Rules and principles for calculating and highlighting wins

Version 2 | April 2022

About this document

This document explains how we handle the kinds of results and wins we see most commonly at PT. We originally drafted it for in-house use, to assemble institutional knowledge and promote consistency in our practices. In the spirit of transparency, we've decided to make this document public so that any interested party can understand the principles and practices that underlie PT reports and the deliverables we base on those reports.

If you'd like to skip the details, jump to Part 6: A summary for readers of PT reports.

If you have questions about any of the practices we outline in this document or any of the results or methodologies in a PT report, please email info@principledtechnologies.com.

PART 1

The principles behind the way we report our wins

A key element of our fact-based marketing work lies in finding results for our clients—but our work is almost always more than delivering a spreadsheet of numbers. When we create marketing collateral, we translate our results into wins, then translate those wins into meaningful benefits for our target audience(s).

As data scientists, we thus must interpret our findings into results that are clear and relevant to readers.

As marketers, we use compelling writing and design to show the wins in ways that engage our audience.

As principled experts, we must calculate and highlight these wins in ways that are accurate, provable, and defensible, and never exaggerated, misleading, or deceptive.

With those components of our mission in mind, this document explains how to handle the kinds of results and wins we see most commonly at PT. A single document can't cover every possible situation; our work is too varied. A few key principles, however, apply across the board.

- **Present wins in a way that reflects how much they matter to the audience.** Every microsecond might count when you're counting on high-speed trading software, but when you're opening a webpage, your experience doesn't change if the website opens a millisecond faster or slower. If we don't believe a win really matters, we shouldn't use words or graphics to suggest that it does.
- **Don't inflate wins.** There are many ways to make wins look bigger than they are—and we have to avoid all of them. Be fair and reasonable, and don't exaggerate the importance of modest wins in words or graphics.
- **Don't be deceptive, misleading, or confusing.** We should present the data in a way that is logical, fair, accurate, clear, and defensible.

One policy we follow might appear to be at odds with the third point above: We often publish only a subset of the results from our testing. For example, if our testing compares the way that two products handle seven use cases, and our client's product wins on four of them, our report might feature only these four use cases. This is one of many reasons that we must always be very careful to keep our language specific when discussing wins, and to avoid generalizing.

There are countless other factors to consider. The background of our audience is important—different audiences understand numbers differently, and we should express wins in ways that make them understandable and relevant to our target audience. We must also adhere to EULAs and other legal restrictions, which may constrain the ways we can show results. And of course, we must carefully listen to our clients' preferences and honor them if possible without violating our principles.

These issues are complicated, so for each report and each win, we need to think carefully about how to follow our principles.

From the moment we start talking about our wins, everyone on the PT team should feel comfortable objecting when they don't agree with how we're presenting a result. This includes the writer, designer, technical people, project manager, or anyone else. This might create tension within the team, and that's okay. Similarly, our clients may ask us to say or show more than we believe is appropriate, and that can create tension between us and the client. That's okay, too. Getting to the clearest, most principled way to treat the wins we find requires thought and discussion. If the team members are having trouble resolving an issue, someone on the team should escalate the issue to their area lead(s).

From the moment we start talking about our wins, everyone on the PT team should feel comfortable objecting when they don't agree with how we're presenting a result.

To aid in those discussions, this document offers guidelines for putting our principles to work in recording and truncating results; calculating wins, including selecting which results to use in that process; and writing and designing to highlight wins.

We've also created an accompanying spreadsheet, PT results handling guidelines: Using Excel, that provides examples of how to use some of these principles in Microsoft Excel, including how to use formulas appropriately and how to create simple Excel charts.

These documents won't answer every question, but no document could! Keep these principles and guidelines in mind, and when you aren't sure of the right path, talk it through with the team. We do our best work when we work and think together.

Recording and rounding results

Our testing work often yields numerical results with multiple decimal places, which begs the question: How and where should you shorten these results and the wins based on them?

Don't round, truncate

Rounding is a popular way to shorten numbers, and the "5 and over round up" formula is the approach that Excel uses by default. Depending on the number of decimal places showing, 3.1415926 becomes 3.14, 3.142, 3.1416, etc. We do not round results or wins because doing so can cause us to overstate a win, which can reflect negatively on both our clients and PT.

Instead, we **truncate** results and wins by simply cutting off all the digits after a certain number of decimal places. If we were truncating to two decimal places, 3.1415926 would become 3.14; if we were truncating to three decimal places, it would become 3.141; and so on.

By truncating rather than rounding, we present wins that are conservative and thus easily defensible.

So how does this work in practice?

When we are testing and recording results, we record whatever the hardware or software tool gives us, with however many digits the tool provides—whether the tool is a benchmark or a hand timer.

When it's time to include those results in the Science section of our report and use them to calculate wins, we may need to truncate the results for one of two reasons:

- To show numbers that are reasonably easy to read and understand. If a tool generates results with 10 decimal places, we might truncate them in the Science to make them simpler to read. Unless tech strongly objects, truncate to four decimal places.
- To reflect the precision of the tool. In these cases, we cut off all digits after N, where N is the level of precision that makes sense for that result or win. We often need to do this with hand-timed results.

We do not round results or wins because doing so can cause us to overstate a win... instead, we truncate.

How can we know whether to truncate results to reflect the precision of a tool? As data scientists, it's our job to take the results a testing tool generates and interpret them into task-relevant, human-understandable measurements.

With data that hardware or software generates, the element of human variation is not present, and we typically do not need to truncate results for precision reasons. (Of course, we may have other reasons, such as relevance, for truncating some digits of such results.)

With hand-timed data, we must reckon with human variation, because multiple people hand-timing an event could come up with different times. We define hand-timing data as any data points we get from a hand-timer, stopwatch, or clock.

If most of our hand-timed results for a test are under a minute and we believe a difference of tenths of seconds matters to the user, we truncate results to the tenth of a second. (We cannot go smaller than a tenth of a second, because human reaction time is generally between one-tenth and two-tenths of a second.[1])

If most of our hand-timed results for a test are over a minute and we believe a difference at the level of tenths of seconds does not matter to the user, we truncate to the second.

