Introduction

Allan

2023-06-14

Contents

1 Exectutive Summary

2 Introduction

2.1 Electricity

2.2 Swiss Grid

2.3 The Network

2.4 Overview of the network

2.5 My questions

3 Preliminary Data Anylsis

3.1 Overview of the data

First, we have data from Swissgrid regarding the energy network :

- Date : A file per year since 2009
- Time : Timestamp of 15min or 1hour (only overall for 1h)
- Total cons for end user only : are not included : grid losses, energy consumed for power plant's own requirements or to drive the pumps in pumped storage hydro power plant.
- Total prod/cons for Switzerland : everything consumed/fed in the network
- Secundary control : positive and negative energy within 15min
- Tertiary control : positive and negative energy after the first 15min
- Vertical load Swiss transmission grid : will not be used
- Net outflow of the Swiss transmission grid : will not be used
- Grid feed-in Swiss transmission grid : will not be used
- Control energy prices : average price in CHF for the last 15min of control energy (rounded to 2 decimals) for second and tertiary control

- Cross border exchange : energy exchanged with bordered country (Austria, Germany, France, Italy)
- Import/export/transit : Transit is not included in the cross border exchange
- Cantons : Details of prod/cons per cantons (starts after 2015) grouped in 19 cantons
- Foreign territories : prod and cons for regions within the control zone of Switzerland but do not belong to its territory

We have a total of 65 variables with a total of 487872 observations.

Second, we got data from gadm providing maps and spatial data for many countries on 4 different levels. In our project, we will use the data for Switzerland on a Canton level.

3.2 Data limitations

As said, we only have data available for each cantons since 2015. Also, due to the density of each cantons, certain areas have been grouped together, meaning we do not have the detail for every cantons. (26 cantons grouped into 19 zones)

3.3 Data cleaning and pre-processing

Due to the importance of the number of variables, we have decided to create sub-dataset for every type of Data :

- Overall
- Cantons
- Borders
- Foreign
- Price

Every DataSet had been transformed into a tsibble, a new data structure that help and support with temporal data. One should look at this paper for further references. We used the timestamp of 15min to get a better understanding of the data and help us building stronger models and forecast.

All the different cantons where set as variable (horizontal), in order to perform the anaylsis we needed to transform our Data-set in a vertical shape.

Here is quick overlook of our dataset :

Dataset	No. observations		Name of the DS
Initial dataset	277'536	Combined all SwissGris's files from 2015	General_df
Monthly data	277'536	Monthly version of General_DF	General_dfM
Cantons'Data	4'995'648	Combined Data for Cantons from 2015	Canton_df_long
Swiss map's data	123,156	Contain data to map Switzerland	gadmCHE1

A new tidy data structure to support exploration and modeling of temporal data

4 EDA

4.1 Overall consumption and production for the country

4.1.1 Quick visualisation

Now that we have better view of the network and where the data come from, let's dive int and see exactly what kind of data we will use. You can see below the first 10 rows of the general_df data containing all the general information at the country level :

time	end_users_cons	energy_prod	energy_cons	pos_second	neg_second	pos_tertiary	neg_tertiary
2015-01-01 00:15:00	1790683	1697772	1922526	37500	0	0	0
2015-01-01 00:30:00	1777126	1686388	1907138	22200	0	0	0
2015-01-01 00:45:00	1807976	1724777	1940146	36100	0	0	0
2015-01-01 01:00:00	1784944	1690007	1918599	16400	0	0	0
2015-01-01 01:15:00	1813997	1681642	1954830	52700	0	0	0

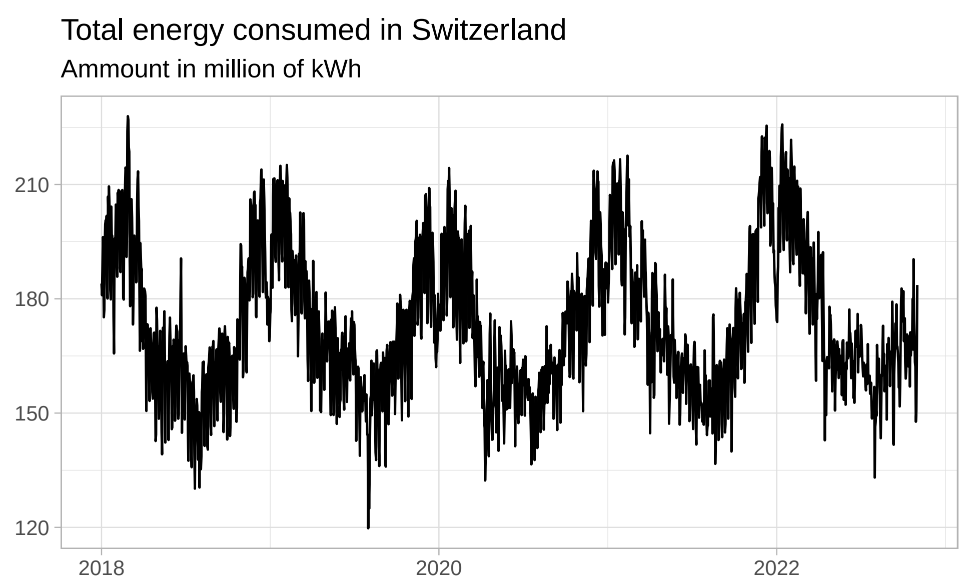
In the following exploration data analysis, we will only take the `energy_cons` and `energy_prod` as total of consumption and production. A important component when working with times series is the missing value. Indeed, many packages used in R can not deal NA. It is also the case for the `tsible` package which we will use throughout this paper. For further information on this new tidy data structure, one can refer to this article on the subject.

After a quick check, we see that we have a total of missing value of :

```
sum(is.na(general_df))
#> [1] 0
```

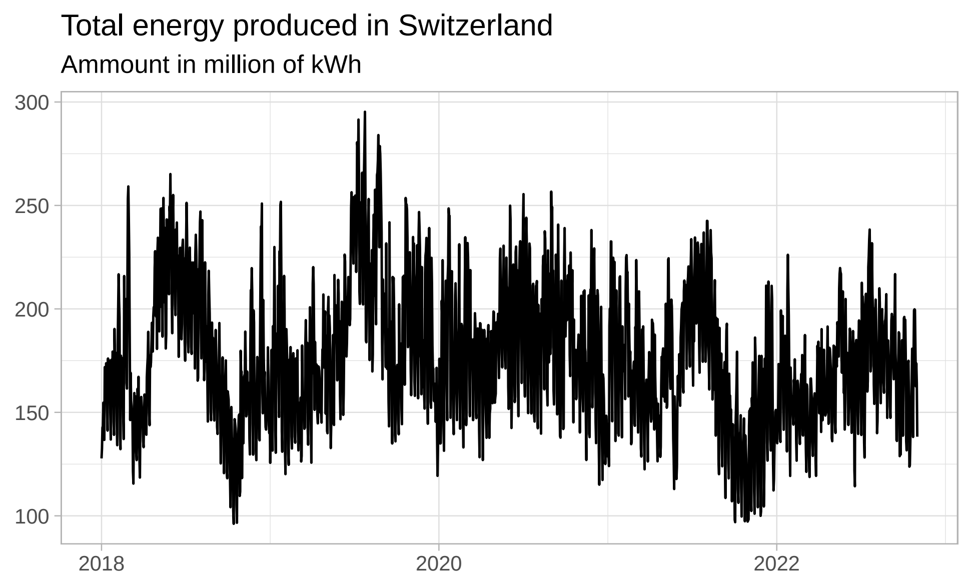
We now know that no data is missing and that we have an observation for every 15 minutes from January 1, 2015 to the end of 2022. In the subsequent analysis, we will also transform our data into hourly (`hourly`), daily (`date`) and monthly (`month`) observations to obtain different insight.

We can now have a first look on the Daily Consumption :



We can already see a strong yearly seasonality effect : for each year, we see peaks at the end/begging of the year (winter season) with big decrease during summer. We can also observe numerous variations over shorter periods, indicating possible seasonality on several levels. Giving the nature of the electricity, this results makes perfect sense : we can expect variation during day and night, weekday and weekend, winter and summer. We will have to work with different time period to understand each of these seasonal effects. Finally, the chart shows high volatility with several peaks each year. Considering them as outliers would not make sense since they occur very often, we will instead try to find out if we can explain a pattern of increase/decrease in consumption on the total from the data at hand.

Let's now have a look on the Daily Production to see if we get similar results :

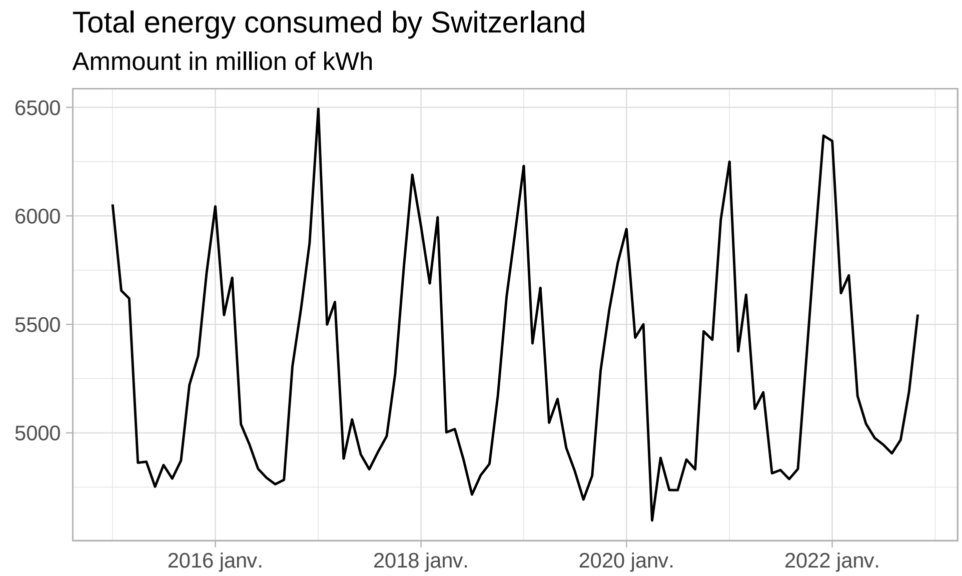


Exactly as for consumption, we can see a strong seasonality over time, also operating on different time period. One difference to note is the possible downward trend after mid-2019. They yearly seasonality pattern seems also less consistent over the years. Clearly, if we can affirm correlation between consumption and production, we already see they behave differently. We will try to understand and explain where the difference come from. Finally, we see the same high volatility with several peaks each year we had for consumption.

