Reproducibility for everyone

https://tinyurl.com/plantbio-repo
CC BY 4.0
Why does reproducibility matter to you?
Have you ever had problems reproducing your own or someone else's research?

60 responses

- My own research: 7 (11.7%)
- Someone else's: 15 (25%)
- Both: 28 (46.7%)
- I have never had problems reproducing research: 11 (18.3%)
Goals and objectives

• ‘Reproducibility’ framework
• ‘Reproducibility’ tools
• Starting point of a ‘lifelong’ journey
What does reproducibility mean?
What are the different modes of reproducibility?
Is reproducibility all that matters?

‘Reproducibility’ tool shed.
  o organization
  o documentation
  o analysis
  o dissemination
What does reproducibility mean?
What does reproducibility mean?

**Reproducible research**: Authors provide all the necessary data and the computer codes to run the analysis again, re-creating the results.

**Replication**: A study that arrives at the same scientific findings as another study, collecting new data and completing new analyses.
What are the different modes of ‘reproducibility’?

<table>
<thead>
<tr>
<th>Methods</th>
<th>Same experimental system</th>
<th>Different experimental system</th>
</tr>
</thead>
<tbody>
<tr>
<td>Same methods</td>
<td>Reproducibility</td>
<td>Replicability</td>
</tr>
<tr>
<td>Different methods</td>
<td>Robustness</td>
<td>Generalizability</td>
</tr>
</tbody>
</table>

Schloss, 2018
10.1128/mBio.00525-18
What are the different modes of ‘reproducibility’?

<table>
<thead>
<tr>
<th>Methods</th>
<th>Same experimental system</th>
<th>Different experimental system</th>
</tr>
</thead>
<tbody>
<tr>
<td>Same methods</td>
<td>Reproducibility</td>
<td>Replicability</td>
</tr>
<tr>
<td>Different methods</td>
<td>Robustness</td>
<td>Generalizability</td>
</tr>
</tbody>
</table>

Reproducibility is the minimum standard for science.
Is reproducibility all that matters?
Introduction
No one is perfect!

Every little helps!

Transparent and open science!

Everyone starts somewhere!

Casadevall and Fang, 2016
10.1128/mBio.01902-16
Introduction

Factors decreasing reproducibility
Introduction

Factors decreasing reproducibility

- Technical Factors
- Human Factors
- Study Design & Statistics
- Rewards & Incentives
Factors decreasing reproducibility

**Technical Factors**
- Natural variability
- False cell lines
- Sensitivity to conditions, equipment
- Batch effects

**Human Factors**

**Study Design & Statistics**

**Rewards & Incentives**

Introduction
Introduction

Factors decreasing reproducibility

- Natural variability
- Bad reagents
- False cell lines

**Technical Factors**

- Sensitivity to conditions, equipment
- Batch effects

- Confirmation bias
- Poor record keeping
- Poor sharing (reagents, methods, data, code)

**Human Factors**

**Study Design & Statistics**

**Rewards & Incentives**
Introduction

Factors decreasing reproducibility

Technical Factors
- Natural variability
- False cell lines
- Bad reagents
- Sensitivity to conditions, equipment
- Batch effects
- Lack of version control

Study Design & Statistics
- Design flaws
- P-hacking
- HARKing
- Selective reporting
- Misunderstanding statistics

Human Factors
- Confirmation bias
- Poor record keeping
- Poor sharing (reagents, methods, data, code)

Rewards & Incentives
- Misunderstanding statistics
Introduction

Factors decreasing reproducibility

- Natural variability
- False cell lines
- Bad reagents

**Technical Factors**
- Sensitivity to conditions, equipment
- Batch effects
- Lack of version control
- Design flaws

**Study Design & Statistics**
- P-hacking
- HARKing
- Selective reporting
- Misunderstanding statistics

**Human Factors**
- Confirmation bias
- Poor record keeping
- Poor sharing (reagents, methods, data, code)

**Rewards & Incentives**
- Lack of training
Factors decreasing reproducibility

Introduction

Factors Decreasing Reproducibility

- Natural Variability
- False Cell Lines
- Bad Reagents

Technical Factors

- Sensitivity to Conditions, Equipment
- Batch Effects
- Lack of Training
- Lack of Version Control
- Design Flaws
- Misleading Statistics

