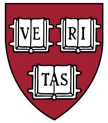


Research software review as part of the publication process

SciCodes - Consortium of scientific software registries and repositories
October 21, 2021

Ana Trisovic, Harvard University

Why review research code?

- Necessary for research verification
- Critical for reproducibility
- It can enhance transparency of research
- Facilitates reuse and building upon previous results

Presentation agenda

Researchers

Research data
and code

Research data or
software repositories

Reviewers

Journals

Institutes

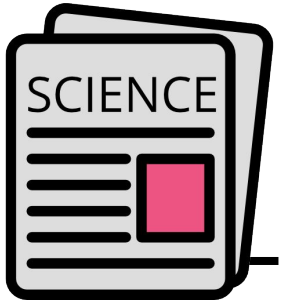
Presentation agenda

Researchers

Research data and code

Research data or software repositories

Reviewers



Journals

Institutes

What can repositories do to facilitate the code review process?

Research code on Harvard Dataverse

```
### TABLE 4: Proportion of MEPs Giving a Speech as a Function of
### Voting and Candidate Selection Mechanisms
### (Vote-speech sample from 6th EP, 2004-2005)

        table(natrebel,centralizedcandidateselection,gavespeech)

### TABLE 4: two-sample test of proportions

        t.test(gavespeech[natrebel==1 & centralizedcandidateselection==0],
        gavespeech[natrebel==0 & centralizedcandidateselection==0],alternative="two.sided",
        var.equal=TRUE,conf.level=0.95)

        t.test(gavespeech[natrebel==1 & centralizedcandidateselection==1],
        gavespeech[natrebel==0 ¢ralizedcandidateselection==1],alternative="two.sided",
        var.equal=TRUE,conf.level=0.95)

### WEBAPPENDIX Table 1: Relationship between Voting Decisions
### and Legislative Speeches (Vote-speech sample from 6th EP, 2004-2005)

        table(natrebel,gavespeech)


#### WEBAPPENDIX Table 1: two-sample test of proportions

        t.test(gavespeech[natrebel==0],gavespeech[natrebel==1],
        alternative="two.sided",var.equal=TRUE,conf.level=0.95)


detach(ep6data)



################################
#### AGGREGATE ANALYSIS EP5 #####
################################


attach(ep5data)


######## TABLE 5: Negative Binomial MODEL A
model.A<-zelig(epwebsitespeechcount~ pernpepg+ perepgnp+  tenure+absent+rapporteurshipcount+ partyleader +epleader+ numcommem+ numcomle
ader+groupsize+ natpartyperc +candidateselection+  pernpepg:candidateselection,model="negbin",data= ep5data ,save.data = TRUE)
summary(model.A)
```

```
#Generate new indicators and fill in values
data$gate<-NA
data$building<-NA
data$split<-NA
data$multi<-NA
data$country_1<-NA
data$country_2<-NA

data$borderid[data$odd==1]<-paste(data$borderid[data$odd==1],"_1",sep="")
data$gate[data$odd==1]<-data$gate_1[data$odd==1]+1
data$building[data$odd==1]<-data$numbld_1[data$odd==1]+1
data$split[data$odd==1]<-data$split_1[data$odd==1]+1
data$multi[data$odd==1]<-data$multilane_1[data$odd==1]+1
data$country_1[data$odd==1]<-data$country1[data$odd==1]
data$country_2[data$odd==1]<-data$country2[data$odd==1]

data$borderid[data$odd==0]<-paste(data$borderid[data$odd==0],"_2",sep="")
data$gate[data$odd==0]<-data$gate_2[data$odd==0]+1
data$building[data$odd==0]<-data$numbld_2[data$odd==0]+1
data$split[data$odd==0]<-data$split_2[data$odd==0]+1
data$multi[data$odd==0]<-data$multilane_2[data$odd==0]+1
data$country_1[data$odd==0]<-data$country2[data$odd==0]
data$country_2[data$odd==0]<-data$country1[data$odd==0]

data$country1<-data$country_1
data$country2<-data$country_2
data$country_1<-NULL
data$country_2<-NULL

####
#1.2: Interpolating missing values
####
data<-arrange(data,borderid,coder,assignment,year)
for(jj in 41:44){
  for(ii in 2:nrow(data)){
    data[ii,jj]<-ifelse(is.na(data[ii,jj]) & !is.na(data[ii-1,jj]) &
                          data$borderid[ii]==data$borderid[ii-1] &
                          data$coder[ii]==data$coder[ii-1] &
                          data$assignment[ii]==data$assignment[ii-1],
                        data[ii-1,jj],data[ii,jj])
  }}
data<-arrange(data,borderid,coder,assignment,-year)
for(jj in 41:44){
  for(ii in 2:nrow(data)){
    data[ii,jj]<-ifelse(is.na(data[ii,jj]) & !is.na(data[ii-1,jj]) &
                          data$borderid[ii]==data$borderid[ii-1] &
                          data$coder[ii]==data$coder[ii-1] &
                          data$assignment[ii]==data$assignment[ii-1],
                        data[ii-1,jj],data[ii,jj])
```

