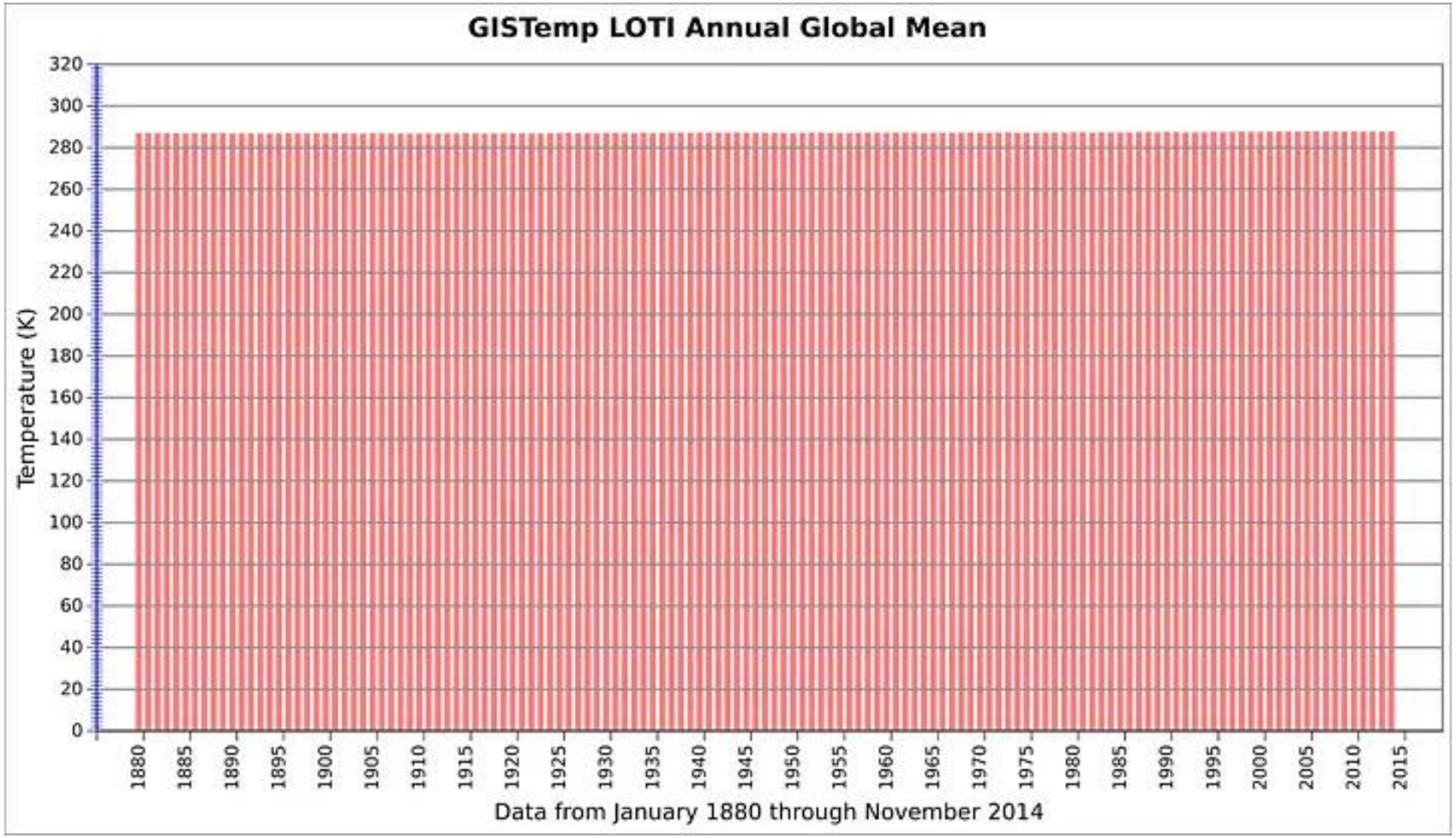
Friends Don’t Let Friends Make Bad Graphs

Slides inspired from Chenxin Li’s post,
https://github.com/cxli233/FriendsDontLetFriends/tree/main?tab=readme-ov-file

Part 1

Introduction

Global warming is fake news?



https://towardsdatascience.com/why-is-this-chart-bad-5f16da298afa

Part 1

Introduction

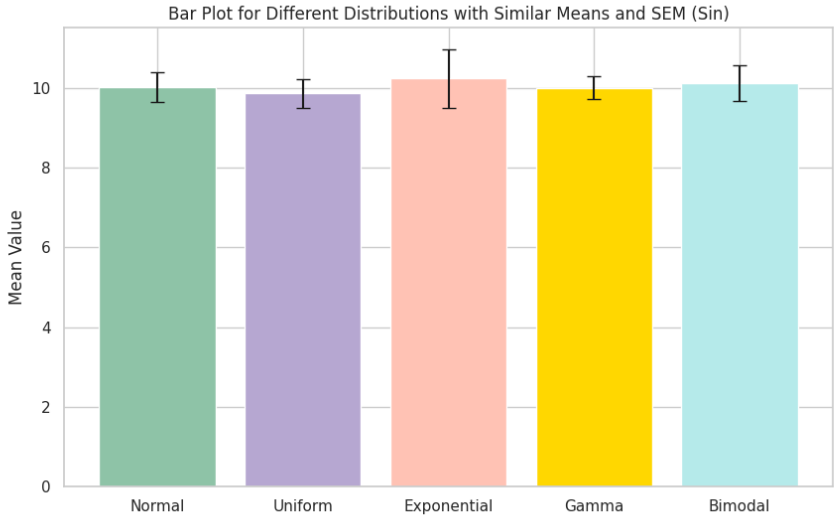
Obamacare is killing US

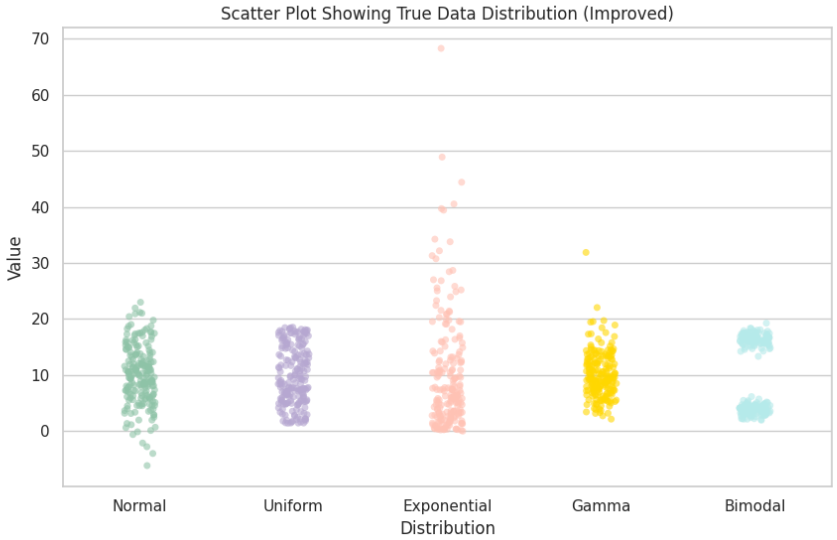
https://www.vizzu.io/blog/data-visualization-examples

