Advanced Data Visualization and Analysis in R

Matthew Galbraith
Linda Crnic Institute for Down Syndrome

Data Science for Developing Scholars in Down Syndrome Research (DS3) 2025

INCLUDE
INvestigation of Co-occurring conditions across the Lifespan to Understand Down syndromE

Linda Crnic Institute for Down Syndrome
UNIVERSITY OF COLORADO ANSCHUTZ MEDICAL CAMPUS

GLOBAL
DOWN SYNDROME FOUNDATION®

Code links for this session

https://github.com/DS3-2025/HTP_linear_regression_example

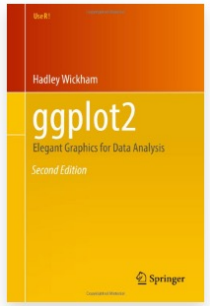
https://github.com/DS3-2025/HTP_DESeq2_analysis

https://github.com/DS3-2025/HTP_single_cell_dataviz_mass_cytometry

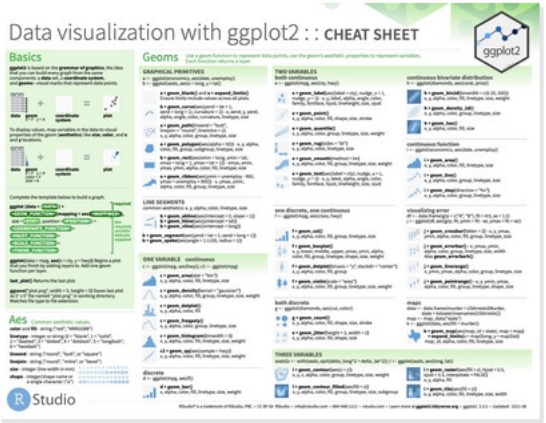
Visualization using ggplot2

Publication-quality data visualization

- Implements a “grammar of graphics”
- Start by defining the data to be plotted (“aesthetics”):
  `ggplot(aes(x, y, color, fill, shape, alpha, linetype))`
- Then add layers (“geoms”) to specify how data is plotted, eg:
  `+ geom_point()`
- Add additional geom layers, eg:
  `+ geom_boxplot()`
- Can split into separate plots, eg male vs. female, by “faceting”:
  `+ facet_wrap(~ Sex)`
- Finally add title and modify theme:
  ```
  + labs(title = “Plot title”, subtitle = “plot details”
  + theme(aspect.ratio = 1)
  ```



https://ggplot2-book.org/



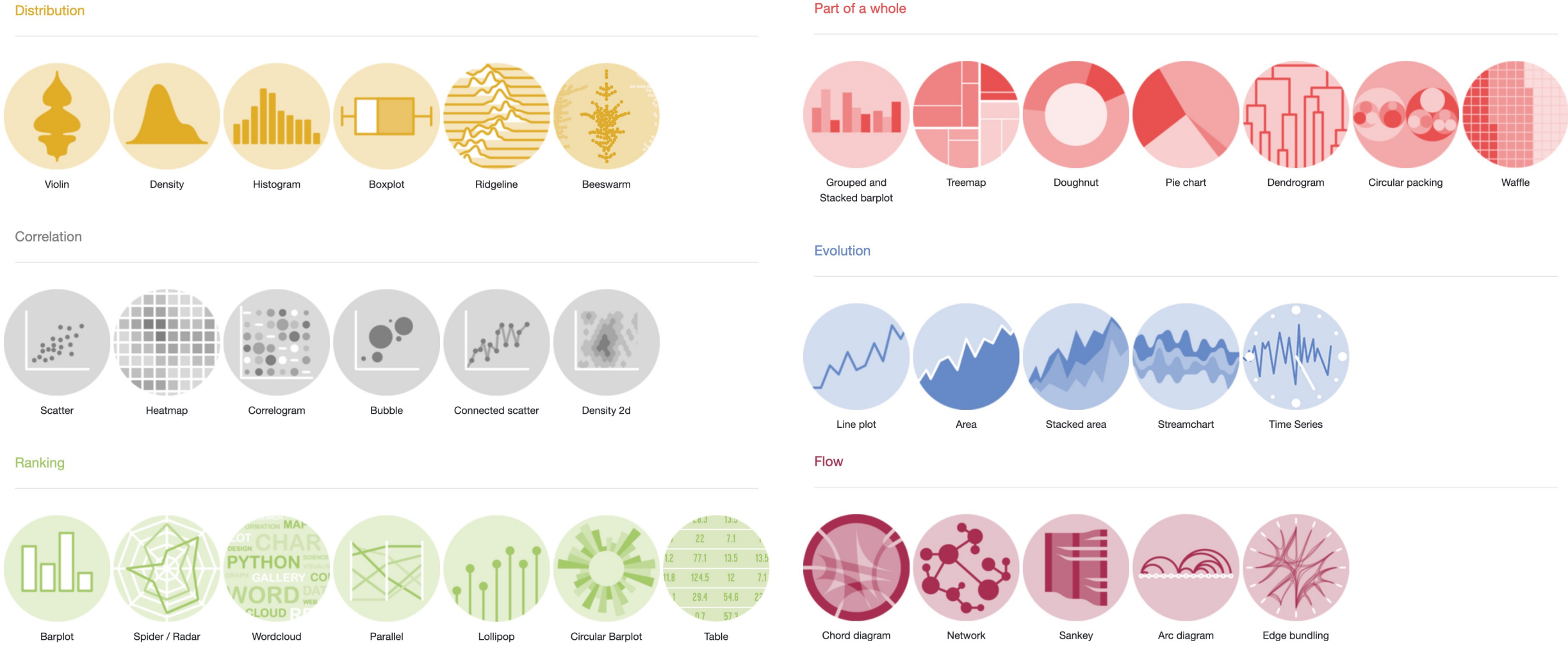
https://raw.githubusercontent.com/rstudio/cheatsheets/main/data-visualization.pdf

https://www.data-to-viz.com/caveats.html

https://r-graph-gallery.com/index.html

Choose your plot type carefully

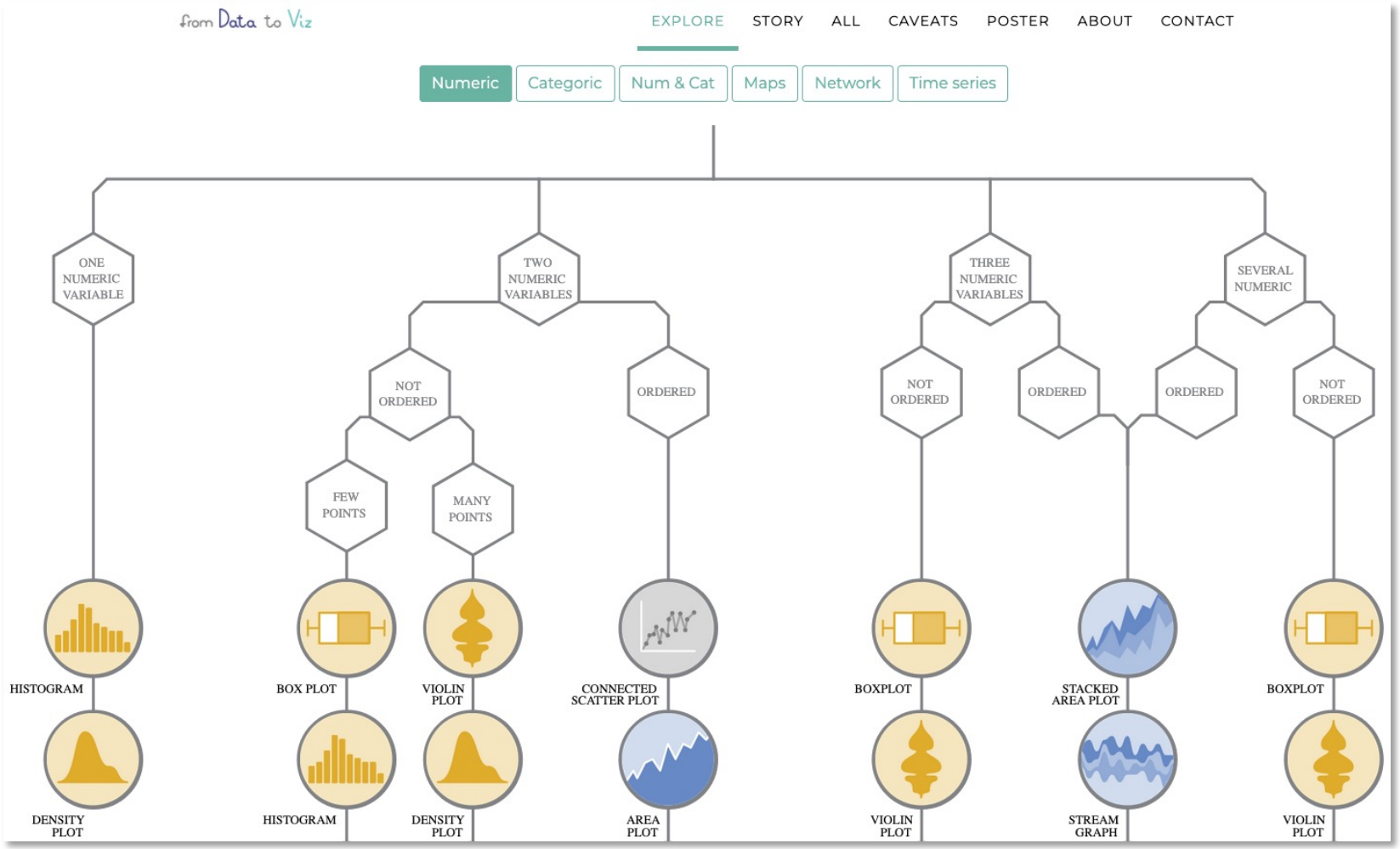
The R Graph Gallery: code examples



https://r-graph-gallery.com/index.html

Choose your plot type carefully

Data to Viz: Decision Tree



https://www.data-to-viz.com/index.html

Choose your plot type carefully
Data to Viz: Caveats

Order your data

When displaying the value of several entities, ordering them makes the graph much more insightful.

