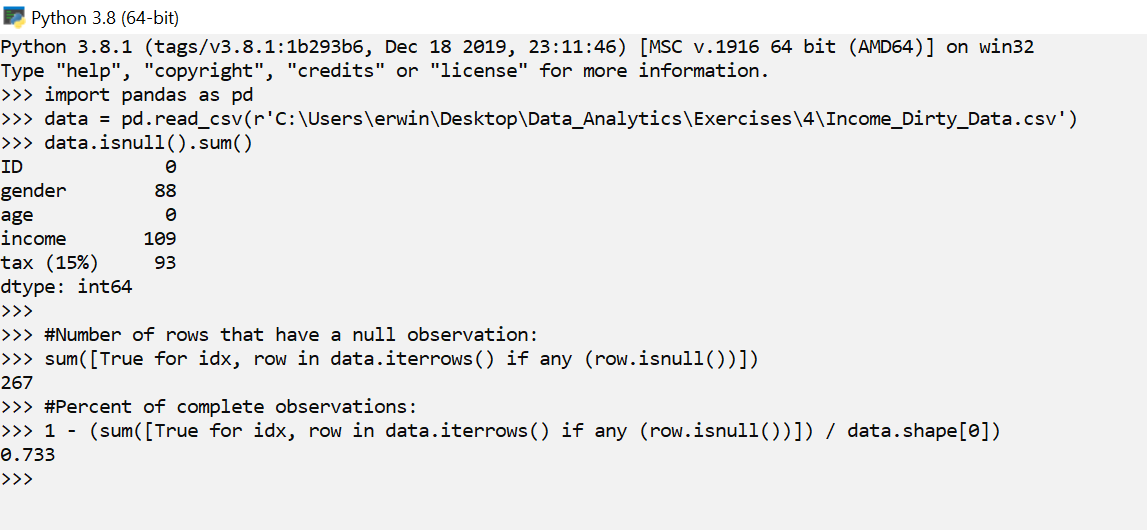
Michael Erwin

Data Analytics

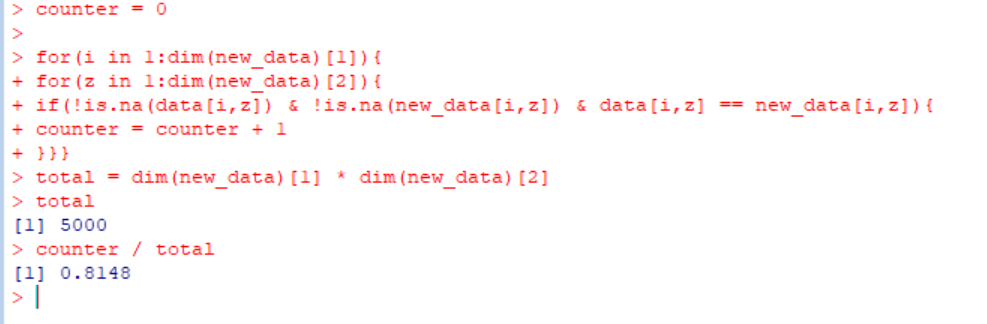
Exercise 4

Dr. Husari

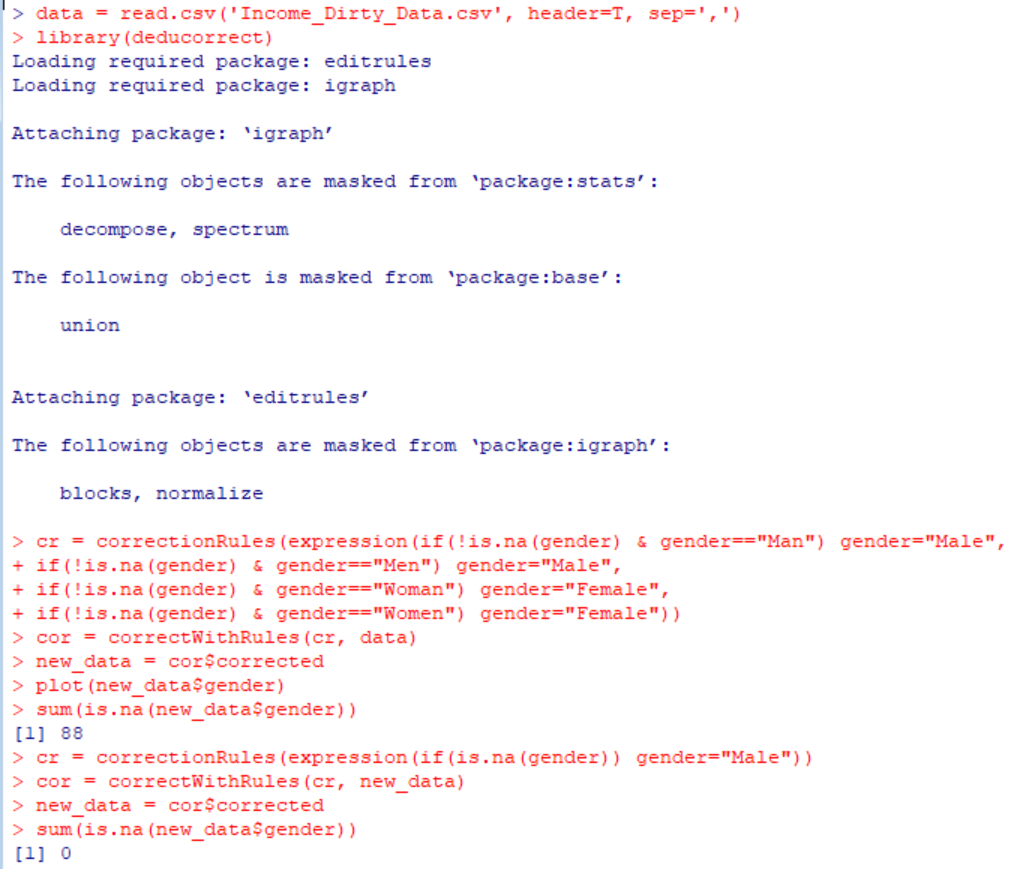
2/17/20



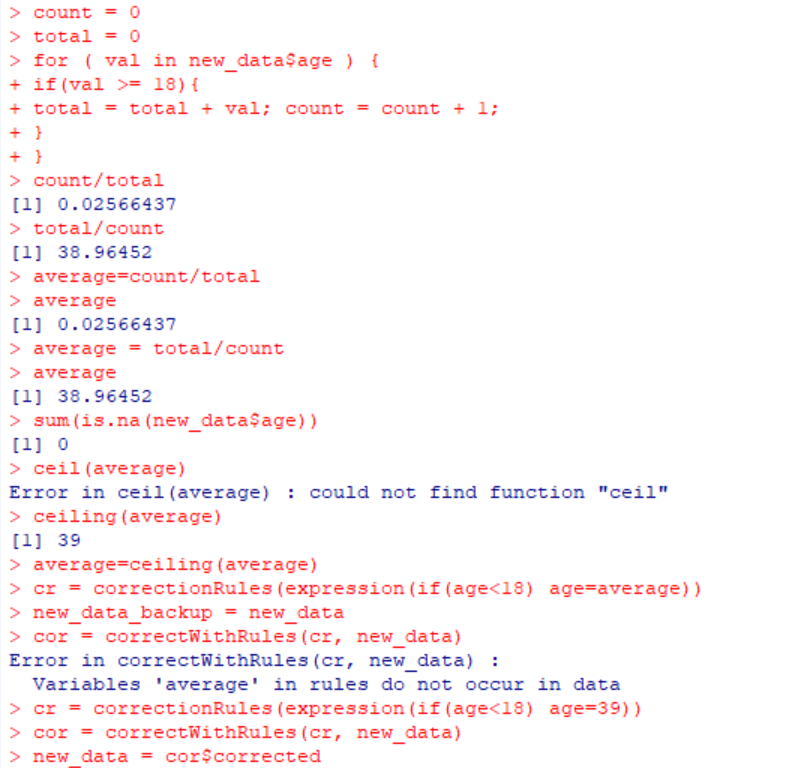
Opening the file in Excel reveals that it is a .csv file. The data has about 73.3% complete rows.



After finishing step 3, I returned to step 2 to see how many individual entries are the same between the original and final data frame.

