Whenever I want to practice something or a new technique in machine learning, I select a new data set and about six or seven months ago, I came across the  California housing dataset. I selected this data set because I thought to myself that there could be multiple correlating relationships, which could help predict the housing prices, such as crime rate, median income, location and more. The relationship between the dependent variable (median house values) and the independent variables (like median income or housing median age) should be linear. Through initial scatterplots, I was able to view the correlation and the scale of these relationships. For instance, while median income might have a seemingly linear relationship with value of the home, the relationship between housing median age and house value may not be as straightforward as some of the other relationships. I had to remove some outliers which our reading touched on the relationship of the equation becomes more complex and that is something that is overlooked greatly when working on models (Osbourne & Waters, 2002).  Nonlinearities had to be addressed, either by transformations or by considering polynomial terms. I ensured that the variance of residuals remained constant across various levels of the independent variables. A glance at a scatterplot of residuals against predicted values helped. For instance, if the model's predictions are more scattered for districts with higher median incomes than those with lower incomes, that would indicate potential heteroscedasticity.    
  
Refrences:

Osbourne, J. W., & Waters, E. (2002). Four assumptions of multiple regression that researchers should always test. *Practical Assessment, Research and Evaluation*, *8*(2), 2. https://doi.org/10.7275/r222-hv23