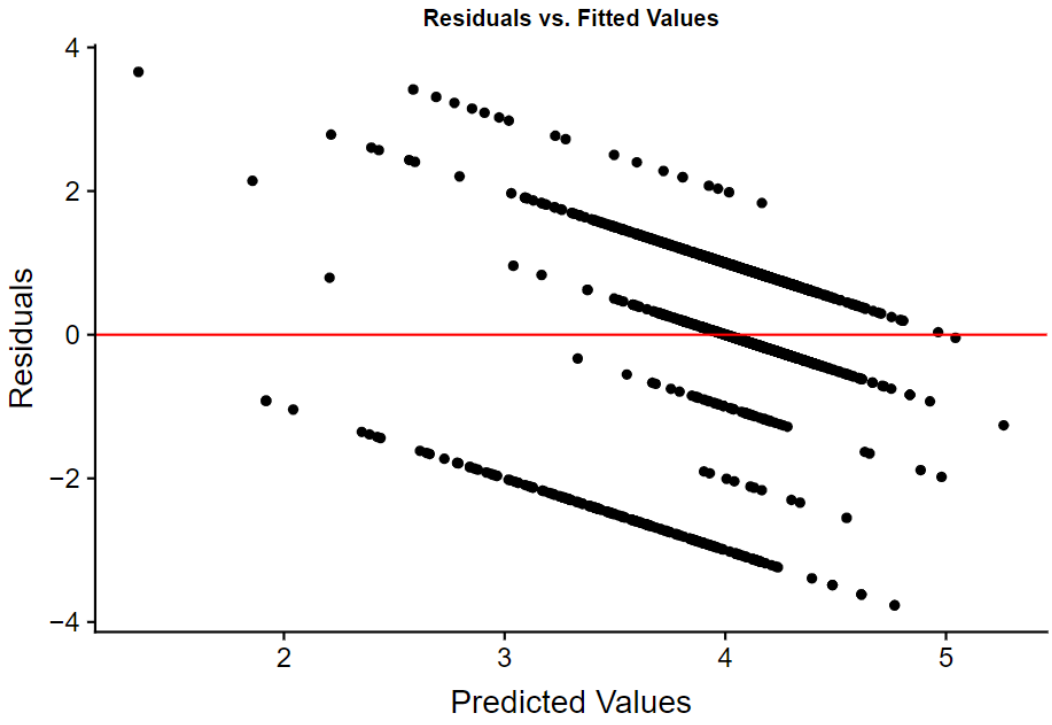
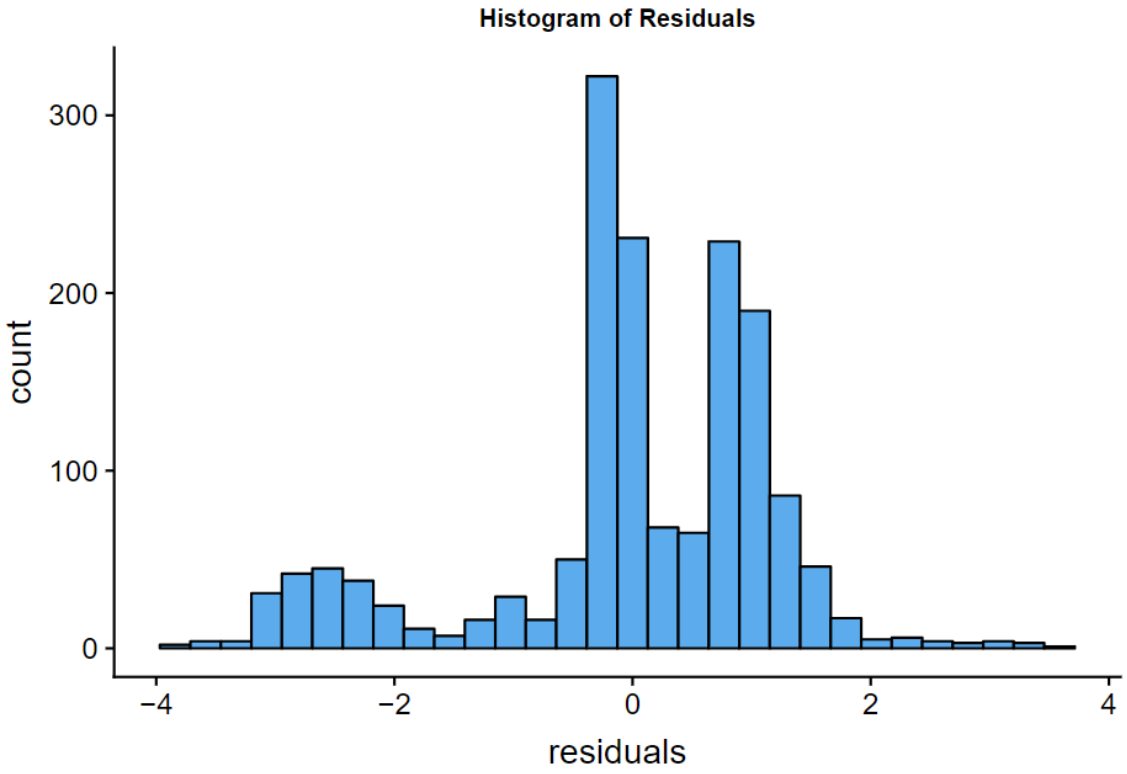
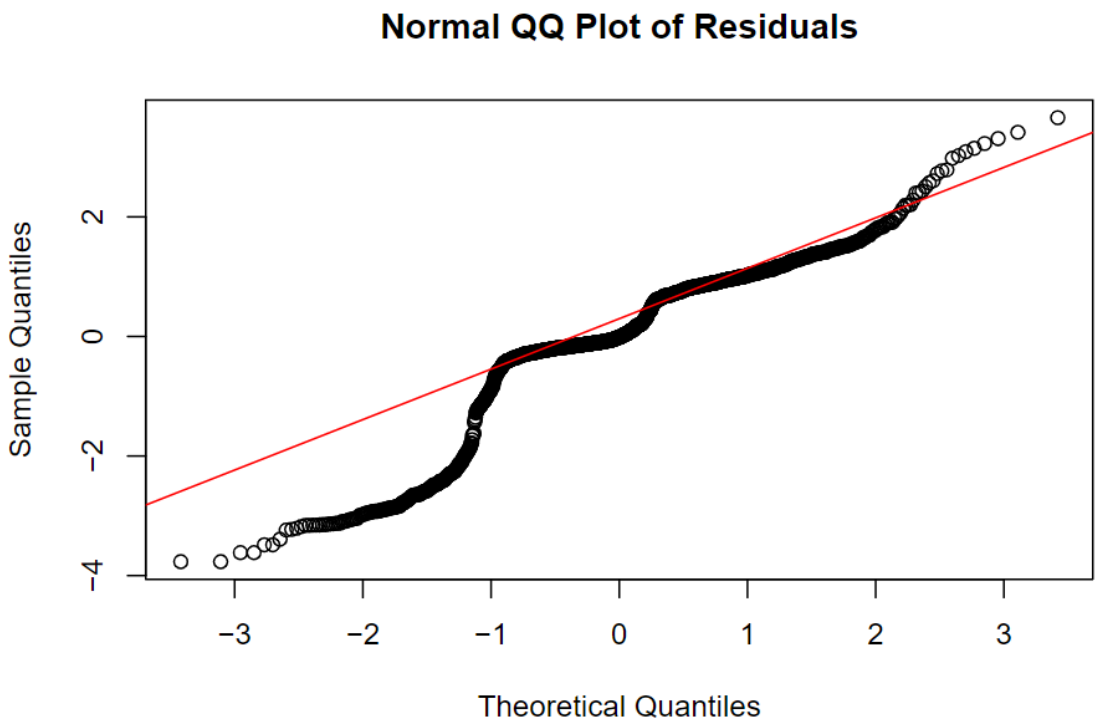
**Multiple Regression Analysis**

