Seasonality Analysis

Nicholas Bradley

Seasonality is one of the main components of time series data and plays a pivotal role in forecasting process of future values of the series as it contains structural patterns. This analysis focuses on methods and approaches for identifiying and classifying seasonal patterns of a series. This draws on descriptive statistics, such as summary statistics and data visualisations. This analysis covers the following topics:

* Single and multiple seasonality patterns
* Descriptive statistic methods to identify seasonality patterns
* Data visualisation tools to explore and identify seasonality patterns

# Seasonality Types

A seasonal pattern exists in a time series whenever it is possible to tie a repeated event in the series with a specific frequency unit, for example, average temperature in New York during January. Therefore, there is a strong relationship between seasonal pattern and series frequency. Seasonality can be classified into one of two categories:

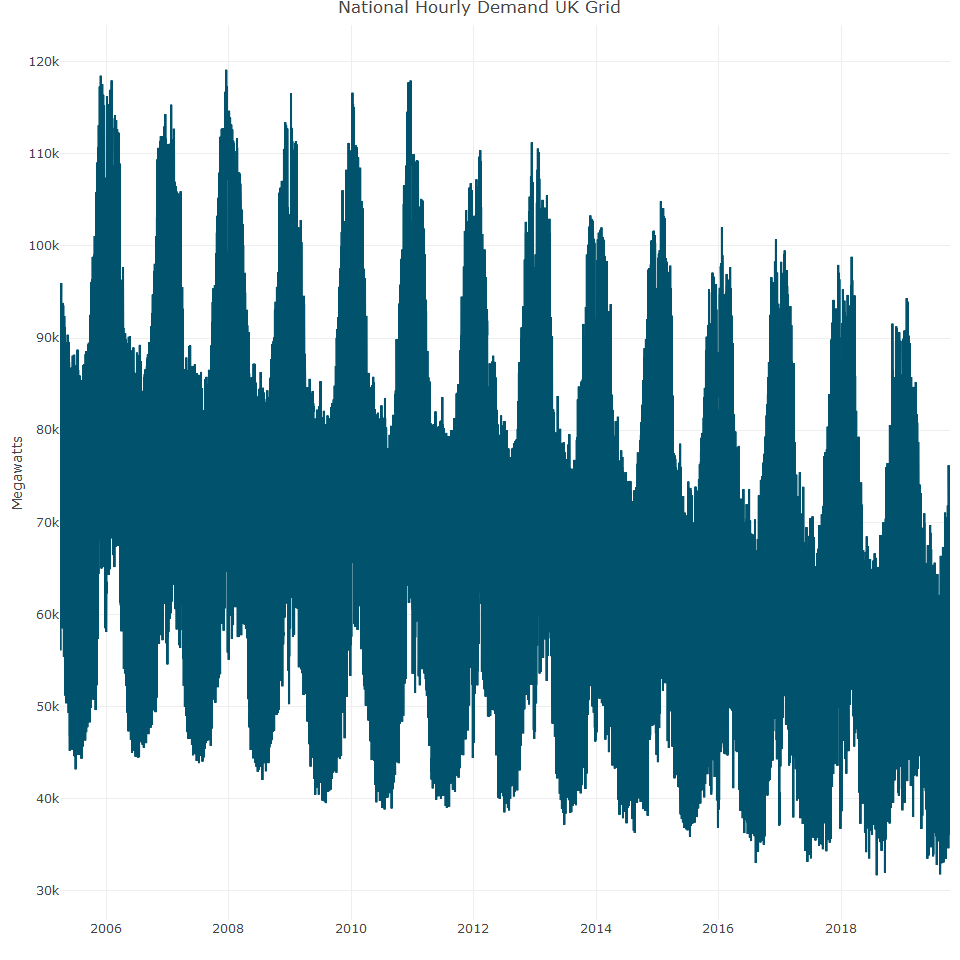
* **Single seasonal pattern:** occurs when there is only one dominant seasonal pattern
* **Multiple seasonal pattern:** occurs when there is more than one dominant seasonal pattern

Generally, multiple seasonal patterns are more likely to occur when the series has a high frequency, so for example, daily, hourly, half-hourly etc, as there are more options to aggregate the series to a lower frequency. A typical example is hourly demand for elecricity. Single seasonal patterns are more evident when the series frequency is lower, so for example, using data aggregated to monthly, quarterly etc.

We can use the UK Grid time series to illustrate seasonality. This dataset indicates the national demand for UK high voltage power transmission since 2011 in a half-hourly frequency.

## The UKgrid series is a xts object with 1 variable and 127296 observations  
## Frequency: hourly   
## Start time: 2005-04-01   
## End time: 2019-10-08 23:00:00

Here is the plot of the UK Grid time series:



In the plot, the national demand on the UK Grid national network has a clear seasonal pattern. This time series has 8,760 observations per cycle unit (24 x 365).

# Seasonal analysis with descriptive statistics

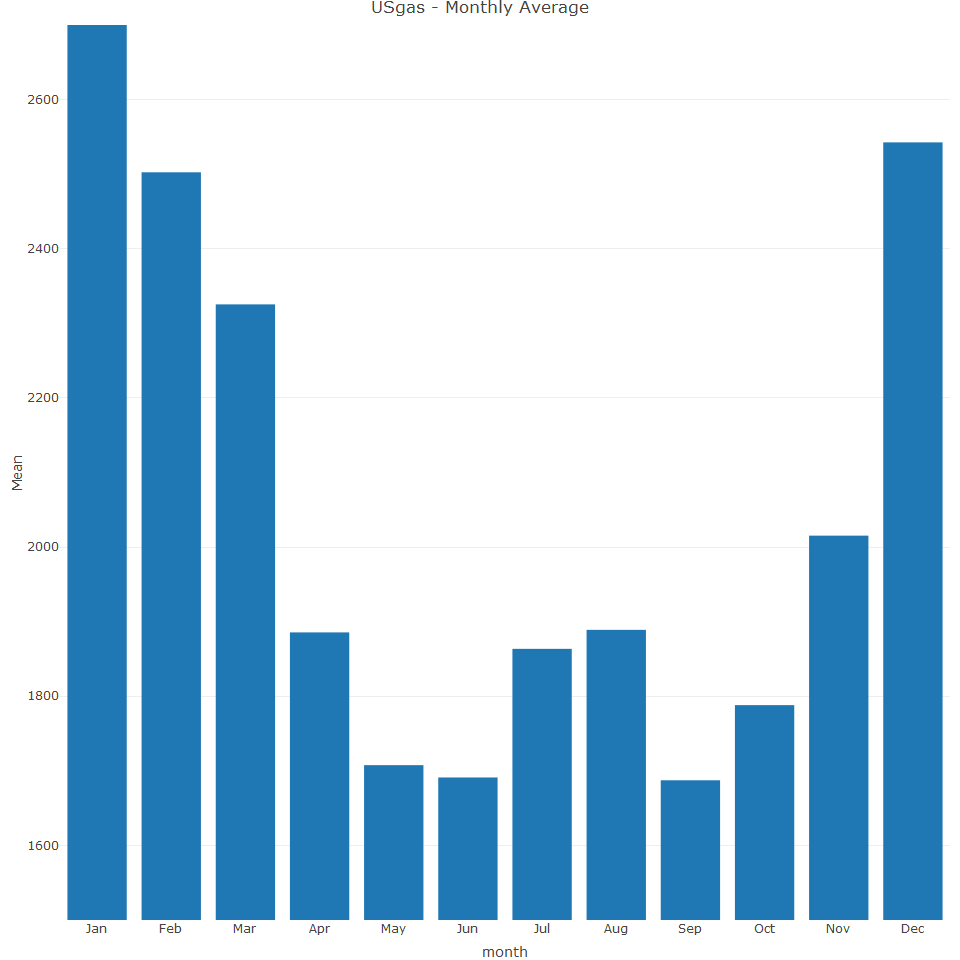
Descriptive statistics are a useful way to understand the data characteristics through summary statistics tables, and summary of the key statistical indicators, such as mean, median, quantile, standard deviation, box plots and bar charts etc.

