Final Project Proposal

Adith Arun and Chloe Warren

We will be investigating weather data over the last ten years in the state of California. We will gather daily mean temperature values for many locations within the state – probably between two to three hundred (or as many as we have access to). The data will come from NOAA where they have data from measurement locations across the nation. We will use their API and query data every day for the past ten years across these locations and record the daily mean temperature (All analysis to be done in R). To make the data more manageable, we will aggregate temperature data to be a monthly value by taking a median across daily values. We are mostly considering latitude and longitude but will also consider other variables besides these like altitude to include multiple data types.

We will then fit a variogram for each month's worth of data. Depending on the scale, we may observe non-stationarity (California has coastal, arid, and temperate areas). We will experiment with different variogram model fits. From here, for each year's worth of data, we will formulate a few key questions in appropriate kriging (ok, sk, kriging of the mean) frameworks:

- attempt to estimate the state-wide mean temperature for a given month
- mean temperature for a given month in locations not queried (note: we will still query data for these locations, but will not include them in the data we use in creating the kriging expressions such that we can assess accuracy)

We can compare how these estimates compare across year. For example, we can look at how the mean temperature in January 2011 California compared to January 2012. We can compare our estimates for unknown locations across different kriging methods and non-kriging methods (e.g., IDW).