MSiA 421 – Data Mining Final
Steven Lin

Problem 1

Note: after inspecting the data, I decided NOT to standardize because all the variables are genes and it makes sense they are measured in some common scale (have the same commensurate units). A quick summary of the data also reveals that the ranges for the values of the gene variables are very close.

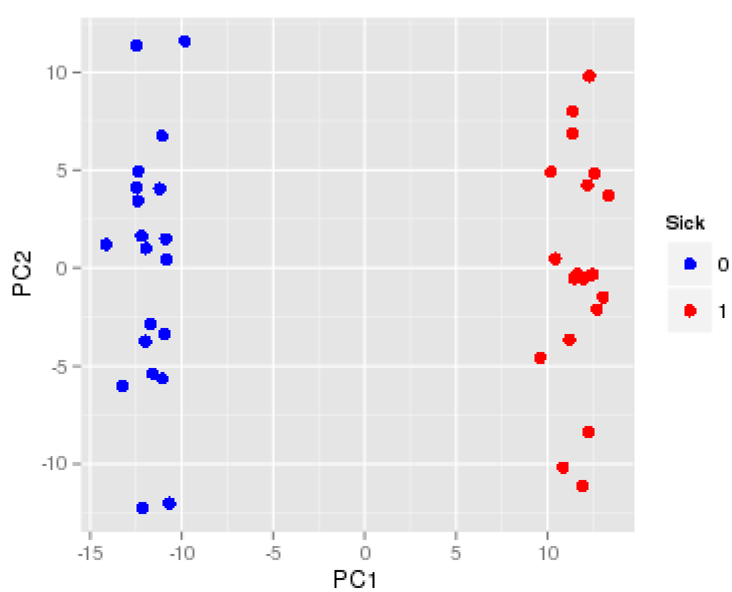
Part a

The proportion of variance explained by first 2 PCS is 15.93% (*note: the answer if scaled is 11.5%*)

```
> summary(fit_pca)$imp[,1:2]
                            PC1      PC2
Standard deviation     11.94090 6.068177
Proportion of Variance  0.12668 0.032710
Cumulative Proportion   0.12668 0.159390
```

Part b

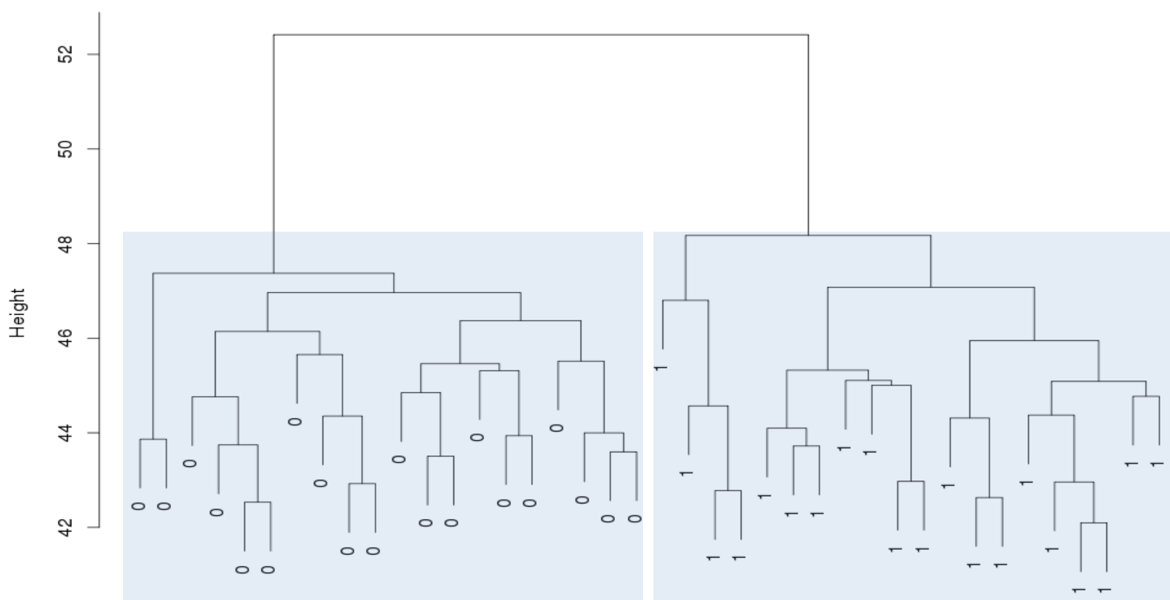
The scatter plot reveals that PC1 can separate apart healthy from sick people, in which low values of PC1 correspond to healthy (blue) and high values of PC1 corresponding to sick people (red). PC2 does not seem to differentiate these two groups of people (there seems to be similar variation along PC2 for both healthy and sick groups).



Part c

The dendogram shows that the two-cluster solution separates healthy from sick people (the left cluster contains all and only healthy, while the right cluster contains all and only sick people)

