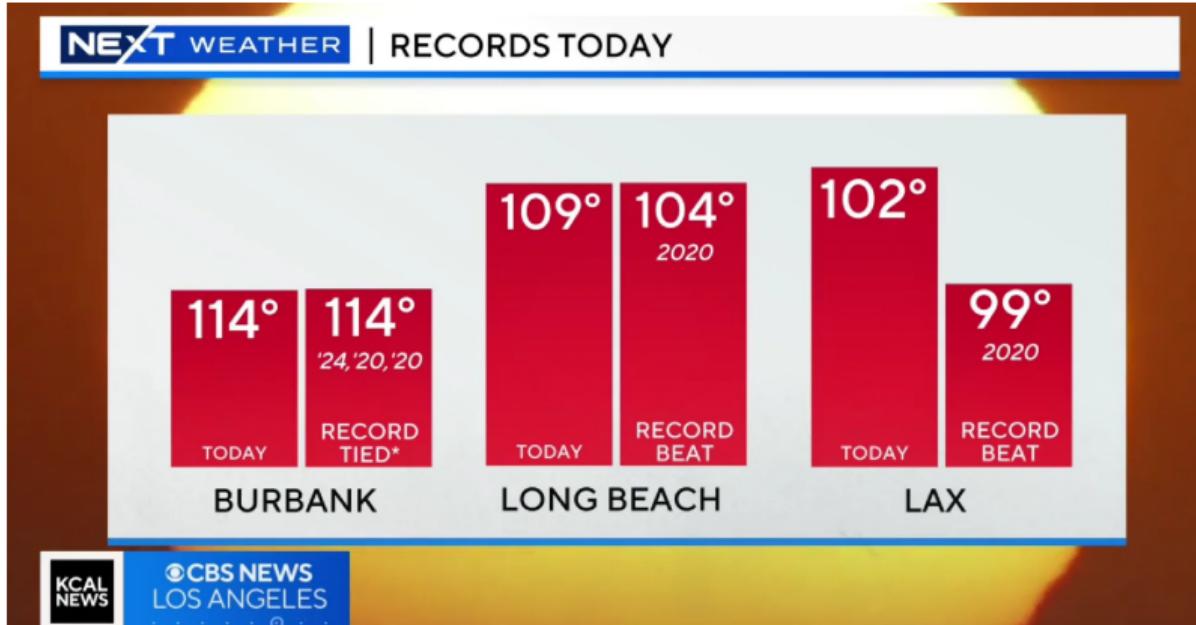


Outline

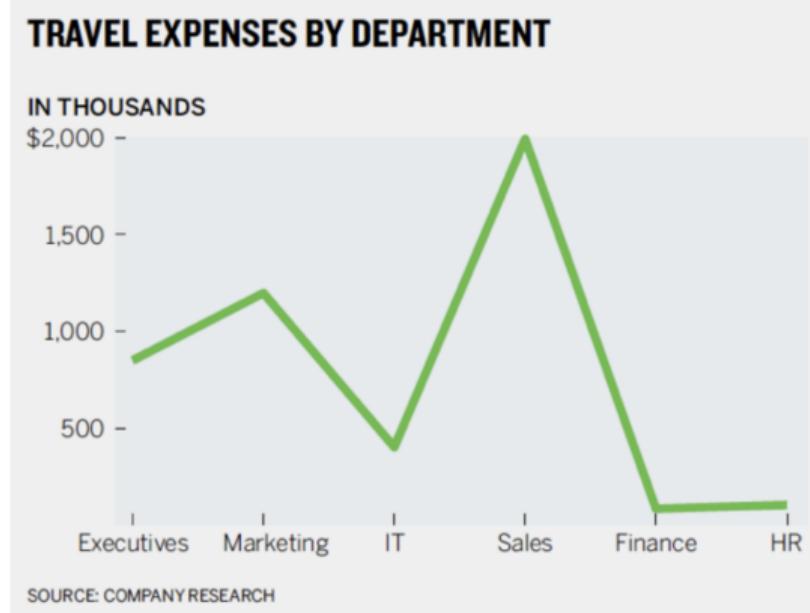
- What makes a data visualization bad?
- What makes a data visualization good?
- The grammar of graphics
- Some rules for making good data visualizations
- Two data visualizations economists should know

