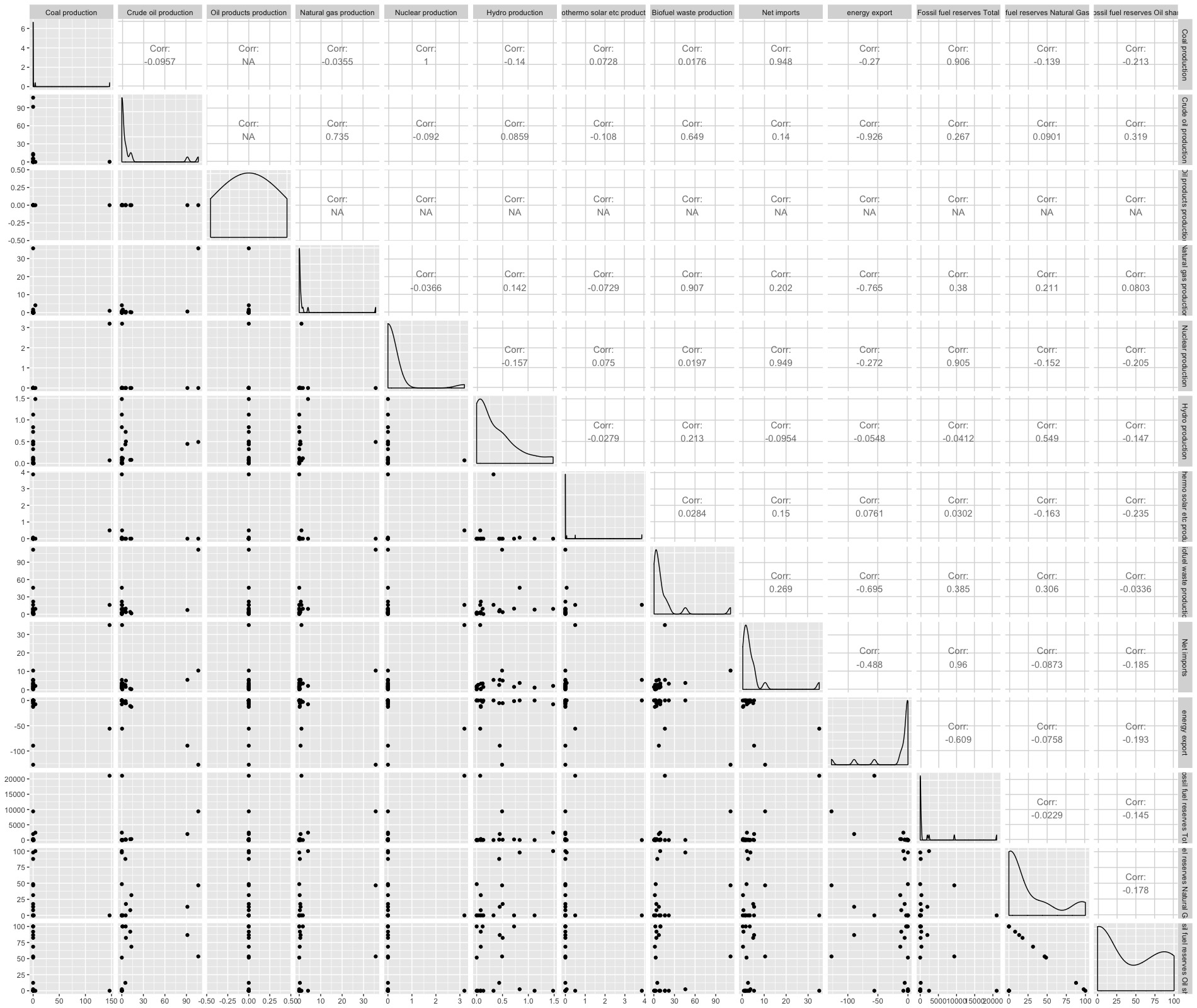


We can see the water resources availability, variability, and extreme weather events are highly correlated.

