Effective Reporting – Make Sure You Tell the Story

Jim Horne

All reporting is not created equal. It is not enough simply to publish data, you must make sure what you publish has meaning. Newspapers have been doing this for centuries. The concepts of Who, What, When, Where, Why and How are as valid in technical and management reporting as they are in journalism. This paper will explore using those ideas to help make your reporting as effective as possible.

Introduction

Management and technical reporting is often not as effective as it could be. The meaning is unclear or unknown and a presentation of raw data is often seen as being all that is needed – by the person presenting the data. But that is not enough. The data you publish should tell a story and convey meaningful information to the reader. It may not always be a story you like or agree with but the story should always be present. In some cases a simple presentation of raw data may be enough to tell a meaningful story but usually more work is needed so the reader can understand what you have to say, and understand it quickly and easily.

Newspapers have been doing this for many years. Their guidelines of Who, What, Where, When, Why and How can serve as a model for management and technical reporting. Every report may not answer every question but all of these questions should be a consideration when writing reports and graphs. If the written portion of the report or graphs will be given to the audience in the form of a presentation, the presenter should have answers to Why and How because someone in the audience will ask those questions.

Even canned or automatic reports should answer at least some of these questions. If you simply show a table of CPU use by time for a system by time you are presenting information but are you telling a story? If you graph the data, making it easy for your audience to see when CPU use was high or low, you convey the same information in a much more meaningful context. They can easily see the relationship between time of day and use. If the day was important to the business or some reason, they will know that and that will layer more meaning into what they see.

The rest of this paper will concentrate on making sure graphs you use tell a story. We will not be covering specifics of how to write good graphs and we will not be covering report writing in general. Many papers have been written on graph techniques to produce good graphs rather than bad graphs and many papers have been written on what constitutes good writing. What has not been covered in detail is making sure you tell a story, that your graphs and reporting convey meaning and not just information to your target audience. By their very nature, pictures are more powerful than words and that is where this paper will focus. We will do this by studying examples and how effective they are. The graphs we will look at do not represent the best way to do things. Many of them have been modified to only show a subset of relevant data or the way it is presented. They show ways we want to use the tools at our command but not necessarily the best ways to use those tools because we want to concentrate on what the tools are instead of making the perfect graph to illustrate a point. With that understanding, the tools we will use to make sure we tell the story are the journalism tools of Who, What, Where, When, Why and How.

Who

In journalism, this is often expressed as, “Who is involved?” In technical and especially management reporting we can think of this as, “Who is our target audience?” If your target audience is management, especially upper management, drowning them in data they do not need is a mistake you want to avoid. On the other hand, if you are presenting to a technical audience, they may want and need that same high quantity of data in order to answer the key questions of Why and How for themselves.

Figure 1 is an example of a graph that can only be considered effective in the context of its target audience. If you are a mainframe systems programmer wanting summary information telling you how busy each of the 256 possible I/O channels is on one of your systems, you can have on a system you will find the information useful and informative. If you are almost anyone else you may well ask, “What is that mess?” In fact, when the author showed this graph to the systems programmer who had asked if there was an easy way to identify high use

channels he did not have a single question, he simply started addressing the issues he saw in the graph. When our manager asked what it was, we took about five minutes to explain it to him after which he left us alone. If he had been the target audience the graph would have to be judged a failure. As it was, the graph was a success because it told a meaningful story to the intended audience.

