Please answer the following questions by analyzing the associated datasets. For all tests, please:

* check whether the data meet the requirements/assumptions of the test you plan to run
* complete any transforms needed to make the data meet the required assumptions
* run the test
* interpret the results, do NOT only include the R output.
* check model fit in the case of linear regressions and/or glms
* if you have the option between running a linear model with a transformed y variable or a glm, choose the linear model with a transformed y variable. only run a glm when you have to.

Provide all answers in R or R markdown (similar to the take home quiz 4). Exams are due at noon on Sunday, April 23. Use the following scripts to load the datasets. The dataset to be used for each question is provided in bold at the end of the question.

Dataset If you download the datasets manually and load them, please use the following notation. Alternatively Oscar will host these datasets on GitHub soon so you can read them in directly like you did for the last exam.

flying = read.csv('/Users/mehajain/Desktop/flying.csv')

college = read.csv('/Users/mehajain/Desktop/college.csv')

happy = read.csv('/Users/mehajain/Desktop/happy.csv')

cancer = read.csv('/Users/mehajain/Desktop/cancer.csv')

1. Is there a significant association between gender (gender) and whether people think it’s rude to bring an unruly child on the plane (unruly\_child)? If yes, which gender tend to think bring an unruly child is more rude. **flying**
2. Is there a significant difference in tuition (tuition) by type of institution (type)? If yes, which type has higher tuition? **college**
3. Is there a significant difference in happiness (Hscore) by region (Region)? **happy**
4. What factors are significantly associated with a country’s corruption levels (Corruption)? Choose three continuous variables to include in your model. **happy**
5. Choose one of the continuous variables that was significant in the model for Question 4 and interact it with region (Region) to predict corruption (Corruption). This model should only include one continuous variable and its interaction with region. Does the influence of your continuous variable on corruption vary by region? If yes, how do you interpret this interaction? **happy**
6. Which factors are significantly associated with whether a breast cancer tumor is malignant or not? Choose three continuous variables to include in your model. [Bonus] What is the “probability” of getting a malignant tumor in your model when all factors are zero? **cancer**
7. BONUS/EXTRA CREDIT: Which variables are the most important in explaining whether a breast cancer tumor is malignant or not? Use the same 3 continuous variables you chose for question 6. **cancer.**