Group A

You are doing a research project on bird skins that you found in the museum's collection. These birds are not catalogued so it is unclear what species they belong to. You believe that they are of the species *Parus lundyensis*. This species is clearly sexually dimorphic in plumage, so you think you can use that to identify the sex of the skins. You want to use statistics to support your idea that these bird skins are from this species.

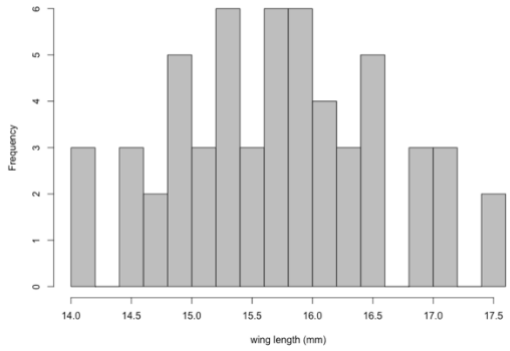
A litereature review brings up the following in the "Handbook of the mysterious birds of the world"

"*Parus lundyensis* shows a plumage dimorphism. Both sexes have a wing length of 15.5-17.0mm."

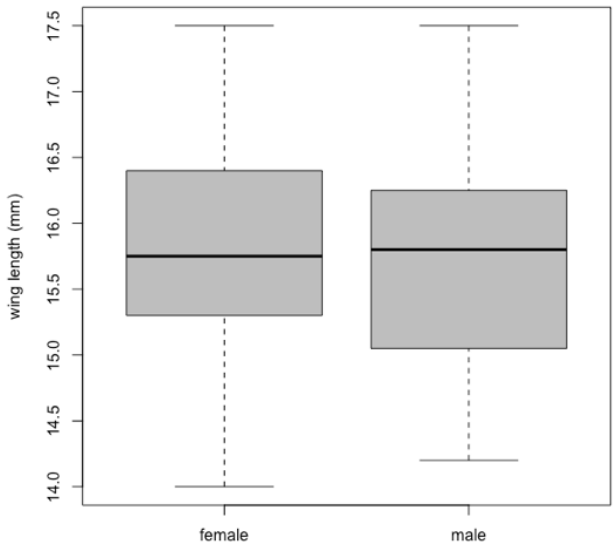
You conduct some exploratory data analysis, and then run the main test. Below is the R output from these analyses.

```
'data.frame': 57 obs. of  3 variables:
 $ Catalogue_Nr   : int  1 2 3 4 5 6 7 8 9 10 ...
 $ wing_length.mm.: num  16.5 15.2 16.1 16.6 14.8 16.6 15.9 15.8 15 15.9 ...
 $ sex            : Factor w/ 2 levels "female","male": 2 1 2 1 2 1 1 2 1 1 ...
```

```
Catalogue_Nr wing_length.mm.    sex
1            16.5   male
2            15.2 female
3            16.1   male
4            16.6 female
5            14.8   male
6            16.6 female
```




```
 Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  14.00   15.20   15.80   15.77   16.30   17.50
```




```
Welch Two Sample t-test

data:  data$wing_length.mm by data$sex
t = 0.074339, df = 55, p-value = 0.941
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -0.4422484  0.4763225
sample estimates:
mean in group female   mean in group male
             15.78000             15.76296


One Sample t-test

data:  data$wing_length.mm
t = -4.2154, df = 56, p-value = 9.175e-05
alternative hypothesis: true mean is not equal to 16.25
95 percent confidence interval:
 15.54474 15.99912
sample estimates:
mean of x
 15.77193
```

Group B

You have collected data on the elusive unicorns, roaming in a forest far away from any known land. You want to know what determines unicorn horn length, and you suspect that unicorns that have a lot of food, and are fat, are able to grow longer horns. Unicorn body mass is measured in g, and unicorn size is measured in cm.

The analysis was done with body mass (in kg) z-standardised, with a mean of 0 and standard deviation of 1. Hornlength (in meter) was not standardized as it was the response variable. You get this output:

```
> summary(d$Bodymass)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  67.39   81.68   96.86  103.92  120.39  173.05

> summary(d$Hornlength)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  3.365   4.660   5.885   5.709   6.332   8.030
> length(d$Bodymass)
[1] 20

> var(d$Hornlength)
[1] 1.510912

> summary(d$z.BodyMass)
Min.   :-1.3107
 1st Qu.:-0.7980
 Median :-0.2531
 Mean   : 0.0000
 3rd Qu.: 0.5912
 Max.   : 2.4807

> summary(lm(d$Hornlength~d$z.BodyMass))

Call:
lm(formula = d$Hornlength ~ d$z.BodyMass)

Residuals:
    Min      1Q  Median      3Q     Max
-1.5492 -0.4333 -0.1230  0.6734  1.3997

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)    5.7090     0.1757  32.493  < 2e-16 ***
d$z.BodyMass   0.9623     0.1803   5.338 4.49e-05 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.7858 on 18 degrees of freedom
Multiple R-squared:  0.6129,     Adjusted R-squared:  0.5914
F-statistic:  28.5 on 1 and 18 DF,  p-value: 4.492e-05
```

Group C

You are interested in the effects of climate change on bird's timing of breeding. You spend the last 4 years collecting data on the date the birds lay their first egg of the first clutch of an individual female in a given year. Your study species is the climate-change sensitive golden phoenix (*Phoenix potterus fawkes*), whose eggs burst into flames and smoke after if they didn't hatch by April 20. Eggs laid after April 20 are infertile. So, with ongoing climate change the hope is that more golden phoenixes may lay earlier when spring starts earlier every year, and that may aid the species' survival.

