```
 rsquare <- function(given, predicted) {
    sse <- sum((predicted - given)^2)
    sst <- sum(given^2)
    rsq <- 1 - sse / sst

    # For this post, impose floor...
    if (rsq < 0) rsq <- 0

    return (rsq)
  }
msd<-function(given,predicted) {
sqrt(mean((given-predicted)^2))
}
 mtcarshp<-mtcars[,c("mpg","wt","drat","hp")]

 ntrain<-round(0.8*nrow(mtcarshp))
 ntest<-nrow(mtcarshp)-ntrain
 allidx<-1:nrow(mtcarshp)
 trainidx<-sample(allidx,ntrain,rep=FALSE)
 testidx<-allidx[-trainidx]

 traindata<-mtcarshp[trainidx,]
 testdata<-mtcarshp[testidx,]

lm.model<-lm(hp~mpg+wt+drat,data=traindata)
traindata$train.predicted.hp<-predict(lm.model,traindata[,c("mpg","wt","drat")])
#training error
train_error<-rsquare(traindata$hp,traindata$train.predicted.hp)
train_msd<-sqrt(mean((traindata$hp-traindata$train.predicted.hp)^2))
#test error for lm
testdata$test.predicted.hp<-predict(lm.model,testdata[,c("mpg","wt","drat")])
test_error<-rsquare(testdata$hp,testdata$test.predicted.hp)
test_msd<-sqrt(mean((testdata$hp-testdata$test.predicted.hp)^2))
test_error/ntest

model_stats<-data.frame(#run_names=c("proportion","rsquare","msd"),
lm.train=c(proportion=1,train_error,train_msd),
lm.test=c(proportion=1,test_error,test_msd),stringsAsFactors=F)
model_stats

#require(magrittr)
#require(purrr)
#require(broom)
#require(tidyverse)
require(ggplot2)
require(glmnet)
y<-traindata$hp

x<-as.matrix(traindata[,c("mpg","wt","drat")])
lambdas<-10^seq(3,-2,by=-0.1)
glm.fit<-glmnet(x,y,alpha=0,lambda=lambdas)
all_coef<-coef(glm.fit)
betas<-all_coef[2:4,]
fitval<-x%*%betas

cv.glm.fit<-cv.glmnet(x,y,alpha=0,lambda=lambdas,nfolds=5)

cv.glm.fit$lambda.min

traindata$train.penalized.hp<-predict(glm.fit,newx=as.matrix(traindata[,c("mpg","wt"
,"drat")]),s=cv.glm.fit$lambda.min,type='response')
```

```
model_stats<-cbind(model_stats,train.penalized.hp=c(proportion=1,rsquare=rsquare(tra
indata$hp,traindata$train.penalized.hp),msd=msd(traindata$hp,traindata$train.penaliz
ed.hp)))

#model_stats<-rbind(model_stats,c("train.penalized.hp",rsquare(traindata$hp,traindat
a$train.penalized.hp),
#+ msd(traindata$hp,traindata$train.penalized.hp))
proportion<-c(0,0,0,1) #c(NA,NA,NA,1)

traindata$tr.p.hp2<-predict(glm.fit,newx=as.matrix(traindata[,c("mpg","wt","drat")])
,s=0.98*cv.glm.fit$lambda.min,type='response')
sum((traindata$tr.p.hp2-traindata$hp)^2)
model_stats<-cbind(model_stats,tr.p.hp2=c(0.98,rsquare=rsquare(traindata$hp,traindat
a$tr.p.hp2),msd=msd(traindata$hp,traindata$tr.p.hp2)))
proportion<-c(proportion,0.98)
traindata$tr.p.hp3<-predict(glm.fit,newx=as.matrix(traindata[,c("mpg","wt","drat")])
,s=0.88*cv.glm.fit$lambda.min,type='response')
