

# Data Visualization

---



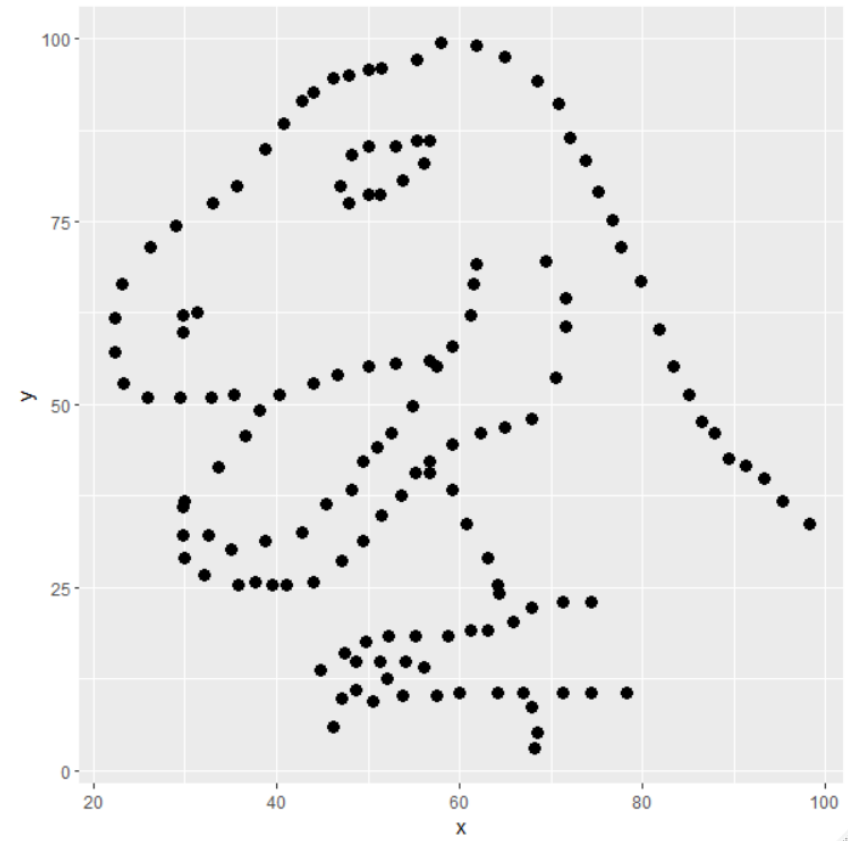
# Data visualization

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**

*Copyright 2019 Russell A. Poldrack*

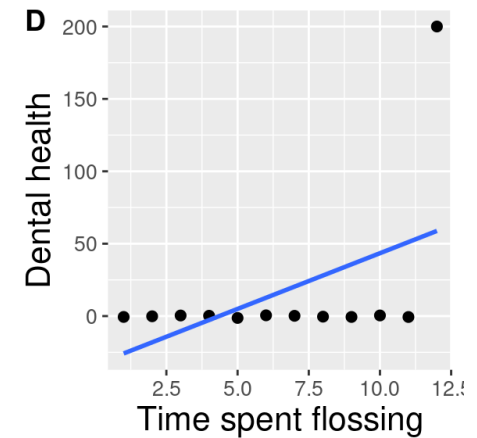
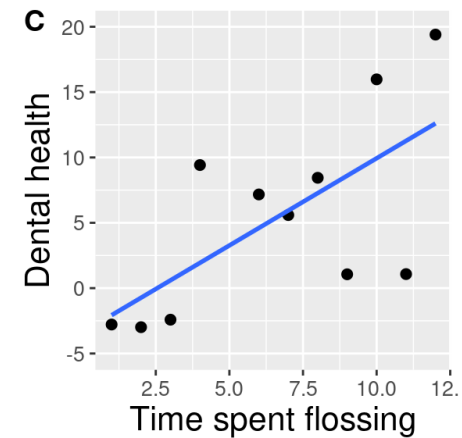
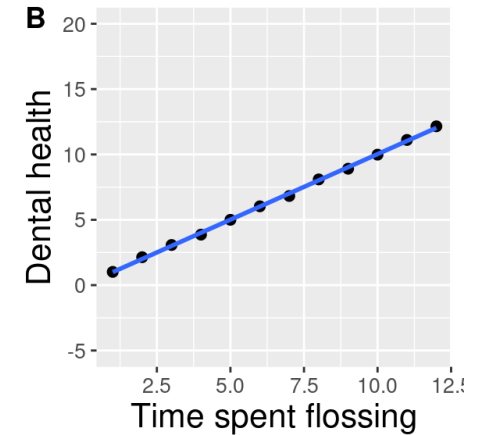
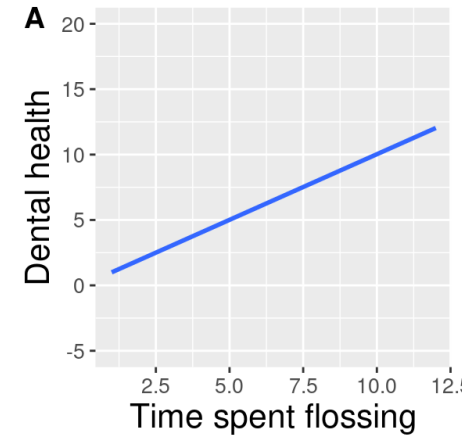
*Draft: 2022-12-22*