Research code on Harvard Dataverse

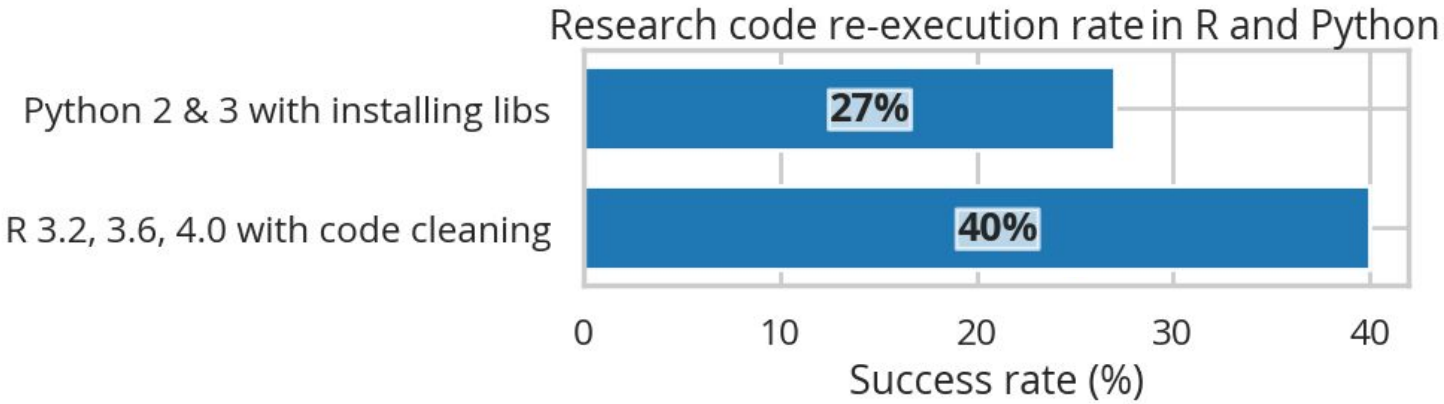
```
### TABLE 4: Proportion of MEPs Giving a Speech as a Function of

#effect of isp (through proportion renewable, electricity per gdp [itself direct and through price], and direct effect)
no.isp.price<-mod1$coef[1]+mod1$coef[2]*mean(POLICYII$rpsprop)+mod1$coef[3]*0+mod1$coef[4]*mean(POLICYII$ipp)+mod1$coef[5]*0+mod1$coef[6]*mean(POLICYII$gasPrice);no.isp.price
no.isp.renew<-pnorm(mod2$coef[1]+mod2$coef[2]*mean(POLICYII$rpsprop)+mod2$coef[3]*0+mod2$coef[4]*mean(POLICYII$ipp)+mod2$coef[5]*0+mod2$coef[6]*mean(POLICYII$gdp)+mod2$coef[7]*mean(POLICYII$gasPrice));no.isp.renew
no.isp.eGDP<-mod3$coef[1]+mod3$coef[2]*0+mod3$coef[3]*mean(POLICYII$ipp)+mod3$coef[4]*0+mod3$coef[5]*no.isp.price+mod3$coef[6]*mean(POLICYII$gasPrice);no.isp.eGDP
no.isp<-exp(mod4$coef[1]+mod4$coef[2]*mean(POLICYII$rpsprop)+mod4$coef[3]*0+mod4$coef[4]*mean(POLICYII$ipp)+mod4$coef[5]*0+mod4$coef[6]*mean(POLICYII$gdp)+mod4$coef[7]*no.isp.renew+mod4$coef[8]*no.isp.eGDP+mod4$coef[9]*mean(POLICYII$gasPrice))

yes.isp.price<- mod1$coef[1]+mod1$coef[2]*mean(POLICYII$rpsprop)+mod1$coef[3]*0+mod1$coef[4]*mean(POLICYII$ipp)+mod1$coef[5]*1+mod1$coef[6]*mean(POLICYII$gasPrice);yes.isp.price
yes.isp.renew<-pnorm(mod2$coef[1]+mod2$coef[2]*mean(POLICYII$rpsprop)+mod2$coef[3]*0+mod2$coef[4]*mean(POLICYII$ipp)+mod2$coef[5]*1+mod2$coef[6]*mean(POLICYII$gdp)+mod2$coef[7]*mean(POLICYII$gasPrice));yes.isp.renew
yes.isp.eGDP<- mod3$coef[1]+mod3$coef[2]*0+mod3$coef[3]*mean(POLICYII$ipp)+mod3$coef[4]*0+mod3$coef[5]*yes.isp.price+mod3$coef[6]*mean(POLICYII$gasPrice);yes.isp.eGDP
yes.isp<-exp(mod4$coef[1]+mod4$coef[2]*mean(POLICYII$rpsprop)+mod4$coef[3]*0+mod4$coef[4]*mean(POLICYII$ipp)+mod4$coef[5]*1+mod4$coef[6]*mean(POLICYII$gdp)+mod4$coef[7]*yes.isp.renew+mod4$coef[8]*yes.isp.eGDP+mod4$coef[9]*mean(POLICYII$gasPrice))

no.isp;yes.isp
100*(no.isp-yes.isp)/no.isp

        gavespeech[natrebel==0 &ralizedcandidateselection==1],alternative="two.sided",
        var.equal=TRUE,conf.level=0.95)

### WEBAPPENDIX Table 1: Relationship between Voting Decisions
### and Legislative Speeches (Vote-speech sample from 6th EP, 2004-2005)

        table(natrebel,gavespeech)


#### WEBAPPENDIX Table 1: two-sample test of p

        t.test(gavespeech[natrebel==0],gavespe
        alternative="two.sided",var.equal=TRUE


detach(ep6data)



################################
#### AGGREGATE ANALYSIS EP5 #####
################################


attach(ep5data)


