

# Introduction to R

Duke Economic Analytics Laboratory

Jenna Goldberg

Spring 2023

A dark blue diagonal gradient bar that starts from the bottom left and extends towards the top right, covering the lower half of the slide.

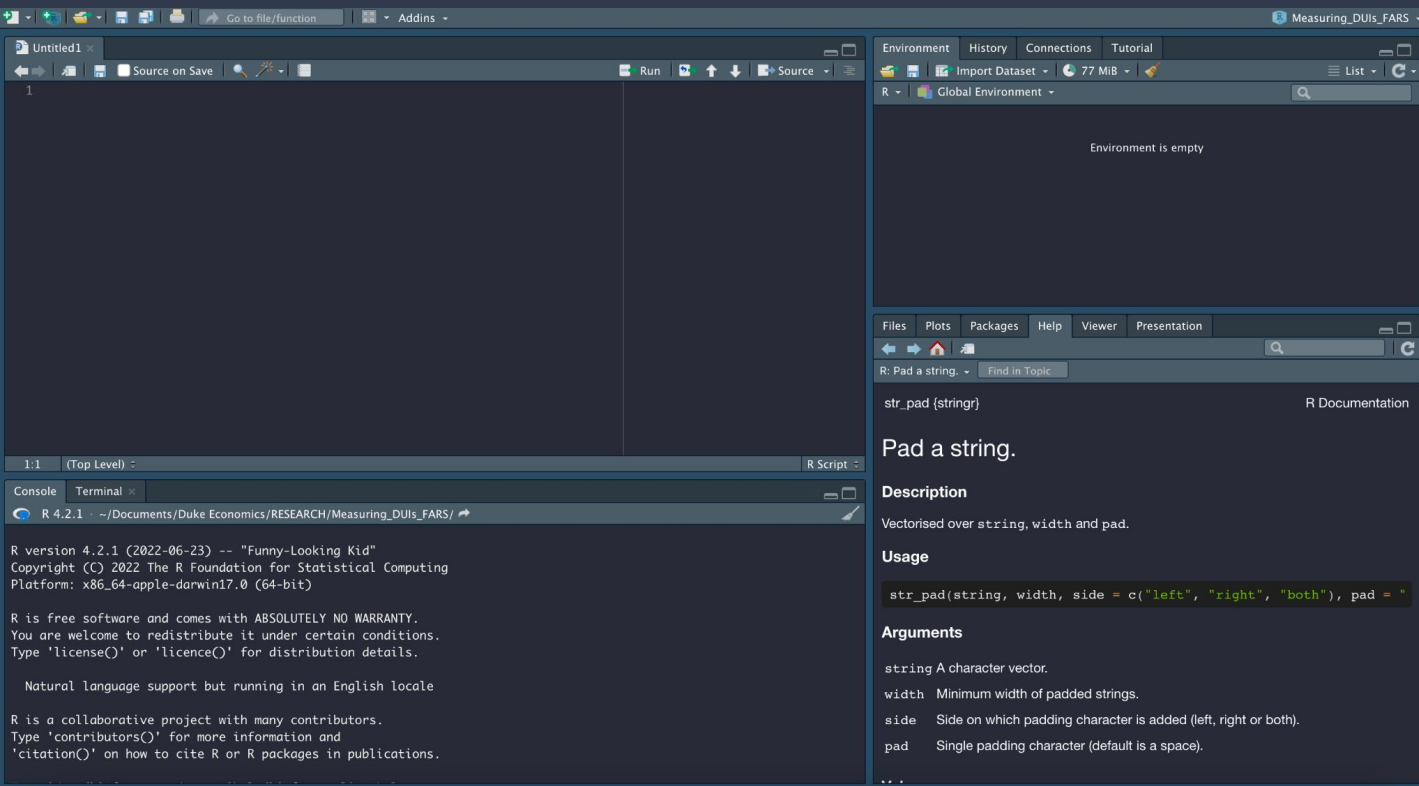
# What is R?

