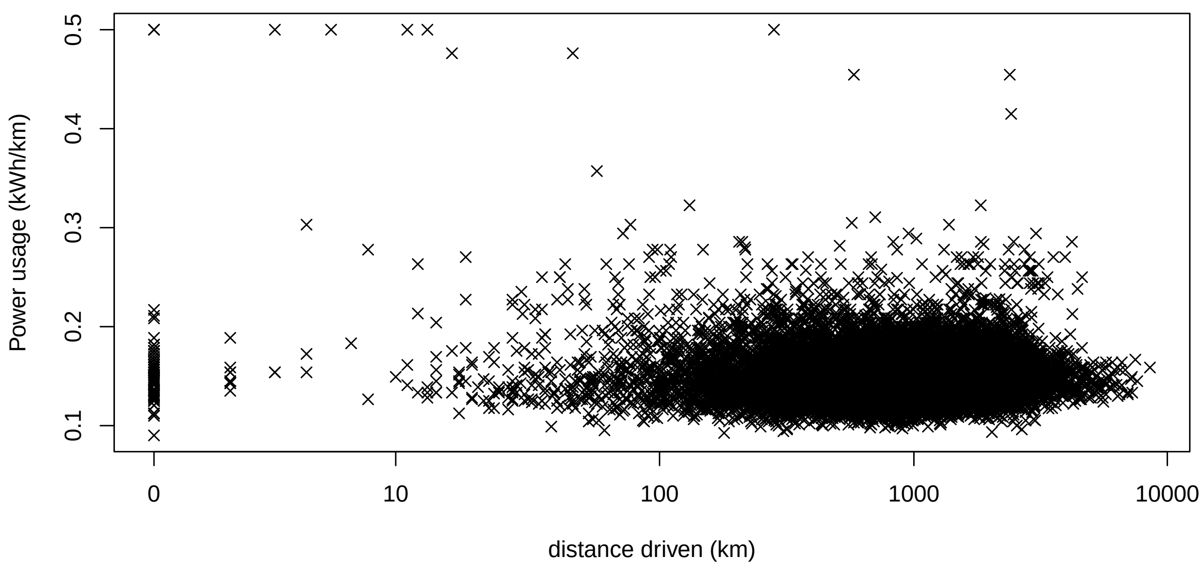
EV data findings

pablo paulsen

23/11/2021

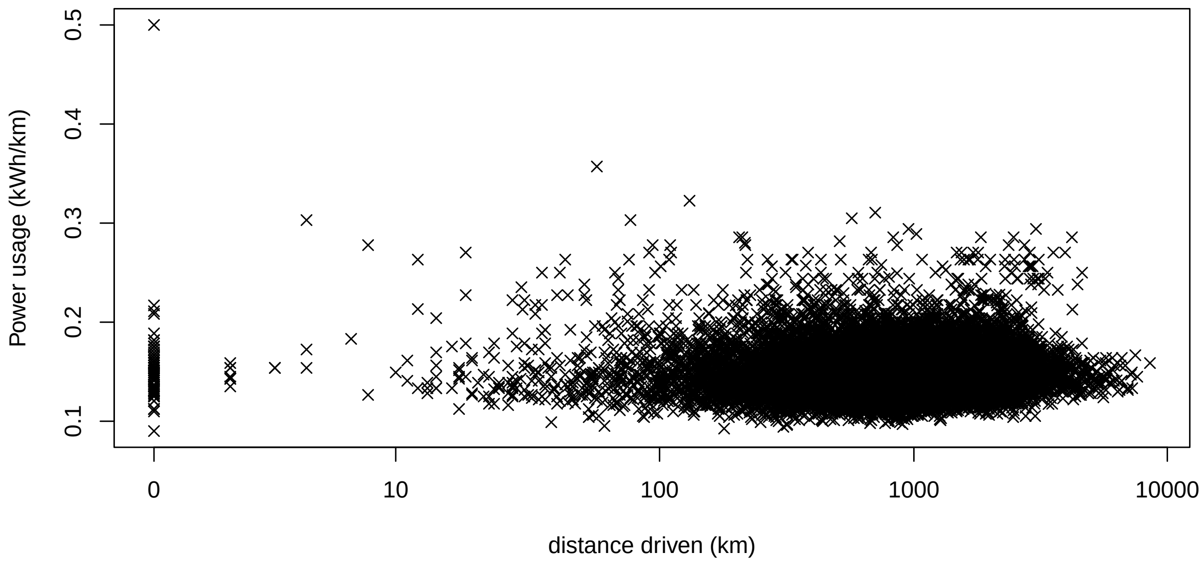
```
load("processed_data/EV_weather_data.rda")
#EV_data = EV_data[EV_data$efficiency > 4,]
```

```
#not strictly a log x axis so that zero can be represented. axis labels are right
plot(EV_data$distance + 1,1/EV_data$efficiency, pch = 4,log = "x", ylab = "Power usage (kWh/km)", xlab =
axis(1, c(0,10,100,1000,10000), at = c(1,9,99,999,9999))
```




```
#bad data vehicles
EV_data = EV_data[EV_data$vehicle != "38f81643" & EV_data$vehicle != "d6082525" & EV_data$vehicle != "ad
```

```
plot(EV_data$distance + 1,1/EV_data$efficiency, pch = 4,log = "x", ylab = "Power usage (kWh/km)", xlab =
axis(1, c(0,10,100,1000,10000), at = c(1,9,99,999,9999))
```




```
EV_data %>%
  group_by(region, weather_region) %>%
  summarise(count =  n_distinct(vehicle)) %>%
  arrange(-count)
```

```
## # A tibble: 25 x 3
## # Groups:   region [25]
##    region           weather_region count
##    <chr>            <fct>          <int>
##  1 Auckland         Auckland         327
##  2 Wellington       Upper Hutt       237
##  3 Christchurch     Christchurch     146
##  4 Coastal Otago    Dunedin          129
##  5 Waikato          Hamilton          64
##  6 Bay of Plenty    Rotorua           52
##  7 North Canterbury Christchurch      34
##  8 Central Otago    Clyde             31
##  9 Mid Canterbury   Christchurch      31
## 10 Nelson           Nelson            31
## # ... with 15 more rows
```

