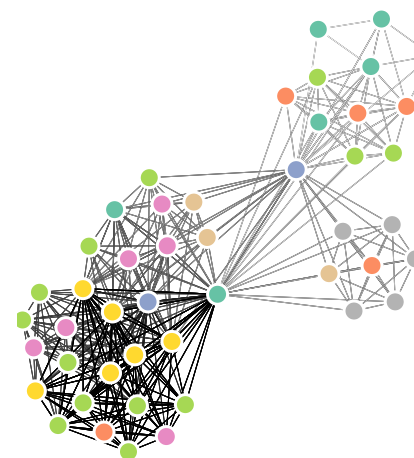


# 劉厚伯

興趣為學習R、Python做資料處理、可視化與創客(arduino、3D print) 平時將興趣學習的東西寫成文章分享在網路上



## EDUCATION

2020  
|  
2017

### 環境工程所-碩士

國立成功大學

📍 台南.東區

- 論文題目: 垃圾焚化飛灰產製鈣矽水合材料及氯鹽穩定化之探討
- ——指導教授:張祖恩

2017  
|  
2013

### 環境工程與科學系-學士

國立屏東科技大學

📍 屏東.內埔

- 專題題目:以奈米碳管/PLS 及石墨烯/PLS 修飾玻璃碳電極測定乙醯胺酚與氧氟沙星
- ——指導教授:黃國林

📄 Download a PDF of this CV

## RESEARCH EXPERIENCE

Current  
|  
2015

### Graduate Research Assistant

TBILab (Yaomin Xu's Lab)

📍 Vanderbilt University

- Primarily working with large EHR and Biobank datasets.
- Developing network-based methods to investigate and visualize clinically relevant patterns in data.

2018  
|  
2017

### Data Science Researcher

Data Science Lab

📍 Johns Hopkins University

- Building R Shiny applications in the contexts of wearables and statistics education.
- Work primarily done in R Shiny and Javascript (node and d3js).

2015  
|  
2013

### Undergraduate Researcher

Rubenstein Ecosystems Science Laboratory 📍 University of Vermont

- Analyzed and visualized data for CATOS fish tracking project.
- Head of data mining project to establish temporal trends in population densities of *Mysis diluviana* (Mysis).
- Ran project to mathematically model the migration patterns of Mysis (honors thesis project.)

## CONTACT

✉ [gtgrthrst4577@gmail.com](mailto:gtgrthrst4577@gmail.com)



<https://www.facebook.com/LHB02>

🔗 <https://hackmd.io/@LHB-0222/AllWritings>

## LANGUAGE SKILLS

Excel

R

Arduino

Python, MicroPython

SQL

Made with the R package  
[pagedown](#).

The source code is available on  
[github.com/nstrayer/cv](https://github.com/nstrayer/cv).

Last updated on 2021-02-21.

2015  
|  
2015

### Human Computer Interaction Researcher

LabInTheWild (Reineke Lab)

📍 University of Michigan

- Led development and implementation of interactive data visualizations to help users compare themselves to other demographics.

2014  
|  
2013

### Undergraduate Researcher

Bentil Laboratory

📍 University of Vermont

- Developed mathematical model to predict the transport of sulfur through the environment with applications in waste cleanup.

2013  
|  
2012

### Research Assistant

Adair Laboratory

📍 University of Vermont

- Independently analyzed and constructed statistical models for large data sets pertaining to carbon decomposition rates.



## INDUSTRY EXPERIENCE

Current  
|  
2020

### Software Engineer

RStudio

📍 Remote

- Helping make programming web applications with R easier and more beautiful on the Shiny team

2016  
|  
2016

### Data Journalist - Graphics Department

New York Times

📍 New York, New York

- Reporter with the graphics desk covering topics in science, politics, and sport.
- Work primarily done in R, Javascript, and Adobe Illustrator.

2015  
|  
2015

### Engineering Intern - User Experience

Dealer.com

📍 Burlington, VT

- Built internal tool to help analyze and visualize user interaction with back-end products.

2015  
|  
2015

### Data Science Intern

Dealer.com

📍 Burlington, VT

- Worked with the product analytics team to help parse and visualize large stores of data to drive business decisions.

