CSCI 491: Data Visualization

4- From Data to Visualization – Color, Size and Position

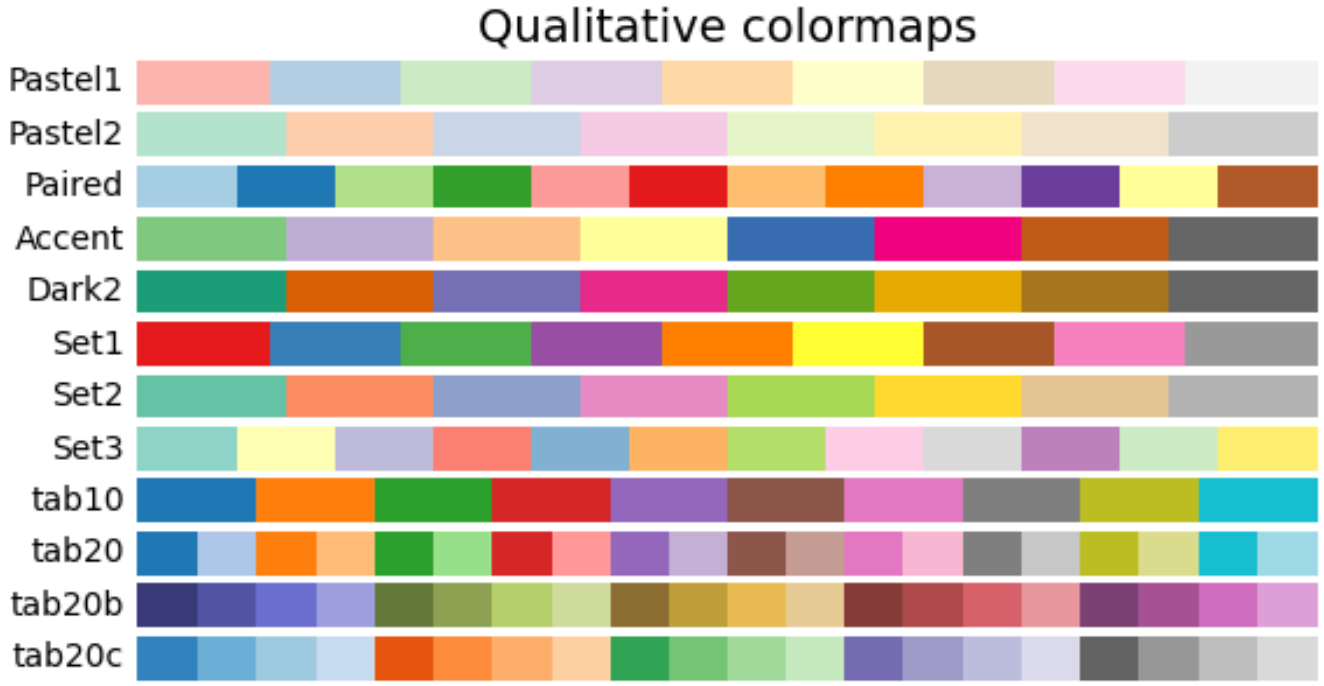
Purpose of Color in Data Visualization

We can use color to:

1. to label : distinguish groups of data from each other (unordered sets)

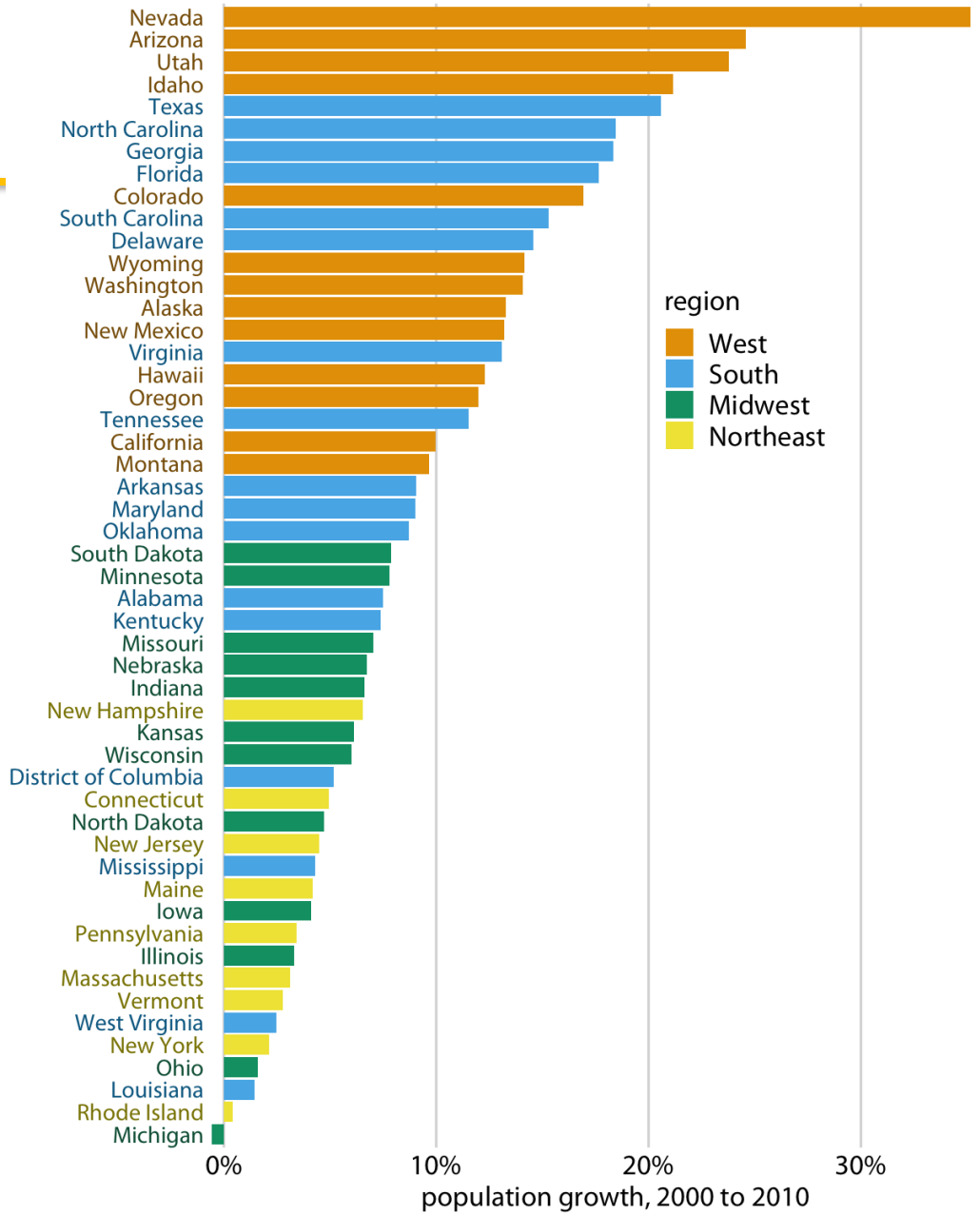
Color scale to distinguish groups of data

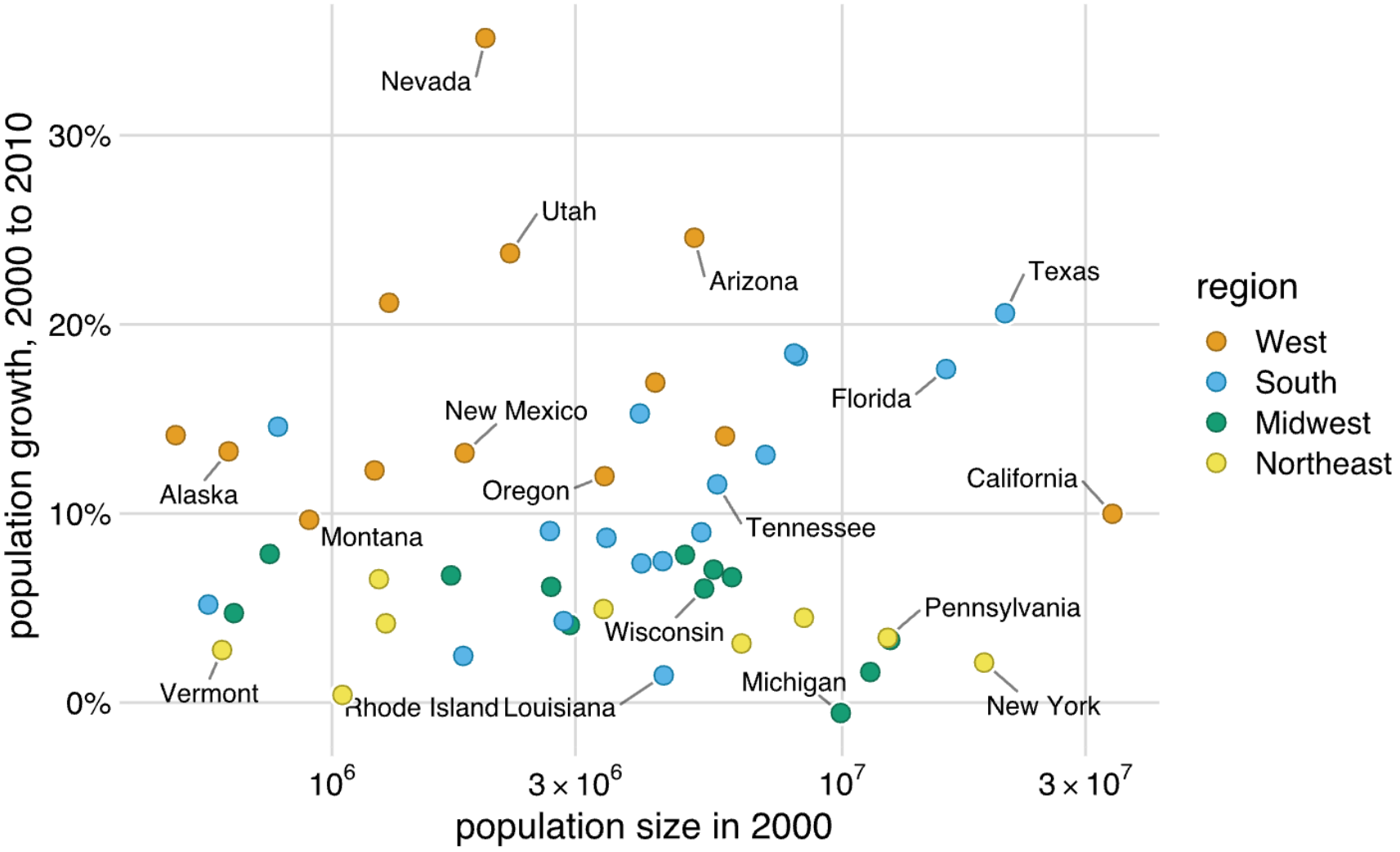
Qualitative color scale with colors that are chosen to look clearly distinct from each other while also being equivalent to each other (no order, no standout color)

