BIMM 143

Course Summary & GitHub Portfolio

Class 20

Barry Grant

UC San Diego

http://thegrantlab.org/bimm143

Today’s Menu

- Summary of major learning goals,
- Course discussion and feedback (https://etherpad.wikimedia.org/p/bimm143_w26)
- SET/CAPs evaluation (Link)
- Final exam
  - ➡ Test structure, guidelines and rules
  - ➡ Topics and example questions
  - ➡ Exam preparation, discussion and open study
- Polish our GitHub content and publish your own website portfolios

Today's Menu

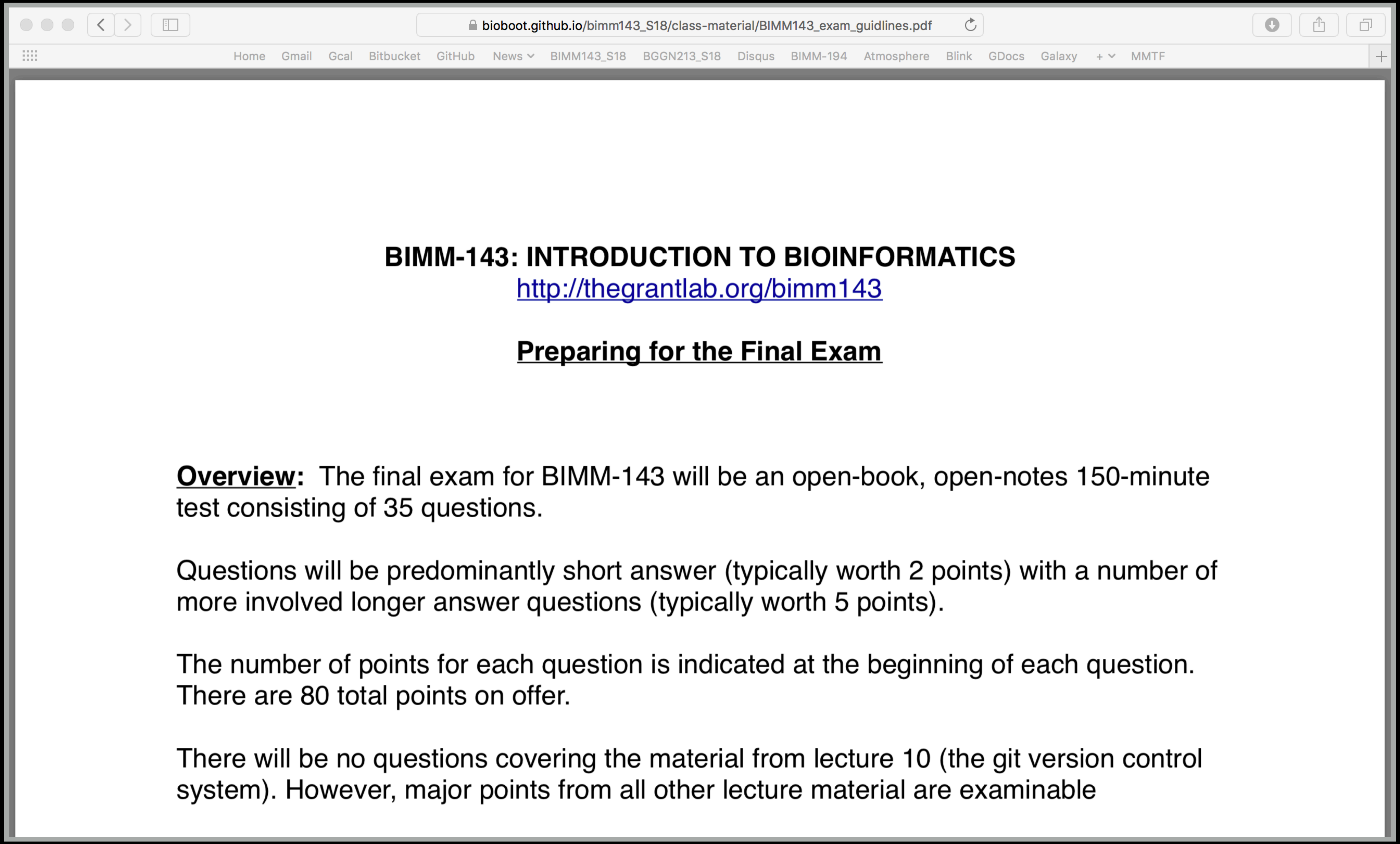
- Summary of major learning goals,
- Course discussion and feedback (https://etherpad.wikimedia.org/p/bimm143_w26)
- SET/CAPs evaluation (Link)
- Final exam
  - Test structure, guidelines and rules
  - Topics and example questions
  - Exam preparation, discussion and open study

- Polish our GitHub content and publish your own website portfolios

Today's Menu

- Summary of major learning goals,
- Course discussion and feedback (https://etherpad.wikimedia.org/p/bimm143_w26)
- SET/CAPs evaluation (Link)
- Project troubleshooting
  - Test structure, guidelines and rules
  - Topics and example questions
  - Exam preparation, discussion and open study
- Polish our GitHub content and publish your own website portfolios

https://bioboot.github.io/bimm143_F22/class-material/BIMM143_exam_guidlines.pdf



BIMM-143: INTRODUCTION TO BIOINFORMATICS

http://thegrantlab.org/bimm143

Preparing for the Final Exam

Overview: The final exam for BIMM-143 will be an open-book, open-notes 150-minute test consisting of 35 questions.

Questions will be predominantly short answer (typically worth 2 points) with a number of more involved longer answer questions (typically worth 5 points).

The number of points for each question is indicated at the beginning of each question. There are 80 total points on offer.

There will be no questions covering the material from lecture 10 (the git version control system). However, major points from all other lecture material are examinable

Q1. Did you enjoy this course? (Rank in relation to others you have experienced at UCSD)?

Q2. Should this course be offered again?

Q3. If so what changes would you recommend for this course? (e.g. more/less DataCamp & Projects)

Q4. Was the course effectively organized (lecture and lab material online vs handout or TritonEd site)?

Q5. What advice would you give to another student who is considering taking this course?

Q6. Considering both the limitations and possibilities of the subject matter and the course, how would you rate the overall effectiveness of this course and instructor?

Q7. Do you agree or disagree - The course developed my abilities and skills for the subject?

Q8. On average, how many hours per week have you spent on this course, including attending classes, doing homework's and assignments?

Q9. Any other comments you would like to share?

EtherPad: **https://etherpad.wikimedia.org/p/bimm143_w26**

Form Version: https://tinyurl.com/bimm143-end

Thank you very much!

Announcement!

Bonus:

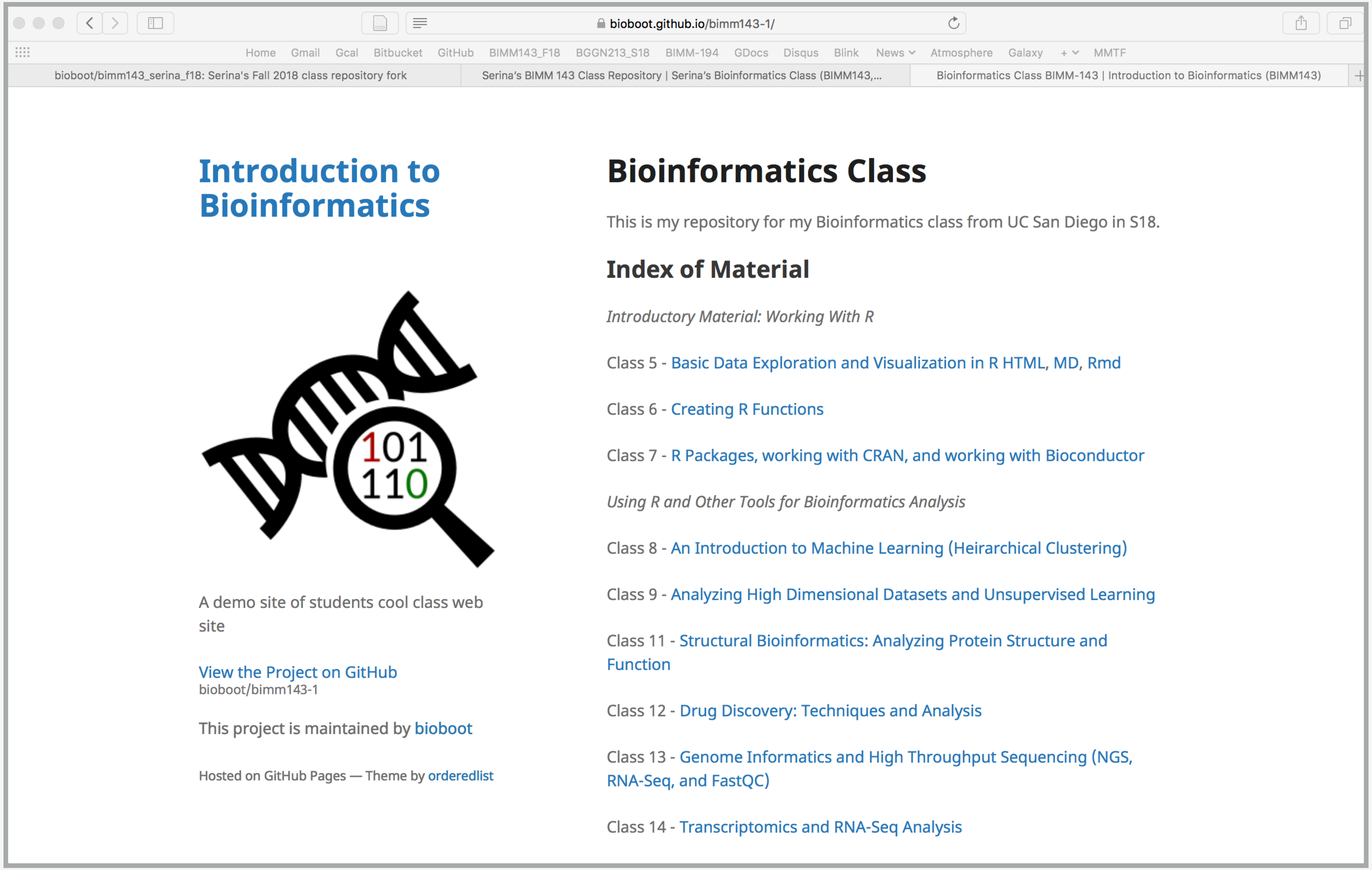
Bioinformatics & Genomics in Industry

Live Stream Video

Enjoy a set of short open ended guest lectures from leading genomic scientists at **Illumina Inc.**, **Synthetic Genomics Inc.**, and the **La Jolla Institute for Allergy and Immunology**. Feel free to contact these scientists for networking and to have your questions about industry careers in Bioinformatics and Genomics answered.

Bonus:

GitHub Spit & Polish



Introduction to Bioinformatics

A demo site of students cool class web site

View the Project on GitHub
bioboot/bimm143-1

This project is maintained by bioboot

Hosted on GitHub Pages — Theme by orderedlist

Bioinformatics Class

This is my repository for my Bioinformatics class from UC San Diego in S18.

Index of Material

Introductory Material: Working With R

Class 5 - Basic Data Exploration and Visualization in R HTML, MD, Rmd

Class 6 - Creating R Functions

Class 7 - R Packages, working with CRAN, and working with Bioconductor

Using R and Other Tools for Bioinformatics Analysis

Class 8 - An Introduction to Machine Learning (Heirarchical Clustering)

Class 9 - Analyzing High Dimensional Datasets and Unsupervised Learning

Class 11 - Structural Bioinformatics: Analyzing Protein Structure and Function

Class 12 - Drug Discovery: Techniques and Analysis

Class 13 - Genome Informatics and High Throughput Sequencing (NGS, RNA-Seq, and FastQC)

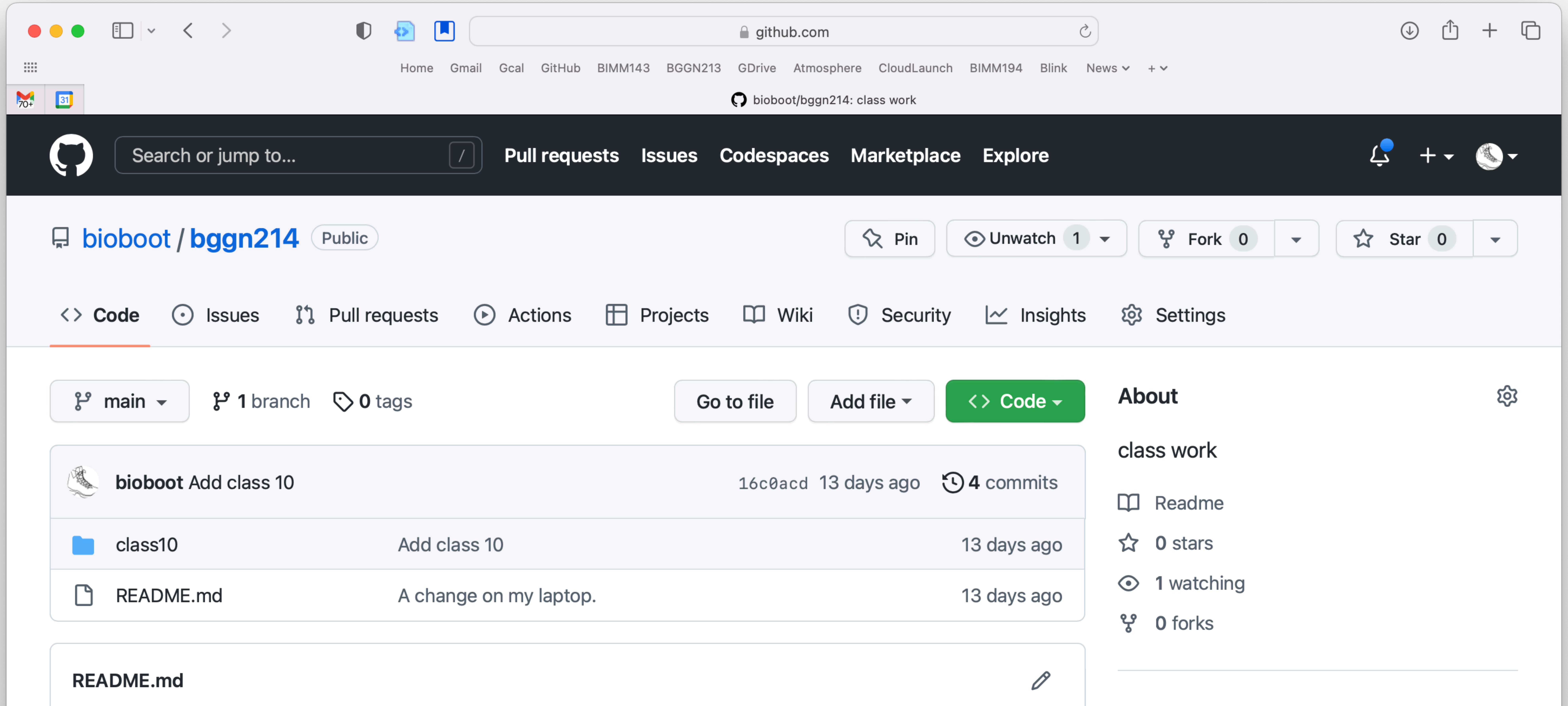
Class 14 - Transcriptomics and RNA-Seq Analysis

