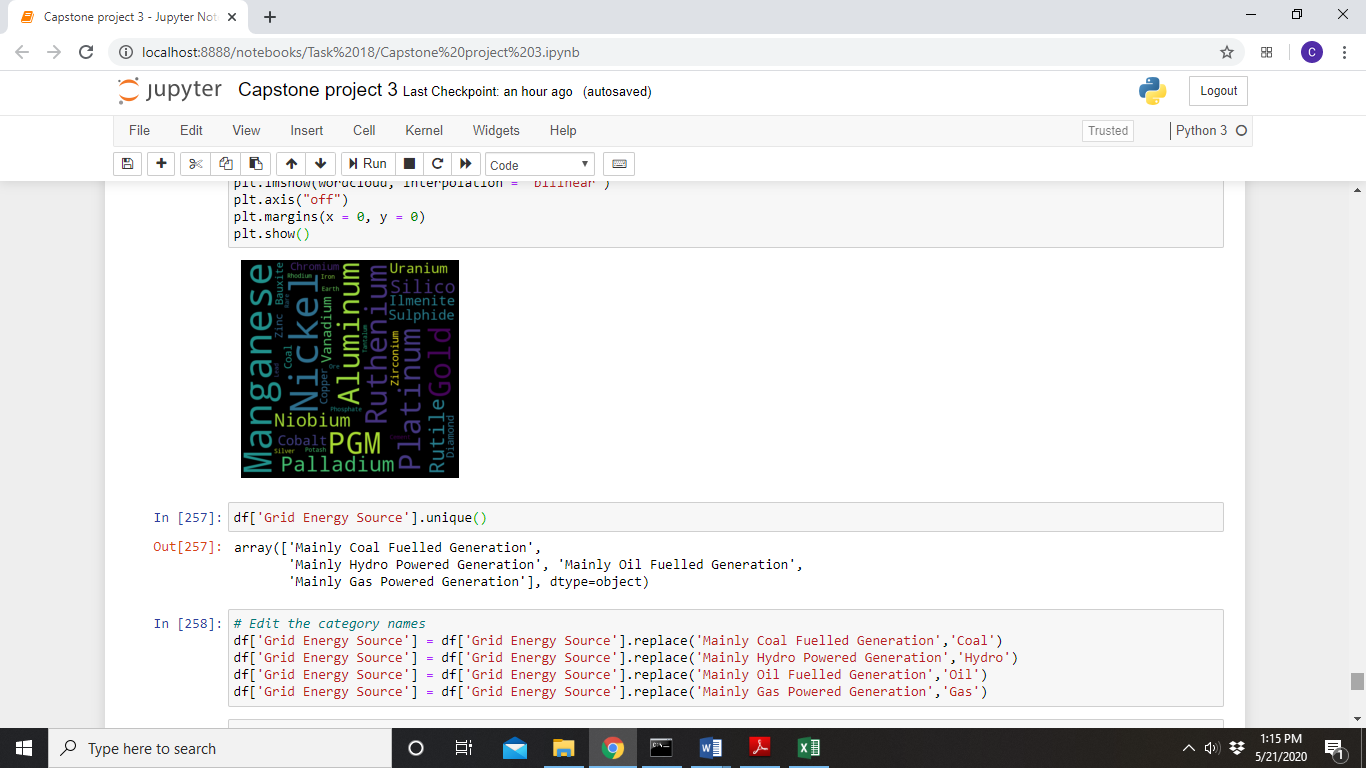


**TASK**

**Exploratory Data Analysis on the Africa power mining projects Data Set**



[](http://www.hyperiondev.com/portal/)

