

# Cleaning & Transforming Data

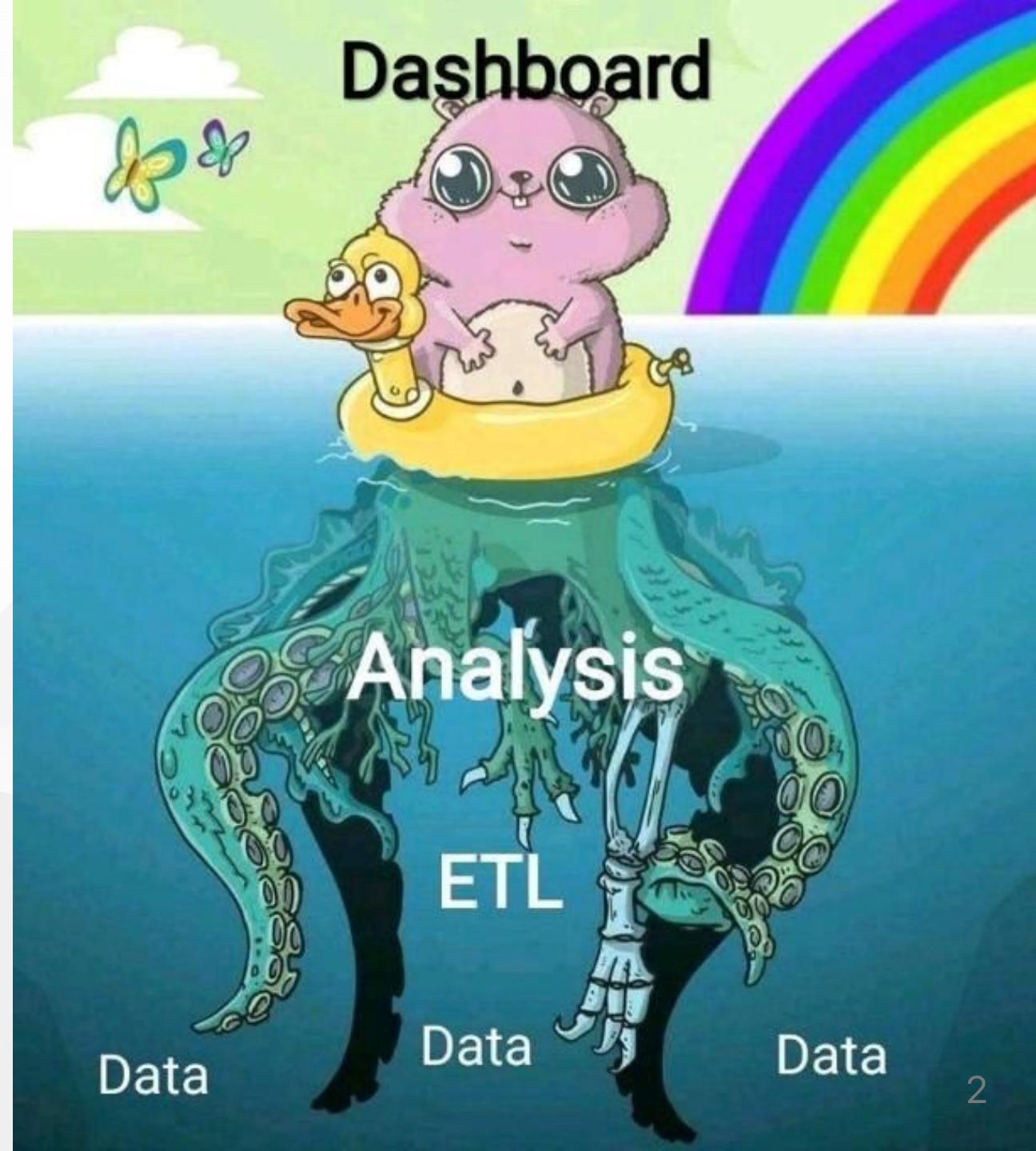
## Learning objectives

- Learn the principles of tidy data.
- Be familiar with some common data transformations.
- Practice cleaning data and using formulas in Tableau.