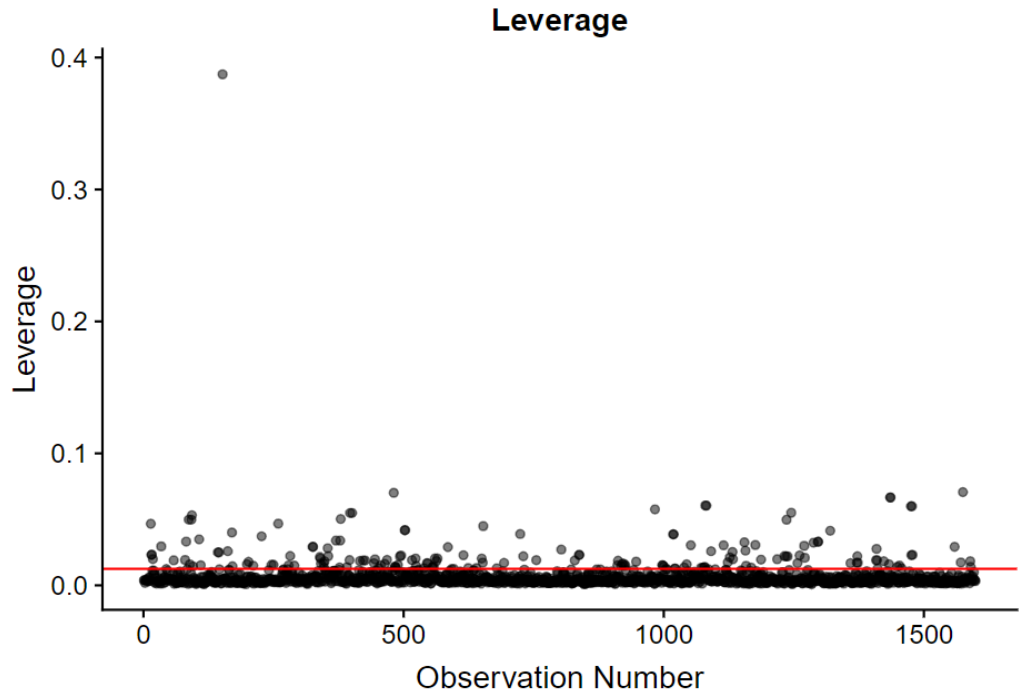



This residual plot is a result of having categorical response variable.

