

Null Hypothesis

Sophie Wulfinf

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 before). 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$log_Nitrate == "-Inf"] <- NA
quantile(test_dataLOGS$DOC_RESULT)
```

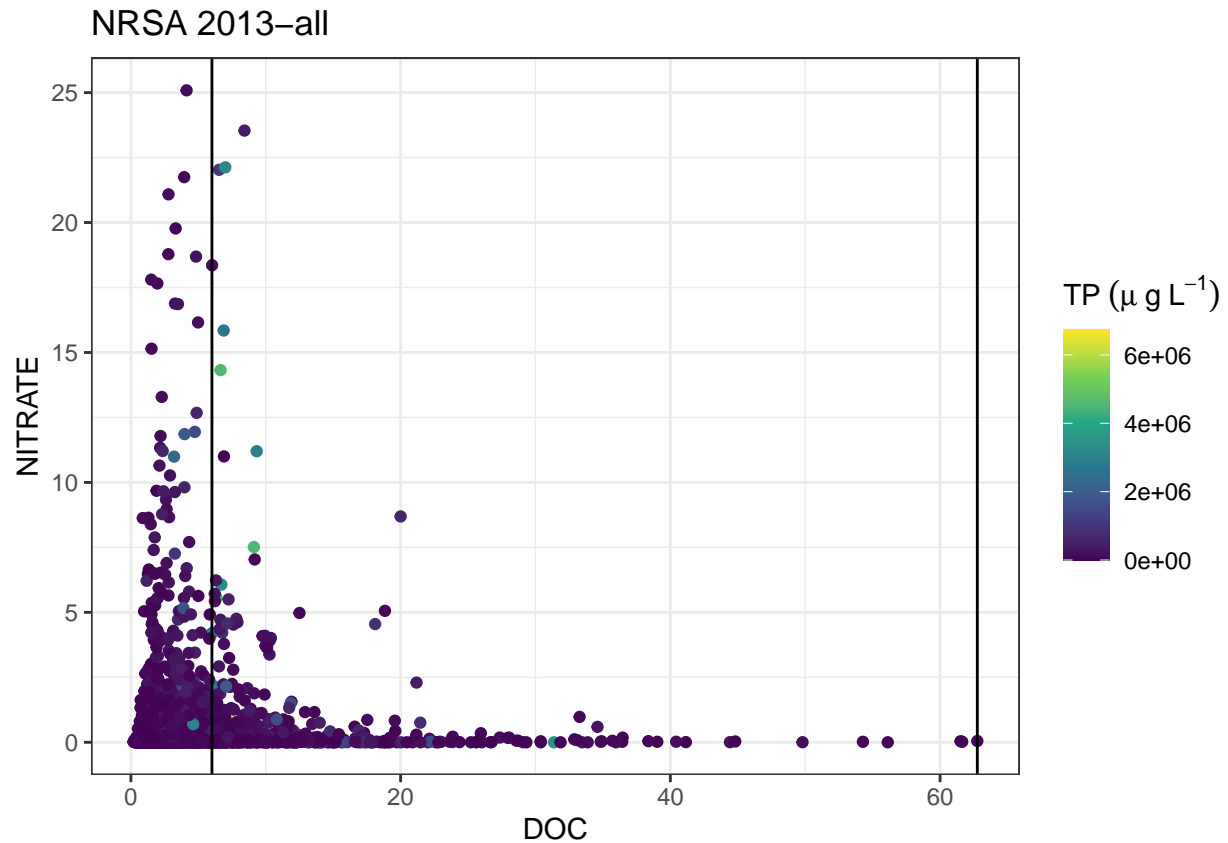
```
##      0%   25%   50%   75%  100%
## 0.19  1.81  3.42  6.01 62.76
```

```