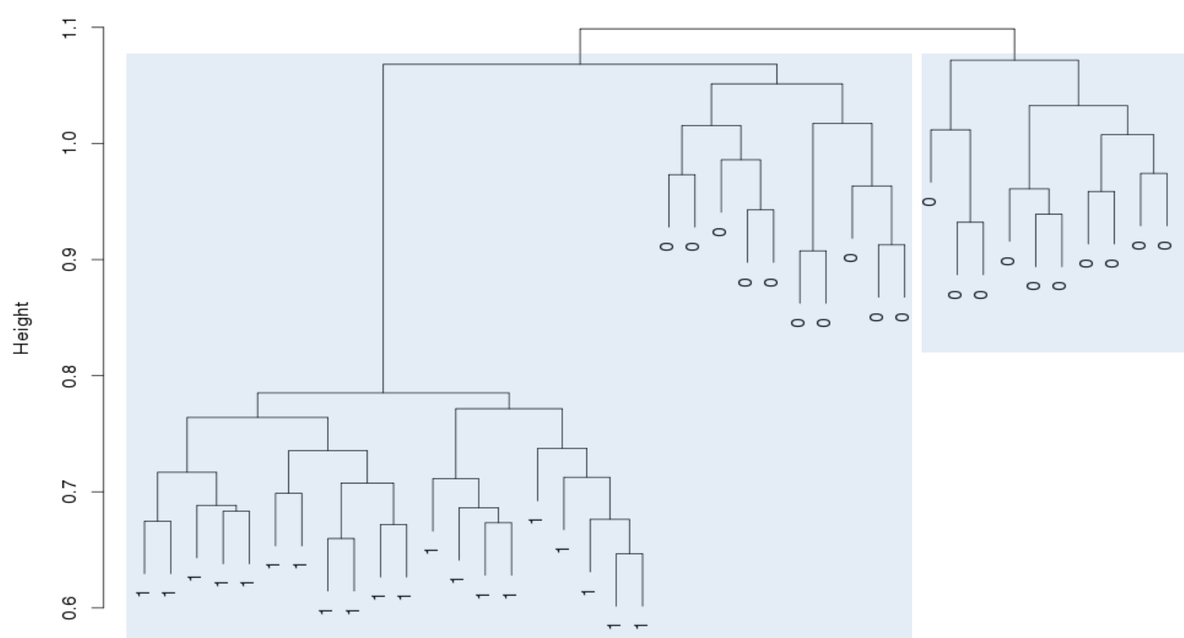
Compete Linkage (Euclidean metric)



Part d

The dendogram shows that the 2-cluster solution does NOT separate healthy from sick people (the left cluster contains all sick and some healthy, while the right cluster contains the rest of the healthy people)

Compete Linkage (Correlation metric)



Part e

The cross tabulation between the cluster assignment versus sick show that the two-cluster solution separates healthy from sick people (cluster1 contains all and only sick, while cluster2 contains all and only healthy people).

```
> crosstab = table(fit_kmeans$cluster,sick)
> rownames(crosstab) = c("cluster1","cluster2")
> colnames(crosstab) = c("healthy","sick")
> crosstab
           healthy   sick
  cluster1       0     20
  cluster2      20      0
```

Part f

i) Based on the scatterplot in part (b), the recommended variance structure is EEI.

ii) The reason is that the two clusters seem to have equal shape and volume, with the distribution being diagonal oriented along the coordinate axes. Also note that the number of parameters grows quickly with the number of clusters and variables, so a lower-complexity model, such as EEI, is recommended when possible.

iii) The parameters for the size, center and dispersion of the clusters are given below. The proportion 50/50 indicates equal size (20 observations in each cluster). The means of PC1 and PC2 for cluster 1 are the negative of those for cluster 2, with the difference in cluster means being greater for PC1 compared to PC2. The covariance matrix is the same for cluster 1 and 2, with PC1 showing a smaller variance than PC2. The covariance between PC1 and PC2 is zero.

```
> fit_mclust$parameter$pro
[1] 0.5 0.5
> fit_mclust$parameter$mean
            [,1]        [,2]
PC1 -11.75209258 11.75209258
PC2   0.03970815 -0.03970815
> fit_mclust$parameter$variance$sigma
, , 1

          PC1      PC2
PC1 0.9087589  0.00000
PC2 0.0000000 35.90063

, , 2

          PC1      PC2
PC1 0.9087589  0.00000
PC2 0.0000000 35.90063
```

iv) The Gaussian mixture seems to separate diseased from healthy. This can be seen in the results shown below from the clustering, where the first 20 observations (healthy) were assigned to cluster 1 (blue), and the second set of 20 observations (sick) were assigned to cluster 2 (red).

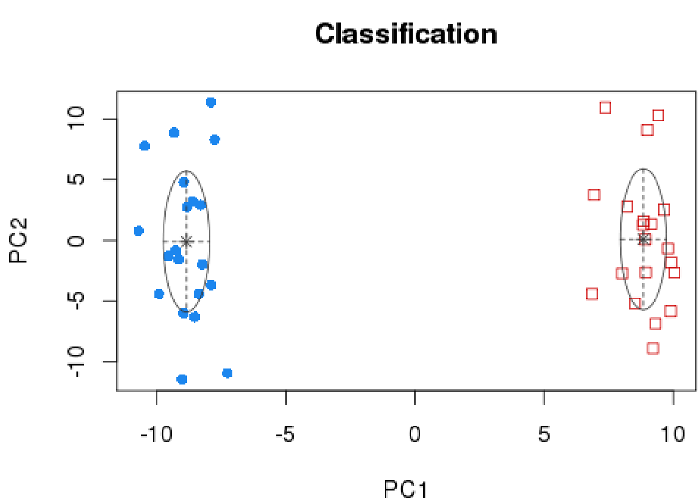
```
> summary(fit_mclust)
----------------------------------------------------
Gaussian finite mixture model fitted by EM algorithm
----------------------------------------------------

Mclust EEI (diagonal, equal volume and shape) model with 2 components:

 log.likelihood  n df       BIC       ICL
      -206.9913 40  7 -439.8048 -439.8048

Clustering table:
 1  2
20 20
```

```
> fit_mclust$classification
 V1  V2  V3  V4  V5  V6  V7  V8  V9 V10 V11 V12 V13 V14 V15 V16 V17 V18 V19 V20 V
21 V22 V23 V24 V25 V26 V27 V28 V29 V30 V31 V32 V33 V34 V35 V36
  1   1   1   1   1   1   1   1   1   1   1   1   1   1   1   1   1   1   1   1
2   2   2   2   2   2   2   2   2   2   2   2   2   2   2   2
V37 V38 V39 V40
  2   2   2   2
```



