Analyzing Country Data

MATH NSPIRED

Math Objectives

- Students will determine and interpret percents to summarize data.
- Students will interpret medians and quartiles to summarize sample data. Use measures of center and measures of variability to draw informal comparative inferences about two populations.
- Students will compare and contrast different sample data summarized by box plots and histograms. They will assess the degree of visual overlap of two numerical data distributions.
- Students will construct viable arguments and critique the reasoning of others.
- Students will use appropriate tools strategically.

Vocabulary

- box plot
- dot plot
- population
- quartiles
- median
- sample
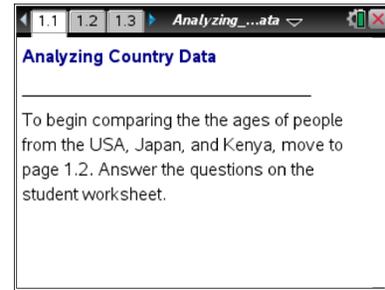- histogram

About the Lesson

This activity involves sample listings of the ages of 200 people from each of three countries: USA, Kenya, and Japan. Students will:

- Consider why a summary of ages is important and how a summary of the ages can be connected to a better understanding of the populations of each country.
- Examine sample ages from the three countries displayed in a spreadsheet and in histograms with percent on the vertical scale that highlight the distinctive features of the distribution of the ages from each sample.
- Connect descriptions of the histograms to summaries based on box plots.
- Learn that samples are not exact pictures of the population but sample statistics can provide a reasonable way to develop estimates of the total populations of each country.
- As an extension, students create a sample of their "own" country and put together a summary of the country based on the sample.

Prerequisite Knowledge

Students should have prior experience with the following:

- Interpreting histograms and box plots
- Working with random samples

Analyzing Country Data

To begin comparing the the ages of people from the USA, Japan, and Kenya, move to page 1.2. Answer the questions on the student worksheet.

Tech Tips:

- This activity includes screen captures taken from the TI-Nspire CX handheld. It is also appropriate for use with the TI-Nspire family of products including TI-Nspire software and TI-Nspire App. Slight variations to these directions may be required if using other technologies besides the handheld.
- Watch for additional Tech Tips throughout the activity for the specific technology you are using.
- Access free tutorials at http://education.ti.com/calculators/pd/US/Online-Learning/Tutorials

Lesson Files:
Student Activity
- Analyzing_Country_Data_Student.pdf
- Analyzing_Country_Data_Student.doc

TI-Nspire document
- Analyzing_Country_Data.tns

TI-Nspire™ Navigator™ System

- Send out the *Analyzing_Country_Data.tns* file.
- Monitor student progress using Class Capture.
- Use Quick Poll to compare student sample means.

Activity Materials

- Compatible TI Technologies: TI-Nspire™ CX Handhelds,   TI-Nspire™ Apps for iPad®,   TI-Nspire™ Software

Discussion Points and Possible Answers

> **Teacher Tip:** Middle grade teachers may choose to stop after question 7.

Move to page 1.2.

1. The spreadsheet shows a sample of the ages of 200 people from each of the United States, Kenya and Japan. The numbers have been rounded to the nearest integer not larger than the number, so a "2" in a cell indicates someone who is two years old but not yet three.

    a. Identify the age of the youngest person and the age of the oldest person in each of the three samples.

    > **Tech Tip:** Students could use the **max** command to get the maximum age in each list. For example, in cell D1, type the following: =max(us_pop). In cell D2, type =max(kenya_pop). In cell D3, type =max(japan_pop). You can use [var] to select the list instead of typing it.

    > **Tech Tip:** For hints to use a List & Spreadsheet press [ctrl] [trig] for [?]. Shortcuts listed include [ctrl] [9] for page up and [ctrl] [3] for page down. Also [ctrl] [1] will take you to the end of the data set; [ctrl] [7] is a shortcut to return to the beginning.