######## TABLE 5: Negative Binomial MODEL A
model.A<-zelig(epwebsitespeechcount~ pernpepg+
ader+groupsize+ natpartyperc +candidateselecti
summary(model.A)
```

```
se.est.plot <- c(sd(brazil.data$voteintent[brazil.data$education < 4 & brazil.data$cred_vs_less==0], na.rm=T)/
  sqrt(sum(is.na(brazil.data$voteintent[brazil.data$education < 4 & brazil.data$cred_vs_less==0])==0)),
 sd(brazil.data$voteintent[brazil.data$education < 4 & brazil.data$cred_vs_less==1], na.rm=T)/
  sqrt(sum(is.na(brazil.data$voteintent[brazil.data$education < 4 & brazil.data$cred_vs_less==1])==0)),
 sd(brazil.data$voteintent[brazil.data$education==4 & brazil.data$cred_vs_less==0], na.rm=T)/
  sqrt(sum(is.na(brazil.data$voteintent[brazil.data$education==4 & brazil.data$cred_vs_less==0])==0)),
 sd(brazil.data$voteintent[brazil.data$education==4 & brazil.data$cred_vs_less==1], na.rm=T)/
  sqrt(sum(is.na(brazil.data$voteintent[brazil.data$education==4 & brazil.data$cred_vs_less==1])==0)),
 sd(argentina.data$voteintent[argentina.data$education < 3 & argentina.data$cred_vs_less==0], na.rm=T)/
  sqrt(sum(is.na(argentina.data$voteintent[argentina.data$education < 3 & argentina.data$cred_vs_less==0])==0)),
 sd(argentina.data$voteintent[argentina.data$education < 3 & argentina.data$cred_vs_less==1], na.rm=T)/
  sqrt(sum(is.na(argentina.data$voteintent[argentina.data$education < 3 & argentina.data$cred_vs_less==1])==0)),
 sd(argentina.data$voteintent[argentina.data$education==3 & argentina.data$cred_vs_less==0], na.rm=T)/
  sqrt(sum(is.na(argentina.data$voteintent[argentina.data$education==3 & argentina.data$cred_vs_less==0])==0)),
 sd(argentina.data$voteintent[argentina.data$education==3 & argentina.data$cred_vs_less==1], na.rm=T)/
  sqrt(sum(is.na(argentina.data$voteintent[argentina.data$education==3 & argentina.data$cred_vs_less==1])==0)))
```

```
#Generate new indicators and fill in values

data$country_2[data$odd==1]<-data$country2[data$odd==1]

data$borderid[data$odd==0]<-paste(data$borderid[data$odd==0],"_2",sep="")
data$gate[data$odd==0]<-data$gate_2[data$odd==0]+1
data$building[data$odd==0]<-data$numbld_2[data$odd==0]+1
data$split[data$odd==0]<-data$split_2[data$odd==0]+1
data$multi[data$odd==0]<-data$multilane_2[data$odd==0]+1
data$country_1[data$odd==0]<-data$country2[data$odd==0]
data$country_2[data$odd==0]<-data$country1[data$odd==0]

a[ii-1,jj]) &
erid[ii-1] &
-1] &
signment[ii-1],

a[ii-1,jj]) &
erid[ii-1] &
-1] &
data$assignment[ii-1],
data[ii-1,jj],data[ii,jj])
```

