

简介GitHub

曲兆鹏

2025年7月23日

Introduction

- 无论是写论文还是coding 在实践中都会涉及到一个很实际的问题：
- 永远会反复修改，但是如何处理和保存这些不同版本的论文或者Dofiles，以便我们能够找到和区别它们？
- 最简单的也是最笨拙的方式——文件名称和文件夹。
 - Eg. DataClean_2010_0724_Qu.do
 - Happiness_2021_0710_Gao.do

Introduction

Name	Date Modified	Size	Kind
>  Figures	✓ July 19, 2021 at 00:48	--	Folder
>  RawData	✓ July 21, 2021 at 18:39	--	Folder
>  WorkData	✓ Yesterday at 23:21	--	Folder
>  References	✓ Today at 05:04	--	Folder
>  Logfiles	✓ Today at 18:22	--	Folder
>  SaveData	✓ Today at 18:28	--	Folder
▼  Dofiles	✓ Today at 21:55	--	Folder
 DataClean_10_0610.do	✓ June 11, 2021 at 14:36	23 KB	Stata Do-file
 DataClean_10_18_0611.do	✓ June 15, 2021 at 14:40	51 KB	Stata Do-file
 DataClean_2012_2018_0615_Qu_MAC.do	✓ June 19, 2021 at 12:22	51 KB	Stata Do-file
 DataClean_2010_0610_Qu_MAC.do	✓ June 19, 2021 at 23:00	21 KB	Stata Do-file
 DataClean_2010_0619_Qu_MAC.do	✓ June 19, 2021 at 23:13	21 KB	Stata Do-file
 DataClean&Merge_2010_0619_Qu_MAC.do	✓ June 19, 2021 at 23:54	20 KB	Stata Do-file
 DataClean&Merge_2010_0621_Qu_MAC.do	✓ June 21, 2021 at 21:34	23 KB	Stata Do-file
 DataClean&Merge_2012_0622_Qu_MAC.do	✓ June 23, 2021 at 09:24	23 KB	Stata Do-file
 DataFilling_0629.do	✓ June 29, 2021 at 22:31	16 KB	Stata Do-file
 DataFilling_0630.do	✓ June 30, 2021 at 22:15	47 KB	Stata Do-file
 DataClean&Merge_10_12_0627.do	✓ July 1, 2021 at 15:04	41 KB	Stata Do-file
 DataClean&Merge_14_18_0701.do	✓ July 1, 2021 at 17:18	41 KB	Stata Do-file
 CFPS2010.do	✓ July 2, 2021 at 18:59	29 KB	Stata Do-file
 CFPS2016.do	✓ July 2, 2021 at 21:42	18 KB	Stata Do-file
 DataClean_child_edu_0702.do	✓ July 2, 2021 at 21:45	18 KB	Stata Do-file
 CFPS2014.do	✓ July 2, 2021 at 21:45	17 KB	Stata Do-file
 DataFilling_parents_0711.do	✓ July 12, 2021 at 17:04	73 KB	Stata Do-file
 DataFilling_parents_0714.do	✓ July 14, 2021 at 08:51	72 KB	Stata Do-file
 CFPS2012.do	✓ July 17, 2021 at 20:19	23 KB	Stata Do-file
 CFPS2018.do	✓ July 18, 2021 at 20:45	20 KB	Stata Do-file
▼  NewMatch	✓ Today at 16:03	--	Folder

Introduction

Logline	June 29, 2021 at 22:21	--	Folder
Writing	July 17, 2021 at 21:59	--	Folder
农民工资趋同_0511_Qu——无分组.docx	May 15, 2021 at 13:00	2.4 MB	Microso...(docx)
农民工资趋同_0515_Qu——无分组.docx	May 16, 2021 at 12:03	3 MB	Microso...(docx)
城市职工的结果.docx	May 16, 2021 at 13:41	279 KB	Microso...(docx)
农民工资趋同_0515_by_QU.docx	May 16, 2021 at 18:59	1.9 MB	Microso...(docx)
农民工资趋同_0516_by_QU.docx	May 18, 2021 at 12:19	1.9 MB	Microso...(docx)
农民工资趋同_0518_qu.docx	May 19, 2021 at 11:34	1.3 MB	Microso...(docx)
农民工资趋同_20210606_qu02.docx	June 7, 2021 at 02:09	827 KB	Microso...(docx)
农民工资趋同_20210606_中国工业经济.docx	June 7, 2021 at 08:19	829 KB	Microso...(docx)
农民工资趋同_20210606.docx	June 7, 2021 at 08:27	826 KB	Microso...(docx)
农民工资趋同_20210606.pdf	June 7, 2021 at 08:28	2.2 MB	PDF Document
农民工资趋同_20210607.docx	June 7, 2021 at 08:50	826 KB	Microso...(docx)
最终版本	July 17, 2021 at 09:51	--	Folder
农民工资趋同_20210607_final.docx	June 7, 2021 at 08:50	826 KB	Microso...(docx)