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])

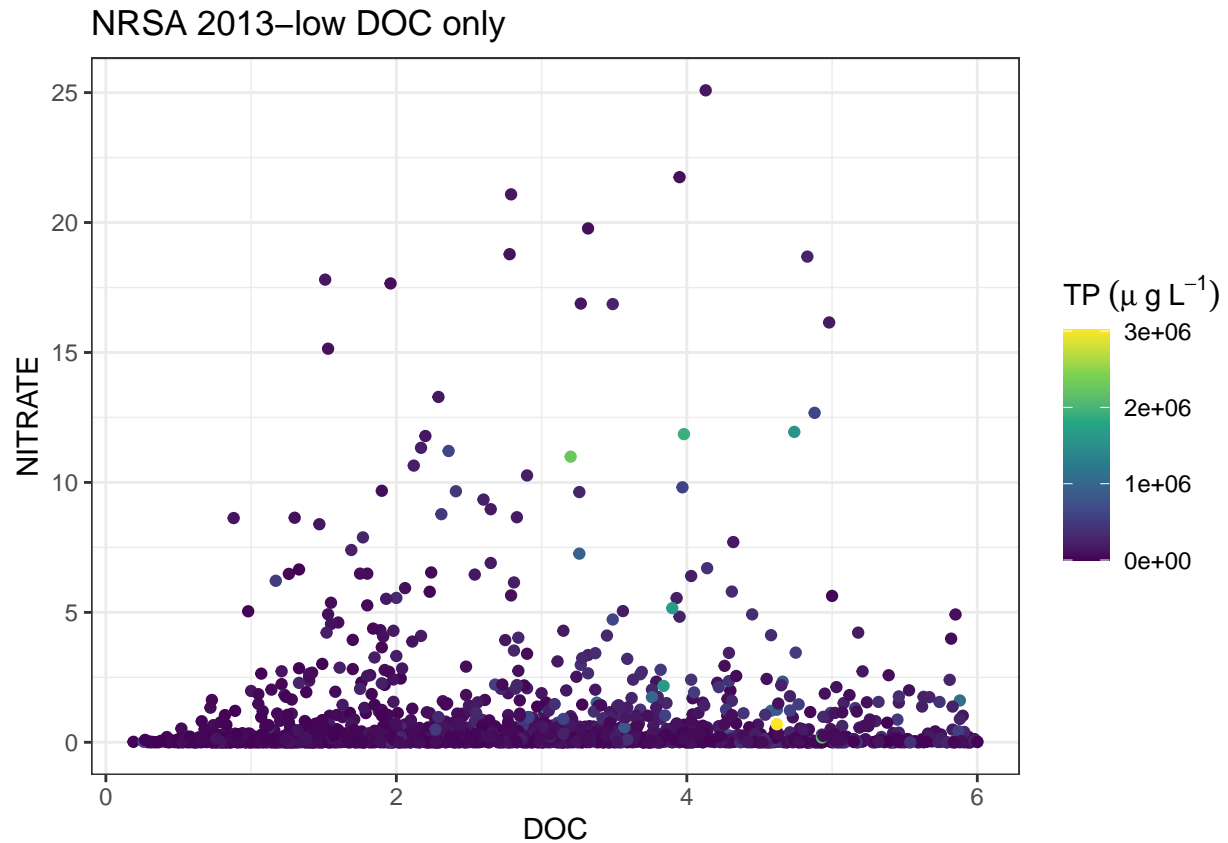
par(mfrow = c(1,2))
p <- ggplot(test_data) +
  geom_point(aes(DOC_RESULT, NITRATE_N_RESULT, color = PTL_RESULT * 1000)) +
  scale_color_viridis_c("TP" ~ (mu ~ g ~ L ^ -1)) +
  theme_bw() +
  labs(title = "NRSA 2013-all",
       x = "DOC", y = "NITRATE")
