

ANT 6973: DATA VISUALIZATION AND EXPLORATION
COLOR, SAVING OUTPUT,
AND GEOSPATIAL DATA

TODAY'S TOPICS

- Using color effectively
- Saving your plots
- More about visualizing geospatial data
- Small Project 1

USING COLOR EFFECTIVELY

TYPES OF COLOR PALETTES

- There are three main types of color palettes:
 - Qualitative
 - Sequential
 - Diverging
- Each is appropriate for particular kinds of data.

QUALITATIVE COLOR PALETTES

- Use when color shows discrete categories that don't have *intrinsic order*.
- Limitations: unlikely to work well with more than ~12 categories (3–5 ideal)

ggplot2 default



Example: `scale_color_hue()`

QUALITATIVE COLOR PALETTES

RColorBrewer



Example: `scale_color_brewer(palette = "Set1")`

QUALITATIVE COLOR PALETTES

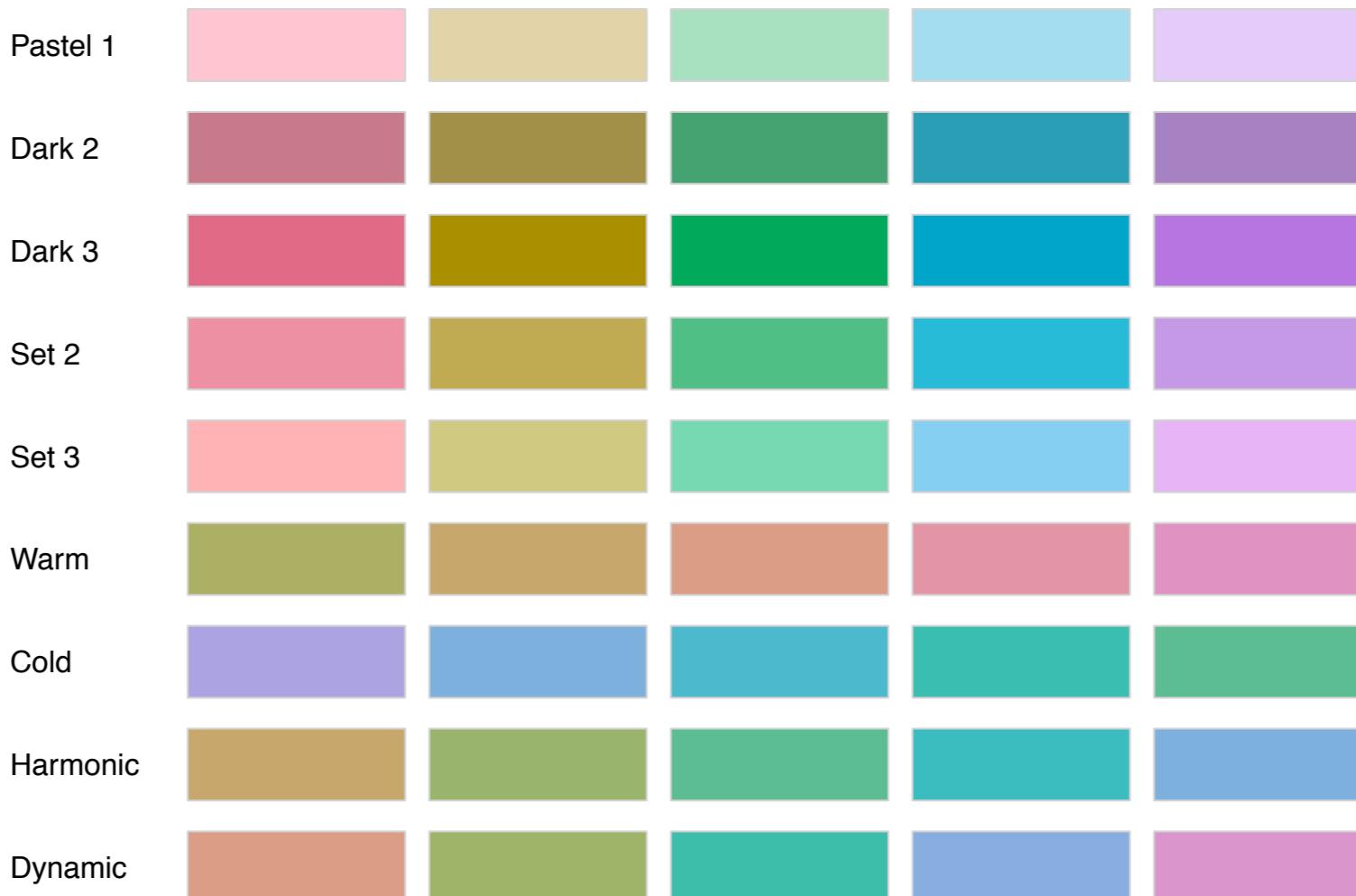
rcartocolor



Example: `scale_color_carto_d(palette = "Bold")`

QUALITATIVE COLOR PALETTES

colorspace



Example: `scale_color_discrete_qualitative(palette = "Cold")`

SEQUENTIAL COLOR PALETTES

- Use when color represents data values that have a natural order.
- Can be based on a single hue (e.g., from dark blue to light blue) or on multiple hues (e.g., from dark red to light yellow)

SEQUENTIAL COLOR PALETTES

- Two usage cases:

SEQUENTIAL COLOR PALETTES

- Two usage cases:
 - When you have ordered categories

ggplot2 default



Example: `scale_color_ordinal()`

SEQUENTIAL COLOR PALETTES

- Two usage cases:
 - When you have ordered categories
 - When you have continuously varying values

ggplot2 default

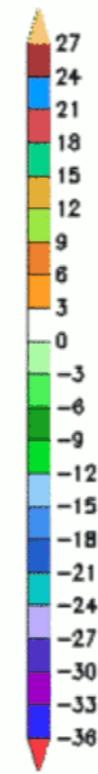
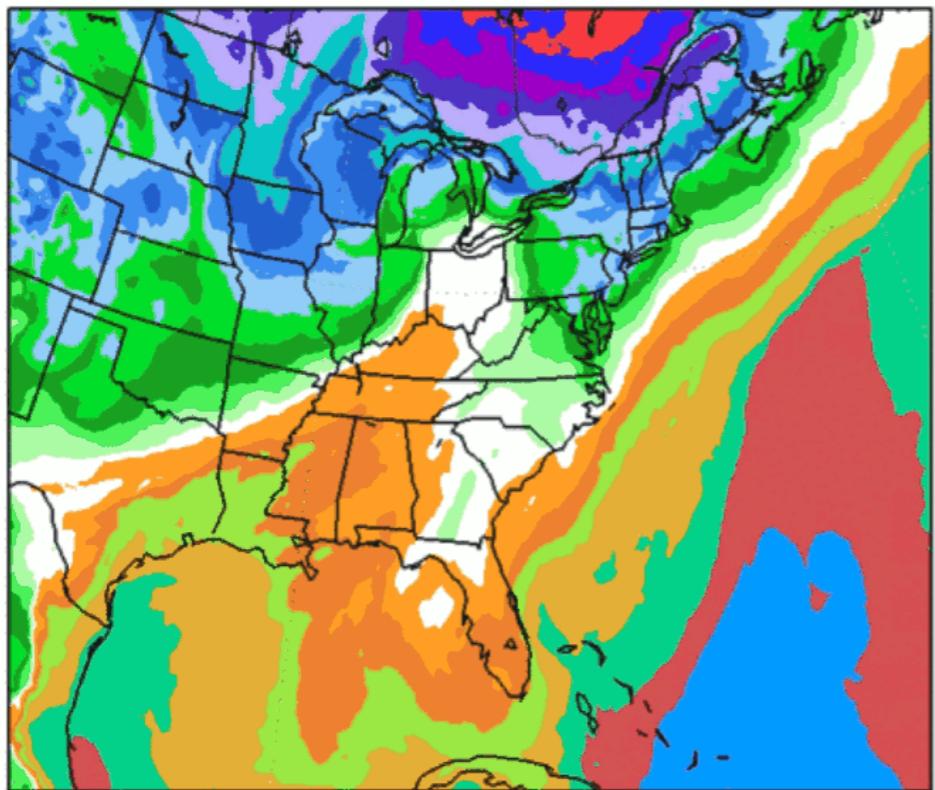


