Visualization Project

Civil Engineering: Cement Manufacturing Dataset

Shubhangi Sanyal
Rollno: MDS202238
Email: shubhangi@cmi.ac.in (mailto:shubhangi@cmi.ac.in)

Date: 29th October, 2022

Abstract

In civil engineering, concrete forms a major part of most projects. Among all materials used in construction, concrete is the most crucial to any structure. Measuring and understanding the concrete compressive strength, thus, is vital. It has been observed that the concrete compressive strength is a highly nonlinear function of age and ingredients. Some of these ingredients are cement, blast furnace slag, fly ash, water, superplasticizer, coarse aggregate, and fine aggregate.

Dataset Description

This dataset has been taken from *Kaggle*. The actual concrete compressive strength (MPa - Mega Pascal) for a given mixture under a specific age (days) was determined from a laboratory. The data has 8 quantitative input variables, 1 quantitative output variable, and 1030 instances (observations). The variables can be described as:

- `cement` (Cement) : measured in kg in a m^3 mixture
- `slag` (Blast) : measured in kg in a m^3 mixture
- `ash` (Fly ash) : measured in kg in a m^3 mixture
- `water` (Water) : measured in kg in a m^3 mixture
- `superplastic` (Superplasticizer) : measured in kg in a m^3 mixture
- `coarseagg` (Coarse Aggregate) : measured in kg in a m^3 mixture
- `fineagg` (Fine Aggregate) : measured in kg in a m^3 mixture
- `age` (Age) : day (1~365)
- `strength` (Concrete compressive strength): measured in MPa

Objectives

Using the data described above, data visualization can be used to determine the following:

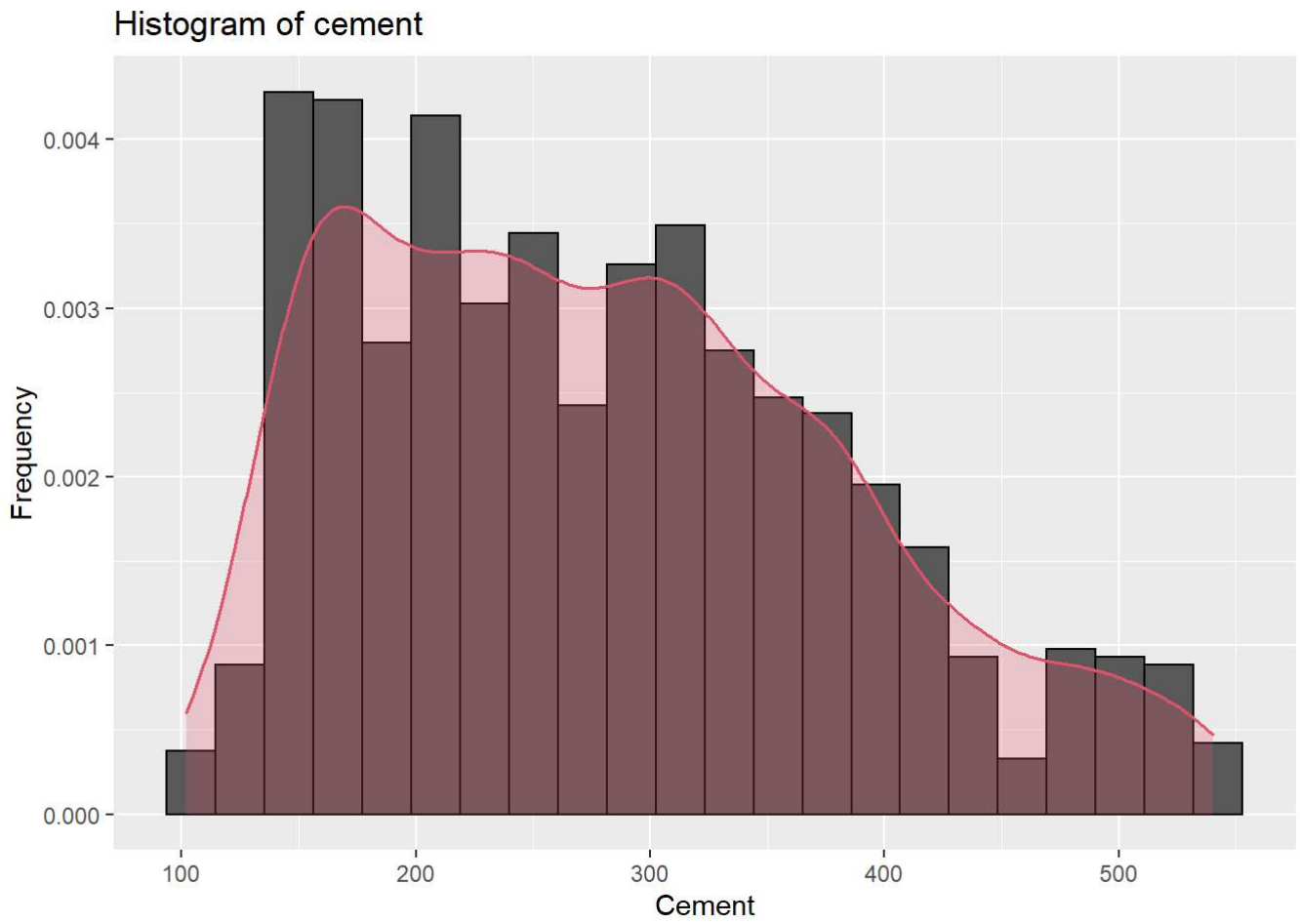
- How each ingredient is distributed
- How the proportion of each ingredient varies with different values of concrete compressive strength
- How the proportion of each ingredient varies with age
- How each ingredient proportion varies with each other when any two are taken at a time
- Finally, how any two ingredient behaves when compared to the concrete compressive strength.

