



*Using Open Science Framework (OSF) and
GitHub to Promote Student Training and
Research Transparency*



Who we are

Data Services @ VT:

<https://lib.vt.edu/research-teaching/data-services.html>

DataBridge:

<https://www.databridge.dev/>

BBL:

<https://bevanbrownlab.com>



Dr. Anne M. Brown
ambrown7@vt.edu
@Biochem_Anne



Jonathan Briganti
jonbrig@vt.edu
@data_jonb

Brown Experiential Learning Lab

Bevan & Brown Lab, DataBridge

We are a computational research lab that applies computational, bio/cheminformatics, and data science techniques to data-heavy projects and consults.

- 30+ students
- Freshman → Senior
- Credit & wage options



Brown Experimental Learning (BEL)

A collection of labs, research groups, and dedicated students working to improve student experiences, collaborate on research, and integrate novel technologies into existing workflows.

All students will:

- Submit weekly updates
- Create and present final presentation
- Create or update a final paper

Trained Students

- Undergraduate research projects
- Community collaborations
- Grant-driven projects

Bevan & Brown Lab

Introduction to Bioinformatics

- Introduction to the command line
- Data management
- Data ethics

Molecular Visualization

- Foundations in Chimera and Blender
- Scientific communication
- Data visualization and dashboard synthesis

Molecular Docking Track

- Molecular docking and protein-ligand interactions
- Pharmacophore building
- Lipinski's Rule of Five

Sequence and Structure Modeling

- Homology Modeling
- Sequence Alignments
- Structural Overlays

Molecular Dynamics

- Molecular dynamics basics
- GROMACS tutorial (Lysozyme in water)
- Analysis Techniques

DataBridge

Datatables and Spreadsheets

- Introduction to data logic
- Excel & Google Sheets
- Tidy Data

Data Visualization

- Design and accessibility principles
- Tableau Desktop
- R & R Studio

Data Engineering

- Python
- JavaScript
- AWS Cloud
- Google App Scripts

Code Fundamentals

- Git-Hub
- Introduction to programming
- Open Access Code

Trained Students

- Undergraduate research projects
- Community collaborations
- Grant-driven projects

What steps do students engage in research?

- All of them!
- Data-...
 - Collector
 - Cleaner
 - Analyzer
 - Programmer
 - Visualizer
 - Steward
- Literature, communication

