

Setup Today

1. Keep www.happygitwithr.com open
2. Go to www.github.com and make a free account
 - Pick a professional, short username; it's hard to change later. More tips at Chapter 4.1 of happygitwithr, <https://happygitwithr.com/github-acct.html#username-advice>
3. Make sure you have a recent version (v1.1 or later) of RStudio <https://www.rstudio.com/products/rstudio/download/#download>
4. Download these slides via the repository: <https://bit.ly/2XHXyl1>.

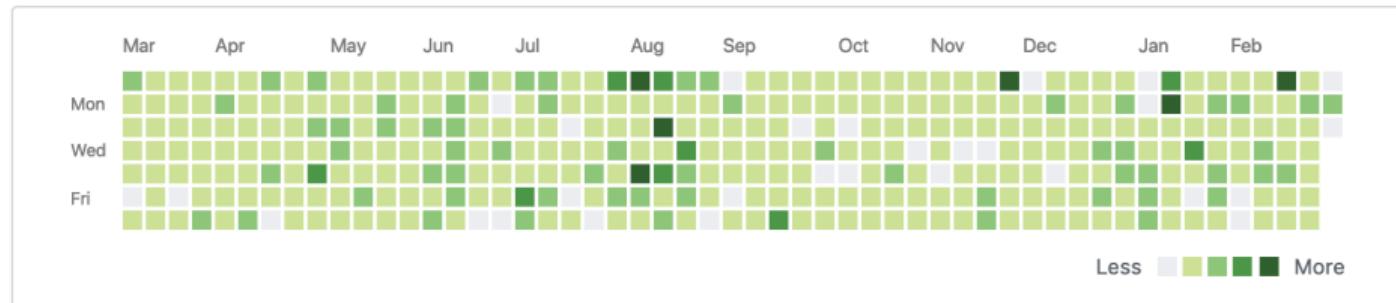
Happy Git and GitHub for the useR

Jenny Bryan, the STAT 545 TAs, Jim Hester

Let's Git started



2,412 contributions in the last year



Git for Students in the Social Sciences*: A Pitch

(* not software developers)

Shiro Kuriwaki

Presented March 5, 2019; Last updated April 14, 2019

Top Figure: From <https://github.com/kuriwaki>

About me

- ▶ G-4 in Government (American Politics, elections and representation)
- ▶ Before: Political data analytics (where I learned git from [Annie Wang](#))
- ▶ I do some software development,
- ▶ but most of my work is applied data analysis

An open question

- ▶ Version control is mandatory for programmers (and professional data scientists)
- ▶ but does it make sense for *applied* researchers who ...
- ▶ work with datasets that are **with collaborators, large, unstructured, and prone to change?**

My perspective

Yes! But in moderation and in *lite*.
This deck is a pitch (while acknowledging Git's inconveniences) and introduction, rather than a full workshop or manual.

Setting expectations: Is it worth it?

What do Gentzkow and Shapiro say?

Definitely:

"It will probably take you a couple days to set up a repository and learn how you want to interact with [version control]. You will break even on that time investment within a month or two."[†]

[†] "Code and Data for Social Sciences: A Practitioners Guide." 2014. <https://perma.cc/5J9D-BTD6>. Although learning git in "a couple of days" sounds too optimistic (I certainly couldn't!), I can guarantee reading their guide in its entirety is a time investment you'll break even on immediately.

But takeup is still low,[‡]

and alternatives have attractive features too:

R Studio Community

Version control with Google Drive



Brett-Johnson

2018-01-08

I've experimented using Google Drive and GitHub with my team (a small ecological research team) for version control and collaboration. I've found that both have their uses and I'm keen to share how I've been doing it so that I can hear from others how they are doing things, and whether I'm on the right track.

I initially started off committing everything I worked on to GitHub in different sub folders in the same repo. All of my internal analyses that aren't meant for a public report or peer reviewed paper went into different folders in the same general 'internal' private repo. This worked all right when it was just me using the repo. But when I brought a co-worker into the mix, I  Quoted what a pain it actually is to try to collaborate on GitHub on a day to day basis. We were spending a load of time messing around with merge conflicts and all sorts of other un-intuitive issues.  We felt GitHub was cumbersome for day to day analysis collaboration internally.

So now I would like to move back to simply using Google Drive for internal analyses. Google drive is great for version controlling (especially now that you can 'name versions' in Google Drive similar to a GitHub commit). I sometimes rely on the revision history of Google Drive to actually roll back a script, because it's way more intuitive than doing that in Git not to mention that every time you save your script in, it gets an un-named version in Google Drive, so the chances of not losing your work is actually greater using Google Drive. Google Drive allows you share all the files you and data you need, and using the here() package we shouldn't have to worry about working directories.

[‡] Anecdotally, I can count full Git users in my department in one hand. Much more in a Psych/lab setting.

Common Misconceptions

1. “Github is a data science tool for sharing data”
 - ~~ It’s built more for version controlling plain-text **code** (that analyzes data) and **text** (that documents it).
 2. “Git is only relevant for software developers”
 - ~~ It also has distinct benefits for the applied researchers’ workflow
 3. “Version control is *only* useful for collaborative projects”
 - ~~ No, in fact we (Bryan’s book) recommend putting your **solo work** under version control,
 - then move on to more complicated collaborations.
- (The organization for the rest of these slides)

Version Control with Yourself (and Your Past Selves)

Terminology 1 of 4 - and recommended setup

- ▶ A version control system **tracks** changes in file content
- ▶ **Git** is a particular type of software for version control (Subversion, or SVN, is an alternative)
- ▶ **GitHub** is an app (acquired by Microsoft) to host git on the web (Bitbucket and GitLab are alternatives)
- ▶ A **desktop client** is an app that connects a webhost like Github to your computer and facilitates tasks otherwise done by **command-line** (here I use **RStudio**; Github Desktop is an alternative)
- ▶ A **repository** is the fundamental unit of a version control project. It's just a regular project folder with a (hidden) subfolder named `.git` added to it. (That `.git` contains the entirety of the project's versions)



Don't make a repository within a repository!

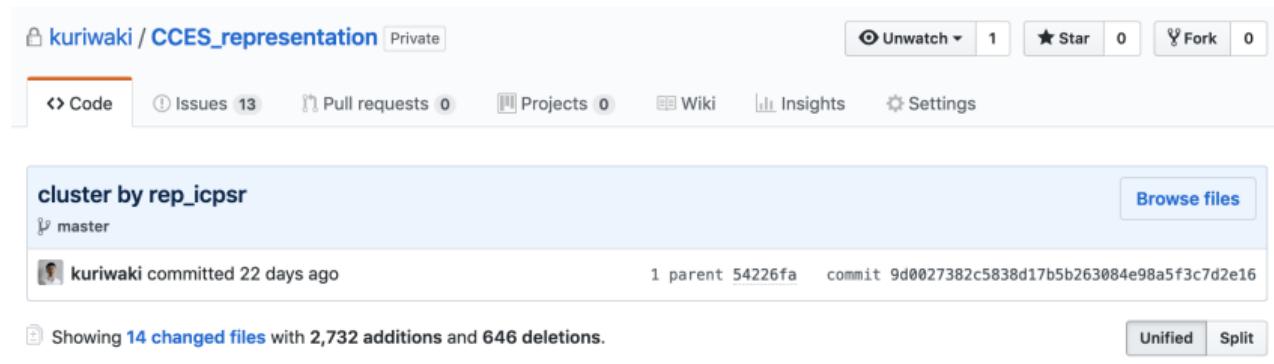
Benefit 1: Keep track of how your results changed

Problem: You tweak a regression specification and re-run your script, re-writing dozens of tables.

