Forecasting Dengue Spread

Srihari Seshadri, Siddhartha Khetawat

June 3, 2019

```
# Load libraries and install them, if necessary
tmp.library.list <- c("haven", "zoo", "fUnitRoots", "tseries", "urca", "lmtest", "forecast", "data.table", "readxl","reshap
e", "quantmod", "ggplot2", "reshape2", "plyr","scales", "hts", "fpp2", "lubridate","stargazer","GGally", "dplyr", "lattice",
"skimr")
for (i in 1:length(tmp.library.list)) {
  if (!tmp.library.list[i] %in% rownames(installed.packages())) {
    install.packages(tmp.library.list[i])
  }
  library(tmp.library.list[i], character.only = TRUE)
}
rm(tmp.library.list)
```

Loading the Data

```
trainraw <- data.table(read.csv("dengue_features.csv"), sep=',', stringsAsFactors = F)
trainlraw <- data.table(read.csv("dengue_labels.csv"), sep=',', stringsAsFactors = F)

head(trainraw)
```

```
##    city year weekofyear week_start_date   ndvi_ne   ndvi_nw   ndvi_se
## 1:   sj 1990         18       4/30/1990 0.1226000 0.1037250 0.1984833
## 2:   sj 1990         19        5/7/1990 0.1699000 0.1421750 0.1623571
## 3:   sj 1990         20       5/14/1990 0.0322500 0.1729667 0.1572000
## 4:   sj 1990         21       5/21/1990 0.1286333 0.2450667 0.2275571
## 5:   sj 1990         22       5/28/1990 0.1962000 0.2622000 0.2512000
## 6:   sj 1990         23        6/4/1990        NA 0.1748500 0.2543143
##      ndvi_sw precipitation_amt_mm reanalysis_air_temp_k
## 1: 0.1776167                12.42              297.5729
## 2: 0.1554857                22.82              298.2114
## 3: 0.1708429                34.54              298.7814
## 4: 0.2358857                15.36              298.9871
## 5: 0.2473400                 7.52              299.5186
## 6: 0.1817429                 9.58              299.6300
##    reanalysis_avg_temp_k reanalysis_dew_point_temp_k
## 1:              297.7429                    292.4143
## 2:              298.4429                    293.9514
## 3:              298.8786                    295.4343
## 4:              299.2286                    295.3100
## 5:              299.6643                    295.8214
## 6:              299.7643                    295.8514
##    reanalysis_max_air_temp_k reanalysis_min_air_temp_k
## 1:                     299.8                     295.9
## 2:                     300.9                     296.4
## 3:                     300.5                     297.3
## 4:                     301.4                     297.0
## 5:                     301.9                     297.5
## 6:                     302.4                     298.1
##    reanalysis_precip_amt_kg_per_m2 reanalysis_relative_humidity_percent
## 1:                           32.00                              73.36571
## 2:                           17.94                              77.36857
## 3:                           26.10                              82.05286
## 4:                           13.90                              80.33714
## 5:                           12.20                              80.46000
## 6:                           26.49                              79.89143
##    reanalysis_sat_precip_amt_mm reanalysis_specific_humidity_g_per_kg
## 1:                        12.42                               14.01286
## 2:                        22.82                               15.37286
## 3:                        34.54                               16.84857
```

```
## 4:                       15.36                            16.67286
## 5:                        7.52                            17.21000
## 6:                        9.58                            17.21286
##    reanalysis_tdtr_k station_avg_temp_c station_diur_temp_rng_c
## 1:          2.628571           25.44286                6.900000
## 2:          2.371429           26.71429                6.371429
## 3:          2.300000           26.71429                6.485714
## 4:          2.428571           27.47143                6.771429
## 5:          3.014286           28.94286                9.371429
## 6:          2.100000           28.11429                6.942857
##    station_max_temp_c station_min_temp_c station_precip_mm sep
## 1:               29.4               20.0              16.0   ,
## 2:               31.7               22.2               8.6   ,
## 3:               32.2               22.8              41.4   ,
## 4:               33.3               23.3               4.0   ,
## 5:               35.0               23.9               5.8   ,
## 6:               34.4               23.9              39.1   ,
```

```
head(trainlraw)
```

```
##    city year weekofyear total_cases sep
## 1:   sj 1990         18           4   ,
## 2:   sj 1990         19           5   ,
## 3:   sj 1990         20           4   ,
## 4:   sj 1990         21           3   ,
## 5:   sj 1990         22           6   ,
## 6:   sj 1990         23           2   ,
```

```
dim(trainraw)
```

```
## [1] 936  25
```

```
dim(trainlraw)
```

```
## [1] 936   5
```

Merge the features data with the labels data

```
sjdata <- merge(trainraw, trainlraw, by=c("city", "year", "weekofyear"), all.x = T)
nrow(sjdata) == nrow(trainraw) # if true, we did not accidentally merge many-to-one
```

```
## [1] TRUE
```

```
dim(sjdata)
```

```
## [1] 936  27
```

Omit reanalysis_sat_precip_amt_mm variable because of large amount of missing data

```
sjdata[, reanalysis_sat_precip_amt_mm := NULL]
```

Sort the dataset so it's in a clear order

```
setkeyv(sjdata, c("city", "year", "weekofyear"))

skim(sjdata)
```

```
## Skim summary statistics
##  n obs: 936
##  n variables: 26
##
## -- Variable type:character -----------------------------------------------------
##  variable missing complete   n min max empty n_unique
##     sep.x       0      936 936   1   1     0        1
##     sep.y       0      936 936   1   1     0        1
##
## -- Variable type:factor --------------------------------------------------------
##         variable missing complete   n n_unique
##             city       0      936 936        1
##  week_start_date       0      936 936      936
##                     top_counts ordered
##                  sj: 936, NA: 0   FALSE
##  1/1: 1, 1/1: 1, 1/1: 1, 1/1: 1   FALSE
##
## -- Variable type:integer -------------------------------------------------------
##     variable missing complete   n    mean    sd   p0     p25    p50
##  total_cases       6      930 936   34.34 51.5     0    9      19
##   weekofyear       0      936 936   26.5  15.02    1   13.75   26.5
##         year       0      936 936 1998.83  5.21 1990 1994    1999
##      p75 p100     hist
##    37     461 <U+2587><U+2581><U+2581><U+2581><U+2581><U+2581><U+2581><U+2581>
##    39.25   53 <U+2587><U+2587><U+2587><U+2587><U+2587><U+2587><U+2587><U+2587>
##  2003    2008 <U+2587><U+2585><U+2585><U+2587><U+2585><U+2585><U+2585><U+2586>
##
## -- Variable type:numeric -------------------------------------------------------
##                               variable missing complete   n    mean     sd
##                                ndvi_ne     191      745 936   0.058  0.11
##                                ndvi_nw      49      887 936   0.067  0.092
##                                ndvi_se      19      917 936   0.18   0.057
##                                ndvi_sw      19      917 936   0.17   0.056
##                   precipitation_amt_mm       9      927 936  35.47  44.61
##                  reanalysis_air_temp_k       6      930 936 299.16   1.24
##                  reanalysis_avg_temp_k       6      930 936 299.28   1.22
##             reanalysis_dew_point_temp_k       6      930 936 295.11   1.57
##               reanalysis_max_air_temp_k       6      930 936 301.4    1.26
##               reanalysis_min_air_temp_k       6      930 936 297.3    1.29
```

```
##        reanalysis_precip_amt_kg_per_m2       6      930 936  30.47  35.63
##   reanalysis_relative_humidity_percent       6      930 936  78.57   3.39
##  reanalysis_specific_humidity_g_per_kg       6      930 936  16.55   1.56
##                      reanalysis_tdtr_k       6      930 936   2.52   0.5
##                     station_avg_temp_c       6      930 936  27.01   1.42
##                station_diur_temp_rng_c       6      930 936   6.76   0.84
##                     station_max_temp_c       6      930 936  31.61   1.72
##                     station_min_temp_c       6      930 936  22.6    1.51
##                      station_precip_mm       6      930 936  26.79  29.33
##       p0      p25     p50    p75   p100     hist
##   -0.41    0.0045   0.058   0.11   0.49 <U+2581><U+2581><U+2581><U+2586><U+2587><U+2582><U+2581><U+2581>
##   -0.46    0.016    0.068   0.12   0.44 <U+2581><U+2581><U+2581><U+2582><U+2587><U+2583><U+2581><U+2581>
##   -0.016   0.14     0.18    0.21   0.39 <U+2581><U+2581><U+2585><U+2587><U+2586><U+2582><U+2581><U+2581>
##   -0.063   0.13     0.17    0.2    0.38 <U+2581><U+2581><U+2582><U+2587><U+2587><U+2583><U+2581><U+2581>
##    0       0       20.8    52.18 390.6  <U+2587><U+2582><U+2581><U+2581><U+2581><U+2581><U+2581><U+2581>
##  295.94  298.2    299.25  300.13 302.2  <U+2581><U+2583><U+2586><U+2586><U+2586><U+2587><U+2583><U+2581>
##  296.11  298.3    299.38  300.23 302.16 <U+2581><U+2582><U+2586><U+2586><U+2586><U+2587><U+2585><U+2581>
##  289.64  293.85   295.46  296.42 297.8  <U+2581><U+2581><U+2582><U+2583><U+2583><U+2585><U+2587><U+2583>
##  297.8   300.4    301.5   302.4  304.3  <U+2581><U+2582><U+2585><U+2587><U+2586><U+2587><U+2585><U+2581>
##  292.6   296.3    297.5   298.4  299.9  <U+2581><U+2581><U+2581><U+2585><U+2585><U+2587><U+2587><U+2582>
##    0      10.83    21.3    37    570.5  <U+2587><U+2581><U+2581><U+2581><U+2581><U+2581><U+2581><U+2581>
##   66.74   76.25    78.67   80.96  87.58 <U+2581><U+2581><U+2582><U+2586><U+2587><U+2586><U+2583><U+2581>
##   11.72   15.24    16.85   17.86  19.44 <U+2581><U+2581><U+2583><U+2585><U+2583><U+2586><U+2587><U+2582>
##    1.36    2.16     2.46    2.8    4.43 <U+2581><U+2585><U+2587><U+2586><U+2583><U+2582><U+2581><U+2581>
##   22.84   25.84    27.23   28.19  30.07 <U+2581><U+2581><U+2585><U+2586><U+2585><U+2587><U+2586><U+2581>
##    4.53    6.2      6.76    7.29   9.91 <U+2581><U+2582><U+2586><U+2587><U+2585><U+2582><U+2581><U+2581>
##   26.7    30.6     31.7    32.8   35.6  <U+2581><U+2582><U+2585><U+2586><U+2587><U+2587><U+2583><U+2581>
##   17.8    21.7     22.8    23.9   25.6  <U+2581><U+2581><U+2583><U+2586><U+2583><U+2587><U+2587><U+2582>
##    0       6.82    17.75   35.45 305.9  <U+2587><U+2582><U+2581><U+2581><U+2581><U+2581><U+2581><U+2581>
```

```
head(sjdata)
```

```
##    city year weekofyear week_start_date   ndvi_ne   ndvi_nw   ndvi_se
## 1:   sj 1990         18       4/30/1990 0.1226000 0.1037250 0.1984833
## 2:   sj 1990         19        5/7/1990 0.1699000 0.1421750 0.1623571
## 3:   sj 1990         20       5/14/1990 0.0322500 0.1729667 0.1572000
## 4:   sj 1990         21       5/21/1990 0.1286333 0.2450667 0.2275571
## 5:   sj 1990         22       5/28/1990 0.1962000 0.2622000 0.2512000
## 6:   sj 1990         23        6/4/1990        NA 0.1748500 0.2543143
##      ndvi_sw precipitation_amt_mm reanalysis_air_temp_k
## 1: 0.1776167                12.42              297.5729
## 2: 0.1554857                22.82              298.2114
## 3: 0.1708429                34.54              298.7814
## 4: 0.2358857                15.36              298.9871
## 5: 0.2473400                 7.52              299.5186
## 6: 0.1817429                 9.58              299.6300
##    reanalysis_avg_temp_k reanalysis_dew_point_temp_k
## 1:              297.7429                    292.4143
## 2:              298.4429                    293.9514
## 3:              298.8786                    295.4343
## 4:              299.2286                    295.3100
## 5:              299.6643                    295.8214
## 6:              299.7643                    295.8514
##    reanalysis_max_air_temp_k reanalysis_min_air_temp_k
## 1:                     299.8                     295.9
## 2:                     300.9                     296.4
## 3:                     300.5                     297.3
## 4:                     301.4                     297.0
## 5:                     301.9                     297.5
## 6:                     302.4                     298.1
##    reanalysis_precip_amt_kg_per_m2 reanalysis_relative_humidity_percent
## 1:                           32.00                              73.36571
## 2:                           17.94                              77.36857
## 3:                           26.10                              82.05286
## 4:                           13.90                              80.33714
## 5:                           12.20                              80.46000
## 6:                           26.49                              79.89143
##    reanalysis_specific_humidity_g_per_kg reanalysis_tdtr_k
## 1:                              14.01286          2.628571
## 2:                              15.37286          2.371429
## 3:                              16.84857          2.300000
```

```
## 4:                            16.67286          2.428571
## 5:                            17.21000          3.014286
## 6:                            17.21286          2.100000
##    station_avg_temp_c station_diur_temp_rng_c station_max_temp_c
## 1:           25.44286                6.900000               29.4
## 2:           26.71429                6.371429               31.7
## 3:           26.71429                6.485714               32.2
## 4:           27.47143                6.771429               33.3
## 5:           28.94286                9.371429               35.0
## 6:           28.11429                6.942857               34.4
##    station_min_temp_c station_precip_mm sep.x total_cases sep.y
## 1:               20.0              16.0     ,           4     ,
## 2:               22.2               8.6     ,           5     ,
## 3:               22.8              41.4     ,           4     ,
## 4:               23.3               4.0     ,           3     ,
## 5:               23.9               5.8     ,           6     ,
## 6:               23.9              39.1     ,           2     ,
```

Splitting the data into training and testing datasets

Transform the Date column into the appropriate type

```
sjdata[,week_start_date := as.Date(week_start_date,format = "%m/%d/%Y")]
```

Look at date ranges

```
min(sjdata$week_start_date)
```

```
## [1] "1990-04-30"
```

```
max(sjdata$week_start_date)
```

```
## [1] "2008-04-22"
```

```
# Preserve 5 years of data for model validation
sjdata.train <- sjdata[week_start_date < "2003-01-01"]
sjdata.test <- sjdata[week_start_date >= "2003-01-01"]
```

Descriptive analytics and data visualisation

```
head(sjdata.train)
```

```
##    city year weekofyear week_start_date   ndvi_ne   ndvi_nw   ndvi_se
## 1:   sj 1990         18      1990-04-30 0.1226000 0.1037250 0.1984833
## 2:   sj 1990         19      1990-05-07 0.1699000 0.1421750 0.1623571
## 3:   sj 1990         20      1990-05-14 0.0322500 0.1729667 0.1572000
## 4:   sj 1990         21      1990-05-21 0.1286333 0.2450667 0.2275571
## 5:   sj 1990         22      1990-05-28 0.1962000 0.2622000 0.2512000
## 6:   sj 1990         23      1990-06-04        NA 0.1748500 0.2543143
##      ndvi_sw precipitation_amt_mm reanalysis_air_temp_k
## 1: 0.1776167                12.42              297.5729
## 2: 0.1554857                22.82              298.2114
## 3: 0.1708429                34.54              298.7814
## 4: 0.2358857                15.36              298.9871
## 5: 0.2473400                 7.52              299.5186
## 6: 0.1817429                 9.58              299.6300
##    reanalysis_avg_temp_k reanalysis_dew_point_temp_k
## 1:              297.7429                    292.4143
## 2:              298.4429                    293.9514
## 3:              298.8786                    295.4343
## 4:              299.2286                    295.3100
## 5:              299.6643                    295.8214
## 6:              299.7643                    295.8514
##    reanalysis_max_air_temp_k reanalysis_min_air_temp_k
## 1:                     299.8                     295.9
## 2:                     300.9                     296.4
## 3:                     300.5                     297.3
## 4:                     301.4                     297.0
## 5:                     301.9                     297.5
## 6:                     302.4                     298.1
##    reanalysis_precip_amt_kg_per_m2 reanalysis_relative_humidity_percent
## 1:                           32.00                              73.36571
## 2:                           17.94                              77.36857
## 3:                           26.10                              82.05286
## 4:                           13.90                              80.33714
## 5:                           12.20                              80.46000
## 6:                           26.49                              79.89143
##    reanalysis_specific_humidity_g_per_kg reanalysis_tdtr_k
## 1:                              14.01286          2.628571
## 2:                              15.37286          2.371429
## 3:                              16.84857          2.300000
```

```
## 4:                            16.67286          2.428571
## 5:                            17.21000          3.014286
## 6:                            17.21286          2.100000
##    station_avg_temp_c station_diur_temp_rng_c station_max_temp_c
## 1:           25.44286                6.900000               29.4
## 2:           26.71429                6.371429               31.7
## 3:           26.71429                6.485714               32.2
## 4:           27.47143                6.771429               33.3
## 5:           28.94286                9.371429               35.0
## 6:           28.11429                6.942857               34.4
##    station_min_temp_c station_precip_mm sep.x total_cases sep.y
## 1:               20.0              16.0     ,           4     ,
## 2:               22.2               8.6     ,           5     ,
## 3:               22.8              41.4     ,           4     ,
## 4:               23.3               4.0     ,           3     ,
## 5:               23.9               5.8     ,           6     ,
## 6:               23.9              39.1     ,           2     ,
```

```
# Summary
summary(sjdata.train)
```

