

**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**

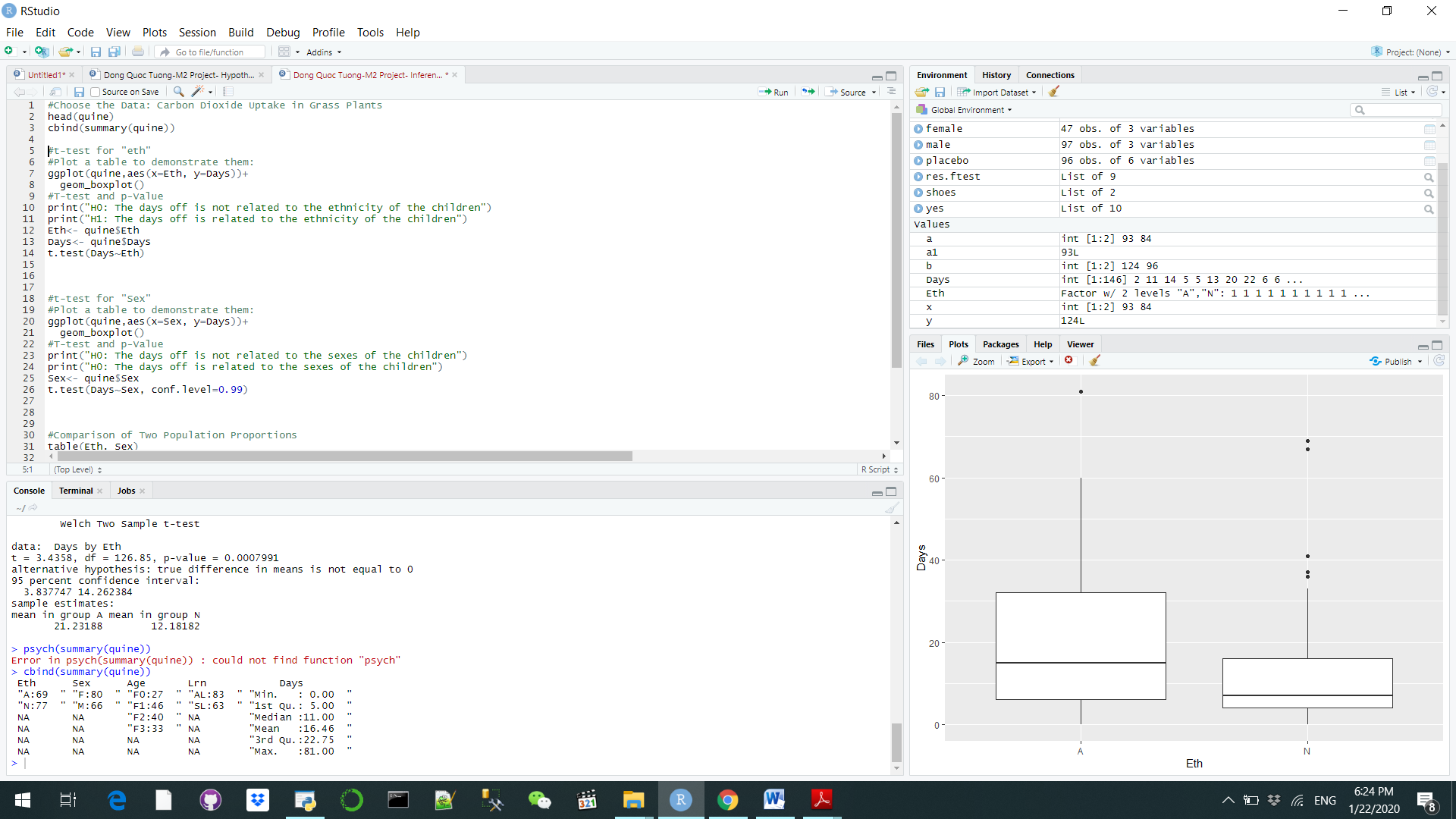
|  |
| --- |
| **Module 2: Inferential Statistics**  **Completion Date: January 22st Due Time:12:00am** |

