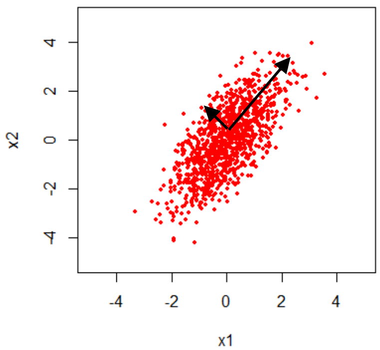
CSC 424 Homework 2

3.

Correlated Dataset are a) and c)

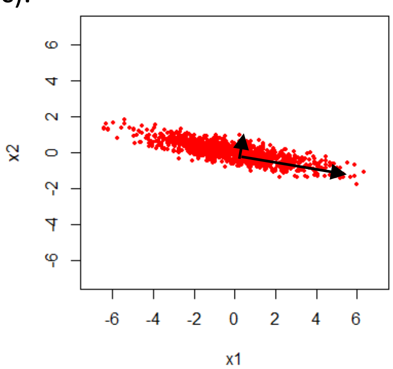
a).



The length of long vector is $\sqrt{2^2 + 4^2} = 2\sqrt{5}$

The length of short vector is $\sqrt{(-2)^2 + 1^2} = \sqrt{5}$

c).



The length of long vector is $\sqrt{6^2 + (-1)^2} = \sqrt{37}$

The length of short vector is $\sqrt{\frac{1}{6}^2 + 1^2} = \frac{\sqrt{37}}{37}$

4.

Homework 2

4.

a). $Mx = \lambda x \Rightarrow MIx = (\lambda \cdot I)x \Rightarrow Mx = (\lambda \cdot I)x$

$\Rightarrow (M - \lambda I)x = 0 \Rightarrow \det(M - \lambda I) = 0$

$\det\left(\begin{bmatrix} 1 & 1 \\ 1 & \frac{5}{2} \end{bmatrix} - \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}\lambda\right) = \det\left(\begin{bmatrix} 1-\lambda & 1 \\ 1 & \frac{5}{2}-\lambda \end{bmatrix}\right) = 0$

$(1-\lambda)(\frac{5}{2}-\lambda) - 1 = 0 \Rightarrow \lambda^2 - \frac{7}{2}\lambda + \frac{3}{2} = 0 \Rightarrow (\lambda - \frac{1}{2})(\lambda - 3) = 0 \Rightarrow \lambda = \frac{1}{2}, \lambda = 3$

so eigenvalues of M are $\frac{1}{2}$ and 3

$\lambda = \frac{1}{2}$: $\begin{bmatrix} 1-\frac{1}{2} & 1 \\ 1 & \frac{5}{2}-\frac{1}{2} \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = 0 \Rightarrow \begin{bmatrix} \frac{1}{2} & 1 \\ 1 & 2 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = 0 \Rightarrow \begin{bmatrix} \frac{1}{2}x_1 + x_2 \\ x_1 + 2x_2 \end{bmatrix} = 0$

$\Rightarrow x = \begin{bmatrix} -2 \\ 1 \end{bmatrix}$

$\lambda = 3$: $\begin{bmatrix} 1-3 & 1 \\ 1 & \frac{5}{2}-3 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = 0 \Rightarrow \begin{bmatrix} -2 & 1 \\ 1 & -\frac{1}{2} \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = 0 \Rightarrow \begin{bmatrix} -2x_1 + x_2 \\ x_1 - \frac{1}{2}x_2 \end{bmatrix} = 0$

$\Rightarrow x = \begin{bmatrix} \frac{1}{2} \\ 1 \end{bmatrix}$

So eigenvectors of M are $x_{\lambda=\frac{1}{2}} = \begin{bmatrix} -2 \\ 1 \end{bmatrix}$, $x_{\lambda=3} = \begin{bmatrix} \frac{1}{3} \\ 1 \end{bmatrix}$

b). $Mv = \begin{bmatrix} 0.4 & 0.88 & -0.28 \\ 0.88 & 1.1 & -0.98 \\ -0.28 & -0.98 & 2.26 \end{bmatrix}\begin{bmatrix} 1 \\ 2 \\ -3 \end{bmatrix} = \begin{bmatrix} 0.4 + 0.88 \times 2 + 0.28 \times 3 \\ 0.88 + 1.1 \times 2 + 0.98 \times 3 \\ -0.28 - 0.98 \times 2 - 2.26 \times 3 \end{bmatrix} = \begin{bmatrix} 3 \\ 6.02 \\ -9.02 \end{bmatrix}$

c). the corresponding eigenvalue is 3.

