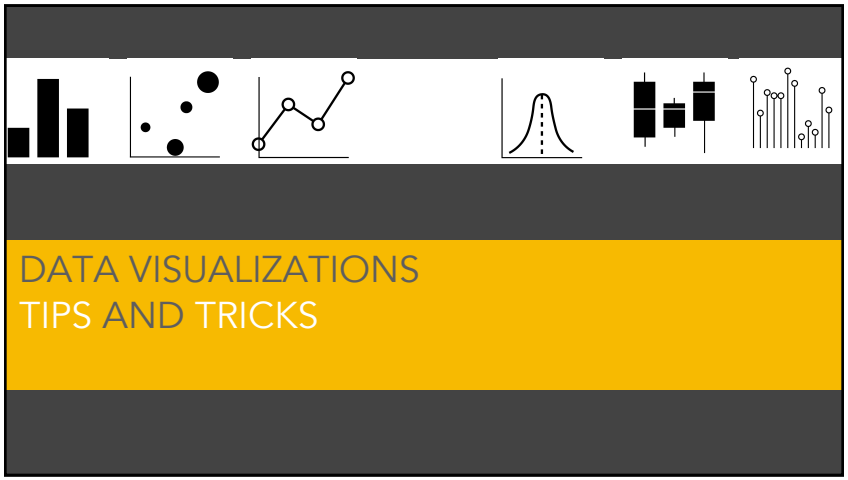


DATA VISUALIZATIONS TIPS AND TRICKS

1 Data Types

Types of Data: Numeric

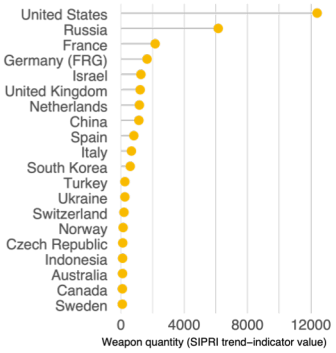
These data have meaning as a measurement, such as a person's height or blood pressure; or a count: how many teeth a dog has (*quantitative data*.)

- DISCRETE data represent items that can be counted; they take on possible values that can be listed out. Coin flips.
- CONTINUOUS data represent measurements; their possible values cannot be counted and can only be described using intervals on the real number line. Amount of gas being pumped

Types of Data: Categorical

Categorical data represent characteristics: hometown or the types of movies they like. Categorical data can take on numerical values (such as "1" indicating right handed and "2" indicating left handed), but those numbers don't have mathematical meaning. You couldn't add them together, for example (*qualitative data* or *Yes/No data*).

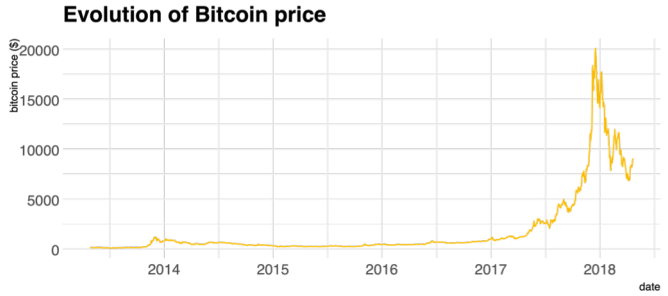
Types of Data:Numerical & Categorical



5

Types of Data: Time Series

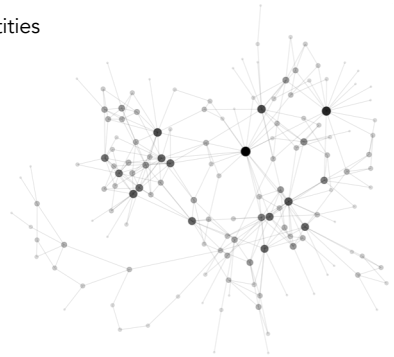
Evolution of one or several numerical variables



