STAT4051Hw6

Mingming Xu

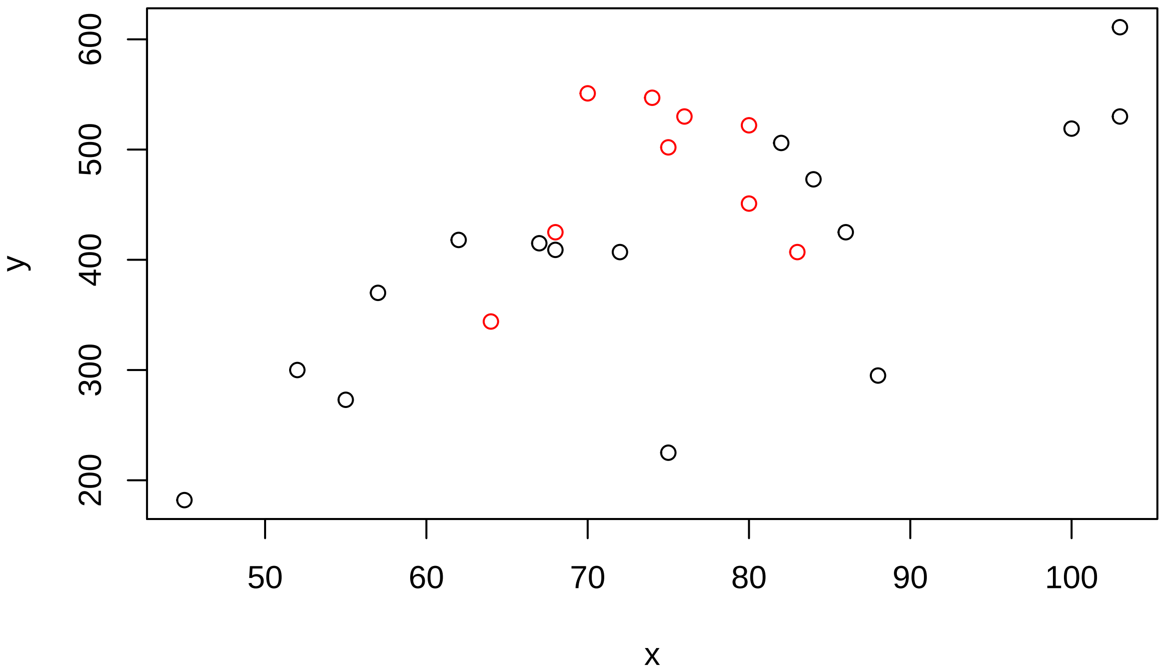
11/17/2019

∗ ∗ ∗Cathedral Problem

```
library (faraway)
data("cathedral",package = "faraway")
```

a.

```
plot(y~x,col=cathedral$style,data=cathedral)
```



b.

```
##Check to determine if treatment affects covariate
cathedral$style=as.factor(cathedral$style)
model.0=lm(x~style,data=cathedral)
anova(model.0)
```

```
## Analysis of Variance Table
##
## Response: x
##           Df Sum Sq Mean Sq F value Pr(>F)
## style      1    1.4    1.40   0.006 0.9389
## Residuals 23 5365.2  233.27
```

```
#The treatment does not affect the covariate
#So we move on to testing the treatment by covariate interaction
model.1<-lm(y~style*x,data = cathedral)
anova(model.1)
```

```
## Analysis of Variance Table
```

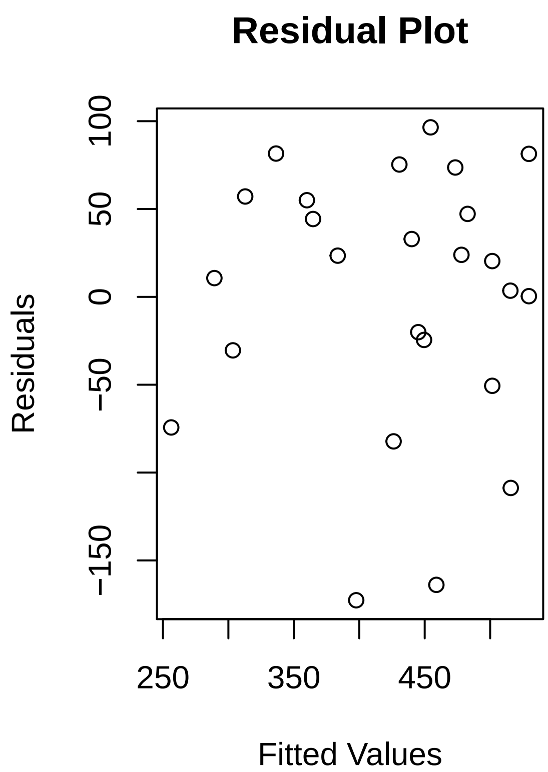
```
##
## Response: y
##           Df Sum Sq Mean Sq F value   Pr(>F)
## style      1  35106   35106  5.6100 0.027527 *
## x          1 119103  119103 19.0329 0.000273 ***
## style:x    1    810     810  0.1294 0.722657
## Residuals 21 131413    6258
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

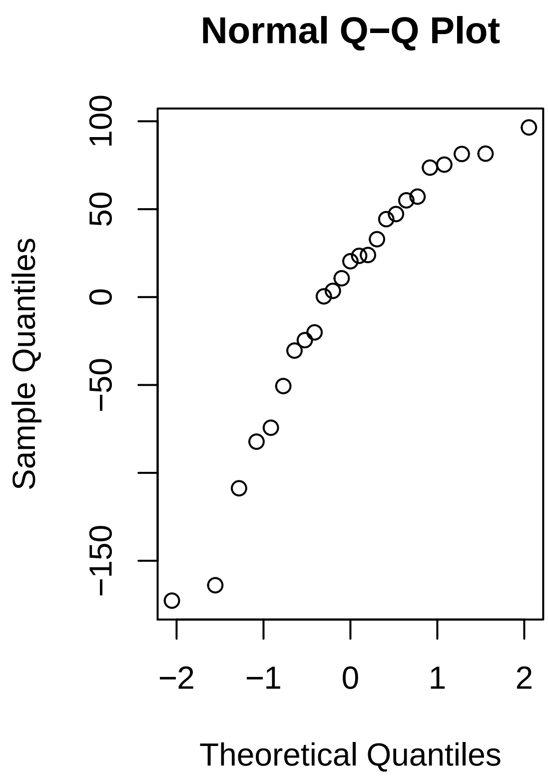
```r
#The treatment by covariate interaction is not statistically significant
#So we will go with a parallel lines
model.2<-lm(y~x+style,data=cathedral )
summary(model.2);anova(model.2)
```

```
##
## Call:
## lm(formula = y ~ x + style, data = cathedral)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -172.67  -30.44   20.38   55.02   96.50
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   44.298     81.648   0.543   0.5929
## x              4.712      1.058   4.452   0.0002 ***
## styler        80.393     32.306   2.488   0.0209 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 77.53 on 22 degrees of freedom
## Multiple R-squared:  0.5384, Adjusted R-squared:  0.4964
## F-statistic: 12.83 on 2 and 22 DF,  p-value: 0.0002028
```

```
## Analysis of Variance Table
##
## Response: y
##           Df Sum Sq Mean Sq F value    Pr(>F)
## x          1 116992  116992 19.4659 0.0002205 ***
## style      1  37217   37217  6.1924 0.0208871 *
## Residuals 22 132223    6010
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
#check assumptions:
par(mfrow=c(1,2))
plot(model.2$residuals~model.2$fitted.values,
     xlab="Fitted Values", ylab="Residuals",
     main="Residual Plot")
qqnorm(model.2$residuals)
```






```
par(mfrow=c(1,1))
#Assumptions look good
```

c.