Source Code:

```
#a.
M <- matrix(c(1, 1, 1, 5/2),
        nrow = 2,
        ncol = 2,
        byrow = T)
eigen(M)
```

Output:

```
> eigen(M)
$values
[1] 3.0 0.5

$vectors
          [,1]       [,2]
[1,] 0.4472136 -0.8944272
[2,] 0.8944272  0.4472136
```

```
#b, c.
N <- matrix(c(0.4, 0.88, -0.28, 0.88, 1.1, -0.98, -0.28, -0.98, 2.26),
        nrow = 3,
        ncol = 3,
        byrow = T)
eigen(N)
```

Output:

```
> eigen(N)
$values
[1]  3.0071447  1.0000000 -0.2471447

$vectors
           [,1]      [,2]       [,3]
[1,] -0.2666746 0.5773503  0.7717197
[2,] -0.5349916 0.5773503 -0.6168068
[3,]  0.8016662 0.5773503 -0.1549129
```

5.

Source Code:

```
#Problem 5
employment <- read.table("/Users/Yiyang/Documents/CSC 
424/homeworkData/Employment.txt", sep = "\t", header = TRUE)
head(employment)

employmentNumbric <- employment[, c(2:10)]
employmentNumbric

employment.pr <- prcomp(employmentNumbric)
employment.pr
summary(employment.pr)
```

Output:

```
> employment.pr
Standard deviations:
[1] 17.4200491  6.6107228  3.8996608  2.3747336  1.5631376  1.0227575  0.6487274  0.2548071  0.0437262

Rotation:
             PC1          PC2          PC3         PC4          PC5         PC6          PC7          PC8        PC9
Agr  0.891758406 -0.006826746  0.118466699  0.09676712  0.180043781 -0.15262561 -0.091621401  0.068678066 -0.3354111
Min  0.001922618  0.092347069  0.079379068  0.01015633 -0.001121643  0.45636121  0.766470364  0.290464275 -0.3239614
Man -0.271271411  0.770269221  0.184678991  0.01040077  0.335999746 -0.20093094 -0.161983468  0.074117735 -0.3374633
PS  -0.008388285  0.012015922 -0.006768322 -0.01814178 -0.002459689  0.23086414  0.062936752 -0.909183254 -0.3398982
Con -0.049594016  0.068988571 -0.077312766  0.08292614 -0.724262390 -0.55835746  0.194294560 -0.004457936 -0.3253270
SI  -0.191798409 -0.234416513 -0.579612752  0.60760858  0.265863007 -0.02157242 -0.087935421  0.104435658 -0.3366529
Fin -0.031128614 -0.130082403 -0.469969939 -0.78119316  0.121062046 -0.05528170 -0.079976564  0.122754675 -0.3343621
SPS -0.298046310 -0.566777401  0.597745181 -0.04833726  0.235915950 -0.24786088 -0.004543731  0.052137300 -0.3323638
TC  -0.045364280 -0.009888386  0.159415225  0.03783527 -0.434890328  0.54593853 -0.567476088  0.223813567 -0.3342147
> summary(employment.pr)
Importance of components:
                           PC1    PC2     PC3     PC4     PC5     PC6     PC7     PC8     PC9
Standard deviation     17.4200 6.6107 3.89966 2.37473 1.56314 1.02276 0.64873 0.25481 0.04373
Proportion of Variance  0.8158 0.1175 0.04088 0.01516 0.00657 0.00281 0.00113 0.00017 0.00001
Cumulative Proportion   0.8158 0.9333 0.97415 0.98931 0.99588 0.99869 0.99982 0.99999 1.00000
```

a).

