

## Part 1.8 | Time Series

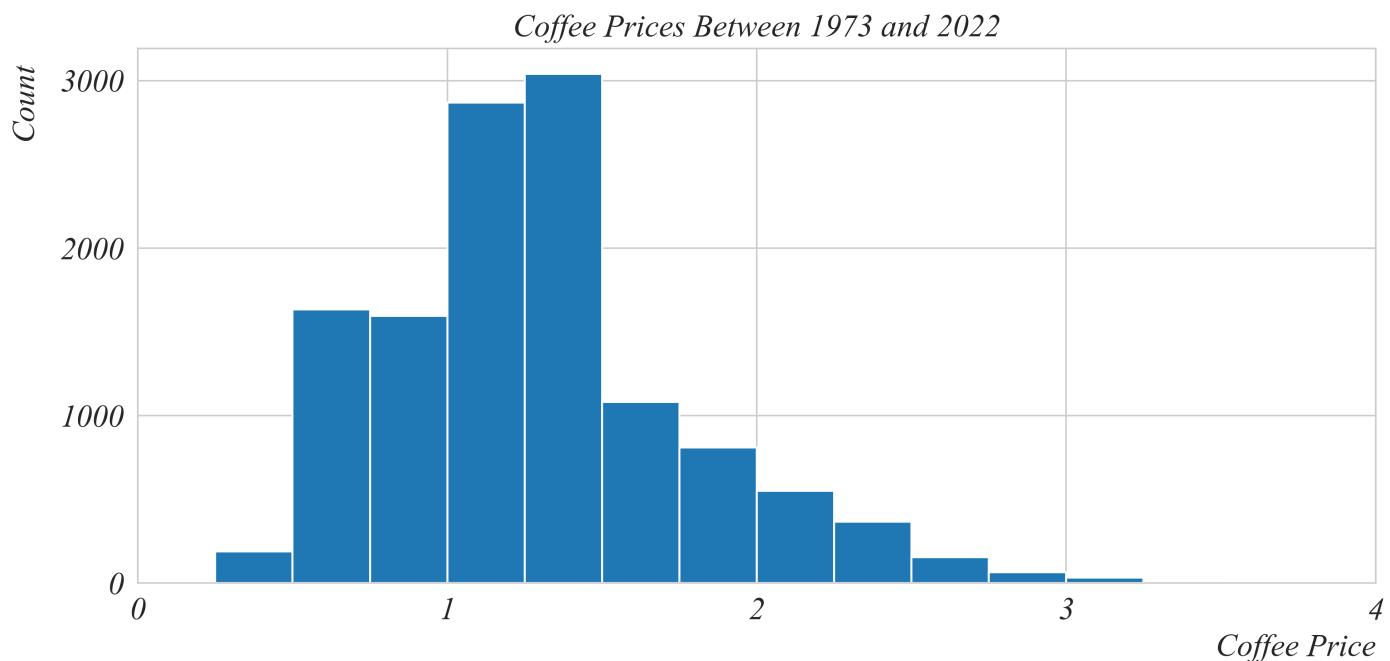
---

### *Long-Term Behavior*

The price of coffee beans is always on the mind of coffee shop owners. How have these prices changed over time?

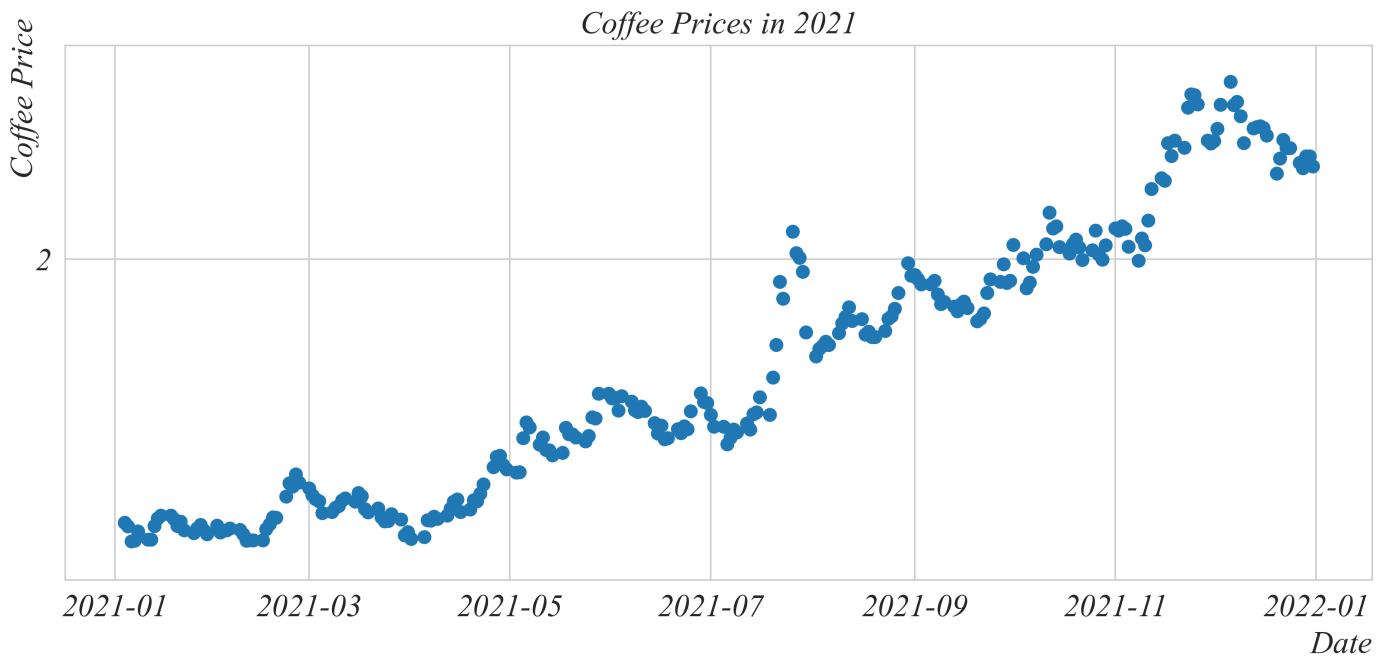
This table contains coffee prices in US dollars per pound recorded on every workday between 2001 and 2021.

We could plot this data on a histogram to get a sense of the types of prices we see frequently.

