RStudio 2020 Internship Application

Riccardo Esclapon

Contents

Chapter 1

Overview

Video intro here

https://education.rstudio.com/blog/2020/02/applications-for-2020-intern-program-are-now-open/

APPLICATIONS END ON MARCH 5TH BE SURE TO APPLY BEFORE THEN!!

Chapter 2

Why me?

Here are some of the things I believe make me a great fit for the internship:

2.1 I .Rmd files

I was completely blown away by the R Markdown file format when I first discovered it, and I definitely felt a bit cheated by the fact that none of the courses I took during my undergrad in R mentioned it at all or the tidyverse. I have spent a lot of my time learning R Markdown and digging through books and amazing resources made available by RStudio, so here are some of my favorite formats that I would love to make more content around and teach people about:

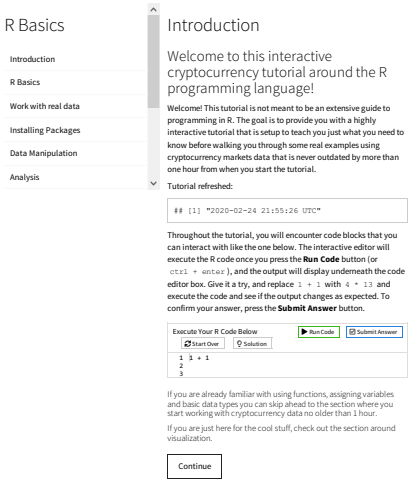
2.1.1 Learnr

I first discovered the ***learnr*** (Schloerke et al., 2019) package in late 2018 and was really impressed by the functionality it provides. My first real project using learnr was centered around teaching my young Italian cousins to program in R by allowing them to compare their Fortnite stats in real time to each other and the best players in the world, and be able to learn more about the game through working with data, for example finding the best weapon based on their damage and range. The GitHub repository associated with that project can be found here: https://github.com/ries9112/R-Tutorial

Today, I use learnr to offer tutorials on my website using learnr where every time the tutorial is opened, users learn to program in R using data from the cryptocurrency markets that is never outdated by more than 1 hour:

(this takes about 45 seconds to load, give it more time if it's blank)

```
knitr::include_app("https://predictcrypto.shinyapps.io/R_Basics/",
  height = "600px")
```


R Basics
Introduction
R Basics
Work with real data
Installing Packages
Data Manipulation
Analysis
Introduction
Welcome to this interactive cryptocurrency tutorial around the R programming language!
Welcome! This tutorial is not meant to be an extensive guide to programming in R. The goal is to provide you with a highly interactive tutorial that is setup to teach you just what you need to know before walking you through some real examples using cryptocurrency markets data that is never outdated by more than one hour from when you start the tutorial.
Tutorial refreshed:
[1] "2020-02-24 21:55:26 UTC"
Throughout the tutorial, you will encounter code blocks that you can interact with like the one below. The interactive editor will execute the R code once you press the Run Code button (or ctrl + enter), and the output will display underneath the code editor box. Give it a try, and replace 1 + 1 with 4 * 13 and execute the code and see if the output changes as expected. To confirm your answer, press the Submit Answer button.
Execute Your R Code Below
Run Code
Submit Answer
Start Over
Solution
1 1 + 1
2
3
If you are already familiar with using functions, assigning variables and basic data types you can skip ahead to the section where you start working with cryptocurrency data no older than 1 hour.
If you are just here for the cool stuff, check out the section around visualization.
Continue

I post these on my website: https://predictcrypto.org/tutorials

2.1.2 Bookdown

At one point I was very close to paying for a monthly subscription on gitbook.com because I thought it was such an amazing format to provide documentation through, so I was particularly impressed by the bookdown (Xie, 2020) package and these days it's my go to for organizing most things I work on.

This document is obviously an example of a bookdown document in itself, but here's another guide I put together using bookdown: https://predictcryptodb-quickstart.com/ <- *MAKE SURE THIS ACTUALLY REFRESHES WITH GITHUB ACTIONS BEFORE APPLYING*

I also found that documentation done in bookdown can work really great when working within a large company as well, and I put together some very thorough documentation for a project using bookdown that was very well received. In my particular case it worked really well because I could send the link to the html index of the bookdown document and when opened it would behave like a website hosted on the shared folders within the secure network which ended up being particularly simple and effective.

2.1.3 Presentations

I am a **big** fan of ioslides and revealjs in particular as R Markdown outputs. I find the revealjs output to be incredibly cool with the rotating cube animation, and the ability to not only move forward but move downward adds a surprisingly useful tool to break down topics; ioslides is just really clean, well made and easy to use and looks great with widescreen enabled.