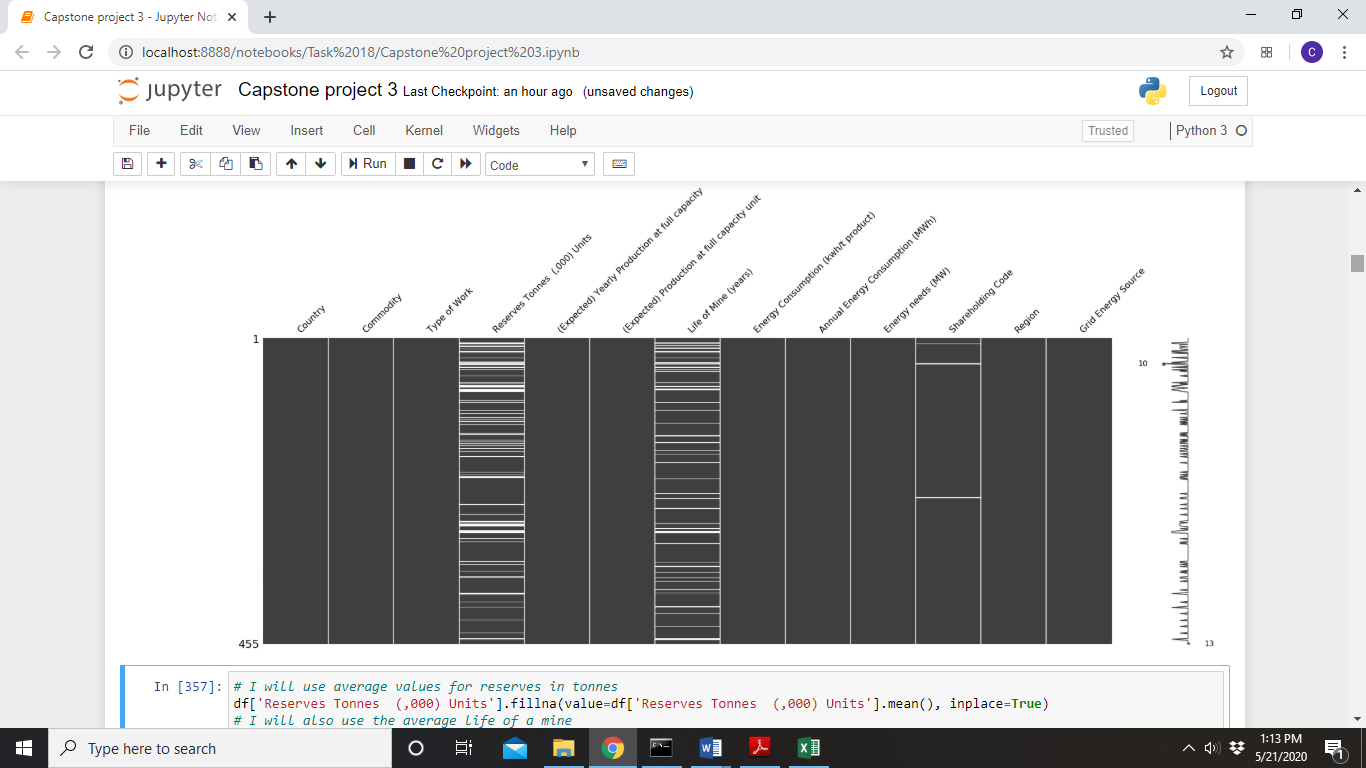
**Introduction**

The data chosen for this analysis was found in the World Bank Data repository. It contains data on 455 mining projects throughout Africa, and it has 39 columns with data including the type of commodity mined, the reserves of each commodity and form of ownership of the mine. The dataset includes quantitative data (for example energy needs) and qualitative, categorical data (for example power source).

While I investigated the relationships between the different quantitative variables, I was mainly focused on identifying the number of mining projects in each country and bringing out the categorical data in the dataset. I also look at the commodities with the largest reserves across the continent and I also visualize the commodities mined in each country.

**DATA CLEANING**

The first step was to remove 26 unnecessary columns using the pandas drop method, leaving me with 13 columns. Of these columns, there were 3 that had missing values, which are displayed below.



The data cleaning also involved looking at the unique values under the Country column and merging the two versions of the name for Democratic Republic of the Congo using a method from the fuzzywuzzy library. I also checked data types to ensure that each column was classified under the correct data type. When visualizing the distributions of the continuous variables, most were not normally distributed, so I normalized them before plotting.

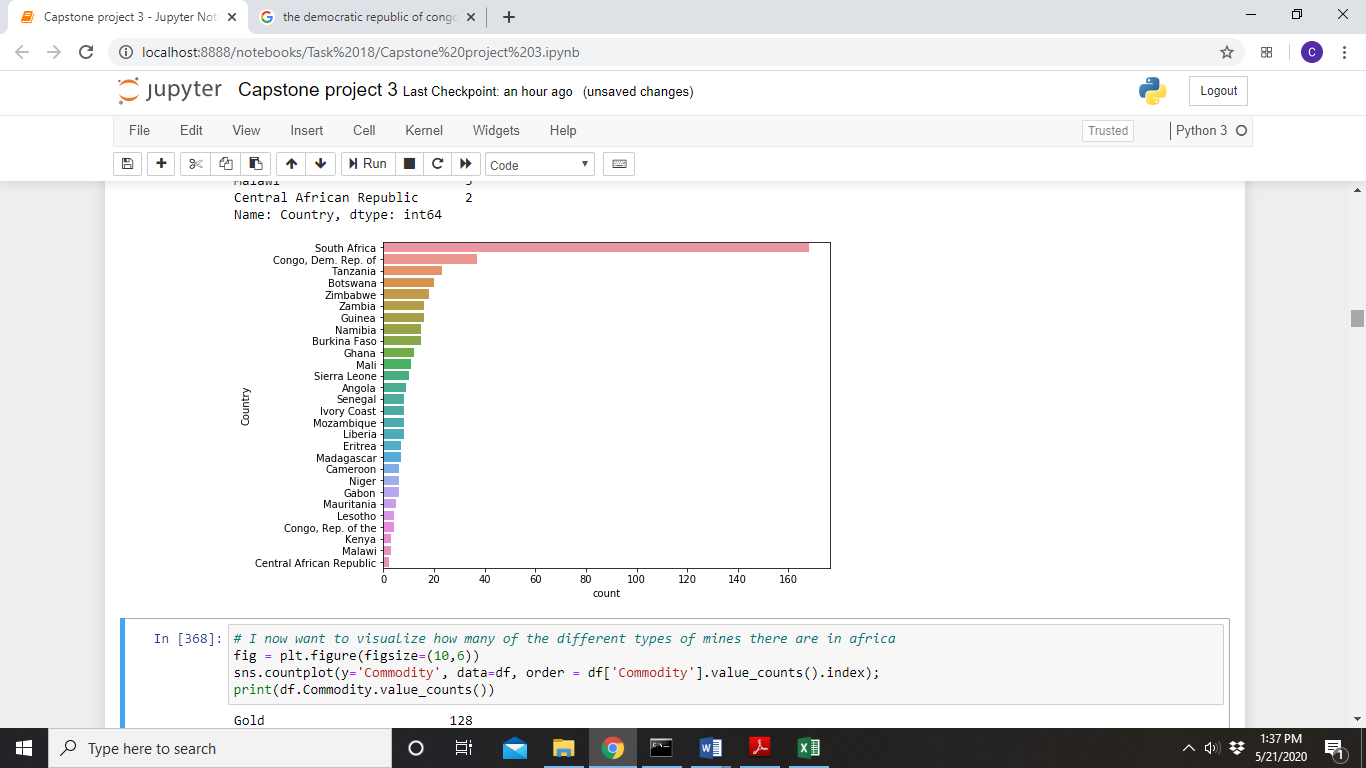
MISSING DATA

I used average values to replace the missing values under the ‘reserves in thousands of tons’ column and the ‘Life of Mine (years)’ column. I then replaced missing values with the modal category under the ‘Shareholding Code’ column.