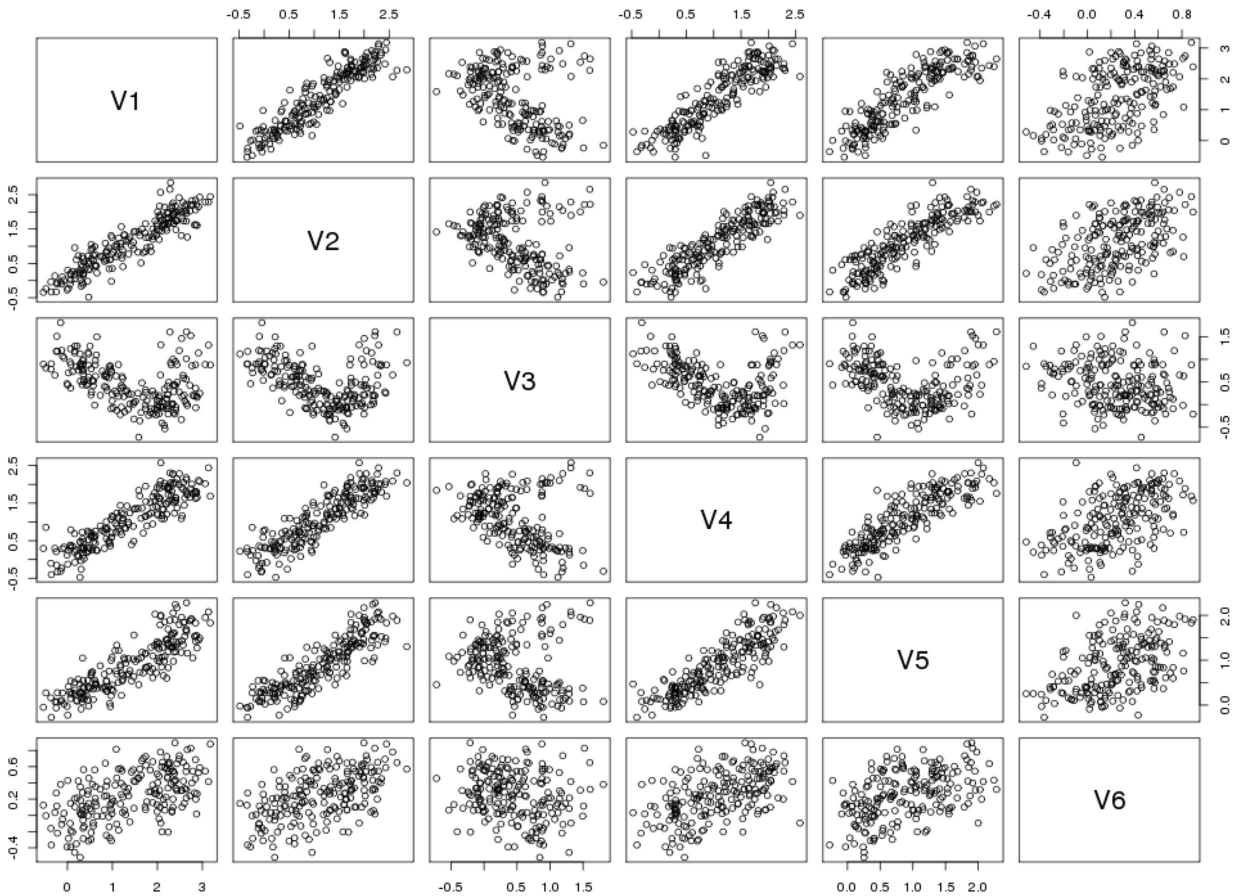
v) If the goal is to separate diseased from healthy, PC1 is the only principal component need as it captures the variation. Running the Gaussian Mixture model with only PC1 still shows the model can perfectly separate diseased from healthy. If other characteristics are important, then other principal components can be used as they can reveal additional features of patients based on genes and contrasts between certain genes.

Problem 2

Part a

This is a list of some of the most important characteristics from the scatterplot:

- V1, V2, V4, V5 look like they have a linear relationship among each other pairwise (e.g. V1 & V2, V1 & V4, V1 & V5, V2 & V4, V2 & V5, V4 & V5)
- V3 seems to have a non-linear relationship (possibly quadratic) with each of V1,V2,V4,V5 (e.g. V3 & V1, V3 &V2, V3 & V4, V3 & V5).
- V6 does not seem to have any relationship with any other variable



Part b

Note: for all the answers here, the data was NOT scaled and NOT centered

The sum of squared errors from the estimated principal curves models is 59.81782.

Part c

The distance along the curve (arc length) between observations 1 and 2 (assume this refers to observations in the data and not points 1 and 2 in the curve) is 0.1966599.