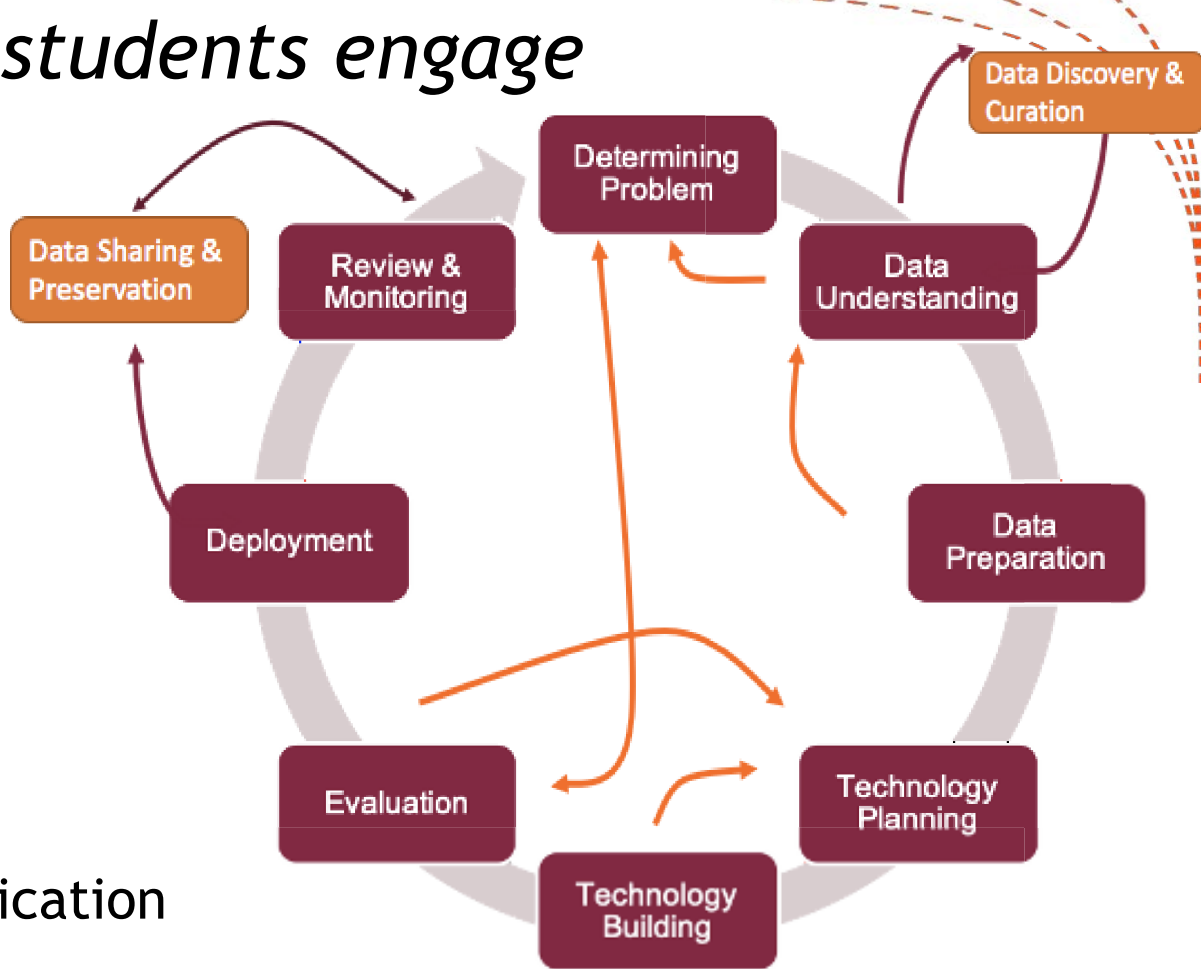


Figure modified from DALM

Preservation, Documentation, & Structure

With remote work, additional emphasis was needed on:

- Maintaining a record of project version history
- Documenting work completed and assigned
- Compartmentalizing work tasks
- Consistent data/code/file storage structure

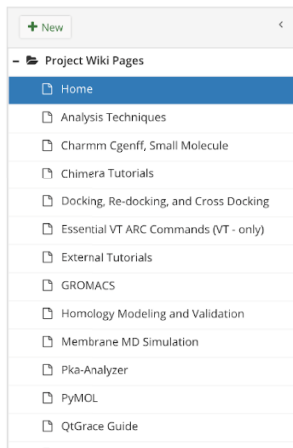
By utilizing industry-standard tools, we could **expose students to these ideas and ensure work was recorded and captured**

Bringing in Open Access + Reproducibility

- Emphasis on creating knowledge and tools that can be used freely
- Encourages students to create reusable, realistic tools, and solutions
- FAIR data practices

The screenshot displays the OSFHOME interface for the 'Bevan & Brown Lab, Public' project. The top navigation bar includes 'OSFHOME', 'My Quick Files', 'My Projects', 'Search', 'Support', 'Donate', and a user profile for 'Anne M. Brown'. The project title is 'Bevan & Brown Lab, Public' with a file size of '175.7MB' and options to 'Make Private', 'Public', 'P', and a menu icon. Below the title, it lists contributors: Anne M. Brown, Jonathan Briganti, Amanda Sharp, Connor Kelly, Carter Gottschalk, and Kelsie Marie King King. It also shows the affiliated institution as Virginia Tech, the fork date (2017-05-24 11:58 AM), and the creation date (2017-05-24 11:58 AM). The description states: 'We are a collaborative molecular modeling and bioinformatics lab at Virginia Tech. We work with individuals all over campus and at other universities to provide computational support to their research, as well as pursuing our own work on amyloid and IDPs. With this OSF, we give you information on our most commonly used techniques and example methods for generating a shared group, multi-project OSF. For information about our current projects or to view our tutorials, head to the Wiki page. To see some of our published papers and images, along with scripts and files to help you with your own molecular modeling, check our Files. Like what we are doing? Feel free to use our templates to make your research more transparent and open. If you have any questions, please contact Dr. Anne M. Brown at ambrown7@vt.edu.' The license is 'CC-BY Attribution 4.0 International'. Below the description is a 'Wiki' section with a 'Read More' link and a 'Citation' section. The wiki content reads: 'Welcome to Bevan & Brown Lab Molecular Modeling and Bioinformatics @ Virginia Tech' followed by a logo for 'MOLECULAR MODELING VIRGINIA TECH'. The logo consists of a stylized hexagonal structure with multiple layers in blue, purple, and pink, next to the text 'MOLECULAR MODELING VIRGINIA TECH'. The 'Citation' section has a dropdown menu and 'Add Component' and 'Link Projects' buttons. The 'Components' section has a text area with the instruction 'Add components to organize your project.' and 'Add Component' and 'Link Projects' buttons. The 'Tags' section shows a list of tags: amyloid, bioinformatics, biomolecular visualization, Computational Biology, data analysis, data visualization, docking, drug design, in silico, Linux, MD simulations, molecular dynamics, PPAR, proteins, protein structure-function, PyMOL.