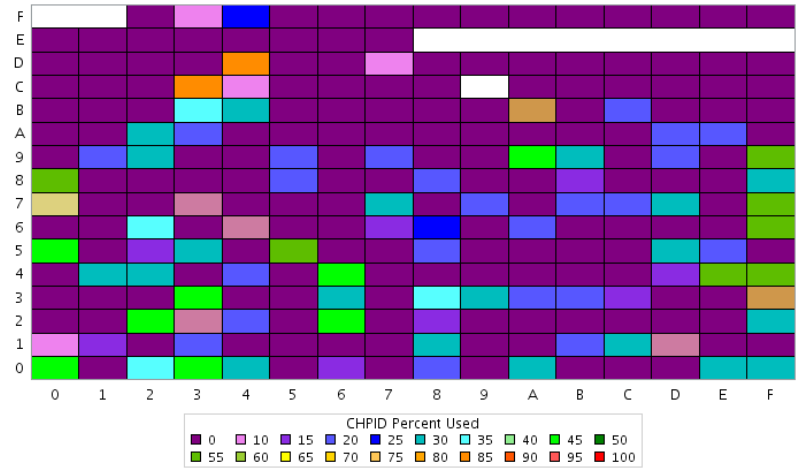

Figure 1. Mainframe Channel Use for a Single System

What

"What happened?" is such an obvious question that you might think answering it is automatic. Any display of data can and probably does answer the question to some degree but how effectively does it answer it? In 2012, Forbes magazine conducted a bad graph contest. They were looking for the worst graph possible to be drawn with commercially available software they could find. The winning graph, shown in Figure 2, can be drawn in Microsoft Excel and is deceiving in its simplicity. While the graph is charting four numbers the challenge is to understand what those numbers are. The simpler chart in Figure 3 gives the amazingly easy answer. "What happened?" becomes much clearer once we know what the actual data is. The story told in Figure 2 is not quite a lie but is misleading to say the least. The simpler story told in Figure 3 leaves no doubt what the story is. Simplest of all would have been to say A=2, B=4, C=6 and D=8 and not even bother with a graph. But, as an illustration of both a bad graph and the importance of clearly telling the audience what happened – or not, in this case – this graph ranks supreme.

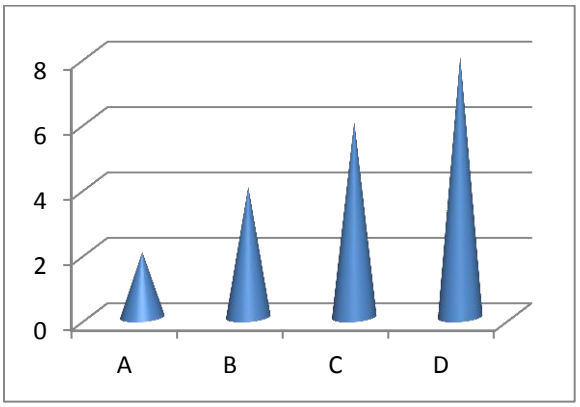

Figure 2. Forbes Magazine Bad Graph Contest Winning Entry

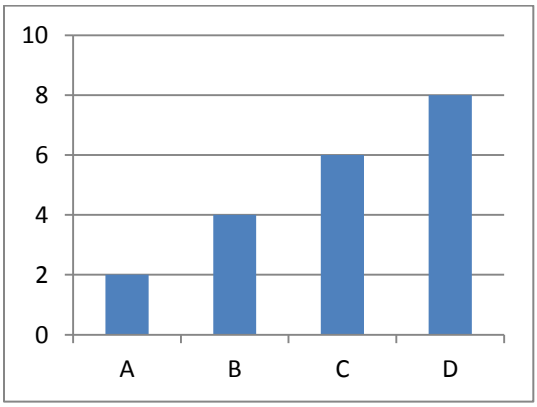

Figure 3. Alternate view of the winning entry

Where

"Where did it take place?" Location is important. If ten murders are committed in Los Angeles in a year it is an unfortunate statistic that directly affects only the people involved; not the rest of the Angelinos. If ten murders happen in a year in French Lick, Indiana (population in 2012: 1,798), it is a tragedy affecting the entire community. In technical reporting similar principles apply. If something happens on a test machine only accessed by one or two people in the company it is not as important as the same thing happening in a major production environment. So it is not enough to know what happened, we also need to know where it happened.

