PCA provides a tool to find a low-dimensional representation for a data set that contains as much of the variation as possible. It aims to provide a low-dimensional representation of the data that captures as much the information as possible. The first principal component of a data set has the largest variance.

> apply(dispute, 2, var)

Medical Surgical Pediatric Gynecologic

37.877470 135.450593 3.173913 2.857708

The variance of Surgical is 135.45093, while the variance of medical and pediatric and gynecological is respectively 37,877470; 3.173913; 2.857708.

The surgical data isn’t necessarily more variable, it is simply on a different scale relative to others.

To calculate principal components, we stat by using the cov( ) function to calculate the covariance matrix, then we use eigen command to calculate the eigenvalues of the matrix. Eigen produces an object that contains both the ordered eigenvalues ($values) and the corresponding eigenvector matrix($vectors). We will take the first two sets of loadings and store them in the matrix phi

dispute.cov <- cov(scaled\_df)

> dispute.eigen <- eigen(dispute.cov)

> str(dispute.eigen)

List of 2

$ values : num [1:4] 2.55 0.738 0.439 0.273

$ vectors: num [1:4, 1:4] -0.43 -0.518 -0.545 -0.499 0.798 ...

- attr(\*, "class")= chr "eigen"

Eigenvectors that are calculated in any software package are unique u p to a sign flip. By default, eigenvectors in R point in the negative direction. We would prefer the eigenvectors point in the positive direction because it leads to more logical interpretation of graphical results. To use the positive- pointing vector, we multiply the default loadings by -1.

The set of loadings for the first principal component(PC1) and second principal component(PC2) are shown below:

> phi <- -phi

> row.names(phi) <- c("Medical","Surgical","Pediatric","Gynecologic")

> colnames(phi) <- c("PC1","PC2")

> phi

PC1 PC2

Medical 0.4303044 -0.79758622

Surgical 0.5184091 -0.05858001

Pediatric 0.5446281 0.15369687

Gynecologic 0.4994700 0.58034634

> PC1 <- as.matrix(scaled\_df)%\*%phi[,1]

> PC2 <- as.matrix(scaled\_df)%\*%phi[,2]

> PC <- data.frame(Factors = row.names(dispute), PC1,PC2)

> PC

Factors PC1 PC2

1 Malpractise -0.6880152 -0.03697162

2 Reoperation -1.0066844 -0.48686032

3 Consultation -1.0257465 -0.72897236

4 Malfunction -0.8948329 0.72729344

5 Option -0.3136041 0.43071321

6 Inspection -0.2395409 0.16559160

7 Tomography -0.4877384 0.43790386

8 Indication -1.7379918 0.32594133

9 Dose -0.8481234 0.49111366

10 Bill 1.2958187 -1.09154125

11 Hospitalization 0.5364168 -0.77482753

12 Efficiacy 2.7666265 1.31841598

13 Complication -1.2198141 0.64407808

14 Fatality -2.0786576 -0.74154044

15 Adverser -0.8400459 0.46522800

16 Miscommunication 0.8226273 -2.53234677

17 Charge -1.5405425 0.30275168

18 Attitude 0.6912792 1.31305314

19 Bureaucracy -0.6980449 -0.02510776

20 Confusion 2.2442988 -0.95620898

21 Ignorance -0.1758285 0.28008616

22 Expectation 4.7947896 0.50398118

23 Hysteria 0.6433542 -0.03177427

Sort PC1

> PC1sorted <- PC[order(-PC1),]

> PC1sorted

Factors PC1 PC2

22 Expectation 4.7947896 0.50398118

12 Efficiacy 2.7666265 1.31841598

20 Confusion 2.2442988 -0.95620898

10 Bill 1.2958187 -1.09154125

16 Miscommunication 0.8226273 -2.53234677

18 Attitude 0.6912792 1.31305314

23 Hysteria 0.6433542 -0.03177427

11 Hospitalization 0.5364168 -0.77482753

21 Ignorance -0.1758285 0.28008616

6 Inspection -0.2395409 0.16559160

5 Option -0.3136041 0.43071321

7 Tomography -0.4877384 0.43790386

1 Malpractise -0.6880152 -0.03697162

19 Bureaucracy -0.6980449 -0.02510776

15 Adverser -0.8400459 0.46522800

9 Dose -0.8481234 0.49111366

4 Malfunction -0.8948329 0.72729344

2 Reoperation -1.0066844 -0.48686032

3 Consultation -1.0257465 -0.72897236

13 Complication -1.2198141 0.64407808

17 Charge -1.5405425 0.30275168

8 Indication -1.7379918 0.32594133

14 Fatality -2.0786576 -0.74154044

Sort PC2

> PC2sorted <- PC[order(-PC2),]

