Data Manipulation in R

Lesson 1

API 209: Advanced Quantative Methods
TF: Rony Rodriguez-Ramirez
Summer 2024

Plan for today

1. Getting to know each other
2. The layout of this summer camp (just the R part)
3. Why R, RStudio, Positron?
4. Data manipulation in R

Getting to know each other

Who am I?

Rony Rodriguez-Ramirez (G2)

Ed Policy (Economics of Education) Program

Previous exp: The World Bank

Like coding!

My role (?)

What should you expect from me?

1. My job if that you feel confident in R.
2. I should be there if you have any questions.
3. Reply to your emails or slack messages.

At the end of this summer camp?

1. You should be ready for the semester.
2. Know enough about R.
3. Know how to craft questions and where to look for answers.
4. Be happy (?)

Course assistants for Math Camp

Shan

Ayush

Sara

The layout

The layout

What are we going during math camp (R Part)?

8 sessions over the next weeks:

1. 4 Lessons (2 hours)
    - I will discuss about coding, strategies, and implementation
2. 4 Labs (1.5 hours)
    - It will be a hands-on session. I will provide you with exercises and we will solve them together.
3. Optional: Office hours

The layout

There is a website for this summer camp:

website

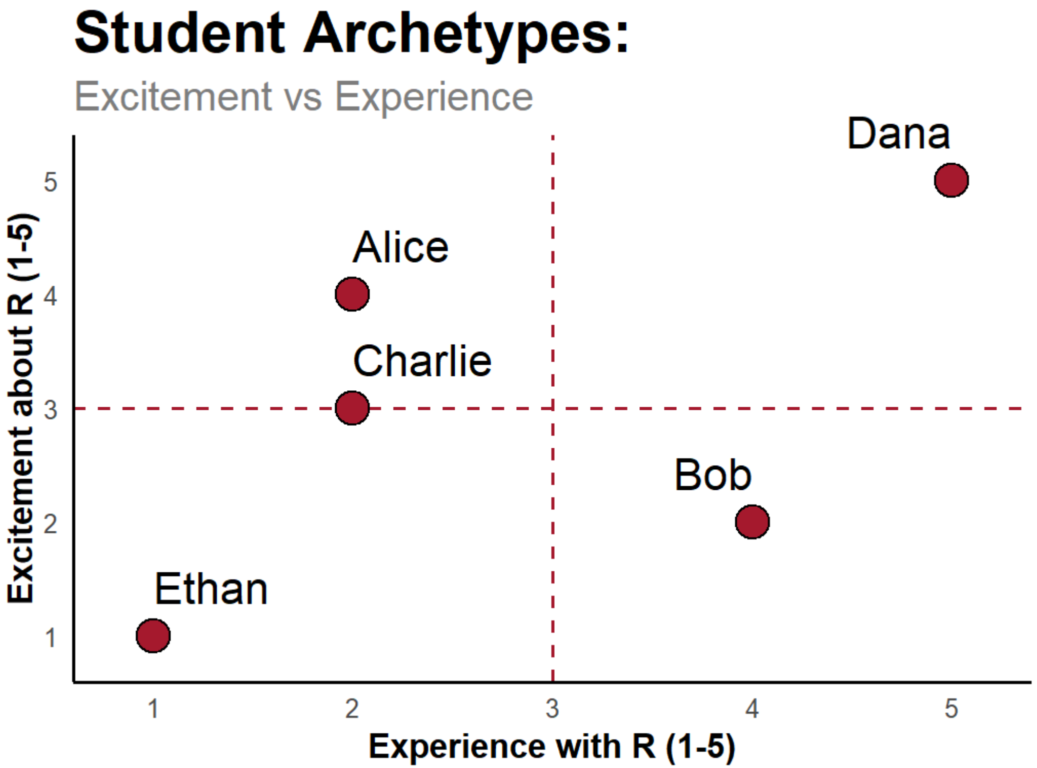
It is not up-to-date; but every week, you will have the materials for that respective week, i.e., you should have already your lab for tomorrow.

Coding

Coding is a Skill

It is most fun to practice a skill with people you know.

