# Strategies for achieving reproducible research

ANTdrew D. Nguyen,
Evolutionary Physiologist

*NATURE* | CORRESPONDENCE

# Research papers: Journals should drive data reproducibility

**Gregorio Santori**

*Nature* **535**, 355 (21 July 2016) | doi:10.1038/535355b
Published online 20 July 2016

Journals would then publish only papers that are accompanied online by full experimental protocols, raw data and source code, as in the Protocol Exchange repository

For manuscripts containing statistical analyses, journals should peer review only those papers that use statistics environments based on source code, enforcing the ban on 'point-and-click' statistical software (see go.nature.com/29pdpcl).

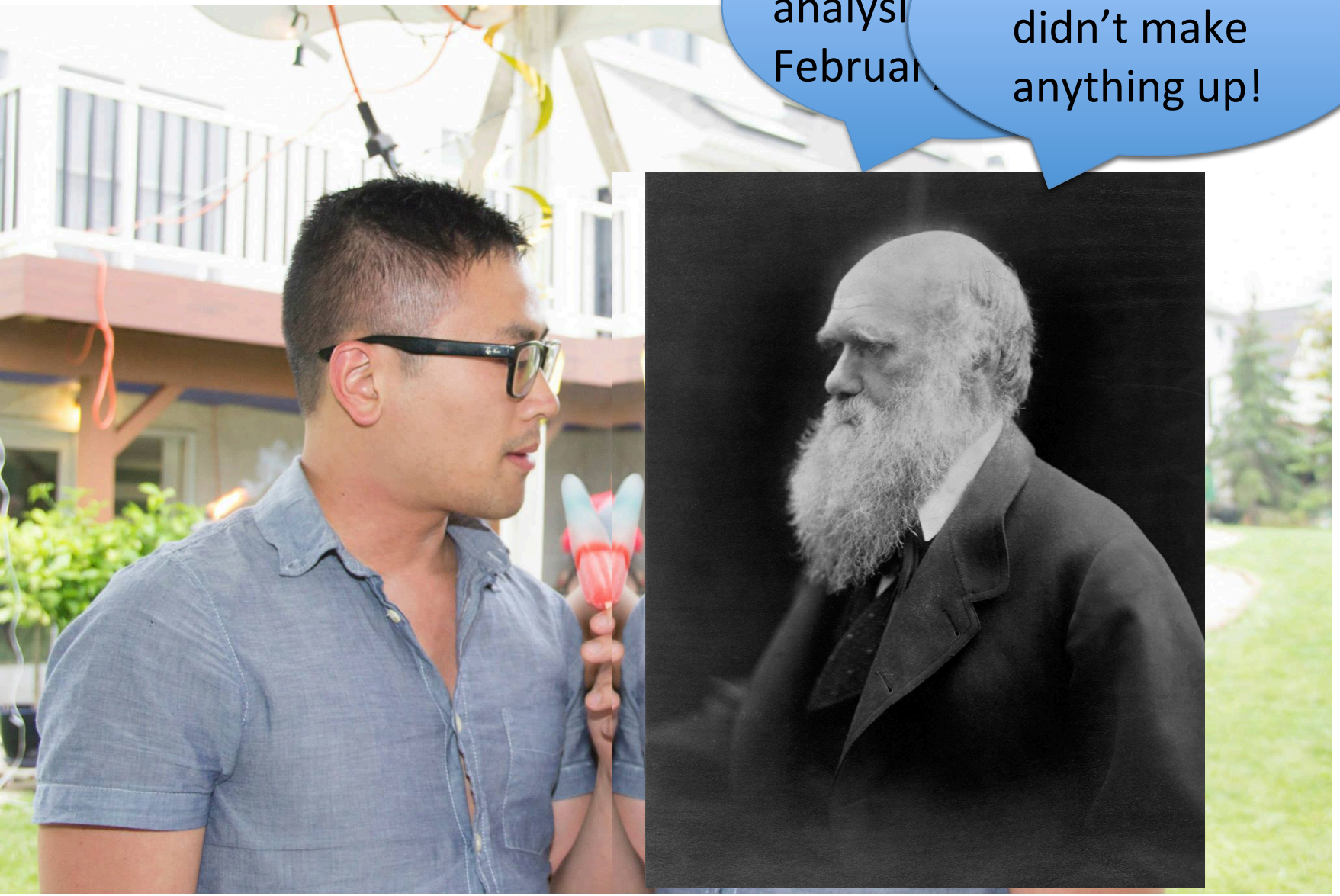
Simply Statistics A statistics blog by Rafa Irizarry, Roger Peng, and Jeff Leek

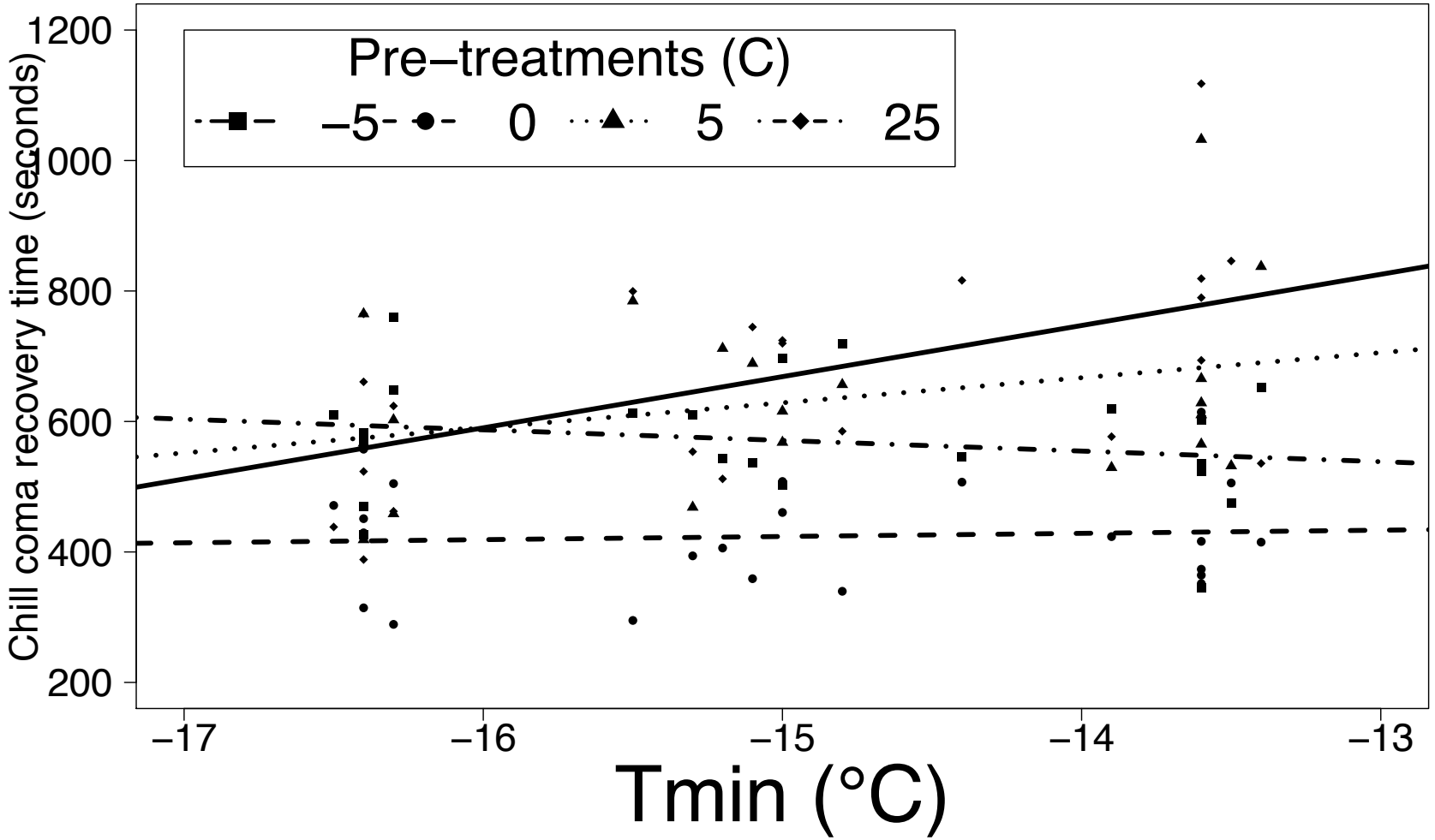
# Instead of research on reproducibility, just do reproducible research

