How NOT To Lie With Graphics and Tables – Don't Convey the Wrong Message!

SHARE session 17370

Greg Caliri
BMC Software, Inc.
Lexington MA, USA
Reference....

Darrell Huff (1913–2001)

“There is terror in numbers..”
Disclaimer and objective...

- This is INTRO stuff...

- Sure, you can lie, but you don’t WANT to lie

- And you don’t want to inadvertently “lie”

- An example of a deliberate lie with statistics...
Misrepresentation and Embellishment

2010
- 70% to organization
- 30% to Promoter

2014
- 20%
- 80% to the Promoter
Now what if …

- 2010 revenue was $2 million…
  - 80% to organization = $1.6 million

- 2014 revenue was $40 million…???
  - 20% to organization = $8 million
  - A “Fair share” argument might be made – politically, but also there is an organizational uptick of over 300 percent
  - Also note that I embellished the 80 percent pie slice with an exaggerated description
Mistakes happen ...

- Executive suite reporting – you can’t pull a trick like that
- You AVOID doing it accidentally
- Easy to trip up with z/OS charts, numbers, etc.
- Other examples follow
In performance and capacity reporting ---

- Generally, don’t use pie charts
- They show shares, not necessarily numbers for decision-making or trending
The “Gee Whiz” graph

- Most common method of confusion
- Easiest way to lie with a graph
- But you don’t to mislead anyone …
What is wrong here?
What is really happening...
The differences?

- Graph y-axis set to make a “good fit” – standard deviation

- But – for consistency, “fix” the Y-axis to capacity (MIPs, % of processor, MSUs)

- Same data, different presentation – and the presentation can be everything
Highly critical

- Users do comparisons
- Scaling variations can confuse the reader
- It will generate questions

AVOID CONFUSION!

IF YOU GET ASKED FOR EXPLANATIONS OF THE REPORT ITSELF – YOU’VE CONFUSED THINGS!
Let’s take a breath here

- Mainframe Information from SMF and RMF and other sources.....it’s an ocean in z/OS

- 50+ years, metrics for everything – most of which you’ll never need ...

- Unlike other platforms – YOU must report on what’s relevant....
... before going back

- Metrics chosen and reported
  - Units of Work, organized along lines of business
  - Units the audience can relate to
  - So it’s easy to report “irrelevant stuff”
  - Back on topic...
Including too much information on a graph

- It ISN’T better to report too much
- Get to the point
- Examples .... Here comes a bad one
Too much and irrelevant...attempt at correlation
Correlative, right amount to report
Correlations...

- OK to do them if the metrics are related to each other

- NOT OK – to associate non-related metrics

- Sometimes I/O, DB2 Calls, Paging, CPU spikes are related

- Non-related activities confuse – and MISLEAD
Correlations

- Separate coincidences from related events
- Avoid the “Stork Correlation”
- AFC/NFC Super Bowl correlation vs. Dow Jones Industrial Average
“Harum Scarum” – data out of context

- Analysis of anything with pure numbers can be very dangerous – tell the wrong story

- Large population = relatively small percentage of population (Nielsen ratings)

- Small population = relatively large percentage sampling (New Hampshire Presidential primary)
Oh my! We missed 100 percent...! Every hour had something wrong!
Details of misses? All the important ones “hit”.