OSF -- Project Sites & Lab Based



View

Wiki Version: (Current) Carter Gottschalk: 2019-08-15 14:16:55+00:00 UTC

Welcome to Bevan & Brown Lab Molecular Modeling and Bioinformatics @ Virginia Tech



MOLECULAR MODELING VIRGINIA TECH

In this OSF, a series of tutorials is provided to get an idea of the language and programs that we commonly use in our lab. The more you work to learn the "how" and "why" of the programs now, the easier it will be when you utilize them for your own research. Understand what the command does and how you can utilize it; do not just type it in and move forward without understanding. This will greatly benefit you in the future. You must complete all tutorials before moving into your own research; however, this does not mean you cannot begin searching the literature about your project and briefly learning more while doing tutorials. These tutorials are provided under the assumption that you are a member of Bevan & Brown lab and are using a lab computer and/or Virginia Tech resources. HOWEVER, these tutorials are easily adapted to your own environment provided you have downloaded the software and installed some source code as in the instructions provided by them for your local environment. For example, most of these tutorials can be done remotely, on your personal computer, but will require download and installation, which is not always described here. We are not responsible for your ability to download other open-source programs. If these programs will be run using Unix/Linux. With some small tweaking of syntax, you will be able to use ar Dr. Brown (ambrown7@vt.edu) or Jonathan Briganti (jonbrig@vt.edu) if you run into any issues or need assistance.

Want to know even more about our research? Click on the Files button to see our usage files for GROMACS how we do docking on that page, some helpful papers, as well as a demo page for our PPAR project. Like our own research projects and collaborations. Feel free to ask us any questions, we want work towards making Researching!

Citation

APA

Brown, A. M., Briganti, J., Sharp, A., Kelly, C., Gottschalk, C., & King, K. M. K. (2020, September 2). Bevan & Brown Lab, Public. Retrieved from osf.io/82n73

[Customize](#)

MLA

Brown, Anne M et al. "Bevan & Brown Lab, Public." OSF, 2 Sept. 2020. Web.

