The code reads the data and performs preliminary checks, including data shape, missing values, repeated row statistics, and basic descriptive statistics. Next, we preprocess the data, rename some columns, and divide the data into 5 age groups (age\_groups) based on the quantiles of the "Age\_Ma" column, and calculate the time interval for each group. Next, the code analyzes the distribution of Max\_Diameter, Min\_Diameter, and elongation by drawing histograms, and uses the IQR (interquartile range) method to detect and remove outliers to improve data quality. After removing outliers, draw box plots and scatter plots to observe the trend of each feature over time. Finally, calculate the correlation matrix between Age\_Ma and each feature and use heatmap visualization to help identify the relationship between variables. I canceled the line chart because it did not show useful data information. I also tried to use resampling to extract data samples from each age group to draw new scatter plots, but no new findings.

Data Overview:

**TOP\_DEPTH** and **BOTTOM\_DEPTH** variables are **0** and **2 , respectively** , and the maximum values are **148** and **150** , respectively. The data distribution is relatively uniform, and the standard deviation is large, indicating that the sample depth varies greatly. The quartiles of the data show that the samples are mainly concentrated between **36 and 114** meters.

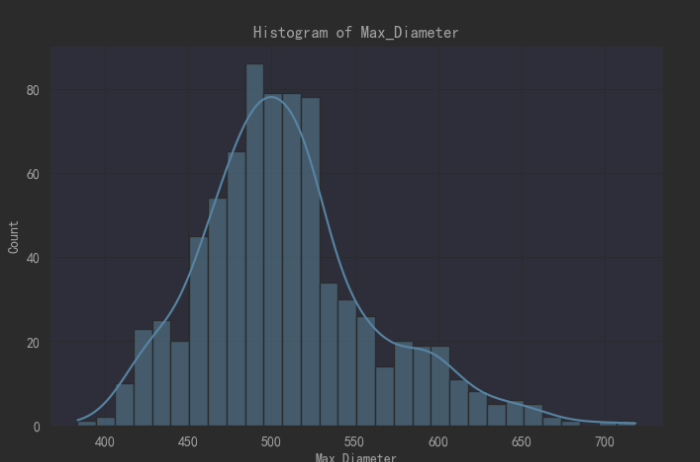
**Max\_Diameter\_μm** : mean **508.73** μm, standard deviation **52.11** , range **383.56 – 718.44** μm. The data is relatively concentrated, and the difference between the maximum and minimum values is about **335** μm.

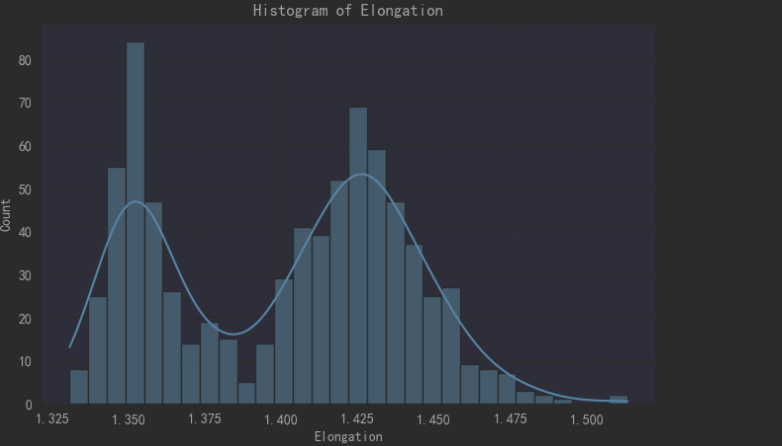
**Elongation** : The mean is **1.40** , the standard deviation **is 0.039** , the distribution is relatively compact, the maximum value **is 1.51** , and the minimum value is **1.33** , indicating that the morphological changes of the data are small.

**Age\_Ma** : mean **1.82** Ma, standard deviation **1.26** , minimum **0.01** Ma, maximum **5.00** Ma, with a large time span.

