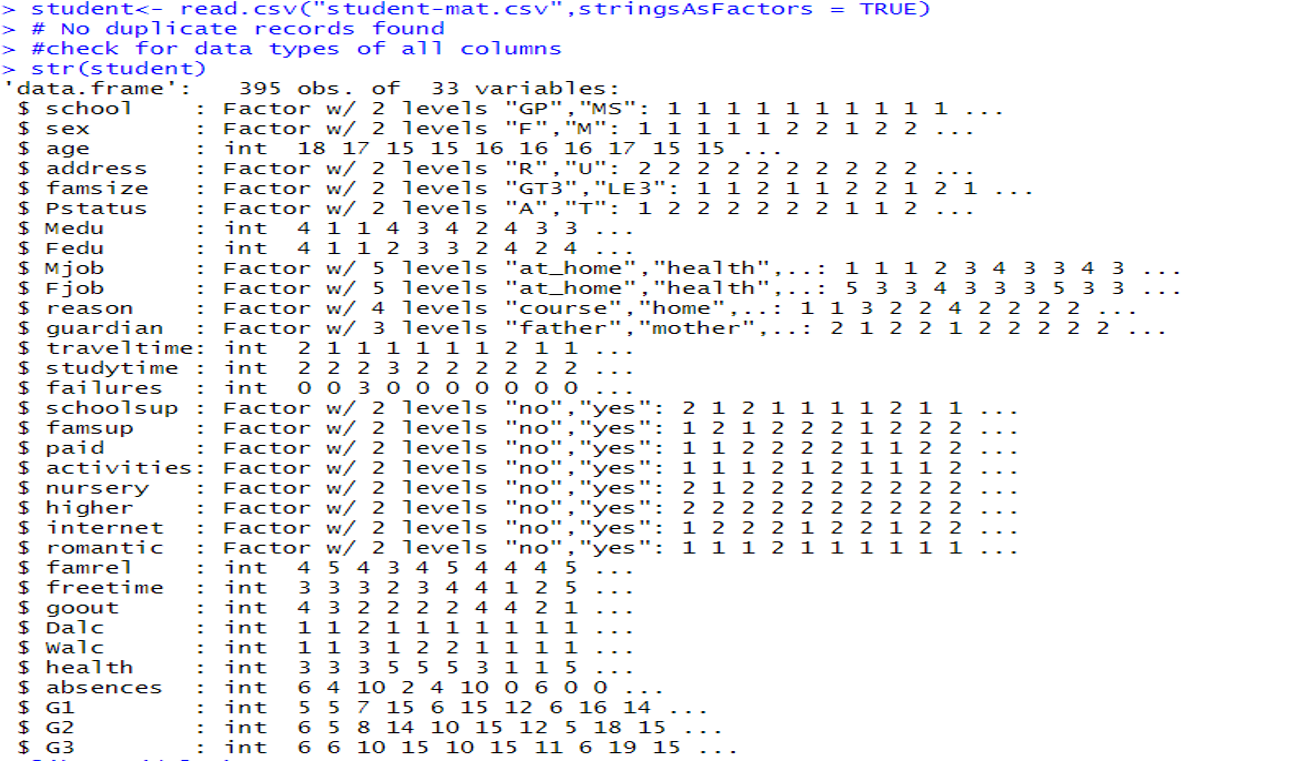
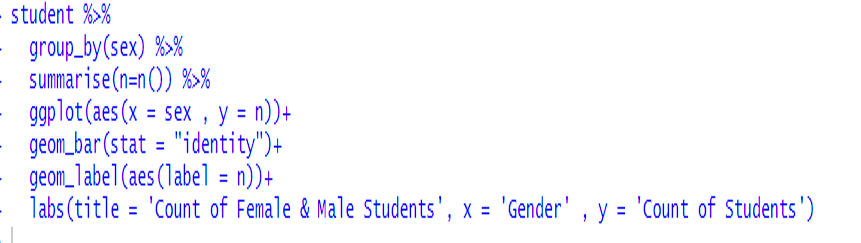