Making presentations in R Markdown is what really got me working with .Rmd files, because I started working towards a very specific project using an idea I haven't really seen elsewhere of creating presentations that give the user options and as they make their way through the slides, those options affect not only what they see in the slides that come afterwards, but also the options they are given. For example, the user could choose to do an analysis for a particular asset, then choose the main category of the analysis to perform, then the sub-category of the analysis and so on, until by the end of the presentation the user has performed an analysis that was completely unique and tailored to their preferences and interests. See the gif below for an example of what this looks like:

FIGURE OUT WHAT IS WRONG WITH THE GIF

2.1.4 Blogdown

Blogdown(Xie, 2019) and bookdown work very similarly, so most of what I mentioned in the bookdown section applies here. Because my website predictcrypto.com only shows the latest data based on the current date, I leverage blogdown to create weekly snapshots of the visualizations over the last 7 day period: https://predictcryptoblog.com/. Because all these systems work so well with automation, as I keep adding new interesting content to my website I can also add archives of that content using blogdown.

2.2 I Automation

Automation is at the center of everything I do and my one true passion. One of my big goals for RStudio::conf 2020 was to learn more about automating things through GitHub using CI since I always had a hard time figuring that out, and the things I learned about especially relating to GitHub actions and using Netlify were above my expectations in terms of the ease of use, capabilities and free tier offerings, and I am super excited to share how crazy simple automating a very complex process can be through RStudio, GitHub Actions and Netlify.

The bookdown example from earlier https://predictcryptodb-quickstart.com/ for example uses those tools to refresh the guide daily in order to show the latest data for the *useful tables* section https://predictcryptodb-quickstart.com/useful-tables.html

It's pretty mindblowing that these frameworks allow a user to create an interactive book with complex javascript, HTML, CSS, TeX, etc… from scratch, deploy it to an https secured website and create an automated process around it, all in less than 10 minutes with minimal code involved. What's even more mindblowing, is that the same methodologies can be applied to make other interfaces, like making a blogdown website, and I can't wait to see what Yihui will bless us all with next!

2.3 I RStudio

I really wanted to go to RStudio::conf 2019 but was not able to make it out and after all the videos got posted I watched most of them and immediately knew I had to come to RStudio::conf 2020 and it was a truly incredible experience.

JJ's talk and BCorp announcement really resonated with me and there is no other company who's mission I agree with more and I would always do my very best in carrying forward those values. I fundamentally believe the most straightforward way to success is to help other people succeed, and I love the values that RStudio holds dear as a company.

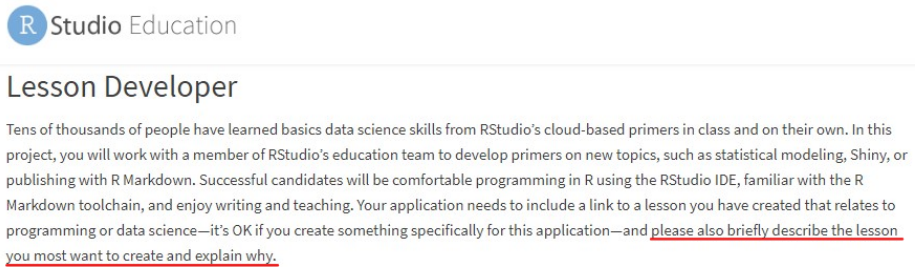
Put pictures with JJ and Hadley here

Chapter 3

About Me

Chapter 4

Ideal Tutorial



Chapter 5

Cool Charts

5.1 Disable while working on bookdown, takes too long to render!

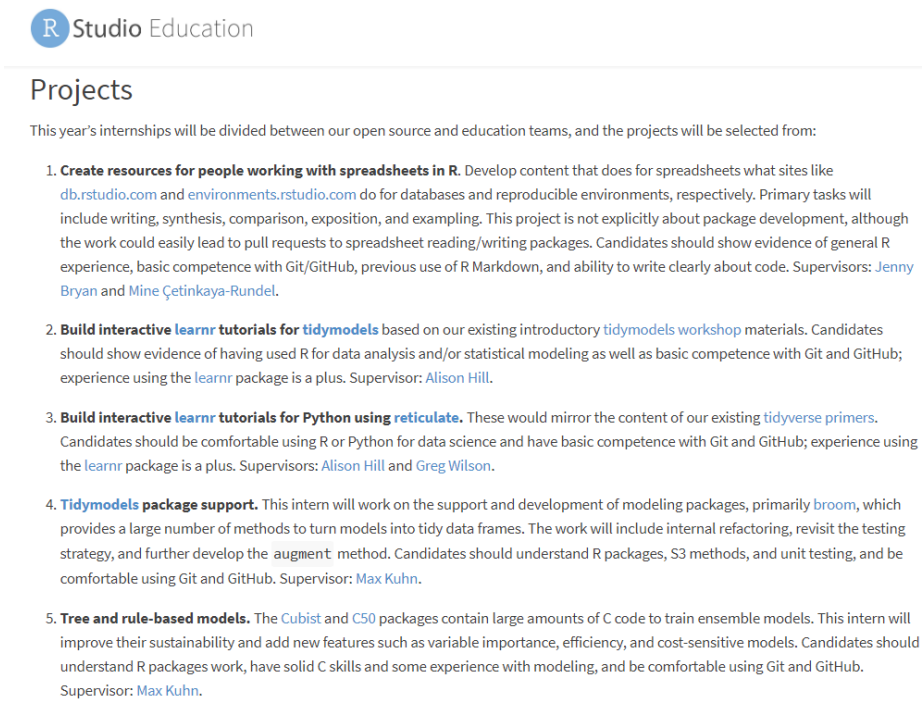
Here are some examples of charts, which refresh daily using GitHub actions and Netlify for automation.

```
## [1] TRUE
```

```
## list()
```

Chapter 6

Projects Well Suited For



RStudio Education

Projects

This year's internships will be divided between our open source and education teams, and the projects will be selected from:

1. **Create resources for people working with spreadsheets in R**. Develop content that does for spreadsheets what sites like db.rstudio.com and environments.rstudio.com do for databases and reproducible environments, respectively. Primary tasks will include writing, synthesis, comparison, exposition, and exampling. This project is not explicitly about package development, although the work could easily lead to pull requests to spreadsheet reading/writing packages. Candidates should show evidence of general R experience, basic competence with Git/GitHub, previous use of R Markdown, and ability to write clearly about code. Supervisors: Jenny Bryan and Mine Çetinkaya-Rundel.

2. **Build interactive learnr tutorials for tidymodels** based on our existing introductory tidymodels workshop materials. Candidates should show evidence of having used R for data analysis and/or statistical modeling as well as basic competence with Git and GitHub; experience using the learnr package is a plus. Supervisor: Alison Hill.

3. **Build interactive learnr tutorials for Python using reticulate.** These would mirror the content of our existing tidyverse primers. Candidates should be comfortable using R or Python for data science and have basic competence with Git and GitHub; experience using the learnr package is a plus. Supervisors: Alison Hill and Greg Wilson.

4. **Tidymodels package support.** This intern will work on the support and development of modeling packages, primarily broom, which provides a large number of methods to turn models into tidy data frames. The work will include internal refactoring, revisit the testing strategy, and further develop the `augment` method. Candidates should understand R packages, S3 methods, and unit testing, and be comfortable using Git and GitHub. Supervisor: Max Kuhn.

5. **Tree and rule-based models.** The Cubist and C50 packages contain large amounts of C code to train ensemble models. This intern will improve their sustainability and add new features such as variable importance, efficiency, and cost-sensitive models. Candidates should understand R packages work, have solid C skills and some experience with modeling, and be comfortable using Git and GitHub. Supervisor: Max Kuhn.

6.1 Create resources for people working with spreadsheets in R

Here make a guide using github environments making a repo. Maybe a learnr tutorial?