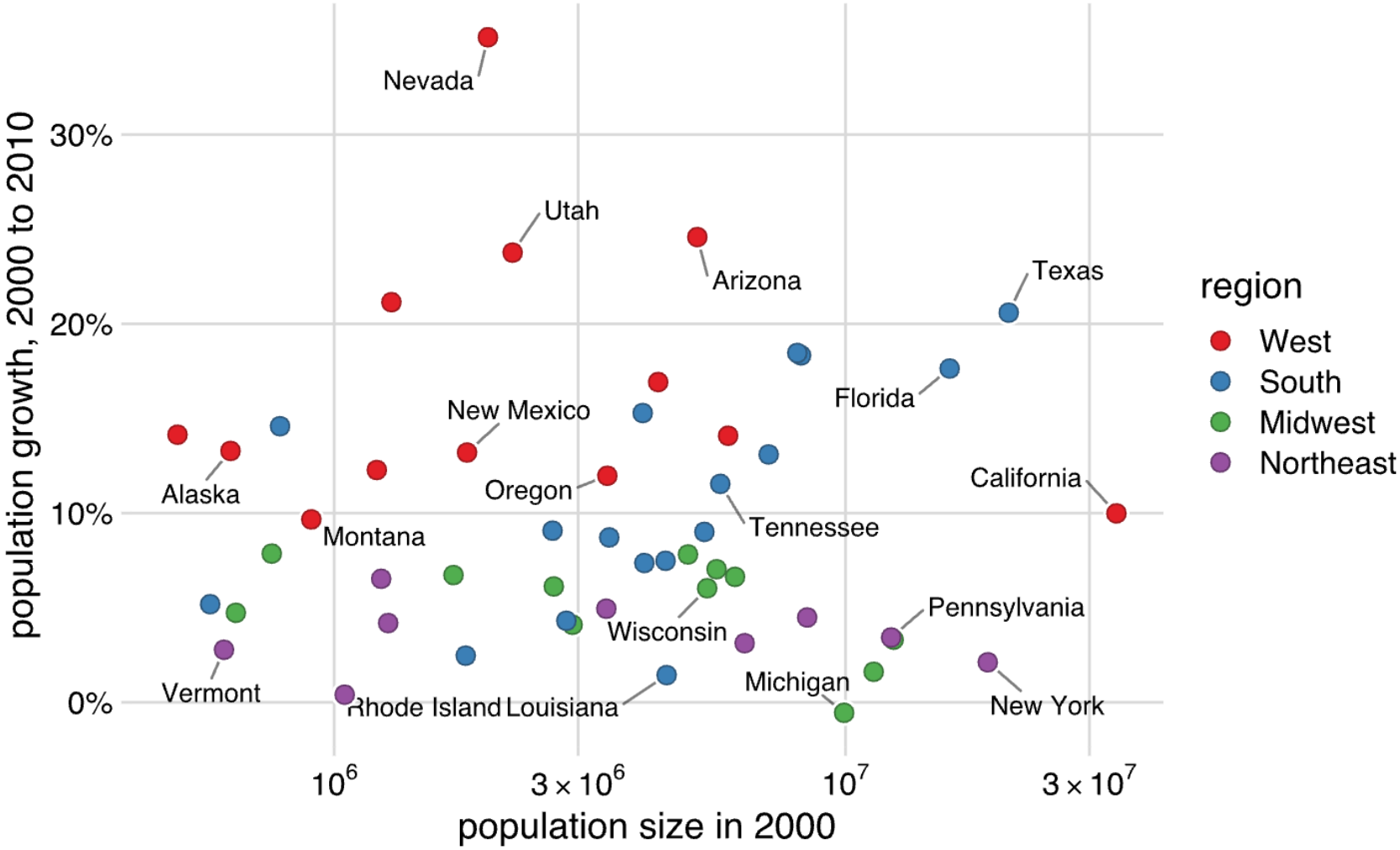


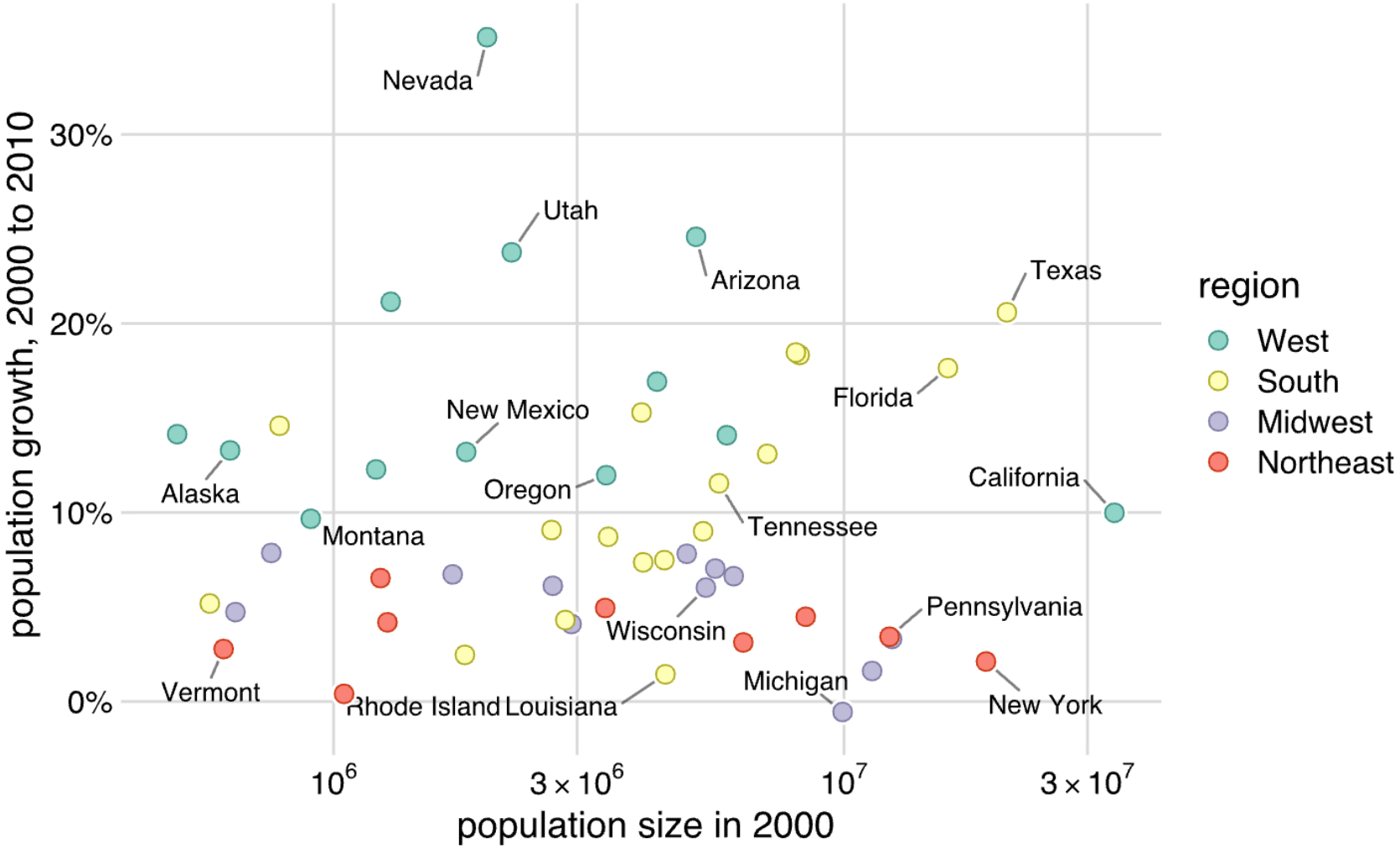
Example of Matplotlib Qualitative colormaps

