I'm using data going back to 2000 until 2018 for the MLB teams. My model is trying to predict game attendance. So far I've looked at some factors that could have a high correlation, like winning percentage of team, day game vs. night game, day of week, how a team is doing relative to the leading team in the league (gb), etc. Here is a heat map with some correlations for attendance and a few plots with lines of best fit.