

**ASSIGNMENT FRONT SHEET**

**Course Name: ALY6015 20904 Intermediate Analytics**

**Professor Name: ChuanLi Jiang,**

**Student Name: Dong Quoc Tuong (Lukas)**

**Student Class: Fall 2019 CPS Term: A. 2020**

|  |
| --- |
| **Module 2: Hypothesis Testing**  **Completion Date: January 22st Due Time:12:00am** |

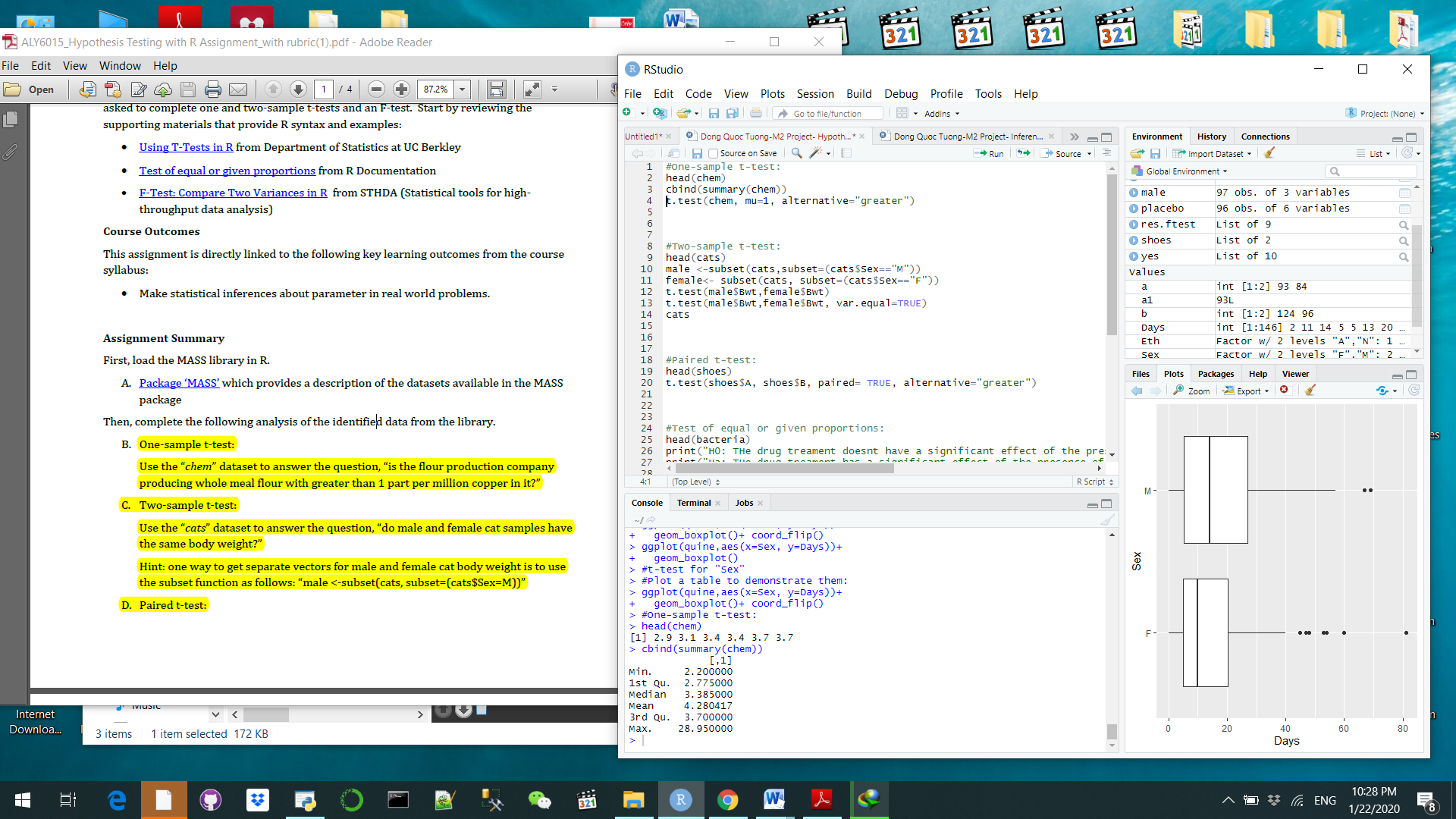
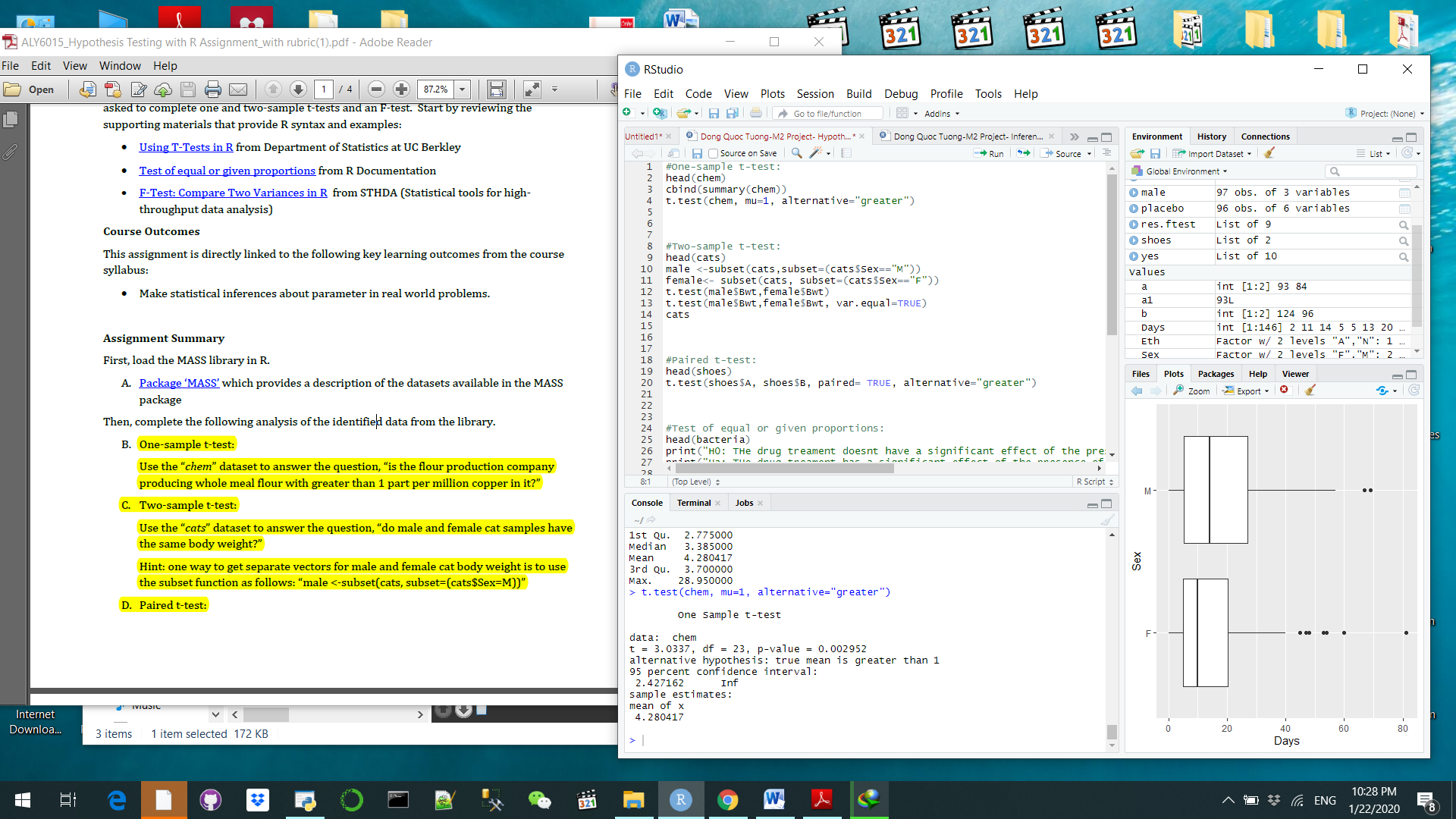
**Statement of Authorship**

