Lab 2: Impact of Stay at Home Orders on Individual Mobility

w203: Statistics for Data Science

Elaine Chang, Dom Dillingham, Jesse Miller, Michael Wang

1. Introduction

2. Model Building

EDA

```
library(dplyr)
```

```
##
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':
##
##     filter, lag
```

```
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
library(magrittr)
library(ggplot2)
library(openxlsx)
library(lmtest)
```

```
## Loading required package: zoo
```

```
##
## Attaching package: 'zoo'
```

```
## The following objects are masked from 'package:base':
##
##     as.Date, as.Date.numeric
```

```
library(sandwich)
```

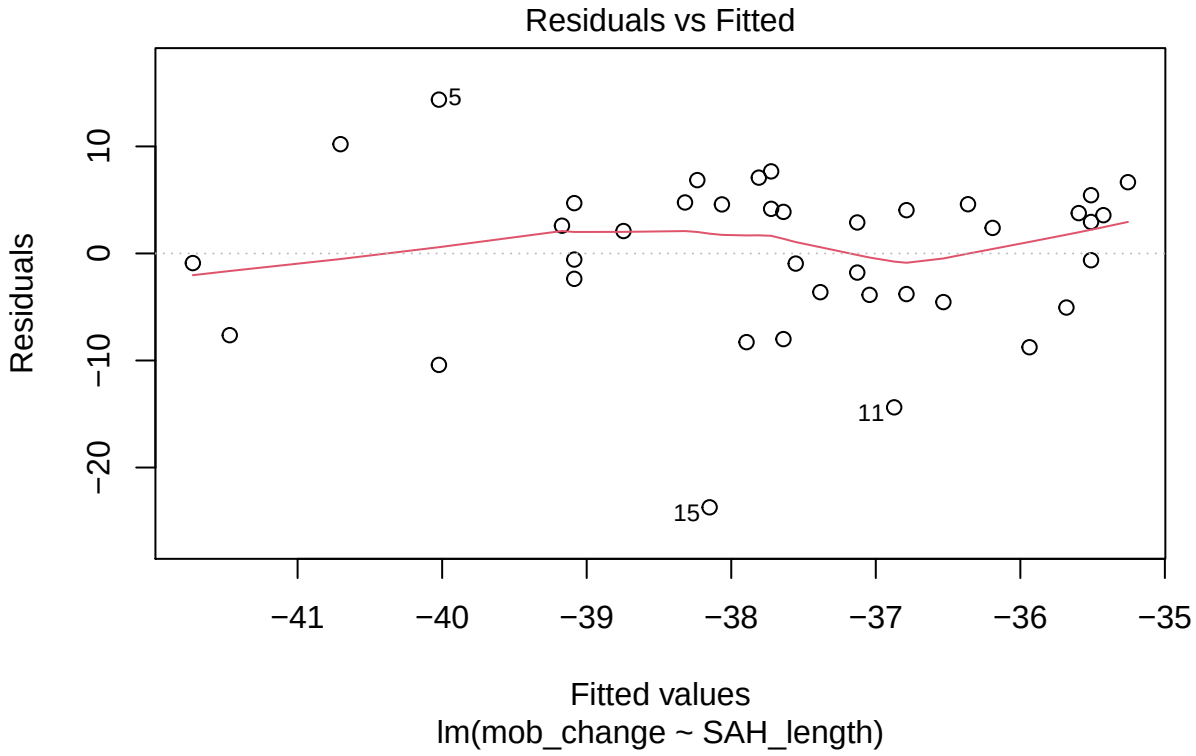
```
sah = read.csv(file = 'SAH_data.csv')
sah$SAH_start <- as.Date(sah$SAH_start, format="%m/%d/%Y")
sah$SAH_end <- as.Date(sah$SAH_end, format="%m/%d/%Y")
sah = sah[sah$SAH == 1, ]
# sah

mob = read.csv(file = 'Region_Mobility_Report_CSVs/2020_US_Region_Mobility_Report.csv')
mob = mob[mob$sub_region_2 == '',]
mob$date <- as.Date(mob$date, format="%Y-%m-%d")
# mob

# mob[mob$sub_region_1 == 'California',]
```