p + geom_vline(aes(xintercept = q4)) + geom_vline(aes(xintercept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

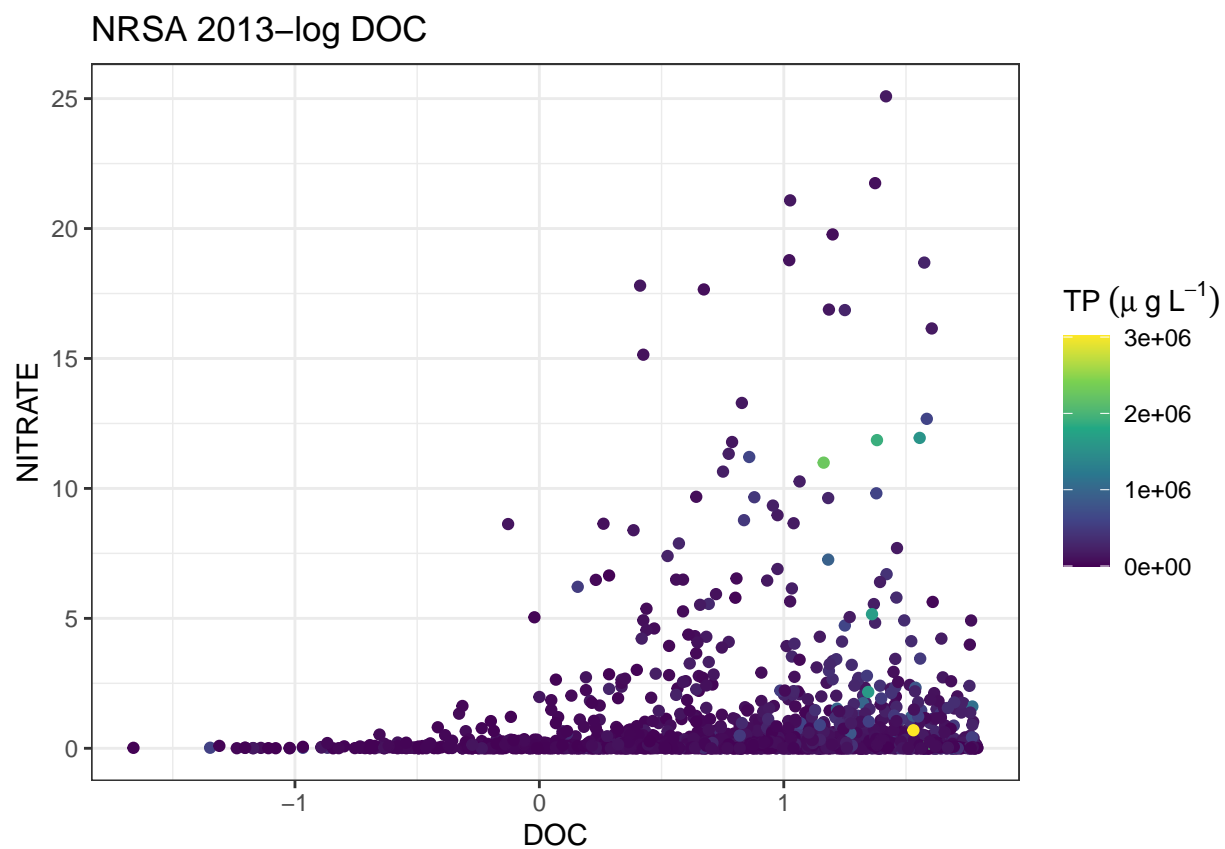
Warning: Removed 4 rows containing missing values (geom_point).



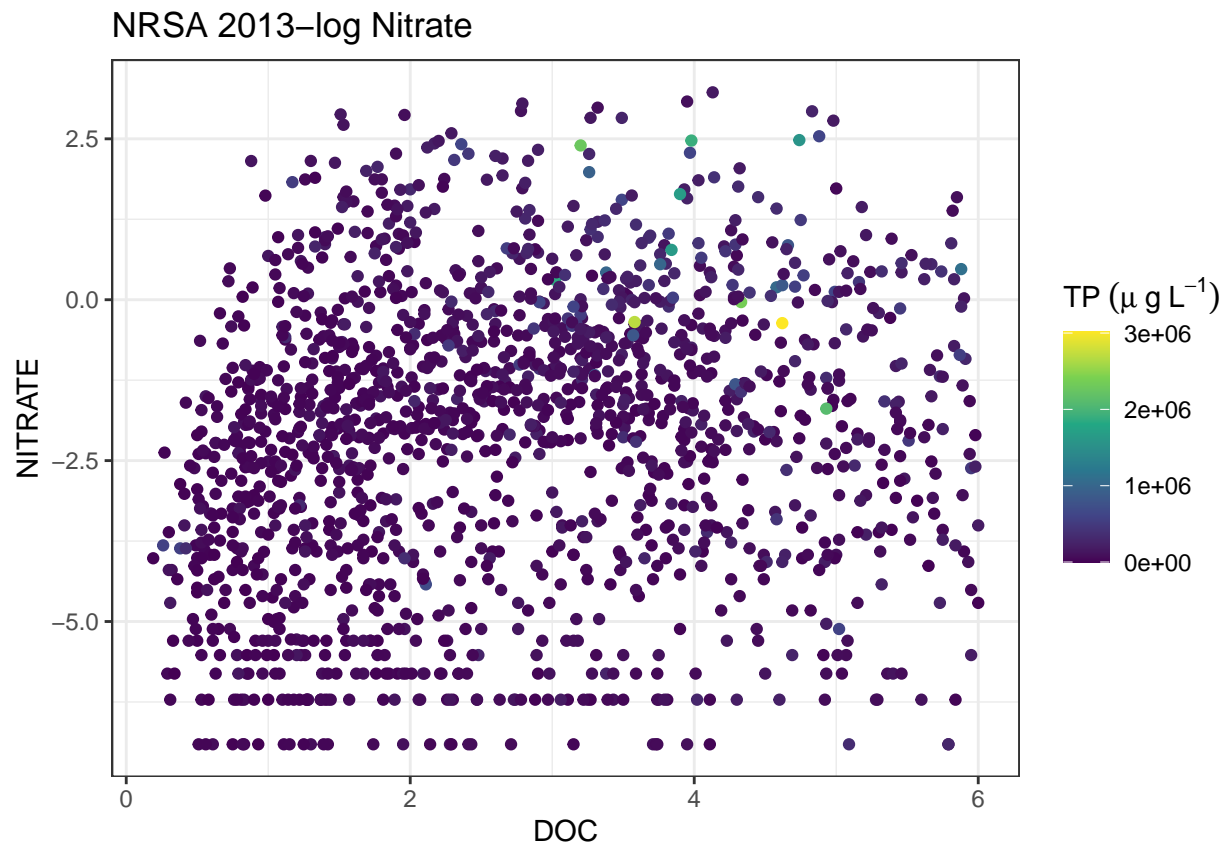
Note:

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