    Sample Answers: The ages in the spreadsheet are listed in ascending order, so students should observe that the oldest age would be found as the last entry of the samples.
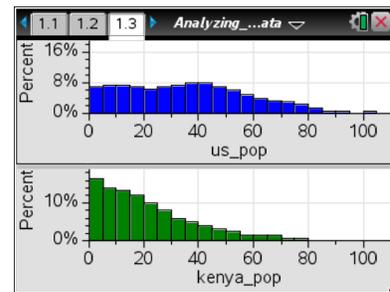
Country	Youngest	Oldest
USA	0 (or not yet 1)	100
Kenya	0	77
Japan	0	100

b. Based on the list of ages from each sample, how do you think a summary of the population of the USA differs from a summary of the population of Kenya? In a similar way, how would a summary of the USA population differ from a summary of the population of Japan? Explain your answers.

Sample Answers: As the ages are listed in ascending order, students observe that Kenya has noticeably younger ages. Within a few entries in the lists, the ages for Kenya lag behind the ages listed for the USA and Japan. Although more subtle, students might observe that the ages in the Japanese list are generally older as they scroll through the list. The 198th entry, for example, has the USA entry as 87 years old and the Japanese entry as 88 years old.

Move to page 1.3.

The histogram in the upper region of the screen displays the sample population data for the USA and the histogram in the lower region of the screen displays the sample population data for Kenya. Note that the vertical scale for each histogram is percent.

2. Describe at least two differences between the histogram of the USA sample and the histogram of the Kenya sample.

Sample Answers: The histograms provide a more visual representation of the comments provided in 1b. The percentage of people from 0 to 9 years old is greater in the Kenya sample; the greatest percentage for Kenya is in the first bar of the histogram (0 to 4 year olds). Possible answers would include the following: 1) higher percentage of the sample in the younger age intervals for Kenya, 2) bulge of the USA sample, representing the higher percentage of people in the middle age categories (30 to 44 year olds), 3) noticeably lower percentage of people in the older ages for Kenya (indicating how few people in the sample were "old"), or 4) other summaries indicating the overall younger distribution of ages in the Kenya sample compared to the USA sample.

Tech Tip: If you change the data or bin width by mistake, select `ctrl` `Z` to return the data to the original state or to undo an action; select `ctrl` `Y` to redo an action.

> **Tech Tip:** If you change the data by mistake, select the undo/redo icon to return the data to the original state or to undo an action; select and hold on the icon and select **Redo** to redo an action.

> **Tech Tip:** To display each of the five-number summary values on the boxplot, the students should tap the line or region on the boxplot.

3. Move the cursor over the bins in the histogram.
   a. Stop the cursor at the first bin for the USA. What ages are represented by this bin? How many people from the sample does this bin represent?
      Answer: This bin represents people who are between 0 and 4 years old. Since 7% of people in the USA sample are 0 - 4 years old, this bin represents 14 people.

      > **Teacher Tip:** Students might not be familiar with the notation [0,5), which means 0 is included in the bin as well as the numbers up to 5, but 5 is not a member; $0 \leq x < 5$.

   b. What percent of the Kenya sample represents 0, 1, 2, 3, and 4-year olds? How many people does that percentage represent?

      Answer: 16.5% of people in the Kenya sample are 0 - 4 years old. This percentage represents 33 people.

   c. What percent of the sample population in each of the two countries is under 30 years old? Explain how you found your answer.

      Answer: USA: 42.5%; Kenya: 74.0%. Students could find the percents by moving the cursor over each of the bars of the histogram and adding the percents. Students could also count the number of people in the spreadsheets who are under 30 and calculate the corresponding percent.

   d. What percent of the sample population in each of the two countries is 65 and older?

      Answer: USA: 12.0%; Kenya: 2.50%

   e. Based on the differences in the percents derived from the samples, would you change your previous descriptions of the population distributions of these two countries? Why or why not?

