Numeric Variable: A quantitative variable that takes numerical values for which it makes sense to find an average. These variables can be either continuous or discrete

Part 1: Game of Greed

Rules:

Everyone stands. Someone throws a die twice and totals the numbers. This is everybody's current score. Those that are happy with that score sit down; they have finished this round. They record their score.

For the others, the die is rolled again. Those still standing get to add the number to their total, UNLESS it is a 2. If it is a 2, the game is over and all those standing receive 0 for that round.

Keep throwing the die until a 2 comes up or everyone has sat down and recorded their score for that round. A game consists of 5 rounds.

At the end of the game, the students add their 5 scores to get their total.

Individual Results:

Round 1	Round 2	Round 3	Round 4	Round 5

My total score is: _____

Class Results:

Part 2: Stemplots

A simple graphical display for fairly small data sets of a quantitative variable is a stemplot (also called a stem-and-leaf plot). We made a stemplot to display the scores for the game of greed played at the beginning of this lesson.

Rules for making a stemplot:

- Each number is separated into...
    - Stem: consists of all but the final digit
    - Leaf: the final digit
- Write the stems in a vertical column with the smallest at the top.
    - Do NOT skip any stems, even is there is not data value for a particular stem.
- Draw a vertical line at the right of this column
- Write each leaf, in increasing order, in the row to the right of its stem

Example 1: The points for the 30 NHL teams from the 2013 regular season are recorded below:

72, 56, 55, 49, 48, 63, 62, 57
56, 48, 57, 51, 42, 40, 36, 77
60, 56, 55, 41, 59, 55, 45, 42
39, 66, 59, 57, 51, 48

Display the data using a stemplot:

Stem	Leaf
3	6 9
4	0 1 2 2 5 8 8 8 9
5	1 1 5 5 5 6 6 6 7 7 7 9 9
6	0 2 3 6
7	2 7

Part 3: Boxplots (5 number summary)

Let's start by watching a video introducing boxplots:

http://www.learner.org/courses/againstallodds/unitpages/unit05.html

While watching the video, fill in answers to the following five questions:

1. What variable is used to compare different brands of hot dogs?

The different brands of hot dogs were compared by their calories

2. What name do we give to the value for which one-quarter of the data values falls at or below it?

Q_1 - The first quartile

3. What numbers make up a five-number summary?

Minimum
Q_1 - first quartile
Q_2 - median
Q_3 - third quartile
Maximum

4. How do you calculate the interquartile range?

$IQR = Q_3 - Q_1$

5. Boxplots show that poultry hot dogs as a group differ from all-beef hot dogs. Compare the distribution of calories between the two types of hot dogs.

The median of the poultry hot dogs is below the minimum for the beef hot dogs. So, half of the brands of poultry hot dogs have fewer calories than the lowest calorie brand of all beef hot dogs.

A boxplot is a <u>five number summary</u> that shows the distribution of a set of quantitative data. The five numbers a boxplot displays are:

Min: smallest data value
Q_1: median of the lower half of the data set (median of data to the left of the median)
Median (Q_2): median of the data set
Q_3: median of the upper half of the data set (median of the data to the right of the median)
Max: largest data value

Note: The median of a set of data is the 'middle most' piece of data. If there are an even number of data points in a set, the median is the <u>average</u> of the two middle most pieces of data.

In a boxplot, the box contains the <u>median</u> of the data and its width represents the <u>middle half</u> of the data.