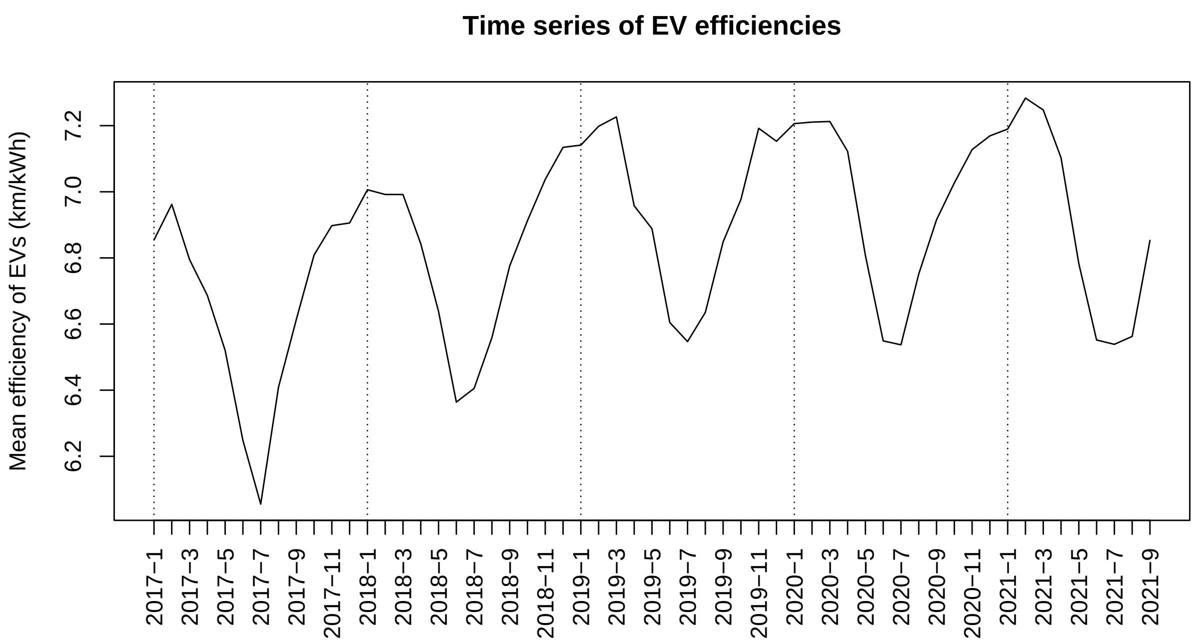
```
EV_data %>%
  group_by(weather_region) %>%
  summarise(count =  n_distinct(vehicle)) %>%
  arrange(-count)
```

```
## # A tibble: 13 x 2
##    weather_region   count
##    <fct>            <int>
##  1 Auckland           382
```

```
##   2 Upper Hutt          248
##   3 Christchurch        226
##   4 Dunedin             136
##   5 Hamilton             64
##   6 Rotorua              52
##   7 Nelson               52
##   8 Clyde                31
##   9 Palmerston North     28
##  10 Stratford            19
##  11 Napier               17
##  12 Invercargill          8
##  13 <NA>                  4
```

```
plot(monthly_EV_data$m, monthly_EV_data$mean_ef, type = 'l', xaxt = "n", xlab = "", ylab = "Mean effici
axis(1,labels = paste(monthly_EV_data$year, monthly_EV_data$month, sep = "-"), at = monthly_EV_data$m, l

yearly_line()
```



simple linear model with mean_eff $= t + \ln t + t^2$+month (as factor). negative squared term means can not use for long term efficiency trend as it will got negative but allows it to better fit the seasonal trend

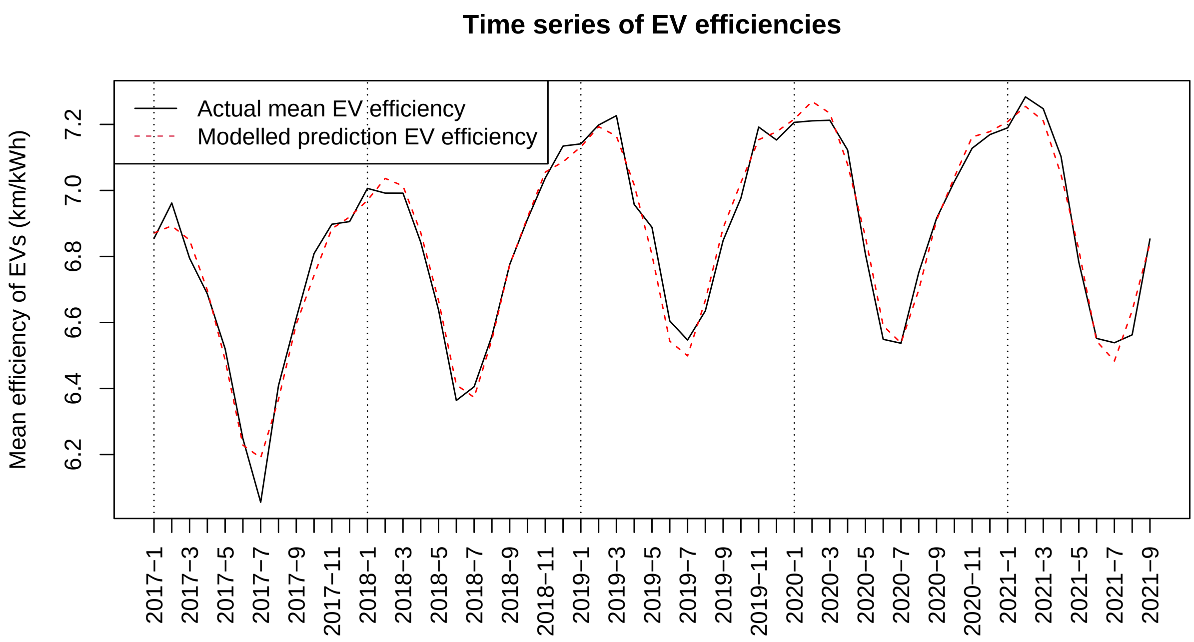
```
monthly_eff_qm_log = lm(mean_ef ~ m+I(log(m))+I(m^2)+factor(month),data = monthly_EV_data)
summary(monthly_eff_qm_log)
```

```
##
## Call:
## lm(formula = mean_ef ~ m + I(log(m)) + I(m^2) + factor(month),
##     data = monthly_EV_data)
##
## Residuals:
##       Min        1Q    Median        3Q       Max
```

```
## -0.134875 -0.029075 -0.000789  0.036190  0.083195
##
## Coefficients:
##                   Estimate Std. Error t value Pr(>|t|)
## (Intercept)      6.840e+00  3.842e-02 178.043  < 2e-16 ***
## m                3.144e-02  4.486e-03   7.010 1.41e-08 ***
## I(log(m))       -8.563e-02  3.391e-02  -2.525   0.0154 *
## I(m^2)          -3.494e-04  5.162e-05  -6.769 3.12e-08 ***
## factor(month)2   5.077e-02  3.166e-02   1.604   0.1163
## factor(month)3   1.276e-02  3.195e-02   0.399   0.6916
## factor(month)4  -1.442e-01  3.220e-02  -4.480 5.66e-05 ***
## factor(month)5  -3.645e-01  3.239e-02 -11.254 2.94e-14 ***
## factor(month)6  -6.339e-01  3.255e-02 -19.475  < 2e-16 ***
## factor(month)7  -6.860e-01  3.268e-02 -20.990  < 2e-16 ***
## factor(month)8  -5.243e-01  3.280e-02 -15.986  < 2e-16 ***
## factor(month)9  -3.113e-01  3.291e-02  -9.458 5.76e-12 ***
## factor(month)10 -1.794e-01  3.439e-02  -5.217 5.25e-06 ***
## factor(month)11 -5.435e-02  3.439e-02  -1.580   0.1216
## factor(month)12 -3.459e-02  3.439e-02  -1.006   0.3204
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.04968 on 42 degrees of freedom
## Multiple R-squared:  0.9775, Adjusted R-squared:  0.9701
## F-statistic: 130.6 on 14 and 42 DF,  p-value: < 2.2e-16
```

```
plot(monthly_EV_data$m, monthly_EV_data$mean_ef, type = 'l', xaxt = "n", xlab = "", ylab = "Mean effici
lines(monthly_EV_data$m, predict(monthly_eff_qm_log), col = 'red', lty = 2)

yearly_line()
axis(1,labels = paste(monthly_EV_data$year, monthly_EV_data$month, sep = "-"), at = monthly_EV_data$m, l
legend("topleft", legend = c("Actual mean EV efficiency", "Modelled prediction EV efficiency"), lty = 1
```