In your web browser navigate to your GitHub class repository < https://github.com/ >

Side-note:

To find the link to your GitHub repository from RStudio, open one of your past class projects and in the terminal type:

```
git remote -v
```



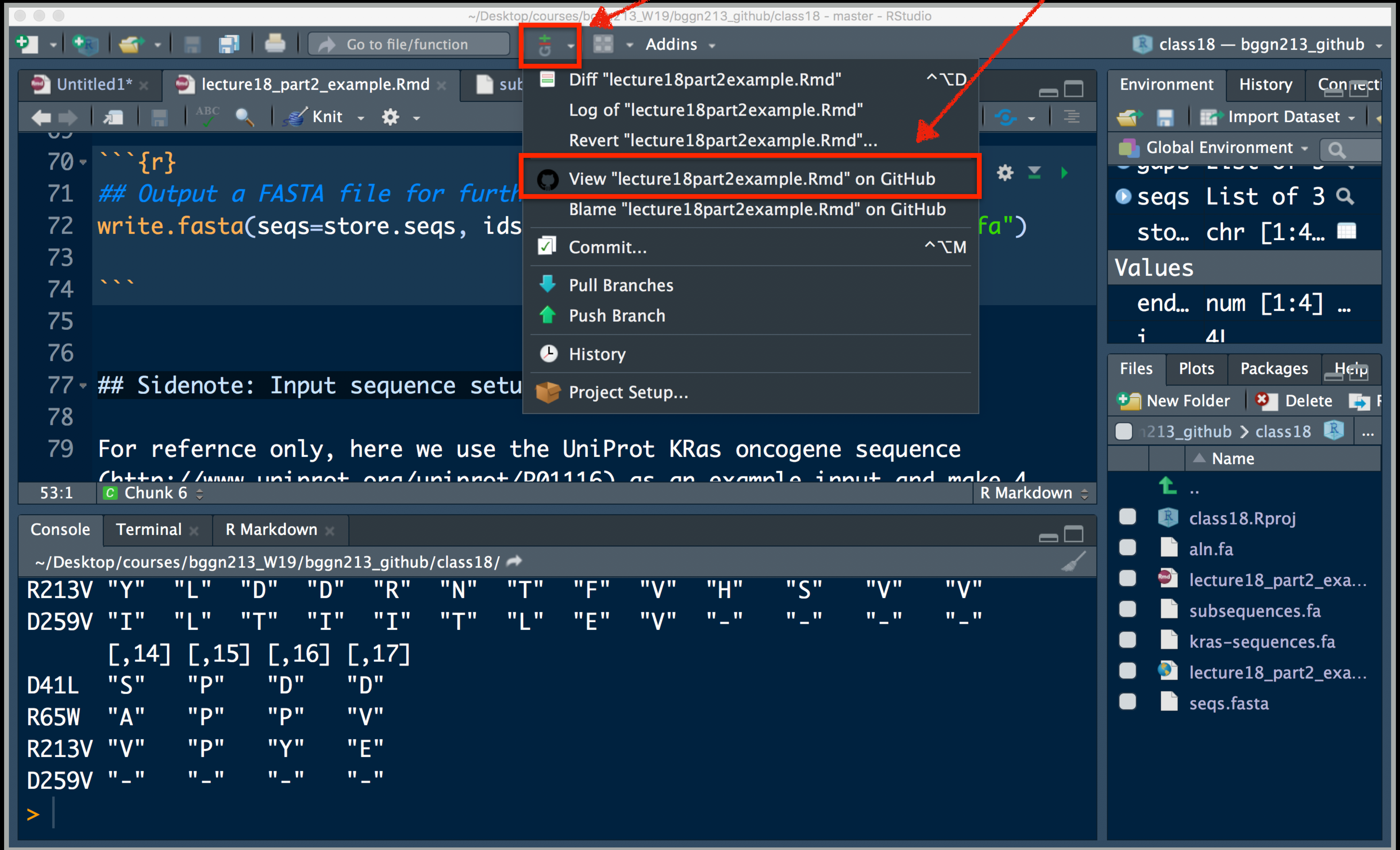
bioboot Add class 10 16c0acd 13 days ago 4 commits

