



# The Southern California Data Science Program

University of California Irvine: Babak Shahbaba, Mine Dogucu, Robert Pelayo  
California State University Fullerton: Sam Behseta and Jessica Jaynes  
Cypress College: Alma Castro



## Goals

- Obj 1: Create more opportunities for underrepresented students in data science;
- Obj 2: Develop and implement modern data science curricula at the three participating institutes;
- Obj 3: Provide undergraduate students with the required skills to solve real-life problems through hands-on research projects;
- Obj 4: Prepare students for successful careers through direct engagement with academic and non-academic partners.

## Who Are We?

**Why UCI?** Uniquely positioned to become a leader in the field of data science since it is one of the few universities in the U.S. where statistics and computer science are joined together both organizationally and physically. Additionally, a Hispanic-serving institution (HSI), where half of all undergraduates are the first in their families to attend college.

**Why CSUF?** Serving more than 40,000 students as an HSI, it is top-ranked in California and third in the nation for awarding bachelor's degrees to underrepresented students in all disciplines.

**Why Cypress College?** Also an HSI with an average of 16,000 students. It was identified by Niche, Inc. as the 2020 best community college in California.



## Structure of the Program

- Aim: Recruit, train, and dispatch diverse workforce
- Recruit Data science Fellows: 20 from CSUF, 6 from UCI, and 6 from Cypress; 87% of fellows women/underrepresented
- Train: Students take data science related courses at each institution during Winter/Spring
- Research: All fellows participate in Summer Research Experience at UCI with industry partners

## Year One!

- Curriculum:** Four new data science courses have been developed at the three participating institutions.
- Pathways:** Newly created data science course at Cypress College gives rise to a degree program that will contribute to a more effective student transfer to CSU and UC programs.
- Workshops:** Multiple workshops, led by the PIs, were held at CSUF and Cypress College
- Community:** To connect our program to a broader HDR ecosystem, we organized an invited panel session with invited speakers from CSUF, UCR, and UCSB at the 2022 Joint Statistical Meetings at Washington D.C.
- Connection to K-12 Education:** To expose students to data science at early age, we offered two summer programs designed for high school students: one through COSMOS (The California State Summer School for Mathematics and Science) with 20 students and the other one called the UCI DATA (Data Analysis, Theory and Application) Academy with 22 students through the Schools of Information and Computer Science.

## Bootcamp and Research Experience

**Summer Bootcamp:** In a span of a week, a host of technical topics and basic skills were introduced.

	Day 1	Day 2	Day 3	Day 4
Morning Session	Git, GitHub, R Markdown	Exploratory Data Analysis	Linear Regression & Model Evaluation	Generalized Linear Models
Afternoon Session	<i>Ethics in Data Science</i> with Rochelle Tractenberg	<i>Finding a Job in Data Science - UCI</i> Division of Career Pathways	<i>Is grad school right for you?</i> - UCI Graduate Student Panel	<i>Guest Speaker: Dr. Hadley Wickham</i>

**Summer Research:** In partnership with various research entities, Fellows got engaged with an intensive 6-week research program.

**Research Symposium:** Fellows presented their work to faculty and students, friends and family, and community members.

**Publication:** Fellows have continued working on their projects with the aim of submitting their manuscripts for publication.

## Research Projects

**Neurology:** Modeling Non-spatial Sequence Memory Tasks Using Neural Decoding. **New Skills:** Advanced classification models, variable selection strategies, statistical methods in neuroscience.

