1. Calculate a 95% confidence interval for the mean Hours worked based on your sample. Give a statement indicating how the confidence interval should be interpreted.

Mean with 95% Confidence Interval

Lower	Mean	Upper
32.45	34.65	36.85

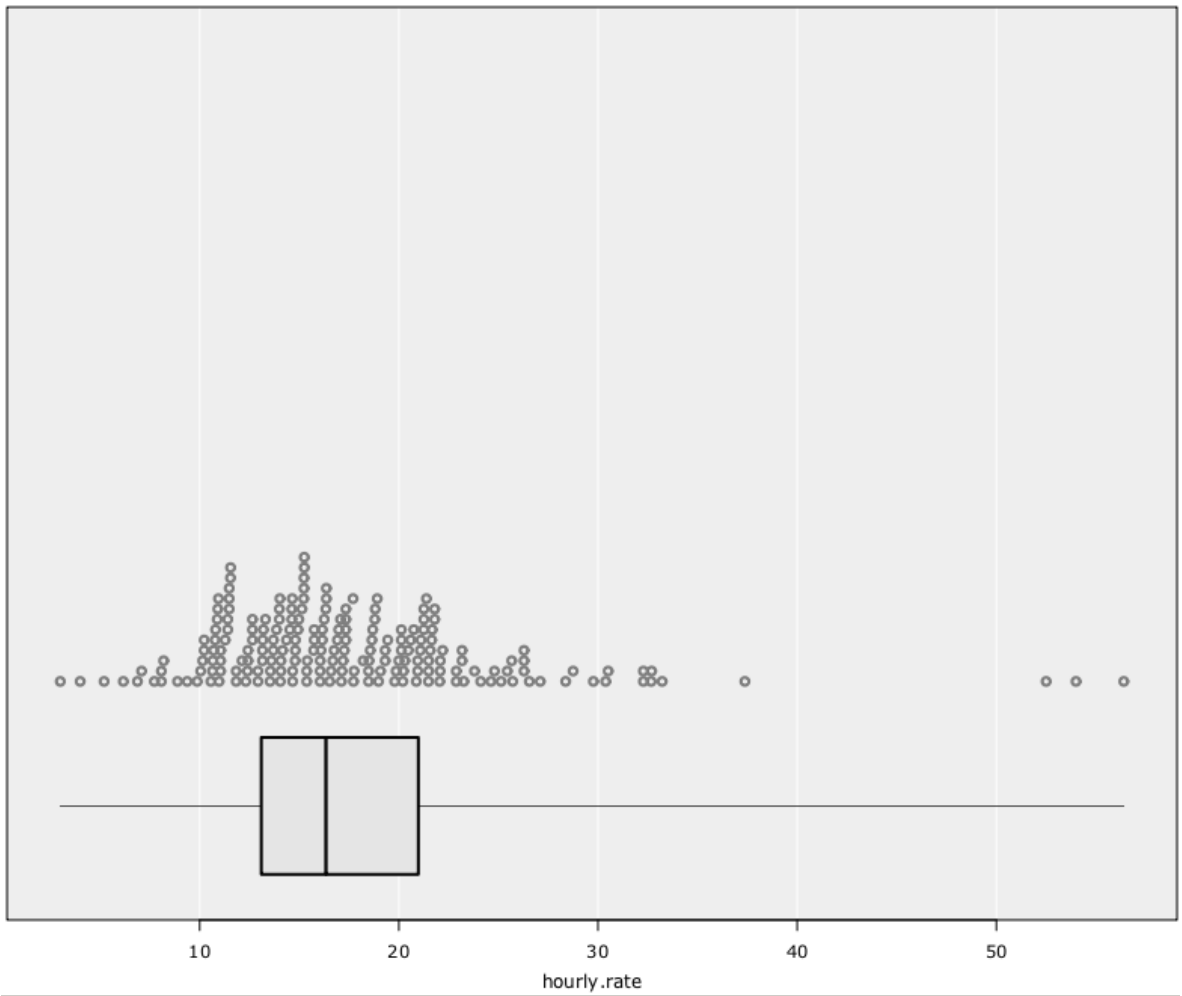
We are 95% confident that the mean hours lie between 32.45-36.85 hours

2.
In iNZight, create a new variable, Hourly Rate = Income / Hours (use the menu item: Variables -> Create new variables). [Alternatively, create the variable in Excel before importing the file into iNZight.] Identify NZ's current Living Wage (e.g. google "living wage NZ"), and use iNZight to test for evidence that the mean hourly rate is below the living wage. Copy and paste the iNZight output for the test, and use the p-value to draw a conclusion. State clearly whether or not you are able to reject the null hypothesis, and why. Provide a graph that supports your finding, explaining how it does so.