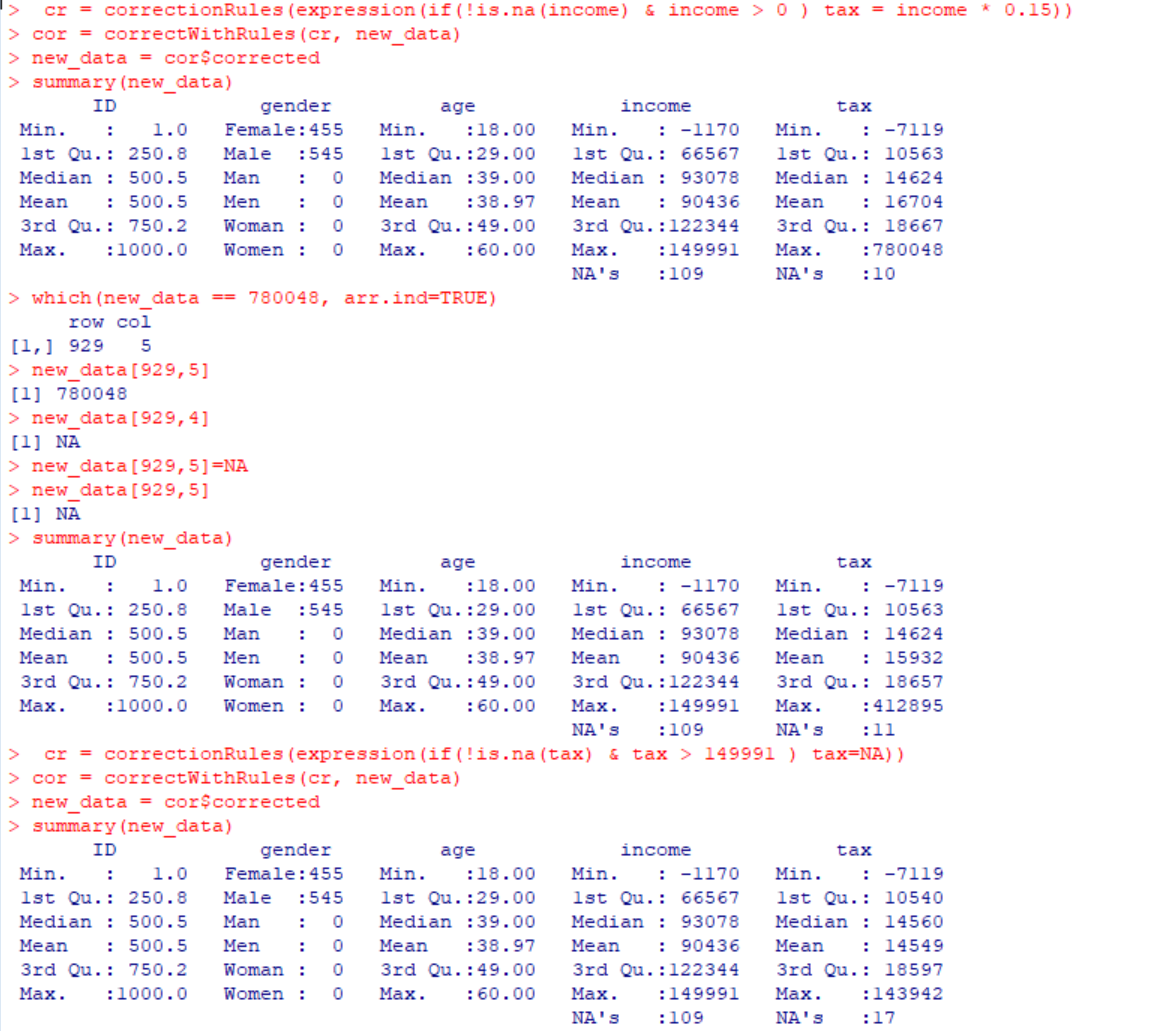


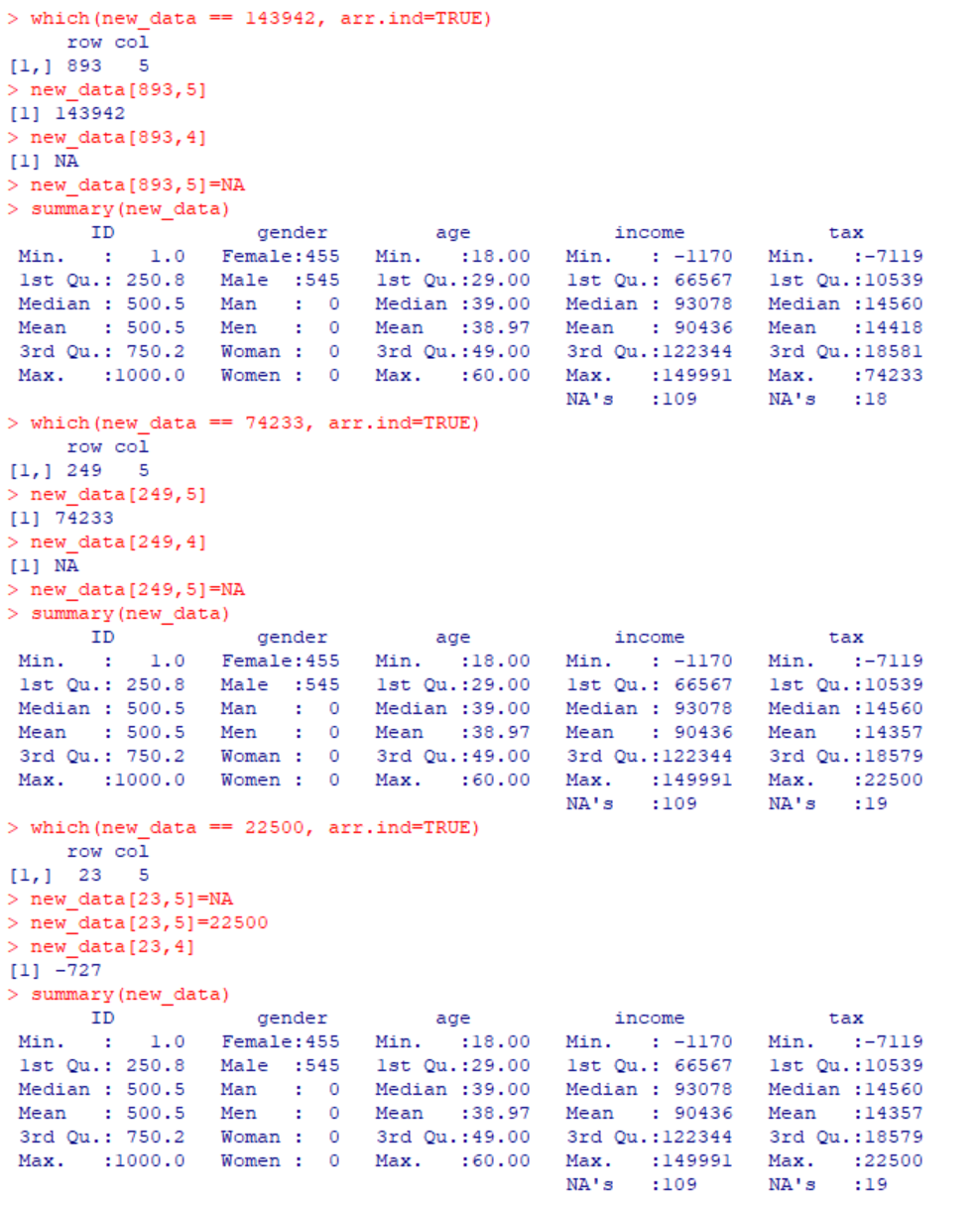
Here, the gender options are set to male and female, instead of having other options such as man, men, woman, and women.



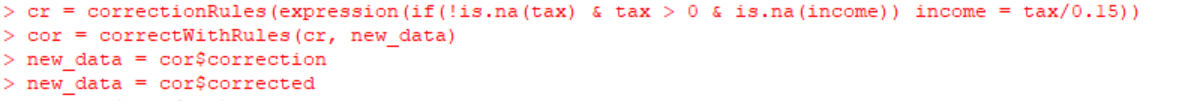
Here, we set invalid ages (ages below 18) to the average age calculated with the for loop.

Next, we set tax based on income. Then, we remove some outliers for tax.

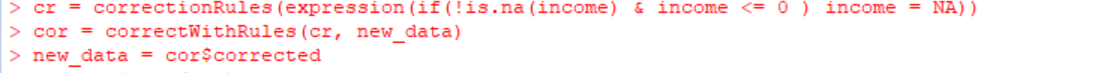




This sets income from tax if tax isn’t null and greater than 0:



This sets income to null if it is below or equal to 0:



Then I remove more outliers and view the summary:

