**Q1: Tanzania Student Data**

This builds on the bonus from the previous Stata assignment. We downloaded the PSLE data of students from 138 schools in Arusha District in Tanzania (previously we only had data of just 1 school) You can build on your code from the previous assignment to create a student level dataset for these 138 schools.

**Q2: Côte d'Ivoire Population Density**

We have household survey data and population density data of Côte d'Ivoire. Merge departmente-level density data from the excel sheet (CIV\_populationdensity.xlsx) into the household data (CIV\_Section\_O.dta) i.e. add population density column to the CIV\_Section\_0 dataset.

**Q3: Enumerator Assignment based on GPS**

We have the GPS coordinates for 111 households from a particular village. You are a field manager and your job is to assign these households to 19 enumerators (~6 surveys per enumerator per day) in such a way that each enumerator is assigned 6 households that are close to each other (this would reduce the amount of time they spend walking from one house to another.) Manually assigning them for each village will take you a lot of time. Your job is to write an algorithm that would auto assign each household (i.e. add a column and assign it a value 1-19 which can be used as enumerator ID). Note: Your code should still work if I run it on data from another village.

**Q4: 2010 Tanzania Election Data cleaning**

2010 election data (Tz\_election\_2010\_raw.xlsx) from Tanzania is not usable in its current format. You have to create a dataset in the wide form, where each row is a unique ward, and votes received by each party are given in separate columns. You can check the following dta file as a template for your output: Tz\_elec\_template. Your objective is to clean the dataset in such a way that it resembles the format of the template dataset.

**Q5: Tanzania PSLE data**

PSLE dataset contains data of 17,329 schools. We have the region and district of each school but for our analysis we need the ward information. There is another dataset (q5\_school\_location) that has the ward information of 19,733 schools. Your job is to identify ward information for 17,329 schools on the PSLE dataset using the q5\_school\_location.dta. Note: Final dataset should be the PSLE dataset + ward column (i.e. N = 17,329). Hint: You might have to try different methods to get the best results, even then you might have some schools where we can’t find ward information.

**Q6: Tanzania Election data Merging (Bonus Question)**

Between 2010 and 2015, the number of wards in Tanzania went from 3,333 to 3,944. This happened by

dividing existing ward into 2 (or in some cases three or more) new wards. You have to create a dataset where each row is a 2015 ward matched with the corresponding parent ward from 2010. It’s a trivial task to match wards that weren’t divided, but it’s impossible to match wards that were divided without additional information. Thankfully, we had access to shapefiles from 2012 and 2017. We used ArcGIS to create a new dataset that tells us the percentage area of 2015 ward that overlaps a 2010 ward. You can use information from this dataset to match wards that were divided.