Introduction to Data & Numerical Summaries

Emile Latour, Nicky Wakim, Meike Niederhausen

January 5, 2026

Learning objectives (today)

By the end of class, you should be able to:

1. Define and distinguish a target population and a sample
2. Describe common sampling methods and why bias can happen
3. Compare experiments vs observational studies (and what each can conclude)

Why we care about study design

- Statistics helps answer questions with data.
- But the design determines what the data can support:
  - Who is included?
  - How are they selected?
  - What kind of study is it?
  - What comparisons are valid?

These decisions determine what conclusions we can (and cannot) draw.

Data collection principles (1.3)

- Population vs. sample
- Sampling methods
- Experiments vs. Observational studies

Population vs. sample

(Target) Population

- Group of interest being studied
- Group from which the sample is selected
  - studies often have *inclusion* and/or *exclusion* criteria
- Almost always too expensive or logistically impossible to collect data for every case in a population

Sample

- Group on which data are collected
- A subset (of measurements) from the population

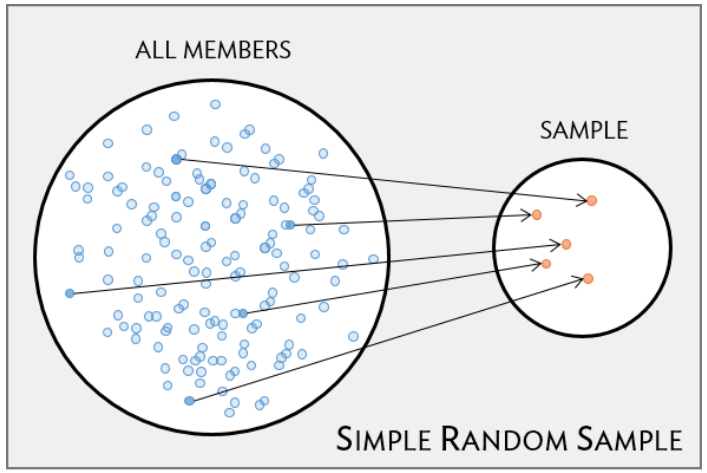
We use information from a sample to learn about the population from which it was drawn.

Sampling methods (1/4)

A good sampling method produces a **representative** sample: one whose characteristics are similar to those of the population.

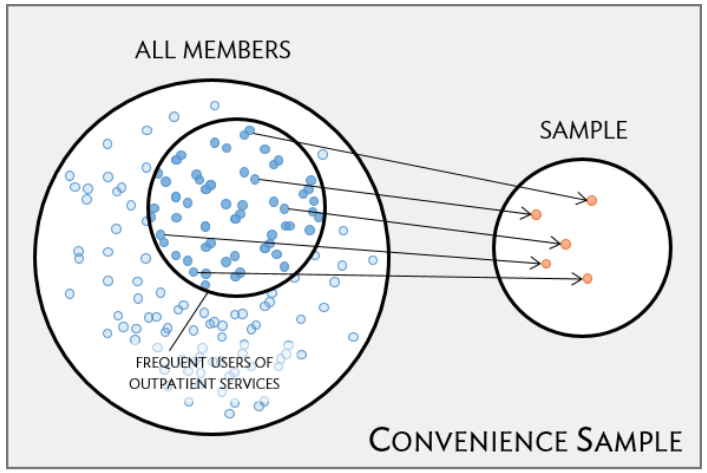
Simple random sample (SRS)

- Each individual of a population has the *same chance* of being sampled
- Randomly sampled
- Considered best way to sample



Convenience sample

- Easily accessible individuals are *more likely* to be included in the sample than other individuals
- Considered a common “pitfall”



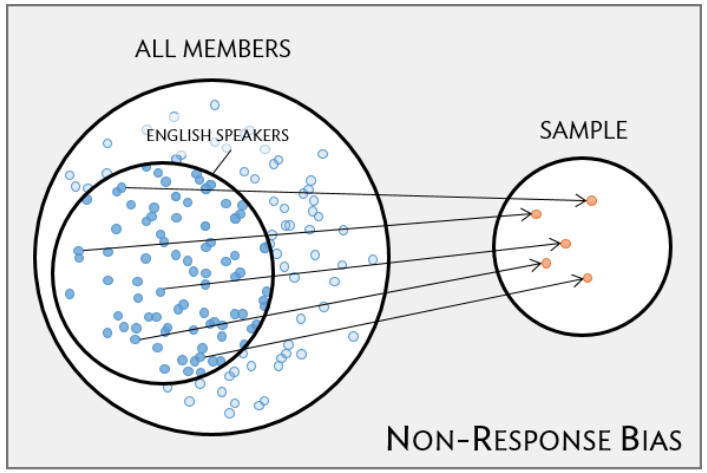
Sampling methods (2/4): Reality check

Even good sampling plans don't guarantee representative samples

Non-response bias

- Non-response rates can be high
- Are all groups within a population being reached?
- Unrepresentative sample

Can lead to skewed results

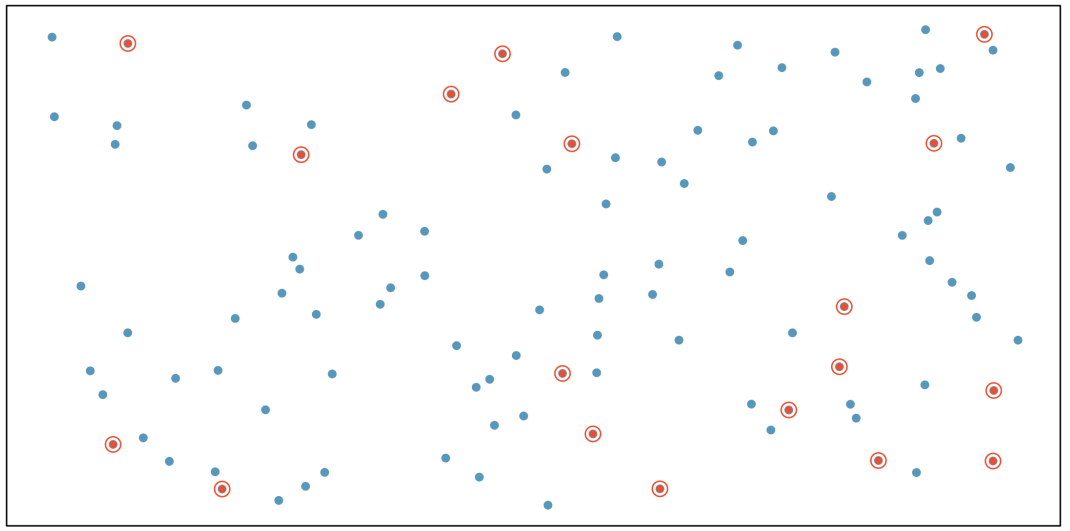


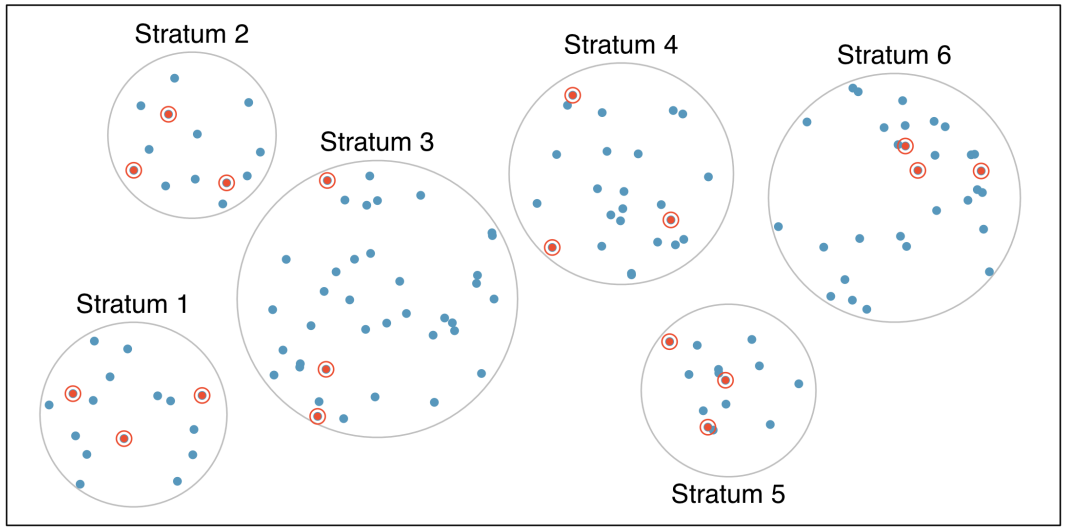
"Random" samples can be unrepresentative by random chance

