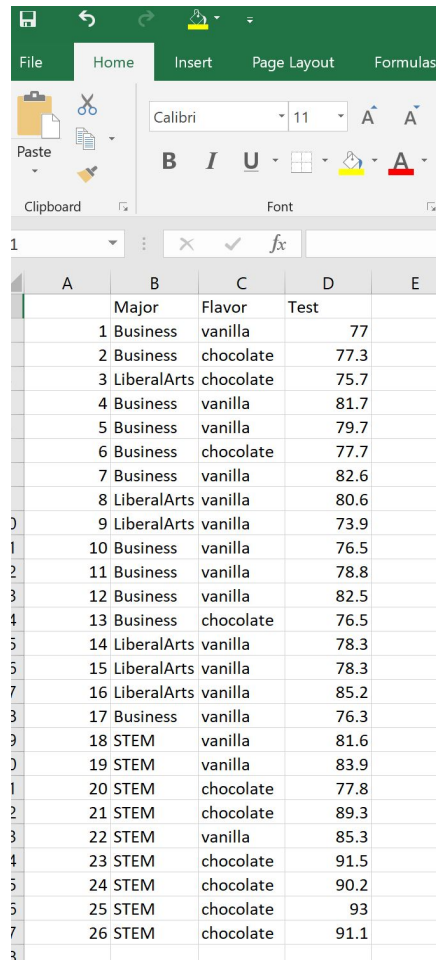


ANOVA Tutorial

We collect the following data from 26 students in our math class:



	A	B	C	D	E
		Major	Flavor	Test	
1	1	Business	vanilla	77	
2	2	Business	chocolate	77.3	
3	3	LiberalArts	chocolate	75.7	
4	4	Business	vanilla	81.7	
5	5	Business	vanilla	79.7	
6	6	Business	chocolate	77.7	
7	7	Business	vanilla	82.6	
8	8	LiberalArts	vanilla	80.6	
9	9	LiberalArts	vanilla	73.9	
10	10	Business	vanilla	76.5	
11	11	Business	vanilla	78.8	
12	12	Business	vanilla	82.5	
13	13	Business	chocolate	76.5	
14	14	LiberalArts	vanilla	78.3	
15	15	LiberalArts	vanilla	78.3	
16	16	LiberalArts	vanilla	85.2	
17	17	Business	vanilla	76.3	
18	18	STEM	vanilla	81.6	
19	19	STEM	vanilla	83.9	
20	20	STEM	chocolate	77.8	
21	21	STEM	chocolate	89.3	
22	22	STEM	vanilla	85.3	
23	23	STEM	chocolate	91.5	
24	24	STEM	chocolate	90.2	
25	25	STEM	chocolate	93	
26	26	STEM	chocolate	91.1	

Statistical Goal. We want to know if this data provides **statistically significant evidence**¹ that a student's major and/or favorite flavor has a predictive effect on how such a student will, on average, score on the math test. The key thing is that we are going to make **inferences** about students in general based upon how our 26 students (our **sample**, $n = 26$) did.

Terminology. The way such a question is typically answered is by performing what is called an **ANOVA**, which stands for "Analysis of Variance". In this example we have two **factors**: Major, which has three **levels**: STEM, Liberal Arts, and Business; and Flavor, which has 2 levels: chocolate or vanilla. Since there are two factors the ANOVA in this example is

¹ **Understanding statistical significance and p-values: an example.** Suppose you want to know who will win the next election.

Case 1. You ask three people how they will vote and two say "Democrat". This would not be considered statistically significant evidence (exact binomial test, $p\text{-value } p = 0.5 > 0.05$) that the Democrats will win because the $p\text{-value} > 0.05$. Explanation: even though 2 out of 3 favored the Democrats, the sample size was too small. Hence, the $p\text{-value}$ ended up being too big.

Case 2. You ask 100 people and 51 of them say they'll vote "Democrat". This would not be considered statistically significant evidence (exact binomial test, $p\text{-value } p = 0.4602 > 0.05$) that the Democrats will win because the $p\text{-value} > 0.05$. Explanation: even though the sample size is not small, only 51% favored the Democrats, 51% is too close to 50%. Hence, the $p\text{-value}$ ended up being too big.