*I confirm that this work is my own. Additionally, I confirm that no part of this coursework, except where clearly quoted and referenced, has been copied from material belonging to any other person e.g. from a book, handout, another student. I am aware that it is a breach of Northeastern University’s regulations to copy the work of another without clear acknowledgement and that attempting to do so renders me liable to disciplinary procedures. To this effect, I have uploaded my work onto Turnitin and have ensured that I have made any relevant corrections to my work prior to submission.*

**Tick here** to confirm that your paper version is identical to the version submitted through Turnitin

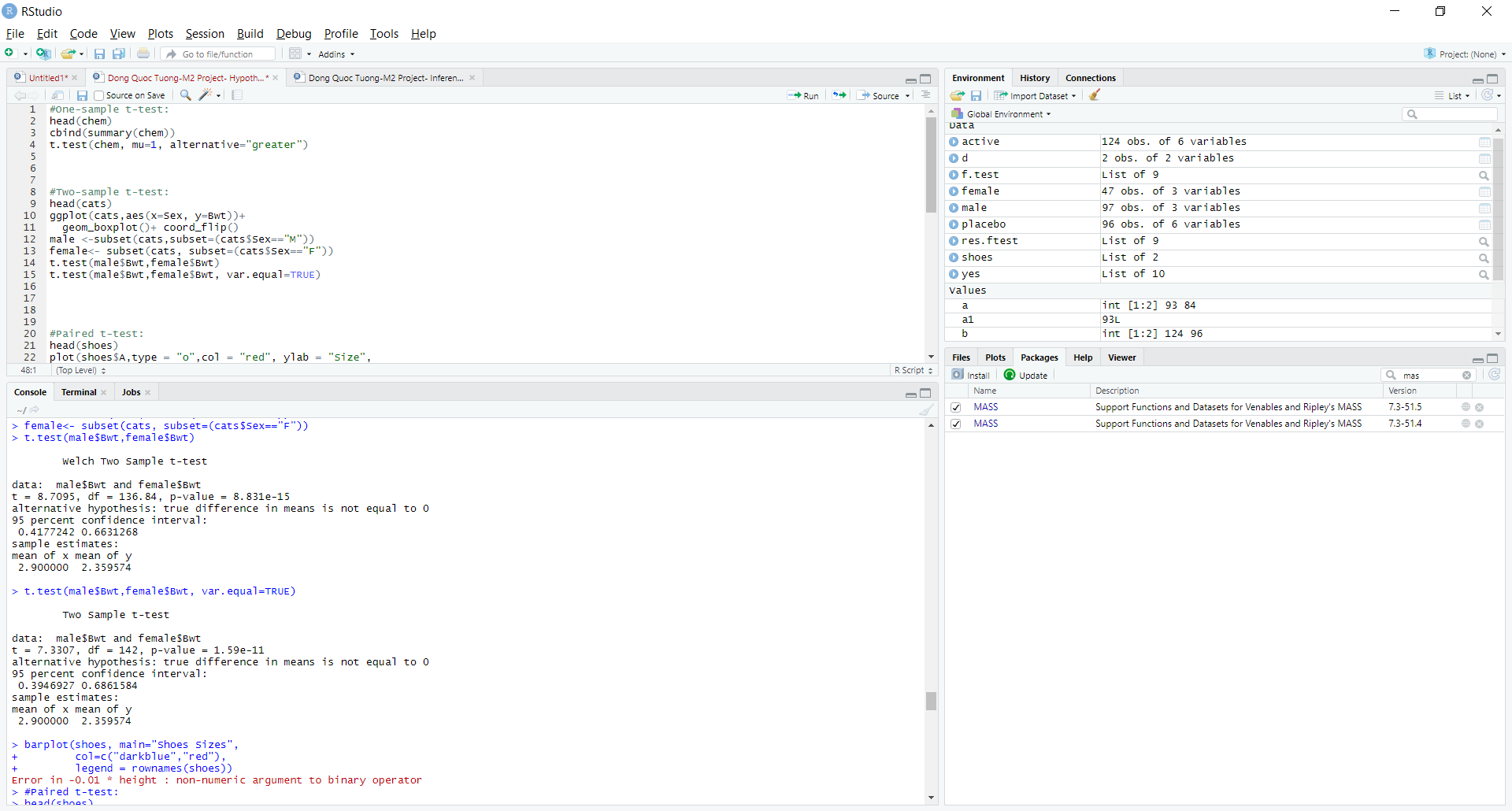
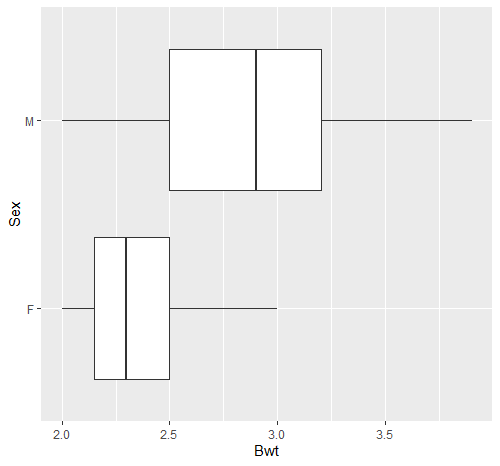
**Part B**

