Lab10a_Philip_Nancy_Neil

Philip Lee (phl6cw) Nancy Zhou (nz2ay) Neil Mathur (nkm7da)

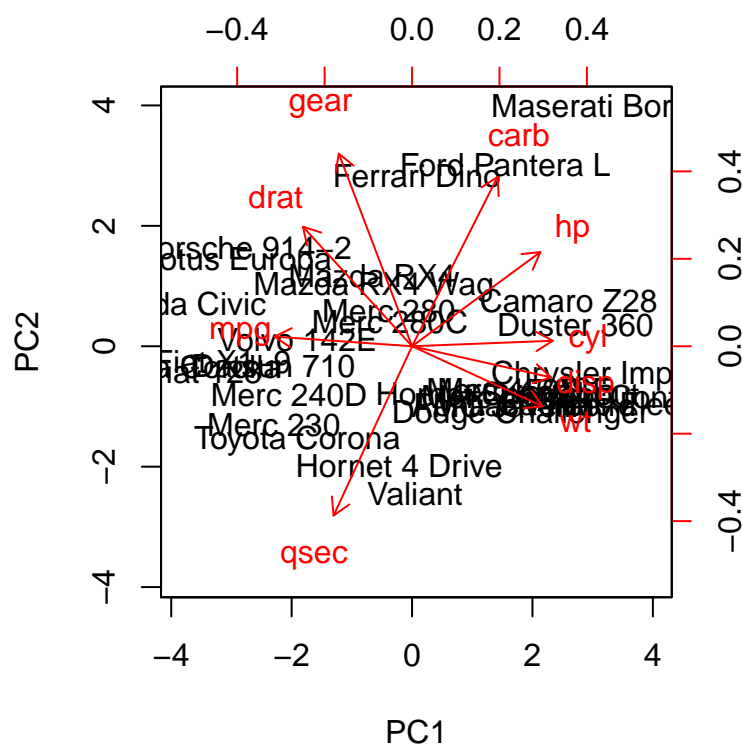*11/5/2019*

##1

Engine and transmission should be removed as they are binary variable

##2

```
pr.out<-prcomp(newdata, scale=TRUE)
biplot(pr.out,scale=0)
```


##3

```
pr.out$rotation[,c(1:2)]
```

```
##               PC1          PC2
## mpg   -0.3931477   0.02753861
## cyl    0.4025537   0.01570975
## disp   0.3973528  -0.08888469
## hp     0.3670814   0.26941371
## drat  -0.3118165   0.34165268
## wt     0.3734771  -0.17194306
## qsec  -0.2243508  -0.48404435
## gear  -0.2094749   0.55078264
```

```
## carb   0.2445807   0.48431310
```

4

mpg, cyl, disp, hp, and wt are the features that have high loadings for the first PC.

Increasing the values for number of cylinders, displacement, gross horsepower, and weight will produce lower values of miles per gallon. If interpretted contextually, the lower the values for these features, the less "sporty" a car is.

5

qsec, gear, and carb are the features that have high loadings for the second PC.

Higher values for number of forward gears and carburetors result in lower value of quarter mile time.

6

Maserati Bora has much higher than average values for number of cylinders, displacement, gross horsepower, weight, number of forward gears and carburetors while having much lower than average value for miles per gallon and quarter mile time. This car is a very sporty car in context.

7

```r
pr.var<-pr.out$sdev^2
pve<-pr.var/sum(pr.var)
par(mfrow=c(1,2))
plot(pve, xlab="Principal Component", ylab="Proportion of Variance Explained",
     main="Scree Plot", ylim=c(0,1),type='b')
```
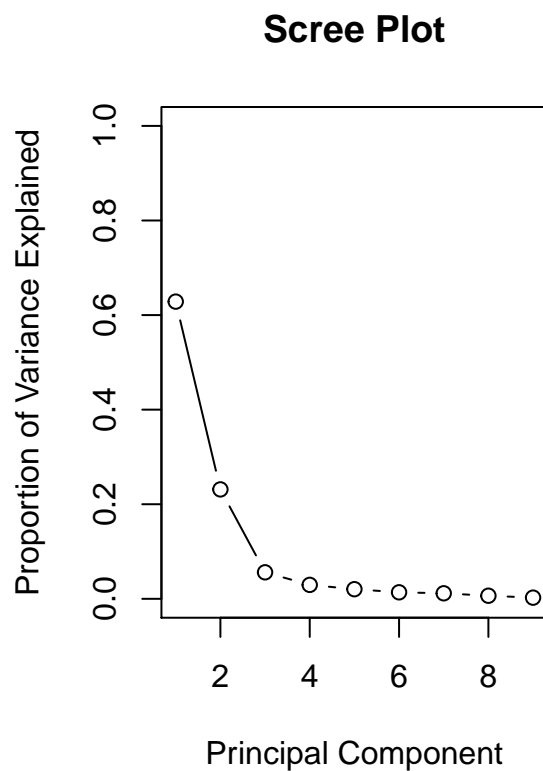

Based on the scree plot, we would be using 3 PCs as it seems like PVE stops being large after third PC.