***Ladesha Jenkins***

***Aggressive Behavior in Children***

**Introduction**

Children learn, adapt, and take on behaviors that are taught from the many influences that he/she may encounter. The number of influences that a person has is not limited to the family members in one’s house hold. A child’s many influence can include, but are not limited to, immediate family, friends, school, neighbors, celebrities, and the biggest of all the media. The media would be considered as all forms of communication outlets like the internet, social media, news media, print media, and in specific to this project video viewing and electronic games. The media can offer good and bad influences for all viewers watching it. It is a parent’s duty to protect the child from negative influences as much as possible to ensure healthy growth and implement good behavior.

The media can introduce children to early exposure of media violence. Early exposure to media violence is a contributing factor to the development of aggression. The aggression then could lead to aggressive and violent behavior in children. Children are increasingly becoming heavy media consumers. Research indicates that much of the media directed at children contains violent content. Early exposure to media violence could place both male and female children at risk for the development of aggressive and violent behavior in adulthood. Children that watch shows that have violence that is realistic, consistently repeated and unpunished, are more likely to imitate what they see. Children with emotional, behavioral, learning or impulse control problems may be more easily influenced by video aggressiveness. The impact of TV violence may show immediately in the child's behavior or may surface years later. Due to the media presenting violence at a very high level, scientists have found ways parents can protect children from excessive media violence. Each day there are multiple things and people that a child would observe where behaviors can be learned from. These multiple things would be considered as the factors of independent variables that affect behavior in children. In this project, the factors that are tested that affect behavior in children are electronics, siblings (older and younger), parental approach, and nutrition. These factors tend to be the major affects in children because these typically are things and people that the child would encounter every day. The examples set by adults, older siblings and children are the most powerful influences shaping a child's behavior and personality. Electronics a nutrition would be the next major influences on children behavior because electronic offer exposure to all types of human behavior from the negative to positive behavior. The elements or ingredients of a child's diet like thiamin treatment reverses aggressiveness in thiamin-deficient children, and sucrose and additives are suspected to induce hyperactivity.

Of all types of behavior a child can display, the focus of the project would be on aggressive behavior. The level of aggressiveness would be determined according to the different circumstances of the household that the child lives in. it is know that all people's circumstances are different or have their own uniqueness. At a general level, realistic scenarios or cases that are close to what a child's family can fit in are projected according to the different levels of the project. Being that children would be the main focus, parents would be the audience of this project. The purpose of the project would be to predict the level of aggression with these given levels according to each factor or independent variable.

**Research Question**

What are the most significant factors that affect aggressive behavior in children?

**Data frame Original**

Kaggle user, Jonathan A., has presented data used for this research on kaggle.com. The dataset is titled, “Aggression in children with older children.” The original purpose behind this data was to determine how these potential factors affected the aggression in children that has older siblings in the household. The dataset provides information to project the aggression level of child according to the levels of each independent variable combined into one. There includes 334 children that are from the United Kingdom, and there are 333 children that are from the United States. Give that there is one child more in the United Kingdom than in the United States there allows room for error because it is not the same amount for both countries. The factors being used to evaluate aggression in children were Parental Approach, Electronic Games, Video Viewing, Nutrition, and Sibling Aggression. The values of each factor are described in terms of what the values represented numerically. High score for parental approach represents poor parenting. High scores for electronic games represents more time spent playing electronic games. High scores for video viewing represents more time spent watching various mediums of videos. High scores in nutrition represents a healthy diet. High scores in Sibling Aggression represents more aggression seen by older siblings. It has been assumed that high scores in aggression results in more aggression noticed in the child’s behavior. Originally, the approach was geared to the effect of television on violent behavior in children. This was the original idea without the dataset, but with the dataset the spectrum of television could be included with the variable of video viewing. With the new variable presented in the dataset it has stretched and strengthened the data to be more reliable because aggressiveness is not only caused by one factor but a multitude of factors.

**Data Wrangling Techniques**

The original dataset was transformed into a clean dataset to be used for the research. In making the dataset clean the necessary packages for the cleaning had to be installed and loaded if they had not been. The packages that were loaded were plyr, dplyr, tidyr, and readxl. Next, the data has then been read and called a new file. The names of the columns in the original dataset had been changed into a cleaner form. AggUk was changed to Aggression. VideoUK was changed to Videos. ElectUK was changed to Electronig games. SibUK was changed to Siblings. NutUK was changed to Nutrition. ParUK was changed to Parental Approach. The names of the columns had to become broader to fit the data of both the United States and the United Kingdom. Lastly, another column was added to give value to the column that contained all factored variables to make it numeric. The numeric values will allow the United States and the United Kingdom to be compared for the charts and graphs. The United States was given the value of zero, while the United Kingdom was given the value of one.

