Tell It Like It Is!

David J. Edwards
Virginia Commonwealth University
dedwards7@vcu.edu

Published: April 2012

Overview of Lesson
One of the most important aspects of statistics is writing a conclusion. It can be compared to reading the last chapter in a book or writing the final paragraph in an essay. As students progress through their statistical education, they usually develop very strong graphing skills. They also have a high understanding of measures of central tendency (mean, mode and median) and measures of spread (range, interquartile range, and standard deviation). After this lesson, students should be able to use correct statistical language in a conclusion, recognize the need to look at the data and graphs, and think about the story the data tells.

GAISE Components
This investigation follows the four components of statistical problem solving put forth in the *Guidelines for Assessment and Instruction in Statistics Education (GAISE) Report*. The four components are: formulate a question, design and implement a plan to collect data, analyze the data by measures and graphs, and interpret the results in the context of the original question. This is a GAISE Level B activity.

Common Core State Standards for Mathematical Practice
1. Make sense of problems and persevere in solving them.
2. Reason abstractly and quantitatively.
3. Construct viable arguments and critique the reasoning of others.
6. Attend to precision.

Common Core State Standards Grade Level Content (Grades 6-8)
6. SP. 5. Summarize numerical data sets in relation to their context, such as by:
a. Reporting the number of observations.
b. Describing the nature of the attribute under investigation, including how it was measured and its units of measurement.
c. Giving quantitative measures of center and variability, as well as describing any overall pattern and any striking deviations from the overall pattern with reference to the context in which the data were gathered.
d. Relating the choice of measures of center and variability to the shape of the data distributions and the context in which the data were gathered.
7. SP. 1. Understand that statistics can be used to gain information about a population by examining a sample of the population; generalizations about a population from a sample are valid only if the sample is representative of that population. Understand that random sampling tends to produce representative samples and support valid inferences.
4. Use measures of center and measures of variability for numerical data from random samples to draw informal comparative inferences about two populations.

Contact Author for permission to use materials from this STEW lesson in a publication

8. SP. 1. Construct and interpret scatter plots for bivariate measurement data to investigate patterns of association between two quantities. Describe patterns such as clustering, outliers, positive or negative association, linear association, and nonlinear association.
4. Understand that patterns of association can also be seen in bivariate categorical data by displaying frequencies and relative frequencies in a two-way table. Construct and interpret a two-way table summarizing data on two categorical variables collected from the same subjects.

NCTM Principles and Standards for School Mathematics
Data Analysis and Probability Standards for Grades 6-8
    Select and use appropriate statistical methods to analyze data:
- find, use, and interpret measures of center and spread, including mean and interquartile range;
- discuss and understand the correspondence between data sets and their graphical representations, especially histograms, stem-and-leaf plots, box plots, and scatter plots.

    Develop and evaluate inferences and predictions that are based on data:
- use observations about differences between two or more samples to make conjectures about the populations from which the samples were taken;
- make conjectures about possible relationships between two characteristics of a sample on the basis of scatterplots of the data and approximate lines of fit;
- use conjectures to formulate new questions and plan new studies to answer them.

Prerequisites
Students should have prior knowledge of:
- the calculation and interpretation of measures of center (mean, median, mode) and spread (range, variance, standard deviation, etc.)
- the interpretation of graphical displays (e.g., histogram, bar chart, pie chart, boxplot)
- the language of population and sample
- the identification of potential outliers in a dataset

Learning Targets
Students will be able to:
- use correct statistical language in a conclusion
- recognize the need to look at the data and associated graphs
- think about the overall story that the data conveys

Time Required
One class period (60-75 minutes)

Materials Required
- Student worksheet
- Dataset A handout
- Dataset B handout

Instructional Lesson Plan

The GAISE Statistical Problem Solving Procedure

I. Formulate Question(s)
Begin by discussing the nature of the activity. A brief review of the language of population and sample may be beneficial at this point. For instance, you may wish to start this activity by talking about why companies/statisticians/etc. take a sample of the population rather than taking a full census. Some of the main reasons are:

- Time
- Cost (reduces cost)
- Practicality (near impossible to ask population)
- Resources (people and materials)

Most data you work with will come from a sample of the population rather than the whole population. Remind students that one of the most important aspects of statistics is writing a conclusion. A **statistical** conclusion links the evidence from the sample to a generalization for the population. It begins with what seems to be true about the sample. Then it answers the question "How do you know this is true?" Usually there are two sorts of evidence: numerical values of statistics and a description of graph(s). The conclusion then sums up what is expected to be true for the population based upon what is seen in the sample.

