The purpose of this project is to experiment with various machine learning concepts in order to find the optimal approach for identifying anomalies in physical commodities prices at any given geographical location. We will be using futures data from commodities exchanges around the world, relying primarily on the QUANDL continuous futures data (CHRIS). With accurate models built around price behavior we can then begin building a model that can accurately detect inventory anomalies at any given geographical location. Once anomalies have been identified they will be categorized. With this new categorized the data we can finally start work on building a model that identifies trading opportunities in real-time. With the outputs from this model a trader will have the ability to identify location arbitrage opportunities slightly before his competitor. The nature of commodities trading often leads to zero sum outcomes. Having such an advantage crucial.

1. Reshape QUANDL CHRIS data to account

Explanation of QUANDL’s Continuous Futures Series:

<https://quantoisseur.wordpress.com/2018/02/05/constructing-continuous-futures-price-series/>

<https://github.com/coltonfsmith/BlogProjects/blob/master/Continuous_Futures_Quandl.py> - fork this for reshape QUANDL

<https://www.quandl.com/data/CHRIS-Wiki-Continuous-Futures/documentation?anchor=continuous-contracts>

<https://blog.revolutionanalytics.com/2014/06/constructing-a-continuous-futures-series-from-quandl.html> - older and in R

1. Collect, clean and shape inventory data.

<https://unstats.un.org/unsd/trade/data/tables.asp#monthlytotal> – UN International Trade Inventories, monthly – Was looking great until the API server crashed

<https://www.census.gov/data/developers/data-sets/international-trade.html> - US Import/Export Inventories and Values, monthly – Solid data. I have already cleaned this and have it in a database. Only for America

<https://wits.worldbank.org/witsapiintro.aspx?lang=en> – This is a difficult API to use. Hopefully the UN API is back online soon. It does have tariff information which could be added later as a feature

1. Collect location data.

We will likely need to divide the world up into trade regions. This has been done by several organizations and is industry practice for international trading. Still looking for an api that already has the lines drawn for us.

1. Picking the models. XGBOOST for sure. I would like to explore using a GANs model to create a “fake” of what the world is supposed to look like at any given moment. Then compare that to what the world actually looks like. Lots to work out here.