The Only Reason to Collect Data is to Take Action!

William Scherkenbach and Mitchell White

While I have been saying this for decades, and while K'ung Fu-tzu implied it millennia ago when he called for a balance of knowledge and action, it takes a while to sink in. Deming showed how simply taking a pencil with paper and plotting the data makes action possible. Tukey made exploring data fun and graphic. Wheeler and Chambers have taken data analysis to the next level. It is a moronically simple concept…but is hardly practiced anywhere.

I have made numerous trips to China where I helped some first and second tier electronic suppliers improve their processes, and thus their products. In fact there is so much to do here that I moved to Taiwan last year. All of the charts that I will show were plotted from tables of data dutifully documented on paper or stored in a computer data base. None of the data were plotted before I liberated them for action with a simple graph.

Example One: Take for instance the data for a heat sink. The supplier provided the data in tabular format. No part was out of specification and so no action was taken. I plotted the data as in Figure 1, showed it to the supplier quality manager, and he immediately saw that there was a pattern to the data; Down… Up… Down… Up… Down… Up…. Then the pattern reverses at Point 16. What causes this special pattern? Different shifts? Different Lines? Different Machines? Different measurement systems? Find out what causes the pattern and take action to fix it. ***Plot the data. Take action.***

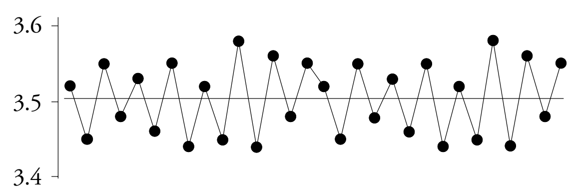

Figure 1: 30 Successive Values for Heat Sink Data

Example Two: Another supplier dutifully recorded the time and the solder temperature for a wave solder operation. The numbers were stored on data sheets and accumulated for weeks and months.

They came to life when I plotted them using boxplots for each of the five different times of day as shown in Figure 2. Nobody knew that there were start-up/warm-up differences at 8:30 compared with the other times. Even if there were no additional clean-up costs after wave solder, the cold solder temperatures could still result in reliability problems later. ***Plot the data and you see what action to take.***

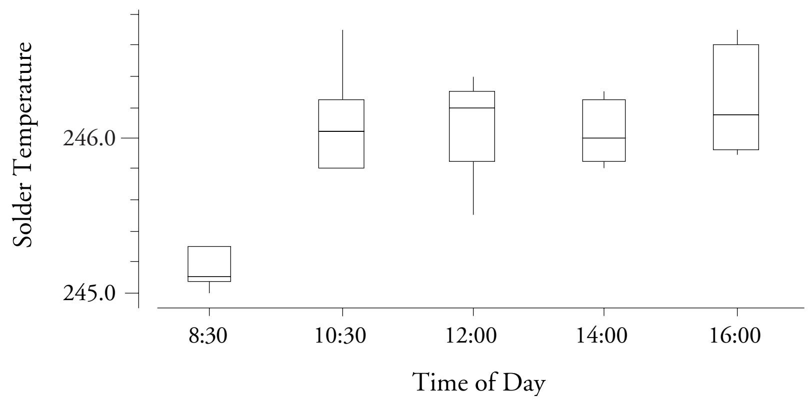

Figure 2: Solder Temperatures versus Time of Day

Example Three: Another supplier had a re-flow solder machine where the centering of the nozzle that delivered the solder paste was a critical process characteristic. Data were collected and recorded on a sheet of paper, and the paper was filed away. No action was taken because no values were out of specification. The supplier was satisfied with the C_{pk} value of 28.