```
##Yes.I include a covariate in my final model
model.n<-lm(y~style,data=cathedral)
summary(model.n);anova(model.n)
```

```
##
## Call:
## lm(formula = y ~ style, data = cathedral)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -215.38  -68.44   17.62   71.56  213.62
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   397.37      26.13  15.206 1.72e-13 ***
## styler         78.07      43.56   1.792   0.0862 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 104.5 on 23 degrees of freedom
## Multiple R-squared:  0.1226, Adjusted R-squared:  0.08441
## F-statistic: 3.213 on 1 and 23 DF,  p-value: 0.08623
```

```
## Analysis of Variance Table
##
## Response: y
```

```
##           Df Sum Sq Mean Sq F value  Pr(>F)
## style      1  35106   35106  3.2127 0.08623 .
## Residuals 23 251326   10927
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
##Only style:df is 23;MSE is 10927;p-value is 0.08623
##style+x:df is 22;MSE is  6010;p-value is 0.0209
##Benefit: There is a varaiance reduction due to addition of the covaraite x.And we observe that style l
```

d.

```
##beta*10
4.712*10
```

```
## [1] 47.12
```

e.

```
## 80.393
```

∗ ∗ ∗Oehlert Problem 17.1

```
prob17.1<- read.table("http://www.stat.umn.edu/~gary/book/fcdae.data/pr17.1",header=TRUE)
head(prob17.1)
```

```
##   pesticide diameter calcium.conc
## 1         1     2.48        10.41
## 2         2     3.10        12.10
## 3         3     2.57        10.33
## 4         4     2.60        10.46
## 5         1     2.81        11.82
## 6         2     2.61        10.38
```

a.

```
#Check to determine if treatment affects covariate
prob17.1$pesticide=as.factor(prob17.1$pesticide)
model_0=lm(diameter~pesticide,data=prob17.1)
anova(model_0)
```

```
## Analysis of Variance Table
##
## Response: diameter
##           Df  Sum Sq  Mean Sq F value Pr(>F)
## pesticide  3 0.23653 0.078845  1.8461 0.1618
## Residuals 28 1.19586 0.042709
```

```
#The p-value is 0.1618 which is larger than 0.05, the treatment does not affect the covariate.
#Testing the treatment by covariate interaction
model_1<-lm(calcium.conc~pesticide*diameter,data=prob17.1)
anova(model_1)
```

```
## Analysis of Variance Table
##
## Response: calcium.conc
##                    Df  Sum Sq Mean Sq   F value    Pr(>F)
## pesticide           3 10.8607  3.6202  437.9085 < 2.2e-16 ***
## diameter            1 18.5668 18.5668 2245.8653 < 2.2e-16 ***
## pesticide:diameter  3  0.1255  0.0418    5.0587  0.007413 **
```

```
## Residuals          24  0.1984  0.0083
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
#The treatment by covariate interaction is statistically significant
#So we will use seperate lines model
```

b.

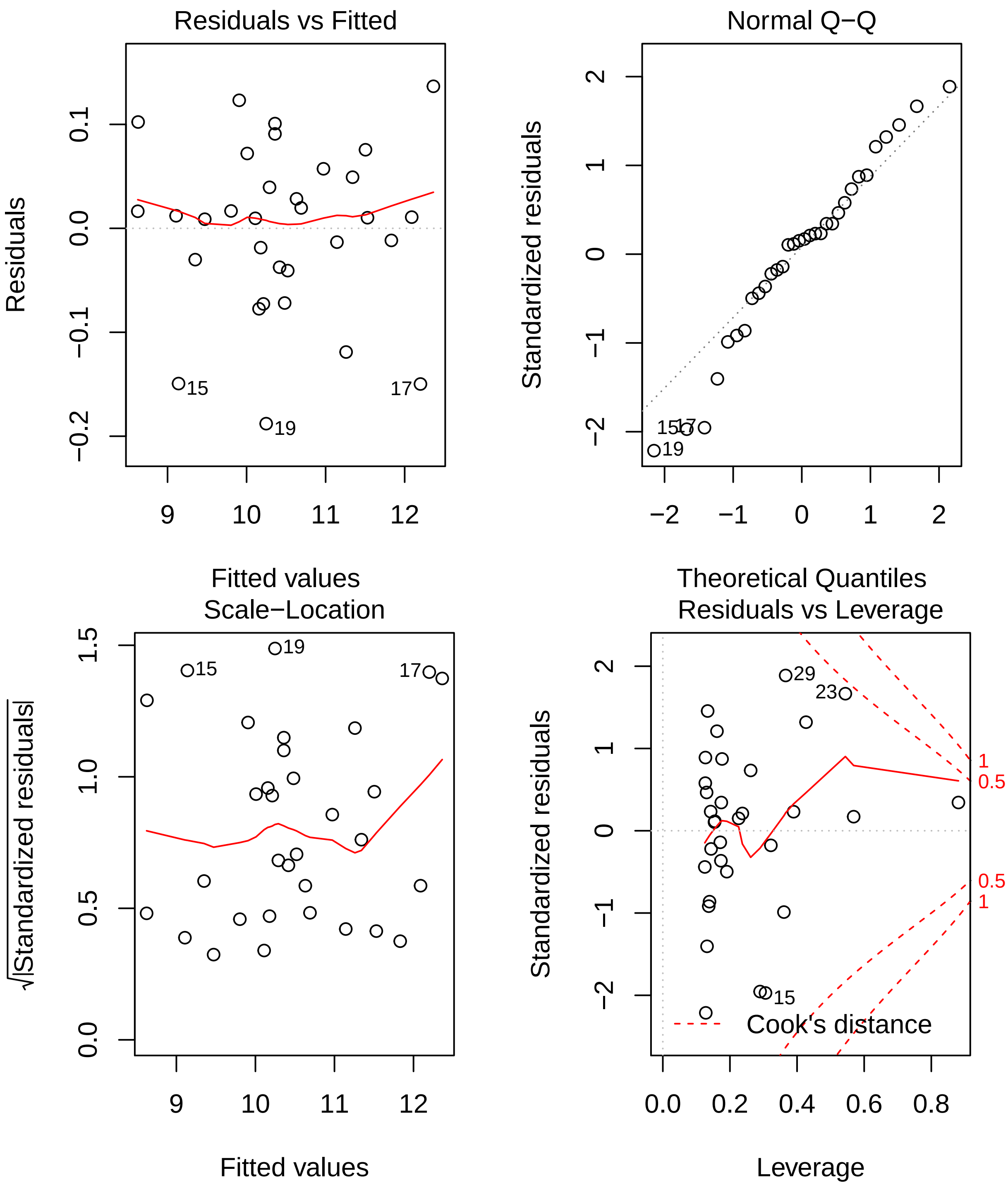
```r
##final model:
model_2<-lm(calcium.conc~pesticide*diameter,data=prob17.1)
summary(model_2);anova(model_2)
```