Mistakes

- P-Hacking
- Selective Reporting

Study Design & Statistics

- Poor Record Keeping
- Confirmation Bias
- Poor Sharing (Reagents, Methods, Data, Code)
- Hypercompetition & Pressure
- Wrong Incentives
- Paywalls
- Fraud

Rewards & Incentives

- Lack of Training
- Lack of Technical Skills
“So what’s a careful scientist to do? First and foremost, be aware of the conditions around you that may increase the risk of irreproducible results, whether they are bad ingredients, dubious statistical traditions, or outside pressures that can shape behavior. Also take heart. This reproducibility “crisis” isn’t really a crisis at all. These are not new problems. Rather, I think of this moment as an awakening. And that’s a good thing, because we need to recognize that a problem exists before we can seek solutions.”

https://cen.acs.org/articles/95/i47/Reproducibility-issues.html
There is good news!

“So what’s a careful scientist to do? First and foremost, be aware of the conditions around you that may increase the risk of irreproducible results, whether they are bad ingredients, dubious statistical traditions, or outside pressures that can shape behavior. Also take heart. This reproducibility “crisis” isn’t really a crisis at all. These are not new problems. Rather, I think of this moment as an awakening. And that’s a good thing, because we need to recognize that a problem exists before we can seek solutions.”

https://cen.acs.org/articles/95/i47/Reproducibility-issues.html
What can we do by the end of the century?

Cori Bargmann, HHMI investigator, President CZI Science

"82 years ago, there were no antibiotics & we didn't know that smoking causes lung cancer... We can expect a lot from the next 82 years."
What can we do by the end of the century?

Cori Bargmann, HHMI investigator, President CZI Science

"82 years ago, there were no antibiotics & we didn't know that smoking causes lung cancer... We can expect a lot from the next 82 years."

"Where can we be in 82 years if we accelerate science?"
Where is your greatest potential for growth?
Where is your greatest potential for growth?

More detailed methods, analysis and record keeping
More publicly available data including meta-data

Fewer incentives to be first rather than right
Better reagent sharing e.g. plasmids, antibodies ...
<table>
<thead>
<tr>
<th>organization</th>
<th>documentation</th>
<th>analysis</th>
<th>dissemination</th>
</tr>
</thead>
</table>

Data management

http://kbroman.org/dataorg/

https://dmptool.org/
I cannot find this file!

Where is my file?

What version was it?

What did I call that file again?

Was this the wild type picture or the mutant one?

Where is my RAW!!! data?

Have a plan! Be happy!
Think about….

- **What** data will be produced as a part of the project
- **How** each type of data will be organized, documented, standardized, stored, protected, shared and archived
- **Who** will take responsibility for carrying out the activities listed above, and
- **When** these activities will take place over the course of the project (and beyond)
- **Metadata**

http://guides.lib.purdue.edu/c.php?g=353013&p=2378292

https://www.dataone.org/best-practices
<table>
<thead>
<tr>
<th>organization</th>
<th>documentation</th>
<th>analysis</th>
<th>dissemination</th>
</tr>
</thead>
</table>

Project directory structure

- Project_1
  - methods
  - raw_data

- analysis

- scripts
- manuscript

readme and/or ELN link

Inspired by 'Bioinformatic data skills' Vincent Buffalo
Project directory structure

Project_1
  methods
  raw_data
    readme
  analysis
    analysis_method_1
      2017
      2018
    analysis_method_2
  scripts
  manuscript
    text
      version_1
  readme and/or ELN link

Inspired by ‘Bioinformatic data skills’ Vincent Buffalo
Always keep raw data!

Always backup your data! X 3

Inspired by ‘Bioinformatic data skills’ Vincent Buffalo
Always keep raw data!