- 命名虽然有一定规则，但比较杂乱。
- 更重要的是完全无法记录，版本与版本之间的差别。
- 时间长了之后，很容易忘记版本之间的差别，甚至连最终投稿的版本也可能会忘记。

Introduction

- 如果是多人合作，那么所有的问题都会放大数倍。
- 比如我和王祎、赵铭心之前的合作。她们在**Onedrive**上传了她们的**code**，但我完全看不出来她们在什么地方修改了，需要一条代码一条代码的对照来看才知道。费时费力又容易出错。
- 怎么办？是否有合适的工具或平台来解决这些问题？

Version Control

- “版本控制”：持续记录追踪文档各个版本改变的过程。
- 因为每次改变系统都会自动记录编号，所以可以将文档回复到任意一次修改前的状态。
- 将这个版本控制系统变成一个网络平台，那么就可以多人合作编辑文档。

GitHub简介

- GitHub就是提供这一功能的一个代码托管平台。Git是最常用的版本控制方法，Hub顾名思义就是一个汇集平台。
- 现在已经是全世界最大的代码存放网站和开源社区。
- 上面有无数代码库供其他人复制、模仿、修改和评论等。
- 全世界各地的人可以不见面不认识，通过GitHub可以无缝合作编程，一起完成项目。
- 现在又继续推出了个人网页(page) 教育等功能。

GitHub安装

1. 在电脑上安装Git.

- macOS自带
- Win可以安装GitHub Desktop
<https://desktop.github.com/>

2. 注册账号 <https://github.com/>

GitHub使用

1.创建或者加入一个库(repository)

- 可以认为是关于一个项目的所有文件和文件夹。

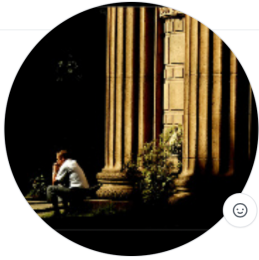
2. 每次使用的既定步骤。

3.使用方式及相应软件

- 网页使用
- 使用**Github Desktop or Rstudio**等。

创建或者加入库(repository)

[Marketplace](#) [Explore](#)



Frank Qu
byelenin

Economist and Number Cruncher.
Using data to explore the economic
and social logic in China.

[Edit profile](#)

[Overview](#) [Repositories 42](#) [Projects](#) [Packages](#)

Pinned

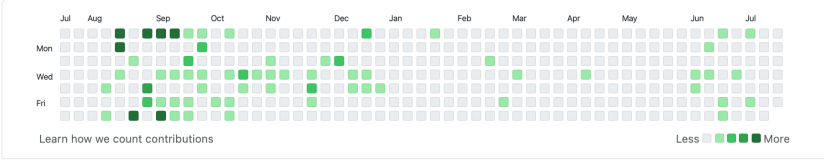
[Econometrics/MBA_BigData](#) [Template](#)

Management in Big Data

☆ 2 🍴 1

329 contributions in the last year

Contribution settings ▾



Learn how we count contributions

Less  More

Contribution activity

July 2021 [2021](#)