sum((traindata$tr.p.hp3-traindata$hp)^2)
model_stats<-cbind(model_stats,tr.p.hp3=c(0.88,rsquare=rsquare(traindata$hp,traindat
a$tr.p.hp3),msd=msd(traindata$hp,traindata$tr.p.hp3)))
proportion<-c(proportion,0.88)

traindata$tr.p.hp4<-predict(glm.fit,newx=as.matrix(traindata[,c("mpg","wt","drat")])
,s=0.80*cv.glm.fit$lambda.min,type='response')
model_stats<-cbind(model_stats,tr.p.hp4=c(0.80,rsquare=rsquare(traindata$hp,traindat
a$tr.p.hp4),msd=msd(traindata$hp,traindata$tr.p.hp4)))
proportion<-c(proportion,0.80)
sum((traindata$tr.p.hp4-traindata$hp)^2)

traindata$tr.p.hp5<-predict(glm.fit,newx=as.matrix(traindata[,c("mpg","wt","drat")])
,s=0.78*cv.glm.fit$lambda.min,type='response')
model_stats<-cbind(model_stats,tr.p.hp5=c(0.78,rsquare=rsquare(traindata$hp,traindat
a$tr.p.hp5),msd=msd(traindata$hp,traindata$tr.p.hp5)))
proportion<-c(proportion,0.78)
sum((traindata$tr.p.hp5-traindata$hp)^2)

traindata$tr.p.hp6<-predict(glm.fit,newx=as.matrix(traindata[,c("mpg","wt","drat")])
,s=0.76*cv.glm.fit$lambda.min,type='response')
model_stats<-cbind(model_stats,tr.p.hp6=c(0.76,rsquare=rsquare(traindata$hp,traindat
a$tr.p.hp6),msd=msd(traindata$hp,traindata$tr.p.hp6)))
proportion<-c(proportion,0.76)
sum((traindata$tr.p.hp6-traindata$hp)^2)

traindata$tr.p.hp7<-predict(glm.fit,newx=as.matrix(traindata[,c("mpg","wt","drat")])
,s=0.66*cv.glm.fit$lambda.min,type='response')
model_stats<-cbind(model_stats,tr.p.hp7=c(0.66,rsquare=rsquare(traindata$hp,traindat
a$tr.p.hp7),msd=msd(traindata$hp,traindata$tr.p.hp7)))
proportion<-c(proportion,0.66)
sum((traindata$tr.p.hp7-traindata$hp)^2)

traindata$tr.p.hp8<-predict(glm.fit,newx=as.matrix(traindata[,c("mpg","wt","drat")])
,s=0.50*cv.glm.fit$lambda.min,type='response')
model_stats<-cbind(model_stats,tr.p.hp8=c(0.50,rsquare=rsquare(traindata$hp,traindat
a$tr.p.hp8),msd=msd(traindata$hp,traindata$tr.p.hp8)))
proportion<-c(proportion,0.50)
sum((traindata$tr.p.hp8-traindata$hp)^2)

traindata$tr.p.hp9<-predict(glm.fit,newx=as.matrix(traindata[,c("mpg","wt","drat")])
,s=0.40*cv.glm.fit$lambda.min,type='response')
model_stats<-cbind(model_stats,tr.p.hp9=c(0.40,rsquare=rsquare(traindata$hp,traindat
a$tr.p.hp9),msd=msd(traindata$hp,traindata$tr.p.hp9)))
```

```
proportion<-c(proportion,0.40)
sum((traindata$tr.p.hp9-traindata$hp)^2)

traindata$tr.p.hp10<-predict(glm.fit,newx=as.matrix(traindata[,c("mpg","wt","drat")]
),s=0.20*cv.glm.fit$lambda.min,type='response')
model_stats<-cbind(model_stats,tr.p.hp10=c(0.20,rsquare=rsquare(traindata$hp,trainda
ta$tr.p.hp10),msd=msd(traindata$hp,traindata$tr.p.hp10)))
proportion<-c(proportion,0.20)
sum((traindata$tr.p.hp10-traindata$hp)^2)

traindata$tr.p.hp11<-predict(glm.fit,newx=as.matrix(traindata[,c("mpg","wt","drat")]