class10	Add class 10	13 days ago
README.md	A change on my laptop.	13 days ago

class work

README.md

Or for a given GitHub tracked file click GIT icon and "VIEW on GITHUB"



~~~
```{r}
## Output a FASTA file for furth
write.fasta(seqs=store.seqs, ids                                     fa")

~~~

Sidenote: Input sequence setu

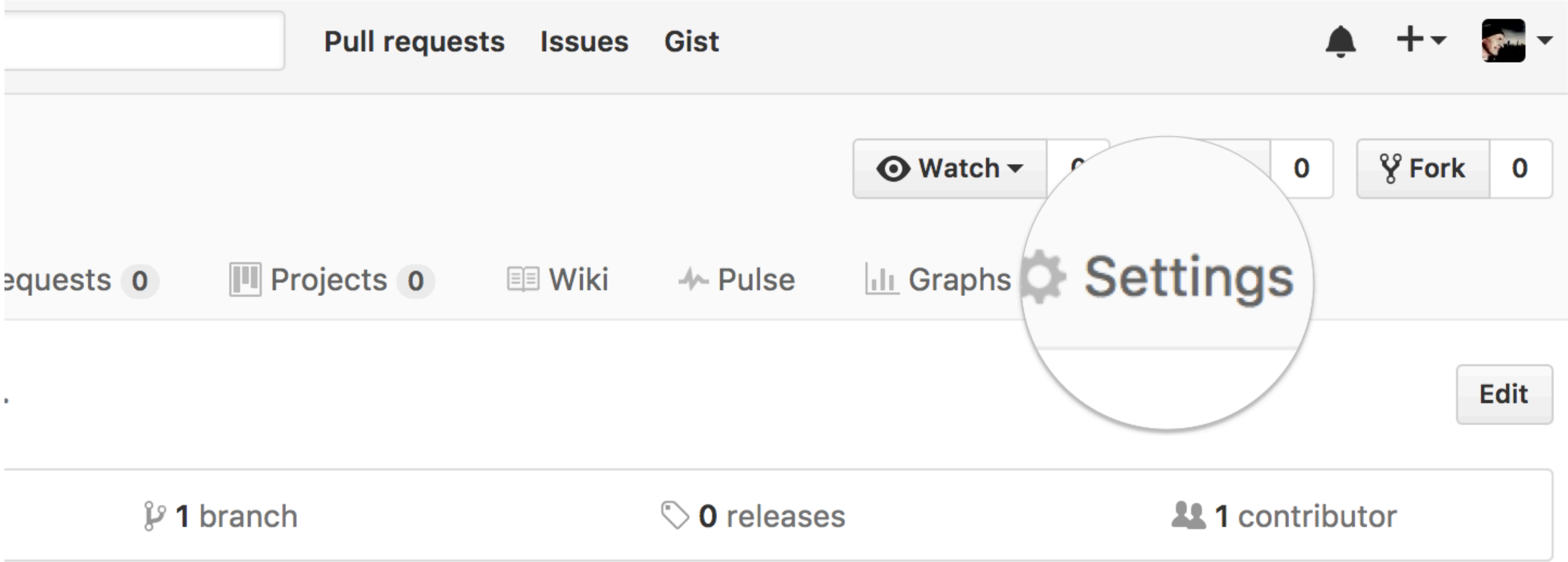
For refernce only, here we use the UniProt KRas oncogene sequence
(http://www.uniprot.org/uniprot/P01116) as an example input and make 4
```

```
R213V "Y"  "L"  "D"  "D"  "R"  "N"  "T"  "F"  "V"  "H"  "S"  "V"  "V"
D259V "I"  "L"  "T"  "I"  "I"  "T"  "L"  "E"  "V"  "-"  "-"  "-"  "-"
      [,14] [,15] [,16] [,17]
D41L  "S"   "P"   "D"   "D"
R65W  "A"   "P"   "P"   "V"
R213V "V"   "P"   "Y"   "E"
D259V "-"   "-"   "-"   "-"
>
```
```

1

# Repository Settings

**Head over to your GitHub class repository (where you are "pushing" all your class work)**

**Click on the Settings tab.**

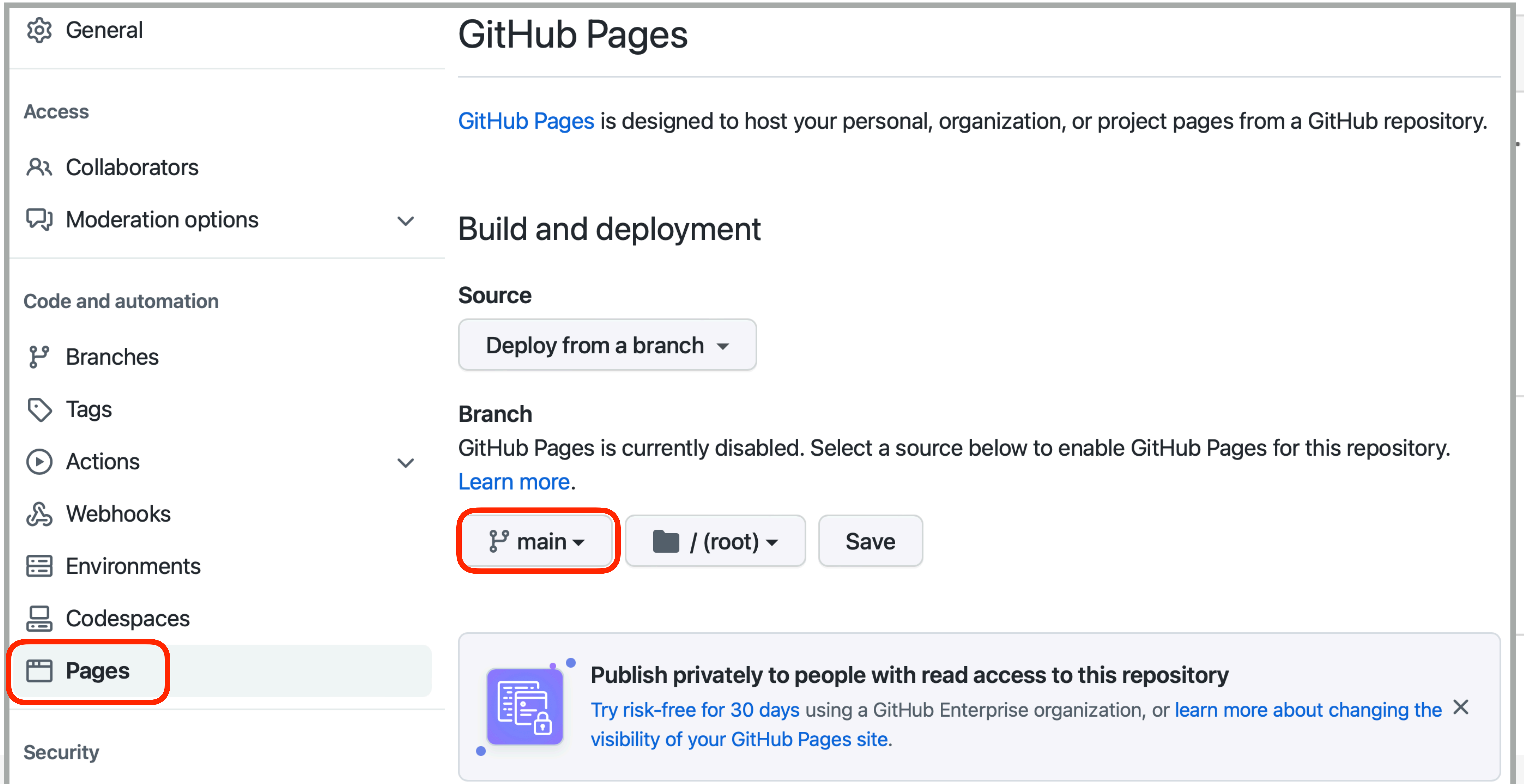


2

# Theme chooser

Scroll down to the **GitHub Pages** section. ~~Press **Choose a theme**.~~

**And set the Source to “main branch"**



## GitHub Pages

GitHub Pages is designed to host your personal, organization, or project pages from a GitHub repository.

### Build and deployment

**Source**

Deploy from a branch

**Branch**

GitHub Pages is currently disabled. Select a source below to enable GitHub Pages for this repository. Learn more.

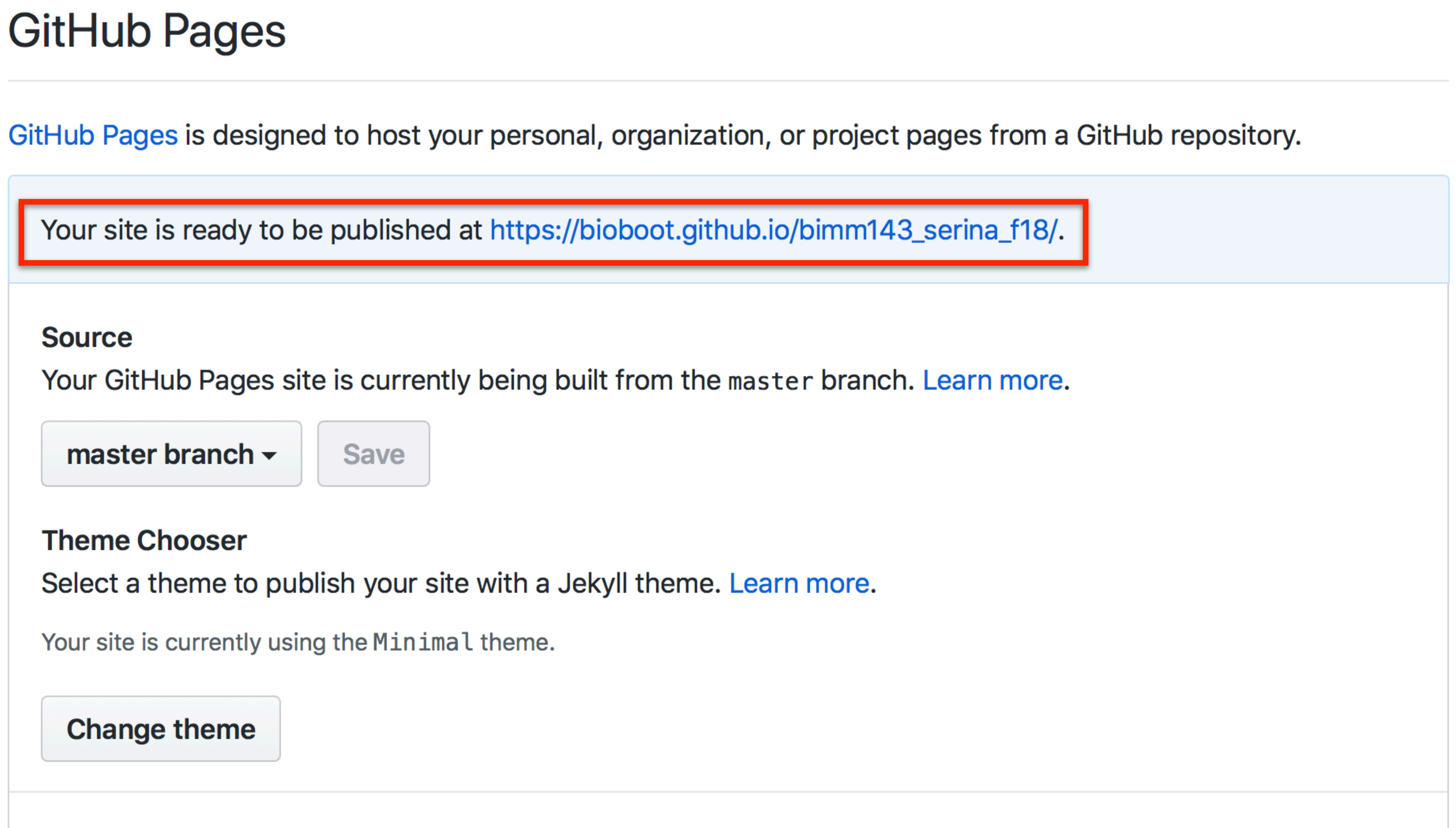
**Publish privately to people with read access to this repository**

Try risk-free for 30 days using a GitHub Enterprise organization, or learn more about changing the visibility of your GitHub Pages site.

4

# Side-note:

**Scroll down again to the GitHub Pages section to find the link to your new website. Open this link in a New Tab of your browser:**



## GitHub Pages

GitHub Pages is designed to host your personal, organization, or project pages from a GitHub repository.

Your site is ready to be published at https://bioboot.github.io/bimm143_serina_f18/.

**Source**

Your GitHub Pages site is currently being built from the `master` branch. Learn more.

**Theme Chooser**

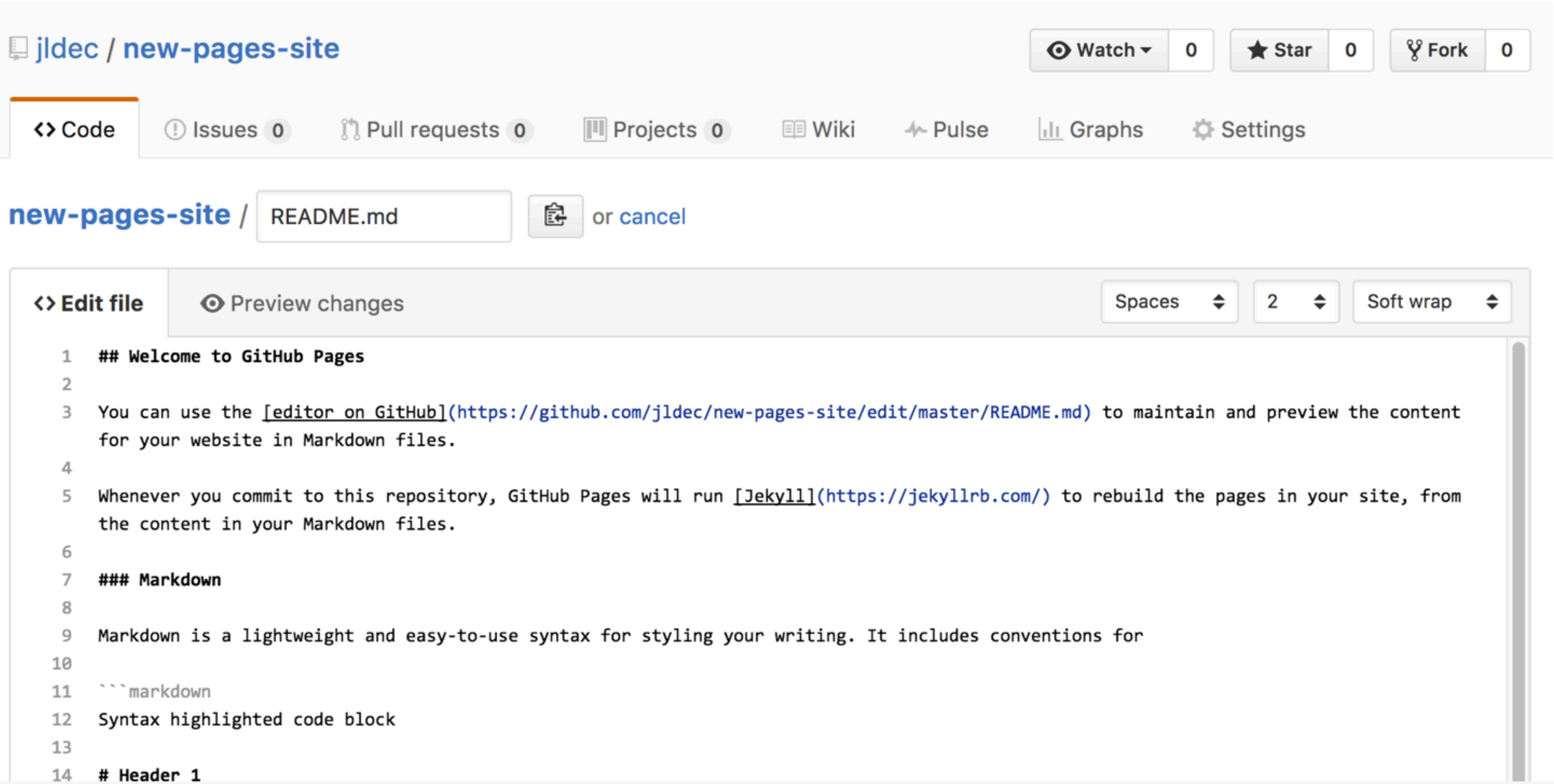
Select a theme to publish your site with a Jekyll theme. Learn more.

Your site is currently using the `Minimal` theme.

4

# Edit content

**Back on the repository main page use the GitHub online editor to add content. In particular, add links to each classes .MD file**

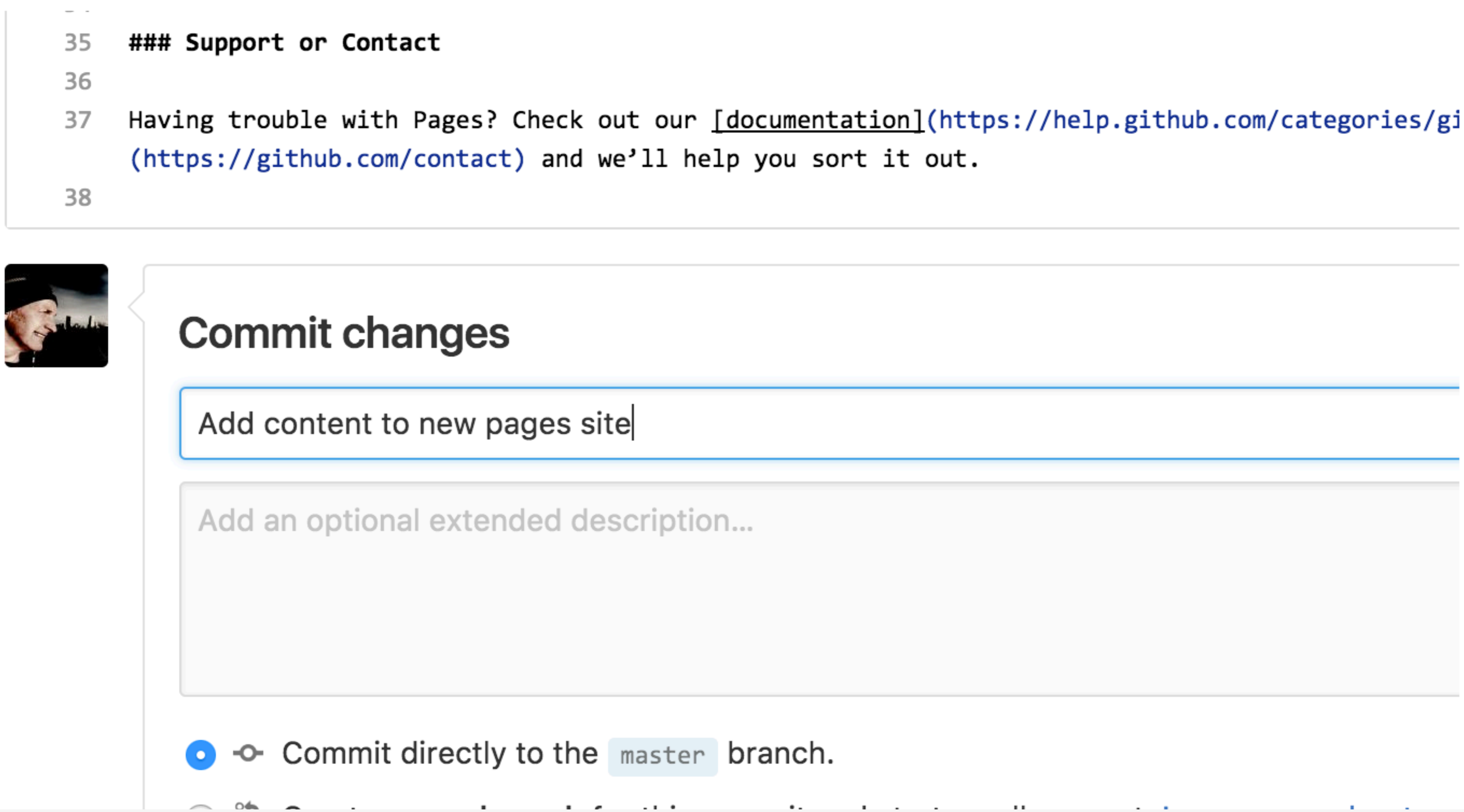


~~~
## Welcome to GitHub Pages

You can use the [editor on GitHub](https://github.com/jldec/new-pages-site/edit/master/README.md) to maintain and preview the content for your website in Markdown files.

Whenever you commit to this repository, GitHub Pages will run [Jekyll](https://jekyllrb.com/) to rebuild the pages in your site, from the content in your Markdown files.

### Markdown

Markdown is a lightweight and easy-to-use syntax for styling your writing. It includes conventions for

```markdown
Syntax highlighted code block

# Header 1
~~~

5

# Commit

Enter a commit comment and click on **Commit changes** below the editor.



```
Support or Contact

Having trouble with Pages? Check out our [documentation](https://help.github.com/categories/gi
(https://github.com/contact) and we’ll help you sort it out.