To cut or not to cut?

Cutting the Y-axis is one of the most controversial practice in data viz. See why.

The spaghetti chart

A line graph with too many lines becomes unreadable: it is called a spaghetti graph.

Pie chart

The human eye is bad at reading angles. See how to replace the most criticized chart ever.

Play with histogram bin size

Always try different bin sizes when you build a histogram, it can lead to different insights.

Do boxplots hide information?

Boxplots are a great way to summarize a distribution but hide the sample size and their distribution.

The problem with error bars

Barplots with error bars must be used with great care. See why and how to replace them.

Too many distributions.

If you need to compare the distributions of many variables, don't clutter your graphic.

https://www.data-to-viz.com/caveats.html

Choose your color schemes carefully

Perspective | Open access | Published: 28 October 2020

The misuse of colour in science communication

Fabio Crameri ✉, Grace E. Shephard & Philip J. Heron

Nature Communications **11**, Article number: 5444 (2020) | Cite this article

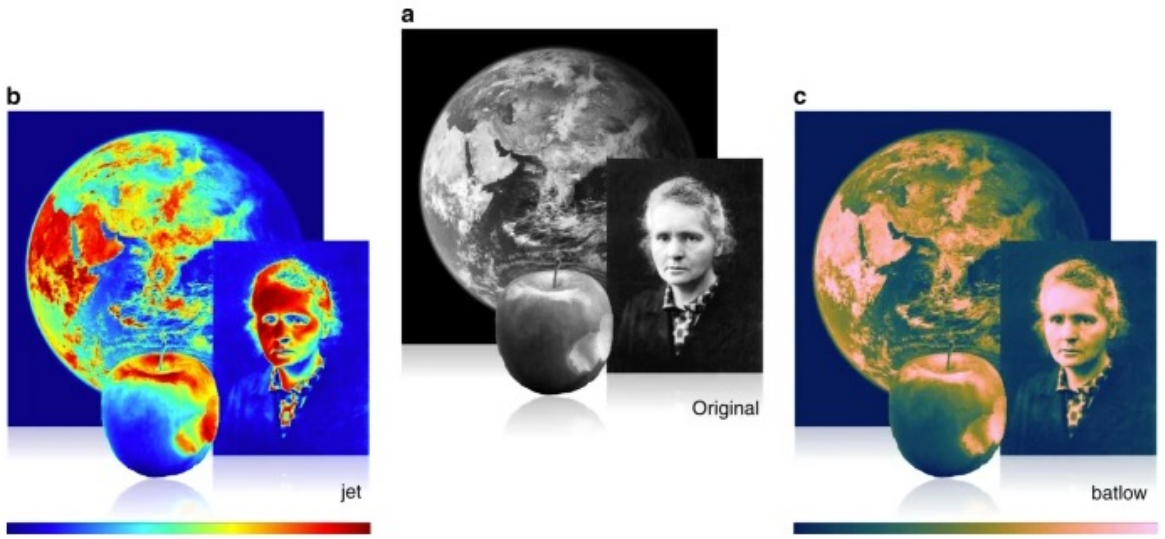
Abstract

The accurate representation of data is essential in science communication. However, colour maps that visually distort data through uneven colour gradients or are unreadable to those with colour-vision deficiency remain prevalent in science. These include, but are not limited to, rainbow-like and red–green colour maps. Here, we present a simple guide for the scientific use of colour. We show how scientifically derived colour maps report true data variations, reduce complexity, and are accessible for people with colour-vision deficiencies. We highlight ways for the scientific community to identify and prevent the misuse of colour in science, and call for a proactive step away from colour misuse among the community, publishers, and the press.

"It is (or should be) in every scientist's best intention to make figures and their content as accurate and easily understandable as possible"

- Misuse of color = effective manipulation of data
- Using an uneven colour gradient is not an action without consequences
- Example: the unscientific rainbow-like colour palette

Fig. 1: The superiority of scientifically derived colour maps.



https://www.nature.com/articles/s41467-020-19160-7

Choose your color schemes carefully

Fig. 4: Perceptual uniformity and order.

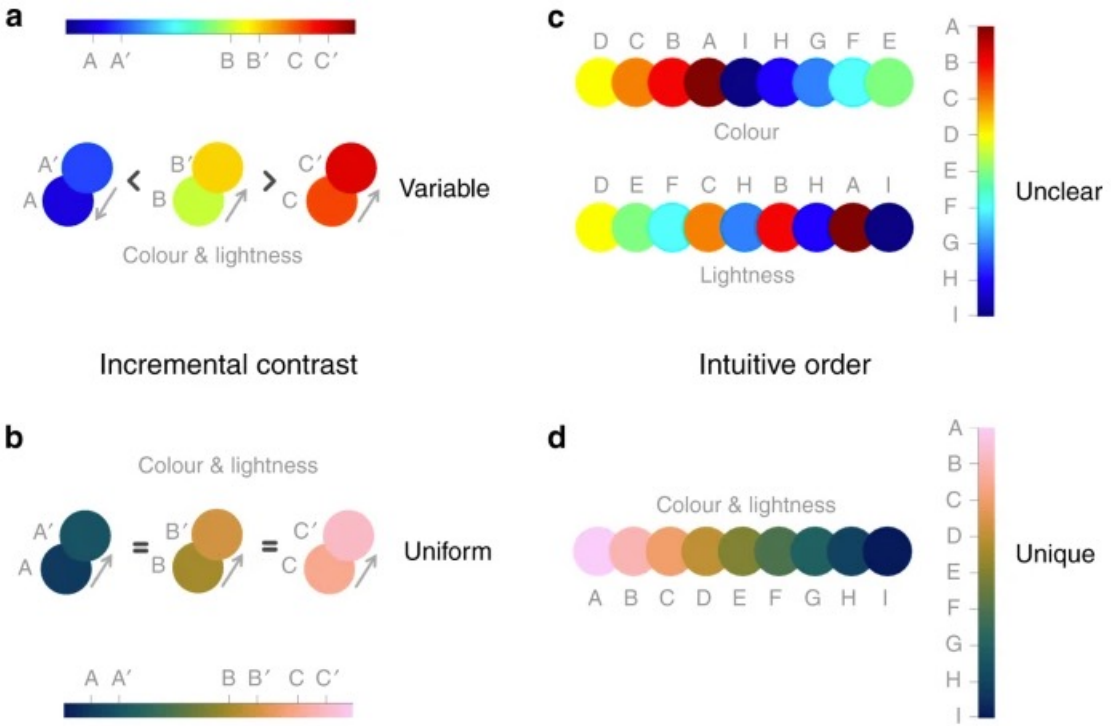
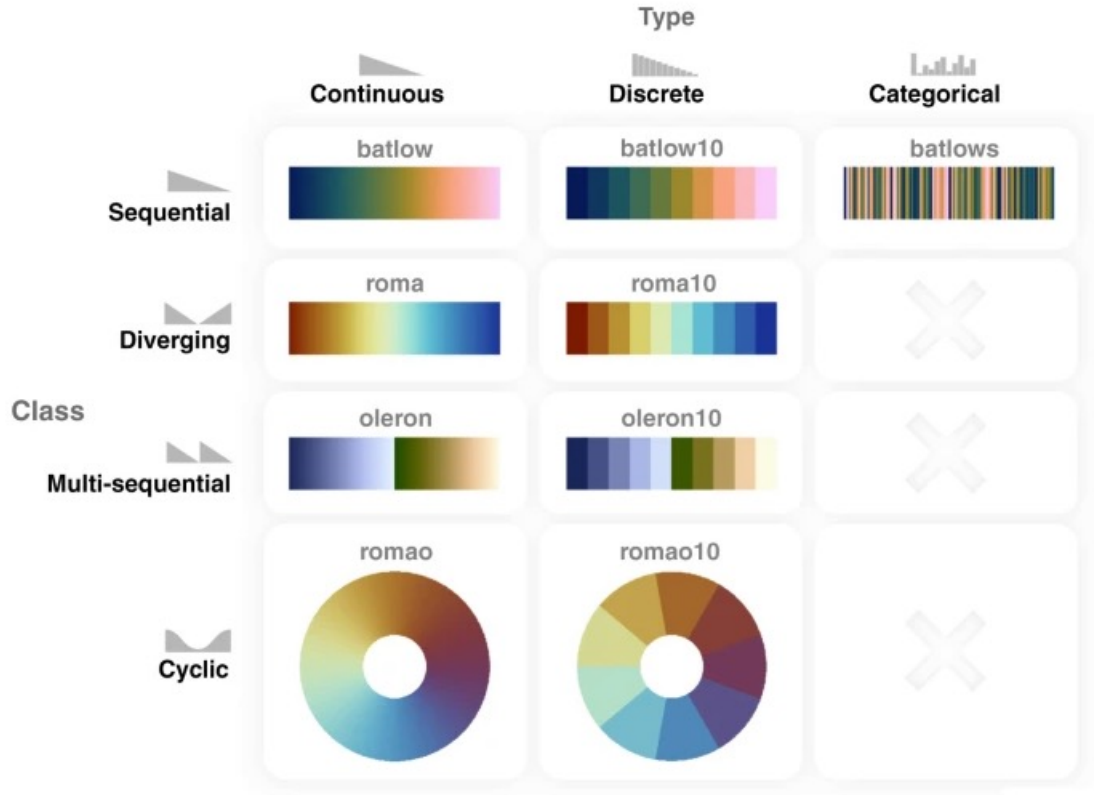