- A programming language designed specifically for statistical computing & graphics
- Open source & free, with a large community of users & package developers
- Most popular for data science & statistical analysis, gaining popularity in Economics

# What is RStudio?

- RStudio is the most popular integrated development environment (IDE) for R users
- Like R, RStudio is open source & free, although commercial enhancements do exist
- Car analogy: R is the engine, RStudio is the dashboard

# RStudio Tour

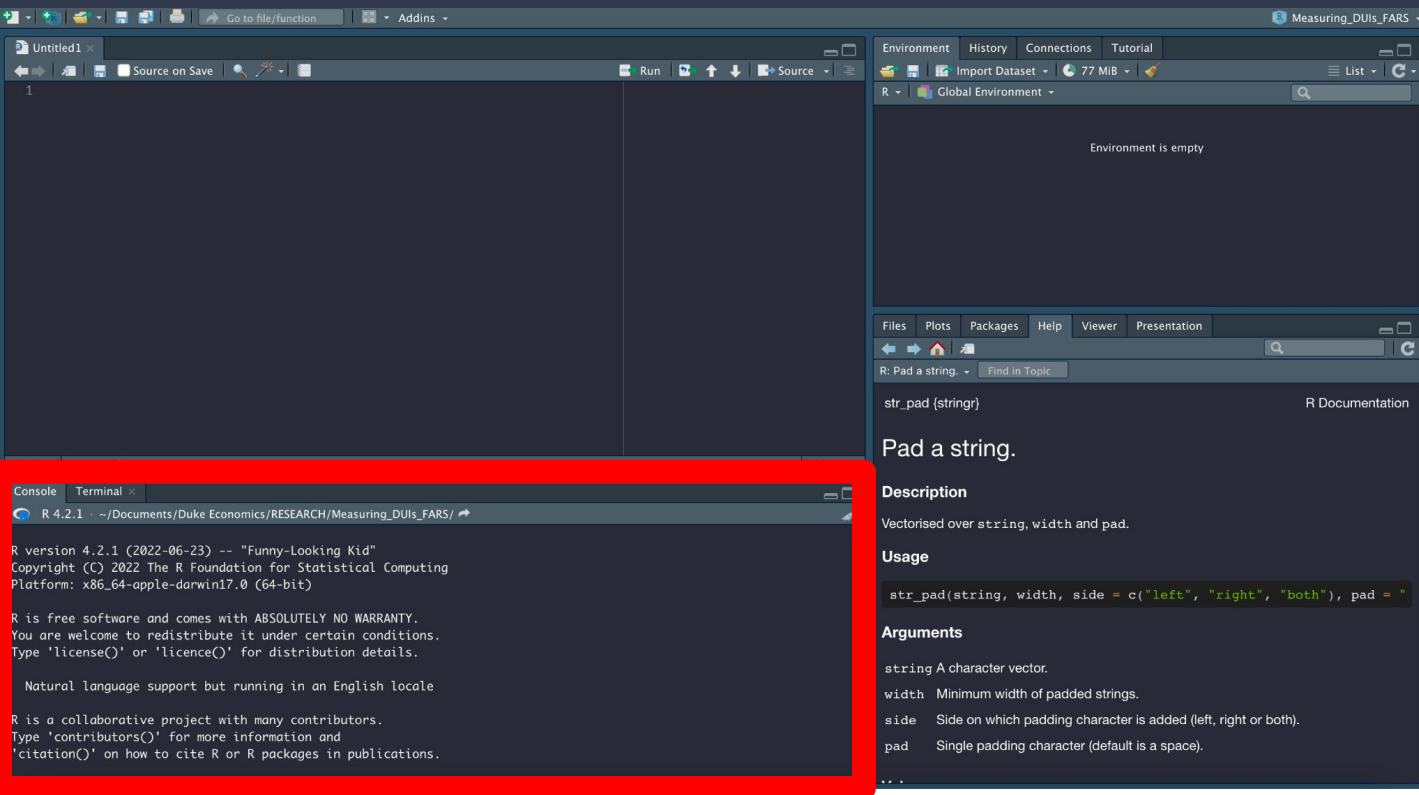


This is RStudio.

I'm going to give you a brief tour and then we'll go to the live demo.

There is no separate window for .do files, everything is here in one screen.