DATA STORIES AND VISUALIZATIONS

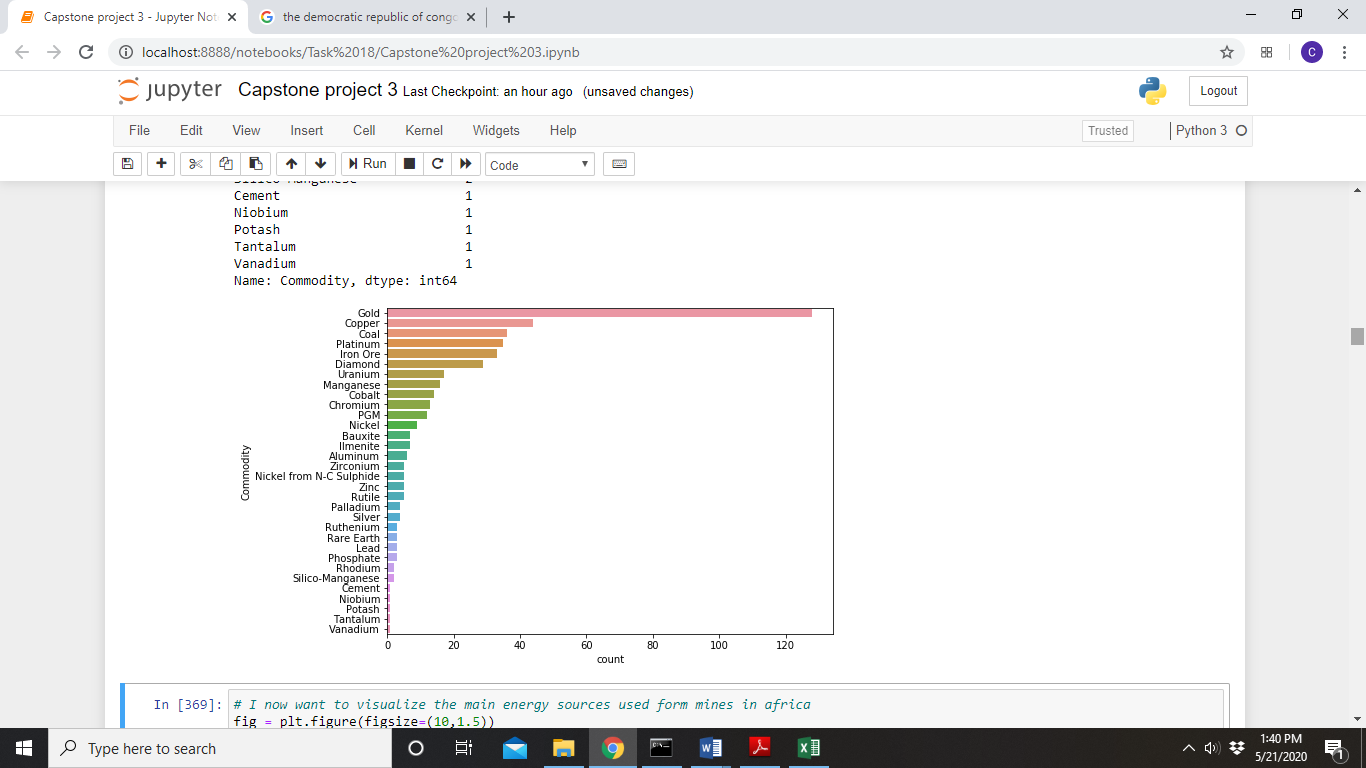
**Which African country is leading the way?**

I started off by getting an overview of the number of mining projects recorded per country. We clearly see that South Africa leads the continent in terms of total mining projects, followed by Democratic Republic of the Congo, Tanzania and Botswana.

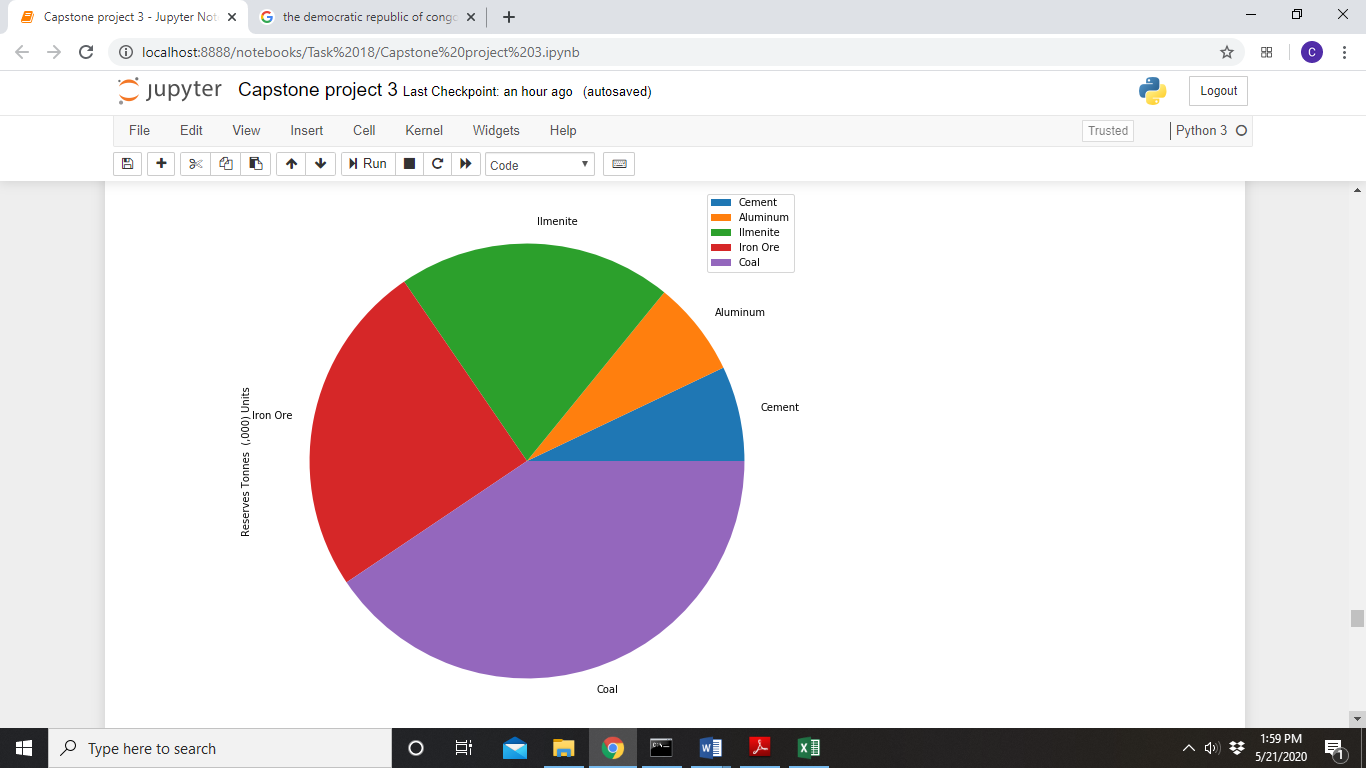


**Which commodities do mining projects target?**

Next, I wanted to get a summary of the commodities that are most commonly mined across Africa as a whole. It becomes apparent that gold is the most commonly mined commodity, followed by Copper, Coal and Platinum.