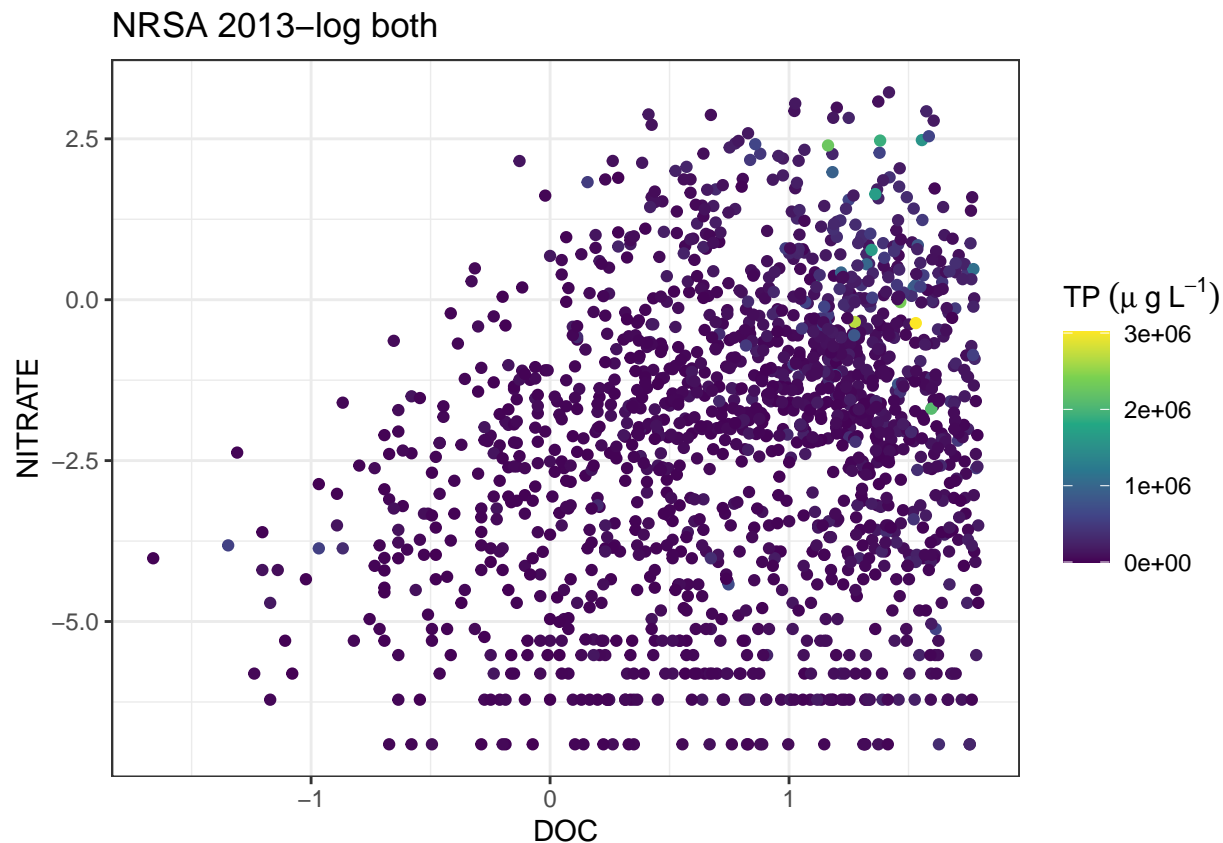
```
## Warning: Removed 4 rows containing missing values (geom_point).
```



