

Thinking with Data in R

Tidy Data, Transformation, and Exploration

Matthew McDonald

Where We Are in the Workflow

What Analysis Textbooks Assume

- observations are well-defined
- variables are clearly represented
- data is already analysis-ready

What This Course Makes Explicit

- data must be:
 - tidy
 - interpretable
 - constructed with intent
- visualization is a thinking tool
 - not decoration

What Is the Tidyverse?

The Role of the Tidyverse

- structuring data for analysis
- constructing meaningful variables
- exploring patterns and variation

ISLR's View of Data

- observations
- variables
- outcomes

What ISLR Leaves Unsaid

- how observations are defined
- how variables are constructed
- how datasets are assembled

Tidy Data Defines Observations

- what does one row represent?
- what is the unit of analysis?
- what comparisons are valid?

Transformation Creates Variables

- variables are constructed, not given
- scale and representation matter
- transformations encode meaning

Exploration Reveals Structure

- distributions and scale
- relationships between variables
- variation across time and groups

How These Pieces Fit Together

- tidy data clarifies observations
- transformation defines variables
- exploration builds intuition

How We'll Spend the Rest of Today

- exploratory visualization
- data wrangling and transformation
- revisiting questions with cleaner structure