Explore the relationship between transit time and number of commutes for subways and roads

```
library(tidyverse)
```

```
## -- Attaching packages --------------------------------------- tidyverse 1.3.2 --
## v ggplot2 3.4.0      v purrr   0.3.5
## v tibble  3.1.8      v dplyr   1.0.10
## v tidyr   1.2.1      v stringr 1.4.1
## v readr   2.1.3      v forcats 0.5.2
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
driving <- readr::read_csv('data/nta-driving-weights.csv')
```

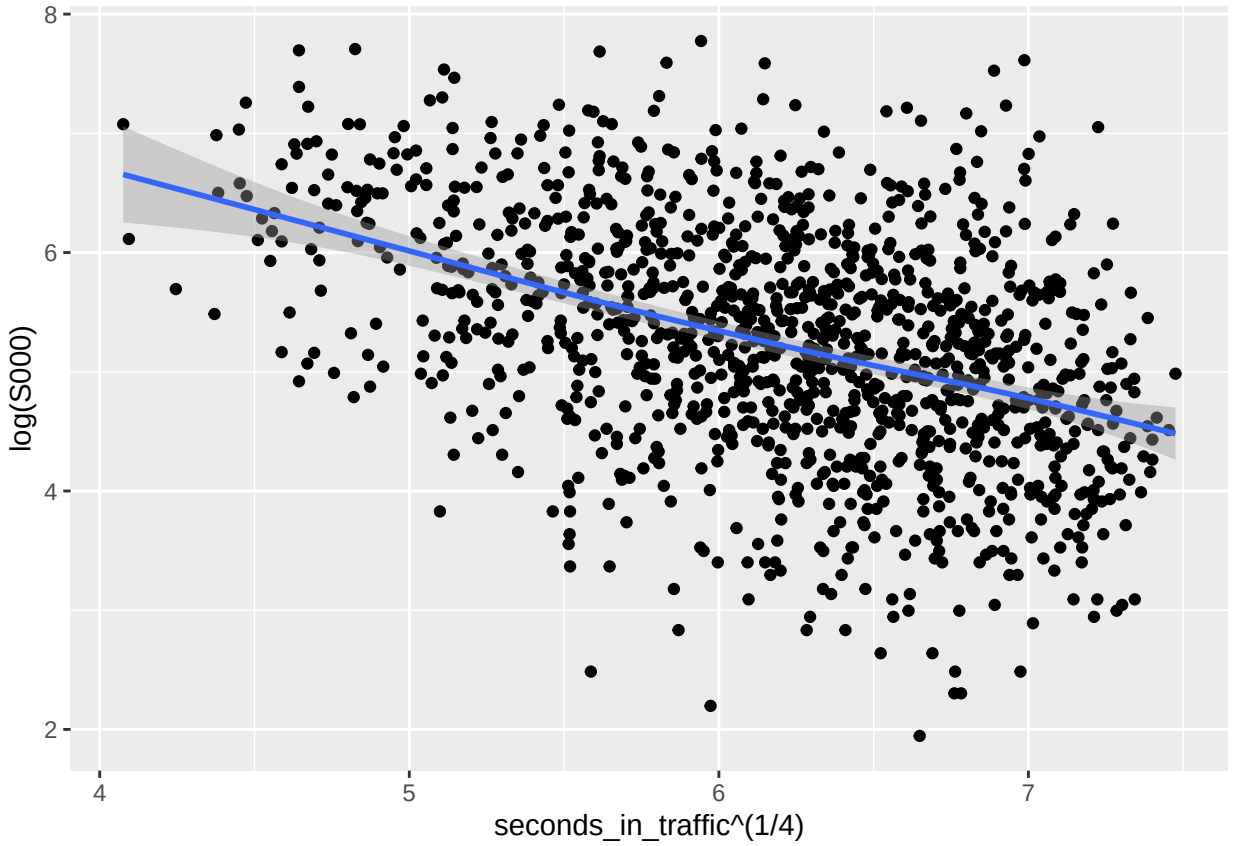
```
## Rows: 1225 Columns: 5
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr (3): trip, nta_one, nta_two