# Guideline 1: Show the data

## Show as much of the data as possible

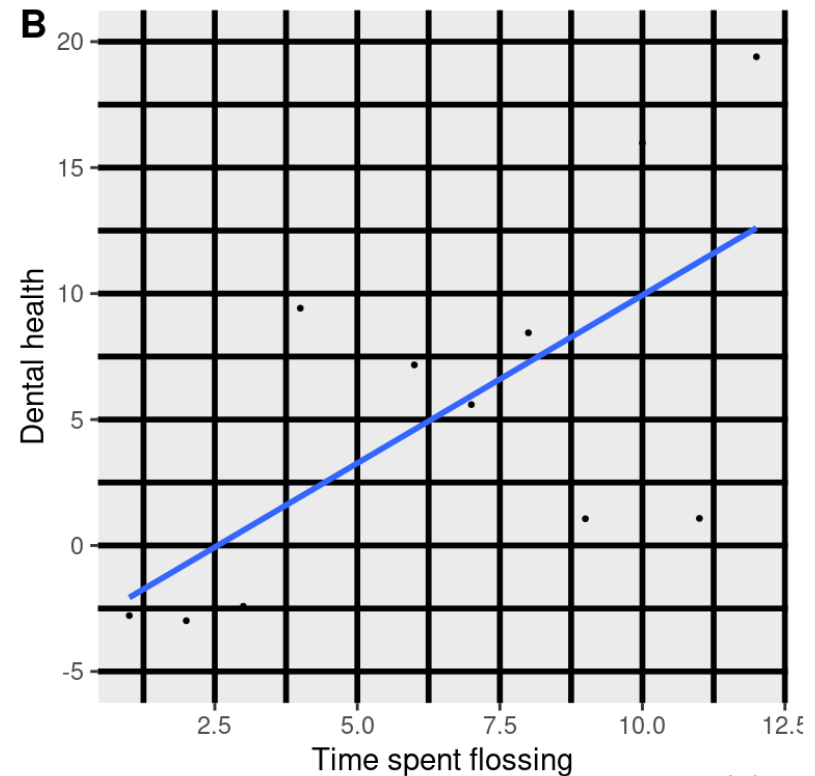
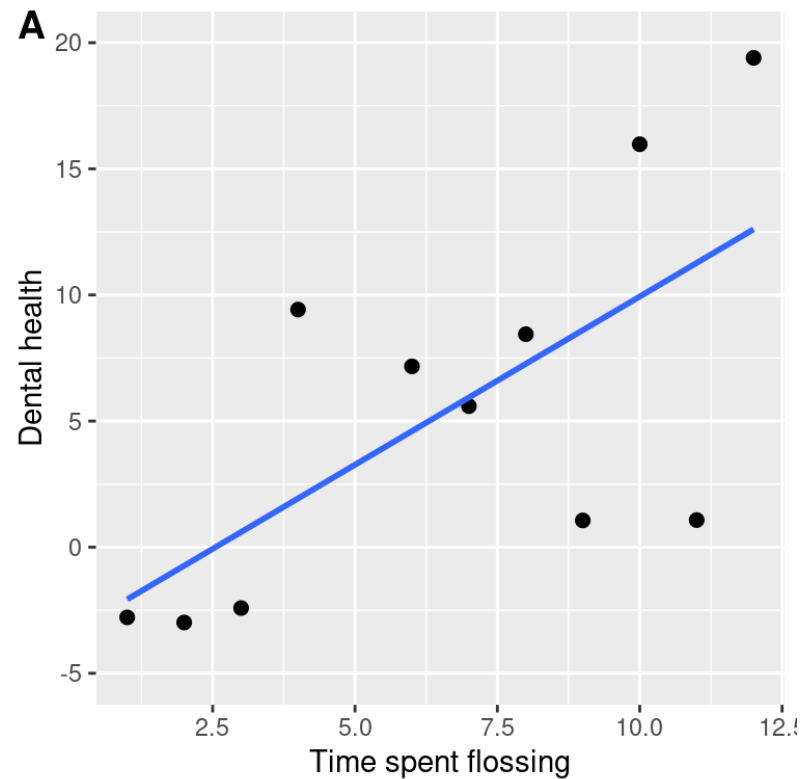
- Show the raw data
- *Not* just aggregate trends
- Make the distribution clear

