DATA471 Group Report

Group 15

2023-08-22

Data Preparation

Source:

Data is downloaded from 'COVID-19 data portal' at https://www.stats.govt.nz/experimental/covid-19-data-portal?tab=Economic&category=Activity&indicator=New%20Zealand%20Activity%20Index%20(NZAC)

File format is Excel including 2 data tabs: metadata and data.

Data Description:

tab 'metadata'
ResourceID: A unique ID generated to represent the indicator.
Subject: The indicator category, such as activity or confidence.
Title: The indicator name.
Description: A short description about the indicator.
Notes: A brief additional description about the indicator.
Caveats: Includes any footnotes or specific/ detailed description of the indicator.
Source: Denotes the source name of the indicator.
Source URL: Denotes the source url of the indicator.
Modified: Denotes the last modified date of the indicator.
Frequency: Indicates the reccurrence of the updates made to the indicator.

tab 'data'
ResourceID: A unique ID generated to represent the indicator.
Geo: Denotes the factor associated with the indicator series. Not applicable for this dataset.
Period: Indicates the date (in the format yyyy-mm-dd) for the timeseries data.
Label1: Indicator used to calculate the NZAC.
Label2: NA.
Label3: NA.
Value: The value of the series.
Unit: The units associated with this series.
Measure: Describes the values measured by the indicator.
Mutliplier: Denotes the factor associated with the indicator series.

Data Analysis

There is only one row in 'metadata' corresponding to the single ResourceID for NZAC in the 'data' tab. In addition, it indicates the data is monthly frequency. Thus, we will load data from tab 'data' directly without joining to tab 'metadata' for the indicator time series.

```
indicators_data <- readxl::read_xlsx("data/covid_19_data_portal.xlsx", sheet = "data")

cat("Columns Name:\n")
print(names(indicators_data))

cat("\nList of Indicators:\n")
print(unique(indicators_data$Label1))


cat("\nPeriod range: ")
cat(paste0(min(indicators_data$Period), " - ", max(indicators_data$Period)))

cat("\nThe first 6 rows:")
print(head(indicators_data))
```

```
## Columns Name:
##  [1] "ResourceID" "Geo"        "Period"     "Label1"     "Label2"
##  [6] "Label3"     "Value"      "Unit"       "Measure"    "Multiplier"
##
## List of Indicators:
## [1] "NZAC component - Card transaction spend"
## [2] "NZAC component - Electricity grid demand"
## [3] "NZAC component - Manufacturing index"
## [4] "NZAC component - New jobs posted online"
## [5] "NZAC component - Traffic index (heavy)"
## [6] "NZAC component - Traffic index (light)"
## [7] "New Zealand Activity Index (NZAC)"
##
## Period range: 2004-01-31 - 2023-06-30
## The first 6 rows:# A tibble: 6 x 10
##   ResourceID Geo   Period     Label1  Label2 Label3  Value Unit  Measure Multi~1
##   <chr>      <lgl> <chr>      <chr>   <lgl>  <lgl>   <dbl> <chr> <chr>     <dbl>
## 1 CPACT12    NA    2004-01-31 NZAC c~ NA     NA     11.2   %     Annual~       0
## 2 CPACT12    NA    2004-01-31 NZAC c~ NA     NA      0.528 %     Annual~       0
## 3 CPACT12    NA    2004-01-31 NZAC c~ NA     NA     55.8   %     Annual~       0
## 4 CPACT12    NA    2004-01-31 NZAC c~ NA     NA     25.2   %     Annual~       0
## 5 CPACT12    NA    2004-01-31 NZAC c~ NA     NA      8.38  %     Annual~       0
## 6 CPACT12    NA    2004-01-31 NZAC c~ NA     NA      3.7   %     Annual~       0
## # ... with abbreviated variable name 1: Multiplier
```

Cleaning data

```
# 1. Remove unused columns
indicators_cleaned <- indicators_data[,c(3, 4, 7)]

# 2. Clean Label1: to remove 'NZAC component - ' part
```

```r
indicators_cleaned$Label1 <- gsub('NZAC component - ', '', indicators_cleaned$Label1)

# 3. Convert to date
indicators_cleaned$Period <- as.Date(indicators_cleaned$Period)

cat("\nThe first 6 rows:")
```

```
##
## The first 6 rows:
```

```r
knitr::kable(head(indicators_cleaned))
```

Period	Label1	Value
2004-01-31	Card transaction spend	11.2260
2004-01-31	Electricity grid demand	0.5283
2004-01-31	Manufacturing index	55.7939
2004-01-31	New jobs posted online	25.2309
2004-01-31	Traffic index (heavy)	8.3823
2004-01-31	Traffic index (light)	3.7000

Load package for analysis

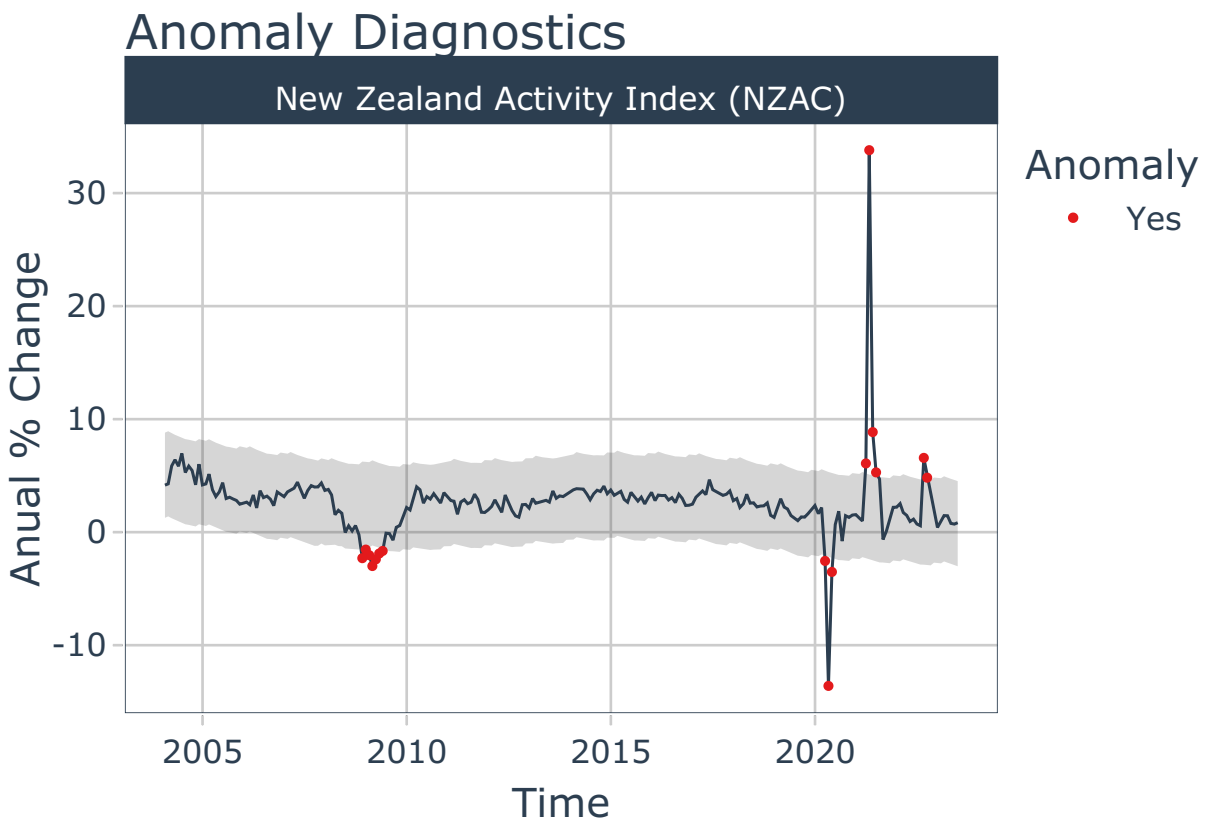
```r
library(ggplot2)
library(dplyr)
library(tidyr)
library(forecast)
library(lubridate)
library(tsibble)
library(zoo)
library(timetk)
library(tsbox)
```

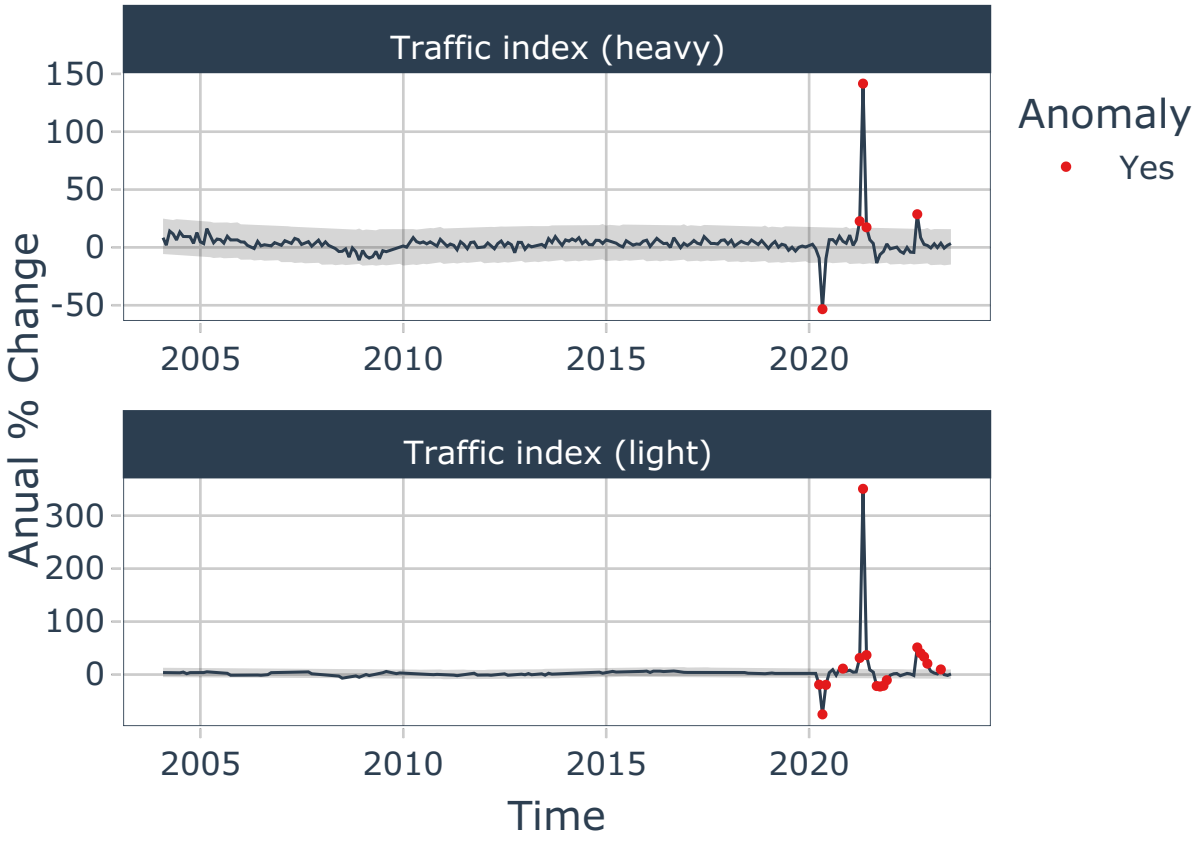
Create time series objects for indicators