## Summary Statistics Tables

## year month USgas  
## 1 2000 Jan 2510.5  
## 2 2000 Feb 2330.7  
## 3 2000 Mar 2050.6  
## 4 2000 Apr 1783.3  
## 5 2000 May 1632.9  
## 6 2000 Jun 1513.1

Next, we summarise and plot the data

## # A tibble: 12 × 3  
## month mean sd  
## <fct> <dbl> <dbl>  
## 1 Jan 2806. 309.  
## 2 Feb 2502. 223.  
## 3 Mar 2325. 241.  
## 4 Apr 1886. 175.  
## 5 May 1708. 194.  
## 6 Jun 1691. 216.  
## 7 Jul 1864. 261.  
## 8 Aug 1889. 237.  
## 9 Sep 1687. 242.  
## 10 Oct 1788. 260.  
## 11 Nov 2015. 284.  
## 12 Dec 2543. 278.



It is evident in the summary statistic table that on average, each month is different from the next consecutive month by its standard devation with the exception of two pairs: May/June and July/August. It is therefore possible to categorise some months as different from the rest, i.e. January, February, March and November.

Let’s now examine the UK Grid time series data

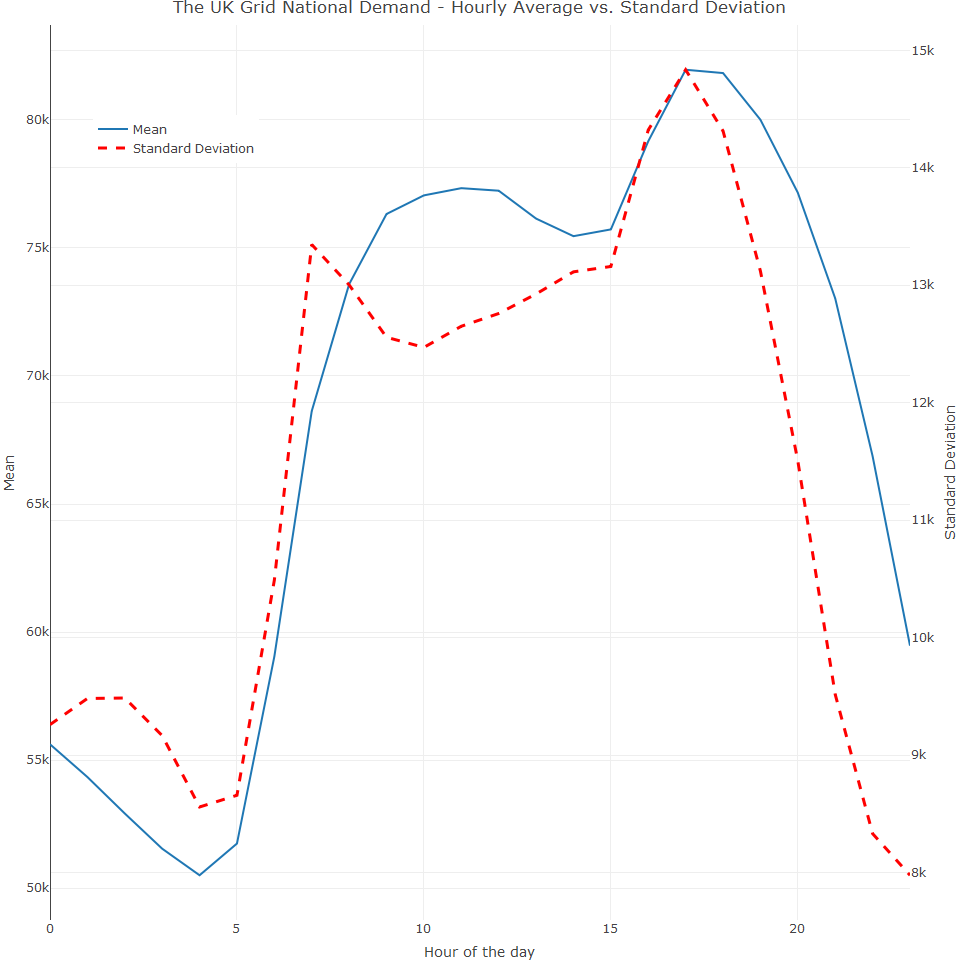
## 'data.frame': 127296 obs. of 2 variables:  
## $ time : POSIXct, format: "2005-04-01 00:00:00" "2005-04-01 01:00:00" ...  
## $ UKgrid: num 65080 68207 69172 66769 64660 ...

It is then necessary to create seasonal features based on the periods we wish to check, such as hours of the day, day of the week and month.

## time UKgrid hour weekday month  
## 1 2005-04-01 00:00:00 65080 0 Fri Apr  
## 2 2005-04-01 01:00:00 68207 1 Fri Apr  
## 3 2005-04-01 02:00:00 69172 2 Fri Apr  
## 4 2005-04-01 03:00:00 66769 3 Fri Apr  
## 5 2005-04-01 04:00:00 64660 4 Fri Apr  
## 6 2005-04-01 05:00:00 65209 5 Fri Apr

It is perhaps more efficient to start to explore the most granular layer of the series first as it could indicate direction of the aggregations and so we summarize the series by its hourly cycle

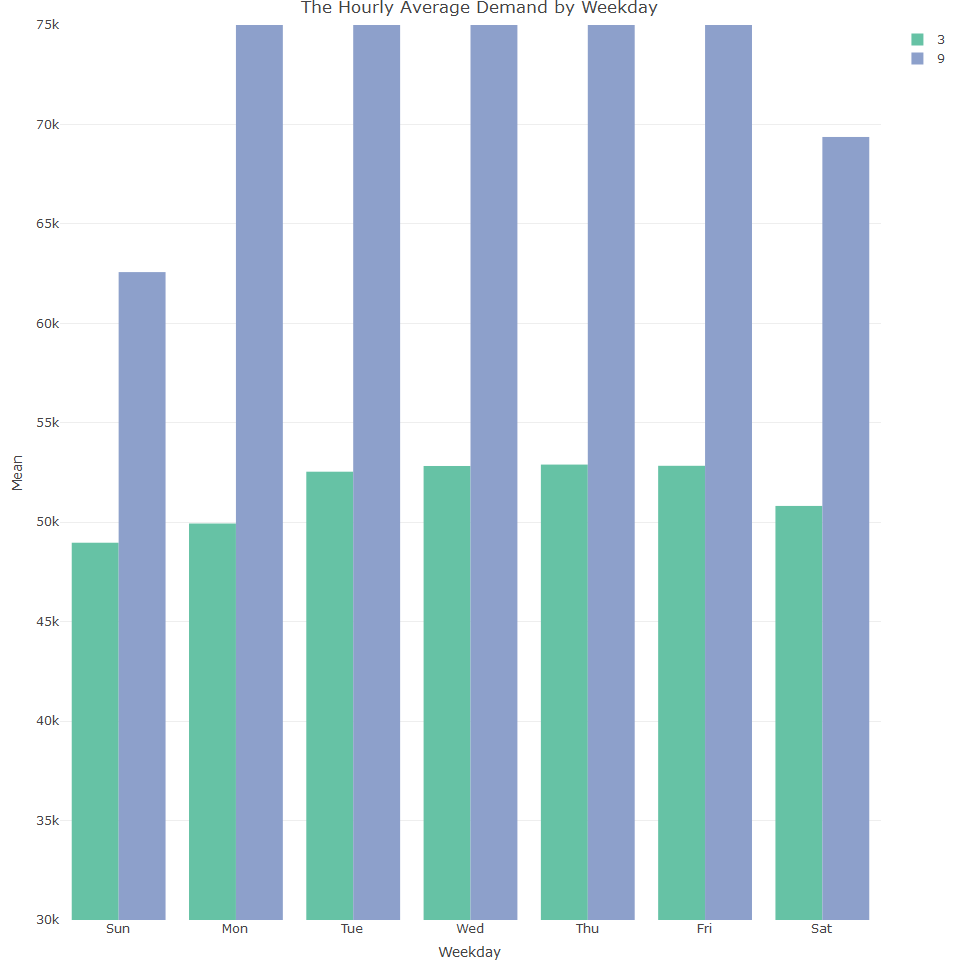
We can now plot the hourly mean and its standard deviation.