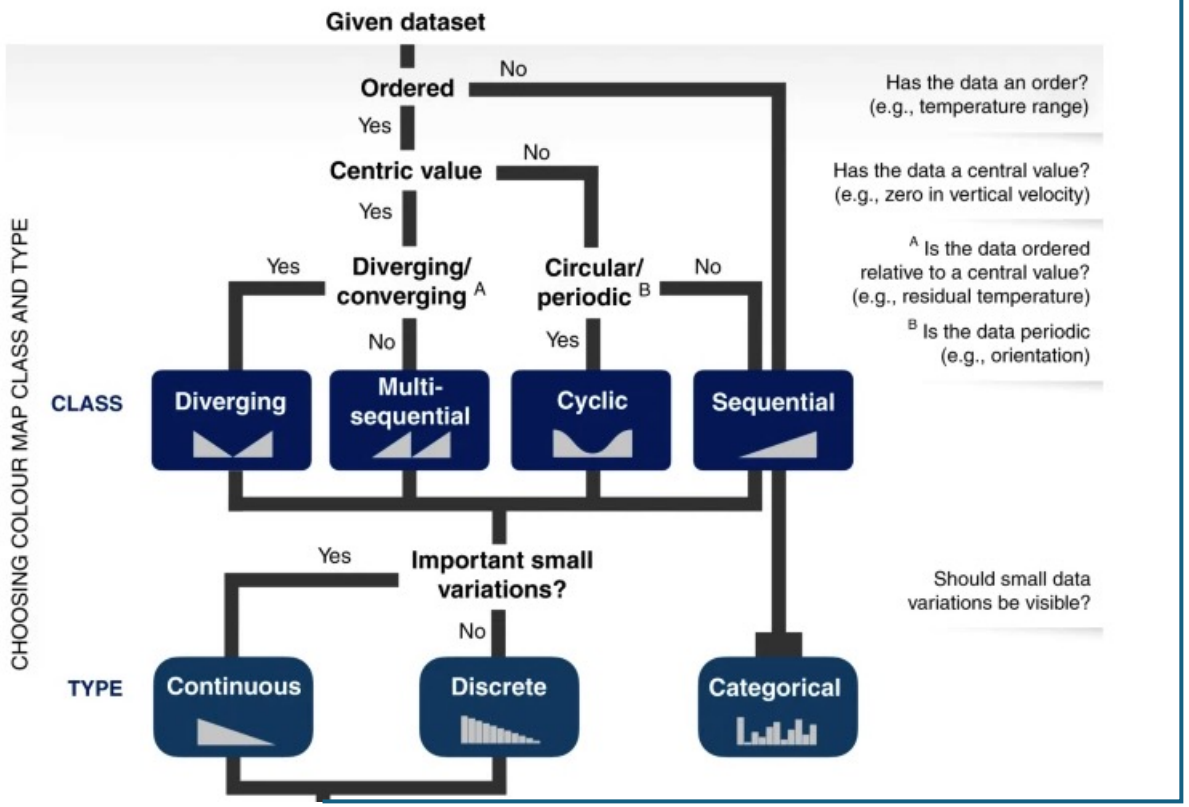
Fig. 5: Colour map classes and types.

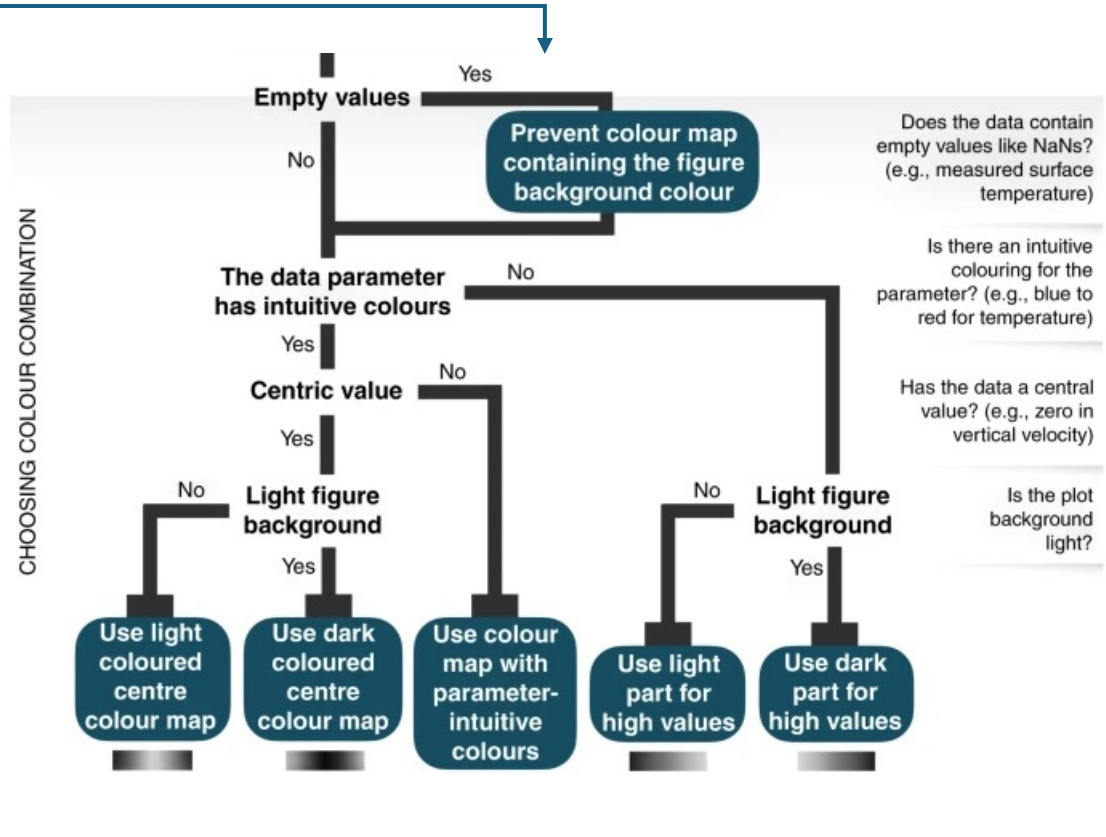


https://www.nature.com/articles/s41467-020-19160-7

Choose your color schemes carefully

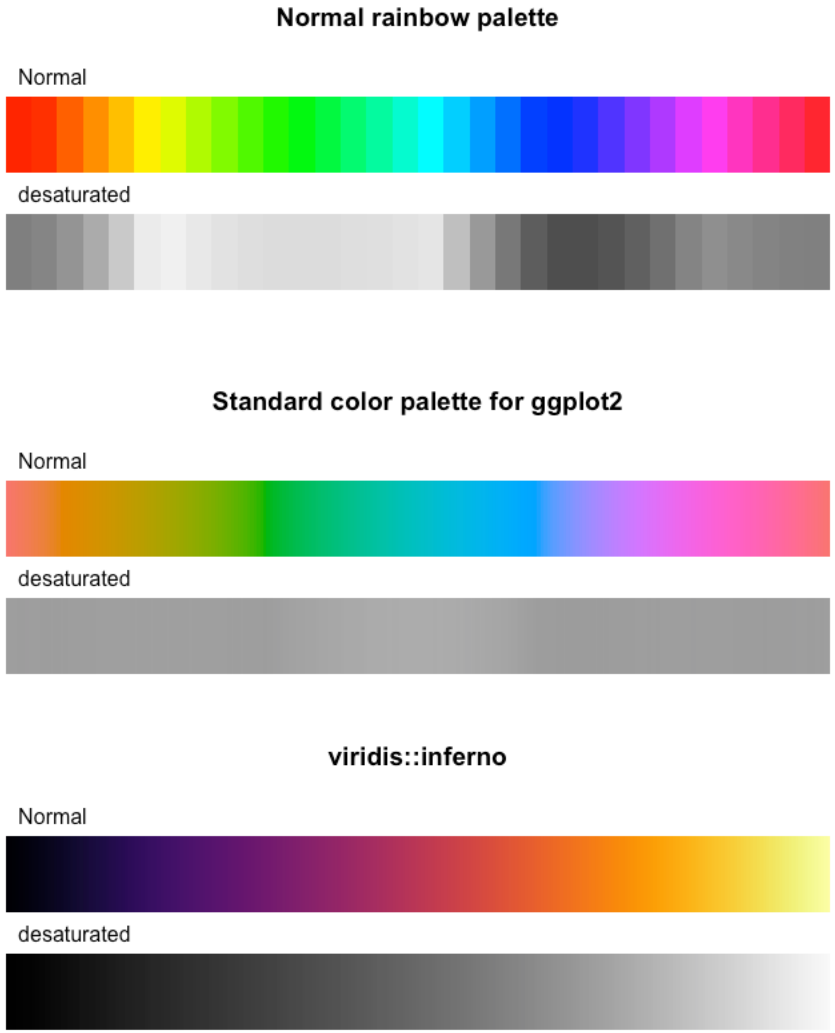
Fig. 6: Guideline for choosing the right scientific colour map.





