# Stream Graph

## Content of the Chapter

This chapter provides a well-founded introduction to the world of data visualization, with a particular focus on the stream graph. It begins by discussing the line chart as a fundamental tool for representing temporal changes in data. Building on this foundation, the area chart is introduced to illustrate the cumulative development of multiple data series. These two types of charts form the essential basis for understanding the more complex concept of the stream graph. As the central focus of this chapter, the stream graph enables dynamic and intuitive visualization of large and multilayered datasets. Through the step-by-step introduction and comparison of these chart types, a deep understanding of the potential and applications of the stream graph is developed.

## Your Competencies

▶The ability to create and interpret line charts in order to accurately represent temporal trends and changes.

▶ Proficiency in the methodology of area charts to illustrate the cumulative development of multiple data series and to understand their application scenarios.

▶A deep understanding of the construction and interpretation of stream graphs, enabling you to recognize their advantages in visualizing complex and multilayered datasets.

Research Project
Vis4Schools

This textbook chapter was developed as part of the research project Vis4Schools at the University of Applied Sciences St. Pölten. The aim of the project is to support students in reading, understanding, and constructing data visualizations more effectively. To this end, the project investigates how learning materials should be designed and which pedagogical methods are necessary to enhance data visualization competencies during school education.

# The Line Chart

## Description

The line chart is a central tool in data visualization, commonly used to represent changes in data values over time. It offers a clear and straightforward method for illustrating developments and trends across a specific time period.

## Functionality and Representation

When creating a line chart, data points, such as values representing specific points in time, are arranged within a coordinate system. These points are placed along the X-axis according to their chronological order and along the Y-axis according to their values, then connected by lines. This allows for a visual traceability of value developments over time.

## Example

A practical example of using a line chart is the visualization of average daily temperatures in March. The following table illustrates how temperatures can be represented in tabular form:

| Date (x-Axis) | Temperature (y-Axis) |
|---|---|
| 1. March | 9,0°C |
| 2. March | 7,5°C |
| 3. March | 8,0°C |
| 4. March | 10,5°C |
| 5. March | 9,5°C |

These and additional values can now be plotted in a coordinate system (Figure 2):

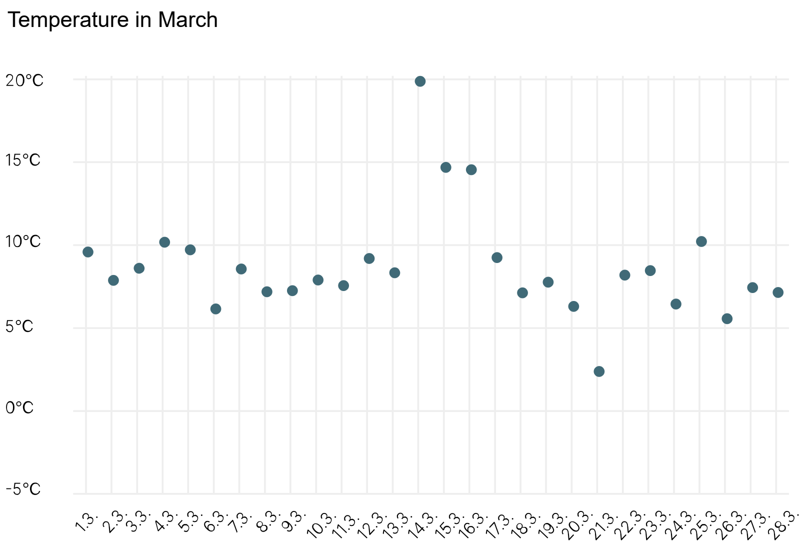
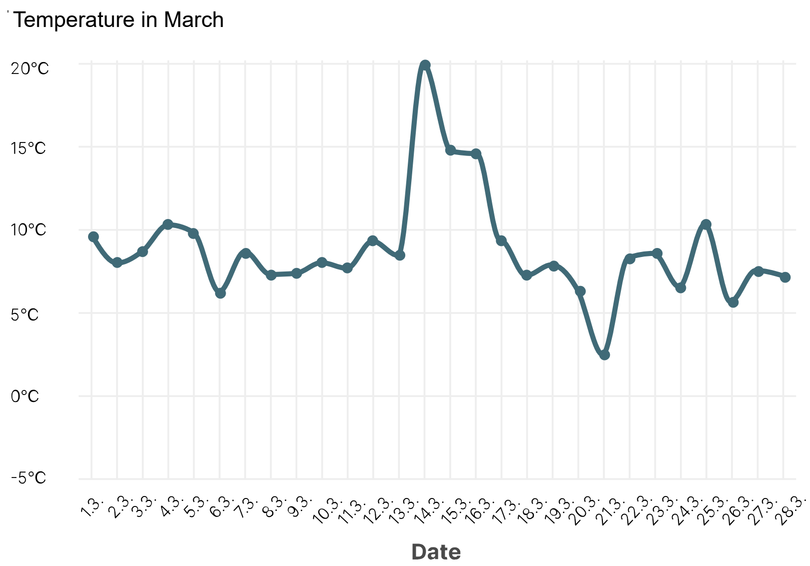



