

EDA overall time series

Chloé Lepert

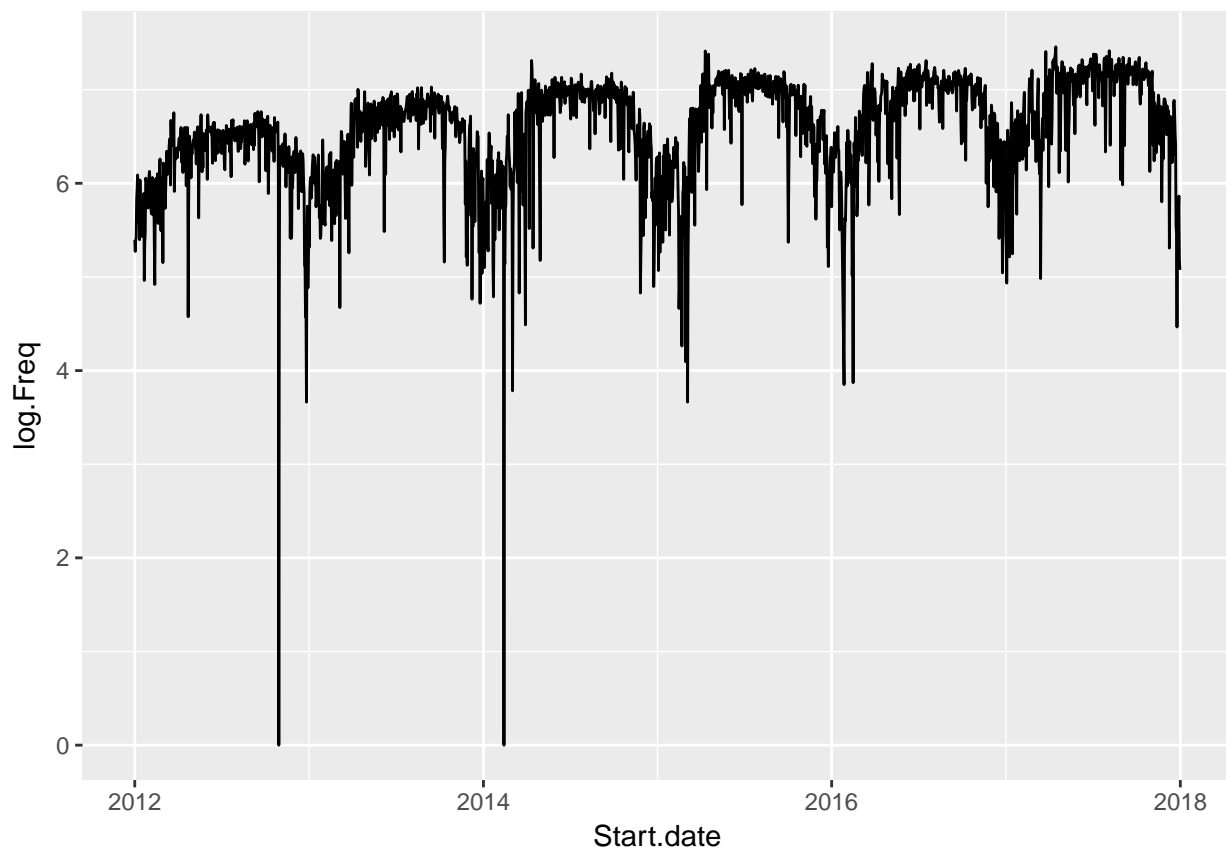
5/9/2018

Notes: # Excluding data prior to 2012

```
overall.daily = aggregate.daily.strong.no.type(df10)
overall.daily = overall.daily[overall.daily$Start.date >= "2012-01-01", ]
overall.daily$log.Freq = log(overall.daily$Freq)
```

Daily traffic

```
ggplot(overall.daily, aes(x = Start.date, y = log.Freq)) + geom_line()
```



