

Core Concepts

Tidyverse is an opinionated collection of R packages designed for data science. They share an underlying design philosophy, grammar, and data structures.

- Pipe operator (`%>` or `▷`): Passes the result of one function as the first argument of the next. `data %>% filter(...) %>% map(...)`.
- Tidy data: 1. Each variable is a column. 2. Each observation is a row. 3. Each value is a cell.

Data Wrangling (dplyr)

- `filter()`: Pick observations based on their values.
- `select()`: Pick variables by their names.
- `mutate()`: Create new variables or transform existing ones.
- `arrange()`: Reorder rows.
- `summarize()` / `summarise()`: Collapse many values down to a single summary.
- `group_by()`: Group data for performing operations within groups.
- `rename()`: Rename columns.
- `relocate()`: Change column order.

Data Transformation (tidyr)

- `pivot_longer()`: Convert from wide to long format.
- `pivot_wider()`: Convert from long to wide format.
- `separate()` / `separate_wider_delim()`: Split a single column into several.
- `unite()`: Combine multiple columns into one.
- `drop_na()`: Remove rows with missing values.
- `replace_na()`: Replace NAs with specified values.

Data Import (readr)

- `read_csv()`: Read comma-separated files.
- `read_tsv()`: Read tab-separated files.
- `write_csv()`: Save data to CSV.

Iteration and Functional Programming (purrr)

- `map()`: Apply a function to each element and return a list.
- `map_dbl()`, `map_chr()`, `map_lgl()`: Return specific atomic vector types.
- `walk()`: For functions with side effects (like plotting).

String Manipulation (stringr)

- `str_detect()`: Check if a pattern exists.
- `str_replace()`: Replace matches with new text.
- `str_sub()`: Extract substrings.
- `str_trim()`: Remove leading/trailing whitespace.

Factors (forcats)

- `fct_reorder()`: Reorder factor levels based on another variable.
- `fct_lump()`: Collapse infrequent levels into "Other".
- `fct_rev()`: Reverse levels.

Dates and Times (lubridate)

- `ymd()`, `mdy()`, `dmy()`: Parse dates.
- `now()`, `today()`: Get current date/time.
- `year()`, `month()`, `day()`: Extract components.

Useful Workflow Example

```
library(tidyverse)

result <- raw_data %>%
  as_tibble() %>%
  filter(status == "Active") %>%
  group_by(category) %>%
  summarize(
    mean_value = mean(score, na.rm =
TRUE),
    n = n()
  ) %>%
  arrange(desc(mean_value)) %>%
  mutate(rank = row_number())
```

```
ggplot(result, aes(x = reorder(category,
```

```
mean_value), y = mean_value)) +
  geom_col() +
  coord_flip() +
  labs(title = "Average Scores by
Category")
```

Tibbles

- Modern version of data frames.
- Only print first 10 rows and columns that fit.
- Do not change variable names or types automatically.
- Never drop dimensions when indexing.