PCA Analysis

```
library(readr)
expect<-read_csv("D:/MITA 2019/Semester 2/Multivariate Analysis/lfe.csv")

## Parsed with column specification:
## cols(
##   .default = col_double(),
##   Country = col_character(),
##   Status = col_character()
## )

## See spec(...) for full column specifications.

head(expect)

## # A tibble: 6 x 22
##   Country  Year Status `Life expectanc~ `Adult Mortalit~ `infant deaths`
##   <chr>   <dbl> <chr>             <dbl>            <dbl>           <dbl>
## 1 Afghan~  2015 Devel~             65                263              62
## 2 Afghan~  2014 Devel~             59.9              271              64
## 3 Afghan~  2013 Devel~             59.9              268              66
## 4 Afghan~  2012 Devel~             59.5              272              69
## 5 Afghan~  2011 Devel~             59.2              275              71
## 6 Afghan~  2010 Devel~             58.8              279              74
## # ... with 16 more variables: Alcohol <dbl>, `percentage
## #   expenditure` <dbl>, `Hepatitis B` <dbl>, Measles <dbl>, BMI <dbl>,
## #   `under-five deaths` <dbl>, Polio <dbl>, `Total expenditure` <dbl>,
## #   Diphtheria <dbl>, `HIV/AIDS` <dbl>, GDP <dbl>, Population <dbl>,
## #   `thinness 1-19 years` <dbl>, `thinness 5-9 years` <dbl>, `Income
## #   composition of resources` <dbl>, Schooling <dbl>

sapply(expect, function(x) sum(is.na(x)))

##                         Country                            Year
##                               0                               0
##                          Status                 Life expectancy
##                               0                              10
##                 Adult Mortality                   infant deaths
##                              10                               0
##                         Alcohol          percentage expenditure
##                             194                               0
##                     Hepatitis B                         Measles
##                             553                               0
##                             BMI               under-five deaths
##                              34                               0
##                           Polio               Total expenditure
##                              19                             226
##                      Diphtheria                        HIV/AIDS
```

```
##                               19                               0
##                              GDP                      Population
##                              448                             652
##             thinness  1-19 years              thinness 5-9 years
##                               34                              34
## Income composition of resources                       Schooling
##                              167                             163
```

```r
expect <- expect[complete.cases(expect),]  ## to remove which has null values
sapply(expect, function(x) sum(is.na(x)))
```

```
##                          Country                            Year
##                                0                               0
##                           Status                 Life expectancy
##                                0                               0
##                  Adult Mortality                   infant deaths
##                                0                               0
##                          Alcohol          percentage expenditure
##                                0                               0
##                      Hepatitis B                         Measles
##                                0                               0
##                              BMI               under-five deaths
##                                0                               0
##                            Polio               Total expenditure
##                                0                               0
##                       Diphtheria                        HIV/AIDS
##                                0                               0
##                              GDP                      Population
##                                0                               0
##             thinness  1-19 years              thinness 5-9 years
##                                0                               0
## Income composition of resources                       Schooling
##                                0                               0
```

```r
dim(expect)
```

```
## [1] 1649   22
```

```r
#Get the Correlations between the measurements
cor(expect[,5:14])
```

```
##                        Adult Mortality infant deaths     Alcohol
## Adult Mortality            1.000000000    0.04245024 -0.17553509
## infant deaths              0.042450237    1.00000000 -0.10621692
## Alcohol                   -0.175535086   -0.10621692  1.00000000
## percentage expenditure    -0.237609890   -0.09076463  0.41704736
## Hepatitis B               -0.105225443   -0.23176894  0.10988939
## Measles                   -0.003966685    0.53267983 -0.05011023
## BMI                       -0.351542478   -0.23442515  0.35339621
## under-five deaths          0.060365026    0.99690562 -0.10108216
## Polio                     -0.199853000   -0.15692881  0.24031453
```

```
## Total expenditure         -0.085226535   -0.14695112  0.21488509
##                        percentage expenditure Hepatitis B      Measles
## Adult Mortality                    -0.23760989 -0.10522544 -0.003966685
## infant deaths                      -0.09076463 -0.23176894  0.532679832
## Alcohol                             0.41704736  0.10988939 -0.050110235
## percentage expenditure              1.00000000  0.01676017 -0.063070789
## Hepatitis B                         0.01676017  1.00000000 -0.124799993
## Measles                            -0.06307079 -0.12479999  1.000000000
## BMI                                 0.24273824  0.14330179 -0.153245464
## under-five deaths                  -0.09215806 -0.24076603  0.517505563
## Polio                               0.12862605  0.46333080 -0.057850133
## Total expenditure                   0.18387236  0.11332668 -0.113582738
##                               BMI under-five deaths       Polio
## Adult Mortality        -0.3515425        0.06036503 -0.19985300
## infant deaths          -0.2344252        0.99690562 -0.15692881
## Alcohol                 0.3533962       -0.10108216  0.24031453
## percentage expenditure  0.2427382       -0.09215806  0.12862605
## Hepatitis B             0.1433018       -0.24076603  0.46333080
## Measles                -0.1532455        0.51750556 -0.05785013
## BMI                     1.0000000       -0.24213740  0.18626797
## under-five deaths      -0.2421374        1.00000000 -0.17116419
## Polio                   0.1862680       -0.17116419  1.00000000
## Total expenditure       0.1894690       -0.14580310  0.11976798
##                        Total expenditure
## Adult Mortality              -0.08522653
## infant deaths                -0.14695112
## Alcohol                       0.21488509
## percentage expenditure        0.18387236
## Hepatitis B                   0.11332668
## Measles                      -0.11358274
## BMI                           0.18946896
## under-five deaths            -0.14580310
## Polio                         0.11976798
## Total expenditure             1.00000000
```

```
#Using prcomp to compute the principal components (eigenvalues and
eigenvectors). With scale=TRUE, variable means are set to zero, and variances
set to one
expect_pca <- prcomp(expect[,5:14],scale=TRUE)
expect_pca
```

```
## Standard deviations (1, .., p=10):
##  [1] 1.71092337 1.34116282 1.10839094 0.97208242 0.90665887 0.82121793
##  [7] 0.77206427 0.72554886 0.69207207 0.05149214
##
## Rotation (n x k) = (10 x 10):
##                               PC1        PC2         PC3         PC4
## Adult Mortality        -0.1834960  0.3703644 -0.01041610  0.62257953
## infant deaths          -0.4770872 -0.3712794 -0.08404310  0.05506979
## Alcohol                 0.2532057 -0.4167450  0.18255591  0.23580006
```

```