<table>
<thead>
<tr>
<th>Workload Name</th>
<th>Number Of Logical Systems Running On</th>
<th>Performance Status</th>
<th>Number Of Occurrences</th>
<th>Number Of Occurrences With Bad Performance Index</th>
<th>Percentage Of Occurrences With Bad Performance Index</th>
</tr>
</thead>
<tbody>
<tr>
<td>BATLOW</td>
<td>1</td>
<td></td>
<td>5</td>
<td>0</td>
<td>0.00%</td>
</tr>
<tr>
<td>BATNRM</td>
<td>1</td>
<td></td>
<td>24</td>
<td>0</td>
<td>0.00%</td>
</tr>
<tr>
<td>BATNRM CICUTIL</td>
<td>1</td>
<td></td>
<td>3</td>
<td>0</td>
<td>0.00%</td>
</tr>
<tr>
<td>BATNRM OMVS</td>
<td>1</td>
<td></td>
<td>1</td>
<td>0</td>
<td>0.00%</td>
</tr>
<tr>
<td>BATPROD</td>
<td>1</td>
<td></td>
<td>25</td>
<td>0</td>
<td>0.00%</td>
</tr>
<tr>
<td>BATPROD CICUTIL</td>
<td>1</td>
<td></td>
<td>5</td>
<td>0</td>
<td>0.00%</td>
</tr>
<tr>
<td>DBTASKS</td>
<td>1</td>
<td></td>
<td>1</td>
<td>1</td>
<td>100.00%</td>
</tr>
<tr>
<td>DBTASKS DB2 DFS</td>
<td>1</td>
<td></td>
<td>24</td>
<td>11</td>
<td>45.83%</td>
</tr>
<tr>
<td>DBTASKS DB2 DFSZ</td>
<td>1</td>
<td></td>
<td>24</td>
<td>11</td>
<td>45.83%</td>
</tr>
<tr>
<td>DBNRMN OMVS</td>
<td>1</td>
<td></td>
<td>25</td>
<td>0</td>
<td>0.00%</td>
</tr>
<tr>
<td>STCHI</td>
<td>1</td>
<td></td>
<td>25</td>
<td>21</td>
<td>84.00%</td>
</tr>
<tr>
<td>STCLOW</td>
<td>1</td>
<td></td>
<td>25</td>
<td>0</td>
<td>0.00%</td>
</tr>
<tr>
<td>STCNRM</td>
<td>1</td>
<td></td>
<td>25</td>
<td>0</td>
<td>0.00%</td>
</tr>
<tr>
<td>STCNRM DB2 DFS</td>
<td>1</td>
<td></td>
<td>24</td>
<td>0</td>
<td>0.00%</td>
</tr>
<tr>
<td>STCNRM DB2 DFSZ</td>
<td>1</td>
<td></td>
<td>24</td>
<td>0</td>
<td>0.00%</td>
</tr>
<tr>
<td>STCNRM HTTP</td>
<td>1</td>
<td></td>
<td>24</td>
<td>0</td>
<td>0.00%</td>
</tr>
<tr>
<td>STCPAS</td>
<td>1</td>
<td></td>
<td>25</td>
<td>19</td>
<td>76.00%</td>
</tr>
<tr>
<td>STCPROD</td>
<td>1</td>
<td></td>
<td>25</td>
<td>1</td>
<td>4.00%</td>
</tr>
</tbody>
</table>
NOT = “Failures in every interval”

BUT = “Six areas – major or minor ones – need addressing”

And analyze and report impact on the business

Also note time of day .... Development batch at 3 am? One job misses?
Outliers

- Historical shift right now
- At one time – you’d drop ‘em
- Examples....
## Use to properly calculate average response time for the day

