Multivariate Statistical Inference Homework #4

Steven Francis

March 21, 2018

Problem 1

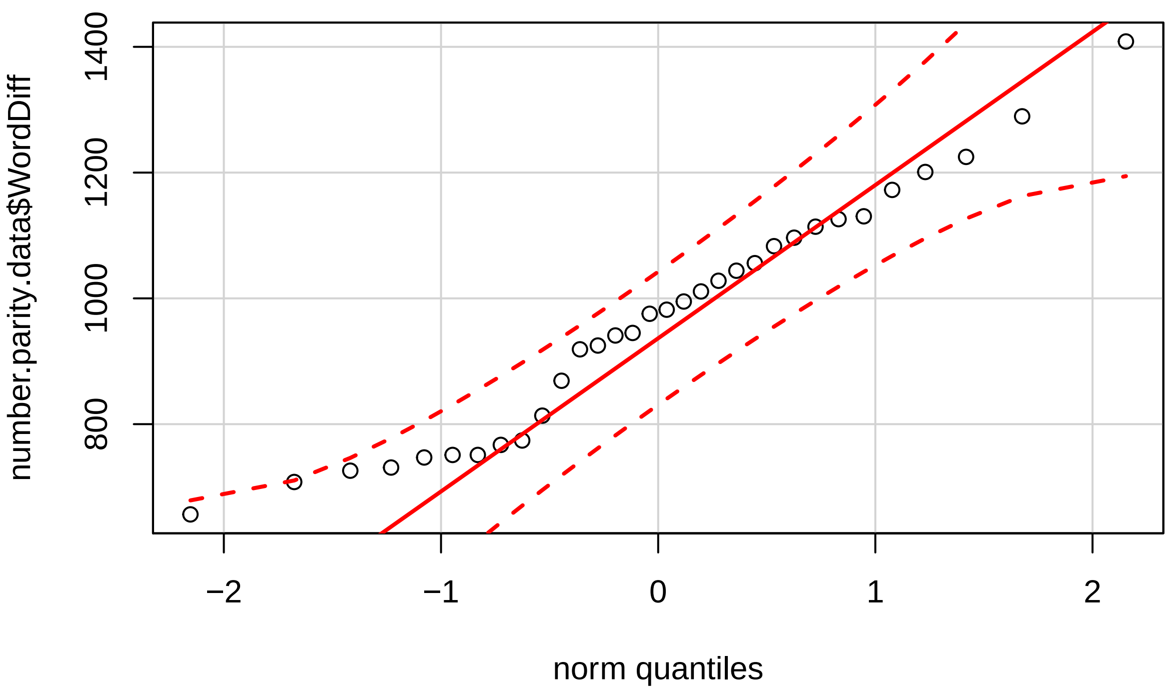
```
number.parity.data <- read.csv("Number_Parity.csv", header = T)
```

Part a)

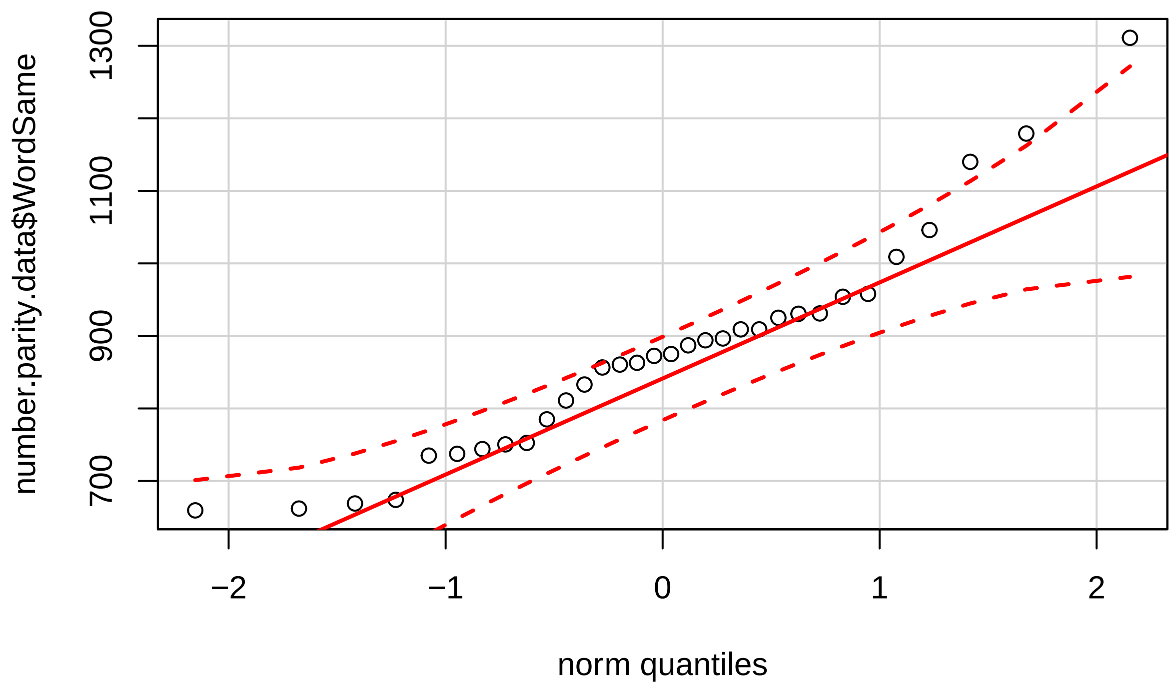
```
library('car')
```

```
## Warning: package 'car' was built under R version 3.4.1
```

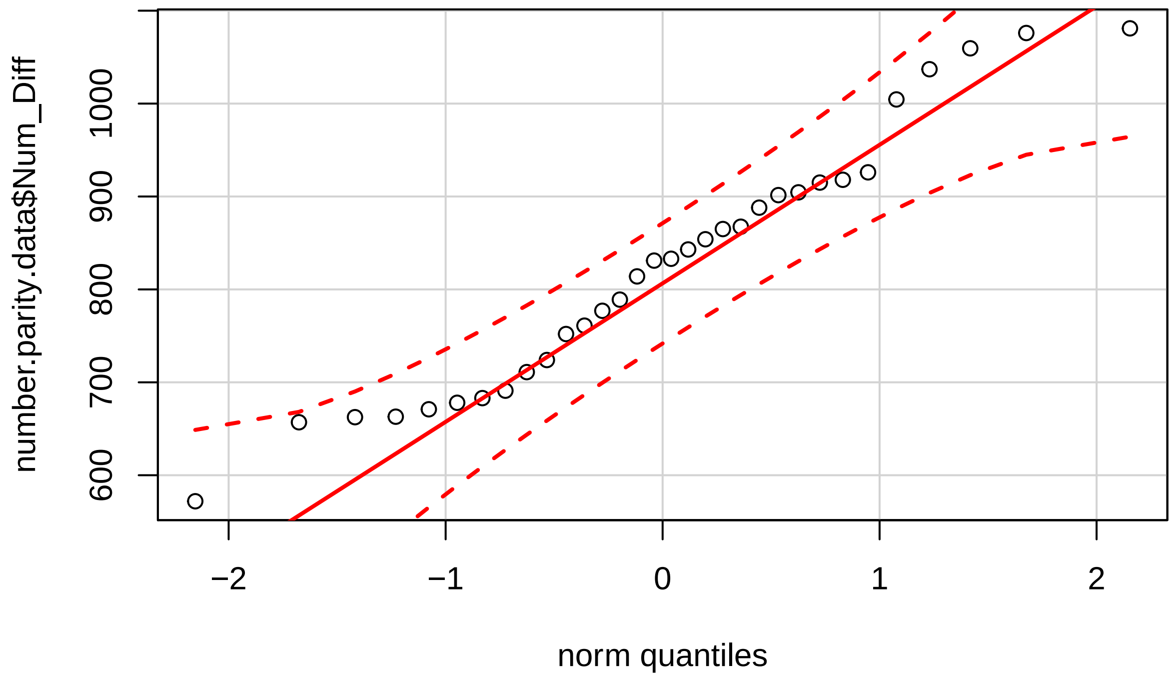
```
qqPlot(number.parity.data$WordDiff)
```