From the cumulative proportion, 2 principle components are required to explain 90% of the total variation.

b).

PC1 = 0.892 Agr + 0.002 Min – 0.271 Man – 0.008 PS – 0.050Con – 0.192SI – 0.031Fin – 0.298SPS – 0.045TC

PC2 = -0.007Agr + 0.092Min + 0.770Man + 0.012PS + 0.069Con – 0.234SI -0.130Fin – 0.567SPS – 0.010TC

For the PC1, the variables Man, SI and SPS are related, Agr is highly related. For the PC2, the variable SI, Fin, SPS are related, Man is highly related.

Rotation:

Source Code:

```
> library(psych)
> employment.pr2 <- principal(employmentNumbric, rotate = "varimax", nfactors = 2, score = TRUE)
> print(employment.pr2$loadings, cutoff = 0.4, sort = TRUE)

Loadings:
    RC1    RC2
Agr -0.905
Man  0.778
PS   0.572
Con  0.600
SPS  0.587  0.532
TC   0.743
Min        -0.858
SI   0.514  0.706
Fin         0.673

                 RC1   RC2
SS loadings    3.356 2.261
Proportion Var 0.373 0.251
Cumulative Var 0.373 0.624
```

PC1 = -0.905Agr + 0.778Man + 0.572 PS + 0.6Con + 0.587SPS + 0.743TC + 0.514 SI

PC2 = 0.532SPS – 0.858Min + 0.706SI + 0.673Fin

After the rotation, for the PC1, variables Agr, Man, PS, Con, SPS, TC and SI are all related with PC1, Agr is also the highly related; for the PC2, variables SPS, Min, SI and Fin are all related, Min is highly related.

c).

Source Code:

scores <- as.matrix(employmentNumbric) %*% as.matrix(employment.pr$rotation) print(scores)

Output:

```
> print(scores)
            PC1        PC2       PC3        PC4       PC5       PC6        PC7      PC8       PC9
 [1,] -17.067780  1.4651937  7.982860  7.0277695 12.898285 -13.42651  -8.815525 7.432995 -33.43634
 [2,] -11.047781 -5.2703442 13.340169  4.3009123 12.145507 -15.01391  -8.626196 7.284704 -33.54256
 [3,]  -8.679779  4.2231401  7.587835  6.7048505 12.781116 -14.75439  -8.339492 7.099479 -33.44156
 [4,] -13.944517 11.4389160 10.543830  5.6193959 14.986129 -14.27501  -9.170549 7.285925 -33.40042
 [5,]   4.907081  0.2598572  8.414140 10.3146748 12.755616 -13.47387  -8.430540 6.750611 -33.51667
 [6,]  -3.577777  6.0025269  7.932279 11.6438313 12.160618 -15.51583  -8.538764 7.223128 -33.47855
 [7,] -11.640845  8.7237909  5.710074  8.7991159 12.375088 -12.92886  -7.080947 7.829138 -33.46540
 [8,] -13.451548 -3.3321681  8.628327  6.1520018 10.896607 -14.84960  -8.273755 6.898469 -33.37749
 [9,] -18.279768  3.0596327 10.927524  5.8263706 14.707148 -13.57089  -8.339886 6.899377 -33.41133
[10,]  -6.022511  9.7480518  5.581538  8.1079132 11.889568 -13.18226  -9.293480 6.915718 -33.42863
[11,]  -6.388140  2.4150815 10.405713  5.8697207 12.681393 -13.33247  -9.716017 6.891997 -33.48378
[12,]  25.875990  4.5867450  7.422009  9.6905445 10.528018 -13.41489  -9.627877 7.309086 -33.42806
[13,] -10.523112 -2.4673556 10.564351  7.4261313 10.400991 -12.59702  -9.670185 7.542634 -33.38539
[14,]   9.852772  6.3057216  9.101578  9.0140878 12.475362 -15.04418  -9.311864 7.096180 -33.46192
[15,]   6.223880 12.5500976  5.459096  2.3528413  9.367169 -15.71732  -8.530548 7.132395 -33.40466
[16,] -14.863068 -2.1353221 14.260242  4.1998341 13.824777 -14.74871  -8.750588 7.146910 -33.42539
[17,] -12.234932 16.1706227  4.648928  7.8776659 13.652905 -15.17456 -10.046351 7.175743 -33.45307
[18,]  52.564551 -2.2502376 13.303295  8.6298099 15.583763 -14.37913  -8.728364 7.287882 -33.38611
[19,]   4.605698 13.0982727 15.278097  6.9709692 12.968098 -14.56201  -9.009601 7.355823 -33.42085
[20,]  -2.797220 15.6261020 14.120396  6.9815607 12.324985 -13.77196  -8.208166 7.048150 -33.37889
[21,] -16.966620 17.1237531 15.233788  6.5883886 13.772133 -12.73912  -9.225372 7.317880 -33.49045
[22,]   3.584644 11.3783732 13.321682  7.5644472 11.092140 -12.13293  -8.210574 6.600079 -33.44094
[23,]  13.764616  9.3362492 13.927087  7.3192242 11.046165 -13.68723  -8.127741 7.188634 -33.43561
[24,]  17.460243 15.5166524 12.919665  6.5957530 12.367935 -15.47439  -8.230583 7.144579 -33.49167
[25,]   5.035950  5.5194518 18.786896  5.8548995  9.475003 -14.09223  -9.413424 7.370376 -33.47904
[26,]  35.281555  7.0841719  3.390255  0.2452743 13.439972 -12.49862  -8.772813 7.230649 -33.48910
```

The highest value in PC1 is 18th country Turkey: 52.56455
The lowest value in PC1 is 9th country United State: -18.27977
The highest value in PC2 is 21st country E. Germany: 17.12375
The lowest value in PC2 is 2nd country Denmark: -5.270344

d).
Source Code:

```
library(psych)
options("scipen" = 100, "digits" = 5)
round(cor(employmentNumbric), 2)
corrTest = corr.test(employmentNumbric, adjust = "none")
corrTest

E = corrTest$p
E

ETest = ifelse(E < 0.1, T, F)
ETest
colSums(ETest) – 1
```

Output:

```
> colSums(ETest) - 1
Agr Min Man  PS Con  SI Fin SPS  TC
  6   4   5   4   4   5   2   3   5
```

From this result, there are total 9 variables, 75% of other fields indicates that (9-1) * 0.75 = 6. Agr = 6, then Agr should be removed. Then get a new matrix.

```
employment.pr3 <- prcomp(employmentNumbric[2: 9], scale = TRUE)
employment.pr3
summary(employment.pr3)
```

Output:

```
> employment.pr3
Standard deviations:
[1] 1.60028 1.45312 1.04643 0.99661 0.71416 0.60575 0.47494 0.36983

Rotation:
          PC1       PC2      PC3       PC4      PC5       PC6       PC7       PC8
Min 0.0821290 -0.616079  0.19406 -0.064858  0.14057 -0.116730 -0.728580 -0.087292
Man 0.4290591 -0.285668  0.15249  0.353660  0.41965 -0.425447  0.462734 -0.121390
PS  0.3409533 -0.201274  0.58888 -0.368932 -0.20209  0.368959  0.259393  0.340720
Con 0.3938579  0.023648 -0.14161  0.681014 -0.39843  0.163153 -0.219631  0.356205
SI  0.3771273  0.426574  0.13765  0.065810  0.44985  0.496992 -0.234071 -0.385200
Fin 0.0031875  0.460201  0.59566  0.069895 -0.31710 -0.510080 -0.195166 -0.172299
SPS 0.4087930  0.302467 -0.28495 -0.409450  0.24556 -0.359973 -0.205234  0.510446
TC  0.4779572 -0.111953 -0.34249 -0.306532 -0.49125 -0.076528  0.067275 -0.544106
> summary(employment.pr3)
Importance of components:
                        PC1   PC2   PC3   PC4    PC5    PC6    PC7    PC8
Standard deviation     1.60 1.453 1.046 0.997 0.7142 0.6058 0.4749 0.3698
Proportion of Variance 0.32 0.264 0.137 0.124 0.0638 0.0459 0.0282 0.0171
Cumulative Proportion  0.32 0.584 0.721 0.845 0.9088 0.9547 0.9829 1.0000
```

There are five principle components needed to account for more than 90% total variance.