      Sample Answers: The histograms essentially continue to summarize the differences in the distributions in ages as previously described, namely, the greater percentage of younger people

in Kenya. The percents allow students to more specifically compare and contrast the differences in the countries and to support the observation that a larger percentage of the people in the sample from Kenya are younger than the people in the sample from the United States (or the larger percentage of older people are in the USA).

f. How is using histograms to compare the population samples from the countries different from using a spreadsheet?

Sample Answers: Histograms are more visual than the list of ages in the spreadsheet. The histograms summarize the raw numbers in various age intervals. The bins provide students the opportunity to identify specific age intervals that are noticeably different when looking at each of the countries.

> **Teacher Tip:** Mathematically proficient students use appropriate tools strategically. They consider the available tools when solving a mathematical problem. For this activity, these tools include the using spreadsheets and the statistical package on the TI-Nspire. Students will also learn the difference and advantages of box plots and histograms.

> **Teacher Tip:** Mathematically proficient students construct viable arguments. This activity is beneficial for students since they need to justify their conclusions and communicate them to others. Students reason inductively about data, making plausible arguments that take into account the context from which the data arose.

4. a. What ages are summarized by [10,15)?

    Answer: The ages summarized are 10, 11, 12, 13, and 14-year olds, or 10- to 14-year olds.

    b. What percent of each sample is in that age interval?

    Answer: USA: 7.5%; Kenya: 13.5%

    c. Without using the spreadsheet of the actual data, estimate the percent of people in this interval who are 13 or 14 years old. Explain how you derived your estimate.

    Sample Answers: Students should observe that people 13 and 14 years old might be approximately 2/5 of the percent in the interval [10,15). Based on this summary:
    - USA: 2/5 of 7.50%, or 3.0% of the sample.
    - Kenya: 2/5 of 13.50%, or 5.4% of the sample.

d. Based on your estimate in c, what percent of the USA sample are teenagers (13 to 19 years old)?

Answer: 10.00%. The estimated percent of 13 and 14 year olds in the sample is 3.00%, and 7.00% of the sample is 15 to 19 years old.

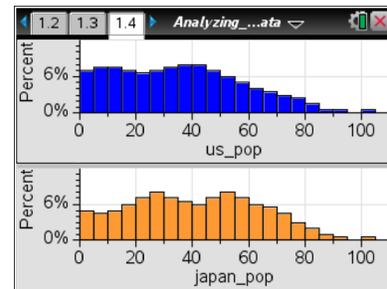e. In a similar way, estimate the percent of the Kenya sample who are teenagers.

Answer: 17.4%. The estimated percent of 13 and 14 year olds in the sample is 5.4%, and 12.00% of the sample is 15 to 19 years old.

> **Teacher Tip:** You might want to have students check their estimates against the actual data. If so, point out that in most cases, the actual data are not available and the observer has to make inferences from the representation and summaries of the data.

Move to page 1.4.

5. a. What percent of the Japanese sample population is in the age bracket [0,5) years old?

Answer: The histogram indicates 5.00% of the sample is 0 to 4 years old.

b. What percent of the Japanese sample population is under 30 years old?

Answer: Using the histogram, 35.5% of the sample is less than 30 years old.

c. What percent of the Japanese sample population is 65 and older?

Answer: Using the histogram, 17.00% of the sample is 65 and older.

d. Using the numbers in the USA and Japan who are in the [0,5) age range and those 65 and over, compare the ages of people in the USA and in Japan. Explain your thinking.

Sample Answers: Based on the samples, in Japan 5% of the people are 0 – 4 years old compared to 7% in the USA. In Japan 17% of the people are people 65 and older, compared to 12% in the USA. This suggests the older profile of the people sampled in Japan.
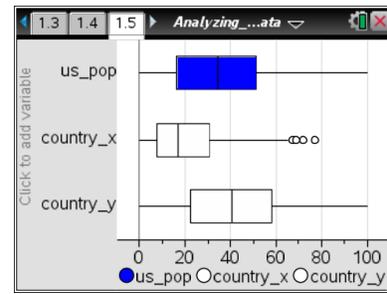
6. Estimate the percent of teenagers in Japan in a way similar to the one you used to estimate the percent of teenagers for the USA and Kenya.

   Answer: Using the same method, 2/5 of the interval containing 10 – 14 year olds would provide an estimate of the 13 and 14 year olds. Based on the histogram, this would be approximated by 2/5 of 5%, or 2%. The percent in the 15 to 19 year old interval is 6%. Therefore, an estimate of the number of teenagers in the Japan sample would be 8%. (This compares to 10% for the USA and 17.4% for Kenya.)

Move to page 1.5.

7. Which country do you think is represented by the box plot for *country_x*? Which country do you think is represented by the box plot for *country_y*? Explain your reasoning.

   Sample Answers: Using the general descriptions students developed in their previous answers, country_x has a larger percentage of younger people in the sample than the USA; therefore, students should connect this to Kenya. The box plot for country_y suggests the ages in the sample are just a bit older than in the USA, so that would fit with the sample of ages from Japan.

8. Move the cursor over the box plots on page 1.5, and fill out the tables for each country.

Answer:

a. USA sample:

Minimum age	Q1	Median Age	Q3	Maximum age
0 years old	16.5 years old	34.5 years old	51.5 years old	100 years old

b. *country_x:*

Minimum age	Q1	Median Age	Q3	Maximum age
0 years old	7.5 years old	17 years old	30.5 years old	77 years old

c. *country_y:*

Minimum age	Q1	Median Age	Q3	Maximum age
0 years old	22.5 years old	40.5 years old	58 years old	100 years old

9. a. What additional information do the box plots provide that were not provided in the histograms?

Sample Answers: Box plots provide an estimate of the median ages of the people represented in the samples, a visual representation of the distribution of age intervals where the age intervals between the quartiles represent 25% of the population. 25% of the sample from Japan is 22.5 years old and younger, compared to 7.5 years old and younger for Kenya and 16.5 years old and younger for the USA. As students note these age ranges capture 25% of the total sample, they begin to see how much younger the age profile of Kenya is compared to the USA and Japan. Similarly, 25% of the sample is 58 years old and older in Japan, 25% of the USA sample is 51.5 year old and older, and 25% of the Kenya sample is 30.5 years old and older. Box plots allow students to compare the median ages and quartiles of the samples.

10. According to the US Census Bureau in the year 2000, the population of the United States was approximately 285 million people.
    a. If the USA sample provided in this investigation were representative of the United States, what is an estimate of the number of people 0 to 9 years old?

    Sample Answers: The histogram indicates that in the sample 14.50% of the people are 0 to 9 years old. Based on the sample, 14.50% of 285 million people would be approximately 41.3 million people.

b. If another sample of 200 people were obtained from the USA population, do you think you would get the same percent of people 0 to 9 years old? Why or why not?

Sample Answers: Students should recognize that samples from the same population will vary and that the outcome from one sample is likely to be different that the outcome from another sample.

> **Teacher Tip:** One of the most difficult things for students to recognize is that if it were possible to draw many samples from a population, they would vary and thus, the result from one sample must be interpreted as one of many possibilities. Statistical inference is about how to make sense of these possibilities with some degree of confidence.

c. Statistical techniques suggest that the variation in estimates from sample to sample for a sample size of 200 is about + or - 7%. For the percent of the 0 to 9-year olds from the sample, what interval of percents might be seen in multiple samples of size 200?

Sample Answers: Based on the summary from the USA sample, other samples of similar size from the USA might reasonably have 7.5% to 21.5% of the people in the 0 to 9 years old range.

d. Given that there are about 285 million people in the USA, use your answer in part c to determine a reasonable range of estimates for the number of people in the USA who are 0 to 9 years old.

Sample Answers: 7.5% of 285 million is approximately 21.38 million people. 21.5% of 285 million is approximately 61.28 million people. Therefore, an interval with a reasonable range of estimates for the number of people 0 to 9 years old in the USA is 21.38 to 61.28 million people.