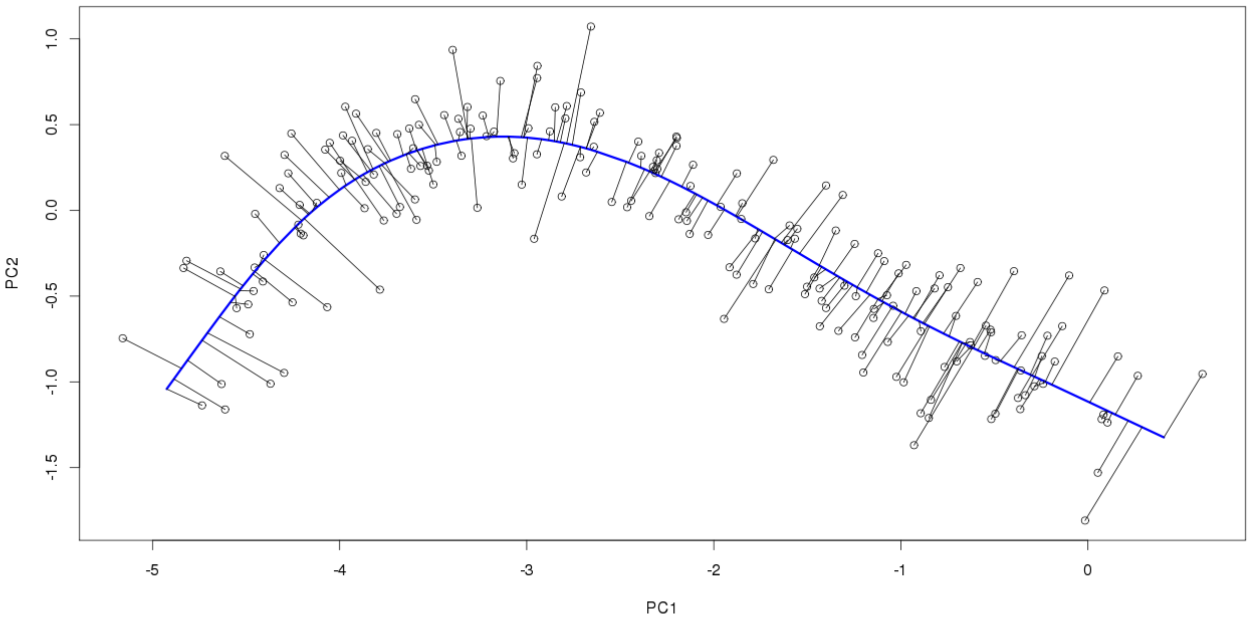
Part d

The fraction of variance is accounted for by the first PCs is 97 % .

```
> fit_pca2 = prcomp(pcurve,scale=F,center=F)
> summary(fit_pca2)$imp[,1:2]
                            PC1       PC2
Standard deviation     2.746863 0.6070504
Proportion of Variance 0.924900 0.0451700
Cumulative Proportion  0.924900 0.9700800
```

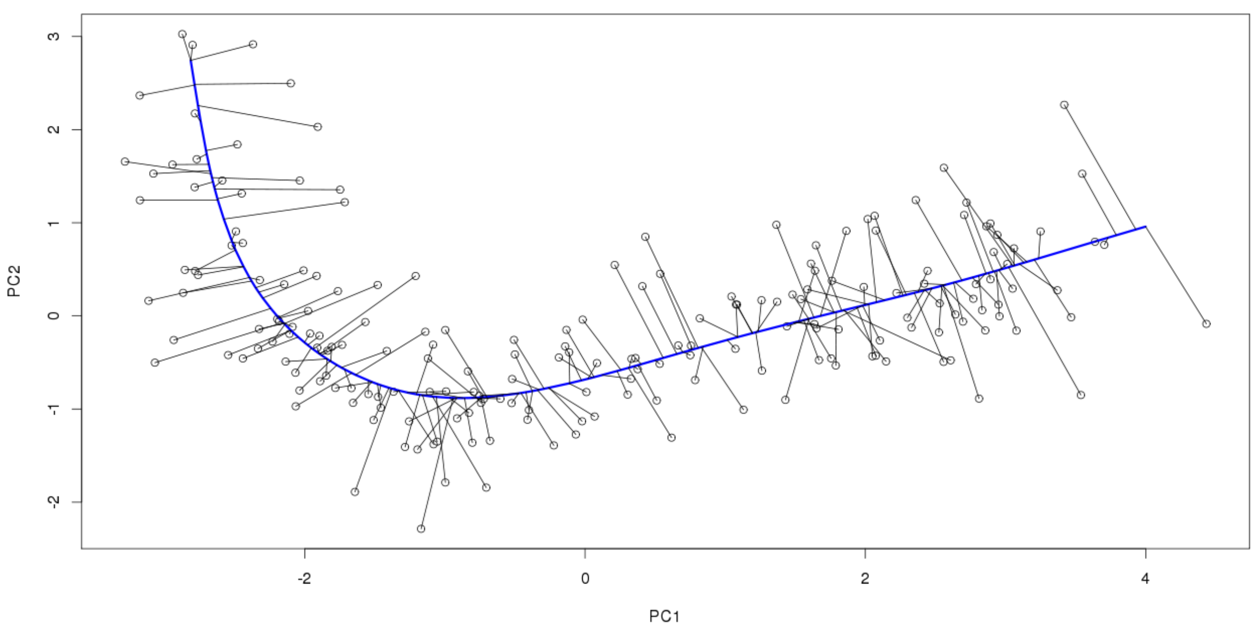
Part e

```
> plot(fit_pca2$x[,"PC1"],fit_pca2$x[,"PC2"],
+      xlab="PC1",
+      ylab="PC2")
>
> # Part e
> dim(fit_pcurve$s)            # fitted curves for 6 dimesions
[1] 200   6
> dim(fit_pca2$rotation[,1:2]) # vectors of PC1 and PC2 from PCA
[1] 6 2
>
> # project curves (points <PC1,PC2> ) into subpace of the PCA
   (vectors of PC1 an   d PC2 of PCA)
> proj = fit_pcurve$s %*% fit_pca2$rotation[,1:2] # projection onto PC1 and PC2
                                                    from PCA
>
> # order by scores (tag has index of small to largest of scores) when plotting
    line
> lines(proj[fit_pcurve$tag,],col="blue",lwd=3)
>
> # from : PC scores of PCA, to: projected points of principal curve into
    PCA sub  space
> segments(fit_pca2$x[,"PC1"],
+          fit_pca2$x[,"PC2"],
+          proj[,"PC1"],
+          proj[,"PC2"])
```



Just in case, here are the answers for scaled and centered: b) 233.8939, c) 0.3250526, d) 83.75%

e) Same code, different graph:



Problem 3

This question was done in SAS using method = P (PCA) for the PROC FACTOR.

