CS 260 – Foundations of Data Science

Lecture 14 - Visualization

10/22/2026

Adam Poliak

Announcements

Logistic Regression Overview

Visualization

Linear Algebra: eigenvectors & eigenvalues

Announcements

HW5 due Tuesday night

Midterm 2 – November 19th

Midterm 1 – go over in lab today

3 important pieces to SGD

- Hypothesis function (prediction)

$$h_{\boldsymbol{w}}(\boldsymbol{x}) = p(y = 1|\boldsymbol{x}) = \frac{1}{1 + e^{-\boldsymbol{w}\cdot\boldsymbol{x}}}$$

- Cost function (want to minimize)

$$J(\boldsymbol{w}) = -\sum_{i=1}^{n} y_i \log h_{\boldsymbol{w}}(\boldsymbol{x_i}) + (1 - y_i) \log(1 - h_{\boldsymbol{w}}(\boldsymbol{x_i}))$$

- Gradient of cost wrt single data point x_i

$$\nabla J_{\boldsymbol{x}_i}(\boldsymbol{w}) = (h_{\boldsymbol{w}}(\boldsymbol{x_i}) - y_i)\boldsymbol{x_i}$$

Cost function as Cross Entropy

probability distribution

$$J(\vec{w}) = -[y \log(h_{\vec{w}}(x)) + (1-y) \log(1 - h_{\vec{w}}(x))]$$

probability distribution

entropy

$$H(Y) = -\sum_{y \in vals(Y)} p(y) \log p(y)$$

Cross entropy

$$H(p,q) = -\sum_{y \in vals(Y)} p(y) \log q(y)$$

Ugly, bad, wrong visualizations

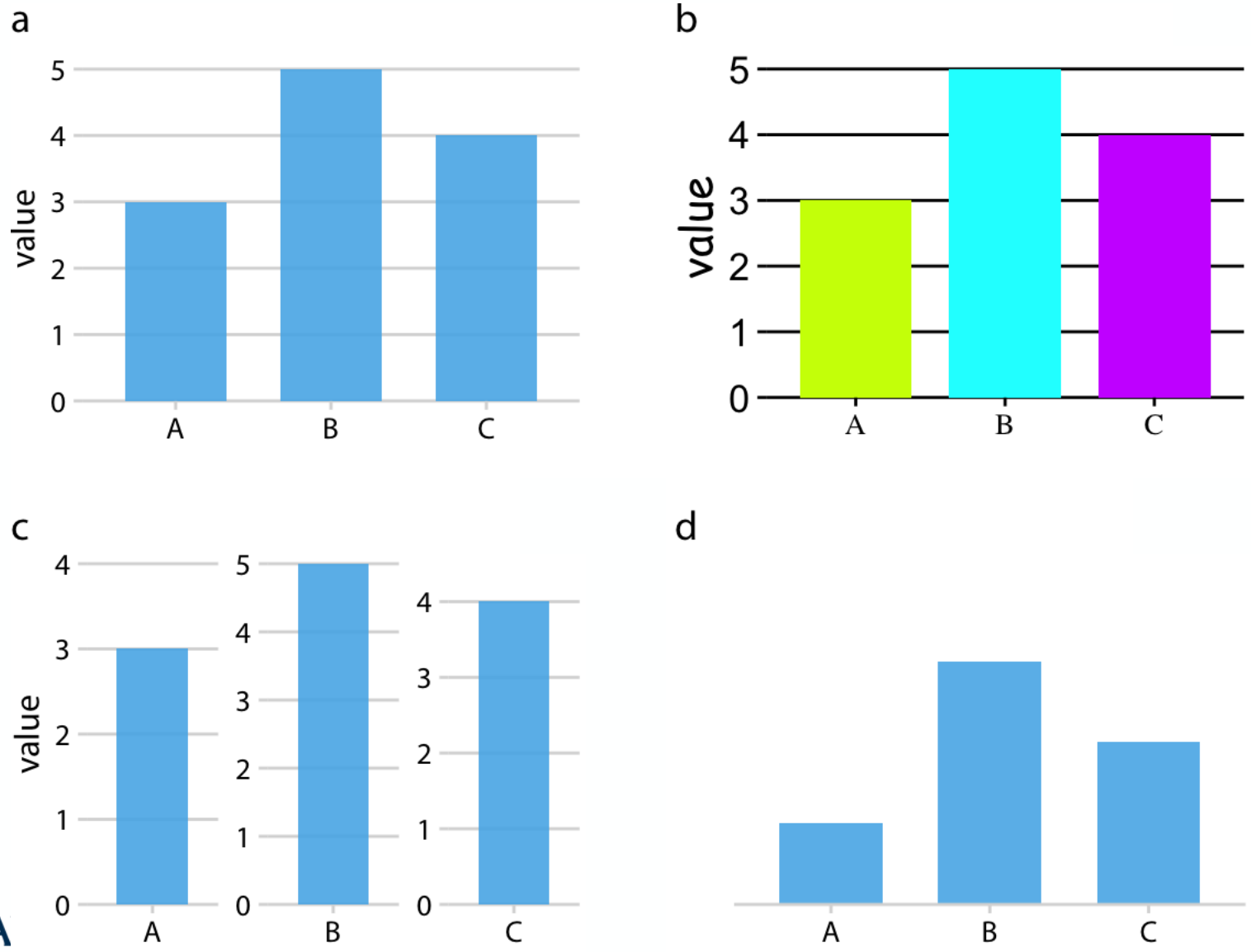
- **ugly**—A figure that has aesthetic problems but otherwise is clear and informative.

- **bad**—A figure that has problems related to perception; it may be unclear, confusing, overly complicated, or deceiving.

- **wrong**—A figure that has problems related to mathematics; it is objectively incorrect.

From “Fundamentals of Data Visualization” by Claus Wilke

Ugly, bad, wrong visualizations



BRYN MA COLLEGE

Fig 1.1 from “Fundamentals of Data Visualization” by Claus Wilke

Ugly, bad, wrong visualizations

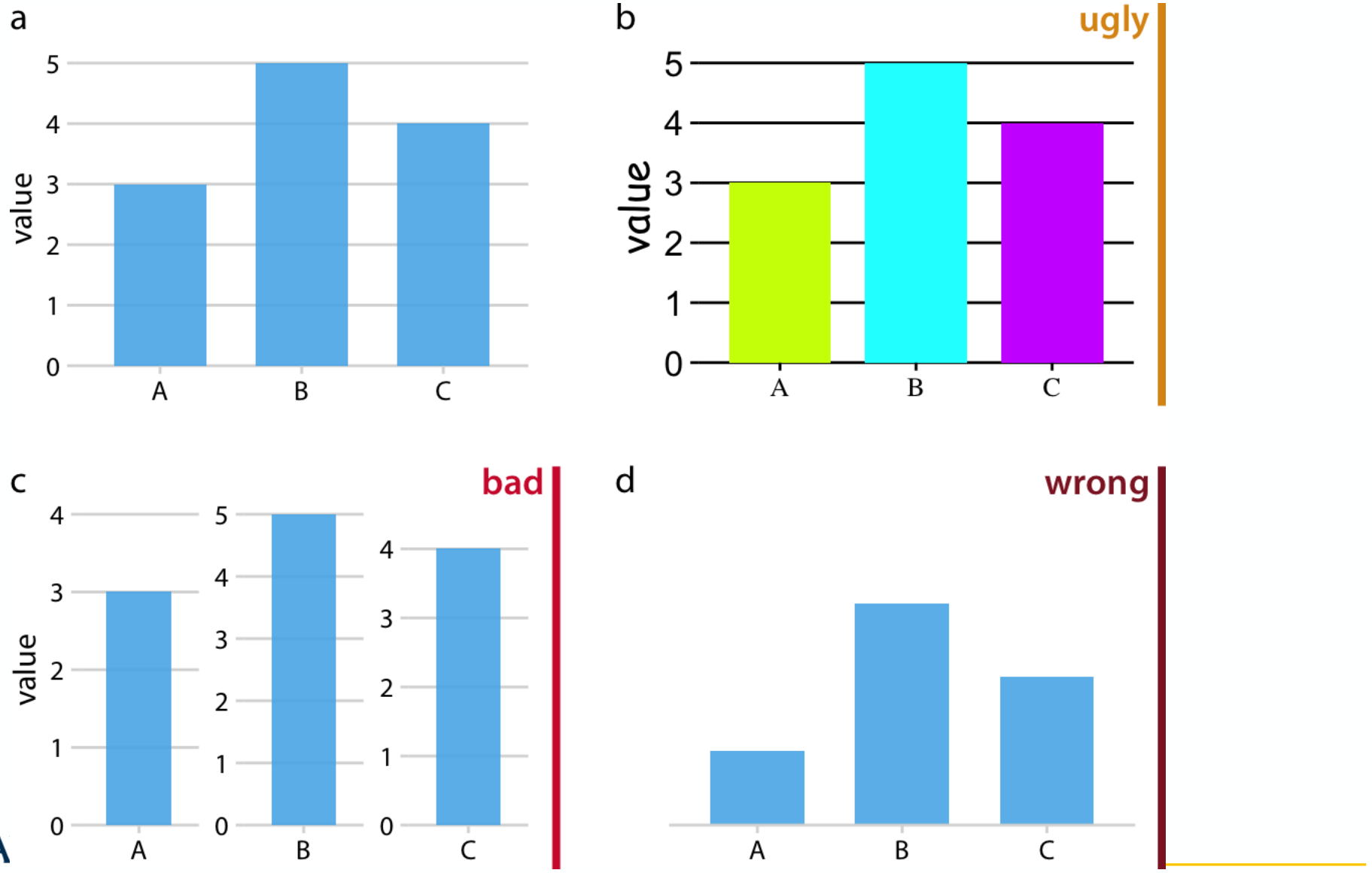


Fig 1.1 from “Fundamentals of Data Visualization” by Claus Wilke

Data Types: continuous vs. discrete

Table 2.1: Types of variables encountered in typical data visualization scenarios.

Type of variable	Examples	Appropriate scale	Description
quantitative/numerical continuous	1.3, 5.7, 83, $1.5x10^{-2}$	continuous	Arbitrary numerical values. These can be integers, rational numbers, or real numbers.
quantitative/numerical discrete	1, 2, 3, 4	discrete	Numbers in discrete units. These are most commonly but not necessarily integers. For example, the numbers 0.5, 1.0, 1.5 could also be treated as discrete if intermediate values cannot exist in the given dataset.
qualitative/categorical unordered	dog, cat, fish	discrete	Categories without order. These are discrete and unique categories that have no inherent order. These variables are also called *factors*.
qualitative/categorical ordered	good, fair, poor	discrete	Categories with order. These are discrete and unique categories with an order. For example, “fair” always lies between “good” and “poor”. These variables are also called *ordered factors*.
date or time	Jan. 5 2018, 8:03am	continuous or discrete	Specific days and/or times. Also generic dates, such as July 4 or Dec. 25 (without year).
text	The quick brown fox jumps over the lazy dog.	none, or discrete	Free-form text. Can be treated as categorical if needed.