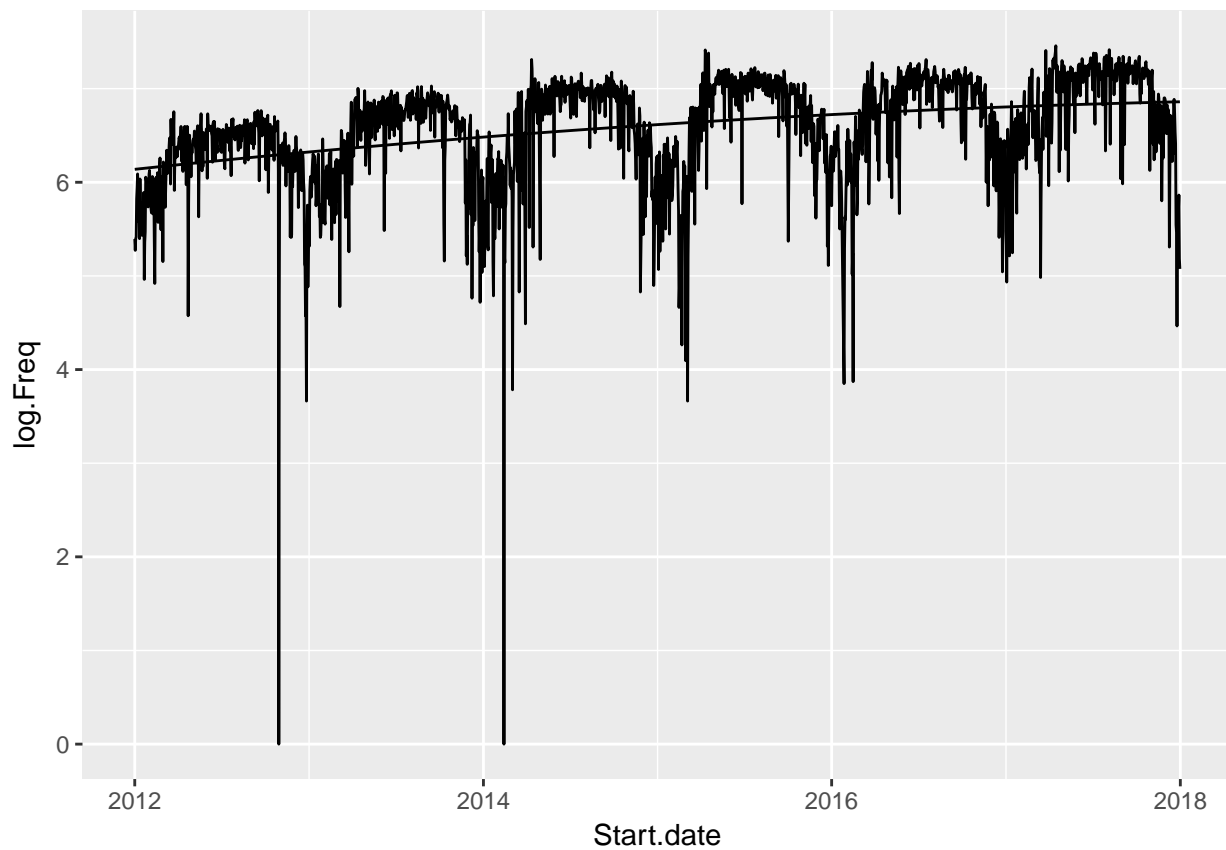
```
trend = lm(log.Freq ~ poly(Start.date, 2), overall.daily)
summary(trend)
```

```
##
## Call:
## lm(formula = log.Freq ~ poly(Start.date, 2), data = overall.daily)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
```