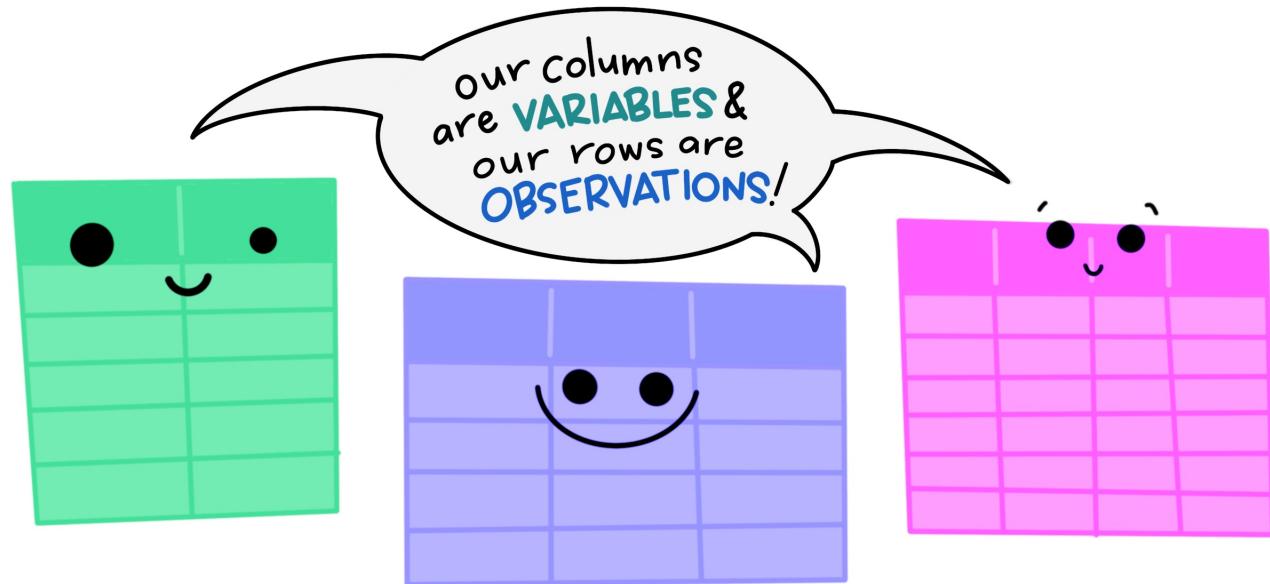
## Data Cleaning vs Data Transformation

**Data cleaning** is the process of fixing or removing incorrect, corrupted, incorrectly formatted, duplicate, or incomplete data within a dataset.