Explain to students that this lesson will help them answer the question, "What can I conclude from the sample data?" and as a consequence "Tell it like it is."

II. Design and Implement a Plan to Collect Data
The activity is focused on writing a conclusion and therefore data has already been collected and presented for writing a conclusion. Tell students that the data they will be using has been collected from a database called CensusAtSchool. The population is students living in the country of New Zealand. Students will be provided handouts with data summarized in graphs and tables. These handouts will then be utilized for learning effective writing of statistical conclusions.

III. Analyze the Data

Part A
Provide students Handout 1: Dataset A (page 11). Encourage students to make "I notice…" statements about each graph and table. These statements could then be collected on the board for later reference as students complete the worksheet. Next, provide the Tell It Like It Is! Activity Sheet (page 9) and ask students to fill in the blanks using their observations made from the dataset A graphs and tables.

Solutions to Fill-in-the-Blank:

From this particular 1**sample** , the 2**table of data** suggests that there are 3**slightly** more girls than boys in the CensusAtSchool 4**population**. The 5**bar graph** indicates that 6**Auckland** had the most respondents while 7**Southland** and 8**West Coast** had the least. From this data, it can be seen that the 9**median** age of a student in the sample was 12. If another sample of 200 students was taken, we could expect the age to be around 10**11-13** years old. Travelling to school by 11**motor car** appears to be the most popular method. If another sample was taken, it is likely that the most popular method of travelling to school would be 12**motor car** . The data also showed that cell phone ownership is high for students with the data showing that 13 **64%** of the 14**sample** own a cell phone. This 15**suggests** that more students own a cell phone than don't own a cell phone.

Part B
Provide students Handout 2: Dataset B (pages 12 and 13). Ask students, "What questions could we answer using this dataset?" Ask students to pose questions for each graph and data set. Explain that when students make a conclusion they should:
- Answer the posed questions – a good beginning is "Based on the evidence from the data/bar graph/table, it seems…"
- Quantify their answer – quote numerical values (with units) of statistics/proportions and describe graph features.
- Generalize the answer (make an inference) to all elements in the population. One possible starter for the generalization is "I expect…"

This part allows students to create and write their own conclusions. Remind students that when writing a conclusion, it is important that they avoid making causality statements (e.g., eating ice cream in the summer causes shark attacks) in their findings. They should stress that the data is from a sample and that if another sample was taken the results may be different. A list of suitable sentence starters has been included to assist students in creating their conclusions. Write the following on the board (or transparency):
- From these data…
- From this sample…
- It appears that…
- The sample data collected from… (Name source of data)
- There is evidence to suggest…
- From the (type) graph …
- From these data, it is likely that the population may …(inference)
- Resampling the population may result in…
- From this sample, the (average value) was… This suggests that…
- The data is…
- Based on these data …
- I noticed that …
- Further investigation may assist …
- I wonder if …

Now ask students to write their own conclusions for each table and graph of the data in the space provided on the activity worksheet.

Below is a model answer. Students are not expected to be as complete with their conclusions, especially if they have minimal prior experience with writing a statistical conclusion.

MODEL ANSWER for Part B

It appears that there are similar amounts of boys and girls from this sample. The sample also indicates that around 70% of both boys and girls play sports. Cell phone ownership appears more important to girls than boys, with about 20% more girls (70%) owning a cell phone than boys (50%). Boys, on the other hand are more likely to own their own TV than a girl with 51% of the boys owning their own TV compared to 40% for girls. The table of lunchtime activities indicates that as students get older they get less active during lunchtime. The main activity for Year 5 to Year 8 students is running around and playing. For Year 9 and Year 10 students, the most common activity is sitting around and talking. The bar graph shows cell phone ownership by region as a percentage. All the students from Taranaki own a cell phone in this sample while no students from Southland or the Hawkes Bay regions owned cell phones. A larger sample may give a better view of the population as smaller regions are likely to be misrepresented due to having very few students representing their region. The scatter plot shows students' heights against age. The height of students is generally increasing as age increases. This is what we would expect. There is an interesting potential outlier, a student who is 13 years old and has a height of 100 cm. This is much lower than expected for the average 13 year old. I wonder if this is a correct measurement or an error.

