

Education

Drexel University

Ph.D in Information Science

Domain: Natural language processing (NLP), computational social science

Advisor: Dr.Shadi Rezapour

Drexel Graduate Honors: Blue Fellow

Philadelphia, PA

Sept 2021 – Present

University of Miami

M.S. in Public Health

Thesis: A Survey of Imputation Methods for Cervical Cancer Registry Data

Final grade: 3.89/4

Miami, FL

Aug 2015 – Dec 2016

Georgia Institute of Technology

B.S. in Industrial Engineering

Major: Industrial & Systems Engineering

Minor: French

Atlanta, GA

Aug 2009 – Jun 2014

Professional Experience

Drexel University

Research Assistant

2021 – Present

Philadelphia, PA

- Social NLP Lab
 - Using NLP methodologies (document classification, topic modeling, sentiment analysis, keyword extraction, summarization) to understand and classify the social complexities of recovery from substance use disorder (SUD)
 - Research collaboration studying the personal, academic, or professional adoption of generative AI and society
 - Foster interdisciplinary scholarship between software engineering, clinical psychology, and human-computer interaction
 - Mentor Masters and Undergraduate students of Data Science on NLP methodologies and participating in reproducible research
- Explainable AI Lab
 - Project management of lab functions including, developing SOPs for all tasks, meeting notes, scheduling stand-ups, documenting tasks on Jira and issues on GitHub
 - Developed PMCFetchR: retrieve full-text, machine readable articles, as a tidy dataframe meant for text mining. The resulting datasets are used to do named entity recognition (NER) using BioBERT to formulate our agent's explanations.
 - Unit testing of NIH Biomedical Translator initiative components, i.e. bugs, duplication

Creative Artists Agency

Data Science Apprentice

May 2022 – Aug 2022

Los Angeles, CA

- Apply NLP methodologies to discover insight on data unique to entertainment
- Lead discussions on novel project pathways
- Spearheaded academic collaboration between Drexel and CAA
- Analyzed Reddit discussions on HBO's Euphoria

University of Miami Miller School of Medicine*Lead Research Analyst*

2019 – 2021

Miami, FL

- NIH HEALing Communities Study Analyst: study on opioid use disorder (OUD)
 - Built Shiny application to visualize community characters and map of naloxone facilities
 - Designed and disseminated statistical reports and visualizations by community with R and R Mark-down
 - Processed population data for NY communities using the US Census Bureau via API
 - Performed geospatial analysis from shapefiles and by overlaying various population health risk factors
- CTN0094 Analyst: Predict response to OUD treatments
 - Developed spell-checker Shiny application to harmonize and standardize drug names from clinical trial data
- SCAN360.com Project Manager
 - University of Miami COVID-19 Contact Tracking and Tracing Committee Principal Analyst: developed the official UM COVID-19 public-facing dashboard
 - Observed [COVID-19 policy adoption in collaboration with the University of Miami Institute for Advanced Study of the Americas](#), led by Dr.Felicia Knaul, and Universidad Anáhuac in Mexico

University of Miami Sylvester Comprehensive Cancer Center*Biostatistician Programmer*

2017 – 2019

Miami, FL

- Responsible for developing data ingestion procedures
- Calculated various measures of cancer burden including age-adjusted incidence, mortality, survival rates
- Prototyped application's novel data sources and features through the development of Shiny applications
- Fostered interdisciplinary collaborations across departments

Université Sidi Mohamed Ben Abdellah Faculté de Médecine et de Pharmacie*Statistical Consultant*

2014 – 2015

Fes, Morocco

- Assisted faculty and students on developing analysis plans for epidemiological research
- Performed literature reviews and assisted in academic writing

Teaching**INFO 103 Introduction to Data Science***Graduate Teaching Assistant*

Spring 2022

Drexel University

DSCI 501 Quantitative Foundations of Data Science*Graduate Teaching Assistant*

Winter 2022

Drexel University

BST 692 Data Science & Machine Learning for Health Research*Co-Instructor of Record*

Summer 2019-2021

University of Miami

Developed curriculum for computational data science methods to graduate students in a modular format that follows the basic components of a data pipeline. Primary topics taught: version control with git and GitHub, building Shiny applications - theory and applied, building interactive visualizations, and working with geospatial data

BST 625 Statistical Computing*Graduate Teaching Assistant*

Spring 2016

University of Miami

Assisted students with SAS code-based assignments.

