

One way to not be a physicist

Laura Sampson

4/26/2025

Where have I been?



- Bachelor's degree: CU Boulder, Physics



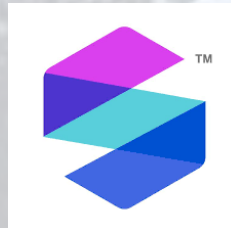
- PhD: Montana State University, Physics



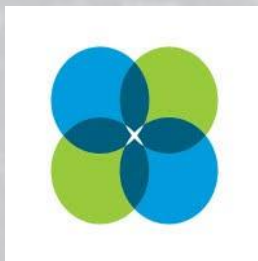
- Postdoc: CIERA Fellow, Northwestern University, Astrophysics



- Postdoc: Center for Infectious Disease Dynamics, Penn State University



- Bioinformatics Scientist: SomaLogic, Boulder, CO



- Biostatistician: Natera, Denver, CO

What have I been doing?



- Bachelor's degree: wind tunnel calibration, femto-second laser pulses, high-energy physics data analysis...



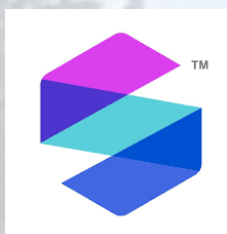
- PhD: tests of general relativity with gravitational waves



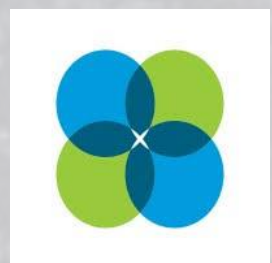
- Postdoc: astrophysical inference with gravitational waves



- Postdoc: optimizing surveillance and vaccination efforts for infectious disease eradication

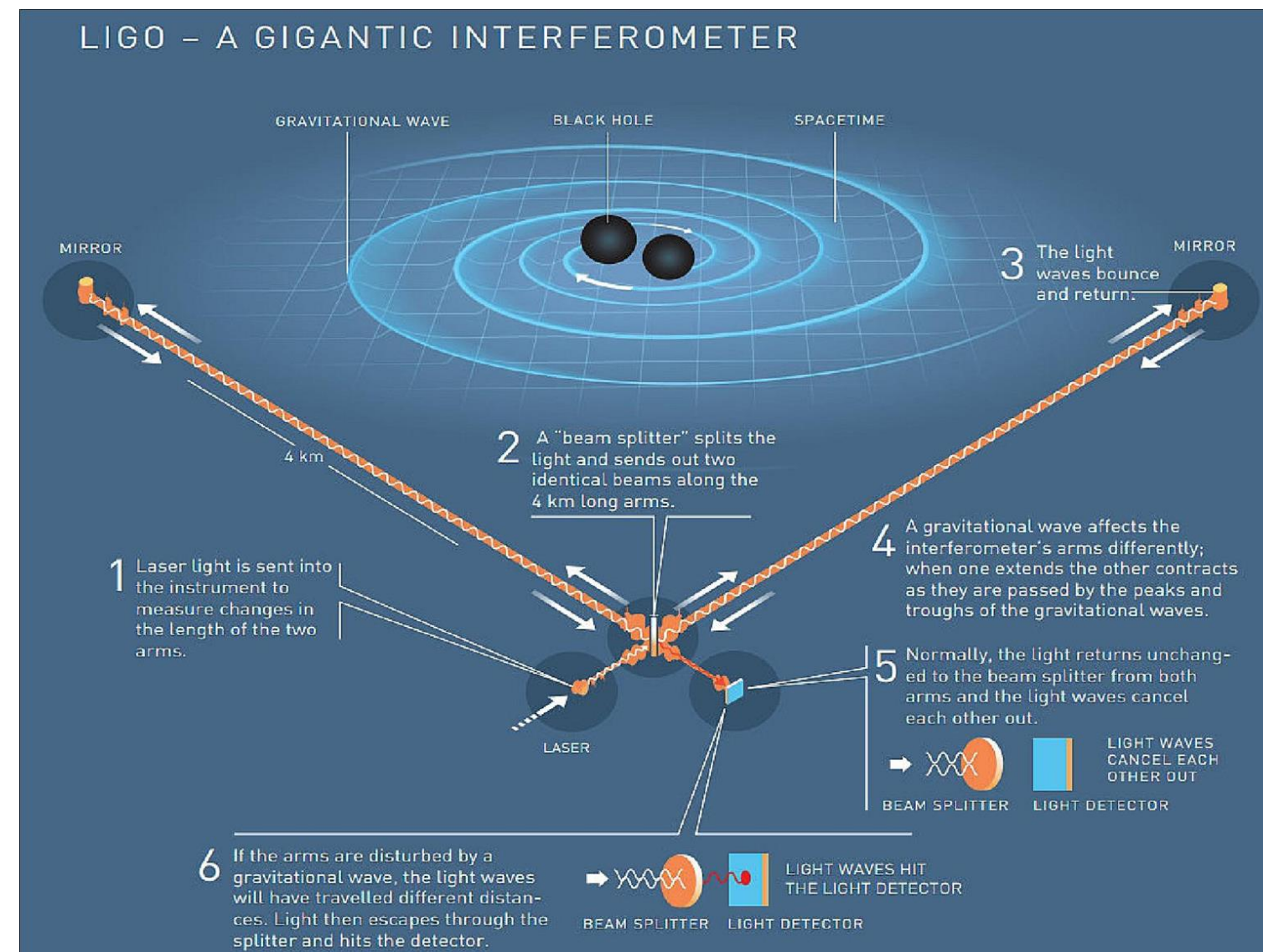


- Bioinformatics Scientist: predictive models for health outcomes based on proteomics data