https://www.nature.com/articles/s41467-020-19160-7

Choose your color schemes carefully



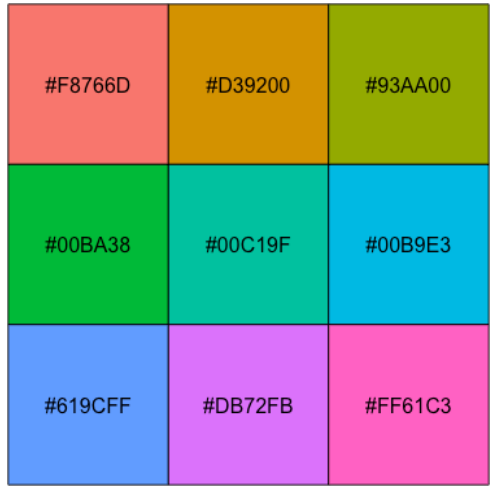
https://github.com/EmilHvitfeldt/r-color-palettes

Default ggplot colors will not always be suitable

ggplot default color colors:

```
gg_color_hue <- function(n) {
  hues = seq(15, 375, length = n + 1)
  hcl(h = hues, l = 65, c = 100)[1:n]
}
```

```
scales::show_col(gg_color_hue(3))
scales::show_col(gg_color_hue(9))
scales::show_col(gg_color_hue(30))
```

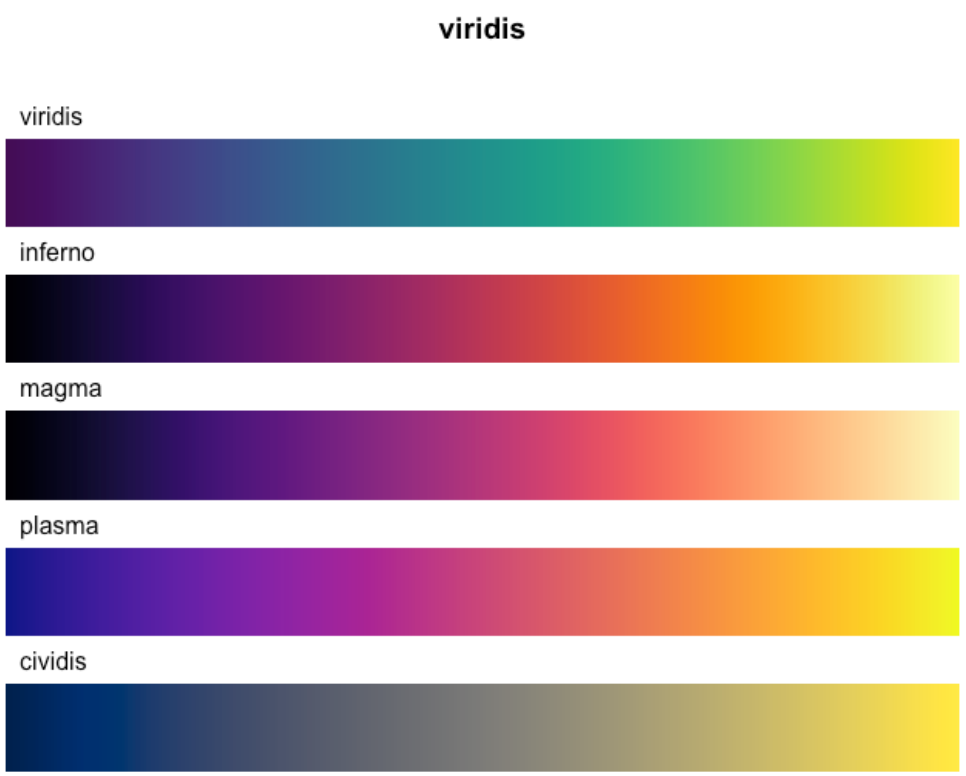


HEX codes obtained elsewhere can be used in ggplot with `*_color_manual()` functions

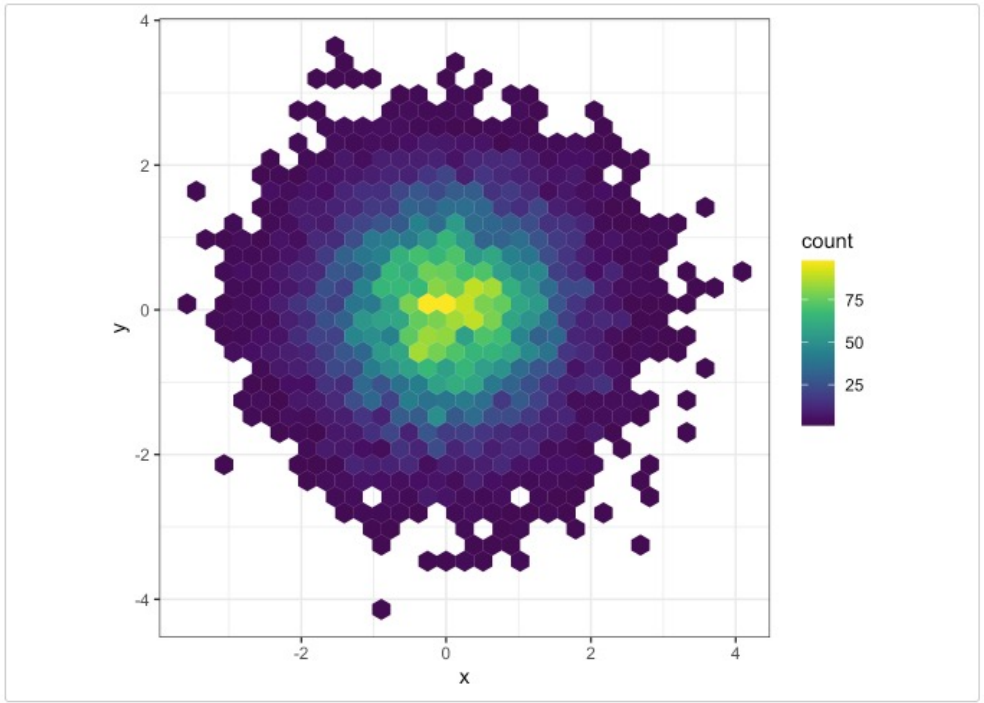
Useful R packages: Viridis

Base R: `install.packages("viridis")`
renv: `renv::install("viridis")`

With ggplot, use `scale_color_viridis()` and `scale_fill_viridis()`:

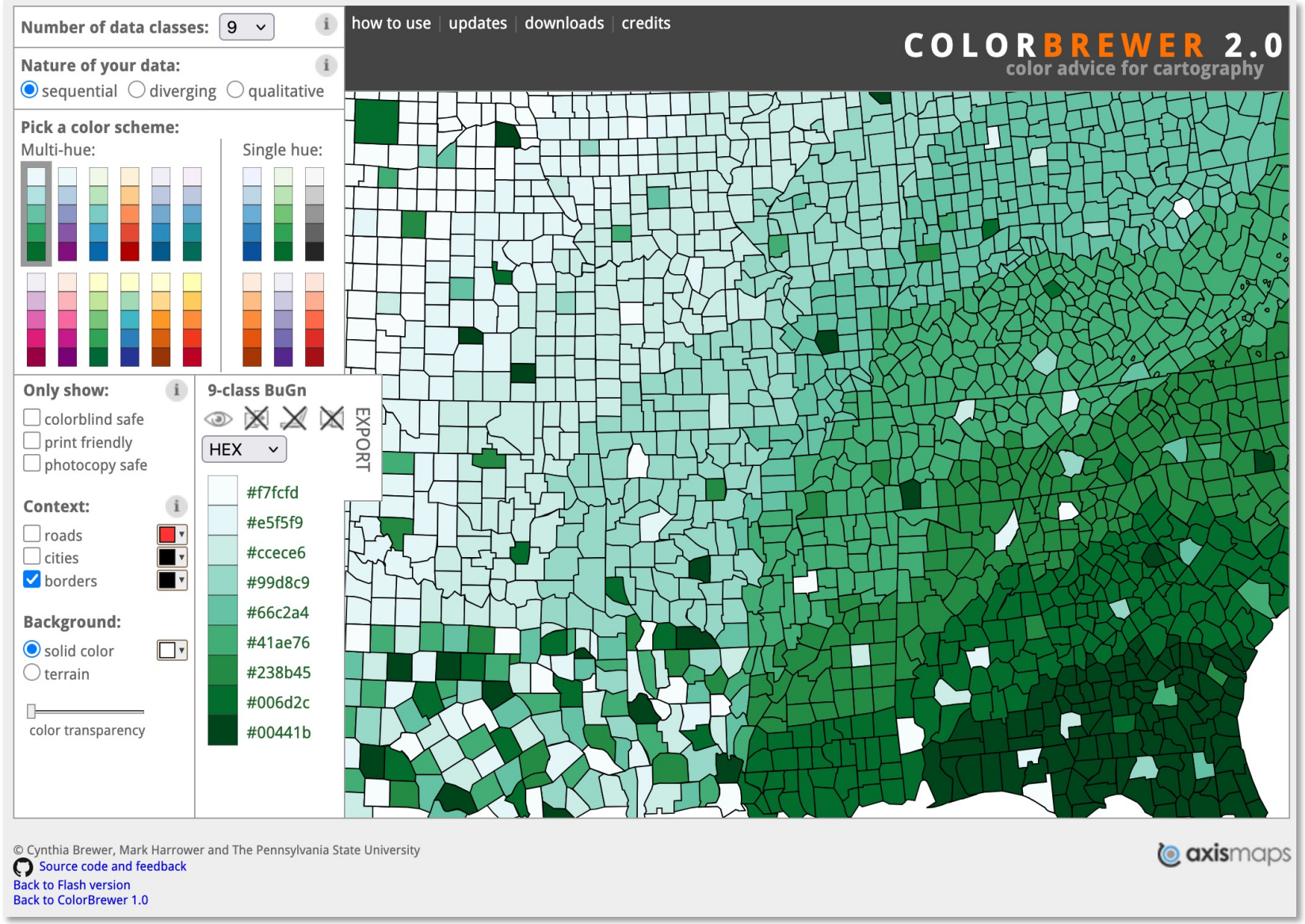



```
library(ggplot2)
ggplot(data.frame(x = rnorm(10000), y = rnorm(10000)), aes(x = x, y = y)) +
  geom_hex() + coord_fixed() +
  scale_fill_viridis() + theme_bw()
```