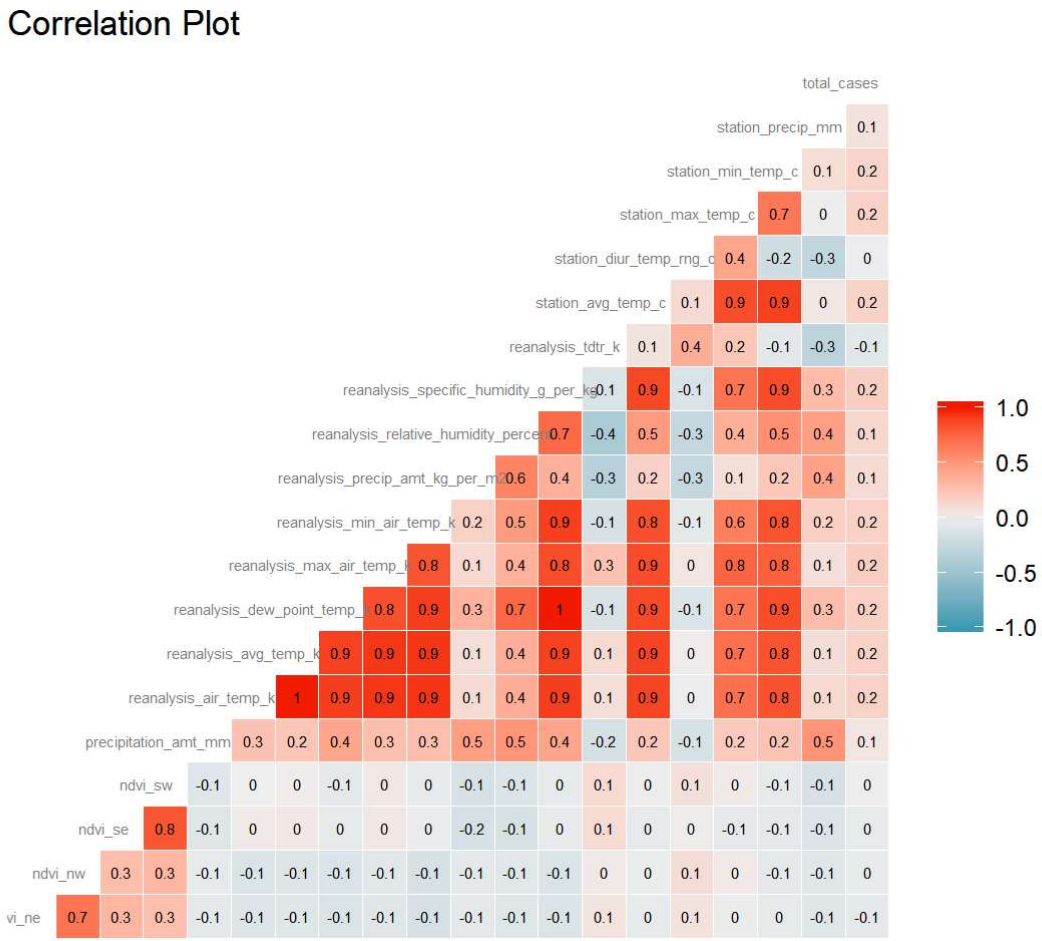
```
##  city          year        weekofyear    week_start_date
##  sj:659   Min.   :1990   Min.   : 1.00   Min.   :1990-04-30
##           1st Qu.:1993   1st Qu.:14.00   1st Qu.:1993-06-28
##           Median :1996   Median :27.00   Median :1996-08-26
##           Mean   :1996   Mean   :26.95   Mean   :1996-08-26
##           3rd Qu.:1999   3rd Qu.:40.00   3rd Qu.:1999-10-25
##           Max.   :2002   Max.   :53.00   Max.   :2002-12-24
##
##     ndvi_ne            ndvi_nw            ndvi_se          ndvi_sw
##  Min.   :-0.29020   Min.   :-0.25280   Min.   :0.0360   Min.   :-0.06346
##  1st Qu.: 0.03332   1st Qu.: 0.04777   1st Qu.:0.1451   1st Qu.: 0.13303
##  Median : 0.08085   Median : 0.08838   Median :0.1806   Median : 0.16887
##  Mean   : 0.08504   Mean   : 0.09475   Mean   :0.1814   Mean   : 0.16859
##  3rd Qu.: 0.12755   3rd Qu.: 0.13681   3rd Qu.:0.2136   3rd Qu.: 0.20444
##  Max.   : 0.44627   Max.   : 0.43710   Max.   :0.3931   Max.   : 0.38142
##  NA's   :139        NA's   :39         NA's   :18       NA's   :18
##  precipitation_amt_mm reanalysis_air_temp_k reanalysis_avg_temp_k
##  Min.   :  0.0000     Min.   :295.9         Min.   :296.1
##  1st Qu.:  0.9975     1st Qu.:298.0         1st Qu.:298.2
##  Median : 21.1500     Median :299.2         Median :299.3
##  Mean   : 34.2270     Mean   :299.0         Mean   :299.1
##  3rd Qu.: 50.5800     3rd Qu.:300.0         3rd Qu.:300.1
##  Max.   :287.5500     Max.   :301.3         Max.   :301.4
##  NA's   :7            NA's   :4             NA's   :4
##  reanalysis_dew_point_temp_k reanalysis_max_air_temp_k
##  Min.   :289.6               Min.   :298.2
##  1st Qu.:293.8               1st Qu.:300.4
##  Median :295.4               Median :301.4
##  Mean   :295.1               Mean   :301.3
##  3rd Qu.:296.4               3rd Qu.:302.3
##  Max.   :297.5               Max.   :303.9
##  NA's   :4                   NA's   :4
##  reanalysis_min_air_temp_k reanalysis_precip_amt_kg_per_m2
##  Min.   :292.6             Min.   :  0.00
##  1st Qu.:296.2             1st Qu.: 11.75
##  Median :297.5             Median : 22.26
##  Mean   :297.2             Mean   : 31.99
##  3rd Qu.:298.2             3rd Qu.: 40.35
##  Max.   :299.5             Max.   :570.50
```

```
##  NA's   :4                NA's   :4
##  reanalysis_relative_humidity_percent
##  Min.   :66.74
##  1st Qu.:76.84
##  Median :79.16
##  Mean   :79.00
##  3rd Qu.:81.42
##  Max.   :87.30
##  NA's   :4
##  reanalysis_specific_humidity_g_per_kg reanalysis_tdtr_k
##  Min.   :11.72                         Min.   :1.357
##  1st Qu.:15.24                         1st Qu.:2.086
##  Median :16.83                         Median :2.357
##  Mean   :16.51                         Mean   :2.437
##  3rd Qu.:17.81                         3rd Qu.:2.693
##  Max.   :19.04                         Max.   :4.100
##  NA's   :4                             NA's   :4
##  station_avg_temp_c station_diur_temp_rng_c station_max_temp_c
##  Min.   :22.84      Min.   :4.529          Min.   :27.20
##  1st Qu.:25.93      1st Qu.:6.364          1st Qu.:30.60
##  Median :27.29      Median :6.857          Median :32.20
##  Mean   :27.05      Mean   :6.866          Mean   :31.72
##  3rd Qu.:28.19      3rd Qu.:7.386          3rd Qu.:32.80
##  Max.   :30.07      Max.   :9.914          Max.   :35.60
##  NA's   :4          NA's   :4              NA's   :4
##  station_min_temp_c station_precip_mm    sep.x            total_cases
##  Min.   :17.80      Min.   :  0.00    Length:659         Min.   :  0.00
##  1st Qu.:21.70      1st Qu.:  6.55    Class :character   1st Qu.: 11.00
##  Median :22.80      Median : 16.60    Mode  :character   Median : 23.00
##  Mean   :22.59      Mean   : 25.12                       Mean   : 39.73
##  3rd Qu.:23.90      3rd Qu.: 32.40                       3rd Qu.: 44.00
##  Max.   :25.60      Max.   :305.90                       Max.   :461.00
##  NA's   :4          NA's   :4                            NA's   :6
##     sep.y
##  Length:659
##  Class :character
##  Mode  :character
##
##
```

```
##
##
```

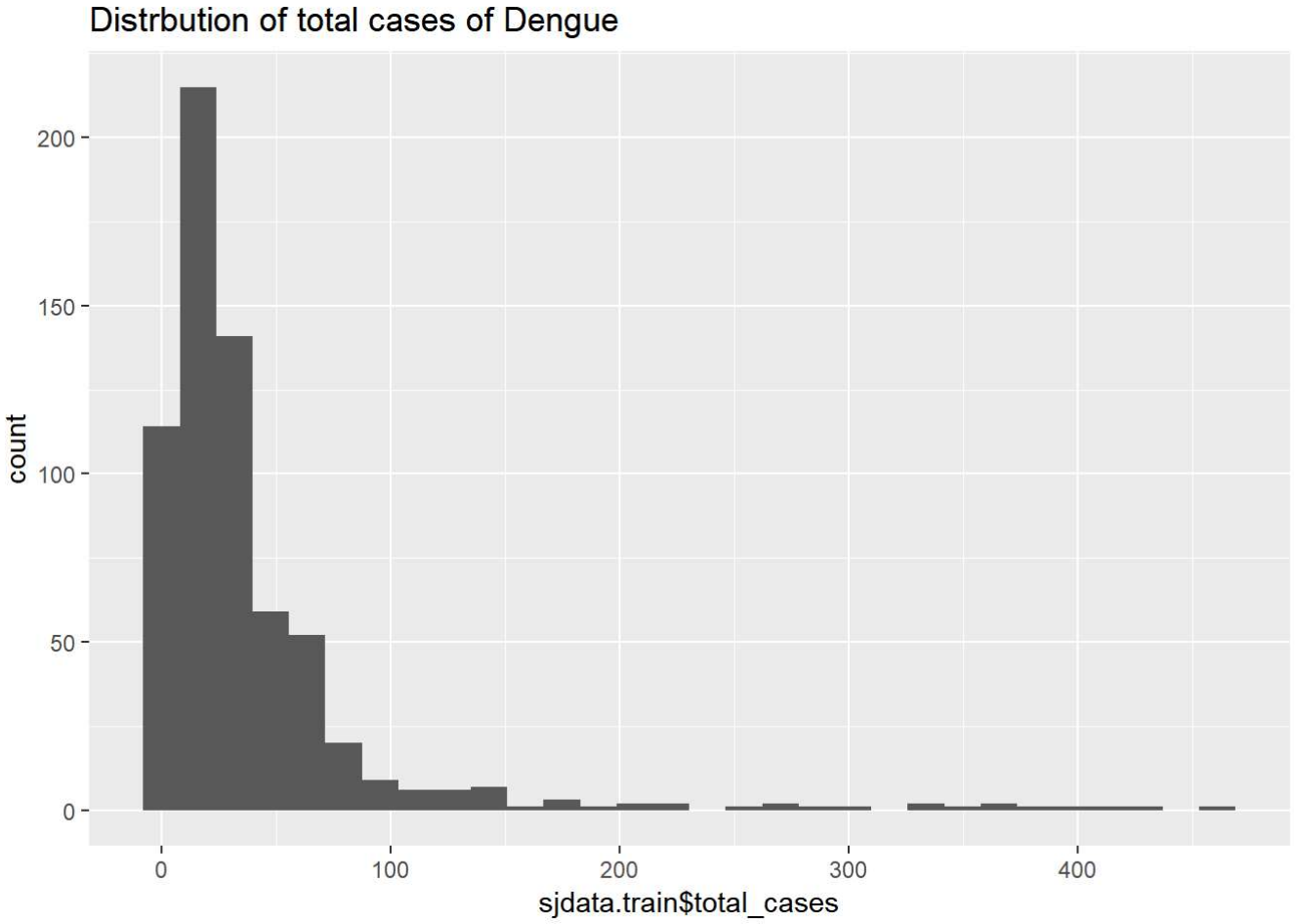
```
# Correlation plots
ggcorr(sjdata.train[, 5:26], label = TRUE, hjust = 0.85, size = 2,color = "grey50",
       label_size = 2) +
  ggplot2::labs(title = "Correlation Plot")
```

```
## Warning in ggcorr(sjdata.train[, 5:26], label = TRUE, hjust = 0.85, size =
## 2, : data in column(s) 'sep.x', 'sep.y' are not numeric and were ignored
```



```
ggsave("./Charts/EDA_correlation.png", width = 8, height = 6)
```

```
ggplot(data=sjdata.train, aes(sjdata.train$total_cases)) +
  geom_histogram() +
  ggtitle("Distrbution of total cases of Dengue")
```




```
ggsave("./Charts/total_cases_distribution.png", width = 8, height = 6)
```

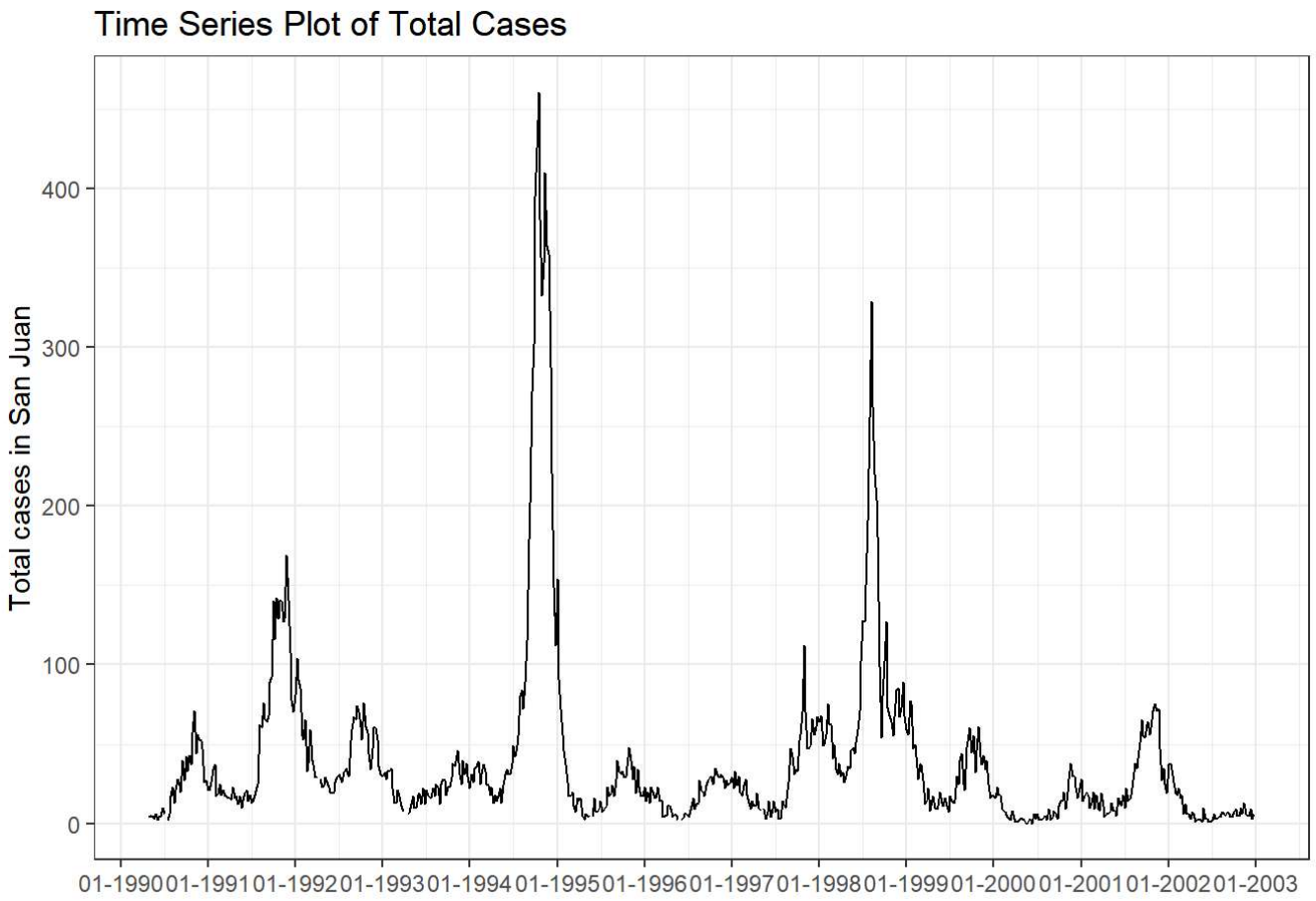
```
(missing_idxs <- c(which(is.na(sjdata.train$total_cases), arr.ind=TRUE)))
```

```
## [1]  11 102 153 263 317 366
```

```
sjdata.train[missing_idxs,] %>% select(year, weekofyear, total_cases)
```

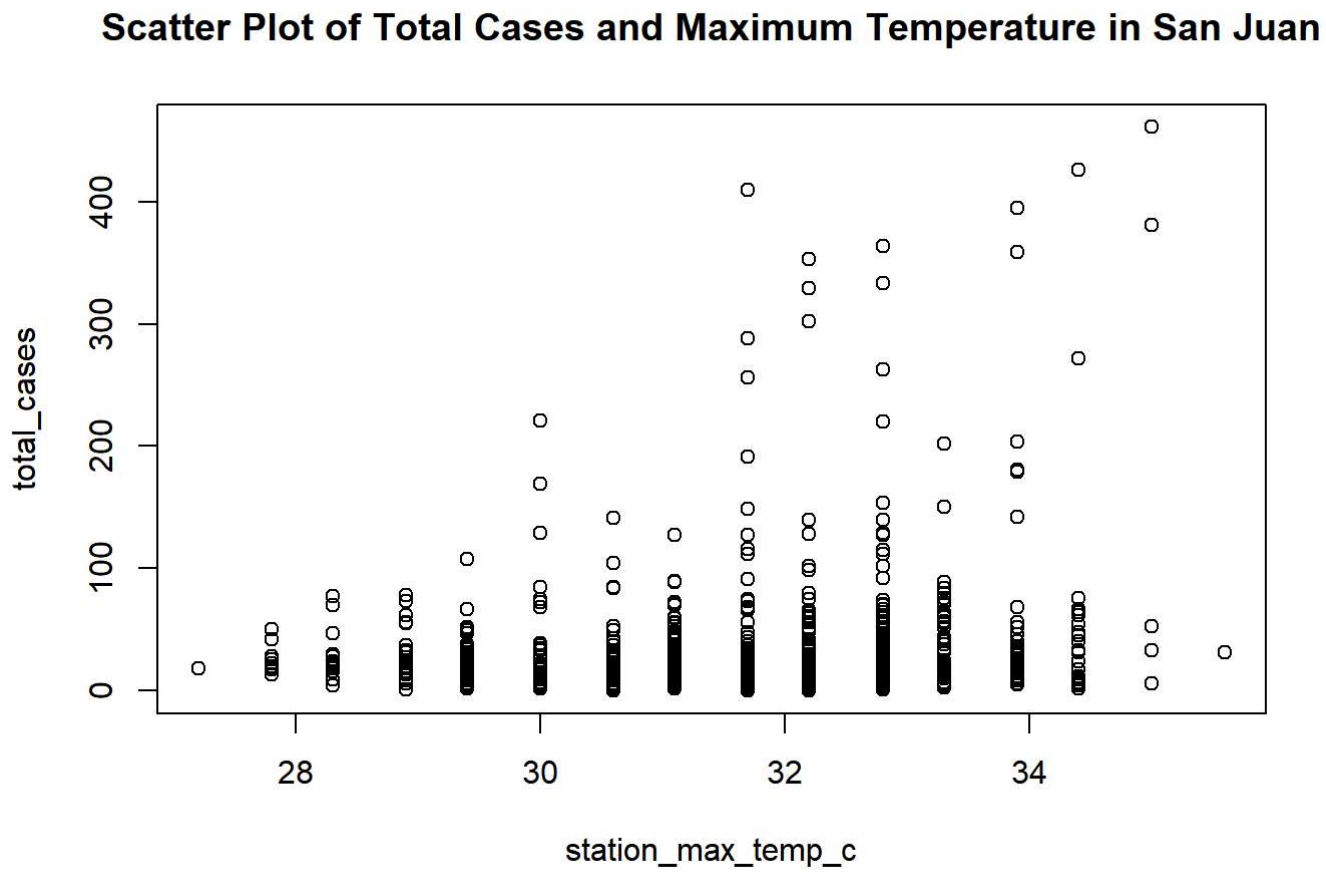
```
##    year weekofyear total_cases
## 1: 1990         28          NA
## 2: 1992         15          NA
## 3: 1993         14          NA
## 4: 1995         20          NA
## 5: 1996         22          NA
## 6: 1997         19          NA
```

```
# Plot total cases with timeline -> seasonal effect
ggplot(sjdata.train , aes(week_start_date, total_cases)) +
  geom_line() +
  scale_y_continuous() +
  scale_x_date(breaks = date_breaks("year"), labels = date_format("%m-%Y")) +
  ylab("Total cases in San Juan") +
  xlab("")+
  theme(axis.text.x=element_text(angle=-70, hjust=0.001)) +
  labs(title = "Time Series Plot of Total Cases") +
  theme_bw()
```