IV. Interpret the Results

Remind students that this activity is focused on writing a conclusion and effective interpretation of statistical data. Students will now spend time as a class combining their own conclusions to write a model interpretative conclusion of the data. If needed, spend additional time going over the fill-in-the-blank answers to Part A and clarifying any difficult to understand answers. In Part B, students wrote their own interpretative conclusions for each dataset B graph and table. For the first graph, have students work in cooperative groups and have each group present their own conclusions. (This gets students to talk to each other first and encourages them to use statistical vocabulary). Then, collect conclusions from each group and write these on the board. Use these to write a model statistical conclusion as a class. Repeat this for the second graph and tables.

Assessment

College applications take time for both applicants and administrators. A university conducted an investigation of the amount of time required to enter the information contained in an application for admission into the university computer system (Peck and Devore (2012)). Completion times (in minutes) for 50 randomly selected applications forms were noted. The data is summarized below with a histogram and several descriptive statistics.
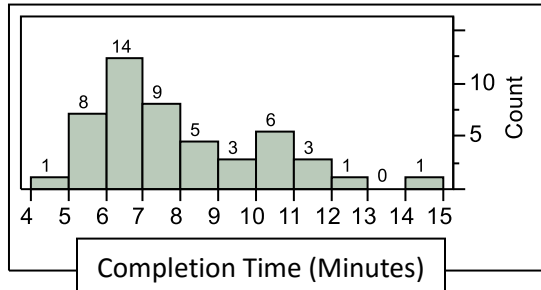
Mean	7.85
Median	7.40
Standard Deviation	2.13

1. Fill in the blanks to summarize what the descriptive statistics tell us about application entry times. Select from the options to the right of each blank.

Based on the evidence from the data, the average time required to enter admissions data was

_____ (7.85 / 7.40), but the relatively large standard deviation suggests that _____ (few /

many) observations differ substantially from the _____ (mean / median). The _____

(mean / median) tells us that half of the applications required less than _____ (7.40 / 7.85)

minutes to enter. The fact that the _____ (mean / median) exceeds the _____ (mean /

median) suggests that some unusually _____ (small / large) values in the _____ (sample /

population) affected the value of the mean.

2. Write an interpretive conclusion about application entry times based on the histogram graph. Start with "Based on the evidence from the histogram…"

Answers to the Assessment

1. Based on the evidence from the data, the average time required to enter admissions data was

_**7.85**_ (7.85 / 7.40), but the relatively large standard deviation suggests that _**many**_ (few /

many) observations differ substantially from the _**mean**_ (mean / median). The _**median**_

(mean / median) tells us that half of the applications required less than _**7.40**_ (7.40 / 7.85)

minutes to enter. The fact that the _**mean**_ (mean / median) exceeds the _**median**_ (mean /

median) suggests that some unusually _**large**_ (small / large) values in the dataset affected the

value of the mean.

2. Answers will vary. Here is one possibility:

Based on the evidence in the histogram, the most common application entry time is between 6
and 7 minutes. The histogram has a right skewed shape which indicates that there are some
unusually large values in the dataset. A potential outlier between 14 and 15 minutes appears to
be present. I wonder if this is an error or if the person entering the application was interrupted or
distracted. A larger sample may give a better view of the population.

Possible Extensions (to GAISE Level C)

Appropriate interpretation of statistical results and the effective writing of conclusions remain important for all levels. To extend this activity, consider writing conclusions for more advanced statistical techniques, such as simple linear regression or two-sample t-tests. For example, in regression analysis, proper interpretation of the slope/intercept parameters and the fit of the line are crucial for understanding the implications of a fitted model.

References

1. *Guidelines for Assessment and Instruction in Statistics Education* (GAISE) Report, ASA, Franklin et al., ASA, 2007 http://www.amstat.org/education/gaise/.