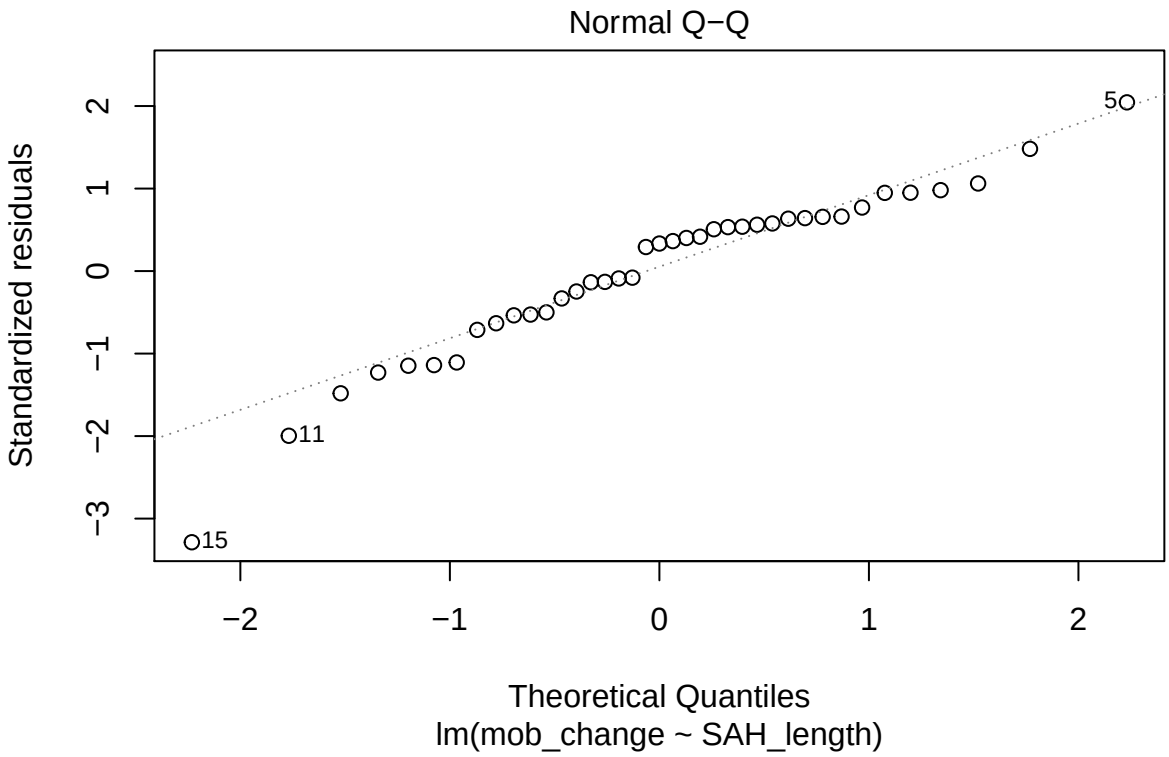
```
mob_change = c()
for (row in 1:nrow(sah)) {
  state = sah[row, 'State']
  start = sah[row, 'SAH_start']
  end = sah[row, 'SAH_end']
  mob_change[row] = mean(mob[mob$sub_region_1 == state & mob$date >= start & mob$date <= end, 'retail_a
}


sah$mob_change = mob_change
```