```

**Commit changes**

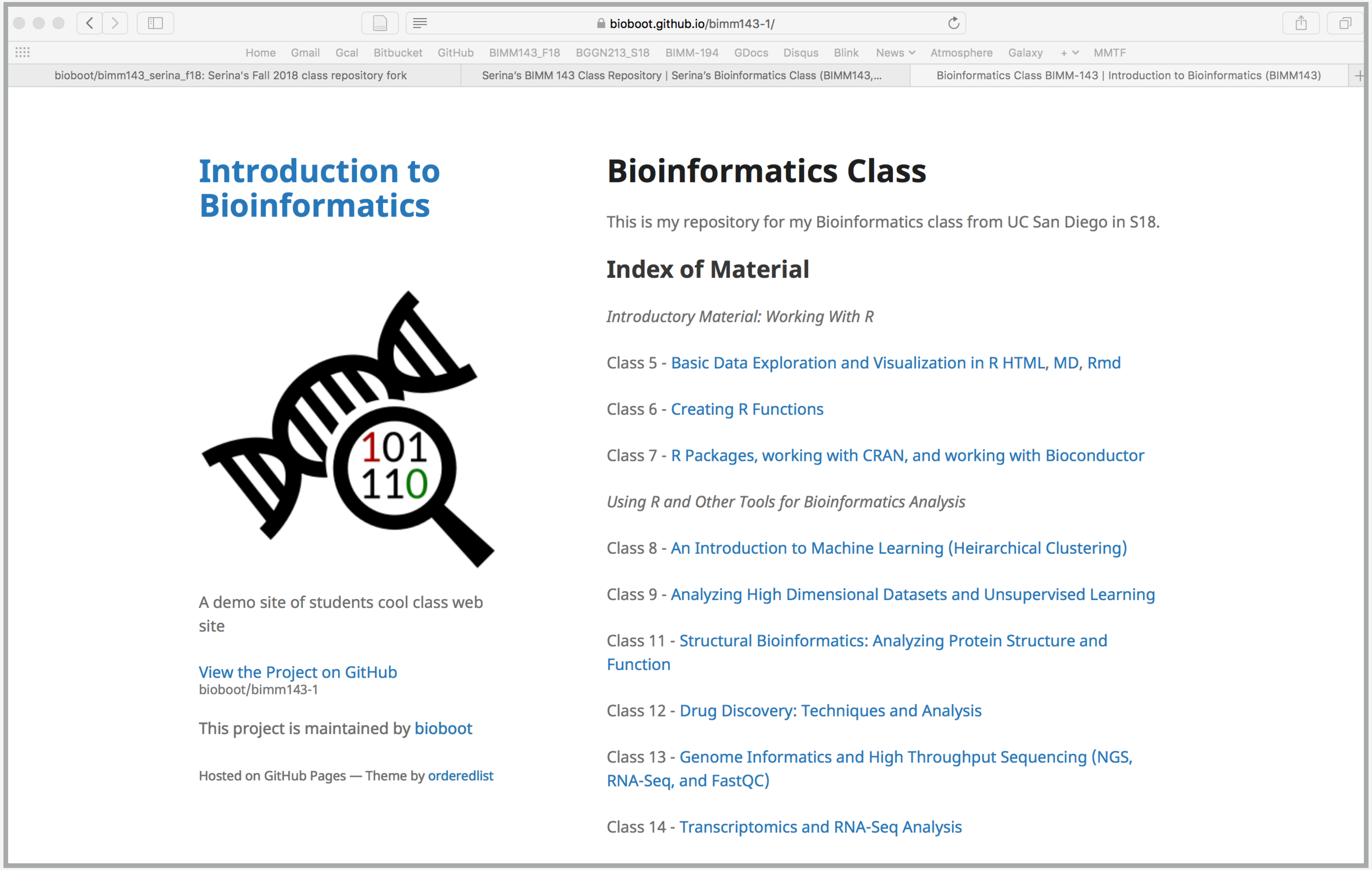
Add content to new pages site

● Commit directly to the `master` branch.

6

# …and you're done!

Fire up a browser and go to **http://*username*.github.io/*repository***.



# Introduction to Bioinformatics

A demo site of students cool class web site

View the Project on GitHub
bioboot/bimm143-1

This project is maintained by bioboot

Hosted on GitHub Pages — Theme by orderedlist

# Bioinformatics Class

This is my repository for my Bioinformatics class from UC San Diego in S18.

## Index of Material

*Introductory Material: Working With R*

Class 5 - Basic Data Exploration and Visualization in R HTML, MD, Rmd

Class 6 - Creating R Functions

Class 7 - R Packages, working with CRAN, and working with Bioconductor

*Using R and Other Tools for Bioinformatics Analysis*

Class 8 - An Introduction to Machine Learning (Heirarchical Clustering)

Class 9 - Analyzing High Dimensional Datasets and Unsupervised Learning

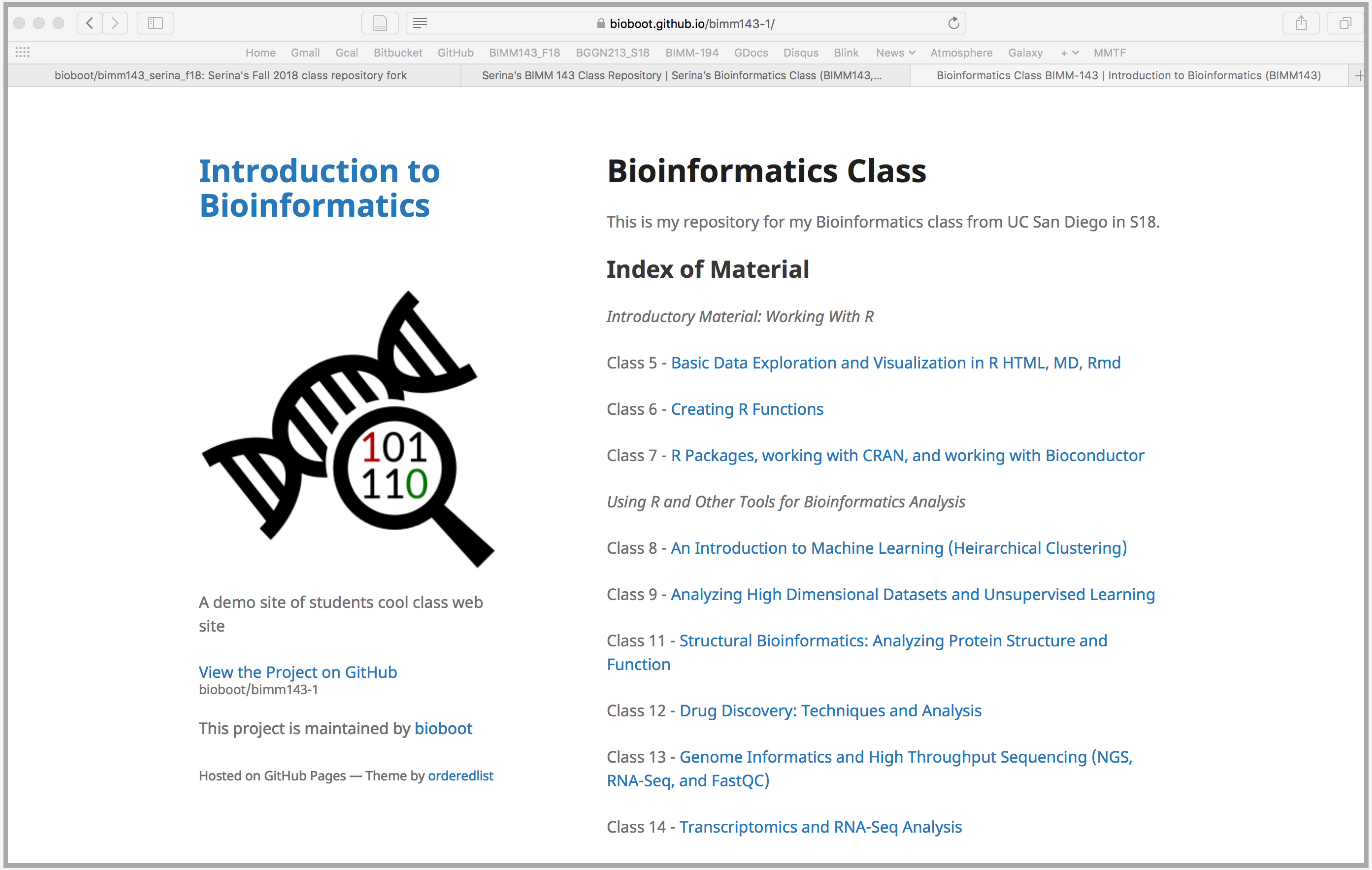
Class 11 - Structural Bioinformatics: Analyzing Protein Structure and Function

Class 12 - Drug Discovery: Techniques and Analysis

Class 13 - Genome Informatics and High Throughput Sequencing (NGS, RNA-Seq, and FastQC)

Class 14 - Transcriptomics and RNA-Seq Analysis

**Here I: (1) forked Serina’s Repo, (2) Chose the “minimal” theme, (3) Edited _config.yml (adding logo and title)**



## Introduction to Bioinformatics

A demo site of students cool class web site

View the Project on GitHub
bioboot/bimm143-1

This project is maintained by bioboot

Hosted on GitHub Pages — Theme by orderedlist

# Bioinformatics Class

This is my repository for my Bioinformatics class from UC San Diego in S18.

## Index of Material

*Introductory Material: Working With R*

Class 5 - Basic Data Exploration and Visualization in R HTML, MD, Rmd

Class 6 - Creating R Functions

Class 7 - R Packages, working with CRAN, and working with Bioconductor

*Using R and Other Tools for Bioinformatics Analysis*

Class 8 - An Introduction to Machine Learning (Heirarchical Clustering)

Class 9 - Analyzing High Dimensional Datasets and Unsupervised Learning

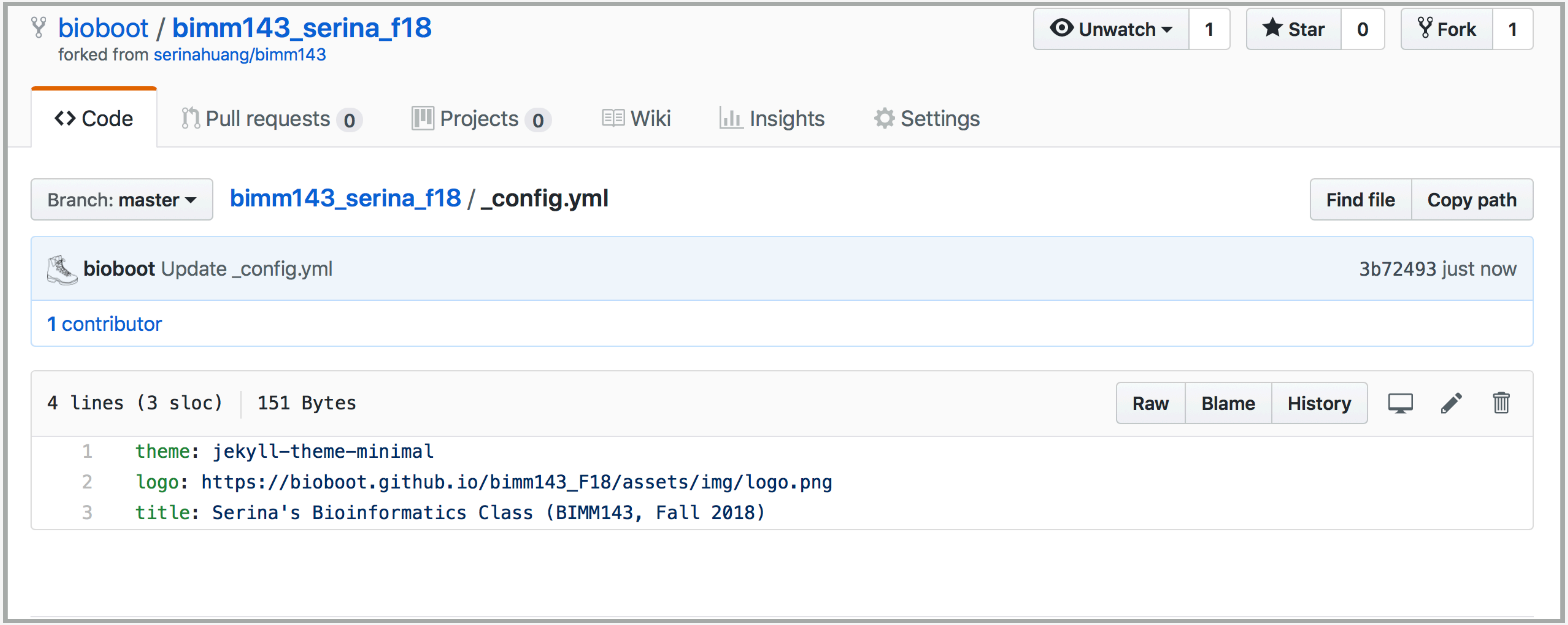
Class 11 - Structural Bioinformatics: Analyzing Protein Structure and Function

Class 12 - Drug Discovery: Techniques and Analysis

Class 13 - Genome Informatics and High Throughput Sequencing (NGS, RNA-Seq, and FastQC)