**Statement of Authorship**

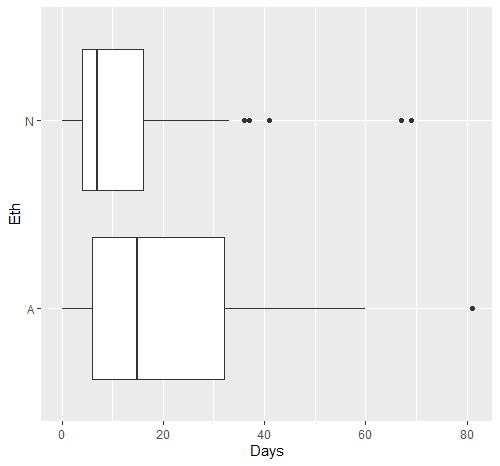
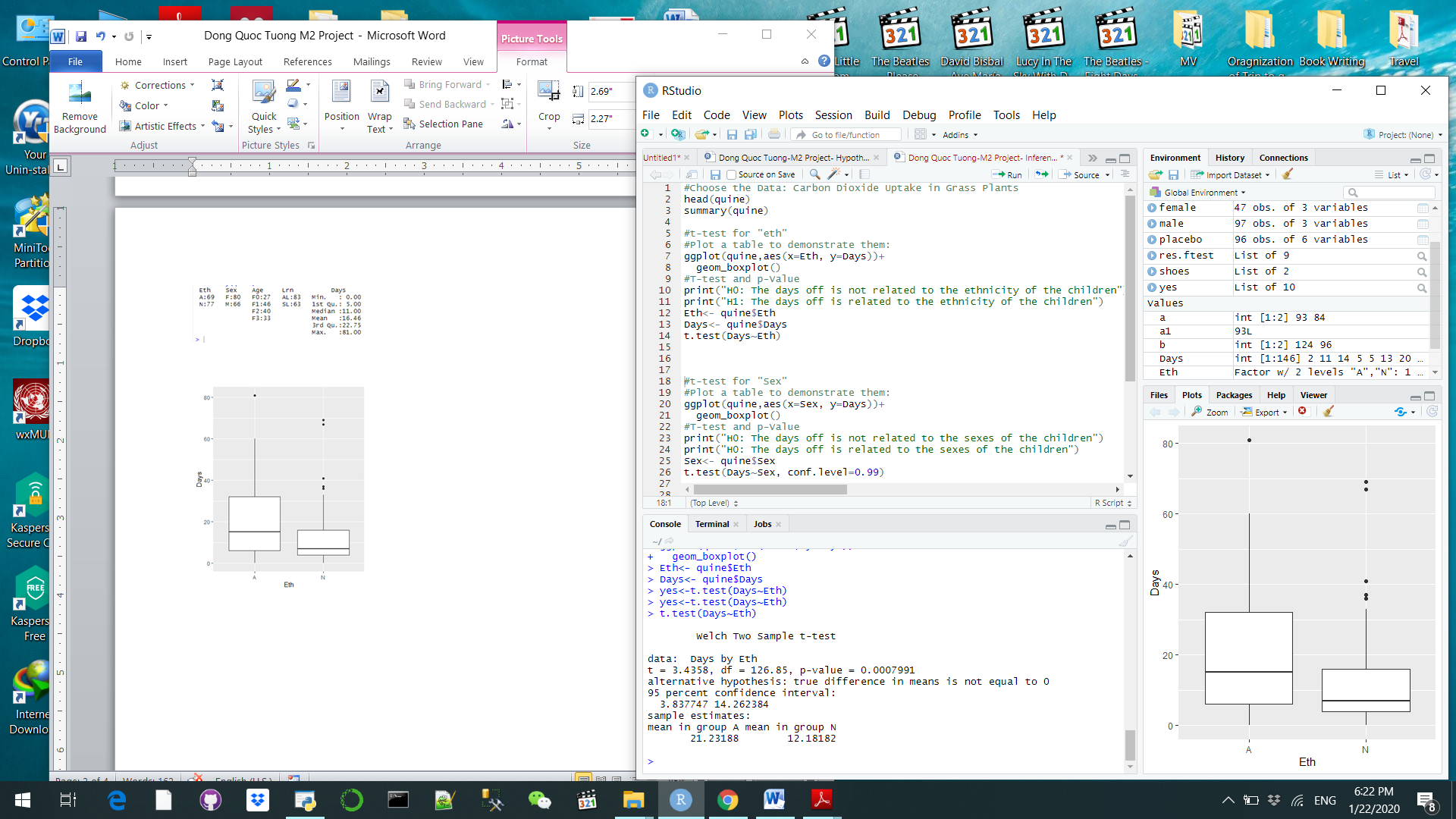
*I confirm that this work is my own. Additionally, I confirm that no part of this coursework, except where clearly quoted and referenced, has been copied from material belonging to any other person e.g. from a book, handout, another student. I am aware that it is a breach of Northeastern University’s 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

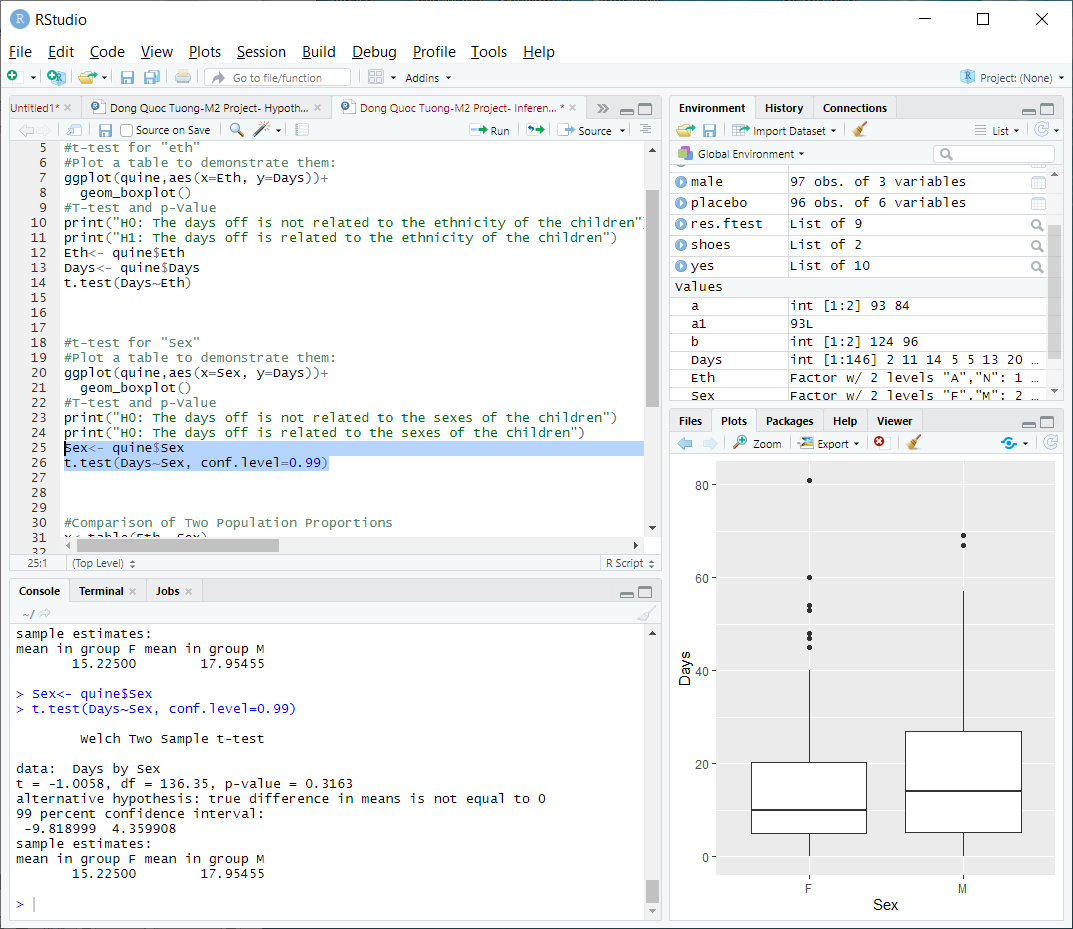
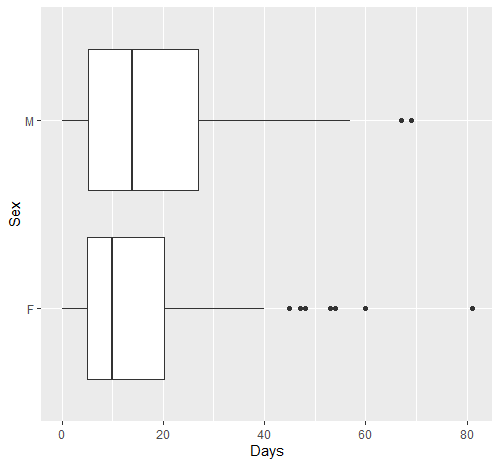
We chose the dataset “quine” from R, which gave readers a deep look into the demographic of students in Walgett, New South Wales, Australia. The dataset has 146 rows and 5 columns. It includes: Ethnicity, Age, Sex, Learner Status and Number of Absent days in a school year. Eth, Sex, Lrn are classified by binary data whereas Age is classified by ordinal and Days is by number. I am the principle of this school so I am trying to figure out are there any contributors to the total absent days of a student, given to the fact that students have an average 11 days absent annually.



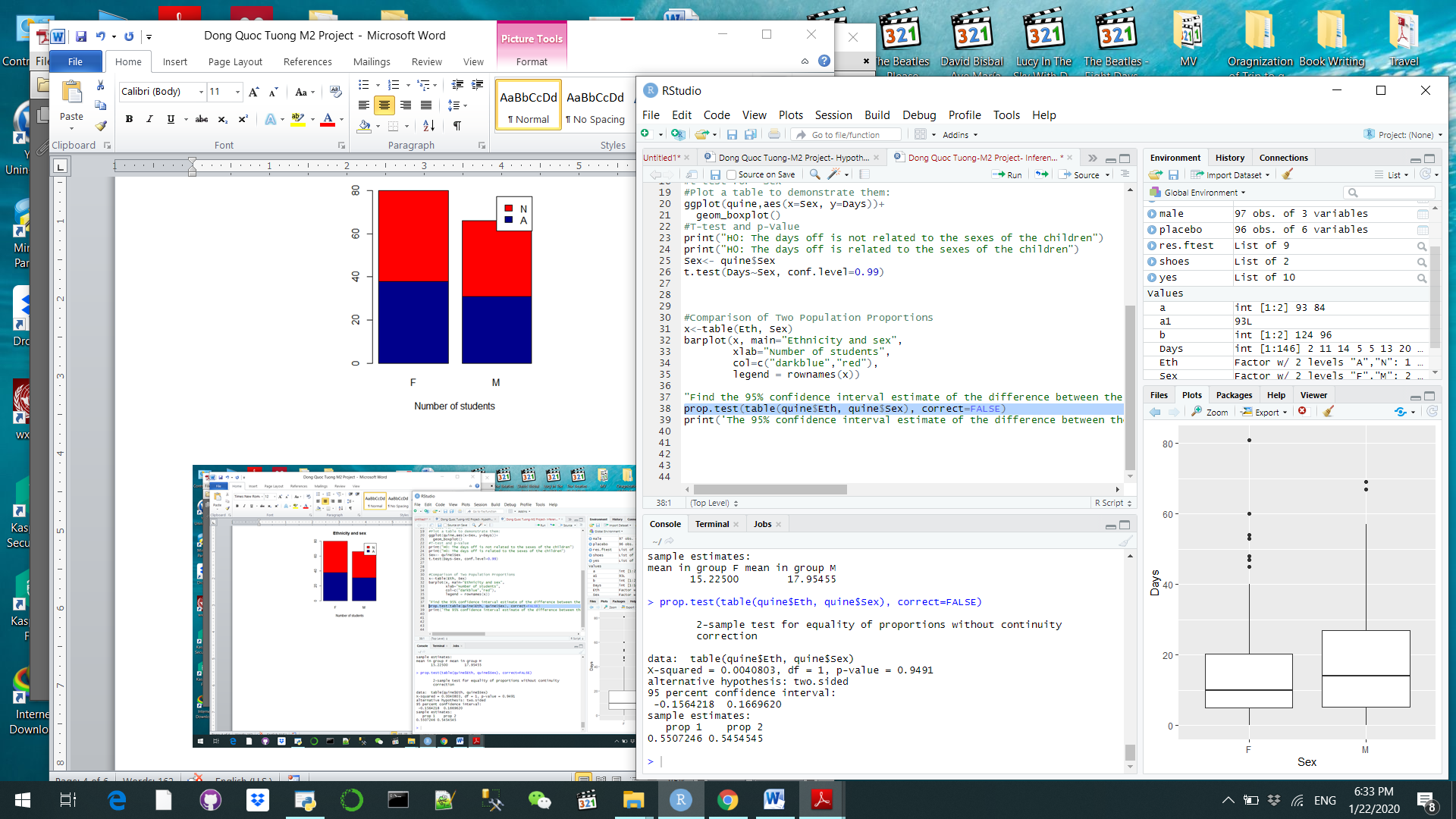
We set the null hypothesis that absents days are not related to the ethnicity of the children because we believe that the all of the students would have similar backgrounds and there is no difference in the absent days’ means. However, when we conducted the t test, we saw that the P-value is 0.0007, which is substantially lower than the significant level of 0.05. Therefore, we can reject the null hypothesis and say that ethnicity does play a big impact on the days of absent. From the box plot, we can see that students with Aboriginal origin skip classes more often than, 14 days on average compared to 8 days by their ethnic counterparts. Even the Aboriginal’s outliner is noticeably bigger. I believe that it is because the Aboriginal students’ lives are not as good as the national standard, thus consequently prevents them from going to class(“Australia failing to improve Aboriginal lives,” 2017). The school can do check on the background check on these students to see if there is any way to ensure their likelihood.