The dataset chem is dataset of the Chemical process experiment with 16 observations divided into 4 variables. We set the null hypothesis as the flour created by the production company has lower than 1 part per million copper in it. Since the data is already per million, we can set mu at 1 and put the alternative as “greater”. Since the null hypothesis is 0.002 which is less than the significant level of 0.05, we can conclude that the flour has greater than 1 part per million copper in it. IF we look at summary number, we see that the minimum is 2 part per million, and the maximum is 28. Therefore, our prediction is correct.

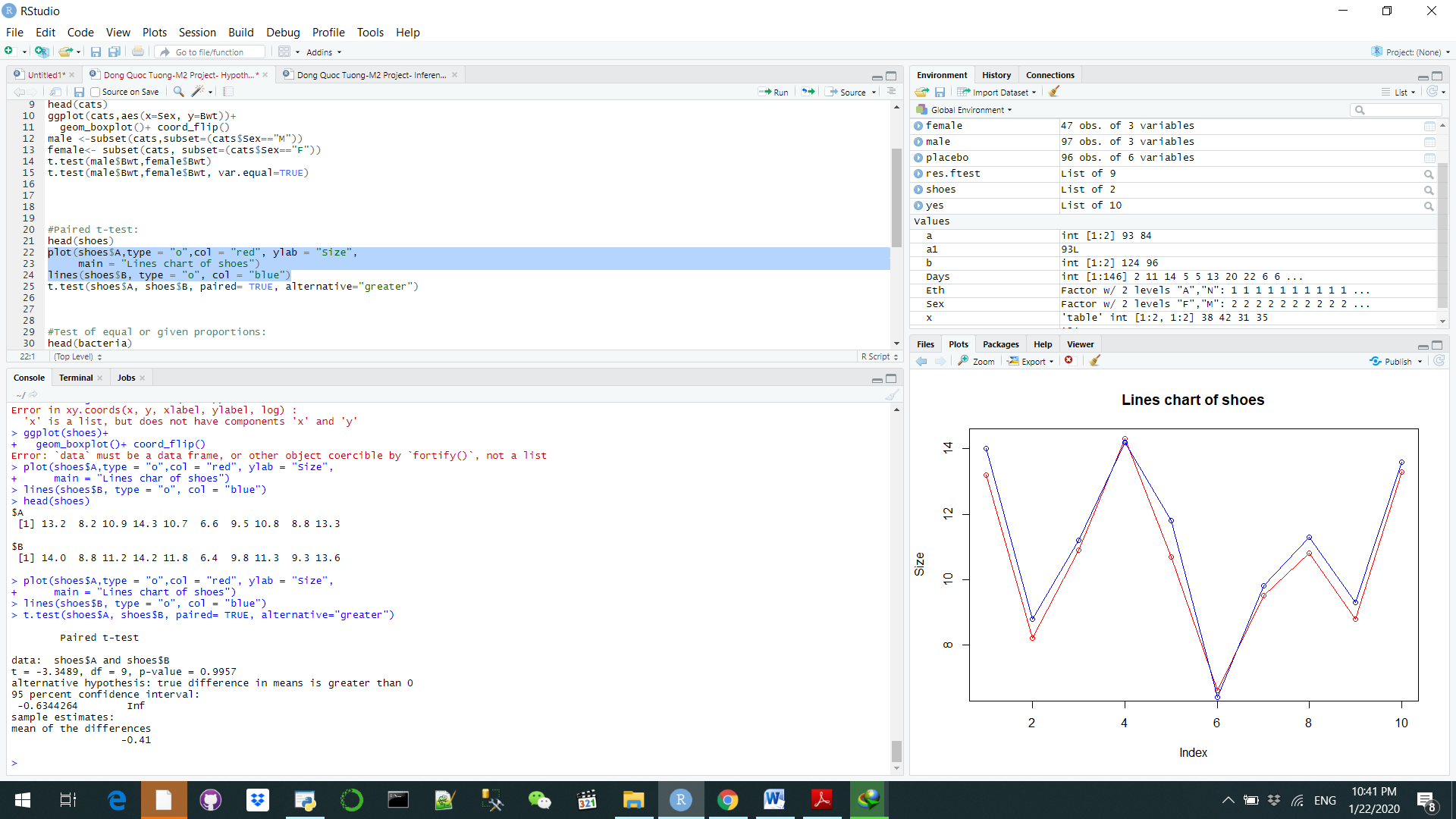
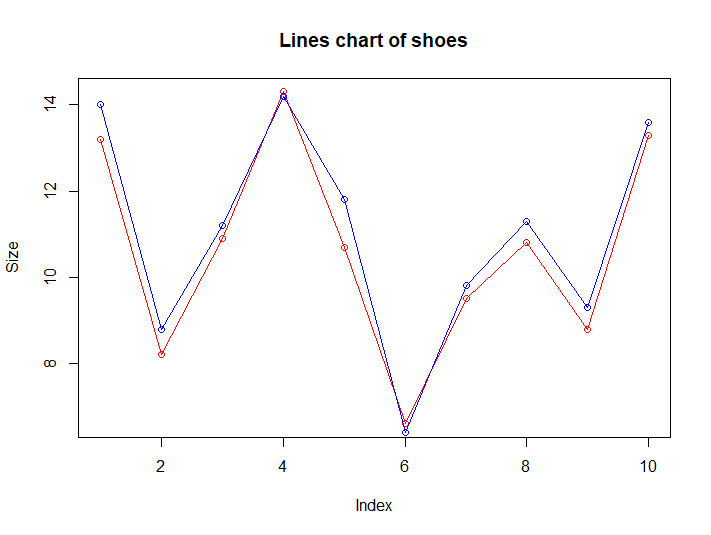
**Part C**

Next, we look at the dataset Cats and measure their body weight and their heart weight. Cats has 97 entries with columns. We will do the two sample t-test to see if the male and female cats share similarity in body weights. Despite the fact that we did two tests: variances are not equal and then equal, the results are similar, which is substantially lower than the significant level of 5%. Meaning that the mean of these two sample are different from each other. Thus, we can conclude that with 95% confidence that genders affect the body weights of the cats. When we look at the boxplot, we can see that the amount of male cats is greater than the amount of female ones according to our prediction.



