R Packages

DS 303

1/26/2022

Contents

Preface

These notes were assembled to accompany a Data Science 303 course at Calvin University.

1 Ethics for Data Science

1.1 Ethical Issues in Data Science

1.1.1 Introduction

Sara Baase, in her classic textbook *A Gift of Fire*, defines ethics as "the study of what it means to do the right thing" [1]. Michael Quinn [2] further specifies that this "right thing" relates to various situations within a particular society. Here we include a list of broad categories of these various situations and call them ethical issues. We present our working definitions of these ethical issues in the context of the computer and data sciences. Each was created to be a one sentence definition that captures the spirit of what we've seen throughout a wide range of texts and conversations.

1.1.2 What is an ethical dilemma?

An ethical dilemma may occur whenever there is a **conflict between two or more ethical issues** or aspects of the same issue. For example, if your company employs cheap labor overseas they may profit economically, but lack in dignifying and promoting physical/mental health of the workers. As a result of these conflicting issues, the company is forced to make a decision: Should they continue to hire international workers for poor wages to keep a high profit, or should they invest more in their employees at the expense of losing revenue.

1.1.3 Some ethical issues in Data Science

Media Literacy How do we tell whether an article, video, etc. is factual, reliable, and unbiased?

Accessibility Who can use the products and analyses we produce? Does this include people with economic, physical, mental, and emotional disadvantages?

Data Visualization Do our visualizations represent data in a clear, accessible way that accurately reflects the meaning of underlying data or model?

Transparency Do we represent data, algorithms, or any information that affects a stakeholder in an open, explicit, and easily comprehensible format?

Diversity Are a variety of people (race and ethnicity, age, gender, skills and education level, sexual orientation, religion, etc.) contributing to the development of an idea or product in an effort to avoid bias and increase the breadth of insight into product needs?

Sustainability Do we employ practices that protect the environment, including living creatures and other naturally occurring resources and features?

Intellectual Property Who owns the work? Who controls development, dissemination, and pricing? Are all those who worked on the project fairly acknowledged and compensated? Are non-owners also treated fairly?

Decision Making for AI Who is responsible for how an artificially intelligent machine makes (ethical) decisions, especially those involving life-altering actions?

Freedom of Speech What are the rights and limitations of the public to express themselves and correspond through electronic communication? Who enforces the rights and limitations?

Dignity How do we value human and animal life, well-being, and connection relative to other goals and desires?

Physical and Mental Health How do we proactively take care to protect the health of employees (including oneself) and consumers of one's product?

Professional Citizenship How do we use our professional expertice wisely both among others "in the guild" and with those outside it? What do we say? How must we listen? What guidelines does our profession have for this? Should we follow them?

Reliability What is our obligation to make reliable products (and to measure and assess that reliability)?

Hospitality What is our obligation to make our work easily understandable and usable by peers in a similar field? To those with different expertise and experience?

Privacy and Security How do we balancing the need for protecting personal and group data with the use of that data for the welfare of the public? Who "owns" data about people? How may it be used?

Algorithmic Bias Might our algorithms represent bias (intential or unintential) arising from ignorance, assumptions, or past societal patterns? What do we do about that?

Data Integrity How does they way we handle data affect potential for bias or false results?

Professional Ethics What are our obligations in the workplace toward our co-workers, supervisors, subordinates, the company, shareholders, and stakeholders? What do we do when these are in conflict?

1.2 Simulating Loan Thresholds

1.2.1 A simplified situation

Imagine a machine learning algorithm that takes information about loan applicants and assigns them a credit score from 0 to 100. The credit scores are not perfect, however. Applicants with higher credit scores are more likely to repay their loan, but some people with lower credit scores will repay loans, and some with higher scores will not.

Your task is to devise a system that uses these credit scores to decide who gets a loan.

1. How will you measure how "good" your system is? What sorts of things could you measure/quantify as indication of how well your system works? We'll call those things metrics. Come up with at least 3 metrics.
2. Which of your metrics do you like better? Why?

Now let's experiment a bit with some simulated data. Go to the "Simulating Loan Thresholds" page at http://research.google.com/bigpicture/attacking-discrimination-in-ml/. The graphs and figures provide metrics, and you can slide the credit score threshold and see how these metrics change. It is based on issues raised Hardt et al. (2016).

