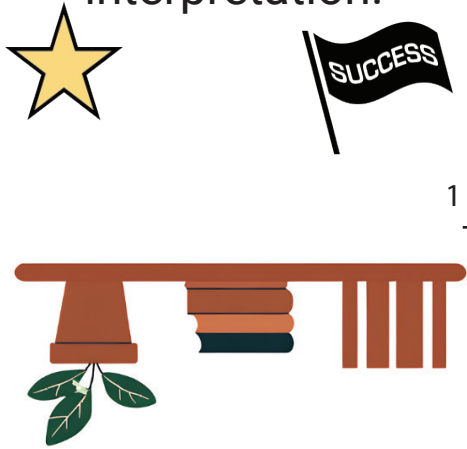


# What is IDA?

IDA (=Initial Data Analysis) aims to provide reliable knowledge about the data to enable valid statistical analyses and interpretation.



Initial Data Analysis (IDA)  
What? Why?



STATS

# Know your data

How was the data collected?

What is missing? (missing values, patterns of missingness)

Univariable descriptions of variables in the planned models

Multivariable descriptions

Timing of measurements or follow-up period



Marianne Huebner  
2025-06-02



# IDA results -oops!

Suspicious values (outliers, inconsistent time differences)

Suspicious rows (eligibility)

Missing data

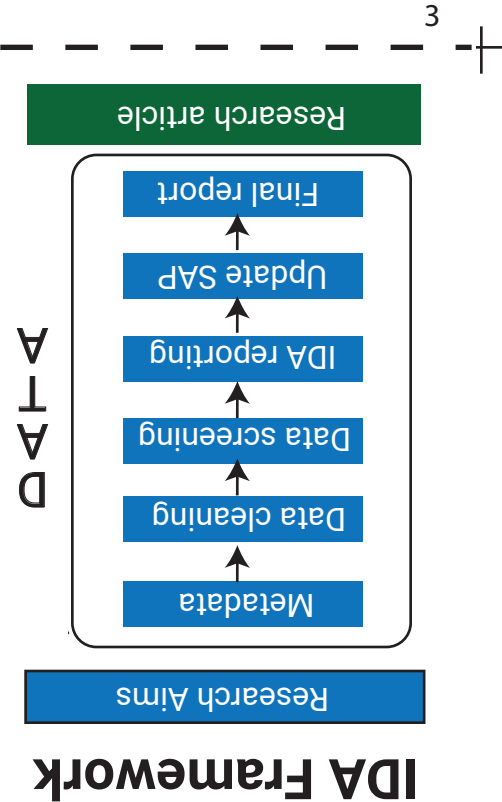
Unexpected distributions of variables



Unexpected heterogeneity of the analysis population

Highly correlated variables

Update analysis plan!



# Avoid sneak peaks

Stick to the research aims: IDA does not generate hypotheses ✓

Change the analysis plan based on data properties ✓

Explore associations with the outcome ⚡

Data driven selection of analyses ⚡

It saves more time later!!!



IDA takes time

repeatable

Make data cleaning code

Keep track of changes  
(data and program scripts)  
from ALL team members!  
Document metadata



Make IDA reproducible