Assessment 1

Your Student Number Here

A heading

A sub-heading

A lower-level subheading

```
# A standard R chunk - use these to put your code in
library(tidyverse)
```

```
## -- Attaching packages --------------------------------------- tidyverse 1.3.2 --
## v ggplot2 3.3.6      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()
```

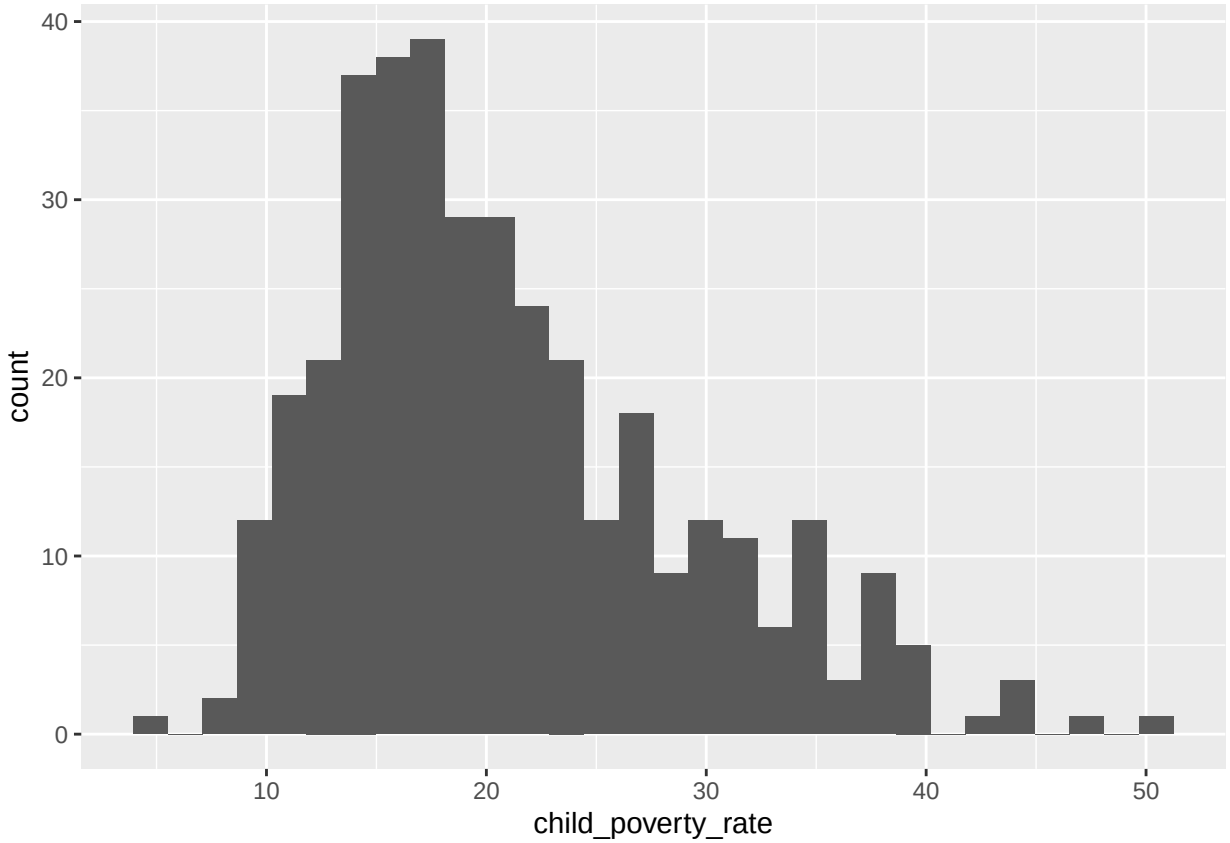
```
dat <- read_csv("dwp-mhclg-dat.csv")
```

```