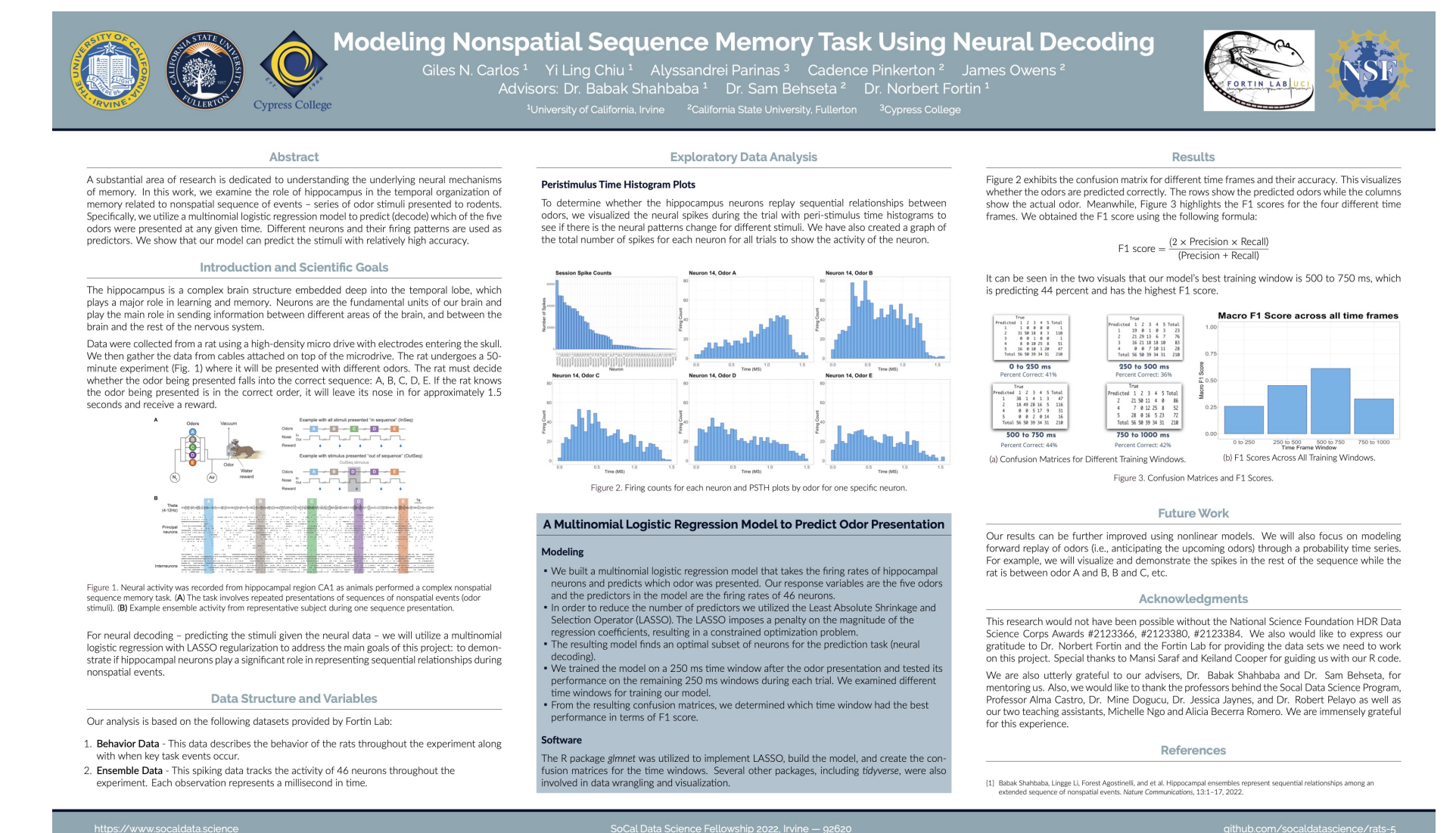
**COVID-19:** Measuring COVID-19 Severity in Children With and Without Cystic Fibrosis. **New Skills:** Generalized linear models, advanced predictive models, variable selection strategies.

**COVID-19:** Modeling COVID-19 Severity in Pediatric Cystic Fibrosis. **New Skills:** Generalized linear models, advanced predictive models, variable selection strategies.

**Oncology:** Impact of Mediterranean and DASH Diets on MPN Symptoms; **New Skills:** Longitudinal models and power analysis.

**Environmental:** An Analysis of Rainfall Over Time. **New Skills:** Time series analysis, ARIMA, Forecasting.

**Education:** How Smartphones Negatively Affect Infant Language Development. **New Skills:** Longitudinal models, design of experiments, statistical methods in education.



## Where Are We Going?

- Continue with curricular development
- Strengthening ties with similar programs across the country for building a community of data science learners
- Engage new industrial partners to introduce our students to more career opportunities.
- Expand our activities and recruit a new cohort of students mentored by the previous fellows to build a diverse and dynamic community of learners in data science.

## Acknowledgements

This program was funded by National Science Foundation HDR Data Science Corps Awards 2123366, 2123380, 2123384.