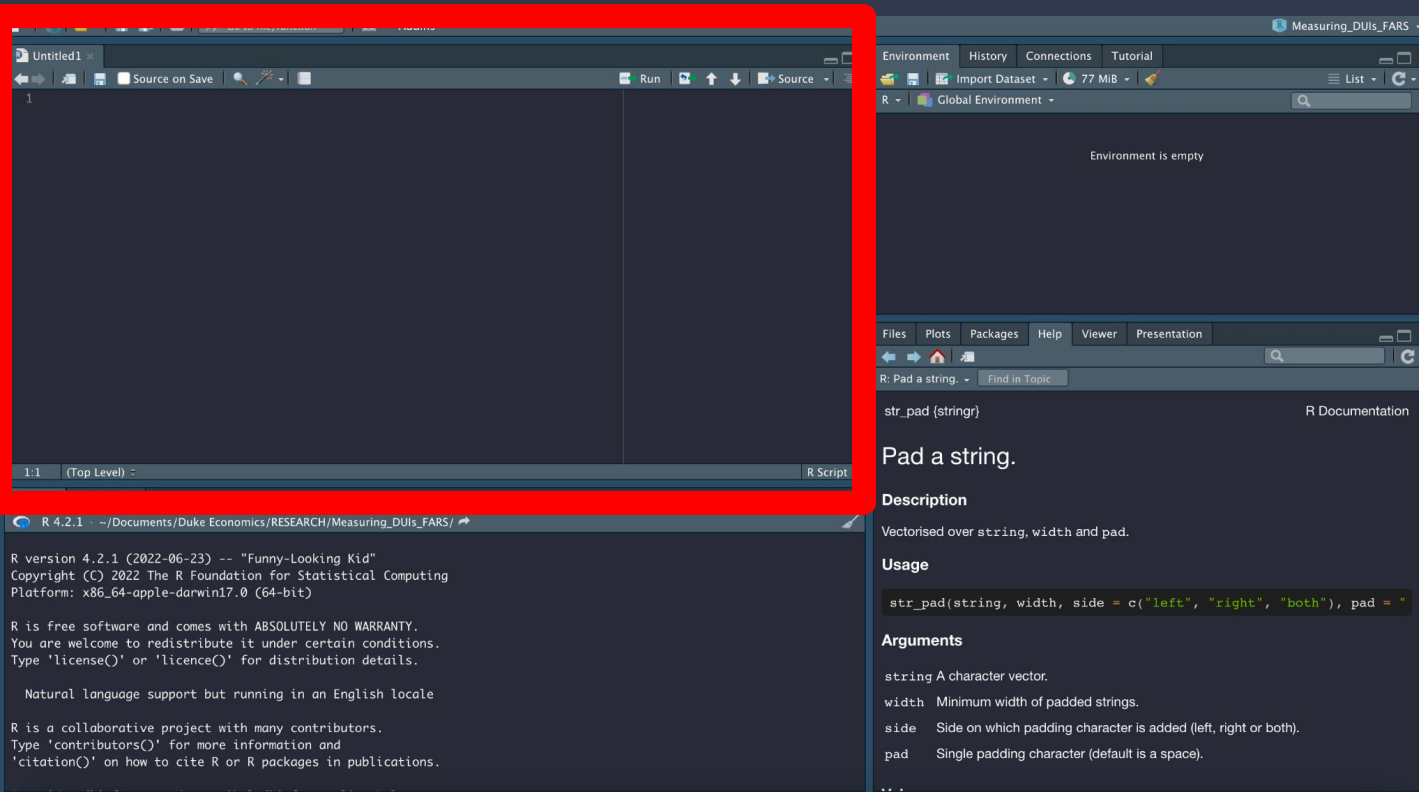
# RStudio Tour



This is the console - everything you see here is directly interfacing with the R engine.

Mostly, you'll see outputs from your code & error messages.

