Second Exercise
Data set visualization

Karolis Kvedaravičius

Vilnius 2025

1. Data set

In this task a wine quality data set is visualized using three different direct visualization methods.

The data set has the attributes that describe a wines chemical composition and its subjective quality.

1. fixed acidity
2. volatile acidity
3. citric acid
4. residual sugar
5. chlorides
6. free sulfur dioxide
7. total sulfur dioxide
8. Density
9. pH
10. sulphates
11. alcohol
12. quality (rating between 0 and 10)

2. Visualizations

The three visualizations are provided below. When performing all these visualizations the data was normalized using min/max scaling to a range of 0 to 1.

2.1.   Radial Visualization
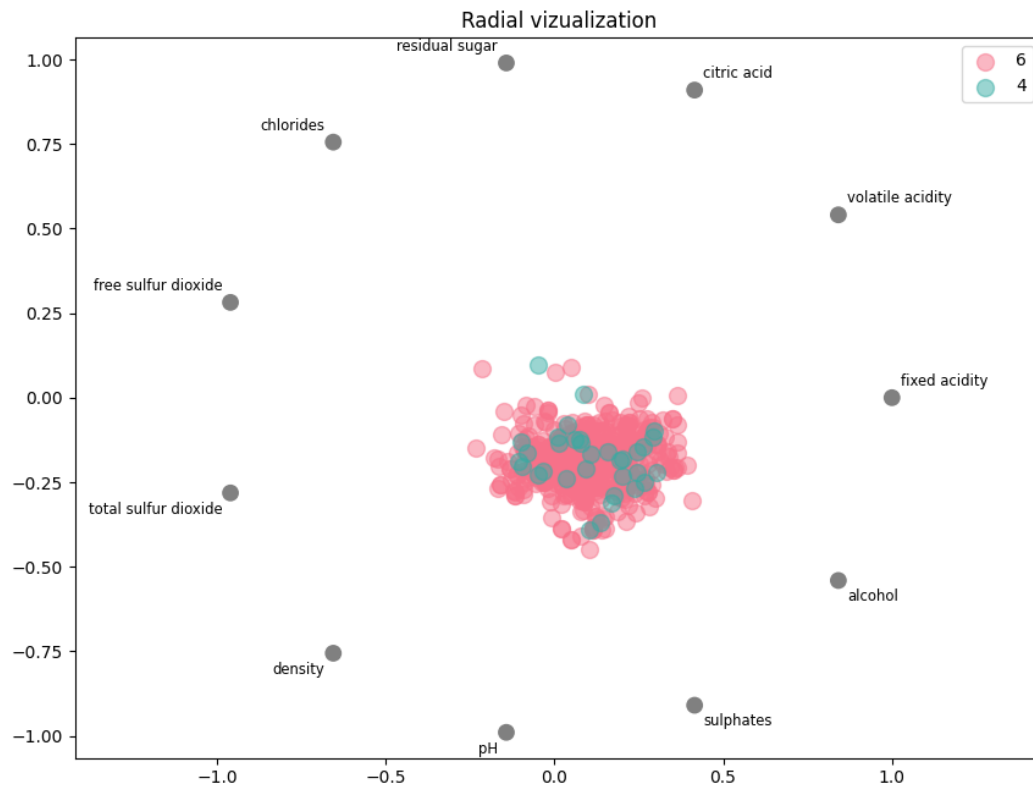
Figure 1 radial visualization of wines rated 6 and 4