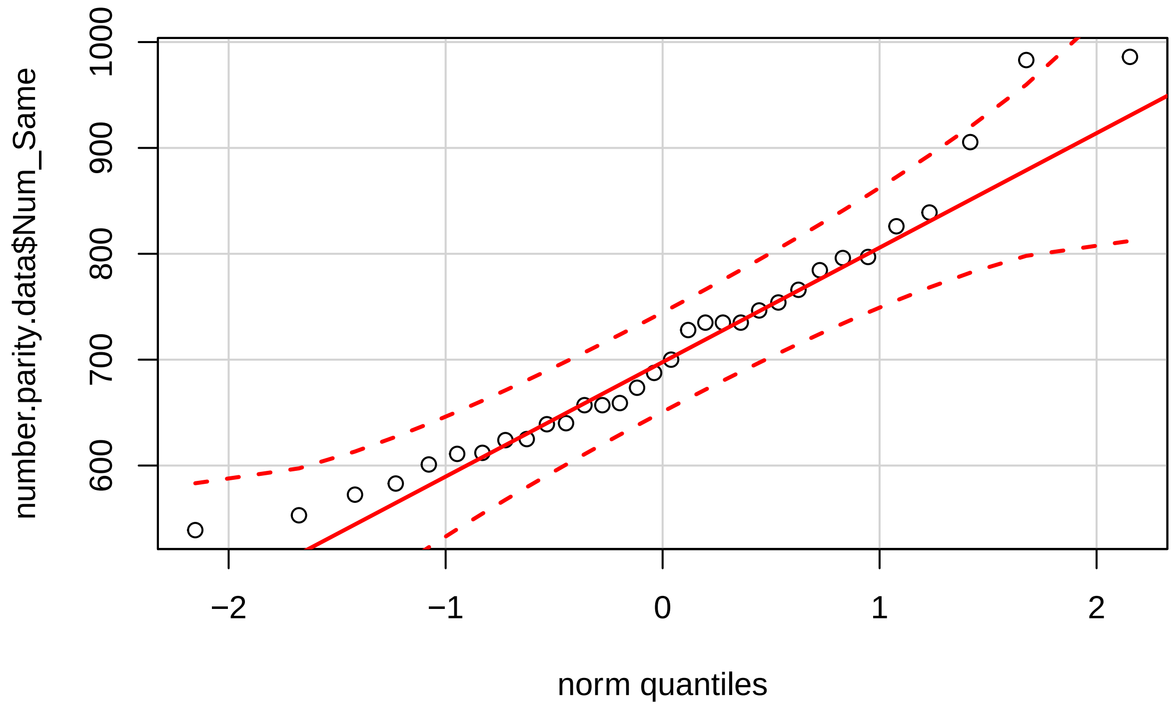

```
qqPlot(number.parity.data$WordSame)
```




```
qqPlot(number.parity.data$Num_Diff)
```




```
qqPlot(number.parity.data$Num_Same)
```




```
shapiro.test(number.parity.data$WordDiff)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  number.parity.data$WordDiff
## W = 0.96543, p-value = 0.3836
```

```
shapiro.test(number.parity.data$WordSame)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  number.parity.data$WordSame
## W = 0.93262, p-value = 0.04635
```

```
shapiro.test(number.parity.data$Num_Diff)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  number.parity.data$Num_Diff
## W = 0.96061, p-value = 0.2853
```

```
shapiro.test(number.parity.data$Num_Same)
```

```
##
##  Shapiro-Wilk normality test
##
```

```
## data:  number.parity.data$Num_Same
## W = 0.94345, p-value = 0.09385
```

By looking at the Q-Q plots, it is apparent that treating this data as a random sample from a multivariate normal distribution is reasonable. All three plots show the data points hover around the "normality line" and are within the "normality boundaries" but could be better. This is also confirmed by the Shapiro-Wilk tests of normality for the various data columns as all p-values (when rounded) are or above 0.05.

Part b)

```
C1 <- array(data = c(-1, -1, -1, 1, -1, 1, -1, 1, 1, 1, 1, -1),
                         dim = c(3,4))

x.bar1 <- colMeans(number.parity.data)

S1 <- cov(number.parity.data)

#Now calculating the T^2 statistic with
n1 = 32
p1 = 4
C1.xbar1 <- as.vector(C1 %*% x.bar1)
C1.xbar1
```

```
## [1] -206.32812 -306.92188   22.42188
```

```
T2.1 <- n1 * sum(C1.xbar1 * solve(C1 %*% S1 %*% t(C1), C1.xbar1))
T2.1
```

```
## [1] 153.7275
```

```
#To obtain a p-value from Hotelling's T^2 distribution
1 - pf((n1-p1+1)/((n1-1)*(p1-1))*T2.1, df1 = p1-1, df2 = n1-p1+1)
```

```
## [1] 2.328437e-11
```

The p-value obtained is essentially zero meaning that the word format/Arabic numeral and parity have some treatment effect on reaction time

Part c)

```
b1.1 <- C1[1,]
b2.1 <- C1[2,]
b3.1 <- C1[3,]

q1 <- 3
alpha <- 0.05


t.mult1 <- qt(1-alpha/(2*q1), df = n1-1)
sum(b1.1*x.bar1) + c(-1,1) * t.mult1 * sqrt(sum(b1.1*S1 %*% b1.1)/n1)
```

```
## [1] -268.9293 -143.7269
```

```r
sum(b2.1*x.bar1) + c(-1,1) * t.mult1 * sqrt(sum(b2.1*S1 %*% b2.1)/n1)
```

```
## [1] -396.7043 -217.1394
```

```r
sum(b3.1*x.bar1) + c(-1,1) * t.mult1 * sqrt(sum(b3.1*S1 %*% b3.1)/n1)
```

```
## [1] -22.33541  67.17916
```

It appears that both main effects (same parity vs. different parity) and (word format vs. Arabic numeral) are statistically significant sense the intervals do not contain 0. Also, the same parity vs, different parity is of higher magnitude.

The interval for the interaction contrast contains zero, therefore there is no evidence of difference in parity effect for word format versus parity effect given Arabic digits.

Problem 2a

```r
turtles <- read.csv("Turtles.csv", header = T)

female_turtles <- turtles[1:24,]
male_turtles <- turtles[25:48,]

n2.1 <- nrow(female_turtles)
n2.2 <- nrow(male_turtles)
p2 <- ncol(male_turtles)

x.bar2.1 <- colMeans(female_turtles[, 1:3])
x.bar2.2 <- colMeans(male_turtles[, 1:3])

S2.1 <- cov(female_turtles[,1:3])
S2.2 <- cov(male_turtles[, 1:3])

S2 <- 1/(n2.1 + n2.2 - 2) * ((n2.1 - 1)*S2.1 + (n2.2 - 1)*S2.2)

T2.2 <- 1/(1/n2.1 + 1/n2.2) * sum((x.bar2.1 - x.bar2.2) *
                                    solve(S2, x.bar2.1 - x.bar2.2))
T2.2
```

```
## [1] 72.38162
```

```r
# To obtain the p-value
1 - pf((n2.1 + n2.2 - p2 - 1)/((n2.1 + n2.2 - 2)* p2) * T2.2,
     df1 = p2, df2 = n2.1 + n2.2- p2 -1)
```

```
## [1] 2.110107e-08
```

Since the p-value is close to zero, the data indicates that the mean vectors are different.