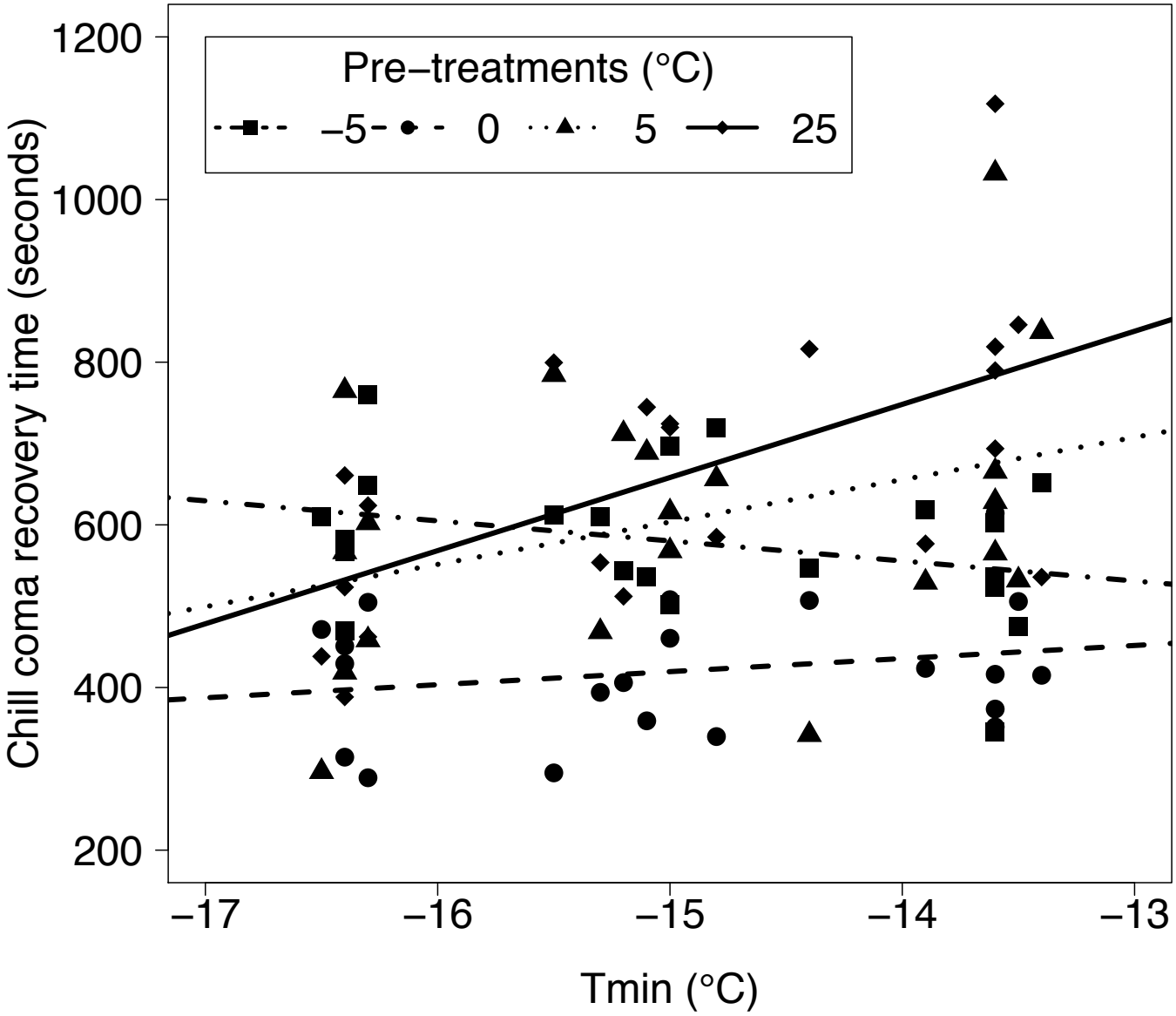
11 Dec 2015

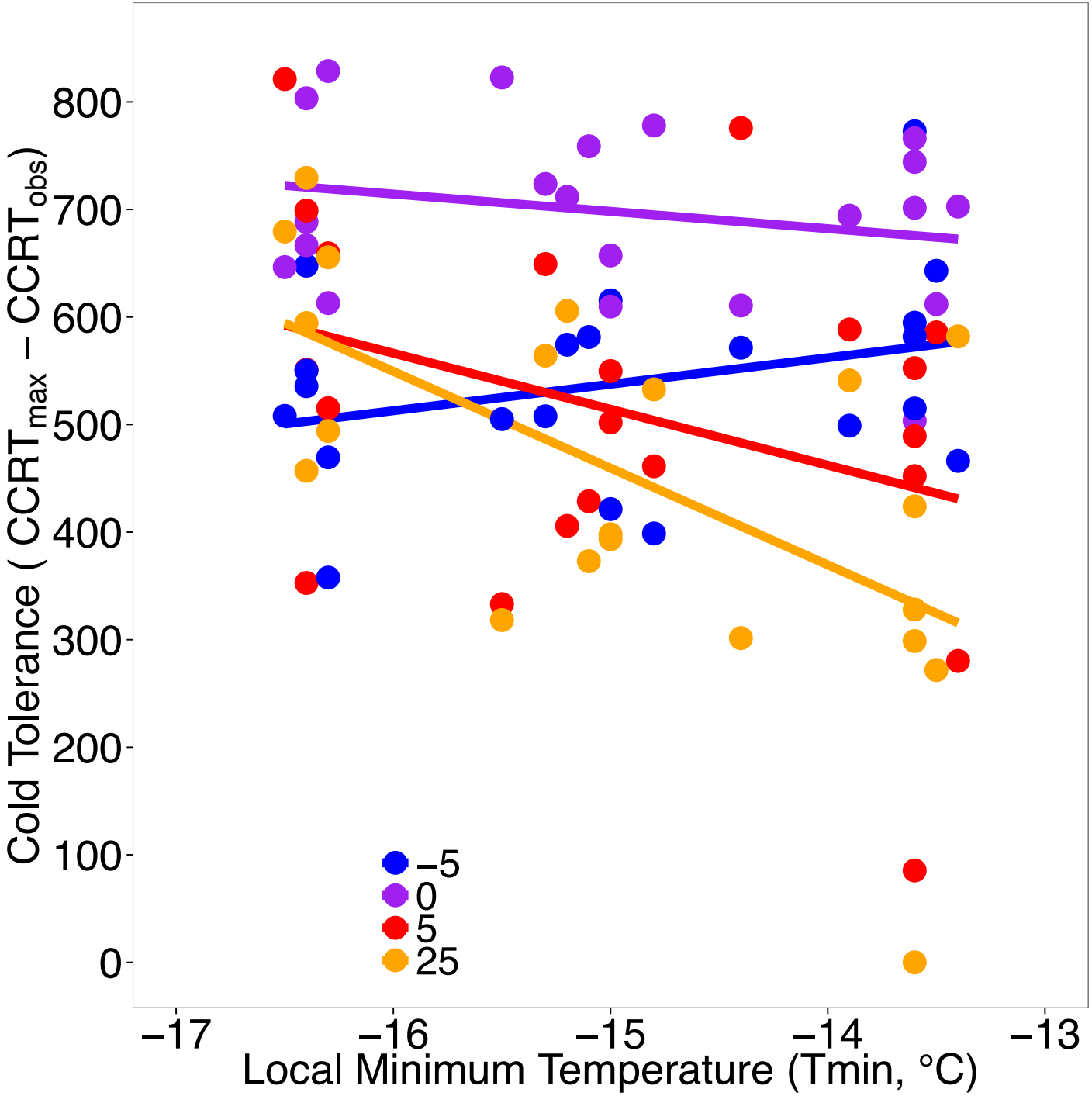
How...?!?!?!

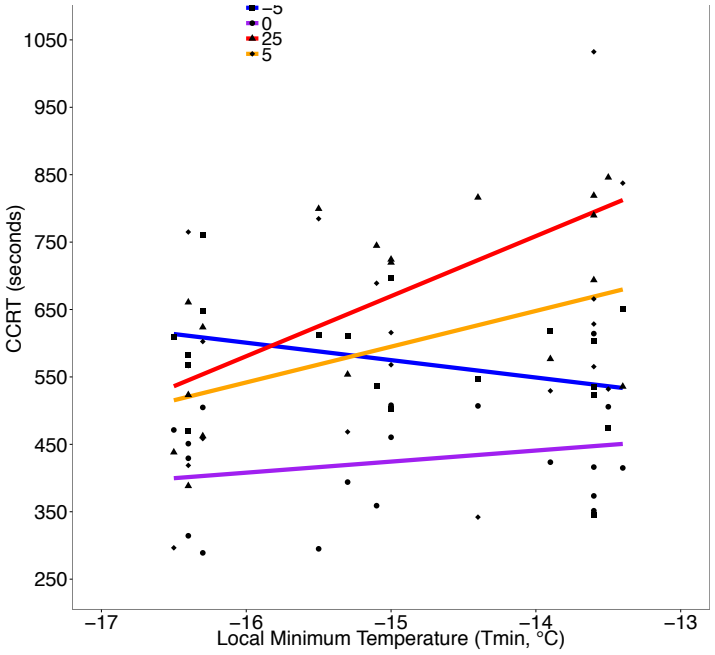
1. Use Github for version control.
2. Use rmarkdown or iPython notebooks for your analysis code
3. When your paper is done post it to arxiv or biorxiv.
4. Post your data to an appropriate repository like SRA or a general purpose site like figshare.
5. Send any software you develop to a controlled repository like CRAN or Bioconductor.
6. Participate in the post publication discussion on Twitter and with a Blog