),s=0.10*cv.glm.fit$lambda.min,type='response')
model_stats<-cbind(model_stats,tr.p.hp11=c(0.10,rsquare=rsquare(traindata$hp,trainda
ta$tr.p.hp11),msd=msd(traindata$hp,traindata$tr.p.hp11)))
proportion<-c(proportion,0.10)
sum((traindata$tr.p.hp11-traindata$hp)^2)
cv.glm.fit$lambda.min



traindata$tr.p.hp12<-predict(glm.fit,newx=as.matrix(traindata[,c("mpg","wt","drat")]
),s=0.05*cv.glm.fit$lambda.min,type='response')
model_stats<-cbind(model_stats,tr.p.hp12=c(0.05,rsquare=rsquare(traindata$hp,trainda
ta$tr.p.hp12),msd=msd(traindata$hp,traindata$tr.p.hp12)))
proportion<-c(proportion,0.05)
sum((traindata$tr.p.hp12-traindata$hp)^2)

traindata$tr.p.hp13<-predict(glm.fit,newx=as.matrix(traindata[,c("mpg","wt","drat")]
),s=0*cv.glm.fit$lambda.min,type='response')
model_stats<-cbind(model_stats,tr.p.hp13=c(0.00,rsquare=rsquare(traindata$hp,trainda
ta$tr.p.hp13),msd=msd(traindata$hp,traindata$tr.p.hp13)))
proportion<-c(proportion,0.0)
sum((traindata$tr.p.hp13-traindata$hp)^2)
traindata$tr.p.hp14<-predict(glm.fit,newx=as.matrix(traindata[,c("mpg","wt","drat")]
),s=-0.05*cv.glm.fit$lambda.min,type='response')
model_stats<-cbind(model_stats,tr.p.hp14=c(-0.05,rsquare=rsquare(traindata$hp,traind
ata$tr.p.hp14),msd=msd(traindata$hp,traindata$tr.p.hp14)))
proportion<-c(proportion,-0.05)
sum((traindata$tr.p.hp14-traindata$hp)^2)
traindata$tr.p.hp15<-predict(glm.fit,newx=as.matrix(traindata[,c("mpg","wt","drat")]
),s=-0.8*cv.glm.fit$lambda.min,type='response')
model_stats<-cbind(model_stats,tr.p.hp15=c(-0.8,rsquare=rsquare(traindata$hp,trainda
ta$tr.p.hp15),msd=msd(traindata$hp,traindata$tr.p.hp15)))
proportion<-c(proportion,-0.80)
sum((traindata$tr.p.hp15-traindata$hp)^2)


traindata$tr.p.hp16<-predict(glm.fit,newx=as.matrix(traindata[,c("mpg","wt","drat")]
),s=-0.90*cv.glm.fit$lambda.min,type='response')
model_stats<-cbind(model_stats,tr.p.hp16=c(-0.9,rsquare=rsquare(traindata$hp,trainda
ta$tr.p.hp16),msd=msd(traindata$hp,traindata$tr.p.hp16)))
proportion<-c(proportion,-0.90)
sum((traindata$tr.p.hp16-traindata$hp)^2)

traindata$tr.p.hp17<-predict(glm.fit,newx=as.matrix(traindata[,c("mpg","wt","drat")]
),s=-1*cv.glm.fit$lambda.min,type='response')
model_stats<-cbind(model_stats,tr.p.hp17=c(-1,rsquare=rsquare(traindata$hp,traindata
$tr.p.hp17),msd=msd(traindata$hp,traindata$tr.p.hp17)))
proportion<-c(proportion,-1.0)
sum((traindata$tr.p.hp17-traindata$hp)^2)
```

```
traindata$tr.p.hp18<-predict(glm.fit,newx=as.matrix(traindata[,c("mpg","wt","drat")]
),s=-2*cv.glm.fit$lambda.min,type='response')
model_stats<-cbind(model_stats,tr.p.hp18=c(-2,rsquare=rsquare(traindata$hp,traindata
