****

**Last Name: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ First Name: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**Lab 3: Looking at Data: Relationships**

Write your name, class section, lab title and date. Print out commands and outputs. Make sure to answer all parts of each question.

1. Open the “Debt” data set from Website and save as a csv file. Import data set into Rstudio. Don’t forget to attach. The data you imported is a spreadsheet giving the government debt for 33 countries that have data for the years 2005 to 2010. The amount of debt owed by a country is a measure of its economic health. The Organization for Economic Cooperation and Development collects data on the central government debt for many countries. Since countries that have large economies tend to have large debts, we have chosen a table that expresses the debt as a percent of the gross domestic product (GDP).

We are interested in determining if there is a relationship between the debt in 2009 and in 2010. That is, we want to know if debt in 2009 is a good predictor of debt in 2010.

a) Make a scatterplot of the data, and draw the regression line on the scatterplot.

b) What is the correlation value? Describe the relationship (strength and direction) between the debt in 2009 and the debt in 2010 using this value.

c) Find the least-squares regression line and write it in the form ŷ = b0 + b1*x.*

d) Use equation in part (c) to predict the percent of debt in 2010 of a country with a 2009 debt of 50%.

e) Find the residuals and plot them against the explanatory variable. Explain what the residual plot tells you about the relationship of debt owed in consecutive years. (Use definition of residuals and residual plot).

2. Find two variables with perfect linear relationship.

a) Generate/create two vectors with 10 data points for two variables that follow a PERFECT positive linear relationship (They do not necessarily need to be variables that exist or make sense in real life).

b) Draw a scatterplot of the variables you created in part (a).

c) What do you expect the correlation value to be? Why?

d) Find correlation.

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\* Rstudio Code goes Here \*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*