- Biostatistician: designing and executing studies to show effectiveness of personalized genomic cancer tests

Gravitational Waves

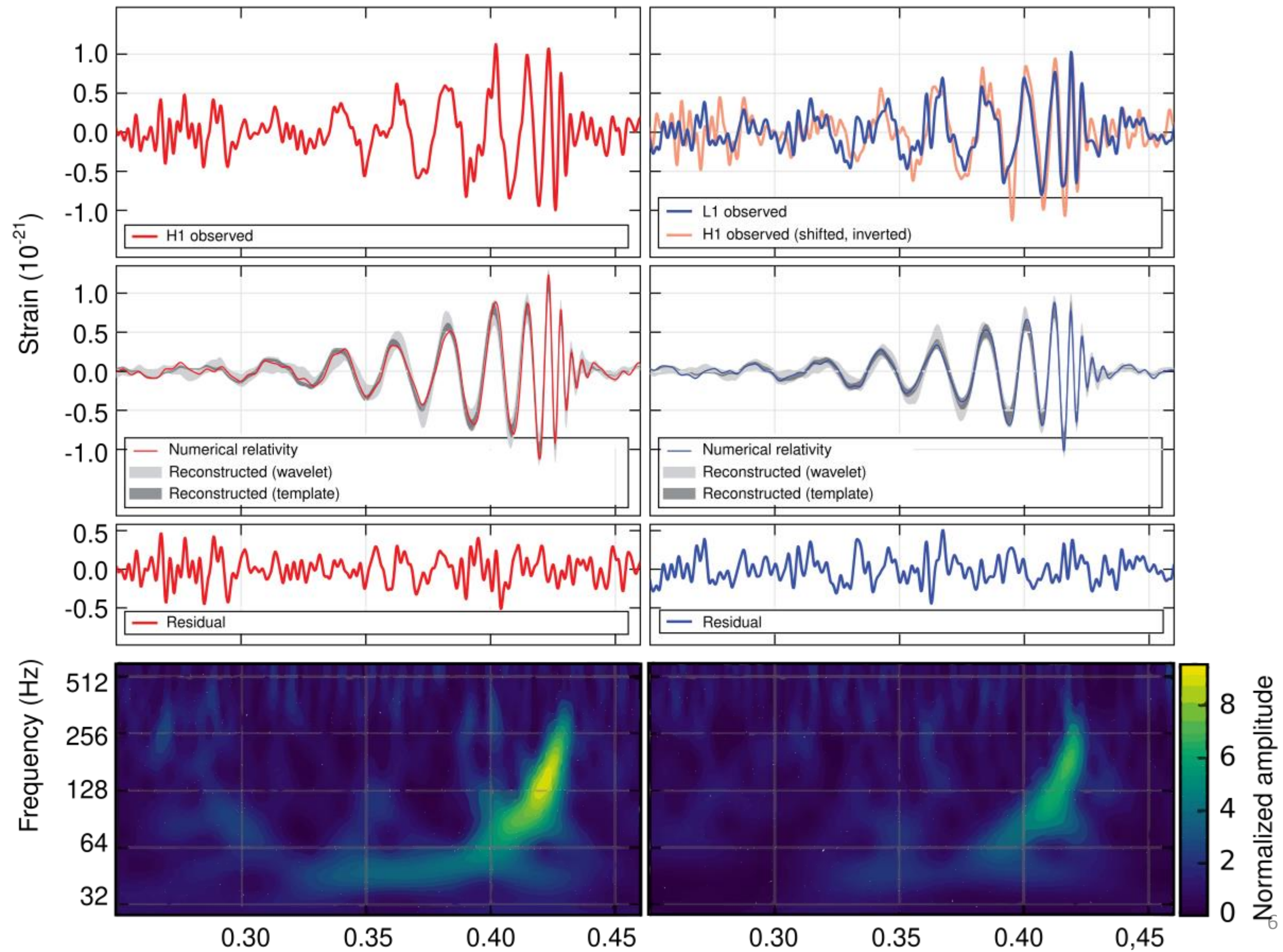


- tests of GR and astrophysical inference with gravitational waves
 - time-series analysis
 - Bayesian inference
 - Gaussian processes
 - coding in C and Python
- Member of LIGO and NANOGrav
 - working in large collaborations
 - presentations
 - publications

One week after I started was finished...