THE FOUR STAGES OF DATA LOSS
DEALING WITH ACCIDENTAL DELETION OF MONTHS OF HARD-EARNED DATA

Inspired by ‘Bioinformatic data skills’ Vincent Buffalo
How did you call the last file you generated?

Did you have a plan?
File naming convention (FNC)

- Test_data_2013
- Project_Data
- Design for project.doc
- Lab_work_Eric
- Second_test
- Meeting Notes Oct 23

http://guides.lib.purdue.edu/c.php?g=353013&p=2378292
File naming convention (FNC)

- Include date in yyyy-mm-dd format
- Use meaningful abbreviations
- Have group identifiers
- Document your decisions
- Be consistent
**File naming convention (FNC)**

- Include data in yyyy-mm-dd format
- Use meaningful abbreviations
- Have group identifiers
- Document your decisions
- Be consistent

**Example:**

```
20130825_DOEProject_Ex1Test1_Data_Gonzalez_v3-03.xlsx
```

<table>
<thead>
<tr>
<th>Date</th>
<th>Project</th>
<th>Type</th>
<th>Version</th>
<th>Experiment</th>
<th>ID</th>
</tr>
</thead>
<tbody>
<tr>
<td>Specificity</td>
<td>Specific</td>
<td>General</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Get organized! Be happy!

Findable
Accessible
Interoperable
Reusable

https://tinyurl.com/plantae-FAIR
Electronic Notebooks
Paper Lab-notebooks - in use since the 15th Century!

Good record keeping is important for

- Dissemination of ideas, findings
- Legally binding record that protects intellectual property

Not searchable!

Can be easily damaged, misplaced.
Not easy to back up.

Hard to share with collaborators.
Why should you use an **Electronic Lab Notebook**?

Store text electronically, Attach Images, 15 GB data limit, Multiple authors possible

Good for sharing data, 2 GB cloud storage limit (free version)

**+ Many more Features!**

- Searchable
- Export data as PDF (must back-up data regularly)
- Easily shareable
- Embed high res images, protocols etc
- Easily accessible world over
- Use the mobile App to quickly upload images
Cost considerations - Softwares available

Paid for— Bio-Itech, LabArchives, LabGuru

Paid (with free version)— SciNote, Benchling

Open source— Open wet ware, ELOG

Free— OSF (Open Science Framework), LocalWiki
One-size does not fit all!

Parameters to consider

ELN Features Matrix 42

https://datamanagement.hms.harvard.edu/electronic-lab-notebooks
Basic features of an Electronic Lab notebook

- Data shared with authorized personnel e.g.: Supervisor
- Office compatible
- Search by keyword, date
- Can print, share
- Attachment
- Each entry is dated
- Organize as needed
General tips on electronic record keeping

- Back-up data regularly
- Maintain a physical notebook in parallel
- Mobile apps provide added portability
- If using free ELNs, check privacy policies
‘Wet lab’ protocol sharing
Looking for protocol in 1997 paper: "as described in (x) et al '96". Finds '96 paper: "as described in (x) '87." Finds '87 paper: Paywall.
Looking for protocol in 1997 paper: "as described in (x) et al '96. Finds '96 paper: "as described in (x) '87." Finds '87 paper: Paywall.
2017: “Devices were fabricated as previously described [ref 8]”

[ref 8] 2015: “Devices were fabricated as previously described [ref 4]”

[ref 4] 2013: “Devices were fabricated as previously described [ref 2]”

[ref 2] 2009: “Devices were fabricated with conventional methods”
How to overcome this problem? - Don’t contribute to it!