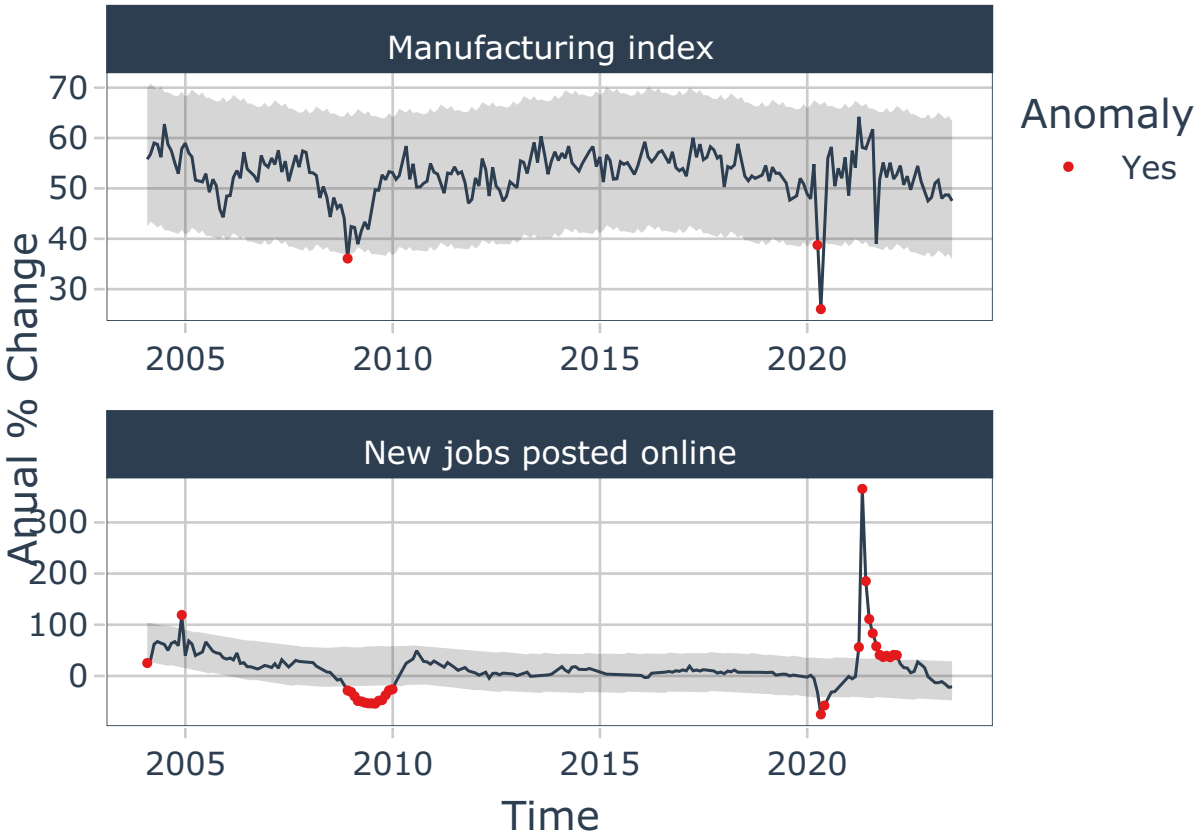
```r
# Convert dataset tsibble, zoo
indicators_ts <- indicators_cleaned %>% as_tsibble(index=Period, key=Label1)
indicators_zoo <- indicators_cleaned %>% pivot_wider(names_from="Label1",
                                                     values_from="Value") %>% read.zoo()
```

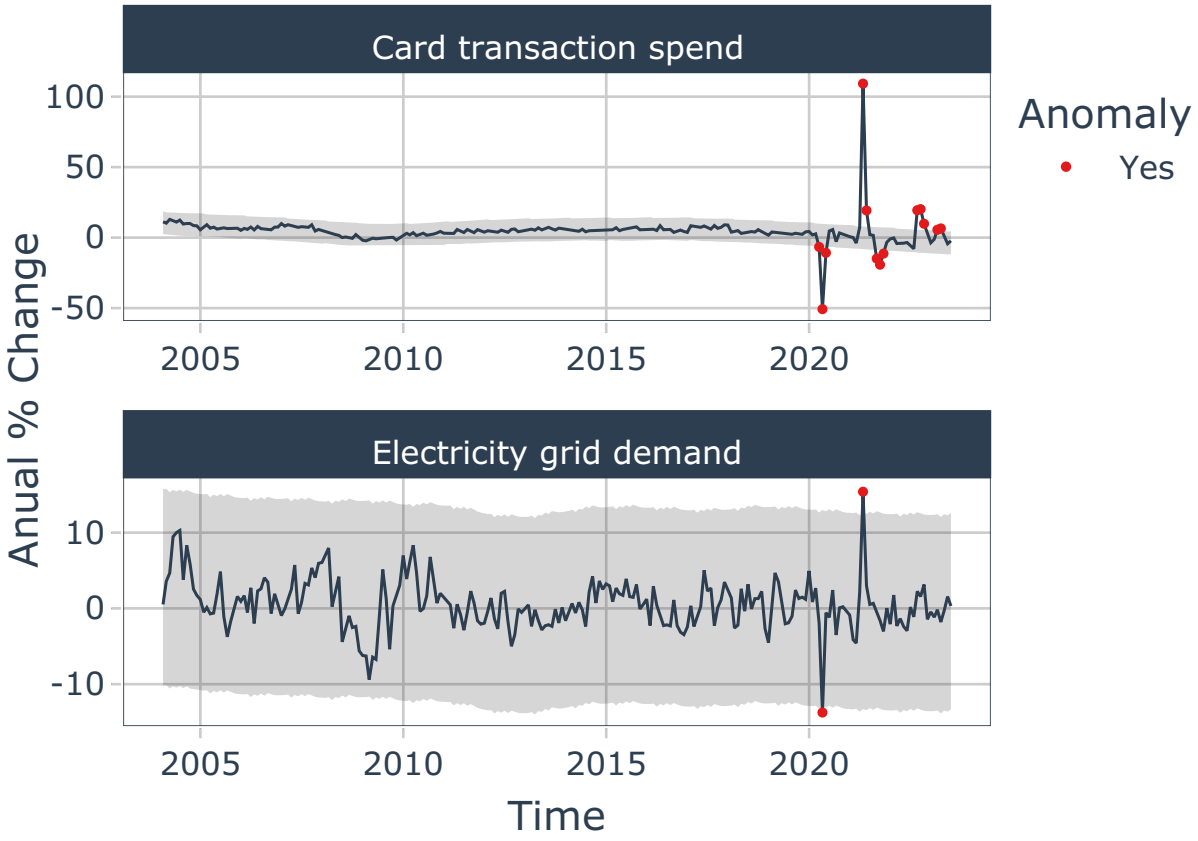
Plot the Anomaly Diagnostics:

- The maximum percent of anomalies permitted to be identified is set to 0.2
  There were 2 main periods with anomalies, economic recession 2008 and pandemic 2020, in the following time series: NZAC index, Manufacturing, and new job posted online.
  While the other time series indicators have anomalies during the pandemic 2020 only. They are heavy traffic, ligh traffic, card transaction, and electricity demand.
  New Job posted online had another periods with identified anomalies during 2004 and 2005.


Anomaly Diagnostics
New Zealand Activity Index (NZAC)
Anual % Change
30
20
10
0
-10
2005
2010
2015
2020
Time
Anomaly
Yes

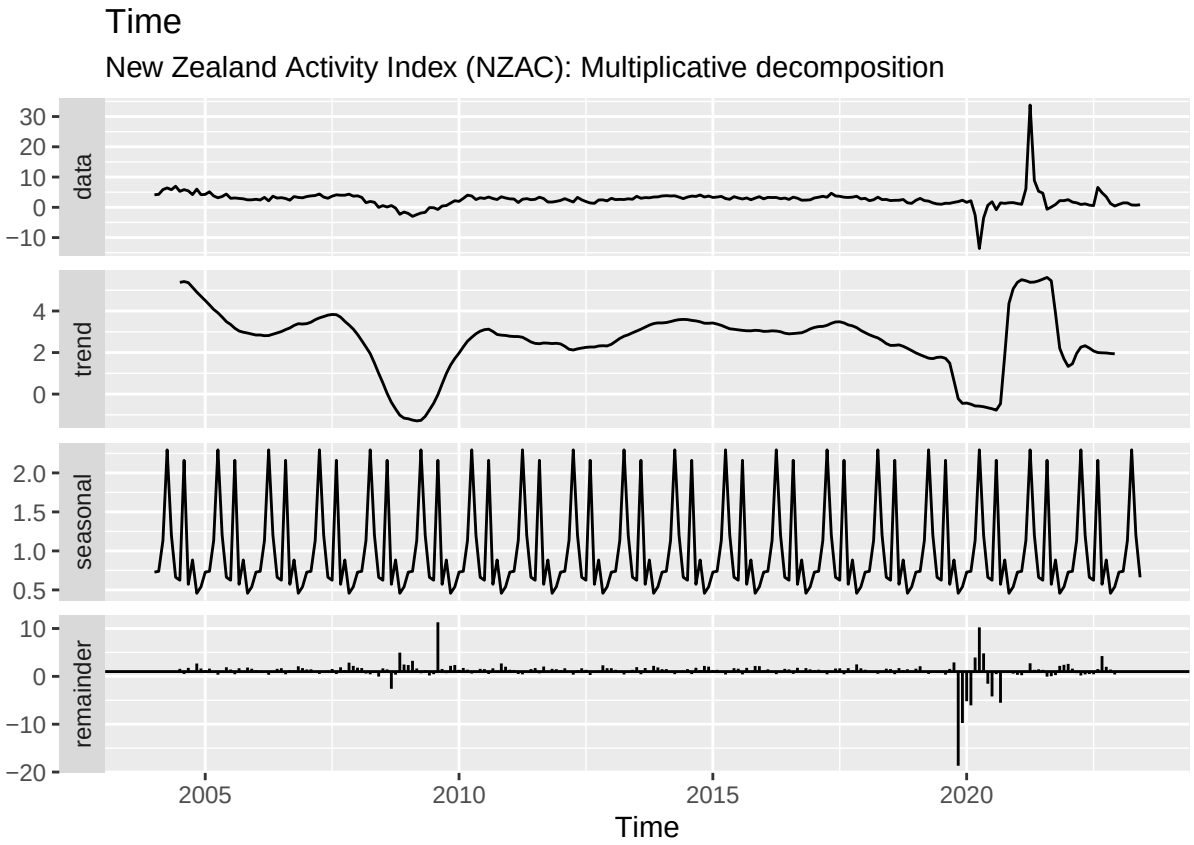

Traffic index (heavy)
150
100
50
0
-50
2005
2010
2015
2020
Traffic index (light)
300
200
100
0
2005
2010
2015
2020
Anual % Change
Time
Anomaly
Yes