2015  
|  
2014

### Data Artist In Residence

Conduce

📍 Carpinteria, CA

- Envisioned, prototyped and implemented visualization framework in the course of one month.
- Constructed training protocol for bringing third parties up to speed with new protocol.

I have worked in a variety of roles ranging from journalist to software engineer to data scientist. I like collaborative environments where I can learn from my peers.

2014  
|  
2014

## Software Engineering Intern

Conduce

📍 Carpinteria, CA

- Incorporated d3.js to the company's main software platform.



## TEACHING EXPERIENCE

2020

### Javascript for Shiny Users

RStudio::conf 2020

- Served as TA for two day workshop on how to leverage Javascript in Shiny applications
- Lectured on [using R2D3 package to build interactive visualizations](#).

2019  
|  
2019

### Data Visualization Best Practices

DataCamp

- Designed from bottom up course to teach best practices for scientific visualizations.
- Uses R and ggplot2.
- In top 10% on platform by popularity.

2019  
|  
2019

### Improving your visualization in Python

DataCamp

- Designed from bottom up course to teach advanced methods for enhancing visualization.
- Uses python, matplotlib, and seaborn.

2018  
|  
2017

### Advanced Statistical Learning and Inference

Vanderbilt Biostatistics Department

📍 Nashville, TN

- TA and lectured
- Topics covered from penalized regression to boosted trees and neural networks
- Highest level course offered in department

2018  
|  
2018

### Advanced Statistical Computing

Vanderbilt Biostatistics Department

📍 Nashville, TN

- TA and lectured
- Covered modern statistical computing algorithms
- 4th year PhD level class

I am passionate about education. I believe that no topic is too complex if the teacher is empathetic and willing to think about new methods of approaching task.

2017  
|  
2017

## Statistical Computing in R

Vanderbilt Biostatistics Department

📍 Nashville, TN

- TA and lectured
- Covered introduction to R language for statistics applications
- Graduate level class



## SELECTED DATA SCIENCE WRITING

2019

### [Using AWK and R to Parse 25th](#)

LiveFreeOrDichotomize.com

- Story of parsing large amounts of genomics data.
- Provided advice for dealing with data much larger than disk.
- Reached top of HackerNews.

I regularly blog about data science and visualization on my blog [LiveFreeOrDichotomize](#).

2018

### [Classifying physical activity from smartphone data](#)

RStudio Tensorflow Blog

- Walk through of training a convolutional neural network to achieve state of the art recognition of activities from accelerometer data.
- Contracted article.

2018

### [The United States of Seasons](#)

LiveFreeOrDichotomize.com

- GIS analysis of weather data to find the most 'seasonal' locations in United States
- Used Bayesian regression methods for smoothing sparse geospatial data.

2017

### [A year as told by fitbit](#)

LiveFreeOrDichotomize.com

- Analyzing a full years worth of second-level heart rate data from wearable device.
- Demonstrated visualization-based inference for large data.

2017

### [MCMC and the case of the spilled seeds](#)

LiveFreeOrDichotomize.com

- Full Bayesian MCMC sampler running in your browser.
- Coded from scratch in vanilla Javascript.

2017

### [The Traveling Metallurgist](#)

LiveFreeOrDichotomize.com

- Pure javascript implementation of traveling salesman solution using simulated annealing.
- Allows reader to customize the number and location of cities to attempt to trick the algorithm.



## SELECTED PRESS (ABOUT)

2017  
|  
2017

### [Great paper? Swipe right on the new ‘Tinder for preprints’ app](#)

Science

- Story of the app [Paprr](#) made with Jeff Leek and Lucy D’Agostino McGowan.

2017  
|  
2017

### [Swipe right for science: Papr app is ‘Tinder for preprints’](#)

Nature News

- Second press article for app Papr.

2016  
|  
2016

### [The Deeper Story in the Data](#)

University of Vermont Quarterly

- Story on my path post graduation and the power of narrative.



## SELECTED PRESS (BY)

2016  
|  
2016

### [The Great Student Migration](#)

The New York Times

- Most shared and discussed article from the New York Times for August 2016.