Figure 4 shows CPU use on a number of systems. It shows the consumption but it is not a good example of effectively presenting the information it contains. All it really says is how busy each machine was during each shift of the day. The bars labeled “Capacity” and “Machine” are present for every machine but what they mean is not clear. Once you know that “Capacity is how much work the machine could do and “Machine” is how much work the machine did the chart makes more sense but this is not obvious. This chart tells us nothing that could not be better shown in a table. So while it tells us where the event happened it fails the test of telling us what happened in an effective manner.

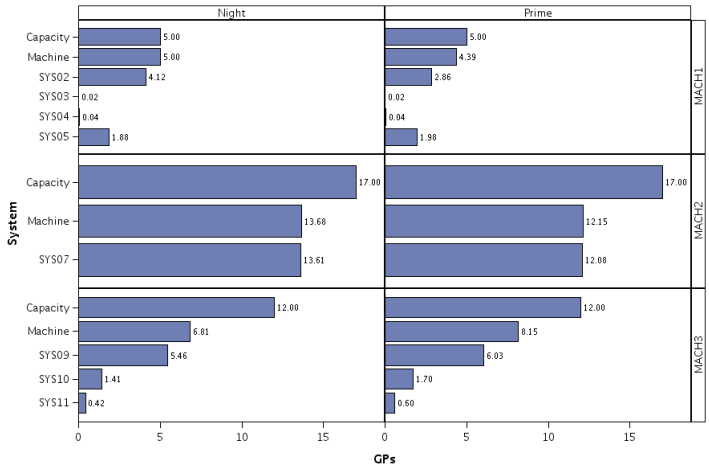

Figure 4. Summary CPU use for a single day by shift.

Contrast that chart with Figure 5. Again, the key metric is CPU consumption but now we see it in the context of how much was available to each system and machine. To compare it to our newspaper analogy, an informed audience can see if something happened in Los Angeles or Muscle Shoals. That enables them to determine what is important to them and use that information more effectively.

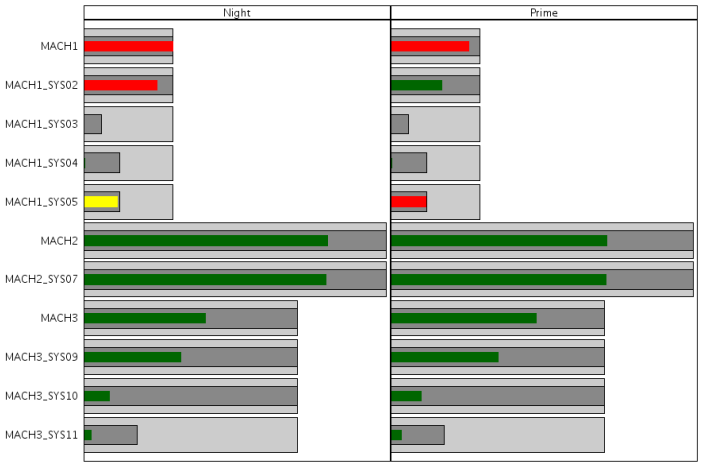

Figure 5. Summary CPU use for a single day by shift in containers.

Another piece of the story we see in Figure 5 that was not available to us in Figure 4 is if we have a potential problem. Color coding the bars in a red-yellow-green traffic light signal pattern draws our eyes immediately to any yellow or red areas. It may be for a system or machine that we don’t care about but our attention is immediately drawn to the issue. Focusing our attention on an area may cause us to reconsider our priorities. While the raw information, CPU use for systems and machines, remains unchanged, our awareness of how it fits in the context of our environment has changed – because of how we chose to present the information.

When

Where something happened gives us part of the story. When something happened adds additional information. In the previous section we saw various machines and systems use and could see one machine in particular that

was interesting. We know it was busy, maybe too busy, but knowing when it was busy would tell us more about whether we have a problem and not just an issue.

Knowing when the incident happened may help determine if it is a problem or not and there are multiple ways to look at that data, some telling the story more effectively than others.

