8 Missing Data

18 Missing Values
  1. Understanding Missing Values
     - missing values in R are represented by NA
     - NAs propagate through most operations
     - special care is required when summarizing or filtering data
  2. Detecting Missing Values
     - use is.na() to check for missingness
     - comparisons like x == NA do not work
     - sum(is.na(x)) is useful for counting NAs
  3. Handling Missingness In Summaries
     - most summary functions require na.rm = TRUE
     - mean(), sum(), and other stats return NA unless instructed to remove NAs
     - always check for missingness before interpreting results
  4. Filtering With Missing Values
     - logical comparisons involving NA return NA, not TRUE or FALSE
     - NA rows may be dropped unintentionally in filter()
     - use is.na() inside filter() to explicitly keep or remove NA rows
  5. Replacing Or Imputing Missing Values
     - use replace_na() to fill in missing values
     - choose replacement strategies based on domain knowledge
     - imputation should be documented and justified
  6. Visualizing Missingness
     - missing values can cause ggplot to drop observations
     - some geoms warn you when rows are removed due to NA
     - visualizing patterns of missingness can reveal data issues
  7. Understanding Why Values Go Missing
     - missingness can reflect data entry issues, censoring, or true absence
     - distinguishing between MCAR, MAR, and MNAR matters in modeling
     - different types of missingness require different handling strategies