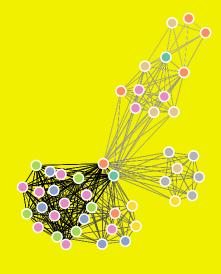
# NICK STRA<mark>YER</mark>

I have made visualizations viewed by hundreds of thousands of people<sup>7</sup>, sped up query times for 25 terabytes of data by an average of 4,800 times<sup>2</sup>, and built packages for R<sup>3</sup> that let yo<mark>u do magic<sup>4</sup>.</mark>

Currently searching for a position that allows me to build tools leveraging a combination of visualization, machine learning, and software engineering to help people explore and understand their data in new and useful ways.



# **EDUCATION**

Current | 2015

#### PhD. Candidate, Biostatistics

Vanderbilt University

Nashville, TN

- Working on Bayes<mark>ian network models & interactive visualization platforms</mark>
- University Graduate Fellow

View this CV online with links at nickstrayer.me/cv/

2015 | 2011

#### B.S., Mathematics, Statistics (minor C.S.)

University of Vermont

Paurlington, VT

• Thesis: An agent based model of Diel Vertical Migration patterns of Mysis diluviana

# 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 Res<mark>earcher</mark>

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

- nick.strayer@gmail.com
- **y** NicholasStrayer
- github.com/nstrayer
- Ø nickstrayer.me
- in linkedin.com/in/nickstrayer

## LANGUAGE SKILLS

R	
Javascript (d3.js)	
C++	
Python	
Bash	
SQL	
AWK	

Made with the R package pagedown.

The source code is available on github.com/nstrayer/cv.

Last updated on 2021-11-17.

Human Computer Interaction Researcher 2015 University of Michigan LabInTheWild (Reineke Lab) 2015 · Led development and implementation of interactive data visualizations to help users compare themselves to other demographics. Undergraduate Researcher 2014 University of Vermont Bentil Laboratory 2013 · Developed mathematical model to predict the transport of sulfur through the environment with applications in waste cleanup. Research Assistant 2013 University of Vermont Adair Laboratory 2012 · Independently analyzed and constructed statistical models for large data sets pertaining to carbon decomposition rates. INDUSTRY EXPERIENCE 2016 Data Journalist - Graphics Department • New York, New York **New York Times** 2016 · Reporter with the graphics desk covering topics in science, politics, and · Work primarily done in R, Javascript, and Adobe Illustrator. Engineering Intern - User Experience 2015 Burlington, VT Dealer.com 2015 · Built internal tool to help analyze and visualize user interaction with back-end products. Data Science Intern 2015 Burlington, VT Dealer.com 2015 · Worked with the product analytics team to help parse and visualize large stores of data to drive business decisions. Data Artist In Resid<mark>ence</mark> 2015 O Carpinteria, CA Conduce 2014 • 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. Software Engineering Intern 2014 O Carpinteria, CA Conduce 2014 · Incorporated d3.js to the company's main software platform.

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.

## ♣ 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.<sup>5</sup>

#### 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.

### • Improving your visualization in Python

DataCamp

2019

2019

2018

2017

2018

2017

2017

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

#### 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 • Advanced Statistical Computing

Vanderbilt Biostatistics Department

Nashville, TN

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

#### Statistical Computing in R

Vanderbilt Biostatistics Department

Nashville, TN

- · TA and lectured
- · Covered introduct<mark>ion to R language for statistics applications</mark>
- · Graduate level class

## SELECTED DATA SCIENCE WRITING

#### 2019 • Using AWK and R to Parse 25tb<sup>7</sup>

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 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.

I regularly blog about data science and visualization on my blog LiveFreeOrDichotomize.<sup>6</sup>

### 2018 • Classifying physical activity from smartphone data<sup>8</sup>

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<sup>9</sup>

LiveFreeOrDichotomize.com

- GIS analysis of weather data to find the most 'seasonal' locations in United States
- · Used Bayesian reg<mark>ression methods for smoothing sparse geospatial data.</mark>

#### 2017 • A year as told by fitbit<sup>10</sup>

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<sup>11</sup>

LiveFreeOrDichotomize.com

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

#### 2017 • The Traveling Metallurgist<sup>12</sup>

Science

2017

2017

2017

2016

2016

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 • Great paper? Swipe right on the new 'Tinder for preprints' app's

• Story of the app Papr<sup>14</sup> made with Jeff Leek and Lucy D'Agostino McGowan.

