This blog looks at lyme disease through a GEO series study, GSE145974, that analyzed patients with lyme disease over two years. Patients came and went during the trial study with few dropping out or not being there for the six month blood extraction of peripheral blood mononucleic cells (PBMC) blood samples after antibiotics. There were healthy samples totalling 21, patients with acute lyme disease infection totalling 28 samples, one month after antibiotics there are 27 patients who gave their blood, and six months after antibiotics treatment there were only 10 patients who gave their blood. More details are in the NCBI gene summary and also on the published research article that donated these samples. The processing was done using Affymetrix and the human genome 37 data using expression profiling by array. The ages and genders weren't given as identifiers for these samples and so can't be compared like a previous analysis on COVID-19 using GSE152418 data. The data that was readily available did have all the platform gene identifiers which was great, but was already log2 normalized. There are various methods for log2 normalizing and I used a typical one of subtracting the min of each vector then dividing by the max-min of same vector, then taking the log base 2 of the results for the output values to be scaled between 0 and 1. The rpubs document highlights and tracks all the steps to these results. The charts were then made to see visually any results by class and gene. All images are links to the Tableau chart and can be downloaded by clicking the upper right down arrow for downloading the chart and underlying data table. But you have to have the Tableau Public Server that is free downloaded and installed on your personal computer for it to work. You can also use your work computer if you are allowed. The rpubs document is available at: <https://rpubs.com/janisharris/lymeDiseaeGSE145974> . And what files were the right size to upload into github are located at: <https://github.com/JanJanJan2018/LymeDiseaseGSE145974> . Otherwise, if the files are too large, the original NCBI access link will provide the underlying original data, and you can use the Rmarkdown file to run the chunk by chunk commands to create those other supplemented tables. There are 46,000+ genes and after aggregating by grouping by genes on mean values for each gene, it was reduced to 19,000+ genes.

The Rpubs document of mine has the links to NCBI:

"This data analysis is on lyme disease using GEO series data made readily available in its normalized state from [GSE145974](https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE145974) on ncbi.nlm.nih.gov as the accession number. The data is from the [platform GPL13667](https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GPL13667) and the [series data](https://ftp.ncbi.nlm.nih.gov/geo/series/GSE145nnn/GSE145974/matrix/). There are also some CEL/TAR files for Ubuntu but I couldn’t get my ubuntu machines to recognize it, and the instructions and tutorials for accessing the SRAtoolkit and using the Windows Ubuntu app, didn’t avail, so I am using the text files only.If you have a windows 10 tutorial on running SRAtoolkit using the ubuntu app for windows or getting the cel files to work on ubuntu with a VirtualBox disk image of ubuntu that works because you tried it within 24 hours and it worked exactly as explained, please share. I have yet to get those up and running. Possibly the new updates to virtualBox or my other apps, like docker or MongoDB or Tableau are interfering. I am not going to waste time figuring it out, its a time trap.

All the data was there as far as being filled out with values for the feature names, because I do recall exploring some platforms and series downloadable text files and only the header information was there and none of the values. The method used for processing was expression profiling by microarray on peripheral blood mononucleated cells (PBMC). The values seem to be scaled or normalized already as the values are inclusive of negative values."

<https://public.tableau.com/profile/janis5126#!/vizhome/LymeDiseaseDashboardGSE145974/LymeDiseaseDashboardGSE145974?publish=yes>