New repository

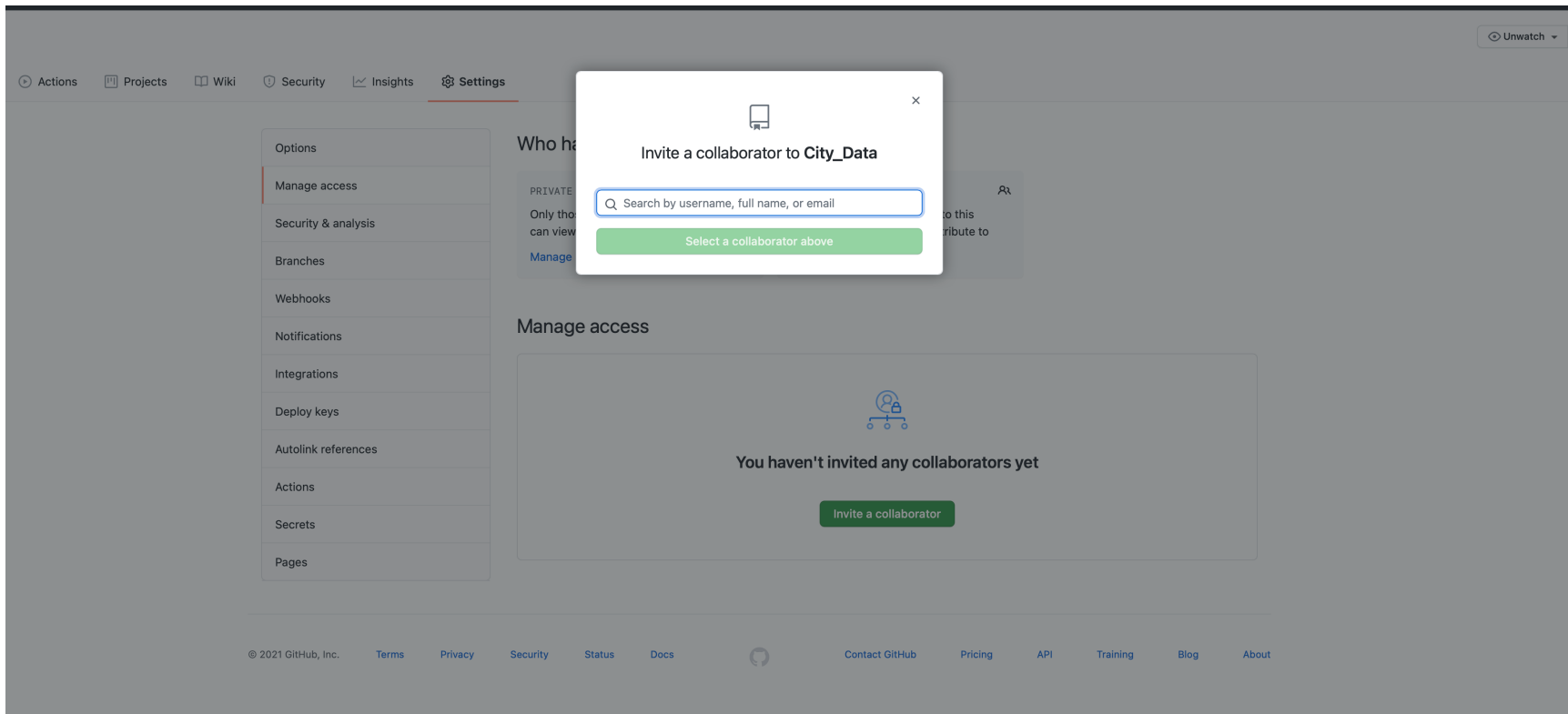
Import repository

New gist

New organization

New project

The owner can invite others to collaborate.



被邀请者接受

 [byelenin](#) / [City_Data](#) Private

👁 Watch 1 ☆ Star 0 🍴 Fork 0

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

 + 

byelenin invited you to collaborate

[Accept invitation](#) [Decline](#)

 [Owners](#) of City_Data will be able to see:

- Your public profile information
- [Certain activity](#) within this repository
- Country of request origin
- Your access level for this repository
- Your IP address

Is this user sending spam or malicious content?

[Block byelenin](#)

© 2021 GitHub, Inc. [Terms](#) [Privacy](#) [Security](#) [Status](#) [Docs](#) 

[Contact GitHub](#) [Pricing](#) [API](#) [Training](#) [Blog](#) [About](#)

接受之后打开库

The screenshot shows a GitHub repository page for 'City_Data' by user 'byelenin'. The repository has 1 branch and 0 tags. The file list includes '.gitignore', 'LICENSE', and 'README.md', all marked as 'Initial commit'. The README.md content is visible, showing the repository name 'City_Data' and its description 'Clean City Level Data from Statistic Books by NBS'. A modal window is open over the repository, showing options to 'Clone' (via HTTPS, SSH, or GitHub CLI) and 'Open with GitHub Desktop' (which is circled in red). The 'Download ZIP' option is also visible. The right sidebar contains sections for 'About' (Clean City Level Data from Statistic Books by NBS, Apache-2.0 License), 'Releases' (No releases published), and 'Packages' (No packages published).

Projects Wiki Security Insights Settings

main 1 branch 0 tags

byelenin Initial commit

- .gitignore Initial commit
- LICENSE Initial commit
- README.md Initial commit

README.md

City_Data

Clean City Level Data from Statistic Books by NBS

Go to file Add file Code

Clone

HTTPS SSH GitHub CLI

https://github.com/byelenin/City_Data

Use Git or checkout with SVN using the web URL.

