2. bc

7.(b) To understand reliability of the tests, we need to check the assumptions underlying the usage of each inferential procedure and the models used.

Assumption behind 6(b) is reasonable because we can check assumptions behind model 3:

$$mpg = \beta_0 + \beta_1 weight + \beta_2 transmission\_typeM + \beta_3 weight \cdot transmission\_typeM + \varepsilon.$$

Check on the residual plot of model 3, the plot generally satisfies constant spread, zero mean and linearity without alternatively possible trend. Check on the normal quantile plot, which is very along the line with small deviation on two sides as a possible result of randomization.

Therefore test in 6(b) is considered reliable.

However, for test in 7(a), in addition to condition for full model, we also need to check assumptions for reduced model:

$$mpg = \beta_0 + \beta_1 weight + \varepsilon$$

Checking on the residual plots for model 1, we can observe suspicious upward alternative trend in residual plot, which is upwards on two sides and lower in the middle. Also, spread of data is not quite constant where data is thin for smaller fitted values. There's large deviation on lower end for normal quantile plot, which means that the distribution of error is not close to normal.

Therefore, test in 7(a) does not meet well linearity, constant spread and normality assumptions. Therefore, test in 7(a) is not as reliable as the test in 6(b).

8. First, we can plot added variable plot with displacement as the potential added predictor. To do so, we model displacement against weight and transmission_typeM, record residuals as resids2, and plot resids2 against resids1 (which is residuals of the original plot). In the added-variable plot, we should observe a weak linear correlation with mostly constant spread and no non-linear curvature. In this case, we might know that we can have additional information added to model and the model likely still meets the model assumptions.

Next, we can try adding the variable into the model. We can compare adjusted R^2, which suggests variability explained by the model as well as number of variables (or whether the addition of the variable overfits the data). If adjusted R^2 increases, we consider adding the predictor.

Finally, we can plot residual v.s. fitted plot to check that there's constant spread, zero mean and no alternative curvature. We also need to plot normal quantile plot to check whether data are along the line, which suggests reasonable normality.