Hw15

20150056 국윤범

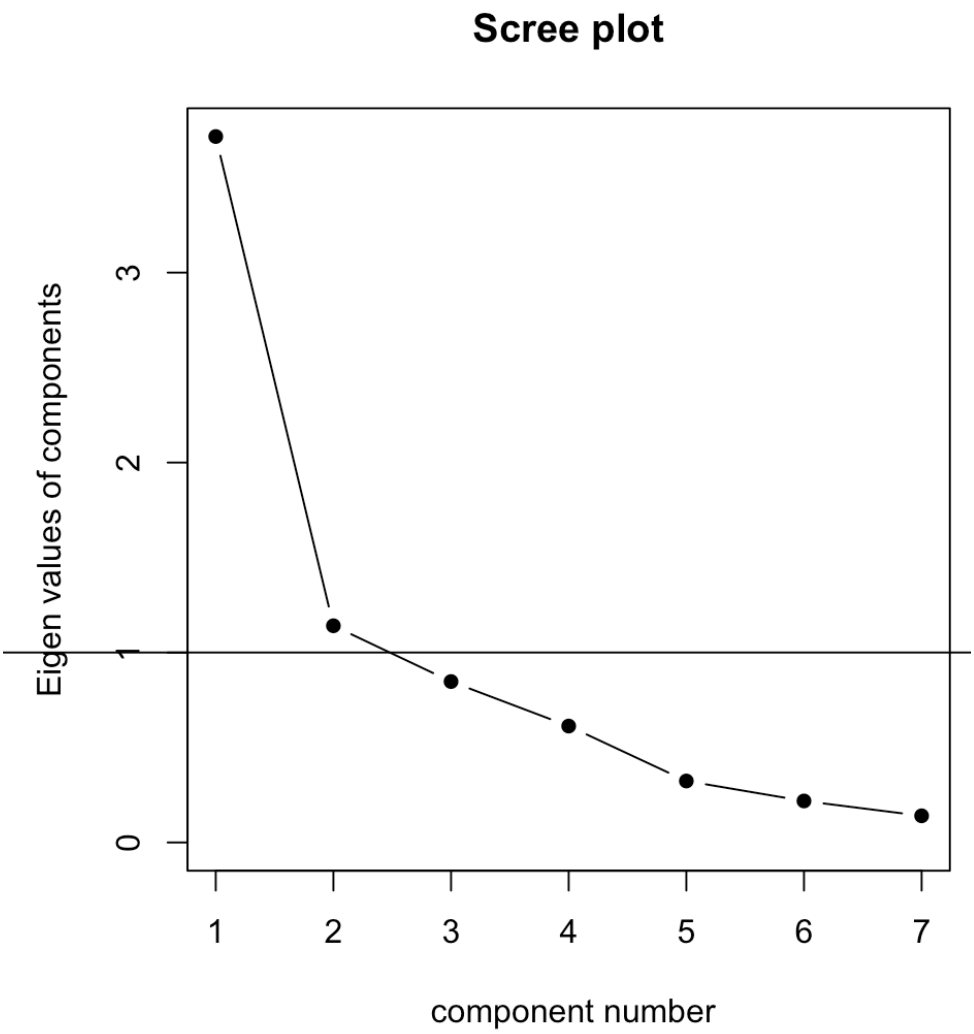
Problem 1

(a)

```
> paf_dat$KMO
[1] 0.76328
> paf_dat$Bartlett
[1] 98.753
> pcacor = cor(USArrests)
> cortest.bartlett(pcacor, n=186)
$chisq
[1] 344.67

$p.value
[1] 2.1515e-71

$df
[1] 6

> det(pcacor)
[1] 0.15181
```

From above, KMO > 0.7, p-value < 0.05, and the determinant of covariance matrix is positive. Hence, the given dataset is suitable for PCA.

(b)



Hence, 2 principal components have eigenvalues equal or greater than 1.0.

