Data Challenge 02

R Functions and Tidyverse Intro

Jake Moore

Objective

Create small functions and practice tidyverse pipelines with the built-in `mpg` dataset. Knit to PDF in an RStudio Project. Use only `tidyverse` and `readxl`. Use snake_case. Un-comment and fill in the code as instructed.

Rules

- Keep object and function names exactly as specified. My test script expects them.
- Use project-relative paths only (none needed here).
- Use clear, one-action-per-line code in pipelines.

0) Setup

```r
library(tidyverse)
```

```
## Warning: package 'tidyverse' was built under R version 4.5.1

## Warning: package 'ggplot2' was built under R version 4.5.1

## Warning: package 'tibble' was built under R version 4.5.1

## Warning: package 'tidyr' was built under R version 4.5.1

## Warning: package 'readr' was built under R version 4.5.1

## Warning: package 'purrr' was built under R version 4.5.1

## Warning: package 'dplyr' was built under R version 4.5.1

## Warning: package 'stringr' was built under R version 4.5.1

## Warning: package 'forcats' was built under R version 4.5.1

## Warning: package 'lubridate' was built under R version 4.5.1
## -- Attaching core tidyverse packages ------------------------ tidyverse 2.0.0 --
## v dplyr     1.1.4     v readr     2.1.5
## v forcats   1.0.0     v stringr   1.5.1
## v ggplot2   3.5.2     v tibble    3.3.0
## v lubridate 1.9.4     v tidyr     1.3.1
## v purrr     1.1.0
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors
```

```
library(readxl)
```

```
## Warning: package 'readxl' was built under R version 4.5.1
```

1) Inspect mpg (5 pts)

Create two objects with the row and column counts of `mpg`.

- `mpg_nrow`
- `mpg_ncol`

```
mpg_nrow <- nrow(mpg)
mpg_ncol <- ncol(mpg)
```

2) Simple math function (5 pts)

Write a function that returns the sum of two numbers.

- `sum_func` returns `a + b` where `a` and `b` are the defined parameters.

```
sum_func <- function(a, b) {
  result <- a + b
  return(result)
}
```

3) head-like function, fixed 6 (10 pts)

Return the first 6 rows of a data frame without calling `head()`.

- `head_func(df)` returns first 6 rows of the dataframe `df`.

```
head_func <- function(df) {
  result <- df[1:6, ]
  return(result)
}
```

4) head-like with default n (15 pts)

Return the first `n` rows, default 6. If `n <= 0`, return zero rows.

- `head_func2(mpg, 10)` would return the first 10 rows of the dataframe `mpg`

```
head_func2 <- function(df, n = 6) {
  if (n <= 0) {
    result <- df[0, ]
  } else {
    result <- df[1:n, ]
  }
  return(result)
}
```

5) Select three columns (15 pts)

Return only `manufacturer`, `model`, and `year` in that order.

- `select_make_model_year(df)`

```
select_make_model_year <- function(df) {
  result <- df %>%
    select(manufacturer, model, year)
  return(result)
}
```

6) Filter by manufacturer (20 pts)

Keep only rows where `manufacturer` equals a supplied name (case-sensitive).

- `filter_manufacturer(df, manufacturer_name)`

```
filter_manufacturer <- function(df, manufacturer_name) {
  result <- df %>%
    filter(manufacturer == manufacturer_name)
  return(result)
}
```

7) Compose a readable pipeline (30 pts)

In the real world, we often automate the cleaning process in what is called a pipeline, where we have data of known format coming in with a specified data formatting coming out using transformations applied.

Create `land_rover_models` with this pipeline:

1. Start from `mpg`.
2. Keep only `manufacturer`, `model`, `year` (use your function from Q5).
3. Keep only rows where manufacturer is `"land rover"` (use your function from Q6).
4. Sort by `year` ascending.

- Object to create: `land_rover_models`

```r
land_rover_models <- mpg %>%
  select_make_model_year() %>%
  filter_manufacturer("land rover") %>%
  arrange(year)
```

Submission

- Knit this file to PDF.
- Submit both the .Rmd and the PDF.