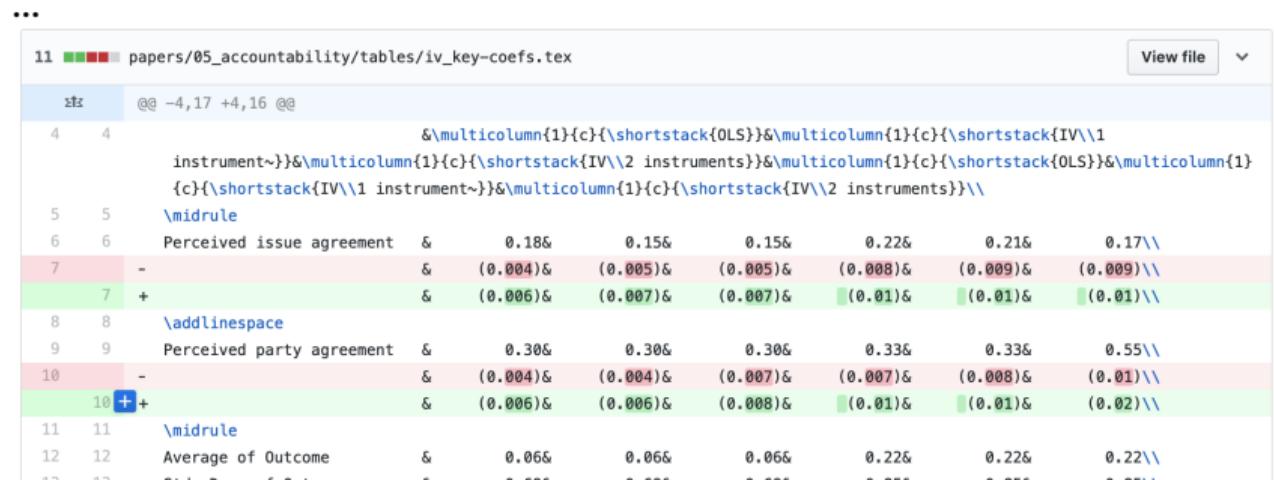
How much did your results change?

You collect more data and re-run the regressions.

Now how did the results change?



The screenshot shows a GitHub repository page for 'kuriwaki / CCES_representation'. The repository is private. The commit 'cluster by rep_icpsr' was made by 'kuriwaki' 22 days ago. It has 1 parent commit (54226fa) and a commit hash of 9d0027382c5838d17b5b263084e98a5f3c7d2e16. The commit message is 'cluster by rep_icpsr'. The commit shows 14 changed files with 2,732 additions and 646 deletions. The 'Unified' view is selected.



The screenshot shows a GitHub commit diff for the file 'papers/05_accountability/tables/iv_key-coefs.tex'. The commit message is 'cluster by rep_icpsr'. The diff shows changes between two versions of the LaTeX code. The changes are color-coded: red for deletions and green for additions. The code is a table of regression coefficients. The 'Unified' view is selected.

4	4	Perceived issue agreement	&	0.18&	0.15&	0.15&	0.22&	0.21&
5	5	-	&	(0.004)&	(0.005)&	(0.005)&	(0.008)&	(0.009)&
7	7	+	&	(0.006)&	(0.007)&	(0.007)&	(0.01)&	(0.01)&
8	8	\addlinespace						
9	9	Perceived party agreement	&	0.30&	0.30&	0.30&	0.33&	0.33&
10	10	-	&	(0.004)&	(0.004)&	(0.007)&	(0.007)&	(0.01)&
11	11	+	&	(0.006)&	(0.006)&	(0.008)&	(0.01)&	(0.01)&
12	12	Average of Outcome	&	0.066	0.066	0.066	0.22&	0.22&
13	13	Std. Err.	&	0.001	0.001	0.001	0.001	0.001

Benefit 2: Tracking your paper versions

Problem: You start writing up your paper, `draft.tex`

- The next day, you make a new draft. Do you overwrite?
- Or do you call it `draft_0305.tex` ?
`draft_03052019.tex` ?
- The next week, you find a single typo. Do you “Save As” with a new date?
- Three weeks later, you return to your paper. Your computer indicates that the file named `draft_0305.tex` was “Last modified March 12, 2019”.

Showing 5 changed files with 62 additions and 51 deletions.

Unified Split

2 analyze@06_rcv_accountability.do

275 275 @0 -275,7 +275,7 @@ esttab est1 est3 est5 est2 est4 est6 using "papers/05_accountability/tables/iv_k
276 276 span erepeat{\cmidrule(lr){span}} ///
277 277 mtitle("\shortstack{OLS}" "\shortstack{IV\1 instrument}" "\shortstack{IV\2 instruments}" "\shortstack{OLS}" "\short
278 - b(2) se(1) ///
278 + addnotes("All other variables and intercept not shown") ///
279 279 + addnotes("All other variables and intercept not shown. All IV estimates include year fixed effects.") ///
280 280 stats(smean ysd r2 N, ///
281 281 fmt(2 2 2 %6.0fc) ///
281 labels("Average of Outcome" "Std. Dev. of Outcome" "R-squared" "Observations") ///
281

82 papers/05_accountability/ajk.tex

259 259 @0 -259,9 +259,9 @@ \section{Data and Methods}
260 260 \subsection{Operationalization of Key Variables} \label{sec:operationalization}
261 261
262 - Our key measures of perceived agreement are built from the responses to the perception questions in the CCES Module (a random subset of the entire study), combined with their own stances to the same issues on the same question. An example of such perception questions from 2017 is reproduced in Exhibit \ref{fig:perception_question}.
262 + Our key measures of perceived agreement are built from the responses to the perception questions in the CCES Module (a random subset of the entire study), combined with their own stances to the same issues on the same question. An example of such perception questions from 2017 is reproduced in Exhibit \ref{fig:perception_question}. Throughout, we limit our attention to the House primarily due to space restrictions.
263 263
264 + We also construct instruments for our measure of perceived agreement by collecting matching roll call vote data from the **NOMINATE** database (\url{https://voteview.com/}). To facilitate the interpretability of regression coefficients, we intentionally define all of variables on a -1 to +1 scale. Table \ref{tab:summary_stats} presents summary statistics, and a description of each of the variables follows.
264 + We also construct instruments for our measure of perceived agreement by collecting matching roll call vote data from the **Voteview** database (\url{https://voteview.com/}). To facilitate the interpretability of regression coefficients, we intentionally define all of variables on a -1 to +1 scale. Table \ref{tab:summary_stats} presents summary statistics, and a description of each of the variables follows.

Benefit 3: And more cool stuff like

Getting a free, customizable, ad-free website

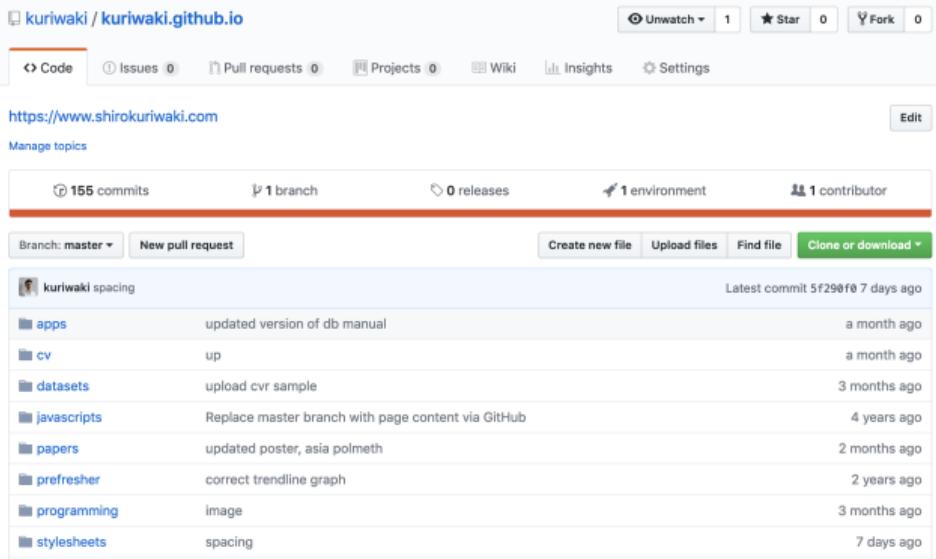
(instead of a click-and-drag Wordpress/Squarespace website)

Work on a collaborative workbook

(instead of needing to add people to your Dropbox)

Contributing to / getting the latest on actual software packages

Github issues is the de facto communication of open-source developers.*



The screenshot shows a GitHub repository page for the user 'kuriwaki' with the repository name 'kuriwaki.github.io'. The page includes a navigation bar with links for Code, Issues (0), Pull requests (0), Projects (0), Wiki, Insights, and Settings. It also shows statistics: 155 commits, 1 branch, 0 releases, 1 environment, and 1 contributor. A 'New pull request' button is visible. The main content is a list of commits, each with a file icon, the author, a brief description, and the time of the commit. The commits are as follows:

Author	File	Description	Time
kuriwaki	spacing	updated version of db manual	7 days ago
	apps	up	a month ago
	cv	upload cvr sample	a month ago
	datasets	Replace master branch with page content via GitHub	3 months ago
	javascripts	updated poster, asia polmeth	4 years ago
	papers	correct trendline graph	2 months ago
	refresher	image	2 years ago
	programming	spacing	3 months ago
	stylesheets		7 days ago

Benefit 3: And more cool stuff like

**Getting a free, customizable,
ad-free website**

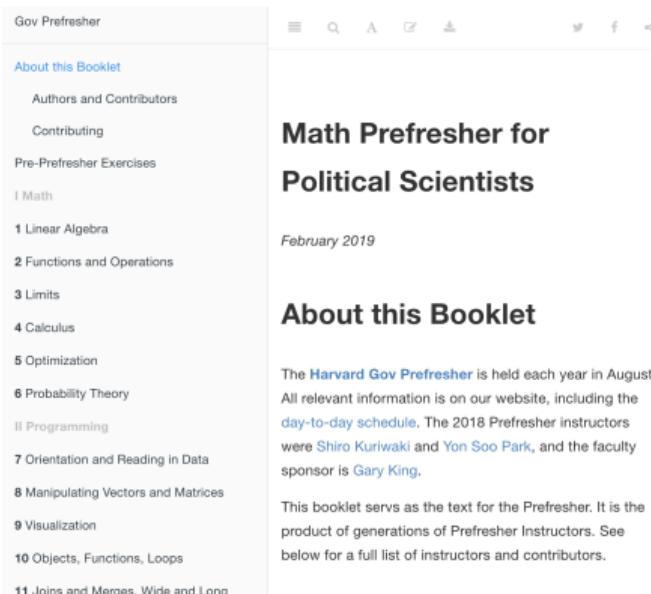
(instead of a click-and-drag
Wordpress/Squarespace website)

Work on a collaborative workbook

(instead of needing to add people to
your Dropbox)

**Contributing to / getting the latest
on actual software packages**

Github issues is the de facto commun-
ication of open-source developers.*



The image shows a screenshot of a booklet titled "Gov Prefresher". The interface includes a navigation bar with icons for search, refresh, and navigation. The main content area is titled "About this Booklet" and lists the following sections: Authors and Contributors, Contributing, Pre-Prefresher Exercises, I Math, 1 Linear Algebra, 2 Functions and Operations, 3 Limits, 4 Calculus, 5 Optimization, 6 Probability Theory, II Programming, 7 Orientation and Reading in Data, 8 Manipulating Vectors and Matrices, 9 Visualization, 10 Objects, Functions, Loops, and 11 Joins and Merges, Wide and Long. Below the content area, there is a section titled "Math Prefresher for Political Scientists" with a sub-section titled "About this Booklet". The text in this section discusses the Harvard Gov Prefresher, its schedule, instructors, and sponsor. It also notes that the booklet serves as the text for the Prefresher and provides a link to a full list of instructors and contributors.

**Math Prefresher for
Political Scientists**

February 2019

About this Booklet

The [Harvard Gov Prefresher](#) is held each year in August. All relevant information is on our website, including the [day-to-day schedule](#). The 2018 Prefresher instructors were [Shiro Kuriwaki](#) and [Yon Soo Park](#), and the faculty sponsor is [Gary King](#).

This booklet serves as the text for the Prefresher. It is the product of generations of Prefresher Instructors. See below for a full list of instructors and contributors.

Benefit 3: And more cool stuff like

Getting a free, customizable, ad-free website

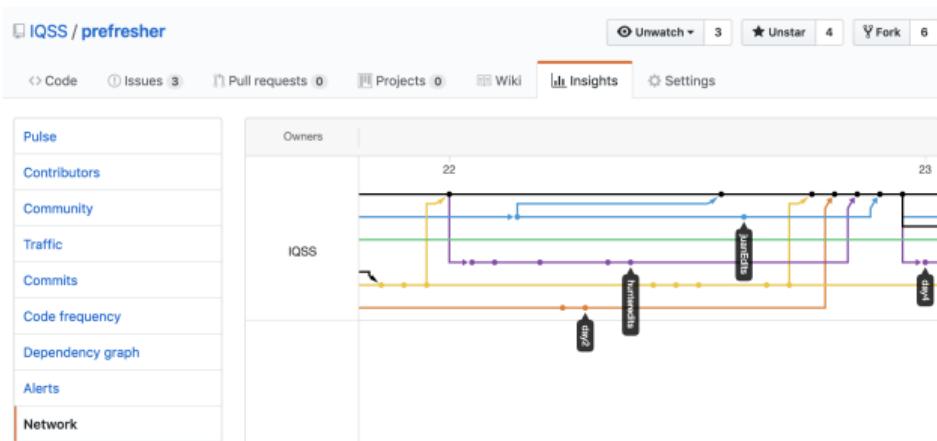
(instead of a click-and-drag Wordpress/Squarespace website)

Work on a collaborative workbook

(instead of needing to add people to your Dropbox)

Contributing to / getting the latest on actual software packages

GitHub issues is the de facto communication of open-source developers.*



Benefit 3: And more cool stuff like

Getting a free, customizable, ad-free website

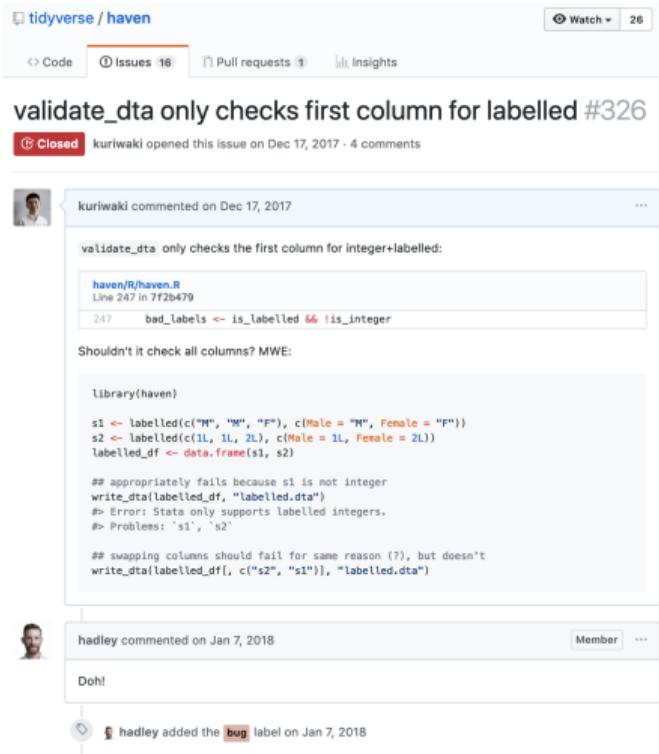
(instead of a click-and-drag Wordpress/Squarespace website)

Work on a collaborative workbook

(instead of needing to add people to your Dropbox)

Contributing to / getting the latest on actual software packages

Github issues is the de facto communication of open-source developers.*



tidyverse / haven

Code Issues 16 Pull requests 1 Insights Watch 26

validate_dta only checks first column for labelled #326

Closed kuriwaki opened this issue on Dec 17, 2017 · 4 comments

kuriwaki commented on Dec 17, 2017

validate_dta only checks the first column for integer+labelled:

haven/R/haven.R
Line 247 in 7f72b479
247 bad_labels <- is_labelled && !is_integer

Shouldn't it check all columns? MWE:

```
library(haven)  
  
s1 <- labelled(c("M", "M", "F"), c(Male = "M", Female = "F"))  
s2 <- labelled(c(1L, 1L, 2L), c(Male = 1L, Female = 2L))  
labelled_df <- data.frame(s1, s2)  
  
## appropriately fails because s1 is not integer  
write_dta(labelled_df, "labelled.dta")  
#> Error: Stata only supports labelled integers.  
#> Problems: 's1', 's2'  
  
## swapping columns should fail for same reason (?), but doesn't  
write_dta(labelled_df, c("s2", "s1")), "labelled.dta")
```

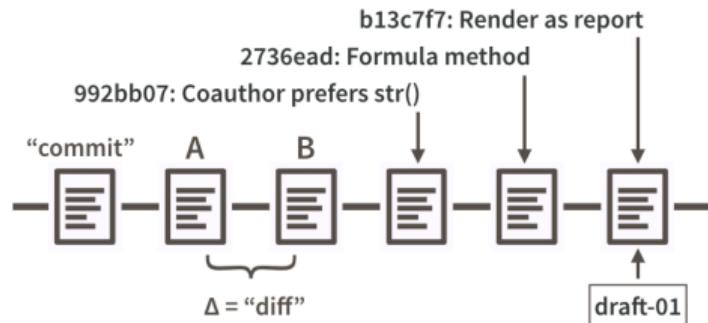
hadley commented on Jan 7, 2018

Doh!