## Make the data clear and easy to understand



# Guideline 2: Maximize the data/ink ratio

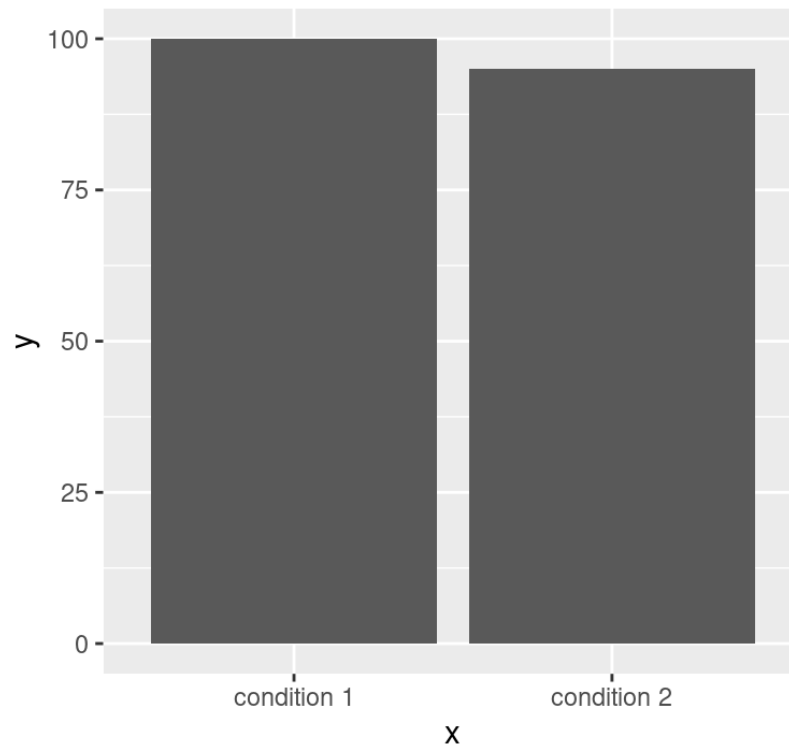
$$\text{data/ink ratio} = \frac{\text{ink used on data}}{\text{total ink}}$$