You collected data from individual birds attending nests, recording the egg laying data in days from 1st March. This way, 14 is March 14, and 36 is April 6, and so forth. You collected this data over the course of four years, between 2006 and 2009. You want to analyse whether laying date changed over the course of the years, in particular, whether it decreased. You use two main approaches for data analysis.

```
> length(PhoenixData$LayingDate)
[1] 108

> var(PhoenixData$LayingDate)
[1] 539.0041

> summary(PhoenixData$LayingDate)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
   6.00   31.75   43.00   48.62   65.25  114.00

> table(PhoenixData$year)
2006 2007 2008 2009
  46   33   10   19

> summary(lm(LayingDate~as.factor(year), data=PhoenixData))
Call:
lm(formula = LayingDate ~ as.factor(year), data = PhoenixData)

Residuals:
    Min      1Q  Median      3Q     Max
-29.804 -16.000  -2.452  12.397  61.697

Coefficients:
                  Estimate Std. Error t value Pr(>|t|)
(Intercept)         56.804      3.076  18.467  < 2e-16 ***
as.factor(year)2007   -4.501      4.759  -0.946  0.34643
as.factor(year)2008  -20.704      7.279  -2.844  0.00536 **
as.factor(year)2009  -27.804      5.689  -4.887 3.73e-06 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 20.86 on 104 degrees of freedom
Multiple R-squared:  0.2152,  Adjusted R-squared:  0.1925
F-statistic: 9.505 on 3 and 104 DF,  p-value: 1.325e-05

> summary(lm(LayingDate~year, data=PhoenixData))
Call:
lm(formula = LayingDate ~ year, data = PhoenixData)

Residuals:
    Min      1Q  Median      3Q     Max
-31.247 -15.118  -4.021  11.753  65.205

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 19017.846   3635.037   5.232 8.52e-07 ***
year           -9.451      1.811  -5.218 9.03e-07 ***
```

---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 20.81 on 106 degrees of freedom
Multiple R-squared:  0.2044,  Adjusted R-squared:  0.1969
F-statistic: 27.23 on 1 and 106 DF,  p-value: 9.026e-07

Group D

You are curious whether the gender of a marker affects a student's project mark. Therefore, you get hold of the EECs database and have a look at whether the gender of the first marker explains the mark students get for their final project. You then remember an article in the newspaper about marks inflation, and thought you'd have a look at that, too.

```
> length(a$Project_Mark)
[1] 653

> summary(a$Project_Mark)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  35.00   63.20   68.00   68.02   72.00   90.00

> var(a$Project_Mark)
[1] 55.07945

> summary(a$Assessor.Gender)
  male    female     NA's
 293          56     304

> summary(lm(a$Project_Mark~a$Assessor.Gender))

Call:
lm(formula = a$Project_Mark ~ a$ Assessor.Gender)

Residuals:
    Min      1Q  Median      3Q     Max
-33.976  -4.776   0.069   3.824  20.024

Coefficients:
                          Estimate Std. Error t value Pr(>|t|)
(Intercept)               68.9763     0.4096 168.390   <2e-16 ***
a$ Assessor.Gender female  0.4654     1.0226   0.455    0.649
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 7.012 on 347 degrees of freedom
  (304 observations deleted due to missingness)
Multiple R-squared:  0.0005965,  Adjusted R-squared:  -0.002284
F-statistic: 0.2071 on 1 and 347 DF,  p-value: 0.6493


> summary(lm(a$Overall_Mark~a$Year))
Call:
lm(formula = a$Overall_Mark ~ a$Year)

Residuals:
     Min       1Q   Median       3Q      Max
-25.3532  -3.6797  -0.0532   3.5937  17.7734

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) -482.76223  142.72293  -3.383 0.000776 ***
a$Year         0.27344    0.07096   3.853 0.000132 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 5.99 on 489 degrees of freedom
  (162 observations deleted due to missingness)
Multiple R-squared:  0.02947,    Adjusted R-squared:  0.02748
F-statistic: 14.85 on 1 and 489 DF,  p-value: 0.0001321


> summary(lm(a$Overall_Mark~as.factor(a$Year)))
Call:
lm(formula = a$Overall_Mark ~ as.factor(a$Year))

Residuals:
     Min       1Q   Median       3Q      Max
-25.1622  -3.9071  -0.1279   4.0488  18.4864

Coefficients:
                     Estimate Std. Error t value Pr(>|t|)
(Intercept)           64.4615     1.6556  38.936  < 2e-16 ***
```

```
as.factor(a$Year)2005   3.3051     2.1066   1.569  0.11733
as.factor(a$Year)2006   1.5228     1.9633   0.776  0.43834
as.factor(a$Year)2007   2.0836     1.9724   1.056  0.29133
as.factor(a$Year)2008   2.2664     1.8894   1.200  0.23091
as.factor(a$Year)2009   2.4456     1.8946   1.291  0.19738
as.factor(a$Year)2010   0.4897     1.9000   0.258  0.79673
as.factor(a$Year)2011   0.8521     1.8844   0.452  0.65133
as.factor(a$Year)2012   3.8538     2.0277   1.901  0.05795 .
as.factor(a$Year)2013   2.3748     1.9724   1.204  0.22919
as.factor(a$Year)2014   3.3006     1.9246   1.715  0.08700 .
as.factor(a$Year)2015   4.1994     1.9000   2.210  0.02756 *
as.factor(a$Year)2016   5.2237     1.8844   2.772  0.00579 **
as.factor(a$Year)2017   4.4440     1.8796   2.364  0.01846 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 5.969 on 477 degrees of freedom
  (162 observations deleted due to missingness)
Multiple R-squared:  0.05975,    Adjusted R-squared:  0.03413
F-statistic: 2.332 on 13 and 477 DF,  p-value: 0.005158
```