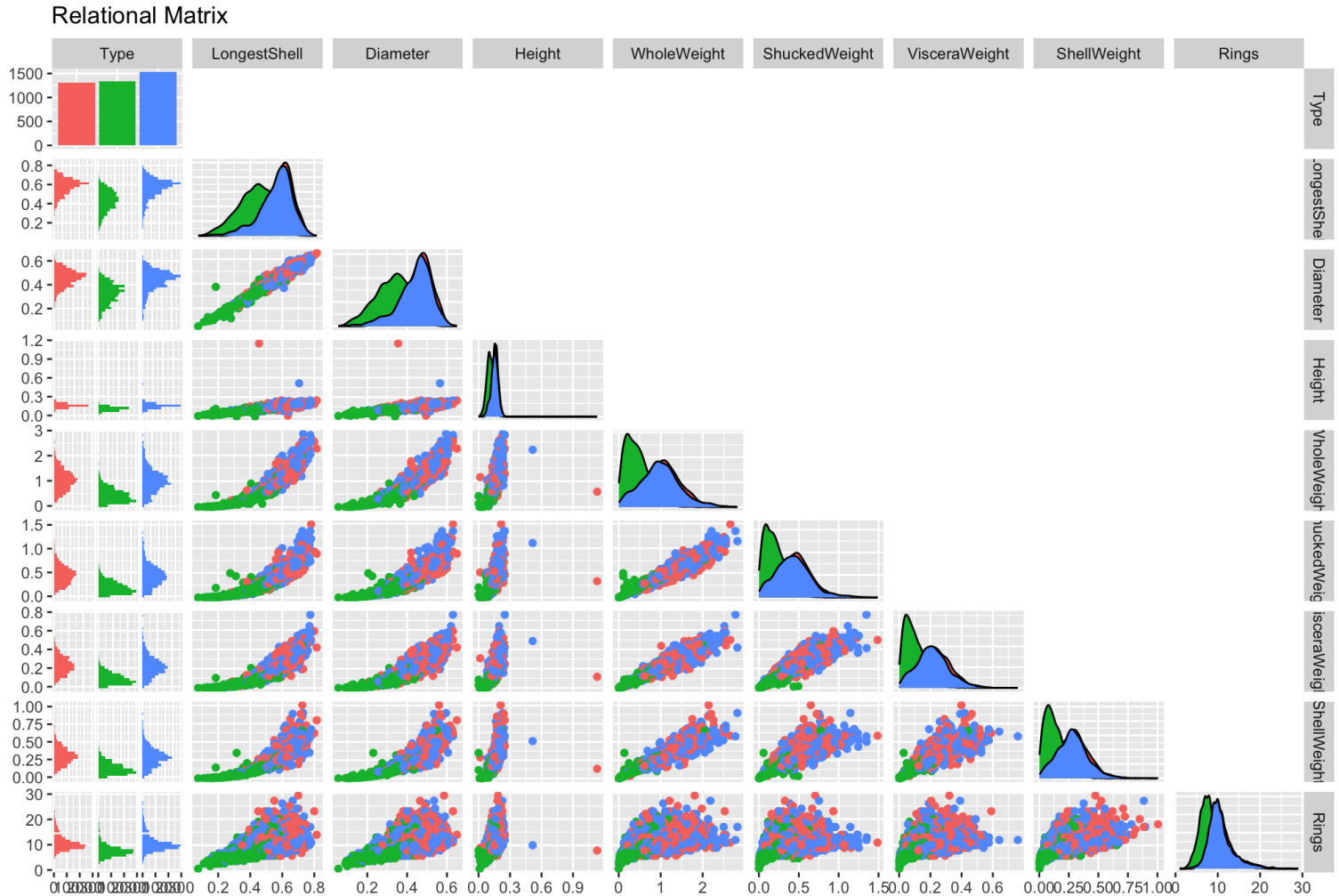
Streamlining Abalone Age Determination with Predictive Models

Hongyu Chen

STAT844, Spring 2024

Exploratory Data Analysis

	LongestShell	Diameter	Height	WholeWeight	ShuckedWeight	VisceraWeight	ShellWeight	Rings
Min	0.07	0.06	0.00	0.00	0.00	0.00	0.00	1.00
Mean	0.52	0.41	0.14	0.83	0.36	0.18	0.24	9.93
Max	0.81	0.65	1.13	2.83	1.49	0.76	1.00	29.00



Methods / Results

Model / Method	Details	RMSE (Train)	RSME (Test)
Linear Model		2.213922	2.229341
Linear Model	AIC & BIC	2.213955	2.228616
PCA	Top 3 Principal Components	2.619	4.508
Top Linear Model	Fit All, (GCV)	2.213955	2.228616
Linear Model	Remove high VIF	2.262283	2.286281
Ridge Regression	Cross-validation, Min_{λ}	2.305265	2.312115
Lasso Regression	Cross-validation, Min_{λ}	2.214153	2.230656
Elastic Net	Linear Model	2.214103	2.230499
Top Additive Model	Fit All	2.063426	2.16213
Polynomial Regression with degree of 4	AIC	2.074328	2.193002
Support Vector Machine	Radial Basis Function	2.08917	2.197734
Random Forest	5000 Trees	1.722188	2.248034
Regression Tree	With Pruning	2.272167	2.42185
Neural Network	4x2 Hidden Layers	0.07	0.169