**Statistical Analysis**

It has been interesting to learn about how children work. This is a definite topic that is one to relate for all people. Children are the most interesting individuals to learn about in my opinion because things learned about children can be used to understand why adults act as they do. The most important factor to consider in this project was the surroundings or the household of each child that was tested in the dataset. It is important to consider a child’s household and living situation when referring to the behavior of the child. The household is the first place where all children learn the foundation of behavior and of life in general. Different household up-bringing results in different children behavior. For a child there has to be different avenues of influences that will encourage both good and bad behavior, so that one can be introduced to both. Typically it is up to the individual child to decide between which behavior he/she will decide to partake in. The early exposure can also help a child in determining which behaviors they see to be good or bad, healthy or unhealthy. The aggressiveness of the child will be strictly determined by what factors they have around that have influenced them to be either very aggressive or less aggressive. It is believed that all children demonstrate some level of aggressiveness but according to these findings some children have not expressed any level of aggressiveness yet.

Aggression has been measured according to the factors used. The factors are videos, electronic games, parental approach, nutrition, and siblings. The child’s projected aggression level has been counted for after the different factors has been measured. The range of aggression of the tested children include and is between -1.3 and 1.8. The highest possible value for aggression (1.8) means a child displays the most amount of aggression. Also the number of children tested in each country can be accounted for.

In comparison, the highest level of aggression is to a child belonging to the United States, and the lowest level of aggression belongs to a child of the United Kingdom. Also the most levels of aggression is centered on the different factors being used.

A box plot has been made to compare the aggression levels of the Unites States and the United Kingdom into one graph. In the graph the box plot for the United Kingdom is slightly higher than the one for the United States.

Each variable can be compared for each country due to each variable having numerical values. Most importantly each variable can be compared for each country due to each child being tested in each of the variables.

Multiple scatterplots have been made to compare each viable to the level of aggression. In the scatterplots there are colors to distinguish the countries for the levels of regression.

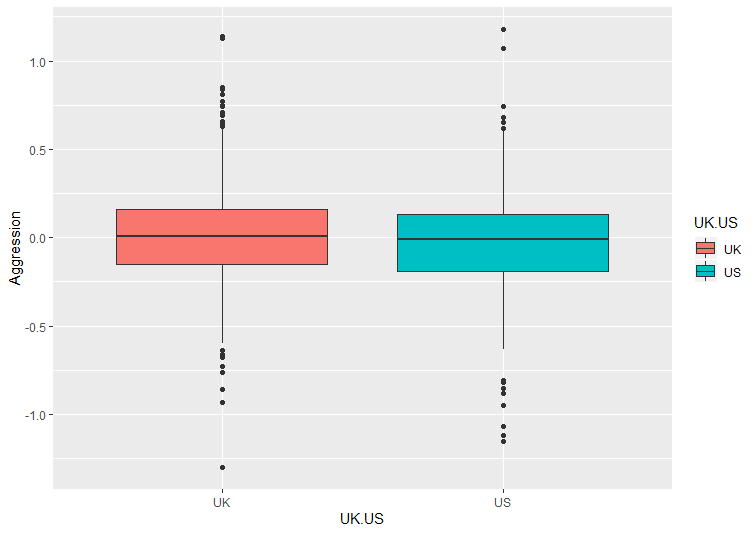
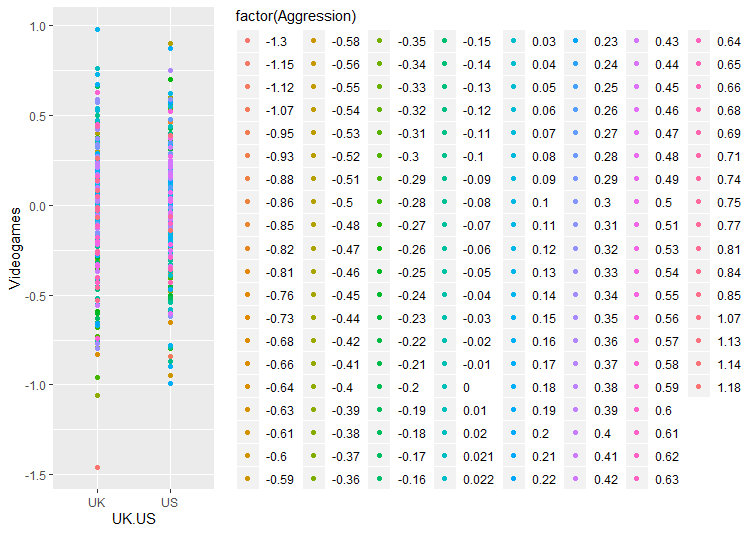
A time series plot cannot be made with the data that I have presented because there is no time involved in the data.