Examples of Bad Data Visualizations

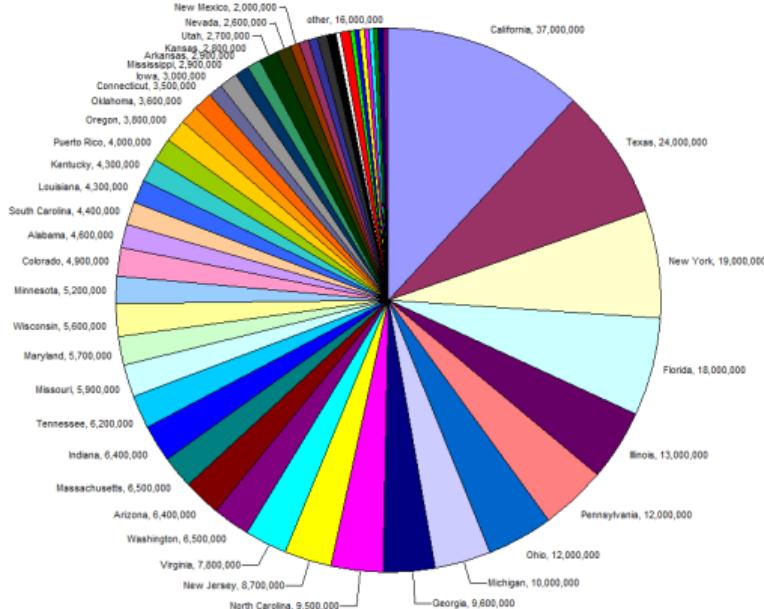


Source: Alessa Somer

Examples of Bad Data Visualizations

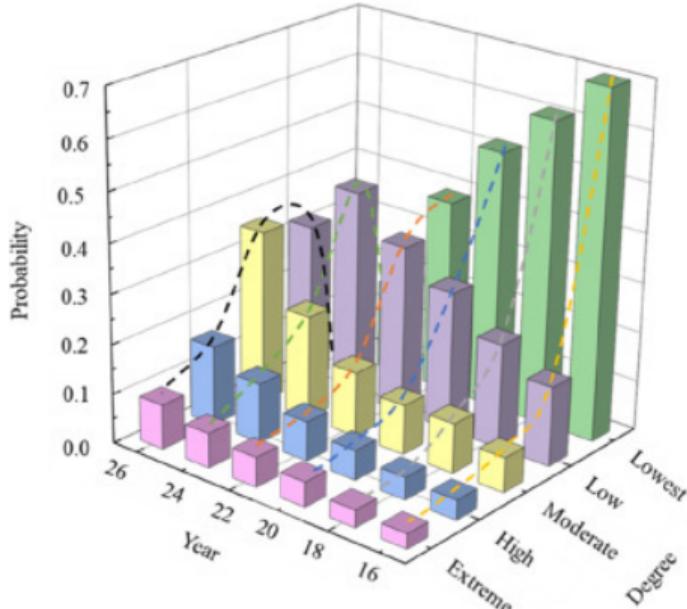


Examples of Bad Data Visualizations



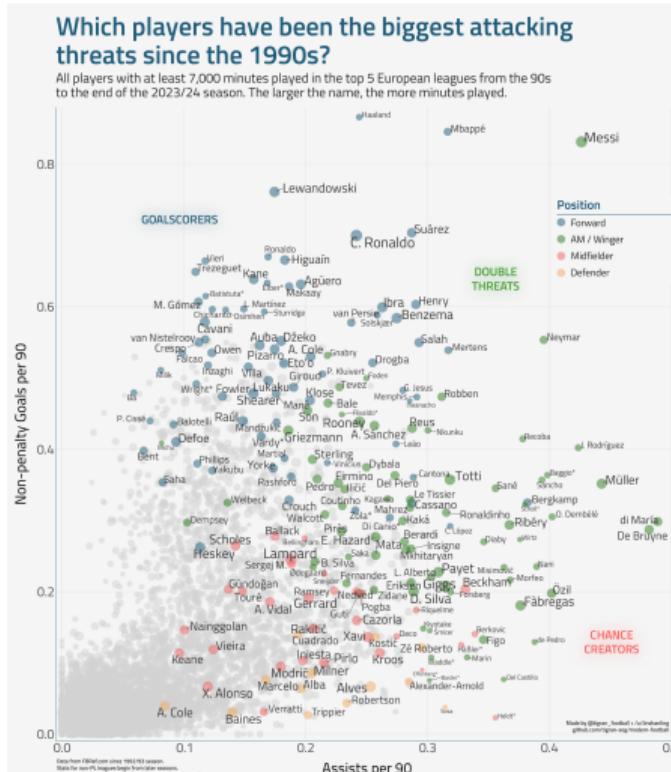
Source: Charlie Venci

Examples of Bad Data Visualizations



Source: Han et al. (2024)

