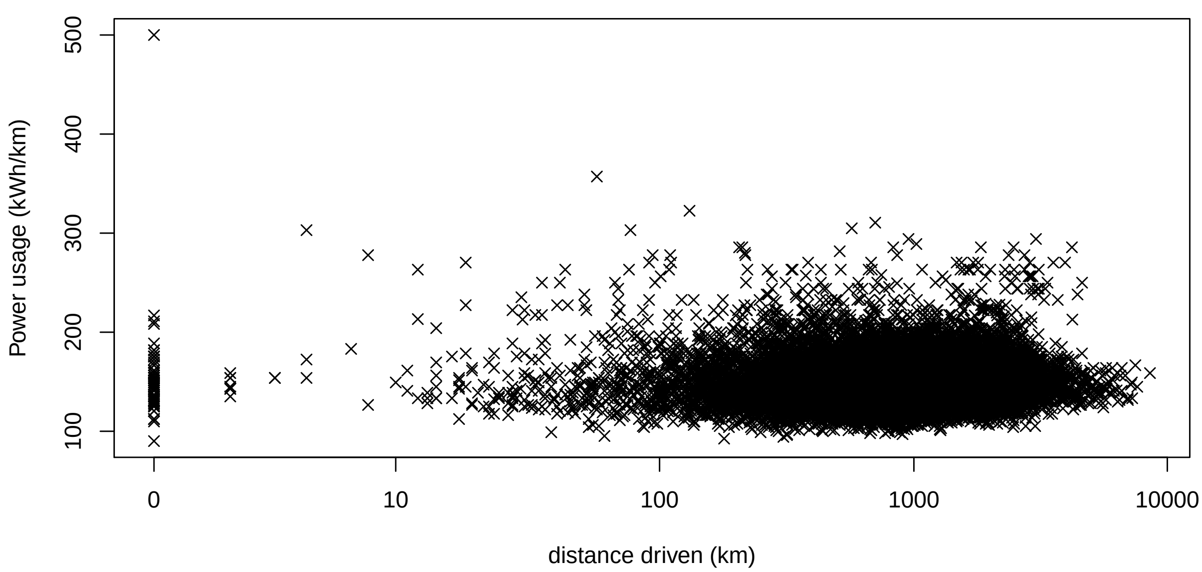
EV data findings

pablo paulsen

23/11/2021

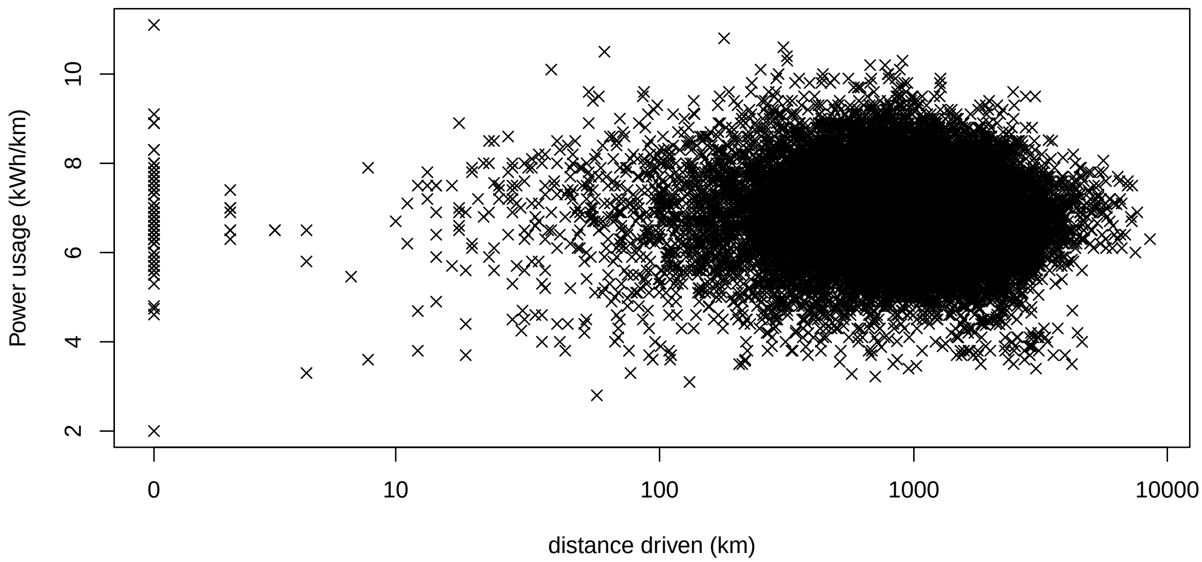
```r
load("processed_data/EV_weather_data.rda")
#EV_data = EV_data[EV_data$efficiency > 4,]
```

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




```r
#bad data vehicles
EV_data = EV_data[EV_data$vehicle != "38f81643" & EV_data$vehicle != "d6082525" &
                    EV_data$vehicle != "ae05b6a6" & EV_data$vehicle != "667822bd" &
                    EV_data$vehicle != "ad7a2a1e" & EV_data$vehicle != "88ee3e2d" &
                    EV_data$vehicle != "e6959846"& EV_data$vehicle != "8a170585" &
                    EV_data$vehicle != "ea66a66d", ]
```

```r
plot(EV_data$distance + 1,EV_data$efficiency, pch = 4,log = "x",
     ylab = "Power usage (kWh/km)", xlab = "distance driven (km)", xaxt = "n")
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: 23 x 3
## # Groups:   region [23]
##    region            weather_region count
##    <chr>             <fct>          <int>
##  1 Auckland          Auckland         326
##  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 13 more rows
```

