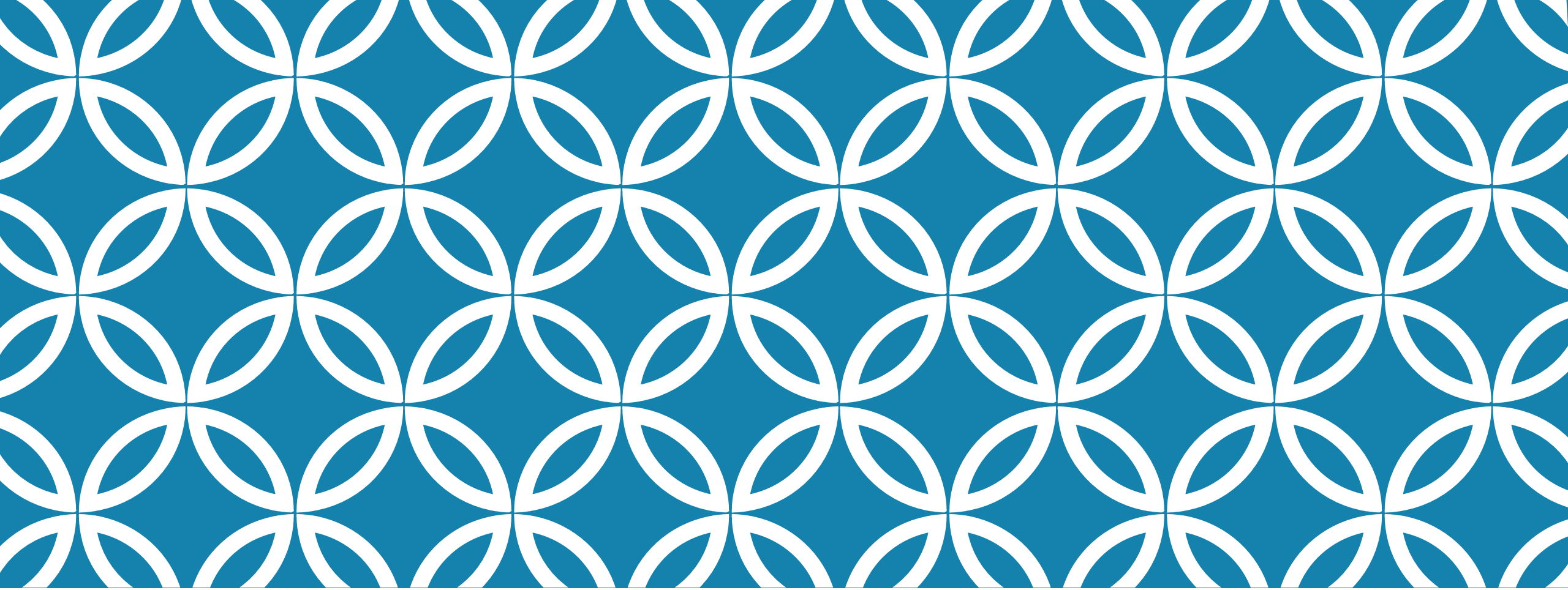


Introduction to R and R Studio

Session 1

Joseph Rudolf

July 16, 2020

July 16 2020	**Session**	**Instructor**
1:00 pm - 1:30 pm	Instructor Introductions, Introduction to technology	Amrom Obstfeld
1:30 pm - 2:15 pm	Introduction to R and RStudio	Joe Rudolf
2:30 pm - 3:15 pm	Reproducible Reporting	Patrick Mathias
3:30 pm - 5:00 pm	Data Visualization	Stephan Kadauke
July 17 2020		
1:00 pm - 2:30 pm	Data Transformation	Amrom Obstfeld
2:45 pm - 4:15 pm	Statistical Analysis	Dan Herman
4:30 pm - 5:00 pm	Advanced Reporting	Patrick Mathias

Lesson Goals

1. Get oriented to R and RStudio
2. Learn some fundamentals of coding

Lesson Objectives

1. Log in and tour RStudio Cloud
2. Execute code at the console
3. Define and use functions
4. Define and create objects in the environment
5. Load data into R and interact with a dataframe

Getting Oriented to R

What is R?

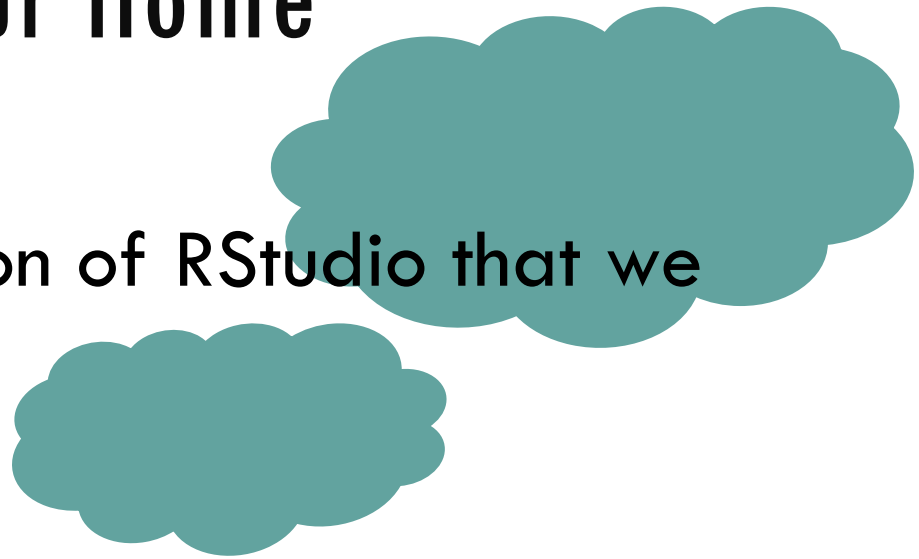
- R is a statistical programming language.
- Using R you can load, analyze, and visualize data.
- R also provides an environment in which we can conduct reproducible data analysis.
  - Documented
  - Revisable
  - Shareable

RStudio: The Portal to R

- RStudio is an integrated development environment (IDE)

- Using RStudio we can interact with the R programming language to:
  - Write and execute code interactively
  - View data
  - Debug and fix errors
  - Author our code

RStudio: In the Cloud… In Your Home



- RStudio Cloud: An online hosted version of RStudio that we will use for these course sessions

- RStudio Desktop: A locally installed version of RStudio that you will use when you get home to continue your learning

Note: Use Rstudio Cloud only for this course. Do not upload protected health information to the cloud!