Problem 2b

```r
q2 <- 3
```

```r
mean(female_turtles$x1) + c(-1,1) * qt(1-alpha/(2*q2), df = n2.1 - 1) *
  sd(female_turtles$x1)/sqrt(n2.1)
```

```
## [1] 124.8423 147.2410
```

```r
mean(female_turtles$x2) + c(-1,1) * qt(1-alpha/(2*q2), df = n2.1 - 1) *
  sd(female_turtles$x2)/sqrt(n2.1)
```

```
## [1]  95.6765 109.4902
```

```r
mean(female_turtles$x3) + c(-1,1) * qt(1-alpha/(2*q2), df = n2.1 - 1) *
  sd(female_turtles$x3)/sqrt(n2.1)
```

```
## [1] 47.80103 56.28230
```

```r
mean(male_turtles$x1) + c(-1,1) * qt(1-alpha/(2*q2), df = n2.2 - 1) *
  sd(male_turtles$x1)/sqrt(n2.2)
```

```
## [1] 107.1664 119.5836
```

```r
mean(male_turtles$x2) + c(-1,1) * qt(1-alpha/(2*q2), df = n2.2 - 1) *
  sd(male_turtles$x2)/sqrt(n2.2)
```

```
## [1] 84.56329 92.02004
```

```r
mean(male_turtles$x3) + c(-1,1) * qt(1-alpha/(2*q2), df = n2.2 - 1) *
  sd(male_turtles$x3)/sqrt(n2.2)
```

```
## [1] 38.93984 42.47683
```

Problem 3a

```r
track_records <- read.csv("Track_Records.csv", header = T)

R3.1 <- cor(track_records[,2:8])
R3.1
```

```
##           x1        x2        x3        x4        x5        x6        x7
## x1 1.0000000 0.9410886 0.8707802 0.8091758 0.7815510 0.7278784 0.6689597
## x2 0.9410886 1.0000000 0.9088096 0.8198258 0.8013282 0.7318546 0.6799537
## x3 0.8707802 0.9088096 1.0000000 0.8057904 0.7197996 0.6737991 0.6769384
## x4 0.8091758 0.8198258 0.8057904 1.0000000 0.9050509 0.8665732 0.8539900
## x5 0.7815510 0.8013282 0.7197996 0.9050509 1.0000000 0.9733801 0.7905565
## x6 0.7278784 0.7318546 0.6737991 0.8665732 0.9733801 1.0000000 0.7987302
## x7 0.6689597 0.6799537 0.6769384 0.8539900 0.7905565 0.7987302 1.0000000
```

```r
ev3.1 <- eigen(R3.1)
evalues <- ev3.1$values
evectors<- ev3.1$vectors

evalues
```

```
## [1] 5.80762446 0.62869342 0.27933457 0.12455472 0.09097174 0.05451882
## [7] 0.01430226
```

```r
evectors
```

```
##             [,1]       [,2]       [,3]        [,4]        [,5]        [,6]
```

```
## [1,] -0.3777657 -0.4071756 -0.1405803  0.58706293 -0.16706891  0.53969730
## [2,] -0.3832103 -0.4136291 -0.1007833  0.19407501  0.09350016 -0.74493139
## [3,] -0.3680361 -0.4593531  0.2370255 -0.64543118  0.32727328  0.24009405
## [4,] -0.3947810  0.1612459  0.1475424 -0.29520804 -0.81905467 -0.01650651
## [5,] -0.3892610  0.3090877 -0.4219855 -0.06669044  0.02613100 -0.18898771
## [6,] -0.3760945  0.4231899 -0.4060627 -0.08015699  0.35169796  0.24049968
## [7,] -0.3552031  0.3892153  0.7410610  0.32107640  0.24700821 -0.04826992
##             [,7]
## [1,]  0.08893934
## [2,] -0.26565662
## [3,]  0.12660435
## [4,] -0.19521315
## [5,]  0.73076817
## [6,] -0.57150644
## [7,]  0.08208401
```

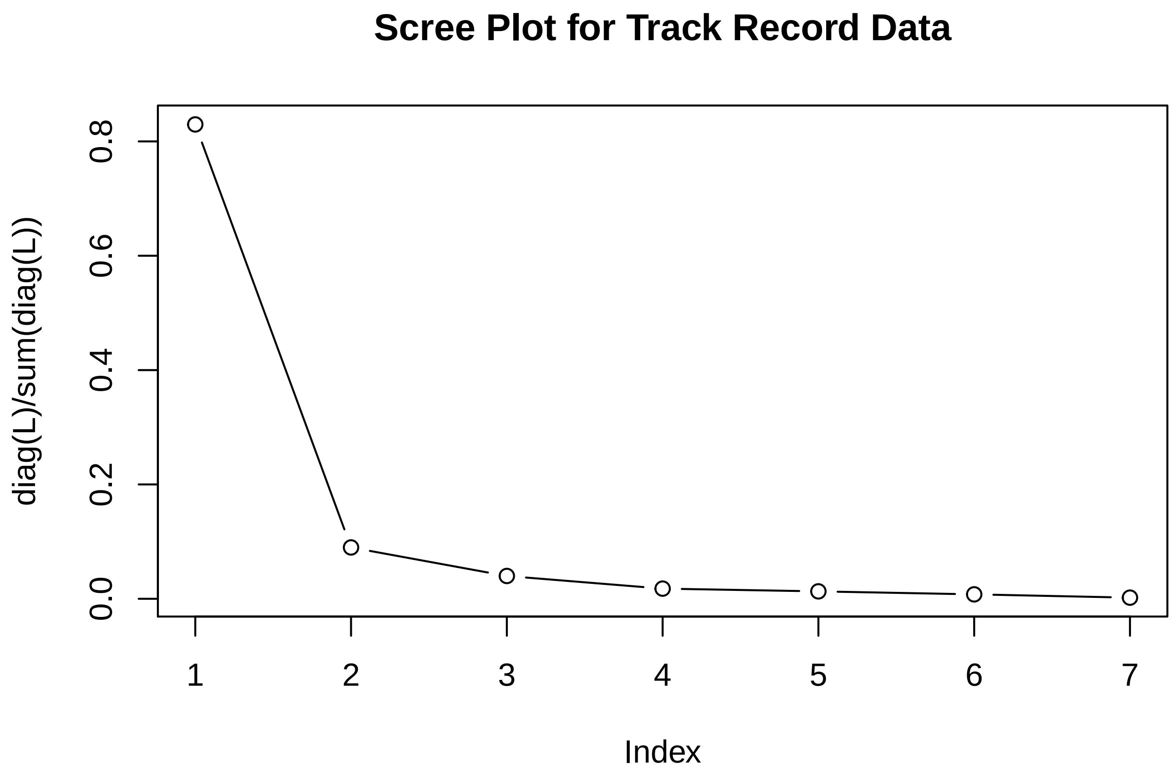
Problem 3b

```r
#Places eigenvalues in a diagonal matrix
L <- round(diag(evalues),6)

#Proportion of variance explained by lambda1,...,lambda7
diag(L)/sum(diag(L))
```

```
## [1] 0.829660571 0.089813286 0.039905000 0.017793571 0.012996000 0.007788429
## [7] 0.002043143
```

```r
#Screeplot of variance
plot(diag(L)/sum(diag(L)), type = "b",
     main = "Scree Plot for Track Record Data")
```




```
#Cumulative proportion of variance explained by lambda1,...,lambda7
cumsum(diag(L))/sum(diag(L))
```

```
## [1] 0.8296606 0.9194739 0.9593789 0.9771724 0.9901684 0.9979569 1.0000000
```

According to the results obtained from the cumulative proportion of explained variance, we would only need to include 2 NPC's to account for 90% of the total standarized variance explained.