The upper and lower limits for the box are found by finding the <u>median</u> for the upper and lower half of the data set. The median value itself is not included in the lower or upper half.
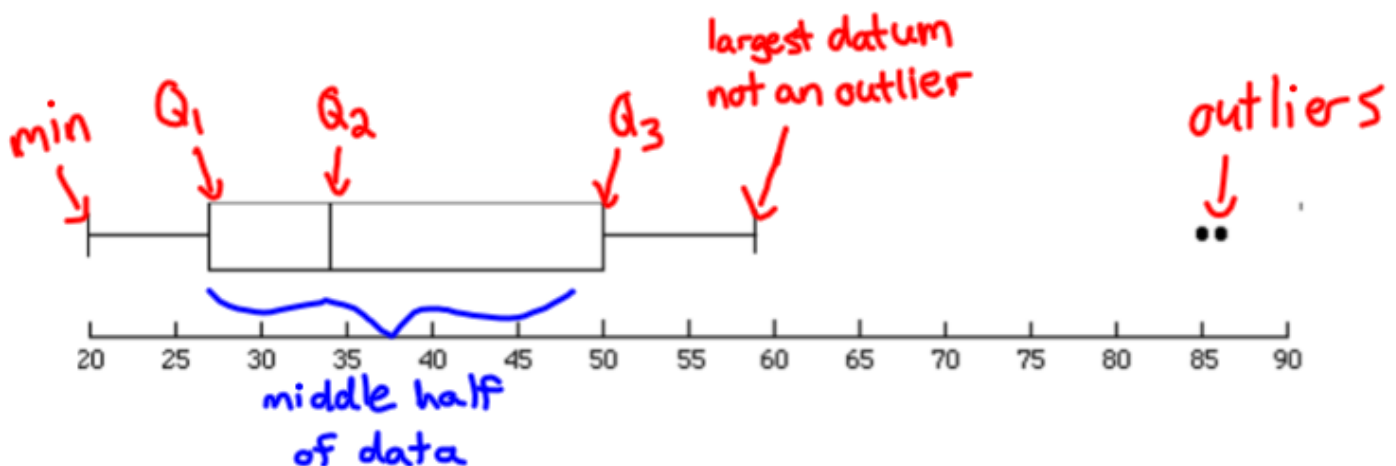
From the sides of the box, horizontal lines are drawn extending to the <u>minimum</u> and <u>maximum</u> data points that are NOT outliers.

<u>Threshold</u> values are used to determine which pieces of data are <u>outliers</u>.

Lower threshold = $Q_1 - 1.5 \times IQR$
Upper threshold = $Q_3 + 1.5 \times IQR$

Note: IQR stands for interquartile range is = $Q_3 - Q_1$

Steps to drawing a boxplot:

- Make sure the data points are in order from smallest to largest
- Find the 5-number summary (1-Var Stats), identify outliers.
- Draw a scale (number line) above which plot will be drawn—include numbers and units: can be vertical or horizontal.
- Draw rectangular box with ends at Q_1 and Q_3.
- Draw line through box at median (Q_2).
- Draw two "whiskers" from corresponding ends of box to most extreme data value that is <u>not</u> an outlier—inside thresholds. Put dots or other marks for each outlier value.

Note: Don't draw thresholds on boxplot—they are not data values. Only use them to identify outliers.

Example 2: A random survey of people at a golf course asked them how many times they had seen Happy Gilmore. The results are shown below in ascending order

$$1, 2, 2, 3, 3, 3, 4, 4, 5, 5, 6, 7, 8, 9, 10, 10, 12, 15, 26$$

a) Find the five number summary of the data (can use 1-var stats on calculator)

min = 1 Q_1 = 3   median (Q_2) = 5 Q_3 = 10   max = 26
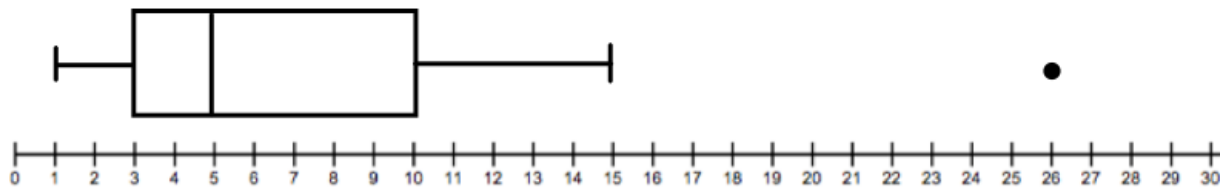
b) Identify outliers

IQR = $Q_3 - Q_1 = 7$ 1.5 * IQR = 10.5

lower threshold = $Q_1 - 1.5 \times IQR = 3 - 10.5 = -7.5$

upper threshold = $Q_3 + 1.5 \times IQR = 10 + 10.5 = 20.5$

Therefore, 26 is an outlier

c) Create a boxplot of the data

Number of Times Golfers Watched Happy Gilmore

Example 3:

The times, in minutes, it took ten police officers to complete routine paperwork after their shift are given below

$$10, 32, 36, 38, 41, 43, 44, 48, 80, 89$$

Find the five-number summary all ten officers. Draw a full (modified) boxplot, showing how you identified any outliers.