Figure 1

When, as shown in Figure 1, only points are displayed, the chart can sometimes be difficult to interpret, especially when many points are located close together. However, when a line is drawn from one point to the next, it significantly improves the readability and clarity of the chart. Compare Figure 1 (points only) with Figure 2 (points connected by lines) to see this difference.

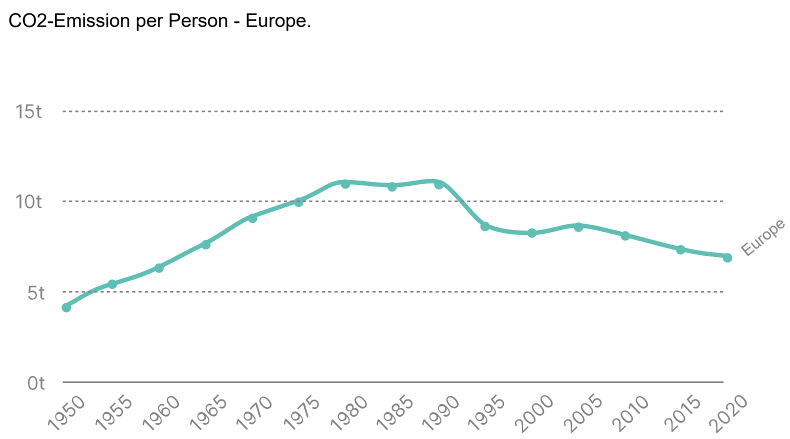



## Figure 2

The chart in Figure 2 clearly shows that March 14th was significantly warmer than the other days. Additionally, it is evident from the diagram that March 21st was the coldest day, followed by a sharp temperature increase on March 22nd and a slight rise on March 23rd.

The chart illustrates that March 14th was significantly warmer than the other days.
It also clearly shows that March 21st was the coldest day, followed by a sharp temperature increase on March 22nd and only a slight rise on March 23rd.


Trends, meaning likely developments of a value, can also be identified in a line chart. The chart in Figure 3 shows how the amount of $CO_2$ per person has changed over time in Europe.




## Figure 3

From the early 1950s until the late 1970s, Figure 3 shows a continuous increase in emissions, which is visualized by the rising curve in the chart. In the following decade, from 1980 to 1990, the emission levels stabilize, reflected by the nearly horizontal direction of the line. From 1990 onward, a declining trend in emissions can be observed, which is traceable through the downward slope of the line in the chart up to the year 2020.

CO2 Emissions

CO2 emissions refer to the release of carbon dioxide into the atmosphere. Carbon dioxide is a greenhouse gas that is mainly produced by the combustion of fossil fuels such as coal, oil, and gas, as well as by other human activities like deforestation and industrial processes.

When analyzing data with a line chart, it is possible for some values to fall below zero. This occurs, for example, in temperature measurements when values drop below freezing during the winter months.
The versatility of the line chart is also evident in its ability to display multiple time series within a single chart. This allows for the comparison of parallel developments.

This chart not only illustrates the average CO2 emissions in Europe but also relates them to other world regions, providing a more comprehensive insight into global trends.

As shown in Figure 4, all individual values from a table are first plotted in the chart again. For better understanding, the different world regions are depicted in distinct colors:

Figure 4

When a chart is reduced to a mere collection of points, clarity suffers and identifying trends can become difficult. To avoid this and ensure a clearer representation of temporal developments, points belonging to the same category are connected with lines of uniform colors. This method significantly facilitates tracking and interpreting developments over time. Compare again the chart in Figure 4 (points only) with Figure 5 (points connected by lines).

