

Color Palettes for Data Visualization



The first step to finding the right color combination is know your data



Qualitative palette is suitable for describing different categories of data. These data have various labels and are not in a fixed order. A high degree of discrimination between hues is required.



Sequential palette is suitable for presenting ordered/continuous data. Use the color gradient effect to display the data from low to high.



Diverging palette is suitable for data containing extremely low and extremely high values. There is usually a midpoint. Use two different gradient colors to emphasize the contrast.

Qualitative

Here is an example of ggplot2 provides a qualitative color scale by default

```
ggplot(iris, aes(Petal.Length, Sepal.Length, colour = Species)) + geom_point()
```

Here is an example of using package "RcolorBrewer"

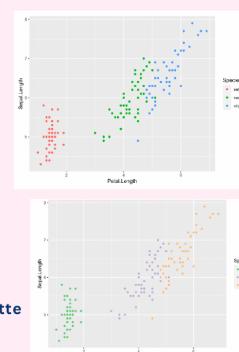
```
ggplot(data, aes()) + <GEO_FUNCTION>+ scale_color_brewer(palette = "Accent")
```

We can replace the palette with the following qualitative color scales from package RColorBrewer:

Accent, Dark2, Paired, Pastel1, Pastel2, Set1, Set2, Set3

Our favorites

INDUSTRY / ACADEMY



Sequential

In R, we can set color to a numeric variable in data. It will automatically produce a gradient of color based on the data range. We can use the built-in package "RcolorBrewer".

```
ggplot(data, aes()) + <GEO_FUNCTION>+ scale_color_brewer(palette = "Purples")
```

We can replace the palette with the following sequential color scales from package RColorBrewer: YlOrRd, Reds, RdPu, PuBuGn, Oranges, Greys, BuGn,

We can also use the "viridis" package. This package can help the colorblindness read the plot better.

```
ggplot(data, aes()) + <GEO_FUNCTION>+ scale_color_viridis(option = "A")
```

We can replace the option with the following sequential color scales from package viridis: magma, plasma, inferno, viridis

Our favorites

INDUSTRY / ACADEMY



Diverging

Like the two palettes above, RcolorBrewer contains different diverging color schemes.

```
ggplot(data, aes()) + <GEO_FUNCTION>+ scale_color_brewer(palette = "RdBu")
```

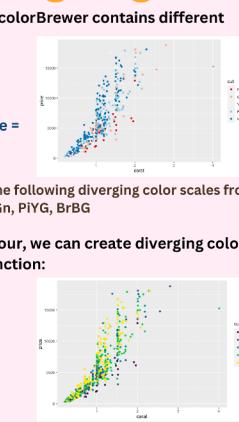
We can replace the palette with the following diverging color scales from package RColorBrewer: PuOr, PRGn, PIYG, BrBG

Instead of customizing the colour, we can create diverging color gradient by using following function:

```
ggplot(data, aes()) + <GEO_FUNCTION>+ scale_fill_gradient2()
```

Our favorites

INDUSTRY / ACADEMY



Now that you have looked at the three palettes, you know how to use them in R and what kind of data to apply them. Let's apply these three palettes to our commonly used plots to see how they work.

Plots with our favorites custom colors

```
ggplot(data, aes()) + <GEO_FUNCTION>+ scale_fill_manual(values = c("#4F596D", "#2E86C1", "#FAC00F", "#A73030", "#5FC6C9", "#7E318A"))
```



Other commonly used plots:
Alluvial diagrams, Biplots, Boxplots, Cleveland dot plots, Mosaic plots, Parallel Coordinates plots, Scatter plots, Smooth plots

Tips

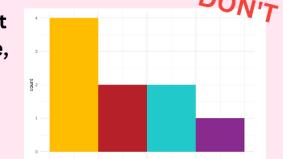
- The data should be unordered nominal data because no ordered colors correspond to no ordered categories, in which any color is not more important than the other color.



In this example, we can't tell which state has the highest crime rate.

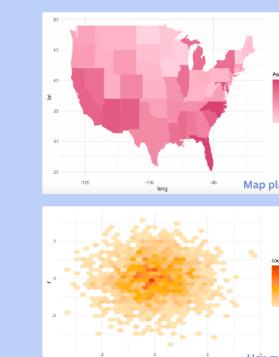
Qualitative PLOTS AND TIPS

- When presenting different values of the same variable, avoid using the qualitative palette because it may not highlight the key point of the data.



Plots with our favorites custom colors

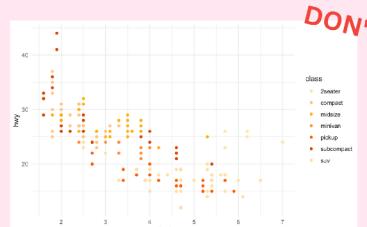
```
ggplot(data, aes()) + <GEO_FUNCTION>+ scale_fill_gradientn(colors = c("#FFD5E0", "#F6BBCA", "#ECA0B4", "#E2869F", "#D66B89", "#CA4E74"))
```



Other commonly used plots:
Mosaic plots, Heatmaps, Parallel Coordinates plots, Tile plots, Contour plot

Tips

- The data should be numeric or ordered
- It is best to use light colors for low values and dark colors for high values.
- Only one or two hues can be used in the sequential palette.

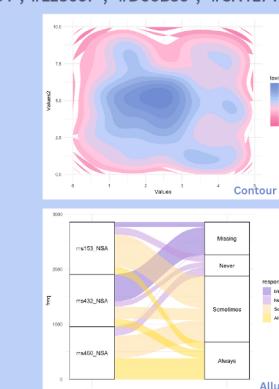


Sequential PLOTS AND TIPS

- Avoid using this type of color palette in category data, which will convey more important information about one particular category to the viewer, just like the figure above.

Plots with our favorites custom colors

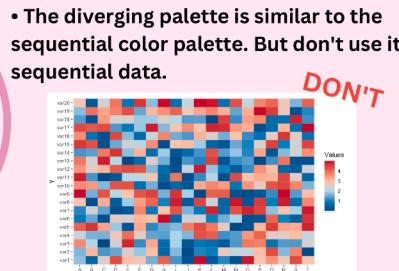
```
ggplot(data, aes()) + <GEO_FUNCTION>+ scale_fill_gradientn(colors = c("#FFD5E0", "#F6BBCA", "#ECA0B4", "#E2869F", "#D66B89", "#CA4E74"))
```



Other commonly used plots:
Mosaic, Heatmaps, Hexmaps, Density plot, Map plots

Tips

- Used the diverging palette when data contains values with opposing relationships. Therefore the two extremes should be the two darkest colors with contrasts. The ends fade to the middle, with the median value being the lightest color.



Diverging PLOTS AND TIPS

- The diverging palette is similar to the sequential color palette. But don't use it with sequential data.