Research code re-execution

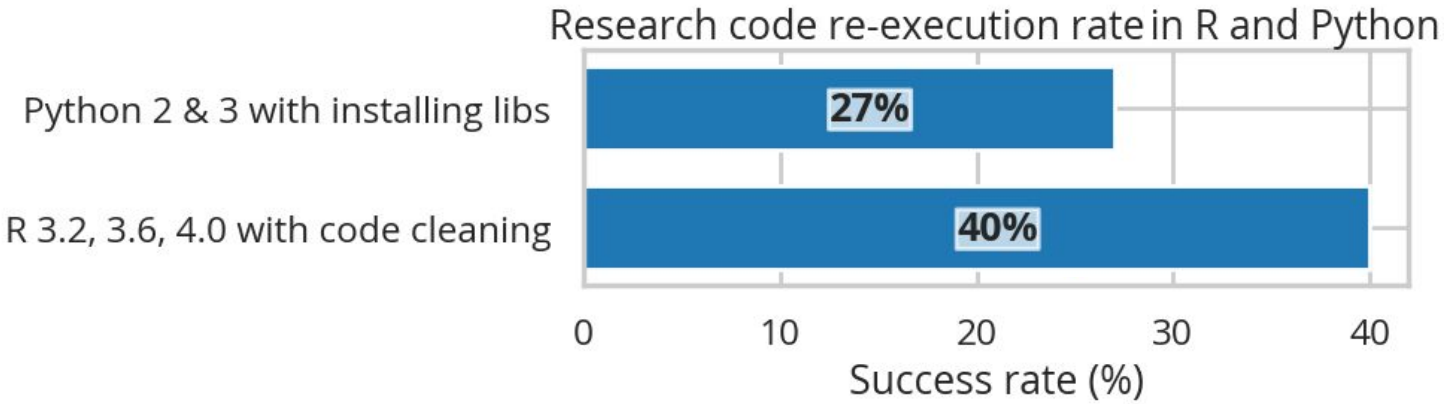
Most code files fail when re-executed out-of-the-box, even with the pre-installation of used libraries [1,2].



[1] Trisovic, Ana, et al. "Repository Approaches to Improving Quality of Shared Data and Code." Data 6.2 (2021): 15.
[2] Trisovic, Ana, et al. "A large-scale study on research code quality and execution." arXiv preprint arXiv:2103.12793 (2021).

Research code re-execution

Most code files fail when re-executed out-of-the-box, even with the pre-installation of used libraries [1,2].

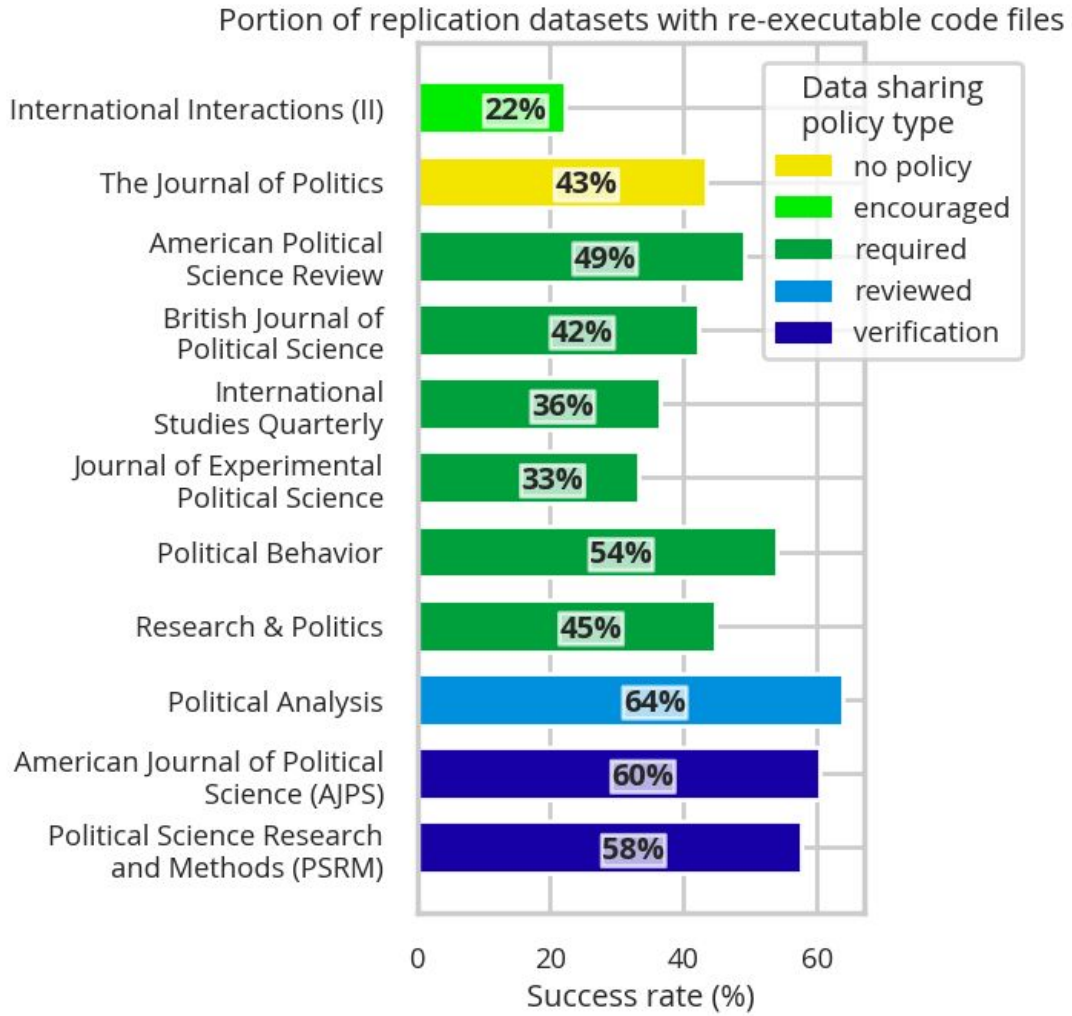


It is hard to re-execute “old” code, but many common errors can be avoided!

