Analyzing a subset of data.

Making a subset based on the columns with values across all the years.

\* Descriptive analysis-

We have 219 rows and 7 columns in the data now.

After visualizing for missing values, we see that we're having one missing value at the start of co2\_growth\_prct. But as this variable equals to the percentage change in co2 emissions relative to the previous year, the very first value is Null.

\* By comparing the mean and median values of each column, we can get a rough estimate of the distributions of values in each column.

\* The columns co2, population show that the mean values are much larger than the median values.

\* Checking for outliers in co2 column

Values of all the points are shown to the left. As the max value does not go beyond the whisker line, we should estimate there are no outliers. We can confirm it by changing the *\_points\_* parameter to 'outliers'. It will not show any points.

\* Visualizing co2 values

The sudden declines such as (1929-1933, 1978-1982) might be due to the great depression era, the inflations in the late 70s, respectively.

The decades 1960s and 1980s show a steep rise. The economy was also rising at that time.

After the peak of 6130.55 in 2005, the values seem to be decreasing after 2007. Maybe due to Obama gov policies.

\* Comparing co2 values and co2\_per\_capita values

As expected, we see can see that both the co2 and co2\_per\_capita variables show mostly similar trends.

But the co2\_per\_capita variable peaks at 22.133 in 1973. After 2007, it keeps on decreasing.

\* Global share of co2 emissions in the USA

Years 1926 to 1938 show a steady decline in the global share of co2 emissions for the USA. It may relate to the downfalls in the economy during that period.

After which it starts to increase rapidly and peaks at 54.35 in the year 1945. It may mean that many other countries started emitting co2 more rapidly than before.

\* Looking for relationships between variables.

We can estimate that co2 and population are the only independent variables in the subset data.

The global share values may depend on the co2 values in the USA and co2 values globally.

A curve with a slope between linear and logarithmic should fit on the graph here.