

Class 3: RMarkdown, Loading Data, Basic Dplyr and Loops

MGSC 310

Prof. Jonathan Hersh

Class 3: Announcements

1. Data Analytics Club Meeting
2. Data Analytics Week! October
3. Problem Set 1 posted – Due Sept 15
 - Must submit compiled HTML file using RMarkdown
4. Quiz due Thursday @ midnight
 - Will post later today

First Meeting DAA

Come listen to our guest speakers

Dr. Seth Benzell & Dave Holtz

Using Data Science to Fight COVID-19

Two researchers discuss their
studies and publications on their
analysis of COVID-19 w/ Q&A



Sept.
8

@

7:00pm
PST

Zoom Meeting ID: <https://chapman.zoom.us/j/5094919512>

Password: DAAFA2020



THE CORONAVIRUS PANDEMIC

**MIT STUDY: "CHAOTIC AND UNCOORDINATED" REOPENING
OF STATES ACROSS AMERICA TAKES A "DEVASTATING" TOLL**

 **MSNBC**



Data Analytics Industry Week

Register on Handshake to get access to the following virtual events!

Careers in Data Analytics

Tuesday, October 6 | 12 p.m. PST

Hear from the renowned authors of Build a Career in Data Science, Jacqueline Nolis and Emily Robinson about careers in data analytics.

Data Analytics Industry Panel

Thursday, October 8 | 4:30 p.m. PST

This data analytics panel will feature industry experts in analytics from entertainment, healthcare, technology, and more.

Entertainment Analytics: Turning Data Into Insights

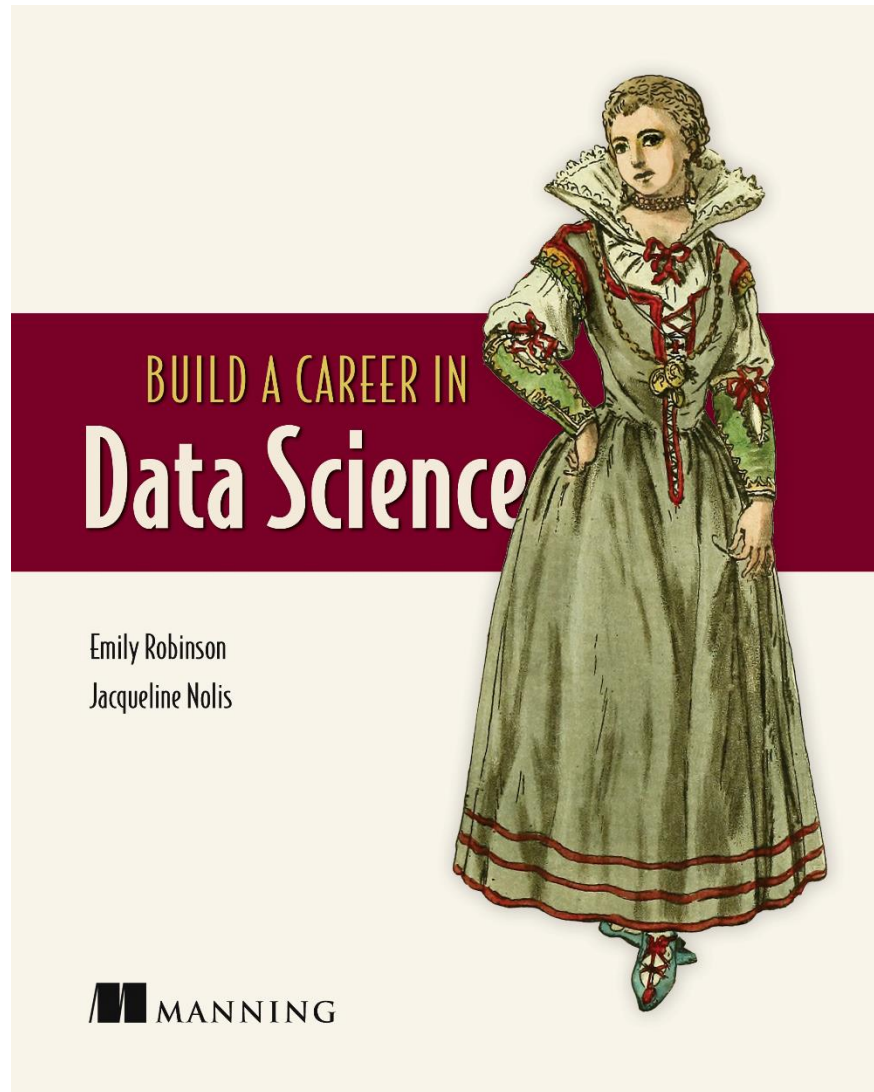
Friday, October 9 | 12 p.m. PST

Come see a live demo and learn about turning data into actionable insights in Entertainment Analytics with Andre Vargas Head of the data department at leading entertainment and sports agency, Creative Artists Agency (CAA).



