# CP321 Data Visualization

Jiashu (Jessie) Zhao

# Visualizing data: Mapping data onto aesthetics

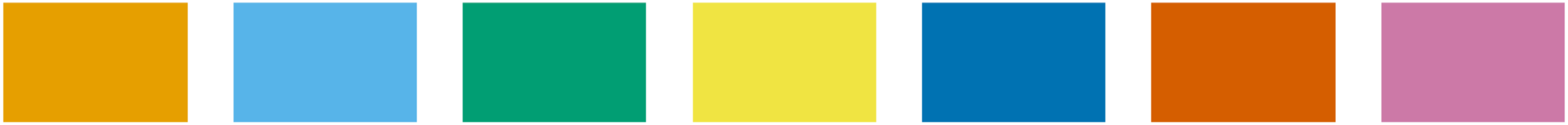
- Position - Coordinate systems
- Color Scales
- Shape – *Markers in Python*
- Line Type

# Data Visualization in Practice: Color Scales

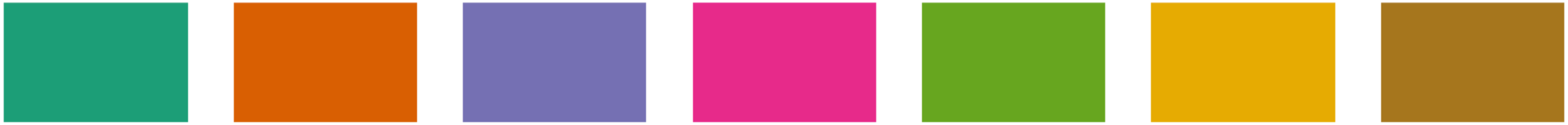
- There are three fundamental use cases for color in data visualizations:
  - (i) distinguish groups of data from each other
  - (ii) represent data values
  - (iii) highlight

- Color as a tool to distinguish
    - We frequently use color to distinguish discrete items or groups that do not have an intrinsic order, such as different countries on a map or different manufacturers of a certain product.
    - In this case, we use a *qualitative* color scale. Such a scale contains a finite set of specific colors that are chosen to look clearly distinct from each other while also being equivalent to each other.
    - The second condition requires that no one color should stand out relative to the others.

