

Semester Project: 2012 United Nations Data from a Sample of 36 Countries

David Gurney, Math 241-61, Apr. 29, 2021

The data for this project came from the U.N. website, “Statistics and indicators on women and men”, located at

<https://unstats.un.org/unsd/demographic/products/indwm/default.htm>

and from the U.N. “Demographic Yearbook” website located at

https://unstats.un.org/unsd/demographic/products/dyb/2010_round.htm

All of the data was accessed on July 5, 2017. The variables being considered are the percent of population under 15 years old, the percent of population that is 60 years old or more, the percent of women in the at least 60 age group, and the continent or sea where the country is located. A complete listing of the data can be found on the last page of this project.

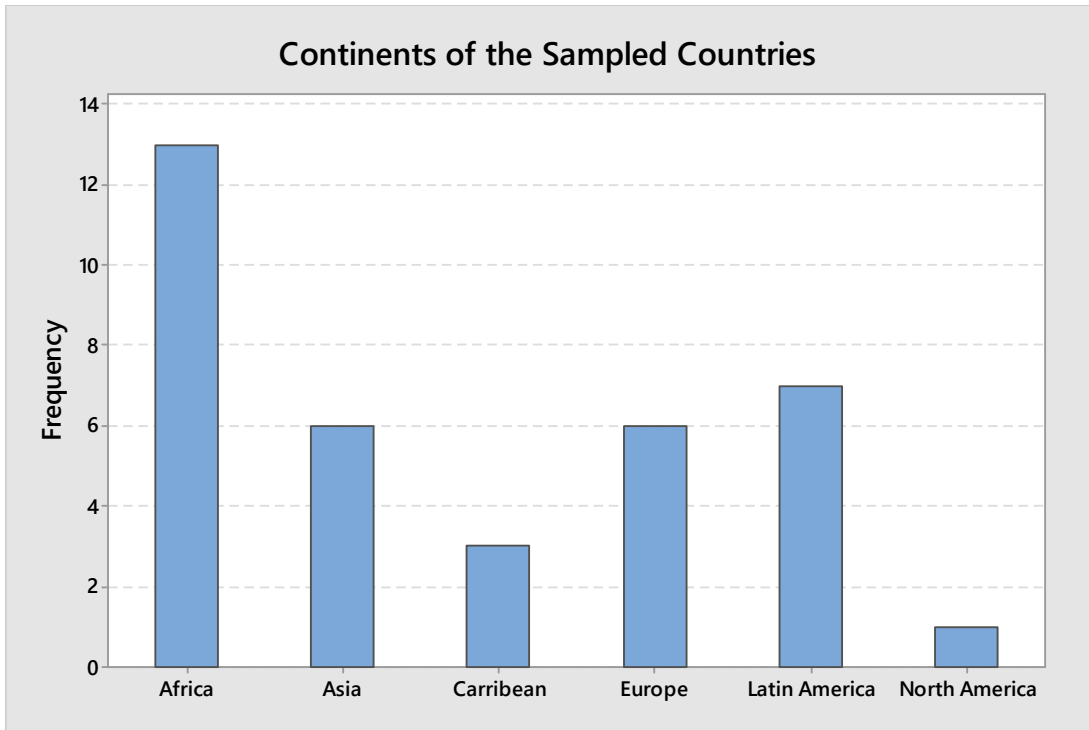
A frequency distribution of the country locations follows.