4.1.2 Different levels of seasonality

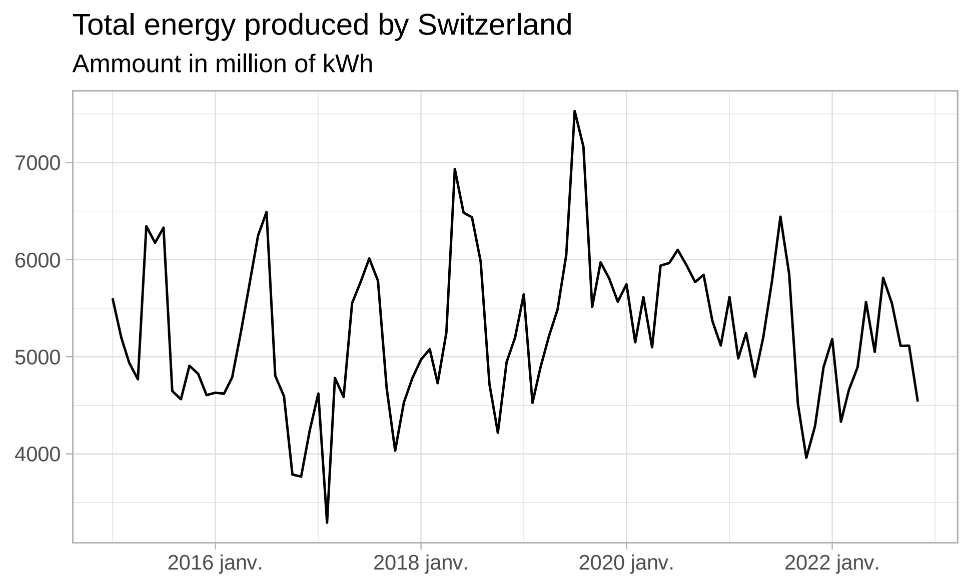
One of the conclusions we were able to draw based on the previous graphs was the presence of different seasonality in the data. To understand the patterns, we will "zoom in" to see what is happening on shorter time period. (starting with a long period and then gradually reducing it)

4.1.2.1 Seasonality of the month over the year Our analysis starts with data grouped by month. Here is the graph monthly consumption :



This clearly confirms what we have said previously :

- The levels of consumption does not vary over time.
- Significant difference between summer and winter, with a peak for consumption in December/January and the lowest in July/August.
- The covid had almost no impact on consumption (2020 is slightly lower than other years, but we can not see any significant difference).

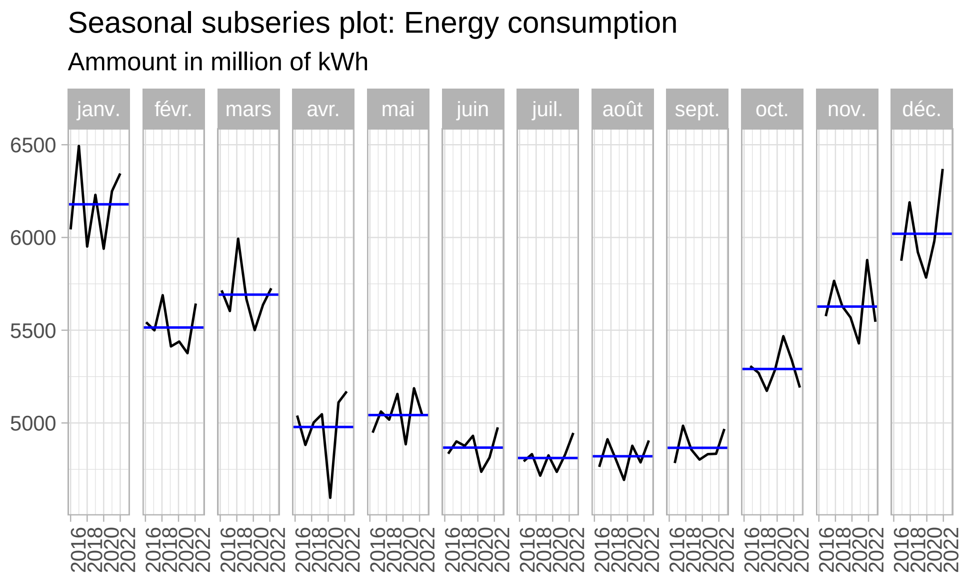


What we can extract from the chart :

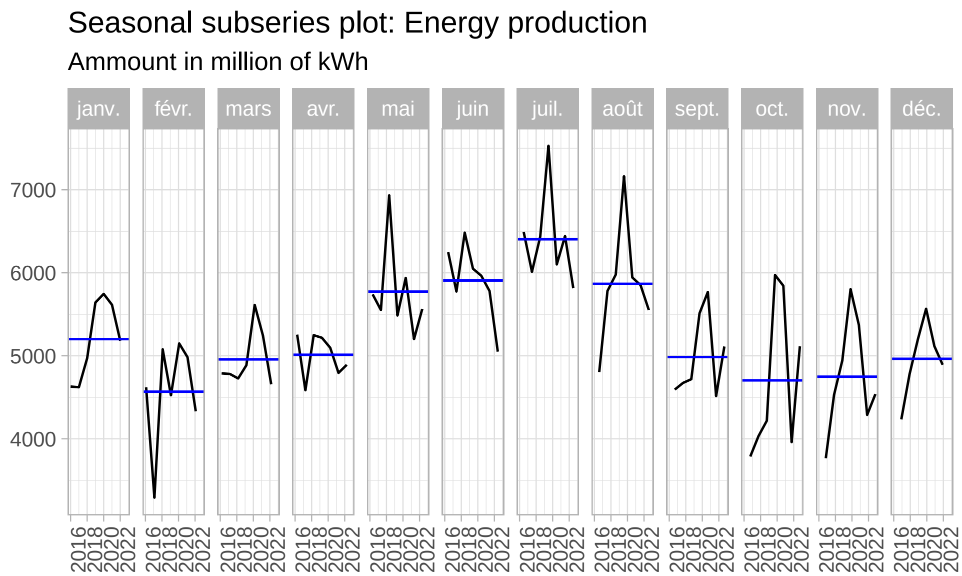
- We can see more clearly the difference with consumption : we do not have the same regularity with seasonality. We can confirm peaks and drops with the season but with less regularity.

- If the general level of production is quite the same between the beginning and the end of the time perdiod, we see a higher variance among the year.
- For the covid period (year 2020 in Switzerland) : we see a redudction in the seasonilty, with significantly smaller peaks and falls throughout this year, with the old pattern appearing to return right after 2020.
- Finally, as mentioned in the previous section, there might be a general decrease in production after the highest point in 2019.

Here is an other view of the same data, grouped by month over the years :



This graph allows us to better compare the differences between the months for each year. The consistency of dips and peaks between summer and winter is clearer than ever. We can also note that the average consumption value (blue line on the graph) shows us a generally low level over a longer period than we might have thought: we can see a drop from April to September, contrary to our first analysis which suggested only the months of July and August. As for the differences between the years themselves, the variance is not very large, so we can say that overall consumption has changed little since 2016. Finally, we note a significant drop between March and June 2020, most likely linked to Covid, which we were unable to see clearly in the previous graphs.

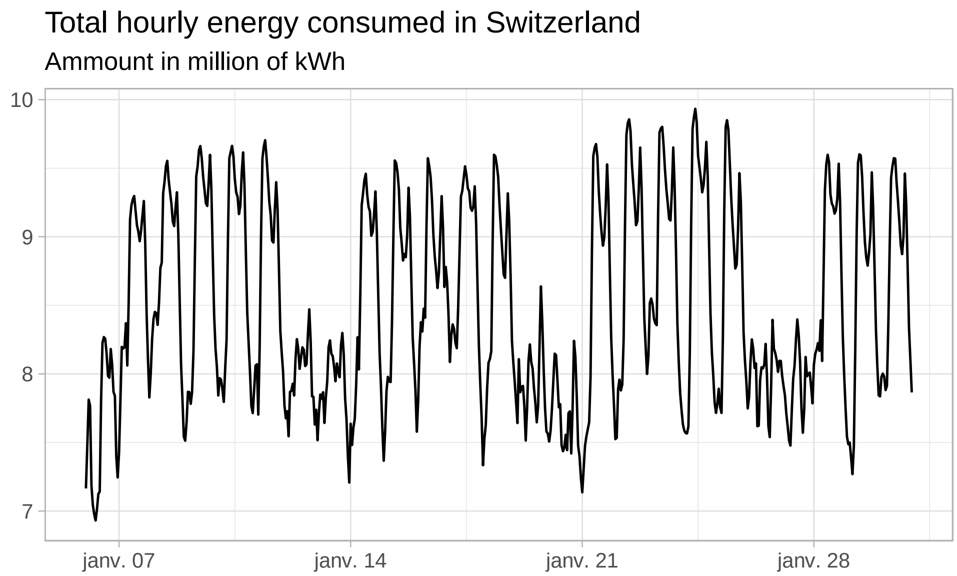


The first thing we notice is the very high variance within months among each year. Indeed, we can see many big drops and peaks for several periods, especially for February 2017 and for the end of 2019 in general. Clearly, the year has a significant impact on production, indicating that it depends on more than just population density or labor.