## percentage expenditure  0.2140378 -0.3834422  0.34580661  0.12594468
## Hepatitis B             0.2649221 -0.0567973 -0.65426151  0.10496091
## Measles                -0.3365688 -0.3228056 -0.15791130 -0.01068973
## BMI                     0.3179295 -0.2877457  0.15671493 -0.31360244
## under-five deaths      -0.4789776 -0.3625045 -0.07115429  0.06951117
## Polio                   0.2639952 -0.2314229 -0.58308223  0.06789989
## Total expenditure       0.2124554 -0.1705419  0.13159750  0.64728553
##                                 PC5          PC6          PC7         PC8
## Adult Mortality        -0.351776663 -0.410661325  0.006840613 -0.17821310
## infant deaths           0.055199591 -0.040233504 -0.339430463 -0.02641364
## Alcohol                -0.389050398 -0.378992608  0.034806248  0.28550238
## percentage expenditure -0.416255561  0.513983441 -0.027196893 -0.37429860
## Hepatitis B            -0.009123371  0.001912539 -0.113222919 -0.60601595
## Measles                 0.016614222 -0.057324250  0.856638622 -0.13336166
## BMI                     0.235921412 -0.627446408 -0.068997207 -0.27042754
## under-five deaths       0.046955972 -0.048132459 -0.357483845 -0.02482776
## Polio                  -0.152691104  0.078832447 -0.023914423  0.53310198
## Total expenditure       0.683403002  0.128117745  0.054887611  0.05427242
##                                 PC9          PC10
## Adult Mortality         0.342255320  0.0105228852
## infant deaths           0.022933151  0.7094653595
## Alcohol                -0.543580842  0.0091310487
## percentage expenditure  0.305660217 -0.0008370615
## Hepatitis B            -0.327638935 -0.0025471263
## Measles                 0.045680114 -0.0142794538
## BMI                     0.407383227 -0.0040672777
## under-five deaths       0.013252633 -0.7043918640
## Polio                   0.466677391 -0.0082532569
## Total expenditure       0.008282982  0.0010715226
```

```
summary(expect_pca)
```

```
## Importance of components:
##                           PC1    PC2    PC3     PC4    PC5     PC6     PC7
## Standard deviation     1.7109 1.3412 1.1084 0.97208 0.9067 0.82122 0.77206
## Proportion of Variance 0.2927 0.1799 0.1229 0.09449 0.0822 0.06744 0.05961
## Cumulative Proportion  0.2927 0.4726 0.5955 0.68995 0.7722 0.83959 0.89920
##                            PC8    PC9    PC10
## Standard deviation     0.72555 0.6921 0.05149
## Proportion of Variance 0.05264 0.0479 0.00027
## Cumulative Proportion  0.95184 0.9997 1.00000
```

```
# sample scores stored in sparrows_pca$x
# singular values (square roots of eigenvalues) stored in sparrow_pca$sdev
# loadings (eigenvectors) are stored in sparrows_pca$rotation
# variable means stored in sparrows_pca$center
# variable standard deviations stored in sparrows_pca$scale
# A table containing eigenvalues and %'s accounted, follows
# Eigenvalues are sdev^2
```

```
(eigen_expect <- expect_pca$sdev^2)

##  [1] 2.92725879 1.79871772 1.22853048 0.94494424 0.82203031 0.67439889
##  [7] 0.59608324 0.52642115 0.47896375 0.00265144

names(eigen_expect) <- paste("PC",1:10,sep="")
eigen_expect

##        PC1        PC2        PC3        PC4        PC5        PC6
## 2.92725879 1.79871772 1.22853048 0.94494424 0.82203031 0.67439889
##        PC7        PC8        PC9       PC10
## 0.59608324 0.52642115 0.47896375 0.00265144

sumlambdas <- sum(eigen_expect)
sumlambdas

## [1] 10

#shows the percent variance each variable PC1, PC2...PC5 holds
propvar <- eigen_expect/sumlambdas
propvar

##         PC1         PC2         PC3         PC4         PC5         PC6
## 0.292725879 0.179871772 0.122853048 0.094494424 0.082203031 0.067439889
##         PC7         PC8         PC9        PC10
## 0.059608324 0.052642115 0.047896375 0.000265144

cumvar_expect <- cumsum(propvar)
cumvar_expect

##       PC1       PC2       PC3       PC4       PC5       PC6       PC7
## 0.2927259 0.4725977 0.5954507 0.6899451 0.7721482 0.8395880 0.8991964
##       PC8       PC9      PC10
## 0.9518385 0.9997349 1.0000000

matlambdas <- rbind(eigen_expect,propvar,cumvar_expect)
rownames(matlambdas) <- c("Eigenvalues","Prop. variance","Cum. prop.
variance")
round(matlambdas,4)

##                         PC1    PC2    PC3    PC4    PC5    PC6    PC7
## Eigenvalues          2.9273 1.7987 1.2285 0.9449 0.8220 0.6744 0.5961
## Prop. variance       0.2927 0.1799 0.1229 0.0945 0.0822 0.0674 0.0596
## Cum. prop. variance 0.2927 0.4726 0.5955 0.6899 0.7721 0.8396 0.8992
##                         PC8    PC9   PC10
## Eigenvalues          0.5264 0.4790 0.0027
## Prop. variance       0.0526 0.0479 0.0003
## Cum. prop. variance 0.9518 0.9997 1.0000

summary(expect_pca)
```

```
## Importance of components:
##                           PC1    PC2    PC3     PC4    PC5     PC6     PC7
## Standard deviation     1.7109 1.3412 1.1084 0.97208 0.9067 0.82122 0.77206
## Proportion of Variance 0.2927 0.1799 0.1229 0.09449 0.0822 0.06744 0.05961
## Cumulative Proportion  0.2927 0.4726 0.5955 0.68995 0.7722 0.83959 0.89920
##                            PC8    PC9    PC10
## Standard deviation     0.72555 0.6921 0.05149
## Proportion of Variance 0.05264 0.0479 0.00027
## Cumulative Proportion  0.95184 0.9997 1.00000
```

```
expect_pca$rotation
```

```
##                               PC1        PC2         PC3         PC4
## Adult Mortality        -0.1834960  0.3703644 -0.01041610  0.62257953
## infant deaths          -0.4770872 -0.3712794 -0.08404310  0.05506979
## Alcohol                 0.2532057 -0.4167450  0.18255591  0.23580006
## percentage expenditure  0.2140378 -0.3834422  0.34580661  0.12594468
## Hepatitis B             0.2649221 -0.0567973 -0.65426151  0.10496091
## Measles                -0.3365688 -0.3228056 -0.15791130 -0.01068973
## BMI                     0.3179295 -0.2877457  0.15671493 -0.31360244
## under-five deaths      -0.4789776 -0.3625045 -0.07115429  0.06951117
## Polio                   0.2639952 -0.2314229 -0.58308223  0.06789989
## Total expenditure       0.2124554 -0.1705419  0.13159750  0.64728553
##                                 PC5          PC6          PC7         PC8
## Adult Mortality        -0.351776663 -0.410661325  0.006840613 -0.17821310
## infant deaths           0.055199591 -0.040233504 -0.339430463 -0.02641364
## Alcohol                -0.389050398 -0.378992608  0.034806248  0.28550238
## percentage expenditure -0.416255561  0.513983441 -0.027196893 -0.37429860
## Hepatitis B            -0.009123371  0.001912539 -0.113222919 -0.60601595
## Measles                 0.016614222 -0.057324250  0.856638622 -0.13336166
## BMI                     0.235921412 -0.627446408 -0.068997207 -0.27042754
## under-five deaths       0.046955972 -0.048132459 -0.357483845 -0.02482776
## Polio                  -0.152691104  0.078832447 -0.023914423  0.53310198
## Total expenditure       0.683403002  0.128117745  0.054887611  0.05427242
##                                 PC9          PC10
## Adult Mortality         0.342255320  0.0105228852
## infant deaths           0.022933151  0.7094653595
## Alcohol                -0.543580842  0.0091310487
## percentage expenditure  0.305660217 -0.0008370615
## Hepatitis B            -0.327638935 -0.0025471263
## Measles                 0.045680114 -0.0142794538
## BMI                     0.407383227 -0.0040672777
## under-five deaths       0.013252633 -0.7043918640
## Polio                   0.466677391 -0.0082532569
## Total expenditure       0.008282982  0.0010715226
```

```
print(expect_pca)
```

```
## Standard deviations (1, .., p=10):
##  [1] 1.71092337 1.34116282 1.10839094 0.97208242 0.90665887 0.82121793
##  [7] 0.77206427 0.72554886 0.69207207 0.05149214
```

```