```
##
## Call:
## lm(formula = calcium.conc ~ pesticide * diameter, data = prob17.1)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.18791 -0.03824  0.01049  0.05121  0.13663
##
## Coefficients:
##                     Estimate Std. Error t value Pr(>|t|)
## (Intercept)          0.33804    0.52384   0.645   0.5248
## pesticide2           1.17449    0.67692   1.735   0.0956 .
## pesticide3          -1.00515    0.67558  -1.488   0.1498
## pesticide4          -0.47351    0.64229  -0.737   0.4681
## diameter             4.09025    0.19302  21.190   <2e-16 ***
## pesticide2:diameter -0.67841    0.25291  -2.682   0.0130 *
## pesticide3:diameter  0.17343    0.25599   0.677   0.5046
## pesticide4:diameter -0.05382    0.24389  -0.221   0.8272
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.09092 on 24 degrees of freedom
## Multiple R-squared:  0.9933, Adjusted R-squared:  0.9914
## F-statistic: 510.7 on 7 and 24 DF,  p-value: < 2.2e-16
```

```
## Analysis of Variance Table
##
## Response: calcium.conc
##                    Df  Sum Sq Mean Sq   F value    Pr(>F)
## pesticide           3 10.8607  3.6202  437.9085 < 2.2e-16 ***
## diameter            1 18.5668 18.5668 2245.8653 < 2.2e-16 ***
## pesticide:diameter  3  0.1255  0.0418    5.0587  0.007413 **
## Residuals          24  0.1984  0.0083
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

c.

```r
par(mfrow=c(1,2))
plot(model_2)
```




```
#Assumptions look good
```

d.

```
##pesticide1:
0.33804+4.09025*3
```

```
## [1] 12.60879
```

```
##pesticide2:
0.33804+1.17449+(4.09025-0.67841)*3
```

```
## [1] 11.74805
```

```
##pesticide3:
0.33804-1.00515+(4.09025+0.17343)*3
```

```
## [1] 12.12393
```

```
##pesticide4:
0.33804-0.47351+(4.09025+0.052382)*3
```

```
## [1] 12.29243
```

e.

```
model_n<-lm(calcium.conc~pesticide,data=prob17.1)
summary(model_n);anova(model_n)
```

```
##
## Call:
## lm(formula = calcium.conc ~ pesticide, data = prob17.1)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.39000 -0.45000 -0.05937  0.54750  1.66125
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  11.4175     0.2904  39.316  < 2e-16 ***
## pesticide2   -0.9787     0.4107  -2.383 0.024187 *
## pesticide3   -1.2975     0.4107  -3.159 0.003774 **
## pesticide4   -1.5275     0.4107  -3.719 0.000887 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.8214 on 28 degrees of freedom
## Multiple R-squared:  0.365,  Adjusted R-squared:  0.297
## F-statistic: 5.366 on 3 and 28 DF,  p-value: 0.004784
```

```
## Analysis of Variance Table
##
## Response: calcium.conc
##           Df Sum Sq Mean Sq F value   Pr(>F)
## pesticide  3 10.861  3.6202   5.366 0.004784 **
## Residuals 28 18.891  0.6747
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
##Only pesticide: df is 28;MSE is 0.6747;p-value is 0.004784.
##pesticide+diameter: df is 24;MSE is 0.0083;p-value is < 2.2e-16.
##Benefit: There is a varaiance reduction due to addition of the covaraite x.
```

Glass Problem

```
glass<- read.csv("https://archive.ics.uci.edu/ml/machine-learning-databases/glass/glass.data")
attach(glass)
# Name the variables
```

```r
colnames(glass) <- c("id","ri","Na","Mg","Al", "Si", "K",
                     "Ca", "Ba", "Fe", "Type")
head(glass)
```

```
##   id      ri    Na   Mg   Al    Si    K   Ca Ba   Fe Type
## 1  2 1.51761 13.89 3.60 1.36 72.73 0.48 7.83  0 0.00    1
## 2  3 1.51618 13.53 3.55 1.54 72.99 0.39 7.78  0 0.00    1
## 3  4 1.51766 13.21 3.69 1.29 72.61 0.57 8.22  0 0.00    1
## 4  5 1.51742 13.27 3.62 1.24 73.08 0.55 8.07  0 0.00    1
## 5  6 1.51596 12.79 3.61 1.62 72.97 0.64 8.07  0 0.26    1
## 6  7 1.51743 13.30 3.60 1.14 73.09 0.58 8.17  0 0.00    1
```

```r
str(glass)
```

```
## 'data.frame':    213 obs. of  11 variables:
##  $ id  : int  2 3 4 5 6 7 8 9 10 11 ...
##  $ ri  : num  1.52 1.52 1.52 1.52 1.52 ...
##  $ Na  : num  13.9 13.5 13.2 13.3 12.8 ...
##  $ Mg  : num  3.6 3.55 3.69 3.62 3.61 3.6 3.61 3.58 3.6 3.46 ...
##  $ Al  : num  1.36 1.54 1.29 1.24 1.62 1.14 1.05 1.37 1.36 1.56 ...
##  $ Si  : num  72.7 73 72.6 73.1 73 ...
##  $ K   : num  0.48 0.39 0.57 0.55 0.64 0.58 0.57 0.56 0.57 0.67 ...
##  $ Ca  : num  7.83 7.78 8.22 8.07 8.07 8.17 8.24 8.3 8.4 8.09 ...
##  $ Ba  : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ Fe  : num  0 0 0 0 0.26 0 0 0 0.11 0.24 ...
##  $ Type: int  1 1 1 1 1 1 1 1 1 1 ...
```

```r
##remove id and Type:
glass.remove<-glass[,-c(1,11)]
```

