

Your Name: _____

Names of people you worked with: _____

Instructions: Work on this problem in class with your group. Do your best. This piece of paper will be collected during class.

Task: Consider a dataset which contains a predictor variable (among many other predictor variables) measuring the date.

1. Name as many variables as you can that you could create from a date variable.
2. What are some concerns that might arise from modeling any/all of the variables you suggested?

Solution:

1.
 - season
 - weekday / weekend
 - day of week
 - month
 - holiday / not
 - year
2. Some of the variables might not vary (at all!) across the observations.
Also, the variables might not do a great job of predicting, and if we create a zillion new variables, we might be tempted to use a variable that shouldn't be used in the final model (which describes the population).