## 
## Rotation (n x k) = (10 x 10):
##                               PC1        PC2         PC3         PC4
## Adult Mortality        -0.1834960  0.3703644 -0.01041610  0.62257953
## infant deaths          -0.4770872 -0.3712794 -0.08404310  0.05506979
## Alcohol                 0.2532057 -0.4167450  0.18255591  0.23580006
## percentage expenditure  0.2140378 -0.3834422  0.34580661  0.12594468
## Hepatitis B             0.2649221 -0.0567973 -0.65426151  0.10496091
## Measles                -0.3365688 -0.3228056 -0.15791130 -0.01068973
## BMI                     0.3179295 -0.2877457  0.15671493 -0.31360244
## under-five deaths      -0.4789776 -0.3625045 -0.07115429  0.06951117
## Polio                   0.2639952 -0.2314229 -0.58308223  0.06789989
## Total expenditure       0.2124554 -0.1705419  0.13159750  0.64728553
##                                 PC5          PC6          PC7         PC8
## Adult Mortality        -0.351776663 -0.410661325  0.006840613 -0.17821310
## infant deaths           0.055199591 -0.040233504 -0.339430463 -0.02641364
## Alcohol                -0.389050398 -0.378992608  0.034806248  0.28550238
## percentage expenditure -0.416255561  0.513983441 -0.027196893 -0.37429860
## Hepatitis B            -0.009123371  0.001912539 -0.113222919 -0.60601595
## Measles                 0.016614222 -0.057324250  0.856638622 -0.13336166
## BMI                     0.235921412 -0.627446408 -0.068997207 -0.27042754
## under-five deaths       0.046955972 -0.048132459 -0.357483845 -0.02482776
## Polio                  -0.152691104  0.078832447 -0.023914423  0.53310198
## Total expenditure       0.683403002  0.128117745  0.054887611  0.05427242
##                                 PC9          PC10
## Adult Mortality         0.342255320  0.0105228852
## infant deaths           0.022933151  0.7094653595
## Alcohol                -0.543580842  0.0091310487
## percentage expenditure  0.305660217 -0.0008370615
## Hepatitis B            -0.327638935 -0.0025471263
## Measles                 0.045680114 -0.0142794538
## BMI                     0.407383227 -0.0040672777
## under-five deaths       0.013252633 -0.7043918640
## Polio                   0.466677391 -0.0082532569
## Total expenditure       0.008282982  0.0010715226

# Sample scores stored in expect_pca$x
#expect_pca$x
#head(5)

# Identifying the scores by their status of the country
expecttyp_pca <- cbind(data.frame(expect$Status),expect_pca$x)
#expecttyp_pca

# Means of scores for all the PC's classified by Status
tabmeansPC <-
aggregate(expecttyp_pca[,2:11],by=list(Status=expect$Status),mean)
tabmeansPC

##       Status        PC1        PC2         PC3         PC4        PC5
## 1  Developed  1.5878806 -1.4542792  0.39620163  0.18058910 -0.4176836
```

```
## 2 Developing -0.2731109  0.2501319 -0.06814555 -0.03106081  0.0718404
##            PC6          PC7          PC8         PC9          PC10
## 1 -0.027444838 -0.022914827  0.042983587 -0.28999984  0.0010817265
## 2  0.004720434  0.003941285 -0.007393055  0.04987915 -0.0001860539
```

```
tabmeansPC <- tabmeansPC[rev(order(tabmeansPC$Status)),]
tabmeansPC
```

```
##       Status        PC1        PC2         PC3         PC4        PC5
## 2 Developing -0.2731109  0.2501319 -0.06814555 -0.03106081  0.0718404
## 1  Developed  1.5878806 -1.4542792  0.39620163  0.18058910 -0.4176836
##            PC6          PC7          PC8         PC9          PC10
## 2  0.004720434  0.003941285 -0.007393055  0.04987915 -0.0001860539
## 1 -0.027444838 -0.022914827  0.042983587 -0.28999984  0.0010817265
```

```
tabfmeans <- t(tabmeansPC[-1])
tabfmeans
```

```
##                  2            1
## PC1  -0.2731109490  1.587880600
## PC2   0.2501318913 -1.454279219
## PC3  -0.0681455533  0.396201626
## PC4  -0.0310608119  0.180589101
## PC5   0.0718403987 -0.417683640
## PC6   0.0047204341 -0.027444838
## PC7   0.0039412850 -0.022914827
## PC8  -0.0073930548  0.042983587
## PC9   0.0498791484 -0.289999842
## PC10 -0.0001860539  0.001081726
```

```
colnames(tabfmeans) <- t(as.vector(tabmeansPC[1]))
tabfmeans
```

```
##         Developing    Developed
## PC1  -0.2731109490  1.587880600
## PC2   0.2501318913 -1.454279219
## PC3  -0.0681455533  0.396201626
## PC4  -0.0310608119  0.180589101
## PC5   0.0718403987 -0.417683640
## PC6   0.0047204341 -0.027444838
## PC7   0.0039412850 -0.022914827
## PC8  -0.0073930548  0.042983587
## PC9   0.0498791484 -0.289999842
## PC10 -0.0001860539  0.001081726
```

```
# Standard deviations of scores for all the PC's classified by Status Of the
country
tabsdsPC <- aggregate(expecttyp_pca[,2:11],by=list(Status=expect$Status),sd)
tabfsds <- t(tabsdsPC[,-1])
colnames(tabfsds) <- t(as.vector(tabsdsPC[1]))
tabfsds
```

```
##        Developed Developing
## PC1  0.671590927 1.68676240
## PC2  0.946777063 1.23613640
## PC3  0.935451300 1.12175075
## PC4  0.848803565 0.98869348
## PC5  1.176829004 0.83122639
## PC6  1.230550802 0.72853846
## PC7  0.200669114 0.83166855
## PC8  0.909384517 0.68913319
## PC9  0.784432348 0.66254393
## PC10 0.009330449 0.05561157
```

```r
t.test(PC1~expect$Status,data=expecttyp_pca)
```

```
##
##  Welch Two Sample t-test
##
## data:  PC1 by expect$Status
## t = 29.854, df = 871.76, p-value < 2.2e-16
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  1.738643 1.983340
## sample estimates:
##  mean in group Developed mean in group Developing
##                1.5878806               -0.2731109
```

```r
t.test(PC2~expect$Status,data=expecttyp_pca)
```

```
##
##  Welch Two Sample t-test
##
## data:  PC2 by expect$Status
## t = -24.626, df = 397.19, p-value < 2.2e-16
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -1.840476 -1.568346
## sample estimates:
##  mean in group Developed mean in group Developing
##               -1.4542792                0.2501319
```

```r
t.test(PC3~expect$Status,data=expecttyp_pca)
```

```
##
##  Welch Two Sample t-test
##
## data:  PC3 by expect$Status
## t = 6.9142, df = 371.06, p-value = 2.066e-11
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  0.3322874 0.5964070
## sample estimates:
```

```
##  mean in group Developed mean in group Developing
##              0.39620163            -0.06814555
```

```r
t.test(PC4~expect$Status,data=expecttyp_pca)
```

```
##
##  Welch Two Sample t-test
##
## data:  PC4 by expect$Status
## t = 3.4928, df = 363.21, p-value = 0.0005369
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  0.09248663 0.33081319
## sample estimates:
##  mean in group Developed mean in group Developing
##              0.18058910            -0.03106081
```

```r
t.test(PC5~expect$Status,data=expecttyp_pca)
```

```
##
##  Welch Two Sample t-test
##
## data:  PC5 by expect$Status
## t = -6.21, df = 283.78, p-value = 1.878e-09
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -0.6446865 -0.3343616
## sample estimates:
##  mean in group Developed mean in group Developing
##               -0.4176836               0.0718404
```

```r
t.test(PC6~expect$Status,data=expecttyp_pca)
```

```
##
##  Welch Two Sample t-test
##
## data:  PC6 by expect$Status
## t = -0.3949, df = 270.77, p-value = 0.6932
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -0.1925257  0.1281952
## sample estimates:
##  mean in group Developed mean in group Developing
##             -0.027444838             0.004720434
```

```r
t.test(PC7~expect$Status,data=expecttyp_pca)
```

```
##
##  Welch Two Sample t-test
##
## data:  PC7 by expect$Status
## t = -1.047, df = 1509.8, p-value = 0.2953
```

```
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -0.07717214  0.02345992
## sample estimates:
##  mean in group Developed mean in group Developing
##              -0.022914827              0.003941285
```

```r
t.test(PC8~expect$Status,data=expecttyp_pca)
```

```
##
##  Welch Two Sample t-test
##
## data:  PC8 by expect$Status
## t = 0.82212, df = 290.47, p-value = 0.4117
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -0.07022549  0.17097877
## sample estimates:
##  mean in group Developed mean in group Developing
##               0.042983587             -0.007393055
```

```r
t.test(PC9~expect$Status,data=expecttyp_pca)
```

```
##
##  Welch Two Sample t-test
##
## data:  PC9 by expect$Status
## t = -6.3613, df = 302.99, p-value = 7.368e-10
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -0.4450185 -0.2347395
## sample estimates:
##  mean in group Developed mean in group Developing
##               -0.28999984               0.04987915
```

```r
t.test(PC10~expect$Status,data=expecttyp_pca)
```

```
##
##  Welch Two Sample t-test
##
## data:  PC10 by expect$Status
## t = 0.79271, df = 1646.6, p-value = 0.4281
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -0.001869110  0.004404671
## sample estimates:
##  mean in group Developed mean in group Developing
##              0.0010817265            -0.0001860539
```

```r
# F ratio tests
var.test(PC1~expect$Status,data=expecttyp_pca)
```

```