$tr.p.hp18),msd=msd(traindata$hp,traindata$tr.p.hp18)))
proportion<-c(proportion,-2.0)
sum((traindata$tr.p.hp18-traindata$hp)^2)


traindata$tr.p.hp19<-predict(glm.fit,newx=as.matrix(traindata[,c("mpg","wt","drat")]
),s=-3*cv.glm.fit$lambda.min,type='response')
model_stats<-cbind(model_stats,tr.p.hp19=c(-3,rsquare=rsquare(traindata$hp,traindata
$tr.p.hp19),msd=msd(traindata$hp,traindata$tr.p.hp19)))
proportion<-c(proportion,-3.0)
sum((traindata$tr.p.hp19-traindata$hp)^2)

traindata$tr.p.hp20<-predict(glm.fit,newx=as.matrix(traindata[,c("mpg","wt","drat")]
),s=-4*cv.glm.fit$lambda.min,type='response')
model_stats<-cbind(model_stats,tr.p.hp20=c(-4,rsquare=rsquare(traindata$hp,traindata
$tr.p.hp20),msd=msd(traindata$hp,traindata$tr.p.hp20)))
proportion<-c(proportion,-4.0)
sum((traindata$tr.p.hp20-traindata$hp)^2)

testdata$test.penalized.hp<-predict(glm.fit,newx=as.matrix(testdata[,c("mpg","wt","d
rat")]),s=cv.glm.fit$lambda.min,type='response')
model_stats<-cbind(model_stats,test.penalized.hp=c(1,rsquare=rsquare(testdata$hp,tes
tdata$test.penalized.hp),msd=msd(testdata$hp,testdata$test.penalized.hp)))

#model_stats<-rbind(model_stats,c("test.penalized.hp",rsquare(testdata$hp,testdata$t
est.penalized.hp),
#+ msd(testdata$hp,testdata$test.penalized.hp))
proportion<-c(NA,NA,NA,1)

testdata$tst.p.hp2<-predict(glm.fit,newx=as.matrix(testdata[,c("mpg","wt","drat")]),
s=0.98*cv.glm.fit$lambda.min,type='response')
sum((testdata$tst.p.hp2-testdata$hp)^2)
model_stats<-cbind(model_stats,tst.p.hp2=c(0.98,rsquare=rsquare(testdata$hp,testdata
$tst.p.hp2),msd=msd(testdata$hp,testdata$tst.p.hp2)))
proportion<-c(proportion,0.98)
testdata$tst.p.hp3<-predict(glm.fit,newx=as.matrix(testdata[,c("mpg","wt","drat")]),
s=0.88*cv.glm.fit$lambda.min,type='response')
sum((testdata$tst.p.hp3-testdata$hp)^2)
model_stats<-cbind(model_stats,tst.p.hp3=c(0.88,rsquare=rsquare(testdata$hp,testdata
$tst.p.hp3),msd=msd(testdata$hp,testdata$tst.p.hp3)))
proportion<-c(proportion,0.88)

testdata$tst.p.hp4<-predict(glm.fit,newx=as.matrix(testdata[,c("mpg","wt","drat")]),
s=0.80*cv.glm.fit$lambda.min,type='response')
model_stats<-cbind(model_stats,tst.p.hp4=c(0.80,rsquare=rsquare(testdata$hp,testdata
$tst.p.hp4),msd=msd(testdata$hp,testdata$tst.p.hp4)))
proportion<-c(proportion,0.80)
sum((testdata$tst.p.hp4-testdata$hp)^2)

testdata$tst.p.hp5<-predict(glm.fit,newx=as.matrix(testdata[,c("mpg","wt","drat")]),
s=0.78*cv.glm.fit$lambda.min,type='response')
model_stats<-cbind(model_stats,tst.p.hp5=c(0.78,rsquare=rsquare(testdata$hp,testdata
$tst.p.hp5),msd=msd(testdata$hp,testdata$tst.p.hp5)))
proportion<-c(proportion,0.78)
sum((testdata$tst.p.hp5-testdata$hp)^2)