Problem 3c

The interpretation of the first NPC seems to be an average of the data since there are no major differences between magnitude or sign. The second NPC appears to have large negative associations with the 100 meter, 200 meter and 400 meter events, so this primarily measures significance of the shorter distances over the longer distances ran by women from their respective countries.

Problem 3d

```
x.bar3.1 <- colMeans(track_records[,2:8])
ones <- rep(1, 54)
X <- as.matrix(track_records[,2:8])

rownames(X) <- track_records$Country
S <- cov(X)
```

```r
D.5 <- diag(1/sqrt(diag(S)))


X.S <- (X - ones%*%t(x.bar3.1))%*%D.5

Z <- X.S %*% evectors
Z
```

```
##              [,1]          [,2]         [,3]         [,4]         [,5]
## ARG   -0.393240234 -0.1316106539 -0.259802233 -0.031900109 -0.300330428
## AUS    1.931642887  0.4910673439 -0.166054687  0.327424485 -0.374654082
## AUT    1.262520373  0.1931483517  0.253382424  0.301303970  0.518269705
## BEL    1.291730279 -0.0024053163 -0.168257439 -0.299729592  0.172461174
## BER   -1.396108552  0.7607805514  0.579260433 -0.044305670  0.299390815
## BRA    1.006778878  0.3795169129 -0.320196905  0.017777114  0.236274580
## CAN    1.734340591  0.2625382896  0.235841588  0.024704466  0.011650507
## CHI   -0.811838204 -0.8689689997 -0.265488361  0.189713853  0.432449790
## CHN    2.989466907  0.0515565410  0.355419942 -0.317729387  0.055408083
## COL   -0.001927672  0.9440511396 -0.452442460  0.376719348 -0.298779150
## COK   -7.906227224 -0.5205487107  1.265650516 -0.168032877  0.390638278
## CRC   -2.166811506  0.3329829275 -0.438831821  0.352569021 -0.211090760
## CZE    2.406030321  0.7596584086 -0.397701406  0.700069078  0.509513490
## DEN    0.082495533 -0.7134670147  0.131816752 -0.110153005 -0.092950809
## DOM   -2.192409809  0.4313474208 -0.341433434  0.170352518  0.032532066
## FIN    1.266731340  0.4263465242  0.050343872 -0.039246225 -0.354342078
## FRA    2.518345696  1.1230568367  0.090577000  0.016227025  0.172556935
## GER    3.047516603  0.9345292649 -0.135098924  0.203123131  0.051857596
## GBR    2.442706280 -0.0333740439 -0.305122395  0.084076503 -0.142380154
## GRE    1.197800425  0.7754294368  0.272767562 -0.416666120  0.144863110
## GUA   -3.294123799 -0.5291973432  0.234255556 -0.001889906 -0.276800645
## HUN    0.788251063 -0.5905189337  0.184140792  0.290919620 -0.106225638
## INA   -1.741942057 -0.5146702995 -0.181528129 -0.645546686 -0.145879192
## IND    0.354256642  0.2542124561  0.226685313  0.386789905  0.124826611
## IRL    1.035907216 -0.7726532308  0.050038407  0.257270930 -0.541479724
## ISR   -0.574161730  0.2181299839 -0.027932891 -0.018756438 -0.322151152
## ITA    1.547452839 -0.2725521643  0.077378876 -0.150588068  0.158682740
## JPN    0.481657610 -0.6557135033 -0.475388171 -0.132470627 -0.233421848
## KEN    0.917735409 -1.3818382037 -0.232365887  0.556085225 -0.193399850
## KOR_S -0.830794629 -0.7687520619 -0.234032763 -0.459980258 -0.551267454
## KOR_N -1.455347346 -2.3771213453 -0.482918445  0.031615673  0.869340755
## LUX   -1.721467731 -1.2782741127 -0.292485944 -0.538676137 -0.015083828
## MAS   -1.495210140  0.5386190883  0.420869393  0.025089791 -0.529723810
## MRI   -1.749727754 -0.5254636441  0.407692531  0.119662243  0.179723993
## MEX    0.995766285  0.4905095362 -0.529409716  0.213763135 -0.478995207
## MYA   -0.815981458 -0.5990664129  0.132119354  0.393743648  0.052982487
## NED    1.544760622 -0.2873591443 -0.087089799 -0.122495478  0.491586187
## NZL    0.755235487 -0.4320195250 -0.136681986  0.134512500  0.207746807
## NOR    0.553003461 -0.9934747091  0.001734929 -0.169756594 -0.459854603
## PNG   -5.257449747  1.1953938028  2.090691992  0.793147042 -0.204640140
## PHI   -1.763533682  0.5797417480  0.250081358 -0.979311794 -0.002008406
## POL    2.273765780  0.4911613673  0.047423287 -0.005819132  0.024631652
## POR    1.175249957 -0.7069615582  0.147263238 -0.063080091 -0.036108058
## ROM    2.123005711 -0.3810120022  0.043757384  0.550078351  0.141165211
## RUS    3.042948214  0.4460682284  0.180508472 -0.132222935  0.302659540
```

```
## SAM   -8.213415123  2.0282582323 -1.974934563  0.067142796  0.254917965
## SIN   -3.093919517 -0.9564211276 -0.886275405  0.197538576 -0.300690828
## ESP    1.889462264  0.2470324869  0.125450420  0.154681156 -0.035096575
## SWE    0.839149567  0.0001607055  0.163946494 -0.184628957  0.146808980
## SUI    1.113545239 -0.5263585776  0.113011526  0.176688501 -0.051220414
## TPE   -0.659093139  1.0063775050 -0.026267033 -0.689453341 -0.102154968
## THA   -1.223805050  0.8469872902 -0.282742485 -0.255653679  0.329380534
## TUR    0.850127798 -0.5785810419  0.760703139 -0.414865140 -0.028468377
## USA    3.299148823  1.1897213000  0.207670730 -0.719831356  0.076878584
##                [,6]         [,7]
## ARG    0.4756553792  0.131258131
## AUS    0.0986621205 -0.057190426
## AUT   -0.0539558884  0.031544437
## BEL    0.2039437451  0.061276014
## BER    0.3920410597 -0.139103026
## BRA    0.0617215562  0.102306317
## CAN   -0.3124603517 -0.081799011
## CHI    0.0179928207 -0.058189850
## CHN   -0.0776050451 -0.030892043
## COL   -0.1505024322  0.100975917
## COK    0.1831260527 -0.074301839
## CRC   -0.1445124782  0.114687910
## CZE    0.2829168306  0.048418589
## DEN   -0.0705133096  0.047962187
## DOM   -0.2216268010  0.170891828
## FIN    0.0669560511  0.100512910
## FRA   -0.2884701503  0.038372423
## GER    0.0195619203  0.019008649
## GBR    0.2393210718  0.111204923
## GRE   -0.4569385240 -0.193437876
## GUA   -0.0925367369  0.050020176
## HUN    0.0554337718 -0.030661901
## INA   -0.0432711006 -0.023002373
## IND    0.2447809879 -0.143168478
## IRL    0.0614651878 -0.086060067
## ISR    0.1289614907 -0.117069965
## ITA    0.0969913564 -0.062379927
## JPN   -0.2079931547  0.053925279
## KEN    0.1276780721 -0.122957771
## KOR_S -0.2567497747 -0.087194933
## KOR_N -0.6344713796  0.063305015
## LUX    0.1839049615  0.260391607
## MAS   -0.0901664041  0.224307626
## MRI    0.0715112359  0.254812053
## MEX   -0.6905244417 -0.121911549
## MYA    0.0223116277 -0.012133729
## NED   -0.2054559118  0.161855818
## NZL   -0.0984761273  0.036085205
## NOR   -0.0419215458 -0.195684242
## PNG   -0.3376207608  0.056762641
## PHI   -0.0281886206 -0.028656689
## POL   -0.0262068788  0.020573945
## POR    0.1320345729 -0.113018916
## ROM    0.3824825713  0.007125618
```

```
## RUS    0.0024926593 -0.090741768
## SAM    0.1060738894 -0.166350888
## SIN    0.2066720007 -0.012289253
## ESP   -0.0499548860  0.075306192
## SWE   -0.0543472171  0.005811323
## SUI    0.1473864264 -0.165420221
## TPE    0.3300803490  0.191580787
## THA   -0.0003953265 -0.251221301
## TUR    0.3205848849 -0.170853706
## USA   -0.0278794061  0.095408226
```