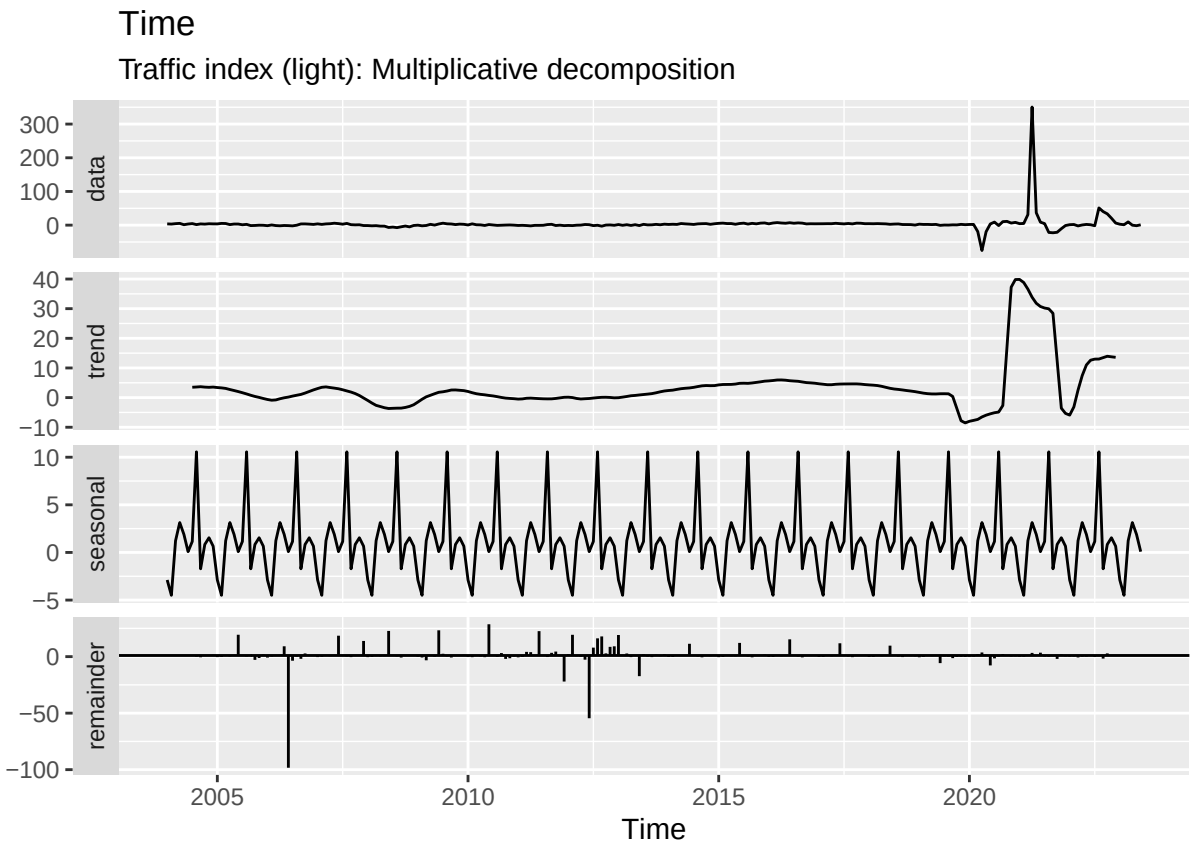

Manufacturing index
New jobs posted online
Anual % Change
70
60
50
40
30
300
200
100
0
2005
2010
2015
2020
Time
Anomaly
Yes

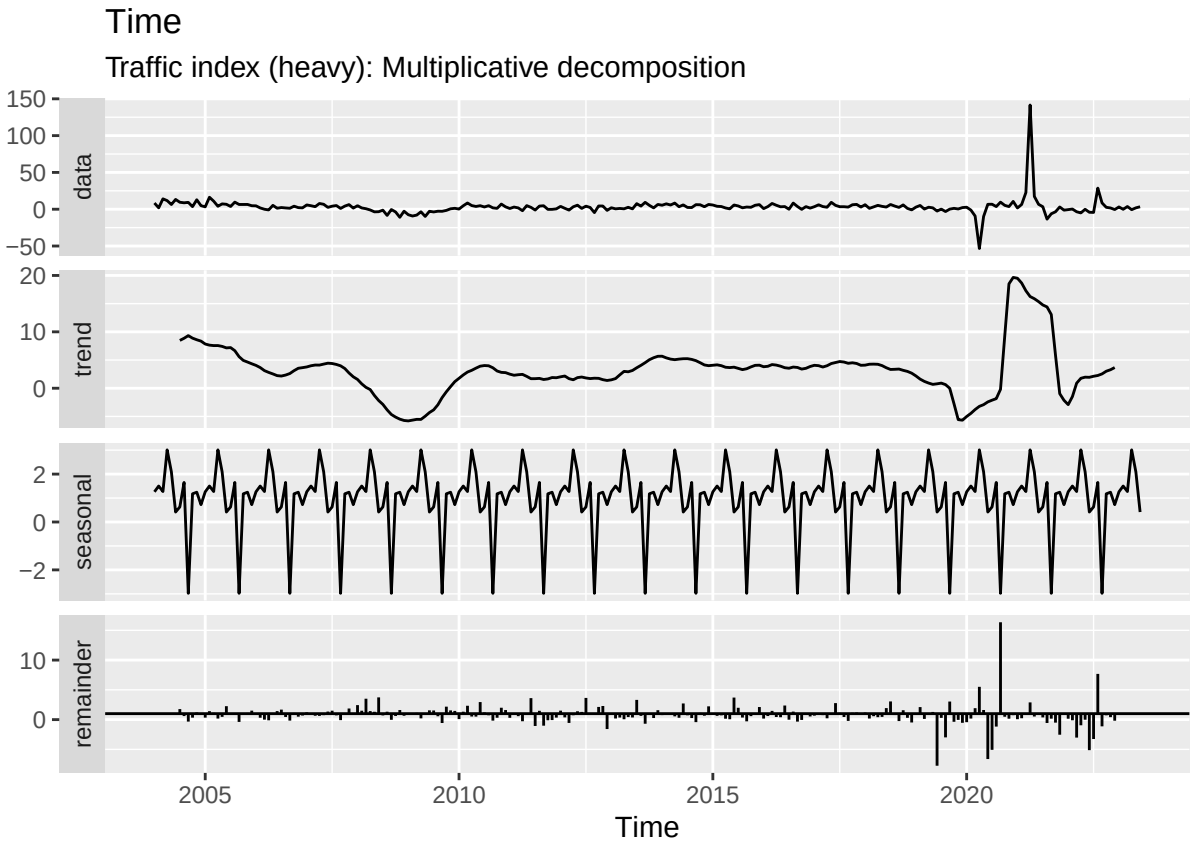


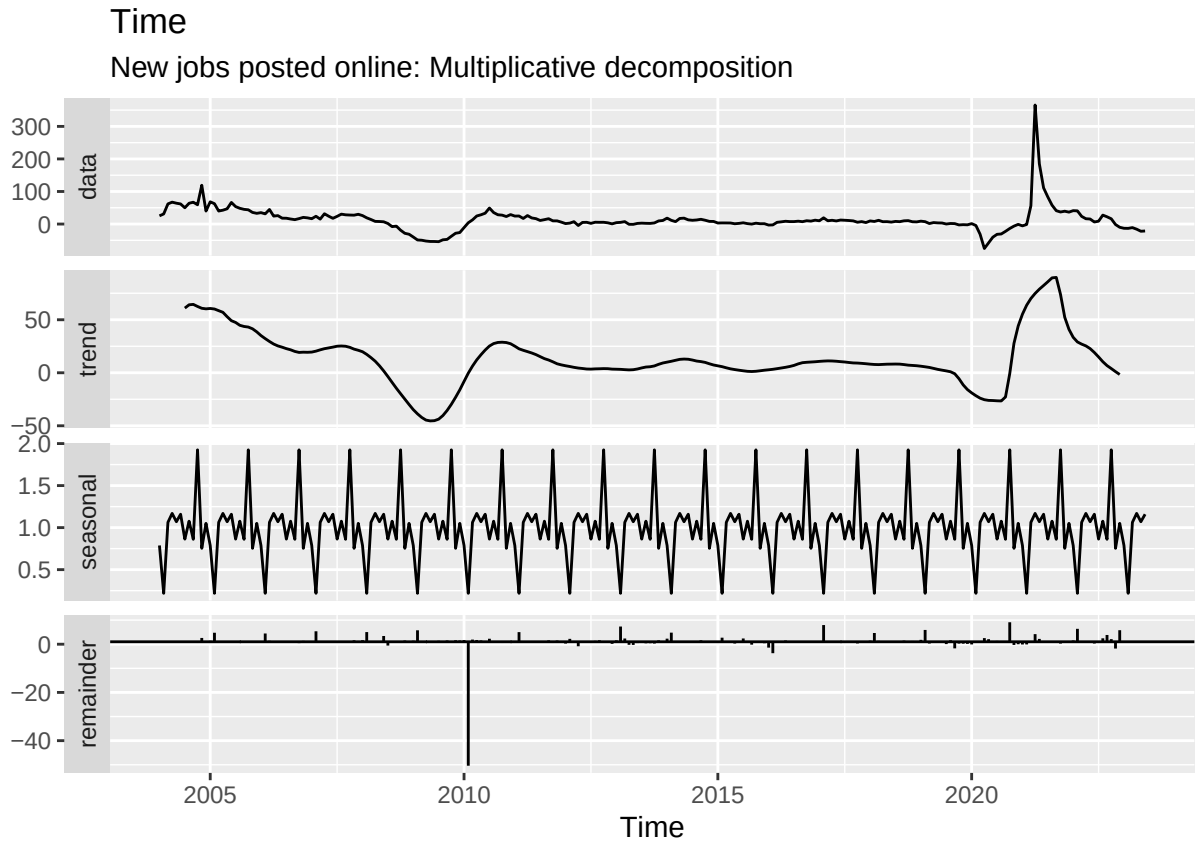
Decomposition

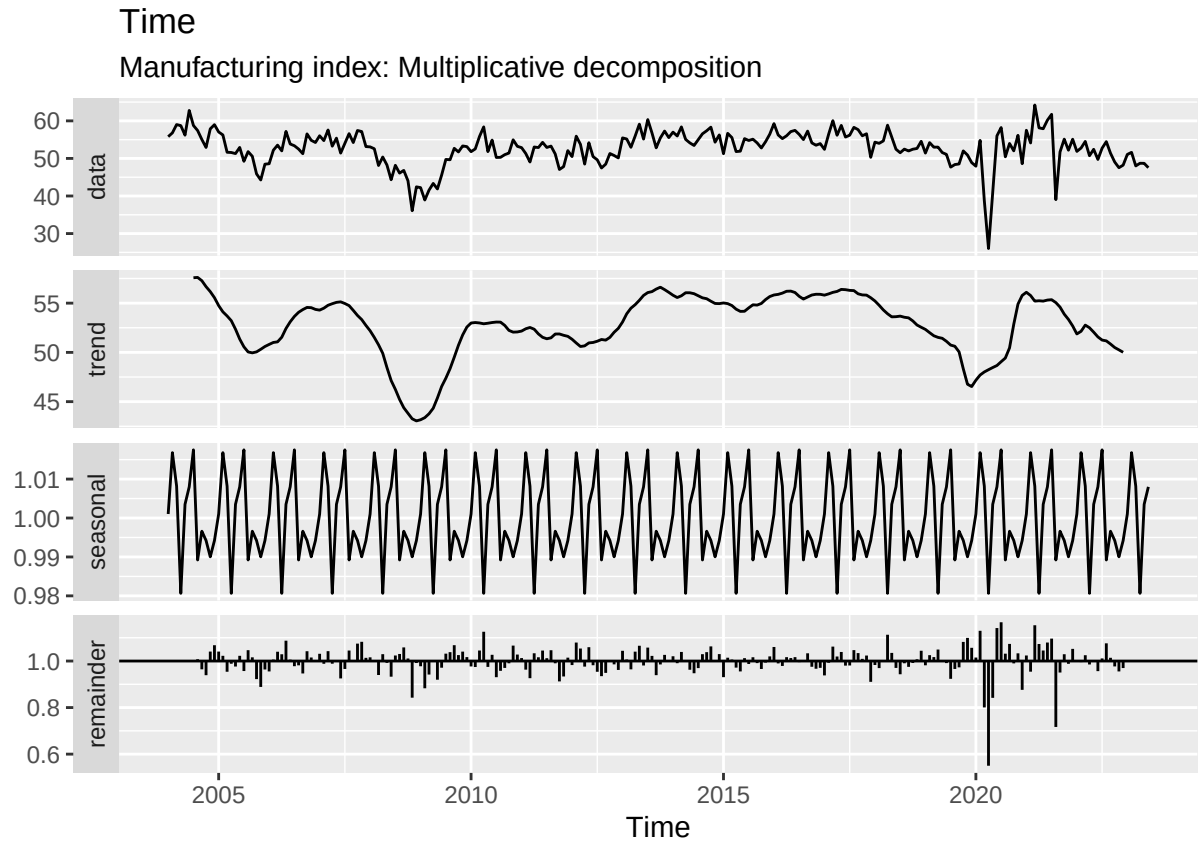
NZAC:
- Trend:
- Seasonal: There were high fluctuation in the first half of every year
- Noise: significant noise during 2008 recession and 2020 pandemic