Your Turn

Navigate to: https://bit.ly/api-r-cloud

Enter your log in credentials

Join Space

Make a copy of the Core Exercises for yourself

Join Space?

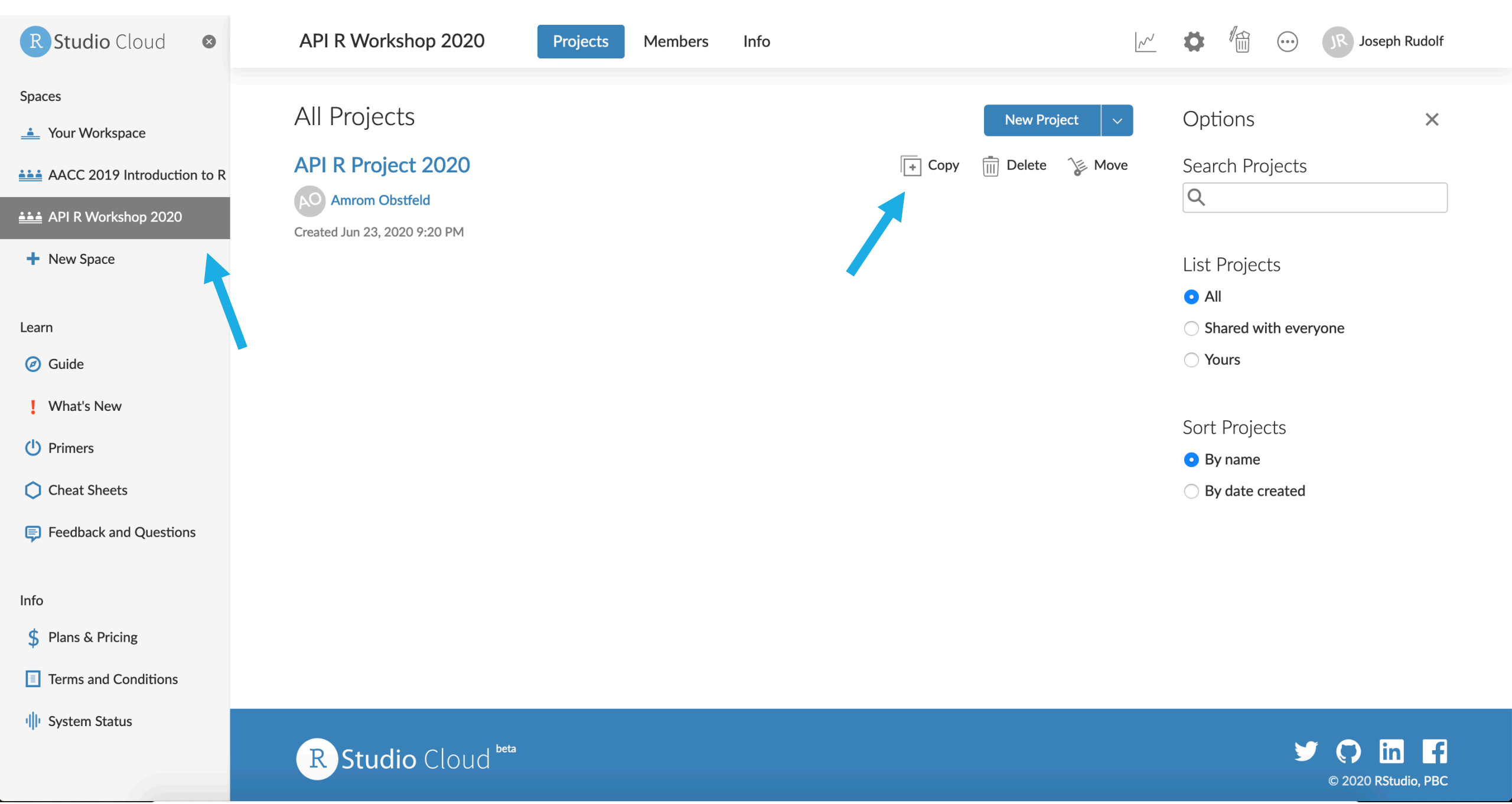
Joining a space gives you access to it and to its contents.

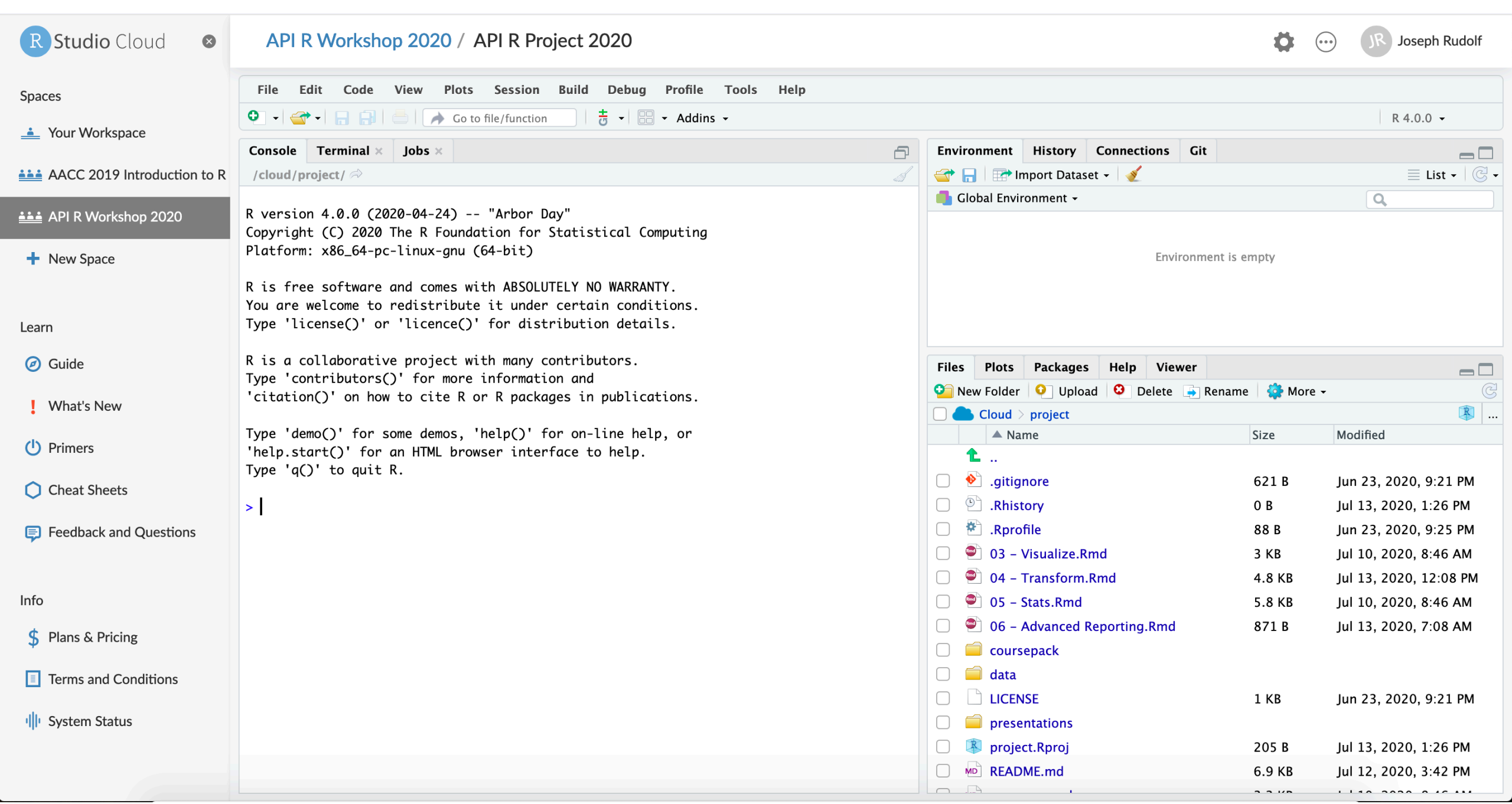
Once you join, admins will be able to see your email address.

