Working with Data Frames: Takeaways

Syntax

INSTALLING AND LOADING PACKAGES

- Install packages:

```
install_packages("package_name")
```

- Load packages:

```
library(package_name)
```

IMPORTING DATA INTO R

- Save data as a data frame (data in .csv format)

```
new_data_frame <- read_csv("data_set.csv")
```

WORKING WITH DATA FRAME COLUMNS

- Select data frame columns:

```
data_frame_2 <- data_frame %>%

select(column_2, column_4, column_6)
```

- Add a new column to a data frame:

```
data_frame_2 <- data_frame_1 %>%

mutate(new_column = (column_2/column_4)*100)
```

FILTERING A DATA FRAME BY A SINGLE CONDITION

- Numeric data

```
data_frame_2 <- data_frame_1 %>%

  filter(column_2 < 70)
```

- Character data

```
data_frame_2 <- data_frame_1 %>%

filter(column_3 == "Variable Name")
```

FILTERING A DATA FRAME BY MULTIPLE CONDITIONS

- Meeting at least one criterion (the | operator):

```
data_frame_2 <- data_frame_1 %>%

filter(column_6 == "Variable Name" | column_4 > 1000)
```

- Meeting multiple criteria (the & operator):

```
data_frame_2 <- data_frame_1 %>%

    filter(column_6 == "Variable Name" & column_4 > 1000)
```

ARRANGING DATA FRAMES BY VARIABLES

- Arrange by a variable from smallest to largest:

```
data_frame_2 <- data_frame_1 %>%

  arrange(column_2)
```

- Arrange by a variable from largest to smallest:

```
data_frame_2 <- data_frame_1 %>%

  arrange(desc(column_2))
```

- Arrange by multiple variables:

```
data_frame_2 <- data_frame_1 %>%

  arrange(column_2, desc(column_4))
```

Concepts

- In R, packages consist of user-contributed functions, code and data that extend R's capabilities.
- The tidyverse is a collection of packages designed to make using R for data science more effective.

- Tibbles are a specialized type of data frame. They are a feature of packages in the tidyverse family that have been introduced to extend R's functionality for modern data science tasks.
- The pipe operator (%>%) is used to write code that chains series of operations together

Resources

- CRAN repository, which contains R packages
- the tidyverse family of packages
- readr package documentation
- Documentation on tibbles
- dplyr package documentation