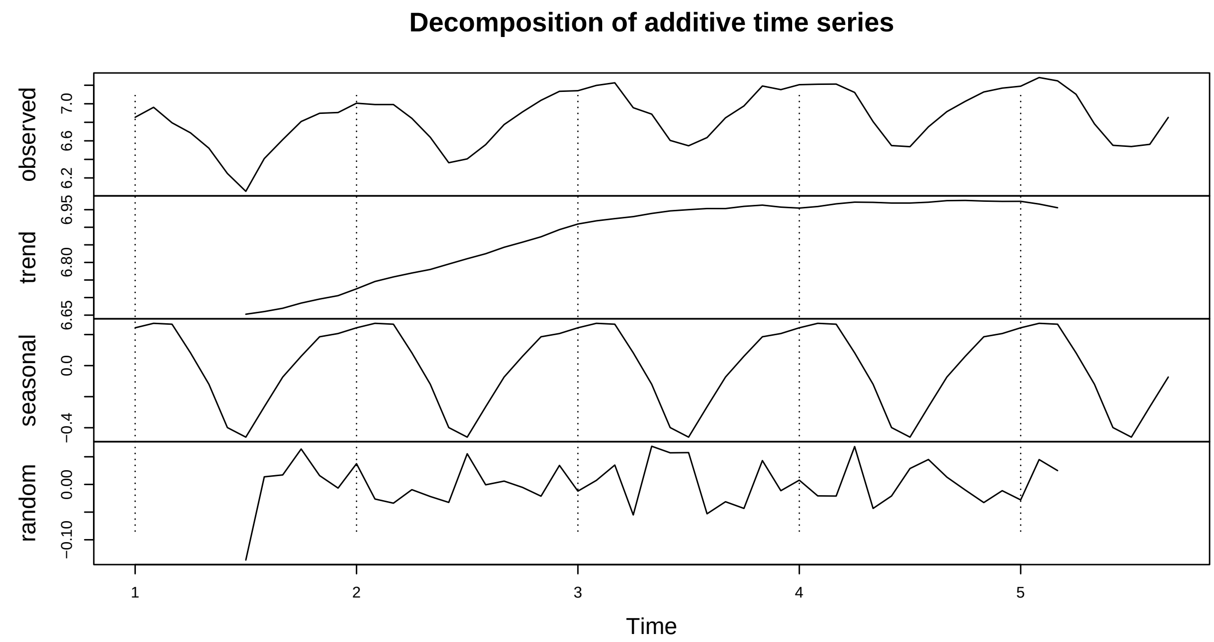

```r
eff_series = ts(monthly_EV_data$mean_ef, frequency = 12)
adf.test(eff_series, alternative = "stationary")
```

```
## Warning in adf.test(eff_series, alternative = "stationary"): p-value smaller
## than printed p-value
```

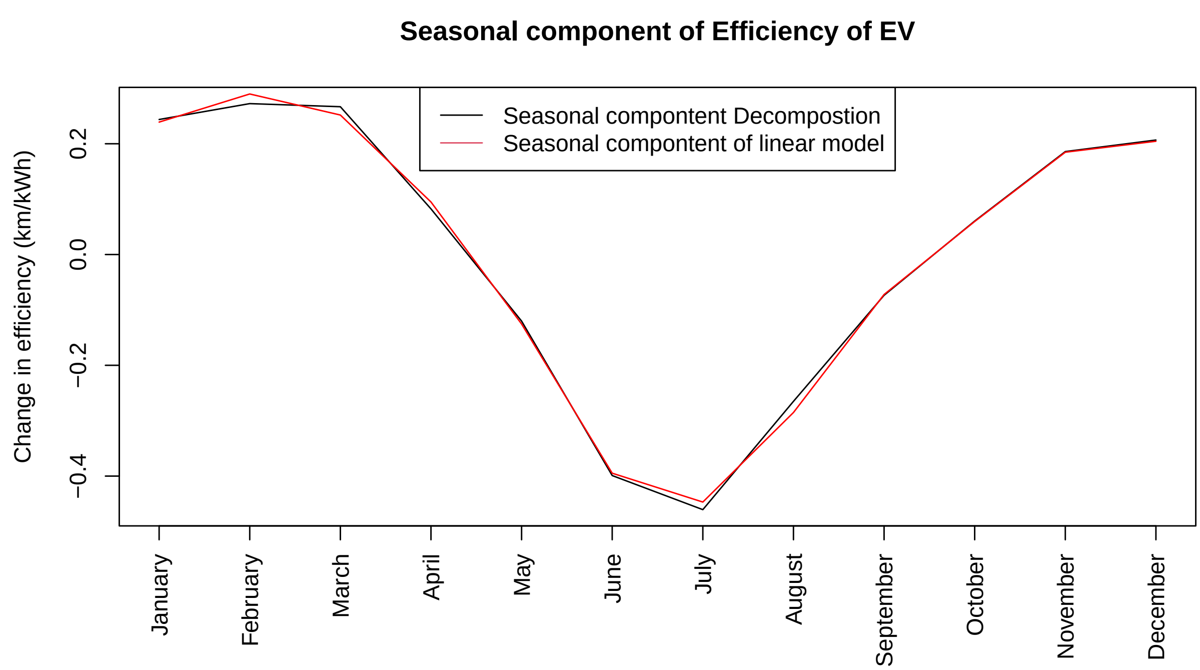
```
##
##  Augmented Dickey-Fuller Test
##
## data:  eff_series
## Dickey-Fuller = -4.6508, Lag order = 3, p-value = 0.01
## alternative hypothesis: stationary
```

we can reject null hypothesis that data is not-stationary. this makes sense as average efficiency should not have significantly changed in a couple of years. use multiplicative instead of additive as preferable to know estimated extra power use? or should i know total extra power used in season?

```r
#decomp_eff = decompose(eff_series, "multiplicative")
decomp_eff = decompose(eff_series,"additive")
plot(decomp_eff)
yearly_line(period = 1)
```




```
plot(decomp_eff$figure, type = 'l', main = "Seasonal component of Efficiency of EV", xaxt = "n",
     xlab = "", ylab = "Change in efficiency (km/kWh)")
points(1:12, scale(c(0,monthly_eff_qm_log$coefficients[paste("factor(month)", 2:12, sep = "")]),scale =
       type = 'l', col = "red")
axis(1, labels = month.name, at = 1:12, las = 3)
legend("top", legend = c("Seasonal compontent Decompostion", 'Seasonal compontent of linear model'), lty
```