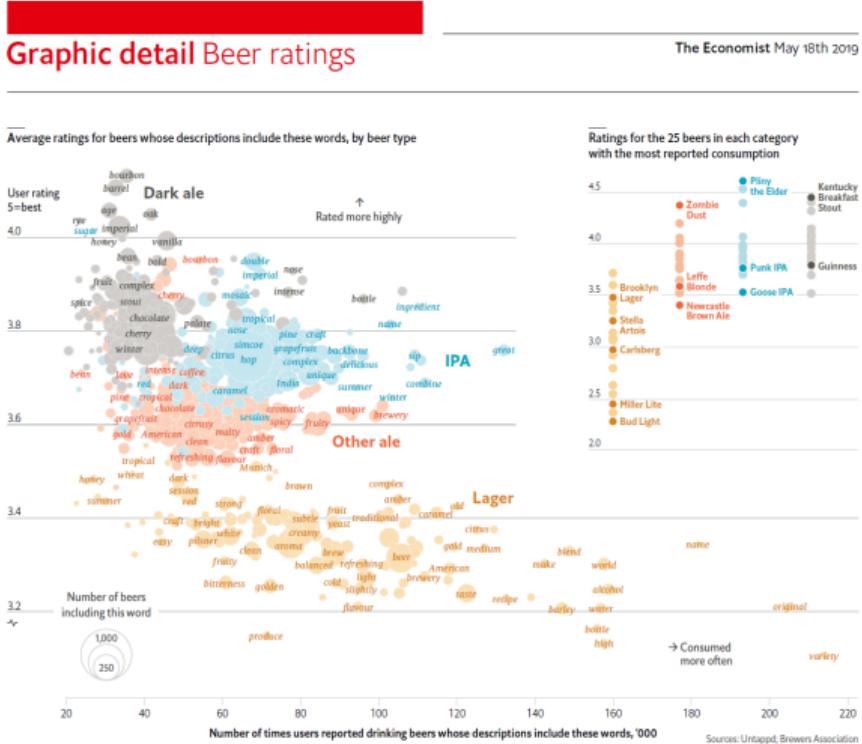
Good Data Visualizations Are All Alike



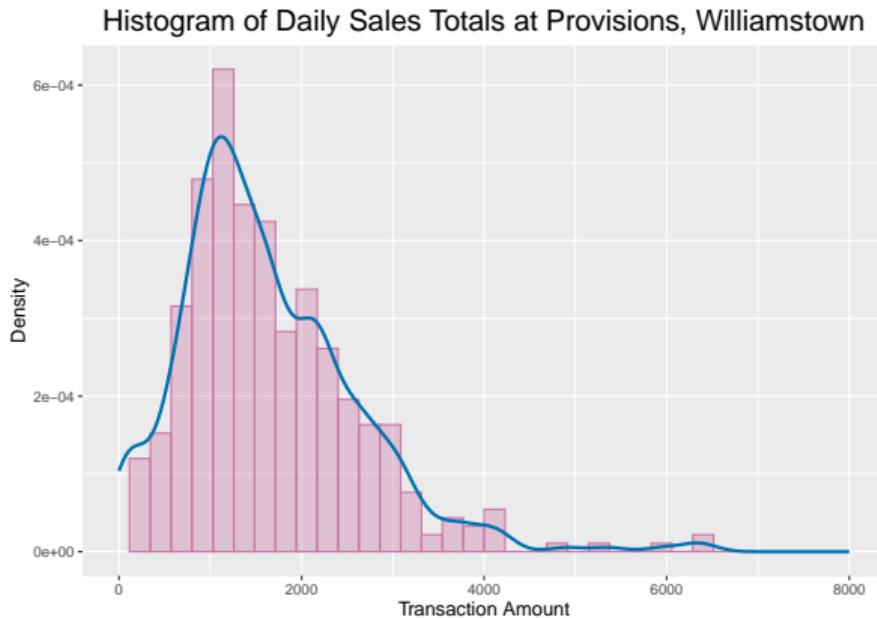
The Grammar of Graphics

- The core elements of a data visualization (Wickham 2010):
 - ▶ **Data** (in the form of a rectangular data frame)
 - ▶ **Aesthetics** (the data is mapped to the aesthetics using scales)
 - ▶ Position (usually in terms of x and y axes)
 - ▶ Other visual attributes (color, shape, size, line style, labels, etc.)
 - ▶ **Geometries**
 - ▶ The type of plot: bar, scatter, line, etc.
- Optional elements: statistics, facets, coordinates

