**Introduction / Preface**

Wow. What an adventure. This week I began looking at the set of all predictors together instead of just picking a few. I found out some pretty interesting things while working through the project this week, but perhaps the most interesting tidbit:

**You can calculate speed from stride and cadence (guess how I found out!)**

The first time I ran through the calculations, I was getting condition numbers in the 50s/60s. This didn’t clue me into an issue until I got to the bottom and I arrived at a rather simplistic model:

**Avgspeed ~ runner + avgcadence + avgstride**

This model had an adjusted r squared of .99 which was alarming, so I googled it and found out that speed = 2 \* stride \* cadence. Who would have thought? I decided to remove both of those predictors from my model and run everything again. This time around I got a much less perfect model. Phew!

**VIF and Condition Numbers**

After correcting the issue I mentioned above, all VIFs returned were less than 3. The condition numbers were all less than 7. This tells me there are no collinearity issues present.

**AIC / BIC**

All of the models returned from the AIC selection were identical. The same result happened when I looked at BIC selection, although the increased tolerance did result in a different model compared to AIC.

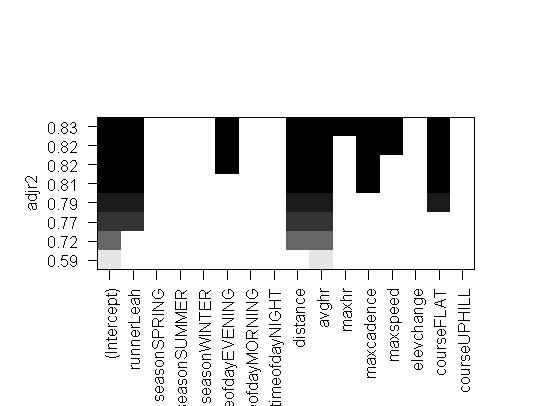
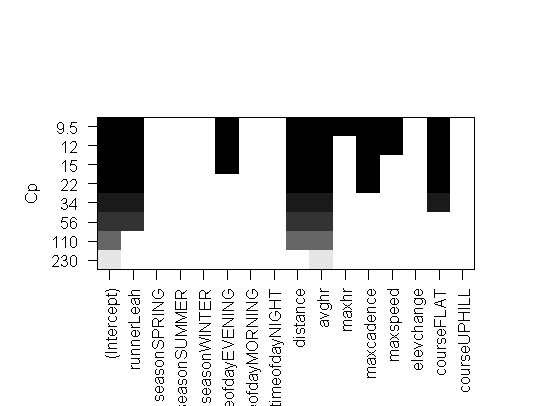
Of the two choices, I feel safer with the BIC selection process as it better protects against overfitting.

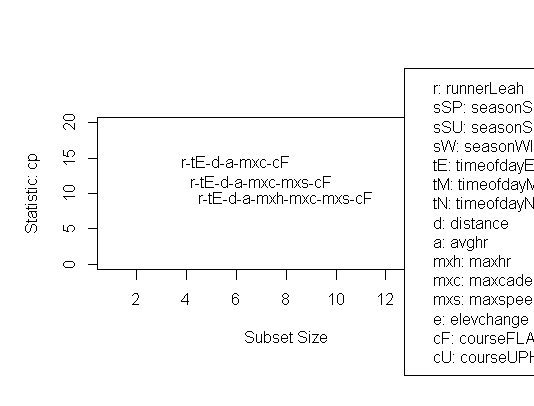
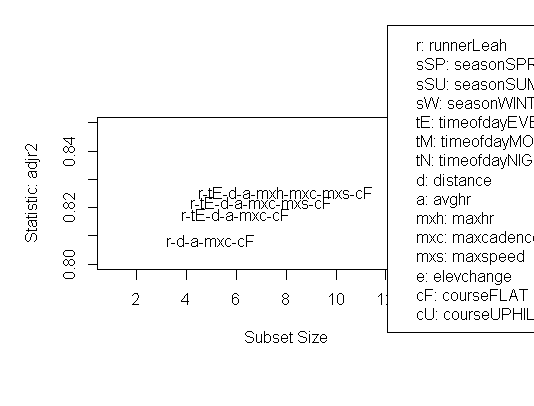
**CP / AdjR2**