## Rows: 376 Columns: 4
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr (2): la_code, la_name
## dbl (2): child_poverty_rate, homelessless_per_10k
##
## 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.
```

```
dat %>%
  ggplot() +
  geom_histogram(aes(x = child_poverty_rate))
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```

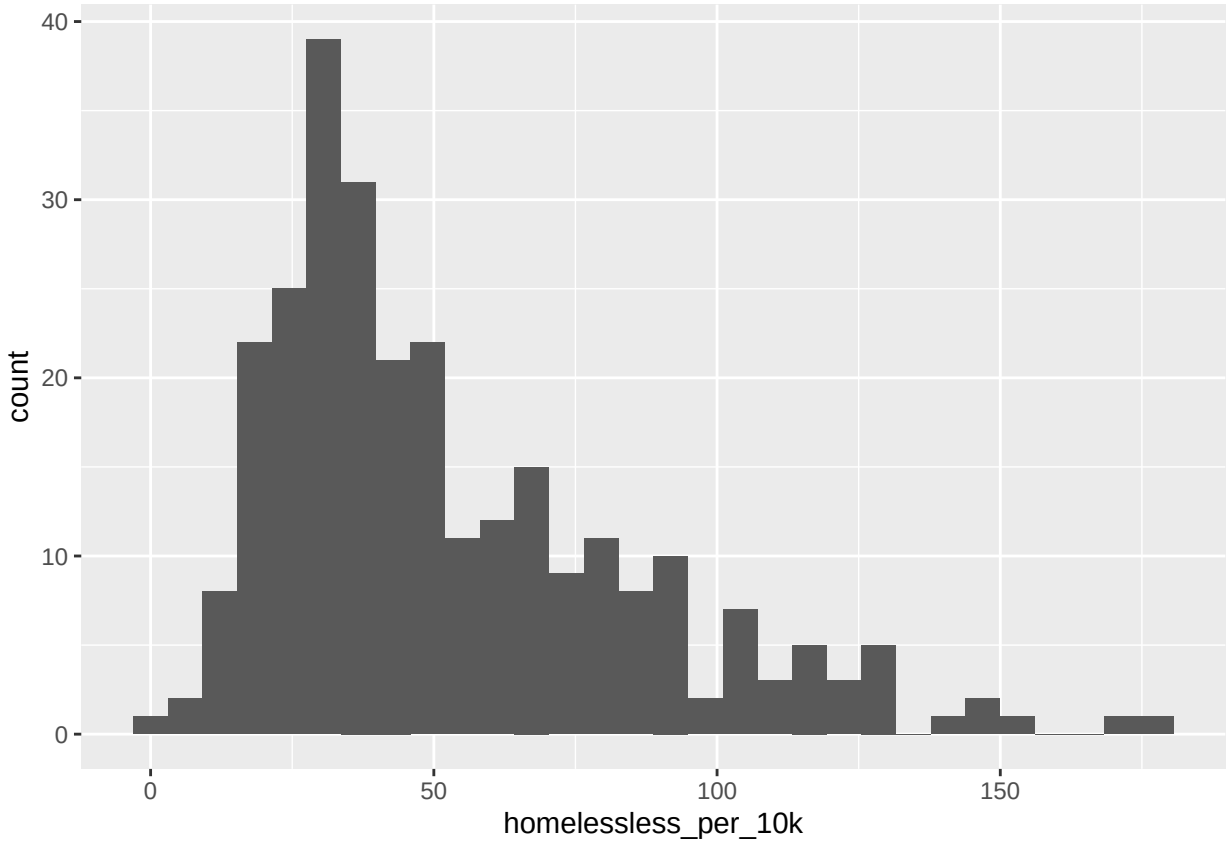
```
## Warning: Removed 1 rows containing non-finite values (stat_bin).
```




