Load in the data "Cat_Hwt.csv".
Data was collected on deceased male and female adult cats used for experiments:
Sex: 1 for male and 2 for female.
Bwt: body weight in kg.
Hwt: heart weight in g.
Height: Height in cm
Age: Age in years.
Outdoor: kept outdoors; 1 = Always, 2 = Frequently, 3 = Never

1. Relationships

a) Investigate if any of the <u>continuous numerical</u> variables have a linear relationship by producing scatterplots, interpret these plots.

b) What is the response variable and why? What is the research question of interest in this dataset?

c) Investigate if any of the <u>categorical /discrete</u> variables seem to have a relationship with the response variable (selected in part b) using boxplots. Interpret these plots.

d) Using part c), <u>choose one categorical /discrete</u> variable with at least 3 categories to test to see if there is any difference between the means of the response variable (selected in part b) using an <u>one-way ANOVA te</u>st. (Note: if categorical variables are not set up as factors, so change to factor prior to running analysis (as.factor())

e) Create a boxplot to see if there is any difference between the means of the response variable (selected in part b) across the two variables Sex and Outdoor. Also, create a plot to see if there is an interaction effect for these two variables with the response variable. Interpret these plots. Test to see if any of these factors and/or interactions have a significant relationship with the response variable (selected in part b) using an <u>two-way ANOVA test</u>. (Note: categorical variables are not set up as factors, so change to factor prior to running analysis (as.factor())

2. Multiple Linear Regression

a) Fit the "best" simple linear regression model, based on your answers from question 1, justify your choice. Comment on whether the assumptions are satisfied, interpretation of the results and the fit of the model.

b) Fit a suitable multiple linear regression model, based on your answers from question 1, justify your choice. Comment on whether the assumptions are satisfied, interpretation of the results and the fit of the model.

c) Are there any variables you would like to add/remove from the model and why? Re-run the multiple linear regression model with these variables, if any, added/removed. (You can do this more than once, in a stepwise process if you think appropriate). Compare the fit of this model to the model in part b. Perform an F-test to compare the two models, stating your hypothesis and the conclusion of this test. (Hint, use the anova() command). If no variables are removed compare model ran in part a with model ran in part b.

d) Conclude your overall results. Use your preferred model to predict the fitted values for your response variable and calculate the residual term if you are given the following data:

Sex	Bwt	Hwt	Height	Age	Outdoor
2	2.6	11.2	23.5	10	2
1	3.1	12.5	25.4	13.2	2