Table 2.1 from “Fundamentals of Data Visualization” by Claus Wilke

Aesthetics in Data Visualization

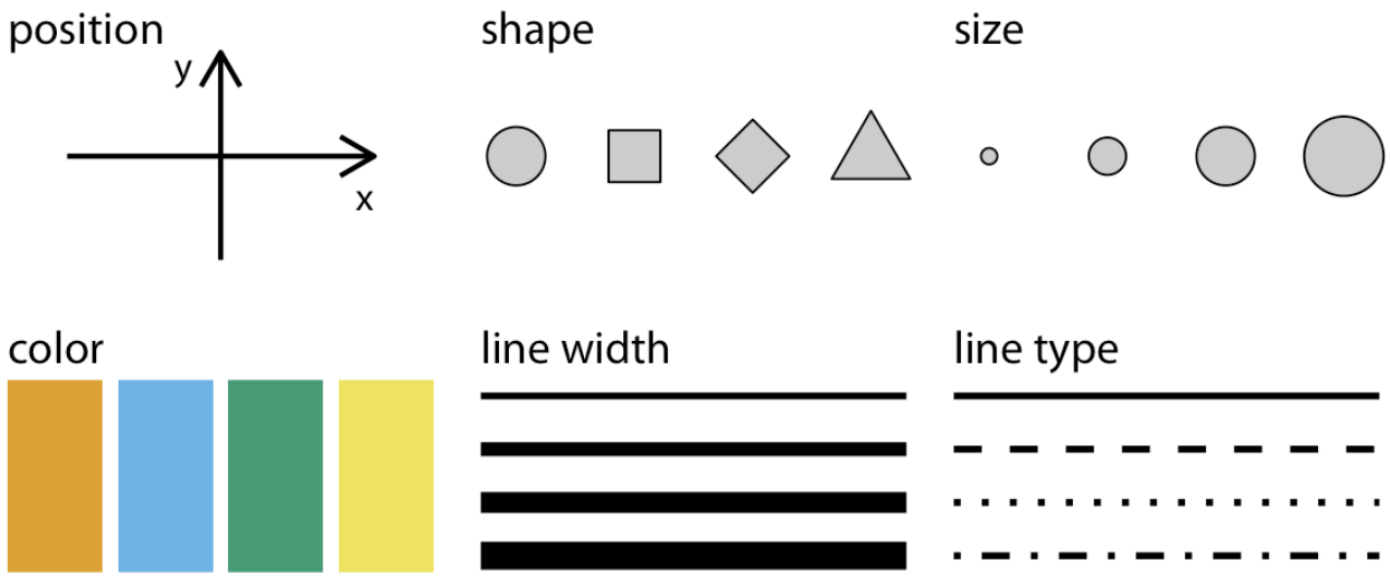
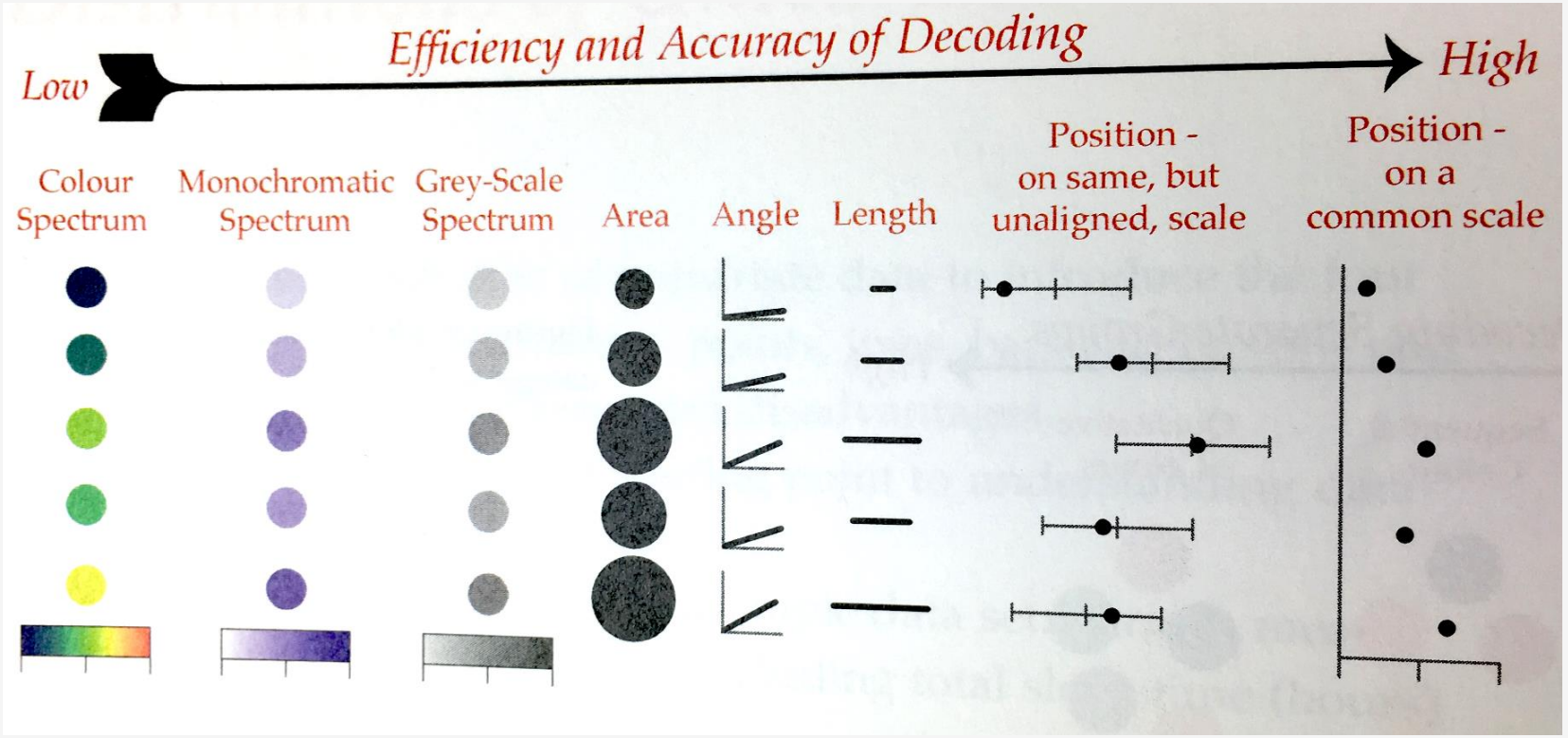


Figure 2.1: Commonly used aesthetics in data visualization: position, shape, size, color, line width, line type. Some of these aesthetics can represent both continuous and discrete data (position, size, line width, color) while others can usually only represent discrete data (shape, line type).

