Tutorial DNS Analyzer

DNS Analyzer - Shiny App

This document shows how to use the Shiny app to visualize DNS data.

Sidebar

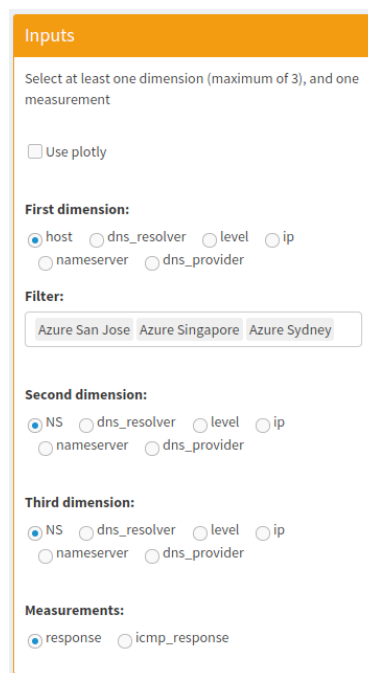On the sidebar there are two options, Plot and Data.

Data

In this panel you can check how the data looks after it was transformed and merged together. This process has to be done before deploying the app. To do that just add the csv files to the data folder and run the prepare_data.R script.

You can use the search bar to search for specific things in the data and sort by column by clicking on the column name. You can also check how many entries are by going to the bottom of the table.

Plot

This is the main panel of the app, it is where you can change inputs to generate different visualizations and see the visualizations itself, together with some statistics about the data.
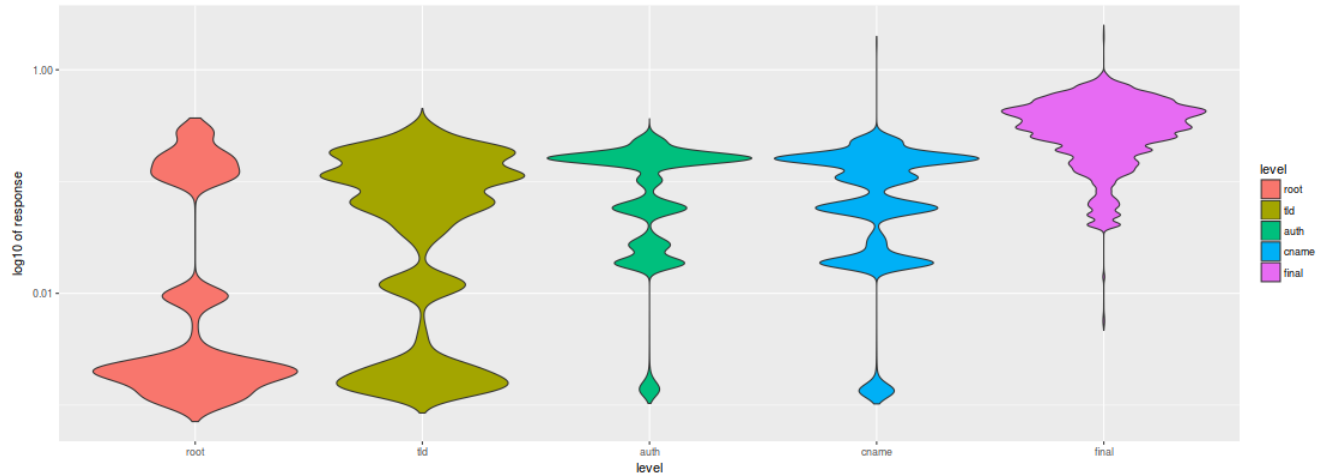
The main point of interaction here is the inputs box.

In the input box you can select dimensions to plot the data, select the value that you want to be plotted and filter the dimensions.

The first dimension is the one that will be plotted on the X axis. For example, if you choose level as first dimension you would see the following plot.