min = 10 Q_1 = 36   med = 42 Q_3 = 48   max = 89

IQR = 48 – 36 = 12   1.5 x IQR = 1.5 x 12 = 18

Thresholds: 36 – 18 = 18   any value below 18 is an outlier
(whisker drawn to smallest non-outlier—here value was 34)

48 + 18 = 66   any value above 66 is an outlier
(no upper whisker—largest non-outlier is at Q_3)
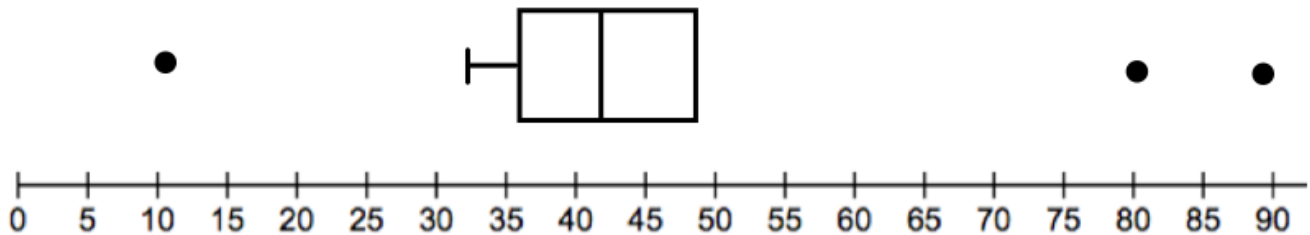
Using Ti 83/84 for Boxplots

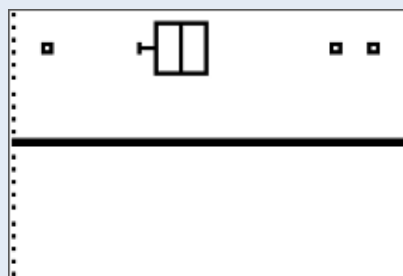Below shows how the previous example can be completed using the ti-83/84 calculator.

- input data in to list L1: STAT → EDIT
- Determine values for 5 number summary: STAT → CALC → 1-VARSTATS → List: L1 → CALCULATE

- turn on statplot: 2nd → y= → ENTER → ENTER
- select modified boxplot
- view graph: GRAPH → ZOOM → ZOOMSTAT

Part 4: Histograms

Quantitative variables often take many values. A graph of the distribution is clearer if nearby values are grouped together. A <u>histogram</u> is a frequency distribution where the horizontal access is divided into equal class <u>intervals</u> in to which data have been divided. The heights of the rectangles (**that have no spaces between them**) represent the frequencies associated with the corresponding intervals. It is important that each interval have the same width. Histograms are most appropriate for <u>continuous</u> variables but you will see them for <u>discrete</u> variables as well.

Example 4:

a) Is the following graph a bar graph or a histogram? How do you know?

It is a histogram because there are no spaces between the bars

Height of High School Students
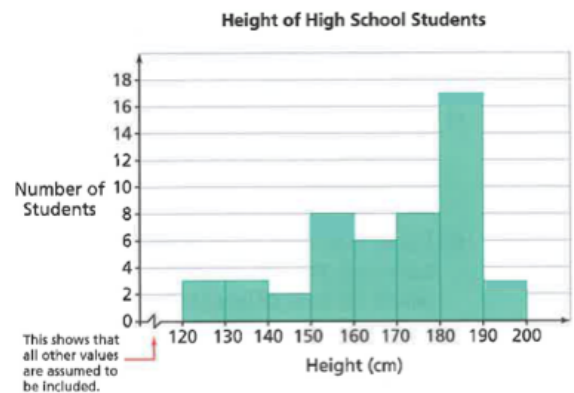
b) Which height interval has the highest frequency? What is the frequency?

The interval between 180 and 190 has the highest frequency. The frequency is 17. This means there are 19 students who have a height between 180 and 190 cm.

Steps for making a histogram:

1. Choose the number of intervals (if the question doesn't specify, choose between 5 and 10)

2. Calculate the range of your data (largest data point – smallest data point)

3. Round your range UP to a number that is easily divided by the number of intervals you chose.

4. Calculate your bin width; $\boldsymbol{bin\ width} = \dfrac{range}{number\ of\ intervals}$

5. Determine the first value for your first interval; $\boldsymbol{lowest\ value} - \dfrac{rounded\ range - actual\ range}{2}$

6. If any data points fall on the border of any of the intervals, add a decimal place to ensure that this doesn't happen.

