

# Strategies for achieving reproducible research

ANTdrew D. Nguyen,  
Evolutionary Physiologist

# 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

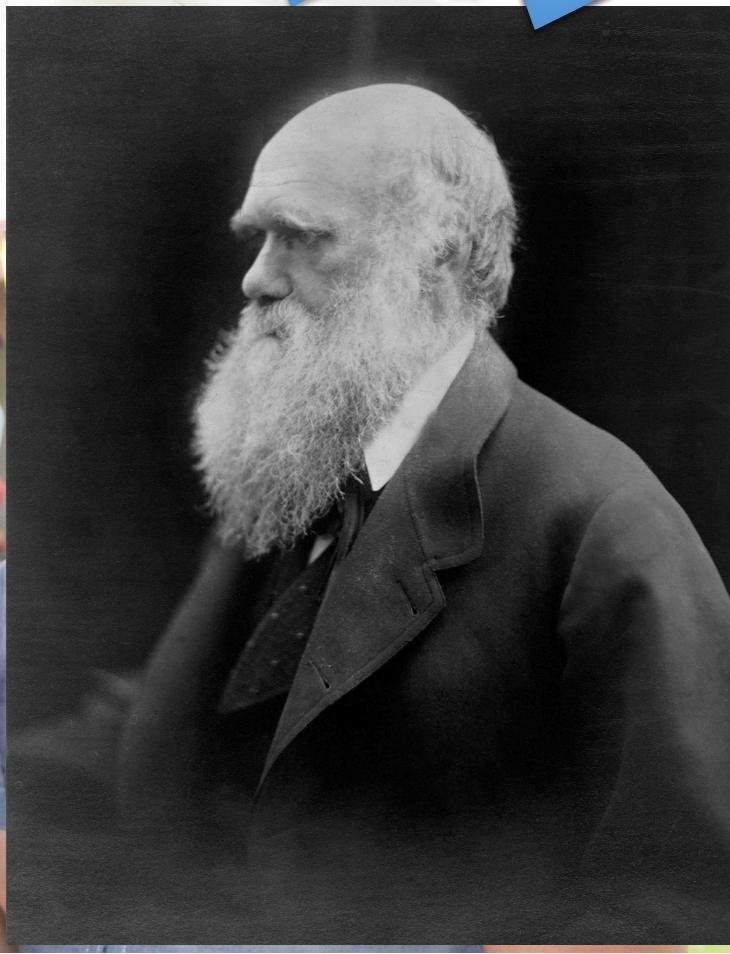
For manuscripts containing statistical analyses, journals should peer review only those papers that use statistics environments based on source code, enforcing the ban on 'point-and-click' statistical software (see [go.nature.com/29pdpc1](http://go.nature.com/29pdpc1)).

# Instead of research on reproducibility, just do reproducible research

11 Dec 2015

How...?!?!?

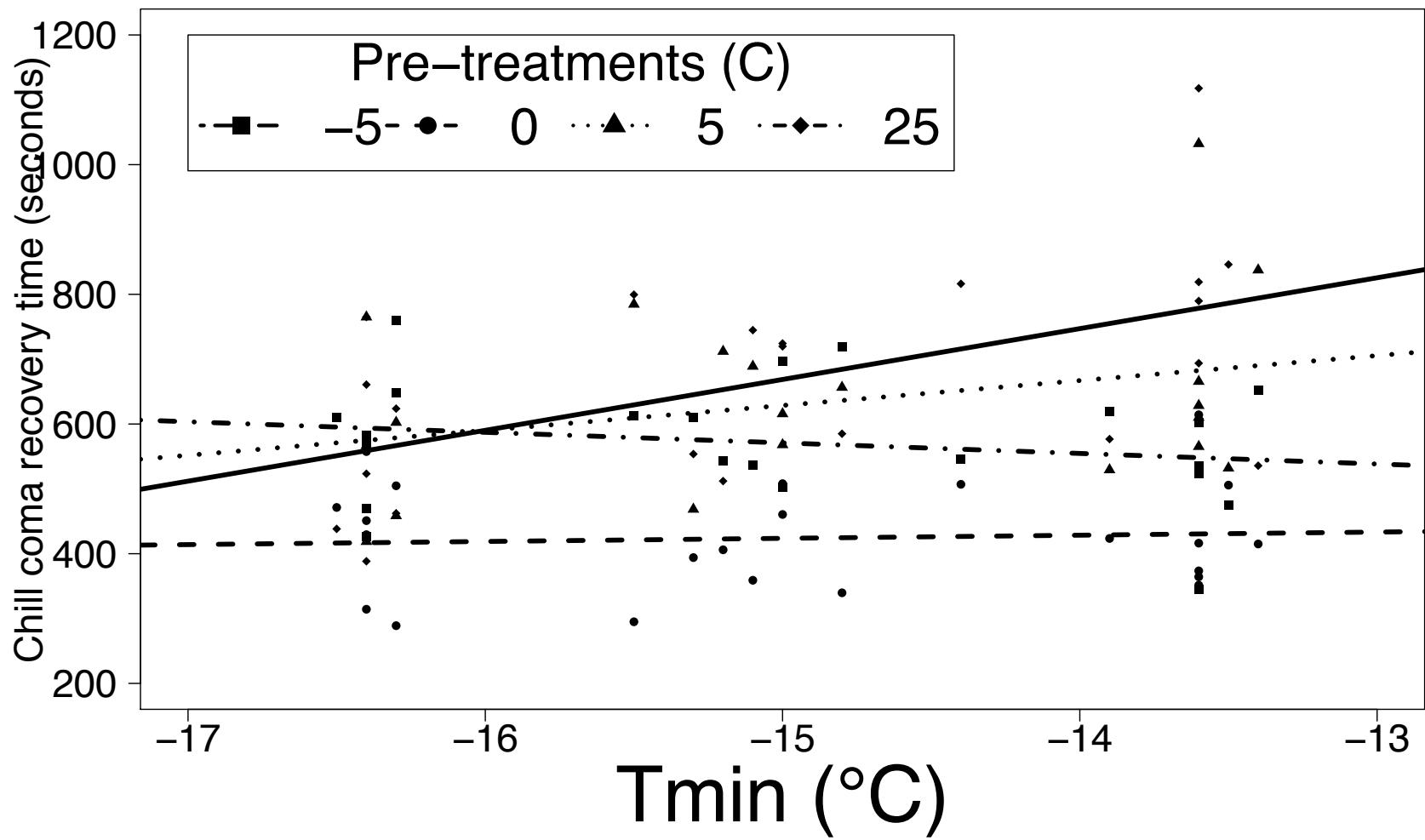
1. Use [Github](#) for version control.
2. Use [rmarkdown](#) or [iPython notebooks](#) for your analysis code
3. When your paper is done post it to [arxiv](#) or [biorxiv](#).
4. Post your data to an appropriate repository like [SRA](#) or a general purpose site like [figshare](#).
5. Send any software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