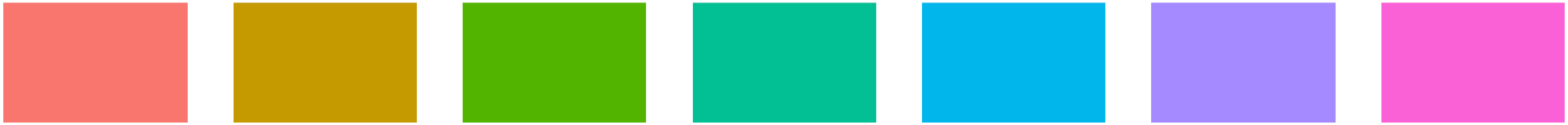
Okabe Ito

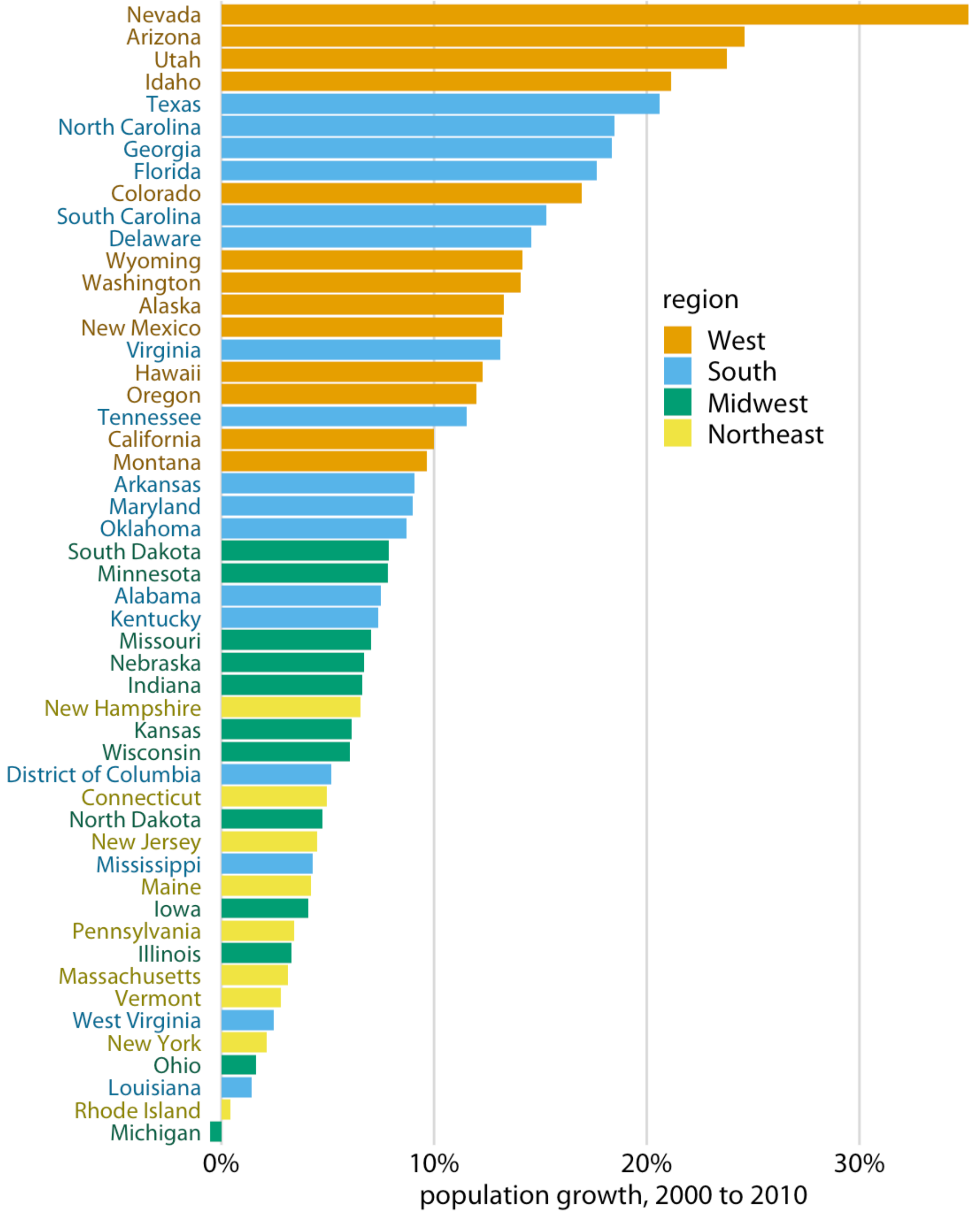


ColorBrewer Dark2



ggplot2 hue

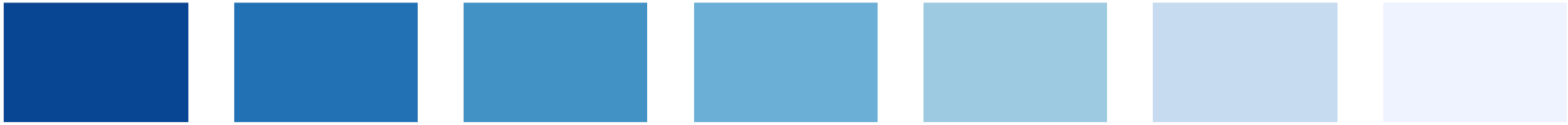



Nevada
Arizona
Utah
Idaho
Texas
North Carolina
Georgia
Florida
Colorado
South Carolina
Delaware
Wyoming
Washington
Alaska
New Mexico
Virginia
Hawaii
Oregon
Tennessee
California
Montana
Arkansas
Maryland
Oklahoma
South Dakota
Minnesota
Alabama
Kentucky
Missouri
Nebraska
Indiana
New Hampshire
Kansas
Wisconsin
District of Columbia
Connecticut
North Dakota
New Jersey
Mississippi
Maine
Iowa
Pennsylvania
Illinois
Massachusetts
Vermont
West Virginia
New York
Ohio
Louisiana
Rhode Island
Michigan
0%
10%
20%
30%
population growth, 2000 to 2010
region
West
South
Midwest
Northeast

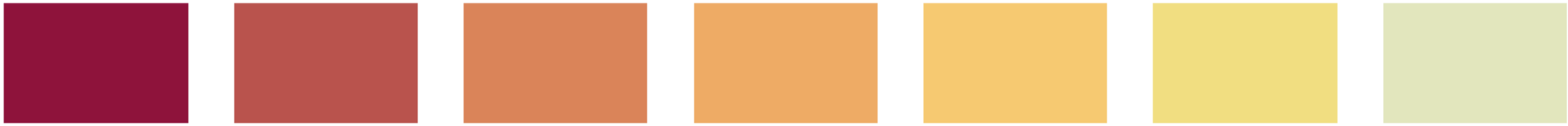
# Color to represent data values

- Color can also be used to represent data values, such as income, temperature, or speed.
- In this case, we use a *sequential* color scale. Such a scale contains a sequence of colors that clearly indicate
  - (i) which values are larger or smaller than which other ones and
  - (ii) how distant two specific values are from each other.
- Sequential scales 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).