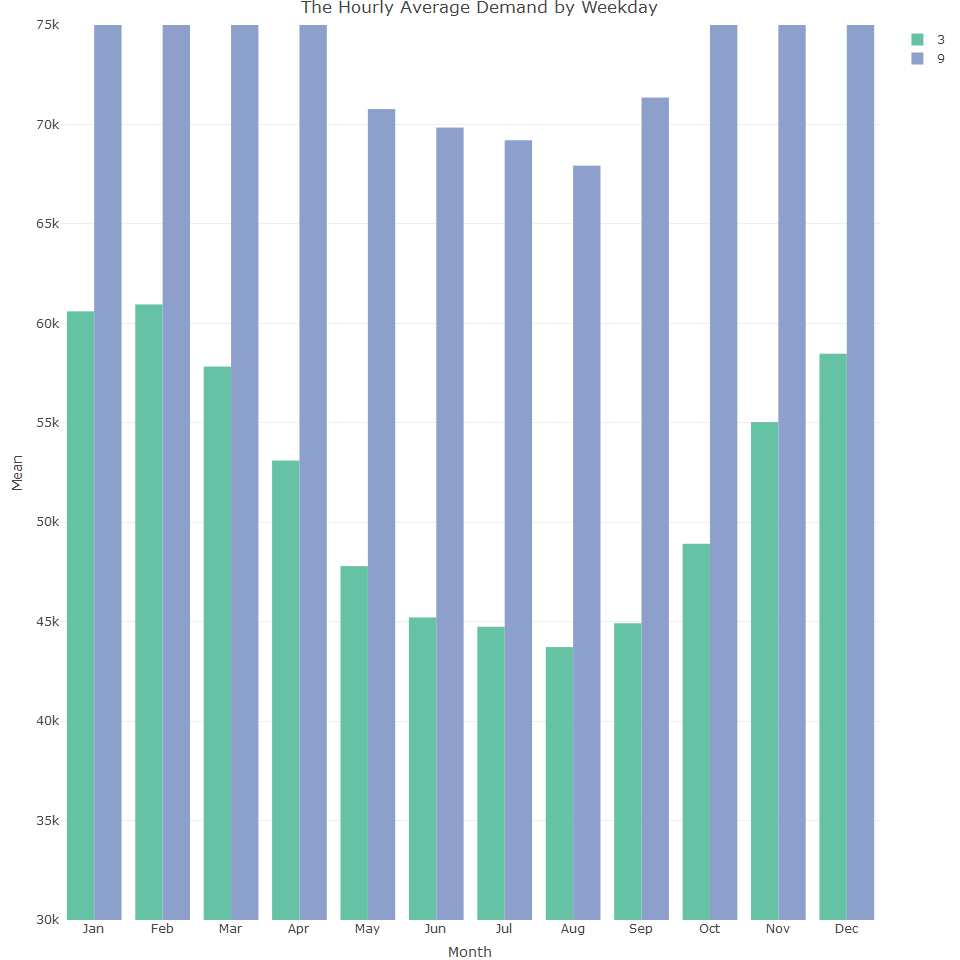
The following is evident in the summary statistic table:

* There is low demand during the nighttime (midnight and 6am) and high demand in the morning and early evening
* There is strong correlation between average demand and its standard deviation
* The relatively low standard deviation of the demand average during the nighttime could indicate strong sub-seasonal effect during those hours beside the hourly seasonality. This could make sense as those are normal sleep hours and therefore, on average, the demand is reasonably the same throughout the weekdays
* However, the high standard deviation throughout the high-demand hours could indicate that demand is distributed differently on different periodocity views (weekday or month)

To further investigate the last point, we can subset the series into two groups representing the demand in the middle of the night and the demand throughout the day (3am and 9am)



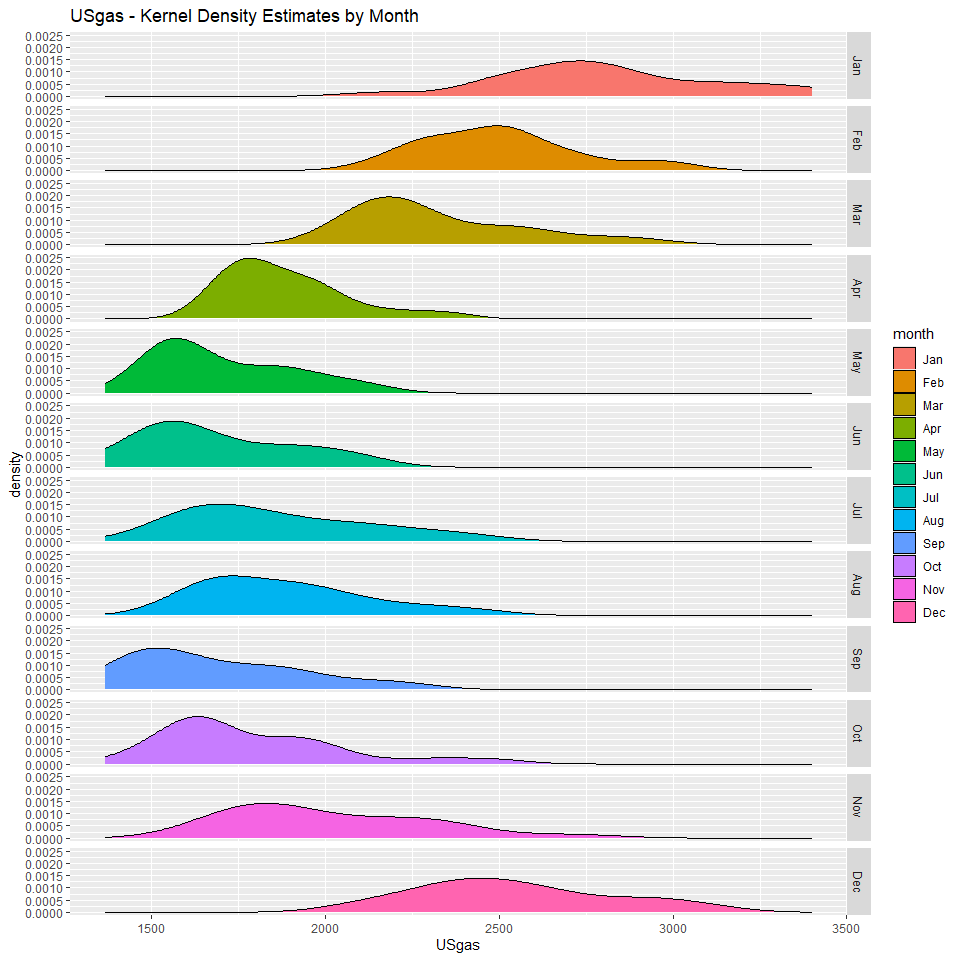
The bar chart above indicates that electricity demand at 3am is relatively stable throughout all the days of the week, with a slight difference between the average during the weekdays and the days in the weekend (about 2% difference). Yet, there is a significant difference between the weekday and the weekend demand at 9am. (demand on Monday is higher on average y 28% from the one on Sunday). We can now examine whether there is a monthly seasonal pattern by looking at the same hours but on a monthly basis.