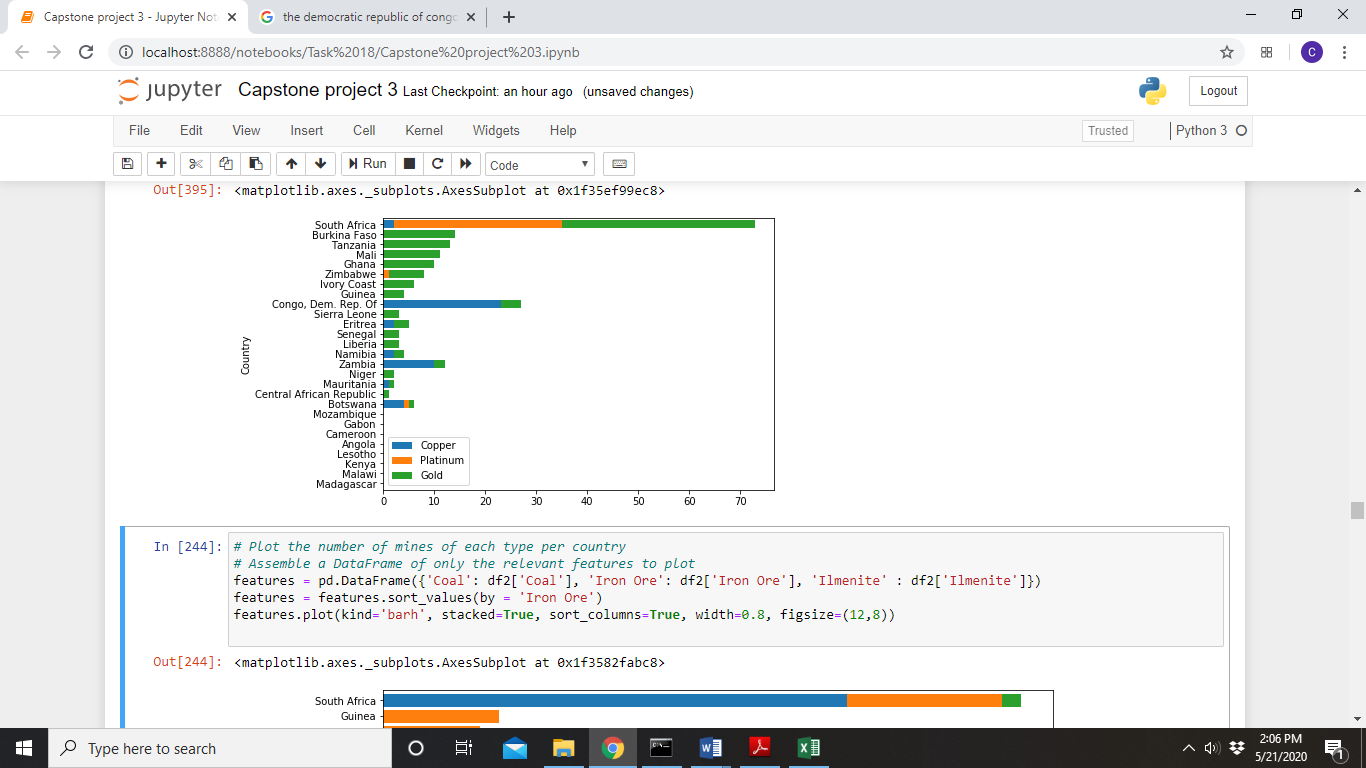
**Which commodities are most abundant?**



Similarly, one might ask the question of which commodities we have the most reserves of. The pie chart above displays the 5 most abundant commodities in terms of documented reserves across the continent. Coal is the most abundant, followed by Iron Ore, Ilmenite and Aluminum.

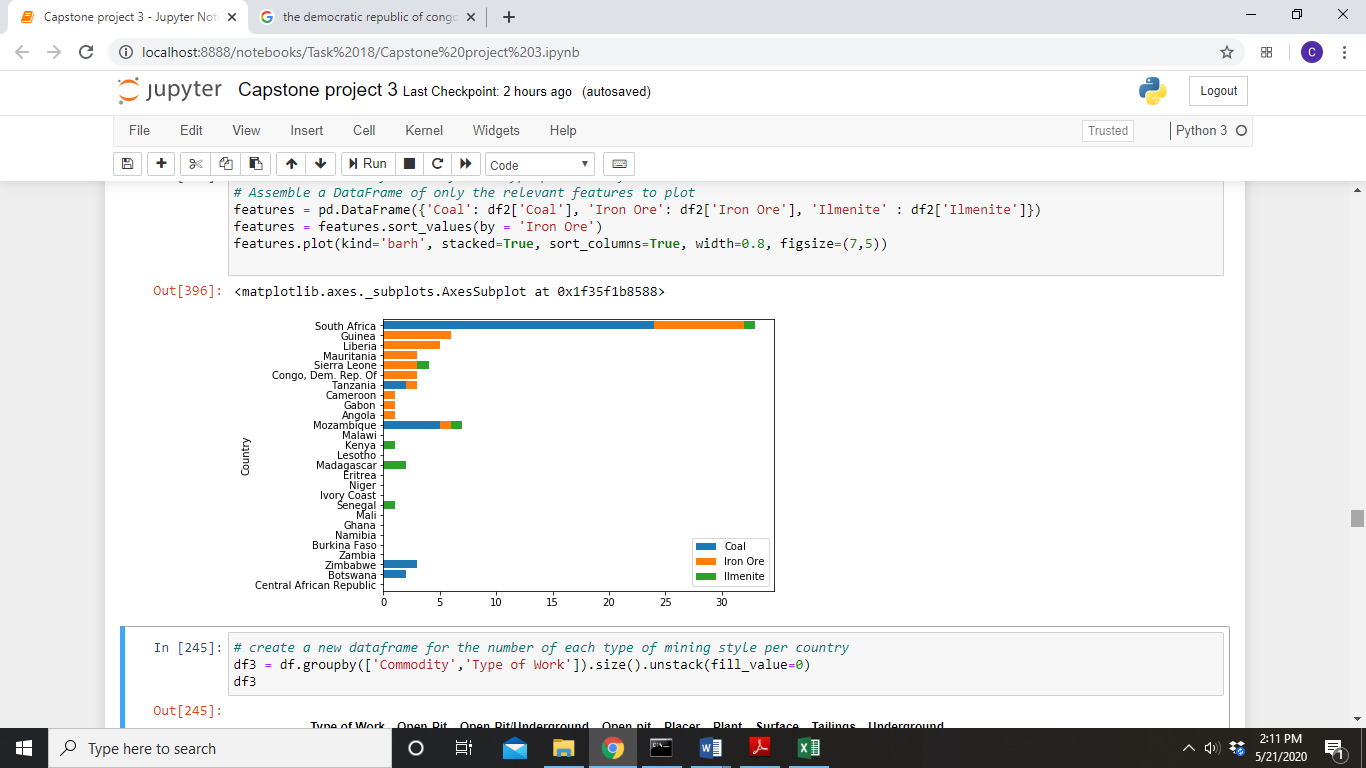
**Which countries are mining these commodities?**

Now that we have an idea of what is commonly mined and what is most abundant, we can see which countries are mining these commodities.



There is some interesting information here:

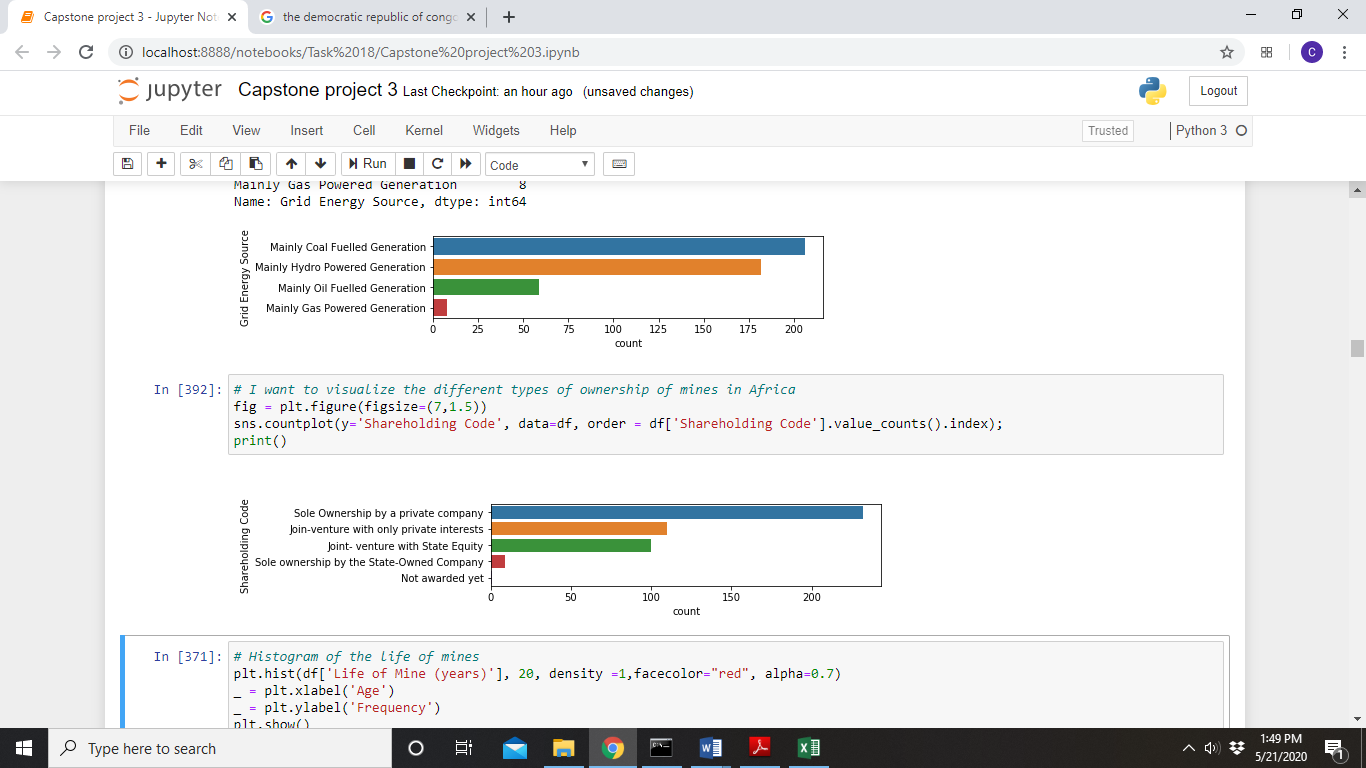
* South Africa has the most gold mines, followed by Burkina Faso, Tanzania and Mali.
* Platinum is only mined by South Africa, Zimbabwe and Botswana and
* Congo has the most copper mines.