3. What metrics are included on this page? How many of them are the same (or equivalent to)[1] the ones you came up with?
4. What credit score threshold maximizes the number of correct decisions? Does that same threshold also minimize the number of incorrect decisions?
5. What threshold maximizes profit?

   Note: The app assumes a marginal gain of $300 for each loan repaid and a loss of $700 for each loan default. This is mentioned later in the text, but you can also figure it from the interactive graphic. How?
6. Are your answers to #4 and #5 the same? Why or why not?
7. The app doesn't let you experiment with this, but suppose repaid loans had a marginal benefit of $400 and defaulted loans a marginal cost of only $600.

   a. Would that lead a profit-maximizer to increase or decrease the threshold? Why?

   b. How would the threshold change if you are maximizing correct decisions.

1.2.2 Different groups

Don't click the red buttons until directed. Instead, just drag the threshold sliders.

8. The text says: "In this case, the distributions of the two groups are slightly different, even though blue and orange people are equally likely to pay off a loan."

   a. In what way are the distributions (of credit scores) different for blue people and orange people? (Remember your statistics? You might think about things like center, spread, shape, etc. But remeber that we are comparing distributions of *two* variables: credit score and whether or not they repay the loan.)

   b. How might you improve the graphical display to make it easier to see how the two groups differ?

   c. Does the credit score do a better job for one group than for the other? What do you mean by better?

[1] For our purposes, two metrics are equivalent if they rank any two things the same – that is, they always agree about which option is the better one.

9. Adjust the thresholds to maximize the profit that the bank gets for each group individually. (a) What are the resulting thresholds for the two groups?

(b) At those thresholds, which group is offered more loans? (c) Of the loans that were offered, which group repays more of those loans? (d) What is the resulting profit for the bank? (e) Is that fair? (explain.) (Check your answer by clicking on the red "Max Profit" button.)

10. Adjust the thresholds to maximize profit under the constraint that blue and orange must have the same threshold. (This is tricky in the UI; hint: it must be between the two thresholds you found in the previous exercise.) Answer the same five questions (a-e) as in the previous exercise. Then check your answer by clicking the "Group Unaware" button.

11. Starting at the "Group Unaware" thresholds, change the threshold on the orange group to increase its Positive Rate to match the Positive Rate of the blue group. (a) What happens to the True Positive Rate? (b) What happens to the bank's profit? (c) Is that good? Explain.

12. Explore the "Demographic Parity" and "Equal Opportunity" objectives by clicking their buttons. The authors of the interactive visualization seem to have a favorite among these objectives. What reasons might someone have to agree or disagree with their assessment?

1.2.3 Thinking about your thinking

There are a lot of value-laden words in the questions above: words like good, better, fair, etc. Compare two ways of awarding loans (from the ones above or from other way you come up with), one of which you think is "better" than the other. This will be more interesting if you choose two that are close rather than one that you think is great and one that you think is terrible – make the decision difficult.

13. What makes one better than the other? Imagine you had to defend the choice, what would be the main components of your argument?

14. Are there general principles that would apply to other situations that have nothing to do with loans and credit scores?

15. If you could wish for any feature(s) to be added to this simulation, what would it/they be?

1.2.4 The article

This simple simulation is a by-product of a paper: Hardt et al. (2016).

Here is the abstract:

> We propose a criterion for discrimination against a specified sensitive attribute in supervised learning, where the goal is to predict some target based on available features. Assuming data about the predictor, target, and membership in the protected group are available, we show how to optimally adjust any learned predictor so as to remove discrimination according to our definition. Our framework also improves incentives by shifting the cost of poor classification from disadvantaged groups to the decision maker, who can respond by improving the classification accuracy. We enourage readers to consult the more complete manuscript on the arXiv.

16. For each of the following words/phrases, say what the correspond to in the simulation we just looked at.

    a. specified sensitive attribute
    b. some target
    c. available features
    d. predictor
    e. membership in the protected group
    f. adjust (any) learned predictor

1.3 Ethical Frameworks and Tests

When making ethical decisions, what do we base them upon? Below are brief descriptions of four frameworks for making ethical decisions. As you have ethical discussions, see if you can identify when you and your conversation partners are using these (or other) frameworks. Keep in mind that it is possilbe for more than one framework to be invovled at the same time. When this happens, see if you can identify which framework has the stronger influence or how the different frameworks are shaping each other.

Bottom line for all frameworks: How would you honestly and rationally defend your decision?

