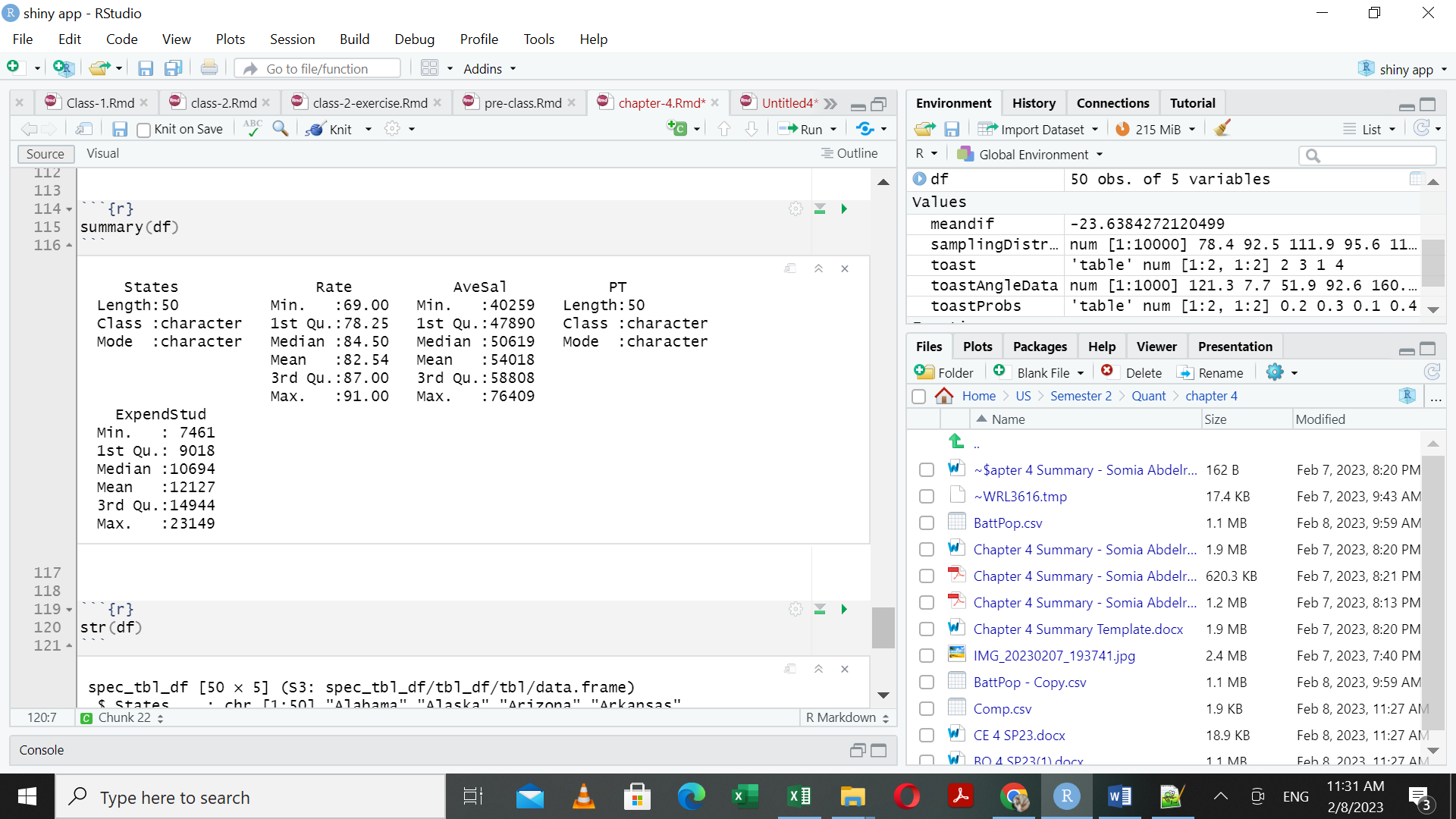
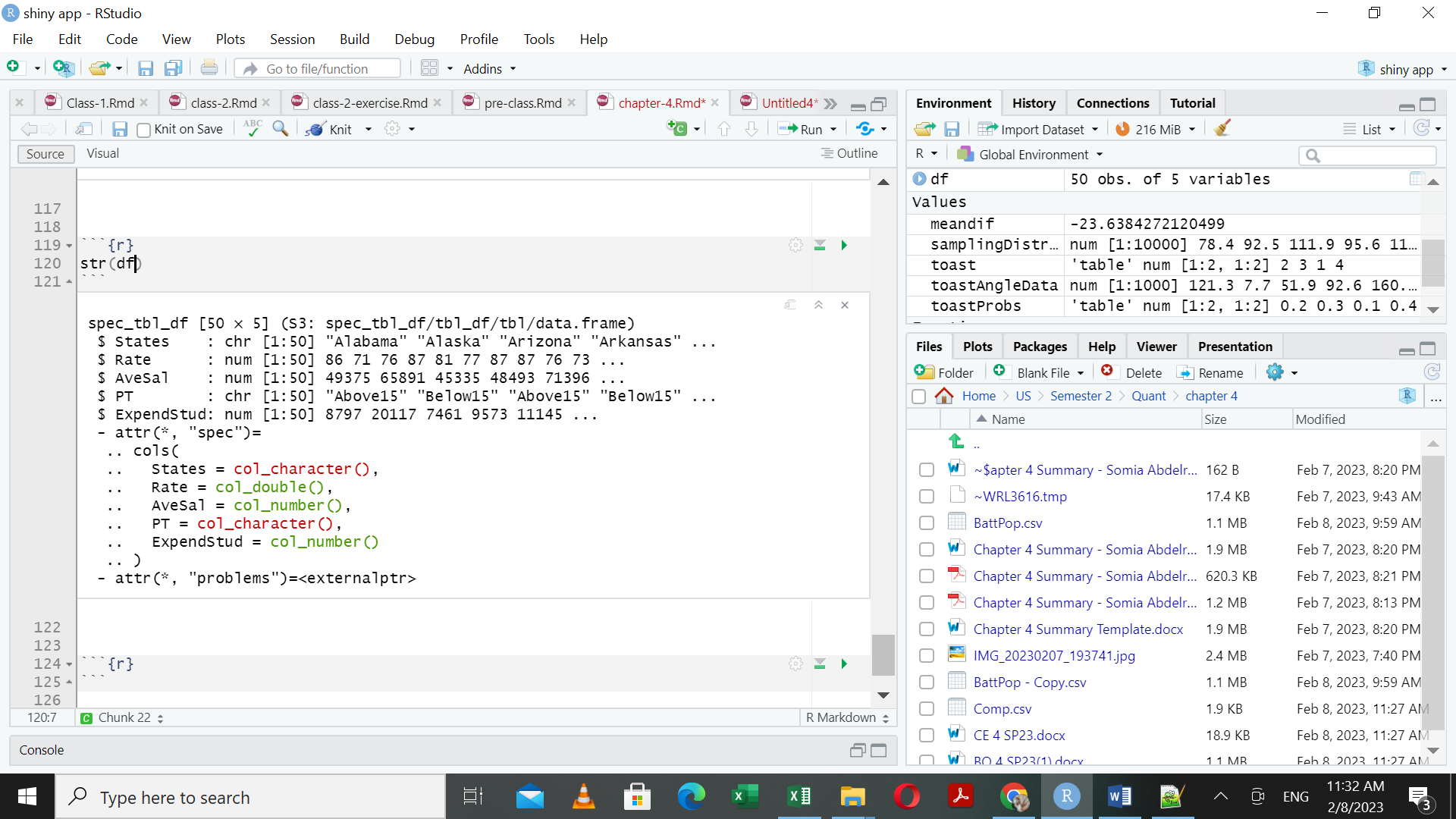
**IST772 Week 4 Class Exercise – Confidence Intervals**