We always include the numbers that we use for win calculations in the Science section of the report document.

When those numbers—both the results themselves and the wins that come from them—include an unwieldy number of decimal places, we help our readers by using shortened versions of them in the report body and front page. To make the numbers shorter, we truncate them rather than round them, as we explained earlier.

For some reports, a reader reviewing only the report, and not the results in the science, might notice that certain calculations appear slightly off or that percentages representing parts of a whole don't add to 100. In the section of the Science where we present our policies regarding calculating and presenting results, we explain that these apparent inconsistencies are due to our truncating numbers in the report body.

In PT results handling guidelines: Using Excel, we show how to use the Excel truncate function to shorten results and wins. You must actively check to see when Excel is rounding data behind the scenes and not use those rounded numbers. The easiest way to do this is to go to the number formatting panel in the home tab and increase the number of decimal places.

1 David A Faux and Janet Godolphin, "Manual timing in physics experiments: error and uncertainty," accessed October 26, 2020, http://epubs.surrey.ac.uk/850094/1/AJP_Faux_revision2.pdf.

PART 3

Choosing representative results to use for analysis

We always aim to present results and wins that are defensible and repeatable. To this end, for most of our performance testing projects, we conduct multiple runs (or iterations) of each test. When we do so, we expect to get different results each time. We refer to the range in these results as variance or variability.

Usually, we troubleshoot and retest until we have three or more test runs with results that vary to a degree that is reasonable for that type of test. For example, we might expect little to no variance on a short networking test, while we might expect high variance on a test scenario involving storage rebalancing. The tech team should understand the characteristics of each case and how much variance we should accept.

Once we assemble a configuration that results in at least three runs with acceptable variance, we must determine which of the three results we will use to calculate wins. We also need to explain our reasoning for this choice in the Science.

For some projects, particularly those in the human usability or management domain, conducting multiple iterations is less appropriate, so parts of this section may not apply.

To understand how to use Excel to find the median or mean of a data set, refer to PT results handling guidelines: Using Excel.

We always aim to present results and wins that are defensible and repeatable.

Median

As a default, we use the median of three test runs to calculate wins. The median is the middle value separating the greater and lesser halves of a data set. (Use the Excel formula MEDIAN to calculate the median of a data set.) This approach allows us to base our wins on an actual test result, which we—and many of our clients—prefer. Depending on the details of the project and the client's preference, we may report all three test runs in the science, or report only the median.

Medians work best when you have an odd number of results. With an even number of results, the median is the midway point between the middle two results, which is not an actual test result. For this reason, we try to avoid having an even number of runs or using the median when we must have an even number of runs.

For some tests, we record more than one metric (e.g., throughput and response time). In these cases, the median throughput and response time we report **must be from the same run**, because the two are related. For each test, we should identify a primary metric (e.g., throughput) and use that to determine the median run. We then report the secondary metric from that run, even if it isn't the median of all the runs. We should include language that explains this.

For example:

> *"We conducted three test runs and report the results of the median run. We use throughput as our primary metric to determine the median run. Any secondary metrics we report, such as response time, come from the median run we have identified based on throughput."*

Arithmetic mean

Sometimes we use the arithmetic mean, or average, to show some results. (Use the Excel formula AVERAGE to calculate the mean of a data set.) We might use the arithmetic mean to compare trends across groups of devices or show average performance across multiple VMs of the same size.

To calculate the arithmetic mean, you add all of the values in a set and then divide by the number of values in the set. The arithmetic mean of a set of results is typically not an actual result. (Think of the average number of children per US family; in 2019, this was 1.93, a number that no family has.[2]) For this reason, we use the arithmetic mean less frequently than the median.

Other kinds of averages

The median is our first choice, followed by the arithmetic mean. But some situations require a different approach, such as when we have multiple data sets of varying types, sizes, and/or units. If our workload contains multiple benchmarks, or the systems under test have multiple VM sizes, an aggregate or geometric mean might be the best way to represent the data.

These exceptions are rare and complicated; you can see one example of geometric mean usage in the 2019 report Improve performance and minimize latency for IOintensive apps by pairing Intel NVMe SSDs with Intel Virtual RAID on CPU (VROC). If you are not sure if you should be using a different average, raise the issue with your team and, if need be, your area lead.

2 "Average number of own children under 18 in families with children in the United States from 1960 to 2019," accessed October 26, 2020, https://www.statista.com/statistics/718084/average-number-of-own-children-per-family/.

PART 4

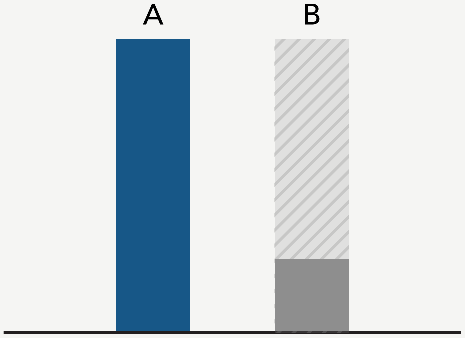
Calculating wins and writing about them

Frequently, the first step in translating our results into benefits for our audience—and our client—is calculating how much better one product performed a task than another and then using that number to write a claim. We might say Product A took 20 percent less time than Product B to open an application, or that Product A did five times the transactional database work as Product B.

When we calculate those wins and write about them, it's critical that we be accurate and specific in the formulas and phrasings we use. **We can portray our results in many different ways, and the words we use around a quantitative win must match the math we use to calculate that win.**

Below, we describe how we calculate wins in a variety of different situations, why we use the rules and formulas we do, and the language we should use to describe the wins in each situation. We also offer examples. For more help and examples around using these formulas in Excel, see PT results handling guidelines: Using Excel.

As a rule, we calculate the improvement or advantage between two results by finding the difference between them (via subtraction), and then calculating the relationship of that difference to the losing result (via division).



In this chart, the shaded area represents the difference between the two results. The formula to find the value of that area is A – B.