##
##  F test to compare two variances
##
## data:  PC1 by expect$Status
## F = 0.15853, num df = 241, denom df = 1406, p-value < 2.2e-16
## alternative hypothesis: true ratio of variances is not equal to 1
## 95 percent confidence interval:
##  0.1315019 0.1936868
## sample estimates:
## ratio of variances
##          0.1585265
```

```
var.test(PC2~expect$Status,data=expecttyp_pca)
```

```
##
##  F test to compare two variances
##
## data:  PC2 by expect$Status
## F = 0.58663, num df = 241, denom df = 1406, p-value = 4.377e-07
## alternative hypothesis: true ratio of variances is not equal to 1
## 95 percent confidence interval:
##  0.4866234 0.7167386
## sample estimates:
## ratio of variances
##          0.5866278
```

```
var.test(PC3~expect$Status,data=expecttyp_pca)
```

```
##
##  F test to compare two variances
##
## data:  PC3 by expect$Status
## F = 0.69542, num df = 241, denom df = 1406, p-value = 0.0004588
## alternative hypothesis: true ratio of variances is not equal to 1
## 95 percent confidence interval:
##  0.5768727 0.8496651
## sample estimates:
## ratio of variances
##          0.6954239
```

```
var.test(PC4~expect$Status,data=expecttyp_pca)
```

```
##
##  F test to compare two variances
##
## data:  PC4 by expect$Status
## F = 0.73704, num df = 241, denom df = 1406, p-value = 0.003076
## alternative hypothesis: true ratio of variances is not equal to 1
## 95 percent confidence interval:
##  0.6113943 0.9005114
## sample estimates:
```

```
## ratio of variances
##            0.73704
```

```r
var.test(PC5~expect$Status,data=expecttyp_pca)
```

```
##
##  F test to compare two variances
##
## data:  PC5 by expect$Status
## F = 2.0044, num df = 241, denom df = 1406, p-value = 1.954e-14
## alternative hypothesis: true ratio of variances is not equal to 1
## 95 percent confidence interval:
##  1.662717 2.448986
## sample estimates:
## ratio of variances
##           2.004417
```

```r
var.test(PC6~expect$Status,data=expecttyp_pca)
```

```
##
##  F test to compare two variances
##
## data:  PC6 by expect$Status
## F = 2.8529, num df = 241, denom df = 1406, p-value < 2.2e-16
## alternative hypothesis: true ratio of variances is not equal to 1
## 95 percent confidence interval:
##  2.366598 3.485718
## sample estimates:
## ratio of variances
##            2.85295
```

```r
var.test(PC7~expect$Status,data=expecttyp_pca)
```

```
##
##  F test to compare two variances
##
## data:  PC7 by expect$Status
## F = 0.058218, num df = 241, denom df = 1406, p-value < 2.2e-16
## alternative hypothesis: true ratio of variances is not equal to 1
## 95 percent confidence interval:
##  0.04829374 0.07113096
## sample estimates:
## ratio of variances
##         0.05821843
```

```r
var.test(PC8~expect$Status,data=expecttyp_pca)
```

```
##
##  F test to compare two variances
##
## data:  PC8 by expect$Status
## F = 1.7414, num df = 241, denom df = 1406, p-value = 1.783e-09
```

```
## alternative hypothesis: true ratio of variances is not equal to 1
## 95 percent confidence interval:
##  1.444505 2.127585
## sample estimates:
## ratio of variances
##           1.741361
```

```r
var.test(PC9~expect$Status,data=expecttyp_pca)
```

```
##
##  F test to compare two variances
##
## data:  PC9 by expect$Status
## F = 1.4018, num df = 241, denom df = 1406, p-value = 0.0003377
## alternative hypothesis: true ratio of variances is not equal to 1
## 95 percent confidence interval:
##  1.162819 1.712694
## sample estimates:
## ratio of variances
##           1.401786
```

```r
var.test(PC10~expect$Status,data=expecttyp_pca)
```

```
##
##  F test to compare two variances
##
## data:  PC10 by expect$Status
## F = 0.02815, num df = 241, denom df = 1406, p-value < 2.2e-16
## alternative hypothesis: true ratio of variances is not equal to 1
## 95 percent confidence interval:
##  0.02335098 0.03439323
## sample estimates:
## ratio of variances
##         0.02814977
```

```r
# Levene's tests (one-sided)
library(car)
```

```
## Warning: package 'car' was built under R version 3.6.2
```

```
## Loading required package: carData
```

```r
(LTPC1 <- leveneTest(PC1~expect$Status,data=expecttyp_pca))
```

```
## Warning in leveneTest.default(y = y, group = group, ...): group coerced to
## factor.
```

```
## Levene's Test for Homogeneity of Variance (center = median)
##         Df F value    Pr(>F)
## group    1  29.983 5.031e-08 ***
##       1647
```

```
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
(LTPC1 <- leveneTest(PC1~expect$Status,data=expecttyp_pca))
```

```
## Warning in leveneTest.default(y = y, group = group, ...): group coerced to
## factor.
```

```
## Levene's Test for Homogeneity of Variance (center = median)
##         Df F value    Pr(>F)
## group    1  29.983 5.031e-08 ***
##       1647
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
(p_PC1_1sided <- LTPC1[[3]][1]/2)
```

```
## [1] 2.515447e-08
```

```
(LTPC2 <- leveneTest(PC2~expect$Status,data=expecttyp_pca))
```

```
## Warning in leveneTest.default(y = y, group = group, ...): group coerced to
## factor.
```

```
## Levene's Test for Homogeneity of Variance (center = median)
##         Df F value Pr(>F)
## group    1  2.4365 0.1187
##       1647
```

```
(p_PC2_1sided=LTPC2[[3]][1]/2)
```

```
## [1] 0.05936535
```

```
(LTPC3 <- leveneTest(PC3~expect$Status,data=expecttyp_pca))
```

```
## Warning in leveneTest.default(y = y, group = group, ...): group coerced to
## factor.
```

```
## Levene's Test for Homogeneity of Variance (center = median)
##         Df F value  Pr(>F)
## group    1   2.767 0.09642 .
##       1647
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
(p_PC3_1sided <- LTPC3[[3]][1]/2)
```

```
## [1] 0.04820763
```

```
(LTPC4 <- leveneTest(PC4~expect$Status,data=expecttyp_pca))
```

```
## Warning in leveneTest.default(y = y, group = group, ...): group coerced to
## factor.
```

```
## Levene's Test for Homogeneity of Variance (center = median)
##         Df F value  Pr(>F)
## group    1  5.5114 0.01901 *
##       1647
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
(p_PC4_1sided <- LTPC4[[3]][1]/2)
```

```
## [1] 0.009505464
```

```r
(LTPC5 <- leveneTest(PC5~expect$Status,data=expecttyp_pca))
```

```
## Warning in leveneTest.default(y = y, group = group, ...): group coerced to
## factor.
```

```
## Levene's Test for Homogeneity of Variance (center = median)
##         Df F value  Pr(>F)
## group    1  19.126 1.3e-05 ***
##       1647
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
(p_PC5_1sided <- LTPC5[[3]][1]/2)
```

```
## [1] 6.500641e-06
```

```r
(LTPC6 <- leveneTest(PC6~expect$Status,data=expecttyp_pca))
```

```
## Warning in leveneTest.default(y = y, group = group, ...): group coerced to
## factor.
```

```
## Levene's Test for Homogeneity of Variance (center = median)
##         Df F value    Pr(>F)
## group    1  76.426 < 2.2e-16 ***
##       1647
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
(p_PC6_1sided <- LTPC6[[3]][1]/2)
```

```
## [1] 2.763661e-18
```

```r
(LTPC7 <- leveneTest(PC7~expect$Status,data=expecttyp_pca))
```

```
## Warning in leveneTest.default(y = y, group = group, ...): group coerced to
## factor.
```

```
## Levene's Test for Homogeneity of Variance (center = median)
##         Df F value    Pr(>F)
## group    1  12.419 0.0004365 ***
##       1647
```

```
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
(p_PC7_1sided <- LTPC7[[3]][1]/2)
```

```
## [1] 0.0002182458
```

```
(LTPC8 <- leveneTest(PC8~expect$Status,data=expecttyp_pca))
```

```
## Warning in leveneTest.default(y = y, group = group, ...): group coerced to
## factor.
```

```
## Levene's Test for Homogeneity of Variance (center = median)
##         Df F value    Pr(>F)    
## group    1  18.635 1.677e-05 ***
##       1647                      
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
(p_PC8_1sided <- LTPC8[[3]][1]/2)
```

```
## [1] 8.383888e-06
```

```
(LTPC9 <- leveneTest(PC9~expect$Status,data=expecttyp_pca))
```

```
## Warning in leveneTest.default(y = y, group = group, ...): group coerced to
## factor.
```

```
## Levene's Test for Homogeneity of Variance (center = median)
##         Df F value  Pr(>F)  
## group    1  4.8289 0.02813 *
##       1647                  
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
(p_PC9_1sided <- LTPC9[[3]][1]/2)
```

```
## [1] 0.01406279
```

```
(LTPC10 <- leveneTest(PC10~expect$Status,data=expecttyp_pca))
```