6

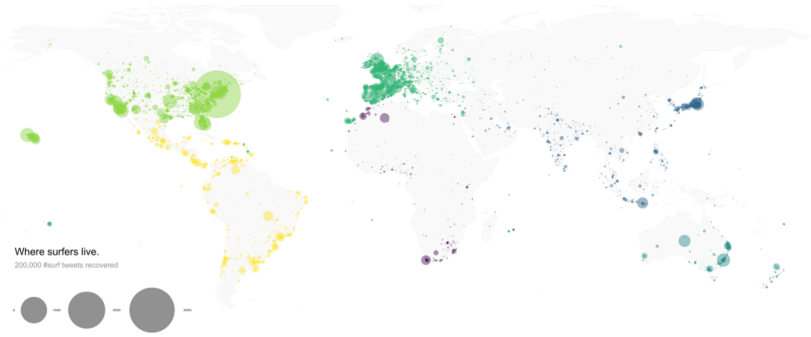
Types of Data: Network

Interconnections between set of entities



7

Types of Data: Map



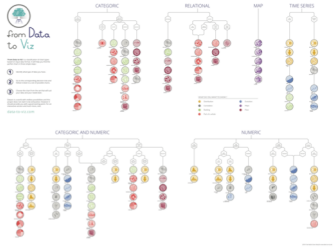
8

Two Visualization Resources

https://serialmentor.com/dataviz/directory-of-visualizations.html

5.1 Amounts

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



9

2 Color Uses

10

Uses of Color

1. Distinguish Groups
2. Represent Data Values
3. Highlighting

The types of colors we use and the way in which we use them are different for these cases

11

Distinguishing Groups

- Distinguishing discrete items or groups that have no order: Use QUALITATIVE color scale

- Qualitative scale: finite set of specific colors

1. Look distinct from each other
2. Being equivalent to each other

12

Distinguishing Groups

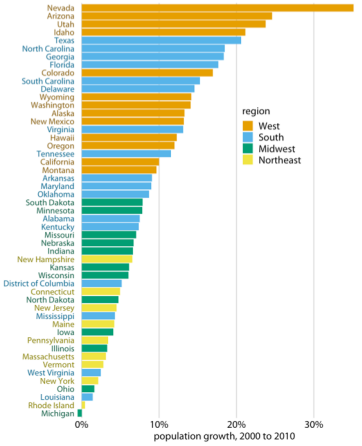
Okabe Ito

ColorBrewer Dark2

ggplot2 hue

13

Distinguishing Groups



14

Represent Data Values

- Color to represent data values (income, temperature, speed): Use SEQUENTIAL color scale

- Sequential scale: sequences of colors that indicates

1. Which values are larger or smaller
2. How distant 2 specific values are from each other (vary uniformly across range)

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). Multi-hue scales tend to follow color gradients that can be seen in the natural world, such as dark red, green, or blue to light yellow, or dark purple to light green. The reverse, e.g. dark yellow to light blue, looks unnatural and doesn't make a useful sequential scale.

15

Represent Data Values

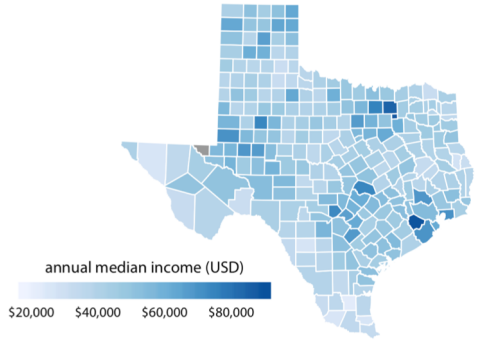
ColorBrewer Blues

Heat

Viridis

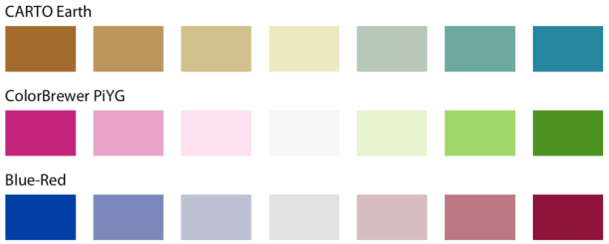
16

Represent Data Values



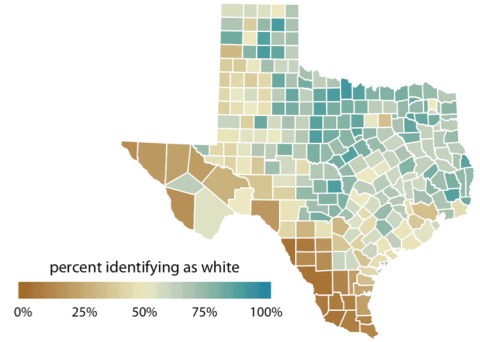
17

Represent Data Values



18

Represent Data Values



19

Color as a Tool to Highlight

- Color can be effective tool to highlight specific elements in data

- May be specific categories or values that have key information and want to strength them by showing their relevant to the reader

- Achieve this by coloring elements in color or set of colors that vividly stand out against rest of the data

