Null Hypothesis

Sophie Wulfing

2/12/2022

Note

This is what I've come up with regarding how to tackle the null hypothesis. I know we've taken NLA data out of circulation but I'm going to continue to use it here only because I was using it earlier as an example set and I wanted to get feedback from you all about the methods and learn a little more about the changes to data before I applied it to other datasets.

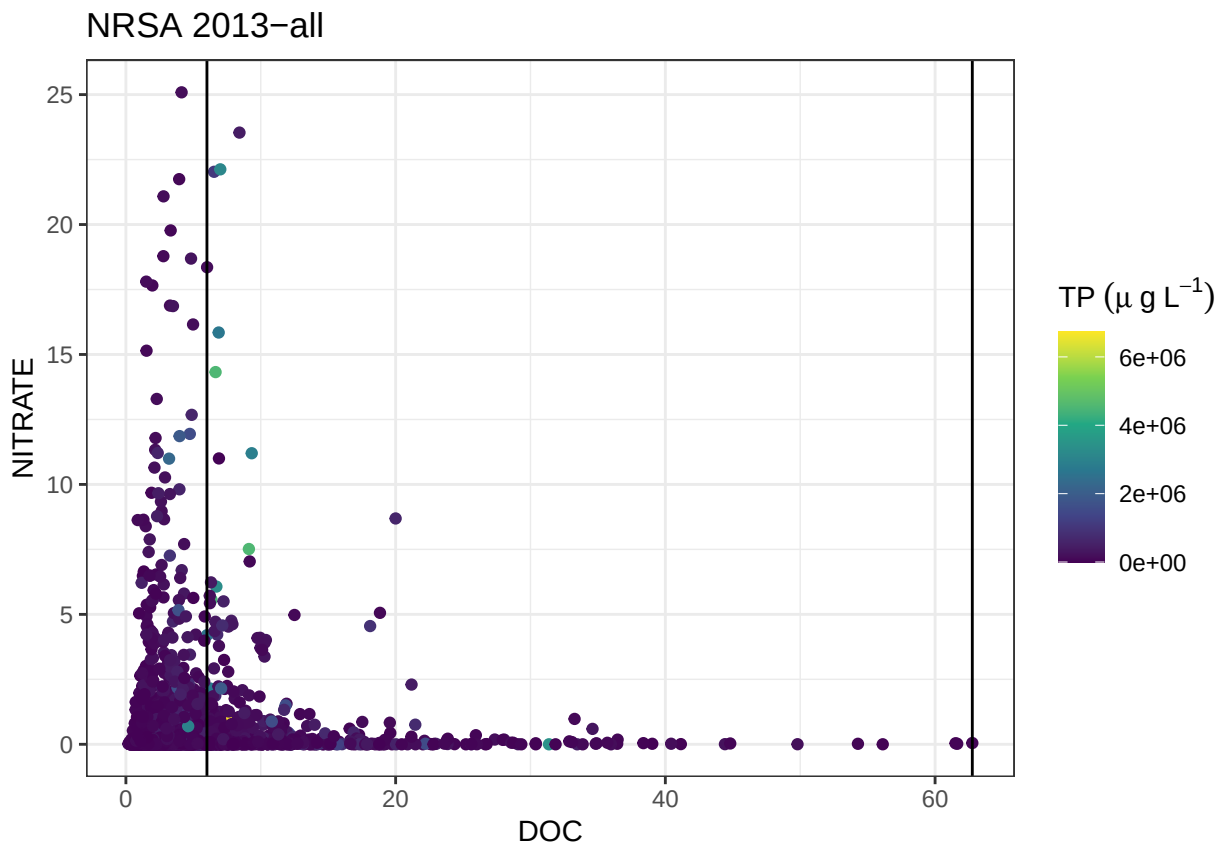
Basically what I plan on doing is trying to linearize data to fit a model to it (I know we've talked about that beore). Once we talk about data, I was going to add other aspects of the water quality as covariates and see what would be the most relevant. I'm pretty new to all this stuff (finally taking a class on it this semester) so I'm open to critiques!

```
test_dataLOGS <- test_data %>%
  mutate(log_Nitrate = log(NITRATE_N_RESULT)) %>%
  mutate(log_TP = log(PTL_RESULT)) %>%
  mutate(log_DOC = log(DOC_RESULT))
                                  # Duplicate data
test_dataLOGS$log_Nitrate[is.na(test_dataLOGS$log_Nitrate) | test_dataLOGS$log_Nitrate == "-Inf"] <- NA
test_dataLOGS$log_TP[is.na(test_dataLOGS$log_TP) | test_dataLOGS$log_TP == "-Inf"] <- NA

quantile(test_dataLOGS$DOC_RESULT)
```

```
##    0%   25%   50%   75%  100%
##  0.19  1.81  3.42  6.01 62.76
```