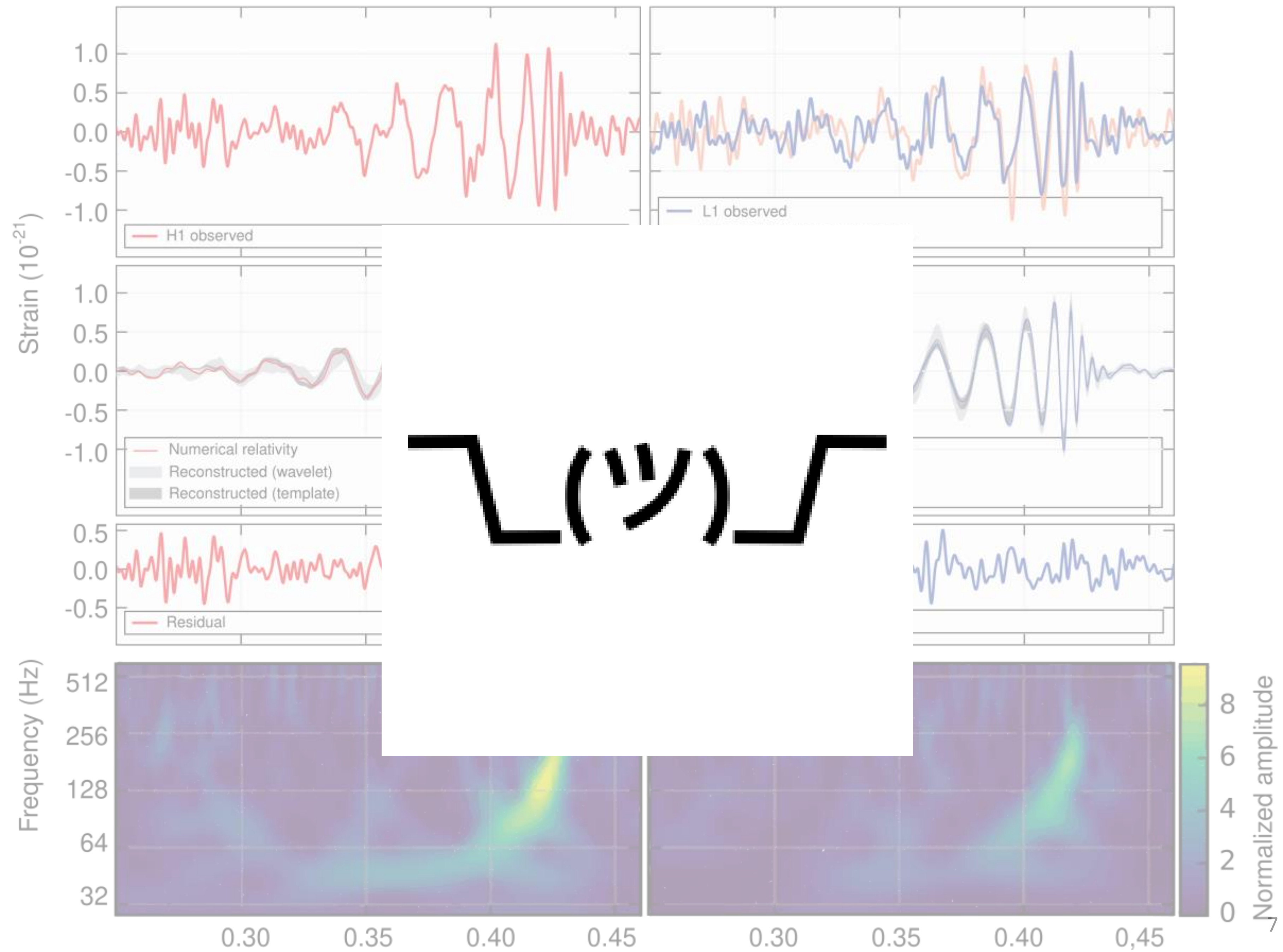
Hanford, Washington (H1)

Livingston, Louisiana (L1)



Hanford, Washington (H1)

Livingston, Louisiana (L1)



So I went to Penn State! How?

- Luck! I had a friend who knew someone who was hiring in a position that I wanted
- Opportunity: I had a fellowship that I could take with me for partial funding
- Skill: I knew how to present the skills I had learned as generalizable to new research

Disease Ecology at Penn State

What did I work on?



Children with central nervous
system infections in central Africa

Due to malaria or cerebral meningitis

antivirals

antibiotics

**Very under-resourced setting! How do we make this decision with
cheap/quickly available data?**

The Data



480 children from Mbarare Hospital

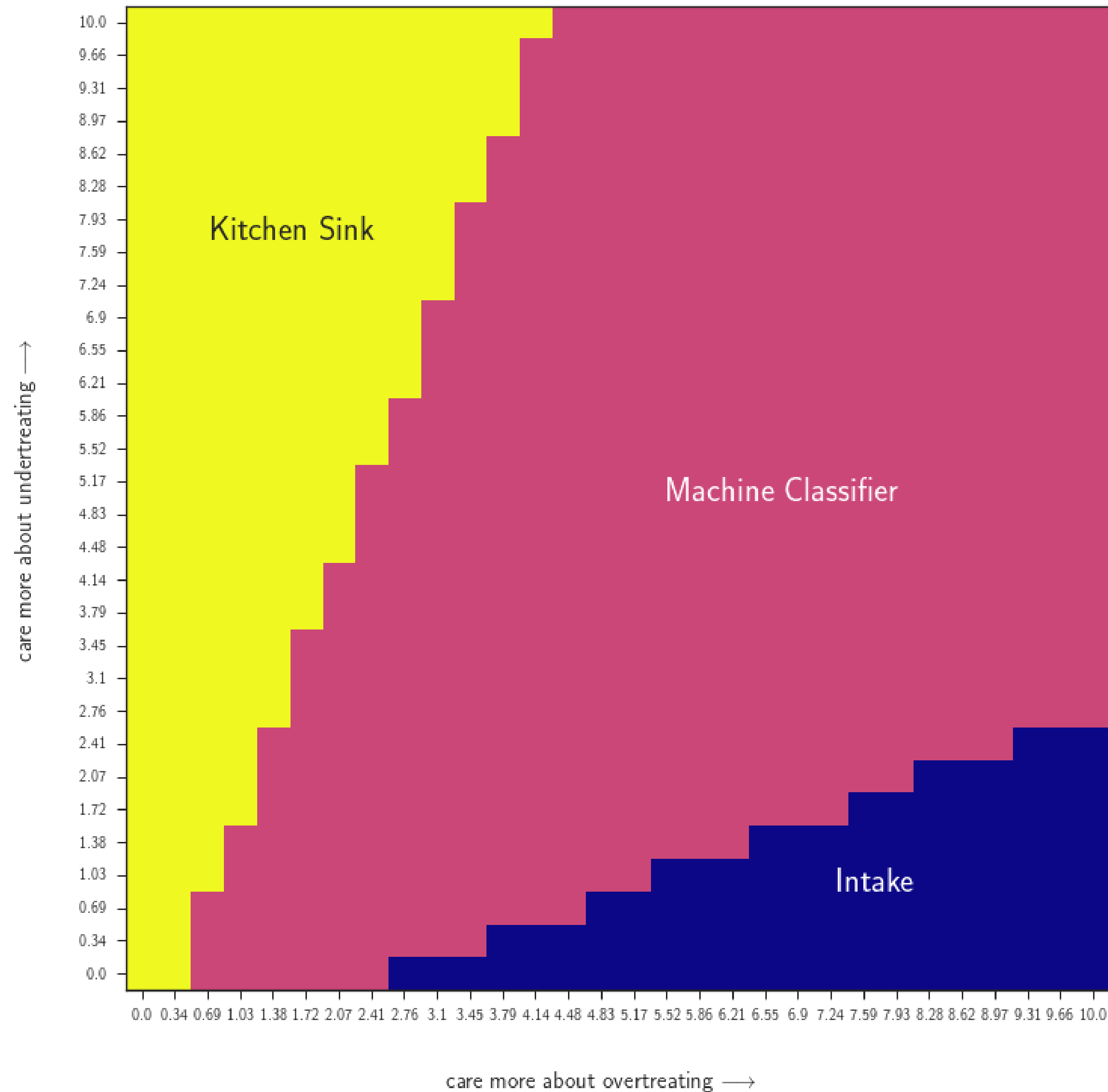
Extensive diagnostics, health
history, treatment information,
and *followup*