Methods section could read

*We draw the owl on 60.2 gsm white paper of the A4 dimension (210mm by 297mm), using 3H and 6B graphite pencils (Derwent, Cumbria, UK). We did so by looking at owls, and drawing what we saw on paper. This protocol yielded one drawn owl.*

https://medium.com/@tpoi/do-the-rest-of-the-fucking-analysis-8fcef22fd991
Write Detailed Protocols

- Think of a protocol as a brief, modular and self-contained scientific publication.
- Include a 3-4 sentence abstract that puts the methodology in context.
- Include as much detail as possible (Duration/time per step, Reagent Amount, vendor name, Catalog number, Expected result, Safety information, Software package)
- Chronology of steps.
- Notes, recipes, tips, and tricks

https://www.protocols.io/view/how-to-make-your-protocol-more-reproducible-discov-g7vzn6
Share protocols on the right platforms

Methods
A detailed version of the protocol described here is available at protocols.io [21].

Availability of supporting source code and requirements

— Project name: PRIMAL, Pipeline of Root Image analysis using MACHine Learning
— Project home page: https://plantmodelling.github.io/primal/
— Operating system(s): platform independent
— Programming language: R
— Other requirements: none
— License: GPL
Share protocols on the right platforms

1. **Bio-protocol** (free to read & publish, but need invitation or pre-submission inquiry)
2. **JOVE** - Journal Of Visual Experiments (nice videos but costly and not open access)
3. **protocols.io** (free to read & publish but not peer-reviewed)
What should Ben do?

Ben is really excited to join a new team that is performing a chemical screen of plant growth regulators on root architecture. However,

- The previous Postdoc started a new job and refuses to respond to his emails.
- The technician on the project was only involved in the data acquisition steps.
- Unfortunately, the lab notebook went missing in a recent move to a new floor.
  - The methods section in a previous paper reads like this -

  **Materials and Methods**
  Plants were grown on appropriate media and roots photographed. Images were analyzed using WinRhizo (Arsenault, J-L., et al. 1995) and data presented as graphs.

Identify the problem(s)?

Suggest a solution.
‘Wet lab’ reagent sharing
Problems with wet-lab reagent availability

Scientist creates & publishes on a reagent

Scientist leaves the lab and stores reagent in freezer

Other scientists request the reagents, but no one remaining remembers where they are
Problems with wet-lab reagent availability

- Wasted time, money, and resources when reagents are recreated
- Mistakes in recreation can lead to spurious results
- Individual labs don’t usually have the resources to:
  - Keep track of all reagents created in lab
  - Consistently validate all reagents in the lab
  - Properly label and store all reagents
  - (Legally) distribute all reagents to interested researchers
- Reagents repositories are part of the solution!
Functions of reagent repositories

They:
- Verify reagents
- Curate reagents
- Facilitate and track shipping
- Protect IP

Process is easier if you:
- Record how a reagent was created
- Provided associated publications
- Provided associated protocols

(All of these are facilitated by other tools discussed in this workshop)
Examples of Reagent repositories

- Addgene
- DNASU
- ATCC
- NCI Mouse Repository
- Coriell Institute
- ABRC
- The Bloomington Drosophila Stock Center
- Developmental Studies Hybridoma bank
Incentivizing reagent sharing

Direct

- Archiving
- Reducing time spend sending out reagents
- Occasional monetary benefits

Indirect

- Creation of educational content
- Direct promotion
- Analysis of reagent distribution
Addgene: The nonprofit plasmid repository

**Goal:** To accelerate science by improving access to research materials and information

**Issues Addressed:** Difficulties in obtaining, verifying, and using plasmids from other labs

**Audience:** Academic and nonprofit institutions doing biology research and using plasmids

**Services:**
- Stores and distributes plasmids and viral vectors
- Verifies plasmids and viral vectors through DNA sequencing with some functional testing
- Collates/curates information about plasmids and viral vectors
- Produces and freely distributes educational content to make it easier for scientists to learn about and use new technologies
Bioinformatic tools
Dependency hell

What version of the program, data etc… did I use?