The above objectives can be achieved using histograms, and boxplots for univariate analysis, scatter plots and line plots for bivariate analysis, and gradient scatter plots for multivariate analysis.

Univariate Analysis

Histogram

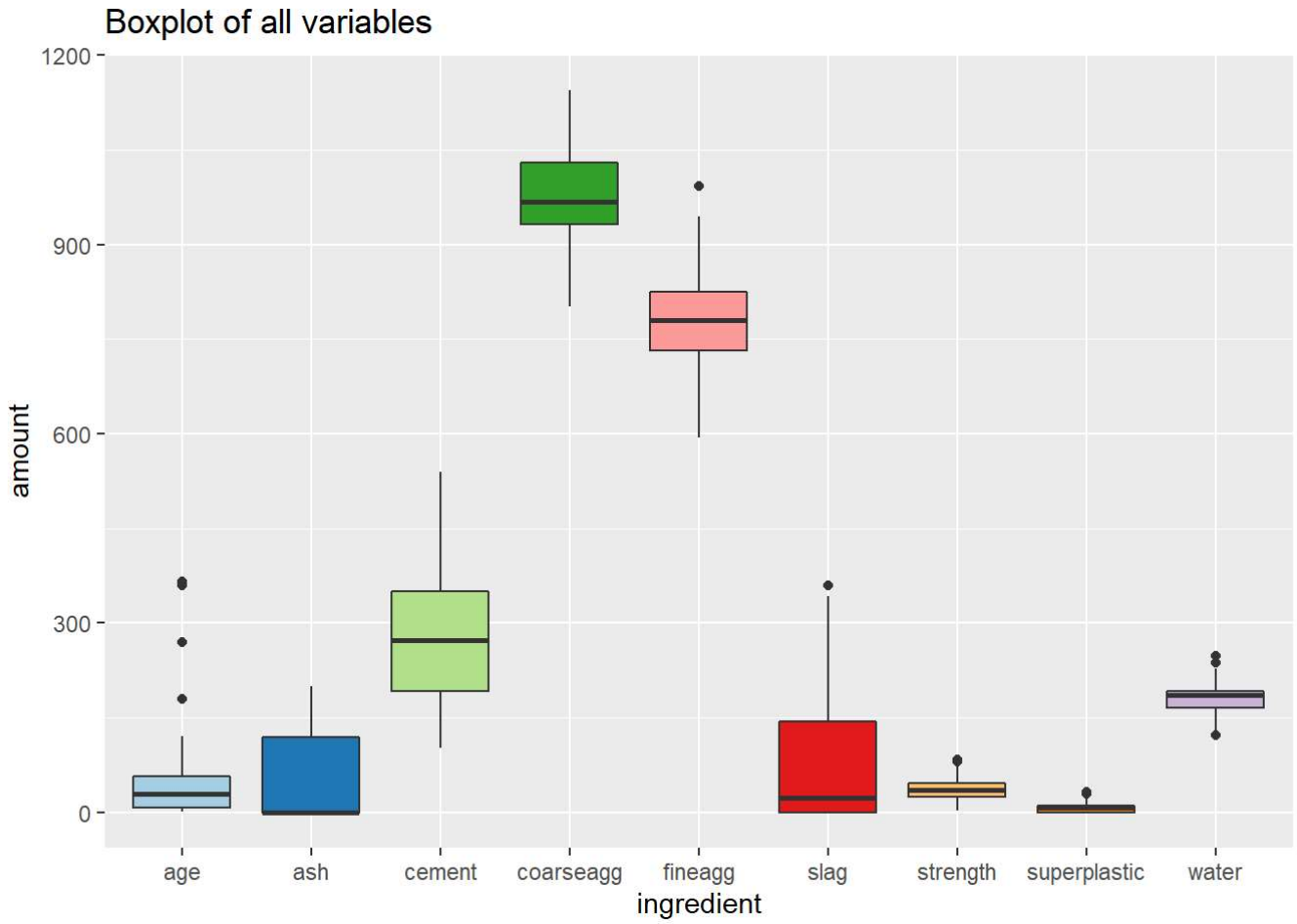
A histogram can be quite useful in visualizing the distribution of each ingredient across a 1030 instances. The following shows a sample of such histograms by plotting the cement values of all observations.



