Multivariate General Linear Models

Lecture 08.2: Multivariate GLMs

Lauren Sullivan

Module: Multivariate Models

Readings

Required for class:

- NA

Optional:

- Tabachnick and Fidell (2012) Using Multivariate Statistics (6th Edition).

Multivariate Analysis

There are several ways to look at multivariate patterns from a matrix of **Y**'s.

1. **Linear models: MANOVA/regression to test patterns**
2. Ordination: PCA, nMDS, etc to visulazie patterns
3. Permutation tests: PERMANOVA to test patterns

Multivariate GLM

All of our previous linear models (e.g.: ANOVA, regression, ANCOVA, etc) we can run as a general linear model (GLM) with multivariate data.

	1 Cat. X	>1 Cat. X	1 Cont. X	>1 Cont. X	Both
1 Cont. Y	ANOVA	Factorial ANOVA	Regression	Multiple Regression	ANCOVA
>1 Cont. Y	MANOVA	Factorial MANOVA	Multivariate Regression	Multivariate Multiple Regression	MANCOVA

Sparrow Data

Data from Bumpus (1898) - where he measured ~136 sparrows after a bad February storm. Half the birds were dead. Bumpus wanted to investigate natural selection and determine if there was a difference between dead or alive birds.

Sparrow Data

Data includes: sex, age, survived (TRUE/FALSE), total length (TL), wing extent (AE), mass (WT), beak-head length (BHL), humerus length (HL), femur length (FL), tibiotarsal length (TTL), skull weight (SW), sternum-keel length (SKL).

```
sparrow
```

```
## # A tibble: 136 x 12
##    sex   age   survived    TL    AE    WT   BHL    HL    FL   TTL    SW   SK
##    <chr> <chr> <lgl>    <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl
##  1 m     a     TRUE       154   241  24.5  31.2  17.4  17.0  26.0  14.9  21.
##  2 m     a     FALSE      165   240  26.5  31    18.7  17.9  27.8  15.4  21.
##  3 m     a     FALSE      160   245  26.1  32    18.7  18.0  28.2  15.5  21.
##  4 m     a     TRUE       160   252  26.9  30.8  18.7  18.0  30.0  15.3  21.
##  5 m     a     TRUE       155   243  26.9  30.6  18.6  17.9  29.2  15.3  21.
##  6 m     a     FALSE      161   249  25.6  32.3  18.9  18.2  28.7  15.3  21.
##  7 m     a     TRUE       154   245  24.3  31.7  18.8  17.5  29.1  14.8  21.
##  8 m     a     FALSE      162   246  25.9  32.3  18.7  18.0  28.8  15.4  22.
##  9 m     a     TRUE       156   247  24.1  31.5  18.2  17.9  28.7  14.6  20.
## 10 m     a     FALSE      163   250  25.5  32.5  19.1  18.6  30.4  15.8  22.
## # ... with 126 more rows
```

Scale Data

You will want to make sure your data are all on similar scales. You can see here that we have some values in mm and some in g, which are on different orders of magnitude and can make analyses tricky.

