Just hit like!

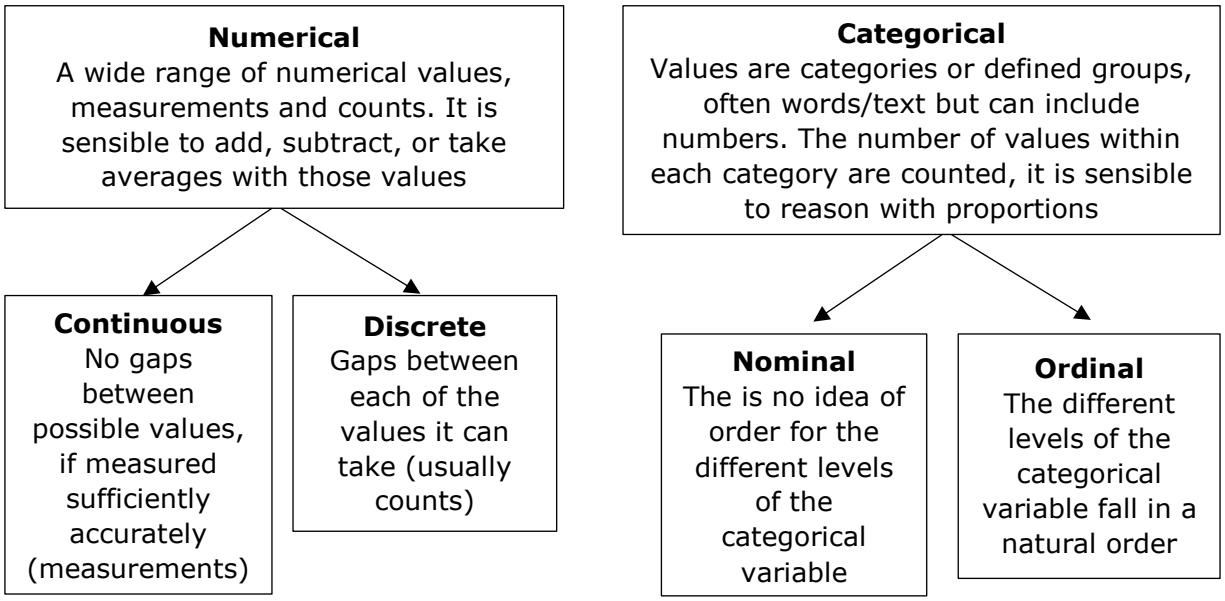
This activity was developed by combining and adapting features of these two resources developed by Pip Arnold: Engaging with shape http://new.censusatschool.org.nz/resource/engaging-with-shape/ and using data cards http://new.censusatschool.org.nz/resource/using-data-cards/

For more information about data types and displays, see the first chapter of the textbook OpenIntro: https://www.openintro.org/stat/textbook.php

1. You have been given a copy of a survey given to NZ students about their use of Facebook. Try completing the survey yourself. Were there any questions that you found hard to answer? Why?

2. Sort the question cards you have been given into groups of similar questions. Think about the answers each question will generate to help you to do this.

3. Read the definitions of each type of data (continuous, discrete, nominal, ordinal) in the diagram below and compare these to the groups you have created from the question cards. Do any of the groups you have formed match these types of data? Are there any questions you might now reconsider which group you have put them into?



We can view the classifying of data into types of variables as a relatively easy activity. However, students can struggle with identifying variables, particularly when measurement is involved. It is helpful to be able to link back to how the data was collected (in this example, by considering the question that was asked, in practical situations, by considering the tool that was used). It is also helpful to reflect on what kind of analysis would be sensible to perform with the values obtained. Go through each of the question cards and check with students which type of variable they have sorted it as.

One possible reason for students misidentifying continuous variables as categorical (or vice versa) is the level of measurement that is used for the misidentified variable. Real continuous data are discrete because we can't measure each observation to an infinite number of decimal places. This means we have to decide on a level of accuracy to use when measuring and recording that data. However, we would still consider the data continuous and treat it as such in our analysis, because our measurements represent an approximation of the underlying continuum.

Other common mistakes are categorizing **age** in years as a categorical variable, similarly **cost** in dollars as a categorical variable. One way to help students with this misunderstanding is to ask them to consider if the value they have observed represents a value from an underlying continuum, such as time or "worth". On the other hand, if **age** is measured in age categories (e.g. 5 to 10 years) then this representation of **age** really is a categorical variable. For this survey, students may want to classify **Q2** as a numerical discrete variable. However, it does not make sense to use this variable to calculate means etc. to summarize the variable, it makes much more sense to consider the proportion of students within each year level. Conversely, students may want to classify **Q10** as a categorical ordinal variable since there are words/text used to help describe the scale used. This is fine. However, the data from this question can also be meaningfully treated as numerical.

4. Try to match the question cards you have been given to the display cards. The variable names have been purposefully left off the displays so you will need to consider what values make sense for the question asked/variable. For some of the displays, two of the question cards have been used. When you are finished all questions cards should be matched to each display card.

5. For each of the question cards listed below, describe one feature of the display that stands out to you and something else you would like to know related to the variable (e.g. something that may be hidden in this display or not clear in the data, or new questions you have).

Question	**One feature that stands out is …**	**What else I would like to know is …**
Q1 & Q3		
Q2 & Q4		

Question	One feature that stands out is …	What else I would like to know is …
Q5		
Q7 & Q10		
Q8		
Q9		
Q11		
Q12		
Q14		
Q15		

While it is important for students to understand how each display is constructed, we haven't focused on that in this activity. Instead, by linking the display to the question card, we are focusing on what feature stands out to students from each particular display, to build understanding of how each display can be useful. Of course we also need to make students aware of what can be hidden in each display and to encourage students to ask more questions about the data. For example, box plots are great at giving an overall summary of dispersion, but you cannot see individual values and you cannot see how much data you have. Similarly histograms are great for highlighting the shape of the distribution, but the "binning" or group process can hide more specific features of the data.

