Data Manipulation with dplyr

aes() (aesthetics): A ggplot2 function that maps data columns to visual properties such as x, y, color, and size for use by geoms

aggregation: The process of summarizing many observations into fewer observations (or single values) using operations like counts, sums, averages, minima, or maxima

arrange(): A dplyr verb that sorts the rows of a table by one or more columns, with desc() used to request descending order

assignment operator (<-): The R operator used to assign values or objects to a name, commonly written as a leftward arrow

count(): A dplyr convenience verb that tallies observations and returns a column named n, and which accepts arguments like sort and wt for ordering and weighting

dplyr: An R package in the Tidyverse that provides a concise set of verbs for data manipulation, including selecting, filtering, arranging, and transforming tables

filter(): A dplyr verb that returns only the rows of a table that meet one or more logical conditions

glimpse(): A dplyr/tibble function that shows the first few values and data types of each column to help you quickly inspect a dataset

group_by(): A dplyr verb that defines groups of rows so subsequent summary or window operations are performed within each group

grouped mutate: A mutate performed after group_by() that computes new column values within each group rather than across the whole table

lag(): A window function that shifts a vector down by one position so each element can be compared to the previous element, with the first position becoming NA

mutate(): A dplyr verb that adds new columns or transforms existing ones by computing values from other columns

n (count column): The default column name produced by count() that stores the number of observations (or the weighted total if wt is used) for each group

observation (row): A single record in a dataset that contains values for each variable, often representing one real-world entity like a county or person

pipe operator (%>%): An operator that passes the output of one expression as the first argument to the next, enabling readable chains of data transformations

relocate(): A dplyr verb that moves columns to a new position in the table without dropping any columns, using before/after semantics

rename(): A dplyr verb that changes one or more column names while keeping all other columns unchanged

select helpers: Functions used inside select() (e.g., contains(), starts_with(), ends_with(), matches(), last_col()) to choose columns by name patterns or position

select(): A dplyr verb that extracts a subset of columns from a table, optionally renaming or reordering them in the same call

slice_min()/slice_max(): dplyr verbs that extract the rows with the smallest or largest values of a specified column within each group, with an n argument to control how many rows to keep

summarize(): A dplyr verb that reduces multiple rows to a single row per group by computing aggregate statistics such as sum(), mean(), min(), or max()

tibble: A modern reimagining of the data.frame in R that prints more clearly and preserves column types, commonly produced and used by Tidyverse packages

Tidyverse: A collection of R packages designed for data science that share common design principles and data structures, including dplyr and ggplot2

ungroup(): A dplyr verb that removes grouping metadata from a table so later operations run on the entire dataset rather than by group

variable (column): A named field in a dataset that stores values of the same type for each observation, often represented as a table column

window function: A function that returns a vector of the same length as its input, allowing calculations that compare elements to other elements in the vector (e.g., lag())

wt argument: An argument used in count() and similar functions to weight observations by a numeric column so the aggregated result reflects sums rather than raw counts