What statistical techniques have you used and what were the specific applications? Why did you select those techniques?

During my Bachelor's and Master's degrees I focused on deterministic population dynamics models for making future bird population forecasts and classic frequentist statistics to compare experimental treatments. I learned the standard suite of tools like the t-test, ANOVA, and chi-squared comparisons that worked well for my simple data and questions at the time. My population dynamics skills got me a position in Brett Melbourne's lab in the EBIO department at CU for a Ph.D. and he taught me how to translate my frequentist skills into Bayesian and stochastic thinking. We built complex population dynamics models together (e.g. differential equation sets) that were either partially or fully stochastic and we fit those models to data using MCMC parametrization and then used those parameterized models to make beetle population forecasts and spatial spread forecasts. We ran 7 years of experiments together and used Bayesian experimental statistics to compare treatments in that work. I will be back in Brett's lab for the summer finishing up some publications together and fitting some new stochastic spatial spread models to our last 10 years of experimental data using MCMC.

I started incorporating spatial analyses into my work during my Master's and then I made it the centerpiece of my Ph.D. and then both of my postdocs. I took my first GIS class with Babs Buttenfield in the CU geography department during my Master's and then she joined my Ph.D. committee and helped me predicted bird migration patterns. I ran S-splines on MODIS data to get a vertical profile of the air matching each individual bird's GPS track and then identified the abiotic niche dimensions of those individuals to predict mechanism. I built a 4d model of Earth's rotation to show that the birds were staying still relative to the sun and the planet was moving under them in an analemma to create a figure-8 shaped migratory path. I also devised a statistic for that model where I used the geodesic distance between the model and the data in the same way that sum of squares measure the distance between the model and the data.

I continued working with global raster data sets during my first postdoc where I worked with my team at the Max Planck Institute to hindcasted and forecasted the global distribution of agriculture using climate simulations based on the WorldClim dataset. During that postdoc I also learned to do phylogenetic analyses. I was given a phylogeny of language, which I used to build macroevolutionary models about the warming of the human niche, the origins of land ownership, and the origins and subsequent spread of agricultures. In that work, I personally modeled data using PGLS to run phylogenetically weighted linear regressions, PCA to collapse environmental variables into model parameters, Species distribution models (SDM and MSDM) of humans and agricultural species (e.g. 116 domesticated species) to identify the niche interdependence between humans and their domesticated species, SEMs to predict the origins of human social inequality, and network-dispersal MCMC simulations to identify the mechanisms behind the spread of agricultural ideas and technology. I used random forest Machine Learning extensively during this postdoc to classify simulations and compare them to real data. Much of this work was done on a collection of three super-computers that supplied us with 2700 continuously running cores for 18 months. I became very proficient at working and teaching using those HPCs.

My current postdoc puts me at the head of a team tasked with designing a data pipeline and visualization dashboard to help municipal decision makers make more sustainability decisions. This process has involved designing computer pipelines for combining many different types of data that describe complex interactions between natural systems, human social systems, transport networks, and the human-built environment using interactive visualizations. I have made Open Street Map the centerpiece of this platform and that decision has opened-up a new world of opportunities for analyzing networks and origin/destination routing data. I've used measures of network centrality and betweenness to judge the efficacy of different transport networks, used census aggregation to report on the equity implications of Covid19 pedestrian corridors, built custom GTFS feeds for 6 different potential subway routes so we could calculate travel time matrixes and equality metrics from each line. I've designed my own green space networks using optimal transport theory and Constructal law and I've used routing and elevation models to calculate $CO_2$ emissions across entire city. I've pushed myself to go beyond the standard raster datasets beloved by ecology and added vector data that urban planners and sociologists find so useful. I have also found that I naturally gravitate towards setting up digital microcosms to experiment with. I also frequently use digital twins to experiment with different mechanisms driving the behavior of a system.