

DUNCAN GATES

As a soon-to-be graduate of Oregon State University with focuses in a variety of fields I have found my interests primarily coalesce around translating data into insight, through the application of statistics, coding, and deep understanding of individual issues. After my internship experience this summer working for Data Science for the Public Good¹ I have enhanced my focus on team based projects that will have significant positive impacts on the lives of others with the skills I have developed over the course of college.

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

2020
|
2017


- **B.S. in Economics w/ minors in Actuarial Science and Political Science**
Corvallis, OR  Oregon State University
 - Thesis: The Relationship Between Income Taxation, Social Benefits, and Wealth Inequality in OECD Countries: Exploring the Factors Contributing to Unequal Household Wealth
 - Honors College Student
 - GPA: 3.82

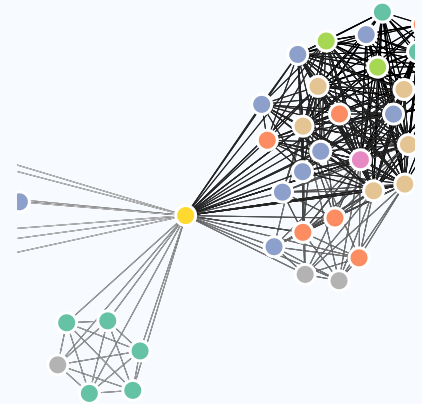
RESEARCH EXPERIENCE

2020
|
2020

- **Data Science Intern**
Data Science for the Public Good  Oregon State University
 - Internship was split into training and research. The first half focused on getting interns up to speed on statistics, machine learning, visualization skills, and teamwork based coding. The second half was based on projects that were assigned to teams composed of two undergraduates, one graduate student, and one mentor (generally a professor).
 - Work was done to measure the impact of the installation of a Selective Water Withdrawal Tower at the Pelton Dam on the Deschutes River. This investigation primarily was based on modelling river temperatures, but also allowed for learning, the team built an R Shiny application, presented at scientific symposiums, and worked collaboratively with officials from PGE, ODFW, ODEQ, and cities on the Deschutes.






2020
|
2019

- **Undergraduate Research Assistant**
OSU Economics Research  Oregon State University
 - Conducting economic research under the guidance of numerous professors
 - Field research to gather panel and time series data

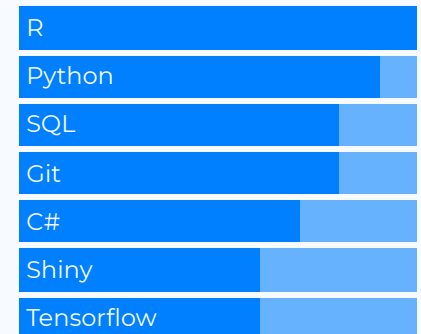


View this CV online with links at duncangates.me/cv/

CONTACT

 gatesdu@oregonstate.edu
 [gates_duncan](https://twitter.com/gates_duncan)
 github.com/itsthesnake
 duncangates.me
 linkedin.com/in/duncan-gates

LANGUAGE SKILLS




Made with the R package [page-down](https://github.com/josiahpage/page-down).


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

Last updated on 2020-09-09.


- 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.)
- 2015
|
2015

Undergraduate 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


- 2019
|
2015

Rental Location Manager
 Portland, Oregon  Alder Creek Kayak & Canoe


 - Spent summers working as a rental location manager for a kayak company.
- 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.²

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

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

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.



SELECTED DATA SCIENCE WRITING

- 2019 ● **Using AWK and R to Parse 25tb⁴**
LiveFreeOrDichotomize.com
- Story of parsing large amounts of genomics data.
 - Provided advice for dealing with data much larger than disk.
 - Reached top of HackerNews.
- 2018 ● **Classifying physical activity from smartphone data⁵**
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 Seasons⁶**
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 fitbit⁷**
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 seeds⁸**
LiveFreeOrDichotomize.com
- Full Bayesian MCMC sampler running in your browser.
 - Coded from scratch in vanilla Javascript.
- 2017 ● **The Traveling Metallurgist⁹**
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.

I regularly blog about data science and visualization on my blog LiveFreeOrDichotomize.³



SELECTED PRESS (ABOUT)

- 2017
|
2017 ● **Great paper? Swipe right on the new "Tinder for preprints" app¹⁰**
Science
- Story of the app Papr¹¹ made with Jeff Leek and Lucy D'Agostino McGowan.

2017 | 2017 ● **Swipe right for science: Papr app is “Tinder for preprints”¹²**

Nature News

- Second press article for app Papr.

2016 | 2016 ● **The Deeper Story in the Data¹³**

University of Vermont Quarterly

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



SELECTED PRESS (BY)

2016 | 2016 ● **Wildfires are Getting Worse, The New York Times¹⁴**

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?¹⁵**

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?¹⁶**

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 Block¹⁷**

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.