Location	Frequency	Relative Frequency
Africa	13	36.1%
Asia	6	16.7%
Caribbean	3	8.3%
Europe	6	16.7%
Latin America	7	19.4%
North America	1	2.8%

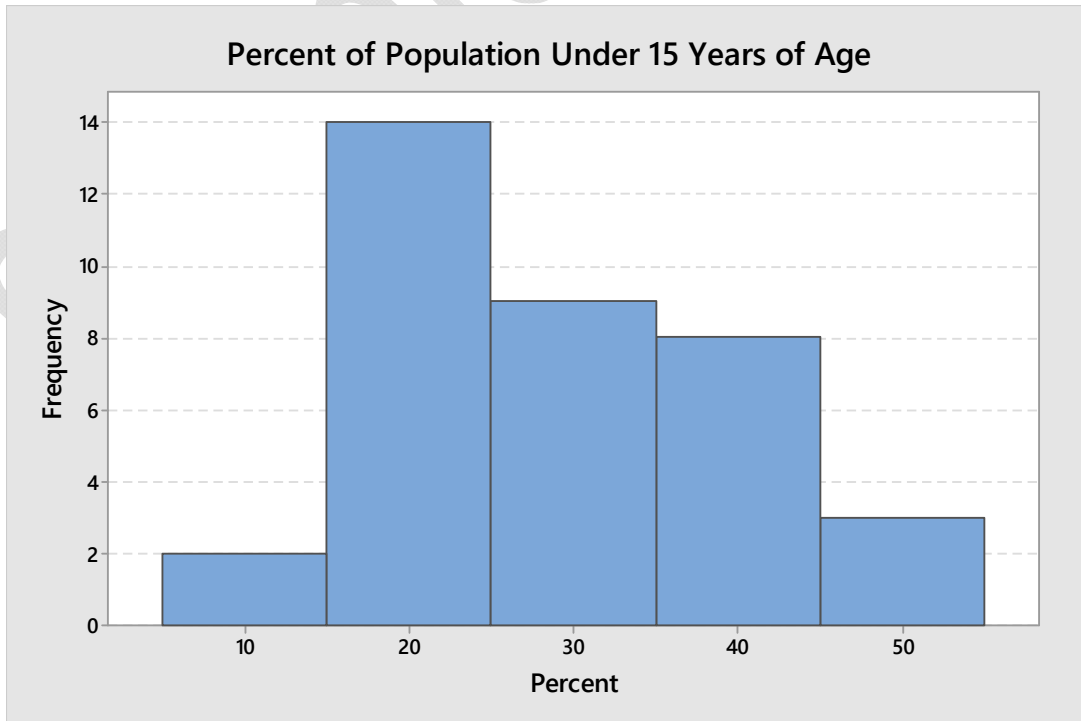
A bar chart of this data appears on the next page. Africa has the highest frequency of 13.

North America has the lowest frequency of 1. Note that while most of the countries are located on a continent, 3 of the countries are located in the Caribbean Sea.

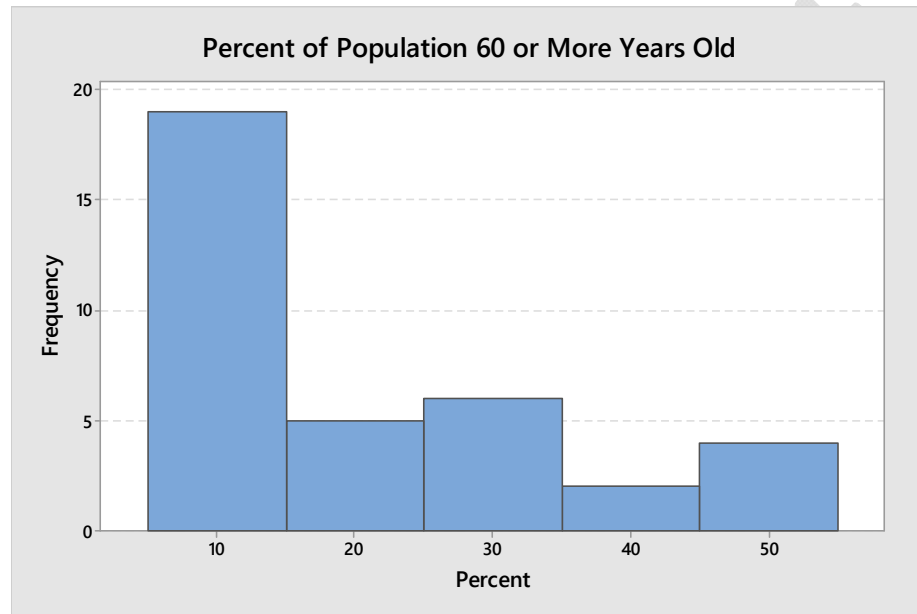
A chi-square test was run to see if all the locations are equally likely. The resulting test statistic was 14.00, and the corresponding P-value was 0.016. Thus, at 95% confidence, the evidence is strong enough to say that not all locations are equally likely.