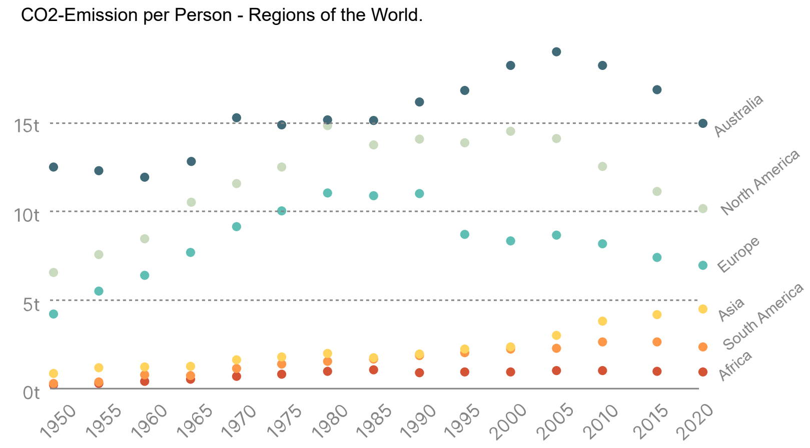
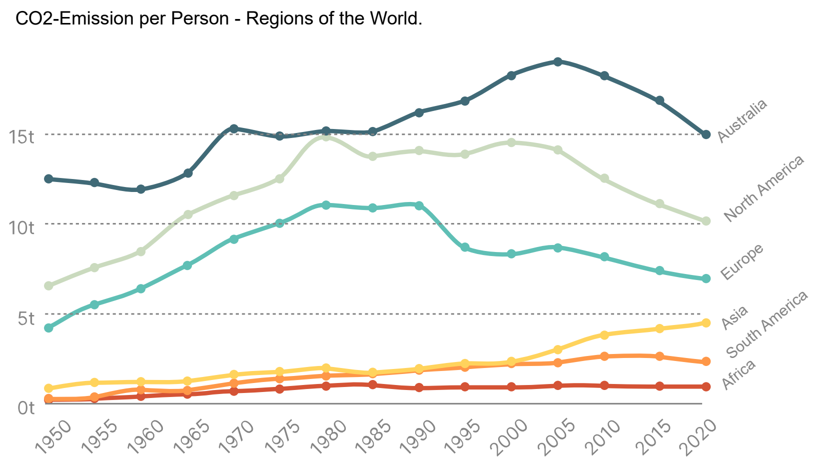



Figure 5

The clear visualization in the line chart allows for a direct comparison of CO2 emissions per person across different world regions. It becomes evident that regions such as Australia, North America, and Europe have significantly higher CO2 levels than Asia, South America, and Africa. Furthermore, the chart provides an immediate overview of the dynamics of emission trends over the years in all regions.



# The Area Chart

## Description

At first glance, the area chart resembles a line chart, but it differs in a key aspect: the values of individual data series are "stacked" on top of each other. This stacked visualization emphasizes the overall development of all categories, in contrast to the line chart, which displays each data series separately.