2. This activity is based on the activity by Joanne Woodward, titled *Tell is Like it is!* on the Census at School website http://www.censusatschool.org.nz/.

3. Peck, R. and Devore, J. L. (2012). *Statistics: The Exploration and Analysis of Data*, Cengage Learning, Boston, MA. (Assessment Data)

Tell It Like It Is! Activity Sheet
One of the most important aspects of statistics is writing your conclusion. It can be compared to reading the last chapter in a book or writing your final paragraph in an essay. In this activity, you will learn how to write an effective statistical conclusion.
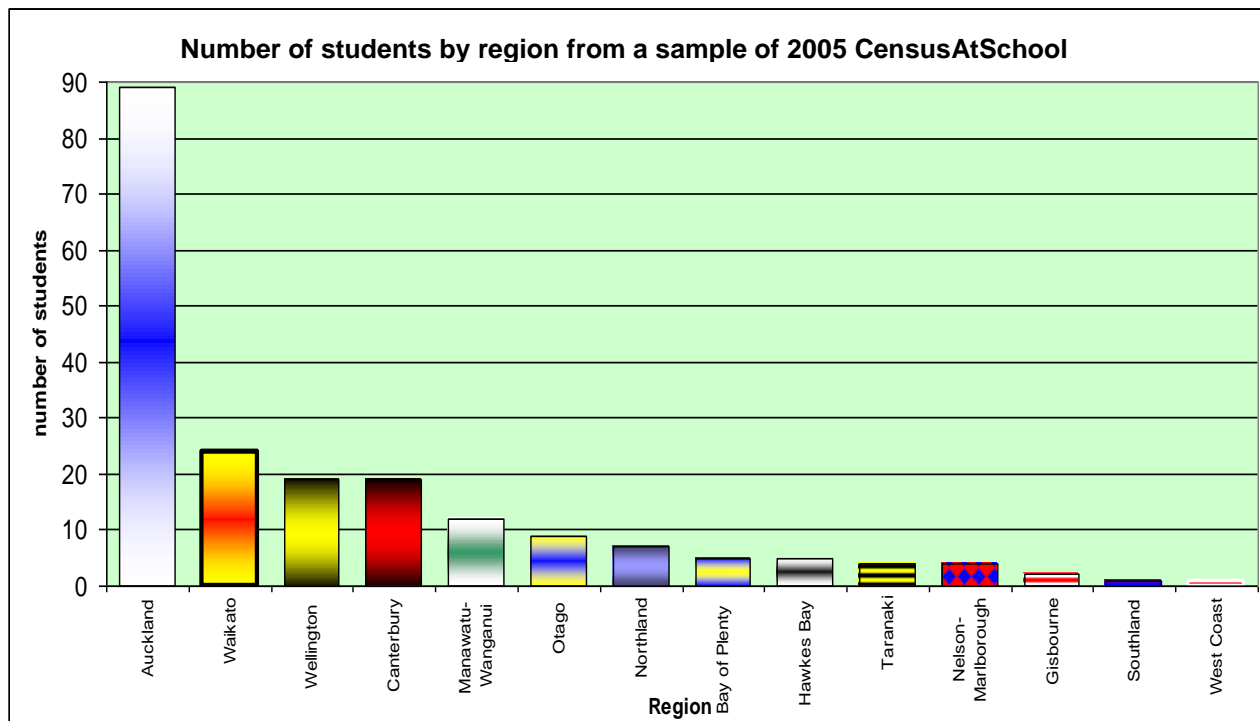
Part A
Fill in the blanks in the following paragraph about dataset A. Choose from the words on the list to the right.

