

Representing Data with Figures

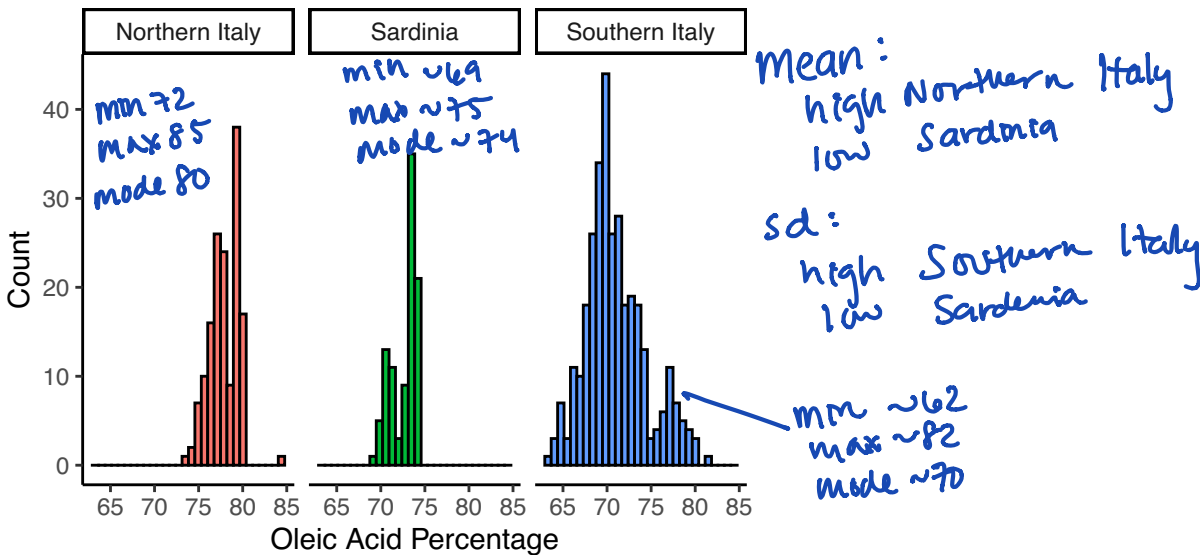
BIOL01104 Fall 2019, Dr. Spielman

This worksheet uses a dataset of information about olives collected from Italy. The data contains information about the fatty acid content for 572 olives from Northern Italy, Southern Italy, and Sardinia (a Mediterranean island off the coast of Italy). Answer questions about each plot with your table. **For all plots**, identify the type of data (quantitative or categorical) on each axis.

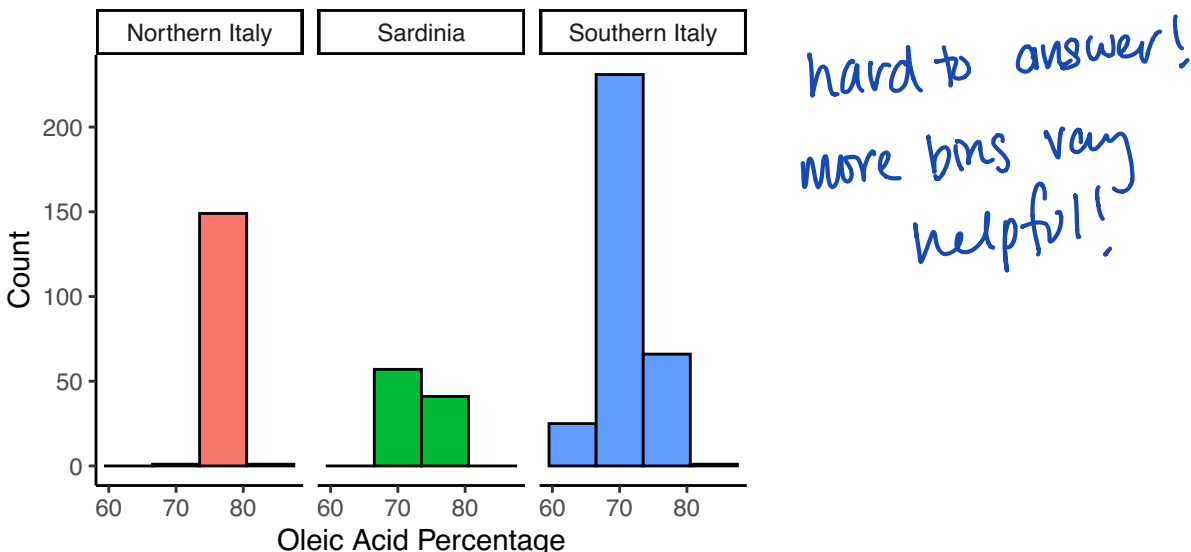
Histograms

The histograms below show the distribution of oleic acid percentages for olives collected from the three regions.