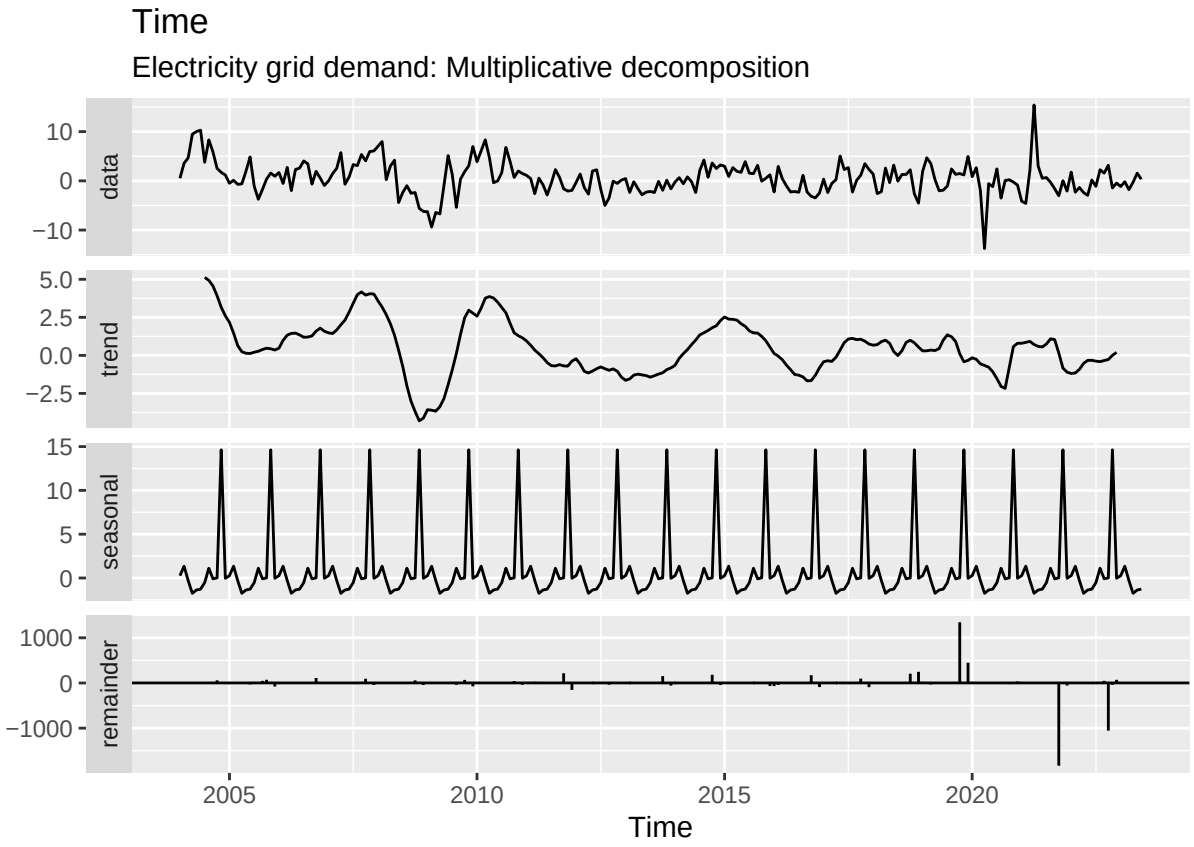

Time
New Zealand Activity Index (NZAC): Multiplicative decomposition
data
trend
seasonal
remainder
30
20
10
0
−10
4
2
0
2.0
1.5
1.0
0.5
10
0
−10
−20
2005
2010
2015
2020
Time

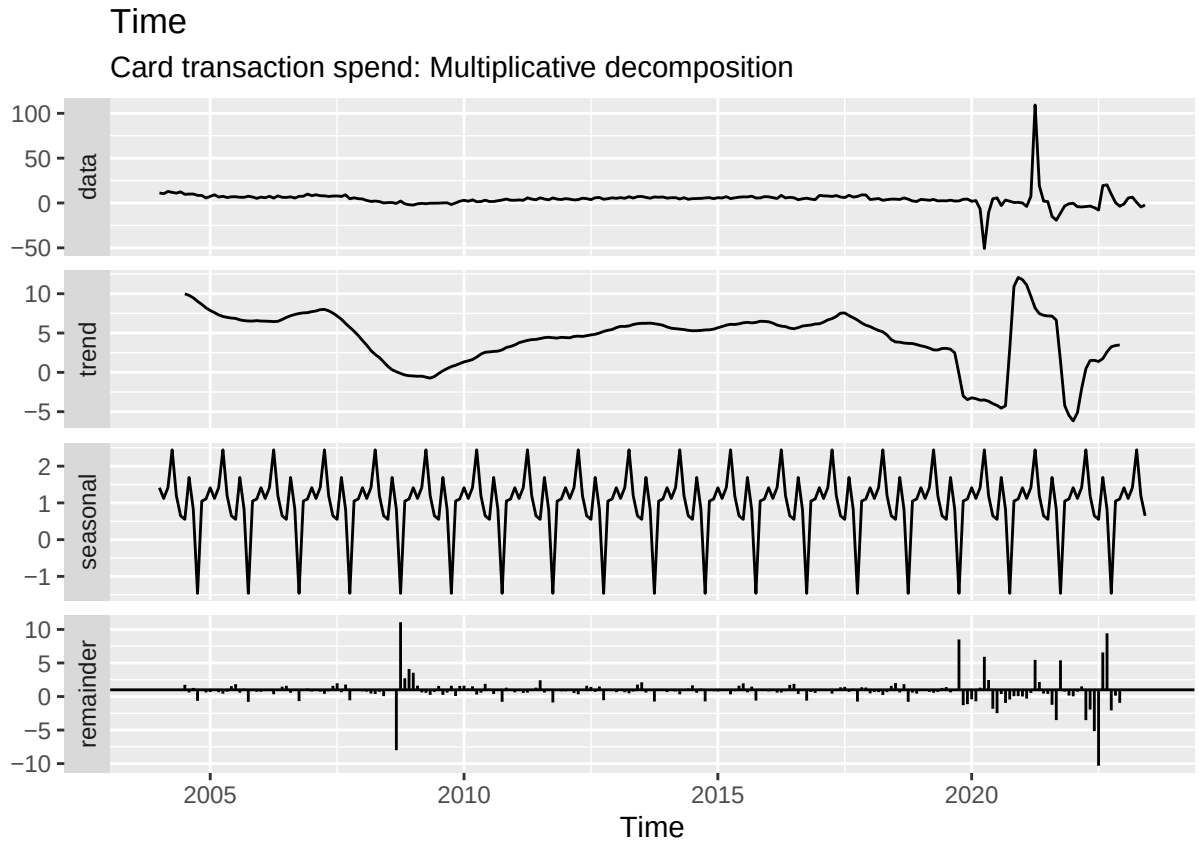

Time
Traffic index (light): Multiplicative decomposition
data
trend
seasonal
remainder
300
200
100
0
40
30
20
10
0
−10
10
5
0
−5
0
−50
−100
2005
2010
2015
2020
Time


Time
Traffic index (heavy): Multiplicative decomposition
data
trend
seasonal
remainder
150
100
50
0
−50
20
10
0
2
0
−2
10
0
2005
2010
2015
2020
Time

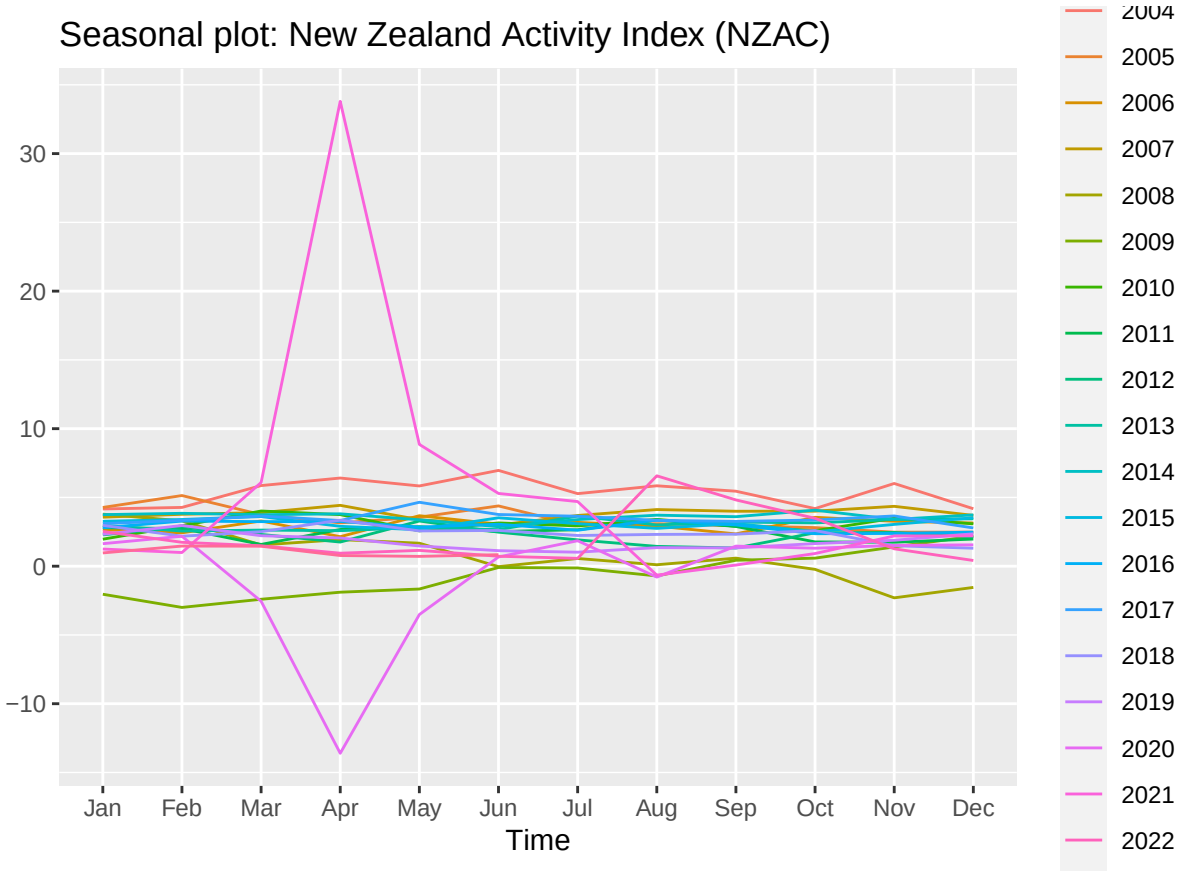

Time
New jobs posted online: Multiplicative decomposition
data
trend
seasonal
remainder
300
200
100
0
50
0
−50
2.0
1.5
1.0
0.5
0
−20
−40
2005
2010
2015
2020
Time

