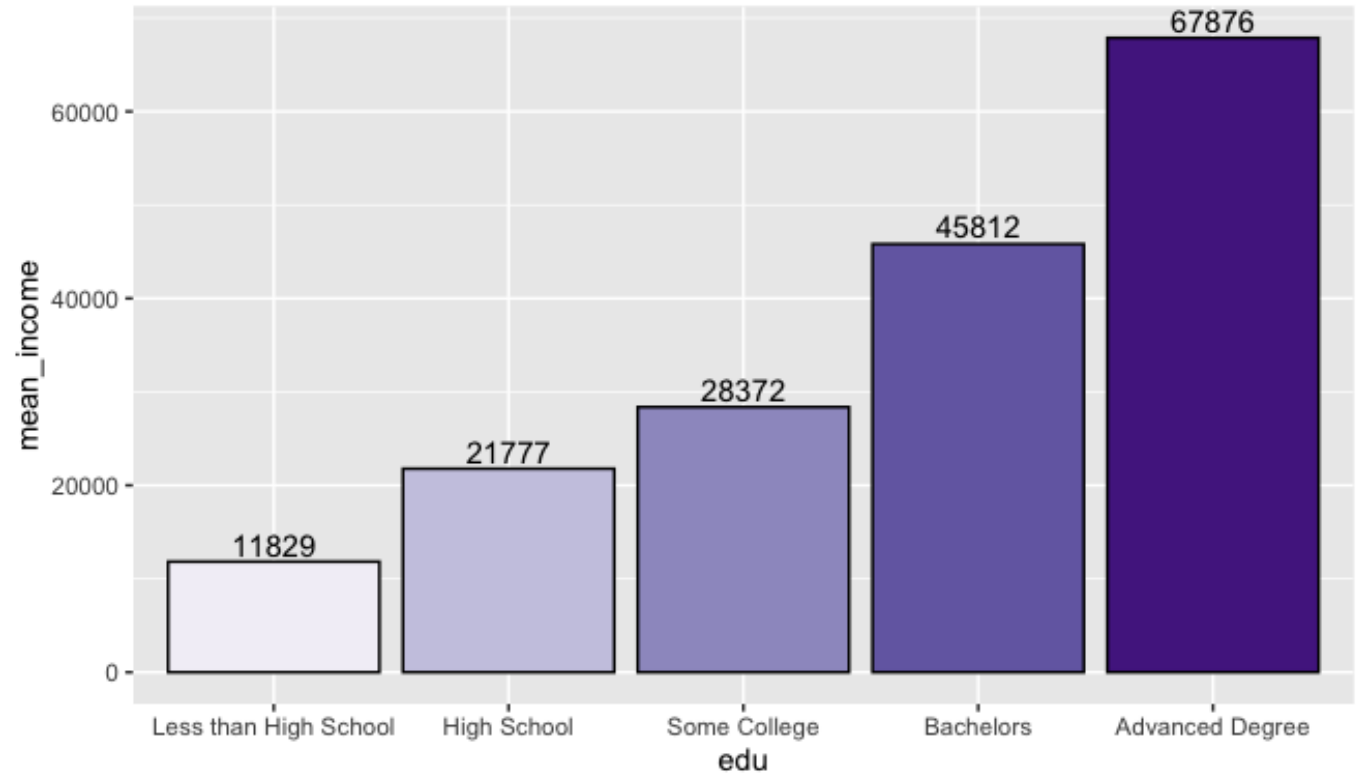
<https://pollev.com/vsovero>



# Means by group

We can take a similar approach to plotting mean income by education:

- First, create a table of mean income by education
- Next, use `geom_col()`
- Add labels using `geom_text()`