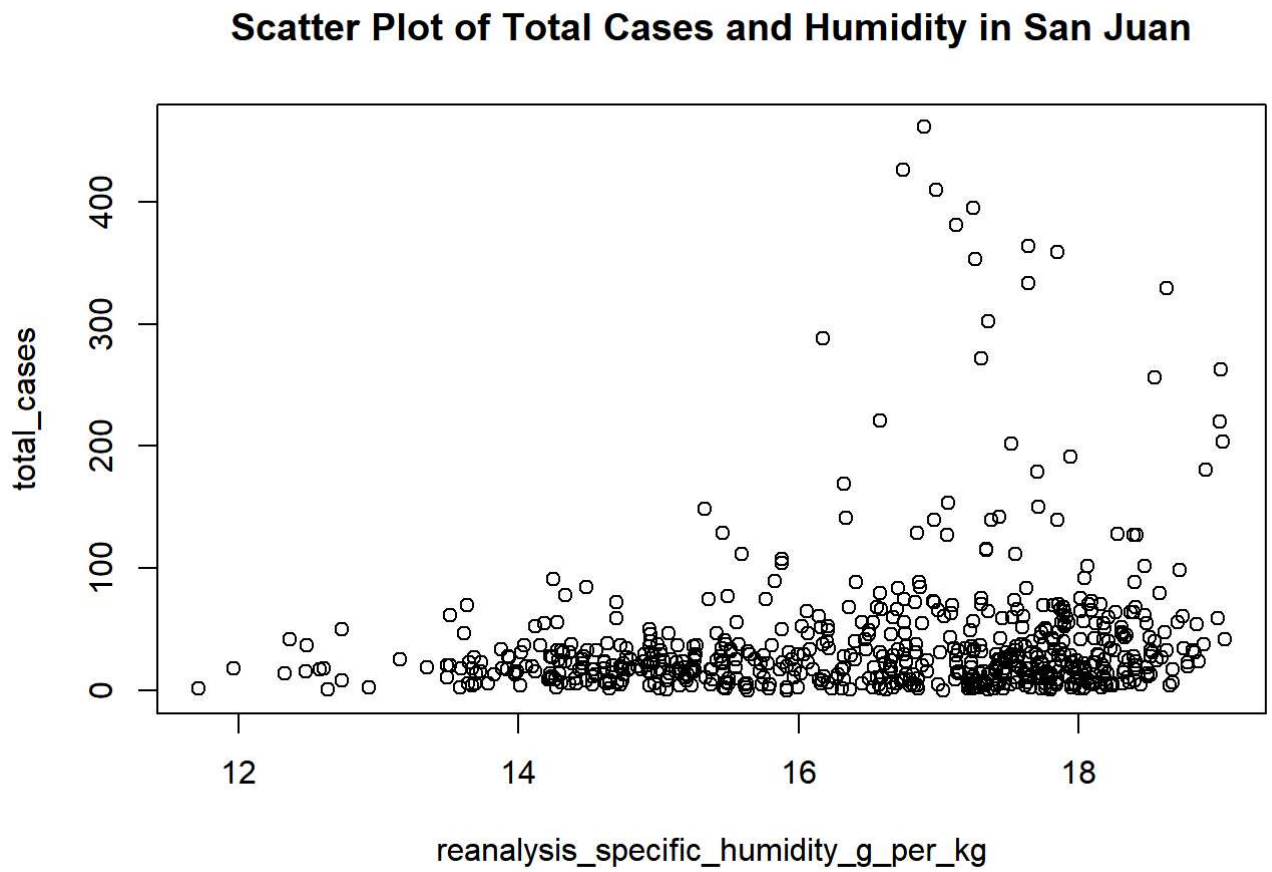



```
ggsave("./Charts/time_series_totalcases.png", width = 8, height = 6)
```

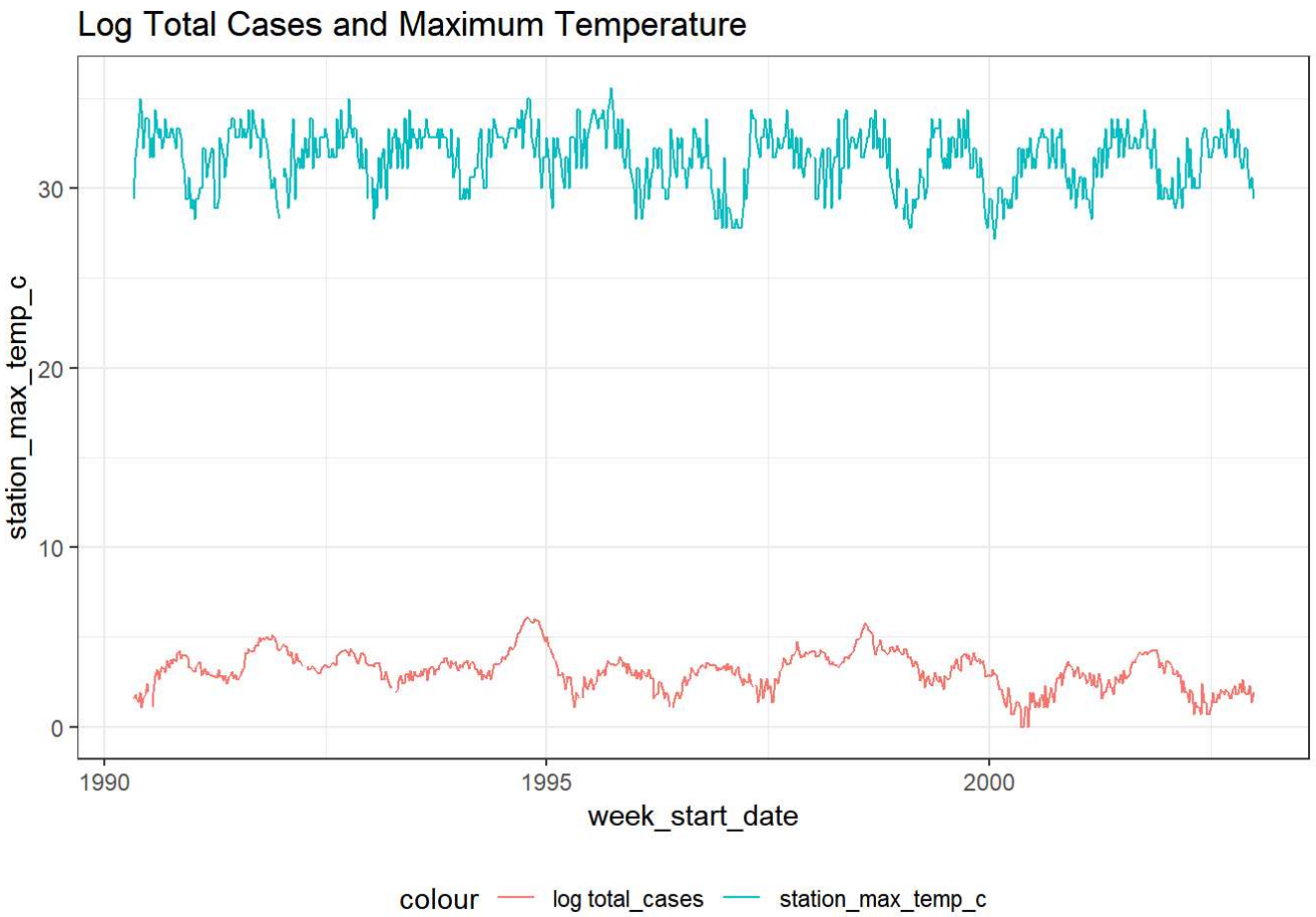
```
# Scatter Plot Total cases vs. Max Temperature
plot(total_cases ~ station_max_temp_c, data=sjdata.train
     ,main = "Scatter Plot of Total Cases and Maximum Temperature in San Juan")
```




```
# Scatter Plot Total cases vs. Humidity
plot(total_cases ~ reanalysis_specific_humidity_g_per_kg, data=sjdata.train,
     main = "Scatter Plot of Total Cases and Humidity in San Juan")
```

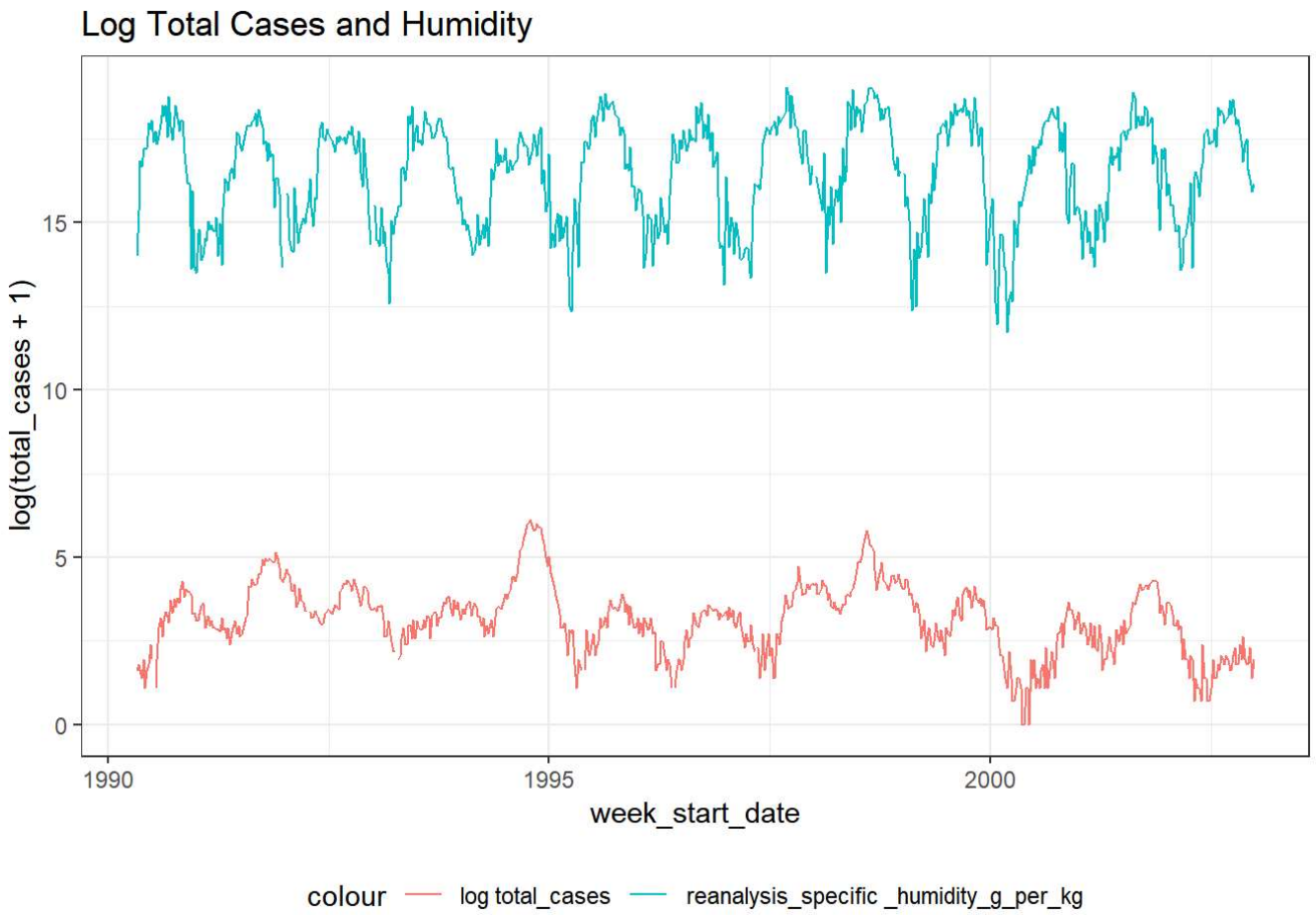



```
# Time Series Plot of log(Total cases) & Max Temperature
ggplot(sjdata.train, aes(week_start_date)) +
  geom_line(aes(y = station_max_temp_c, colour = "station_max_temp_c")) +
  geom_line(aes(y = (y = log(total_cases+1)),
                colour = "log total_cases")) +
  labs(title = "Log Total Cases and Maximum Temperature ") +
  theme_bw() +theme(legend.position="bottom")
```




```
ggsave("./Charts/time_series_logtotalcases_maxtemp.png", width = 8, height = 6)
```

```
# Time Series Plot of log(Total cases) & Humidity
ggplot(sjdata.train, aes(week_start_date)) +
  geom_line(aes(y = log(total_cases+1),
                colour = "log total_cases")) +
  geom_line(aes(y = reanalysis_specific_humidity_g_per_kg,
                colour = "reanalysis_specific _humidity_g_per_kg")) +
  labs(title = "Log Total Cases and Humidity") +
  theme_bw() + theme(legend.position="bottom")
```




```
ggsave("./Charts/time_series_logtotalcases_humidity.png", width = 8, height = 6)
```

Data Imputation

```
# Build a function to impute data with the most recent that is non missing (Using LOCF)
na.locf.data.frame <-
  function(object, ...) replace(object, TRUE, lapply(object, na.locf, ...))

# Fill in NAs
sjdata.train.imputed <- na.locf.data.frame(sjdata.train)
summary(sjdata.train.imputed)
```

```
##  city          year        weekofyear    week_start_date
##  sj:659   Min.   :1990   Min.   : 1.00   Min.   :1990-04-30
##           1st Qu.:1993   1st Qu.:14.00   1st Qu.:1993-06-28
##           Median :1996   Median :27.00   Median :1996-08-26
##           Mean   :1996   Mean   :26.95   Mean   :1996-08-26
##           3rd Qu.:1999   3rd Qu.:40.00   3rd Qu.:1999-10-25
##           Max.   :2002   Max.   :53.00   Max.   :2002-12-24
##     ndvi_ne            ndvi_nw            ndvi_se          ndvi_sw
##  Min.   :-0.29020   Min.   :-0.25280   Min.   :0.0360   Min.   :-0.06346
##  1st Qu.: 0.03343   1st Qu.: 0.04863   1st Qu.:0.1405   1st Qu.: 0.13411
##  Median : 0.07740   Median : 0.08460   Median :0.1787   Median : 0.16953
##  Mean   : 0.08379   Mean   : 0.09341   Mean   :0.1792   Mean   : 0.16885
##  3rd Qu.: 0.12387   3rd Qu.: 0.13375   3rd Qu.:0.2127   3rd Qu.: 0.20316
##  Max.   : 0.44627   Max.   : 0.43710   Max.   :0.3931   Max.   : 0.38142
##  precipitation_amt_mm reanalysis_air_temp_k reanalysis_avg_temp_k
##  Min.   :  0.000      Min.   :295.9         Min.   :296.1
##  1st Qu.:  0.245      1st Qu.:298.0         1st Qu.:298.2
##  Median : 20.800      Median :299.2         Median :299.3
##  Mean   : 33.962      Mean   :299.0         Mean   :299.1
##  3rd Qu.: 50.025      3rd Qu.:300.0         3rd Qu.:300.1
##  Max.   :287.550      Max.   :301.3         Max.   :301.4
##  reanalysis_dew_point_temp_k reanalysis_max_air_temp_k
##  Min.   :289.6               Min.   :298.2
##  1st Qu.:293.8               1st Qu.:300.4
##  Median :295.4               Median :301.4
##  Mean   :295.1               Mean   :301.3
##  3rd Qu.:296.4               3rd Qu.:302.2
##  Max.   :297.5               Max.   :303.9
##  reanalysis_min_air_temp_k reanalysis_precip_amt_kg_per_m2
##  Min.   :292.6             Min.   :  0.00
##  1st Qu.:296.2             1st Qu.: 11.85
##  Median :297.5             Median : 22.10
##  Mean   :297.2             Mean   : 31.92
##  3rd Qu.:298.2             3rd Qu.: 40.30
##  Max.   :299.5             Max.   :570.50
##  reanalysis_relative_humidity_percent
##  Min.   :66.74
##  1st Qu.:76.80
##  Median :79.15
```

```
##  Mean   :78.99
##  3rd Qu.:81.41
##  Max.   :87.30
##  reanalysis_specific_humidity_g_per_kg reanalysis_tdtr_k
##  Min.   :11.72                         Min.   :1.357
##  1st Qu.:15.23                         1st Qu.:2.086
##  Median :16.82                         Median :2.357
##  Mean   :16.50                         Mean   :2.436
##  3rd Qu.:17.80                         3rd Qu.:2.686
##  Max.   :19.04                         Max.   :4.100
##  station_avg_temp_c station_diur_temp_rng_c station_max_temp_c
##  Min.   :22.84      Min.   :4.529           Min.   :27.20
##  1st Qu.:25.92      1st Qu.:6.357           1st Qu.:30.60
##  Median :27.27      Median :6.843           Median :32.20
##  Mean   :27.04      Mean   :6.863           Mean   :31.71
##  3rd Qu.:28.19      3rd Qu.:7.386           3rd Qu.:32.80
##  Max.   :30.07      Max.   :9.914           Max.   :35.60
##  station_min_temp_c station_precip_mm    sep.x            total_cases
##  Min.   :17.80      Min.   :  0.00    Length:659         Min.   :  0.00
##  1st Qu.:21.70      1st Qu.:  6.60    Class :character   1st Qu.: 11.00
##  Median :22.80      Median : 16.60    Mode  :character   Median : 23.00
##  Mean   :22.58      Mean   : 25.09                       Mean   : 39.46
##  3rd Qu.:23.90      3rd Qu.: 32.40                       3rd Qu.: 44.00
##  Max.   :25.60      Max.   :305.90                       Max.   :461.00
##     sep.y
##  Length:659
##  Class :character
##  Mode  :character
##
##
##
```

```
sjdata.train <- sjdata.train.imputed
```

Diagnostic study

```
# Convert data to TS with weekly frequency
sjtrain.ts <- ts(sjdata.train$total_cases, frequency = 52, start = c(1990,18))
head(sjtrain.ts)
```

```
## Time Series:
## Start = c(1990, 18)
## End = c(1990, 23)
## Frequency = 52
## [1] 4 5 4 3 6 2
```

Stationarity test

```
# ADF Test
adf.test(sjdata.train$total_cases)
```

```
## Warning in adf.test(sjdata.train$total_cases): p-value smaller than printed
## p-value
```