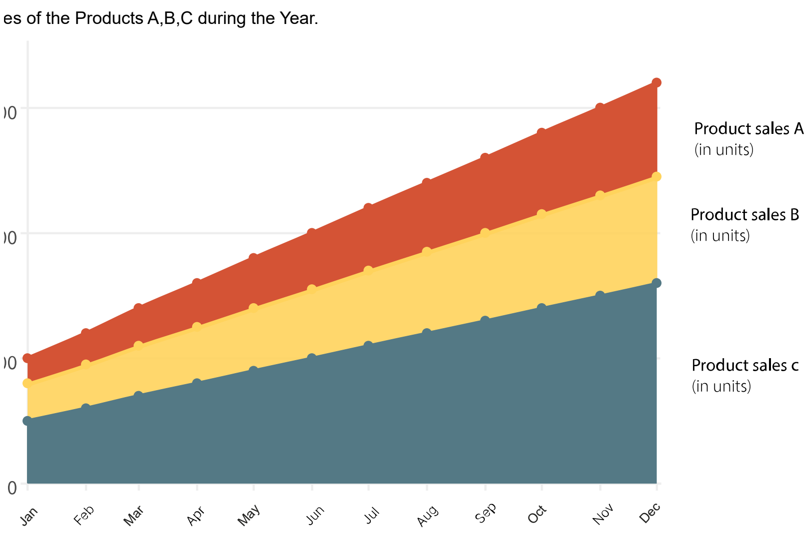
## Functionality and Representation

Bei der Erstellung eines Flächendiagramms werden die Werte wie im Liniendiagramm entlang der X-Achse zeitlich und entlang der Y-Achse nach Wertigkeit eingetragen. Allerdings werden die Datenreihen im Flächendiagramm übereinandergestapelt. Diese Darstellungsweise zeigt, wie sich alle Datenreihen zusammenaddieren und macht klar, wie jede Kategorie zum Gesamtergebnis beiträgt.

## Example

The following area chart serves as an illustration, showing the sales figures of three products over four months (January to April):

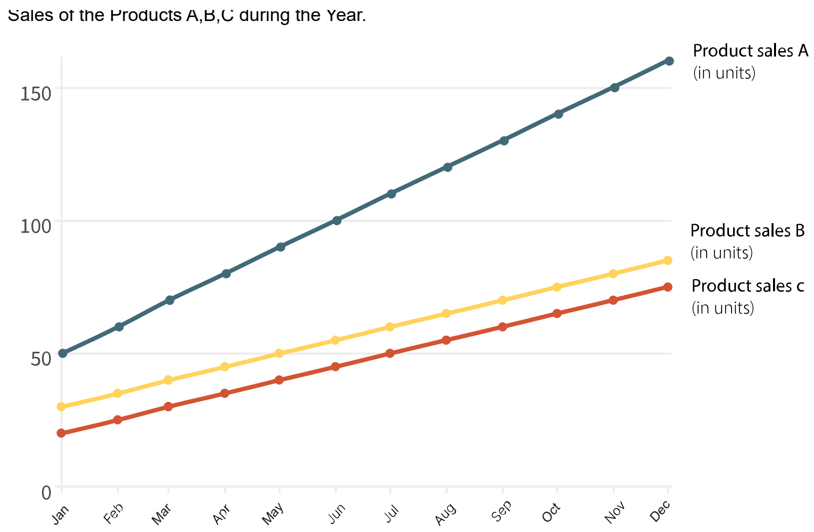
| Month | Sales A | Sales B | Sales C |
| --- | --- | --- | --- |
| January | 50 | 30 | 20 |
| Feburary | 60 | 35 | 25 |
| March | 70 | 40 | 30 |
| April | 80 | 45 | 35 |




### Figure 6

Figure 6 shows that the orange line, representing the sales of Product C, starts at a value of 100, even though only 20 units were sold in January according to the data. The reason for this apparent contradiction lies in the stacking nature of the area chart: the sales of Product C are added to those of Product B (shown in yellow) and Product A (shown in blue), which already amount to 30 and 50 units, respectively. This addition results in the total value of 100, where the line for Product C begins.
These data illustrate how total sales increase monthly and how each product type contributes to the overall result. In the area chart, the stacked representation makes the total sales effect visible, an aspect that would not be apparent in a line chart.

The same data would appear like this in a line chart, as shown in Figure 7:

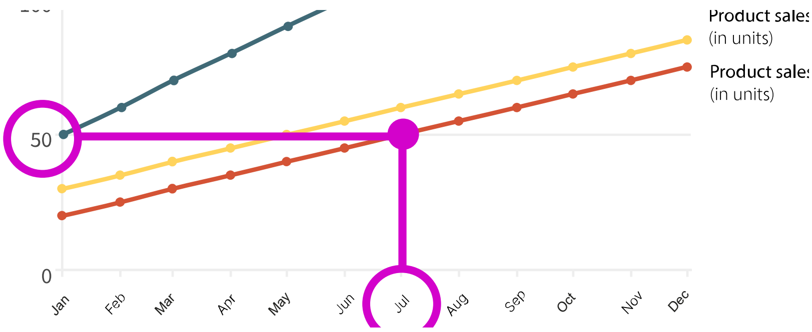



## Figure 7

In contrast to the stacked area chart in Figure 6, the line chart in Figure 7 shows unstacked data. Here, the orange line for Product C actually starts at the value of 20, the yellow line for Product B at 30, and the blue line for Product A at 50. This illustrates that in the line chart, the values are displayed separately and not added together.

