Can’t use regsubsets 🡪 only for linear models

CV is better than AIC

CV shouldn’t be used to pick the number of predictors in the best model 🡪 should just be evaluating the performance of variable selection

Bestglm function is for best subset selection

**Flowchart**:

1) Variable selection methods 🡪 best subset, forward stepwise, backward stepwise

* Have to find a different function besides regsubsets
* If can’t find a different function, use lasso and ridge regression 🡪 use lasso because predictors aren’t correlated (none more than 0.4) 🡪 reference correlation matrix that Anna made
* Could do ridge

2) Compare the test errors for each of the variable selection methods (or lasso vs. ridge) using CV

* Set a random seed
* Do different cross validation methods 🡪 LOOCV, k-fold CV (k = 5 and k = 10), validation set approach

3) Pick the “best” model (the one that minimizes test error) 🡪 does this produce a logistic regression model already? Or is this just telling us which predictors go into the “best” model

4) Perform logistic regression, LDA, and QDA, and then compare the test errors for each of these three methods using k-fold CV (k = 10)