Fig 2.1 from “Fundamentals of Data Visualization” by Claus Wilke

Data Types

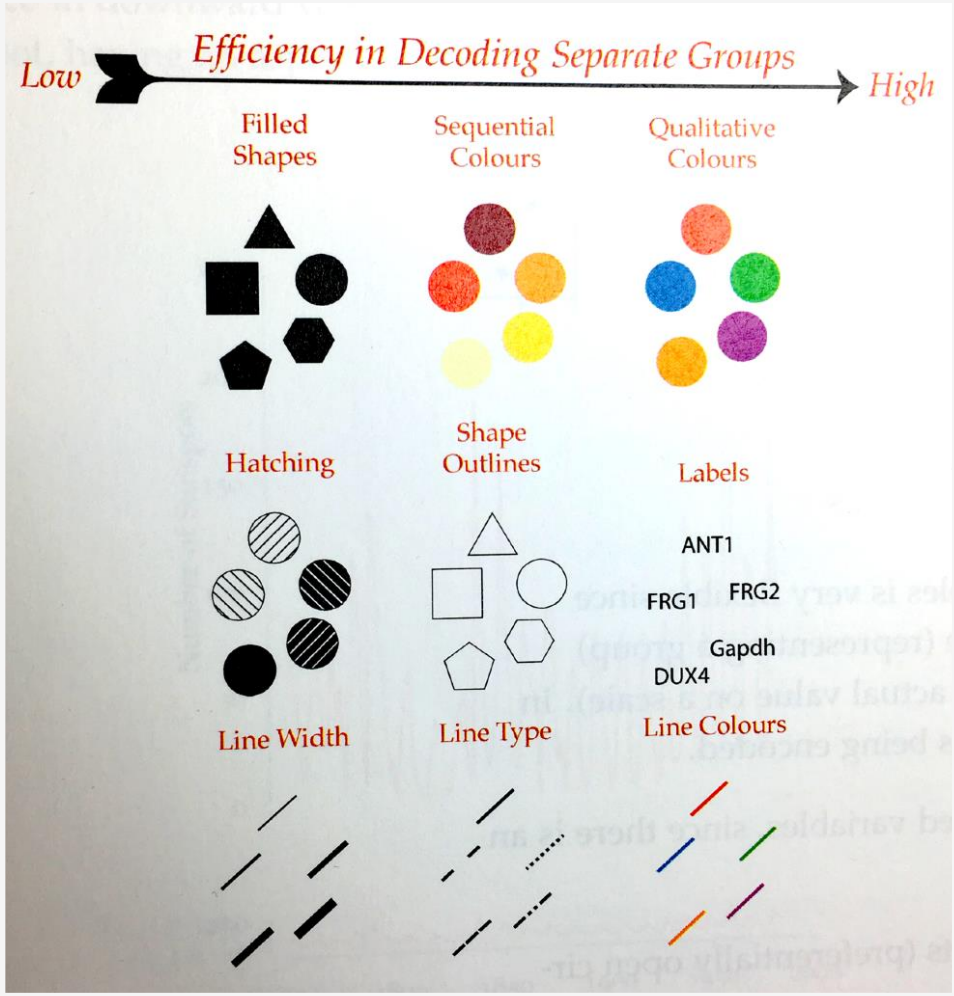
Continuous



Slide from: Eric Miller

Data Types

Discrete



Slide from: Eric Miller

Two different visualizations of the same data

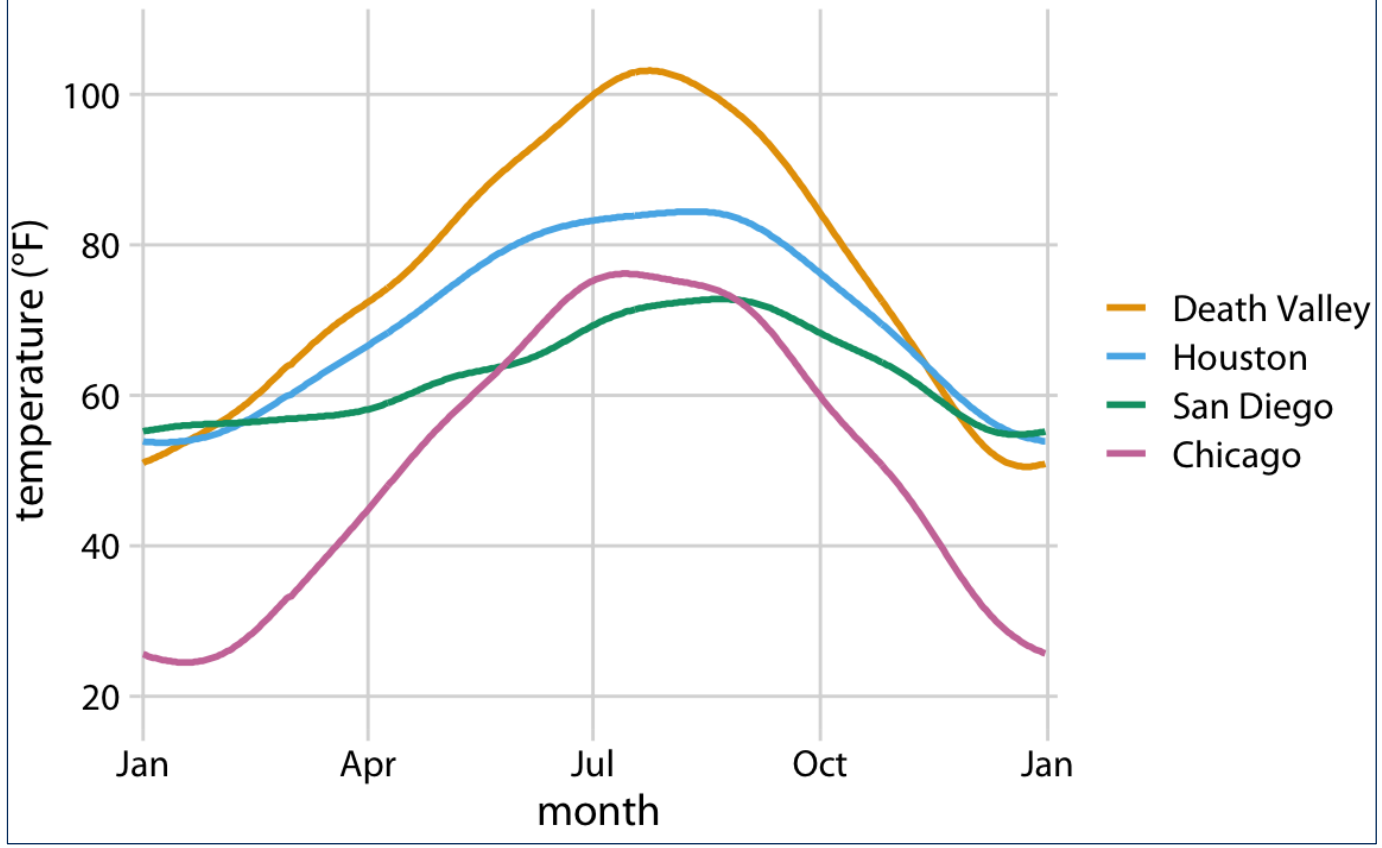
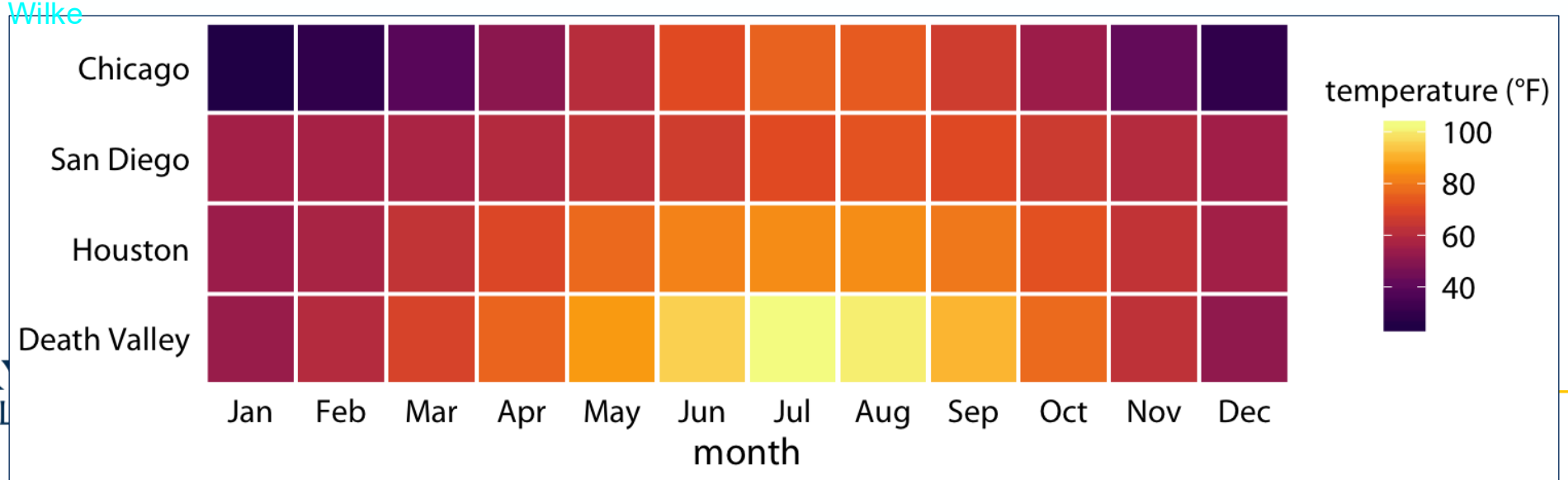


Fig 2.3/2.4 from "Fundamentals of Data Visualization" by Claus Wilke



All the same data – what do we want to convey?

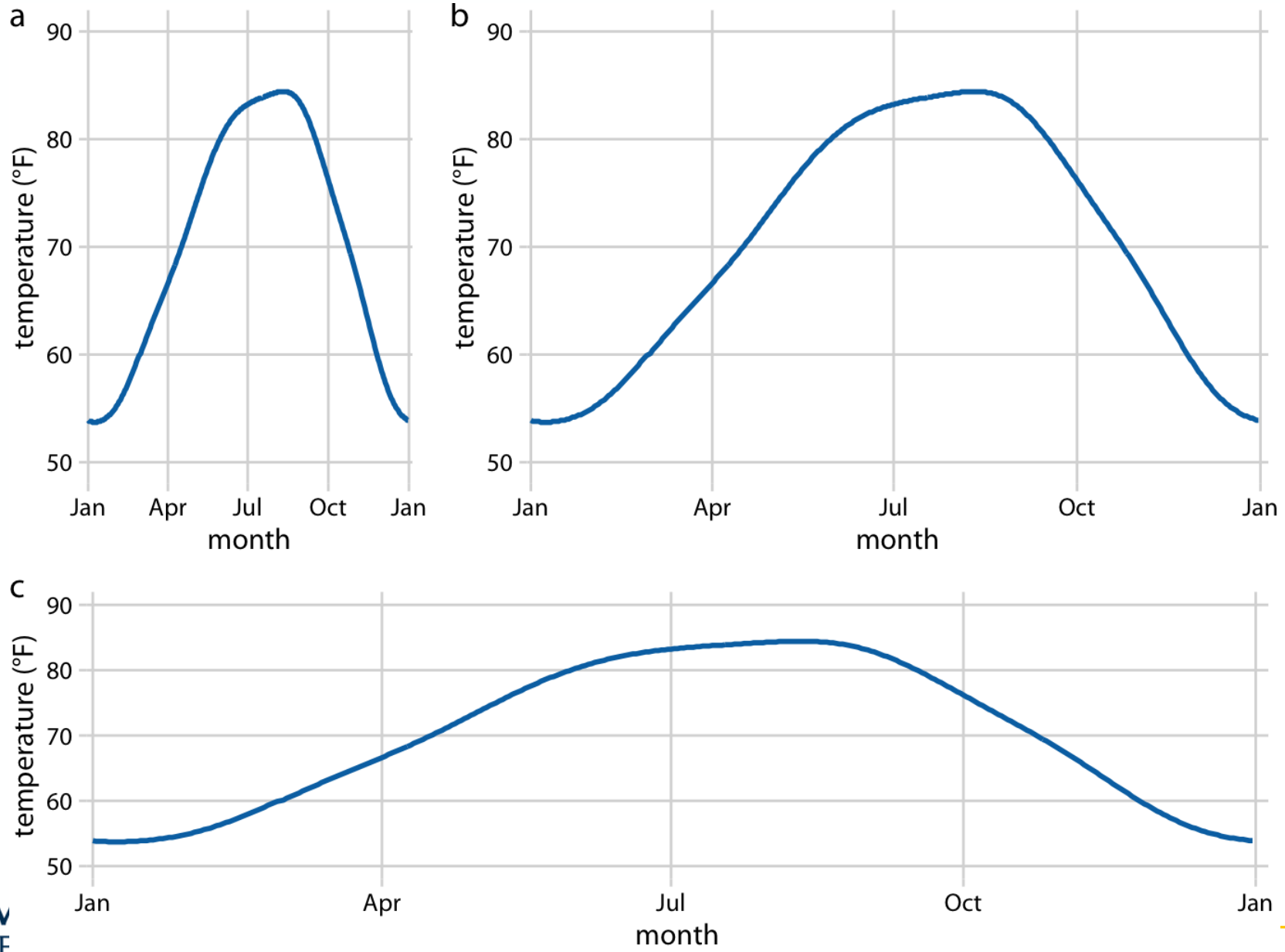


Fig 3.2 from “Fundamentals of Data Visualization” by Claus Wilke

Color

Why use color?

Which colors?