## dbl (2): S000, seconds_in_traffic
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```
subway <- readr::read_csv('data/nta-subway-weights.csv')
```

```
## Rows: 1225 Columns: 5
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr (3): trip, nta_one, nta_two
## dbl (2): S000, seconds_in_transit
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```
ggplot(data = driving, aes(x = seconds_in_traffic**(1/4), y = log(S000))) +
  geom_point() +
  stat_smooth()
```

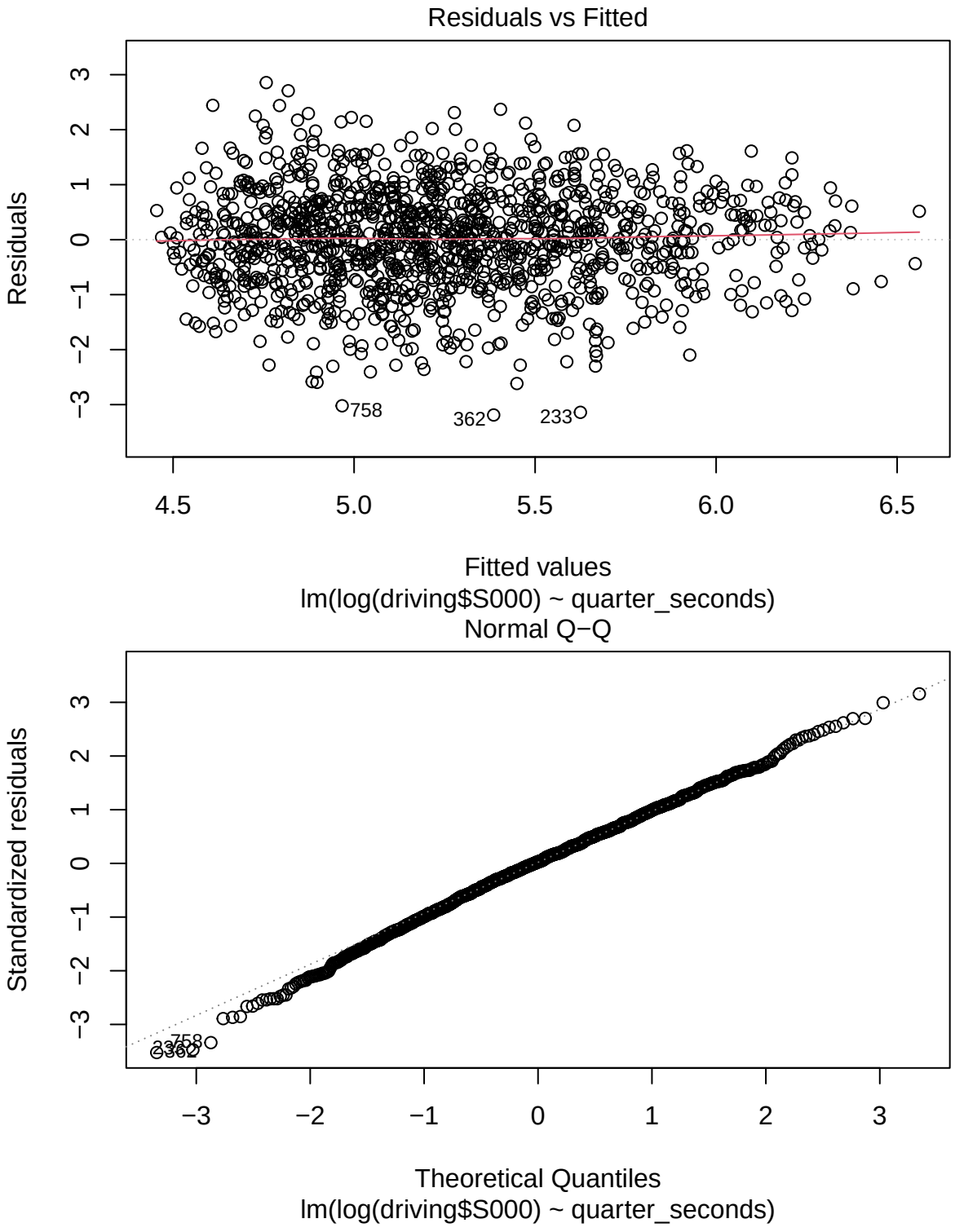
```
## `geom_smooth()` using method = 'gam' and formula = 'y ~ s(x, bs = "cs")'
```

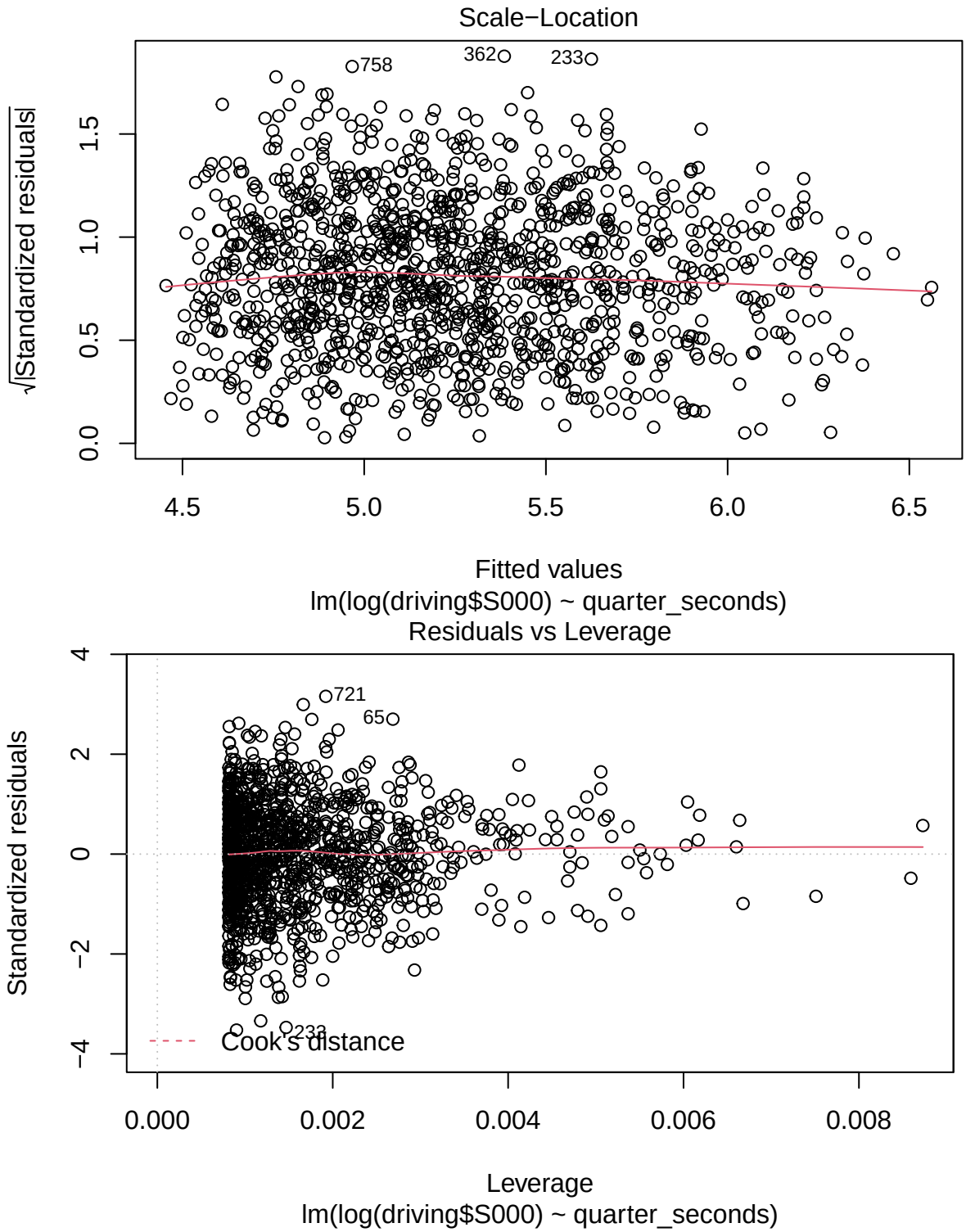



```
quarter_seconds <- driving$seconds_in_traffic^(1/4)