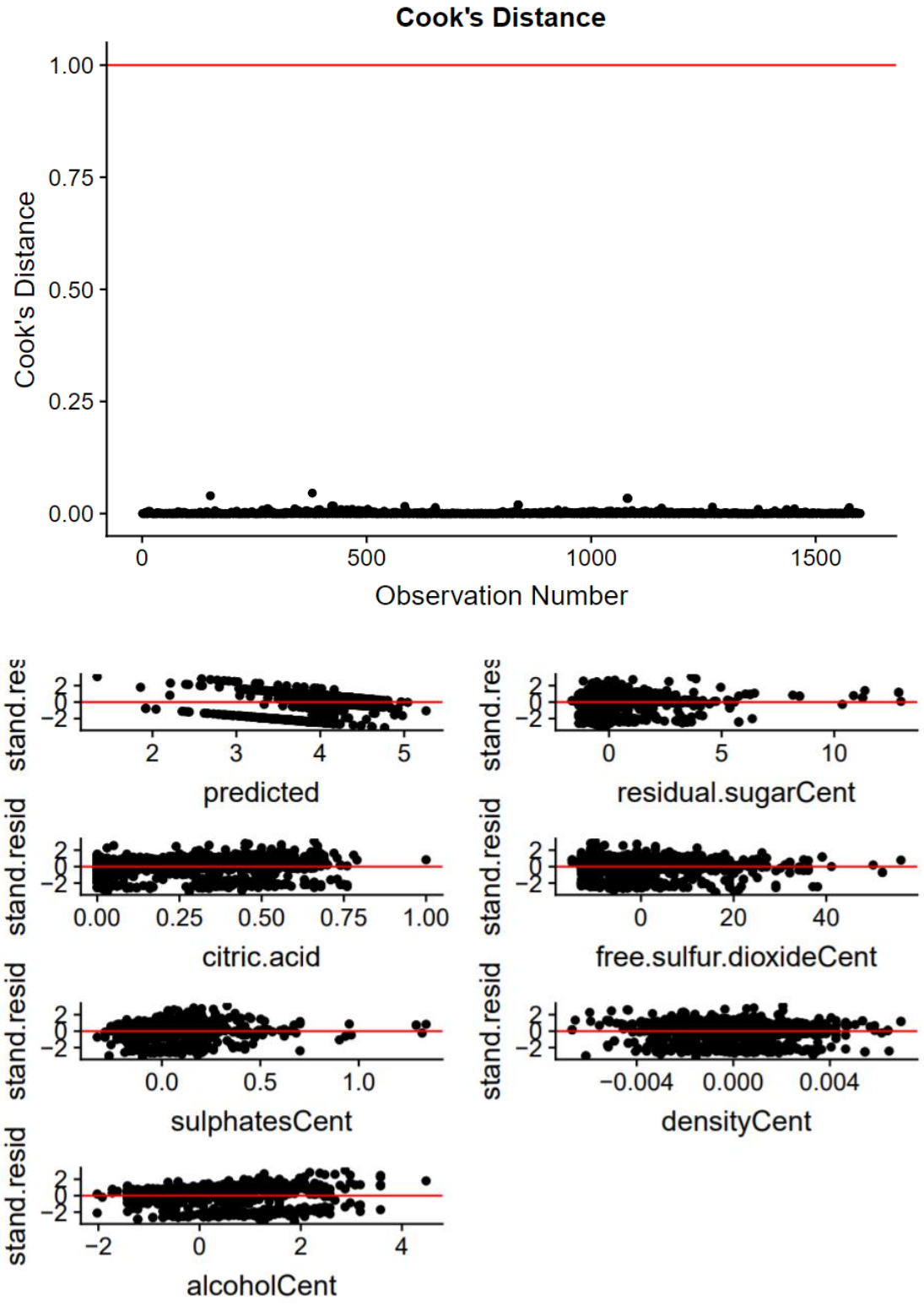





The distribution of residuals appear to be bimodal. The QQ plot also suggests the same conclusion, since we can see a very prominent deviation from the diagonal normal line on the left side. Overall, the Normaltiy Assumption seems to be violated. This could be due to the fact that we are dealing with a categorical response varaible quality.




We can see that there is one point with a significantly high leverage around 0.4, comparing with other observations. This could be an outlier and might be an influence point.