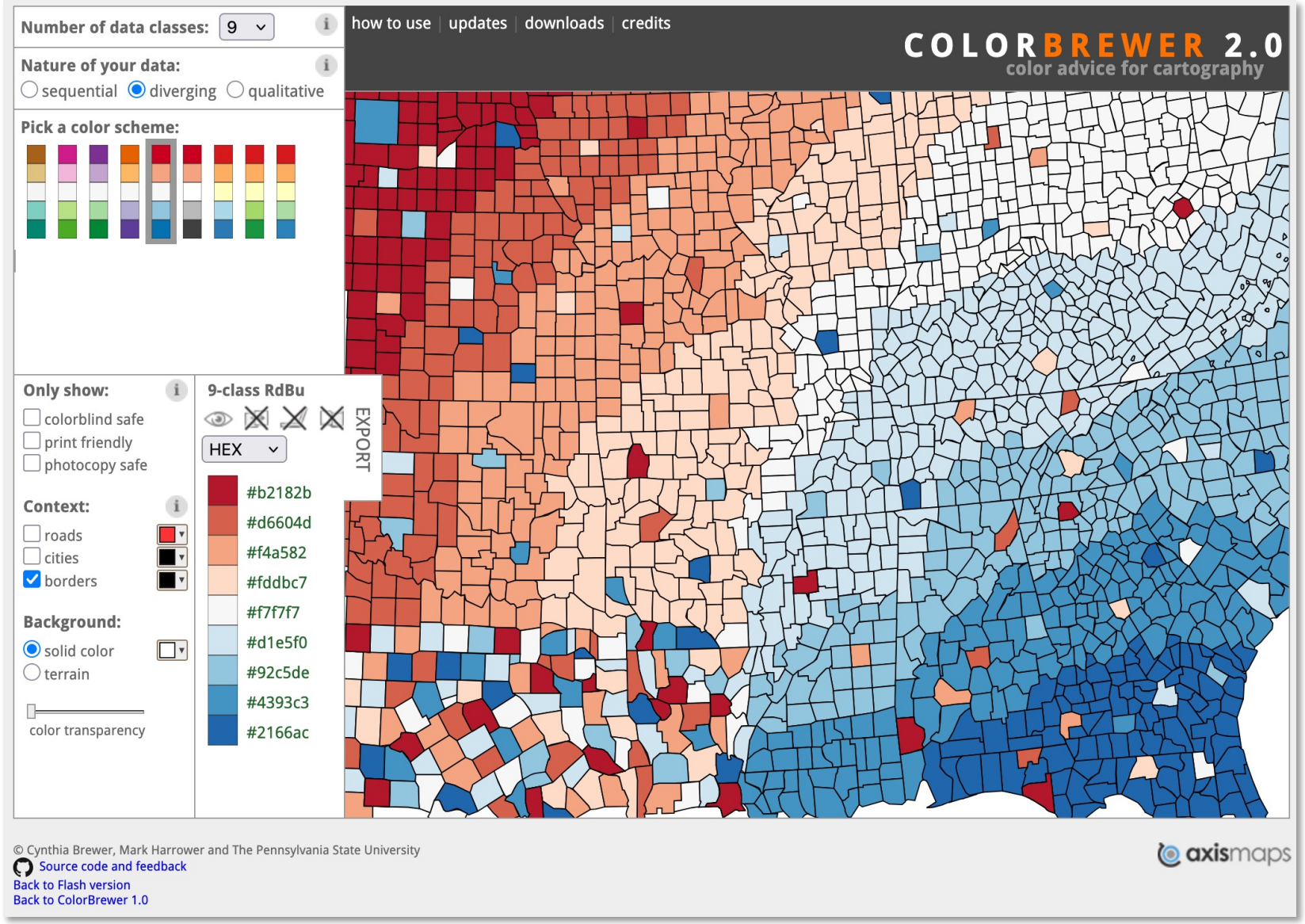
https://cran.r-project.org/web/packages/viridis/vignettes/intro-to-viridis.html

Useful R packages/tools: Color Brewer / RColorBrewer



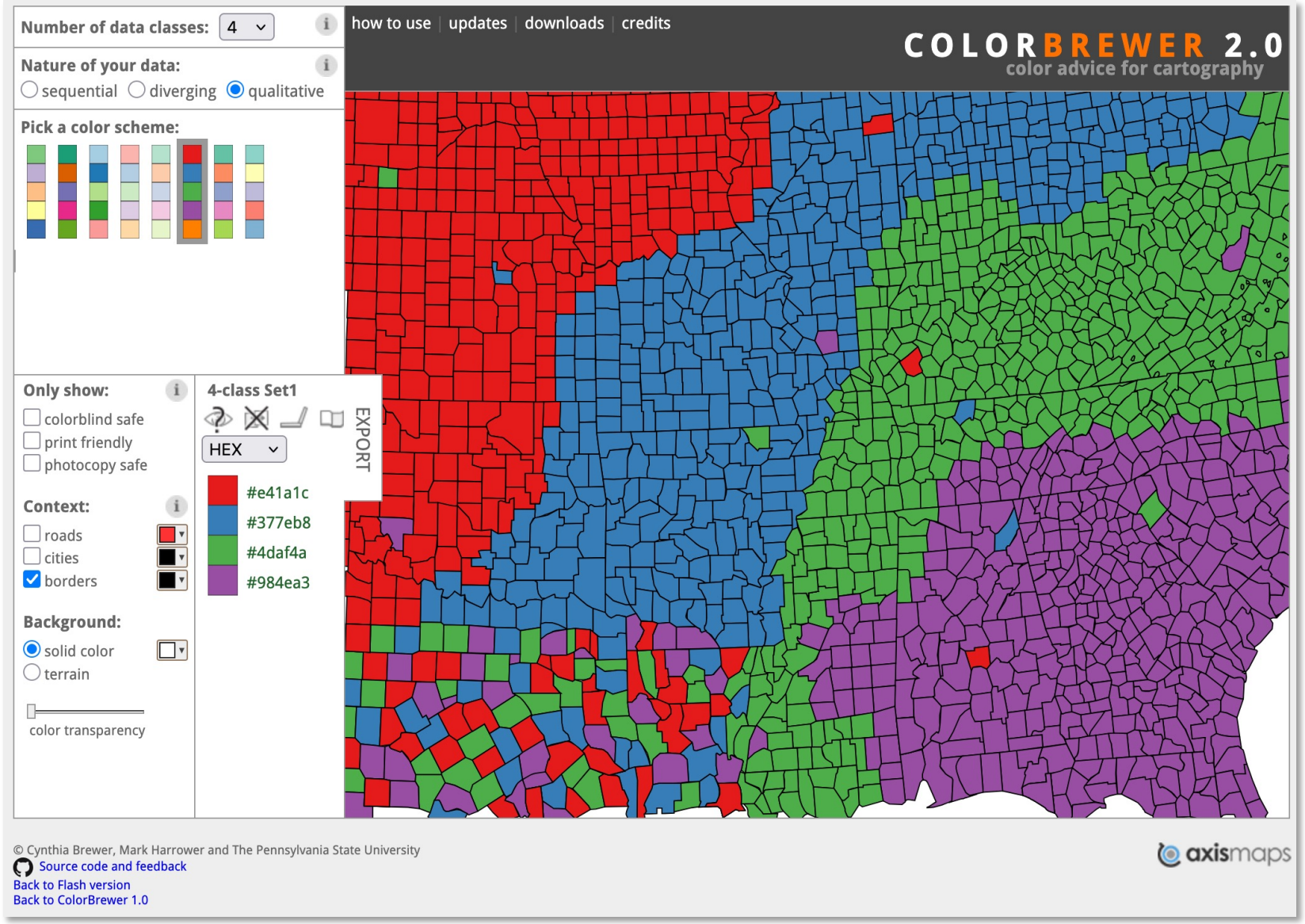
https://colorbrewer2.org

Useful R packages/tools: Color Brewer / RColorBrewer



https://colorbrewer2.org

Useful R packages/tools: Color Brewer / RColorBrewer



https://colorbrewer2.org

Useful R packages/tools: Color Brewer / RColorBrewer

Choice of Sequential vs. Diverging vs. Qualitative color schemes
is especially important for heatmaps