Case 3. You ask 100 people and 59 of them say they'll vote "Democrat". This would be considered statistically significant evidence (exact binomial test, $p\text{-value } p = 0.0443 < 0.05$) that the Democrats will win because the $p\text{-value} < 0.05$. Explanation: the sample size is large enough and 59% is not too close to 50%. Hence, the $p\text{-value}$ ended up being small enough.

Note. Just because you have statistically significant evidence that some claim is true (e.g., that the democrats will win), it doesn't mean the claim is true. For example, your data might not be representative of the population from which it is drawn; or your respondents might be lying; or your data might be corrupted due to faulty equipment, or your data isn't measuring what you think it is, etc.

The **p-value** is a mathematical measure of how "gullible" we need to be in order to accept a given claim as true based upon the given data. So, the smaller the $p\text{-value}$, the less gullible we'd need to be, and less gullible is good. The cut-off for statistical significance is a $p\text{-value}$ of .05. **If the $p\text{-value} < .05$ we can say we have statistical significance.** On a practical level, the $p\text{-value}$ is the first hurdle your study needs to clear. It tells your readers that if they are to dispute your results, they shouldn't say things like your sample size was too small, but rather they should attack your selection process as being biased or not-random, or that your survey was poorly worded, or that your respondents aren't answering honestly, and so on.

called a two-way ANOVA. The Test scores are called the **dependent variable** because we think of them as depending on the factors.

We can multiply the factors together to get **interactions**. In our example, if we multiply the factors Major and Flavor we get $3 \times 2 = 6$ interaction groups or cells (one group would be “STEM majors who prefer vanilla,” another would be “STEM majors who prefer chocolate”).

Analyzing the Data in R

The work flow for statistical analysis is like this. You keep your data in a data file, which for us will be a spreadsheet, e.g., Excel² or OpenOffice³, which is free. When you want to analyze your data, you import it into statistics software, which for us will be the free, open-source statistical computing environment called R⁴. We will use an R script, see below, to analyze our data. R will not modify the data file, unless you explicitly tell it to.

We will use the following R script to analyze the data from the example and to perform the (factorial, 2 way) ANOVA.

```
# ANOVA for SoTL      anovaSoTL.R

#-----
# Import Data From Spreadsheet (Step 1)
df = read.table(file = "clipboard", sep = "\t", header = TRUE);
df; # if you did this right, you will see the students' data in the R console.
#-----
# Boxplots of Data (Step 2)
boxplot(df$Test ~ df$Major,col = "cyan",
main = "Boxplots of students' test scores by major.");

boxplot(df$Test ~ df$Flavor,col = c("brown", "yellow"),
main = "Boxplots of students' test scores by favorite flavor.");

boxplot(df$Test ~ df$Flavor*df$Major,col = "green",
main = "Boxplots of students' test scores by major x favorite flavor.");
#-----
# Perform the ANOVA (Step 3)
dfANOVA <- aov(Test ~ Major * Flavor, data = df);
summary(dfANOVA);
# you should see the ANOVA table in the R console.
#-----
# Make a table with means and standard deviations sd (Step 4)

# for main effects
agMajor = aggregate(Test~Major, df, function(x) c(mean = mean(x), sd = sd(x),n = length(x)));
agMajor;

agFlavor = aggregate(Test~Flavor, df, function(x) c(mean = mean(x), sd = sd(x), n = length(x)));
agFlavor ;

# for interaction effects
agMajorFlavor = aggregate(Test~Major*Flavor, df, function(x) c(mean = mean(x), sd = sd(x),n = length(x)));
agMajorFlavor;
#-----
# End of script
```

Step 1. How to import the data into R from Excel by copy and paste. Open the anovaSoTL.R script file in R. Then, in the spreadsheet copy the entire Major, Flavor, and Test columns to the “clipboard”.

²Microsoft Office

³<https://www.openoffice.org/> An excellent, free, open-source version of Microsoft Office.

⁴<https://www.r-project.org/> We will explore and analyze our data using the very powerful, open source statistical software R. R is probably the most widely used statistics software by professional statisticians. You can **download R for free** and find **R tutorials** at: <https://www.r-project.org/>. R is very easy to use. You can easily find free resources online to help you learn R. Also, see the Appendix of these notes for a short introduction to using R.