Part 2

Common Pitfalls of Data Visualization

1. No Bar Plots for Mean Separation



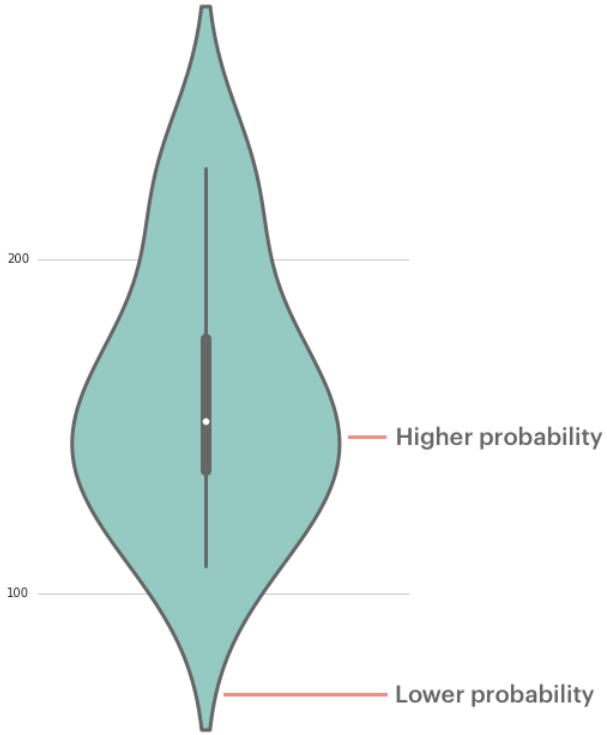


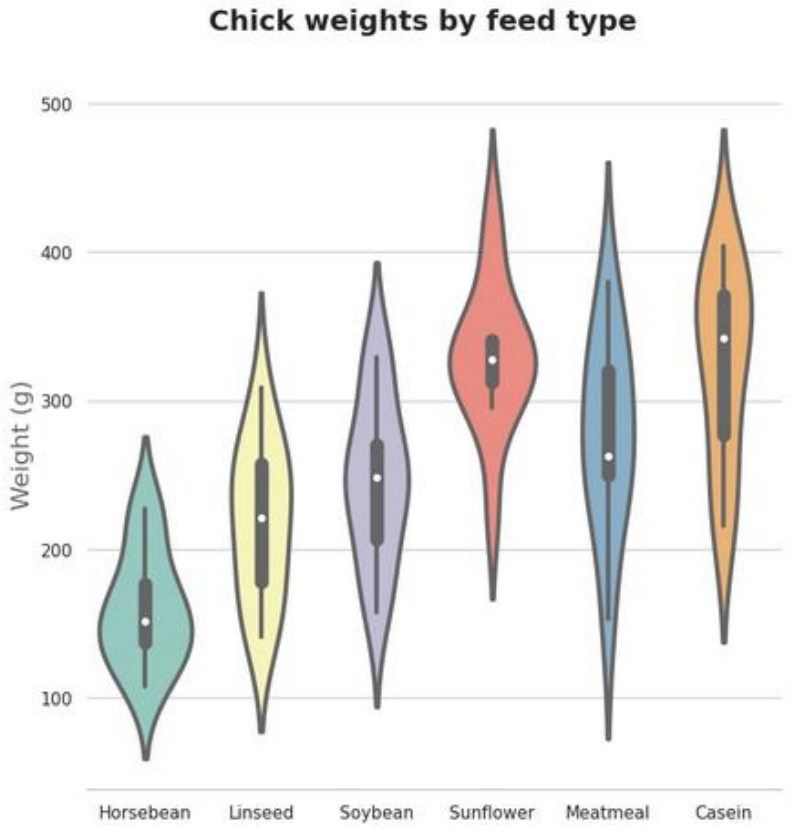
➔ Bar plots seems similar (means, standard deviations), but data shows varied distribution.

Part 2

Common Pitfalls of Data Visualization

2. Violin Plots?

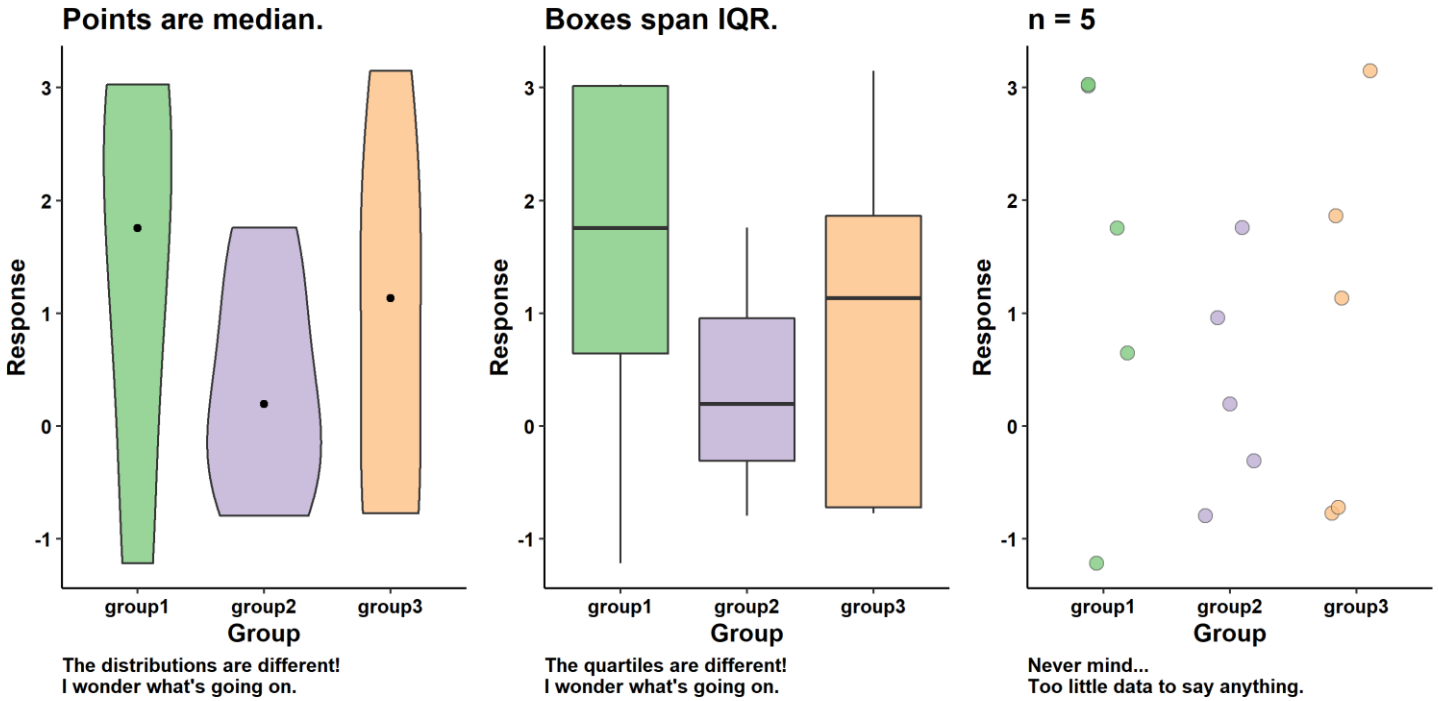




➔ Shows distribution of numerical data

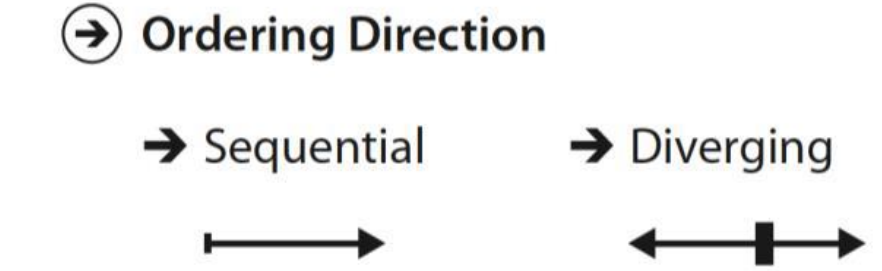
https://mode.com/blog/violin-plot-examples

2. Beware of Violin Plots for Small Sample Sizes



➔ Small N can say anything, even violin plots seems similar.
Distributions are only meaningful with large n (i.e. 〉50)

3. Sequential / Diverging Data



- Sequential: there is a homogeneous range from a minimum to a maximum value, such as height.
- Diverging: data can be deconstructed into two sequences pointing in opposite directions that meet at a common zero point, such as elevation.

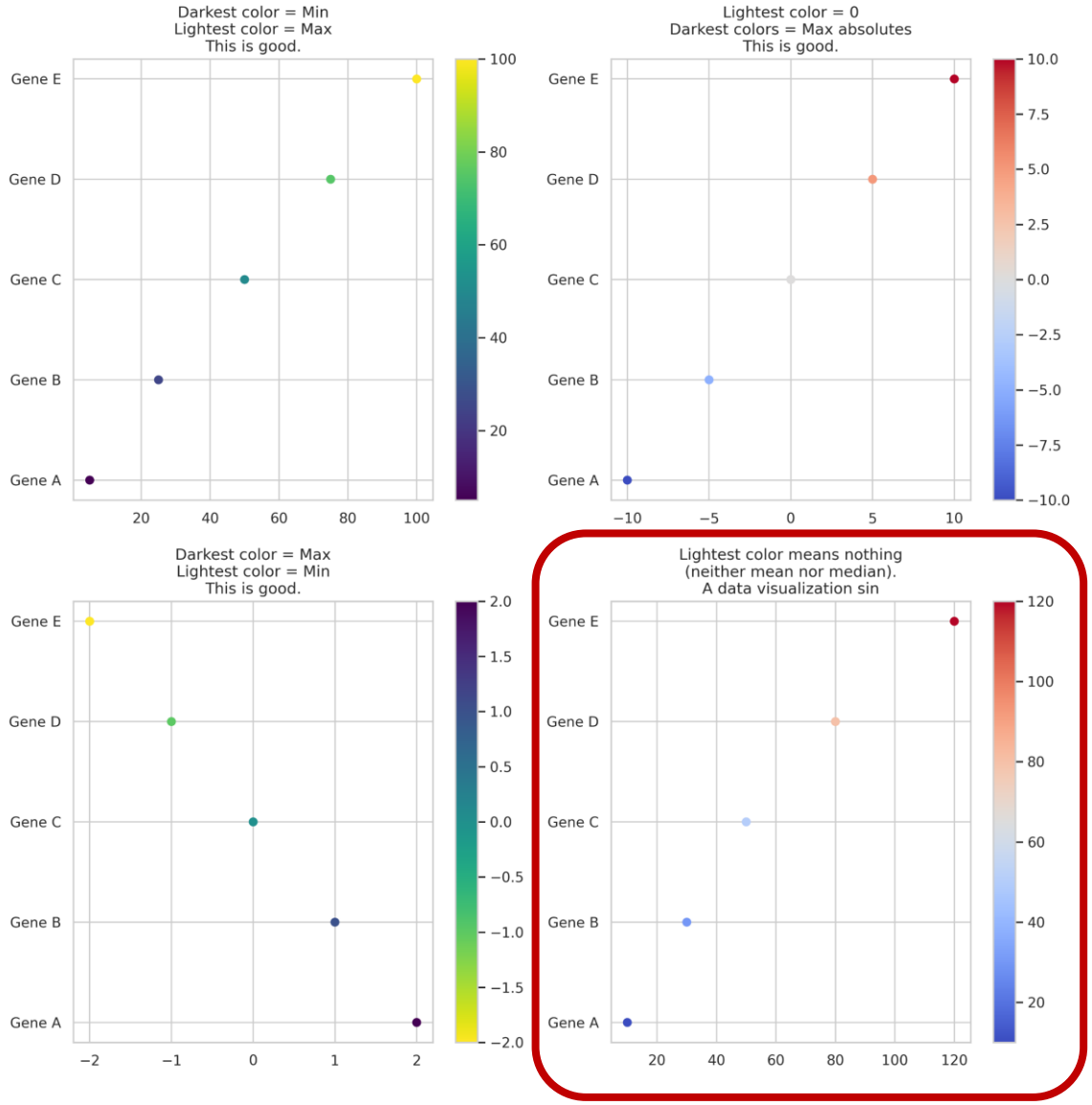
3. Sequential / Diverging Colormaps

- Sequential colormaps range from a minimum value to a maximum value.
- Diverging colormaps map a midpoint to an off-white color and two ends to two fully saturated colors with different hues.

