🔨 Tools for Integrating Data Science in the Introductory and Intermediate Statistics Classroom

🔗 http://bit.ly/tools4ds

🌈 Dr. Allison Theobold
🎲 Dr. Kelly Bodwin
February 19, 2021

you...

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

Tools

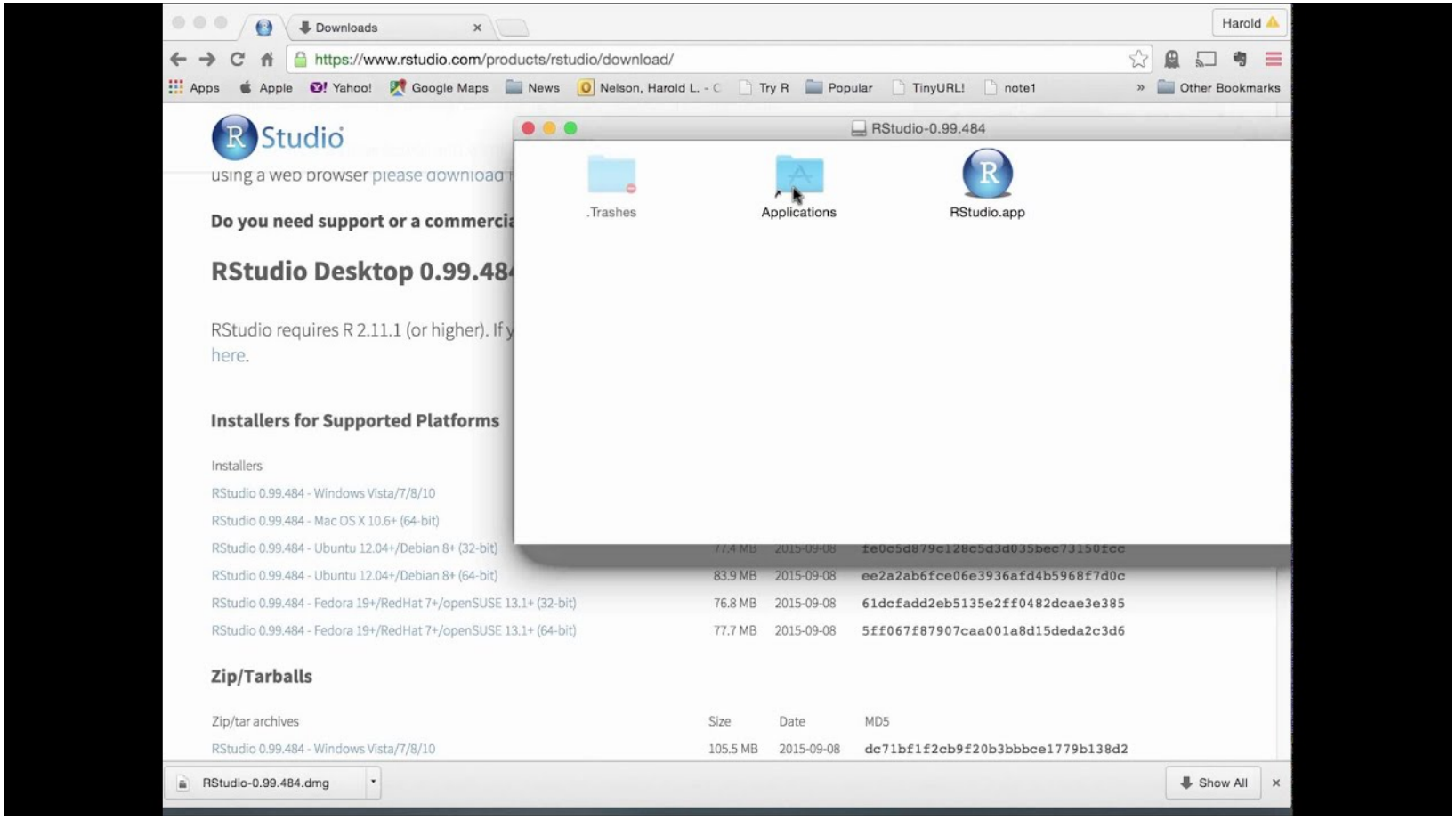
1. At-home Install
2. RStudio Cloud
3. Shiny Apps
4. learnr Tutorials

At-home Installs

Cautionary Tale 1: Old versions of Mac

- Most of what you want to teach will work with older versions of R...
- ... but earlier than **R 3.4** is a problem.
- **Solution**: Have students on old machines use RS Cloud.

