Linear/Logistic HW: Due 3/14/23

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

1. Explore the first two columns which contain real numbers:
    a. Plot first column (Y) against second column (X). **Save the plot to a pdf file.**
    b. Try fitting these two columns with a linear model lm(). Hint: You might want to review the linear regression lab.
    c. As in the linear regression lab, visualize the model with the commands, where **m** is the variable you used to hold the model:

```
par(mfrow=c(2,2))
plot(m)
```

    Save this plot to a pdf file.

    d. **Explain the top left figure.** What does this tell us about the fit of our model?
    e. **Do the residuals have the property of homoscedasticity? Explain!**
    f. Visualize the predicted and observed y values similar to what we did in slide 6 of the linear regression lab. **Save this graph to a pdf file.**

2. Explore column 4 versus columns 1 and 2.
    a. Plot column 4 (Y) against column 1 (X). **Save this plot to a pdf file.**
    b. Plot column 4 (Y) against column 2 (X). **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:

```
plot(gold_target1$V4~gold_target1$V2)
lines(gold_target1$V2,lrm1$fitted,type="l", col="red")
```

    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.
    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$V2. What are the two values? How have the confidence values for these estimates changed? (Hint: look at the significance codes.)
3. If the probability of rain tomorrow is 25%, what are the odds of rain tomorrow?