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

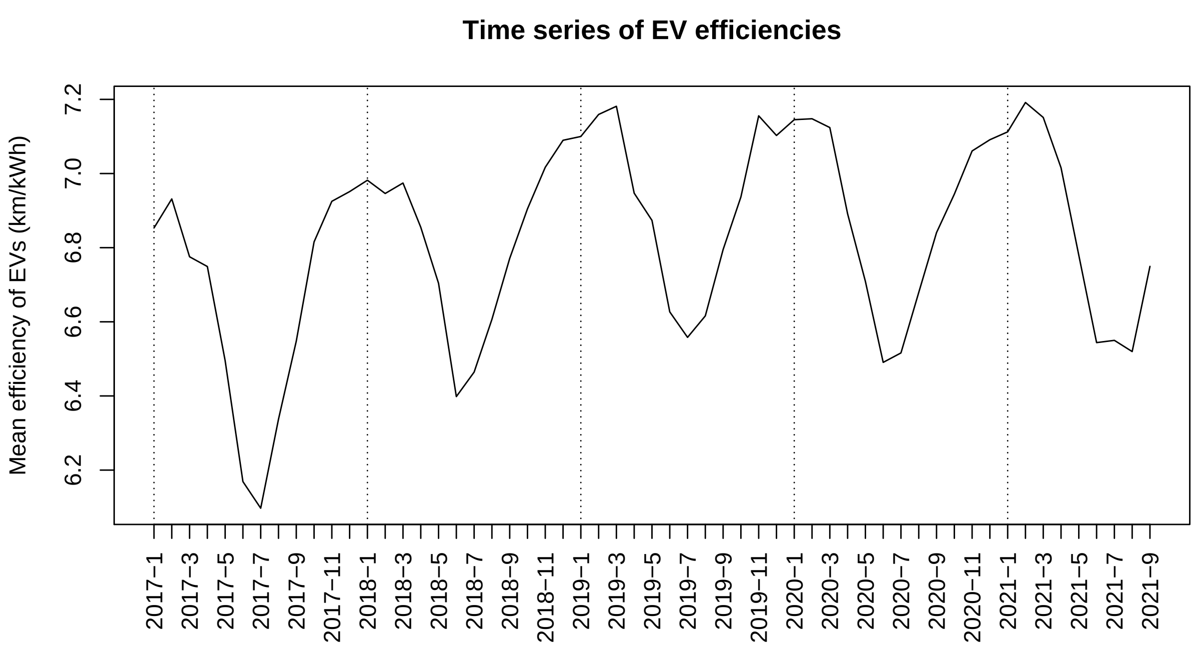
```
## `summarise()` ungrouping output (override with `.groups` argument)
```

```
## # A tibble: 12 x 2
##    weather_region  count
```

```
##    <fct>              <int>
##   1 Auckland            381
##   2 Upper Hutt          248
##   3 Christchurch        226
##   4 Dunedin             136
##   5 Hamilton             64
##   6 Nelson               52
##   7 Rotorua              52
##   8 Clyde                31
##   9 Palmerston North     28
##  10 Stratford            19
##  11 Napier               17
##  12 Invercargill          8
```

```r
plot(monthly_EV_data$m, monthly_EV_data$mean_ef, type = 'l', xaxt = "n", xlab = "",
     ylab = "Mean efficiency of EVs (km/kWh)", main = "Time series of EV efficiencies")
axis(1,labels = paste(monthly_EV_data$year, monthly_EV_data$month, sep = "-"),
     at = monthly_EV_data$m, las = 2,srt = 35)

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 become negative but allows it to better fit the seasonal trend.

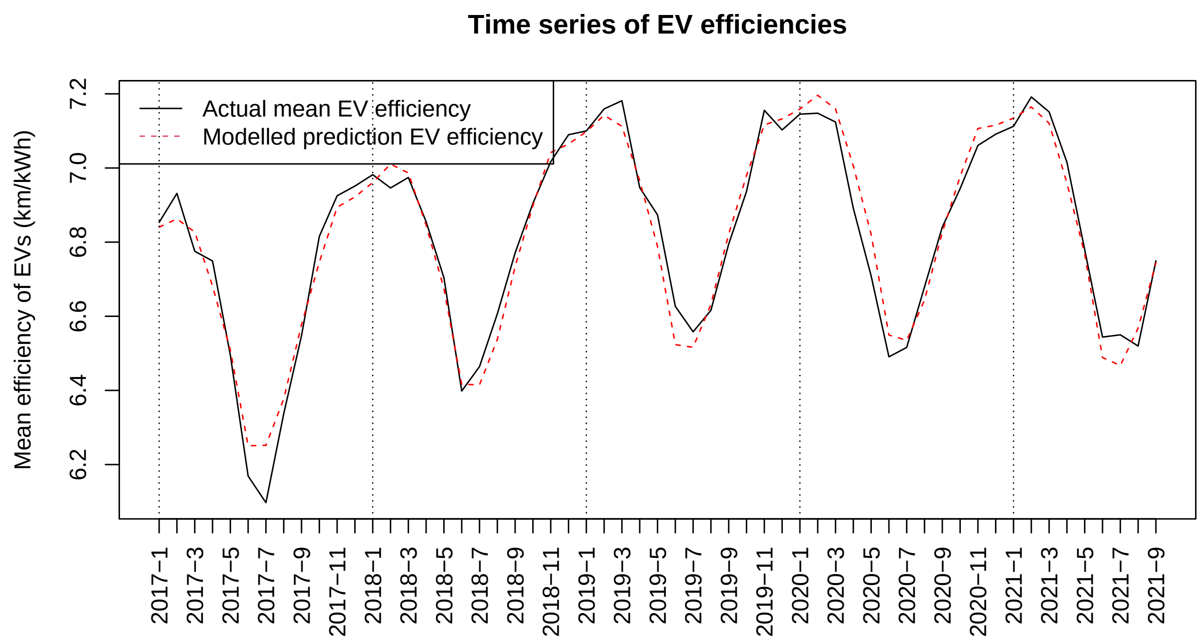
```r
monthly_eff_qm_log = lm(mean_ef ~ m+I(log(m))+I(m^2)+factor(month),
                        data = monthly_EV_data)
summary(monthly_eff_qm_log)
```

```
##
## Call:
## lm(formula = mean_ef ~ m + I(log(m)) + I(m^2) + factor(month),
```

```
##     data = monthly_EV_data)
##
## Residuals:
##       Min        1Q    Median        3Q       Max
## -0.154673 -0.029845  0.005689  0.035518  0.103239
##
## Coefficients:
##                   Estimate Std. Error t value Pr(>|t|)
## (Intercept)      6.814e+00  4.690e-02 145.300  < 2e-16 ***
## m                2.660e-02  5.476e-03   4.858 1.69e-05 ***
## I(log(m))       -5.677e-02  4.140e-02  -1.371 0.177559
## I(m^2)          -3.174e-04  6.301e-05  -5.037 9.44e-06 ***
## factor(month)2   3.604e-02  3.864e-02   0.933 0.356389
## factor(month)3  -1.377e-03  3.901e-02  -0.035 0.971999
## factor(month)4  -1.553e-01  3.931e-02  -3.951 0.000292 ***
## factor(month)5  -3.391e-01  3.954e-02  -8.575 8.95e-11 ***
## factor(month)6  -6.099e-01  3.973e-02 -15.350  < 2e-16 ***
## factor(month)7  -6.225e-01  3.989e-02 -15.604  < 2e-16 ***
## factor(month)8  -5.116e-01  4.004e-02 -12.777 4.62e-16 ***
## factor(month)9  -3.258e-01  4.018e-02  -8.108 3.95e-10 ***
## factor(month)10 -1.718e-01  4.198e-02  -4.091 0.000190 ***
## factor(month)11 -3.824e-02  4.198e-02  -0.911 0.367619
## factor(month)12 -2.451e-02  4.198e-02  -0.584 0.562429
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.06064 on 42 degrees of freedom
## Multiple R-squared:  0.9611, Adjusted R-squared:  0.9481
## F-statistic: 74.03 on 14 and 42 DF,  p-value: < 2.2e-16
```

