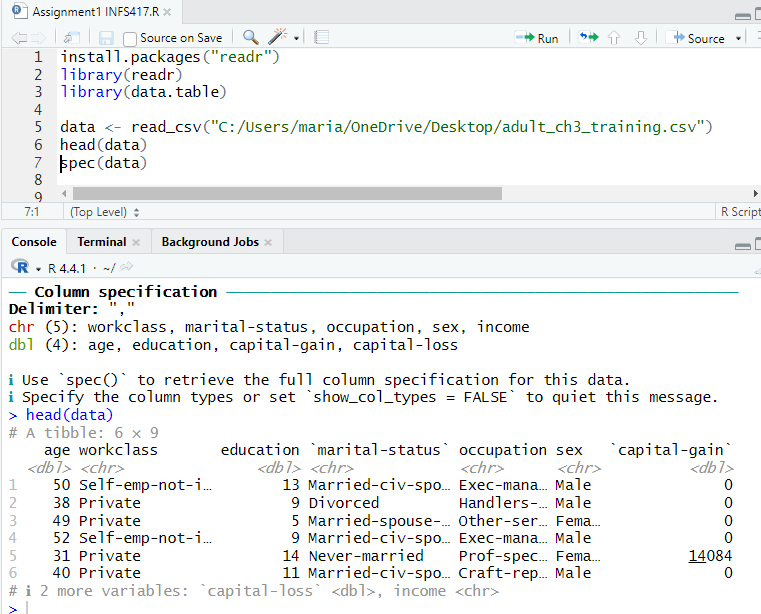
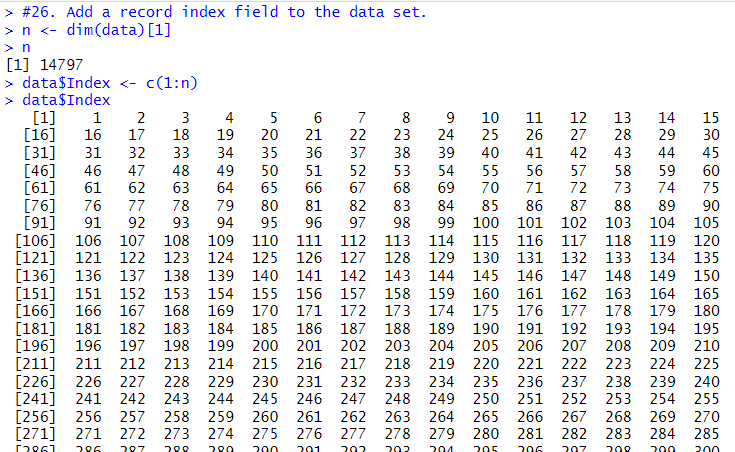
**INFS417**