Slide from: Eric Miller

Color: qualitative

Fig 4.1 from “Fundamentals of Data Visualization” by Claus Wilke

Qualitative example

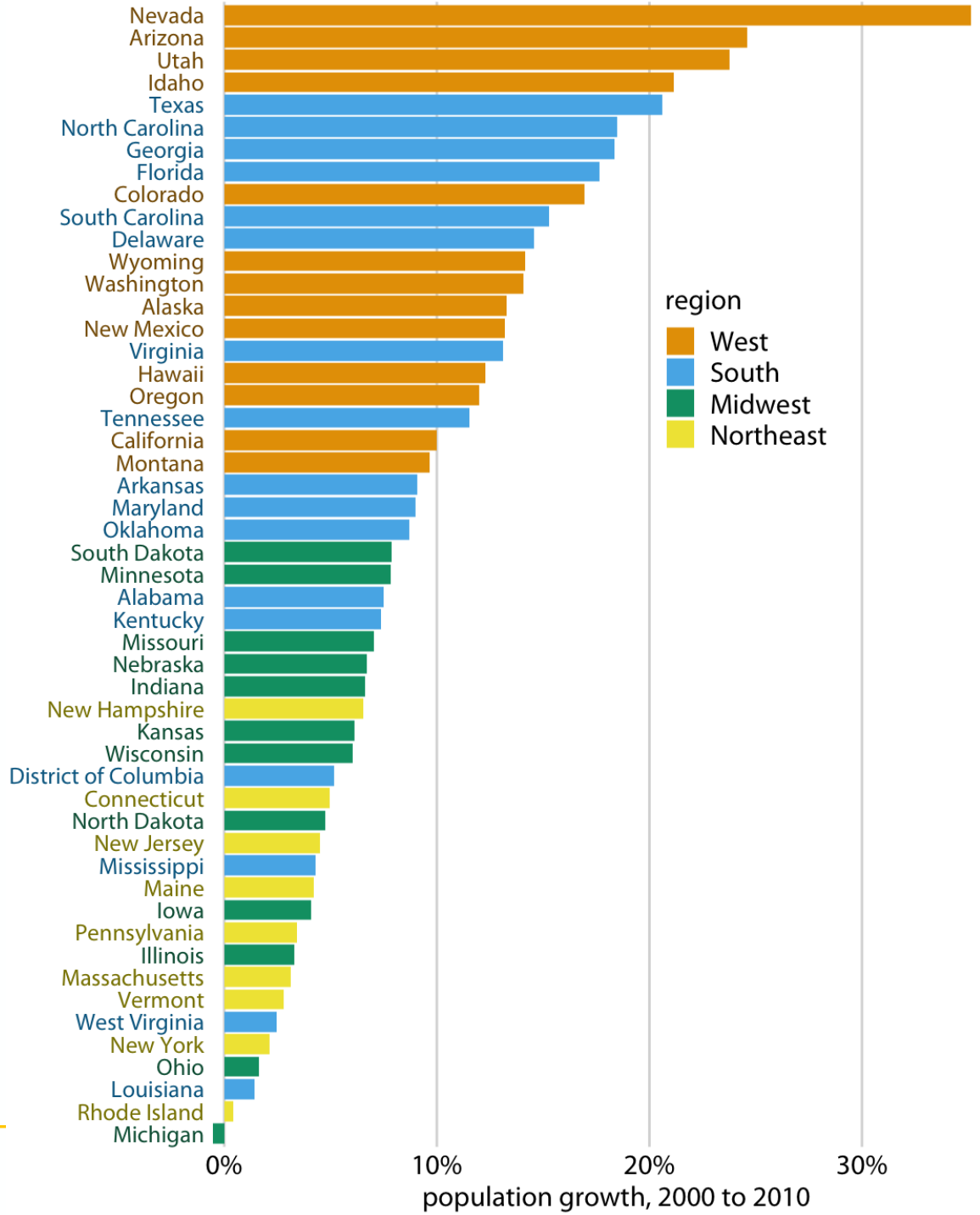


Fig 4.2 from "Fundamentals of Data Visualization" by Claus Wilke

Color: sequential

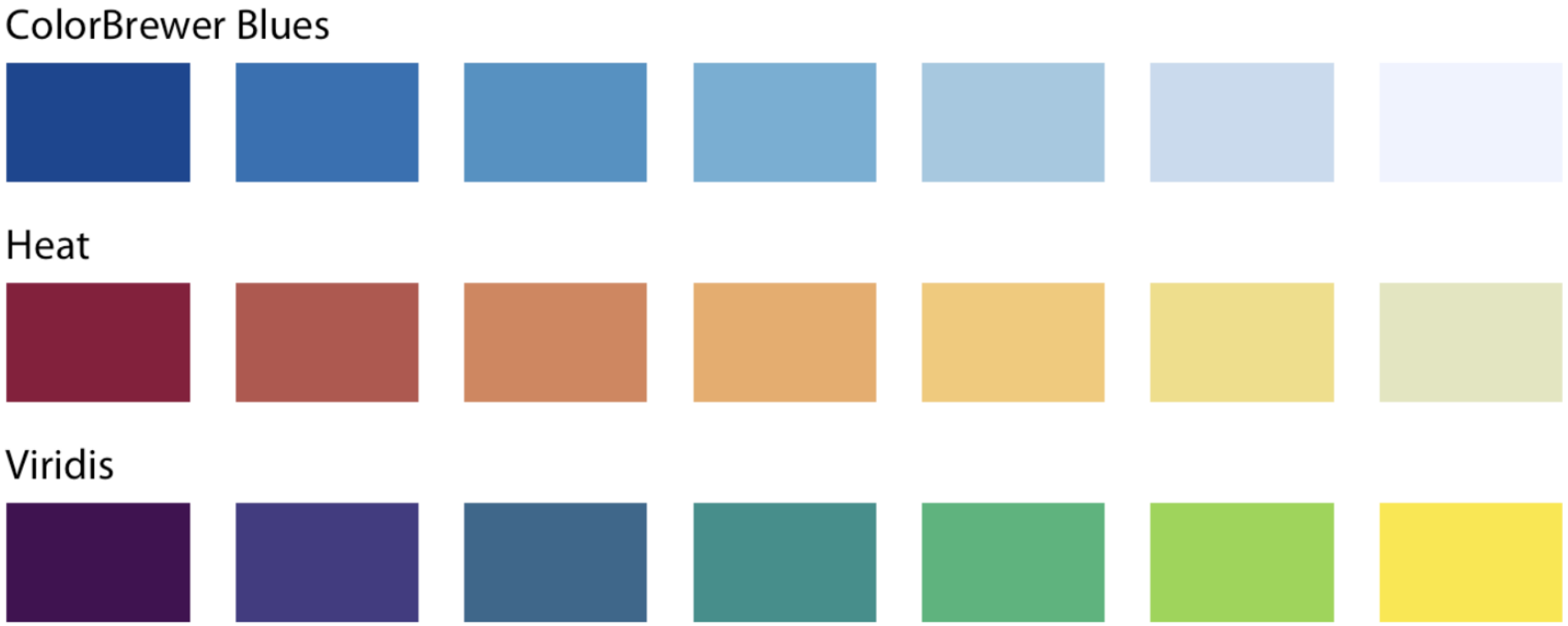
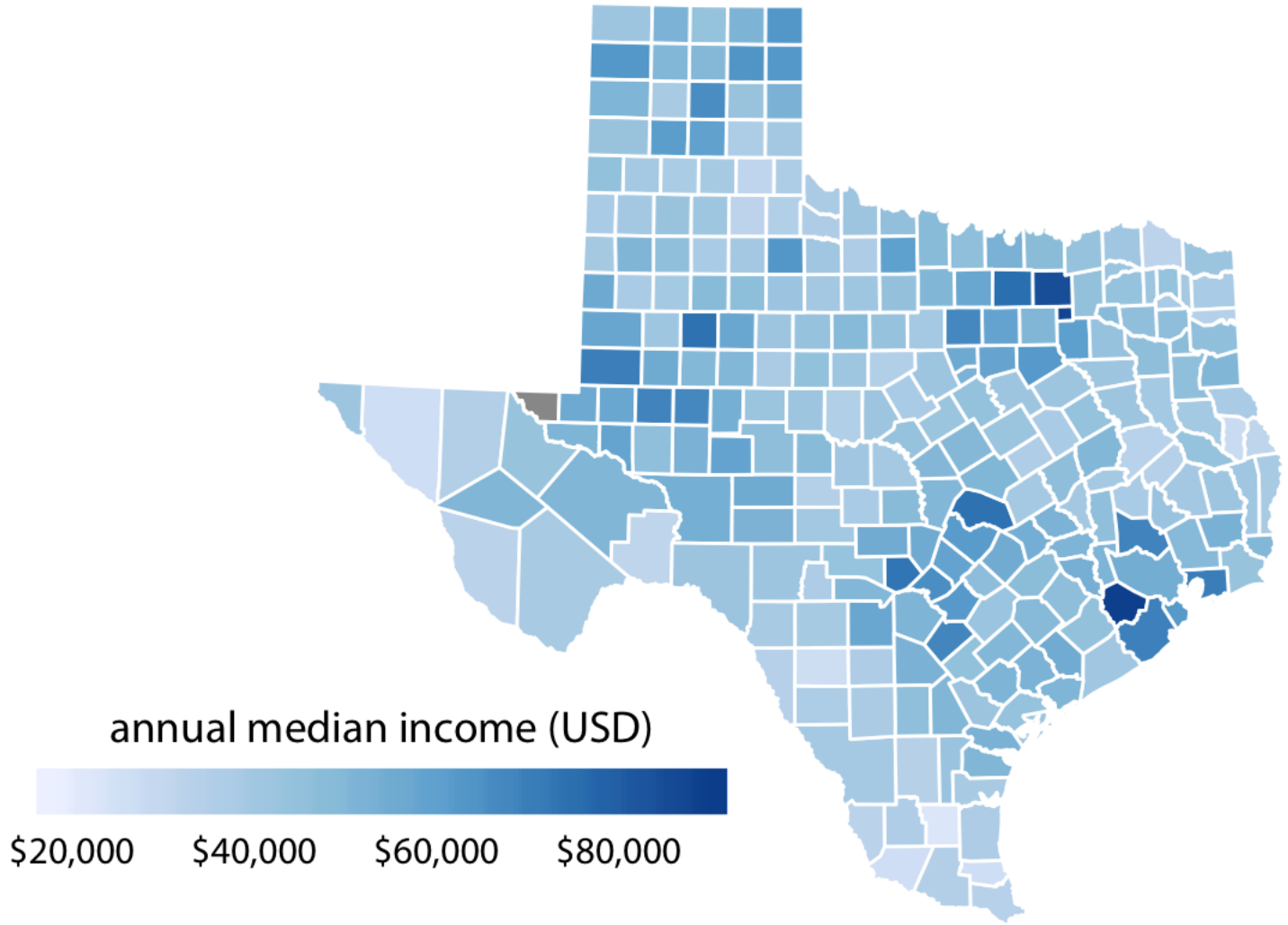


