

# How R Markdown + Git + GitHub changed my (teaching?) life

R Summit & Workshop 2015

Dr. Jennifer (Jenny) Bryan

Dept. of Statistics & Michael Smith Laboratories, UBC

[jenny@stat.ubc.ca](mailto:jenny@stat.ubc.ca)

<http://stat545-ubc.github.io>

<http://www.stat.ubc.ca/~jenny/>

 [@JennyBryan](https://twitter.com/JennyBryan)

 [@STAT545](https://twitter.com/STAT545)

 [@jennybc](https://github.com/jennybc)

GitHub repo for this talk  
gives a bunch of links + PDF of slides

[https://github.com/jennybc/2015-06-28\\_r-summit-talk](https://github.com/jennybc/2015-06-28_r-summit-talk)

# Data wrangling, exploration, and analysis with R

UBC STAT 545A and 547M

Learn how to

- explore, groom, visualize, and analyze data
- make all of that reproducible, reusable, and shareable
- using R

<http://stat545-ubc.github.io>

<https://github.com/STAT545-UBC/STAT545-UBC.github.io>

MOST YEARS OF R  
USE BEFORE  
DEVELOPING AND  
DISTRIBUTING A  
PACKAGE



What's so great about  
(R) Markdown + Git(Hub)?

**weak links in the chain:  
process, packaging and  
presentation**



# R + markdown

Do your work

Get a presentable, web-friendly version for free

Present-ability is BAKED IN

... not a separate process you never get around to

stuff you  
need to  
write



stuff people  
like to  
read



stuff you  
need to  
write



stuff people  
like to  
read



stuff you  
need to  
write



stuff people  
like to  
read

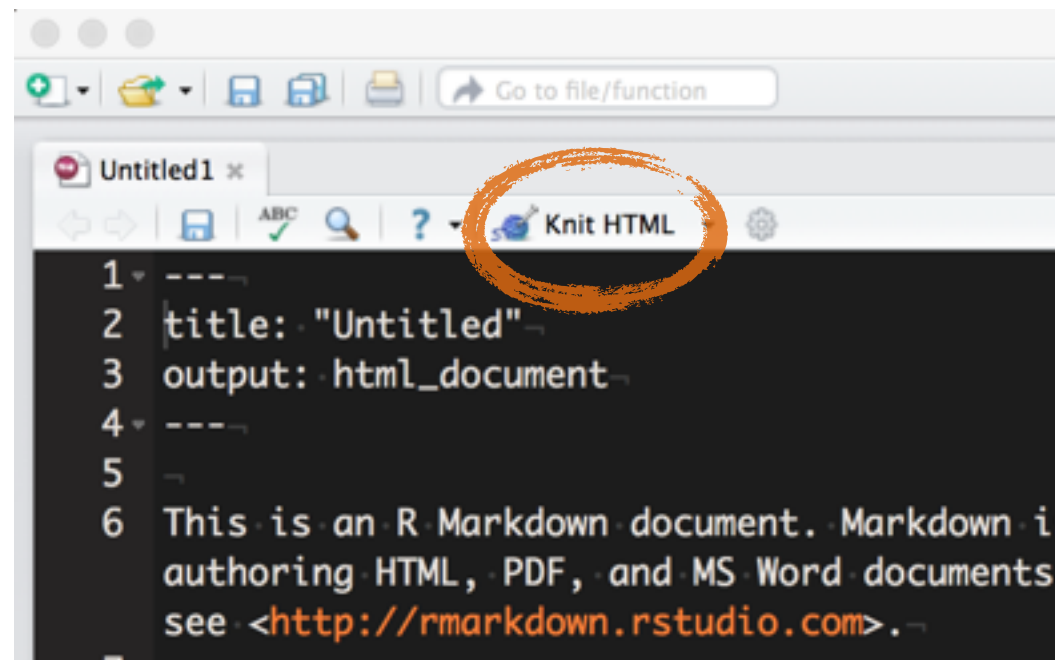
foo.R  
foo.Rmd



foo.md  
foo.html

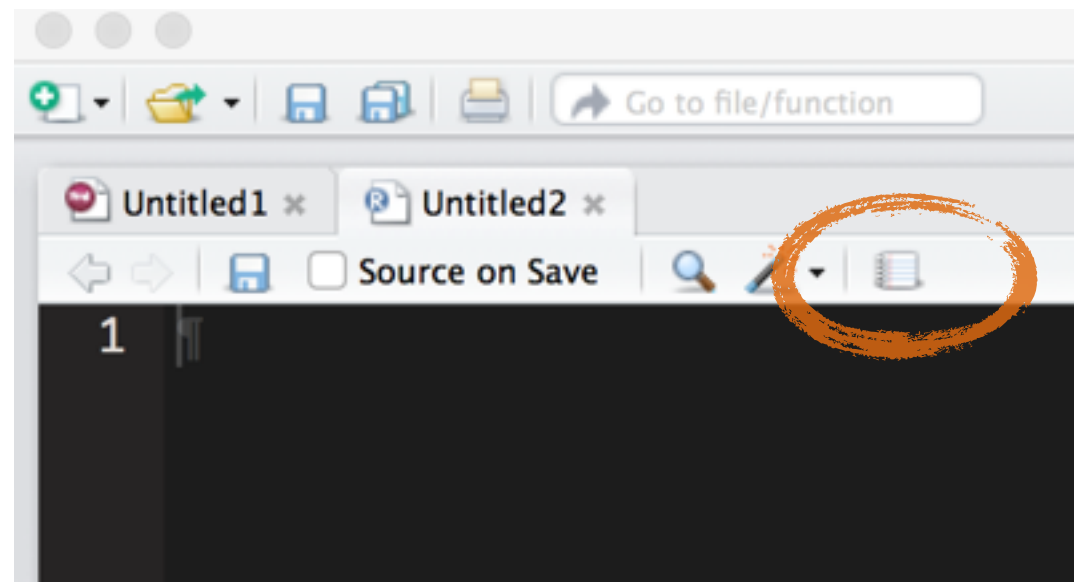


```
library(rmarkdown)  
render("foo.Rmd")
```





```
library(rmarkdown)  
render("foo.R")
```