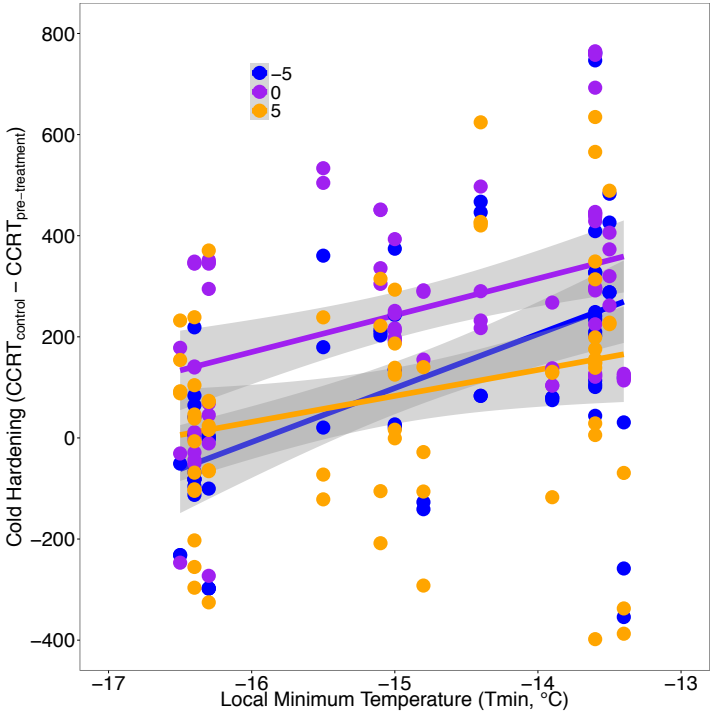

How did
analysi
Februa
Show me that you didn’t make anything up!

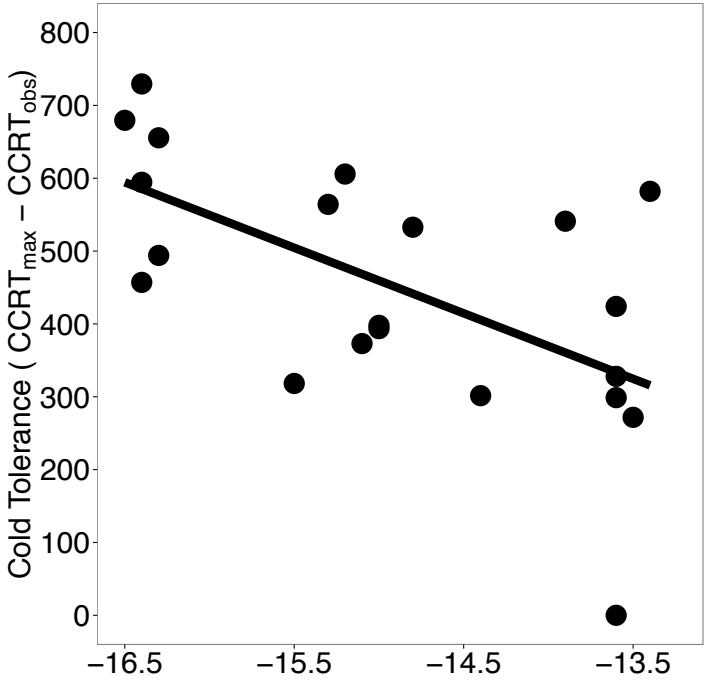

Pre–treatments (C)
–5
0
5
25
Chill coma recovery time (seconds)
1200
1000
800
600
400
200
–17
–16
–15
–14
–13
Tmin (°C)

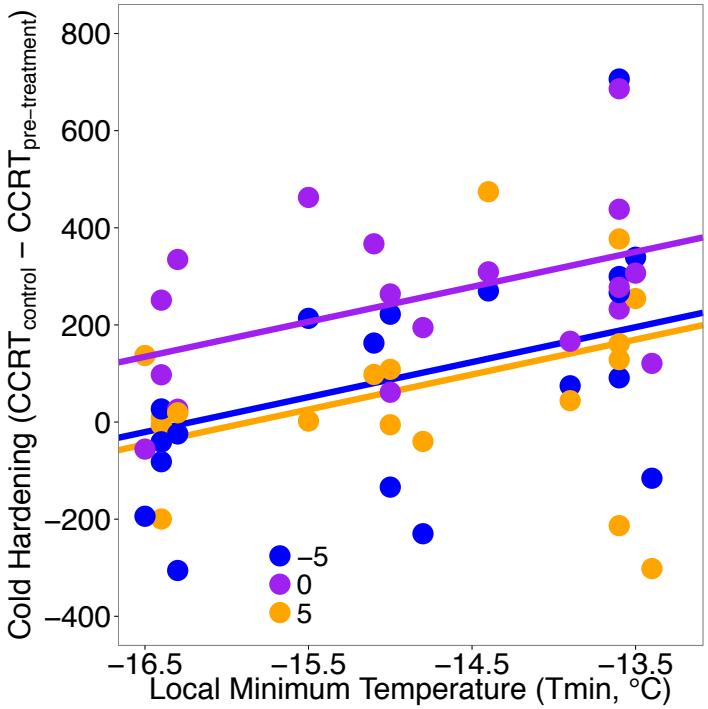

Pre−treatments (°C)
−5
0
5
25
Chill coma recovery time (seconds)
1200
1000
800
600
400
200
−17
−16
−15
−14
−13
Tmin (°C)

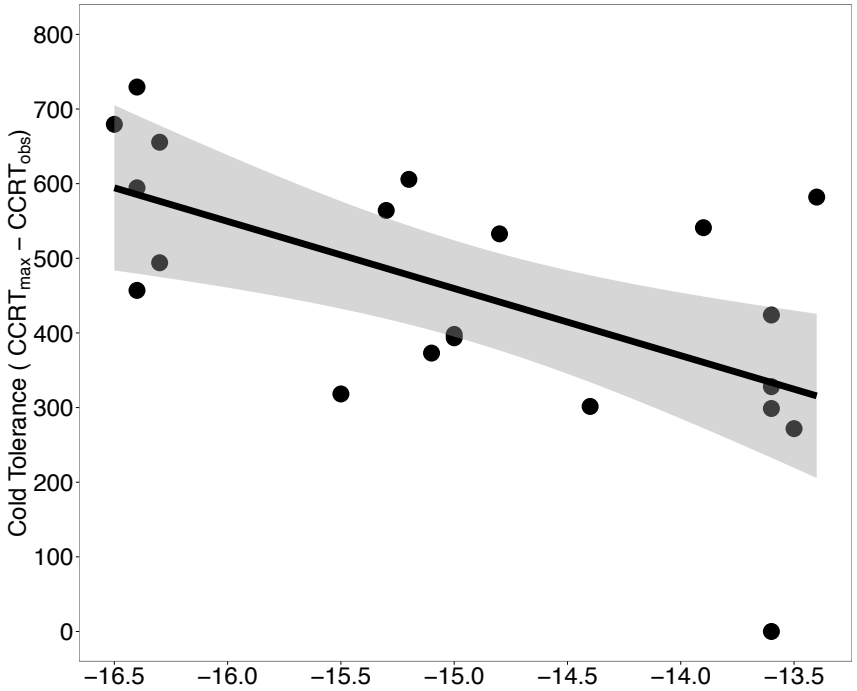

800
700
600
500
400
300
200
100
0
Cold Tolerance ( CCRTmax – CCRTobs)
−5
0
5
25
−17
−16
−15
−14
−13
Local Minimum Temperature (Tmin, °C)

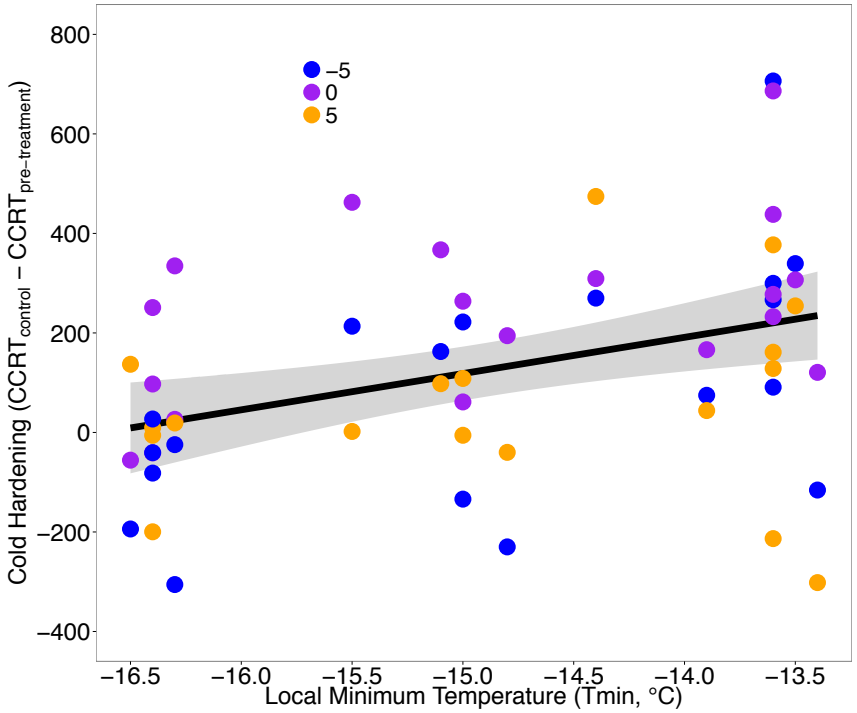

−5
0
25
5
1050
950
850
750
650
550
450
350
250
CCRT (seconds)
−17
−16
−15
−14
−13
Local Minimum Temperature (Tmin, °C)