Cautionary Tale 2: Installing on any Mac



Cautionary Tale 3: All the package headaches

```
> library(multcomp)
Error: package 'mvtnorm' required by 'multcomp' could not be found
> install.packages("multcomp")
also installing the dependency 'mvtnorm'


  There is a binary version available but the source version is
  later:
        binary source needs_compilation
mvtnorm  1.1-0  1.1-1              TRUE

Do you want to install from sources the package which needs compilation? (Yes/no/cancel)
trying URL 'https://cran.rstudio.com/bin/macosx/contrib/4.0/multcomp_1.4-13.tgz'
Content type 'application/x-gzip' length 728457 bytes (711 KB)
==================================================
downloaded 711 KB

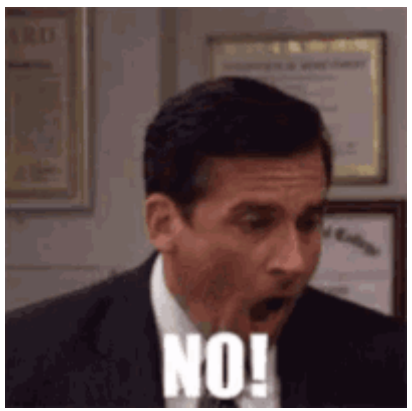

The downloaded binary packages are in
        /var/folders/dd/21svxs313wjfmtnhm2qvchkh0000gn/T//RtmpWNrxCo/downloaded_packages
installing the source package 'mvtnorm'

trying URL 'https://cran.rstudio.com/src/contrib/mvtnorm_1.1-1.tar.gz'
Content type 'application/x-gzip' length 163902 bytes (160 KB)
==================================================
downloaded 160 KB
```

Cautionary Tale 3: Continued...

Do you want to install from sources the package which needs compilation?

Do you want to install from **developer version** sources **as opposed to CRAN version** the package which needs compilation?



Cautionary Tale 3: Continued...

```
Error in loadNamespace(name) : there is no package called 'here'
```

or

```
Error in library(here) : there is no package called 'here'
```

- This means that a dependency didn't auto-install like it should have.
- Why?
- No idea. Just install the package directly

```
install.packages("here")
```

Cautionary Tale 3: Continued...

```
Error: package or namespace load failed for 'tidyverse' in loadNamespace(j <-
i[[1L]], c(lib.loc, .libPaths()), versionCheck = vI[[j]]): namespace 'tibble'
2.1.3 is already loaded, but >= 3.0.0 is required
```

This means you're trying to update a package that's in use!

Clear out the **environment**, including "hidden objects"

Close other R/RStudio windows.

Try to install again.

Cautionary Tale 4: Gen Z doesn't understand folders

Things I have seen students do...

- Download a dataset to a random folder, expect it to be read into R automatically.
- Never **save** or **knit** their R Markdown file. Lose everything when something crashes.
- Re-download the lab every time they want to work on it, open it from their downloads folder, copy paste code from a Google Doc.
- Re-download R or RStudio every time they want to open it.

Cautionary Tale 5: Equality of assessment

- Using a Studio computer means students taking exams out of their comfort zone...
- ... but troubleshooting individual laptops during an exam is impossible!
- Not all students can afford the most powerful computer.
- Not all students have a laptop computer they can bring to class.