foo.Rmd → foo.html

```
---
title: "Untitled"
output: html_document
---
```

foo.Rmd → foo.md → foo.html

```
---
title: "Untitled"
output:
  html_document:
    keep_md: yes
---
```

foo.Rmd → foo.md

```
---
output:
  md_document:
    variant: markdown_github
---
```

foo.R → foo.html

```
#' —  
#′ title: "Untitled"  
#′ output: html_document  
#′ ---
```

foo.R → foo.md → foo.html

```
#′ ---  
#′ title: "Untitled"  
#′ output:  
#′   html_document:  
#′     keep_md: yes  
#′ ---
```

foo.R → foo.md

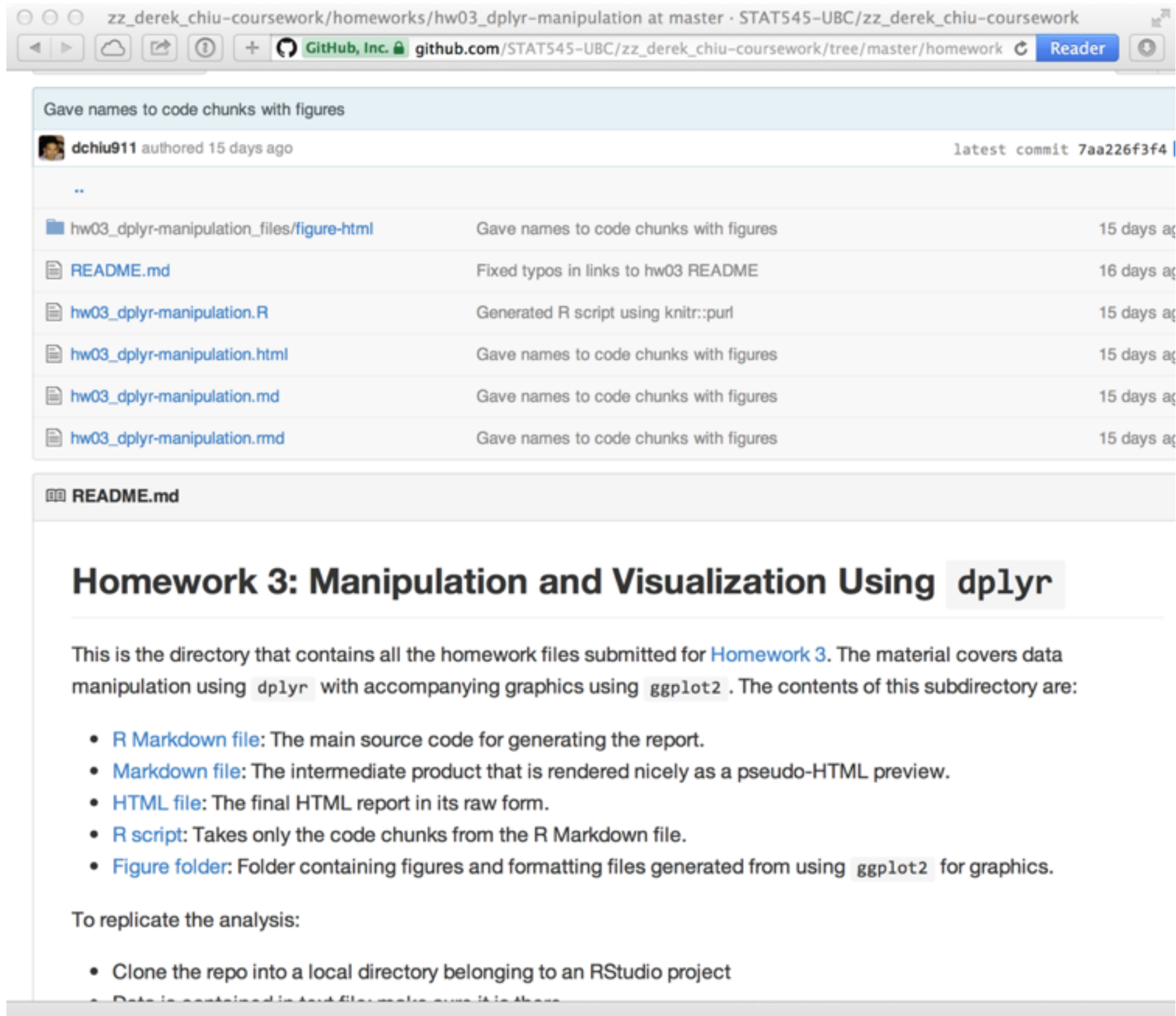
```
#′ ---  
#′ output:  
#′   md_document:  
#′     variant: markdown_github  
#′ ---
```