- In a SRS each case in the population has an equal chance of being included in the sample
- But by random chance alone a random sample might contain a higher proportion of one group over another
- Ex: a SRS might by chance include 70% men (unlikely, but theoretically possible)

Sampling methods (3/4)

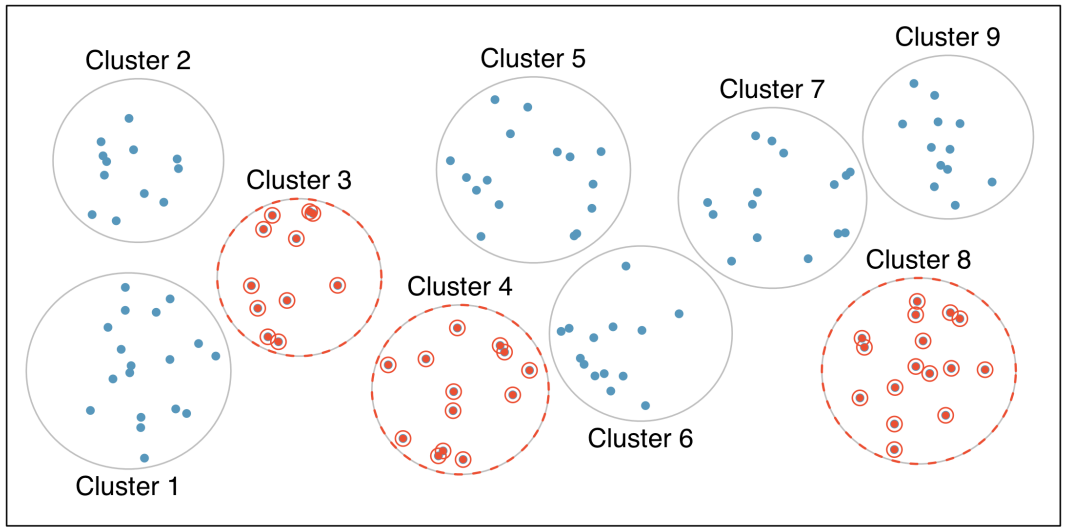
- **Simple random sample (SRS)**
  - Each individual of a population has the *same chance* of being sampled
  - *Statistical methods taught in this class assume a SRS!*
- **Stratified sampling**
  - Divide population into groups (strata) before selecting cases within each stratum (often via SRS)
  - Usually cases within a strata are similar, but are different from other strata with respect to the outcome of interest, such as gender or age groups

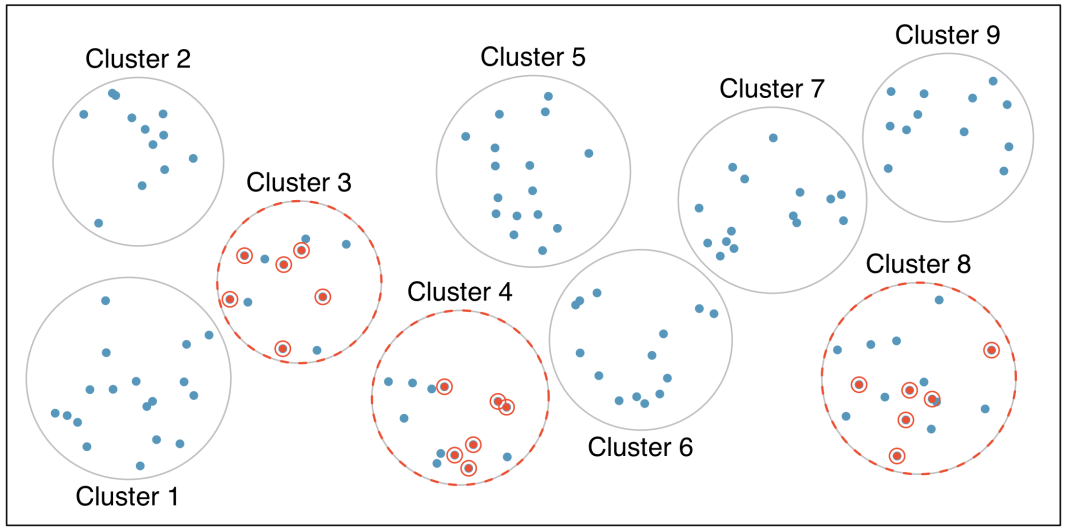




Sampling methods (4/4)

- **Cluster sample**
  - First divide population into groups (clusters)
  - Then sample a fixed number of clusters, and include *all* observations from chosen clusters
  - Clusters are often hospitals, clinicians, schools, etc., where each cluster will have similar services/ policies/ etc.
  - Cases within clusters usually very diverse
  - Example: sample zip codes in Oregon, then include **all** households in the sampled zip codes.
- **Multistage sample**
  - Similar to a cluster sample, but randomly sample individuals within each selected cluster instead of including everyone
  - Example: sample zip codes in Oregon, then randomly sample households in the zip codes (i.e. include **some** households in the sampled zip codes).





Two basic study designs

Experiment

Researchers directly influence how data arise

- Such as: assigning groups of individuals to different treatments and assessing how the outcome varies across treatment groups
- Three major parts to an experiment
  - Control
  - Randomization
  - Replication

Observational study

Researchers merely observe and record data, without interfering with how the data arise

- For example, to investigate why certain diseases develop, researchers might collect data by conducting surveys, reviewing medical records, or following a cohort of many similar individuals.
- Often the only available way to study your research question
  - Due to ethical considerations, funds, or availability of data

Experiments (1/2)

- **Control**
  - Researchers limit variability by carefully selecting participants
  - Inclusion and exclusion criteria reduce the influence of extraneous variables
  - Helps ensure the sample is appropriate and relevant to the research question
- **Randomization**
  - Group assignment is usually random to ensure similar (balanced) study arms for all variables (observed and unobserved)
  - Randomization allows study arm differences in outcomes to be attributed to treatment rather than variability in patient characteristics
    - Treatment is the only systematic difference between groups
    - Establish causality
  - **Blocking (stratification)**: group individuals into blocks (strata) before randomizing if there are certain characteristics that may influence the outcome other than treatment (i.e. gender, age group)
  - **Different than random sampling**: sampling decides *who enters the study*; randomization decides *which treatment they receive*

Experiments (2/2)

- **Replication**
  - Accomplished by collecting a sufficiently large sample
  - Results usually more reliable with a large sample size
    - Often less variability
    - More likely to be representative of population

Observational studies

Some research questions **cannot be studied experimentally** due to ethical, practical, or logistical constraints.

- Data are observed and recorded without interference
  - Researchers do not assign treatments or exposures
- Often done via surveys, electronic health records, or medical chart reviews
- Associations between variables can be established, **but not causality**
  - Individuals with different characteristics may also differ in other ways that influence response
- Confounding variables (lurking variable)
  - Variables associated with both the explanatory and response variables

Observational studies: prospective vs. retrospective studies

Many observational studies follow **cohorts** — groups of individuals observed over time.

Prospective

- Identifies participants and collects information **at scheduled times or as events unfold**.