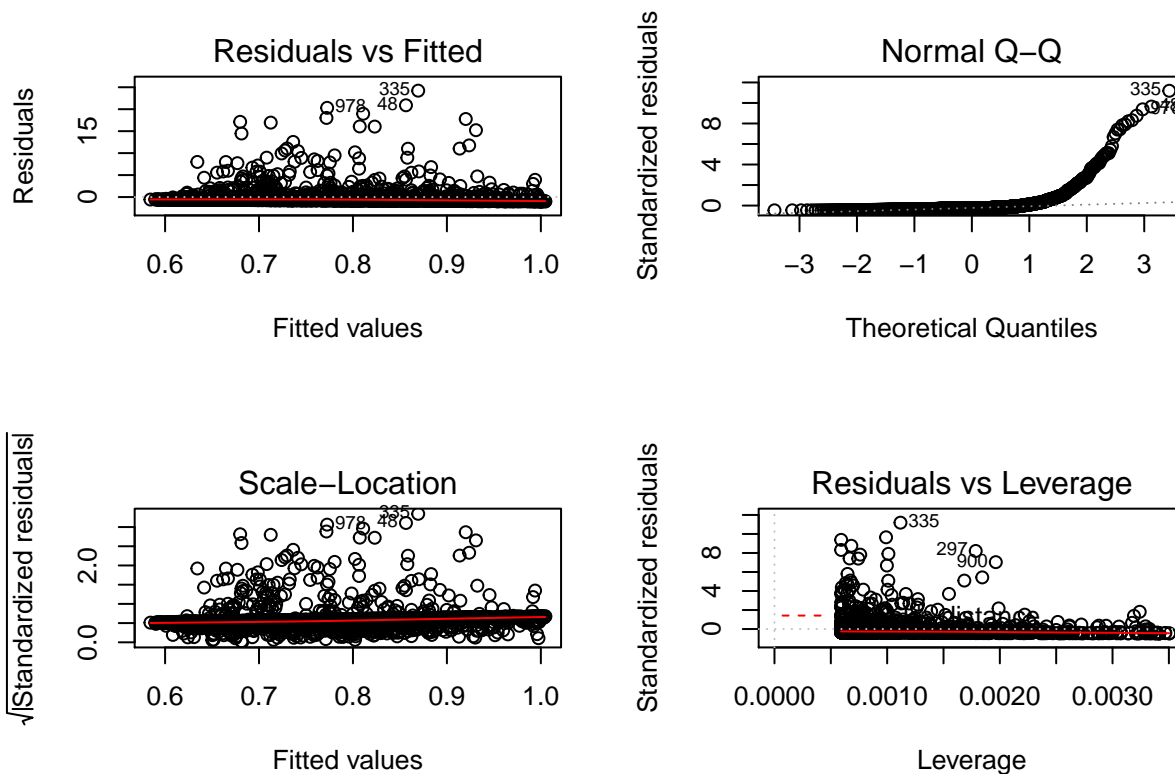
Warning: Removed 150 rows containing missing values (geom_point).