2016  
|  
2016

### [Wildfires are Getting Worse, The New York Times](#)

The New York Times

- GIS analysis and modeling of fire patterns and trends
- Data in collaboration with NASA and USGS

2016  
|  
2016

### [Who’s Speaking at the Democratic National Convention?](#)

The New York Times

- Data scraped from CSPAN records to figure out who talked and past conventions.

2016  
|  
2016

### [Who’s Speaking at the Republican National Convention?](#)

The New York Times

- Used same data scraping techniques as Who’s Speaking at the Democratic National Convention?

2016  
|  
2016

### [A Trail of Terror in Nice, Block by Block](#)

The New York Times

- Led research effort to put together story of 2016 terrorist attack in Nice, France in less than 12 hours.
- Work won Silver medal at Malofiej 2017, and gold at Society of News and Design.



## SELECTED PUBLICATIONS, POSTERS, AND TALKS

2020

### **Building a software package in tandem with machine learning methods research can result in both more rigorous code and more rigorous research**

ENAR 2020

- Invited talk in Human Data Interaction section.
- How and why building an R package can benefit methodological research

2020

### **Stochastic Block Modeling in R, Statistically rigorous clustering with rigorous code**

RStudio::conf 2020

- Invited talk about new [sbmR package](#).
- Focus on how software development and methodological research can improve both benefit when done in tandem.

2020

### **PheWAS-ME: A web-app for interactive exploration of multimorbidity patterns in PheWAS**

Bioinformatics

- Manuscript detailing application for the exploration of multimorbidity patterns in PheWAS analyses
- See [landing page](#) for more information.

2019

|  
2019

### **Charge Reductions Associated with Shortening Time to Recovery in Septic Shock**

Chest

- Authored with Wesley H. Self, MD MPH; Dandan Liu, PhD; Stephan Russ, MD, MPH; Michael J. Ward, MD, PhD, MBA; Nathan I. Shapiro, MD, MPH; Todd W. Rice, MD, MSc; Matthew W. Semler, MD, MSc.

2019

|  
2019

### **Multimorbidity Explorer | A shiny app for exploring EHR and biobank data**

RStudio::conf 2019

- Contributed Poster. Authored with Yaomin Xu.

2019

|  
2019

### **Taking a network view of EHR and Biobank data to find explainable multivariate patterns**

Vanderbilt Biostatistics Seminar Series

- University wide seminar series.

2019

### **Patient-specific risk factors independently influence survival in Myelodysplastic Syndromes in an unbiased review of EHR records**

Under-Review (copy available upon request.)

- Bayesian network analysis used to find novel subgroups of patients with Myelodysplastic Syndromes (MDS).
- Analysis done using method built for my dissertation.

2019	<b>Patient specific comorbidities impact overall survival in myelofibrosis</b> Under-Review (copy available upon request.) <ul style="list-style-type: none"> <li>• Bayesian network analysis used to find robust novel subgroups of patients with given genetic mutations.</li> <li>• Analysis done using method built for my dissertation.</li> </ul>
2018   2018	<b>R timelineViz: Visualizing the distribution of study events in longitudinal studies</b> Under-Review (copy available upon request.) <ul style="list-style-type: none"> <li>• Authored with Alex Sunderman of the Vanderbilt Department of Epidemiology.</li> </ul>
2017   2017	<b>Continuous Classification using Deep Neural Networks</b> Vanderbilt Biostatistics Qualification Exam <ul style="list-style-type: none"> <li>• Review of methods for classifying continuous data streams using neural networks</li> <li>• Successfully met qualifying examination standards</li> </ul>
2015   2015	<b>Asymmetric Linkage Disequilibrium: Tools for Dissecting Multiallelic LD</b> Journal of Human Immunology <ul style="list-style-type: none"> <li>• Authored with Richard Single, Vanja Paunic, Mark Albrecht, and Martin Maiers.</li> </ul>
2015   2015	<b>An Agent Based Model of Mysis Migration</b> International Association of Great Lakes Research Conference <ul style="list-style-type: none"> <li>• Authored with Brian O'Malley, Sture Hansson, and Jason Stockwell.</li> </ul>
2015   2015	<b>Declines of Mysis diluviana in the Great Lakes</b> Journal of Great Lakes Research <ul style="list-style-type: none"> <li>• Authored with Peter Euclide and Jason Stockwell.</li> </ul>