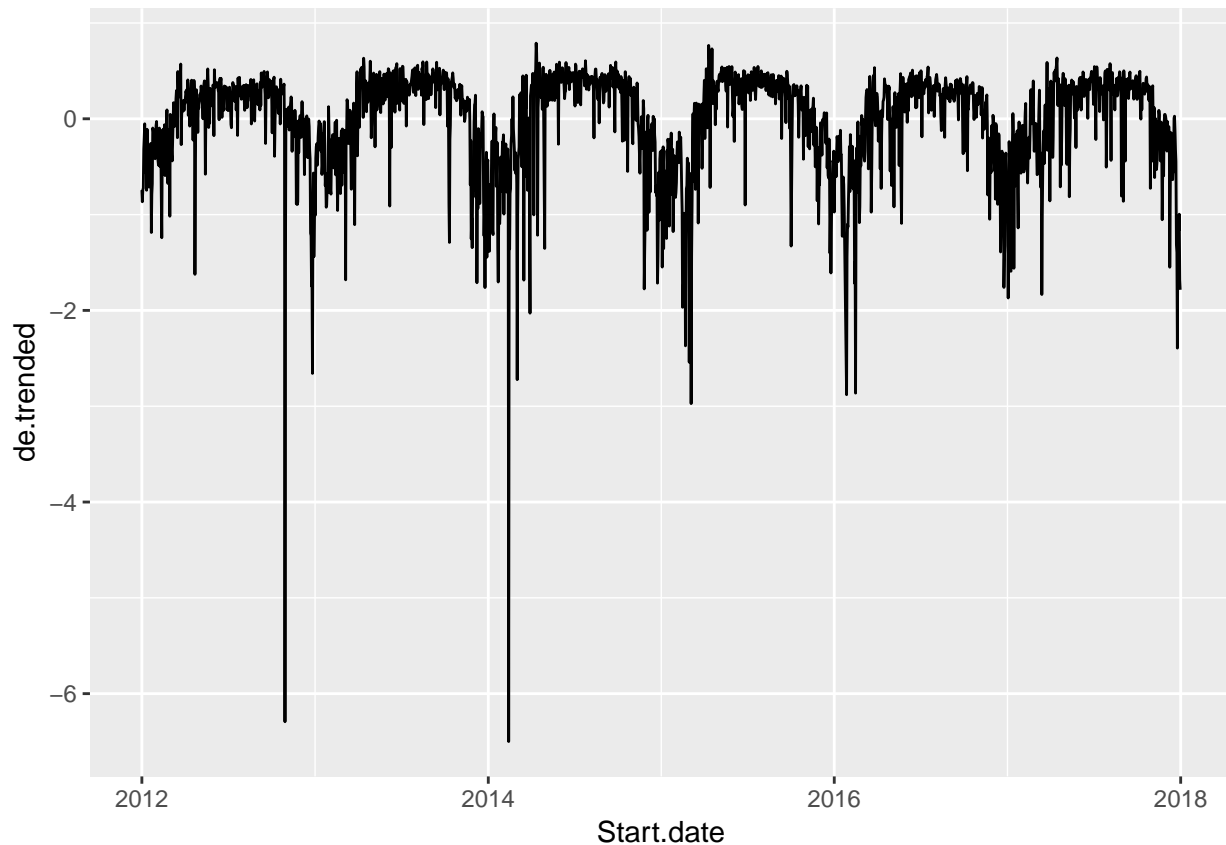
```
## -6.4997 -0.2323 0.1905 0.3602 0.7891
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      6.57645    0.01142 575.764 < 2e-16 ***
## poly(Start.date, 2)1  9.74880    0.53428 18.247 < 2e-16 ***
## poly(Start.date, 2)2 -1.62612    0.53428 -3.044 0.00237 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.5343 on 2185 degrees of freedom
## Multiple R-squared:  0.1354, Adjusted R-squared:  0.1346
## F-statistic: 171.1 on 2 and 2185 DF,  p-value: < 2.2e-16
```

```
overall.daily$trend = fitted(trend)
```

```
ggplot(overall.daily, aes(x = Start.date)) + geom_line(aes(y = log.Freq)) +
  geom_line(aes(y = trend))
```



```
overall.daily$de.trended = overall.daily$log.Freq - overall.daily$trend
ggplot(overall.daily, aes(x = Start.date)) + geom_line(aes(y = de.trended))
```



```
overall.daily$month = month(overall.daily$Start.date)
overall.daily$year = year(overall.daily$Start.date)
season = lm(de.trended ~ as.factor(month), overall.daily)
overall.daily$season = fitted(season)
summary(season)
```

```
##
## Call:
## lm(formula = de.trended ~ as.factor(month), data = overall.daily)
##
## Residuals:
```