From this particular 1_____, the 2_____ suggests that there are 3_____ more girls than boys in the CensusAtSchool 4_____. The 5_____ indicates that 6_____ had the most respondents while 7_____ and 8_____ had the least. From this data, it can be seen that the 9_____ age of a student in the sample was 12. If another sample of 200 students was taken, we could expect the age to be around 10_____ years old. Travelling to school by 11_____ appears to be the most popular method. If another sample was taken, it is likely that the most popular method of travelling to school would most likely be 12_____. The data also showed that cell phone ownership is high for students with the data showing that 13 _____ of the 14_____ own a cell phone. This 15_____ that more students own a cell phone than don't own a cell phone.	1 **Population/sample** 2 **Table of data/ graph** 3 **Slightly / a lot** 4 **Population /sample** 5 **Histogram / bar graph** 6 **Auckland / North Island** 7 **Southland / Canterbury** 8 **West Cost / Gisbourne** 9 **Mean / median** 10 **9 -11 / 11 – 13** 11 **Motor car / Train** 12 **Motor car / walking** 13 **36% /64%** 14 **sample / population** 15 **proves / suggests**

Part B

Use dataset B to write your own conclusion for each graph and table of data.

_____

_____

_____

_____

_____

_____

_____

_____

_____

_____

_____

Handout 1: Dataset A

Number of students by region from a sample of 2005 CensusAtSchool

Region	Count	Own TV
Auckland	89	37
Waikato	24	9
Wellington	19	9
Canterbury	18	12
Manawatu-Wanganui	12	5
Otago	9	5
Northland	7	5
Bay of Plenty	5	3
Hawkes Bay	5	4
Taranaki	4	2
Nelson-Marlborough	4	1
Gisborne	2	2
Southland	1	1
West Coast	1	1
TOTAL	**200**	**96**

Gender	count
girls	107
Boys	93

Age	
Mean	12.115
Median	12
Mode	12

Own a Cell phone	Count	Percentage
Yes	127	64%
No	73	36%

How students travelled to school

Age	walk	motor	Bus	Train	bicycle	Other	Total
8 to 10	10	20	3	0	2	2	**37**
11 to 13	28	40	35	1	9	2	**115**
14 to 16	11	22	9	2	1	3	**48**
Total	**49**	**82**	**47**	**3**	**12**	**7**	**200**

Handout 2: Dataset B (Page 1)

Percenatge of cell phone ownerhip by region of 100 randomly selected CensusAtSchool NZ students