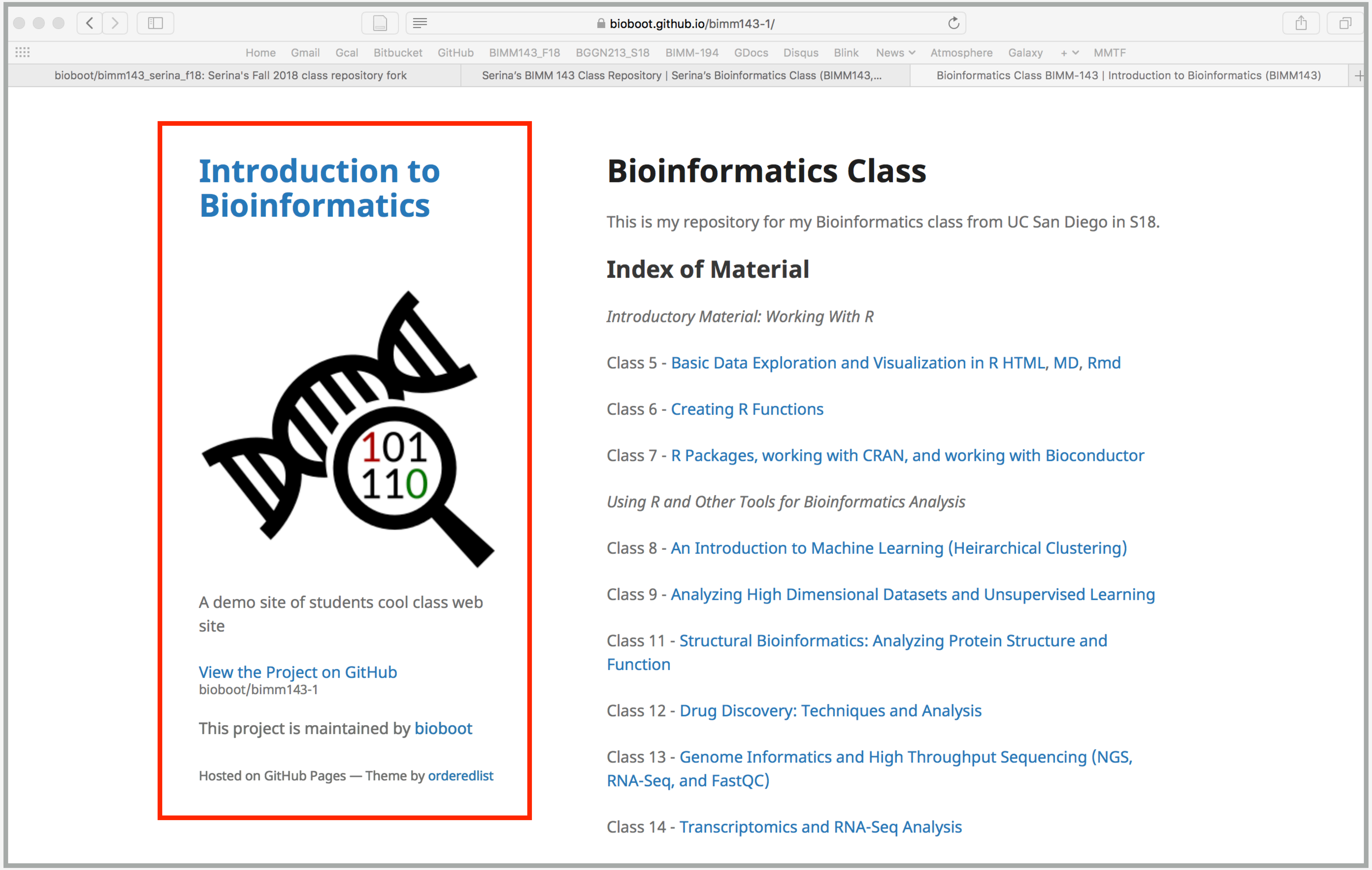
Class 14 - Transcriptomics and RNA-Seq Analysis

**Here I: (1) forked Serina's Repo, (2) Chose the "minimal" theme, (3) Edited _config.yml (adding logo and title)**



```yaml
theme: jekyll-theme-minimal
logo: https://bioboot.github.io/bimm143_F18/assets/img/logo.png
title: Serina's Bioinformatics Class (BIMM143, Fall 2018)
```

Here I: (1) Chose the “minimal” theme, (3) Edited _config.yml (adding logo and title), (4) Edited README.md



Introduction to Bioinformatics

A demo site of students cool class web site

View the Project on GitHub
bioboot/bimm143-1

This project is maintained by bioboot

Hosted on GitHub Pages — Theme by orderedlist

Bioinformatics Class

This is my repository for my Bioinformatics class from UC San Diego in S18.

Index of Material

Introductory Material: Working With R

Class 5 - Basic Data Exploration and Visualization in R HTML, MD, Rmd

Class 6 - Creating R Functions

Class 7 - R Packages, working with CRAN, and working with Bioconductor

Using R and Other Tools for Bioinformatics Analysis

Class 8 - An Introduction to Machine Learning (Heirarchical Clustering)

Class 9 - Analyzing High Dimensional Datasets and Unsupervised Learning

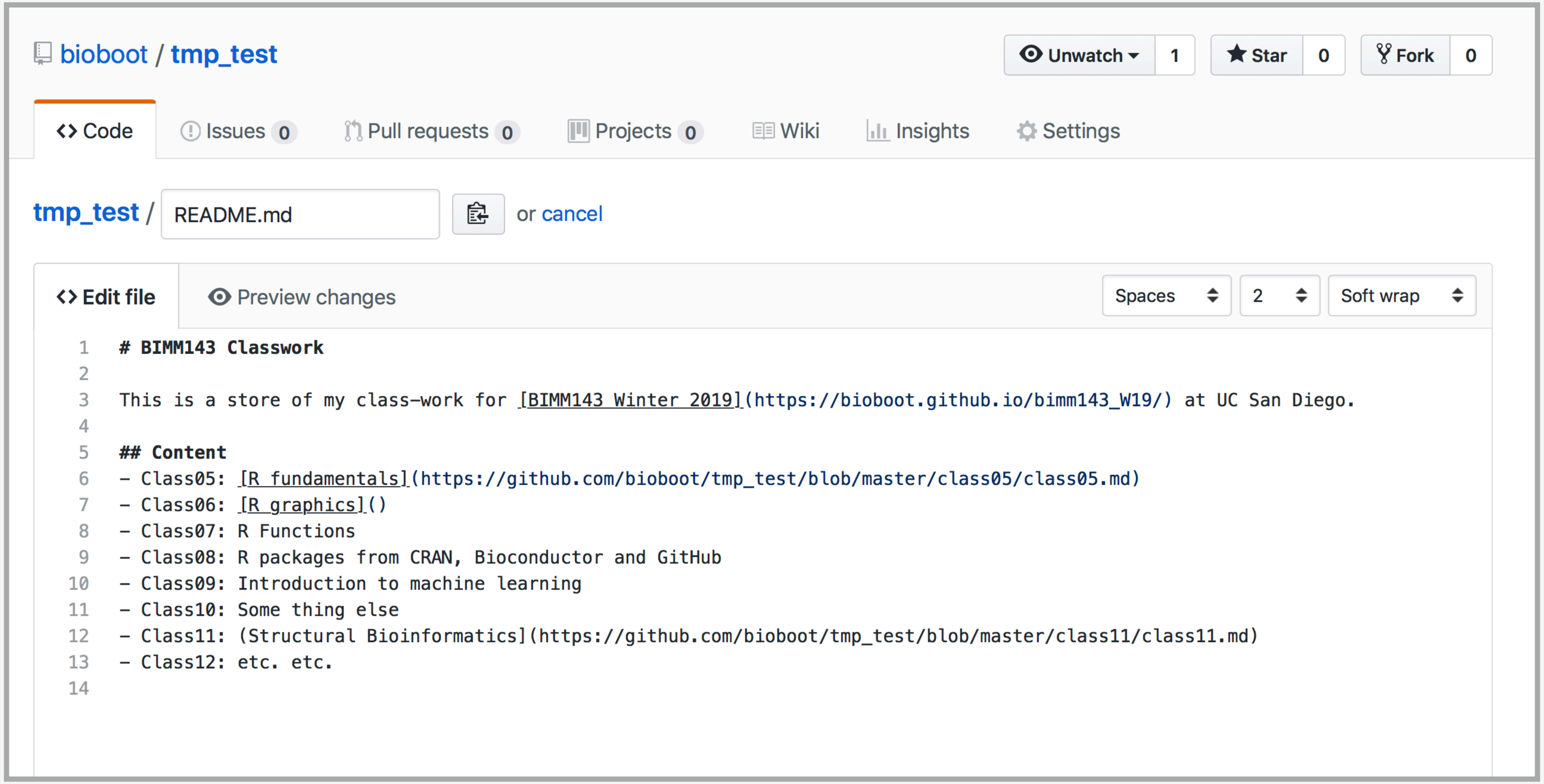
Class 11 - Structural Bioinformatics: Analyzing Protein Structure and Function

Class 12 - Drug Discovery: Techniques and Analysis

Class 13 - Genome Informatics and High Throughput Sequencing (NGS, RNA-Seq, and FastQC)

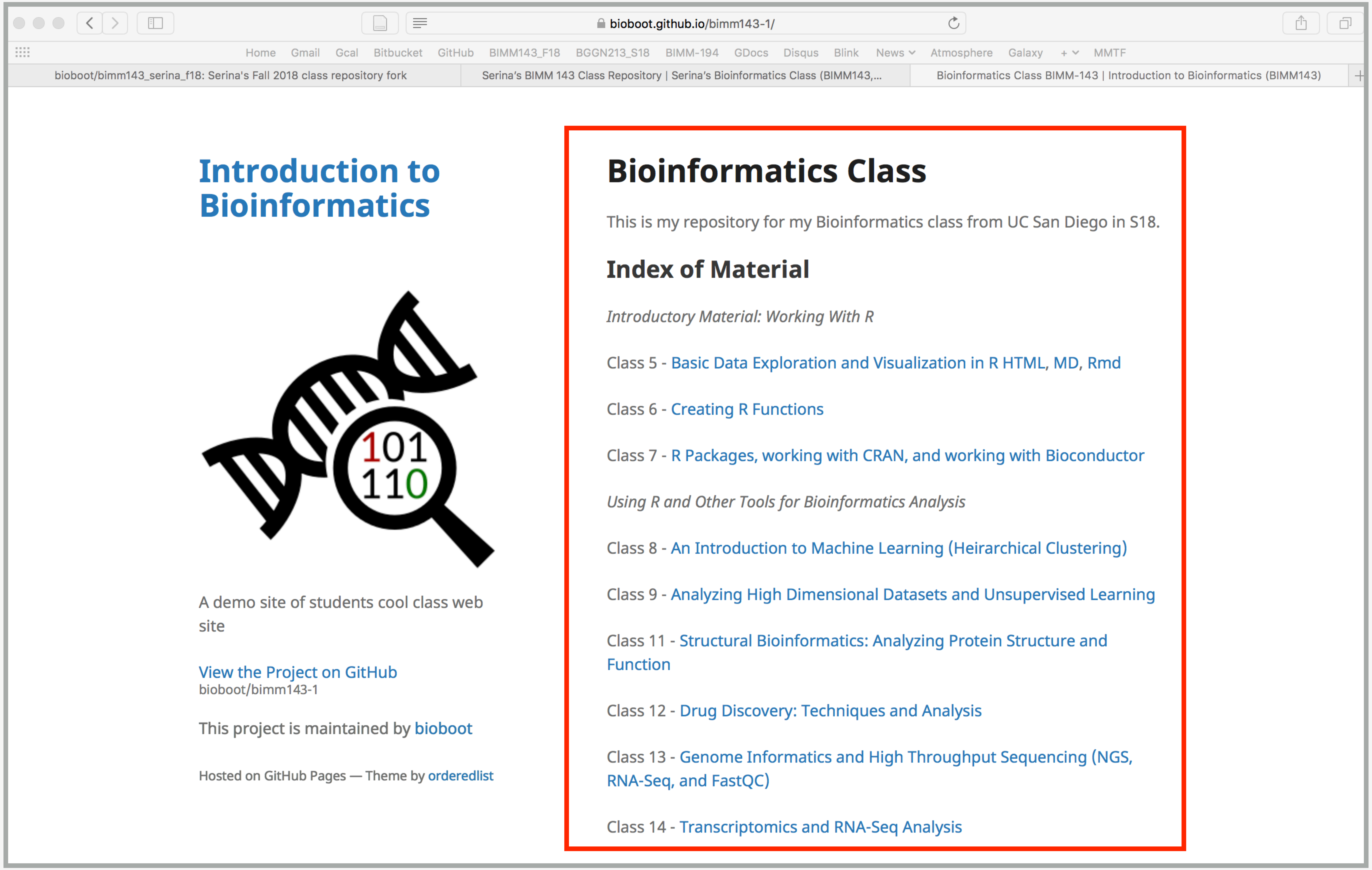
Class 14 - Transcriptomics and RNA-Seq Analysis

Here I: (1) Chose the “minimal” theme, (3) Edited _config.yml (adding logo and title), (4) Edited README.md



```
# BIMM143 Classwork

This is a store of my class-work for [BIMM143 Winter 2019](https://bioboot.github.io/bimm143_W19/) at UC San Diego.

## Content
- Class05: [R fundamentals](https://github.com/bioboot/tmp_test/blob/master/class05/class05.md)
- Class06: [R graphics]()
- Class07: R Functions
- Class08: R packages from CRAN, Bioconductor and GitHub
- Class09: Introduction to machine learning
- Class10: Some thing else
- Class11: (Structural Bioinformatics](https://github.com/bioboot/tmp_test/blob/master/class11/class11.md)
- Class12: etc. etc.

```

Here I: (1) Chose the “minimal” theme, (3) Edited _config.yml (adding logo and title), (4) Edited README.md



Introduction to Bioinformatics

A demo site of students cool class web site

View the Project on GitHub
bioboot/bimm143-1

This project is maintained by bioboot

Hosted on GitHub Pages — Theme by orderedlist

Bioinformatics Class

This is my repository for my Bioinformatics class from UC San Diego in S18.