Fig 4.3 from “Fundamentals of Data Visualization” by Claus Wilke

Sequential example



BRYN MA COLLEGE

Fig 4.4 from “Fundamentals of Data Visualization” by Claus Wilke

Color: diverging

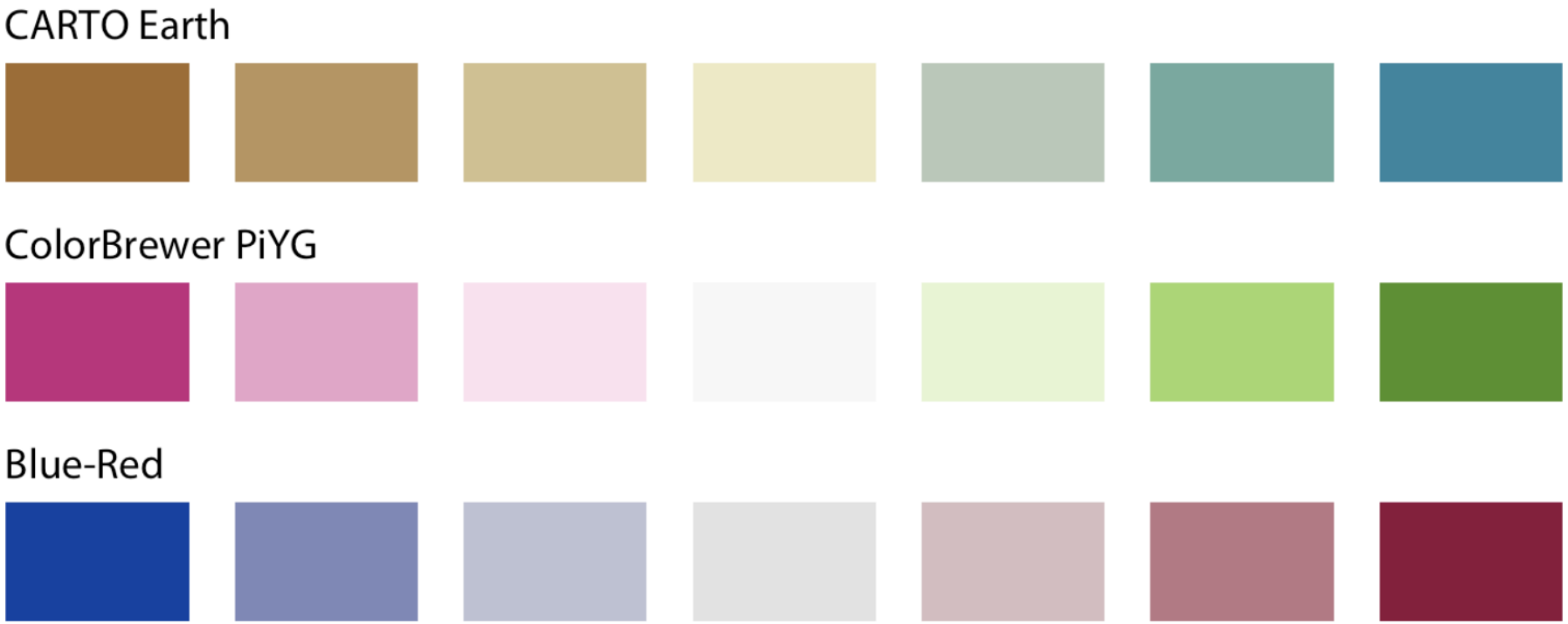


Fig 4.5 from “Fundamentals of Data Visualization” by Claus Wilke

Diverging example

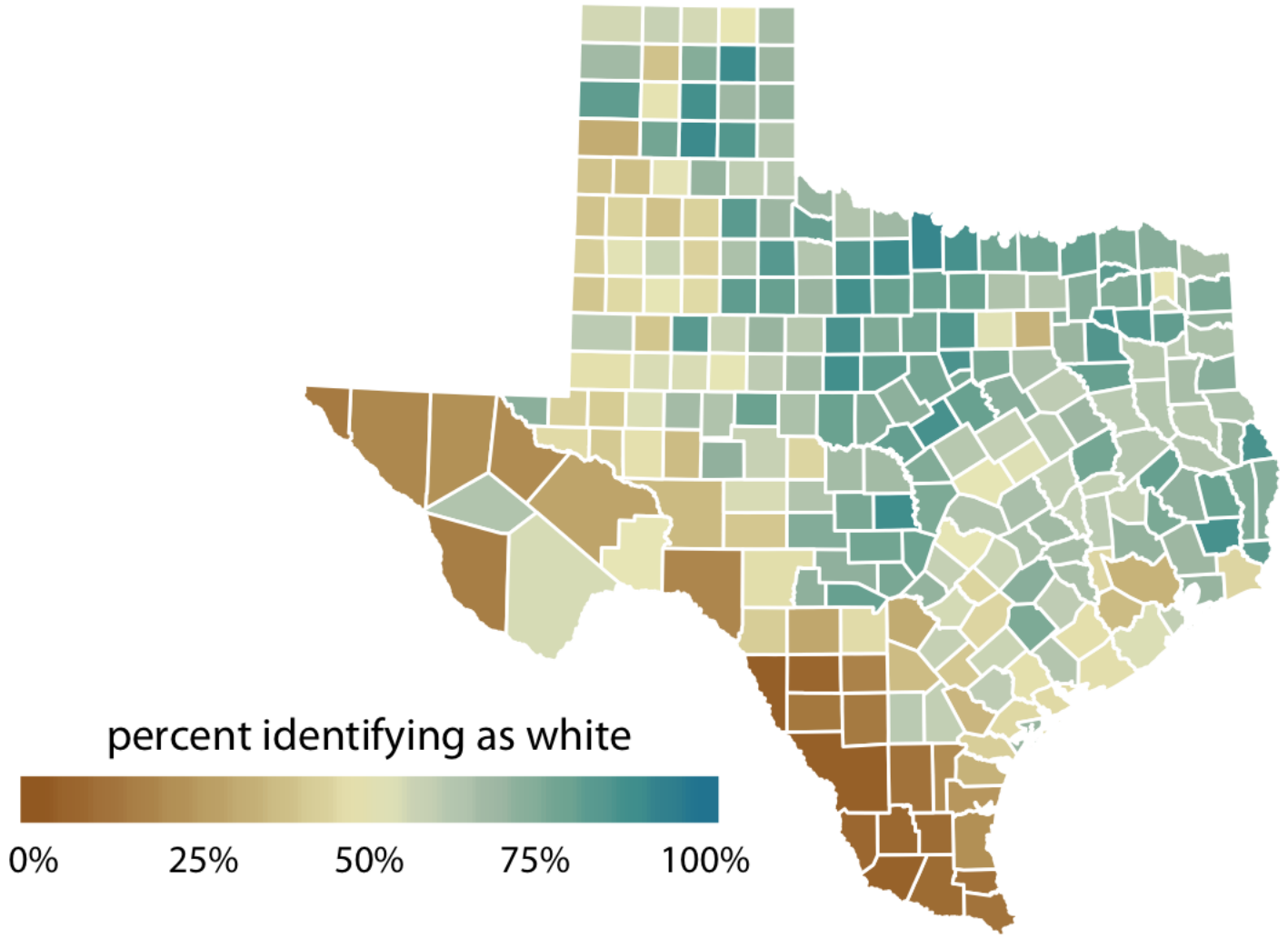
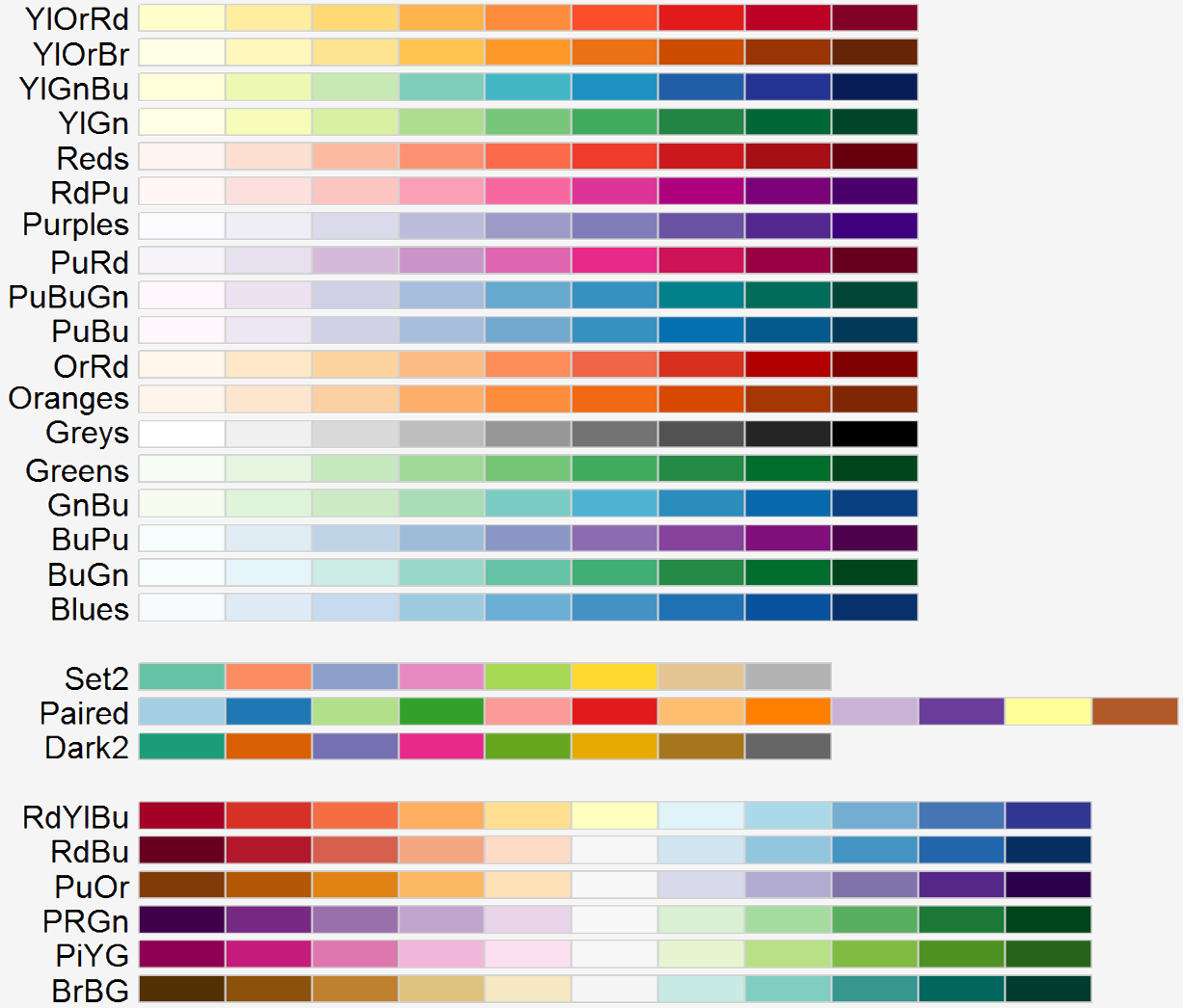


