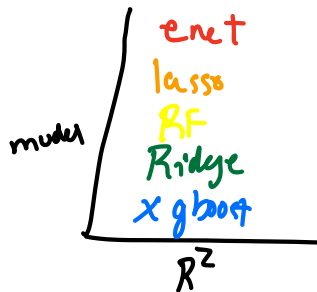


# Testing $R^2$ metrics

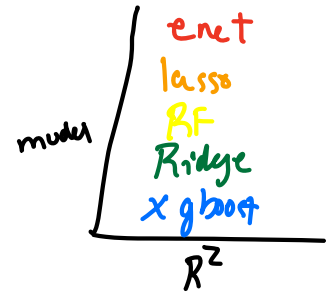
Genus (all)



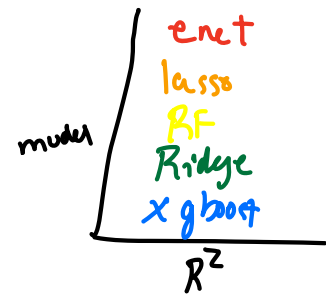
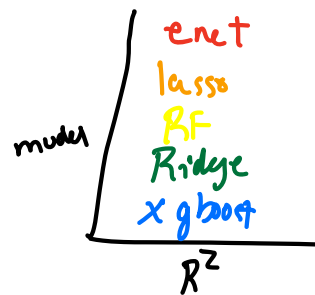
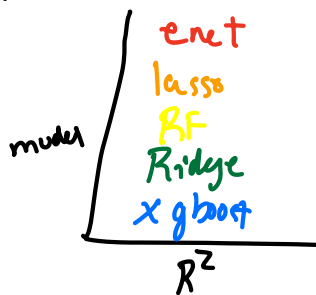
(Redundant Removed)



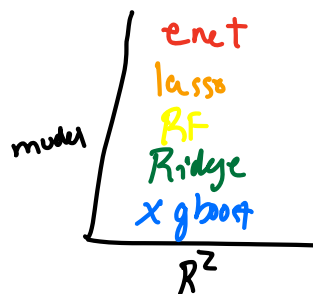
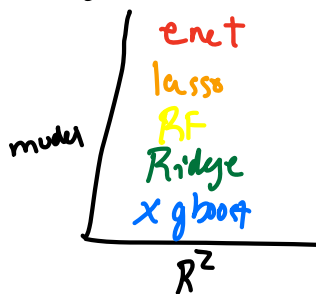
(Removed Clinical)