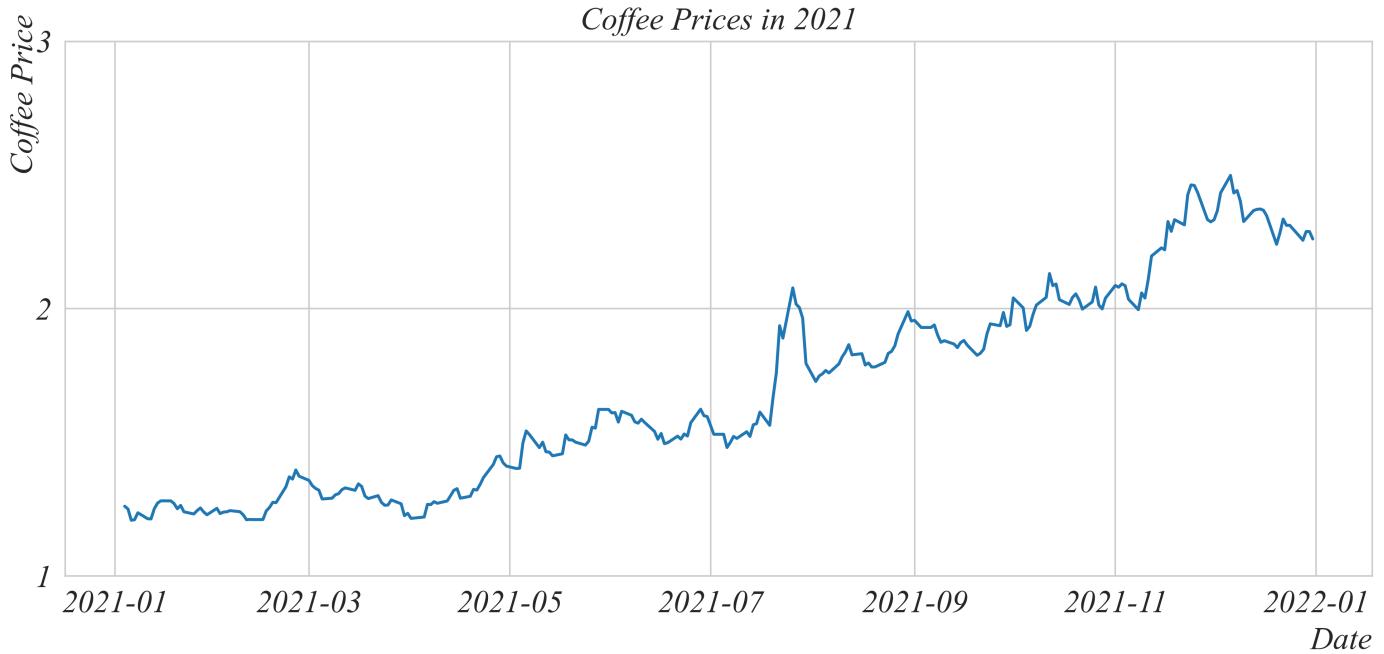


But this isn't great for a couple of reasons. First, a dollar today isn't comparable to a dollar in 1980. Prices change through inflation so as we'll do later, we would need to do a data transformation. But second, histograms do not show the relationship between time and the values. In some sense, year is a second variable and maybe we actually care about the *relationship* between the two variables of time and prices.

What have we done so far with bi-variate data? We've used a scatterplot. Lets scatter price on the vertical axis and year on the horizontal axis. And for now, lets focus on 2021.



This gives us a better picture of prices. We can see that prices are rising throughout the year. A scatterplot is nice for showing relationships between variables (time and price here) but we can improve the figure by showing the relationship between points using a line graph. Last time we used scatterplots to show distinct and unrelated points. A line graph on two dimensions gives us the relationship between related points.

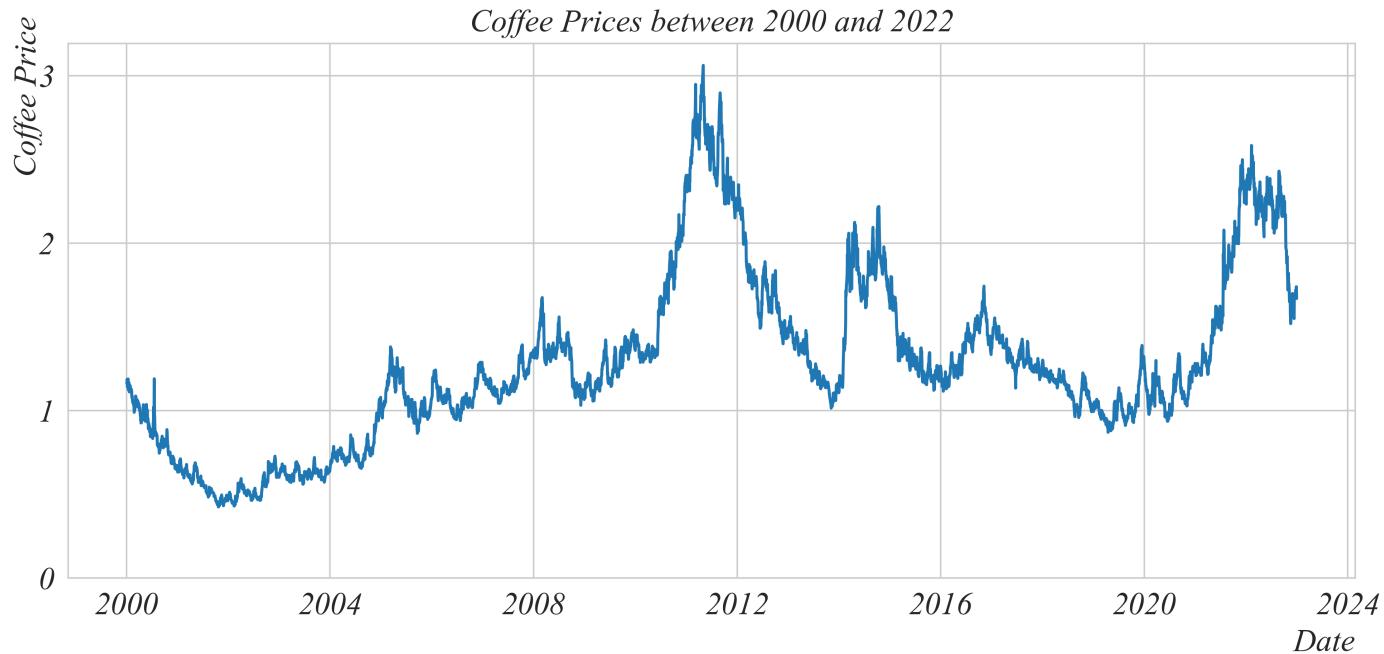


The resulting line graph illustrates how coffee prices changed over 2021. An ordered sequence of values taken at equally spaced points in time is called a **time series**.