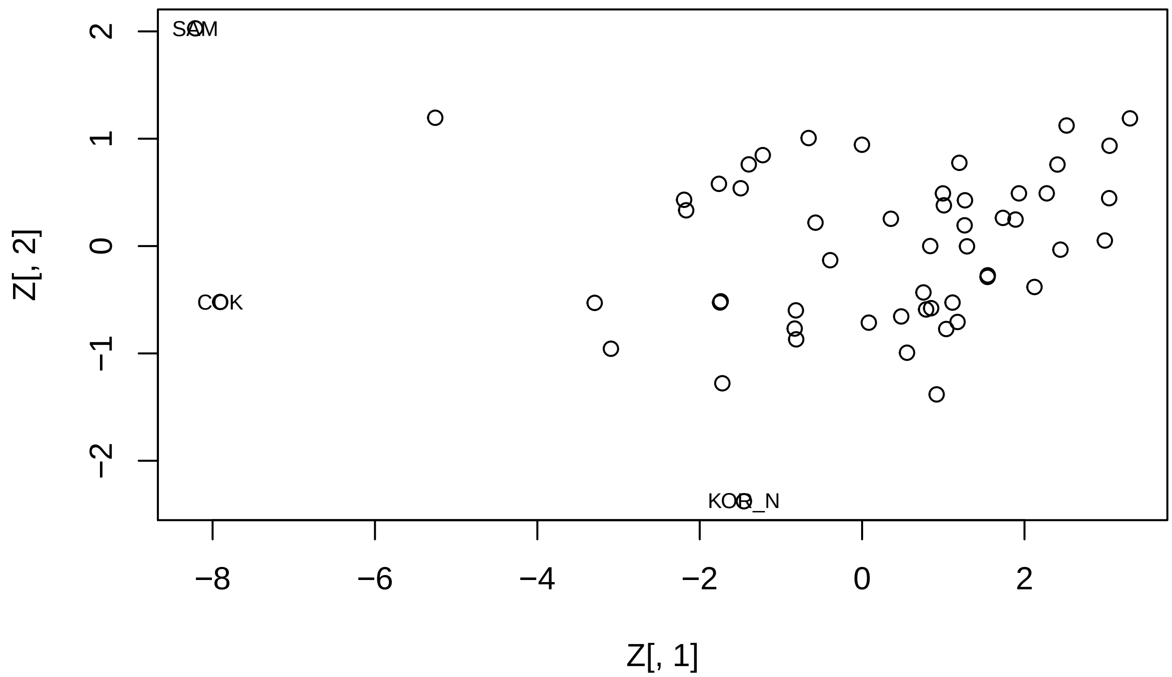
```r
sort(Z[,1], decreasing = T)
```

```
##          USA          GER          RUS          CHN          FRA 
##  3.299148823  3.047516603  3.042948214  2.989466907  2.518345696 
##          GBR          CZE          POL          ROM          AUS 
##  2.442706280  2.406030321  2.273765780  2.123005711  1.931642887 
##          ESP          CAN          ITA          NED          BEL 
##  1.889462264  1.734340591  1.547452839  1.544760622  1.291730279 
##          FIN          AUT          GRE          POR          SUI 
##  1.266731340  1.262520373  1.197800425  1.175249957  1.113545239 
##          IRL          BRA          MEX          KEN          TUR 
##  1.035907216  1.006778878  0.995766285  0.917735409  0.850127798 
##          SWE          HUN          NZL          NOR          JPN 
##  0.839149567  0.788251063  0.755235487  0.553003461  0.481657610 
##          IND          DEN          COL          ARG          ISR 
##  0.354256642  0.082495533 -0.001927672 -0.393240234 -0.574161730 
##          TPE          CHI          MYA        KOR_S          THA 
## -0.659093139 -0.811838204 -0.815981458 -0.830794629 -1.223805050 
##          BER        KOR_N          MAS          LUX          INA 
## -1.396108552 -1.455347346 -1.495210140 -1.721467731 -1.741942057 
##          MRI          PHI          CRC          DOM          SIN 
## -1.749727754 -1.763533682 -2.166811506 -2.192409809 -3.093919517 
##          GUA          PNG          COK          SAM 
## -3.294123799 -5.257449747 -7.906227224 -8.213415123
```

No, this ranking does not exactly correspond with my previous notion of athletic excellence for various countries. This is true because even thought that the country with the fastest time (USA - row 54) is ranked number one by the first principal component, I would have figured that the rest of the rankings would have corresponded with the female runners from the countries that have the next fastest time, and so on.

Problem 3e

```r
plot(Z[,1], Z[,2])
text(-1.455347346, -2.3771213453, labels = "KOR_N", cex = 0.7)
text(-8.213415123, 2.0282582323, labels = "SAM", cex = 0.7)
text(-7.906227224, -0.5205487107, labels = "COK", cex = 0.7)
```



The things that make these points stand out the way they do is because the female world records for those respective countries are close to being (or are) the slowest for the 100-meter and 200-meter track events.