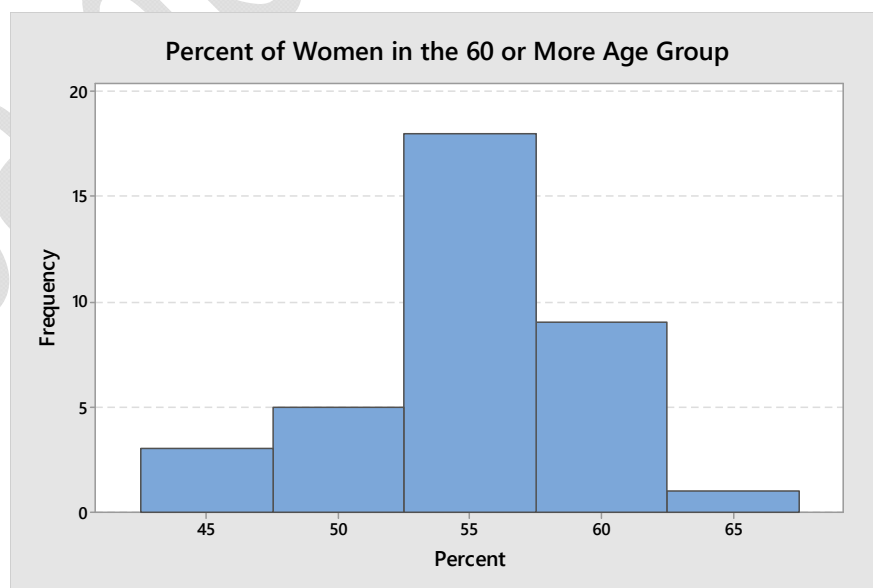
The following histogram is of the percent of the population under 15 years of age. The maximum data value is about 55% and the minimum data value is about 5%. The histogram has one peak at 20% and is skewed to the right. There are no gaps or extreme values.



The histogram below is of the percent of the population that is 60 or more years old. The minimum is about 5% and the maximum is about 55%. There is a major peak at 10% and a smaller peak at 30%. The histogram is skewed to the right. There are no gaps or extreme values.



The next histogram is of the percentage of women in the 60 or more age group.