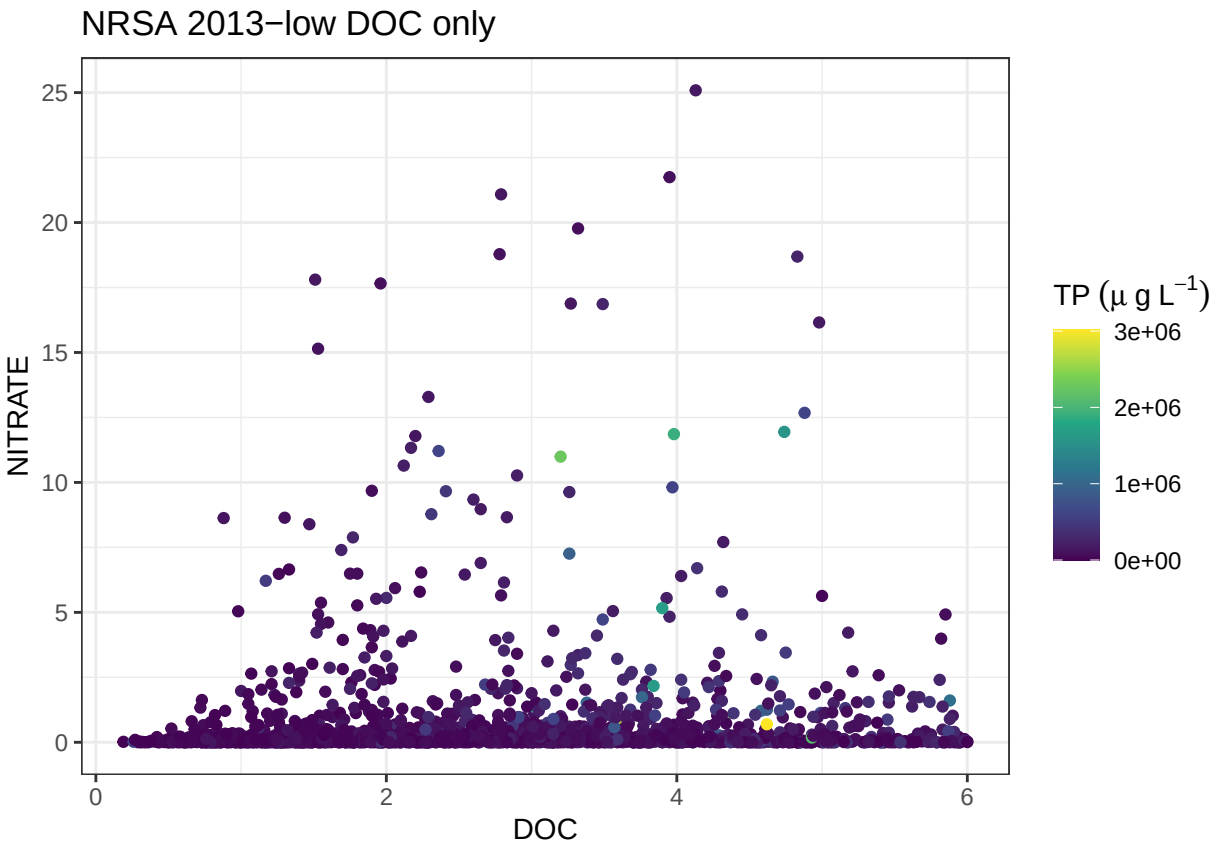
```
q4 <- quantile(test_dataLOGS$DOC_RESULT)[4]
q5 <- quantile(test_dataLOGS$DOC_RESULT)[5]

lowDOC_data <- test_dataLOGS %>%
  filter(DOC_RESULT < quantile(test_data$DOC_RESULT)[4])

par(mfrow = c(1,2))
p <- ggplot(test_data) +
  geom_point(aes(DOC_RESULT, NITRATE_N_RESULT, color = PTL_RESULT * 1000)) +
  scale_color_viridis_c("TP"~(mu~g~L^-1)) +
  theme_bw() +
  labs(title = "NRSA 2013-all",
       x = "DOC", y = "NITRATE")
p + geom_vline(aes(xintercept = q4)) + geom_vline(aes(xintercept = q5))
```

```
## Warning: Removed 21 rows containing missing values (geom_point).
```




```
ggplot(lowDOC_data) +
  geom_point(aes(DOC_RESULT, NITRATE_N_RESULT, color = PTL_RESULT * 1000)) +
  scale_color_viridis_c("TP"~(mu~g~L^-1)) +
  theme_bw() +
  labs(title = "NRSA 2013-low DOC only",
       x = "DOC", y = "NITRATE")
```