Living wage=21.15

```
================================================================================================
                              iNZight Inference using Normal Theory
------------------------------------------------------------------------------------------------
   Primary variable of interest: hourly.rate (numeric)

   Total number of observations: 200
================================================================================================

Inference of hourly.rate:
-------------------------

Mean with 95% Confidence Interval

   Lower    Mean   Upper
    16.6   17.64   18.68

One Sample t-test

   t = -6.656, df = 199, p-value = 1.3363e-10

         Null Hypothesis: true mean is equal to 21.15
   Alternative Hypothesis: true mean is less than 21.15


================================================================================================
```

The p-value is 1.3363e-10
Which means that the Null Hypothesis: true mean is equal to 21.15 is very unlikely (practically 0%)- hence we can reject the null hypothesis as it is clearly below 5%
Therefore we are 95% confident that the mean hourly income is below the living wage



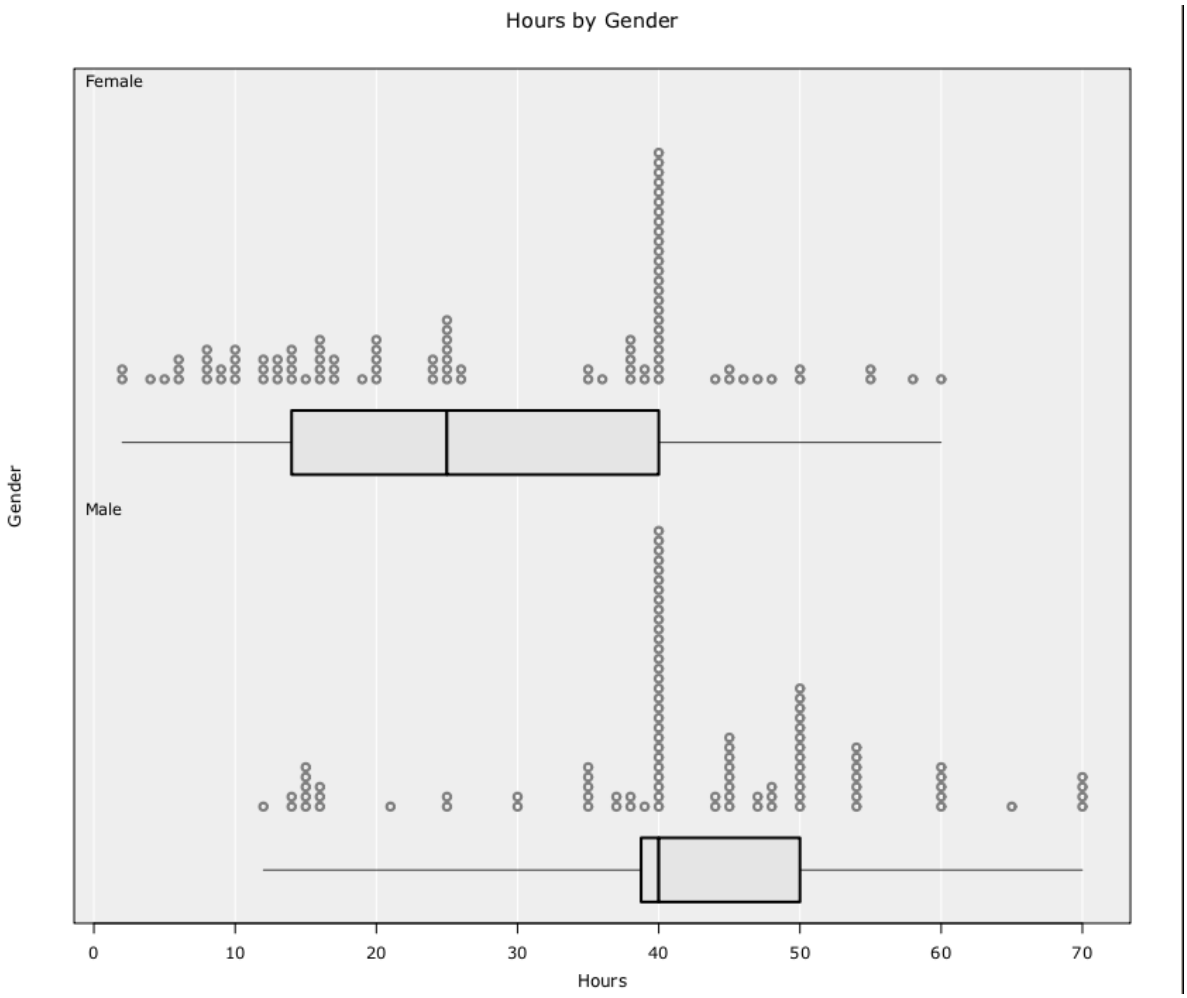
The graph supports this as the mean lies below the living wage of 21.15 and so does most of the interquartile range (most accurate 50%) additionally the graph is skewed to the right which indicates that income earned above the upper quartile is inconsistent