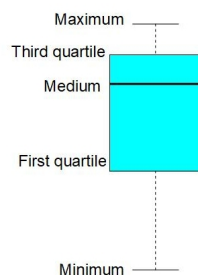
	B	C	D
1	Major	Flavor	Test
1	Business	vanilla	77
2	Business	chocolate	77.3
3	LiberalArts	chocolate	75.7
4	Business	vanilla	81.7
5	Business	vanilla	79.7
6	Business	chocolate	77.7
7	Business	vanilla	82.6
8	LiberalArts	vanilla	80.6
9	LiberalArts	vanilla	73.9
10	Business	vanilla	76.5
11	Business	vanilla	78.8
12	Business	vanilla	82.5
13	Business	chocolate	76.5
14	LiberalArts	vanilla	78.3
15	LiberalArts	vanilla	78.3
16	LiberalArts	vanilla	85.2
17	Business	vanilla	76.3
18	STEM	vanilla	81.6
19	STEM	vanilla	83.9
20	STEM	chocolate	77.8
21	STEM	chocolate	89.3
22	STEM	vanilla	85.3
23	STEM	chocolate	91.5
24	STEM	chocolate	90.2
25	STEM	chocolate	93
26	STEM	chocolate	91.1

Then, back in R, in the `anovaSoTL.R` script, run the following lines of code:

```
# Import Data From Spreadsheet (Step 1)
df = read.table(file = "clipboard", sep = "\t", header = TRUE);
df; # if you did this right, you will see the students' data in the R console.
```

Step 2. Produce boxplots of the groups.

Perhaps the quickest way to get a feel for how the different groups perform is to compare boxplots of their test scores. Also, boxplots are visually appealing. They will make your article look nicer!

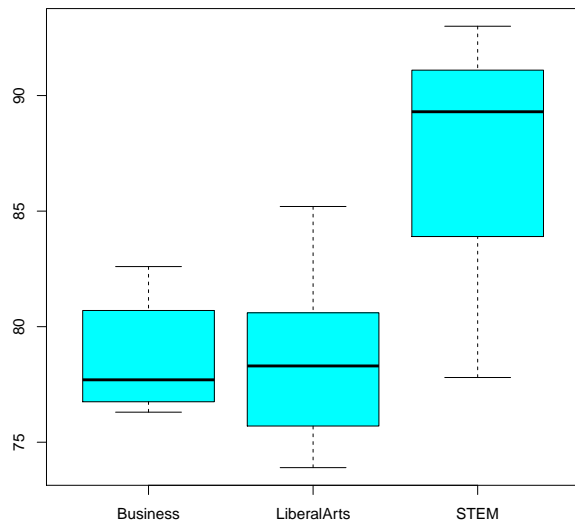


Once you've imported the data into R, run the following lines of code from the `anovaSoTL.R` script to produce the images of boxplots show below:

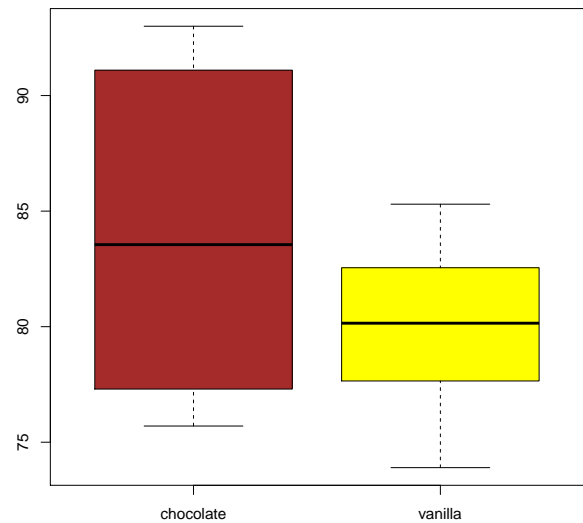
```
# Boxplots of Data (Step 2)
boxplot(df$Test ~ df$Major,
        col = "cyan", main = "Boxplots of students' test scores by major.");

boxplot(df$Test ~ df$Flavor,
        col = c("brown", "yellow"), main = "Boxplots of students' test scores by favorite flavor.");
```

Boxplots of students' test scores by major.



Boxplots of students' test scores by favorite flavor.