RStudio Cloud

Lots of Friction

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

Less Friction

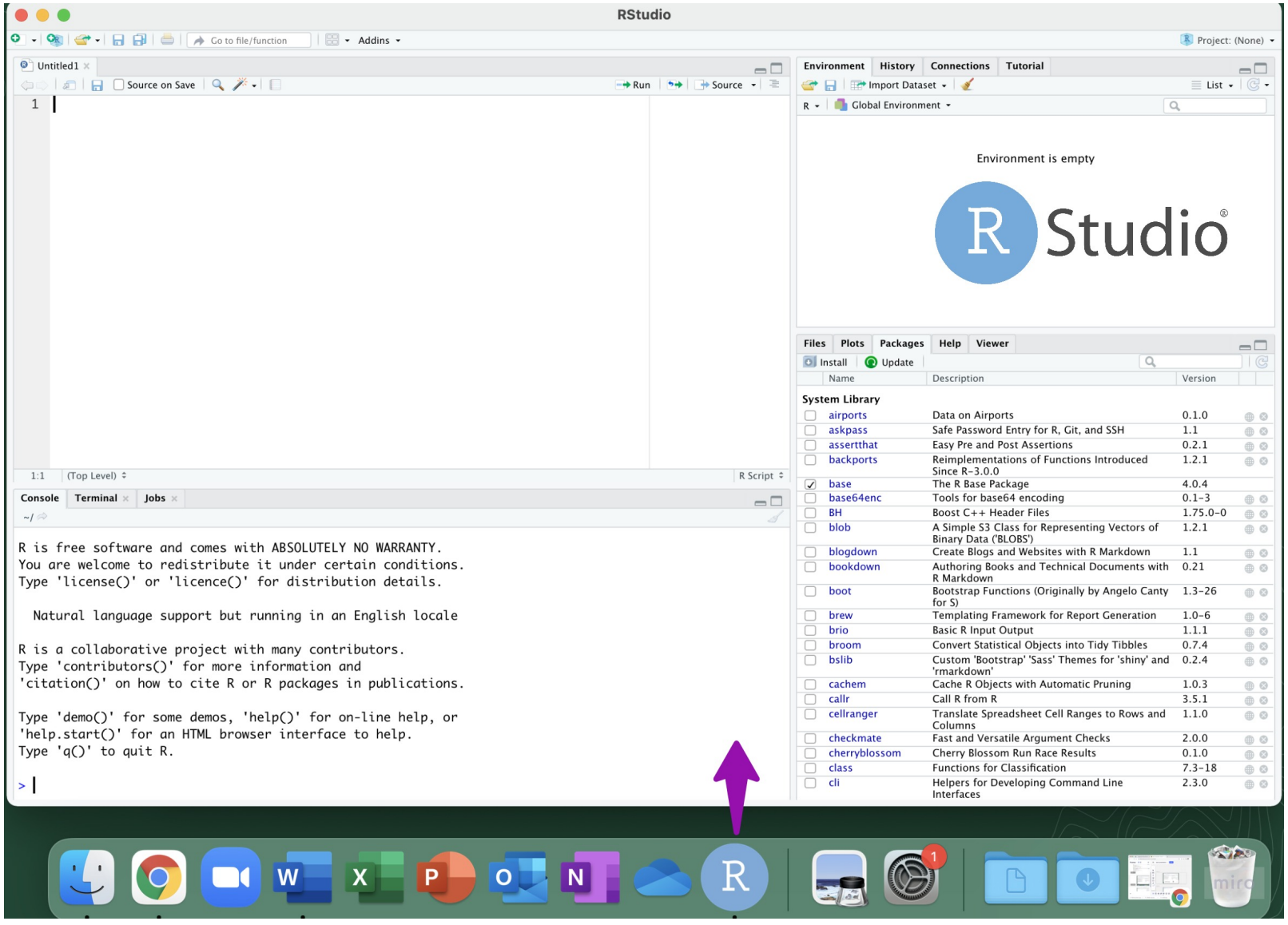
- Go to rstudio.cloud
- Log in

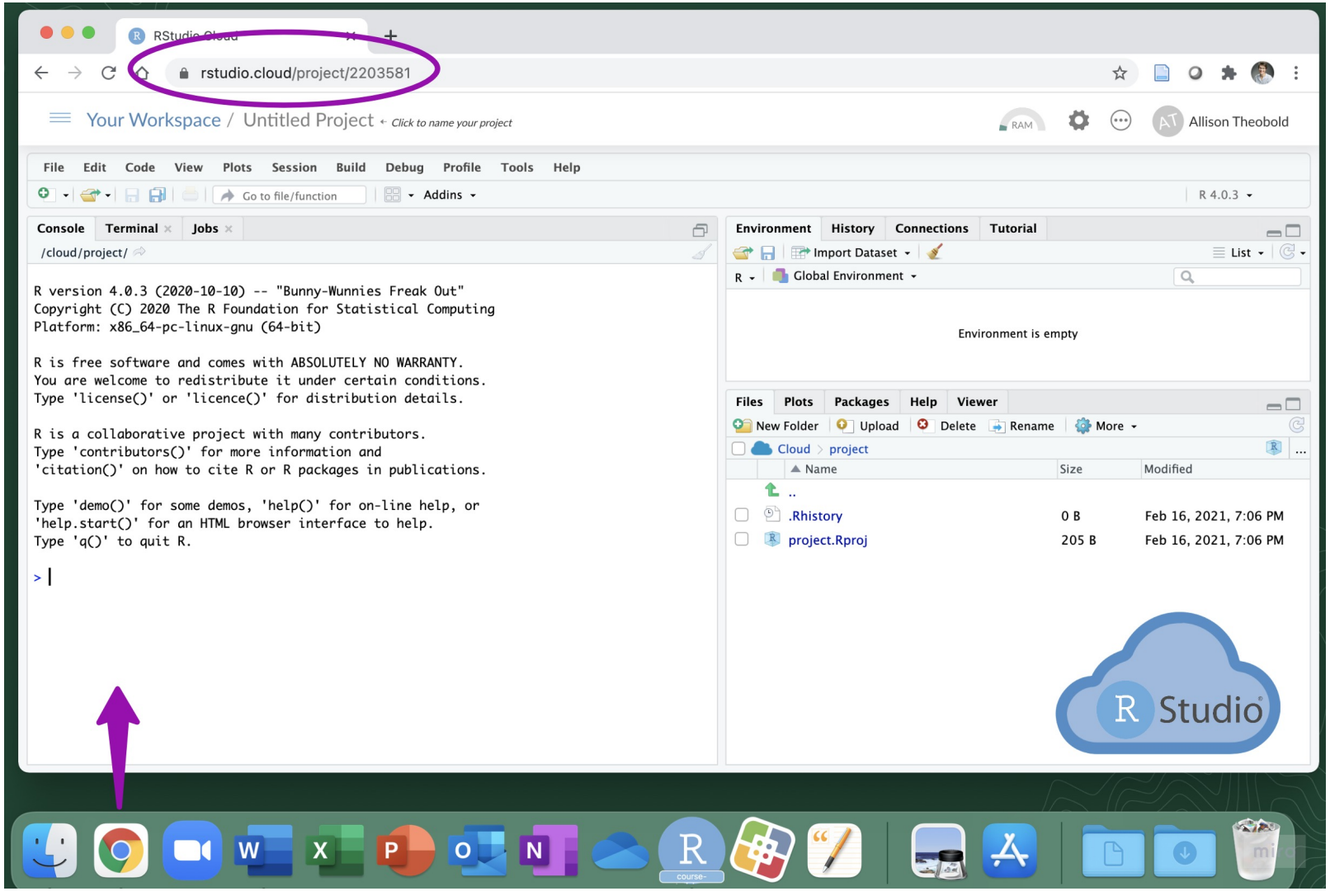
```
>hello R!
```

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

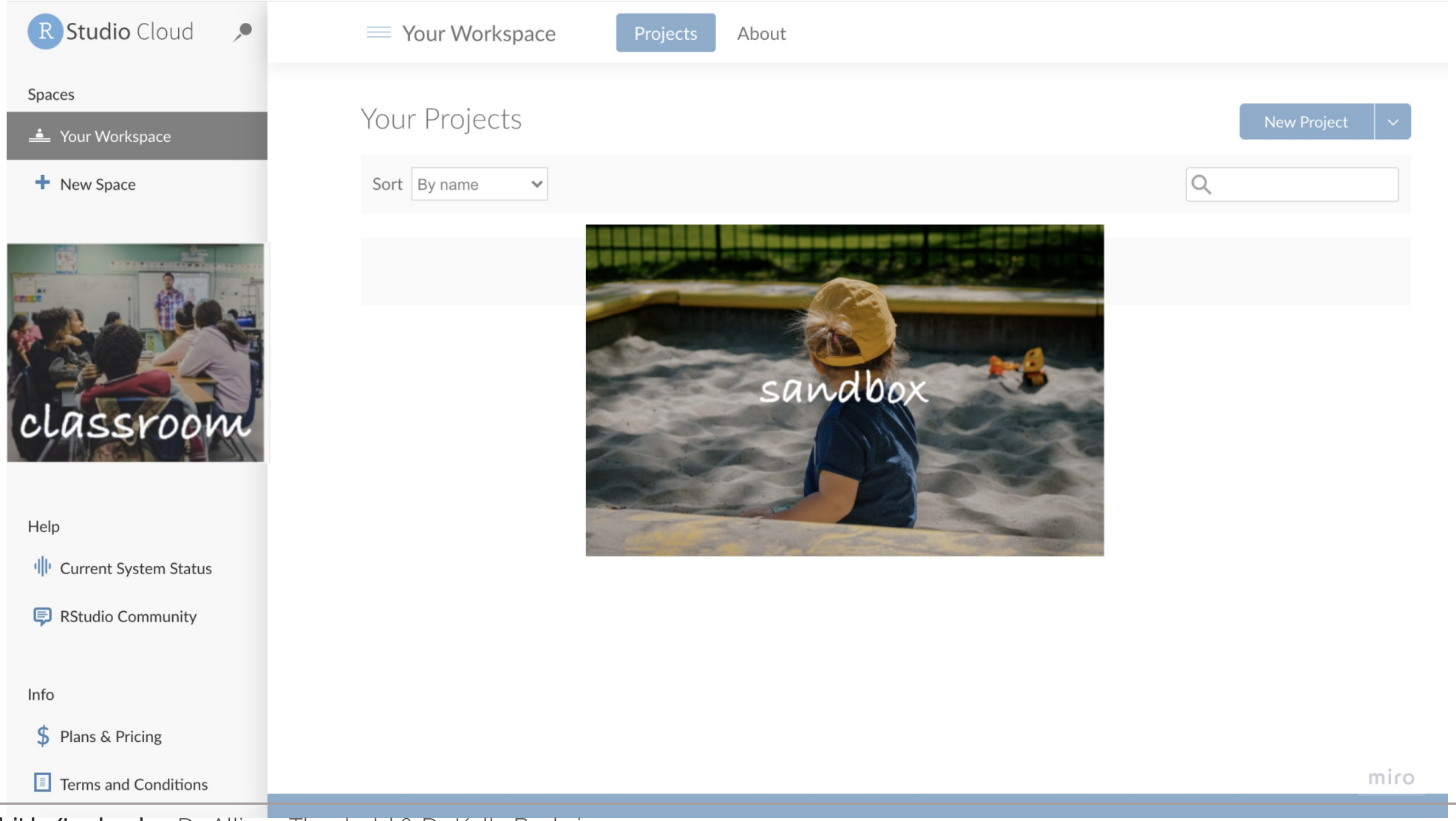
What is RStudio Cloud?

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.


RStudio
Go to file/function
Addins
Project: (None)
Untitled1
Source on Save
Run
Source
1
1:1
(Top Level)
R Script
Environment
History
Connections
Tutorial
Import Dataset
List
R
Global Environment
Environment is empty
RStudio
Files
Plots
Packages
Help
Viewer
Install
Update
Name
Description
Version
System Library
airports Data on Airports 0.1.0
askpass Safe Password Entry for R, Git, and SSH 1.1
assertthat Easy Pre and Post Assertions 0.2.1
backports Reimplementations of Functions Introduced Since R-3.0.0 1.2.1
base The R Base Package 4.0.4
base64enc Tools for base64 encoding 0.1-3
BH Boost C++ Header Files 1.75.0-0
blob A Simple S3 Class for Representing Vectors of Binary Data ('BLOBS') 1.2.1
blogdown Create Blogs and Websites with R Markdown 1.1
bookdown Authoring Books and Technical Documents with R Markdown 0.21
boot Bootstrap Functions (Originally by Angelo Canty for S) 1.3-26
brew Templating Framework for Report Generation 1.0-6
brio Basic R Input Output 1.1.1
broom Convert Statistical Objects into Tidy Tibbles 0.7.4
bslib Custom 'Bootstrap' 'Sass' Themes for 'shiny' and 'rmarkdown' 0.2.4
cachem Cache R Objects with Automatic Pruning 1.0.3
callr Call R from R 3.5.1
cellranger Translate Spreadsheet Cell Ranges to Rows and Columns 1.1.0
checkmate Fast and Versatile Argument Checks 2.0.0
cherryblossom Cherry Blossom Run Race Results 0.1.0
class Functions for Classification 7.3-18
cli Helpers for Developing Command Line Interfaces 2.3.0
Console
Terminal
Jobs
~/
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.
Natural language support but running in an English locale
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.
>

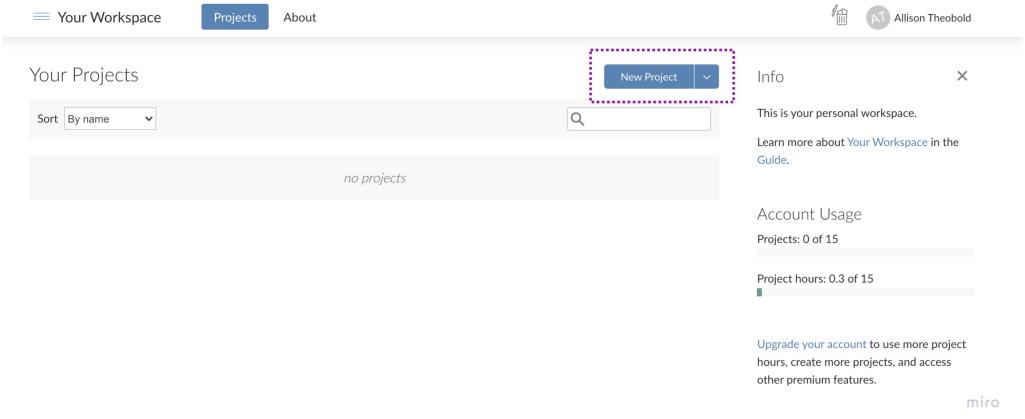

rstudio.cloud/project/2203581
Your Workspace / Untitled Project
Click to name your project
Allison Theobold
File Edit Code View Plots Session Build Debug Profile Tools Help
Go to file/function
Addins
R 4.0.3
Console Terminal Jobs
/cloud/project/
R version 4.0.3 (2020-10-10) -- "Bunny-Wunnies Freak Out"
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 Tutorial
Import Dataset
List
R
Global Environment
Environment is empty
Files Plots Packages Help Viewer
New Folder
Upload
Delete
Rename
More
Cloud
project
Name
Size
Modified
..
.Rhistory
0 B
Feb 16, 2021, 7:06 PM
project.Rproj
205 B
Feb 16, 2021, 7:06 PM
RStudio

Workspaces

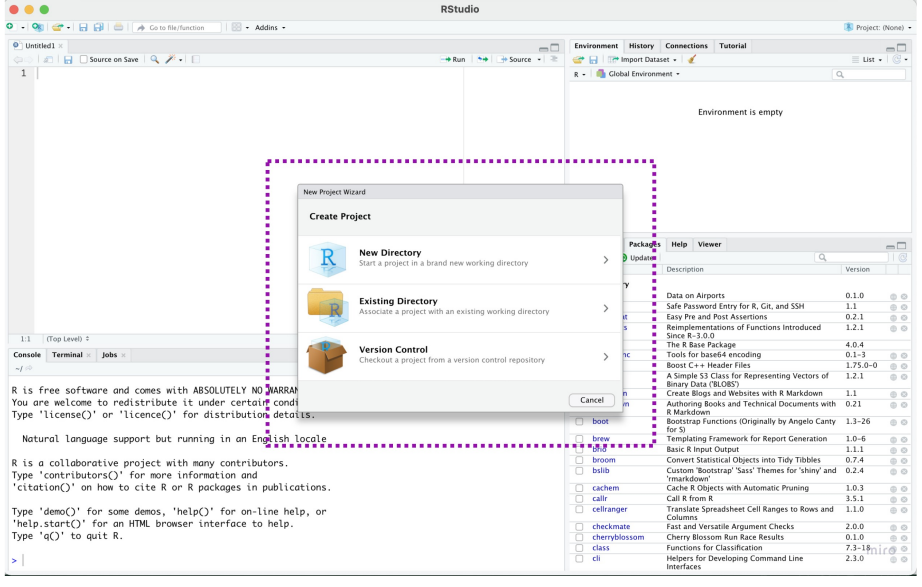

Studio Cloud
Your Workspace
Projects
About
Spaces
Your Workspace
New Space
Your Projects
New Project
Sort
By name
classroom
sandbox
Help
Current System Status
RStudio Community
Info
Plans & Pricing
Terms and Conditions
miro

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.

Sharing Options

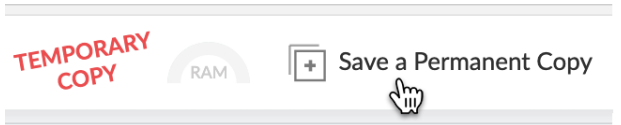
- Option 1: Share a single project
- Option 2: Invite users to a workspace (presumably with many projects)

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, reprexes, 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

Sharing a Workspace

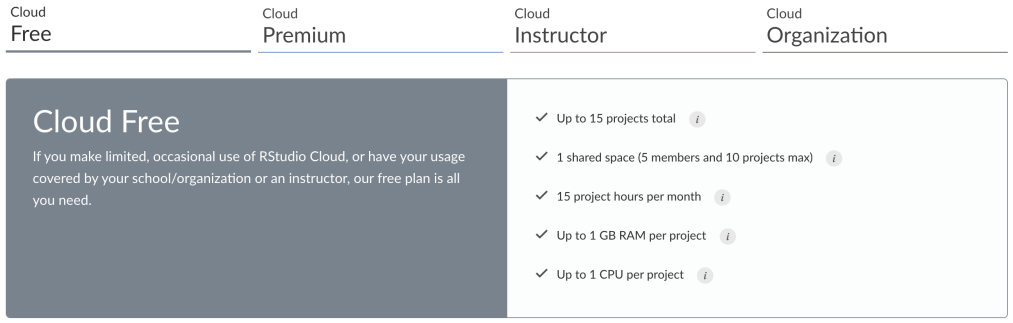
Good!

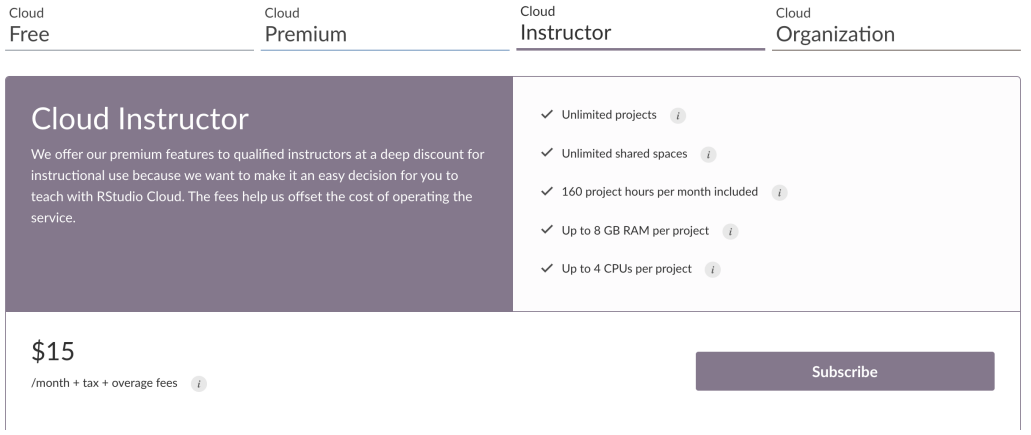
- Base projects with desired packages installed
- Assignments -- no more "make a copy of the project before starting work"
- Peek into students' projects

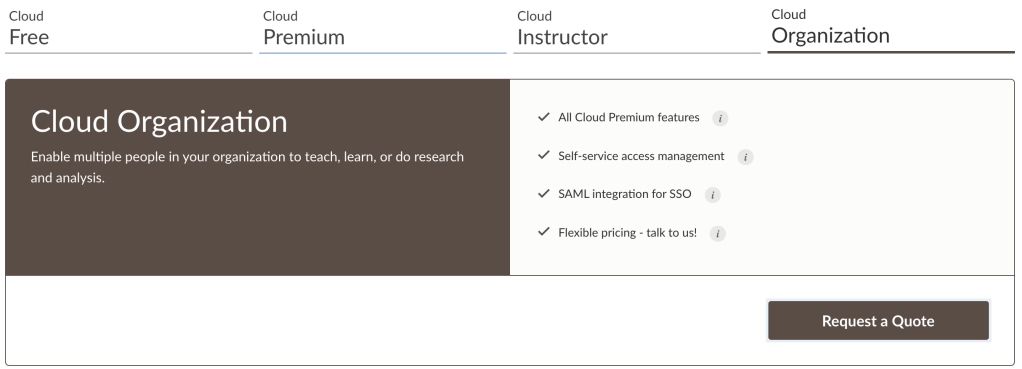
Not so good...

- 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?