Height versus age for 100 randomly selected students from CensusAtSchool NZ

Contact Author for permission to use materials from this STEW lesson in a publication
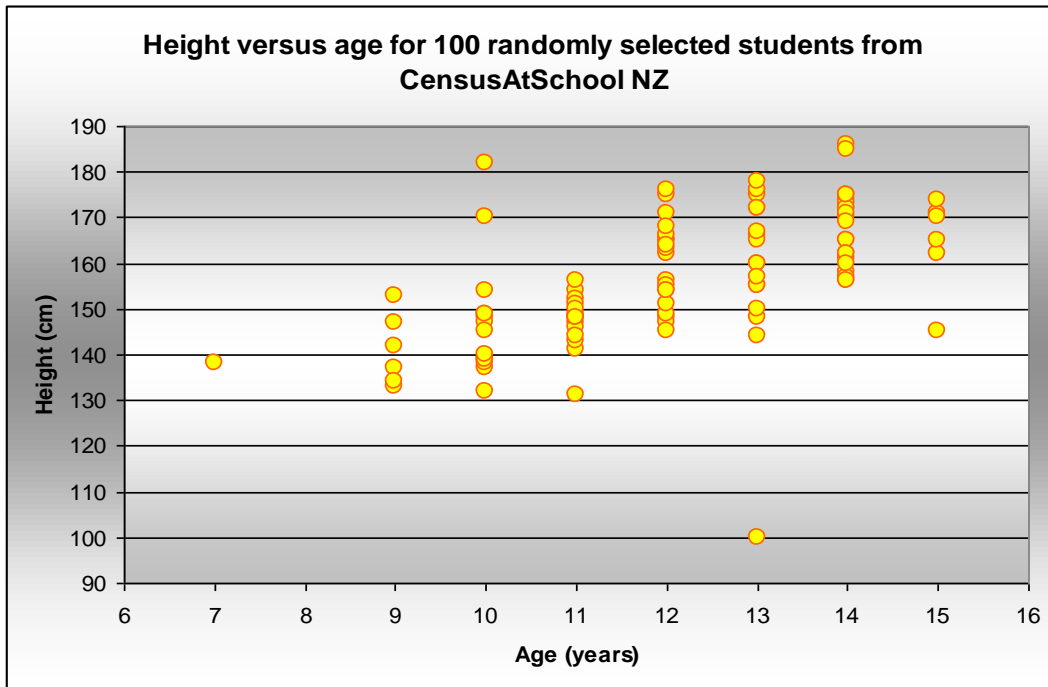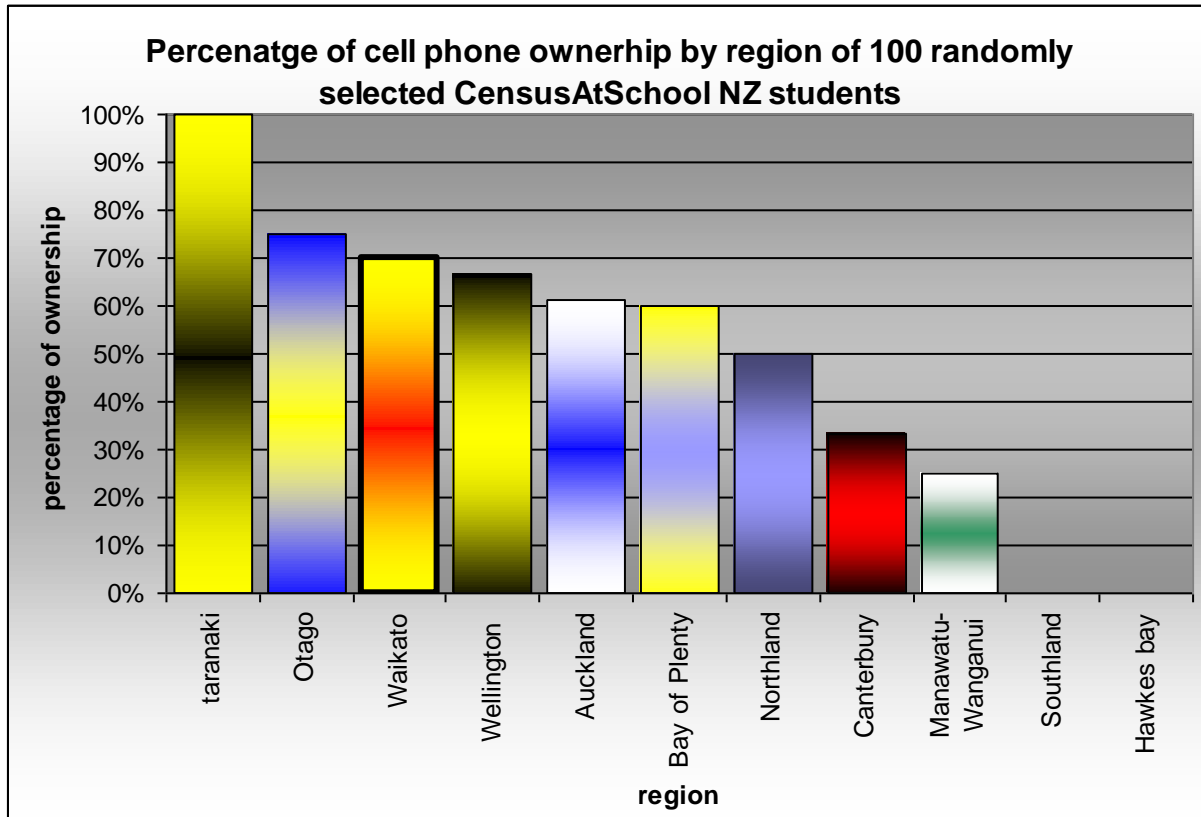
Handout 2: Dataset B (Page 2)

Percentages of ownership and sport by
Gender (random sample of 100 students)

	gender	Play Sport	Cell phone	Own TV
Boys	51%	70%	50%	51%
Girls	49%	70%	70%	40%

Count of lunchtime activities

Year	Ran around	Walked	Stood	Sat	**Total**
Y5 & Y6	18	1	1	1	**21**
Y7 & Y8	24	4	2	5	**35**
Y9 & Y10	13	12	4	15	**44**
Total	**55**	**17**	**7**	**21**	**100**