This is big data for this setting!

Techniques

- Bayesian logistic regression
- Feature engineering - do we combine different information?
Code it in a particular way?
- Feature selection - what do we drop, what do we keep?
- Defining a cost function - how important is it to not over-treat?
To not under-treat?

Comparison of Policies



This work resulted in actual recommendations to these hospitals on treatment strategies.

We also worked with the government of the DRC on vaccination strategies for measles and rubella. Lots of real-world impact!

Why didn't I keep doing that?

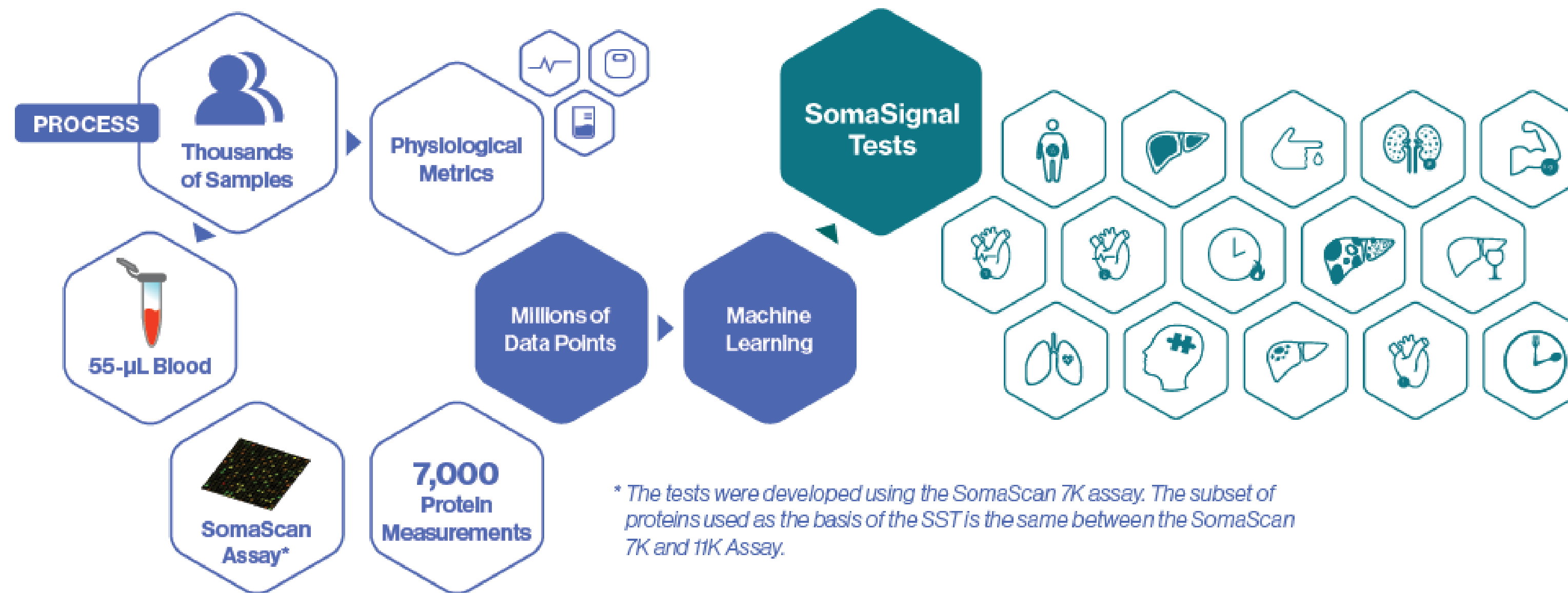


Where there are
in disease ecology



Where I am willing to live

More Luck, More Skill -----> Boulder, CO, working in Proteomics

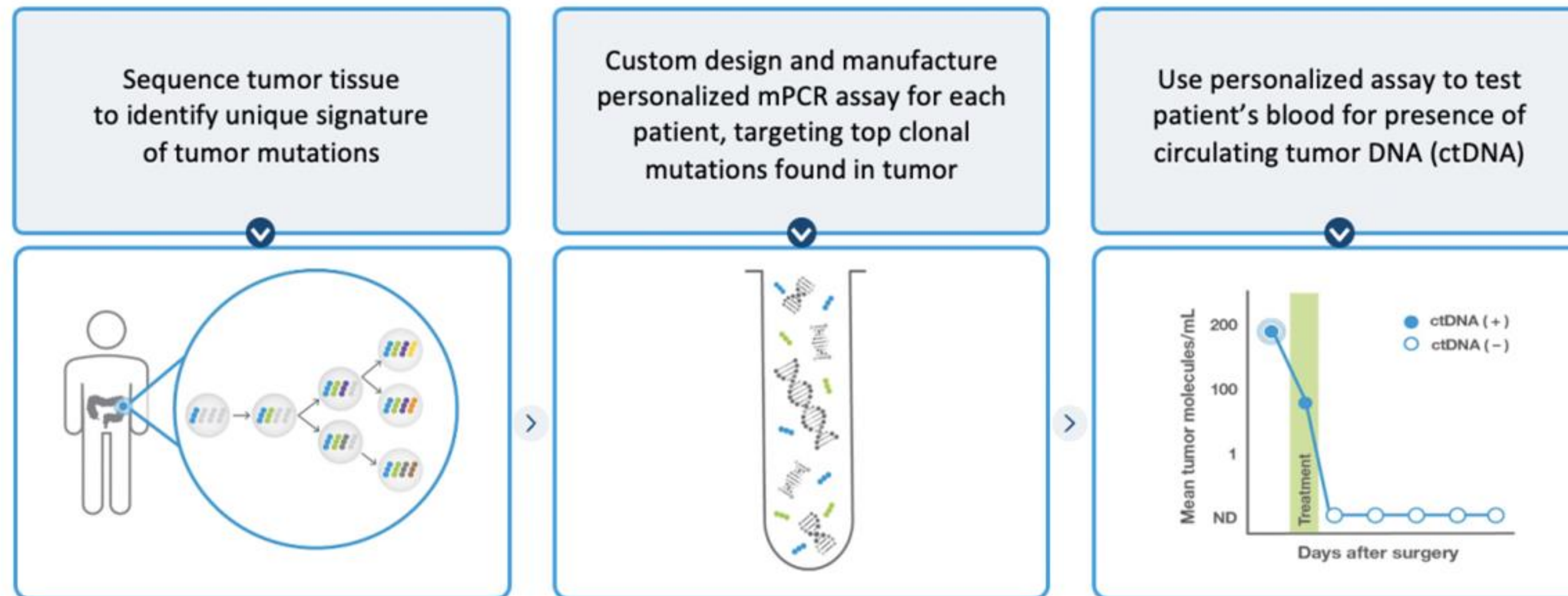


Proteomics: measuring thousands of proteins in your blood, making health predictions from them

Techniques: more Bayesian inference, *survival modeling*, feature engineering, feature selection, *frequentist statistics*

Worse Luck ----> Denver, CO, working as a Biostatistician at Natera

 genomics tests in organ health, women's health, and **oncology**



My job: prove that this is working! (or not!)

New skills: power analysis, sample selection and study design, managing non-technical teammates

Technical Roles at Natera

- Bioinformatics: algorithms for reading DNA (noise modeling), predictive models (classifiers, feature selection, feature engineering, etc.)
- Data Science: pipeline building, data cleaning, data engineering, developing data QCs
- Biostatistics: data analysis, study design, clinical trials

Resumes are Skills Focused and Specific

LAURA SAMPSON

Denver, CO • (303) 883 2372 • lamaesampson@gmail.com

Experienced data scientist with expertise in using multiple statistical and machine learning techniques on both real world and simulated data to draw inferences and provide insight into solutions for highly technical problems in bioinformatics, disease dynamics and epidemiology, predictive modeling, diagnostics, and physics.

AREAS OF EXPERTISE

Data Science and Research

- Data science methods including Bayesian parameter estimation and model selection, Markov chain Monte Carlo, Gaussian process regression and classification, supervised learning, natural language processing, simulation of complicated systems, statistical hypothesis tests, regression models, Cox and AFT survival modeling, feature engineering, predictive modeling, and time series analysis.