Markdown, and even Rmarkdown,  
are very privileged on GitHub

Essentially rendered as HTML



# When I mark homework ... this is what I see.



The screenshot shows a web browser displaying a GitHub repository page. The browser's address bar shows the URL: `github.com/STAT545-UBC/zz_derek_chiu-coursework/tree/master/homework`. The repository name is `zz_derek_chiu-coursework/homeworks/hw03_dplyr-manipulation` at master. The page shows a commit history table with columns for file changes, commit messages, and commit times. The latest commit is by `dchiu911` 15 days ago, with commit hash `7aa226f3f4`. The commit message is "Gave names to code chunks with figures". The table lists several files: `hw03_dplyr-manipulation_files/figure-html`, `README.md`, `hw03_dplyr-manipulation.R`, `hw03_dplyr-manipulation.html`, `hw03_dplyr-manipulation.md`, and `hw03_dplyr-manipulation.rmd`. Below the table, the `README.md` file is displayed. The README has a title "Homework 3: Manipulation and Visualization Using `dplyr`" and a description: "This is the directory that contains all the homework files submitted for Homework 3. The material covers data manipulation using `dplyr` with accompanying graphics using `ggplot2`. The contents of this subdirectory are:". The README lists five items: "R Markdown file", "Markdown file", "HTML file", "R script", and "Figure folder". It also includes a section "To replicate the analysis:" with a list of steps.

zz\_derek\_chiu-coursework/homeworks/hw03\_dplyr-manipulation at master · STAT545-UBC/zz\_derek\_chiu-coursework

github.com/STAT545-UBC/zz\_derek\_chiu-coursework/tree/master/homework

Reader

Gave names to code chunks with figures

dchiu911 authored 15 days ago latest commit 7aa226f3f4

..

hw03_dplyr-manipulation_files/figure-html	Gave names to code chunks with figures	15 days ago
README.md	Fixed typos in links to hw03 README	16 days ago
hw03_dplyr-manipulation.R	Generated R script using knitr::purl	15 days ago
hw03_dplyr-manipulation.html	Gave names to code chunks with figures	15 days ago
hw03_dplyr-manipulation.md	Gave names to code chunks with figures	15 days ago
hw03_dplyr-manipulation.rmd	Gave names to code chunks with figures	15 days ago

README.md

## Homework 3: Manipulation and Visualization Using `dplyr`

This is the directory that contains all the homework files submitted for [Homework 3](#). The material covers data manipulation using `dplyr` with accompanying graphics using `ggplot2`. The contents of this subdirectory are:

- [R Markdown file](#): The main source code for generating the report.
- [Markdown file](#): The intermediate product that is rendered nicely as a pseudo-HTML preview.
- [HTML file](#): The final HTML report in its raw form.
- [R script](#): Takes only the code chunks from the R Markdown file.
- [Figure folder](#): Folder containing figures and formatting files generated from using `ggplot2` for graphics.

To replicate the analysis:

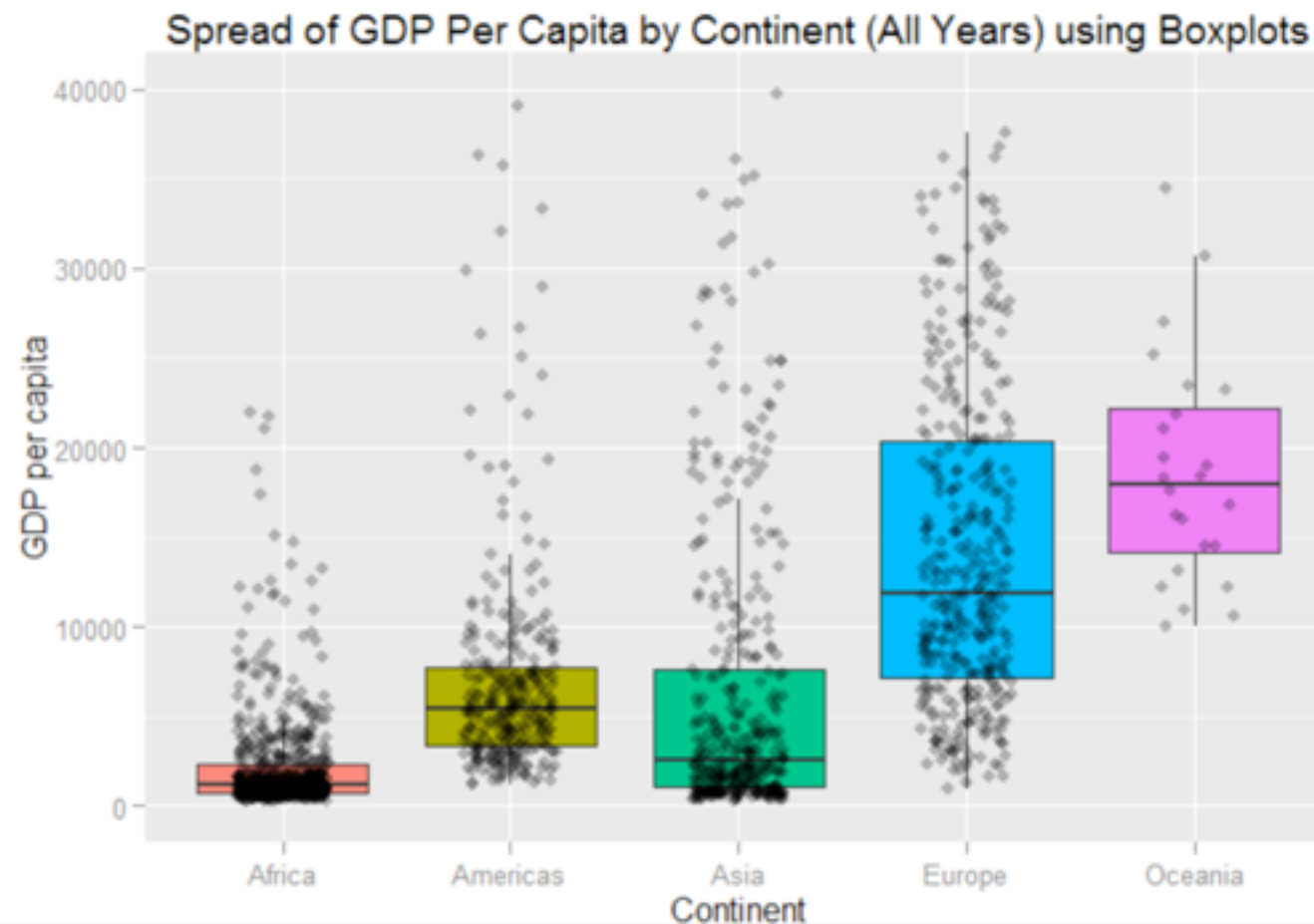
- Clone the repo into a local directory belonging to an RStudio project
- Data is contained in text files, make sure it is there