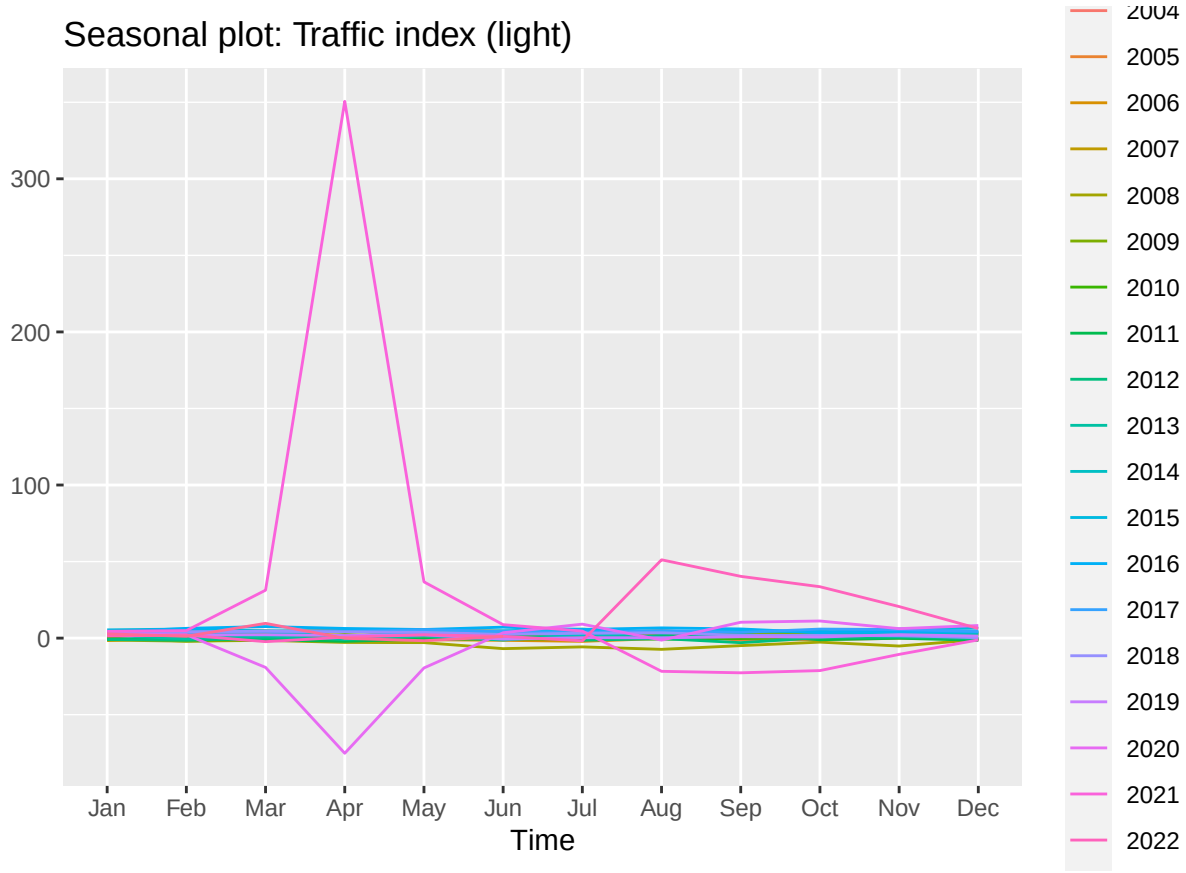

Time
Manufacturing index: Multiplicative decomposition
data
trend
seasonal
remainder
60
50
40
30
55
50
45
1.01
1.00
0.99
0.98
1.0
0.8
0.6
2005
2010
2015
2020
Time

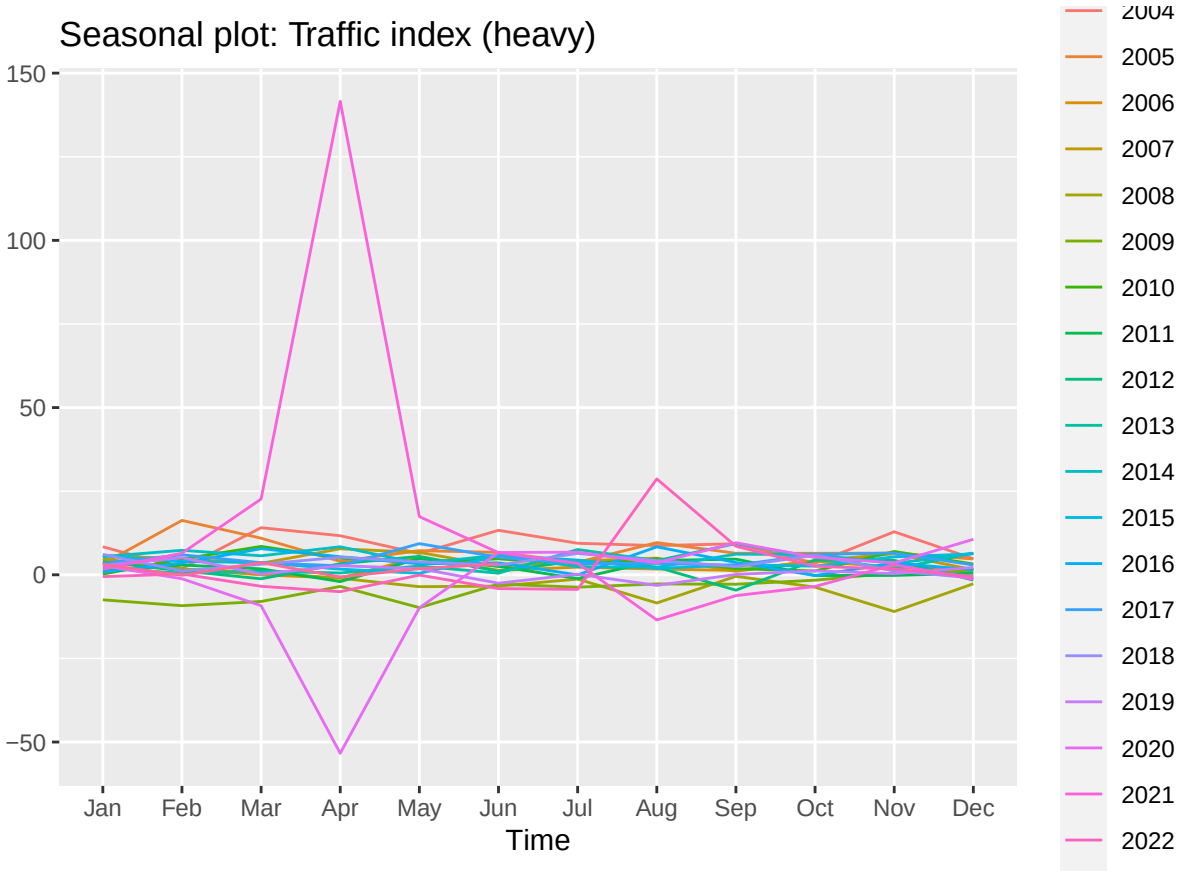

Time
Electricity grid demand: Multiplicative decomposition
data
trend
seasonal
remainder
10
0
−10
5.0
2.5
0.0
−2.5
15
10
5
0
1000
0
−1000
2005
2010
2015
2020
Time