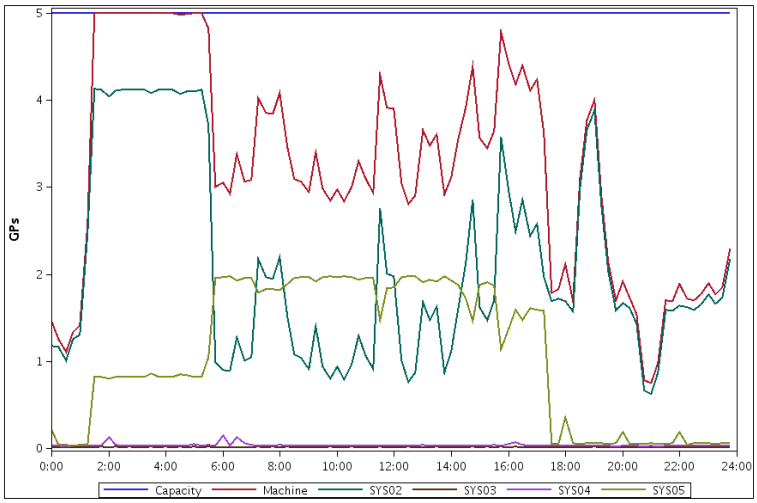

Figure 6. CPU use by all systems on a machine daily detail.

Figure 6 shows us the daily detail for on type of CPU engine on the system and day we want to investigate. It shows us how much work the machine can do, how much work the machine is doing and how much work each system on the machine is doing. It has all of the information we need and even presents it in a logical manner but how clear is it? How easy to understand is it?

Contrast it with Figure 7. Figure 7 takes the same data but stacks the systems on top of each other instead of showing each one relative to zero. While this makes the impact of small use systems on the machine slightly harder to see, it dramatically drives home the story telling how busy the machine is and tells us which system or systems caused it. Both views tell the same story, but which one tells it more effectively? Different people will have different answers to this but, if the story you want to tell is, “How busy was this machine during the day and what systems caused it?” the stacked view tells that story faster and quicker.

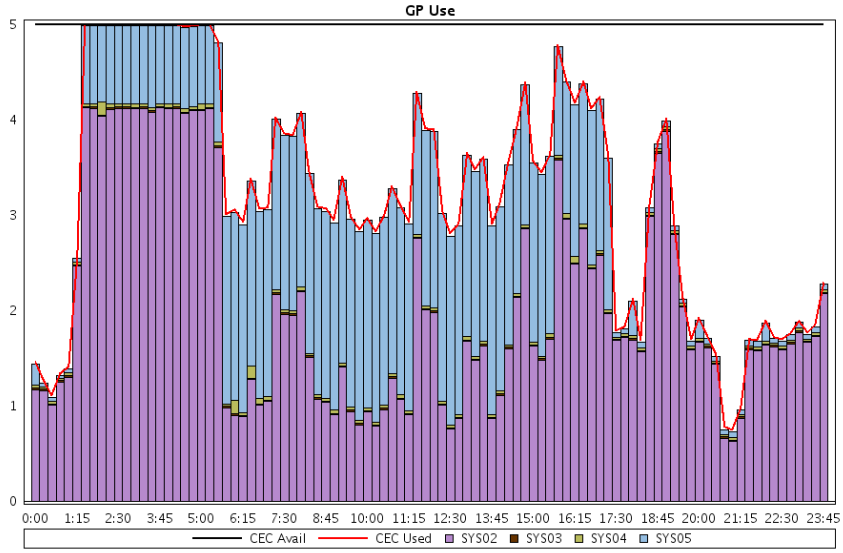

Figure 7. CPU Use by all systems on a single machine daily detail – stacked view.

Why

"Why did it happen?" For routine, automatically generated reports this may not always be an important question, and it never impacts whether or not the report is created. If a question is asked every day, if something looks different or there is a problem somewhere, why it happened becomes of paramount importance. In that case, special reports will be needed. When you create those reports and graphs, "WHY did it happen?" needs to be in your thoughts. If what you produce does not answer that question in your mind it probably will not answer it for anyone else. Even if it does answer it to your satisfaction, will it answer it as clearly for your intended audience?

