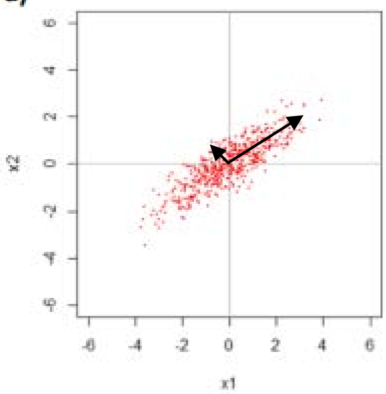
Keiland Pullen
Homework 3

DSC 424 | Fall 2021

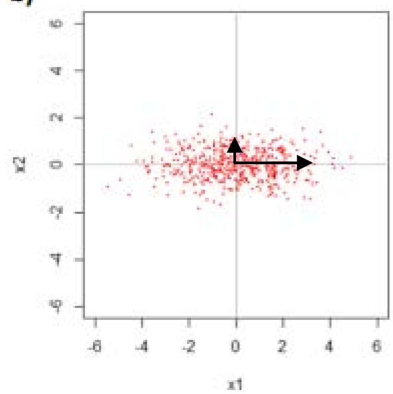
2.)

a)



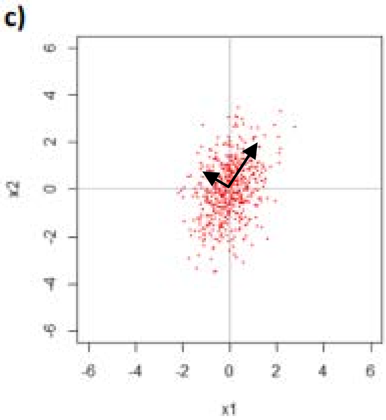
The estimated length of the long vector is 4
The estimated length of the short vector is 1

b)



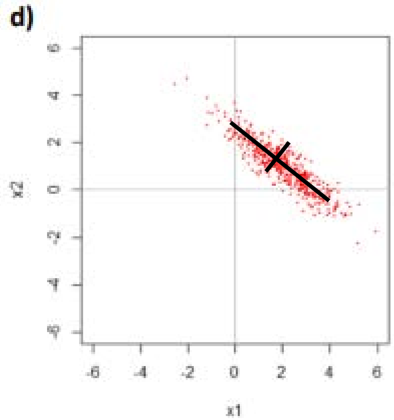
The estimate length of the long vector is 4
The estimated length of the short vector is 1

c)



The estimated length of the long vector is 2
The estimated length of the short vector is 1

d)



The estimated length of the long vector is 5
The estimated length of the short vector is 1

Keiland Pullen
Homework 3

DSC 424 | Fall 2021

3.)
a.)

Keiland Pullen

#3a.) $M = \begin{bmatrix} 5 & -1 \\ -1 & 5 \end{bmatrix}$

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

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

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

$\det = (ad - bc) \quad = (5-\lambda)(5-\lambda) - (-1 \cdot -1)$

$= 25 - 5\lambda - 5\lambda + \lambda^2 - 1$

$= \lambda^2 - 10\lambda + 25 - 1$

$= \lambda^2 - 10\lambda + 24$

$= (\lambda - 6)(\lambda - 4)$

Eigenvalues = $\underline{\lambda = 6, 4}$

$M - \lambda I = \begin{bmatrix} 5-\lambda & -1 \\ -1 & 5-\lambda \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$

$= \begin{bmatrix} 5-6 & -1 \\ -1 & 5-6 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \qquad \begin{bmatrix} 5-4 & -1 \\ -1 & 5-4 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$

$= \begin{bmatrix} -1 & -1 \\ -1 & -1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \qquad \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$

$-x_1 - x_2 = \emptyset \qquad x_1 - x_2 = \emptyset$

$x_1 = -x_2 \qquad x_1 = x_2$

if $x_2 = 1$, then $x_1 = -1$ — if $x_2 = 1$, then $x = 1$

$\underline{v_1 = \begin{bmatrix} -1 \\ 1 \end{bmatrix}} \qquad \underline{v_2 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}}$

```
> M <- matrix(c(5, -1, -1, 5), 2,2, byrow=TRUE)
> M
     [,1] [,2]
[1,]    5   -1
[2,]   -1    5
> eigen(M)
eigen() decomposition
$values
[1] 6 4

$vectors
           [,1]       [,2]
[1,] -0.7071068 -0.7071068
[2,]  0.7071068 -0.7071068
```

b.)