```r
plot(monthly_EV_data$m, monthly_EV_data$mean_ef, type = 'l', xaxt = "n", xlab = "",
     ylab = "Mean efficiency of EVs (km/kWh)", main = "Time series of EV efficiencies")
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, las = 2,srt = 35)
legend("topleft", legend = c("Actual mean EV efficiency",
                             "Modelled prediction EV efficiency"), lty = 1:2, col = 1:2)
```



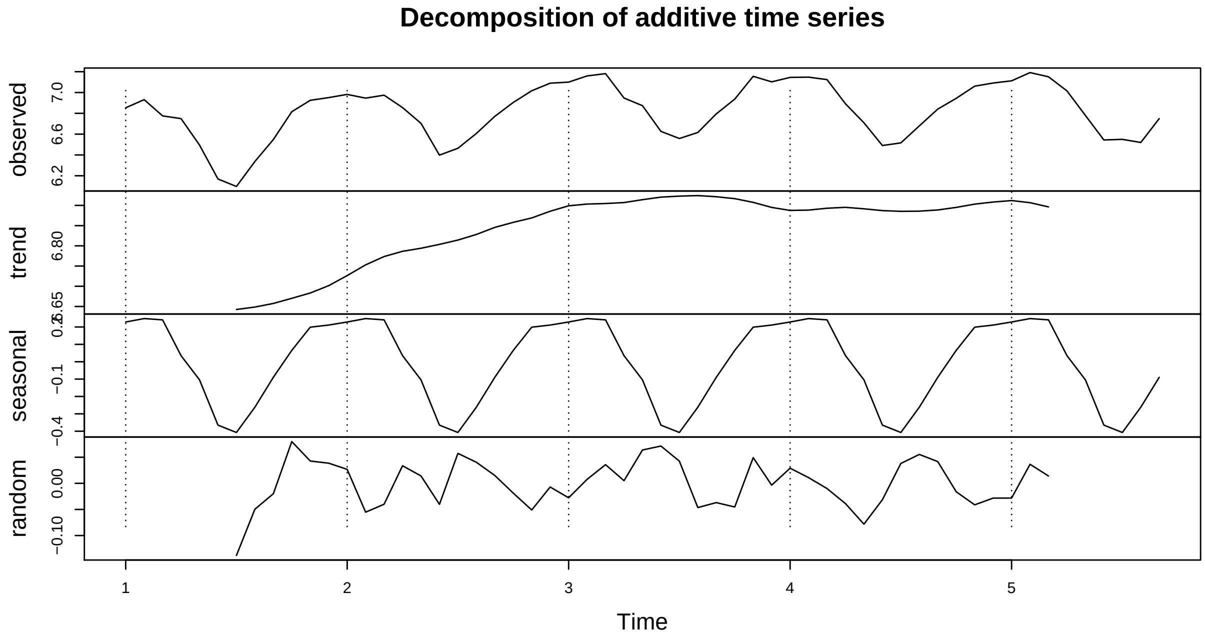

```r
eff_series = ts(monthly_EV_data$mean_ef, frequency = 12)
adf.test(eff_series, alternative = "stationary")
```

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

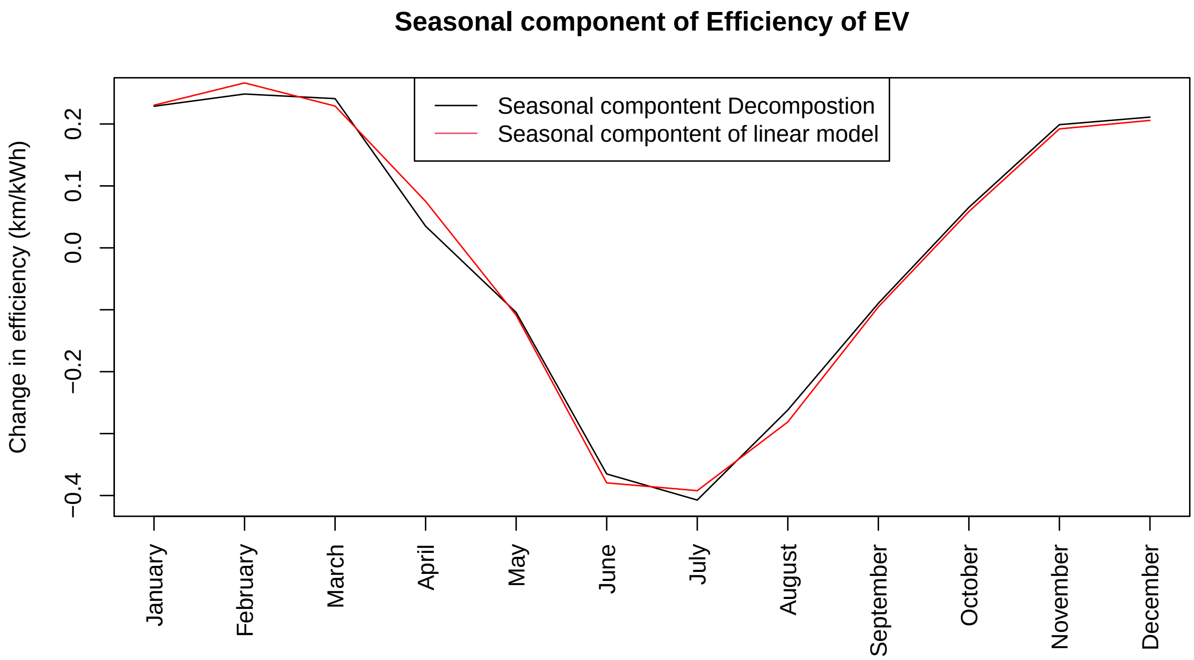
```
##
##  Augmented Dickey-Fuller Test
##
## data:  eff_series
## Dickey-Fuller = -4.4715, Lag order = 3, p-value = 0.01
## alternative hypothesis: stationary
```

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)
```




```r
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 = F),
       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 = 1, col = 1:2)
```