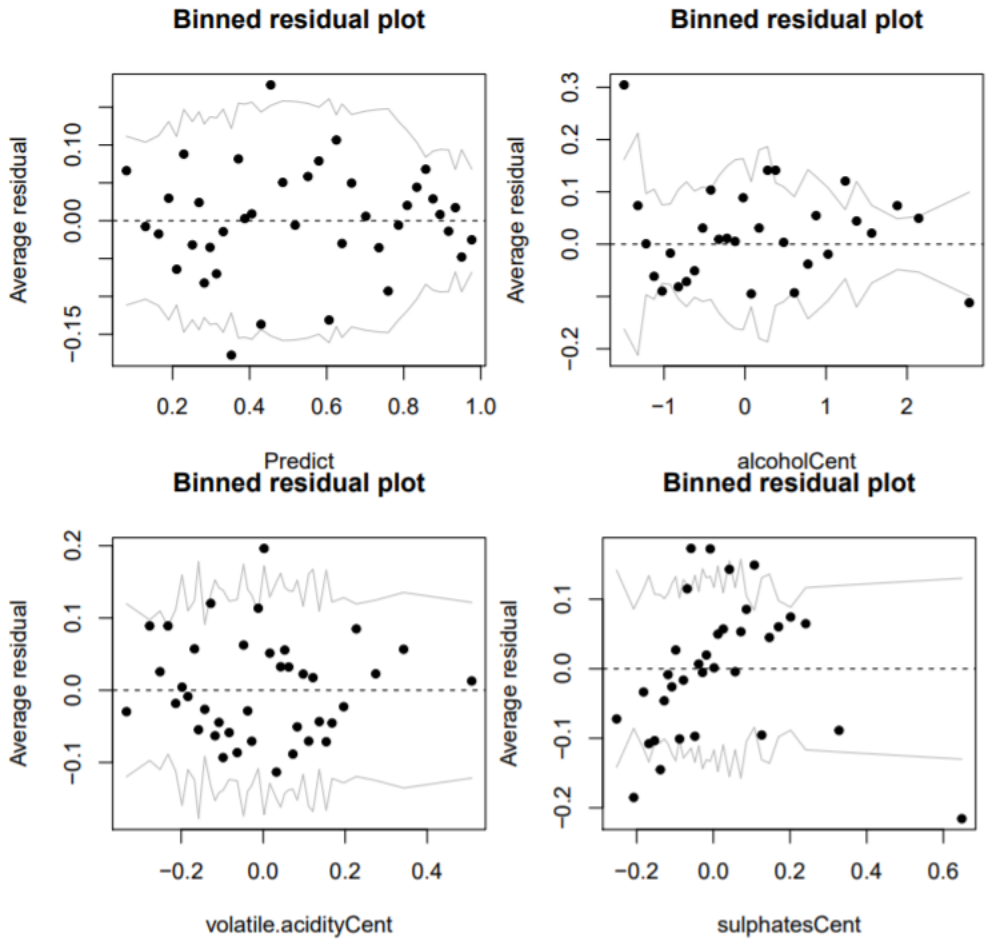

The standardized residuals show some points with magnitude greater than 2, but overall, in combination with our observation from Cook's Distance and just one data point with high leverage away from other points, we can conclude there isn't any obvious influential points in this model.

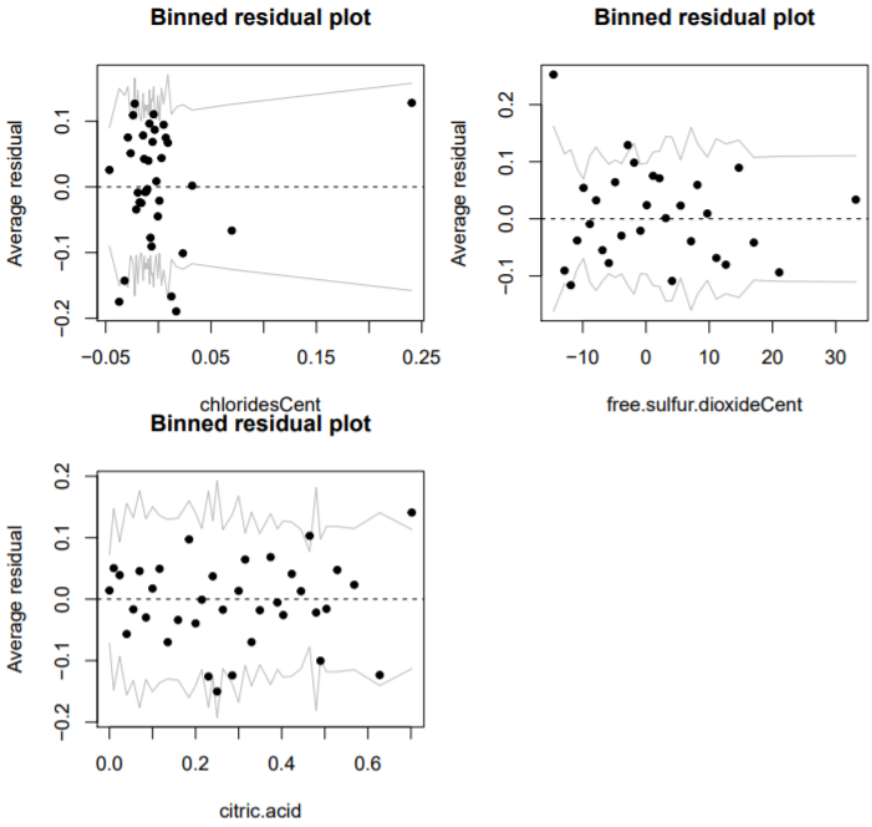
```
kable(tidy(vif(final.model)))
```

```
## Warning: 'tidy.numeric' is deprecated.
## See help("Deprecated")
```

| names | x |
|---|---|
| residual.sugarCent | 1.402748 |
| citric.acid | 1.425081 |
| free.sulfur.dioxideCent | 3.169724 |
| sulphatesCent | 3.922514 |
| densityCent | 2.274231 |
| alcoholCent | 1.825881 |
| citric.acid:free.sulfur.dioxideCent | 3.249458 |
| citric.acid:sulphatesCent | 3.832337 |
| free.sulfur.dioxideCent:alcoholCent | 1.097922 |
| sulphatesCent:alcoholCent | 1.187895 |