This high variance can also lead to misinterpretation of the graph: one might think that the average value of production (blue lines) does not vary significantly from one month to the next, compared with consumption. This is mainly due to the higher scale of the y-axis, caused by high volatility. The range of production (4,500 to 6,500) is even higher than that of consumption (4,600 to 6,400), also showing the strong importance of seasons.

4.1.2.2 Seasonality of the day over the week Now that we have a better understanding of the monthly seasonality, let's zoom in on weekly seasonality. To do so, we will take the hourly data in a shorter time period.

You can see here the chart of the consumption per hour for the month of January 2019 (2019-01-06 ~ 2019-01-30, to be exact)



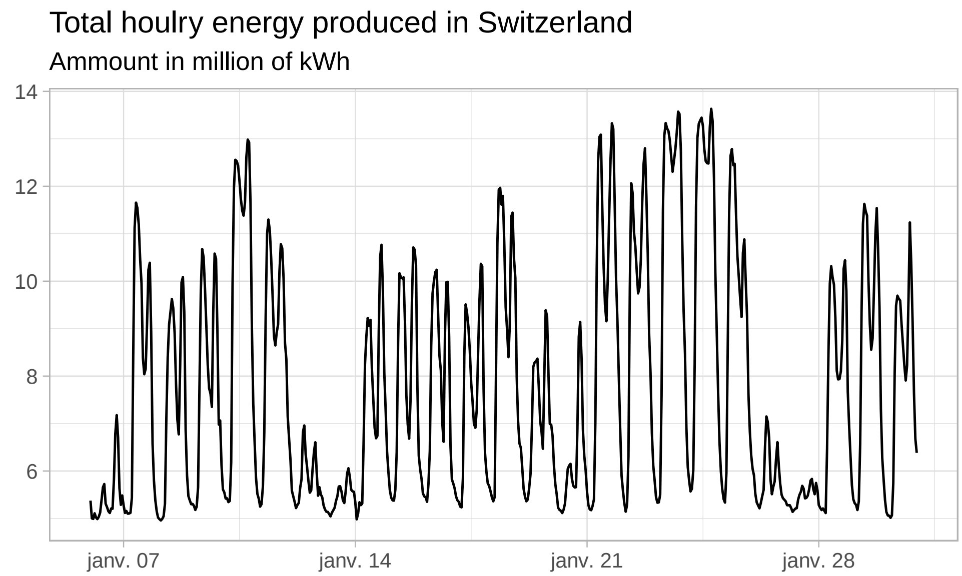
This graph shows us both weekly and daily seasonality :

We have peaks during the day (in the morning and a second peak in the late afternoon) and dips during the night, but also during the week, with higher volumes on weekdays with less than half for the weekend. There is no significant difference between the days of the week themselves. This can easily be explained but the weekly and daily agenda of most people : the peaks happens when poeple are awake and are working.

An interesting comparison is the small difference between weekend days and weeknights: the total amount consumed is almost the same, with some nights even exceeding weekend consumption. This suggests that the majority of consumption comes from companies and large institutes, and not from private individuals. Indeed, these two times of the week are marked by the absence of work.

It's important to note that we only have data for a specific month over a 6-year period : so while this gives us a general idea of consumption behavior throughout days and weeks, and we can expect it to be constant over the whole period, we can not yet be sure that this pattern is generalizable to every period.

As usual, let's now displays the same chart but this time for the production :



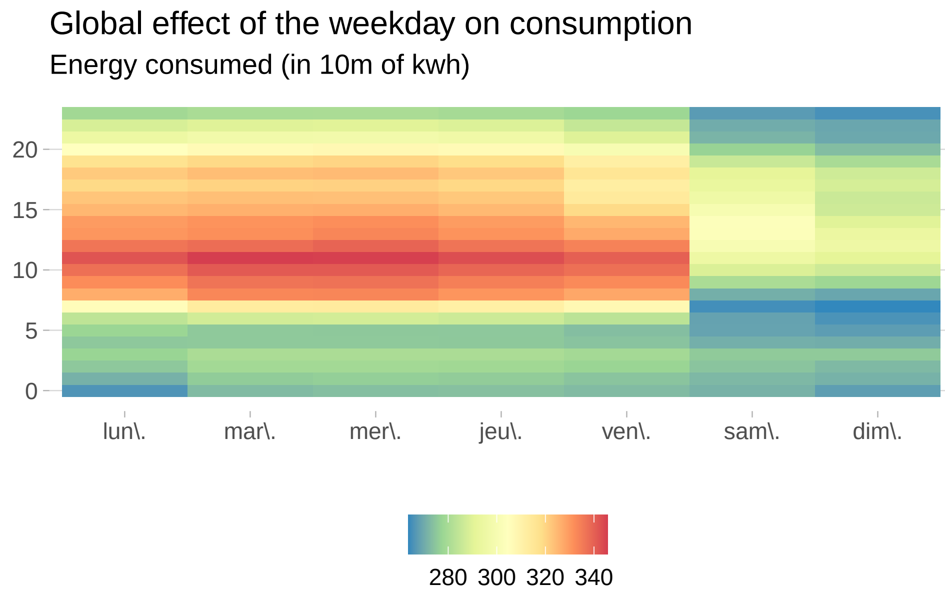
It gives us some very important indications. First of all, the production range is higher, which means that we produce too much during the day and not enough at night. This lack of electricity at night could be a choice, given that its price is very low and that importing electricity could be cheaper than producing it during a period of low activity.

In contrast to consumption, we have similar production on every night of the week (around 5 millions). Even lower than during the weekdays, the weekend is always higher than nights: the fact that production is low at weekends highlights human dependency. Indeed, if factories were 100% autonomous, we shouldn't see a drop in production at weekends. However, it remains to be seen whether this drop is due to a reduction in manpower or simply a choice on the part of the factories, since demand is much lower at weekends.

Generally speaking, we can see that production for this period is in line with consumption.

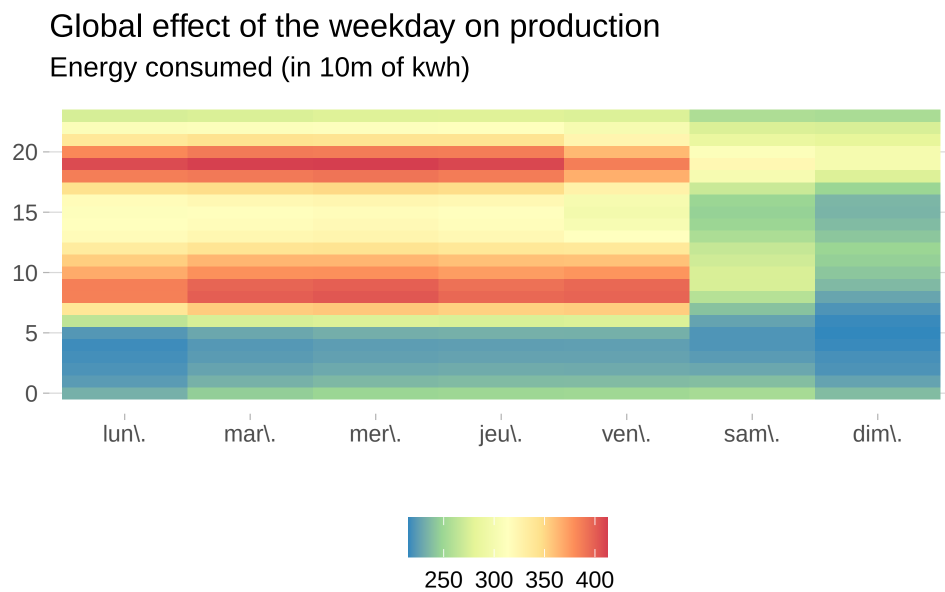
4.1.2.3 Seasonality of the hour over the day Our final zoom will focus on hourly/daily seasonality. As usual, we'll display the consumption graph first, before the production graph. Instead of a conventional graph, given the short period between observations (every hour), we have decided to use a heat map this time. This has enabled us to use all the data at our disposal, while giving us a clear model output. This will thus enable us to confirm the models studied over the whole period.

You can see below the first heat map, containing the data for consumption.



We can conclude that the trend is generalized over the whole period, with peaks around midday and a gradual decrease as the day progresses. Minimum consumption can also be generalized between nights and weekends. Electricity consumption in Switzerland is clearly in line with human activity.

Here is the second map, this time for production:



The production pattern differs from the previous one: unlike an increase that would occur progressively, there are clear peaks and drops during the day, with a maximum around 7 p.m. (with a value 18% higher than maximum consumption).

We can also confirm the absence of night-time production, with overall values more than 20% lower.