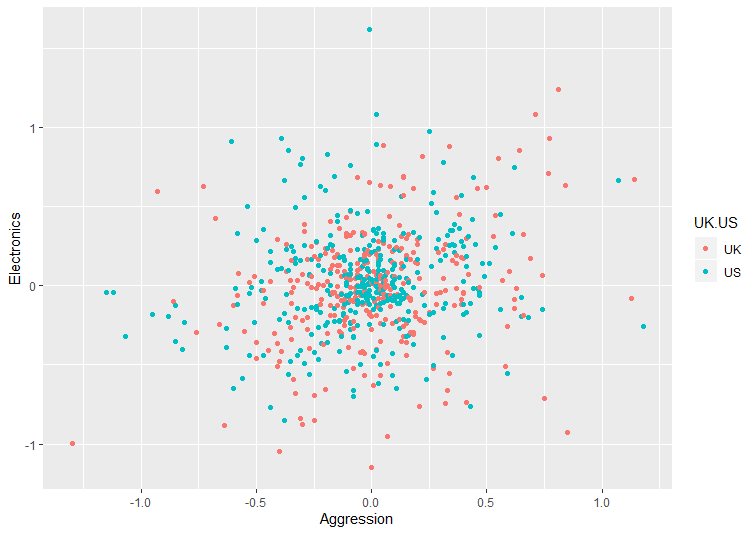
It has been interesting to find the different levels of aggression that has been presented according to the levels of each factor. The levels of aggression in both the United States and the United Kingdom has varied form very high levels to very low levels. The high levels include the positive values, and the low levels include the negative values.

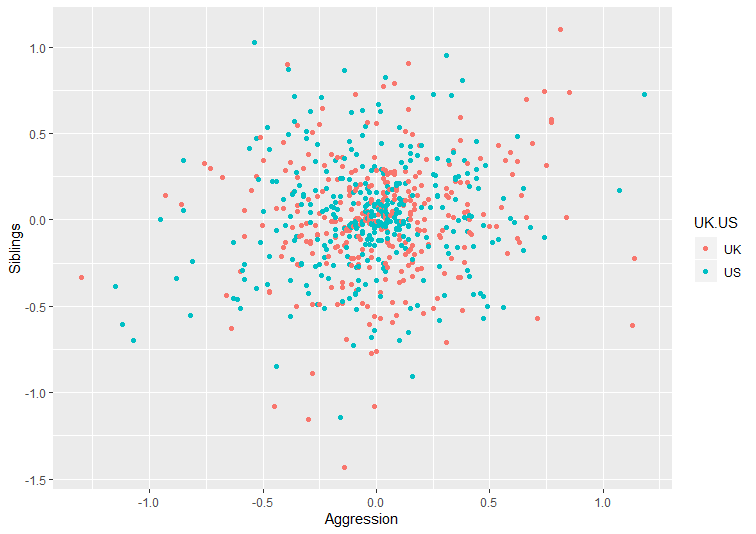
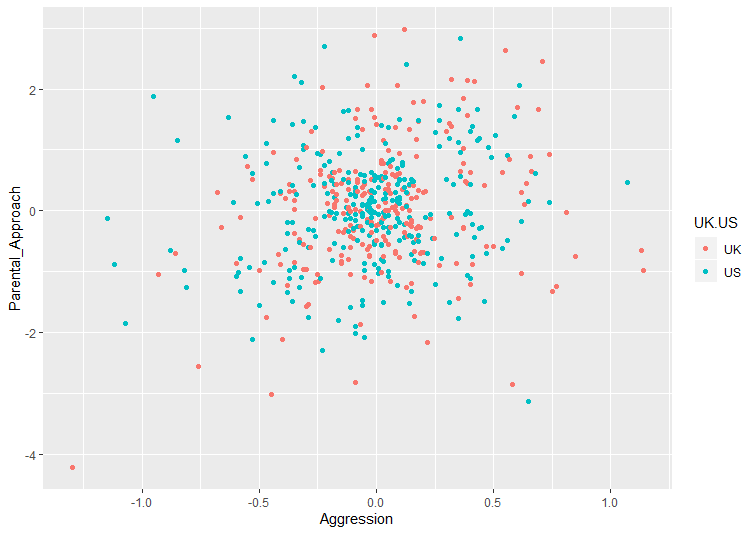
After a box plot has been created to compare the aggression levels in the United States and the United Kingdom, it has been discovered that the United Kingdom has higher levels of aggression in children than in the United States. The box plot for the United Kingdom is slightly higher than the box plot for the United States. Later it has been proven that once the mean of aggression has been taken for each country the United Kingdom’s mean is higher than the mean for the United States.

