STAT375: Statistical Aspects of Competition **with Dr. McShane**
Final Project Proposal **Due April 23rd at 10pm** **Sunny Lin**

Context

Which sport, game, etc will you be analyzing?
I will be analyzing basketball.

Overall Proposal

What do you propose to do?
I will be looking at tracking data to determine where players of the defensive team are when they miss a shot attempt and it is a turnover (how far players are from the basket) and how that affects the next possession (whether they get "back on defense" faster). I would like to examine if player position has a significant effect on next possession as well as the outcome of the game overall (W/L). If applicable, I'd also like to examine how the position of offensive players affects turnover vs offensive rebounds ratio on their offensive attempt.

Data

What data do you plan to use? (Provide a link to the page or pages where this is contained or described). Scrape/read-in the data here. Describe the origin of the data, who collected it, how large is it (# rows, columns, tables, file size, etc as appropriate)?

Exploratory Data Analysis

Explore your data with `ggpairs` or as you think is appropriate. If using `ggpairs`, no more than 6 variables should be included at a time. `ggduo` can be used to fill in the "off-diagonal" pairs in an overall scatterplot matrix. If you have too many columns (say, over 18) to make this meaningful, select at least 12 columns to explore, including your candidate response variable or variables. This does not need to be submitted as part of your proposal, but you'll need to do this eventually (if your project is based on data).

```
## Rows: 5,533
## Columns: 13
## $ player_id  <chr> "201143", "203145", "2594", "202694", "201952", "203083"...
## $ lastname   <chr> "Horford", "Bazemore", "Korver", "Morris", "Teague", "Dr...
## $ firstname  <chr> "Al", "Kent", "Kyle", "Marcus", "Jeff", "Andre", "Kentav...
## $ jersey     <chr> "15", "24", "26", "13", "0", "0", "5", "4", "23", "1", N...
## $ position   <chr> "F-C", "G", "G", "F", "G", "C", "G", "F", "F", "G", NA, ...
## $ team_id    <dbl> 1610612737, 1610612737, 1610612737, 1610612765, 16106127...
## $ x_loc      <dbl> 24.87104, 25.51108, 28.58620, 29.75039, 33.45193, 33.695...
## $ y_loc      <dbl> 24.38831, 32.80861, 21.64312, 30.73939, 25.15548, 25.295...
## $ radius     <dbl> 0.00000, 0.00000, 0.00000, 0.00000, 0.00000, 0.00000, 0....
## $ game_clock <dbl> 717.41, 717.41, 717.41, 717.41, 717.41, 717.41, 717.41, ...
## $ shot_clock <dbl> 21.40, 21.40, 21.40, 21.40, 21.40, 21.40, 21.40, 21.40, ...
## $ quarter    <dbl> 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1,...
## $ event.id   <dbl> 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2,...
```

Methods

What statistical methods will you be using? (Simulation counts). Will you be assessing models by technique?

We will be using similar methods to those employed in decision-making analysis as well as common graph theory analysis.

US-RESP or US-CLAP

Which competition would your project be eligible for, do you think? If US-CLAP, describe which course you've seen the method in. If US-RESP, confirm that you have not taken a class that has covered that topic.

Depending on the types of questions we ask and the answers we get, we may be able to be eligible for US-RESP. Otherwise we are eligible for US-CLAP and we have seen the described methods employed in this class.

Number of group members

To adequately address this problem, how many people do you think would be needed in a group to address it? (1 to 3 group members is feasible, but again, additional people means the scope of the project needs to expand.)
Sunny Lin and Michael Pitts are working together

Would you be willing to join someone else's project?

We can add one more if somebody wants to join.

Submit this on Gradescope, individually, by Friday at 10pm