The Grammar of Graphics in Practice

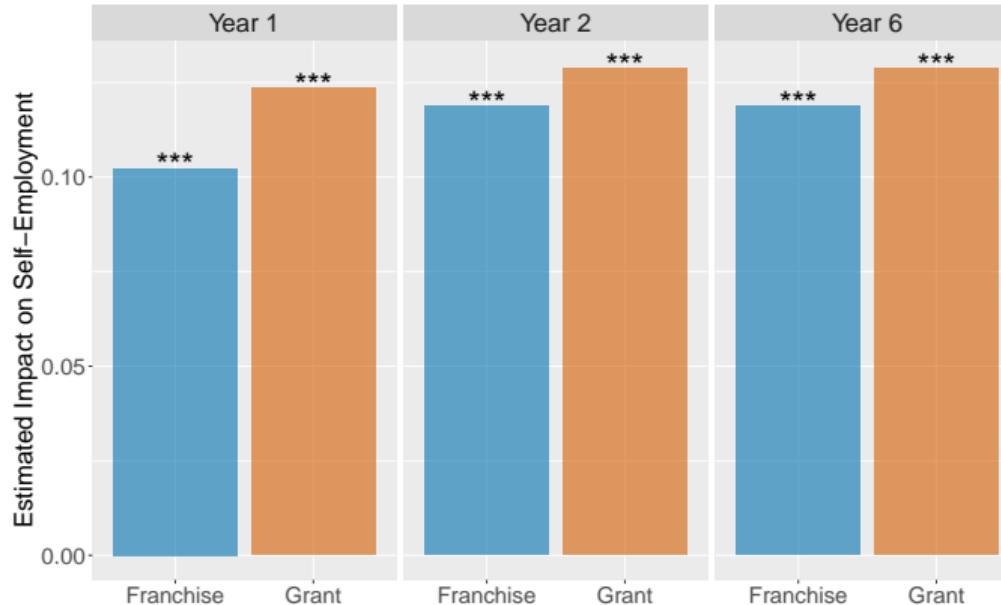


The Grammar of Graphics in Practice: Statistics



data → statistics → aesthetic mappings: here, the densities in each bin + kernel densities (x)

The Grammar of Graphics in Practice: Facets



Why Is the Grammar of Graphics a Grammar?

The grammar of graphics places limits on the set of valid data visualizations

- “I am going to the store later” vs. “Am later I store to the going”

There are (usually) infinitely many grammatically correct data visualizations

- “I am going to the store later”
- “I am going to the store in an hour”
- “I am driving to the store later”

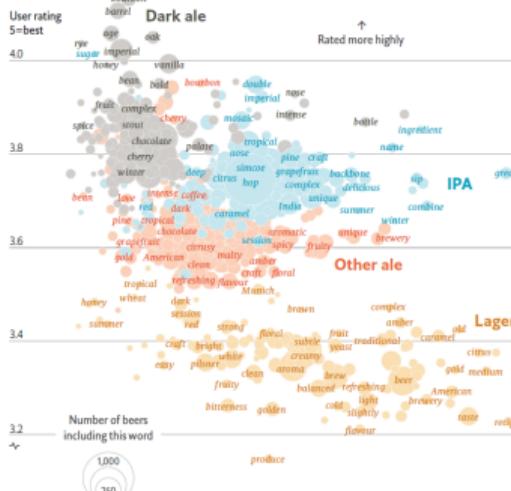
The choice of aesthetic mappings and geometrics is dictated by the message we wish to convey

