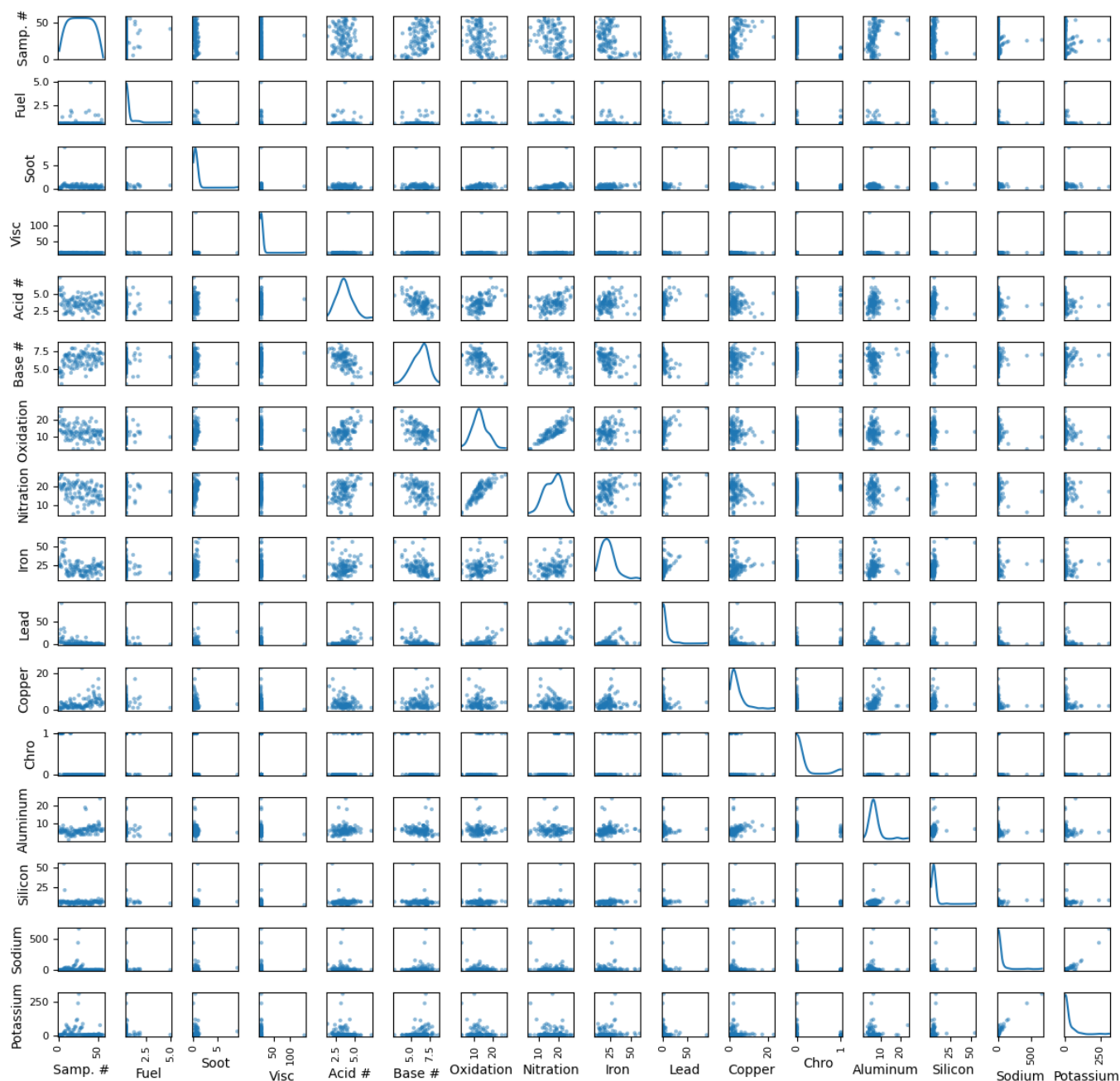# Research Questions

```
- Does an increase in degredation indicators  (oxidation/nitration) effect
engine wear metals?
- Does soot increasing correspond to more engine wear metals
- Do wear metals increase linearly or non-linearly
```

Data cleaning

The primary cleaning I had to do for this dataset was converting NaN values to numerics and wrangling the dataset to fit neatly into a dataframe by dropping the unnamed column and making the headers appear as proper headers. I will likely have to do more cleaning in the future, as there are clear outliers in the data, though further investigation is required to determine if they are erroneous or just extreme values. Future data cleaning I will need to do will be for joining the two excel sheets together in order to determine how wear will effect long run frequency at the fleet level. another data cleaning issue that exist that will need to be fixed is with the truck naming schema. Each truck is separated into #, and #B. One is erroneously named 1b instead of 1B.

Graph 1 & 2: Pairplot and Correlation Matrix

I created a pairplot for each of my covariates. This has been instrumental in deciding which variables might have a connection and will certainly help with modeling. This pairs nicely with a correlation heatmap. Looking at the diagnals we can see the distributions of each of the covariates. The vast majority of distributions are strongly right skewed, however some such as Base, and Nitration are left skewed. Another interesting detail is that it looks like base and acidity are mirrors of each other.

This pairplot also shows that many of the wear metals have many 0's, which indicates that they are rarely detected. This may raise class imbalance issues in the future for certain kinds of modeling, such as logistic regression.
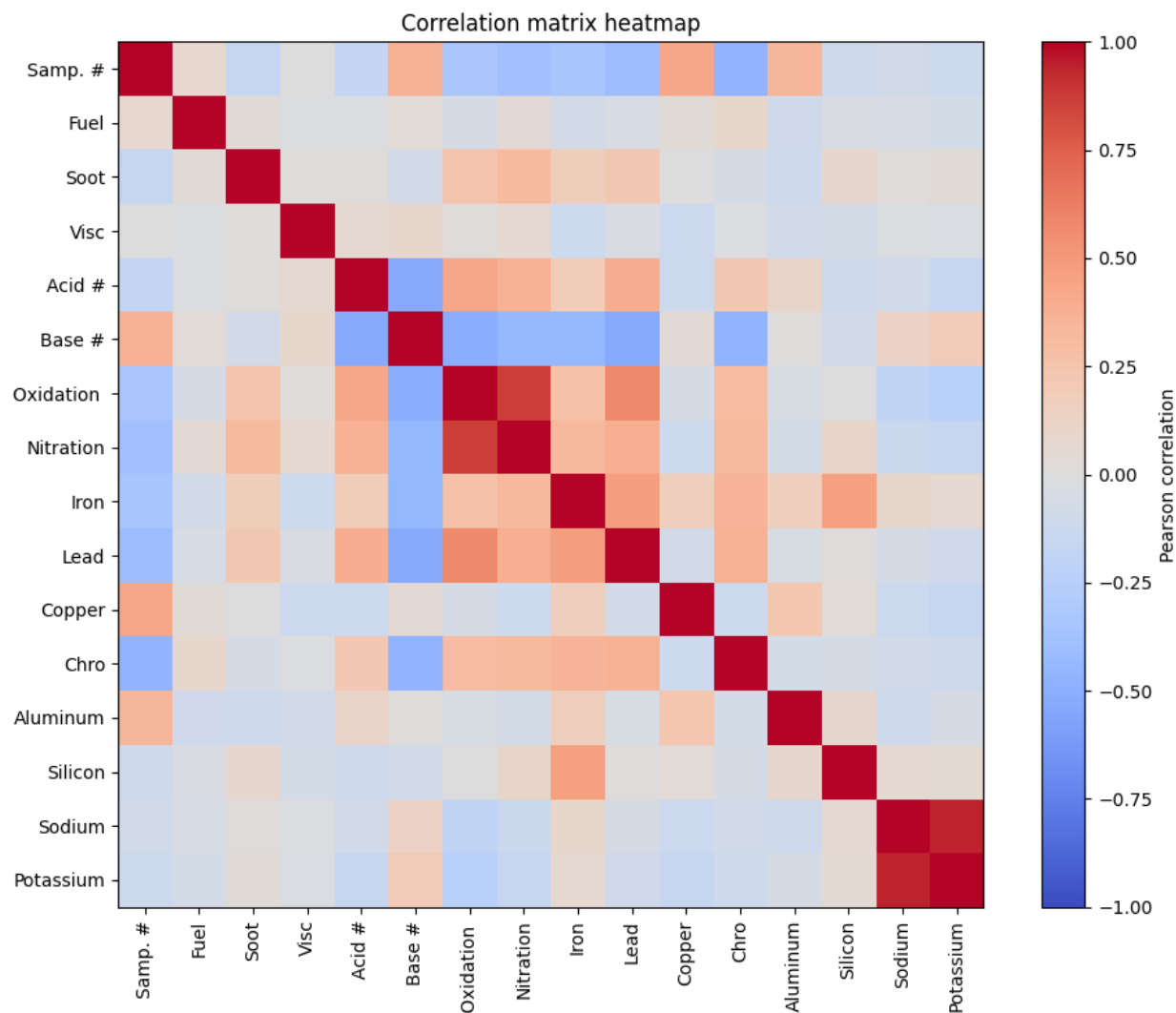
We also are able to see that there are many extreme values, which may need to be removed in the future. We also observe heavy heteroskedasticity, which will require non-linear methods to overcome.

For my first research question, "Does an increase in degredation indicators (oxidation/nitration) effect engine wear metals?", we can observe a clear positive association with iron. Relationships between degredation

indicators and copper and lead are weak, which may imply alternative wear mechanisms or interaction effects. We can also see Oxygen and Nitration are highly correlated, which we will address later on.

For my second research queston comparing soot vs engine wear metals we can see that soot shows a weak/moderate positive realtionship with iron and aluminum. The relationship is non-linear with lots of noise.

For research question three we can observe non-linear behaviors, specifically curving. Their diagnal distribution plots are heavily tailed, which would violate typical linear normality assumptions. These assumptions are also likely violated due the homeoscedasticity issues previously addressed.

This Pearson correlation heatmap helps see the strength between covariates. This graph shows mostly moderate relationships, which reinforces the idea that the true relationship between covariates is non-linear.

For my first research question we can see that oxidation and nitration show strong positive correlations with one another. This supports the hypothesis that chemical breakdown has some association with mechanical wear. Correlations with Copper and Lead are weaker, which suggests that not all wear metals respond similarly to chemical conditions.

For research question 2 we see that soot has a moderate/weak positive correlation with iron and lead. Overall the interaction with soot is modest at best, which signals that there is likely interactions with other factors, such as the chemical composition of the oil. In future modeling I would like to try and uncover how soot impacts wear metals when interacting with acidity/base and oxidation/nitration.
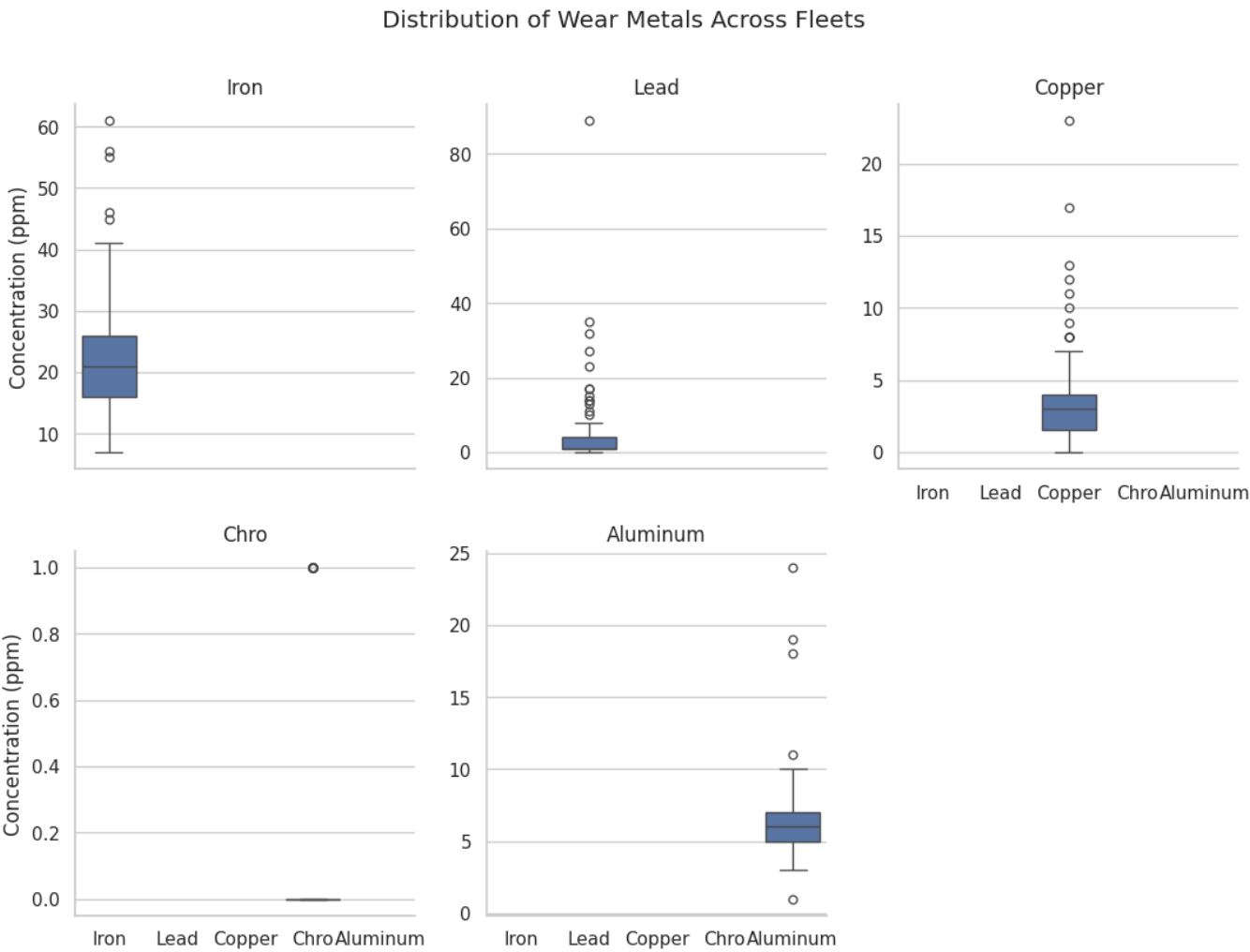
Table 3: Wear Metals Correlation

```
         Iron   Lead    Copper Chro   Aluminum
Iron     1.000  0.469    0.159  0.355      0.163
Lead     0.469  1.000   -0.078  0.360     -0.046
Copper   0.159 -0.078    1.000 -0.116      0.232
Chro     0.355  0.360   -0.116  1.000     -0.064
Aluminum 0.163 -0.046    0.232 -0.064      1.000
```