Part 2

Common Pitfalls of Data Visualization

3. Are You Using the Right Color Scale for Sequential/Diverging Data?

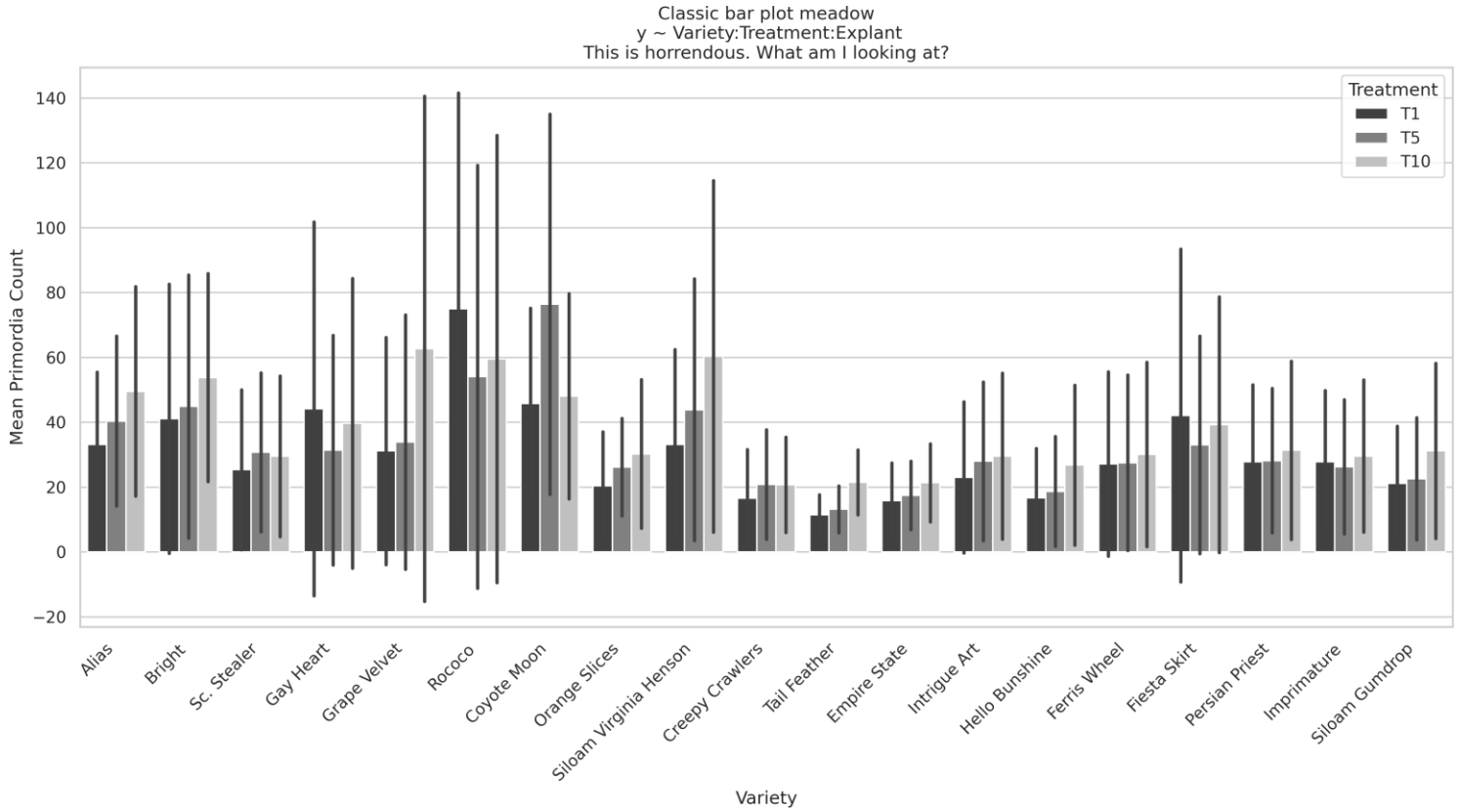


Type	Sequential Colormaps	Diverging Colormaps
Sequential Data	O	O
Diverging Data	O	X

Part 2

Common Pitfalls of Data Visualization

4. Avoid Bar Plot Meadows



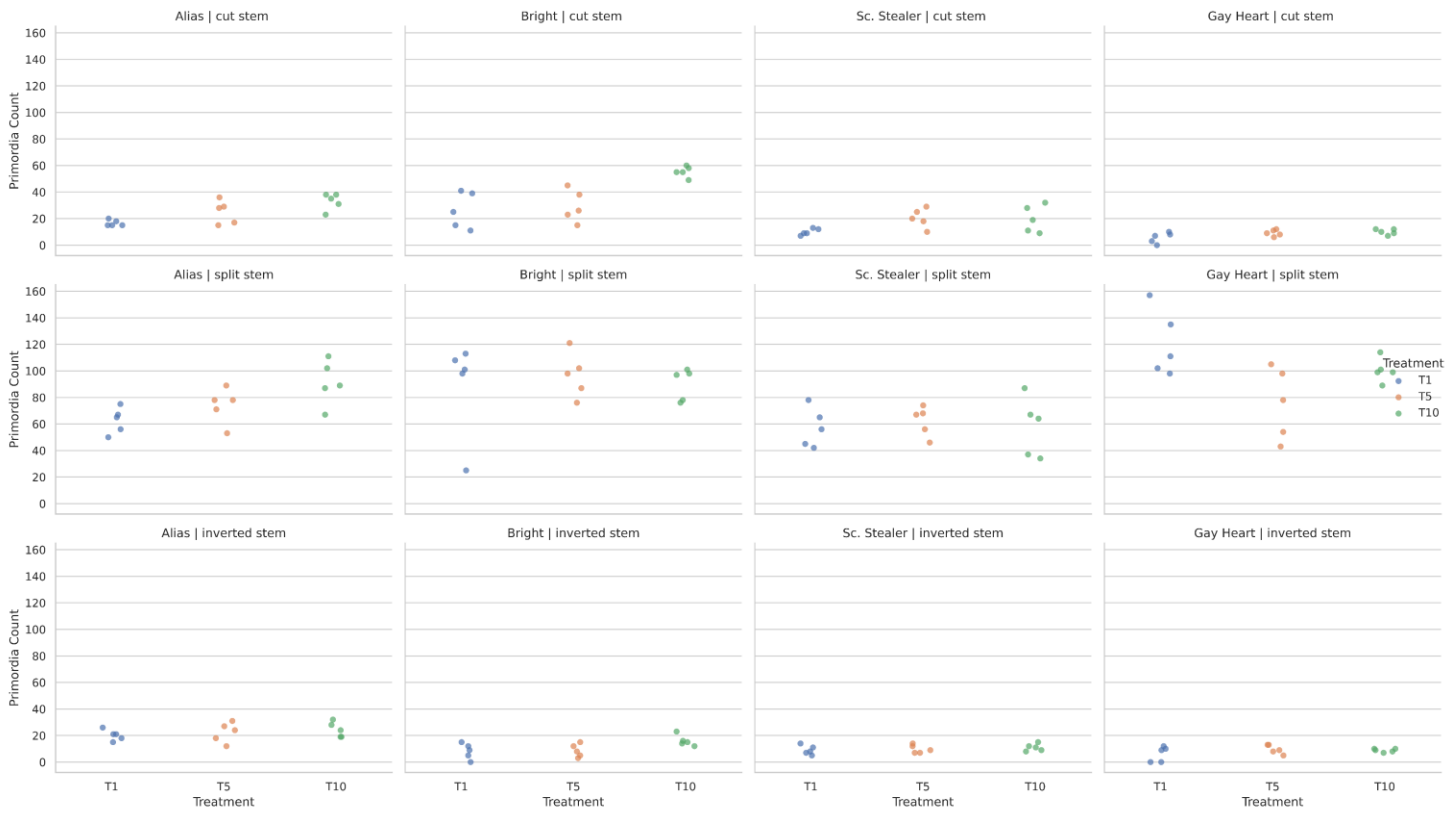
➔ For multi-faceted experiments, large(cluttered) bar plots are ineffective.