- Which region has the highest mean oleic acid percentage? The lowest?
- Which region has the highest standard deviation oleic acid percentage? The lowest?
- What is the minimum, maximum, and mode for each distribution? ← approximate ok!

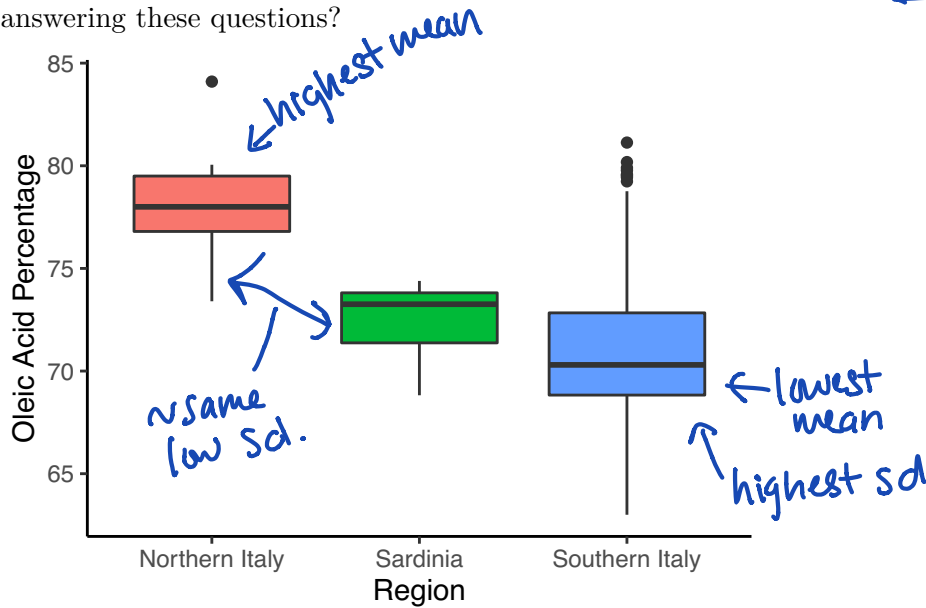


The histograms below show the *same* distributions as above, but with different *bin widths*. How do your answers to the questions change? What does this tell you about the importance of choosing good histogram bin widths?



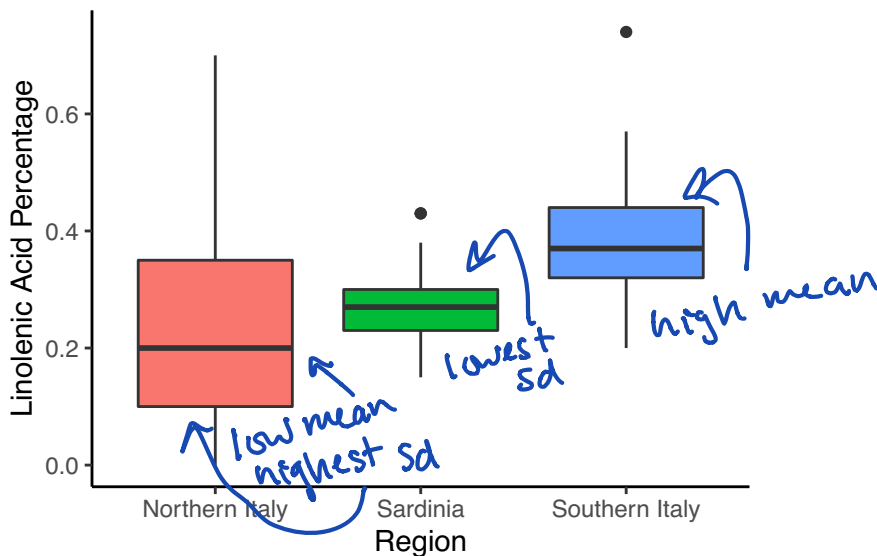
Boxplots

The boxplots below again shown the distribution of oleic acid percentages for olives collected from the three regions. Answer the same questions (mean, standard deviation). Were boxplots or histograms more informative for answering these questions?