- Data cleaning and aggregation, including survey data, natural language data, incomplete/incorrectly entered data, and longitudinal and clinical data.
- Data visualization with matplotlib in Python and ggplot in R.
- Independent research, including design of analyses, interpretation of results of statistical tests, gathering and understanding literature from diverse fields.
- Mathematical skills, including statistics, probability, linear algebra, calculus, dynamical systems.
- Bioinformatics with proteomic and clinical data, translation of data across platforms and matrices

Computer Programming

- Highly experienced with R, Python, C, Linux, shell scripting; coursework using

Matlab, Mathematica, SQL, mapping software.

- Experienced with publicly available libraries Pandas, scikit-learn, basemap, tidyverse.

Communication

- Adept at presenting highly technical material to expert and non-expert audiences from tens to hundreds in size, producing scientific writing in multiple fields, writing successful grant applications.
- Experienced with presentation of complex information to leadership, including c-suite executives.

Leadership

- Experienced in direct management, leading multiple group research projects, and leading a mathematical research group from conceptualizing to completing projects.
- Skilled in organizing professional/scientific meetings, conceptualizing and establishing groups aimed at public engagement, supervising research projects with graduate students, undergraduate students, and employees with credentials from bachelor's degrees to PhDs.

EXPERIENCE

SomaLogic, Inc., Boulder, CO

Bioinformatics Scientist II

April 2019 - present

Performed statistical analyses of high-dimensional proteomic data using primarily R. Analyses included statistical hypothesis tests, regression models, and survival analysis. Built multivariate models for clinical outcomes for continuous, classification, and survival data using elastic net regression, FDA, SVM, Gaussian processes, and neural nets. Led a mathematics research team through projects on cure modeling, visualization of performance metrics, and gaussian processes and neural networks for data translation. Managed a direct report for a year, including mentorship through research projects on model calibration, prediction uncertainty, and developing multiple predictive models for clinical outcomes. Lead author on multiple IP filings for both diagnostic models and novel mathematical techniques.