testdata$tst.p.hp6<-predict(glm.fit,newx=as.matrix(testdata[,c("mpg","wt","drat")]),
s=0.76*cv.glm.fit$lambda.min,type='response')
```

```
model_stats<-cbind(model_stats,tst.p.hp6=c(0.76,rsquare=rsquare(testdata$hp,testdata
$tst.p.hp6),msd=msd(testdata$hp,testdata$tst.p.hp6)))
proportion<-c(proportion,0.76)
sum((testdata$tst.p.hp6-testdata$hp)^2)

testdata$tst.p.hp7<-predict(glm.fit,newx=as.matrix(testdata[,c("mpg","wt","drat")]),
s=0.66*cv.glm.fit$lambda.min,type='response')
model_stats<-cbind(model_stats,tst.p.hp7=c(0.66,rsquare=rsquare(testdata$hp,testdata
$tst.p.hp7),msd=msd(testdata$hp,testdata$tst.p.hp7)))
proportion<-c(proportion,0.66)
sum((testdata$tst.p.hp7-testdata$hp)^2)

testdata$tst.p.hp8<-predict(glm.fit,newx=as.matrix(testdata[,c("mpg","wt","drat")]),
s=0.50*cv.glm.fit$lambda.min,type='response')
model_stats<-cbind(model_stats,tst.p.hp8=c(0.50,rsquare=rsquare(testdata$hp,testdata
$tst.p.hp8),msd=msd(testdata$hp,testdata$tst.p.hp8)))
proportion<-c(proportion,0.50)
sum((testdata$tst.p.hp8-testdata$hp)^2)

testdata$tst.p.hp9<-predict(glm.fit,newx=as.matrix(testdata[,c("mpg","wt","drat")]),
s=0.40*cv.glm.fit$lambda.min,type='response')
model_stats<-cbind(model_stats,tst.p.hp9=c(0.40,rsquare=rsquare(testdata$hp,testdata
$tst.p.hp9),msd=msd(testdata$hp,testdata$tst.p.hp9)))
proportion<-c(proportion,0.40)
sum((testdata$tst.p.hp9-testdata$hp)^2)

testdata$tst.p.hp10<-predict(glm.fit,newx=as.matrix(testdata[,c("mpg","wt","drat")])
,s=0.20*cv.glm.fit$lambda.min,type='response')
model_stats<-cbind(model_stats,tst.p.hp10=c(0.20,rsquare=rsquare(testdata$hp,testdat
a$tst.p.hp10),msd=msd(testdata$hp,testdata$tst.p.hp10)))
proportion<-c(proportion,0.20)
sum((testdata$tst.p.hp10-testdata$hp)^2)

testdata$tst.p.hp11<-predict(glm.fit,newx=as.matrix(testdata[,c("mpg","wt","drat")])
,s=0.10*cv.glm.fit$lambda.min,type='response')
model_stats<-cbind(model_stats,tst.p.hp11=c(0.10,rsquare=rsquare(testdata$hp,testdat
a$tst.p.hp11),msd=msd(testdata$hp,testdata$tst.p.hp11)))
proportion<-c(proportion,0.10)
sum((testdata$tst.p.hp11-testdata$hp)^2)
cv.glm.fit$lambda.min



testdata$tst.p.hp12<-predict(glm.fit,newx=as.matrix(testdata[,c("mpg","wt","drat")])
,s=0.05*cv.glm.fit$lambda.min,type='response')
model_stats<-cbind(model_stats,tst.p.hp12=c(0.05,rsquare=rsquare(testdata$hp,testdat
a$tst.p.hp12),msd=msd(testdata$hp,testdata$tst.p.hp12)))
proportion<-c(proportion,0.05)
sum((testdata$tst.p.hp12-testdata$hp)^2)

testdata$tst.p.hp13<-predict(glm.fit,newx=as.matrix(testdata[,c("mpg","wt","drat")])
,s=0*cv.glm.fit$lambda.min,type='response')
model_stats<-cbind(model_stats,tst.p.hp13=c(0.00,rsquare=rsquare(testdata$hp,testdat
a$tst.p.hp13),msd=msd(testdata$hp,testdata$tst.p.hp13)))