Also put some code here:

```
library(googlesheets4)
practice_sheet <- read_sheet("https://docs.google.com/spreadsheets/d/1_zRBFrB1au7qhxuDl
practice_sheet
```

```
## # A tibble: 2,421 x 17
##    Id    Name  Symbol  Rank `Price Usd` `Price Btc` `24h Volume Usd`
##    <chr> <chr> <chr>  <dbl>       <dbl>       <dbl>            <dbl>
##  1 bitc~ Bitc~ btc        1    9769.      1             40463828067
##  2 ethe~ Ethe~ eth        2     269.      0.0275        19542499911
##  3 xrp   XRP   xrp        3       0.274   0.0000280      2377865581
##  4 bitc~ Bitc~ bch        4     383.      0.0392         4421385407
##  5 bitc~ Bitc~ bsv        5     284.      0.0291         2228030328
##  6 lite~ Lite~ ltc        6      75.5     0.00773        5638873067
##  7 teth~ Teth~ usdt       7       1.00    0.000102      47741567290
##  8 eos   EOS   eos        8       4.14    0.000424       3697914797
##  9 bina~ Bina~ bnb        9      22.6     0.00231         409613428
## 10 tezos Tezos xtz       10       3.19    0.000327        196594975.
## # ... with 2,411 more rows, and 10 more variables: `Market Cap Usd` <dbl>,
## #   `Circulating Supply` <dbl>, `Total Supply` <dbl>, `Max Supply` <dbl>,
## #   `Percent Change 1h` <dbl>, `Percent Change 24h` <dbl>, `Percent Change
## #   7d` <dbl>, last_updated <dttm>, `24h Volume Btc` <dbl>, `Market Cap
## #   Btc` <dbl>
```

```
library(data.table)
data.table(practice_sheet)
```

```
##                                      Id                          Name Symbol
##    1:                           bitcoin                       Bitcoin    btc
##    2:                          ethereum                      Ethereum    eth
##    3:                               xrp                           XRP    xrp
##    4:                      bitcoin-cash                  Bitcoin Cash    bch
##    5:                        bitcoin-sv                    Bitcoin SV    bsv
##   ---
## 2417:                            ftoken                        FToken     ft
## 2418:                          eosblack                      eosBLACK  black
## 2419: airline-and-life-networking-token Airline & Life Networking Token   alln
```

```
## 2420:                          harcomia                          Harcomia    hca
## 2421:                             blocs                             Blocs  blocs
##       Rank     Price Usd    Price Btc 24h Volume Usd Market Cap Usd
##    1:    1 9769.48991700 1.0000000000    40463828067   178146277391
##    2:    2  269.05298170 0.0275401258    19542499911    29551156294
##    3:    3    0.27378778 0.0000280248     2377865581    11978054912
##    4:    4  382.85920450 0.0391892727     4421385407     7004887720
##    5:    5  284.46044790 0.0291172262     2228030328     5203799131
##   ---
## 2417: 2417    0.04225120 0.0000043248             NA             NA
## 2418: 2418    0.01052418 0.0000010773             NA             NA
## 2419: 2419    0.02015197 0.0000020627             NA             NA
## 2420: 2420            NA           NA             NA             NA
## 2421: 2421   55.94342404 0.0057263403             NA             NA
##       Circulating Supply Total Supply   Max Supply Percent Change 1h
##    1:           18234962     18234962     21000000             -0.02
##    2:          109833967    109833967           NA             -0.29
##    3:        43749413421  99991077044 100000000000             -0.11
##    4:           18296250     18296250     21000000             -1.36
##    5:           18293577     18293577     21000000             -0.80
##   ---
## 2417:                 NA   2510925464           NA              0.00
## 2418:                 NA   3000000000           NA              0.00
## 2419:                 NA   3500000000           NA              0.00
## 2420:                 NA       122000           NA                NA
## 2421:                 NA     21000000           NA              0.00
##       Percent Change 24h Percent Change 7d        last_updated 24h Volume Btc
##    1:              -0.99              1.19 2020-02-24 15:07:35      4141856.8
##    2:              -0.62              6.30 2020-02-24 15:07:35      2000360.3
##    3:              -2.63             -2.21 2020-02-24 15:07:35       243397.1
##    4:              -3.06             -2.63 2020-02-24 15:07:35       452570.8
##    5:              -2.18             -1.17 2020-02-24 15:07:35       228060.0
##   ---
## 2417:               0.00              0.00 2020-02-24 15:07:35             NA
## 2418:               0.00              0.00 2020-02-24 15:07:35             NA
## 2419:               0.00              0.00 2020-02-24 15:07:35             NA
## 2420:                 NA                NA 2020-02-24 15:07:35             NA
## 2421:               0.00              0.00 2020-02-24 15:07:35             NA
##       Market Cap Btc
##    1:     18234962.0
##    2:      3024841.3
##    3:      1226067.6
##    4:       717016.7
##    5:       532658.2
##   ---
## 2417:             NA
```

```
## 2418:            NA
## 2419:            NA
## 2420:            NA
## 2421:            NA
```

COULD MAKE THIS FIRST SECTION FROM GOOGLE SHEETS USING DATA ACTUALLY PREDICTING % CHANGE AND WHATNOT IF I LOAD THAT OTHER DATA INTO HERE. COULD THEN WORK ON NEXT SECTION WHILE MAKING PROGRESS ON BOTH INTERNSHIP AND RESEARCH PAPER!

6.2 Build interactive learnr tutorials for tidymodels

https://education.rstudio.com/blog/2020/02/conf20-intro-ml/

https://conf20-intro-ml.netlify.com/materials/01-predicting/

6.3 Build interactive learnr tutorials for Python using reticulate

Replace this with the Python one:

Could make a very simple xgboost model maybe?

Bibliography

Schloerke, B., Allaire, J., and Borges, B. (2019). learnr: Interactive tutorials for r. R package version 0.10.0.

Xie, Y. (2019). blogdown: Create blogs and websites with r markdown. R package version 0.17.

Xie, Y. (2020). *bookdown: Authoring Books and Technical Documents with R Markdown*. R package version 0.17.