```
chil_pov_summ <- dat %>%
  summarise(
    Variable = "Child Poverty Rate",
    mean = mean(child_poverty_rate, na.rm = TRUE),
    med = median(child_poverty_rate, na.rm = TRUE),
    sd = sd(child_poverty_rate, na.rm = TRUE),
    q10 = quantile(child_poverty_rate, probs = 0.1, na.rm = TRUE),
    q90 = quantile(child_poverty_rate, probs = 0.9, na.rm = TRUE)
  )


dat %>%
  ggplot() +
  geom_histogram(aes(x = homelessless_per_10k))
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```

```
## Warning: Removed 98 rows containing non-finite values (stat_bin).
```

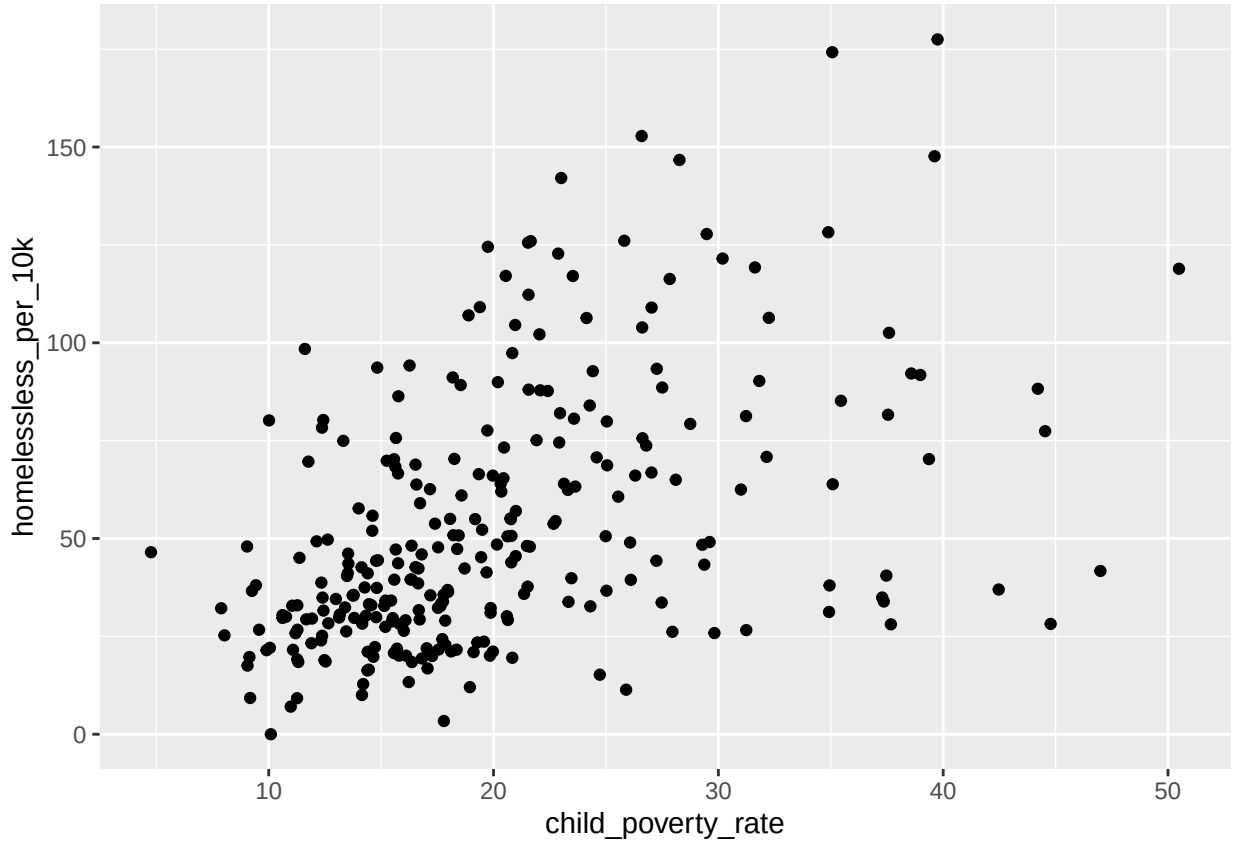



```
homeless_summ <- dat %>%
  summarise(
    Variable = "Homelessness Rate per 10k",
    mean = mean(homelessless_per_10k, na.rm = TRUE),
    med = median(homelessless_per_10k, na.rm = TRUE),
    sd = sd(homelessless_per_10k, na.rm = TRUE),
    q10 = quantile(homelessless_per_10k, probs = 0.1, na.rm = TRUE),
    q90 = quantile(homelessless_per_10k, probs = 0.9, na.rm = TRUE)
  )


bind_rows(chil_pov_summ, homeless_summ) %>%
  knitr::kable(digits = 1,
               col.names = c("Variable", "Mean", "Median", "SD", "10th Percentile", "90th Percentile"))
```