Every Picture Tells a Story

Graphic detail Beer ratings

The Economist May 18th 2019

Average ratings for beers whose descriptions include these words, by beer type



Ratings for the 25 beers in each category with the most reported consumption

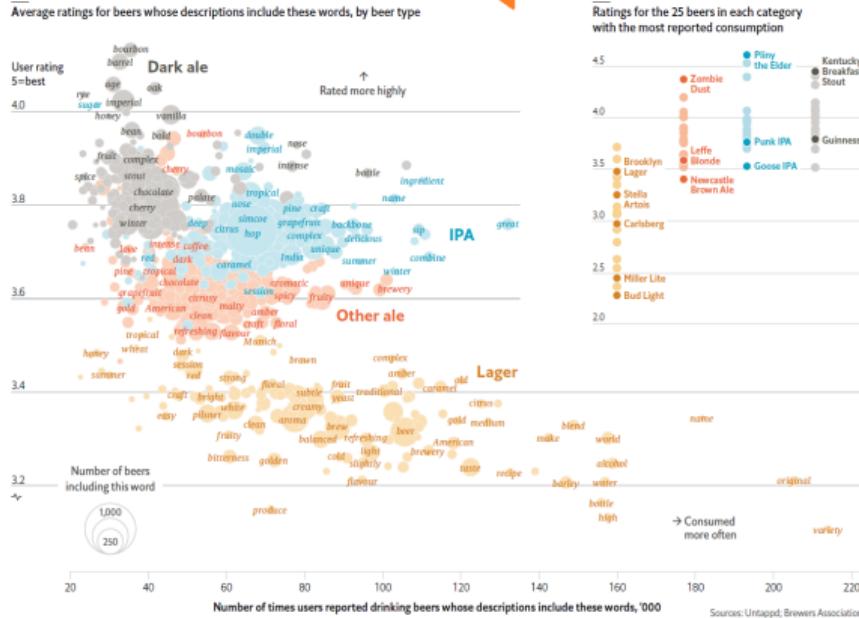


Every Picture Tells a Story

Graphic detail Beer ratings

The Economist May 18th 2019

The beers online raters drink most are the ones they claim to like least



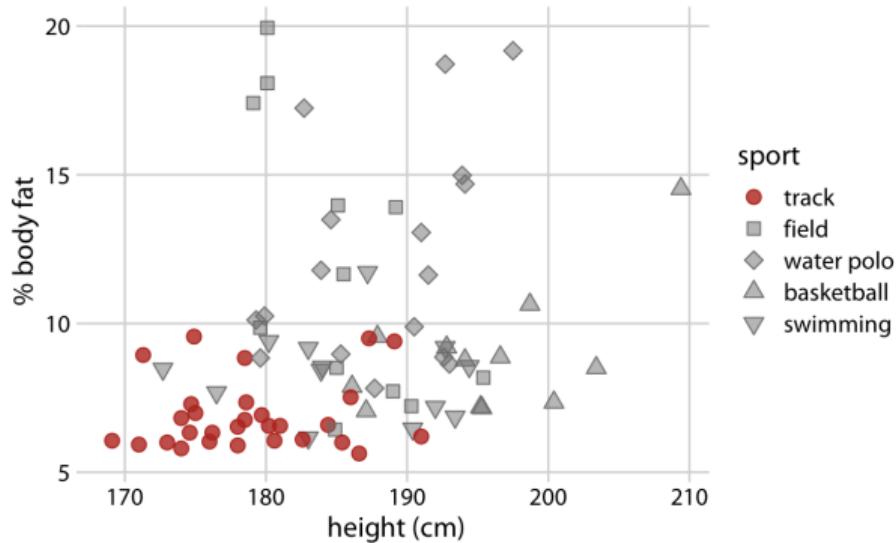
Economics 370: Data Science for Economics (Professor Jakiela)

Lecture 3: Data Visualization, Slide 14

Rules for Data Visualizations

1. Know what you are saying: deploy aesthetics intentionally to make your point
2. Use aesthetic mappings faithfully
 - 2.1 The x and y axes are aesthetics indicating magnitude or quantity
 - 2.2 Color is an aesthetic mapping – use it accordingly or hold it constant
3. Be kind to humans
 - 3.1 Use colorblind-friendly colors
 - 3.2 Use larger text for your axis labels (yes, I'm talking to you!)
 - 3.3 Avoid 3D plots, adding too many aesthetics to process
4. Don't do anything else (i.e. don't add visual elements that don't serve a purpose)

Using Color to Highlight



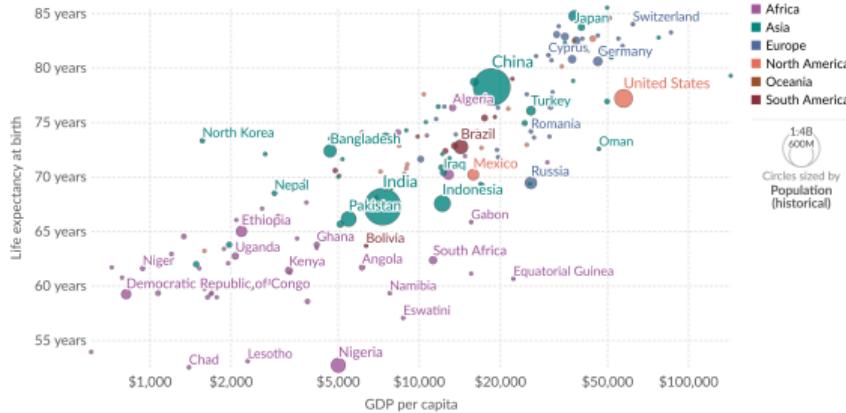
Source: Wilke (2019)

Using Color to Illustrate Clusters

Life expectancy vs. GDP per capita, 2021



The period life expectancy¹ at birth, in a given year. GDP per capita is adjusted for inflation and differences in the cost of living between countries.