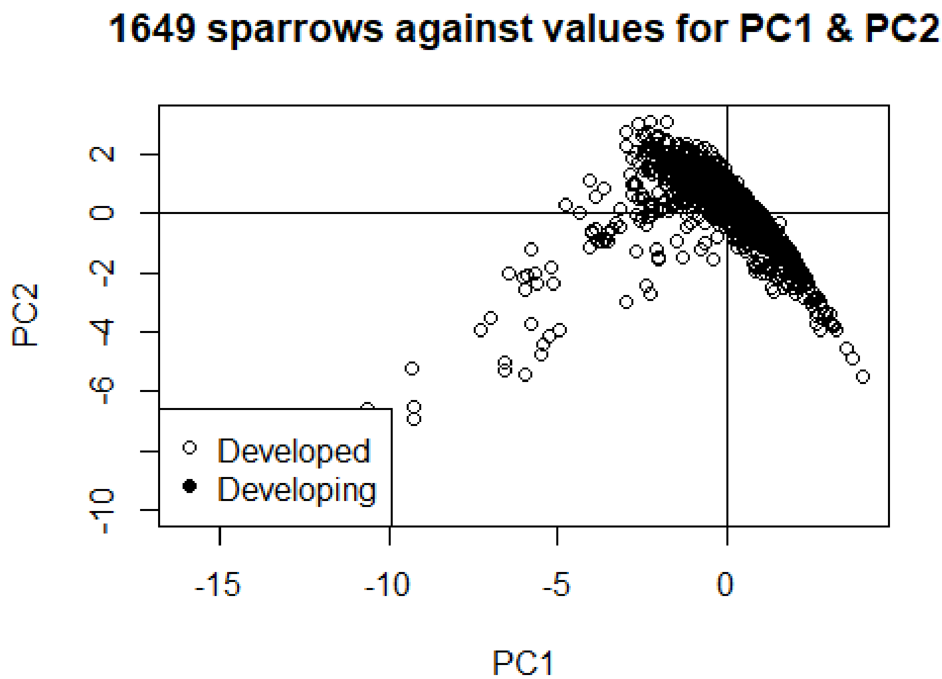
```
## Warning in leveneTest.default(y = y, group = group, ...): group coerced to
## factor.
```

```
## Levene's Test for Homogeneity of Variance (center = median)
##         Df F value    Pr(>F)    
## group    1  26.144 3.539e-07 ***
##       1647                      
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

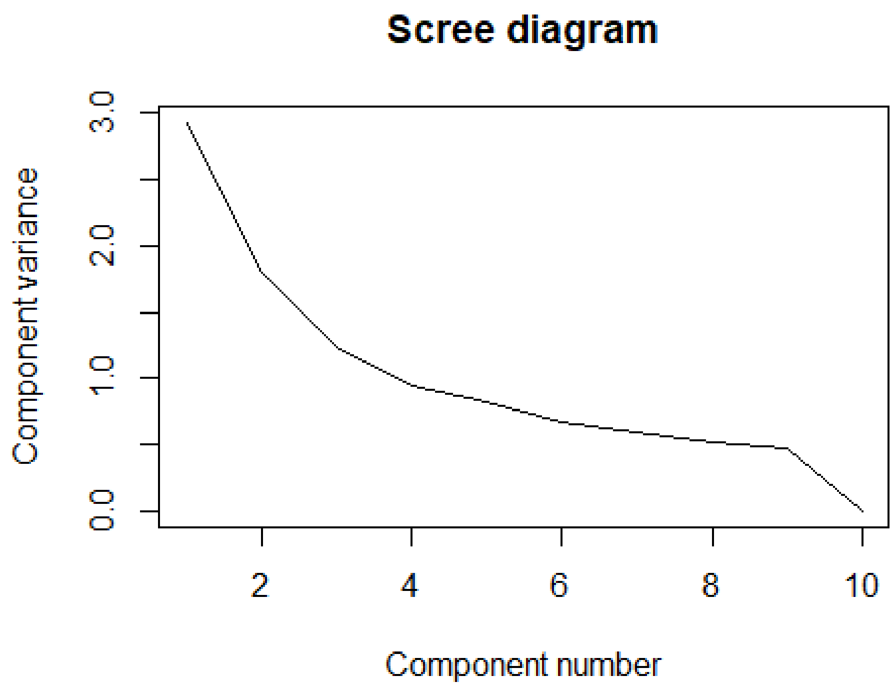
```
(p_PC10_1sided <- LTPC10[[3]][1]/2)
```