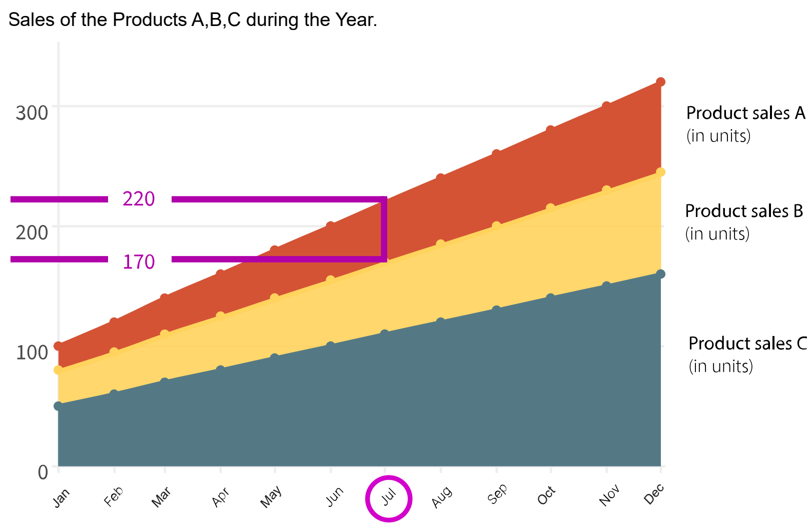
## Figure 8

To clarify, Product C is used as an example: In July, 50 units of Product C were sold. In the line chart, this sales figure is directly visible.




## Figure 9

In the area chart, two horizontal purple lines are added within the orange segment (sales of Product C) to mark the value for July for a more detailed view. These lines indicate that the value for July is 220 at the top and 170 at the bottom. By subtracting 170 from 220, we get 50, confirming that 50 units of Product C were sold in July. This method is the only way to identify individual sales figures in an area chart.



The use of area charts offers significant advantages, especially when illustrating the relationship between parts and the whole. These charts make it clear how each individual category contributes to the total. By stacking the values, the distribution of quantities becomes immediately visible, allowing viewers to quickly grasp how many products were sold in total and what share each category represents.

To better illustrate this form of representation, the example in Figure 10 shows CO2 emissions from industrial fuels in Austria. It highlights how emissions from gas, oil, and coal have developed over time and how they relate to one another.

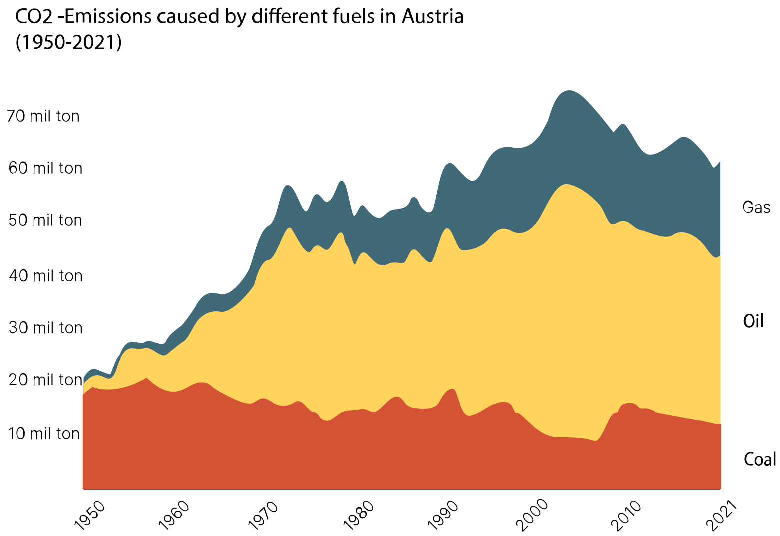



**Figure 10**

The area chart in Figure 10 impressively illustrates the changes in CO2 emissions caused by different fuels in Austria. Particularly striking is the decline in coal usage between 2000 and 2010. At the same time, oil usage has steadily increased since the 1960s. During the period from 2000 to 2010, peak levels in total emissions were reached. It is also evident that, compared to gas and coal, oil is a major source of emissions—clearly reflected in the large yellow area in the chart.