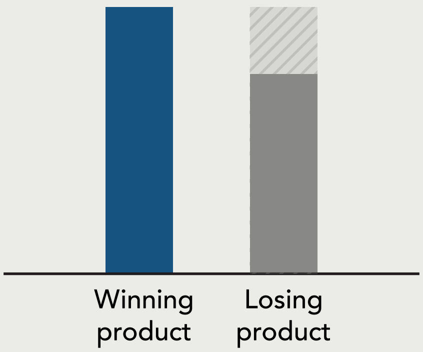
The words we use around a quantitative win must match the math we use to calculate that win.

When larger numbers are better, wins are increases

Larger numbers are desirable when more of a thing is beneficial—more work in a given amount of time, more users supported, higher benchmark scores, longer battery life, etc. In these situations, we calculate wins in terms of the increase from the losing result to the winning result.

When the winning result is less than twice the losing result, the improvement (the shaded area) is between **1 percent and 99 percent of the losing result**. In these cases, we usually use a percentage to express the win:

Larger results are better
Improvement = increase



Formula

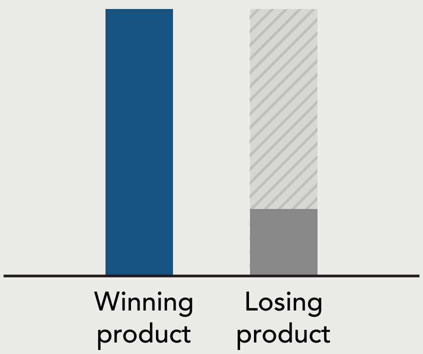
$$\frac{\text{larger (winning) result} - \text{smaller (losing) result}}{\text{smaller (losing) result}}$$

Phrasing

"Winning product achieved 40 percent more OPM than losing product"

When the winning result is more than twice the losing result, the improvement (the shaded area) is **greater than 100 percent of the losing result**. In these cases, we usually use a multiplier ("times") because many people find percentages in this range harder to understand. As we explain on the next page, there are two ways to phrase these types of wins, and each has its own formula.

Larger results are better
Improvement = increase



Formula

$$\frac{\text{larger (winning) result}}{\text{smaller (losing) result}}$$

Phrasing

"Winning product achieved **2.6 times the OPM that** *losing product achieved"*

Formula

$$\left(\frac{\text{larger (winning) result}}{\text{smaller (losing) result}}\right) - 1$$

Phrasing

"Winning product achieved **1.6 times more OPM than** *losing product achieved"*

Important note about phrasing wins with "times"

In casual conversation, we sometimes use multipliers inaccurately. Referring to two cars, one that costs $25K and one that costs $100K, we might say the second one is "four times more" than the first one. That's technically incorrect, but that doesn't stop people from saying it.

Let's change the numbers from car prices in dollars to OPM database servers delivered. The difference between the numbers is 75K, which is three times the 25K OPM the losing server achieves. That means the winning server achieves **three times more OPM** than the first server. We can say that, or we say the second server achieves "four times as many OPM" or does "four times the work." If we want to use the word "more" (or if a client wants us to), we must subtract 1.[3]

When smaller numbers are better, wins are decreases or savings

Smaller numbers are more desirable when less of a thing is better. The most common examples in our work involve time (how long a person needs to perform a task, or how long a system takes to respond) and money. We all want things to be as quick and inexpensive as possible.

When the winning result is more than half the losing result, the improvement (the shaded area) is between **1 percent and 49 percent of the losing result**. In these cases, we usually use a percentage to express the win.

Formula to calculate percentage savings:

$$\frac{\text{larger (losing) result} - \text{smaller (winning) result}}{\text{larger (losing) result}}$$

Phrasing

"Winning product performed the task in **28 percent less time** *than losing product"*

3 If this is still hard to grasp, the following passage from "Common Errors in Forming Arithmetic Comparisons" by Milo Schield might help: "If B is three times as much as A, then B is two times more than A – not three times more than A. The essential feature is the difference is between 'as much as' and 'more than.' 'As much as' indicates a ratio; 'more than' indicates a difference. 'More than' means 'added onto the base'. This essential difference is ignored by those who say that 'times' is dominant so that 'three times as much' is really the same as 'three times more than'." You can find this passage at https://web.augsburg.edu/~schield/MiloPapers/984OfSigCompare3.pdf (accessed October 26, 2020).

When the winning result is less than half the losing result, the improvement (the shaded area) is **greater than 50 percent of the losing result**. In these cases, it's fine to use a percentage to express the decrease, or savings. Using an appropriate fraction can make the savings easier for many readers to understand.

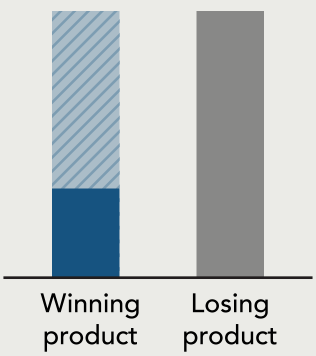
Say the winning product took 15 seconds, and the losing one took 45 seconds. We often focus on the shaded area and make claims such as:

- The winning product needed 66 percent less time
- The winning product cut the time needed by two-thirds

We can also flip this and focus on the time used rather than the time saved:

- The winning product did the job in one-third the time

Smaller results are better
Improvement = decrease



Formula for percentage savings

$$\frac{\text{larger (losing) result} - \text{smaller (winning) result}}{\text{larger (losing) result}}$$

Formula for fraction

$$\frac{\text{smaller (winning) result}}{\text{larger (losing) result}}$$

Phrasing

"Winning product cut the time to do the task ***by 66 percent****"*

OR

"Winning product cut the time to do the task ***by two-thirds****"*

Phrasing

"Winning product did the job in ***one-third the time****"*

Note that regardless of the numbers involved, the savings (the shaded area) will always be less than 100 percent of the larger result. **(If you reduce something by more than 100 percent, the result is a negative number.) If you end up with savings greater than 100 percent, you've made an error.**

"Less time" vs. "faster"

It's tempting to use "faster" in our win statements because it's intuitive and easy to grasp. Doing so, however, can be dangerous; for example, the following two statements are not interchangeable:

- "Winning product did the job in 33% less time than losing product"
- "Winning product did the job 33% faster than losing product"

The first statement refers to elapsed time. This is the approach we prefer and use most often, because it is the clearest and least open to misinterpretation. To calculate it, we use the percentage savings formula:

- (Larger – smaller)/larger

The second statement refers to a rate. We strongly prefer to avoid this, because different people can reasonably interpret "33% faster" in different ways, making it a less clear and less precise way to provide the result. We will, however, use it and explain it if a client insists. To calculate it, we use the alternative formulas:

- Larger/smaller
- (Larger/smaller) – 1

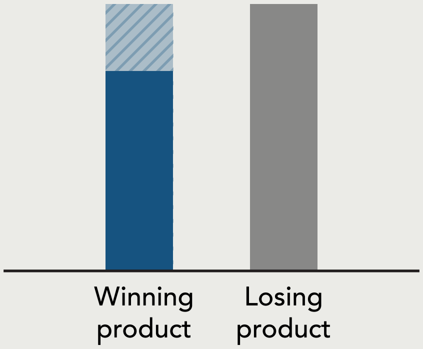
Remember our discussion earlier about the difference between "3 times as much" and "3 times more"? The rule was that if you use the word "more," you must subtract 1. In this situation, think of "faster" as being the equivalent of "more fast," and apply the same rule.

If the time to perform a task is 150 seconds for the winning product and 225 seconds for the losing product, we would say, "Winning product performed the task in 33% less time than losing product."

If a client insists that we use "faster," we can say "Winning product performed the task 50% faster than losing product."

We can also say simply, "Winning product was faster than losing product"—provided, of course, that's what we found.

Smaller results are better
Improvement = decrease



Formula	**Formula**
$$\frac{\text{larger (losing) result}}{\text{smaller (winning) result}}$$	$$\left(\frac{\text{larger (losing) result}}{\text{smaller (winning) result}}\right) - 1$$
Phrasing	**Phrasing**
"Winning product did the job ***1.5 times as fast as*** *losing product"*	*"Winning product did the job* ***50 percent faster*** *than losing product"*

Note: The following phrasing is mathematically correct, but sounds odd:

- *"Winning product did the job 150 percent as fast as losing product" (first formula)*
- *"Winning product did the job 0.5 times faster than losing product" (second formula)*

Special situations

Occasionally we measure temperature or sound level, which require us to use specific rules for calculating wins. **We do not use our usual formulas to quantify differences in temperature or sound level.**

For temperature, quantify the difference by saying, "X degrees [Fahrenheit or Celsius] [cooler or warmer]." **Do not use a percentage.**

We typically measure sound level in decibels. Decibel changes use a logarithmic scale, and the math involved in calculating differences and wins is complex. To measure the difference between two decibel levels and to find the win, we recommend using an online calculator such as this one: http://www.sengpielaudio.com/calculator-levelchange.htm.

Normalizing results

Sometimes, rather than reporting the actual numbers we found, we must normalize our results. This means we present our findings in relative terms by assigning the value of 1 to one product's result (the reference result) and showing the results of other products as relatively greater or lesser.

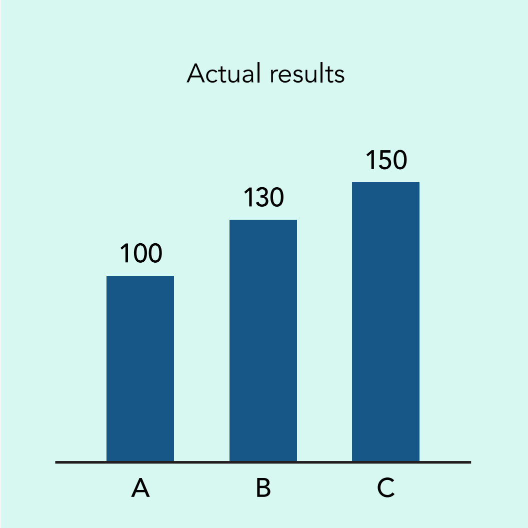
There are a number of reasons why we might need to do this:

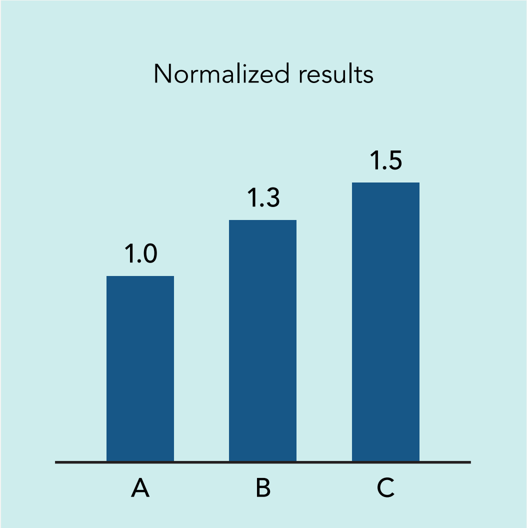
- To show widely varying sets of values, with different units and metrics, on a single chart
  Example: In one chart, we need to show both latency and throughput differences between two systems
- To avoid using actual numbers
  - Because of licensing restrictions, such as those from Oracle, whose EULA expressly prohibits publishing results
  - Because doing so focuses the data more squarely on key areas, such as price/performance
    Example: A product might offer a better price/performance metric but have a price that the client would prefer not to highlight
- To recast a "smaller is better" win as an easier to visualize "higher is better" result

The formula we use is:

$$\frac{\text{result}}{\text{reference result}}$$

In the charts to the right, Product A is the reference product. We assign its result a value of 1 and express the results of the other products relative to Product A.





PART 5

Highlighting wins with graphic design

When we're using graphics to show results or wins, our goals are the same as at every other stage of the project—to fairly and accurately show what we found in a way that makes the results easy for our readers to understand and that meets the client's goals. Below, we discuss the types of charts we typically use in our deliverables, when and how to use them, and how to deal with a few common challenges.