Part a

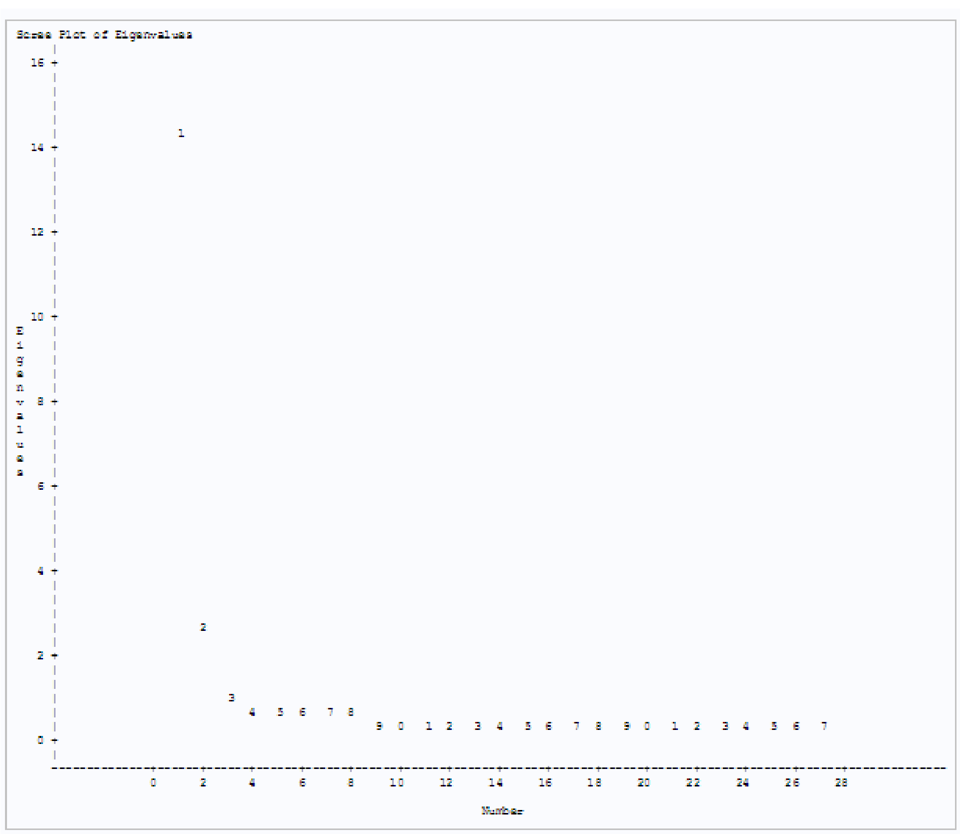
Using the “eigenvalue-greater-than-one” rule (Kaiser criterion), it seems that the number of factors should be 3-5 (the eigenvalue of 5 is close to 4). The scree plot shows that the elbow occurs with number of factors equal to 3. Running PCA for N=3,4 and 5 shows that factors 4 and 5 do not have any questions withe loading (absolute value) greater than 0.4 in the rotated axes using varimax. Therefore, the chosen number of factors is 3.

The FACTOR Procedure
Initial Factor Method: Principal Components

Prior Communality Estimates: ONE

Eigenvalues of the Correlation Matrix: Total = 27 Average = 1

	Eigenvalue	Difference	Proportion	Cumulative
1	14.3878122	11.8256951	0.5329	0.5329
2	2.5621172	1.4922098	0.0949	0.6278
3	1.0699073	0.3211021	0.0396	0.6674
4	0.7488053	0.0387638	0.0277	0.6951
5	0.7100415	0.0316233	0.0263	0.7214
6	0.6784182	0.1211723	0.0251	0.7466
7	0.5572459	0.0384308	0.0206	0.7672
8	0.5188151	0.0319861	0.0192	0.7864
9	0.4868290	0.0186672	0.0180	0.8044
10	0.4681618	0.0534180	0.0173	0.8218



Part b

Varimax was used to check for small loadings and substantial cross-loadings (large loadings on more than one factor). The following criteria were used to purify (get rid of items that do not fall into construct domain) the measures:

- Content Validity: was ensured by having at least 3 questions per factor so that the construct domain is captured by the measure
- Discriminant Validity: the questions 54, 51, 28, 40, 53, 31 were removed sequentially since they were found to have substantial cross loadings and large cross loadings diminish discriminant validity. Question 44 was not removed because at least 3 questions are required per factor, and removing question 44 also create additional cross loadings.
- Reliability: was evaluated by looking at Cronbach's coefficient alpha. None of the variables increased alpha if removed, and alpha was ensured to be greater than 0.8 for all factors.
- Face validity: was evaluated by looking if all items in the factors made sense. Questions 35 and 48 do not seem to make sense in factor 2. Similarly, Q 47 was removed from factor 1.

Questions number assigned to each factor:

- Factor 1: V29, V30, V32, V33, V34, V43, V46, V50, V52
- Factor 2: V36, V37, V38, V39, V41, V42, V44
- Factor 3: V44, V45, V49

Note that there is one cross loading in question 44, but the question seems to make sense be relevant for the measuring of factor 2 and 3 (more for factor 3).

