Here is the code that was run in order to generate the graph.

|  |
| --- |
| > #Read in the data set GFCLOCKS.csv  > GFCLOCKS <- read.csv(file="C:/Users/buchh/OneDrive/Desktop/regression/hw6/GFCLOCKS.csv",header = TRUE)  >  > head(GFCLOCKS,5L)  AGE NUMBIDS PRICE AGE.BID  1 127 13 1235 1651  2 115 12 1080 1380  3 127 7 845 889  4 150 9 1522 1350  5 156 6 1047 936  >  > lmod<-lm(PRICE ~ AGE + NUMBIDS, data=GFCLOCKS)  >  > ols\_plot\_cooksd\_chart(lmod)  > ols\_plot\_dfbetas(lmod)  > ols\_plot\_dffits(lmod)  > ols\_plot\_hadi(lmod)  > ols\_plot\_resid\_pot(lmod) |
|  |
| |  | | --- | |  | |

Cook’s D Chart was generated by this command,

> ols\_plot\_cooksd\_chart(lmod)

It basically helps us detecting the influential points. A data point having a large cook’s d indicates that the data point strongly influences the fitted values. Here the threshold is 0.125 and there are several values above that number. Hence, we can say that those values are influential as they affect the other fitted values

Just like above function, this command,

> ols\_plot\_dffits(lmod)

helps us to plot detecting influential observations using DFFITs. This command basically returns two things: outliers and threshold. By looking at the graph, we can say that the values that are above or very below the threshold are outliers and affect the fitted values. We see that the values that are above 0.61 and below -0.61 are outliers and influential observations. Removing them will give better understanding of the data.

The plot for Hadi’s influence measure was generated by this command,

> ols\_plot\_hadi(lmod)

The plot is made under the assumption of influential variables existing in the data, either in response variable or the in the predictors or in both. Potential Residual Plot also works like Hadi’s measurement and helps classifying unusual observations as high-leverage points, outliers, or a combination of both.