Color scale to distinguish groups of data(example)




Nevada
Utah
Arizona
Texas
Florida
New Mexico
Alaska
Oregon
California
Montana
Tennessee
Pennsylvania
Wisconsin
Vermont
Rhode Island
Louisiana
Michigan
New York
30%
20%
10%
0%
population growth, 2000 to 2010
10^6
3×10^6
10^7
3×10^7
population size in 2000
region
West
South
Midwest
Northeast


Nevada
Utah
Arizona
Texas
Florida
New Mexico
Alaska
Oregon
California
Montana
Tennessee
Pennsylvania
Wisconsin
Vermont
Rhode Island
Louisiana
Michigan
New York
30%
20%
10%
0%
population growth, 2000 to 2010
10^6
3×10^6
10^7
3×10^7
population size in 2000
region
West
South
Midwest
Northeast


Nevada
Utah
Arizona
Texas
Florida
New Mexico
Alaska
Oregon
California
Montana
Tennessee
Pennsylvania
Wisconsin
Vermont
Rhode Island
Louisiana
Michigan
New York
30%
20%
10%
0%
population growth, 2000 to 2010
10^6
3×10^6
10^7
3×10^7
population size in 2000
region
West
South
Midwest
Northeast

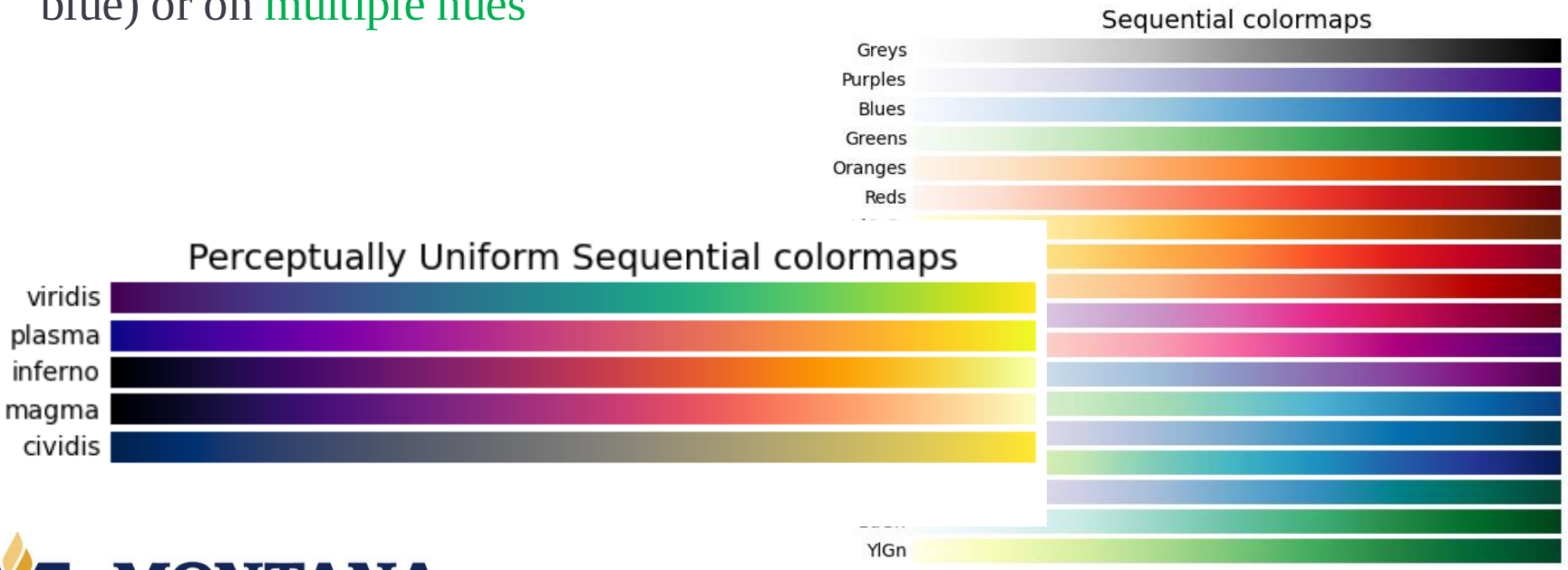
Color Scales Application

We can use color to:

1. to label: distinguish groups of data from each other (unordered sets)
2. to measure: to represent data values

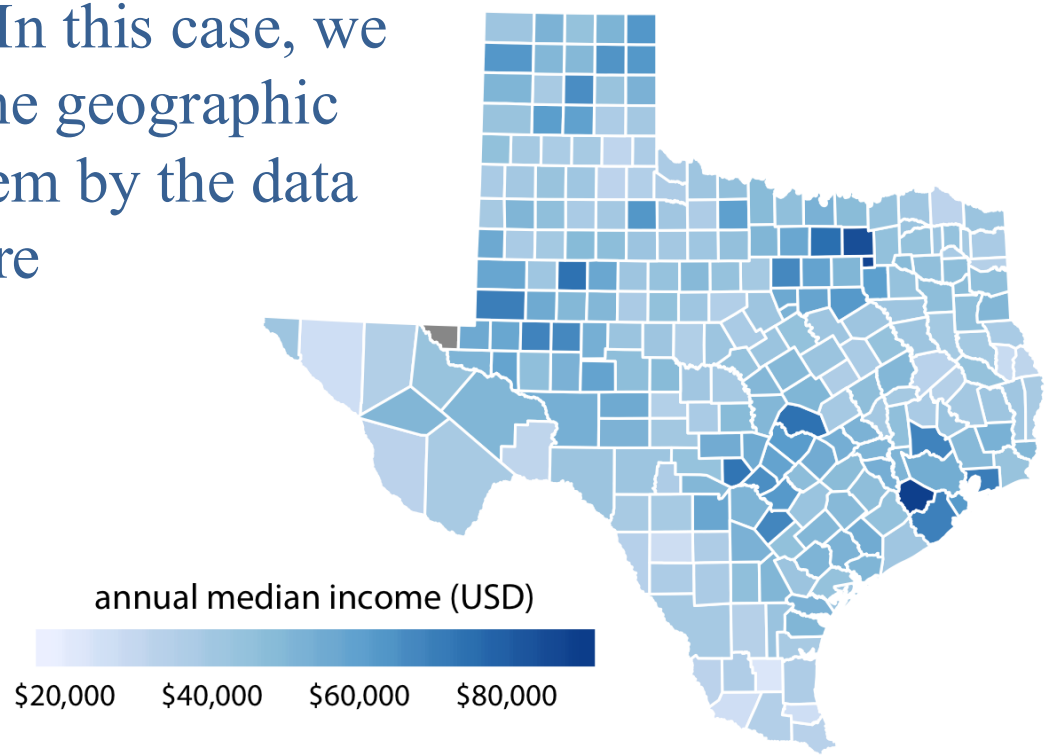
Color scale to represent data values

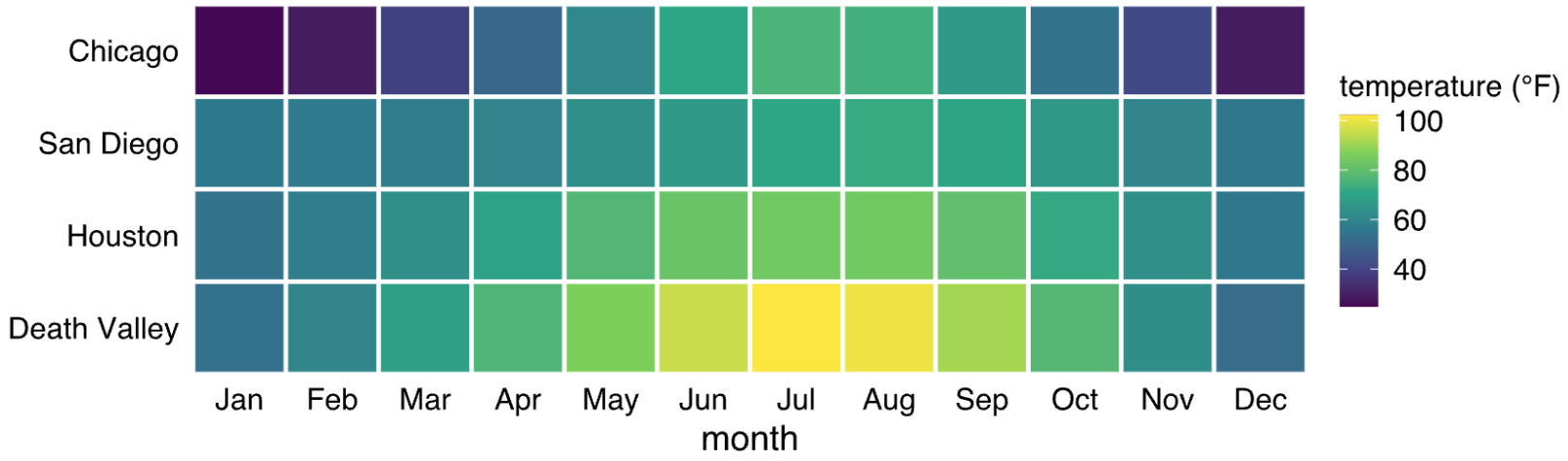
In this case, we use a sequential color scale.
Such a scale contains a sequence of colors that clearly indicate which values are larger or smaller than which other ones, and how distant two specific values are from each other. The second point implies that the color scale needs to be perceived to vary uniformly across its entire range.
Sequential scales can be based on a single hue (e.g., from dark blue to light blue) or on multiple hues