The strongest relationships between wear metals are Iron-Lead (.469), Iron-Chromium (.355), and Lead-Chromium (.36). Other relationships are very weak. Even though we did not get high correlation coefficients, this is still very significant for answering the research question because it implies that the relationship is not linear. This could imply that there are independent wear mechanisms, as if they were dependent we would expect the wear metals to increase at about the same rate, leading to large coefficients. This could also mean that there are differing wear rates, where different metals would start to be detectable at different parts of an engines lifespan. The weak correlation between copper and other metals might imply a non-linear relationship. Another interesting implication is due to the mixture negative, weak, and moderate correlation coefficients that some wear metals might increase together, like Iron-Lead-Chromium, and others wear independently of others.

Given these findings, my next steps will be to try and identify if there is a way to model these with linear models with interaction terms, as well as polynomial regression. I would also like to investigate if tree-based models would be suitable for predicting engine wear given metals.

Graph 4: Wear Metals Boxplot
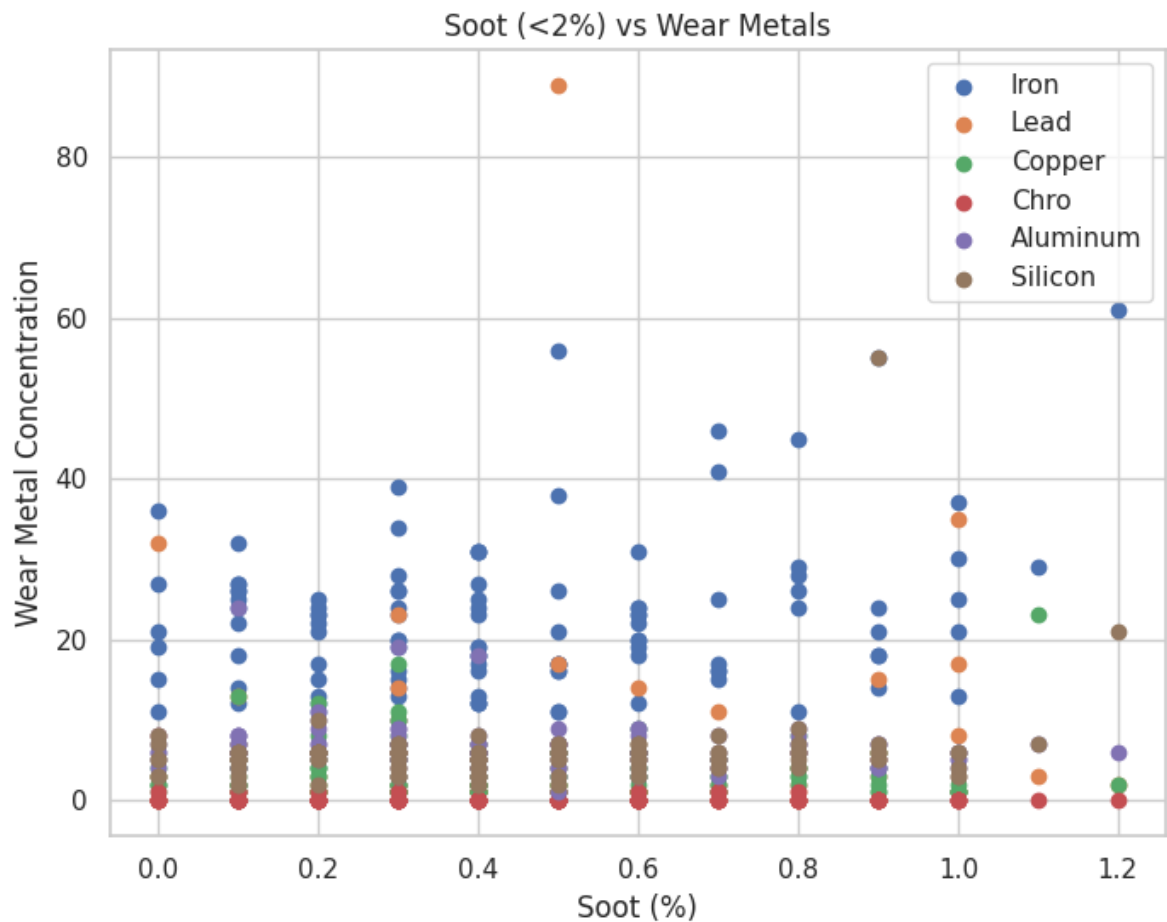
Distribution of Wear Metals Across Fleets

The most notable thing for this boxplot is the Chromium graph. All values are either 0 or 1 ppm. This indicates that Chromium is extremely rare. For this reason, it might be worth considering entirely dropping chromium from any modeling being done in the future. The other thing we can easily see is that all are right skewed and all have many extreme points beyond the threshold for outliers. This is consistent with our previous findings from the first graph where we determined they had very long tails, so these are likely not erroneous. The only suspect point may be the extreme outlier for lead, which is more than double the nearest data point.