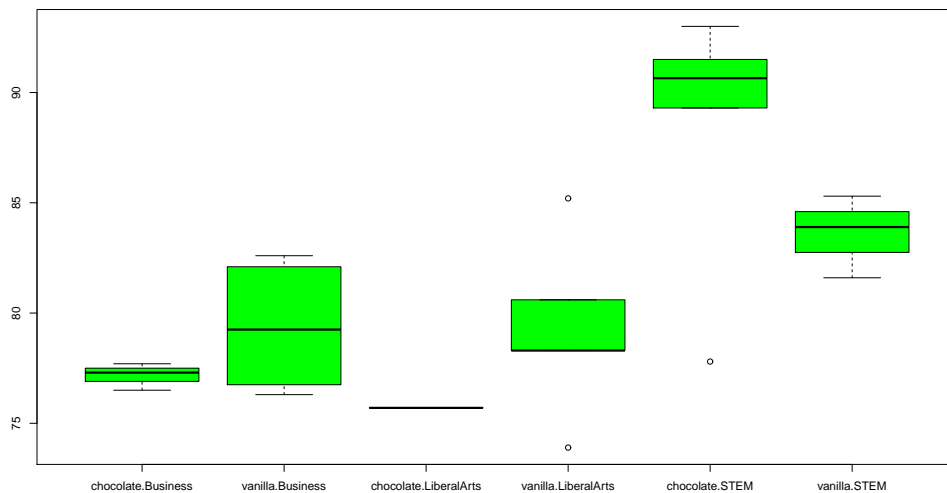


Main effects boxplots: Factors one at a time. See the above boxplots. It looks like the STEM majors (in our data) are scoring higher on the test than the students with other majors. However, favorite flavor doesn't seem to have much of an impact on average scores. It is worth mentioning that students who prefer chocolate seem to have more variation in their test scores. However, in this tutorial, we won't discuss variation very much.

Interaction effects boxplots: Factors two or more at a time. See the boxplots below. To see the interaction effects as boxplots run the following lines of code from the `anovaSoTL.R` script:

```
boxplot(df$Test ~ df$Flavor*df$Major,
        col = "green", main = "Boxplots of students' test scores by major x favorite flavor.");
```

Boxplots of students' test scores by major x favorite flavor.



Interaction Effects (Combinations of factors). Based on the above boxplots, it looks like the STEM majors who prefer chocolate did the best, STEM majors who prefer vanilla did the second best, and all other groups did about the same. However. . .

Warning! We have 6 interaction groups ($3 \text{ majors} \times 2 \text{ flavors}$), but our sample size n is only 26. In some of the interaction groups we have hardly any data. As a result, we should not be surprised if it turns out that our data doesn't provide statistically significant evidence about interaction effects.

Note. The little circles \circ appearing above and/below some of the boxplots indicate outliers, and/or are artifacts of the groups not having enough data in them.

Step 3. Perform the ANOVA

To run the ANOVA run the following lines of code from the `anovaSoTL.R` script:

```
# Perform the ANOVA (Step 3)
dfANOVA <- aov(Test ~ Major * Flavor, data = df);
summary(dfANOVA); # if you did this right, you will see an ANOVA table in the R console.
```

Here's the ANOVA table that R should produce for you:

```
> summary(dfANOVA);
          Df Sum Sq Mean Sq F value    Pr(>F)
Major      2  409.0   204.51   14.704 0.000118 ***
Flavor     1    1.4    1.37    0.098 0.756889
Major:Flavor 2   74.4    37.19    2.674 0.093524 .
Residuals 20  278.2    13.91
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Interpreting the ANOVA table. The stars * tell us if there is statistically significant evidence of an effect.

The row for Major has stars *. This indicates that there is statistically significant evidence that Major has an “effect” on how students, on average, will do on the test.

The row for Flavor has no stars. We interpret this as indicating that the data does not provide statistically significant evidence that students who prefer one of the flavors will score higher than students who prefer the other flavor (on average).

The row for the Major:Flavor interaction has no stars. We interpret this as indicating that the data does not provide statistically significant evidence that students with a particular major and flavor preference will score higher than students with a different combination of major and flavor preference (on average). It does have a period . at the end. We'll ignore that for now.

Step 4. Create a table of means and standard deviations.

The following commands from the `anovaSoTL.R` script will produce a table of means, standard deviations (sd), and level (or group or cell) size (n).

