# Data Visualization

UCSB SUMMER STATS WORKSHOP 2023

WEEK 4

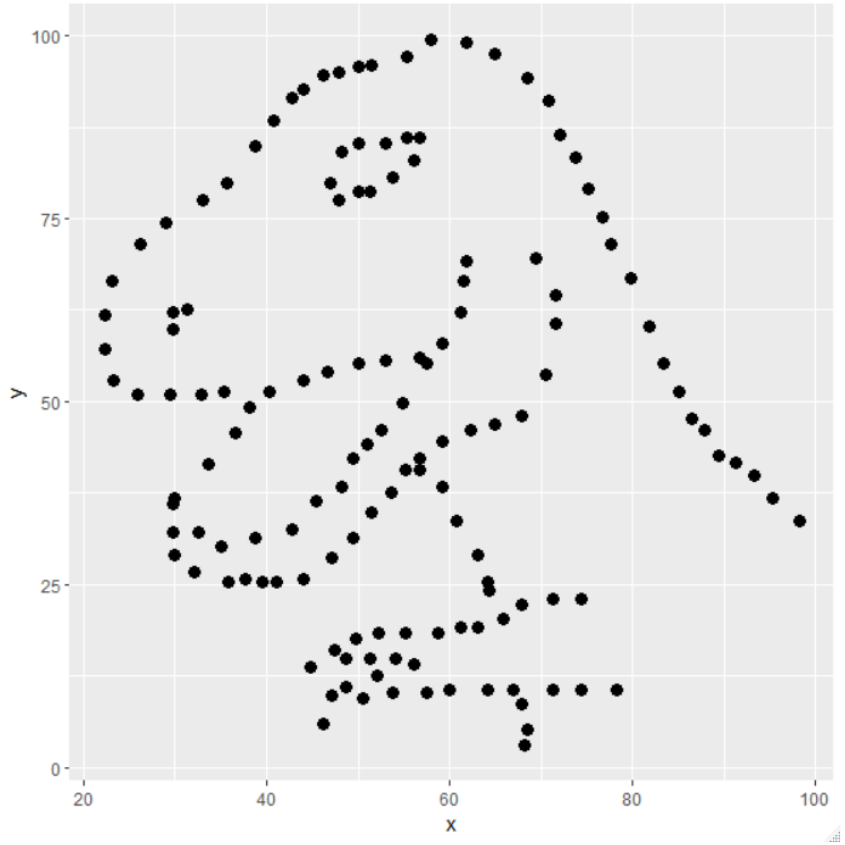
# Data visualization

Graphing your data is *essential*

```
> summary(data$x)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  22.31   44.10   53.33   54.26   64.74   98.21

> summary(data$y)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  2.949  25.288  46.026  47.832  68.526  99.487

> cor(data$x,data$y)
[1] -0.06447185
```