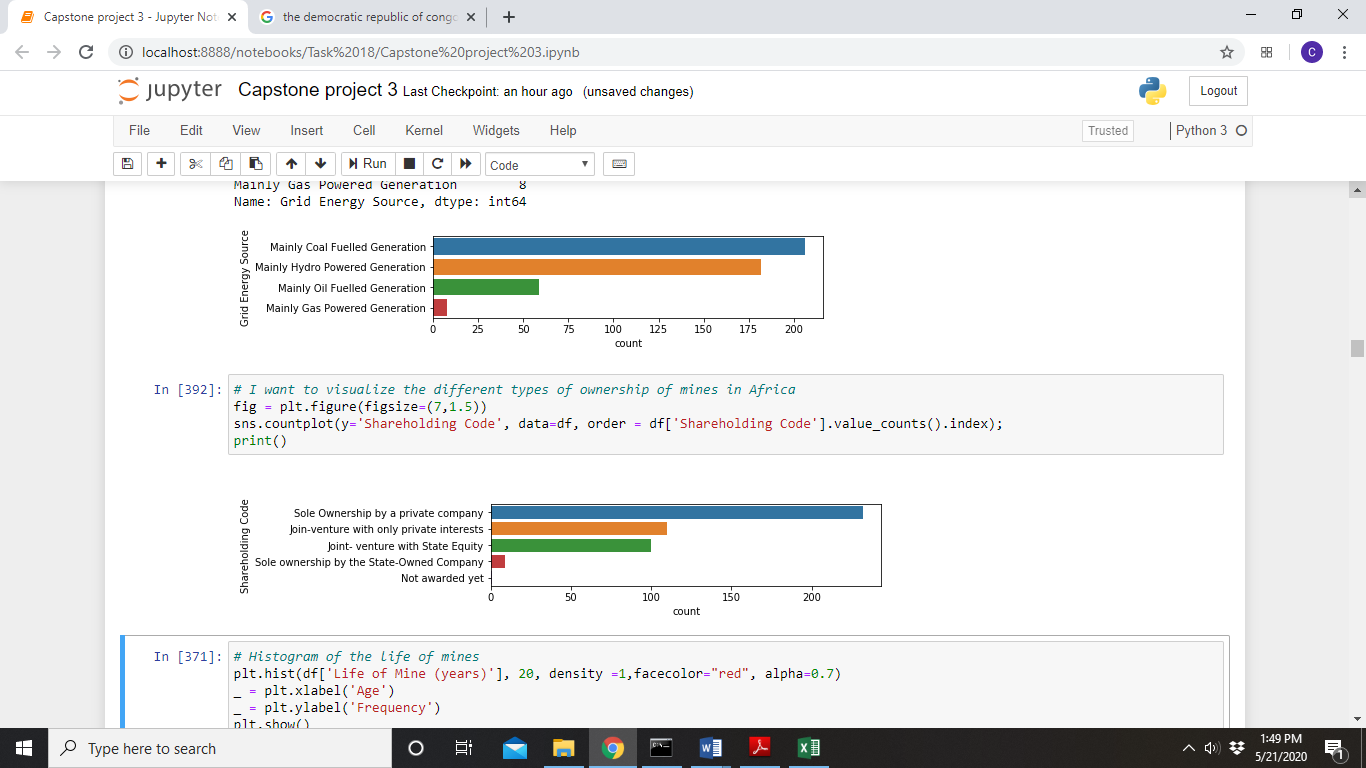


Coal and Ilmenite are abundant, but only mined by a few countries.

**Ownership of mines and sources of power**

I generated count plots to display the number of mining projects that fall under the different forms of ownership as well as the different sources of power used.

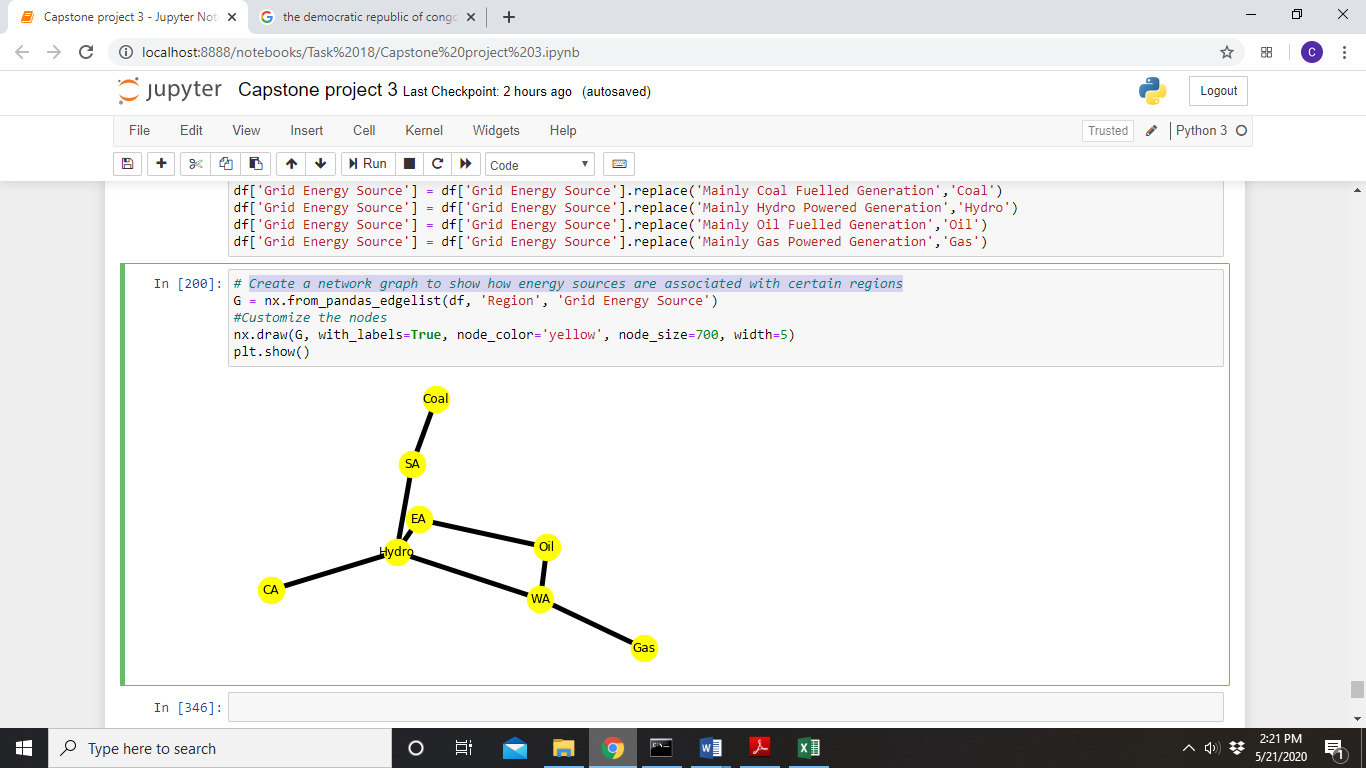




As one might assume, the majority of mines are privately owned and most rely on coal fueled generation. Few mines rely on energy from gas and state owned companies rarely have sole ownership over mines. It is interesting that many mines use hydro power.