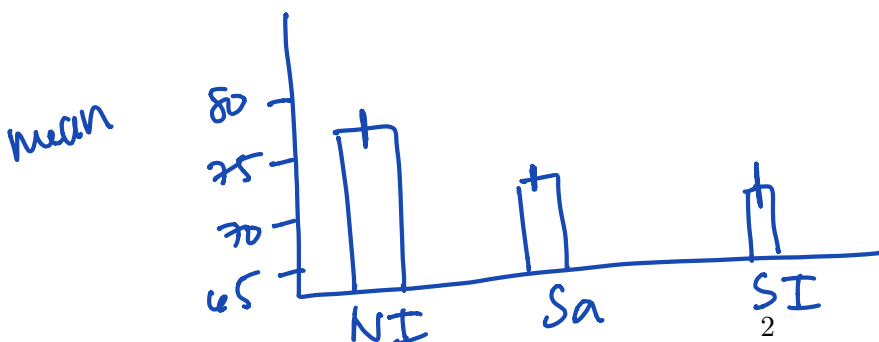
your personal discretion!

The boxplots below shown the distribution of linolenic acid percentages for olives collected from the three regions. Answer the same questions (mean, standard deviation). Then, compare the trends you observed for the previous boxplots of oleic acid to these linolenic acid boxplots. What are the similarities and differences between these two types of acids in olives?



reversed.

Finally, draw by hand (with your table!) a *bar plot* showing the mean oleic acid percentage for each region. Include rough estimates of standard deviation error bars.

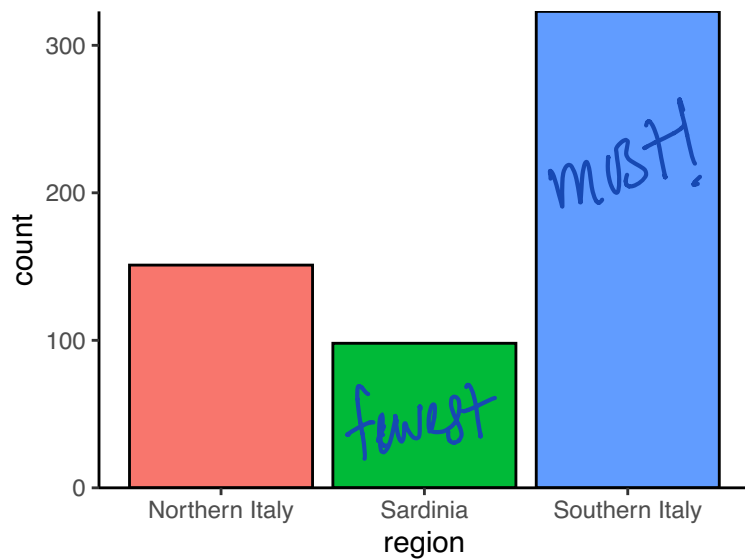


Barplots

The barplot below shows the number of olives collected from each region.

From which region were the most olives collected? The fewest?

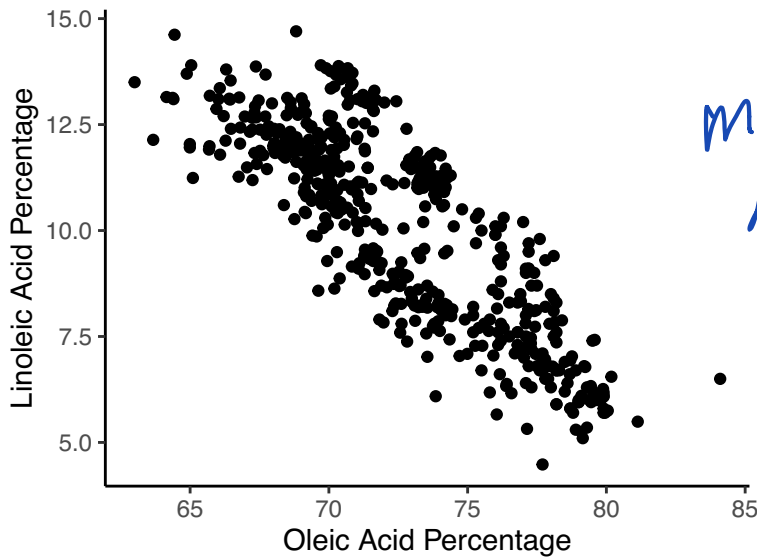
In addition, compare the general look of this barplot to the first histogram in this document. Although barplots and histograms look similar, they have some key differences. Take this opportunity to determine what those differences are.



Scatterplots, Part 1

The scatterplot below shows the *relationship* between percentages of oleic and linoleic acid in each olive. Thus, there are 572 points in this figure (1 point per olive).

