While the "two-lines" model is a very good way to begin your journey into multiple linear regression, there is a vast number of other possible models that we could use, infintely many actually. This problem will allow you to practice what is called a "high dimensional model."

A researcher is interested in knowing what controllable factors influence the rate of vehicle accidents on highways. Data was collected in 1973 on highways in Minnesota in an effort to answer this question.

> library(car)
> ?Highway1
> View(Highway1)

Use the **Highway1** dataset in R to determine if the speed limit (**slim**), width of the shoulder on the roadway (**shld**), and the percentage of vehicle volume that are trucks (**trks**) have any significant effect on the rate of accidents.

The mathematical model for this regression is given by

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i3} + \epsilon_i$$

where Y_i is the Rate of accidents , X_{i1} is the Speed limit , X_{i2} is the Width of shoulder , and X_{i3} is the Percentage of trucks .

Use R to find an estimated regression equation of the model defined above.

$\hat{Y}_i =$ 17.69 $+$ -0.2048 $X_{i1} + 0.02015 X_{i2} + -0.28175 X_{i3}$.

Select the variables that have a significant effect (non-zero β_i) on the average rate of accidents when all variables stated above are included in the model.

- [x] Speed limit
- [ ] Width of shoulder
- [x] Percentage of trucks

Check the appropriateness of this regression model by creating the Residuals vs. fitted-values and normal Q-Q plots. One point is labeled as the most extreme outlier in these plots? What number is next to this point?

Point # 27

The linear relation assumption for this model is perhaps somewhat questionable, but constant variance and normality of the errors don't pose any real concerns. Continuing with the regression despite the possible difficulty, use your math skills or R skills to predict the rate of accidents if the speed limit = 55, width of shoulder = 6, and Percentage of trucks = 10.

The predicted rate of accidents is 3.727934