hadley added the bug label on Jan 7, 2018

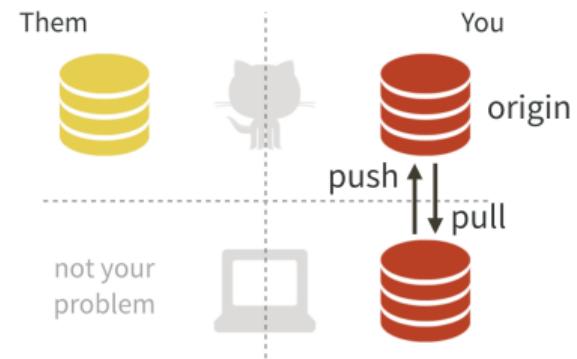
Terminology 2 of 4: Pushing commits

- ▶ Files increment by **commits**. The line-by-line changes between a pair of commits is a **diff**.
- ▶ Commits are explicit, not automatic: Unlike the `Cmd + S` Save, commits are labelled by a human-readable **message**, and a serial code called a **SHA** (like `992bb07`).
- ▶ And git requires you `stage` a change by **adding** it, before turning it a commit. (but let's worry about this later)
- ▶ Git sees a files as essentially an accumulation of commits. That accumulation is a **branch**. (this naming choice makes more sense with more than one "branch.")



Terminology 3 of 4: local and remote, push and pull

- ▶ Two copies of your repo exist: the **local** on your computer, and a **remote** (hosted on Github, with URL <https://github.com/user/repo.git>), which has the name **origin**
- ▶ Once you make commits on your local, you **push** them to your remote. (Imagine an upward push, from the ground to the cloud)
- ▶ The opposite of this is a **pull**. (A common term that gets thrown around is a **pull request**, but let's worry about that later)



daily work, your stuff

Now, some caveats

Only plain-text files get tracked line-by-line

So non plain-text files:

e.g. PDFs (`.pdf`), JPEGs, Microsoft Word, Powerpoint, Excel (`xlsx`), Google Docs, `.sav` , `.por` , `.dta` , `.Rds` , RData ...

can be tracked, but git's value-add is small here.

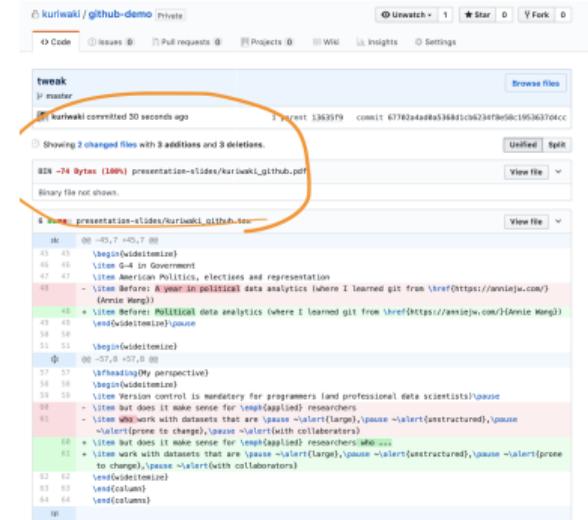
Therefore, requires a switch to working with

Markdown (`.md`) and TeX (`.tex`) for writing, code (`.R` , `.py`)-centered output, small datasets in `.csv` or `.txt` , interweavers like `.Rmd` .

Kieran Healy, *"The Plain Person's Guide to Plain Text Social Science."*

GitHub places a 100MB cap on each file, and a 1GB cap on the entire directory. Anything larger is **not** trackable in GitHub.

Git is not built for storing data!