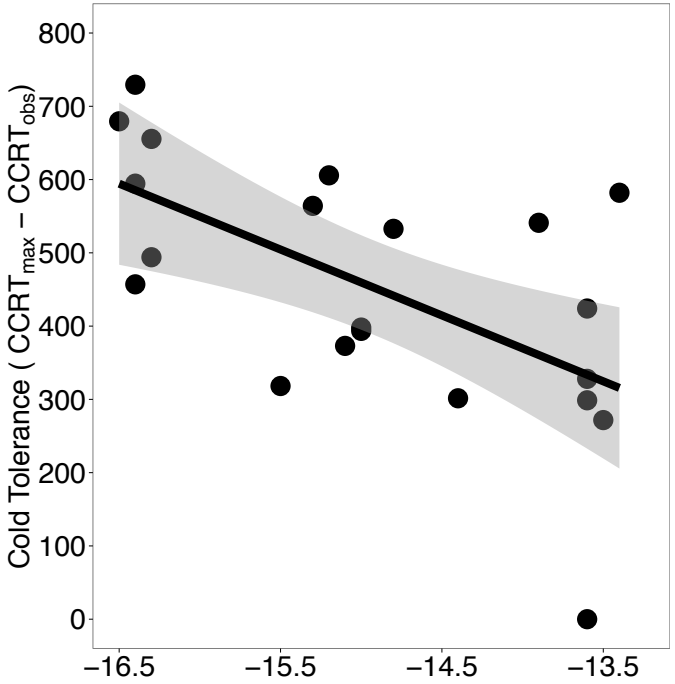

−5
0
5
800
600
400
200
0
−200
−400
Cold Hardening (CCRTcontrol − CCRTpre−treatment)
−17
−16
−15
−14
−13
Local Minimum Temperature (Tmin, °C)

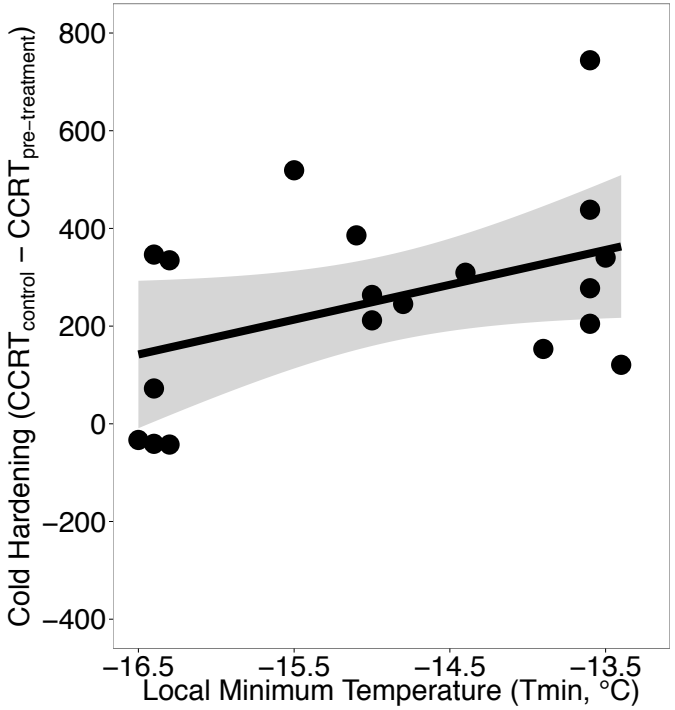

Cold Tolerance ( CCRTmax – CCRTobs)
800
700
600
500
400
300
200
100
0
−16.5
−15.5
−14.5
−13.5

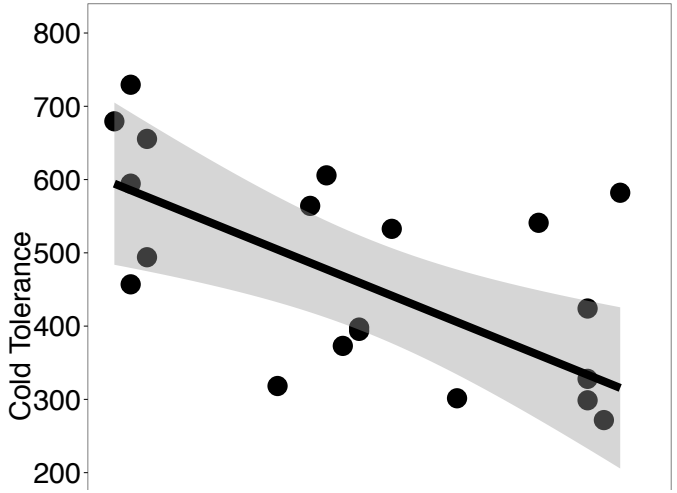

Cold Hardening (CCRTcontrol – CCRTpre−treatment)
800
600
400
200
0
−200
−400
−5
0
5
−16.5
−15.5
−14.5
−13.5
Local Minimum Temperature (Tmin, °C)


Cold Tolerance ( CCRTmax – CCRTobs)
800
700
600
500
400
300
200
100
0
−16.5
−16.0
−15.5
−15.0
−14.5
−14.0
−13.5

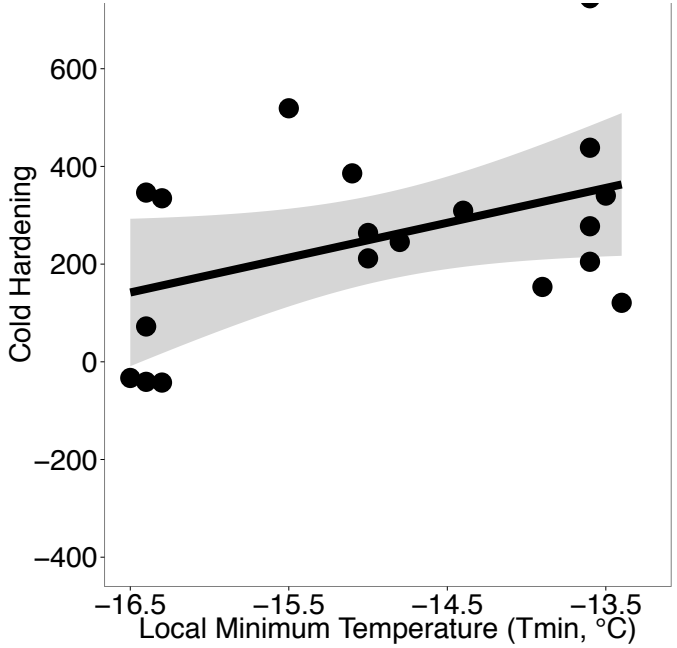

Cold Hardening (CCRTcontrol – CCRTpre−treatment)
800
600
400
200
0
−200
−400
−5
0
5
−16.5
−16.0
−15.5
−15.0
−14.5
−14.0
−13.5
Local Minimum Temperature (Tmin, °C)


Cold Tolerance ( CCRTmax – CCRTobs)
800
700
600
500
400
300
200
100
0
–16.5
–15.5
–14.5
–13.5