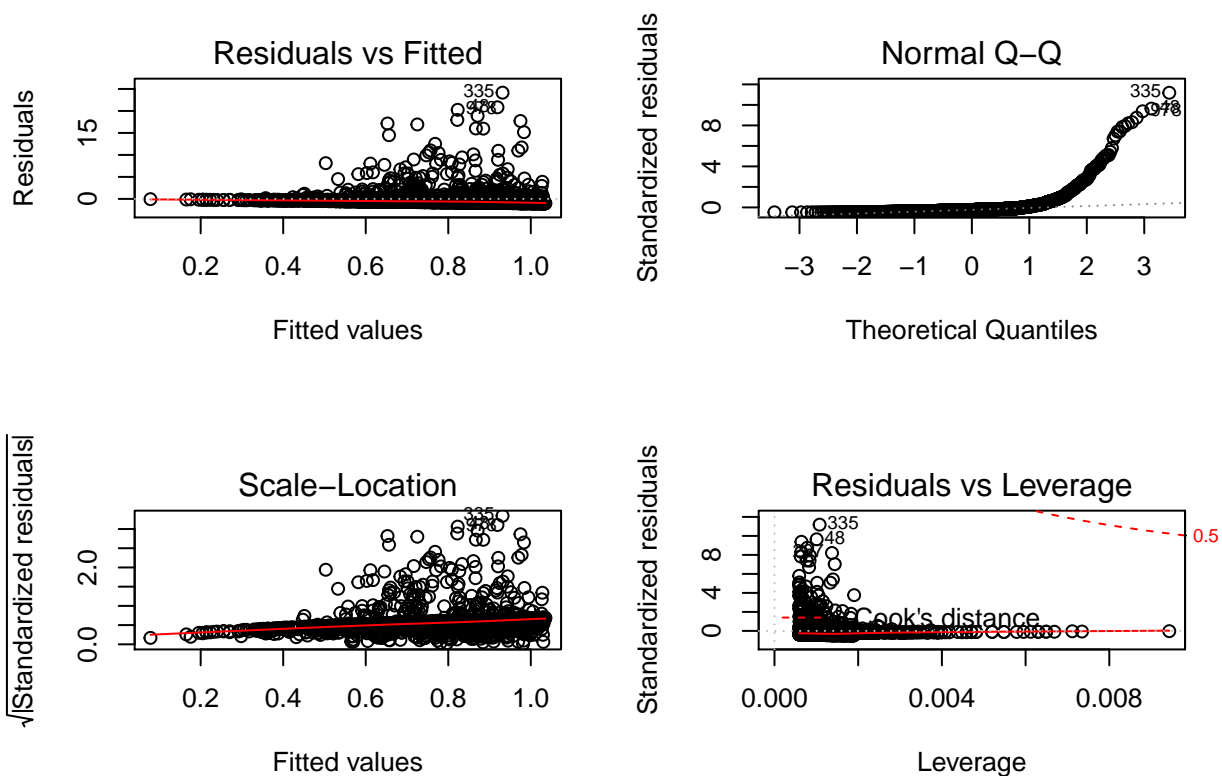
Warning: Removed 150 rows containing missing values (geom_point).