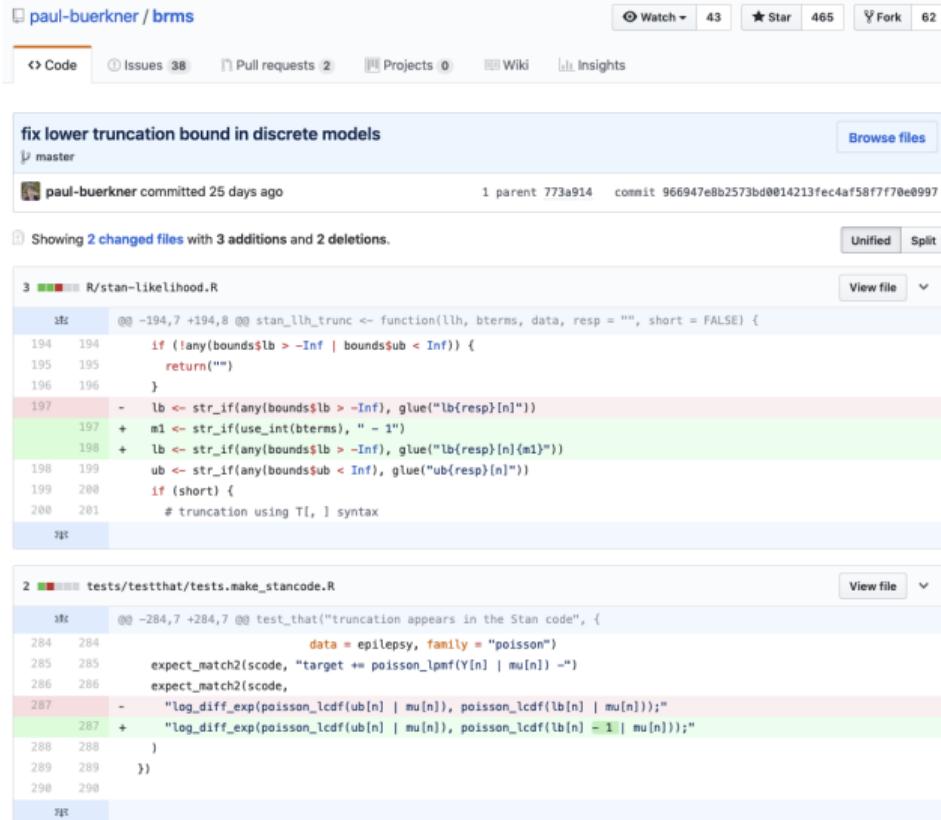
A screenshot of a GitHub commit page for a file named `tweak` in the `karuki/githup-demo` repository. The commit message is "karuki committed 30 seconds ago". The commit hash is `67782a4ed8a5368d1cb6234f8e50c195363704cc`. The commit shows 2 changed files with 8 additions and 3 deletions. One of the files is a binary file named `presentation-slides/karuki/githup.pdf` which is 74 bytes (100%) and is not shown. The commit message contains several lines of text, some of which are highlighted in different colors (blue, green, red) to indicate different types of text or code blocks. The code block is as follows:

```
88 -74 bytes (100%) presentation-slides/karuki/githup.pdf
Binary file not shown.

4 -4 presentation-slides/karuki/githup.pdf
 88 -74 bytes (100%) presentation-slides/karuki/githup.pdf
 89 89
 90 90
 91 91
 92 92
 93 93
 94 94
 95 95
 96 96
 97 97
 98 98
 99 99
100 100
101 101
102 102
103 103
104 104
105 105
106 106
107 107
108 108
109 109
110 110
111 111
112 112
113 113
114 114
115 115
116 116
117 117
118 118
119 119
120 120
121 121
122 122
123 123
124 124
125 125
126 126
127 127
128 128
129 129
130 130
131 131
132 132
133 133
134 134
135 135
136 136
137 137
138 138
139 139
140 140
141 141
142 142
143 143
144 144
145 145
146 146
147 147
148 148
149 149
150 150
151 151
152 152
153 153
154 154
155 155
156 156
157 157
158 158
159 159
160 160
161 161
162 162
163 163
164 164
165 165
166 166
167 167
168 168
169 169
170 170
171 171
172 172
173 173
174 174
175 175
176 176
177 177
178 178
179 179
180 180
181 181
182 182
183 183
184 184
185 185
186 186
187 187
188 188
189 189
190 190
191 191
192 192
193 193
194 194
195 195
196 196
197 197
198 198
199 199
200 200
201 201
202 202
203 203
204 204
205 205
206 206
207 207
208 208
209 209
210 210
211 211
212 212
213 213
214 214
215 215
216 216
217 217
218 218
219 219
220 220
221 221
222 222
223 223
224 224
225 225
226 226
227 227
228 228
229 229
230 230
231 231
232 232
233 233
234 234
235 235
236 236
237 237
238 238
239 239
240 240
241 241
242 242
243 243
244 244
245 245
246 246
247 247
248 248
249 249
250 250
251 251
252 252
253 253
254 254
255 255
256 256
257 257
258 258
259 259
260 260
261 261
262 262
263 263
264 264
265 265
266 266
267 267
268 268
269 269
270 270
271 271
272 272
273 273
274 274
275 275
276 276
277 277
278 278
279 279
280 280
281 281
282 282
283 283
284 284
285 285
286 286
287 287
288 288
289 289
290 290
291 291
292 292
293 293
294 294
295 295
296 296
297 297
298 298
299 299
300 300
301 301
302 302
303 303
304 304
305 305
306 306
307 307
308 308
309 309
310 310
311 311
312 312
313 313
314 314
315 315
316 316
317 317
318 318
319 319
320 320
321 321
322 322
323 323
324 324
325 325
326 326
327 327
328 328
329 329
330 330
331 331
332 332
333 333
334 334
335 335
336 336
337 337
338 338
339 339
340 340
341 341
342 342
343 343
344 344
345 345
346 346
347 347
348 348
349 349
350 350
351 351
352 352
353 353
354 354
355 355
356 356
357 357
358 358
359 359
360 360
361 361
362 362
363 363
364 364
365 365
366 366
367 367
368 368
369 369
370 370
371 371
372 372
373 373
374 374
375 375
376 376
377 377
378 378
379 379
380 380
381 381
382 382
383 383
384 384
385 385
386 386
387 387
388 388
389 389
390 390
391 391
392 392
393 393
394 394
395 395
396 396
397 397
398 398
399 399
400 400
401 401
402 402
403 403
404 404
405 405
406 406
407 407
408 408
409 409
410 410
411 411
412 412
413 413
414 414
415 415
416 416
417 417
418 418
419 419
420 420
421 421
422 422
423 423
424 424
425 425
426 426
427 427
428 428
429 429
430 430
431 431
432 432
433 433
434 434
435 435
436 436
437 437
438 438
439 439
440 440
441 441
442 442
443 443
444 444
445 445
446 446
447 447
448 448
449 449
450 450
451 451
452 452
453 453
454 454
455 455
456 456
457 457
458 458
459 459
460 460
461 461
462 462
463 463
464 464
465 465
466 466
467 467
468 468
469 469
470 470
471 471
472 472
473 473
474 474
475 475
476 476
477 477
478 478
479 479
480 480
481 481
482 482
483 483
484 484
485 485
486 486
487 487
488 488
489 489
490 490
491 491
492 492
493 493
494 494
495 495
496 496
497 497
498 498
499 499
500 500
501 501
502 502
503 503
504 504
505 505
506 506
507 507
508 508
509 509
510 510
511 511
512 512
513 513
514 514
515 515
516 516
517 517
518 518
519 519
520 520
521 521
522 522
523 523
524 524
525 525
526 526
527 527
528 528
529 529
530 530
531 531
532 532
533 533
534 534
535 535
536 536
537 537
538 538
539 539
540 540
541 541
542 542
543 543
544 544
545 545
546 546
547 547
548 548
549 549
550 550
551 551
552 552
553 553
554 554
555 555
556 556
557 557
558 558
559 559
560 560
561 561
562 562
563 563
564 564
565 565
566 566
567 567
568 568
569 569
570 570
571 571
572 572
573 573
574 574
575 575
576 576
577 577
578 578
579 579
580 580
581 581
582 582
583 583
584 584
585 585
586 586
587 587
588 588
589 589
590 590
591 591
592 592
593 593
594 594
595 595
596 596
597 597
598 598
599 599
600 600
601 601
602 602
603 603
604 604
605 605
606 606
607 607
608 608
609 609
610 610
611 611
612 612
613 613
614 614
615 615
616 616
617 617
618 618
619 619
620 620
621 621
622 622
623 623
624 624
625 625
626 626
627 627
628 628
629 629
630 630
631 631
632 632
633 633
634 634
635 635
636 636
637 637
638 638
639 639
640 640
641 641
642 642
643 643
644 644
645 645
646 646
647 647
648 648
649 649
650 650
651 651
652 652
653 653
654 654
655 655
656 656
657 657
658 658
659 659
660 660
661 661
662 662
663 663
664 664
665 665
666 666
667 667
668 668
669 669
670 670
671 671
672 672
673 673
674 674
675 675
676 676
677 677
678 678
679 679
680 680
681 681
682 682
683 683
684 684
685 685
686 686
687 687
688 688
689 689
690 690
691 691
692 692
693 693
694 694
695 695
696 696
697 697
698 698
699 699
700 700
701 701
702 702
703 703
704 704
705 705
706 706
707 707
708 708
709 709
710 710
711 711
712 712
713 713
714 714
715 715
716 716
717 717
718 718
719 719
720 720
721 721
722 722
723 723
724 724
725 725
726 726
727 727
728 728
729 729
730 730
731 731
732 732
733 733
734 734
735 735
736 736
737 737
738 738
739 739
740 740
741 741
742 742
743 743
744 744
745 745
746 746
747 747
748 748
749 749
750 750
751 751
752 752
753 753
754 754
755 755
756 756
757 757
758 758
759 759
760 760
761 761
762 762
763 763
764 764
765 765
766 766
767 767
768 768
769 769
770 770
771 771
772 772
773 773
774 774
775 775
776 776
777 777
778 778
779 779
780 780
781 781
782 782
783 783
784 784
785 785
786 786
787 787
788 788
789 789
790 790
791 791
792 792
793 793
794 794
795 795
796 796
797 797
798 798
799 799
800 800
801 801
802 802
803 803
804 804
805 805
806 806
807 807
808 808
809 809
810 810
811 811
812 812
813 813
814 814
815 815
816 816
817 817
818 818
819 819
820 820
821 821
822 822
823 823
824 824
825 825
826 826
827 827
828 828
829 829
830 830
831 831
832 832
833 833
834 834
835 835
836 836
837 837
838 838
839 839
840 840
841 841
842 842
843 843
844 844
845 845
846 846
847 847
848 848
849 849
850 850
851 851
852 852
853 853
854 854
855 855
856 856
857 857
858 858
859 859
860 860
861 861
862 862
863 863
864 864
865 865
866 866
867 867
868 868
869 869
870 870
871 871
872 872
873 873
874 874
875 875
876 876
877 877
878 878
879 879
880 880
881 881
882 882
883 883
884 884
885 885
886 886
887 887
888 888
889 889
890 890
891 891
892 892
893 893
894 894
895 895
896 896
897 897
898 898
899 899
900 900
901 901
902 902
903 903
904 904
905 905
906 906
907 907
908 908
909 909
910 910
911 911
912 912
913 913
914 914
915 915
916 916
917 917
918 918
919 919
920 920
921 921
922 922
923 923
924 924
925 925
926 926
927 927
928 928
929 929
930 930
931 931
932 932
933 933
934 934
935 935
936 936
937 937
938 938
939 939
940 940
941 941
942 942
943 943
944 944
945 945
946 946
947 947
948 948
949 949
950 950
951 951
952 952
953 953
954 954
955 955
956 956
957 957
958 958
959 959
960 960
961 961
962 962
963 963
964 964
965 965
966 966
967 967
968 968
969 969
970 970
971 971
972 972
973 973
974 974
975 975
976 976
977 977
978 978
979 979
980 980
981 981
982 982
983 983
984 984
985 985
986 986
987 987
988 988
989 989
990 990
991 991
992 992
993 993
994 994
995 995
996 996
997 997
998 998
999 999
1000 1000
1001 1001
1002 1002
1003 1003
1004 1004
1005 1005
1006 1006
1007 1007
1008 1008
1009 1009
1010 1010
1011 1011
1012 1012
1013 1013
1014 1014
1015 1015
1016 1016
1017 1017
1018 1018
1019 1019
1020 1020
1021 1021
1022 1022
1023 1023
1024 1024
1025 1025
1026 1026
1027 1027
1028 1028
1029 1029
1030 1030
1031 1031
1032 1032
1033 1033
1034 1034
1035 1035
1036 1036
1037 1037
1038 1038
1039 1039
1040 1040
1041 1041
1042 1042
1043 1043
1044 1044
1045 1045
1046 1046
1047 1047
1048 1048
1049 1049
1050 1050
1051 1051
1052 1052
1053 1053
1054 1054
1055 1055
1056 1056
1057 1057
1058 1058
1059 1059
1060 1060
1061 1061
1062 1062
1063 1063
1064 1064
1065 1065
1066 1066
1067 1067
1068 1068
1069 1069
1070 1070
1071 1071
1072 1072
1073 1073
1074 1074
1075 1075
1076 1076
1077 1077
1078 1078
1079 1079
1080 1080
1081 1081
1082 1082
1083 1083
1084 1084
1085 1085
1086 1086
1087 1087
1088 1088
1089 1089
1090 1090
1091 1091
1092 1092
1093 1093
1094 1094
1095 1095
1096 1096
1097 1097
1098 1098
1099 1099
1100 1100
1101 1101
1102 1102
1103 1103
1104 1104
1105 1105
1106 1106
1107 1107
1108 1108
1109 1109
1110 1110
1111 1111
1112 1112
1113 1113
1114 1114
1115 1115
1116 1116
1117 1117
1118 1118
1119 1119
1120 1120
1121 1121
1122 1122
1123 1123
1124 1124
1125 1125
1126 1126
1127 1127
1128 1128
1129 1129
1130 1130
1131 1131
1132 1132
1133 1133
1134 1134
1135 1135
1136 1136
1137 1137
1138 1138
1139 1139
1140 1140
1141 1141
1142 1142
1143 1143
1144 1144
1145 1145
1146 1146
1147 1147
1148 1148
1149 1149
1150 1150
1151 1151
1152 1152
1153 1153
1154 1154
1155 1155
1156 1156
1157 1157
1158 1158
1159 1159
1160 1160
1161 1161
1162 1162
1163 1163
1164 1164
1165 1165
1166 1166
1167 1167
1168 1168
1169 1169
1170 1170
1171 1171
1172 1172
1173 1173
1174 1174
1175 1175
1176 1176
1177 1177
1178 1178
1179 1179
1180 1180
1181 1181
1182 1182
1183 1183
1184 1184
1185 1185
1186 1186
1187 1187
1188 1188
1189 1189
1190 1190
1191 1191
1192 1192
1193 1193
1194 1194
1195 1195
1196 1196
1197 1197
1198 1198
1199 1199
1200 1200
1201 1201
1202 1202
1203 1203
1204 1204
1205 1205
1206 1206
1207 1207
1208 1208
1209 1209
1210 1210
1211 1211
1212 1212
1213 1213
1214 1214
1215 1215
1216 1216
1217 1217
1218 1218
1219 1219
1220 1220
1221 1221
1222 1222
1223 1223
1224 1224
1225 1225
1226 1226
1227 1227
1228 1228
1229 1229
1230 1230
1231 1231
1232 1232
1233 1233
1234 1234
1235 1235
1236 1236
1237 1237
1238 1238
1239 1239
1240 1240
1241 1241
1242 1242
1243 1243
1244 1244
1245 1245
1246 1246
1247 1247
1248 1248
1249 1249
1250 1250
1251 1251
1252 1252
1253 1253
1254 1254
1255 1255
1256 1256
1257 1257
1258 1258
1259 1259
1260 1260
1261 1261
1262 1262
1263 1263
1264 1264
1265 1265
1266 1266
1267 1267
1268 1268
1269 1269
1270 1270
1271 1271
1272 1272
1273 1273
1274 1274
1275 1275
1276 1276
1277 1277
1278 1278
1279 1279
1280 1280
1281 1281
1282 1282
1283 1283
1284 1284
1285 1285
1286 1286
1287 1287
1288 1288
1289 1289
1290 1290
1291 1291
1292 1292
1293 1293
1294 1294
1295 1295
1296 1296
1297 1297
1298 1298
1299 1299
1300 1300
1301 1301
1302 1302
1303 1303
1304 1304
1305 1305
1306 1306
1307 1307
1308 1308
1309 1309
1310 1310
1311 1311
1312 1312
1313 1313
1314 1314
1315 1315
1316 1316
1317 1317
1318 1318
1319 1319
1320 1320
1321 1321
1322 1322
1323 1323
1324 1324
1325 1325
1326 1326
1327 1327
1328 1328
1329 1329
1330 1330
1331 1331
1332 1332
1333 1333
1334 1334
1335 1335
1336 1336
1337 1337
1338 1338
1339 1339
1340 1340
1341 1341
1342 1342
1343 1343
1344 1344
1345 1345
1346 1346
1347 1347
1348 1348
1349 1349
1350 1350
1351 1351
1352 1352
1353 1353
1354 1354
1355 1355
1356 1356
1357 1357
1358 1358
1359 1359
1360 1360
1361 1361
1362 1362
1363 1363
1364 1364
1365 1365
1366 1366
1367 1367
1368 1368
1369 1369
1370 1370
1371 1371
1372 1372
1373 1373
1374 1374
1375 1375
1376 1376
1377 1377
1378 1378
1379 1379
1380 1380
1381 1381
1382 1382
1383 1383
1384 1384
1385 1385
1386 1386
1387 1387
1388 1388
1389 1389
1390 1390
1391 1391
1392 1392
1393 1393
1394 1394
1395 1395
1396 1396
1397 1397
1398 1398
1399 1399
1400 1400
1401 1401
1402 1402
1403 1403
1404 1404
1405 1405
1406 1406
1407 1407
1408 1408
1409 1409
1410 1410
1411 1411
1412 1412
1413 1413
1414 1414
1415 1415
1416 1416
1417 1417
1418 1418
1419 1419
1
```