# Data visualization

Not all graphs are good

No single algorithm for good graphs

- Every situation is different
- Have some room for creative freedom

But there are some general guidelines

**Statistical Thinking for the 21st Century**



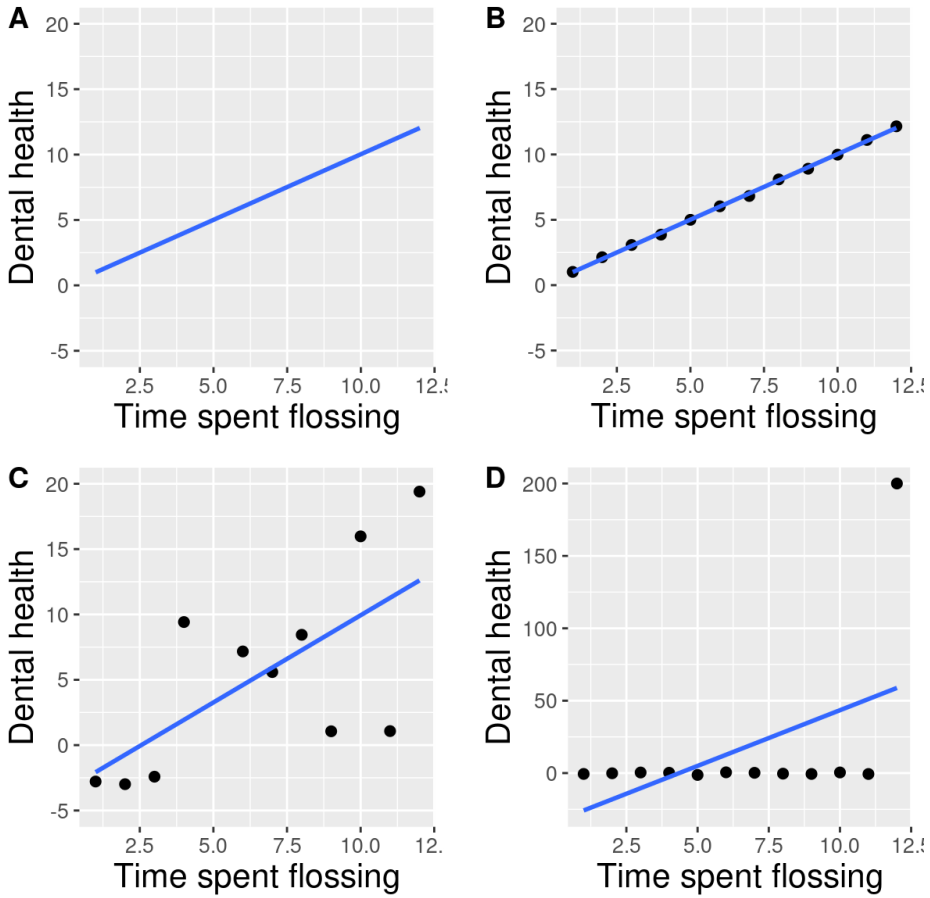
*Draft: 2022-12-22*

# Guideline 1: Show the data

**Show as much of the data as possible**

- Show the raw data
- *Not* just aggregate trends
- Make the distribution clear

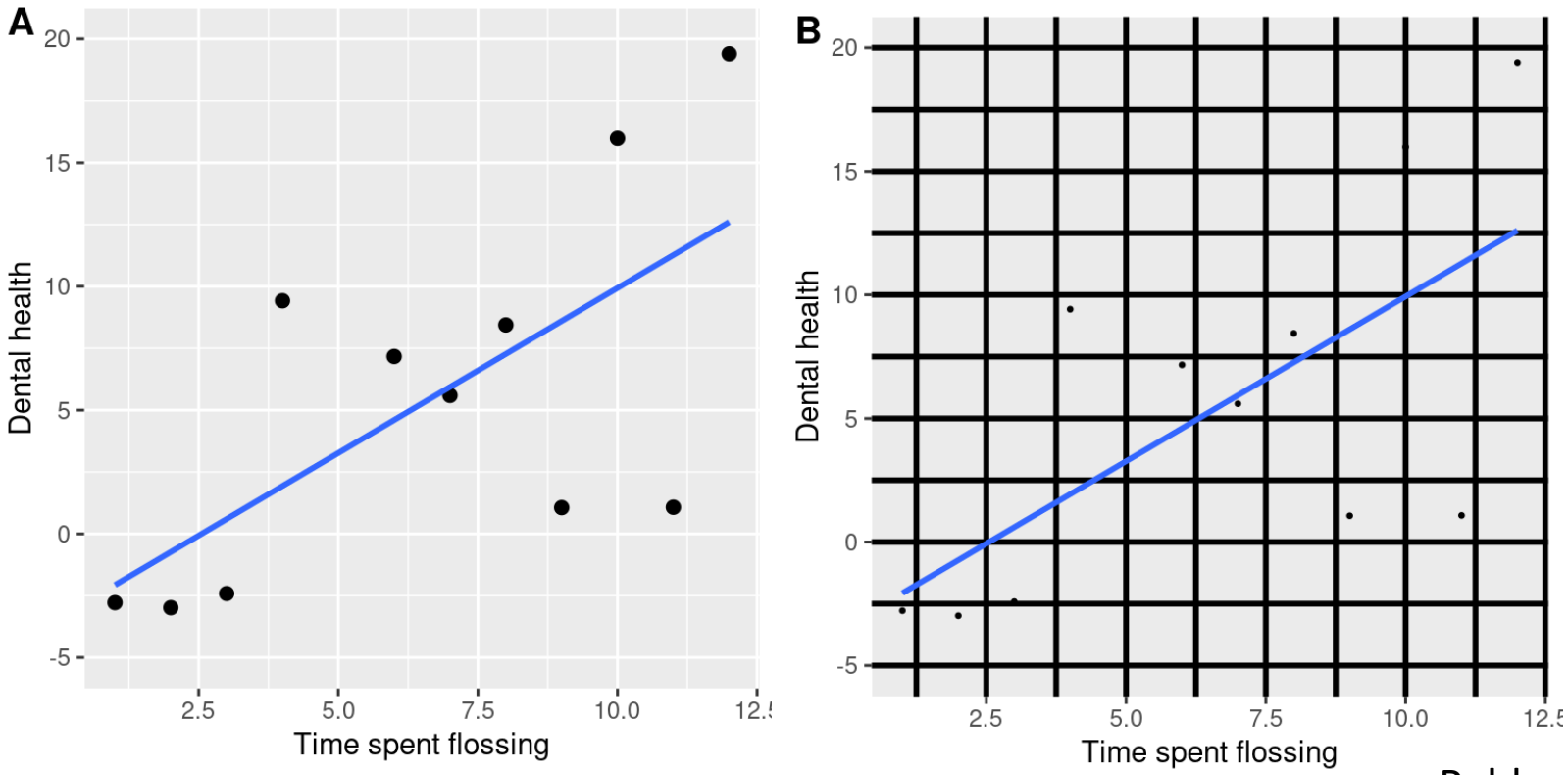
Make the data clear and easy to understand