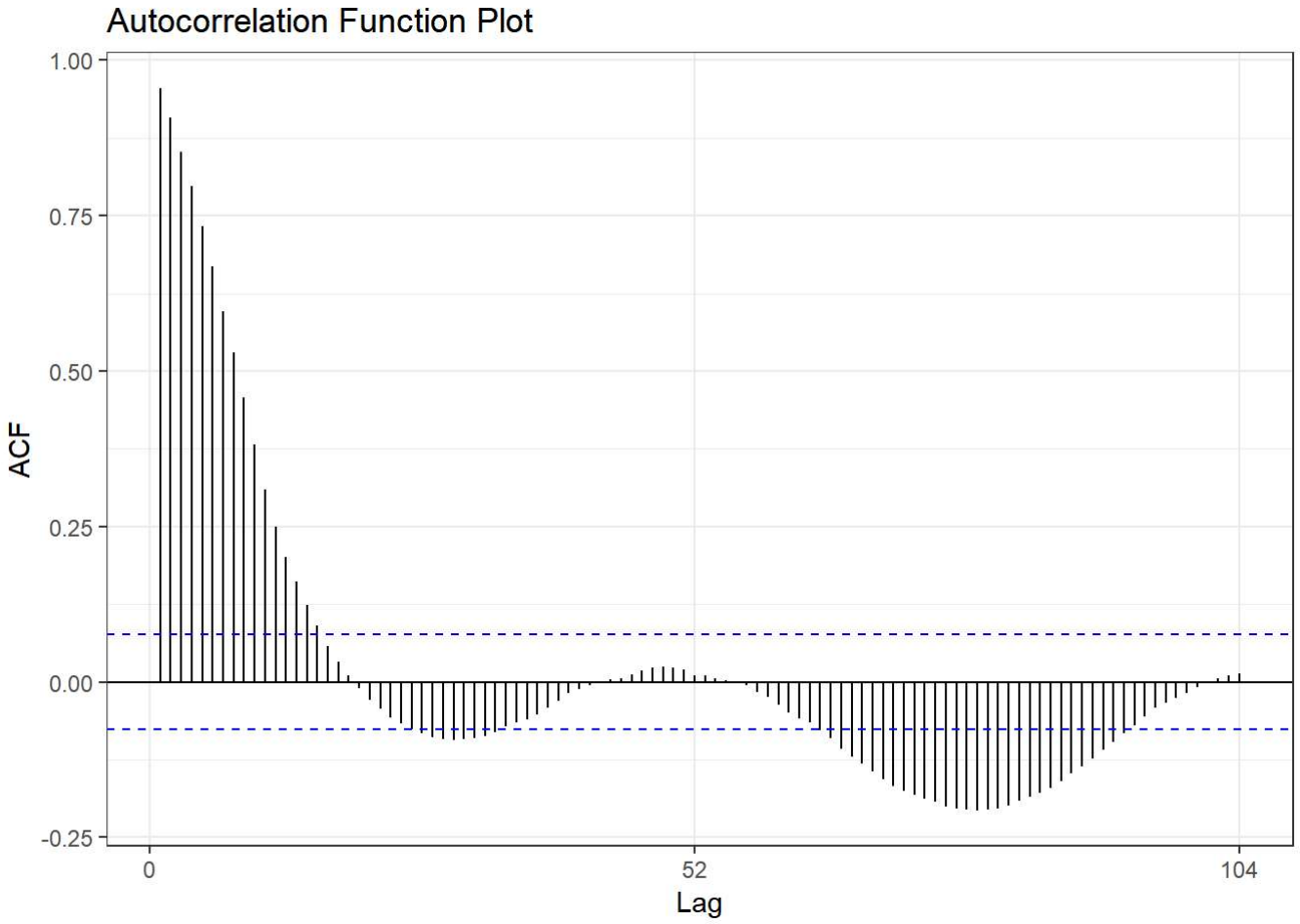
```
##
##  Augmented Dickey-Fuller Test
##
## data:  sjdata.train$total_cases
## Dickey-Fuller = -5.8906, Lag order = 8, p-value = 0.01
## alternative hypothesis: stationary
```

```
# KPSS Test
kpss.test(sjdata.train$total_cases)
```

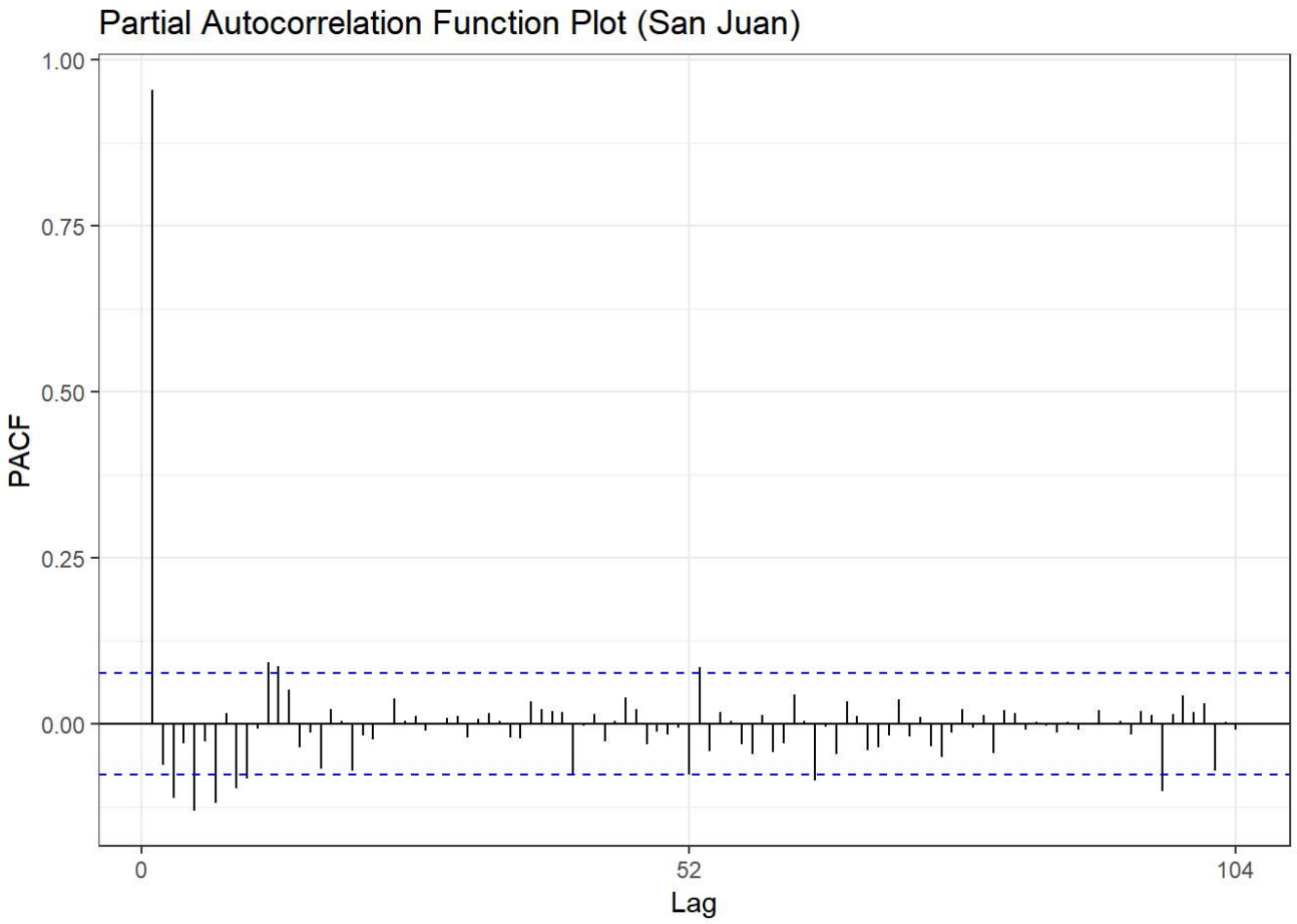
```
##
##  KPSS Test for Level Stationarity
##
## data:  sjdata.train$total_cases
## KPSS Level = 0.3684, Truncation lag parameter = 6, p-value =
## 0.09078
```

The time series is stationary.

```
# ACF Plot
ggAcf(sjtrain.ts) +
  labs(title = "Autocorrelation Function Plot") +
  theme_bw()
```



```
ggsave("./Charts/ACF.png", width = 8, height = 6)


# PACF Plot
ggPacf(sjtrain.ts)  +
  labs(title = "Partial Autocorrelation Function Plot (San Juan)") +
  theme_bw()
```

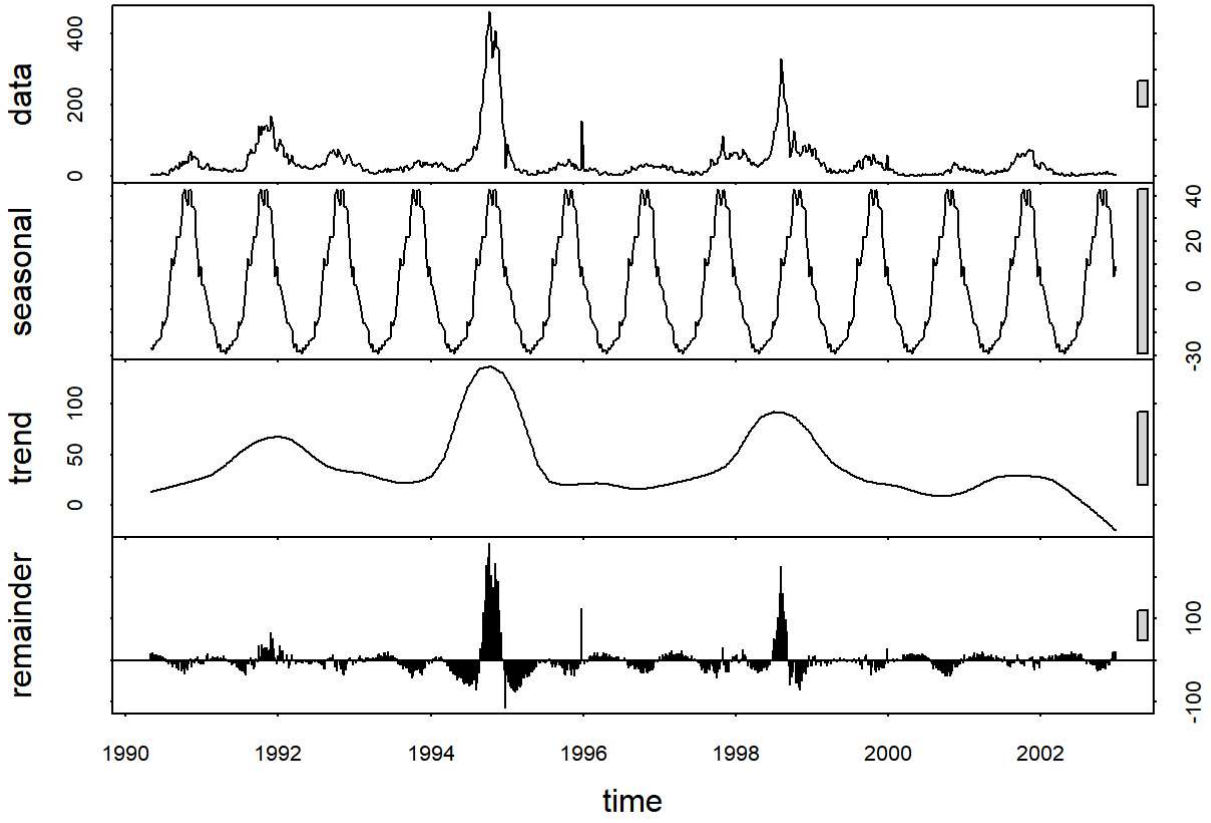



```
ggsave("./Charts/PACF.png", width = 8, height = 6)
```

We can observe significant autocorrelation and seasonality.

Decompose the time series

```
decomp = stl(sjtrain.ts, s.window="periodic")
plot(decomp)
```



Modeling

The following models were used for forecasting dengue spread.

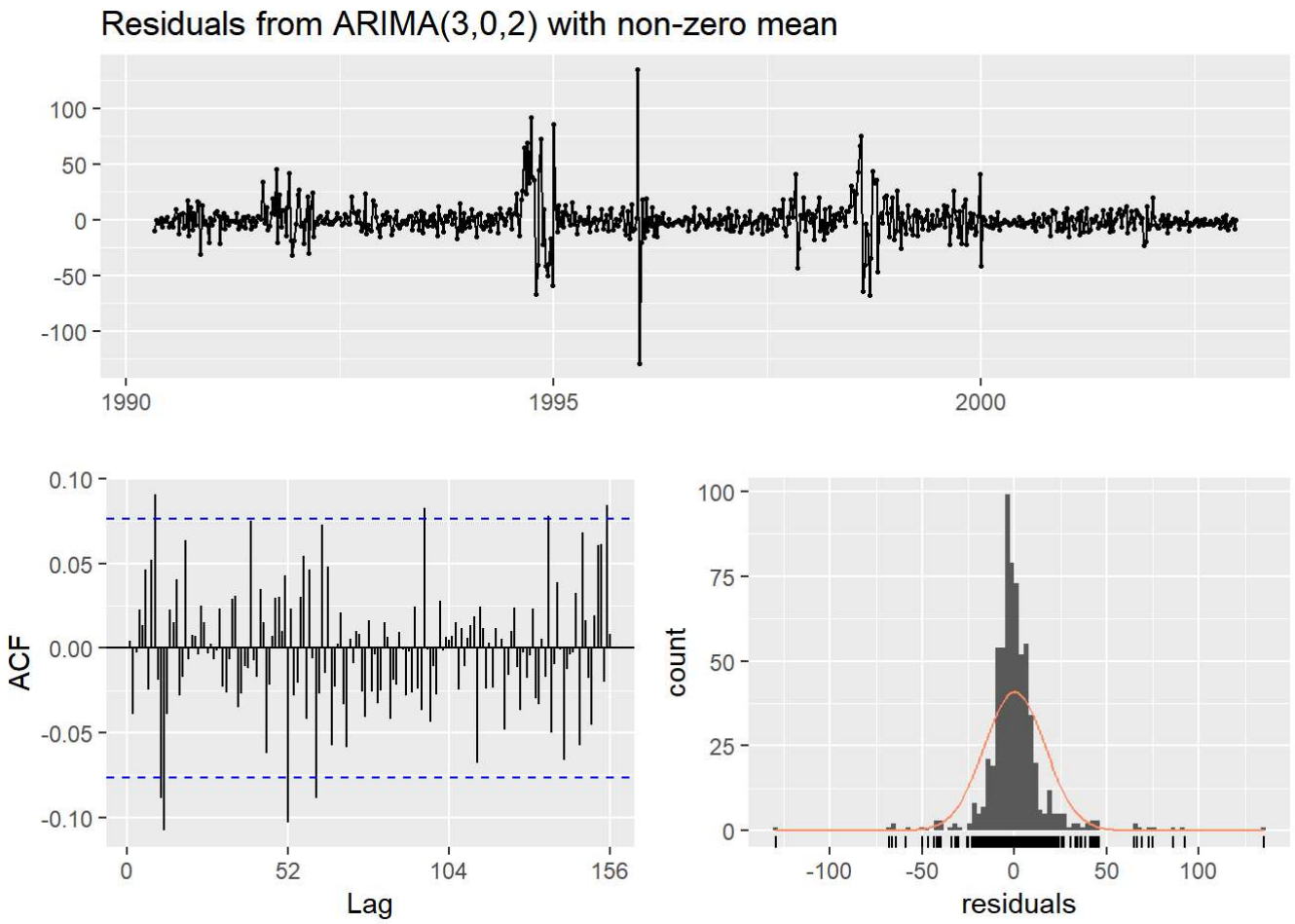
ARIMA

Use ARIMA without seasonality on the training data and analyse the residuals.

```
sj.fit1 <- auto.arima(sjtrain.ts, seasonal = F)
sj.fit1
```

```
## Series: sjtrain.ts
## ARIMA(3,0,2) with non-zero mean
##
## Coefficients:
##          ar1     ar2      ar3     ma1      ma2     mean
##       0.9208  0.8686  -0.8161  0.0720  -0.7303  38.9242
## s.e.  0.0601  0.0767   0.0543  0.0743   0.0707   8.2134
##
## sigma^2 estimated as 276:  log likelihood=-2785.33
## AIC=5584.67   AICc=5584.84   BIC=5616.1
```

```
checkresiduals(sj.fit1)
```




```
##
##  Ljung-Box test
##
## data:  Residuals from ARIMA(3,0,2) with non-zero mean
## Q* = 96.623, df = 98, p-value = 0.5204
##
## Model df: 6.   Total lags used: 104
```

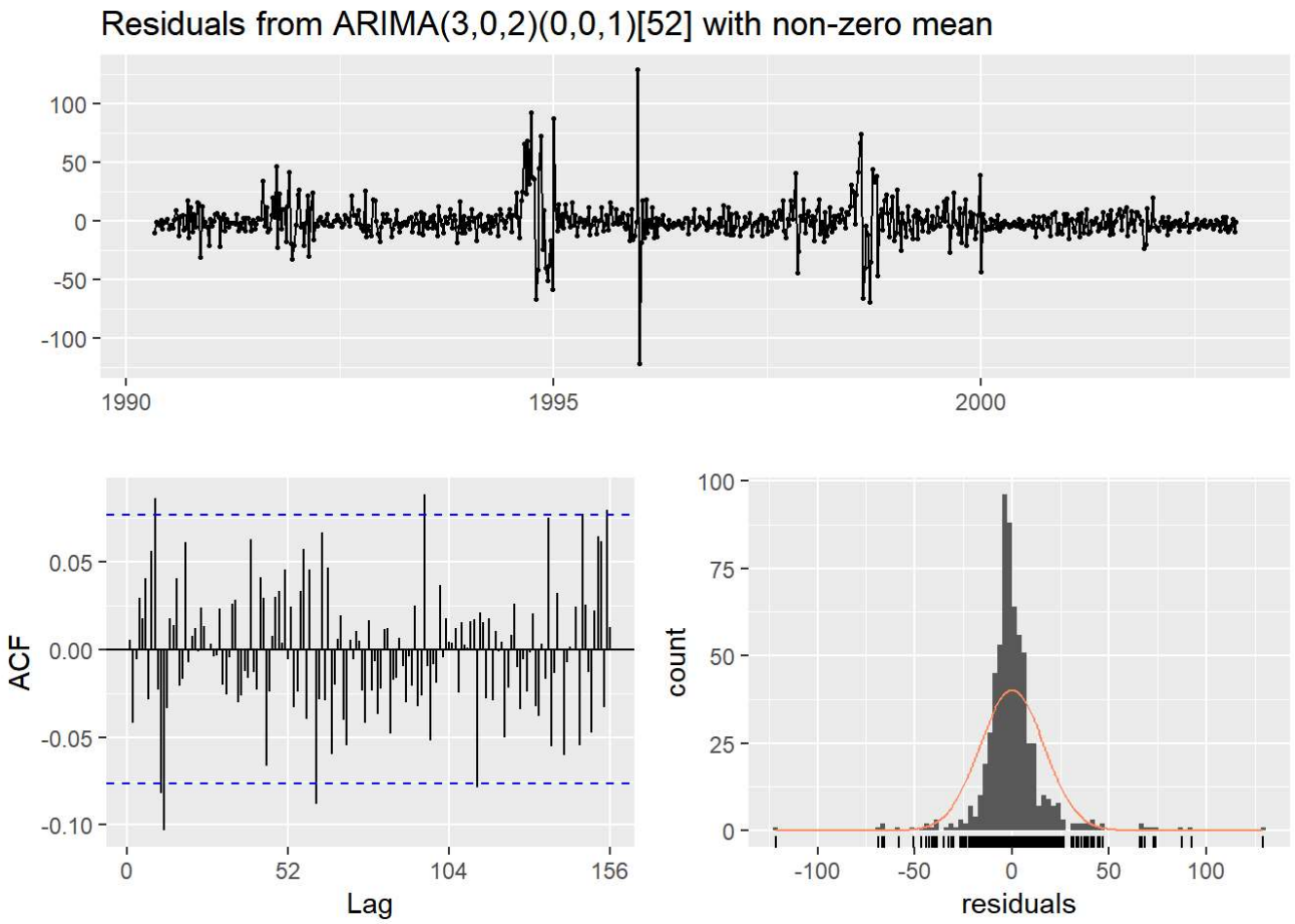
SARIMA

Apply Seasonal ARIMA to the data and analyse the residuals.

```
sj.fit2 <- auto.arima(sjtrain.ts, seasonal = T)
sj.fit2
```

```
## Series: sjtrain.ts
## ARIMA(3,0,2)(0,0,1)[52] with non-zero mean
##
## Coefficients:
##          ar1     ar2      ar3     ma1      ma2     sma1     mean
##       0.9226  0.8788  -0.8282  0.0761  -0.7418  -0.1004  39.3919
## s.e.  0.0570  0.0763   0.0522  0.0715   0.0686   0.0382   7.2312
##
## sigma^2 estimated as 273.3:  log likelihood=-2781.92
## AIC=5579.85   AICc=5580.07   BIC=5615.77
```

```
checkresiduals(sj.fit2)
```