a.

```r
cov(glass.remove)
```

```
##                ri            Na            Mg            Al           Si
## ri  9.232899e-06 -0.0004810133 -0.0005607631 -0.0006161839 -0.001270109
## Na -4.810133e-04  0.6697314377 -0.3259293449  0.0646066171 -0.043403933
## Mg -5.607631e-04 -0.3259293449  2.0749060723 -0.3456529166 -0.178797548
## Al -6.161839e-04  0.0646066171 -0.3456529166  0.2498822128 -0.003569729
## Si -1.270109e-03 -0.0434039330 -0.1787975485 -0.0035697294  0.599156249
## K  -5.712780e-04 -0.1419005448  0.0088393901  0.1059247165 -0.099926158
## Ca  3.521613e-03 -0.3213883271 -0.9134416755 -0.1856581540 -0.232024480
## Ba  1.608227e-06  0.1334290061 -0.3532144831  0.1192855789 -0.040248487
## Fe  4.323181e-05 -0.0192315041  0.0122165692 -0.0037297878 -0.007378238
##                K           Ca            Ba            Fe
## ri -0.000571278  0.003521613  1.608227e-06  4.323181e-05
## Na -0.141900545 -0.321388327  1.334290e-01 -1.923150e-02
## Mg  0.008839390 -0.913441676 -3.532145e-01  1.221657e-02
## Al  0.105924717 -0.185658154  1.192856e-01 -3.729788e-03
## Si -0.099926158 -0.232024480 -4.024849e-02 -7.378238e-03
## K   0.426455333 -0.296825908 -1.424804e-02 -6.109310e-04
## Ca -0.296825908  2.034716467 -8.039678e-02  1.735520e-02
## Ba -0.014248042 -0.080396782  2.482479e-01 -2.904228e-03
## Fe -0.000610931  0.017355198 -2.904228e-03  9.523682e-03
```

b.

```r
pca.cov.glass<-prcomp(glass.remove,scale=FALSE)
pca.cov.glass
```

```
## Standard deviations (1, .., p=9):
## [1] 1.7339331189 1.2888890549 0.8230159016 0.8037778804 0.4575380503
## [6] 0.3187644761 0.0949679118 0.0385199521 0.0009851739
##
## Rotation (n x k) = (9 x 9):
##              PC1          PC2          PC3           PC4           PC5
## ri  0.0009426369 -0.001507361  0.001383446 -0.0002752632  0.0007133913
## Na  0.0170585332  0.402691960  0.644059231  0.3631166029 -0.3974921649
## Mg -0.7200216947 -0.546147610  0.133769823  0.1029265396  0.0775884298
## Al  0.0448613746  0.258926607 -0.051299444 -0.2727088141  0.3128099378
## Si -0.0107005971  0.190491530 -0.707296148  0.5518892468 -0.1043215227
## K  -0.0801861468  0.101256248 -0.200349383 -0.6829911414 -0.5084138505
## Ca  0.6833747419 -0.612479194  0.078809480  0.0562028676 -0.0655418724
## Ba  0.0756193904  0.224846621  0.134167084 -0.0940209804  0.6809833206
## Fe  0.0008264559 -0.017032204 -0.006351187 -0.0111776607  0.0269778571
##             PC6           PC7         PC8           PC9
## ri  0.001823733  0.0003009644 0.004157102  9.999868e-01
## Na -0.015763146  0.0380574751 0.362192473 -1.404974e-03
## Mg -0.048189195  0.0754757143 0.375181470 -1.851128e-03
## Al -0.780393765  0.0751709644 0.376068581 -4.199496e-05
## Si  0.060831016  0.0585208547 0.375154335 -1.860404e-04
## K   0.265698672  0.0601905917 0.379467106 -1.400096e-03
## Ca -0.028956560  0.0740543163 0.371309118 -3.127292e-03
## Ba  0.559710066  0.0791000907 0.365520895 -2.993772e-03
## Fe -0.001237927 -0.9840812044 0.174351862 -4.663624e-04
```

```r
summary(pca.cov.glass)
```

```
## Importance of components:
##                           PC1    PC2    PC3    PC4     PC5    PC6     PC7
## Standard deviation     1.7339 1.2889 0.8230 0.8038 0.45754 0.3188 0.09497
## Proportion of Variance 0.4763 0.2632 0.1073 0.1023 0.03316 0.0161 0.00143
## Cumulative Proportion  0.4763 0.7394 0.8467 0.9491 0.98224 0.9983 0.99976
##                            PC8       PC9
## Standard deviation     0.03852 0.0009852
## Proportion of Variance 0.00024 0.0000000
## Cumulative Proportion  1.00000 1.0000000
```

The first eigenvector is related to the first eigenvalue 1.7339^2

c.

```r
cor(glass.remove)
```

```
##              ri          Na           Mg           Al           Si
## ri  1.000000000 -0.19343619 -0.128118295 -0.405670651 -0.540009928
## Na -0.193436186  1.00000000 -0.276486480  0.157927946 -0.068518634
## Mg -0.128118295 -0.27648648  1.000000000 -0.480035475 -0.160358613
## Al -0.405670651  0.15792795 -0.480035475  1.000000000 -0.009225663
## Si -0.540009928 -0.06851863 -0.160358613 -0.009225663  1.000000000
## K  -0.287899989 -0.26551982  0.009396937  0.324483684 -0.197684386
## Ca  0.812494939 -0.27531369 -0.444559250 -0.260372076 -0.210141492
## Ba  0.001062271  0.32723299 -0.492148823  0.478935953 -0.104360629
```

```
## Fe  0.145791387 -0.24080220  0.086905565 -0.076456429 -0.097674263
##               K         Ca           Ba           Fe
## ri -0.287899989  0.8124949  0.001062271  0.145791387
## Na -0.265519820 -0.2753137  0.327232988 -0.240802205
## Mg  0.009396937 -0.4445593 -0.492148823  0.086905565
## Al  0.324483684 -0.2603721  0.478935953 -0.076456429
## Si -0.197684386 -0.2101415 -0.104360629 -0.097674263
## K   1.000000000 -0.3186494 -0.043790065 -0.009586342
## Ca -0.318649382  1.0000000 -0.113121169  0.124673790
## Ba -0.043790065 -0.1131212  1.000000000 -0.059729016
## Fe -0.009586342  0.1246738 -0.059729016  1.000000000
```

```
##The correlation of Na and Ca and the correlation of Si and ri are large.
```

d.

```r
pca.cor.glass<-prcomp(glass.remove,scale=TRUE)
pca.cor.glass
```

```
## Standard deviations (1, .., p=9):
## [1] 1.5844012 1.4338398 1.1866007 1.0715515 0.9566406 0.7265090 0.6075153
## [8] 0.2527219 0.0401607
##
## Rotation (n x k) = (9 x 9):
##           PC1         PC2          PC3         PC4         PC5         PC6
## ri  0.5462809 -0.28309516  0.088556294  0.14508159 -0.07485001  0.11630236
## Na -0.2590903 -0.27319107 -0.375130240  0.50006762  0.14862038 -0.55600354
## Mg  0.1050559  0.59395205  0.015332525  0.38038101  0.12007329  0.30905278
## Al -0.4262025 -0.29739624  0.328138258 -0.14925214  0.01761676 -0.02143635
## Si -0.2275631  0.15608309 -0.470977869 -0.64442150  0.01384647  0.08425032
## K  -0.2182629  0.15297373  0.661936363 -0.05234636 -0.30679244 -0.24211783
## Ca  0.4954747 -0.34086413 -0.007150813 -0.27550770 -0.18589436 -0.14867282
## Ba -0.2476471 -0.48561315  0.076673447  0.13026778  0.25154794  0.65851237
## Fe  0.1884573  0.06502946  0.278595459 -0.22588580  0.87489148 -0.24474351
##            PC7         PC8         PC9
## ri  0.08170153 -0.75237306  0.02596627
## Na  0.14889157 -0.12839556 -0.31235634
## Mg -0.20828297 -0.07872130 -0.57597689
## Al -0.69750889 -0.27515725 -0.19223927
## Si  0.21797966 -0.37861427 -0.29765779
## K   0.50538672 -0.10923977 -0.26070815
## Ca -0.09929384  0.39874009 -0.58038575
## Ba  0.35020314  0.14499997 -0.19854510
## Fe  0.07631077 -0.01549403 -0.01474062
```

```r
summary(pca.cor.glass)
```

