Q1.

Two variable types in the datasets include Categorical and Numerical Data.

Unique example of the categorical data is the Gender column under Section 1: Household Roster of the Dataset

For numerical data, an example in the dataset is the discrete numeric data of Quantity of Crops harvested in the 2004/2005 season under Section 4: Details of Farming Activities.

Q2.

The population was sampled using a Multi-stage stratified sampling method.

Given the purpose of the survey, the sampling method seems quite appropriate as it satisfactorily covers the areas of the investigation linked with the purpose of data gathering.

However, certain elements associated with the method of data collection could cause conceivable bias between the actual intended population size and sample size. A possible source of bias could include:

Self-selection Bias: In the summary of the survey background information, it was stated that countries in the various geographic zones were added to the survey pool by country climate change affiliated institutions opting to participate in the survey. It is possible that this selection process may misrepresent the actual population of agricultural household as not all countries were given an equal chance of being selected.

Q3.

Two research questions that can be obtained from the dataset includes:

1. Is there a relationship between changes in climate conditions and output that causes an adaptation farming method?
2. Is the average age of farmers in East Africa higher than that in West Africa?

Q4.

A statistical analysis to evaluate research question **2** could go like this:

The Null Hypothesis: The average age of farmers in East Africa is greater than the average age of farmers in Western Africa.

The Alternate Hypothesis: The average age of farmers in East Africa is less than the average age of farmers in Western Africa

Selected Hypothesis Test: We will use the two sample T-test to test the hypothesis

Reason: This is used for independent samples and used to compare the means of two populations

How to determine significance: Significance would be determined by the p-value. With a p-value greater than the significance level of 0.05 we would accept the null hypothesis. If the p-value is lower the significance level of 0.05 we reject the null hypothesis.

Q5.

If the results from the tests indicates that the hypothesis was significant we conclude that the average age of farmers in East Africa is higher than that of West Africa. If the results were not significant we would conclude that the average age of farmers in East Africa is less than that in West Africa.

Q6.

1. Bar Charts – used to show the comparison of dataset of quantity a certain crop (rice or beans or maize…etc) produced between the 4 geographical zones of North, South, East and West Africa.
2. Contingency Tables – Used to show the distribution of gender of farmers across the 4 African geographical zones
3. Pie Charts – used to show the percentage quantity of production of a certain crop between the 4 Africa geographical zones

Q7.

The data provided on fertilizer use, pesticide use, and irrigated data is important to have been compared to a reference point to investigate reasons as to why there may or may not be biases in a dataset so that further research can be carried out to improve accuracy of research outcomes.