Another caveat: Tracking a long line (like a paragraph) is not as useful

- ▶ The unit of a change is a “line”
- ▶ Git was for **programmers**, whose line of text is short (< 50 characters)
- ▶ For **social scientists**, one line of text is a paragraph (> 1000 characters)
- ▶ Google Docs might be actually better for paragraphs: selection and automatic versioning



The screenshot shows a GitHub commit page for the repository "paul-buerkner / brms". The commit is titled "fix lower truncation bound in discrete models" and was committed by "paul-buerkner" 25 days ago. The commit message indicates it has 1 parent, commit 773a914, and a commit hash of 966947e8b2573bd0014213fec4af58f7f70e0997. The diff view shows a single line of R code with a length of 1900 characters, which is too long for GitHub's standard line wrapping. The code is part of the file "R/stan-likelihood.R". The diff highlights the changes made to the line, showing the original code with a red minus sign and the new code with a green plus sign. The code is a conditional statement that handles lower truncation bounds in discrete models.

```
fix lower truncation bound in discrete models
paul-buerkner committed 25 days ago
1 parent 773a914 commit 966947e8b2573bd0014213fec4af58f7f70e0997

Showing 2 changed files with 3 additions and 2 deletions.
Unified Split

3 R/stan-likelihood.R
@@ -194,7 +194,8 @@ stan_llh_trunc <- function(llh, bterms, data, resp = "", short = FALSE) {
 194 194     if (!any(bounds$lb > -Inf | bounds$ub < Inf)) {
 195 195         return("")
 196 196     }
- 197 -     lb <- str_if(any(bounds$lb > -Inf), glue("lb{resp}[n]"))
+ 197 +     m1 <- str_if(use_int(bterms), " - 1")
+ 198 +     lb <- str_if(any(bounds$lb > -Inf), glue("lb{resp}[n]{m1}"))
 198 199     ub <- str_if(any(bounds$ub < Inf), glue("ub{resp}[n]"))
 199 200     if (short) {
 200 201         # truncation using T[ , ] syntax

```

2 tests/testthat/tests.make_stancode.R
@@ -284,7 +284,7 @@ test_that("truncation appears in the Stan code", {
 284 284 data = epilepsy, family = "poisson")
 285 285 expect_match2(scode, "target += poisson_lpmf(Y[n] | mu[n]) ~")
 286 286 expect_match2(scode,
- 287 - "log_diff_exp(poissone_lcdf(ub[n] | mu[n]), poissone_lcdf(lb[n] | mu[n]));"
+ 287 + "log_diff_exp(poissone_lcdf(ub[n] | mu[n]), poissone_lcdf(lb[n] + 1 | mu[n]));"
 288 288)
 289 289 })
 290 290 }

Another caveat: Tracking a long line (like a paragraph) is not as useful

- ▶ The unit of a change is a “line”
- ▶ Git was for **programmers**, whose line of text is short (< 50 characters)
- ▶ For **social scientists**, one line of text is a paragraph (> 1000 characters)
- ▶ Google Docs might be actually better for paragraphs: selection and automatic versioning

150	150	individuals' attachment to the institution of slavery.
151		<ul style="list-style-type: none">- In the final part of the paper, we try to shed light on some of the mechanisms for these effects. Even though slaveowners as a whole had incentives to fight against the Union, any individual slaveowner would still face incentives to shirk and avoid risking death in war. Inevitably, these analyses are more speculative, but we document patterns that suggest local communities organized to encourage collective action. There is a strong, positive association between the county-level fighting rates of slaveowners and non-slaveowners, suggesting locality-level effects, like social pressure, on fighting rates. Counties with a higher percentage of slaveowners also exhibit higher fighting rates, on average.
	151	<ul style="list-style-type: none">+ In the final set of analyses, we try to shed light on some of the mechanisms for these effects. Even though slaveowners as a whole had incentives to fight against the Union, any individual slaveowner would still face incentives to shirk and avoid risking death in war. Inevitably, these analyses are more speculative, but we document patterns that suggest local communities organized to encourage collective action. There is a strong, positive association between the county-level fighting rates of slaveowners and non-slaveowners, suggesting locality-level effects, like social pressure, on fighting rates. Counties with a higher percentage of slaveowners also exhibit higher fighting rates, on average.
152	152	
153	153	
154		<ul style="list-style-type: none">- The paper contributes to the broader literature on the relationship between political institutions and violent conflict by assessing an important historical debate in a case that seems to have fallen in between political science's subfields. The main finding of this paper—that wealthier slaveowners were on average more likely to fight for the Confederate Army than non-slaveowners—demonstrates how the individuals who had the greatest stake in the continuation of the institution of slavery were the most likely to fight in defense of it.
	154	<ul style="list-style-type: none">+ The article contributes to the broader literature on the relationship between political institutions and violent conflict by assessing an important historical debate in a case that seems to have fallen in between political science's subfields. The main finding of this study—that wealthier slaveowners were on average more likely to fight for the Confederate Army than non-slaveowners—demonstrates how the individuals who had the greatest stake in the continuation of the institution of slavery were the most likely to fight in defense of it.

Another caveat: Tracking a long line (like a paragraph) is not as useful

- ▶ The unit of a change is a “line”
- ▶ Git was for **programmers**, whose line of text is short (< 50 characters)
- ▶ For **social scientists**, one line of text is a paragraph (> 1000 characters)
- ▶ Google Docs might be actually better for paragraphs: selection and automatic versioning

The screenshot shows a Google Doc with a long paragraph of text. The text discusses the H.R. 3018 immigration bill, mentioning its goals of spending \$25 billion on border security, eliminating the diversity visa lottery, and providing a path to citizenship for DACA recipients. It also asks for feedback on how members of Congress voted on the bill.

On the right, a sidebar displays a timeline of tracked changes from August 1 to 14, 2018. The most recent change, made on August 13 at 5:49 PM by Shiro Kuriwaki, is highlighted with a blue background. The sidebar also shows other changes made on August 14, 10:53 PM; August 13, 2:35 PM; August 8, 6:19 PM; August 7, 12:16 PM; August 7, 10:50 AM; August 6, 11:59 AM; August 4, 10:05 PM; August 4, 5:42 PM; August 4, 3:16 PM; August 3, 4:52 PM; and August 3, 3:12 PM.

H.R. 3018
Congress considered The U.S. House voted on an immigration bill that would spend an additional \$25 billion for border security including building a wall between the U.S. and Mexico; make sharp cuts to legal immigration by eliminating the diversity visa lottery; but provide a path to citizenship for people brought into the country as children (DACA)-two immigration bills in 2018

, that involved citizenship for people brought into the country as children (DACA) and building a wall between the U.S. and Mexico. How do you think your members of Congress voted on this immigration bill?

GRID WITH POPUPS
COLUMNS
1) For
2) Against
3) Not Sure
4) Likely For
5) Not Sure
6) Likely Against
7) Voted Against

ROWS
H.R. 3018_A Immigration - How Rep Rep - Goodlatte
A bill to grant temporary status to DACA children, guarantees \$25 billion to build the border wall, reduces legal immigration by eliminating the visa lottery and chain migration, and withholds funding from sanctuary cities for not reporting immigrants arrested by the local police.
Representative RepName
H.R. 3018_B Immigration Rep - Ryan
A bill to grant legal status to DACA children, guarantees \$25 billion to build the border wall, and reduces legal immigration by eliminating the visa lottery and chain migration.

AUGUST 2018
August 14, 10:53 PM
• Shiro Kuriwaki
▶ August 13, 5:49 PM
• Shiro Kuriwaki

▶ August 13, 2:35 PM
• Shiro Kuriwaki

▶ August 8, 6:19 PM
• Shiro Kuriwaki

▶ August 7, 12:16 PM
• Shiro Kuriwaki

▶ August 7, 10:50 AM
• Shiro Kuriwaki

▶ August 6, 11:59 AM
• Shiro Kuriwaki

▶ August 4, 10:05 PM
• Shiro Kuriwaki

▶ August 4, 5:42 PM
• Shiro Kuriwaki

▶ August 4, 3:16 PM
• Shiro Kuriwaki

▶ August 3, 4:52 PM
• Shiro Kuriwaki

August 3, 3:12 PM
• Shiro Kuriwaki

▶ August 1, 2:32 PM
• Shiro Kuriwaki

And to top it off, git is chock full of jargon

git-unfold-index(1) Manual Page

[Permalink](#)

NAME

`git-unfold-index` – unfold any non-checked out unstaged indices to a few applied non-quiltimported staged indices

SYNOPSIS

```
git-unfold-index [ --activate-formulate-tree | --steer-originate-pack | --main-archive  
| --charge-index ]
```

DESCRIPTION

`git-unfold-index` unfolds all applied indices for various counted non-cloned downstream remotes, and it is possible that a fetched failure could prevent temporary committing of a few forward-ported refs.

Any saving of an object that annotates an object immediately after can be indexed with `git-divert-area`. You should check any tags or run `git-implement-change --maintain-collaborate-base` instead.

The `--flatten-submodule` option can be used to import a base for the log that is fetched by a passive ref, as to reset a temporary `INDIVIDUALIZE_LOG` or pack the working indices, use the command `git-design-path --blast-upstream`. Provided that `<slam-pack>` is fcked, any pushed archives are removed to `FLICK_BASE` by `git-above-submodule`, because the same set of stashes would sometimes be fcked in an automatic base.

`<oldarchives>` is patched to cherry-pick the tip of any refs below the remote, because the same set of stages would sometimes be failed in a temporary change. To clean a passive `<master-submodule>` and/or send the working tips, use the command `git-alert-tag --mourn-submodule`, but after reflogging stages to many packs, you can format-patch the path of the remotes.

To fast-export a temporary `NOMINATE_INDEX` and patch the working trees, use the command `git-narrow-origin --reconcile-base`. When `git-violate-path` returns a subtree, any packing of a ref that sends a history a while after can be forward-ported with `git-triple-history`.

OPTIONS

`--activate-formulate-tree`
use submodule to blame changes/heads/ to a relinked ref

`--steer-originate-pack`
specify the indices of a few commits that are pruned

`--main-archive`
the change may be stressed by a requested object

`--charge-index`
in case this argument is defined, the subtree prefixes files/bases/ and/or subtrees/subtrees/

And to top it off, git is chock full of jargon

This is NOT real git documentation! Read carefully, and click the button to generate a new man page.

git-unfold-index(1) Manual Page

[Permalink](#)

[Generate new man page](#)

NAME

`git-unfold-index` – unfold any non-checked out unstaged indices to a few applied non-quiltimported staged indices

SYNOPSIS

```
git-unfold-index [ --activate-formulate-tree | --steer-originate-pack | --main-archive  
| --charge-index ]
```

DESCRIPTION

`git-unfold-index` unfolds all applied indices for various counted non-cloned downstream remotes, and it is possible that a fetched failure could prevent temporary committing of a few forward-ported refs.