Fig 4.6 from “Fundamentals of Data Visualization” by Claus Wilke

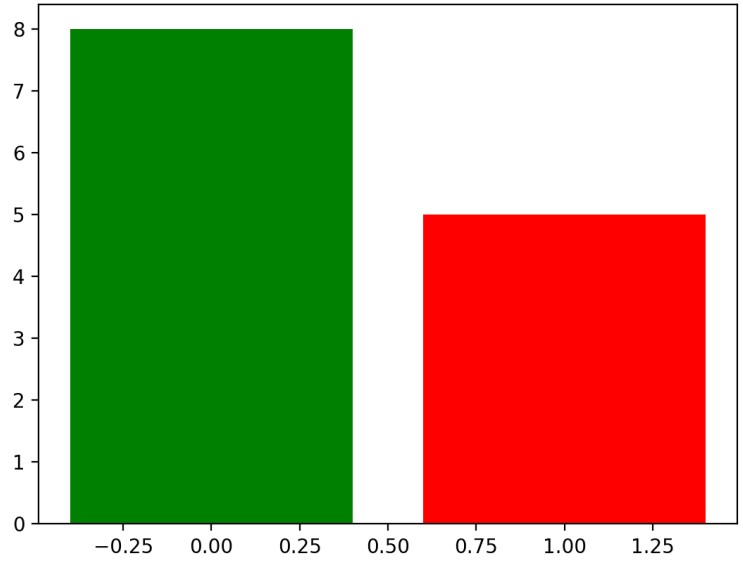
Color

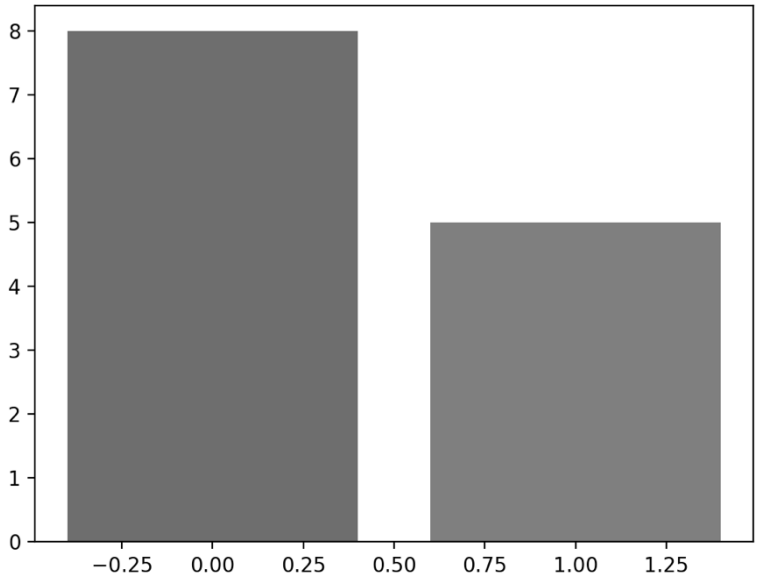
Color-blind friendly



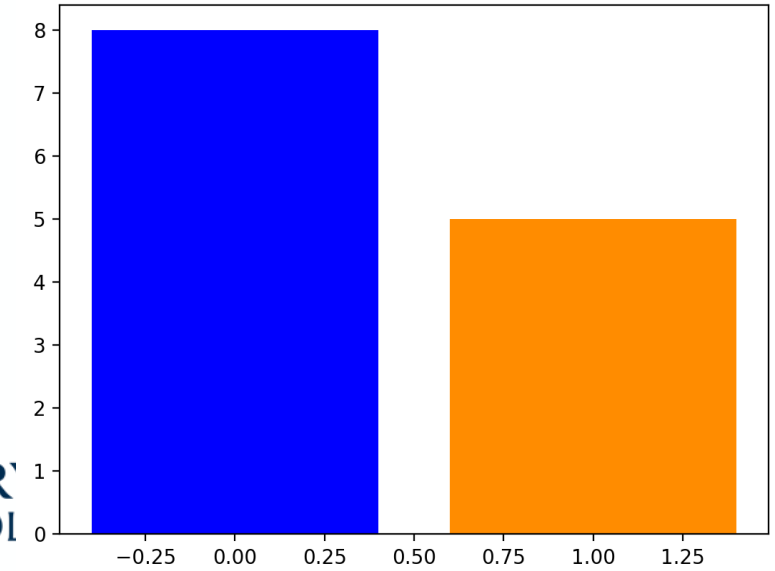
Slide from: Eric Miller

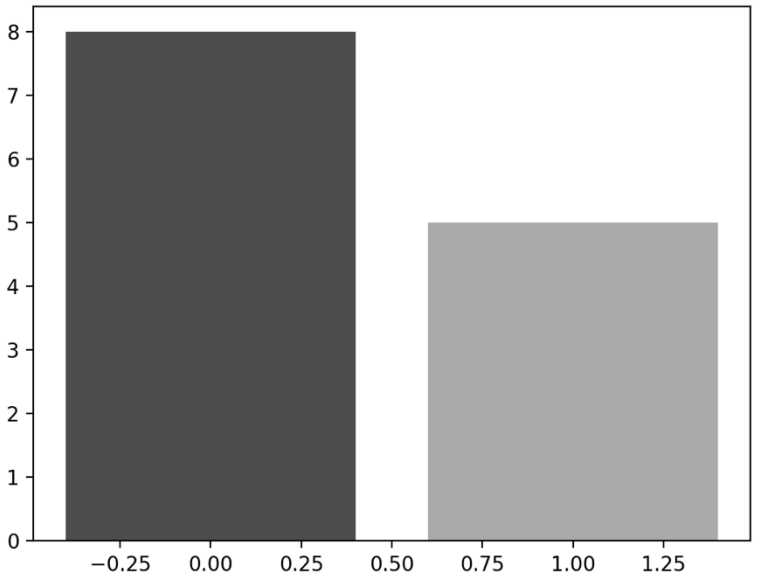
Red/green vs. blue/orange



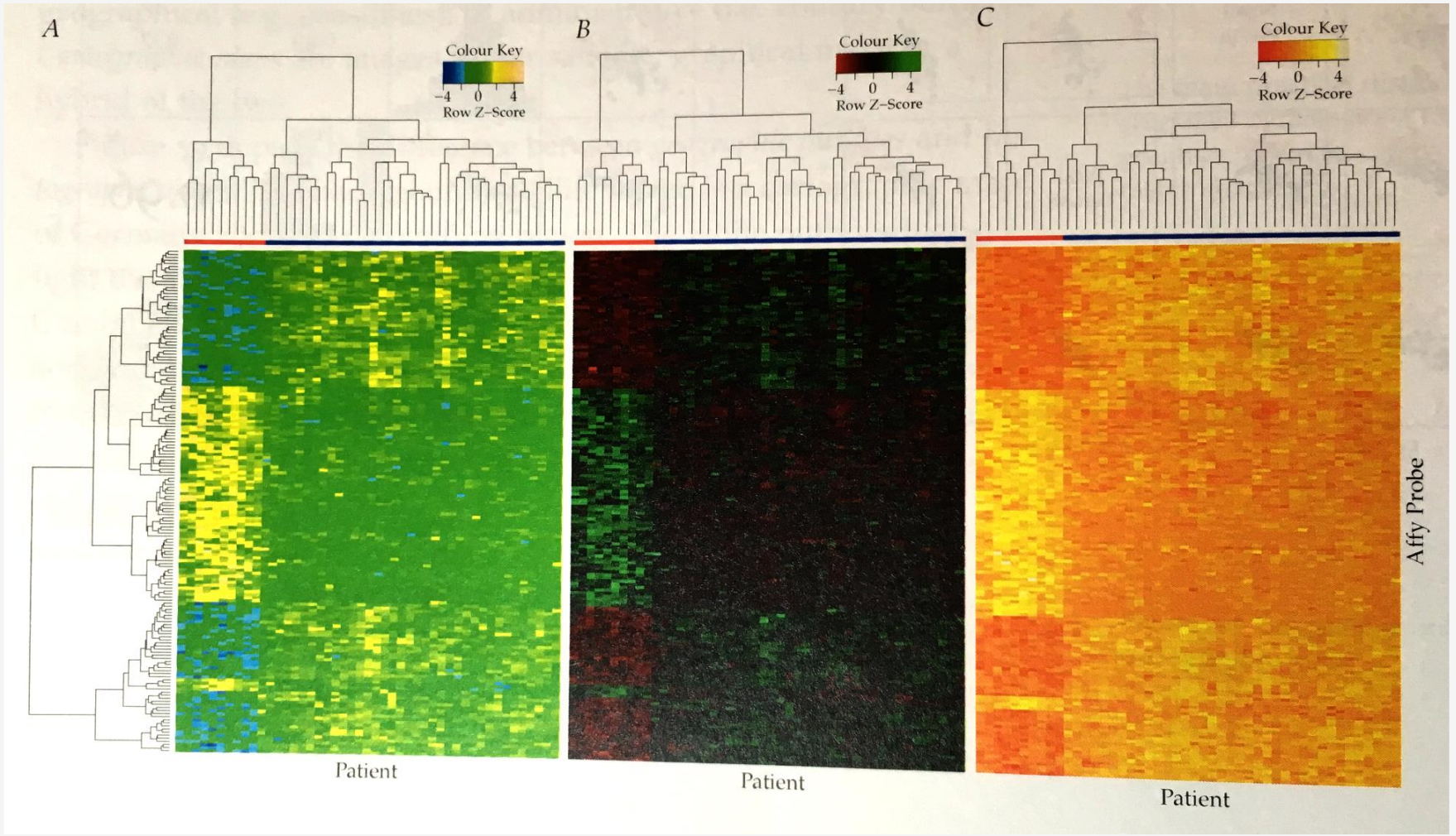


To black and white



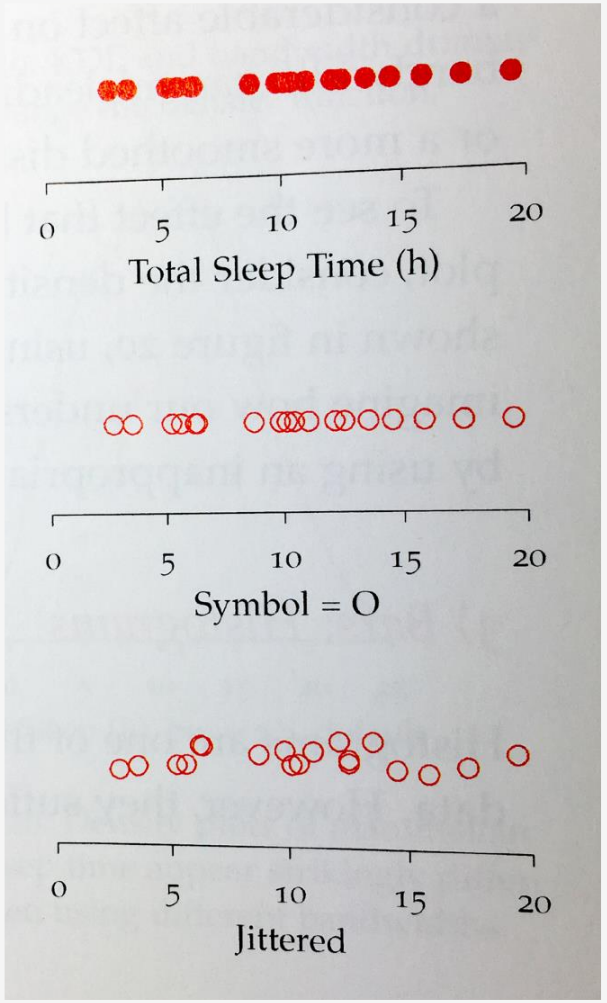


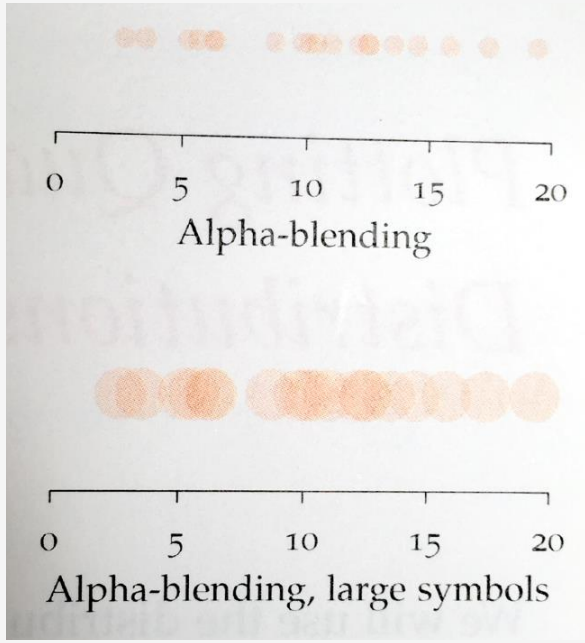
Color



Slide from: Eric Miller