Example: `scale_color_continuous()`

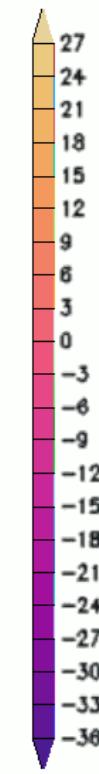
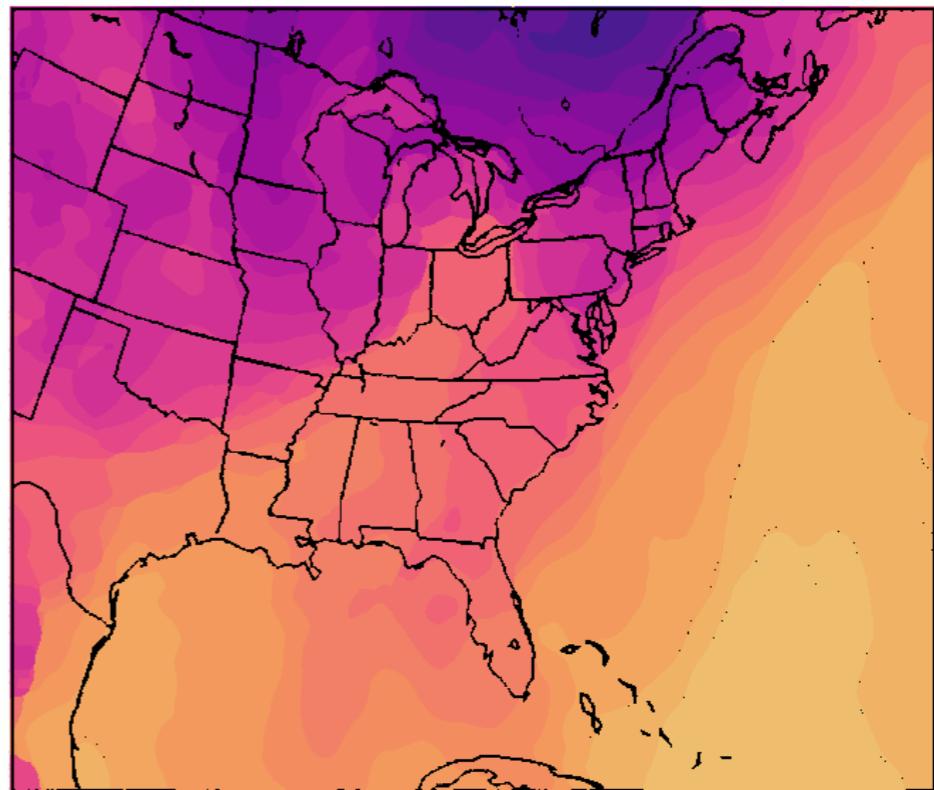
SEQUENTIAL COLOR PALETTES

- A good sequential color scale accurately conveys:
 - Which values are larger or smaller
 - How distant two values are from each other
- This means that it should be *perceptually uniform* across its entire range.

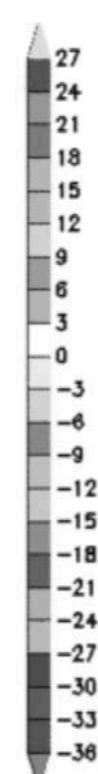
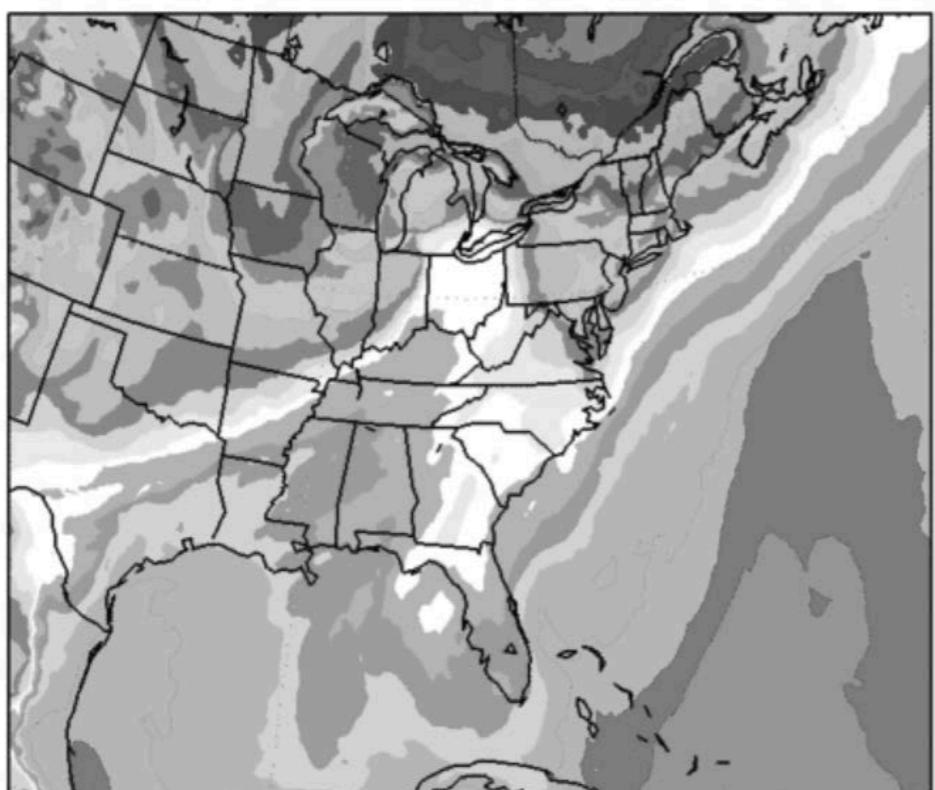
2-M DEW PT TEMP FV3GFS 00H FCST VALID 00Z 29 JAN 2019



2-M DEW PT TEMP FV3GFS 00H FCST VALID 00Z 29 JAN 2019

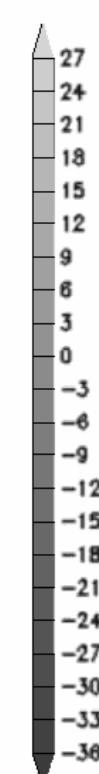
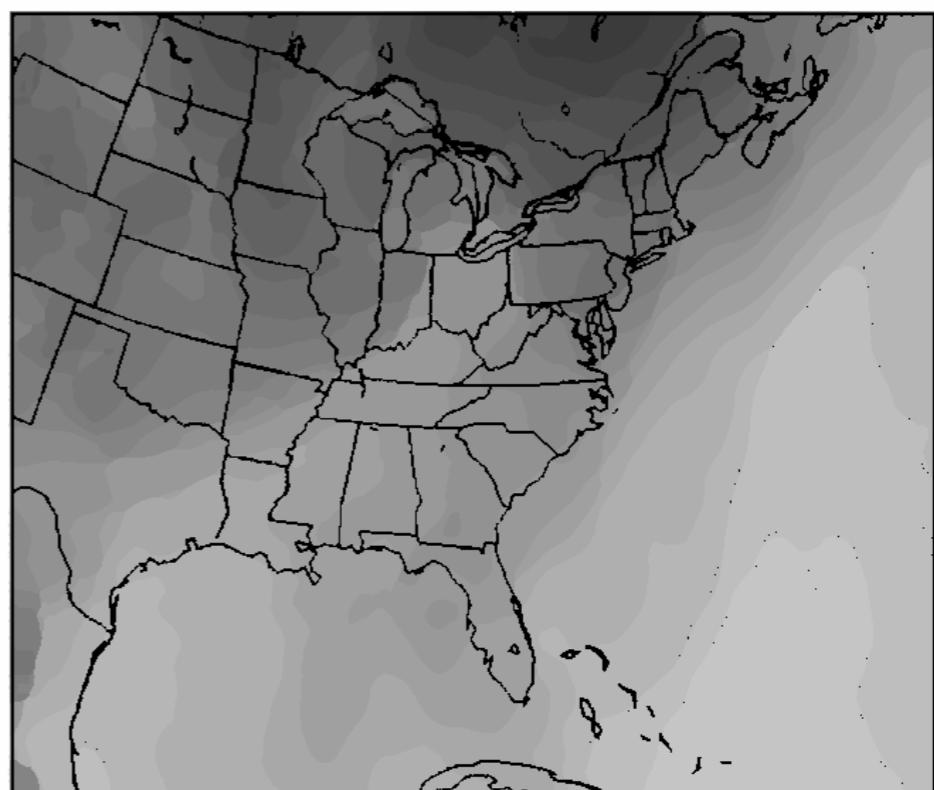