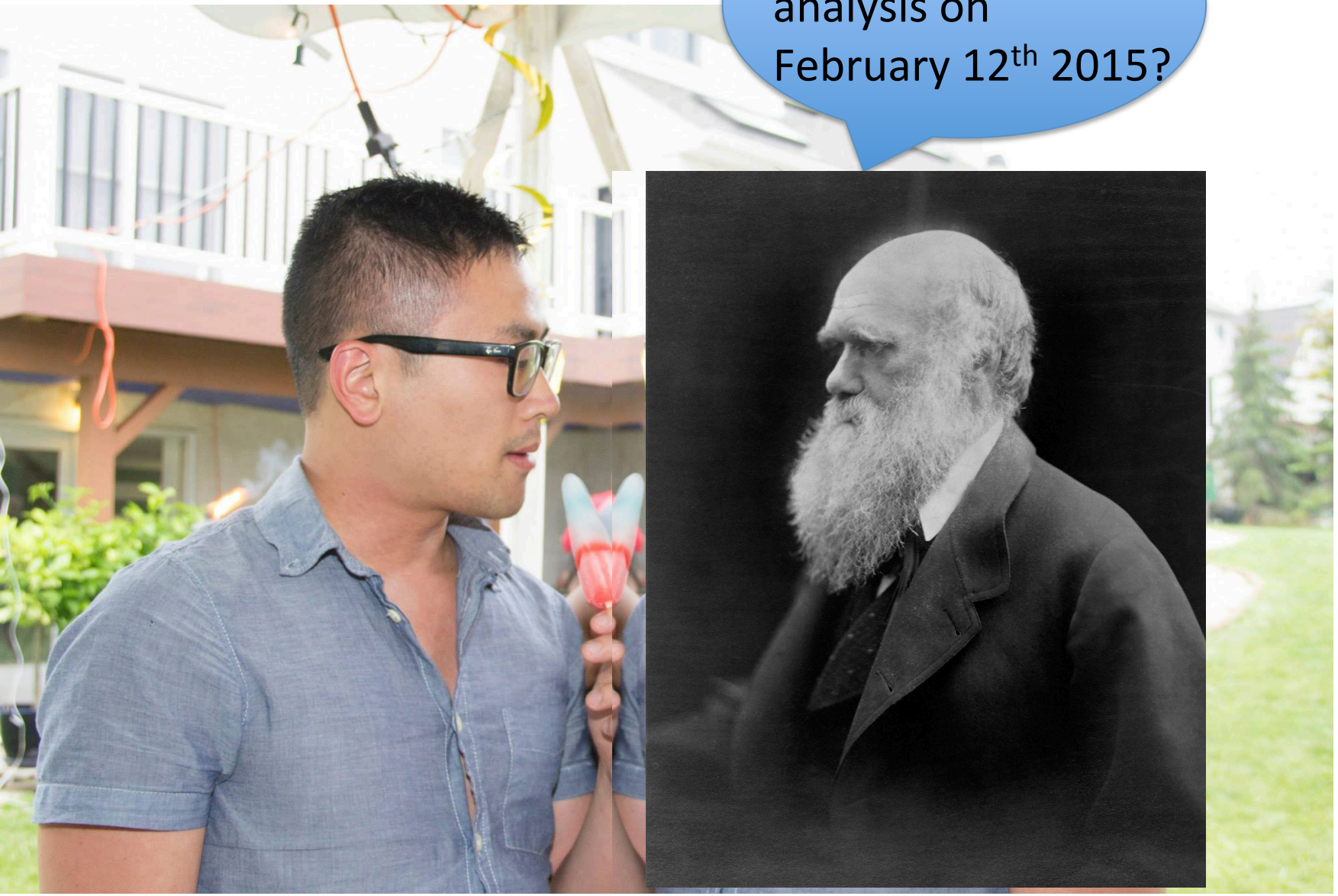

Cold Hardening (CCRTcontrol – CCRTpre–treatment)
800
600
400
200
0
–200
–400
–16.5
–15.5
–14.5
–13.5
Local Minimum Temperature (Tmin, °C)




# Observation: You're always going to repeat your work

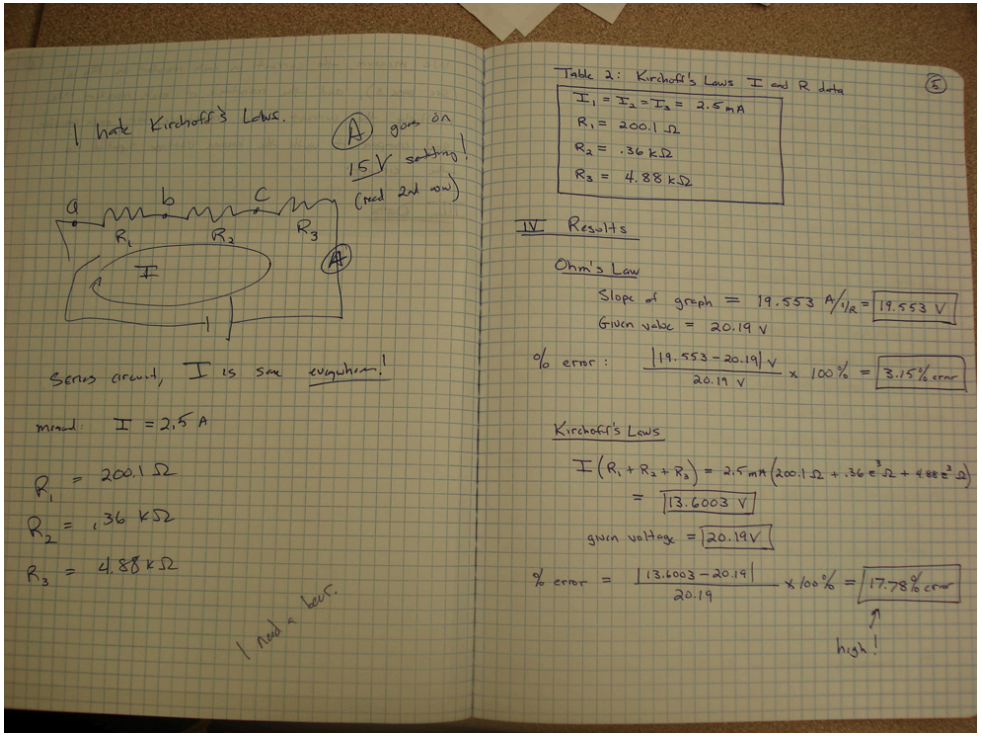



How did you do this analysis from February 12th 2015?

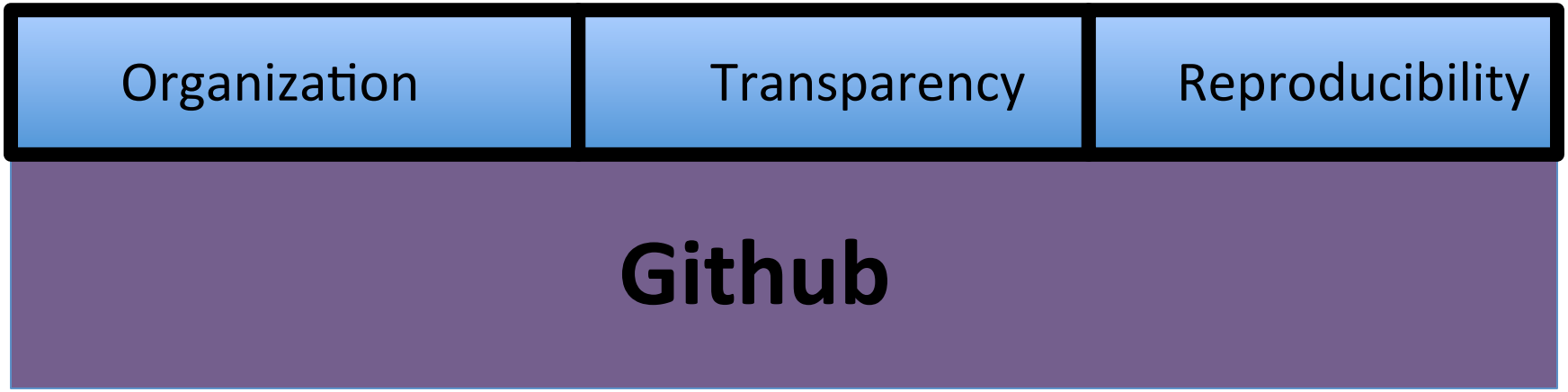

How did you do this analysis on February 12th 2015?

# Science needs to be reproducible, repeatable! How?

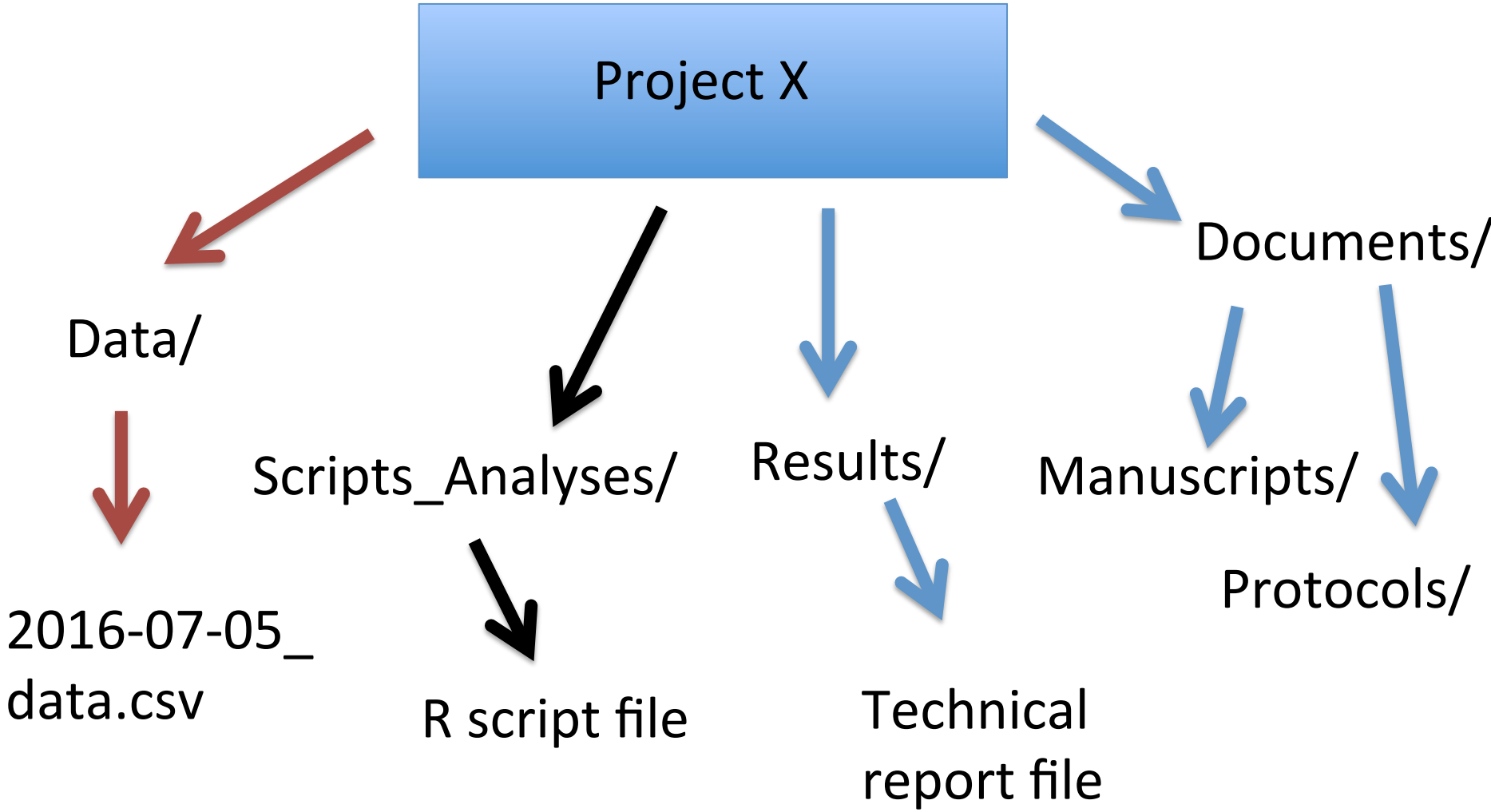
**Capture** the *process* of science as it is happening and *share* it.