Base R: `install.packages("RColorBrewer")`
renv: `renv::install("RColorBrewer")`

```
brewer.pal(9, "Blues")
```

```
display.brewer.pal(9, "Blues")
display.brewer.pal(9, " RdBu")
display.brewer.pal(9, " RdBu")
```

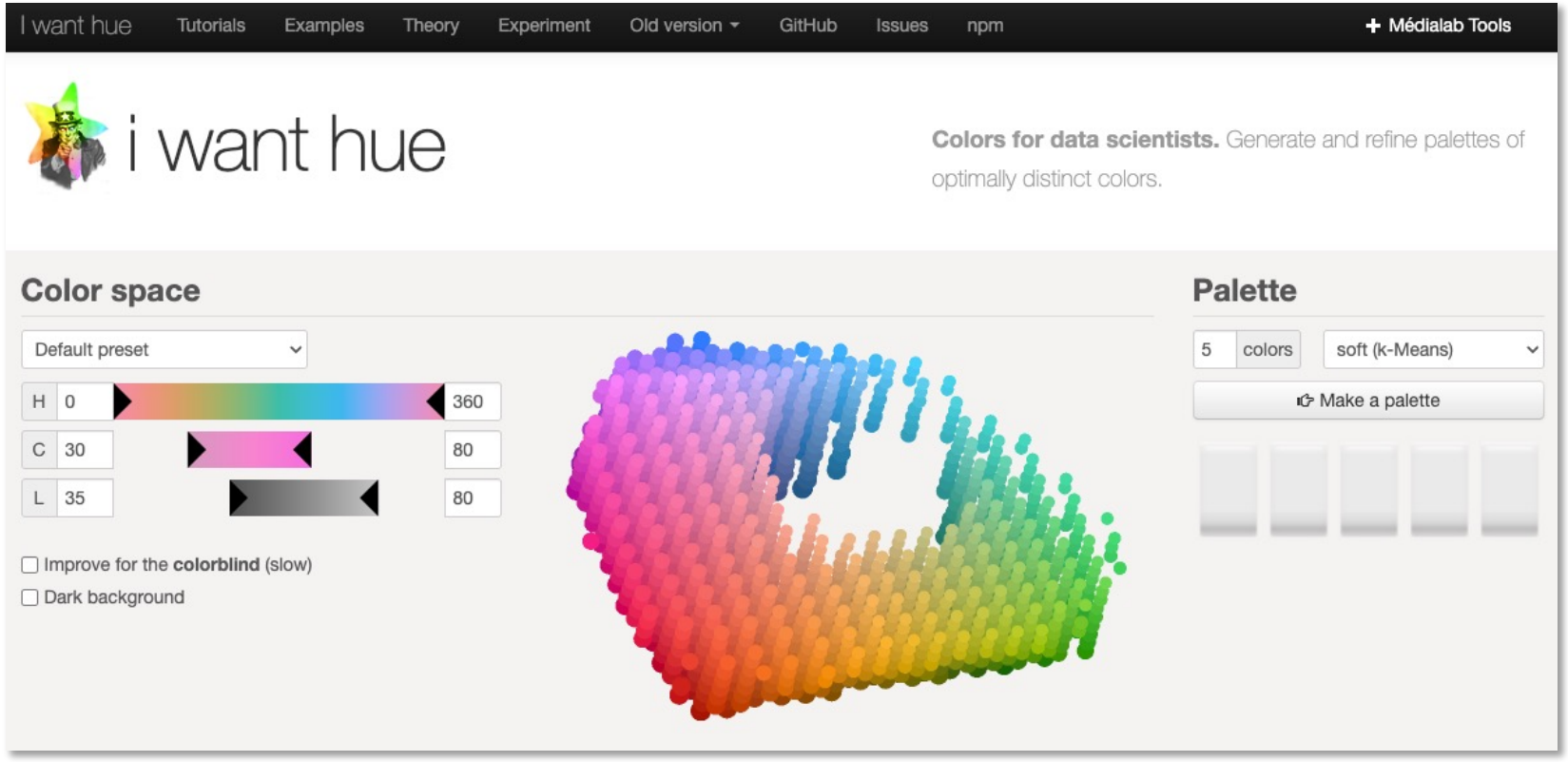
Blues (sequential)

RdBu (divergent)

Set1 (qualitative)

https://cran.r-project.org/web/packages/RColorBrewer/index.html

Useful R packages/tools: I want hue / hues



https://medialab.github.io/iwanthue/

Useful R packages/tools: I want hue / hues

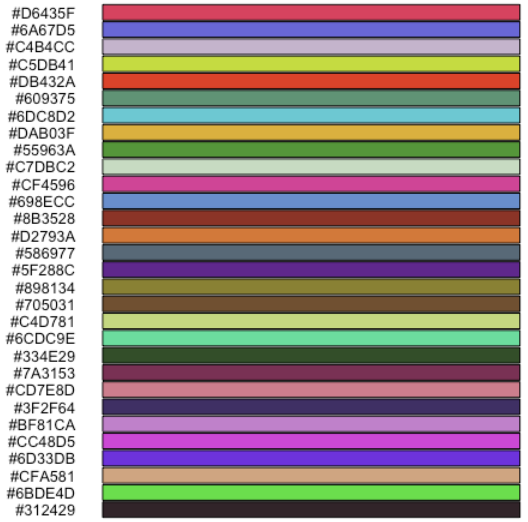
Base R: `install.packages("hues")`
renv: `renv::install("hues")`

`iwanthue(5, plot=TRUE)`

"#503F44" "#964FB8" "#97B2B7" "#94BF58" "#C6624D"