• Swipe right for science: Papr app is 'Tinder for preprints's'
Nature News

· Second press artic<mark>le for app Papr.</mark>

The Deeper Story in the Data<sup>16</sup>
 University of Vermont Quarterly

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

# ■ SELECTED PRESS (BY)

2016 2016 The Great Student Migration 17

The New York Times

· Most shared and discussed article from the New York Times for August

2016 2016 Wildfires are Getting Worse, The New York Times<sup>18</sup>

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?<sup>19</sup>

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?<sup>20</sup>

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<sup>21</sup>

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<sup>22</sup>

RStudio::conf 2020

- Invited talk about new sbmR package<sup>23</sup>.
- · Focus on how software development and methodological research can improve both benefit when done in tandem.

Charge Reductions Associated with Shortening Time to Recovery in 2019 Septic Shock<sup>24</sup> 2019 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. Multimorbidity Explorer | A shiny app for exploring EHR and biobank 2019 data<sup>25</sup> 2019 RStudio::conf 2019 · Contributed Poster, Authored with Yaomin Xu. Taking a network view of EHR and Biobank data to find explainable 2019 multivariate patterns<sup>26</sup> 2019 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. Patient specific comorbidities impact overall survival in myelofibrosis 2019 Under-Review (copy available upon request.) · Bayesian network analysis used to find robust novel subgroups of patients with given genetic mutations. · Analysis done using method built for my dissertation. R timelineViz: Visualizing the distribution of study events in longitudinal 2018 studies 2018 Under-Review (copy available upon request.) · Authored with Alex Sunderman of the Vanderbilt Department of Epidemiology. Continuous Classification using Deep Neural Networks<sup>27</sup> 2017 Vanderbilt Biostatistics Qualification Exam 2017 · Review of methods for classifying continuous data streams using neural networks · Successfully met qualifying examination standards Asymmetric Linkage Disequilibrium: Tools for Dissecting Multiallelic LD 2015 Journal of Human Immunology 2015

· Authored with Richard Single, Vanja Paunic, Mark Albrecht, and Martin

Maiers.

An Agent Based Model of Mysis Migration<sup>28</sup>
International Association of Great Lakes Research Conference
• Authored with Brian O'Malley, Sture Hansson, and Jason Stockwell.

Declines of Mysis diluviana in the Great Lakes

Journal of Great Lakes Research
• Authored with Peter Euclide and Jason Stockwell.



- 1: https://www.nytimes.com/interactive/2016/08/26/us/college-student-migration.html
- 2. https://livefreeordichotomize.com/2019/06/04/using\_awk\_and\_r\_to\_parse\_25tb/
- 3. https://github.com/nstrayer/shinysense
- 4: http://nickstrayer.me/dataDayTexas/
- 5: http://nickstrayer.me/js4shiny\_r2d3/slides
- 6. https://livefreeordichotomize.com/
- 7: https://livefreeordichotomize.com/2019/06/04/using\_awk\_and\_r\_to\_parse\_25tb/
- 8: https://blogs.rstudio.com/tensorflow/posts/2018-07-17-activity-detection/
- 9: https://livefreeordichotomize.com/2018/02/12/the-united-states-of-seasons/
- 10: https://livefreeordichotomize.com/2017/12/27/a-year-as-told-by-fitbit/
- 11: https://livefreeordichotomize.com/2017/10/14/mcmc-and-the-case-of-the-spilled
- 12: https://livefreeordichotomize.com/2017/09/25/the-traveling-metallurgist/
- 13: https://www.sciencemag.org/news/2017/06/great-paper-swipe-right-new-tinder-preprints-app
- 14: https://jhubiostatistics.shinyapps.io/papr/
- 15: https://www.nature.com/news/swipe-right-for-science-papr-app-is-tinder-for-preprints-1.22163
- 16: https://www.uvm.edu/uvmnews/news/deeper-story-data
- 17: https://www.nytimes.com/interactive/2016/08/26/us/college-student-migration.html?smid=pl-share
- 18. https://www.nytimes.com/interactive/2016/07/25/us/wildfire-seasons-los-angeles
- 19. https://www.nytimes.com/2<mark>016/07/26/upshot/democrats-may-not-be-unified-but-their-convention-speakers-are.html</mark>
- 20: https://www.nytimes.com/2<mark>016/07/19/upshot/whos-not-speaking-how-this-republican-convention-differs.html?smid=pl-share</mark>
- 21. https://www.nytimes.com/interactive/2016/07/14/world/europe/trail-of-terror-france.html
- 22. http://nickstrayer.me/rstudioconf\_sbm
- 23. https://tbilab.github.io/sbmR/
- 24: https://www.ncbi.nlm.nih.gov/pubmed/30419234
- 25: http://nickstrayer.me/rstudioconf19\_me-poster/
- 26. http://nickstrayer.me/biostat\_seminar/
- 27. http://nickstrayer.me/qualifying\_exam/
- 28: https://www.semanticscholar.org/paper/An-Agent-Based-Model-of-the-Diel -Vertical-Migration-Strayer-Stockwell /40493c78e8ecf22bd882d17ec99fd913ec4b9820