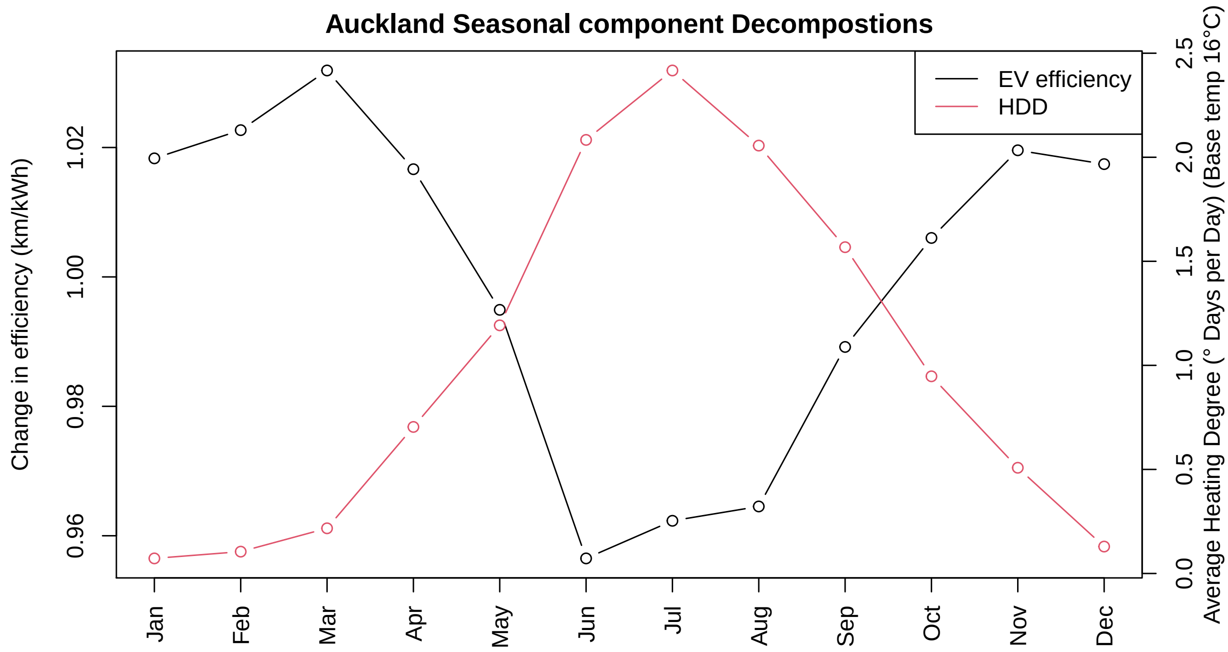
will only do for Auckland as too many lines would get crowded

```
auck_eff_series = ts(monthly_reg_EV_data$mean_ef[
  which(monthly_reg_EV_data$weather_region == "Auckland")], frequency = 12)
decomp_auck_eff = decompose(auck_eff_series, "multiplicative")
auck_HDD_series = ts(monthly_reg_EV_data$HDD[
  monthly_reg_EV_data$weather_region == "Auckland"], frequency = 12)
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 -0.4937 -0.0253  0.4884  3.7192
##
## Coefficients:
##                                       Estimate Std. Error t value Pr(>|t|)
## (Intercept)                           7.558816   0.010264 736.415  < 2e-16 ***
## HDD                                  -0.111227   0.001858 -59.873  < 2e-16 ***
## modelNissan Leaf (30 kWh)            -0.137165   0.012600 -10.886  < 2e-16 ***
## modelNissan Leaf (24 kWh) 2011-2012 -0.604363   0.015152 -39.886  < 2e-16 ***
## modelNissan Leaf (40 kWh)            -0.523063   0.026735 -19.565  < 2e-16 ***
## modelNissan e-NV200 (24 kWh)        -1.272899   0.023890 -53.282  < 2e-16 ***
## modelHyundai Ioniq (EV)               0.839786   0.035130  23.905  < 2e-16 ***
## modelBMW i3                          -0.163967   0.039719  -4.128 3.67e-05 ***
## modelHyundai Kona (EV)               -0.050030   0.047318  -1.057 0.290379
## modelRenault Zoe                     -0.457309   0.044056 -10.380  < 2e-16 ***
## modelTesla Model 3                   -0.591040   0.053875 -10.970  < 2e-16 ***
## modelNissan Leaf (62 kWh)            -1.031878   0.083452 -12.365  < 2e-16 ***
## modelKia Niro (EV)                   -0.483850   0.061146  -7.913 2.63e-15 ***
## modelTesla Model S                   -2.085572   0.082924 -25.150  < 2e-16 ***
## modelVolkswagen e-Golf               -0.072617   0.067158  -1.081 0.279585
## modelTesla Model-X                   -3.065957   0.082905 -36.982  < 2e-16 ***
```

```
## modelKia Soul                          -0.431046   0.068833  -6.262 3.86e-10 ***
## modelMG ZS EV                          -0.739405   0.192003  -3.851 0.000118 ***
## modelRenault Kangoo (van)              -2.159740   0.093186 -23.177  < 2e-16 ***
## modelJaguar I-PACE                     -2.523611   0.129577 -19.476  < 2e-16 ***
## modelPeugeot e-208                     -0.346002   0.332446  -1.041 0.297990
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.7431 on 22571 degrees of freedom
## Multiple R-squared:  0.3323, Adjusted R-squared:  0.3317
## F-statistic: 561.6 on 20 and 22571 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)
```