Open with GitHub Desktop

Download ZIP

About

Clean City Level Data from Statistic Books by NBS

Readme

Apache-2.0 License

Releases

No releases published

[Create a new release](#)

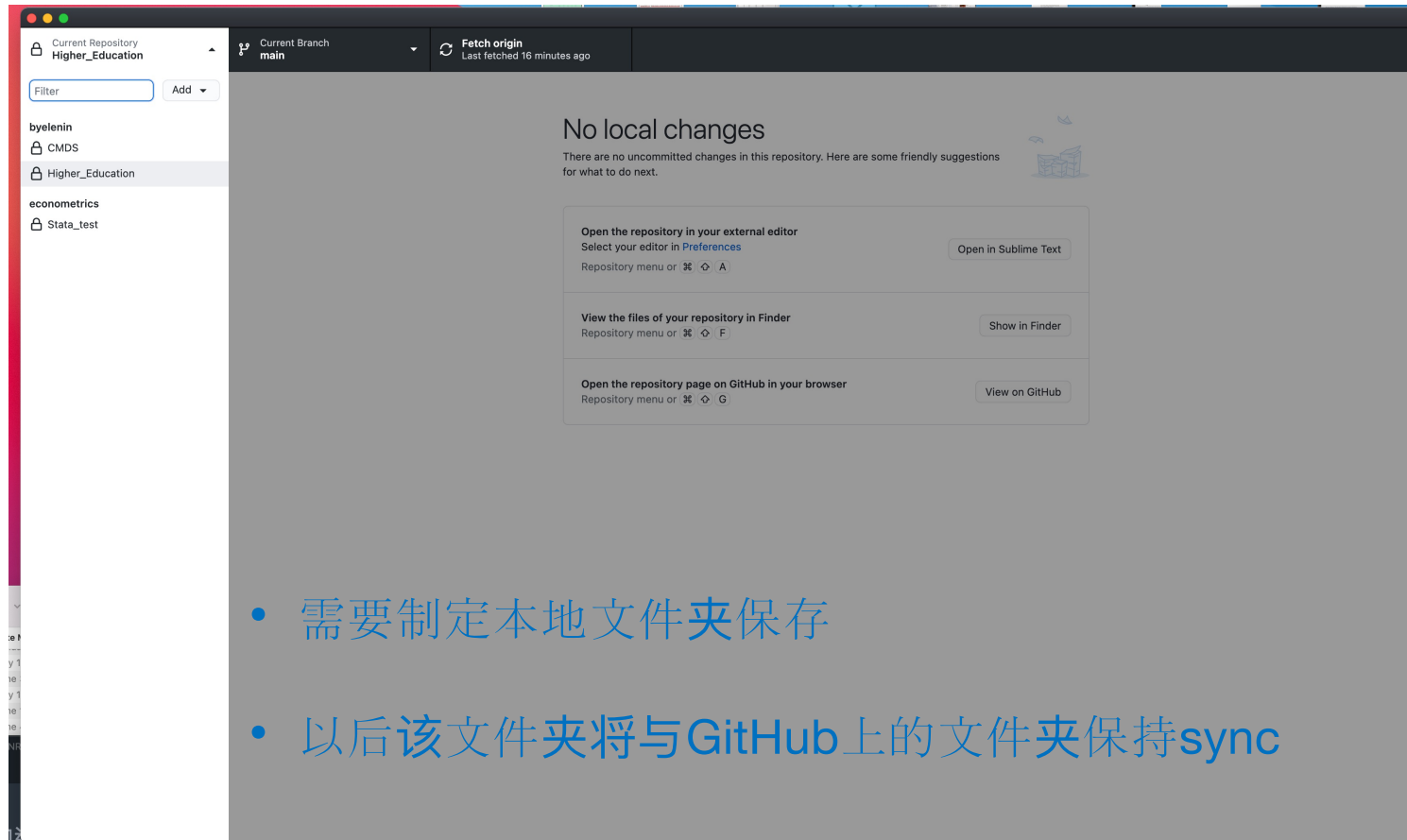
Packages

No packages published

[Publish your first package](#)

© 2021 GitHub, Inc. Terms Privacy Security Status Docs Contact GitHub Pricing API Training Blog About