6. Below is a table with the names of the different displays for univariate data which were used in the cards. Complete the table with the details the type of data that would be appropriate to use with each display.

Display name	**Suitable data**	**Additional information**
Stem-and-leaf plot		
Dot plot		
Box plot		
Histogram		
Frequency table		
Bar chart		
Segmented/stacked bar chart		
Time series plot		

7. Now that you have familiarized yourself with the survey questions, the type of data collected through the survey, and the kinds of displays that could be useful for exploring the data, write three ***I wonder*** questions that could be investigated with this data. Two examples of questions have been given below as an example.

I wonder how much time spent on Facebook each day is a useful predictor of how much time is spent on homework for students who completed this survey?

I wonder how the number of close friends reported by students who completed this survey compares for female and male students?

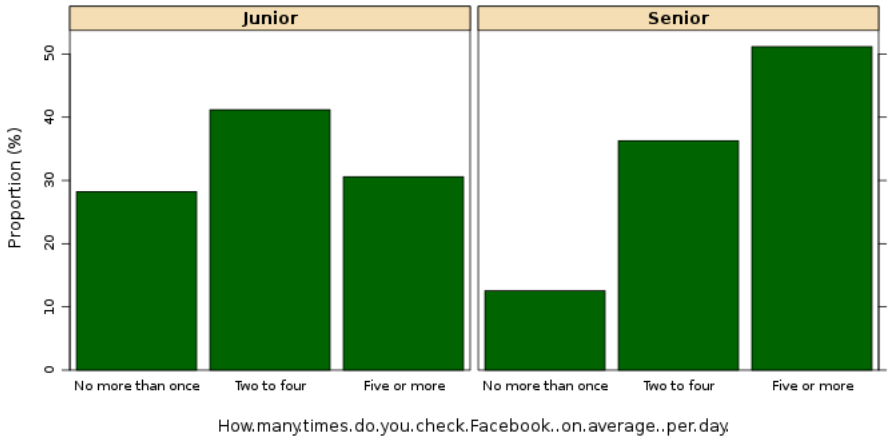
I wonder ….

I wonder ….

I wonder ….

Note for the questions above that we do not have a random sample from a well-defined population, and so these questions are all framed in terms of "the students who completed this survey". The focus for exploring this data set would be of a descriptive nature: to find patterns and relationships for these students.

8. Using only the information shown in the bar charts below, which one of the following statements cannot be made?



A. A higher percentage of senior students surveyed check their Facebook account five or more times a day than junior students

B. A higher percentage of junior students surveyed check their Facebook account no more than once a day than senior students

C. A higher number of junior students surveyed check their Facebook account two to four times a day

D. Over 80% of the senior students surveyed check their Facebook account at least twice a day

Stick person meet world

This activity was developed by combining and adapting features of these two resources developed by Pip Arnold: Engaging with shape http://new.censusatschool.org.nz/resource/engaging-with-shape/ and using data cards http://new.censusatschool.org.nz/resource/using-data-cards/. The stick people data cards and data set used in this activity are based on NZ census at school data from 2015 and can be found here: http://teaching.statistics-is-awesome.org/census-at-school-stick-people-data-cards/. For more information about calculating and describing measures/summaries, see the first chapter of the textbook OpenIntro: https://www.openintro.org/stat/textbook.php

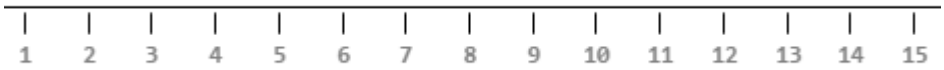
1. You have been given one stick person data card for this activity. Complete the missing row in the table below with the details of your stick person.

gender	name	name length	age	bag weight	tech phone	tech facebook	tech snapchat	tv time	reading time
male	ROBBIE	6	13	5.1	yes	yes	yes	0	1.25
female	AVA	3	12	5	yes	yes	no	1.5	1.75
female	TIANA	5	15	4.4	yes	yes	yes	0	0.5
female	SABINE	6	13	0.5	yes	yes	yes	3	0.25
female	AMBER	5	11	0.7	no	no	no	1.25	0.75
male	BENJAMIN	8	13	7.7	yes	yes	yes	0.5	0.25
male	TANISHQ	7	13		yes	yes	no	0	0.5
female	ANDY	4	13	8.3	yes	no	yes	1.25	1.25
male	DOMINIC	7	13		no	no	no	1.25	0.5

2. Based on the information shown in the table above about nine other stick people from the same world, how does your stick person compare to other stick people in this world?

Choose two different types of variables to base your comparison on i.e. one categorical variable and one numerical variable.

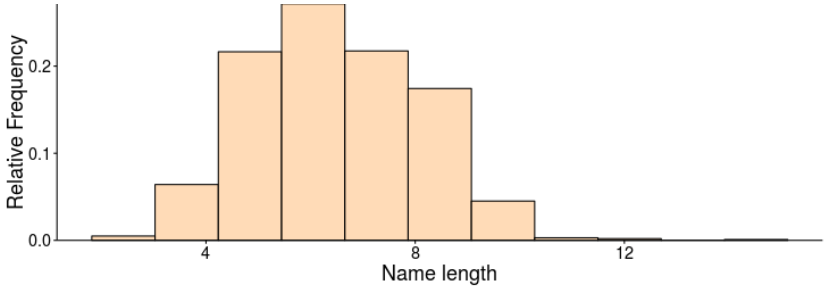
3. On the blank axis below, sketch what you think the shape of the distribution of name lengths might look like for the whole world of stick people.