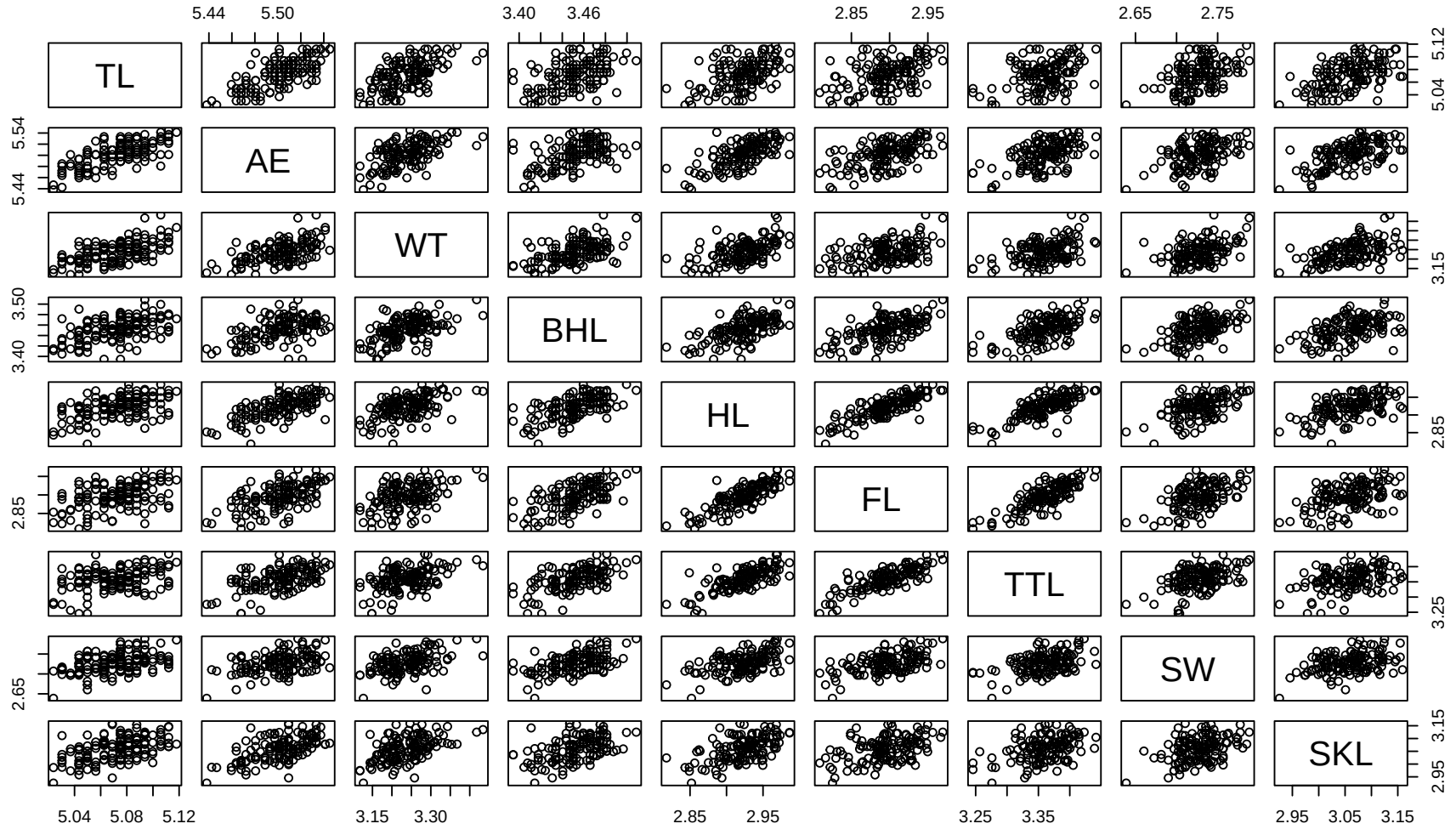
- A simple way to scale your data is to simply log everything.

```
sparrow_log <- log(sparrow[,4:12])
sparrow <- cbind(sparrow[,1:3], sparrow_log)
sparrow[1:9, 1:8]
```

```
##   sex age survived       TL       AE       WT      BHL       HL
## 1   m   a     TRUE 5.036953 5.484797 3.198673 3.440418 2.859328
## 2   m   a    FALSE 5.105945 5.480639 3.277145 3.433987 2.930938
## 3   m   a    FALSE 5.075174 5.501258 3.261935 3.465736 2.928224
## 4   m   a     TRUE 5.075174 5.529429 3.292126 3.427515 2.928224
## 5   m   a     TRUE 5.043425 5.493061 3.292126 3.421000 2.924140
## 6   m   a    FALSE 5.081404 5.517453 3.242592 3.475067 2.937690
## 7   m   a     TRUE 5.036953 5.501258 3.190476 3.456317 2.934995
## 8   m   a    FALSE 5.087596 5.505332 3.254243 3.475067 2.930938
## 9   m   a     TRUE 5.049856 5.509388 3.182212 3.449988 2.899276
```

