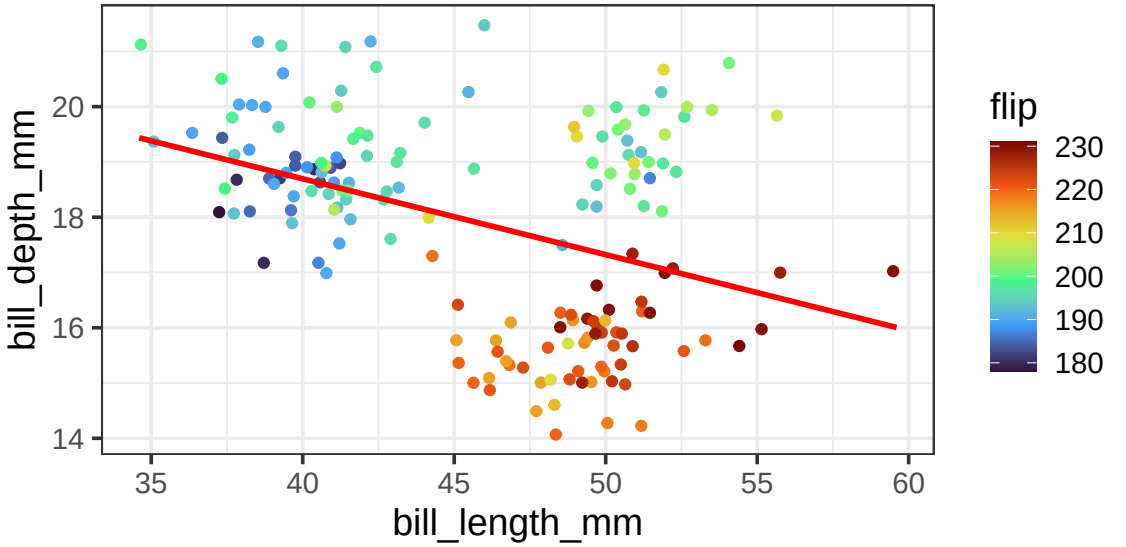
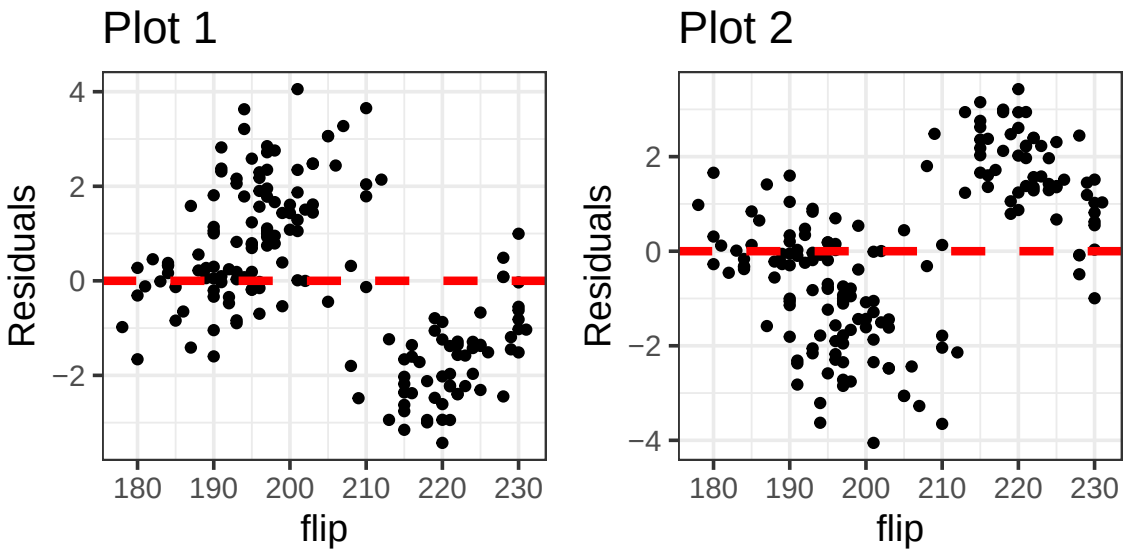
MLR

2025-12-12

Question 1

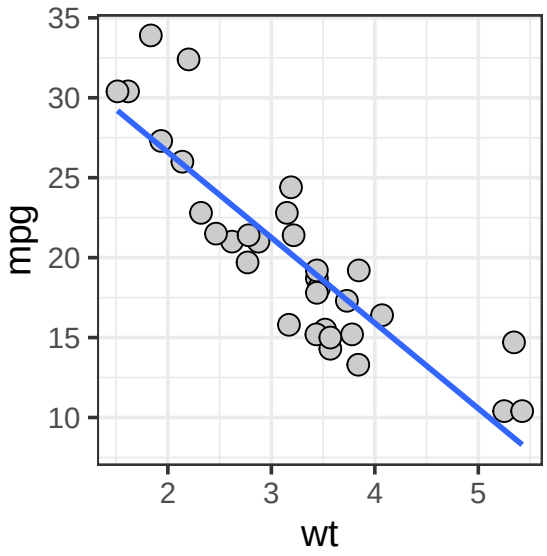
Below is a plot showing the linear relationship between bill length and depth. Each observation is colored by flipper lenght. Which of Plot 1 or Plot 2 represents the correct residual plot? Justify your answer





Question 2

Below is a model illustrating the relationship between vehicle weight and miles per gallon:

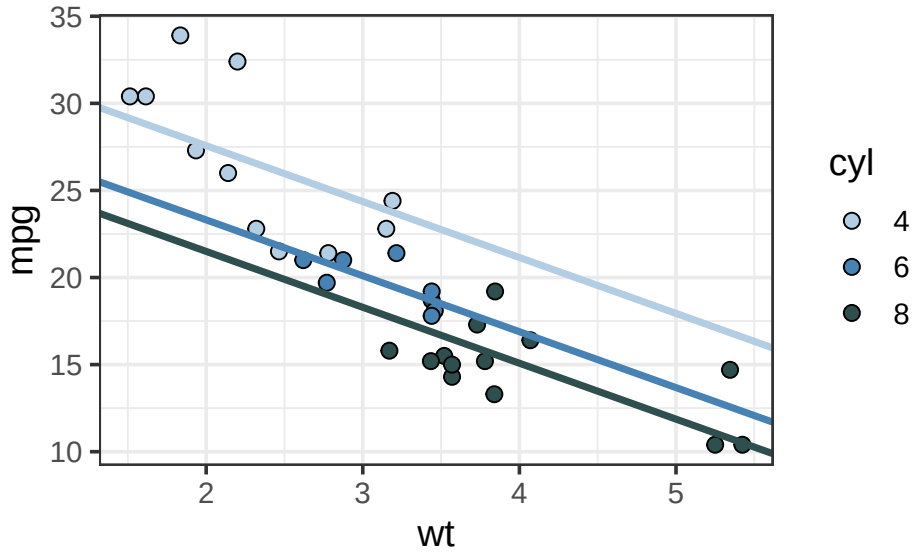



```
##
## Call:
## lm(formula = mpg ~ wt, data = mtcars)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -4.543 -2.365 -0.125  1.410  6.873
##
## Coefficients:
##             Estimate Std. Error t value             Pr(>|t|)
## (Intercept)   37.285      1.878   19.86 < 0.0000000000000002 ***
## wt            -5.344      0.559   -9.56        0.00000000013 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.05 on 30 degrees of freedom
## Multiple R-squared:  0.753,  Adjusted R-squared:  0.745
## F-statistic: 91.4 on 1 and 30 DF,  p-value: 0.000000000129
```

1. What is the null hypothesis associated with the coefficient for weight?
2. Provide an interpretation of this coefficient. Do we have evidence to reject the null?
3. How much variance is explained by this model?

Question 3

Below is a model predicting miles per gallon with weight and number of cylinders as a categorical variable

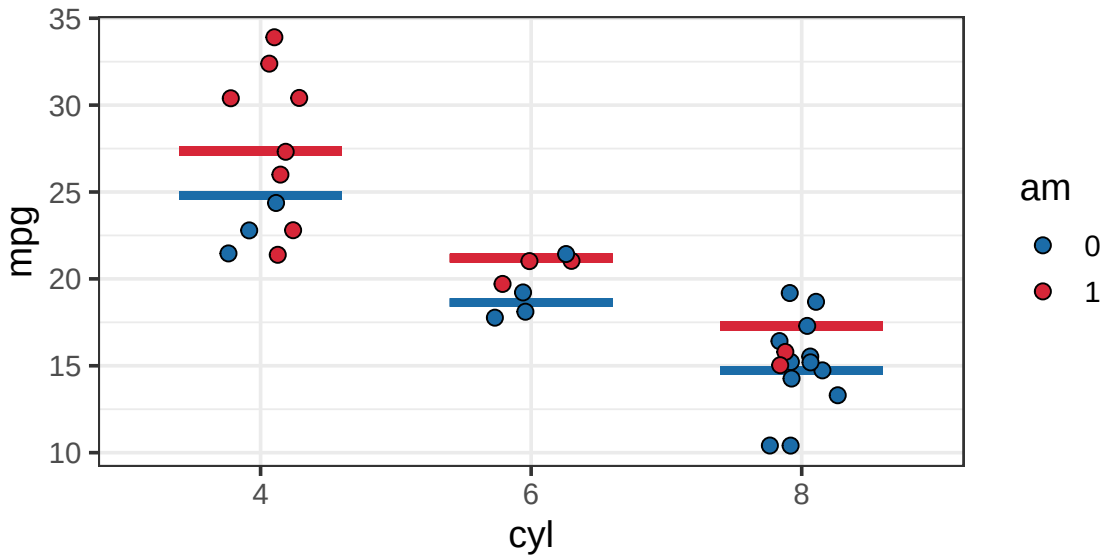



```
##
## Call:
## lm(formula = mpg ~ wt + cyl, data = mtcars)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -4.589 -1.236 -0.516  1.384  5.792
##
## Coefficients:
##             Estimate Std. Error t value             Pr(>|t|)
## (Intercept)   33.991      1.888   18.01 < 0.0000000000000002 ***
## wt            -3.206      0.754   -4.25              0.00021 ***
## cyl6          -4.256      1.386   -3.07              0.00472 **
## cyl8          -6.071      1.652   -3.67              0.00100 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.56 on 28 degrees of freedom
## Multiple R-squared:  0.837,  Adjusted R-squared:  0.82
## F-statistic: 48.1 on 3 and 28 DF,  p-value: 0.0000000000359
```

1. Write out the linear equation represented by this model
2. For the coefficient for cylinder 6, what is the null hypothesis? Interpret the coefficient returned by the model. Do we have evidence to reject?
3. Using the model output, provide an assessment of how the model including weight and cylinders compares to the model with just weight. Which would you prefer?

Question 4

For this, `am` represents the variable automatic/manual, with 0 corresponding to an automatic transmission




```
##
## Call:
## lm(formula = mpg ~ cyl + am, data = mtcars)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -5.962 -1.497 -0.206  1.891  6.538
##
## Coefficients:
##             Estimate Std. Error t value             Pr(>|t|)
## (Intercept)    24.80       1.32   18.75 < 0.0000000000000002 ***
## cyl6           -6.16       1.54   -4.01              0.00041 ***
## cyl8          -10.07       1.45   -6.93           0.00000015 ***
## am1             2.56       1.30    1.97              0.05846 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.07 on 28 degrees of freedom
## Multiple R-squared:  0.765,  Adjusted R-squared:  0.74
## F-statistic: 30.4 on 3 and 28 DF,  p-value: 0.00000000596
```

1. Write out the linear equation for this model. Explain what variables are included and how they are being used.
2. Does it appear the manual or automatic transmissions get better gas mileage? Testing at $\alpha = 0.05$, do we have enough evidence to conclude that there is a difference?
3. How does this model compare to the one with cylinders and weight? Justify your answer

Question 5

Model using weight, engine displacement, and transmission

```
##
## Call:
## lm(formula = mpg ~ wt + disp + am, data = mtcars)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -3.489 -2.411 -0.723  1.750  6.329
##
## Coefficients:
##             Estimate Std. Error t value       Pr(>|t|)
## (Intercept) 34.67591    3.24061   10.70 0.000000000021 ***
## wt          -3.27904    1.32751   -2.47          0.020 *
## disp        -0.01780    0.00937   -1.90          0.068 .
## am1          0.17772    1.48432    0.12          0.906
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.97 on 28 degrees of freedom
## Multiple R-squared:  0.781,  Adjusted R-squared:  0.758
## F-statistic: 33.3 on 3 and 28 DF,  p-value: 0.00000000225
```

1. Write out the linear equation for this model.
2. Consider the F Statistic at the bottom of this output. What is the null hypothesis associated with this?