Discussion Activity: Archetypes



Credit: @Dom | Former TF.

- **Intros**: Name, where you’re from, favorite midday snack or superhero.
- **Experience**: With statistics, programming, and/or R? (Yes/No)
- **Which quadrant describes you?**

Classroom Norms

Class is a collective enterprise!

- Allow everyone the chance to speak.
- Be mindful of thoughts and actions.
- Understand differing levels of knowledge and experience.
- Help others in your group!
- Don't be afraid to ask questions!
- Respect others' opinions and suggestions.
- Try questions on your own first, then come together.
- Take breaks and have fun!

Why R, RStudio, Positron?

Why R, RStudio, and Positron?

Why R?

- Open-source and free.
- Extensive ecosystem for statistical analysis.
- Wide range of packages for data manipulation and visualization.
- Active and supportive community.

RStudio and Positron: The IDEs for R

Why RStudio?

- Integrated development environment (IDE) that simplifies coding in R.
- Built-in tools for code development, debugging, and collaboration.
- Seamless integration with RMarkdown for dynamic report generation.
- Powerful tools for data visualization and manipulation.

Why Positron?

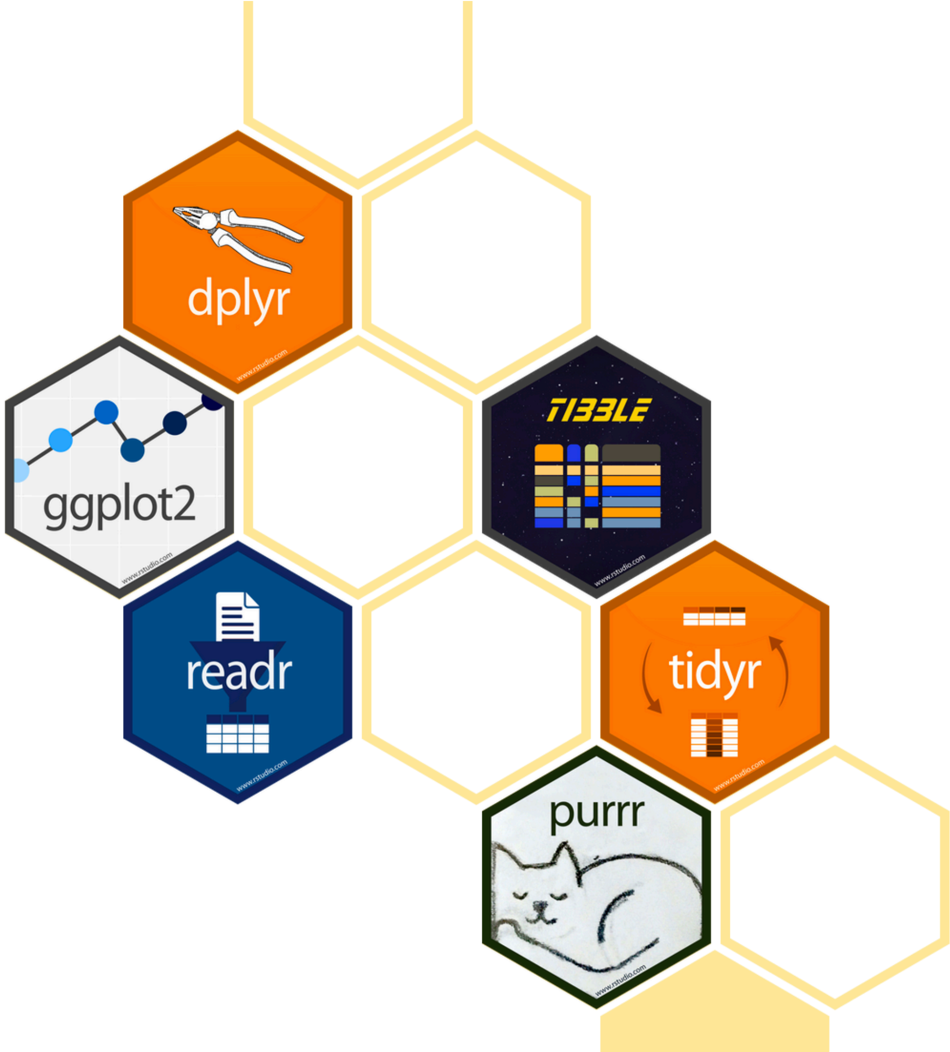
- New, modern IDE designed to enhance the R programming experience.
- Sleeker interface with enhanced performance and features.
- Supports the latest R packages and workflows.
- Focuses on integrating modern development tools and practices.