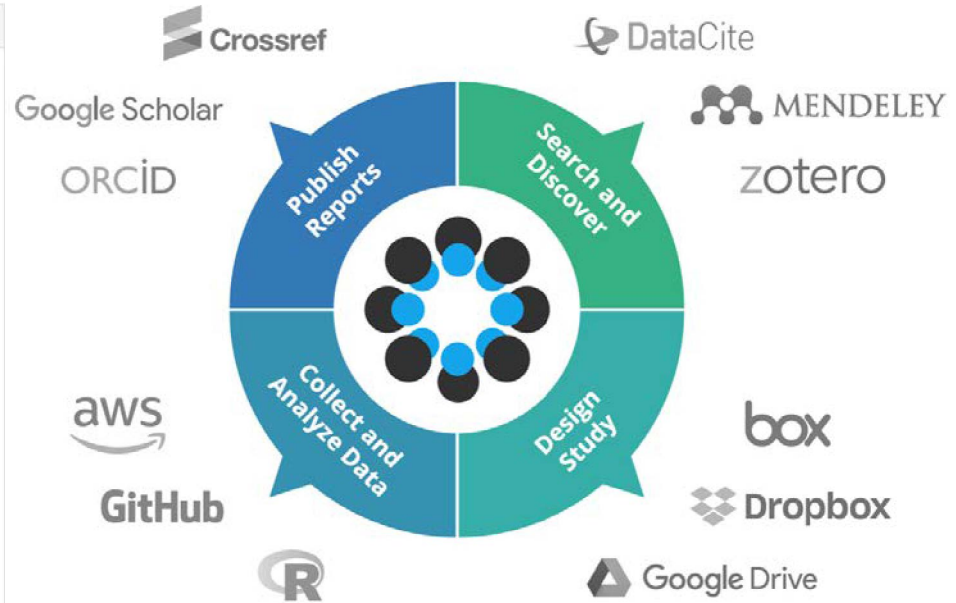
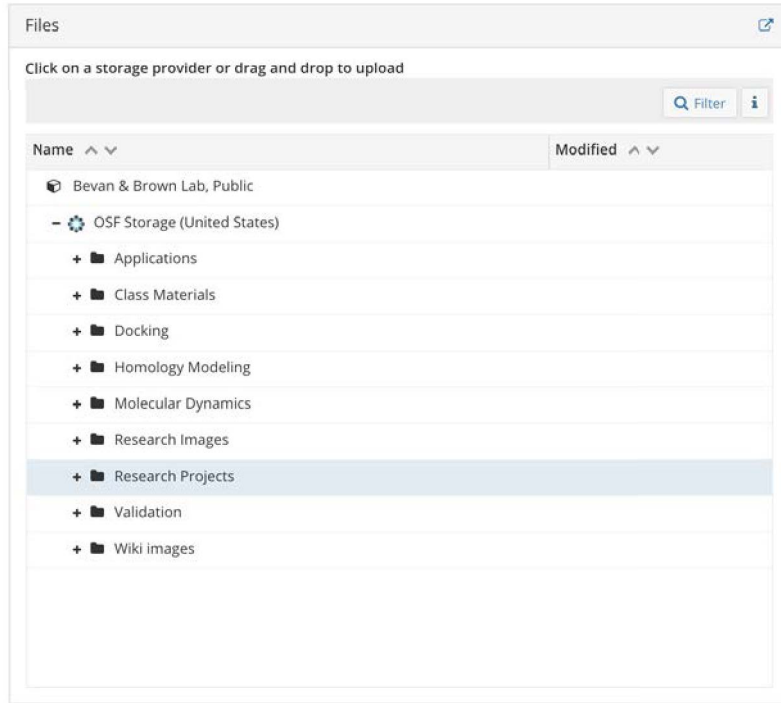
Chicago

Brown, Anne M, Jonathan Briganti, Amanda Sharp, Conor Kelly, Carter Gottschalk, and Kelsie M K King. 2020. "Bevan & Brown Lab, Public." OSF. September 2. osf.io/82n73.

Get more citations

Enter citation style (e.g. "APA")

And it connects with GitHub!



<https://www.cos.io/products/osf>


Github

Code/data repository system for students on programming-heavy projects

Facilitators maintained all ownership, students could submit push requests that were manually reviewed

Often was student's first time conceptualizing a project with multiple active developers

<https://github.com/databridgevt>

 Introductions + kaggle data download Fall 2020 #68
chendaniely opened this issue on Aug 31 · 0 comments

As you are joining the team and getting acquainted with the project (and learning git).

Part 1: Introduce yourself to the repository

Please do the following to add your name to the "Teams" section of the `README.md`.

Follow the non-maintainer steps in: https://chendaniely.github.io/training_ds_r/help-faq.html#general-workflow

You should see your changes when you go to the repository page under the "Teams" section. This task also serves as your understanding of Git and makes sure the settings on this repository are correct. So, please let me know if you run into issues.

If you are having issues

Depending on when people `clone` the repository, when you try to `push` your changes you may be blocked for one of 2 reasons

1. Permissions (403 error): let me know your GH username so I can add you to this repository as a maintainer
2. Something about the remote having changes you don't have: if you keep reading the error message it's essentially telling you that you'd need to `pull` first before `push`ing again.

