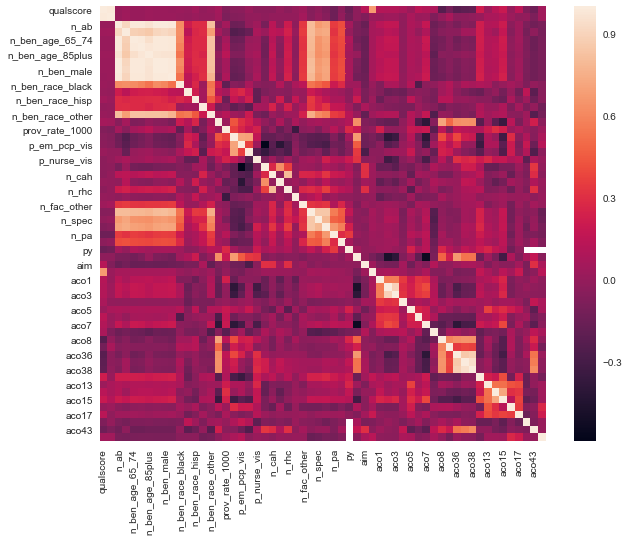
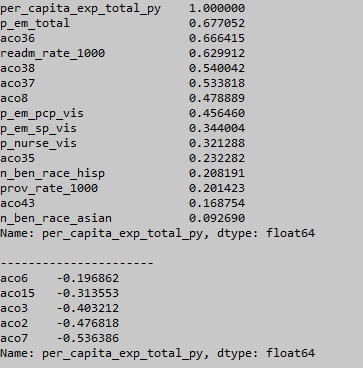


Now, we are interested to learn about the correlation behavior of numeric variables. Out of 38 variables, I presume some of them must be correlated. If found, we can later remove these correlated variables as they won't provide any useful information to the model.



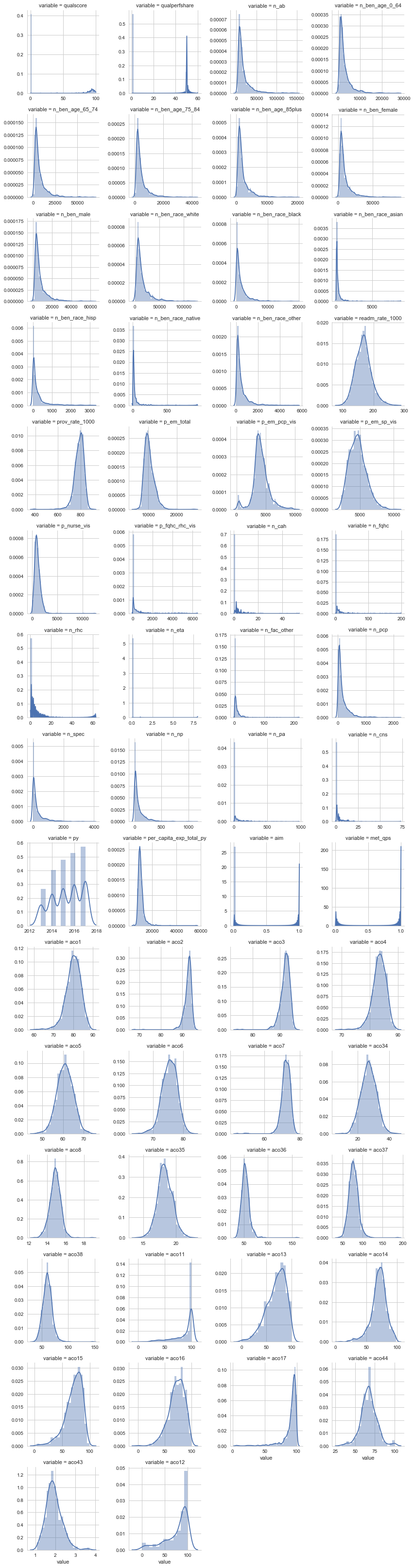
Here, a numeric correlation score will help us understand the graph better. We can see the correlation of all variables against per\_capita\_exp\_total\_py. As you can see, some variables seem to be strongly correlated with the target variable.



Here we see that the p\_em\_total feature is 67% correlated with the target variable.Total number of primary care services per 1,000 person-years in the performance year. This is because Primary care services are counted regardless of physician specialty.

Now, we'll move forward and explore categorical features.

Now, we'll define a function which calculates p values. From those p values, we'll calculate a disparity score. Higher the disparity score, better the feature in predicting.



As you can see, most of the variables are right skewed. We'll have to transform them in the next stage.