The goal of this visualization was to check if wines of different classes lean more heavily to one direction or another. In the case of wines which were given a score of 4 and 6 there is a lot of overlap, which makes it difficult to notice any differences, but it does visualize that there are significantly more wines of quality of 6 than quality of 4.
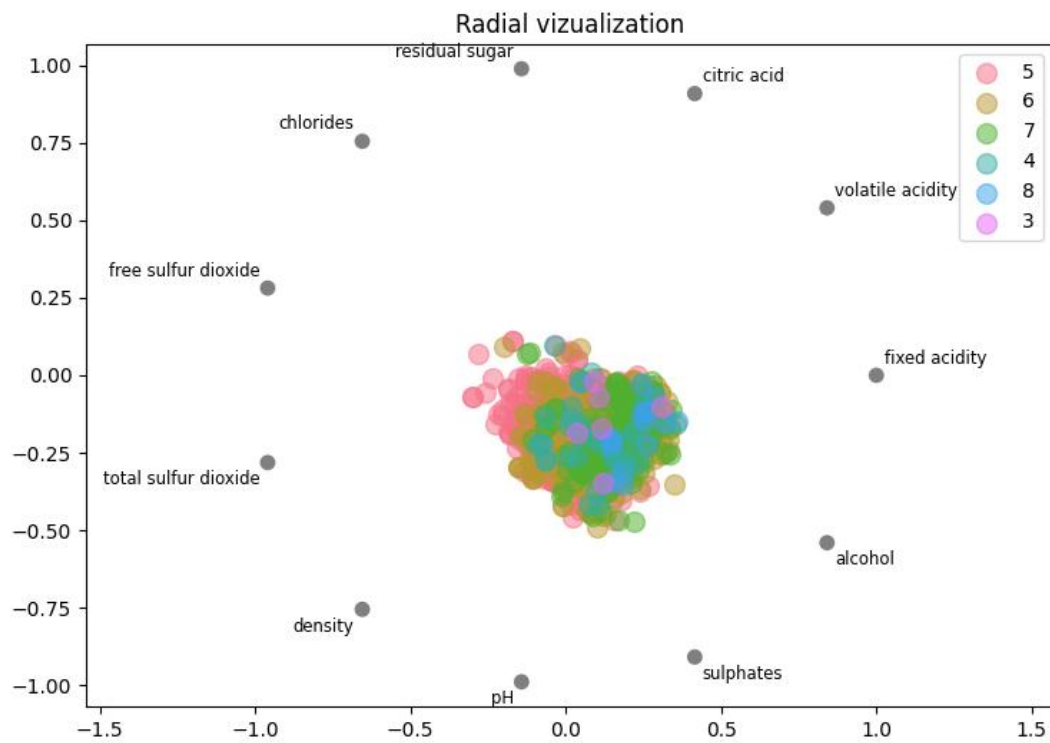
Figure 2 radial visualization of the entire wine data set

The rest of the wines where also visualized, but diagram becomes quite hard to parse. Maybe a lower number of features would improve this visualization.
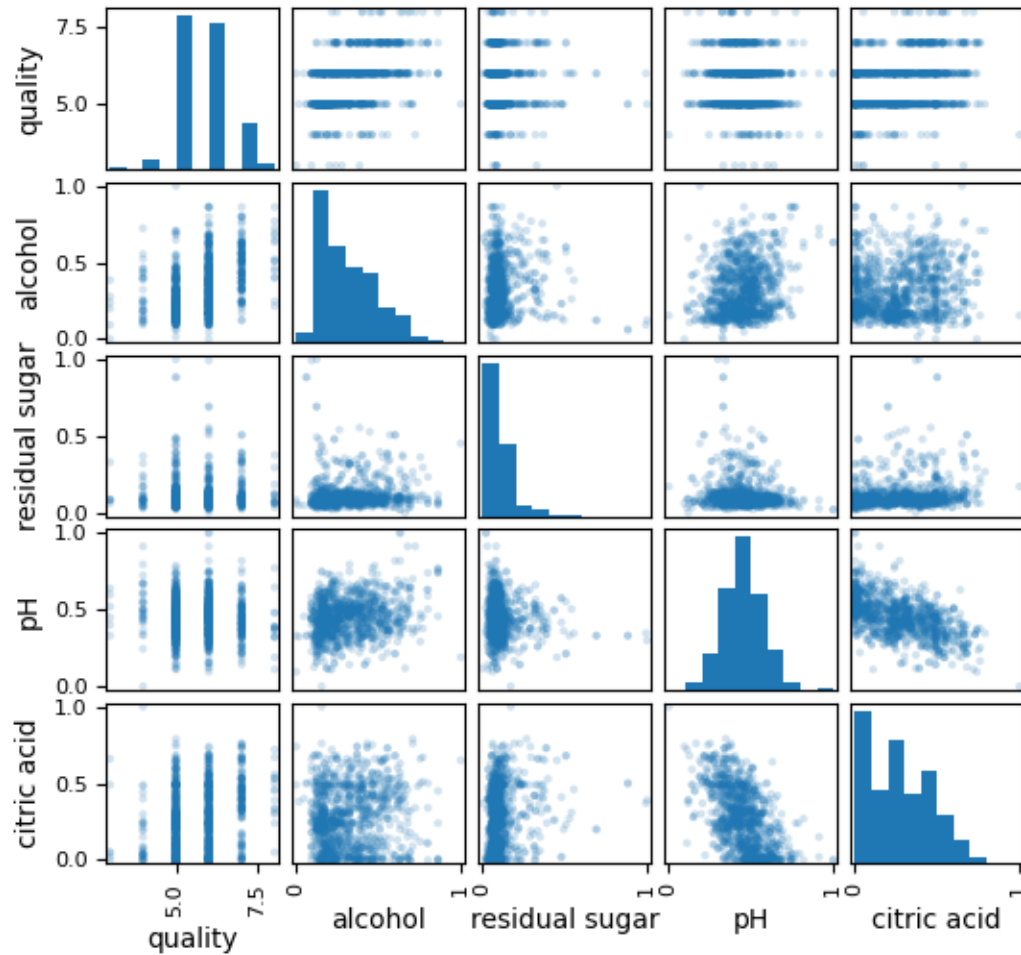
2.2.   Scatter plot matrix