2016 | 2016 ● **The Great Student Migration¹⁸**

The New York Times

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



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
- **PheWAS-ME: A web-app for interactive exploration of multimorbidity patterns in PheWAS¹⁹**
MedRXiv
 - Manuscript detailing application for the exploration of multimorbidity patterns in PheWAS analyses
 - See landing page²⁰ for more information.
- 2020
- **Stochastic Block Modeling in R, Statistically rigorous clustering with rigorous code²¹**
RStudio::conf 2020
 - Invited talk about new sbmR package²².
 - Focus on how software development and methodological research can improve both benefit when done in tandem.
- 2019
|
2019
- **Taking a network view of EHR and Biobank data to find explainable multivariate patterns²³**
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
- **Patient specific comorbidities impact overall survival in myelofibrosis**
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.
- 2019
|
2019
- **Charge Reductions Associated with Shortening Time to Recovery in Septic Shock²⁴**
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 data²⁵**
RStudio::conf 2019
 - Contributed Poster. Authored with Yaomin Xu.
- 2018
|
2018
- **R timelineViz: Visualizing the distribution of study events in longitudinal studies**
Under-Review (copy available upon request.)
 - Authored with Alex Sunderman of the Vanderbilt Department of Epidemiology.
- 2017
|
2017
- **Continuous Classification using Deep Neural Networks²⁶**
Vanderbilt Biostatistics Qualification Exam
 - Review of methods for classifying continuous data streams using neural networks
 - Successfully met qualifying examination standards
- 2015
|
2015
- **An Agent Based Model of Mysis Migration²⁷**
International Association of Great Lakes Research Conference
 - Authored with Brian O'Malley, Sture Hansson, and Jason Stockwell.
- 2015
|
2015
- **Declines of Mysis diluviana in the Great Lakes**
Journal of Great Lakes Research
 - Authored with Peter Euclide and Jason Stockwell.
- 2015
|
2015
- **Asymmetric Linkage Disequilibrium: Tools for Dissecting Multiallelic LD**
Journal of Human Immunology
 - Authored with Richard Single, Vanja Paunic, Mark Albrecht, and Martin Maier.

LINKS

- 1: <https://workspace.oregonstate.edu/data-science-for-the-public-good>
- 2: http://nickstrayer.me/js4shiny_r2d3/slides
- 3: <https://livefreeordichotomize.com/>
- 4: https://livefreeordichotomize.com/2019/06/04/using_awk_and_r_to_parse_25tb/
- 5: <https://blogs.rstudio.com/tensorflow/posts/2018-07-17-activity-detection/>
- 6: <https://livefreeordichotomize.com/2018/02/12/the-united-states-of-seasons/>
- 7: <https://livefreeordichotomize.com/2017/12/27/a-year-as-told-by-fitbit/>
- 8: <https://livefreeordichotomize.com/2017/10/14/mcmc-and-the-case-of-the-spilled-seeds/>
- 9: <https://livefreeordichotomize.com/2017/09/25/the-traveling-metallurgist/>
- 10: <https://www.sciencemag.org/news/2017/06/great-paper-swipe-right-new-tinder-preprints-app>
- 11: <https://jhubiostatistics.shinyapps.io/papr/>
- 12: <https://www.nature.com/news/swipe-right-for-science-papr-app-is-tinder-for-pre-prints-1.22163>

- 13: <https://www.uvm.edu/uvmnews/news/deeper-story-data>
- 14: <https://www.nytimes.com/interactive/2016/07/25/us/wildfire-seasons-los-angeles.html>
- 15: <https://www.nytimes.com/2016/07/26/upshot/democrats-may-not-be-unified-but-their-convention-speakers-are.html>
- 16: <https://www.nytimes.com/2016/07/19/upshot/whos-not-speaking-how-this-republican-convention-differs.html?smid=pl-share>
- 17: <https://www.nytimes.com/interactive/2016/07/14/world/europe/trail-of-terror-france.html>
- 18: <https://www.nytimes.com/interactive/2016/08/26/us/college-student-migration.html?smid=pl-share>
- 19: <https://www.medrxiv.org/content/10.1101/19009480v4>
- 20: https://prod.tbilab.org/phewas_me_info/
- 21: http://nickstrayer.me/rstudioconf_sbm
- 22: <https://tbilab.github.io/sbmR/>
- 23: http://nickstrayer.me/biostat_seminar/
- 24: <https://www.ncbi.nlm.nih.gov/pubmed/30419234>
- 25: http://nickstrayer.me/rstudioconf19_me-poster/
- 26: http://nickstrayer.me/qualifying_exam/
- 27: <https://www.semanticscholar.org/paper/An-Agent-Based-Model-of-the-Diel-Vertical-Migration-Strayer-Stockwell/40493c78e8ecf22bd882d17ec99fd913ec4b9820>