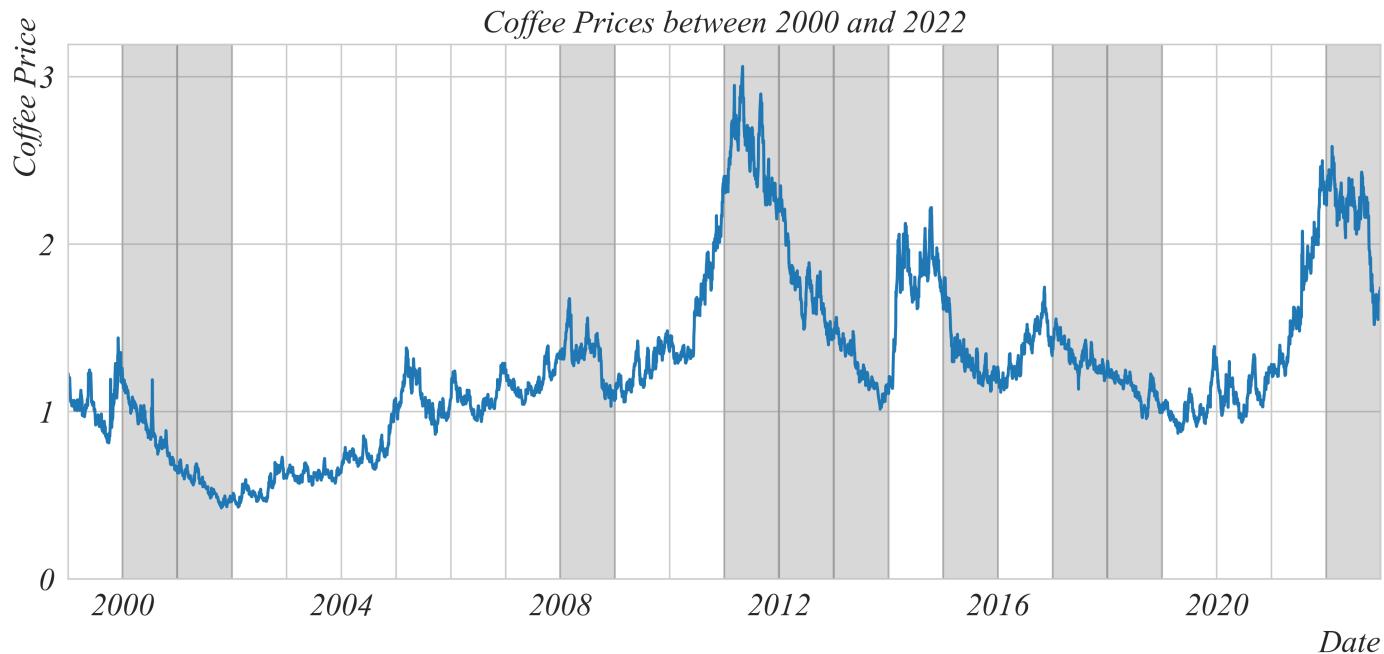
How did coffee prices change during 2021? Although daily prices in 2021 oscillated, the general pattern was increasing.

The general increase we see in the line graph is the **trend**. A **trend** is the general, long-term pattern of a time series. A time series can have an increasing or a decreasing trend, or no trend at all.

Let's see if the 20-year trend was also increasing.

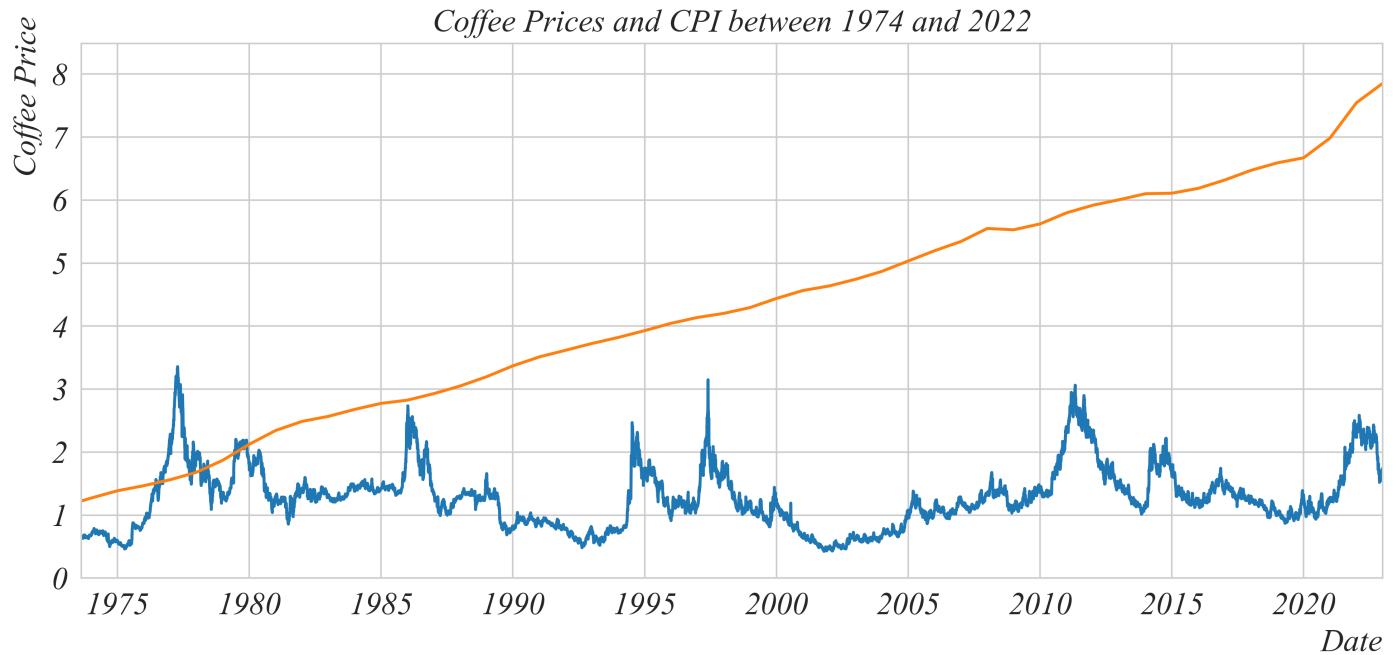


This line graph represents daily coffee prices in US dollars per pound recorded from 2001 to 2021. Which intervals was the trend **negative**?