We'll need to deal with unique issues as they arise; in those cases, we should use the principles below as guidelines and talk through the problem until we come up with a good solution.

We typically do not use Excel charts in our final deliverables; however, it's often helpful to create preliminary versions of charts in Excel before the designers create final versions using an Adobe® program. For help on how to create simple charts in Excel, see PT results handling guidelines: Using Excel.

Our goal is to fairly and accurately show what we found in a way that makes the results easy for our readers to understand and that meets the client's goals.

Standard bar charts

Most of our wins are simple comparisons of two to three data points that represent benchmark scores, timings, steps, or other metrics among two or three solutions. Usually, the clearest expression for these is a bar chart.

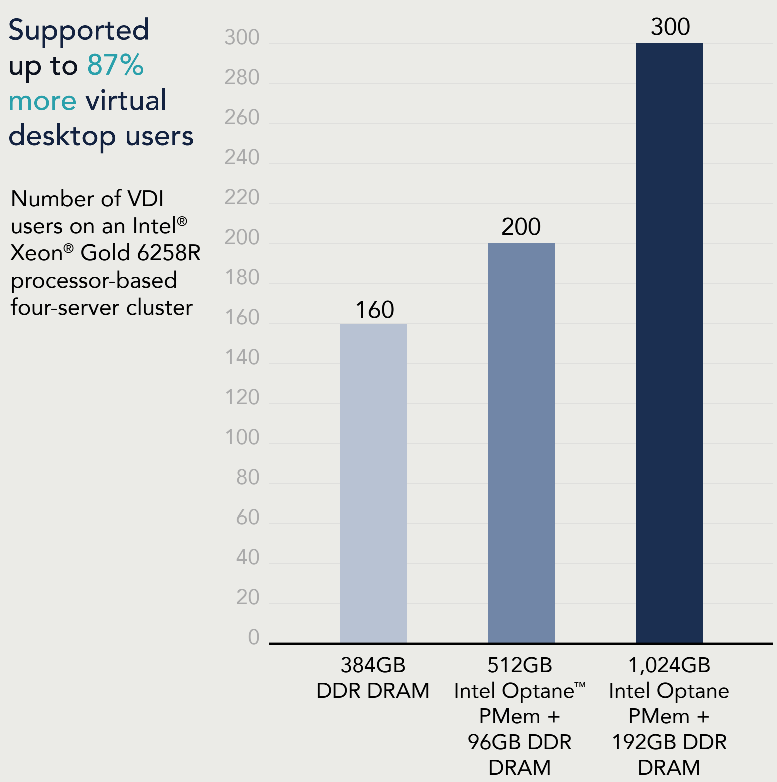
When you're working with bar charts:

Do

- Start the y-axis at 0
- Try to make all of the bars representing the same kind of data use the same scale
  - This makes it easier for a reader to understand the importance of a win and avoids misrepresenting small wins
  - Bar charts representing different kinds of data can use different scales

Do not

- Put bars representing different metrics on the same baseline, e.g., time and steps or OPM and latency
- Magnify sections of charts to emphasize a small difference



(Source: Principled Technologies.)

How should we handle reports where we have many different data sets with the same metric but widely varying results?

Many of our end-user reports include a large number of timing comparisons, where all results are in seconds, but the range of seconds varies considerably from task to task. For example, for one task, Product A took 3 seconds and Product B took 4 seconds, while for another task, Product A took 300 seconds and Product B took 400 seconds. Both wins matter, and they use the same metric—seconds—but one range is 100 times the size of other.

In the past, we've sometimes ignored the rule that all bar charts representing the same kind of data should use the same scale. This has helped us make reports that are visually engaging and easy to read, but it can represent the data in a potentially misleading way. We need to change this practice.

First, as we discussed earlier, we must make sure the differences we're highlighting would really matter to the user. If they don't matter, we shouldn't call them out. For some tasks, a difference of a second could be noticeable and frustrating; for others, it might not be.

Not every approach will work for every report, so we should discuss this challenge each time it comes up and decide how to handle it.

Assuming we're confident in the importance of all our wins, we have several options for how to handle the wins graphically. Not every approach will work for every report, so we should discuss this challenge each time it comes up and decide how best to handle it. The options we have include:

- **Using different units on different pages.** If you can express some of the wins in minutes and others in seconds, use minutes as the scale on one page and seconds as the scale on another.
- **Grouping results and using different scales on different pages.** Say you have three wins with results under 10 seconds and three wins with results between 40 and 60 seconds. You can group the first three wins on one page with one scale and the second three wins on another page with a different scale. This is not ideal, but is appropriate for some reports.
    - **When you do this, use words to call out the wins on top of the graphs.** Calling out the number of seconds saved can help readers notice the different scales of the wins.
- **Using words for some or all wins.** Instead of showing a bar graph at all, note that Product A took 3 seconds and Product B took 4 seconds, and then call out the percentage win.
- **Presenting the numbers in tables instead of graphs.** In a table, we don't need to worry about exaggerating the importance of a small win.
- **Normalizing the numbers.** Normalization enables us to put everything on a single scale.

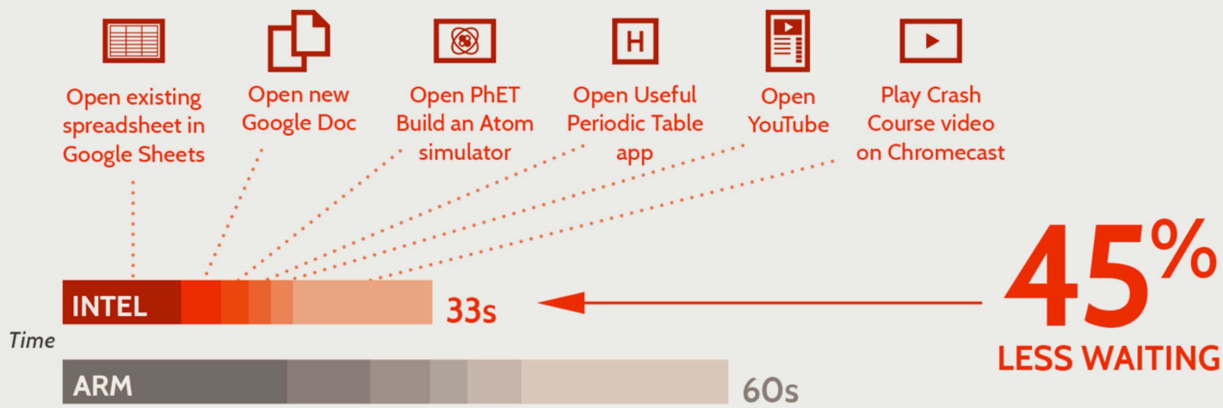
This change in practice will make it more difficult to design a report that has a chart for every win, will make some wins in some reports be difficult to see, and might affect the attractiveness of our design overall. Some clients may push back on this practice, because they want us to show every win with the same level of detail. We'll need to explain our reasoning and work with the client to try to find a compromise using some of the techniques above.