Problem 4a

```
X
```

```
##          x1    x2    x3   x4   x5    x6     x7
## ARG   11.57 22.94 52.50 2.05 4.25  9.19 150.32
## AUS   11.12 22.23 48.63 1.98 4.02  8.63 143.51
## AUT   11.15 22.70 50.62 1.94 4.05  8.78 154.35
## BEL   11.14 22.48 51.45 1.97 4.08  8.82 143.05
## BER   11.46 23.05 53.30 2.07 4.29  9.81 174.18
## BRA   11.17 22.60 50.62 1.97 4.17  9.04 147.41
## CAN   10.98 22.62 49.91 1.97 4.00  8.54 148.36
## CHI   11.65 23.84 53.68 2.00 4.22  9.26 152.23
## CHN   10.79 22.01 49.81 1.93 3.84  8.10 139.39
## COL   11.31 22.92 49.64 2.04 4.34  9.37 155.19
## COK   12.52 25.91 61.65 2.28 4.82 11.10 212.33
## CRC   11.72 23.92 52.57 2.10 4.52  9.84 164.33
## CZE   11.09 21.97 47.99 1.89 4.03  8.87 145.19
## DEN   11.42 23.36 52.92 2.02 4.12  8.71 149.34
## DOM   11.63 23.91 53.02 2.09 4.54  9.89 166.46
## FIN   11.13 22.39 50.14 2.01 4.10  8.69 148.00
## FRA   10.73 21.99 48.25 1.94 4.03  8.64 148.27
```

```
## GER   10.81 21.71 47.60 1.92 3.96  8.51 141.45
## GBR   11.10 22.10 49.43 1.94 3.97  8.37 135.25
## GRE   10.83 22.67 50.56 2.00 4.09  8.96 153.40
## GUA   11.92 24.50 55.64 2.15 4.48  9.71 171.33
## HUN   11.41 23.06 51.50 1.99 4.02  8.55 148.50
## INA   11.56 23.86 55.08 2.10 4.36  9.50 154.29
## IND   11.38 22.82 51.05 2.00 4.10  9.11 158.10
## IRL   11.43 23.02 51.07 2.01 3.98  8.36 142.23
## ISR   11.45 23.15 52.06 2.07 4.24  9.33 156.36
## ITA   11.14 22.60 51.31 1.96 3.98  8.59 143.47
## JPN   11.36 23.33 51.93 2.01 4.16  8.74 139.41
## KEN   11.62 23.37 51.56 1.97 3.96  8.39 138.47
## KOR_S 11.49 23.80 53.67 2.09 4.24  9.01 146.12
## KOR_N 11.80 25.10 56.23 1.97 4.25  8.96 145.31
## LUX   11.76 23.96 56.07 2.07 4.35  9.21 149.23
## MAS   11.50 23.37 52.56 2.12 4.39  9.31 169.28
## MRI   11.72 23.83 54.62 2.06 4.33  9.24 167.09
## MEX   11.09 23.13 48.89 2.02 4.19  8.89 144.06
## MYA   11.66 23.69 52.96 2.03 4.20  9.08 158.42
## NED   11.08 22.81 51.35 1.93 4.06  8.57 143.43
## NZL   11.32 23.13 51.60 1.97 4.10  8.76 146.46
## NOR   11.41 23.31 52.45 2.03 4.01  8.53 141.06
## PNG   11.96 24.68 55.18 2.24 4.62 10.21 221.14
## PHI   11.28 23.35 54.75 2.12 4.41  9.81 165.48
## POL   10.93 22.13 49.28 1.95 3.99  8.53 144.18
## POR   11.30 22.88 51.92 1.98 3.96  8.50 143.29
## ROM   11.30 22.35 49.88 1.92 3.90  8.36 142.50
## RUS   10.77 21.87 49.11 1.91 3.87  8.38 141.31
## SAM   12.38 25.45 56.32 2.29 5.42 13.12 191.58
## SIN   12.13 24.54 55.08 2.12 4.52  9.94 154.41
## ESP   11.06 22.38 49.67 1.96 4.01  8.48 146.51
## SWE   11.16 22.82 51.69 1.99 4.09  8.81 150.39
## SUI   11.34 22.88 51.32 1.98 3.97  8.60 145.51
## TPE   11.22 22.56 52.74 2.08 4.38  9.63 159.53
## THA   11.33 23.30 52.60 2.06 4.38 10.07 162.39
## TUR   11.25 22.71 53.15 2.01 3.92  8.53 151.43
## USA   10.49 21.34 48.83 1.94 3.95  8.43 141.16
```

```r
#Calculation of Speed for various events:
X[,1] <- 100/X[,1]
X[,2] <- 200/X[,2]
X[,3] <- 400/X[,3]
X[,4] <- 800/(X[,4]*60)
X[,5] <- 1500/(X[,5]*60)
X[,6] <- 3000/(X[,6]*60)
X[,7] <- 42195/(X[,7]*60)

X
```

```
##             x1       x2       x3       x4       x5       x6       x7
## ARG   8.643042 8.718396 7.619048 6.504065 5.882353 5.440696 4.678353
## AUS   8.992806 8.996851 8.225375 6.734007 6.218905 5.793743 4.900355
## AUT   8.968610 8.810573 7.902015 6.872852 6.172840 5.694761 4.556203
## BEL   8.976661 8.896797 7.774538 6.768190 6.127451 5.668934 4.916113
## BER   8.726003 8.676790 7.504690 6.441224 5.827506 5.096840 4.037490
```

```
## BRA   8.952551 8.849558 7.902015 6.768190 5.995204 5.530973 4.770708
## CAN   9.107468 8.841733 8.014426 6.768190 6.250000 5.854801 4.740159
## CHI   8.583691 8.389262 7.451565 6.666667 5.924171 5.399568 4.619654
## CHN   9.267841 9.086779 8.030516 6.908463 6.510417 6.172840 5.045197
## COL   8.841733 8.726003 8.058018 6.535948 5.760369 5.336179 4.531542
## COK   7.987220 7.719027 6.488240 5.847953 5.186722 4.504505 3.312061
## CRC   8.532423 8.361204 7.608902 6.349206 5.530973 5.081301 4.279499
## CZE   9.017133 9.103323 8.335070 7.054674 6.203474 5.636979 4.843653
## DEN   8.756567 8.561644 7.558579 6.600660 6.067961 5.740528 4.709053
## DOM   8.598452 8.364701 7.544323 6.379585 5.506608 5.055612 4.224739
## FIN   8.984726 8.932559 7.977663 6.633499 6.097561 5.753740 4.751689
## FRA   9.319664 9.095043 8.290155 6.872852 6.203474 5.787037 4.743036
## GER   9.250694 9.212345 8.403361 6.944444 6.313131 5.875441 4.971721
## GBR   9.009009 9.049774 8.092252 6.872852 6.297229 5.973716 5.199630
## GRE   9.233610 8.822232 7.911392 6.666667 6.112469 5.580357 4.584420
## GUA   8.389262 8.163265 7.189073 6.201550 5.580357 5.149331 4.104652
## HUN   8.764242 8.673027 7.766990 6.700168 6.218905 5.847953 4.735690
## INA   8.650519 8.382230 7.262164 6.349206 5.733945 5.263158 4.557975
## IND   8.787346 8.764242 7.835455 6.666667 6.097561 5.488474 4.448134
## IRL   8.748906 8.688097 7.832387 6.633499 6.281407 5.980861 4.944456
## ISR   8.733624 8.639309 7.683442 6.441224 5.896226 5.359057 4.497634
## ITA   8.976661 8.849558 7.795751 6.802721 6.281407 5.820722 4.901722
## JPN   8.802817 8.572653 7.702677 6.633499 6.009615 5.720824 5.044473
## KEN   8.605852 8.557980 7.757952 6.768190 6.313131 5.959476 5.078717
## KOR_S 8.703220 8.403361 7.452953 6.379585 5.896226 5.549390 4.812825
## KOR_N 8.474576 7.968127 7.113640 6.768190 5.882353 5.580357 4.839653
## LUX   8.503401 8.347245 7.133940 6.441224 5.747126 5.428882 4.712524
## MAS   8.695652 8.557980 7.610350 6.289308 5.694761 5.370569 4.154360
## MRI   8.532423 8.392782 7.323325 6.472492 5.773672 5.411255 4.208810
## MEX   9.017133 8.646779 8.181632 6.600660 5.966587 5.624297 4.881647
## MYA   8.576329 8.442381 7.552870 6.568144 5.952381 5.506608 4.439149
## NED   9.025271 8.768084 7.789679 6.908463 6.157635 5.834306 4.903089
## NZL   8.833922 8.646779 7.751938 6.768190 6.097561 5.707763 4.801652
## NOR   8.764242 8.580009 7.626311 6.568144 6.234414 5.861665 4.985467
## PNG   8.361204 8.103728 7.249003 5.952381 5.411255 4.897160 3.180112
## PHI   8.865248 8.565310 7.305936 6.289308 5.668934 5.096840 4.249758
## POL   9.149131 9.037506 8.116883 6.837607 6.265664 5.861665 4.877584
## POR   8.849558 8.741259 7.704160 6.734007 6.313131 5.882353 4.907879
## ROM   8.849558 8.948546 8.019246 6.944444 6.410256 5.980861 4.935088
## RUS   9.285051 9.144947 8.144981 6.980803 6.459948 5.966587 4.976647
## SAM   8.077544 7.858546 7.102273 5.822416 4.612546 3.810976 3.670790
## SIN   8.244023 8.149959 7.262164 6.289308 5.530973 5.030181 4.554433
## ESP   9.041591 8.936550 8.053151 6.802721 6.234414 5.896226 4.800014
## SWE   8.960573 8.764242 7.738441 6.700168 6.112469 5.675369 4.676175
## SUI   8.818342 8.741259 7.794232 6.734007 6.297229 5.813953 4.833001
## TPE   8.912656 8.865248 7.584376 6.410256 5.707763 5.192108 4.408262
## THA   8.826125 8.583691 7.604563 6.472492 5.707763 4.965243 4.330624
## TUR   8.888889 8.806693 7.525870 6.633499 6.377551 5.861665 4.644060
## USA   9.532888 9.372071 8.191685 6.872852 6.329114 5.931198 4.981935
```

Problem 4b

```
S4.1 <- cov(X)

ev4.1 <- eigen(S4.1)
new_evalues <- ev4.1$values
new_evectors<- ev4.1$vectors


#Places eigenvalues in a diagonal matrix
L4.1 <- round(diag(new_evalues),6)

#Proportion of variance explained by lambda1,...,lambda7
diag(L4.1)/sum(diag(L4.1))
```

```
## [1] 0.828538320 0.097403869 0.037774667 0.016948807 0.010016285 0.006980052
## [7] 0.002338001
```

```
#Cumulative proportion of variance explained by lambda1,...,lambda7
cumsum(diag(L4.1))/sum(diag(L4.1))
```

```
## [1] 0.8285383 0.9259422 0.9637169 0.9806657 0.9906819 0.9976620 1.0000000
```

We should retain 2 principal components if our goal is to account for 90% of total sample variance.