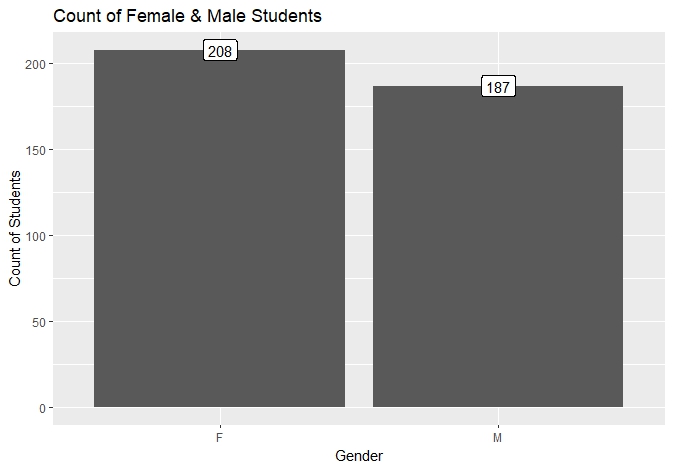
**Convert data types of the required variables**



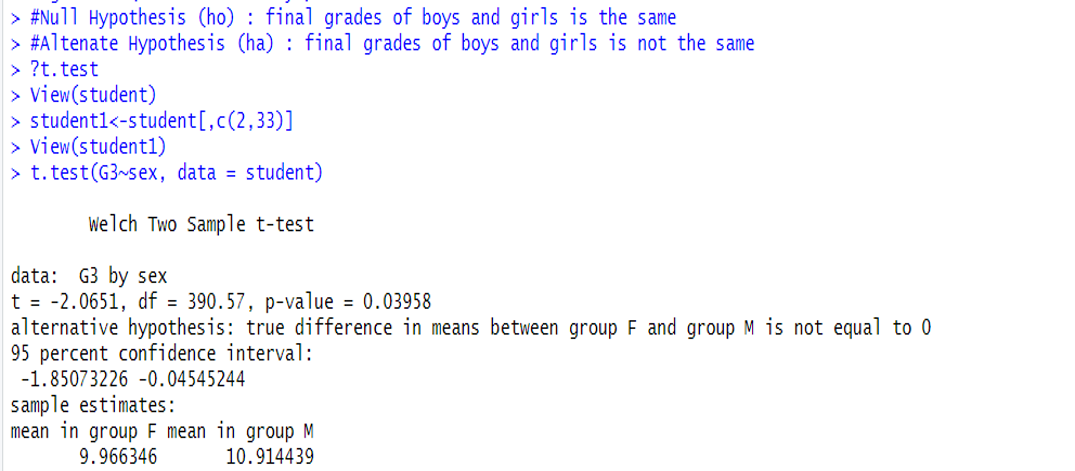
**Run libraries dplyr, ggplot2, tidyverse, tidyr**

**Find out the count of male vs female students**



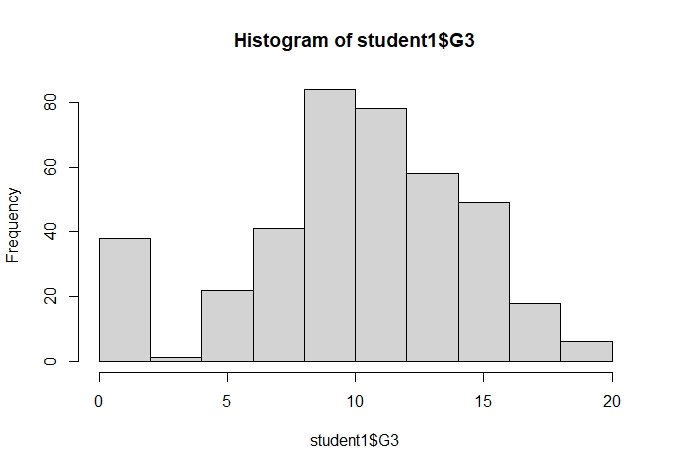


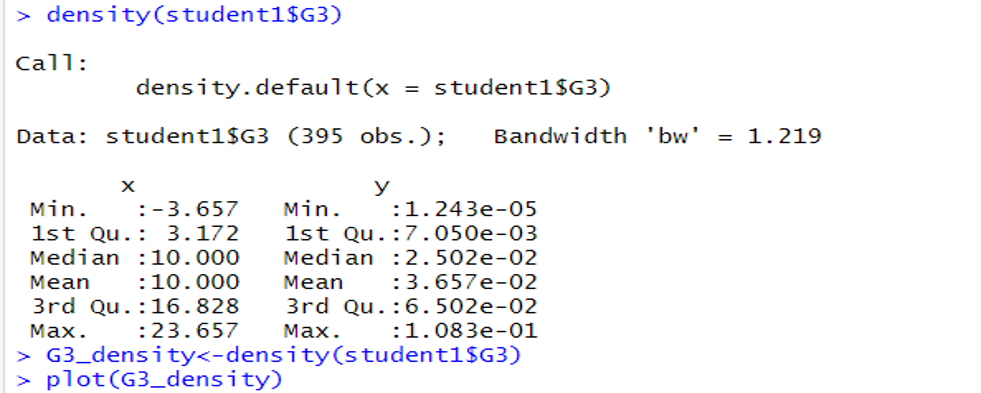
We keep only two columns namely 'Sex' and 'G3' and remove the other columns

