

RStudio 2020 Internship Application

Riccardo Esclapon

Contents

1	Overview	5
2	What makes me a good fit	7
2.1	I .Rmd files	7
2.2	I Automation	9
2.3	I RStudio	10
3	Projects Well Suited For	13
3.1	Create resources for people working with spreadsheets in R . . .	14
3.2	Build interactive learnr tutorials for tidyverse	17
3.3	Build interactive learnr tutorials for Python using reticulate . . .	17
4	About Me	19
5	Ideal Tutorial	21
6	Cool Charts	23
6.1	Disable while working on bookdown, takes too long to render! . . .	23

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!!

For video:

Start off with overview of projects I am suited for showing work I did for this application specifically. Then go on to talk about ways I have applied the broad RMarkdown ecosystem and automation in my work. Then talk a bit more about myself. Talk about ideal tutorial overview and close things by mentioning cool charts/visualizations section (outline this at a high level under 2 minutes in the video at the start here)

Chapter 2

What makes me a good fit

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> (the apps themselves are down but the repo has some gifs illustrating the past functionality)

More recently, I have been using 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 showing up blank)

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 and grateful for the bookdown (Xie, 2020) package, and these days it's my go to for organizing most things I work on, so why not my application?

This document is obviously an example of a bookdown document in itself, but here's another guide I put together using bookdown:

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 (but I can't show here). 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. I aspire to be an expert in Xaringan one day but am not currently.

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:

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.

2.1.5 Pagedown

Pagedown(Xie et al., 2020) is yet another awesome way to create html outputs and I used Nick Strayer's repository <https://github.com/nstrayer/cv> to build my cv and resume using his template:

2.1.6 Flexdashboard

Flexdashboards(Iannone et al., 2018) were my first introduction to shiny apps and I was completely blown away by that framework and have used it for several projects and is one of my absolute favorite tools.

For my first flexdashboard, I converted some of the content found in Tidy Text Mining by Julia Silge and David Robinson and made it into a flexdashboard. I made no changes to the code found within the book, this was simply an experiment to learn more about flexdashboards and semantic analysis:

I made the code available through RStudio Cloud here as well: https://rstudio.cloud/spaces/9369/join?access_code=pkfhGuOMRhleNIHSHH6YOQPEWstEdg0e7Pi6Ue3q

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 in the *useful tables* section

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 speak highly enough of all the work Yihui blessed us all with.

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.

I tried really hard to not be too much of a fanboy at the conference, but I couldn't help but get a picture with JJ and Hadley.





Would have loved one with Yihui as well, so I will try and get one with him at next year's conference if I get the chance

Chapter 3

Projects Well Suited For



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 exemplifying. 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](#).

3.1 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_zRBFrB1au7qhhuD")
knitr::kable(head(practice_sheet, 30))
```

date	Symbol	AdrActCnt	BlkCnt	BlkSizeByte	BlkSizeMeanByte	CapMVRVCu
2020-02-24	BTC	758122	141	161000000	1141214.00	1.657654
2020-02-24	ETH	320273	6481	153000000	23637.33	29000000000.00000
2020-02-23	BTC	622397	142	131000000	924118.80	1.713841
2020-02-23	ETH	311622	6532	130000000	19886.52	30200000000.00000
2020-02-22	BTC	660812	148	136000000	919925.90	1.663798
2020-02-22	ETH	314925	6528	131000000	20003.89	28700000000.00000
2020-02-21	BTC	768521	154	164000000	1065423.00	1.668813
2020-02-21	ETH	323587	6475	144000000	22264.03	29200000000.00000
2020-02-20	BTC	766448	126	158000000	1254015.00	1.654352
2020-02-20	ETH	317504	6480	152000000	23416.70	28400000000.00000
2020-02-19	BTC	792783	141	164000000	1164743.00	1.657941
2020-02-19	ETH	333141	6557	151000000	23009.82	28700000000.00000
2020-02-18	BTC	783570	150	169000000	1129523.00	1.754254
2020-02-18	ETH	344572	6480	157000000	24219.29	31100000000.00000
2020-02-17	BTC	773751	157	160000000	1021988.00	1.670618
2020-02-17	ETH	333123	6447	156000000	24137.83	29300000000.00000
2020-02-16	BTC	655188	142	135000000	951565.30	1.716954
2020-02-16	ETH	337792	6552	145000000	22162.91	28700000000.00000
2020-02-15	BTC	736640	142	154000000	1081882.00	1.710757
2020-02-15	ETH	343999	6472	149000000	23067.16	29100000000.00000
2020-02-14	BTC	741471	135	149000000	1103415.00	1.788310
2020-02-14	ETH	323942	6493	169000000	26040.92	31200000000.00000
2020-02-13	BTC	815808	141	171000000	1213825.00	1.768516
2020-02-13	ETH	374630	6481	166000000	25626.90	29400000000.00000
2020-02-12	BTC	822547	142	165000000	1159784.00	1.792084
2020-02-12	ETH	394534	6463	171000000	26520.93	29200000000.00000
2020-02-11	BTC	885934	148	185000000	1251850.00	1.779233
2020-02-11	ETH	319771	6534	151000000	23172.66	26100000000.00000
2020-02-10	BTC	703356	122	146000000	1198981.00	1.713470
2020-02-10	ETH	318635	6592	143000000	21715.29	24600000000.00000

3.1. CREATE RESOURCES FOR PEOPLE WORKING WITH SPREADSHEETS IN R15

This data is sourced from the website coinmetrics.io

They also provide a data dictionary to go along with the data, let's also read that in with `googlesheets4`:

```
data_dictionary <- read_sheet("https://docs.google.com/spreadsheets/d/1\_zRBFrB1au7qhxuDDfDuh\_bPLC")
```

```
knitr::kable(data_dictionary)
```

id	name	category	sub
AdrActCnt	Addresses, active, count	Addresses	Ac
BlkCnt	Block, count	Blockchain / ledger	Bl
BlkSizeByte	Block, size, bytes	Blockchain / ledger	Bl
BlkSizeMeanByte	Block, size, mean, bytes	Blockchain / ledger	Bl
CapMVRVCur	Capitalization, MVRV, current supply	Market	Ma
CapMrktCurUSD	Capitalization, market, current supply, USD	Market	Ma
CapRealUSD	Capitalization, realized, USD	Market	Ma
DiffMean	Difficulty, mean	Mining	Mi
FeeMeanNtv	Fees, transaction, mean, native units	Fees and revenue	Fee
FeeMeanUSD	Fees, transaction, mean, USD	Fees and revenue	Fee
FeeMedNtv	Fees, transaction, median, native units	Fees and revenue	Fee
FeeMedUSD	Fees, transaction, median, USD	Fees and revenue	Fee
FeeTotNtv	Fees, total, native units	Fees and revenue	Fee
FeeTotUSD	Fees, total, USD	Fees and revenue	Fee
HashRate	Hash rate, mean	Mining	Mi
IssContNtv	Issuance, continuous, native units	Supply	Iss
IssContPctAnn	Issuance, continuous, percent, annualized	Supply	Iss
IssContUSD	Issuance, continuous, USD	Supply	Iss
IssTotNtv	Issuance, total, native units	Supply	Iss
IssTotUSD	Issuance, total, USD	Supply	Iss
NVTAdj	NVT, adjusted	Valuation	Va
NVTAdj90	NVT, adjusted, 90d MA	Valuation	Va
PriceBTC	Price, BTC	Market	Pri
PriceUSD	Price, USD	Market	Pri
ROI1yr	ROI, percent, 1yr	Market	Re
ROI30d	ROI, percent, 30d	Market	Re
SplyCur	Supply, current	Supply	Cu
TxCnt	Transactions, count	Transactions	Tr
TxTfrCnt	Transactions, transfers, count	Transactions	Tr
TxTfrValAdjNtv	Transactions, transfers, value, adjusted, native units	Transactions	Tr
TxTfrValAdjUSD	Transactions, transfers, value, adjusted, USD	Transactions	Tr
TxTfrValMeanNtv	Transactions, transfers, value, mean, native units	Transactions	Tr
TxTfrValMeanUSD	Transactions, transfers, value, mean, USD	Transactions	Tr
TxTfrValMedNtv	Transactions, transfers, value, median, native units	Transactions	Tr
TxTfrValMedUSD	Transactions, transfers, value, median, USD	Transactions	Tr
TxTfrValNtv	Transactions, transfers, value, native units	Transactions	Tr
TxTfrValUSD	Transactions, transfers, value, USD	Transactions	Tr
VtyDayRet180d	Volatility, daily returns, 180d	Market	Re
VtyDayRet30d	Volatility, daily returns, 30d	Market	Re
VtyDayRet60d	Volatility, daily returns, 60d	Market	Re

I am comfortable writing packages in R as well as using testthat and showing code coverage for a repository. I attended the building tidy tools workshop working with Charlotte and Hadley at RStudio::conf

2020.

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!

3.2 Build interactive learnr tutorials for tidy-models

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

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

Create parsnip model

```
library(parsnip)
linear_reg() %>%
  set_engine("glmnet") %>%
  set_mode("regression")

## Linear Regression Model Specification (regression)
##
## Computational engine: glmnet

# List of models to refer to: https://tidymodels.github.io/parsnip/articles/articles/Models.html

xgboost_parsnip <- boost_tree() %>%
  set_engine("xgboost") %>%
  set_mode("regression")
```

3.3 Build interactive learnr tutorials for Python using reticulate

Replace this with the Python one:

Could make a very simple xgboost model maybe?

Could also show using Shrimpy API to pull latest data, manipulate in pandas and visualize

Mention experience/courses taken in Python and how it's never clicked with me very much but how I am taking a basic Python course in my Master's in Data

Science and I am looking to take it as an opportunity to create a lot of content using reticulate.

```
library(reticulate)
```

Chapter 4

About Me

Chapter 5

Ideal Tutorial



Lesson Developer

Tens of thousands of people have learned basics data science skills from RStudio's cloud-based primers in class and on their own. In this project, you will work with a member of RStudio's education team to develop primers on new topics, such as statistical modeling, Shiny, or publishing with R Markdown. Successful candidates will be comfortable programming in R using the RStudio IDE, familiar with the R Markdown toolchain, and enjoy writing and teaching. Your application needs to include a link to a lesson you have created that relates to programming or data science—it's OK if you create something specifically for this application—and please also briefly describe the lesson you most want to create and explain why.

Chapter 6

Cool Charts

6.1 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
```


Bibliography

Iannone, R., Allaire, J., and Borges, B. (2018). *flexdashboard*: R markdown format for flexible dashboards. R package version 0.5.1.1.

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.

Xie, Y., Lesur, R., and Thorne, B. (2020). *pagedown*: Paginate the html output of r markdown with css for print. R package version 0.8.