Data Source: https://link.springer.com/article/10.1186/s12870-020-2243-7

Part 2

Common Pitfalls of Data Visualization

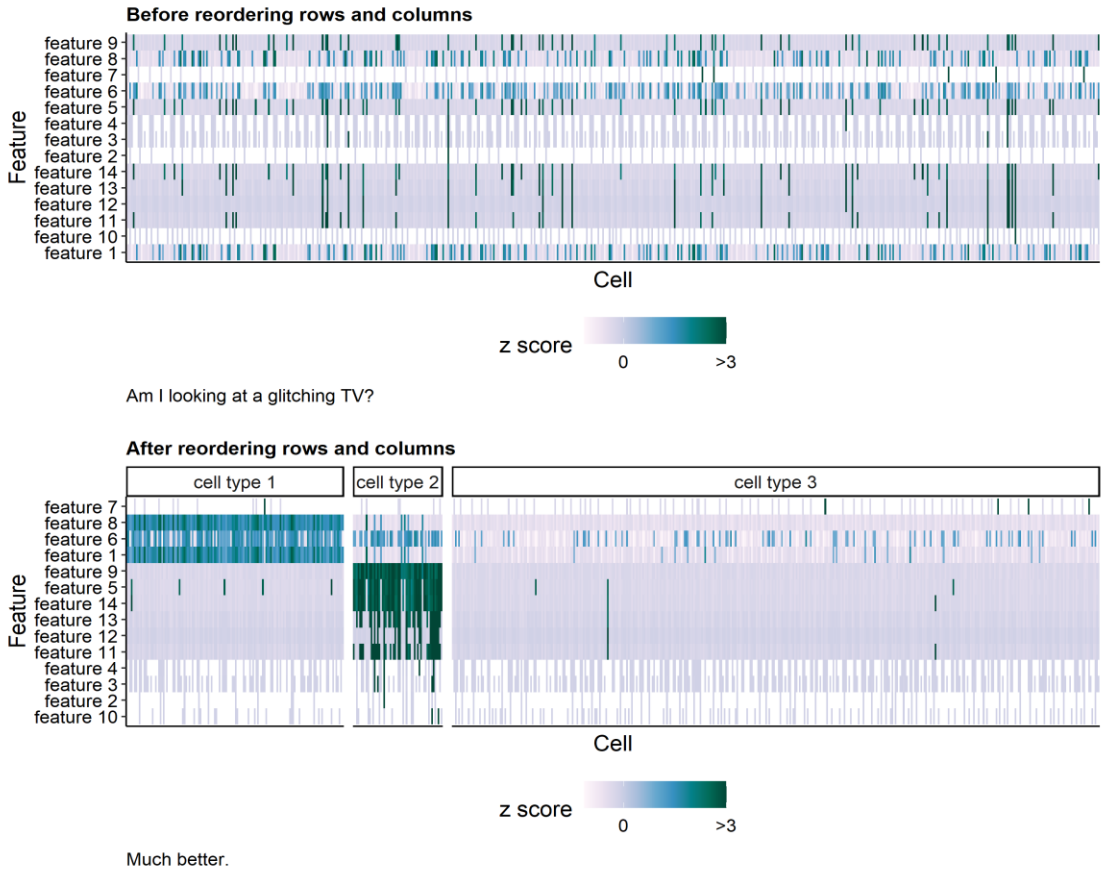
4. Avoid Bar Plot Meadows



➔ More eye-catching than large bar plot.

Data Source: https://link.springer.com/article/10.1186/s12870-020-2243-7

5. Consider Reordering Rows & Columns in Heatmaps



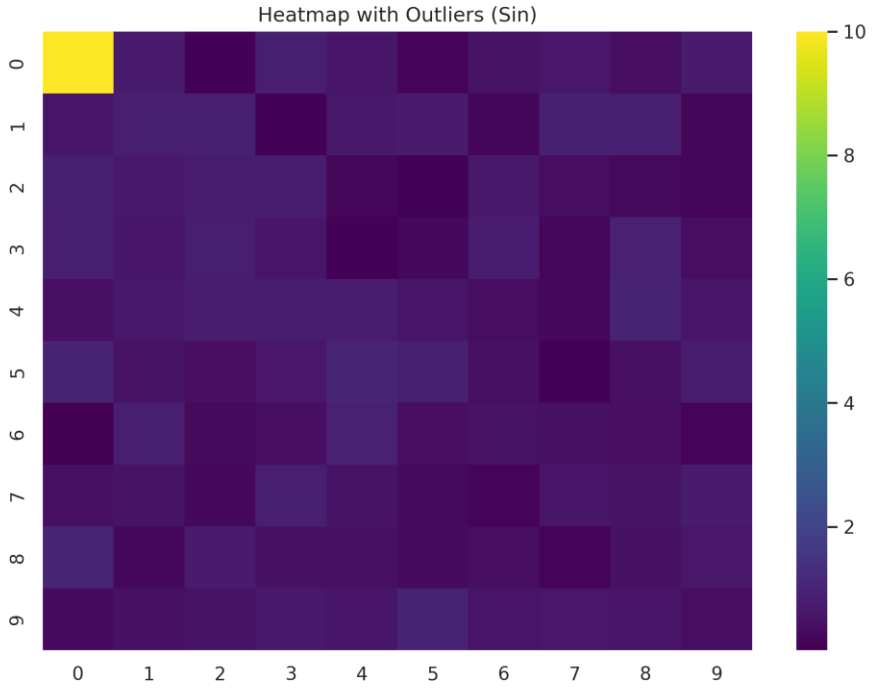
➔ Use clustering or strategic reordering for effective heatmap representation.

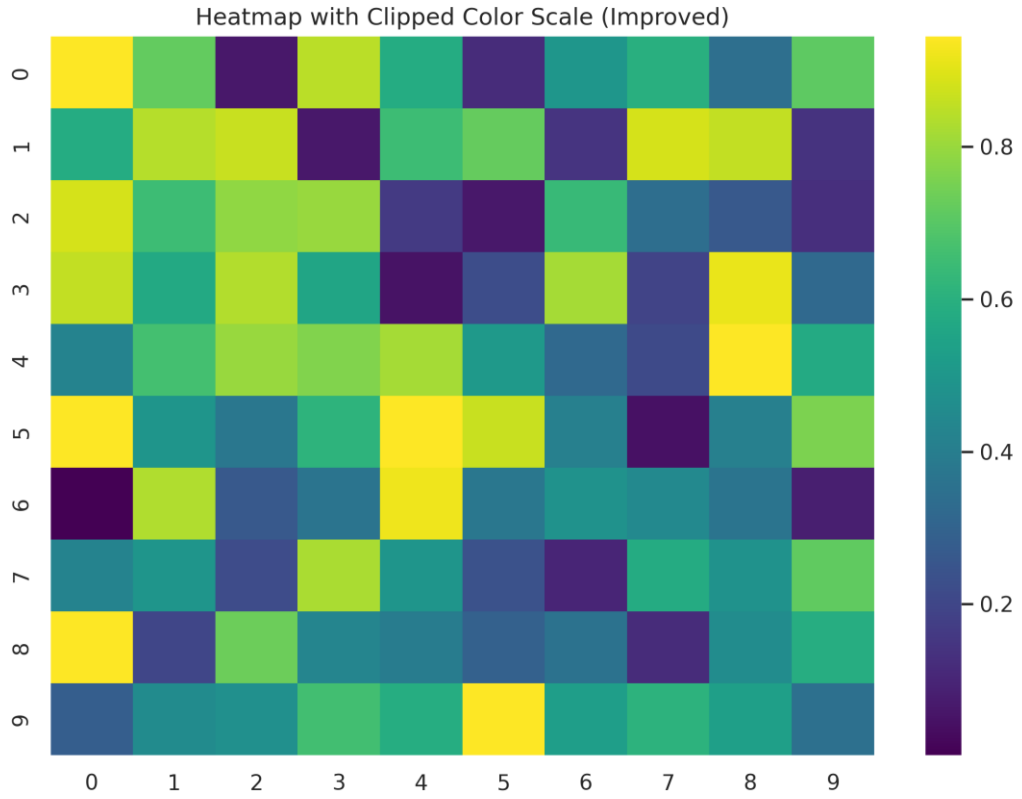
https://github.com/cxli233/FriendsDontLetFriends/blob/main/Results/Reorder_rows_and_columns_for_heatmap.png

