**Introduction**

* Introduce goals of project – trying to identify factors that connect with outcomes for people who are part of The Connection’s programs
* Discuss literature (copy over information from proposal)

**Research Questions**

1. What factors influence successful, neutral, and unsuccessful outcomes for people in rehabilitation programs at the Connection?
   1. What is the relationship between risk of recidivism, as measured through the Ohio Risk Assessment Survey, and successful, neutral, and unsuccessful outcomes?
   2. How do experiences suicide and homicide risk levels, criminal thinking profile traits, experiences with drug and alcohol usage, and demographic variables relate to successful or unsuccessful outcomes?

**Methods**

*Sample*

* Describe all of the datasets I’m using in this project, how many observations in each one, and how many observations there are in the final dataset after all merges (mention that participants could go into the program multiple times as well)
  + SRT, RT, and CST Assessment Scores
  + Client Data
  + Risk Data for suicide and homicide risk
  + ASUS Data
    - Mention problems with missing data in this dataset that will require it to be analyzed separately from the other variables of interest
  + Criminal thinking profile
  + Episode data – provides information on what program they were in and for how long

*Measures*

* Describe all of the variables that I’m using in analysis, provide basic information on each one, some brief univariate statistics

*Detailed Methodology*

* Multiple methods needed to be used because of the complexity of the merged datasets and the sheer number of variables
  + Multinomial regressions and moderation analysis – briefly discuss interest in understanding specific factors that were of interest to The Connection
    - Mention programs with this method including the lack of inclusion of all variables of interest at the same time, doesn’t create a full picture
    - ASUS variables can be run in this section as well (since they don’t work in the machine learning model), but need to include a note on how much missing data they have
      * Note to self: might be better to just remove these measures since they’re mostly missing – need to think about this more
    - Combine this with machine learning stuff to do regressions based on what’s most important (maybe?)
  + Survival analysis – briefly discuss interest in understanding how length of stay impacted outcomes
    - Mention problems with this method including the required lengths of stay for some programs
  + Machine learning methods – used to attempt to better understand all variables that are important for analysis and to narrow down which ones are the most important utilizing lasso, gradient boosting, and random forest methods
    - Discuss which variables were able to be ruled out through other investigations, which variables were of focus with these techniques, which techniques are being prioritized and why
    - Discuss correlation matrices and how those were used to reduce dimensions on ASUS data

**Results**

*Multinomial/Interaction analysis results*

* Brief description of what I found through this
  + Discuss interesting findings with ORAS risk level, suicide risk (especially its moderation), and program
  + Include a few basic visualizations (moderation graph for suicide risk and ORAS risk level, maybe another graph with program and outcomes)
  + Create and include at least one multinomial visualization

*Survival analysis*

* Brief description of what I found here – mostly the result is that there is a significant result, but only after stratification and only for highly supervised programs, so this should be a really short part
  + Include the graphs and the assumptions that were tested
  + Think about looking into the AFT model instead – need to verify that it works with clustered standard errors as well and that it would even make sense to do this

*Machine learning methods*

* Main focus of results – discuss what each method found and which variables turned out to be most important
  + Depending on these results, maybe take a closer look into the variables that are most important in a multinomial model (just to see how it goes)
  + Create some graphs (variable importance graph from forest, some other things, maybe there are some nice ones that can be made)

**Discussion**

* Talk about the results, why they’re important, what they’re telling us
* Link findings back to literature
* Talk about future research

**Conclusion**

* Summarize the whole thing

**References**

* What it says

**Appendix**

* Alluvial graph of living situations (I know it’s not really necessary, but I like the graph and want to include it somewhere)
* Other visualizations that are cool and part of exploring but shouldn’t really be in the main paper

**New idea:** using the machine learning methods to inform logistic/multinomial regressions so that we can better understand which variables are the most important and then apply them to understand how they’re influencing the model

* Multinomial regression is identifying risk level as significant, but random forest doesn’t think they’re as important – variables that it believes are important are not significant (not sure what to do with that finding)

**Story of poster:**

* Looking into how variables of interest vary by program
  + Alluvial facetted by program to show where people are coming in from and where they end up
  + Interaction graph with programs and risk levels and outcomes---providing information on which programs work best for those with higher risk levels and which need to be more careful and provide more resources for those at greater risk
  + Survival analysis curve with programs---shows how people are lasting longer in certain programs with less support but are leaving earlier from the high support programs (maybe just because of how they’re structured, but it could also be something about those populations)
    - Maybe investigate more what the specific outcomes are for these programs – tie in to how many successful outcomes are in each one