This quarter, under the guidance of Ethan Ancell and alongside my fellow mentee Alexis, did a preliminary exploration of survival analysis and its applications with R.

We first learned a brief overview of what survival analysis (and time-to-event data) is, how censoring works and its types, the survival function (the thing we use to apply survival analysis), the hazard function (instantaneous risk), how to implement survival analysis in R, how to manually code (and plot) the naive and kaplan meier estimators of the survival function, how to use R's survival library to code a KM estimator of the survival function, how to code and plot the cox regression in R (with a different teaching dataset).

At the end of the quarter, the project split between mathy/application summary of survival analysis up to what we have learned so far. Alexis did the mathy part, I did the application part. My half of the final project was exploring how survival analysis can be applied to real world data; how to set up/find a survival dataset, what the plots of survival function estimate and hazard function tell us about the dataset, what the plots of each mean, definitively (e.g. value at S(32) = probability of risk after t=32), and how to specifically work/use the Natality teaching dataset (2018 US birth data).