# Mean Income by Education

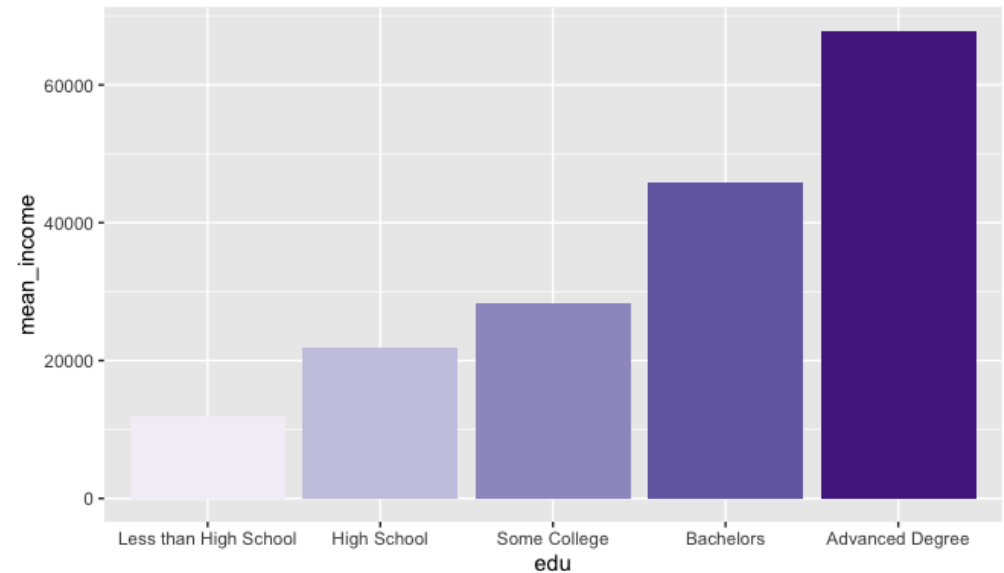
```
income_by_edu<-acs %>%  
  filter(!is.na(edu)) %>%  
  group_by(edu) %>%  
  summarize(mean_income = mean(income, na.rm = TRUE))
```

	edu	mean_income
1	Less than High School	11829.41
2	High School	21777.43
3	Some College	28372.29
4	Bachelors	45811.77
5	Advanced Degree	67875.52

# Bar Plot of Mean Income by Education

```
ggplot(data=income_by_edu,  
mapping=aes(x=edu, y=mean_income))+  
geom_col(aes(fill=edu))+  
scale_color_brewer(palette="Purples") +  
theme(legend.position = 'none') +
```

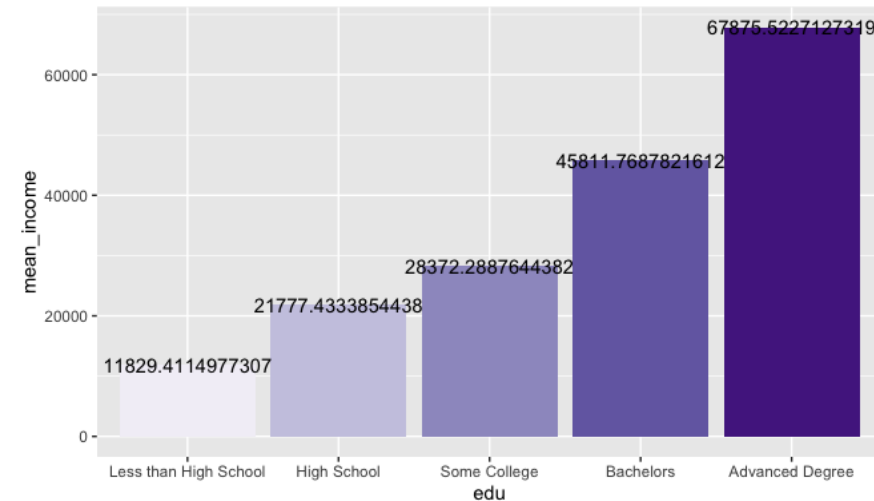
	edu	mean_income
1	Less than High School	11829.41
2	High School	21777.43
3	Some College	28372.29
4	Bachelors	45811.77
5	Advanced Degree	67875.52



# Add text (mean income values)

```
ggplot(data=income_by_edu,  
mapping= aes(x=edu, y=mean_income))+  
geom_col(aes(fill=edu))+  
geom_text(aes(label=mean_income ))+  
scale_color_brewer(palette= "Purples" )+  
theme(legend.position = 'none') +
```