Why did I do this?

https://software-carpentry.org/

‘Bioinformatic data skills’ Vincent Buffalo
Notebooks

- Keep track of analysis
- Interactive coding
- Interactive data exploration
- Imbedded visualization
- Easy access to docstrings
- Mix of code and documentation

https://jupyter.org/documentation  
https://www.rstudio.com/
Notebooks

- Keep track of analysis
- Interactive coding
- Interactive data exploration
- Imbedded visualization
- Easy access to docstrings
- Mix of code and documentation

- Over 40 programming languages
- Easily shared
- Widgets
- Interactive plots
- Run remotely on server

https://jupyter.org/documentation
https://www.rstudio.com/
In this Notebook we explore the Lorenz system of differential equations:

\[
\begin{align*}
\dot{x} &= \sigma(y - x) \\
\dot{y} &= px - y - xz \\
\dot{z} &= -\beta z + xy
\end{align*}
\]

Let's call the function once to view the solutions. For this set of parameters, we see the trajectories swirling around two points, called attractors.

```python
In [4]: from lorenz import solve_lorenz
t, X_T = solve_lorenz(N=10)
```

![Graph of the Lorenz system](image)

```python
# Define the Lorenz derivatives
def lorenz_deriv(x, y, z, t, sigma, beta, rho=28.0):
    '''Compute the time-derivative of a Lorenz system.'''
    return [sigma * (y - x), x * (rho - z) - y, x * y - beta * z]

# Choose random starting points, uniformly distributed from -15 to 15
np.random.seed(1)
x0 = -15 + 30 * np.random.rand(3)
```
Version control

- Records changes
- Keeps track of change history
- Illustrates changes between versions
- Lets you share your code easily
- Lets you collaborate on your code more easily
Version control
Version control

https://git-scm.com/doc

http://smutch.github.io/VersionControlTutorial//
Version control
Version control

Google docs does history tracking.

https://git-scm.com/doc

http://smutch.github.io/VersionControlTutorial//
How to I install all these different software packages???

Dependency hell

Version conflict
Package, dependency, and environment manager

- Handles installs and dependencies
- Allows for multiple independent environments
- Easily configurable
- Allows for manual installs as well
- Runs on all three major systems
- Open source

- You can package your own work and contribute
It is really that simple….

conda install bwa

Or a new environment can be created:

conda create -n aligners bwa bowtie hisat star

…. most of the time
Docker runs images as containers that are

- self contained with all code, programs, libraries included. No subsequent installation required.
- Isolated
- Portable including dissemination
- Lightweight
Containers

Get local blast

$ docker pull biocontainers/blast

Run local blast

$ docker run biocontainers/blast blastp -help
Containers

Turns a GitHub repo with data and notebooks into a collection of interactive notebooks run in the cloud executable

Configuration, preservation, & reuse of executable code using containers for researchers

<table>
<thead>
<tr>
<th>organization</th>
<th>documentation</th>
<th>analysis</th>
<th>dissemination</th>
</tr>
</thead>
</table>

Data sharing
Data sharing

What to share?
- Share research data and code that is necessary to validate findings & reproduce results of research outputs
- Share data and code that might be valuable to other researchers or policy-makers
- Share data and code which cannot be (easily) re-generated

Why share?
- Funder or publisher mandates
- Citation benefits (Piwowar 2013, https://doi.org/10.7717/peerj.175)
- Preserve long-term access to data

How to share?
- Choose open, persistent, and non-proprietary file formats
- Create and share documentation to enable reuse
- Include data citations of source data
- Create rich metadata
Data sharing

- Use a data repository, not your website!
  - Repositories provide
    - Persistent identifiers for your data like a DOI
      - Unique and citable
      - Prevents "link rot"
    - Persistent access
    - Preservation
    - Backup
    - Management of access
    - Versioning
    - Licensing

Specify a data licence:
- Consider Creative Commons licenses for data and text, either CC-0 or CC-BY.
  Guidance on data licenses from the Digital Curation Center: [http://www.dcc.ac.uk/resources/how-guides/license-research-data](http://www.dcc.ac.uk/resources/how-guides/license-research-data)

Specify a code licence:
- Consider an open source license such as the MIT, BSD, or Apache license.
  Guidance on software licenses from Karl Broman: [http://kbroman.org/steps2rr/pages/licenses.html](http://kbroman.org/steps2rr/pages/licenses.html) and Open Source Initiative: [https://opensource.org/licenses](https://opensource.org/licenses)
Data sharing