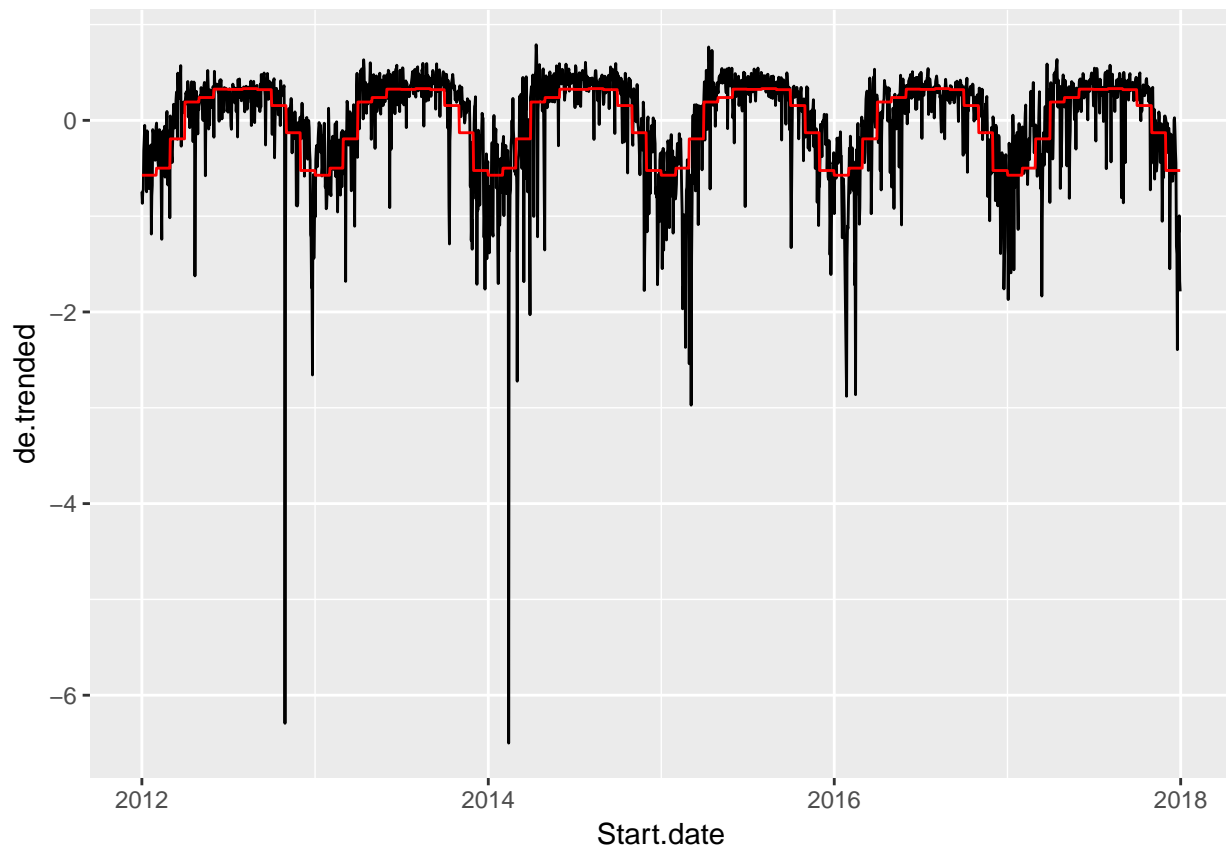
	Min	1Q	Median	3Q	Max
	-6.4479	-0.0964	0.0712	0.1895	0.8939

```
##
## Coefficients:
```

