



American Society of Plant Biology, @ASPB



eLife Science, @eLife

# Reproducibility Resources & Tools

## Data management

**Harvard University Data Management page** <https://datamanagement.hms.harvard.edu/>  
**Kbroman Lab** <http://kbroman.org/dataorg/> (Short primer on data storage and handling from Kbroman)

**Purdue Library** <http://guides.lib.purdue.edu/c.php?g=353013&p=2378292> (Short primer on data management and file naming conventions)

**Data One Best Practices** <https://www.dataone.org/best-practices> (Detailed resource on how to handle data throughout its life-cycle)

**Mantra** <https://mantra.edina.ac.uk/> (Free online course for those who handle digital data)

## Electronic Lab Notebooks (ELN)

**Harvard University ELN guide** <https://tinyurl.com/Harvard-ELN> (Great summary about current ELNs and what they do)

**Benchling** <https://benchling.com/> (free)

**Evernote** <https://evernote.com/> (free and \$\$)

**Labguru** <https://www.labguru.com/> (\$)

**sciNote** <https://scinote.net/> (open source, free)

**Open Science Framework** <https://osf.io/> (free)

## Code

**Github** <https://github.com/> (code repository; free for public repos)

**Jupyter Notebooks** <http://jupyter.org/> (open source web-app for creating & sharing live code, equations, and more)

**Code Ocean** <https://codeocean.com/> (computational reproducibility platform; free to upload, share & publish executable code with DOI; pay for more computing time over freemium limit)

**Conda and BioConda** <https://conda.io/docs/> and <https://bioconda.github.io/> (A operating system independent package environment manager for the command line)

**Docker and Biocontainers** <https://docs.docker.com/> and <http://biocontainers.pro> (A container ecosystem to package code and data on the command line.)

**Binder** <https://mybinder.org/> (A tool to make your github repository an online docker image run in the cloud)

**Galaxy** <https://usegalaxy.org/> (A web and graphic interface based bioinformatics platform. Needs local set-up for larger data handling.)

## Reagents

**Addgene** <https://www.addgene.org/> (nonprofit plasmid repository)

**CiteAb** <https://www.citeab.com/> (antibody search engine with results sorted by citations)

**Quartzy** <https://www.quartzy.com/> (manage lab inventory)

## Methods

**Bio-Protocol** <https://bio-protocol.org/> (A peer-reviewed protocol journal; free to read & publish)

**protocols.io** <http://protocols.io/> (an open access repository of science methods; free to read & publish)

## Data

**DataDryad** <http://datadryad.org/> (curated digital repository; free to access, \$120 to publish dataset up to 20GB)

**Figshare** <http://datadryad.org/> (free digital repository, 5GB per file limit)

**Zenodo** <https://zenodo.org/> (free digital repository; 50GB per dataset limit)

## Data Visualization

**Beyond Bar Graphs: Free Tools & Resources for Creating More Transparent Figures for Small Datasets** <https://tinyurl.com/ecrbeyondbargraph>

**Interactive Dotplot Tool** <http://statistika.mfub.bg.ac.rs/interactive-dotplot/> (create dotplots, box plots, violin plots, show subgroups or display clusters of non-independent data):

**Interactive Linegraph Tool** (examine different summary statistics, focus on groups, time points or conditions of interest, examine lines for any individual in the dataset, view change scores):

<http://statistika.mfub.bg.ac.rs/interactive-linegraph/>

**Other free tools:** [https://twitter.com/T\\_Weissgerber/status/953334933019398145](https://twitter.com/T_Weissgerber/status/953334933019398145)

## R

**Tutorial - Plotting in R:**

[https://www.youtube.com/watch?v=sf\\_li1XV664&list=PL-IR12W3BZkXGfIjRtMgAw1Ff0liWX\\_Aj](https://www.youtube.com/watch?v=sf_li1XV664&list=PL-IR12W3BZkXGfIjRtMgAw1Ff0liWX_Aj)