> PC2sorted

Factors PC1 PC2

12 Efficiacy 2.7666265 1.31841598

18 Attitude 0.6912792 1.31305314

4 Malfunction -0.8948329 0.72729344

13 Complication -1.2198141 0.64407808

22 Expectation 4.7947896 0.50398118

9 Dose -0.8481234 0.49111366

15 Adverser -0.8400459 0.46522800

7 Tomography -0.4877384 0.43790386

5 Option -0.3136041 0.43071321

8 Indication -1.7379918 0.32594133

17 Charge -1.5405425 0.30275168

21 Ignorance -0.1758285 0.28008616

6 Inspection -0.2395409 0.16559160

19 Bureaucracy -0.6980449 -0.02510776

23 Hysteria 0.6433542 -0.03177427

1 Malpractise -0.6880152 -0.03697162

2 Reoperation -1.0066844 -0.48686032

3 Consultation -1.0257465 -0.72897236

14 Fatality -2.0786576 -0.74154044

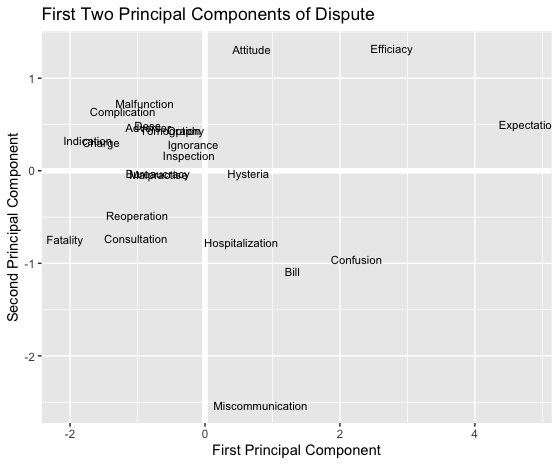
11 Hospitalization 0.5364168 -0.77482753

20 Confusion 2.2442988 -0.95620898

10 Bill 1.2958187 -1.09154125

16 Miscommunication 0.8226273 -2.53234677

Now that we have calculated the first and second principal for each dispute, we can plot them against each other and produce a two-dimensional view of the data.



The first principal component (x-axis) roughly corresponds to the rate of surgical and pediatric. Expectation and Efficiacy have high rate of Surgical and Pediatric, while indication and fatality have low rate of Surgical and Pediatric.

The second principal component (y-axis) roughly corresponds to the rate of pediatric and gynecologic , especially gynecologic. Attitude and Efficiacy have high rate of Surgical and Pediatric, while Miscommunication have low rate of pediatric and gynecologic

Appendix

> Medical <- c(4,7,10,2,6,7,4,1,6,19,15,7,1,6,2,21,2,5,8,18,6,18,13)

> Surgical <- c(21,16,4,6,13,6,12,9,1,15,4,33,14,7,7,28,,5,17,1,31,9,48,10)

Error in c(21, 16, 4, 6, 13, 6, 12, 9, 1, 15, 4, 33, 14, 7, 7, 28, , 5, :

> Pediatric <- c(2,1,2,3,2,4,3,1,2,4,5,7,1,0,4,3,2,3,3,5,4,7,3)

> Gynecologic <- c(1,1,1,2,3,2,2,1,3,3,2,5,2,0,1,0,1,5,2,3,2,7,4)

> dispute <- data.frame(Medical, Surgical, Pediatric, Gynecologic)

Error in data.frame(Medical, Surgical, Pediatric, Gynecologic) :

> Surgical <- c(21,16,4,6,13,6,12,9,1,15,4,33,14,7,7,28,5,17,1,31,9,48,10)

> dispute <- data.frame(Medical, Surgical, Pediatric, Gynecologic)

> dispute

Medical Surgical Pediatric Gynecologic

1 4 21 2 1

2 7 16 1 1

3 10 4 2 1

4 2 6 3 2

5 6 13 2 3

6 7 6 4 2

7 4 12 3 2

8 1 9 1 1

9 6 1 2 3

10 19 15 4 3

11 15 4 5 2

12 7 33 7 5

13 1 14 1 2

14 6 7 0 0

15 2 7 4 1

16 21 28 3 0

17 2 5 2 1

18 5 17 3 5

19 8 1 3 2

20 18 31 5 3

21 6 9 4 2

22 18 48 7 7

23 13 10 3 4

> row.names(dispute) <- c("Malpractise", "Reoperation", "Consultation", "Malfunction", "Option", "Inspection", "Tomography","Indication", "Dose", "Bill", "Hospitalization", "Efficiacy", "Complication", "Fatality", "Adverser", "Miscommunication", "Charge", "Attitude", "Bureaucracy", "Confusion", "Ignorance", "Expectation", "Hysteria")