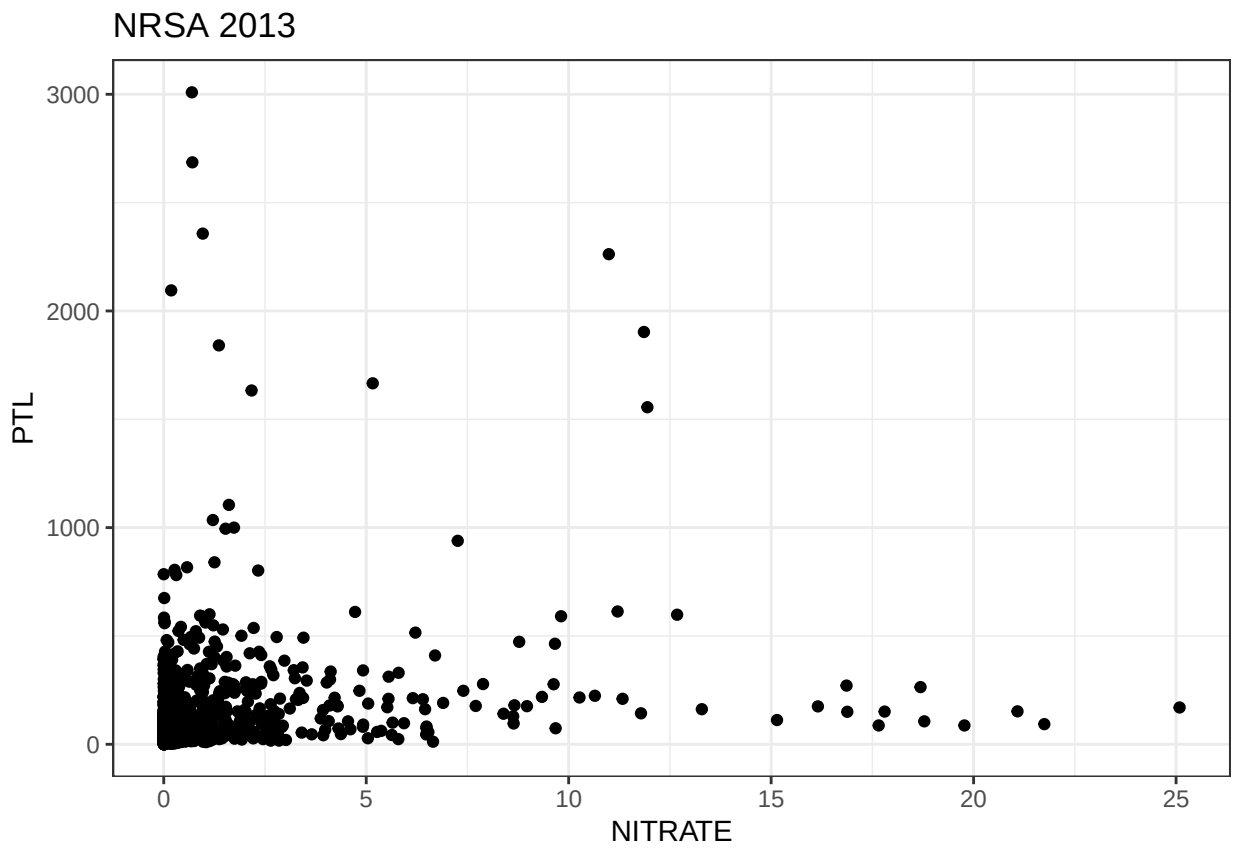
```
## Warning: Removed 4 rows containing missing values (geom_point).
```