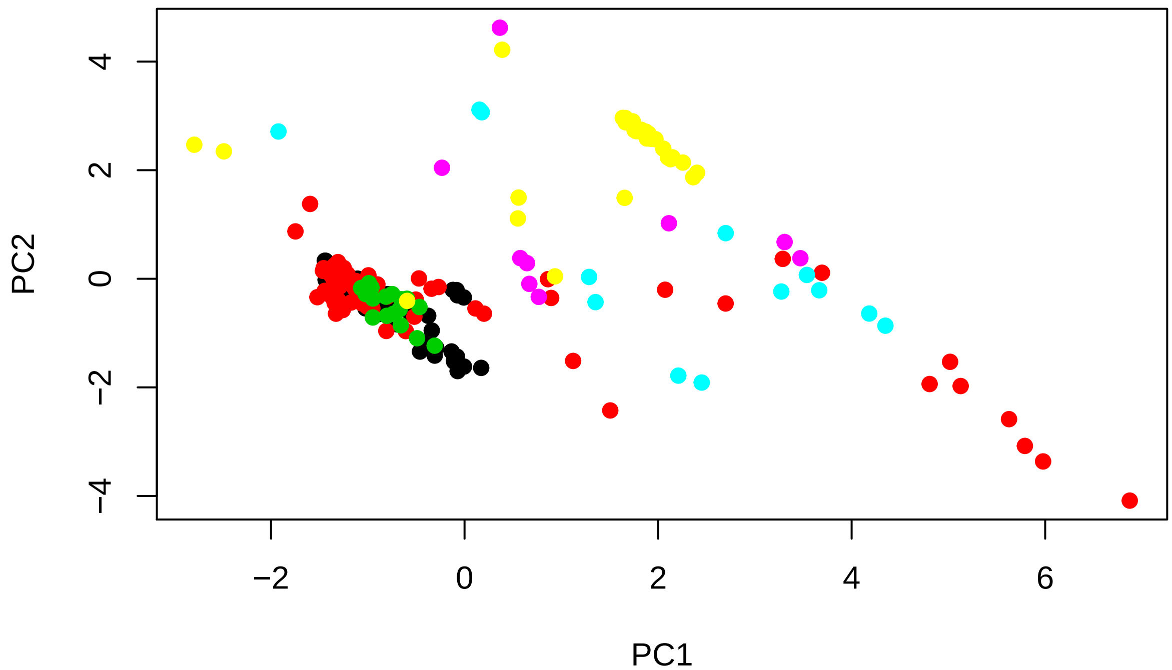
```
## Importance of components:
##                           PC1    PC2    PC3    PC4    PC5     PC6     PC7
## Standard deviation     1.5844 1.4338 1.1866 1.0716 0.9566 0.72651 0.60752
## Proportion of Variance 0.2789 0.2284 0.1565 0.1276 0.1017 0.05865 0.04101
## Cumulative Proportion  0.2789 0.5074 0.6638 0.7914 0.8931 0.95172 0.99272
##                           PC8     PC9
## Standard deviation     0.2527 0.04016
## Proportion of Variance 0.0071 0.00018
## Cumulative Proportion  0.9998 1.00000
```

```r
pca.cor.glass$rotation[,1]
```

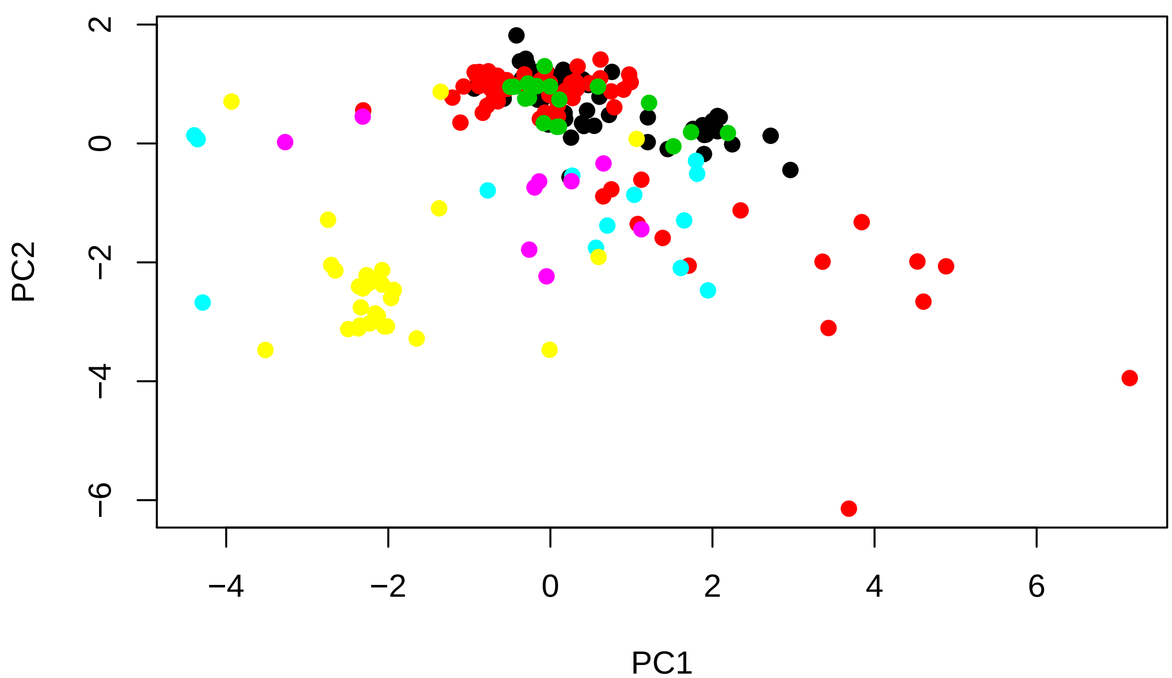
```
##        ri        Na        Mg        Al        Si         K 
##  0.5462809 -0.2590903  0.1050559 -0.4262025 -0.2275631 -0.2182629 
##        Ca        Ba        Fe 
##  0.4954747 -0.2476471  0.1884573
```