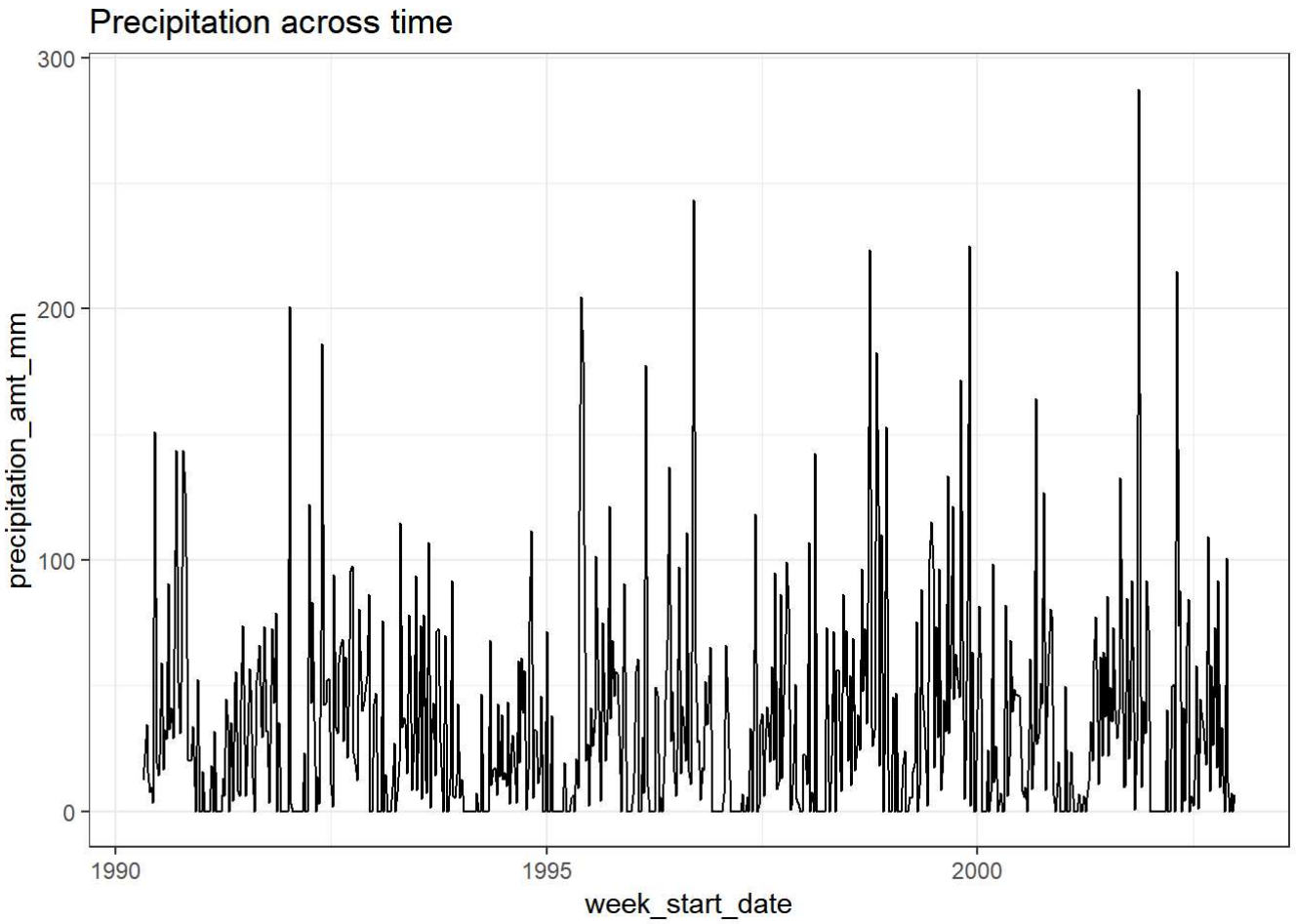

```
##
##  Ljung-Box test
##
## data:  Residuals from ARIMA(3,0,2)(0,0,1)[52] with non-zero mean
## Q* = 88.664, df = 97, p-value = 0.7151
##
## Model df: 7.   Total lags used: 104
```

SARIMAX

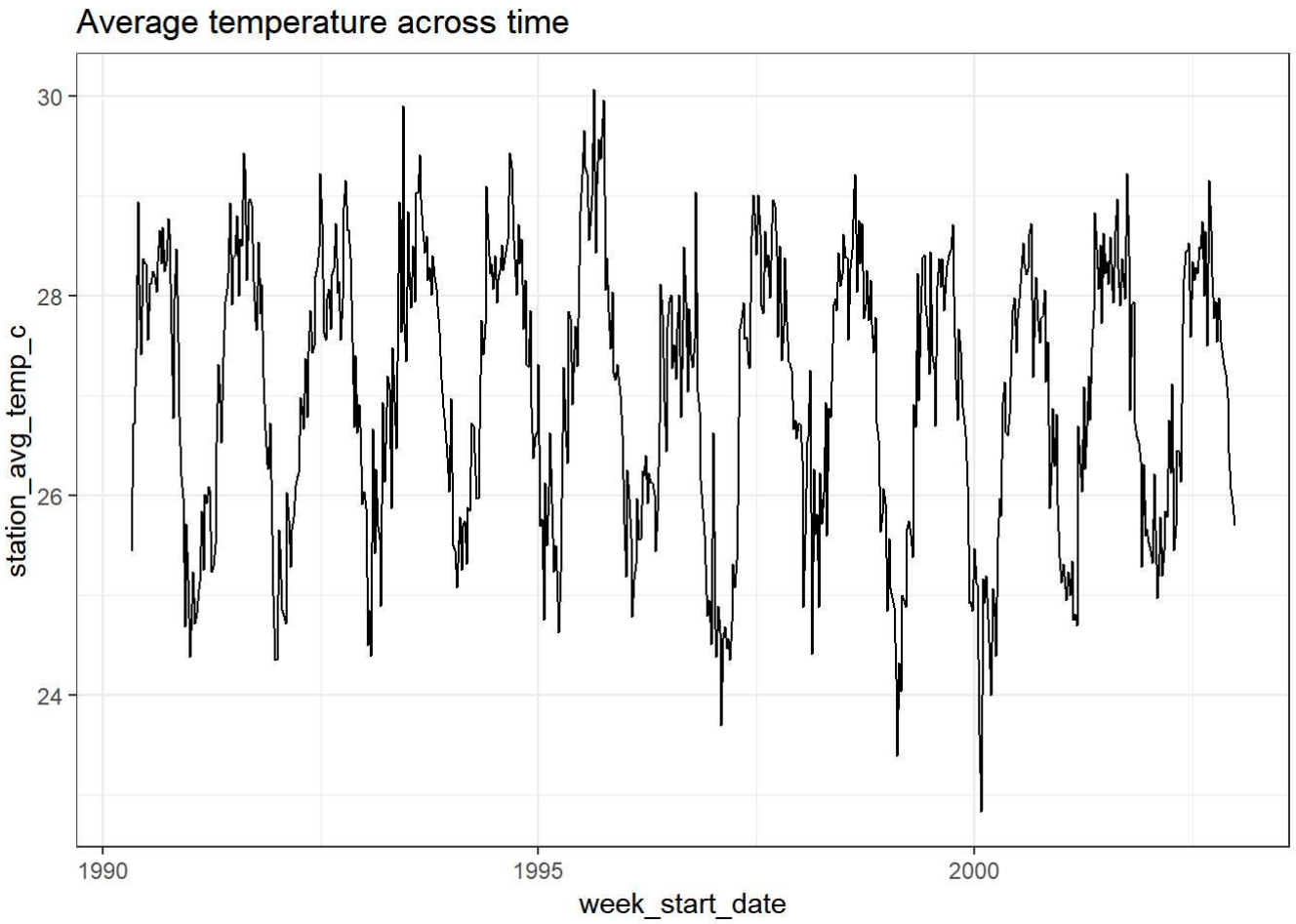
```
# Prepare xreg
sjtrain.ts1 <- ts(sjdata.train.imputed,
                  freq=365.25/7,
                  start=decimal_date(ymd("1990-04-30")))

varlist <- c("precipitation_amt_mm"
             ,"reanalysis_dew_point_temp_k"
             ,"reanalysis_relative_humidity_percent"
             ,"station_avg_temp_c"
             ,"station_diur_temp_rng_c"
             ,"station_max_temp_c"
             ,"station_min_temp_c")

create.tslag <- function(x, dataset) {
  cbind(
    Lag0 = dataset[,x],
    Lag1 = stats::lag(dataset[,x],-1),
    Lag2 = stats::lag(dataset[,x],-2),
    Lag3 = stats::lag(dataset[,x],-3),
    Lag4 = stats::lag(dataset[,x],-4)) %>%
    head(NROW(dataset))
}
```

Plot distributions of the feature vectors

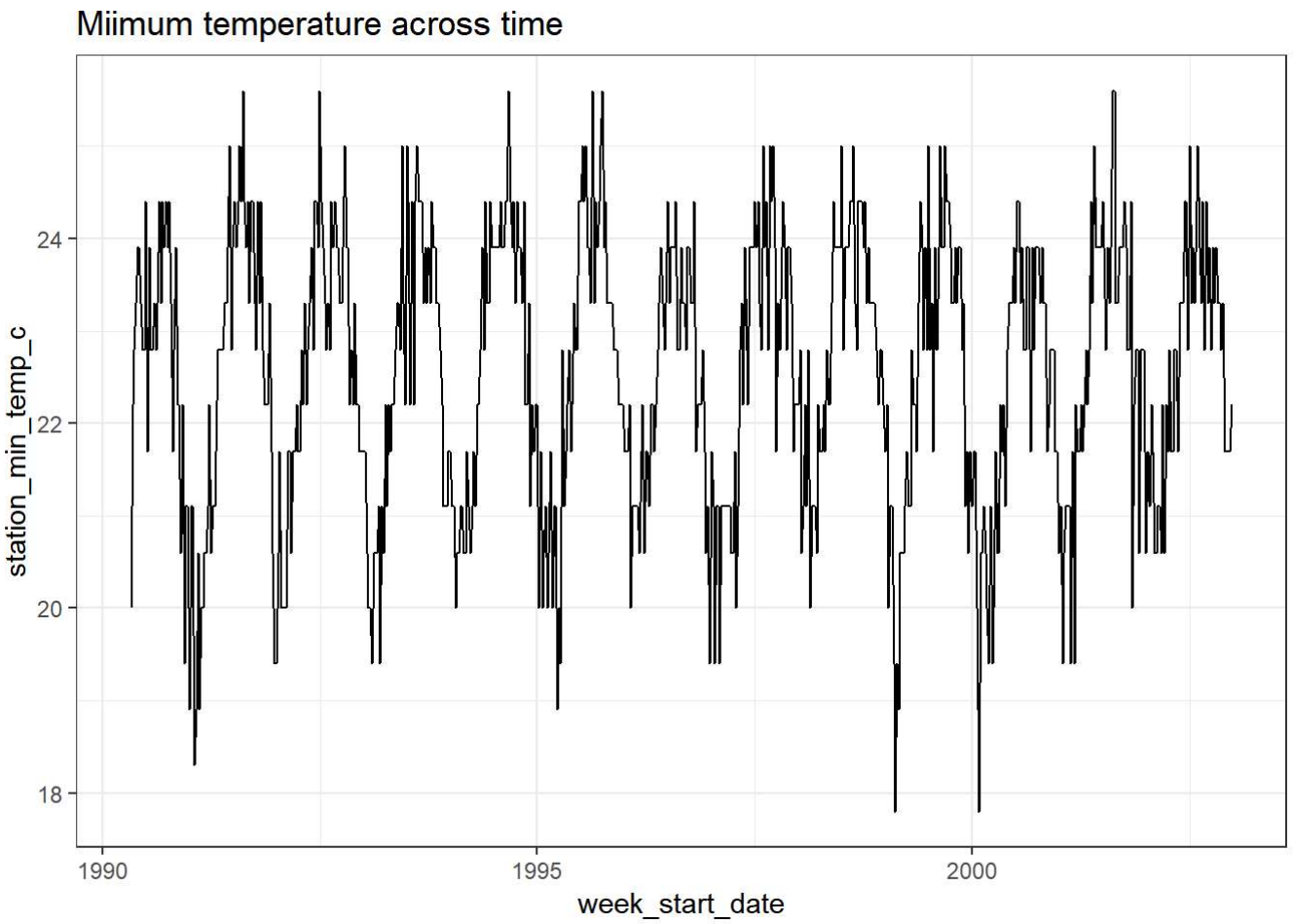
```
ggplot() +
  geom_line(mapping=aes(x=week_start_date, y=precipitation_amt_mm), data=sjdata.train.imputed) +
  theme_bw() +
  ggtitle("Precipitation across time")
```



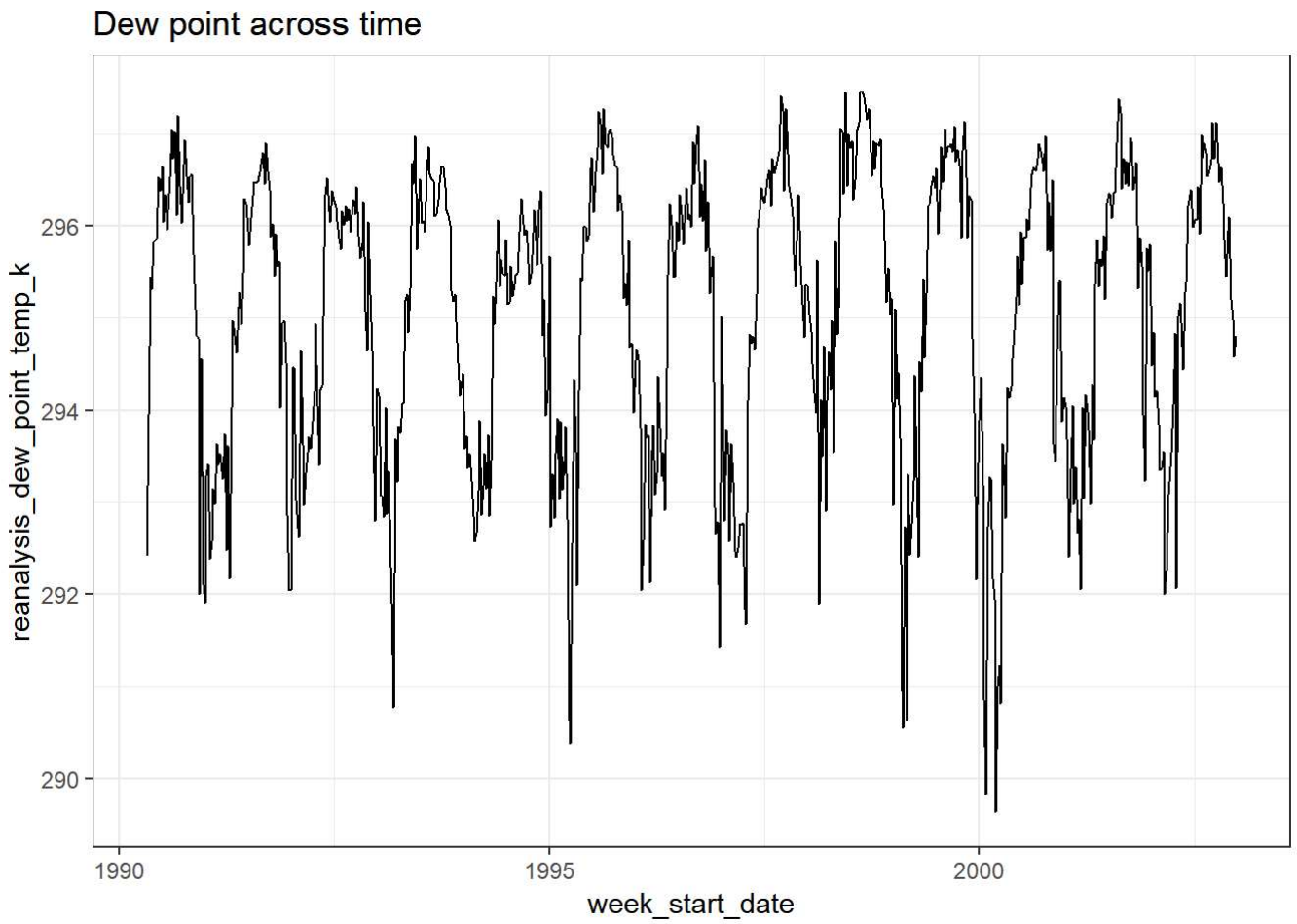

```
ggplot() +
  geom_line(mapping=aes(x=week_start_date, y=station_avg_temp_c), data=sjdata.train.imputed) +
  theme_bw() +
  ggtitle("Average temperature across time")
```



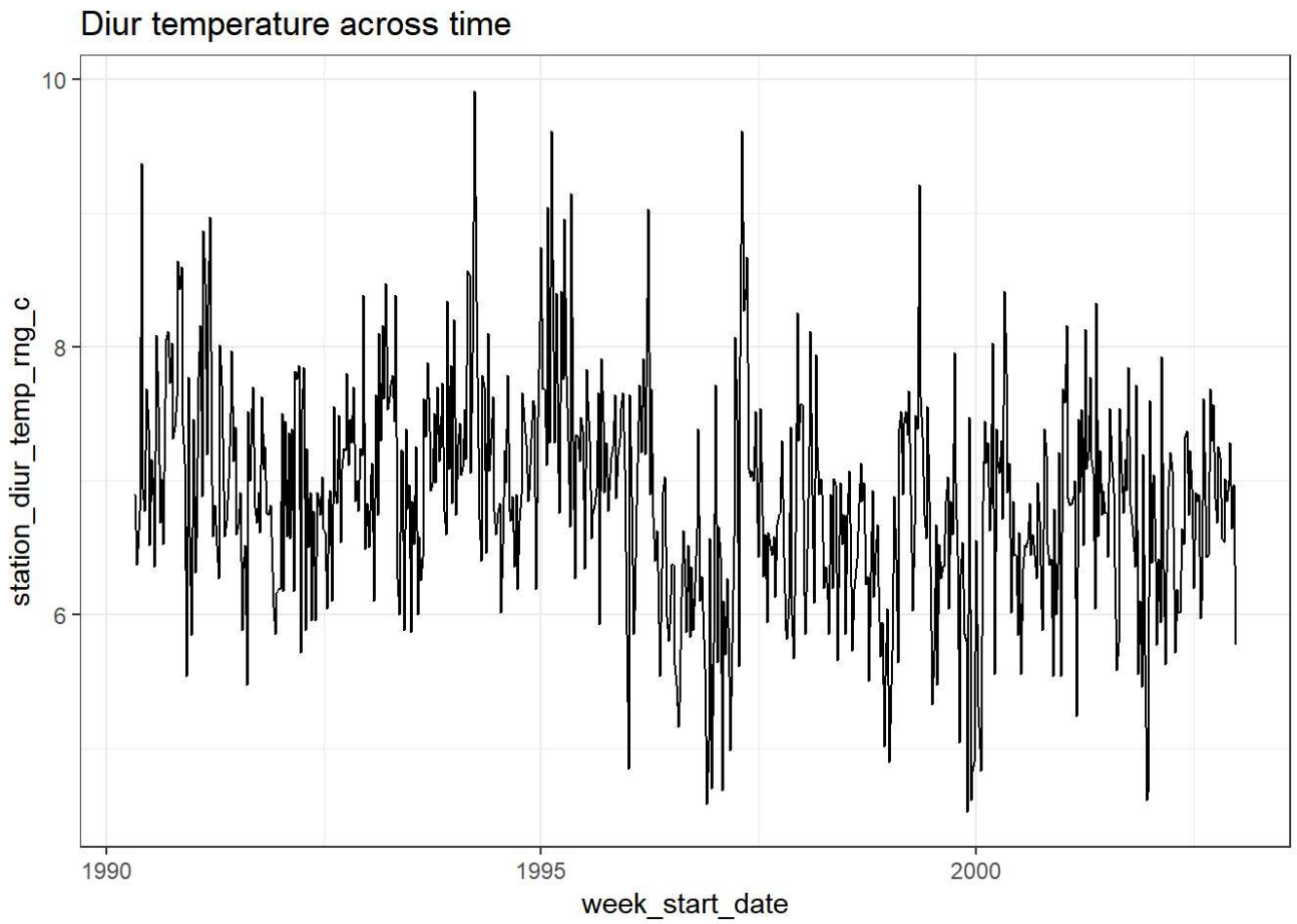

```
ggplot() +
  geom_line(mapping=aes(x=week_start_date, y=station_min_temp_c), data=sjdata.train.imputed) +
  theme_bw() +
  ggtitle("Miimum temperature across time")
```