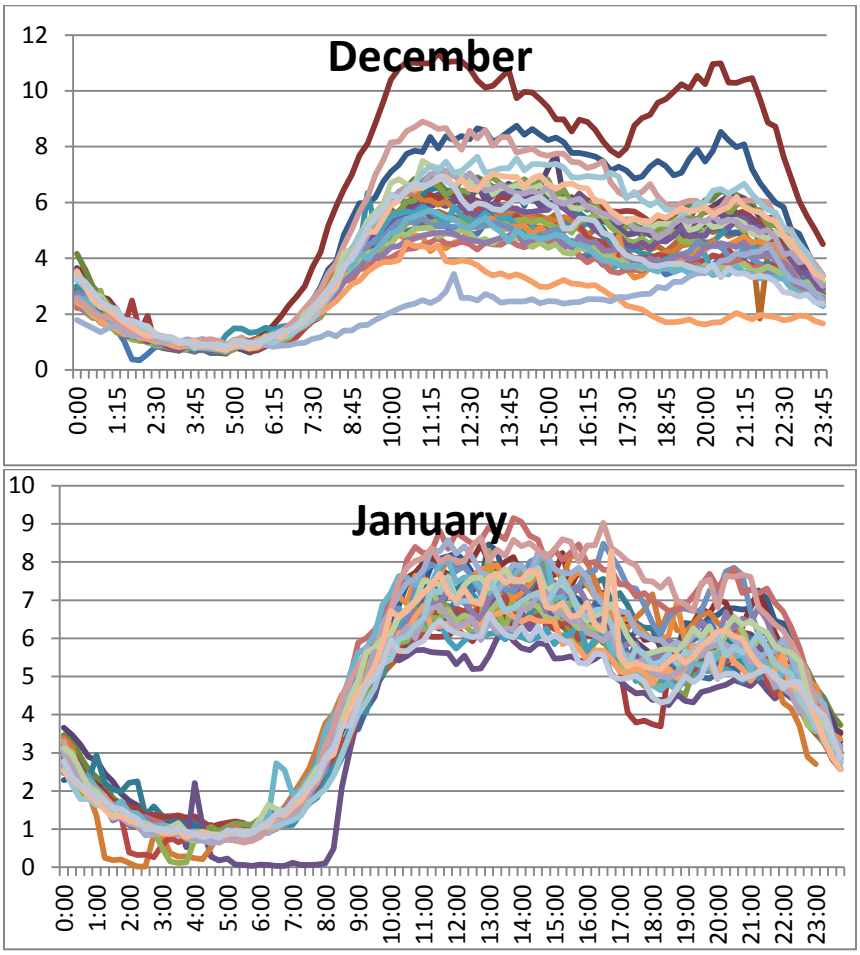

Figure 8. Application Use by Month.

Figure 8 was emailed to the author with the accompanying text, "It looks like there is a 2 processor jump from Dec to Jan. The last code release was June so it wouldn't be from there." Implicit in this statement what the question, "Why did it happen? We know it was not the code."

Looking at the graphs, the fact of a two processor jump is not readily apparent to the casual observer because outliers in December draw the eye away from the data cluster. The graphs are so cluttered with data it is hard to extract any meaning from them unless that meaning is something you are already looking to find. At first view, it was not apparent to the author that the jump was in December as the outliers in December are higher than any line in January.

In an attempt to bring order out of chaos, the author decided to look at one point per day, the 90th percentile peak value for each day, and to graph both months on a single graph. The result is in Figure 9.

Now we see the December values are not always lower than the January values. They start the month higher and finish roughly even. Since there is a drop in the middle of December with the lowest point on the 25th of the month and since use sharply rises immediately after the 25th, could the difference between the two months be related to the Christmas holidays?

