Title: Exploring two or more values

Introduction

In this assignment we will describe two or more values with charts and plots like violin and line. Furthermore, there will be an explanation of each made chart or plot. The main dataset was chosen as the agriculture_8.csv file. The head description is here

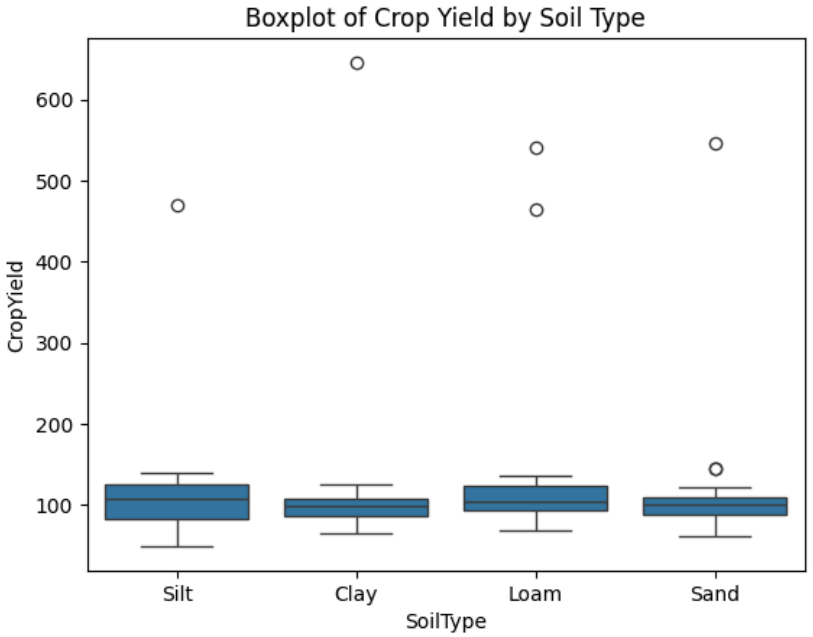
	FarmID	CropYield	WaterUsage	FertilizerAmount	SoilType
0	0	135.281047	688.315070	46.308182	Silt
1	1	108.003144	365.224094	47.606208	Clay
2	2	119.574760	372.951500	60.996596	Loam
3	3	144.817864	596.939671	56.552637	Sand
4	4	137.351160	382.687659	56.401315	Silt

Mostly, we will use one categorical data and one numerical. As numeric data we will take the CropYield feature.