```
## [1] 1.769428e-07
```

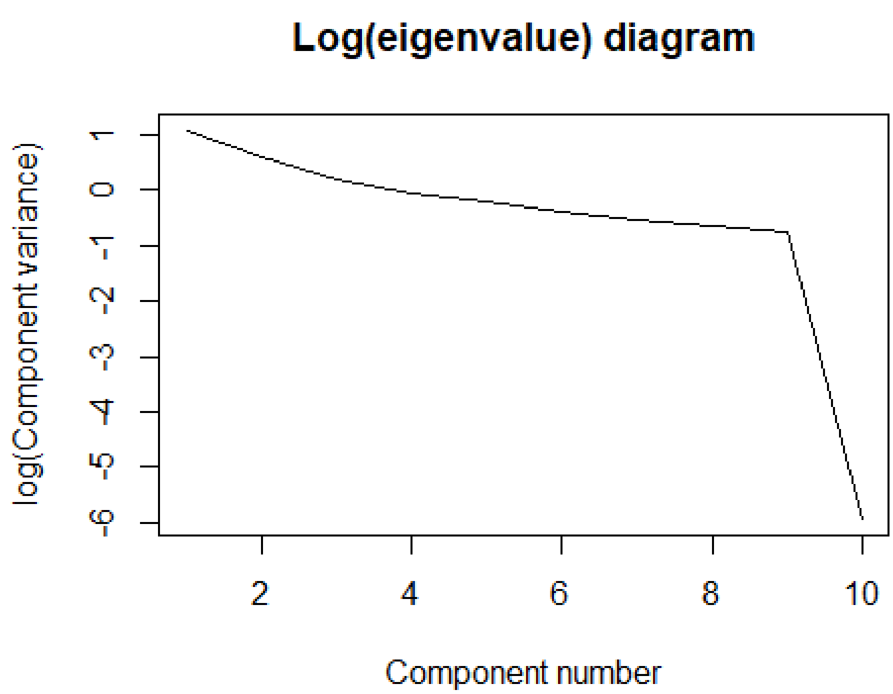
```r
# Levene's tests (one-sided)
# Plotting the scores for the first and second components
plot(expecttyp_pca$PC1, expecttyp_pca$PC2,pch=ifelse(expecttyp_pca$Status ==
"Developed",2,16),xlab="PC1", ylab="PC2", main="1649 sparrows against values
for PC1 & PC2")
abline(h=0)
abline(v=0)
legend("bottomleft", legend=c("Developed","Developing"), pch=c(1,16))
```



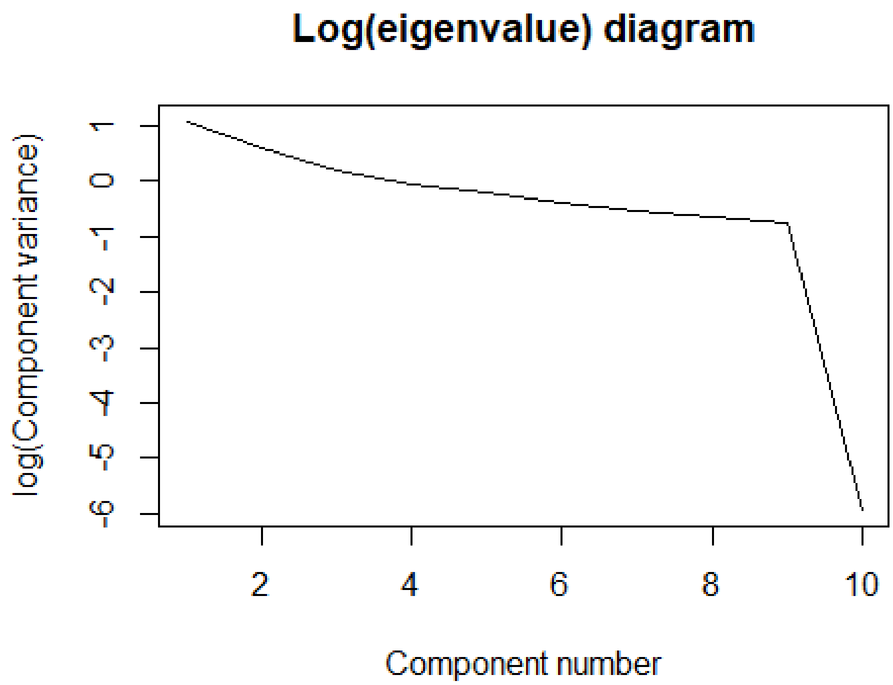