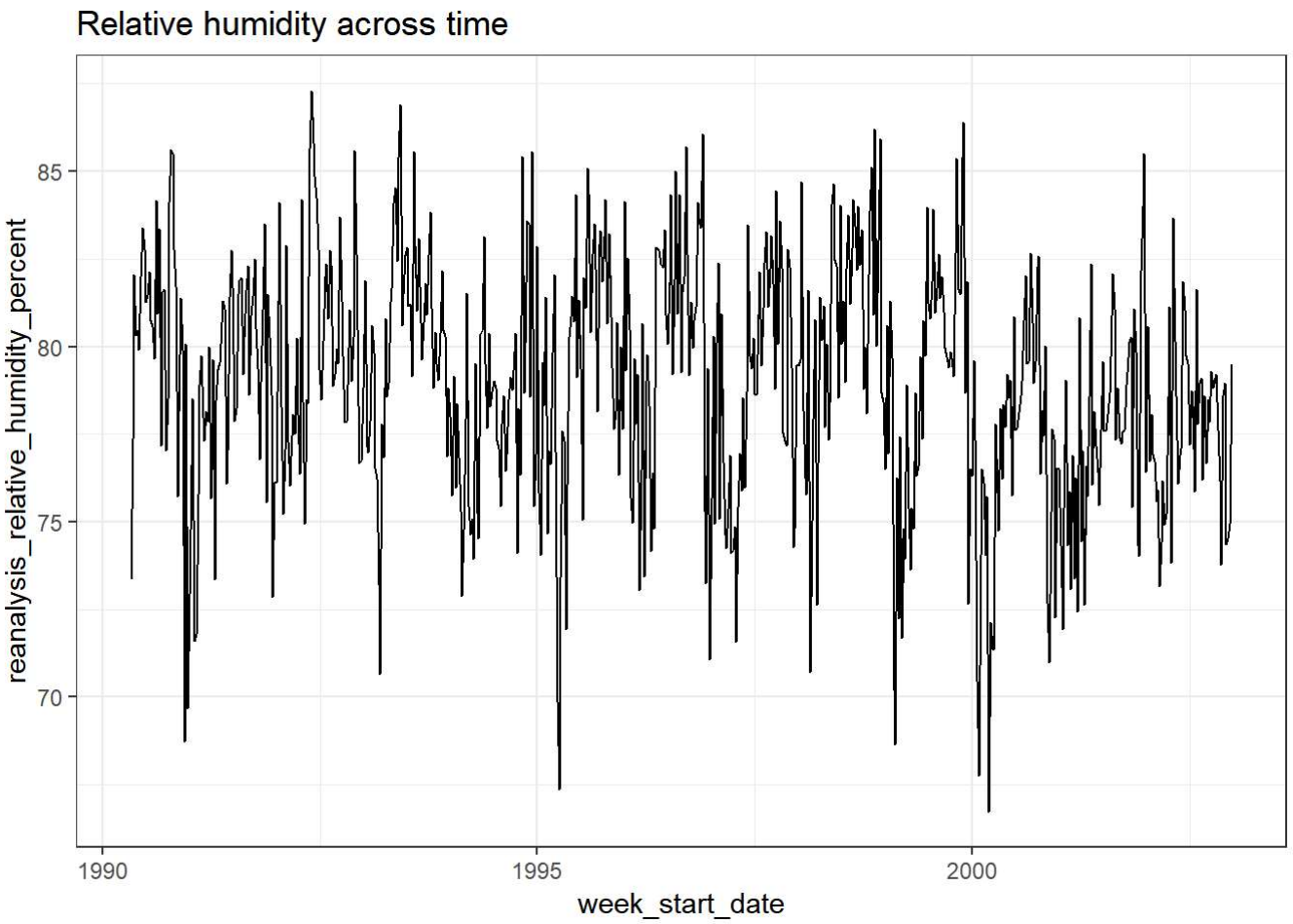

```
ggplot() +
  geom_line(mapping=aes(x=week_start_date, y=reanalysis_dew_point_temp_k), data=sjdata.train.imputed) +
  theme_bw() +
  ggtitle("Dew point across time")
```



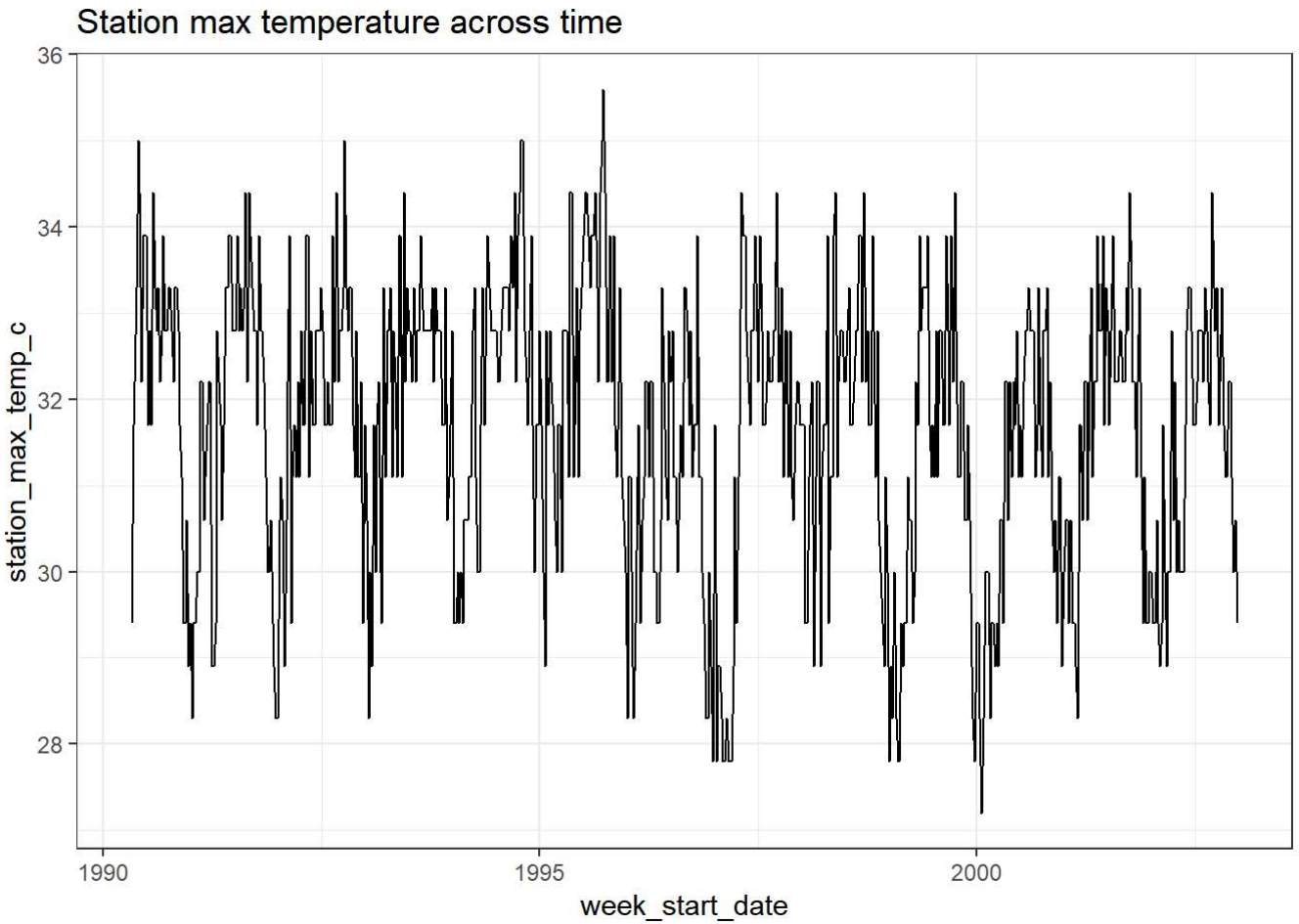

```
ggplot() +
  geom_line(mapping=aes(x=week_start_date, y=station_diur_temp_rng_c), data=sjdata.train.imputed) +
  theme_bw() +
  ggtitle("Diur temperature across time")
```




```
ggplot() +
  geom_line(mapping=aes(x=week_start_date, y=reanalysis_relative_humidity_percent), data=sjdata.train.imputed) +
  theme_bw() +
  ggtitle("Relative humidity across time")
```




```
ggplot() +
  geom_line(mapping=aes(x=week_start_date, y=station_max_temp_c), data=sjdata.train.imputed) +
  theme_bw() +
  ggtitle("Station max temperature across time")
```


Station max temperature across time
station_max_temp_c
36
34
32
30
28
1990
1995
2000
week_start_date

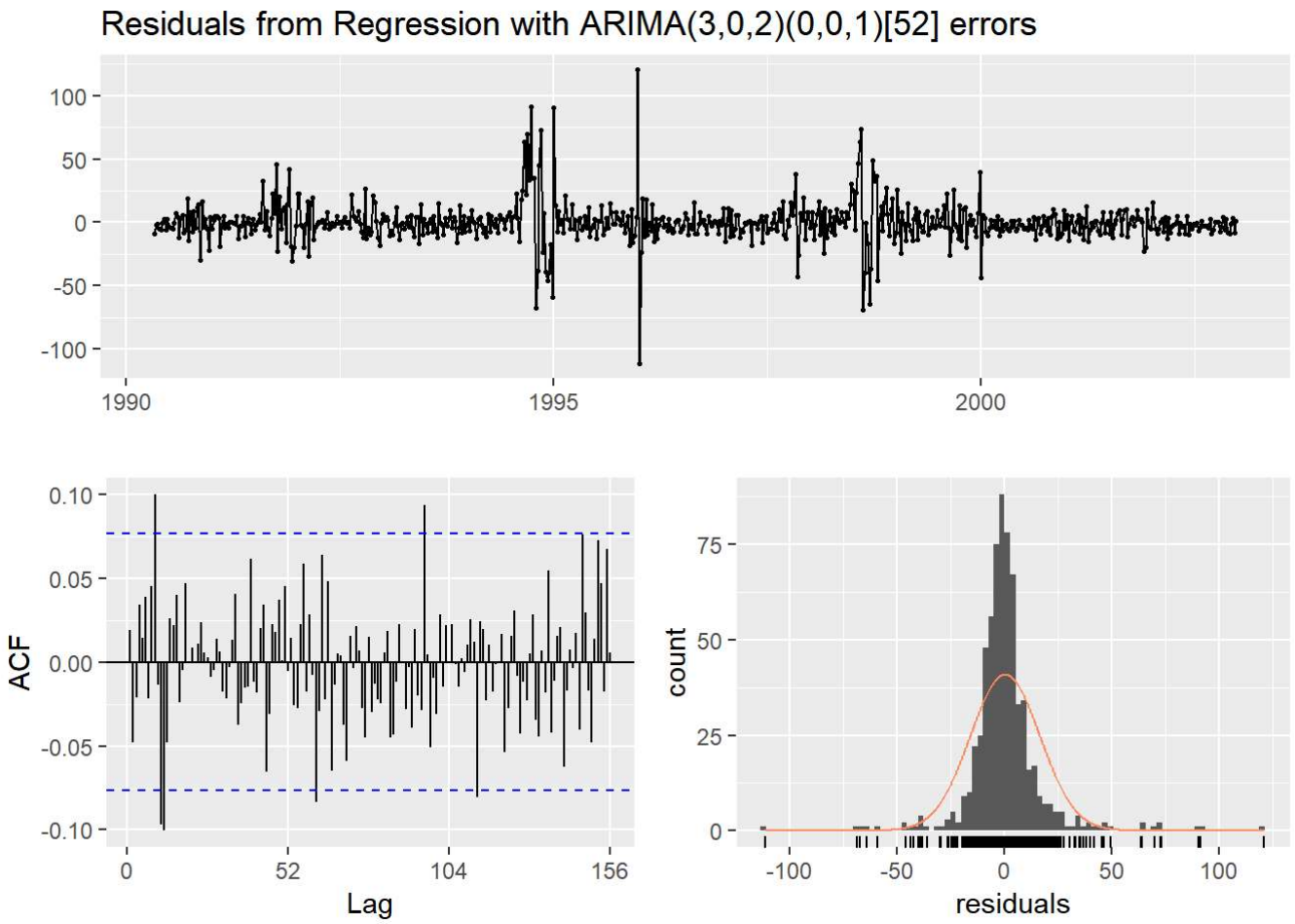
```
precipitation <- create.tslag(x = "precipitation_amt_mm"
                              ,sjtrain.ts1)

dew.temp <- create.tslag(x = "reanalysis_dew_point_temp_k"
                         ,sjtrain.ts1)

relative.humidity <- create.tslag(x = "reanalysis_relative_humidity_percent"
                                  ,sjtrain.ts1)

avg.temp <- create.tslag(x = "station_avg_temp_c"
                         ,sjtrain.ts1)

diur.temp <- create.tslag(x = "station_diur_temp_rng_c"
                          ,sjtrain.ts1)

max.temp <- create.tslag(x = "station_max_temp_c"
                         ,sjtrain.ts1)

min.temp <- create.tslag(x = "station_min_temp_c"
                         ,sjtrain.ts1)
```

Apply Seasonal ARIMAX to the data and analyse the residuals.

```
sj.fit3 <- auto.arima(sjtrain.ts, xreg=cbind(precipitation[,1],relative.humidity[,1]
                                             ,dew.temp[,1],avg.temp[,1]
                                             ,diur.temp[,1],max.temp[,1]
                                             ,min.temp[,1]), seasonal=TRUE)
sj.fit3
```

```
## Series: sjtrain.ts
## Regression with ARIMA(3,0,2)(0,0,1)[52] errors
##
## Coefficients:
##          ar1     ar2      ar3     ma1      ma2     sma1
##       0.9109  0.8935  -0.8323  0.0909  -0.7407  -0.1005
## s.e.  0.0529  0.0599   0.0487  0.0676   0.0655   0.0388
##       precipitation[, 1]  relative.humidity[, 1]  dew.temp[, 1]
##                  -0.0025                  0.0223        -0.1753
## s.e.              0.0137                  0.2363         0.1100
##       avg.temp[, 1]  diur.temp[, 1]  max.temp[, 1]  min.temp[, 1]
##              4.3603          0.5835         0.0055        -1.4431
## s.e.         1.6925          0.9835         0.7211         0.9264
##
## sigma^2 estimated as 268.6:  log likelihood=-2773.17
## AIC=5574.35   AICc=5575   BIC=5637.22
```

```
checkresiduals(sj.fit3)
```




```
##
##  Ljung-Box test
##
## data:  Residuals from Regression with ARIMA(3,0,2)(0,0,1)[52] errors
## Q* = 90.464, df = 91, p-value = 0.4961
##
## Model df: 13.   Total lags used: 104
```

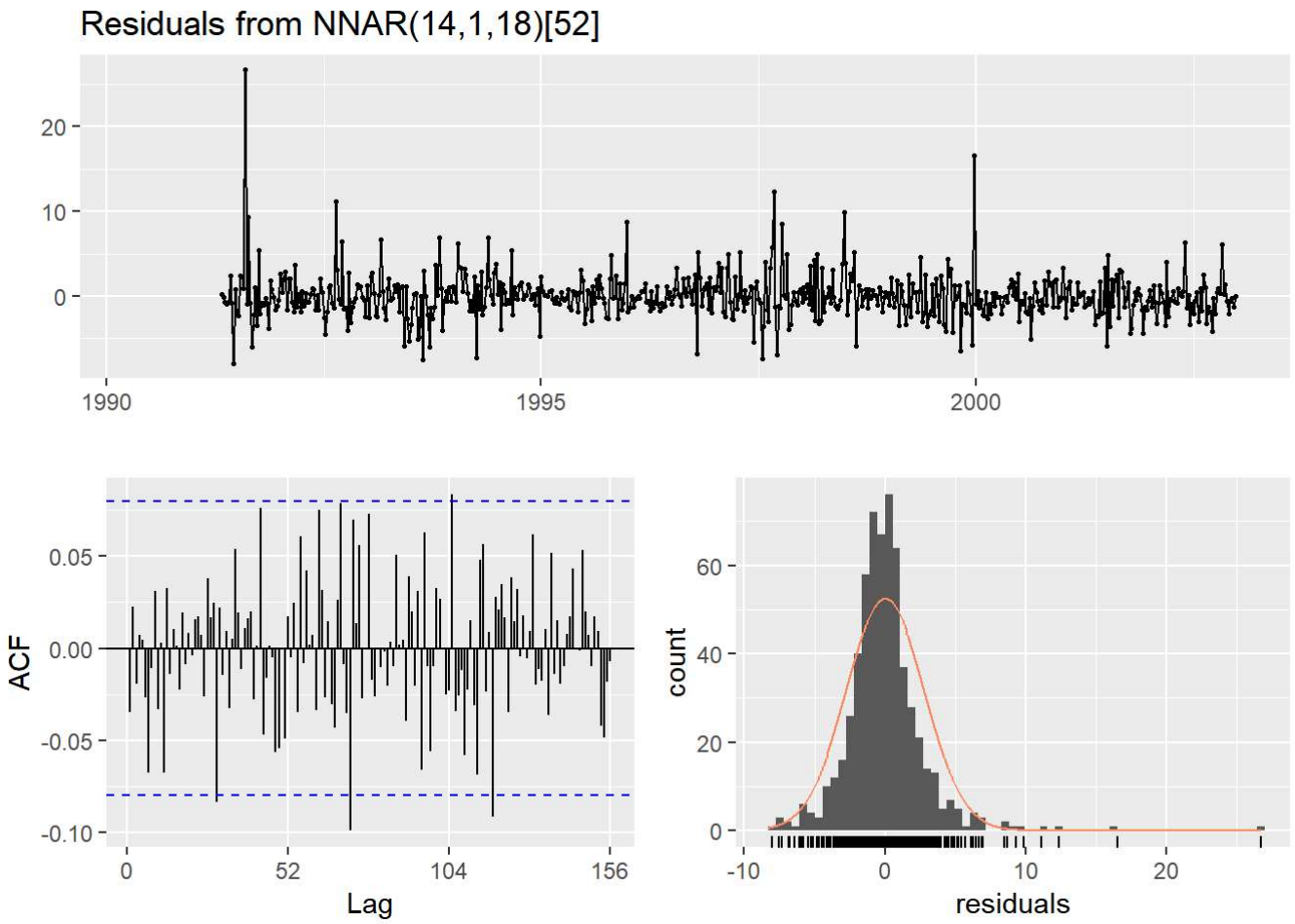
Neural Network

```
# Fit Neural Network
sj.fit4 <- nnetar(sjtrain.ts,xreg=sjtrain.ts1[,5:23])

sj.fit4
```

```
## Series: sjtrain.ts
## Model:  NNAR(14,1,18)[52]
## Call:   nnetar(y = sjtrain.ts, xreg = sjtrain.ts1[, 5:23])
##
## Average of 20 networks, each of which is
## a 34-18-1 network with 649 weights
## options were - linear output units
##
## sigma^2 estimated as 7.629
```

```
checkresiduals(sj.fit4)
```

```
## Warning in modeldf.default(object): Could not find appropriate degrees of
## freedom for this model.
```




```
Box.test(sj.fit4$residuals, type = "Ljung-Box")
```

```
##
##  Box-Ljung test
##
## data:  sj.fit4$residuals
## X-squared = 0.73509, df = 1, p-value = 0.3912
```