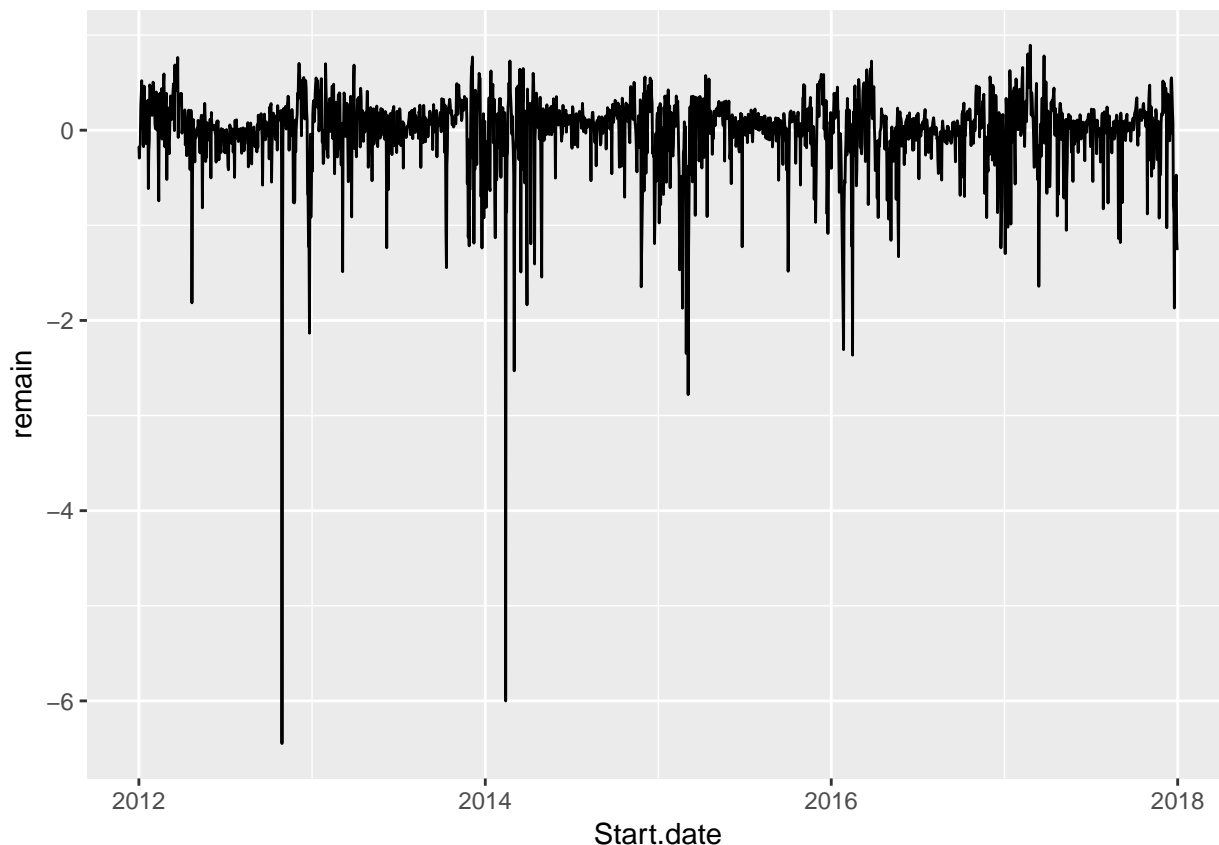
	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	-0.57277	0.03022	-18.951	<2e-16 ***
as.factor(month)2	0.07393	0.04349	1.700	0.0893 .
as.factor(month)3	0.37975	0.04251	8.933	<2e-16 ***
as.factor(month)4	0.76428	0.04286	17.831	<2e-16 ***
as.factor(month)5	0.81112	0.04251	19.080	<2e-16 ***
as.factor(month)6	0.89846	0.04286	20.962	<2e-16 ***
as.factor(month)7	0.89571	0.04251	21.069	<2e-16 ***
as.factor(month)8	0.90417	0.04251	21.268	<2e-16 ***
as.factor(month)9	0.89320	0.04286	20.839	<2e-16 ***
as.factor(month)10	0.72744	0.04251	17.111	<2e-16 ***
as.factor(month)11	0.44361	0.04286	10.350	<2e-16 ***

```
## as.factor(month)12 0.04916 0.04251 1.156 0.2476
## ---
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.4077 on 2176 degrees of freedom
## Multiple R-squared: 0.42, Adjusted R-squared: 0.4171
## F-statistic: 143.2 on 11 and 2176 DF, p-value: < 2.2e-16
```

```
ggplot(overall.daily, aes(x = Start.date)) + geom_line(aes(y = de.trended)) +
  geom_line(aes(y = season), color = "red")
```



```
overall.daily$remain = overall.daily$de.trended - overall.daily$season
ggplot(overall.daily, aes(x = Start.date, y = remain)) + geom_line()
```



```
trend.season = lm(log.Freq ~ poly(year, 2) + as.factor(month), overall.daily)
summary(trend.season)
```

```
##
## Call:
## lm(formula = log.Freq ~ poly(year, 2) + as.factor(month), data = overall.daily)
##
## Residuals:
```