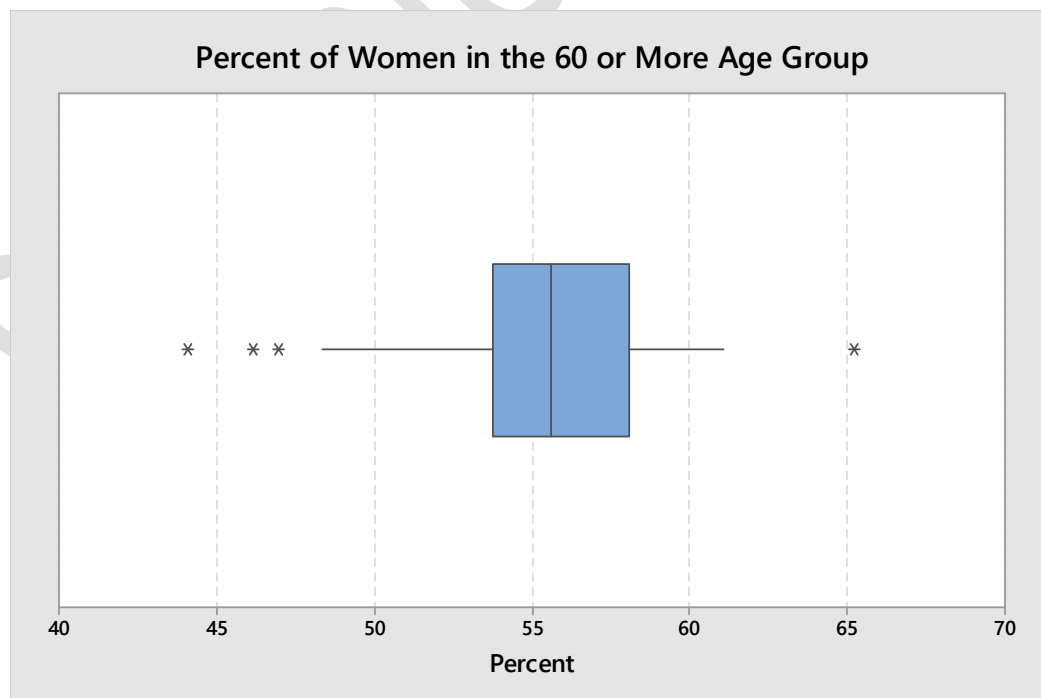


The “Percent of Women ...” histogram has a minimum of about 42% and a maximum of about 68%. There is a single peak at 55% and the histogram is skewed to the left slightly. The histogram has no gaps or extreme values.

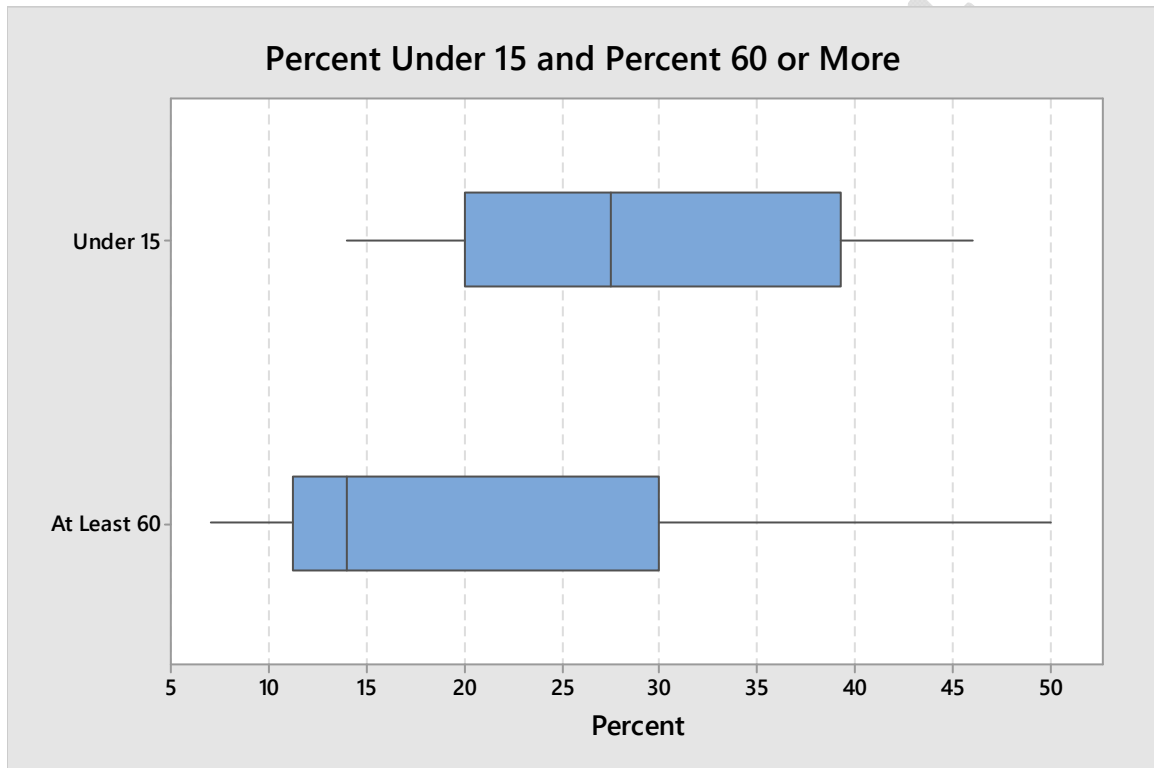
The table below shows some basic statistics for the three quantitative variables.