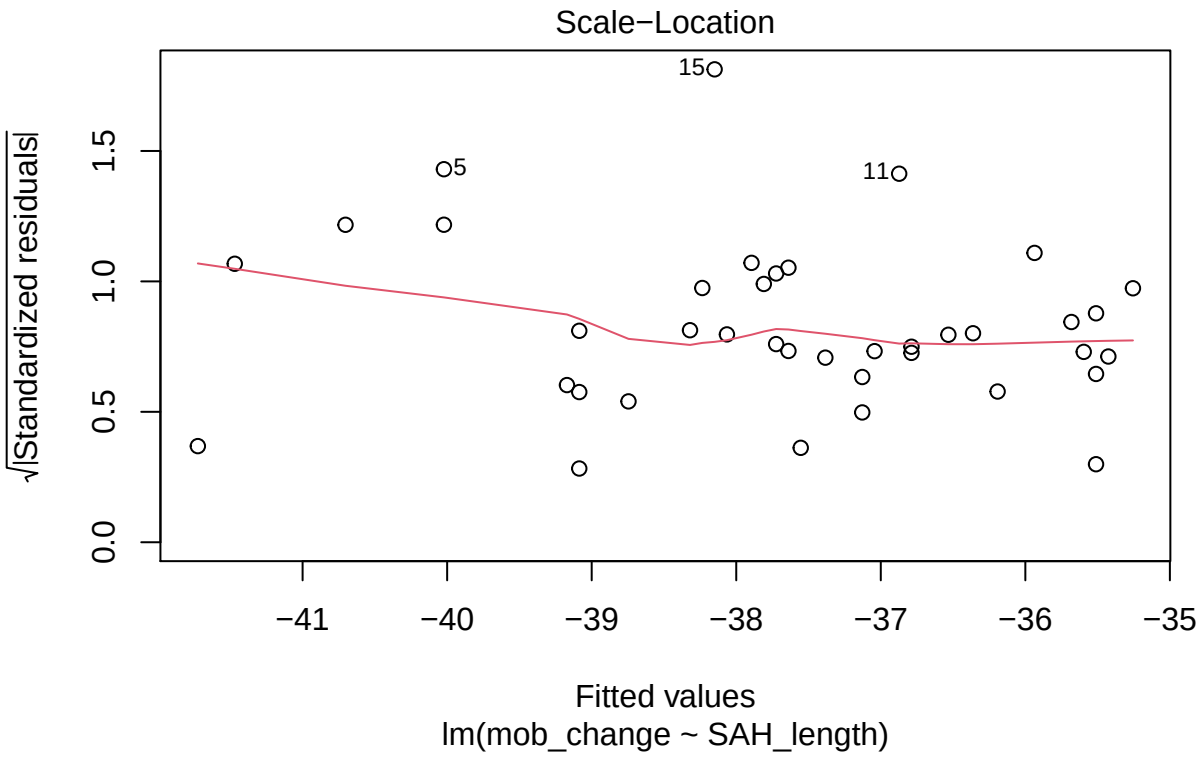
```r
model1 <- lm(mob_change ~ SAH_length, data=sah)
coeftest(model1, vcov = vcovHC(model1))
```