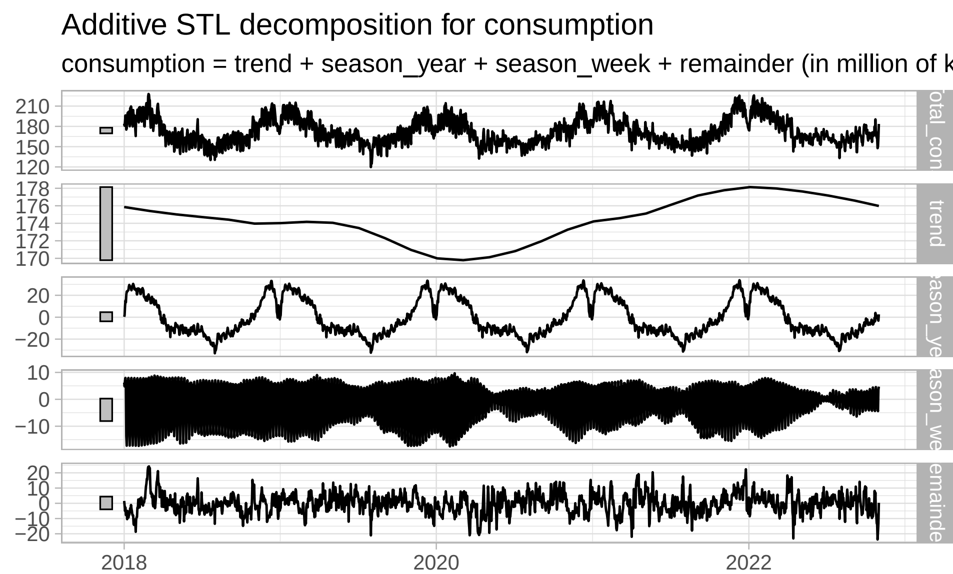
4.1.3 STL Decomp

Now that we know exactly how our data works, we can go a step further and try to find the exact origin of these pattern. To do so, we will use the "Seasonal and Trend decomposition using Loess" (STL), one of the best smoothing methods. It allows us to divide the data into trend, seasonality and residuals, and analyze them separately.

Before doing it, we have to chose the right model for the decomposition (additive vs multiplicative). Since the overall level of data, seasonality included, does not change magnitudes from one year to the next, we have decided to chose the additive model.

We also have reduce the scope to five years (2018 to 2022) to get a better output. We have shown that seasonality is constant throughout the years, so we will not miss any important insight.

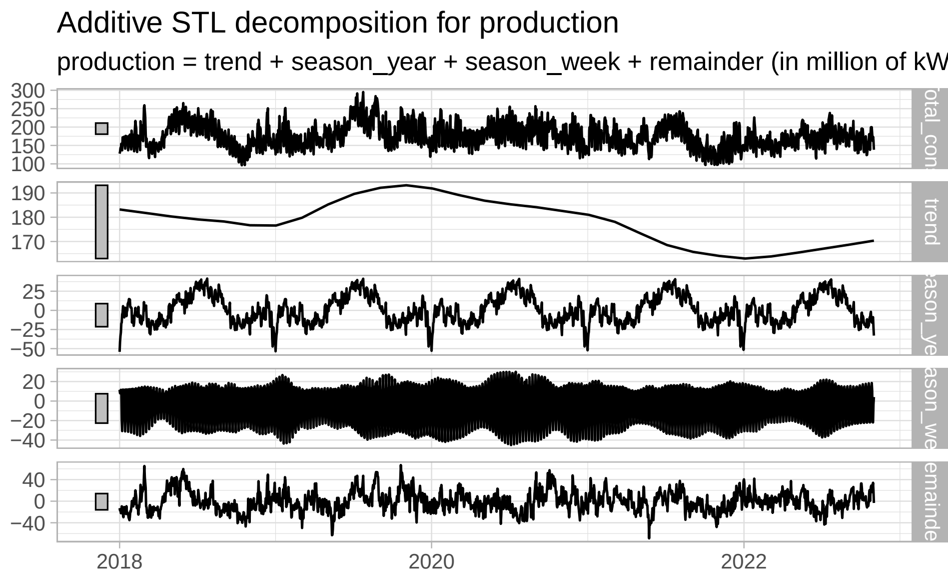
Here is what the decomposition looks like for the consumption :



Although the second is not very visible on the chart, STL has highlighted the annual and weekly seasonality as expected.

For the rest, the trend appears to be a cycle, with the shape of an oscillation. Looking at the monthly data, there were peaks in December and January. While this is true for the month as a whole, we see a sharp drop around Christmas and New Year, as well as in August. This is due to the fact that many people leave the country on vacation, causing daily consumption to drop significantly.

Moving on to the decomposition of production, we get :



Decomposition has proved very useful for production: whereas it was more difficult to extract patterns from previous graphs, the seasonality is obvious and regular once isolated.

Production is highest in summer, when the weather is fine and the rivers have their maximum flow, and lowest in winter, with a sharp drop-off around the end-of-year holiday periods, when the power plants are less active and their staff are on vacation. Production may also have adjusted to a lower consumption forecast for the same period.

Finally, as in the first decomposition, the weekly seasonality, although illegible on the graph, is clearly there.

4.1.4 Main insight from the overall Data

Similarities

- 3 levels of seasonality : daily, weekly and yearly.
- Low impact of the Covid on both variable.
- They are correlated to the weather, but for opposites reasons.
- When possible, production tries to match consumption

Results from consumption

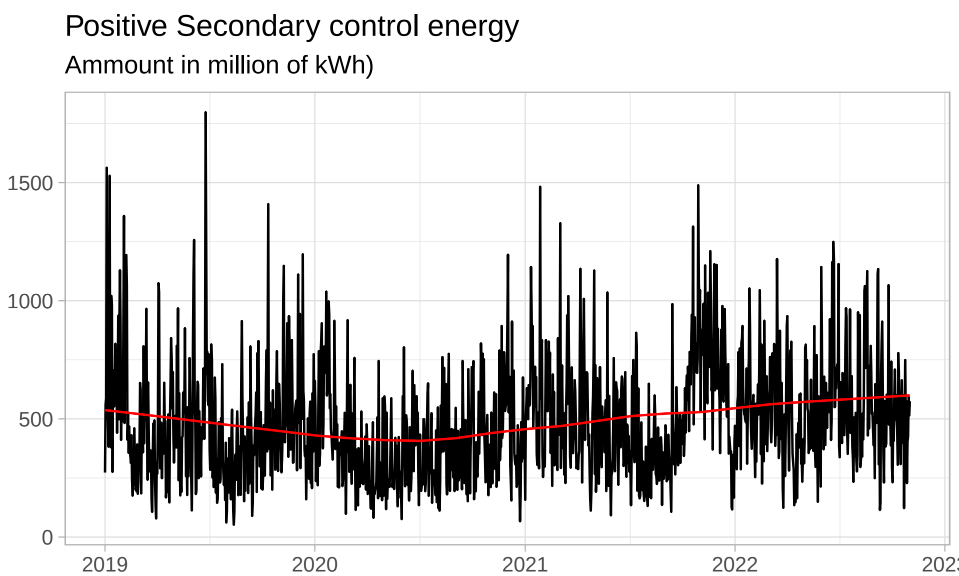
- Very regular data between years.
- Low in summer, high in Winter, with falls during high vacation periods.
- High volatility during winter, caused by winters of varying severity and low one during summer.
- Depends mainly on the weather: people use little electricity when it is warm and sunny, and much more when they need heating or light.

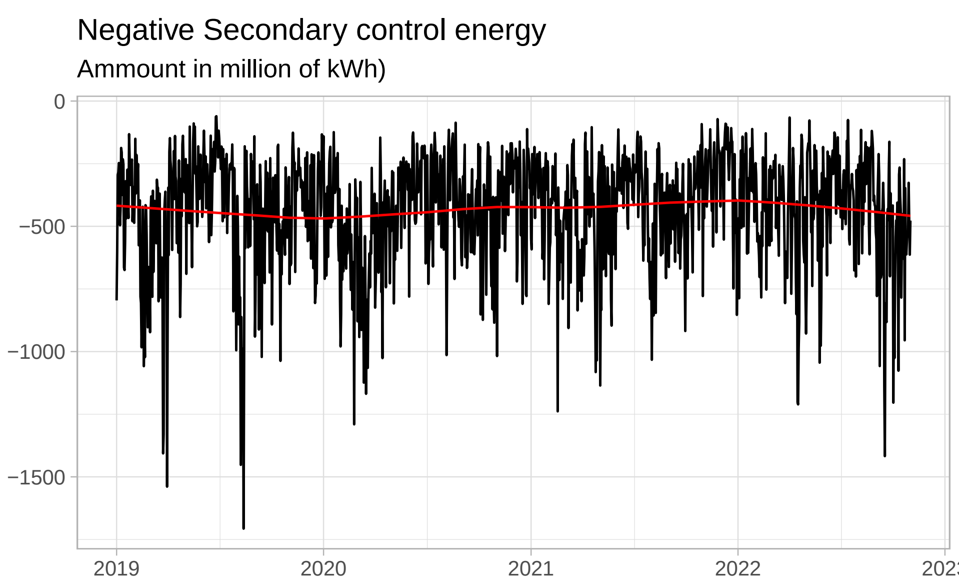
Results from production

- No clear pattern among years when we looked at the total energy produced.
- High in summer, low in Winter, with higher peaks and dips.
- Higher volatility among the year, since it depends more heavily on meteorological conditions, which vary from year to year.
- Depends mainly on the weather: due to the nature of the production in Switzerland (hydroelectric and solar more specifically), it can produce more electricity during periods of high heat and rainfall, with higher river levels, than in winter, when it relies mainly on nuclear power plants.