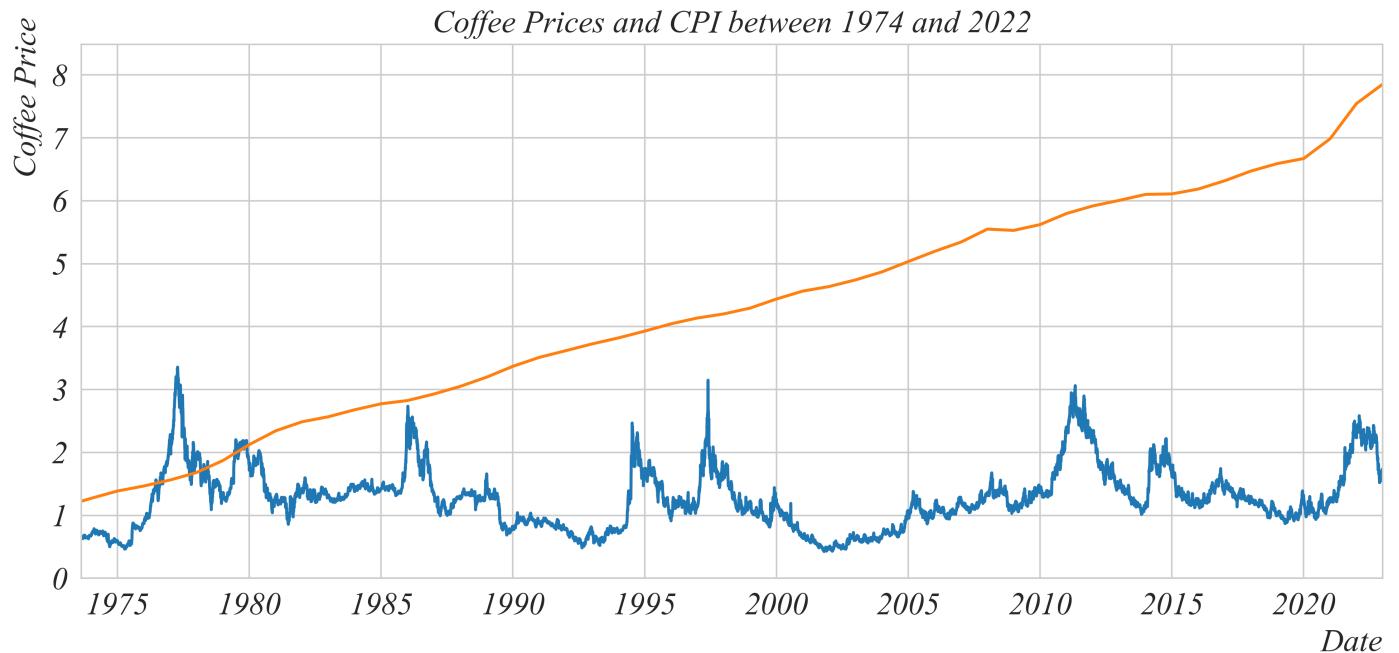


The trend is negative when the prices generally fall over a given period. Parts of a time series can have different trends than the overall trend of the time series.

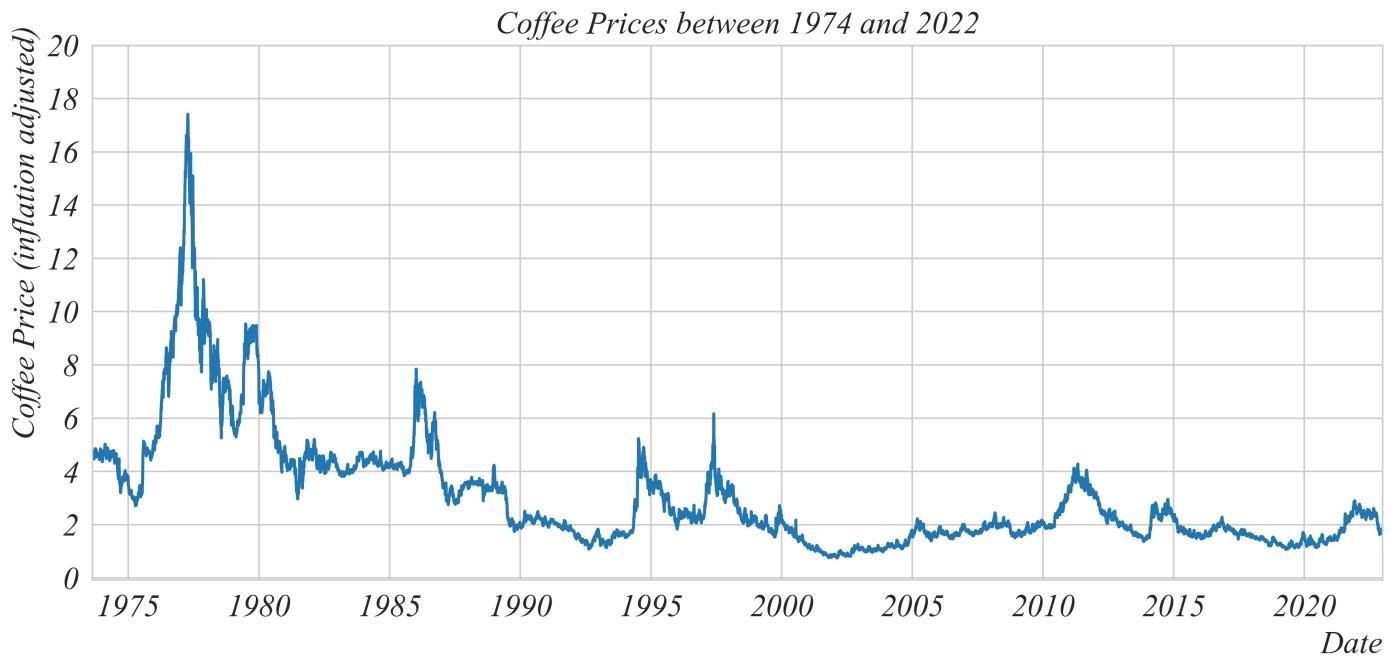
Lets open this up going back to the 70s. What do you notice about this figure?



If we plot the value of a dollar in 1970 over this period we can see that prices earlier in time mean something very different.



To adjust for inflation, we divide the value of today's dollar yesterday by the price yesterday.



This tells a very different picture. Prices of coffee have fallen since the 1970s but have remained relatively stable since 2000.