<table>
<thead>
<tr>
<th>Time</th>
<th>Tx Count</th>
<th>Avg Rpt</th>
<th>Tx &gt; 5 Sec</th>
<th>Tx &gt; 10</th>
<th>Tx &gt; 1000</th>
<th>Outliers</th>
<th>Factor (Tx Count * Rpt)</th>
</tr>
</thead>
<tbody>
<tr>
<td>9:00 AM</td>
<td>22516</td>
<td>0.23</td>
<td>34</td>
<td>4</td>
<td>0</td>
<td></td>
<td>5178.68</td>
</tr>
<tr>
<td>10:00 AM</td>
<td>33534</td>
<td>0.26</td>
<td>36</td>
<td>13</td>
<td>0</td>
<td></td>
<td>8718.84</td>
</tr>
<tr>
<td>11:00 AM</td>
<td>66221</td>
<td>0.33</td>
<td>55</td>
<td>22</td>
<td>1</td>
<td>1123</td>
<td>21852.93</td>
</tr>
<tr>
<td>12:00 PM</td>
<td>62134</td>
<td>0.31</td>
<td>78</td>
<td>21</td>
<td>0</td>
<td></td>
<td>19261.54</td>
</tr>
<tr>
<td>1:00 PM</td>
<td>44986</td>
<td>0.89</td>
<td>45</td>
<td>33</td>
<td>1</td>
<td>3215</td>
<td>40037.54</td>
</tr>
<tr>
<td>2:00 PM</td>
<td>57810</td>
<td>0.3</td>
<td>77</td>
<td>29</td>
<td>0</td>
<td></td>
<td>17343</td>
</tr>
<tr>
<td>3:00 PM</td>
<td>65007</td>
<td>0.33</td>
<td>49</td>
<td>19</td>
<td>0</td>
<td></td>
<td>21452.31</td>
</tr>
<tr>
<td>4:00 PM</td>
<td>58711</td>
<td>0.31</td>
<td>58</td>
<td>15</td>
<td>0</td>
<td></td>
<td>18200.41</td>
</tr>
<tr>
<td>5:00 PM</td>
<td>49867</td>
<td>0.39</td>
<td>35</td>
<td>8</td>
<td>0</td>
<td></td>
<td>19448.13</td>
</tr>
<tr>
<td>Tot Tx</td>
<td>460786</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>171493.38</td>
</tr>
<tr>
<td>Rpt Day</td>
<td>0.372176</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Norm !</td>
<td>0.362761</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Pulling the one outlier recalc to .81s, so there are further problems, possible looping tx impacting others.*
Whither monthly license charges?

- With MLC licensing
  - Outlying CPU Usage – requiring capping / time shifting of work, now is more critical
  - Cannot be ignored, and it must be handled –
There’s a spike I might want to adjust – you might call it an outlier.
Numbers behind it...
MSU/MLC – look for those outliers!

BMC Cost Analyzer for zEnterprise

CICS TS for z/OS V3 MSU Utilization on BRYALS

<table>
<thead>
<tr>
<th>LPAR Name</th>
<th>Group Name</th>
<th>Sysplex Name</th>
<th>R4HA MSU Utilization</th>
<th>R4HA MSU Utilization Weight</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>11.46 %</td>
<td>29</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>42.25 %</td>
<td>107</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>30.83 %</td>
<td>78</td>
</tr>
</tbody>
</table>

Workload Reports
- Workloads by Suites
- Workloads by Service Class
- Workloads by Subsystem Address Space
- Workloads by Importance
- Workloads by Workload Manager
- Workloads by Report Class

CPC and LPAR R4HA MSU Utilization

BRYALS: 7/29/2015 10:00 AM, 253(MSU), R4HA First Peak
QACI: 7/29/2015 10:00 AM, 0(MSU)
JPP: 7/29/2015 10:00 AM, 0(MSU)
SJS: 7/29/2015 10:00 AM, 107(MSU)
SYSM: 7/29/2015 10:00 AM, 78(MSU)
Back to (another) deep breath

New analysts to mainframe – you need to organize

- Identify units of work, organize those units along lines of business
- Establish accountability!
Epilog

- Understand what is going on within the system
  - Your own graphs can “throw” you
  - Outliers may have low or high impact, know them
  - Be able to explain everything possible
Know your audience!

- Executives don’t care about paging or I/O rates, but the bottom line ($$$) – know their lingo, speak in their terms
- Users want to know their own successes/failures
- Techs want to know those system intricacies
- Consider a modular approach
Paper on modular approach.

- Gregory V. Caliri:

  Performance reporting in the 21st century – changes in scope and direction. Int. CMG Conference 2006

- Greg_caliri@bmc.com

- 781 257 3120
Epilog

- Know your metrics!
  - Know what you’ll be explaining
  - Practice practice, and more practice!
  - Take suggestions but
  - Don’t get taken down side roads
Epilog

- If you are asked “what does this mean?” over and over again – next time make it simpler
- Don’t present something you can’t explain
- Don’t present irrelevant info
- Do explain everything that *is* relevant
Thank you!

“I didn’t know my parents could dress so cool”

– Jill Caliri Patruno