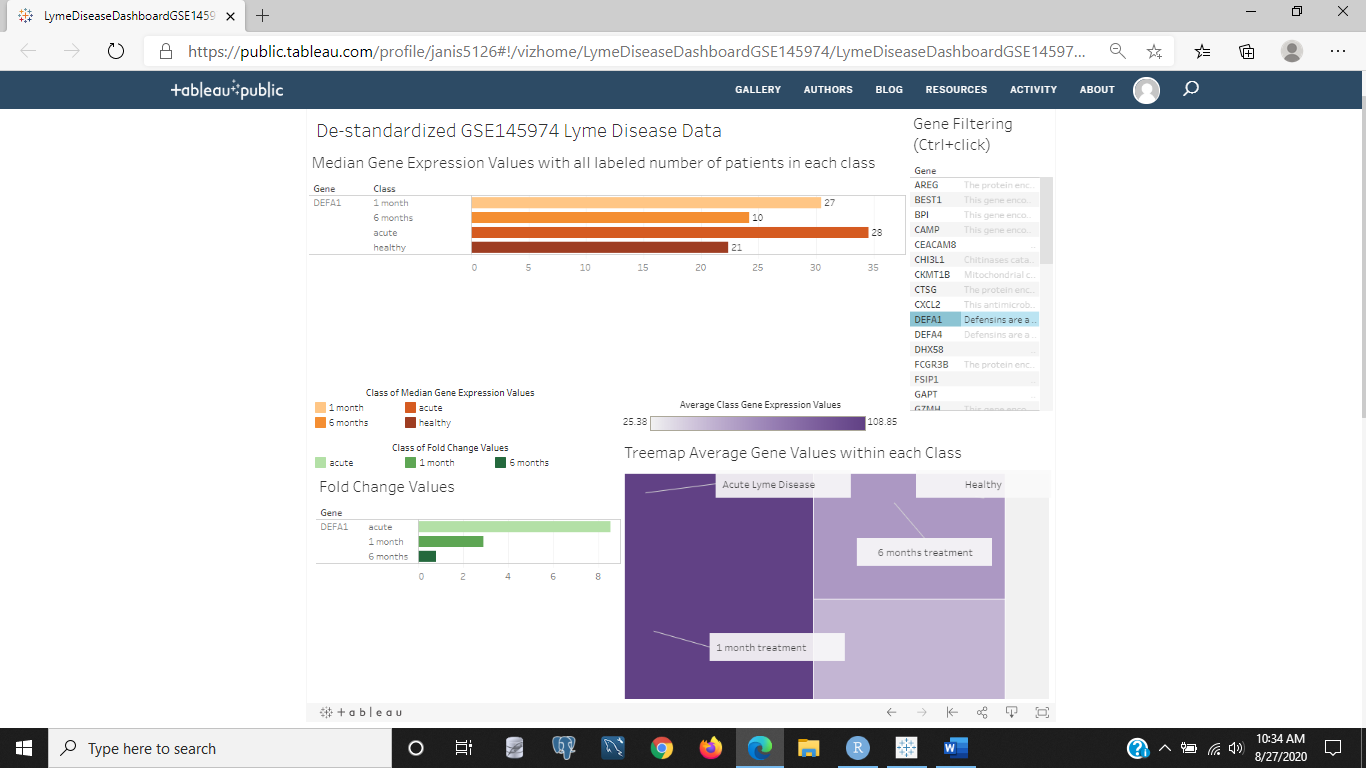
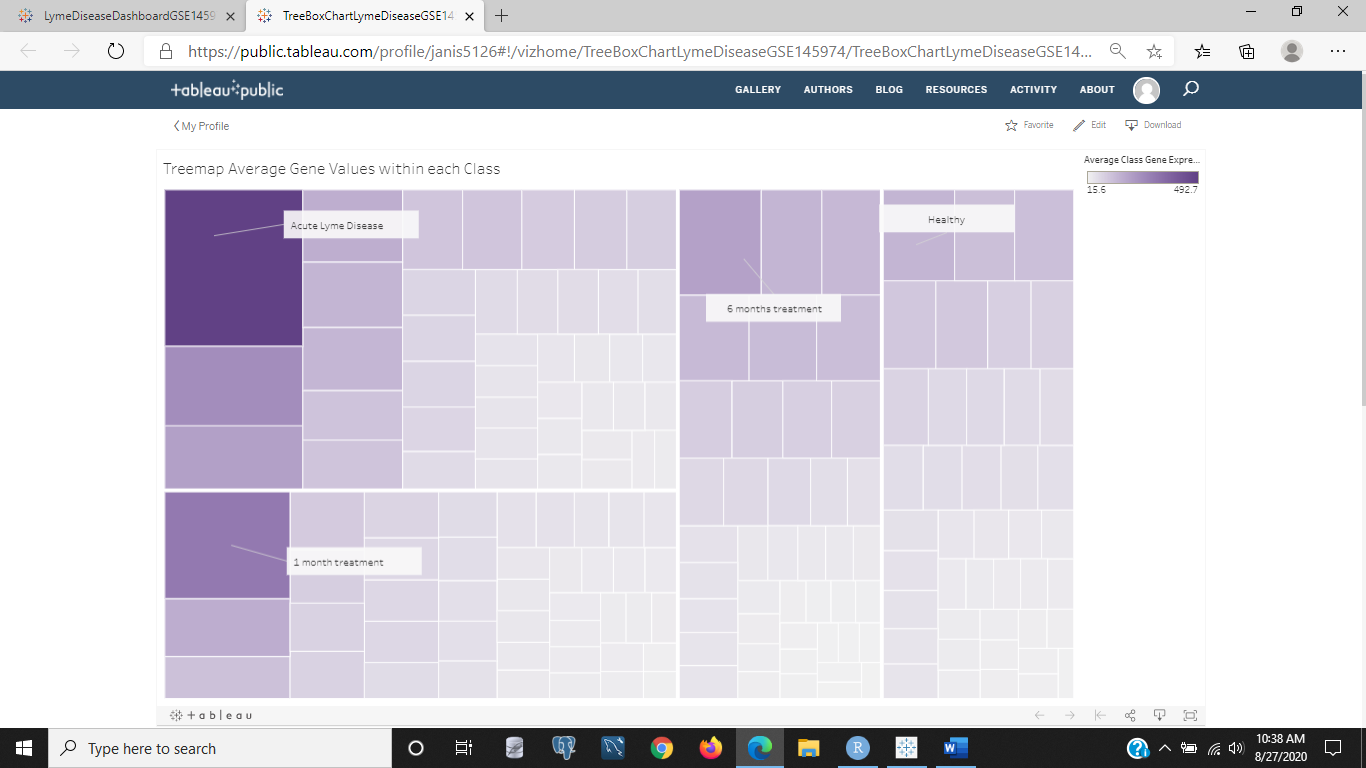
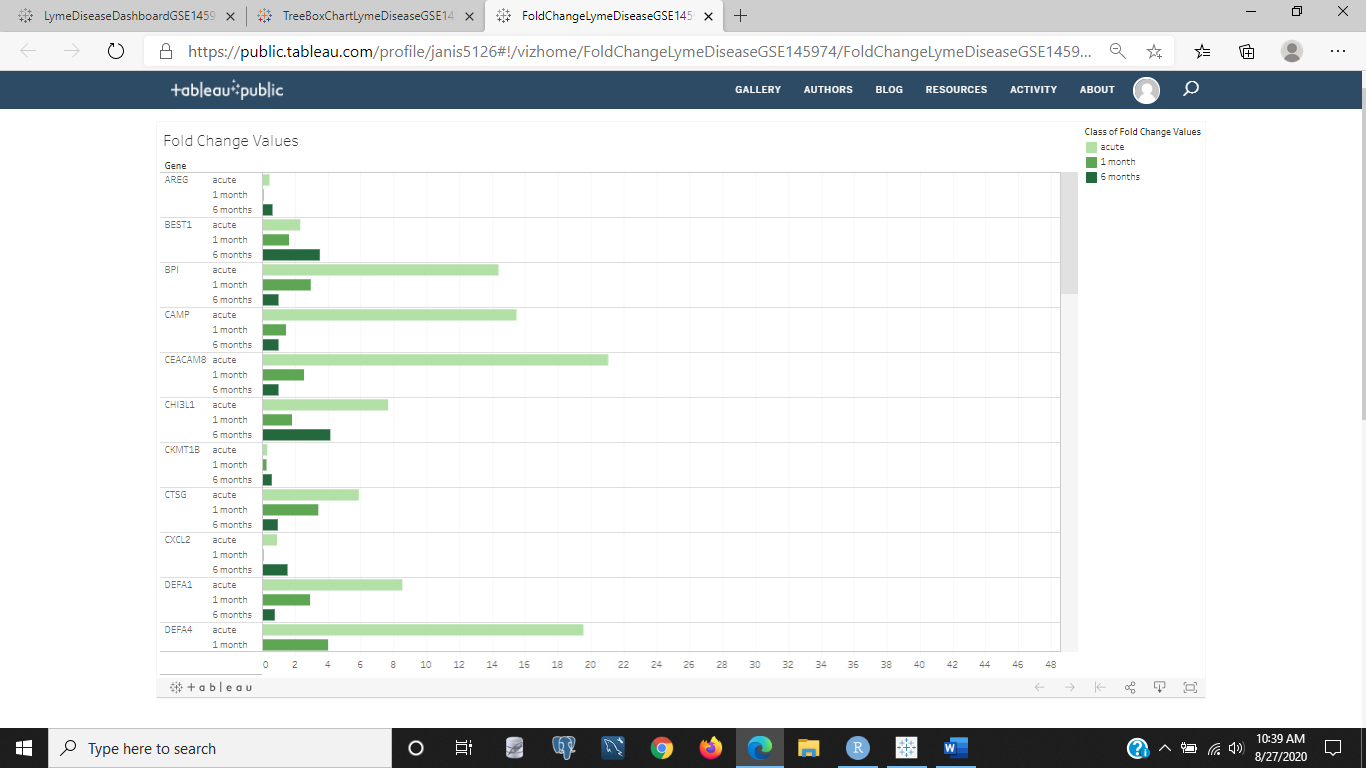