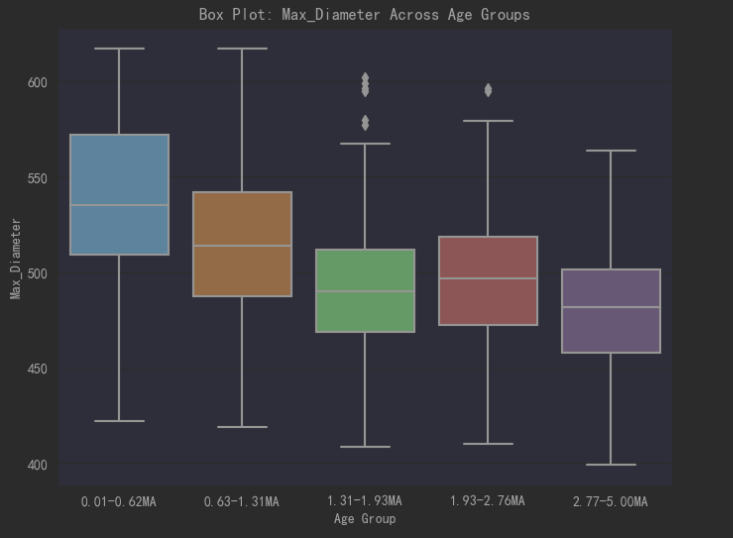
**Min\_Diameter\_μm** : mean **147.41** μm, standard deviation **9.91** , range **109.32 – 170.02** μm, the data distribution is relatively uniform.

Sphericity : **The 75% quantile is only 1.71** , but the maximum value is **3546.00** , which is far beyond the normal range, indicating that the data may have outliers or measurement errors. Since the mean **(293.35)** is much larger than the median **(0.70)** , and the standard deviation **(644.26)** is far greater than the mean, it indicates that the data distribution is severely **right-skewed** .

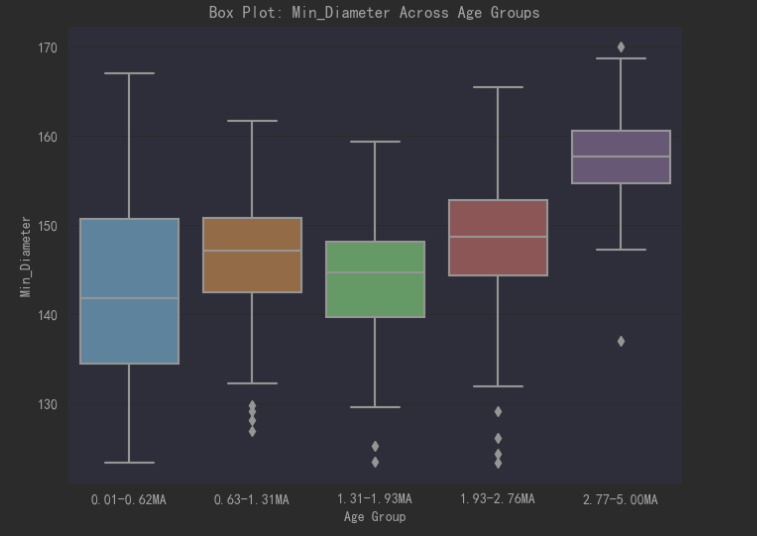
The mean of the data **is slightly higher than the median** , and the histogram shows a right-skewed distribution.