The demand at 3am and 9am varies throught the months of the year and there is also significant change in demand at night compared to the weekday aggregation. This variation indicates existence of monthly seasonality

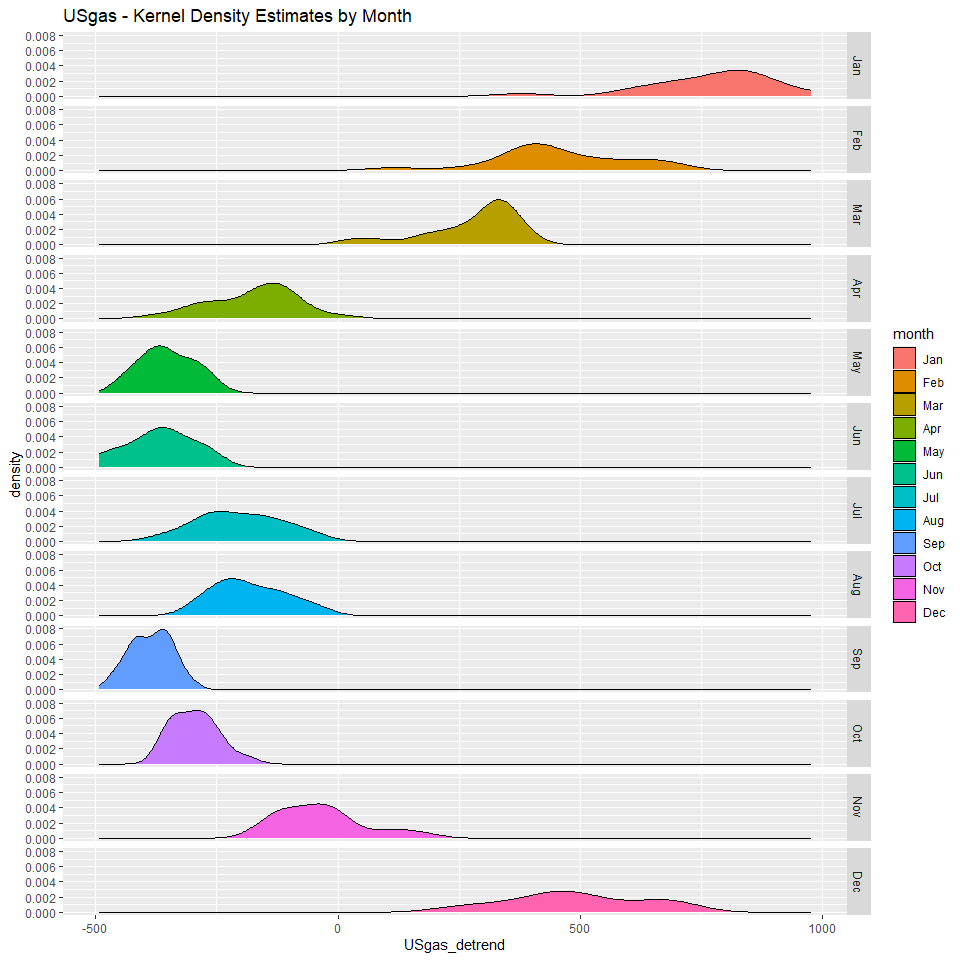
# Seasonal Analysis with density plots

It is also possible to plot the distribution of the frequency units by density plots, which enables the identification whether each frequency unit has a unique distribution that distinguishes itself from the rest of the units

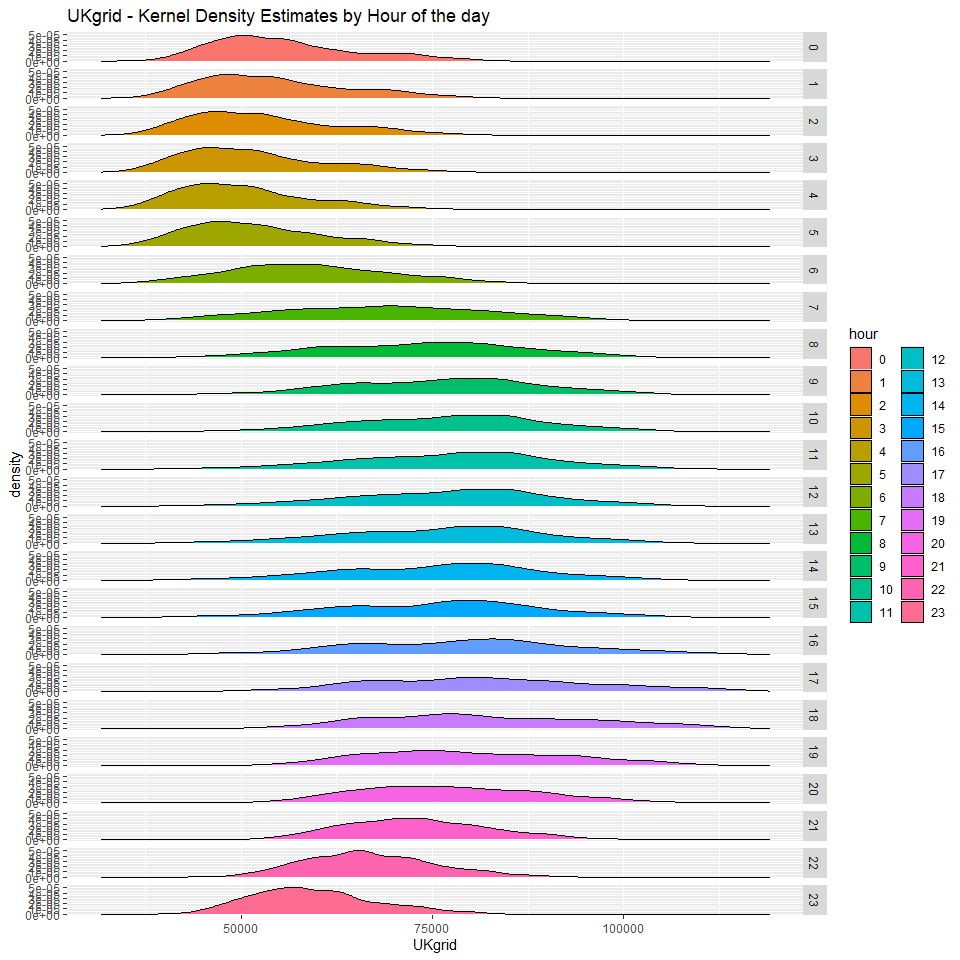


We can see some indication of a seasonal pattern as the density plots don’t overlap each other with the exception of May and June. Additionally, in some months, the distribution shape is flatter with long tails (November, December, January). This could be caused by volatility in some of the exogenous factors, for instance, a combination of weather patterns along with the elasticity or sensitivity of the series for changes in the weather.For example, in gas supply, there is higher elasticity during winter due to dependency of the heating systems for this resource.

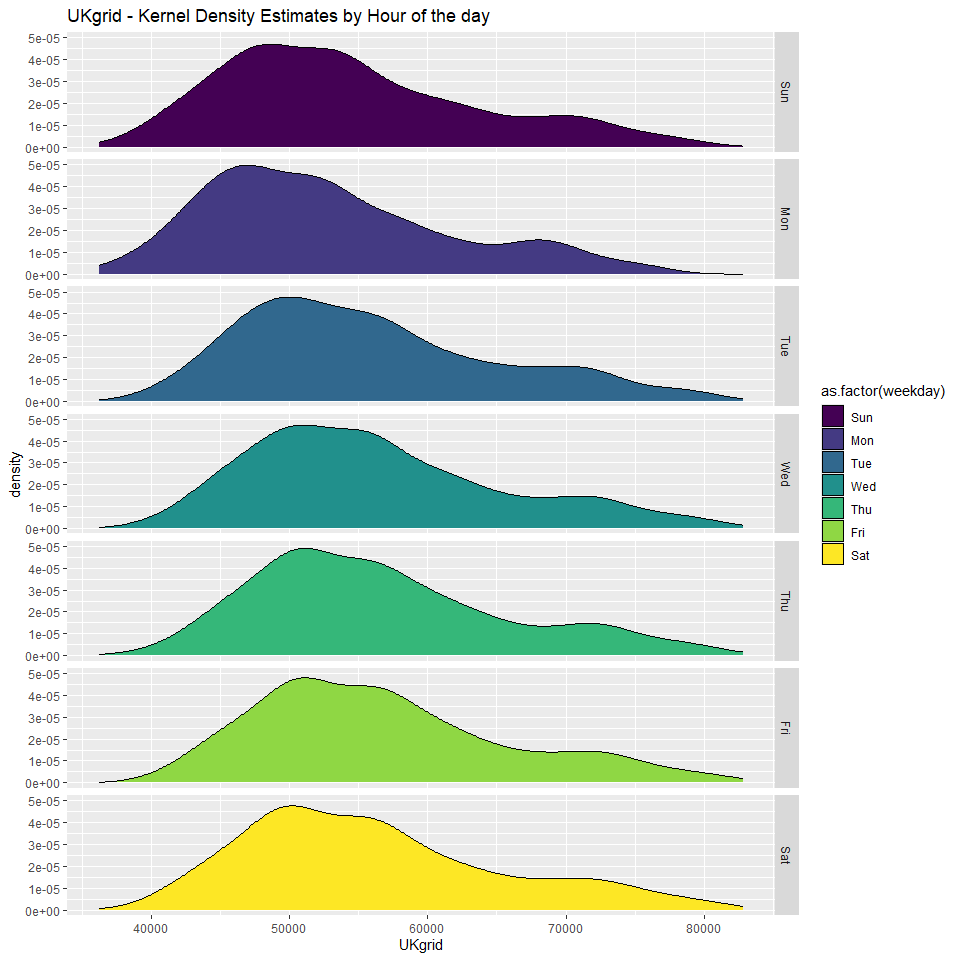
It is important not to forget about the trend effect or the growth year to year. Let’s detrend the US Gas series by computing the series trend and then subtract it from the series