# Overall workflow



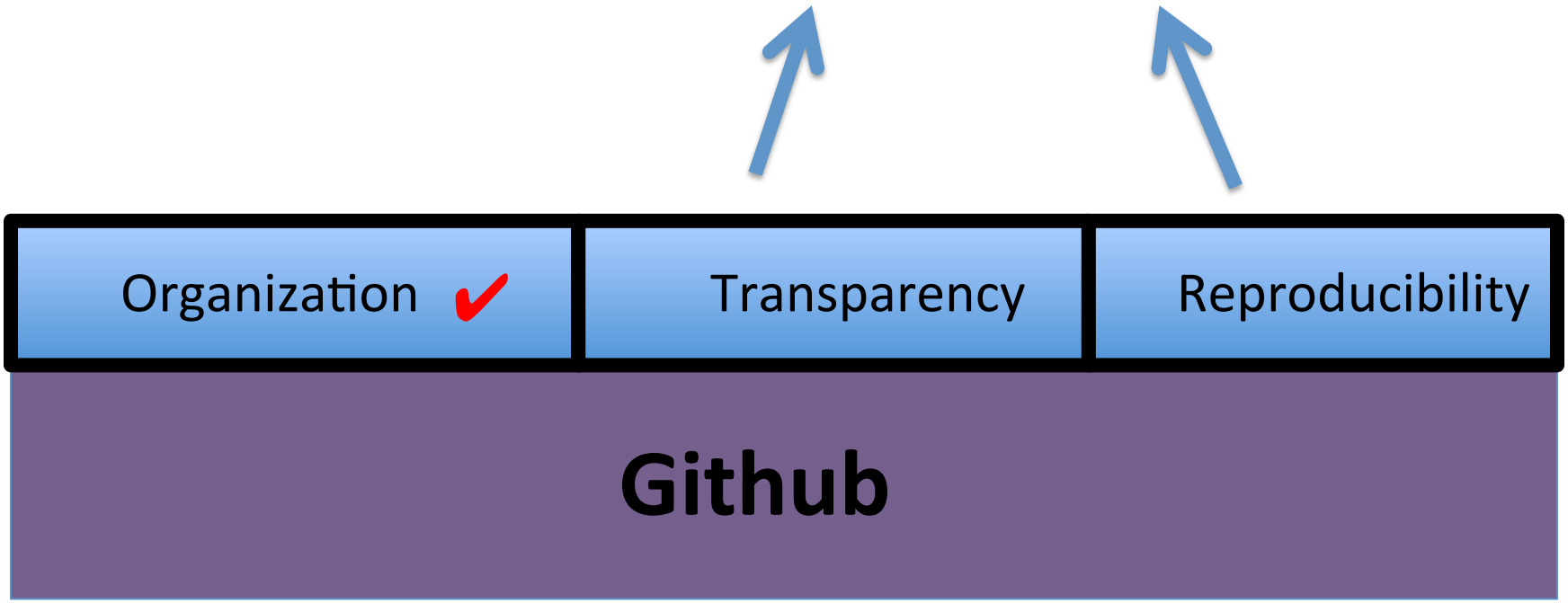
# Organize by projects using a tree-like structure



