



# Tools for Integrating Data Science in the Statistics Classroom



Dr. Allison Theobold

August 6, 2022

# you...

- know R
- are familiar with R Markdown
- are interested in integrating R into your course(s)

# Tools

1. RStudio Cloud
2. learnr Tutorials

# RStudio Cloud

## Lots of Friction

- Install R
- Install RStudio
- Install the following packages:  
rmarkdown, tidyverse, ...
- Load these packages
- Install git
- Install MiKTeX

## Less Friction

- Go to rstudio.cloud
- Log in

```
> hello R!
```

tl;dr

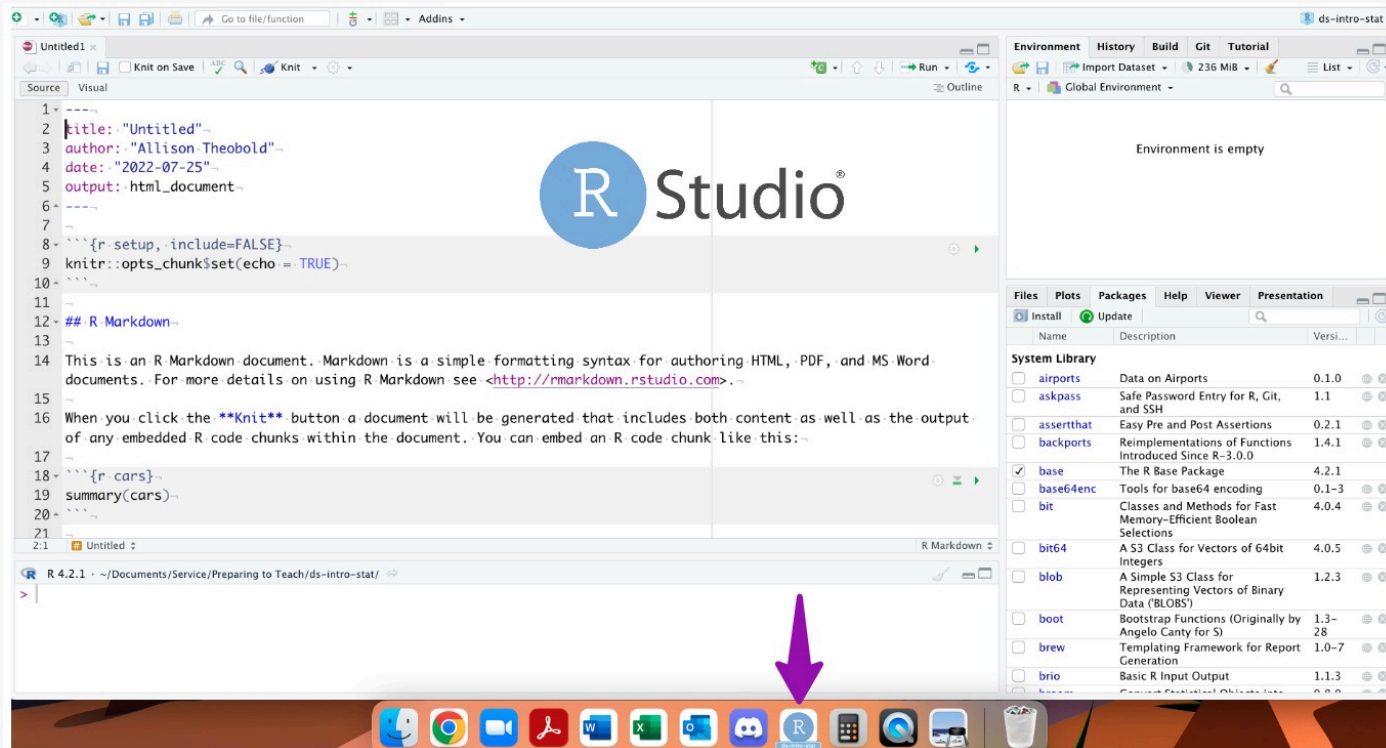
\* Çetinkaya-Rundel, Mine. "Teaching R online with RStudio Cloud." RStudio Webinar.