[1] Trisovic, Ana, et al. "Repository Approaches to Improving Quality of Shared Data and Code." Data 6.2 (2021): 15.
[2] Trisovic, Ana, et al. "A large-scale study on research code quality and execution." arXiv preprint arXiv:2103.12793 (2021).

Journals with stricter data policies have higher rate of executable code

Had some code review!



Can we learn from industry code review?

Research code vs. industry code

Industry development teams use advanced tools to facilitate code review such as continuous integration and containers

In most cases, students and early-career researchers write research code (sometimes completely new to programming)

Research code vs. industry code

Industry development teams use advanced tools to facilitate code review such as continuous integration and containers

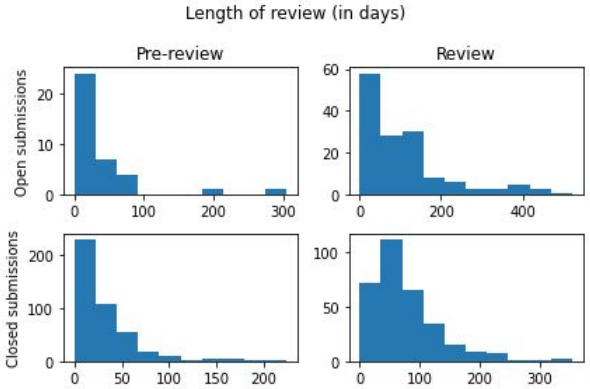
In most cases, students and early-career researchers write research code (sometimes completely new to programming)

Volunteers review research code (seeing it for the first time), while development team members (who are already familiar with software) review new (small) code contributions

Lifetime of research code is less than industry code so there are less incentives to keep it clean

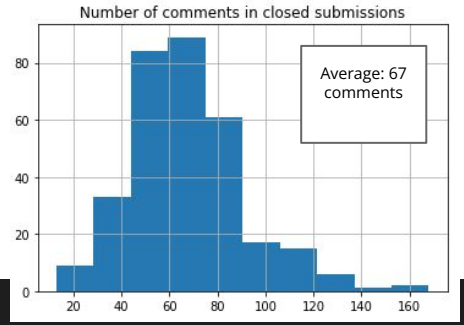
Publishing research software with the Journal of Open Source Software (JOSS)

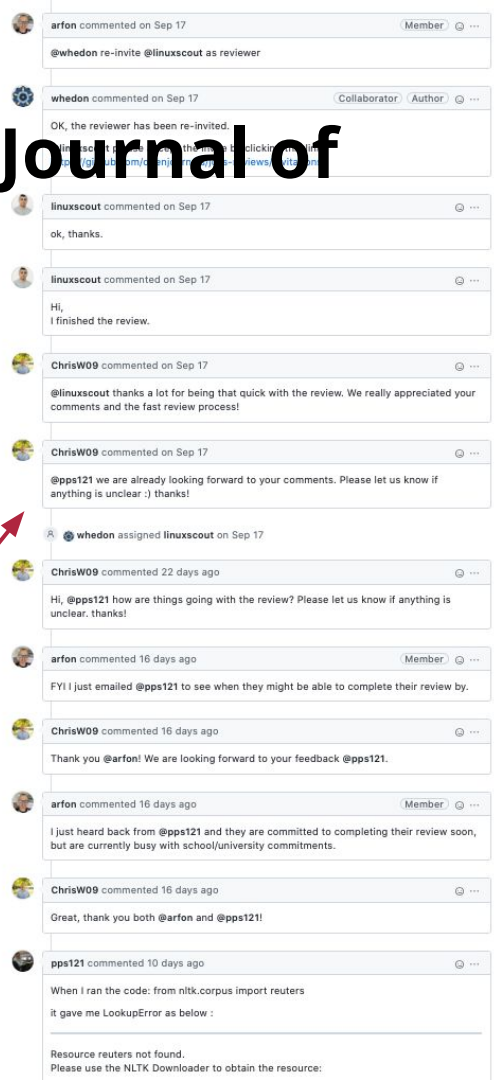
- Software as a recognized output of research
- Retrieved 1000 closed and 182 open issues (paper submission and review) containing 25,382 comments.



