Title: Who is FIT?

Data Summary

The data, *fitnessR*, contains 20 observations and 6 variables. The variables are physical measurements of weight, waste size and pulse rate correspond to exercise related measurements of number of chin-ups, sit-ups, and jumps performed in a timed period.

```
> summary(fitnessR)
     WEIGHT           WAISST          PULSE          CHINUPS           SITUPS
 Min.   :138.0   Min.   :31.0   Min.   :46.0   Min.   : 1.00   Min.   : 50.0
 1st Qu.:160.8   1st Qu.:33.0   1st Qu.:51.5   1st Qu.: 4.75   1st Qu.:101.0
 Median :176.0   Median :35.0   Median :55.0   Median :11.50   Median :122.5
 Mean   :178.6   Mean   :35.4   Mean   :56.1   Mean   : 9.45   Mean   :145.6
 3rd Qu.:191.5   3rd Qu.:37.0   3rd Qu.:60.5   3rd Qu.:13.25   3rd Qu.:210.0
 Max.   :247.0   Max.   :46.0   Max.   :74.0   Max.   :17.00   Max.   :251.0
     JUMPS
 Min.   : 25.00
 1st Qu.: 39.50
 Median : 54.00
 Mean   : 70.30
 3rd Qu.: 85.25
 Max.   :250.00
```

Statement of Problem

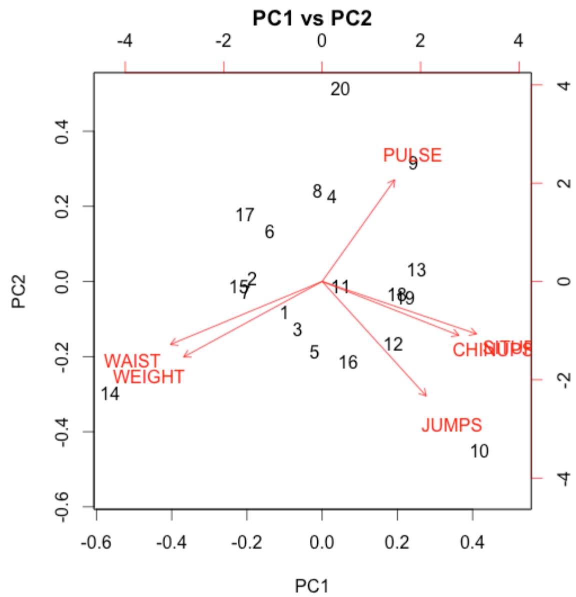
For this dataset, we want to find the relationship between variables using principal component and we want examine the relationships between physical variables and exercise variables and canonical correlation.

Principal Component

In this analysis, we would find the relationship in the data.

```
> pca = prcomp(fitnessR, scale. = T)
> summary(pca)
Importance of components:
                          PC1    PC2    PC3     PC4     PC5     PC6
Standard deviation     1.8026 1.1201 0.8557 0.68183 0.48475 0.25254
Proportion of Variance 0.5416 0.2091 0.1220 0.07748 0.03916 0.01063
Cumulative Proportion  0.5416 0.7507 0.8727 0.95021 0.98937 1.00000

> round(pca$rotation[,1:3],2)
          PC1   PC2   PC3
WEIGHT  -0.44 -0.38  0.32
WAIST   -0.48 -0.32  0.35
PULSE    0.23  0.52  0.82
CHINUPS  0.43 -0.27 -0.04
SITUPS   0.49 -0.27  0.07
JUMPS    0.33 -0.58  0.31
```



From the cumulative proportion of variance, we can see that the first three components accord for about 87% of the data. Therefore, we use first three components to exam the relationship between variables.

The first component explains about 54% of the variation in the data. It is mainly compose of variable waist, weight, chin-ups, and sit-ups. The weight and waist is negative correlated with chin-ups and sit-ups.

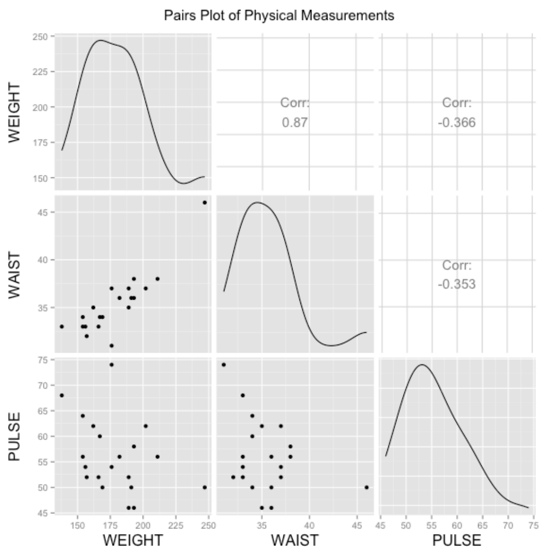
The second component explains 21% of the variation in the data. It is mainly measure of pulse and jump. The pulse is negative correlated with jump.