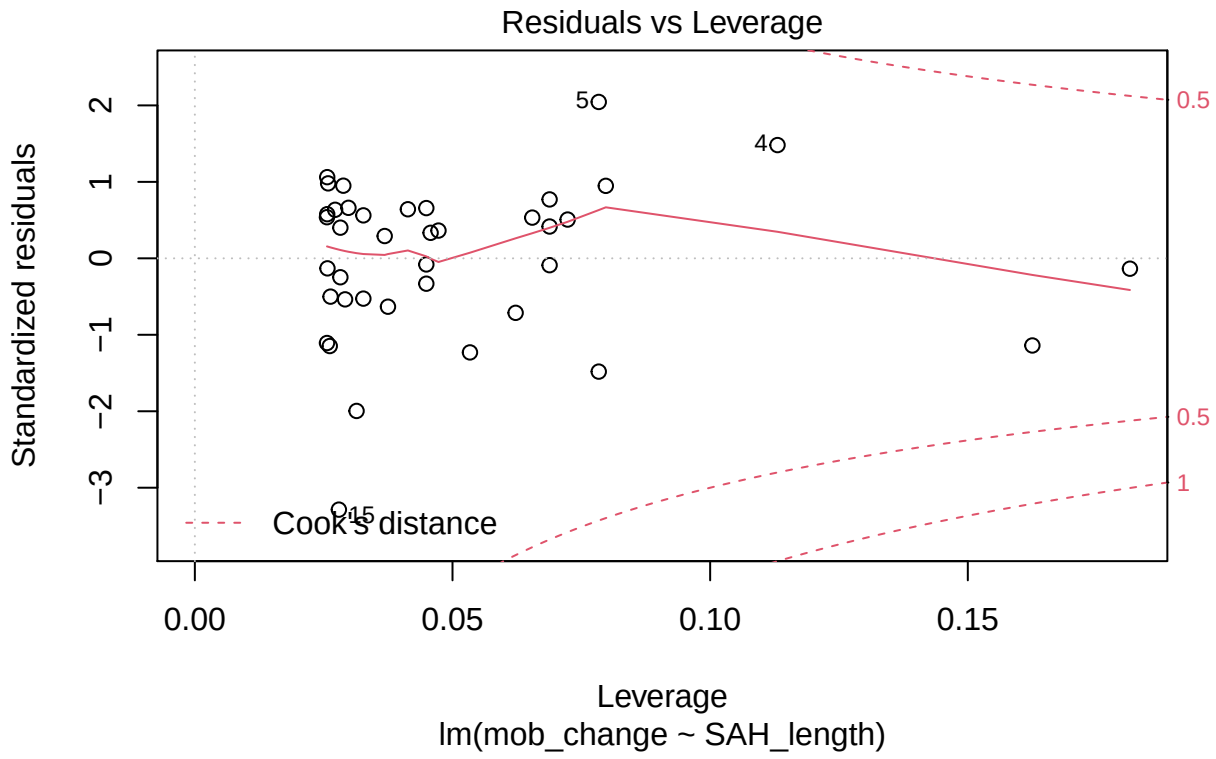
```
##
## t test of coefficients:
##
##               Estimate Std. Error  t value  Pr(>|t|)
## (Intercept) -33.212626   3.053995 -10.8751 4.439e-13 ***
## SAH_length   -0.085121   0.062374  -1.3647    0.1806
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
plot(model1)
```




Normal Q−Q
Standardized residuals
Theoretical Quantiles
lm(mob_change ~ SAH_length)
5
11
15
−3
−2
−1
0
1
2


Scale−Location
15
5
11
√|Standardized residuals|
0.0
0.5
1.0
1.5
−41
−40
−39
−38
−37
−36
−35
Fitted values
lm(mob_change ~ SAH_length)