<https://bit.ly/ptt2022-tools4ds>

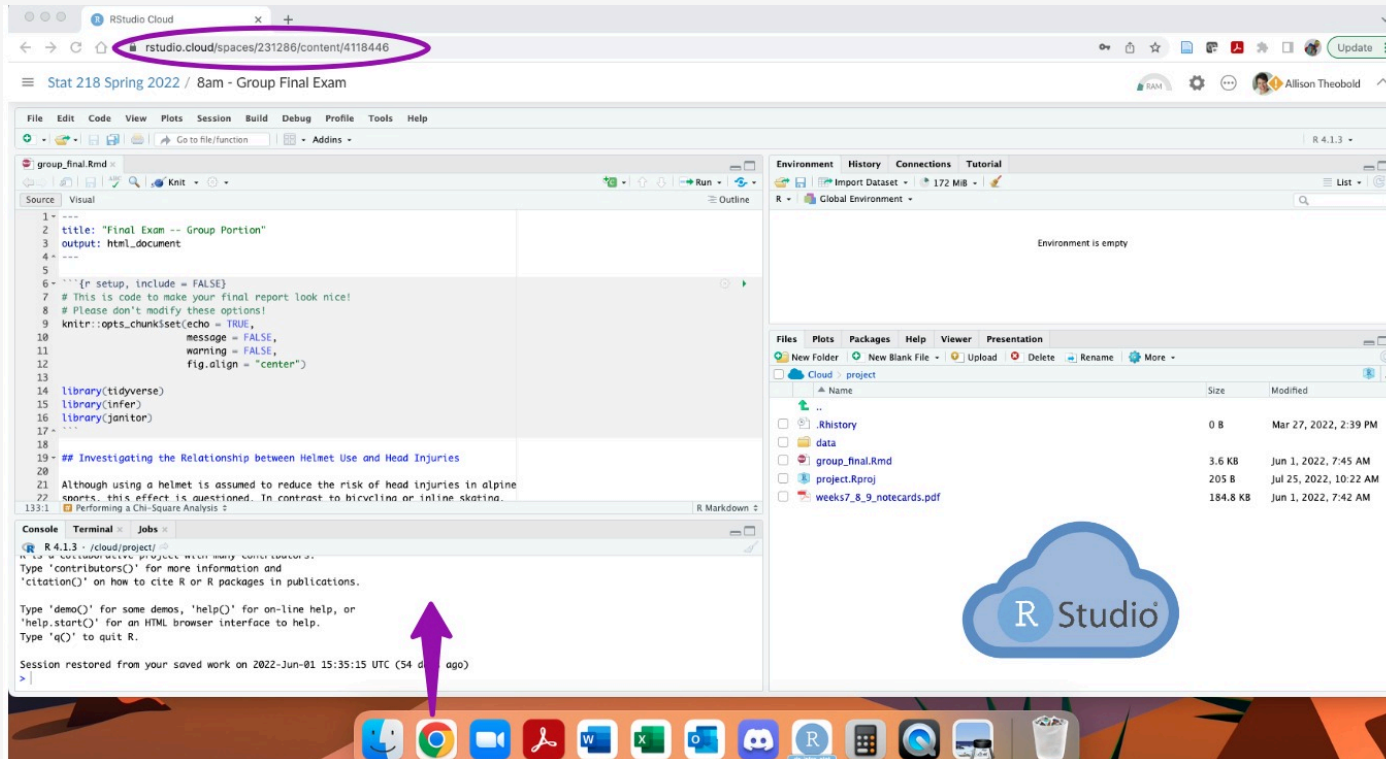
# What is RStudio Cloud?



We created **RStudio Cloud** to make it easy for professionals, hobbyists, trainers, teachers, and students to do, share, teach, and learn data science using R.