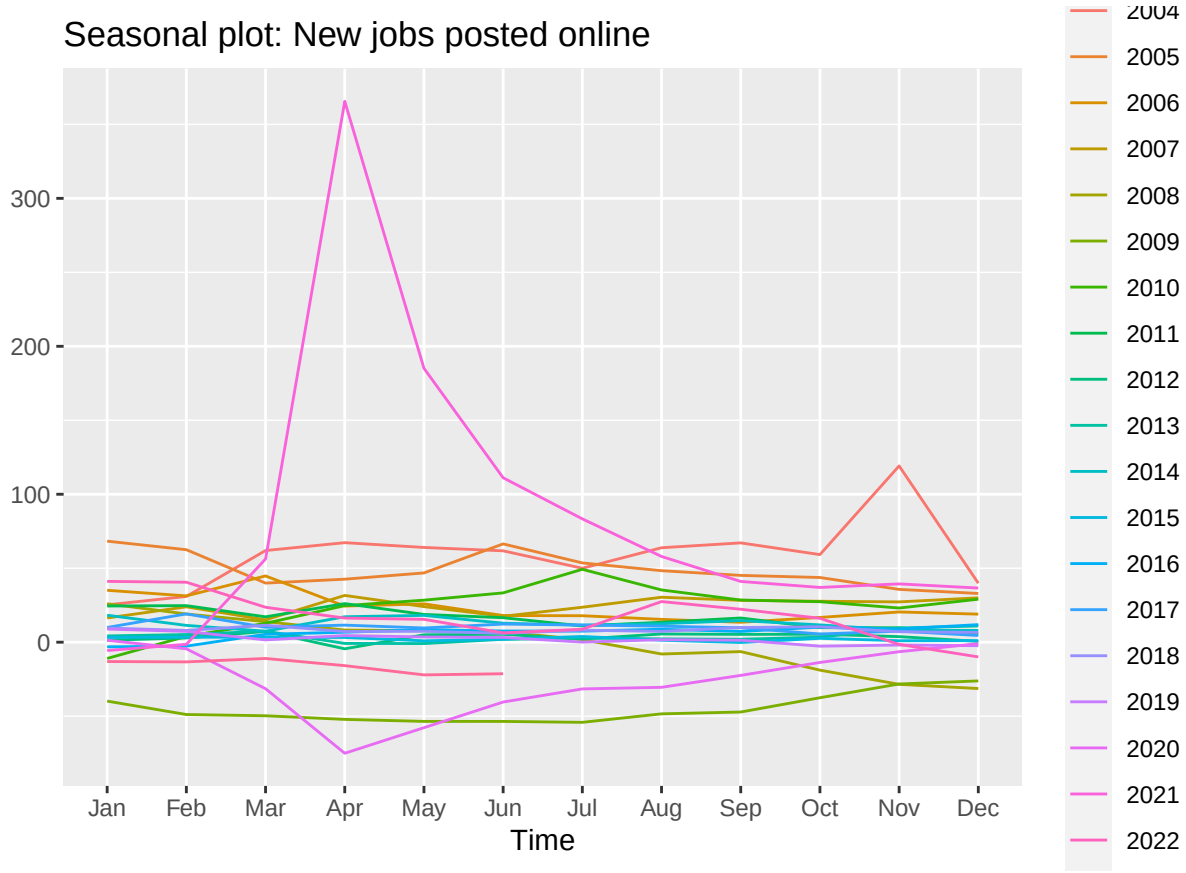


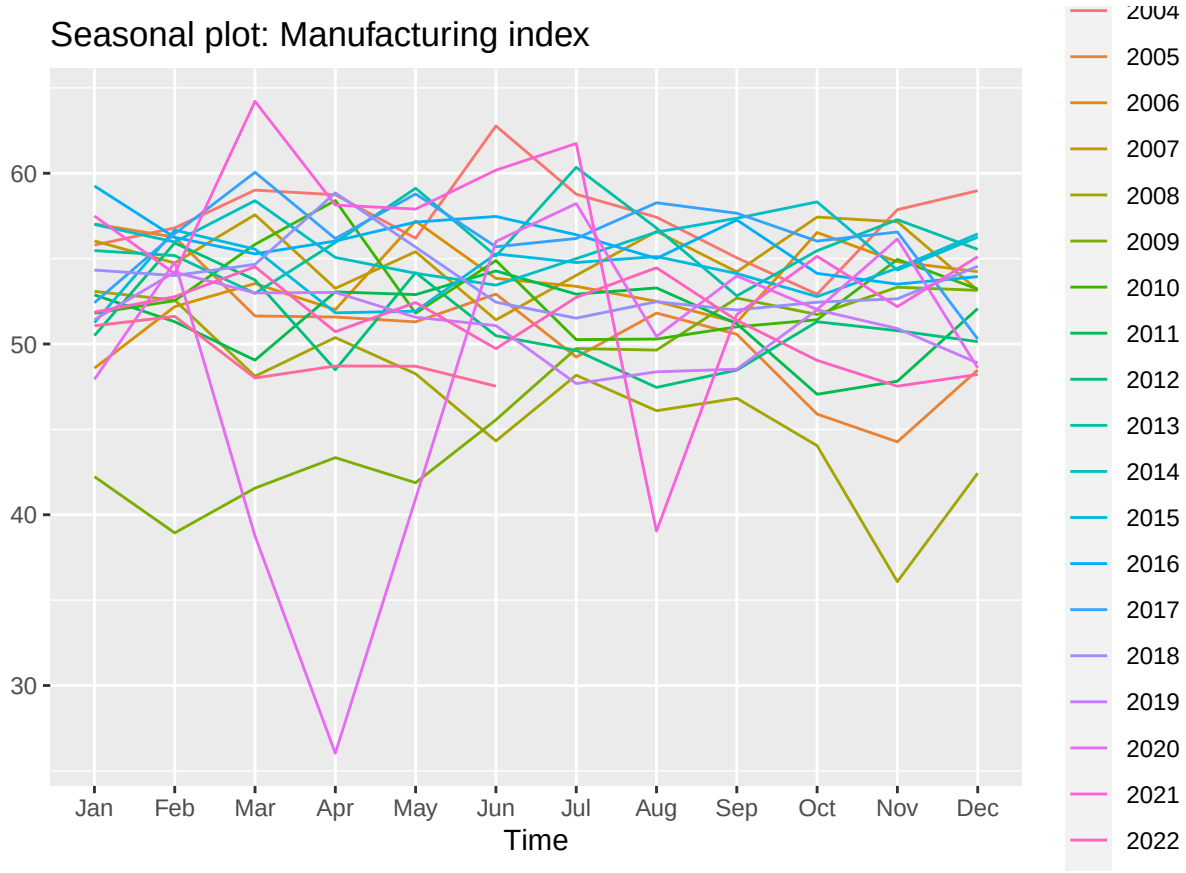
Plot seasonal

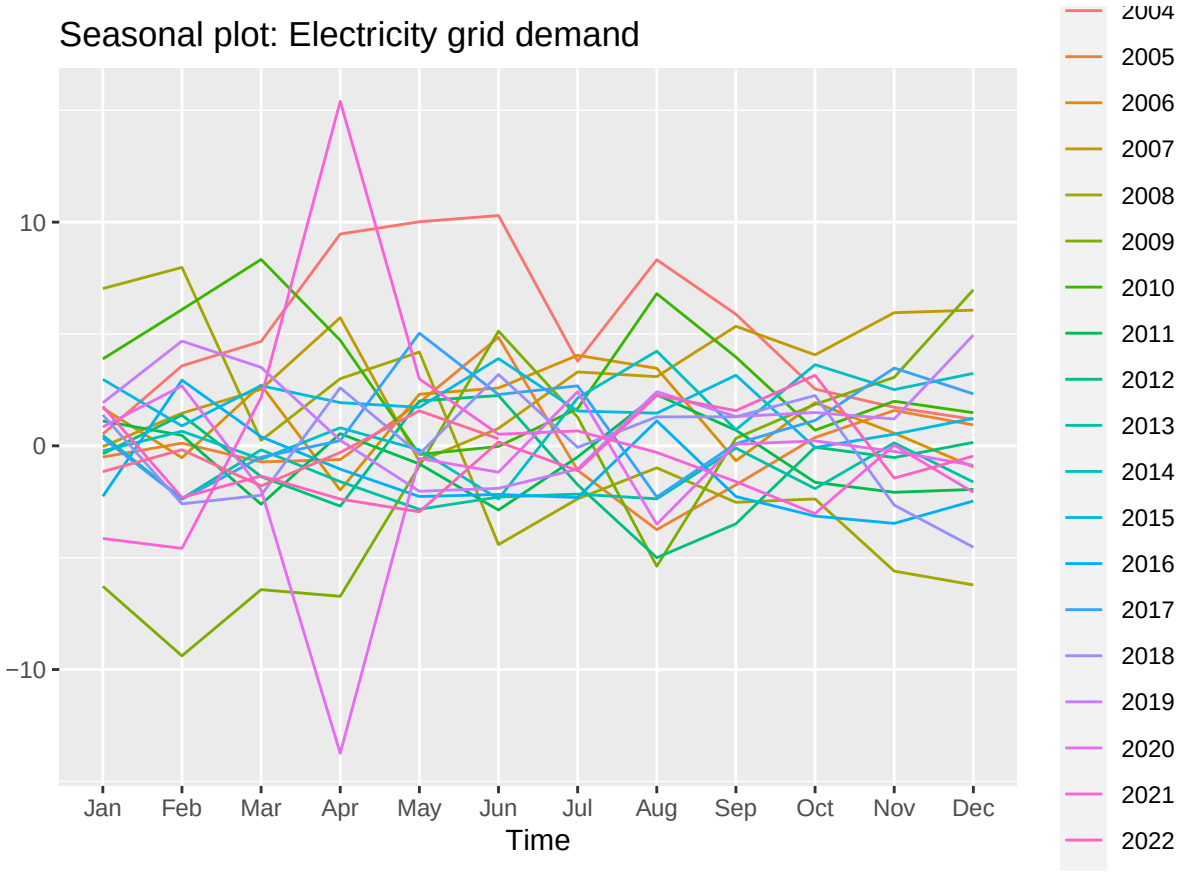

Seasonal plot: New Zealand Activity Index (NZAC)
30
20
10
0
−10
Jan
Feb
Mar
Apr
May
Jun
Jul
Aug
Sep
Oct
Nov
Dec
Time
2004
2005
2006
2007
2008
2009
2010
2011
2012
2013
2014
2015
2016
2017
2018
2019
2020
2021
2022

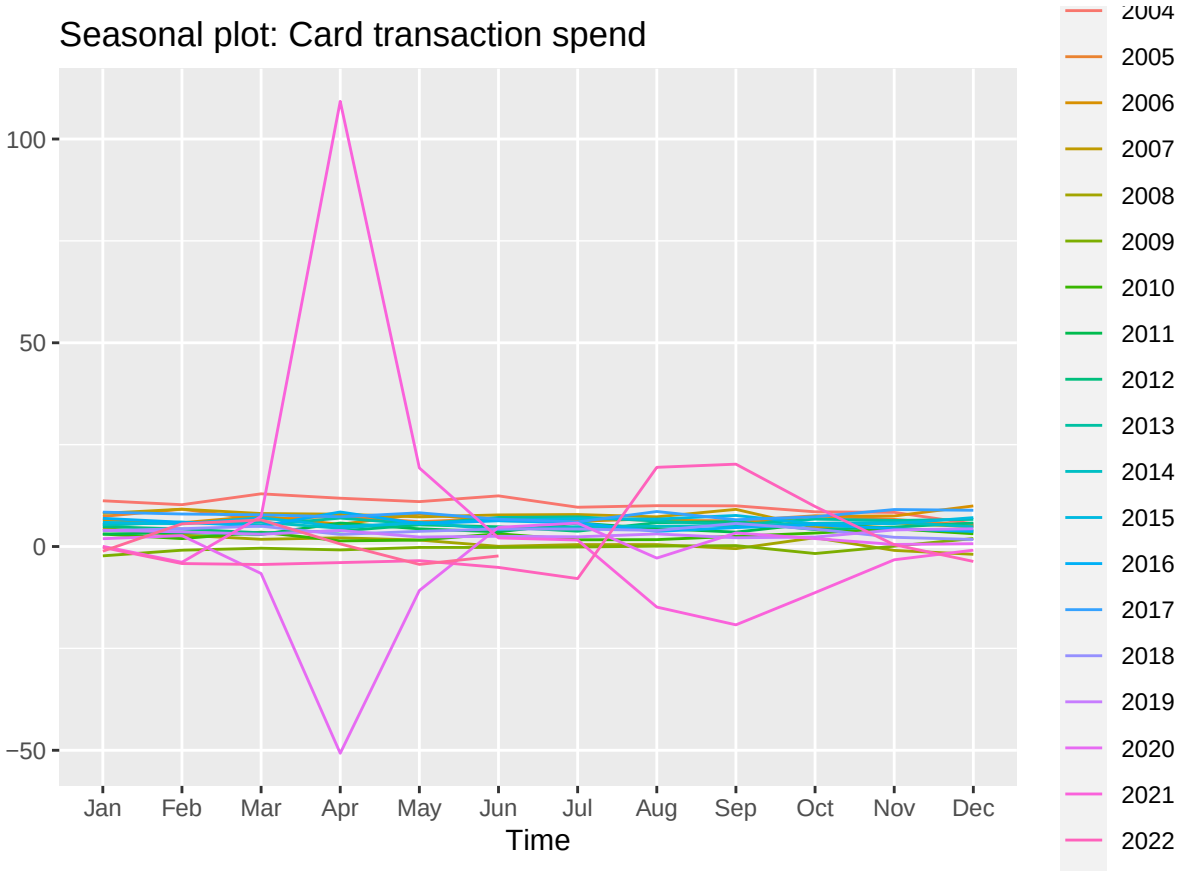

Seasonal plot: Traffic index (light)
300
200
100
0
Jan
Feb
Mar
Apr
May
Jun
Jul
Aug
Sep
Oct
Nov
Dec
Time
2004
2005
2006
2007
2008
2009
2010
2011
2012
2013
2014
2015
2016
2017
2018
2019
2020
2021
2022


Seasonal plot: Traffic index (heavy)
150
100
50
0
−50
Jan
Feb
Mar
Apr
May
Jun
Jul
Aug
Sep
Oct
Nov
Dec
Time
2004
2005
2006
2007
2008
2009
2010
2011
2012
2013
2014
2015
2016
2017
2018
2019
2020
2021
2022

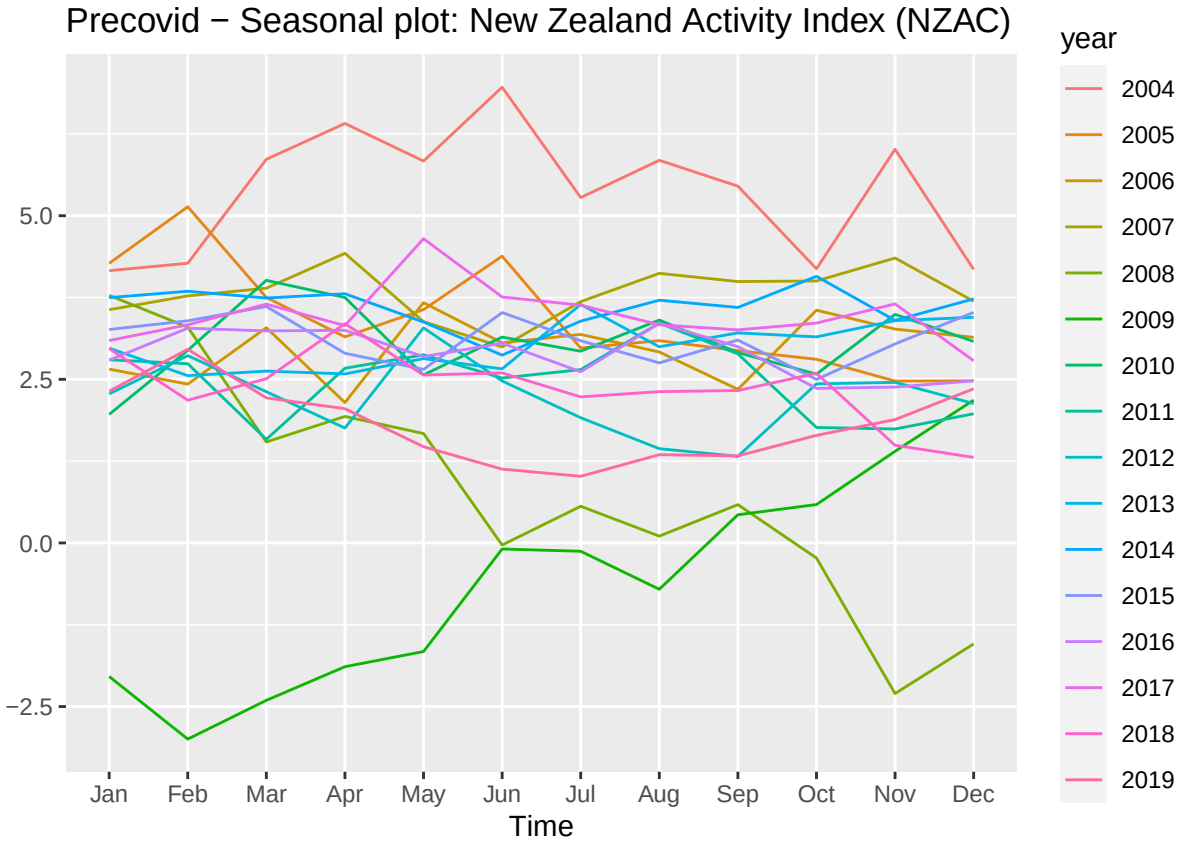

Seasonal plot: New jobs posted online
300
200
100
0
Jan
Feb
Mar
Apr
May
Jun
Jul
Aug
Sep
Oct
Nov
Dec
Time
2004
2005
2006
2007
2008
2009
2010
2011
2012
2013
2014
2015
2016
2017
2018
2019
2020
2021
2022