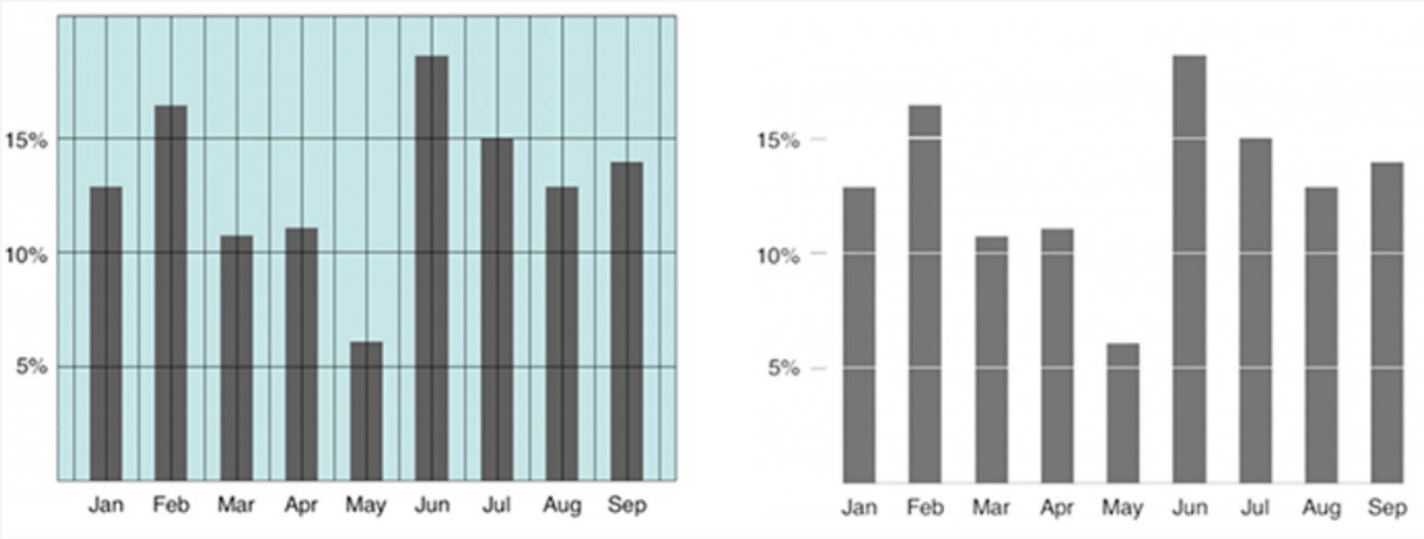
Overplotting





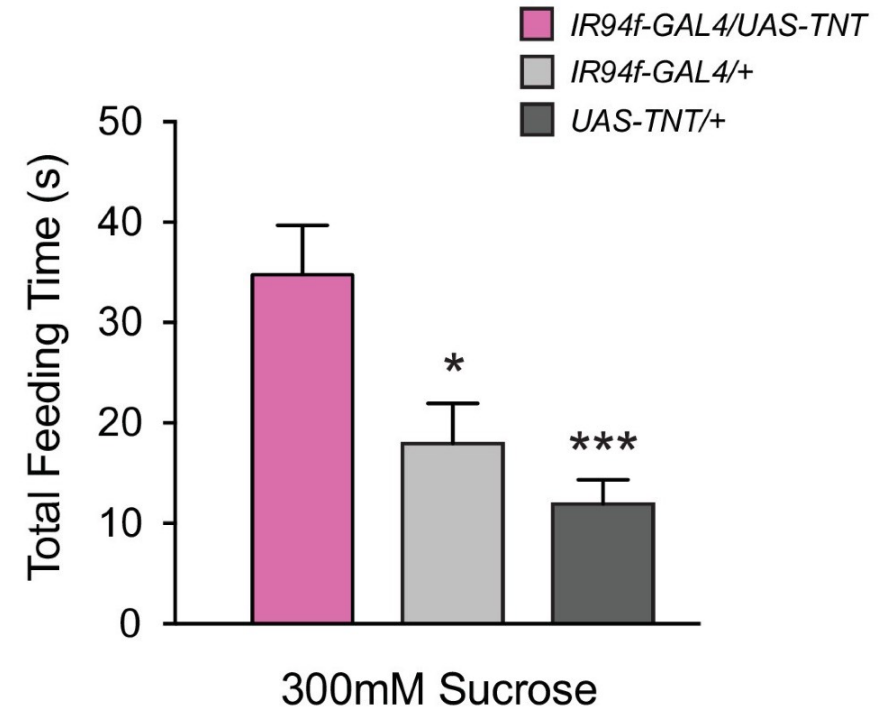
Slide from: Eric Miller

Data::Ink



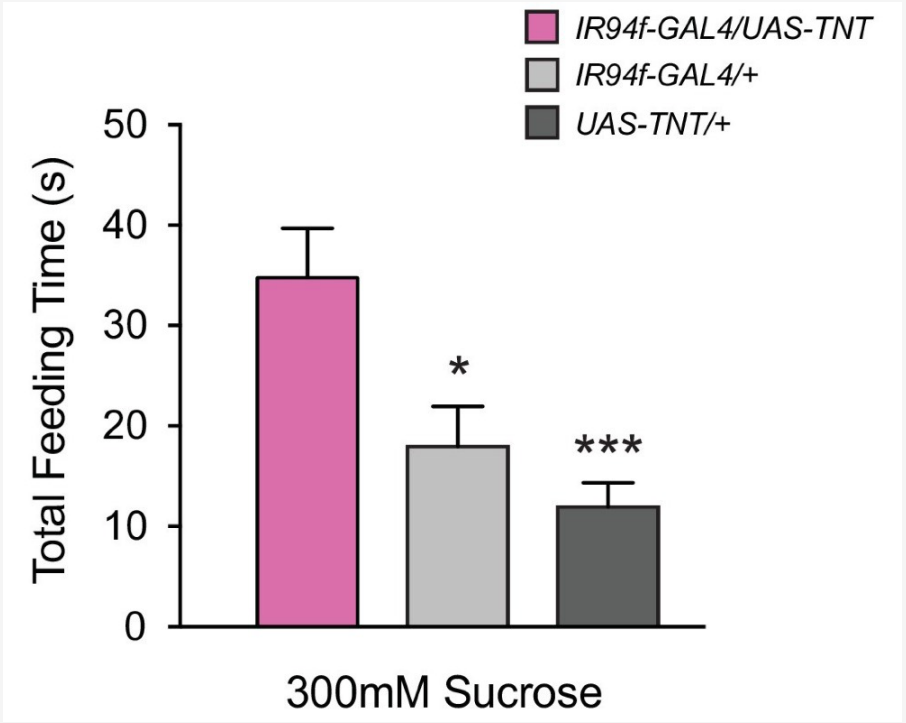
Slide from: Eric Miller

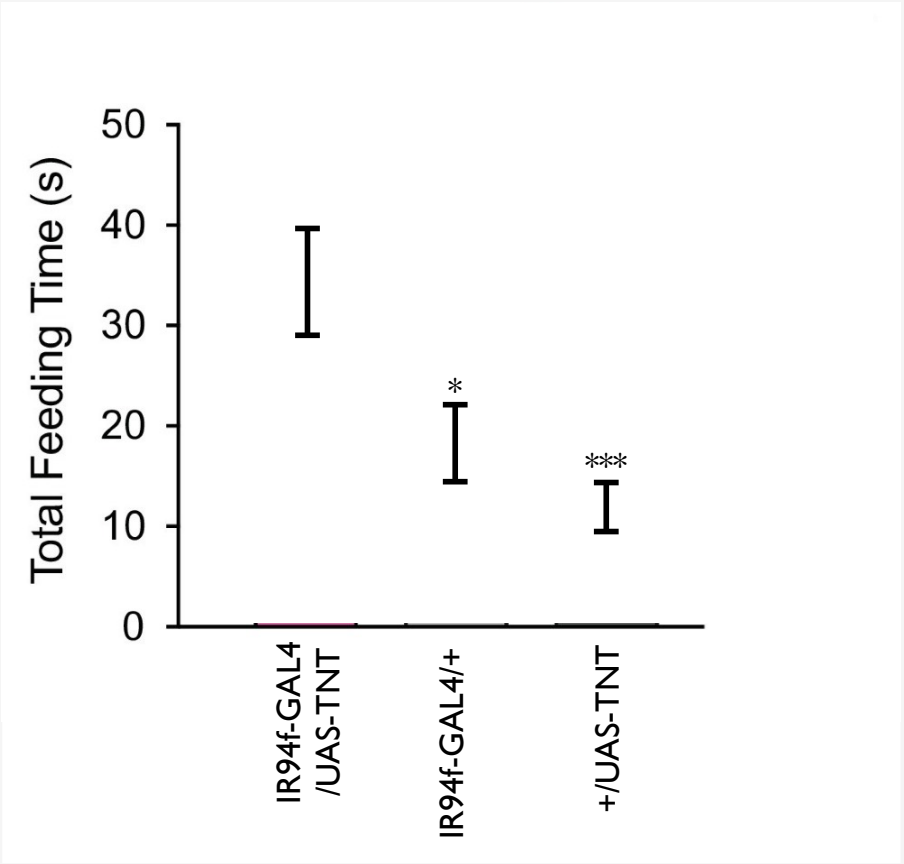
Data::Ink



Slide from: Eric Miller

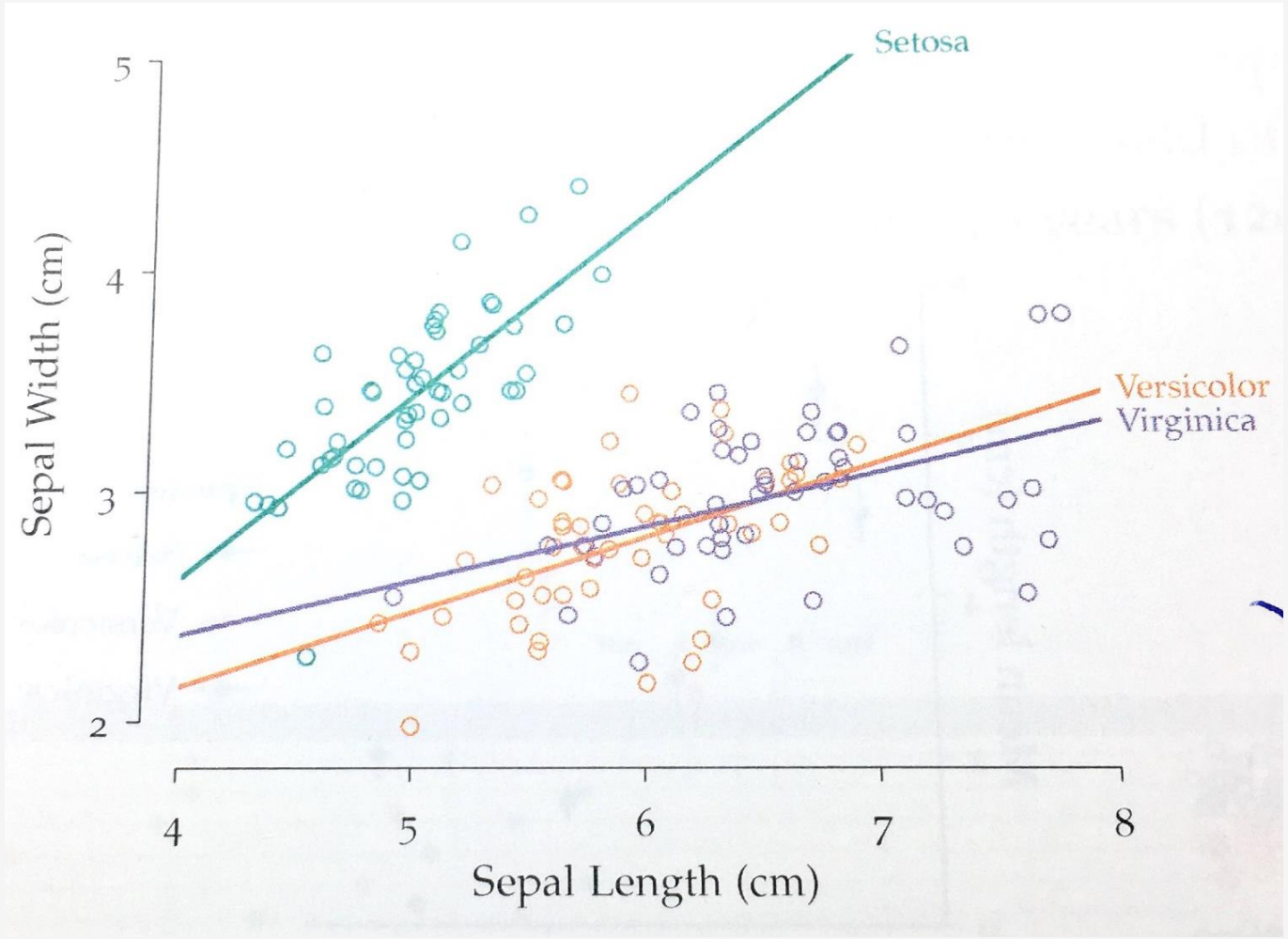
Data::Ink





Slide from: Eric Miller

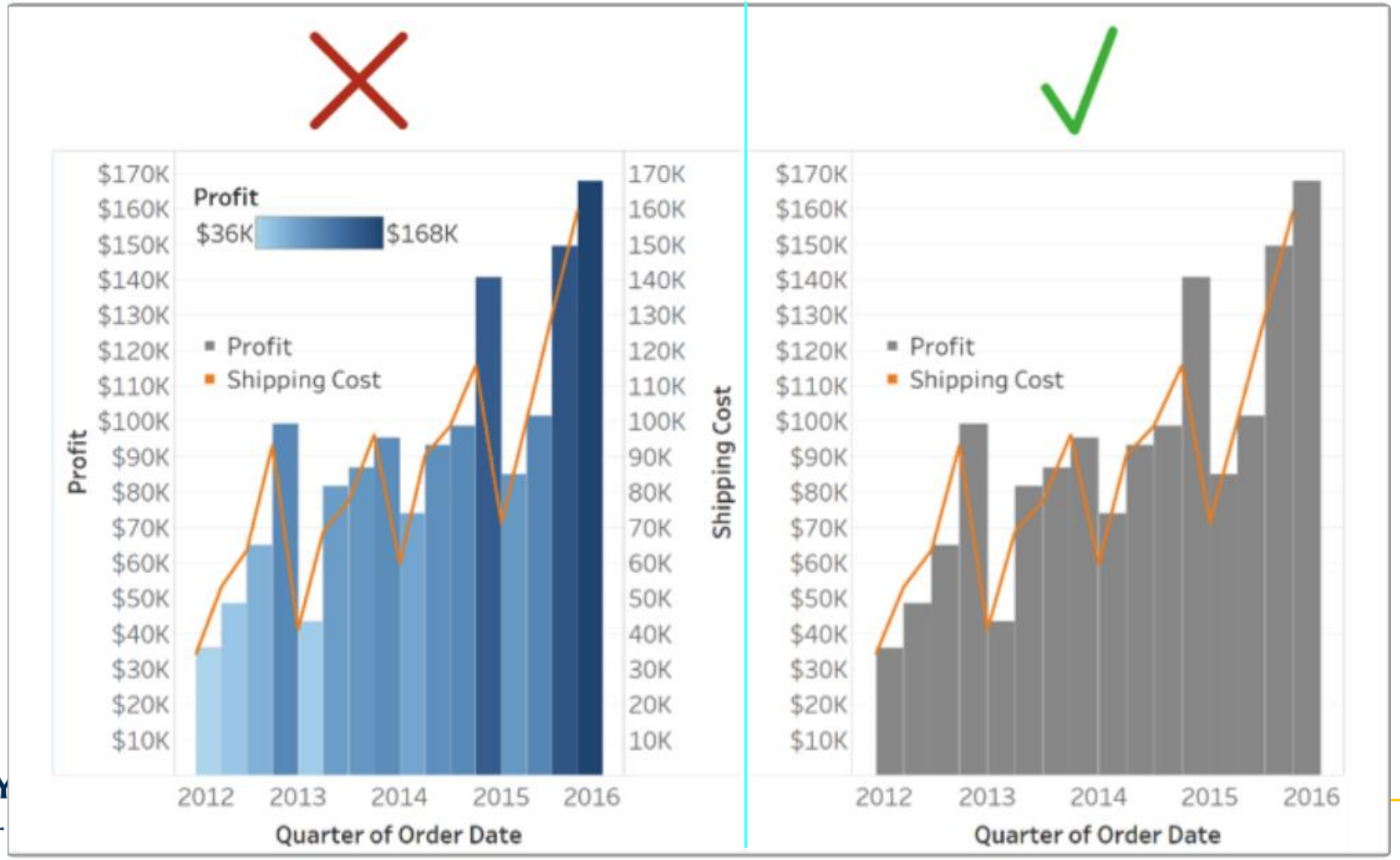
Data::Ink



Where is the legend?

Slide from: Eric Miller

Double encoding



Double encoding