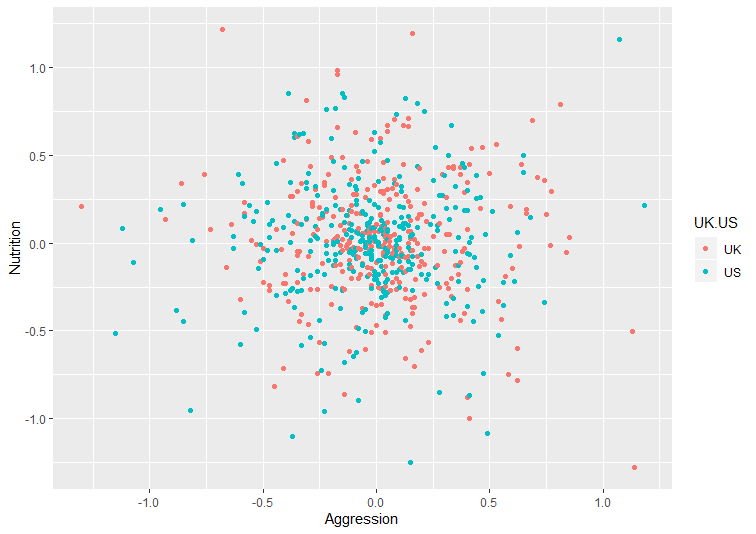
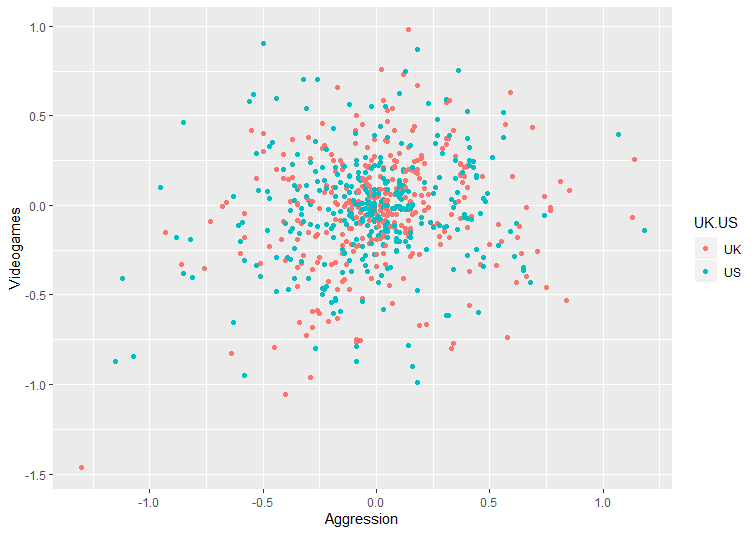
All scatter plots that were made was used to compare each independent variable to the dependent variable. Also when using and making observation of the scatter plot one can tell the relationships between the x- axis variable and the y-axis variable. The independent variables include each factor: Nutrition, Parental Approach, Siblings, Electronic Games, and Video. The dependent variable is Aggression. There were five scatter plot graphs made to compare each independent variable with the independent variable. In the first scatter plot, electronic games was compared to aggression, and it has a positive linear association between the two variables. In the second scatter plot, videos was compared to aggression, and it has a positive linear association between the two models. In the third scatter plot, siblings was compared to aggression, and it has a positive linear association between the two variables. The third scatter plot seems to have the highest positive linear slope due to the points on the scatter plot spreading from one side of the graph to the other. Also, it doesn’t have as many points focused or centered in the middle of the plot. In the fourth scatter plot, nutrition was compared to aggression, and it is hard to decipher whether or not it has a negative or positive linear association or no association between the two variables. If the scatter plot included a linear slope, it could possibly be a negative slope. In the final scatter plot, parental approach was compared to aggression, and it has a positive linear association between the two variables.

