Introduction to Data Visualization

How to refine our plots?
"colors & text"

Halil Bisgin, Ph.D.

The default look with ggplot

- The default settings in ggplot should be pretty good to work with.
- It's only when we have some specific plot in mind that the question of polishing the results comes up.

Refining a plot can mean several things

- We might want to get the look of it just right, based on our own tastes and our sense of what needs to be highlighted.
- We might want to format it in a way that will meet the expectations of a journal, or of a conference audience, or the general public.
- We might want to tweak this or that feature of the plot or add an annotation or additional detail not covered by the default output.
- We might want to completely change the look of the entire thing, without distorting the original plot.

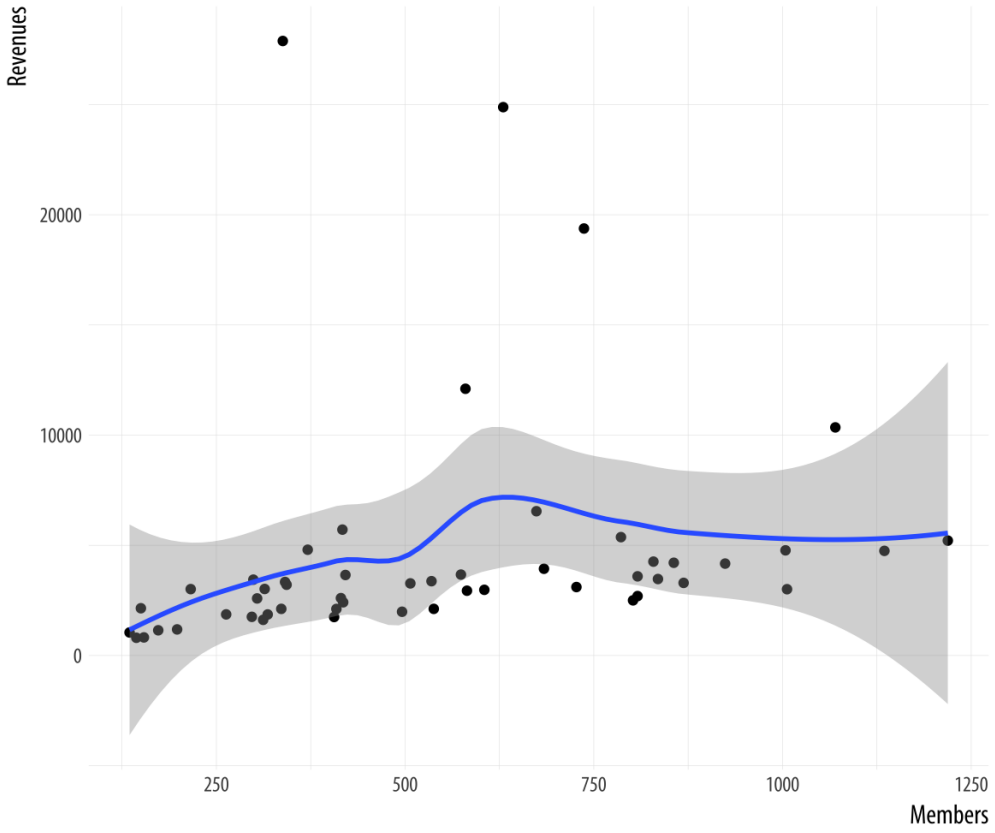
Use case- American Sociological Association

- Data becomes available when you install socviz in case you don't have it on your system.
- Membership data for each section over a ten year period, but the data on section reserves and income (the Beginning and Revenues variables) is for the 2015 year only.

```
> head(asasec)
                              Section         Sname Beginning Revenues Expenses Ending Journal Year Members
1      Aging and the Life Course (018)         Aging     12752    12104    12007  12849      No 2005     598
2     Alcohol, Drugs and Tobacco (030) Alcohol/Drugs     11933     1144      400  12677      No 2005     301
3 Altruism and Social Solidarity (047)      Altruism      1139     1862     1875   1126      No 2005      NA
4            Animals and Society (042)       Animals       473      820     1116    177      No 2005     209
5             Asia/Asian America (024)          Asia      9056     2116     1710   9462      No 2005     365
6            Body and Embodiment (048)          Body      3408     1618     1920   3106      No 2005      NA
>
```