The descriptions below are based on draft materials for ethics in data science being developed by authors at Point Loma Nazarene University.

Some of the test questions were derived from material on the following websites (accessed 6/2019):

- https://blink.ucsd.edu/finance/accountability/ethics/ask.html
- http://www.refreshleadership.com/index.php/2014/01/6-questions-confronted-ethical-issues/
- https://www.ncoss.org.au/sites/default/files/public/resources/A%20Guide%20to%20Ethical%20Decision%20Making.pdf

1.3.1 Utilitarian Ethics

Basic idea: The utilitarian framework determines the best solution by choosing the course of action which seems the best for the most people affected by the solution (for the most stakeholders).

Ethical tests based on the framework:

- Who will be affected by this decision? (who are the stakeholders?)
- Who benefits from this decision?
- Who will be harmed by this decision?
- Do the benefits outweigh the harms?
- What are all the long-term and short-term consequences?
- Does this option do less harm than the alternatives?

Example: There are five ill patients in a hospital who desperately need organ transplants; however, there are no available organs. By chance, a healthy traveler comes in for a check-up and happens to be a match for each of the five patients. Assuming that there would be no legal repercussions, the doctor could kill the traveler and use his organs to save the 5 lives. The stakeholders would be each of the patients and their families. From a utilitarian perspective, the right thing to do could be to kill the one healthy person in order to save the five, because this option brings about the greatest utility (or happiness) despite causing instrumental harm.

1.3.2 Virtue Ethics

Basic idea: The virtue ethics framework determines the best solution by choosing the solution that is most in agreement with our social virtues (moral values such as honesty, compassion, diversity.)

Ethical tests based on the framework:

- Does this decision agree or disagree with my moral values?
- Are any of my virtues in conflict with each other because of this dilemma?
- Think of someone whose moral judgement you respect, would this decision line up with his or her values?
- Would a child or your parental figure be proud of your decision?
- Is this decision in line with my company's core values?

Example: If you are unsure of whether or not it is right to punish a poor person for stealing food, your virtue of compassion may say that it isn't correct while your virtue of justice may say that it is.

1.3.3 Ethical Analogies

Basic idea: The analogy approach compares two completely different things that might have similarities in concepts. Generally the concept in the analogy is better understood, and might help in understanding the concept in the other and determining its ethicality.

Ethical tests based on the framework:

- Can I think of a simpler scenario to compare this situation to?
- Is this situation similar to anything I have seen before?
- Can I use these similar scenarios to help clarify or make decisions about the current dilemma?

Example: If you are unsure if it is ethical to hack someone without stealing/manipulating any of their information, you may compare it to breaking into someone's home and only looking around.

1.3.4 Deontologcal Ethics

Basic idea: The deontological framework is built on a set of rules or principles. While more rigid, it can sometimes clarify the gray areas and subjectivity left by the other frameworks.

Ethical tests based on the framework:

- Does this decision break any laws?
- Would I be going against any codes of ethics?
- What would my profession's governing body think of this decision?
- Does my decision go against any implied, unspoken moral laws?
- If I am setting a standard, would I want it to be followed by other people (including my competitors)?

Example: If you are unsure of whether or not it is ethical to murder someone, you can refer to the US law which says that it is wrong.

1.3.5 Identifying Frameworkds

1. Return to your discussion of using credit scores to decide who gets a loan. Identify places where the four ethical frameworks were used to evaluate the situation.

2. The ASA (American Statistical Association) and ACM (Association for Computing Machinery) have each published ethical guidelines for their professions.

- ACM: poster, full version
- ASA

Identify the ethical framework(s) being used in these statements.

1.3.6 Additional Resources

- A Framework for Making Ethical Decisions (STS @ Brown)

- *Artificial Morality*, an essay by Bruce Sterling

1.4 Values Survey

1.4.1 What is a value?

> "A value is something you are willing to act on" – Ayn Rand, philosopher
>
> "Values are a matter of what guides you through every day, every task, every encounter with another human being." – Richard Bolles, *What Color is Your Parachute?*
>
> "a preference that is felt and/or considered to be justified – morally, by reasoning, or by aesthetic judgments, usually by two or all three of these" – Sarah Spiekermann, *Ethical IT*

1. Write the definition of "value" in your own words. (You might find it handy to look at the list of values below to see how well your definition works compared to that list.)

