# Testosterone, diversity, and group project performance

*Cathy Su*

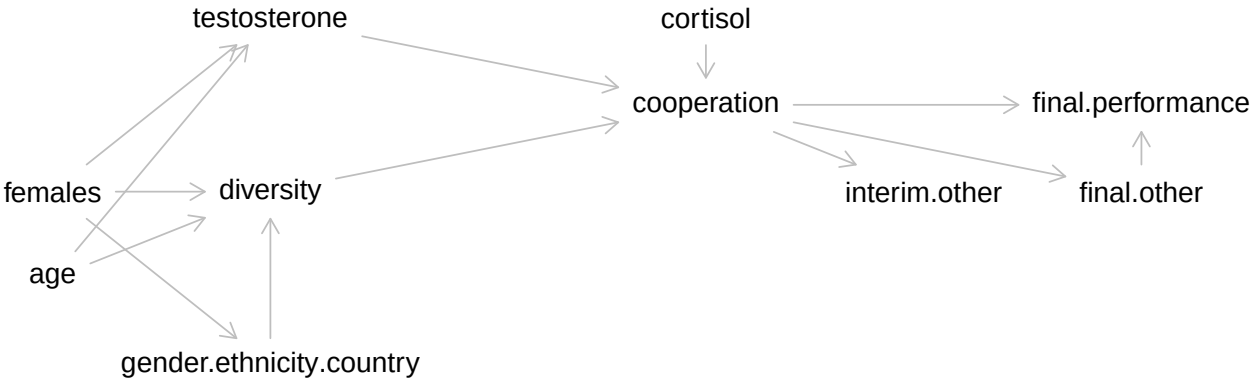
*9/10/2019*

We have both individual-level and team-level data for MBA students working on a competitive task. Our objective in the project is to further examine the relationship between testosterone and group performance, adopting the author's hypothesis that:

> diversity is beneficial for performance, but only if group-level testosterone is low; diversity has a negative effect on performance if group-level testosterone is high.

## Hypothesized causal diagram

Based on the preamble we may guess that the effects of testosterone and diversity on performance are mediated by their opposite effects on 'cooperation' in the group. Then our hypothesized causal diagram as follows:

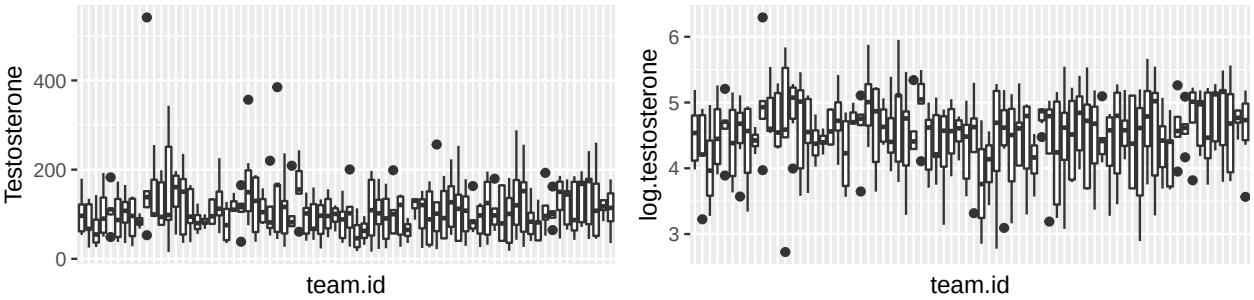



## Handle missing data

Before calculating additional team level statistics, we saw there were <10 individuals with missing data. We preferred to not remove any individuals since we are trying to look at team level performance. We checked that not everything was missing and just calculated group average measurements e.g. average hormone measurements from other members. When we did this, there were no teams with missing data aside from in the 'interim' variables.

## Log vs raw values of testosterone and cortisol

Since we are interested in how the the levels of these two hormones may be related to the other variables, we first wanted to check whether we should use their log vs raw values in our EDA.



(Cortisol looks similar to the above, except for scale.) It was clear when for both hormones that the log transformed values were distributed with less skew across teams which would be our preference.