Which IDE Should You Use?

RStudio vs. Positron

- **RStudio** is well-established with a large user base and extensive support.
- **Positron** offers cutting-edge features for those looking to adopt the latest tools.
- Consider trying both to see which fits your workflow best.
- During the pre-summer assignment, we used **Posit Cloud**.
    - For those who haven't installed, either RStudio not Positron in your computer, there is a Posit Cloud Project here

The tidyverse



Data manipulation

Advanced Data Manipulation

Mastering Data Manipulation in R

- **Advanced Filtering and Selection**:
  - Use of conditional filtering and dynamic column selection.
- **Complex Mutate Operations**:
  - Creating conditional columns, using lag and lead.
- **Data Reshaping**:
  - Pivoting data, advanced grouping.
- **Efficient Data Handling**:
  - Joining datasets, parallel processing.

Recap: The Tidyverse

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

Core Packages:

- `ggplot2` - Data visualization
- `dplyr` - Data manipulation
- `tidyr` - Data tidying
- `readr` - Data import
- `purrr` - Functional programming
- `tibble` - Modern data frames
- `stringr` - String manipulation

dplyr: Key Functions

Commonly Used Functions:

- `filter()` - Subset rows based on conditions
- `select()` - Choose columns by names
- `mutate()` - Create new columns or modify existing ones
- `arrange()` - Reorder rows
- `summarize()` - Aggregate data
- `group_by()` - Group data for summary operations

dplyr: Example

```r
# Example of dplyr in action
library(dplyr)

# Filter and select
filtered_data <- starwars |>
  filter(height > 180) |>
  select(name, height, hair_color)

filtered_data
```

```
## # A tibble: 39 × 3
##    name               height hair_color
##    <chr>               <int> <chr>
##  1 Darth Vader           202 none
##  2 Biggs Darklighter     183 black
##  3 Obi-Wan Kenobi        182 auburn, white
##  4 Anakin Skywalker      188 blond
##  5 Chewbacca             228 brown
##  6 Boba Fett             183 black
##  7 IG-88                 200 none
##  8 Bossk                 190 none
##  9 Qui-Gon Jinn          193 brown
## 10 Nute Gunray           191 none
## # ℹ 29 more rows
```

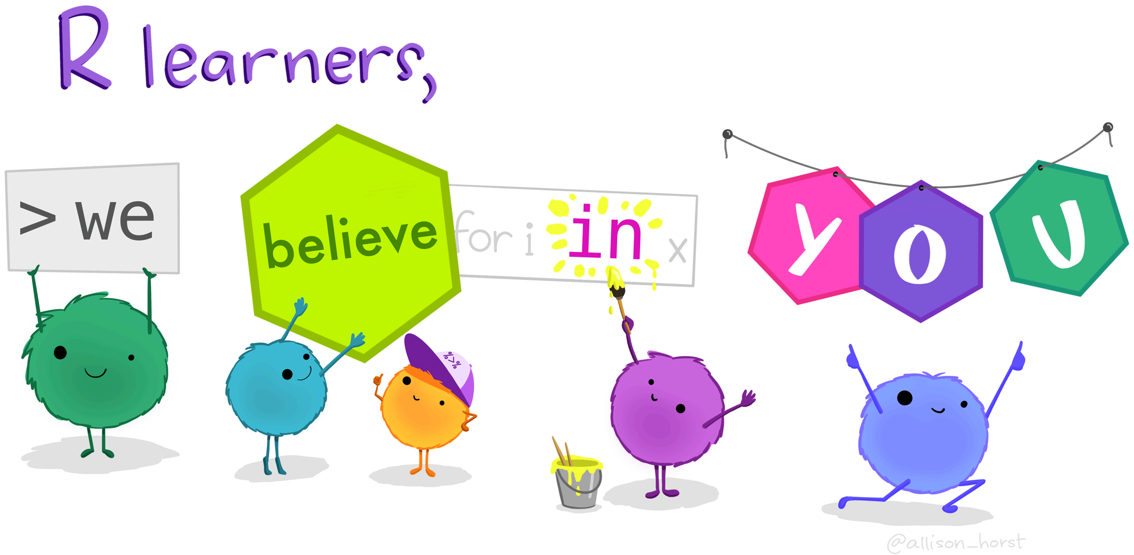
dplyr: Example

```r
# Mutate
starwars |>
  mutate(
    tatooine = ifelse(
      homeworld == "Tatooine",
      "Tatooine",
      "Others"
    )
  ) |>
  group_by(tatooine) |>
  summarize(
    mean = mean(height)
  )
```

```
## # A tibble: 3 × 2
##   tatooine  mean
##   <chr>    <dbl>
## 1 Others    177.
## 2 Tatooine  170.
## 3 <NA>       NA
```