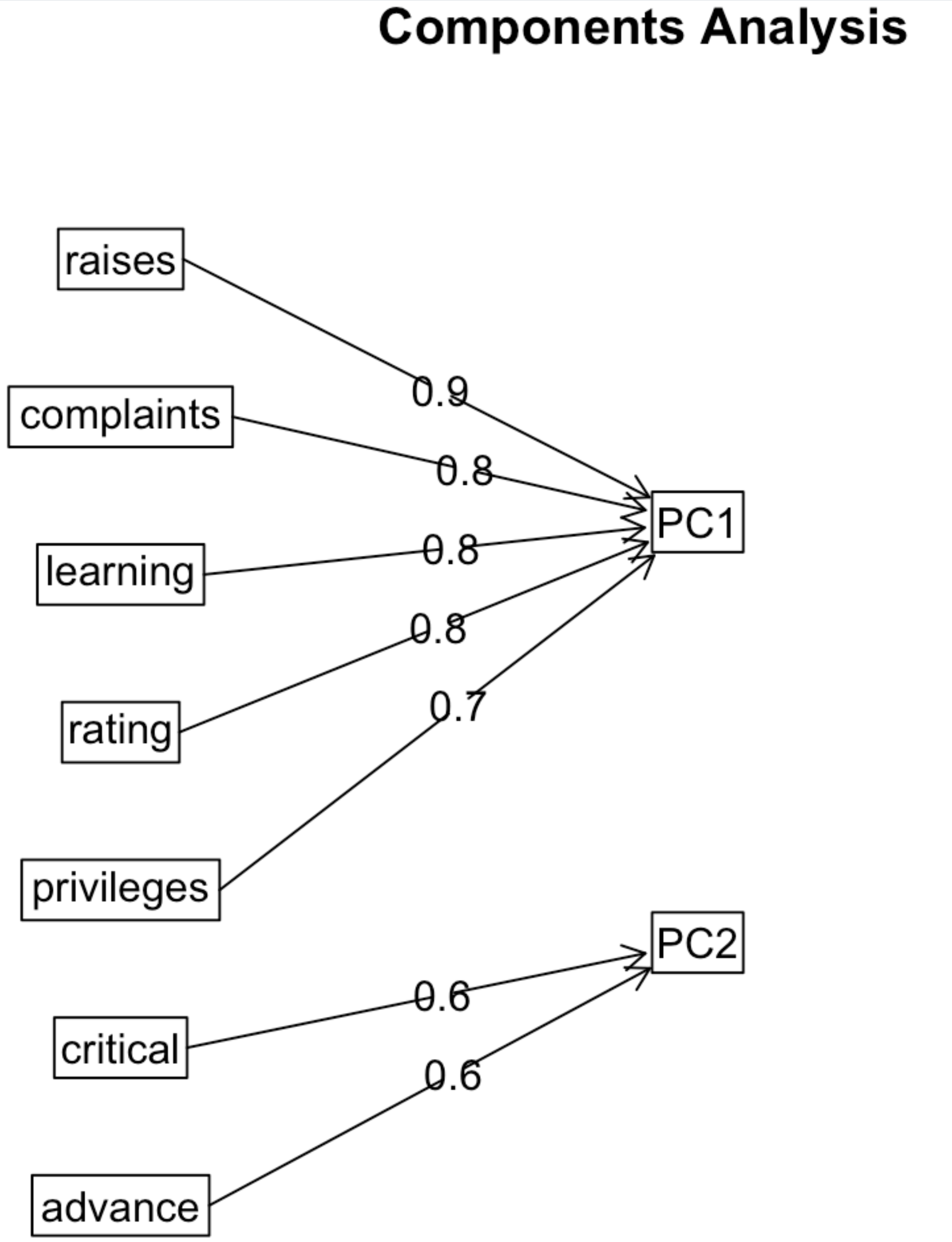
(c)

```
Principal Components Analysis
Call: principal(r = attitude, nfactors = 2, rotate = "none")
Standardized loadings (pattern matrix) based upon correlation matrix
            PC1   PC2   h2   u2 com
rating     0.80 -0.42 0.81 0.19 1.5
complaints 0.85 -0.36 0.85 0.15 1.3
privileges 0.68 -0.10 0.48 0.52 1.0
learning   0.83 -0.05 0.68 0.32 1.0
raises     0.86  0.19 0.78 0.22 1.1
critical   0.36  0.64 0.54 0.46 1.6
advance    0.58  0.61 0.71 0.29 2.0

                       PC1  PC2
SS loadings           3.72 1.14
Proportion Var        0.53 0.16
Cumulative Var        0.53 0.69
Proportion Explained  0.77 0.23
Cumulative Proportion 0.77 1.00

Mean item complexity =  1.4
Test of the hypothesis that 2 components are sufficient.

The root mean square of the residuals (RMSR) is  0.11
 with the empirical chi square  14.83  with prob <  0.063

Fit based upon off diagonal values = 0.95
```

We should select variable whose coefficient is equal or greater than 0.5. Then, the principal component equations to generate the scores is

Equation = 0.8*rating + 0.85*complaints + 0.68*privileges + 0.83*learning + 0.86*raises + 0.58*advance

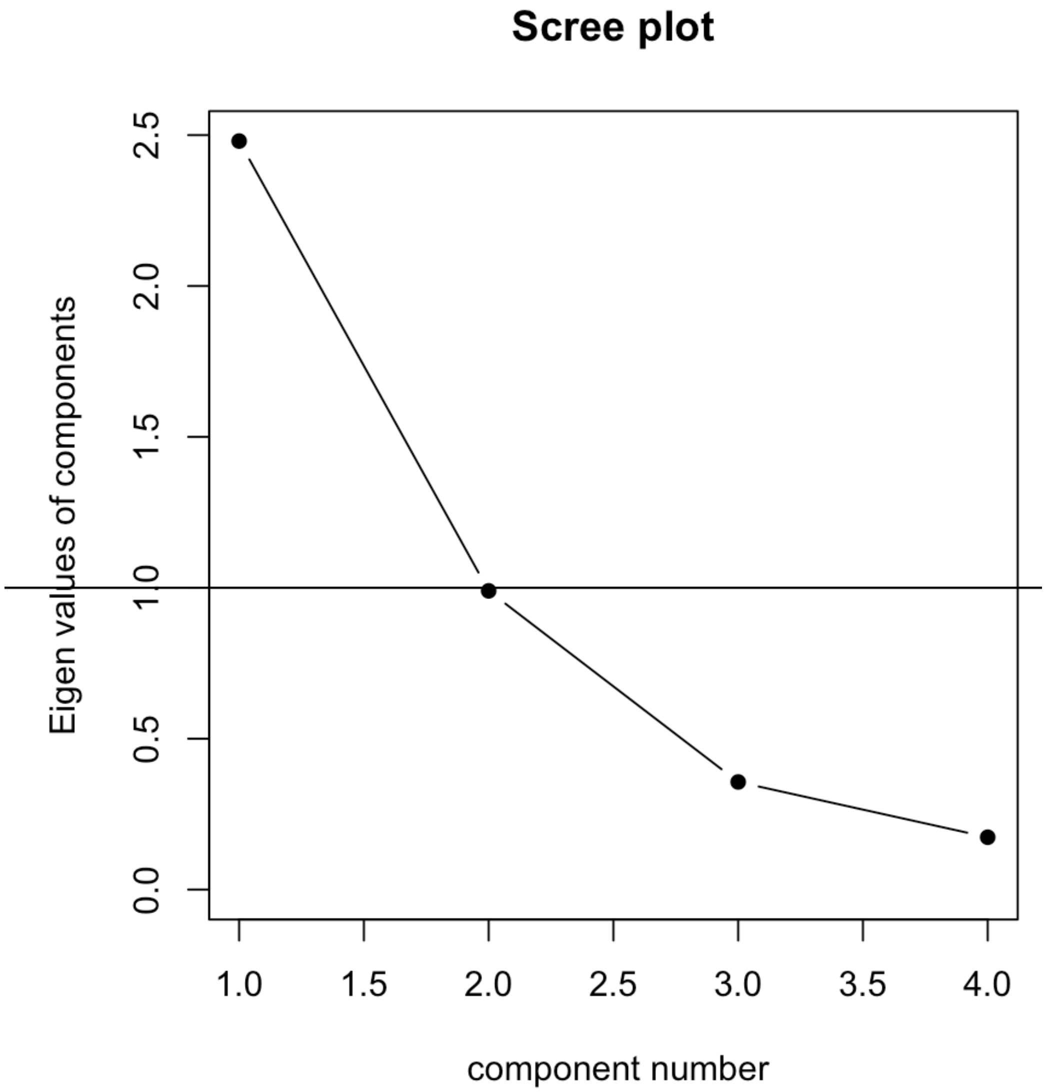
(d)


Components Analysis
raises
complaints
learning
rating
privileges
critical
advance
PC1
PC2
0.9
0.8
0.8
0.8
0.7
0.6
0.6