Variable	Mean	Median	SD	10th Percentile	90th Percentile
Child Poverty Rate	21.0	19.2	8.1	12.3	33.1
Homelessness Rate per 10k	53.1	43.0	32.9	20.5	102.3

```
#Bivariate statistics

dat %>%
  ggplot() +
  geom_point(aes(x = child_poverty_rate, y = homelessless_per_10k))
```

```
## Warning: Removed 98 rows containing missing values (geom_point).
```




```
cor(dat$child_poverty_rate, dat$homelessless_per_10k, method = "pearson", use = "complete.obs")
```

```
## [1] 0.472233
```

```
cor.test(dat$child_poverty_rate, dat$homelessless_per_10k, method = "pearson", use = "complete.obs")
```

```
##
##  Pearson's product-moment correlation
##
## data:  dat$child_poverty_rate and dat$homelessless_per_10k
## t = 8.9002, df = 276, p-value < 2.2e-16
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  0.3754480 0.5588302
## sample estimates:
##      cor
## 0.472233
```

```
# Linear regression
homeless_model <- lm(data = dat, homelessless_per_10k ~ child_poverty_rate)
summary(homeless_model)
```

```
##
## Call:
## lm(formula = homelessless_per_10k ~ child_poverty_rate, data = dat)
##
```

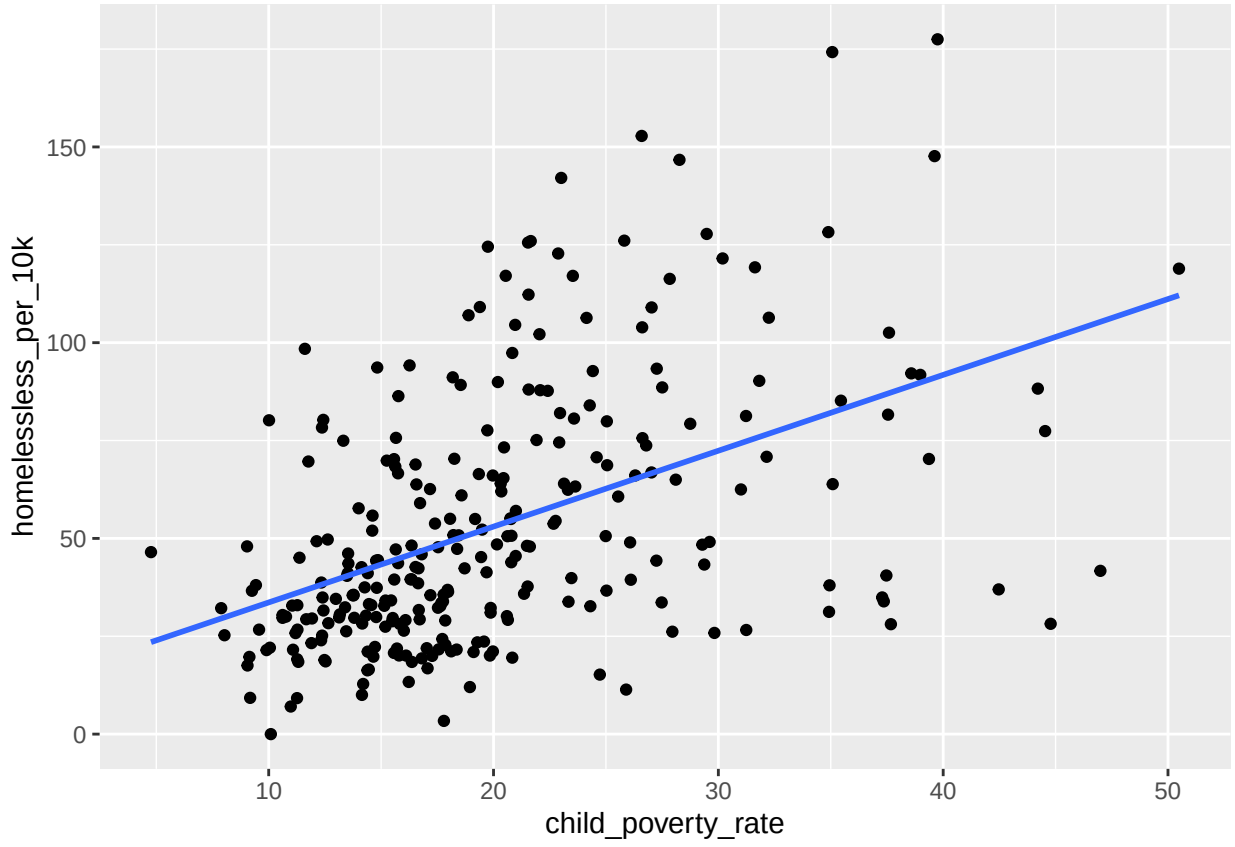
```
## Residuals:
##     Min      1Q  Median      3Q     Max
## -72.855 -18.353  -5.046  13.201  92.023
##
## Coefficients:
##                    Estimate Std. Error t value Pr(>|t|)
## (Intercept)         14.2610     4.6949   3.038  0.00261 **
## child_poverty_rate   1.9377     0.2177   8.900  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 29.05 on 276 degrees of freedom
##   (98 observations deleted due to missingness)
## Multiple R-squared:  0.223,  Adjusted R-squared:  0.2202
## F-statistic: 79.21 on 1 and 276 DF,  p-value: < 2.2e-16
```

```
# Assumptions checking

dat %>%
  ggplot() +
  geom_point(aes(x = child_poverty_rate, y = homelessless_per_10k)) +
  geom_smooth(aes(x = child_poverty_rate, y = homelessless_per_10k), method = "lm", se = FALSE)
```