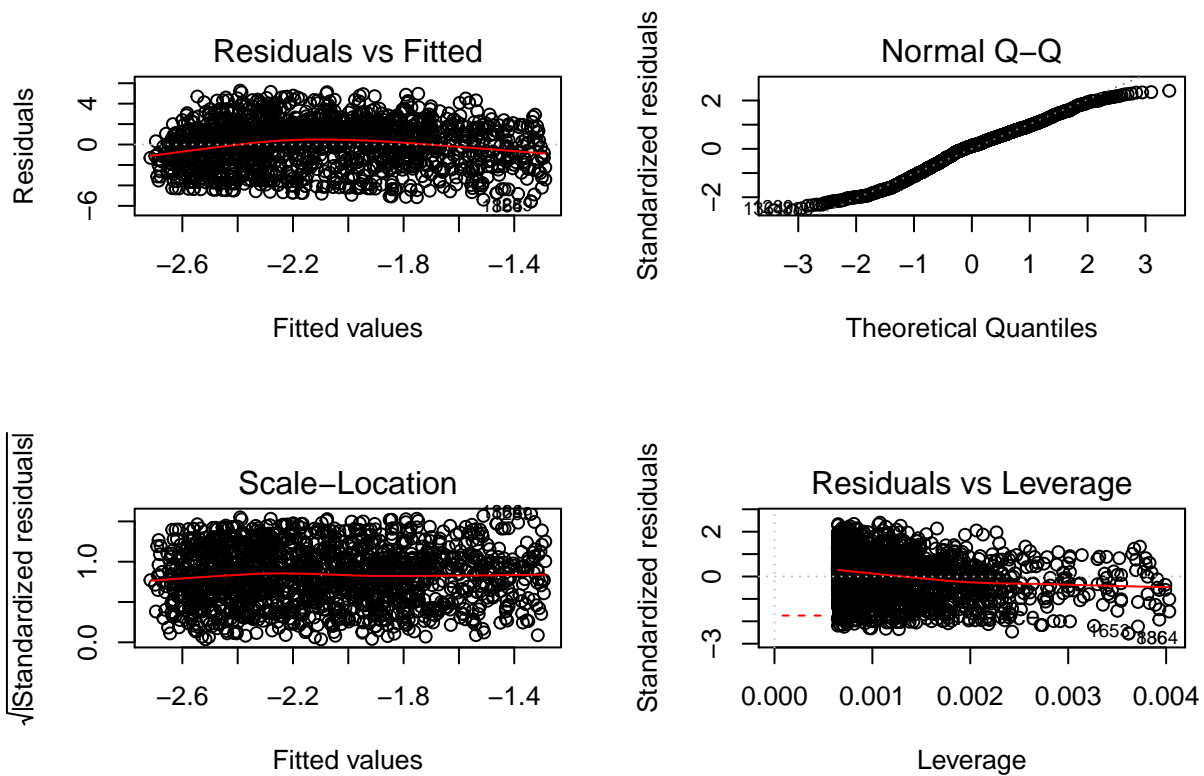
```
##
## Call:
## lm(formula = NITRATE_N_RESULT ~ DOC_RESULT, data = lowDOC_data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.0011 -0.7271 -0.5937 -0.2642  24.2176
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.57070    0.11180   5.105 3.69e-07 ***
## DOC_RESULT   0.07234    0.03592   2.014  0.0442 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.166 on 1689 degrees of freedom
## (4 observations deleted due to missingness)
## Multiple R-squared:  0.002396, Adjusted R-squared:  0.001805
## F-statistic: 4.056 on 1 and 1689 DF, p-value: 0.04417
```



```
##
## Call:
## lm(formula = NITRATE_N_RESULT ~ log_DOC, data = lowDOC_data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.0328 -0.7660 -0.5477 -0.2348  24.1554
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   0.53856    0.08584   6.274 4.45e-10 ***
## log_DOC       0.27711    0.08150   3.400 0.000689 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.161 on 1689 degrees of freedom
## (4 observations deleted due to missingness)
## Multiple R-squared:  0.006798, Adjusted R-squared:  0.00621
## F-statistic: 11.56 on 1 and 1689 DF, p-value: 0.0006892
```



```
##
## Call:
## lm(formula = log_Nitrate ~ DOC_RESULT, data = lowDOC_data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -5.5684 -1.5163  0.2724  1.5141  5.2703
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -2.76199    0.11783  -23.440  < 2e-16 ***
## DOC_RESULT   0.24570    0.03858   6.369  2.5e-10 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.195 on 1543 degrees of freedom
## (150 observations deleted due to missingness)
## Multiple R-squared:  0.02562,    Adjusted R-squared:  0.02499
## F-statistic: 40.57 on 1 and 1543 DF,  p-value: 2.5e-10
```

```
##
## Call:
## lm(formula = log_Nitrate ~ log_DOC, data = lowDOC_data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -5.5115 -1.5031  0.2327  1.4955  5.2783
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -2.70642    0.08877  -30.49  <2e-16 ***
## log_DOC      0.74607    0.08566   8.71  <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.171 on 1543 degrees of freedom
## (150 observations deleted due to missingness)
## Multiple R-squared:  0.04686,    Adjusted R-squared:  0.04624
## F-statistic: 75.86 on 1 and 1543 DF,  p-value: < 2.2e-16
```