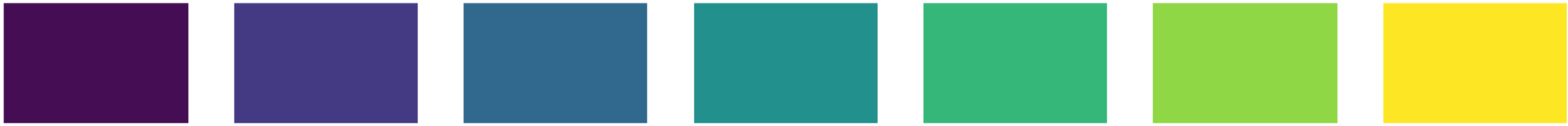
ColorBrewer Blues

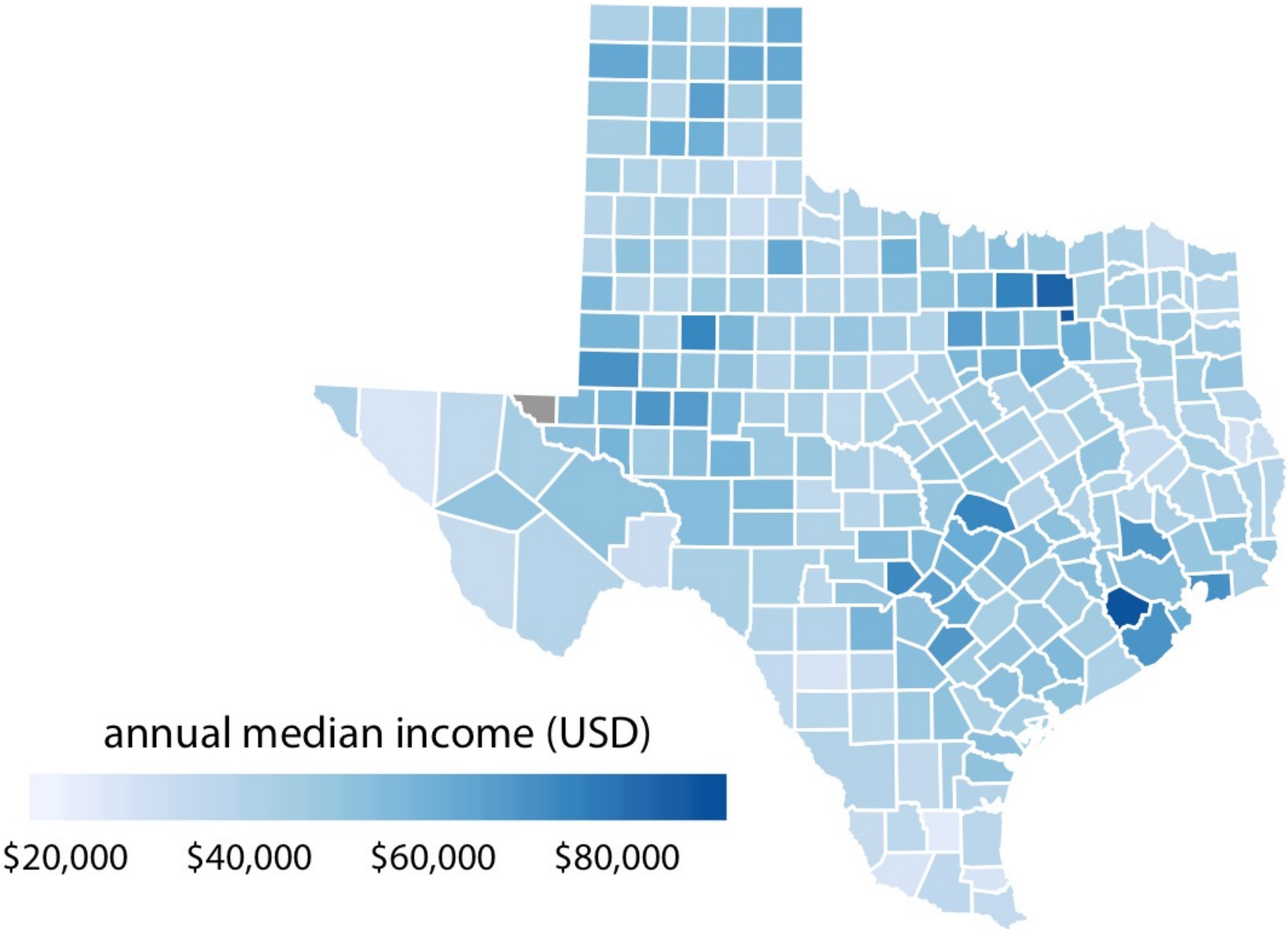


Heat



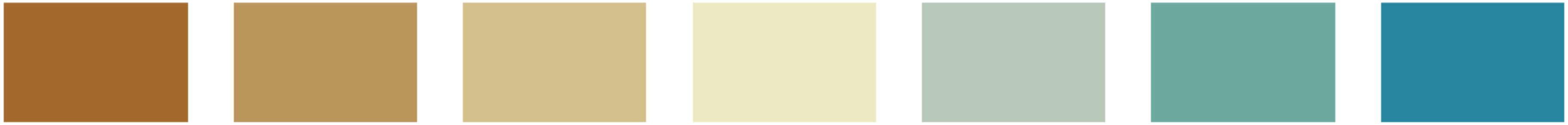
Viridis



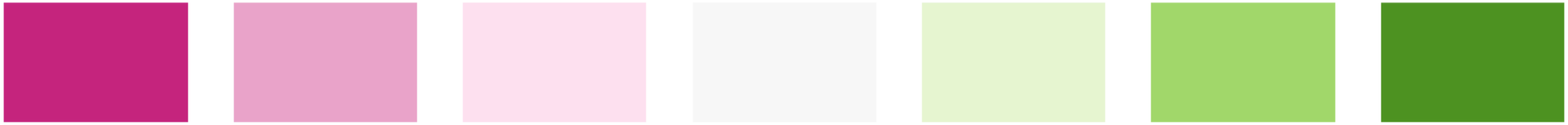

annual median income (USD)
$20,000
$40,000
$60,000
$80,000

- In some cases, we need to visualize the deviation of data values in one of two directions relative to a neutral midpoint.
- The appropriate color scale in this situation is a *diverging* color scale.
- Diverging scales need to be balanced, so that the progression from light colors in the center to dark colors on the outside is approximately the same in either direction.