proportion<-c(proportion,0.0)
sum((testdata$tst.p.hp13-testdata$hp)^2)
testdata$tst.p.hp14<-predict(glm.fit,newx=as.matrix(testdata[,c("mpg","wt","drat")])
,s=-0.05*cv.glm.fit$lambda.min,type='response')
model_stats<-cbind(model_stats,tst.p.hp14=c(-0.05,rsquare=rsquare(testdata$hp,testda
ta$tst.p.hp14),msd=msd(testdata$hp,testdata$tst.p.hp14)))
proportion<-c(proportion,-0.05)
```

```
sum((testdata$tst.p.hp14-testdata$hp)^2)
testdata$tst.p.hp15<-predict(glm.fit,newx=as.matrix(testdata[,c("mpg","wt","drat")])
,s=-0.8*cv.glm.fit$lambda.min,type='response')
model_stats<-cbind(model_stats,tst.p.hp15=c(-0.80,rsquare=rsquare(testdata$hp,testda
ta$tst.p.hp15),msd=msd(testdata$hp,testdata$tst.p.hp15)))
proportion<-c(proportion,-0.80)
sum((testdata$tst.p.hp15-testdata$hp)^2)


testdata$tst.p.hp16<-predict(glm.fit,newx=as.matrix(testdata[,c("mpg","wt","drat")])
,s=-0.90*cv.glm.fit$lambda.min,type='response')
model_stats<-cbind(model_stats,tst.p.hp16=c(-0.90,rsquare=rsquare(testdata$hp,testda
ta$tst.p.hp16),msd=msd(testdata$hp,testdata$tst.p.hp16)))
proportion<-c(proportion,-0.90)
sum((testdata$tst.p.hp16-testdata$hp)^2)

testdata$tst.p.hp17<-predict(glm.fit,newx=as.matrix(testdata[,c("mpg","wt","drat")])
,s=-1*cv.glm.fit$lambda.min,type='response')
model_stats<-cbind(model_stats,tst.p.hp17=c(-1.0,rsquare=rsquare(testdata$hp,testdat
a$tst.p.hp17),msd=msd(testdata$hp,testdata$tst.p.hp17)))
proportion<-c(proportion,-1.0)
sum((testdata$tst.p.hp17-testdata$hp)^2)

testdata$tst.p.hp18<-predict(glm.fit,newx=as.matrix(testdata[,c("mpg","wt","drat")])
,s=-2*cv.glm.fit$lambda.min,type='response')
model_stats<-cbind(model_stats,tst.p.hp18=c(-2.0,rsquare=rsquare(testdata$hp,testdat
a$tst.p.hp18),msd=msd(testdata$hp,testdata$tst.p.hp18)))
proportion<-c(proportion,-2.0)
sum((testdata$tst.p.hp18-testdata$hp)^2)


testdata$tst.p.hp19<-predict(glm.fit,newx=as.matrix(testdata[,c("mpg","wt","drat")])
,s=-3*cv.glm.fit$lambda.min,type='response')
model_stats<-cbind(model_stats,tst.p.hp19=c(-3.0,rsquare=rsquare(testdata$hp,testdat
a$tst.p.hp19),msd=msd(testdata$hp,testdata$tst.p.hp19)))
proportion<-c(proportion,-3.0)
sum((testdata$tst.p.hp19-testdata$hp)^2)

testdata$tst.p.hp20<-predict(glm.fit,newx=as.matrix(testdata[,c("mpg","wt","drat")])
,s=-4*cv.glm.fit$lambda.min,type='response')
model_stats<-cbind(model_stats,tst.p.hp20=c(-4.0,rsquare=rsquare(testdata$hp,testdat
a$tst.p.hp20),msd=msd(testdata$hp,testdata$tst.p.hp20)))
proportion<-c(proportion,-4.0)
sum((testdata$tst.p.hp20-testdata$hp)^2)

#model_stats<-rbind(model_stats,proportion=proportion)
ggplot(as.data.frame(t(model_stats[,20])),aes(x=proportion,y=rsquare))+geom_point()
```