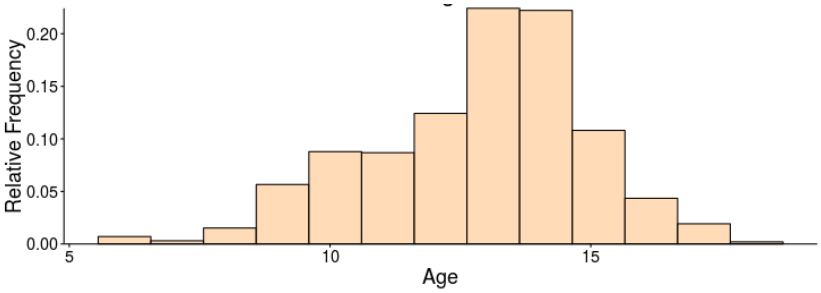


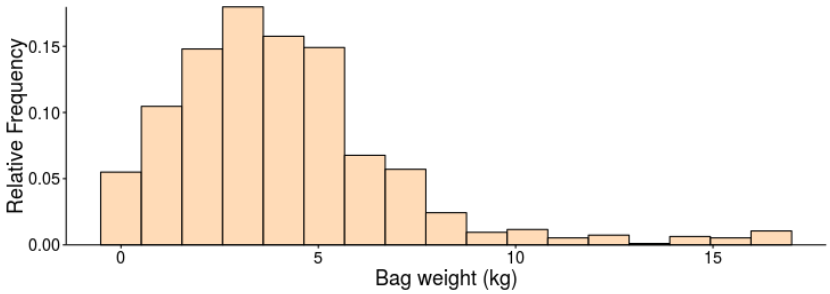
Name length (number of letters in first name)

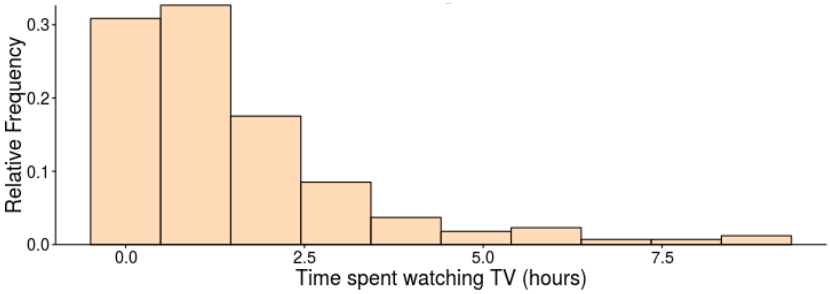
The first few questions of this activity are to help students familiarize themselves with the data card and the rectangular data structure we use to represent data sets. Q2 should prompt students to think about a *location* type comparison (my stick person is in the higher/lower half of values) and a *proportional* type comparison (my stick person is with the majority/minority). Q3 uses a variable that should be familiar to students – the lengths of names (although this data is NZ data!) – which means they can use personal contextual knowledge to help with their sketch. There are of course no right or wrong answers for Q3 – the point is to engage students in thinking what the distributional shape might look like for all the stick people.

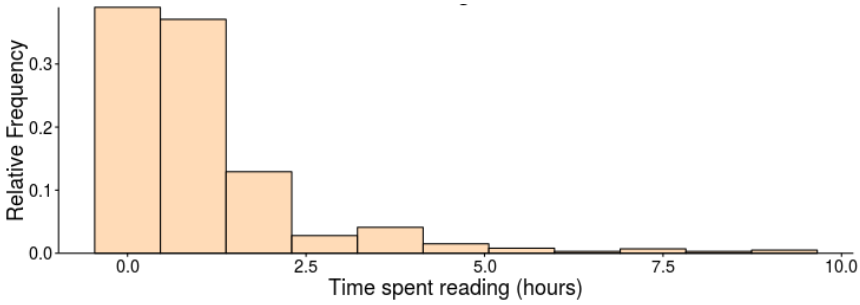
4. Below are histograms for each of the numerical variables measured for these stick people.