**Data transformation** is the process of converting data from one format or structure into another.

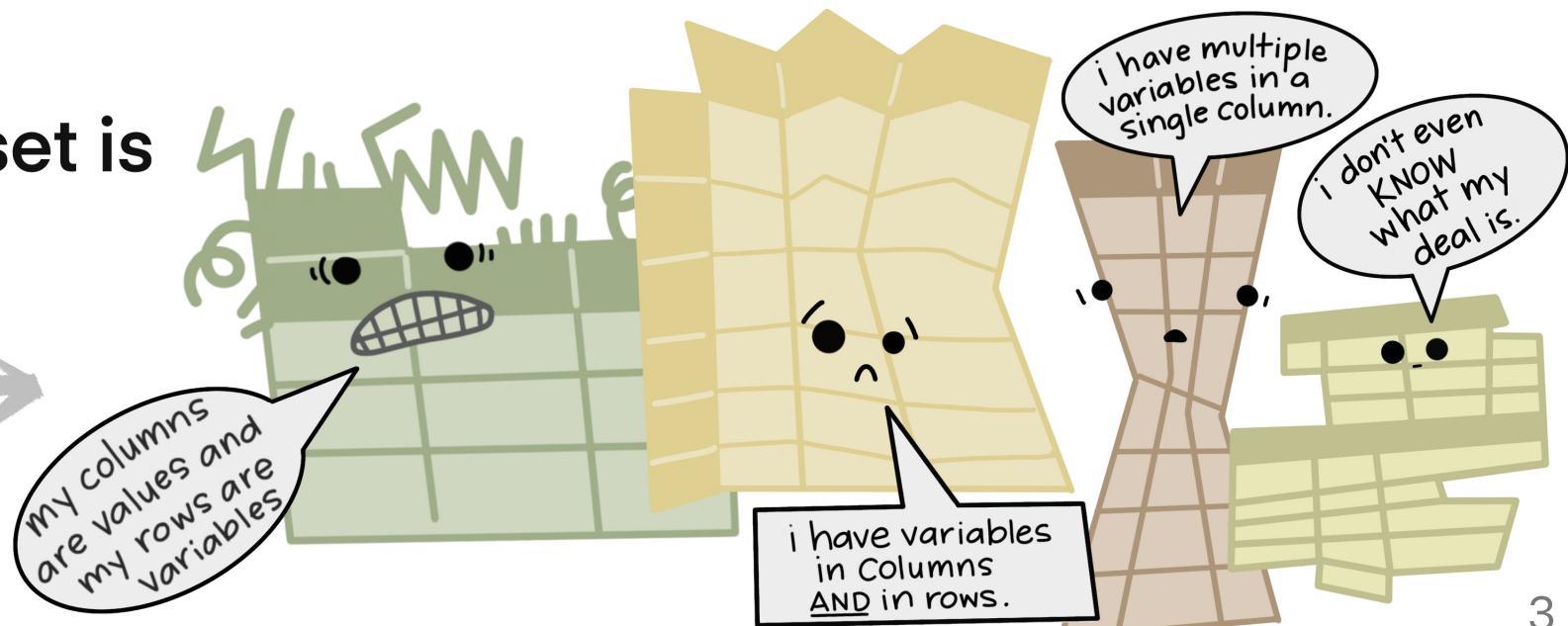


The standard structure of  
tidy data means that  
“tidy datasets are all alike...”



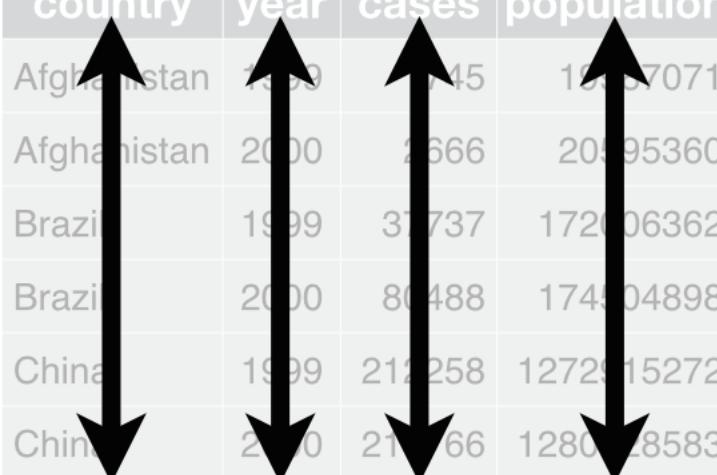
“...but every messy dataset is  
messy in its own way.”

-HADLEY WICKHAM



| country     | year | cases  | population |
|-------------|------|--------|------------|
| Afghanistan | 1999 | 745    | 19087071   |
| Afghanistan | 2000 | 2666   | 20595360   |
| Brazil      | 1999 | 37737  | 172006362  |
| Brazil      | 2000 | 80488  | 174504898  |
| China       | 1999 | 212258 | 1272915272 |
| China       | 2000 | 213766 | 128042583  |

variables



| country     | year | cases  | population |
|-------------|------|--------|------------|
| Afghanistan | 1999 | 745    | 19087071   |
| Afghanistan | 2000 | 2666   | 20595360   |
| Brazil      | 1999 | 37737  | 172006362  |
| Brazil      | 2000 | 80488  | 174504898  |
| China       | 1999 | 212258 | 1272915272 |
| China       | 2000 | 213766 | 128042583  |