When using the summary function of all the variables in the dataset, it gives an overall of each variable. It has shown that the highest level of aggression with a score of 1.18 was in the United States, and the lowest level of aggression with a score of -1.3 was in the United Kingdom. The highest and lowest levels can be described as the minimum and the maximum. The minimum of the variable nutrition is -1.28, and the maximum is 1.22. The minimum of the variable parental approach is -4.46, and the maximum is 3.99. The minimum of the variable siblings is -1.43, and the maximum is 1.1. The minimum of the variable electronic games is -1.15, and the maximum is 1.62. The minimum of the variable video viewing is -1.46, and the maximum is 0.48. Based on each variable separately, the minimum of each variable was found in the United Kingdom except for the variable parental approach. Based on each variable separately, the maximum of the variables nutrition, parental approach, and siblings were found in the United Kingdom, while the maximum of electronic games and video viewing were found in the United States. These levels of maximum and minimum has been used in proving or disapproving the hypothesis.







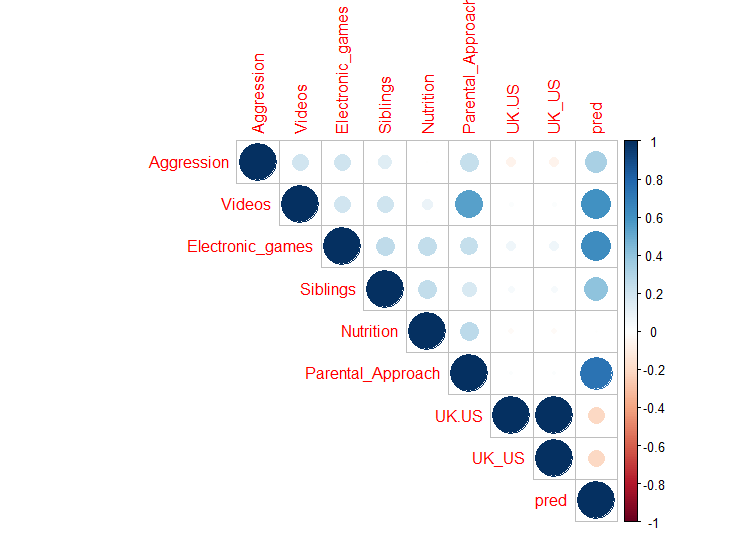


**Machine Learning**

A linear regression technique was used to create the models of the cleaned data. To begin the structure of the data was vied by using the str() function, to inform us that there were 534 variables and 8 variables used. . To create the linear regression models, the dataset had to be first split into 80 percent of the data and 20 percent of the data. For the first model, the structure function and the summary function was used to determine the p-value of the model with using one independent variable to the dependent variable. The p value is the observed significance level of the hypothesis. The idea is to get the p-value to be small because p values range from 0 to 1. A small p-value (typically ≤ 0.05) indicates strong evidence against the null hypothesis, so you reject the null hypothesis .A large p-value (> 0.05) indicates weak evidence against the null hypothesis, so you fail to reject the null hypothesis. In the first model the p-value was 1.763e-05. The second model included the same functions summary, and structure but with the variables videos and electronic games, and the p-value was 6.29e-08. In the third model, the variables included all the independent variables, and the p-value was 4.074e-09. The p-values of the models determined that these were not good models to use for this search so new models would need to be used to get a better understanding of the data.

**Results**

After the variables in the model have been correlated, it is shown that each variable perfectly correlates to itself for each gives a value of 1. In making a linear model of the variables, it shows how each variable is correlated to the other variables. The correlation between the prediction model and the original dataset was slightly similar. The r squared values for both the testing set and the training set were very close after the cross validation was ran on each set which told us that the model was consistent, but due to the value being far from 1 and close to 0 is tell us that the model could be tweaked to become better.



**Future Work**

In enhancing the data more variables could be used to compare the aggression levels of children in each country. Other factors that could be used could be used are school environment and neighborhood environment. Also more countries around the world, at least one country one ach continent can be used to compare the data of children of each country. More in depth, the factors can be compared in countries that are in poverty in compared to more developed countries.

**Acknowledgements**

I would like to acknowledge my Springboard mentor, Mr. Blaine Bateman, and my faculty mentor, Dr. Torina Lewis for their expressed support. This research opportunity was sponsored by the National Science Foundation.