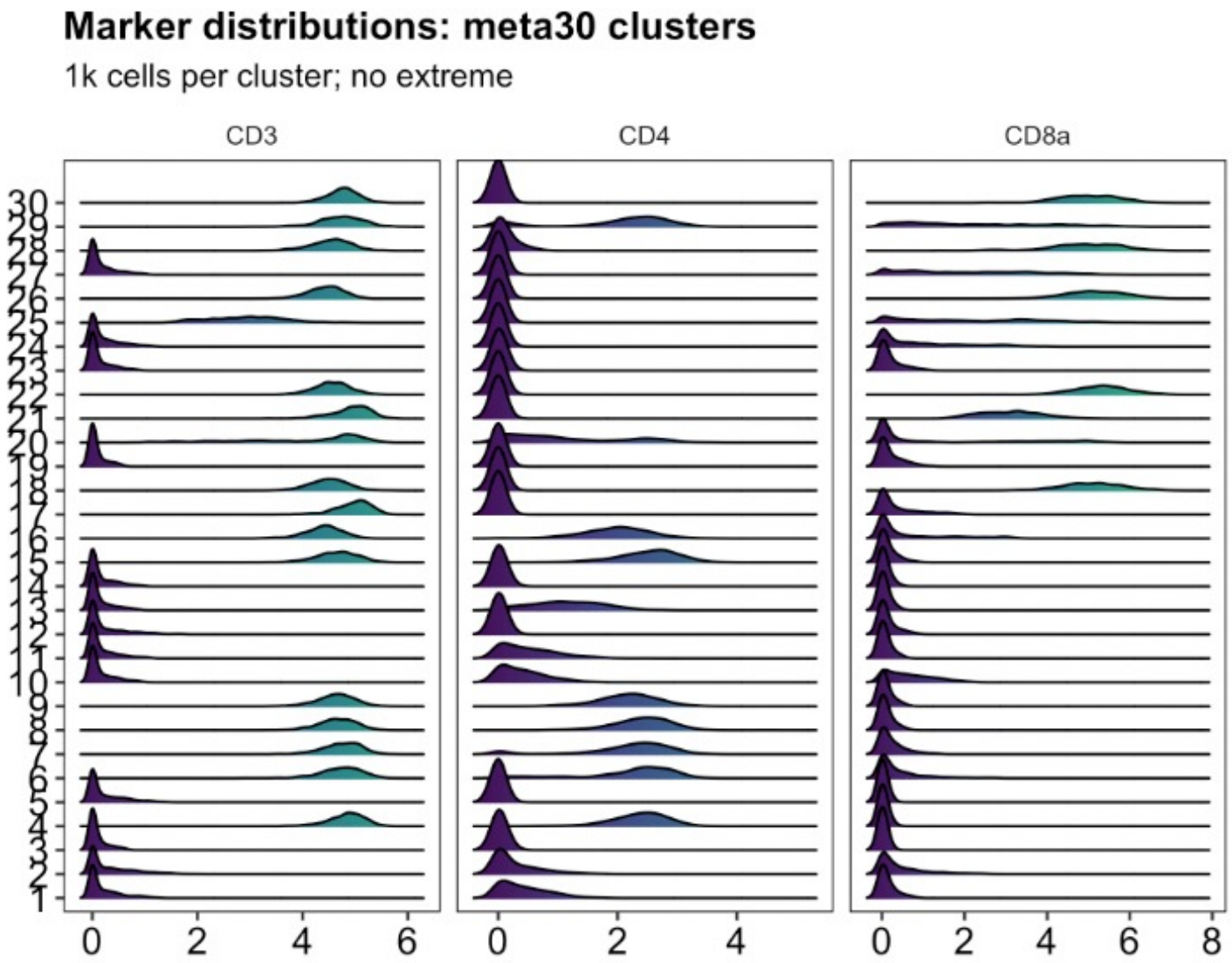
`iwanthue(30, plot=TRUE)`

"#D6435F" "#6A67D5" "#C4B4CC" "#C5DB41" "#DB432A"
"#609375" "#6DC8D2" "#DAB03F" "#55963A" "#C7DBC2"
"#CF4596" "#698ECC" "#8B3528" "#D2793A" "#586977"
"#5F288C" "#898134" "#705031" "#C4D781" "#6CDC9E"
"#334E29" "#7A3153" "#CD7E8D" "#3F2F64" "#BF81CA"
"#CC48D5" "#6D33DB" "#CFA581" "#6BDE4D" "#312429"

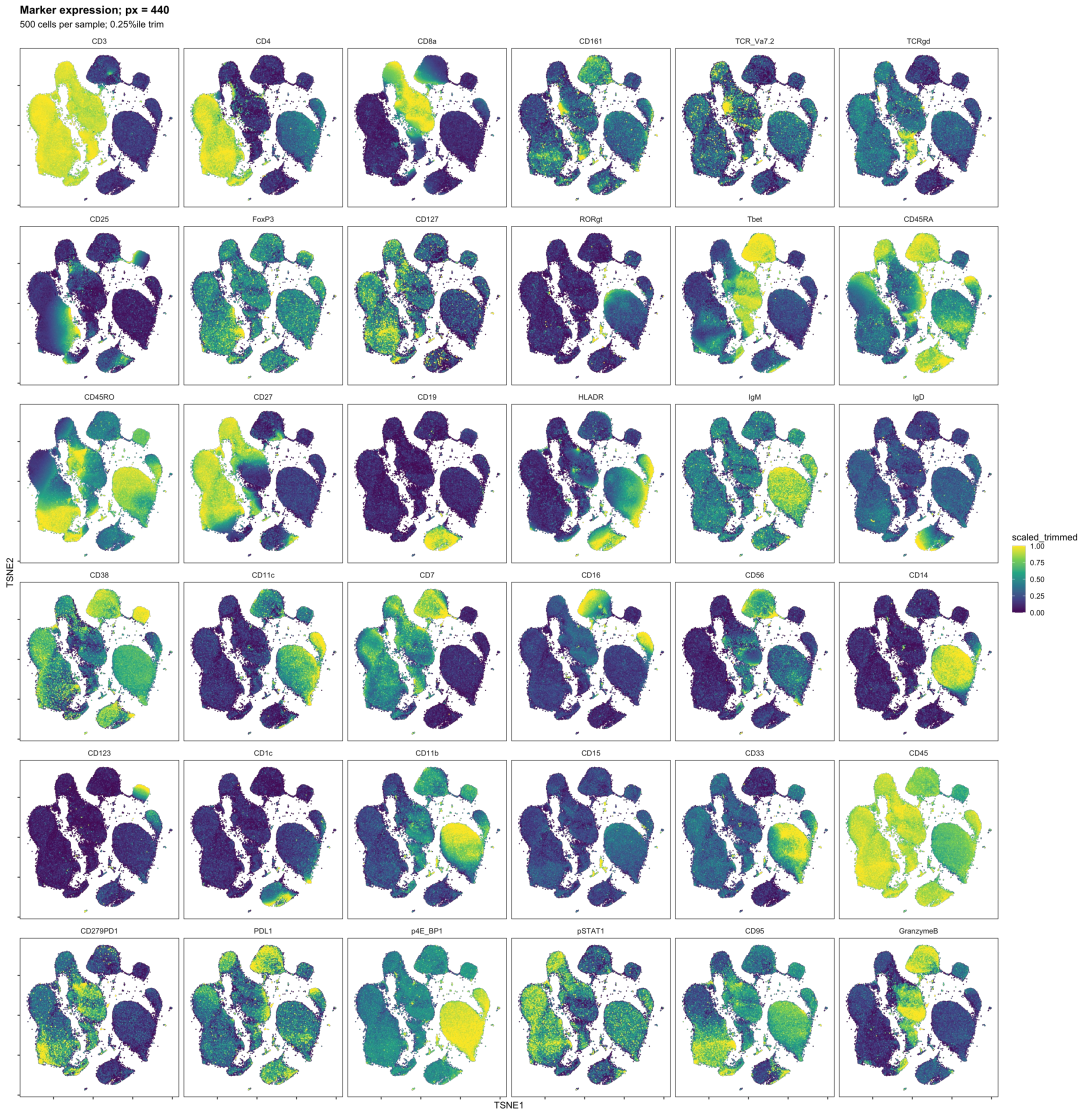


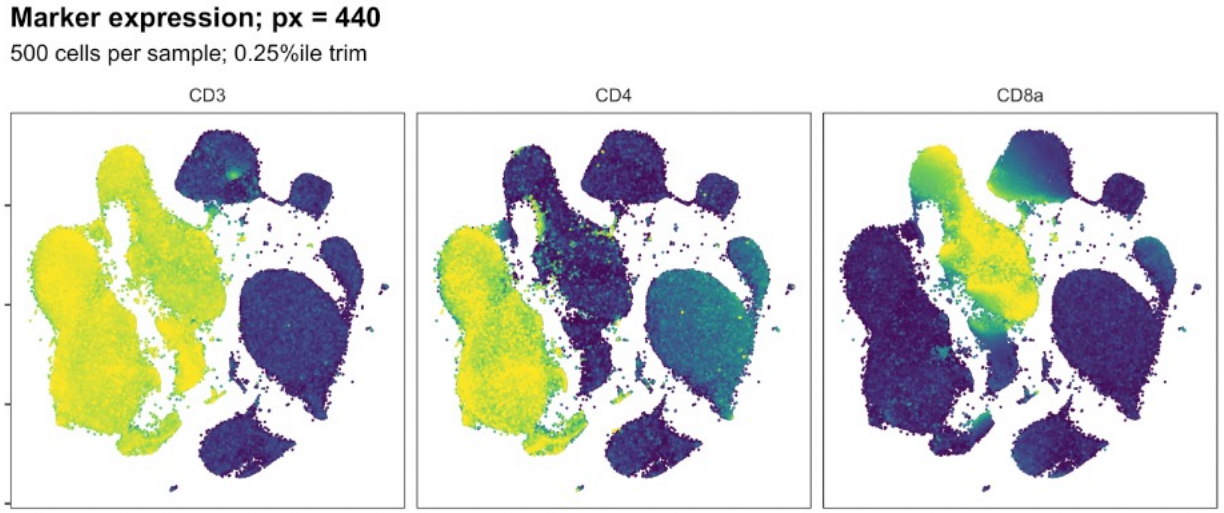
https://cran.r-project.org/web/packages/hues/index.html