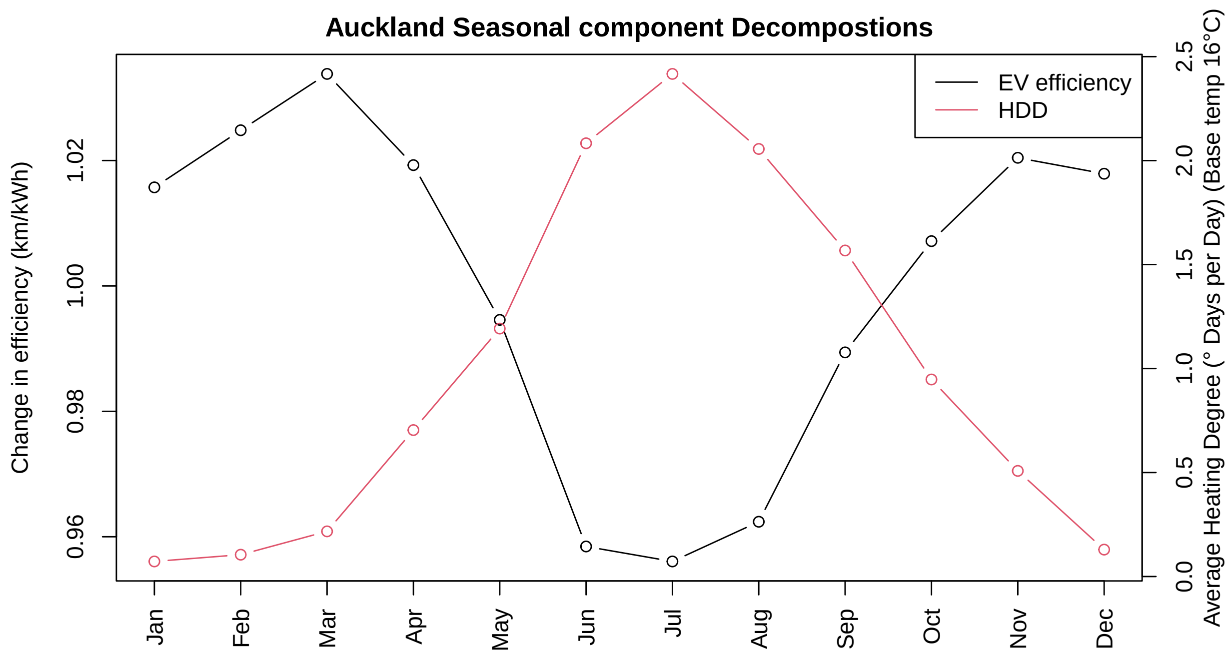
will only do for Auckland as too many lines would get crowded

```r
auck_eff_series = ts(monthly_reg_EV_data$mean_ef[which(monthly_reg_EV_data$weather_region == "Auckland")
decomp_auck_eff = decompose(auck_eff_series,"multiplicative")
auck_HDD_series = ts(monthly_reg_EV_data$HDD[monthly_reg_EV_data$weather_region == "Auckland"], frequen
decomp_auck_HDD = decompose(auck_HDD_series,"multiplicative")

par(mar = c(4, 4, 2, 4))
plot(decomp_auck_eff$figure, type = 'b', main = "Auckland Seasonal component Decompostions", xaxt = "n"
     xlab = "", ylab = "Change in efficiency (km/kWh)")
axis(1, labels = month.abb, at = 1:12, las = 3)

par(new=TRUE)
plot(decomp_auck_HDD$figure, type = 'b', col = 2, xlab="", ylab="",axes=FALSE)
mtext("Average Heating Degree (° Days per Day) (Base temp 16°C)",side=4,line=2, cex = 1)
axis(4)

legend("topright", legend = c("EV efficiency", "HDD"), lty = 1, col = 1:2)
```



intercept base line is Nissan Leaf (24 kWh) 2013-2016

```r
eff_lm = lm(efficiency ~ HDD + model, data = EV_data, na.action=na.omit)
summary(eff_lm)
```

```
##
## Call:
## lm(formula = efficiency ~ HDD + model, data = EV_data, na.action = na.omit)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.5896 -0.4938 -0.0252  0.4884  3.7193
##
```

```
## Coefficients:
##                                    Estimate Std. Error t value Pr(>|t|)    
## (Intercept)                         7.558781   0.010265 736.374  < 2e-16 ***
## HDD                                -0.111218   0.001858 -59.867  < 2e-16 ***
## modelNissan Leaf (30 kWh)          -0.137165   0.012601 -10.885  < 2e-16 ***
## modelNissan Leaf (24 kWh) 2011-2012 -0.604362   0.015153 -39.883  < 2e-16 ***
## modelNissan Leaf (40 kWh)          -0.523064   0.026737 -19.563  < 2e-16 ***
## modelNissan e-NV200 (24 kWh)       -1.272898   0.023892 -53.278  < 2e-16 ***
## modelHyundai Ioniq (EV)             0.839789   0.035133  23.903  < 2e-16 ***
## modelBMW i3                        -0.163968   0.039722  -4.128 3.67e-05 ***
## modelHyundai Kona (EV)             -0.050028   0.047322  -1.057 0.290437    
## modelRenault Zoe                   -0.457310   0.044060 -10.379  < 2e-16 ***
## modelTesla Model 3                 -0.591039   0.053880 -10.970  < 2e-16 ***
## modelNissan Leaf (62 kWh)          -1.031890   0.083459 -12.364  < 2e-16 ***
## modelKia Niro (EV)                 -0.483856   0.061151  -7.912 2.64e-15 ***
## modelTesla Model S                 -2.085562   0.082930 -25.148  < 2e-16 ***
## modelVolkswagen e-Golf             -0.072622   0.067164  -1.081 0.279590    
## modelTesla Model-X                 -3.065955   0.082911 -36.979  < 2e-16 ***
## modelKia Soul                      -0.431039   0.068838  -6.262 3.88e-10 ***
## modelMG ZS EV                      -0.739402   0.192018  -3.851 0.000118 ***
## modelRenault Kangoo (van)          -2.159741   0.093193 -23.175  < 2e-16 ***
## modelJaguar I-PACE                 -2.523603   0.129587 -19.474  < 2e-16 ***
## modelAudi A3 e-tron                -1.395609   0.262848  -5.310 1.11e-07 ***
## modelPeugeot e-208                 -0.346034   0.332473  -1.041 0.297985    
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## 
## Residual standard error: 0.7431 on 22578 degrees of freedom
##   (67 observations deleted due to missingness)
## Multiple R-squared:  0.3325, Adjusted R-squared:  0.3319 
## F-statistic: 535.5 on 21 and 22578 DF,  p-value: < 2.2e-16
```

different city weather stations may be measuring colder or warmer regions of the city and therefore may need a slightly different scaling. interesting that Rotorua has lower effect on HDD, could be cause Rotorua is inland Bay of Plenty so its temperature change is more significant that coastal Tauranga which would also be included in bay of plenty.

```r
eff_reg_lm = lm(efficiency ~ HDD + CDD + weather_region + model, data = EV_data, na.action=na.omit)
summary(eff_reg_lm)
```