observations



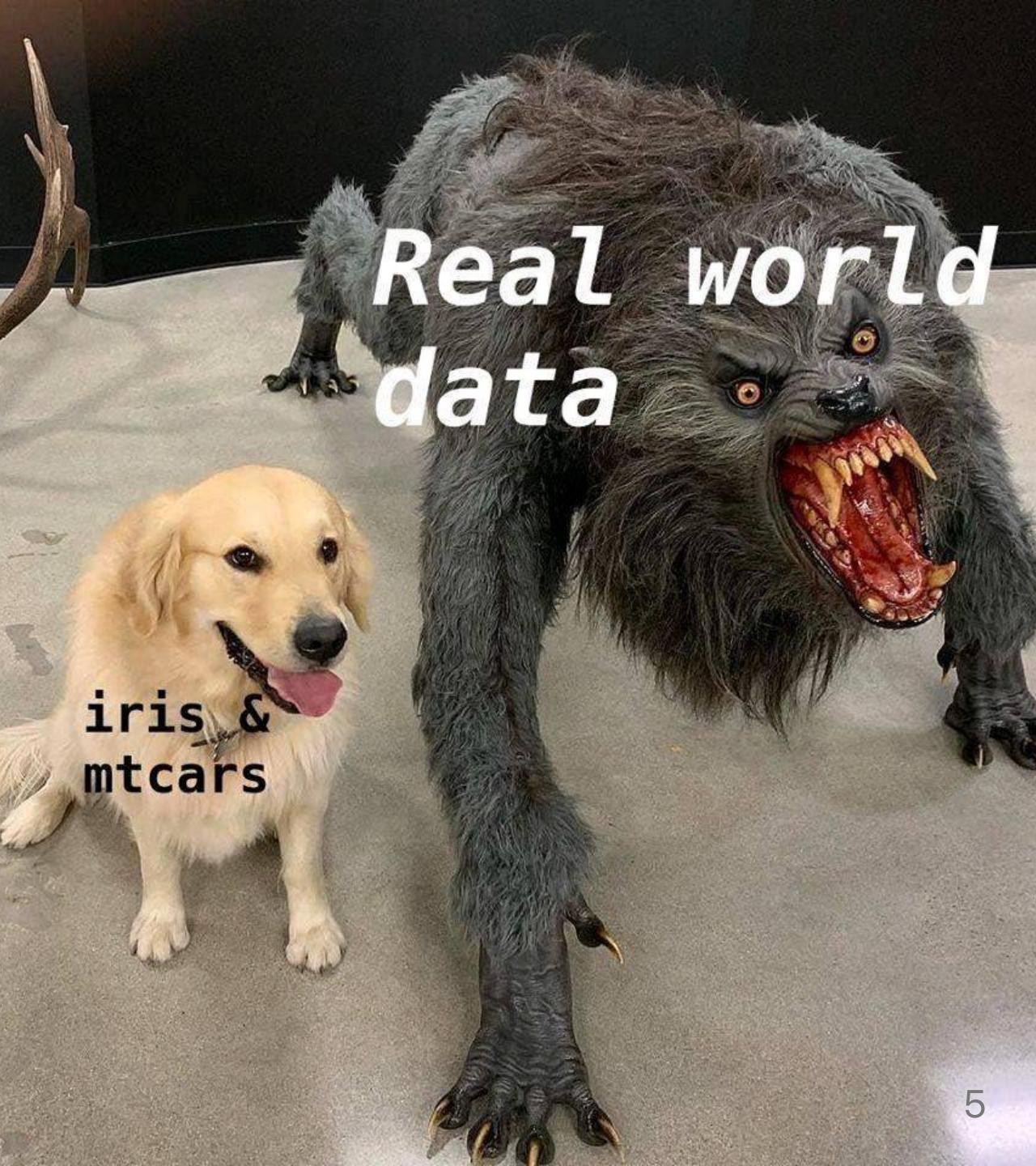
| country     | year | cases  | population |
|-------------|------|--------|------------|
| Afghanistan | 1999 | 745    | 19087071   |
| Afghanistan | 2000 | 2666   | 20595360   |
| Brazil      | 1999 | 37737  | 172006362  |
| Brazil      | 2000 | 80488  | 174504898  |
| China       | 1999 | 212258 | 1272915272 |
| China       | 2000 | 213766 | 128042583  |

values



## Data Cleaning

- Filter irrelevant rows.
- Remove unnecessary columns.
- Filter duplicates.
- Fix structural errors.
- Standardize labels.
- Handle outliers.
- Handle missing data.
- Validate data.



## Data Transformation

- Adding Columns.
- Splitting Columns.
- Pivoting & Unpivoting.
- Detect Data Type.
- Aggregate.
- Sort by Column.
- Concatenate Columns.