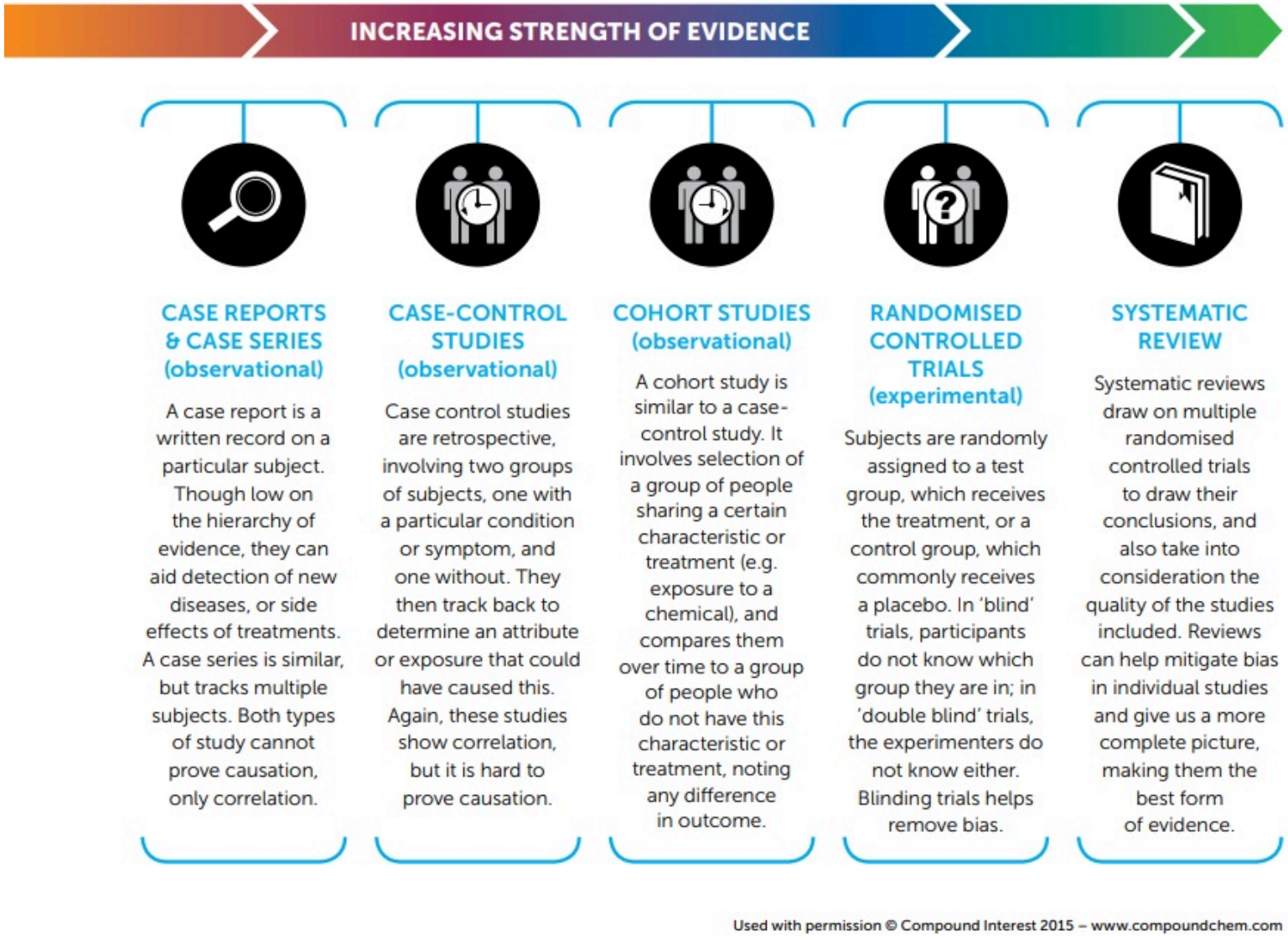
Retrospective

- Collect data **after events have taken place**, such as from medical records

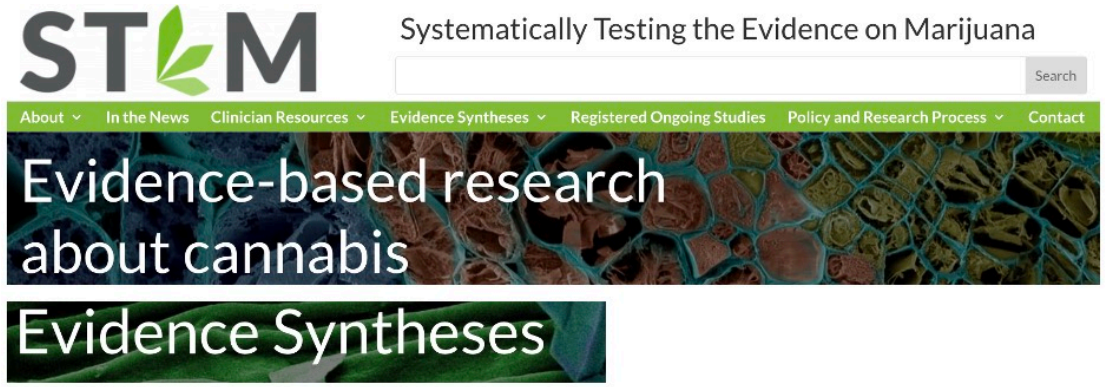
Some studies can have prospective and retrospective data.

Example: The Cancer Care Outcomes Research and Surveillance Consortium (CanCORS) enrolled participants with lung or colorectal cancer, collected information about diagnosis, treatment, and previous health behavior (retrospective), but also maintained contact with participants to gather data about long-term outcomes (prospective).

Comparing study designs



Systematic Reviews example



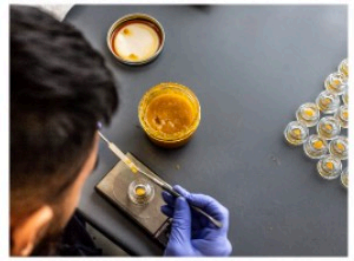
STEM is an independent, methodologically rigorous, and updated cannabis evidence resource for the health care sector that synthesizes what is known from research and what is left to learn about the health effects of cannabis.

Cannabis Use in Pregnancy

Cannabis for Chronic Pain

Cannabis for PTSD

Pharmaceuticals for Cannabis Use Disorder



Prevalence and Incidence of Cannabis Use Disorder

STEM: Systematically Testing the Evidence on Marijuana

STEM is a collaborative project between the US Department of Veterans Affairs and the **Center for Evidence-based Policy** at Oregon Health & Science University.
The project is funded by the US Department of Veterans Affairs: Office of Rural Health.