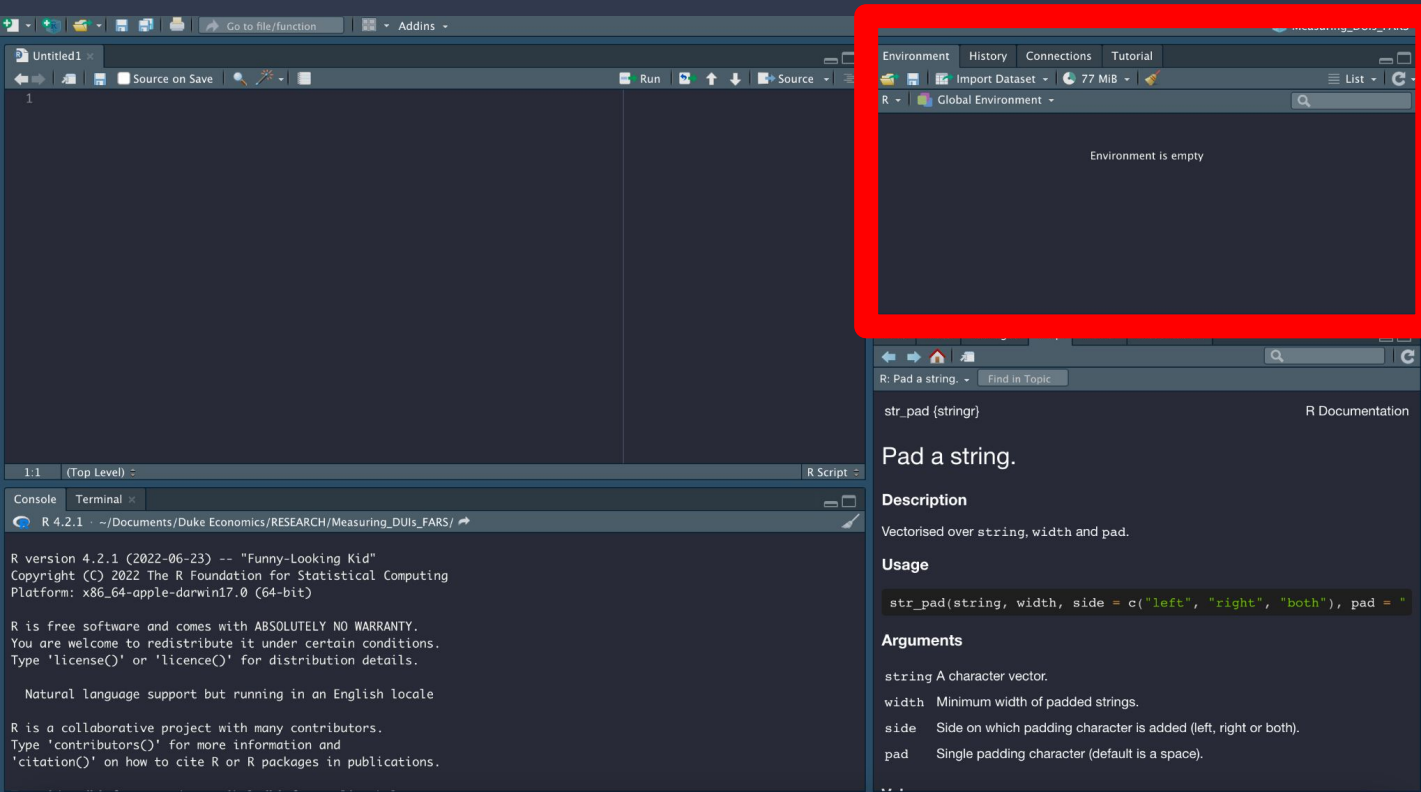
# RStudio Tour



This is where you actually write your code.

