* For report, might be interesting to read in multiple csv files for each data type (efficacy, laser, etc.) by binding rows into one huge csv file that’s uploaded, then run that through the functions for cleaning and plot on single plot
* Review the “log” decision – why not keep raw numbers?
* To convert from animal level to drug level: start with cfu, convert to log, then take the mean, SD, counts, etc.
* -dosing schedule: how many times per day
* For lesion pk, there should be duration column

Independent tab:

* Clustering across all drug types: how closely are all the results related to one another?
* Want to know: what is the most important feature for each drug type
* Correlation plot to see which of the features correlate, and for a particular drug class, do you have to do all of the assays or are they well correlated enough that you can do fewer assays and find out enough information?
* In-vitro distribution of drugs: in app, need ability to select which parameters to plot (as inputs to function)
* Critical to know cmax of 50 (twice a day) vs 100 (once a day) for doses
* Standard lung does not include the lesions
* 3 graphs: correlation plot, summary, cluster graph: have tabs w/in tab for independent for each plot type
* Start by inputting defaults for the graphs, then work on the radio button options after

Independent/Dependent:

* Linear model func mapped over each dependent variable
* ELU most important: with what does this value correlate the best: drug levels? in-vitro feature? Etc
* Want to know which assays are more critical for each drug class: is there a common feature that drives the efficacy?
* Not need to see drug separately, but rather entire drug class
* Penalized maximum likelihood model
* Looks for autocorrelations
* Classification (regression) tree: taking all independent variables and one dependent variable: which is most important independent variable and at what level it splits
* Random forest algorithm to split based on most important variables: repeat random selection many many times to create more sturdy model
* Higher numbers are most predictive of the variables
* Lung better with trough, spleen better with cmax