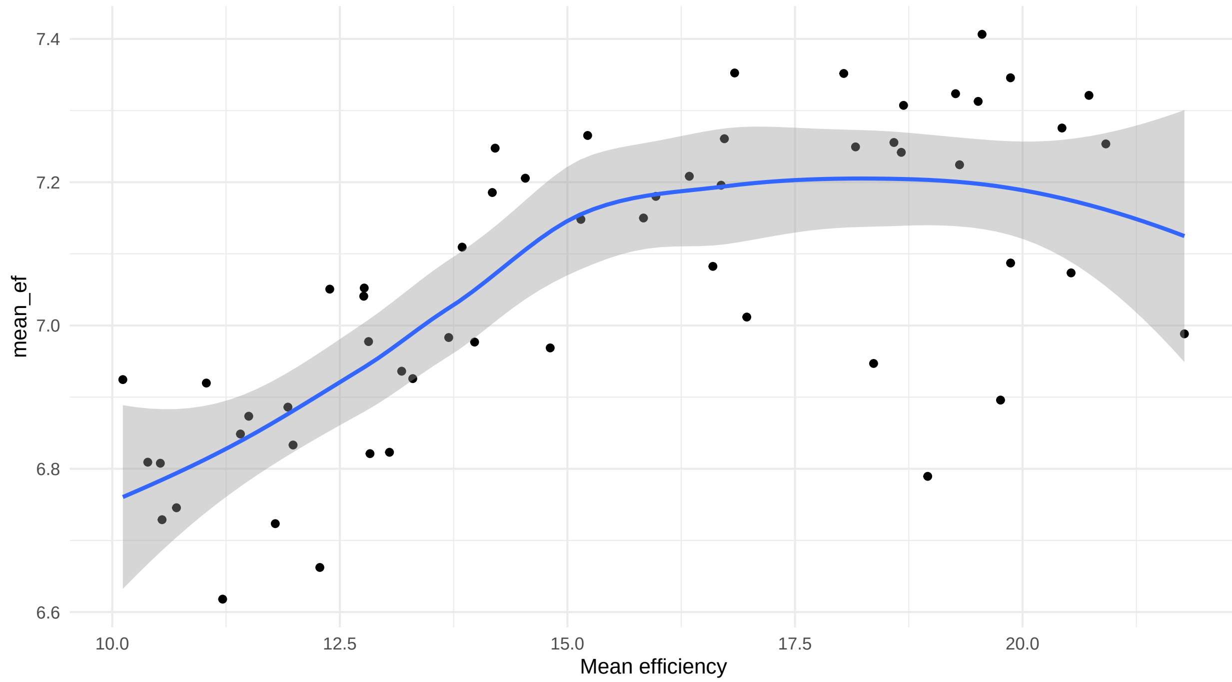
```
##
## Call:
## lm(formula = efficiency ~ HDD + CDD + weather_region + model,
##     data = EV_data, na.action = na.omit)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.7120 -0.4689 -0.0179  0.4665  3.9680
##
## Coefficients:
##                                         Estimate Std. Error t value Pr(>|t|)
## (Intercept)                             7.671636   0.013483 568.984  < 2e-16 ***
## HDD                                    -0.108337   0.002435 -44.490  < 2e-16 ***
## CDD                                    -0.114679   0.027887  -4.112 3.93e-05 ***
## weather_regionUpper Hutt               -0.052323   0.014546  -3.597 0.000322 ***
## weather_regionChristchurch              0.090741   0.015347   5.912 3.42e-09 ***
## weather_regionDunedin                  -0.552394   0.017564 -31.450  < 2e-16 ***
## weather_regionHamilton                 -0.349068   0.025974 -13.439  < 2e-16 ***
## weather_regionNelson                   -0.018558   0.022133  -0.838 0.401778
## weather_regionRotorua                  -0.047164   0.026654  -1.770 0.076820 .
## weather_regionClyde                    -0.145878   0.037585  -3.881 0.000104 ***
## weather_regionPalmerston North         -0.702942   0.033467 -21.004  < 2e-16 ***
## weather_regionStratford                -0.205385   0.044714  -4.593 4.39e-06 ***
## weather_regionNapier                   -0.336149   0.040453  -8.310  < 2e-16 ***
## weather_regionInvercargill             -0.061850   0.069822  -0.886 0.375719
## modelNissan Leaf (30 kWh)              -0.165899   0.012226 -13.569  < 2e-16 ***
## modelNissan Leaf (24 kWh) 2011-2012 -0.632914   0.014652 -43.197  < 2e-16 ***
## modelNissan Leaf (40 kWh)              -0.519590   0.025746 -20.181  < 2e-16 ***
## modelNissan e-NV200 (24 kWh)           -1.305704   0.022977 -56.825  < 2e-16 ***
## modelHyundai Ioniq (EV)                 0.940443   0.034589  27.189  < 2e-16 ***
## modelBMW i3                            -0.048304   0.038681  -1.249 0.211758
## modelHyundai Kona (EV)                 -0.095884   0.045495  -2.108 0.035080 *
```

```
## modelRenault Zoe                        -0.532539   0.042443 -12.547  < 2e-16 ***
## modelTesla Model 3                      -0.598143   0.052153 -11.469  < 2e-16 ***
## modelNissan Leaf (62 kWh)               -1.130704   0.080292 -14.082  < 2e-16 ***
## modelKia Niro (EV)                      -0.509501   0.059808  -8.519  < 2e-16 ***
## modelTesla Model S                      -2.198272   0.079779 -27.555  < 2e-16 ***
## modelVolkswagen e-Golf                  -0.139703   0.065544  -2.131 0.033064 *
## modelTesla Model-X                      -3.143650   0.081130 -38.748  < 2e-16 ***
## modelKia Soul                           -0.472299   0.067696  -6.977 3.10e-12 ***
## modelMG ZS EV                           -0.766802   0.184451  -4.157 3.23e-05 ***
## modelRenault Kangoo (van)               -2.223584   0.092770 -23.969  < 2e-16 ***
## modelJaguar I-PACE                      -2.646534   0.124517 -21.254  < 2e-16 ***
## modelPeugeot e-208                      -0.569080   0.319217  -1.783 0.074643 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.7132 on 22559 degrees of freedom
## Multiple R-squared:  0.3852, Adjusted R-squared:  0.3844
## F-statistic: 441.8 on 32 and 22559 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)) +
  geom_point() +
  geom_smooth(method = 'loess') +
  theme_minimal() +
  labs(x = "Mean efficiency", "Average temperature")
```

