# DATA CLEANING

INTRO TO R

## 80% of data work is data cleaning!

In this lesson, we covered

- What is data cleaning, and why is it important?
- How do I identify and handle missing or incomplete data in R?
- How can I transform or reformat my data to make it easier to analyze or visualize in R?
- What are some applicable R packages and functions for data cleaning, such as tidyr and dplyr?

To refresh, Overall, data cleaning is an essential step in data analysis, and R provides several powerful tools to help you do it. tidyr and dplyr are two examples of packages that can help you organize and manipulate your data, and functions like data(), View(), class(), glimpse(), and unique() can help you explore and understand your data better. By ensuring that your data is accurate and ready for analysis, you can increase the effectiveness of your data analysis and ultimately make better decisions based on your results.

- R also provides several functions to help you explore and understand your data.
  - use "data()" to load a data set into R,
  - "View()" to open it in a spreadsheet-like format.
  - "class()" can tell you what type of data you are working with, and
  - "glimpse()" can give you a quick summary of your data structure, including the number of observations and variables.
  - Additionally, "unique()" can help you identify any duplicate values in your data.
- tidyr helps you clean your data by organizing it into a tidy format, meaning each variable has its own column and each observation has its own row.
  - use functions like "gather()" to convert wide data into tidy data, and
  - "spread()" to convert tidy data into wide data.
- dplyr is another package that can help you clean your data by manipulating it with functions like "filter()", "arrange()", "mutate()", and "summarize()". These functions allow you to extract, filter, arrange, and summarize data.

Try this guided course to strengthen your learnings:

Apply your importing and data cleaning skills to real-world soccer data.

45 minutes  Data Manipulation, Importing & Cleaning Data...  Erin LaBrecque  Project
https://www.datacamp.com/projects/758