e.

```r
plot(PC2~PC1,pch=19,col=glass$Type,data=pca.cov.glass$x)
```



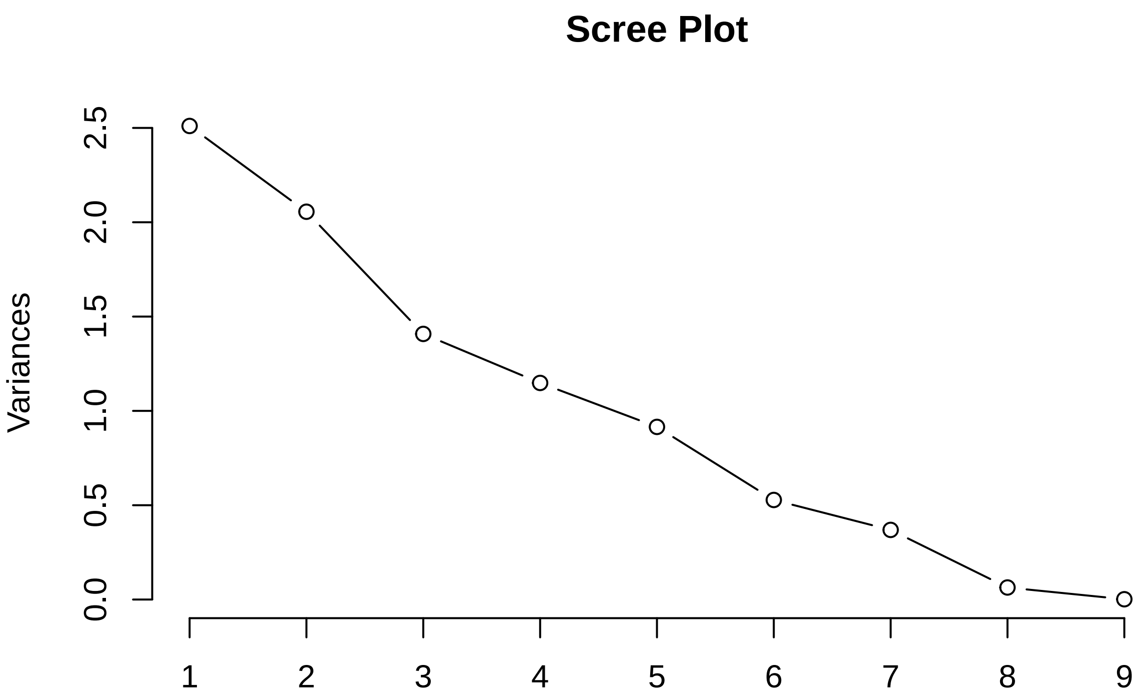

```r
plot(PC2~PC1,pch=19,col=glass$Type,data=pca.cor.glass$x)
```



```
##Correlation matrix should principal components be performed on.
```

f.

```
screeplot(pca.cor.glass,type = "lines",main = "Scree Plot")
```




```
##5 components should be retained.
```

∗ ∗ ∗Turtle Problem

```
turtles<-read.csv("turtles.csv",header = TRUE)
attach(turtles)
head(turtles)
```

```
##   sex length width height
## 1   1     98    81     38
## 2   1    103    84     38
## 3   1    103    86     42
## 4   1    105    86     42
## 5   1    109    88     44
## 6   1    123    92     50
```

```
turtles.no.sex<-turtles[,-1]
```

a.

```
cor(turtles.no.sex)
```

```
##           length     width    height
## length 1.0000000 0.9778869 0.9628899
## width  0.9778869 1.0000000 0.9599055
## height 0.9628899 0.9599055 1.0000000
```

```
##My findings: the correlations of theses variables are so closed
```

b.

```r
pca.cor.turtles<-prcomp(turtles.no.sex,scale=TRUE)
summary(pca.cor.turtles)
```

```
## Importance of components:
##                           PC1     PC2     PC3
## Standard deviation     1.7128 0.21027 0.14825
## Proportion of Variance 0.9779 0.01474 0.00733
## Cumulative Proportion  0.9779 0.99267 1.00000
```

```r
pca.cor.turtles
```

```
## Standard deviations (1, .., p=3):
## [1] 1.712837 0.210268 0.148249
##
## Rotation (n x k) = (3 x 3):
##              PC1        PC2         PC3
## length 0.5787409 -0.3504098 -0.73639113
## width  0.5781514 -0.4605453  0.67352724
## height 0.5751520  0.8155434  0.06394651
```

```r
PCA1=pca.cor.turtles$x[,1]
PCA2=pca.cor.turtles$x[,2]
prod(PCA1,PCA2)
```

```
## [1] -3.379165e-50
```

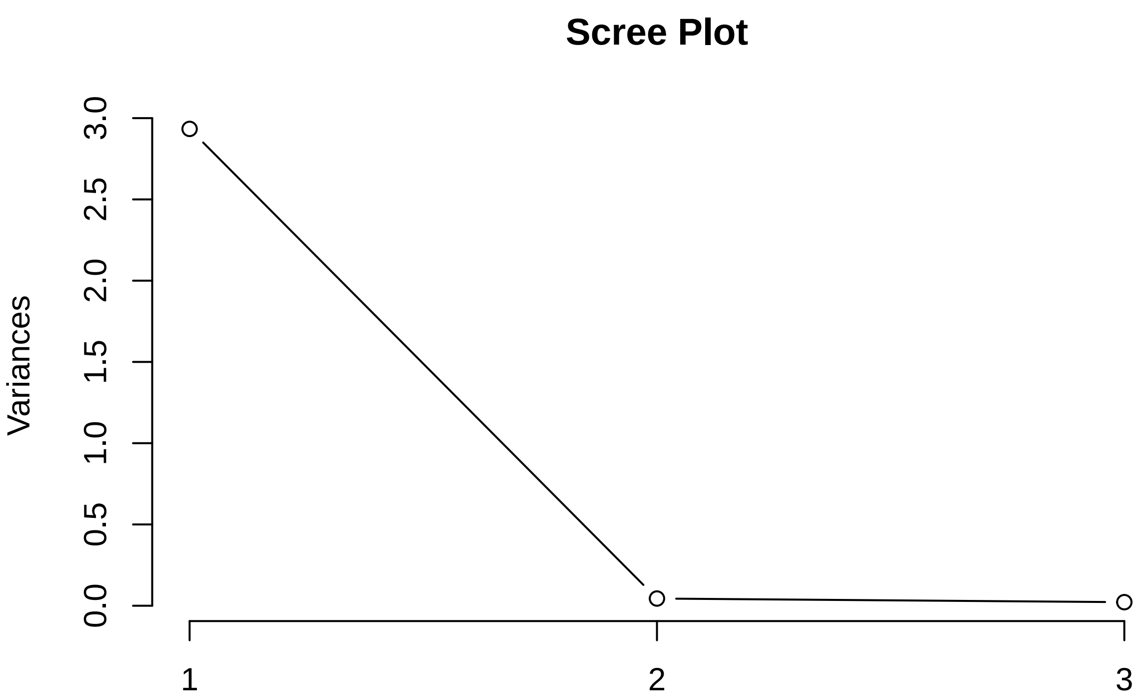
```r
##The value is amolst 0.
```

c.

```r
##PC1=0.5787409length+0.5781514*width+0.5751520*height
```

d.

```r
screeplot(pca.cor.turtles,type = "lines",main = "Scree Plot")
```



Two components should be retained.

e.