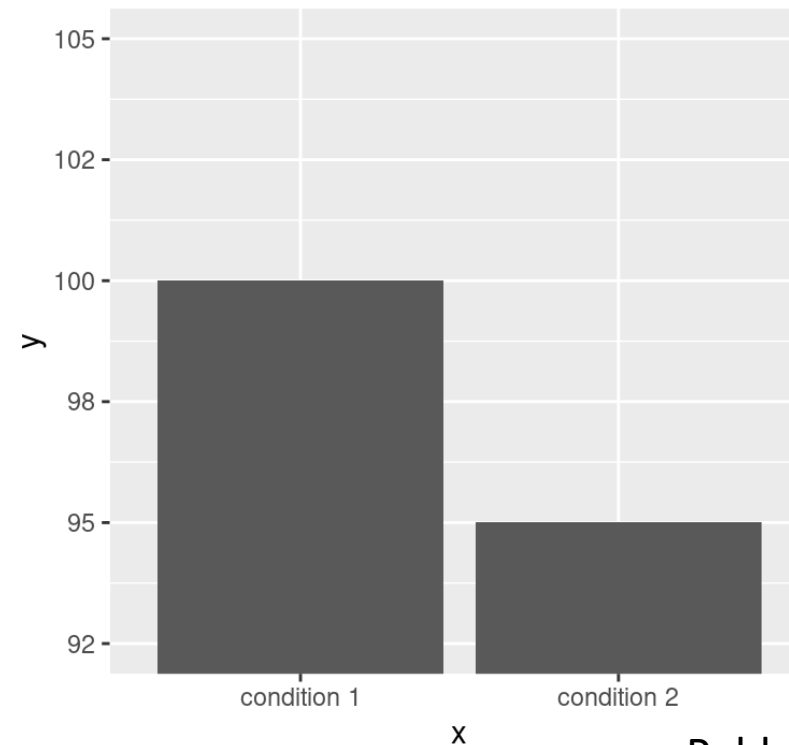
# Guideline 3: Minimize the lie factor

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

**A** lie factor = 1

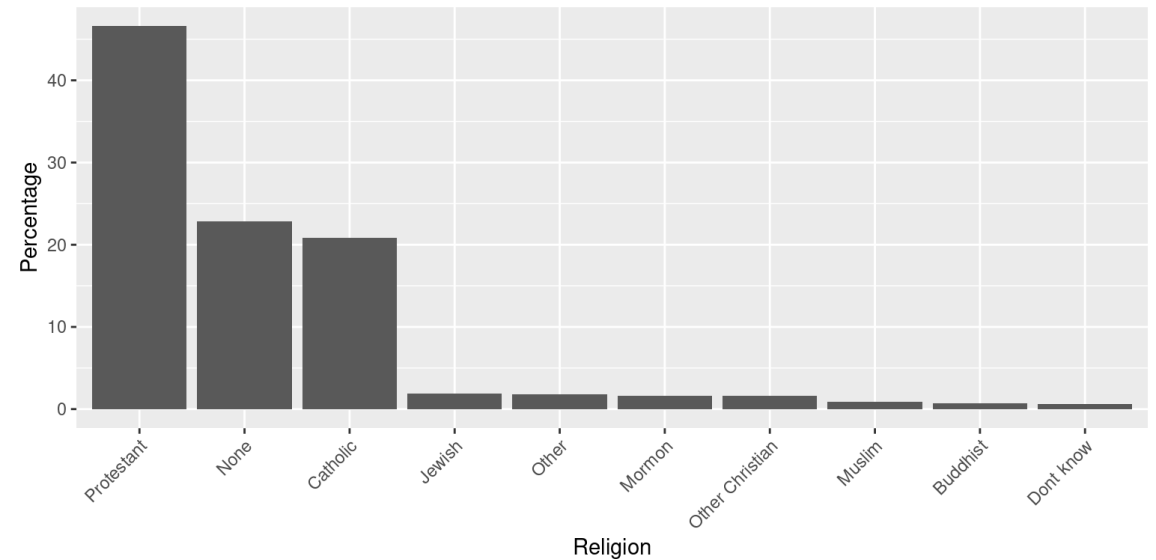
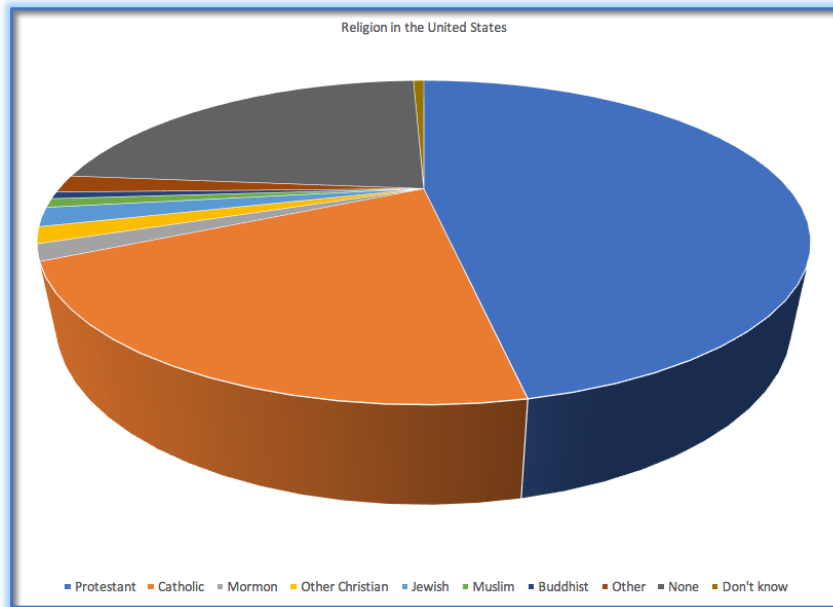