From study design to data

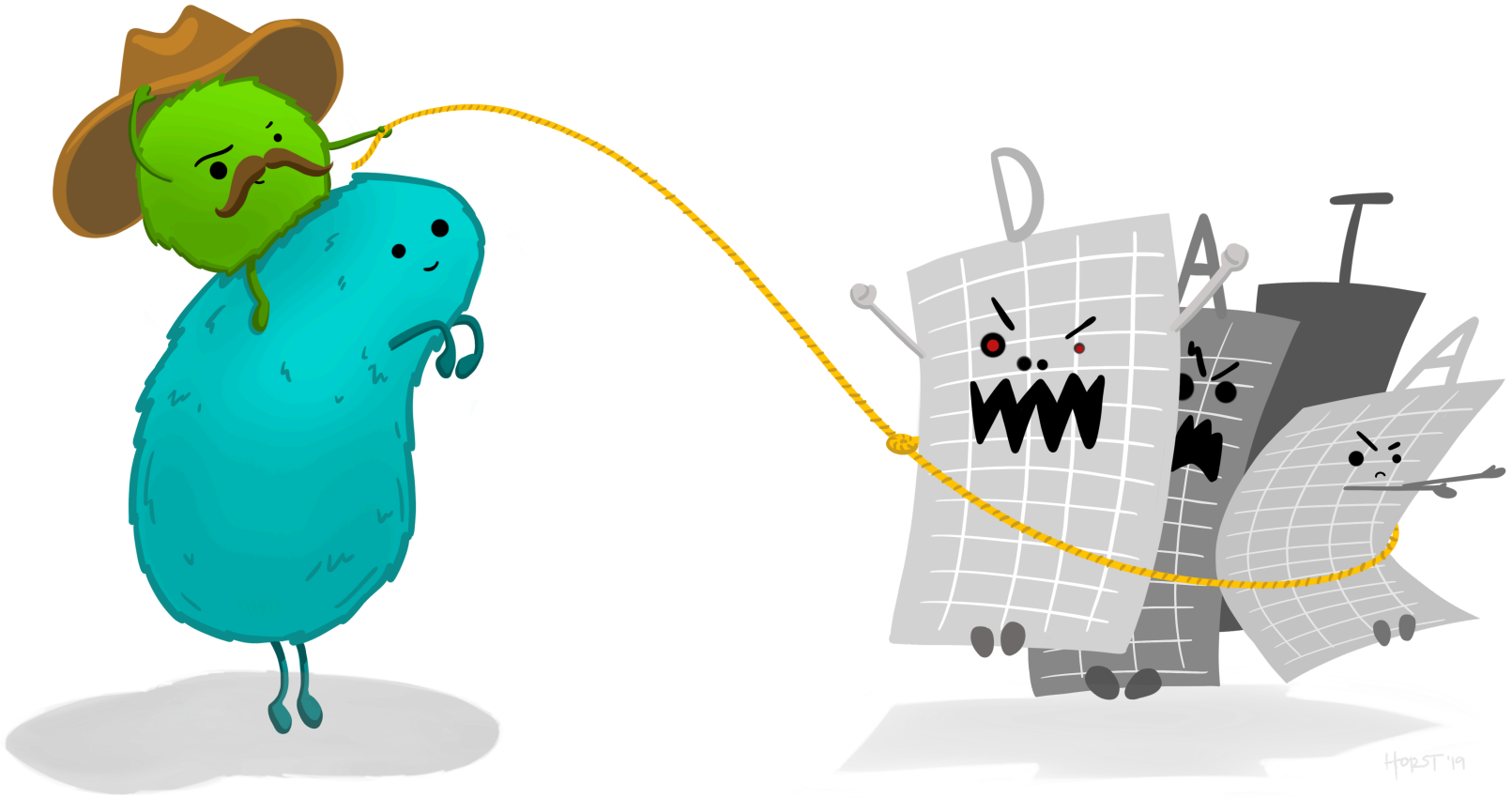
So far, we've focused on:

- how studies are designed
- how data are collected

Now we shift to:

- what data look like once we have them
- how data are stored and summarized

Intro to Data (1.2)



Artwork by @allison_horst

A first look at data in R

- Today is about *exposure*, not mastery
- We will revisit all of this slowly
- Right now: focus on concepts, not syntax

How are data stored, how do we use them?

- Often, data are in an Excel sheet, or a plain text file (.csv, .txt)
- .csv files open in Excel automatically, but actually are plain text
- Usually, columns are variables/measures and rows are observations (i.e. a person's measurements)

Data in R

- We can import data from many file types, including .csv, .txt., and .xlsx
  - We will cover this on a later date
- Once imported, R typically stores data as **data frames**, or **tibbles** if using the `tidyverse` package (more on this later).
  - For our purposes, these are essentially the same, and I will tend to use the terms interchangeably.
  - These are examples of what we call **object types** in R.

Data frame example

```r
df <- data.frame(
  IDs=1:3,
  gender=c("male", "female", "Male"),
  age=c(28, 35.5, 31),
  trt = c("control", "1", "1"),
  Veteran = c(FALSE, TRUE, TRUE)
  )
df
```

```
  IDs gender  age     trt Veteran
1   1   male 28.0 control   FALSE
2   2 female 35.5       1    TRUE
3   3   Male 31.0       1    TRUE
```

- **Vectors** vs. **data frames**
  - a data frame is a collection (or array or table) of vectors
- Different columns can be of different data types (i.e. numeric vs. text)
- Both numeric and text can be stored within a column (stored together as *text*).
- Vectors and data frames are examples of ***objects*** in R.
  - There are other types of R objects to store data, such as matrices, lists.

Observations & variables

```
1 df
```

```
  IDs gender  age     trt Veteran
1   1   male 28.0 control   FALSE
2   2 female 35.5       1    TRUE
3   3   Male 31.0       1    TRUE
```

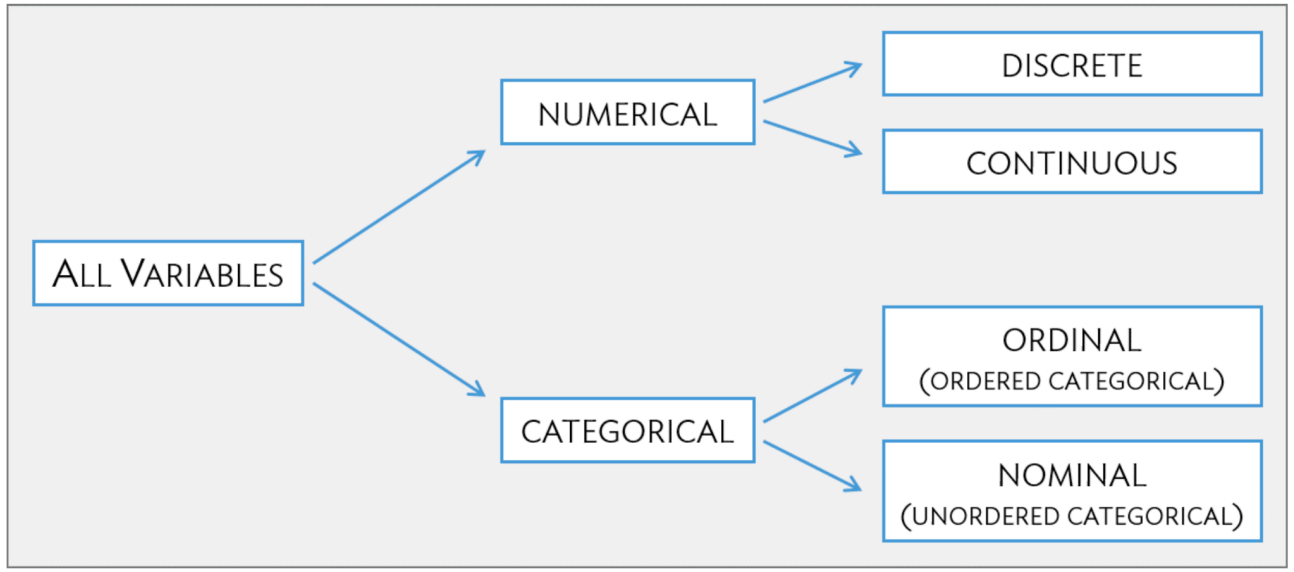


Figure 1.8: Breakdown of variables into their respective types.

ISLBS

- Book refers to a dataset as a *data matrix*
- Rows are usually **observations**
- Columns are usually **variables**
- **How many observations are in this dataset?**
- **What are the variable types in this dataset?**

Variable (column) types

R type	variable type	description
integer	discrete	integer-valued numbers
double or numeric	continuous	numbers that are decimals
factor	categorical	categorical variables stored with levels (groups)
character	categorical	text, “strings”
logical	categorical	boolean (TRUE, FALSE)

- View the **structure** of our data frame to see what the variable types are:

```
1 str(df)
```

```
'data.frame':   3 obs. of  5 variables:
 $ IDs    : int  1 2 3
 $ gender : chr  "male" "female" "Male"
 $ age    : num  28 35.5 31
 $ trt    : chr  "control" "1" "1"
 $ Veteran: logi  FALSE TRUE TRUE
```

Fisher's (or Anderson's) Iris data set

Data description:

- n = 150
- 3 species of Iris flowers (Setosa, Virginica, and Versicolour)
  - 50 measurements of each type of Iris
- **variables**:
  - sepal length, sepal width, petal length, petal width, and species

Can the iris species be determined by these variables?