3.
Now, test whether or not there is evidence of a difference in average Hours worked for males vs females. Copy and paste the relevant iNZight output for the test, and use the p-value to draw a conclusion. State clearly whether or not you are able to reject the null hypothesis, and why. Provide a graph that supports your finding, explaining how it does so. Also, justify your choice of test, in particular, whether or not you've used an assumption of equal variance. Provide graphical evidence or formal inference justifying your choice.

```
                Secondary variable: Hours (numeric)

   Total number of observations: 200
================================================================================================

Inference of Hours by Gender:
-----------------------------

Group Means with 95% Confidence Intervals

            Lower    Mean   Upper
   Female   24.56   27.53   30.50
     Male   39.15   41.77   44.39

Difference in group means with 95% Confidence Interval

                   Lower     Mean    Upper
   Female - Male   -18.17   -14.24   -10.31

Welch Two Sample t-test

   t = -7.1389, df = 195.01, p-value = 1.8132e-11

          Null Hypothesis: true difference in means is equal to 0
   Alternative Hypothesis: true difference in means is not equal to 0


================================================================================================
```

Yes we are able to reject the null hypothesis: true group means are equal as the p-value for this is p-value = 1.8132e-11 (practically zero) which is well below 5% Hence we are 95% confident that the true group means are not equal

No I have not made an assumption of equal variance as the lower quartile, mean and upper quartile appear to be significantly different so I have used a welch two sample t-test instead of an anova



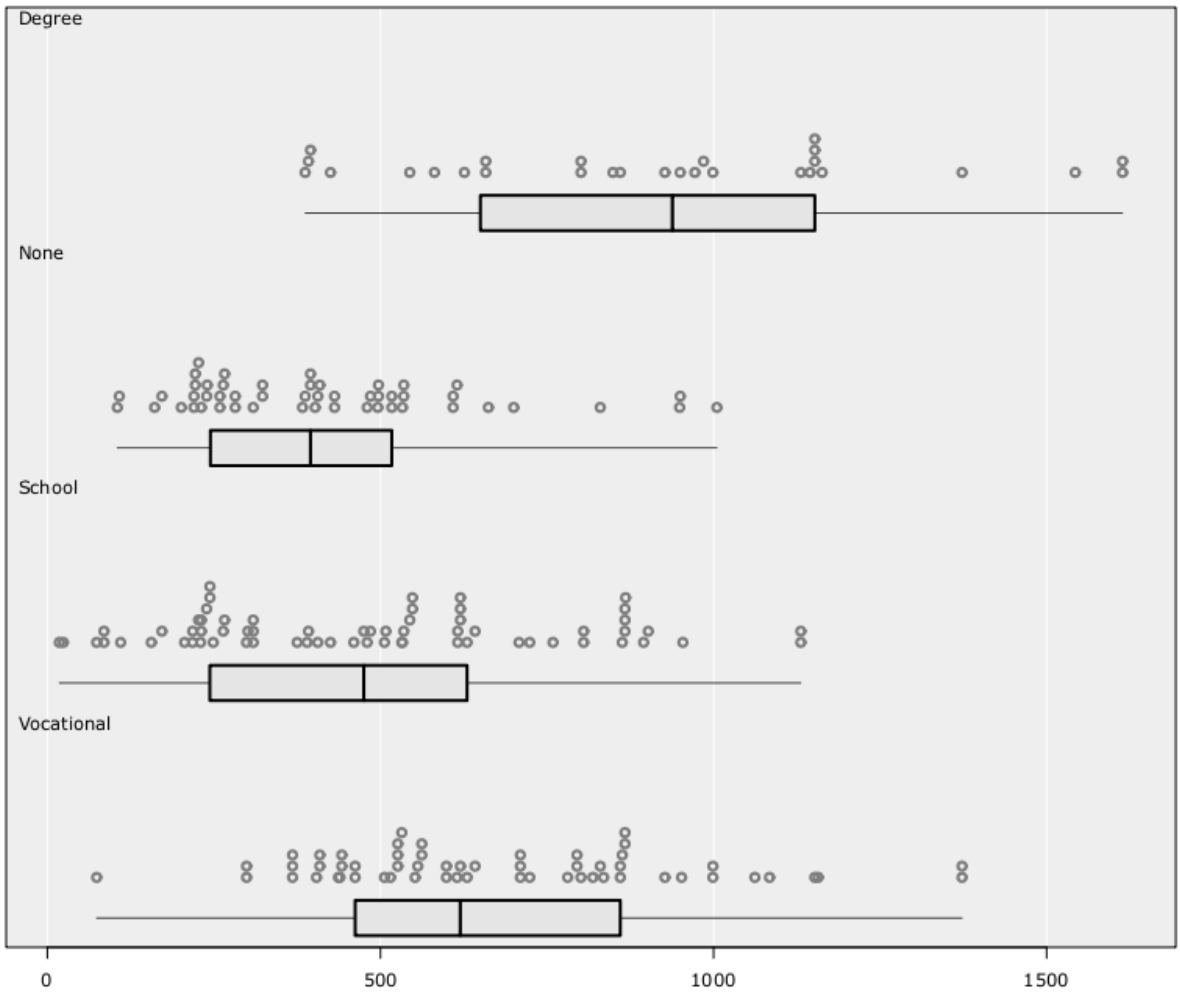
from the graph we can see that the Male box is shifted to the right which suggest that the male's mean is larger additionally The female mean lies outside the IQR (Most accurate 50%) of the male sample - this suggest that males are more likely to work longer hours than females