	Min	1Q	Median	3Q	Max
	-6.4522	-0.0978	0.0688	0.1919	0.9331

```
##
## Coefficients:
```

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	5.94622	0.03019	196.966	< 2e-16 ***
poly(year, 2)1	8.93005	0.40727	21.926	< 2e-16 ***
poly(year, 2)2	-1.22267	0.40727	-3.002	0.002712 **
as.factor(month)2	0.08519	0.04344	1.961	0.050009 .
as.factor(month)3	0.40162	0.04246	9.458	< 2e-16 ***
as.factor(month)4	0.79675	0.04281	18.610	< 2e-16 ***
as.factor(month)5	0.85399	0.04246	20.111	< 2e-16 ***
as.factor(month)6	0.95156	0.04281	22.227	< 2e-16 ***
as.factor(month)7	0.95886	0.04246	22.581	< 2e-16 ***
as.factor(month)8	0.97735	0.04246	23.016	< 2e-16 ***
as.factor(month)9	0.97606	0.04281	22.799	< 2e-16 ***
as.factor(month)10	0.81980	0.04246	19.306	< 2e-16 ***
as.factor(month)11	0.54530	0.04281	12.737	< 2e-16 ***
as.factor(month)12	0.15999	0.04246	3.768	0.000169 ***

```
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.4073 on 2174 degrees of freedom
## Multiple R-squared:  0.5002, Adjusted R-squared:  0.4972
## F-statistic: 167.3 on 13 and 2174 DF,  p-value: < 2.2e-16
```

Weather data

```
library(stringr)
weather[is.na(weather) == TRUE] = 0

convert.weather.date = function(date.str) {
  spl = str_split(date.str, "/")[[1]]
  if(nchar(spl[1]) == 1){
    spl[1] = paste("0", spl[1], sep = "")
  }
  if(nchar(spl[2]) == 1){
    spl[2] = paste("0", spl[2], sep = "")
  }
  date = as.Date(paste(spl[3], spl[1], spl[2], sep = "-"))
  return(date)
}

date.list = as.list(weather$DATE)
new.date.list = list()
for(i in 1:length(date.list)) {
  new.date.list[[i]] = convert.weather.date(date.list[[i]])
}

weather$date = as.Date(unlist(new.date.list))
daily.weather = merge(overall.daily, weather, by.x = "Start.date",
                      by.y = "date", all.x = T)
```

