TAKE IT TO THE MAT

A NEWSLETTER ADDRESSING THE FINER POINTS OF MATHEMATICS INSTRUCTION

Southern Nevada Regional Professional Development Program
October 2003 — High School Edition

rpdp.ccsd.net

Welcome back to school! This issue of *Take It to the MAT* begins the fifth year of the Regional Professional Development Program's newsletter on mathematics instruction.

Box-and-whisker plots are a topic in the Data Analysis strand of the Nevada State Standards. Their presence in the recently released *Nevada High School Proficiency Examination Review Guide* suggests that students are ultimately responsible for understanding how to create and interpret them. In fact, the State Standard that includes box-and-whisker plots is classified as "enduring" knowledge.

Box-and-whisker plots (also known as boxplots) are graphs used to analyze the center and spread of data. Their use began in the 1960's after their creation by John Tukey, a leader in exploratory data analysis. In this and subsequent issues of *Take It to the MAT*, we will look at the construction of the box-and-whisker plot, where their real power lies, and misconceptions students have about them.

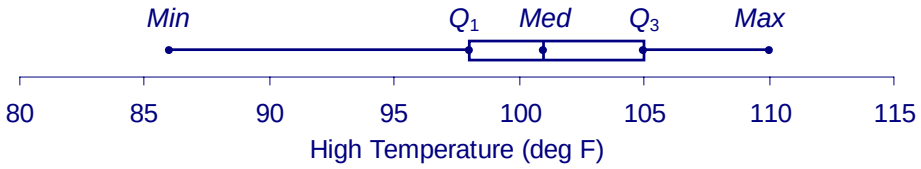
The data table at right shows the official daily high temperatures (in ºF) for Las Vegas, NV in August 2003. The right pair of columns is the raw data by date; the left pair shows the data sorted from lowest to highest temperature.

Las Vegas High Temperatures August 2003

Sorted		Unsorted	
Day	Temp	Day	Temp
1	**86**	1	86
27	92	2	96
16	93	3	101
20	95	4	103
26	95	5	104
2	96	6	103
21	97	7	105
15	**98**	8	107
28	98	9	108
23	100	10	109
24	100	11	110
3	101	12	106
17	101	13	107
22	101	14	105
25	101	15	98
29	**101**	16	93
30	101	17	101
31	101	18	109
19	102	19	102
4	103	20	95
6	103	21	97
5	104	22	101
7	105	23	100
14	**105**	24	100
12	106	25	101
8	107	26	95
13	107	27	92
9	108	28	98
10	109	29	101
18	109	30	101
11	**110**	31	101

The boxplot shows the data broken into four equally sized groups using the five-number summary as the basis for its construction. The five number summary consists of the minimum, first quartile, median, third quartile, and maximum. The median of our data set is the 16th observation, in this case *Med* = 101º. The quartiles, Q_1 and Q_3, split the lower and upper halves of the data in half. They are computed by finding the median of all values above and below the median. In our example, there are 15 data points below the overall median of 101º; the median of those values is Q_1 = 98º. Similarly, Q_3 = 105º. The extremes are *Min* = 86º and *Max* = 110º.

We now plot those values on a number line, draw a box between Q_1 and Q_3, split the box at the median, then extend whiskers from the ends of the box to the extreme points.



Each of the four segments of the graph represents the spread of one-fourth of the data. Any two segments represent one-half. Thus, the hottest quarter of temperatures was between 105º and 110º. The middle half of high temperatures was between 98º and 105º.

What other conclusions can we draw? What questions can we ask? What does the graph really tell us? What does the graph *not* tell us?