```
# Make a table with means and standard deviations sd (Step 4)

# for main effects
agMajor = aggregate(Test~Major, df, function(x) c(mean = mean(x), sd = sd(x), n = length(x)));
agMajor;

agFlavor = aggregate(Test~Flavor, df, function(x) c(mean = mean(x), sd = sd(x), n = length(x)));
agFlavor ;

# for interaction effects
agMajorFlavor = aggregate(Test~Major*Flavor, df, function(x) c(mean = mean(x), sd = sd(x), n = length(x)));
agMajorFlavor;
```

which outputs:

```
> agMajor;
      Major Test.mean  Test.sd   Test.n
1   Business 78.781818  2.467314 11.000000
2 LiberalArts 78.666667  3.955081  6.000000
3      STEM  87.077778  5.182127  9.000000
```

```
> agFlavor ;
      Flavor Test.mean  Test.sd   Test.n
1 chocolate 84.010000  7.471048 10.000000
2  vanilla  80.137500  3.349801 16.000000
```

```
> agMajorFlavor;
      Major  Flavor Test.mean  Test.sd   Test.n
1   Business chocolate 77.1666667  0.6110101 3.0000000
2 LiberalArts chocolate 75.7000000      NA 1.0000000
3      STEM  chocolate 88.8166667  5.5394645 6.0000000
4   Business  vanilla 79.3875000  2.6556880 8.0000000
5 LiberalArts  vanilla 79.2600000  4.1125418 5.0000000
6      STEM   vanilla 83.6000000  1.8681542 3.0000000
```

Note. The NA in row 2 stands for “Not Available”. This occurred because there is only one Liberal Arts Major who prefers chocolate, i.e., $\text{Test.n} = 1$ for row 2. To calculate the standard deviation (sd), the sample size n needs to be at least 2.

Step 4. How to write up these results in APA form.

Reporting results of major tests in factorial ANOVA; non-significant interaction:

Math test scores were subjected to a two-way analysis of variance having three levels of major (Business, Liberal Arts, STEM) and two levels of favorite flavor (chocolate, vanilla). The factorial ANOVA revealed that there was no main effect of favorite flavor on math test scores ($F(1,20) = 0.098$, $p > .1$). There was, however, a significant main effect of major on math test scores ($F(2,20) = 14.704$, $p < .001$), suggesting that STEM majors ($M = 87.08$, $SD = 5.182$) scored higher on average than Business majors ($M = 78.78$, $SD = 2.467$) and Liberal Arts majors ($M = 78.67$, $SD = 3.955$). There was no statistically significant interaction between the two factors ($F(2,20) = 2.674$, $p > .05$).

Where did the numbers in the report come from?

They came from the ANOVA summary table

```
> summary(dfANOVA);
      Df Sum Sq Mean Sq F value  Pr(>F)
Major    2  409.0   204.51  14.704 0.000118 ***
Flavor    1    1.4    1.37   0.098 0.756889
Major:Flavor 2   74.4   37.19   2.674 0.093524 .
Residuals 20  278.2   13.91
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

and from the table of means and standard deviations (Step 4) for Major:

See next page.

```
> agMajor;
      Major Test.mean  Test.sd  Test.n
1   Business 78.781818  2.467314 11.000000
2 LiberalArts 78.666667  3.955081  6.000000
3      STEM 87.077778  5.182127  9.000000
```

In the report we wrote: The factorial ANOVA revealed that there was no main effect of favorite flavor on math test scores ($F(1,20) = 0.098$, $p > .1$).

Explanation From the ANOVA summary we know that the factor Flavor does not have a statistically significant effect on test scores. We know this two ways. (1) There are no stars * at the end of the Flavor row. (2) In the Flavor row, the p -value = $\Pr(>F) = 0.756889$. If the p -value $> .05$ then we don't have statistical significance for that factor (row). The Signif. codes in the ANOVA summary indicate what we should put for the p -value p . Since $0.1 < 0.756889 < 1$ we put $p > .1$. When we don't have significance we put $p > .$ The $F(1,20) = 0.098$ comes from Flavor having 1 Df and the residuals having 20 Df. The 0.098 comes from the F value. Note. The effects that factors have are called main effects.

In the report we wrote: There was, however, a significant main effect of major on math test scores ($F(2,20) = 14.704$, $p < .001$), suggesting that STEM majors ($M = 87.08$, $SD = 5.182$) scored higher on average than Business majors ($M = 78.78$, $SD = 2.467$) and Liberal Arts majors ($M = 78.67$, $SD = 3.955$).