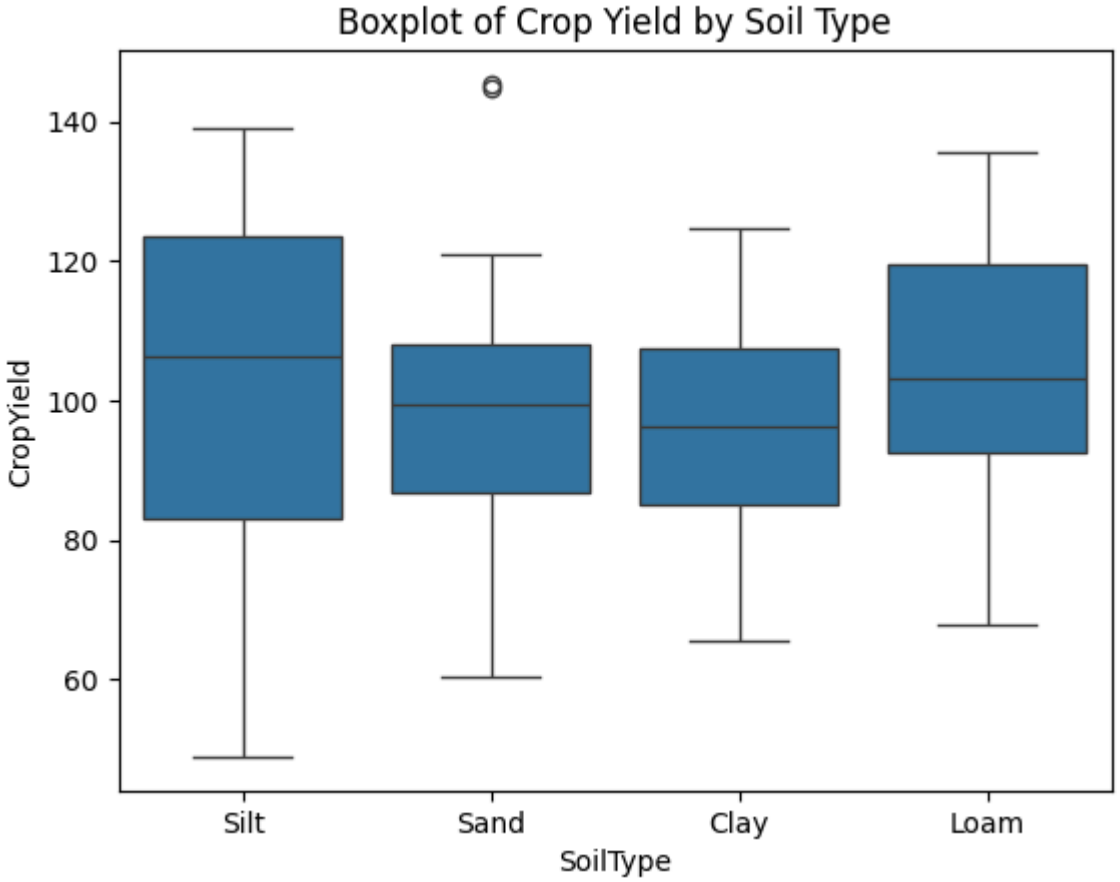
Methods

Firstly, will show the box plots grouped by categorical data, looking at this we can identify how far our outliers are from normal values. Then, I will draw violin plots to see the distribution as well with box plot information. With violin plots there will be clear the shape of distribution. Then, pair plot, to see the relations between some numeric features' values.

