STATISTICAL RETHINKING WINTER 2019
HOMEWORK, WEEK 5

When/how is homework due? This assignment is due Friday February 1. You know where and how to deliver it.

1. Consider the `data(Wines2012)` data table. These data are expert ratings of 20 different French and American wines by 9 different French and American judges. Your goal is to model `score`, the subjective rating assigned by each judge to each wine. I recommend standardizing it.

In this first problem, consider only variation among judges and wines. Construct index variables of `judge` and `wine` and then use these index variables to construct a linear regression model. Justify your priors. You should end up with 9 judge parameters and 20 wine parameters. Use `ulam` instead of `quap` to build this model, and be sure to check the chains for convergence. If you'd rather build the model directly in Stan or PyMC3, go ahead. I just want you to use Hamiltonian Monte Carlo instead of quadratic approximation.

How do you interpret the variation among individual judges and individual wines? Do you notice any patterns, just by plotting the differences? Which judges gave the highest/lowest ratings? Which wines were rated worst/best on average?

2. Now consider three features of the wines and judges:
- (1) `flight`: Whether the wine is red or white.
- (2) `wine.amer`: Indicator variable for American wines.
- (3) `judge.amer`: Indicator variable for American judges.

Use indicator or index variables to model the influence of these features on the scores. Omit the individual judge and wine index variables from Problem 1. Do not include interaction effects yet. Again use `ulam`, justify your priors, and be sure to check the chains. What do you conclude about the differences among the wines and judges? Try to relate the results to the inferences in Problem 1.

3. Now consider two-way interactions among the three features. You should end up with three different interaction terms in your model. These will be easier to build, if you use indicator variables. Again use `ulam`, justify your priors, and be sure to check the chains. Explain what each interaction means. Be sure to interpret the model's predictions on the outcome scale (`mu`, the expected score), not on the scale of individual parameters. You can use `link` to help with this, or just use your knowledge of the linear model instead.

What do you conclude about the features and the scores? Can you relate the results of your model(s) to the individual judge and wine inferences from Problem 1?