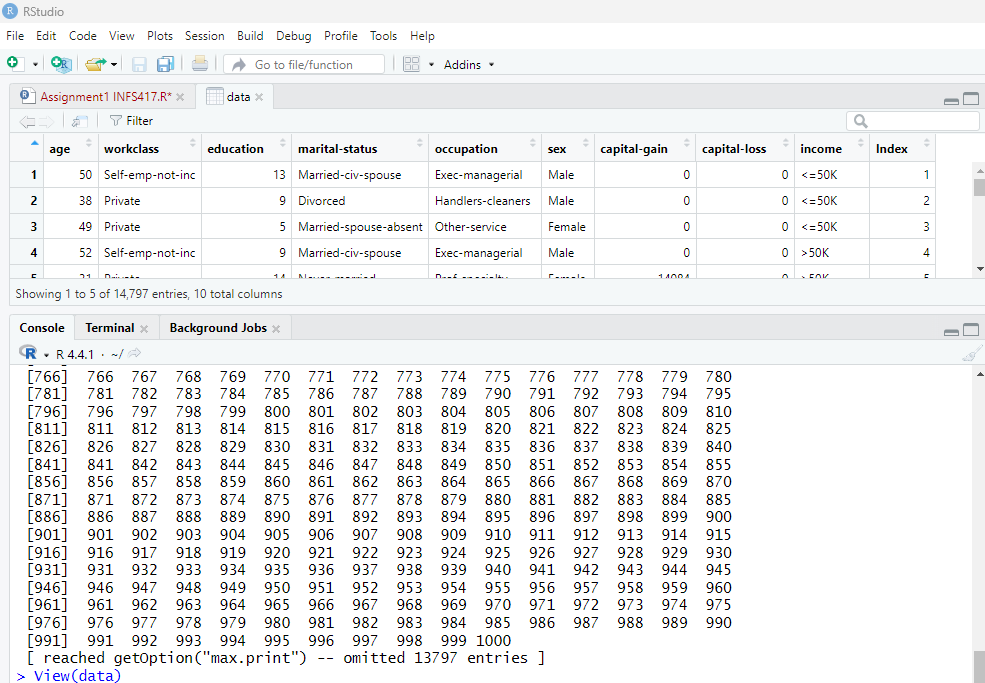
**CHAPTER 3-ASSIGNMENT 1**

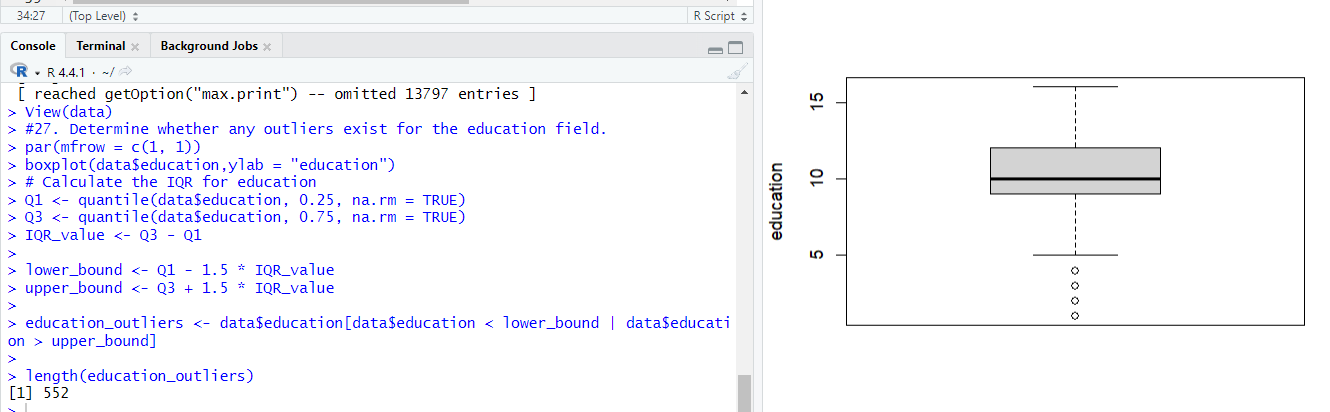
**STEP 1: INSTALLING LIBRARIES & IMPORTING DATA**



**STEP 2: CODE FOR QUESTIONS**

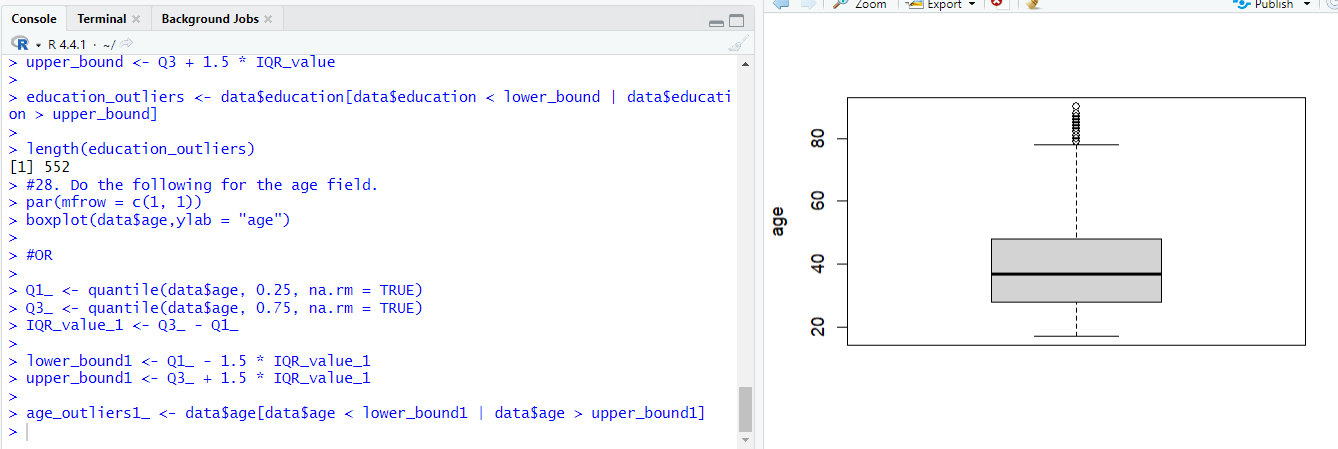
**26. Add a record index field to the data set.**