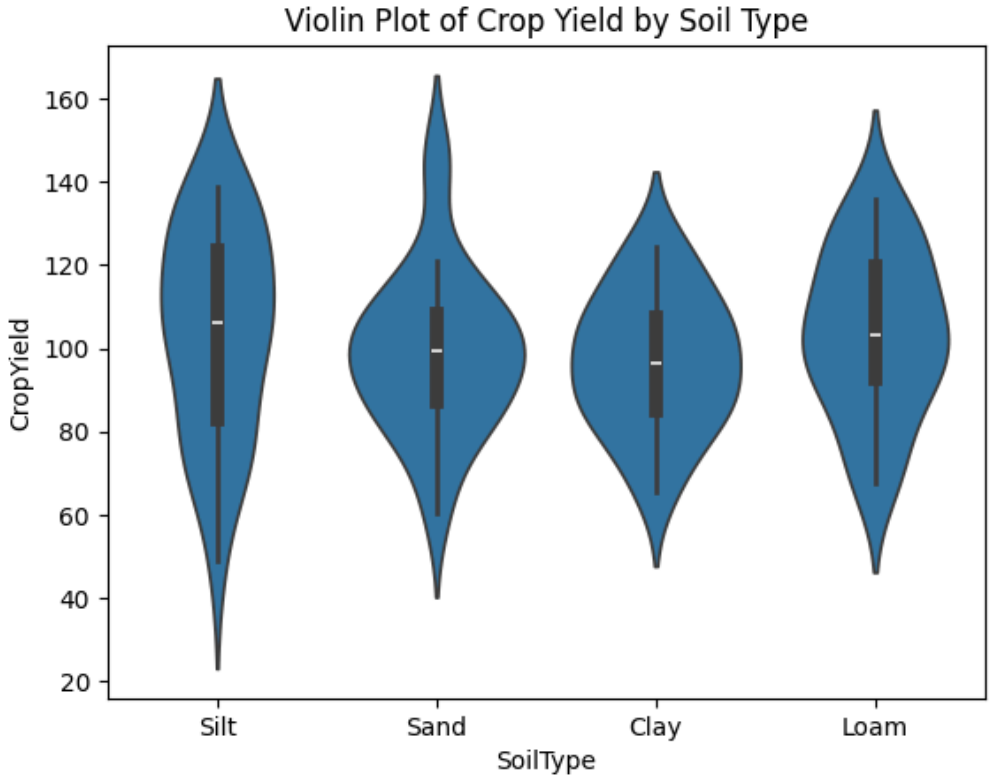
Findings



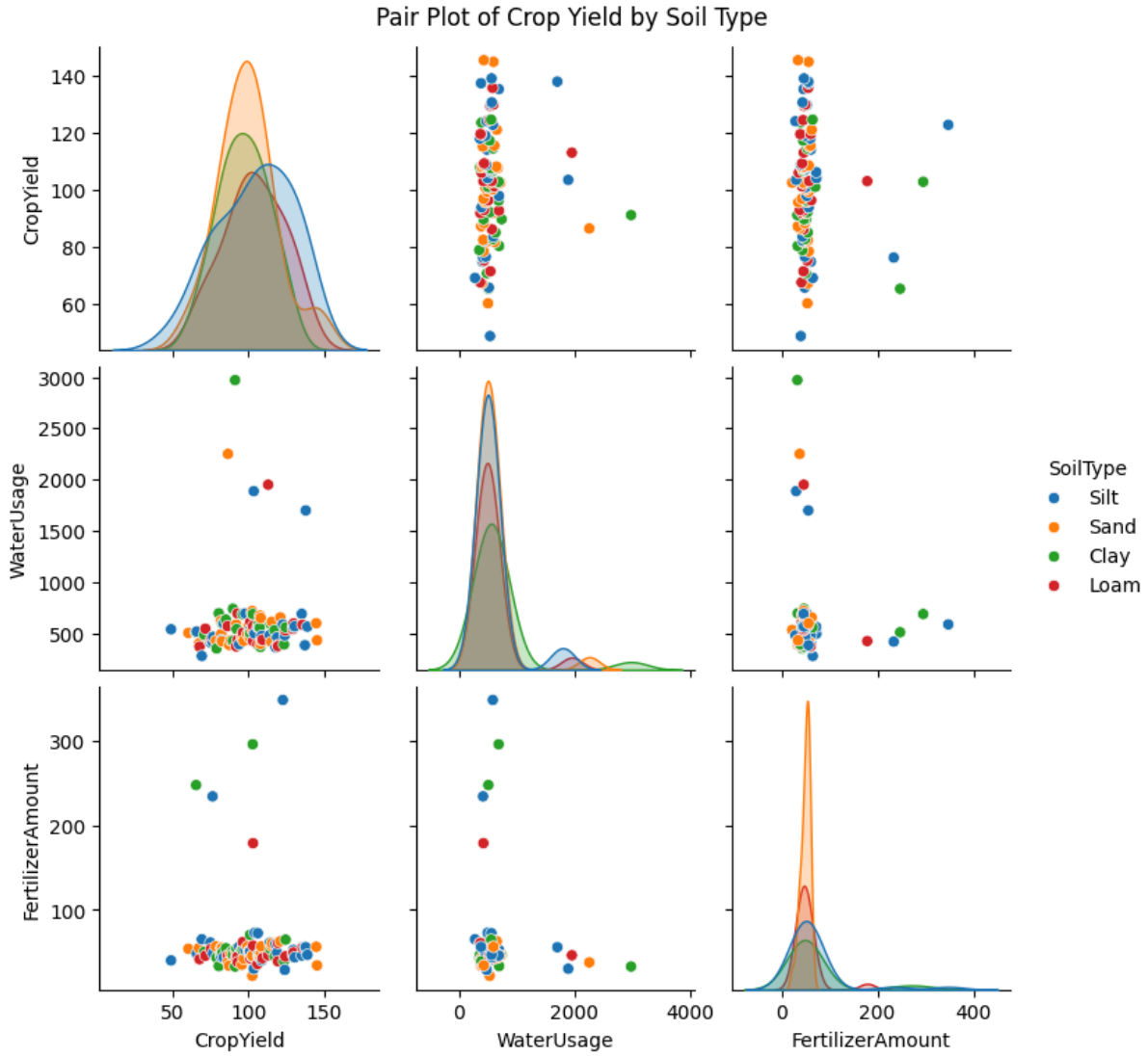
The box plot from the simple dataframe does not look so good because of extremely big outliers. Let's trim them and looks one more to box plot



Now, our data normalized and shows a normal box plot with only two not far outliers in the sand category. The biggest range between two bounders is in the Silt category. So, more clarification, let's see the violin plot



Silt and Loam soil types have wider distributions compared to the others. And the media of each category is almost similar and it means that the central crop yields are close to each other.Because of trimming our df we don't have unusual skewed tails of violins.
Now, let’s look at to pair plot



There I took features for comparison like CropYield, WaterUsage and FertilizerAmount and as a category use SoilType. Crop yields for all soil categories follow a roughly normal distribution, with Sand having a sharper peak compared to others. And there is no clear linear relationship between any features.

Conclusion

There is no strong correlation between crop yield and either water or fertilizer usage, and it indicates that increased resource use does not necessarily result in higher yields.