Keiland Pullen

3b.)

$$\begin{array}{cc} N & v \\ \begin{bmatrix} 21 & -2 & 1 \\ -3 & 10 & -11 \\ 3 & -22 & -1 \end{bmatrix} & \begin{bmatrix} -1 \\ 1 \\ -1 \end{bmatrix} \end{array}$$

$$\begin{bmatrix} (21 * -1) + (-2 * 1) + (1 * -1) \\ (-3 * -1) + (10 * 1) + (-11 * -1) \\ (3 * -1) + (-22 * 1) + (-1 * -1) \end{bmatrix}$$

$$\begin{bmatrix} (-21) + (-2) + (-1) \\ (3) + (10) + (11) \\ (-3) + (-22) + (1) \end{bmatrix}$$

$$\begin{bmatrix} -24 \\ 24 \\ -24 \end{bmatrix} = 24 \cdot \begin{bmatrix} -1 \\ 1 \\ -1 \end{bmatrix}$$

```
> N <- matrix(c(21, -2, 1, -3, 10, -11, 3, -22, -1), nrow=3, byrow=T)
> N
     [,1] [,2] [,3]
[1,]   21   -2    1
[2,]   -3   10  -11
[3,]    3  -22   -1
> v <- matrix(c(-1, 1, -1), byrow=F)
> v
     [,1]
[1,]   -1
[2,]    1
[3,]   -1
> N%*%v
     [,1]
[1,]  -24
[2,]   24
[3,]  -24
> 24*v
     [,1]
[1,]  -24
[2,]   24
[3,]  -24
```

c.)

The corresponding eigenvalue should be <u>1</u>.

Keiland Pullen DSC 424 | Fall 2021
Homework 3

4.)
a.)

```
setwd("C:/Users/Home/Desktop/DePaul/DSC-424-AdvancedDataAnalysis/Week-4/Homework")

census = read.csv("Census2.csv")
View(census)

head(census)

c = prcomp(census)
print(c)
summary(c)
```

Standard deviations (1, .., p=5):
[1] 56446.885008 10.206857 6.218887 2.246707 1.559823

Rotation (n x k) = (5 x 5):

	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

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

In this dataset, 100% of the variance is accounted for in PC1. This may be due to the units in the dataset. The Median home value dataset is listed in dollars, where as there are three variables with units in percentages.

b.)

```
census2 <- census[,c(1:4)]
census2$MedianHomeValue <- census$MedianHomeVal/100000
head(census2)
c2 = prcomp(census2)
print(c2)
summary(c2)
> c2 = prcomp(census2)
> print(c2)
```

Standard deviations (1, .., p=5):
[1] 10.3448177 6.2985820 2.8932449 1.6934798 0.3933104

Rotation (n x k) = (5 x 5):

	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
MedianHomeValue	-0.009122262	-0.01474342	-0.12498114	0.08170118	0.988637470

\> summary(c2)
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 adjusting the Median Home Value variable, we can now see that three PCs can explain 98% of the variance.

c.)

After scaling the Median Home Value variable, there isn't a need for any additional scaling.

> head(census2)

	Population	Professional	Employed	Government	MedianHomeValue
1	2.67	5.71	69.02	30.3	1.48
2	2.25	4.37	72.98	43.3	1.44
3	3.12	10.27	64.94	32.0	2.11
4	5.14	7.44	71.29	24.5	1.85
5	5.54	9.25	74.94	31.0	2.23
6	5.04	4.84	53.61	48.2	1.60

d.)

```
> c3 = prcomp(census2, scale=TRUE)
> print(c3)
Standard deviations (1, .., p=5):
[1] 1.4113534 1.1694129 0.9296006 0.7314787 0.4912604

Rotation (n x k) = (5 x 5):
```

