1. Please walk us through how you cleaned up this dataset. Do you see any interesting trends?
   1. clinical.csv: first I assessed the data types for each column.
   2. In some cases R loaded features in as a data type that was not appropriate, so I changed it. A couple examples were with ID and Radiation columns. Both were loaded in as dbls so I changed them to factors.
   3. I also change the name of T to “Tumor.Stage” because T is a built-in function name in R
   4. I then assessed the range and/or unique values of each column to make sure there were no typos or unexpected value and I cleaned the values where necessary
      1. For example, in the Grade column, 9 was used to denote no grade assessment so I changed 9 to NA.

2. Tell us how you decided which features to use for your model. Are there any new features

it might be productive to engineer from the current features?

3. Which algorithm did you use for the prediction and why?

4. How did you assess the predictive model’s quality? Summarize your findings.

5. Next steps? What might you do with more time or access to additional data or expertise?