Name: _____

Correlation Investigation

Before you begin the written report,

1. In a sport of your choice, select a season that you are interested in. For this project, you don't have to use a sport.
2. Pick two explanatory variables and a response variable. Pick variables that you think might be correlated. For example, the number of rushing yards and the number of wins per team in a football season. You could choose to compare athletes instead of teams. You are encouraged to have at least 20 data points.
   You can use http://sports-reference.com.
3. For a more thorough investigation, study more explanatory variables, and compare the results. Another extension would be to do a hypothesis test for two of your explanatory variables.

To complete the written report,

1. Write an introduction that gives a brief summary of the season you have chosen, a list of the teams/athletes you will be investigating, and why you chose to focus on the chosen variables.
2. Predict, before doing your investigation, which of the two explanatory variables you think is a better predictor of the response. Make a prediction of whether the stronger correlation is significant.
3. Record the pairs of *PERFORMANCES* for each team/athlete in a table.

Team Name	Response Variable (y)	Explanatory Variable 1 (x_1)	Explanatory Variable 2 (x_2)
A			
B			
…			

4. Graph two scatter plots, one for each explanatory variable. Then describe and compare the two scatter plots, including a comparison of the correlation (r) using each explanatory variable.
5. Do a hypothesis test for the variable with the stronger relationship to the response variable to see if the association is significant. Write down the hypotheses (H_0 and H_a). Simulate the distribution of the correlation assuming that the true correlation is 0. Use the results to estimate the p-value, interpret the p-value, and make an appropriate conclusion.
6. Discuss what we know about what causes a correlation.

Grading Correlation Investigation	A complete report	Points
Introduction and Data Collection	• Describes the context for the research, including why a particular season and variables were chosen • Uses one response variable and two explanatory variables. • Has a clearly stated question of interest • States your prediction of which explanatory variable has the stronger correlation. • States your prediction of whether the stronger correlation is significant (as determined using the hypothesis test). • Describes how the data were collected • Includes a table of all the data	(4)
Scatter Plots and Correlation	• Includes appropriate types of scatter plots • Graphs are clearly labeled and easy to compare • Correlation (r) is calculated • Graphs are discussed/compared correctly	(4)
Hypothesis Test and Analysis	• The test statistic for the hypothesis test is defined • Hypotheses are clearly stated • Clearly describes method for performing a simulation and includes an large number of trials • Displays results of simulation in a clear, well labeled dot plot • Estimates the p-value correctly	(4)
Conclusions	• Correctly interprets p-value in context • Correctly uses the results of the simulation to draw an appropriate conclusion about the question of interest • Shows evidence of critical reflection (discusses possible errors, shortcomings, limitations, alternate explanations, etc.) • Explains what can be learned about cause and effect from a correlation	(4)
Overall Presentation/ Communication	• Investigation is well organized, neat and easy to read • Ideas are well communicated, including appropriate transitions between sections • Investigation is focused on answering the question of interest	(4)
Total		(20)