Index of Material

Introductory Material: Working With R

Class 5 - Basic Data Exploration and Visualization in R HTML, MD, Rmd

Class 6 - Creating R Functions

Class 7 - R Packages, working with CRAN, and working with Bioconductor

Using R and Other Tools for Bioinformatics Analysis

Class 8 - An Introduction to Machine Learning (Heirarchical Clustering)

Class 9 - Analyzing High Dimensional Datasets and Unsupervised Learning

Class 11 - Structural Bioinformatics: Analyzing Protein Structure and Function

Class 12 - Drug Discovery: Techniques and Analysis

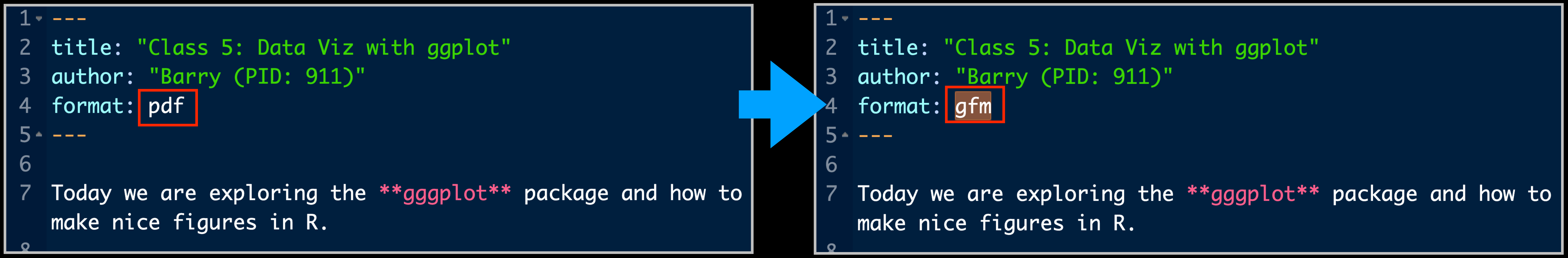
Class 13 - Genome Informatics and High Throughput Sequencing (NGS, RNA-Seq, and FastQC)

Class 14 - Transcriptomics and RNA-Seq Analysis

Here I: (1) Chose the "minimal" theme, (3) Edited _config.yml (adding logo and title), (4) Edited README.md

Formats for Portfolio Lab Reports

- Your Quarto lab reports (input for quarto render) likely produced pdf format outputs:
- Ideally we want something nicer for online viewing. Enter the gfm format (short for GitHub Flavored Markdown).
- We can edit our YMAL header and quarto render, git add, git commit and git push to test this out on GitHub… <demo>



```
---
title: "Class 5: Data Viz with ggplot"
author: "Barry (PID: 911)"
format: pdf
---

Today we are exploring the **ggplot** package and how to
make nice figures in R.
```

```
---
title: "Class 5: Data Viz with ggplot"
author: "Barry (PID: 911)"
format: gfm
---

Today we are exploring the **ggplot** package and how to
make nice figures in R.
```

Formats: HTML, GFM or htmlpreview

- There are three options to get “nice” lab reports on GitHub Pages

1. Link to a HTML file e.g. `[class05](class05/class05.html)`

2. Create and link to a special “gfm” format (GitHub Flavored Markdown)”:

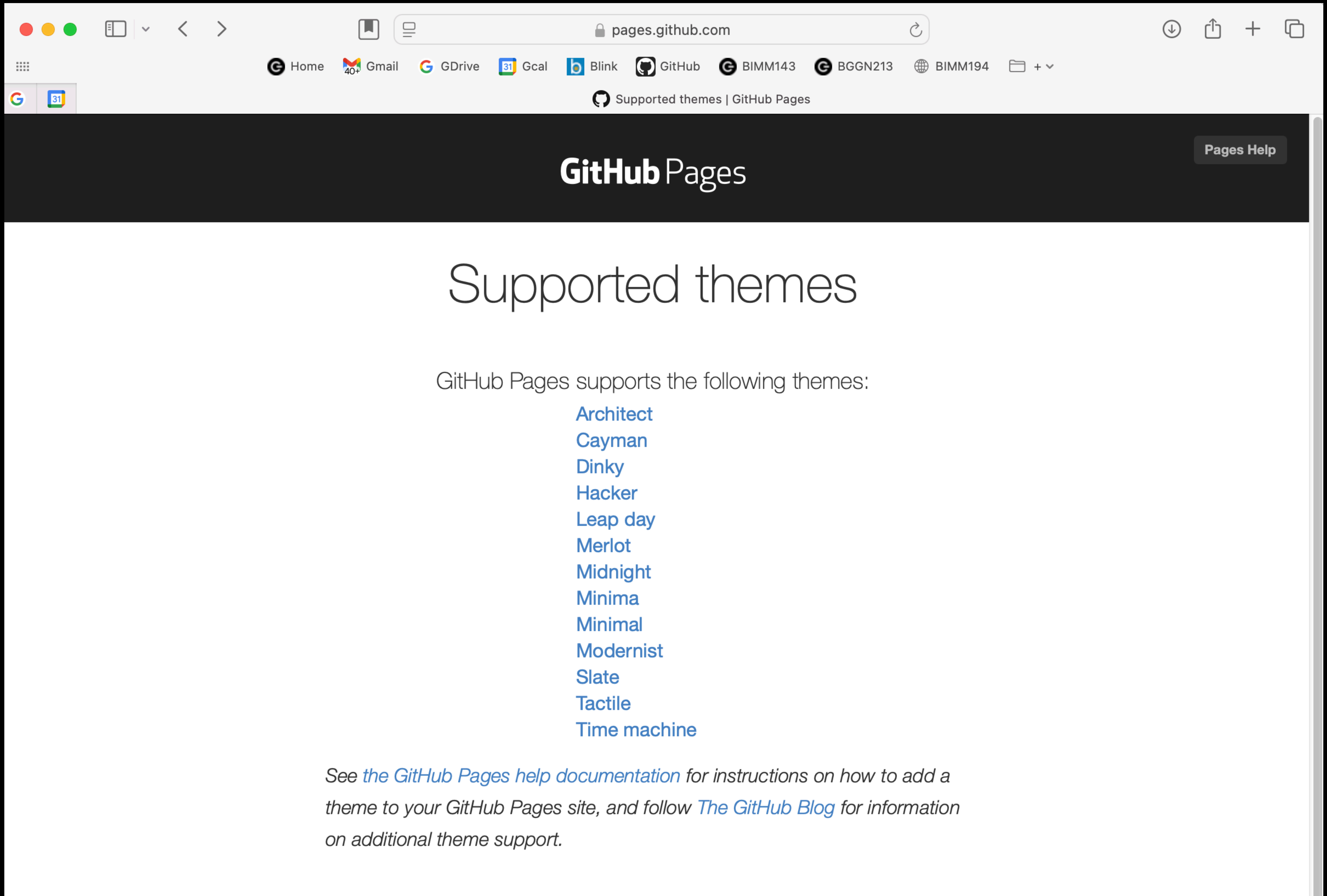
```
quarto render class05.qmd --to gfm
```

3. Or you can prepend https://tmlpreview.github.io/?:
    - Go to the HTML file in the GitHub repository
    - Click on the "Raw" button to get the raw file URL
    - Construct your link like so: https://htmlpreview.github.io/?[raw-file-url]
    - This will work on both the repo README and the GitHub website…

Why Vim Opens During `git pull`

- You edited a file locally and also on GitHub.com.
- Git detects the two versions don't match, so it needs to merge them.
- Git opens Vim to confirm the merge with a commit message.
- To finish the merge: type :wq and press Enter.
- To avoid this: Always run git pull before editing locally.

https://pages.github.com/themes/



Supported themes

GitHub Pages supports the following themes:

- Architect
- Cayman
- Dinky
- Hacker
- Leap day
- Merlot
- Midnight
- Minima
- Minimal
- Modernist
- Slate
- Tactile
- Time machine

See the GitHub Pages help documentation for instructions on how to add a theme to your GitHub Pages site, and follow The GitHub Blog for information on additional theme support.

Thank you very much!

Please do fill out your CAPs evaluation (Link!) if you get a change.
It is important to the courses we offer in the future and how we teach them!

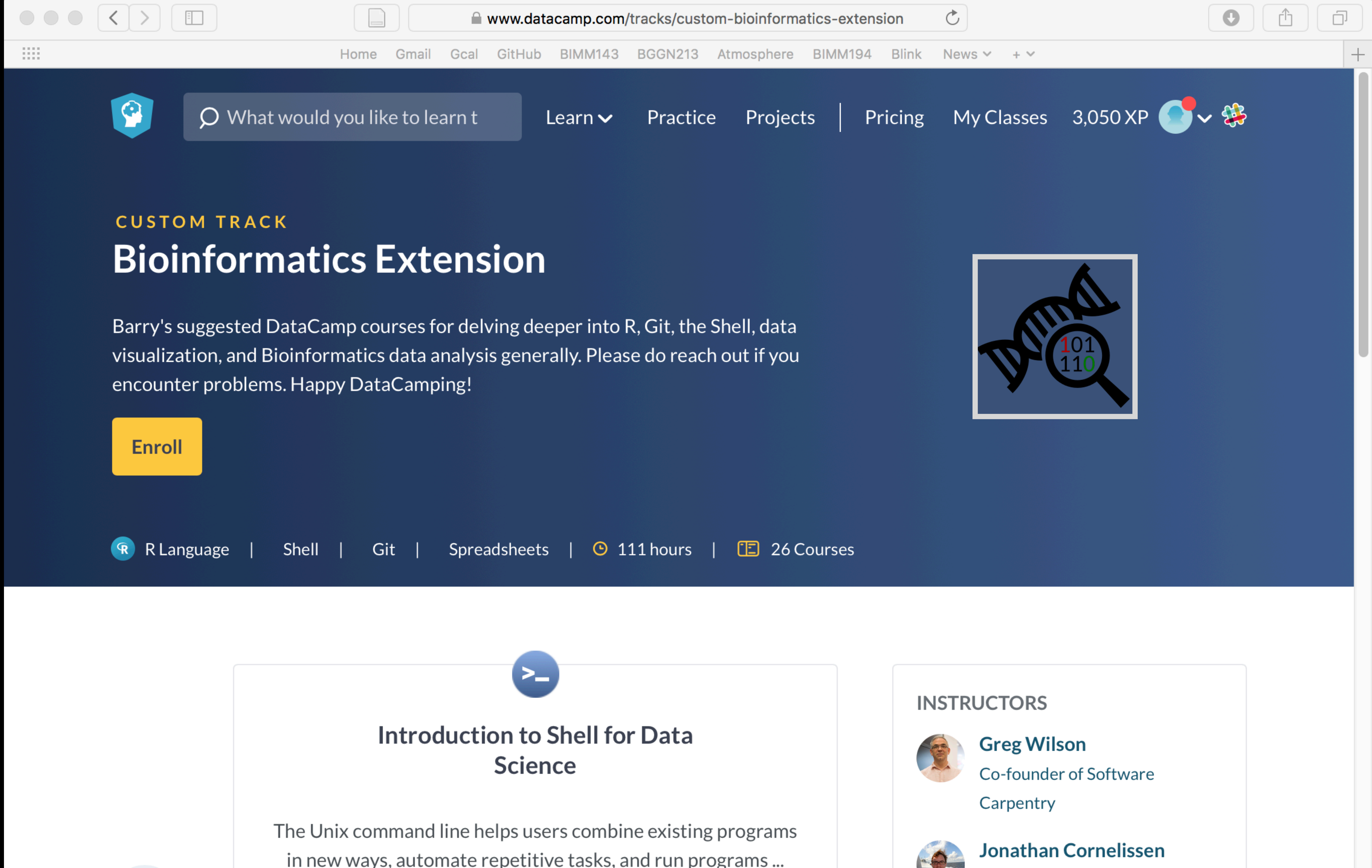
Thank you very much!

Please do fill out your CAPs evaluation (Link!) if you get a change.
It is important to the courses we offer in the future and how we teach them!

Post to GradeScope your GitHub Pages portfolio URL to GradeScope!

Going Further With DataCamp

https://bioboot.github.io/bimm143_F19/class-material/datacamp_extras.pdf



CUSTOM TRACK

Bioinformatics Extension

Barry's suggested DataCamp courses for delving deeper into R, Git, the Shell, data visualization, and Bioinformatics data analysis generally. Please do reach out if you encounter problems. Happy DataCamping!

R Language | Shell | Git | Spreadsheets | 111 hours | 26 Courses

Introduction to Shell for Data Science

The Unix command line helps users combine existing programs in new ways, automate repetitive tasks, and run programs ...

INSTRUCTORS

Greg Wilson
Co-founder of Software Carpentry

Jonathan Cornelissen

Announcement!

Bonus:

Bioinformatics & Genomics in Industry

Live Stream Video

Enjoy a set of short open ended guest lectures from leading genomic scientists at **Illumina Inc.**, **Synthetic Genomics Inc.**, and the **La Jolla Institute for Allergy and Immunology**. Feel free to contact these scientists for networking and to have your questions about industry careers in Bioinformatics and Genomics answered.