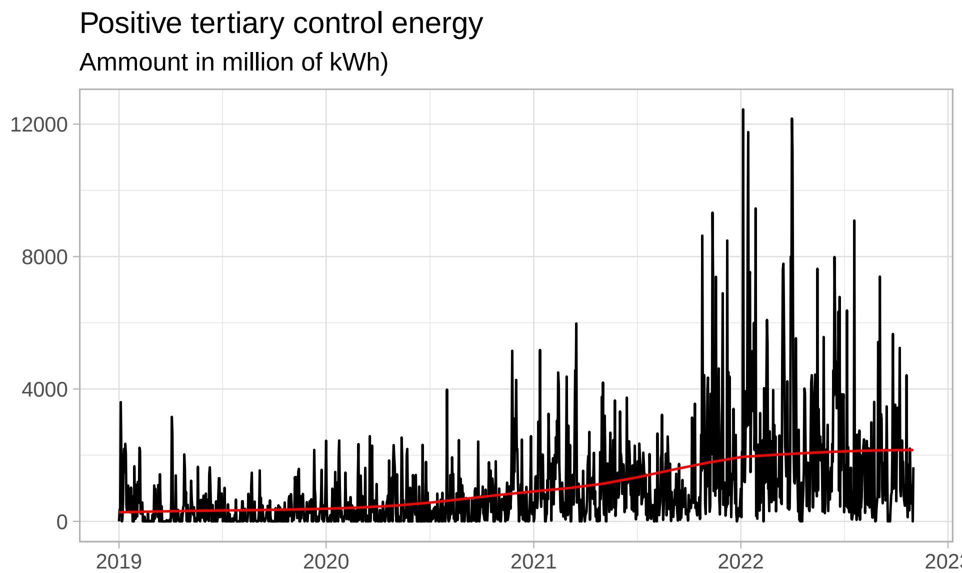
4.2 Secondary and Tertiary control

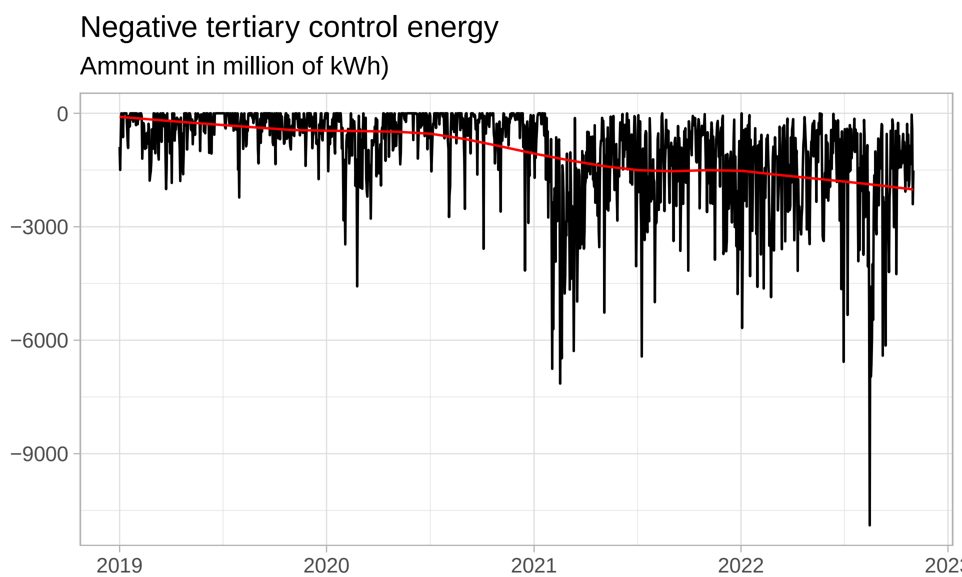
4.2.1 Positive and Negative Secondary control with trend





4.2.2 Positive and Negative Tertiary control with trend





4.3 Border

4.4 Foreign

4.5 Canton

4.5.1 Quick visualisation

Now that we have seen how consumption and production behave for the country as a whole, we can move on to the cantonal level. The idea is to understand exactly where electricity production and consumption stand: are they similar for each canton, or do they depend on a number of factors? If so, are they the same for each category? Does Switzerland depend on just a few cantons for its electricity supply?

These are just some of the questions we'll try to answer in this section.

To do so, we will use the `Canton_df_long` dataset, which contains all the necessary information at the cantonal level.

Here is what the final version looks like :

time	hourly	month	date	Cantons	production	consumption
2015-01-01 00:15:00	2015-01-01	janv	2015-01-01	argovie	511742	151008
2015-01-01 00:15:00	2015-01-01	janv	2015-01-01	fribourg	6657	82368
2015-01-01 00:15:00	2015-01-01	janv	2015-01-01	glaris	56449	12761
2015-01-01 00:15:00	2015-01-01	janv	2015-01-01	grisons	196507	89631
2015-01-01 00:15:00	2015-01-01	janv	2015-01-01	lucerne	4576	104484

We can see that we have now for every time, the production and consumption for each canton.

A quick check of the total of missing value gives us :

```r
sum(is.na(canton_df_long))
#> [1] 0
```

As said in introduction, some cantons have been grouped together. You can see the breakdown here:

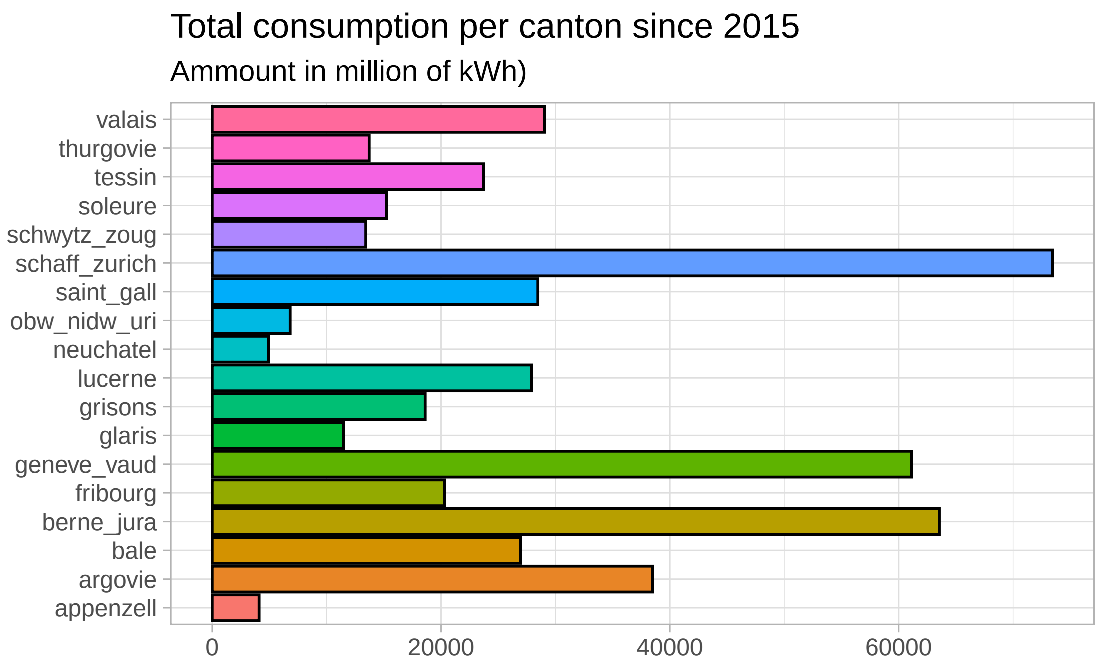
```
#>  [1] "argovie"       "fribourg"      "glaris"        "grisons"
#>  [5] "lucerne"       "neuchatel"     "soleure"       "saint_gall"
#>  [9] "tessin"        "thurgovie"     "valais"        "appenzell"
#> [13] "bale"          "berne_jura"    "schwytz_zoug"  "obw_nidw_uri"
#> [17] "geneve_vaud"   "schaff_zurich"
```

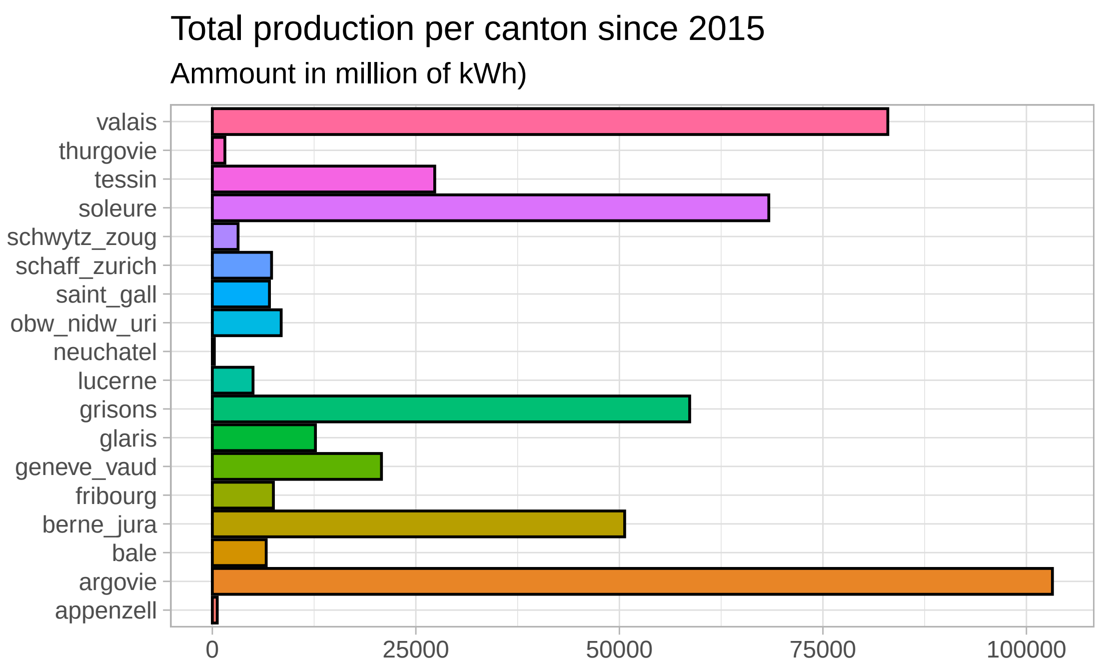
In all, we have 24 cantons (they do not differentiate between half-cantons) spread over 18 values. For further analysis, we will split the value to get the 24 cantons. Methods and results will be presented in an other section.

Regarding the time period, as we have done previously, we will only represent data since 2018 to get a better view of on graphs.

Since we have so many cantons, showing them all on one chart would make it incomprehensible. So we'll try to see whether certain cantons are major contributors to consumption or production, and then focus our analysis on them.