Mean review time 30.5 + 77.4 = 107.9 days
Median review time 21.0 + 63.0 = 84.0 days

Assigning reviewers and the review is happening in GitHub issue comments

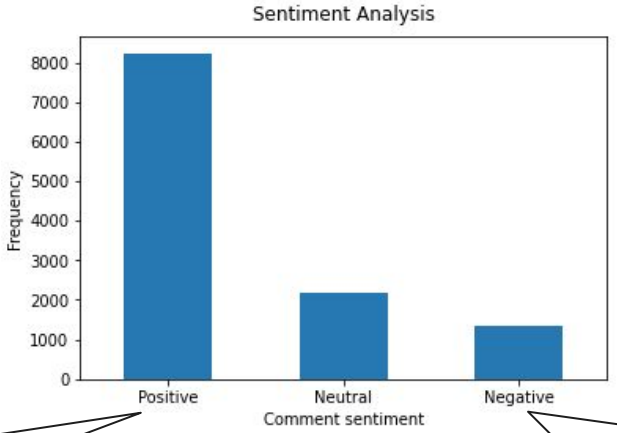


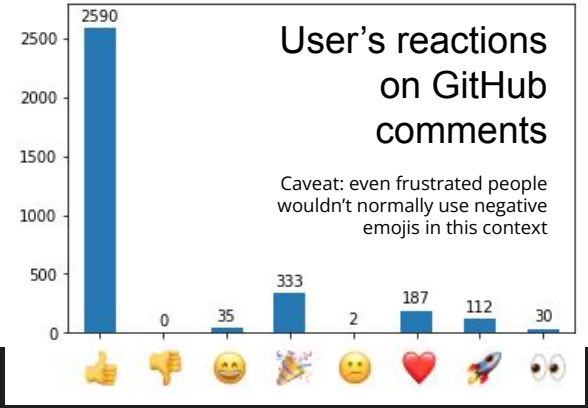


https://github.com/atrisovic/joss-reviews

How reviewers feel about reviewing software?

- Basic NLP analysis of the content of JOSS issue comments in completed submissions.
  - Comments by the bot @whedon were excluded.
- The sentiment seems mostly positive!





xuanxu: Great! Thank you both!
bradkav: I'm also now happy to recommend `sntools` for publication. Well done !
neuromusic: Changes look great!
Yurlungur: Perfect!

danielskatz: This may have just been very bad timing
SteveMacenski: Hi, just wanted to touch base on this - any progress?
simonom: I have completed my review, but my invitation to review has expired so I can't check off the checklist :(
sgrieve: I'm afraid I don't have any capacity this month. Apologies!

Word cloud of JOSS submission comments (reviews)

Ideas to make code review easy at research repositories

1) Checklist for code reviewers / data curators

Review checklist

✨ Important: Please do not use the *Convert to issue* functionality when working through this checklist, instead, please open any new issues associated with your review in the software repository associated with the submission. ✨

Conflict of interest

- [ ] I confirm that I have read the JOSS conflict of interest (COI) policy and that: I have no COIs with reviewing this work or that any perceived COIs have been waived by JOSS for the purpose of this review.

Code of Conduct

- [ ] I confirm that I read and will adhere to the JOSS code of conduct.

General checks

- [ ] **Repository:** Is the source code for this software available at the repository url?
- [ ] **License:** Does the repository contain a plain-text LICENSE file with the contents of an OSI approved software license?
- [ ] **Contribution and authorship:** Has the submitting author (**@hcdenbakker**) made major contributions to the software? Does the full list of paper authors seem appropriate and complete?
- [ ] **Substantial scholarly effort:** Does this submission meet the scope eligibility described in the JOSS guidelines

Functionality

- [ ] **Installation:** Does installation proceed as outlined in the documentation?
- [ ] **Functionality:** Have the functional claims of the software been confirmed?
- [ ] **Performance:** If there are any performance claims of the software, have they been

