5.
  a. Given Data Sample Size: 119,968 patients
     Data Classification(Sizes):
     - RIght Handed: 85%
     - Left Handed: 13%
     - Ambidextrous/Mixed-Handed: 2%

     Observation:
     The data provides 3 classes. The given 3 classes do not have equal constitution in the entire dataset. It is also noted that the 3 data classes mutually exclusive to each other (i.e. Right handed(85%) + Left handed(13%) + Ambidextrous/Mixed-Handed(2%) = 100%)

     Sampling Method used: Stratified Sampling

     Justification:
     In this situation, we will utilize stratified sampling methods for preliminary analysis. We will use the following method as in this situation the data classes don't have an equal contribution in the whole dataset. Stratified sampling will help in selecting an equal amount of data from each data class and thus help in better preliminary analysis.
     In case of random sampling, the situation will end up picking up most data points from the largest class, because of its highest probability of being selected. This will end up in the other 2 data classes (left handed and ambidextrous) being neglected in the preliminary analysis.

  b. Given Data Sample Size: 1 million+
     Data Classification: 15 classes

     Observation:
     The data sample is larger than the available space to perform the analysis. Hence there is a need of selecting 100,000 words for analysis of the data to draw a conclusion.

     Sampling Method used: Stratified Sampling

     Justification:
     As there is a need to have efficient use of space, it is necessary to have an equal amount of data from all the 15 different classes. Stratified sampling will help in taking in equal amount of data points from all the different classes and thus help in better analysis of the data

  c. Sampling Method used: Progessive Sampling

Justification:
Here we would use progressive sampling technique because it would help to determine the optimum size of the sample data required. The testing of it, starts with a small sample size and gradually increases till it reaches the true frequency. At a point if the increasing sample size starts decreasing the accuracy, we stop and revert to the size providing the best result, and helping determine the most optimal sample size.