driving_model <- lm(log(driving$S000) ~ quarter_seconds)
summary(driving_model)
```

```
##
## Call:
## lm(formula = log(driving$S000) ~ quarter_seconds)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.1883 -0.5621  0.0265  0.5979  2.8560
##
## Coefficients:
##                 Estimate Std. Error t value Pr(>|t|)
## (Intercept)       9.0874     0.2368   38.37   <2e-16 ***
## quarter_seconds  -0.6197     0.0380  -16.31   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.9053 on 1223 degrees of freedom
## Multiple R-squared:  0.1786, Adjusted R-squared:  0.178
## F-statistic:   266 on 1 and 1223 DF,  p-value: < 2.2e-16
```

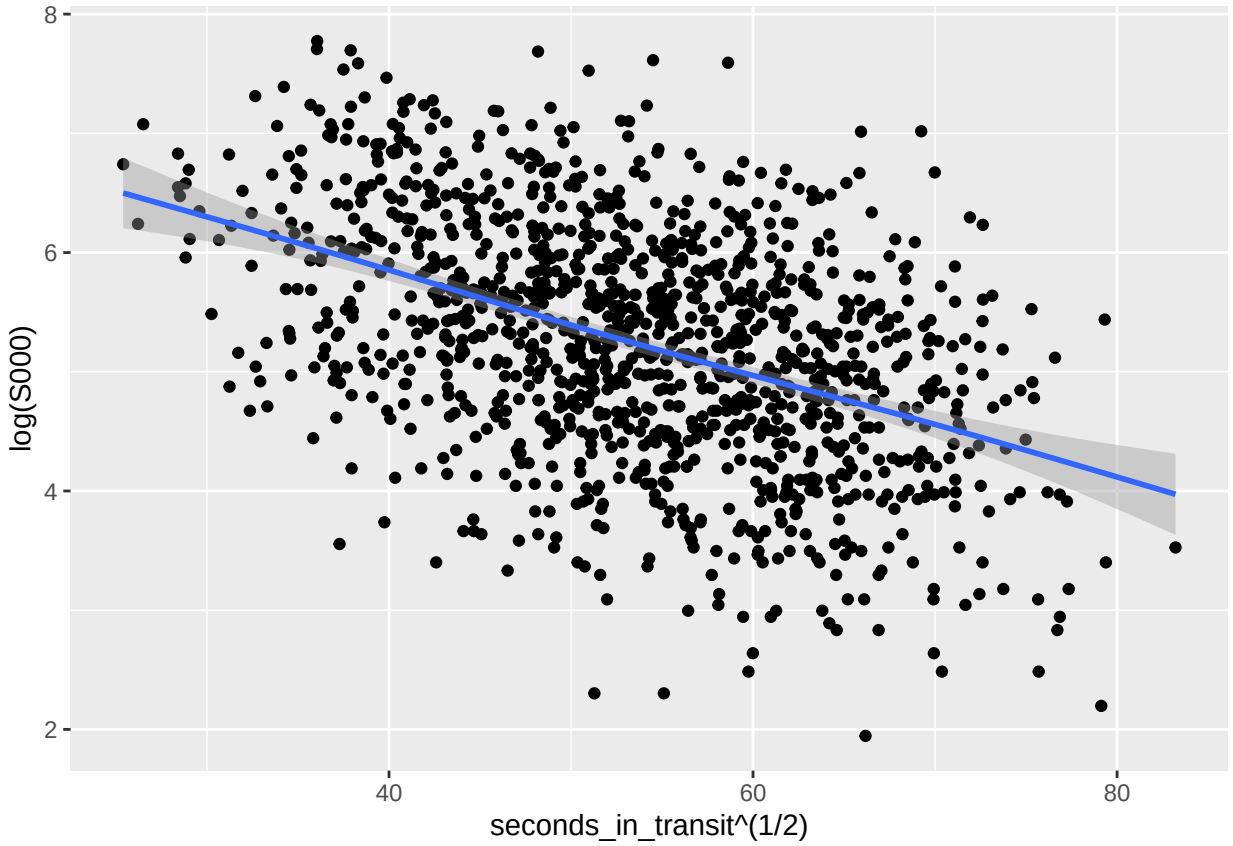