The removal of the series trend sharpens the seasonal effect as the tail of the distribution becomes shorter. In other time series, indication of the seasonal trend may not be so clear before the removal of the trend and so, it is very important that the series trend is removed whenever the trend is non linear or when the distribution has a long tail. Where the distribution of most of the frequency units is flat with a long tail, it could indicate multiple seasonal patterns in the series. We can plot the UK grid series



The distribution of the net demand for electricity during nighttime is relatively stable (hence the non flat distribution with short tails as opposed to flat distribution with long tails during the day). If we subset one of the hours during the day, and plot its distribution by the day of the week, we should expect an overlapping during the nighttime and be able to distinguish between the distribution during the weekdays and weekend as opposed to just the weekday. For example, the following plot indicates the distribution during the weekdays is distinguished from the one at the weekend



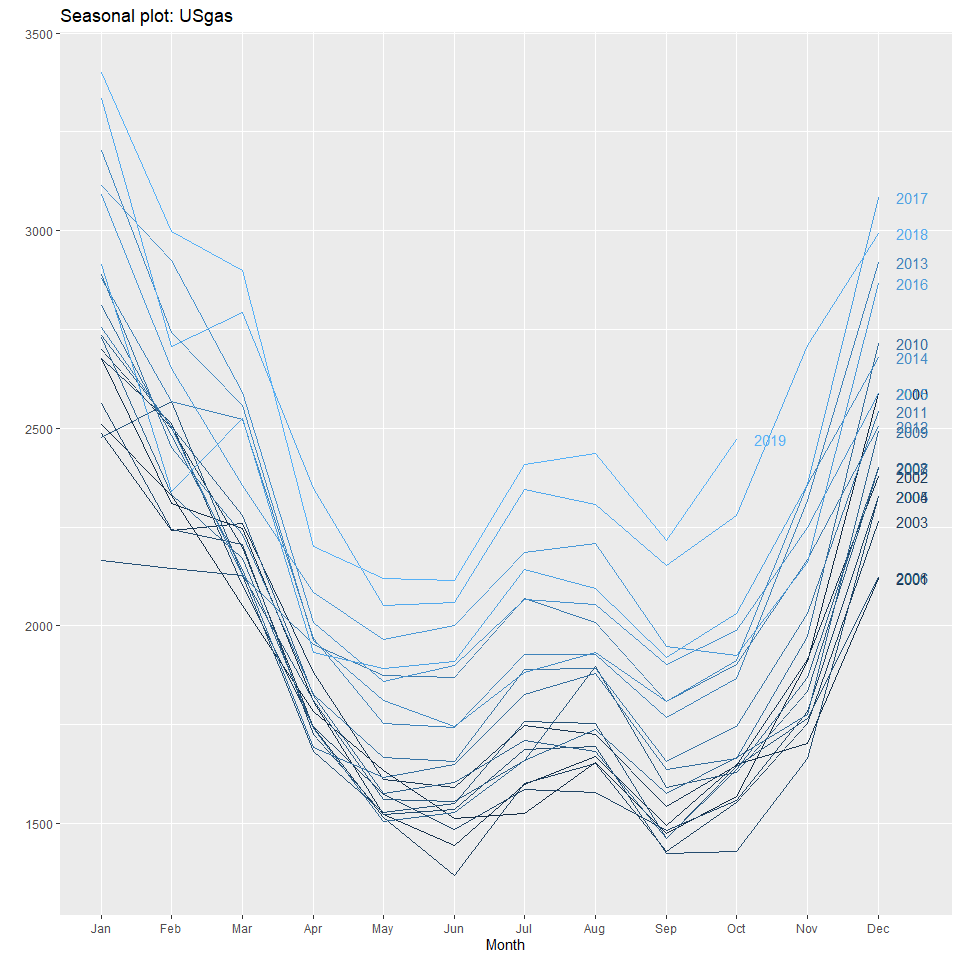
Lets now examine structural tools for seasonal analysis

## Structural tools for seasonal analysis

Data visualisation is important in time series and there are various tools in the forecast and TSstudio packages.

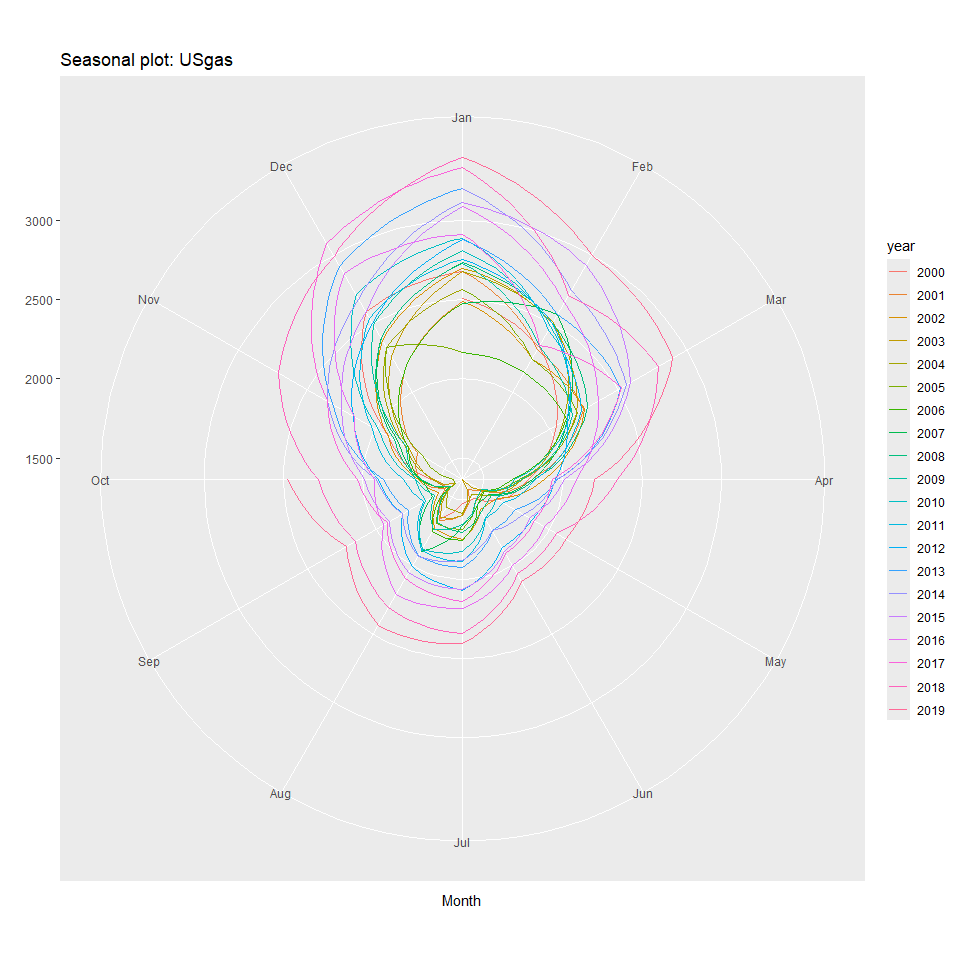
## Seasonal analysis with Forecast

We can use the forecast package to create a seasonal plot of the series by splitting and plotting each year as a seperate line.