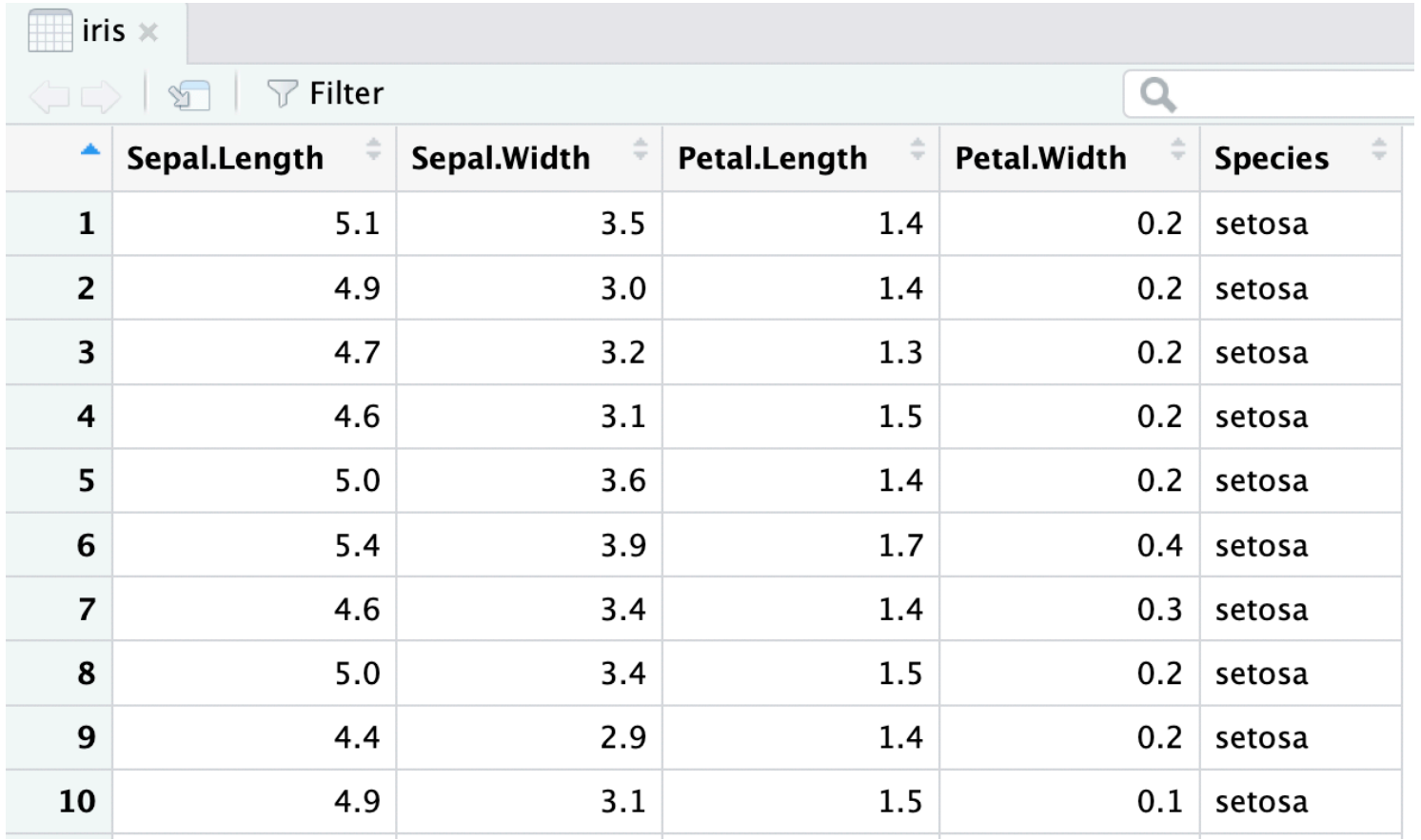
Gareth Duffy

View the `iris` dataset

- The `iris` dataset is already pre-loaded in *base* R and ready to use.

```
1 View(iris)
```

A new tab in the scripting window should appear with the `iris` dataset.



iris

Filter

	Sepal.Length	Sepal.Width	Petal.Length	Petal.Width	Species
1	5.1	3.5	1.4	0.2	setosa
2	4.9	3.0	1.4	0.2	setosa
3	4.7	3.2	1.3	0.2	setosa
4	4.6	3.1	1.5	0.2	setosa
5	5.0	3.6	1.4	0.2	setosa
6	5.4	3.9	1.7	0.4	setosa
7	4.6	3.4	1.4	0.3	setosa
8	5.0	3.4	1.5	0.2	setosa
9	4.4	2.9	1.4	0.2	setosa
10	4.9	3.1	1.5	0.1	setosa

Data structure

- What are the different **variable types** in this data set?

```r
str(iris)   # structure of data
```

```
'data.frame':   150 obs. of  5 variables:
 $ Sepal.Length: num  5.1 4.9 4.7 4.6 5 5.4 4.6 5 4.4 4.9 ...
 $ Sepal.Width : num  3.5 3 3.2 3.1 3.6 3.9 3.4 3.4 2.9 3.1 ...
 $ Petal.Length: num  1.4 1.4 1.3 1.5 1.4 1.7 1.4 1.5 1.4 1.5 ...
 $ Petal.Width : num  0.2 0.2 0.2 0.2 0.2 0.4 0.3 0.2 0.2 0.1 ...
 $ Species     : Factor w/ 3 levels "setosa","versicolor",..: 1 1 1 1 1 1 1 1 1 1 ...
```

Data set summary

```
1 summary(iris)
```

```
 Sepal.Length    Sepal.Width     Petal.Length    Petal.Width
Min.   :4.300   Min.   :2.000   Min.   :1.000   Min.   :0.100
1st Qu.:5.100   1st Qu.:2.800   1st Qu.:1.600   1st Qu.:0.300
Median :5.800   Median :3.000   Median :4.350   Median :1.300
Mean   :5.843   Mean   :3.057   Mean   :3.758   Mean   :1.199
3rd Qu.:6.400   3rd Qu.:3.300   3rd Qu.:5.100   3rd Qu.:1.800
Max.   :7.900   Max.   :4.400   Max.   :6.900   Max.   :2.500
      Species
setosa    :50
versicolor:50
virginica :50
```

Data set info

```r
dim(iris)
```

```
[1] 150   5
```

```r
nrow(iris)
```

```
[1] 150
```

```r
ncol(iris)
```

```
[1] 5
```

```r
names(iris)
```

```
[1] "Sepal.Length" "Sepal.Width"  "Petal.Length" "Petal.Width"  "Species"
```

View the beginning or end of a dataset

```
1 head(iris)
```

```
  Sepal.Length Sepal.Width Petal.Length Petal.Width Species
1          5.1         3.5          1.4         0.2  setosa
2          4.9         3.0          1.4         0.2  setosa
3          4.7         3.2          1.3         0.2  setosa
4          4.6         3.1          1.5         0.2  setosa
5          5.0         3.6          1.4         0.2  setosa
6          5.4         3.9          1.7         0.4  setosa
```

```
1 tail(iris)
```

```
    Sepal.Length Sepal.Width Petal.Length Petal.Width   Species
145          6.7         3.3          5.7         2.5 virginica
146          6.7         3.0          5.2         2.3 virginica
147          6.3         2.5          5.0         1.9 virginica
148          6.5         3.0          5.2         2.0 virginica
149          6.2         3.4          5.4         2.3 virginica
150          5.9         3.0          5.1         1.8 virginica
```

Specify how many rows to view at beginning or end of a dataset

```r
head(iris, 3)
```

```
  Sepal.Length Sepal.Width Petal.Length Petal.Width Species
1          5.1         3.5          1.4         0.2  setosa
2          4.9         3.0          1.4         0.2  setosa
3          4.7         3.2          1.3         0.2  setosa
```

```r
tail(iris, 2)
```

```
    Sepal.Length Sepal.Width Petal.Length Petal.Width   Species
149          6.2         3.4          5.4         2.3 virginica
150          5.9         3.0          5.1         1.8 virginica
```

The $

- Suppose we want to single out the column of petal width values.
- One way to do this is to use the `$`
  - `DatSetName$VariableName`

```
1 iris$Petal.Width
```

```
  [1] 0.2 0.2 0.2 0.2 0.2 0.4 0.3 0.2 0.2 0.1 0.2 0.2 0.1 0.1 0.2 0.4 0.4 0.3
 [19] 0.3 0.3 0.2 0.4 0.2 0.5 0.2 0.2 0.4 0.2 0.2 0.2 0.2 0.4 0.1 0.2 0.2 0.2
 [37] 0.2 0.1 0.2 0.2 0.3 0.3 0.2 0.6 0.4 0.3 0.2 0.2 0.2 0.2 1.4 1.5 1.5 1.3
 [55] 1.5 1.3 1.6 1.0 1.3 1.4 1.0 1.5 1.0 1.4 1.3 1.4 1.5 1.0 1.5 1.1 1.8 1.3
 [73] 1.5 1.2 1.3 1.4 1.4 1.7 1.5 1.0 1.1 1.0 1.2 1.6 1.5 1.6 1.5 1.3 1.3 1.3
 [91] 1.2 1.4 1.2 1.0 1.3 1.2 1.3 1.3 1.1 1.3 2.5 1.9 2.1 1.8 2.2 2.1 1.7 1.8
[109] 1.8 2.5 2.0 1.9 2.1 2.0 2.4 2.3 1.8 2.2 2.3 1.5 2.3 2.0 2.0 1.8 2.1 1.8
[127] 1.8 1.8 2.1 1.6 1.9 2.0 2.2 1.5 1.4 2.3 2.4 1.8 1.8 2.1 2.4 2.3 1.9 2.3
[145] 2.5 2.3 1.9 2.0 2.3 1.8
```

Example using the $

The $ is helpful if you want to create a new dataset for just that one variable, or, more commonly, if you want to calculate summary statistics for that one variable.

```
1 mean(iris$Petal.Width)
```

```
[1] 1.199333
```

```
1 sd(iris$Petal.Width)
```

```
[1] 0.7622377
```

```
1 median(iris$Petal.Width)
```

```
[1] 1.3
```

Inline code

- With markdown you can also report **R code output inline** with the text instead of using a chunk.

Text in editor:

```
The mean petal width for all 3 species combined
is `r round(mean(iris$Petal.Width),1)`
(SD = `r round(sd(iris$Petal.Width),1)`) cm.
```

Output:

The mean petal width for all 3 species combined is 1.2 (SD = 0.8) cm.

- Reporting summary statistics this way in a report, makes the numbers computationally reproducible.
- For example, if this were for an abstract and a year later you are wondering where the numbers came from, your R code will tell you exactly which dataset was used to calculate the values.

Summarizing numerical data (1.4)

Measures of center & spread

THE PROBLEM WITH AVERAGING STAR RATINGS

https://xkcd.com/937/

Table 1 example

Table 1. Patient characteristics, overall and by concordance

		Total	Discordant	Concordant	p-value
		N=204	N=40	N=164	
Site, n (%)	OHSU	122 (62.7%)	26 (65.0%)	96 (62.2%)	0.86
	VA	76 (37.3%)	14 (35.0%)	62 (37.8%)	
Gender, n (%)	Male	85 (41.7%)	18 (45.0%)	67 (40.9%)	0.72
	Female	119 (58.3%)	22 (55.0%)	97 (59.1%)	
Age (years), mean (SD)		57.2 (14.2)	58.2 (15.1)	56.9 (14.0)	0.62
Language, n (%)	English	168 (84.4%)	35 (92.1%)	133 (82.6%)	0.21
	Spanish	31 (15.6%)	3 (7.9%)	28 (17.4%)	
Limited English language proficiency, n (%)		30 (15.1%)	3 (7.9%)	27 (16.8%)	0.17
Coupled, n (%)		110 (57.9%)	22 (61.1%)	88 (57.1%)	0.71
Education, n (%)	High school or less	60 (31.6%)	15 (40.5%)	45 (29.4%)	0.24
	Some college or more	130 (68.4%)	22 (59.5%)	108 (70.6%)	
Income, >$40,000, n (%)	Less than $40,000	85 (45.5%)	12 (33.3%)	73 (48.3%)	0.14
	Greater than $40,000	102 (54.5%)	24 (66.7%)	78 (51.7%)	
People in household, median (IQR)		2 (2-4)	2 (2-3)	2 (2-4)	0.92
Race/Ethnicity, n (%)	White	123 (68.3%)	25 (78.1%)	98 (66.2%)	0.62
	Black	6 (3.3%)	0 (0.0%)	6 (4.1%)	
	Latinx/Hispanic	39 (21.7%)	6 (18.8%)	33 (22.3%)	
	Other	12 (6.7%)	1 (3.1%)	11 (7.4%)	
Limited health literacy, n (%)		55 (28.6%)	13 (35.1%)	42 (27.1%)	0.42
Disease duration (years), median (IQR)		8 (4-16)	13 (5-21)	7 (4-15)	0.039
Number of medications, median (IQR)		1 (1-2)	1 (0-2)	1 (1-2)	0.10
Depressive symptoms, n (%)		38 (20.8%)	3 (8.1%)	35 (24.0%)	0.040
PTSD, n (%)		13 (7.1%)	2 (5.6%)	11 (7.5%)	1.00
Self-efficacy score, mean (SD)		6.3 (2.1)	6.3 (2.1)	6.3 (2.1)	0.96
Trust in Physician, n (%)		106 (53.8%)	19 (51.4%)	87 (%)	0.74
Disease activity score (CDAI), mean (SD)		12.8 (10.5)	10.5 (9.7)	13.2 (10.8)	0.21
Medication Adherence, n (%)	High	63 (33.5%)	7 (20.6%)	56 (36.4%)	0.11

Are We on the Same Page?: A Cross-Sectional Study of Patient-Clinician Goal Concordance in Rheumatoid Arthritis
J Barton et al.
Arthritis Care & Research.
2021 Sep 27
https://pubmed.ncbi.nlm.nih.gov/34569172/

Measures of center: mean

Sample mean: the average value of observations

$$\bar{x} = \frac{x_1 + x_2 + \cdots + x_n}{n} = \sum_{i=1}^{n} \frac{x_i}{n}$$

where $x_1, x_2, \ldots, x_n$ represent the n observed values in a sample

Example: What is the mean age in the toy dataset `df` defined earlier?

```
1 df
```

```
  IDs gender  age     trt Veteran
1   1   male 28.0 control   FALSE
2   2 female 35.5       1    TRUE
3   3   Male 31.0       1    TRUE
```

```
1 mean(df$age)
```

```
[1] 31.5
```

Measures of center: median

- The **median** is the middle value of the observations in a sample.
- The median is the 50th percentile, meaning
  - 50% of observations lie below and
  - 50% of observations lie above the median.
- If the number of observations is
  - odd: the median is the middle observed value
  - even: the median is the average of the two middle observed values

```
1 df$age
```

```
[1] 28.0 35.5 31.0
```