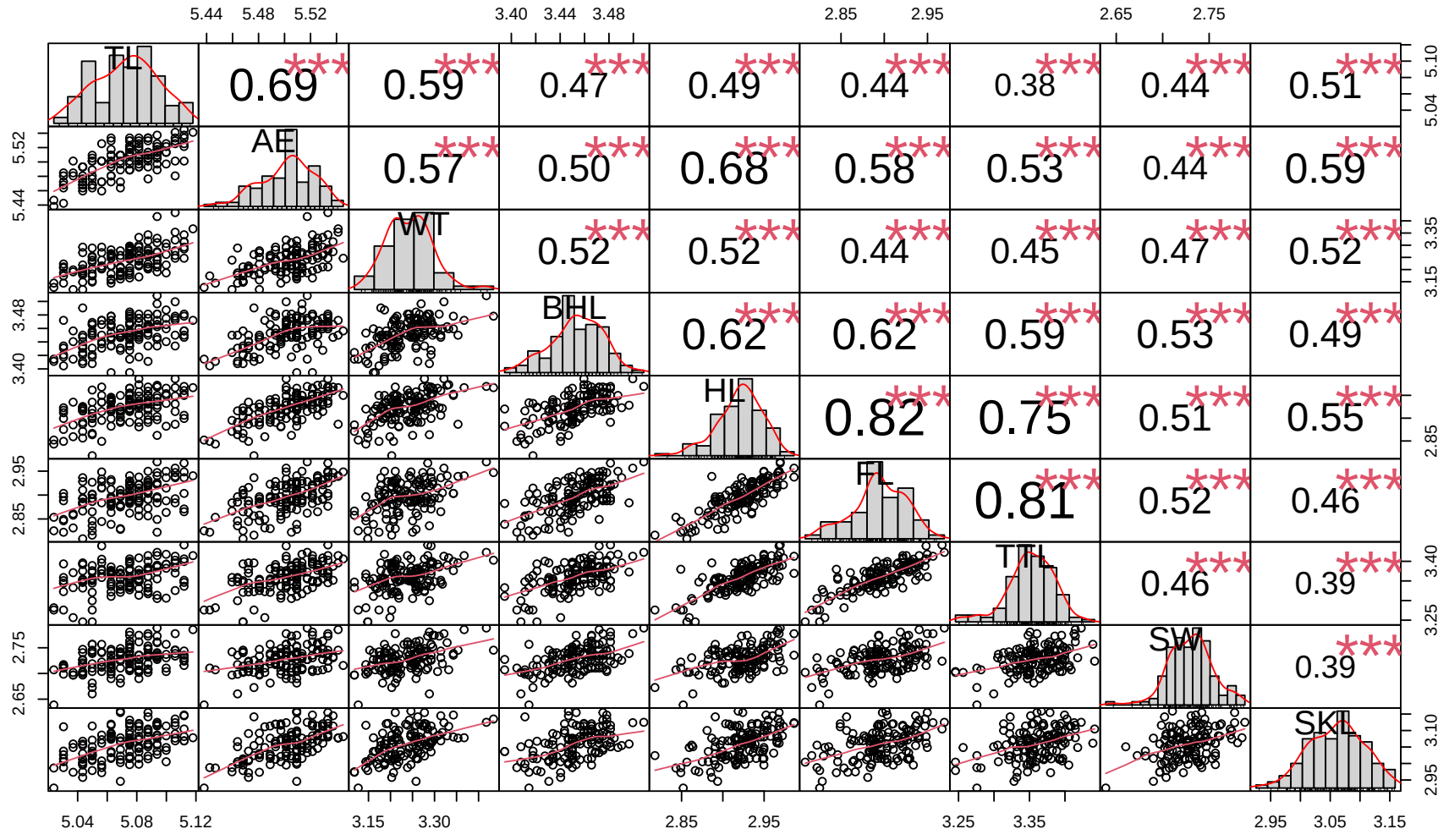
Plot the Correlations

```
pairs(sparrow[,4:12])
```



Plot the Correlations with Values

```
chart.Correlation(sparrow[,4:12], histogram = TRUE, pch = 19)
```



Create a Correlation Table

```
scorr <- round(cor(sparrow[,4:12]),3)

#Hides Upper Triangle (You hide lower triangle with lower.tri())
upper<-scorr
upper[upper.tri(scorr)]<-""
upper<-as.data.frame(upper)
upper
```

```
##        TL    AE    WT   BHL    HL    FL   TTL    SW SKL
## TL      1
## AE  0.693     1
## WT  0.587 0.574     1
## BHL 0.471 0.504 0.524     1
## HL  0.486 0.679  0.52 0.623     1
## FL  0.444 0.578 0.443 0.616 0.822     1
## TTL 0.378 0.534 0.455 0.585 0.749 0.811     1
## SW  0.438 0.436  0.47 0.535 0.512 0.523 0.461     1
## SKL 0.505 0.586 0.519 0.494 0.552 0.457 0.389 0.389   1
```

Check out this website for code.