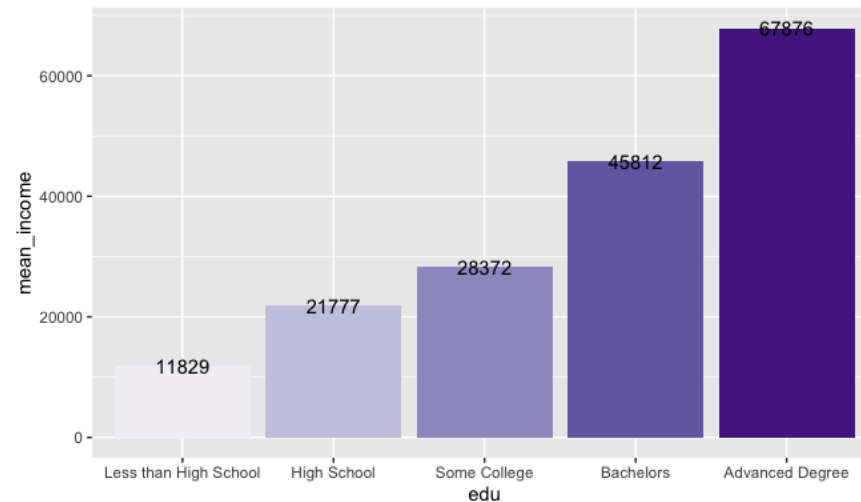
	edu	mean_income
1	Less than High School	11829.41
2	High School	21777.43
3	Some College	28372.29
4	Bachelors	45811.77
5	Advanced Degree	67875.52



# Add text (rounded mean income values)

```
ggplot(data=income_by_edu,  
mapping= aes(x=edu, y=mean_income))+  
geom_col(aes(fill=edu))+  
geom_text(aes(label=round(mean_income)))+  
scale_color_brewer(palette= "Purples" )+  
theme(legend.position = 'none') +
```

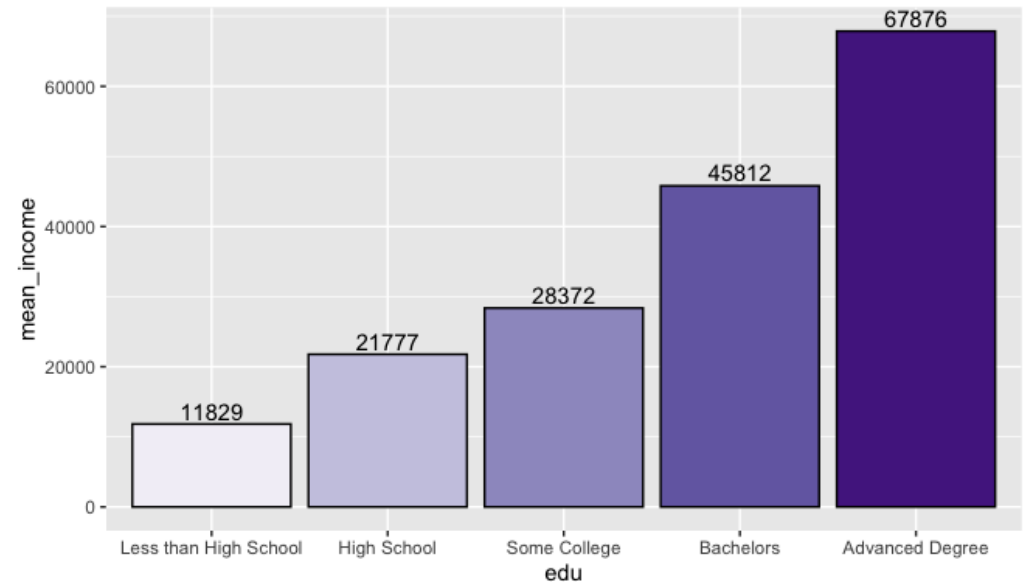
	edu	mean_income
1	Less than High School	11829.41
2	High School	21777.43
3	Some College	28372.29
4	Bachelors	45811.77
5	Advanced Degree	67875.52



# Add text (adjust position)

```
ggplot(data=income_by_edu,  
mapping= aes(x=edu, y=mean_income))+  
geom_col(aes(fill=edu))+  
geom_text(aes(label=round(mean_income), vjust=-.3))+  
scale_color_brewer(palette= "Purples" )+  
theme(legend.position = 'none') +
```

	edu	mean_income
1	Less than High School	11829.41
2	High School	21777.43
3	Some College	28372.29
4	Bachelors	45811.77
5	Advanced Degree	67875.52