GitHub Copilot

Hands-on Lab Session

Barry Grant

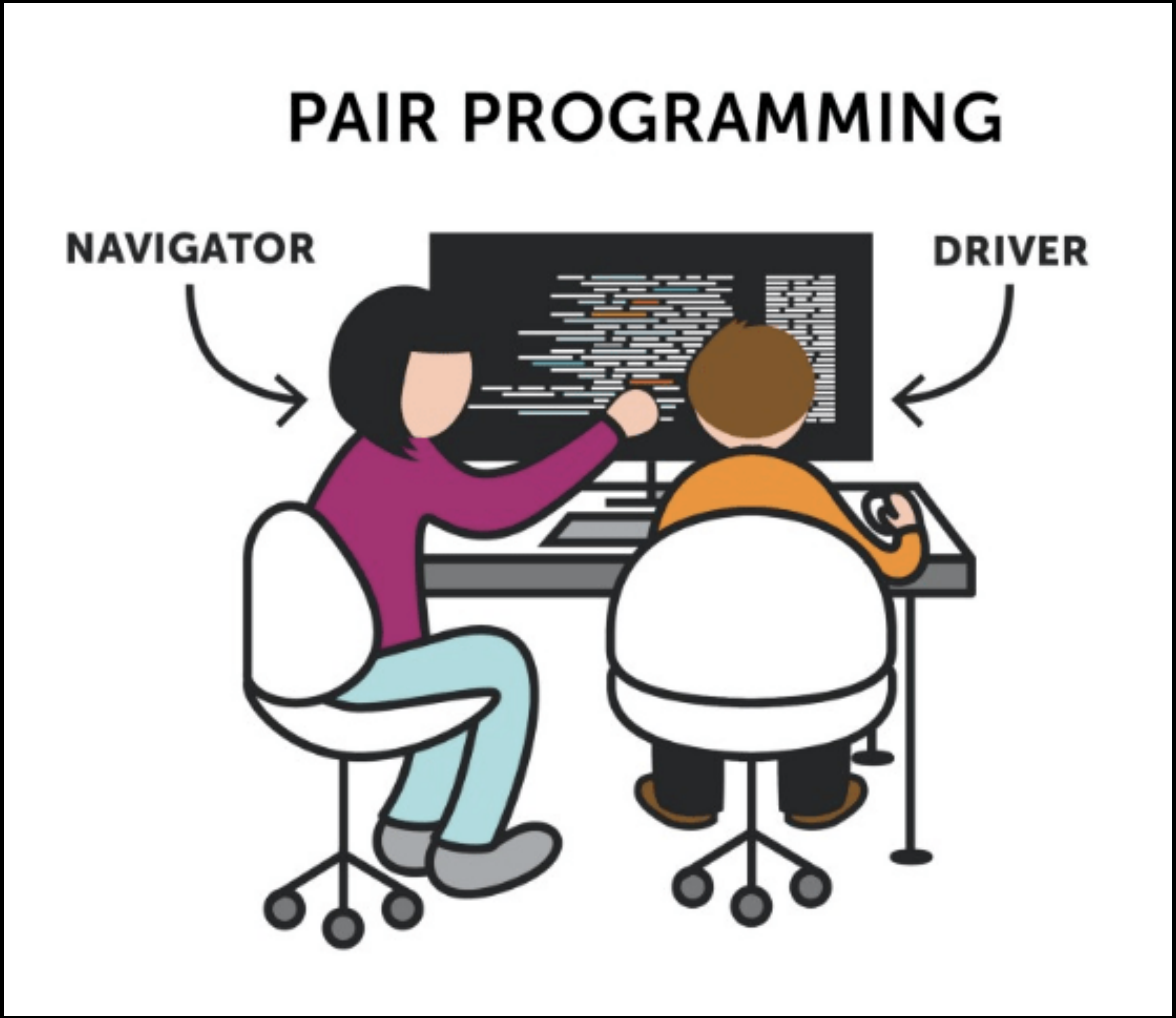
UC San Diego

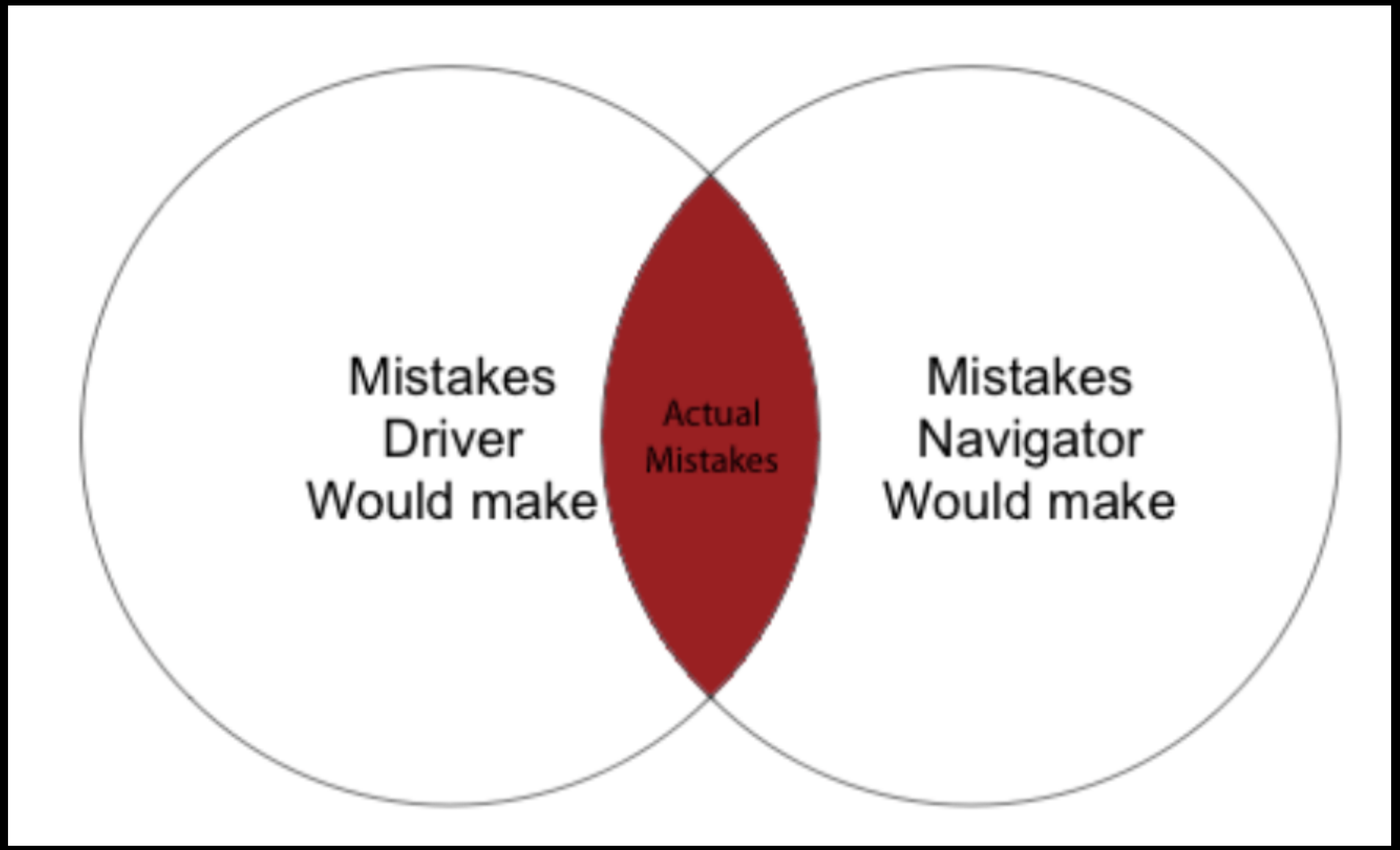
http://thegrantlab.org

GitHub Copilot

GitHub Copilot

Your AI “pair programmer”





Source: https://unruly.co/blog/article/2019/08/27/what-is-pair-programming/

GitHub Copilot

AI powered add-on that aims to give helpful suggestions when writing code or documentation

AI pair programmer

GitHub Copilot is an AI pair programmer that offers autocomplete-style suggestions and real-time hints for the code you are writing by providing suggestions as “ghost text” based on the context of the surrounding code

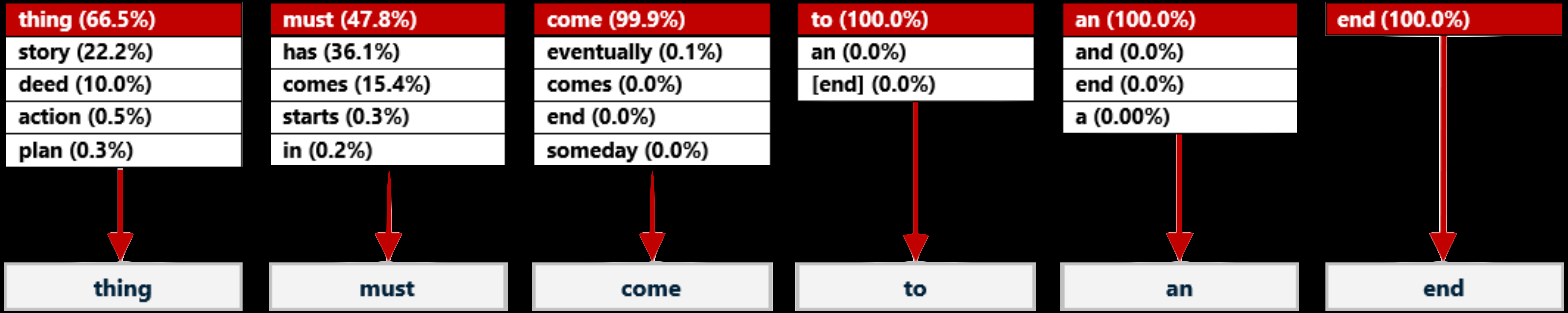
What can GitHub Copilot do?

- It provides autocomplete suggestions while you code.
- You can get suggestions by typing code or describing it in natural language (i.e. code comments or markdown text).
- Copilot analyzes your file (and related files), offering suggestions in your editor.
- It uses OpenAI Codex to derive context from your code and comments, and then suggests new lines or entire functions.

Generative AI Models

Generative AI

- For text generation, Generative AI just wants to predict the next word/token/string!
- I might ask ChatGPT: > “Complete the sentence every good…”



thing (66.5%)
story (22.2%)
deed (10.0%)
action (0.5%)
plan (0.3%)

must (47.8%)
has (36.1%)
comes (15.4%)
starts (0.3%)
in (0.2%)

come (99.9%)
eventually (0.1%)
comes (0.0%)
end (0.0%)
someday (0.0%)

to (100.0%)
an (0.0%)
[end] (0.0%)

an (100.0%)
and (0.0%)
end (0.0%)
a (0.00%)

end (100.0%)

thing must come to an end

OpenAI

Generative AI

GPT-3

Prompt:

Write a tagline for an ice cream shop.

Response:

We serve up smiles with every scoop!

Codex

Prompt:

Table customers, columns = [CustomerId, FirstName, LastName, Company, Address, City, State, Country, PostalCode]

Create a SQL query for all customers in Texas named Jane
query =

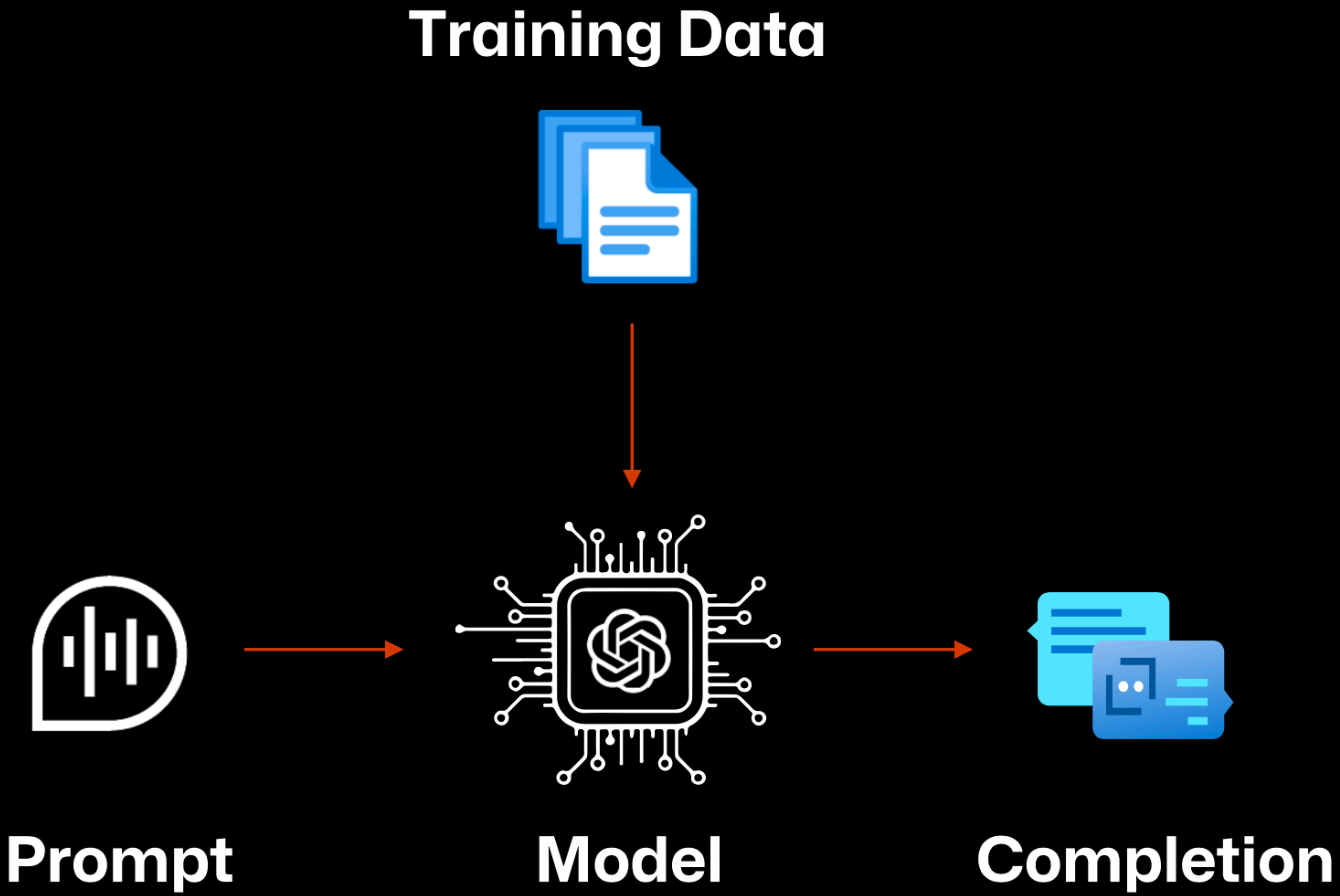
Response:

```
SELECT *
FROM customers
WHERE State = 'TX' AND FirstName
= 'Jane'
```

DALL·E

Prompt: A white Siamese cat

Response:



Training Data

Prompt

Model

Completion

Generative AI can:

Generate text, images and code

Different models are trained on different corpuses, depending on the application.

Generate “humanlike” output

What is a likely continuation of the prompt, given the training data?

Extract information

The continuation is likely to be similar to text frequently represented in the training data.

Create novel content

Text, images and code not contained in its training set. Translations. “Creative” works.

Generative AI is not:

Intelligent

It’s just a predictive system, designed to give a likely continuation of the prompt given the training data.

Deterministic

Run the same prompt. Get back a different response (probably).

Trustworthy

It can “hallucinate” facts and confidently assert them to be true.

