Collaborative notes from the talk "Five selfish reasons to working reproducibly":
https://www.youtube.com/watch?v=S8bU1CyEkRM&t=186s

write down tips useful for your future self

0. Intro + general thoughts

- no miracles in science -> be explicit instead :-)
- how to get from data to paper
- Descibe explicitely how to get from data to conclusions
- How to get from very dense data to a conscise/specific result
- Duke(?) disaster (bio-statisticians)
- Keep in mind: Scientific journals can make it to the press eventually
- cell lines + drugs -> see how they react -> claim can be used for practical results (transfer to real humans)
- bio-statisticians tried to reproduce (forensics)
- Make the point of the paper the result itself, not the unclear way you get to the result
- mistakes in analysis via excel ("the devils tool") -> shift via copy paste
- Do your data cleaning and preparation carefully. And deterministic -> Explore tools and choose wisely according to the nature of data/analysis
- other less innocent problems: a lot of duplicated samples in dataset + incoherent data -> no idea how researchers got to their data
- journal editors and original researchers did not really respond to criticism
- -> first author of paper has faked data and analysis!
- At any moment you can get contacted about your analysis or data. Be prepared!
- main problem is intransparency
- --> can happen to everybody
- **Take the time to check the results** (even if you are not the first author, your name is in there) coherence of the data lenght on the analsysis and your results,
- There are realistic (not idealistic) reasons to follow reproducibility. Specially the ones related to what it can do for you ;D

- Tidy up your work BEFORE submission
- It might take much time, but it worth the trouble at the end

1. Avoid disaster

- do back-ups! Physical storage devices are not always reliable, clouds aren't either -> multiple back-ups
- Write your report along with the experiments. Keep notes and document everything -> document from day 1
- Starting early reproducibility SAVES time
- easier to clean up mistakes

2. Easier to write papers

- supporting information in paper (e.g sweave file)
- Well documented and acccesible code+data **helps to**...
- reference easily
- update results
- make reading more engaging
- spot mistakes
- verify your analysis by others

3. Easier to talk to reviewers

- Make your paper reproducible to answer properly to Reviewer 2
- get reviewes engaged

4. Continuity of your work

- handing over work to those who follow you much easier
- continuing your own work is much easier

- make work for your future-self easier

5. Reputation

- packages and repos in CVs as outcome of research
- Journals care more and more about reproducibility -> this is the highest level of transparency asked

Baby steps to reproducibility

- Simplest thing: CLEAN the folder structure of your projects -> Tidy folder structure
- data:
- raw,
- clean,
- augmented
- scripts

When to worry about reproducibility:

1. Before starting
2. While doing the analysis
3. When writing your paper
4. When coatuthoring
5. After published
6. When reviewiing
7. always in general X)

- results
- docs
- Tidy data, uniform units and format for all entries. (tidy data standards)
- https://cran.r-project.org/web/packages/tidyr/vignettes/tidy-data.html