Answers to Part 7 (Try It yourself) Exercise

1. What is the size of the largest cluster?

The largest cluster (#2) has 6200 observations.

1. Compare the characteristics (RegionDensityPercentile, MedianHouseholdIncome, and AverageHouseholdSize) of cluster 3 to the population as a whole?

Cluster 3 is less densely populated than the overall population, has a lower median household income, but a larger average household size than the overall population.

1. What is the range of withinss error for those 7 clusters (lowest to highest)?

The withinss ranges from 2286.613 (cluster 6) to 4367.405 (cluster 7).

1. Is the cohesion generally higher or lower than the 5 cluster solution?

Cohesion is generally higher (lower withinss) than the 5 cluster solution.

1. What is the average betweenss error for those 7 clusters?

The average betweenss is 7182.23.

1. Is the separation higher or lower than the 10 cluster solution?

Separation is generally higher (higher betweenss) than the 10 cluster solution.

1. Based on the analyses we have done so far, use your own words to summarize how the number of clusters can affect withinss error (cohesion) and betweenss error (separation).

Increasing the number of clusters often results in (1) lower withinss error = higher cohesion (good) but (2) lower betweenss error = lower separation (bad).