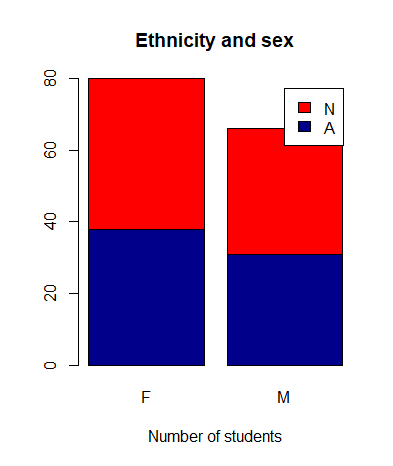


Next, we will test the likelihood that students go to school regularly regardless of sexes. When we conducted our analysis, we saw that the sample’s P value yielded 0.3163 with the confidence level of 99%. Ultimately, we cannot reject the null hypothesis and the number of absent days does not in fact, depend on the genders of the students. However, we see that the mean of the two groups are slightly different from each other (15 and 18 days). Thus, we use a boxplot to check what is going on and we see that the number of male students is significantly larger than the number of female ones. So this can be explained by the sample size that we have is not representative enough because if we have more girls, mean of days off could potentially equals to each other. But I would do more analysis as a principle to see if it is genders do not play a role in this or not



Now we wonder if that there might be a link between the Ethnicity and the Sex of the students. If the majority of Aboriginal students are males, it could contribute to the fact that Aboriginal students are more likely to have absent days like we saw below. Thus, it is best to conduct a test of equality of proportions without continuity correction test to see how likely male students are Aboriginal. P-value is 0.94 which is larger than the significant level. We can conclude that the ethnicity does not dictate the sexes of the students and they are approximately equally distributed in the graph bar chart below. It is likely that our suggestion above is correct and we need to have a look into it and solve this issue of racial injustice.





**References**

Australia failing to improve Aboriginal lives. (2017). Retrieved from https://www.aljazeera.com/news/2017/02/australia-failing-improve-aboriginal-lives-170214170102126.html