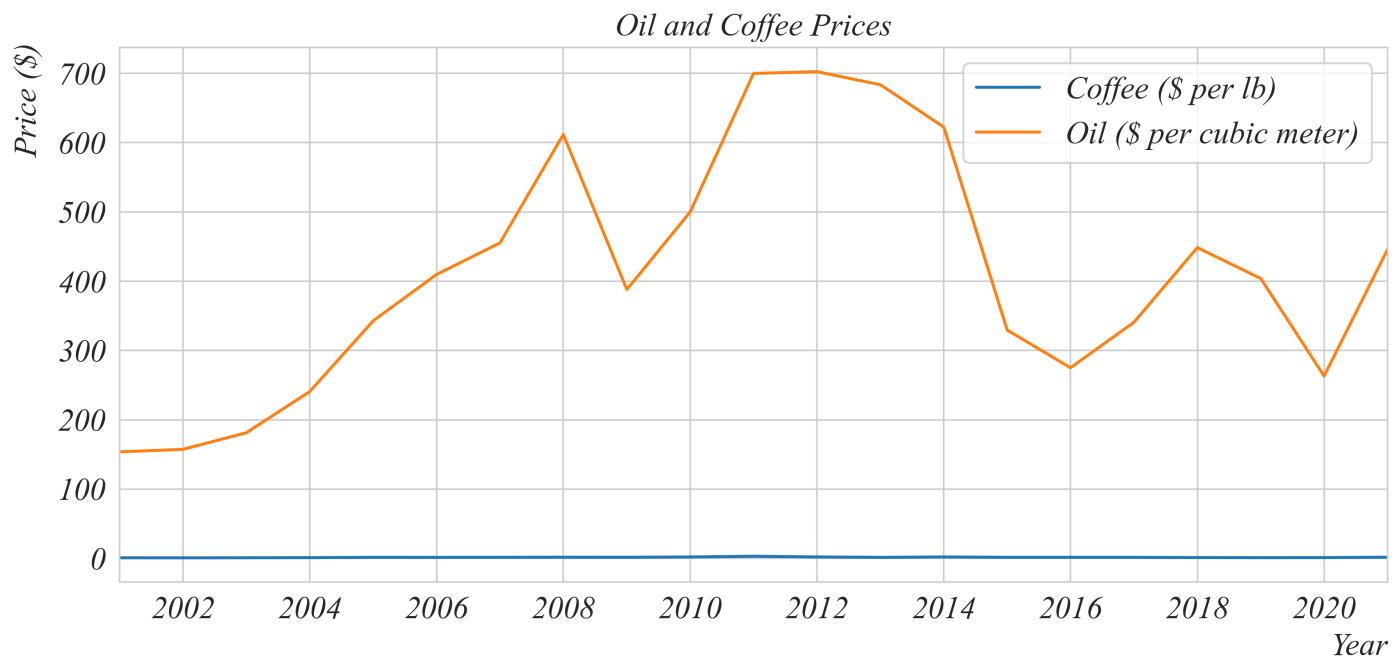
## *Coffee vs. Oil*

Next, we'll compare coffee prices with the prices of another commodity: oil. A **commodity** is a product that can be traded or sold, such as coffee or oil.

Coffee and oil prices depend on many factors, from transportation costs to geopolitical events to economic conditions. Let's see how the prices of these two commodities relate to each other.

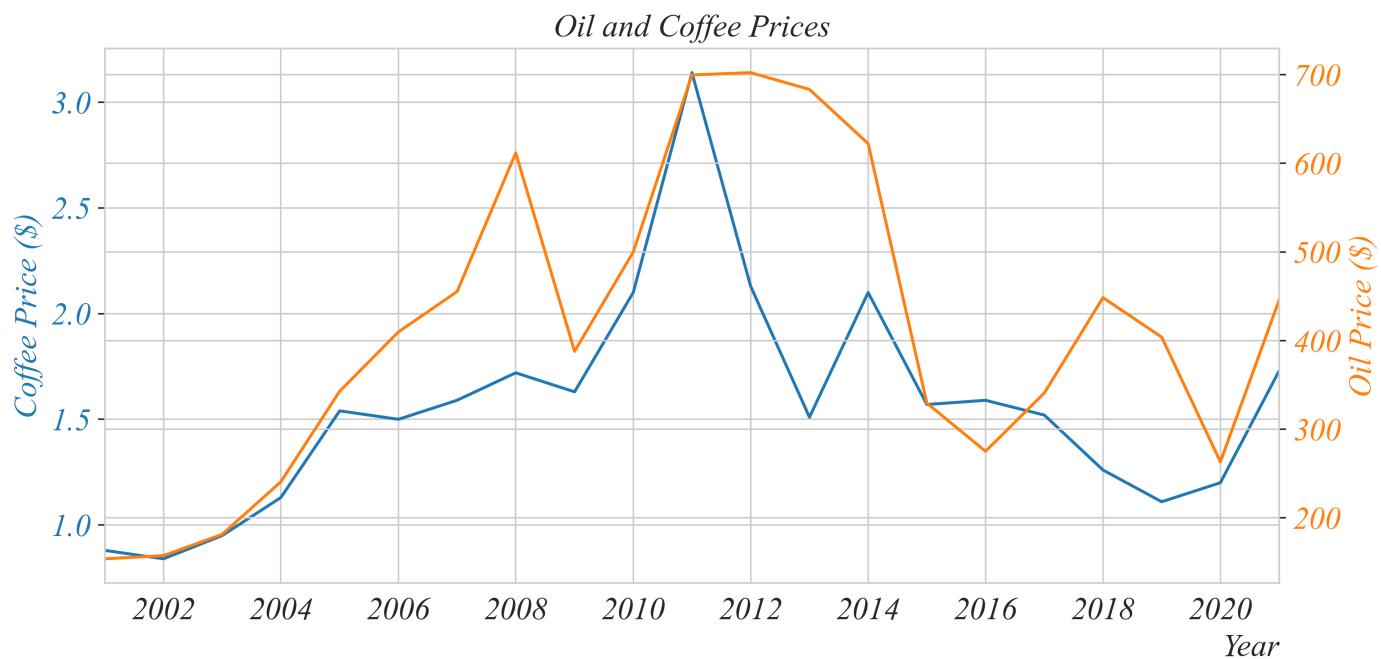
To focus on the bigger picture, we'll consider the average yearly prices of coffee (in US dollars per pound) and crude oil (in US dollars per cubic meter).

We can see that oil is much more expensive in these units. We can see this because the crude oil line is consistently higher than the coffee line, indicating a higher unit price.