We can see from this plot tha it has a strong repeated pattern, which indicates the existence of the monthly seasonal pattern. Additionally, as can be seen from the color scale and the years labels, the series is growing from year to year.

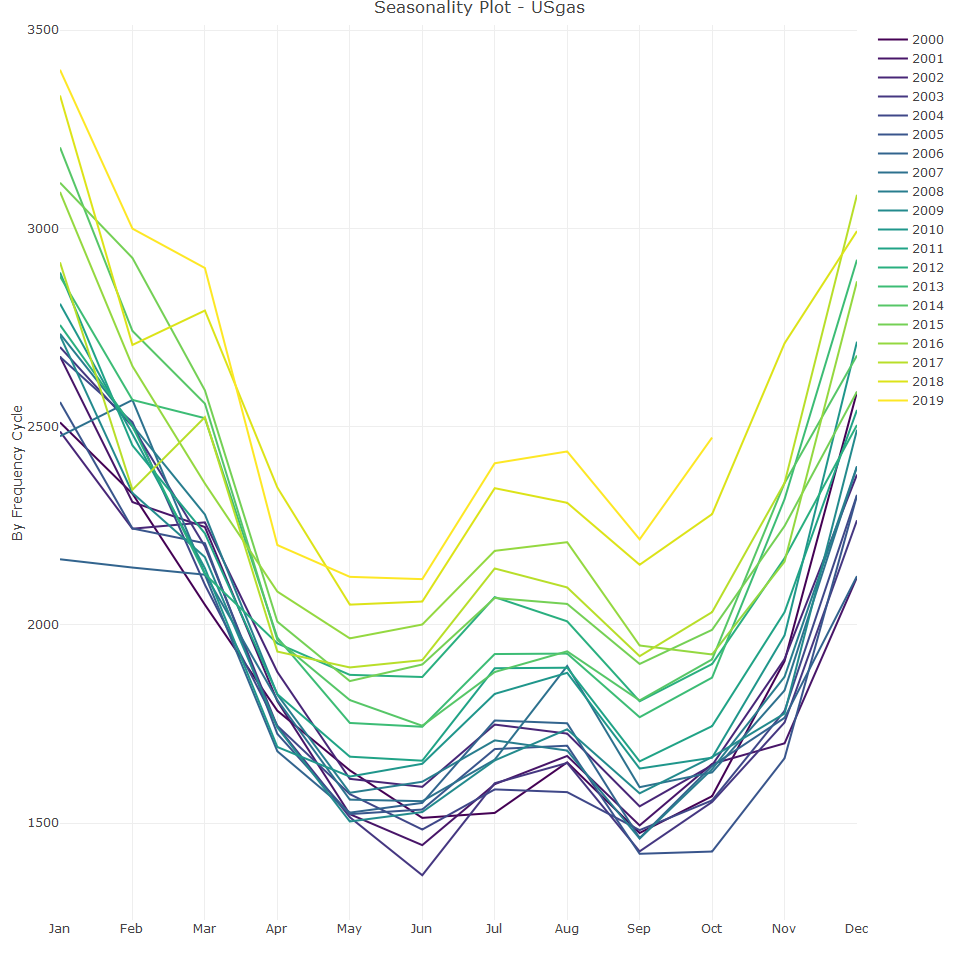
The polar plot is another way to visualise a series by its frequency, which spreads the frequency units across 360 degrees, depending on the number of frequency units (here it is 360/frequency), where the distance from the polar center represents the magnitude of the observations.



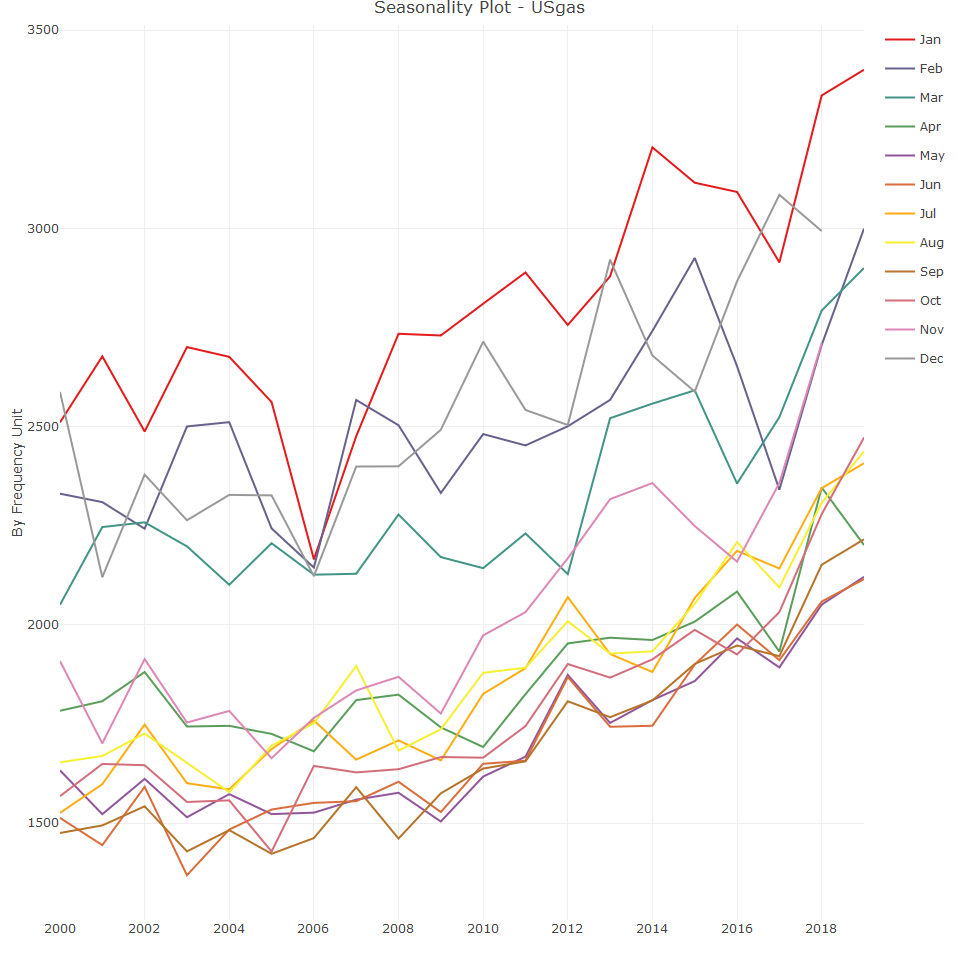
We can see repeated seasonal patterns along with a year to year growth or trend.