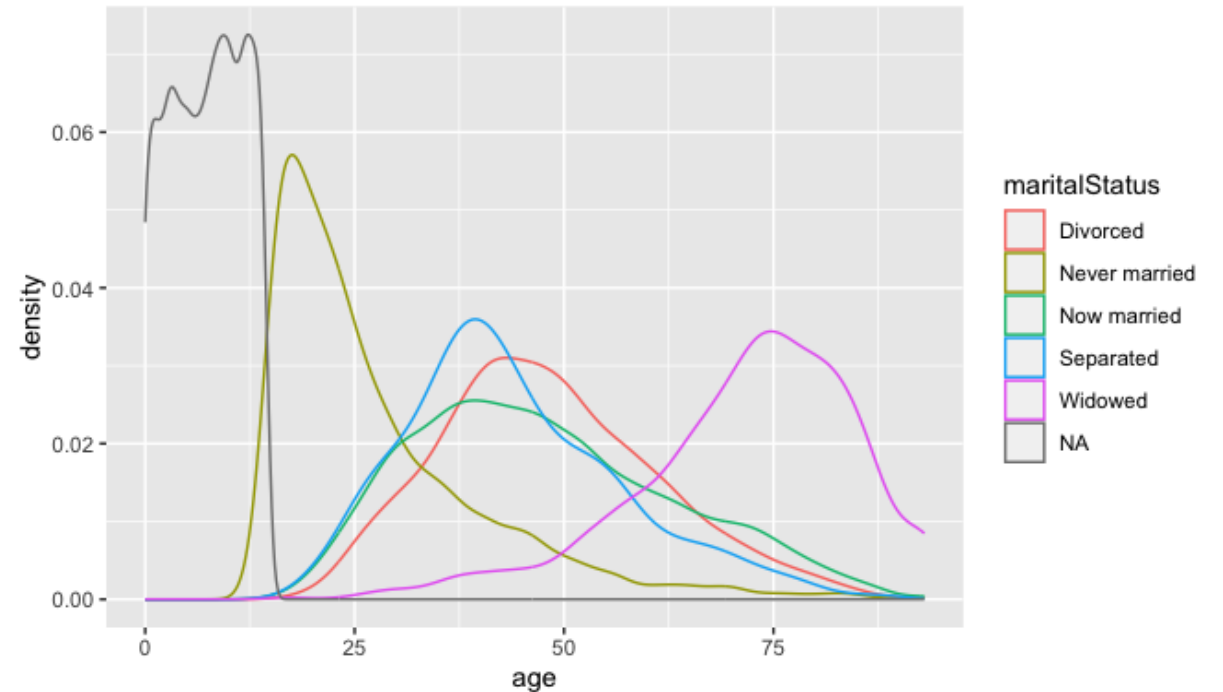
# Visual Decluttering

- We reviewed this earlier, but I'll repeat it again
- Less is more, don't over clutter your graphs

# Visual Decluttering

- Things I don't like:
- NA needs to be removed
- too many lines on a graph

```
ggplot(data=acs,  
mapping= aes(x=age))+  
geom_density(aes(color=maritalStatus))
```