	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
MedianHomeValue	-0.4932213	-0.4996473	-0.06882436	-0.4072024	-0.5801162

```
> summary(c3)
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

Using PCA with the correlation matrix shows that 4 Principal Components can explain up to 95% of the variance. Compared to 2 PCs explaining 93% of the variance and 3 PCs explaining 99% of the variance in answer b.

e.)

Reviewing PC1, it appears that Professional, MedianHomeValue and Government are the significant entries and they each have a negative relationship. In PC2, Employed is the most significant entry and it also has a negative relationship. In PC3, Professional has a positive relationship and is the most significant variable. Doing this allows us to reduce the number of features or predictors in a dataset.

f.)

For this dataset, using the correlation matrix for PCA was a better option that the covariance matrix because if allowed for the data to be scale, in particular, the MedianHomeVal predictor. Thus, allowing for a more accurate view of the predictors.

5.)
a.)
For this dataset, there isn't a need for any data scaling, because the features are all in units of percentage.

b.)

```
setwd("C:/Users/Home/Desktop/DePaul/DSC-424-AdvancedDataAnalysis/Week-4/Homework")
employment = read.delim("Employment.txt")
View(employment)
head(employment)
employment2 <- employment[,c(2:9)]
head(employment2)
emp = prcomp(employment2, scale=TRUE)
print(emp)
summary(emp)
```

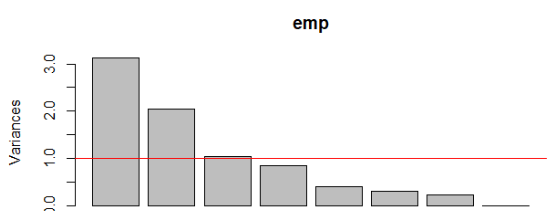
Standard deviations (1, .., p=8):
[1] 1.76670981 1.43073103 1.02302558 0.92131421 0.63073817 0.55959321 0.47287305 0.04195023

Rotation (n x k) = (8 x 8):

	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8
Agr	0.55742936	-0.04856193	0.01524420	0.09971546	-0.1967998	0.009456233	0.01756224	-0.79850528
Min	0.06622652	0.63536406	0.16511982	-0.05879390	0.1410821	-0.022393053	-0.73383738	-0.04777420
Man	-0.34599938	0.45653320	-0.14055175	0.14807077	0.4738874	0.311463276	0.42541638	-0.35724550
PS	-0.24143758	0.32233980	0.67304330	0.02618831	-0.4900020	-0.196300491	0.32145305	-0.04651816
Con	-0.33718653	0.11905508	-0.52502230	0.50338841	-0.3524925	-0.434509486	-0.13309930	-0.11098617
SI	-0.45462924	-0.26130841	0.03764582	-0.02002179	-0.3600305	0.661155624	-0.33265677	-0.21401596
Fin	-0.15996555	-0.40190133	0.47125975	0.55324514	0.4427153	-0.179864882	-0.20196612	-0.12482750
SPS	-0.40113265	-0.19403057	-0.02469014	-0.63569519	0.1637824	-0.453419504	-0.08773167	-0.39574739

> summary(emp)
Importance of components:

	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8
