You will run the hierarchical clustering algorithm and the K-means clustering algorithm implemented on JMP Pro on the iris dataset.

Hierarchical Clustering

1. Start JMP Pro
2. Open iris.jmp
3. Analyze > Clustering > Hierarchical Clustering
4. Select sepallength, sepalwidth, petallength, petalwidth, and click *Y, Columns*
5. Click OK.
6. Click the red triangle next to *Hierarchical Clustering*.
7. Select *Color Clusters*. Capture a top part (or as much as you can) of the *Dendrogram* and include it in your submission.

8. Click the triangle next to *Clustering History* (scroll down if you don't see it).

It will show you a table that shows all merge steps performed by the hierarchical clustering, including the number of clusters and distance. The distance increases as the number of clusters decreases. We can determine an optimal number of clusters by finding an inflection point in the distance (where the increase of distance changes significantly). This method is similar to the elbow method we discussed in the class.

9. Scroll down and observe the last 10 *Distance* values (corresponding to *Number of Clusters* 10 to 1). Capture a screenshot that shows the last 10 distance values and number of clusters, and include it in your submission.

10. Determine an optimal number of clusters from these 10 distance values and show it in your report. You have to justify why you chose that number.

K Means

1. Start JMP Pro
2. Open iris.jmp
3. Analyze > Clustering > K Means Cluster
4. Select sepallength, sepalwidth, petallength, petalwidth, and click *Y, Columns*
5. Click OK.
6. In the *Control Panel*, enter 1 for *Number of Clusters* and 10 for *Range of Clusters*.
7. Click Go.
8. Capture the screenshot of *Cluster Comparison* (which includes *CCC* values of 10 clusters) and include it in your submission.

Under the *Best* column, it shows 3 is an *Optimal CCC*. *CCC* (cubic clustering criterion) is a statistical measure that is used to find an optimal number of clusters.

9. Capture the screenshot of *K Means NCluster*=3 (scroll down if you don't see it), which shows *Cluster Summary*, *Cluster Means*, and *Cluster Standard Deviations*, and include it in your submission.

10. Click the red triangle next to *K Means NCluster*=3, and select *Scatterplot Matrix.*
11. Capture the screenshot of *Scatterplot Matrix* that shows all six scatterplots and include it in your submission.