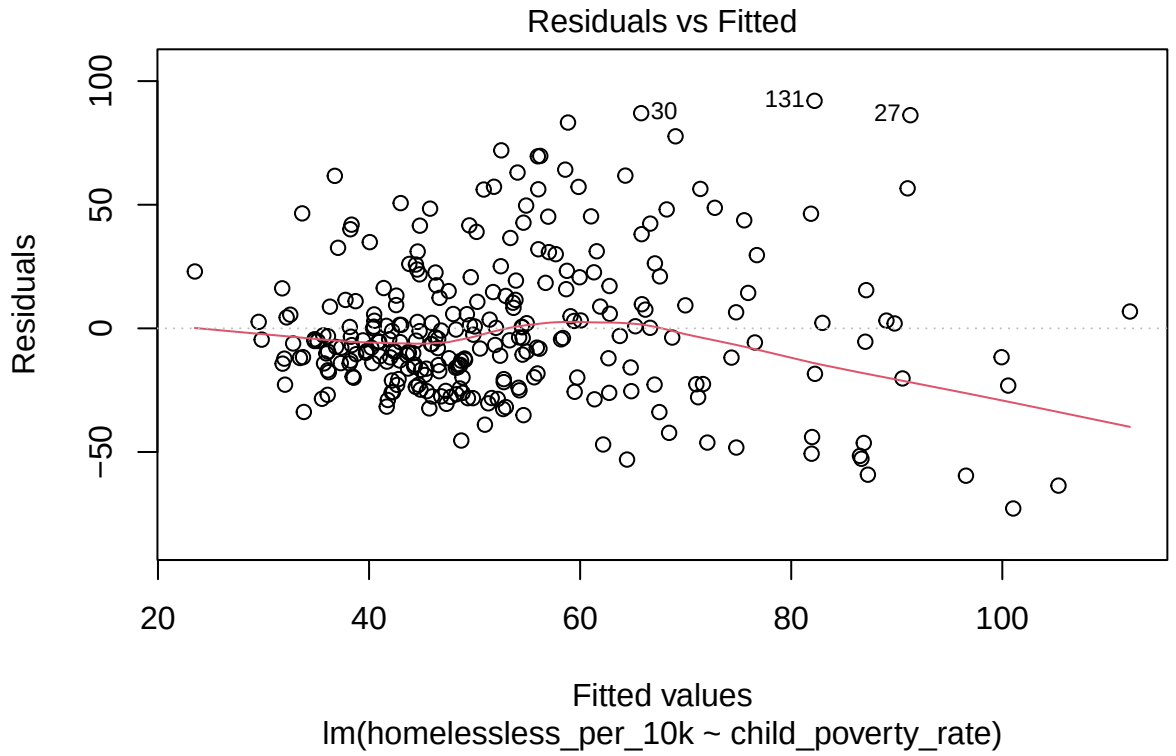
```
## `geom_smooth()` using formula 'y ~ x'
```

```
## Warning: Removed 98 rows containing non-finite values (stat_smooth).
```