Stacked bar charts

Stacked bars aren't ideal for detailed value comparisons, so they shouldn't be a standby option.[4] But they can be useful in two scenarios:

- When you need to show a value that is the sum of other values, and the most important comparison is that between the totals.
- When you need to show the relative importance of one component of a total. For example, in a TCO analysis, a stacked bar chart can illustrate that over a certain period, hardware accounts for a relatively small part of total costs compared to software or administrative costs. Be sure to put the primary category of comparison—hardware in this example—at the bottom of the stack (or at the left when the bars are horizontal) to make it easier to compare values.

This chart emphasizes the total elapsed time for a classroom workflow without specifying how long the individual tasks took.



(Source: Principled Technologies.)

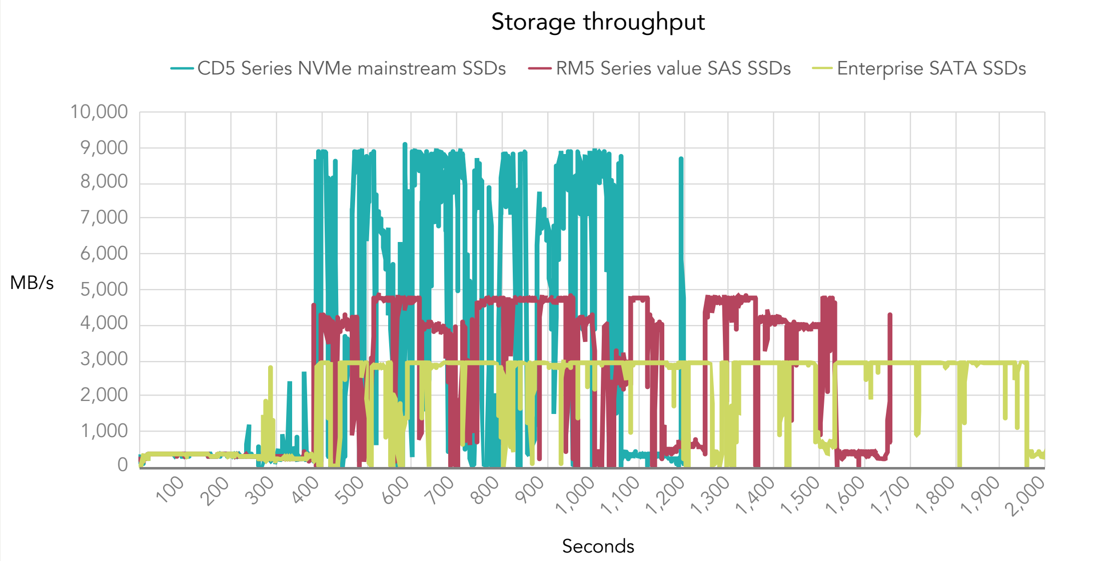
4 Robert Kosara, "Stacked Bars Are the Worst," accessed October 26, 2020, https://eagereyes.org/techniques/stacked-bars-are-the-worst.

Line charts

For comparing trends over time, scaling out, etc., a line chart is the best choice.

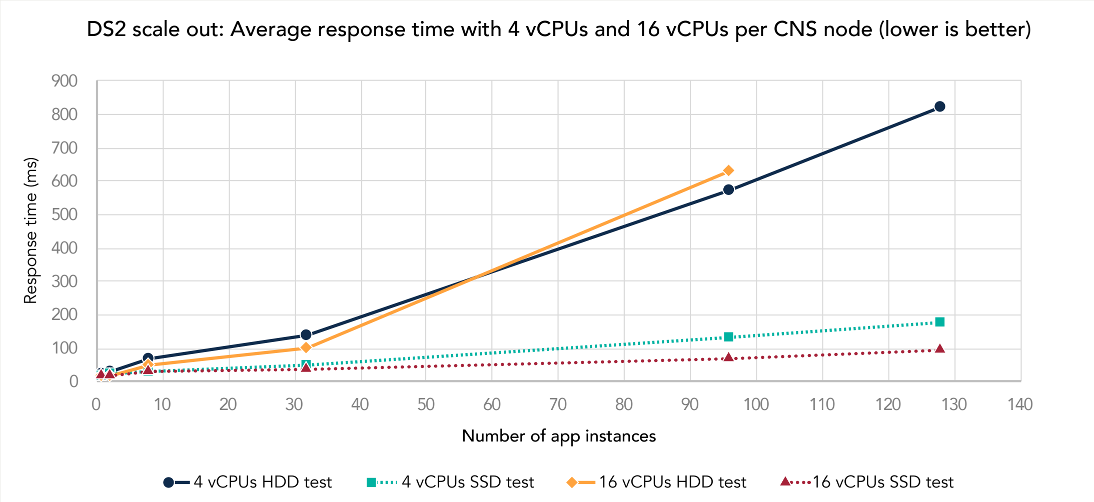
This line chart shows throughput over time.

(Source: Principled Technologies.)



This line chart shows trends across iterations of testing.

(Source: Principled Technologies.)