After rotating:

```
> employment.pr4 <- principal(employmentNumbric[2: 9], rotate = "varimax", nfactors = 5, score = TRUE)
> print(employment.pr4$loadings, cutoff = 0.4, sort = TRUE)

Loadings:
    RC3    RC5    RC4    RC2    RC1
Min  0.698                -0.512
PS   0.889
SI          0.839
SPS         0.792                0.526
Man  0.558         0.627
Con                0.947
Fin                       0.914
TC                               0.881

                 RC3   RC5   RC4   RC2   RC1
SS loadings    1.644 1.601 1.423 1.392 1.211
Proportion Var 0.206 0.200 0.178 0.174 0.151
Cumulative Var 0.206 0.406 0.584 0.757 0.909
```

I think it makes some improvement compare with b), after remove the highest correlated field.

6.

a).

Source Code:

census <- read.csv("/Users/Yiyang/Documents/CSC 424/homeworkData/Census2.csv", sep = ",", header = TRUE)

census

census.pr <- prcomp(census)
census.pr
summary(census.pr)

Output:

```
>
> census.pr <- prcomp(census)
> census.pr
Standard deviations:
[1] 56446.885008    10.206857     6.218887     2.246707     1.559823

Rotation:
                       PC1           PC2           PC3           PC4           PC5
Population     8.537905e-07 -4.108282e-02 -7.059713e-02  4.826860e-01  8.719762e-01
Professional   3.775797e-05  7.080539e-02 -7.460074e-02 -8.714029e-01  4.796648e-01
Employed      -1.367095e-06 -5.126328e-01 -8.542663e-01 -1.524163e-02 -8.487872e-02
Government     3.004471e-05  8.546967e-01 -5.095880e-01  8.624903e-02 -4.873218e-02
MedianHomeVal  1.000000e+00 -2.901832e-05  1.701961e-05  2.987813e-05 -1.750755e-05
>
> summary(census.pr)
Importance of components:
                         PC1   PC2   PC3   PC4  PC5
Standard deviation     56447 10.21 6.219 2.247 1.56
Proportion of Variance     1  0.00 0.000 0.000 0.00
Cumulative Proportion      1  1.00 1.000 1.000 1.00
>
```

From the PCA of this dataset, there are no variables are highly related with PC1. The coefficients of these variables accounted for the first component are very small, the main reason, I think, is the value of the median home value is too large for the other variables.

b).
Source Code:
censusEdit <- census[, c(1:4)]
censusEdit$MedianHomeVal <- census$MedianHomeVal/(100000)
censusEdit

census.pr2 <- prcomp(censusEdit)
census.pr2
summary(census.pr2)

Output:

```
>
> census.pr2 <- prcomp(censusEdit)
> census.pr2
Standard deviations:
[1] 10.3448177  6.2985820  2.8932449  1.6934798  0.3933104

Rotation:
                      PC1         PC2         PC3         PC4          PC5
Population     0.038887287 -0.07114494  0.18789258  0.97713524 -0.057699864
Professional  -0.105321969 -0.12975236 -0.96099580  0.17135181 -0.138554092
Employed       0.492363944 -0.86438807  0.04579737 -0.09104368  0.004966048
Government    -0.863069865 -0.48033178  0.15318538 -0.02968577  0.006691800
MedianHomeVal -0.009122262 -0.01474342 -0.12498114  0.08170118  0.988637470
> summary(census.pr2)
Importance of components:
                          PC1    PC2     PC3     PC4     PC5
Standard deviation     10.345 6.2986 2.89324 1.69348 0.39331
Proportion of Variance  0.677 0.2510 0.05295 0.01814 0.00098
Cumulative Proportion   0.677 0.9279 0.98088 0.99902 1.00000
```

After divide all the value in MedianHomeVal by 100000, the result this time is better, and we could find there are two PCs could explain 90% of the total variation.

c).