Weather as predictor

```
no.weather = lm(log.Freq ~ poly(year, 2) + as.factor(month), daily.weather)
weather = lm(log.Freq ~ poly(year, 2) + as.factor(month) + poly(AWND, 2) + poly(PRCP, 2) +
             poly(SNOW, 2) + poly(TMAX, 2) + poly(TMIN, 2), daily.weather)
summary(weather)
```

```
##
## Call:
## lm(formula = log.Freq ~ poly(year, 2) + as.factor(month) + poly(AWND,
##      2) + poly(PRCP, 2) + poly(SNOW, 2) + poly(TMAX, 2) + poly(TMIN,
##      2), data = daily.weather)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
```

```
## -3.2968 -0.0897 0.0163 0.1208 2.5026
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      6.34724    0.02703 234.808 < 2e-16 ***
## poly(year, 2)1      8.97464    0.27665  32.440 < 2e-16 ***
## poly(year, 2)2     -2.64170    0.27853  -9.484 < 2e-16 ***
## as.factor(month)2    0.02848    0.02950   0.965 0.33456
## as.factor(month)3    0.14113    0.03054   4.621 4.05e-06 ***
## as.factor(month)4    0.29174    0.03410   8.556 < 2e-16 ***
## as.factor(month)5    0.28729    0.03745   7.672 2.55e-14 ***
## as.factor(month)6    0.38349    0.04306   8.907 < 2e-16 ***
## as.factor(month)7    0.39936    0.04754   8.400 < 2e-16 ***
## as.factor(month)8    0.36500    0.04506   8.101 9.00e-16 ***
## as.factor(month)9    0.35875    0.04077   8.799 < 2e-16 ***
## as.factor(month)10   0.31133    0.03519   8.848 < 2e-16 ***
## as.factor(month)11   0.18972    0.03110   6.100 1.26e-09 ***
## as.factor(month)12  -0.01812    0.02961  -0.612 0.54068
## poly(AWND, 2)1     -1.78240    0.29030  -6.140 9.80e-10 ***
## poly(AWND, 2)2     -2.42067    0.28435  -8.513 < 2e-16 ***
## poly(PRCP, 2)1     -8.44105    0.28736 -29.375 < 2e-16 ***
## poly(PRCP, 2)2      0.21438    0.28271   0.758 0.44836
## poly(SNOW, 2)1     -3.99428    0.29270 -13.646 < 2e-16 ***
## poly(SNOW, 2)2     -0.72670    0.28160  -2.581 0.00993 **
## poly(TMAX, 2)1     13.66559    0.89584  15.254 < 2e-16 ***
## poly(TMAX, 2)2     -2.58687    0.53663  -4.821 1.53e-06 ***
## poly(TMIN, 2)1     -2.85855    1.04848  -2.726 0.00646 **
## poly(TMIN, 2)2     -2.73193    0.57607  -4.742 2.25e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.2753 on 2164 degrees of freedom
## Multiple R-squared:  0.7727, Adjusted R-squared:  0.7703
## F-statistic: 319.9 on 23 and 2164 DF, p-value: < 2.2e-16
```

Adding in financials - Note no financials for the weekend (we should detrend with weekends tho)

```
financials$date = as.character(financials$date)
financials$date = paste(substr(financials$date, 1, 4),
                        substr(financials$date, 5, 6),
                        substr(financials$date, 7, 8), sep = "-")
financials$date = as.Date(financials$date)

daily.weather.fin = merge(daily.weather, financials, by.x = "Start.date", by.y = "date", all.x = T)

weather = lm(log.Freq ~ poly(year, 2) + as.factor(month) + poly(AWND, 2) + poly(PRCP, 2) +
              poly(SNOW, 2) + poly(TMAX, 2) + poly(TMIN, 2), daily.weather)
weather.fin = lm(log.Freq ~ poly(year, 2) + as.factor(month) + poly(AWND, 2) +
                 poly(PRCP, 2) + poly(SNOW, 2) + poly(TMAX, 2) + poly(TMIN, 2) +
                 cny + eur + gbp + jpy + snp_index + snp_daily_return + vix +
                 vxo,
```

```

daily.weather.fin)
summary(weather.fin)