Model Selection

```
# Fill in NAs
sjdata.test.imputed <- na.locf.data.frame(sjdata.test)
summary(sjdata.test.imputed)
```

```
##  city          year        weekofyear    week_start_date
##  sj:277   Min.   :2003   Min.   : 1.00   Min.   :2003-01-01
##           1st Qu.:2004   1st Qu.:12.00   1st Qu.:2004-04-29
##           Median :2005   Median :25.00   Median :2005-08-27
##           Mean   :2005   Mean   :25.43   Mean   :2005-08-26
##           3rd Qu.:2006   3rd Qu.:39.00   3rd Qu.:2006-12-24
##           Max.   :2008   Max.   :53.00   Max.   :2008-04-22
##     ndvi_ne            ndvi_nw             ndvi_se
##  Min.   :-0.40625   Min.   :-0.456100   Min.   :-0.01553
##  1st Qu.:-0.05557   1st Qu.:-0.042350   1st Qu.: 0.12332
##  Median : 0.01010   Median : 0.010950   Median : 0.16833
##  Mean   :-0.00306   Mean   : 0.004001   Mean   : 0.16937
##  3rd Qu.: 0.05160   3rd Qu.: 0.053400   3rd Qu.: 0.20832
##  Max.   : 0.49340   Max.   : 0.296000   Max.   : 0.35434
##     ndvi_sw        precipitation_amt_mm reanalysis_air_temp_k
##  Min.   :0.01025   Min.   :  0.00       Min.   :296.1
##  1st Qu.:0.12143   1st Qu.:  0.00       1st Qu.:298.4
##  Median :0.15679   Median : 20.30       Median :299.5
##  Mean   :0.16033   Mean   : 38.15       Mean   :299.5
##  3rd Qu.:0.19631   3rd Qu.: 60.93       3rd Qu.:300.7
##  Max.   :0.31026   Max.   :390.60       Max.   :302.2
##  reanalysis_avg_temp_k reanalysis_dew_point_temp_k
##  Min.   :296.2         Min.   :290.5
##  1st Qu.:298.6         1st Qu.:293.9
##  Median :299.5         Median :295.5
##  Mean   :299.6         Mean   :295.2
##  3rd Qu.:300.8         3rd Qu.:296.6
##  Max.   :302.2         Max.   :297.8
##  reanalysis_max_air_temp_k reanalysis_min_air_temp_k
##  Min.   :297.8             Min.   :293.3
##  1st Qu.:300.7             1st Qu.:296.5
##  Median :301.8             Median :297.6
##  Mean   :301.7             Mean   :297.5
##  3rd Qu.:302.9             3rd Qu.:298.7
##  Max.   :304.3             Max.   :299.9
##  reanalysis_precip_amt_kg_per_m2 reanalysis_relative_humidity_percent
##  Min.   :  0.00                  Min.   :70.64
##  1st Qu.:  8.57                  1st Qu.:75.24
##  Median : 19.10                  Median :77.50
```

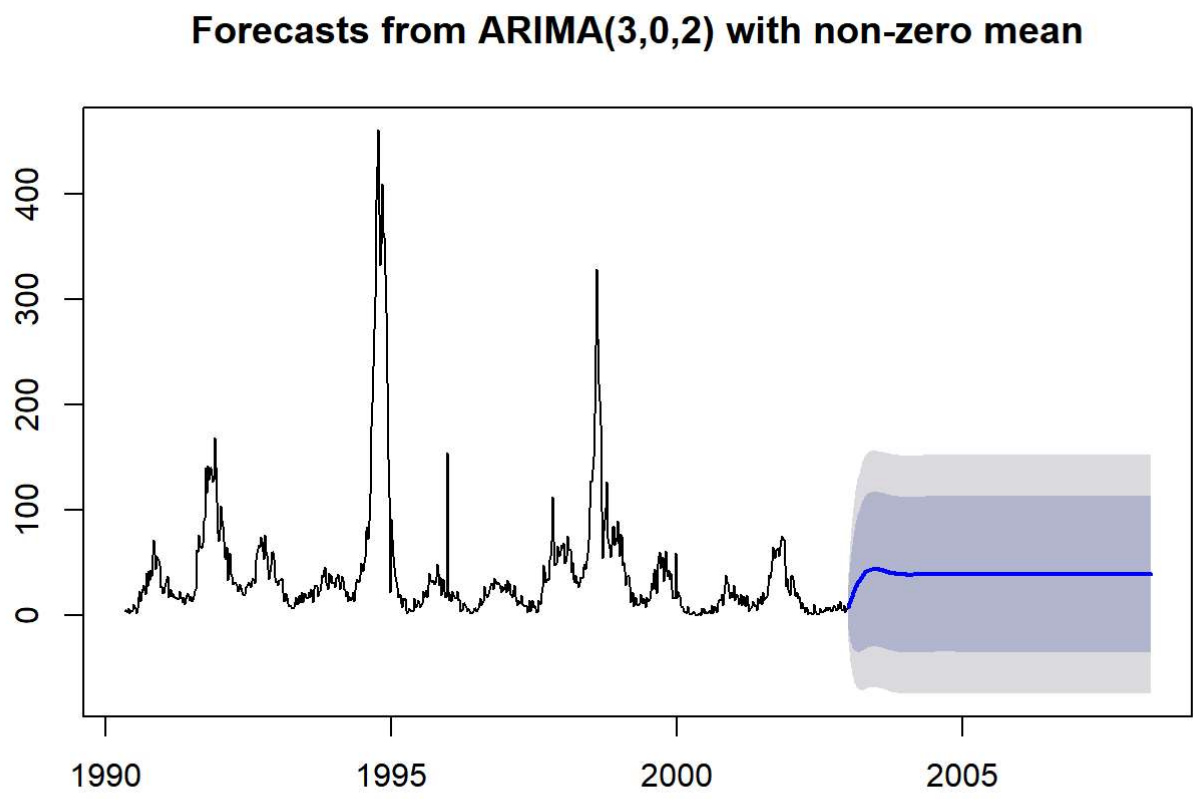
```
##  Mean   : 26.75                  Mean   :77.54
##  3rd Qu.: 32.30                  3rd Qu.:79.64
##  Max.   :254.95                  Max.   :87.58
##  reanalysis_specific_humidity_g_per_kg reanalysis_tdtr_k
##  Min.   :12.36                         Min.   :1.657
##  1st Qu.:15.26                         1st Qu.:2.386
##  Median :16.90                         Median :2.671
##  Mean   :16.65                         Mean   :2.705
##  3rd Qu.:18.05                         3rd Qu.:3.000
##  Max.   :19.44                         Max.   :4.429
##  station_avg_temp_c station_diur_temp_rng_c station_max_temp_c
##  Min.   :23.31      Min.   :4.529           Min.   :26.70
##  1st Qu.:25.56      1st Qu.:6.029           1st Qu.:30.00
##  Median :27.04      Median :6.443           Median :31.70
##  Mean   :26.89      Mean   :6.499           Mean   :31.33
##  3rd Qu.:28.17      3rd Qu.:7.043           3rd Qu.:32.80
##  Max.   :30.03      Max.   :8.671           Max.   :35.00
##  station_min_temp_c station_precip_mm    sep.x            total_cases
##  Min.   :17.80      Min.   :  0.00    Length:277         Min.   :  1.00
##  1st Qu.:21.10      1st Qu.:  7.60    Class :character   1st Qu.:  7.00
##  Median :22.80      Median : 20.90    Mode  :character   Median : 13.00
##  Mean   :22.62      Mean   : 30.68                       Mean   : 21.62
##  3rd Qu.:23.90      3rd Qu.: 41.00                       3rd Qu.: 22.00
##  Max.   :25.60      Max.   :207.50                       Max.   :170.00
##     sep.y
##  Length:277
##  Class :character
##  Mode  :character
##
##
##
```

```
# Convert data to TS
sjtest.ts1 <- ts(sjdata.test.imputed,
                 freq=52,
                 start=decimal_date(ymd("2003-01-01")))
```

```
# Take lags of predictors
precipitation1 <- create.tslag(x = "precipitation_amt_mm"
                               ,sjtest.ts1)

dew.temp1 <- create.tslag(x = "reanalysis_dew_point_temp_k"
                          ,sjtest.ts1)

relative.humidity1 <- create.tslag(x = "reanalysis_relative_humidity_percent"
                                   ,sjtest.ts1)

avg.temp1 <- create.tslag(x = "station_avg_temp_c"
                          ,sjtest.ts1)

diur.temp1 <- create.tslag(x = "station_diur_temp_rng_c"
                           ,sjtest.ts1)

max.temp1 <- create.tslag(x = "station_max_temp_c"
                          ,sjtest.ts1)

min.temp1 <- create.tslag(x = "station_min_temp_c"
                          ,sjtest.ts1)
```

Evaluate the model accuracies

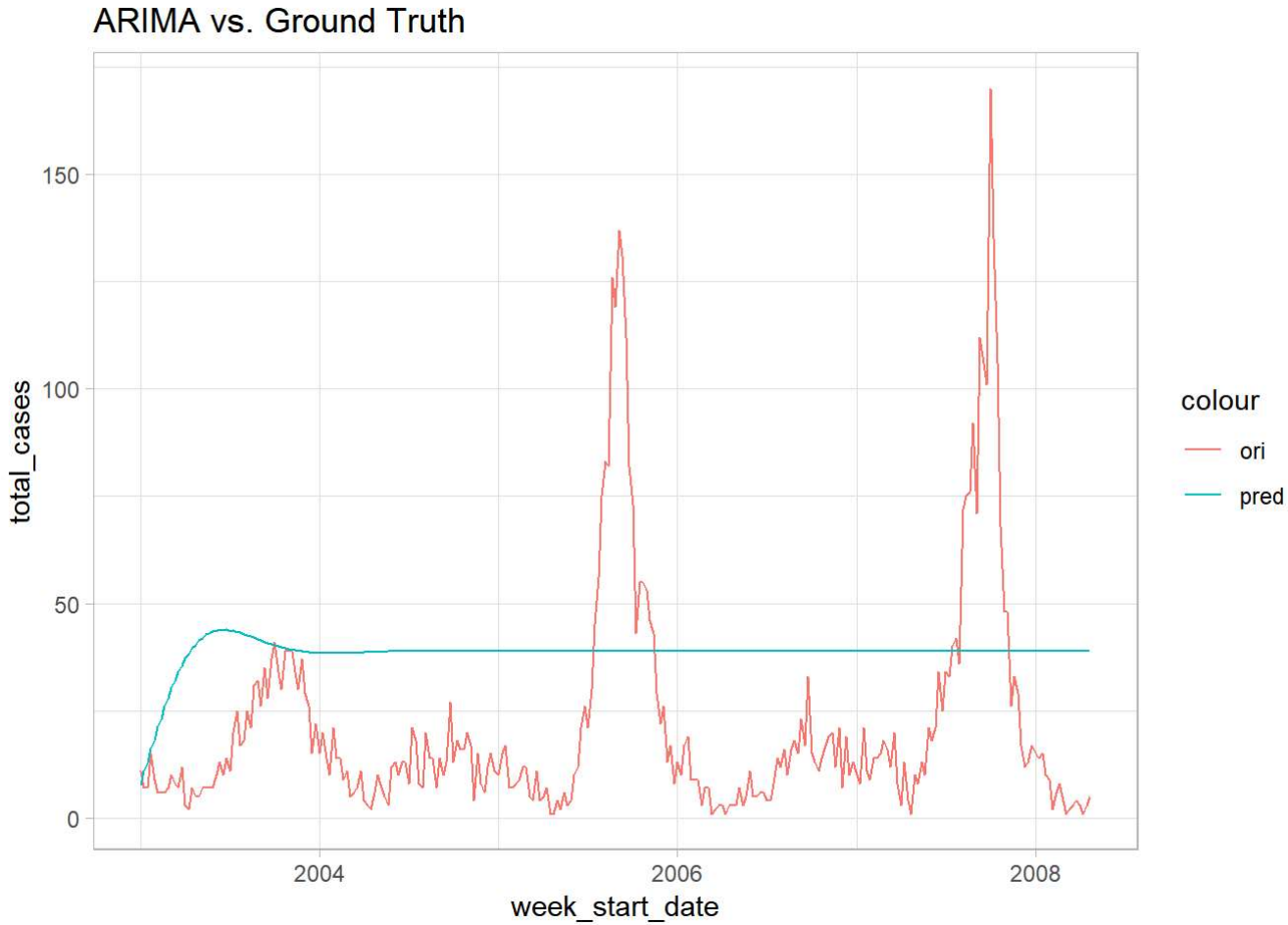
```
# Model 1 (ARIMA)
fc0 <- forecast(sj.fit1,h=277)
plot(fc0)
```




```
accuracy(fc0$mean, as.integer(sjtest.ts1[,"total_cases"]))
```

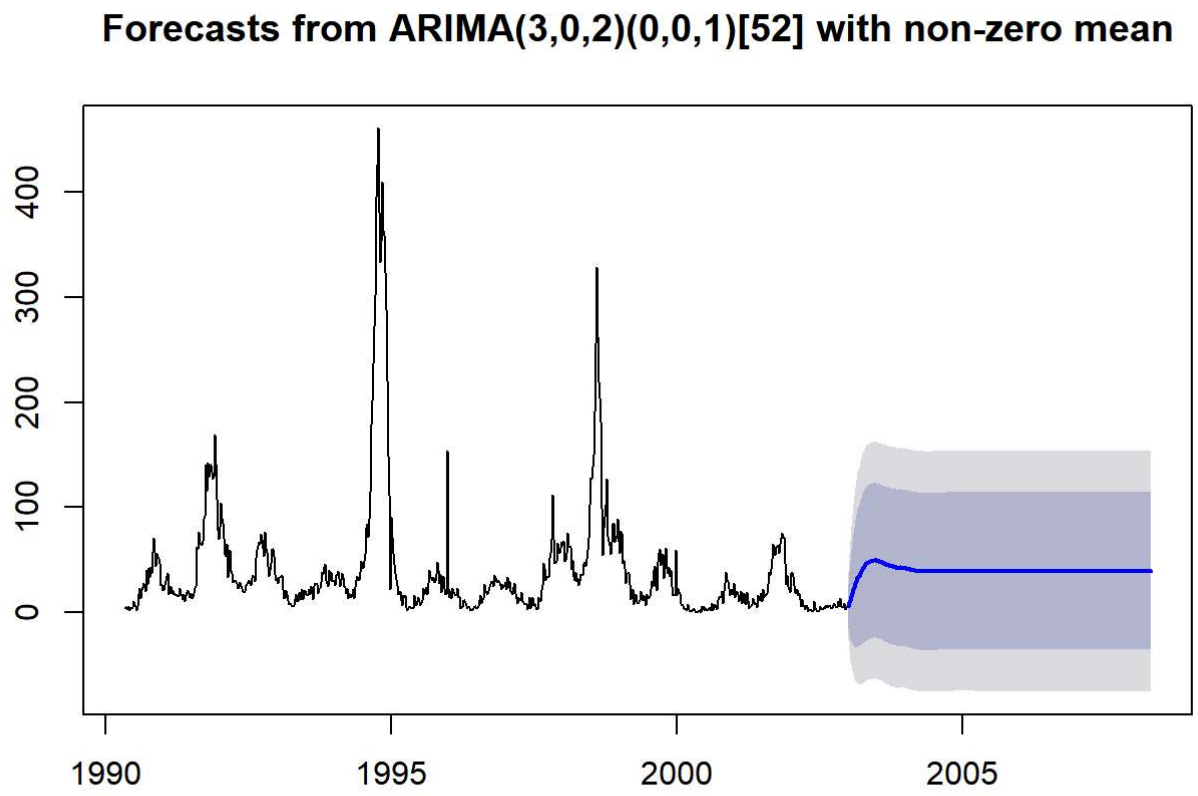
```
##                 ME     RMSE      MAE       MPE     MAPE
## Test set -16.81586 31.55054 27.34825 -399.3748 410.4269
```

```
ggplot() +
  geom_line(mapping=aes(x=week_start_date, y=total_cases, color="ori"), data=sjdata.test) +
  geom_line(mapping=aes(x=sjdata.test$week_start_date, y=fc0$mean, color="pred")) +
  theme_light() + theme(legend.position = "right") +
  ggtitle("ARIMA vs. Ground Truth")
```




```
ggsave("./Charts/ARIMA vs. Ground Truth.png", width = 8, height = 6)
```