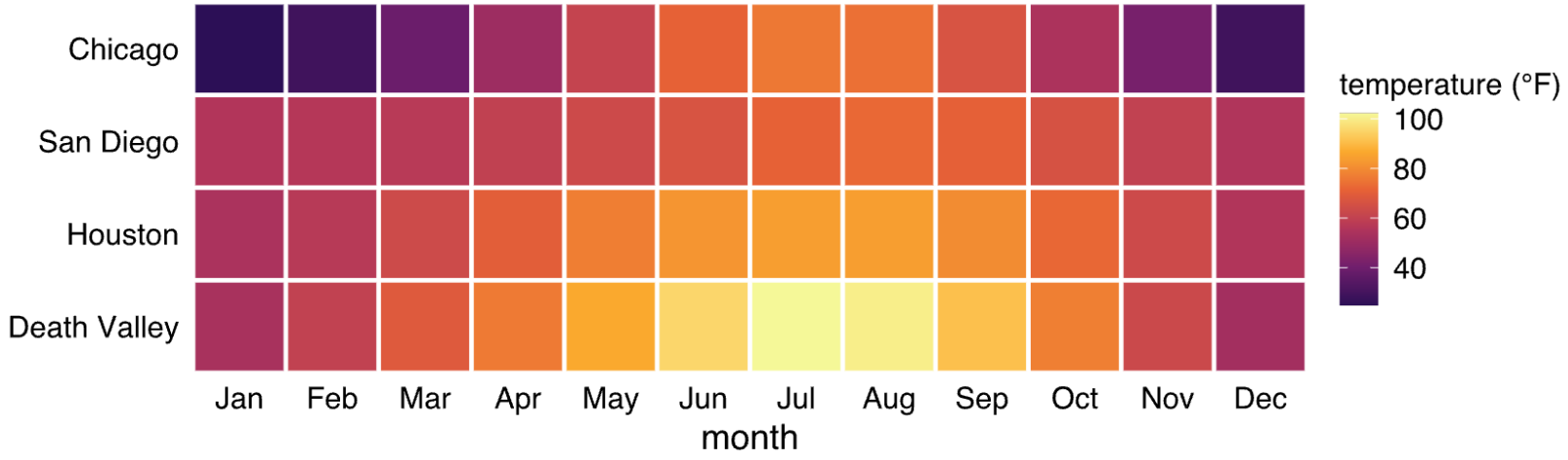


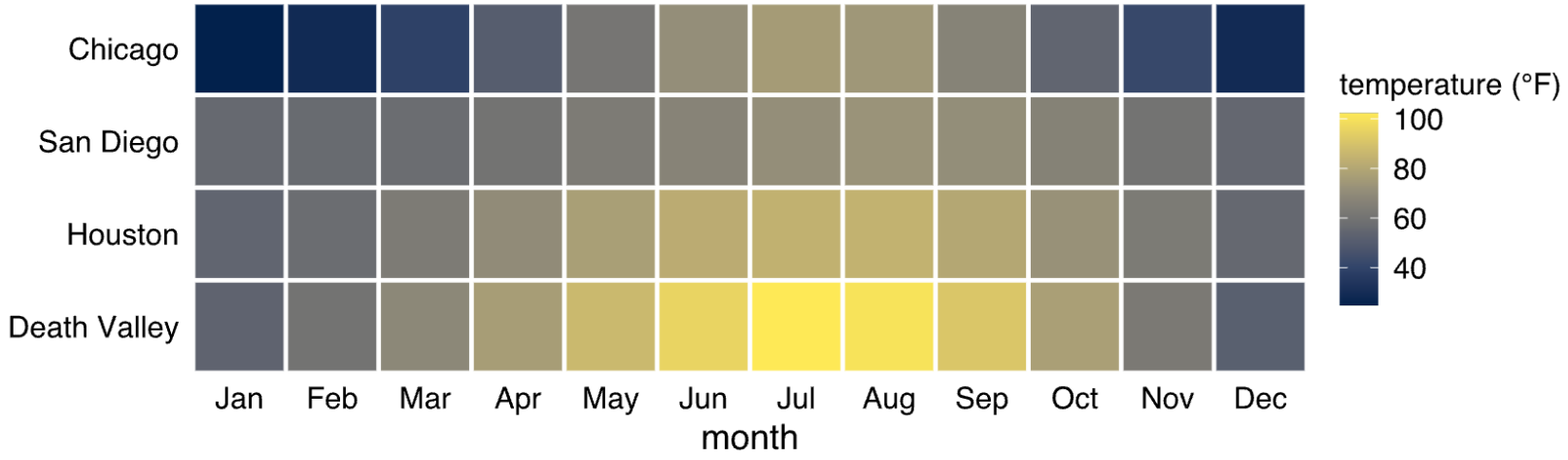
Color scale to represent data values (example)

- Representing data values as colors is particularly useful when we want to show how the data values vary across geographic regions. In this case, we can draw a map of the geographic regions and color them by the data values. Such maps are called choropleths



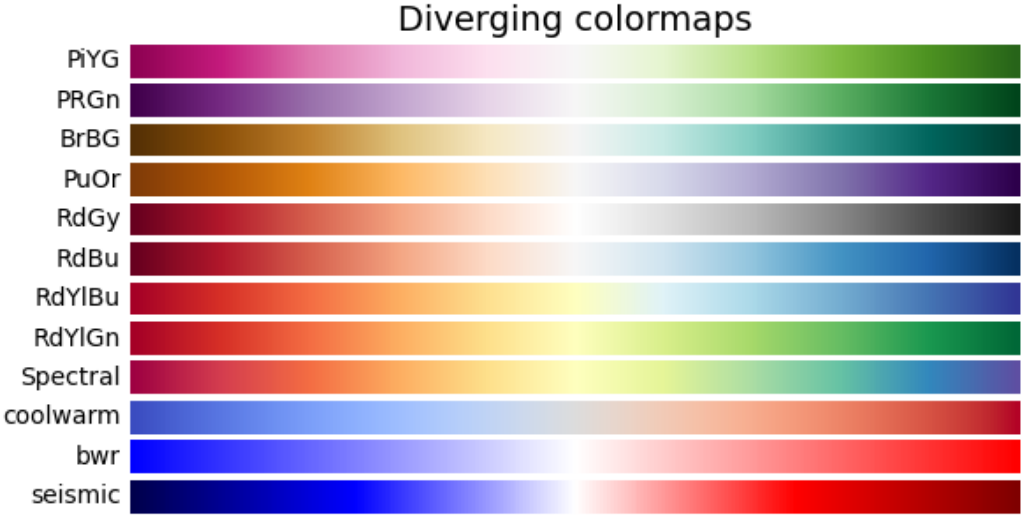

Chicago
San Diego
Houston
Death Valley
Jan
Feb
Mar
Apr
May
Jun
Jul
Aug
Sep
Oct
Nov
Dec
month
temperature (°F)
100
80
60
40


Chicago
San Diego
Houston
Death Valley
Jan
Feb
Mar
Apr
May
Jun
Jul
Aug
Sep
Oct
Nov
Dec
month
temperature (°F)
100
80
60
40