Problem 4c

```
new_evectors
```

```
##            [,1]        [,2]        [,3]        [,4]        [,5]
## [1,] -0.3102442 -0.37596510 -0.09755628  0.58479630  0.04613051
## [2,] -0.3573948 -0.43376925 -0.08896099  0.32287531  0.02977941
## [3,] -0.3787367 -0.51873227  0.27424547 -0.66667306  0.18727340
## [4,] -0.2993405  0.05313551  0.05252266 -0.12808676 -0.89434367
## [5,] -0.3912131  0.21084397 -0.43498609 -0.05510789 -0.12725405
## [6,] -0.4595909  0.39557338 -0.42664455 -0.18388862  0.35674301
## [7,] -0.4227291  0.44458346  0.73031571  0.23675670  0.13639673
##             [,6]        [,7]
## [1,]  0.62433141  0.13775753
## [2,] -0.68870961 -0.31103524
## [3,]  0.12377209  0.13198849
## [4,]  0.13592439 -0.26472817
## [5,] -0.23626094  0.73364469
## [6,]  0.19925854 -0.49948755
## [7,] -0.08106294  0.09516116
```

The interpretation of the first NPC seems to be an average of the data since there are no major differences between magnitude or sign. The second NPC appears to have large negative associations with the 100 meter, 200 meter and 400 meter events, so this primarily measures significance in the speed of runners of the shorter distances over the speed of runners of longer distances ran by women from their respective countries.

These interpretations are very similar to those of the first two NPCs in the previous excerise.

Problem 4d

```
x.bar4.1 <- colMeans(X)
D_0.5 <- diag(1/sqrt(diag(S4.1)))

X.S2 <- (X - ones%*%t(x.bar4.1))%*%D_0.5

Z2 <- X.S2 %*% new_evectors
Z2
```

```
##              [,1]         [,2]         [,3]         [,4]         [,5]
## ARG    0.49629927  0.158702088  0.293509632  0.027216573  0.2334368009
## AUS   -2.02081066 -0.661551534  0.222094377 -0.303564974  0.1111995149
## AUT   -1.10684191 -0.409123770 -0.391260889 -0.162063047 -0.7123249222
## BEL   -1.25445492 -0.031618978  0.195099080  0.441288231 -0.3078923730
## BER    1.74127213 -0.798519917 -0.548486448  0.171309257 -0.0940920465
## BRA   -0.86135883 -0.486554423  0.352964805  0.117209923 -0.3693024035
## CAN   -1.73331742 -0.402955101 -0.310522191  0.009621708 -0.1169578277
## CHI    0.95550854  0.842801687  0.192053802 -0.200551689 -0.4950618751
## CHN   -3.28171979 -0.124673619 -0.487025069  0.456714033 -0.2476437272
## COL    0.20223993 -1.037822745  0.551886563 -0.403110388  0.2872108478
## COK    7.31711331  0.924204600 -0.827316160 -0.133022322  0.6389296137
## CRC    2.36528288 -0.244840961  0.471531574 -0.455951334  0.3794379951
## CZE   -2.37687859 -1.137341474  0.406209050 -0.465118321 -1.0244750438
## DEN   -0.06630281  0.747956123 -0.207011022  0.018897199  0.0856262815
## DOM    2.43774222 -0.334028995  0.365630863 -0.239349627  0.2257972010
## FIN   -1.24836100 -0.514865116 -0.036183271  0.062104144  0.2710966089
## FRA   -2.49346795 -1.445908890 -0.090425800  0.157510381 -0.3079955481
## GER   -3.19604894 -1.256235599  0.182362110 -0.025082794 -0.3754848738
## GBR   -2.70457852  0.005863049  0.387238070  0.034401309 -0.1823431856
## GRE   -1.01130568 -0.921823098 -0.272008955  0.520521712 -0.0517406170
## GUA    3.36487981  0.648588477 -0.211679077 -0.238041644  0.5967797081
## HUN   -0.82713669  0.544512000 -0.313510742 -0.336499909 -0.0809511704
## INA    1.86225672  0.661235108  0.234116569  0.466467917  0.3984762912
## IND   -0.15432957 -0.427642539 -0.300241133 -0.302192429 -0.2817076000
## IRL   -1.25133824  0.829790621 -0.115884225 -0.384790083  0.3308796409
## ISR    0.75101263 -0.220790407  0.062817205 -0.029347095  0.3454329505
## ITA   -1.58662272  0.225471783 -0.140280346  0.242959731 -0.3468612656
## JPN   -0.56974087  0.792846723  0.579996447  0.057222130  0.1744887821
## KEN   -1.19014122  1.450687583  0.158421108 -0.634579723 -0.1690766098
## KOR_S  0.80010778  0.951818543  0.322956037  0.229141352  0.6739155143
## KOR_N  1.36618143  2.395392864  0.369705582 -0.200201768 -0.8120729464
## LUX    1.72419931  1.419270728  0.305978889  0.349039269  0.1922028215
## MAS    1.68505986 -0.480436688 -0.363591665 -0.150192489  0.7646037766
## MRI    1.92974122  0.516992943 -0.485301205 -0.194566895  0.0002397687
## MEX   -0.99939319 -0.539579496  0.698483186 -0.345702159  0.4430362499
## MYA    0.95089073  0.544202756 -0.230159722 -0.455540389 -0.0950806216
## NED   -1.53997125  0.238423813  0.023002986  0.234938333 -0.6416622686
## NZL   -0.71108519  0.392585670  0.083155929 -0.112601880 -0.3576290964
## NOR   -0.73051642  1.128552630 -0.004910874  0.026239850  0.3648207670
## PNG    5.09155478 -0.718047231 -1.379554602 -0.738535697  0.9775421601
## PHI    2.04237208 -0.445605098 -0.164935022  0.912282184  0.4464960891
## POL   -2.33126568 -0.669881357 -0.066845241  0.134682181 -0.2245544544
```

```
## POR   -1.28101024  0.721267299 -0.242372066  0.061262206 -0.1525311102
## ROM   -2.29746224  0.237213187 -0.194678710 -0.448936248 -0.6118343112
## RUS   -3.20485331 -0.665505263 -0.260129220  0.364285443 -0.6004882482
## SAM    7.50383487 -0.836719247  1.610948162 -0.313774882  0.7977557844
## SIN    3.07676036  1.110156058  0.896417920 -0.385050692  0.3877225486
## ESP   -1.93304667 -0.394618112 -0.194508491 -0.094847563 -0.1522797360
## SWE   -0.72801296 -0.068781788 -0.221886211  0.241000727 -0.1851409830
## SUI   -1.17071704  0.478953233 -0.210534138 -0.172170709 -0.1892049412
## TPE    0.94785706 -0.968919079  0.108828210  0.777328634  0.3414754296
## THA    1.57343255 -0.814249145  0.329286052  0.329302285 -0.1101703392
## TUR   -0.87360639  0.549623861 -0.954007179  0.413769256 -0.0281773841
## USA   -3.44990253 -1.458473759 -0.179444534  1.068670784 -0.1438656168
##                [,6]         [,7]
## ARG   -0.5351840179 -0.152488220
## AUS   -0.1609588863  0.076704010
## AUT    0.1817890595 -0.072402362
## BEL   -0.1250698353 -0.071101626
## BER   -0.3470538455  0.089415073
## BRA    0.0157240448 -0.139028433
## CAN    0.3791123605  0.118207820
## CHI    0.0008018837  0.026734094
## CHN    0.2045892468  0.071290247
## COL    0.0854440867 -0.107842348
## COK   -0.2750337149  0.047812056
## CRC    0.0254110434 -0.125773549
## CZE   -0.1824190430 -0.143070945
## DEN    0.0588987651 -0.034632285
## DOM    0.1640473362 -0.184401542
## FIN   -0.0879330359 -0.097968822
## FRA    0.4471555104 -0.037230072
## GER    0.0620120619 -0.029365109
## GBR   -0.2364212868 -0.128265913
## GRE    0.5883778577  0.259552022
## GUA   -0.0492567023 -0.033290498
## HUN   -0.0880232899  0.036730707
## INA    0.0067479708  0.057576793
## IND   -0.2514618964  0.127876062
## IRL   -0.1824943358  0.125186646
## ISR   -0.2085639120  0.129717493
## ITA   -0.0331400856  0.083303431
## JPN    0.1602164815 -0.024435496
## KEN   -0.2300829310  0.144087388
## KOR_S  0.1281429947  0.157962905
## KOR_N  0.6392409958 -0.106480140
## LUX   -0.1849350969 -0.246560359
## MAS   -0.0136036232 -0.204333273
## MRI   -0.0646827049 -0.279391542
## MEX    0.6128288367  0.195046523
## MYA   -0.0704622260 -0.003705457
## NED    0.3736701735 -0.198865045
## NZL    0.1409152933 -0.049986393
## NOR   -0.0654314777  0.275289655
## PNG    0.0862003611  0.004854616
## PHI    0.0904842780  0.052093925
```

```
## POL    0.1018118542 -0.017352382
## POR   -0.1348920197  0.153420274
## ROM   -0.3698195238 -0.038658467
## RUS    0.1461179812  0.119783953
## SAM   -0.1940248915 -0.084773659
## SIN   -0.3678290277 -0.003752396
## ESP    0.0980303547 -0.086769678
## SWE    0.1257521284  0.002059477
## SUI   -0.1765190578  0.201855083
## TPE   -0.2800528697 -0.216123907
## THA    0.0588038853  0.200853517
## TUR   -0.2968773156  0.243460329
## USA    0.2298998073 -0.082824184
```