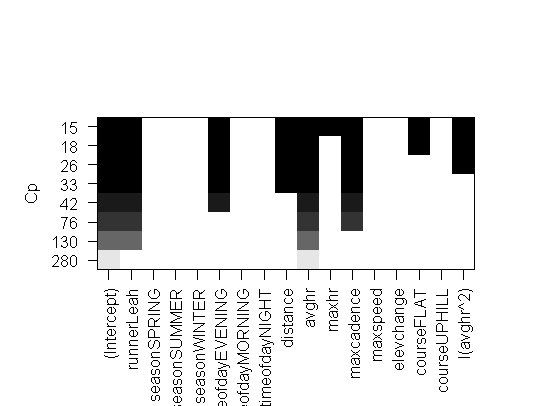
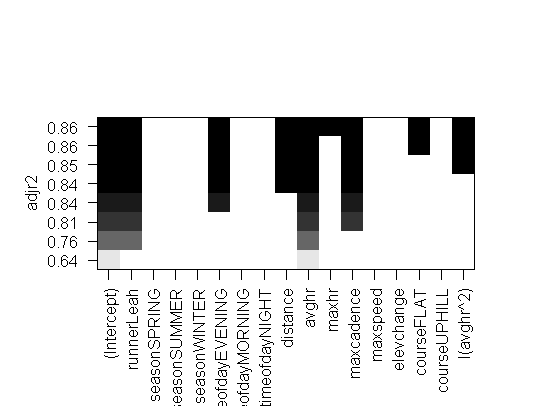
I looked at two models each from the CP/AdjR2 plots – one with the optimal CP/AdjR2 and one with minimal predictors for simplicity.

**Transformations**

In week 5 I looked at doing a logarithmic response variable and adding quadratic terms. I did reinvestigate these options here. I could have spent more time tuning this if I hadn’t waited until the last minute.

**Plots without Transformations**



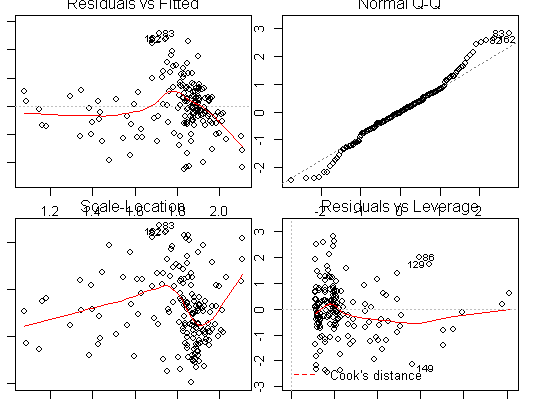


**Plots with Transformations (logarithmic response and quadratic predictor)**

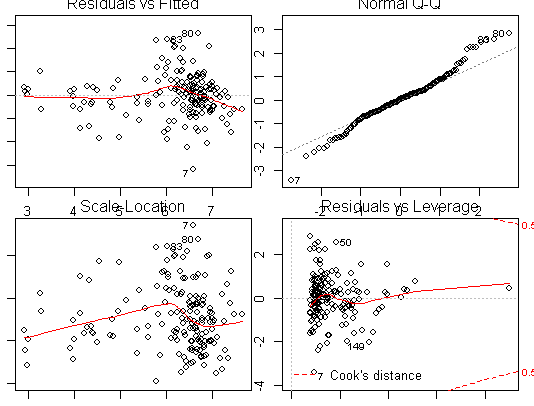
Distance seems to be less important in this new model. RunnerLeah and avghr are both still important.

**Investigating Assumptions with Models**

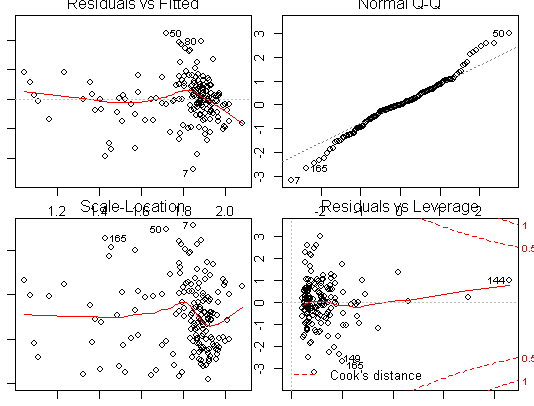
Simple Model



BIC Model



Extra Model



Of the many models investigated this week, I looked at 3 in the end: “Simple”, “BIC” and “Extra”. All of these models have some issues with the linearity assumption. They all exhibit heteroscedasticity, although the “extra” model is the least severe. Normality isn’t perfect but could be much worse.

Correcting for outliers may improve some of the issues I see.