Logistic Regression HW: Due 2/27/24

Download dataset http://www.cse.sc.edu/~rose/590B/CSV/gold_target1.csv The description of dataset is http://www.stat.ufl.edu/~winner/data/gold_target1.txt. The description of each column is shown below:

```
Column 1: As level                # Real value
Column 2: Sb level                # Real value
Column 3: Lineament Proximity     # binary, 1=Present, 0 if absent  (0.5km)
Column 4: Gold deposit proximity # binary, 1=Present, 0=absent  (0.5km)
```

Step 0: start by changing the default column headings from V1, V2, V3, and V4 to `"As_Level", "Sb_Level", "Lineament_Proximity", and "Gold_Deposit_Proximity"`, respectively. If you are flummoxed, review the "Int ro to R-RStudio" slides to see how to change the names of variables.

Step 1: Explore column 4 versus columns 1 and 2.

a. Plot column 4 (Y) against column 1 (X).
   Be sure to label each axis and include a title. Save this plot to a pdf file.
b. Plot column 4 (Y) against column 2 (X).
   Be sure to label each axis and include a title. Save this plot to a pdf file.
c. Try fitting column 4 versus column 2 with a logistic model glm(). Hint: You might want to review the logistic regression lab.
d. Visualize the fit of your model using commands similar to:

```
plot(gold_target1$Gold_Deposit_Proximity~gold_target1$Sb_Level)
lines(gold_target1$Sb_Level,lrm1$fitted,type="l", col="red")
```

   where **lrm1** refers to your logistic regression model. If you name it something else, then substitute the name you used.

   Be sure to label each axis and include a title. Save this plot to a pdf.

e. Now try fitting column 4 versus columns 1 and 2 with the logistic model glm(). How can you accomplish this? When you only have Y versus X, you use Y~X as you did in step c. When you have X1 and X2 then you use Y~X1+X2. **Note: RStudio will give a warning that glm fitted probabilities numerically 0 or 1 occurred.** This is caused by the data in column 1. **Note: you do NOT need to create a plot for this step.**
f. Compare the models from step c with that of step e using the function summary(). In particular, compare the estimated coefficient for **gold_target$Sb_Level** in the two models. What are the two values? How have the confidence values for these estimates changed? (Hint: look at the significance codes.)
g. **(Grad Students only)** Explain the results from part f). Specifically, explain why you think the confidence values for the estimated coefficients changed. Hint: think about the difference between the "reduced model" and the "full model". We discussed this in class.