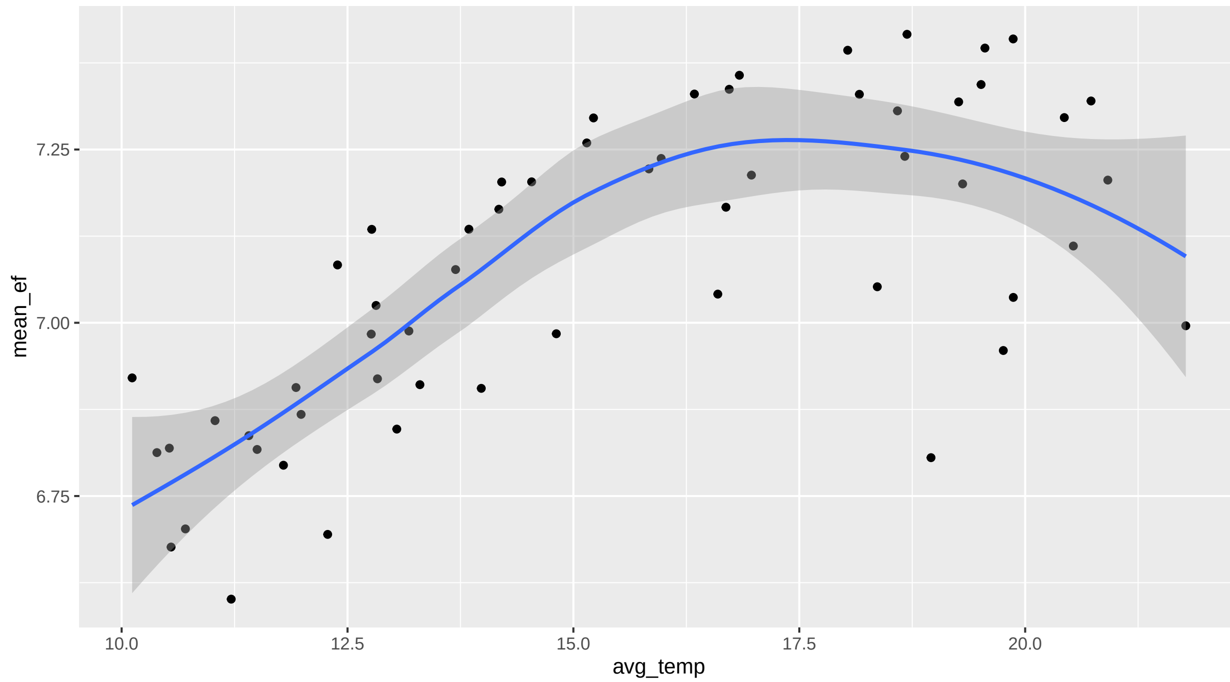
```
## 
## Call:
## lm(formula = efficiency ~ HDD + CDD + weather_region + model, 
##     data = EV_data, na.action = na.omit)
## 
## Residuals:
##     Min      1Q  Median      3Q     Max 
## -3.7121 -0.4691 -0.0179  0.4666  3.9681 
## 
## Coefficients:
##                                    Estimate Std. Error t value Pr(>|t|)    
## (Intercept)                         7.671768   0.013484 568.952  < 2e-16 ***
## HDD                                -0.108359   0.002435 -44.498  < 2e-16 ***
```

```
## CDD                                 -0.115277   0.027888  -4.134 3.59e-05 ***
## weather_regionUpper Hutt            -0.052324   0.014547  -3.597 0.000323 ***
## weather_regionChristchurch           0.090803   0.015349   5.916 3.35e-09 ***
## weather_regionDunedin               -0.552403   0.017566 -31.447  < 2e-16 ***
## weather_regionHamilton              -0.349028   0.025976 -13.436  < 2e-16 ***
## weather_regionRotorua               -0.047141   0.026656  -1.768 0.076995 .
## weather_regionNelson                -0.018554   0.022135  -0.838 0.401909
## weather_regionClyde                 -0.145740   0.037589  -3.877 0.000106 ***
## weather_regionPalmerston North      -0.702947   0.033470 -21.002  < 2e-16 ***
## weather_regionStratford             -0.205407   0.044718  -4.593 4.39e-06 ***
## weather_regionNapier                -0.336085   0.040457  -8.307  < 2e-16 ***
## weather_regionInvercargill          -0.061839   0.069828  -0.886 0.375853
## modelNissan Leaf (30 kWh)           -0.165900   0.012227 -13.568  < 2e-16 ***
## modelNissan Leaf (24 kWh) 2011-2012 -0.632911   0.014653 -43.193  < 2e-16 ***
## modelNissan Leaf (40 kWh)           -0.519599   0.025749 -20.180  < 2e-16 ***
## modelNissan e-NV200 (24 kWh)        -1.305699   0.022980 -56.820  < 2e-16 ***
## modelHyundai Ioniq (EV)              0.940443   0.034592  27.187  < 2e-16 ***
## modelBMW i3                         -0.048308   0.038684  -1.249 0.211766
## modelHyundai Kona (EV)              -0.095884   0.045500  -2.107 0.035098 *
## modelRenault Zoe                    -0.532540   0.042447 -12.546  < 2e-16 ***
## modelTesla Model 3                  -0.598148   0.052158 -11.468  < 2e-16 ***
## modelNissan Leaf (62 kWh)           -1.130718   0.080299 -14.081  < 2e-16 ***
## modelKia Niro (EV)                  -0.509514   0.059814  -8.518  < 2e-16 ***
## modelTesla Model S                  -2.198273   0.079787 -27.552  < 2e-16 ***
## modelVolkswagen e-Golf              -0.139709   0.065550  -2.131 0.033073 *
## modelTesla Model-X                  -3.143645   0.081138 -38.744  < 2e-16 ***
## modelKia Soul                       -0.472298   0.067702  -6.976 3.12e-12 ***
## modelMG ZS EV                       -0.766831   0.184468  -4.157 3.24e-05 ***
## modelRenault Kangoo (van)           -2.223587   0.092778 -23.967  < 2e-16 ***
## modelJaguar I-PACE                  -2.646551   0.124529 -21.253  < 2e-16 ***
## modelAudi A3 e-tron                 -1.507841   0.252394  -5.974 2.35e-09 ***
## modelPeugeot e-208                  -0.569116   0.319248  -1.783 0.074652 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.7132 on 22566 degrees of freedom
##   (67 observations deleted due to missingness)
## Multiple R-squared:  0.3854, Adjusted R-squared:  0.3845
## F-statistic: 428.8 on 33 and 22566 DF,  p-value: < 2.2e-16
```

https://www.geotab.com/blog/ev-range/

based on this AC should also decrease range. not too obvious in NZ as is kind of cold but in Auckland can see such a trend what if we include cooling degree days in analysis too? unlike this direct average temp vs efficiency plot this would allow for cooling and heating in the same month that could reduce efficiency. could explain the couple month that have very bad efficiency, possibley have a few cold and warm days but average is nothing unsual

```
monthly_reg_EV_data[monthly_reg_EV_data$weather_region == "Auckland",] %>% ggplot(aes(avg_temp, mean_ef
```