```
plot(driving_model)
```


Residuals vs Fitted
Residuals
Fitted values
lm(log(driving$S000) ~ quarter_seconds)
Normal Q−Q
Standardized residuals
Theoretical Quantiles
lm(log(driving$S000) ~ quarter_seconds)




```
ggplot(data = subway, aes(x = seconds_in_transit^(1/2), y = log(S000))) +
  geom_point() +
  stat_smooth()
```

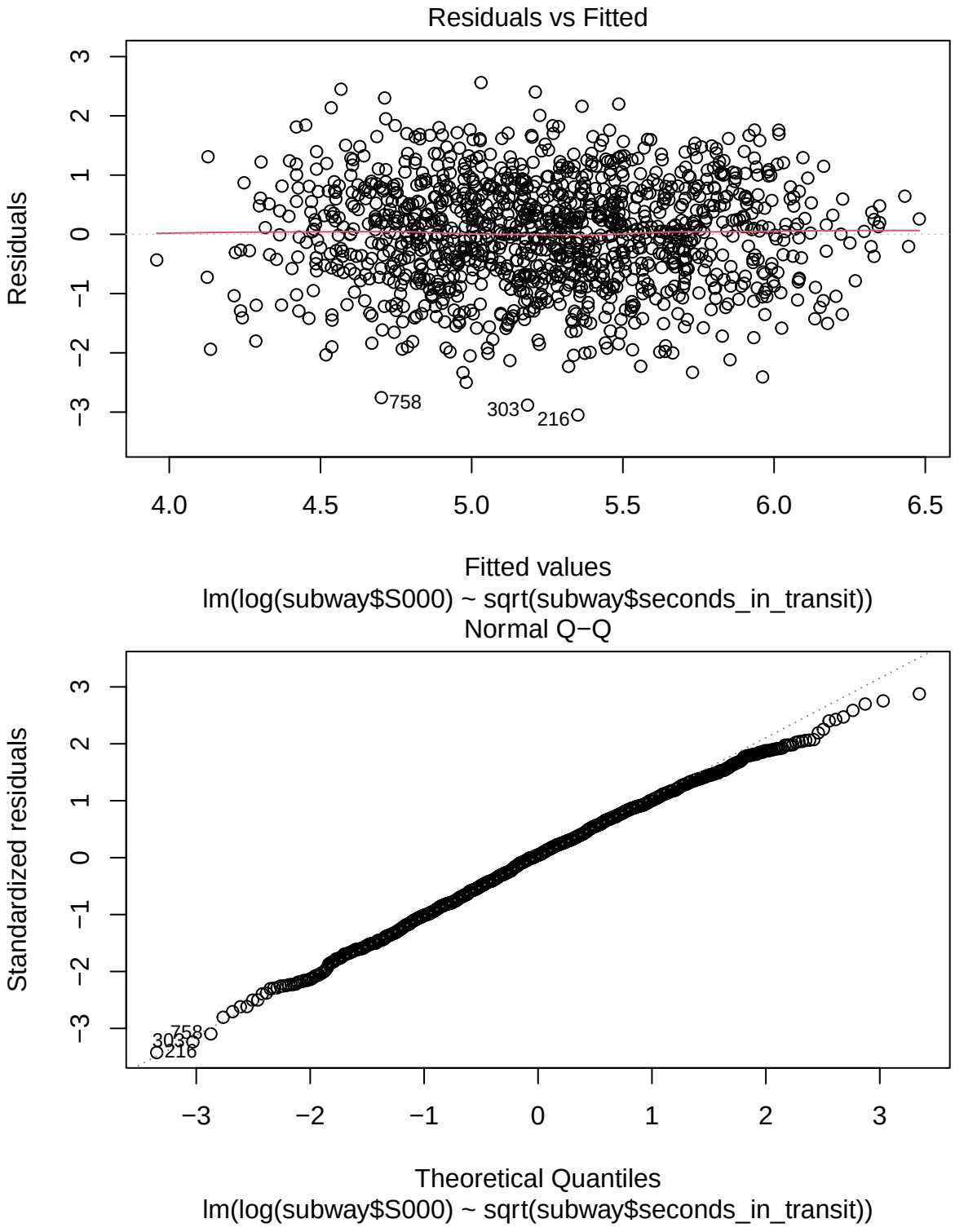
```
## `geom_smooth()` using method = 'gam' and formula = 'y ~ s(x, bs = "cs")'
```

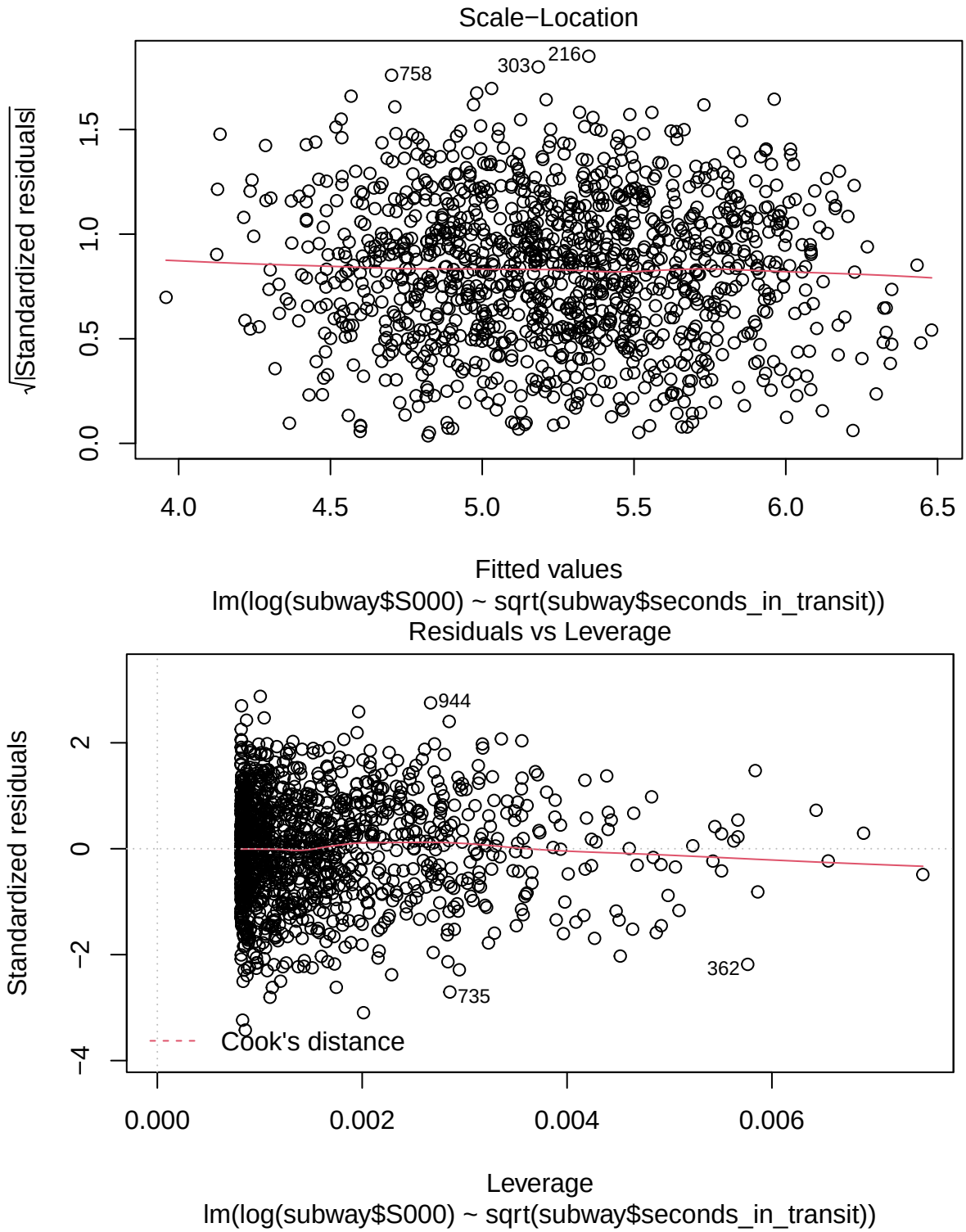



```
subway_model <- lm(log(subway$S000) ~ sqrt(subway$seconds_in_transit))
summary(subway_model)
```

```