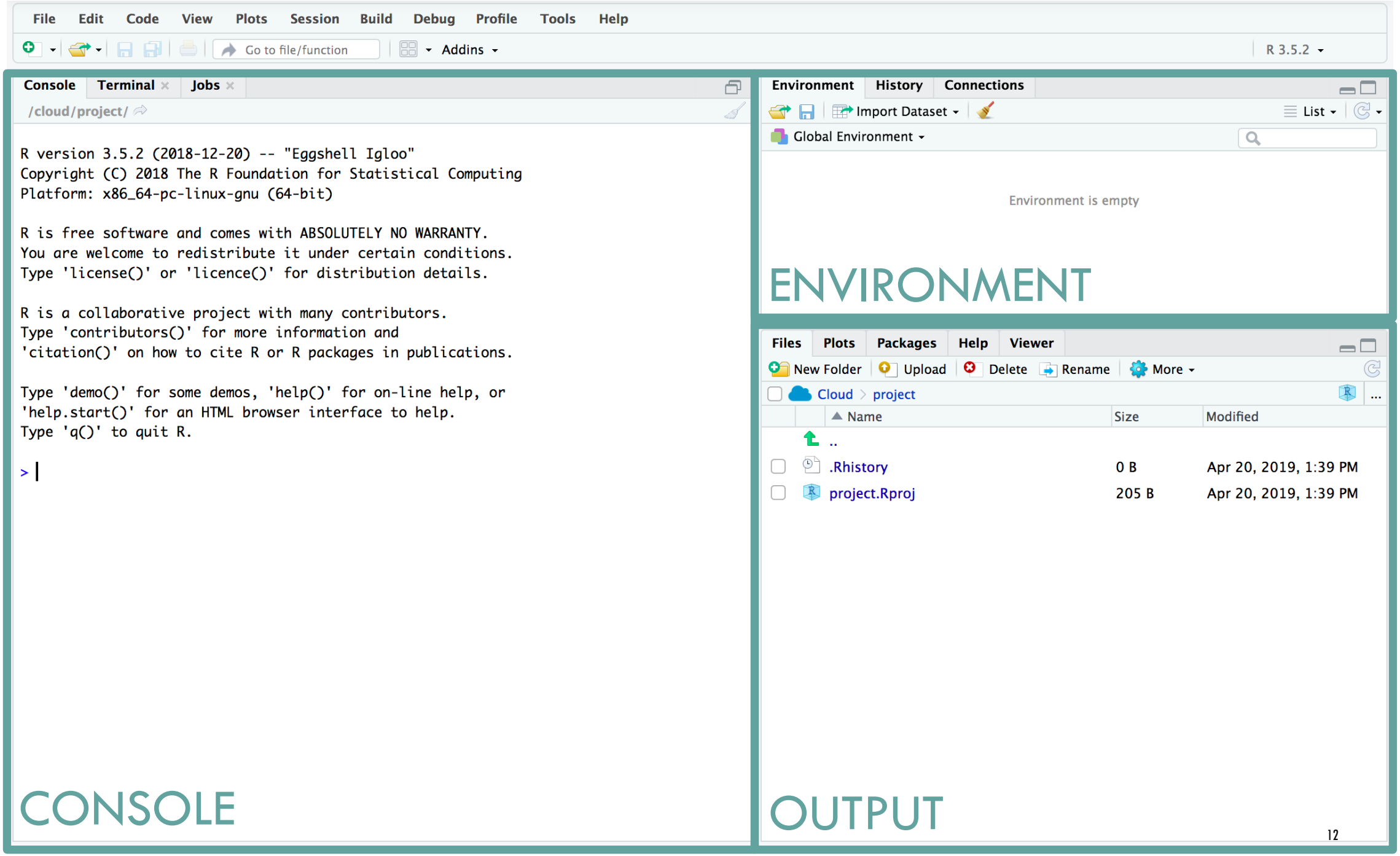
Would you like to join this space?