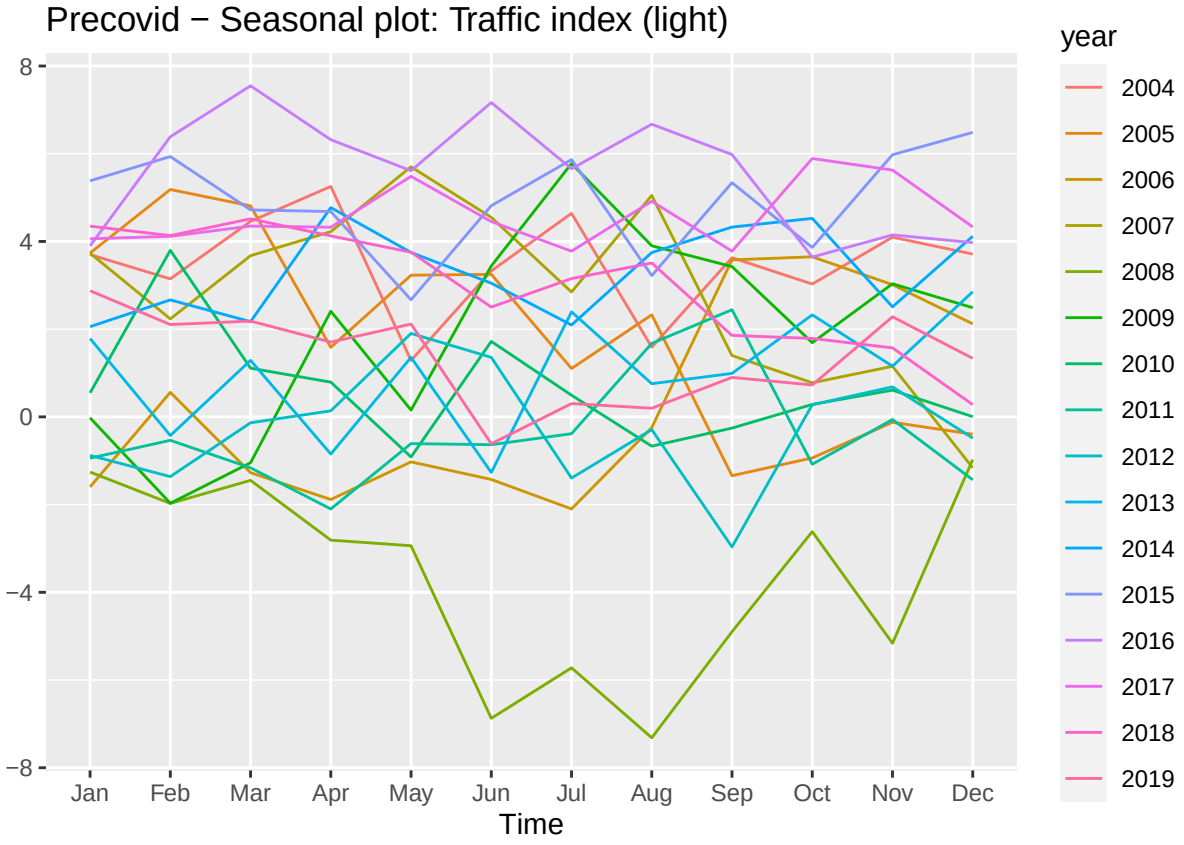

Seasonal plot: Manufacturing index
60
50
40
30
Jan
Feb
Mar
Apr
May
Jun
Jul
Aug
Sep
Oct
Nov
Dec
Time
2004
2005
2006
2007
2008
2009
2010
2011
2012
2013
2014
2015
2016
2017
2018
2019
2020
2021
2022

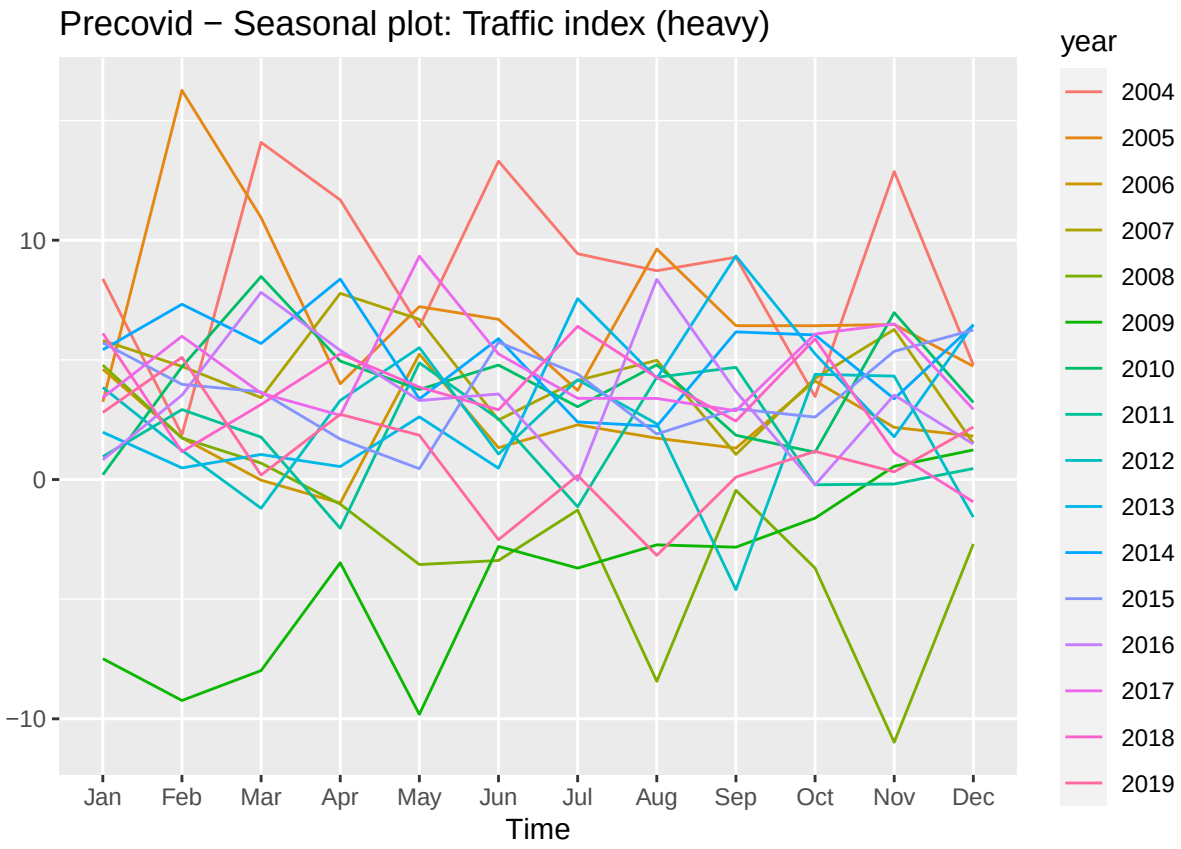

Seasonal plot: Electricity grid demand
10
0
−10
Jan
Feb
Mar
Apr
May
Jun
Jul
Aug
Sep
Oct
Nov
Dec
Time
2004
2005
2006
2007
2008
2009
2010
2011
2012
2013
2014
2015
2016
2017
2018
2019
2020
2021
2022

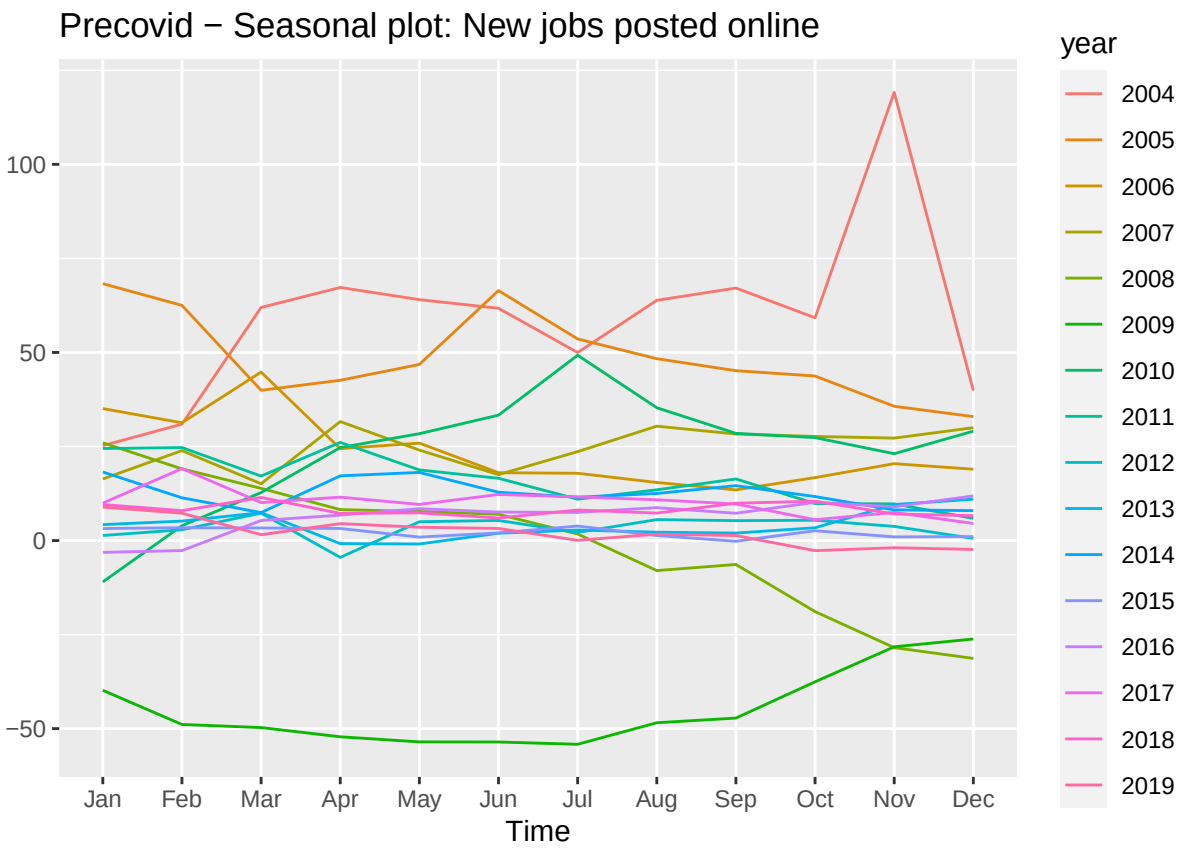

Seasonal plot: Card transaction spend
100
50
0
−50
Jan
Feb
Mar
Apr
May
Jun
Jul
Aug
Sep
Oct
Nov
Dec
Time
2004
2005
2006
2007
2008
2009
2010
2011
2012
2013
2014
2015
2016
2017
2018
2019
2020
2021
2022