```

```

##
## Call:
## lm(formula = log.Freq ~ poly(year, 2) + as.factor(month) + poly(AWND,
##      2) + poly(PRCP, 2) + poly(SNOW, 2) + poly(TMAX, 2) + poly(TMIN,
##      2) + cny + eur + gbp + jpy + snp_index + snp_daily_return +
##      vix + vxo, data = daily.weather.fin)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -3.01135 -0.07246  0.01229  0.09136  2.55633
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    5.006e+00  6.775e-01   7.389 2.54e-13 ***
## poly(year, 2)1    9.592e+00  2.481e+00   3.866 0.000116 ***
## poly(year, 2)2   -1.012e+00  6.322e-01  -1.601 0.109689
## as.factor(month)2 -2.813e-02  3.016e-02  -0.933 0.351153
## as.factor(month)3  6.691e-02  3.227e-02   2.074 0.038306 *
## as.factor(month)4  1.732e-01  3.654e-02   4.740 2.36e-06 ***
## as.factor(month)5  2.098e-01  4.070e-02   5.154 2.91e-07 ***
## as.factor(month)6  2.931e-01  4.614e-02   6.353 2.86e-10 ***
## as.factor(month)7  3.110e-01  5.124e-02   6.070 1.64e-09 ***
## as.factor(month)8  2.874e-01  4.898e-02   5.867 5.52e-09 ***
## as.factor(month)9  2.749e-01  4.625e-02   5.944 3.51e-09 ***
## as.factor(month)10 2.462e-01  4.244e-02   5.801 8.13e-09 ***
## as.factor(month)11 1.743e-01  4.119e-02   4.232 2.46e-05 ***
## as.factor(month)12 2.535e-03  4.328e-02   0.059 0.953305
## poly(AWND, 2)1   -1.448e+00  3.042e-01  -4.759 2.14e-06 ***
## poly(AWND, 2)2   -1.370e+00  3.028e-01  -4.523 6.63e-06 ***
## poly(PRCP, 2)1   -7.183e+00  2.761e-01 -26.022 < 2e-16 ***
## poly(PRCP, 2)2    2.463e+00  3.089e-01   7.974 3.18e-15 ***
## poly(SNOW, 2)1   -4.382e+00  2.715e-01 -16.137 < 2e-16 ***
## poly(SNOW, 2)2   -7.796e-01  2.571e-01  -3.032 0.002472 **
## poly(TMAX, 2)1    1.059e+01  9.191e-01  11.526 < 2e-16 ***
## poly(TMAX, 2)2   -1.076e+00  5.464e-01  -1.969 0.049176 *
## poly(TMIN, 2)1   -6.423e-01  1.088e+00  -0.591 0.554937
## poly(TMIN, 2)2   -3.610e+00  5.786e-01  -6.239 5.83e-10 ***
## cny              -5.670e+00  1.828e+01  -0.310 0.756456
## eur              -1.704e-01  1.322e-01  -1.289 0.197686
## gbp               2.270e-01  1.317e-01   1.723 0.085117 .
## jpy              8.401e+00  3.980e+00   2.111 0.034957 *
## snp_index         9.884e-05  1.468e-04   0.673 0.500830
## snp_daily_return -1.749e-01  8.075e-01  -0.217 0.828538
## vix              -1.723e-02  9.882e-03  -1.743 0.081506 .
## vxo               1.648e-02  1.043e-02   1.581 0.114197
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.2264 on 1397 degrees of freedom
## (759 observations deleted due to missingness)
## Multiple R-squared:  0.8158, Adjusted R-squared:  0.8117

```


F-statistic: 199.6 on 31 and 1397 DF, p-value: < 2.2e-16