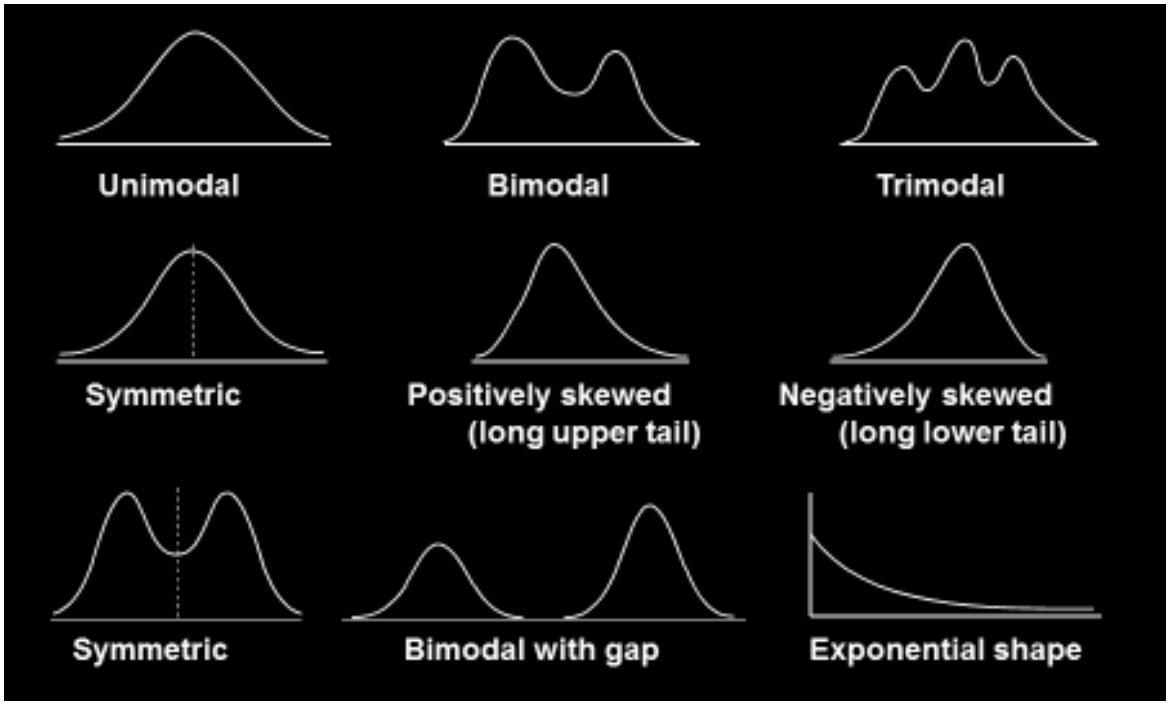




In table below, write which variables you would group together in terms of them having a similar distributional shape. Sketch this common shape. You do not need to use all of the rows of the table.

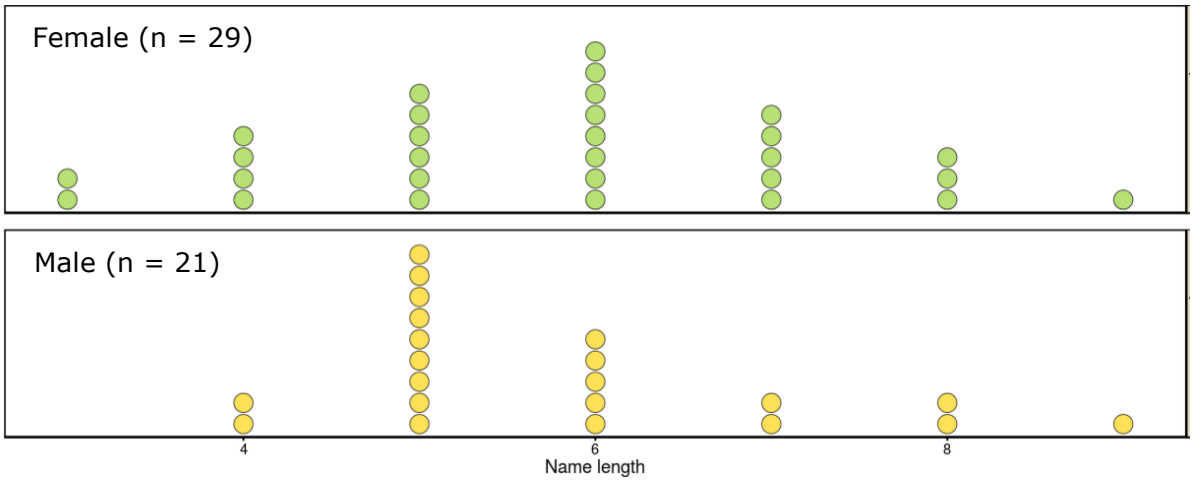
Variables I have grouped together	**My sketch of the common shape**	**Words I can use to describe this particular shape**

Below are a collection of shapes and descriptions that might be useful as a reference.



An extension for this question would be to explore how the shape observed in a histogram changes based on the bins/groupings used, and to discuss the positives and negatives of this with students. The whole stick person world was used in Q3 to create the histograms (n = 998).

5. On dot plots of name length below, locate your stick person as one of the dots by adding a face.



a. Add box plots overtop each of the dot plots.

b. Where does your stick person sit within their group in terms of name length?

c. Circle true or false for each of the statements below:

The mean name length for these male stick people is lower than the median.

True False

The name lengths for these female stick people are symmetric.

True False

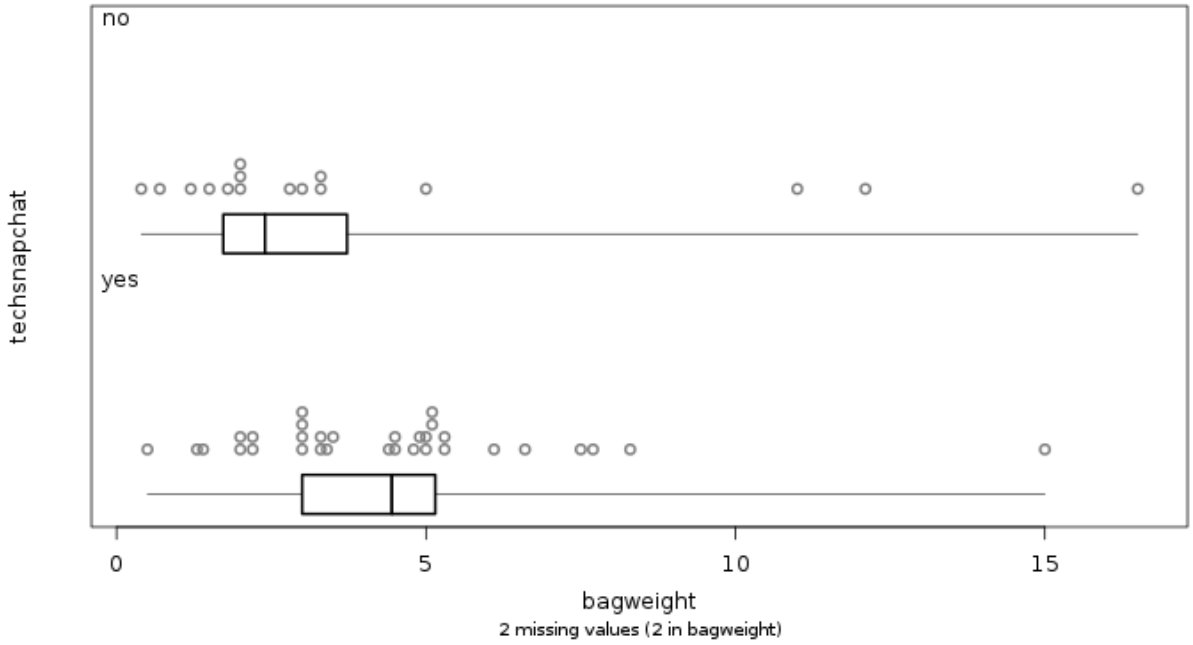
The standard deviation of the name lengths is greater for these male stick people.