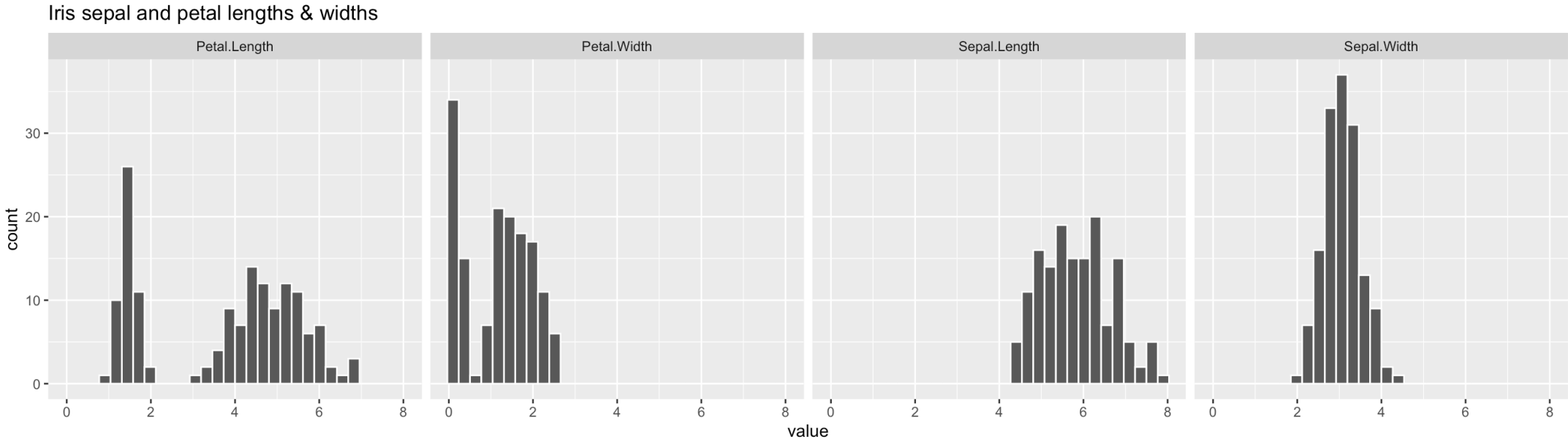
```
1 median(df$age)
```

```
[1] 31
```

```
1 median(c(df$age, 67))
```

```
[1] 33.25
```

Measures of center: mean vs. median

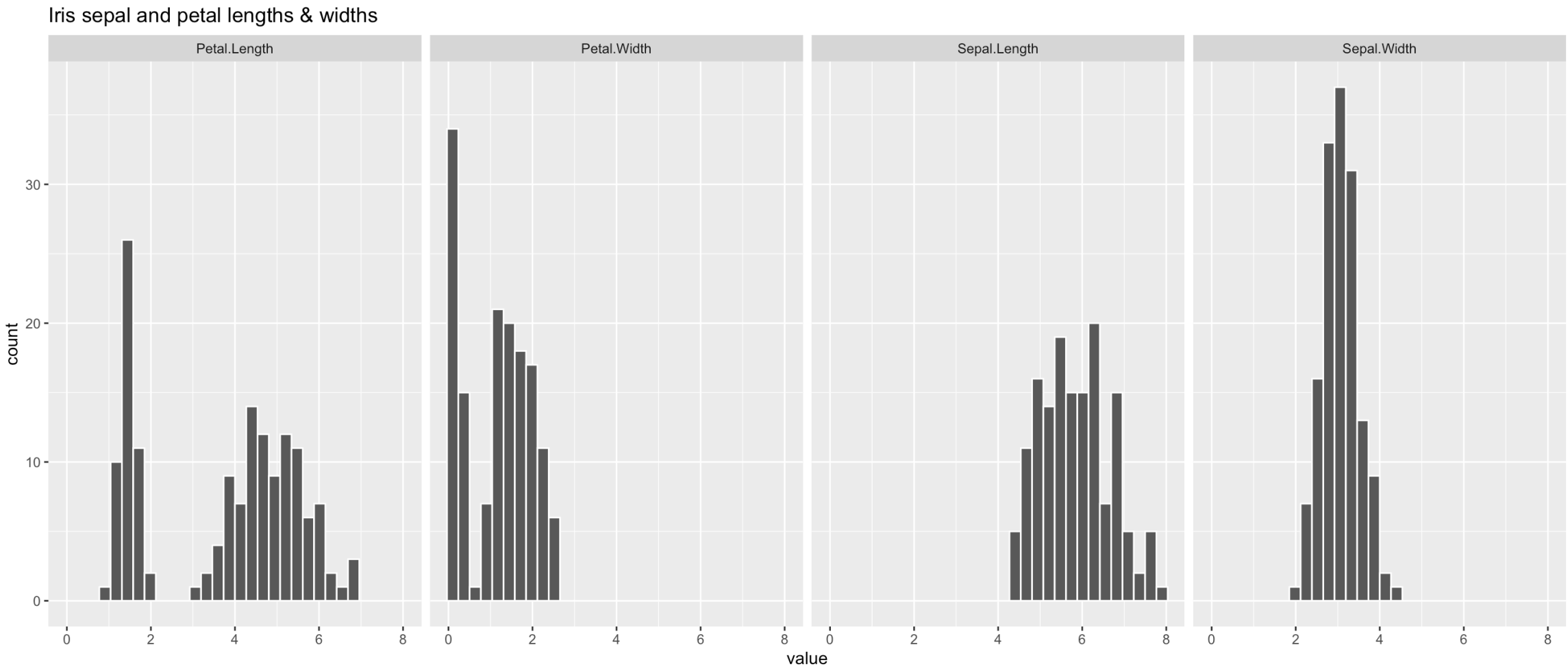



```
1 summary(iris)
```

```
  Sepal.Length    Sepal.Width     Petal.Length    Petal.Width
Min.   :4.300   Min.   :2.000   Min.   :1.000   Min.   :0.100
1st Qu.:5.100   1st Qu.:2.800   1st Qu.:1.600   1st Qu.:0.300
Median :5.800   Median :3.000   Median :4.350   Median :1.300
Mean   :5.843   Mean   :3.057   Mean   :3.758   Mean   :1.199
3rd Qu.:6.400   3rd Qu.:3.300   3rd Qu.:5.100   3rd Qu.:1.800
Max.   :7.900   Max.   :4.400   Max.   :6.900   Max.   :2.500
       Species
setosa    :50
versicolor:50
virginica :50
```

Measures of center: mode

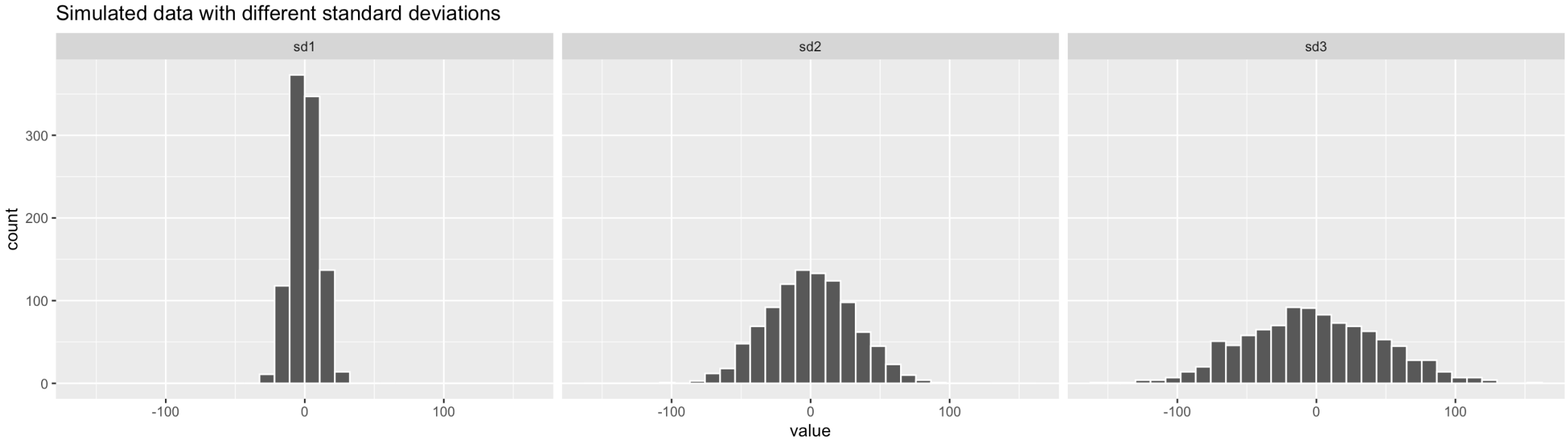
mode: the most frequent value in a dataset



Measures of spread: standard deviation (SD) (1/3)

standard deviation is (approximately) the average distance between a typical observation and the mean

- An observation's **deviation** is the distance between its value x and the sample mean $\bar{x}$: deviation = $x - \bar{x}$.