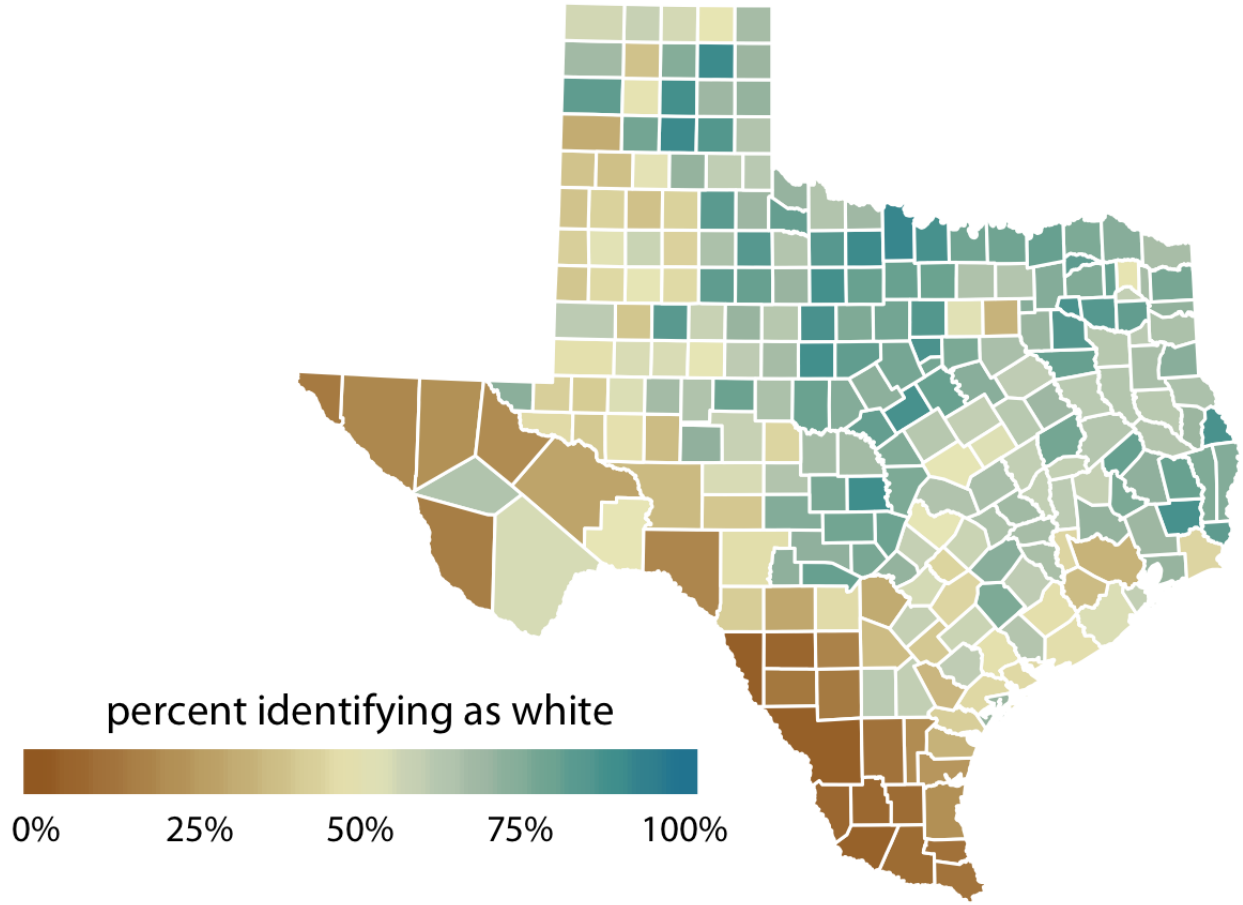

Chicago
San Diego
Houston
Death Valley
Jan
Feb
Mar
Apr
May
Jun
Jul
Aug
Sep
Oct
Nov
Dec
month
temperature (°F)
100
80
60
40

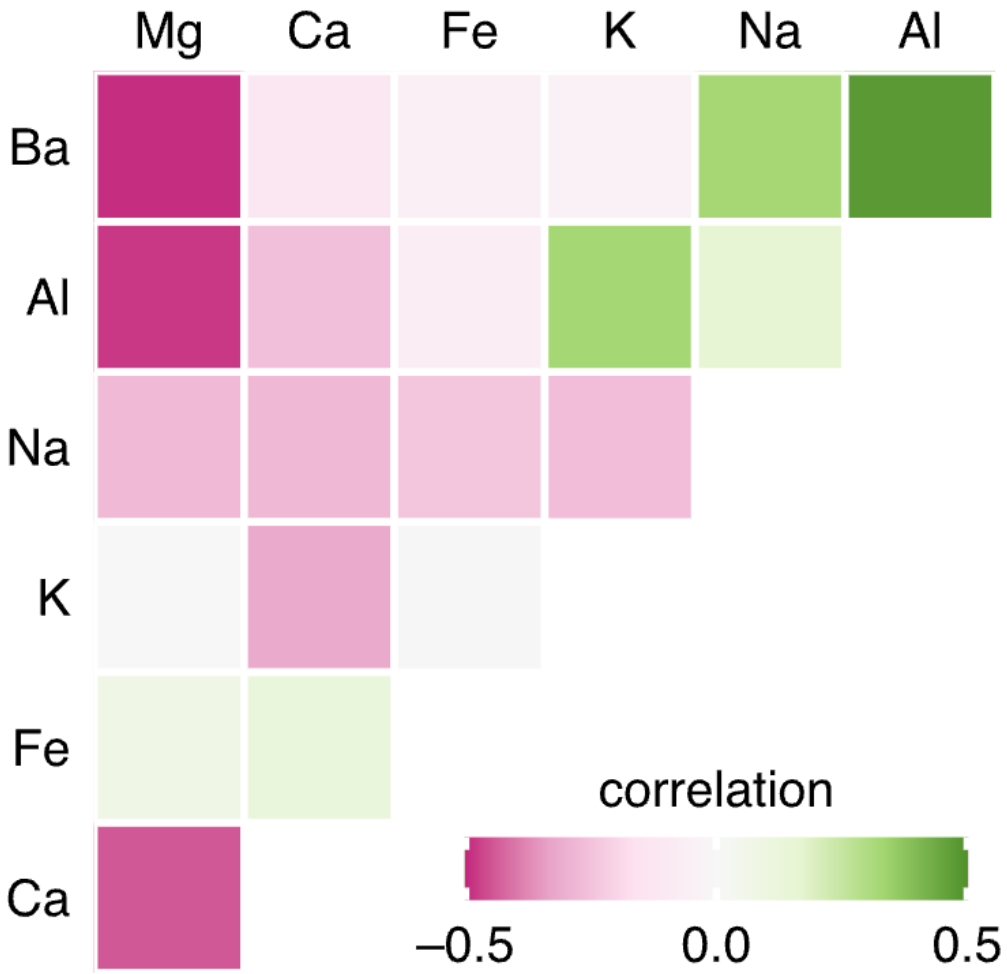
Representing deviation in data values

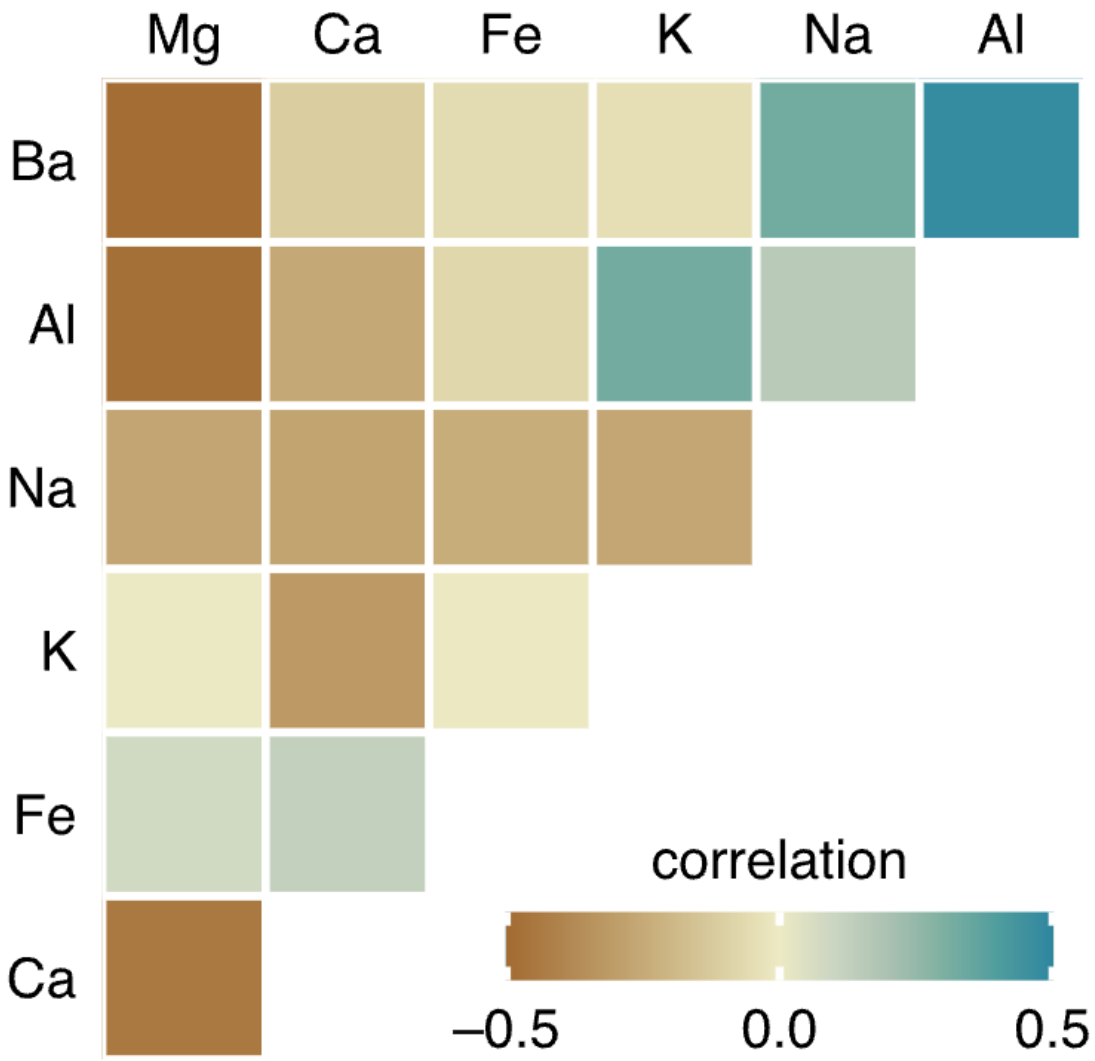
- In some cases, we need to visualize the deviation of data values in one of two directions relative to a neutral midpoint. Example?
- We can think of a diverging scale as two sequential scales stitched together at a common midpoint, which usually is represented by a light color