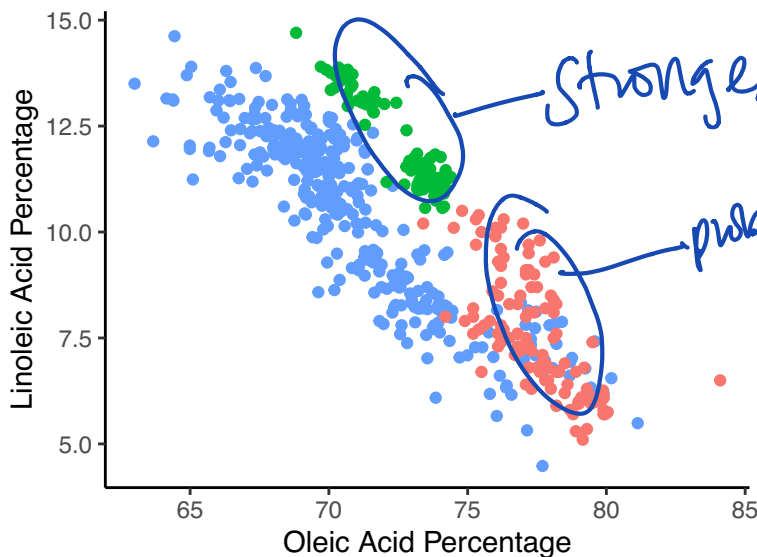
Is there a strong, moderate or weak correlation? Is the correlation positive or negative?



moderate to - strong
negative

The scatterplot below is a modified version of the above scatterplot, where this time points are colored according to which region the olive comes from.

What differences and similarities do you observe among regions? Which region appears to have the highest correlation? The lowest? Are all correlations the same direction (positive or negative)?



strongest

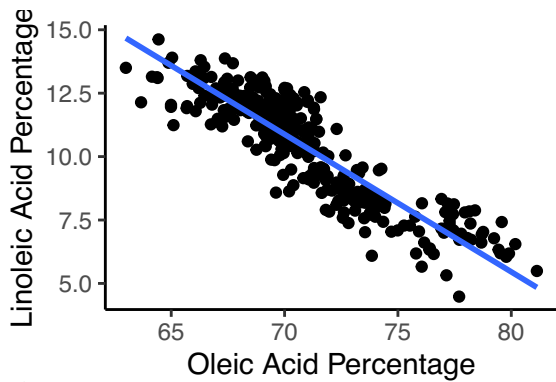
probably weakest.

all negative

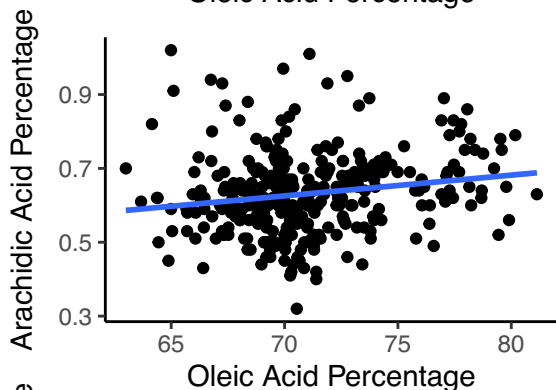
Scatterplots, part 2

Below are four different scatterplots each with a *line of best fit* showing the overall relationship between X and Y variables. Each plot shows the relationship between two different types of fatty acids, for Southern Italy olives only.

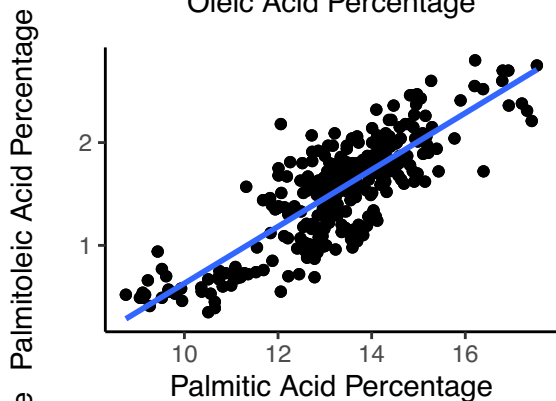
Characterize each plot as having a strong, moderate or weak correlation, as well as whether the correlation is positive or negative.



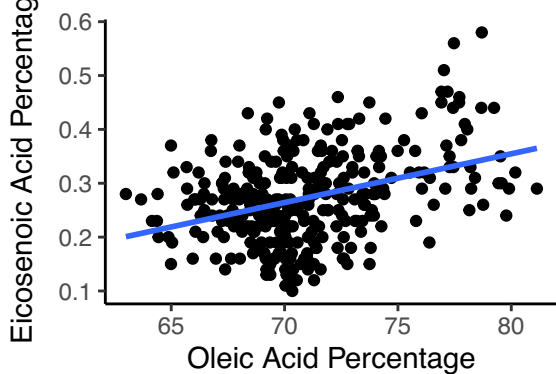
Strong-moderate
negative



Weak positive OR NONE



Strong-moderate
positive



Weak positive