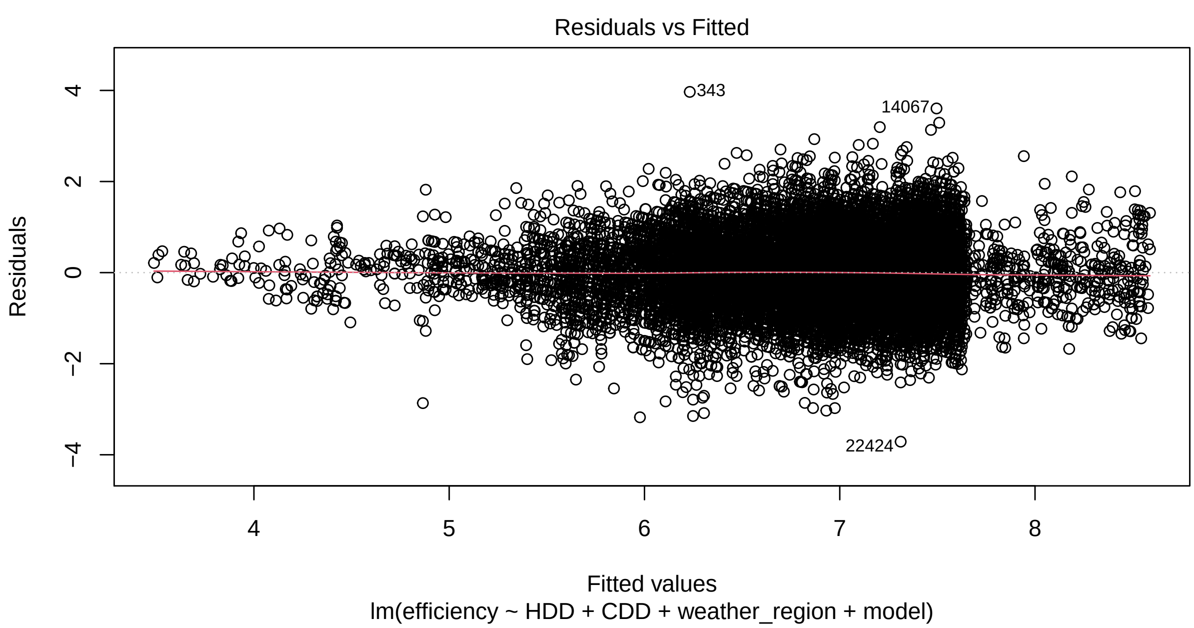


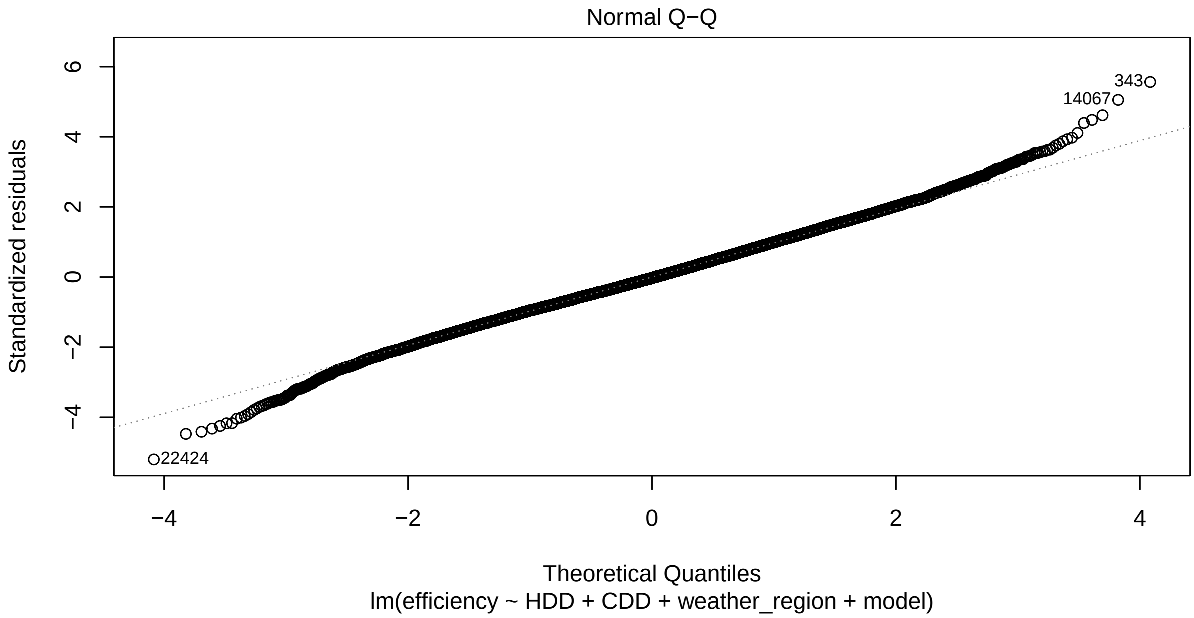

```r
eff_h_c_lm = lm(consumption ~ HDD + CDD + weather_region + model, data = EV_data,
                na.action=na.omit, weights = distance)
summary(eff_h_c_lm)
```

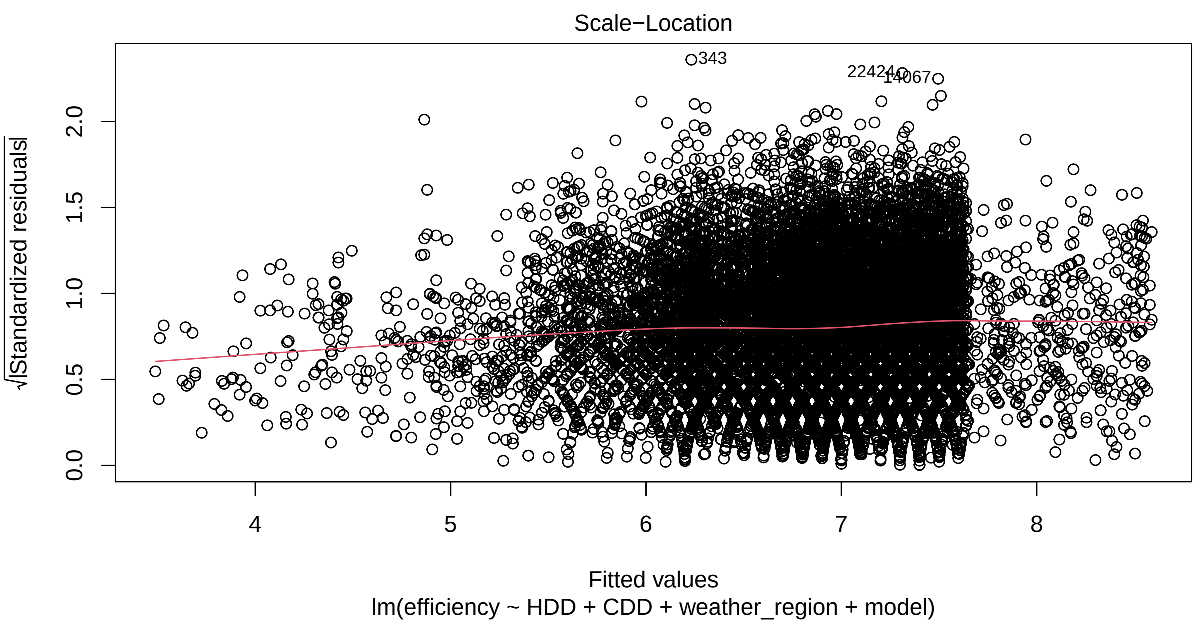
```
##
## Call:
## lm(formula = consumption ~ HDD + CDD + weather_region + model,
##     data = EV_data, weights = distance, na.action = na.omit)
##
## Weighted Residuals:
##     Min      1Q  Median      3Q     Max
## -2330.2  -336.0   -59.0   233.9  4076.5
##
## Coefficients:
##                                    Estimate Std. Error t value Pr(>|t|)
## (Intercept)                       132.11705    0.28672 460.788  < 2e-16 ***
## HDD                                 2.19480    0.05096  43.069  < 2e-16 ***
## CDD                                 2.34713    0.57221   4.102 4.11e-05 ***
## weather_regionUpper Hutt           -0.47962    0.30356  -1.580  0.11412
## weather_regionChristchurch         -0.90734    0.32573  -2.786  0.00535 **
## weather_regionDunedin              12.06016    0.38348  31.449  < 2e-16 ***
## weather_regionHamilton              8.51342    0.52984  16.068  < 2e-16 ***
## weather_regionNelson                2.71145    0.48058   5.642 1.70e-08 ***
## weather_regionRotorua               5.01492    0.54618   9.182  < 2e-16 ***
## weather_regionClyde                 4.53037    0.74912   6.048 1.49e-09 ***
## weather_regionPalmerston North     14.11176    0.66519  21.215  < 2e-16 ***
## weather_regionStratford            10.35783    0.94967  10.907  < 2e-16 ***
## weather_regionNapier                6.31628    0.84725   7.455 9.31e-14 ***
## weather_regionInvercargill          3.19115    1.75793   1.815  0.06949 .
## modelNissan Leaf (30 kWh)           3.40087    0.25243  13.473  < 2e-16 ***
## modelNissan Leaf (24 kWh) 2011-2012 12.38674    0.32455  38.165  < 2e-16 ***
```