Part c

- Factor 1: expression and social sharing
- Factor 2: relaxation and timeout therapy
- Factor 3: deep emotional effect

Part d

Factor 1: Cronbach alpha = 0.931. Deleting any variable will not cause an increase in alpha.

Cronbach Coefficient Alpha	
Variables	Alpha
Raw	0.931438
Standardized	0.931625

Cronbach Coefficient Alpha with Deleted Variable					
	Raw Variables		Standardized Variables		
Deleted Variable	Correlation with Total	Alpha	Correlation with Total	Alpha	Label
V29	0.741933	0.923814	0.743643	0.923863	I like others to know which music I listen to.
V30	0.743414	0.923666	0.745844	0.923730	play songs with others so they understand me
V32	0.808631	0.919612	0.808614	0.919880	give advice/rec to my friends about songs.
V33	0.760729	0.922576	0.758459	0.922961	wear t-shirts name of my favorite musicians.
V34	0.737113	0.924250	0.734370	0.924426	I often Like musicians on Facebook.
V43	0.767464	0.922196	0.766707	0.922457	my role keep friends informed about new music
V46	0.764688	0.922341	0.761844	0.922754	like discuss music on social media sites
V50	0.680820	0.927323	0.682208	0.927568	Being fan of bands like belonging to a club.
V52	0.710111	0.925671	0.712374	0.925756	knowledge of music makes me interesting

Factor 2: Cronbach alpha = 0.937. Deleting Q44 can cause marginal increase in alpha, but this would cause increase in cross loadings.

Cronbach Coefficient Alpha	
Variables	Alpha
Raw	0.937139
Standardized	0.936963

Cronbach Coefficient Alpha with Deleted Variable					
	Raw Variables		Standardized Variables		
Deleted Variable	Correlation with Total	Alpha	Correlation with Total	Alpha	Label
V36	0.832323	0.923959	0.831635	0.923735	I feel energized after listening to music.
V37	0.806884	0.926378	0.806648	0.926051	lose myself in pleasure of listening to music.
V39	0.827407	0.924371	0.826930	0.924173	Listening to music is an escape.
V38	0.843449	0.922841	0.843001	0.922675	I often unwind/relax by listening to music
V41	0.847201	0.922688	0.847131	0.922289	I feel less stress after listening to music.
V42	0.767433	0.929906	0.767205	0.929671	When I listen to music I am worry-free.
V44	0.640817	0.940694	0.640926	0.940961	Music sometimes touches me deep down.

Factor 3: Cronbach alpha = 0.80. Deleting Q44 can cause marginal increase in alpha, but this would cause increase in cross loadings.

Cronbach Coefficient Alpha	
Variables	Alpha
Raw	0.808232
Standardized	0.814108

Cronbach Coefficient Alpha with Deleted Variable					
	Raw Variables		Standardized Variables		
Deleted Variable	Correlation with Total	Alpha	Correlation with Total	Alpha	Label
V44	0.702585	0.698489	0.707560	0.700479	Music sometimes touches me deep down.
V45	0.704322	0.687632	0.713449	0.694249	songs send shivers up my spine/goose bumps.
V49	0.579110	0.828723	0.579030	0.830194	sometimes cry after listening to certain songs.

Part e

Varimax rotated loadings of final set of items:

Rotated Factor Pattern				
		Factor1	Factor2	Factor3
V43	my role keep friends informed about new music	0.80793	.	.
V46	like discuss music on social media sites	0.80648	.	.
V33	wear t-shirts name of my favorite musicians.	0.80215	.	.
V34	I often Like musicians on Facebook.	0.79656	.	.
V32	give advice/rec to my friends about songs.	0.78038	.	.
V30	play songs with others so they understand me	0.68380	.	.
V29	I like others to know which music I listen to.	0.65018	.	.
V50	Being fan of bands like belonging to a club.	0.64571	.	.
V52	knowledge of music makes me interesting	0.62251	.	.
V38	I often unwind/relax by listening to music	.	0.82572	.
V41	I feel less stress after listening to music.	.	0.81018	.
V39	Listening to music is an escape.	.	0.80635	.
V36	I feel energized after listening to music.	.	0.76806	.
V42	When I listen to music I am worry-free.	.	0.75352	.
V37	lose myself in pleasure of listening to music.	.	0.72401	.
V45	songs send shivers up my spine/goose bumps.	.	.	0.76873
V49	sometimes cry after listening to certain songs.	.	.	0.74690
V44	Music sometimes touches me deep down.	.	0.47495	0.72953
Values less than 0.4 are not printed.				

Eigenvalues:

Eigenvalues of the Correlation Matrix: Total = 18 Average = 1				
	Eigenvalue	Difference	Proportion	Cumulative
1	9.84127408	7.77599688	0.5467	0.5467
2	2.06527720	1.10538420	0.1147	0.6615
3	0.95989300	0.34355860	0.0533	0.7148
4	0.61633440	0.09249789	0.0342	0.7490
5	0.52383651	0.04296844	0.0291	0.7781
6	0.48086808	0.04530530	0.0267	0.8049
7	0.43556278	0.05615083	0.0242	0.8291
8	0.37941194	0.01975173	0.0211	0.8501
9	0.35966021	0.02799678	0.0200	0.8701
10	0.33166343	0.02406842	0.0184	0.8885
11	0.30759501	0.02483517	0.0171	0.9056
12	0.28275984	0.00123323	0.0157	0.9213
13	0.28152661	0.02147163	0.0156	0.9370
14	0.26005498	0.01608457	0.0144	0.9514
15	0.24397041	0.01680335	0.0136	0.9650
16	0.22716706	0.01488187	0.0126	0.9776
17	0.21228519	0.02142590	0.0118	0.9894
18	0.19085929		0.0106	1.0000