```r
plot(eigen_expect, xlab = "Component number", ylab = "Component variance",
type = "l", main = "Scree diagram")
```




```
plot(log(eigen_expect), xlab = "Component number",ylab = "log(Component
variance)", type="l",main = "Log(eigenvalue) diagram")
```



```
plot(log(eigen_expect), xlab = "Component number",ylab = "log(Component
variance)", type="l",main = "Log(eigenvalue) diagram")
```



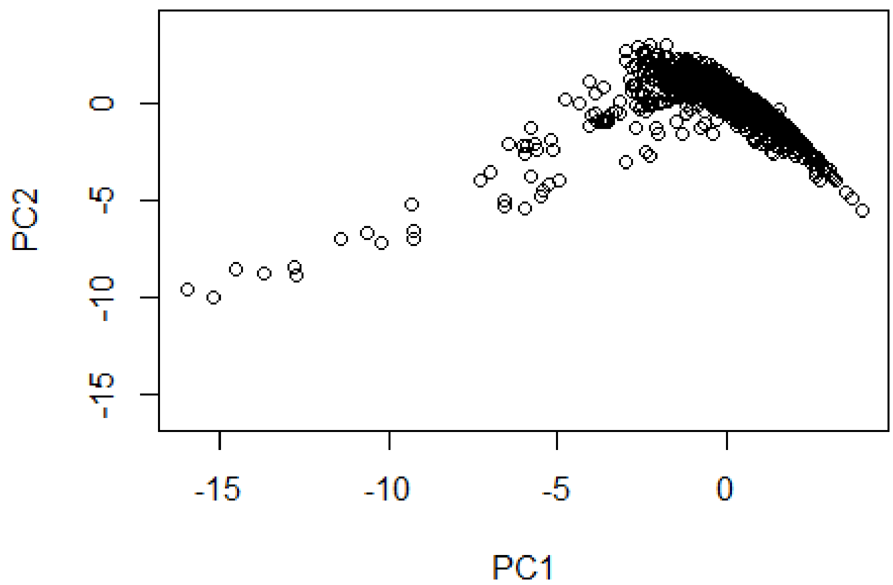