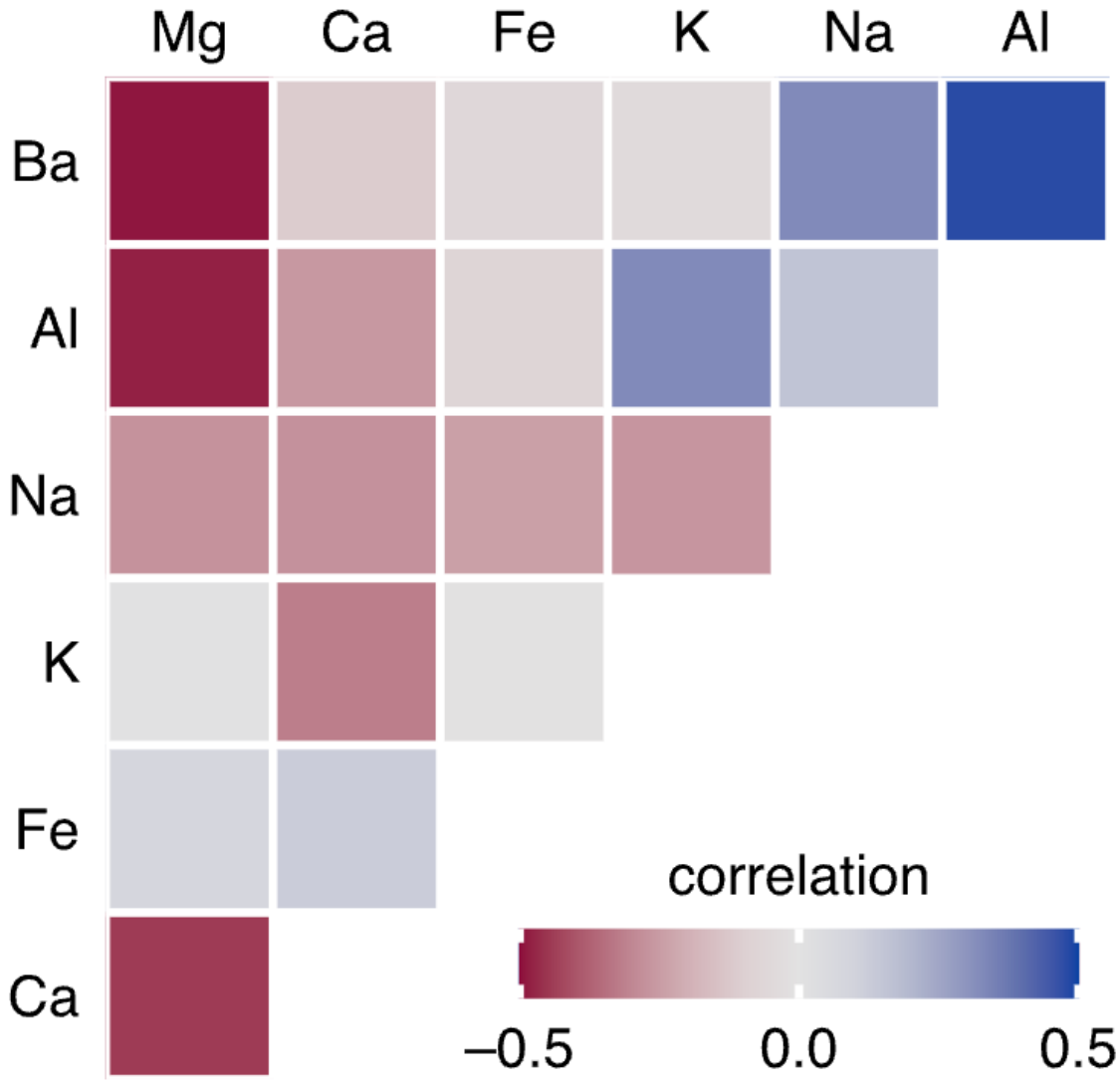


Example of the diverging color scale




Mg
Ca
Fe
K
Na
Al
Ba
Al
Na
K
Fe
Ca
correlation
–0.5
0.0
0.5


Mg
Ca
Fe
K
Na
Al
Ba
Al
Na
K
Fe
Ca
correlation
–0.5
0.0
0.5


Mg
Ca
Fe
K
Na
Al
Ba
Al
Na
K
Fe
Ca
correlation
–0.5
0.0
0.5

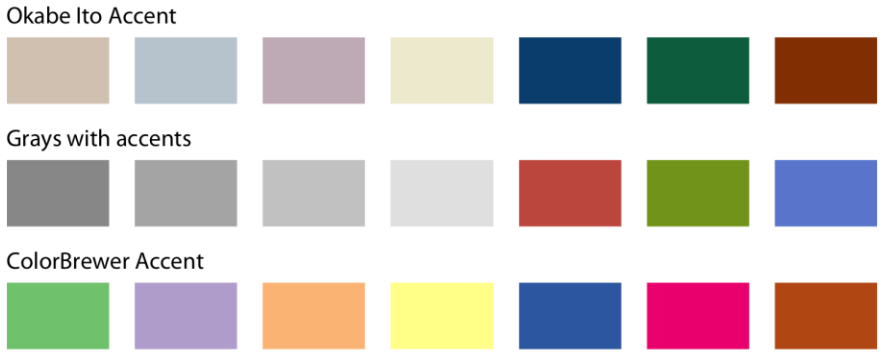
Color Scales Application

We can use color to:

1. to label: distinguish groups of data from each other (unordered sets)
2. to measure: to represent data values
3. to highlight a finding.
4.

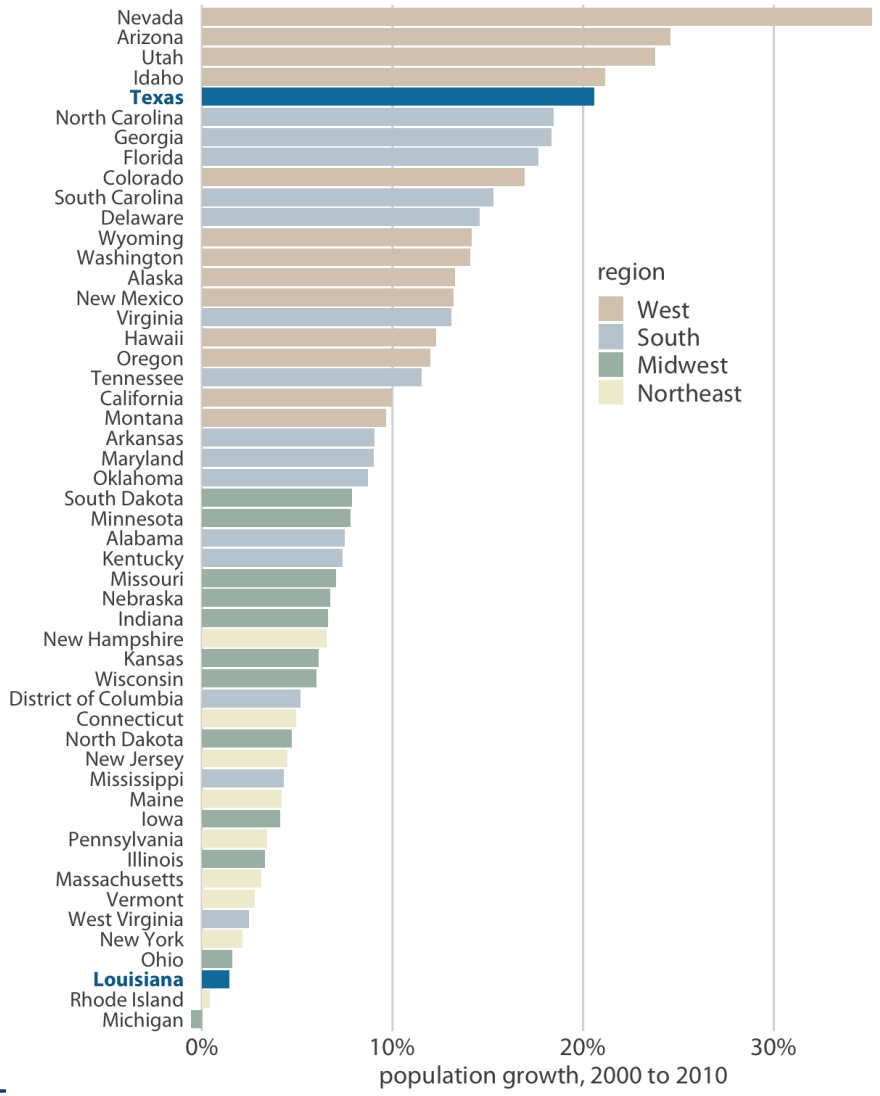
Color as a Tool to Highlight

- Color can also be an effective tool to highlight specific elements in the data. We can strengthen the story by emphasizing the relevant figure elements to the reader.
- An easy way to achieve this emphasis is to color these figure elements in a color or set of colors that vividly stand out against the rest of the figure. This effect can be achieved with accent color scales