**B** lie factor ~ 2.8



# Guideline 4: Be mindful of human perception

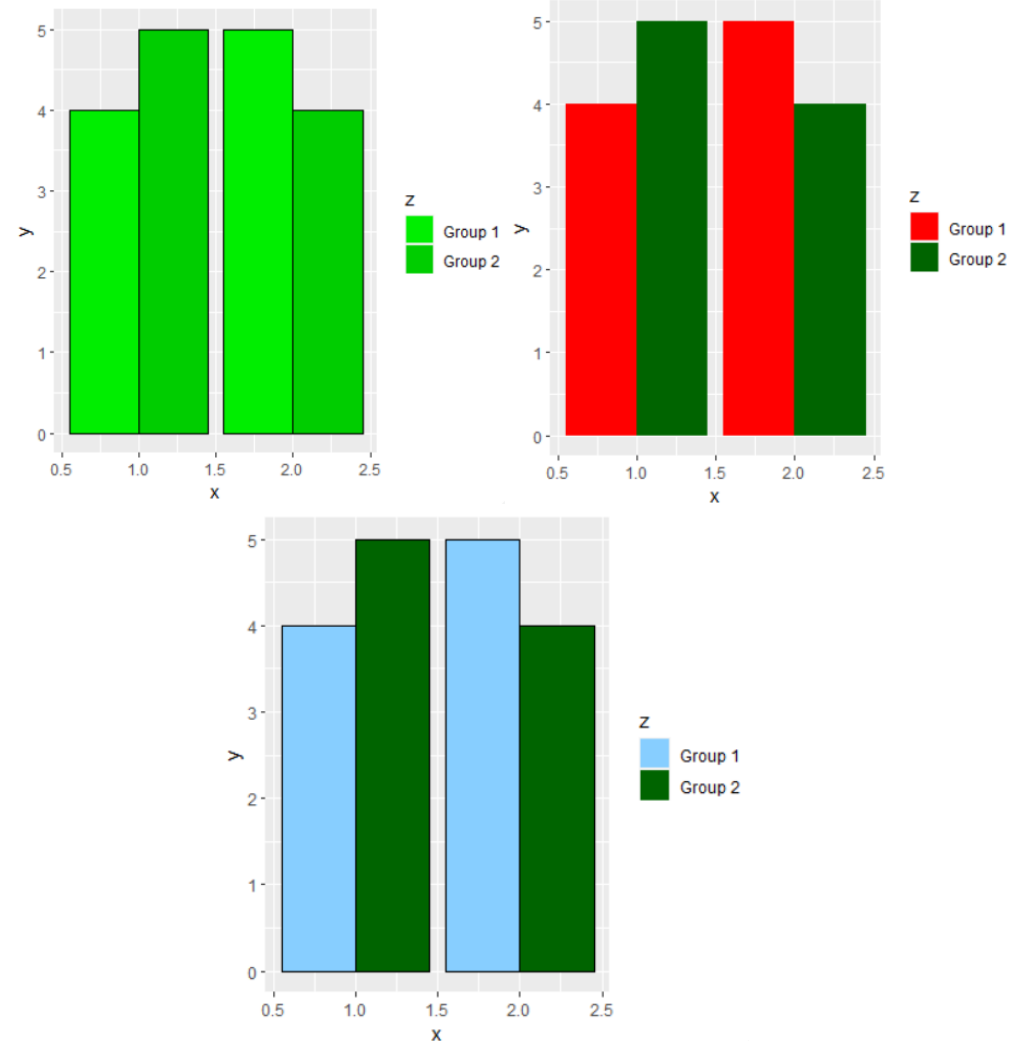
Don't show differences using volumes



# 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