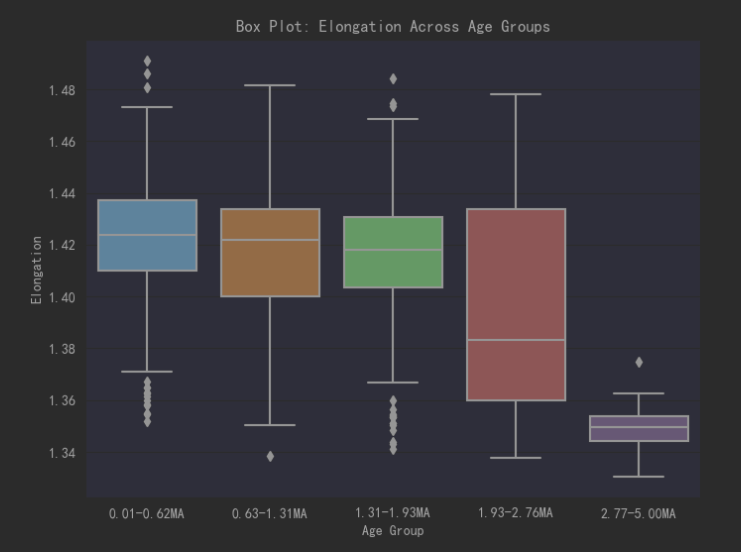
The bimodal distribution of the data may be suggested to indicate that the data come from two different groups of foraminifera.



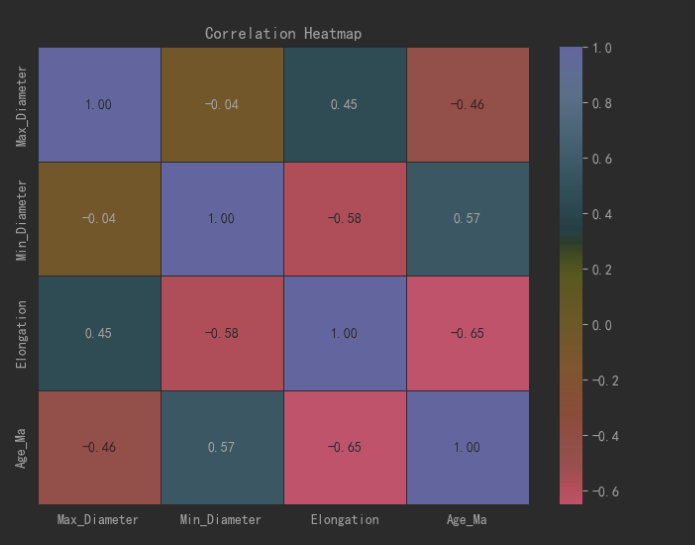
The box plot shows the changing trend of **the maximum diameter of foraminifera** in different age groups, revealing its changing characteristics over time. In general, **the younger foraminifera (0.01-1.31 Ma) have a larger maximum diameter** , with a median of about **520-550 μm** , a wider data distribution, and some large anomalies, indicating that the diversity of individual sizes during this period is high. Over time, **the maximum diameter of the 1.31-2.76 Ma group gradually decreases** , with the median dropping to **480-500 μm** , the data distribution tends to be stable, and the outliers decrease , indicating that the individual sizes of this stage are more uniform. **The oldest foraminifera (2.77-5.00 Ma) have the smallest maximum diameter and the most concentrated data distribution** , with a median of about **470 μm** , indicating that these individuals may have been affected by long-term changes in the ecological environment.



The box plot shows the trend of **the minimum diameter of Foraminifer** in different age groups, which is different from the evolutionary pattern of the maximum diameter. **The young foraminifer (0.01-1.93 Ma) has a smaller individual size and a wider distribution range** , with a median of about **140-145 μm** , and there are many outliers, indicating that the individual morphological diversity is high during this period. Over time, **the oldest foraminifer (2.77-5.00 Ma) has the largest minimum diameter, with a median of nearly 155-160 μm, the most concentrated data distribution** , and fewer outliers, indicating that **Foraminifer** may gradually tend to a larger stable form .