Standard deviation	1.7667	1.4307	1.0230	0.9213	0.63074	0.55959	0.47287	0.04195
Proportion of Variance	0.3902	0.2559	0.1308	0.1061	0.04973	0.03914	0.02795	0.00022
Cumulative Proportion	0.3902	0.6460	0.7769	0.8830	0.93269	0.97183	0.99978	1.00000



This method suggests that there are 3 components that are responsible for 78% of the variance. However, including PC 4 suggests that there are 4 components that are responsible for 88% of the variance. There could be some ambiguity with PC 5 as it would add an additional 5% of variance.

c.)
No, VARIMAX factor rotation was not applied. Prior to using the "principal" function and VARIMAX rotation, we should know the number of components that are significant.

d.)

	PC1	PC2	PC3	PC4
Agr	0.55742936	-0.04856193	0.01524420	0.09971546
Min	0.06622652	0.63536406	0.16511982	-0.05879390
Man	-0.34599938	0.45653320	-0.14055175	0.14807077
PS	-0.24143758	0.32233980	0.67304330	0.02618831
Con	-0.33718653	0.11905508	-0.52502230	0.50338841
SI	-0.45462924	-0.26130841	0.03764582	-0.02002179
Fin	-0.15996555	-0.40190133	0.47125975	0.55324514
SPS	-0.40113265	-0.19403057	-0.02469014	-0.63569519

PC1 = .55 * Agr + .06 * Min + -0.35 * Man + -0.24 * PS + -0.34 * Con + -0.45 * SI + - 0.16 * Fin + -0.40 * SPS
The variable AGR is high contribution while SI, Con, SPS, and Man are also high but contribute negatively.

PC2 = -0.05 * Agr + .64 * Min + .45 * Man + .32 * PS + .12 * Con + -0.26 * SI + -.40 * Fin + -.19 * SPS
The variables Min, Man and PS are significant positive contributors, while Fin is a significantly negative contributor.

PC3 = .015 * Agr + .17 * Min + -0.14 * Man + .67 * PS + -.53 * Con + .038 * SI + .47 * Fin + -0.024 * SPS
The variables PS and Fin are significant contributors while Con is a strong negative contributor

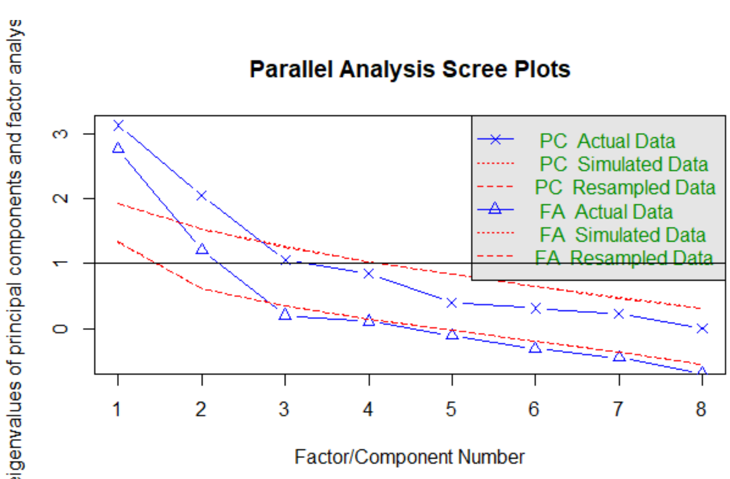
PC4 = .099 * Agr + -0.06 * Min + 0.15 * Man + .026 * PS + .50 * Con + -.02 * SI + .55 * Fin + -0.64 * SPS
The variables Fin and Con are significant positive contributors while SPS is a significantly negative contributor.

For these variables, I would think that they could be cleaned up a bit more as they don't seem to be "easy" in regard to separate and interpret.

e.)

library(psych)
parallel_emp = fa.parallel(employment2, n.iter=500)

> parallel_emp = fa.parallel(employment2, n.iter=500)
Parallel analysis suggests that the number of factors = 2 and the number of components = 2



The Scree Plot suggests that the number of Principal Components needed is 3. Earlier, I selected 4 Principal Components because a case can be made for PC4 being slightly above the knee.

f.)