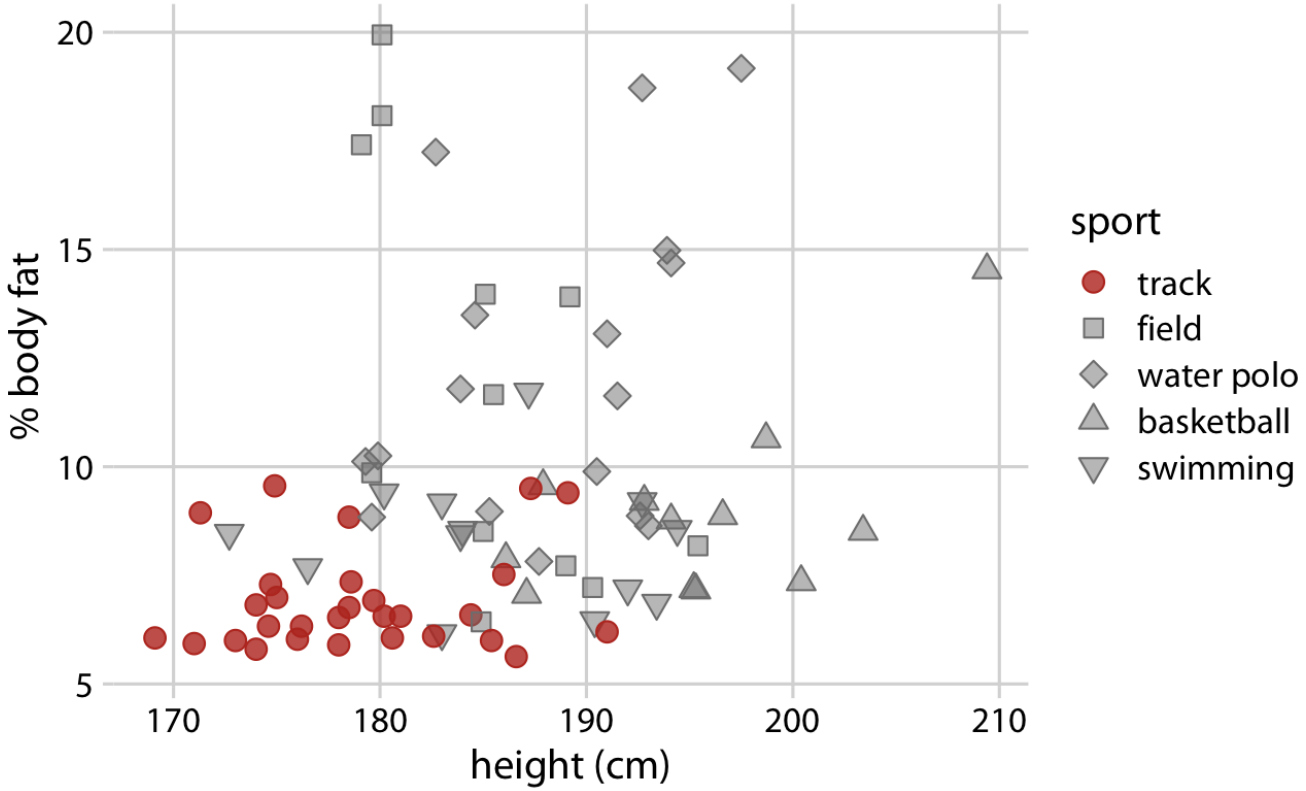
Color as a Tool to Highlight (Example)

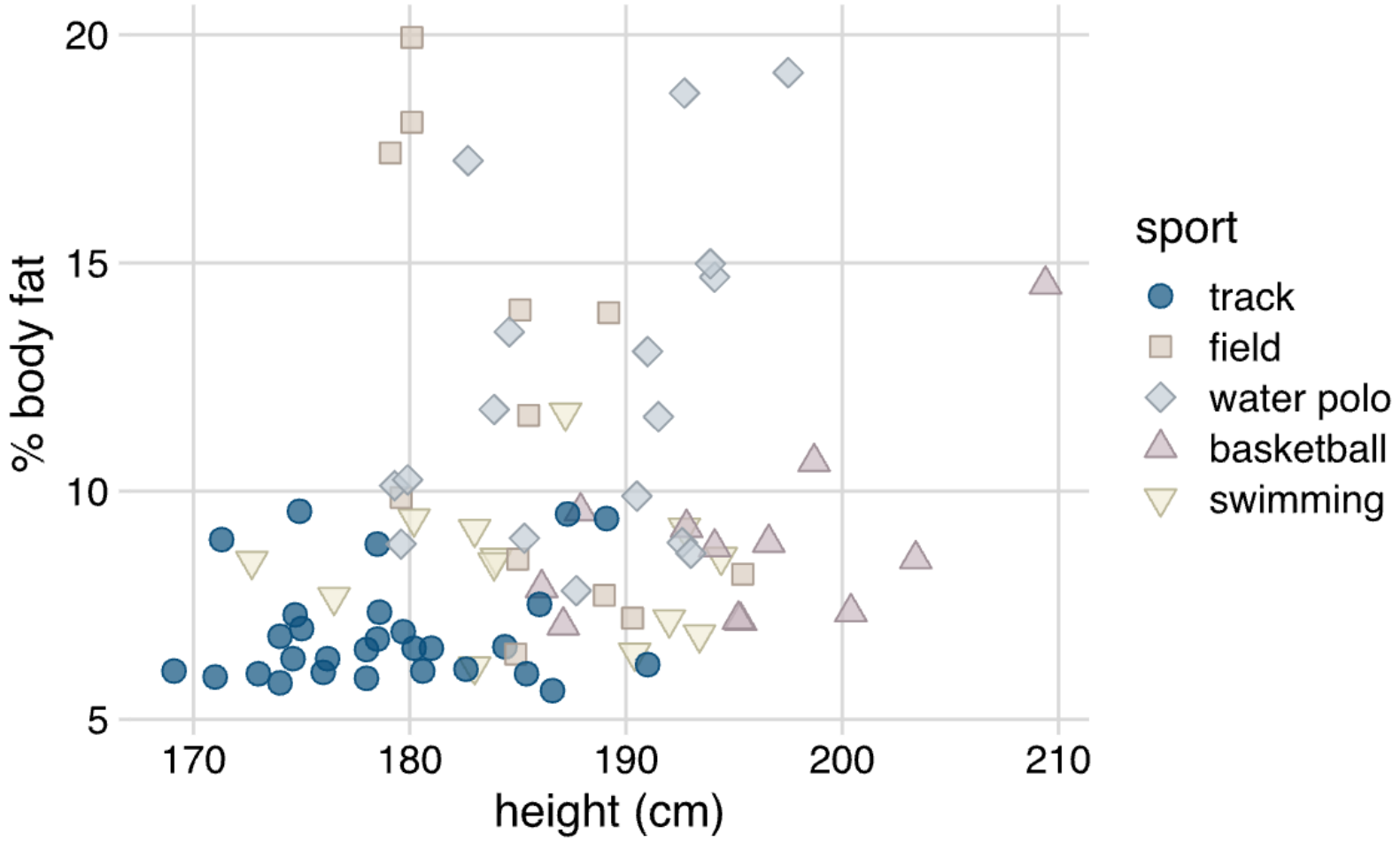
- What is the central message of this plot?