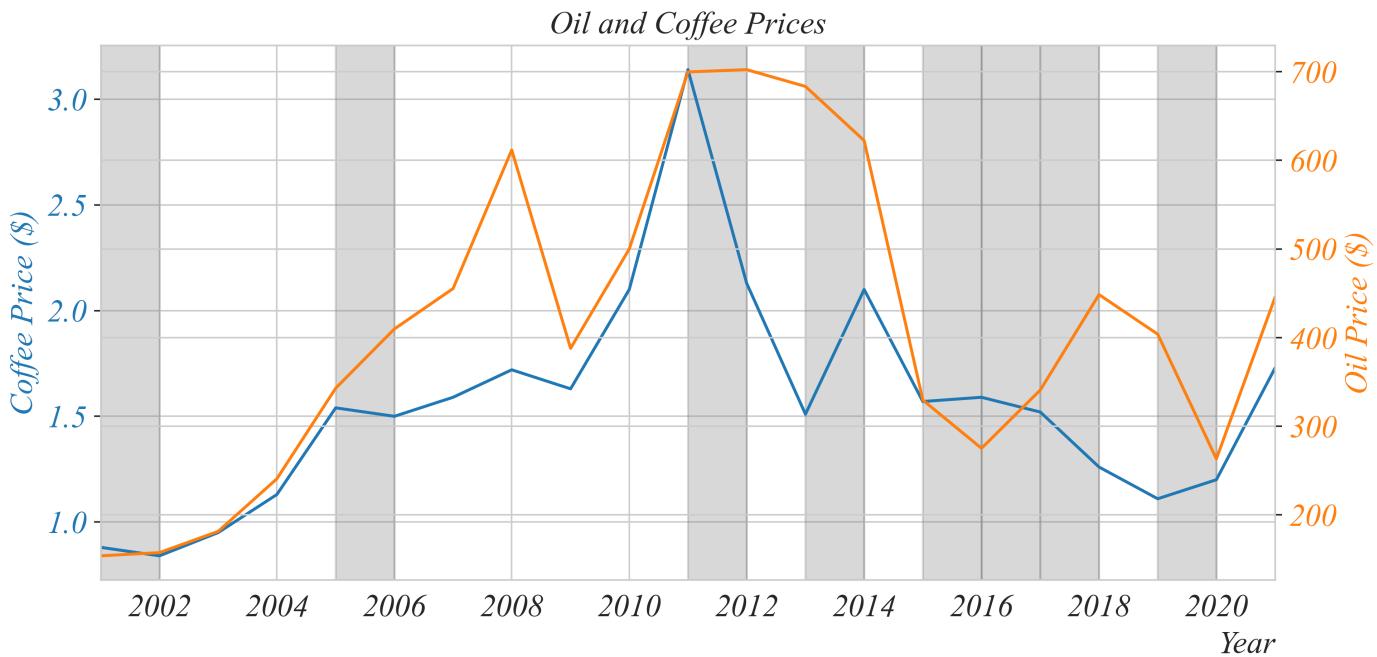


What can be inferred about the two time series? It's difficult to tell. The unit of oil is always much more expensive than the unit of coffee making the variations in coffee prices negligible when compared to oil prices. So, we can't use this graph to analyze the patterns of coffee prices.

Because the values of the two time series differ in orders of magnitude, the previous graph didn't show the comparatively small variations in coffee prices. We can fix this graph by adding a separate *y*-axis.



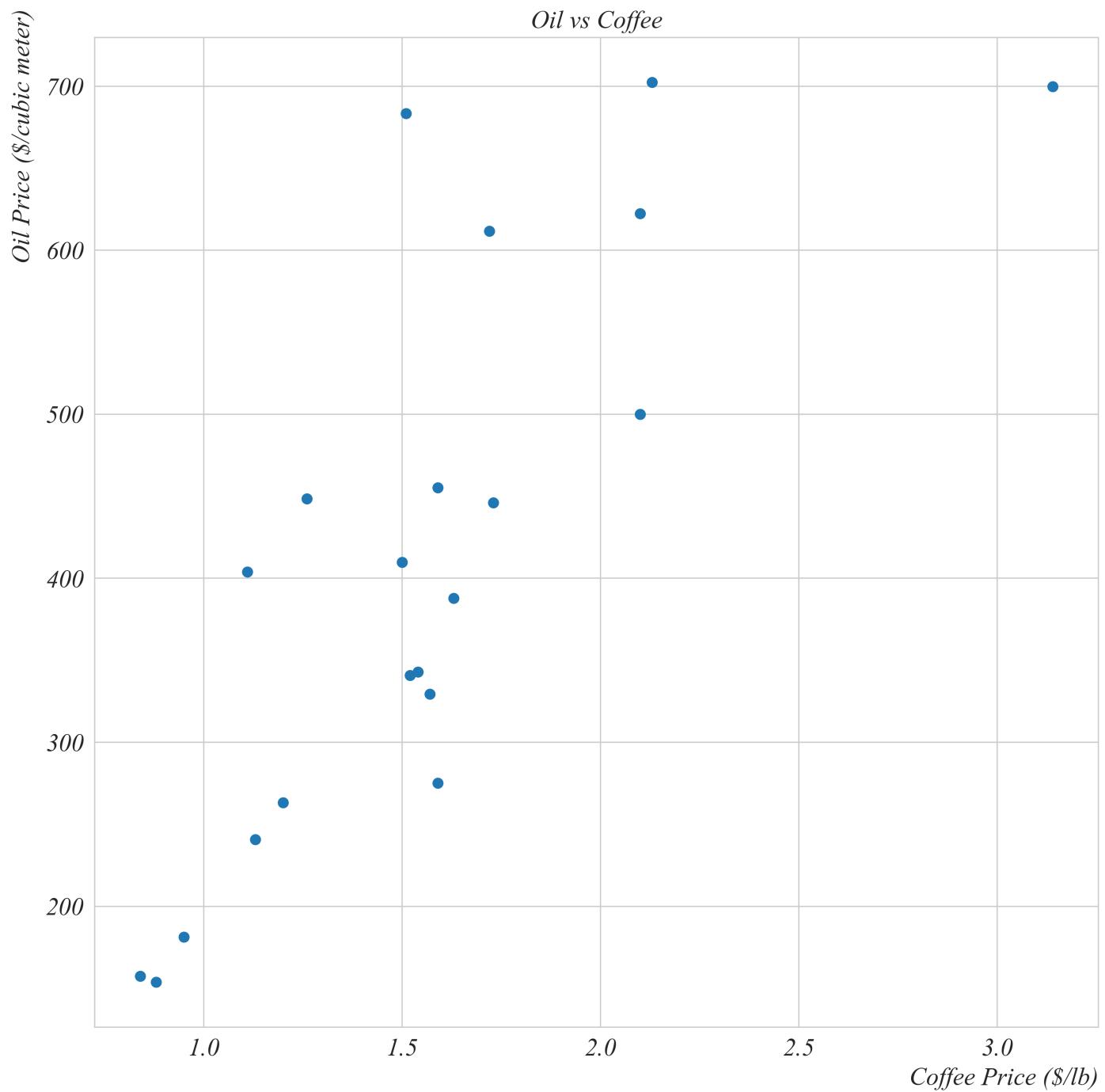
Most of the time, when the price of oil increased, so did the price of coffee, and vice versa. Which years did oil and coffee prices move in **opposite** directions?



In 2001 coffee decreased and oil increased. In 2002, 2003, and 2004, both prices increased. In 2005 oil increased and coffee decreased. ....

This is one way of understanding the relationships between prices. But we've already talked about another one: we can think of each year's price as a separate point in a scatter plot.

The line graph shows that in general, lower coffee prices corresponded to lower oil prices. This is reflected in the scatter plot.



What do you think is true?

- Coffee price determines oil price.
- Oil price determines coffee price.
- Coffee and oil prices are related.

Coffee prices could impact oil prices, oil prices could impact coffee prices, they both could be influenced by a different, unknown process, or their graphs could look similar by chance.