2-M DEW PT TEMP FV3GFS 00H FCST VALID 00Z 29 JAN 2019



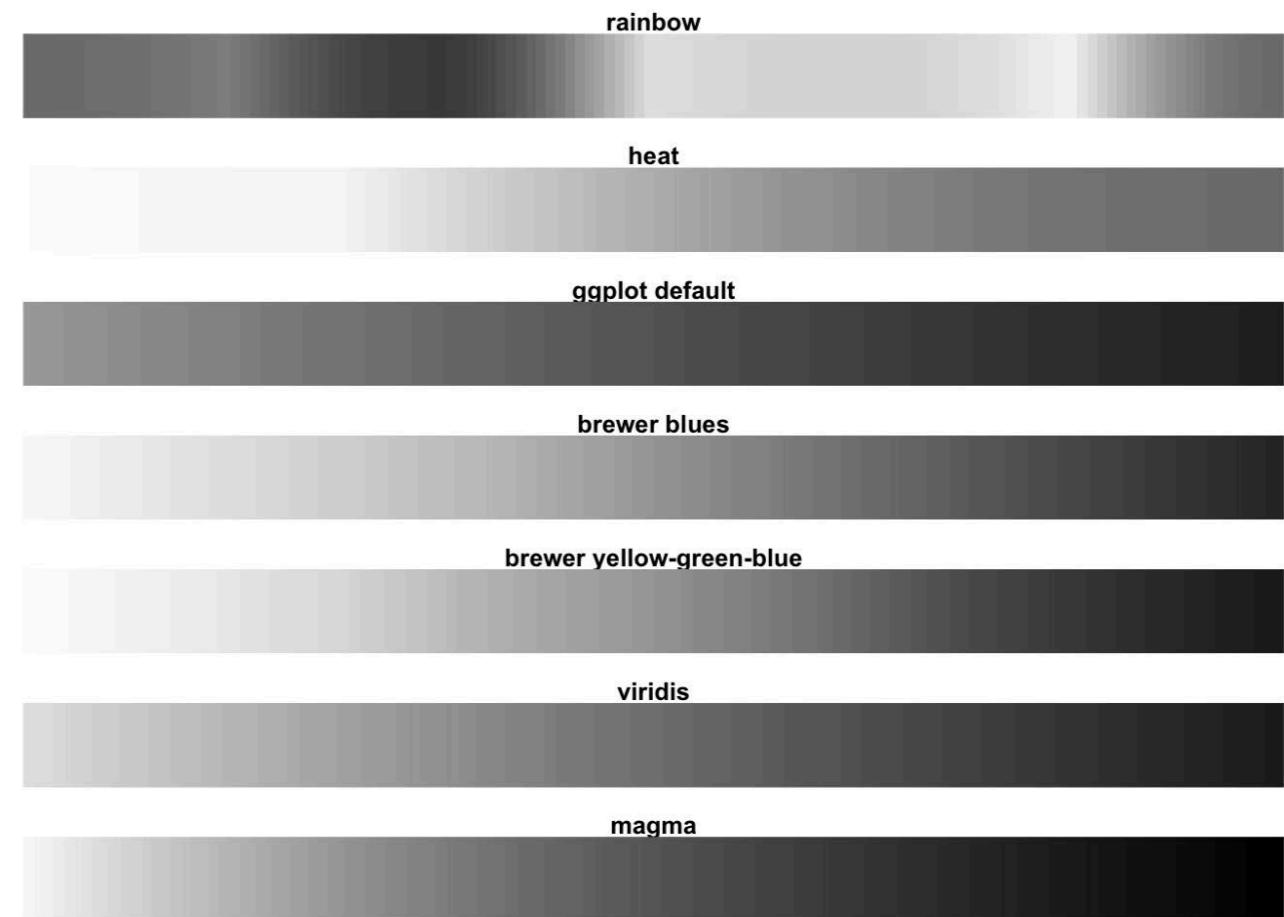
~ 1

2-M DEW PT TEMP FV3GFS 00H FCST VALID 00Z 29 JAN 2019



SEQUENTIAL COLOR PALETTES

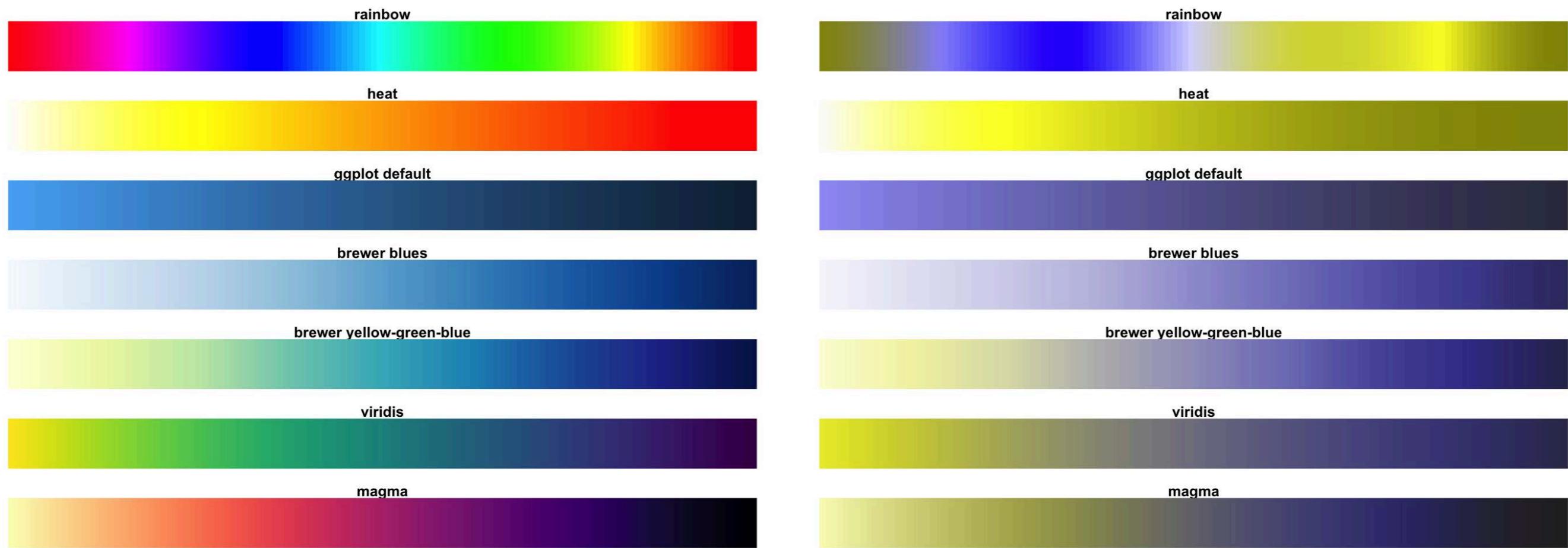
Desaturated



When figures are printed or rendered in grayscale

SEQUENTIAL COLOR PALETTES

Red-green color blindness



~8% of men and 0.5% of women

SEQUENTIAL COLOR PALETTES

Blue-yellow color blindness



Very rare, ~1 in 40,000

SEQUENTIAL COLOR PALETTES

viridis



magma



plasma



inferno



cividis



Ordered categories: `scale_color_viridis_d(option = "magma")`
Continuous values: `scale_color_viridis_c(option = "magma")`

SEQUENTIAL COLOR PALETTES

RColorBrewer

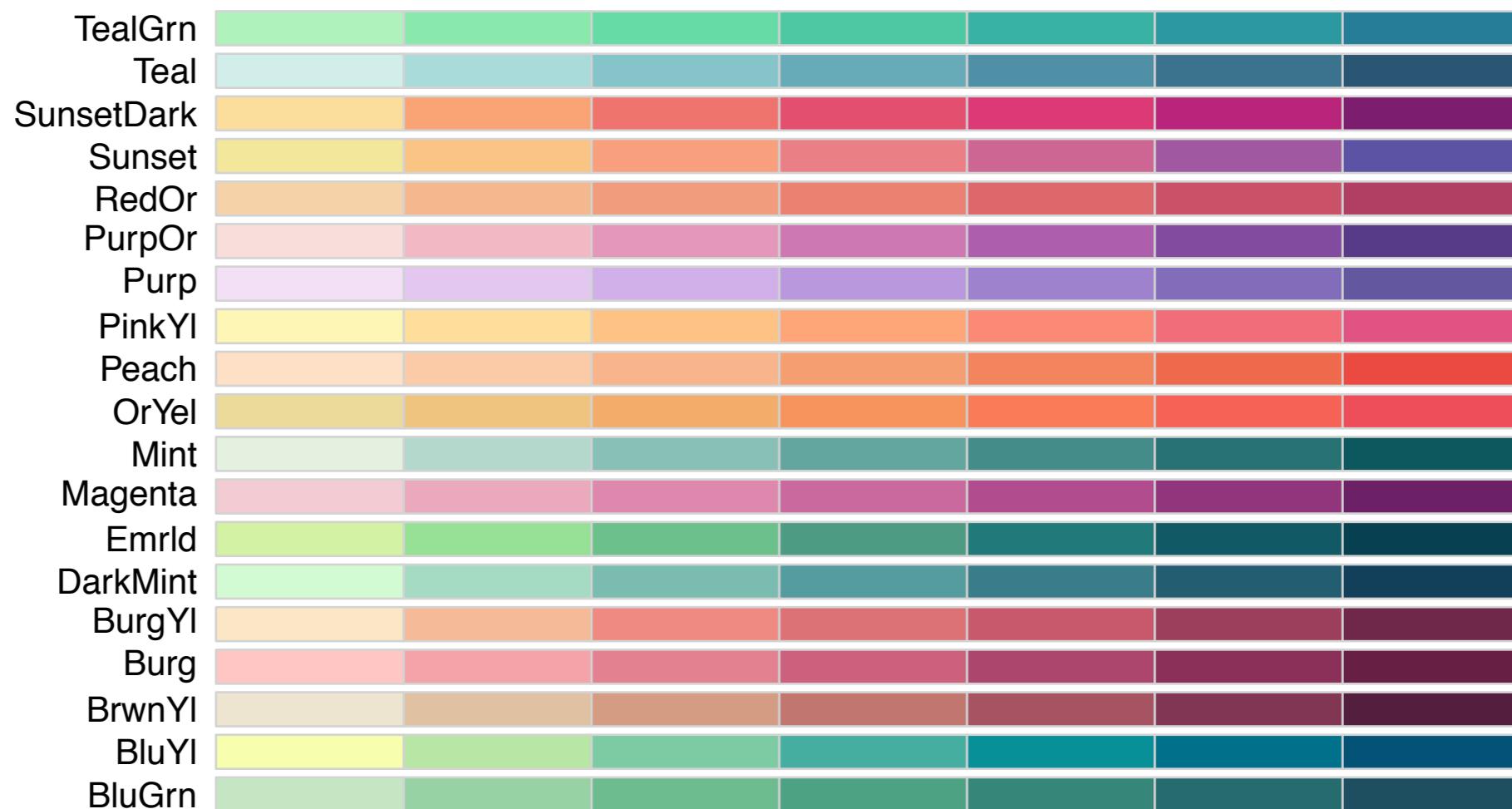


Ordered categories: `scale_color_brewer(palette = "RdPu")`

Continuous values: `scale_color_distiller(palette = "RdPu")`

SEQUENTIAL COLOR PALETTES

rcartocolor



Ordered categories: `scale_color_carto_d(palette = "Mint")`
Continuous values: `scale_color_carto_c(palette = "Mint")`