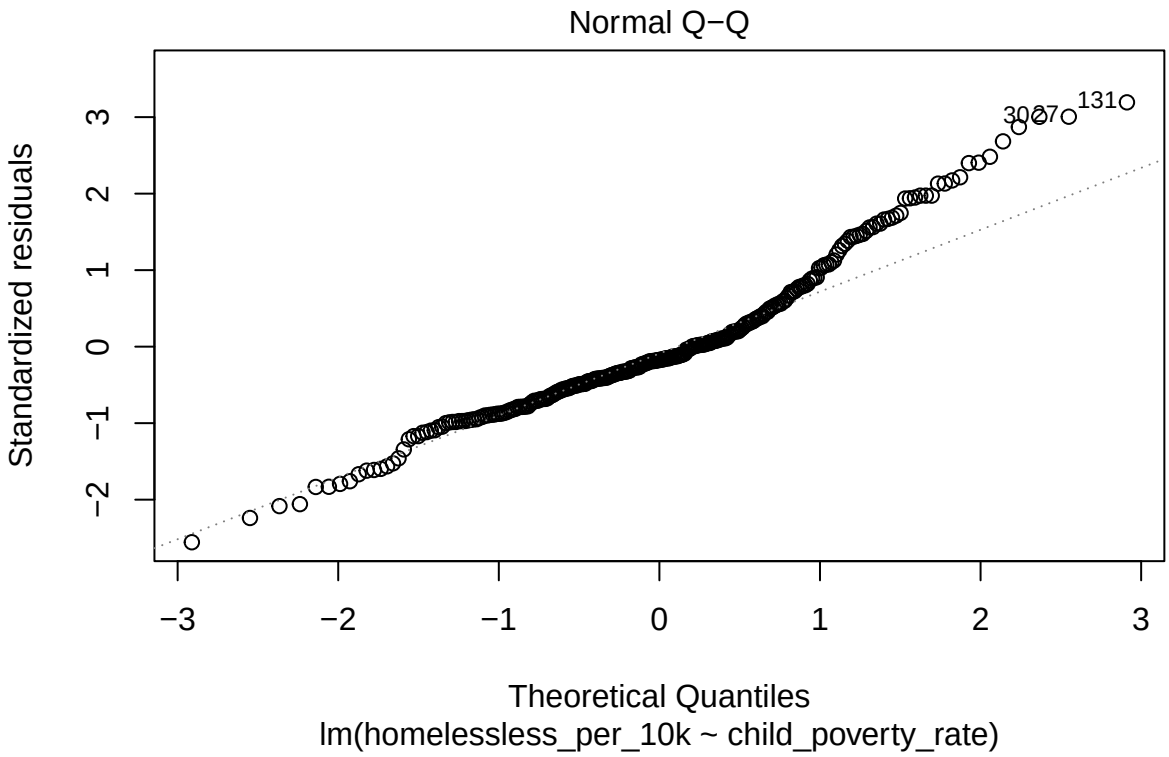
```
## Warning: Removed 98 rows containing missing values (geom_point).
```