Part 2

Common Pitfalls of Data Visualization

6. Check for Outliers in Heatmaps



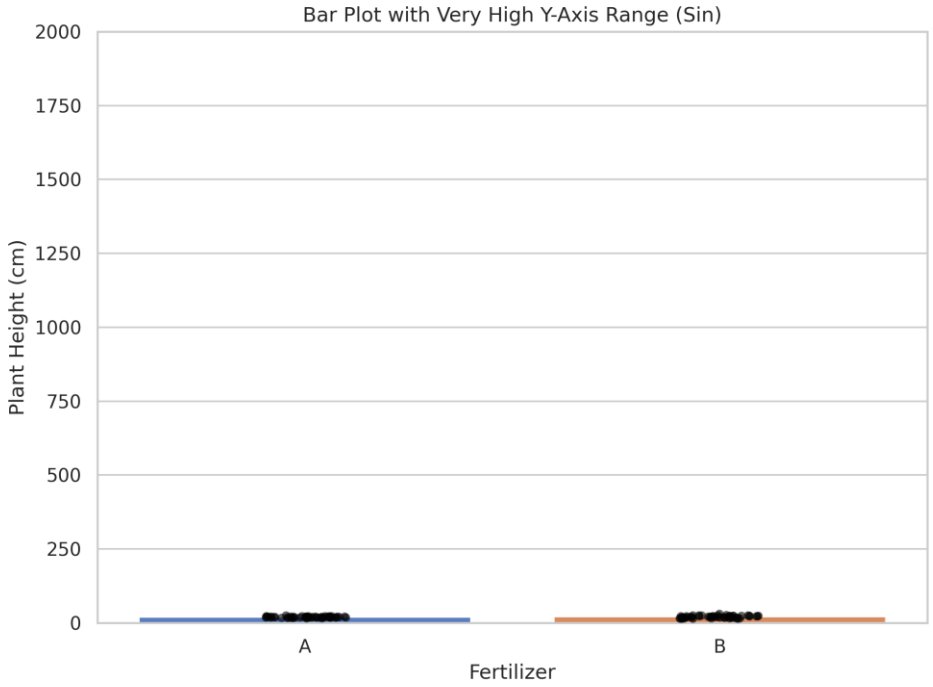


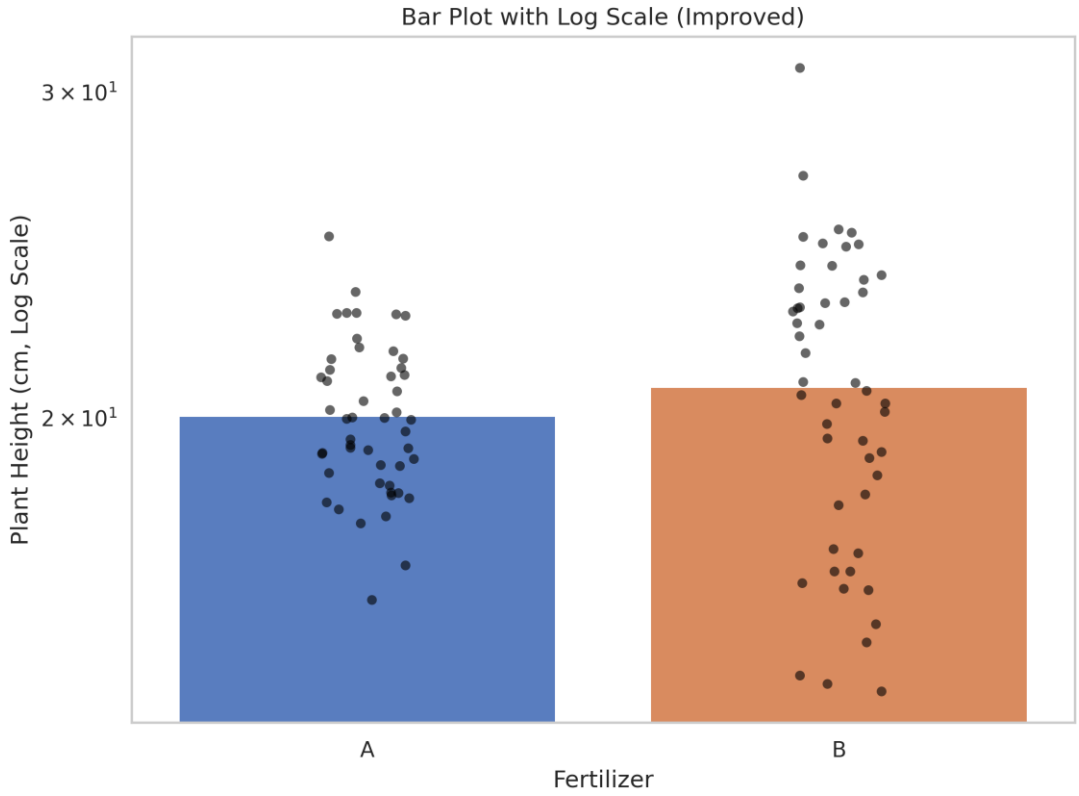
➔ Outliers can skew perceptions; capping color scales.

Part 2

Common Pitfalls of Data Visualization

7. Data Range per Factor Level



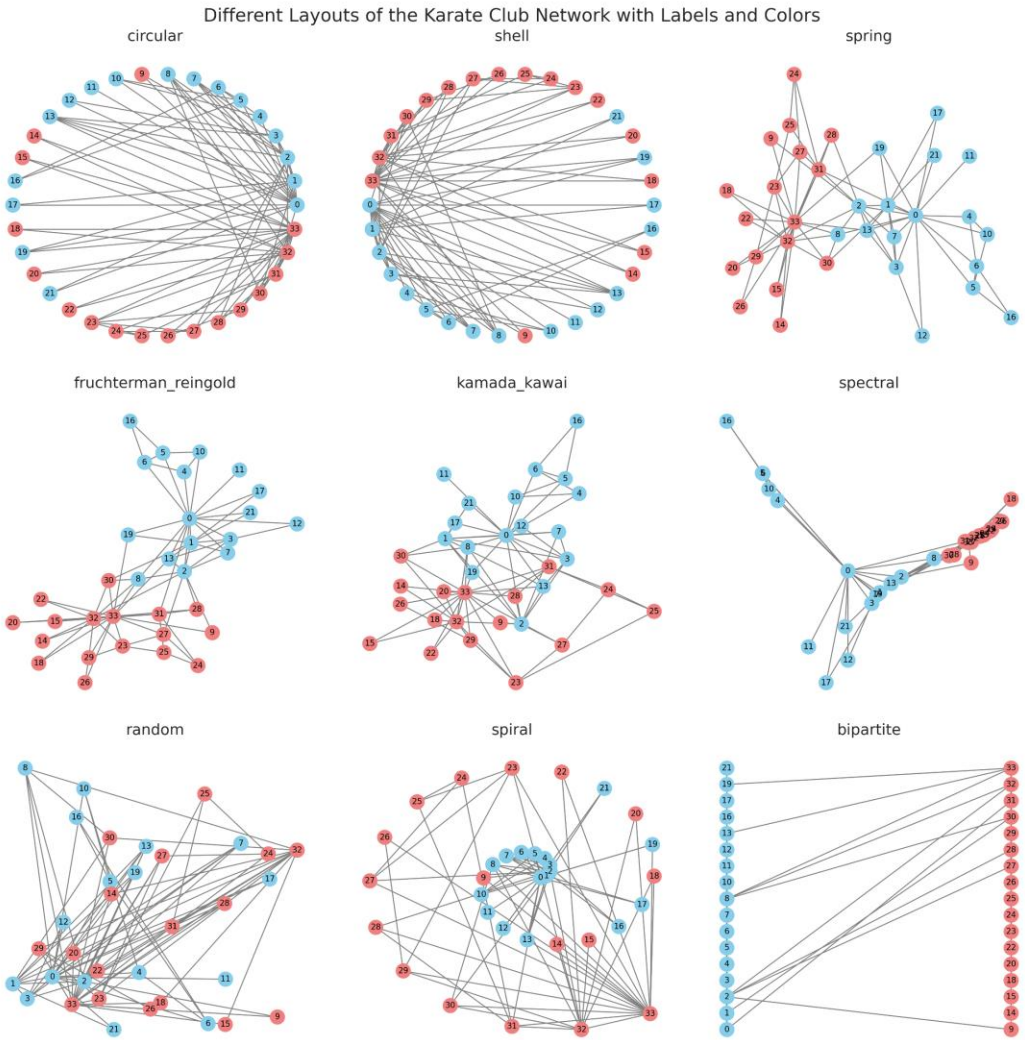


➔ Effectiveness of Fertilizer B underexamined, check range reveals true effects.