# When I mark homework ... this is what I see.

In this section, we will use two similar graphs to visualize spread: the **boxplot** and the **violin plot**. Note that the original unmanipulated data frame `gtbl` will be used here.

```
ggplot(gtbl, aes(continent, gdpPercap))+  
  geom_boxplot(aes(fill = continent), outlier.shape = NA)+  
  geom_jitter(alpha = 0.3, position = position_jitter(width = 0.2))+  
  xlab("Continent")+  
  ylab("GDP per capita")+  
  ylim(c(0,40000))+  
  theme(legend.position = "none")+  
  ggtitle("Spread of GDP Per Capita by Continent (All Years) using Boxplots")
```



# When I mark homework ... this is what I see.

## STAT 545A

- [Homework 1](#): Edit README.md and experiment with Markdown
- [Homework 2](#): Exploring the Gapminder Dataset
- [Homework 3](#): Manipulation and Visualization Using `dplyr`
- [Homework 4](#): Writing and Testing Functions
- [Homework 5](#): Factor Control and File I/O
- [Homework 6](#): *Optional*: Transition Activities

## STAT 547M

- [Homework 7](#): Data Wrangling Grand Finale
- [Homework 8](#): Data Cleaning
- [Homework 9](#): Automating Data Analysis Pipelines
- [Homework 10](#): Building an R Package
- [Homework 11](#): Building a Shiny App
- [Homework 12](#): Getting Data off the Web

What's so great about  
(R) Markdown + Git(Hub)?

# What's so great about Git(Hub)?

search

linky-ness

meta-stuff (comments, issues, ...)

Problem:

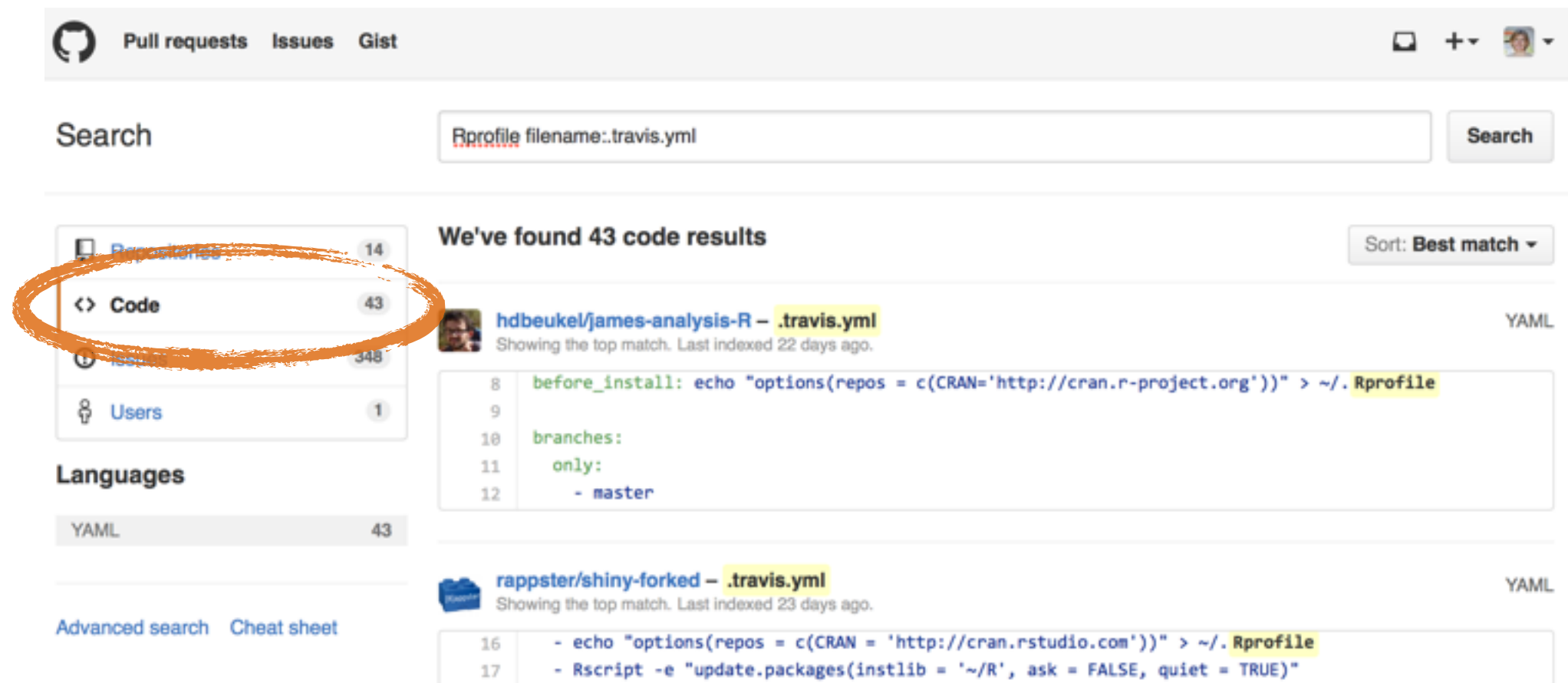
NOTE from R CMD check when testing a package on Travis ... but I *am* specifying my CRAN repo?

```
* checking package dependencies ... NOTE  
No repository set, so cyclic dependency check skipped
```