MANOVA

Used when there is one (or more) categorical X variables (or groups), and more than one Y variable. MANOVA's compare variation within groups to varaition between groups.

MANOVA assumptions include:

1. Multivariate Normal distributions of error variance within groups. To test for this multivariate normality, you can use `m.shapiro.test()` in the `mvnormtest` package.
2. Equal variance (aka - Homoscedasticity)
3. Equal co-variance within groups.

MANOVA

How does sex of the birds alter their total length (TL), wing extent (AE), and femur length (FL)?

```
sparrow.man <- manova(cbind(TL, AE, FL) ~ sex, data = sparrow)
summary(sparrow.man, test.statistic = "F", type = 3)
```

```
##             Df  Pillai approx F num Df den Df    Pr(>F)
## sex          1 0.38852   27.956      3    132 4.631e-14 ***
## Residuals 134
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Note: `manova()` uses Pillai's trace as a test statistic (falls between 0-1), which is converted to an F statistic. A larger Pillai's value = a larger F statistic, which is more likely to reject H_0.

MANOVA

See which specific Y results differ.

```
summary.aov(sparrow.man)
```

```
##  Response TL :
##              Df   Sum Sq   Mean Sq F value    Pr(>F)
## sex           1 0.007460 0.0074597  16.667 7.617e-05 ***
## Residuals   134 0.059975 0.0004476
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##  Response AE :
##              Df   Sum Sq   Mean Sq F value    Pr(>F)
## sex           1 0.019244 0.0192444  51.708 4.089e-11 ***
## Residuals   134 0.049871 0.0003722
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##  Response FL :
##              Df   Sum Sq    Mean Sq F value Pr(>F)
## sex           1 0.000188 0.00018836  0.1617 0.6882
## Residuals   134 0.156092 0.00116487
```

Factorial MANOVA

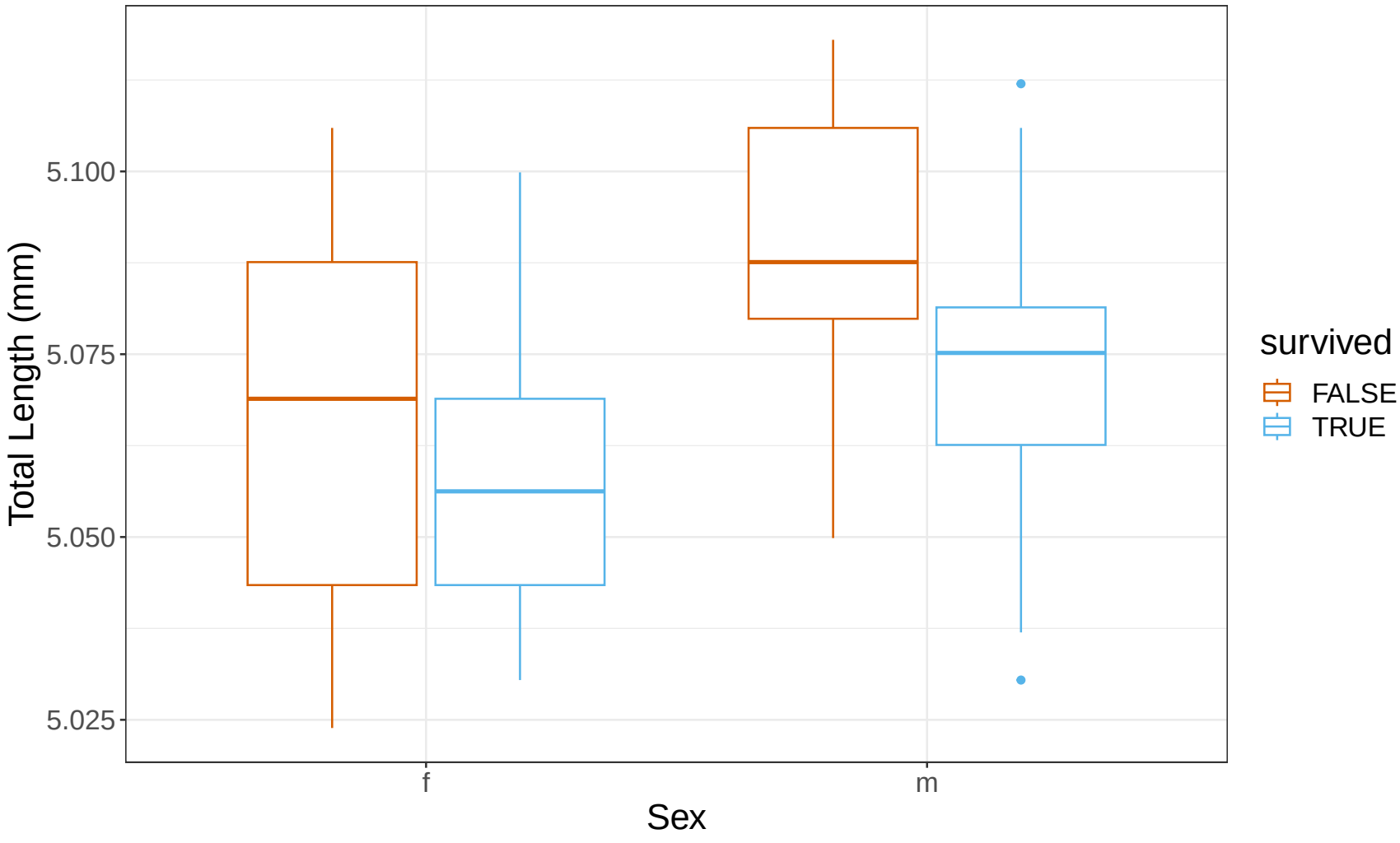
How does sex and survival of the birds alter their total length (TL), wing extent (AE), and femur length (FL)?