You can find below the total consumption per area since 2015 :





Clearly, some cantons are responsible for most of our data.

For the consumption, we have :

- Schaffouse and Zuirch, responsible of 73460 MkWH of the total consumption ()
- Berne and the Jura, accumulating a total of 63551 MkWH ()
- Geneve and Vaud, with a global consumption of 61111 MkWH ()

- Together, they are responsible of 41% of the total consumption

While for production, we see :

- Argovie, responsible of 103194 MkWH of the total prodcution ()
- The Valais, accumulating a total of 82975 MkWH ()
- Soleure, with a global production of 68353 MkWH ()
- Together, they are responsible of 54% of the totel consumption

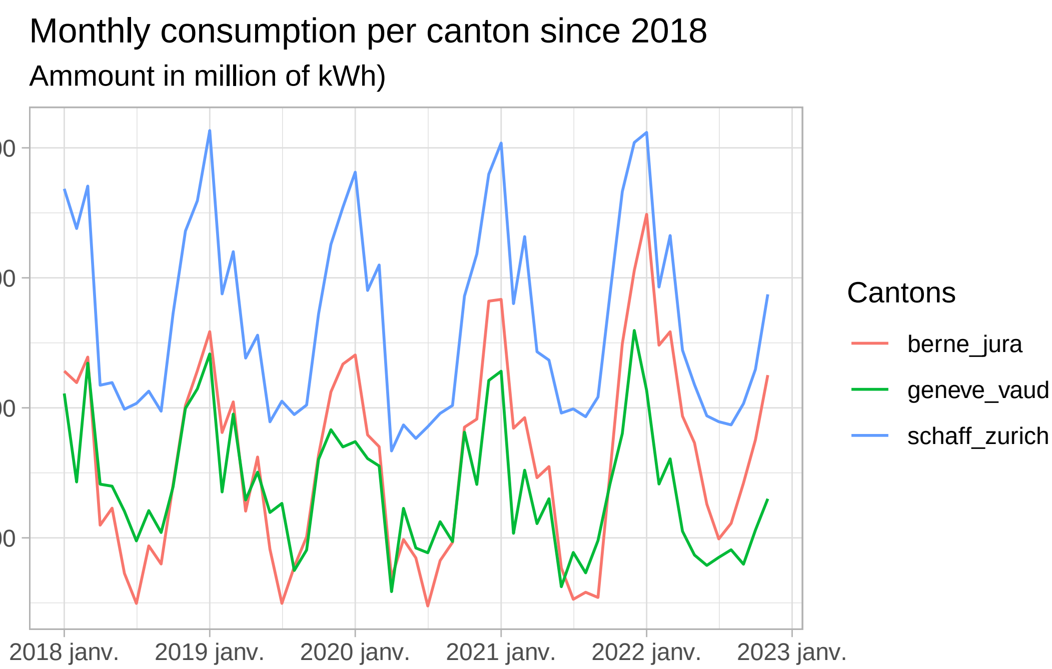
In general, there is greater variance in production, with some cantons producing a great deal and others producing almost no electricity at all.

Consumption and production appear to be independent: large consumers are not necessarily large producers, so it could be said that they are not driven by the same variable.

4.5.2 Analyse of the top 3

For the next few charts, we will focus on the top 3 in each category.

Let's start by ploting the monthly time series for the top 3 consumers :

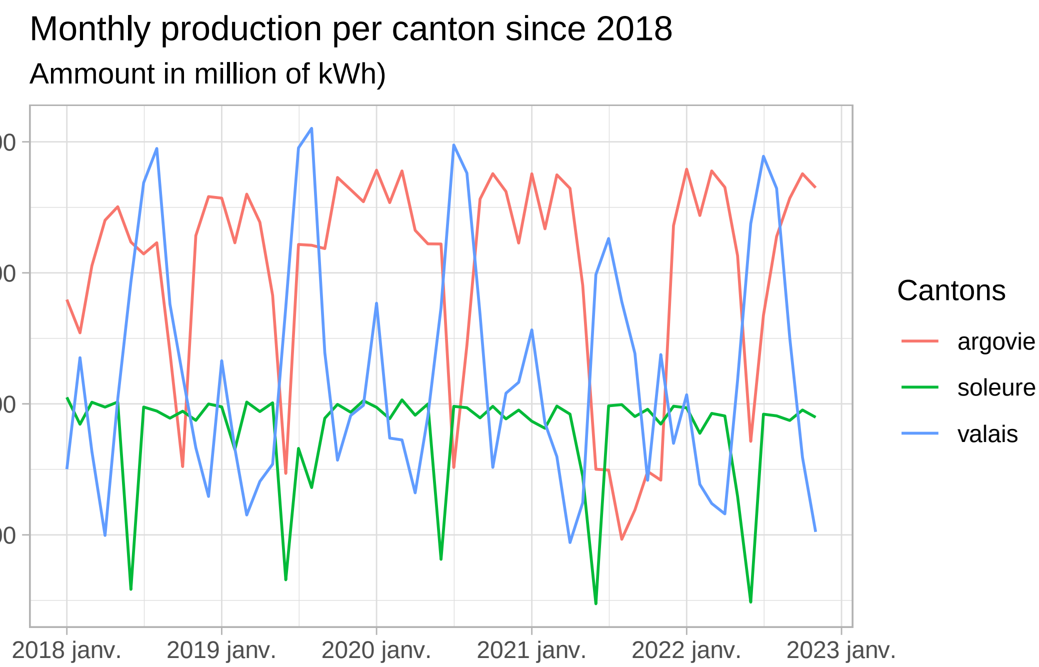


The only noticeable difference among the cantons seems to be the level. Everything else looks the same : the general trend is constant and we have peaks and dips at the same period of the time. If we compare it to the density we have :

- *x for Berne and the Jura*
- *y for Geneve and Vaud*

- *z for Schaffouse and Zurich*

The difference in levels can be explained by the total population of the regions.

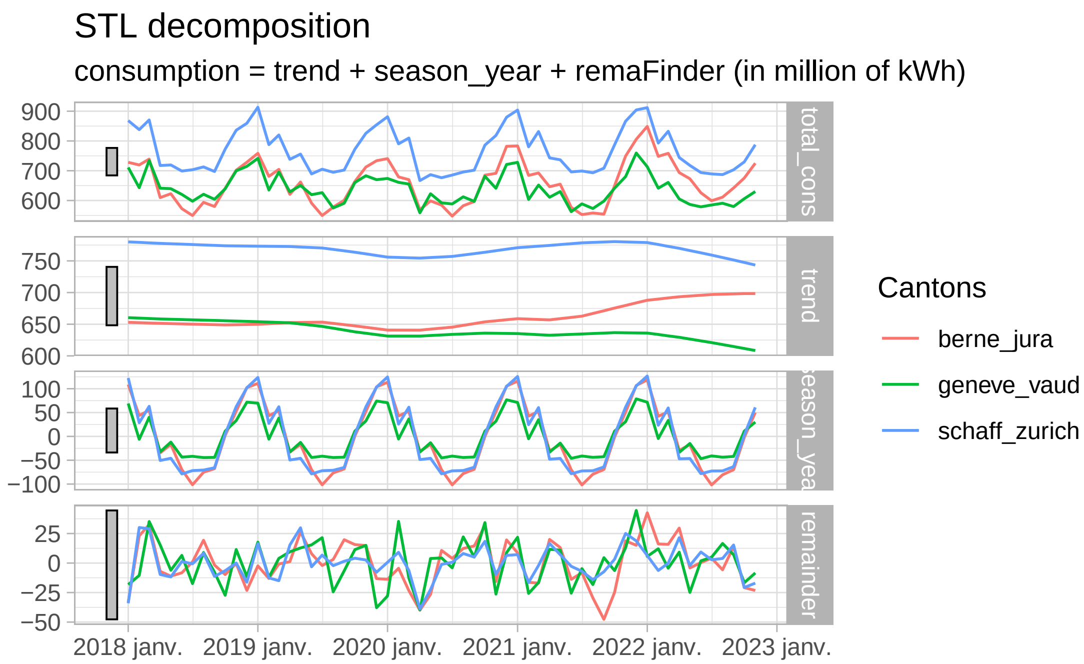


As we suggested earlier, production is not determined by total population. This is clearly confirmed by the graph above. Moreover, seasonality seems to be completely different for each canton, with staggered peaks and dips (exactly the opposite for Argovie and Valais). Finally, while Argovie and Valais show generally low production with peaks, Soleure seems to have generally high production with only dips, happening once a year.

The production pattern really does differ from canton to canton, and we will look at it in more detail later in this section.