This shows the distribution of the log of the response for each level (root, tld, etc.) for all the data. You can also check the statistics for this plot by looking at the statistic box.
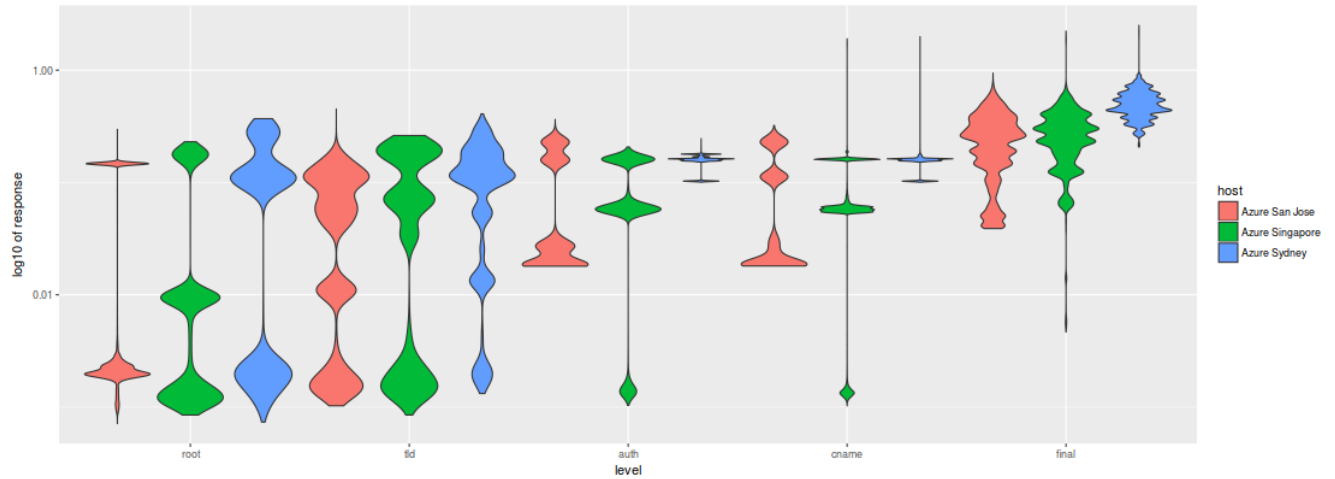

The statistics shows

- mean: the mean response (real value in this case, not the log) for the specific level
- Sd: the standard deviation of the response for the speficic level
- n: the number of data points for the specific level
- se: the standard error of the mean for the specific level
- lower.ci: the lower bound of the 95% confidence interval for the value of the mean.

2

- upper.ci: the upper bound of the 95% confidence interval for the value of the mean.

The confidence interval tells us that the real value of the mean is between the lower.ci and upper.ci with a 95% of confidence.

Looking at the data for just one dimension usually does not give us any insights, so a second dimension is normally needed for that. The second dimension will partition the data. For example, let's keep level as our first dimension and choose host as second dimension.


Now we can compare the response time for each level by each host. You can see, for example, that the Azure San Jose host appears to outperform the others on every level. And you can confirm that by look at the statistics box.