```
sparrow.Fman <- manova(cbind(TL, AE, FL) ~ sex * survived, data = sparrow)
summary(sparrow.Fman)
```

```
##                Df  Pillai approx F num Df den Df    Pr(>F)
## sex             1 0.39229  27.9723      3    130 5.054e-14 ***
## survived        1 0.21796  12.0770      3    130 5.035e-07 ***
## sex:survived    1 0.04598   2.0884      3    130    0.1049
## Residuals     132
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Plot the Data

This figure does not take into account co-variation!



Multivariate Regression

Used when there is one (or more) continuous X variables, and more than one Y variable.

Multivariate regression assumptions include:

1. Relationships are all linear
2. Multivariate Normal distributions of error variance
3. Equal variance (aka - Homoscedasticity)
4. Absence of multicollinearity (Tabachnick & Fidell (2012) suggest no correlation above $r = 0.9$).

Multivariate Regression - Allometry

Does the total length of a bird relate to its other size characteristics?

```
summary(manova(cbind(AE, WT, BHL, HL, FL, TTL, SW, SKL) ~ TL,
               data = sparrow))
```

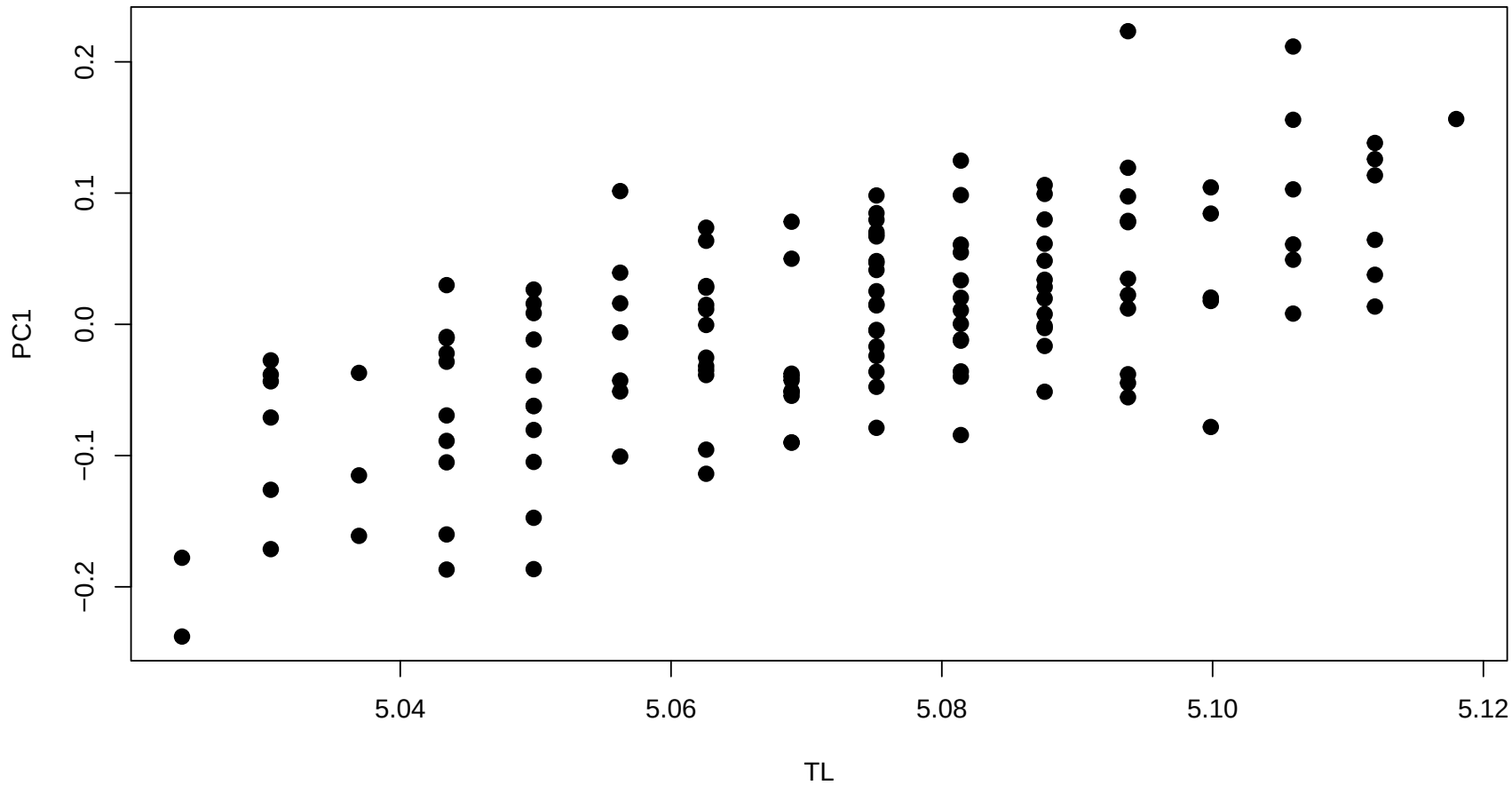
```
##             Df  Pillai approx F num Df den Df    Pr(>F)
## TL           1 0.55629   19.903      8    127 < 2.2e-16 ***
## Residuals 134
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

There is evidence for allometry - there is a significant relationship between total length and the rest of the body variables.

- But plotting this is difficult because of so many Y's!

Plotting the Data

Visualize multivariate regression with summary variables (here with regression PC1 - we will talk about this soon).



MANCOVA

Used when there is one (or more) continuous X variables, one (or more) categorical X variables, and more than one Y variable.

MANCOVA assumptions include:

1. Relationships are all linear
2. Multivariate Normal distributions of error variance within groups
3. Equal variance (aka - Homoscedasticity)
4. Absence of multicollinearity
5. No relationship between covariates (continuous X's) and groups (categorical X's)

MANCOVA

How does a bird's overall body size depend on it's sex * survival (groups), and its total length (covariate)?

```
summary(manova(cbind(AE, WT, BHL, HL, FL, TTL, SW, SKL) ~ TL + sex * survived,
               data = sparrow))
```

```
##                Df  Pillai approx F num Df den Df    Pr(>F)
## TL              1 0.62878  26.2545      8    124 < 2.2e-16 ***
## sex             1 0.40635  10.6098      8    124 2.859e-11 ***
## survived        1 0.26117   5.4791      8    124 6.315e-06 ***
## sex:survived    1 0.08125   1.3708      8    124    0.2157
## Residuals     131
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Total length is a significant predictor of overall bird body size, males and females tend to have different body sizes, and those that survived and those that didn't have different body sizes.

Plotting the Data

Visualizing this multivariate model is even more difficult! But try it with the TL ~ PC1 + groups.