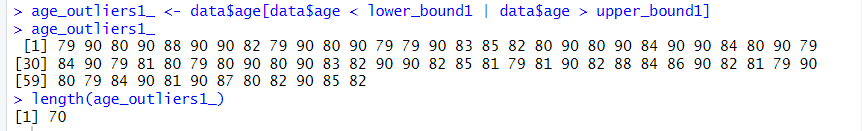


**27. Determine whether any outliers exist for the education field.**

***EXPLANATION:***

Yes, the outlier exists in the education column. There are 552 values in the data that need to be considered while proceeding.

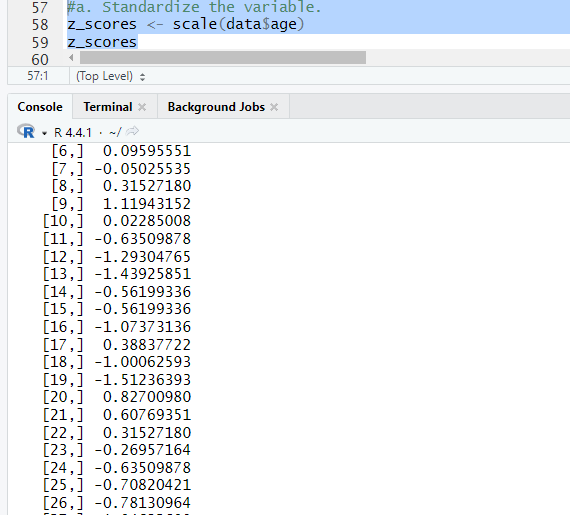
**28. Do the following for the age field.**

******

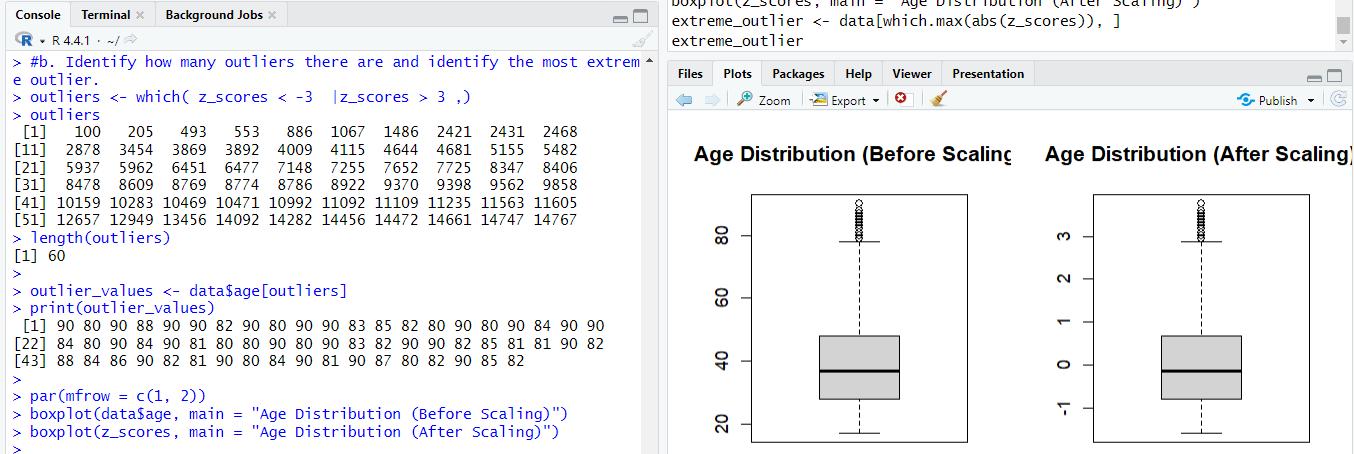
***EXPLANATION:***

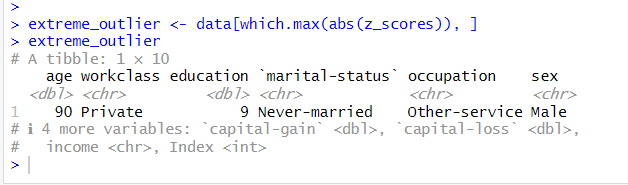
There are 70 values which acts as an outlier before scaling.