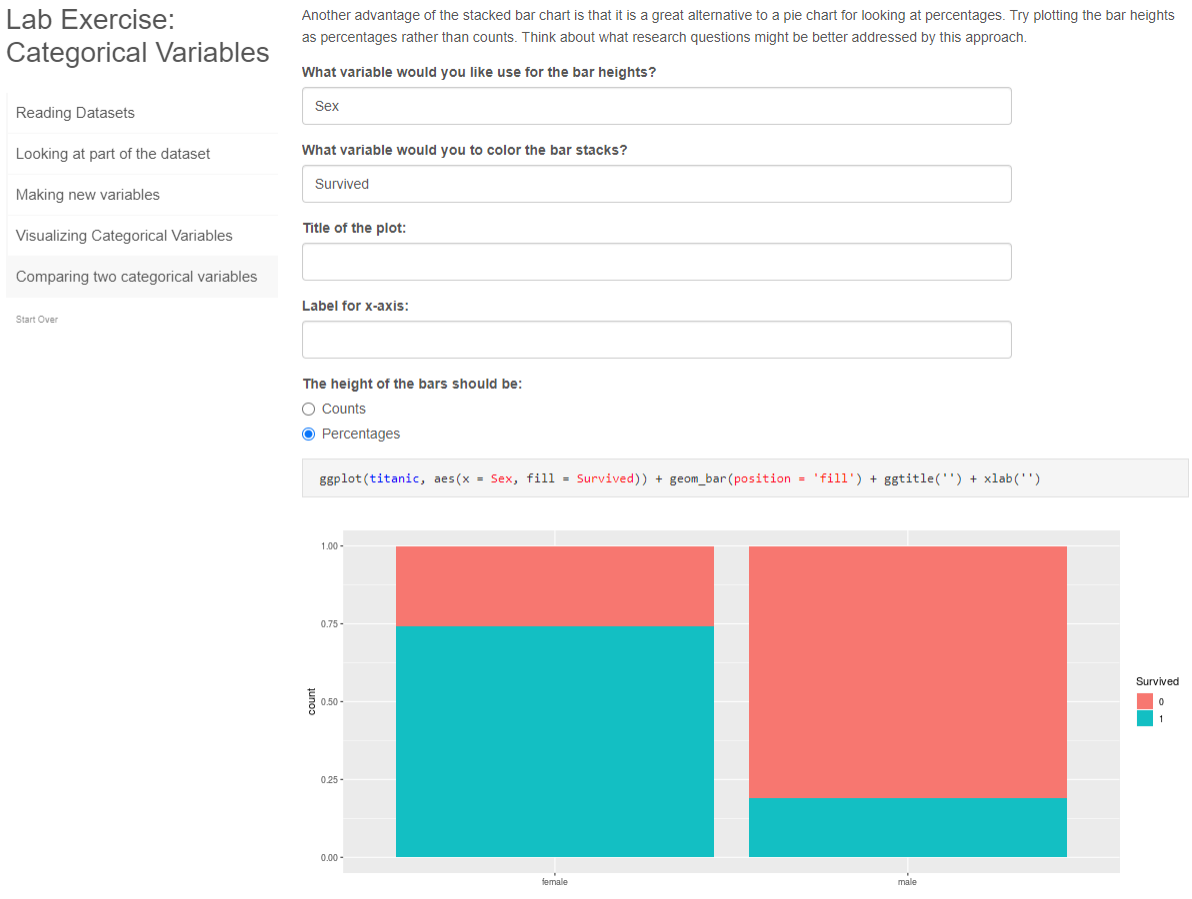




Shiny Apps

demo:

https://kbodwin.shinyapps.io/Lab_Exercise_CatVars2



Good!

- Students get experience with code without typing code
- Gives them an easy "copy-paste" resource
- No need to install or interact with R/RStudio

Not so good...

- Apps need to be hosted somewhere
- Many students using app at once = high computing demand
- Hard to create (for now)

Solutions: shindig

shindig

- R package to create your own exercises
- Distribute exercises to students in package form
- More info: https://rstudio.github.io/learnr/publishing.html

```
penguins %>%
  filter(bill_length_mm > 5) %>%
  ggplot(aes(x  = {quant_var}, fill = {cat_var})) +
  geom_histogram()
```

```
textInput("quant_var",
          "What quantitative variable?")

textInput("cat_var",
          "What categorical variable?")
```

```
decorate_shiny("plot1") %>%
  shindig("renderPlot")
```

```
plotOutput("plot1")
```