1.4.2 Most Important Values

Circling the values in the list below that are most important to you. Try to choose 20-25 top values. Think a value is missing from the list? Feel free to add additional values.

Achievement
Advancement
Adventure
Appreciation
Attentiveness
Authenticity
Autonomy
Balance
Beauty
Beingthebest
Belonging
Challenge
Compassion
Competition
Consistency
Contentment
Creativity
Decisiveness
Dignity
Diversity
Education
Entrepreneurship
EnvironmentalProtection
Equality
Fame
Family
Financialsecurity
Flexibility
Freedom
Fun
Generosity
Gratitude
Growth
Happiness
Harmony
Health
Honesty
Hospitality
Humility
Humor
Inclusivity
Individuality
Independence
Influence
Innovation
Integrity
Intelligence
Kindness
Leadership
Leisure
Listening
Love
Loyalty
Mastery
Mercy
Open-mindedness
Peace
Physical activity
Power
Privacy
Problemsolving
Professionalism
Punctuality
Quality
Recognition
Relationships
Reliability
Respect
Risk-taking
Safety
Security
Self-expression
Service
Simplicity
Spirituality
Stability
Status
Success
Teamwork
Tenacity
Transparency
Trustworthiness
Usefulness
Vision
Wealth

3. After you have circled your 20-25 top values, put a W next to the circled values that might inform where you would choose to work when you graduate.
4. When everyone in your group has finished, each person should choose two of the values that they have put W's next to and describe to the group how that value would influence a job choice for you.

5. How are values related to ethical frameworks? Look over your circled values, and especially the two you selected in the previous question.

    a. If you were asked *why* you value those things, how would explain? Does your explanation sound like it is coming from one of our ethical frameworks (utilitarianisms/consequentialism, deontology, or virtue ethics)?

    b. Does your *set* of values lean toward one of these more than the others?

6. Compare your answers to the previous item with others in your group. Are there notable similarities? notable differences?

    In a workplace setting which do you think matters more: *which* values are most important or *why* those values are most important?

2 Creating R packages

2.1 Resources

- R Packages by Hadley Wickham and Jenny Bryan

2.2 Creating your first package

Following a similar (but not identical) path to chapter 2 of R Packages by Hadley Wickham and Jenny Bryan, we will build a small package to work with regular expressions. There already are packages that do this sort of thing, so this is just for demonstration purposes.

2.2.1 Step 1: Create a GitHub repo

This isn't a requirement for building a package, and it doesn't have to be the first step. But there are many advantages to using github and doing this step first.

As you create your repo:

2.2.1.1 Give it the same name as your package.

- This isn't required, but makes life simpler.
- You may name it whatever you like. If you want to match what is in the example by Hadley and Jenny, name it `regexcite`.
- Note: package names should consist entirely of letters. (Dots are allowed, but please don't use them in package names.) Unless there is a good reason to use capitals, all lower case is usually preferred.
- There is actually a function in R to help you figure out if you have chosen a good name for your package. See `available::available()`.

2.2.1.2 Choose to add a `.gitignore` file and select R as the language.

- This will tell git not to track certain files. But the main reason for this is so there is a file in the repo.

2.2.1.3 Optional stuff

- Optionally, add a README file
  - We will be creating a different README file in a moment.
- Optionally, add a license
  - Again, we will see a different way to do this in a moment.

2.2.1.4 Add collaborators If you are working together with others, you can add them as collaborators to your project. This let's them make changes to your github repository. Alternatively, you could require people to fork your project and issue pull requests to make changes.

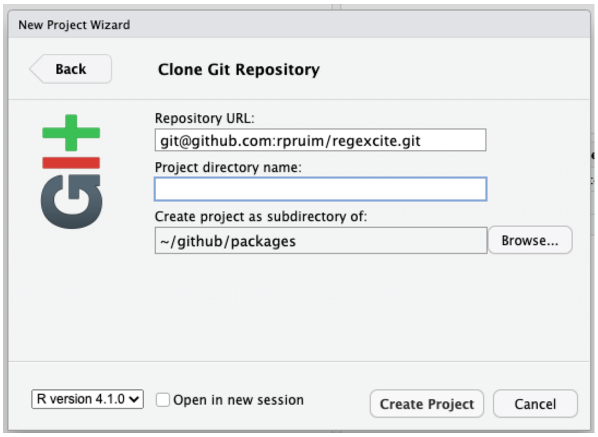
2.2.2 Step 2: Connect R and GitHub

Note: If you are working on your own machine, you will need some things that may or may not be installed.