##
## Call:
## lm(formula = log(subway$S000) ~ sqrt(subway$seconds_in_transit))
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -3.04855 -0.61476  0.04009  0.64111  2.56151
##
## Coefficients:
##                                  Estimate Std. Error t value Pr(>|t|)
## (Intercept)                      7.588132   0.134318   56.49   <2e-16 ***
## sqrt(subway$seconds_in_transit) -0.043620   0.002458  -17.74   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.8908 on 1223 degrees of freedom
## Multiple R-squared:  0.2047, Adjusted R-squared:  0.2041
## F-statistic: 314.8 on 1 and 1223 DF,  p-value: < 2.2e-16
```

```
plot(subway_model)
```


Residuals vs Fitted
Residuals
Fitted values
lm(log(subway$S000) ~ sqrt(subway$seconds_in_transit))
Normal Q−Q
Standardized residuals
Theoretical Quantiles
lm(log(subway$S000) ~ sqrt(subway$seconds_in_transit))
758
303
216




```
subway_and_driving_model <- lm(log(subway$S000) ~ subway$seconds_in_transit + driving$seconds_in_traffi
summary(subway_and_driving_model)
```

```
##
## Call:
## lm(formula = log(subway$S000) ~ subway$seconds_in_transit + driving$seconds_in_traffic)
##
```

```
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.86636 -0.58987  0.03693  0.61302  2.64418
##
## Coefficients:
##                              Estimate Std. Error t value Pr(>|t|)
## (Intercept)                 6.647e+00  7.699e-02  86.337  < 2e-16 ***
## subway$seconds_in_transit  -2.835e-04  2.819e-05 -10.057  < 2e-16 ***
## driving$seconds_in_traffic -3.508e-04  4.984e-05  -7.039 3.23e-12 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.8756 on 1222 degrees of freedom
## Multiple R-squared:  0.2321, Adjusted R-squared:  0.2309
## F-statistic: 184.7 on 2 and 1222 DF,  p-value: < 2.2e-16
```

```
summary(subway$seconds_in_transit)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##     645    2146    2892    2985    3781    6924
```

```
summary(driving$seconds_in_traffic)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##     276    1076    1547    1576    2053    3122
```