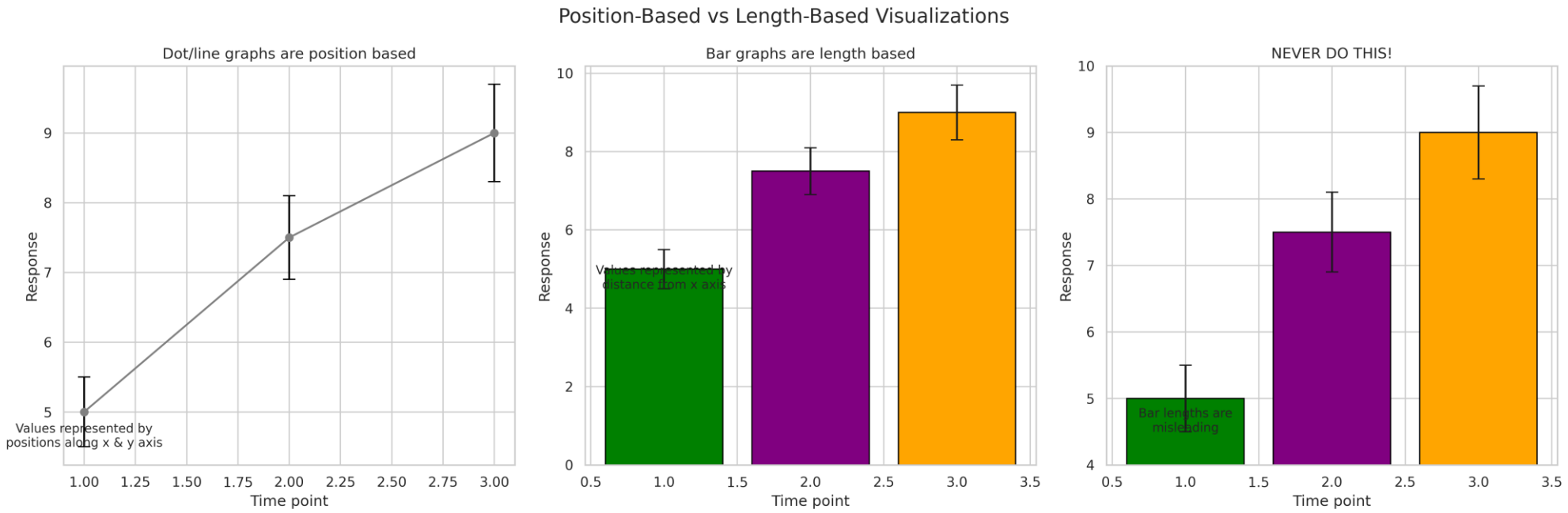
Part 2

Common Pitfalls of Data Visualization

8. Try Different Layouts for Network Graphs



9. Understand Position vs. Length-Based Visualizations

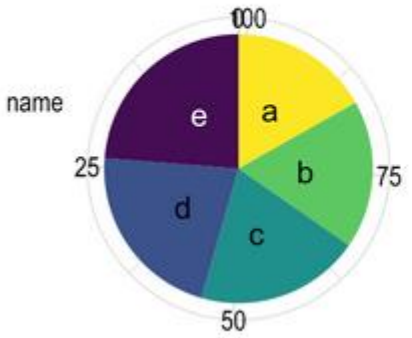


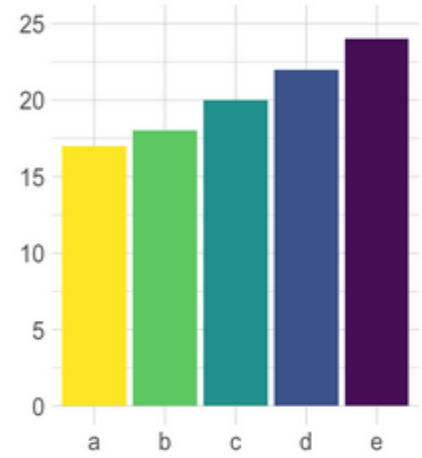
➔ Avoid broken axes in length-based visualizations (i.e. Bar plots).

