|  |
| --- |
| Course: Data Science [Z15811] Class: Data Wrangling [04]  Subject: Discussion topics |

1. What is a flat file?

Plain text file, one record per line, consistent structure.

1. A tidy dataset has to follow 3 basic rules. Which rules?

Each variable must have its own column.

Each observation must have its own row.

Each value must have its own cell.

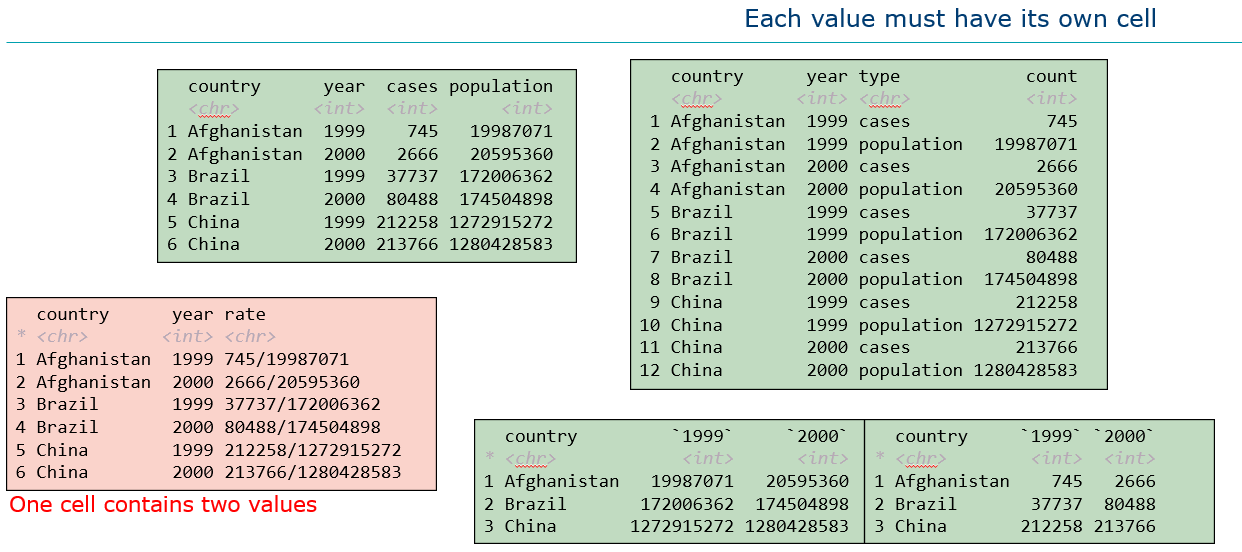
1. Give 3 examples of an untidy dataset and explain why they are untidy.

Explain what steps are necessary to make it tidy.

Afbeelding met tafel

Automatisch gegenereerde beschrijving

Afbeelding met tafel

Automatisch gegenereerde beschrijving

1. Why is it useful to make your dataset tidy?

dpLyr and ggplot2 are designed to work with tidy data

1. Give an example of a variable that should be stored as an ordered factor.

Tshirt size

1. Explain the concepts of parsing and regular expressions, as it relates to data importing and cleaning.
2. Discuss different strategies of handling missing data.

Listwise or case deletion

Pairwise deletion

Mean substitution

…