20

Color as a Tool to Highlight

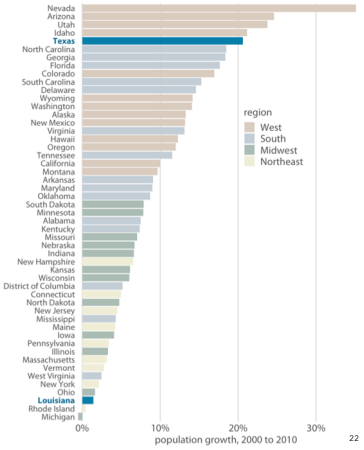
Okabe Ito Accent

Grays with accents

ColorBrewer Accent

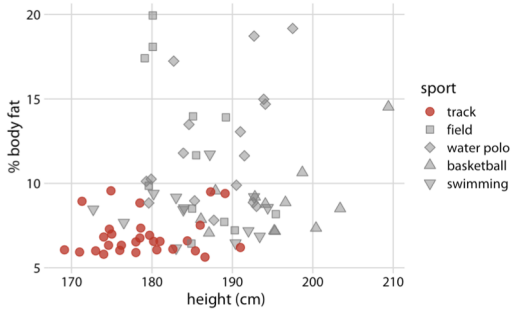
21

Color as a Tool to Highlight



22

Color as a Tool to Highlight

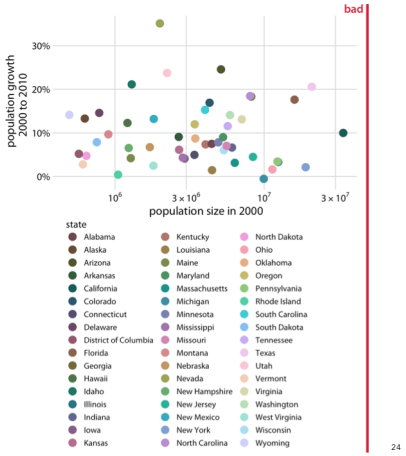


23

Color Pitfalls

Too much or irrelevant information

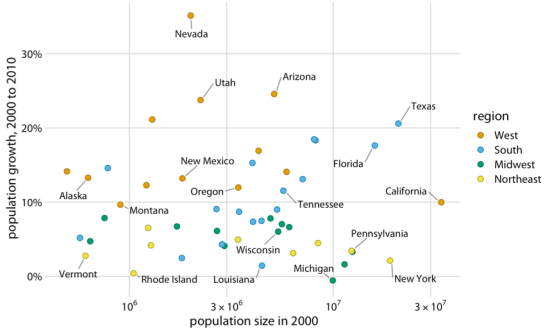
Rule of thumb:
qualitative scales (3-5 categories)



24

Color Pitfalls

Use direct labeling instead of colors when you need to distinguish between more than about eight categorical items.

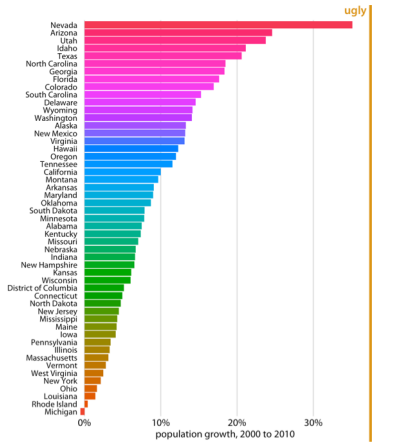


25

Color Pitfalls

Color for the sake of color

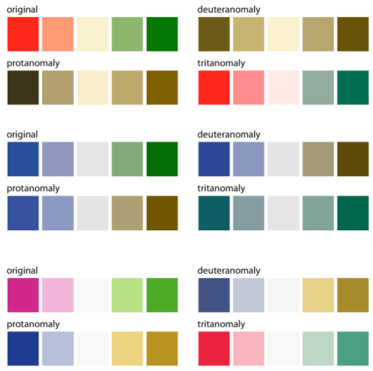
Avoid large filled areas of overly saturated colors. They make it difficult for your reader to carefully inspect your figure.