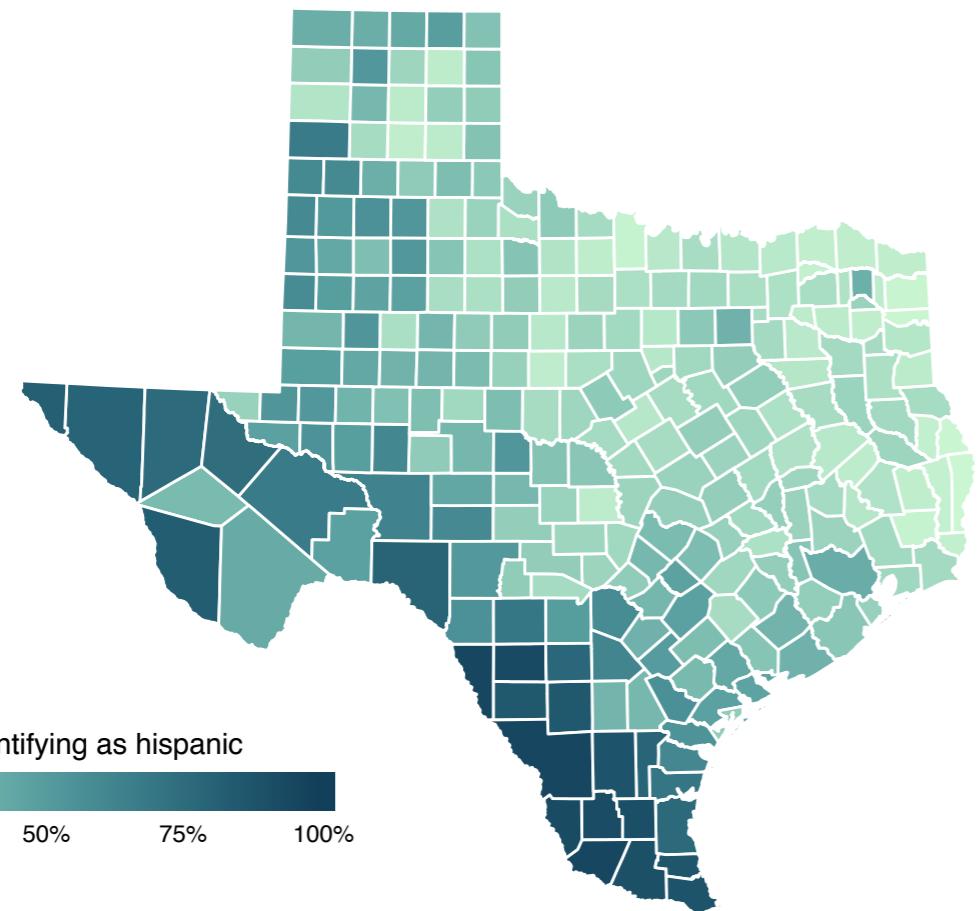
SEQUENTIAL COLOR PALETTES

colorspace

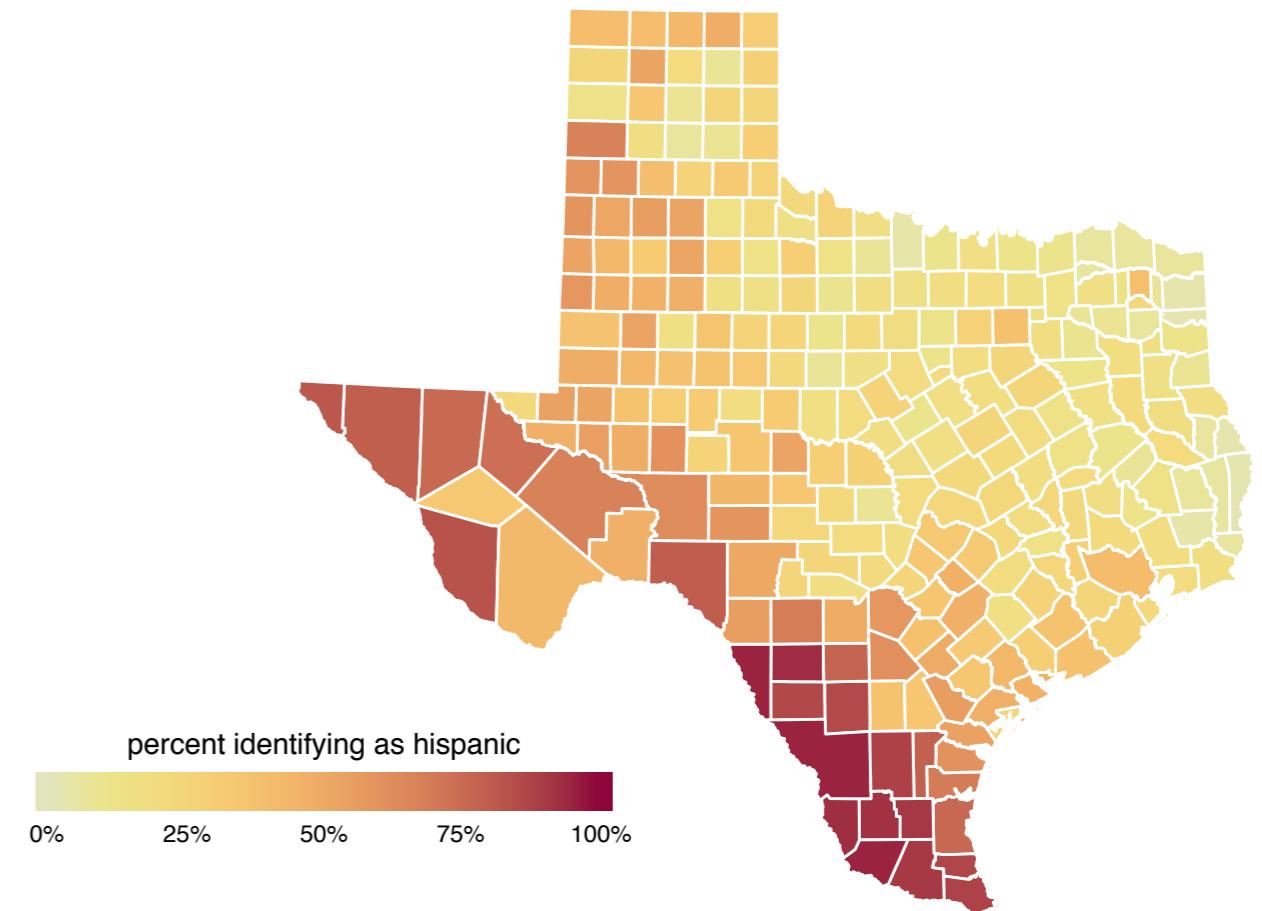


Ordered categories: `scale_color_discrete_sequential(palette = "Heat")`
 Continuous values: `scale_color_continuous_sequential(palette = "Heat")`

SEQUENTIAL COLOR PALETTES



DarkMint from `rcartocolor`

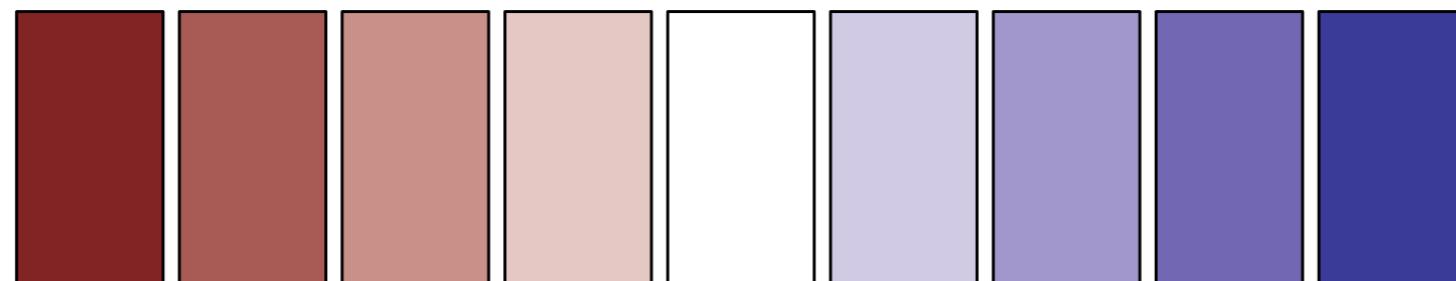


Heat from `colorspace`

DIVERGING COLOR PALETTES

- Use when data values diverge in two directions from a neutral midpoint (e.g., positive and negative numbers).
- Shows both how much and in which direction each value deviates from the midpoint.