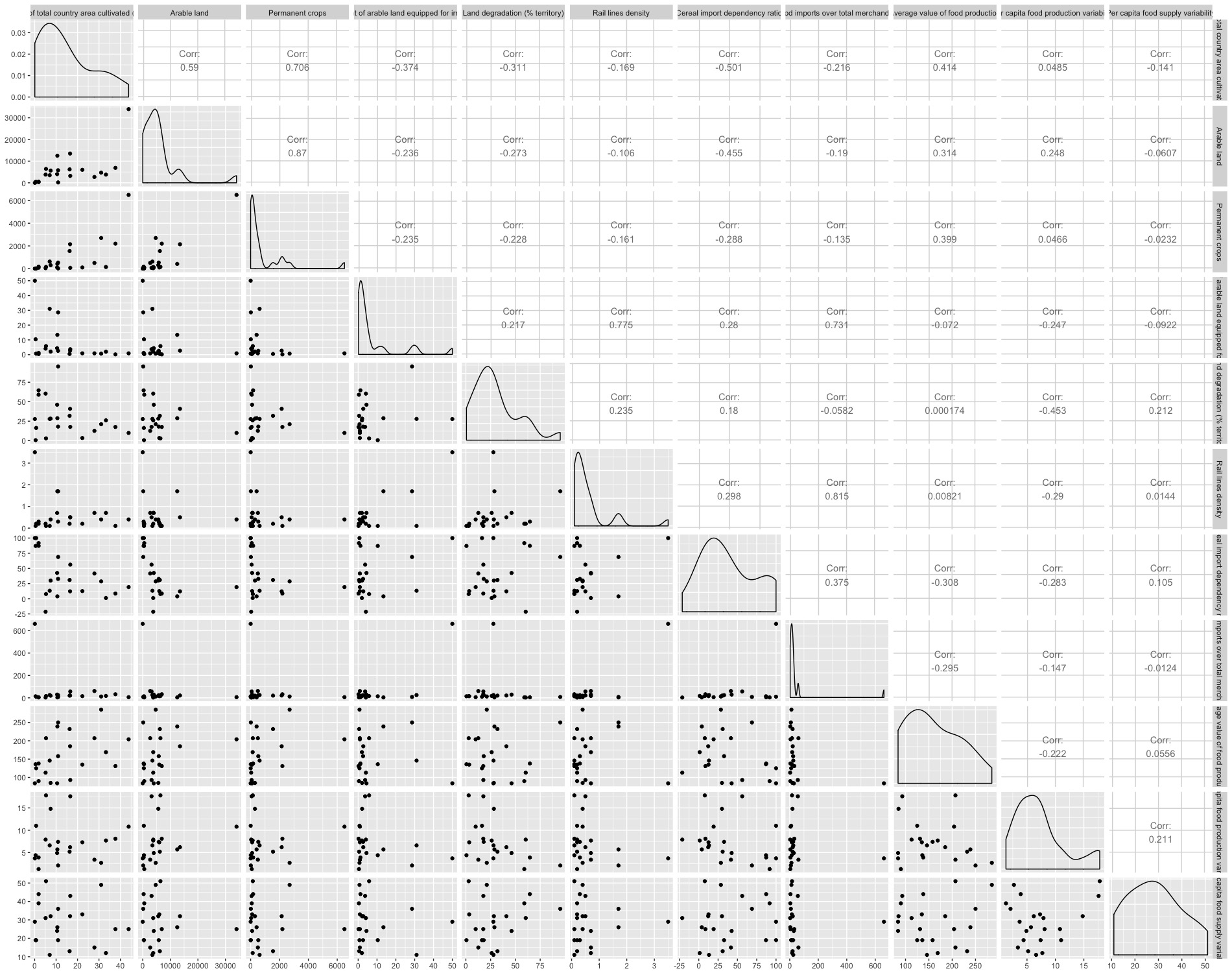


We can see from this correlation matrix of the independent variables of energy resources, many countries have numbers of zero, so we see NA values. The distribution of the energy data are so skewed.

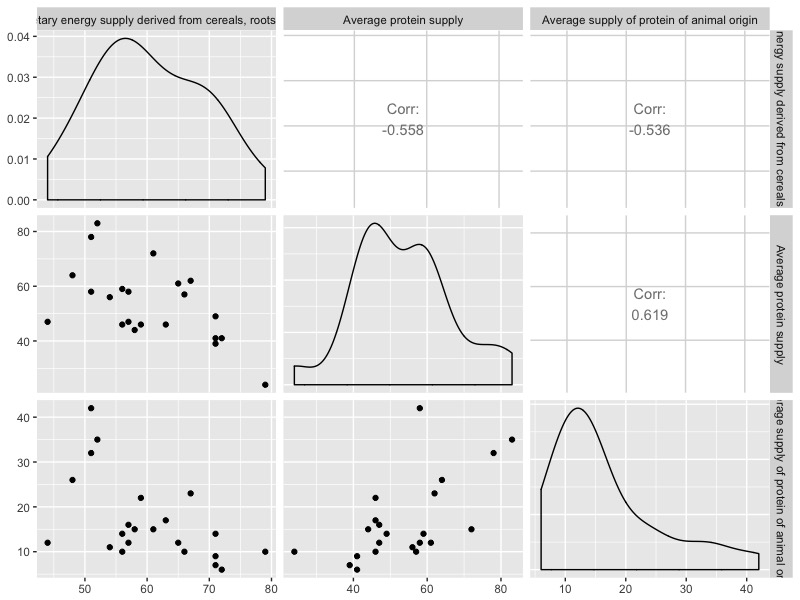
1. For energy, given that a lot of the data is wonky/related, it might be worth it to explore a hierarchical/multi-level regression. Where basically, your y ~ total production + net imports + energy exports + total reserves, where total production and fuel reserves are composed of the different fuel types you have data for.

Since Oil Products that no countries have any, we can eliminate that data in the stepwise regression.

* Total fossil vs total non-fossil

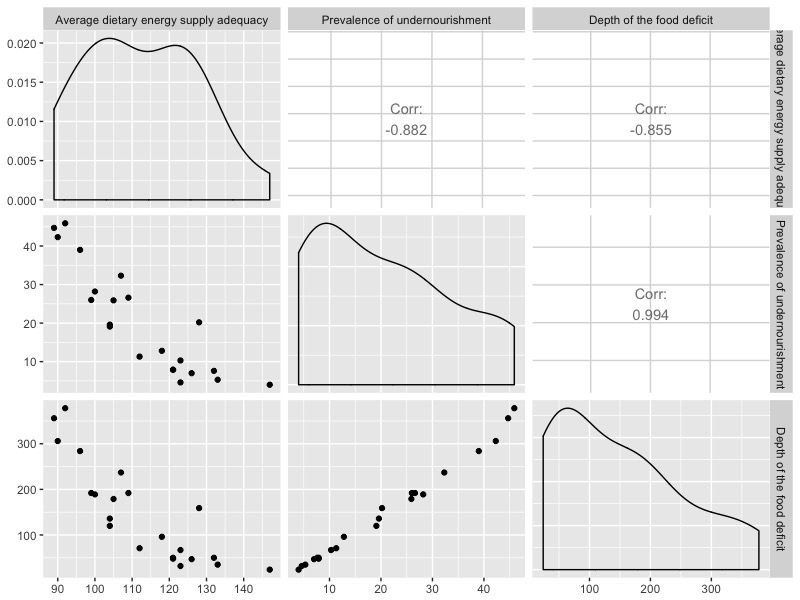


This is the correlation matrix of the food (land resources) independent variables. The land resource availability variables (total area cultivated, permanent crops, arable land, percentage of irrigation, land degradation). Rail line density are highly correlated with most of the variables, which food import and percentage of irrigation land are the most correlated variables.



Share of dietary energy supply derived from cereals roots and tubers, average protein supply, average supply can be interpreted as measures of utilization: nutritional balance.

We can see that the three are also very highly correlated.



Average dietary energy supply adequacy, prevalence of undernourishment, and food deficit can be all interpreted as measures of general utilization of food. The three variables are highly correlated. Food deficit are rather significantly correlated to prevalence of undernourishment.

2. For food, I like how you broke out some of the variables into related groups (I.e., nutritional balance and deficit/lack of availability). These findings are interesting in and on themselves (e.g., prevalence of undernourishment vs depth of food deficit almost look like they were derived from each other! Is this next step to do a PCA on these mini groups?

* The plan is to do pca on the first group of the land resources availability. And the second is to perform the pca for the nutritional balance, the third plan is to choose the deficit of food.
* Capacity variables, include the correlation variables from the outcomes.
* Pull irrigation, pull that out as a capacity variable.

Second word document, Clean Version, include PCA, more into text.

Last document will be the regression.

Once we have all the final variables, make copy of the framework and put the variables into the diagrams. Track Units, and Year.

Table for the variables. A compact way to keep track of everything.