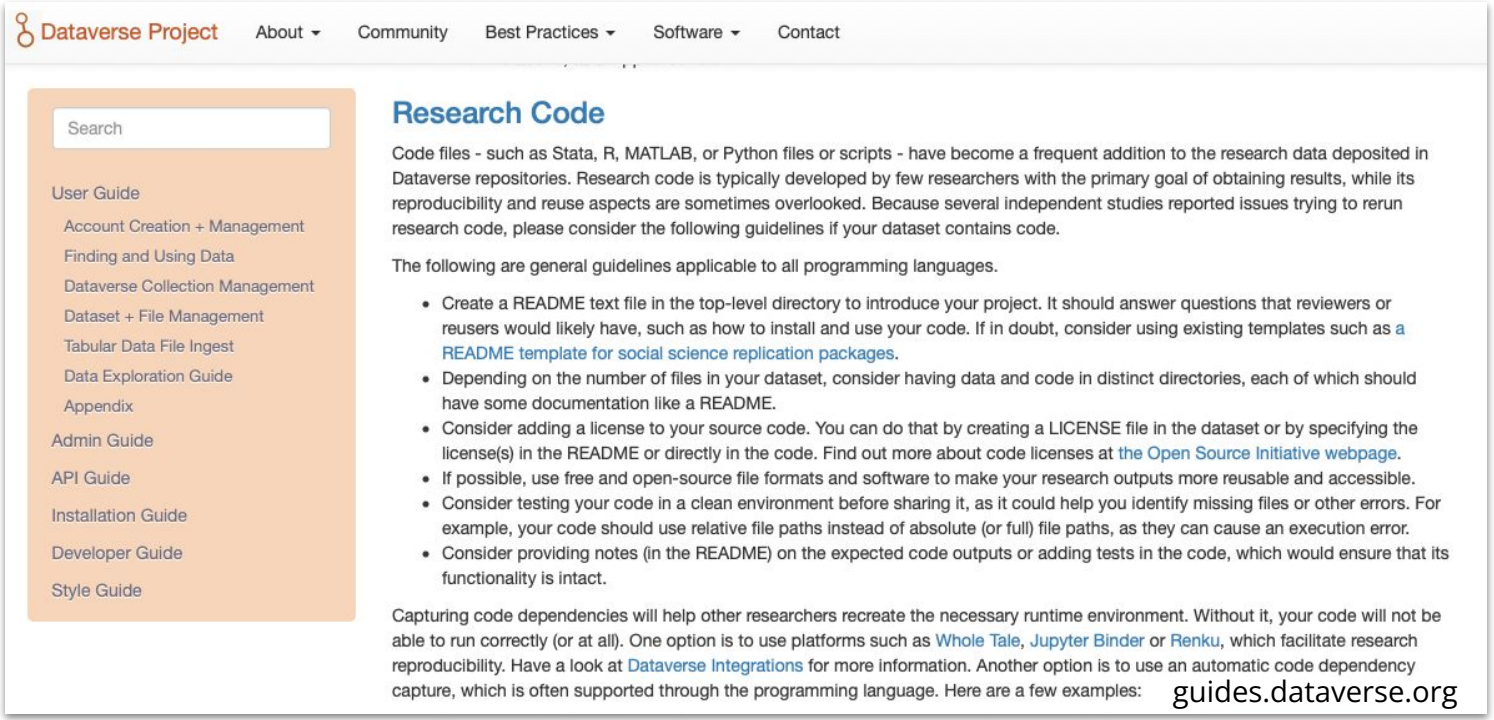
Documentation

- [ ] **A statement of need:** Do the authors clearly state what problems the software is designed to solve and who the target audience is?
- [ ] **Installation instructions:** Is there a clearly-stated list of dependencies? Ideally these should be handled with an automated package management solution.
- [ ] **Example usage:** Do the authors include examples of how to use the software (ideally to solve real-world analysis problems).
- [ ] **Functionality documentation:** Is the core functionality of the software documented to a satisfactory level (e.g., API method documentation)?
- [ ] **Automated tests:** Are there automated tests or manual steps described so that the functionality of the software can be verified?
- [ ] **Community guidelines:** Are there clear guidelines for third parties wishing to 1) Contribute to the software 2) Report issues or problems with the software 3) Seek support

Software paper

- [ ] **Summary:** Has a clear description of the high-level functionality and purpose of the software for a diverse, non-specialist audience been provided?
- [ ] **A statement of need:** Does the paper have a section titled 'Statement of Need' that clearly states what problems the software is designed to solve and who the target audience is?
- [ ] **State of the field:** Do the authors describe how this software compares to other commonly-used packages?
- [ ] **Quality of writing:** Is the paper well written (i.e., it does not require editing for structure, language, or writing quality)?
- [ ] **References:** Is the list of references complete, and is everything cited appropriately that should be cited (e.g., papers, datasets, software)? Do references in the text use the proper citation syntax?