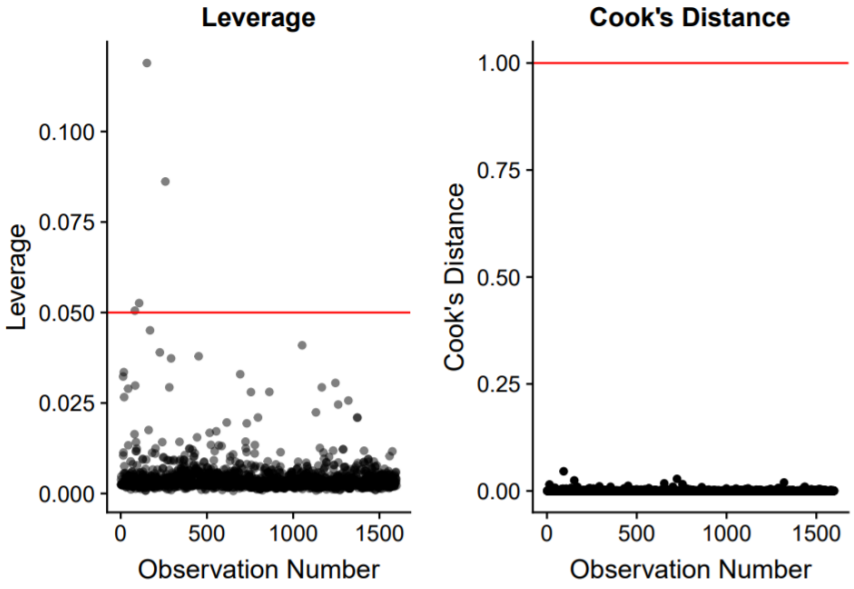
## Logistic Regression






Most of the binned residual plots show random pattern and raises no major concerns of violations. The binned residual plot of alcoholCent and free.sulfur.dioxideCent seems to show an outlier on the top left, suggesting high average residuals for that bin. There is also one outlier on the binned residual plot of chloridesCent on the far right.

The binned residual plot also reveals that sulphatesCent might need a transformation, since there appears to be some linear trend.

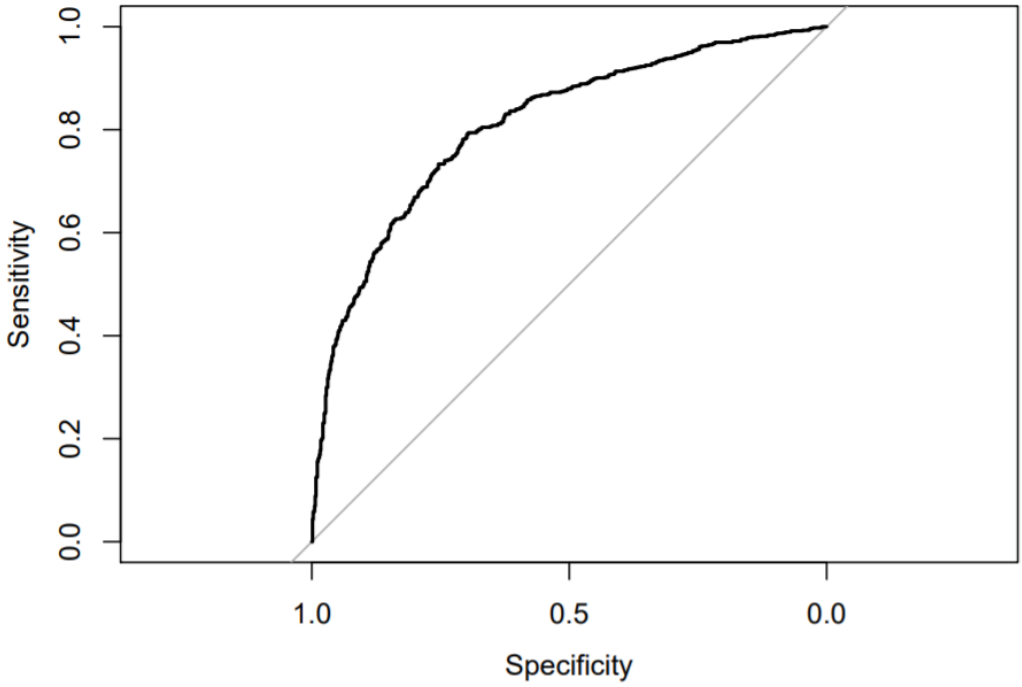



Overall, there is no significant influence point.

```
# multicollinearity
kable(tidy(vif(modellg1.final)))
```

```
## Warning: 'tidy.numeric' is deprecated.
## See help("Deprecated")
```