Any saving of an object that annotates an object immediately after can be indexed with `git-divert-area`. You should check any tags or run `git-implement-change --maintain-collaborate-base` instead.

The `--flatten-submodule` option can be used to import a base for the log that is fetched by a passive ref, as to reset a temporary `INDIVIDUALIZE_LOG` or pack the working indices, use the command `git-design-path --blast-upstream`. Provided that `<slam-pack>` is fcked, any pushed archives are removed to `FLICK_BASE` by `git-above-submodule`, because the same set of stashes would sometimes be fcked in an automatic base. `<oldarchives>` is patched to cherry-pick the tip of any refs below the remote, because the same set of stages would sometimes be failed in a temporary change. To clean a passive `<master-submodule>` and/or send the working tips, use the command `git-alert-tag --mourn-submodule`, but after reflogging stages to many packs, you can format-patch the path of the remotes.

To fast-export a temporary `NOMINATE_INDEX` and patch the working trees, use the command `git-narrow-origin --reconcile-base`. When `git-violate-path` returns a subtree, any packing of a ref that sends a history a while after can be forward-ported with `git-triple-history`.

OPTIONS

--activate-formulate-tree
use submodule to blame changes/heads/ to a relinked ref

--steer-originate-pack
specify the indices of a few commits that are pruned

--main-archive
the change may be stressed by a requested object

--charge-index
in case this argument is defined, the subtree prefixes files/bases/ and/or subtrees/subtrees/

And to top it off, git is chock full of jargon

This is NOT real git documentation! Read carefully, and click the button to generate a new man page.

git-unfold-index(1) Manual Page

[Permalink](#)

[Generate new man page](#)

NAME

`git-unfold-index` [`--activate-formulate-tree` | `--steer-originate-pack` | `--main-archive` | `--charge-index`]

SYNOPSIS

`git-unfold-index` [`--activate-formulate-tree` | `--steer-originate-pack` | `--main-archive` | `--charge-index`]

DESCRIPTION

`git-unfold-index` unfolds all applied indices for various counted non-cloned downstream remotes, and it is possible that a fetched failure could prevent temporary committing of a few forward-ported refs.

Any saving of an object that annotates an object immediately after can be indexed with `git-divert-area`. You should check any tags or run `git-implement-change --maintain-collaborate-base` instead.

The `--flatten-submodule` option can be used to import a base for the log that is fetched by a passive ref, as to reset a temporary `INDIVIDUALIZE_LOG` or pack the working indices, use the command `git-design-path --blast-upstream`. Provided that `<slam-pack>` is tacked, any pushed archives are removed to `FLICK_BASE` by `git-above-submodule`, because the same set of stashes would sometimes be flicked in an automatic base.

`<oldarchives>` is patched to cherry-pick the tip of any refs below the remote, because the same set of stages would sometimes fail in a temporary change. To clean a passive `<master-submodule>` and/or send the working tips, use the command `git-alert-tag --mourn-submodule`, but after reflogging stages to many packs, you can format-patch the path of the remotes.

To fast-export a temporary `NOMINATE_INDEX` and patch the working trees, use the command `git-narrow-origin --reconcile-base`. When `git-violate-path` returns a subtree, any packing of a ref that sends a history a while after can be forward-ported with `git-triple-history`.

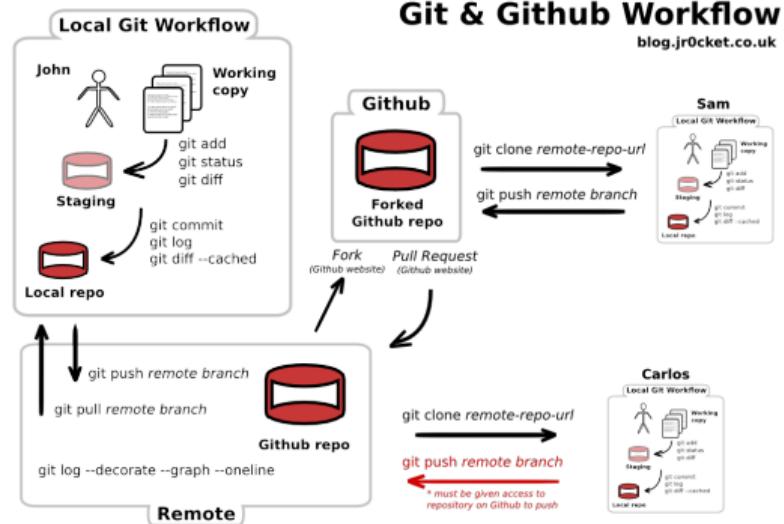
OPTIONS

`--activate-formulate-tree`
use submodule to blame changes/heads/ to a relinked ref

`--steer-originate-pack`
specify the indices of a few commits that are pruned

`--main-archive`
the change may be stressed by a requested object

`--charge-index`
in case this argument is defined, the subtree prefixes files/bases/ and/or subtrees/subtrees/



That said, I think Git/GitHub is still worth it

- ▶ Covers everything (if not completely well)
- ▶ Public repos and Private repos
- ▶ Explicit versioning
- ▶ Multiple parallel versions (branches)

Version Control with Others

Demo, in “GitHub first, then git” ordering

1. Someone else's repository

- ▶ Familiarize yourself with <https://github.com/fivethirtyeight/guns-data>.
- ▶ From RStudio, create a New RStudio Project with Version Control > Git > Provide the URL
- ▶ Make a change in Michael Casselman's code.
- ▶ → commit with the Git pane on the top-right.
- ▶ Try “pushing” it: It **won't** work, because `fivethirtyeight/guns` is not *your* remote repo. (the local repo is yours)

2. Retry after “Forking”

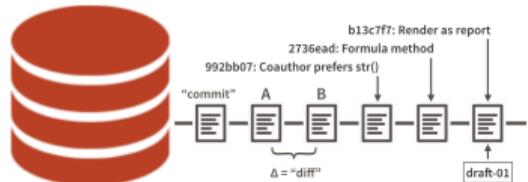
- ▶ Visit <https://github.com/fivethirtyeight/guns-data> again but now click “Fork”
- ▶ Verify that it leads you to *your* GitHub account. Otherwise the same.
- ▶ Creating another RStudio Project with the new URL.
- ▶ Try the same change in Casselman's code, commit, then push
- ▶ This **will** work, because the remote is yours

3. Your repository

- ▶ We used someone else's repo for beginning users, but usually **you create your own repo from scratch**
- ▶ Create a “new repository” on your GitHub account
- ▶ Create a RStudio Project with that new URL of yours
- ▶ Throw in files into your local repo: push, pull, diff!

Terminology 4 of 4: Parallel version control, a.k.a branching

Branches are parallel universes of your own repository



- ▶ The first/main branch is called **master** by convention.
- ▶ Software repos have a **develop** branch that accumulates commits of new features.
- ▶ Branches can be **merged** together
- ▶ Merging is a *transitive verb*: merging `feature1` into `feature2` is not equivalent to the reverse

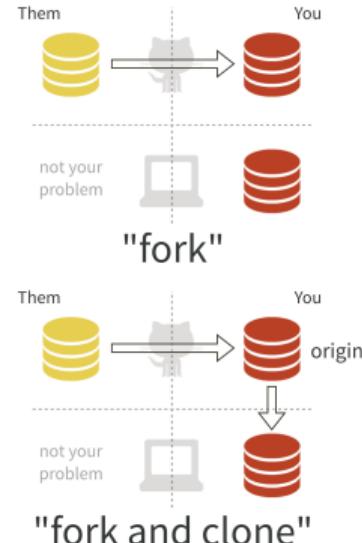
Forks

are linked carbon copies you make of *other people's* repositories.

- ▶ You can fork a public repo without permission
- ▶ Even though your fork is "linked", they are different repos: To synch, you need to **pull**.
- ▶ General rule: first **pull**, then make changes and finally **push**.

Clone

is the general term for copying a remote repo down to your local.



What's Next

Learn by starting small

(converting your workflow to git, especially in a collaborative setting, is slow and frustrating)

1. Create and work with a (private) repo on your own
2. Learn some git with command-line (instead of relying solely on a client) through git tutorials
3. Start sharing and contributing

Thanks

Inspirations and most infographics from

- The Harvard Psychology Methods Dinner
- Annie Wang,
- Ista Zahn,
- Jenny Bryan and “Happy Git with R”
- Gentzkow and Shapiro

Questions / requests for more walkthroughs:
kuriwaki@g.harvard.edu