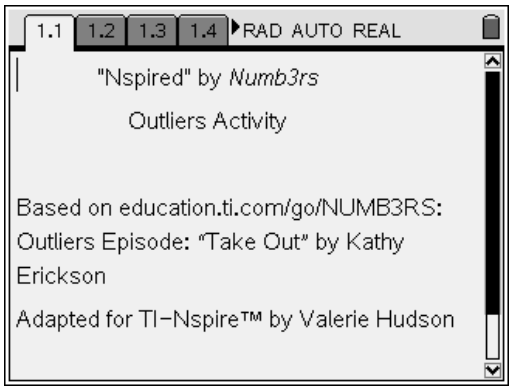

Figure 1

In the *Numb3rs* episode, "Take Out" Charlie investigates various financial transactions using "a target specific outlier model. Something called outlier detection."

In this activity, you will investigate various methods of representing statistical data, as well as identify outliers, or values that are widely separated from the rest of the data as a whole. A data value is determined to be an outlier if it is greater than the upper quartile (Q3) by more than 1.5 times the interquartile range (Q3-Q1), or if it is less than the lower quartile (Q1) by more than 1.5 times the interquartile range.

On page 1.3 of the Outliers doc.tns file (Figure 1), enter the following data in column A, labeled “transfers” which lists the amounts of twenty different wire transfers: $50, $900, $1,000, $1,210, $1,250, $1,300, $1,300, $1,500, $1,500, $1,500, $1,500, $1,850, $2,000, $2,000, $2,100, $2,500, $2,500, $2,800, $3,800, $6,000.

On page 1.4, create a box and whisker plot of the data in the bottom half of the screen. First, click on "Click to add variable" and choose "transfers". Next, choose "Box Plot" in the "Plot Type" menu. Scroll across the graph and notice the various values that appear. (Figure 2)

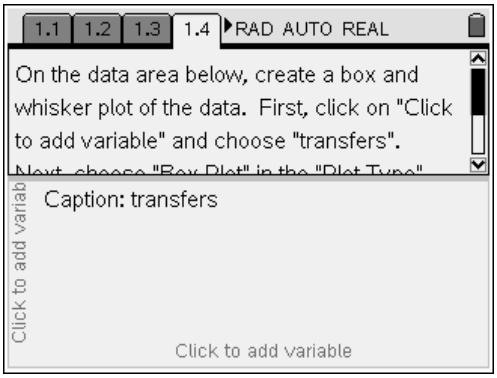

Figure 2

On page 1.5, identify: the minimum, maximum, first quartile, third quartile and median values of the data. Next, find the interquartile range Q3-Q1.

Minimum: _____ Maximum: _____ First Quartile: _____
Third Quartile: _____ Median: _____ Interquartile Range: _____

On page 1.6, find the mean of the data by pressing "Menu," then "Analyze," then "Plot Value". In the text box that appears, type "mean(transfers)". Multiply the interquartile range by 1.5. Is there a point that is greater than the upper quartile (Q3) by more than 1.5 times the interquartile range (Q3-Q1)? _____ Where is it? _____
What is it called? _____ Why is there a difference between the mean and the median? _____

On page 1.7, in column A, labeled “transfers2”, remove the outlier value from the data from page 1.3, and enter the remaining data on the table. The data from page 1.3 is repeated here: $50, $900, $1,000, $1,210, $1,250, $1,300, $1,300, $1,500, $1,500, $1,500, $1,500, $1,850, $2,000, $2,000, $2,100, $2,500, $2,500, $2,800, $3,800, $6,000.

On page 1.9, create a box and whisker plot of the data from "transfers" on the top part of the page, and create a box and whisker plot of the data from "transfers2" on the bottom part of the page. What are the differences between the two graphs?

_____

_____

On page 1.11, recreate a box and whisker plot of the data from "transfers" on the top part of the page, and create a histogram of the same data on the bottom part of the page. For the histogram, click on "Click to add variable" and choose "transfers". Next, choose "Histogram" in the "Plot Type" menu. On page 1.12, recreate a box and whisker plot of the data from "transfers2" on the top part of the page, and create a histogram of the same data on the bottom part of the page. Where is the outlier located on each histogram?

_____

Problem 2: On the list on page 2.2, enter the following information:

Dallas Mavericks 2006-2007 Regular Season Stats

Player	Minutes per Game	Points per Game
Nowitzki	36.2	24.6
Howard	35.1	18.9
Terry	35.1	16.7
Stackhouse	24.1	12
Harris	26	10.2
Dampier	25.2	7.1
George	21.4	6.4
Buckner	18.1	4
Diop	18.3	2.3
Ager	6.7	2.2

From http://www.nba.com/mavericks/stats/2006/index.html

On page 2.4, create a scatterplot of your data which compares the number of minutes each player plays versus the number of points scored. Select the appropriate units for x and y. Select the appropriate "Window Settings" to create an appropriate viewing window.

On page 2.6, select "Stat Calculations" to find the linear regression for this data. Select the appropriate lists for "x" and "y" and save the regression equation to a function location. Note the value of "r" the correlation coefficient. The closer the absolute value of "r" is to 1, the better the graph will match your data. Go back to page 2.2, and eliminate the point that you think is an outlier, then on page 2.6, find another regression equation for the altered data. Note the value of "r" again. Go back to page 2.2, re-enter the previously eliminated point, then eliminate another data point and find the regression equation. Repeat this process until each point has been eliminated. Which point affects the value of "r" the most? _____

Is this the same point that you chose as the outlier? _____