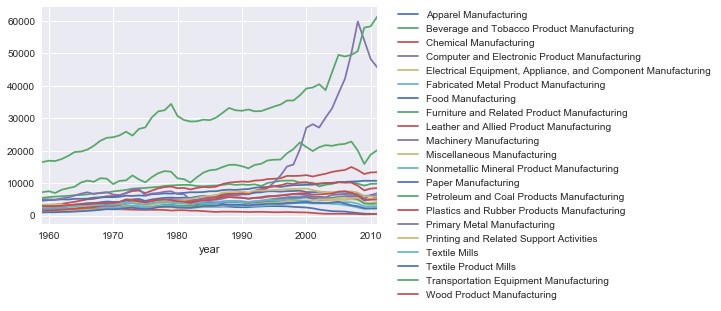
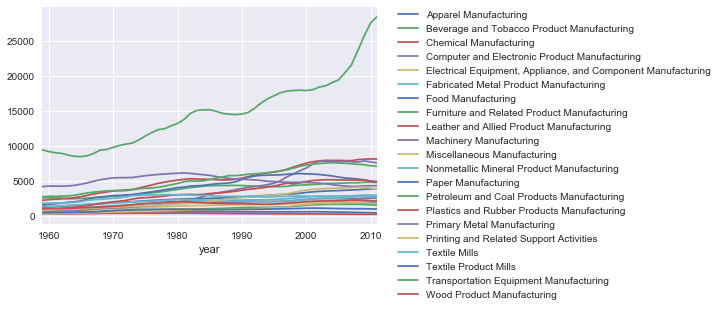
The dataset I’m working with has a total of 25542 observations, covering 473 industries over 54 years. The cleaning of the data was relatively simple. There were about 150 rows that had missing information. In the end, I decided to simply remove the rows as they only accounted for less than one percent of the observations. I also added a total of 5 columns to the dataset. Four of the columns use deflators to convert the nominal value into real values for the purposes of comparing growth over time. The last column uses the first 3 digits of the NAICS code to determine an observation’s main category.

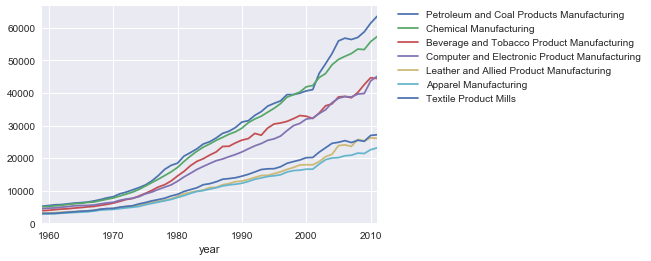
I started my EDA by taking a look at the changes in variables over the 54 years. Immediately, I noticed that employment in the manufacturing business has been on a significant downward trend. This is true of all industries, although the severity of this change depends on the industry. Looking at individual categories, the Petroleum/Coal industry is by far the largest and the fastest growing. For example, the value of shipments in Real GDP has grown from $16 billion to $60 billion.



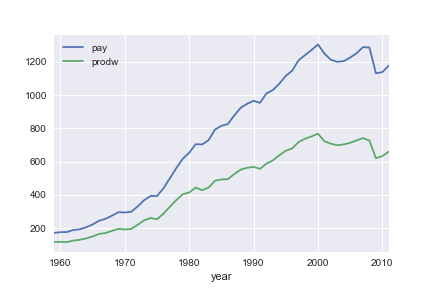
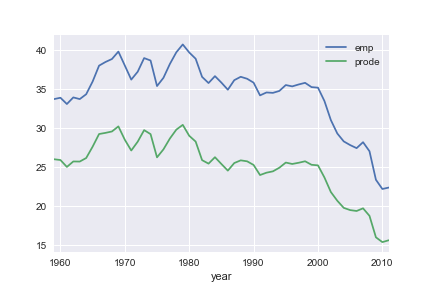
In addition, the total capital investment for Petroleum is three times that of any industry, as shown below.



By dividing the production worker wages and production worker wages, we can find the average wage for each industry. Something that is immediately evident is the disparity in wages between different fields. The industries with the highest wages are Petroleum/Coal, Chemical, Computer/Electronic Product, and Beverage/Tobacco. All of these fields seem to be an emergent field, a field requiring more specialized knowledge, or a combination of the two. Inversely, the three industries with the lowest wages (Leather, Apparel, Textile Product), require less specialized knowledge, and are either on the decline or are in a state of stagnation. It’s also possible that these industries are more easily automated and thus require far less employees.



Meanwhile, as we see an increase in wages, we also see a decrease in general employment as well as production employment. As shown through the charts below, manufacturing employment peaked in 1979 and has been decreasing ever since. On the other hand, wages stagnated for a little around that time before returning to their normal increase. One thing of note is the average payroll decrease of $100 between 2000 and 2004 and the massive decrease in employment starting in the 2000’s. In total, from 2000 to 2011, the manufacturing industry lost 6 million jobs, 4.5 million of those jobs being production jobs.



I decided to use a Linear Regression model in order to determine what variables contributed best to improving Average Yearly Wages. One interesting thing to note is the negative correlation between the number of production workers, and the Average Yearly Wage. The weight of the correlation is very significant, only being eclipsed by the real GDP of energy. I hypothesize this is because the industries with the highest average pay require special skills or a higher level of risk. For example, a job on an oil rig is quite dangerous. As such, the Petroleum industry has one of the highest average Yearly Wages. The other high industries, such as the Chemical and Computer industries, also require high skill or a higher risk of workplace accident than the rest of the industries. The variables with the highest positive effect are the value of shipments and the invested capital. This makes sense as the more investment and profit a particular industry receives, the more they can afford to spend on wages.