> dispute

Medical Surgical Pediatric Gynecologic

Malpractise 4 21 2 1

Reoperation 7 16 1 1

Consultation 10 4 2 1

Malfunction 2 6 3 2

Option 6 13 2 3

Inspection 7 6 4 2

Tomography 4 12 3 2

Indication 1 9 1 1

Dose 6 1 2 3

Bill 19 15 4 3

Hospitalization 15 4 5 2

Efficiacy 7 33 7 5

Complication 1 14 1 2

Fatality 6 7 0 0

Adverser 2 7 4 1

Miscommunication 21 28 3 0

Charge 2 5 2 1

Attitude 5 17 3 5

Bureaucracy 8 1 3 2

Confusion 18 31 5 3

Ignorance 6 9 4 2

Expectation 18 48 7 7

Hysteria 13 10 3 4

> apply(dispute, 2, var)

Medical Surgical Pediatric Gynecologic

37.877470 135.450593 3.173913 2.857708

> scaled\_df <- apply(dispute, 2, scale)

> scaled\_df

Medical Surgical Pediatric Gynecologic

[1,] -0.67819239 0.62014007 -0.61011916 -0.7715863

[2,] -0.19074161 0.19052496 -1.17142878 -0.7715863

[3,] 0.29670917 -0.84055131 -0.61011916 -0.7715863

[4,] -1.00315958 -0.66870526 -0.04880953 -0.1800368

[5,] -0.35322520 -0.06724410 -0.61011916 0.4115127

[6,] -0.19074161 -0.66870526 0.51250009 -0.1800368

[7,] -0.67819239 -0.15316713 -0.04880953 -0.1800368

[8,] -1.16564317 -0.41093619 -1.17142878 -0.7715863

[9,] -0.35322520 -1.09832037 -0.61011916 0.4115127

[10,] 1.75906152 0.10460194 0.51250009 0.4115127

[11,] 1.10912714 -0.84055131 1.07380971 -0.1800368

[12,] -0.19074161 1.65121634 2.19642896 1.5946118

[13,] -1.16564317 0.01867892 -1.17142878 -0.1800368

[14,] -0.35322520 -0.58278224 -1.73273840 -1.3631359

[15,] -1.00315958 -0.58278224 0.51250009 -0.7715863

[16,] 2.08402870 1.22160123 -0.04880953 -1.3631359

[17,] -1.00315958 -0.75462828 -0.61011916 -0.7715863

[18,] -0.51570880 0.27644798 -0.04880953 1.5946118

[19,] -0.02825802 -1.09832037 -0.04880953 -0.1800368

[20,] 1.59657792 1.47937030 1.07380971 0.4115127

[21,] -0.35322520 -0.41093619 0.51250009 -0.1800368

[22,] 1.59657792 2.94006168 2.19642896 2.7777108

[23,] 0.78415995 -0.32501317 -0.04880953 1.0030622

> dispute.cov <- cov(scaled\_df)

> dispute.eigen <- eigen(dispute.cov)

> str(dispute.eigen)

List of 2

$ values : num [1:4] 2.55 0.738 0.439 0.273

$ vectors: num [1:4, 1:4] -0.43 -0.518 -0.545 -0.499 0.798 ...

- attr(\*, "class")= chr "eigen"

> phi <- dispute.eigen$vectors[,1:2]

> phi

[,1] [,2]

[1,] -0.4303044 0.79758622

[2,] -0.5184091 0.05858001

[3,] -0.5446281 -0.15369687

[4,] -0.4994700 -0.58034634

> phi <- -phi

> row.names(phi) <- c("Medical","Surgical","Pediatric","Gynecologic")

> colnames(phi) <- c("PC1","PC2")

> phi

PC1 PC2

Medical 0.4303044 -0.79758622

Surgical 0.5184091 -0.05858001

Pediatric 0.5446281 0.15369687

Gynecologic 0.4994700 0.58034634

> PC1 <- as.matrix(scaled\_df)%\*%phi[,1]

> PC2 <- as.matrix(scaled\_df)%\*%phi[,2]

> PC <- data.frame(Factors = row.names(dispute), PC1,PC2)