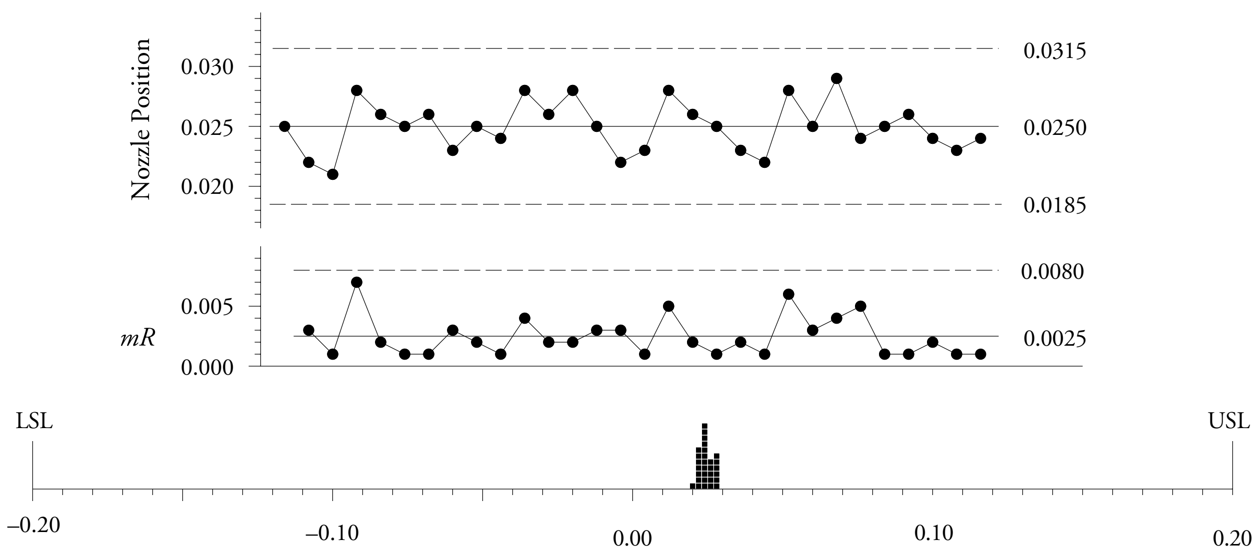

Figure 3: *XmR* Chart and Histogram for Nozzle Centering

As may be seen in Figure 3, this process shows a reasonable degree of predictability and has a very high capability. This process is capable of meeting the next generation of specifications. It does not cost them any more to operate with this capability since they have already own this process. In fact, this capability represents an opportunity for this supplier. They should be pushing for *tighter* specifications in order to create a hurdle for their competitors. ***Plot the data. Take action.***

Example Four: Another supplier recorded C_p and C_{pk} values in their database. Every month they gave these values to managers in tabular form. The managers would look at these tables, their eyes would glaze over, and no action would be taken since most of the values for C_p and C_{pk} were above customer requirements. I used the values from one monthly report to plot the graph in Figure 4 where the diagonal line represents processes that are perfectly centered within the specifications (where $C_{pk} = C_p$). Points to the left of the diagonal line represent processes with centering opportunities. (Centering a process would move its dot horizontally to the right in Figure 4.) Since centering a process is generally easier than reducing the process variation, Figure 4 identifies some of the opportunities for easy improvements. The 13 processes with the smaller C_{pk} values circled in Figure 4 are special candidates for centering efforts. A management dashboard that displayed data in this way would get immediate action to improve these processes. ***Plot the data. Take action.***

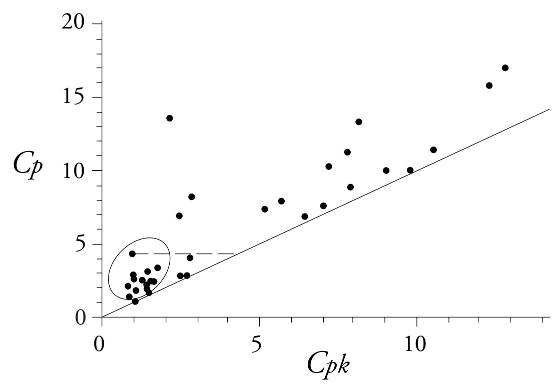

Figure 4: Scatterplot of C_{pk} versus C_p

Example Five: Another supplier recorded an injection molding machine stack drying temperature every four hours and filed the data collection sheets away in a file cabinet. Since no reading was out of specification, no action was taken. (Although 22 out of 126 values below are right at the upper specification limit which is always suspicious, this did not show up on the data collection sheets.) I plotted 21 days of data with an *XmR* chart. The *X* Chart is shown in Figure 5. With six signals of unexplained process changes in Days 1 through 9, it should have been no surprise when the temperatures shifted up by 10 degrees on Day 10.