True False

The mean name length is similar for these male and female stick people.

True False

6. The side-by-side box plots and dot plots below show the bag weights of these stick people compared by whether each stick person uses Snapchat or not.




```
Summary of bagweight by techsnapchat:
-------------------------------------

        Min     25%   Median     75%    Max    Mean      SD   Sample Size
   no   0.4   1.725     2.40   3.725   16.5   4.287   4.681            16
  yes   0.5   3.000     4.45   5.150   15.0   4.444   2.704            32
```

a. Locate your stick person as one of the dots by shading in the dot.

b. Is your stick person in the group with the heavier bag weights on average?

Explain how you have answered this question using features of the graphs and the summary measures.

c. Is your stick person in the group with the most varied bag weights?

Explain how you have answered this question using features of the graphs and the summary measures.

Mathematical practices through a statistical lens

Consider the 8 Mathematical Practices through a Statistical Lens. Which practices were utilized in this workshop? Explain.

Make sense of problems and persevere in solving them	
Construct viable arguments and critique the reasoning of others	
Model with mathematics	
Use appropriate tools strategically	
Attend to precision	
Look for and make use of structure	
Look for and express regularity in repeated reasoning	

Interpreting Categorical and Quantitative Data **S-ID**

Summarize, represent, and interpret data on a single count or measurement variable

1. Represent data with plots on the real number line (dot plots, histograms, and box plots).
2. Use statistics appropriate to the shape of the data distribution to compare center (median, mean) and spread (interquartile range, standard deviation) of two or more different data sets.
3. Interpret differences in shape, center, and spread in the context of the data sets, accounting for possible effects of extreme data points (outliers).
4. Use the mean and standard deviation of a data set to fit it to a normal distribution and to estimate population percentages. Recognize that there are data sets for which such a procedure is not appropriate. Use calculators, spreadsheets, and tables to estimate areas under the normal curve.

Resources

Survey used for *Just hit like!* activity

Facebook survey

Q1	What city or town do you live in?	
Q2	What is your current year level?	7 8 9 10 11 12 13 14
Q3	What is your gender?	Male Female Other:
Q4	Do you have a Facebook account?*	Yes No
Q5	How many friends do you have on Facebook?	
Q6	How many close friends do you in real life?	
Q7	Have you ever received a friend request on Facebook from someone that you did not know?	Yes No
Q8	How much time do you spend on Facebook, on average, per day? *Give your answer in hours and minutes e.g. 3 hours 20 minutes*	
Q9	How many times do you check Facebook, on average, per day?	No more than once Two to four Five or more
Q10	Do you think that teenagers spend too much time socializing on Facebook?	1 2 3 4 5 Strongly disagree ○ ○ ○ ○ ○ Strongly agree
Q11	How much time do you spend completing school work outside of school, on average, per day? *Give your answer in hours and minutes e.g. 3 hours 20 minutes*	
Q12	Which of the following would describe your main use of Facebook?	Posting and viewing photos Chatting with friends Commenting on friends' posts Other:
Q13	How much would you be prepared to pay for Facebook each month (if Facebook was no longer free)? *You do not need to enter a dollar sign for this answer*	

* If the answer was no, respondents did not complete the rest of the survey

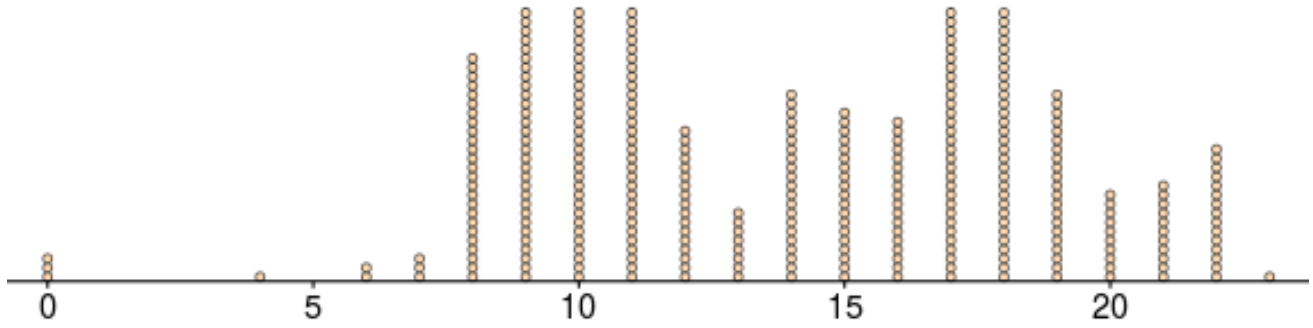
Question cards used for *Just hit like!* activity