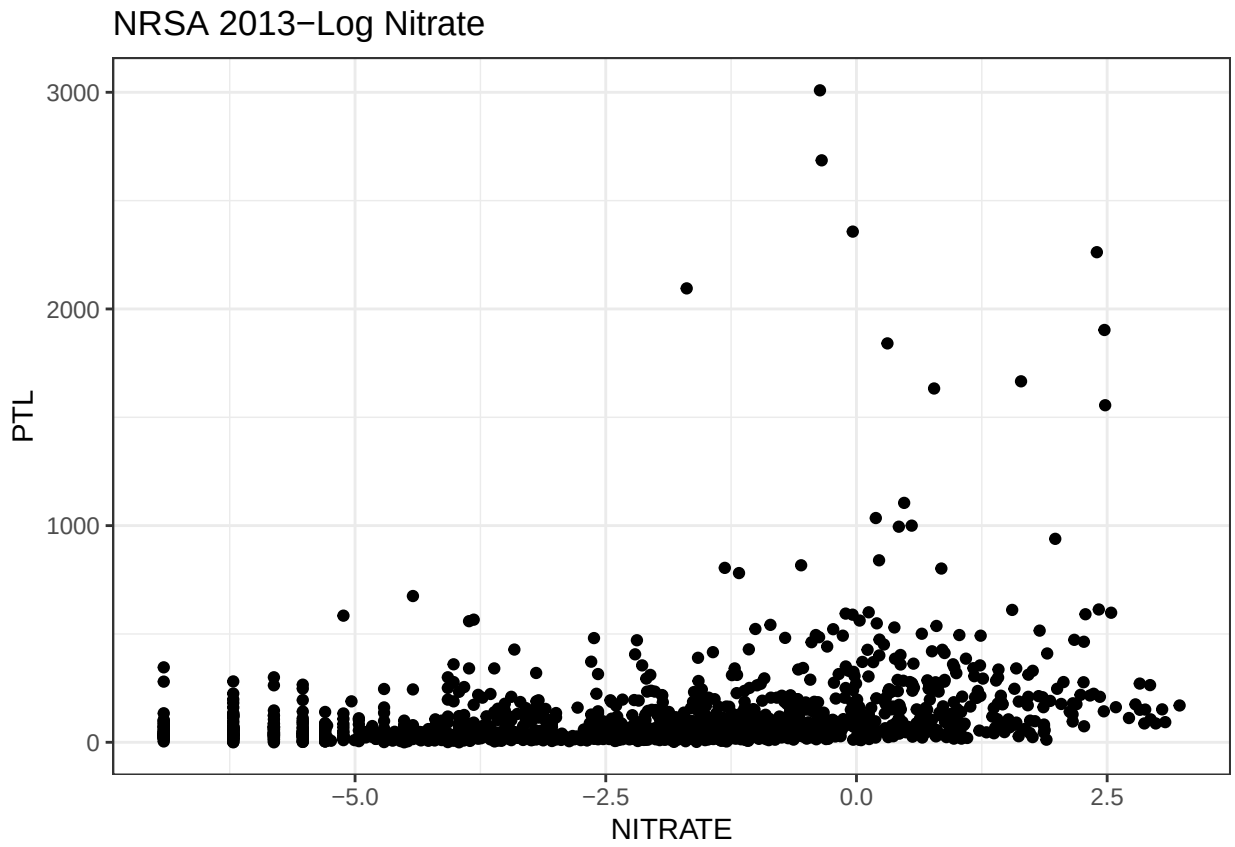
Note:

For now I am using the 4the quantile of this specific dataset. I realize that will change both where the quantile is/where we choose to cut off data

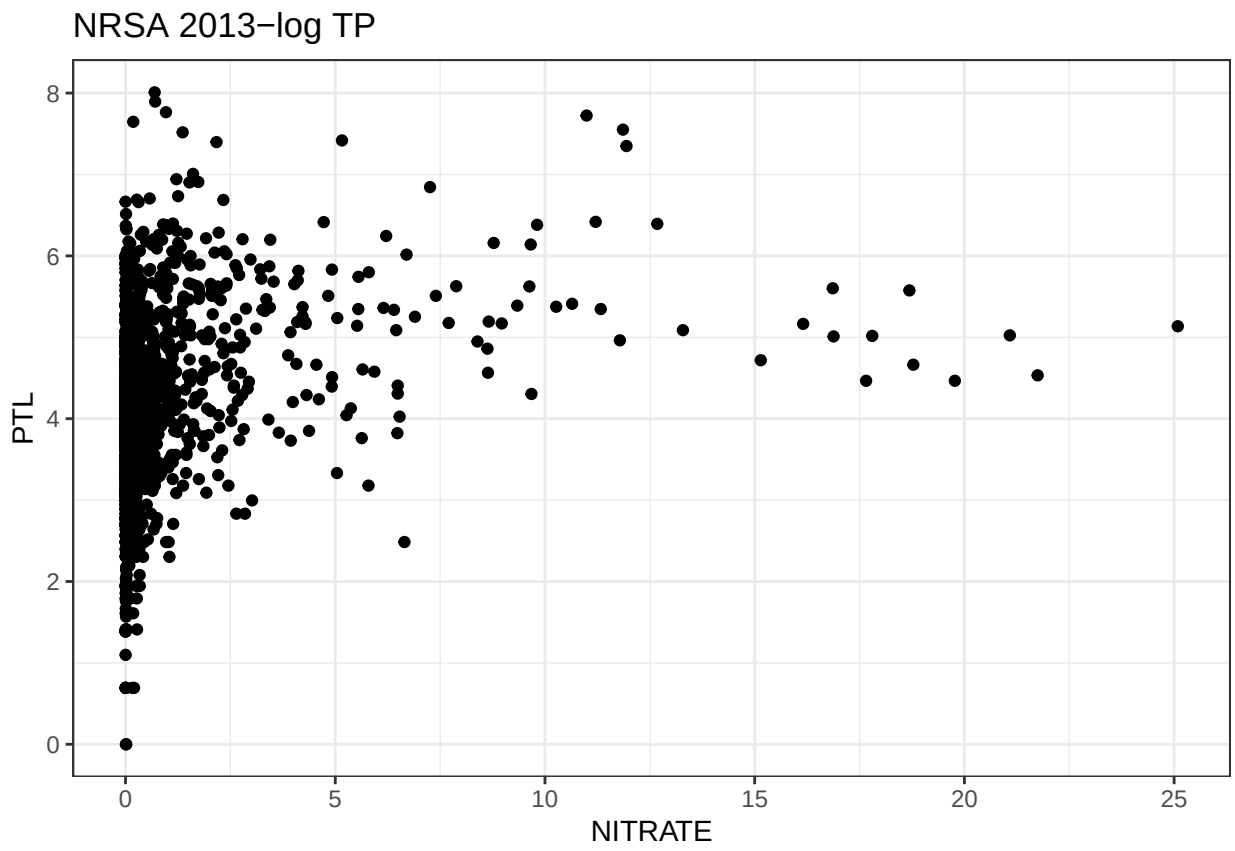
```
## Warning: Removed 4 rows containing missing values (geom_point).
```