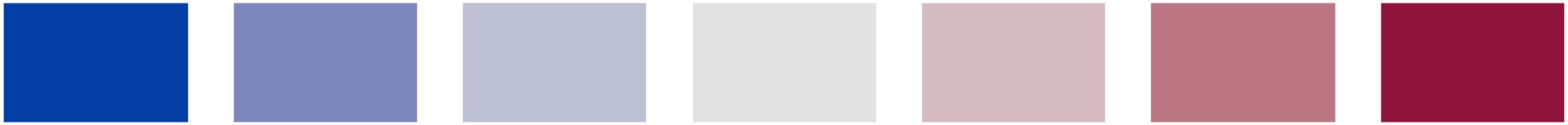
CARTO Earth

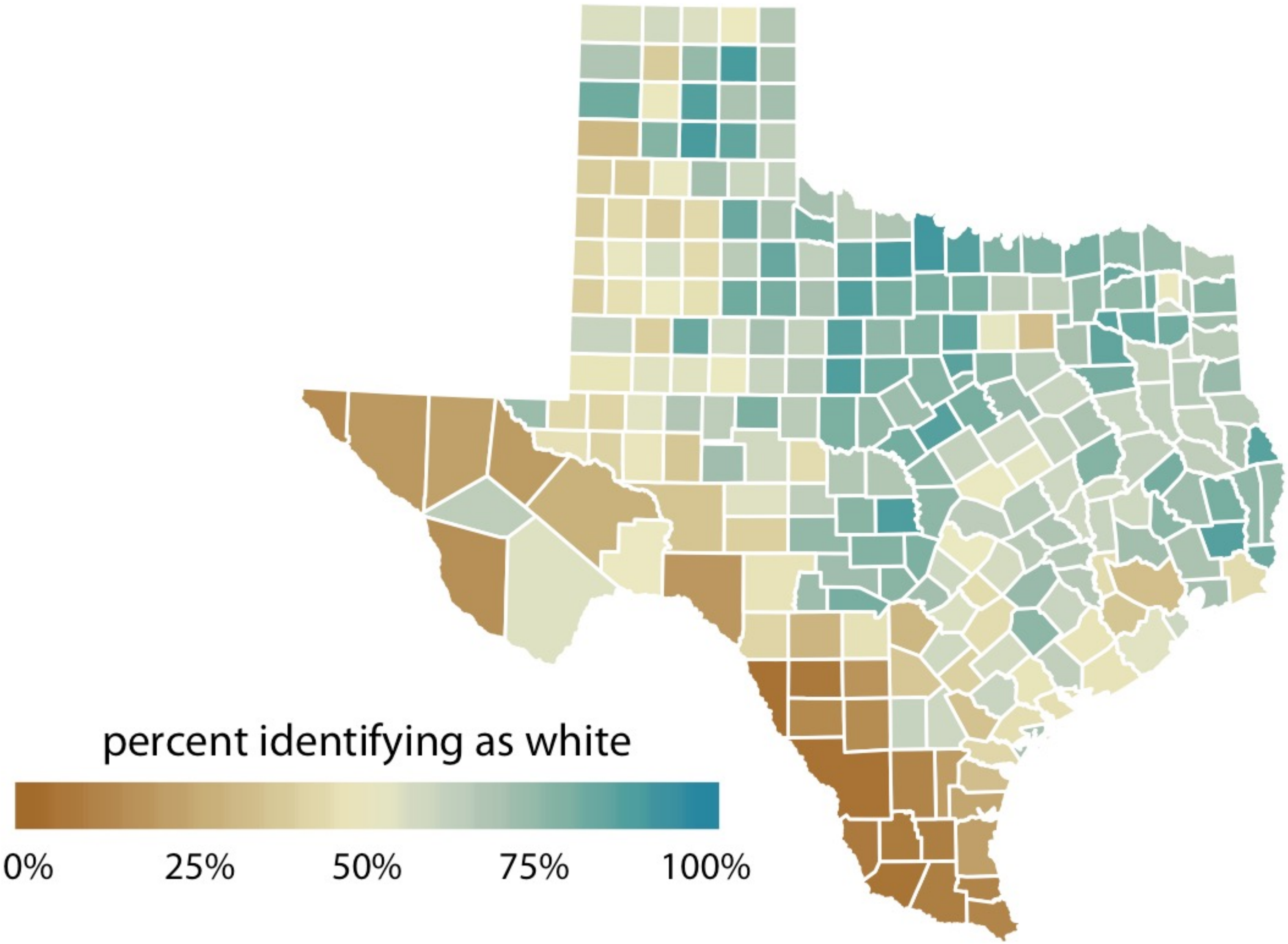


ColorBrewer PiYG



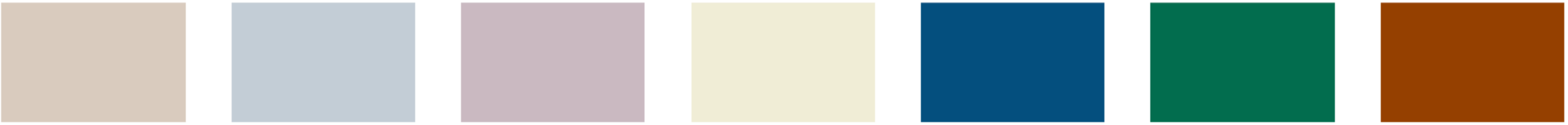
Blue-Red



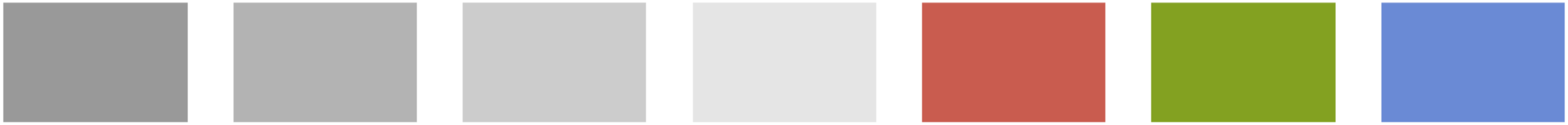

percent identifying as white
0%
25%
50%
75%
100%

- Color as a tool to highlight
- There may be specific categories or values in the dataset that carry key information about the story we want to tell, and we can strengthen the story by emphasizing the relevant figure elements to the reader.
- This effect can be achieved with *accent* color scales, which are color scales that contain both a set of subdued colors and a matching set of stronger, darker, and/or more saturated colors.