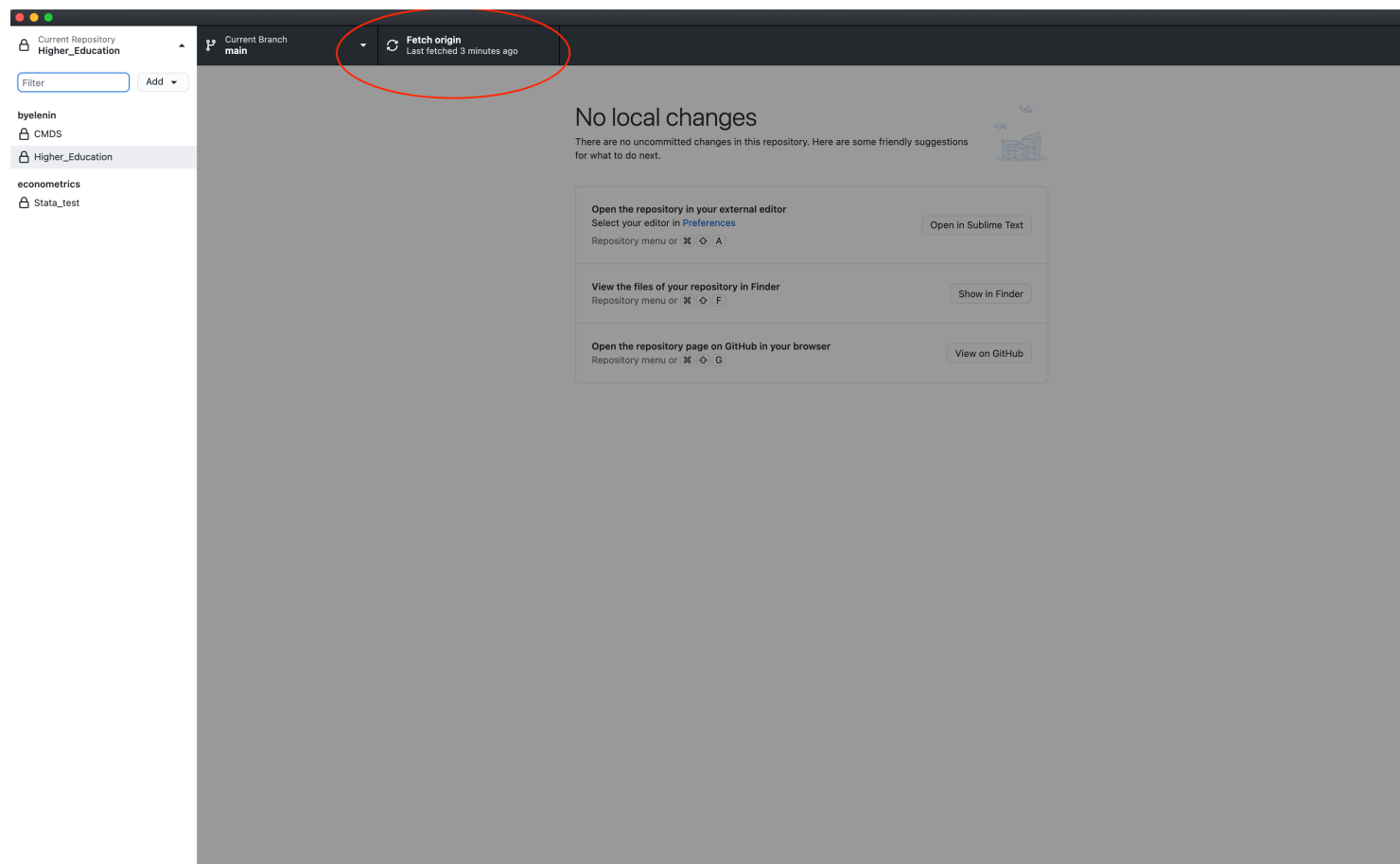
用GitHub Desktop打开



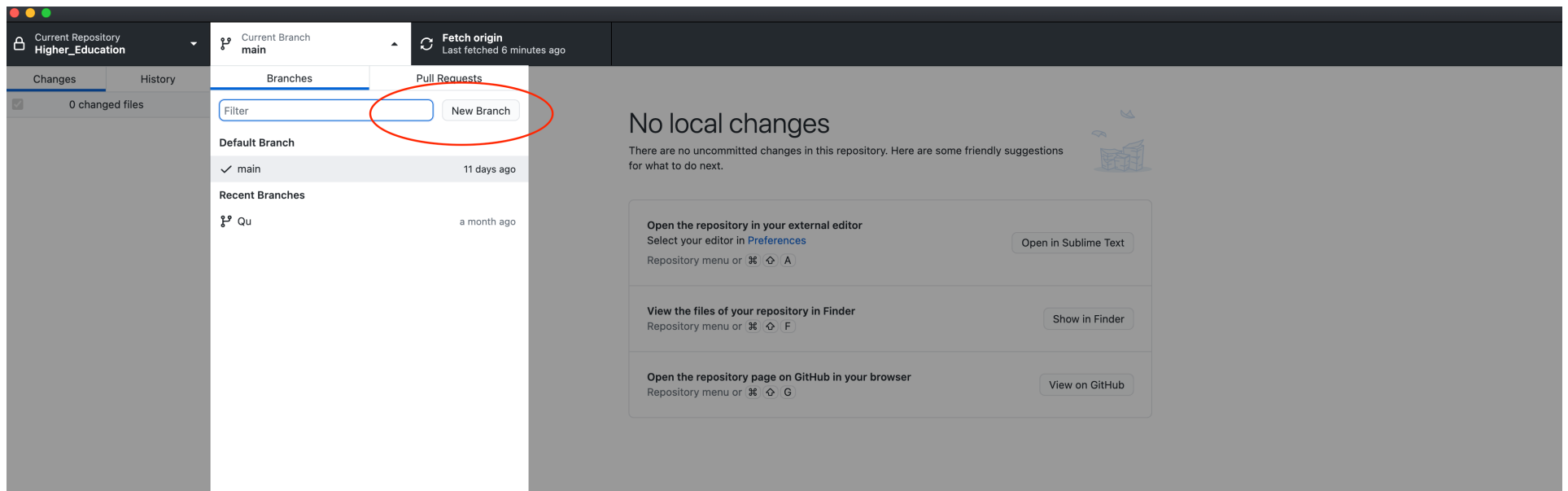
GitHub合作使用步骤

1. 先同步到最新的版本
2. 新建分支(**branch**)
3. 查看他人的修改记录 (如果有的话)。
4. 在分支中进行修改。
 - 增加或删除 **Files**.
 - 修改原有的**Dofile**.
5. 确认修改(**commit**) , 必须写清楚说明文字。
6. 发送修改请求(**pull request**)
7. 将修改合并到主信息库中 (**Merge pull request**)

同步最新版本



新建分支(branch)



查看修改记录

Current Repository
Higher_Education

Current Branch
main

Fetch origin
Last fetched 7 minutes ago

Changes

History

Select Branch to Compare...

Update DataFilling_2010_Parents...

zmx · Jul 12, 2021

Update DataFilling_2010_Parents...

zmx · Jul 12, 2021

Create DataFilling_2010_Parents...

Frank Qu · Jul 9, 2021

最新的文件

Frank Qu · Jul 4, 2021

Data Clean for parents

zmx · Jun 30, 2021

Update DataFilling_0629.do

Frank Qu, zmx · Jun 30, 2021

cleaning and filling in the main var

zmx · Jun 29, 2021

pu

Frank Qu, zmx · Jun 25, 2021

Delete DataClean_0611.do

zmx · Jun 25, 2021

Update DataFilling_2010_Parents_0704.do

zmx · 53ab44f ± 1 changed file

按照生成伪父母匹配回核心库但不drop的思路，对核心库原本就有的父母SES变量做补全。暂时没有管收入变量。

