POL 681 Problem Set 1: Regression and Correlation Review

Objective: To give you experience in hand-calculating the correlation coefficient, regression parameter estimates, and associated quantities of interest from a linear regression model.
Directions: Please use the following data to answer the questions. Show your work and do not use statistical software to solve these problems. Below I've given you the data on two variables: the amount of money the House incumbent spent on his reelection bid and the percentage of votes the incumbent received. These data are for the 6 House incumbents from the state of Arizona from the 1998 House election. Note that money spent is in terms of thousands of dollars. Therefore, a '362' denotes $362,000 dollars.

Table 1: Money Spent and Incumbent Vote Margin: 1998 Election

District	Incumbent	Money	Votes
1	Matt Salmon	362	65
2	Ed Pastor	418	68
3	Bob Stump	246	67
4	John Shadegg	526	65
5	Jim Kolbe	708	52
6	J.D. Hayworth	1839	53

Questions

Political scientists have pointed out the seemingly paradoxical finding that the money spent by an incumbent is inversely related to vote totals. Evaluate this claim with these data. Specifically, answer the following (each question is worth 5 points):

1. What is the correlation between money spent and votes received? Note that when computing the correlation do *not* convert money into dollar units; leave it in terms of thousands of dollars. (Show all work).
2. What does this correlation tell you, substantively?
3. Draw a scatterplot of the data on an xy graph. Clearly label your graph. Assume that y corresponds to votes and x corresponds to money spent.
4. What are the coefficient estimates for the slope and intercept from a regression model where votes were treated as a function of money spent? (Show all work)
5. What does your regression model tell you, substantively?
6. On the scatterplot drawn in problem 3, draw the predicted regression line generated from your regression model.
7. Compute each of the variance components for the regression model; that is compute the following quantities (show all work):
$\sum(y_i - \hat{y})^2$ and $\sum(\hat{y}_i - \overline{y})^2$
8. From your work in problem 7, compute the r^2.
9. Based on the value of the r^2, does knowing the amount of money an incumbent spent seem to predict well the percentage of votes the incumbent received?
10. Do these data lend support to the claim of an inverse relationship between money spent and votes won? Why might this finding be expected (I'm asking you to speculate about the substantive reasons)?