Identify mandated or disciplinary repository:
- Funder specified repository
- Institutionally specified data repository
- Domain or discipline-specific data repository
  - Find and compare disciplinary repositories using the Repository of Research Data Repositories [https://www.re3data.org/](https://www.re3data.org/)

In addition to a specified data repository, you can make a deposit to a general purpose repository:
- DataDryad [http://datadryad.org/](http://datadryad.org/) (curated digital repository; free to access, $120 to publish dataset up to 20GB)
- Figshare [https://figshare.com/](https://figshare.com/) (free digital repository, 5GB per file limit)
- Zenodo [https://zenodo.org/](https://zenodo.org/) (free digital repository; 50GB per dataset limit)
Image handling and analysis
Challenges:

- reproducibility of image capture
- manage data sets
- accurately represent 3D data in a 2D format
- reproducibility of image analysis
Reproducibility of image capture

Issues
- Variability in fluorescent lines/staining
- Variability in laser power detector sensitivity
- Differing sensitivity of markers
- Variations in expression due to time of day, developmental stage, growth conditions

Possible solution
More detailed methods?
Manage image data

Issues
- Storing images and meta data with sufficient information that on how they were captured, what they are and how they were processed.
- Files are often very large and in a commercial format.

Possible solutions

https://www.openmicroscopy.org/

https://www.openmicroscopy.org/omero/
Accurately represent 3D data in a 2D format

Issues
- showing single slices is not representative
- max projections vs average projections give different impression

Possible solutions
- Share raw data
- movies of time-course data
- movies to show full z-stacks
Reproducibility of image analysis

Issues
- can another researcher reproduce your analysis?

Possible solution

https://fiji.sc/
Fiji/ ImageJ - open source software for image analysis

- Enables automated renaming conversion of image formats to reduce human error
- Quantitative image analysis
- Use huge number of plugins/macros available or write own
- Enables sharing of analysis pipeline with the data
- Interacts with OME
MorphoGraphX is an open source application for the visualization and analysis of 4D biological datasets. Developed by researchers, it is primarily used for the analysis and quantification of 4D live-imaged confocal data.

- Enables projections of stacks, movies and 3D rendering of confocal stack
- Quantification of geometry and signal intensity
- Records processing that occurs with the data
- Can automate segmentation with macros
- Can share pipeline of image analysis to enable reproducibility of the analysis

http://www.mpipz.mpg.de/MorphoGraphX
Data Analysis and Visualization
Data presentation is the foundation of our collective scientific knowledge…

Figures are especially important. They often show data for key findings.
What is good DataViz?

Effective figures should:

1. Immediately convey information about the study design
2. Illustrate important findings
3. Allow the reader to critically evaluate the data

Weissgerber et al. 10.1074/jbc.RA117.000147
The usual way and its flaws

Issues:

- Reproducible Workflows?
  - Problems can be avoided by using macros or dashboards
  - However, who uses these?

- Excel Renames Genes
  - 20% of papers in leading genomic journals contain gene list errors

- Default Plots are often Bar Charts and Line Plots
Why does DataViz matter for reproducibility?

<table>
<thead>
<tr>
<th>Test</th>
<th>Symmetric</th>
<th>Outlier</th>
<th>Bimodal</th>
<th>Unequal n</th>
</tr>
</thead>
<tbody>
<tr>
<td>T-test: Equal var.</td>
<td>0.035</td>
<td>0.050</td>
<td>0.026</td>
<td>0.063</td>
</tr>
<tr>
<td>T-test: Unequal var.</td>
<td>0.035</td>
<td>0.050</td>
<td>0.026</td>
<td>0.035</td>
</tr>
<tr>
<td>Wilcoxon</td>
<td>0.054</td>
<td>0.073</td>
<td>0.128</td>
<td>0.103</td>
</tr>
</tbody>
</table>

Weissgerber et al., 10.1371/journal.pbio.1002128
Why does DataViz matter for reproducibility?