- You may run into a merge conflict here depending on what lines were changed. Just let me know if you end up with problems here
- The `file` is you need to open the `README.md` file and remove the `>>>>>>`, `=====`, and `<<<<<<` and clean up the entire file so you're happy with it. Then `add`, `commit`, and `push` again.

Part 2: Download the kaggle dataset

Tasks:

1. Make sure you have Python installed (anaconda or miniconda is preferred, otherwise you'll have to manage your own virtual environment)
 - If you haven't already done so, read about python (conda) virtual environments here: <https://daniel.rbind.io/2020/02/29/python-environments-with-conda/>
 - Setup conda forge as the default repository:

```
# run this in your terminal (anaconda command prompt for windows)
conda config --add channels conda-forge
conda config --set channel_priority strict
```

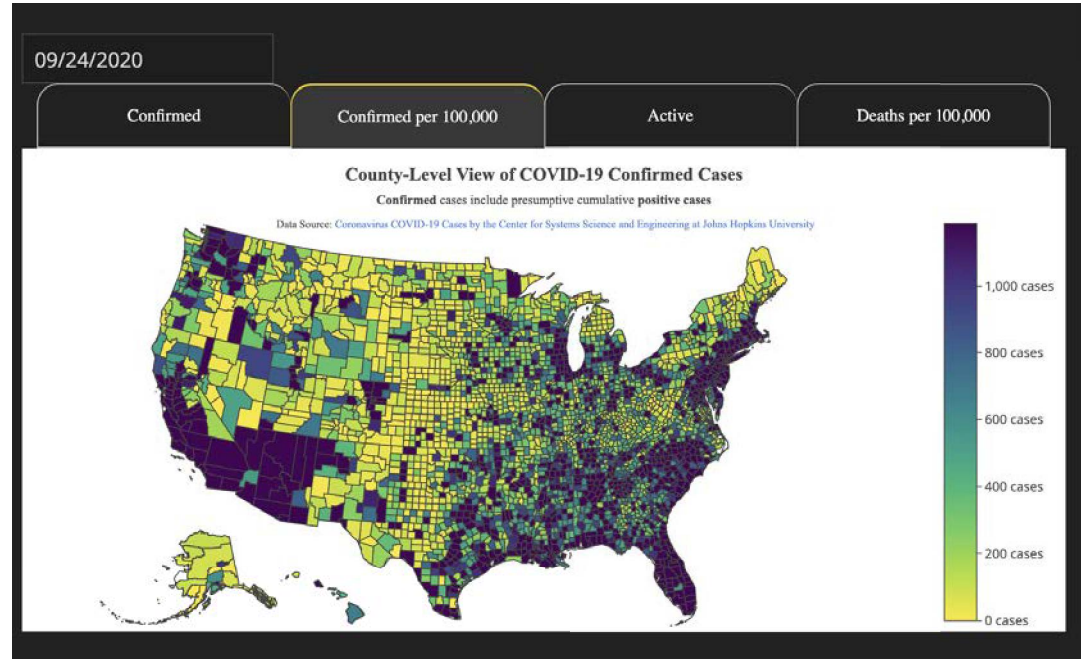
2. `pull` down the new updates from master. What you see on your computer should be what's displayed on GitHub

Example: COVID-19 Dashboard

Live dashboard combining disparate datasets and functions into a single page

Students have:

- Collaborated on code via GitHub
- Used GitHub push & issues to alert group
- Focused on providing “live” data as efficiently as possible



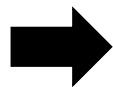
Example: Chicago Elections Project

OFFICE OF THE COMMISSIONER OF THE OFFICE OF GENERAL SERVICES AND GENERAL SERVICES, THE HOUSE OF REPRESENTATIVES, THE SENATE OF MASSACHUSETTS, MASSACHUSETTS HOUSE AND SENATE
LEGISLATIVE SERVICES - BOSTON, MA 02133

