

# MARCELO RODRIGUES RIBEIRO

Experiência em manipulação de dados e relatórios gerenciais utilizando Software Estatístico R.



## EDUCATION

2021  
|  
2019



### Pós-graduação 'lato sensu' – Especialização em Ciência de Dados e Big Data

Belo Horizonte, MG, Brasil

📍 Pontifícia Universidade Católica de Minas Gerais (PUC MINAS)

2011  
|  
2006



### Bacharelado em Estatística

Belo Horizonte, MG, Brasil

📍 Universidade Federal de Minas Gerais (UFMG)



## RESEARCH EXPERIENCE

2021  
|  
2019



### Analista do Seguro Social – Estatístico

Brasília, DF, Brasil

📍 Ministério da Economia (ME)

- Coordenação de Inteligência Previdenciária e trabalhista, precisamente na Coordenação de Informações Estratégicas – CIEST
- Análise de dados para aprimorar a capacidade de prevenção, detecção e análise de grandes ameaças representadas pelos esquemas e organizações criminosas que atuam em detrimento da Previdência e do Trabalho.

2019  
|  
2014



### Analista do Seguro Social – Estatístico

Brasília, DF, Brasil

📍 Instituto Nacional do Seguro Social (INSS)

- Coordenação-Geral de Recursos Logísticos

2014  
|  
2010



### Analista Criminal

Belo Horizonte, MG, Brasil

📍 Polícia Militar de Minas Gerais (PMMG)

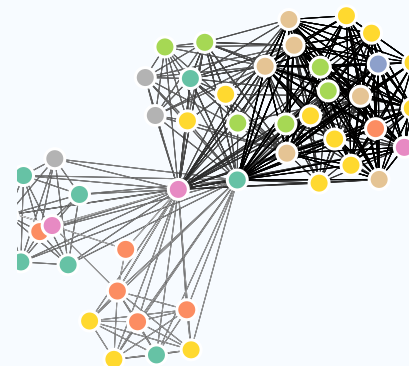
- Desenvolvimento e disseminação do conhecimento sobre a gestão da informação e conhecimentos em estatística aplicada à análise de Sistema Criminal.



## INDUSTRY EXPERIENCE



## TEACHING EXPERIENCE



View this CV online with links at [nickstrayer.me/cv/](https://nickstrayer.me/cv/)

## CONTACT



[mrodriguesribeiro@gmail.com](mailto:mrodriguesribeiro@gmail.com)

[marcelo.ribeiro@economia.gov.br](mailto:marcelo.ribeiro@economia.gov.br)



@mrodriguesrib



<https://github.com/MarceloRibeiro>

-celo

## LANGUAGE SKILLS

R

Python

Made with the R package

[pagedown](#)

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.

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 tasks.

- 2020  
|  
2019

●

**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>7</sup>
- 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



## SELECTED DATA SCIENCE WRITING

- 2019

●

**Using AWK and R to Parse 25tb<sup>3</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 regularly blog about data science and visualization on my blog [LiveFreeOrDichotomize.com](http://LiveFreeOrDichotomize.com).<sup>2</sup>

- 2018 ● **Classifying physical activity from smartphone data<sup>4</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>5</sup>**  
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<sup>6</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>7</sup>**  
LiveFreeOrDichotomize.com
  - Full Bayesian MCMC sampler running in your browser.
  - Coded from scratch in vanilla Javascript.
- 2017 ● **The Traveling Metallurgist<sup>8</sup>**  
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<sup>9</sup>**  
Science
  - Story of the app Papr<sup>10</sup> made with Jeff Leek and Lucy D'Agostino McGowan.
- 2017  
|  
2017 ● **Swipe right for science: Papr app is "Tinder for preprints"<sup>11</sup>**  
Nature News
  - Second press article for app Papr.
- 2016  
|  
2016 ● **The Deeper Story in the Data<sup>12</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<sup>13</sup>**

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<sup>44</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>15</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>16</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>17</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>18</sup>**

RStudio::conf 2020

- Invited talk about new sbmR package<sup>19</sup>.
- 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<sup>20</sup>**  
Bioinformatics
- Manuscript detailing application for the exploration of multimorbidity patterns in PheWAS analyses
  - See landing page<sup>27</sup> for more information.
- 2019 | 2019 ● **Charge Reductions Associated with Shortening Time to Recovery in Septic Shock<sup>22</sup>**  
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<sup>23</sup>**  
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<sup>24</sup>**  
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.
- 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 Networks<sup>25</sup>**  
Vanderbilt Biostatistics Qualification Exam
- Review of methods for classifying continuous data streams using neural networks
  - Successfully met qualifying examination standards

- 2015  
|  
2015
- **Asymmetric Linkage Disequilibrium: Tools for Dissecting Multiallelic LD**  
Journal of Human Immunology  
• Authored with Richard Single, Vanja Paunic, Mark Albrecht, and Martin Maiers.
- 2015  
|  
2015
- **An Agent Based Model of Mysis Migration<sup>26</sup>**  
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.

## LINKS

- 1: [http://nickstrayer.me/js4shiny\\_r2d3/slides](http://nickstrayer.me/js4shiny_r2d3/slides)
- 2: <https://livefreeordichotomize.com/>
- 3: [https://livefreeordichotomize.com/2019/06/04/using\\_awk\\_and\\_r\\_to\\_parse\\_25tb/](https://livefreeordichotomize.com/2019/06/04/using_awk_and_r_to_parse_25tb/)
- 4: <https://blogs.rstudio.com/tensorflow/posts/2018-07-17-activity-detection/>
- 5: <https://livefreeordichotomize.com/2018/02/12/the-united-states-of-seasons/>
- 6: <https://livefreeordichotomize.com/2017/12/27/a-year-as-told-by-fitbit/>
- 7: <https://livefreeordichotomize.com/2017/10/14/mcmc-and-the-case-of-the-spilled-seeds/>
- 8: <https://livefreeordichotomize.com/2017/09/25/the-traveling-metallurgist/>
- 9: <https://www.sciencemag.org/news/2017/06/great-paper-swipe-right-new-tinder-preprints-app>
- 10: <https://jhubiostatistics.shinyapps.io/papr/>
- 11: <https://www.nature.com/news/swipe-right-for-science-papr-app-is-tinder-for-preprints-1.22163>
- 12: <https://www.uvm.edu/uvmnews/news/deeper-story-data>
- 13: <https://www.nytimes.com/interactive/2016/08/26/us/college-student-migration.html?smid=pl-share>
- 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: [http://nickstrayer.me/rstudioconf\\_sbm](http://nickstrayer.me/rstudioconf_sbm)
- 19: <https://tbilab.github.io/sbmR/>
- 20: <https://academic.oup.com/bioinformatics/advance-article-abstract/doi/10.1093/bioinformatics/btaa870/5922817?redirectedFrom=fulltext>
- 21: [https://prod.tbilab.org/phewas\\_me\\_info/](https://prod.tbilab.org/phewas_me_info/)
- 22: <https://www.ncbi.nlm.nih.gov/pubmed/30419234>
- 23: [http://nickstrayer.me/rstudioconf19\\_me-poster/](http://nickstrayer.me/rstudioconf19_me-poster/)
- 24: [http://nickstrayer.me/biostat\\_seminar/](http://nickstrayer.me/biostat_seminar/)
- 25: [http://nickstrayer.me/qualifying\\_exam/](http://nickstrayer.me/qualifying_exam/)

26: <https://www.semanticscholar.org/paper/An-Agent-Based-Model-of-the-Diel-Vertical-Migration-Strayer-Stockwell/40493c78e8ecf22bd882d17ec99fd913ec4b9820>