Data source: UN WPP (2022); HMD (2023); Zijlstra et al. (2015); Riley (2005); Bolt and van Zanden - Maddison Project Database 2023

Note: GDP per capita is expressed in international-\$² at 2011 prices.

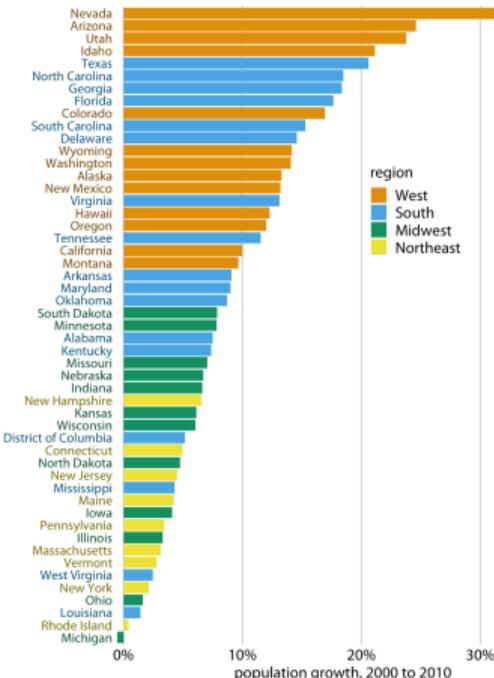
OurWorldinData.org/life-expectancy | CC BY

1. Period life expectancy: Period life expectancy is a metric that summarizes death rates across all age groups in one particular year. For a given year, it represents the average lifespan for a hypothetical group of people, if they experienced the same age-specific death rates throughout their whole lives as the age-specific death rates seen in that particular year. Learn more in our article: "Life expectancy" – What does this actually mean?

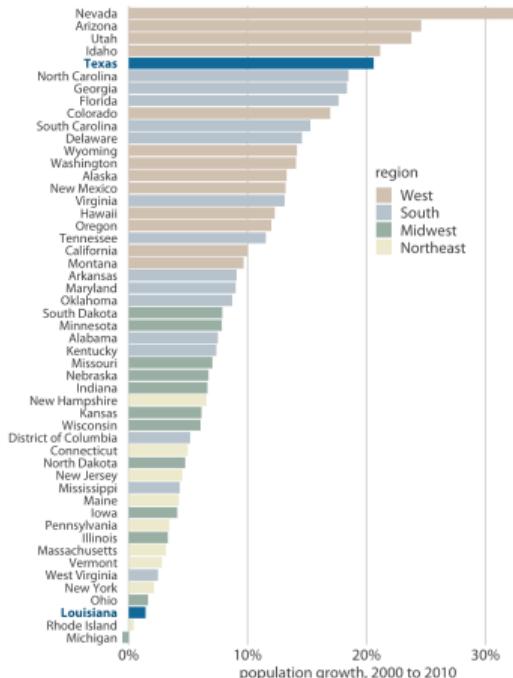
2. International dollars: International dollars are a hypothetical currency that is used to make meaningful comparisons of monetary indicators of living standards. Figures expressed in International dollars are adjusted for inflation within countries over time, and for differences in the cost of living between countries. The goal of such adjustments is to provide a unit whose purchasing power is held fixed over time and across countries, such that one international dollar can buy the same quantity and quality of goods and services no matter where or when it is spent. Read more in our article: What are Purchasing Power Parity adjustments and why do we need them?