Simplifying the presentation made the question of the drop take on new meaning. When the person asking the original question saw this graph, he commented that the relevance of the two months being December and January had not occurred to him. Subsequent analysis showed that December, not January was the odd month. It was lower than normal, due to the Christmas holidays.

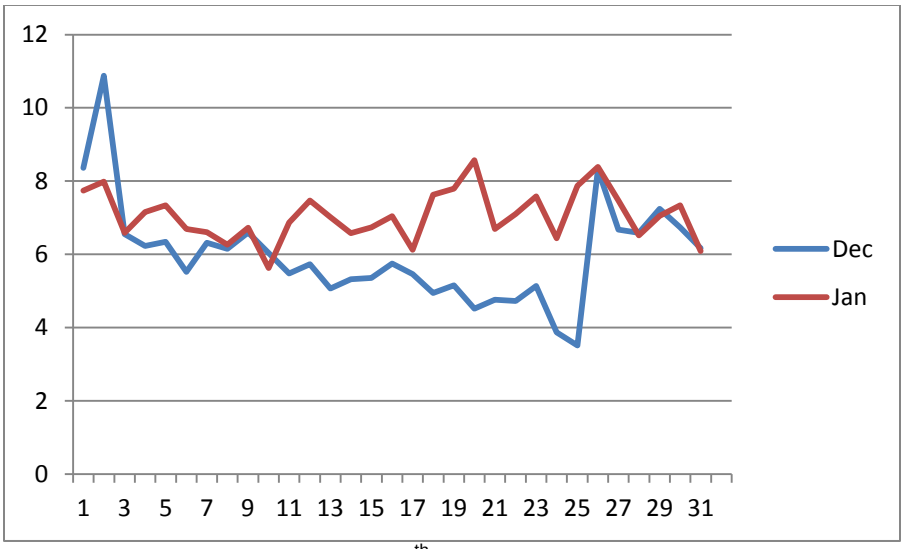

Figure 9. December v. January 90th Percentile Peaks.

How

"How did it happen?" This is a question journalists ask and try to answer. Sometimes we must answer it in technical and management reporting. When we do, it will almost always be as a one-time custom report. While it is very important to do this successfully it is also beyond the scope of this presentation to give it the attention it deserves.

A different HOW question all technical and management reporting should answer is, "How clear is this presentation?" How much room for ambiguity is there, how many questions will it raise and, most importantly, does it answer the basic questions of What, Where, When and Why? If the report or graph fails in this, the message we were trying to convey is lost. There will always be some room for questions and there will always be some level of miscommunication. You should work to make it as little as possible.

Consider the following quote from Larry Niven, a well-known science fiction writer. Larry specializes in "hard" science fiction, where science is integral to the story. Over the course of a writing career spanning decades he has come up with "laws for various things. His law of writing is:

"If you've nothing to say, say it any way you like. Stylistic innovations, contorted story lines or none, exotic or genderless pronouns, internal inconsistencies, the recipe for preparing your lover as a cannibal banquet: feel free. If what you have to say is important and/or difficult to follow, use the simplest language possible. If the reader doesn't get it then, let it not be your fault."

To put it more simply, "If you have nothing to say, say it any way you like. If what you have to say is important, say it as simply as possible." This is nothing more or less than the KISS principle: Keep It Simple.

It is not enough that your target audience understand your message. The quicker they understand it the more likely they are to respond to it and be engaged by it. As Larry says, "If the reader doesn't get it, let it not be your fault." Using overly complex graphs or graphs with a lot of unnecessary frills such as three dimensional effects or data skins do not add value to your core message and, as we have seen, can actually detract from it or change it entirely.

Let us look at one final graph and see how well it answers the journalism questions we have posed. Figure 10 is mainframe CPU growth of time for one company.