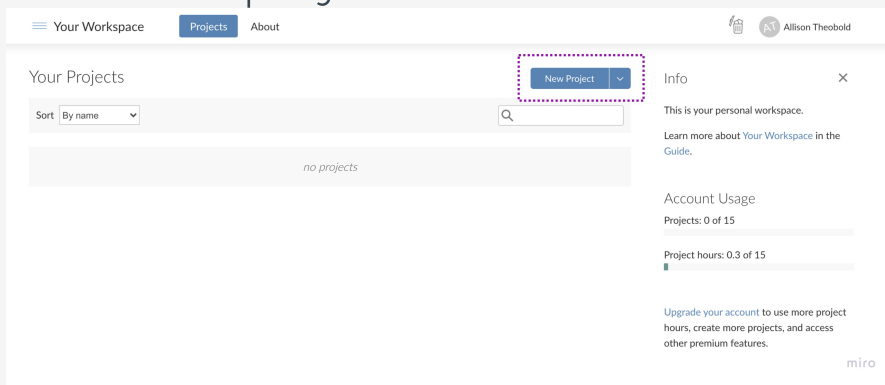
<https://bit.ly/ptt2022-tools4ds>



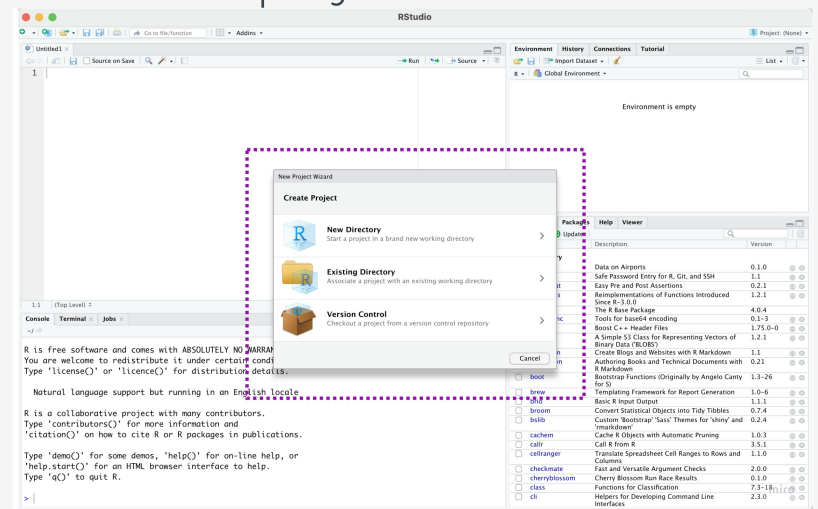
<https://bit.ly/ptt2022-tools4ds>

# Projects

## A new project in RStudio Cloud



## is a new project in RStudio IDE



**PSA:** If you use RStudio, use projects! Trust me, you won't regret it.

<https://bit.ly/ptt2022-tools4ds>



# Workspaces

The screenshot displays the RStudio Cloud interface for a workspace named "Stat 313 Section 01" by Allison Theobald. The interface is divided into a left sidebar, a top navigation bar, and a main content area.

**Left Sidebar:**

- Spaces:** A list of workspaces including "Your Workspace", "Going live - USCOTS 2021", "Stat 218 - W 2022", "Stat 218 Spring 2022", "Stat 313 Section 01" (selected), and "Stat 313 Section 02".
- Classroom:** A section with a "classroom" image and a "What's New" notification.
- Help:** Links to "Primers", "Cheat Sheets", "Current System Status", and "RStudio Community".

**Top Navigation Bar:**

- Stat 313 Section 01:** The workspace name and owner.
- Projects:** The active tab, showing a list of projects.
- Members:** A tab for viewing workspace members.
- Usage:** A tab for viewing usage statistics.
- About:** A tab for workspace details.

**Main Content Area:**

- All Projects (9):** A list of projects in the workspace.
- Final Project:** An assignment project created on Nov 29, 2021.
- Lab 0:** An assignment project created on Sep 21, 2021.
- Lab 1 - Categorical Data:** An assignment project created on Sep 28, 2021.
- Lab 2 - Numerical Data:** An assignment project created on Oct 6, 2021.
- Lab 3 - SLR:** An assignment project created on Oct 12, 2021.
- Lab 4 - Confidence Intervals:** An assignment project created on Nov 3, 2021.
- Lab 5 - Hypothesis Tests:** An assignment project created on Nov 10, 2021.

**Right Sidebar:**

- sandbox:** A section with a "sandbox" image.

<https://bit.ly/ptt2022-tools4ds>

# Sharing Options

Option 1: Share a single project

Option 2: Share a workspace  
(presumably with many projects)

# Sharing a Single Project

## Good!

- Students land directly in a project upon login
- Works well for workshops where all work will be completed in a single project
- Also great for sharing code in general, e.g. collaboration, reprints, etc.