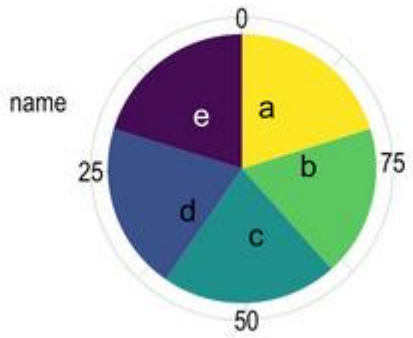
Part 2

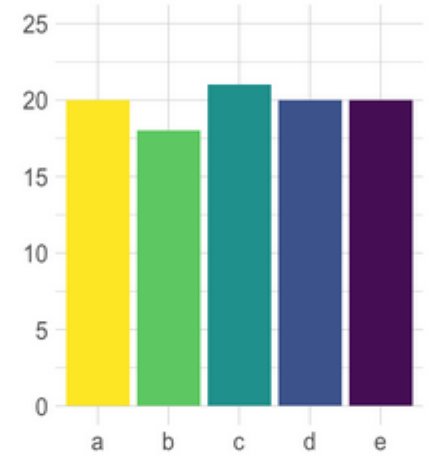
Common Pitfalls of Data Visualization

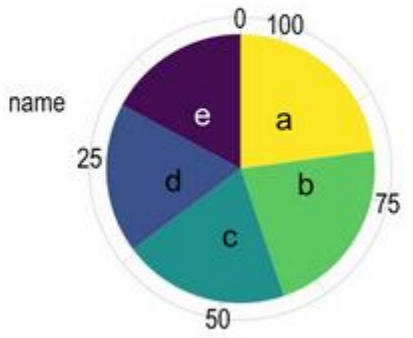
10. No Room for Pie Charts

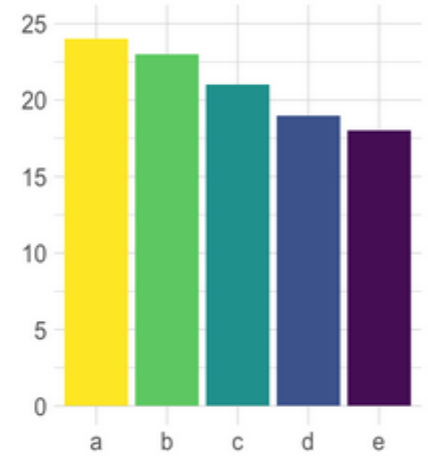






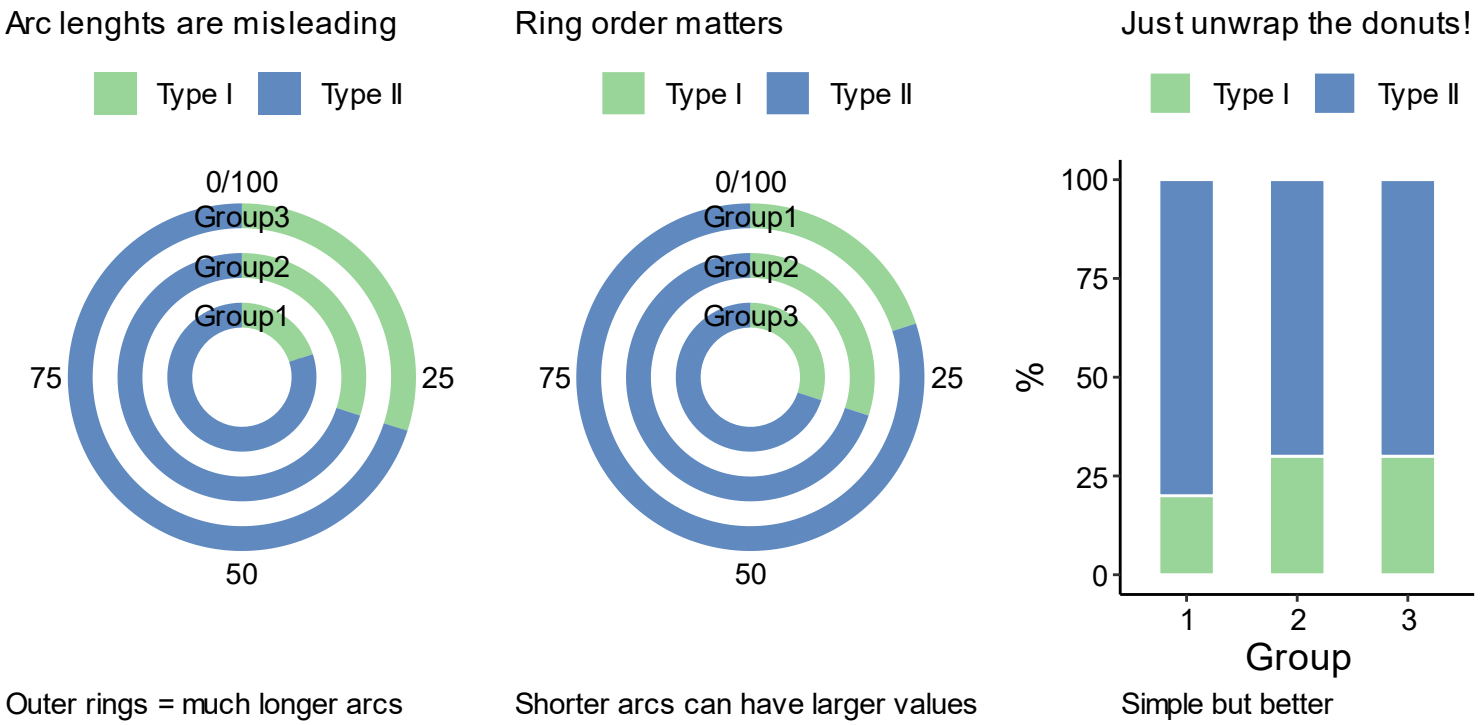






https://www.data-to-viz.com/caveat/pie.html

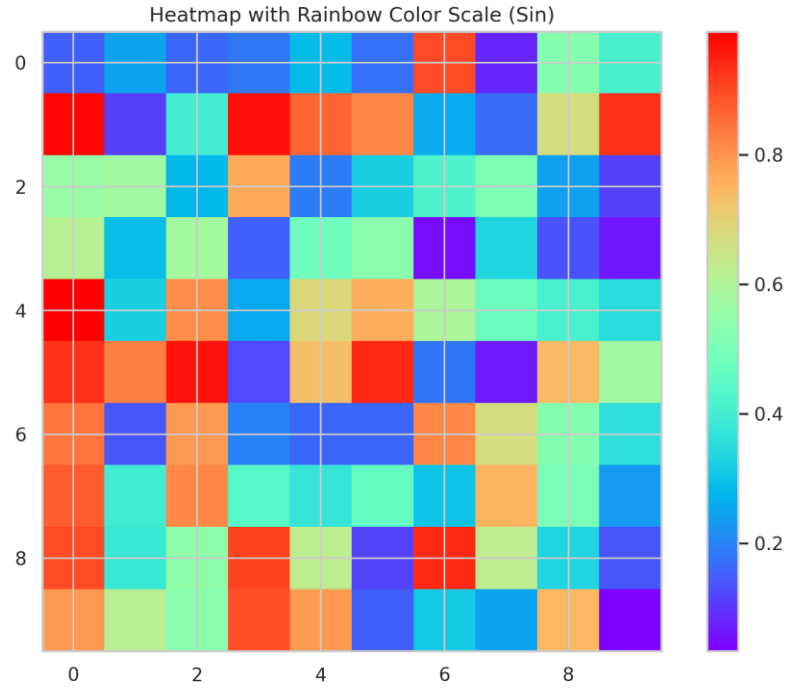
11. Avoid Concentric Donuts

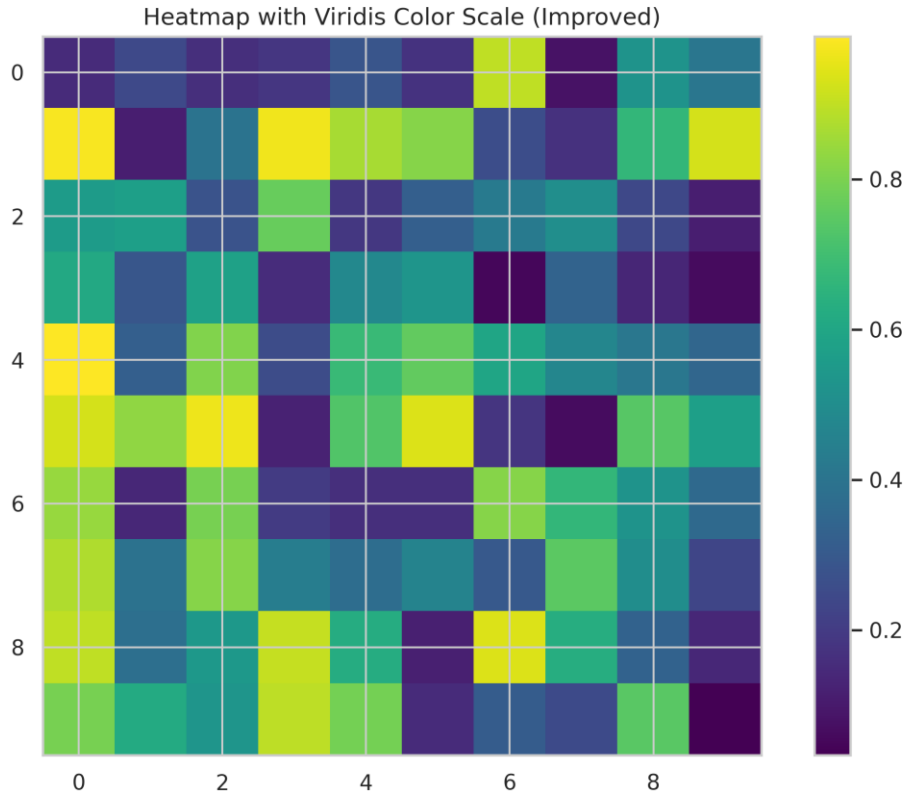


➔ Length of arcs do not accurately represent data; use stacked bars.

https://github.com/cxli233/FriendsDontLetFriends/blob/main/Results/dont_concentric_donuts.svg

12. Use Red/Green Colorblind-Friendly Scales





➔ 1/16 and 1/256 of male and female suffer red/green color blindness; use better color scales (i.e. Viridis)

12. Colorblind-friendly scale

d3.**interpolateViridis**(*t*) <>

Given a number *t* in the range [0,1], returns the corresponding color from the "viridis" perceptually-uniform color scheme designed by van der Walt, Smith and Firing for matplotlib, represented as an RGB string.

d3.**interpolateInferno**(*t*) <>

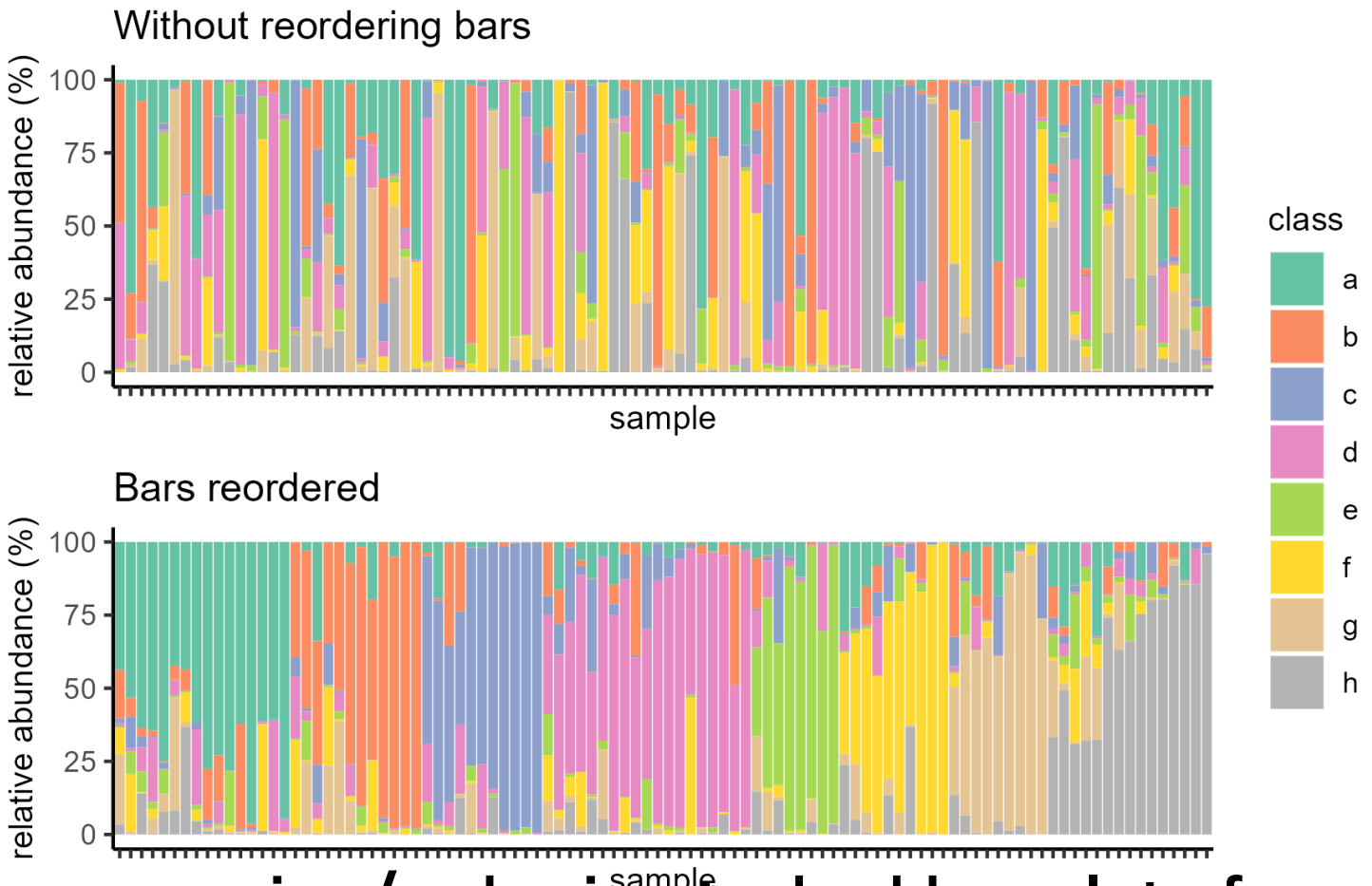
Given a number *t* in the range [0,1], returns the corresponding color from the "inferno" perceptually-uniform color scheme designed by van der Walt and Smith for matplotlib, represented as an RGB string.

d3.**interpolateMagma**(*t*) <>

Given a number *t* in the range [0,1], returns the corresponding color from the "magma" perceptually-uniform color scheme designed by van der Walt and Smith for matplotlib, represented as an RGB string.

➔ Colorful, perceptually uniform, colorblind safe, monotonically increasing luminance

