The preliminary results is to show the primary analysis of what will be your final paper. There is a section on inference between variables and another section of comparing the performance of three different types of models.

Format [20 points]

Breakdown

	Points	Desc.	Points	Desc.
$\leq$ 8 pages	10	yes	0	no
Section headers	6	2 points each		
Typos/grammar errors	4	none to some	0	many

Provide section headers for the following:

Inference [20 points]

First, we want to establish the relationships between important covariates and the response variable. Here you will use LASSO to select variables (transformed and/or interacted as you see fit) using cross-validation to determine the best hyperparameter on the training data set. Then you will run the selected variables through either OLS or logistic regression (binomial/multinomial), whichever applies to your problem. Finally, you will interpret the top three most statistically significant coefficients.

Show code only for the cross-validation LASSO and the best λ. Show code and output for the summary of the final selected model. The output should contain coefficient estimates, standard errors, t or z statistics, and the associated p-values.

Breakdown

	Points	Desc.
Cross-validation	5	showed code
Best λ	4	showed best
Final model code/output	5	2 points for code, 3 points for output
Interpretation	6	2 points each

Prediction [50 points]

The next step is to compare how the model selected in the Inference section performs against two other models of your choosing. Choose two from the following:

Classification	Regression
Boosted Random Forest	Boosted Random Forest
GAM (spline/local/polynomial)	GAM (spline/local/polynomial)
Neural Network	Neural Network
KNN	KNN
SVM	PCR
LDA/QDA	

Train each model on the training data and display the accuracy of each model using either MSE (regression) or the error rate (classification). Show the code and/or output for each part as described in the point breakdown table below.

Describe the two modelling choices made and discuss model output.

BREAKDOWN

	Points	Desc.
Inference model accuracy	8	showed code and output
Chosen model 1 training	8	showed code
Chosen model 1 test accuracy	8	showed code and output
Chosen model 2 training	8	showed code
Chosen model 2 test accuracy	8	showed code and output
Descriptions/discussions	10	not a wall of only code and output

Comparison [10 points]

In one paragraph, compare the interpretability and flexibility of the models, compare the performance of the models, and tell us which model was the most accurate.