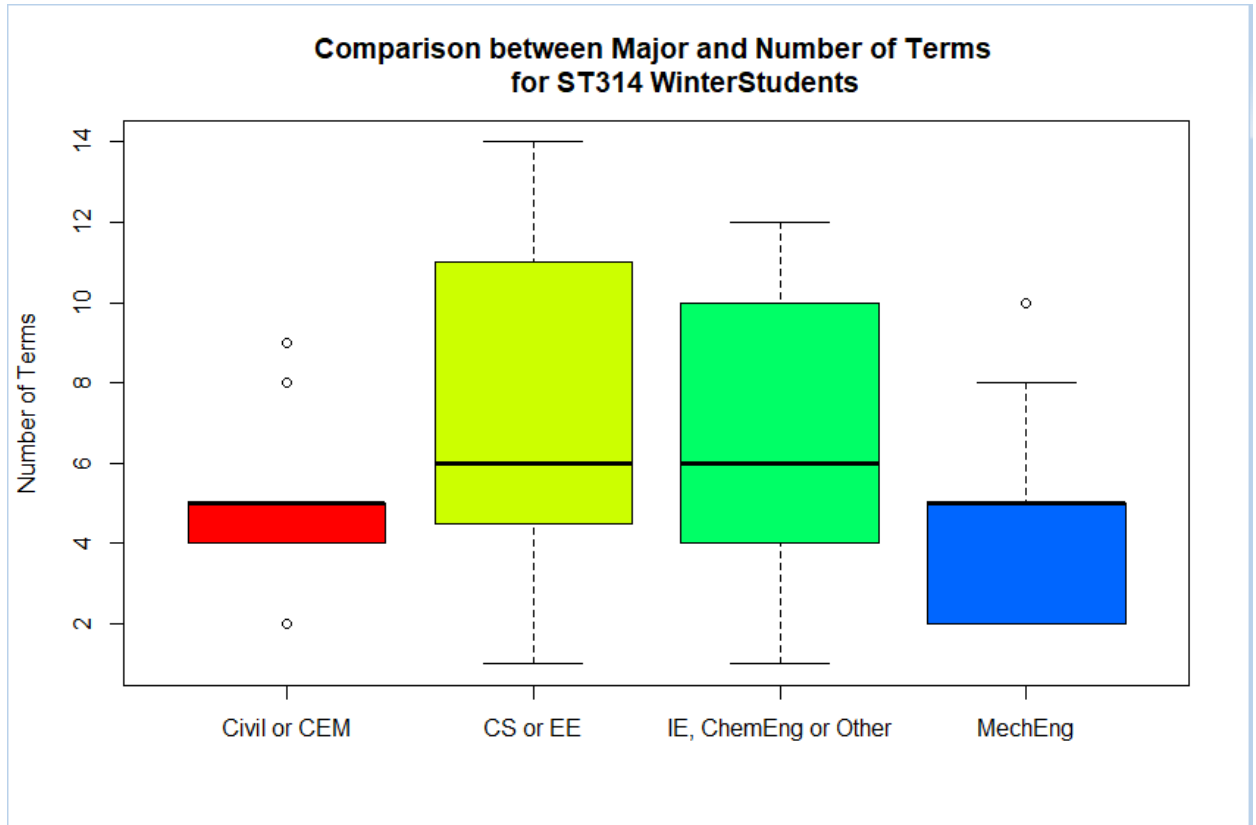


- Data Analysis #6 -

Part 1 – Option 1: Is there evidence the average number of terms a student has been studying at OSU differs between majors?

a.



b. All four majors seem to have very similar median values, however Civil or CEM and Mechanical Engineering both have extremely small third quartile ranges. The other two have much larger third quartiles and similar second quartiles. This likely indicates that there are a lot of values above the median range that are very spread out and that the average number of terms in CS/EE and IE/ChemE/Other is higher than the median.

c. Null Hypothesis - $H_0: \mu_1 = \mu_2 = \mu_3 = \mu_4$

Alternative Hypothesis - H_a : at least two μ 's are unequal.

d. Conditions for single factor ANOVA test:

- 1) Samples are obtained using random mechanism
- 2) Populations are independent
- 3) Populations are normal
- 4) Population standard deviations are the same