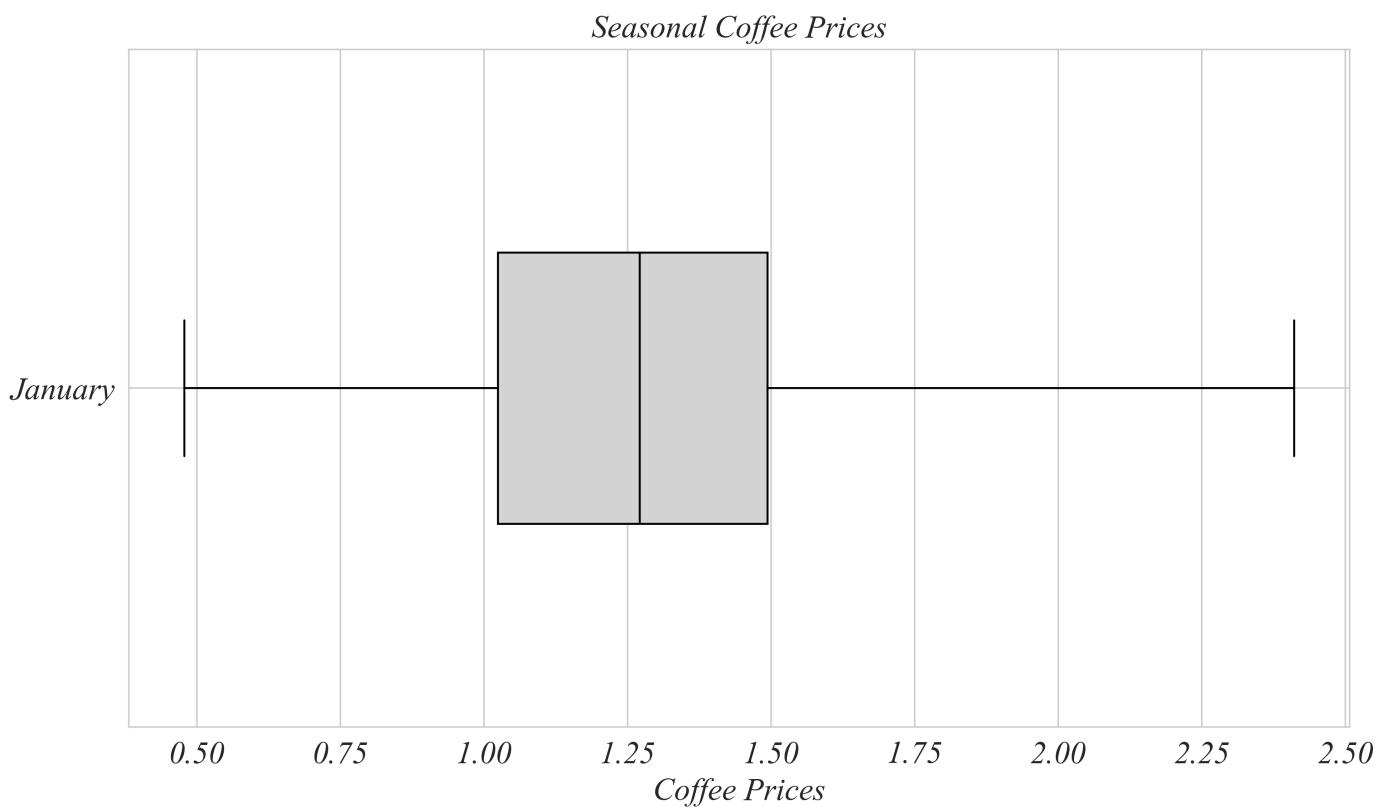
We need to be careful with drawing too many conclusions from the data! We'll talk about how we might disentangle correlation from causation later in the semester. But we always need to be careful.

Next, we'll look closer at seasonal changes in coffee prices.

## *Seasonal Variation*

Coffee prices fluctuate a lot throughout the year. If we're trying to set prices for four months from now, what data do you think we should be looking at? Does coffee tend to be more expensive in some months than in others?

We'll start by gathering all January prices from all years. Filter for January.

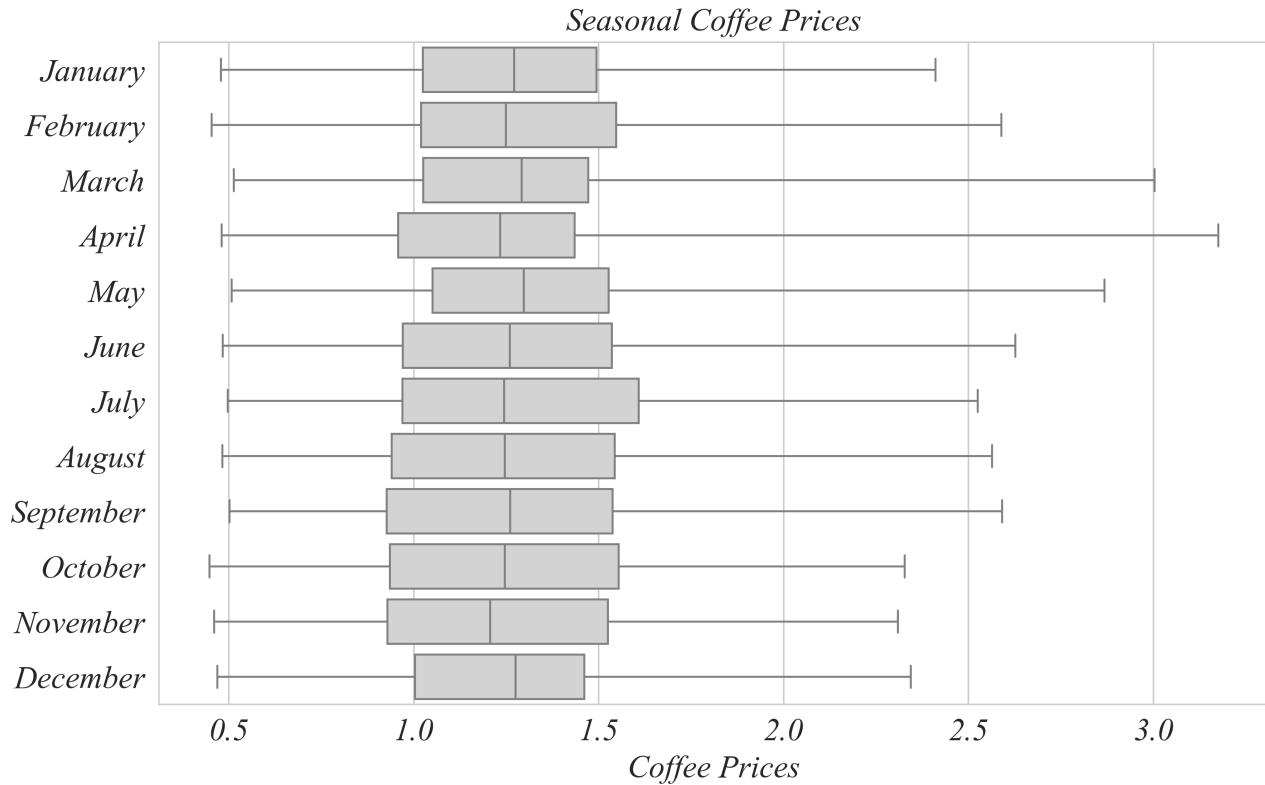


What was the median coffee price per pound across all years in January?