- Xcode Tools (mac) or RTools (PC)
- git
- some R packages (you can install those as we go along)

Assuming those are set up and ready to go, or you are working on https://rstudio.calvin.edu, you should be ready to create an R project for your package.

1. In GitHub, copy the GitHub SSH URL from the **Code** pill (green).
   - It should look something like `git@github.com:rpruim/regexcite.git`
2. In R, create a **new project**
   - Choose **Version Control > Git** and fill out this dialog

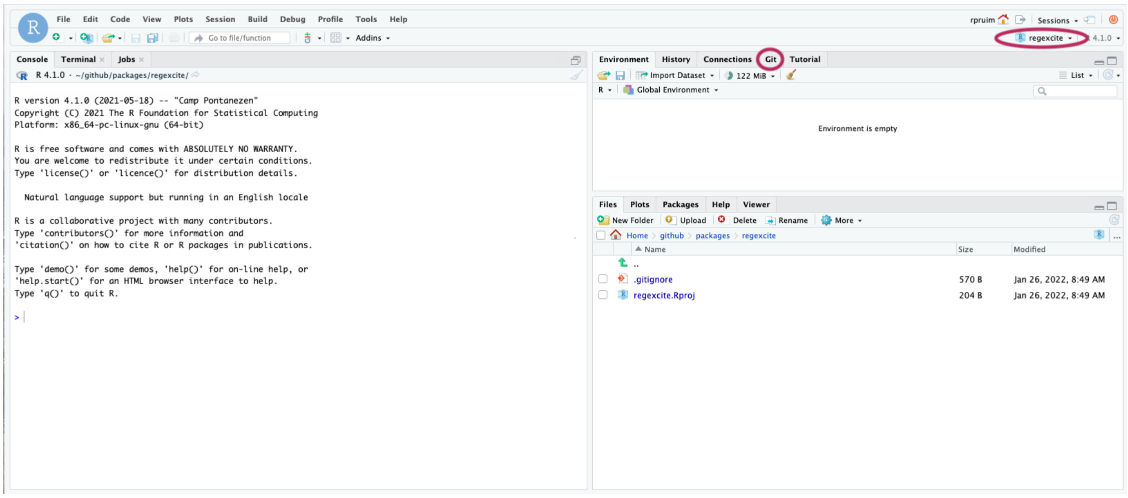


Repository URL: git@github.com:rpruim/regexcite.git

Project directory name:

Create project as subdirectory of: ~/github/packages

   - If you leave the project directory name empty, it will name the directory after your github repo name. That's usually a good option.
   - You can choose any path for your "Create project as a subdirectory of" as long as it isn't already under source control.
   - You should now see something like this:



R 4.1.0 · ~/github/packages/regexcite/

```
R version 4.1.0 (2021-05-18) -- "Camp Pontanezen"
Copyright (C) 2021 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.

  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.

> 
```

Environment is empty

Name	Size	Modified
..		
.gitignore	570 B	Jan 26, 2022, 8:49 AM
regexcite.Rproj	204 B	Jan 26, 2022, 8:49 AM

2.2.3 Step 3: Tell R you want to create a package

The steps so far would be the same for any R-based project we wanted to work on with GitHub. Now we want to set things up for creating a package. The source for an R package is just a bunch of files, some with **specified names and/or locations**. You can maintain all this manually, but it is much easier to let the computer take care of some of the bookkeeping for you. We will do this using the `devtools` package.

devtools is a meta-package As devtools grew, it got unwieldy to maintain and was split into several packages. The main task of `devtools` now is to load all those smaller packages. So you may see reference to packages like `usethis`, `remotes`, etc.

```r
library(devtools)
```

2.2.3.1 Attach the `devtools` package:

```
## Loading required package: usethis
```

```r
# check the package version

packageVersion('devtools')
```

```
## [1] '2.4.3'
```