**Are energy sources associated with certain regions?**

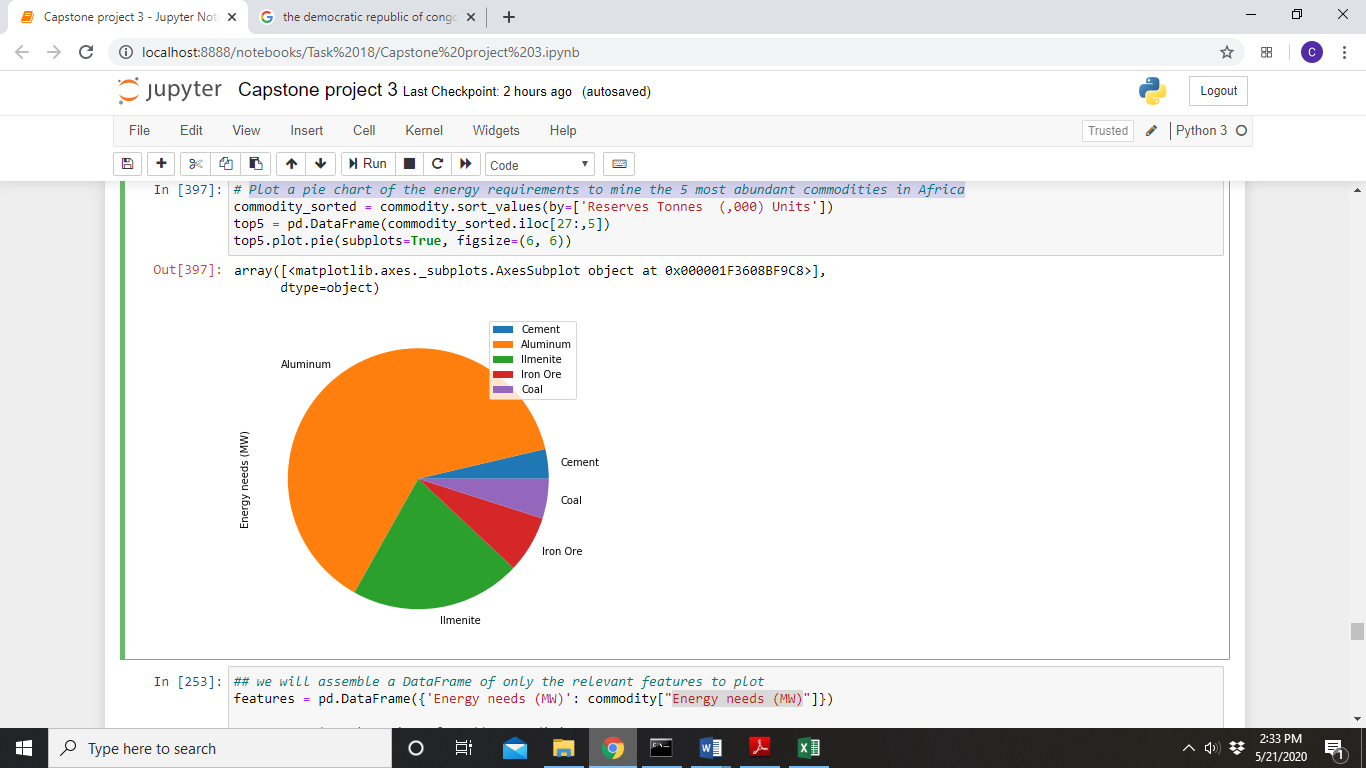
I wanted to investigate whether different energy sources are associated with certain regions, so I generated a network graph to represent this.



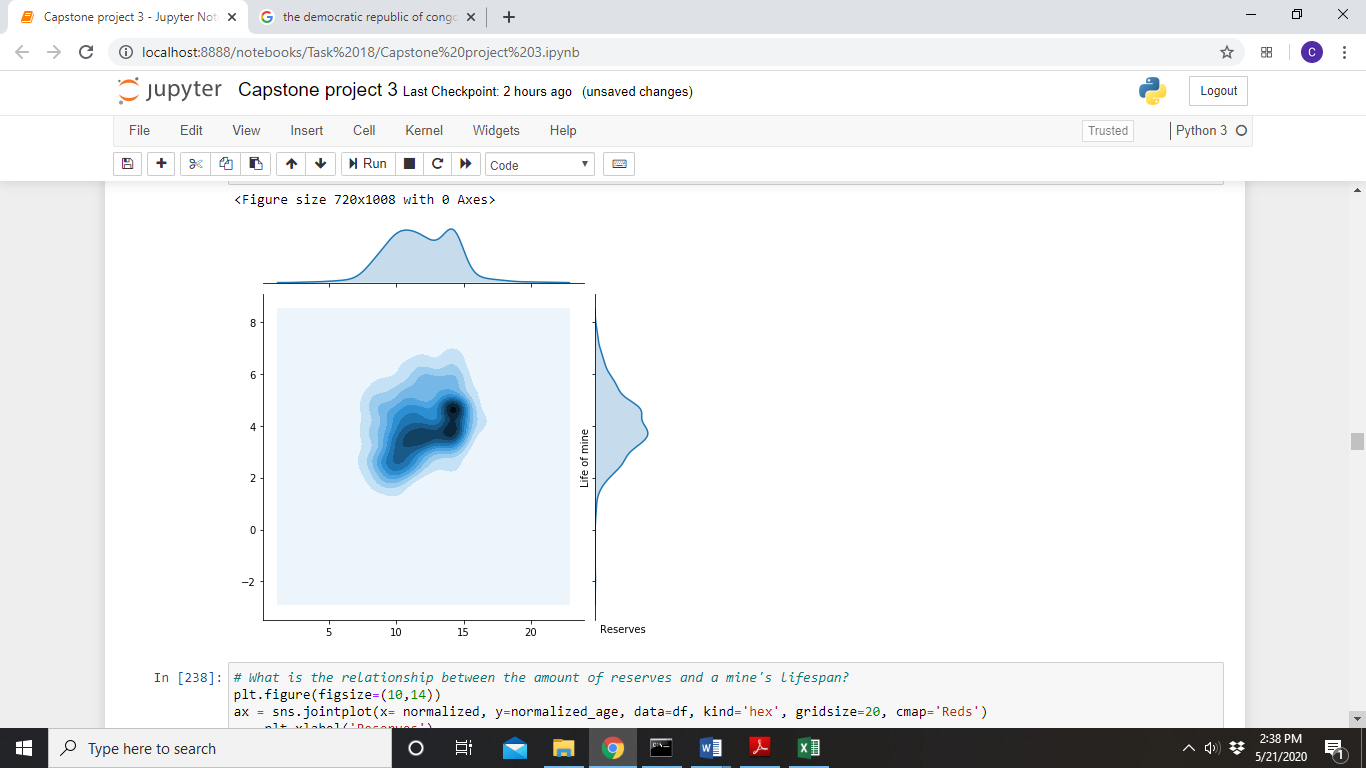
This graph shows that Southern Africa is heavily dependent on coal for its mining, however it is also dependent on hydro power. Mines in East and Central Africa also rely on hydropower, with mines in East Africa also using oil. Mines in West Africa, on the other hand, tend to rely more on energy from gas an oil.

**How do the different commodities differ in the amount of energy required to mine them?**

The pie chart below represents the energy required to mine the 5 most abundant commodities in Africa.

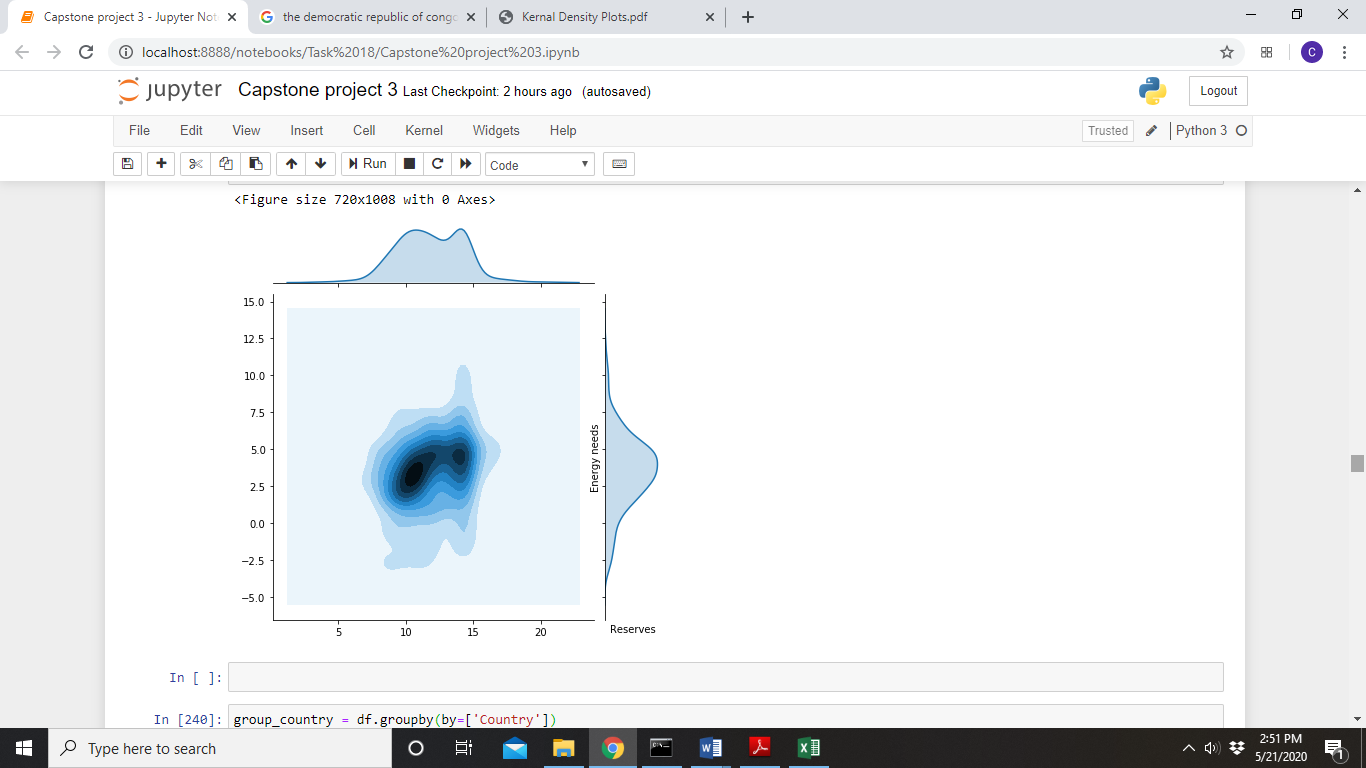


**Is there a relationship between the abundance of a commodity and the lifespan of the mine?**



The above figure is a plot of ‘life of mine’ against ‘reserves’. The distributions were both normalized, as they were both skewed to the right and had very different value ranges. The kernel density estimation type was chosen to show where there is clustering in the data. There seems to be a positive relationship between the abundance of the commodity (reserves) and the lifespan of the mine, however there is a cluster at slightly higher ‘reserve’ and ‘life of mine’ values. A positive relationship would make sense, as a mine with more reserves should be mined for a longer period of time.

**Is there a relationship between the abundance of a commodity and the energy requirements of the mine?**



This figure represents ‘energy needs’ against ‘reserves’. The kernel density estimation seems to show a positive relationship between the abundance of the commodity (reserves) and the energy needed, however there are two distinct clusters at slightly lower and higher ‘reserve’ values. The trend seems to drop off at the second cluster, where energy needs tend to vary more.

CONCLUSION

The exploratory data analysis conducted here revealed which commodities are most commonly mined, most abundant, and which countries mine them. It also gave insight into the most energy intensive commodities and the types of energy used for mining in the different regions. Lastly we found that there is a positive relationship between the abundance of a commodity and both the amount of years a mine will operate for and the energy needed to operate the mine.

**THIS REPORT WAS WRITTEN BY: CHRISTOPHER DELPORT**