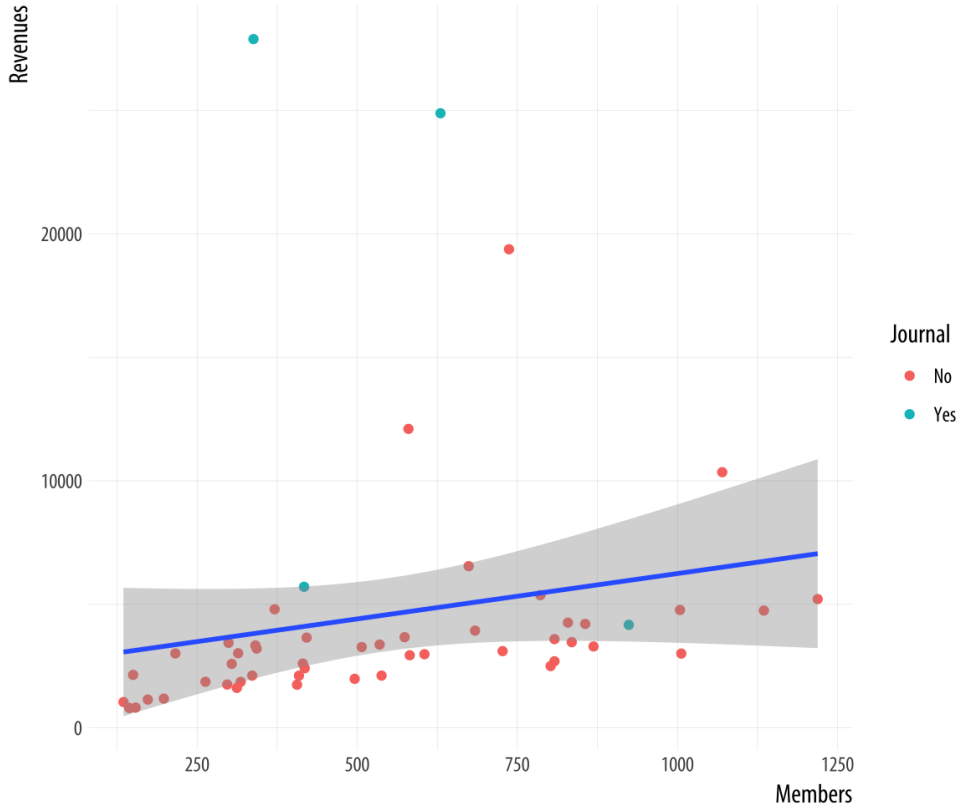
Basic plot w/ ggplot

- Let's look at the relationship between section membership and section revenues for a single year, 2014.
  - *geom_point()*
  - *smooth with loess*



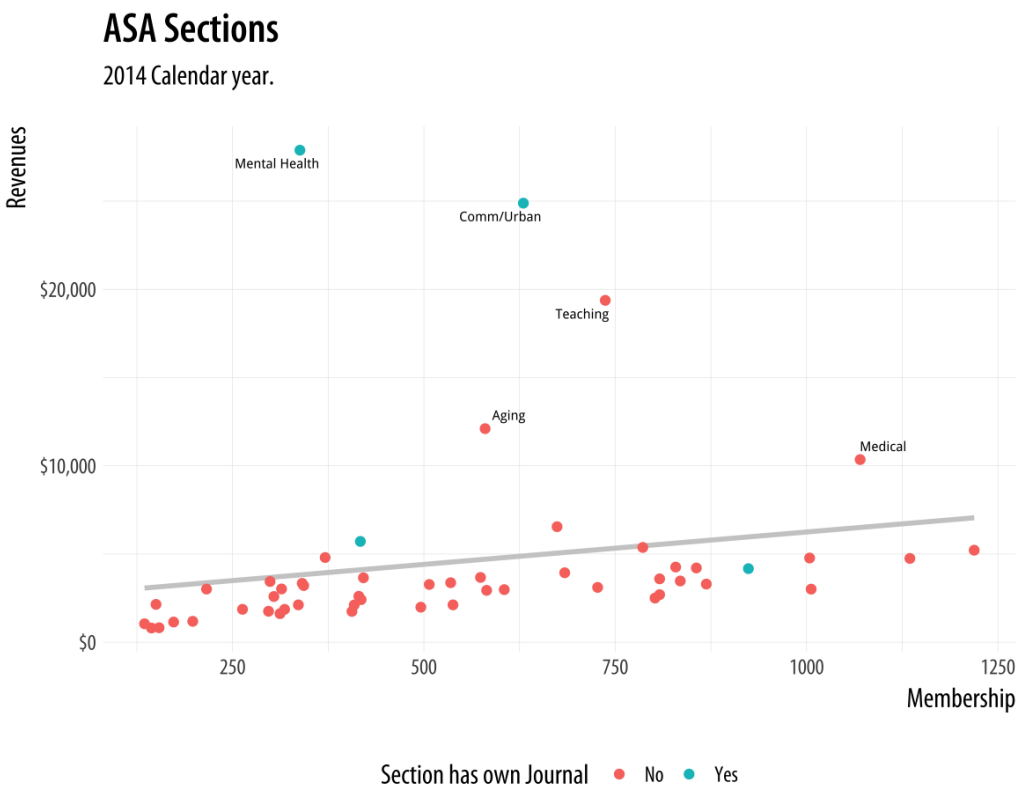
Refinement 1-*color*

- Change the smooth parameter, and add a third variable journal by using color.



Refinement II-adding annotations

- *geom_text_repel ()* **for point annotations.**
- *labs* **and** *scale_y_continuous()* **to add labels.**
- *theme()* **for legend**



Use color to your advantage

- You should choose a color palette in the first place based on its ability to express the data you are plotting.
- An unordered categorical variable like “Country” or “Sex”, for example, requires distinct colors that won’t be easily confused with one another.
- An ordered categorical variable like “Level of Education”, on the other hand, requires a graded color scheme of some kind running from less to more or earlier to later.

There are other considerations

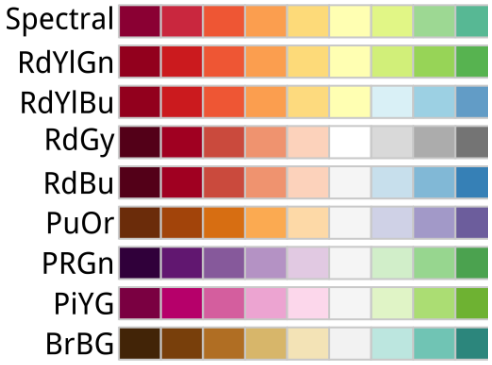
- If your variable is ordered, is your scale centered on a neutral midpoint with departures to extremes in each direction, as in a Likert scale?
- These questions are about ensuring accuracy and fidelity when mapping a variable to a color scale.
- Take care to choose a palette that reflects the structure of your data.
  - *For example, do not map sequential scales to categorical palettes, or use a diverging palette for a variable with no well-defined midpoint.*

Palettes

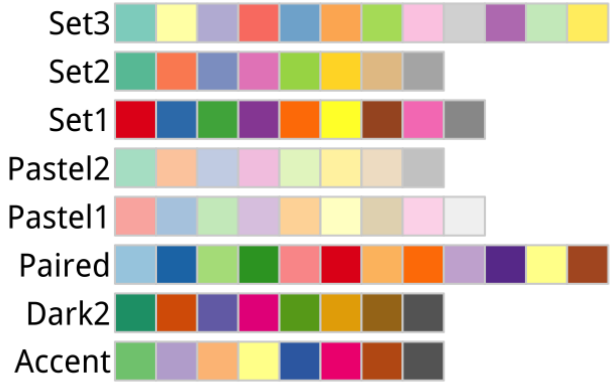
- We choose color palettes for mappings through one of the scale_ functions for color or fill.
- While it is possible to use scale_color_hue(), or scale_fill_hue(), in general this is not recommended.
- Instead we should use the RColorBrewer package to pick a palette.
- When used in conjunction with ggplot, you access these colors by specifying the scale_color_brewer() or scale_fill_brewer() functions, depending on the aesthetic you are mapping

RColorBrewer's palettes

sequential



diverging

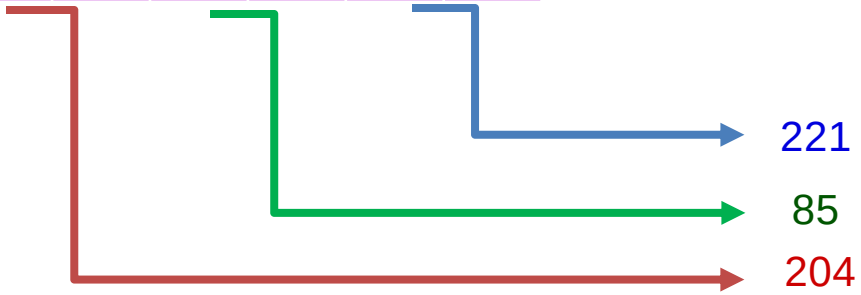


qualitative

You can always color manually

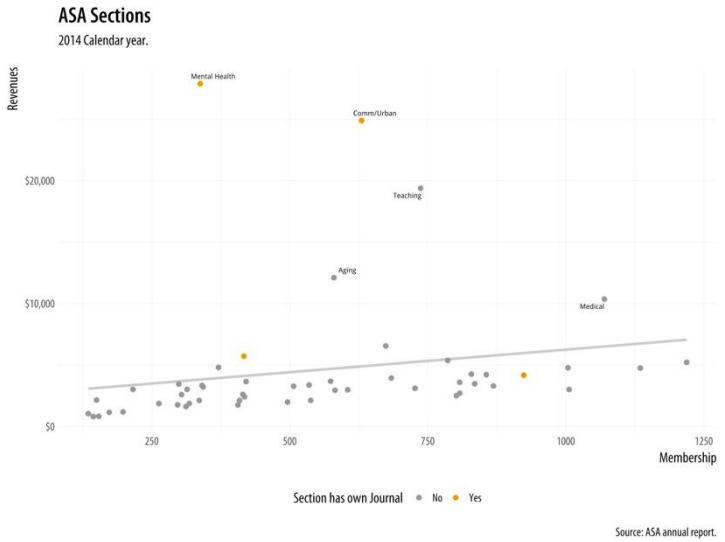
- scale_color_manual() or scale_fill_manual()
- Take a value argument that can be specified as vector of color names or color values that R knows about
  - *demo('colors')*
- Colors can be s… l RGB value.

- Letters represe… base 16



Manually setting colors-example

- Useful when the meaning of a category itself has a strong color association (e.g., Political parties)
- A color-blind-friendly palette (Chang, 2013)
  - *cb_palette <- c("#999999", "#E69F00", "#56B4E9", "#009E73", "#F0E442", "#0072B2", "#D55E00", "#CC79A7")*
  - *… + scale_color_manual(values = cb_palette)*



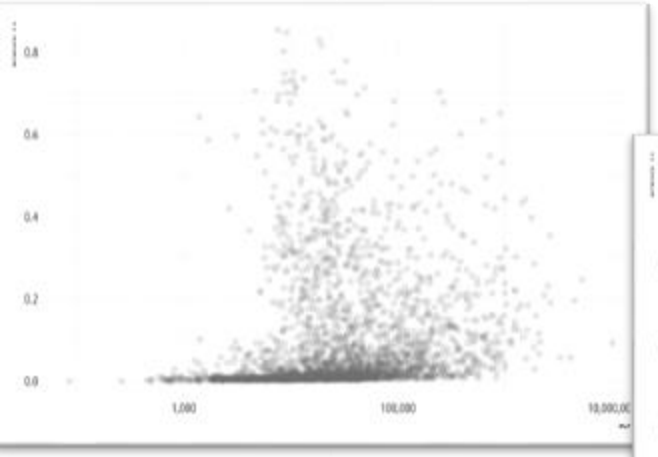
More we can do for color-blind viewers

- dichromat package
- Let's choose 5 colors from sequential palette
  - Default <- **brewer.pal**(5, "Set2")

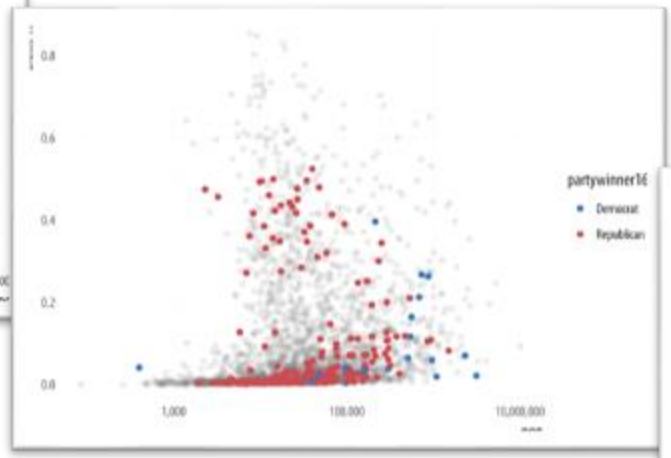

```
Default <- brewer.pal(5, "Set2")

library(dichromat)types <- c("deutan", "protan", "tritan")
names(types) <- c("Deuteronopia", "Protanopia", "Tritanopia")

color_table <- types %>%purrr::map(~ dichromat(Default, .x)) %>%    as_tibble()
%>%add_column(Default, .before = TRUE)

color_comp(color_table)
```