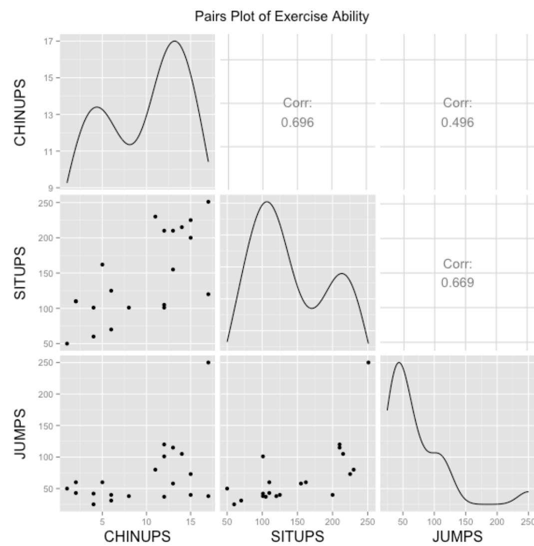
The third component explains 12% of the variation in the data. It is mainly measurement of pulse.

Canonical Correlation Analysis

In this analysis, we want to exam the significant of physical variables in response to exercise variables. Therefore, we would set the exercise variable as response variables and physical variables as predictors. First we check the correlation of variables between each group and within the group.

```
> body = fitnessR[1:3]
> exc = fitnessR[4:6]
> ggpairs(body, title = "Pairs Plot of Physical Measurements")
> ggpairs(exc, titl = "Pairs Plot of Exercise Ability")
```





From pair plots, we can see that there is a positive correlation between weight & waist, chin-ups& sit-ups, and sit-ups and jumps.

```
> round(cor(body, exc),2)
       CHINUPS SITUPS JUMPS
WEIGHT   -0.39  -0.49 -0.23
WAIST    -0.55  -0.65 -0.19
PULSE     0.15   0.23  0.03
```

From the correlation matrix, we can see there is negatively correlation between waist with chin-ups and sit-ups.

Frist, we assign physical measurements into variable body and exercise measurements into variable exc. Then we perform canonical correlation analysis.

```
> body = fitnessR[1:3]
> exc = fitnessR[4:6]
> cc1 = cc(body, exc)

Display the canonical correlations
> cc1$cor
[1] 0.79560815 0.20055604 0.07257029

Tests of canonical dimensions
> (dmat <- cbind(WilksL = w, F = f, df1 = d1, df2 = d2, p = pv))
        WilksL          F df1      df2          p
[1,] 0.3503905 2.04823353   9 34.22293 0.06353094
[2,] 0.9547227 0.17578229   4 30.00000 0.94912025
[3,] 0.9947336 0.08470926   1 16.00000 0.77475327
```

```
Compute canonical loadings
> cc2 = comput(body, exc, cc1)
> cc2[3:6]

$corr.X.xscores
              [,1]       [,2]        [,3]
WEIGHT -0.6206424  0.7723919 -0.13495886
WAIST  -0.9254249  0.3776614 -0.03099486
PULSE   0.3328481 -0.0414842  0.94206752

$corr.Y.xscores
              [,1]       [,2]        [,3]
CHINUPS 0.5789047 -0.0475222 -0.04671717
SITUPS  0.6505914 -0.1149232  0.00395139
JUMPS   0.1290401 -0.1922586 -0.01697689

$corr.X.yscores
              [,1]         [,2]         [,3]
WEIGHT -0.4937881  0.154907853 -0.009794003
WAIST  -0.7362756  0.075742277 -0.002249306
PULSE   0.2648166 -0.008319907  0.068366110

$corr.Y.yscores
              [,1]       [,2]        [,3]
CHINUPS 0.7276254 -0.2369522 -0.64375064
SITUPS  0.8177285 -0.5730231  0.05444915
JUMPS   0.1621905 -0.9586280 -0.23393722

Standardized body canonical coefficients diagonal matrix of body sd's
> s1 <- diag(sqrt(diag(cov(body))))
> s1 %*% cc1$xcoef
             [,1]       [,2]       [,3]
[1,]  0.77539761  1.8843672 -0.1909822
[2,] -1.57934657 -1.1806411  0.5060195
[3,]  0.05912012  0.2311068  1.0507838

Standardized exc canonical coefficients diagonals matrox of exc sd's
> s2 <- diag(sqrt(diag(cov(exc))))
> s2 %*% cc1$ycoef
           [,1]       [,2]       [,3]
[1,]  0.3494969  0.3755436 -1.2965937
[2,]  1.0540110 -0.1234905  1.2367934
[3,] -0.7164267 -1.0621670 -0.4188073
```

The first canonical correlation is 0.7956, which greatest of all the correlations. The probability level for the null hypothesis that all the canonical correlations are zero in the population is only 0.0635, but we want it to be significant because the data is small. The remaining canonical correlations are not worthy of consideration, as can be seen from the probability levels are very high.

The first canonical correlation is mostly related to waist (with loading = -0.93) for the psychical measurements and chin-ups (loading = 0.73) and sit-ups (loading = 0.81) for the exercise measurements. Therefore, waist is negative correlated with chin-ups and sit-ups.

Conclusion

From first principal component, we find that waist is negatively correlated with chip-ups and shit-ups accorded for 54% variation of the data. And for canonical correlation analysis with .80 correlations, we also found that the waist is negatively correlated with chip-ups and shit-ups. This makes sense because a person with bigger waist tends to have more difficulty doing sit-ups. And person with bigger waist tend to be heavier, so less likely to do chip-ups.