Species

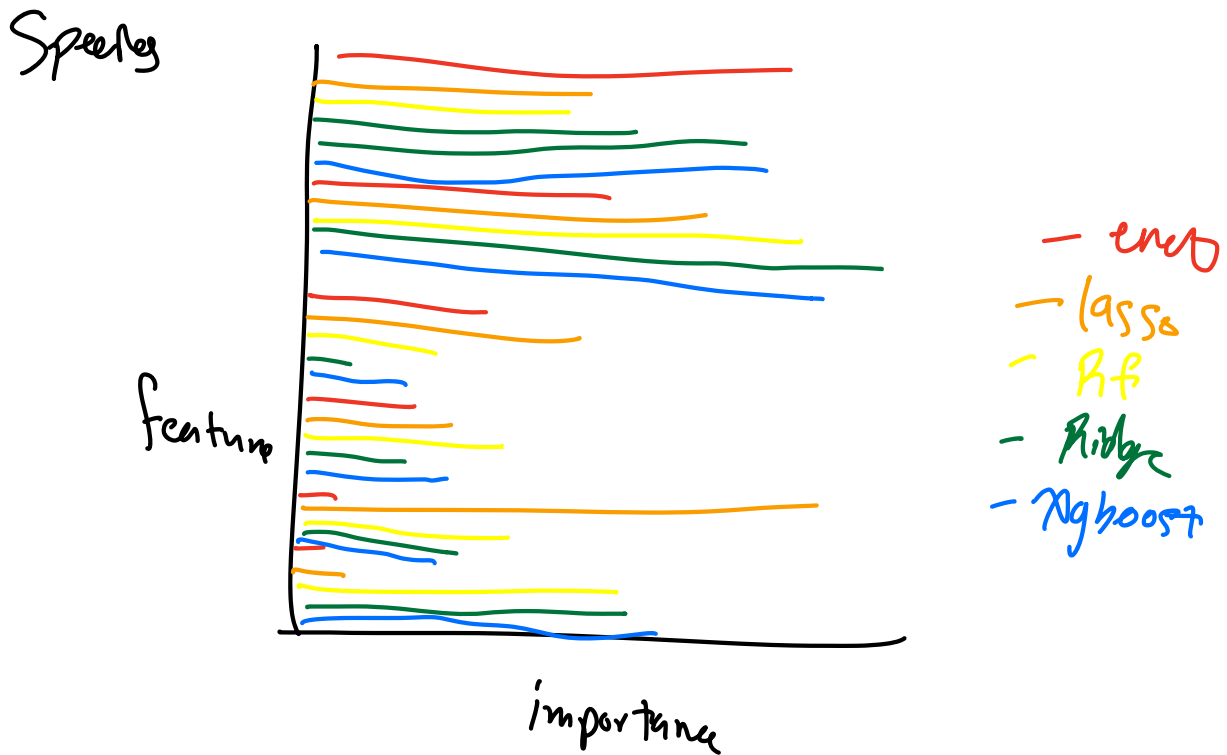
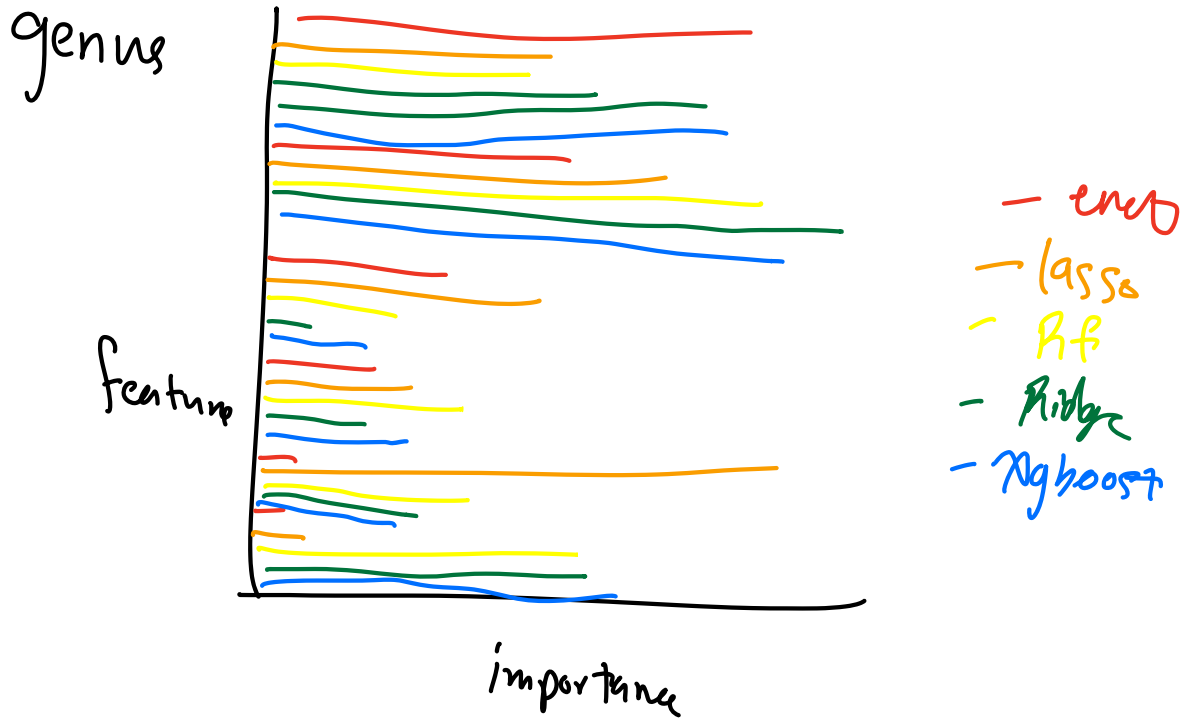


Pathways



$\Rightarrow$  isolate the best performing models based on dataset  
(Highest total  $R^2$  for that dataset)

# Top 20 features

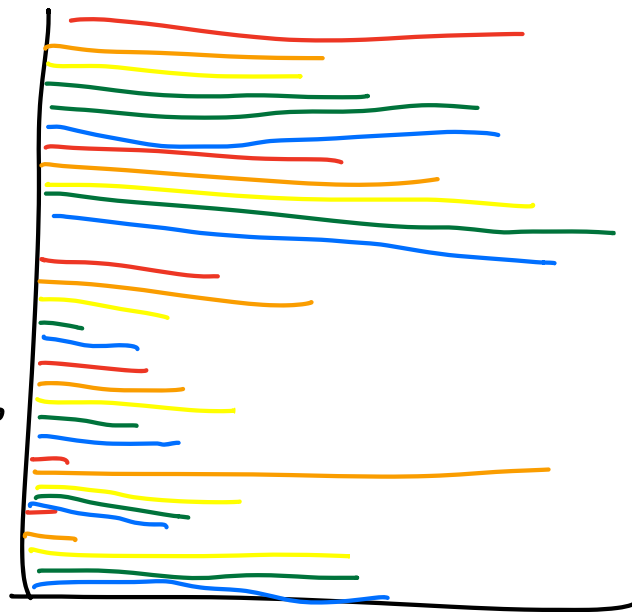


Pathway

feature

importance

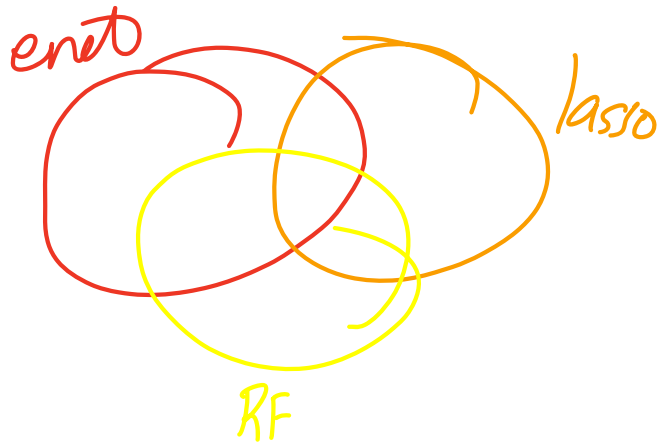
- enet
- lasso
- Rf
- Ridge
- Xgboost



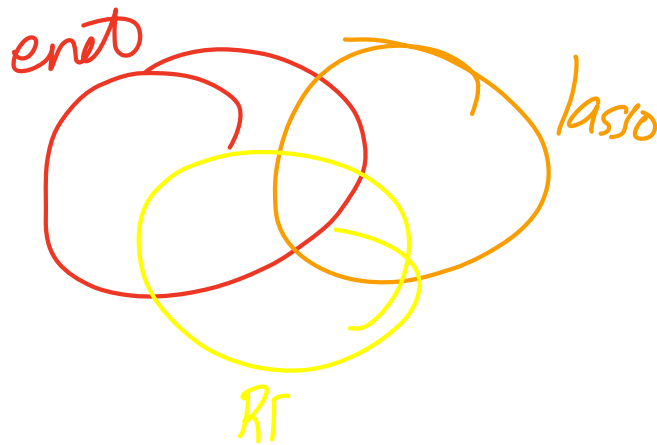
# Venn diagrams of feature

(Show 3 for simplicity)