Poldrack, 2019

# Guideline 2: Maximize the data/ink ratio

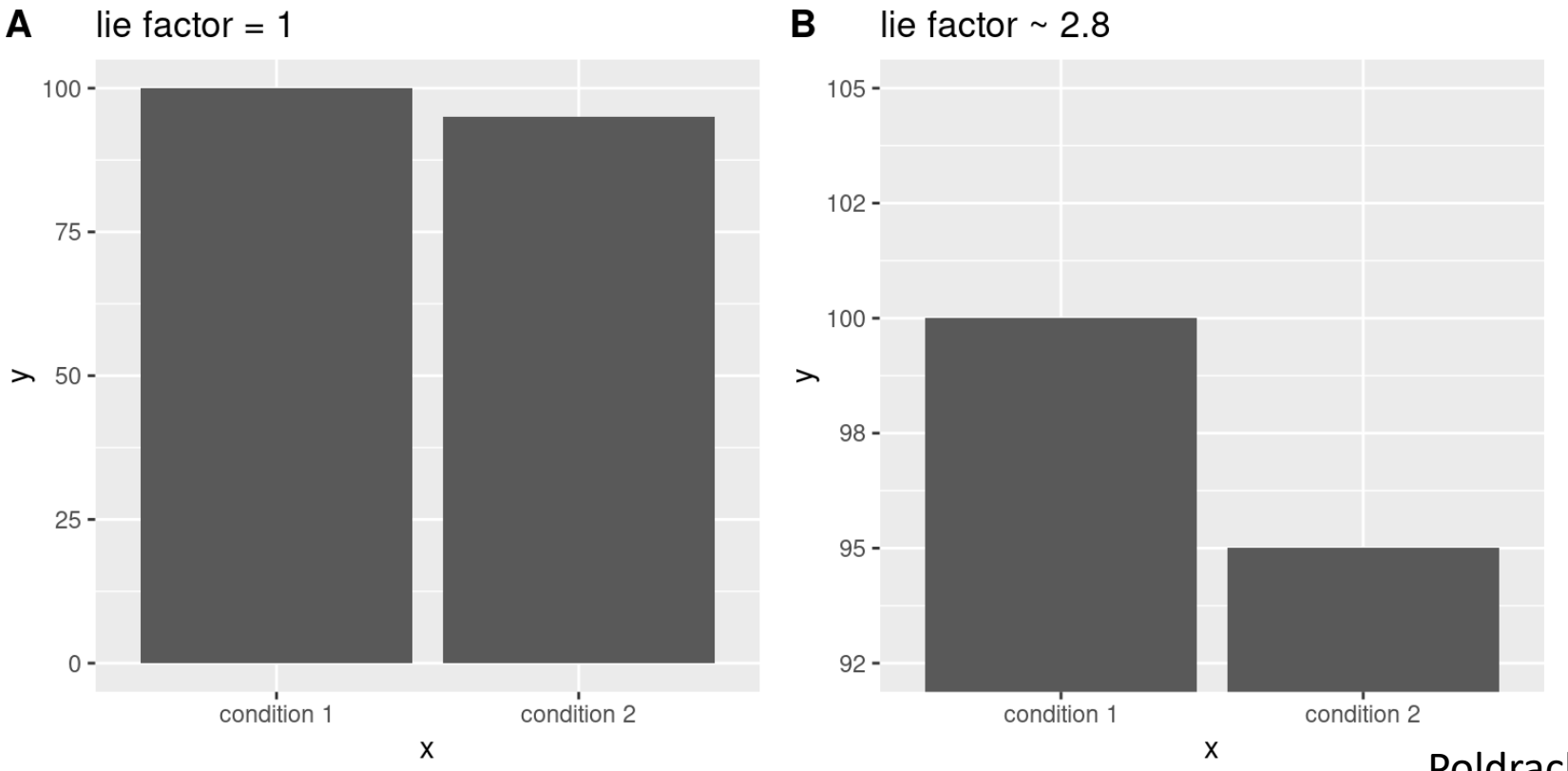
$$data/ink\ ratio = \frac{ink\ used\ on\ data}{total\ ink}$$