BST 601 Introduction to Biostatistics

Graduate Teaching Assistant

Sections include a computing lab - assisted students with SAS code-based assignments.

Fall 2016
University of Miami

Invited Presentations & Guest Lectures

Data Acquisition with Reddit API

INFO 440 Social Media Data Analysis

Apr 2023
Philadelphia, PA

Contributing to COVID-19 Fear Mongering

Data Mishaps

Mar 2023

Virtual

Introduction to Version Control with Git & GitHub

Drexel CCI Data Science Club

Mar 2023
Philadelphia, PA

Sentiment Analysis with Twitter Data

INFO 440 Social Media Data Analysis

Jul 2022
Philadelphia, PA

What Goes Around Comes Back Around: Benefit of Community Involvement

University of Miami

Apr 2022

Virtual

Introduction to Public Health Data Science with the R Programming Language

NSF RESET

Mar 2021

Virtual

Introduction to R Markdown

R-Ladies Miami

Mar 2020

Miami, FL

Tearing Down Barriers to Publication with R Markdown

FIU Workshop

Feb 2020

Miami, FL

Refereed Presentations

Understanding Personal Drug Experiences on Reddit

TADA Conference

Nov 2023
Amherst, MA

Presentation on a multi-level, multi-label classification system to understand personal narratives, particularly around SUD and recovery using few-shot and LLMs.

Evolution of Substance Use Coverage in the Philadelphia Inquirer

ICWSM Data Challenge

June 2023

Virtual

Discourse and Networks Around HBO's Euphoria

AAAS Annual Meeting

Feb 2023

Virtual

DOPE: An R Package for Processing & Classifying Drug Names

R Medicine Conf

Aug 2021

Virtual

Virtual conference lightening talk to demonstrate to international members of the R in Medicine community the functionality of DOPE.

A Web App to Visualize Cancer Risk Factors for 900+ Neighborhoods in Florida

R Medicine Conf

Sept 2018
New Haven, CT

SCAN360

NAACCR Annual Meeting

June 2018
Pittsburgh, PA

Poster on SCAN360.com, a web-platform that visualized multiple SCAN 360 measures of cancer burden for several population subgroups across the state of Florida

Poster Presentations

Real Talk: HBO's Euphoria Viewer Discourse on Substance Use

International Conference on Computational Social Science

July 2023
Copenhagen, DE

pmcFetchR: PMC Full-Text Retrieval for Text Mining

useR Conf

June 2022

Virtual

This poster introduces a novel R package that facilitates full text retrieval from NCBI's PMC OA dataset on AWS.