Explanation From the ANOVA summary we know that the factor Major has a statistically significant effect on test scores. We know this two ways. (1) There are stars * at the end of the Major row. (2) In the Major row, the p -value = $\Pr(>F) = 0.000118$. If the p -value $< .05$ then we have statistical significance for that factor (row). The Signif. codes in the ANOVA summary indicate what we should put for the p -value p . Since $0 < 0.000118 < .001$ we put $p < .001$. When we have significance we put $p < .$ Note that M = mean and SD = standard deviation. These came from the table of means and standard deviations. The $F(2,20) = 14.704$ comes from Major having 2 Df and the residuals having 20 Df. The 14.704 comes from the F value.

In the report we wrote: There was no statistically significant interaction between the two factors ($F(2,20) = 2.674$, $p > .05$).

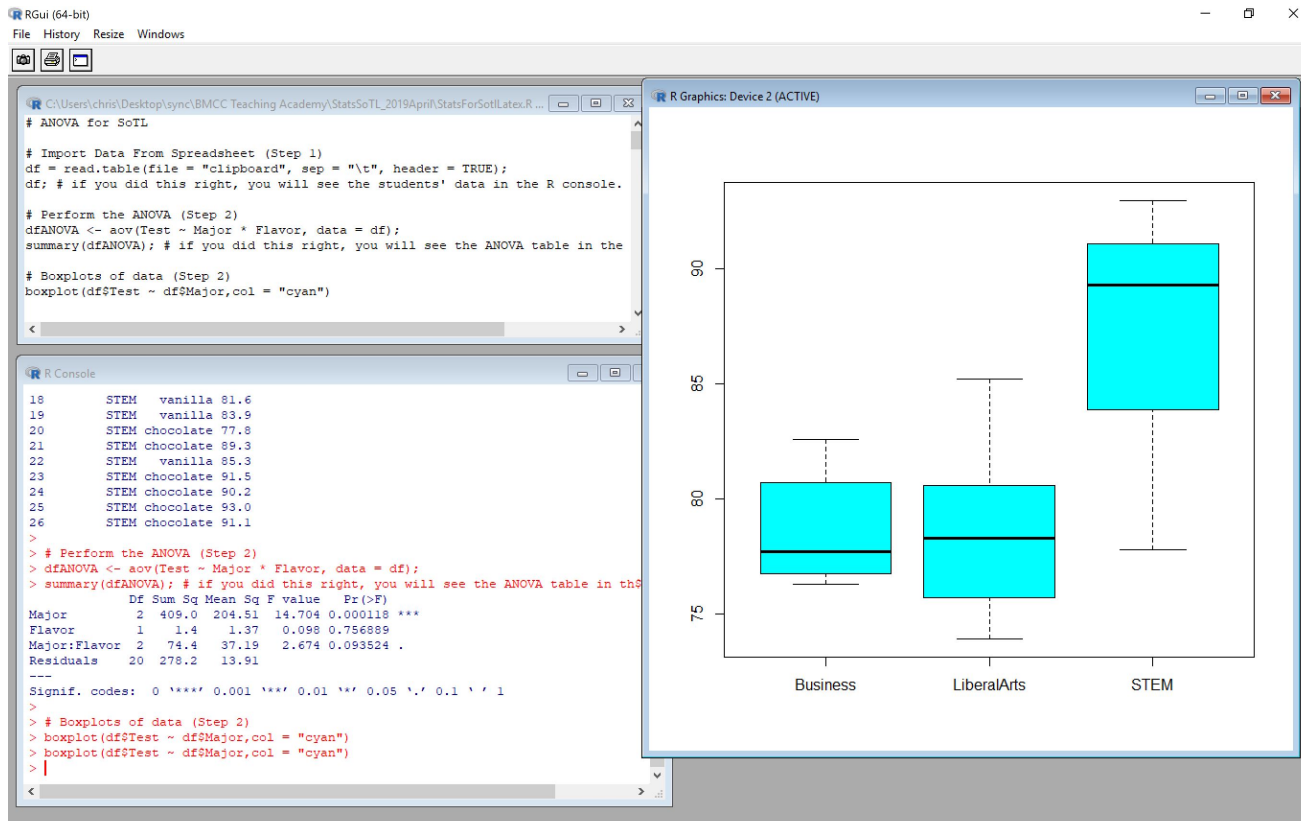
Explanation From the ANOVA summary we know that the interaction Major:Flavor does not have a statistically significant effect on test scores. We know this two ways. (1) There are no stars * at the end of the Major:Flavor row. (2) In the Major row, the p -value = $\Pr(>F) = 0.093524$. If the p -value $> .05$ then we do not have statistical significance for that row. The Signif. codes in the ANOVA summary indicate what we should put for the p -value p . Since $.05 < 0.093524 < .1$ we put $p > .05$. When we don't have significance we put $p > .$ The $F(2,20) = 2.674$ comes from Major:Flavor having 2 Df and the residuals having 20 Df. The 2.674 comes from the F value.