> **Teacher Tip:** This results in a rather large interval. Encourage a discussion that would help students think about what might be involved in obtaining an interval that is not as wide. More of the thinking behind reasoning from samples to the population is developed in an investigation of confidence intervals and margin of error (See Statistics Nspired Activities *Confidence Intervals for Proportions* and *Confidence Levels for Proportions*.)

11. The estimated population of Kenya in 2000 was 31 million people, and the estimated population of Japan in 2000 was 127 million people. Use the information about the samples of 200 people in each country in the TI-Nspire activity and the +/- 7% rule to estimate the number of people within the age categories specified in the table.

Sample Answers:

	United States	Kenya	Japan
Ages 0-9			
Sample percent	**14.5%**	**30.5%**	**9.5%**
Interval	**7.5% to 21.5%**	**23.5% to 37.5%**	**2.5% to 16.5%**
Estimated number of people	**21.4 to 61.3 million**	**7.3 to 11.6 million**	**3.2 to 21 million**
Older than 65			
Sample percent	**12%**	**2.5%**	**17%**
Interval	**5% to 19%**	**0% to 9.5%**	**10% to 24%**
Estimated number of people	**14.3 to 54.2 million**	**0 to 2.9 million**	**12.7 to 30.5 million**
Teenagers (13-19)			
Sample percent	**10%**	**17.4%**	**8%**
Interval	**3% to 17%**	**10.4% to 24.4%**	**1% to 15%**
Estimated number of people	**8.6 to 48.5 million**	**3.2 to 7.6 million**	**1.3 to 19.1 million**

Teacher Tip: Students could be directed to websites such as the US Census Bureau to determine the current or most recent census data.

Extension:

Create your own country that has the following characteristics:

- The total population of your country is approximately 140 million people who are 0 to 100 years old.
- 15% of the people are 0 to 9 years old.
- 25% of the people are 65 to 100 years old.
- The median age is approximately 37 years old.

1. Draw a prediction of your estimation for what the histogram and boxplot would look like.

2. Create a sample of 200 people for your country in the spreadsheet. Be sure to label the country at the top of the column.

3. Insert new pages ([ctrl] [doc▾]), and graph a histogram and box plot of your sample.

4. Compare your country's histogram and box plot to that of the United States. (Don't forget to name your country.)

- Refer to the Statistics Math Nspired Activities *Multiple Boxplots* and *How to Make Histograms* if you need help making these plots.

Sample Answers:

Students can develop their country in several ways. A good starting point would be to address the median age. Since the students are expected to create a sample of 200 people, a median value would be determined by the ages in the 100th and 101st entry (if the ages are entered in ascending order). To create a sample where 15% of the ages were 0 to 9 years old, students need to assign ages 0 to 9 in the first 30 entries. Entries 31 to 99 would range from 10 to 37 years old. To design a sample that has 25% of the people with ages 65 to 100, students would enter ages in the last quarter with ages 65 years old to 100 years old. This would be accomplished with age entries for the 151st to 200th person of 65 to 100 years old. Entries for the 102nd to the 150th person would be in the range of 37 years to 65 years old. Whatever methods students use, teachers should verify whether the students correctly developed a sample to fit the descriptions of this extension.
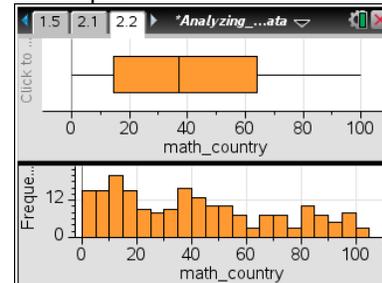
A summary of the above:

Entries 1 to 30: ages 0 to 9 years old

Entries 31 to 100: ages 10 to 37 years old

Entries 101 to 150: ages 37 to 65 years old

Entries 151 to 200: ages 65 to 100 years old

Example

Wrap Up

Upon completion of the lesson, the teacher should ensure that students are able to understand:

- Histograms and box plots can be used to analyze and summarize data.
- Histograms and box plots are good tools for comparing two or more data sets.