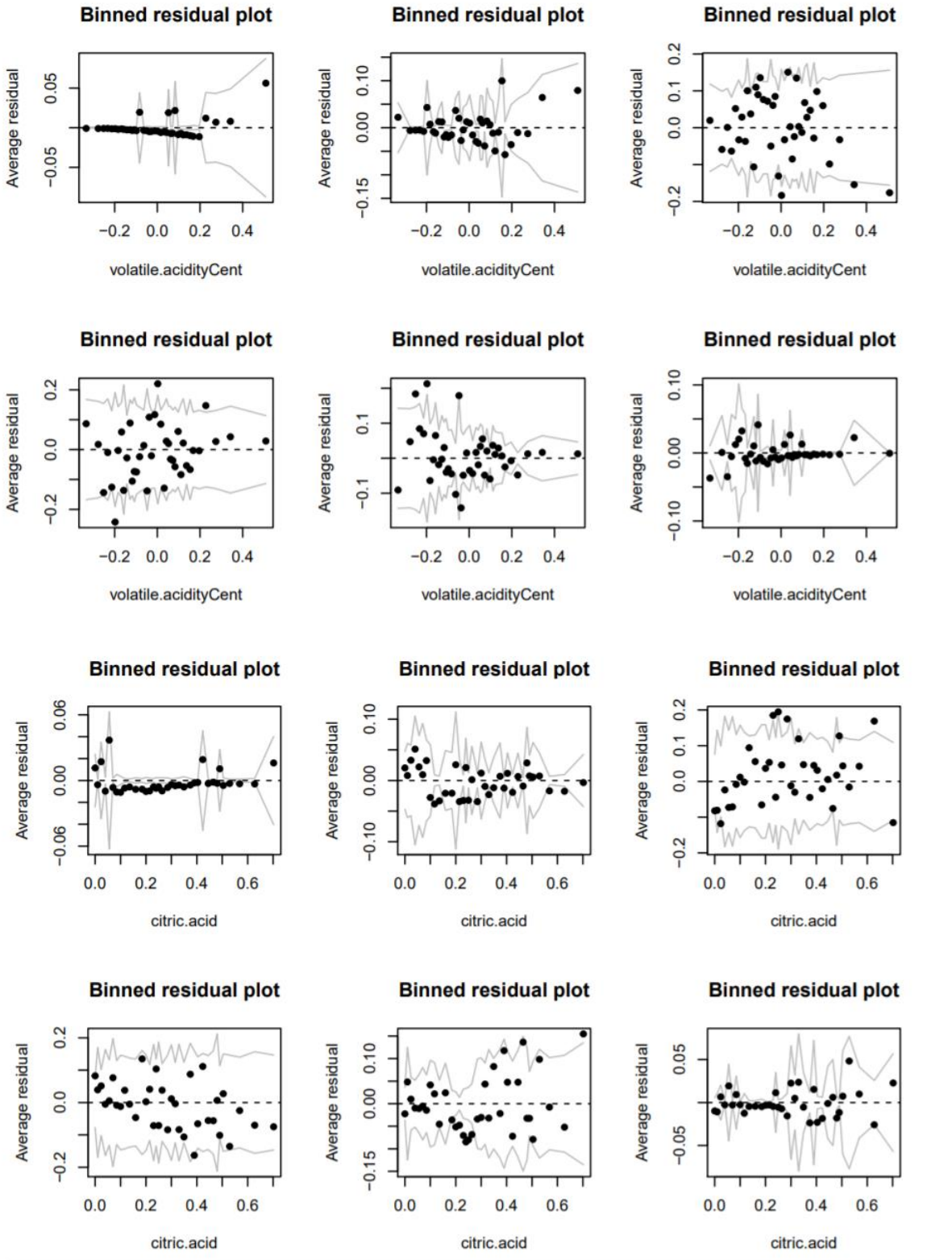
| names | x |
|---|---|
| alcoholCent | 1.050002 |
| volatile.acidityCent | 1.542159 |
| sulphatesCent | 1.416972 |
| chloridesCent | 1.433665 |
| free.sulfur.dioxideCent | 1.018900 |
| citric.acid | 1.650265 |

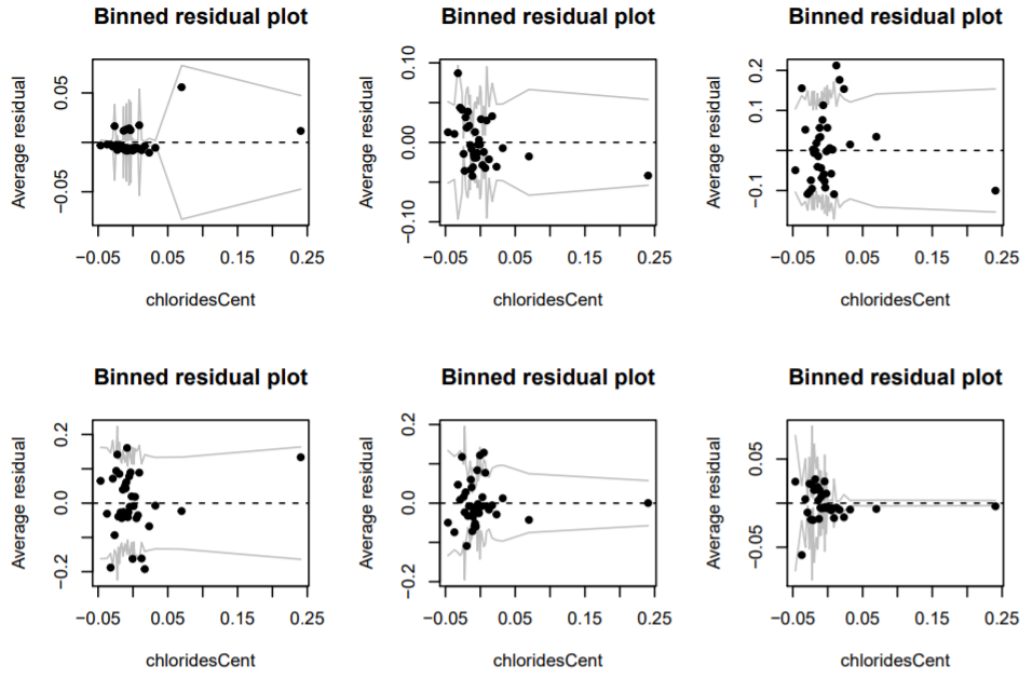



```
## Area under the curve: 0.8092
```