## Not so good...

- Students need to remember to make a copy of the project (which means you need to remember to remind them!)
- You can't keep track which students started their assignment
- You can't easily peek into student projects -- they would need to explicitly share the project with you

# Sharing a Workspace

## Good!

- Base projects with desired packages installed
- Assignments -- no more "make a copy of the project before starting work"
- Collaborate with students *inside* their projects

## Not so good...

- Students land in the workspace, may need to provide instructions for the next steps
- Git config for each project can get tedious and doesn't reflect realistic practice

# Q: How do I access RStudio Cloud?

## Cloud Instructor

We offer our premium features to qualified instructors and their students at a deep discount for instructional use - we want to make it an easy decision for you to teach with RStudio Cloud.

The price for the instructor is \$15 per month - and we offer three convenient ways to cover student costs to fit your budgeting requirements.



Students Pay



Instructor Pays



School Pays

## Requirements

- You are teaching at an accredited institution that meets one of the following qualifications:
  - a public or private university or college (including community, junior or vocational college) that grants degrees requiring not less than the equivalent of two years of full-time study
  - a public or private primary or secondary school providing full-time instruction for grades K-12
  - a hospital that is wholly owned and operated by an education institution
  - a higher-education research laboratory that is a public institution, teaches students, and can provide evidence of relationships with universities
- Or ... you are an RStudio Certified Instructor

---

<https://bit.ly/ptt2022-tools4ds>

# learnr Tutorials

---

<https://bit.ly/ptt2022-tools4ds>



- **learnr** is an R package that makes it easy to create interactive tutorials from R Markdown documents.
- Tutorials can include:
  - Narrative, figures, illustrations, and equations
  - Code exercises (R code chunks that users can edit and execute directly)
  - Multiple choice questions
  - Videos (YouTube, Vimeo)
  - Interactive Shiny components
- learnr is on CRAN

```
install.packages("learnr")
```

← → ↺ 🏠

openintro.shinyapps.io/ims-05-introduction-to-statistical-inference-02/#section-completing-a-randomization-test-gender-discrimination

☆ 📄 🔍 👤 ⋮

Introduction to Statistical Inference: 2 - Randomization Test

Completing a randomization test:

gender discrimination

Distribution of statistics

Why 0.05?

What is a p-value?

Summary of gender discrimination

Congratulations!

Start Over

Gender discrimination hypotheses

Which of the following null and alternative hypotheses are appropriate for the gender discrimination example described in the previous lesson?

☐ H0: gender and promotion are unrelated variables. HA: men are more likely to be promoted.

☐ H0: gender and promotion are unrelated variables. HA: women are more likely to be promoted.

☐ H0: men are more likely to be promoted. HA: gender and promotion are unrelated variables.

☐ H0: women are more likely to be promoted. HA: gender and promotion are unrelated variables.

Submit Answer

Summarizing gender discrimination

As the first step of any analysis, you should look at and summarize the data. Categorical variables are often summarized using proportions, and it is always important to understand the denominator of the proportion.

Do you want the proportion of women who were promoted or the proportion of promoted individuals who were women? Here, you want the first of these, so in your R code it's necessary to group by `gender` before you calculate the proportions!

The discrimination study data are available in your workspace as `gender_discrimination`.

- Using the `count()` function, tabulate the variables `gender` and `decision`.
- Group the data by `gender`.
- Calculate the proportion of those who were and were not promoted in each gender and call this variable `prop`.

R code ⏮ Start Over 🗨 Hints ▶ Run Code

```
1 gender_discrimination %>%
2   count(____, ____)%>%
3   group_by(____)%>%
4   mutate(____ = ____ / ____)
```

<https://bit.ly/ptt2022-tools4ds>



# demo

<https://rstudio.github.io/learnr/articles/examples.html>

[tutorial]

[code]

# Components of a `learnr` tutorial

# YAML

Start with a YAML, just like in R Markdown:

```
---  
title: "Starting with Data"  
output:  
  learnr::tutorial:  
    progressive: true  
    allow_skip: true  
runtime: shiny_prerendered  
---
```

1. Create a new RMarkdown file
2. Select from Template
3. Choose the Interactive Tutorial template from **learnr**
4. Start editing!