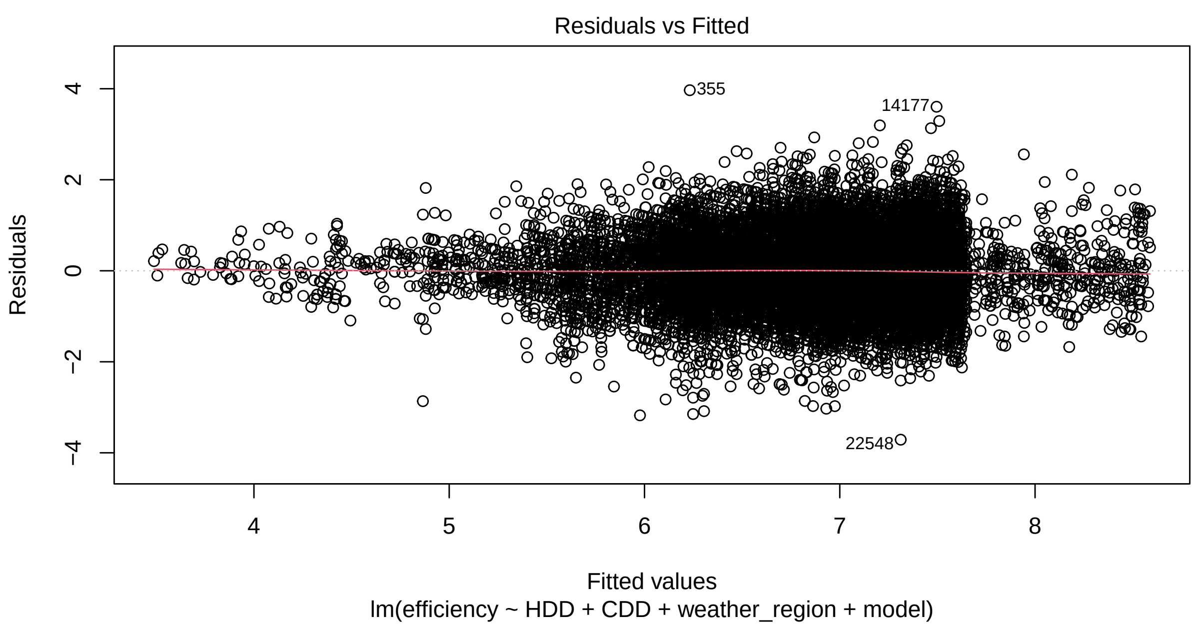


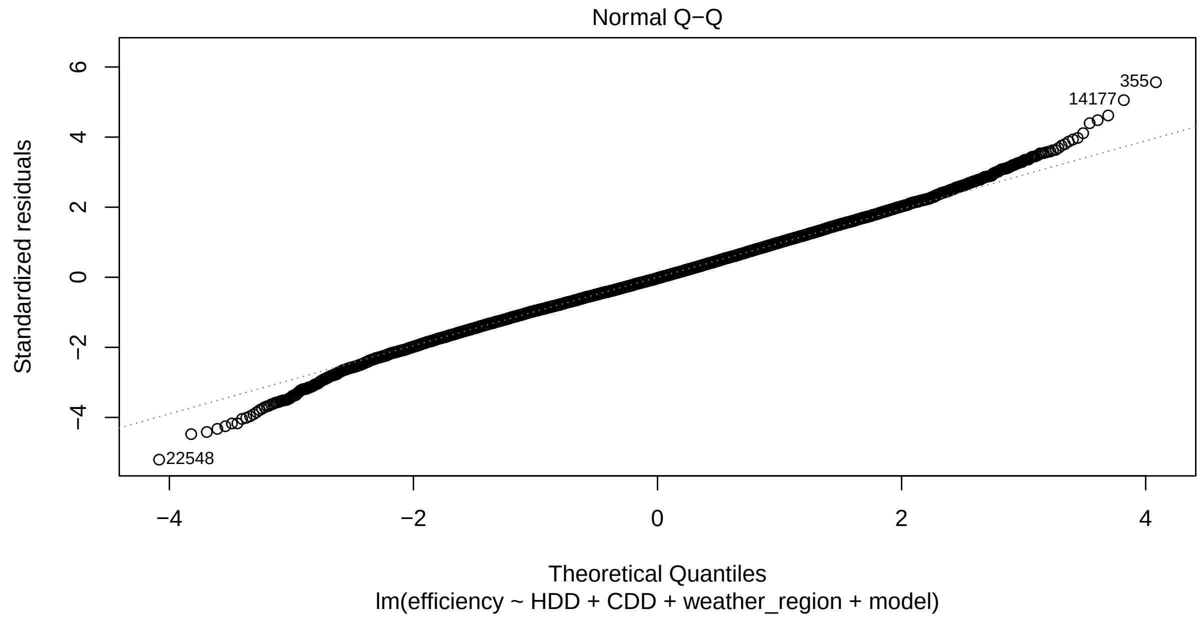

```
eff_h_c_lm = lm((1/efficiency) ~ HDD + CDD + weather_region + model, data = EV_data, na.action=na.omit,
summary(eff_h_c_lm)
```

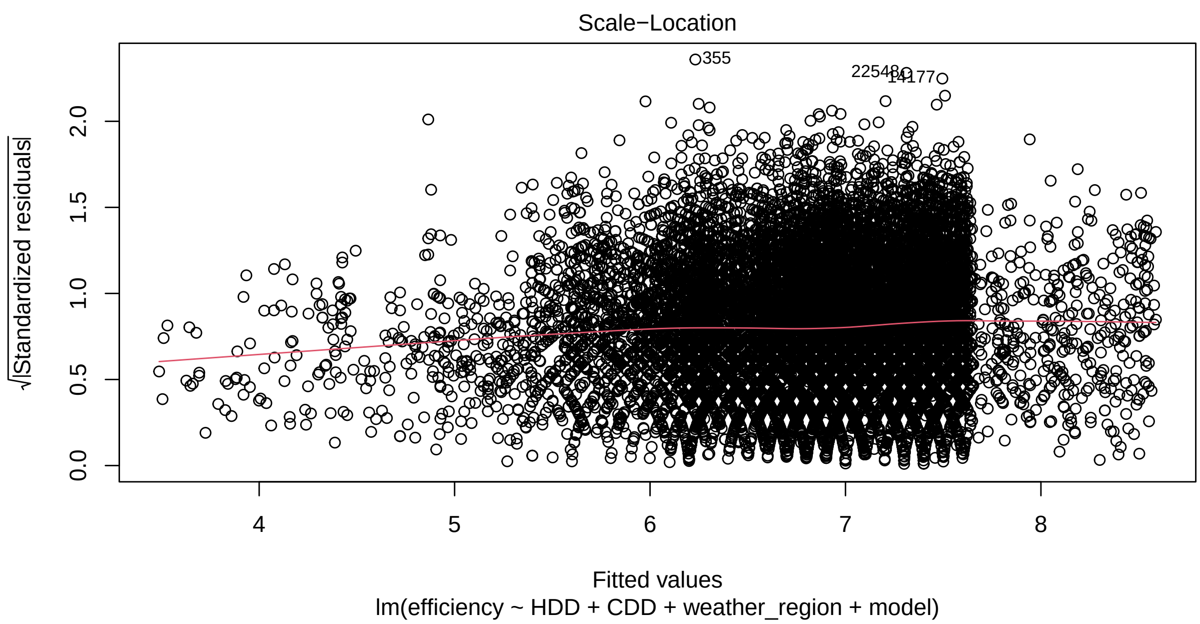
```
##
## Call:
## lm(formula = (1/efficiency) ~ HDD + CDD + weather_region + model,
##     data = EV_data, weights = distance, na.action = na.omit)
##
## Weighted Residuals:
##     Min      1Q  Median      3Q     Max
## -2.3300 -0.3360 -0.0589  0.2337  4.0766
##
## Coefficients:
##                                     Estimate Std. Error t value Pr(>|t|)
## (Intercept)                        1.321e-01  2.868e-04 460.596  < 2e-16 ***
## HDD                                2.196e-03  5.098e-05  43.077  < 2e-16 ***
## CDD                                2.384e-03  5.724e-04   4.166 3.11e-05 ***
## weather_regionUpper Hutt          -4.788e-04  3.037e-04  -1.577  0.11486
## weather_regionChristchurch        -9.105e-04  3.258e-04  -2.794  0.00521 **
## weather_regionDunedin              1.206e-02  3.836e-04  31.441  < 2e-16 ***
## weather_regionHamilton             8.511e-03  5.301e-04  16.058  < 2e-16 ***
## weather_regionRotorua              5.014e-03  5.464e-04   9.176  < 2e-16 ***
## weather_regionNelson               2.712e-03  4.808e-04   5.641 1.71e-08 ***
## weather_regionClyde                4.523e-03  7.494e-04   6.036 1.61e-09 ***
## weather_regionPalmerston North     1.411e-02  6.654e-04  21.208  < 2e-16 ***
## weather_regionStratford            1.036e-02  9.500e-04  10.905  < 2e-16 ***
## weather_regionNapier               6.312e-03  8.476e-04   7.448 9.85e-14 ***
## weather_regionInvercargill         3.192e-03  1.759e-03   1.815  0.06956 .
## modelNissan Leaf (30 kWh)          3.401e-03  2.525e-04  13.468  < 2e-16 ***
## modelNissan Leaf (24 kWh) 2011-2012  1.239e-02  3.247e-04  38.150  < 2e-16 ***
## modelNissan Leaf (40 kWh)          1.068e-02  5.176e-04  20.625  < 2e-16 ***
```