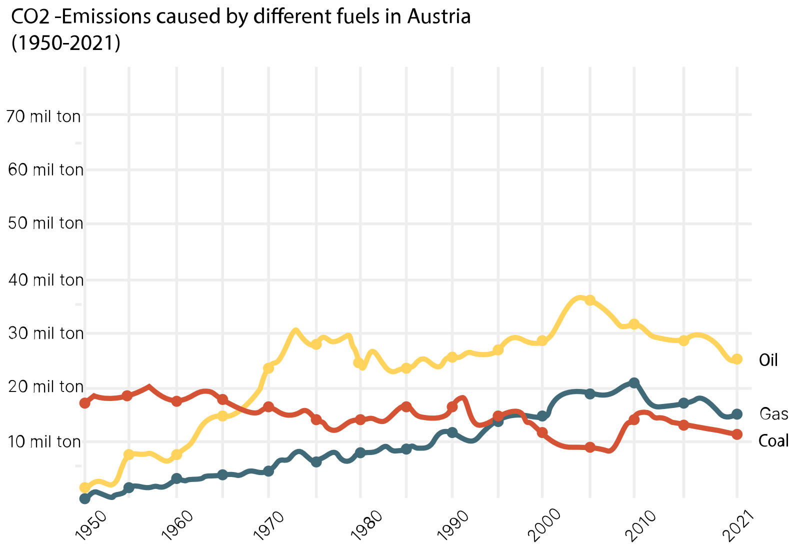



**Figure 11**

Figure 11, in comparison to the area chart in Figure 10, shows a line chart using the example of CO2 emissions in Austria. While a line chart could reveal more detailed differences between the emissions from gas, coal, and oil, the area chart more clearly conveys the overall volume of emissions. Additionally, it illustrates how the contributions of the various fuels relate to one another.

# Stream Graph

## Description

A stream graph is similar to an area chart but is arranged around a central axis, giving it a wave-like appearance, hence the name „stream graph." A stream graph provides a visually engaging, approximate overview of the distribution of data.

## Functionality and Representation

In a stream graph, the size of each wave corresponds to the values of the respective category. The axis along which the stream graph is oriented usually represents time. Colors can be used to distinguish between different categories or to convey additional information by varying the shades. By using colored „waves" with varying heights, the stream graph allows for comparison of developments within different categories over time. This type of visualization helps identify patterns and trends, although exact numerical values cannot be read directly.

Visualization

Due to its complexity, the stream graph is typically created using computer-based tools that not only process the data accurately but also ensure an aesthetically pleasing presentation. These tools help users visualize extensive information effectively and efficiently.

Figure 12 ▶

The example in Figure 12 shows five products with varying sales numbers over the course of a year. However, exact sales figures cannot be read from the chart. What is clear from the diagram, though, is that Product E (shown in red) achieved the highest sales compared to the other products.

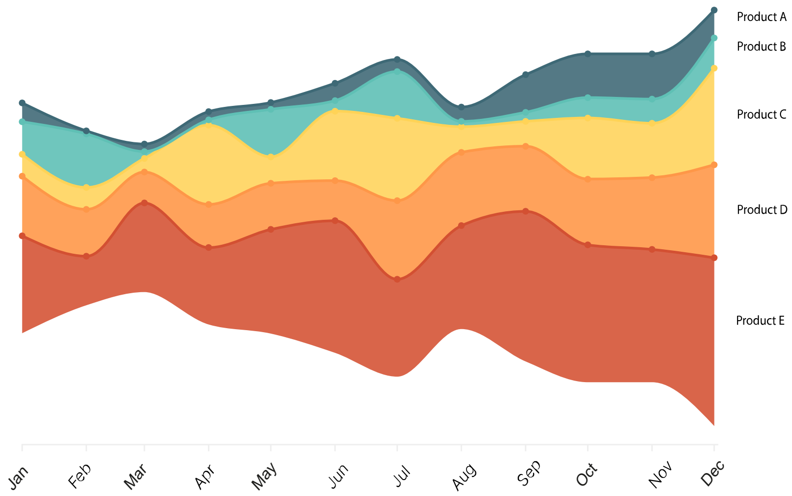



Figure 13 ▶

Figure 13 also shows that sales of Product B (shown in light green) nearly came to a halt between March and April as well as between August and September. This is an important insight, even though the exact values cannot be read.