4.

Using the Income data, conduct a test of the null hypothesis that average Income does not vary by Qualification. Copy and paste the relevant iNZight output for the test, and use the p-value to draw a conclusion. State clearly whether or not you are able to reject the null hypothesis, and why. Provide a graph that supports your finding, explaining how it does so.

```
                                 Insight Inference using Normal Theory
--------------------------------------------------------------------------------
   Primary variable of interest: Income (numeric)
             Secondary variable: Qualification (categorical)

   Total number of observations: 200
================================================================================

Inference of Income by Qualification:
-------------------------------------

Group Means with 95% Confidence Intervals

                Lower    Mean    Upper
       Degree   783.8   923.2   1062.5
         None   355.1   415.9    476.7
       School   407.6   476.5    545.3
   Vocational   606.5   678.2    749.8

One-way Analysis of Variance (ANOVA F-test)

   F = 26.088, df = 3 and 196, p-value = 3.0439e-14

          Null Hypothesis: true group means are equal
   Alternative Hypothesis: true group means are not equal


### Difference in mean Income between Qualification groups
    (col group - row group)

Estimates

                Degree      None   School
         None    507.3
       School    446.7    -60.56
   Vocational    245.0   -262.26   -201.7

95% Confidence Intervals (adjusted for multiple comparisons)

                Degree      None    School
         None   339.49
                675.06
       School   286.03   -194.28
                607.40     73.15
   Vocational    80.98   -399.99   -330.69
                409.07   -124.52    -72.71

P-values

                Degree   None   School
         None    0e+00
       School    0e+00   0.64
   Vocational    8e-04   0.00    4e-04


================================================================================
```

Null Hypothesis: true group means are equal
We are able to reject the null hypothesis (true group means are equal) as the p-value is 3.0439e-14 which is well below 5%



The varying position of the mean for each qualification indicate that average income does vary by qualification

5.
Calculate the sample proportion of people who have a Degree. Use it to test the claim that fewer than 20% of the population have a Degree. State your null and alternative hypotheses clearly. Calculate the test statistic and use Excel to give a rejection region or p-value. State which you choose and what it is, and draw a conclusion.

```
================================================================================================
                                iNZight Inference using Normal Theory
------------------------------------------------------------------------------------------------
   Primary variable of interest: Qualification (categorical)

   Total number of observations: 200
================================================================================================

Inference of the distribution of Qualification:
-----------------------------------------------

Estimated Proportion with 95% Confidence Interval

                Lower   Estimate   Upper
       Degree   0.0919      0.140   0.188
         None   0.1900      0.250   0.310
       School   0.2601      0.325   0.390
   Vocational   0.2224      0.285   0.348

Chi-square test for equal proportions

   X^2 = 15.16, df = 3, p-value = 0.0016849

         Null Hypothesis: true proportions in each category are equal
   Alternative Hypothesis: true proportions in each category are not equal


### Differences in proportions of Qualification
    (col group - row group)

Estimates

               Degree     None   School
         None  -0.110
       School  -0.185   -0.075
   Vocational  -0.145   -0.035     0.04

95% Confidence Intervals

                  Degree       None    School
         None  -0.224680
                0.004681
       School  -0.307440   -0.21577
               -0.062559    0.06577
   Vocational  -0.263570   -0.17130   -0.1055
               -0.026429    0.10130    0.1855

================================================================================================
```

In the sample 14% of participants have a degree
Ho=P=20%=20% of population has a degree
Ha= P < 20%= less than 20% of population has a degree

N =200

Test stat =0.2-0.14/SQRT(0.14*(1-0.14)/200)=-5.505974
P Value = t.dist(-5.505974,199,TRUE)= 5.62138E-08
P value is less than 5% therefore we can reject the null hypothesis that 20% of the population has a degree. This means that the alternative hypothesis that less than 20% of the population have an qualified degree is true.