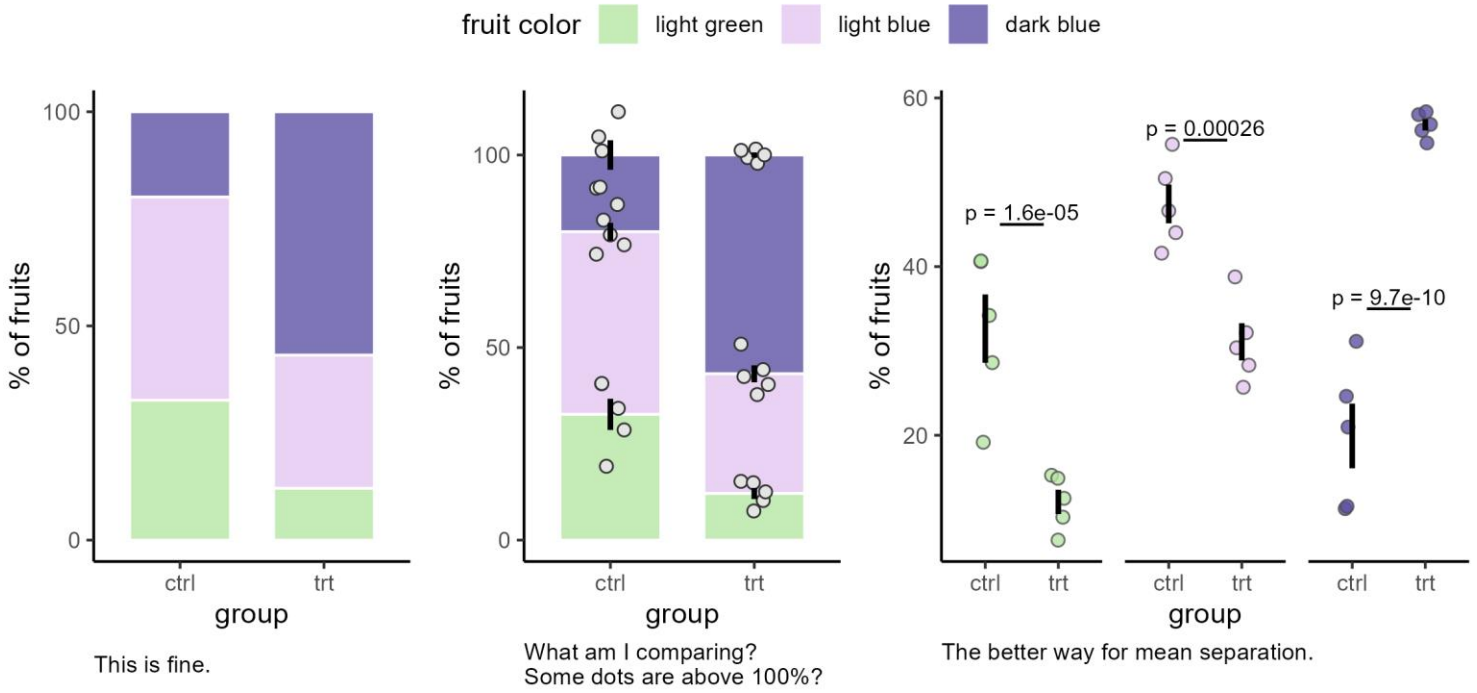
13. Reorder Your Stacked Bars



➔ Optimize grouping/order in stacked bar plots for easier interpretation.

https://github.com/cxli233/FriendsDontLetFriends/blob/main/Results/Reorder_stacked_bars.png

14. Don't Mix Stacked Bars and Mean Separations



➔ Use separate visuals for proportional and mean separation data.

https://github.com/cxli233/FriendsDontLetFriends/blob/main/Results/stacked_bar_vs_jitter.png

Conclusion

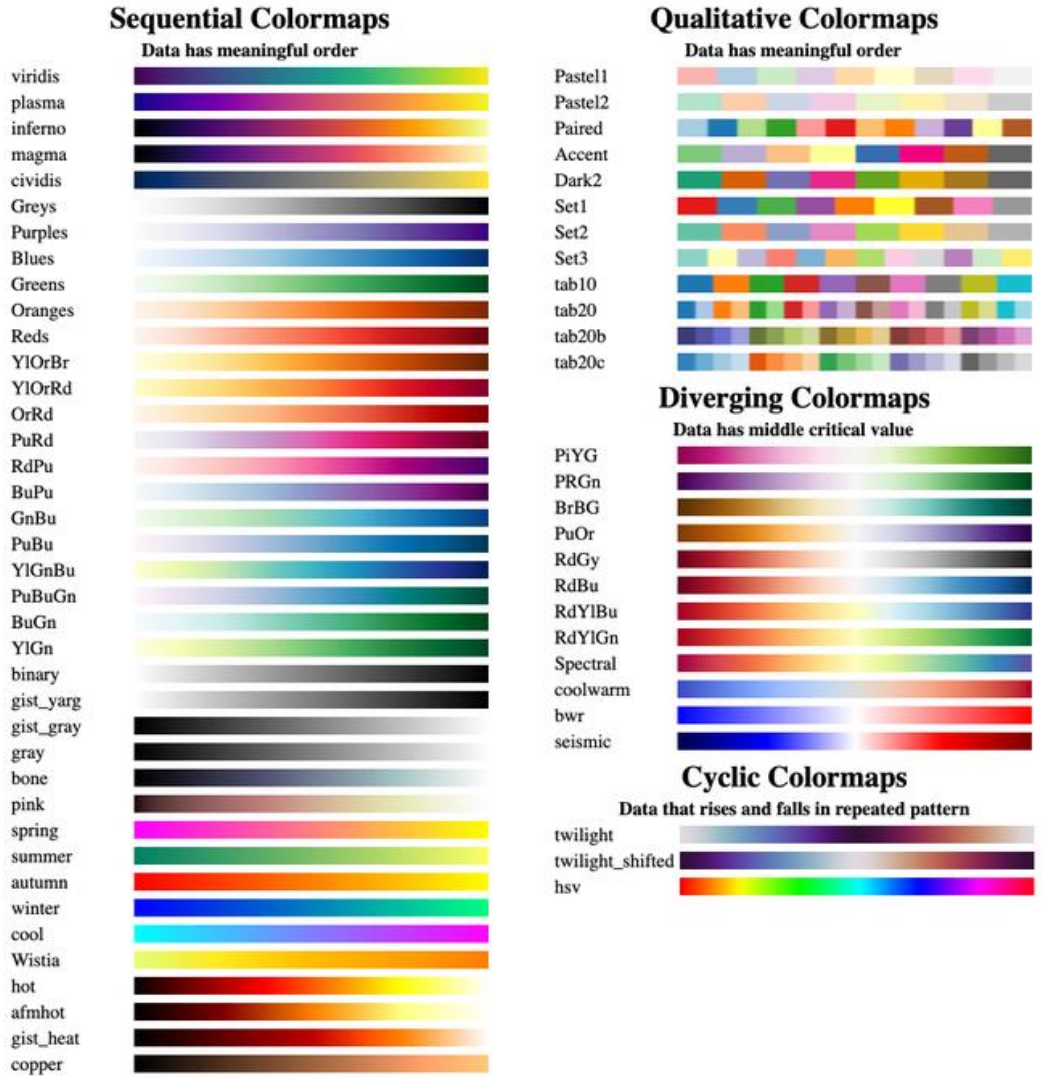
- Use “appropriate” visualization techniques when visualizing result of (paper, report..) to enhance the interpretability of reader.
- Python code for visualizing for 14 cases are: https://for8821.synology.me:5001/d/s/10mReatWt66PCycINFHs4BQydeLXOwxg/jkaGLyakuSIZ7hr9omf1wdPy8N2_4fP--CrJAstFeyws

Thank You! Any Questions?

Part 4

Appendix

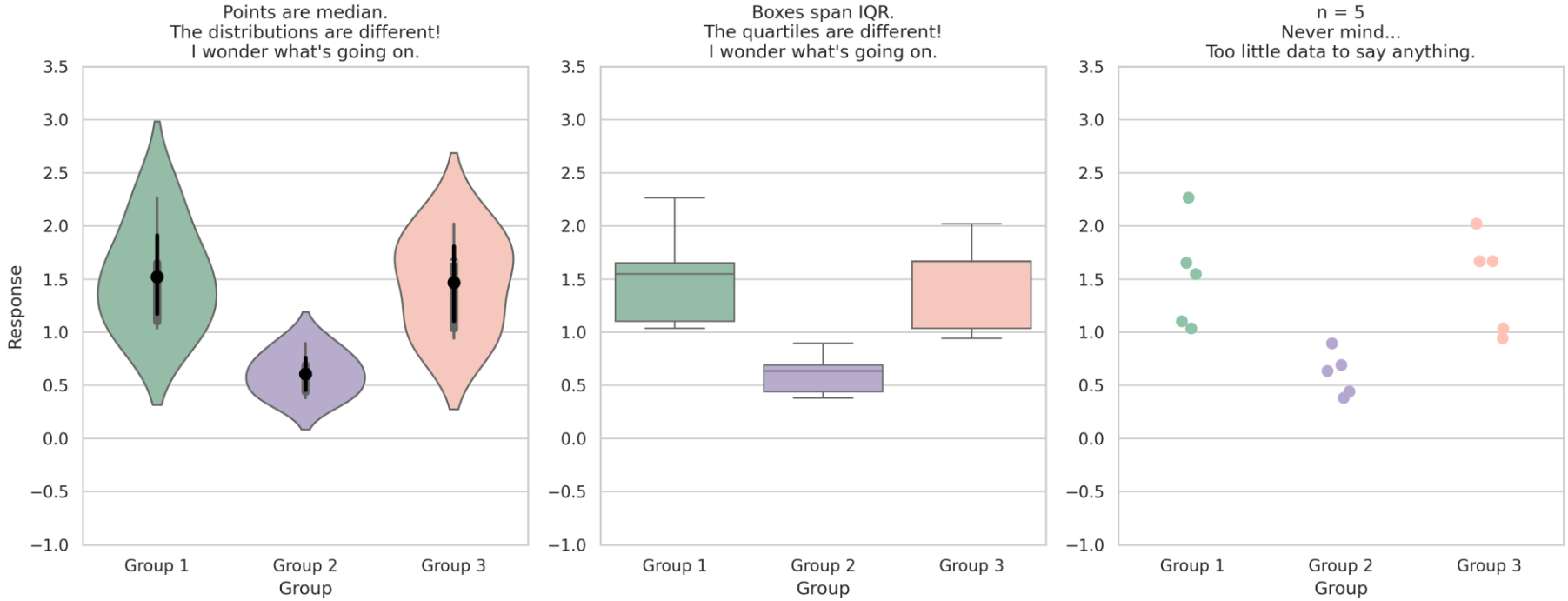
Colormap for matplotlib



https://x.com/TedPetrou/status/1246110879118176257

2. Beware of Violin Plots for Small Sample Sizes

```python
data = pd.DataFrame({
    'Group': np.repeat(groups, 5),
    'Response': np.concatenate([
        np.random.normal(1.2, 0.7, 5),
        np.random.normal(0.5, 0.25, 5),
        np.random.normal(1.9, 0.5, 5)
    ])
})
```



➔ Small N can say anything, even violin plots seems similar.