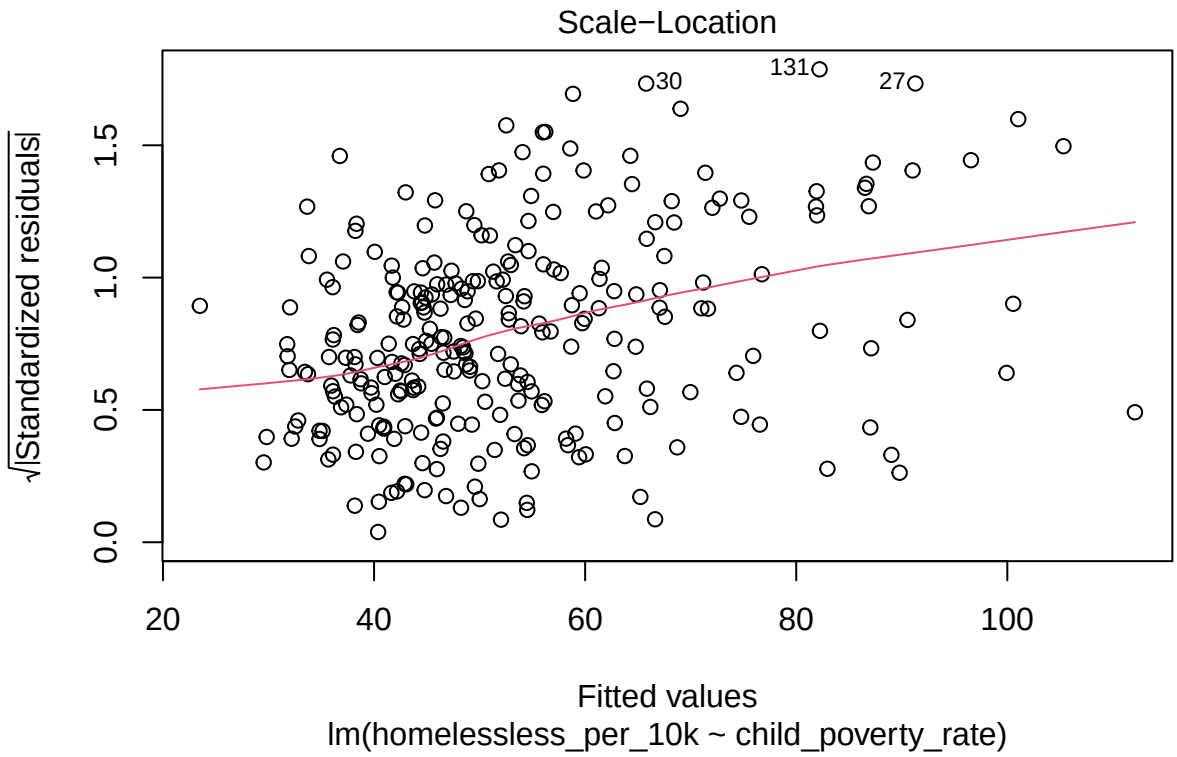
```r
# linearity
plot(homeless_model, which = 1)
```



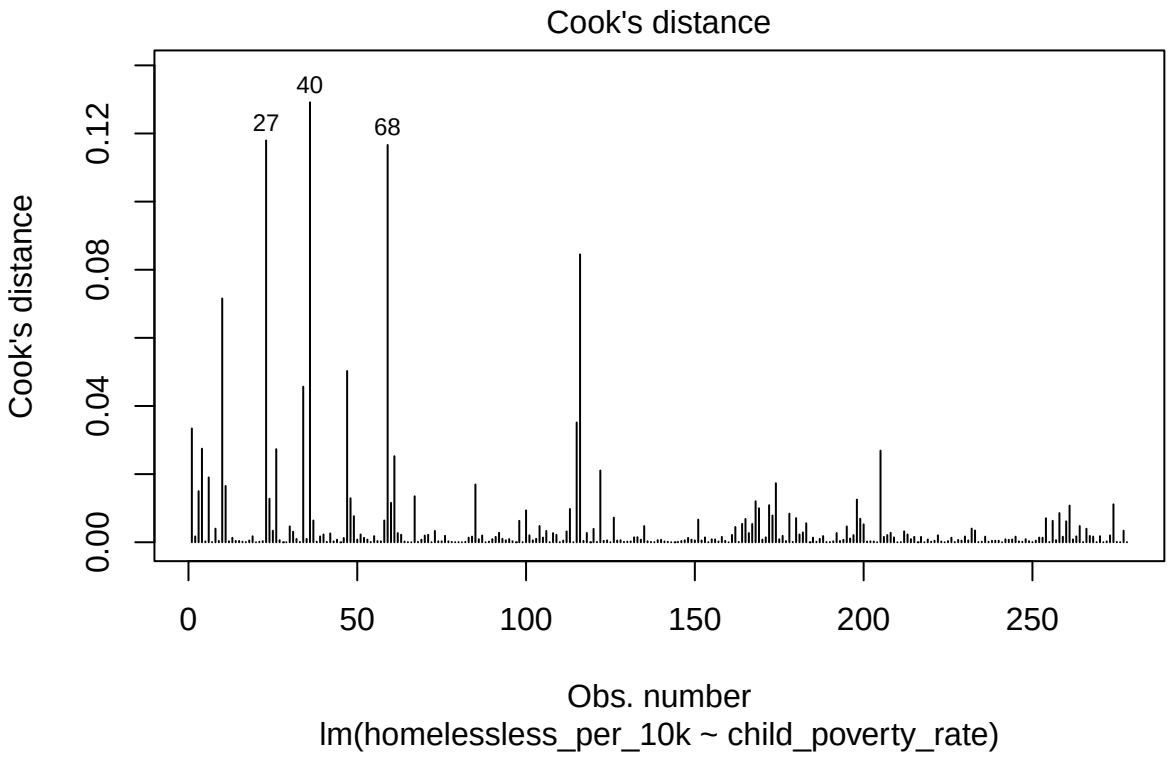

```r
# normality of residuals
plot(homeless_model, which = 2)
```




```
# Heteroscedasticity
plot(homeless_model, which = 3)
```