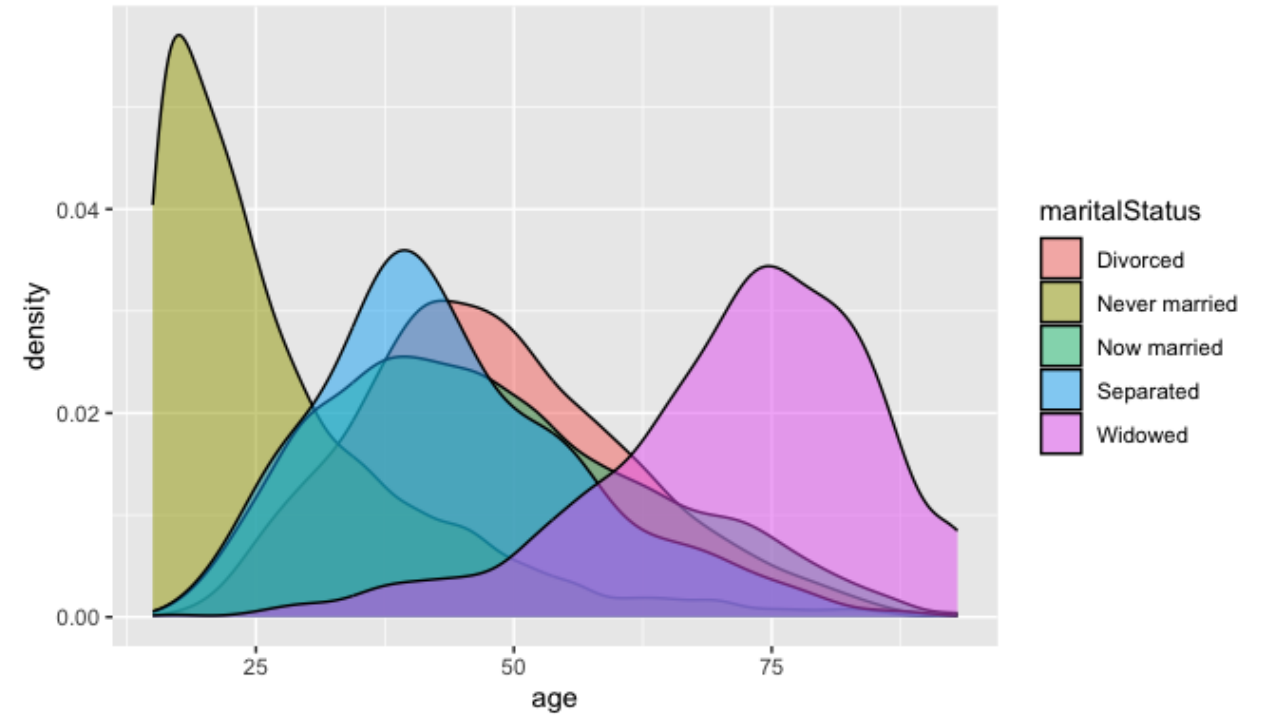




# Try fill instead of color

- We can adjust the transparency as well (alpha)

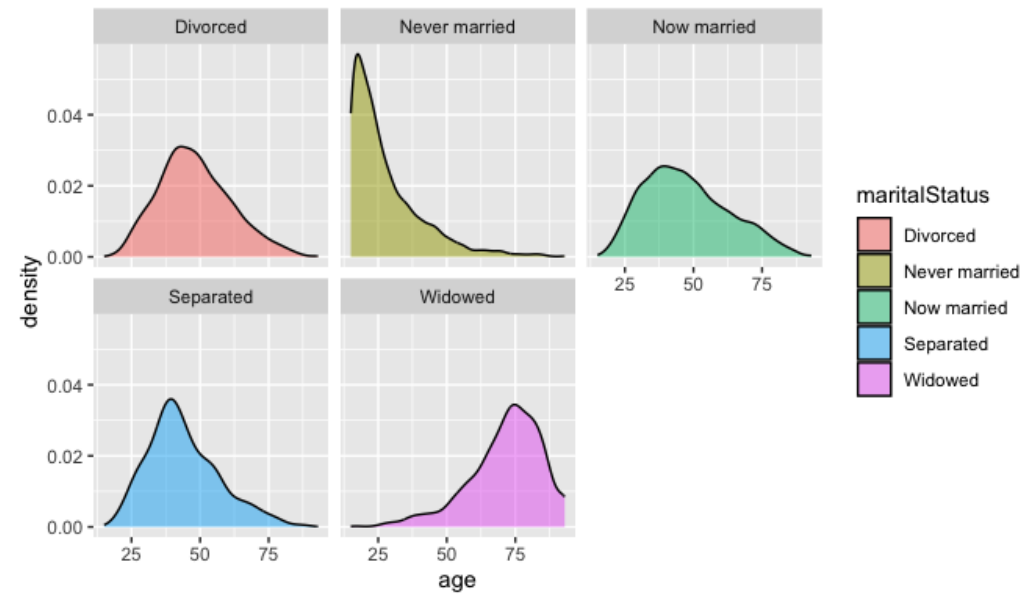
```
ggplot(data=acs_marital,  
mapping= aes(x=age))+  
geom_density(aes(fill=maritalStatus), alpha=.5)
```



# Try Faceting

- Breaks up the graph into separate parts

```
ggplot(data=acs_marital,  
mapping= aes(x=age)) +  
geom_density(aes(fill=maritalStatus), alpha=.5) +  
facet_wrap(. ~ maritalStatus)
```