Center for Infectious Disease Dynamics, Penn State, State College, PA

Postdoctoral Scholar

April 2017 - April 2019

Used machine learning techniques including decision tree classification, Bayesian logistic regression, Gaussian process regression, reinforcement learning, natural language processing, and Markov chain Monte Carlo (MCMC) to explore problems in predicting disease diagnoses and determining optimal response solutions to outbreaks. Wrote thousands of lines of original code in Python, translated R and C++ codes, and implemented existing packages such as scikit-learn. Supervised a graduate student and a recent bachelor's graduate, and worked with colleagues in the computer science department on data visualization. Wrote code to generate simulated data from complicated disease simulations, and analyzed real-world data from health systems ranging from highly developed countries (US) to developing countries (Uganda, Democratic Republic of Congo). Produced four publications, attended three collaborative meetings, and presented at a large conference of disease modelers. Publications and presentations all included data visualization.

Center for Interdisciplinary Exploration and Research in Astrophysics, Northwestern University, IL

Postdoctoral Fellow

September 2015 - April 2017

Used Gaussian process regression, MCMC, and population statistics to use both simulated and actual data to make inferences about populations of astronomical objects based on a small set of observations. Wrote original code (thousands of lines) in Python, and interacted with a large, previously established codebase. Gave three invited colloquia to audiences of undergraduate students, and two invited talks. Presented at five collaboration meetings. Supervised two graduate students and a summer undergraduate. Published six academic papers, all including complicated data visualization and presentation. Co-founded an outreach organization in local businesses and organized a large, multi-media (video, orchestral, and presentation) outreach event for a public audience.

Dept. of Physics, Montana State, Bozeman, MT

Graduate Teaching and Research Assistant

August 2008 - September 2015

Used primarily MCMC methods (code written in C) to learn how to detect extremely quiet signals buried in noise and use these detections to make inferences about astronomy. Taught and graded lab courses of approximately 30 students from various backgrounds in highly technical material. Presented work at more than ten scientific conferences. Founded mentoring organization in the physics department and a university-wide organization for women in science. Helped to organize a scientific meeting of 100 colleagues, paired with an outreach event that reached thousands in Montana.

EDUCATION

Ph.D., Physics, Montana State University, Bozeman, MT (2014)

B.A., *Magna Cum Laude with Distinction*, Physics, University of Colorado, Boulder, CO (2007)

SELECT PUBLICATIONS

15 total publications (full list available at <https://tinyurl.com/54fbzs35>), including:

Demography, education, and research trends in the interdisciplinary field of disease ecology. EE Brandell, DJ Becker, L Sampson, KM Forbes. *Ecology and Evoution* Vol. 11, Issue 24, December, 2021.

Epidemic growth rates and host movement patterns shape management performance for pathogen spillover at the wildlife/livestock interface. Kezia Manlove, Laura Sampson, Benny Borremans, E. Francis Cassirer, Ryan S. Miller, Kim M. Pepin, Thomas E. Besser, and Paul C. Cross. *Phil. Trans. B* Vol. 374, Issue 1782, September, 2019

A novel catalytic model for estimating age-specific measles sero-prevalence using case-incidence data: a case study from the Democratic Republic of the Congo. Spencer Carran, Steve Ahuka, Simon Bezirganian, Matthew Ferrari, and Laura Sampson. *Submitted*.

Constraints on the Dynamical Environments of Supermassive Blackhole Binaries Using Pulsar Timing Arrays. Stephen Taylor, Joseph Simon, and Laura Sampson. *Phys. Rev. Lett.* 118, 181102, March, 2017.

Observation of Gravitational Waves from a Binary Black Hole Merger. The LIGO Scientific Collaboration. *PhysRevLett*.116.061102 (2016).

Resumes are Skills Focused and Specific

LAURA SAMPSON

Denver, CO • (303) 883 2372 • lmaesampson@gmail.com

Experienced data scientist with expertise in using multiple statistical and machine learning techniques on both real world and simulated data to draw inferences and provide insight into solutions for highly technical problems in bioinformatics, disease dynamics and epidemiology, predictive modeling, diagnostics, and physics.

AREAS OF EXPERTISE

Data Science and Research

- Data science methods including Bayesian parameter estimation and model selection, Markov chain Monte Carlo, Gaussian process regression and classification, supervised learning, natural language processing, simulation of complicated systems, statistical hypothesis tests, regression models, Cox and AFT survival modeling, feature engineering, predictive modeling, and time series analysis.
- Data cleaning and aggregation, including survey data, natural language data, incomplete/incorrectly entered data, and longitudinal and clinical data.
- Data visualization with matplotlib in Python and ggplot in R.
- Independent research, including design of analyses, interpretation of results of statistical tests, gathering and understanding literature from diverse fields.
- Mathematical skills, including statistics, probability, linear algebra, calculus, dynamical systems.
- Bioinformatics with proteomic and clinical data, translation of data across platforms and matrices

Computer Programming

- Highly experienced with R, Python, C, Linux, shell scripting; coursework using

Matlab, Mathematica, SQL, mapping software.

- Experienced with publicly available libraries Pandas, scikit-learn, basemap, tidyverse.

Communication

- Adept at presenting highly technical material to expert and non-expert audiences from tens to hundreds in size, producing scientific writing in multiple fields, writing successful grant applications.
- Experienced with presentation of complex information to leadership, including c-suite executives.