## Seasonal Analysis with the TSstudio package

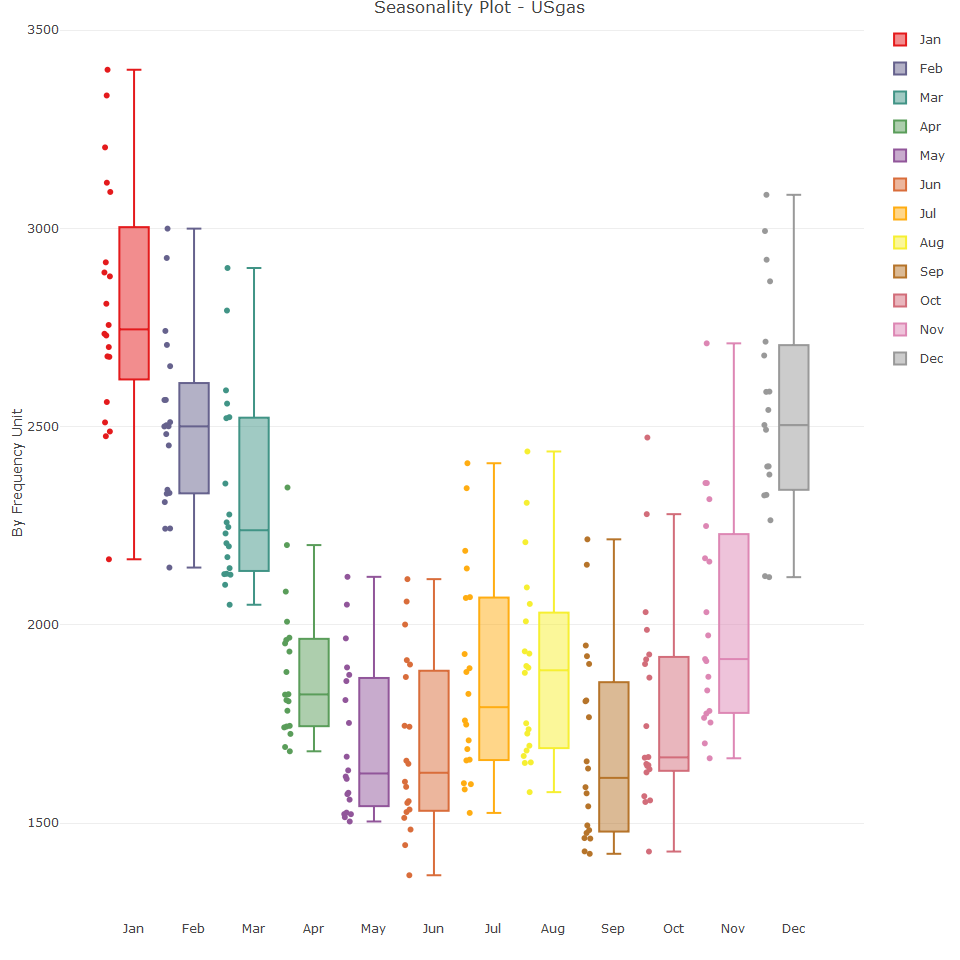
This package provides various interactive data visualisation options.



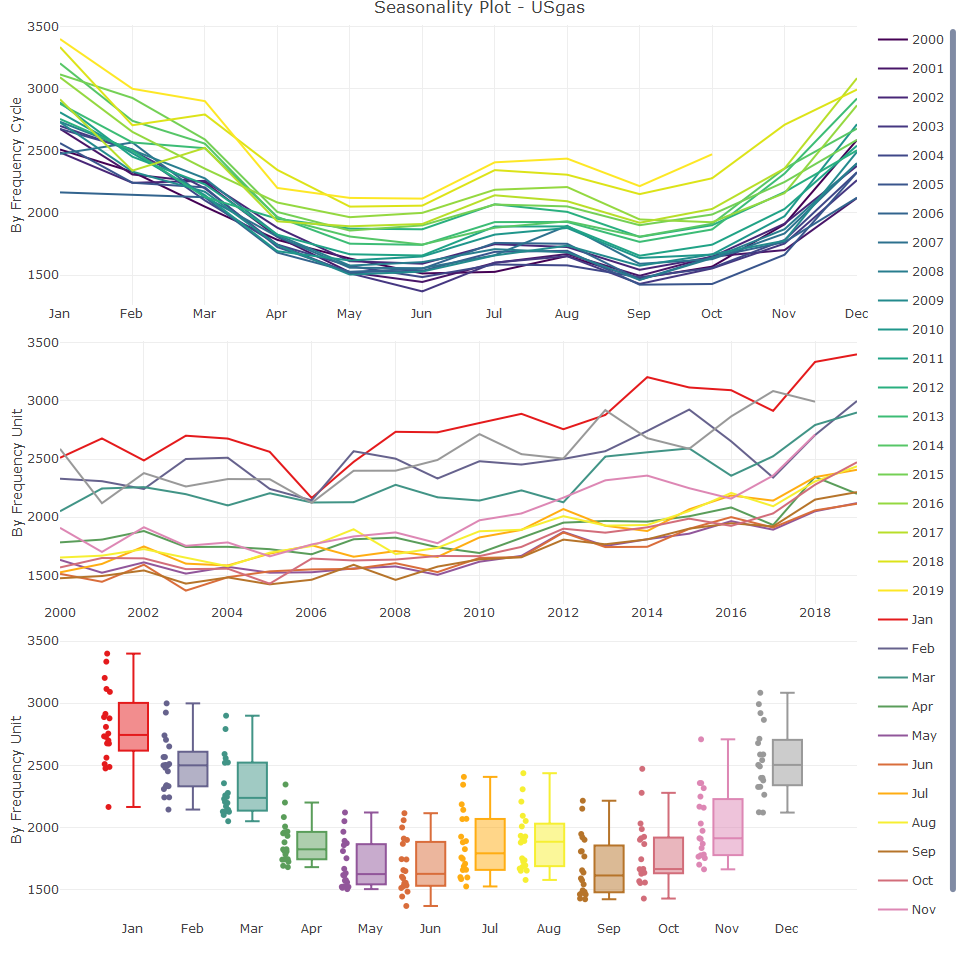
The cycle option plots the series frequency units over time in chronological order; for example, all the observations that occurred during January over a certain amount of time for a monthly series. This enables the researcher to identify seasonal pattern without detrending the series. For instance, the following plot indicates that, despite the growth year on year, in most cases, then order of the months, from higher to lower, remains the same



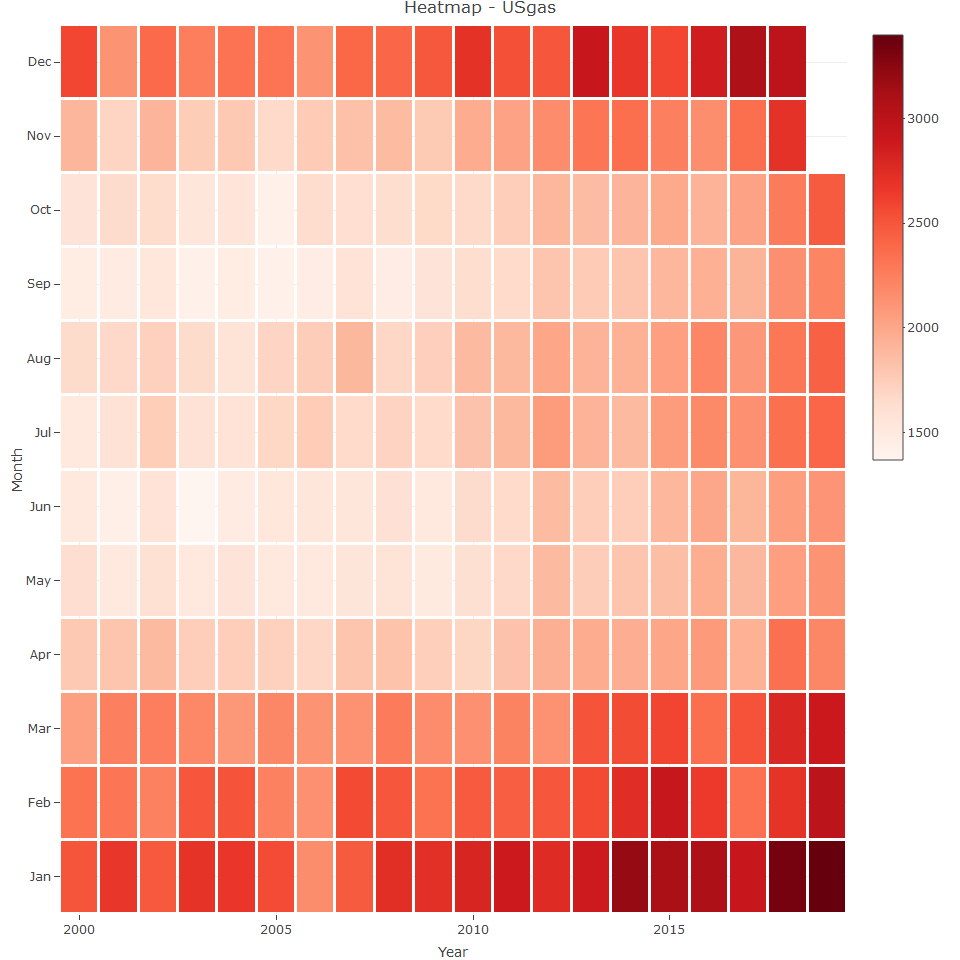
The box plot is also very informative as it provides the range and quartile representations of the observations of each frequency unit. However, some of the oscillation in the box plot representation may occur as results of the series trend and therefore, it is a good idea to use a combination of different plots.