Iron has the largest inner quartile range and highest median. Due to being more continuous and more various than other metals, I will use this as a primary predictor going forwards.

Lead and copper both have low medians but several large outliers, which suggests rare but extreme events causing degredation rather than a continuous degredation. Using my domain knowledge from working at the Department of Transportation Maintenance Bureau, I believe this is likely due to the concentration of lead and copper in bearings used, and potential incidents like crashes or large bumps, which may cause large spikes in wear.

Graph 5: Soot Plot

This is a plot for soot compared to wear metals where soot concentration is below 2%. The soot has few points around the 8% range, but inclusion makes the graph impossible to read. Using this graph we are able to tell that there is no strong clear relationship between soot and wear metals. Wear appears to be driven either by other factors, such as operating conditions, load, contamination, or maintenance related factors than wear metals. I may reconsider this research question as there is not strong evidence to suggest a meaningful relationship.

Iron has the largest spread among the wear metals, and has a mild upward trend with higher soot values, but also larger variance.

Synopsis

My project will analyze oil conditions and wear data from Class 8 Heavy-Duty vehicles to research how degredation indicators relate to engine wear, and whether these relationships are linear or non-linear across fleets.

Keywords

1. Heavy-Duty Vehicles
2. Oil Degredation
3. Engine Wear

4. Vehicle Maintenance
5. Fuel Efficiency
6. Contamination