learnr

Tutorials

learnr

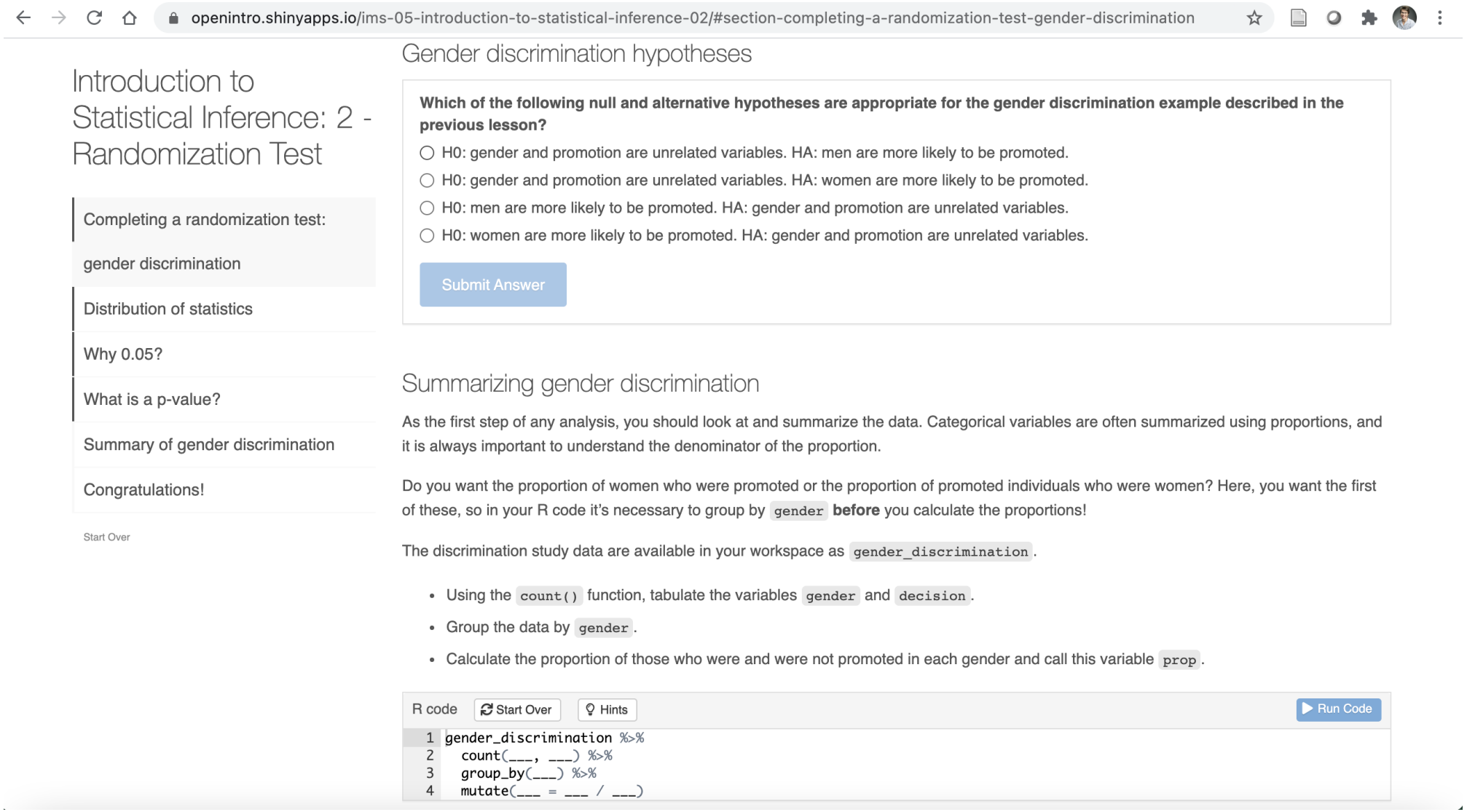
- learnr is an R package that makes it easy to create interactive tutorials from R Markdown documents.

- 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(___ = ___ / ___)

demo

[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!

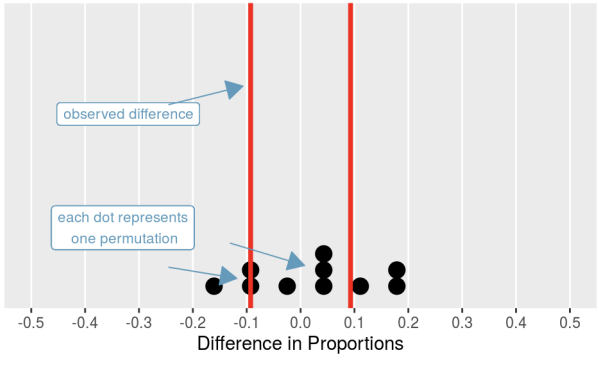
Narrative

- 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.

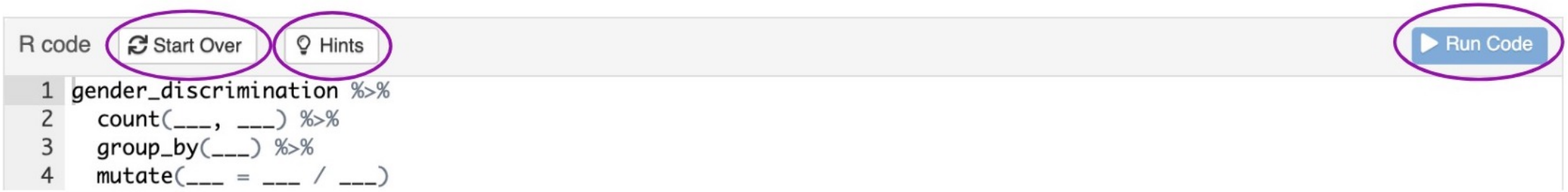


Multiple choice questions

```
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



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

```r
# Calculate the observed difference in promotion rate
diff_orig <- gender_discrimination %>%
  # Group by gender
  group_by(gender) %>%
  # Summarize to calculate proportion promoted
  summarize(prop_promoted = mean(decision == "promoted")) %>%
  # Summarize to calculate difference
  summarize(stat = diff(prop_promoted))

# See the result
diff_orig
```

R code | Start Over | Solution | Run Code

```r
# Calculate the observed difference in promotion rate
diff_orig <- gender_discrimination %>%
  # Group by gender
  group_by(___) %>%
  # Summarize to calculate proportion promoted
  summarize(prop_promoted = mean(decision == "promoted")) %>%
  # Summarize to calculate difference
  ___(stat = ___(___))

# See the result
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

What are my resources?

- Does your university have server time for hosting `learnr/shindig` 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 `shindig` tutorials, `learnr` tutorials with pre-supplied code, or RS Cloud.

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

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

Thank you!

Allison

atheobol@calpoly.edu

@mtstatistics

@atheobold

Kelly

kbodwin@calpoly.edu

@kellybodwin

@kbodwin