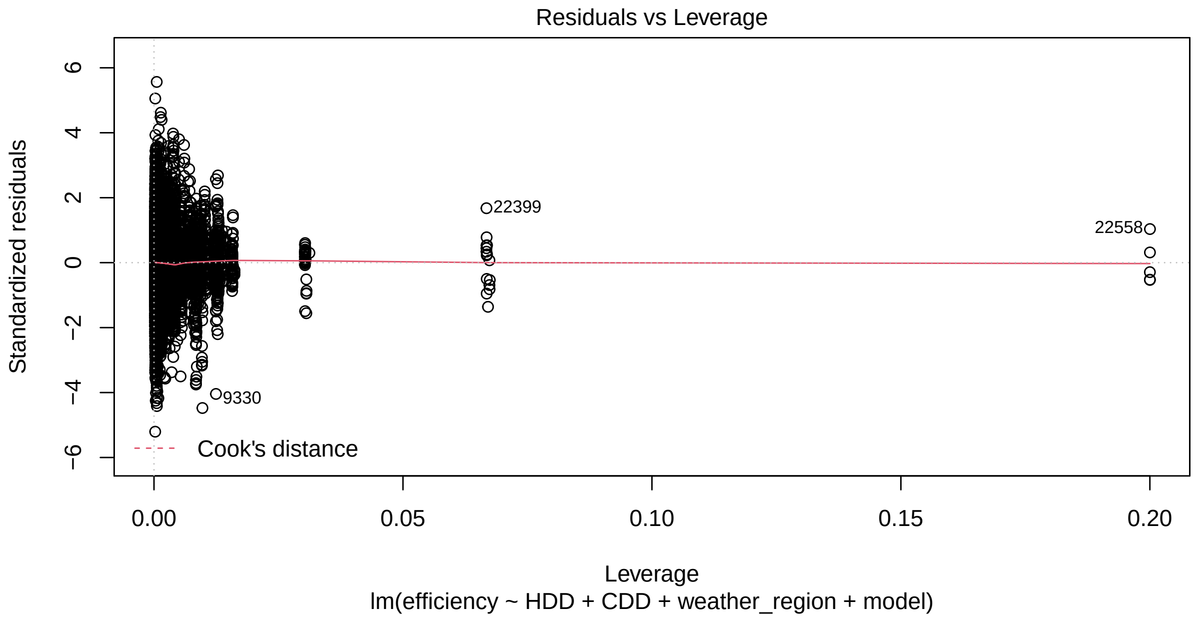
```
## modelNissan Leaf (40 kWh)          10.67531    0.51742  20.632  < 2e-16 ***
## modelNissan e-NV200 (24 kWh)       32.71494    0.53673  60.952  < 2e-16 ***
## modelHyundai Ioniq (EV)           -18.32440    0.68504 -26.750  < 2e-16 ***
## modelBMW i3                        -1.33452    0.78727  -1.695  0.09006 .
## modelHyundai Kona (EV)              0.68218    0.85997   0.793  0.42763
## modelRenault Zoe                   11.54688    0.85068  13.574  < 2e-16 ***
## modelTesla Model 3                 10.54874    1.02214  10.320  < 2e-16 ***
## modelNissan Leaf (62 kWh)          25.45761    1.75213  14.530  < 2e-16 ***
## modelKia Niro (EV)                 11.34383    1.19273   9.511  < 2e-16 ***
## modelTesla Model S                 48.37586    1.68995  28.626  < 2e-16 ***
## modelVolkswagen e-Golf              1.20755    1.53771   0.785  0.43229
## modelTesla Model-X                104.12469    1.29607  80.339  < 2e-16 ***
## modelKia Soul                       6.27576    1.24963   5.022 5.15e-07 ***
## modelMG ZS EV                      22.11736    3.90027   5.671 1.44e-08 ***
## modelRenault Kangoo (van)          56.62855    1.53694  36.845  < 2e-16 ***
## modelJaguar I-PACE                 73.02470    2.95064  24.749  < 2e-16 ***
## modelPeugeot e-208                 10.96380    9.58072   1.144  0.25249
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 492.2 on 22490 degrees of freedom
## Multiple R-squared:  0.4855, Adjusted R-squared:  0.4848
## F-statistic: 663.2 on 32 and 22490 DF,  p-value: < 2.2e-16
```

