

Name: Prakruthi Shetty

UmichID: pshetty

What dataset are you working with:

1. love\_actually\_appearance
2. US\_births\_2000\_2014
3. male\_flight\_attend

List 3 questions that you can ask with your dataset.

- Q1: Do Colin Firth and Hugh Grant appear in any scenes together?  
Q2: Is there no difference in the number of births in the year 2000 and 2010?  
Q3: Do most of the occupation have 50% of males?

List the associated null hypothesis for each question:

- Q1: Colin Firth's and Hugh Grant's appearance don't overlap  
Q2: Number of births in 2000 and 2010 are same  
Q3: Male percentage on an average in any occupation is 50

What statistical test(s) will you use to answer each of the questions:

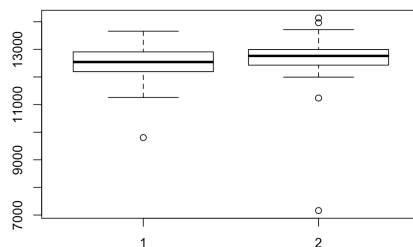
- Q1: chi-square test  
Q2: 2-way T-test  
Q3: 1-way t-test

Make a visual plot showing the relationship that you will analyze statistically (e.g. boxplot for t-test or ANOVA; scatterplot for regression; table for chi-square).

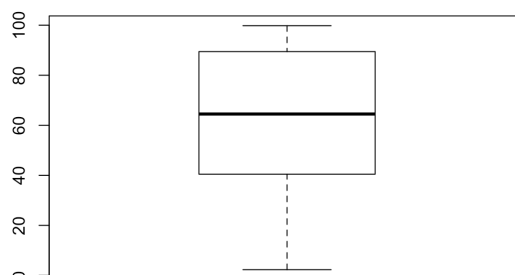
Q1:

	FALSE	TRUE
FALSE	46	12
TRUE	12	1

Q2:



Q3:



Do your data meet the assumptions required for the statistical test you want to run?  
Please state the assumptions you examined and whether or not your data meet those assumptions:

Q1: Its not a randomly collected data as it is from a movie and the observations may not be independent as actors appearance in a scene may have influenced another actors appearance, but the the sample is more than 30.

Q2: Is randomly sampled, independent values, variance between the groups are same, sample size is more than 30.

Q3:

Run the statistical test! Put your results here:

Q1: `chisq.test(love_actually_appearance$colin_firth, love_actually_appearance$hugh_grant)`

Q2: `t.test(birth2000Fri$births, birth2010Fri$births)`

Q3: `t.test (maleperc$percentage_male, mu = 50)`

Interpret your results!

Pearson's Chi-squared test with Yates' continuity correction

```
data: love_actually_appearance$colin_firth and love_actually_appearance$hugh_grant
X-squared = 0.48784, df = 1, p-value = 0.4849
```

Q1: Since the p-value is greater than 0.05, we can reject the hypothesis. Colin Firth and Hugh Grant appear together in a scene.

Welch Two Sample t-test

```
Q2: data: birth2000Fri$births and birth2010Fri$births
t = -0.93782, df = 90.586, p-value = 0.3508
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -452.8016 162.3785
sample estimates:
mean of x mean of y
12524.50 12669.71
```

Since p-value is greater than 0.05, our null hypothesis that the number of births were same is rejected.

Q3:

One Sample t-test

```
data: maleperc$percentage_male
t = 7.7817, df = 319, p-value = 1.004e-13
alternative hypothesis: true mean is not equal to 50
95 percent confidence interval:
 59.17388 65.38237
sample estimates:
mean of x
62.27813
```

Since the p-value is less than 0.05, our null hypothesis that on an average male percentage in any occupation is 50 is true.