Join Space

Cancel


RStudio Cloud
Spaces
Your Workspace
AACC 2019 Introduction to R
API R Workshop 2020
New Space
Learn
Guide
What's New
Primers
Cheat Sheets
Feedback and Questions
Info
Plans & Pricing
Terms and Conditions
System Status
API R Workshop 2020
Projects
Members
Info
Joseph Rudolf
All Projects
New Project
API R Project 2020
Amrom Obstfeld
Created Jun 23, 2020 9:20 PM
Copy
Delete
Move
Options
Search Projects
List Projects
All
Shared with everyone
Yours
Sort Projects
By name
By date created
RStudio Cloud beta
© 2020 RStudio, PBC

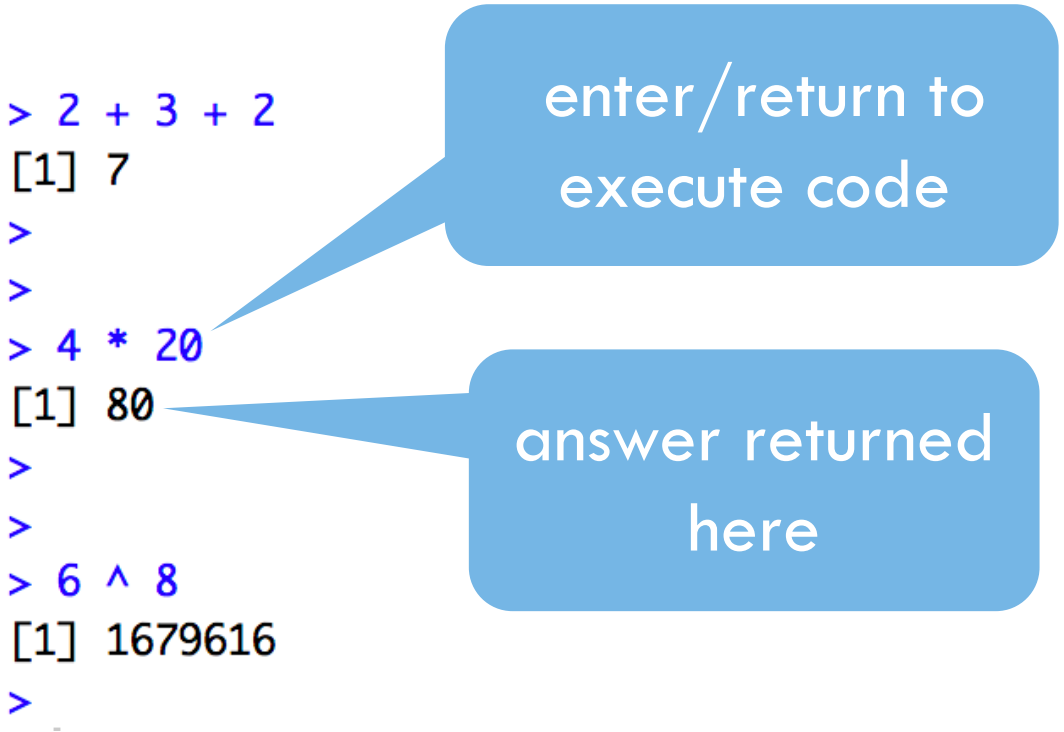

RStudio Cloud
API R Workshop 2020 / API R Project 2020
Joseph Rudolf
Spaces
Your Workspace
AACC 2019 Introduction to R
API R Workshop 2020
New Space
Learn
Guide
What's New
Primers
Cheat Sheets
Feedback and Questions
Info
Plans & Pricing
Terms and Conditions
System Status
File Edit Code View Plots Session Build Debug Profile Tools Help
Go to file/function
Addins
R 4.0.0
Console
Terminal
Jobs
/cloud/project/
R version 4.0.0 (2020-04-24) -- "Arbor Day"
Copyright (C) 2020 The R Foundation for Statistical Computing
Platform: x86_64-pc-linux-gnu (64-bit)
R is free software and comes with ABSOLUTELY NO WARRANTY.
You are welcome to redistribute it under certain conditions.
Type 'license()' or 'licence()' for distribution details.
R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.
Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.
>
Environment
History
Connections
Git
Import Dataset
List
Global Environment
Environment is empty
Files
Plots
Packages
Help
Viewer
New Folder
Upload
Delete
Rename
More
Cloud
project
Name
Size
Modified
..
.gitignore 621 B Jun 23, 2020, 9:21 PM
.Rhistory 0 B Jul 13, 2020, 1:26 PM
.Rprofile 88 B Jun 23, 2020, 9:25 PM
03 – Visualize.Rmd 3 KB Jul 10, 2020, 8:46 AM
04 – Transform.Rmd 4.8 KB Jul 13, 2020, 12:08 PM
05 – Stats.Rmd 5.8 KB Jul 10, 2020, 8:46 AM
06 – Advanced Reporting.Rmd 871 B Jul 13, 2020, 7:08 AM
coursepack
data
LICENSE 1 KB Jun 23, 2020, 9:21 PM
presentations
project.Rproj 205 B Jul 13, 2020, 1:26 PM
README.md 6.9 KB Jul 12, 2020, 3:42 PM


File Edit Code View Plots Session Build Debug Profile Tools Help
Go to file/function
Addins
R 3.5.2
Console
Terminal
Jobs
/cloud/project/
R version 3.5.2 (2018-12-20) -- "Eggshell Igloo"
Copyright (C) 2018 The R Foundation for Statistical Computing
Platform: x86_64-pc-linux-gnu (64-bit)
R is free software and comes with ABSOLUTELY NO WARRANTY.
You are welcome to redistribute it under certain conditions.
Type 'license()' or 'licence()' for distribution details.
R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.
Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.
>
CONSOLE
Environment
History
Connections
Import Dataset
List
Global Environment
Environment is empty
ENVIRONMENT
Files
Plots
Packages
Help
Viewer
New Folder
Upload
Delete
Rename
More
Cloud
project
Name
Size
Modified
..
.Rhistory
0 B
Apr 20, 2019, 1:39 PM
project.Rproj
205 B
Apr 20, 2019, 1:39 PM
OUTPUT

The Basics of Coding

The Basics of Coding: Calculation

- R is a calculator!



Your Turn 1

Place your cursor at the console and click to enter the console.

Complete the following calculation:

-For the date 12-29-1974

-Take the four digit year

-Subtract the month then multiply by the day

What did you get?

- A four digit number? A five digit number?

```
> 1974 - 12 * 29
[1] 1626
>
>
> (1974 - 12) * 29
[1] 56898
```

- Order of operations matters!