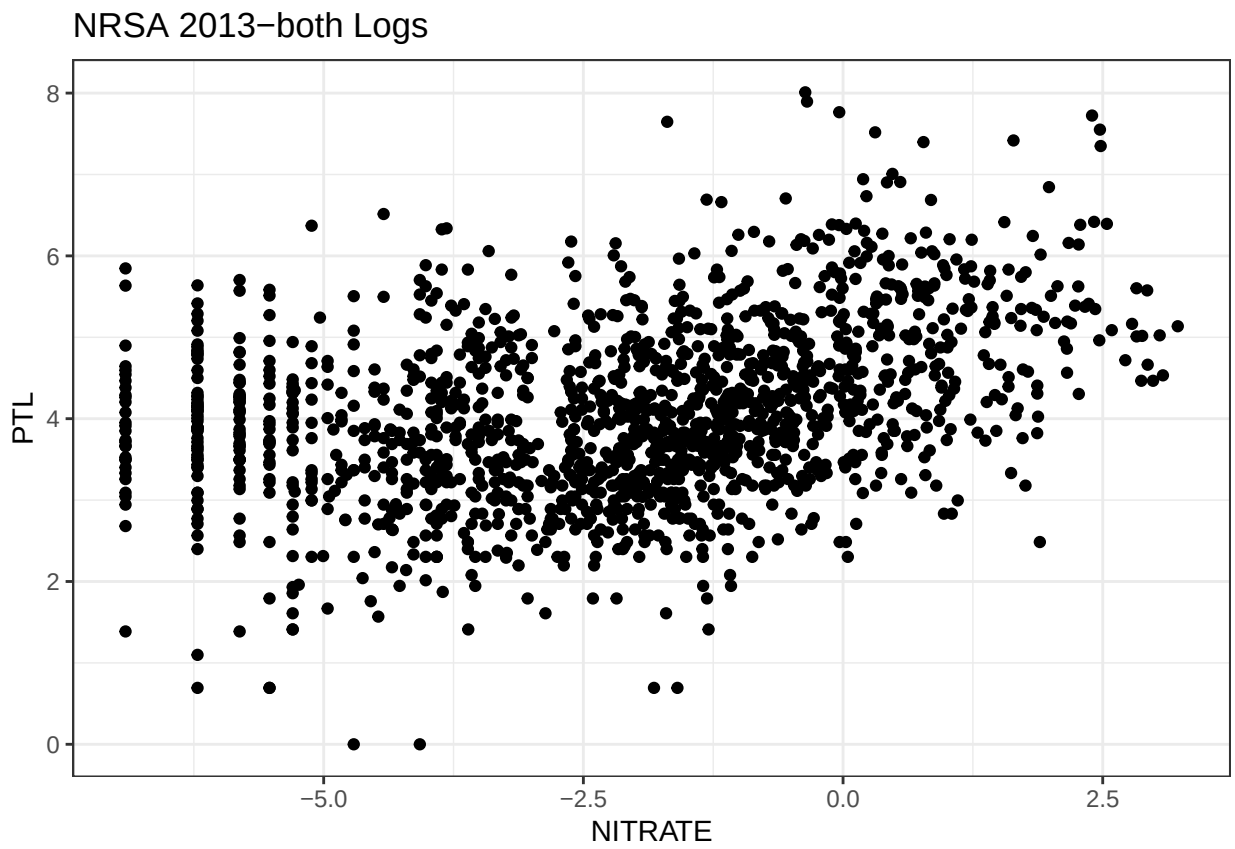

```
## Warning: Removed 150 rows containing missing values (geom_point).
```