Selected Publications

1. **Bouzoubaa, L.**, Young, J., & Rezapour, R. (2023). Exploring the Landscape of Drug Communities on Reddit: A Network Study. <https://doi.org/10.1145/3625007.3629125>
2. Luo SX, Feaster DJ, Liu Y, et al. Individual-Level Risk Prediction of Return to Use During Opioid Use Disorder Treatment. *JAMA Psychiatry*. Published online October 04, 2023. doi:10.1001/jamapsychiatry.2023.3596
3. **Bouzoubaa, L.**, Ehsani, R., Chatterjee, P., & Rezapour, R. (2023). The Evolution of Substance Use Coverage in the Philadelphia Inquirer. <https://doi.org/10.48550/ARXIV.2307.01299>
4. Elkbuli, A., Byrne, M. M., Zhao, W., Sutherland, M., McKenney, M., Godinez, Y., Dave, D. J., **Bouzoubaa, L.**, & Koru-Sengul, T. (2020). Gender disparities in lung cancer survival from an enriched Florida population-based cancer registry. *Annals of Medicine and Surgery*, 60, 680–685. <https://doi.org/10.1016/j.amsu.2020.11.081>
5. Bailey, Z., Balise, R., **Bouzoubaa, L.**, & Kobetz, E. (2020). SCAN360: A Resource for a 360-Degree View of Cancer Prevention, Risk, and Survival. *Preventing Chronic Disease*, 17, 200263. <https://doi.org/10.5888/pcd17.200263>
6. Gauri, A., Messiah, S. E., **Bouzoubaa, L. A.**, Moore, K. J., & Koru-Sengul, T. (2020). Cervical cancer sociodemographic and diagnostic disparities in Florida: A population-based study (1981–2013) by stage at presentation. *Ethnicity & Health*, 25(7), 995–1003. <https://doi.org/10.1080/13557858.2018.1471669>
7. Saper, V. E., Chen, G., Deutsch, G. H., Guillerman, R. P., Birgmeier, J., Jagadeesh, K., Canna, S., Schulert, G., Deterding, R., Xu, J., Leung, A. N., **Bouzoubaa, L.**, Abulaban, K., Baszis, K., Behrens, E. M., Birmingham, J., Casey, A., Cidon, M., Cron, R. Q., ... Mellins, E. D. (2019). Emergent high fatality lung disease in systemic juvenile arthritis. *Annals of the Rheumatic Diseases*, 78(12), 1722–1731. <https://doi.org/10.1136/annrheumdis-2019-216040>
8. Pinheiro, P. S., Callahan, K. E., Koru-Sengul, T., Ransdell, J., **Bouzoubaa, L.**, Brown, C. P., & Kobetz, E. (2019). Risk of Cancer Death Among White, Black, and Hispanic Populations in South Florida. *Preventing Chronic Disease*, 16, 180529. <https://doi.org/10.5888/pcd16.180529>
9. Lopategui, D. M., Balise, R. R., **Bouzoubaa, L. A.**, Wilson, S. K., & Kava, B. R. (2018). The Impact of Immediate Salvage Surgery on Corporeal Length Preservation in Patients Presenting with Penile Implant Infections. *Journal of Urology*, 200(1), 171–177. <https://doi.org/10.1016/j.juro.2018.01.082>
10. Picado, O., Baeker Bispo, J., **Bouzoubaa, L.**, Balise, R. R., Lopes, G., & Kobetz, E. N. (2018). Cancer patterns and trends in Costa Rica: A population-based tumor registry study. *Journal of Clinical Oncology*, 36(15_suppl), e13605–e13605. https://doi.org/10.1200/JCO.2018.36.15_suppl.e13605

Software

PMCFetchR	An R package to retrieve full-text articles from the NCBI PMC Articles Dataset on AWS
DOPE	Drug Ontology Parsing Engine: classifying drug names
tidyREDCap	An R package with functions for processing REDCap data

Issued Patents

U.S. Patent US17/273,704	System and method for analyzing and displaying statistical data geographically	2021
U.S Patent US17/288,097	A dynamic multi-factor representation of health data	2021

Certifications

Google Analytics	Google Academy	2020
Associate Cloud Engineer	Google Cloud	2019
Applied Statistics	Pennsylvania State University	2017

Institutional & Community Service

Drexel CCI Doctoral Student Association	Vice President, Treasurer	2022 – Present
GitHub Campus	Advisor	2020 – Present
Google Women Techmakers	Ambassador	2019 – Present
Google #IamRemarkable	Facilitator	2019 – Present
Google Developer Groups Miami	Organizer	2019 – 2022
R-Ladies Miami	Organizer	2019 – 2021

Technical skills

Programming Languages	Python, R, SQL, HTML, CSS
Software	Git, GitHub, Google Cloud Platform
Algorithms	Supervised and unsupervised classification, NLP, large language models
Development	APIs, web scraping

Honors

Dean's Student Leadership & Service Award	Drexel College of Computing & Informatics	2023
Grad Cohort for IDEALS Workshop Fellow	CRA-WP	2023
Outstanding Graduate Teaching Assistant Award	Drexel College of Computing & Informatics	2022
Drexel Blue Fellow	Drexel University Graduate School	2021