Q1 What city or town do you live in?	**Q9** How many times do you check Facebook, on average, per day? No more than once Two to four Five or more
Q2 What is your current year level? 7 8 9 10 11 12 13 14	**Q10** Do you think that teenagers spend too much time socializing on Facebook? 1 2 3 4 5 Strongly disagree ○ ○ ○ ○ ○ Strongly agree
Q3 What is your gender? Male Female Other:	**Q11** How much time do you spend completing school work outside of school, on average, per day? *Give your answer in hours and minutes e.g. 3 hours 20 minutes*
Q4 Do you have a Facebook account? Yes No	**Q12** Which of the following would describe your main use of Facebook? Posting and viewing photos Chatting with friends Commenting on friends' posts Other:
Q5 How many friends do you have on Facebook?	**Q13** How much would you be prepared to pay for Facebook each month (if Facebook was no longer free)? *You do not need to enter a dollar sign for this answer*
Q6 How many close friends do you in real life?	**Q14*** The time of the day that the survey was completed by the student, as measured by the number of hours since midnight.
Q7 Have you ever received a friend request on Facebook from someone that you did not know? Yes No	**Q15*** The date and time that the survey was completed by the student
Q8 How much time do you spend on Facebook, on average, per day? *Give your answer in hours and minutes e.g. 3 hours 20 minutes*	

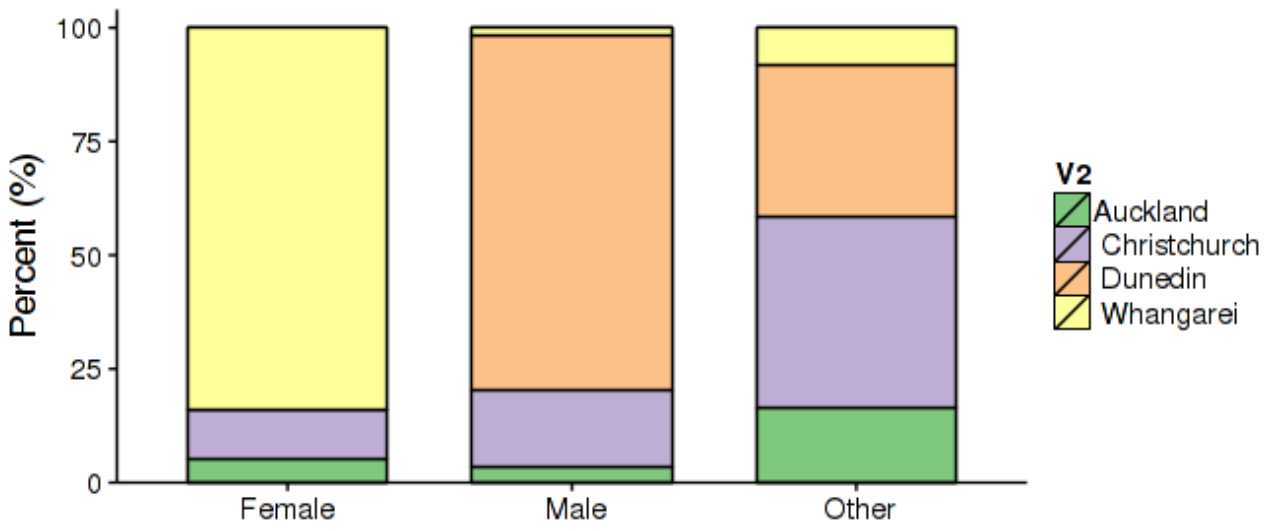
* These are based on the timestamp that was recorded when each student completed the survey

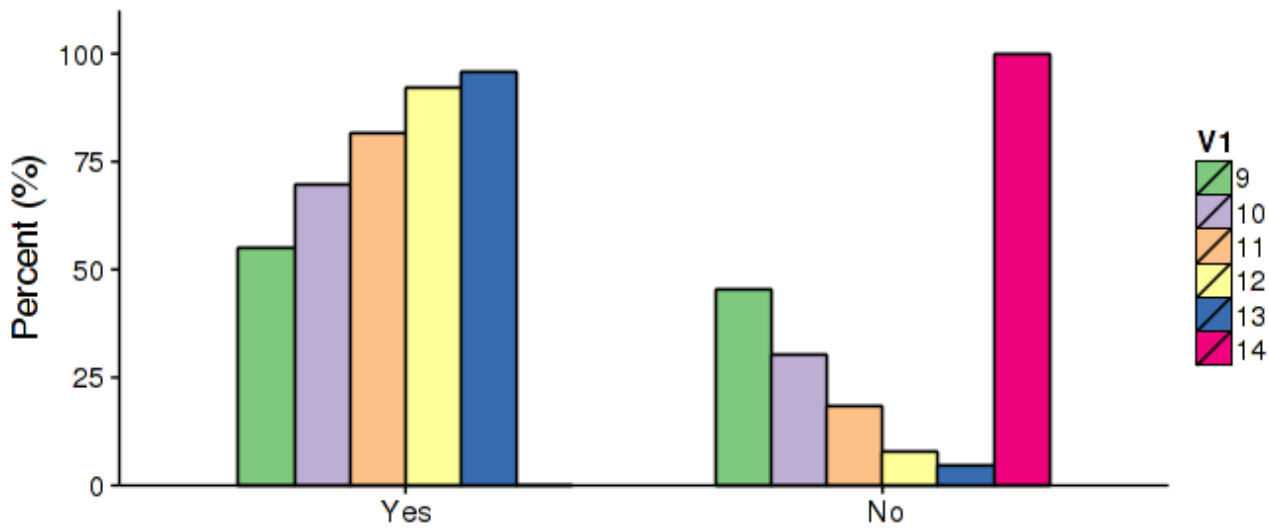
Display cards used for *Just hit like!* activity

The graphs were produced using Microsoft Excel and the online apps from http://www.artofstat.com/webapps.html