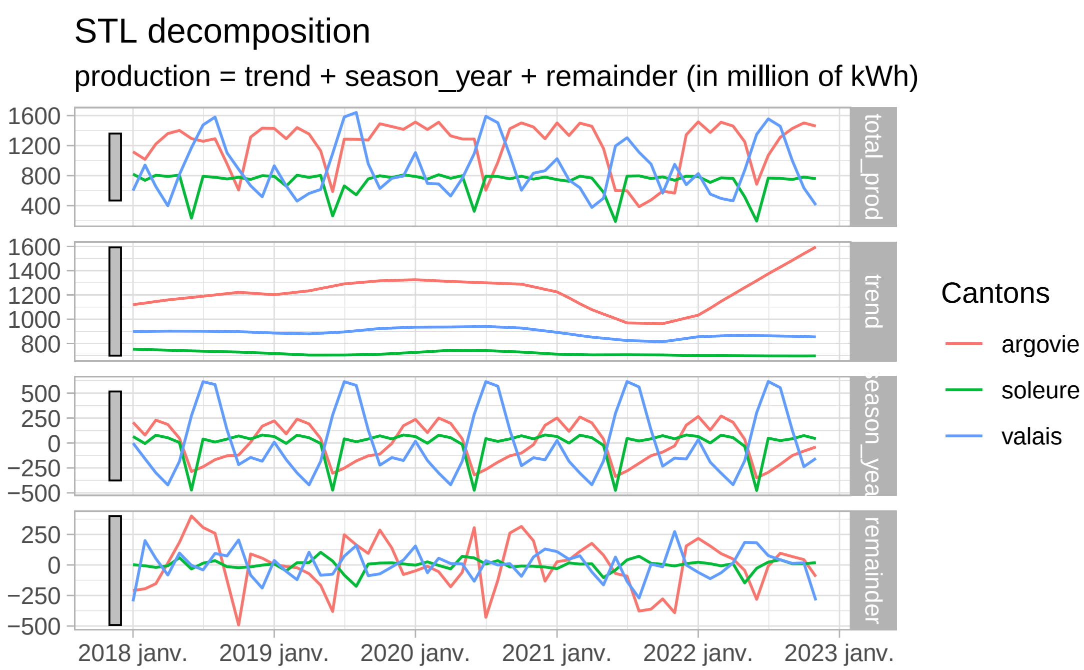
To confirm these similarities/differences, we ran the STL decomposition for the 3 main electricity producers and consumers. Once again, we opted for an additive model.

Top 3 consumers STL decomposition :



The STL decomposition looks exactly as we expected : the trends are almost parallel between the cantons, with similar oscillation shapes for the seasonality and almost superposed remainders. Once again, the main difference lies in the level of the trend, but not its direction.

For the second STL, you can see below the corresponding output :



The STL decomposition reinforces dissimilarity between cantons :

- While Soleure and Argovie have a flat trend, Valais seems to become positive at the end.
- If Valais production seems to peak in summer, the other two cantons have their lows at the same time.
- With the exception of a higher volatility in Argovie's remainder, we can not draw any conclusions.

4.5.3 Facet wrap by Canton

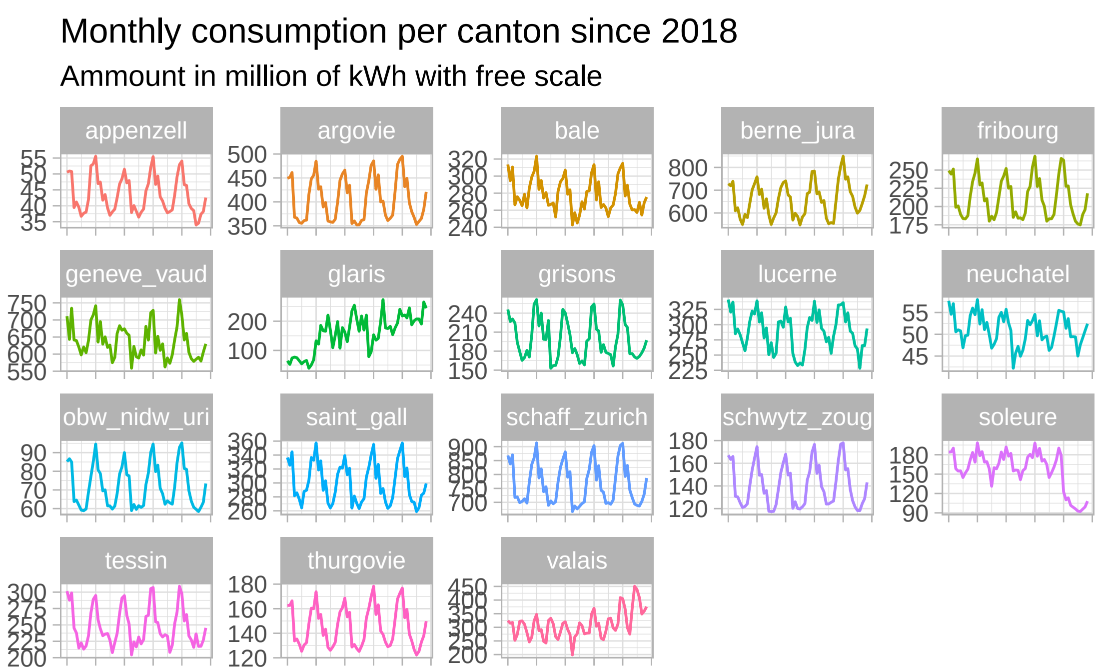
Let's now look at each canton separately. We want to see whether we can find similarities in the data for both production and consumption, and whether we can group them by macroeconomic variable.

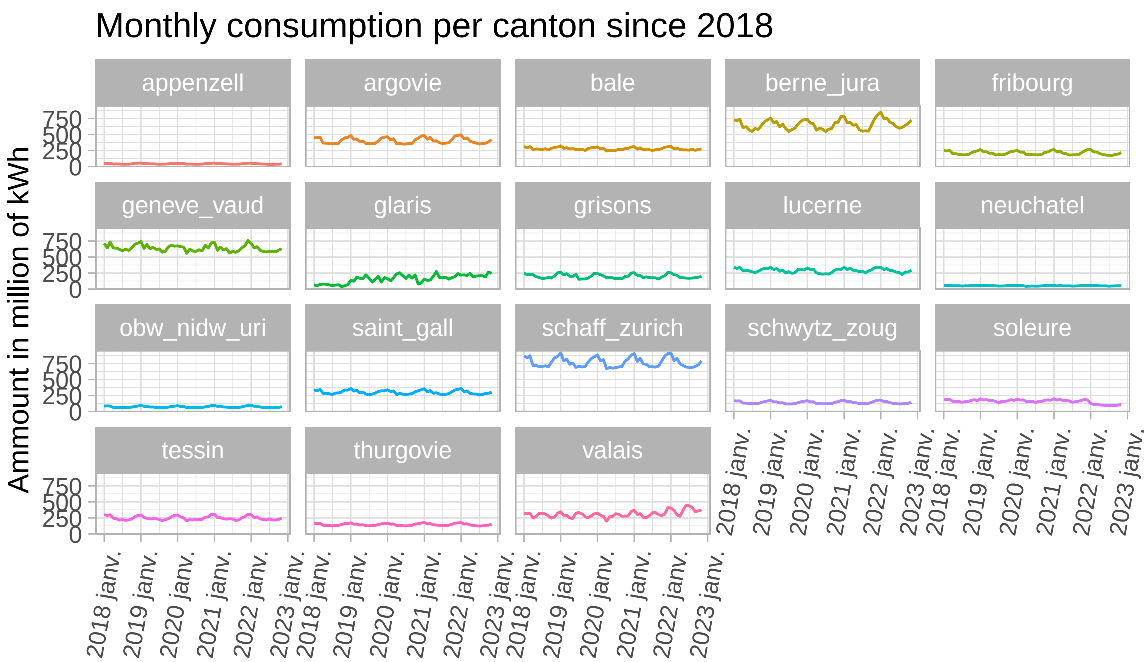
In other words: do we have specific patterns for each canton, some of them, or common to all?

To do this, we analyzed each canton side by side in two different graphs: with and without a free scale.

The first will allow us to compare seasonality and the specific pattern, while the second will show us the total difference, in terms of overall consumption/production.

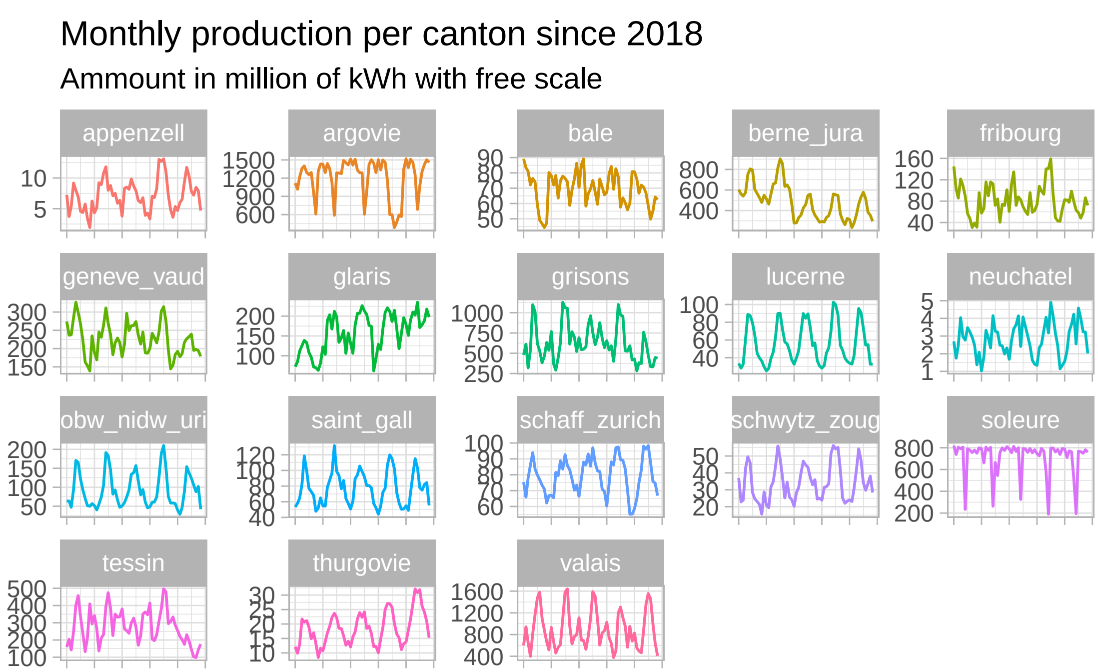
Facet wrap per Cantons with and without free scale.

