Data Science Homework 2

Matt McCullough

1a.

Chart, scatter chart

Description automatically generated

The relationship between fixed acidity and pH is a negative linear relationship.

1b.

Chart, scatter chart

Description automatically generated

I can say that the pH and fixed acidity of the wines with low citric acid content is that they have higher PHs and lower fixed acidity as higher PH values are less acidic

1c.

Chart, scatter chart

Description automatically generated

1d.

Chart, scatter chart

Description automatically generated

1e.

Chart, scatter chart

Description automatically generated

2a.

Chart, bar chart, histogram

Description automatically generated

2b.

Chart, bar chart

Description automatically generated

2c.

Chart, bar chart

Description automatically generated

Q5 has the highest proportion of wines with low alcohol content and Q8 has the highest proportion of wines with high alcohol content

2d.

Chart, bar chart

Description automatically generated

3a.

Chart, histogram

Description automatically generated

3b.

Chart, histogram

Description automatically generated

3c.

Chart, histogram

Description automatically generated

Yes, all of the distributions across the levels of alcohol.cat are roughly the same as the shape of the original distribution

4a.

Chart, box and whisker chart

Description automatically generated

4b.

A screenshot of a computer

Description automatically generated with low confidence

It shows all the levels of citric.acid.cat for all qualities if the fill aesthetic is mapped to citric.acid.cat

5. A highly rated TV news network invited viewers to participate in a poll to determine the percent of Americans who believe in the existence of extraterrestrial intelligence. Viewers participate by texting one of two numbers depending on their response. About 2 million viewers participated in the survey. Do you expect that this sampling method will provide you with a good estimate of the percent of Americans who believe in extraterrestrial intelligence? Explain your answer in a few sentences.

**I expect this sampling method will not give a good estimate of the percent of Americans who believe in extraterrestrial intelligence as this method is bias towards people who invest and watch the TV network to begin with as the survey targeted viewers, even if the sample size is 2 million. It would give a good estimate of the percentage of Americans who watch the TV network, but not the estimate of the population since only the viewers were ever considered with the sampling method used.**

6. At a hospital, a researcher selects a Simple Random Sample (SRS) of 50 patients from among those who underwent surgery for intestinal cancer and another SRS of 50 patients who were treated for intestinal cancer without surgery. The researcher found that patients who received surgery live much longer after treatment than patients who are treated without surgery. Moreover, suppose similar studies from dozens of hospitals nationwide yielded the same result that those who received surgery for intestinal cancer live much longer after treatment than patients who are treated without surgery. On the basis of these studies, can you conclude that surgery lengthens intestinal cancer patients’ lives? Explain your answer in a few sentences.

**From the data given, I would say that no reasonable conclusion can be drawn as while all the patients in the sample did have at one point have intestinal cancer, there are a lot of other factors that are aren’t accounted for, or lurking variables like age, sex, and other health conditions a patient might have. Whether people live or not after the intestinal surgery depends on those things as well as the surgery which are factors to consider with the sample of people that did not receive the surgery as well.**