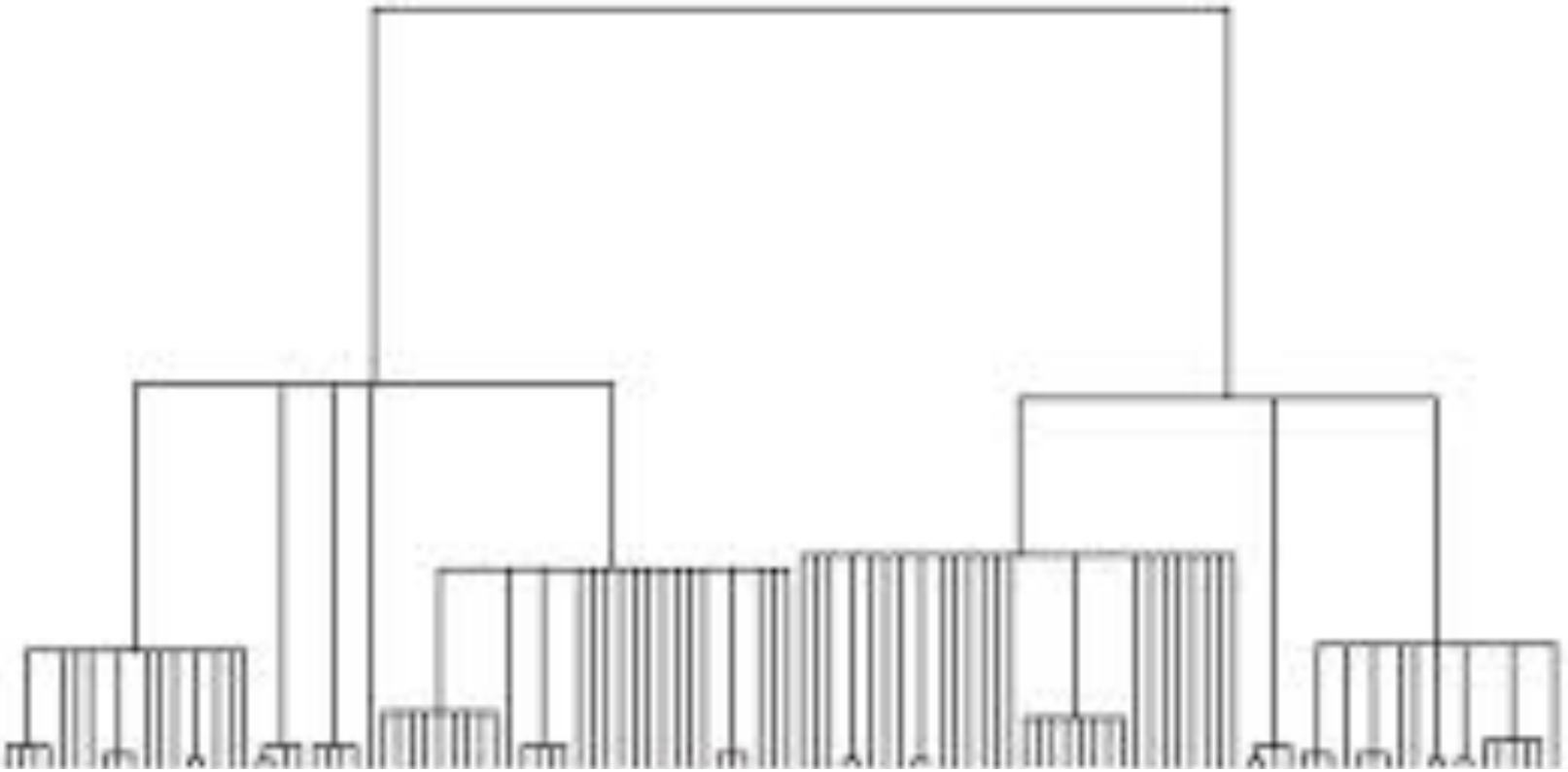

Noble 2012; *PlosOne*

**Track all of the changes to your project**
**Share it online!**

**Let anybody access and repeat your analysis**



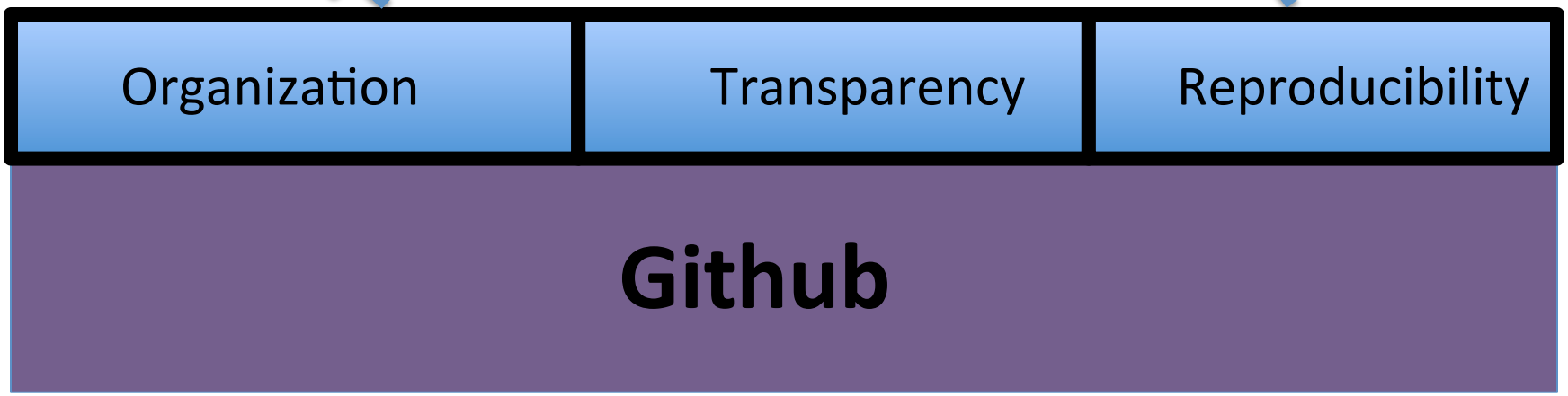
# But what if your project is a large tree?



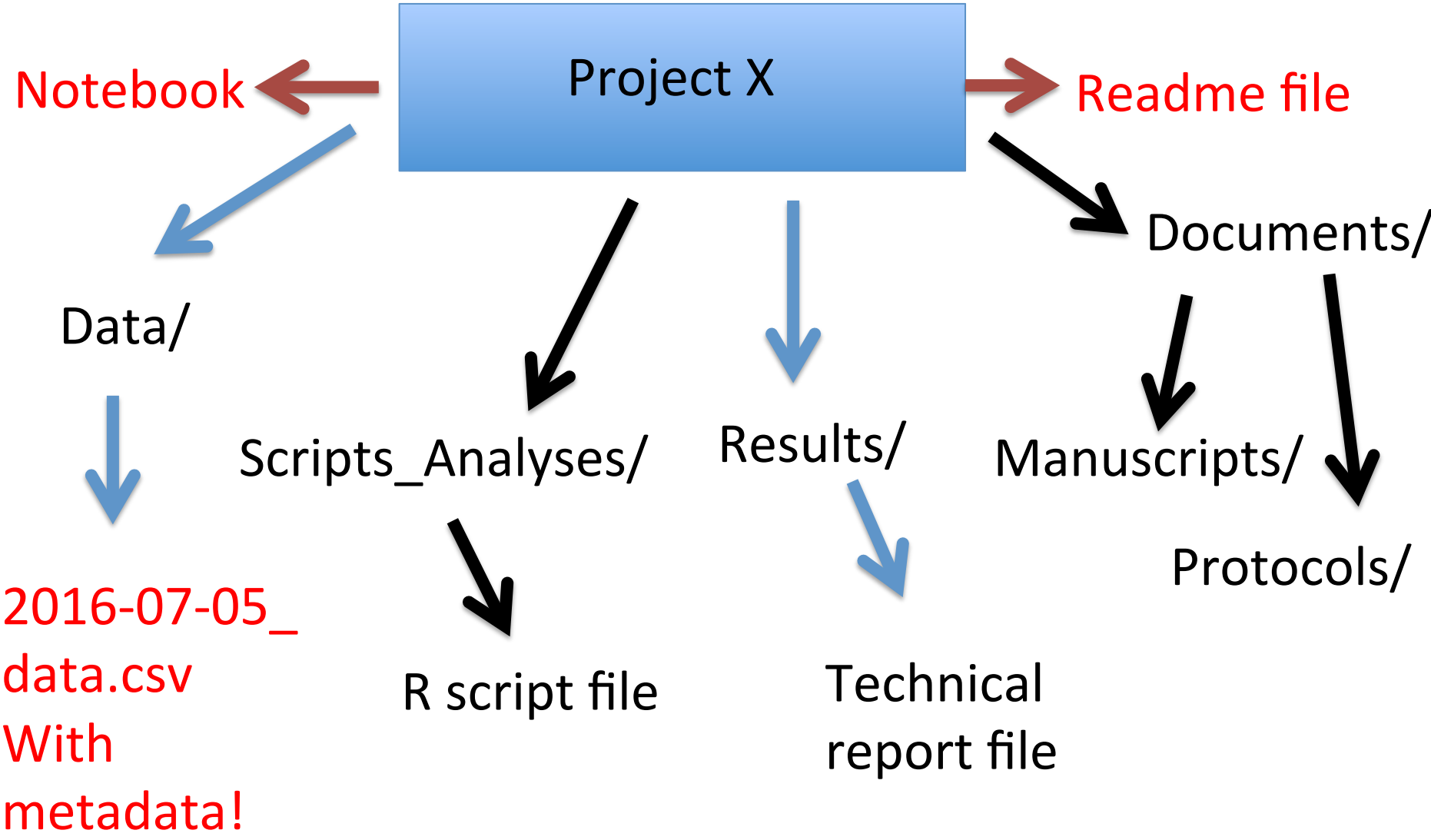
# If you didn't write it down, it didn't happen!

**Track all of the changes to your project**
**Share it online!**
**Let anybody access and repeat your analysis**

**Annotate everything…**



# Organize by projects using a tree-like structure



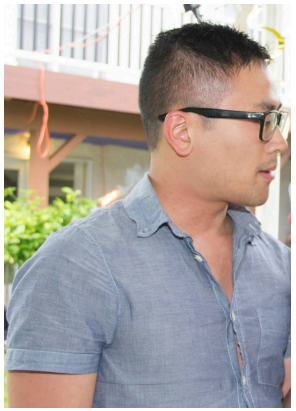

Noble 2012; *PlosOne*

# Now, how do we track changes?

Personal computer

Track changes

Github software



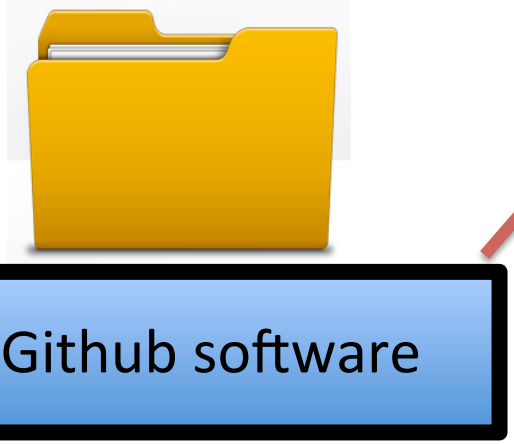
# Connects present and future self

# Connecting with Darwin?

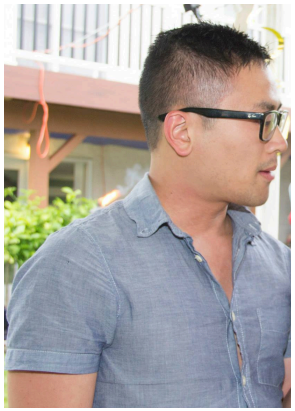
# Now, how do we track changes and share projects?