**Argyros School of
Business and Economics**

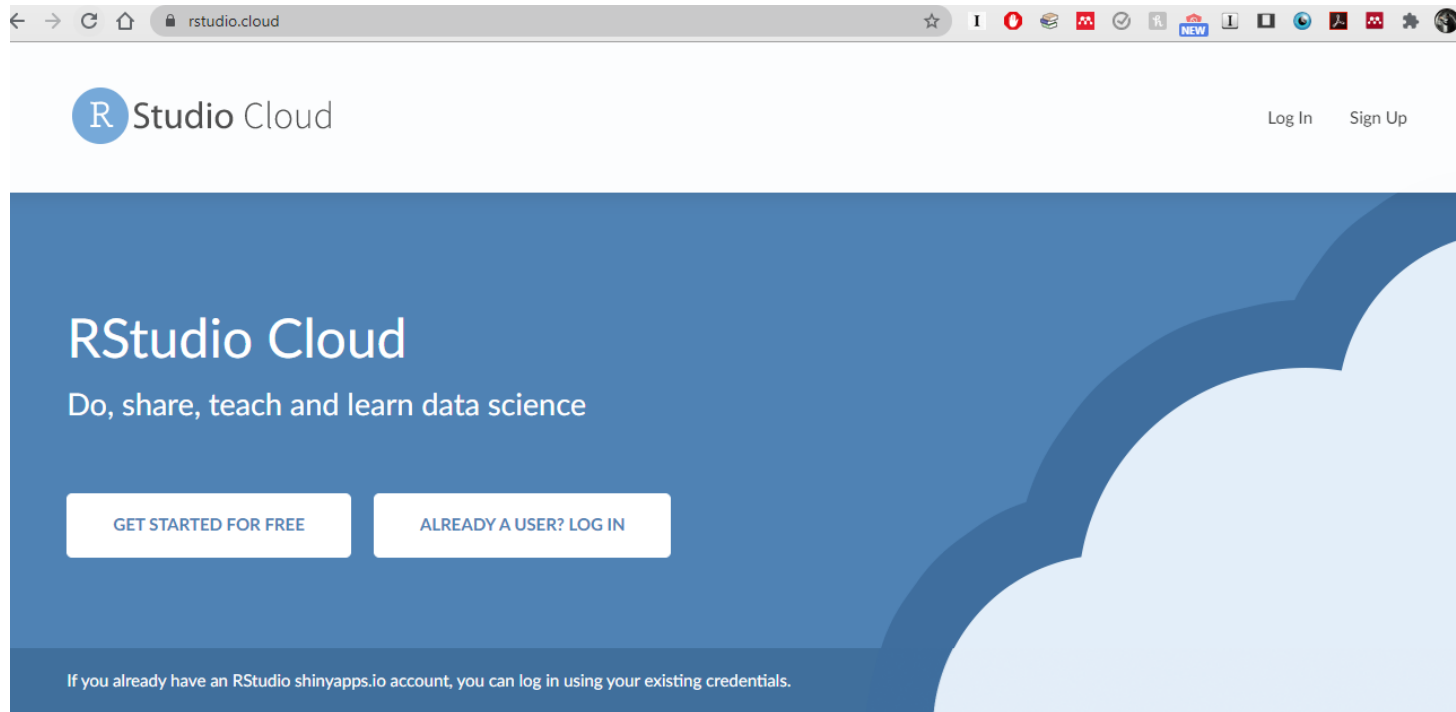
Emily Robinson and Jacqueline Nolis are Awesome



Class 3: Outline

1. Qs from last week?
2. Rmarkdown Lab
3. Loading Data
4. Basic Data Manipulation with DPLYR
 - Glimpse to view
 - Pipe operator
 - slice() to select rows
 - arrange() to order data frame
 - select() to choose variables
- filter() to select rows matching characteristics
5. Basic factors
6. Missing values
7. For loops
8. Dplyr lab!

R Studio Cloud



- Go to rstudio.cloud if your version of R is ever not working

Data science without the hardware hassles

RStudio Cloud is a lightweight, cloud-based solution that allows anyone to do, share, teach and learn data science online.

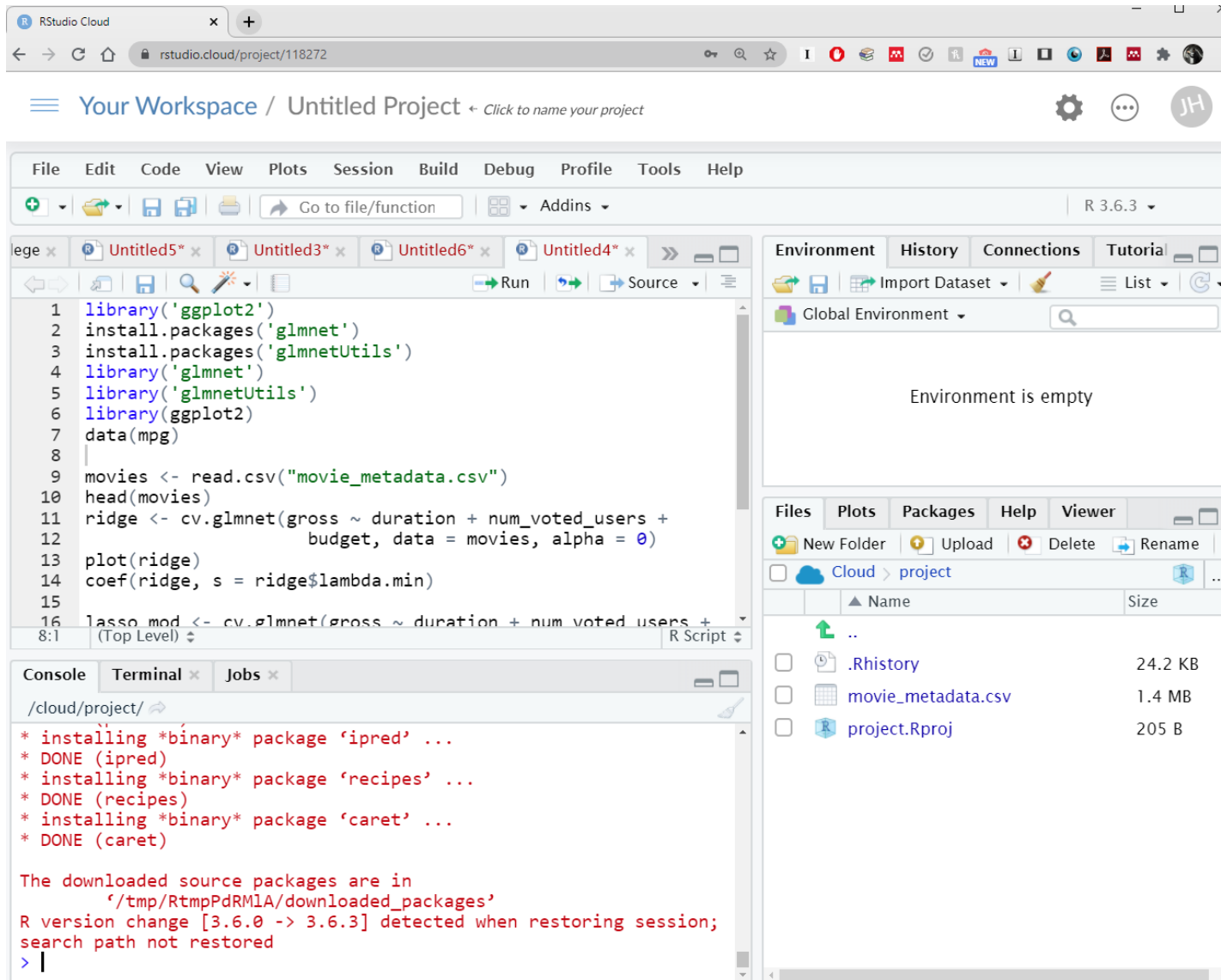
- Analyze your data using the RStudio IDE, directly from your browser.
- Share projects with your team, class, workshop or the world.
- Teach data science with R to your students or colleagues.
- Learn data science in an instructor-led environment or with interactive tutorials.

[\\$ AVAILABLE PRICING PLANS](#)

[RSTUDIO CLOUD GUIDE](#)

[RSTUDIO.COM](#)

R Studio Cloud



- R Studio Cloud is a full featured version of R in your browser!

D Plyr Lab

1. Take the movies dataset and answer the following:
 1. How many movies has George Lucas made?
 2. What are the highest grossing Steven Spielberg films?
 3. Create a dataset with only budget, gross and director
 4. How many NAs are there for each column
 5. Make 2-3 ggplots using different colors and themes