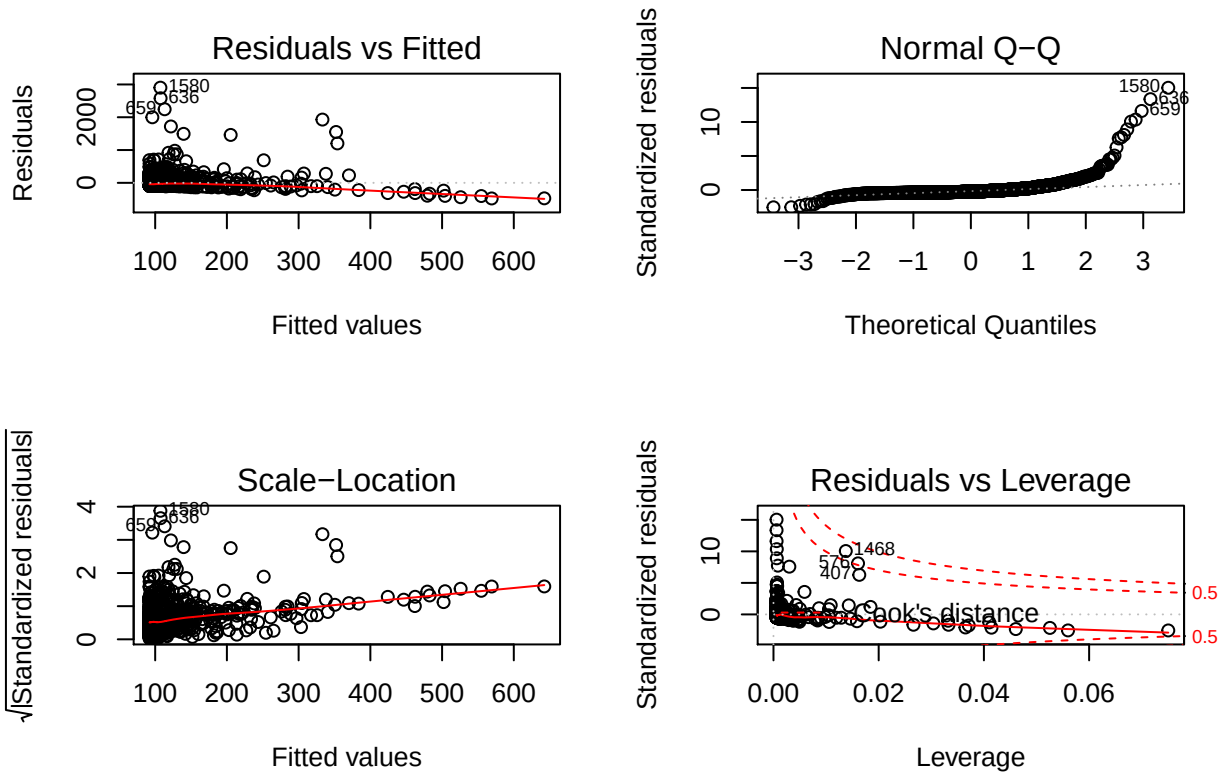

```
## Warning: Removed 9 rows containing missing values (geom_point).
```



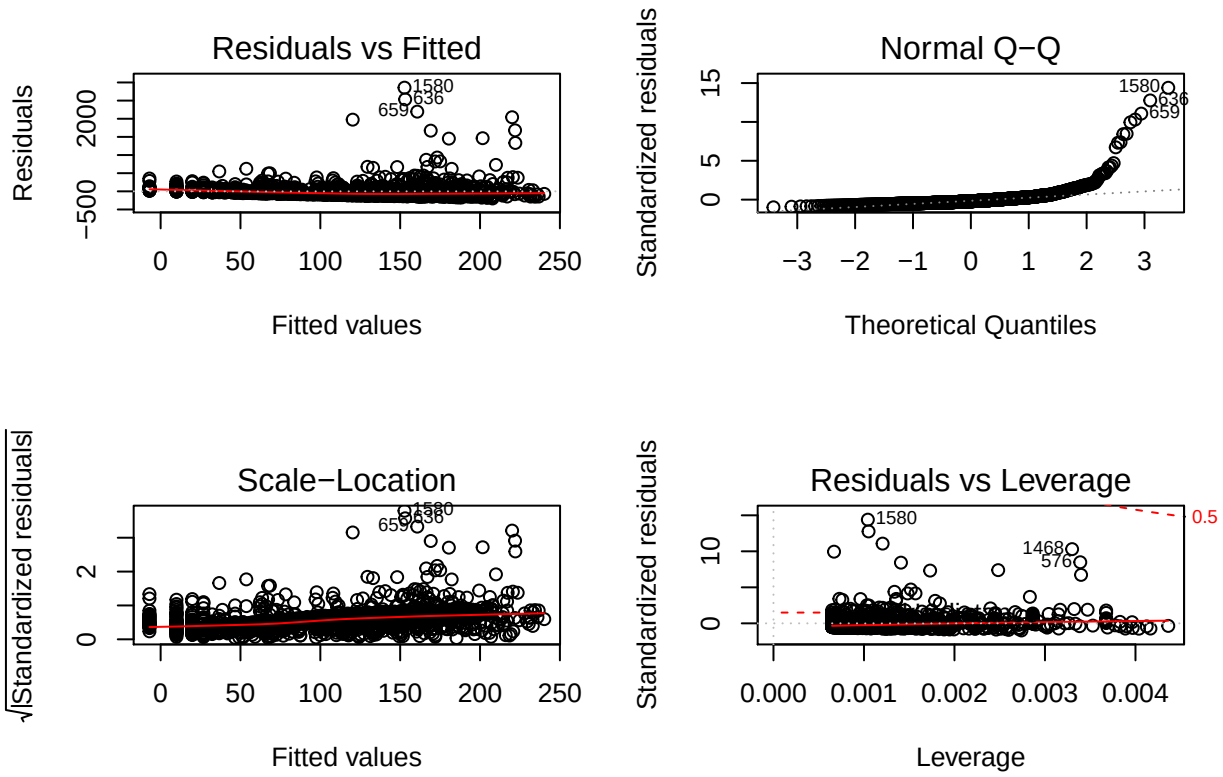

```
## Warning: Removed 154 rows containing missing values (geom_point).
```



