

BST 219

Core Principles of Data Science

Lecture 2: Introduction RStudio, RMarkdown & GitHub

September 5, 2024



Recipe of the Day!

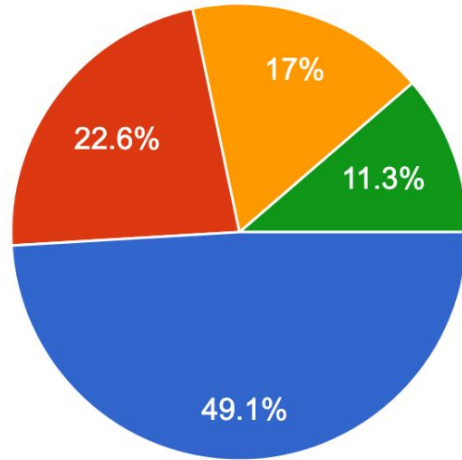
Apple Scones with Maple Cinnamon Glaze



Preliminary Survey Results

How long have you been programming?

53 responses

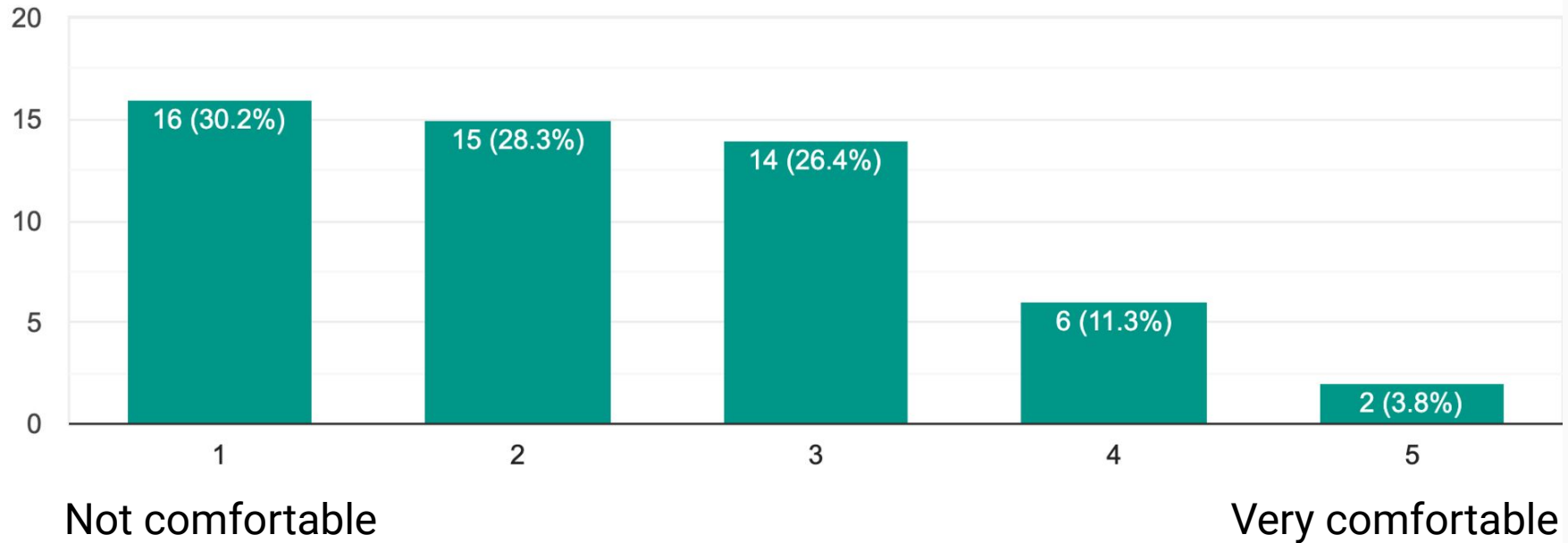


- Less than 6 months
- 6 months - 1 year
- 1-3 years
- Over 3 years

Preliminary Survey Results

Overall, how comfortable are you with programming?