```
principal(employment2, nfactors=3, rotate = "varimax")
> principal(employment2, nfactors=3, rotate = "varimax")
Principal Components Analysis
Call: principal(r = employment2, nfactors = 3, rotate = "varimax")
Standardized loadings (pattern matrix) based upon correlation matrix
```

	RC1	RC2	RC3	h2	u2	com
Agr	-0.83	-0.42	-0.32	0.97	0.025	1.8
Min	0.01	-0.66	0.66	0.87	0.131	2.0
Man	0.72	-0.26	0.49	0.82	0.179	2.1
PS	0.09	0.17	0.91	0.87	0.131	1.1
Con	0.81	-0.08	-0.10	0.67	0.328	1.1
SI	0.56	0.69	0.03	0.79	0.214	1.9
Fin	-0.15	0.79	0.07	0.64	0.357	1.1
SPS	0.54	0.54	0.02	0.58	0.420	2.0

	RC1	RC2	RC3
SS loadings	2.49	2.10	1.63
Proportion Var	0.31	0.26	0.20
Cumulative Var	0.31	0.57	0.78
Proportion Explained	0.40	0.34	0.26
Cumulative Proportion	0.40	0.74	1.00

```
Mean item complexity =  1.6
Test of the hypothesis that 3 components are sufficient.