```
print(summary(expect_pca))

## Importance of components:
##                           PC1    PC2    PC3     PC4    PC5     PC6     PC7
## Standard deviation     1.7109 1.3412 1.1084 0.97208 0.9067 0.82122 0.77206
## Proportion of Variance 0.2927 0.1799 0.1229 0.09449 0.0822 0.06744 0.05961
## Cumulative Proportion  0.2927 0.4726 0.5955 0.68995 0.7722 0.83959 0.89920
##                            PC8    PC9    PC10
## Standard deviation     0.72555 0.6921 0.05149
## Proportion of Variance 0.05264 0.0479 0.00027
## Cumulative Proportion  0.95184 0.9997 1.00000

diag(cov(expect_pca$x))

##        PC1        PC2        PC3        PC4        PC5        PC6
## 2.92725879 1.79871772 1.22853048 0.94494424 0.82203031 0.67439889
##        PC7        PC8        PC9       PC10
## 0.59608324 0.52642115 0.47896375 0.00265144

xlim <- range(expect_pca$x[,1])
#expect_pca$x[,1]
#expect_pca$x
plot(expect_pca$x,xlim=xlim,ylim=xlim)
```




```
expect_pca$rotation[,1]
```

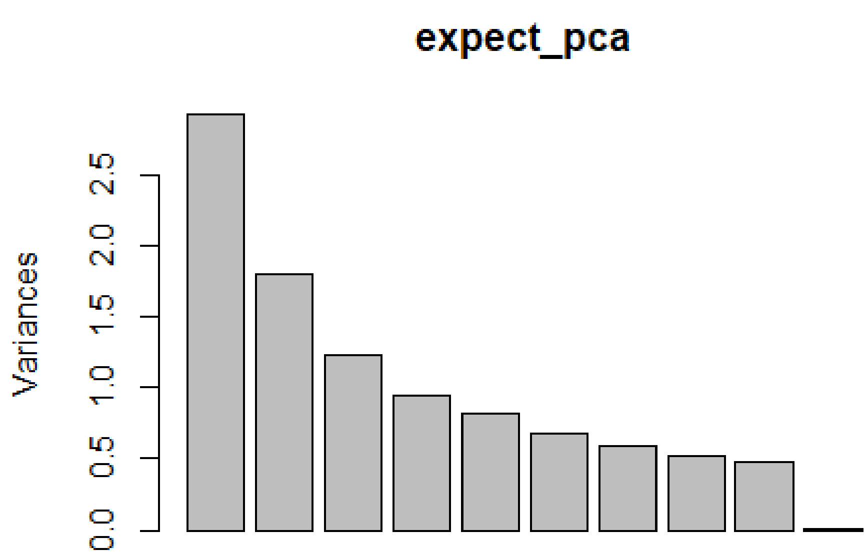
```
##         Adult Mortality            infant deaths                  Alcohol
##              -0.1834960               -0.4770872                0.2532057
## percentage expenditure              Hepatitis B                 Measles
##               0.2140378                0.2649221               -0.3365688
##                     BMI        under-five deaths                   Polio
##               0.3179295               -0.4789776                0.2639952
##       Total expenditure
##               0.2124554
```

```
expect_pca$rotation
```

```
##                               PC1        PC2         PC3         PC4
## Adult Mortality        -0.1834960  0.3703644 -0.01041610  0.62257953
## infant deaths          -0.4770872 -0.3712794 -0.08404310  0.05506979
## Alcohol                 0.2532057 -0.4167450  0.18255591  0.23580006
## percentage expenditure  0.2140378 -0.3834422  0.34580661  0.12594468
## Hepatitis B             0.2649221 -0.0567973 -0.65426151  0.10496091
## Measles                -0.3365688 -0.3228056 -0.15791130 -0.01068973
## BMI                     0.3179295 -0.2877457  0.15671493 -0.31360244
## under-five deaths      -0.4789776 -0.3625045 -0.07115429  0.06951117
## Polio                   0.2639952 -0.2314229 -0.58308223  0.06789989
## Total expenditure       0.2124554 -0.1705419  0.13159750  0.64728553
##                                 PC5          PC6          PC7         PC8
## Adult Mortality        -0.351776663 -0.410661325  0.006840613 -0.17821310
## infant deaths           0.055199591 -0.040233504 -0.339430463 -0.02641364
```

```
## Alcohol                  -0.389050398 -0.378992608  0.034806248  0.28550238
## percentage expenditure -0.416255561  0.513983441 -0.027196893 -0.37429860
## Hepatitis B            -0.009123371  0.001912539 -0.113222919 -0.60601595
## Measles                  0.016614222 -0.057324250  0.856638622 -0.13336166
## BMI                      0.235921412 -0.627446408 -0.068997207 -0.27042754
## under-five deaths        0.046955972 -0.048132459 -0.357483845 -0.02482776
## Polio                   -0.152691104  0.078832447 -0.023914423  0.53310198
## Total expenditure        0.683403002  0.128117745  0.054887611  0.05427242
##                                  PC9          PC10
## Adult Mortality          0.342255320  0.0105228852
## infant deaths            0.022933151  0.7094653595
## Alcohol                 -0.543580842  0.0091310487
## percentage expenditure  0.305660217 -0.0008370615
## Hepatitis B             -0.327638935 -0.0025471263
## Measles                  0.045680114 -0.0142794538
## BMI                      0.407383227 -0.0040672777
## under-five deaths        0.013252633 -0.7043918640
## Polio                    0.466677391 -0.0082532569
## Total expenditure        0.008282982  0.0010715226
```

```
#plot(expect[,-1])
#expect_pca$x
plot(expect_pca)
```




```
#get the original value of the data based on PCA
center <- expect_pca$center
scale <- expect_pca$scale
```

```
new_expect <- as.matrix(expect[,-1])
#new_expect
#drop(scale(new_expect,center=center, scale=scale)%*%expect_pca$rotation[,1])
#predict(expect_pca)[,1]
#The aboved two gives us the same thing. predict is a good function to know.
pairs(expect_pca$x[,1:10], ylim = c(-8,10),xlim = c(-
8,10),panel=function(x,y,...){text(x,y,expect$Status)})
```