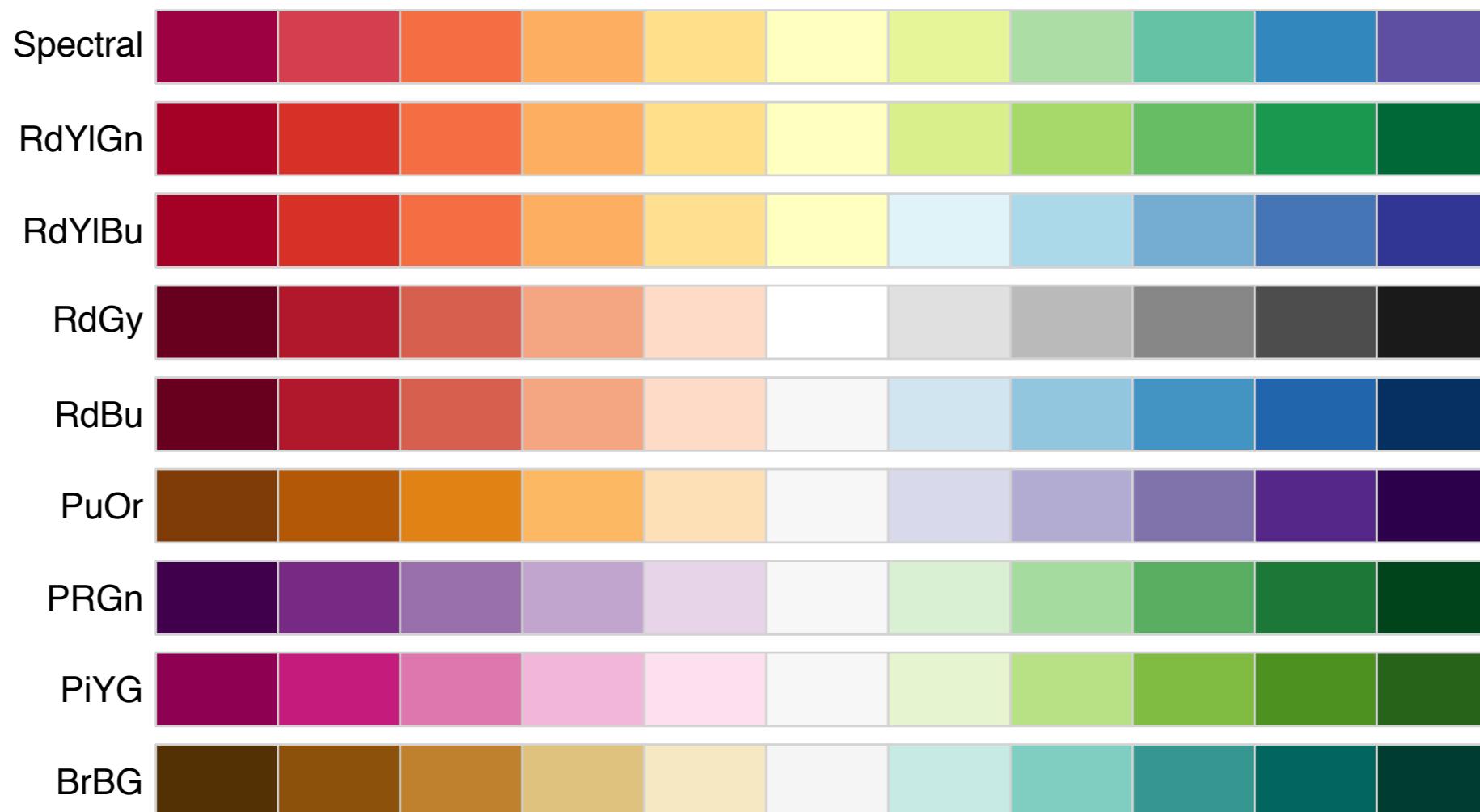
ggplot2 default



Example: `scale_color_gradient2()`

DIVERGING COLOR PALETTES

RColorBrewer

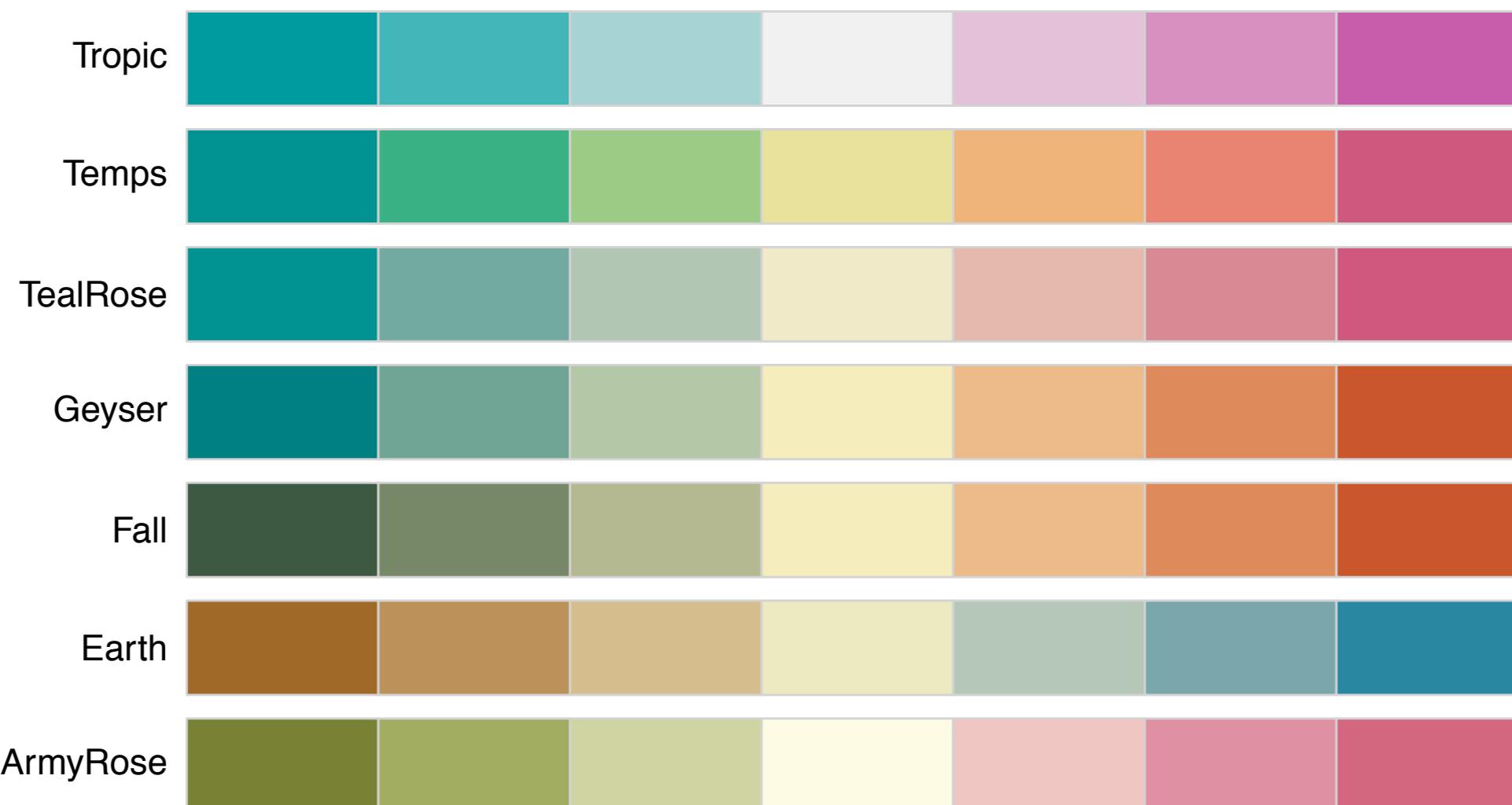


Ordered categories: `scale_color_brewer(palette = "RdYlBu")`

Continuous values: `scale_color_distiller(palette = "RdYlBu")`

DIVERGING COLOR PALETTES

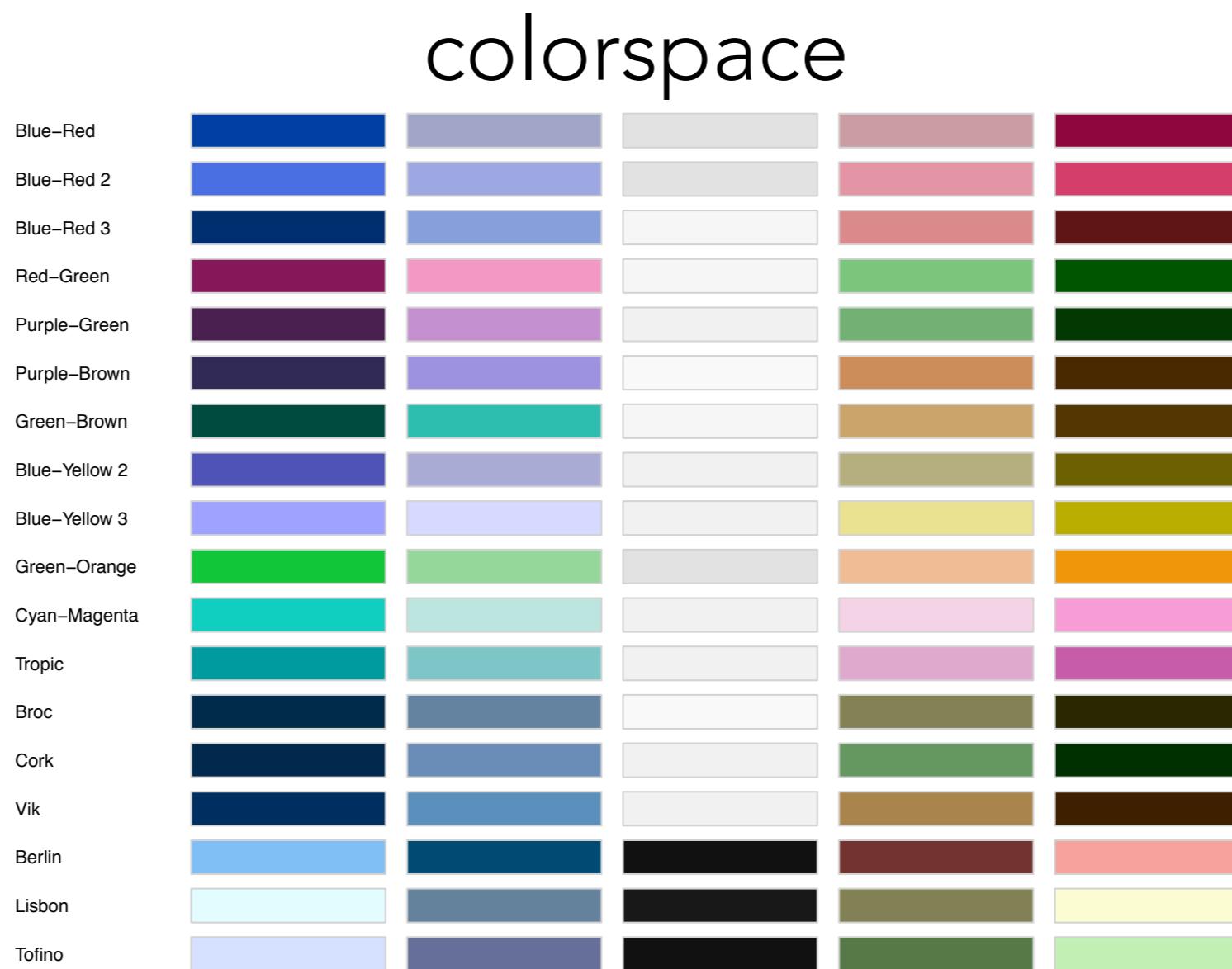
rcartocolor



Ordered categories: `scale_color_carto_d(palette = "Tropic")`

Continuous values: `scale_color_carto_c(palette = "Tropic")`

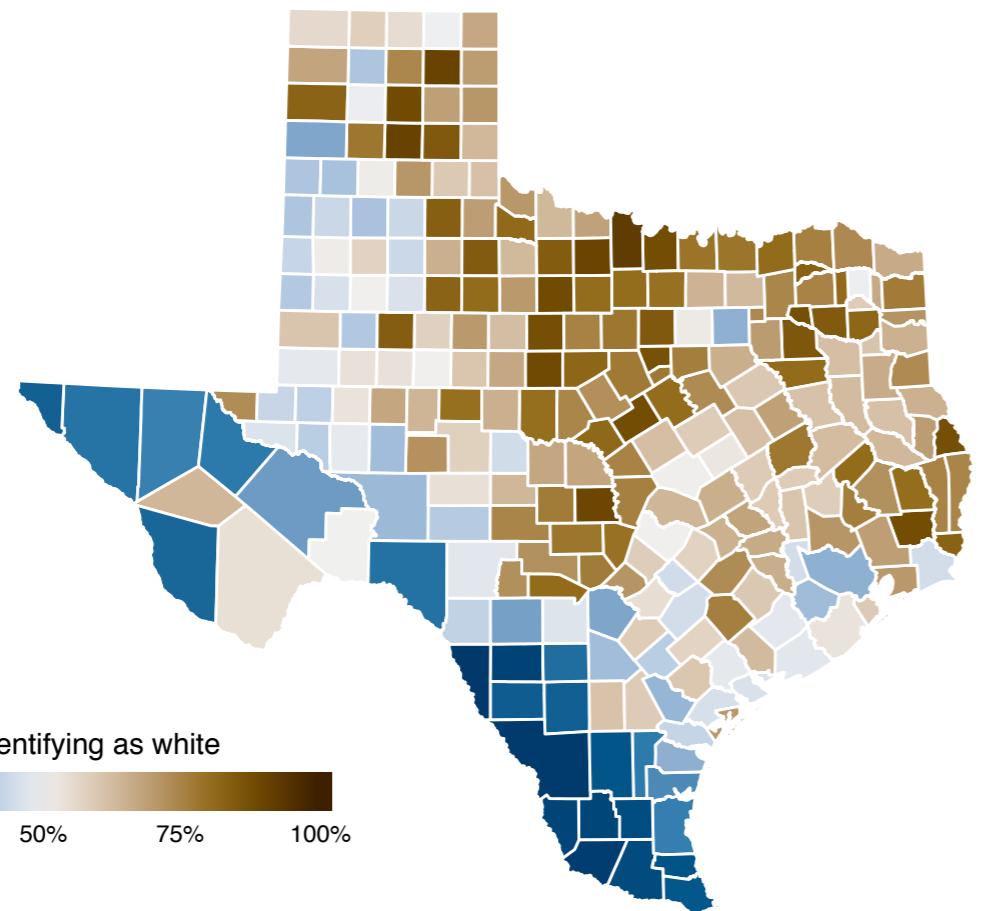
DIVERGING COLOR PALETTES



Ordered categories: `scale_color_discrete_diverging(palette = "Vik")`
Continuous values: `scale_color_continuous_diverging(palette = "Vik")`

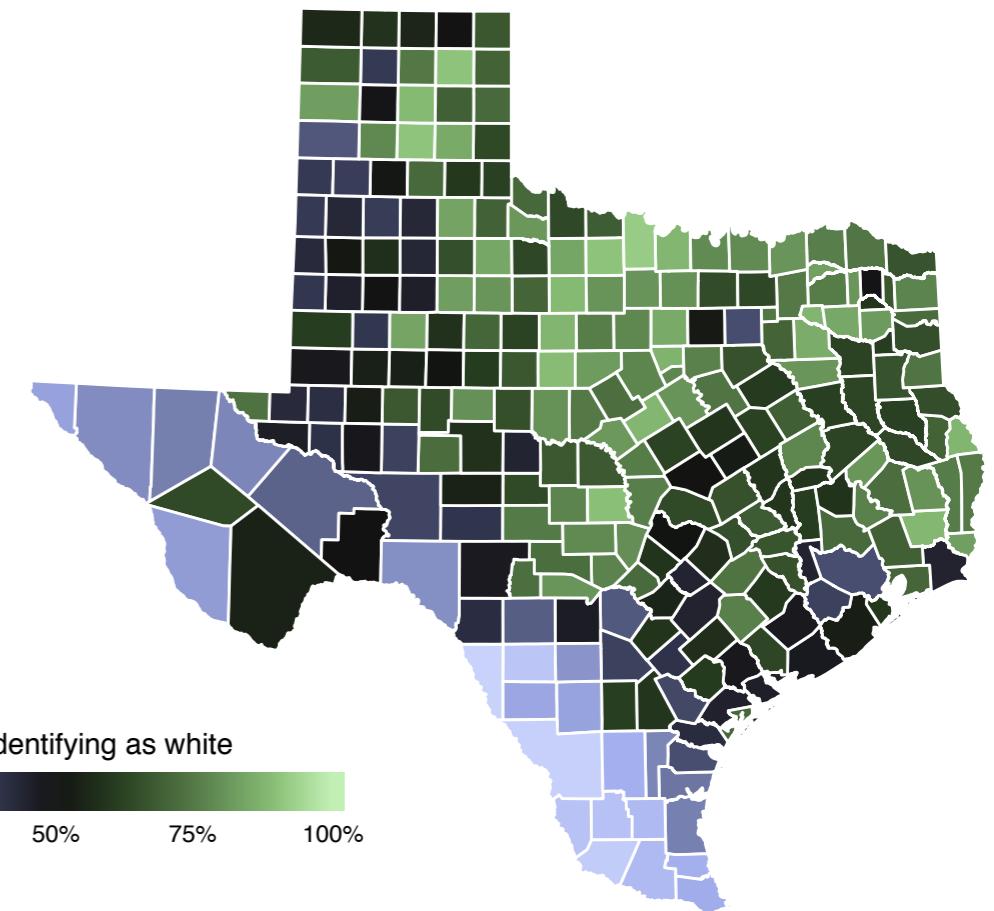
DIVERGING COLOR PALETTES

Light midpoint



Vik from colorspace

Dark midpoint



Tofino from colorspace

COMMON PITFALLS OF COLOR USE

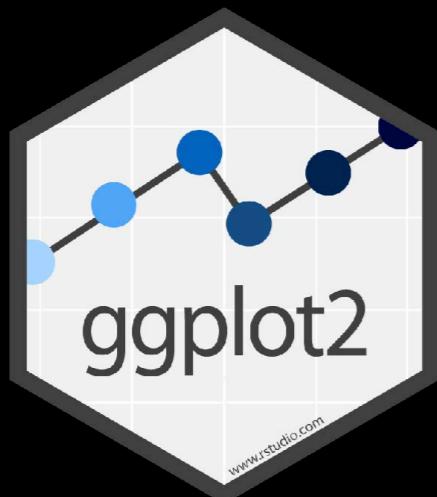
- Too many colors for discrete categories
- Using coloring without clear purpose
- Color gradients with perceptual discontinuities (e.g., rainbows)
- Not considering conversion to grayscale and color-vision deficiency

SAVING YOUR PLOTS

TWO WAYS I RECOMMEND



Let knitr create the plots for you



`ggsave()` function

KNITR



- This should be familiar to you already.
- Figure height and width can be specified in chunk options.

fig.height

fig.width

```
179
180 ````{r, fig.height = 6, fig.width = 10}
181 ggplot(filter(babynames, name == "Lionel" & sex == "M"),
182     aes(x = year, y = prop)) +
183     geom_ribbon(aes(ymax = prop, ymin = 0), alpha = 0.5) +
184     geom_line() +
185     annotate(geom = "rect", xmin = 1950, xmax = 1960, ymin = 0, ymax = Inf, fill = "red", alpha = 0.25) +
186     annotate(geom = "text", x = 1949, y = 3e-4, label = "Some text\nMore text", color = "red", hjust = 1) +
187     labs(title = "Popularity of the name Lionel")
188
189
```