The root mean square of the residuals (RMSR) is  0.09
 with the empirical chi square  13.07  with prob <  0.07

Fit based upon off diagonal values = 0.94
```

In PC1, the variables Man, Con, SI and SPS are all positively significant where as Agr is significant, but negatively.
PC1 = -0.83 * Agr + 0.01 * Min + 0.72 * Man + 0.09 * PS + 0.81 * Con + 0.56 * SI + -0.15 * Fin + 0.54 * SPS

In PC2, the variables SI, Fin and SPS are significant and are related. The variable Min also has a negative significant impact.
PC2 = -.042 * Agr + -0.66 * Min + -0.26 * Man + 0.17 * PS + -0.08 * Con + 0.69 * SI + 0.79 * Fin + 0.54 * SPS

In PC3, the variables PS and Min are positively significant.
PC3 = -0.32 * Agr + 0.66 * Min + 0.49 * Man + 0.91 * PS + -0.10 * Con + 0.03 * SI + 0.07 * Fin + 0.02 * SPS

g.)

as.matrix(employment2) %*% as.matrix(emp$rotation)

	PC1	PC2	PC3
[1,]	-30.978019	1.63425489	-4.3956706
[2,]	-25.945159	-1.92449444	-4.0433338
[3,]	-24.323181	2.70170112	-4.7333883
[4,]	-27.536409	7.90447516	-5.5940761
[5,]	-13.435601	0.71977834	-4.0149248
[6,]	-20.686805	4.28838177	-7.5123419
[7,]	-26.302768	6.60176402	-5.6013208
[8,]	-28.549515	-1.43577346	-4.3960193
[9,]	-31.463096	3.62043590	-4.0290650
[10,]	-21.830302	5.77322645	-5.1255416
[11,]	-21.807745	1.98046580	-3.8409247
[12,]	4.127036	1.45954251	-4.3000250
[13,]	-25.299696	-0.27157837	-4.7356713
[14,]	-9.115201	3.41827136	-5.6157892
[15,]	-11.592496	5.76177670	-5.0116343
[16,]	-28.658620	0.43635648	-4.1532809
[17,]	-27.151019	8.70795077	-6.8354787
[18,]	26.267466	-2.93684074	-0.9588893
[19,]	-11.752877	10.03793676	-7.4284931
[20,]	-17.623417	12.43190352	-7.6907306
[21,]	-28.747239	14.07461228	-7.9215793
[22,]	-12.480689	9.86257847	-5.9907277
[23,]	-4.452149	7.65578923	-6.2207680
[24,]	-1.594612	10.28564872	-6.9728770
[25,]	-11.189921	6.43180556	-7.5775770
[26,]	12.672324	-0.04643955	2.2318519

The countries with the highest values are Turkey[18] and Yugoslavia[26]. The countries with the lowest values are Belgium [1] and Sweden [16].

h.)

emp3 = principal(employment2, nfactors=3, rotate = "varimax")
print(emp3$loadings, cutoff=0)
> print(emp3$loadings, cutoff=0)
Loadings:

	RC1	RC2	RC3
Agr	-0.833	-0.420	-0.325
Min	0.010	-0.659	0.658
Man	0.719	-0.256	0.489
PS	0.086	0.175	0.911
Con	0.810	-0.081	-0.102
SI	0.559	0.687	0.031
Fin	-0.146	0.785	0.069
SPS	0.536	0.541	0.018

	RC1	RC2	RC3
SS loadings	2.494	2.096	1.625
Proportion Var	0.312	0.262	0.203
Cumulative Var	0.312	0.574	0.777

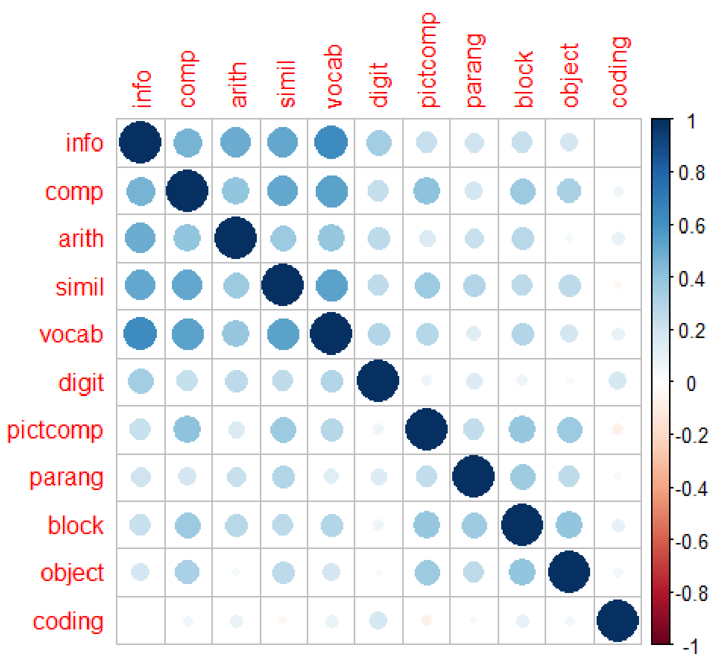
The final number of components to select should be 3.

5.)
a.)

For this dataset, the data does NOT need to be scaled. Each of the 11 variables is of a small integer type.

b.)