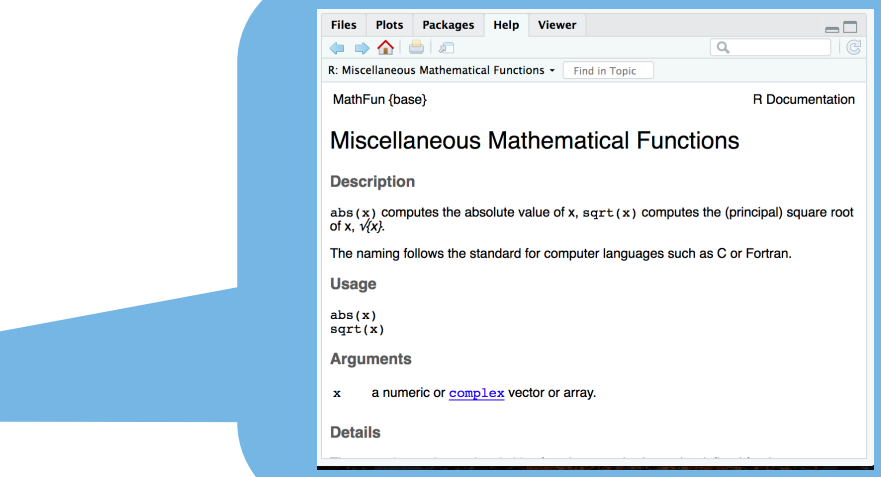
The Basics of Coding: Functions

- Code that extends our reach beyond the basic operators

```
> abs(-77)
[1] 77
>
```

- What if I don't know what a function does?

```
>
> ?abs()
>
```



When you need more help

- The Internet (Stack Overflow: https://stackoverflow.com/)
- Work Aids (RStudio Cheat Sheets: https://www.rstudio.com/resources/cheatsheets/)
- A Good Book (R for Data Science: http://r4ds.had.co.nz/)

Putting Functions to Work

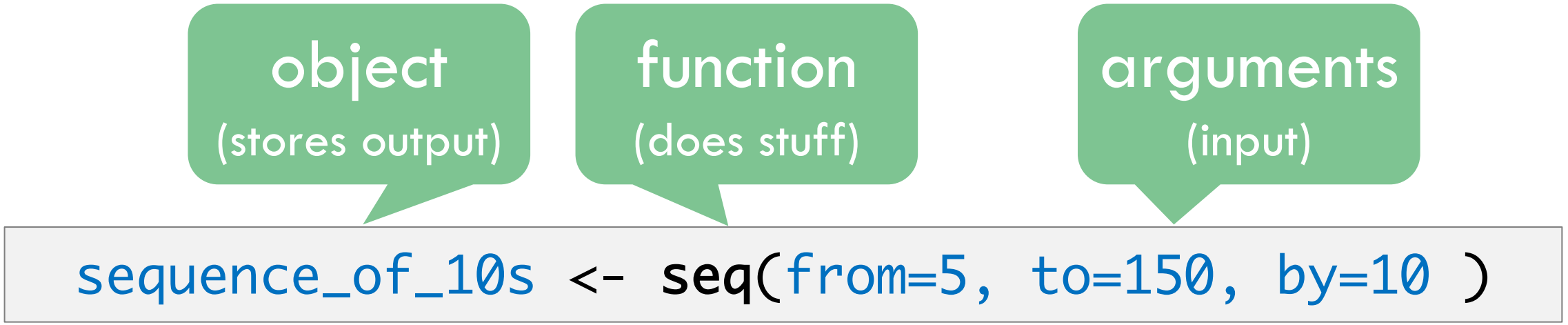
- We can use functions to do more than simple math, we can make things!

- We can create a series of integers (a vector) using the seq() function

```
>
> seq(from=5, to=150, by=10)
 [1]   5  15  25  35  45  55  65  75  85  95 105 115 125 135 145
```

The Basics of Coding: Objects

- Objects are the container for your output



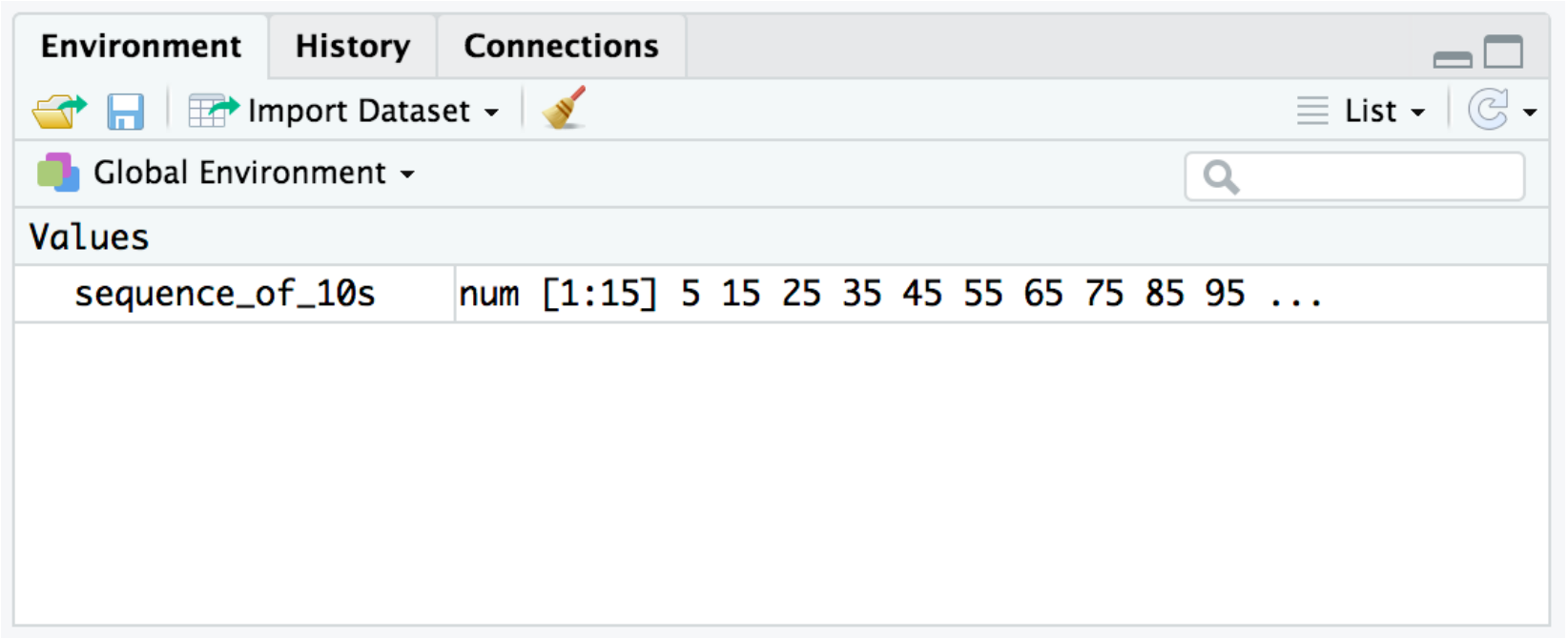
Checking the Contents of an Object

- Entering the object name at the console allows us to output the contents of an object.

```
>
> sequence_of_10s
 [1]   5  15  25  35  45  55  65  75  85  95 105 115 125 135 145
```

Checking the contents of an object

- The environment tab shows us the objects we have created.