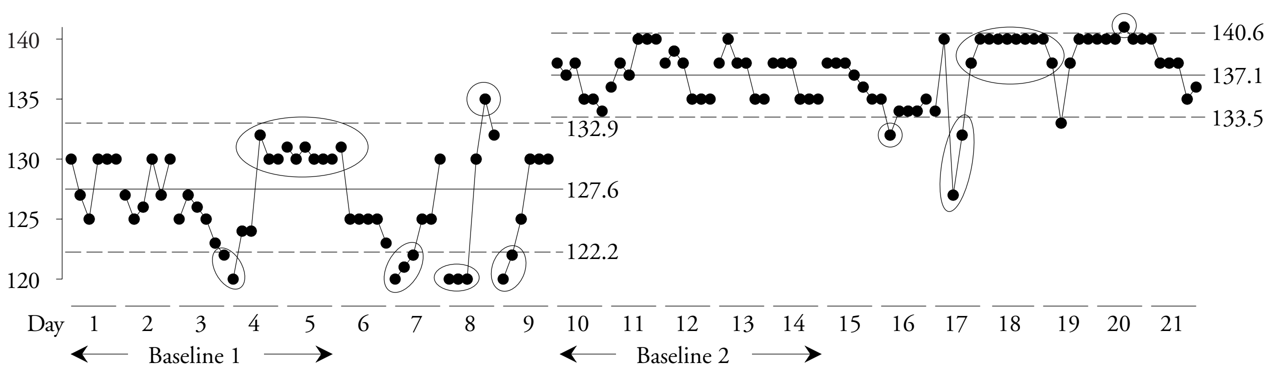

Figure 6: *X* Chart of Stack Drying Temperatures

Even though each of the values in Figure 6 had been written down by hand, no one knew about the changes in the stack drying temperature. No one could say whether a higher temperature would result in better parts. Everyone agreed that the higher temperature resulted in higher costs. ***Plot the data. Take action.***

Example Six: Another supplier measured the verticality of a shaft. Their sampling plan called for a subgroup size of 5. (No real reasons other than there were 5 spaces for readings in the template.) While the supplier did manually plot the data in a control chart, they did not spot the pattern shown by the boxplots in Figure 6. There we see that successive readings in each subgroup are lower than the preceding readings. Something is happening, but they didn't know to look for this until the data were plotted as shown in Figure 6. ***Plot the data. Take action.***

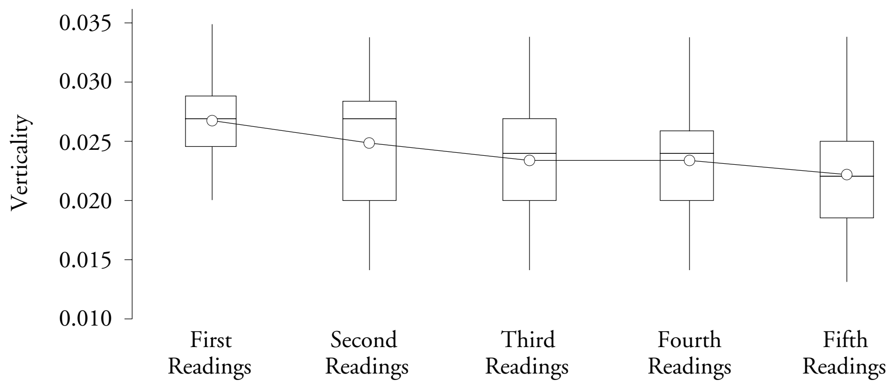

Figure 6: Boxplots of Verticality versus Position in Subgroups

Starting from the top down, I look for differences between suppliers, plants, shifts, lines, machines, spindles, operators, etc. The key to creating useful graphs that reveal the structures within the data is to use the context for the data as a guide. Think about the kinds of differences that could occur, and then draw a graph that will let you see if those differences are present in the data. It is this act of thinking about the data within their context that leads to useful graphs and the ability to take action.

