



Clean: Summary

Cleaning is the third step in the data wrangling process:

- Gather
- Assess
- **Clean**

There are two types of cleaning:

- Manual (not recommended unless the issues are one-off occurrences)
- Programmatic

The programmatic data cleaning process:

1. Define: convert our assessments into defined cleaning tasks. These definitions also serve as an instruction list so others (or yourself in the future) can look at your work and reproduce it.
2. Code: convert those definitions to code and run that code.
3. Test: test your dataset, visually or with code, to make sure your cleaning operations worked.

Always make copies of the original pieces of data before cleaning!

NEXT