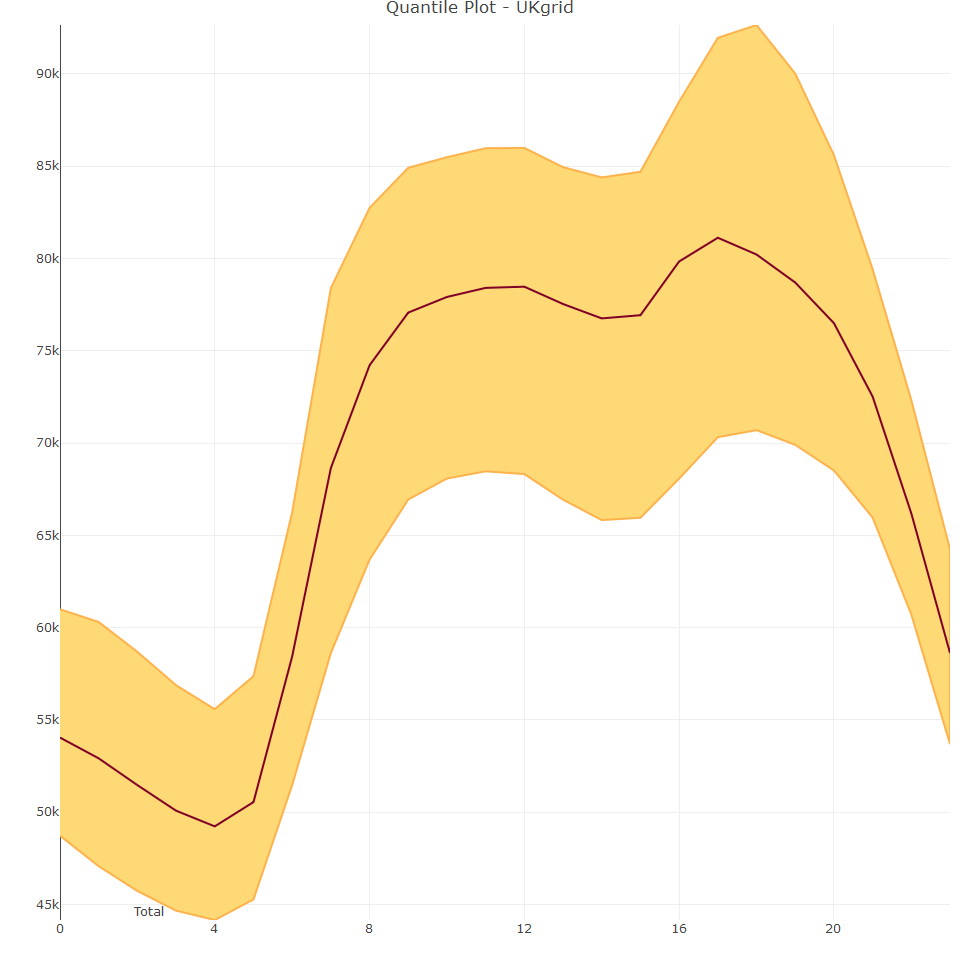
Finally, it is possible to show all the plot types next to each other, to gain a more complete understanding



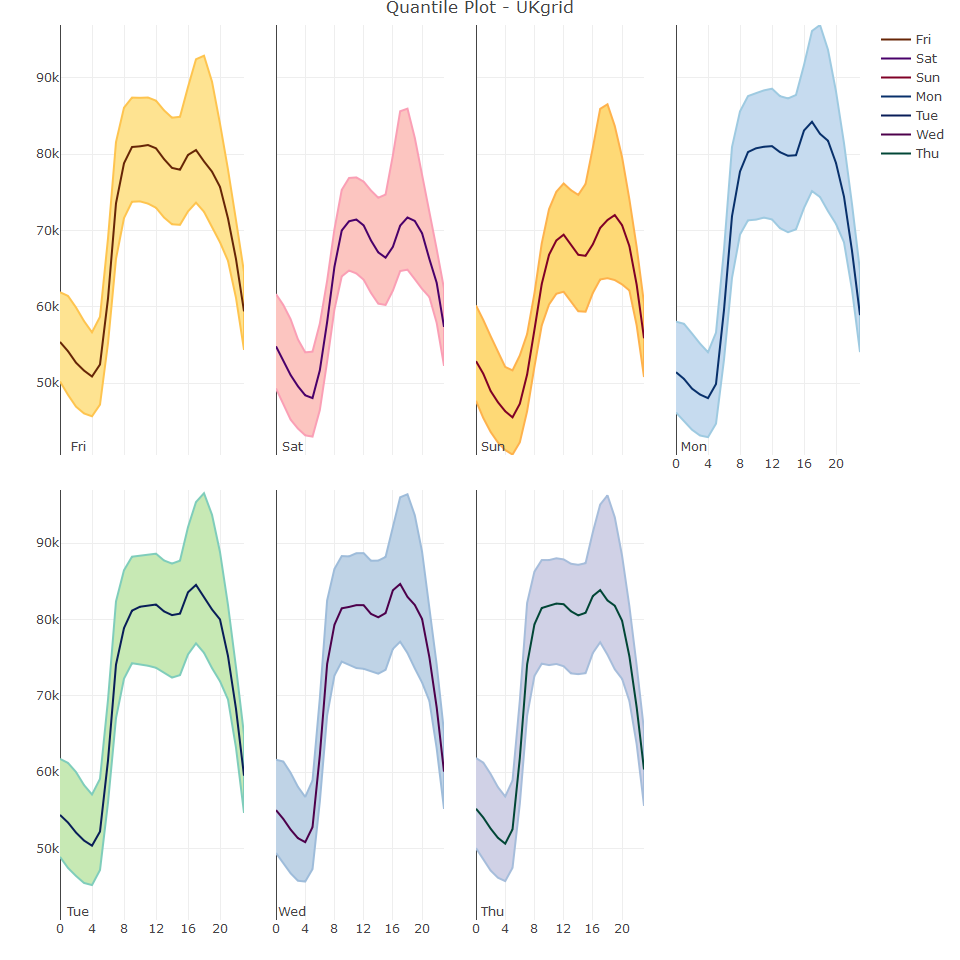
A heatmap is another good way to visualise a time series and to explore seasonality. In the below plot, the magnitude of each observation is represented by colour scale. The darker the colour, the the higher the observations value with respect to the overall values of the series



There are also quantile plots, which are typically used on high frequency time series data such as hourly or daily. A quantile plot of the series frequency units is given, where the middle line represents the median and the lower and upper lines represent the 25th and 75th percentiles. For instance, we can visualise the 24 cycles of the UK grid.



The period argument enables the researcher to investigate whether the seasonal patterns change when using a different subset of time. For example, we plot 24 hour cycle of the UK grid series by the day of the week by setting the period argument to weekdays



The electricity demand during the daytime is higher throughout weekdays than weekends. We can also examine the 24 hour cycle by month:

