Questions

1. Why is this discriminant analysis of three different file formats of paleoclimatology data based off of 2 lakes in the Russian Federation important to perform?
    a) *To understand how to handle various file formats within R programming, data wrangling them, connecting them by a single column variable, merging them into a single dataset, it is then concluded to store the conjoined data into a database.*
2. Why is **Age** the chosen common column variable among the three different file formats?
    a) *The **Age** column was chosen due to the varying ages between the biogenic silica and sediments collected at Lake Baikal and Lake Elgygytgyn.*
3. Do you think **Depth** could have been a chosen common column variable?
    a) *No because the **Depth** column only pertains to the Baikal's data.*
4. Do you think **PANN** could have been a chosen common column variable?
    a) *No because the **PANN** column only pertains to the Lake's data.*
5. Do you think **Picea** could have been a chosen common column variable?
    a) *No because the **Picea** column is only contained with the Lake Elgygytgyn datasets.*
6. Why were no columns removed from any dataset?
    a) *Due to the fact that these datasets had no missing values but rather NA's, there was no point of excluding column variables.*
7. Is renaming the chosen common column variable important to structuring the data into a database?
    a) *Absolutely yes, how you put more than one dataset into a database is based off a common column variable name, thus, you are able to place the combined dataset into a database from the common column variable relationship among the datasets.*
8. Shouldn't the outliers be removed from the chosen common column variable?
    a) *Not necessarily, this discriminant analysis is based off of just performing EDA, then combining the different datasets into a single dataset and transferring into a database.*
9. Why were a txt, JSON, and xls file formats chosen to perform this discriminant analysis?
    a) *As CSV files are the go-to for performing data analysis, working in various other file formats is crucial to potential data scientists to learn.*
10. Do you think a database is efficient when studying the paleoclimatology in the Russian Federation?
    a) *Without a doubt yes, this discriminant analysis is basically working with other file formats that are seldom harnessed in data analysis, structured into a conjoined dataset, then placed into a database, and all this is based off the age of the biogenic silica and sediments collected at Lake Baikal and Lake Elgygytgyn.*