KNITR



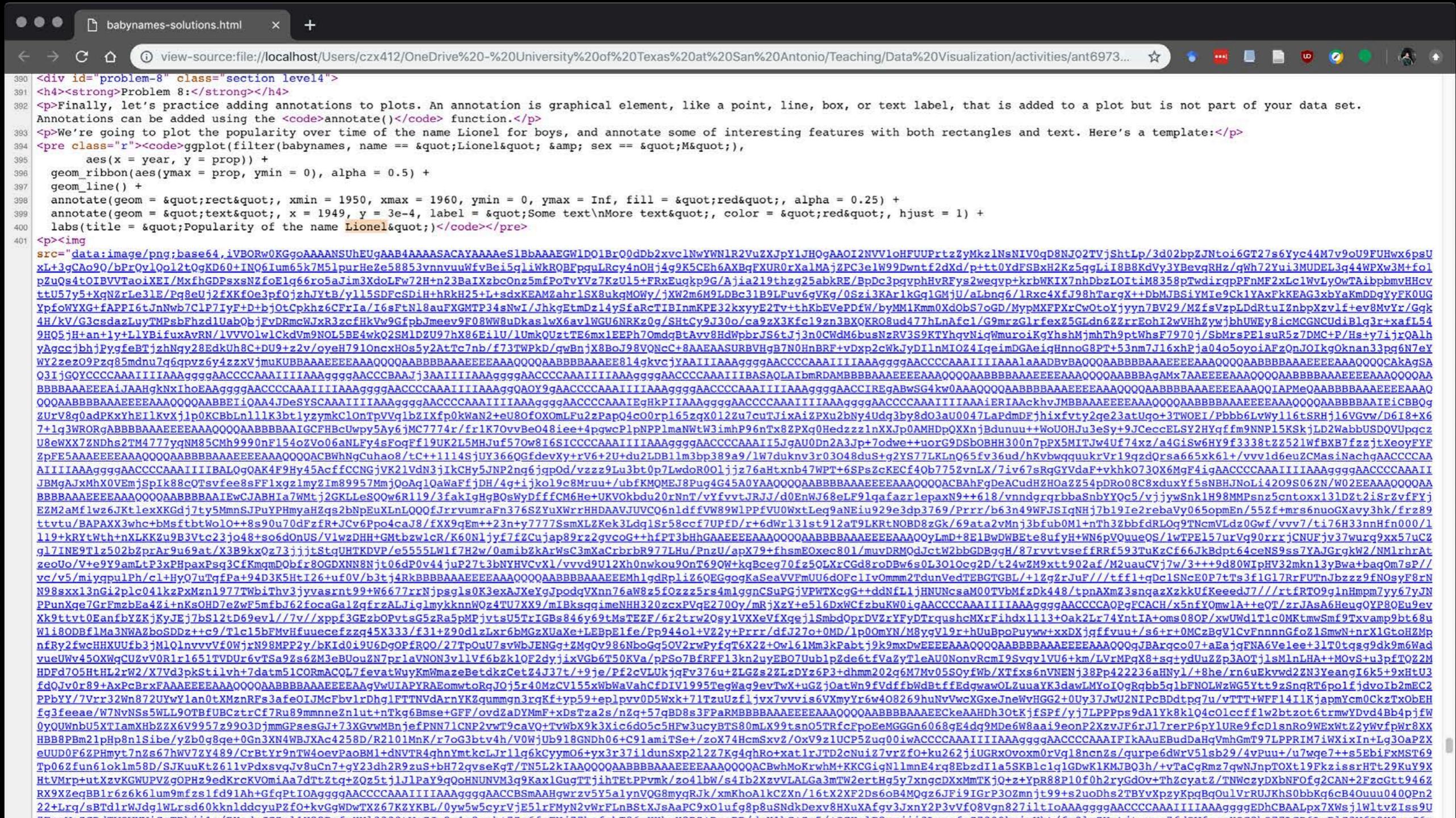
- Where are the image files?

```
babynames-solutions.html
```

Finally, let's practice adding annotations to plots. An annotation is graphical element, like a point, line, box, or text label, that is added to a plot but is not part of your data set. Annotations can be added using the `annotate()` function.

We're going to plot the popularity over time of the name Lionel for boys, and annotate some of interesting features with both rectangles and text. Here's a template:

```
ggplot(filter(babynames, name == "Lionel" & sex == "M"),  
       aes(x = year, y = prop)) +  
  geom_ribbon(aes(ymax = prop, ymin = 0), alpha = 0.5) +  
  geom_line() +  
  annotate(geom = "rect", xmin = 1950, xmax = 1960, fill = "red", alpha = 0.25) +  
  annotate(geom = "text", x = 1949, y = 3e-4, label = "Some text\nMore text", color = "red", hjust = 1) +  
  labs(title = "Popularity of the name Lionel")
```



KNITR



- By default, knitr creates standalone HTML documents by encoding the image and embedding it inline with the document.
- That means that there is no separate image file sitting somewhere on your computer.
- This makes the HTML file portable!
- But if you need to save one of the images as a file, you can do so easily from your web browser.



In-class Activity: Changing Bar

Problem 8:

Finally, let's practice adding annotations to plots. An annotation is graphical element, like a point, line, box, or text label, that is added to a plot but is not part of your data set. Annotations can be added using the `annotate()` function.

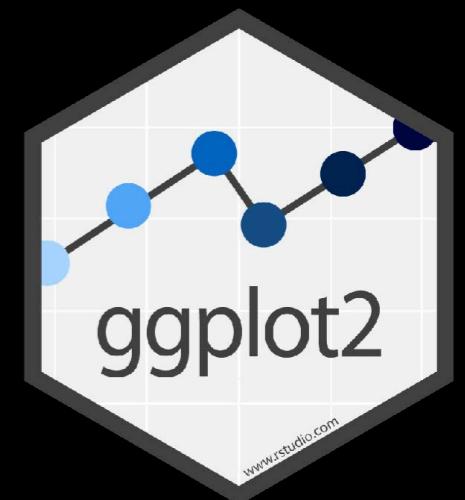
We're going to plot the popularity over time of the name Lionel for boys, and annotate some of interesting features with both rectangles and text. Here's a template:

```
ggplot(filter(babynames, name == "Lionel" & sex == "M"),  
       aes(x = year, y = prop)) +  
  geom_ribbon(aes(ymax = prop, ymin = 0), alpha = 0.5) +  
  geom_line() +  
  annotate(geom = "rect", xmin = 1950, xmax = 1960, ymin = 0, ymax = Inf, fill = "red", alpha = 0.25) +  
  annotate(geom = "text", x = 1949, y = 3e-04, label = "Some text\nMore text", color = "red", hjust = 1) +  
  labs(title = "Popularity of the name Lionel")
```

Popularity of the name Lionel

The chart shows the popularity of the name Lionel from 1880 to 2020. The x-axis is labeled "year" and ranges from 1880 to 2000. The y-axis is labeled "prop" and ranges from 0e+00 to 3e-04. The data is represented by a dark gray filled area and a black line. A red rectangle is drawn from the year 1950 to 1960 and from the y-axis value 0 to Inf. Red text "Some text" and "More text" is placed within this rectangle. A context menu is open over the chart, listing options: Open Image in New Tab, Save Image As..., Copy Image, Copy Image Address, Search Google for Image, Block element, LastPass: Free Password Manager, Save to Zotero, and Inspect.

GGSAVE



- Save a plot as an image file

```
ggsave(<filename>, <plot>, ...)
```

function

Name of file
you want to
create

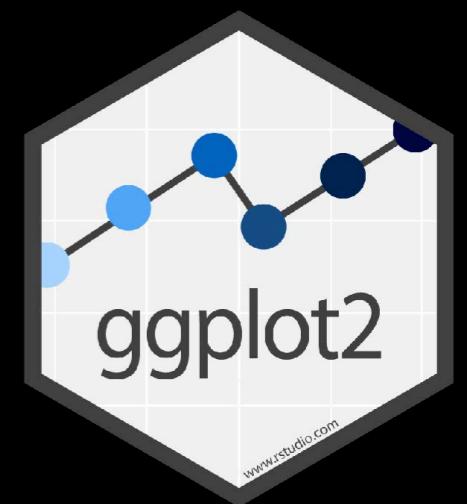
Name of
plot object
in R (if any)

Options including
width, height, and units

You provide the path and
file extension (e.g.,
".png" or ".pdf")

If nothing is specified
here, it'll use the last
plot that you created!