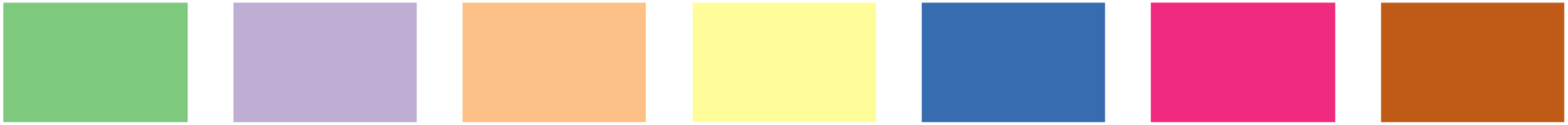
Okabe Ito Accent

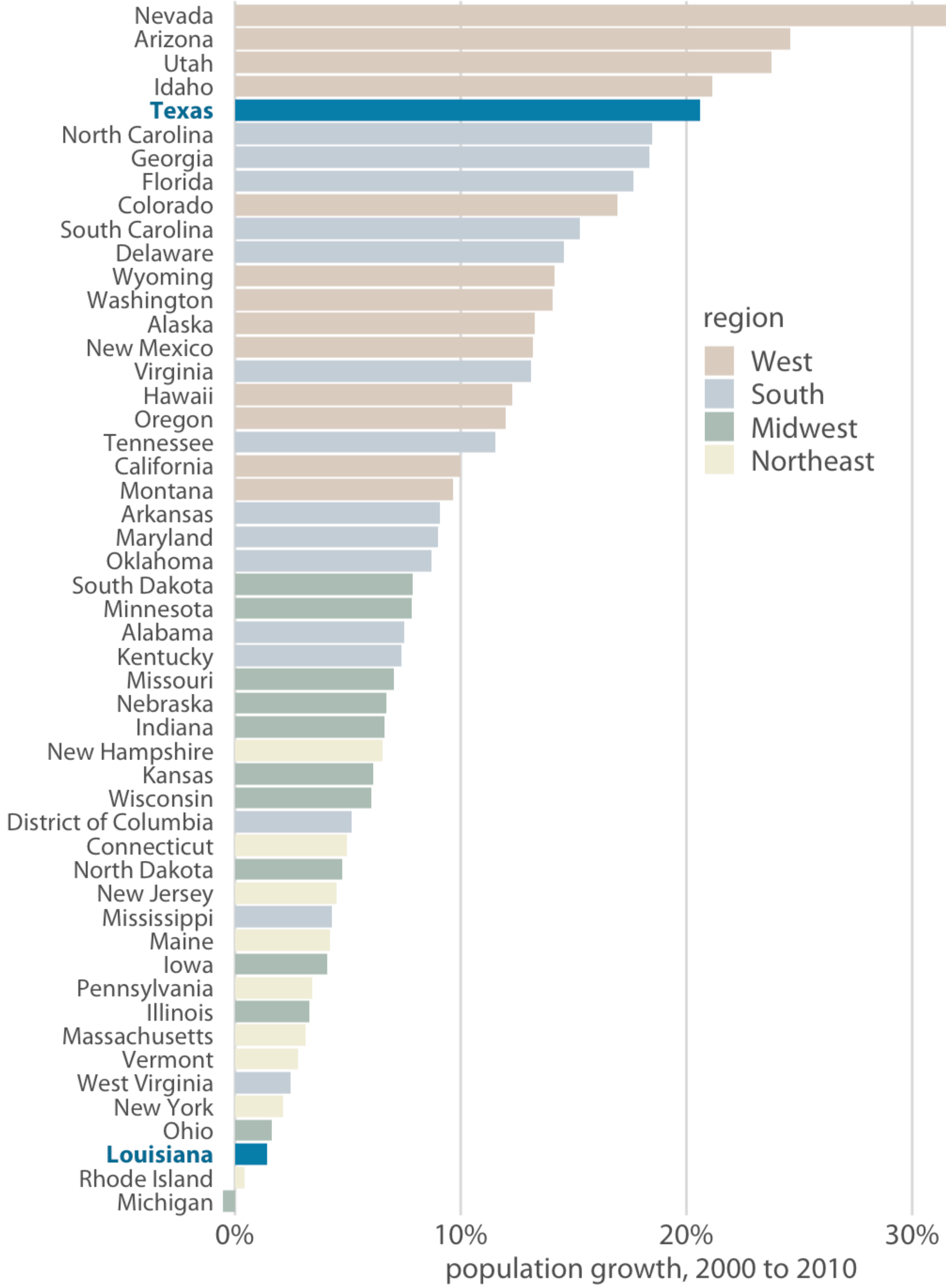


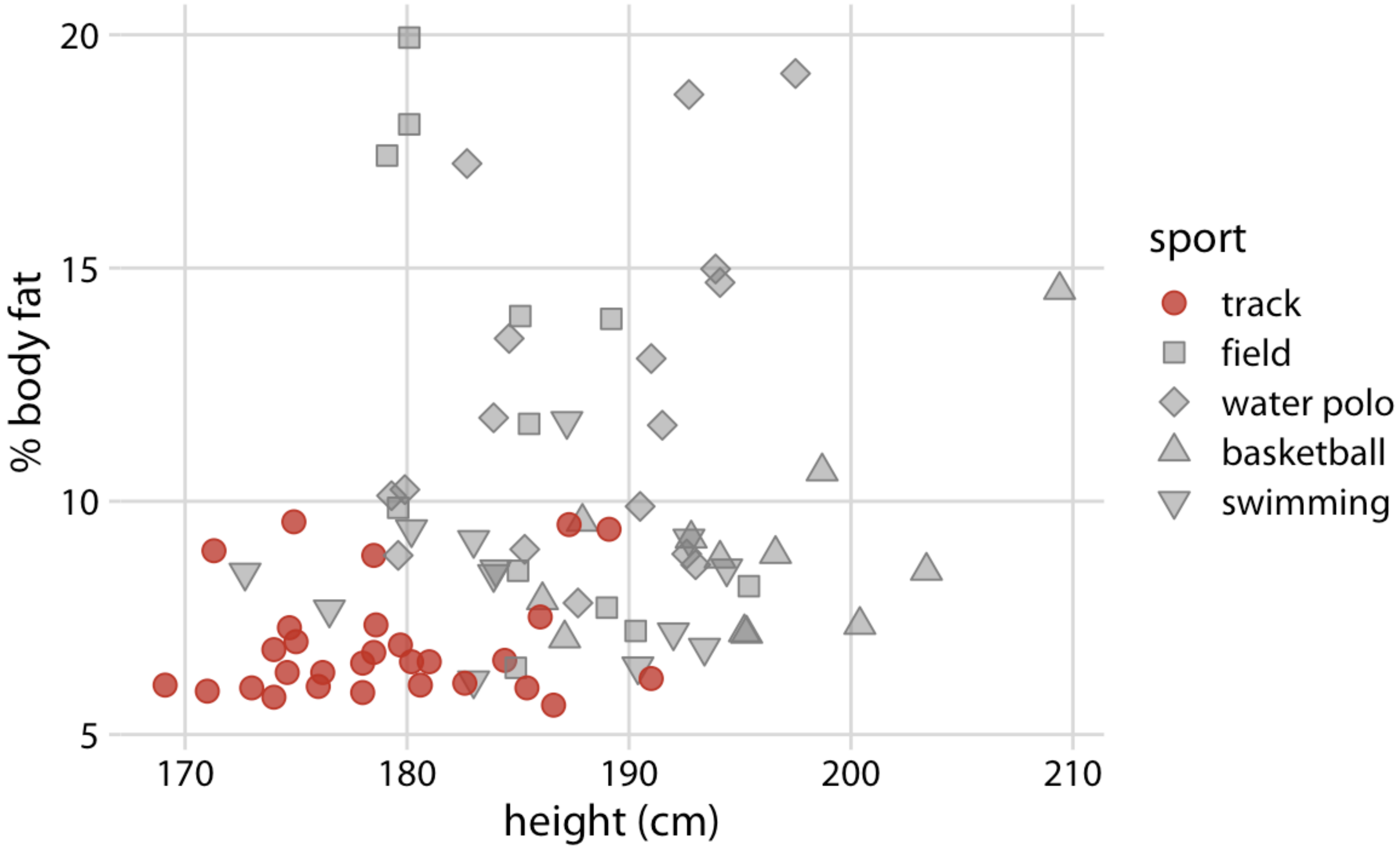
Grays with accents



ColorBrewer Accent




Nevada
Arizona
Utah
Idaho
Texas
North Carolina
Georgia
Florida
Colorado
South Carolina
Delaware
Wyoming
Washington
Alaska
New Mexico
Virginia
Hawaii
Oregon
Tennessee
California
Montana
Arkansas
Maryland
Oklahoma
South Dakota
Minnesota
Alabama
Kentucky
Missouri
Nebraska
Indiana
New Hampshire
Kansas
Wisconsin
District of Columbia
Connecticut
North Dakota
New Jersey
Mississippi
Maine
Iowa
Pennsylvania
Illinois
Massachusetts
Vermont
West Virginia
New York
Ohio
Louisiana
Rhode Island
Michigan
0%
10%
20%
30%
population growth, 2000 to 2010
region
West
South
Midwest
Northeast