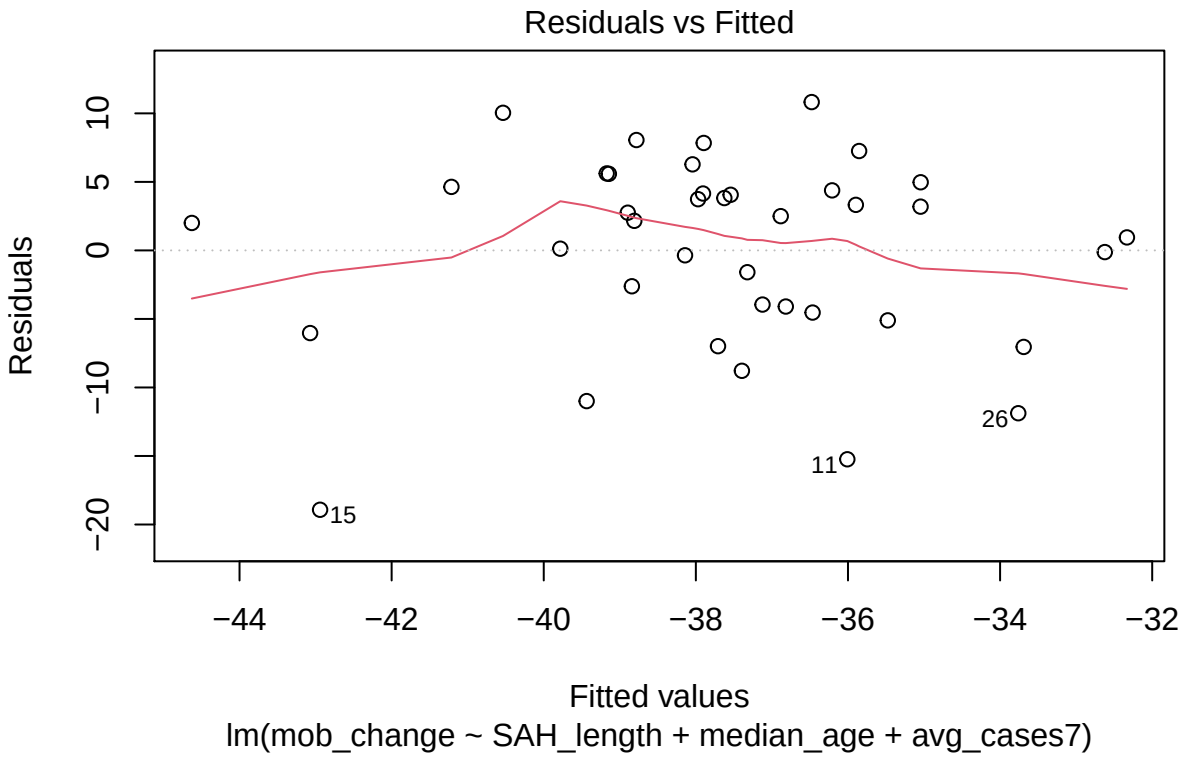


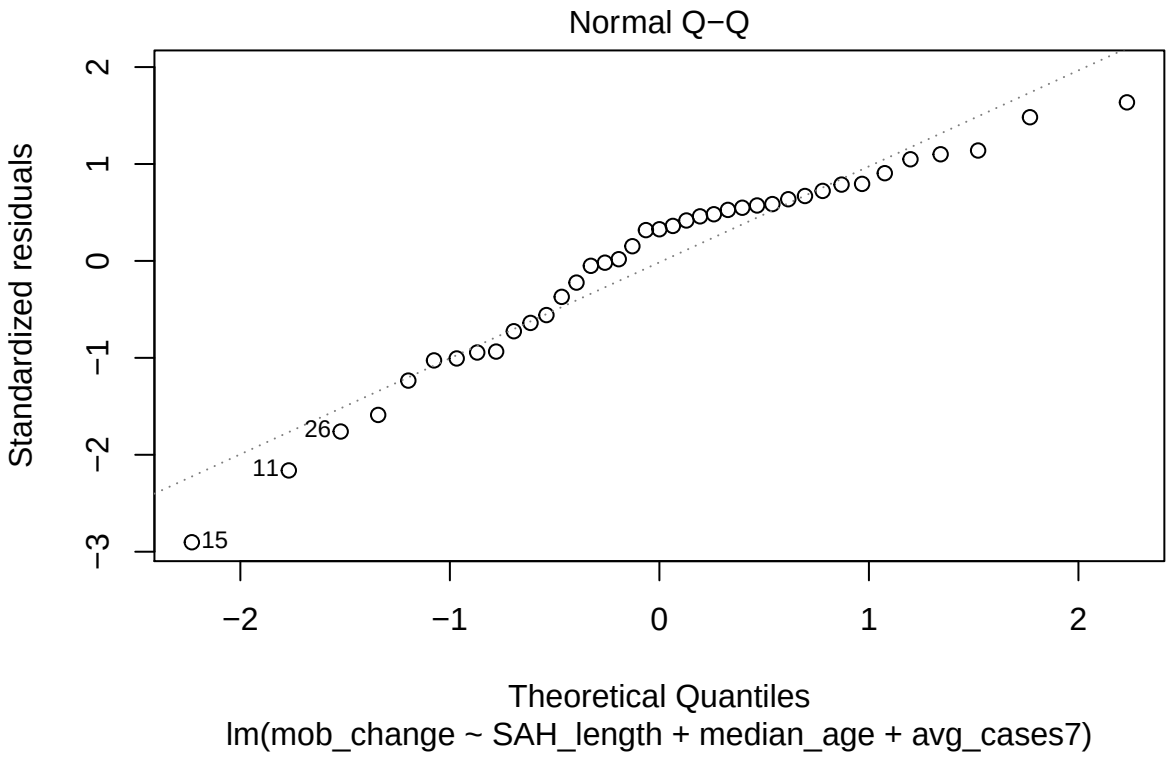

```
cases = read.csv(file='United_States_COVID-19_Cases_and_Deaths_by_State_over_Time.csv')
cases$submission_date <- as.Date(cases$submission_date, format="%m/%d/%Y")


avg_cases = c()
for (row in 1:nrow(sah)) {
  state_code = sah[row, 'State_code']
  start = sah[row, 'SAH_start']
  start_pre7 = start - 7
  avg_cases[row] = mean(cases[cases$state == state_code & cases$submission_date >= start_pre7 & cases$su
}

sah$avg_cases7 = avg_cases
# sah

model2 <- lm(mob_change ~ SAH_length + median_age + avg_cases7, data=sah)
coeftest(model2, vcov = vcovHC(model2))
```