Hmm ... maybe the CRAN environment variable isn't consulted, maybe I need to set the CRAN repo set in `.Rprofile` ... how do I *that* on Travis?

Solution:

Find a ton of other `.travis.yml` files that mention `.Rprofile`!



GitHub interface showing search results for `Rprofile filename:.travis.yml`.

Search bar: `Rprofile filename:.travis.yml` Search

Filters: Pull requests, Issues, Gist

Sort: Best match

We've found 43 code results

Filters: **Code** (43), Issues (348), Users (1)

Languages: YAML (43)

Advanced search Cheat sheet

Top results:

- hdbeukel/james-analysis-R — .travis.yml** (YAML)  
Showing the top match. Last indexed 22 days ago.  
8 `before_install: echo "options(repos = c(CRAN='http://cran.r-project.org'))" > ~/.Rprofile`  
9  
10 `branches:`  
11 `only:`  
12 `- master`
- rappster/shiny-forked — .travis.yml** (YAML)  
Showing the top match. Last indexed 23 days ago.  
16 `- echo "options(repos = c(CRAN = 'http://cran.rstudio.com'))" > ~/.Rprofile`  
17 `- Rscript -e "update.packages(instlib = '~/R', ask = FALSE, quiet = TRUE)"`



**Jennifer Bryan** @JennyBryan · May 29

We have `isTRUE` but not `isFALSE`? 😡 #rstats

RETWEET

1

FAVORITES

9



Replies included discussion of why we have `isTRUE ( )` and the existence + history of various “truthy” things

Included links to where all this occurs in the R source and when it was added

```
120 }
121 #endif
122
123 const static char * const truenames[] = {
124     "T",
125     "True",
126     "TRUE",
127     "true",
128     (char *) NULL,
129 };
130
131 const static char * const falsenames[] = {
132     "F",
133     "False",
134     "FALSE",
135     "false",
136     (char *) NULL,
137 };
138
```

<https://github.com/wch/r-source/blob/trunk/src/main/util.c#L445-L452>

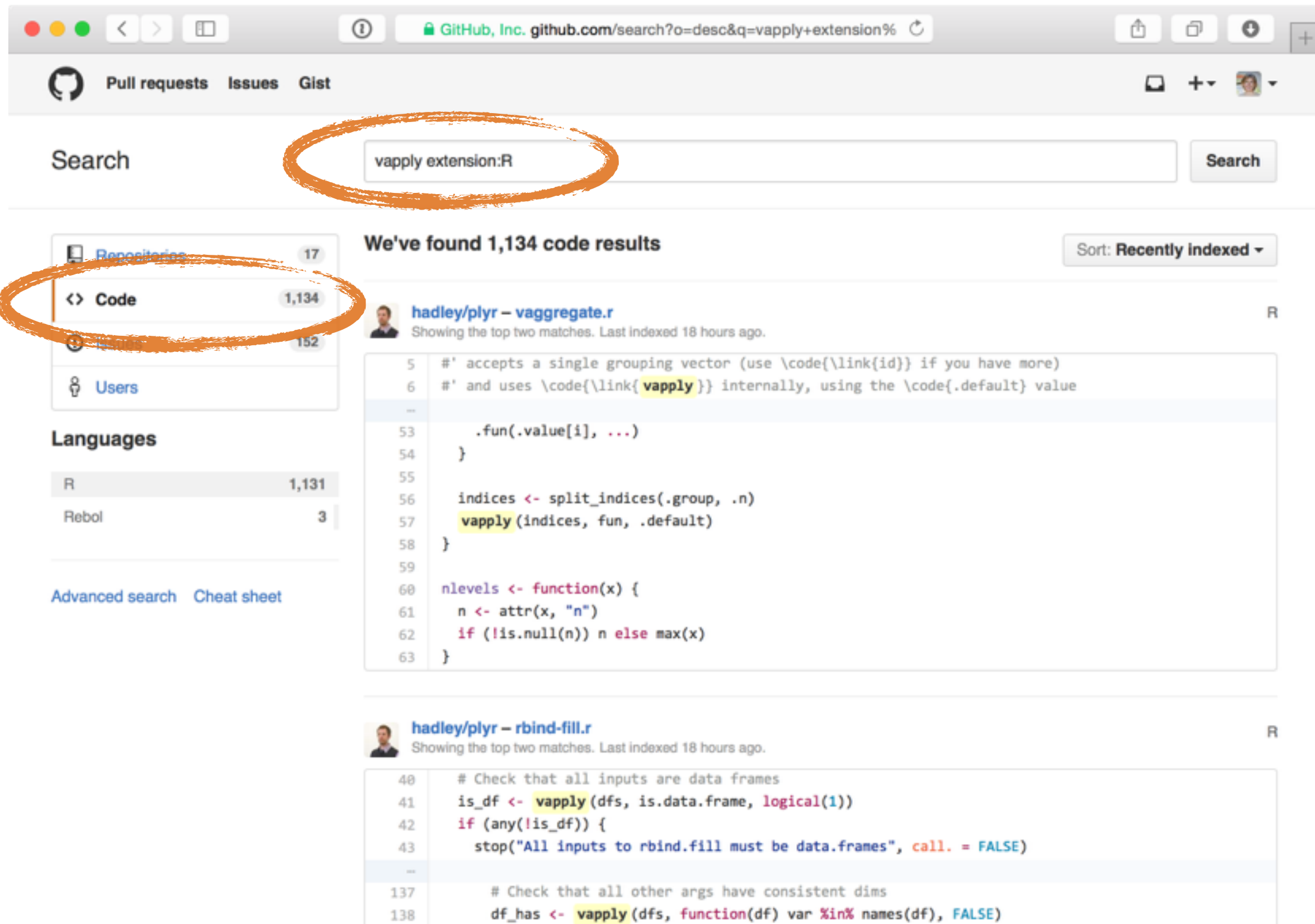
```
442
443 /* Function to test whether a string is a true value */
444
445 Rboolean StringTrue(const char *name)
446 {
447     int i;
448     for (i = 0; truenames[i]; i++)
449         if (!strcmp(name, truenames[i]))
450             return TRUE;
451     return FALSE;
452 }
453
454 Rboolean StringFalse(const char *name)
455 {
456     int i;
457     for (i = 0; falsenames[i]; i++)
458         if (!strcmp(name, falsenames[i]))
459             return TRUE;
460     return FALSE;
461 }
```

<https://github.com/wch/r-source/blob/trunk/src/main/util.c#L123-L129>





Are the official examples for a function kind of thin?  
Search Github to see the function “in the wild”.



GitHub, Inc. github.com/search?o=desc&q=vapply+extension%

Pull requests Issues Gist

Search  Search

We've found 1,134 code results Sort: Recently indexed

**Code** 1,134

**Languages**

Language	Count
R	1,131
Rebol	3

[Advanced search](#) [Cheat sheet](#)

**hadley/plyr – vaggregate.r** R

Showing the top two matches. Last indexed 18 hours ago.

```
5 #' accepts a single grouping vector (use \code{\link{id}} if you have more)
6 #' and uses \code{\link{vapply}} internally, using the \code{.default} value
...
53 .fun(.value[i], ...)
54 }
55
56 indices <- split_indices(.group, .n)
57 vapply(indices, fun, .default)
58 }
59
60 nlevels <- function(x) {
61   n <- attr(x, "n")
62   if (!is.null(n)) n else max(x)
63 }
```

**hadley/plyr – rbind.fill.r** R

Showing the top two matches. Last indexed 18 hours ago.

```
40 # Check that all inputs are data frames
41 is_df <- vapply(dfs, is.data.frame, logical(1))
42 if (any(!is_df)) {
43   stop("All inputs to rbind.fill must be data.frames", call. = FALSE)
...
137 # Check that all other args have consistent dims
138 df_has <- vapply(dfs, function(df) var %in% names(df), FALSE)
```

Are the official examples for a function kind of thin?  
Search Github to see the function “in the wild”.

GitHub, Inc. github.com/search?q=vapply+user%3Acran&type=Code

Pull requests Issues Gist

Search  Search

We've found 894 code results Sort: Best match ▾

**Code** 894

**Languages**

R	866
RMarkdown	4
Text	3
Rebol	2
HTML	2
Org	1
Markdown	1
INI	1

**cran/tumblr - oauth.encode.R** R  
Showing the top match. Last indexed 25 days ago.

```
1  oauth.encode <-  
2  function(x) vapply(x, oauth.encode1, character(1))
```

**cran/tumblr - compact.R** R  
Showing the top match. Last indexed 25 days ago.

```
1  compact <-  
2  function(x) {  
3      null <- vapply(x, is.null, logical(1))  
4      x[!null]  
5  }
```

**cran/sdprisk - almost.equal.R** R  
Showing the top match. Last indexed 27 days ago.

```
1  almost.equal <- function(x, value) {  
2      vapply(Map(all.equal, x, rep.int(value, length(x))), isTRUE, logical(1L))  
3  }
```

[Advanced search](#) [Cheat sheet](#)



# STAT 545 @ University of British Columbia

Course in data wrangling, exploration, and analysis with R

Vancouver, BC

Filters ▾

Find a repository...

+ New repository

## STAT545-UBC.github.io

HTML ★ 47 P 39

Main repository for STAT 545 @ University of British Columbia, a course in data wrangling, exploration, and analysis with R.

