A Guide to Lab Reports in 18-344

Matt Ngaw

October 2023

1 Introduction

The purpose of this guide is to define expectations for 18-344 lab reports.

2 Structure

The structure we hope to see in an 18-344 lab report is inspired by the typical format seen in lab reports for physics or chemistry classes.[1]. All reports flow in a similar fashion:

Define the topic of the report → Describe how you explore it → Show results → Explain and draw conclusions

The following report sections are commonly found in conventional lab reports. We do not strictly require each of the following sections in the same exact order, since different labs may call for different structures. Still, we expect your reports to flow in a fashion that is at least somewhat similar.

2.1 Title

While the title of your report in this class is not a critical feature of your report, a title provides your readers with a first-impression. Instead of a boring title like "Lab 1 Report", a title like "Redstone: An Analysis of Digital and Analog Logic in Minecraft" immediately establishes the topic and direction in the reader's mind.

2.2 Introduction

A good report must start by defining the topic at hand. It can be brief, but it should answer the following:

1. What is the topic?
2. Why is it important?
3. Briefly, what do you plan to explore in your report?

[1] How To Write A Lab Report — Indiana University Libraries

2.3 Methodology

Before showing any data to the reader, it is important that you describe to the reader how you gathered your results. Here are the key points that you must mention when describing your methods.

1. What are you testing? What did you build?
   (a) How did you implement your experiment? Diagrams or brief source-code snippets could be useful here.
   (b) Across which dimensions did you explore your experiment? You do not need to state the exact values you explored.
   (c) If your experiment is a simulation, how could you approximate hardware resources? Equations could help here.
2. What are you testing against? How did you generate your data?
   (a) Did you use a suite of benchmarks? Which benchmarks? Did you omit any benchmarks? Why?
   (b) Did you write any custom benchmarks? How do they work? What do they aim to test?
3. Are there any calculations that you use to generate data in your results section? Equations could be useful here.

2.4 Results

The results section should show the data you collected, but it is expected that any data you present be processed into a reader-friendly format. This means small tables and presentable graphs, with clear labels and titles.

2.5 Discussion

The discussion of your results can be a separate section that follows your results. Discussion can also be interleaved with your results.

This section should answer questions such as:

1. What trends did you notice?
   (a) *Why* do you think you saw these trends?
2. What trade-offs were made as you varied parameters?
3. Which parameters are "ideal"?
4. Were there trends that were you expecting which your results did not capture? *Why*?

2.6 Conclusion

3 Data

In these labs, you have the freedom to portray your data however you want. The presentation of your data should be:

1. Clear—what you are trying to show the reader about your data should be easy to grasp. This means choosing the right visualization for your data.
2. Concise—do not show all the data you collected all at once, and do not clutter your graphs with too much data.
3. Honest—do not bend the presentation of your data to mislead the reader.

3.1 Tables

Tables are best for briefly showing singular data points. In general, a data point in a table should either be a summarizing data point (e.g an average or a sum), or it should be a data point that doesn't have any related points (i.e. it doesn't make sense to graph). An example graph:

Speedup Over Single-Cycle for `gcc_s`		
Scalar Pipelined	2-Way Superscalar Pipelined	2-Way Superscalar Out-of-Order
2.1x	5.7x	12.7x

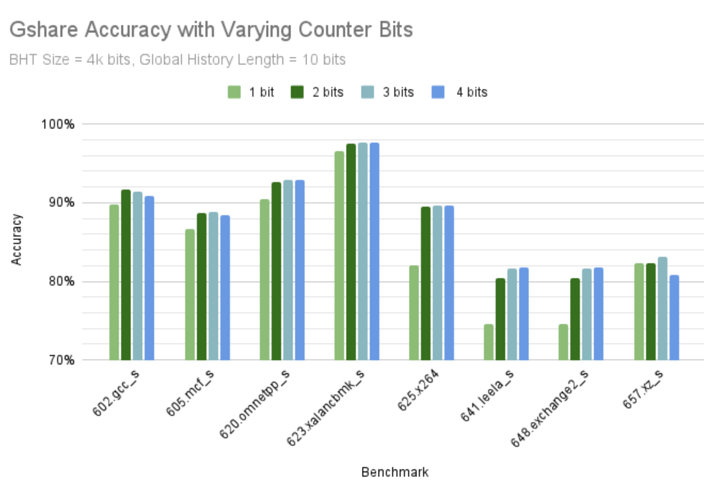
Figure 1: Virtual memory speedup resulting from TLB

3.2 Graphs

Here are some quick tips on graphing your data. The data used in the following examples should not be taken seriously.