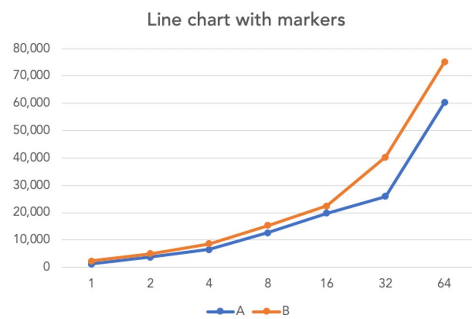
By default, line charts in Excel treat the y-axis values as numbers, and plot them accordingly. Be aware, however, that they treat the x-axis values as text labels, and plot them at even intervals regardless of their values. In the two previous charts, note that each chart's x-axis starts at zero and the increments increase by a fixed amount (100 in the first chart and 10 in the second chart).

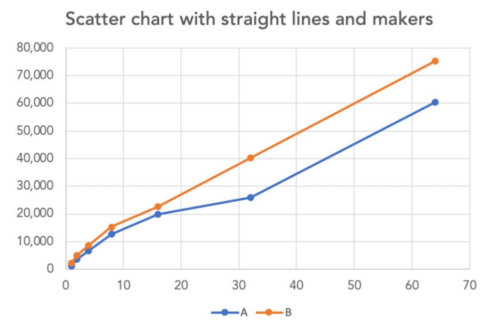
For some sets of results, the increments are not regular as they are in the charts above. In the table below, the number of VMs starts at 1 then doubles several times. If we plot these data points using the default line chart in Excel, the distance between each of the x-axis increments is consistent, even though the number of VMs increases irregularly (see the chart next to the table). As a result, the line presents a misleading picture of how performance increases.

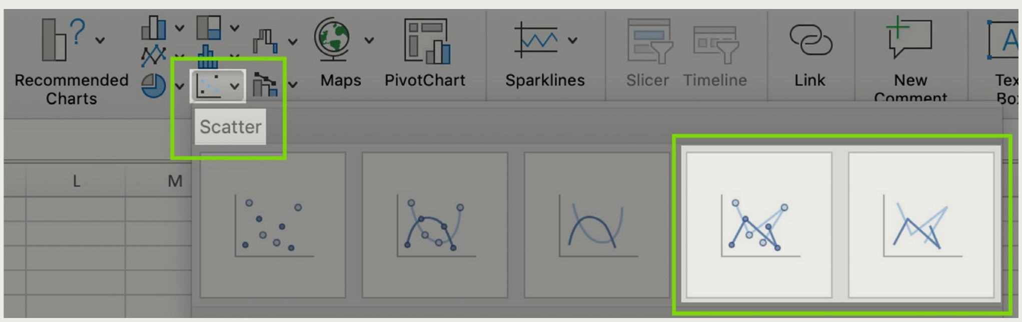
The solution is to use a **scatter chart with straight lines**, which treats x-axis values as numbers and plots them accordingly (see the chart to the right of the line chart). When you're doing this, be sure that the x-axis labels are easy to read, and use whole numbers only.

Performance

VMs	A	B
1	1254.65	2351.65
2	3687.25	4987.32
4	6589.32	8654.65
8	12659.32	15321.86
16	19854.87	22489.65
32	25916.65	40258.96
64	60259.36	75126.54



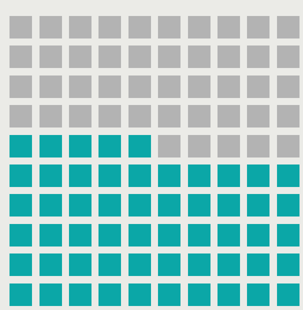




Showing a percentage of a whole

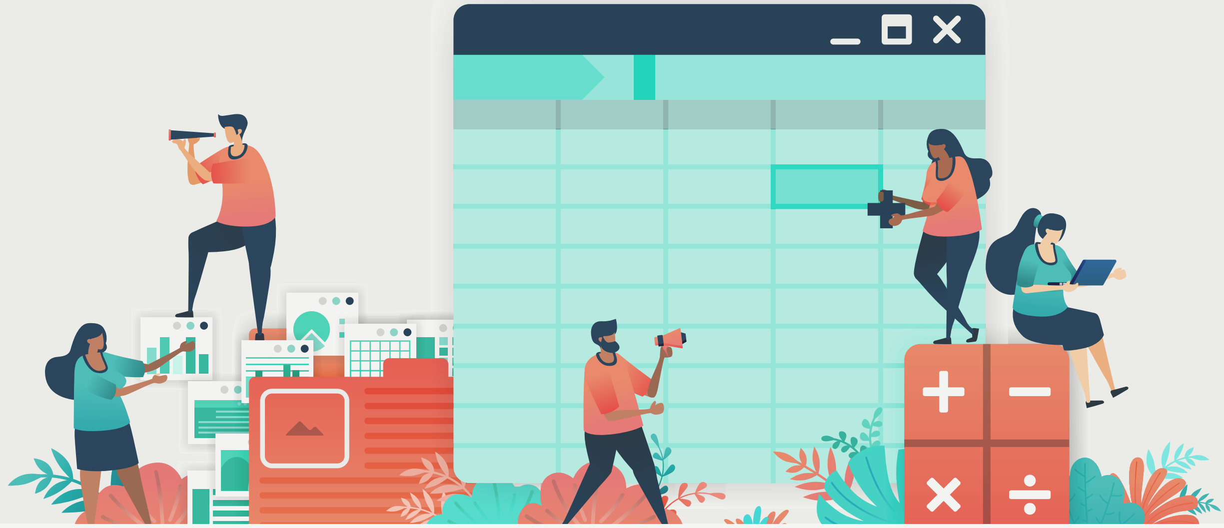
We rarely need to show percentage breakdowns. When we do, you can use a pie or donut chart, segmented bars, or square pies. We recommend using pie and donut charts sparingly, because they have a lot of shortcomings; segmented bars or square pies can be easier to read.[5,6] Percentage charts such as this one show **parts of a whole**, which always add to 100 percent. The percentages we typically highlight in our reports represent **the difference between two values**. Do not use a percentage chart to illustrate the results of head-to-head testing.

This square pie chart shows 55%.