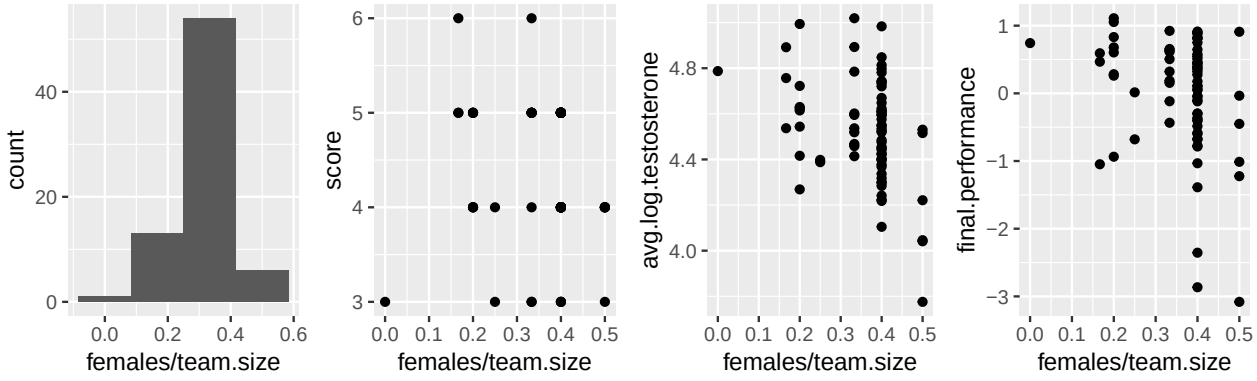
## Variable selection

To choose what other variables to include in further analysis, we visually inspected pairwise correlations. We removed the following:

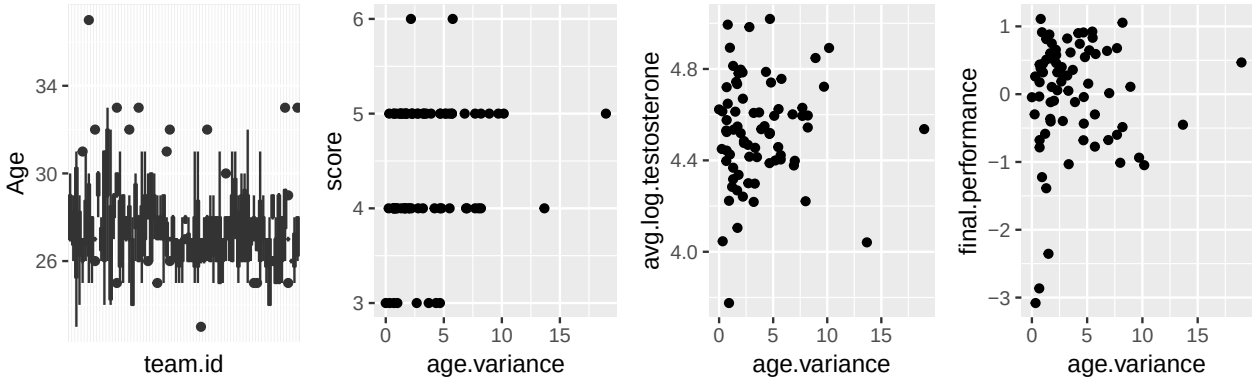
- time.of.day : didn't seem to have a straightforward relationship with other variables.
- 'interim'and other 'final' variables: seemed partly redundant with final.performance (and for former, contain missing values).

We then looked more closely at the variables 'females' and'age'.

Gender is incorporated in the diversity score, but the variable females was measured separately. Since we are looking at average hormone scores, we examined the proportion of females per group. As seen below, proportion of females is not clearly correlated with diversity score yet shows similar negative correlation with both testosterone and team.size. Our hypothesized causal graph suggested proportion of females should influence performance through both diversity and through affecting testosterone level. However since the two scatterplots on the RHS are so similar, proportion of females seems may not affect performance through diversity score but rather through their effect on testosterone.




A similar situation occurs for age. When we plotted the spread of age across groups (below left), it seemed mean age is not too different between groups. Since according to our causal graph, the effects of age may influence performance through changing cooperation, we further plotted the variance of age since different ages in a group could also lead to conflict. It also doesn't correlate clearly with diversity score but does show some positive correlation with testosterone and performance. Once again, since the two scatterplots on the RHS are so similar, variance and mean of age may not affect performance through diversity score but rather through their effect on testosterone.




Therefore, we decided to discard both 'age' and 'females' as extra variables which don't seem to tell us more about the relationship between testosterone and performance.