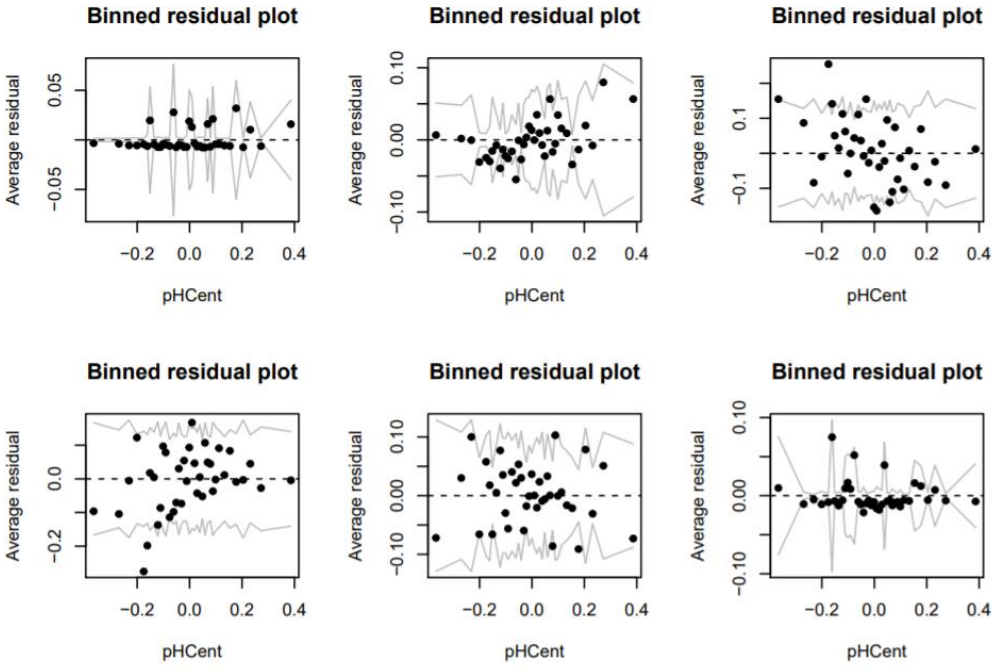
From the ROC curve and AUC calculation, we can see the curve is fairly close to the top left corner (area under the curve is close to 1). This shows that the logistic model is able to distinguish between good and not good quality, so this is a pretty good model.