The box plot shows the changing trend of **the elongation of Foraminifera** in different age groups. **The elongation of younger foraminifera (0.01-1.93 Ma) is relatively stable, with a median of about 1.41-1.42, and the data distribution is relatively concentrated** , indicating that the individual morphology of this period is relatively uniform. **The elongation range of the 1.93-2.76 Ma group is the largest** , and the interquartile range increases, indicating that there are large fluctuations in the individual morphology of this stage, which may be related to environmental changes or species diversity. **The oldest foraminifera (2.77-5.00 Ma) has the smallest elongation and the most concentrated distribution, with a median of about 1.35** , indicating that the individual morphology of this period is more consistent and may have experienced strong selection pressure or adaptive evolution. Overall, the change in elongation may reflect **the differences in the adaptation of foraminifera to the environment in different geological periods** .



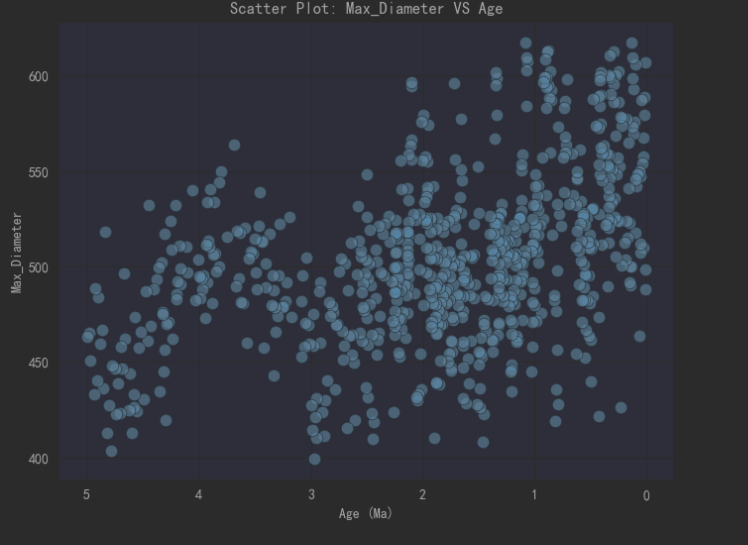
Age\_Ma is strongly negatively correlated with Elongation (-0.65), indicating that as the age of foraminifera increases, their morphology may tend to be less elongated, that is, the shape becomes more round or more compact. This may be related to adaptive evolution or changes in ecological environment.

Age\_Ma was positively correlated with Min\_Diameter (0.57), indicating that older foraminifera individuals tend to have larger minimum diameters, which may be related to population selection or shell preservation bias.

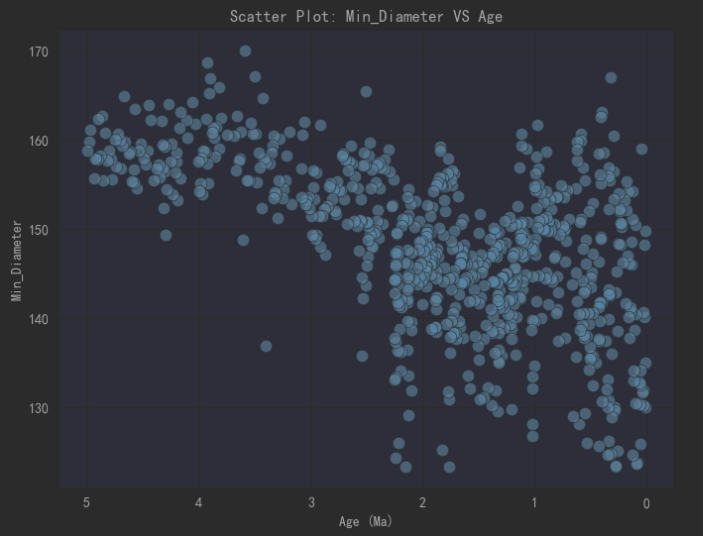
Max\_Diamete is moderately negatively correlated with Age\_Ma (-0.46), which indicates that the maximum diameter of foraminifera individuals decreases to a certain extent with increasing geological age, which may be limited by environmental pressure or nutrient supply.

Elongatio was strongly negatively correlated with Min\_Diameter (-0.58), meaning that individuals with more elongated morphology generally have smaller minimum diameters, while larger individuals may be more rounded.