Source Code:

census.pr3 <- prcomp(census, scale = TRUE)

census.pr3

summary(census.pr3)

```
> census.pr3 <- prcomp(census, scale = TRUE)
> census.pr3
Standard deviations:
[1] 1.4113534 1.1694129 0.9296006 0.7314787 0.4912604

Rotation:
                    PC1        PC2         PC3        PC4        PC5
Population    0.2625829 -0.4629936  0.78390268 -0.2169291  0.2347882
Professional -0.5933541 -0.3256442 -0.16407255  0.1446471  0.7028828
Employed      0.3256978 -0.6051419 -0.22487455  0.6628689 -0.1943206
Government   -0.4792022  0.2524850  0.55070086  0.5716730 -0.2766497
MedianHomeVal -0.4932213 -0.4996473 -0.06882436 -0.4072024 -0.5801162
> summary(census.pr3)
Importance of components:
                          PC1    PC2    PC3    PC4     PC5
Standard deviation     1.4114 1.1694 0.9296 0.7315 0.49126
Proportion of Variance 0.3984 0.2735 0.1728 0.1070 0.04827
Cumulative Proportion  0.3984 0.6719 0.8447 0.9517 1.00000
```

After the PCA of the correlation matrix, I find that the result is similar with the result in b), but the number of PCs that could explain 90% of the total variation changes from two to three, the coefficients of the variables in PC1 and PC2 are changes either.

d).

For PC1, the most significant variable is professional, it is negative relationship with PC1, Government and MedianHomeVal are also significant related with PC1.

For PC2, the most significant variable is Employed, it is negative relationship with PC2, then Population and MedianHomeVal are also significant related with PC2.

For PC3, the most significant variable is Population, it is positive relationship with PC3, then Government is also significant related with PC3.

e).

I think the using of correlation matrix could make the data in different magnitudes to the same level and reduce the error that occurs when using the covariance matrix at this situation. In this dataset, PCA of correlation matrix is appropriate, since it could make the larger value in MedianHomeVal column to the same level of the other columns and then get a relatively accurate result of this dataset.

7.

Source Code:

track <- read.table("/Users/Yiyang/Documents/CSC 424/homeworkData/trackRecord.txt", header = TRUE)

trackNumberic <- track[2: 9]

trackNumberic

library(psych)

options("scipen" = 100, "digits" = 5)

round(cor(trackNumberic), 2)
tcorrTest = corr.test(trackNumberic, adjust = "none")
tcorrTest

Tr = tcorrTest$p
Tr

TTest = ifelse(Tr < 0.1, T, F)
TTest
colSums(TTest) – 1

```
> colSums(TTest) - 1
    m100     m200     m400     m800    m1500    m5000   m10000 Marathon 
       7        7        7        7        7        7        7        7 
```

As the result, all the variables are highly correlated with each other, so don't apply the correlation matrix. Then apply the covariance matrix.

track.pr <- prcomp(trackNumberic)
track.pr
summary(track.pr)

```
> summary(track.pr)
Importance of components:
                         PC1    PC2     PC3     PC4     PC5     PC6     PC7    PC8
Standard deviation     9.193 1.0683 0.47829 0.28384 0.10853 0.07910 0.05536 0.0199
Proportion of Variance 0.983 0.0133 0.00266 0.00094 0.00014 0.00007 0.00004 0.0000
Cumulative Proportion  0.983 0.9961 0.99881 0.99975 0.99989 0.99996 1.00000 1.0000
```

trackEdit <- trackNumberic[1: 3]
trackEdit$m800_Sec <- trackNumberic$m800*(60)
trackEdit$m1500_Sec <- trackNumberic$m1500*(60)
trackEdit$m5000_Sec <- trackNumberic$m5000*(60)
trackEdit$m10000_Sec <- trackNumberic$m10000*(60)
trackEdit$Marathon <- trackNumberic$Marathon
trackEdit