## Ordinal Regression Analysis

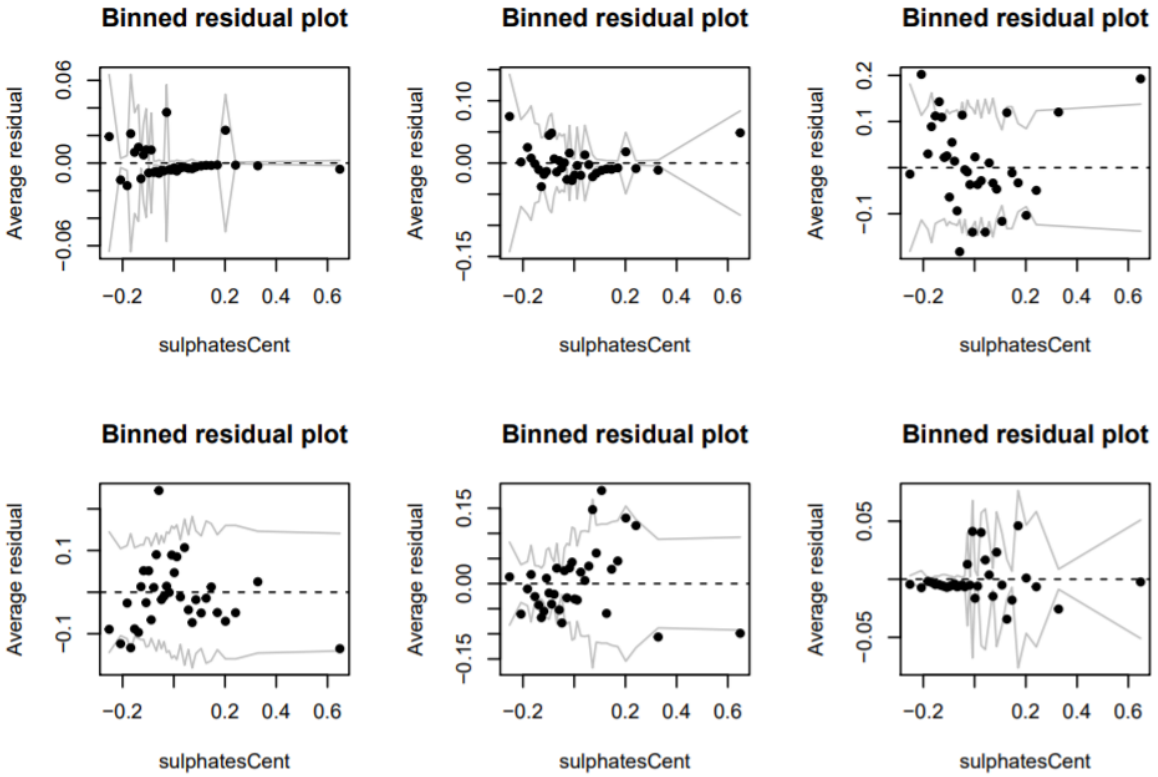




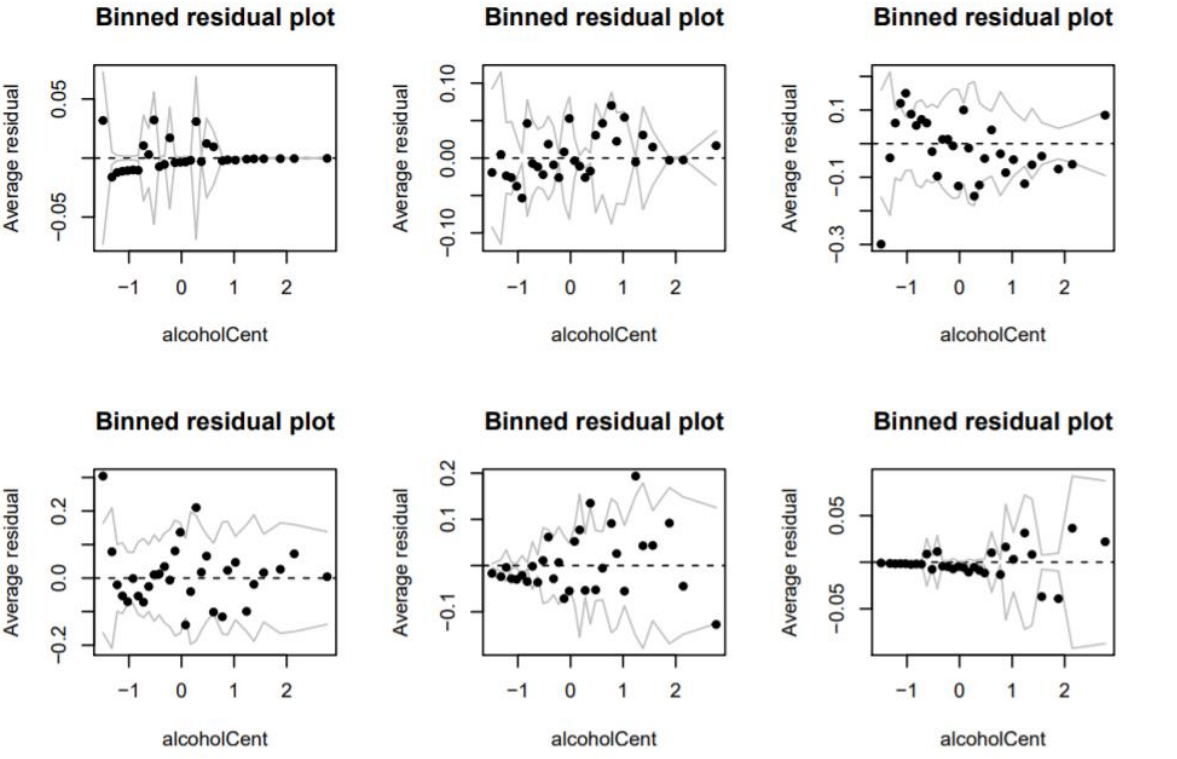

The binned residual plots show a possible chloridesCent outlier.




There is also some linear trends in the binned residual plots for pHCent. We should also tranform this variable.




The binned residual plot shows some potential linear trend and outliers in three categories of quality. We might need to transform sulphatesCent.



```
##
## Re-fitting to get Hessian
##
## z test of coefficients:
##
##                       Estimate Std. Error  z value  Pr(>|z|)
## volatile.acidityCent -3.59868    0.37775  -9.5267 < 2.2e-16 ***
## citric.acid          -0.73977    0.37520  -1.9717 0.0486457 *
## chloridesCent        -5.25929    1.30690  -4.0242 5.716e-05 ***
## sulphatesCent         2.73449    0.36207   7.5524 4.274e-14 ***
## pHCent               -1.47755    0.41452  -3.5645 0.0003646 ***
## alcoholCent           0.94934    0.05799  16.3708 < 2.2e-16 ***
## 3|4                  -6.08448    0.34309 -17.7344 < 2.2e-16 ***
## 4|5                  -4.14345    0.17838 -23.2280 < 2.2e-16 ***
## 5|6                  -0.49521    0.11605  -4.2674 1.978e-05 ***
## 6|7                   2.31842    0.13664  16.9678 < 2.2e-16 ***
## 7|8                   5.33090    0.27003  19.7419 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

p-values of all the variables are extremely small, so they are all significant predictors of the log-odds of the wine falling in or below quality j. We should also not extrapolate beyond quality j=3,4,...,7.