Source: Our World in Data

Different Ways to Use Color

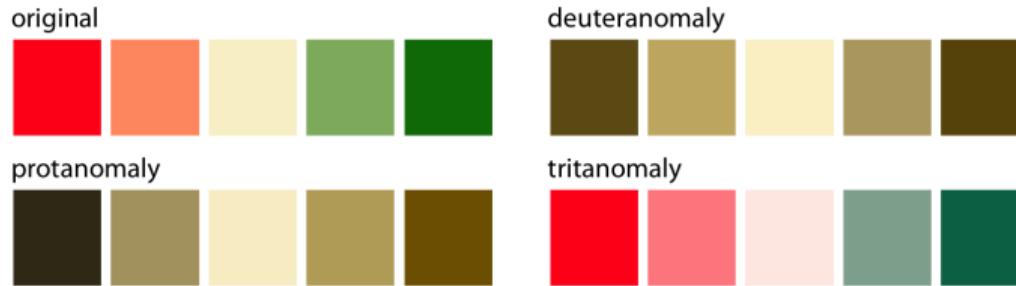


Source: Wilke (2019)



Source: Wilke (2019)

7% of Males in the U.S. Are Color Blind



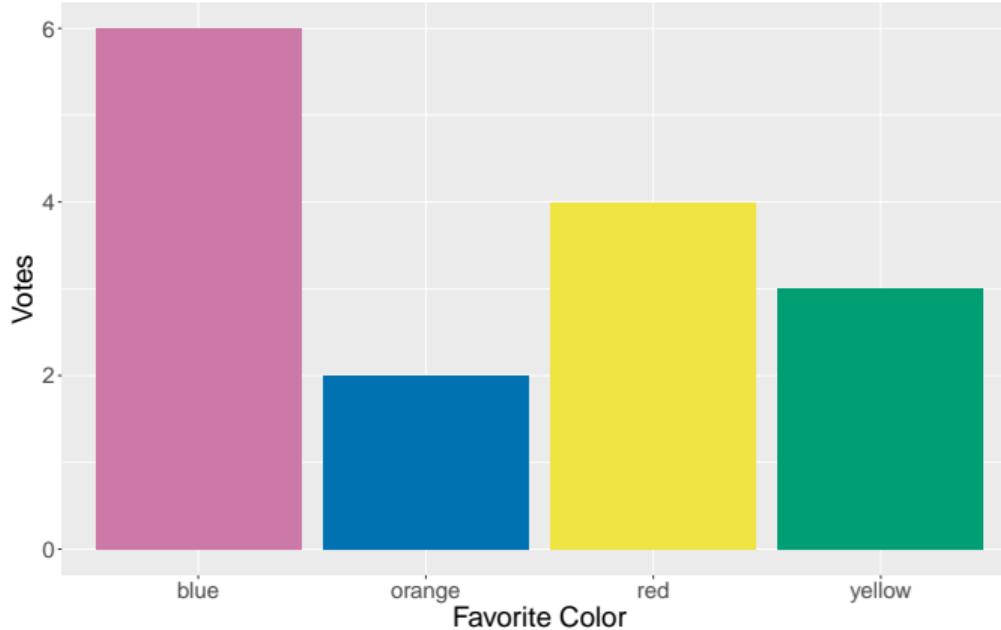
Wilke (2019)

The Okabe-Ito Color Blind Friendly Palette



Wilke (2019)

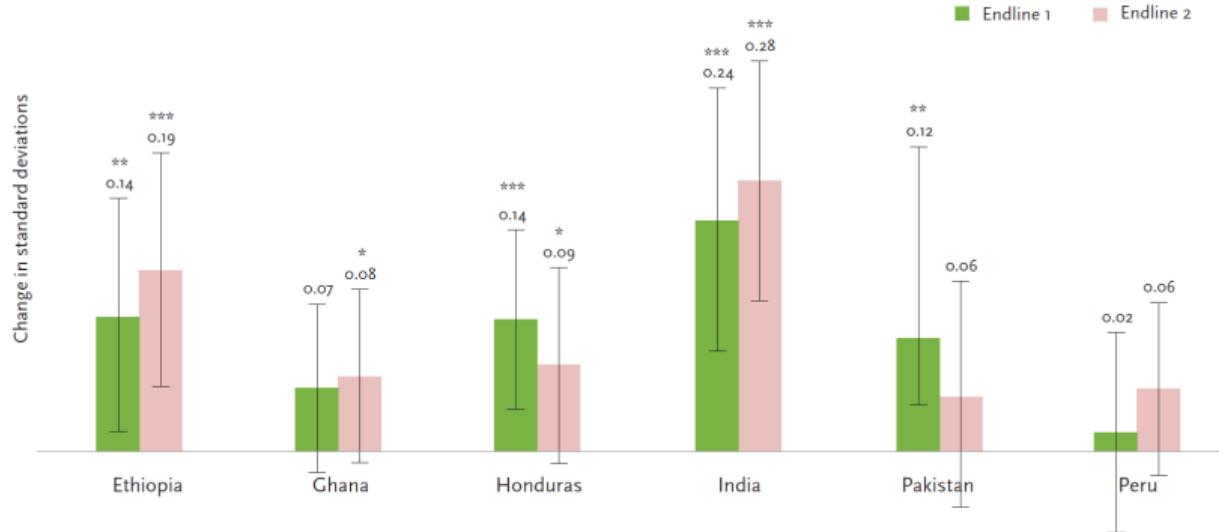
Only Use Color (Variation) as an Aesthetic



