Nicolas Saenz

DSC 530

3/2/25

I started the project wanting to know if economic data can be used to approximate or predict the number of invasive species within a country. The outcome of my EDA is that assuming the data is correct the number of invasive species is highly left skewed many countries have zero invasive species, but there is a long right tail. While the US has the highest number of invasive species as well as highest in many other metrics.

I utilized large general economic data for each country. I think that this approach means I either missed many of the smaller contributors to invasive species like the size of the exotic pet industry or the size of the tourism industry. Besides missing smaller metrics I think using such large metrics may have watered down some valuable data, for instance focusing on food imports may have been a more meaningful number than overall imports.

Like mentioned above I think exotic pet industry size and tourism industry size, and food imports may have been really valuable variables for trying to get more predictive power into a model.

My assumption is that separating data based on whether or not a country is land locked may have too many hidden variables attached to it to be a meaningful to separate high invasive species countries from low invasive species countries.

Many of our examples in class dealt mostly with approximately normal looking distributions. The struggle I had is that most of my data was not normal or even approimately so. Finding the appropriate means to statistically work with the data was one of my most serious challenges in this project.