Are we good here?

As of now, you should have the tools to understand the last code. More resources are available in our website. Now, it's time to make some mistakes!



Artwork by @allison_horst

Takeaways

Sucking (Slide from Prof. Andrew Heiss)

> “There is no way of knowing nothing about a subject to knowing something about a subject without going through a period of much frustration and suckiness.”

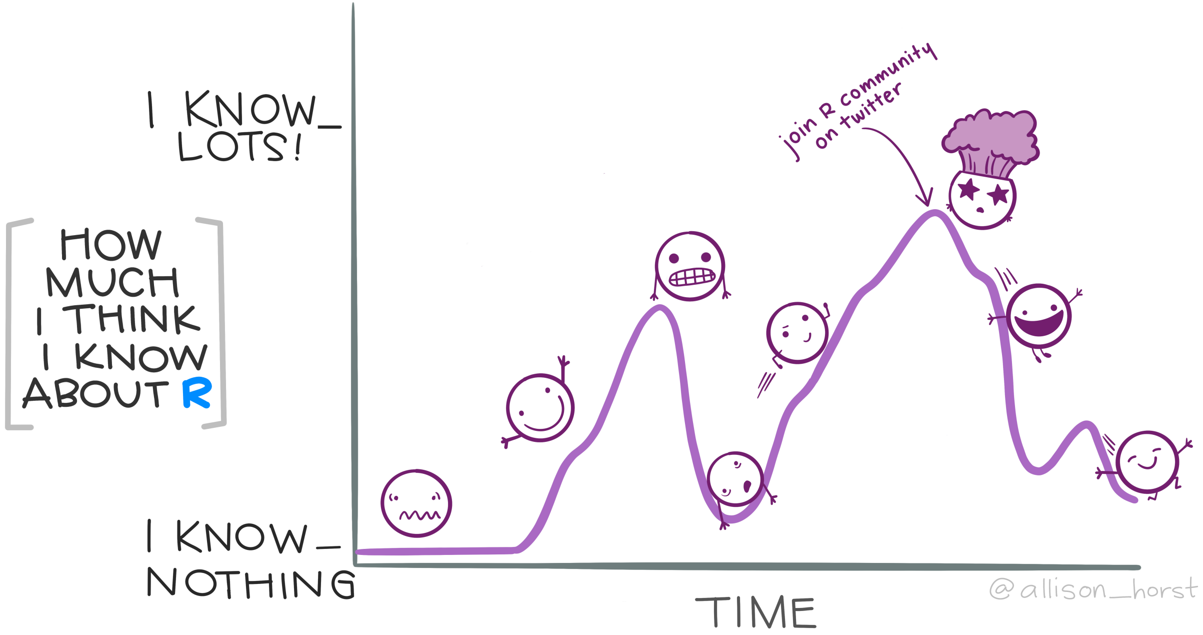
> **“Push through. You’ll suck less.”**

Hadley Wickham, author of {ggplot2}

Sucking

Dude, sucking at something is the first step
towards being sort of good at something

Sucking



Artwork by @allison_horst