You can run an individual line with the keyboard shortcut :  
**CTRL + ENTER**

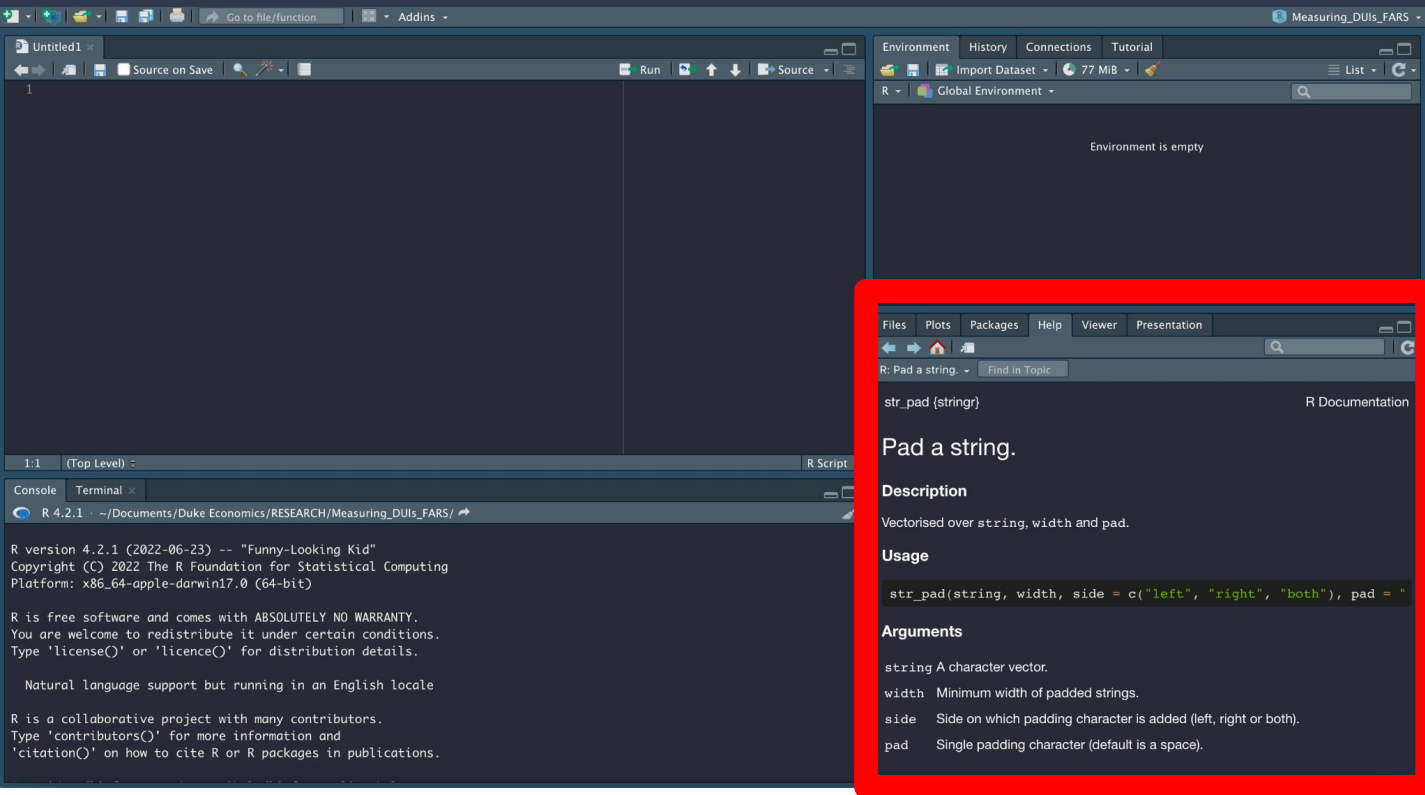
# RStudio Tour



In the environment - you can see which data tables, values and lists are stored & currently accessible in your R instance.

You can click on data frames and other objects to actually view them.

# RStudio Tour



This area of RStudio can serve a few purposes, but there are two main ones I want to tell you now, and the rest you'll see as you learn!

1. If you ask for help on a specific function, you'll see documentation pop up here.
2. If you generate a graph, it will show up here. If you



# Live Demo!

1. Open Introduction.Rproj
2. Open Code/Introduction\_\_to\_\_R.R

# Style: Good vs Bad

Which is better?

Version 1:

```
data<-data%>%group_by(group_name)%>%summarise(mean=mean(var),max=max(var))
```

Version 2:

```
new_data <-  
  data %>%  
    group_by(group_name) %>%  
    summarise(mean_var = mean(var),  
              max_var = max(var))
```

# How to Learn R Effectively

- I can tell when someone had to google every line in a script - stylistic inconsistency is usually a big giveaway
  - Make sure you're understanding and using base R or tidyverse consistently - if you're switching between syntax, make sure you're commenting WHY
  - Make sure to consider legibility in your code - I am intentional about spacing and consistent syntax in my scripts.
- Stackoverflow is a great resource - but be aware that solutions vary in quality & style
- Depending on the task, your primary workflow might be:
  - Find functions with cheat sheets: <https://posit.co/resources/cheatsheets/>
  - Built in R documentation to understand using those functions
  - If you're stuck - google it, but make sure you consider consistency, don't just mindly copy solution structures online, people can tell!

# Next Session's Topics

Date:

3 Basic Data Manipulation Tasks for RA's

- Bulk data import
- Balancing a panel
- Basic String cleaning

I've put some good R resources on the DEAL webpage, but feel free to email me with questions [jennifer.goldberg@duke.edu](mailto:jennifer.goldberg@duke.edu)