**Customized interactive visualizations (Shiny):**

<https://www.frontiersin.org/articles/10.3389/fpsyg.2015.01782/full>

**Ggplot2**

<https://ggplot2.tidyverse.org/>

## Claus Wilke blog post

<http://serialmentor.com/blog/2018/1/23/fundamentals-of-data-visualization> (contains several links to his upcoming book about data visualization)

## Python

### Collection of useful resources:

[https://github.com/schmelling/python\\_materials](https://github.com/schmelling/python_materials)

### Tutorial - Data Analysis and Visualization in Python:

[Data Carpentry: An Introduction to Python for Data Analysis and Visualization - Tracy Teal PyCon 2016 Tutorial](#)

**PyData Packages** (incl. Matplotlib, Seaborn, Numpy, Pandas, and many more important for data analysis and visualization)

<https://pydata.org/downloads.html>

## Statistical Analysis

**Handbook of Biological Statistics!** <http://www.biostathandbook.com/> and <http://rcompanion.org/rcompanion/> (Web page from John H. McDonald and others from University of Delaware with pdf download links to free book on stats in Biology and its R implementation).

**Scipy stats lectures** <https://tinyurl.com/scipystats> (Lecture on stats in python using scipy) see also <https://www.statsmodels.org/stable/index.html> for more stats in python

### Nature Stats for Biologist resources

<https://www.nature.com/collections/qghhqm/content/practical-guides>

# Practical tips for reproducibility

1. Plan for reproducibility before you start
  - a. **Write a study plan or protocol** and track new versions.
  - b. **Set-up a reproducible project** using an electronic lab notebook to organize and track your work. Avoid saving proprietary file formats.
2. Keep track of things
  - a. **Preregister** important study design and analysis information. Free tools to help you make your first registration include [AsPredicted](#), [Open Science Framework](#), and [Registered Reports](#). Clinical trials use [Clinicaltrials.gov](#).
  - b. **Track changes** to your files using version control.
  - c. **Document** everything done by hand in a README file and data dictionary. **Karl Broman's Data Organization module:**  
<http://kbroman.org/dataorg/pages/dictionary.html>
3. Report your research transparently
  - a. **Share your protocols and interventions** explicitly and transparently.



- Constructs: [http://www.addgene.org/Nicole\\_King](http://www.addgene.org/Nicole_King)
- **Implicating candidate genes at GWAS signals by leveraging topologically associating domains**
  - Paper: <https://dx.doi.org/10.1038/ejhg.2017.108>
  - Code: <https://zenodo.org/record/163950#.W0hqTdJKjIU>
  - Docker workflow: <https://zenodo.org/record/166556#.W0hqc9JKjIU>
- **mcSCRB-seq: sensitive and powerful single-cell RNA sequencing**
  - Protocol: [dx.doi.org/10.17504/protocols.io.p9kdr4w](https://dx.doi.org/10.17504/protocols.io.p9kdr4w)
  - Paper: <https://doi.org/10.1101/188367>
  - Code: [https://github.com/cziegenhain/Bagnoli\\_2017](https://github.com/cziegenhain/Bagnoli_2017)
- **TransRate: reference-free quality assessment of de novo transcriptome assemblies**
  - Paper: <https://dx.doi.org/10.1101%2Fgr.196469.115>
  - Code: <https://github.com/Blahah/transrate>
  - Tutorial: <http://hibberdlab.com/transrate/>
- **Genomic insights into members of the candidate phylum Hyd24-12 common in mesophilic anaerobic digesters**
  - Paper: <https://doi.org/10.1038/ismej.2016.43>
  - Code: <https://github.com/Kirk3gaard/Publications>
- **Experimenting with Reproducibility: a case study of Robustness in Bioinformatics**
  - Paper: <https://doi.org/10.1093/gigascience/giy077>
  - Code: <https://github.com/sje30/waverepo>
- **A Bayesian Mixture Modelling Approach For Spatial Proteomics**
  - Paper: <https://doi.org/10.1101/282269>
  - Code: <https://github.com/lgatto/2018-tagm-paper>



@Addgene



protocols.io

@ProtocolsIO



CODE OCEAN  
BETA

@CodeOceanHQ