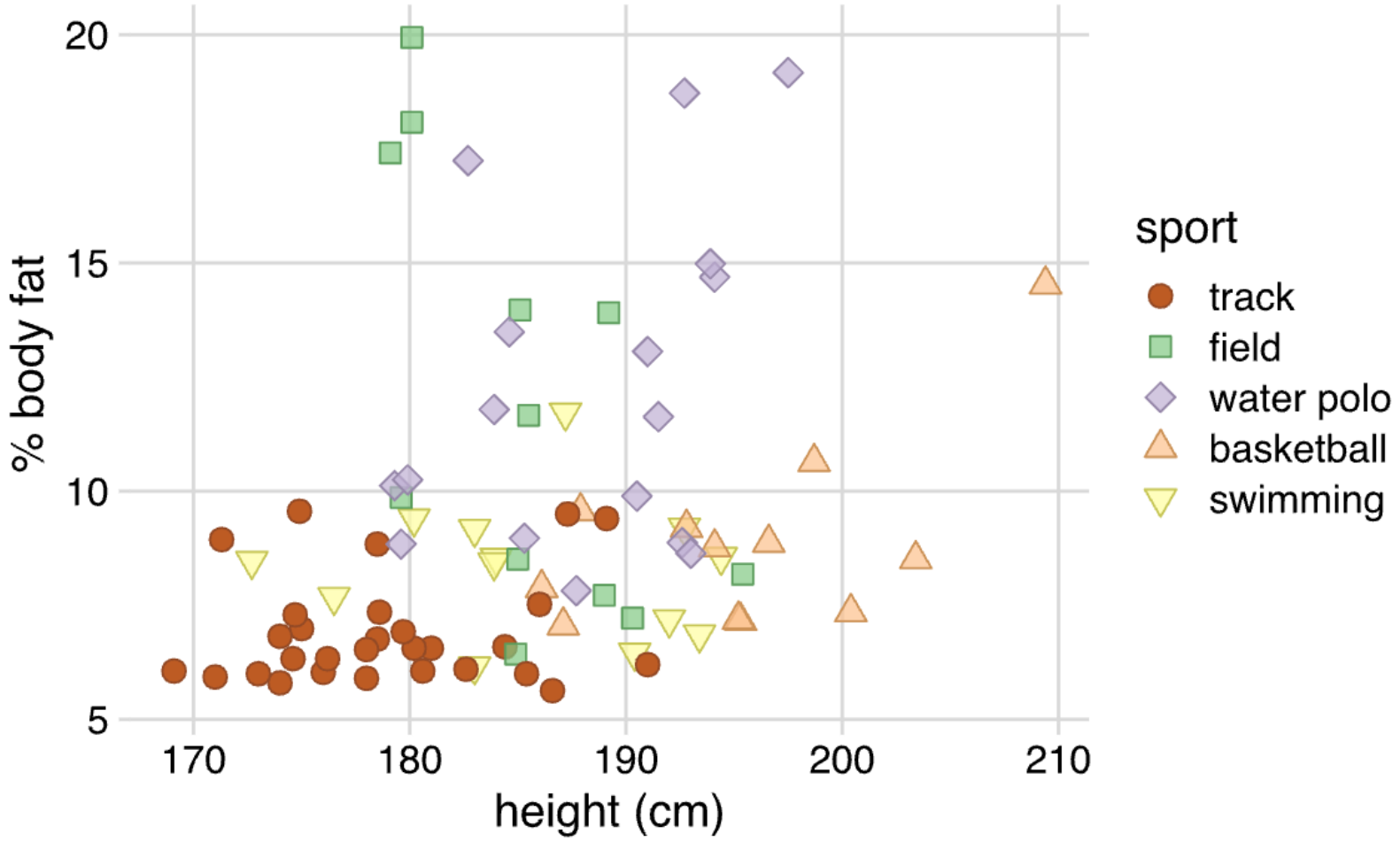


Color as a Tool to Highlight (Example)

- When working with accent colors, it is critical that the baseline colors do not compete for attention.




20
15
10
5
% body fat
170
180
190
200
210
height (cm)
sport
track
field
water polo
basketball
swimming


20
15
10
5
% body fat
170
180
190
200
210
height (cm)
sport
track
field
water polo
basketball
swimming

Color Scales Application

We can use color to:

1. to label: distinguish groups of data from each other (unordered sets)
2. to measure: to represent data values
3. to highlight a finding.
4. to enliven and decorate!

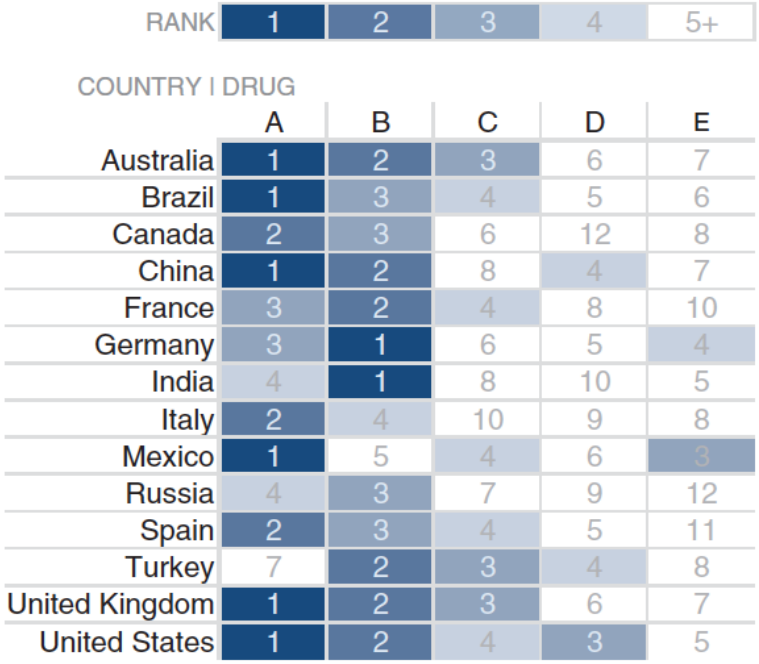
Use color sparingly

Country Level Sales Rank Top 5 Drugs

Rainbow distribution in color indicates sales rank in given country from #1 (red) to #10 or higher (dark purple)

Country	A	B	C	D	E
AUS	1	2	3	6	7
BRA	1	3	4	5	6
CAN	2	3	6	12	8
CHI	1	2	8	4	7
FRA	3	2	4	8	10
GER	3	1	6	5	4
IND	4	1	8	10	5
ITA	2	4	10	9	8
MEX	1	5	4	6	3
RUS	4	3	7	9	12
SPA	2	3	4	5	11
TUR	7	2	3	4	8
UK	1	2	3	6	7
US	1	2	4	3	5

Top 5 drugs: country-level sales rank



Use color consistently

- Your audience will typically take time to familiarize themselves with what colors mean once and then will assume the same details apply throughout the rest of the communication.

Be thoughtful of tone that color conveys

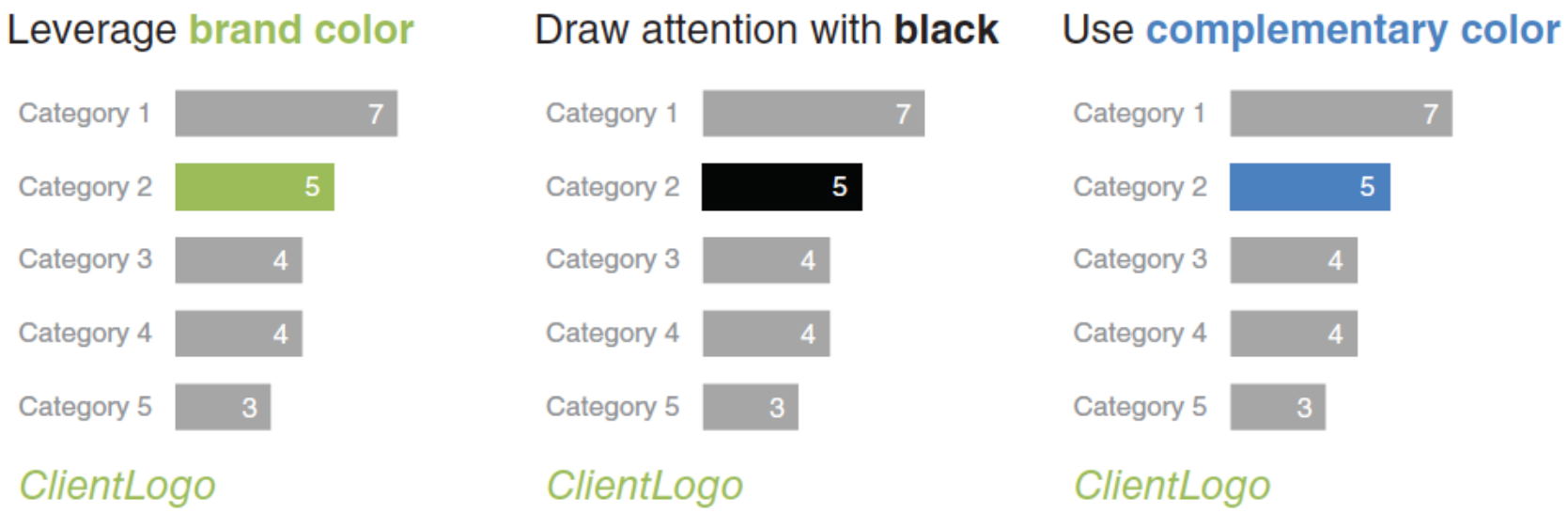
- Color evokes emotion. Consider the tone you want to set with your data visualization or broader communication and choose a color (or colors) that help reinforce the emotion you want to arouse from your audience.
- When picking colors for communications to international audiences, it may be essential to consider the connotations colors have in other cultures.
- https://informationisbeautiful.net/visualizations/colours-in-cultures/

Color vision deficiencies

- Roughly 8% of men and half a percent of women are colorblind.

- The most common form of color vision deficiency involves differentiating between red and green. Blue and orange is a better choice.
- Thus, avoiding colormaps with both red and green will avoid many problems in general.
- https://vischeck.com/examples/
- Color blindness simulator
- https://colororacle.org/

Brand colors: to leverage or not to leverage?



COLORBREWER 2.0

- https://colorbrewer2.org/#type=diverging&scheme=RdGy&n=3