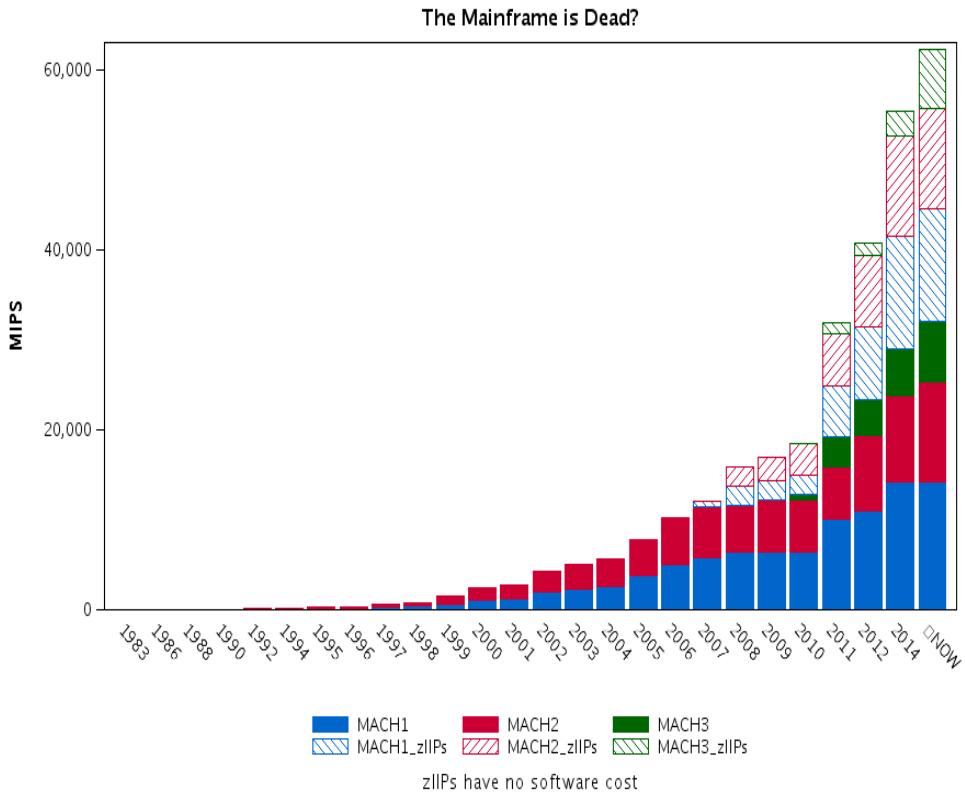

Figure 10. Mainframe growth over time for a single company.

Who is its audience? Its target audience is anyone in IT. Mainframe types will get the joke that the mainframe is not dead; non mainframe types will be shocked that the mainframe is alive and well, at least in one place.

What happened? The mainframe is growing by leaps and bounds in this shop.

Where did it take place? It happened on this company's mainframes.

When did it take place? Since the beginning of record keeping and it gets larger every year.

Why did it happen? Corporations do not spend money on IT they do not have to so the growth is real. Applications are growing on the mainframe. The hatched bars suffixed with "_zIIPs" show a second level of growth. zIIP engines cost no software dollars and have a hardware retail price of about one-fifth the cost of a regular engine. When they became available this company adopted them enthusiastically.

How clear is this presentation? The title of the graph tells us what it wants to say. The graph itself tells the story.

Conclusion

We started with the premise that effective technical and management reporting needs to tell a story. Since journalists have been writing newspaper stories for centuries with good effect we looked at their guidelines to see if they could be equally effective for us.

Using examples, we have seen that using the journalism story guidelines of who, what, where, when, why and how can help us make our reporting more effective. If we keep in mind WHO our audience is, WHAT happened, WHERE it took place, WHEN it took place and HOW clearly we are telling our audience this information we can tell them what they need to know much more effectively than simply posting the data. When we keep the additional questions WHY did it happen and HOW did it happen in our thoughts, we can be ready to explain why our reports say what they say. When we write special reports we can keep all of the questions in mind and give people receiving them the information they need as quickly and painlessly as possible.

And that is a key to being effective.