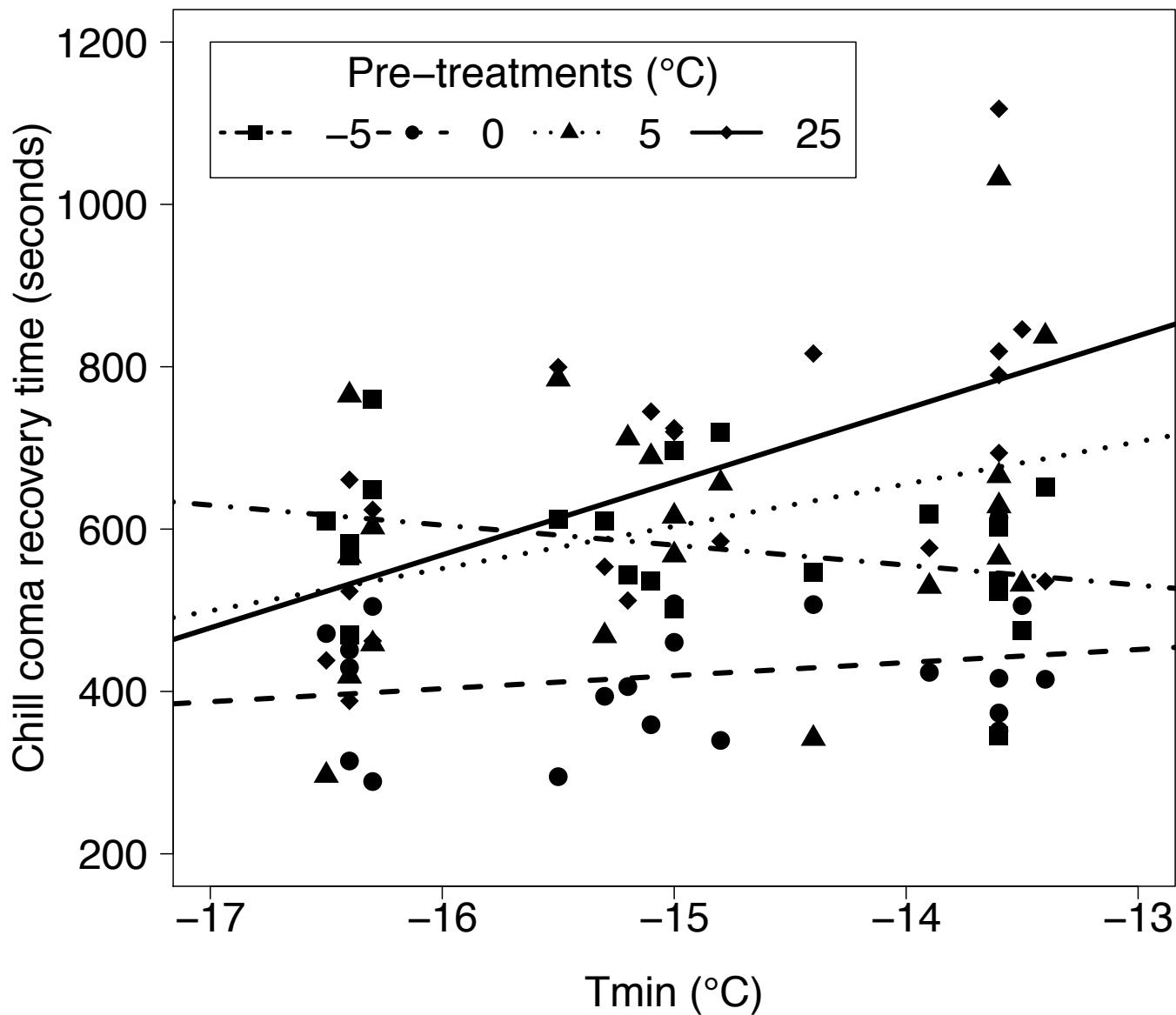


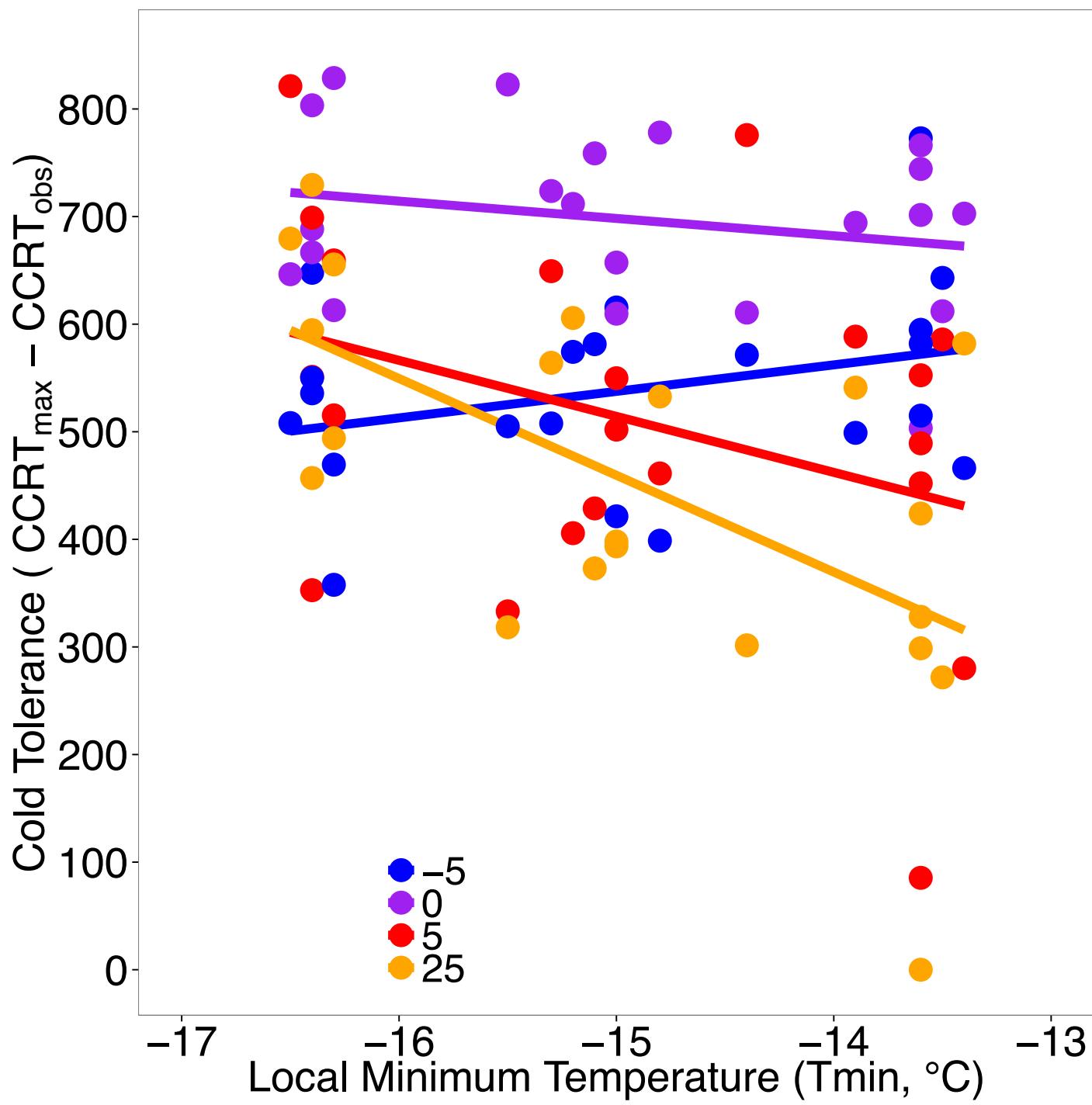
How did you  
analyse the  
February

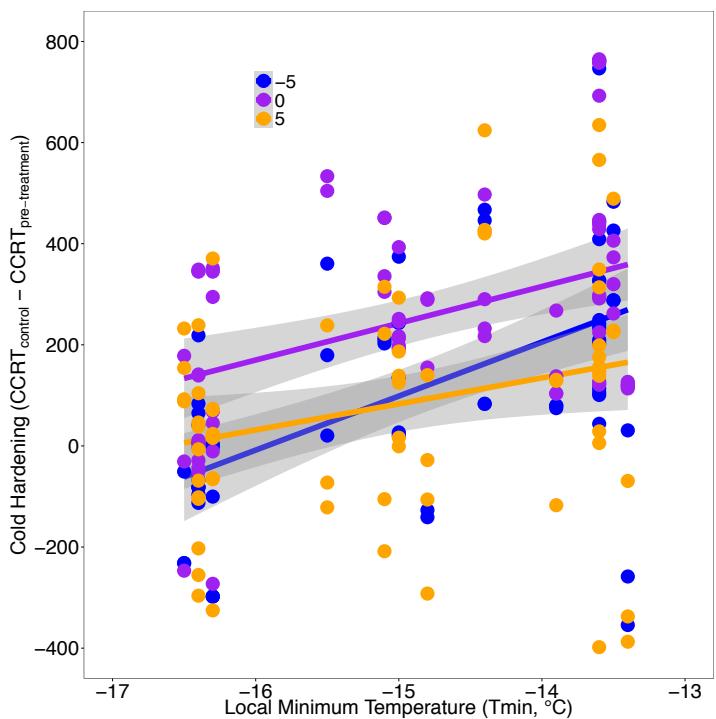
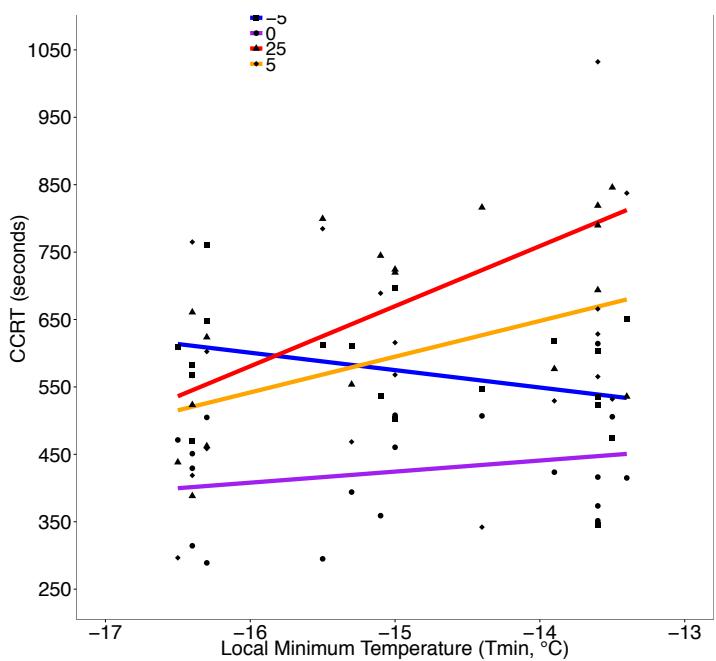
Show me that you  
didn't make  
anything up!

