IC 272: Lab1: Data visualization and statistics from data

A dataset related to red variants of the Portuguese "Vinho Verde" wine is given as a csv file. This dataset contains the values of different physicochemical tests from each samples of red wine [1]. The original goal of the dataset is to model wine quality based on physicochemical tests. The attributes of the dataset based on physicochemical tests are fixed acidity, volatile acidity, citric acid, residual sugar, chlorides, free sulfur dioxide, total sulfur dioxide, density, pH value, sulphates, alcohol content and last attribute is on quality. Each expert graded the wine quality between 0 (very bad) and 10 (very excellent).

Write a python program (with pandas) to read the given data and display following:

- Mean, median, mode, minimum, maximum and standard deviation for the all the attributes.
- Obtain the scatter plot for all the attributes based on physicochemical tests and quality (You can use matplotlib library).
- Find the value of the Pearson correlation coefficient between all the attributes based on physicochemical tests and quality
- Plot the histogram for all the attributes (You may use "hist" function from pandas)
- Group the data according the attribute 'quality' and plot the histogram for attribute 'pH' for each quality value (Use *"groupby"* function)
- Obtain the boxplot for the all the attributes (Use "boxplot" function).

Reference:

[1] P. Cortez, A. Cerdeira, F. Almeida, T. Matos and J. Reis. "*Modeling wine preferences by data mining from physicochemical properties*," In Decision Support Systems, Elsevier, vol. 47, issue 4, pp. 547-553, 2009.