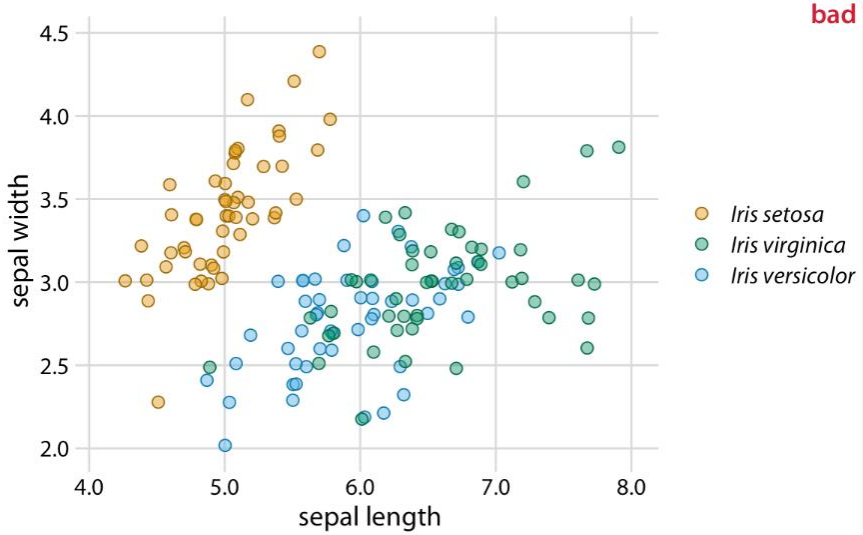
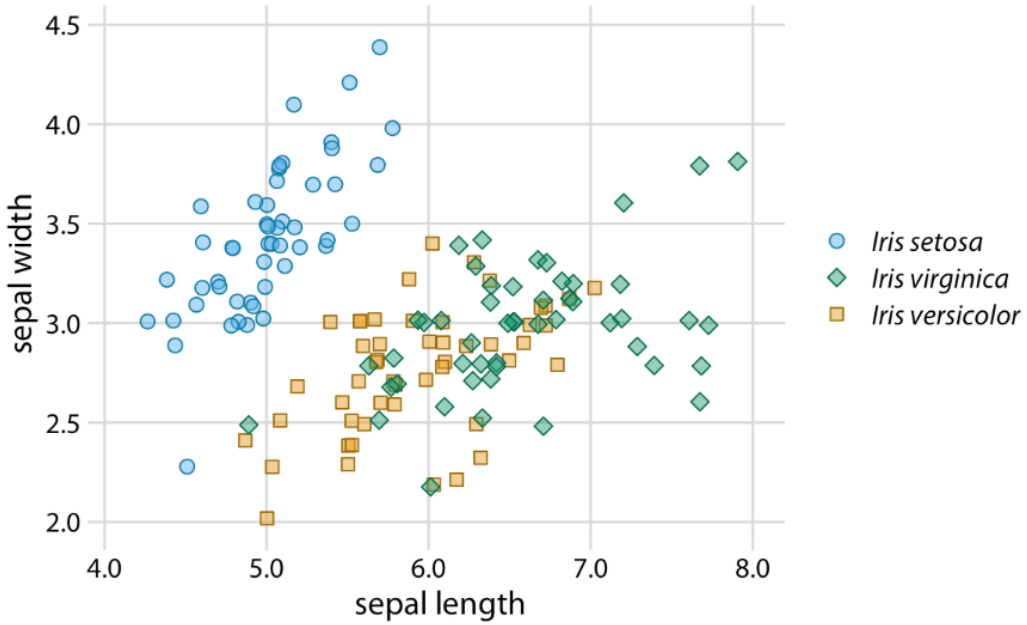
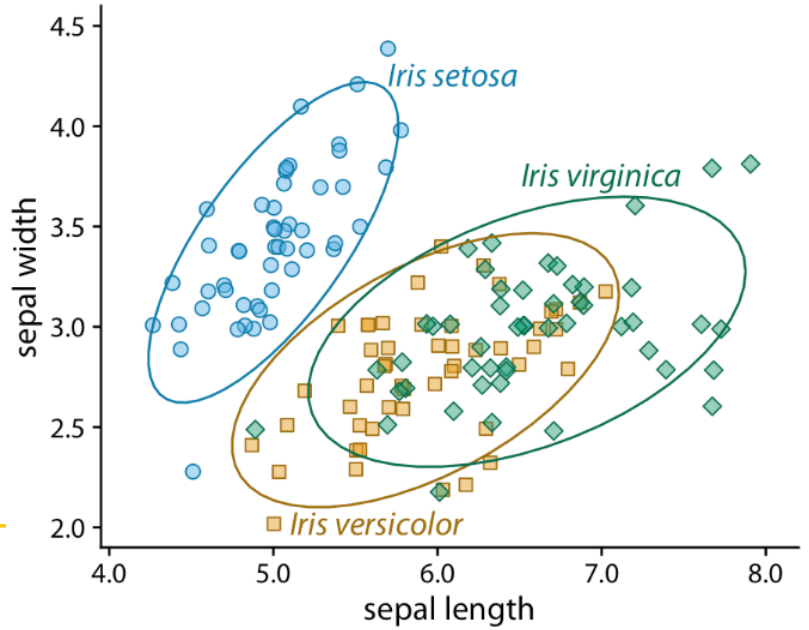


Fig 20.1/20.3/20.9 from "Fundamentals of Data Visualization" by Claus Wilke





Data::Ink

- Remove excess ink
- Show distributions, instead of bars
- Can you remove the legend?
- Remove double encodings when appropriate
- Is a log scale appropriate? https://www.lrs.org/2020/06/17/visualizing-data-the-logarithmic-scale/
- What do the 'error bars' represent?

Slide from: Eric Miller