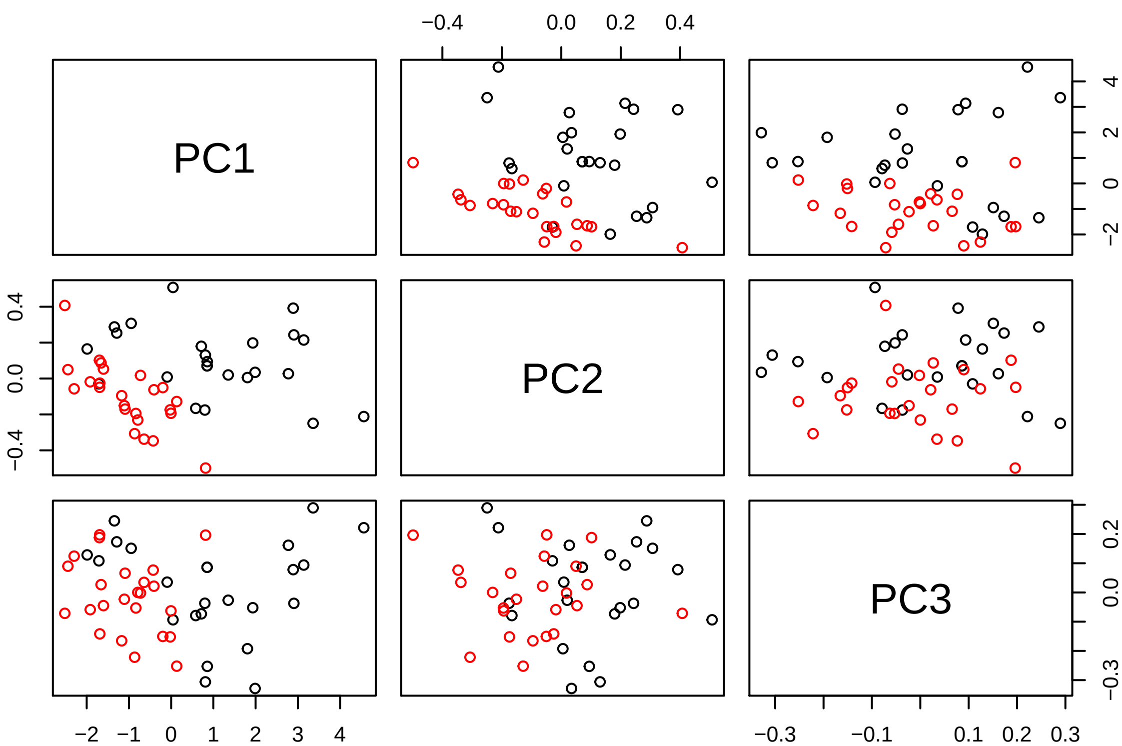
```
pca.cor.turtles
```

```
## Standard deviations (1, .., p=3):
## [1] 1.712837 0.210268 0.148249
##
## Rotation (n x k) = (3 x 3):
##              PC1        PC2         PC3
## length 0.5787409 -0.3504098 -0.73639113
## width  0.5781514 -0.4605453  0.67352724
## height 0.5751520  0.8155434  0.06394651
```

The first principle component appears to wiegh the importance of length and width. The second principle component appears to wiegh the importance of height.

f.

```
pairs(~PC1+PC2+PC3,data=pca.cor.turtles$x,col=turtles$sex)
```



The plot PC1 vs.PC2 seperates the turtles sex the most.

Women Track Problem

```
women<-read.csv("Womens Track Records.csv", header=TRUE)
head(women)
```

```
##    m100  m200  m400 m800 m1500 m3000 marathon  country
## 1 11.61 22.94 54.50 2.15  4.43  9.79   178.52 argentin
## 2 11.20 22.35 51.08 1.98  4.13  9.08   152.37 australi
## 3 11.43 23.09 50.62 1.99  4.22  9.34   159.37  austria
## 4 11.41 23.04 52.00 2.00  4.14  8.88   157.85  belgium
## 5 11.46 23.05 53.30 2.16  4.58  9.81   169.98  bermuda
```

```
## 6 11.31 23.17 52.80 2.10  4.49  9.77   168.75   brazil
```

a.

```r
##remove country
women.no.coun<-women[,-8]
##correlation matrix
cor(women.no.coun)
```

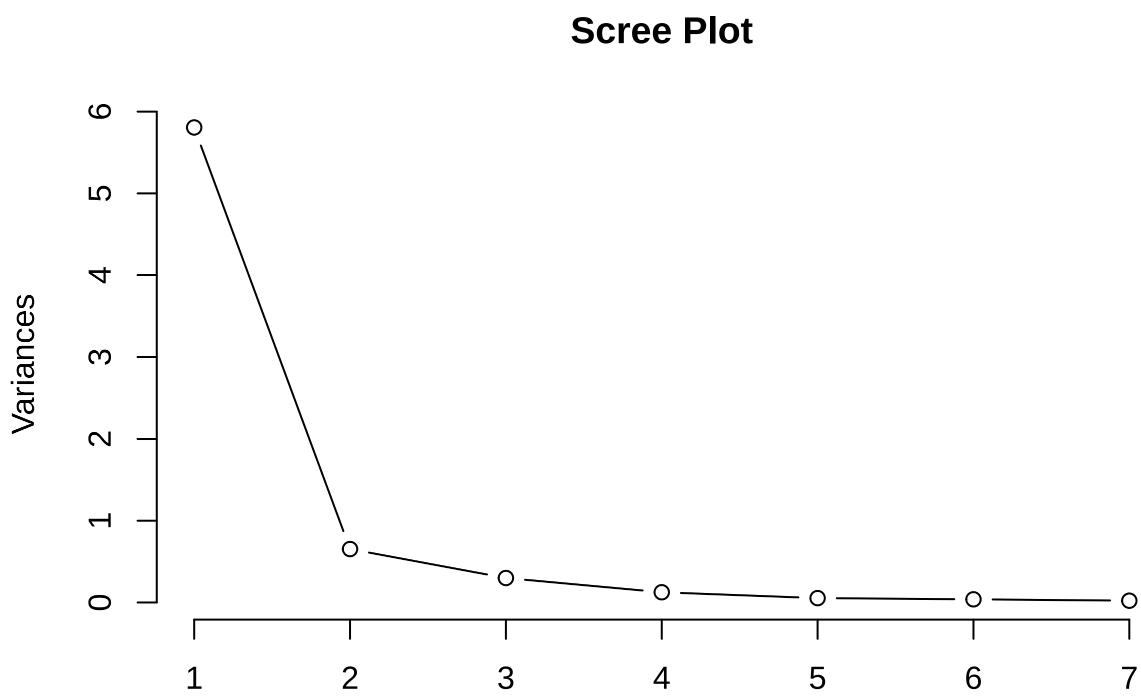
```
##               m100      m200      m400      m800     m1500     m3000
## m100     1.0000000 0.9527911 0.8346918 0.7276888 0.7283709 0.7416988
## m200     0.9527911 1.0000000 0.8569621 0.7240597 0.6983643 0.7098710
## m400     0.8346918 0.8569621 1.0000000 0.8984052 0.7878417 0.7776369
## m800     0.7276888 0.7240597 0.8984052 1.0000000 0.9016138 0.8635652
## m1500    0.7283709 0.6983643 0.7878417 0.9016138 1.0000000 0.9691690
## m3000    0.7416988 0.7098710 0.7776369 0.8635652 0.9691690 1.0000000
## marathon 0.6863358 0.6855745 0.7054241 0.7792922 0.8779334 0.8998374
##           marathon
## m100     0.6863358
## m200     0.6855745
## m400     0.7054241
## m800     0.7792922
## m1500    0.8779334
## m3000    0.8998374
## marathon 1.0000000
```

