

NAME: Oluwasina Olowookere
STUDENT NO.: 1006648068
COURSE: MSE238
TITLE: Project Proposal

Introduction

This proposal will highlight what procedures produce a good engineering alloy by analyzing the Multi Principle Element Alloy Dataset. The objectives of the project are to visualize the data and research the role a materials processing procedure plays in its yield strength. Statistical procedures such as data visualizing, line fitting, standard error and summary statistics will be used to classify and the data will be visualized using functions such as ggplot, geom and facet. The anticipated outcome is a set of ordered visualized data that shows a clear relationship between the processing procedures and the material's yield strength. The project also references the Beaches dataset and gives an index as to which locations are safe to visit based on the level of toxic ecoli present at the location. It will give data such as the mean ecoli toxicity and uses data visualization to show relationships.

1. Problem definition

This project is centred around the question; what procedures produce a good engineering alloy? The generic To answer this is; if a material has a high yield strength. A high Yield Strength implies that a material can undergo a certain amount of stress without deforming plastically. Once materials pass their Ultimate Tensile Stress, they start to neck and then continue necking until fracture. We will be analyzing various factors such as the composition of the alloy, whether it is single or multiphased and the processing method in order to determine their relationship using the data Multi Principle Element Alloy Dataset.

The second part of the project is centred around determining information such as average toxicity levels for ecoli at specific locations and determining which locations are more prone to be more dangerous. The purpose of this aspect of the project is to create an index where users can determine whether a beach is safe to visit based on the levels of e coli present at the beach.

2. Objectives

The objectives for the MPEA dataset are to visualize the data and research the role a materials processing procedure plays in its yield strength. We need to visualize the data so we can determine if there are noticeable relationships between variables. An example of data visualization would be plotting the temperature with respect to the yield strength and colour coding it with respect to the phase types as seen in Figure 1.

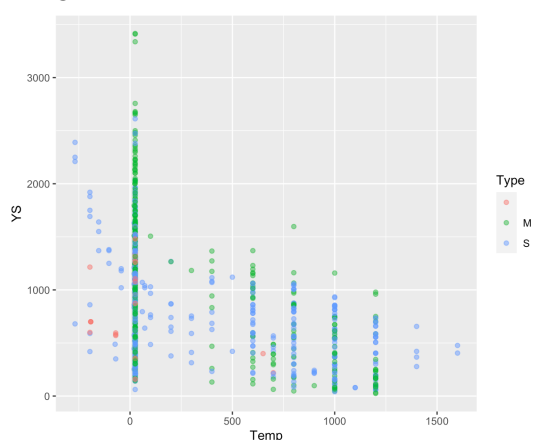


Figure 1. (Plot of Temperature VS Yield Strength)

The plot above shows a general trend that suggests that as test temperature increases, the yield strength decreases. The project will contain more trends like this and after they are identified, research will be conducted to provide some reasoning as to why the trend is observed.

The Objective for the Beaches dataset are to visualize the given data in meaningful ways, construct relevant control charts and make any inferences you can draw from the data. This can be helpful when determining information such as mean ecoli toxicity levels and locations that are more toxic to be in than the others.

3. Description of statistical techniques/theory to be used

Statistical techniques that will be used in the project will are;

- Construction of box plots with respect to the temperature and the yield strength which would show us data such as the outliers and the median. Histograms will be used to get the frequency of a particular phase in an alloy so we can determine the dominant phase of that alloy. Histograms can also be used to determine the dominant processing method.
- Determining the line of least squares in plots and standard error reporting
- Data summary techniques such as determining the average grain size in an alloy with multiple (i.e. mean, median).

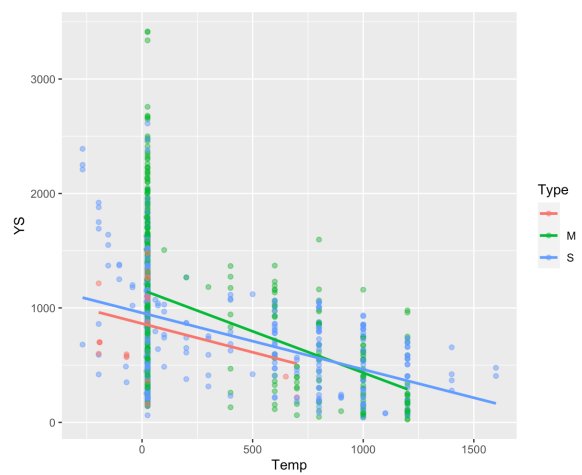


Figure 2. (linear modelling of the plot above)

- A better technique that will be used to explore relationships between the independent variables in and the dependent variable (yield strength) is the multiple regression technique.

4. R packages/modules planning to use

In this project, I plan to use functions such as `read.csv` in order to import the csv files provided and perform analysis on them. There will also be a bit of data visualization, and for this to be possible, I will need to import the 'tidyverse' library. The 'tidyverse' library gives me access to visualization functions such as `ggplot`, `geom`, `facet` and the ability to customize plots in such a way that makes relationships clear.

It may also be possible to write a function that will read the data from the Beaches dataset csv file to extract the data into a new data frame and plot it.

The data above shows the relationship between temperature and yield strength, colour coded based on the type of phases and faceted based on the composition of the phases. This type of data will allow me to reach reasonable conclusions.