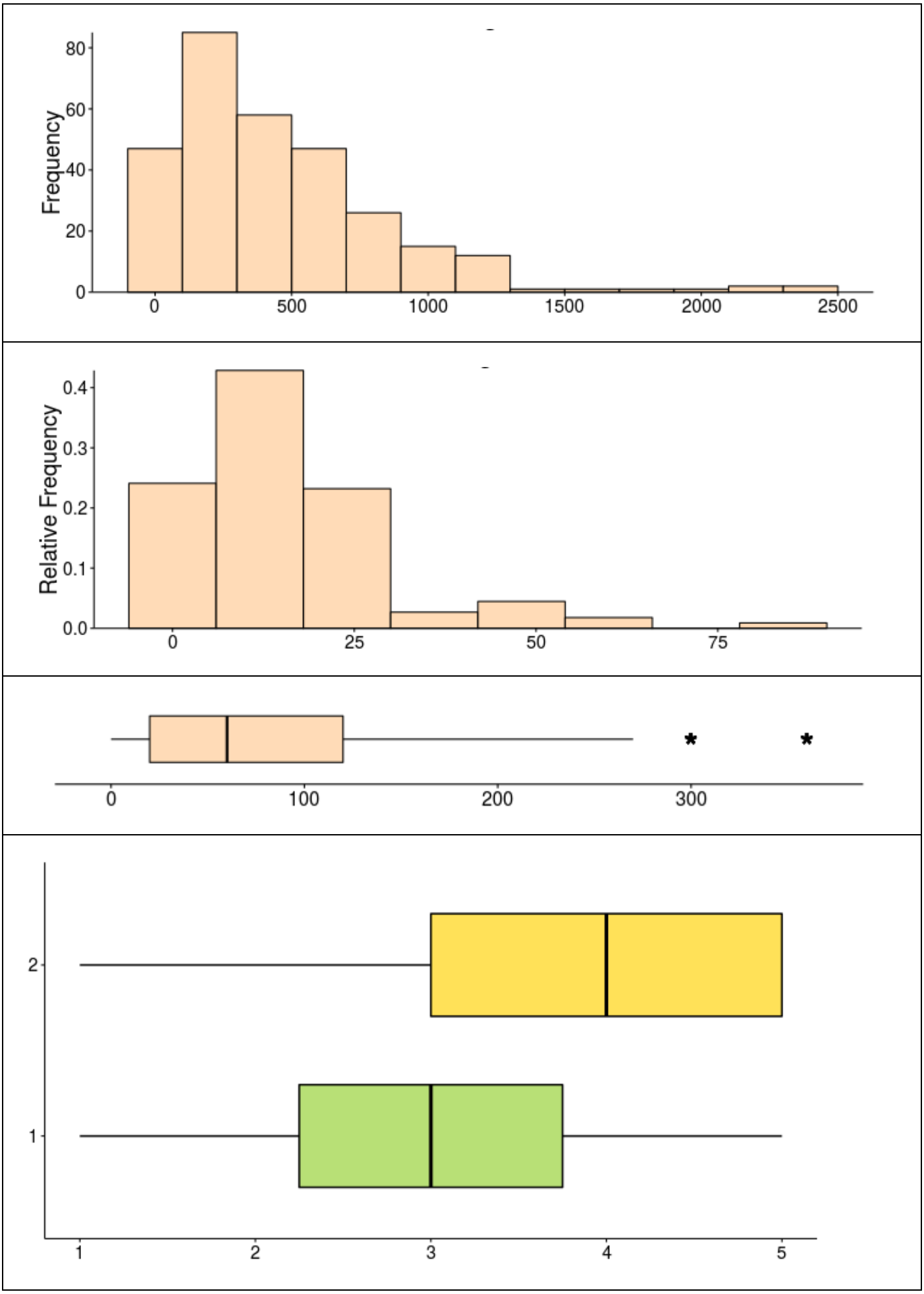


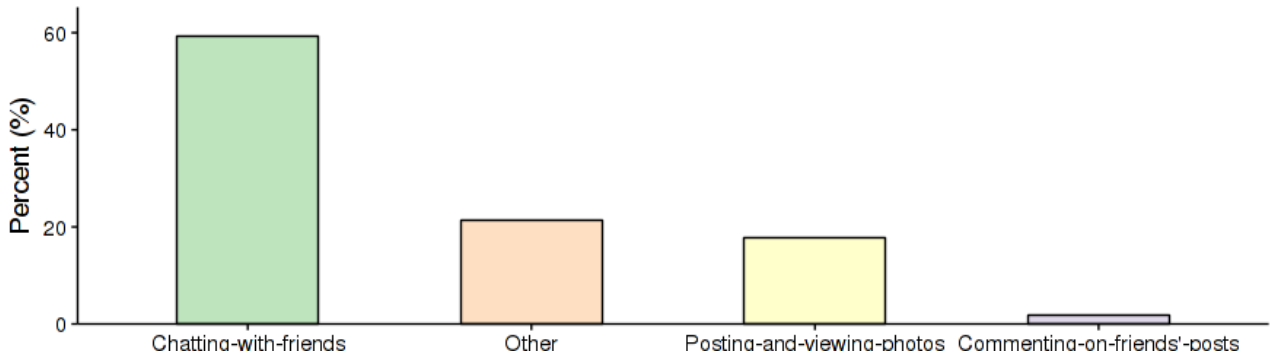

```
1 | 2: represents 12
 leaf unit: 1
            n: 300
   32     0 | 00000000000000000000000000015555
   41     1 | 000055555
   50     2 | 000000000
   86     3 | 00000000000000000000000000000000
   95     4 | 000555555
   96     5 | 0
  (76)    6 | 0000000000000000000000000000000000000000000000000000000000000000000000000
  128     7 | 55
  126     8 | 005
  123     9 | 00000000000000000000005
  100    10 | 55
         11 |
   98    12 | 00000000000000000000000000000000000000000000000000000000
         13 |
   40    14 | 0
   39    15 | 000000
   33    16 | 6
         17 |
   32    18 | 00000000000000000000
         19 |
         20 |
   12    21 | 000
         22 |
         23 |
    9    24 | 000000
HI: 300 300 300
```