DataFilling_2010_Parents_0704.do

@@ -317,70 +317,117 @@ save temp_cfps_individual_sib,replace

317 /* PART IV : combine 父母信息+ 原始成人收入数据 从而让父母数据库获得收入*/

318 /******

319

320 /*-----*/

321 - /* [> 1. 提取父母信息 <] */

322 /*-----*/

323

324 - use temp_cfps_individual_sib,clear

325

326 - keep pid-countyid code_a_f--td8_a_m feduc meduc fbirthy-mparty qb1 sib_num_s sib_alive

327

328 - save temp_adult_fm,replace // 可以先暂存一个数据，用前缀temp标识。

329

317 /* PART IV : combine 父母信息+ 原始成人收入数据 从而让父母数据库获得收入*/

318 /******

319

320 +

321 /*-----*/

322 + /* [> 1. 提取父母信息 <] */

323 /*-----*/

324

325 + use temp_cfps_individual_sib,clear // 克隆需要变量并匹配自算兄弟姐妹个数的原始成人库

326

327 + *-----*

328 + *-- 1.1 父亲:pid/出生年、政治面貌、最高学历、是否健在、同住、婚姻、有效问卷、职业国标码、职业ISCO

329

330 + *06-22备注:父母是否健在、同住、年龄、出生年代(年龄)、教育年限(或水平)、性别、婚姻状况、就业状况(通过

331

332 + clonevar f_pid = pid_f

333 + clonevar f_view = t1_f //父亲是否有有效问卷

334 + clonevar f_alive = alive_a_f //父亲是否健在

在分支中修改

Current Repository: Higher_Education

Current Branch: New_Test

Publish branch: Publish this branch to GitHub

Changes: History

Select Branch to Compare...