The third dimension allows you to plot multiple charts. So let's choose the DNS Resolver as the third dimension.
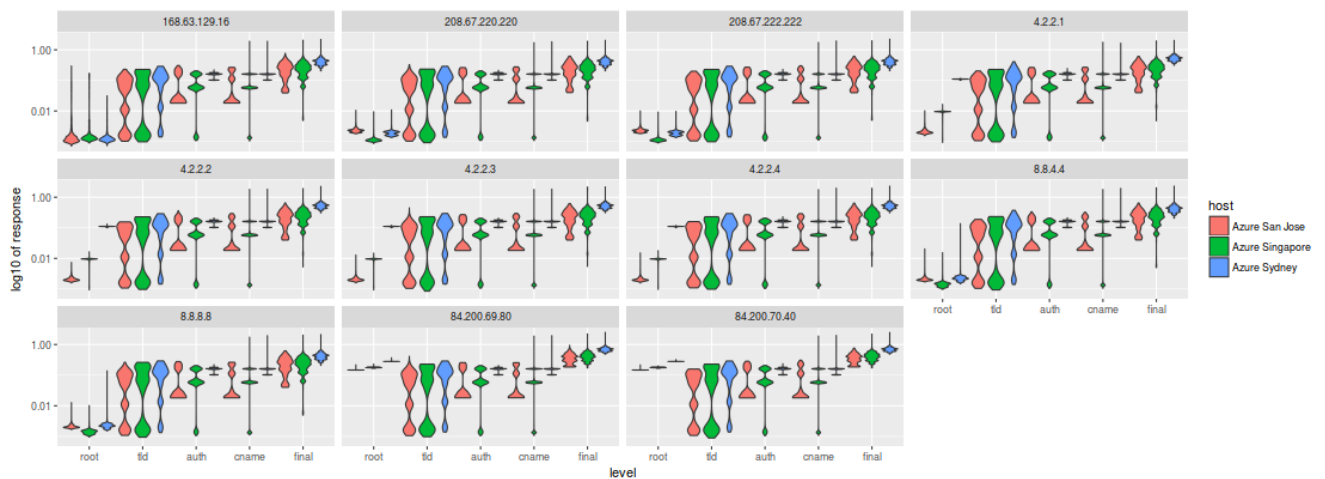
Now we have one chart showing the relation between level and host for each DNS Resolver. But as we can see there are a lot of charts and the analysis is harder. At this point we can use filters to facilitate visualizing the data. For that you go to the filter box below the dimensions and add/remove values as wanted. For example, let's remove DNS resolvers and leave only two (be aware that it might take some time for the final plot to render after playing with the filters):
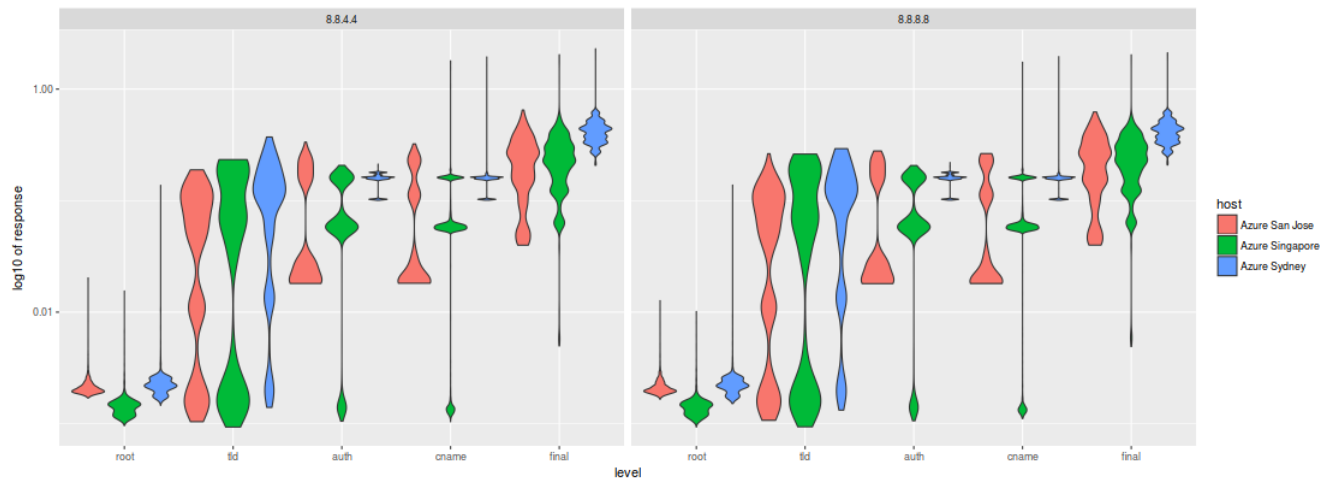
Now we can compare two DNS Resolvers, and its easier to visualize the data. You can also filter the other dimensions, and play with different combinations. For example, let's compare the root nameservers by host. First we select as first dimension level and remove the other levels using the filter box, so only root remains. Then we select nameserver as the second dimension, and host as third dimension. This would be the result:
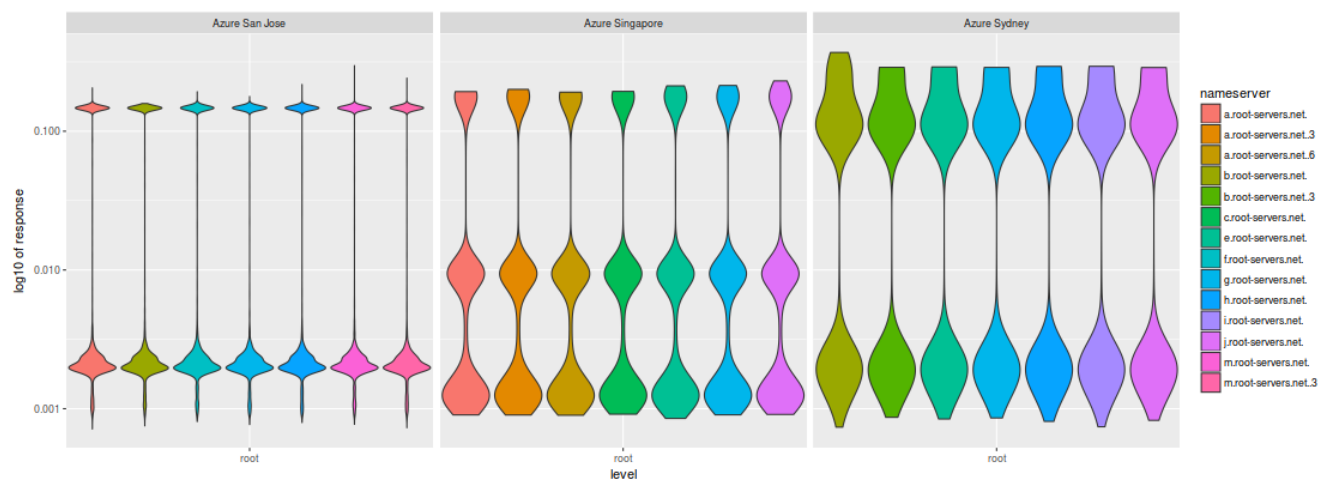
That are a lot of other different combinations that can be done just by choosing dimensions and playing with the filters, and for every combination there will be the statistics box showing the real values for the selection.