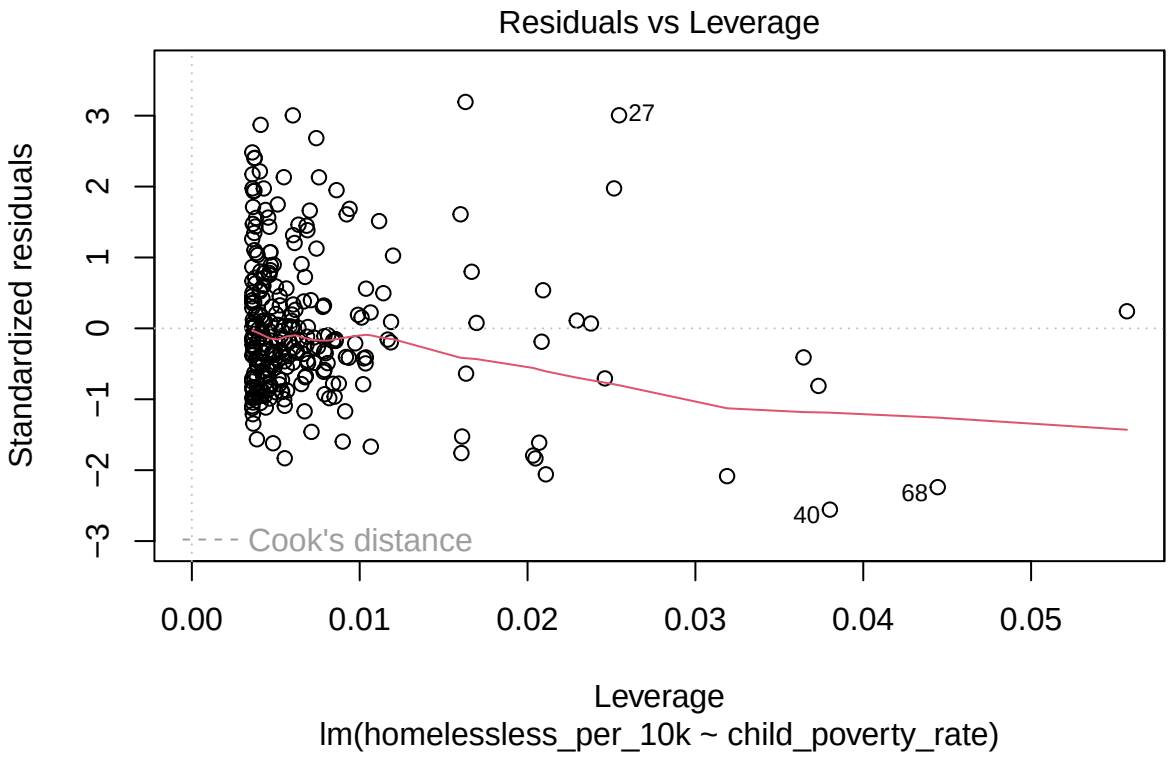

```
# Outliers and leverage points
plot(homeless_model, which = 4)
```




```
4/276
```

```
## [1] 0.01449275
```

```
plot(homeless_model, which = 5)
```


Residuals vs Leverage
Standardized residuals
3
2
1
0
−1
−2
−3
27
68
40
Cook's distance
0.00
0.01
0.02
0.03
0.04
0.05
Leverage
lm(homelessless_per_10k ~ child_poverty_rate)