GGSAVE



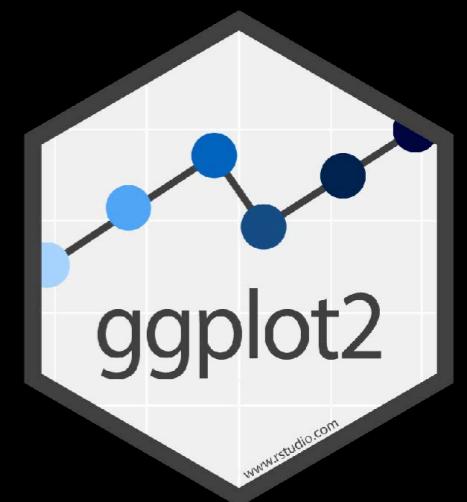
- Example

```
ggplot(starwars, aes(x = eye_color)) +  
  geom_bar()  
  
ggsave("starwars-eye-color.pdf", width = 10, height = 5)
```

No path provided, so
file will be saved in
working directory

Function knows to
create a PDF file

GGSAVE

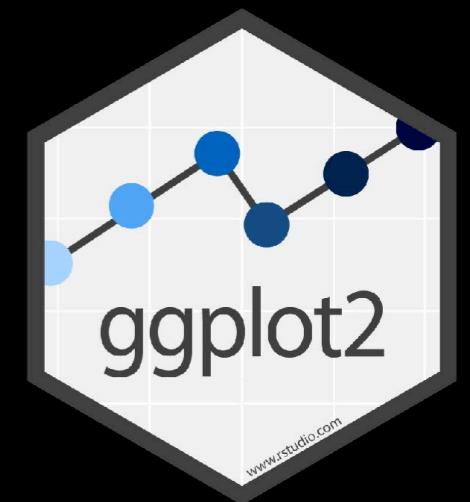


- Example

```
ggplot(starwars, aes(x = eye_color)) +  
  geom_bar()  
  
ggsave("07_viz-spatial/starwars-eye-color.pdf",  
       width = 10, height = 5)
```

Provide a path to save file in
subfolder of project home
(note: folder must exist)

GGSAVE



- Example

```
ggplot(starwars, aes(x = eye_color)) +  
  geom_bar()  
  
ggsave("~/Documents/Teaching/DataViz/activities/07_viz-  
spatial/starwars-eye-color.pdf", width = 10, height = 5)
```

An absolute path will work but it is usually a sign of potential problems. It suggests you're not using a project-based workflow, and it will not work on other computers.

HOW DO I GET THE SIZE RIGHT?

- You can eventually develop an intuition about sizing image files...
- ... but it still involves some trial and error.
- In R Studio, you can experiment using Export > Save as PDF... in the Plots pane.

HOW DO I GET THE SIZE RIGHT?

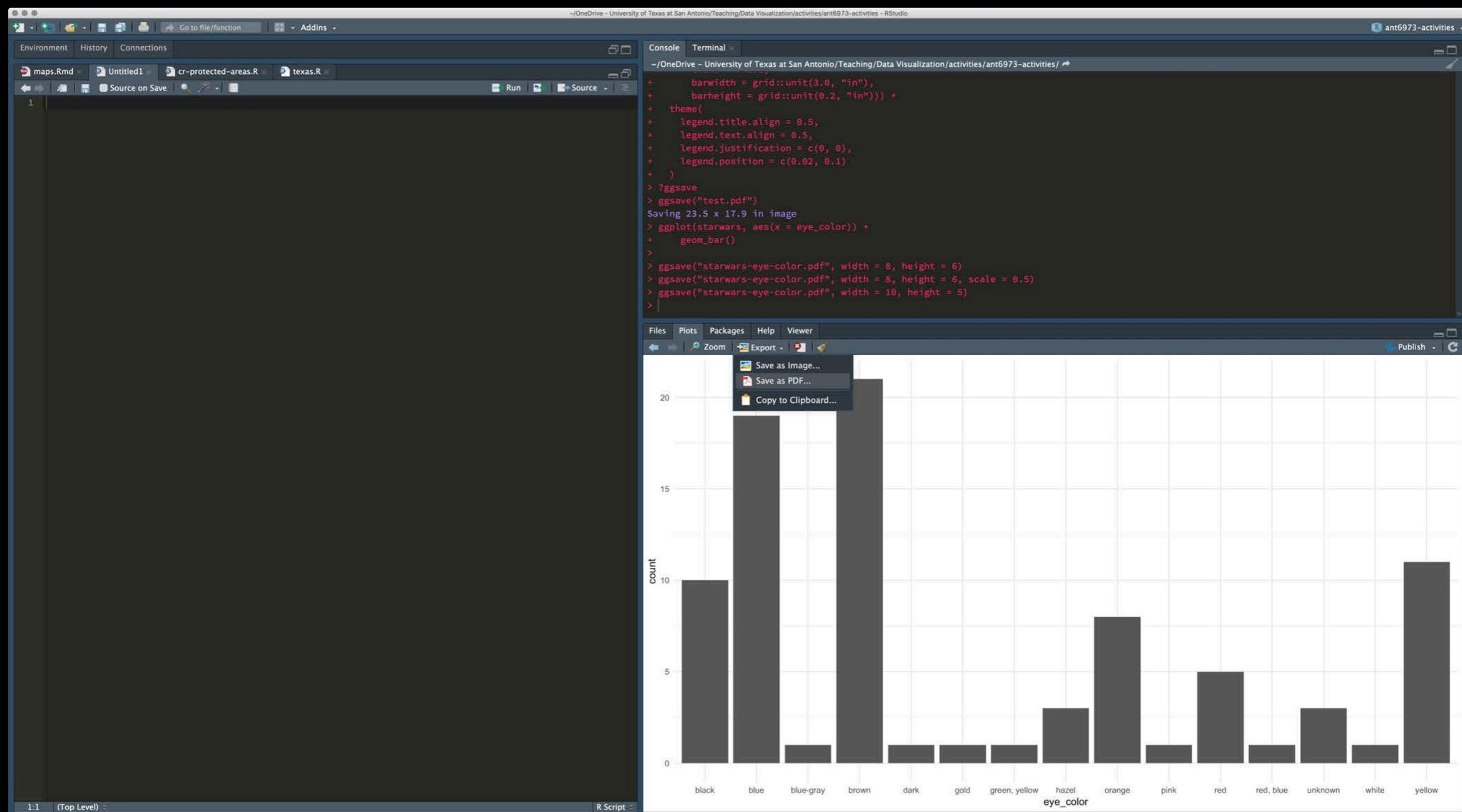
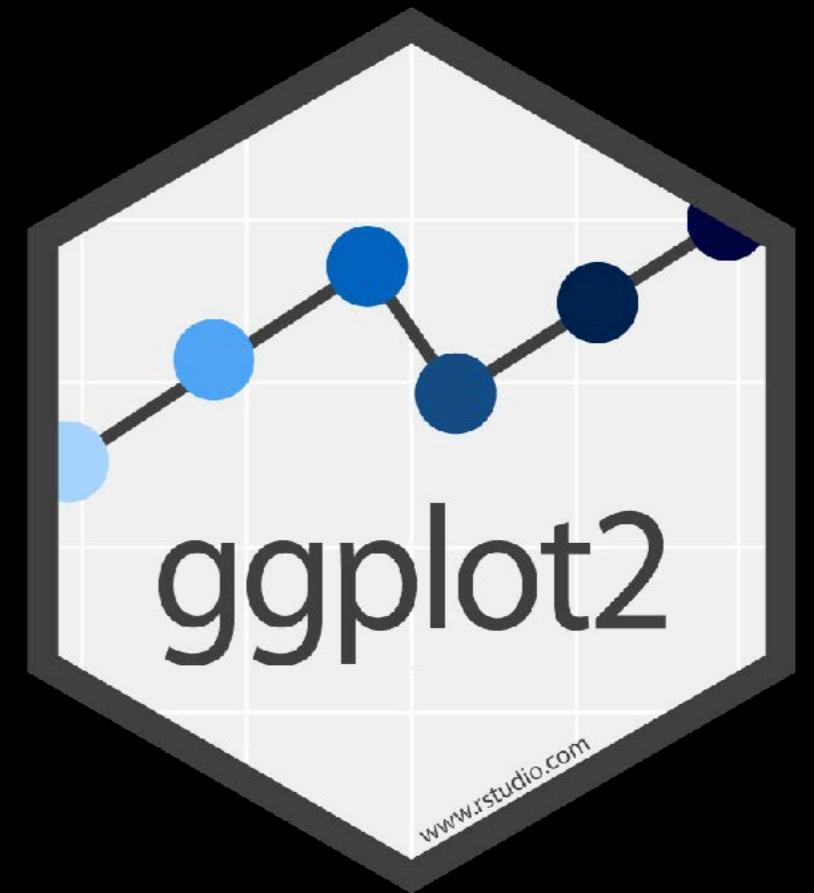


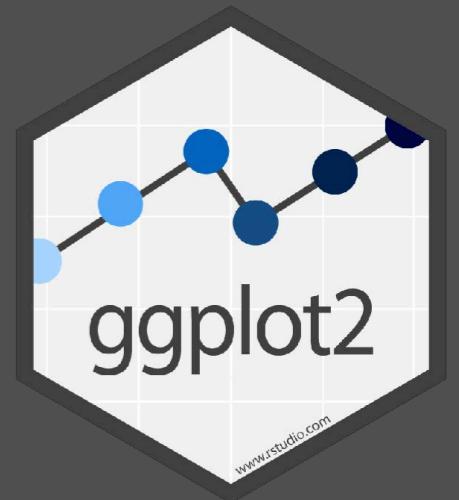
IMAGE FORMATS

- Refer to Wikle Ch. 27, "Understanding the most commonly used image file formats"
- Rules of thumb:
 1. Use PDF whenever possible, especially for creating high-quality, publication-ready files.
 2. Use PNG or TIFF for reports and web content, if required by journal, or for plots with huge amounts of data (e.g., millions of points).
 3. Use JPG only for photographs or if PNG file is too large.

VISUALIZING GEOSPATIAL DATA



ACTIVITY: MAPS



- Go to this week's assignments on the course website.
- Follow the instructions in the maps.Rmd file to learn some map-making skills.
- Submit the html report (not the Rmd) before next class.