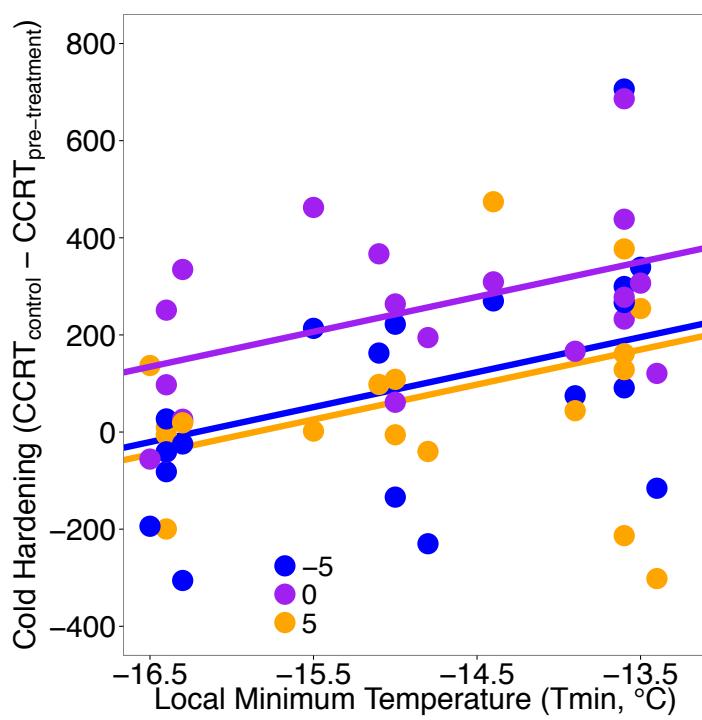
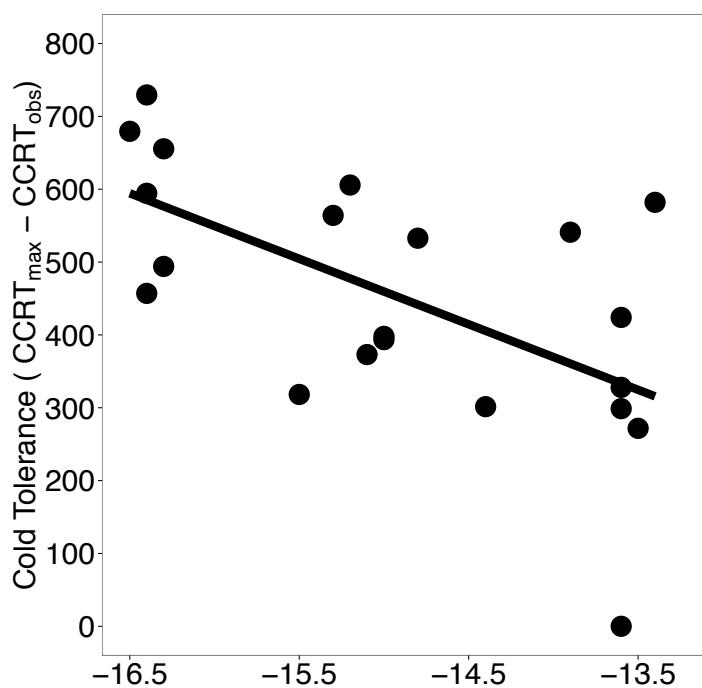
Group meeting with climate cascade group

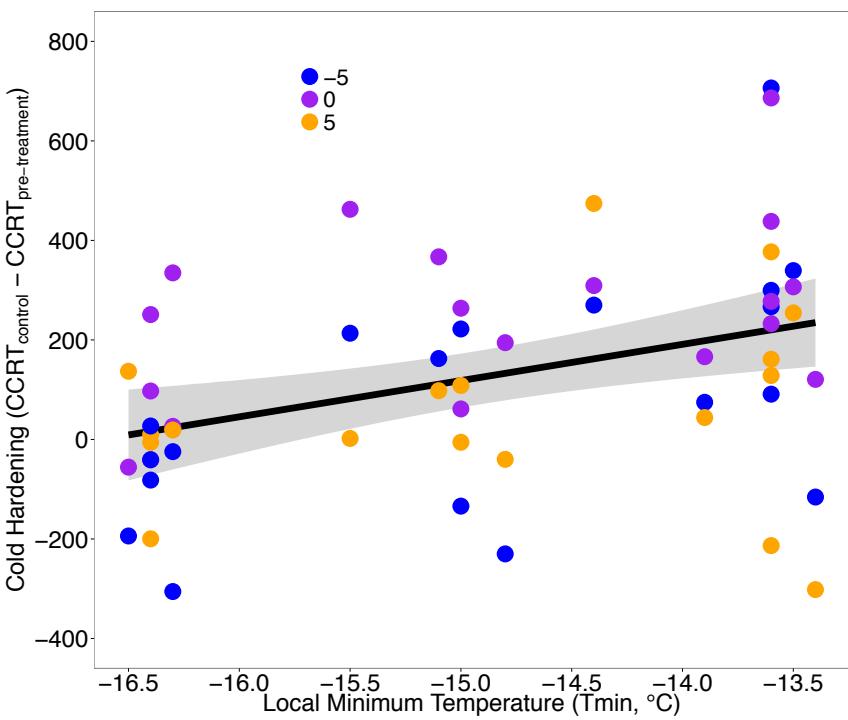
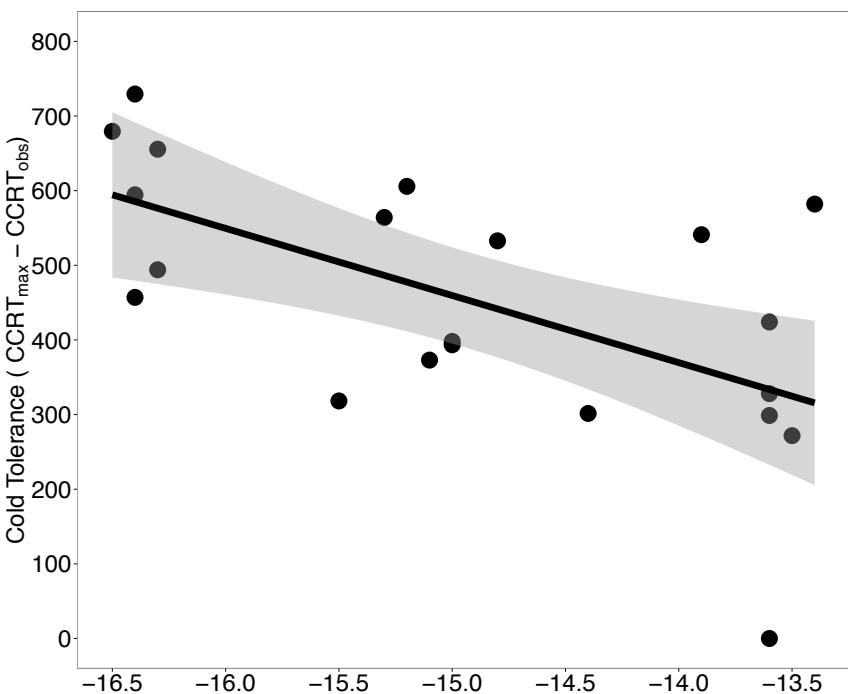