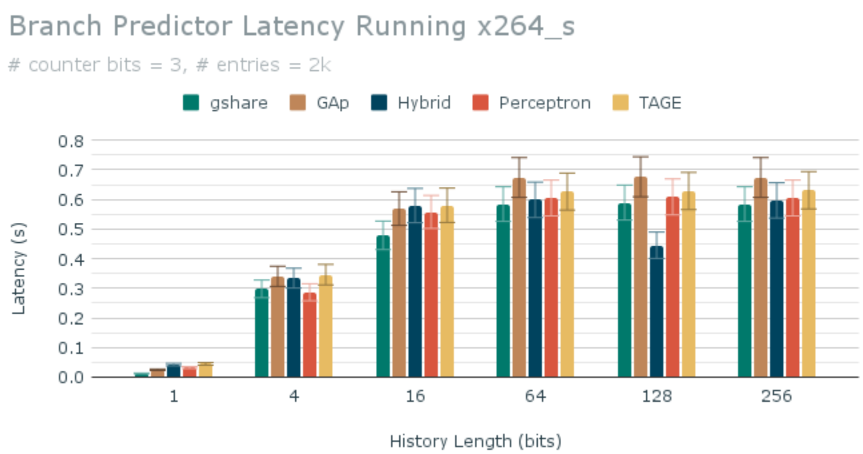
3.2.1 Bar Chart

The simplest graph to use is a bar chart. To get extra data density from a bar chart, a ”third” dimension can be added by plotting multiple data points per x-axis value and labelling the different points with a legend.



Some points to mention about this first example:

- Accuracy is on the y-axis. It is the dependent variable, the variable we want to see change as we vary any other variables.
- The benchmark is intentionally on the x-axis, rather than on the legend. We want to cluster the data points by the number of bits, to see how accuracy changes as we vary the counter bits, *per benchmark*.
  - The alternative is to cluster by benchmark. It does not make sense to cluster by benchmark, at least not while we are also varying by number of bits.
  - We anyways get a good sense of the differences between benchmarks when we compare *between* clusters (as opposed to within clusters).
- The title describes what I (the author) want you (the reader) to take away from this graph.
- The subtitle lists any assumptions made and constants held while collecting this data.

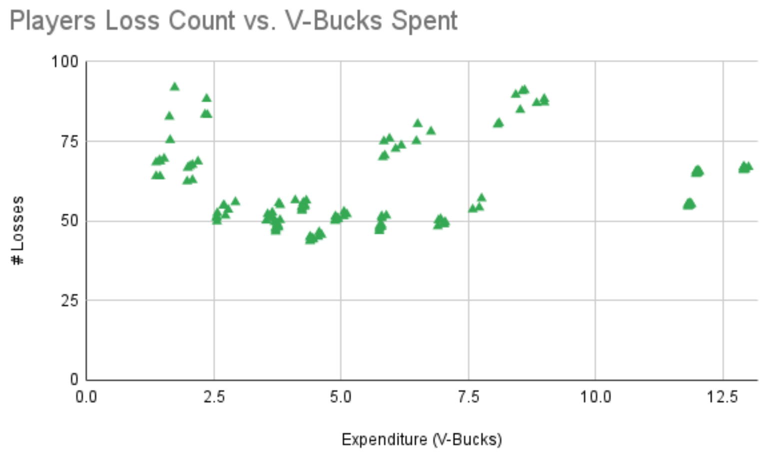


Regarding this second example:

- Latency (a made up metric for this fake graph) is our dependent variable here.
- With the branch predictor type on the legend, we get to see how latency varies as we vary the predictor type, while holding history length constant within each cluster.
- The title provides an important detail: this data *only* pertains to the benchmark x264_s.
- Again the subtitle lists assumptions/constants held.
- Notice the **units** provided with the axes labels.