Variance Explained by Each Factor		
Factor1	Factor2	Factor3
9.8412741	2.0652772	0.9598930

Part f

Estimate scores for factors (promax):

The FACTOR Procedure
Rotation Method: Promax (power = 3)

Target Matrix for Procrustean Transformation				
		Factor1	Factor2	Factor3
V43	my role keep friends informed about new music	0.94766	.	.
V46	like discuss music on social media sites	0.98358	.	.
V33	wear t-shirts name of my favorite musicians.	0.96724	.	.
V34	I often Like musicians on Facebook.	1.00000	.	.
V32	give advice/rec to my friends about songs.	0.83332	.	.
V30	play songs with others so they understand me	0.67565	.	.
V29	I like others to know which music I listen to.	0.57551	.	.
V50	Being fan of bands like belonging to a club.	0.66095	.	.
V52	knowledge of music makes me interesting	0.56801	.	.
V38	I often unwind/relax by listening to music	.	1.00000	.
V41	I feel less stress after listening to music.	.	0.95915	.
V39	Listening to music is an escape.	.	0.97539	.
V36	I feel energized after listening to music.	.	0.86280	.
V42	When I listen to music I am worry-free.	.	0.88885	.
V37	lose myself in pleasure of listening to music.	.	0.77784	.
V45	songs send shivers up my spine/goose bumps.	.	.	0.96858
V49	sometimes cry after listening to certain songs.	.	.	1.00000
V44	Music sometimes touches me deep down.	.	.	0.79768
Values less than 0.4 are not printed.				

Correlation matrix :

Inter-Factor Correlations			
	Factor1	Factor2	Factor3
Factor1	1.00000	0.55991	0.41550
Factor2	0.55991	1.00000	0.58477
Factor3	0.41550	0.58477	1.00000

Part g

Using varimax: All factors are significant predictors of consumption (time1) since p-values < 0.05. Factor 1 is the most predictive, followed by factor 2 and 3. However, the R2 is very low.

The REG Procedure
Model: MODEL1
Dependent Variable: time1 Time today music

Number of Observations Read	1278
Number of Observations Used	1278

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	3	10683	3561.06995	111.89	<.0001
Error	1274	40548	31.82699		
Corrected Total	1277	51231			

Root MSE	5.64154	R-Square	0.2085
Dependent Mean	5.83881	Adj R-Sq	0.2067
Coeff Var	96.62141		

Parameter Estimates						
Variable	Label	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	Intercept	1	5.83881	0.15781	37.00	<.0001
Factor1		1	2.36602	0.15787	14.99	<.0001
Factor2		1	1.45934	0.15787	9.24	<.0001
Factor3		1	0.79884	0.15787	5.06	<.0001

Using promax: Factors 1 and 2 are significant predictors of consumption (time1) since p-values <0.05. Factor 1 is the most predictive, followed by factor 2. Factor 3 is not a significant predictor of consumption since p-value > 0.05. However, the R2 is very low.

The REG Procedure
Model: MODEL1
Dependent Variable: time1 Time today music

Number of Observations Read	1278
Number of Observations Used	1278

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	3	10683	3561.06995	111.89	<.0001
Error	1274	40548	31.82699		
Corrected Total	1277	51231			

Root MSE	5.64154	R-Square	0.2085
Dependent Mean	5.83881	Adj R-Sq	0.2067
Coeff Var	96.62141		

Parameter Estimates						
Variable	Label	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	Intercept	1	5.83881	0.15781	37.00	<.0001
Factor1		1	2.16401	0.19220	11.26	<.0001
Factor2		1	0.90235	0.21551	4.19	<.0001
Factor3		1	0.22177	0.19631	1.13	0.2588

Part h

PROC CALIS using the promax factors was run giving the same results as before. Factors 1 and 2 are significant predictors of consumption (time1) since p-values <0.05. Factor 1 is the most predictive, followed by factor 2. Factor 3 is not a significant predictor of consumption since p-value > 0.05.

The CALIS Procedure
Covariance Structure Analysis: Maximum Likelihood Estimation

Linear Equations																
time1	=		2.1640	(**)	Factor1	+	0.9024	(**)	Factor2	+	0.2218	(ns)	Factor3	+	1.0000	e

Effects in Linear Equations						
Variable	Predictor	Parameter	Estimate	Standard Error	t Value	Pr > |t|
time1	Factor1	b1	2.16401	0.20172	10.7278	<.0001
time1	Factor2	b2	0.90235	0.16277	5.5437	<.0001
time1	Factor3	b3	0.22177	0.20633	1.0748	0.2825

If a CFA/SEM model with PROC CALIS instead of the regression in the previous part is used with the original input dataset, the results change. Bentler's CFI > .9 indicates that the fit is acceptable. The coefficients of the predictors change. Factor 1 remains the most important predictor, followed by Factor 2 and 3. However, Factor 2 now is insignificant at 0.05 level (borderline significant at 0.10 level). Factor 3 remains insignificant.