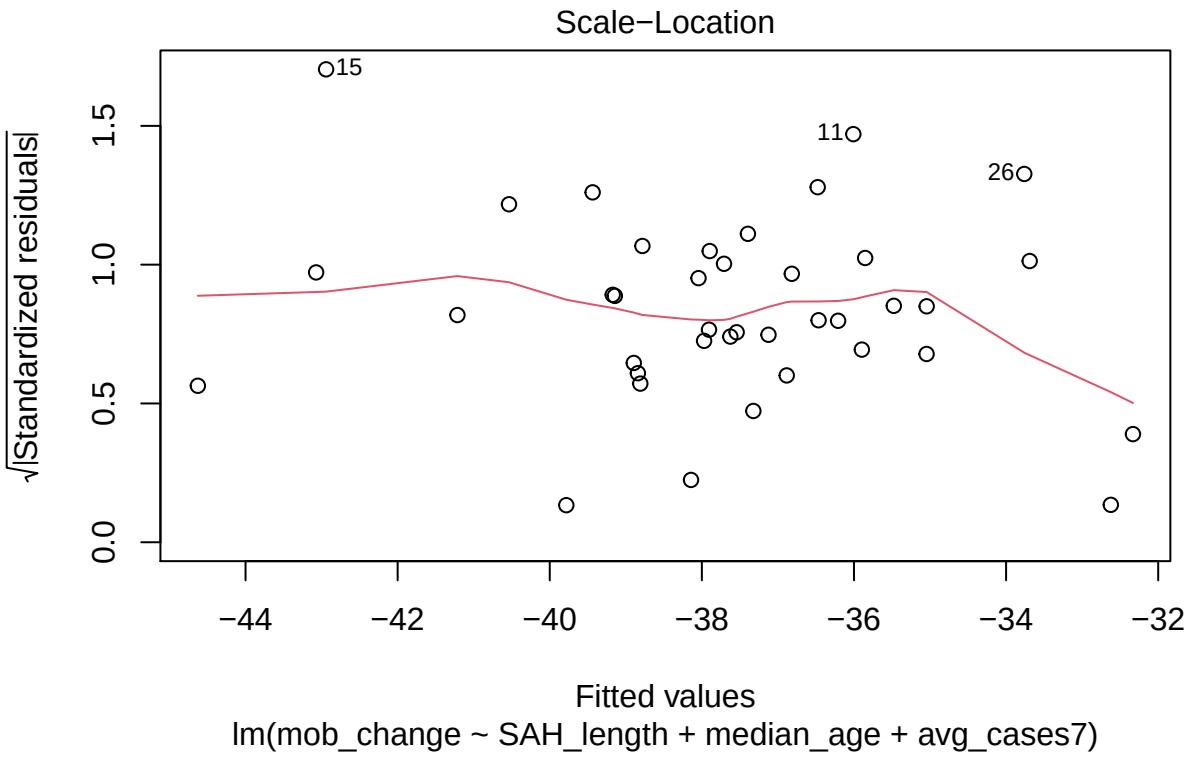
```
##
## t test of coefficients:
##
##                Estimate  Std. Error t value Pr(>|t|)
## (Intercept) -7.0592e+01  3.3944e+01 -2.0797  0.04494 *
## SAH_length  -1.0231e-01  6.2234e-02 -1.6440  0.10913
## median_age   9.8538e-01  8.7763e-01  1.1228  0.26917
## avg_cases7  -5.4658e-04  1.3544e-03 -0.4035  0.68900
```