Update DataFilling_2010_Parents_0704.do

zmx · 84adb36 · 1 changed file

补充了在补全核心变量的基础上第二块：heckman两步法预测收入的部分，最后匹配家庭经济库的样本22142个。

DataFilling_2010_Parents_0704.do

```
@@ -1137,6 +1137,13 @@ tab m_age,m
label var f_age "父亲年龄-自算"
label var m_age "母亲年龄-自算"

1137 label var f_age "父亲年龄-自算"
1138 label var m_age "母亲年龄-自算"
1139
1140 +
1141 + gen cohort = 0 if f_birthy_b<1980|m_birthy_b<1980
1142 + replace cohort=1 if cohort==.
1143 + label define cohrt 0 "出生于1980年以前" 1 "出生于1980年以后"
1144 + label var cohort "出生年代是否在1980年以后"
1145 + label values cohort cohrt
1146 +

1147 /* 官方清理报告之中关于缺失值多重插补的思路是否可以借鉴？下文来自《cfps2012家庭SES综合变量》
1148
1149 在整合数据及生成综合变量时，我们制定的基本规则是：

1150 /* 官方清理报告之中关于缺失值多重插补的思路是否可以借鉴？下文来自《cfps2012家庭SES综合变量》
1151
1152 在整合数据及生成综合变量时，我们制定的基本规则是：

1153 @@ -1158,8 +1165,8 @@ label var m_age "母亲年龄-自算"
1158
1159 对于第一、第三种缺陷，用户可以用 多重插补 处理缺失值。对于第二、第四种缺陷，用户可以自行进行数据清理。
1160 */
1161 - preserve
1162 - foreach v of var childnum_alive-m_age {
1163 drop if `v'==.
1164 display "`v'"
1165 }

1166 对于第一、第三种缺陷，用户可以用 多重插补 处理缺失值。对于第二、第四种缺陷，用户可以自行进行数据清理。
1167 */
1168 +
1169 + foreach v of var childnum_alive-cohort f_live m_live {
1170 drop if `v'==.
1171 display "`v'"
1172 }
```

确认修改

Current Repository: Higher_Education | Current Branch: New_Test | Publish branch: Publish this branch to GitHub

Changes: 1 changed file | History

1 changed file

DataFilling_2010_Parents_0704.do

@@ -7,12 +7,6 @@

7 7 *****

8 8

9 9

-

- /*****/

- /* NOTE : 数据清理指南 */

- /*****/

-

10 11

12 13

14 15

16 10 /*0701 “向上兼并方式”，即使用成人数据库里的父母信息来预测父母的收入****

17 11

18 12 A. 从原始2010年成人库(A)的父母信息汇报中得到父母的相关信息，得到父母库(B)

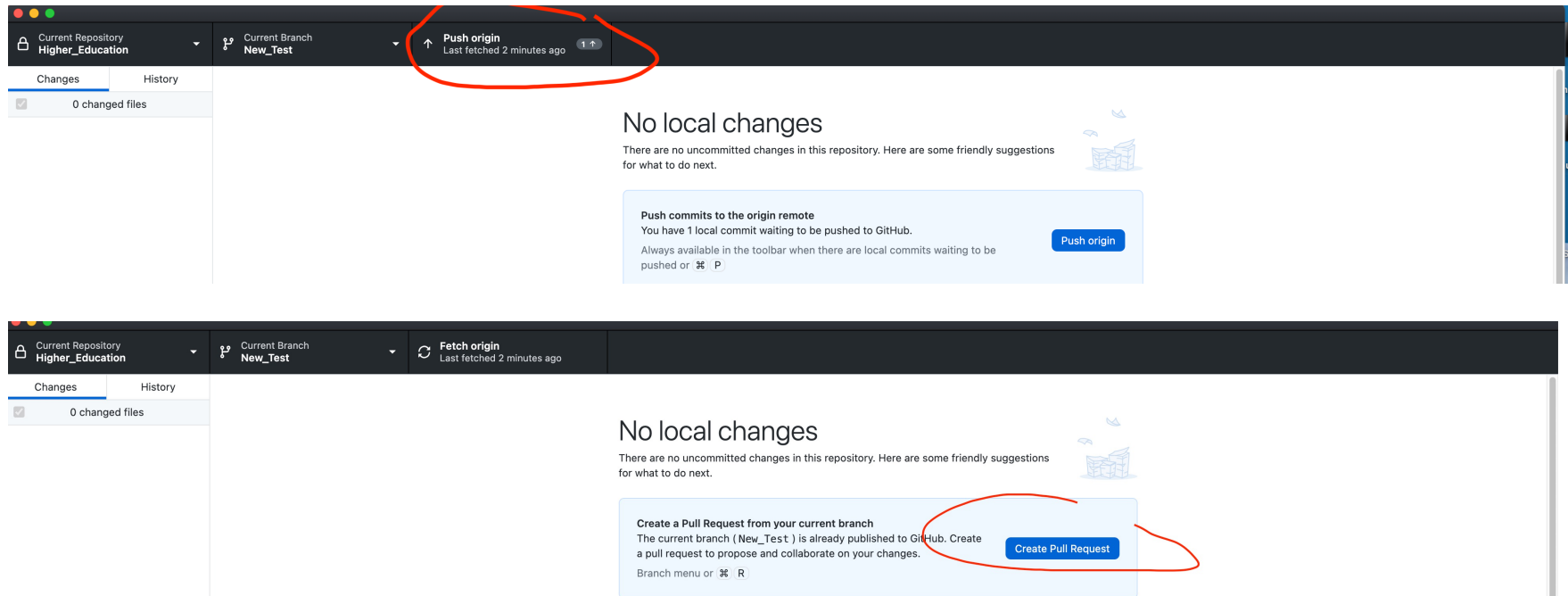
↓ @@ @@

修改

Description

Commit to New_Test

发送请求(push origin and Pull Request)



发送请求(Pull Request)

修改 #1

Edit

Open with ▾

 Open byelenin wants to merge 1 commit into [main](#) from [New_Test](#) 

 Conversation 0

 Commits 1

 Checks 0

 Files changed 1

+0 -6 



byelenin commented now



No description provided.