Example Seven: Another supplier collected reflow peak temperature data by shift, but filed the data away. The boxplots in Figure 7 show an actionable difference between shifts. If the night shift values are correct, the day shift has been tampering with the process. If the day shift has not been overadjusting the process, then the night shift has been faking the data. Either way, something needs to be fixed here. ***Plot the data. Take action.***

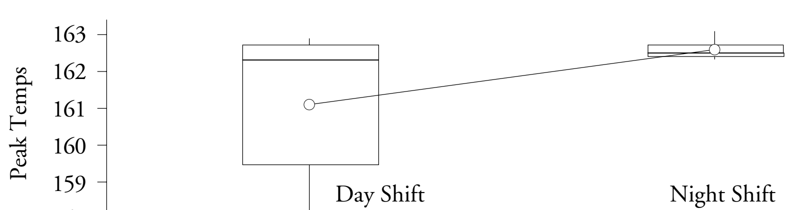

Figure 7: Boxplots of Peak Reflow Temperatures by Shift

Example Eight: Another supplier calculated a C_p of 16.8 and a C_{pk} of 2.0 for a final test voltage. Since both were above 1.33 no action was taken. However, as noted earlier, this discrepancy between the capability ratios represents an opportunity for centering the process at the target value of 5V. The XmR Chart for 50 final test values in Figure 8 shows a process that is subject to upsets. The histogram shows how the process is off target.

One popular piece of software tested these data for normality. Such a test is complete nonsense. According to Walter Shewhart, we are not concerned with whether the histogram might or might not be normal, but merely with the assumption that the data might be homogeneous enough to be characterized by a single probability model. Whenever the process is operated unpredictably, there is not one probability model, but many. Hence, testing the data for normality prior to having a predictable process is merely a triumph of computation over common sense.

Here there is a real need to determine what is causing the process upsets. Otherwise, in spite of having a C_p value of 16.8, the next upset might take this process outside the specifications. In the meantime, centering this process would buy some insurance that the product will continue to meet specifications in spite of the process upsets. ***Plot the data. Take appropriate action.***