TODO

Generative AI does not:

○ **Learn**

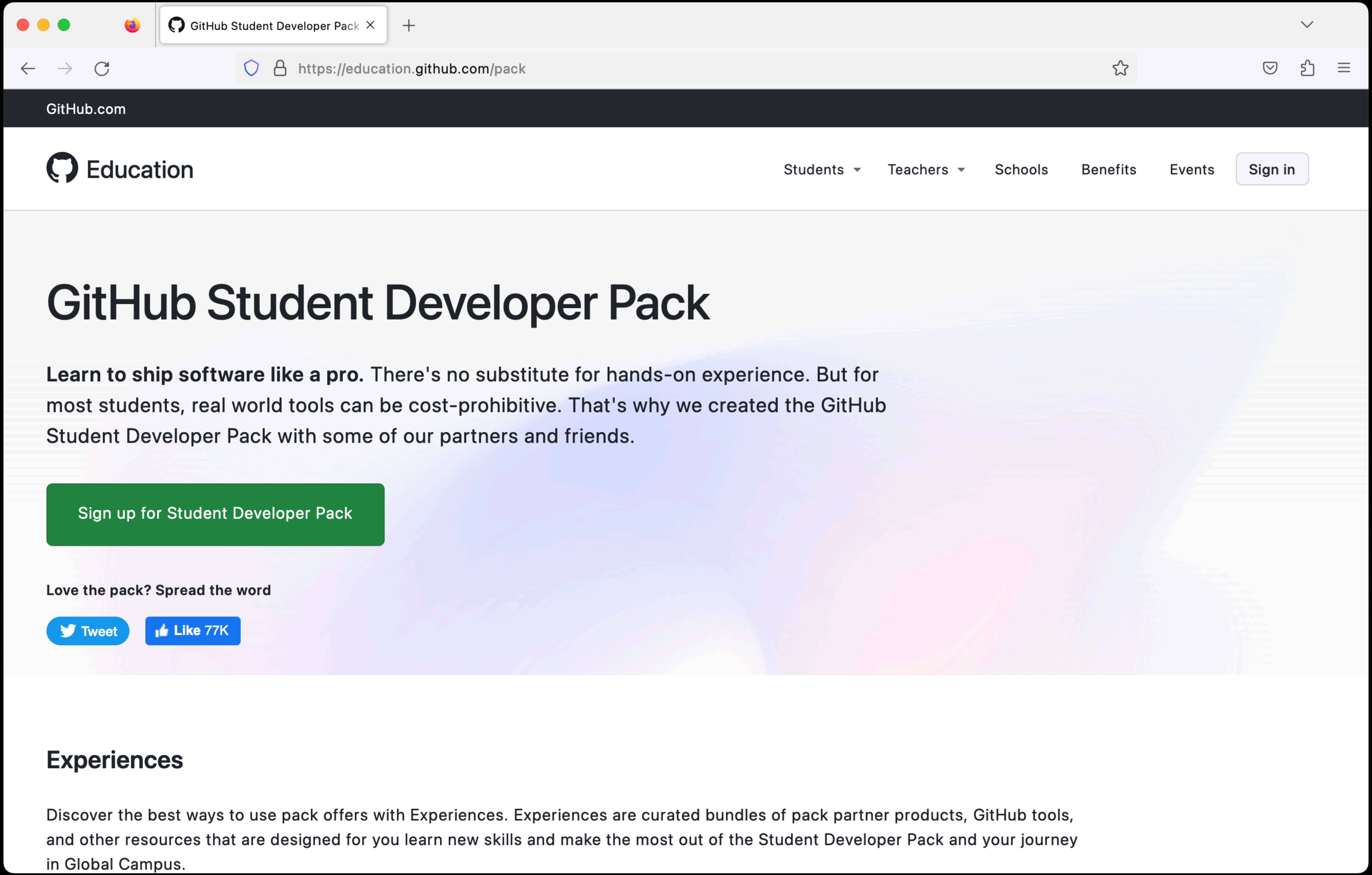
The model is fixed at the time of its training.

○ **Contain all of the information of its training set**

Think: a blurry jpeg of its training data.

○ **Include verbatim copies of its training data**

But it can generate stuff that looks like it.



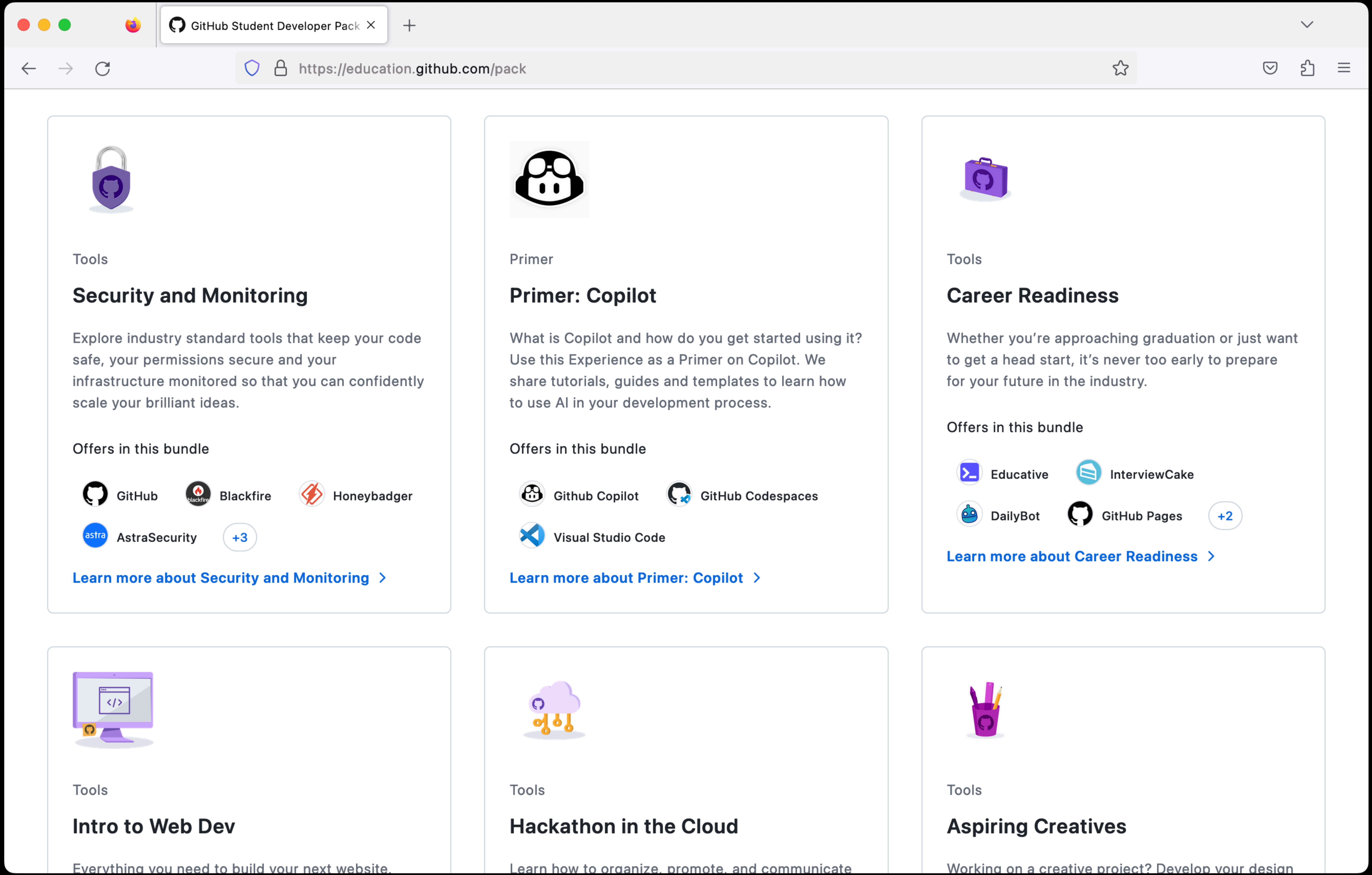
GitHub Student Developer Pack

Learn to ship software like a pro. There's no substitute for hands-on experience. But for most students, real world tools can be cost-prohibitive. That's why we created the GitHub Student Developer Pack with some of our partners and friends.

Love the pack? Spread the word

Experiences

Discover the best ways to use pack offers with Experiences. Experiences are curated bundles of pack partner products, GitHub tools, and other resources that are designed for you learn new skills and make the most out of the Student Developer Pack and your journey in Global Campus.



Tools

Security and Monitoring

Explore industry standard tools that keep your code safe, your permissions secure and your infrastructure monitored so that you can confidently scale your brilliant ideas.

Offers in this bundle

Primer

Primer: Copilot

What is Copilot and how do you get started using it? Use this Experience as a Primer on Copilot. We share tutorials, guides and templates to learn how to use AI in your development process.

Offers in this bundle

Tools

Career Readiness

Whether you're approaching graduation or just want to get a head start, it's never too early to prepare for your future in the industry.

Offers in this bundle

Tools

Intro to Web Dev

Everything you need to build your next website.

Tools

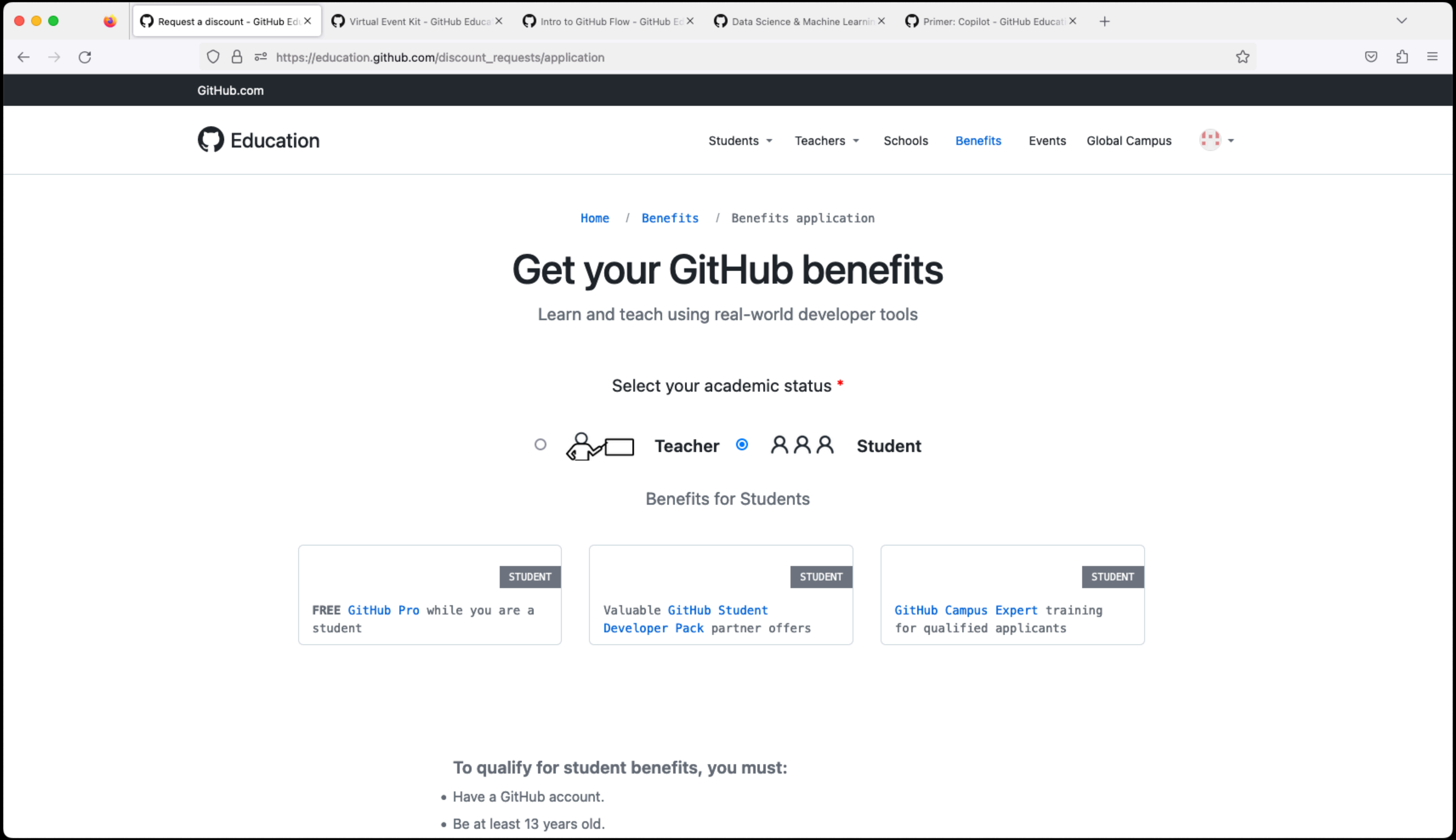
Hackathon in the Cloud

Learn how to organize, promote, and communicate

Tools

Aspiring Creatives

Working on a creative project? Develop your design



Home / Benefits / Benefits application

Get your GitHub benefits

Learn and teach using real-world developer tools

Select your academic status *

○ Teacher ● Student

Benefits for Students

STUDENT

FREE GitHub Pro while you are a student

STUDENT

Valuable GitHub Student Developer Pack partner offers

STUDENT

GitHub Campus Expert training for qualified applicants

To qualify for student benefits, you must:

- Have a GitHub account.
- Be at least 13 years old.

Quick Start Guide

https://docs.github.com/en/copilot/quickstart

Additional Learning Resources

- **Video series from GitHub:**
    - https://learn.microsoft.com/en-us/shows/introduction-to-github-copilot/
- **Blog post series:**
    - https://github.blog/2023-06-20-how-to-write-better-prompts-for-github-copilot/
- **RStudio Docs:**
    - https://docs.posit.co/ide/user/ide/guide/tools/copilot.html

The End!

http://thegrantlab.org

Thank You!

http://thegrantlab.org