5 EagerEyes, "Understanding Pie Charts," accessed October 26, 2020, https://eagereyes.org/pie-charts.

6 Robert Kosara, "A Reanalysis of A Study About (Square) Pie Charts from 2009," accessed October 26, 2020, https://eagereyes.org/blog/2016/a-reanalysis-of-a-study-about-square-pie-charts-from-2009. Also in Dona Wong's *The Wall Street Journal Guide to Information Graphics.*



PART 6

A summary for readers of PT reports

We want every reader of a PT report to understand how we got our results and how those results constitute wins and losses for the products we tested. In this summary, we explain our choices around identifying representative results, calculating the differences between results, and displaying those results and wins.

Identifying a representative run

Typically, for performance claims, we conduct three test runs and base the wins in the report on the results of the median run. The median is the middle value separating the greater and lesser halves of a data set. By calculating wins using the median run, we're basing the wins on actual test results.

Sometimes, we record more than one metric (e.g., throughput and response time) for each test run. In such cases, we select one of these as our primary metric to determine the median run. Because all metrics from a given test run are related, any secondary metrics we report come from the median run we have identified based on the primary metric.

Sometimes we do not conduct three test runs. When we test the amount of time and number of steps it takes to perform a management task, for example, we typically perform a single test run, because repetition could enable our tester to do it faster the second and third times. In addition, sometimes conducting three test runs doesn't make sense with the tool(s) we're using for testing.

The Science section of each report explains the testing we performed for that report.

We want every reader to understand how we got our results and how those results constitute wins and losses for the products we tested.

Shortening numbers with many decimal places

Sometimes the numbers in our results and the calculations we perform using those results have many decimal places, and we shorten those numbers for precision and/or readability.

We do not use rounding to make numbers shorter; instead, we truncate numbers by removing some of the decimal places. This is because our goal is always to present wins that are conservative and defensible, and while rounding can cause us to overstate a win, truncating ensures we never do that.

Our results fall into two primary categories: Those that individuals generate using tools such as hand timers and stopwatches, and those that test hardware or software generates.

- With hand-timed data, we must reckon with human variation: Different people hand-timing an event could come up with slightly different times. Say that a digital stopwatch delivered a result of 3.547 seconds to complete a task. Using that full number in our reports would suggest a level of precision that is not appropriate due to the variability of the human response. Therefore, we truncate such results.
- With machine-generated data, human variability is not present, so we typically do not shorten results for precision reasons. We do sometimes truncate these numbers to make them more relevant for readers.

The tables below present the guidelines we follow when truncating results and wins. As a rule, for ease of reading, we shorten numbers to a greater degree in the report than in the Science behind the report.

Results		
	Report	**Science**
Machine-generated results	For readability, truncate to one or zero decimal places	For precision reasons, truncating is usually unnecessary For readability, when the number of decimal places is greater than four, truncate to four decimal places
Hand-timed results	**If most results are < 60 seconds:** For precision reasons, truncate to one decimal place[7]	
	If most results are 1 to 60 minutes: For precision reasons, truncate to the second	
	If most results are > 1 hour: For precision reasons, truncate to the minute	

Differences between results		
	Report	**Science**
Both machine-generated and hand-timed results	For readability, truncate to one or zero decimal places	If decimal places repeat infinitely (e.g., 33.3333333…), truncate to three decimal places Otherwise, do not truncate

7 We cannot go smaller than a tenth of a second, because human reaction time is generally between one-tenth and two-tenths of a second. Source: David A. Faux and Janet Godolphin, "Manual timing in physics experiments: error and uncertainty," accessed October 26, 2020, http://epubs.surrey.ac.uk/850094/1/AJP_Faux_revision2.pdf.

Calculating differences between results

When more of something is advantageous, we use the larger-is-better formulas in the table below.

Type of calculation	Example phrasing	Formula
Percentage	"Winning product delivered 43 percent greater throughput than losing product"	$\frac{\text{larger (winning) result} - \text{smaller (losing) result}}{\text{smaller (losing) result}}$
Multiplier	"Winning product delivered 6.7 times the performance of losing product"	$\frac{\text{larger (winning) result}}{\text{smaller (losing) result}}$
	"Winning product delivered 5.7 times greater performance than losing product"	$\left(\frac{\text{larger (winning) result}}{\text{smaller (losing) result}}\right) - 1$

When less of something is advantageous (usually time or money), we use the smaller-is-better formulas in the table below. Note that in these examples, we express time differences in terms of *less time*.

Type of calculation	Example phrasing	Formula
Percentage	"Winning product completed the task in 75 percent less time than losing product" "Latency for winning product was 66 percent lower than for losing product" "Three-year cost of ownership was for winning product was 32 percent lower than for losing product"	$\frac{\text{larger (losing) result} - \text{smaller (winning) result}}{\text{larger (losing) result}}$
Fraction	"Winning product completed the task in one-third the time of losing product"	$\frac{\text{smaller (winning) result}}{\text{larger (losing) result}}$

Sometimes, we use time differences to calculate a rate and express the difference using the terms "times as fast as" or "times faster than." In these cases, we use the formulas in the table below.

Type of calculation	Example phrasing	Formula
Multiplier	"Winning product completed the task 2.7 times as fast as losing product"	$\frac{\text{larger (losing) result}}{\text{smaller (winning) result}}$
	"Winning product completed the task 1.7 times faster than losing product"	$\left(\frac{\text{larger (losing) result}}{\text{smaller (winning) result}}\right) - 1$

Normalizing results

For a variety of reasons, we occasionally normalize our results, meaning that we present our findings in relative terms by assigning the value of 1 to one result (the reference result) and showing other results as relatively greater or lesser. We use the following formula for normalization:

$$\frac{\text{result}}{\text{reference result}}$$

Comparing sound levels

When we measure the sound levels of different solutions in decibels, we compare them using an online calculator such as the one at http://www.sengpielaudio.com/calculator-levelchange.htm.

Thank you for reading.

If you have questions about any of the practices we outline in this document or any of the results or methodologies in a PT report, please email info@principledtechnologies.com.