See next page for Appendix: Getting Started with R.

Appendix

Getting Started with R

This section is quick introduction to R to help you get started.



The above shows a screen shot of **R on a Windows PC**. R on an Apple will look slightly different. When you start R the “R Console” will appear (see above, bottom left). You can type commands directly into the R Console, however, it is better to run commands from the “R Editor” (top left). In the R Editor you can create “R scripts” which are just R commands in a text file. When you run R commands from the R Editor you’ll see the output in the “R Console” or in an “R Graphics” window (right).

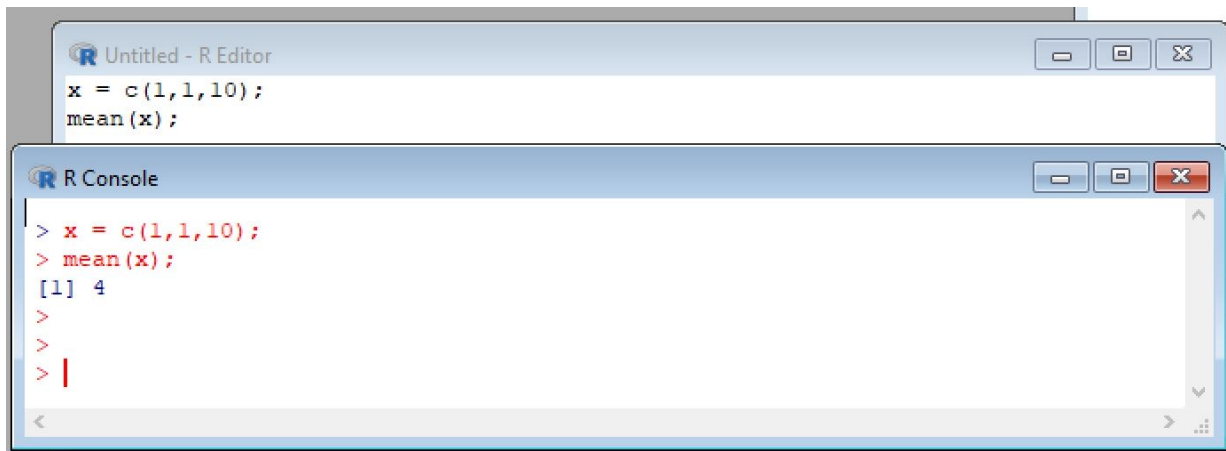
Two minute R Tutorial to Get You Started.

Finding the average of the numbers 1, 1, 10 in R.

1. Open R.
2. Inside of R create the script file by doing the following. Go to the menu, then File, then New script. A blank script will appear in the R Editor. Type, or copy and paste, the following R commands into your R script:

```
x = c(1,1,10);
mean(x);
```

3. Run the script. **On a Windows PC**, to run a line of code in an R script, put your cursor anywhere in that line and do **Ctrl R**. To run a bunch of lines, highlight the lines by doing **Ctrl Shift** and the **Up** or **Down** key, then do **Ctrl R** to run them.
On an Apple Computer, to run a line of code in an R script, put your cursor anywhere in that line and press **Command Enter** at the same time.
4. The results you should get. After you run this script you should see, in the R console window, the commands from the script displayed together with the number 4, since the mean of 1, 1, 10 is $\frac{1+1+10}{3} = \frac{12}{3} = 4$.



Notes:

1. If you want to save your script file you should save it with the .R extension. Something like `MyScript.R`.
2. To load your R script file into R you need to open R and then from inside of R, go to the menu, File, Open. Double clicking on the R file typically won't load it into R.
3. If you don't know how to do something in R just Google it. R is so widely used that just about any question you might have will have already been asked and answered somewhere online.
4. If you want to put a comment in your R script use the # symbol. Anything on a line after the # symbol is considered by R to be "comment", and won't be compiled.

End of R tutorial.