I would say that I am doubtful all conditions are met. The populations are likely normal, and the samples are independent. However, the samples were not necessarily obtained using a random mechanism if online ST314 students were simply surveyed. Additionally, the standard deviations for all populations look like they vary, which would mean only two of the four conditions are “satisfied”.

e.

```
> summary(mod)
              Df Sum Sq Mean Sq F value Pr(>F)
Major          3   82.1    27.37   2.695 0.0534 .
Residuals     63  639.8    10.16
```

From the ANOVA table, the mean square of treatment is $MSTr = 27.37$, the mean square of error is $MSE = 10.16$.

- f. There is convincing evidence that the average amount of terms across majors varies among online ST314 students. We reject the null at the 0.1 significance level, as the F statistic = 2.695 and the p-value = 0.0534. The strength of the evidence is not extremely compelling, as it would not reject the null at a significance level of 0.05. In the context of this problem, there is enough evidence to contradict the null hypothesis, however not necessarily enough to say for certain that the terms do not vary across majors. That said, it makes some logical sense that different majors would have more or less terms under their belt when they take ST314 depending on their major requirements.

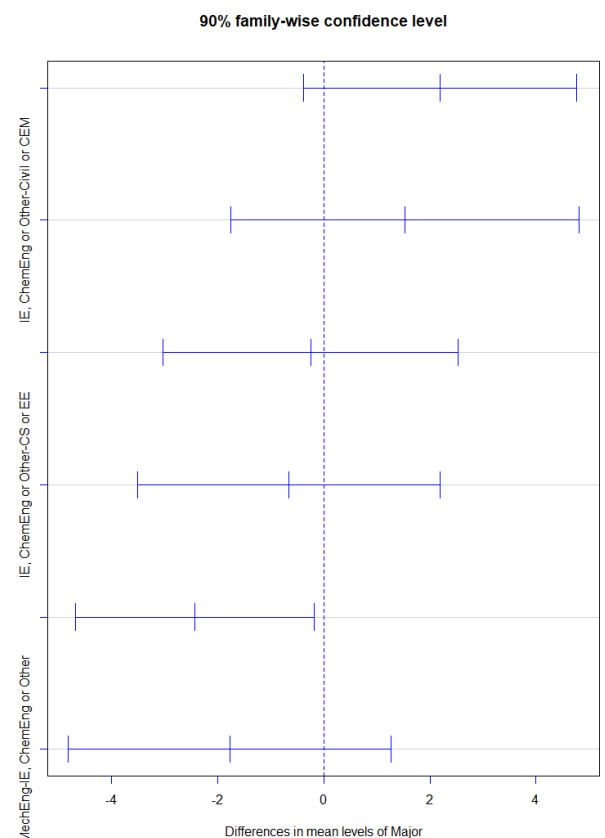
g.

```
Tukey multiple comparisons of means
90% family-wise confidence level

Fit: aov(formula = Terms ~ Major, data = st314data)

$Major
              diff          lwr          upr          p adj
CS or EE-Civil or CEM      2.1904762 -0.3818614  4.7628138  0.2018604
IE, ChemEng or Other-Civil or CEM 1.5277778 -1.7597193  4.8152749  0.6985919
MechEng-Civil or CEM      -0.2500000 -3.0284422  2.5284422  0.9966686
IE, ChemEng or Other-CS or EE  -0.6626984 -3.5194219  2.1940251  0.9481712
MechEng-CS or EE          -2.4404762 -4.6928022  -0.1881501  0.0641721
MechEng-IE, ChemEng or Other -1.7777778 -4.8214087  1.2658531  0.5248387
```

In this instance, the only comparison that is significant at the 90% confidence level is the Mechanical Engineer and CS/EE comparison. The 90% family-wise confidence level for this comparison means that we can conclude with 90% certainty that the mean terms for the two majors are not the same.



On the plot this is represented at the range that fails to include zero (second to lowest). Should we increase the confidence level to 95%, this would likely mean that our range would include zero, and the comparison would not be significant. This is supported by the p-value for the comparison, 0.064 which would not be significant at $\alpha = 0.05$.