3.2.2 Scatterplot

A scatterplot is useful for plotting multi-dimensional data. Furthermore, a scatterplot is ideal when trying to find a Pareto frontier.



To make a graph that has a Pareto frontier, the axes must be chosen carefully so that the two are inversely related—an increase in one axis leads to a decrease in the other. The Pareto curve describes the trade-off being made to move in either direction on the graph.

If identifying a Pareto frontier, it is expected that you identify **three** data points on this curve. Typically:

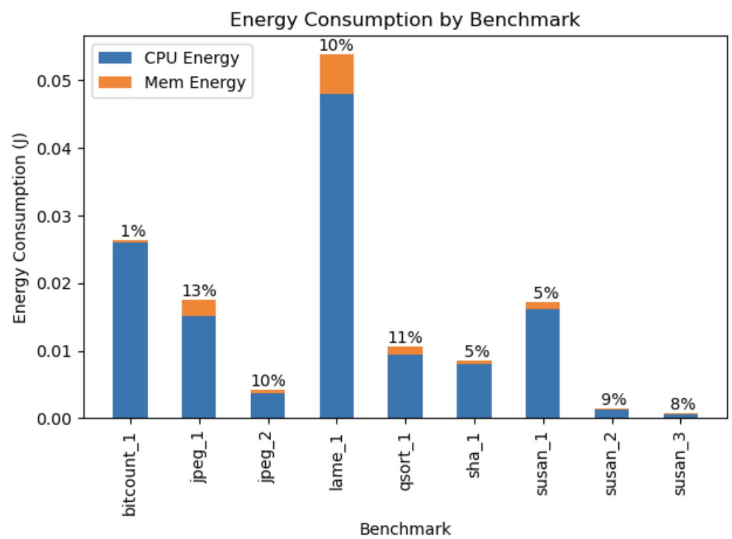
Some points to mention about this first example:

1. An extreme point favoring the x-axis. This point is all the way to the right, and on the low end relative to the y-axis.
2. An opposing point similarly chosen but for the y-axis. All the way up, and on the left side relative to the x-axis.
3. A compromising point that is somewhere in between. This is your all-around, good-for-the-general-case data point.

These special points should be marked differently on the graph (e.g. different colors). Since it is hard to read the exact values of these points, it can help to have a table describing these special points (What are their x and y values? What parameters yield these x and y values?).

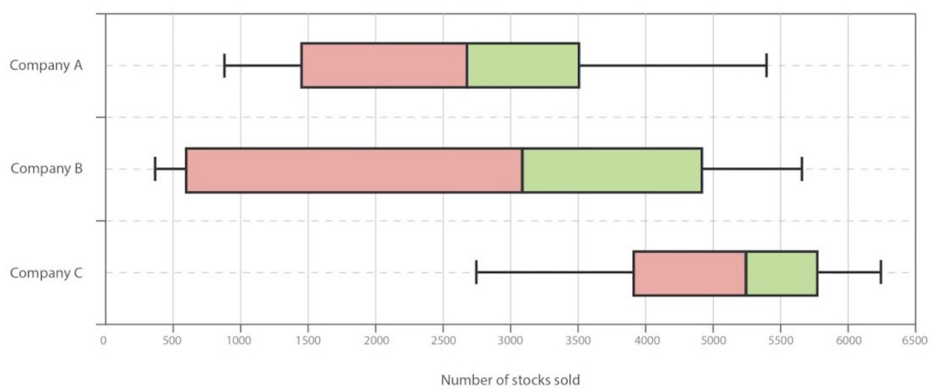
3.2.3 Stacked Bar Plot

On top of clustering bars in a normal bar graph, bars can be stacked to form a stacked bar plot. This is especially useful for data that is made up of components which sum to a whole. For example, if total energy consumption is the sum of CPU energy and memory energy, the following example plot would be appropriate.



3.2.4 Box-and-Whisker Plot

For data where it's important to give a feel of the distribution of the data, a box and whisker plot is appropriate.



Instead of summarizing data into a single data point (e.g. averaging the runtimes of multiple runs of a benchmark), it may be interesting to see the spread of the runtimes.