Bending objects to your will

- Once we have created an object we can start to interact with it.
- This includes passing our objects to other functions… Whoa!

```
>
> min(sequence_of_10s)
[1] 5
>
> max(sequence_of_10s)
[1] 145
>
```

Your Turn 2

Generate a sequence, store it to an object, and ply your object

Type the following code to create a sequence from 0 to 500 in increments of 25 called sequence_of_25s:

```
sequence_of_25s <- seq(from=0, to=500, by=25)
```

Calculate the median value of this series using the `median()` function

The Basics of Coding: Packages

- A package is a collection of functions.

- Packages extend the capabilities of the base R programming language.

- The **tidyverse** includes functions for reading data into the R environment, cleaning and manipulating data, and plotting our results.

Installing and Loading Packages

- Installing a package


```
install.packages("tidyverse")
```

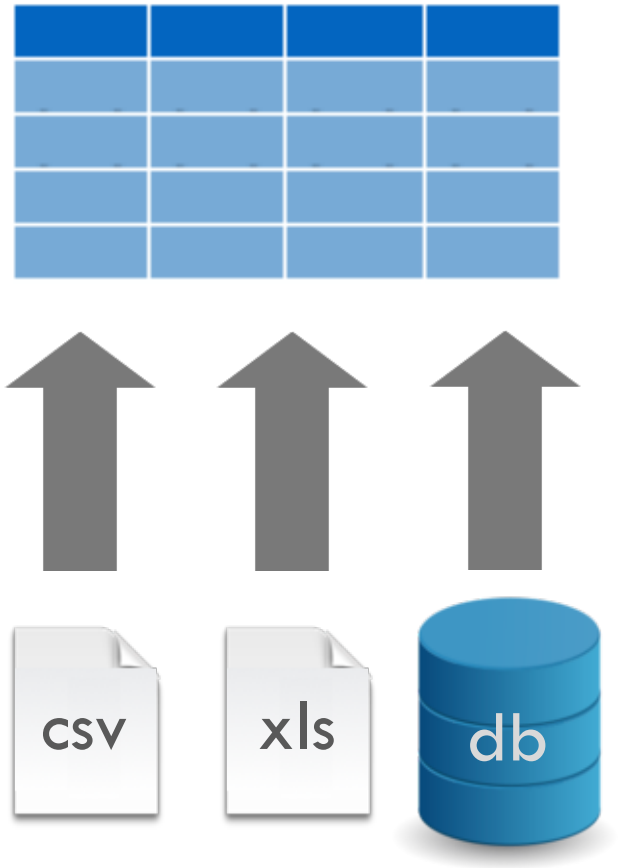
- Loading into your environment

```
library(tidyverse)
```

Importing Data and Working with Dataframes (aka Useful Data)

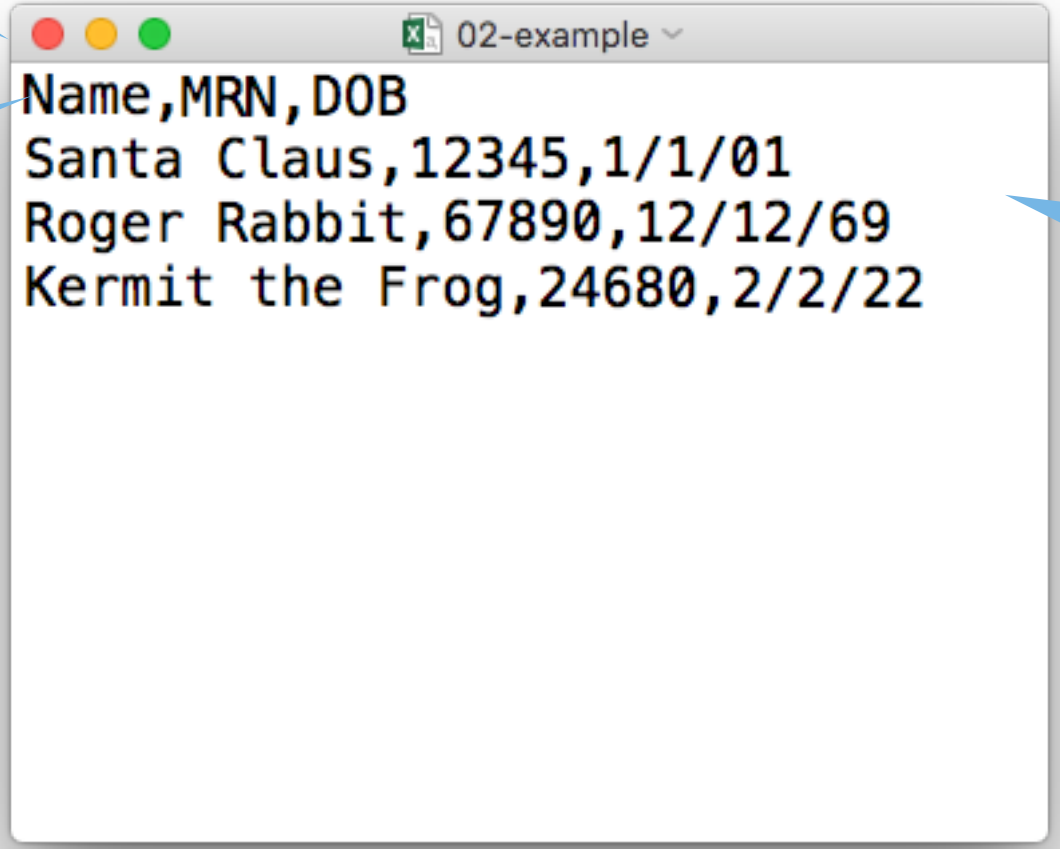
Dataframes: Beyond the Vector

- Dataframe is the term for a table
- Dataframes are composed:
  Columns (Variables)
  Rows (Observations)
- Dataframes are objects and can be acted on like other objects