Measures of spread: SD (2/3)

- The **sample variance** s^2 is the sum of squared deviations divided by the number of observations minus 1.

$$s^2 = \frac{(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \cdots + (x_n - \bar{x})^2}{n-1} = \sum_{i=1}^{n} \frac{(x_i - \bar{x})^2}{n-1}$$

where $x_1, x_2, \ldots, x_n$ represent the n observed values.

- The **standard deviation** s (or sd) is the square root of the variance.

$$s = \sqrt{s^2} = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2}$$

Measures of spread: SD (3/3)

Let's calculate the sample standard deviation for our toy example

```
1 df$age
```

```
[1] 28.0 35.5 31.0
```

$$s = \sqrt{\sum_{i=1}^{n} \frac{(x_i - \bar{x})^2}{n-1}} =$$

```
1 mean(df$age)
```

```
[1] 31.5
```

```
1 sd(df$age)
```

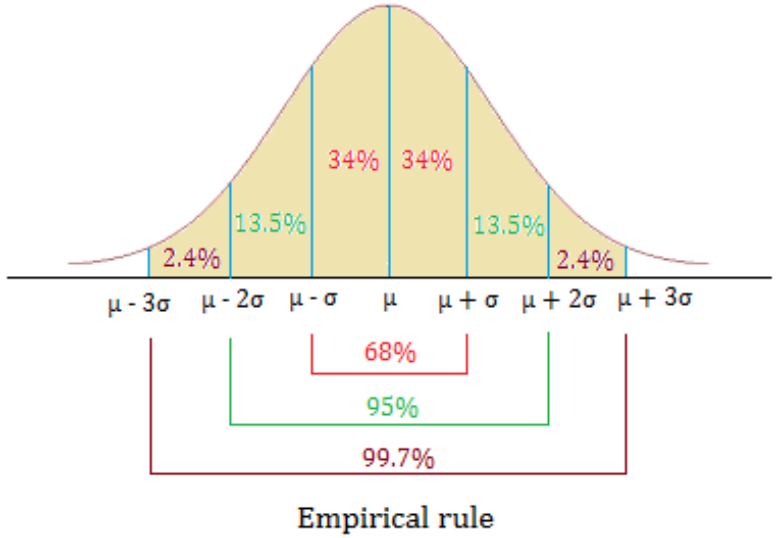
```
[1] 3.774917
```

Empirical Rule: one way to think about the SD (1/2)

For symmetric bell-shaped data, about

- 68% of the data are within 1 SD of the mean
- 95% of the data are within 2 SD's of the mean
- 99.7% of the data are within 3 SD's of the mean

These percentages are based off of percentages of a true normal distribution.

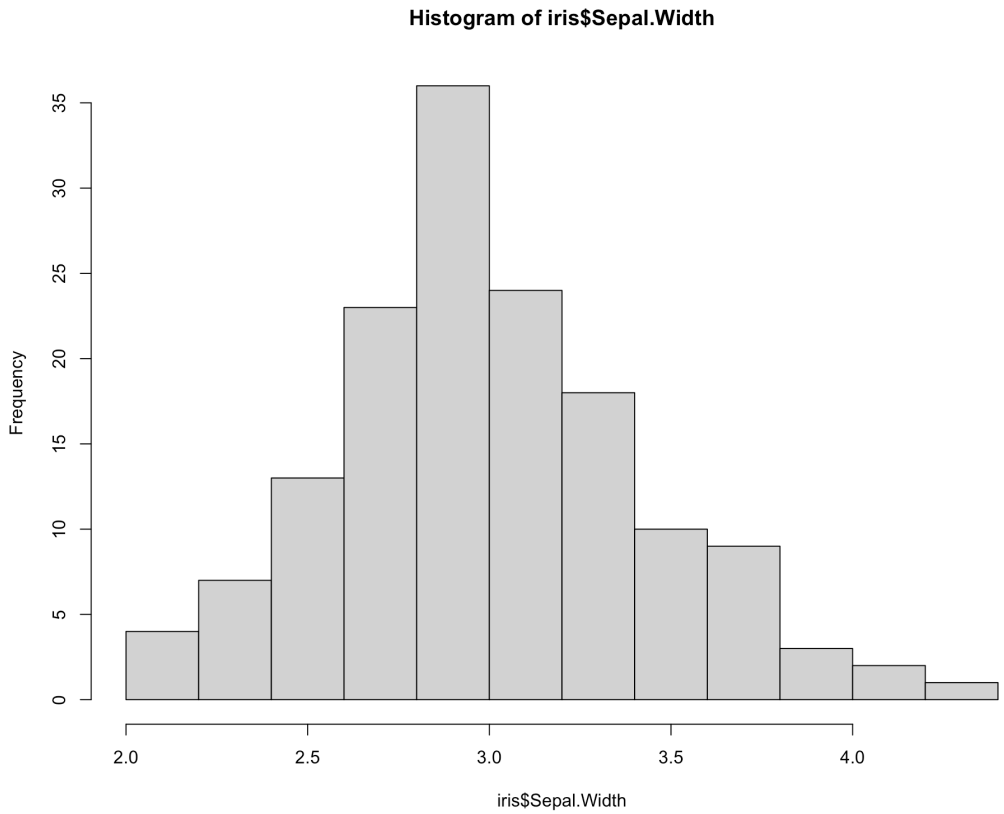


Empirical rule

https://statistics-made-easy.com/empirical-rule/

Empirical Rule: one way to think about the SD (2/2)

```
1 hist(iris$Sepal.Width)
```




```
1 mean(iris$Sepal.Width)
```

```
[1] 3.057333
```

```
1 sd(iris$Sepal.Width)
```

```
[1] 0.4358663
```

Measures of spread: interquartile range (IQR) (1/2)

The p^{th} percentile is the observation such that $p\%$ of the remaining observations fall below this observation.

- The *first quartile* Q_1 is the 25^{th} percentile.
- The *second quartile* Q_2, i.e., the median, is the 50^{th} percentile.
- The *third quartile* Q_3 is the 75^{th} percentile.

The **interquartile range (IQR)** is the distance between the third and first quartiles.

$$IQR = Q_3 - Q_1$$

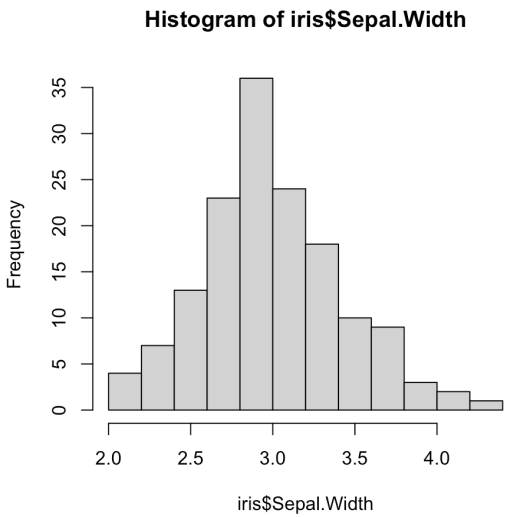
- IQR is the width of the *middle half* of the data

Measures of spread: IQR (2/2)

5 number summary

```
summary(iris$Sepal.Width)
```

```
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  2.000   2.800   3.000   3.057   3.300   4.400
```



What is the IQR of the sepal widths?

```
quantile(iris$Sepal.Width, c(.25, .75))
```

```
25% 75%
2.8 3.3
```

```
diff(quantile(iris$Sepal.Width, c(.25, .75)))
```

```
75%
0.5
```

```
IQR(iris$Sepal.Width)
```

```
[1] 0.5
```

Robust estimates

Summary statistics are called **robust estimates** if extreme observations (outliers) have little effect on their values

Estimate	Robust?
Sample mean	❌
Median	✅
Standard deviation	❌
IQR	✅
Range	❌

- For samples with extreme values or skewed distributions, the **median and IQR** often provide a more stable summary of center and spread than the mean, standard deviation, or range.
- The **range** depends only on the smallest and largest observations, so a single outlier can dramatically change its value.