The CALIS Procedure
Covariance Structure Analysis: Maximum Likelihood Estimation

Fit Summary		
Modeling Info	Number of Observations	1278
	Number of Variables	19
	Number of Moments	190
	Number of Parameters	43
	Number of Active Constraints	0
	Baseline Model Function Value	14.2349
	Baseline Model Chi-Square	18178.0063
	Baseline Model Chi-Square DF	171
	Pr > Baseline Model Chi-Square	<.0001
Absolute Index	Fit Function	0.9622
	Chi-Square	1228.7703
	Chi-Square DF	147
	Pr > Chi-Square	<.0001
	Z-Test of Wilson & Hilferty	26.5167
	Hoelter Critical N	184
	Root Mean Square Residual (RMR)	0.1060
	Standardized RMR (SRMR)	0.0548
	Goodness of Fit Index (GFI)	0.8961
Parsimony Index	Adjusted GFI (AGFI)	0.8657
	Parsimonious GFI	0.7703
	RMSEA Estimate	0.0759
	RMSEA Lower 90% Confidence Limit	0.0720
	RMSEA Upper 90% Confidence Limit	0.0799
	Probability of Close Fit	<.0001
	ECVI Estimate	1.0306
	ECVI Lower 90% Confidence Limit	0.9454
	ECVI Upper 90% Confidence Limit	1.1219
	Akaike Information Criterion	1314.7703
	Bozdogan CAIC	1579.3515
	Schwarz Bayesian Criterion	1536.3515
	McDonald Centrality	0.6549
Incremental Index	Bentler Comparative Fit Index	0.9399
	Bentler-Bonett NFI	0.9324
	Bentler-Bonett Non-normed Index	0.9301
	Bollen Normed Index Rho1	0.9214
	Bollen Non-normed Index Delta2	0.9400
	James et al. Parsimonious NFI	0.8015

The CALIS Procedure
Covariance Structure Analysis: Maximum Likelihood Estimation

Linear Equations

time1	=		2.4034	(**)	f1	+	0.5747	(ns)	f2	+	0.1694	(ns)	f3	+	1.0000		e19
V29	=		0.9502	(**)	f1	+	1.0000		e1								
V30	=		1.0423	(**)	f1	+	1.0000		e2								
V32	=		1.1022	(**)	f1	+	1.0000		e3								
V33	=		1.0403	(**)	f1	+	1.0000		e4								
V34	=		1.0671	(**)	f1	+	1.0000		e5								
V43	=		1.0175	(**)	f1	+	1.0000		e6								
V46	=		1.0622	(**)	f1	+	1.0000		e7								
V50	=		0.8882	(**)	f1	+	1.0000		e8								
V52	=		0.8907	(**)	f1	+	1.0000		e9								
V36	=		0.9841	(**)	f2	+	1.0000		e10								
V37	=		1.0247	(**)	f2	+	1.0000		e11								
V38	=		1.0198	(**)	f2	+	1.0000		e12								
V39	=		0.9935	(**)	f2	+	1.0000		e13								
V41	=		0.9781	(**)	f2	+	1.0000		e14								
V42	=		0.9335	(**)	f2	+	1.0000		e15								
V44	=		0.9401	(**)	f3	+	1.0000		e16								
V45	=		0.9777	(**)	f3	+	1.0000		e17								
V49	=		0.8537	(**)	f3	+	1.0000		e18								

Effects in Linear Equations

Variable	Predictor	Parameter	Estimate	Standard Error	t Value	Pr > |t|
time1	f1	b1	2.40337	0.25000	9.6135	<.0001
time1	f2	b2	0.57471	0.35192	1.6331	0.1025
time1	f3	b3	0.16936	0.30594	0.5536	0.5799
V29	f1	p1	0.95024	0.02917	32.5747	<.0001
V30	f1	p2	1.04227	0.03180	32.7778	<.0001
V32	f1	p3	1.10224	0.02986	36.9166	<.0001
V33	f1	p4	1.04027	0.03149	33.0365	<.0001
V34	f1	p5	1.06710	0.03378	31.5893	<.0001
V43	f1	p6	1.01754	0.03022	33.6760	<.0001
V46	f1	p7	1.06220	0.03212	33.0663	<.0001
V50	f1	p8	0.88820	0.03117	28.4976	<.0001
V52	f1	p9	0.89072	0.02891	30.8065	<.0001
V36	f2	p10	0.98415	0.02567	38.3389	<.0001
V37	f2	p11	1.02466	0.02785	36.7887	<.0001
V38	f2	p12	1.01978	0.02593	39.3265	<.0001
V39	f2	p13	0.99354	0.02600	38.2194	<.0001
V41	f2	p14	0.97812	0.02513	38.9195	<.0001
V42	f2	p15	0.93354	0.02708	34.4690	<.0001
V44	f3	p16	0.94014	0.02655	35.4045	<.0001
V45	f3	p17	0.97767	0.02915	33.5369	<.0001
V49	f3	p18	0.85368	0.03460	24.6763	<.0001

Code for extra credit:

```
proc calis data=SL.music;
lineqs
	time1 = b1 f1 + b2 f2 + b3 f3 + e19,
	V29 = p1 f1 + e1,
	V30 = p2 f1 + e2,
	V32 = p3 f1 + e3,
	V33 = p4 f1 + e4,
	V34 = p5 f1 + e5,
	V43 = p6 f1 + e6,
	V46 = p7 f1 + e7,
	V50 = p8 f1 + e8,
	V52 = p9 f1 + e9,

	V36 = p10 f2 + e10,
	V37 = p11 f2 + e11,
	V38 = p12 f2 + e12,
	V39 = p13 f2 + e13,
	V41 = p14 f2 + e14,
	V42 = p15 f2 + e15,

	V44 = p16 f3 + e16,
	V45 = p17 f3 + e17,
	V49 = p18 f3 + e18;
std
	e1-e19 = vare1-vare19,
	f1 =1,
	f2 =1,
	f3 =1;
cov
	f1 f2= covf1f2,
	f1 f3 = covf1f3,
	f2 f3 = covf2f3;
var
	time1
	V29 V30 V32 V33 V34 V43 V46 V50 V52
				V36 V37 V38 V39 V41 V42
				V44 V45 V49
;
run;
```