SHEET 1

PRESIDENT AND VICE PRESIDENT OF THE UNITED STATES

YEAR	PRESIDENT	VICE PRESIDENT	MASSACHUSETTS SENATOR	MASSACHUSETTS REPRESENTATIVE
1876	Rutherford B. Hayes	William A. Wheeler	William Pitt Fessenden	John D. Phillips
1880	James A. Garfield	Chester A. Arthur	William Pitt Fessenden	John D. Phillips
1884	James G. Blaine	John C. Breckinridge	William Pitt Fessenden	John D. Phillips
1888	Grover Cleveland	Tuckey C. Davis	William Pitt Fessenden	John D. Phillips
1892	Benjamin Harrison	William H. Hunt	William Pitt Fessenden	John D. Phillips
1896	William McKinley	Theodore Roosevelt	William Pitt Fessenden	John D. Phillips
1900	William McKinley	Theodore Roosevelt	William Pitt Fessenden	John D. Phillips
1904	Theodore Roosevelt	Cook	William Pitt Fessenden	John D. Phillips
1908	William Howard Taft	James S. Sherman	William Pitt Fessenden	John D. Phillips
1912	Woodrow Wilson	Thomas R. Marshall	William Pitt Fessenden	John D. Phillips
1916	Woodrow Wilson	Thomas R. Marshall	William Pitt Fessenden	John D. Phillips
1920	Warren G. Harding	Cox	William Pitt Fessenden	John D. Phillips
1924	Calvin Coolidge	Charles McNary	William Pitt Fessenden	John D. Phillips
1928	Herbert Hoover	Charles McNary	William Pitt Fessenden	John D. Phillips
1932	Franklin D. Roosevelt	John N. Garner	William Pitt Fessenden	John D. Phillips
1936	Franklin D. Roosevelt	John N. Garner	William Pitt Fessenden	John D. Phillips
1940	Franklin D. Roosevelt	John N. Garner	William Pitt Fessenden	John D. Phillips
1944	Franklin D. Roosevelt	John A. Sparkman	William Pitt Fessenden	John D. Phillips
1948	Dwight D. Eisenhower	Nathan P. Taylor	William Pitt Fessenden	John D. Phillips
1952	Dwight D. Eisenhower	Nathan P. Taylor	William Pitt Fessenden	John D. Phillips
1956	Dwight D. Eisenhower	Nathan P. Taylor	William Pitt Fessenden	John D. Phillips
1960	John F. Kennedy	Licorne	William Pitt Fessenden	John D. Phillips
1964	Lyndon B. Johnson	Hubert H. Humphrey	William Pitt Fessenden	John D. Phillips
1968	Richard M. Nixon	Hubert H. Humphrey	William Pitt Fessenden	John D. Phillips
1972	Richard M. Nixon	Hubert H. Humphrey	William Pitt Fessenden	John D. Phillips
1976	Gerald R. Ford	Buckley	William Pitt Fessenden	John D. Phillips
1980	Jimmy Carter	Buckley	William Pitt Fessenden	John D. Phillips
1984	Ronald Reagan	Buckley	William Pitt Fessenden	John D. Phillips
1988	George H. W. Bush	Buckley	William Pitt Fessenden	John D. Phillips
1992	Bill Clinton	Al Gore	William Pitt Fessenden	John D. Phillips
1996	Bill Clinton	Al Gore	William Pitt Fessenden	John D. Phillips
2000	George W. Bush	Rice	William Pitt Fessenden	John D. Phillips
2004	George W. Bush	Rice	William Pitt Fessenden	John D. Phillips
2008	Barack Obama	Michelle Obama	William Pitt Fessenden	John D. Phillips
2012	Barack Obama	Michelle Obama	William Pitt Fessenden	John D. Phillips
2016	Donald Trump	Mike Pence	William Pitt Fessenden	John D. Phillips