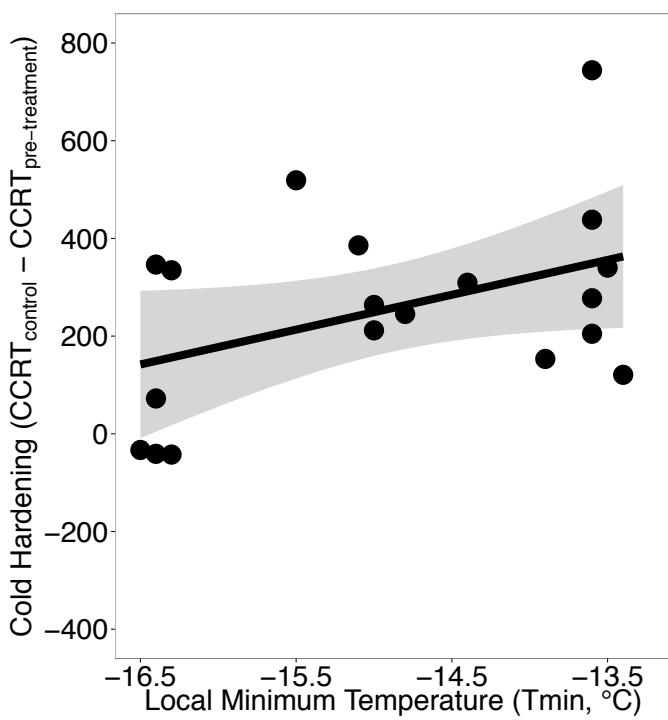
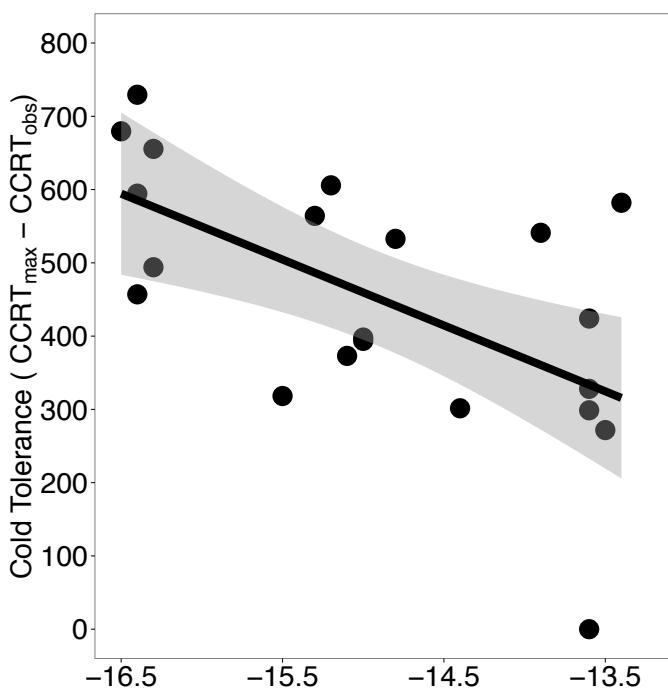