Rules for Data Visualizations

1. Know what you are saying: deploy aesthetics intentionally to make your point
2. Use aesthetic mappings faithfully
 - 2.1 The x and y axes are aesthetics indicating magnitude or quantity
 - 2.2 Color is an aesthetic mapping – use it accordingly or hold it constant
3. Be kind to humans
 - 3.1 Use colorblind-friendly colors
 - 3.2 Use larger text for your axis labels (yes, I'm talking to you!)
 - 3.3 Avoid 3D plots, adding too many aesthetics to process
4. Don't do anything else (i.e. don't add visual elements that don't serve a purpose)

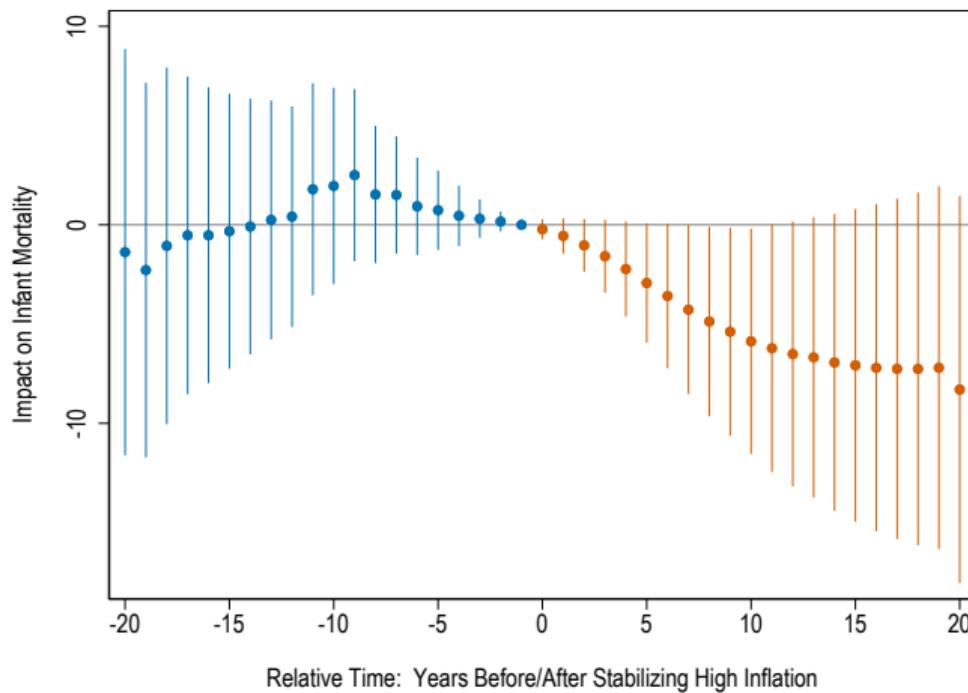
The Two Most Important Data Visualizations for Economists



Note: Error bars represent 95% confidence intervals. Statistical significance relative to comparison households at each endline is noted at the 1% (***) or 5% (**) level.

Source: J-PAL

The Two Most Important Data Visualizations for Economists



Lab #3

Objective: to make a figure summarizing the impact of two labor market interventions on the self-employment of young women in Nairobi using data from Brudevold-Newman et al. (2024)

- RCT with three treatment arms: Control, **Franchise**, **Grant**
- OLS specification for an RCT: $Y = \alpha + \beta F + \gamma G + \varepsilon$
- Three rounds of post-treatment data: Year 1, Year 2, and Year 6

Basic syntax for ggplot in R:

```
group.colors <- c(blue = c1, yellow = c2, orange = c3, red = c4)
ggplot(data, aes(x = colors, y = numbers, fill = colors)) +
  geom_col() +
  scale_fill_manual(values=group.colors)
```

Rafael Irizarry's Data Visualization Principles

- Know when to include 0
- Do not distort quantities
- Order categories by a meaningful value
- Show (all) the data
- Ease comparisons: use common axes and align plots for comparisons
- Visual cues that will be compared must be adjacent
- Think of the color blind
- Avoid 3D

Source: Irizarry (YEAR)