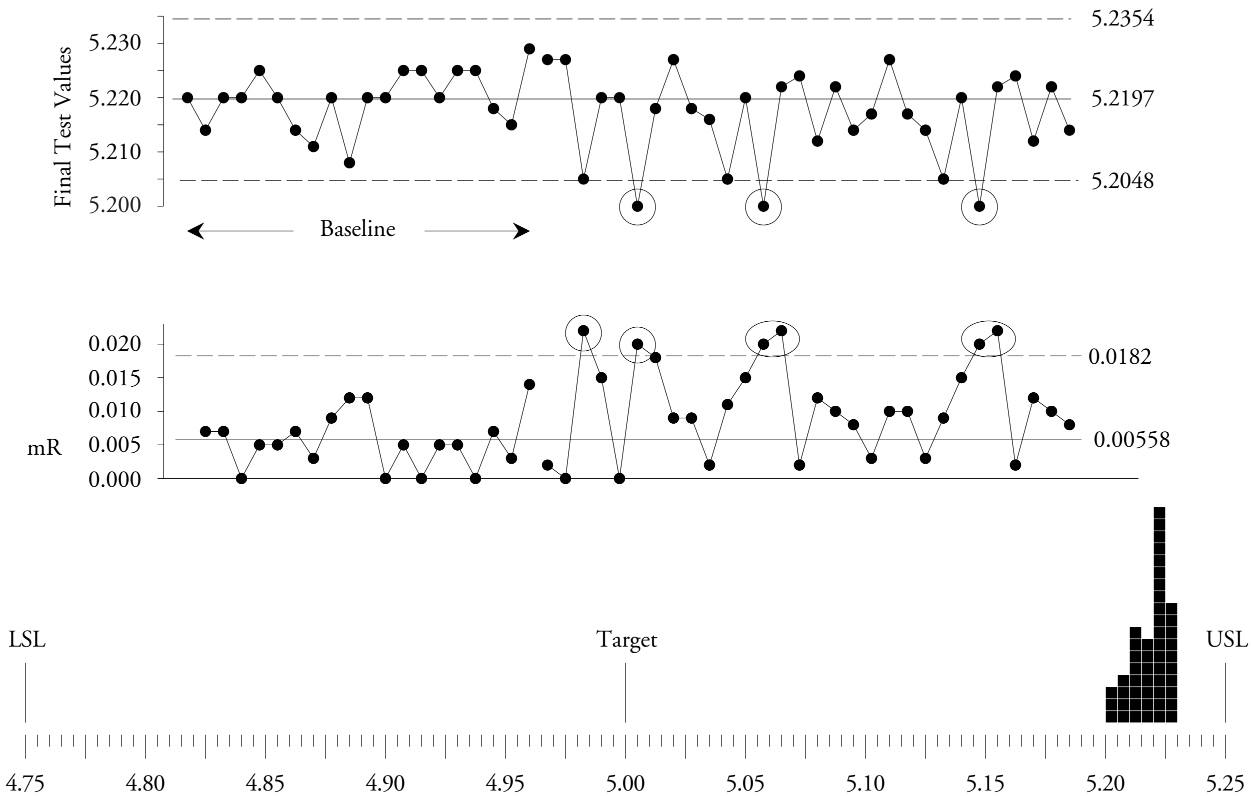

Figure 8: ***XmR*** **Chart and Histogram for Final Test Values**

Example Nine: Another supplier collected data every two hours to manage surface run out. They stored the data in a file cabinet—one sheet of paper for each day. Until the data were plotted, they had no idea that the process was getting worse month by month by month, as may be seen in Figure 9. A little bit here, a little bit there; a gradual drift towards destroying their reputation. ***Plot the data. Take action.***

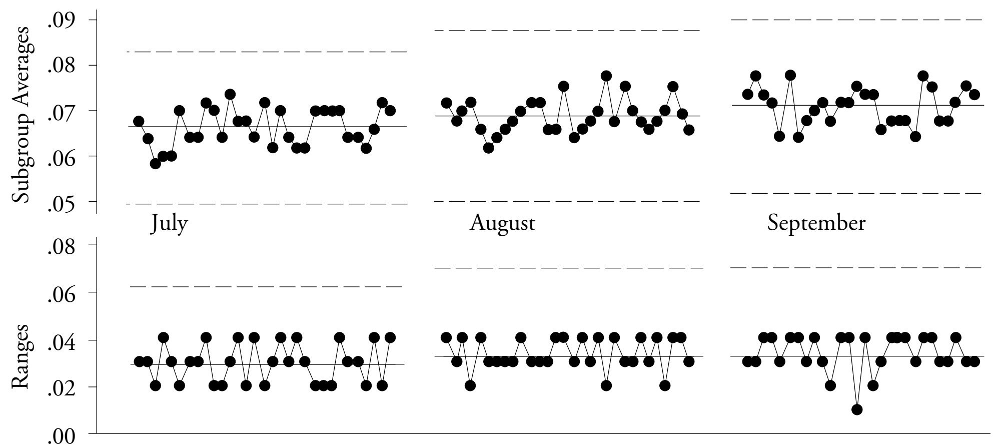

Figure 9: Average and Range Charts for Surface Run-Out

Example Ten: Another supplier sampled solder paste thicknesses at five locations on a printed circuit board (PCB). They used the five values from a single PCB as a subgroup and created an Average and Range Chart. This time they plotted the data but didn't take the correct action, because they failed to correctly diagnose the Average and Range Chart shown in Figure 10.

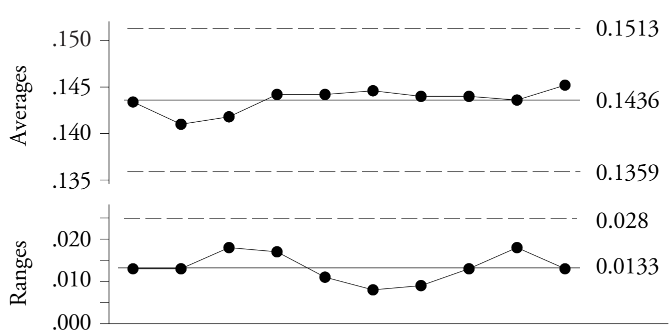

Figure 10: Average and Range Charts for Solder Paste Thickness

When they averaged the solder paste thickness for the five sites on the PCB they masked the individual site thicknesses. The Average and Range Chart shows the data hugging the central line, which is an indication of possible stratification. Stratification occurs when each subgroup combines values that are systematically different, and the high values average out with the low values, making both the Ranges and Averages "hug" their central lines. While Figure 10 looks like the process is in control, it is not a true representation of the process.