Poldrack, 2019

# Guideline 3: Minimize the lie factor

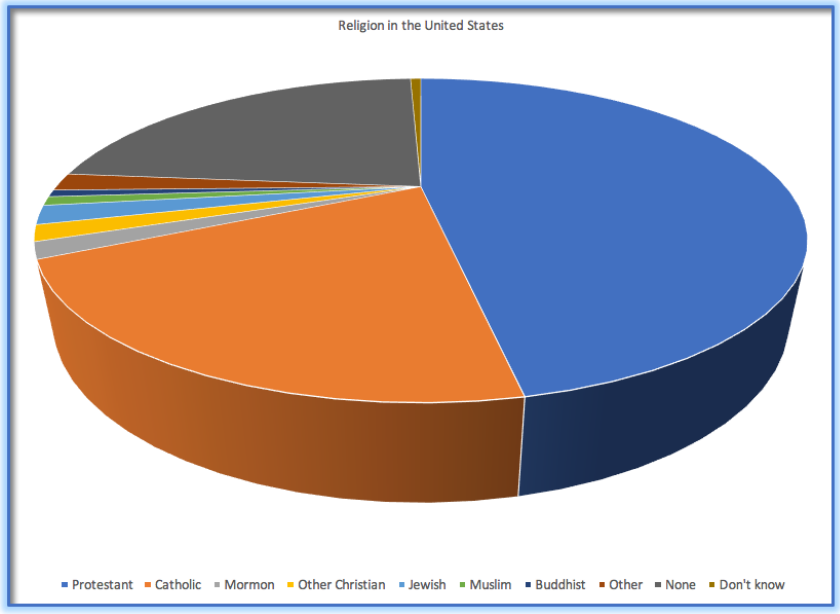
$$lie\ factor = \frac{Physical\ difference\ in\ your\ graph}{Magnitude\ of\ difference\ in\ your\ data}$$