> PC

Factors PC1 PC2

1 Malpractise -0.6880152 -0.03697162

2 Reoperation -1.0066844 -0.48686032

3 Consultation -1.0257465 -0.72897236

4 Malfunction -0.8948329 0.72729344

5 Option -0.3136041 0.43071321

6 Inspection -0.2395409 0.16559160

7 Tomography -0.4877384 0.43790386

8 Indication -1.7379918 0.32594133

9 Dose -0.8481234 0.49111366

10 Bill 1.2958187 -1.09154125

11 Hospitalization 0.5364168 -0.77482753

12 Efficiacy 2.7666265 1.31841598

13 Complication -1.2198141 0.64407808

14 Fatality -2.0786576 -0.74154044

15 Adverser -0.8400459 0.46522800

16 Miscommunication 0.8226273 -2.53234677

17 Charge -1.5405425 0.30275168

18 Attitude 0.6912792 1.31305314

19 Bureaucracy -0.6980449 -0.02510776

20 Confusion 2.2442988 -0.95620898

21 Ignorance -0.1758285 0.28008616

22 Expectation 4.7947896 0.50398118

23 Hysteria 0.6433542 -0.03177427

> PC1.sorted

[,1]

[1,] 4.7947896

[2,] 2.7666265

[3,] 2.2442988

[4,] 1.2958187

[5,] 0.8226273

[6,] 0.6912792

[7,] 0.6433542

[8,] 0.5364168

[9,] -0.1758285

[10,] -0.2395409

[11,] -0.3136041

[12,] -0.4877384

[13,] -0.6880152

[14,] -0.6980449

[15,] -0.8400459

[16,] -0.8481234

[17,] -0.8948329

[18,] -1.0066844

[19,] -1.0257465

[20,] -1.2198141

[21,] -1.5405425

[22,] -1.7379918

[23,] -2.0786576

> PC1sorted <- PC[order(-PC1),]

> PC1sorted

Factors PC1 PC2

22 Expectation 4.7947896 0.50398118

12 Efficiacy 2.7666265 1.31841598

20 Confusion 2.2442988 -0.95620898

10 Bill 1.2958187 -1.09154125

16 Miscommunication 0.8226273 -2.53234677

18 Attitude 0.6912792 1.31305314

23 Hysteria 0.6433542 -0.03177427

11 Hospitalization 0.5364168 -0.77482753

21 Ignorance -0.1758285 0.28008616

6 Inspection -0.2395409 0.16559160

5 Option -0.3136041 0.43071321

7 Tomography -0.4877384 0.43790386

1 Malpractise -0.6880152 -0.03697162

19 Bureaucracy -0.6980449 -0.02510776

15 Adverser -0.8400459 0.46522800

9 Dose -0.8481234 0.49111366

4 Malfunction -0.8948329 0.72729344

2 Reoperation -1.0066844 -0.48686032

3 Consultation -1.0257465 -0.72897236

13 Complication -1.2198141 0.64407808

17 Charge -1.5405425 0.30275168

8 Indication -1.7379918 0.32594133

14 Fatality -2.0786576 -0.74154044

> PC2sorted <- PC[order(-PC2),]

> PC2sorted

Factors PC1 PC2

12 Efficiacy 2.7666265 1.31841598

18 Attitude 0.6912792 1.31305314

4 Malfunction -0.8948329 0.72729344

13 Complication -1.2198141 0.64407808

22 Expectation 4.7947896 0.50398118

9 Dose -0.8481234 0.49111366

15 Adverser -0.8400459 0.46522800

7 Tomography -0.4877384 0.43790386

5 Option -0.3136041 0.43071321

8 Indication -1.7379918 0.32594133

17 Charge -1.5405425 0.30275168

21 Ignorance -0.1758285 0.28008616

6 Inspection -0.2395409 0.16559160

19 Bureaucracy -0.6980449 -0.02510776

23 Hysteria 0.6433542 -0.03177427

1 Malpractise -0.6880152 -0.03697162

2 Reoperation -1.0066844 -0.48686032

3 Consultation -1.0257465 -0.72897236

14 Fatality -2.0786576 -0.74154044

11 Hospitalization 0.5364168 -0.77482753

20 Confusion 2.2442988 -0.95620898

10 Bill 1.2958187 -1.09154125

16 Miscommunication 0.8226273 -2.53234677

> library(ggplot2)

> ggplot(PC, aes(PC1, PC2)) +

+

> ggplot(PC, aes(PC1, PC2)) +

+ modelr::geom\_ref\_line(h = 0) +

+ modelr::geom\_ref\_line(v = 0) +

+ geom\_text(aes(label = Factors), size = 3) +

+ xlab("First Principal Component") +

+ ylab("Second Principal Component") +

+ ggtitle("First Two Principal Components of Dispute")