b.

```r
women.no.coun.pca.cor<-prcomp(women.no.coun,scale=TRUE)
print(summary(women.no.coun.pca.cor))
```

```
## Importance of components:
##                           PC1     PC2     PC3     PC4     PC5     PC6
## Standard deviation     2.4095 0.80848 0.54762 0.35423 0.23198 0.19761
## Proportion of Variance 0.8294 0.09338 0.04284 0.01793 0.00769 0.00558
## Cumulative Proportion  0.8294 0.92276 0.96560 0.98353 0.99122 0.99679
##                            PC7
## Standard deviation     0.14981
## Proportion of Variance 0.00321
## Cumulative Proportion  1.00000
```

```r
##scree plot
screeplot(women.no.coun.pca.cor,type = "lines",main="Scree Plot")
```



2 principle components should be retained.

c.

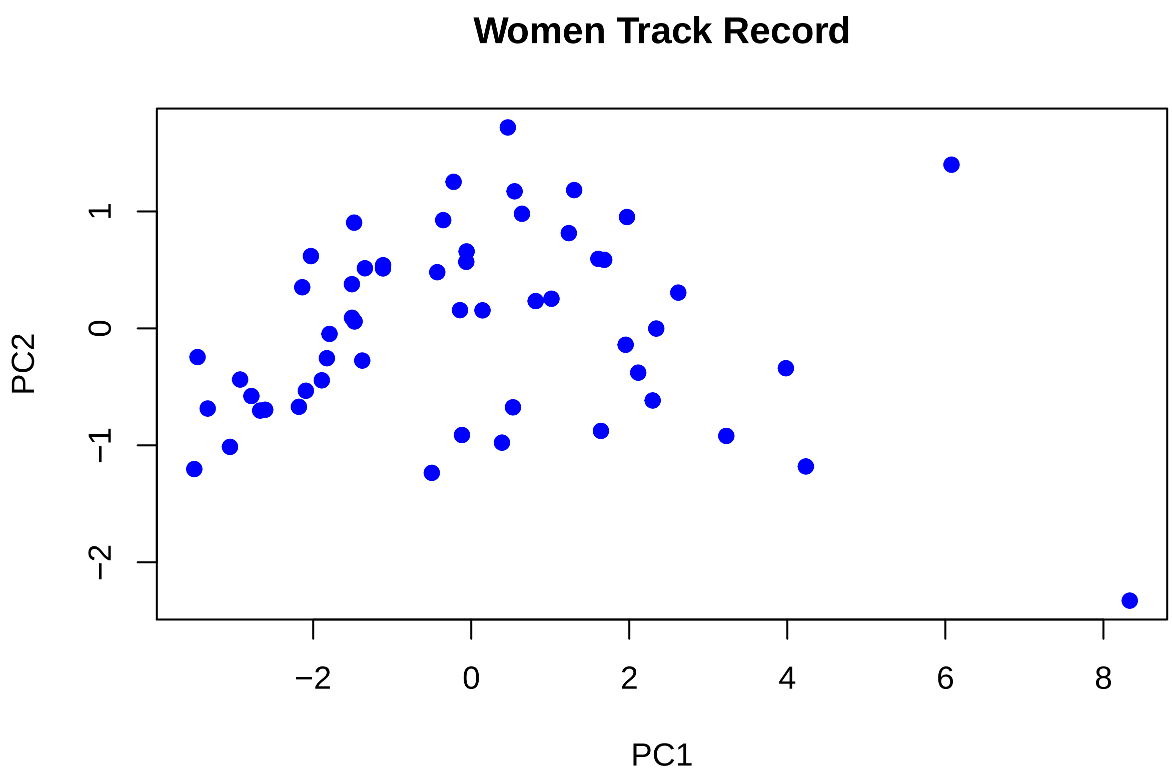
The first principal component descirbes the main items to be what it takes to win. It weights the importance of m400,m800,m1500 and m3000. The second principal component is weighs the importance of m100,m200 and marathon.

```r
women.no.coun.pca.cor
```

```
## Standard deviations (1, .., p=7):
## [1] 2.4094991 0.8084835 0.5476152 0.3542280 0.2319847 0.1976089 0.1498085
##
## Rotation (n x k) = (7 x 7):
##                PC1        PC2         PC3         PC4         PC5
## m100     0.3683561  0.4900597 -0.28601157  0.31938631  0.23116950
## m200     0.3653642  0.5365800 -0.22981913 -0.08330196  0.04145457
## m400     0.3816103  0.2465377  0.51536655 -0.34737748 -0.57217791
## m800     0.3845592 -0.1554023  0.58452608 -0.04207636  0.62032379
## m1500    0.3891040 -0.3604093  0.01291198  0.42953873  0.03026144
## m3000    0.3888661 -0.3475394 -0.15272772  0.36311995 -0.46335476
## marathon 0.3670038 -0.3692076 -0.48437037 -0.67249685  0.13053590
##                   PC6         PC7
## m100      0.619825234  0.05217655
## m200     -0.710764580 -0.10922503
## m400      0.190945970  0.20849691
## m800     -0.019089032 -0.31520972
## m1500    -0.231248381  0.69256151
## m3000     0.009277159 -0.59835943
## marathon  0.142280558  0.06959828
```

d.

```r
plot(PC2~PC1,pch=19,col="blue",main="Women Track Record",data=women.no.coun.pca.cor$x)
```




```
##abline(h=0,lwd=3)
##points(-3.50601681, -1.202500275,pch=4,col="blue",lwd=3)
##points(-3.46468721,-0.245078447,pch=4,col="red",lwd=3)
##points(-3.33581190 ,-0.685104574,pch=4,col="black",lwd=3)
```