Stat. Camp. Oct 6th

What is Statistics all about?

- Science of analysing, interpreting, presenting data
- Model ⟷ Real World

Statistics is about finding the balance between the two

Data - and how it came about - is absolutely crucial for producing results of value

Garbage in → Garbage out

- to generate information from data
- predictive ability

Let's focus on the data side:

- weather data, e.g. for climate models
- survey, like general census
- patient data
- numerical data ⟷ images or news paper corpus
  tabular data
  more complicated data structures
- banking or financial institution data
- operational data from online shops and other service provider

Data may have been collected for the purpose of being analysed for a specific research question, e.g. a dedicated study for a new drug

- should include all important aspects that are needed for answering the research question

- big effort / expensive / slow
- could go wrong for all sorts of biases, e.g. covid vaccination rate of students
  - students were asked in surveys, rates of 80% or even 90% vaccinated were reported
  - Real life: many instances of only 2/3 or even only 50% vaccinated

Assuming the survey targeted a <u>representative sample</u> of students or even <u>all</u> the students, bias will result from differential answering behavior.

Always take a skeptical look at your data and think about what might bias them.

That's even more important if your data were not specifically collected for the purpose of your research question.