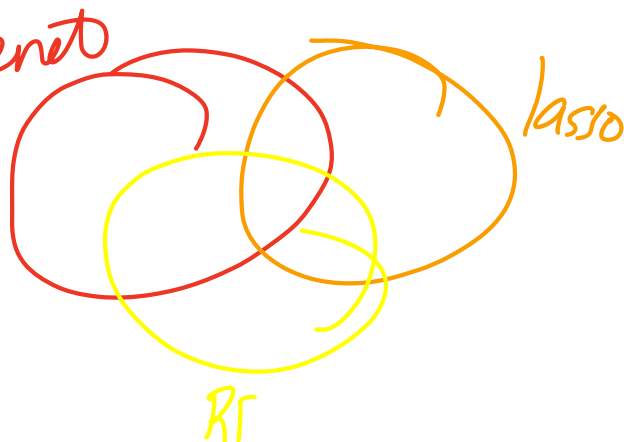
genus



Species



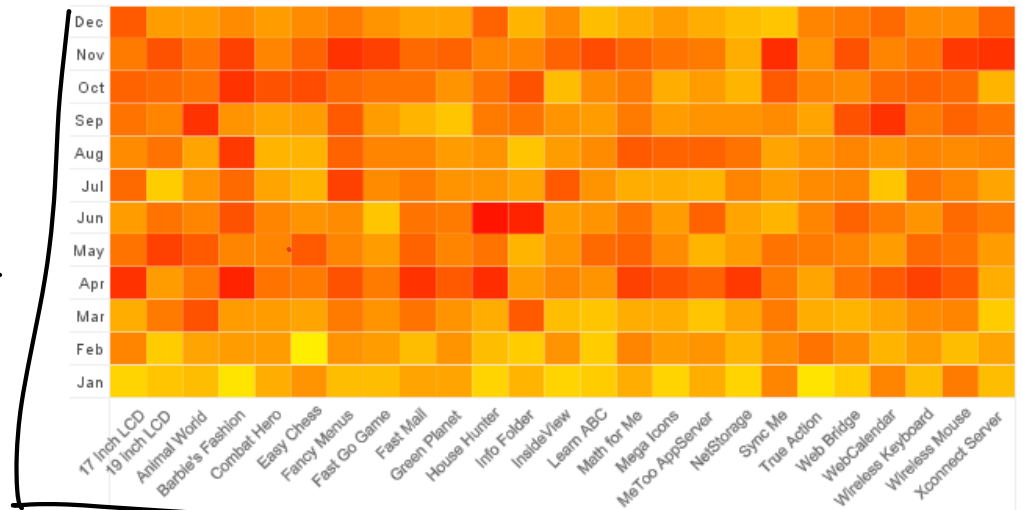
Pathway



# 20 Feature Correlation map

genus

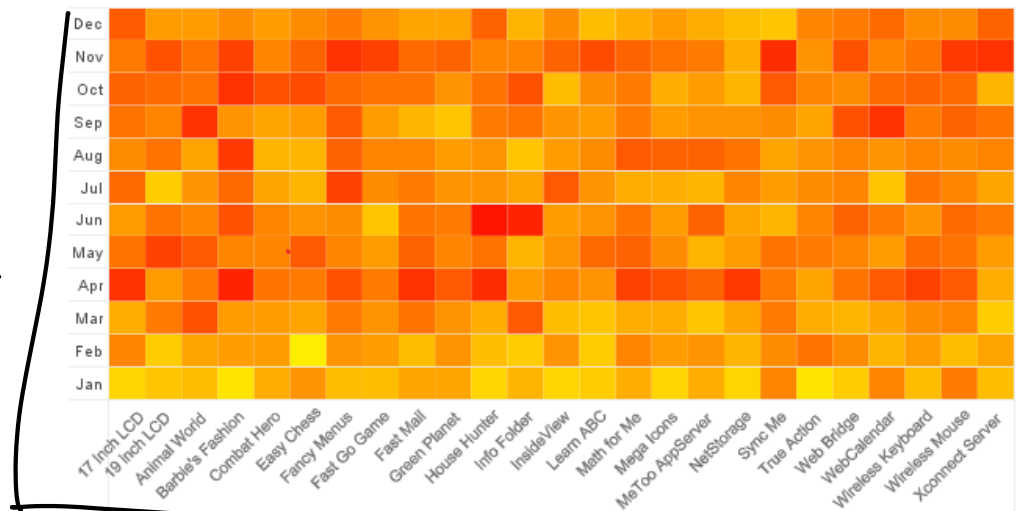
Ridge



lasso

Species

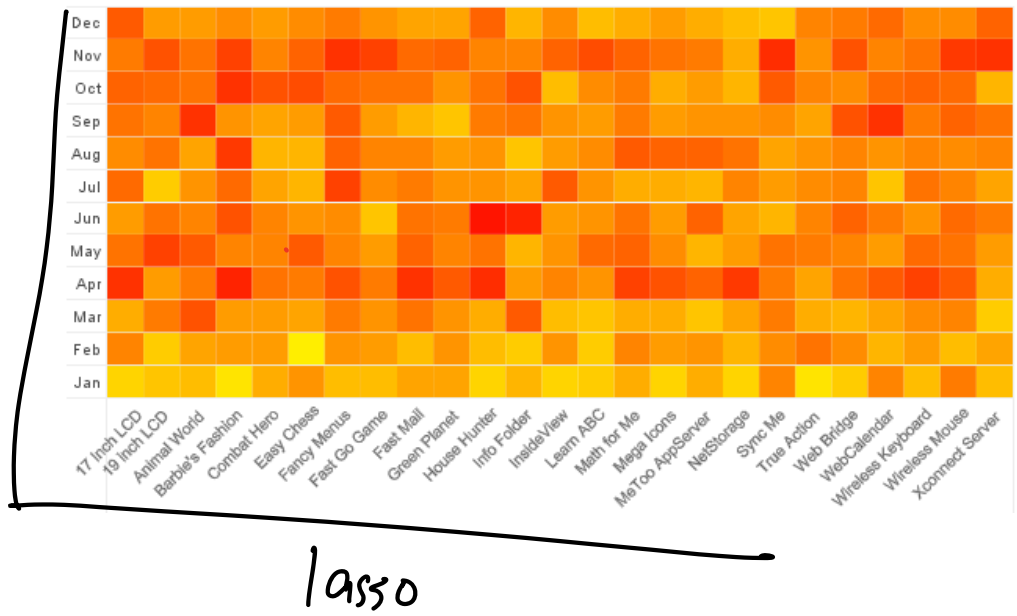
Ridge



lasso

pathway

# Ridge



link Gut microbiota  
Composition to weight  
discrepancy