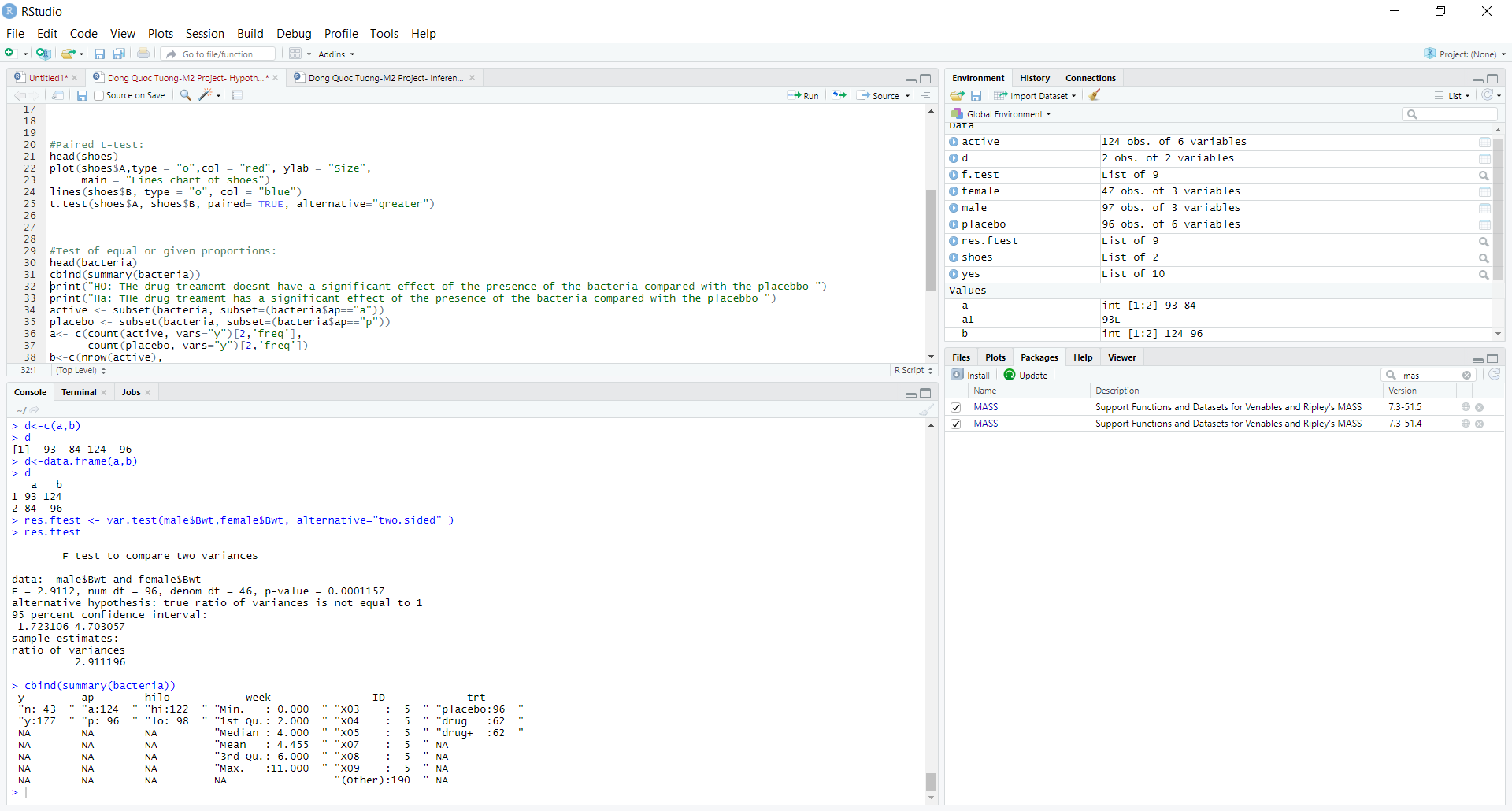
**Part D**

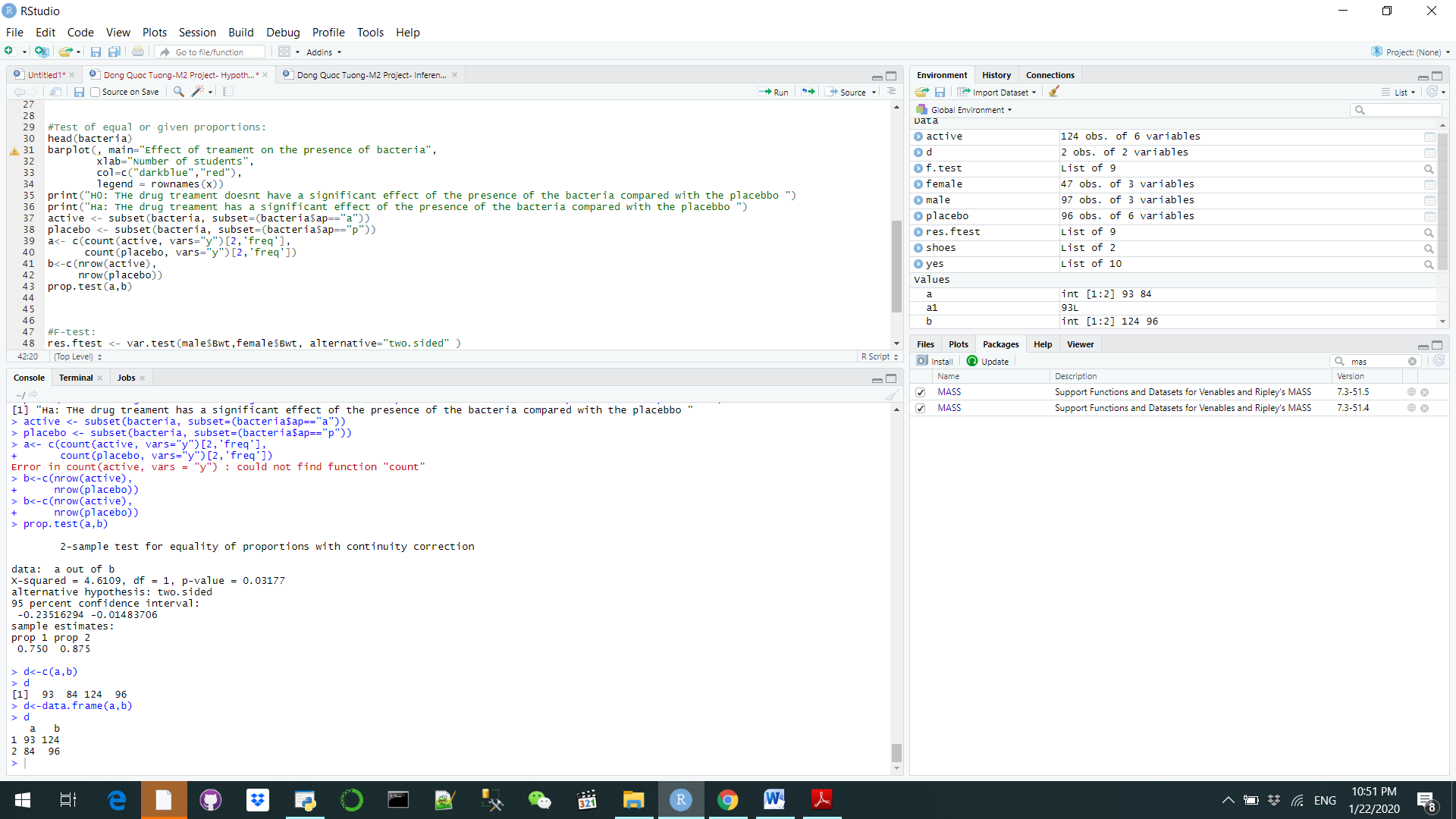
Now, we are looking at the dataset of shoes for boys. Since it is the same types of shoes, but measured in different material. After conducting the research, the P-value yields as 0.99 which is significantly higher than 0.05, we cannot reject the null hypothesis and say there is no difference when children wear shoes that make from material A vs B. To make sure we do not commit Type 2 error, we illustrated the data in the line chart and saw that that the two lines almost overlap each other so our estimation is correct.



**Part E**

The next dataset is about the presence of bacteria H. influenzae in children with otitis media in the Northern Territory of Australia. There are 220 variables, divided into 6 variables. The columns of y, ap, hilo are all binary whereas week is numeric ad ID is unique. We put the analysis to use the test of equal or given proportions to see if the drug treatment or placebo has more significant impact on the bacteria’s presence. Since the P value is 0.03, we can conclude that drug treatment has far more positive impact on the patient than placebo. The researchers should develop more vaccines for this

. 



**Part F**

Finally, we will look at the dataset for the cats and their weight measurement again. This time, we are using the F test to see if the variance of one gender is greater than the other. Since the P- value is extremely small, it is safe to say that there is the male cats body weight vary greater than their females counterparts when it comes to bodies weight regardless of the fact that they have similar mean