# Class Exercise

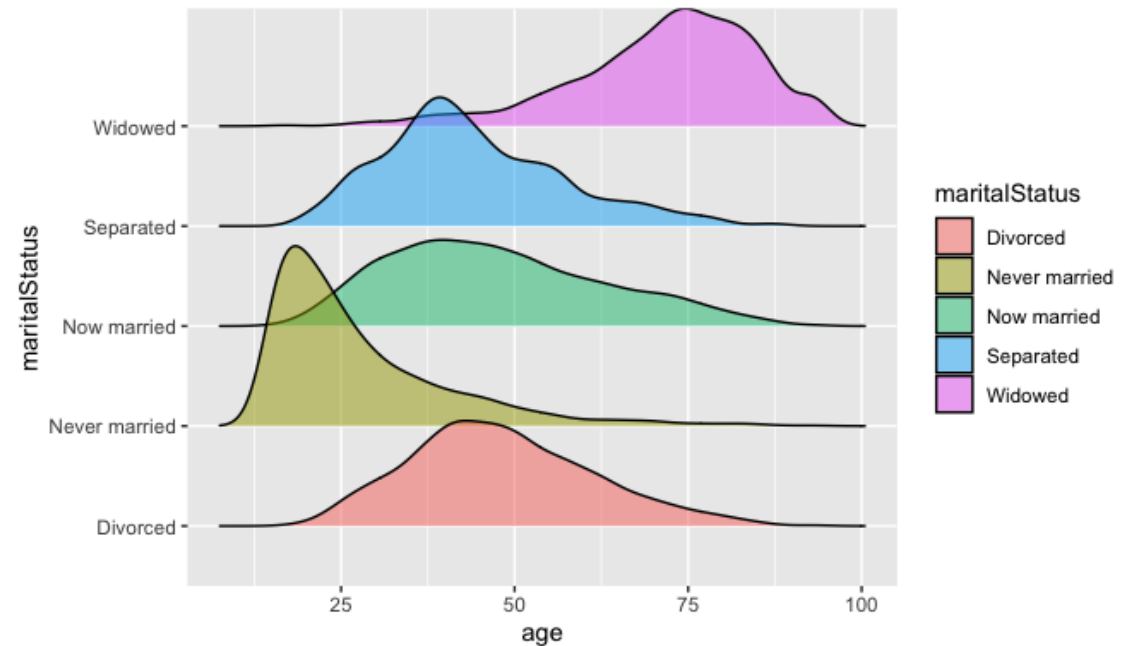
- Select individuals with an advanced degree earning 200,000 or less
- create a density plot of income faceted by race

<https://pollev.com/vsovero>

# Try ggridges

- Breaks up the graph into separate parts, but with a slight overlap

```
ggplot(data=acs_marital,  
mapping= aes(x=age))+  
geom_density_ridges(aes(fill=maritalStatus), alpha=.5)
```