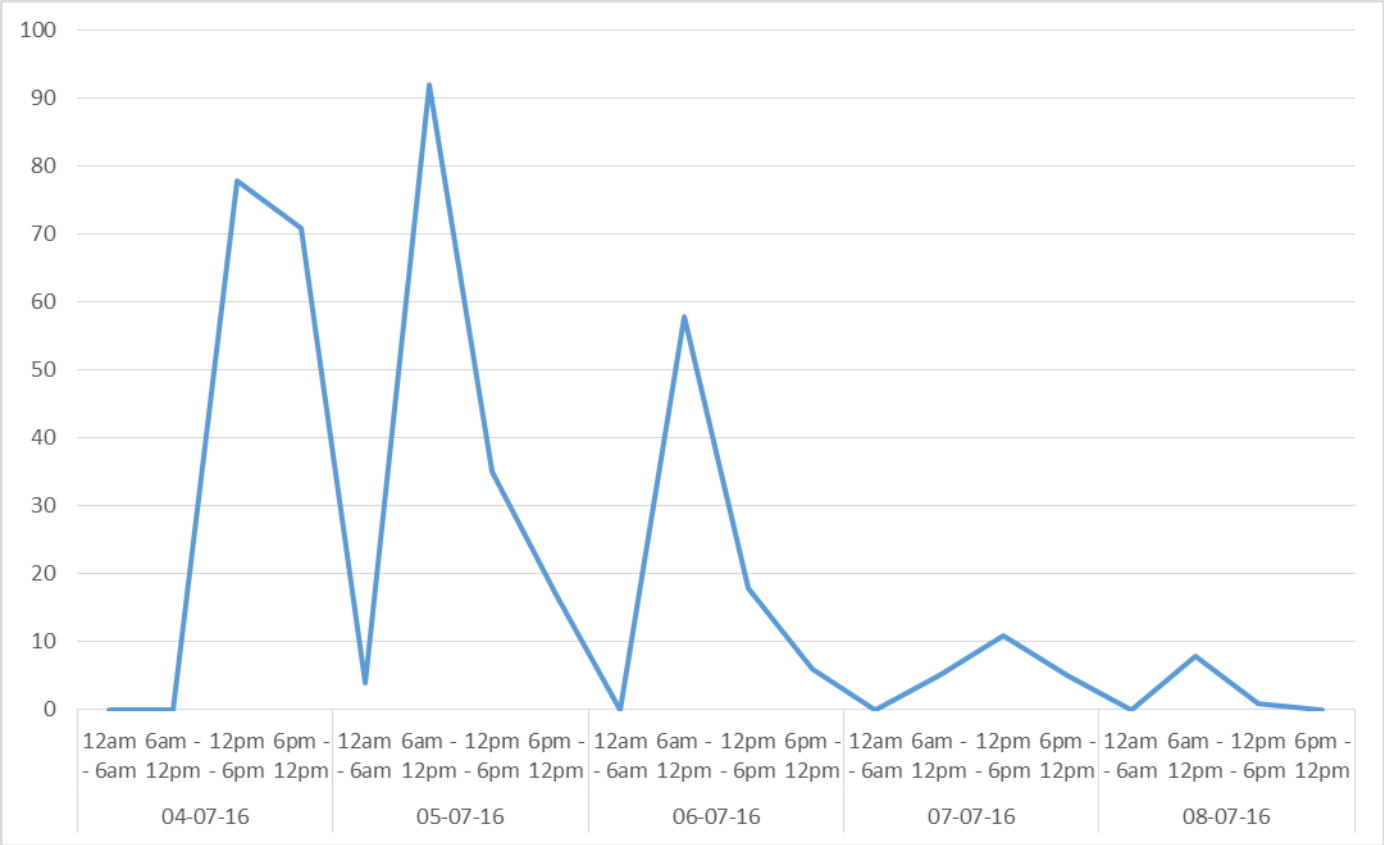


	Frequency	Relative frequency	Cumulative relative frequency
No more than once	51	17.0	17.0
Two to four	113	37.7	54.7
Five or more	136	45.3	100
Total	**300**	**100**	


Frequency
80
60
40
20
0
0
500
1000
1500
2000
2500
Relative Frequency
0.4
0.3
0.2
0.1
0.0
0
25
50
75
0
100
200
300
2
1
1
2
3
4
5



	Frequency	Relative frequency	Cumulative relative frequency
0 - 20	290	0.967	0.967
20 - 40	5	0.017	0.984
40 - 60	3	0.010	0.994
80 - 100	1	0.003	0.997
100 - 200	1	0.003	1
Total	**300**	**1**	



Websites referred to in webinar

A small selection of Pip Arnold resources on Census At School (NZ):

Engaging with shape http://new.censusatschool.org.nz/resource/engaging-with-shape/

Using data cards http://new.censusatschool.org.nz/resource/using-data-cards/

Growing scatterplots http://new.censusatschool.org.nz/resource/growing-scatterplots/

Stats cats data card survey: https://goo.gl/forms/7Uy6P5fZey7iqAwA3

Census At School (US): http://ww2.amstat.org/censusatschool/index.cfm

GAISE report: http://www.amstat.org/asa/files/pdfs/GAISE/GAISEPreK-12_Full.pdf

Statistics teacher: http://www.statisticsteacher.org/

Digital data cards: http://learning.statistics-is-awesome.org/stickland

Interactive dot plot: http://learning.statistics-is-awesome.org/dots

Models of teacher preparation designed around the GAISE framework (Christine Franklin and Gary Kader): https://www.stat.auckland.ac.nz/~iase/publications/icots8/ICOTS8_3E3_FRANKLIN.pdf

Exploring students' conception of standard deviation (Robert delMas and Yan Liu): http://iase-web.org/documents/SERJ/SERJ4(1)_delMas_Liu.pdf