If we look at the five individual sites on the PCBs in Figure 11 we see that Position 5 has consistently higher thickness values than the other four locations. With this understanding of the Solder Paste Thicknesses you can adjust specific nozzles and pressures to eliminate these differences. The problem with the Average and Range Chart in Figure 10 is inappropriate subgrouping. Whenever you subgroup two values together you are making a judgment that those two values were collected under essentially the same conditions. When unlike things are subgrouped together the Average and Range Chart will fail to deliver the message contained within the data. ***Plot the data properly. Take action.***

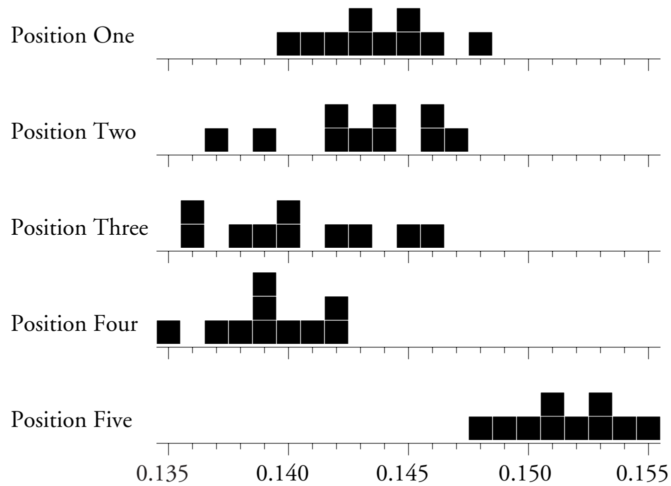

Figure 11: Histograms for Solder Paste Thickness for Each of the Five Positions

Example Eleven: I asked another supplier about their final functional test performance. They had two test stations (FT1 and FT2) which prompted an automatic question: "How different are these two machines?" A just about automatic answer was: "They're both the same manufacturer and model; no difference."

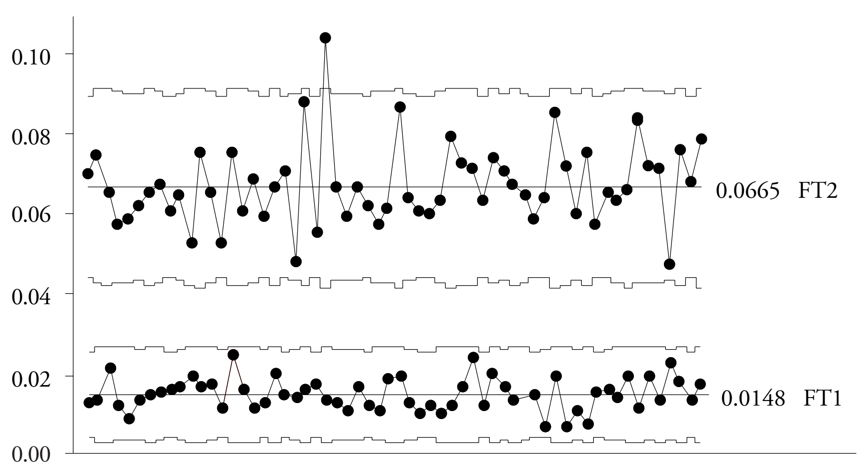

Figure 12: *p*-Charts for Proportion Rejected by FT1 and FT2 During the Same 60 Day Period

I asked how did they know they were both the same and could they show me the data. While finding the person who knew how to get the data, we looked at the logs for both machines which indicated that one had a reject rate of about 1.5% while the other had a reject rate of about 6.5%. They got me the number of rejects per day for about 60 days. Looking at the data over time in a p-chart you can see that the two machines are doing completely different jobs. I asked if they purposely routed the difficult or borderline cases to FT2 and they said no. So now they need to determine if there is a systematic difference in the two production lines or if this difference is merely a difference in the two testing machines. ***Plot the data. Take action.***

I see this everywhere I go. Or more accurately, I don't see the data plotted anywhere I go. The only reason to collect data is to take action, and the key to using data to take action is to create the graphs! *Always plot the data over time with control charts, and always plot the data with time collapsed with either a histogram or a boxplot.* ***Look at the data through the lens of theory and context! Plot the data! Take action!***

I didn't point out all the opportunities that the data show in these graphs. Can you find five special patterns in the data that were not discussed?

William W. Scherkenbach is a former student of Dr. W. Edwards Deming and the author of two popular books on using Dr. Deming's approach to improve quality, productivity, and competitive position. You may contact him at <bill@scherkenbach.com>.