Leadership

- Experienced in direct management, leading multiple group research projects, and leading a mathematical research group from conceptualizing to completing projects.
- Skilled in organizing professional/scientific meetings, conceptualizing and establishing groups aimed at public engagement, supervising research projects with graduate students, undergraduate students, and employees with credentials from bachelor's degrees to PhDs.

EXPERIENCE

SomaLogic, Inc., Boulder, CO Bioinformatics Scientist II

April 2019 - present

Performed statistical analyses of high-dimensional proteomic data using primarily R. Analyses included statistical hypothesis tests, regression models, and survival analysis. Built multivariate models for clinical outcomes for continuous, classification, and survival data using elastic net regression, FDA, SVM, Gaussian processes, and neural nets. Led a mathematics research team through projects on cure modeling, visualization of performance metrics, and gaussian processes and neural networks for data translation. Managed a direct report for a year, including mentorship through research projects on model calibration, prediction uncertainty, and developing multiple predictive models for clinical outcomes. Lead author on multiple IP filings for both diagnostic models and novel mathematical techniques.

Center for Infectious Disease Dynamics, Penn State, State College, PA

Postdoctoral Scholar

April 2017 - April 2019

Used machine learning techniques including decision tree classification, Bayesian logistic regression, Gaussian process regression, reinforcement learning, natural language processing, and Markov chain Monte Carlo (MCMC) to explore problems in predicting disease diagnoses and determining optimal response solutions to outbreaks. Wrote thousands of lines of original code in Python, translated R and C++ codes, and implemented existing packages such as scikit-learn. Supervised a graduate student and a recent bachelor's graduate, and worked with colleagues in the computer science department on data visualization. Wrote code to generate simulated data from complicated disease simulations, and analyzed real-world data from health systems ranging from highly developed countries (US) to developing countries (Uganda, Democratic Republic of Congo). Produced four publications, attended three collaborative meetings, and presented at a large conference of disease modelers. Publications and presentations all included data visualization.

Center for Interdisciplinary Exploration and Research in Astrophysics, Northwestern University, IL Postdoctoral Fellow

September 2015 - April 2017

Used Gaussian process regression, MCMC, and population statistics to use both simulated and actual data to make inferences about populations of astronomical objects based on a small set of observations. Wrote original code (thousands of lines) in Python, and interacted with a large, previously established codebase. Gave three invited colloquia to audiences of undergraduate students, and two invited talks. Presented at five collaboration meetings. Supervised two graduate students and a summer undergraduate. Published six academic papers, all including complicated data visualization and presentation. Co-founded an outreach organization in local businesses and organized a large, multi-media (video, orchestral, and presentation) outreach event for a public audience.

Dept. of Physics, Montana State, Bozeman, MT

Graduate Teaching and Research Assistant

August 2008 - September 2015

Used primarily MCMC methods (code written in C) to learn how to detect extremely quiet signals buried in noise and use these detections to make inferences about astronomy. Taught and graded lab courses of approximately 30 students from various backgrounds in highly technical material. Presented work at more than ten scientific conferences. Founded mentoring organization in the physics department and a university-wide organization for women in science. Helped to organize a scientific meeting of 100 colleagues, paired with an outreach event that reached thousands in Montana.

EDUCATION

Ph.D., Physics, Montana State University, Bozeman, MT (2014)

B.A., *Magna Cum Laude with Distinction*, Physics, University of Colorado, Boulder, CO (2007)

SELECT PUBLICATIONS

15 total publications (full list available at <https://tinyurl.com/54fbzs35>), including:

Demography, education, and research trends in the interdisciplinary field of disease ecology. EE Brandell, DJ Becker, L Sampson, KM Forbes. *Ecology and Evoution* Vol. 11, Issue 24, December, 2021.

Epidemic growth rates and host movement patterns shape management performance for pathogen spillover at the wildlife/livestock interface. Kezia Manlove, Laura Sampson, Benny Borremans, E. Francis Cassirer, Ryan S. Miller, Kim M. Pepin, Thomas E. Besser, and Paul C. Cross. *Phil. Trans. B* Vol. 374, Issue 1782, September, 2019

A novel catalytic model for estimating age-specific measles sero-prevalence using case-incidence data: a case study from the Democratic Republic of the Congo. Spencer Carran, Steve Ahuka, Simon Bezirganian, Matthew Ferrari, and Laura Sampson. *Submitted*.

Constraints on the Dynamical Environments of Supermassive Blackhole Binaries Using Pulsar Timing Arrays. Stephen Taylor, Joseph Simon, and Laura Sampson. *Phys. Rev. Lett.* 118, 181102, March, 2017.

Observation of Gravitational Waves from a Binary Black Hole Merger. The LIGO Scientific Collaboration. *PhysRevLett*.116.061102 (2016).

Resumes are Skills Focused and Specific

LAURA SAMPSON

Denver, CO • (303) 883 2372 • lmaesampson@gmail.com

Experienced data scientist with expertise in using multiple statistical and machine learning techniques on both real world and simulated data to draw inferences and provide insight into solutions for highly technical problems in bioinformatics, disease dynamics and epidemiology, predictive modeling, diagnostics, and physics.

AREAS OF EXPERTISE

Data Science and Research

- Data science methods including Bayesian parameter estimation and model selection, Markov chain Monte Carlo, Gaussian process regression and classification, supervised learning, natural language processing, simulation of complicated systems, statistical hypothesis tests, regression models, Cox and AFT survival modeling, feature engineering, predictive modeling, and time series analysis.
- Data cleaning and aggregation, including survey data, natural language data, incomplete/incorrectly entered data, and longitudinal and clinical data.
- Data visualization with matplotlib in Python and ggplot in R.
- Independent research, including design of analyses, interpretation of results of statistical tests, gathering and understanding literature from diverse fields.
- Mathematical skills, including statistics, probability, linear algebra, calculus, dynamical systems.
- Bioinformatics with proteomic and clinical data, translation of data across platforms and matrices

Computer Programming

- Highly experienced with R, Python, C, Linux, shell scripting; coursework using

Matlab, Mathematica, SQL, mapping software.