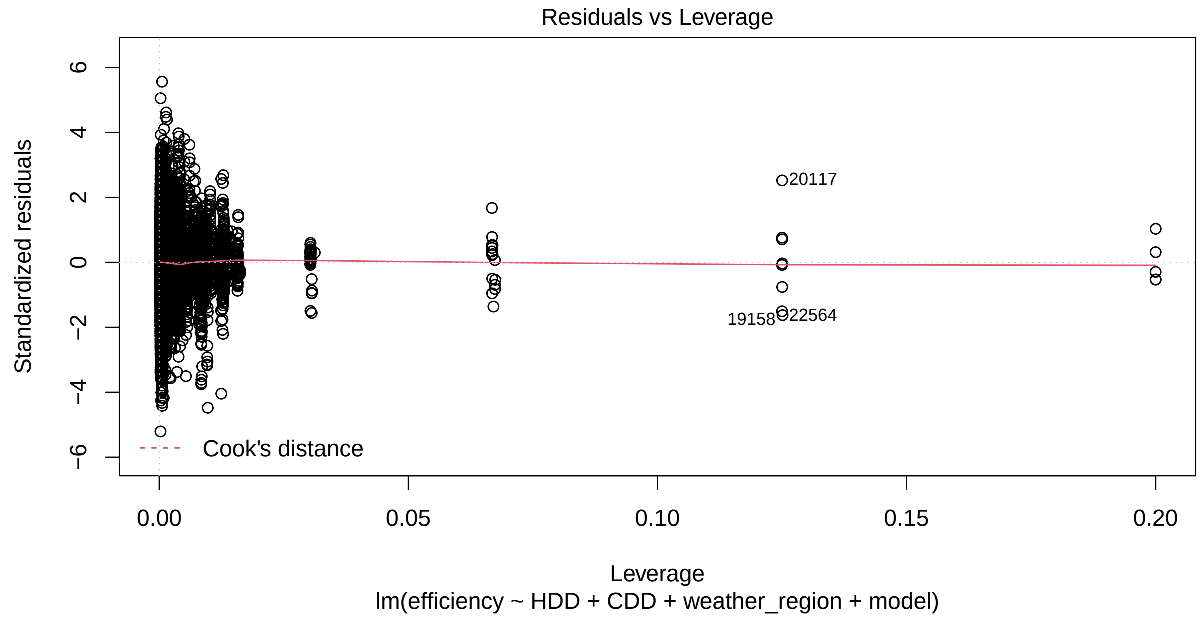
```
## modelNissan e-NV200 (24 kWh)          3.272e-02  5.369e-04  60.929  < 2e-16 ***
## modelHyundai Ioniq (EV)              -1.832e-02  6.853e-04 -26.739  < 2e-16 ***
## modelBMW i3                          -1.334e-03  7.876e-04  -1.694  0.09026 .
## modelHyundai Kona (EV)                6.818e-04  8.603e-04   0.793  0.42808
## modelRenault Zoe                      1.155e-02  8.510e-04  13.568  < 2e-16 ***
## modelTesla Model 3                    1.055e-02  1.023e-03  10.317  < 2e-16 ***
## modelNissan Leaf (62 kWh)             2.546e-02  1.753e-03  14.525  < 2e-16 ***
## modelKia Niro (EV)                    1.134e-02  1.193e-03   9.507  < 2e-16 ***
## modelTesla Model S                    4.838e-02  1.691e-03  28.615  < 2e-16 ***
## modelVolkswagen e-Golf                1.208e-03  1.538e-03   0.785  0.43233
## modelTesla Model-X                    1.041e-01  1.297e-03  80.309  < 2e-16 ***
## modelKia Soul                         6.276e-03  1.250e-03   5.020 5.20e-07 ***
## modelMG ZS EV                         2.212e-02  3.902e-03   5.669 1.45e-08 ***
## modelRenault Kangoo (van)             5.663e-02  1.538e-03  36.832  < 2e-16 ***
## modelJaguar I-PACE                    7.303e-02  2.952e-03  24.740  < 2e-16 ***
## modelAudi A3 e-tron                   3.338e-02  4.855e-03   6.875 6.35e-12 ***
## modelPeugeot e-208                    1.097e-02  9.584e-03   1.144  0.25254
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.4924 on 22497 degrees of freedom
##   (67 observations deleted due to missingness)
## Multiple R-squared:  0.4855, Adjusted R-squared:  0.4847
## F-statistic: 643.2 on 33 and 22497 DF,  p-value: < 2.2e-16
```

```r
#plot(eff_lm)
plot(eff_reg_lm)
```