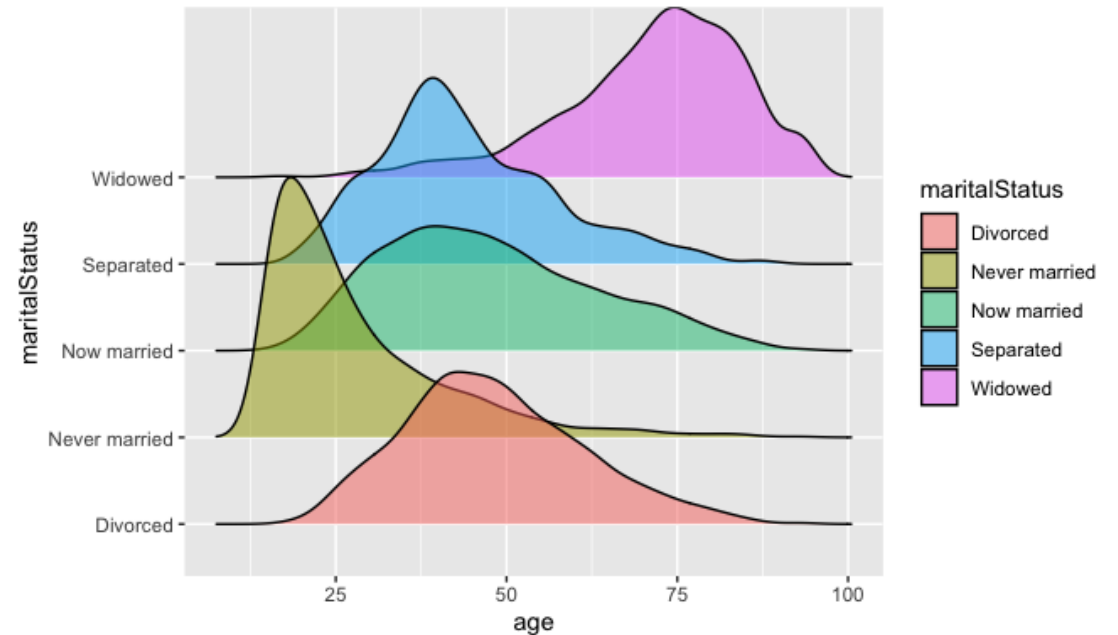


<https://cran.r-project.org/web/packages/ggribbles/vignettes/introduction.html>

# Try ggbridges

- You can make the overlap larger with the scale argument (bigger number gives a greater overlap)

```
ggplot(data=acs_marital,  
mapping=aes(x=age))+  
geom_density_ridges(aes(fill=maritalStatus), alpha=.5, scale=3)
```

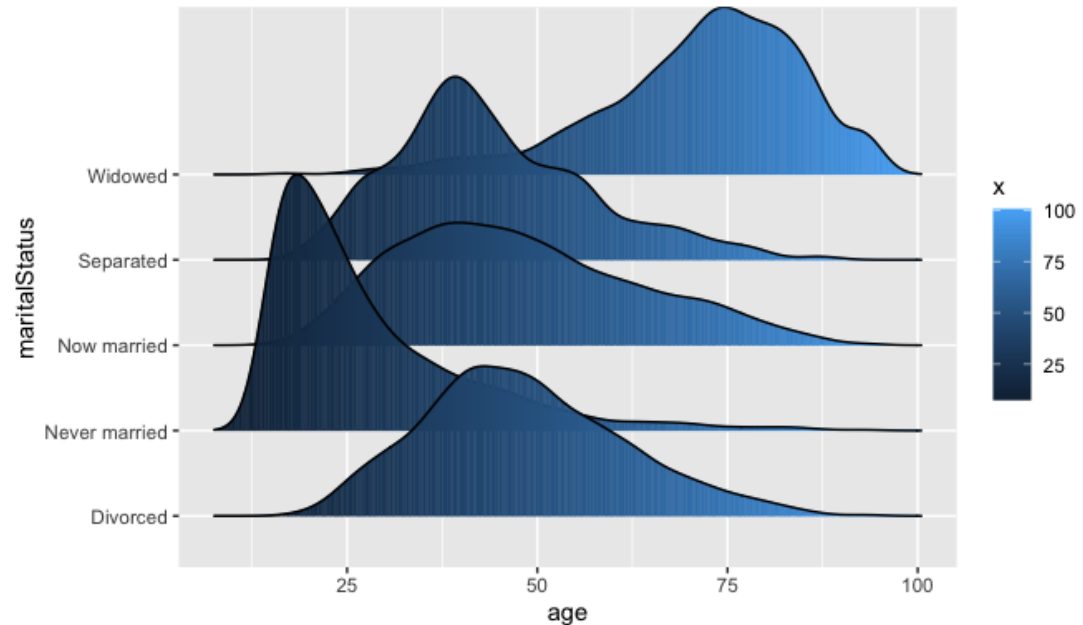


<https://cran.r-project.org/web/packages/ggbridges/vignettes/introduction.html>

# Try ggridges

- Use color to represent values on the x-axis (age)
- Very fancy.

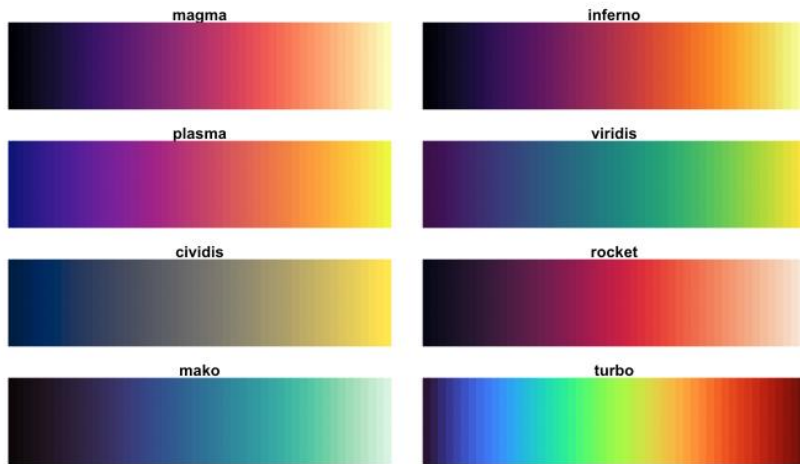
```
ggplot(data=acs_marital,  
mapping= aes(x=age))+  
geom_density_ridges(aes(fill=stat(x)), scale=3)
```



