# Preliminary Research and Question Development

Bioclim

* Climate Envelope Method
* Occurrence points and environment space – draws a box around it
* Look at environmental gradients of various variables and see what conditions a species appears in most. Assume that the highest quality habitat are the most populated areas
* Does not extrapolate much. Quite 2 dimensional
* Finds combinations of environments that do not really exist

Domain

* Climate Envelope Method
* Environment space distance to other data points using a modified Manhattan distance rather than Euclidian
* A math equations and formulas

  Description automatically generated with medium confidence
* The domain score for this combination environment is proportional to the shortest of those distances

GLM

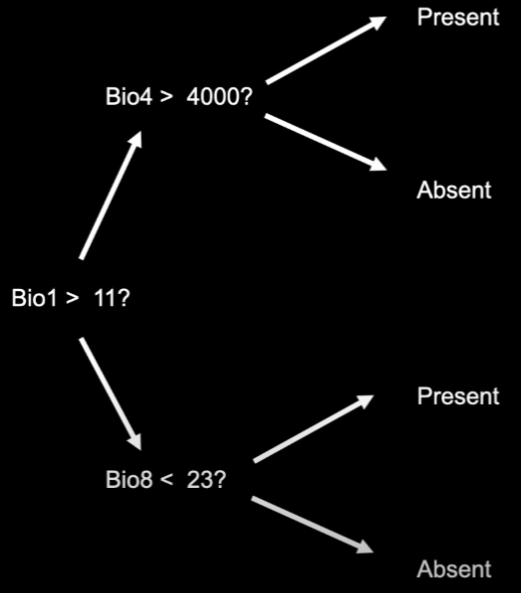
* Extremely flexible and complex
* Generalized Linear Modelling
* Uses background, pseudoabsence, or absence points as well as occurrence data
* Fits models using maximum likelihood
* Very flexible but requires a lot of hands-on work (you have to know what to do, needs great understanding)
* Binomial GLM regression on presence/absence data
* A graph of a function

  Description automatically generated

GAM

* Generalised Additive Model
* Looks quite a bit like GLMs
* The primary difference is that they allow us to fit a species’ response using a nonlinear smoother function of the environmental variables
* Has more localized fits to the equation
* Allows you to control how complex the function will be
* We can combine smoothed predictors with unsmoothed predictors

Random Forests

* Models are based on decision trees
* Decision trees take a set of outcomes and a set of predictors and produce a set of binary decisions intended to predict those outcomes
* Decision trees: Start by choosing the predictor and threshold value that best separates our data into classes (presence vs. absence or pseudoabsense)
  + As only presence data can truly be available, Pseudoabsense is artificial absence data, sometimes called background data
* 
* Tend to fit training data extremely well
* Can often be terrible at predicting other data
* Random forests extends decision trees to compensate for limitations
* For each replicate, start by randomly resampling your occurrences with replacement
  + Helps control overfitting to training data
* At each node, sample a random subset of your predictors
  + Helps control for correlations between trees based on one or a few strong predictors