	% Under 15	% 60 or More	% Women of 60 or More
Minimum	14	7	44
Q1	20.0	11.2	54.0
Median	27.5	14.0	56.0
Q3	39.2	30.0	58.0
Maximum	46	50	65
Mean	28.4	21.1	55.2
Mode	24, 40 (3 times)	14 (4 times)	56 (7 times)
Std. Dev.	10.2	13.3	4.4
Std. Error	1.7	2.2	0.7
Range	32	43	21
IQR	19.2	18.8	4.3
Skewness	0.26	0.97	-0.55

The boxplot for the percent of women in the 60 or more age group is shown below. The boxplot is skewed to the left. There are three outliers on the left side and one on the right.



The boxplots for the percent under 15 years and the percent of 60 or more are shown below. “60 or More” has the smallest and the largest values, and also the largest range. “Under 15” has the highest median, and a slightly larger interquartile range. “60 or More” is much more skewed.



A dependent difference test was run to see if the means of “Under 15” and “60 or More” were the same. The resulting test statistic was 1.96 and the corresponding P-value was 0.057. Thus, at 95% confidence, the evidence is not strong enough to say the means are different.

The 95% confidence intervals for all three quantitative variables are shown below.

Percent Under 15 Years	(25.05, 31.97)
Percent 60 or More Years	(16.67, 25.62)
Percent of Women, 60 or More	(53.78, 56.71)

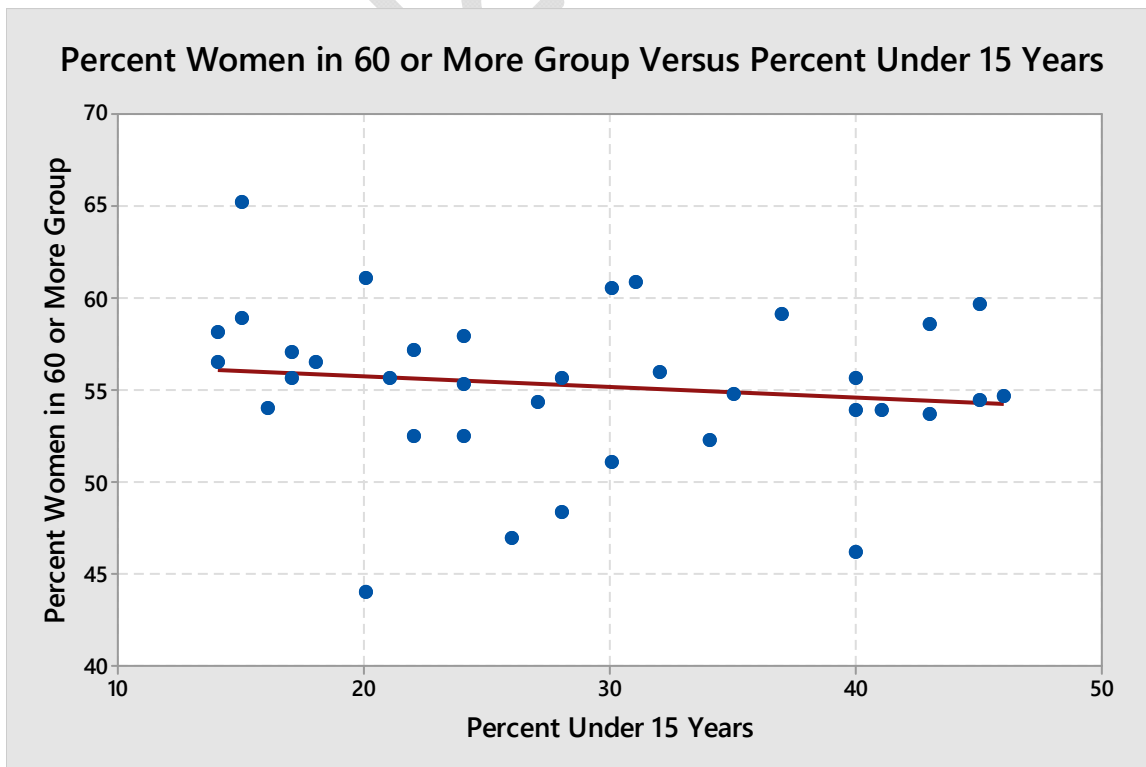
The Ryan-Joiner test was used to determine if each quantitative variable was normally distributed. The test statistics and corresponding P-values are shown below.