7. Make a frequency table using the intervals you have determined.

8. Draw the histogram (no spaces between bars)

Example 5:

Here are a class' scores obtained on a data management exam:

78, 81, 55, 60, 65, 86, 44, 90

77, 71, 62, 41, 80, 72, 70, 64

88, 73, 61, 70, 75, 98, 51, 73

59, 68, 65, 81, 78, 67

a) Determine the range of the data

$Range = 98 - 41 = 57$

b) Determine an appropriate bin (interval) width that will divide the data into 6 intervals.

$Range \sim 60$

$Bin\ Width = \dfrac{Rounded\ Range}{\#\ of\ intervals} = \dfrac{60}{6} = 10$

> **Note:**
>
> Round your range UP to a value that can be divided easily.

c) Determine the first value of your first interval

We added <u>3</u> to 57 when we rounded our range, therefore we should subtract $\dfrac{3}{2} = 1.5$ from our smallest value <u>41</u>; which makes our starting point <u>39.5</u>.

Or use formula:

$$initial\ value = 41 - \dfrac{60 - 57}{2} = 39.5$$

> **Note:**
>
> 1. If you have rounded your range up you should subtract half of the amount you rounded from the smallest value to evenly distribute the 'excess of your range'.
>
> 2. Make sure no data points lie on the border of two intervals. (Do this by subtracting .5 from a whole number, .05 from data with one decimal point, .005 from data with two decimal points and so on)
> **d)** Create a frequency table using your intervals

Notice that the number one interval ends with, the next interval starts with the same number. This is because the data for a histogram is continuous!!!

Grade Interval	Frequency
39.5 - 49.5	2
49.5 - 59.5	3
59.5 - 69.5	8
69.5 - 79.5	10
79.5 - 89.5	5
89.5 - 99.5	2

e) Create a histogram of the data
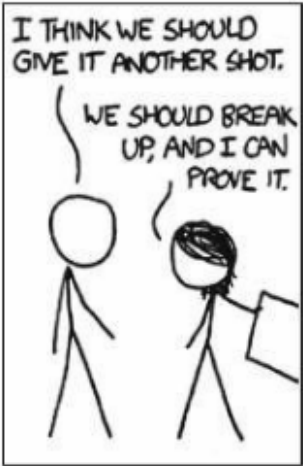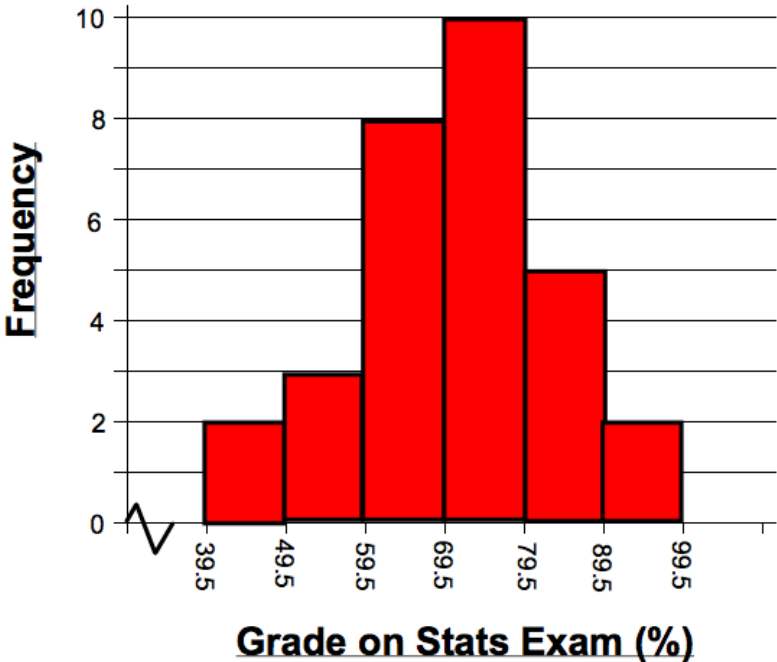
Using Ti 83/84 for Histograms

Below shows how the previous example can be completed using the ti-83/84 calculator.

- input data in to list L1: STAT → EDIT
- turn on statplot: 2nd → y= → ENTER → ENTER
- select histogram
- view graph: GRAPH → ZOOM → ZOOMSTAT

- change class intervals: WINDOW → enter values shown in picture below
- exam class intervals: TRACE → arrow left and right