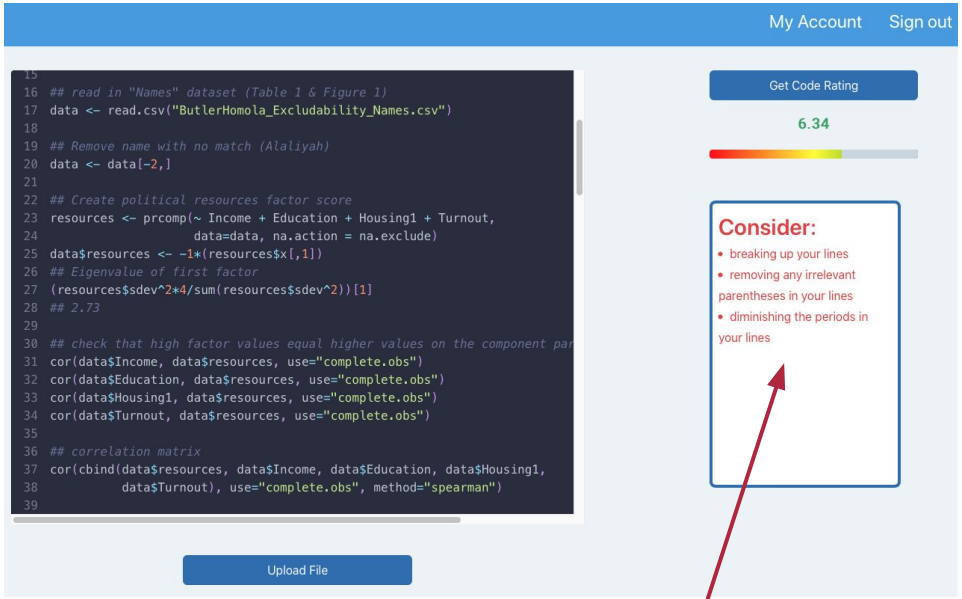
2) Guidelines for code depositors



guides.dataverse.org

3) Automatic code style assessment

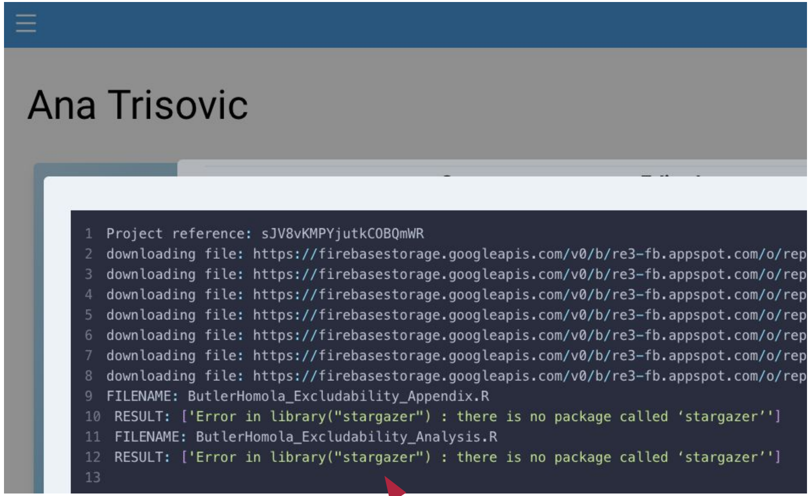
- Automatic code style assessment informs the depositors of the readability of their code.
- There are existing code formatting tools that can be encouraged or recommended (`lintR`, `pycodestyle`, even in-browser tools).



Bahaidarah, Layan, et al. "Toward Reusable Science with Readable Code and Reproducibility." (arXiv:2109.10387)

4) Automatic re-execution (for free software and small-scale studies)

- Enabling an automatic code re-execution test could fast identify missing files and other common errors.
- It could help with the documentation of analysis flow as the researchers would need to add re-execution commands.
- Code would run out-of-the-box for reviewers and future reusers.

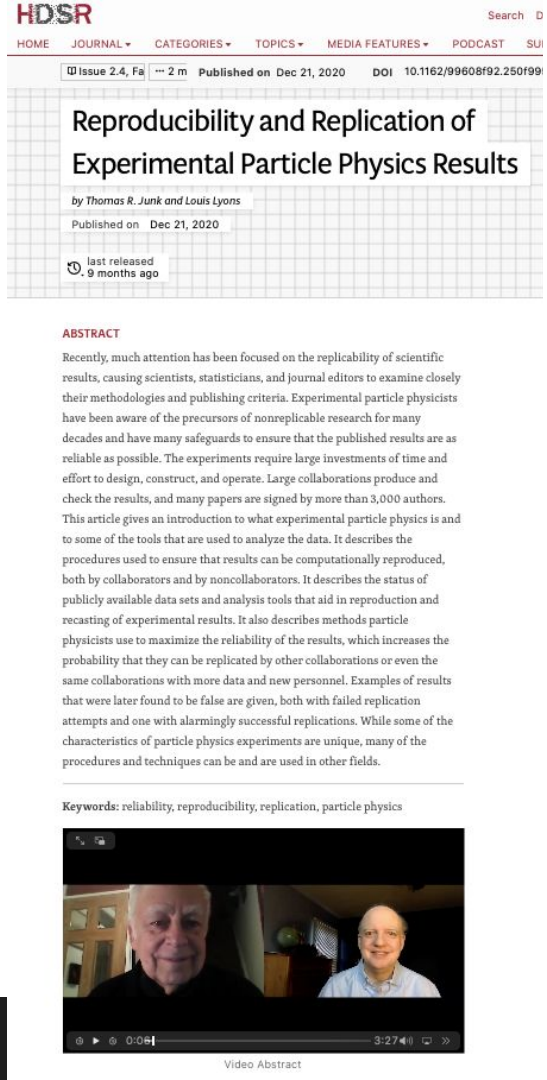


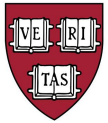
Code re-execution test at upload

Bahaidarah, Layan, et al. "Toward Reusable Science with Readable Code and Reproducibility." (arXiv:2109.10387)

5*) “Walk me through your code” video

- Media files as part of publication (video summaries of articles (i.e., HDSR))
- Creating a video presentation of code could take a few hours for its creator, but probably save twice as much time for each code reviewer and reuser
- Creating video presentations is easy with Zoom!





Thank you! Questions?

Email: anatrisovic@g.harvard.edu
GitHub & Twitter: atrisovic