EXERCISE I: EXPLORING DATA

TOPICS COVERED:
- ❏ Frequency tables
- ❏ Visual Summaries (pie chart and bar graph)
- ❏ Numerical Summaries (mean, median, IQR, standard deviation, boxplot)

- ❏ **REQUIRED DATASET: IPUMS-International**
- ❏ **REQUIRED VARIABLES:**
   1. COUNTRY
   2. YEAR
   3. AGE
   4. SEX
   5. CHBORN (children ever born)

[The only preselected variables that are needed in this exercise are COUNTRY and YEAR. Make sure to remove all of the other preselected variables by unchecking the blue boxes next to them. This will reduce the size of your data file and also make it easier to view the data in R.]

- ❏ **RECOMMENDED SAMPLES:**
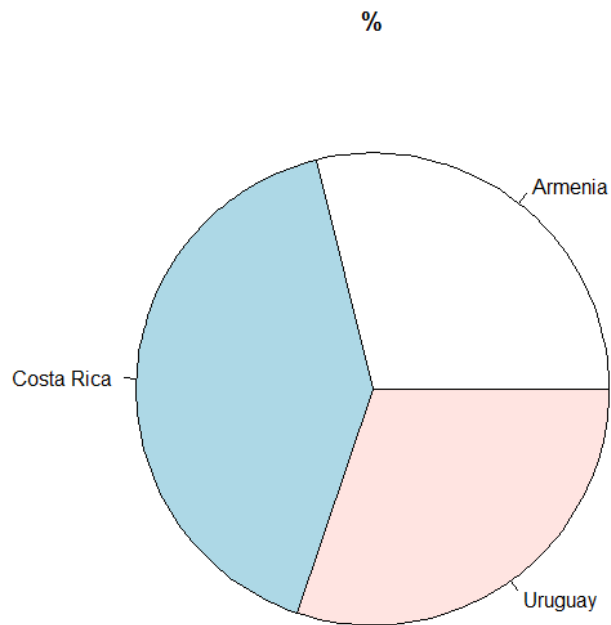   1. Armenia 2011
   2. Costa Rica 2011
   3. Uruguay 2011

❖ Question 1

Complete the following frequency table based on the three countries you chose to work with:

Country	Freq.	Percent	Cum.
Armenia			
Costa Rica			
Uruguay			
Total			

Country	Freq.	Percent	Cum.
Armenia	301,831	28.47	28.47
Costa Rica	430,082	40.56	69.03
Uruguay	328,425	30.97	100
Total	1,060,338	100.00	100.00

❖ **Question 2**

Create a pie chart to represent the data shown above:

%

❖ Question 3

For each of the countries in your dataset, find the (A) Mean and (B) Median of the total number of children ever born [CHBORN]

a) NB! There are missing values in CHBORN. Before calculating the mean and median, you need to address the missing values. What is the difference between the two missing-value categories 98: Unknown and 99: NIU (not in universe)?

Persons who did not provide a response to this census question (but were asked the question) receive a code of 98. Persons who were not asked this census question because they are outside the population for which this question was intended (in this case all males and female children) receive a code of 99.

b) Should you include or exclude cases with the missing values from your calculations?

Persons who were not asked the census question (99: NIU) should always be excluded from analyses. They are not part of the population for which we are calculating summary statistics. For summary statistics, we also want to exclude persons with unknown responses (98: Unknown) and make note of this exclusion when reporting or interpreting our results. For certain analyses and statistical tests, we may choose to include persons with unknown responses or impute these values for these individuals.

c) What is the population universe for this variable in each country? That is, in each census, who was asked this question?

Armenia 2011: Females age 15+
Costa Rica 2011: Females age 12+
Uruguay 2011: Females age 12+

d) To make an accurate comparison of children ever born across these three countries, which cases should you drop from your dataset?

Females age 12-14 in Costa Rica and Uruguay. This provides a comparable population across the three countries: females age 15+.

e) Complete the table for females age 15+

Country	Mean CHBORN	Median CHBORN	N
Armenia			
Costa Rica			
Uruguay			

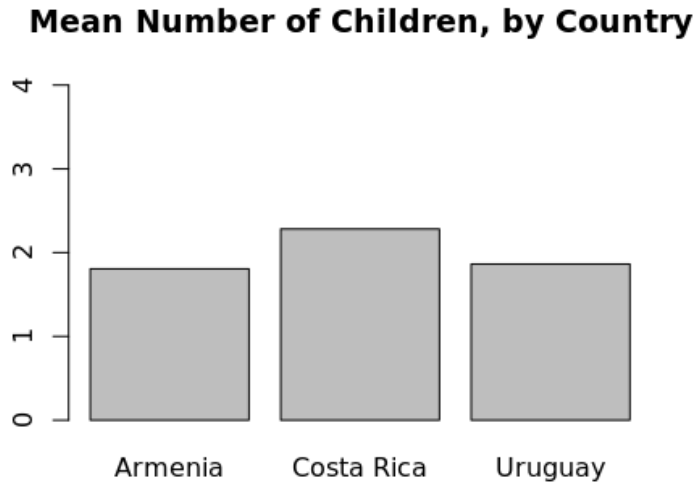| Country	Mean CHBORN	Median CHBORN	N
Armenia	1.806892	2	130,580
Costa Rica	2.282232	2	158,887
Uruguay	1.864225	2	133,279

❖ **Question 4**

For each of the countries in your dataset, find the (A) Q1, (B) Q3, (C) interquartile range [IQR] and (D) standard deviation of the total number of children ever born [CHBORN].