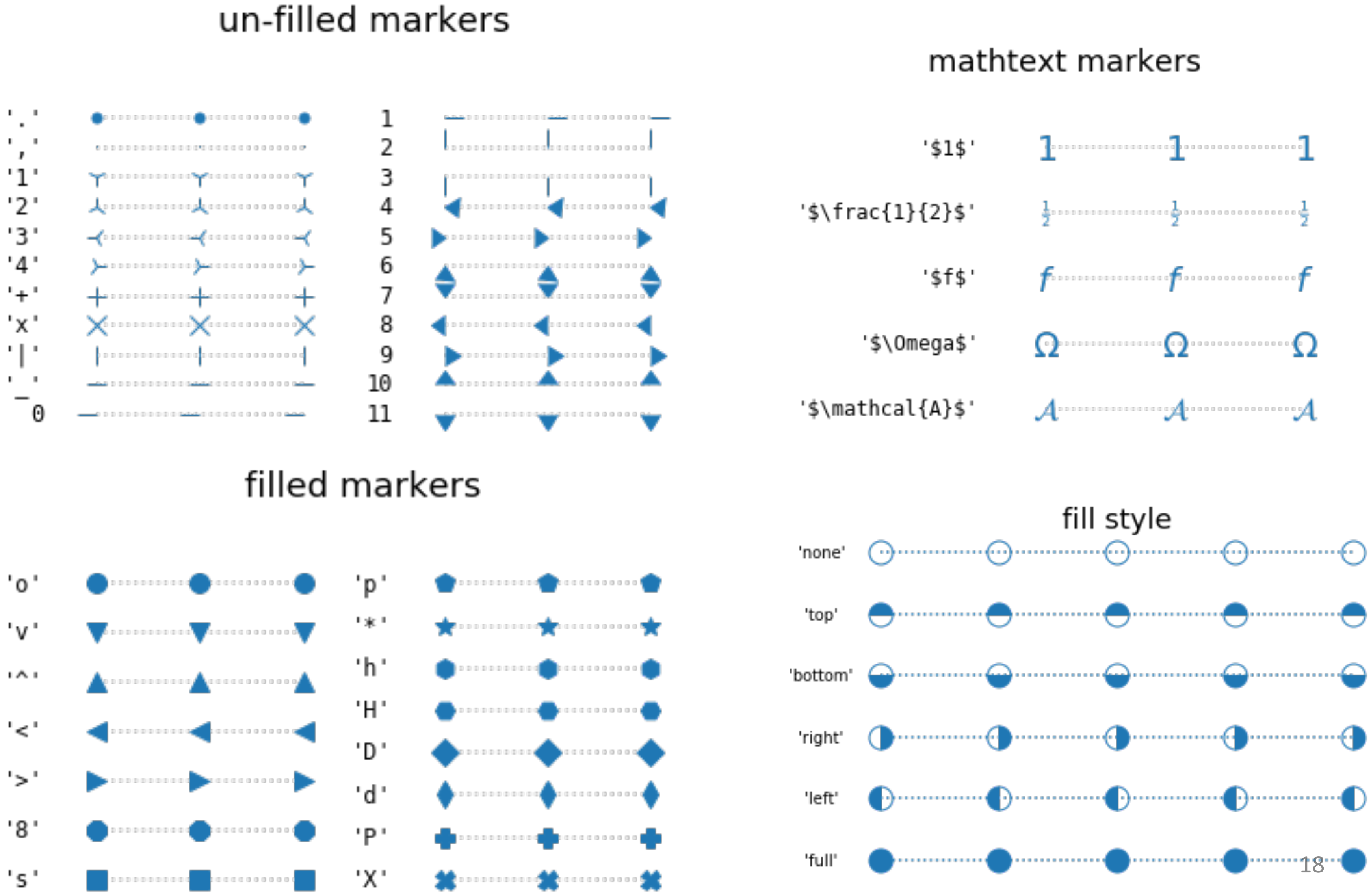

% body fat
20
15
10
5
170
180
190
200
210
height (cm)
sport
track
field
water polo
basketball
swimming

# Color in Python

## Matplotlib Recognizes:

- A single letter string, i.e. one of {'b', 'g', 'r', 'c', 'm', 'y', 'k', 'w'};
- An RGB or RGBA tuple of float values in [0, 1] (e.g. (0.1, 0.2, 0.5) or (0.1, 0.2, 0.5, 0.3)). RGBA is short for Red, Green, Blue, Alpha;
- A hex RGB or RGBA string (e.g., '#0F0F0F' or '#0F0F0F0F’);
- A string representation of a float value in [0, 1] inclusive for gray level (e.g., '0.5’);
- a X11/CSS4 ("html") color name, e.g. "blue";
- A name from the xkcd color survey, prefixed with 'xkcd:' (e.g., 'xkcd:sky blue');
- a "Cn" color spec, i.e. 'C' followed by a number, which is an index into the default property cycle (matplotlib.rcParams['axes.prop_cycle']); the indexing is intended to occur at rendering time, and defaults to black if the cycle does not include color.
- Oneof {'tab:blue’,  'tab:orange', 'tab:green', 'tab:red', 'tab:purple', 'tab:brown','tab:pink', 'tab:gray', 'tab:olive', 'tab:cyan'} which are the Tableau Colors from the 'tab10' categorical palette (which is the default color cycle);

# Markers in python matplotlib



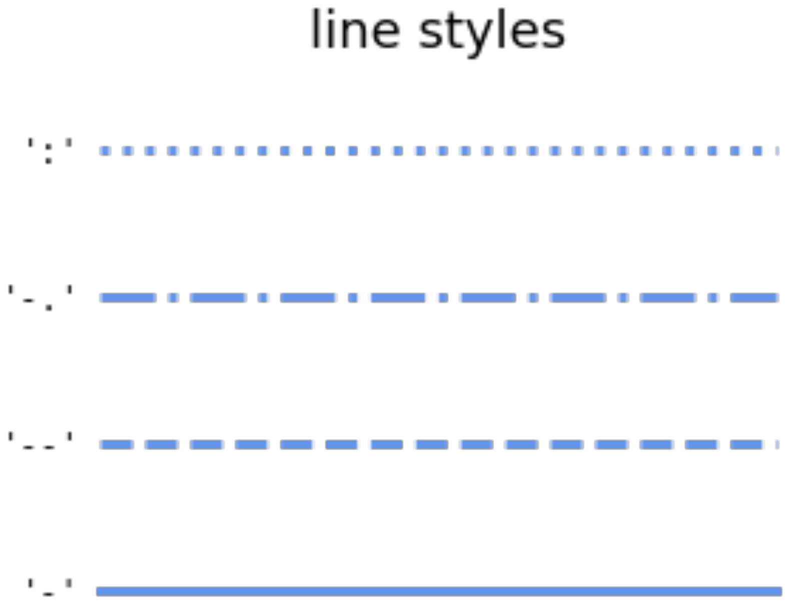
un-filled markers

mathtext markers

filled markers

fill style

# Line Styles in python matplotlib


line styles
':'
'-.'
'--'
'-'