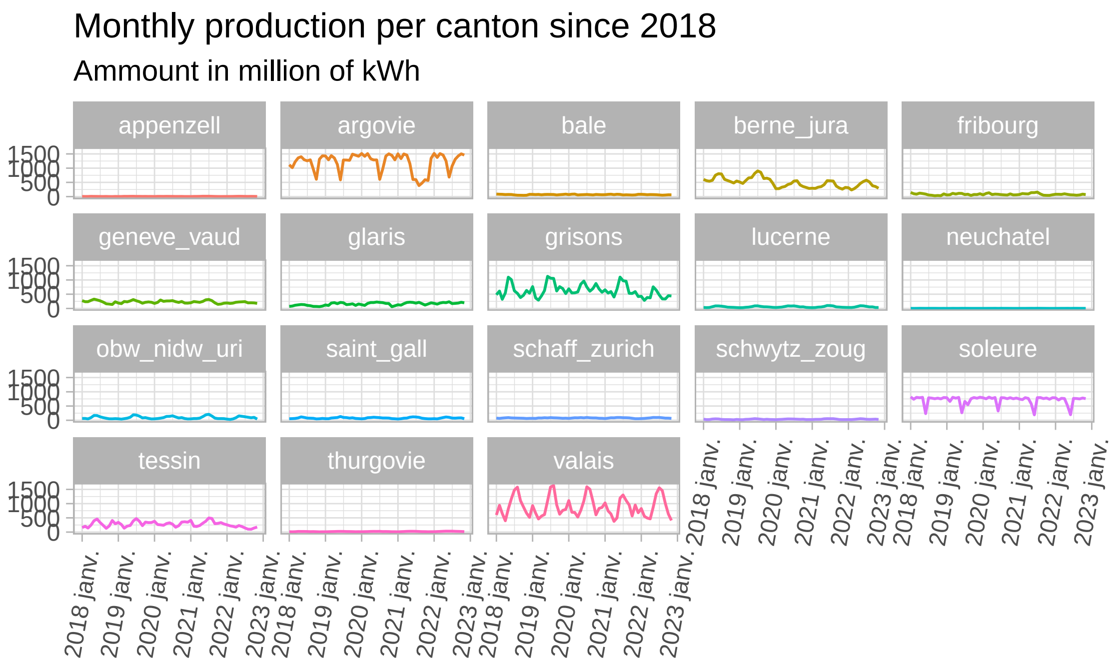




-> gros peak l'hiver lié au ski, surout poste covid.

Conclusion : niveau global et légère tendance liée à la population, changement "soudain" lié à à d'autre facteur comme tourisme, activité industriel ou politique d'économie.

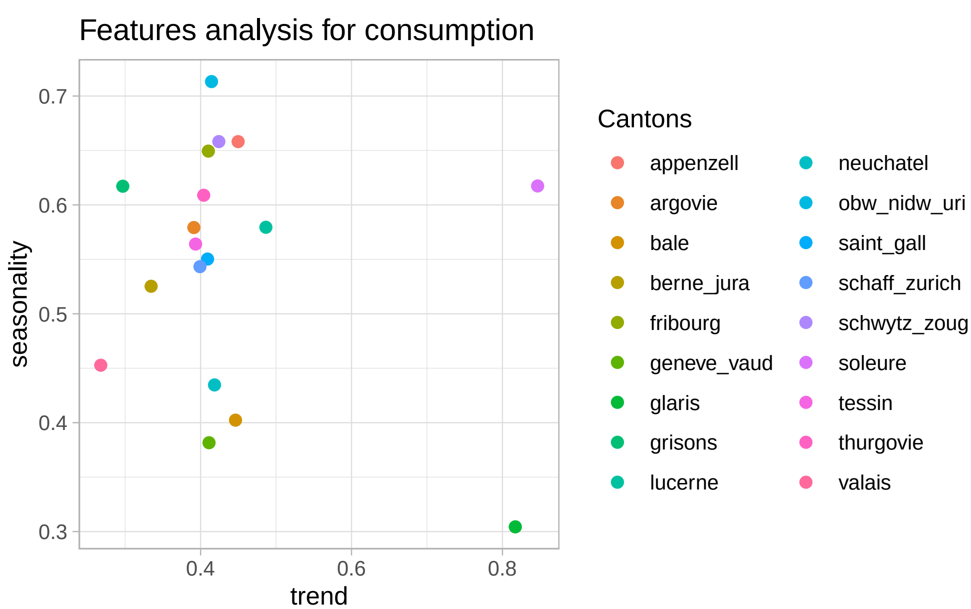




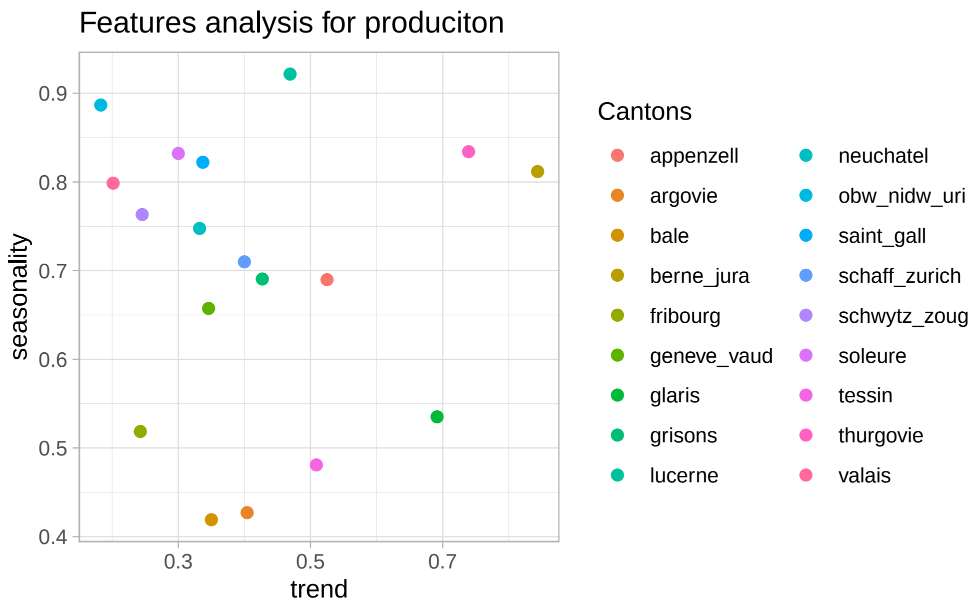
Format de production pour Soleure et argovie provient des centrales nucléaire

4.5.4 Features analysis

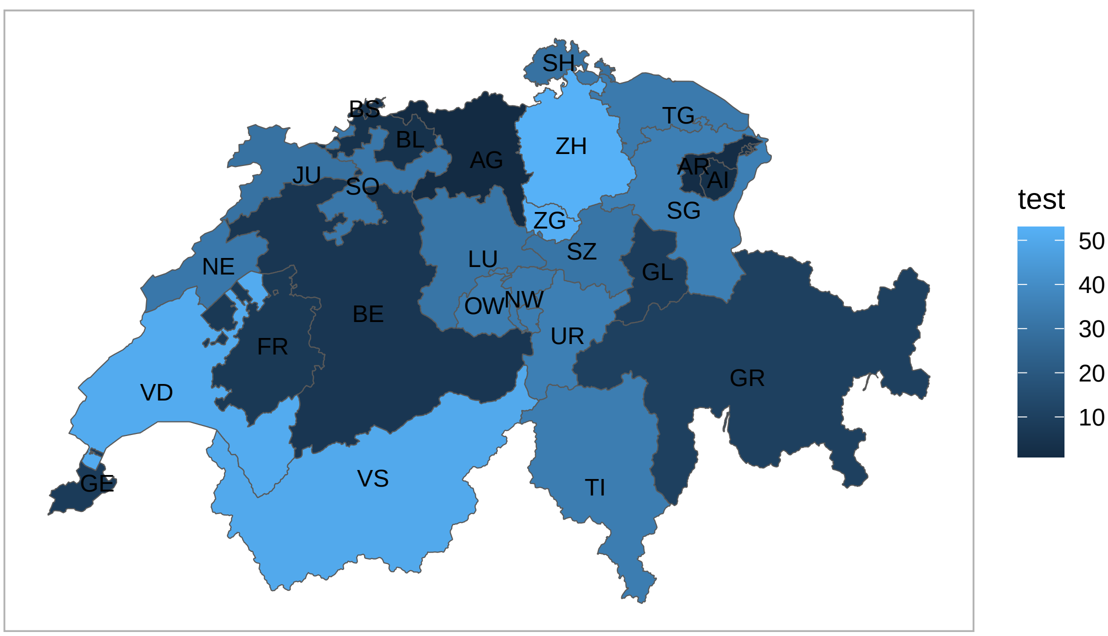
Features analysis to check the strength of the seasonality/trend



test



4.6 Mapping



5 Forecasting

6 Conclusion and possible extension

- Would have been nice to assess the price evolution/prediction.

7 References and Appendix

- https://arxiv.org/pdf/1901.10257.pdf
- https://www.swissgrid.ch/fr/home.html
- https://gadm.org/
- https://www.admin.ch/gov/fr/accueil/documentation/communiques.msg-id-90221.html
- https://www.eda.admin.ch/aboutswitzerland/fr/home/wirtschaft/energie/energie---fakten-und-zahlen.html
- https://www.rts.ch/info/economie/13304056-penurie-delectricite-cet-hiver-les-enjeux-en-cinq-questions.html
- https://www.economiesuisse.ch/fr/articles/energieticker-f
- https://www.kernenergie.ch/fr/les-centrales-nucleaires-suisses-_content---1--1068.html
- https://fr.wikipedia.org/wiki/Hydro%C3%A9lectricit%C3%A9_en_Suisse
- https://www.bfe.admin.ch/bfe/fr/home/approvisionnement/energies-renouvelables/force-hydraulique.html#:~:text=Environ%2063%20%25%20de%20la%20production,quantit%C3%A9s%20consid%C3%A9rables%20d’%C