Figure 3 scatter plot visualization of select attributes

A subset of the data set features (citric acid, pH, residual sugar, quality) where visualized using a scatter plot matrix. From it several insights can be gleamed. Firstly, higher quality wines (score of 7 and 8) seen to have larger citric acid values than lower quality wines, also higher quality wines seem to have higher average alcohol amount. In comparison it's harder to make definitive statements about residual sugar because they are similarly clustered around the bottom of the residual sugar values.
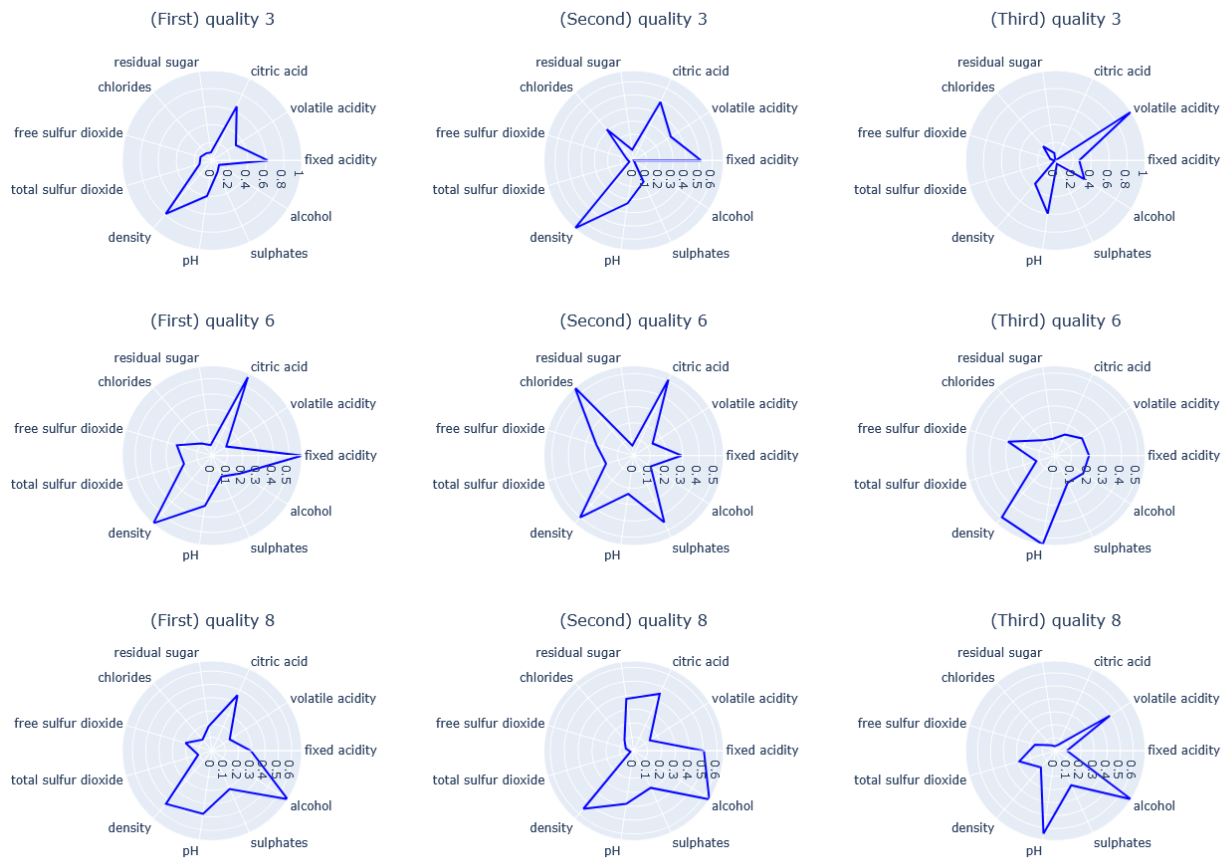
2.3.   Radar visualization

Figure 4 radar visualizations of different quality wines

In this diagram a few select wines where visualized as a matrix of radar graphs. The goal was to see if we can discern a "shape" which wines of different quality take when visualized Like this. We can see that the selected higher quality wines have a higher alcohol amount than the other wines. We can also see the low-quality wines have a lower pH level.

3. Results

The performed visualizations did help some insights into data. For example, from the scatterplot matrix and the radar visualizations we can see that the alcohol amount has a noticeable effect on the quality rating of the wine. The radial visualization was the least useful, it was difficult to differentiate between the different qualities of wine because they were highly clustered, maybe the dataset has too many dimensions for this kind of visualization