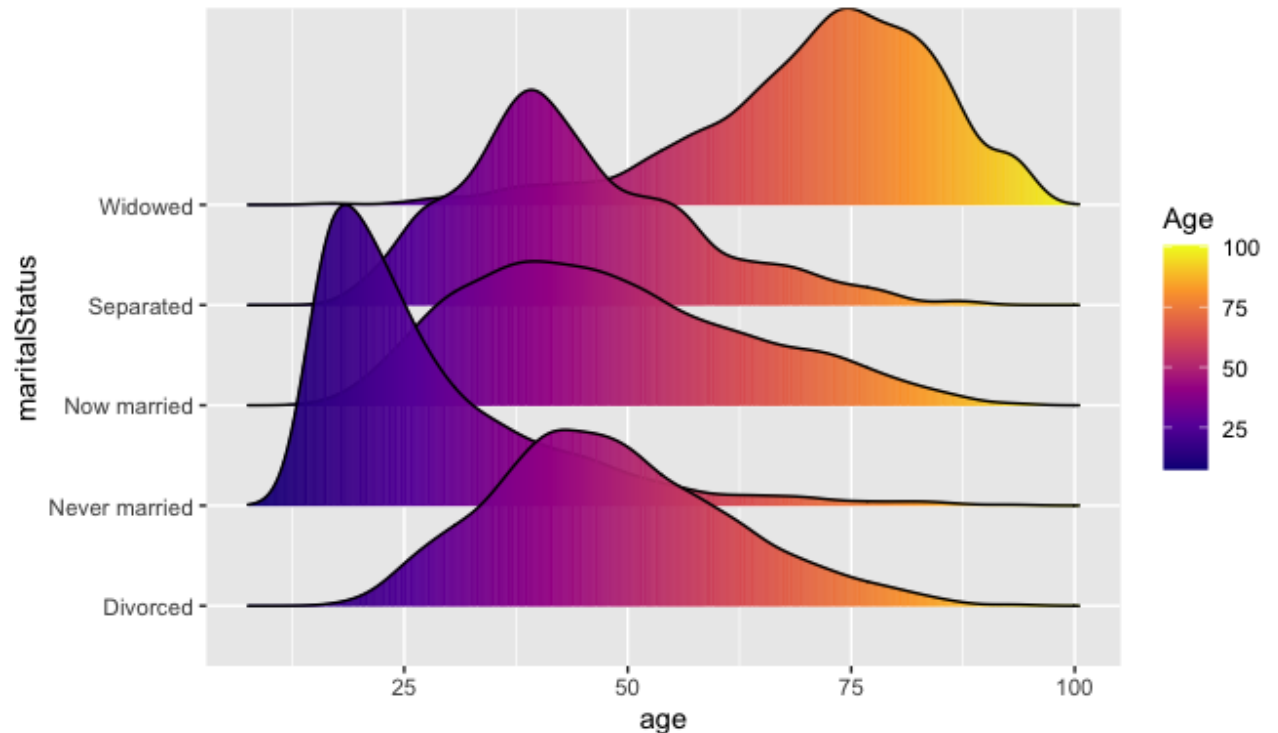
<https://cran.r-project.org/web/packages/ggribes/vignettes/introduction.html>

# Try ggridges

- Use a viridis color palette (plasma)
- Extra fancy.



```
ggplot(data=acs_marital,  
mapping= aes(x=age))+  
geom_density_ridges(aes(fill=stat(x)), scale=3)  
scale_fill_viridis_c(name= "Age", fill= "plasma" )+
```



<https://cran.r-project.org/web/packages/ggribes/vignettes/introduction.html>