- Experienced with publicly available libraries Pandas, scikit-learn, basemap, tidyverse.

Communication

- Adept at presenting highly technical material to expert and non-expert audiences from tens to hundreds in size, producing scientific writing in multiple fields, writing successful grant applications.
- Experienced with presentation of complex information to leadership, including c-suite executives.

Leadership

- Experienced in direct management, leading multiple group research projects, and leading a mathematical research group from conceptualizing to completing projects.
- Skilled in organizing professional/scientific meetings, conceptualizing and establishing groups aimed at public engagement, supervising research projects with graduate students, undergraduate students, and employees with credentials from bachelor's degrees to PhDs.

EXPERIENCE

SomaLogic, Inc., Boulder, CO

Bioinformatics Scientist II

April 2019 - present

Performed statistical analyses of high-dimensional proteomic data using primarily R. Analyses included statistical hypothesis tests, regression models, and survival analysis. Built multivariate models for clinical outcomes for continuous, classification, and survival data using elastic net regression, FDA, SVM, Gaussian processes, and neural nets. Led a mathematics research team through projects on cure modeling, visualization of performance metrics, and gaussian processes and neural networks for data translation. Managed a direct report for a year, including mentorship through research projects on model calibration, prediction uncertainty, and developing multiple predictive models for clinical outcomes. **Lead author on multiple IP filings** for both diagnostic models and novel mathematical techniques.

Center for Infectious Disease Dynamics, Penn State, State College, PA

Postdoctoral Scholar

April 2017 - April 2019

Used machine learning techniques including decision tree classification, Bayesian logistic regression, Gaussian process regression, reinforcement learning, natural language processing, and Markov chain Monte Carlo (MCMC) to explore problems in predicting disease diagnoses and determining optimal response solutions to outbreaks. **Wrote thousands of lines of original code in Python, translated R and C++ codes**, and implemented existing packages such as scikit-learn. Supervised a graduate student and a recent bachelor's graduate, and worked with colleagues in the computer science department on data visualization. Wrote code to generate simulated data from complicated disease simulations, and analyzed real-world data from health systems ranging from highly developed countries (US) to developing countries (Uganda, Democratic Republic of Congo). Produced four publications, attended three collaborative meetings, and presented at a large conference of disease modelers. Publications and presentations all included data visualization.

Center for Interdisciplinary Exploration and Research in Astrophysics, Northwestern University, IL

Postdoctoral Fellow

September 2015 - April 2017

Used Gaussian process regression, MCMC, and population statistics to use both simulated and actual data to make inferences about populations of astronomical objects based on a small set of observations. Wrote original code (thousands of lines) in Python, and interacted with a large, previously established codebase. Gave three invited colloquia to audiences of undergraduate students, and two invited talks. Presented at five collaboration meetings. Supervised two graduate students and a summer undergraduate. **Published six academic papers**, all including complicated data visualization and presentation. Co-founded an outreach organization in local businesses and organized a large, multi-media (video, orchestral, and presentation) outreach event for a public audience.

Dept. of Physics, Montana State, Bozeman, MT

Graduate Teaching and Research Assistant

August 2008 - September 2015

Used primarily MCMC methods (code written in C) to learn how to detect extremely quiet signals buried in noise and use these detections to make inferences about astronomy. Taught and graded lab courses of approximately 30 students from various backgrounds in highly technical material. Presented work at more than ten scientific conferences. Founded mentoring organization in the physics department and a university-wide organization for women in science. Helped to **organize a scientific meeting of 100 colleagues**, paired with an outreach event that reached thousands in Montana.

EDUCATION

Ph.D., Physics, Montana State University, Bozeman, MT (2014)

B.A., Magna Cum Laude with Distinction, Physics, University of Colorado, Boulder, CO (2007)

SELECT PUBLICATIONS

15 total publications (full list available at <https://tinyurl.com/54fbzs35>), including:

Demography, education, and research trends in the interdisciplinary field of disease ecology. EE Brandell, DJ Becker, L Sampson, KM Forbes. *Ecology and Evoution* Vol. 11, Issue 24, December, 2021.

Epidemic growth rates and host movement patterns shape management performance for pathogen spillover at the wildlife/livestock interface. Kezia Manlove, Laura Sampson, Benny Borremans, E. Francis Cassirer, Ryan S. Miller, Kim M. Pepin, Thomas E. Besser, and Paul C. Cross. *Phil. Trans. B* Vol. 374, Issue 1782, September, 2019

A novel catalytic model for estimating age-specific measles sero-prevalence using case-incidence data: a case study from the Democratic Republic of the Congo. Spencer Carran, Steve Ahuka, Simon Bezirganian, Matthew Ferrari, and Laura Sampson. *Submitted*.

Constraints on the Dynamical Environments of Supermassive Blackhole Binaries Using Pulsar Timing Arrays. Stephen Taylor, Joseph Simon, and Laura Sampson. *Phys. Rev. Lett.* 118, 181102, March, 2017.

Observation of Gravitational Waves from a Binary Black Hole Merger. The LIGO Scientific Collaboration. *PhysRevLett*.116.061102 (2016).

You Are Learning a Lot That is Not Physics

Pros

and

Cons

- Compensation
- Project and Role Clarity
- Concrete impact
- Career progression
- Boundaries
- Stability

- Deadlines
- Lack of flexibility
- Lack of research freedom
- Focus on profit
- Stability