Updated 4 days ago

## InstructorsOnly PRIVATE

R ★ 0 P 0

Private repository for STAT 545 instructors

Updated on Jan 25

## Discussion

★ 3 P 1

Public discussion

Updated on Sep 17, 2014

## zz\_derek\_chiu-coursework PRIVATE

HTML ★ 0 P 1

coursework created for zz\_derek\_chiu

Updated on Mar 25

← Github Organization

← Content, website

← Shh... secret

← Discussion forum

← 1 repo per student

# Issues as discussion forum

Issues · STAT545-UBC/Discussion

github.com/STAT545-UBC/Discussion/issues

STAT545-UBC / Discussion

Unwatch 23 Star 1 Fork 1

Issues Pull requests Labels Milestones Filters is:issue is:open New Issue

19 Open 64 Closed

	Author	Labels	Milestones	Assignee	Sort
<input type="checkbox"/> <b>Future plans for your STAT 545 / 547 coursework repository</b> #83 opened on Dec 18, 2014 by jennybc					3
<input type="checkbox"/> <b>Worked example: when two data.frames are almost, but not quite, the same</b> #82 opened on Dec 16, 2014 by jennybc					1
<input type="checkbox"/> <b>Using `rplots` in RMarkdown without including API key inside RMarkdown?</b> #81 opened on Dec 5, 2014 by spencerfrei					5
<input type="checkbox"/> <b>Remixing data -- `dplyr` bug</b> #80 opened on Dec 3, 2014 by aammd					0
<input type="checkbox"/> <b>Small detail: Indented Rstudio code not displaying nicely with github's tab setting of 8 spaces</b> <span>Mac OS</span> <span>Windows</span> #78 opened on Nov 26, 2014 by jooolia					0
<input type="checkbox"/> <b>Deploying our shiny apps to the UBC stats server</b> #76 opened on Nov 25, 2014 by daattali					2



# Issues to submit, peer review, and mark homework

The screenshot shows a web browser window displaying the GitHub interface for the repository 'STAT545-UBC / zz\_derek\_chiu-coursework'. The browser's address bar shows the URL: `github.com/STAT545-UBC/zz_derek_chiu-coursework/issues?q=is%3Aissue+is%3Aclosed`. The repository is marked as 'PRIVATE'. The user 'jennybc' is logged in. The 'Issues' tab is selected, and a search filter 'is:issue is:closed' is applied. The interface shows 0 Open and 33 Closed issues. A list of issues is displayed, including 'Mark homework 12 of Derek-Chiu', 'Peer review of derek\_chiu's hw11 by [redacted]', 'Peer review of derek\_chiu's hw11 by [redacted]', 'Mark homework 11 of Derek-Chiu', and 'Peer review of derek\_chiu's hw10 by [redacted]'. Each issue entry includes a checkbox, a red 'F' icon, the issue title, a comment count, and a date. Labels like 'hw11' and 'peer-review' are visible on some issues. A sidebar on the right contains icons for code, issues, pull requests, and other repository features.

Issues · STAT545-UBC/zz\_derek\_chiu-coursework

github.com/STAT545-UBC/zz\_derek\_chiu-coursework/issues?q=is%3Aissue+is%3Aclosed

This repository Search Explore Gist Blog Help jennybc

STAT545-UBC / zz\_derek\_chiu-coursework PRIVATE

Unwatch 2 Star 0 Fork 1

Issues Pull requests Labels Milestones Filters is:issue is:closed New issue


Clear current search query, filters, and sorts

0 Open ✓ 33 Closed

Author Labels Milestones Assignee Sort

- Mark homework 12 of Derek-Chiu  
#33 opened on Dec 5, 2014 by dchiu911 2
- Peer review of derek\_chiu's hw11 by [redacted] hw11 peer-review 2
- Peer review of derek\_chiu's hw11 by [redacted] hw11 peer-review 2
- Mark homework 11 of Derek-Chiu  
#30 opened on Dec 1, 2014 by dchiu911 2
- Peer review of derek\_chiu's hw10 by [redacted] hw10 peer-review 3
- Peer review of derek\_chiu's hw10 by [redacted] hw10 peer-review

GitHub, Inc. github.com/na




Search GitHub

Pull requests

Issues


Gist


+





Nate Anderson


na

 Predictive Engineering

 Eugene, Or

 nande12@gmail.com

 http://nateanderson.net

 Joined on Feb 28, 2008

12

Followers

79

Starred

11

Following


Contributions

Repositories

Public activity


Follow

Popular repositories

 [django-vim](#)


A set of vim plugins and settings for using vim to develop django.

37 ★

 [Webfaction-Node](#)


Node install script for webfaction.com

18 ★

 [django-metatags](#)


Simple django app to add html metatags to an object.

5 ★

 [codereviewr](#)

A site to make your code lean and mean.