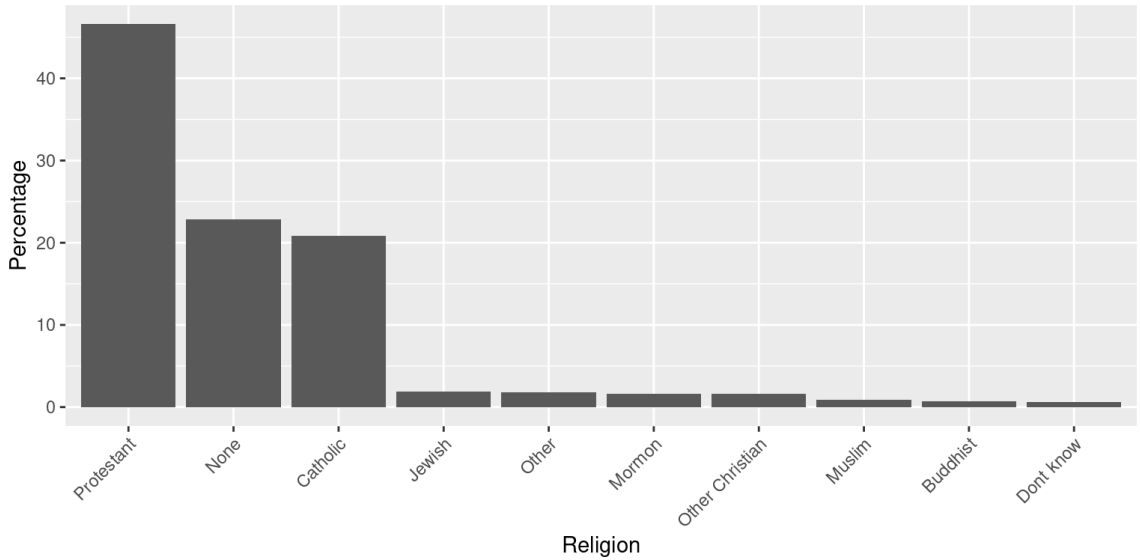



Poldrack, 2019

# Guideline 4: Be mindful of human perception

## Don’t show differences using volumes



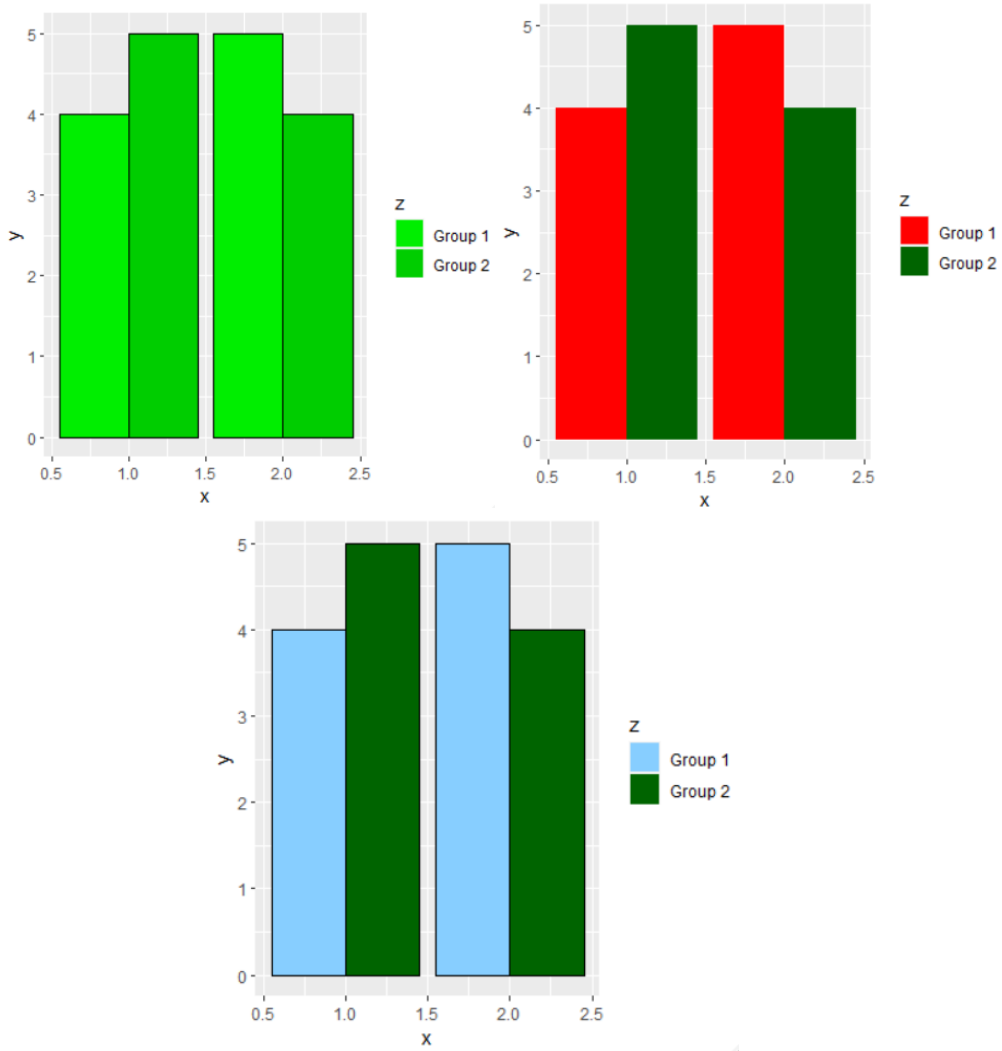




Poldrack, 2019

# Guideline 4: Be mindful of human perception

Choose your color palette carefully:

- High contrast
- Colorblind friendly
- Different in multiple ways



# Guideline 4: Be mindful of human perception

Make your graphs legible:

- Clear, intuitive labels
- Large, easy-to-read text