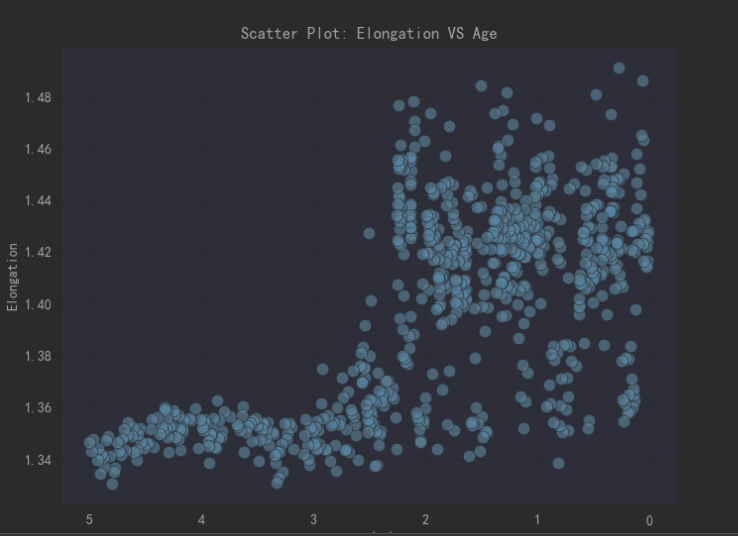
Max\_Diameter was positively correlated with Elongatio (0.45), indicating that larger individuals usually have higher elongation, which may represent some specific growth patterns or species characteristics.



The scatter plot shows the relationship between the maximum diameter **of foraminifera and their age** . From the overall trend, **the maximum diameter of older samples (>3 Ma) is relatively small, mainly concentrated between 400-500 μm, while the maximum diameter of younger samples (<2 Ma) is larger, and some exceed 600 μm** , indicating that the maximum diameter of foraminifera individuals may show an increasing trend over time. However, the distribution of data points is relatively scattered, indicating that the individual sizes vary greatly in different time periods, which may be affected by environmental factors, species diversity or preservation conditions. In addition, the data points in some age groups (such as 1-3 Ma) are dense, indicating that there are more samples in this period, and the relationship between diameter and age can be quantified using one-step **regression analysis .**



The scatter plot shows the relationship between **the minimum diameter and age of foraminifera** . From the overall trend, **the minimum diameter of older samples (>3 Ma) is larger, mainly concentrated in 155-165 μm, while the minimum diameter of younger samples (<2 Ma) individuals is significantly reduced, and some are lower than 130 μm** , indicating that the minimum diameter of foraminifera has been declining over time. This may reflect changes in the ecological environment, such as changes in the **marine environment , reduced nutrient supply, or adjustments in survival strategies** , resulting in smaller individuals being more adaptable. In addition, the data points **are densely distributed between 2-3 Ma** , which may represent an important stage of morphological change. Although the overall trend is relatively obvious, there is still a certain degree of dispersion, indicating that individual size is affected by multiple factors, such as species diversity or the selection effect of the sedimentary environment.



The scatter plot shows the relationship between **the elongation of foraminifera and the year** . From the overall trend, **the older samples (>3 Ma) have lower elongation, mainly concentrated between 1.33-1.36, while the younger samples (<2 Ma) have higher elongation, some exceeding 1.45** , indicating that the morphology of foraminifera has changed significantly over time. **The individual morphology before about 3 Ma was relatively stable, with low elongation and a small range of variation** , which may represent a relatively homogeneous population or stable environmental conditions. However, between **3-2 Ma, the elongation changed dramatically, and the distribution of data points became more dispersed** , which may correspond to an important ecological transition period, such as environmental change or species adaptive evolution. **After 2 Ma, the number of individuals with larger elongation increased and the distribution range became wider** , indicating that the morphological diversity of foraminifera increased during this period .