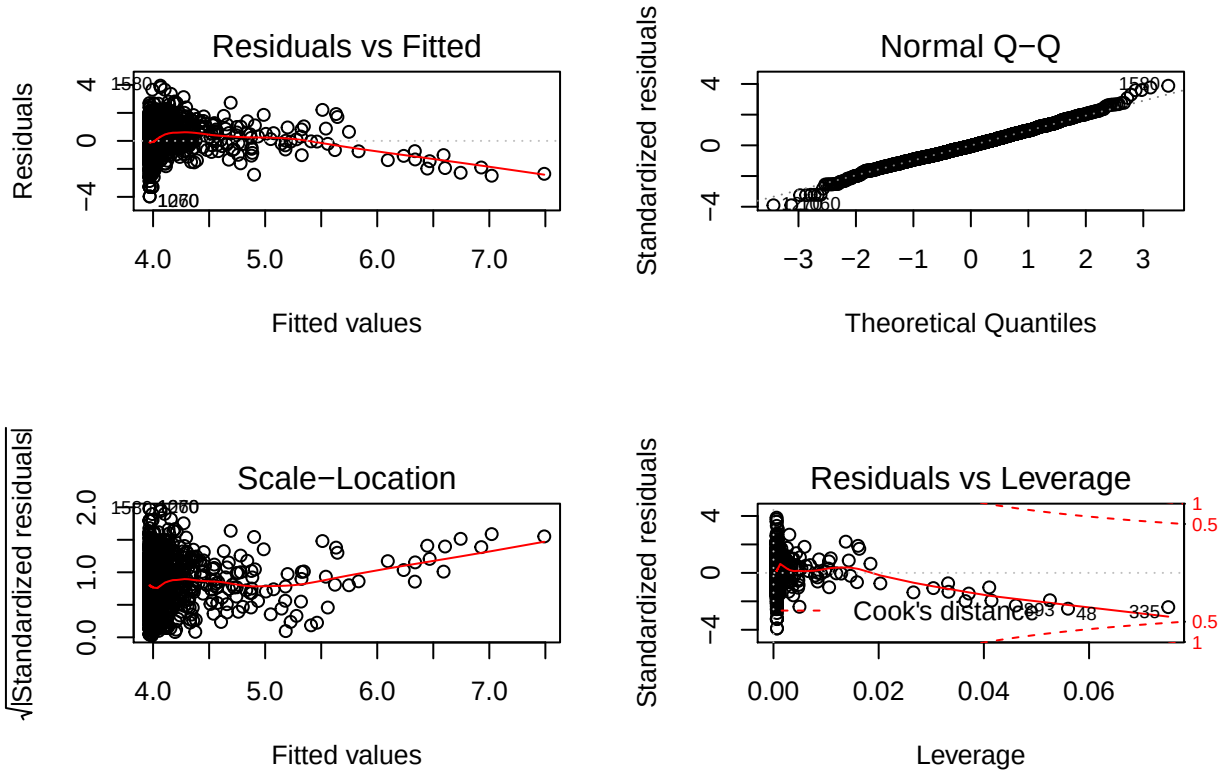

```
##
## Call:
## lm(formula = PTL_RESULT ~ NITRATE_N_RESULT, data = lowDOC_data)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -476.17  -70.09  -46.37    7.66 2901.71
##
## Coefficients:
##                  Estimate Std. Error t value Pr(>|t|)
## (Intercept)        92.042      4.985   18.46   <2e-16 ***
## NITRATE_N_RESULT   21.941      2.167   10.12   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 193.2 on 1689 degrees of freedom
##   (4 observations deleted due to missingness)
## Multiple R-squared:  0.0572, Adjusted R-squared:  0.05664
## F-statistic: 102.5 on 1 and 1689 DF,  p-value: < 2.2e-16
```