Bar Charts Don’t Allow You to Critically Evaluate Continuous Data

Weissgerber et al., 2017 JBC, http://www.jbc.org/content/292/50/20592.full
How to Choose the Right Plot

<table>
<thead>
<tr>
<th></th>
<th>Dotplot</th>
<th>Boxplot with points</th>
<th>Boxplot</th>
<th>Violin plot (with or without points)</th>
<th>Bar graph</th>
</tr>
</thead>
<tbody>
<tr>
<td>Outcome variable</td>
<td>Continuous</td>
<td>Continuous</td>
<td>Continuous</td>
<td>Continuous</td>
<td>Counts &amp; proportions</td>
</tr>
<tr>
<td>Sample size</td>
<td>Small</td>
<td>Medium</td>
<td>Large</td>
<td>Medium to Large</td>
<td>Any</td>
</tr>
<tr>
<td>Data distribution</td>
<td>Any</td>
<td>Any</td>
<td>Do not use for bimodal data</td>
<td>Any</td>
<td>N/A</td>
</tr>
</tbody>
</table>
Making Effective Dotplots

Step 1: Make all data points visible

A: Ineffective graph
B: Decrease point size
C: Semi-transparent points
D: Random jitter
E: Symmetric jitter

Step 2: Emphasize summary statistics

A: Ineffective graphic
B: Increase width
C: Emphasize summary statistics

Tracey Weissgerber Personal Communication
One Step Further

Interactive Dot Plot -
http://statistika.mfub.bg.ac.rs/interactive-dotplot/

Interactive Line Graph -
http://statistika.mfub.bg.ac.rs/interactive-linegraph/
Some Intermediate Options

https://plot.ly/create/#/

https://www.datawrapper.de/
Which programming language should I use?

- Select a language that is used in your lab or community
- *Select a general purpose language such as Python to start with if you don’t have a specific problem. That way you learn basic programming skills, which allows you to switch to other languages more easily, and you can tackle different problems. You usually learn multiple languages anyway.
Programming Languages

- Anaconda (Distribution)
- Numpy & Pandas (Data Wrangling)
- Scipy (Higher Math)
- Matplotlib (Basic Graphs)
- Seaborn, Bokeh, Altair, Plotly (Advanced Statistical & Interactive Graphs)
- Jupyter notebook / lab (Interactive Notebook)

- tidyverse (Distribution)
- dplyr & tidyr (Data Wrangling)
- ggplot / ggplot2 (Basic Graphs)
- shiny / RMarkdown (Interactive Notebook)
- RStudio (Interactive Notebook)
Dealing with Data

- **Provide Open-Source Data** *(Rule 2 of Enable Multi-site Collaborations through Data Sharing)*
- **Keep Raw Data Raw** *(Rule 3 of Digital Data Storage)*
- **Store Data in Open Formats** *(Rule 4 of Digital Data Storage)*
- **Data Should Be Structured for Analysis** *(Rule 5 of Digital Data Storage)*
- **Data Should Be Uniquely Identifiable** *(Rule 6 of Digital Data Storage)*
- **Link Relevant Metadata** *(Rule 7 of Digital Data Storage)*
- **Have a Systematic Backup Scheme** *(Rule 9 of Digital Data Storage)*
- **Archive The Data Appropriately**

---

organization
documentation
analysis
dissemination
Summary

Reproducible research practices enable you to:
- Organize experiments productively
- Accurately analyze results
- Share results with future researchers
- Share techniques
- Share reagents with future researchers
- Accelerate science!

The tools discussed here should provide you with the framework to make your research more reproducible and will save you time and resources in the long term.
Contributors

Sonali Roy
Lenny Teytelman
Sarah Robinson
Benjamin Schwessinger
April Clyburne-Sherin
Nicolas Schmelling
Tracey Weissgerber
Tyler Ford
Joanne Kamens
Steven Burgess

What is one thing that you can do today to start making your research more reproducible?

lenny@protocols.io
sroy@noble.org
benjamin.schwessinger@gmail.com