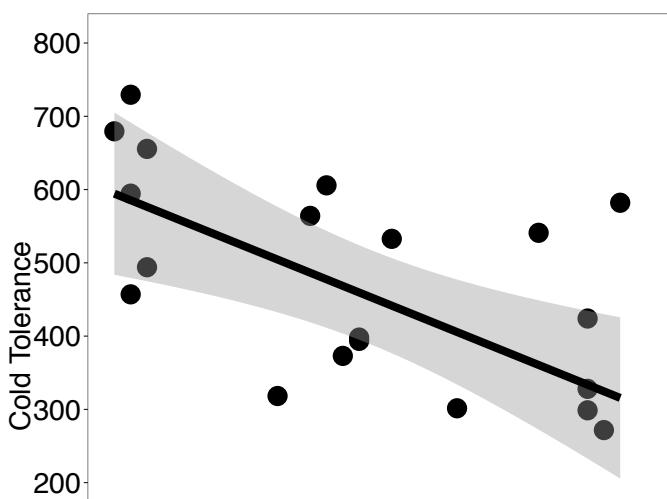




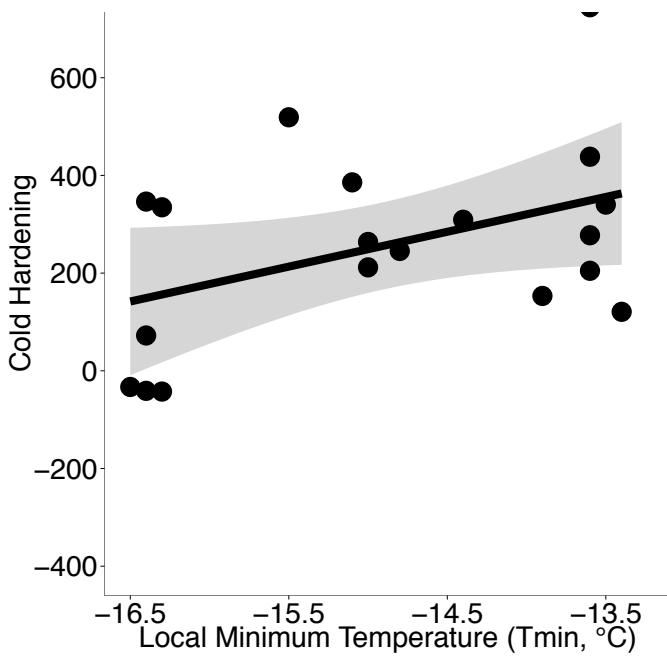


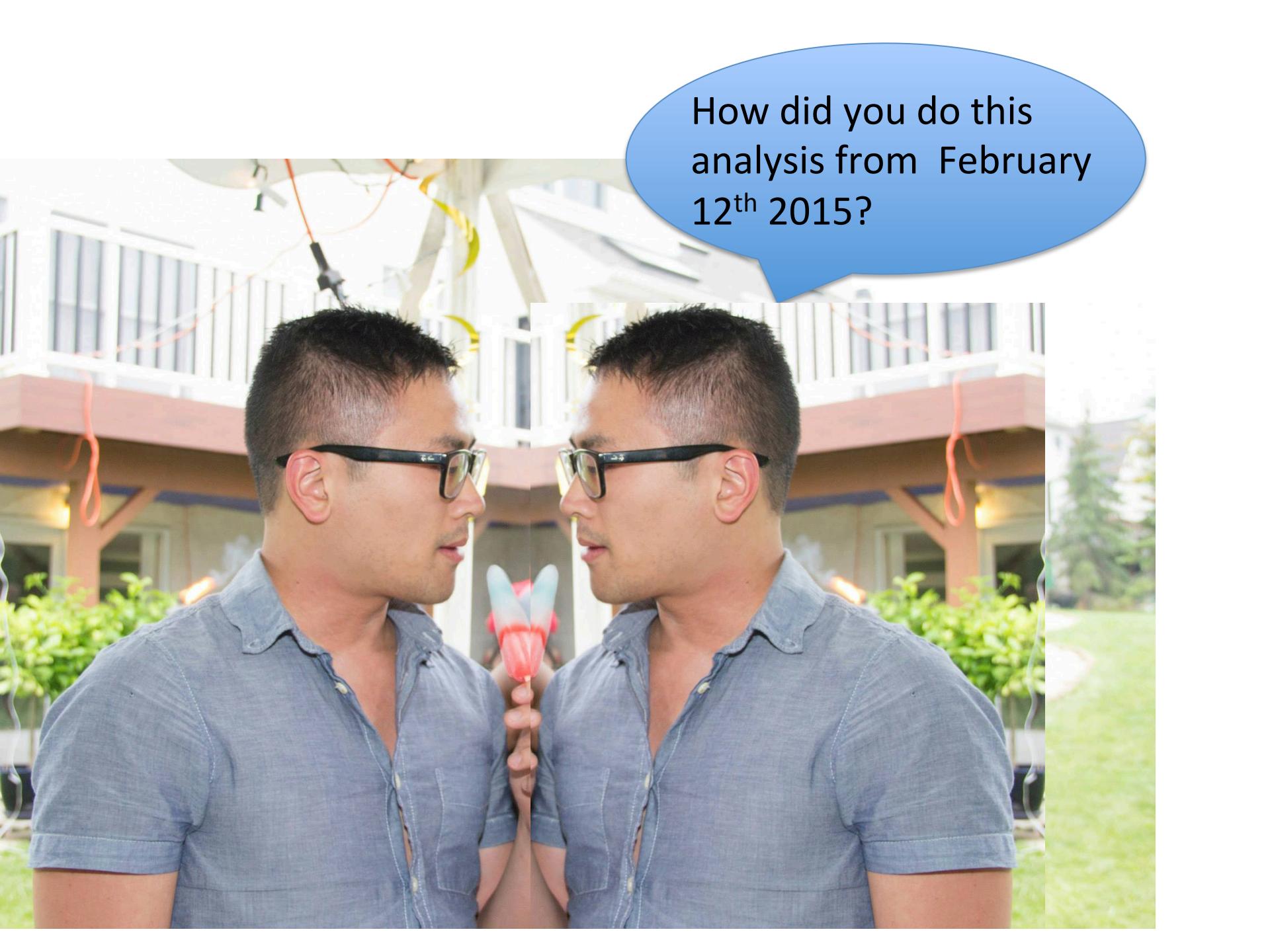




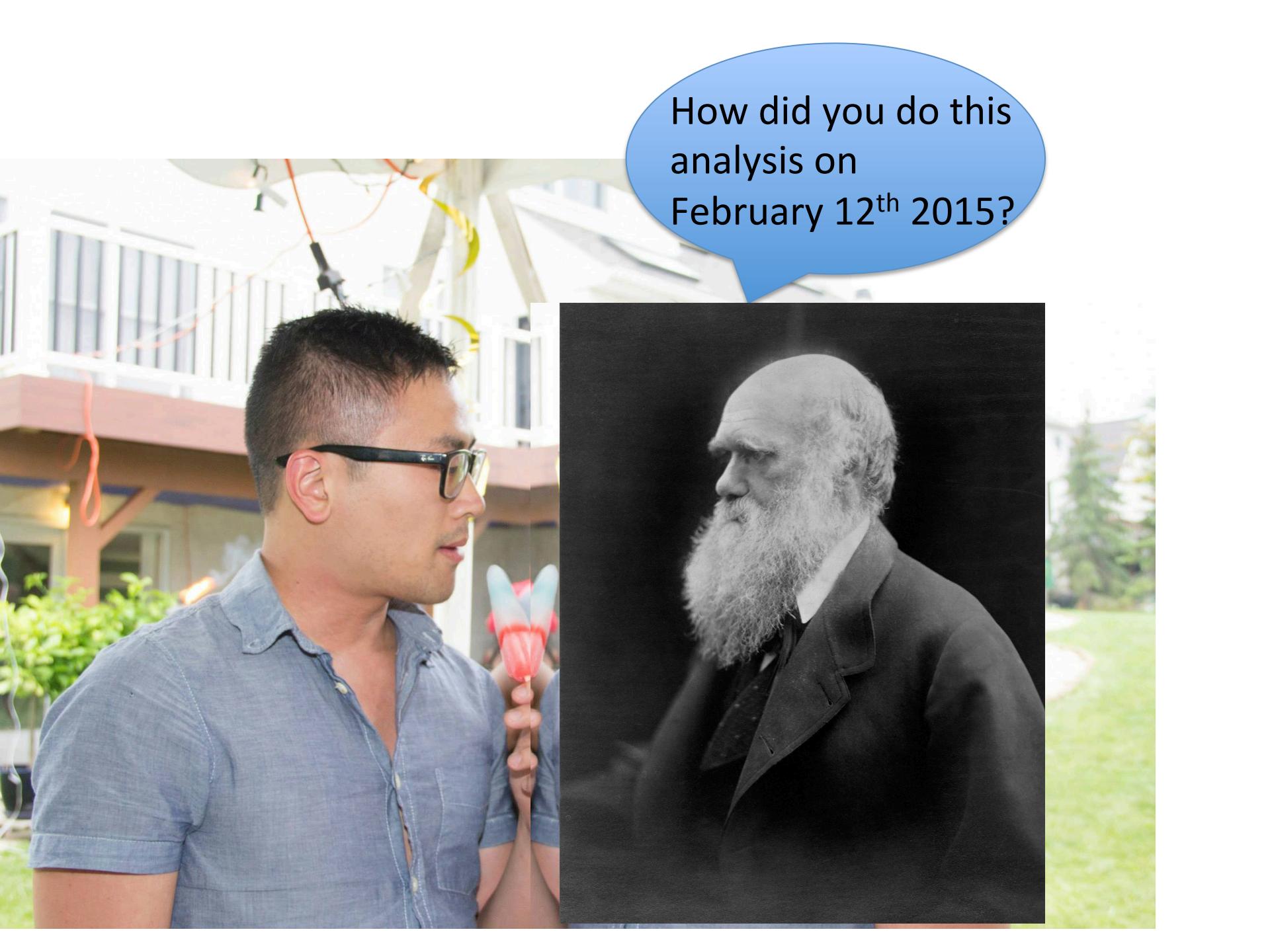


Observation: You're always going to  
repeat your work

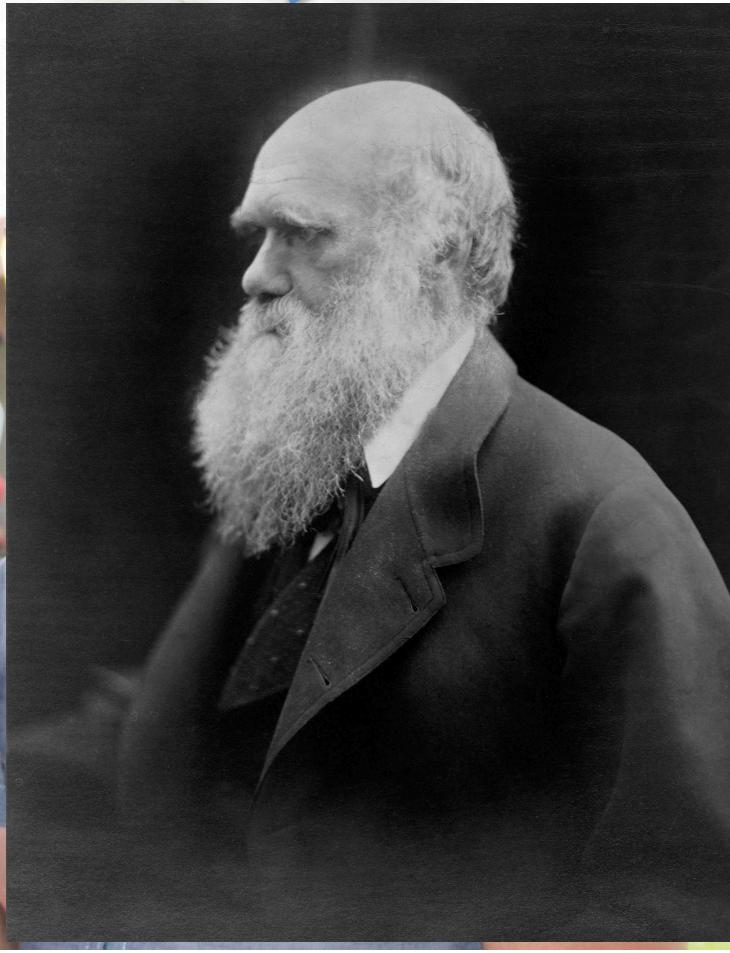




How did you do this analysis from February 12<sup>th</sup> 2015?

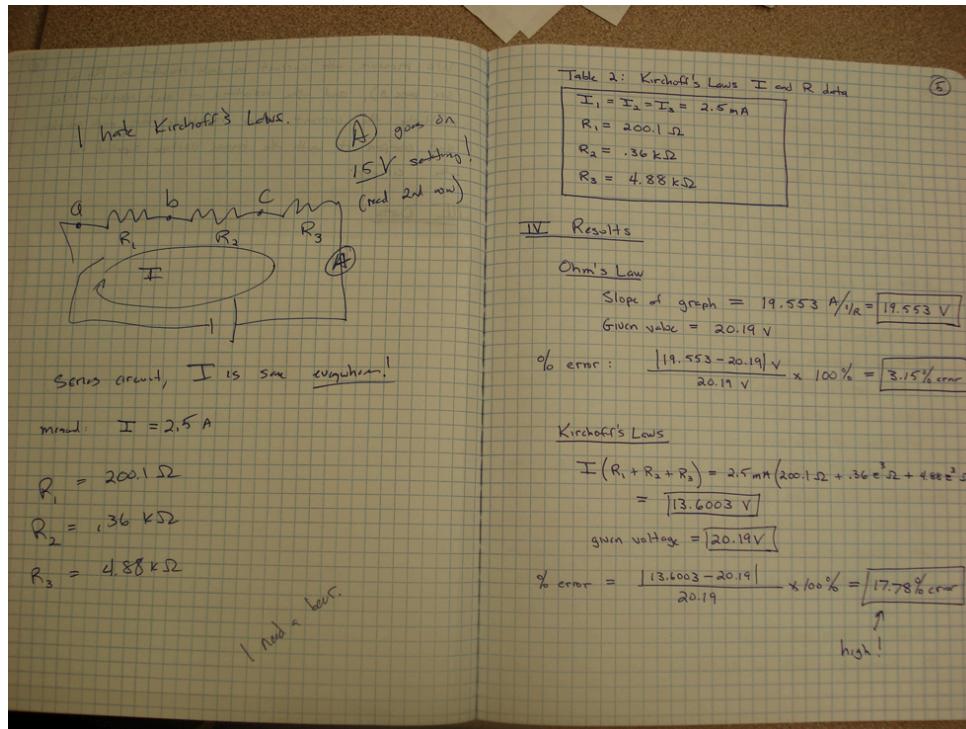


How did you do this  
analysis on  
February 12<sup>th</sup> 2015?