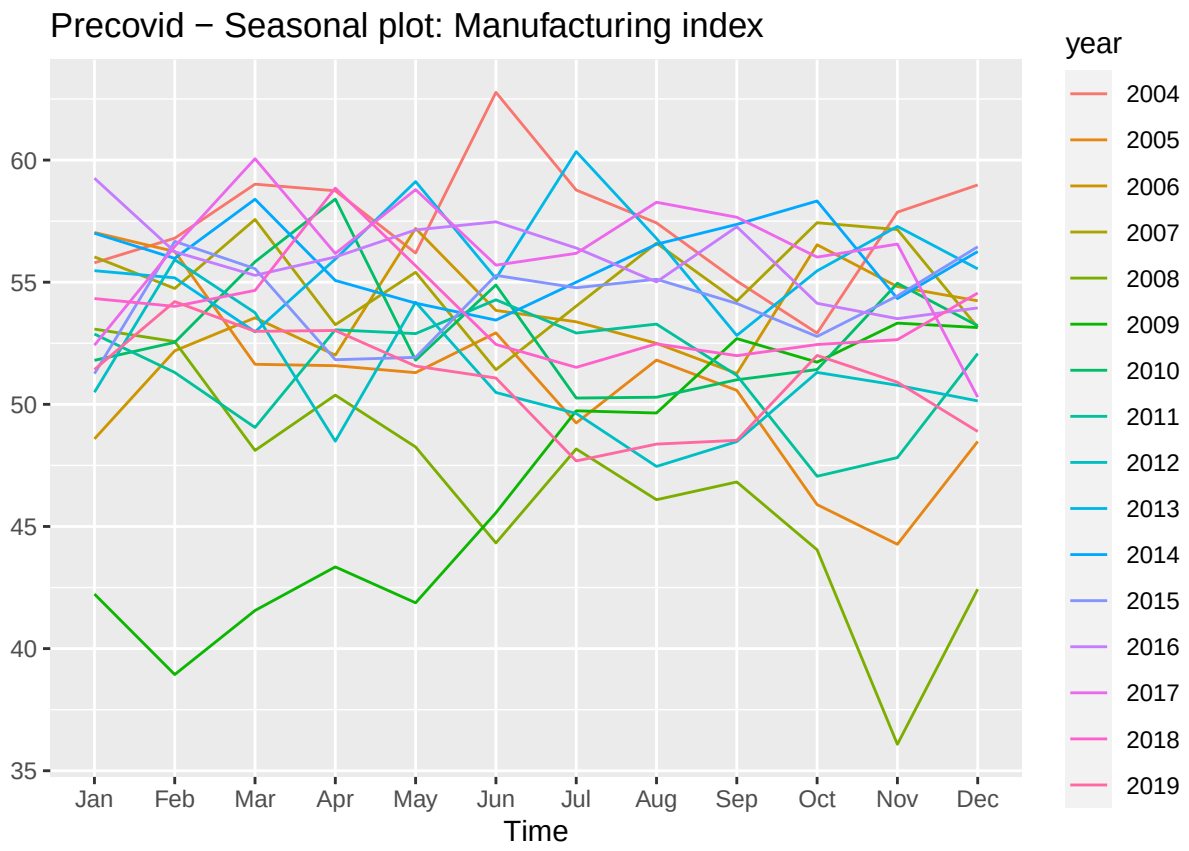
Plot seasonal pre-Covid

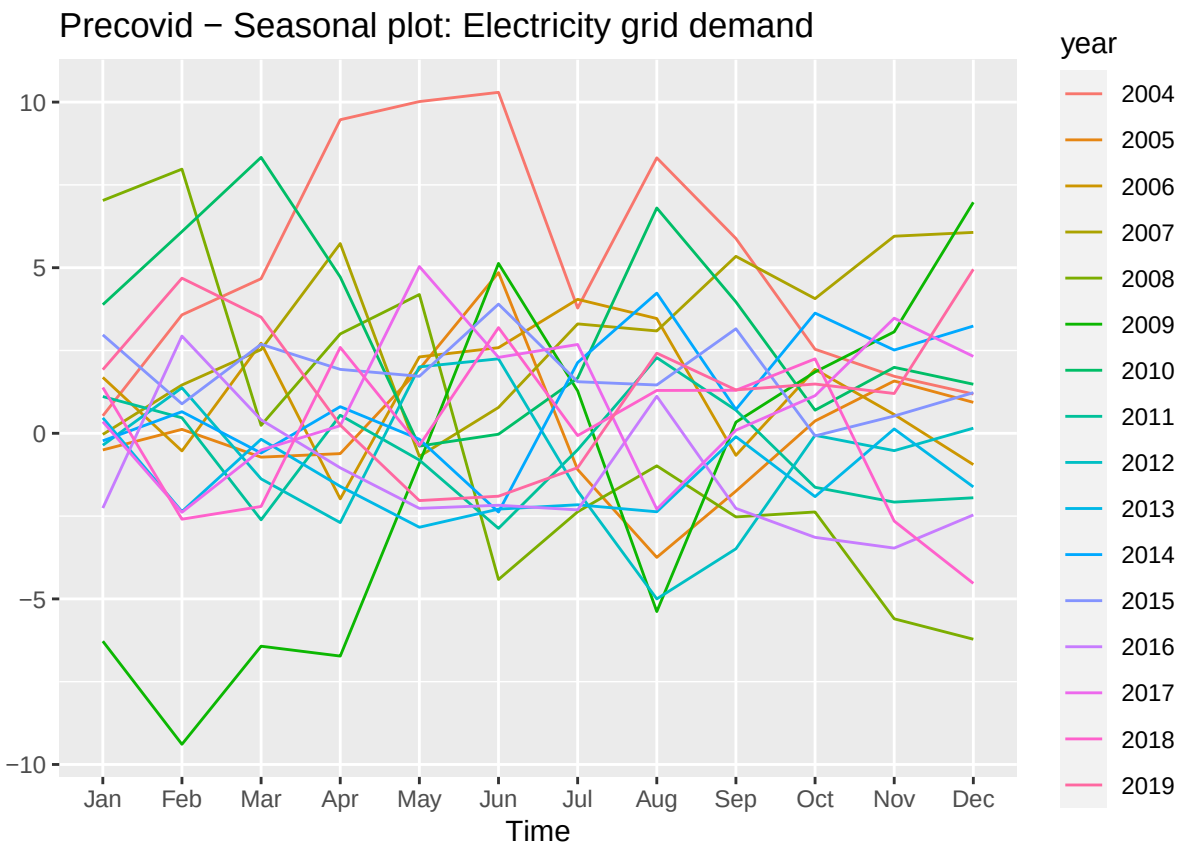


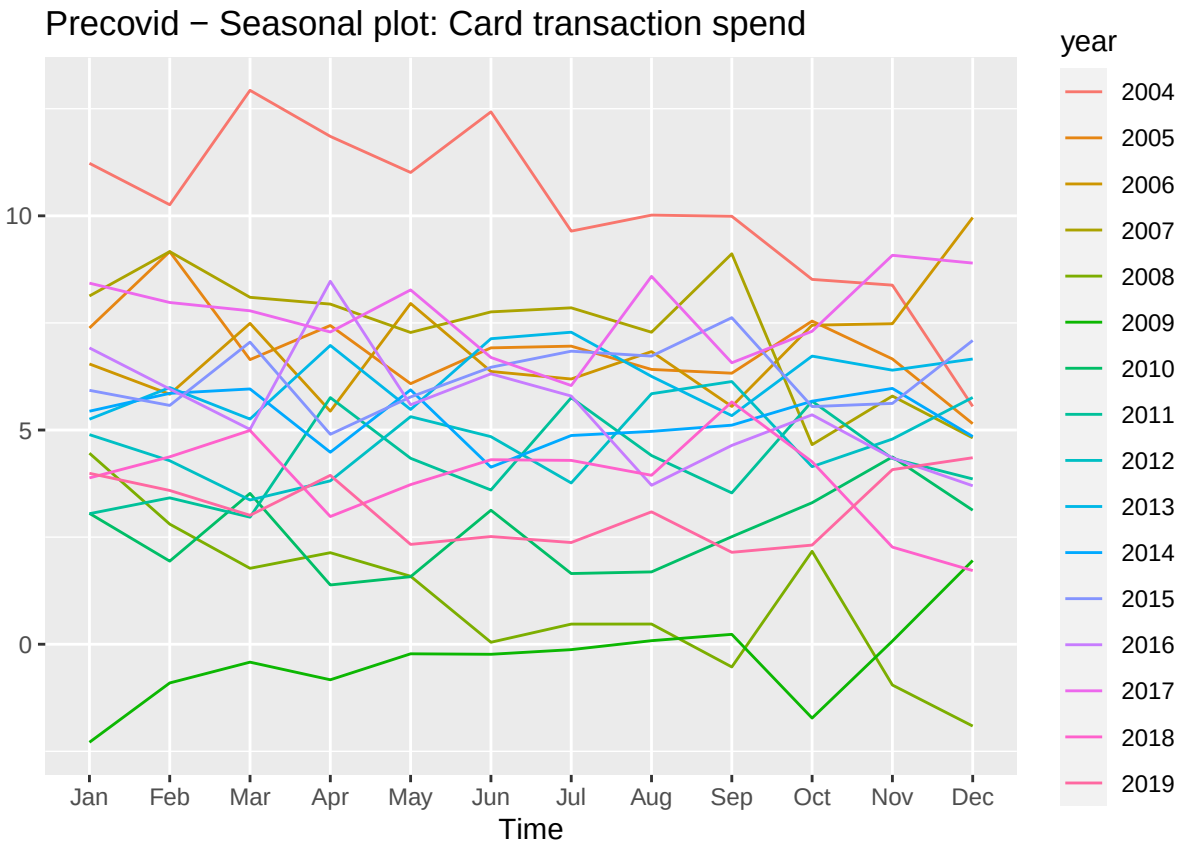



Precovid − Seasonal plot: Traffic index (heavy)
10
0
−10
Jan
Feb
Mar
Apr
May
Jun
Jul
Aug
Sep
Oct
Nov
Dec
Time
year
2004
2005
2006
2007
2008
2009
2010
2011
2012
2013
2014
2015
2016
2017
2018
2019


Precovid − Seasonal plot: New jobs posted online
100
50
0
−50
Jan
Feb
Mar
Apr
May
Jun
Jul
Aug
Sep
Oct
Nov
Dec
Time
year
2004
2005
2006
2007
2008
2009
2010
2011
2012
2013
2014
2015
2016
2017
2018
2019


Precovid − Seasonal plot: Manufacturing index
60
55
50
45
40
35
Jan
Feb
Mar
Apr
May
Jun
Jul
Aug
Sep
Oct
Nov
Dec
Time
year
2004
2005
2006
2007
2008
2009
2010
2011
2012
2013
2014
2015
2016
2017
2018
2019


Precovid − Seasonal plot: Electricity grid demand
10
5
0
−5
−10
Jan
Feb
Mar
Apr
May
Jun
Jul
Aug
Sep
Oct
Nov
Dec
Time
year
2004
2005
2006
2007
2008
2009
2010
2011
2012
2013
2014
2015
2016
2017
2018
2019


Precovid − Seasonal plot: Card transaction spend
10
5
0
Jan
Feb
Mar
Apr
May
Jun
Jul
Aug
Sep
Oct
Nov
Dec
Time
year
2004
2005
2006
2007
2008
2009
2010
2011
2012
2013
2014
2015
2016
2017
2018
2019