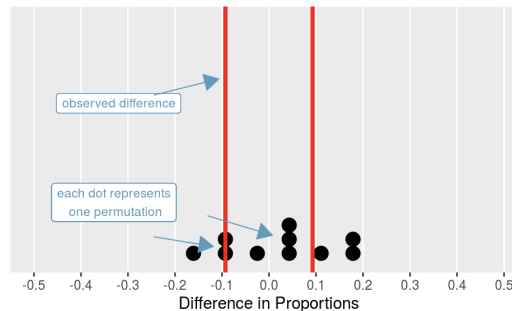
# Narrative

- R Markdown style section and subsection headings with ##, ###, etc.
- Text, figures, illustrations, and equations.
- Videos: supported services include YouTube and Vimeo

## Using the randomization distribution

Recall that the logic of statistical inference is to compare the observed statistic to the distribution of statistics that come from a null distribution. You've now seen how to create the distribution with your own R code. The next question to ask is, **how do we use the information in the null distribution?**

Remember that each dot that gets generated is from a different permutation of the data. We use the null differences, the dots, to define the setting that we are not interested in. The goal is to show that our observed data are not consistent with the differences generated. We want our observed data to be different from the null so that we can claim the alternative research hypothesis to be true.





# Multiple choice questions


```
quiz(  
  question("What position is the letter A in the english alphabet?",  
    answer("8"),  
    answer("14"),  
    answer("1", correct = TRUE),  
    answer("23"),  
    incorrect = "See [here](https://en.wikipedia.org/wiki/English_alphabet) and try again.",  
    allow_retry = TRUE  
  ),  
  
  question("Where are you right now? (select ALL that apply)",  
    answer("Planet Earth", correct = TRUE),  
    answer("Pluto"),  
    answer("At a computing device", correct = TRUE),  
    answer("In the Milky Way", correct = TRUE),  
    incorrect = paste0("Incorrect. You're on Earth, ",  
                        "in the Milky Way, at a computer.")  
  )  
)
```

# Code exercises - rendered

R code

 Start Over

 Hints

 Run Code

```
1 gender_discrimination %>%  
2   count(____, ____)%>%  
3   group_by(____)%>%  
4   mutate(____ = ____ / ____)
```

# Code exercises - code

```
```{r gender-promoted, exercise=TRUE}
gender_discrimination %>%
  count(, ) %>%
  group_by( ) %>%
  mutate( = / )
```

```{r gender-promoted-hint-1}
gender_discrimination %>%
  count(gender, decision) %>%
  group_by( ) %>%
  mutate( = / )
```

```{r gender-promoted-hint-2}
gender_discrimination %>%
  count(gender, decision) %>%
  group_by(gender) %>%
  mutate( = / )
```
```

# Code exercises - solution

**Solution** [Copy to Clipboard](#)

```
1 # Calculate the observed difference in promotion rate
2 diff_orig <- gender_discrimination %>%
3   # Group by gender
4   group_by(gender) %>%
5   # Summarize to calculate proportion promoted
6   summarize(prop_promoted = mean(decision == "promoted")) %>%
7   # Summarize to calculate difference
8   summarize(stat = diff(prop_promoted))
9
10 # See the result
11 diff_orig
```

**R code** [Start Over](#) [Solution](#) [Run Code](#)

```
1 # Calculate the observed difference in promotion rate
2 diff_orig <- gender_discrimination %>%
3   # Group by gender
4   group_by(____) %>%
5   # Summarize to calculate proportion promoted
6   summarize(prop_promoted = mean(decision == "promoted")) %>%
7   # Summarize to calculate difference
8   ____ (stat = ____ (____))
9
10 # See the result
11 diff_orig
```

Continue



# *Q: How do I share with my students?*

- Deploy on
  - shinyapps.io (variety of pricing plans available)
  - RStudio Connect (free for academic use, requires setup)
- Essential reading:
  - Publishing learnr Tutorials on shinyapps.io by Angela Li
  - Teach R with learnr: a powerful tool for remote teaching by Allison Horst
  - See the publishing instructions on the learnr website for step-by-step instructions

# Questions to Ponder

---

<https://bit.ly/ptt2022-tools4ds>

# What are my resources?

- Does your university have server time for hosting **learnr** tutorials?
- If not, does your university have funding sources to host these things elsewhere? (e.g. shinyapps.io)
- Can you acquire funding for RStudio Cloud? Can you charge students to use it?
- How much do **you** enjoy creating / teaching with R resources?

# What are my learning objectives?

- If **learning R** is one of them...

...students should probably have a native install!

- If **understanding how statisticians use code** is one of them ...

... consider **learnr** tutorials with pre-supplied code or pre-made RS Cloud.

- If **software is not a learning objective**...

... consider using R as a back-end only, to make your own life easier.