Chicago Elections Project

HOME Select City

CITIES Chicago

CANDIDATES

GRAPH

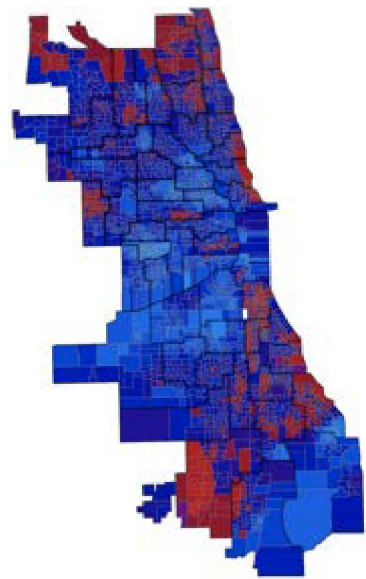
1910

Between 1910 and 1930, the African American population of Chicago increased dramatically, from 44,100 to 233,903. This Great Migration had an immense cultural impact, called the Chicago Black Renaissance, in art, literature, and music. Continuing racial tensions and violence, such as the Chicago Race Riot of 1919, also occurred.



1920

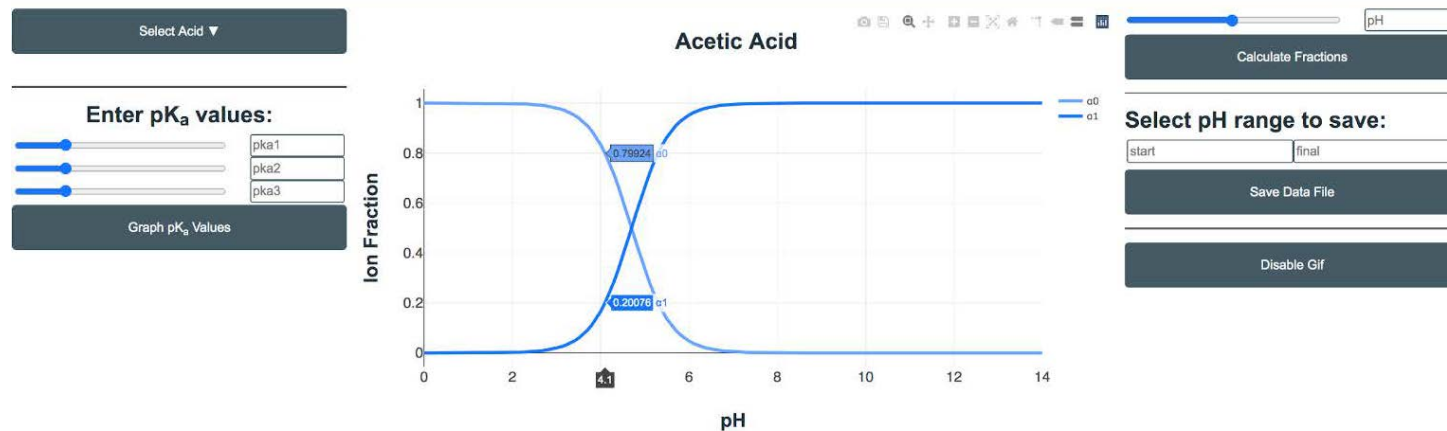
The ratification of the 18th amendment to



<https://github.com/databridgenvt>

Example: pK analyzer

pK_A analyzer 2.1



pH	α_0	α_1	α_2	α_3	pH	α_0	α_1	α_2	α_3
0.0	0.99998	0.00002	-	-	7.0	0.00499	0.99501	-	-
1.0	0.99980	0.00020	-	-	8.0	0.00050	0.99950	-	-
2.0	0.99801	0.00199	-	-	9.0	0.00005	0.99995	-	-
3.0	0.98044	0.01956	-	-	10.0	0.00001	0.99999	-	-
4.0	0.83366	0.16634	-	-	11.0	0.00000	1.00000	-	-
5.0	0.33386	0.66614	-	-	12.0	0.00000	1.00000	-	-

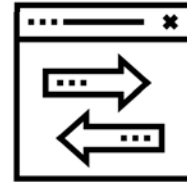
Why does it all matter?



Central Resource



Usefulness of Open!
They can show and
share!



Research Data
reproducibility and
reuse

- I'm not (as frequently) chasing students down for data after they graduate
- UR = hands-on, real world work
 - Train them to be the data stewards of the future

Contact us!



Dr. Anne M. Brown
ambrown7@vt.edu
@Biochem_Anne



Jonathan Briganti
jonbrig@vt.edu
@data_jonb