From the above figure, it is evident that most of the time the amount of cement varies from 140 to 160.

Boxplot of all attributes

The boxplot of all the attributes (input and output) gives us a general idea of the range of their values and provides more descriptive statistical measures like mean, median, maximum, and minimum which can be useful for creating concrete mixtures. Further, it is useful in spotting outliers in the data. The following gives the boxplot described above.

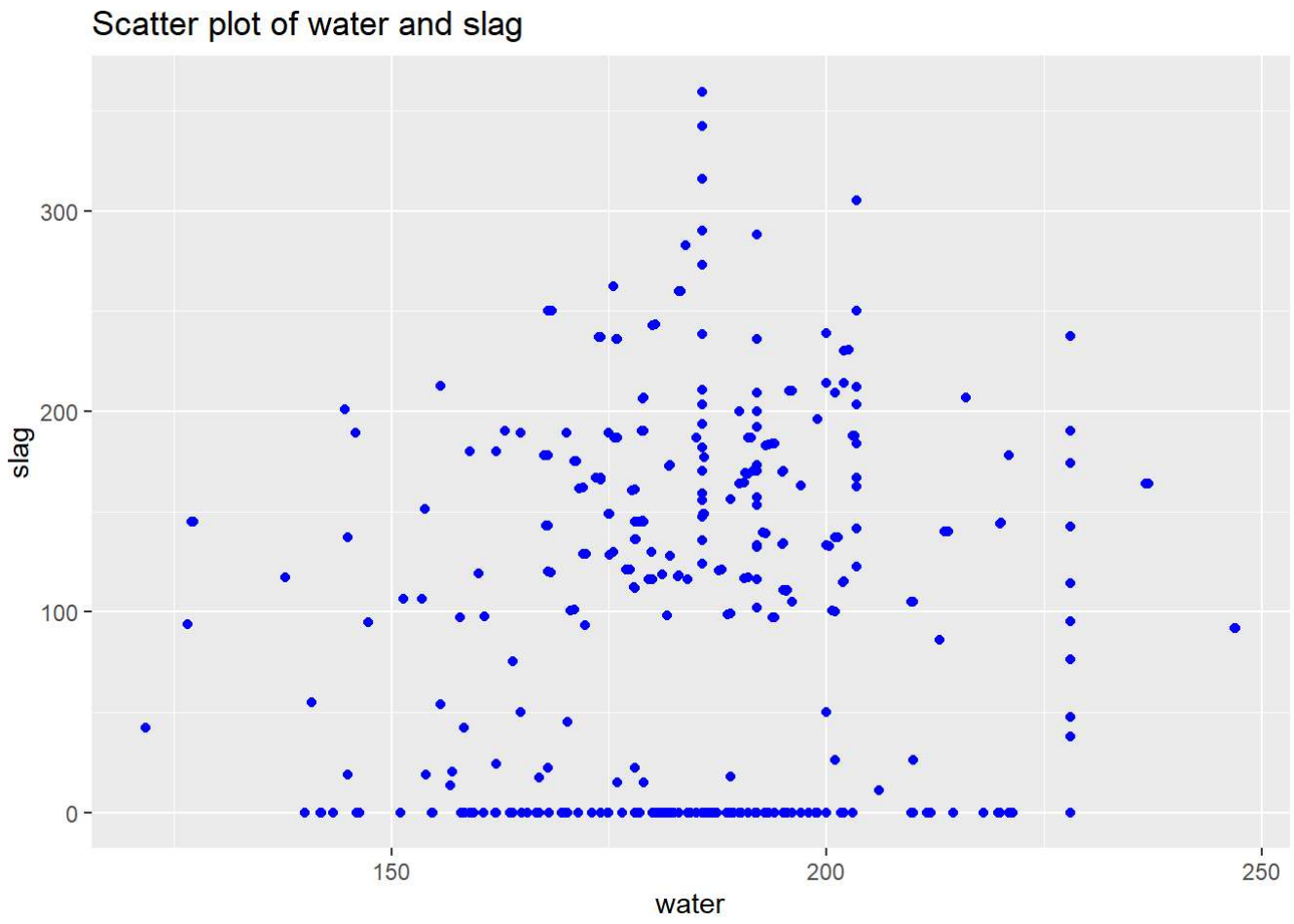


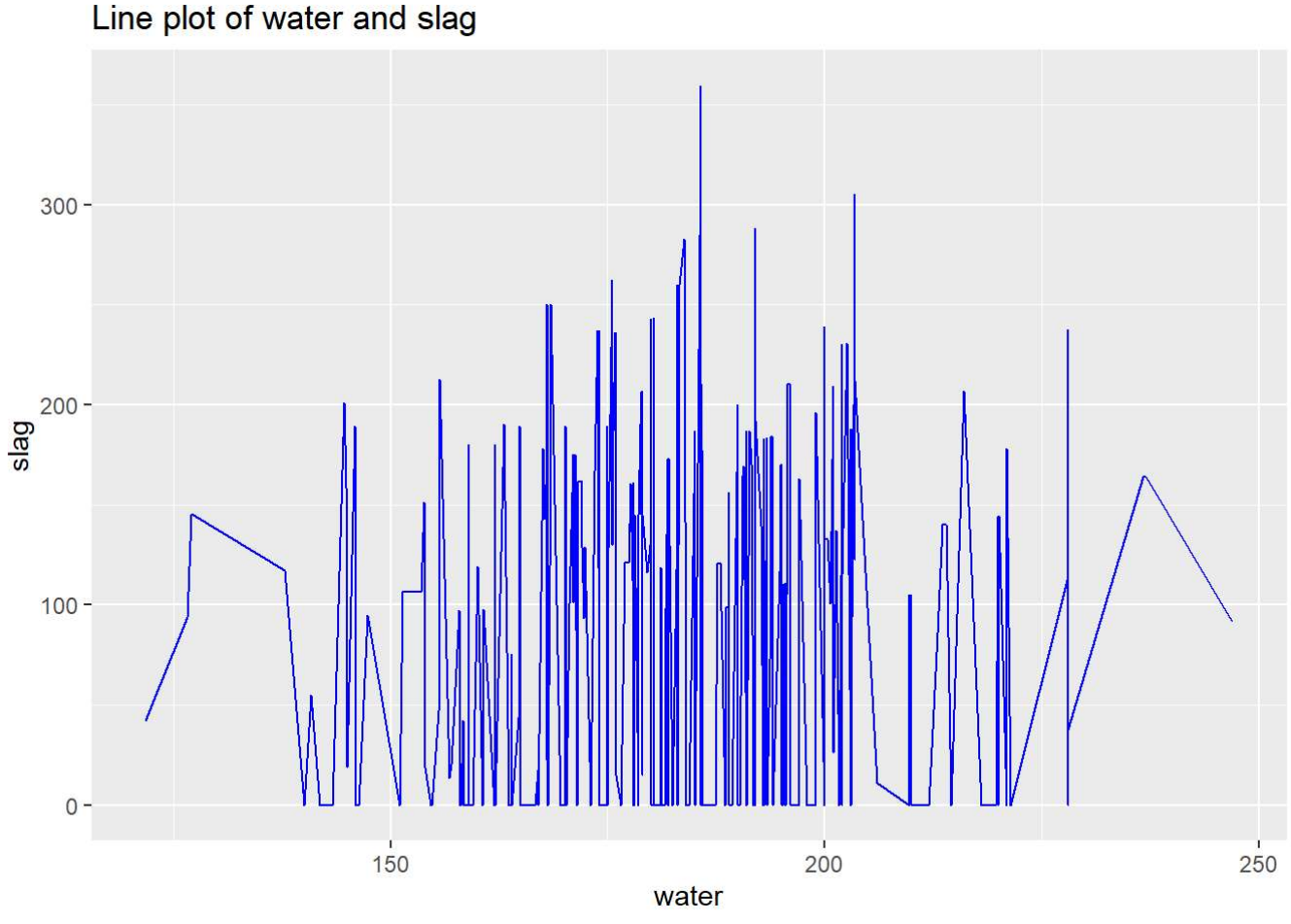
The boxplot above provides certain information about each attribute. For example, in the case of fly ash (`ash`) the maximum, minimum, first quartile, median, and third quartile are 200.1, 0, 0, 0, and 118.3, respectively. Also, `age` , `fineagg` , `slag` , `superplastic` , and `water` show the presence of outliers.

Bivariate Analysis

Relationship between any two ingredients

The relationship between any two ingredients that are used in the composition of concrete can be observed using a scatterplot and a line plot. For further examination, more statistical tests are required. The following gives the scatter and line plot for the ingredients ‘water’ and ‘slag’.