```
setwd("C:/Users/Home/Desktop/DePaul/DSC-424-AdvancedDataAnalysis/Week-4/Homework")
smarts = read.csv("wiscsem.csv")
View(smarts)
head(smarts)
tests <- smarts[,c(3:13)]
View(tests)
library(corrplot)
cor.tests = cor(tests)
corrplot(cor.tests)
```




```
> principal(tests, nfactors=4, rotate="none" )
```

Principal Components Analysis
Call: principal(r = tests, nfactors = 4, rotate = "none")
Standardized loadings (pattern matrix) based upon correlation matrix

	PC1	PC2	PC3	PC4	h2	u2	com
info	0.74	-0.35	-0.17	0.01	0.69	0.31	1.6
comp	0.75	0.01	-0.09	-0.26	0.65	0.35	1.3
arith	0.60	-0.35	0.04	0.26	0.56	0.44	2.0
simil	0.74	-0.06	-0.24	0.04	0.62	0.38	1.2
vocab	0.74	-0.28	-0.12	-0.25	0.71	0.29	1.6
digit	0.42	-0.47	0.28	0.16	0.50	0.50	2.9
pictcomp	0.56	0.47	-0.17	-0.12	0.58	0.42	2.3
parang	0.45	0.32	0.25	0.70	0.86	0.14	2.5
block	0.57	0.45	0.26	0.03	0.60	0.40	2.3
object	0.46	0.58	0.14	-0.23	0.63	0.37	2.4
coding	0.10	-0.19	0.86	-0.31	0.89	0.11	1.4

	PC1	PC2	PC3	PC4
SS loadings	3.83	1.44	1.12	0.89
Proportion Var	0.35	0.13	0.10	0.08
Cumulative Var	0.35	0.48	0.58	0.66
Proportion Explained	0.53	0.20	0.15	0.12
Cumulative Proportion	0.53	0.72	0.88	1.00

Mean item complexity = 1.9
Test of the hypothesis that 4 components are sufficient.

The root mean square of the residuals (RMSR) is 0.08
 with the empirical chi square 123.67 with prob < 3.1e-18

Fit based upon off diagonal values = 0.93

The "coding" variable appears that it could be a single-variable factor as it appears to have no correlation with any of the other variables.

c.)

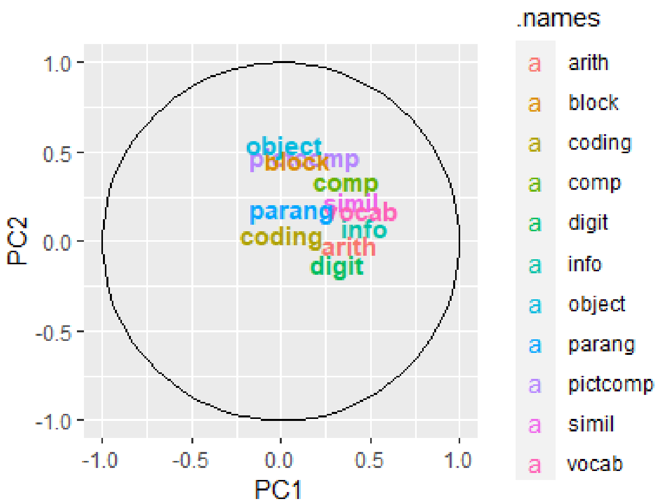
```
testsVarimax = principal(tests, nfactors = 4, rotate = "varimax")
print(testsVarimax)
print(testsVarimax$loadings, cutoff=.4)
```

Loadings:

	RC1	RC2	RC3	RC4
info	0.822			
comp	0.631	0.492		
arith	0.662			
simil	0.681			
vocab	0.785			
digit	0.539			
pictcomp		0.691		
parang				0.882
block		0.660		
object		0.785		
coding			0.941	

	RC1	RC2	RC3	RC4
SS loadings	2.971	2.064	1.124	1.117
Proportion Var	0.270	0.188	0.102	0.102
Cumulative Var	0.270	0.458	0.560	0.662

source("PCA_Plot.R")
PCA_Plot_Psyc(testsVarimax)



The loadings and plot allow us to view which variables are related to which principal components. :

- PC1 would consist of the variables info, comp, arith, simil, vocab, and digit
- PC2 would consist of the variables comp, pictcomp, block and object
- PC3 would consist of the variable coding
- PC4 would consist of the variable parang

d.)

print(testsVarimax$loadings, cutoff=.4, sort=T)

Loadings:

	RC1	RC2	RC3	RC4
info	0.822			
comp	0.631	0.492		
arith	0.662			
simil	0.681			
vocab	0.785			
digit	0.539			
pictcomp		0.691		
block		0.660		
object		0.785		
coding			0.941	
parang				0.882

	RC1	RC2	RC3	RC4
SS loadings	2.971	2.064	1.124	1.117
Proportion Var	0.270	0.188	0.102	0.102
Cumulative Var	0.270	0.458	0.560	0.662

e.)

```
fit = factanal(tests, 4)
print(fit)
print(fit$loadings, cutoff=.4, sort=T)

> print(fit$loadings, cutoff=.4, sort=T)

Loadings:
         Factor1  Factor2  Factor3  Factor4
info      0.788
comp      0.529    0.507
arith     0.570
simil     0.586
vocab     0.725
pictcomp           0.617
block              0.566
object             0.595
coding                      0.994
parang                               0.948
digit     0.438

               Factor1 Factor2 Factor3 Factor4
SS loadings      2.380   1.640   1.041   0.983
Proportion Var   0.216   0.149   0.095   0.089
Cumulative Var   0.216   0.365   0.460   0.549
```

The differences between the two methods are minor. While the Common Factor Analysis makes it a bit easier to choose which factors to select, the Principal Component Analysis method may be better if our goal and focus were on dimension reduction.

7.)

a.)
In this article, it seems that PCA is suitable, because it states that the data can consider genes as variables, the experiements as variables or both. We know that there are thousands of genes, so in this case, PCA may be applicable. However, the report states that the experiments will be used as variables. It seems that their goal is to extract interpretable underlying variables.

b.)
If I read the article correctly, instead of scaling they are simply discarding all components that account for less than (70/n)% of the variance.

c.)
I'm unsure but maybe an orthogonal rotation.

d.)
The article states that the first two components account for over 90% of the variance. However, a third component adds and additional 5% of variance for a total of 95%. It appears that the criteria used were the Eigenvalues that were about the 10% cutoff.

e.)
To evaluate the stability of the PCs, for PC1 the article states that the first component is an average calculation weighted by the variance of an experiment. For PC2, the article explains how the change in expression over time is calculated.

f.)
In summary, the article states that "*PCA on the time-points suggests that much of the observed variability in the experiment can be summarized in just 2 components*."