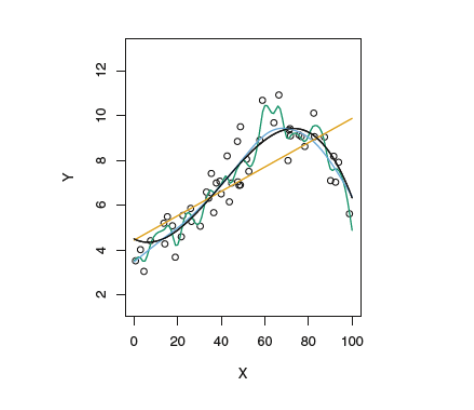
Outline

Methodology:

-Predict game attendance and ticket prices based on characteristics

General idea: we want to compare the performances of using different subset of characteristics and using different models.

We want to get the comparison figure like this:



Since the outcome variable game attendance and ticket prices are quantitative, this is a regression problem:

1. divide the data set into training sample, validation sample and test sample.

2. use the training sample to calculate the correlations between characteristics and outcome variables. Get the correlation matrix to choose characteristics.

3. try **OLS, subset selection, ridge** and **lasso** model.

4. use validation se t and test set to check their performances respectively.

Do the prices vary by team (home, visiting) record, time of year, day of week, time of day, stadium, section, etc.?

Since in this case, team record, time of year, day of week, time of day, stadium and section are categorical variables, we want to think this as a classification problem.

Similarly, we want to try **linear probability model, probit, logit, linear discriminant analysis** and **SVM** model to check if we can use ticket prices to classify these variables.

-For your favorite team, provide a recommendation on seller strategy on Stubhub, including which tickets are best to resell and when you should post your tickets.

1. choose 1 team.

2. find all ticket ids belong to this team.

3. for each ticket id, check its sales volume, price, attendance rate and post timing, etc.

4. use these characteristics, predict each ticket’s future sales volume, price and attendance rate and post timing, etc.

5. the paper we refer to use some econometrics models to predict the price and attendance rate for tickests. Thus, we want to compare our ML results with the econometrics model results.

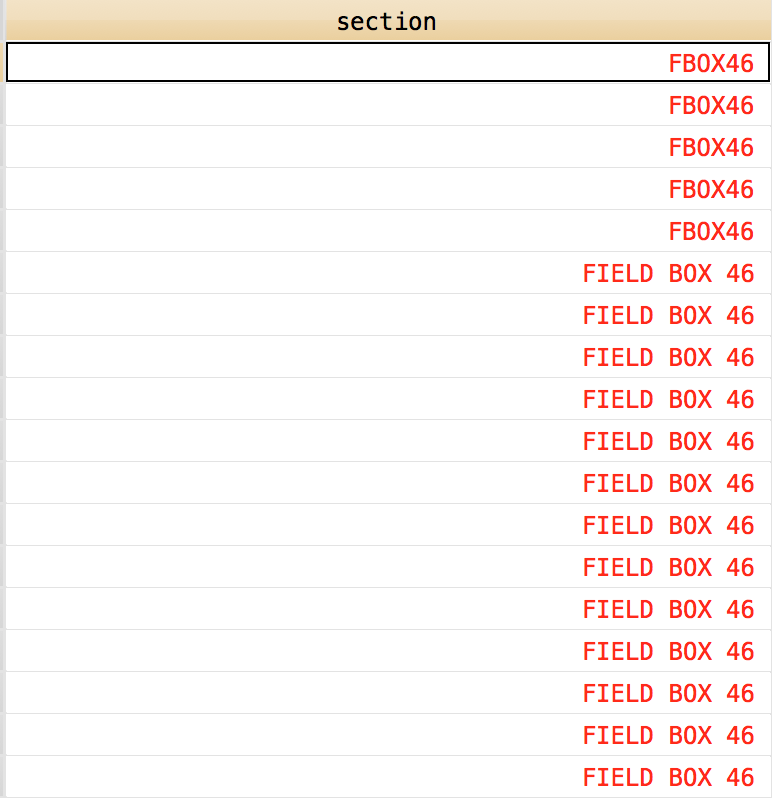
Questions:

1. could you please give us some comments and advices about the methodology?

2. In the Stubhub dataset, many variables have no label (explanation). We plan to find the meaning of these variables from the appendix. Are there better ways to do so?

3. Do we need to include literature review in our report?

4. Some variables’ type is “str”, such as section:



How can we change these variables into number? (ordering them in number seems unreasonable, because we do not know how to decide which section can be assigned to a larger number)