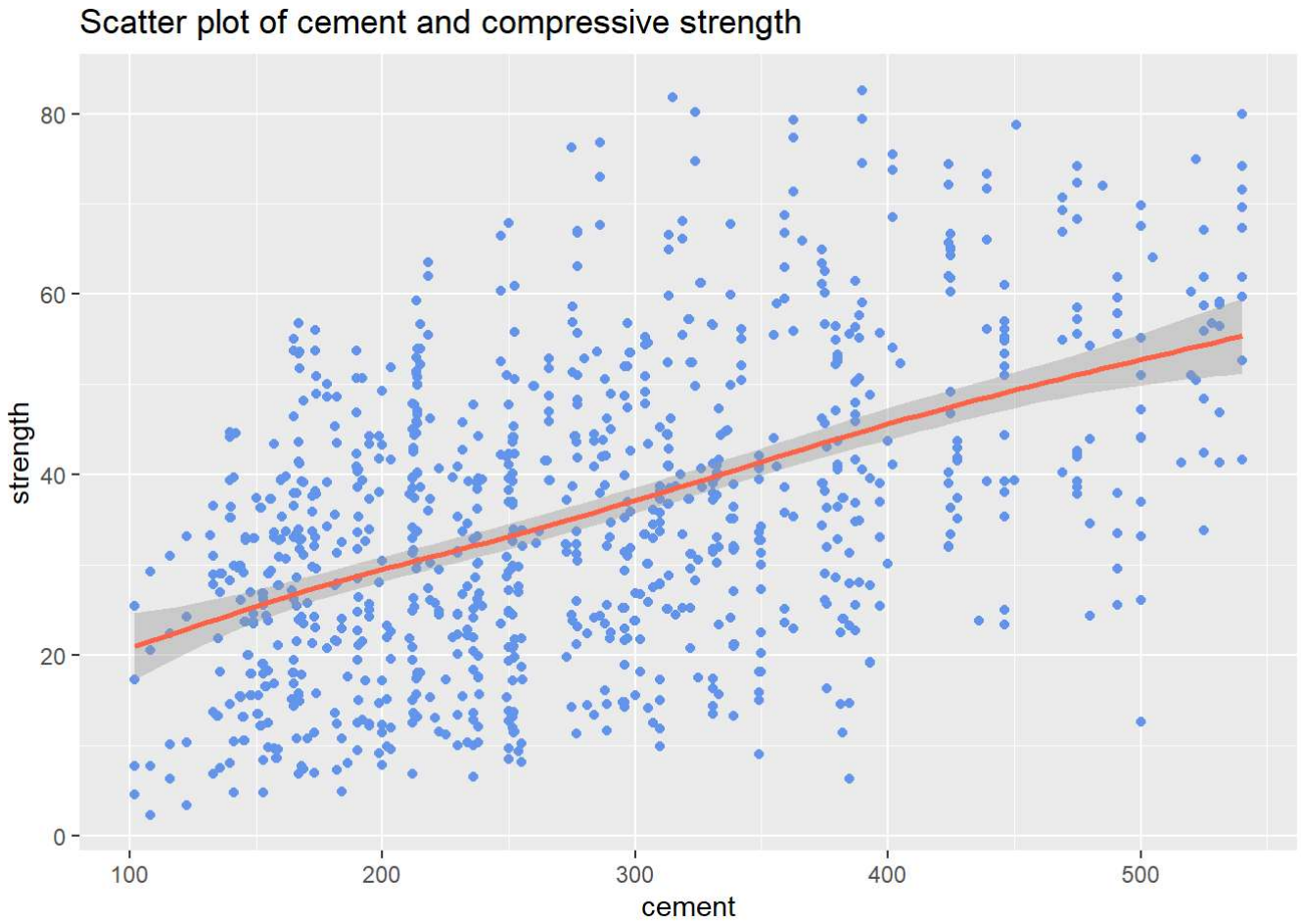

Scatter plot of water and slag
slag
300
200
100
0
150
200
250
water


Line plot of water and slag
slag
300
200
100
0
150
200
250
water

Both the scatter plot and the line plot of `water` vs. `slag` show that there is no significant correlation between the two attributes.

Relationship of an ingredient with concrete compressive strength

Further, an ingredient's change of behaviour with respect to the concrete compressive strength which is the output variable can be observed through a scatter plot. Additionally, a line of best fit has been added to the plots to better conclude their relationship.