 修改

4e7095e

Add more commits by pushing to the [New_Test](#) branch on [byelenin/Higher_Education](#).



✓ This branch has no conflicts with the base branch
Merging can be performed automatically.

Merge pull request ▾

You can also [open this in GitHub Desktop](#) or view [command line instructions](#).



Write

Preview

H B I         

Leave a comment

Attach files by dragging & dropping, selecting or pasting them. 

 Close pull request

Comment

Reviewers



No reviews

Assignees



No one—assign yourself

Labels



None yet

Projects



None yet

Milestone



No milestone

Linked issues



Successfully merging this pull request may close these issues.

None yet

Notifications

Customize

 Unsubscribe

You're receiving notifications because you're watching this repository.

修改合并(Merge Pull Request)

修改 #1

[Open](#) byelenin wants to merge 1 commit into `main` from `New_Test`

Conversation 0 Commits 1 Checks 0 Files changed 1 +0 -6

byelenin commented 2 minutes ago

No description provided.

修改 4e7095e

Add more commits by pushing to the `New_Test` branch on `byelenin/Higher_Education`.

Merge pull request #1 from `byelenin/New_Test`

修改

quzhaopeng@gmail.com

Choose which email address to associate with this commit

[Confirm merge](#) [Cancel](#)

Write Preview

Leave a comment

Attach files by dragging & dropping, selecting or pasting them.

[Close pull request](#) [Comment](#)

Remember, contributions to this repository should follow our [GitHub Community Guidelines](#).

[DnTint!](#) Add comments to specific lines under [File changes](#)

Reviewers

No reviews

Assignees

No one—assign yourself

Labels

None yet

Projects

None yet

Milestone

No milestone

Linked issues

Successfully merging this pull request may close these issues.

None yet

Notifications

[Unsubscribe](#)

You're receiving notifications because you're watching this repository.

1 participant

[Lock conversation](#)

合并成功(Merge Pull Request)

修改 #1

Merged byelenin merged 1 commit into `main` from `New_Test` 1 minute ago

Conversation 0 Commits 1 Checks 0 Files changed 1

byelenin commented 5 minutes ago

No description provided.

修改 4e7095e

byelenin merged commit 79a2d19 into `main` 1 minute ago [Revert](#)

Pull request successfully merged and closed [Delete branch](#)

You're all set—the `New_Test` branch can be safely deleted.

Write Preview H B I < >

成功的尝试了一次完整的GitHub操作体验。|

Attach files by dragging & dropping, selecting or pasting them.

[Comment](#)

① Remember, contributions to this repository should follow our [GitHub Community Guidelines](#).

💡 **ProTip!** Add comments to specific lines under [Files changed](#).

GitHub合作使用步骤

- 至此一次完整的**GitHub**操作就完成了。
- 如果想回到之前的某个版本，那么就要**revert**。
- 然后重复刚才的**pull request** 的操作。

如果想恢复修改之前的版本

Mergedbyelenin merged 1 commit into `main` from `New_Test` 9 minutes ago

Conversation 1

Commits 1

Checks 0

Files changed 1

+0 -6

byelenin commented 13 minutes ago

No description provided.

修改4e7095e

byelenin merged commit 79a2d19 into `main` 9 minutes ago

Revert

Create a new pull request to revert these changes

byelenin commented 7 minutes ago

成功的尝试了一次完整的GitHub操作体验。

Pull request successfully merged and closed

You're all set—the `New_Test` branch can be safely deleted.

Delete branch

WritePreview

H B I

Leave a comment

Attach files by dragging & dropping, selecting or pasting them.

Comment

Remember, contributions to this repository should follow our [GitHub Community Guidelines](#).

Lock conversation

Reviewers

No reviews

Assignees

No one—assign yourself

Labels

None yet

Projects

None yet

Milestone

No milestone

Linked issues

Successfully merging this pull request may close these issues.

None yet

Notifications

Customize

Unsubscribe

You're receiving notifications because you're watching this repository.

1 participant

GitHub

- 你们觉得如何？

供大家参考！