53 responses



Preliminary Survey Results

What are you hoping to get out of this course?

Getting **comfortable coding in R** and with coding in general

Better understanding of data science for the **entire research pipeline**

Become a **coding master!** Mostly to understand the **dialogue/script that appears on the console** and understanding R functions, **how to work with raw data** that isn't already finetuned, how to **problem solve when errors occur and regulating frustration.**

Better at data cleaning, intro to ML, feel better about **generating files to upload as images for manuscript instead of just screenshotting the image in R console**

To be able to **interface better with colleagues** who do most of the programming but being able to suggest approaches to data analysis knowing more about the potential and limitation of R and ML

A practical understanding of data science in order to **collaborate effectively on teams and apply to my own research.**

Course Roadmap



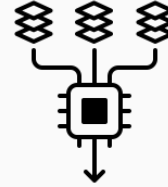
**Importing (loading)
the data**



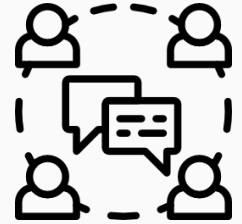
**Processing
(cleaning,
wrangling) the data**



**Visualizing and
summarizing the data**



**Building models
(statistical and ML)**



**Interpretation and
communication of
results**

RStudio and RMarkdown Files

The screenshot displays the RStudio IDE interface for a project named "BST219_2024". The main window is divided into four panes:

- Console:** Shows the R version (4.4.1) and the current directory (~/.Desktop/BST219_2024/). It displays the R startup message, including the copyright notice and the license information. The console output is as follows:

```
R version 4.4.1 (2024-06-14) -- "Race for Your Life"
Copyright (C) 2024 The R Foundation for Statistical Computing
Platform: aarch64-apple-darwin20

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:** Shows the current environment, which is empty. The top bar indicates "Global Environment" and "77 MiB".
- Files:** A file explorer showing the project structure. The files listed are:

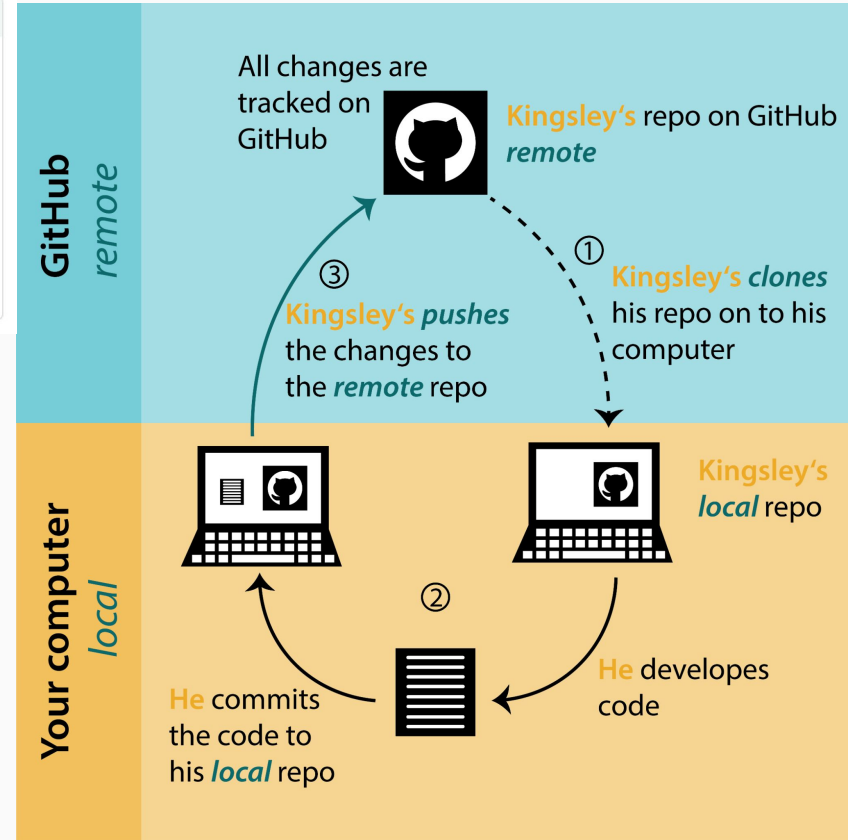
Name	Size	Modified
..		
.gitignore	626 B	Sep 4, 2024, 10:22 PM
00_course_introduction		
01_R-basics		
BST219_2024.Rproj	205 B	Sep 4, 2024, 10:22 PM
LICENSE	1 KB	Sep 4, 2024, 10:22 PM
README.md	3 KB	Sep 4, 2024, 10:22 PM
- Plots:** Empty.
- Packages:** Empty.
- Help:** Empty.
- Viewer:** Empty.
- Presentation:** Empty.

GitHub and RStudio

Definitions




- **repo** - short for repository
- **repository** - the directory or folder that is under version control
- **local** - the repository on your computer
- **remote** - the repository on GitHub
- **commit** - take a snapshot of one or more files in the repository
- **push** - send commits from the local repo to the remote repo
- **pull** - retrieve commits from the remote repo to the local repo
- **.gitignore** a file that tells git which files or types of files you don't want to commit






The screenshot shows the GitHub interface for a repository named 'BST219_2024'. The repository is public and has 1 branch and 0 tags. The file list includes '00_course_introduction', '01_R-basics', '.gitignore', 'LICENSE', and 'README.md'. The README file is selected, showing a welcome message for BST 219: Core Principles of Data Science! and a link to find course materials for Fall 2024.




Source

Homework Repositories

coredatascience / Knabi-2022HW4 








[Code](#) [Issues](#) [Pull requests](#) [Actions](#) [Projects](#) [Security](#) [Insights](#) [Settings](#)



 **Knabi-2022HW4** Private

[Edit Pins](#) [Unwatch](#) **2** [Fork](#) **0** [Star](#) **0**

[master](#) [1 Branch](#) [0 Tags](#) [Add file](#) [Code](#)

 **Knabi** This is my final file. Please grade this. 8fb2327 · 2 years ago 4 Commits

 HW4.Rmd	This is my final file. Please grade this.	2 years ago
 HW4.html	This is my final file. Please grade this.	2 years ago
 README.md	Update README.md	2 years ago
 heart_failure_clinical_records_dataset.csv	Add files via upload	2 years ago







[README](#)  

Homework 4

Due December 12, 2022 by 11:59pm EST

About

2022HW4 created for Knabi


-  [Readme](#)
-  [Activity](#)
-  [Custom properties](#)
-  **0 stars**
-  **2 watching**
-  **0 forks**

Releases




No releases published
[Create a new release](#)






Packages

No packages published
[Publish your first package](#)




Homework Repositories

 coredatascience / Knabi-2022HW4 








[Code](#) [Issues](#) [Pull requests](#) [Actions](#) [Projects](#) [Security](#) [Insights](#) [Settings](#)

 **Knabi-2022HW4** Private

[Edit Pins](#) [Unwatch](#) **2** [Fork](#) **0** [Star](#) **0**

[master](#) [1 Branch](#) [0 Tags](#) [Add file](#) [Code](#)

 **Knabi** This is my final file. Please grade this. 8fb2327 · 2 years ago 4 Commits

 HW4.Rmd	This is my final file. Please grade this.	2 years ago
 HW4.html	This is my final file. Please grade this.	2 years ago
 README.md	Update README.md	2 years ago
 heart_failure_clinical_records_dataset.csv	Add files via upload	2 years ago

README

Homework 4

Due December 12, 2022 by 11:59pm EST

About

2022HW4 created for Knabi

- Readme
- Activity
- Custom properties
- 0 stars
- 2 watching
- 0 forks

Releases

No releases published
[Create a new release](#)

Packages

No packages published
[Publish your first package](#)

Our favorite commit message EVER!