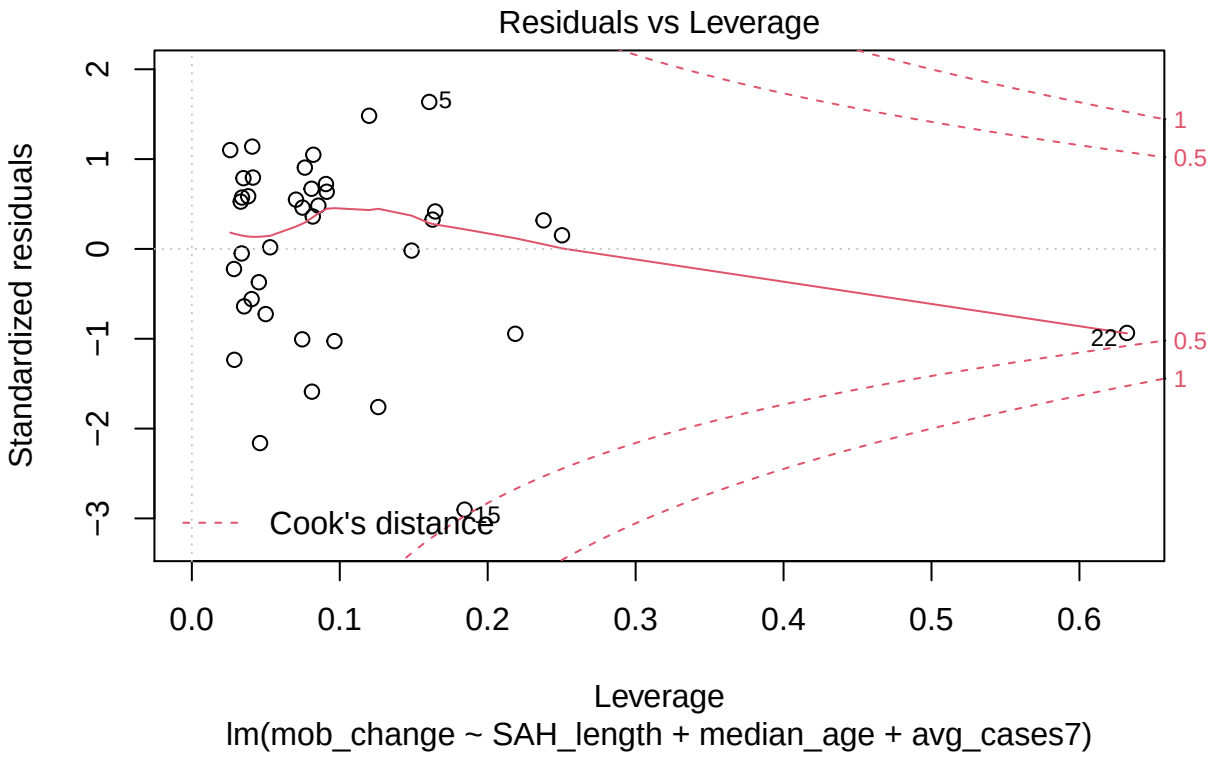
```
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
plot(model2)
```




Normal Q−Q
Standardized residuals
Theoretical Quantiles
lm(mob_change ~ SAH_length + median_age + avg_cases7)
26
11
15


Scale−Location
15
11
26
√|Standardized residuals|
0.0
0.5
1.0
1.5
−44
−42
−40
−38
−36
−34
−32
Fitted values
lm(mob_change ~ SAH_length + median_age + avg_cases7)


Residuals vs Leverage
Standardized residuals
2
1
0
−1
−2
−3
5
22
15
1
0.5
0.5
1
Cook's distance
0.0
0.1
0.2
0.3
0.4
0.5
0.6
Leverage
lm(mob_change ~ SAH_length + median_age + avg_cases7)