Country	Q1	Q3	IQR	SD
Armenia				
Costa Rica				
Uruguay				

Country	Q1	Q3	IQR	SD
Armenia	0	3	3	1.51
Costa Rica	0	3	3	2.39
Uruguay	0	3	3	1.90

❖ Question 5

Create a bar chart that displays the mean number of children born to women age 15 and older by country.

Mean Number of Children, by Country

❖ Question 6

The mean for CHBORN is calculated for women 15 & above. Recalculate the mean limiting your data to: (1) Women age 21 and older (2) Women age 30 and older and (3) Women above 45 and older:
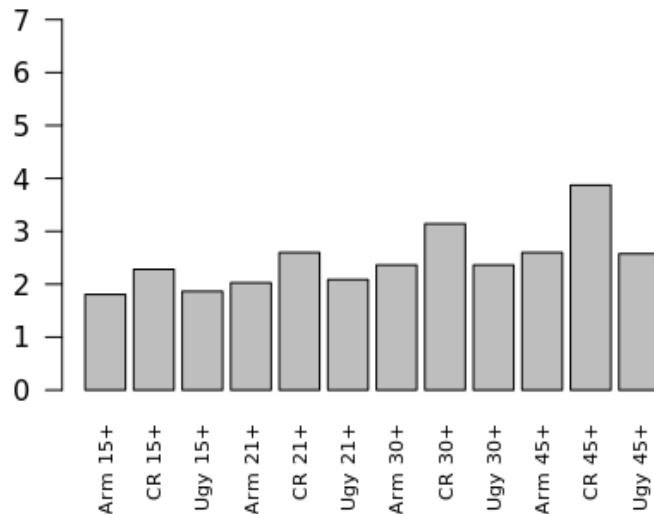
Country	15+	21+	30+	45+

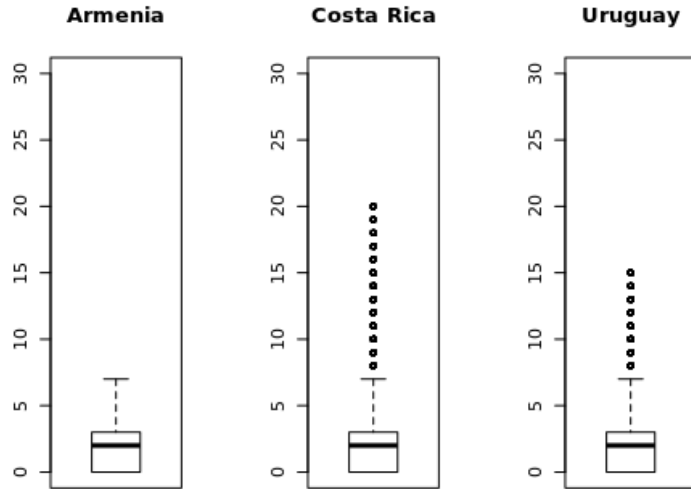| Country	15+	21+	30+	45+
Armenia	1.80	2.03	2.36	2.60
Costa Rica	2.28	2.60	3.14	3.87
Uruguay	1.86	2.09	2.36	2.58

❖ **Question 7**

Create a bar chart of mean number of children born by country and age group.
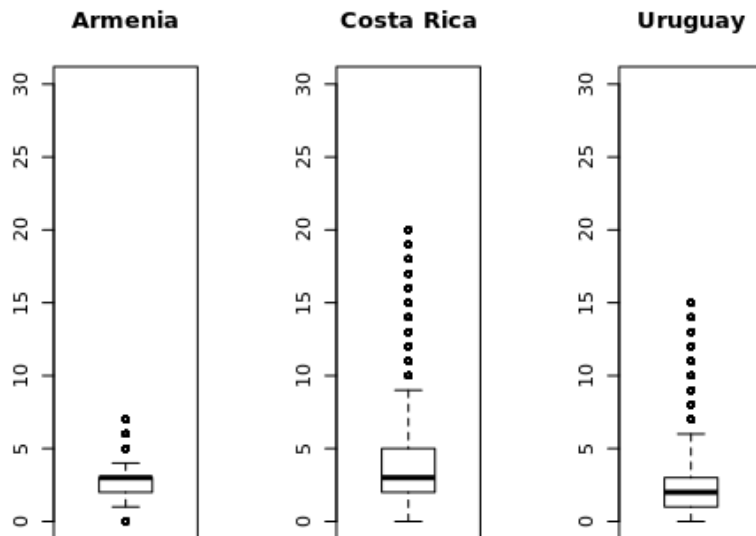

Mean Number of Children, by Country & Age

❖ Question 8

Create a boxplot (A) representing the number of children born (CHBORN) to women age 15 and older by country.

❖ Question 9

Create a boxplot (B) representing the number of children born (CHBORN) by country. Limit your sample to women age 45 and older.

❖ Question 10

Based on both box plots above, which of the following is true?

A) In box plot B, there is no median for Armenia because there not enough outliers.

B) In box plot B, the median in Armenia is the smallest since it has the least dots on top of its upper whisker.

C) In box plot A, the range is equal in all three countries

D) In box plot A, the IQR in Uruguay is the smallest, since its median is the smallest