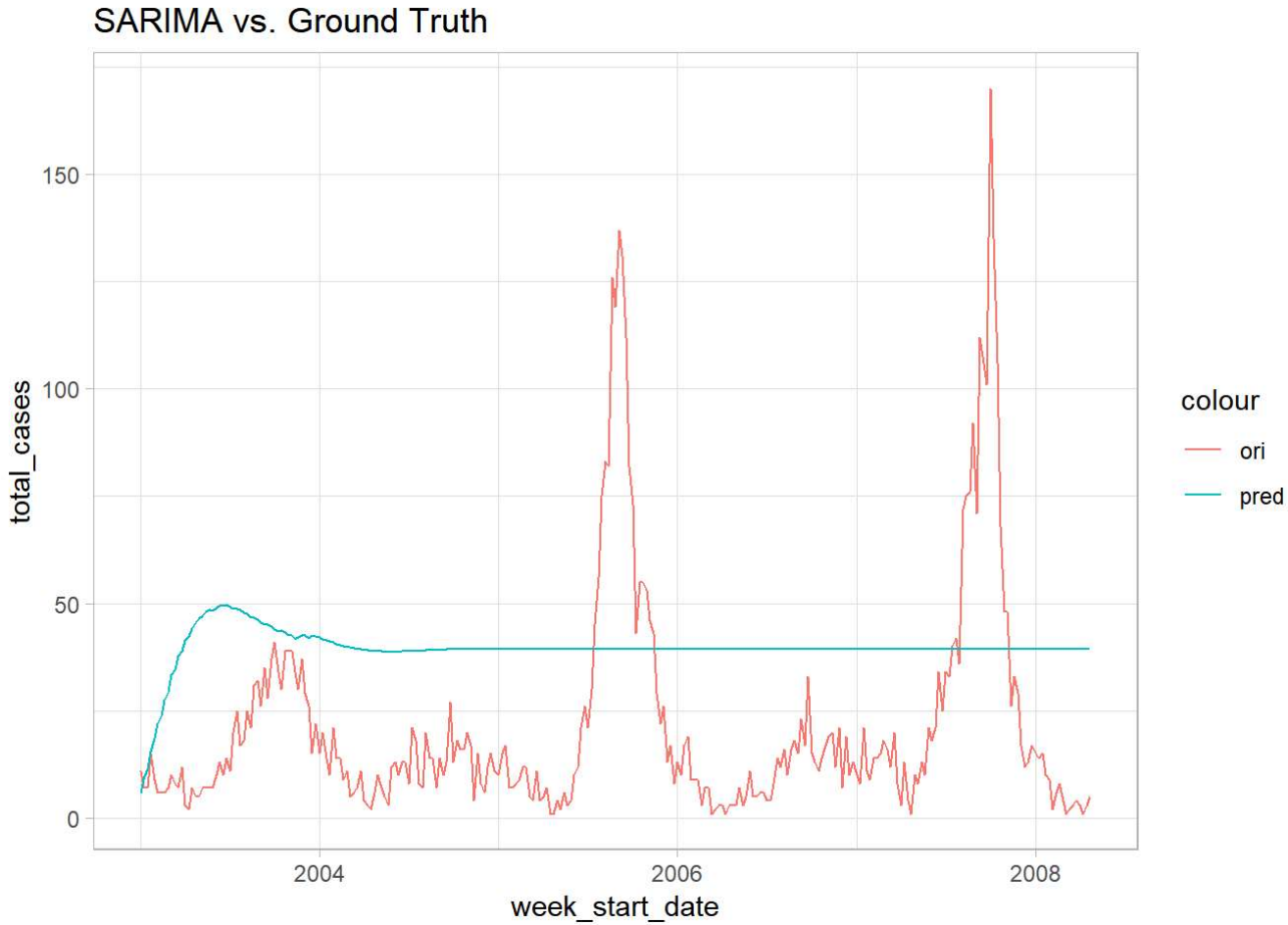
```
# Model 2 (SARIMA)
fc1 <- forecast(sj.fit2, h=277)
plot(fc1)
```




```
accuracy(fc1$mean, as.integer(sjtest.ts1[,"total_cases"]))
```

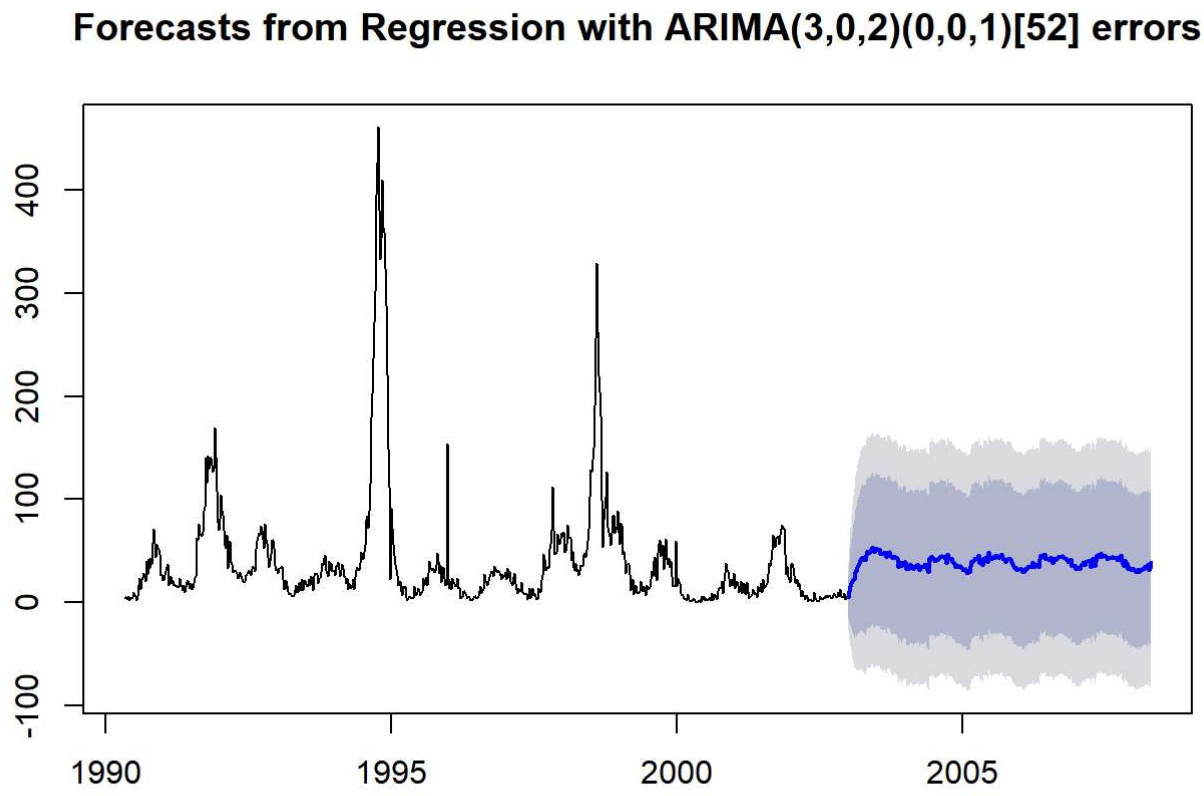
```
##                 ME     RMSE      MAE       MPE    MAPE
## Test set -17.90895 32.29676 28.33627 -411.0115 422.012
```

```
ggplot() +
  geom_line(mapping=aes(x=week_start_date, y=total_cases, color="ori"), data=sjdata.test) +
  geom_line(mapping=aes(x=sjdata.test$week_start_date, y=fc1$mean, color="pred")) +
  theme_light() + theme(legend.position = "right") +
  ggtitle("SARIMA vs. Ground Truth")
```




```
ggsave("./Charts/SARIMA vs. Ground Truth.png", width = 8, height = 6)
```

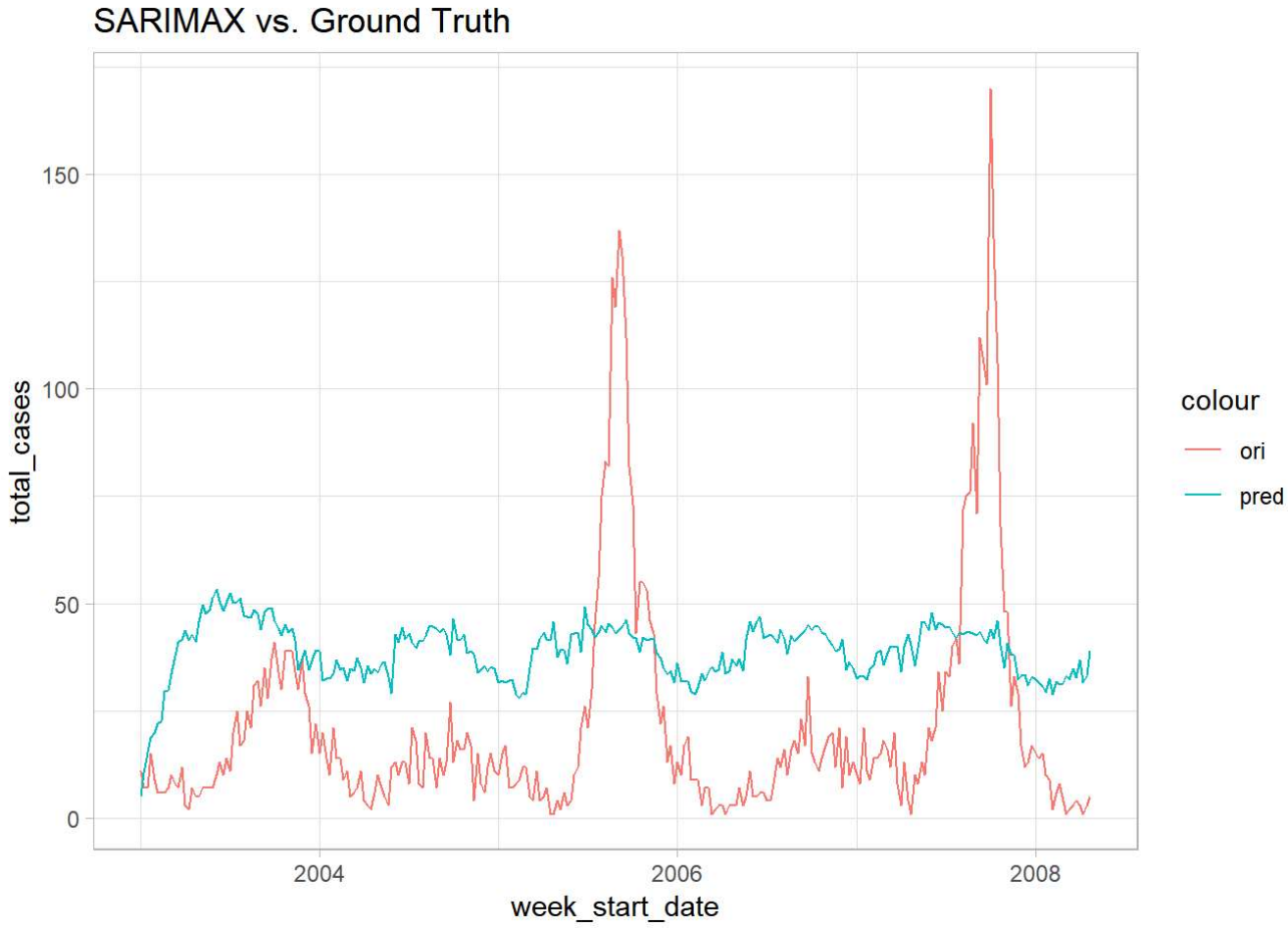
```
# Model 3 (SARIMAX)
fc3 <- forecast(sj.fit3, xreg=cbind(precipitation1[,1],relative.humidity1[,1]
                                    ,dew.temp1[,1],avg.temp1[,1]
                                    ,diur.temp1[,1],max.temp1[,1]
                                    ,min.temp1[,1]), h =277)
plot(fc3)
```




```
accuracy(fc3$mean,as.integer(sjtest.ts1[,"total_cases"]))
```

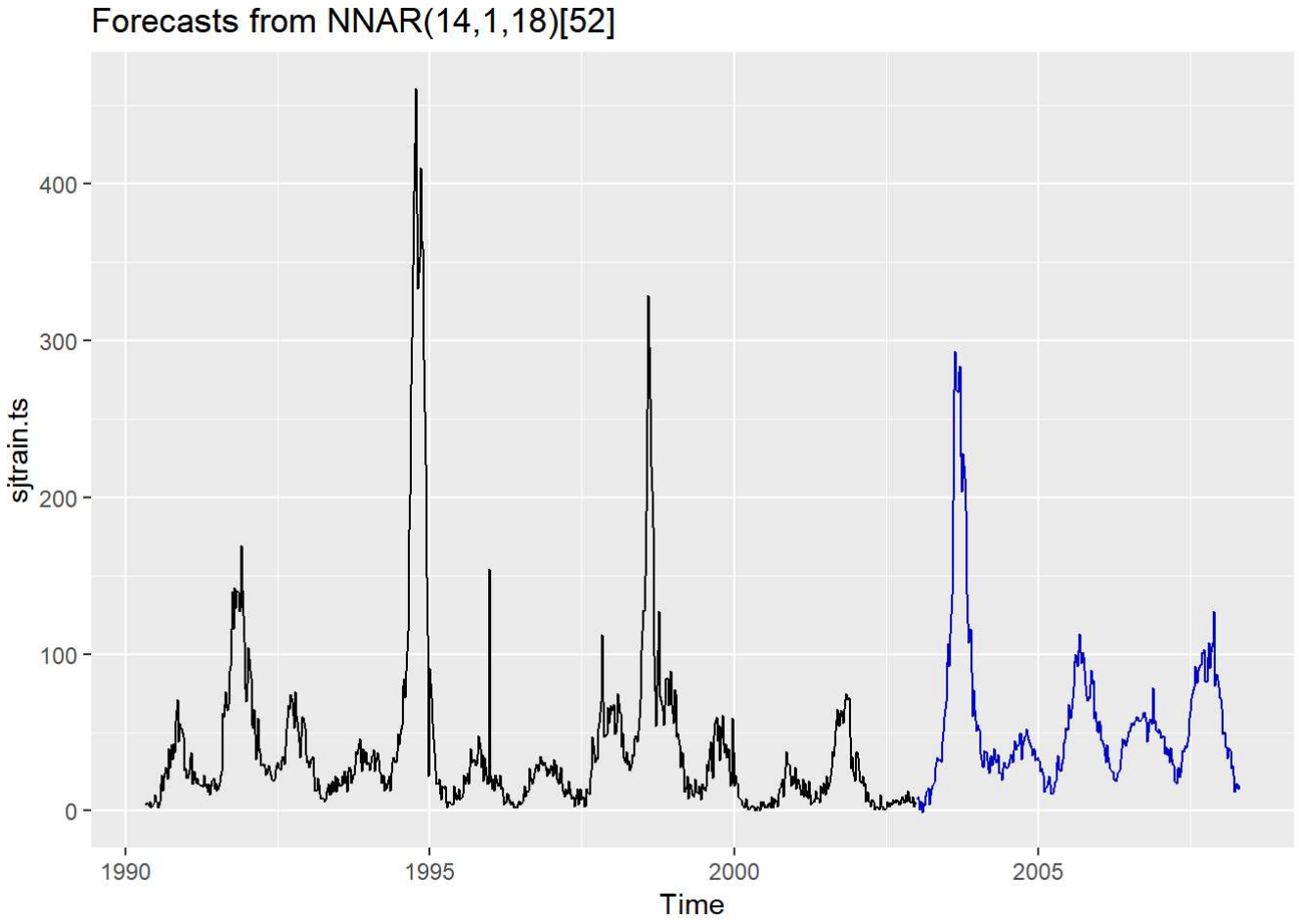
```
##                 ME     RMSE      MAE       MPE     MAPE
## Test set -17.18556 30.79444 26.85361 -385.4345 395.5418
```

```
ggplot() +
  geom_line(mapping=aes(x=week_start_date, y=total_cases, color="ori"), data=sjdata.test) +
  geom_line(mapping=aes(x=sjdata.test$week_start_date, y=fc3$mean, color="pred")) +
  theme_light() + theme(legend.position = "right") +
  ggtitle("SARIMAX vs. Ground Truth")
```



```
ggsave("./Charts/SARIMAX vs. Ground Truth.png", width = 8, height = 6)
```

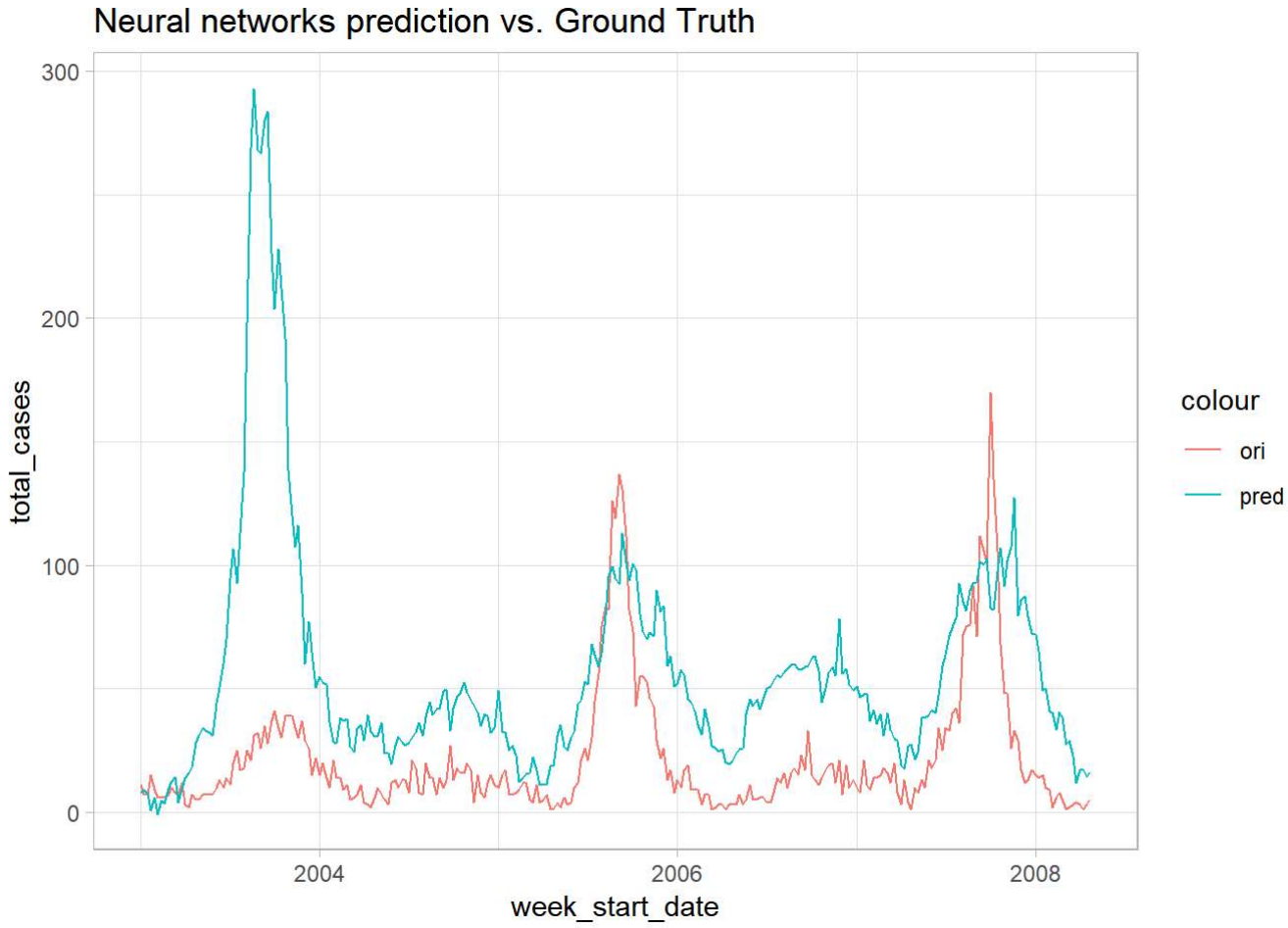
```
# Try NN
fc4 <- forecast(sj.fit4, xreg=sjtest.ts1[,5:23], h=277)
autoplot(fc4)
```




```
accuracy(fc4$mean,as.integer(sjtest.ts1[,"total_cases"]))
```

```
##                 ME     RMSE      MAE      MPE     MAPE
## Test set -35.33398 56.76832 37.78695 -357.191 361.7181
```

```
ggplot() +
  geom_line(mapping=aes(x=week_start_date, y=total_cases, color="ori"), data=sjdata.test) +
  geom_line(mapping=aes(x=sjdata.test$week_start_date, y=fc4$mean, color="pred")) +
  theme_light() + theme(legend.position = "right") +
  ggtitle("Neural networks prediction vs. Ground Truth")
```



```
ggsave("./Charts/Neural networks prediction vs. Ground Truth.png", width = 8, height = 6)
```