More Plot Examples: Ridgeline Plots for Expression



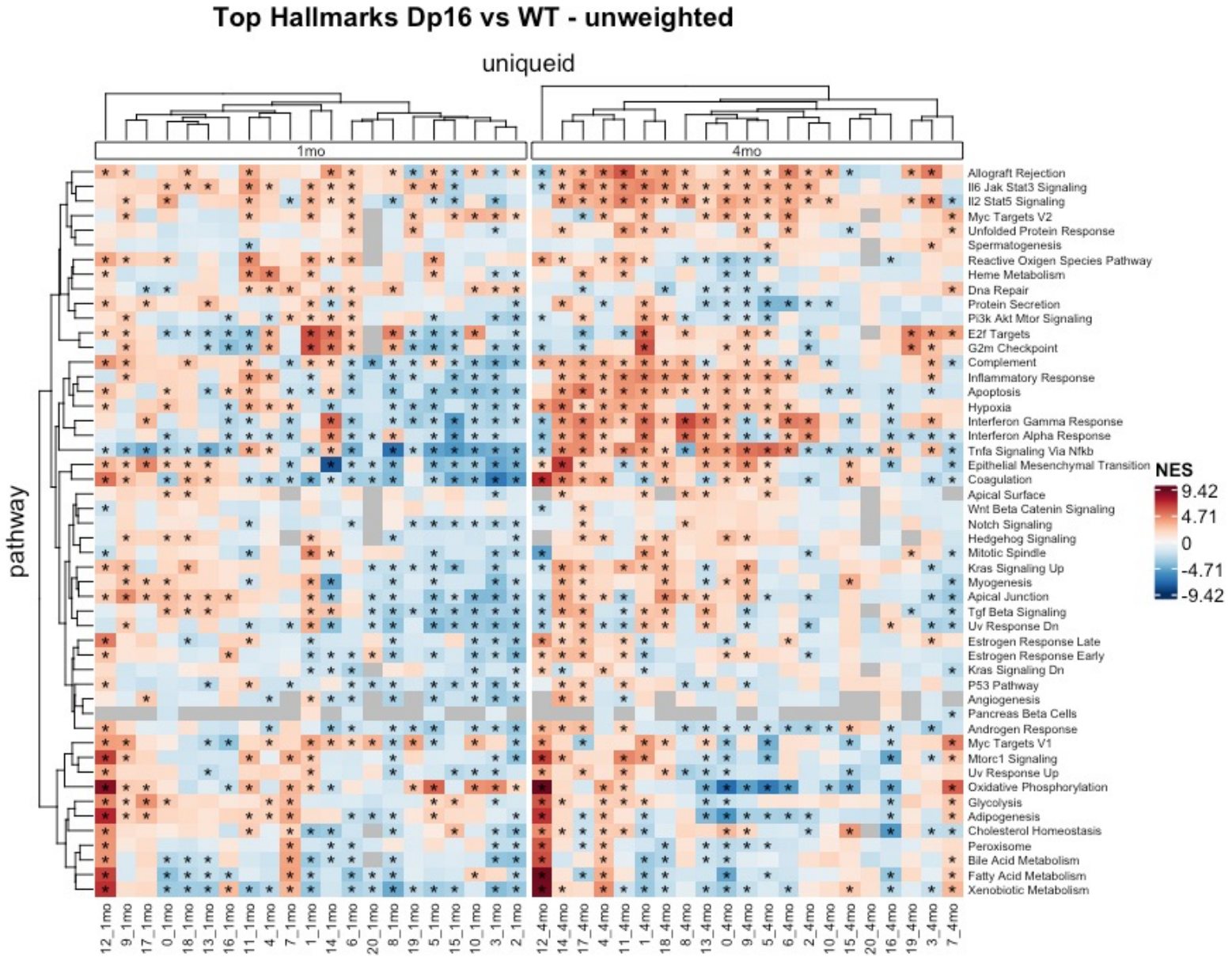
More Plot Examples: tSNE Plots with Expression





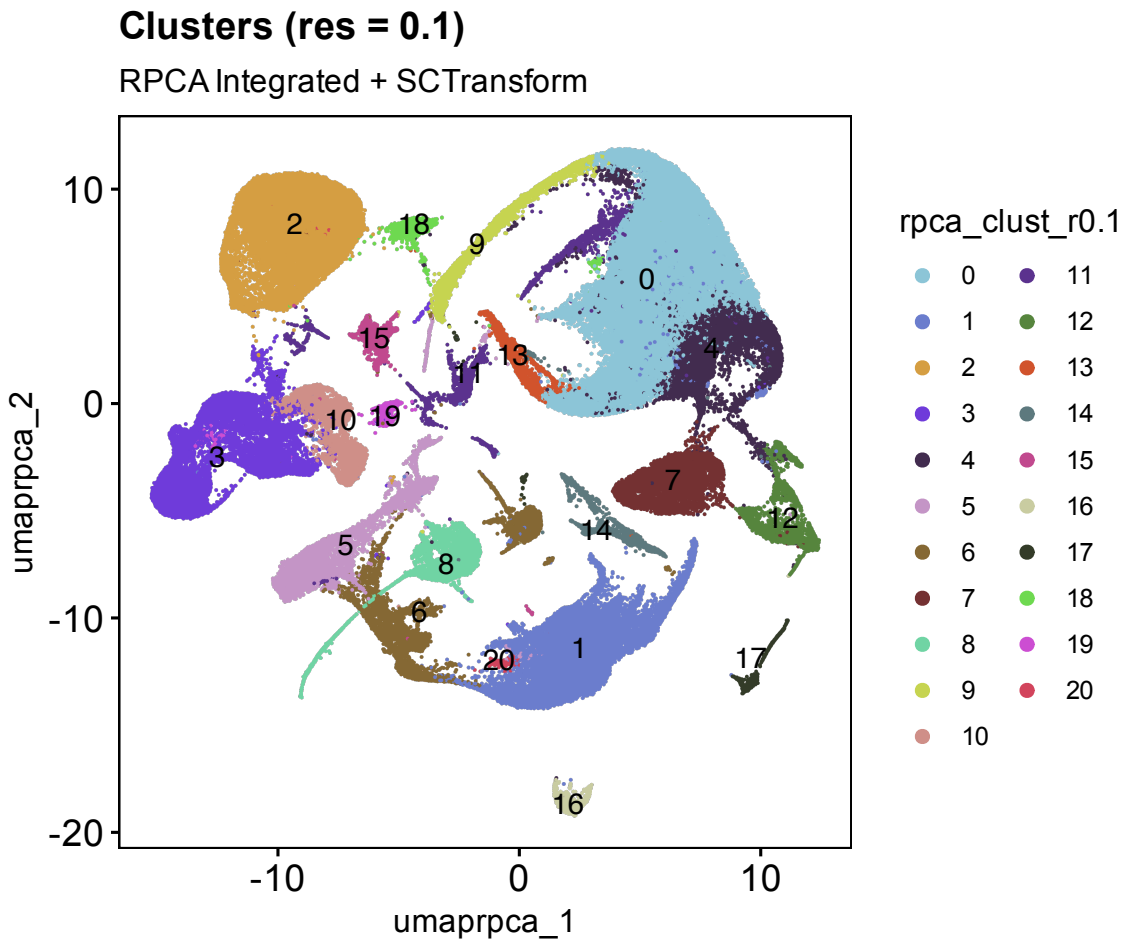
https://towardsdatascience.com/how-to-tune-hyperparameters-of-tsne-7c0596a18868
https://distill.pub/2016/misread-tsne/

More Plot Examples: Heatmap with asterisks for significance



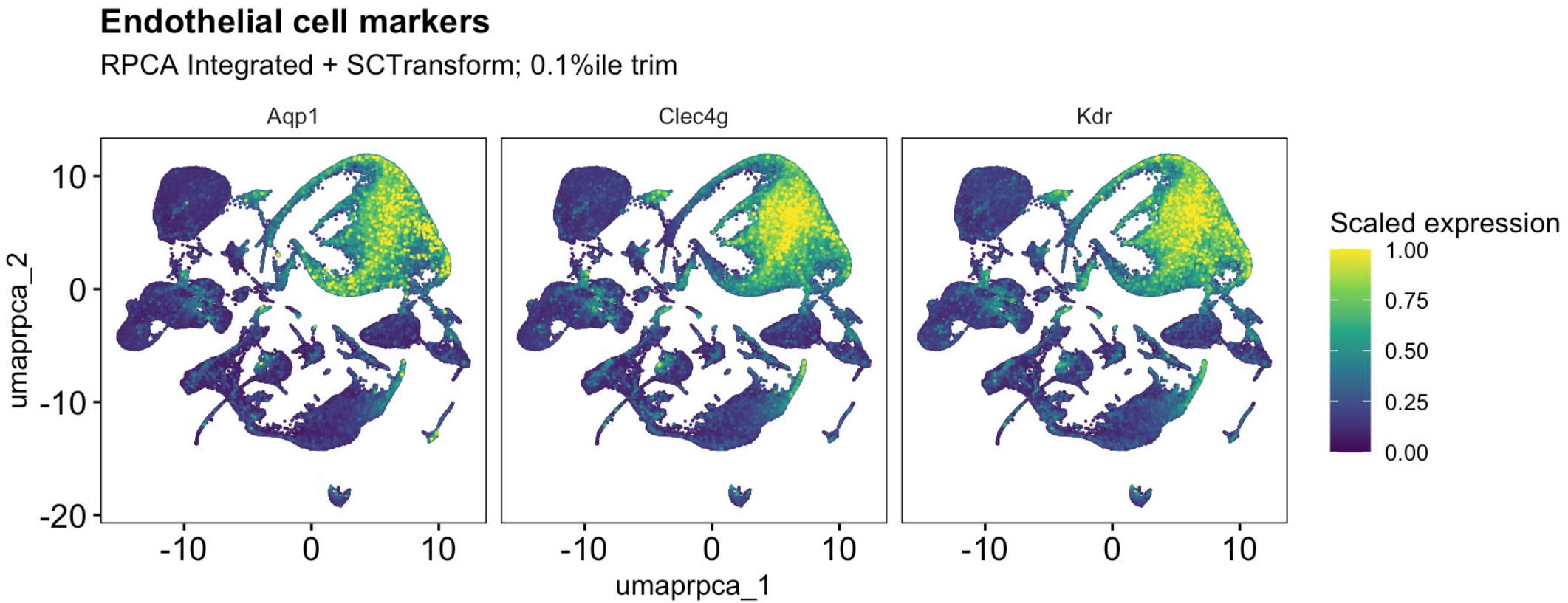
Single Cell RNA

More Plot Examples: UMAP with Clusters



https://pair-code.github.io/understanding-umap/

More Plot Examples: UMAP with Expression



More Plot Examples: UMAP with Fold-change