# Science needs to be reproducible, repeatable! How?

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



# Overall workflow

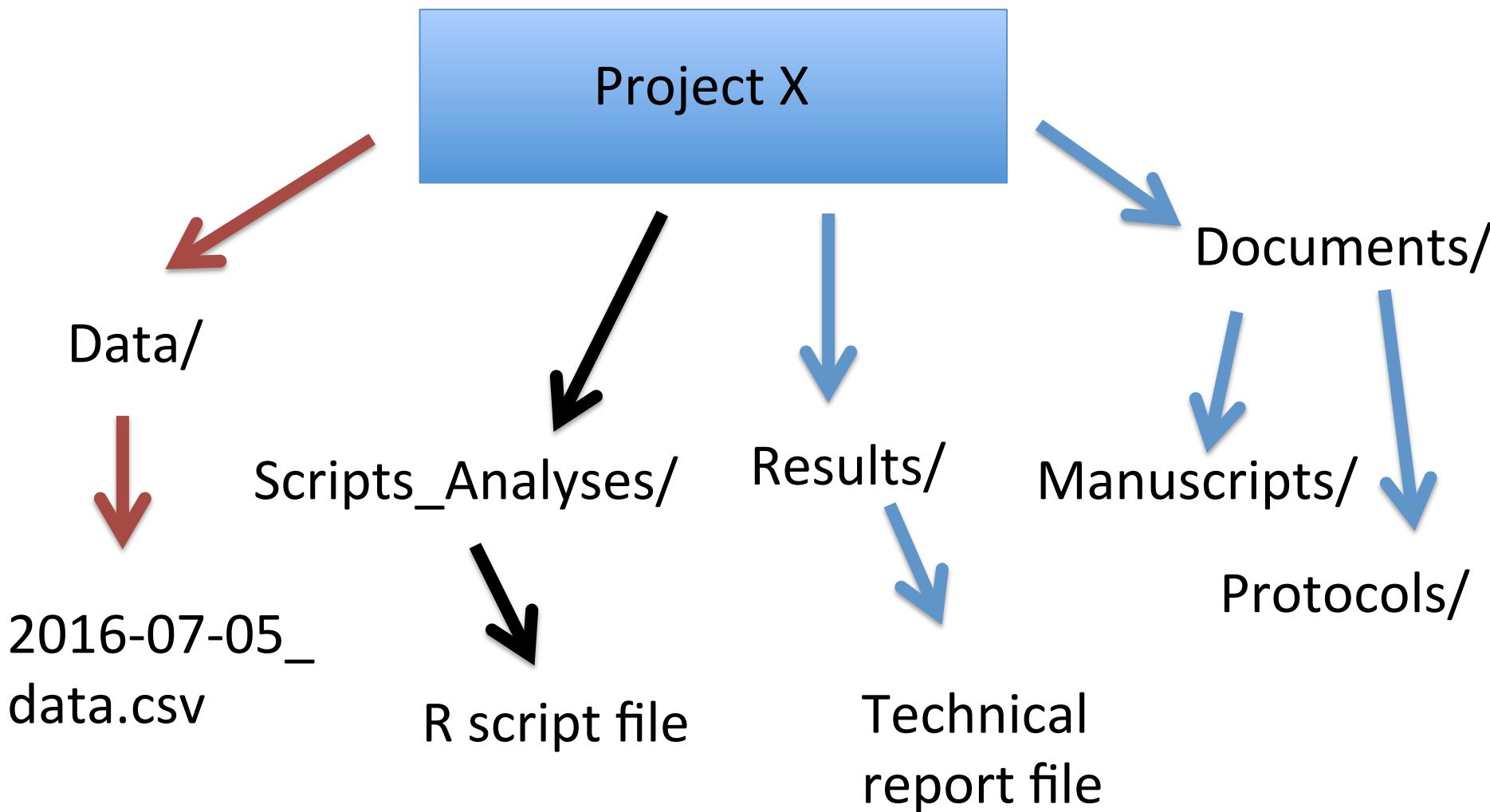
Organization

Transparency

Reproducibility

**Github**

# Organize by projects using a tree-like structure



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

**Let anybody access and repeat your analysis**



Organization ✓

Transparency

Reproducibility

**Github**

# But what if your project is a large tree?



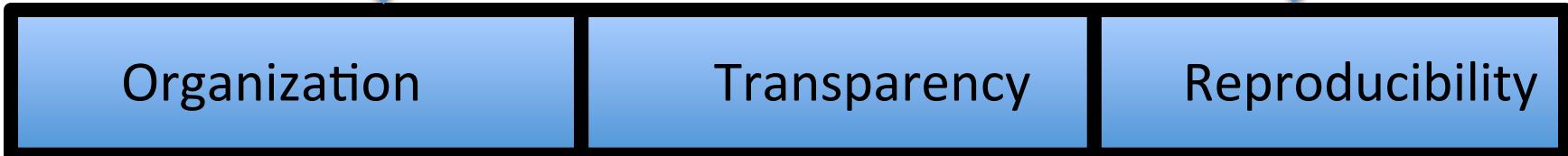
# If you didn't write it down, it didn't happen!

Track all of the changes to your project

Share it online!

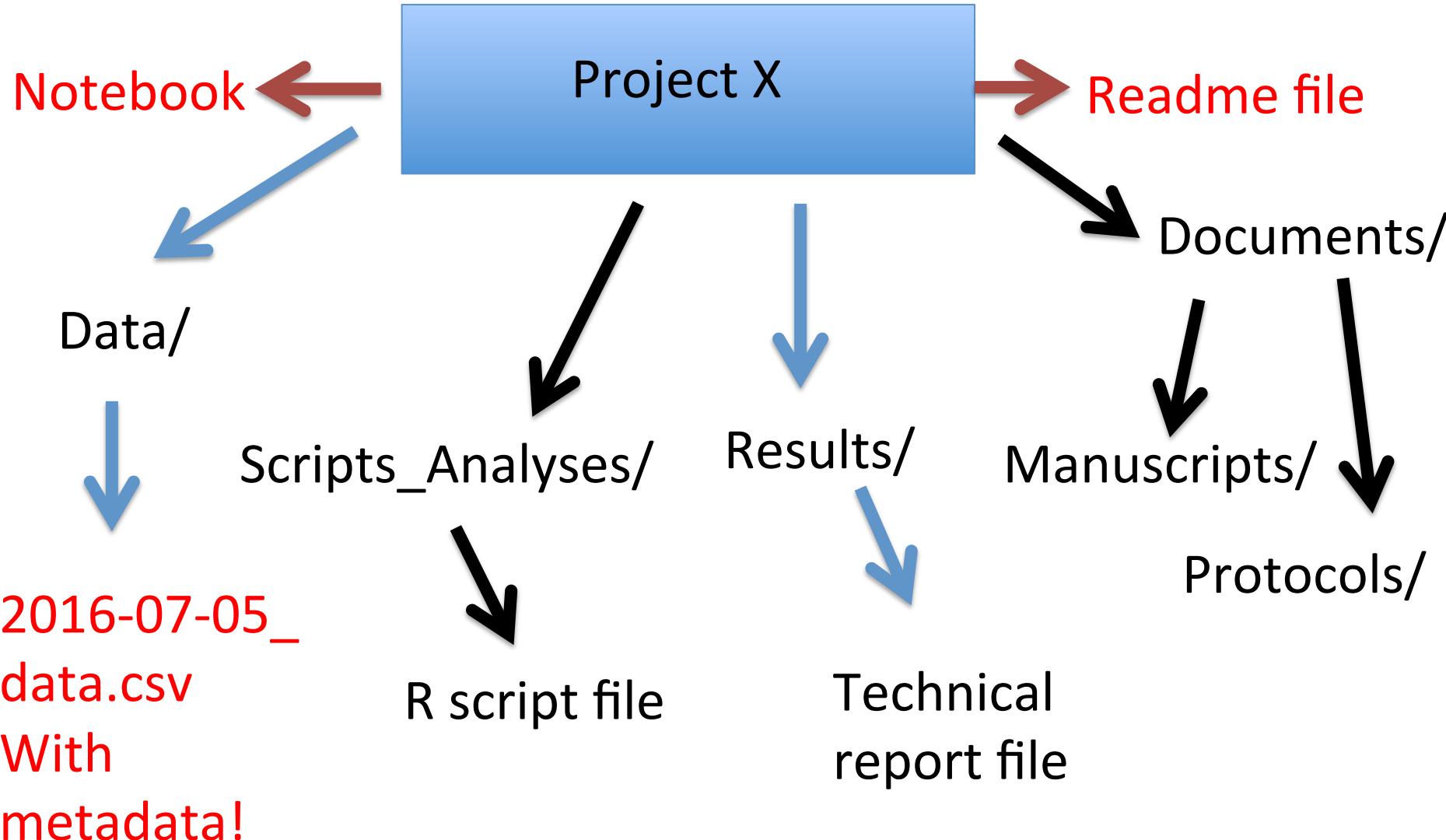
Let anybody access and repeat your analysis

Annotate everything...