	Test Statistic	P-value
Percent Under 15 Years	0.974	0.091
Percent 60 or More Years	0.926	< 0.010
Percent of Women, 60 or More	0.977	> 0.100

Since the P-values for both the “Percent Under 15 ... ” and the “Percent Women ... ” are more than 0.05, we have 95 % confidence that these variables are normally distributed. Since the P-value for “Percent 60 or More ... ” is less than 0.05, we conclude that this variable is not normally distributed with 95% confidence.

A scatter diagram between “Percent Women ... ” and “Percent Under 15 ... ” is shown below. The equation of the regression line is

$$y = 56.92 - 0.05862x.$$



The outlier is at about 20% “Percent Under 15 ...” and 44% “Percent Women ...”. There are no influential observations. The coefficient of determination is about 0.019, which means about 1.9% of the variation is determined by the regression line.

A test for significant linear correlation between the “Percent Under 15 ...” and the “Percent Women ...” was run. The correlation coefficient was -0.139, and the corresponding P-value was 0.420. So at 95% confidence, the evidence is not strong enough to say there is significant linear correlation. This conclusion is supported by the scatter diagram above.

Table of Data

Country	Percent Population Under 15	Percent Population At Least 60	% Women In At Least 60 Portion	Continent
Albania	22	27	52	Europe
Algeria	27	14	54	Africa
Angola	46	8	55	Africa
Argentina	24	30	58	Latin America
Armenia	20	30	61	Asia
Aruba	18	31	56	Carribbean
Austria	14	47	56	Austria
Bahamas	22	22	57	Carribbean
Bahrain	20	8	44	Asia
Bangladesh	30	14	51	Asia
Barbados	17	34	57	Carribbean
Belarus	15	37	65	Europe
Belgium	17	48	56	Europe
Belize	34	12	52	Latin America
Benin	43	9	59	Africa
Bhutan	28	14	48	Asia
Bolivia (Plurinational State of)	35	15	55	Latin America
Botswana	32	13	56	Africa
Brazil	24	22	55	Latin America
Brunei Darussalam	26	13	47	Asia
Bulgaria	14	50	58	Europe
Burkina Faso	45	7	60	Africa
Burundi	37	9	59	Africa
Cambodia	31	13	61	Asia
Cameroon	40	11	54	Africa
Canada	16	42	54	North America
Cape Verde	30	14	60	Africa
Central African Republic	40	12	56	Africa
Chad	45	9	54	Africa
Chile	21	28	56	Latin America
Colombia	28	18	56	Latin America
Comoros	43	8	54	Africa
Congo	41	11	54	Africa
Costa Rica	24	20	52	Latin America
Côte d'Ivoire	40	12	46	Africa
Croatia	15	48	59	Europe