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

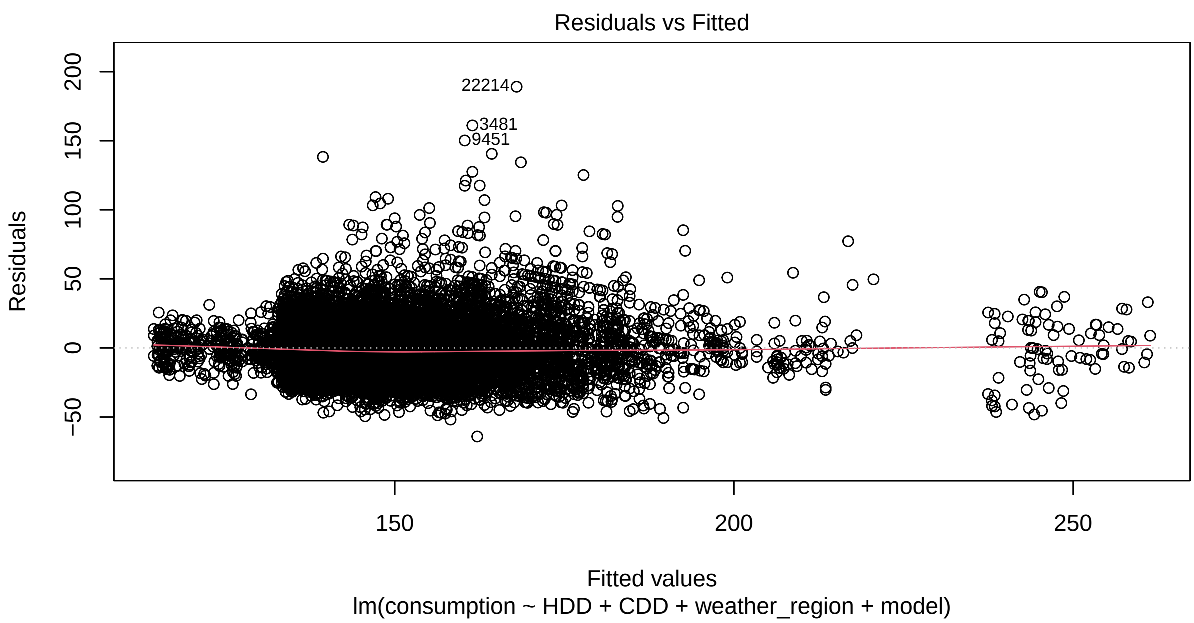


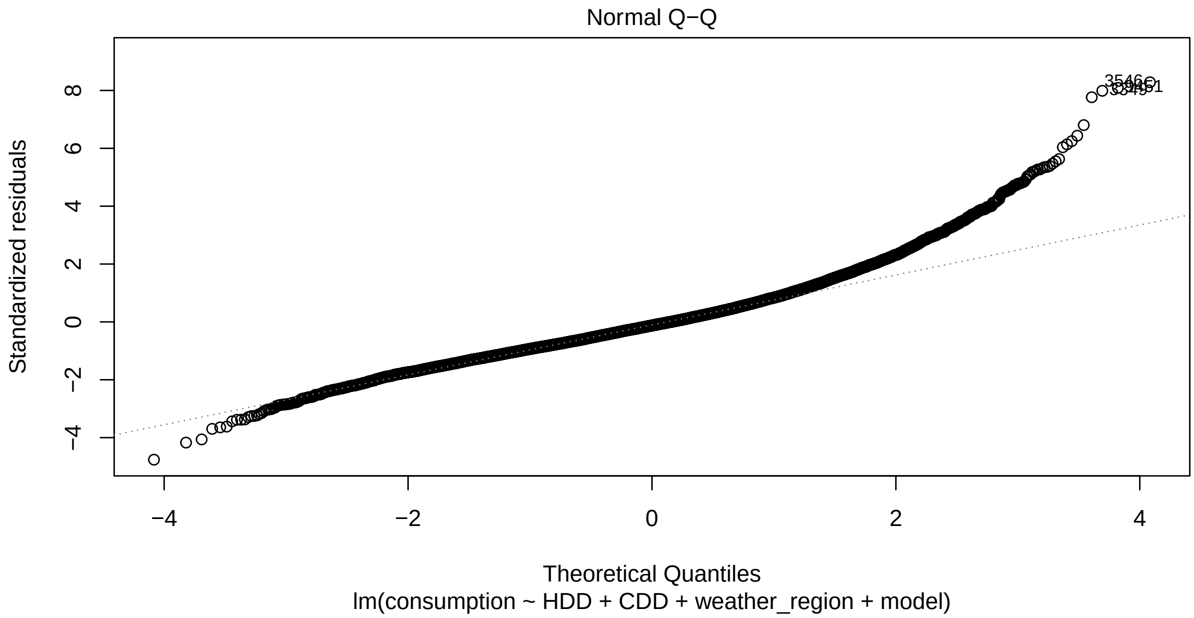

Normal Q−Q
Standardized residuals
Theoretical Quantiles
lm(efficiency ~ HDD + CDD + weather_region + model)
343
14067
22424

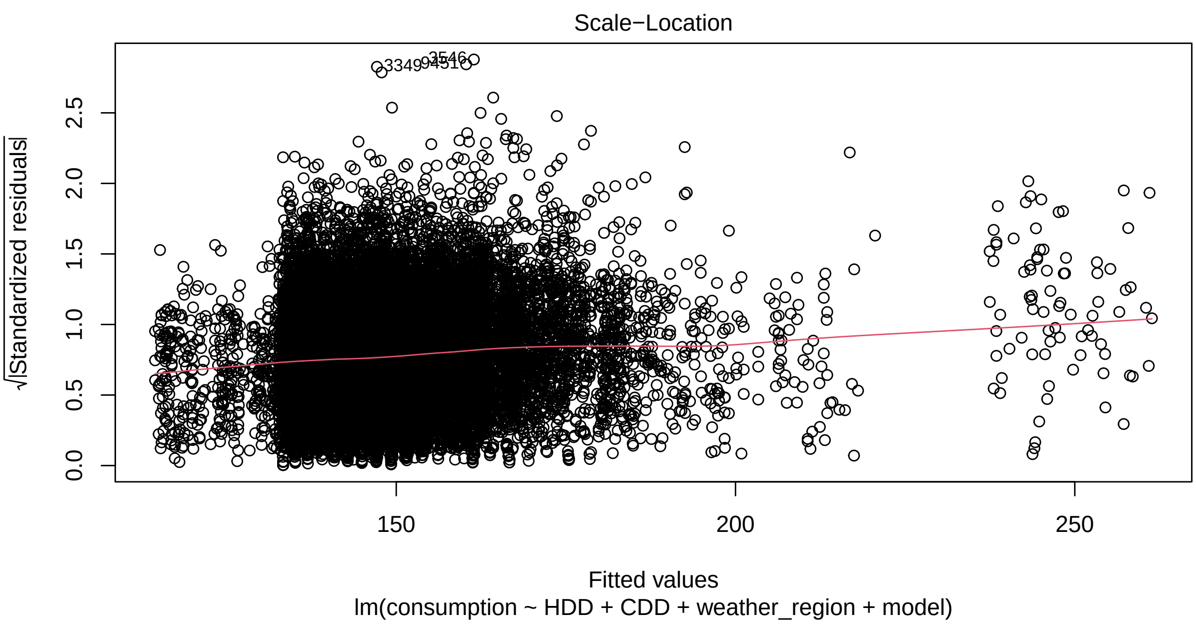

Scale−Location
√|Standardized residuals|
Fitted values
lm(efficiency ~ HDD + CDD + weather_region + model)
343
22424
14067

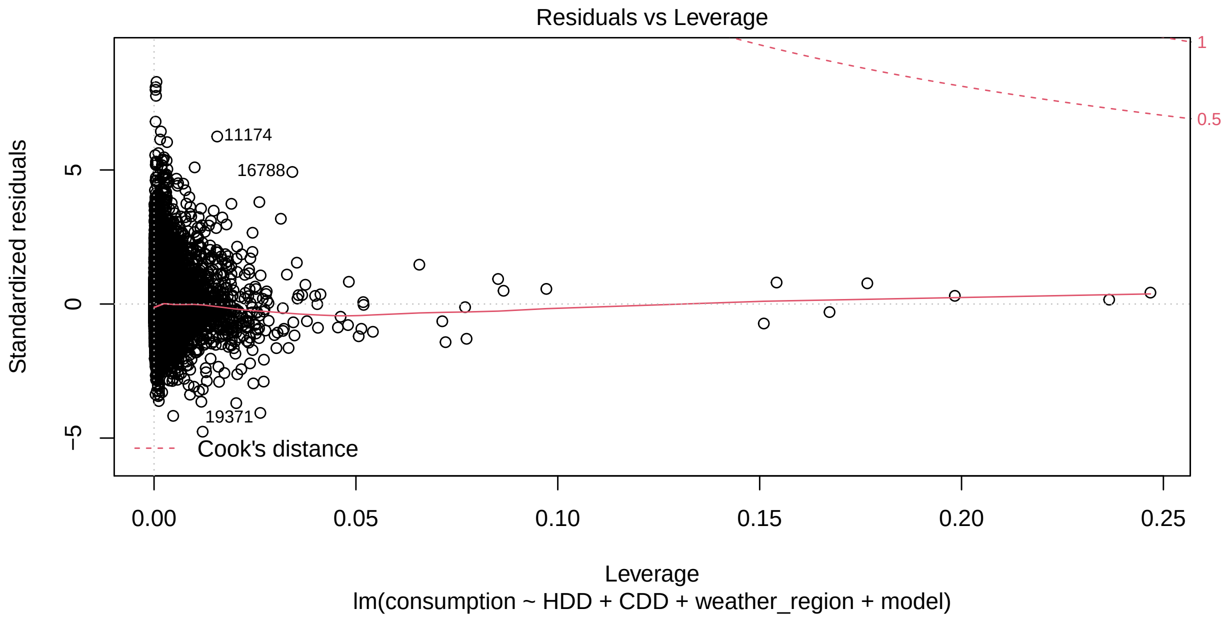



```
plot(eff_h_c_lm)
```




Normal Q−Q
Standardized residuals
Theoretical Quantiles
lm(consumption ~ HDD + CDD + weather_region + model)
3546
3349
4511


Scale−Location
√|Standardized residuals|
Fitted values
lm(consumption ~ HDD + CDD + weather_region + model)
3349
4511
3546



Cooling degree days does explain extra variance but not much. Likely as not many cooling days above 20°C in NZ.

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

```
## Warning in anova.lmlist(object, ...): models with response '"consumption"'
## 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  22571 12462
## 2  22559 11474 12    988.15 161.9 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```