**a. Standardize the variable.**



**b. Identify how many outliers there are and identify the most extreme outlier.**

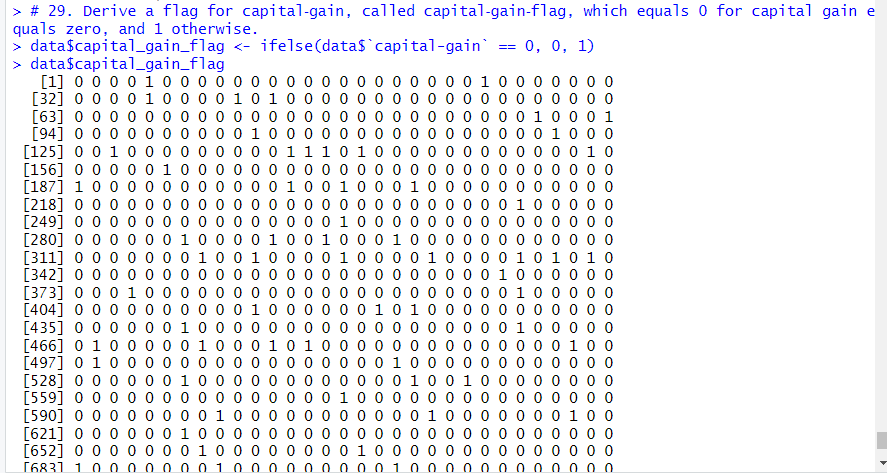




***EXPLANATION:***

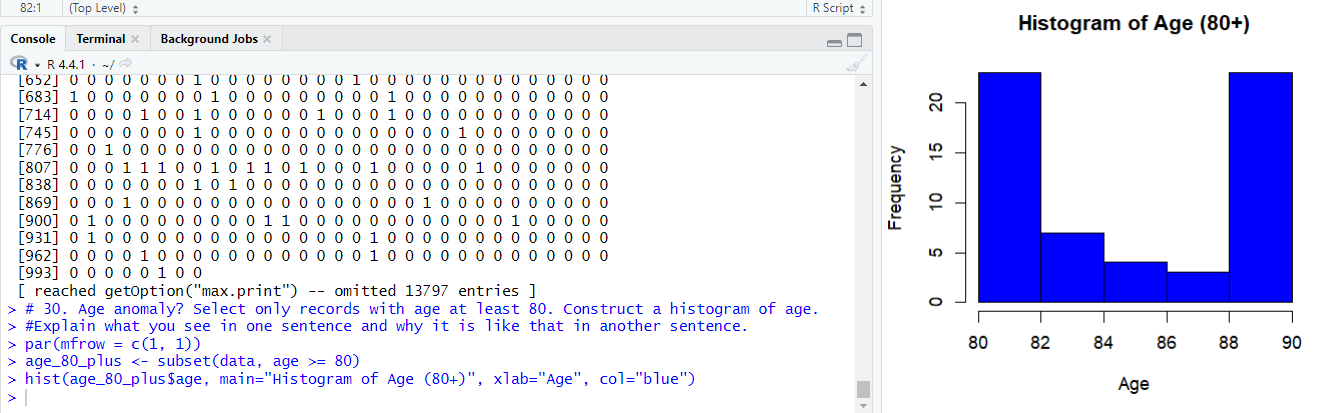
There are 60 values that are outliers and 1 row with extreme value i.e age 90.

**29. Derive a flag for capital‐gain, called capital‐gain‐flag, which equals 0 for capital gain equals zero, and 1 otherwise.**



**30. Age anomaly? Select only records with age at least 80. Construct a histogram of age.**

**Explain what you see in one sentence and why it is like that in another sentence.**



***EXPLANATION:***

This histogram illustrates the age distribution for individuals aged 80 and older.

**The Reason Behind It:** The distribution likely looks this way because there are fewer people in the older age brackets, as those over 80 represent a smaller segment of the overall population.