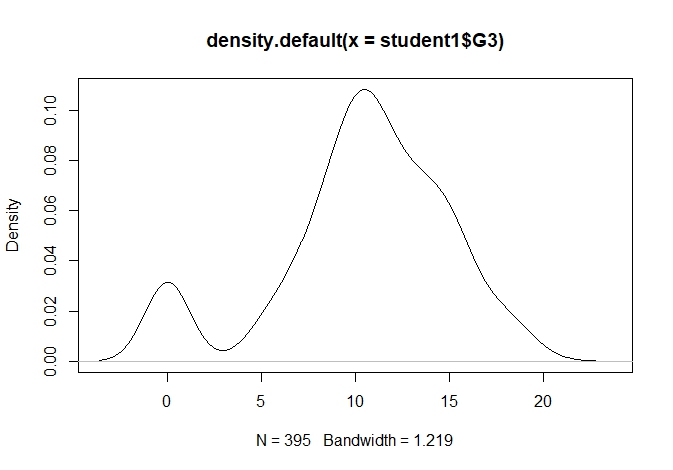
****

* t=-2.0651 indicates the distance from 0
* df = 390.57 is related to the sample size, how many free data points are available for making comparisons
* p value = 0.03958 is the probability value and indicates that we can reject the null hypothesis as it is less than that of alpha (0.05). Hence it is statisticall y significant.
* 95% confidence interval suggests that the true difference in means will lie between -1.85 and -0.04 (95% of time)
* We can see the difference in means between the two groups (10.91-9.96) = 0.95

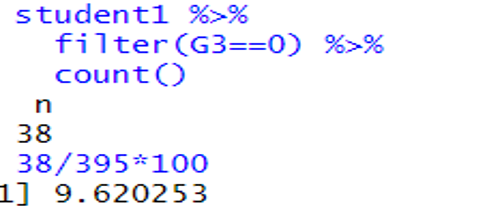
****

****

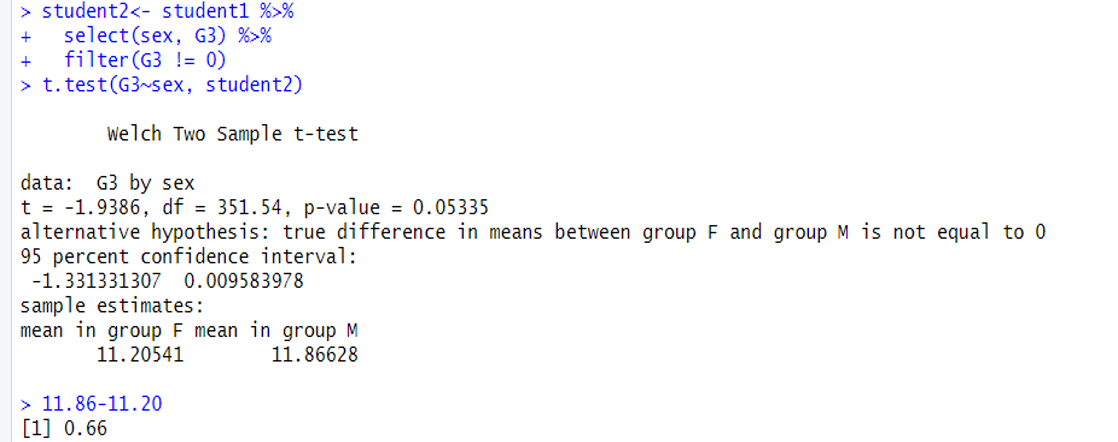
****

****

* Both the histogram and the density plot indicate that there are students who got 0. Could this be due to non attendance of exams. Let us find out the number of students who got 0.

****

* 38 students in total out of 395 have got a score of 0. That is 9.62% students.
* Let us check the mean for both groups by removing students who got zeros.
* We have created a new data frame called student 2 which includes a total of 357 students with no zero marks



* **Conclusion:**
* **mean of females is 11.20 and 11.86 of males. The difference in mean of the two groups is 0.66 as compared to the earlier mean difference of 0.95.**
* **P value is shown as 0.05335. For us to reject the null hypothesis the p value should be less than 0.05.**
* **Therefore, it is difficult to say if it is statistically significant.**