Analyze Data in Stata
To complete this exercise, you will create a customized data extract, request the data formatted for Stata (.dta), and download the file to analyze on your computer. The variables and samples you should include in your extract are listed below. Step-by-step instructions for creating an IPUMS-International data extract are available here: https://youtu.be/-0befCJAcjQ.

Samples:  Ghana 2010  
          South Africa 2007  
          Tanzania 2012  

Variables:  SAMPLE [preselected]  
            SERIAL [preselected]  
            URBAN  
            PERNUM [preselected]  
            PERWT [preselected]  
            AGE  
            SEX  
            EMPSTAT  
            OCCISCO  

Data Exercise 1: Know Your Sample Identifiers
Your analyses will be built around the sample identifier (SAMPLE), which is typically included in your extract by default. To check the sample codes and total number of cases, run a frequency on the sample identifier.

How many individuals are in the Ghana 2010 sample extract? _________  
How many individuals are in the South Africa 2007 sample extract? _________  
How many individuals are in the Tanzania 2012 sample extract? _________  

Stata command:  
tab sample  

Note: In Stata, the default setup shows value labels, but not the values themselves. The values for these samples are 288201001=Ghana, 710200701=South Africa, and 834201201=Tanzania.

Data Exercise 2: Urban/Rural Population and the Importance of Weights  
2a. Get basic frequencies of the URBAN variable by sample.  
How many individuals from the sample lived in urban areas?
What proportion of individuals from the sample lived in urban areas?

Stata command:
\texttt{tab urban sample, column}

2b. **Redo the analysis with the person weight (PERWT) to get accurate estimates.**
You will redo the frequency table you just made, but this time you'll apply the *person weight*.

Using weights, what is the total population of each country? (Compare to Part 1.)

Stata command:
\texttt{tab urban sample [iweight=perwt], column}

Using weights, how many individuals lived in urban areas?

Using weights, what proportion of individuals lived in urban areas? (Compare to Part 2a.)

Stata command:
\texttt{tab urban sample [iweight=perwt], column}

**Data Exercise 3: Labor Force Participation**

3a. **Compare the distribution of occupational activity in the two countries.**

Using weights, which occupational category (from OCCISCO) has the highest percentage of workers in each country?

Stata command:
\texttt{tab occisco sample [iweight=perwt], column}

(Notice the high percentages in NIU. These people were not asked occupational questions.)
3b. Compare the distribution of occupational activity for females in the two countries.

Which occupational group has the highest percentage of female workers in each country?
- Ghana 2010 _____________________ South Africa 2007 _____________________
- Tanzania 2012 _____________________

Stata command:
```
tab occisco sex if sample==288201001 [iweight=perwt], column
```
```
tab occisco sex if sample==710200701 [iweight=perwt], column
```
```
tab occisco sex if sample==834201201 [iweight=perwt], column
```

3c. Compare the distribution of occupational activity among people employed.

Note that in order to do your analysis, you must decide whether you are analyzing the total population or the people participating in the labor force. The previous commands yielded totals and percentages among all people in the population. If you want to know how women's work is distributed among women employed, you have to limit your analysis to people who declare being employed. That is, look at employment status category 1, "employed."

What is the employment rate by gender in each country?

- Female: _____%  Female: _____%  Female: _____%

Stata command:
```
tab empstat sex if sample==288201001 [iweight=perwt], column
```
```
tab empstat sex if sample==710200701 [iweight=perwt], column
```
```
tab empstat sex if sample==834201201 [iweight=perwt], column
```

From your metadata exercise, you may also have noticed that employment questions were only asked of persons above a certain age. Review the universe for EMPSTAT and OCCISCO online. Filter data to include only employed persons who are 15 or older (EMPSTAT = 1 and AGE ≥ 15).

What percentage of employed women is working ... (Compare results to Part 3b.)

...as Skilled agriculture and fisheries workers
```
In Ghana 2010
```
```
In South Africa 2007
```
```
In Tanzania 2012
```

...as Service workers and shop and market sales
In Ghana 2010
In South Africa 2007
In Tanzania 2012

Stata command:

```
tab occisco sex if sample==288201001 & empstat==1 & age>14 [iweight=perwt], column
tab occisco sex if sample==710200701 & empstat==1 & age>14 [iweight=perwt], column
tab occisco sex if sample==834201201 & empstat==1 & age>14 [iweight=perwt], column
```
Answer Key

1. How many individuals are in the sample extract?
   - Ghana 2010: 99,699 persons
   - South Africa 2007: 209,176 persons
   - Tanzania 2012: 179,618 persons

2a. How many individuals from the sample live in urban areas?
   - In Ghana 2010:
     Number: 50,908
     Percent: 51.06%
   - In South Africa 2007:
     Number: 120,802
     Percent: 57.75%
   - In Tanzania 2012:
     Number: 57,185
     Percent: 31.84%

2b. What is the total population according to the weighted data?
   - Ghana 2010: 24,705,412
   - South Africa 2007: 47,142,493
   - Tanzania 2012: 43,960,848

   How many individuals live in urban areas (with properly weighted data)?
   - In Ghana 2010:
     Number: 12,615,002
     Percent: 51.06%
   - In South Africa 2007:
     Number: 28,340,280
     Percent: 60.12%
   - In Tanzania 2012:
     Number: 12,831,777
     Percent: 29.19%

Comparing frequencies and proportions, you can see that unweighted sample data from South Africa and Tanzania does not represent the population properly. The weights in both samples differ across urban and rural areas. Instead, the Ghana dataset has a flat expansion factor.

3a. Labor Force Composition
   - Ghana 2010: 19.75% in Skilled agriculture and fisheries workers
   - South Africa 2007: 4.73% in Elementary occupations
   - Tanzania 2012: 26.93% in Skilled agriculture and fisheries workers

3b. Which occupational category employs the greatest percentage of female workers in each country?
   - Ghana 2010: 17.99% in Skilled agriculture and fisheries workers
   - South Africa 2007: 5.70% in Elementary occupations
   - Tanzania 2012: 26.60% in Skilled agriculture and fisheries workers

3c. What is the employment rate for each gender in each country?
   - Ghana 2010:
     Males: 44.40% employed
     Females: 44.00% employed
   - South Africa 2007:
     Males: 30.30% employed
     Females: 21.40% employed
   - Tanzania 2012:
     Males: 43.90% employed
     Females: 39.10% employed

3d. Among employed women age 15 or older, what percentage is working…
   - In Agriculture
     - Ghana 2010: 38.4%
     - South Africa 2007: 2.9%
     - Tanzania 2012: 68.2%
   - In Services
     - Ghana 2010: 31.3%
     - South Africa 2007: 8.3%
     - Tanzania 2012: 6.6%