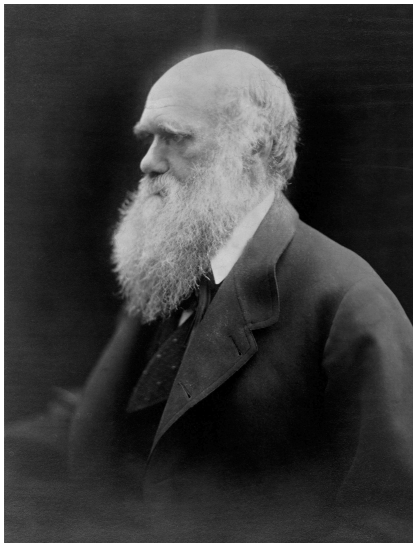


Track changes

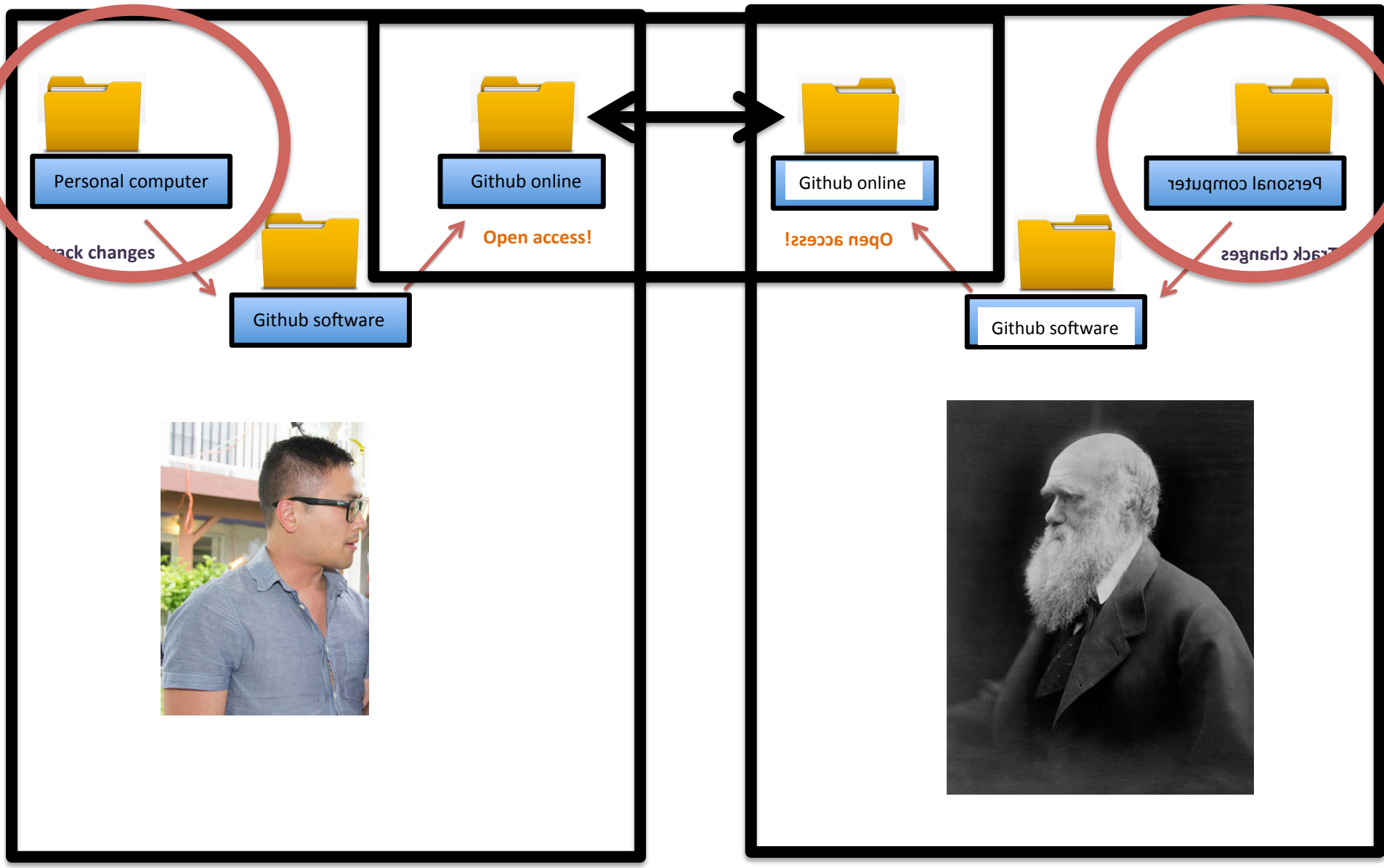



Open access!

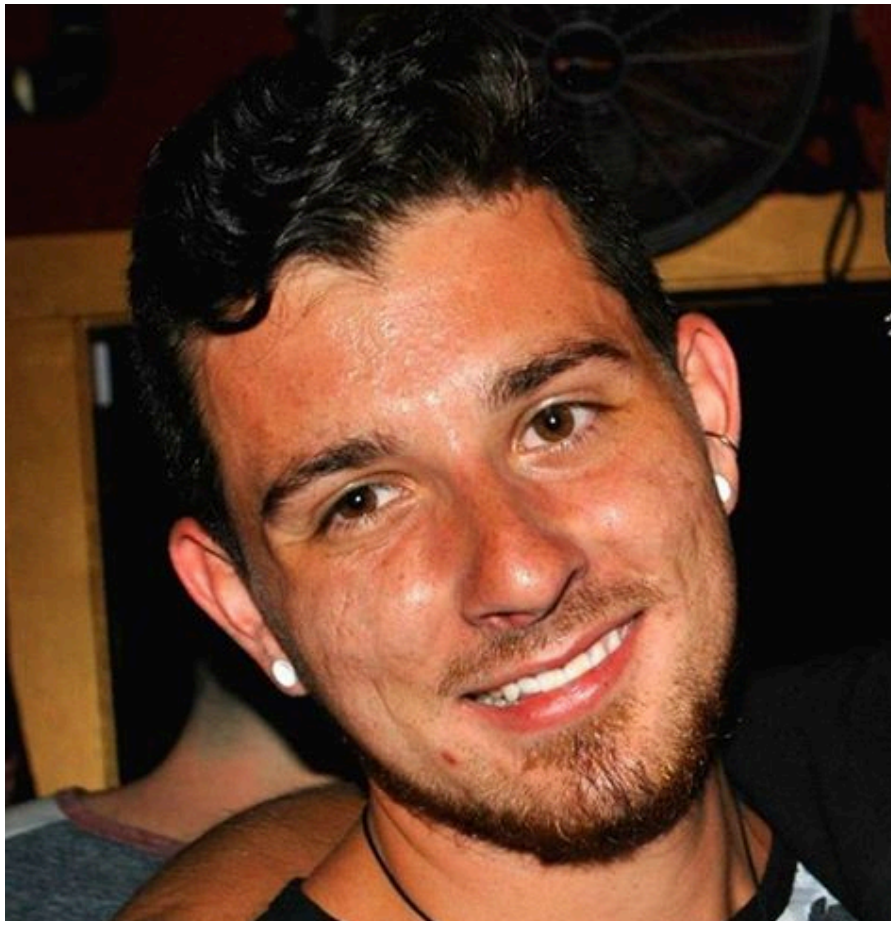




# Connecting AND collaborating with Darwin



# Test case: Undergraduate researcher



http://adnguyen.github.io/2013_stressed_ants_in_warmingchambers/

*NATURE* | CORRESPONDENCE

# Research papers: Journals should drive data reproducibility

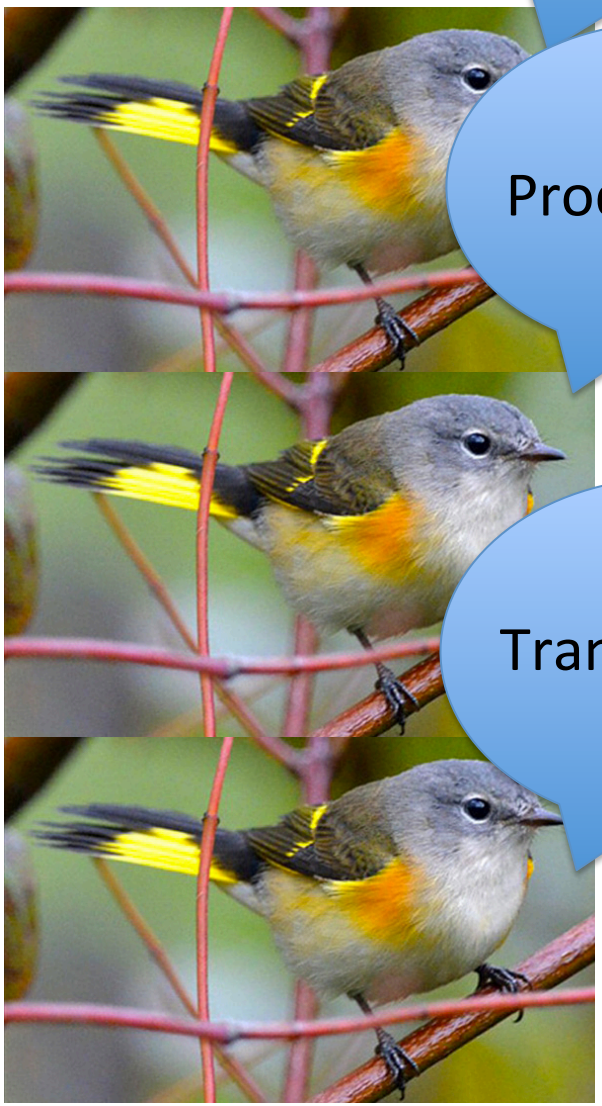
**Gregorio Santori**

*Nature* **535**, 355 (21 July 2016) | doi:10.1038/535355b
Published online 20 July 2016

Journals would then publish only papers that are accompanied online by full experimental protocols, raw data and source code, as in the Protocol Exchange repository

For manuscripts containing statistical analyses, journals should peer review only those papers that use statistics environments based on source code, enforcing the ban on 'point-and-click' statistical software (see go.nature.com/29pdpcl).

Reproducibility



Productivity


Transparency

Organization