Problem 2

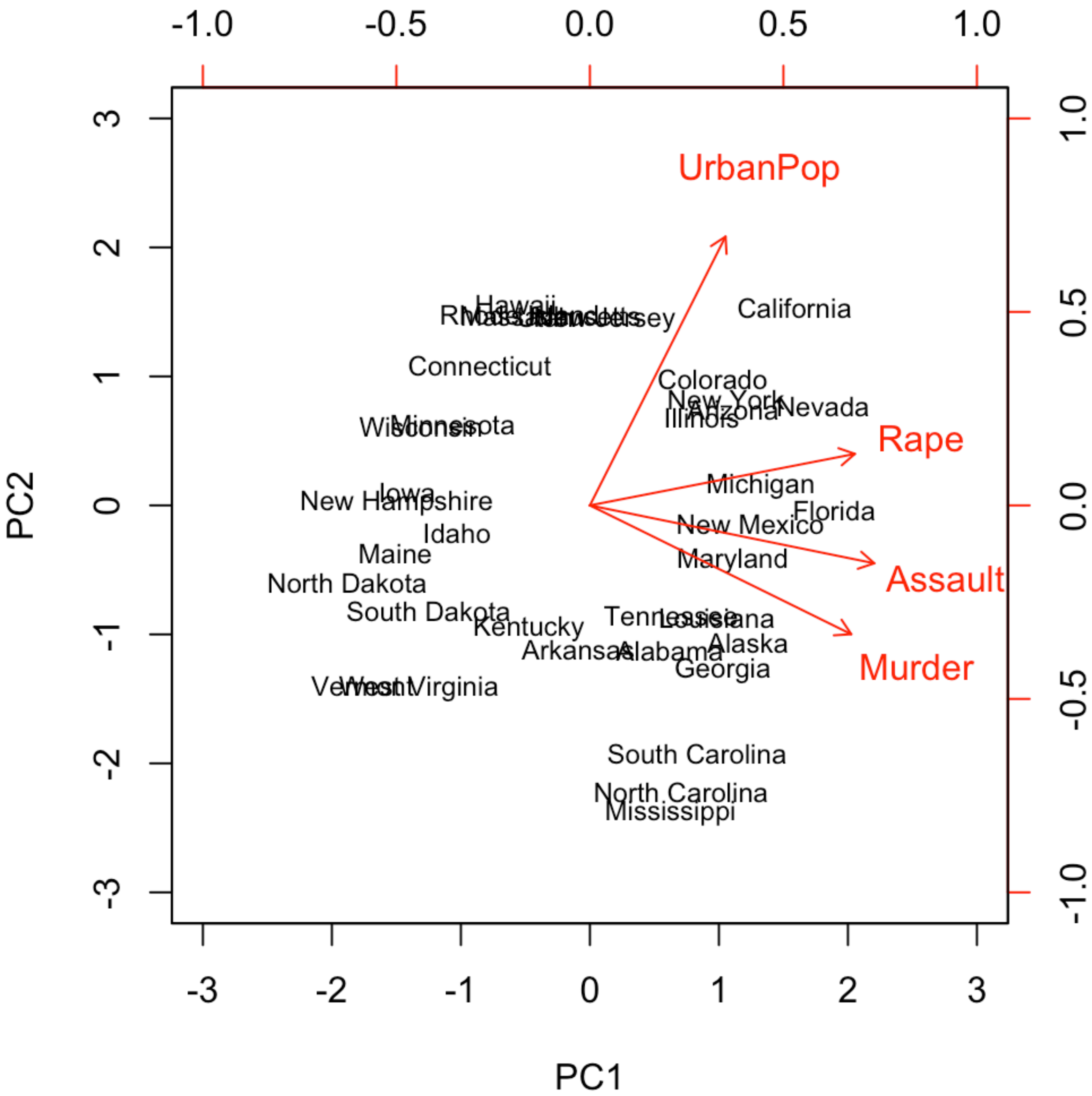
(a)

```
> paf_dat$KMO
[1] 0.65382
> paf_dat$Bartlett
[1] 88.288
> pcacor = cor(USArrests)
> cortest.bartlett(pcacor, n=186)
$chisq
[1] 344.67

$p.value
[1] 2.1515e-71

$df
[1] 6

> det(pcacor)
[1] 0.15181
> scree(USArrests, factors=FALSE, pc=TRUE) # 2 eigenvalues
> (pca <- principal(USArrests, nfactors=2, rotate='none'))
Principal Components Analysis
Call: principal(r = USArrests, nfactors = 2, rotate = "none")
Standardized loadings (pattern matrix) based upon correlation matrix
          PC1   PC2   h2    u2 com
Murder   0.84 -0.42 0.89 0.115 1.5
Assault  0.92 -0.19 0.88 0.121 1.1
UrbanPop 0.44  0.87 0.95 0.054 1.5
Rape     0.86  0.17 0.76 0.240 1.1

                       PC1  PC2
SS loadings           2.48 0.99
Proportion Var        0.62 0.25
Cumulative Var        0.62 0.87
Proportion Explained  0.71 0.29
Cumulative Proportion 0.71 1.00

Mean item complexity =  1.3
Test of the hypothesis that 2 components are sufficient.

The root mean square of the residuals (RMSR) is  0.08
 with the empirical chi square  3.44  with prob <  NA
```



From the some tests, we conclude that the given dataset is suitable for PCA. From the scree plot, we have 2 principal components. Also, from the PCA result, the equation to generate the score is **0.84*Murder + 0.92*Assault + 0.86*Rape.**

(b)



From above plot, we can say Rape, Assault, and Murder are correlated, and these seem to be uncorrelated with UrbanPop.
Since California gets aligned with UrbanPop, the majority of crime in California has come from UrbanPop. However, for Mississippi, it gets aligned much more with Murder. It means the most crime in Mississippi has come from Murder. For North Dakota, it positions at the opposite of all variables (UrbanPop, Rape, Assault, Murder) It means the amount of crime in North Dakota is quite small compared to the other states.