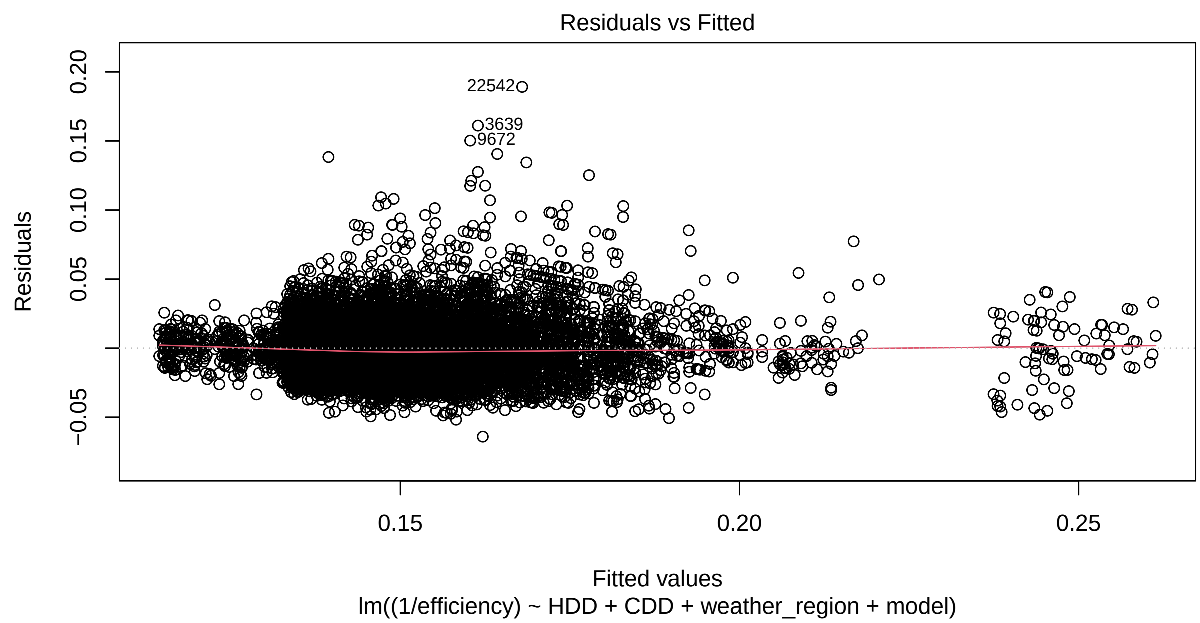


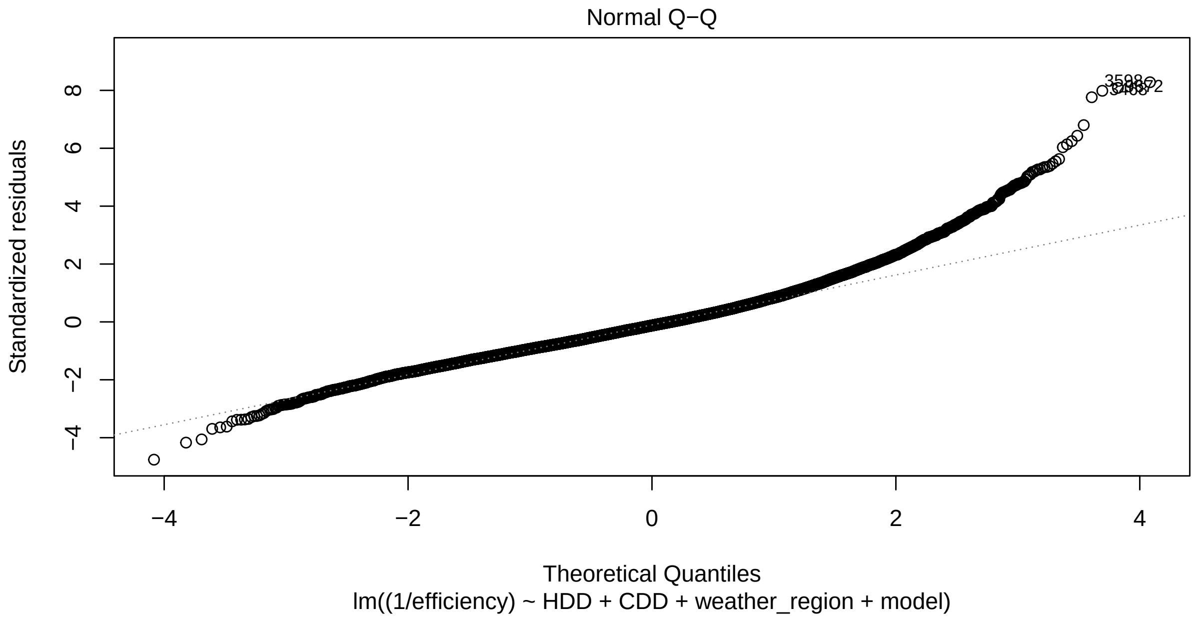

Normal Q−Q
Standardized residuals
Theoretical Quantiles
lm(efficiency ~ HDD + CDD + weather_region + model)
355
14177
22548
−4
−2
0
2
4
6

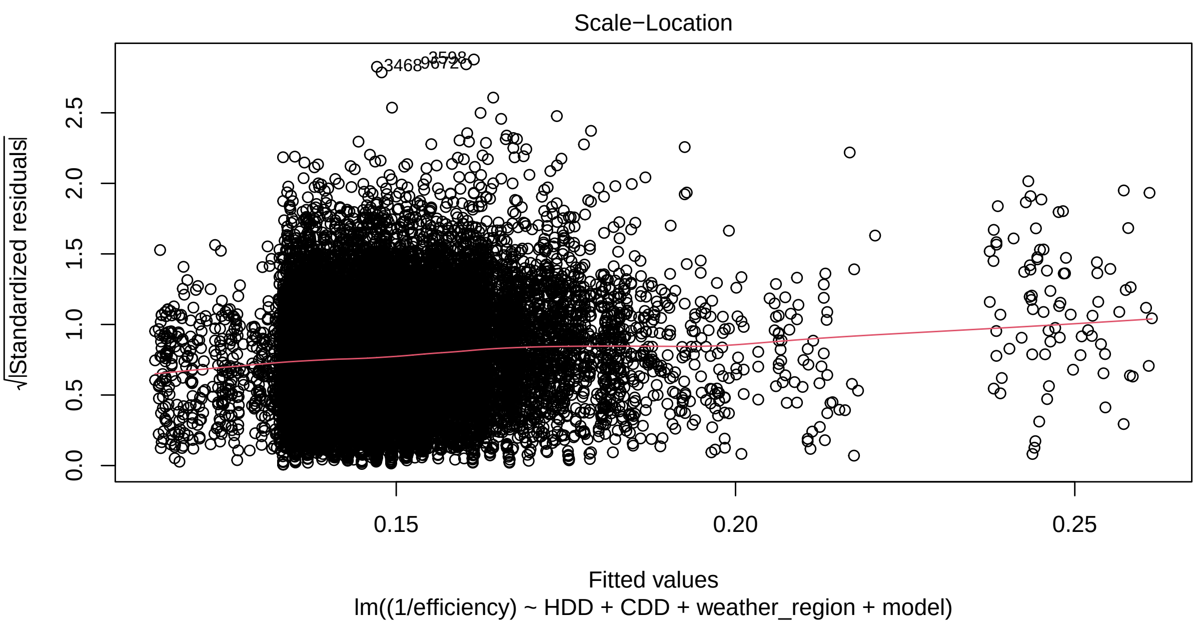

Scale−Location
√|Standardized residuals|
Fitted values
lm(efficiency ~ HDD + CDD + weather_region + model)
355
22548
14177
0.0
0.5
1.0
1.5
2.0
4
5
6
7
8

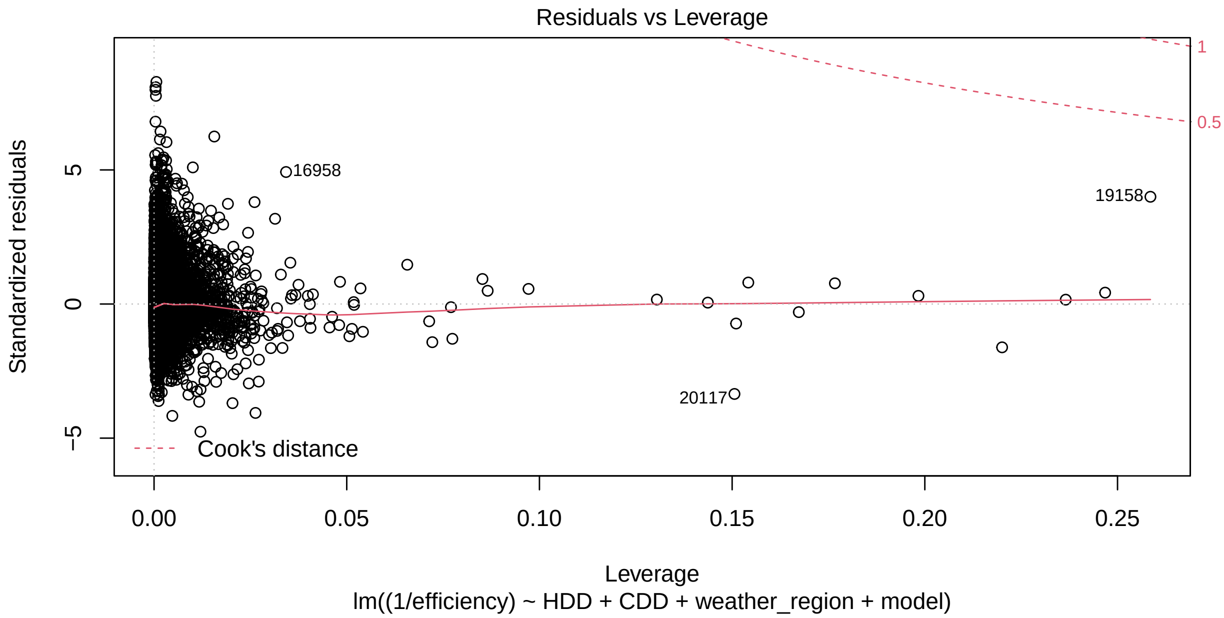



```
plot(eff_h_c_lm)
```




Normal Q−Q
Standardized residuals
−4
−2
0
2
4
6
8
3598
3468
9672
−4
−2
0
2
4
Theoretical Quantiles
lm((1/efficiency) ~ HDD + CDD + weather_region + model)


Scale−Location
√|Standardized residuals|
0.0
0.5
1.0
1.5
2.0
2.5
3468
9672
3598
0.15
0.20
0.25
Fitted values
lm((1/efficiency) ~ HDD + CDD + weather_region + model)



cooling degree days does explain extra variance but not much. likely as not many cooling days above 20 in nz

```
anova(eff_lm, eff_reg_lm, eff_h_c_lm)
```

```
## Warning in anova.lmlist(object, ...): models with response '"(1/efficiency)"'
## removed because response differs from model 1
```

```
## Analysis of Variance Table
##
## Model 1: efficiency ~ HDD + model
## Model 2: efficiency ~ HDD + CDD + weather_region + model
##   Res.Df   RSS Df Sum of Sq      F    Pr(>F)
## 1  22578 12468
## 2  22566 11480 12    988.26 161.89 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```