**Instructions: Paste all code used, all results from R, and all summary statements in this document.**

This exercise provides an opportunity to work with data and develop a confidence interval. The provided data set includes data on High School Completion Rates for the 50 US States. It additionally includes variables that may be factors related to these completion rates. **The research question is whether the High School Completion Rates differ based on the Pupil/Teacher Ratio in the classroom.** The Pupil/Teacher Ratio variable is categorical and is coded as Below15 for classrooms with less than 15 pupils per 1 teacher and Above15 for classrooms with more than 15 pupils per 1 teacher.

1. Download the provided CSV file and read in the data using read csv() (from the readr package) or the Import Data dialog in R-Studio and save as a data frame. **Examine the data with str() and summary() and provide an overview of what you find in a comment.**



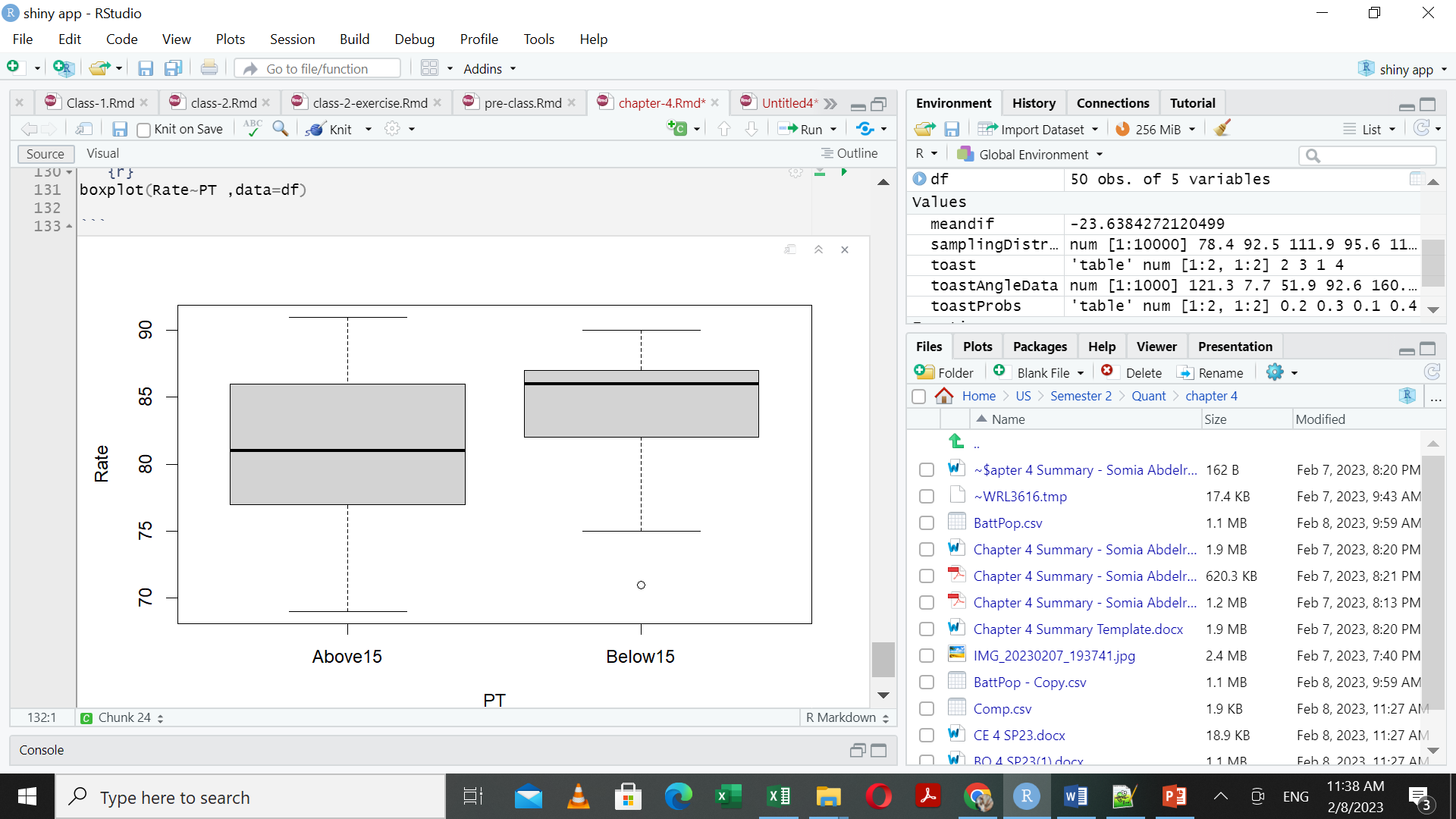


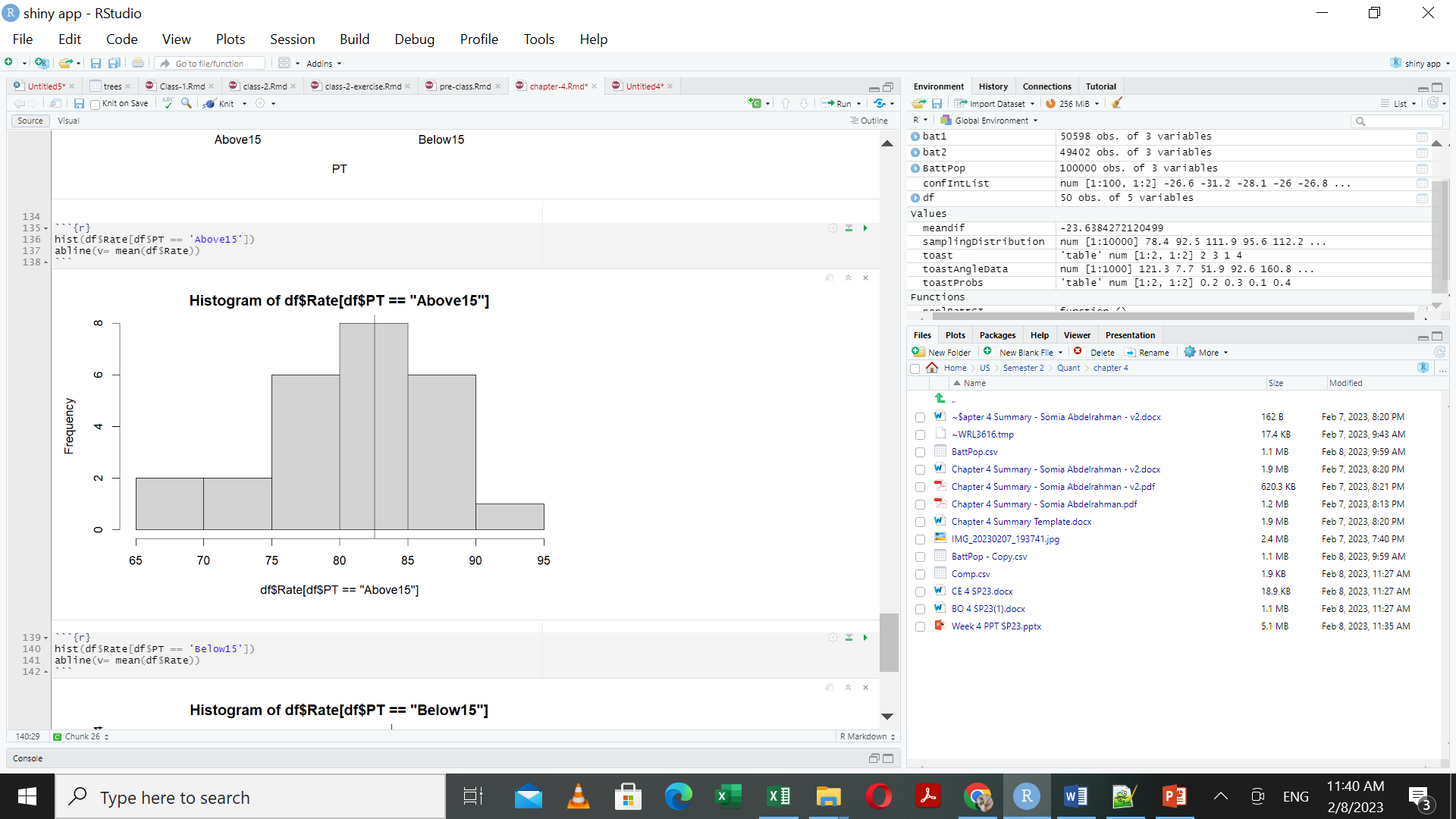
#Comment on the above:

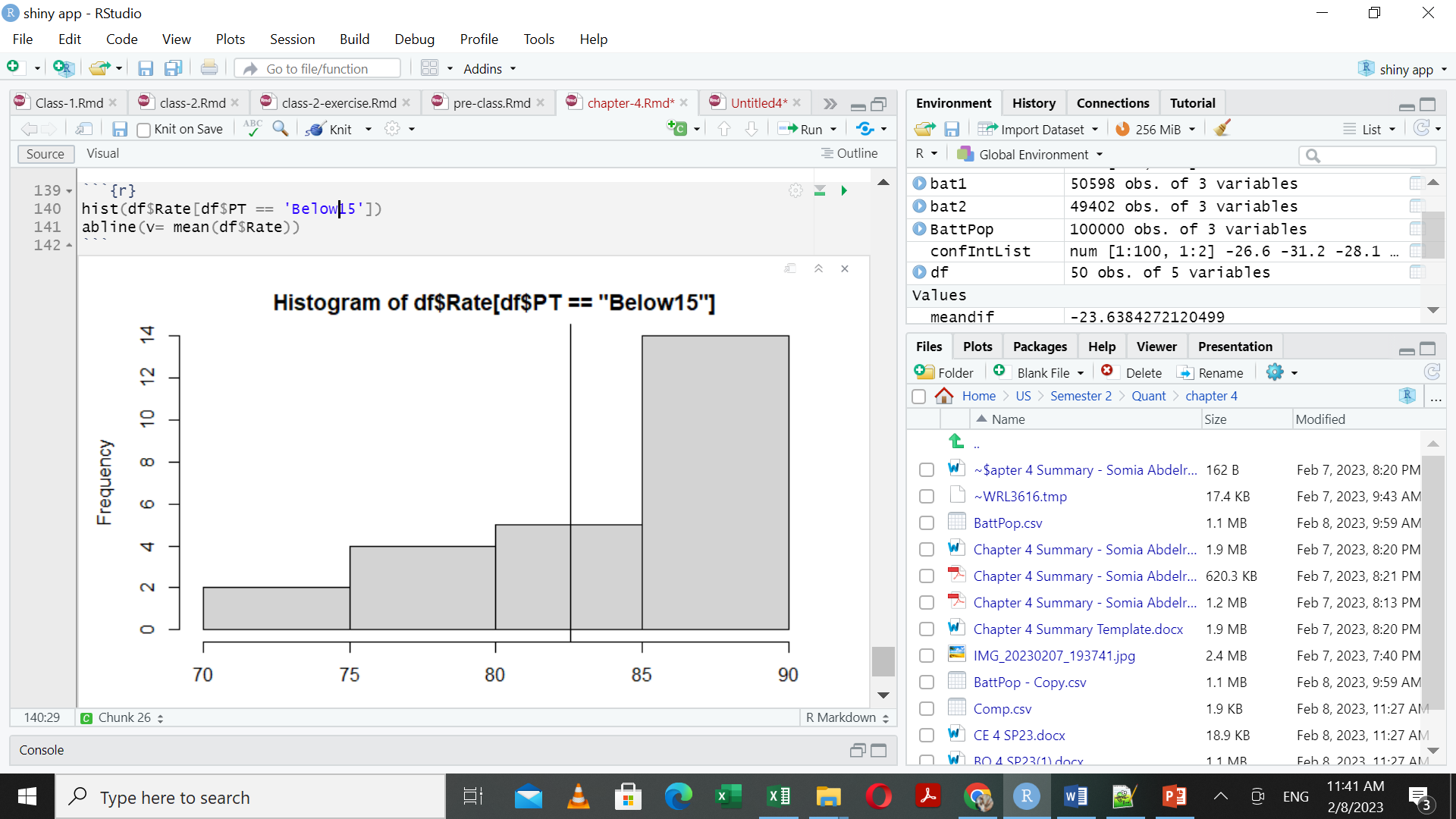
The dataset contains 5 variables and 50 rows. The Rate variable is numeric variable that describes the students’ completion ratio(Min = 68 and Max = 91) and PT is the Pupil to teacher ratio which a categorical attribute (above15 and below15) .

1. **Use graphical diagnostics to explore the data and report what you see in a comment.** Make sure to use histograms and a grouped boxplot to examine the distributions of completion rates based on the pupil/teacher ratios.

Note that boxplot() will take formula language as its first argument, like this: Rates ~ PT