plain text ("flat") file

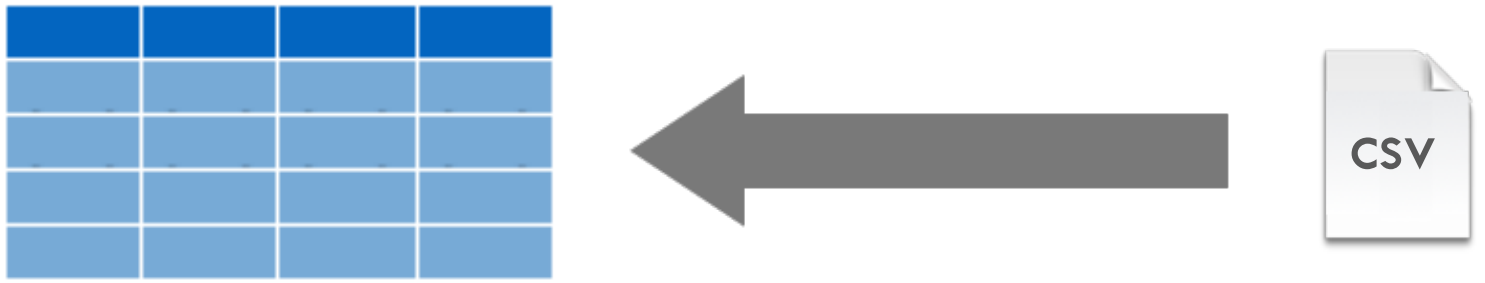
header row



rectangular structure

Loading Data to Create a Dataframe

```
data_frame <- read_csv("file_name")
```



Your Turn 3

Configure environment and load the Covid Testing CSV:

Load the tidyverse library using `library(tidyverse)`

Use the **`read_csv()`** function to load the data

-File_name argument: `"data/covid_testing.csv"`

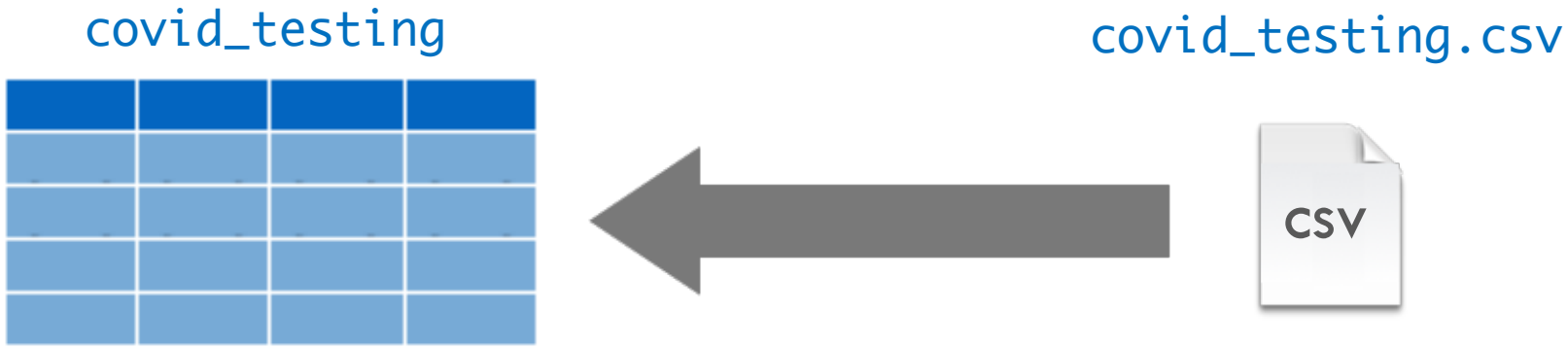
-Object name: `covid_testing`

read_csv()

data frame to read data into

name of CSV file

```
covid_testing <- read_csv("data/covid_testing.csv")
```



What's in a name?

- Capitalization matters

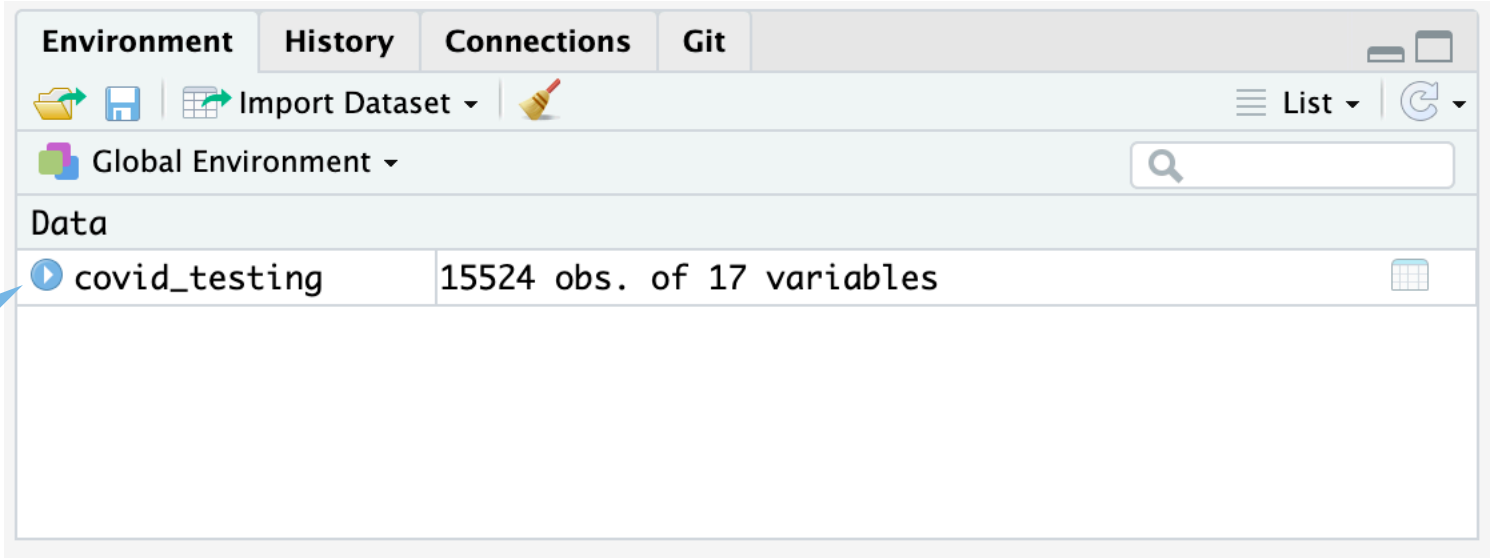
`covid_testing` ≠ `Covid_testing` ≠ `COVID_TESTING`

- Strive for names that are concise and meaningful (not easy!)