2 ★

 [blue-channel](#)

A content management system developed in Django, jQuery and 960

1 ★

Public contributions

Jul

Aug

Sep

Oct

Nov

Dec

Jan

Feb

Mar

Apr

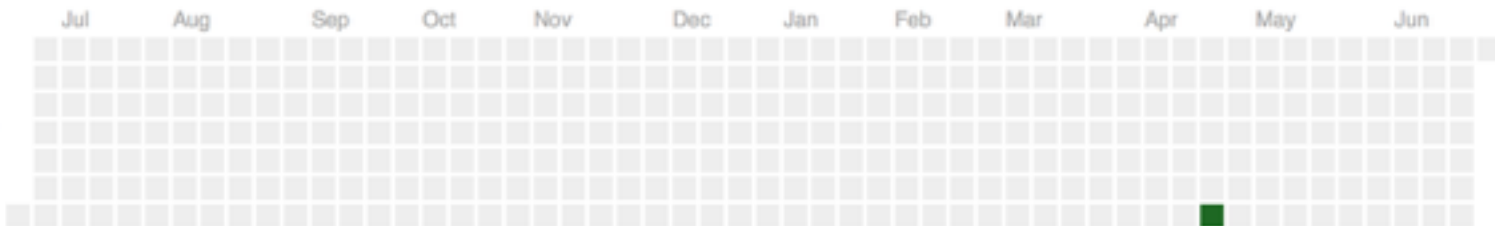
May

Jun


M

W

F



Summary of Pull Requests, issues opened, and commits. [Learn more.](#)

Less  More

Contributions in the last year

2 total

Jun 28, 2014 – Jun 28, 2015

Longest streak

1 day

April 25 – April 25

Current streak

0 days

Last contributed 2 months ago

sorry about all those issues!

# The unreasonable effectiveness of GitHub browsability

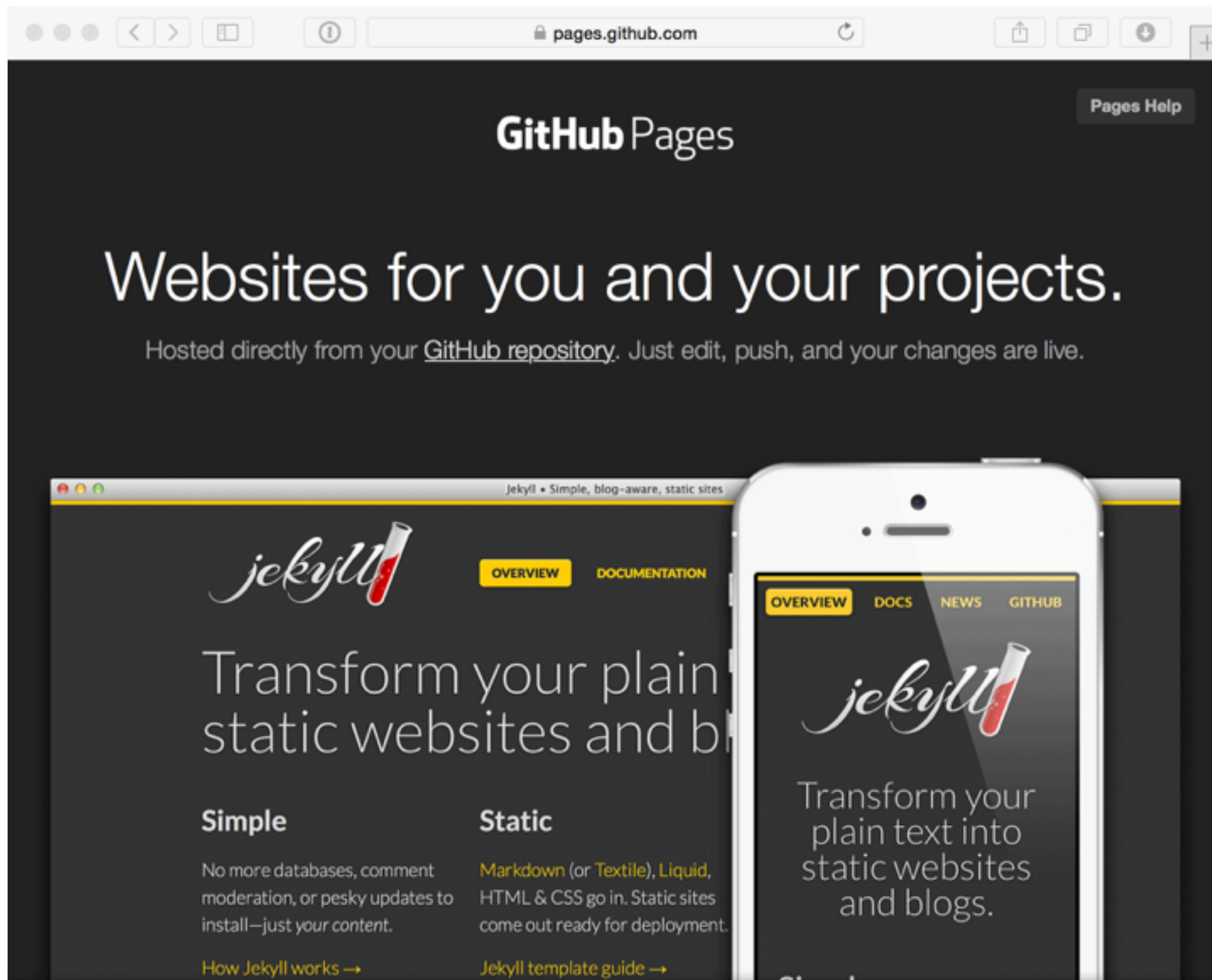


Commit .md files and figures!  
They are not like .o files or executables.  
Get over it.



When (intermediate) Markdown is no longer enough ...

`<my_thing>.github.io`



reproducible  
presentable

search

linky-ness

meta-stuff, e.g. issues

R Markdown v2

+



=



collaboration: “link, don’t attach”  
machine- and human-readable  
meta self-documentation