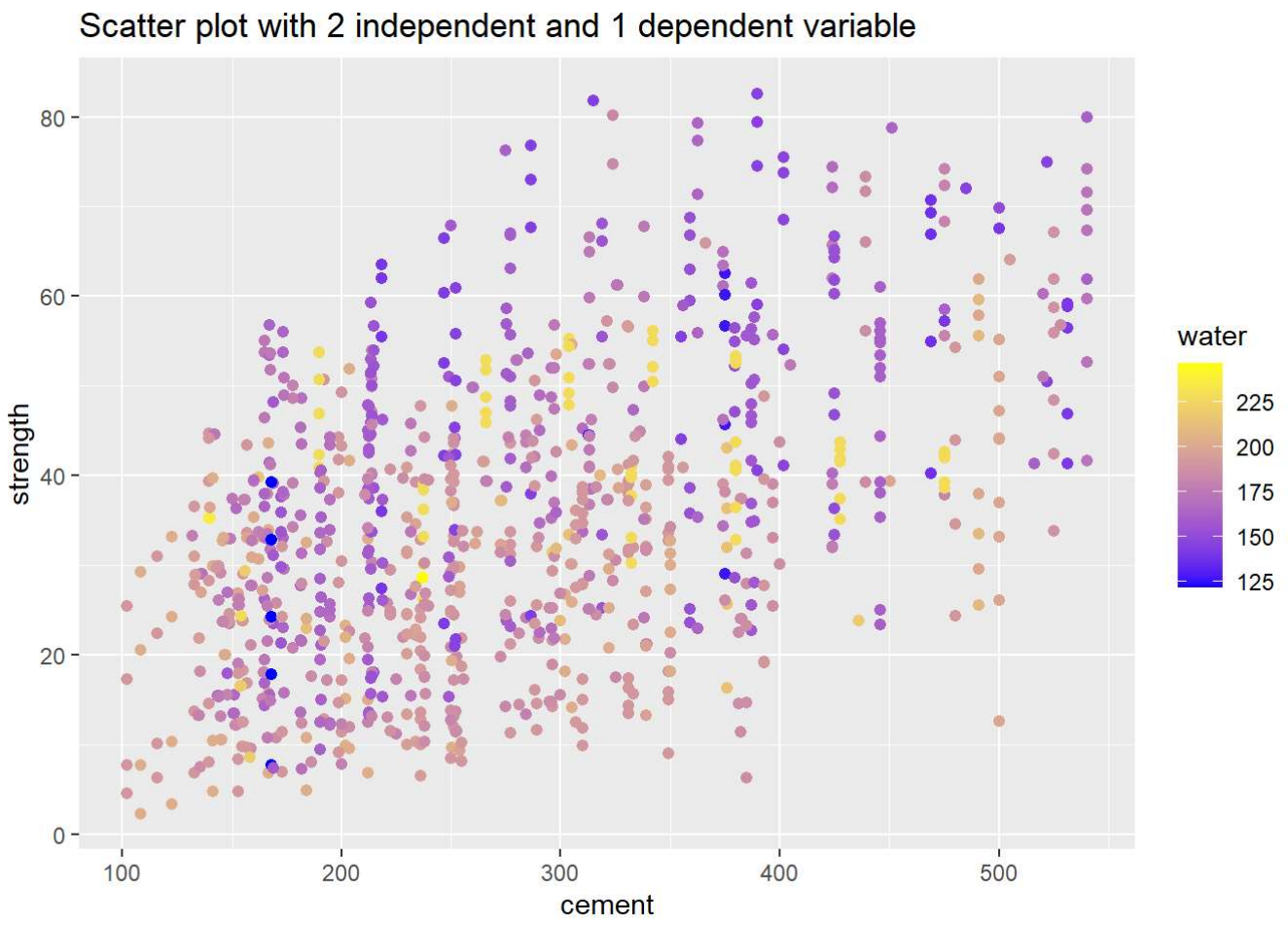


The above plot shows that with the increase in the amount of cement the average concrete compressive strength increases. The best fit line shows that cement is positively correlated with strength.

Multivariate Analysis

Relationship of any two ingredients with concrete compressive strength

In a more complicated analysis, the scatter plot is used to map the behaviour of not just one but two ingredients against the compressive strength. This is done by using a colour gradient on the scatter plot as seen in the following scatter plot.



Along with the positive correlation of cement and compressive strength as seen in the previous plot, it is also observed that water concentration decreases with increase in strength and cement.

Conclusion

The following points summarize the observations in this report:

- Histogram provides the distribution of every attribute. For cement, the histogram observes that cement concentration varies from 140 to 160 units most of the time.
- Correlation between `water` and `slag` is 0.107.
- Correlation between `cement` and `strength` is 0.498.
- Correlation between `water` and `strength` is -0.289.

References

- https://www.kaggle.com/datasets/vinayakshanawad/cement-manufacturing-concrete-dataset (https://www.kaggle.com/datasets/vinayakshanawad/cement-manufacturing-concrete-dataset)