**Next steps for RSF project**

**Goal: draft by the end of February**

-break this down further

**Practical next steps:**

* Set up a version of PRISM data that I will use for modelling
  + Correcting previous scripts to work with new file
  + Extract the proper coordinates from GIS data into PRISM excel file

Notes from Isabel (emails forwarded by Laurent, Nov 25):

-Please, refer to any data from Tyler to find correct Region/Subregion designations for the plots as well as for the plot UTMs.

-Could you, please, include "order" in addition to:

Map\_datum

UTM\_zone

UTM\_easting

UTM\_northing

Comment: anything you think is noteworthy about a plot location, shape, ….

* Define a grid to resample environmental layers and predict onto with model
  + - Define myself – probably 30m? look at my environmental layers
      * Might be able to use what I had created for version 1
* Start trying out modelling process
  + Figure out data needs
  + Figure out computational power needs
  + Sketch an outline of hypotheses/paper to guide decisions, and sketch of how I think I might do the analysis
  + Test out analysis methods to see if they do what I think they will do
    - Path analysis or SEM component
    - Multi-scale component

**Research Questions:**

* + Basic literature search/ introduction/methods for each of the research questions below to frame possible ideas/discuss with Paul
* How does the relative influence of (abiotic and biotic) / (resource, direct and indirect) variables change with spatial scale? How transferrable are these patterns across space and time?
* What time frame do species respond to for annually variable predictor variables (climate, surface water, snowmelt)? Are there time lags?
* What is the best way to identify which predictors will be the best for predicting responses to climate change?

**Modelling Approach:**

<https://damariszurell.github.io/SDM-Intro/>

* Presence-Absence vs Presence-Background: I was always under the impression that PA would be better, but one paper suggested otherwise and I didn’t understand why
* Dispersal barriers: community species pool
* Scale of effect: nested spatial scales, optimized spatial scales
  + Do scale and extent of biodiversity and environmental data match?
* Test for multicollinearity of environmental data, spatial autocorrelation
* Algorithms: GLM, GAM, machine learning, maxent, ensemble modelling, model averaging
* Threshold for binary prediction of presence/absence?
* Model selection: AIC, LRT….?
* Model validation: k-folds cross validation, training&testing split, independent data
* Model performance: AUC, True Skill Statistic etc

-how do I account for regional/annual differences?

-how can you account for social attraction?

-does setting up the habitat data as a percent coverage per plot for each habitat type make sense? Should I use coarser habitat types or finer?  
-how does prevalence (common vs rare species) influence results?

-I keep reading that depending on whether you’re trying to explain causal relationships that drive species distributions or whether you’re trying to predict species distributions should influence how models are fitted and evaluated but I don’t actually know what that means

* + Paul: explanatory = small number of variables that make the most sense, predictive = lots of different things to get the best r2, even if they don’t make sense. BUT then if you want to make it generalizable, this may not be the case