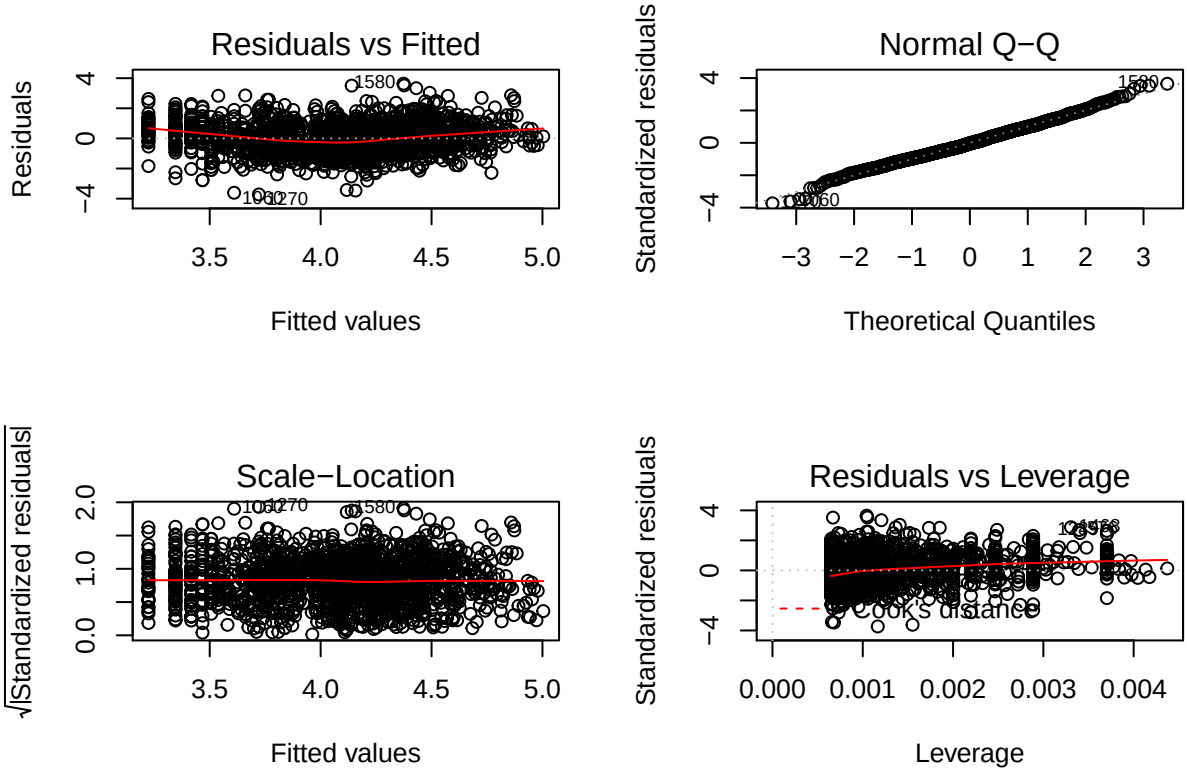

```
##
## Call:
## lm(formula = PTL_RESULT ~ log_Nitrate, data = lowDOC_data)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -195.82  -83.31  -41.45   24.17 2856.29
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  161.588      6.950   23.25   <2e-16 ***
## log_Nitrate   24.403      2.273   10.74   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 198.5 on 1543 degrees of freedom
##   (150 observations deleted due to missingness)
## Multiple R-squared:  0.06952,    Adjusted R-squared:  0.06891
## F-statistic: 115.3 on 1 and 1543 DF,  p-value: < 2.2e-16
```




```
##
## Call:
## lm(formula = log_TP ~ NITRATE_N_RESULT, data = lowDOC_data)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.9701 -0.6748 -0.0407  0.6569  3.9441
##
## Coefficients:
##                  Estimate Std. Error t value Pr(>|t|)
## (Intercept)       3.96773    0.02627  151.03   <2e-16 ***
## NITRATE_N_RESULT  0.14041    0.01141   12.31   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.016 on 1684 degrees of freedom
##   (9 observations deleted due to missingness)
## Multiple R-squared:  0.08256,    Adjusted R-squared:  0.08201
## F-statistic: 151.5 on 1 and 1684 DF,  p-value: < 2.2e-16
```




```
##
## Call:
## lm(formula = log_TP ~ log_Nitrate, data = lowDOC_data)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.7216 -0.6753 -0.0411  0.6801  3.6358
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  4.43750    0.03493  127.03   <2e-16 ***
## log_Nitrate  0.17571    0.01145   15.35   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.9975 on 1539 degrees of freedom
##   (154 observations deleted due to missingness)
## Multiple R-squared:  0.1327, Adjusted R-squared:  0.1321
## F-statistic: 235.5 on 1 and 1539 DF,  p-value: < 2.2e-16
```


Residuals vs Fitted
Residuals
Fitted values
Normal Q−Q
Standardized residuals
Theoretical Quantiles
Scale−Location
√|Standardized residuals|
Fitted values
Residuals vs Leverage
Standardized residuals
Cook's distance
Leverage