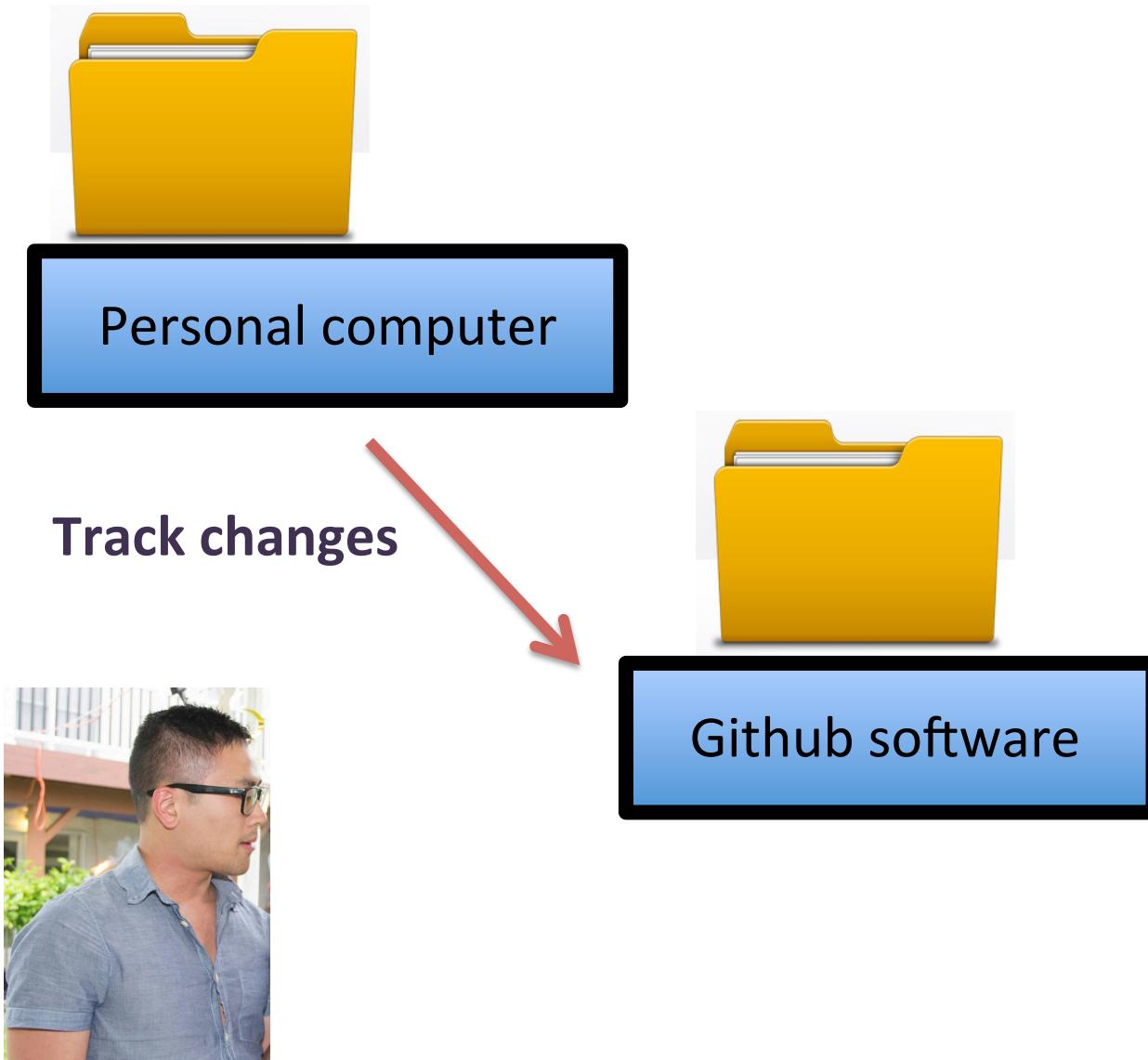


Github

# Organize by projects using a tree-like structure



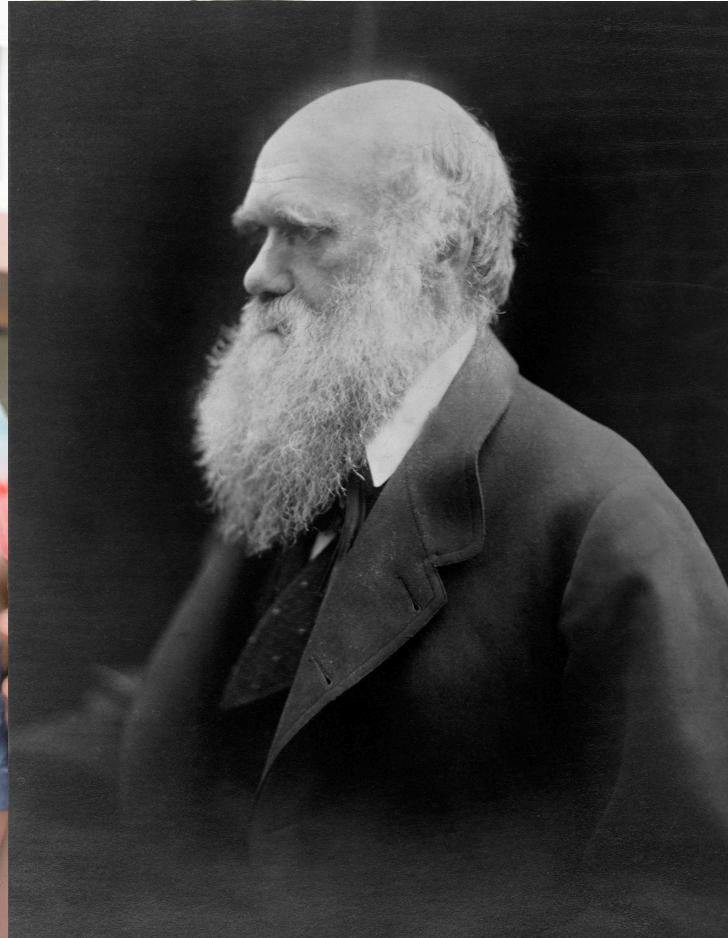
# Now, how do we track changes?