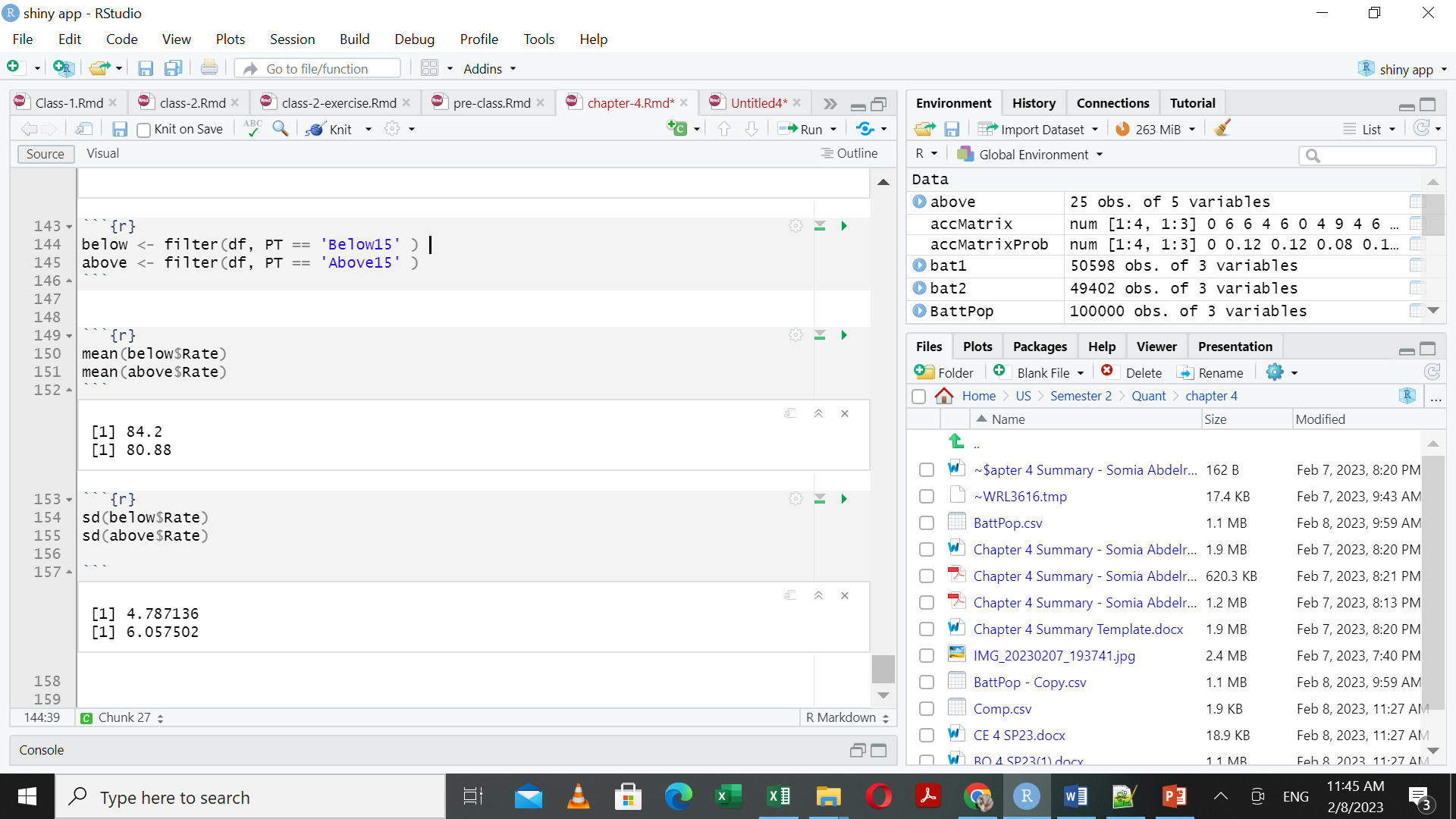




Comment:

From the graphs above, we observed that the median of the completion ratio of PT = below 15 is higher than the median of the completion ratio of PT = above 15

1. Find the mean and standard deviation of completion rates for both Pupil/Teacher ratios. **Add a comment indicating which mean is higher. Comment on the standard deviations for each subsample as well.**



#Comment:

The mean of the schools with below15 PT is 84.2

The mean of the schools with above15 PT is 80.88

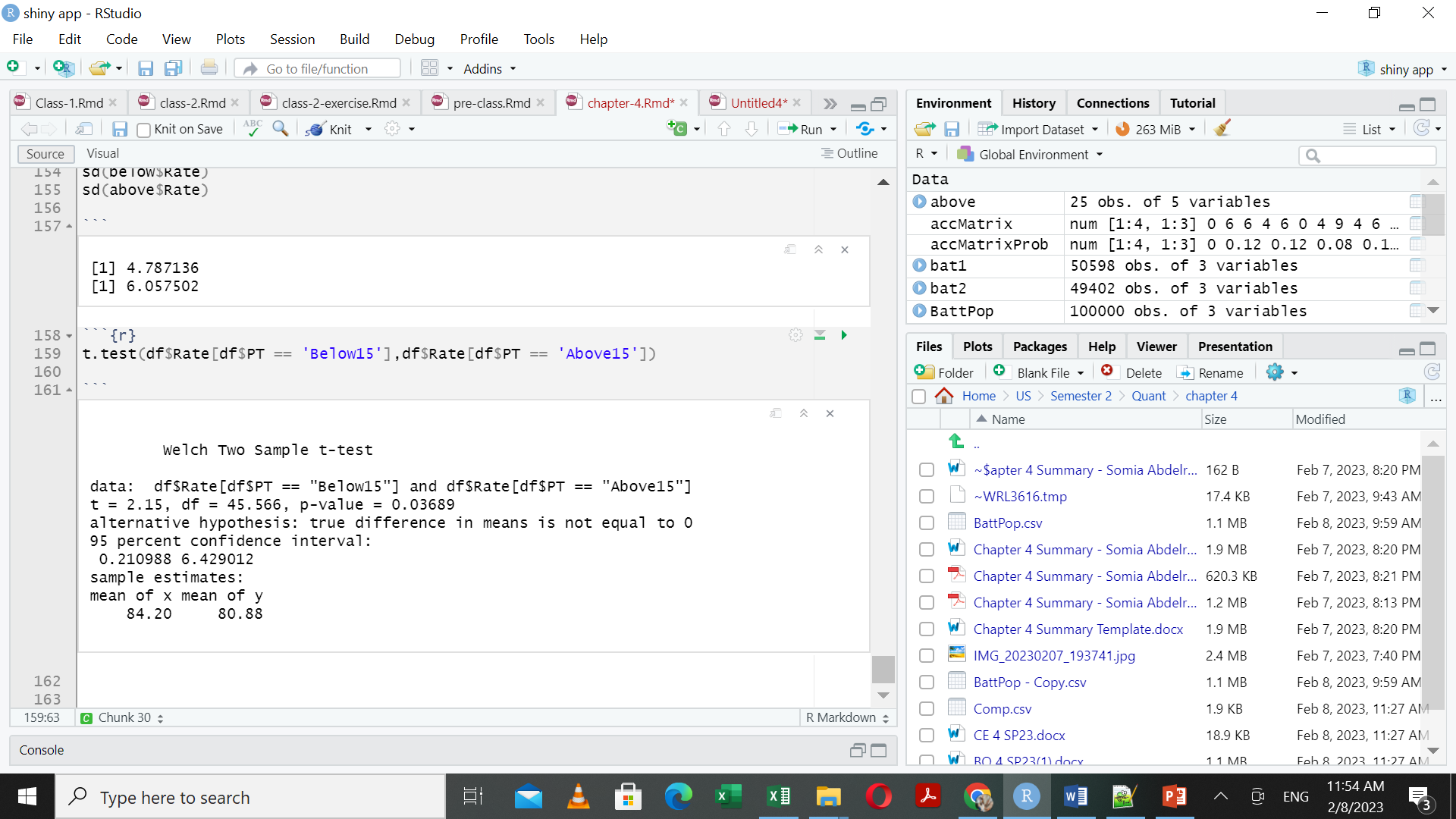
The standard deviation of the schools with below15 PT is 4.79

The standard deviation of the schools with above15 PT is 6.057

The school with PT below15 has mean of higher completion rate than the schools with above15 PT.

The below15 group has lower standard deviation than the above15 group

1. Use the t.test() command to create a confidence interval for the mean difference in completion rates between these two ratios. **Paste the results below and report the numeric values of the upper and lower bounds of the confidence interval.**



The CI is 0.210988 and 6.429012

Lower Bound: 0.210988

Upper Bound: 6.429012

1. Imagine you are now reporting these results back to the Federal Education Committee. The members of the committee generally have no difficulty with quantitative concepts, but they are not statisticians. They need thoughtful guidance on making sense of the results. In a short paragraph, answer the Research Question posed at the beginning for the committee. Explain what a confidence interval is and what you now know based on this confidence interval. Also explain what caveats, cautions, or warnings should accompany this confidence interval.

Research question:

**whether the High School Completion Rates differ based on the Pupil/Teacher Ratio in the classroom**

From the result obtained above, we are 95% confident that the classes with pupils to teacher ratio below 15 has higher completion rate than classes with pupils to teacher ratio above 15 with mean difference between 0.210988 and 6.429012.