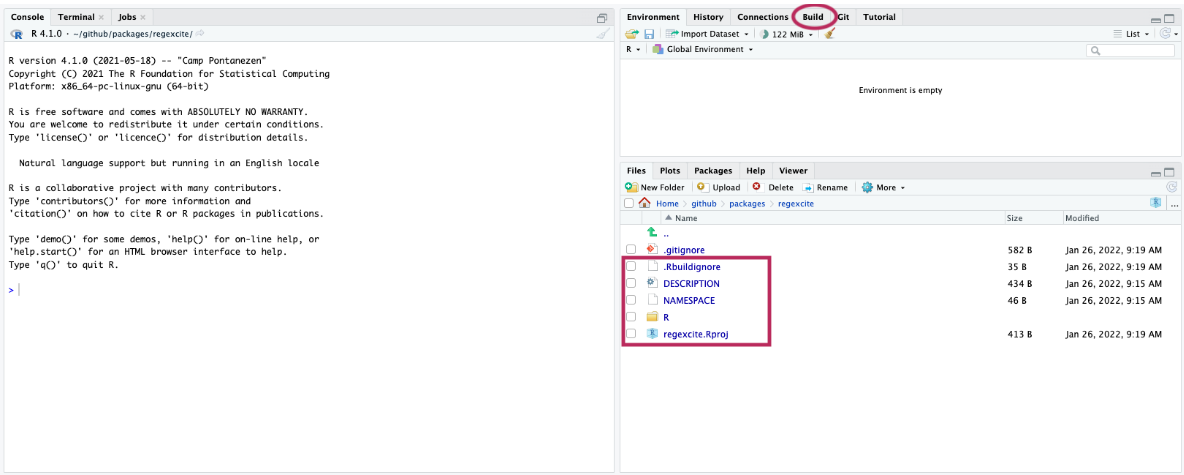
2.2.3.2 Run `create_package()` You will need to specify where your package code will be located. Since we just set up a new project for this, we'll put it at the root of the new project. Assuming you haven't changed your working directory, you can do the following.

R will warn you that you have a project going here already. That's OK, we're just getting started and this is what we intend. But generally you don't want to create a new package inside an existing project, so `devtools` is alerting you to the situation.

```
> create_package(".")
New project 'regexcite' is nested inside an existing project './', which is rarely a good idea.
If this is unexpected, the here package has a function, `here::dr_here()` that reveals why './' is rega
Do you want to create anyway?

1: No way
2: Negative
3: Definitely

Selection: 3
  Setting active project to '/home/rpruim/github/packages/regexcite'
  Creating 'R/'
  Writing 'DESCRIPTION'
Package: regexcite
Title: What the Package Does (One Line, Title Case)
Version: 0.0.0.9000
Authors@R (parsed):
    * First Last <first.last@example.com> [aut, cre] (YOUR-ORCID-ID)
Description: What the package does (one paragraph).
License: `use_mit_license()`, `use_gpl3_license()` or friends to
    pick a license
Encoding: UTF-8
Roxygen: list(markdown = TRUE)
RoxygenNote: 7.1.2
  Writing 'NAMESPACE'
Overwrite pre-existing file 'regexcite.Rproj'?

1: I agree
2: Negative
3: Not now
```

```
1
```

Your project should close and reopen and magically some new things appear:

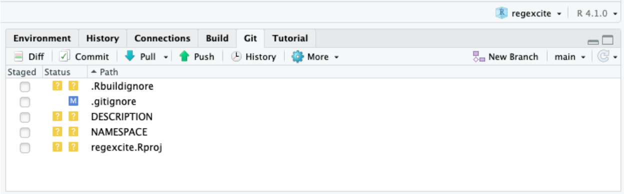


```
R version 4.1.0 (2021-05-18) -- "Camp Pontanezen"
Copyright (C) 2021 The R Foundation for Statistical Computing
Platform: x86_64-pc-linux-gnu (64-bit)

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.

  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.

> 
```

Name	Size	Modified
..		
.gitignore	582 B	Jan 26, 2022, 9:19 AM
.Rbuildignore	35 B	Jan 26, 2022, 9:19 AM
DESCRIPTION	434 B	Jan 26, 2022, 9:15 AM
NAMESPACE	46 B	Jan 26, 2022, 9:15 AM
R		
regexcite.Rproj	413 B	Jan 26, 2022, 9:19 AM

- **.Rbuildignore** lists files that we need to have around but that should not be included when building the R package from source. More in 4.3.1.
- **.Rproj.user**, if you have it, is a directory used internally by RStudio.
- **.gitignore** ignores some standard, behind-the-scenes files created by R and RStudio.
- **DESCRIPTION** provides metadata about your package. We will edit this shortly.
- **NAMESPACE** declares the functions your package exports for external use and the external functions your package imports from other packages. At this point, it is empty, except for a comment declaring that this is a file we will not edit by hand.
- The **R/** directory is the will soon contain .R files with function definitions.
- **regexcite.Rproj** is the file that makes this directory an RStudio Project. This file stores various settings for your project. You were prompted to overwrite this so that `devtools` could store some information here about your package.