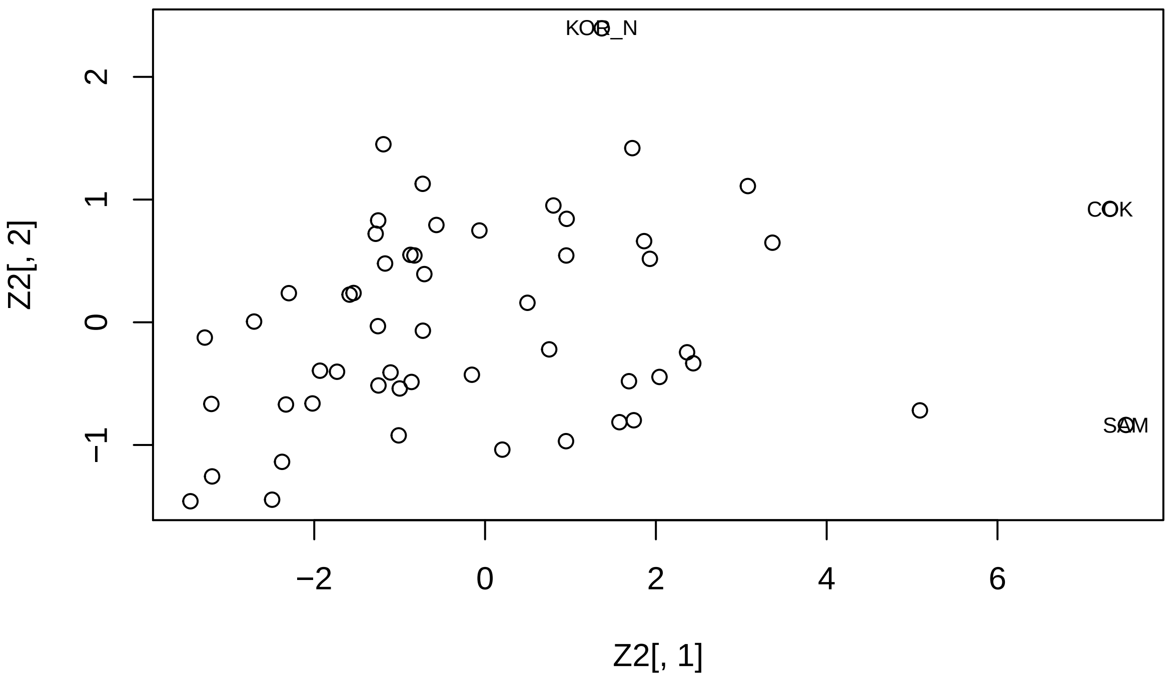
```r
sort(Z2[,1], decreasing = F)
```

```
##         USA         CHN         RUS         GER         GBR         FRA
## -3.44990253 -3.28171979 -3.20485331 -3.19604894 -2.70457852 -2.49346795
##         CZE         POL         ROM         AUS         ESP         CAN
## -2.37687859 -2.33126568 -2.29746224 -2.02081066 -1.93304667 -1.73331742
##         ITA         NED         POR         BEL         IRL         FIN
## -1.58662272 -1.53997125 -1.28101024 -1.25445492 -1.25133824 -1.24836100
##         KEN         SUI         AUT         GRE         MEX         TUR
## -1.19014122 -1.17071704 -1.10684191 -1.01130568 -0.99939319 -0.87360639
##         BRA         HUN         NOR         SWE         NZL         JPN
## -0.86135883 -0.82713669 -0.73051642 -0.72801296 -0.71108519 -0.56974087
##         IND         DEN         COL         ARG         ISR       KOR_S
## -0.15432957 -0.06630281  0.20223993  0.49629927  0.75101263  0.80010778
##         TPE         MYA         CHI       KOR_N         THA         MAS
##  0.94785706  0.95089073  0.95550854  1.36618143  1.57343255  1.68505986
##         LUX         BER         INA         MRI         PHI         CRC
##  1.72419931  1.74127213  1.86225672  1.92974122  2.04237208  2.36528288
##         DOM         SIN         GUA         PNG         COK         SAM
##  2.43774222  3.07676036  3.36487981  5.09155478  7.31711331  7.50383487
```

Other than the negative signs, there aren't really too many differences from this ranking compared to the ranking from Problem 3. There were a few countries that appeared to leap over others in the speed ranking but the female runners from USA still appears to be the fastest runner and the runners from Samoa still seem to be the slowest.

Problem 4e

```r
plot(Z2[,1], Z2[,2])
text(1.36618143, 2.395392864, labels = "KOR_N", cex = 0.7)
text(7.50383487, -0.836719247, labels = "SAM", cex = 0.7)
text(7.31711331, 0.924204600, labels = "COK", cex = 0.7)
```



This plot seems to be the similar to the previous plot but just rotated 180 degrees clockwise. It also appears that, much like the previous plot, runners from North Korea, the Cook Islands and Samoa slow times have also affected their speeds as they seem to be the slowest runners.