trackEdit.pr <- prcomp(trackEdit)
summary(trackEdit.pr)

```
> trackEdit.pr <- prcomp(trackEdit)
> summary(trackEdit.pr)
Importance of components:
                           PC1     PC2     PC3     PC4     PC5     PC6     PC7    PC8
Standard deviation     111.101 6.60057 3.64550 2.62996 1.33409 0.85629 0.26730 0.0762
Proportion of Variance   0.995 0.00351 0.00107 0.00056 0.00014 0.00006 0.00001 0.0000
Cumulative Proportion    0.995 0.99816 0.99923 0.99979 0.99993 0.99999 1.00000 1.0000
```

track.pr2 <- prcomp(trackNumberic, scale = TRUE)
summary(track.pr2)

```
> summary(track.pr2)
Importance of components:
                         PC1    PC2    PC3    PC4    PC5     PC6     PC7     PC8
Standard deviation     2.589 0.7990 0.4770 0.4537 0.3124 0.26587 0.21666 0.09858
Proportion of Variance 0.838 0.0798 0.0284 0.0257 0.0122 0.00884 0.00587 0.00121
Cumulative Proportion  0.838 0.9177 0.9462 0.9719 0.9841 0.99292 0.99879 1.00000
```

library(psych)
track.pr3 <- principal(trackNumberic, rotate = "varimax", nfactors = 8, score = TRUE)
print(track.pr3$loadings, cutoff = 0.4, sort = TRUE)

```
> employment.pr4 <- principal(employmentNumbric[2: 9], rotate = "varimax", nfactors = 5, score = TRUE)
> print(employment.pr4$loadings, cutoff = 0.4, sort = TRUE)

Loadings:
    RC3    RC5    RC4    RC2    RC1
Min  0.698               -0.512
PS   0.889
SI          0.839
SPS         0.792               0.526
Man  0.558         0.627
Con                0.947
Fin                       0.914
TC                              0.881

               RC3   RC5   RC4   RC2   RC1
SS loadings    1.644 1.601 1.423 1.392 1.211
Proportion Var 0.206 0.200 0.178 0.174 0.151
Cumulative Var 0.206 0.406 0.584 0.757 0.909
```

track.fc <- factanal(trackNumberic, 4)
print(track.fc$loadings, cutoff = 0.4, sort = TRUE)

```
> track.fc <- factanal(trackNumberic, 4)
> print(track.fc$loadings, cutoff = 0.4, sort = TRUE)

Loadings:
         Factor1 Factor2 Factor3 Factor4
m1500     0.670   0.491   0.447
m5000     0.843   0.425
m10000    0.871
Marathon  0.837
m100              0.849
m200              0.867
m400      0.475   0.669
m800      0.539   0.440   0.715

               Factor1 Factor2 Factor3 Factor4
SS loadings      3.405   2.838   1.177   0.045
Proportion Var   0.426   0.355   0.147   0.006
Cumulative Var   0.426   0.780   0.927   0.933
```

After these analysis, PC1 and PC2 are found to explain 90% of total variation.

9.
How suitable is their data for PCA?
There data is about sporulation, the procedures of sporulation is in a special law, each data in the sporulation, have some relationship with each other, so these data are suitable for PCA.

How are they applying PCA? Are they trying to extract interpretable underlying variables, or is their goal more along the lines of dimensionality reduction?
They apply the correlation test and then apply the PCA according to the correlation matrix, and then get the largest Principle Components for the data.

What kind of factor rotation do they use if any?
After analyzing from first and second PC, they reduce some dimensions and then get a new PC3.

How many components do they concentrate on in their analysis?
They have 3 principle components, which they concentrate on in their analysis.

Do they evaluate, and how do they evaluate the stability of the components?

What conclusions does PCA allow them to draw?
“In particular, these clustershighlight the potential biases used in analyzing clusters using traditional cognitivecategories. This observation corroborates the original investigators’ finding that the clustersare somewhat arbitrary; many genes were found to have high correlation with multiplecluster representatives (Chu et al. 1998)” From the article, they find some high correlation between many genes according to the clusters they got.