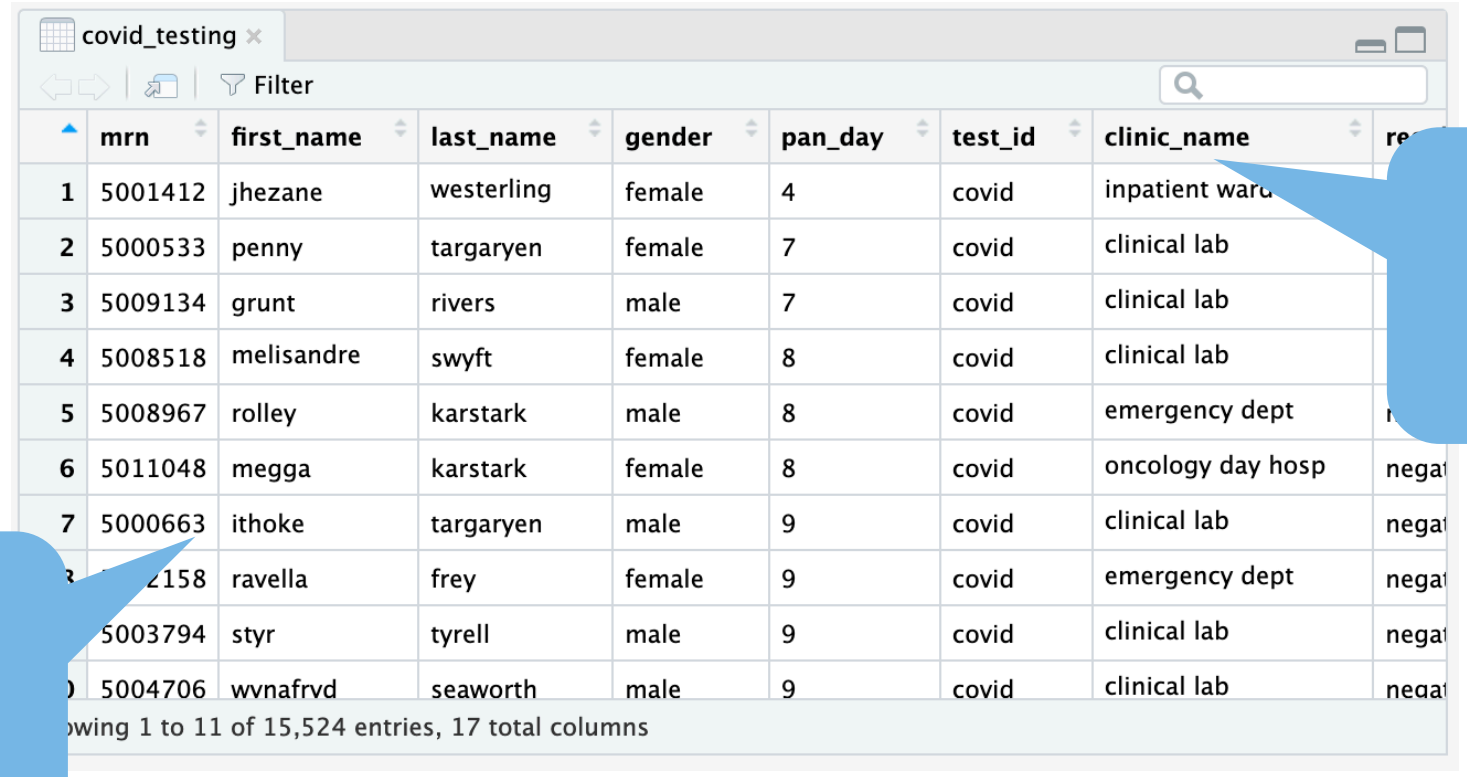
Bad	Still not great	Good
`p`	`name`	`patient_name`

Viewing the Contents of a Dataframe

single click to explore the data



Viewing the Contents of a Dataframe



17 Attributes (Columns)

15,524 Observations (Rows)

Working with Dataframes at the Console

- The head() function is helpful for displaying a snippet of your dataframe

```
head(object_name, n=number of rows to view)
```

Sample of Data in Your Object

Console | Terminal | Jobs

/cloud/project/

```
> head(covid_testing, n=5)
# A tibble: 5 x 17
     mrn first_name last_name gender pan_day test_id clinic_name result demo_group
   <dbl> <chr>      <chr>     <chr>    <dbl> <chr>   <chr>       <chr>  <chr>
1 5.00e6 jhezane    westerli… female       4 covid   inpatient … negat… patient
2 5.00e6 penny      targaryen female       7 covid   clinical l… negat… patient
3 5.01e6 grunt      rivers    male         7 covid   clinical l… negat… patient
4 5.01e6 melisandre swyft     female       8 covid   clinical l… negat… patient
5 5.01e6 rolley     karstark  male         8 covid   emergency … negat… patient
# … with 8 more variables: age <dbl>, drive_thru_ind <dbl>, ct_value <dbl>,
#   orderset <dbl>, payor_group <chr>, patient_class <chr>, col_rec_tat <dbl>,
#   rec_ver_tat <dbl>
>
```

Your Turn 4

Understanding Object Contents

Use the `tail()` function to view the **last 10 rows** of our object `covid_testing`.

-What is the ratio of female to male patients in this subset of data?

A Hero is Born

Lesson Goals

1. Get oriented to R and RStudio
2. Learn some fundamentals of coding

Lesson Objectives

1. Log in and tour RStudio Cloud
2. Execute code at the console
3. Define and use functions
4. Define and create objects in the environment
5. Load data into R and interact with a dataframe