26

Color Pitfalls

- Color blindness

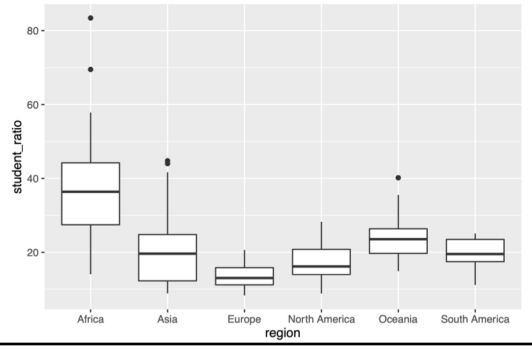


27

3 Evolution of a Plot

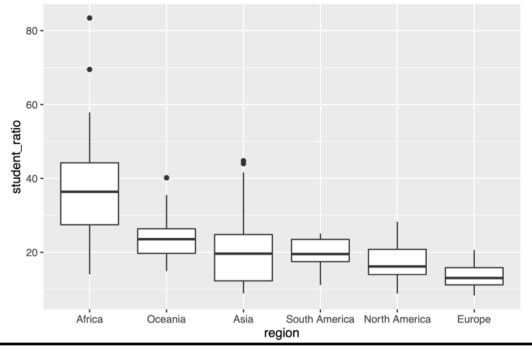
28

Default Boxplot



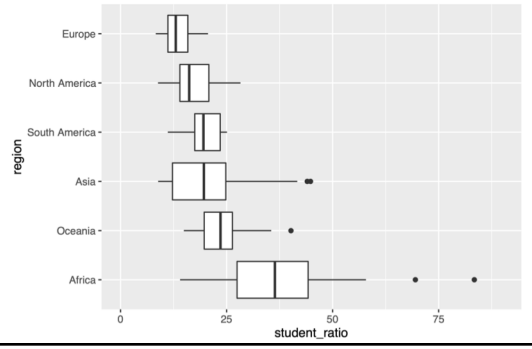
29

Sorted Boxplot



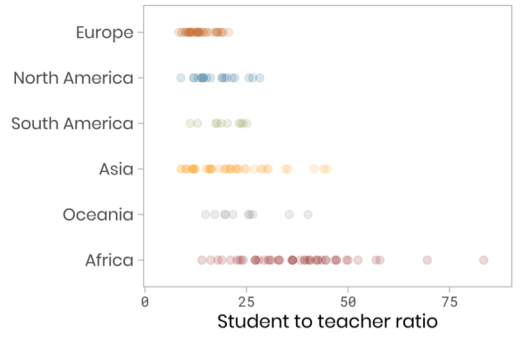
30

Flip/Zero Boxplot



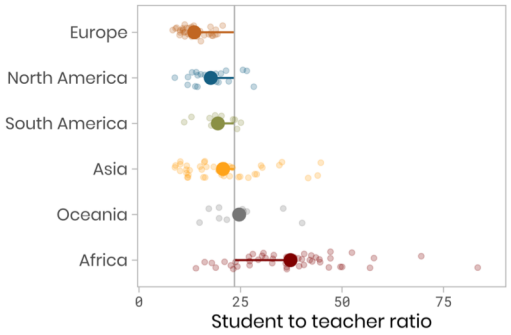
31

Show Data / Aesthetics



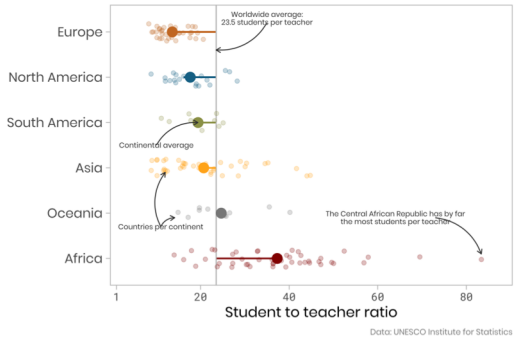
32

Baseline / Mean



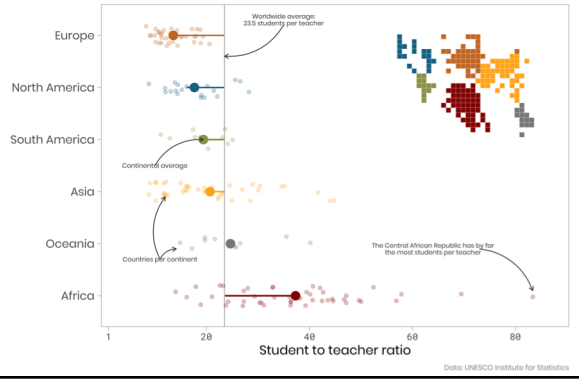
33

Plot Stand Alone



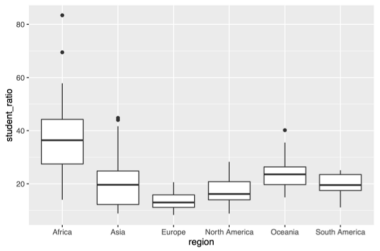
34

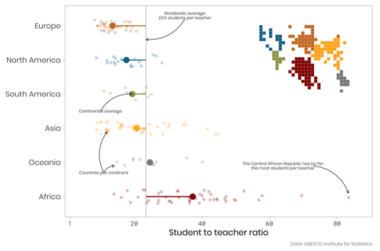
Really Fancy



35

Before / After





36

Data Visualization Take-Aways

- Determine the data type and then figure out plot types that are appropriate for the data type
- Make sure your plot is an accurate representation of the data
- Be strategic with the use of colors
- Often times (most of the time!) less is more