# Connects present and future self



# Connecting with Darwin?



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



Personal computer



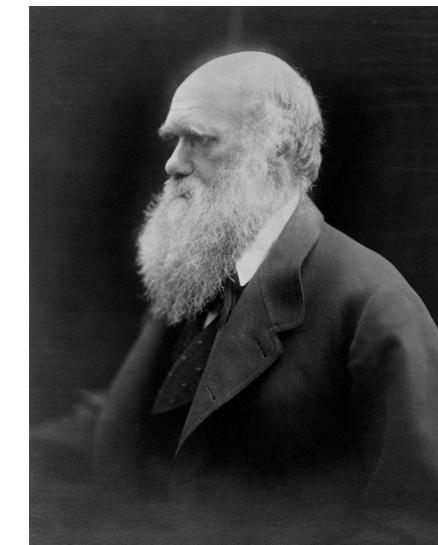
Github online

Track changes

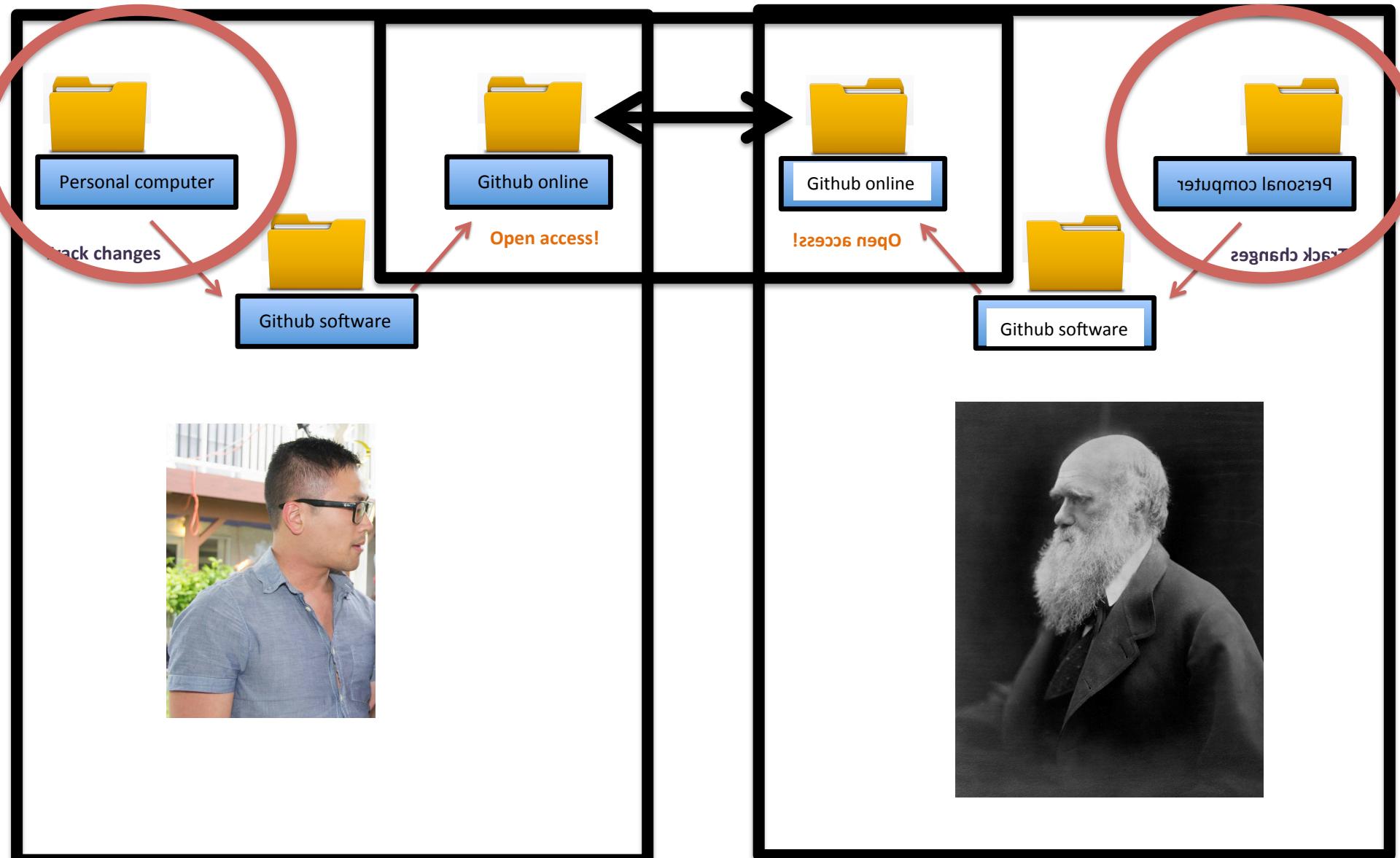


Github software

Open access!



# Connecting AND collaborating with Darwin



Test case: Undergraduate researcher

[http://adnguyen.github.io/2013\\_stressed\\_ants\\_in\\_warmingchambers/](http://adnguyen.github.io/2013_stressed_ants_in_warmingchambers/)

# Research papers: Journals should drive data reproducibility

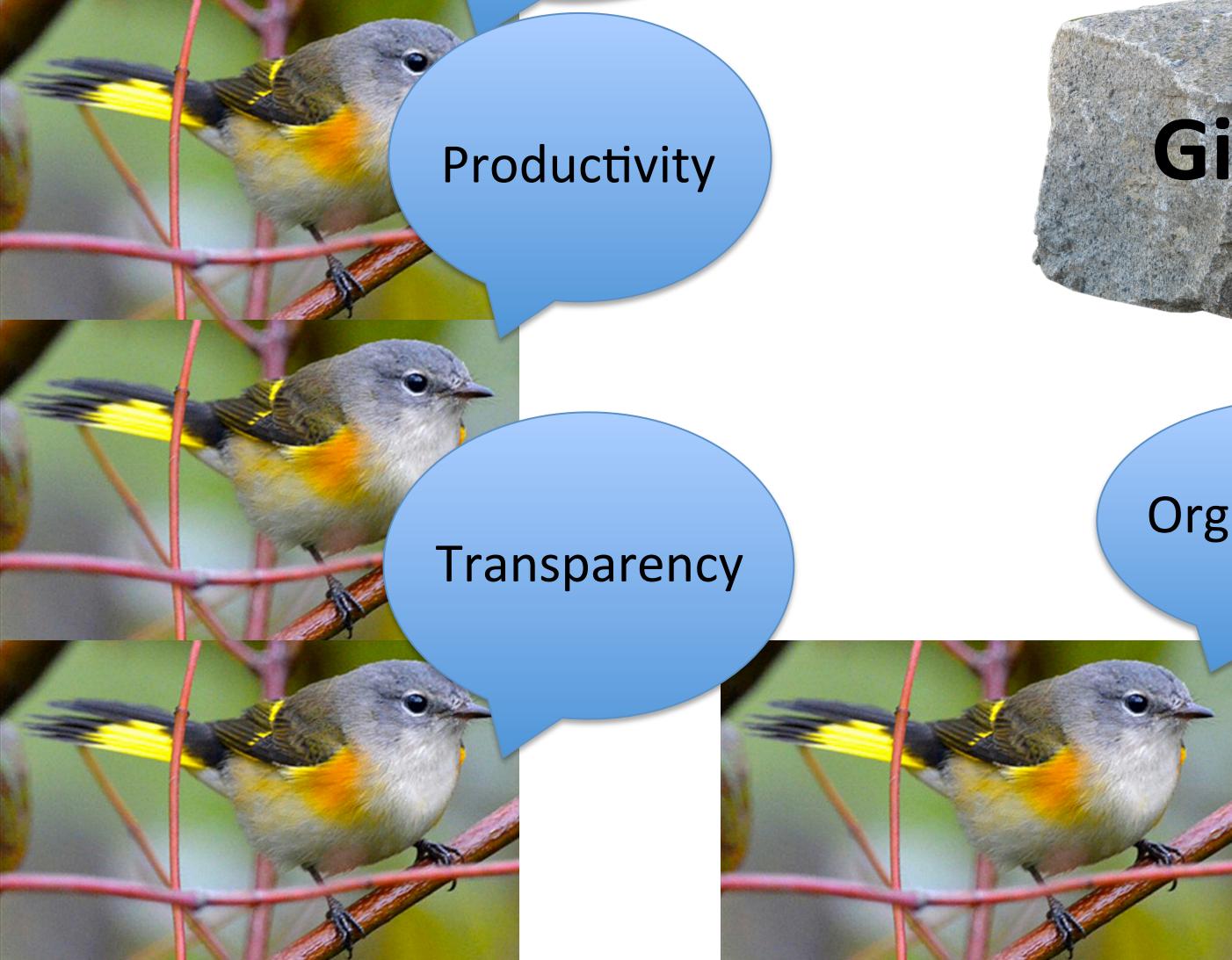
**Gregorio Santori**

*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

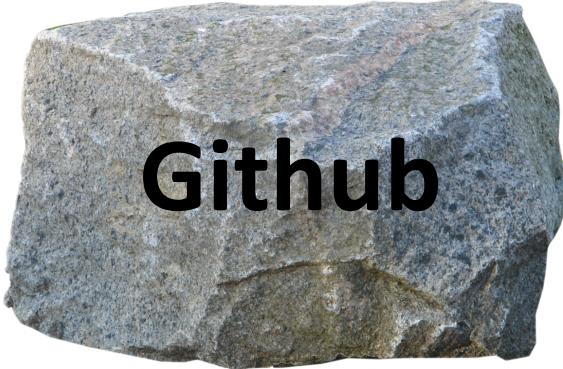
For manuscripts containing statistical analyses, journals should peer review only those papers that use statistics environments based on source code, enforcing the ban on 'point-and-click' statistical software (see [go.nature.com/29pdpc1](http://go.nature.com/29pdpc1)).



Reproducibility

Productivity

Transparency



Github

Organization