Layer color and text together - 1

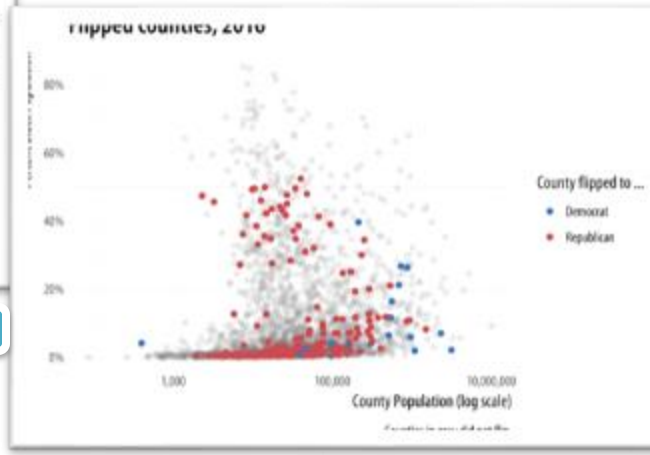
- Color is also very useful when we want to pick out or highlight some aspect of our data.
  - *The layered approach of ggplot can really work to our*



Counties not flipped



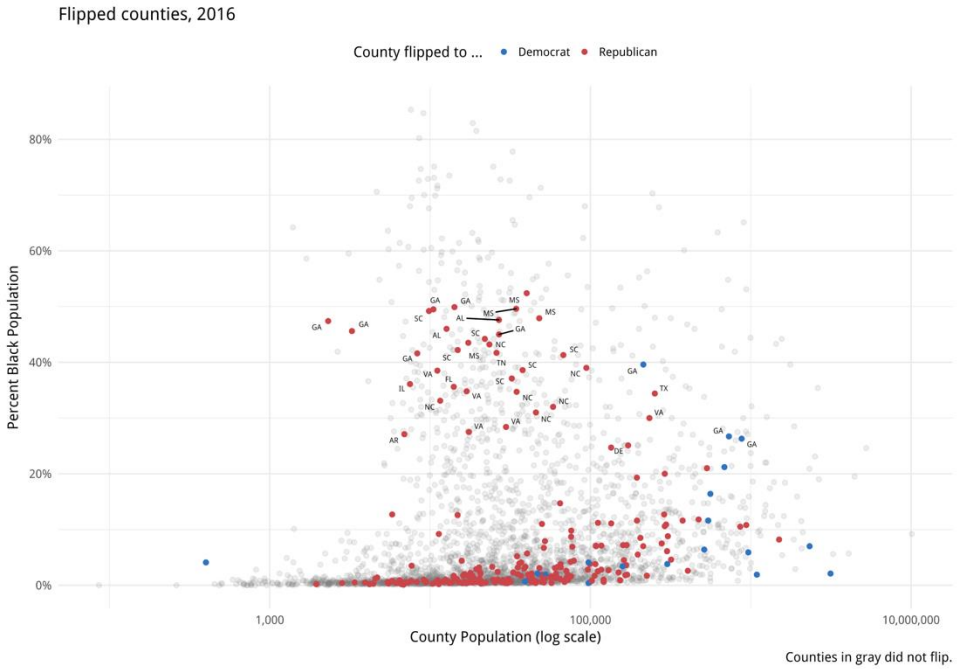
Counties flipped w/ color



Titles and labels added

Layer color and text together - 2

```
p4 <- p3 + geom_text_repel(data = subset(county_data,                    flipped == "Yes" & black > 25),mapping = aes(x=pop,y = black/100,label = state), size = 2)

p4 + theme_minimal() +    theme(legend.position="top")
```



What to pay attention?

- When looking at good plots, we should be able to see the implicit or explicit structure.
  - *The mappings that form the basis of the plot, picking out which variables are mapped to x and y, and which to to color, fill, shape, label, and so on. What geoms were used to produce them?*
  - *How have the scales been adjusted? Are the axes transformed? Are the fill and color legends combined?*
  - *What is the base layer? What has been drawn on top of it, and in what order? Which upper layers are formed from subsets of the data? Which are new datasets? Are there annotations? The ability to evaluate plots in this way, to apply the grammar of graphics in practice, is useful both for looking at plots and for thinking about how to make them.*

Demonstration

- RefineColorTextThemeIntro.R