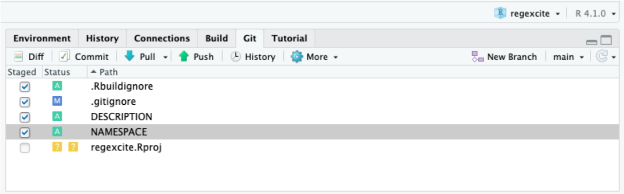
2.2.3.3 Commit your work It is a good idea to commit your work to git frequently. Let's do that now. You can do this using command line git in the Terminal tab, or you can use RStudio's git integration tab:



Staged	Status	Path
□	? ?	.Rbuildignore
□	M	.gitignore
□	? ?	DESCRIPTION
□	? ?	NAMESPACE
□	? ?	regexcite.Rproj

Notice that blue M. That indicates a file that has been modified. We had let GitHub create this file for us, and `devtools` added some more things to it.

The files with the question marks are not currently part of the repository. We should add them all. (Possible exception: the .Rproj file; there are different opinions about whether this belongs in the repo. I usually **do not put it in the repo** so that different collaborators can have different RStudio settings for the project.)



Staged	Status	Path
☑	A	.Rbuildignore
☑	M	.gitignore
☑	A	DESCRIPTION
☑	A	NAMESPACE
□	? ?	regexcite.Rproj

Now click commit and add your commit message – perhaps something like "initialize package with create_package()."

If you don't commit your .Rproj file, git will continue to nag you about it. This is what .gitignore is for. Edit that file so that this file is included and git will know that you want it to be ignored by git.

You should then commit that change as well.

Figure 1: Houston, we have ignition!

We have ignition, but we aren't really going anywhere yet. It's time to add some functionality to the package.

2.2.4 Make the package do something

Now we are ready to join the walk through from Hadley and Jenny at section 2.6.

A few notes:

1. The walk-through emphasizes R functions but many of the these things can also be done from the Build tab.

   So investigate that as well. You will see that if you use the Build tab, the same commands are executed in the console, but sometimes it is faster/easier to just click rather than type the command (and you don't have to remember the exact name of the R function).

2. **Don't do the use_github() section**.

   If you read that section, you will see that there are multiple ways to connect R and github for package development and the recommendation is to start at GitHub, like we have done. So you are ahead of the game here.

2.2.5 Share your work

Finish the example in Chapter 2. Make sure that you push your work to GitHub and that the repo on GitHub is public.

Packages can be installed directly from github with, for example:

```
devtools::install_github('user/package')
```

See the documentation of this function for some optional arguments that can be used to control whether documentation is built, which git branch to use, etc.

You will want/need to set up SSH authentication with github if you have not already done this so you don't have to type your password every time you do a push or a pull.

You can find instructions at GitHub, but there is also an R package that will assist you:

```
library(credentials)
ssh_key_info()
```

I've not used this, but it looks like it is super easy. See https://docs.ropensci.org/credentials/ for details.

3 Resources

3.1 R

- Advanced R by Hadley Wickham
- R for Data Science by Hadley Wickham and Garrett Grolemund
- R Packages by Hadley Wickham and Jenny Bryan
- https://stat545.com/: A textbook / course notes on "Data wrangling, exploration, and analysis with R"

3.2 Visualization

- Fundamentals of Data Visualization is very nice. The author doesn't advertise it plainly, but the source code for all the plots is available; look on the Technical notes page.

Course websites:

- CS 171 at Harvard
- 6.894 at MIT

Hardt, Moritz, Eric Price, Eric Price, and Nati Srebro. 2016. “Equality of Opportunity in Supervised Learning.” In *Advances in Neural Information Processing Systems 29*, edited by D. D. Lee, M. Sugiyama, U. V. Luxburg, I. Guyon, and R. Garnett, 3315–23. Curran Associates, Inc. http://papers.nips.cc/paper/6374-equality-of-opportunity-in-supervised-learning.pdf.