Figure 1: Dashboard showing Lyme Disease data with de-normalized data from a GEO series study, GSE145974, recently published in April of this year 2020. You can see the genes to the right with the gene summaries if you hover over the text to the right of the dashboard in the 'Gene Filtering' box. It will select only the genes you select to show the median gene expression values within each class of healthy, acute lyme disease, one month after antibiotics treatment, and six months after antibiotics treatement, with varying class sizes due to changes in patient participation and methods during the study. The top chart of the warm colors is for the median gene expression values for each gene of 43 genes that were filtered from 19,000 genes as having the most or least fold change in disease or treatment to healthy ratios for all three classes with duplicates removed from the top 10 or bottom 10 genes in each class by fold change. The lower left chart with the greens is the fold change values for each gene within each class of acute lyme disease, one month of treatment, or six months of treatment compared to healthy samples by mean values of all samples in each class. The lower right chart of the purple colors is a tree map that is categorized by class and within each class each box is a gene with the average gene expression value within that class for that gene. The upper right box shows that the gene DEFA1 was selected and it is displayed in all three accompanying charts on the dashboard.

<https://public.tableau.com/profile/janis5126#!/vizhome/TreeBoxChartLymeDiseaseGSE145974/TreeBoxChartLymeDiseaseGSE145974?publish=yes>



<https://public.tableau.com/profile/janis5126#!/vizhome/FoldChangeLymeDiseaseGSE145974/FoldChangeLymeDiseaseGSE145974?publish=yes>



<https://public.tableau.com/profile/janis5126#!/vizhome/medianClassValuesLymeDiseaseGSE145974/medianClassValuesLymeDiseaseGSE145974?publish=yes>