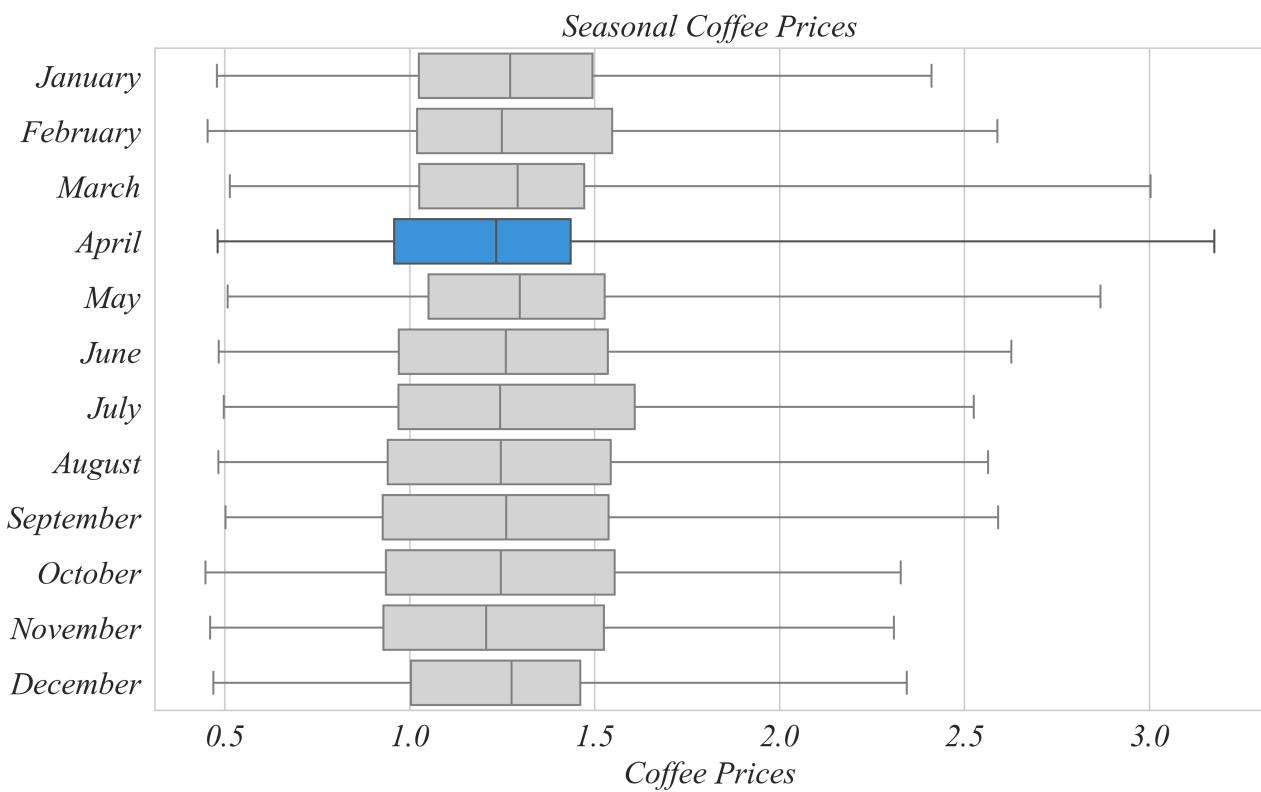
- About \$0.25
- About \$1.25
- About \$1.50
- About \$1.75

The vertical line in the box represents the median. So, the median January price for a pound of coffee was about \$1.25.

To compare the prices between months, let's look at boxplots for each month.

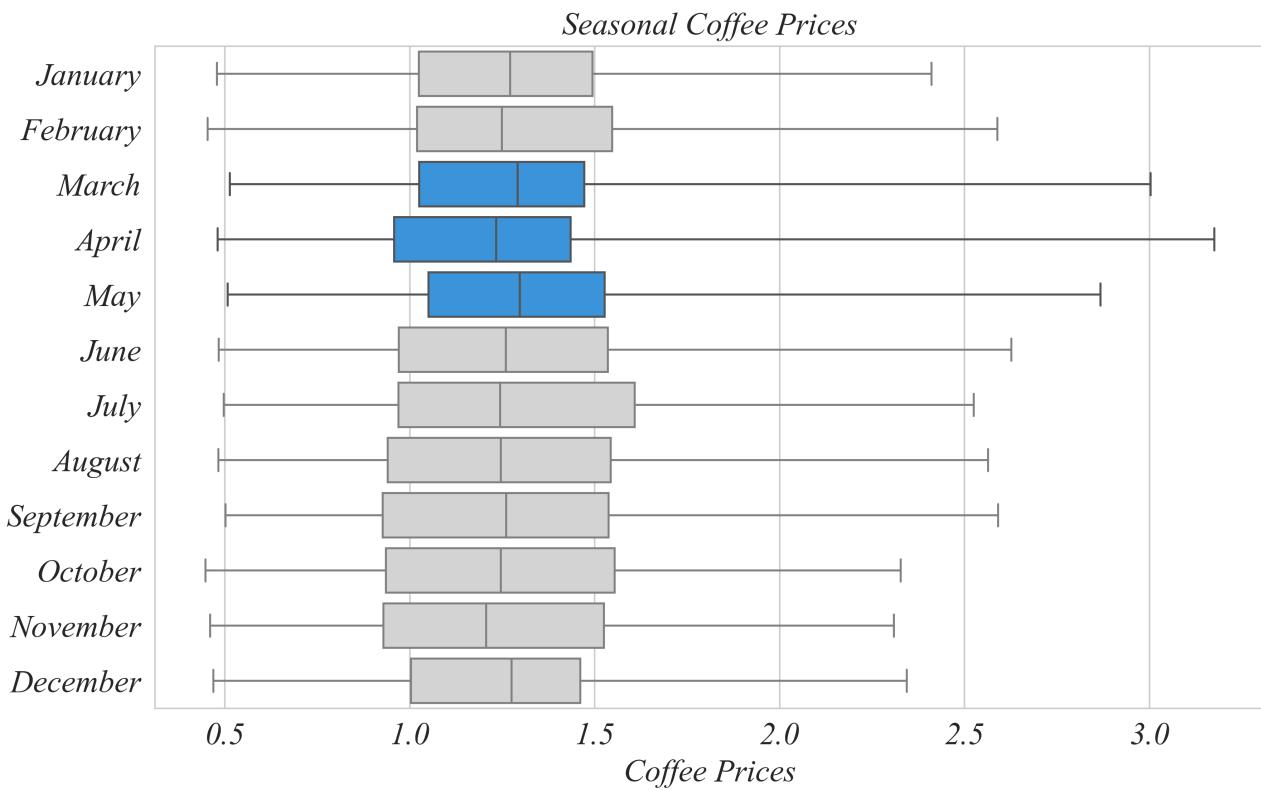


In which month was the record highest price set? The highest price in each month is represented by the upper end of the upper whisker. The largest of these twelve maximum prices occurred in April.



In which season are prices most spread out?

- Spring
- Summer
- Fall
- Winter



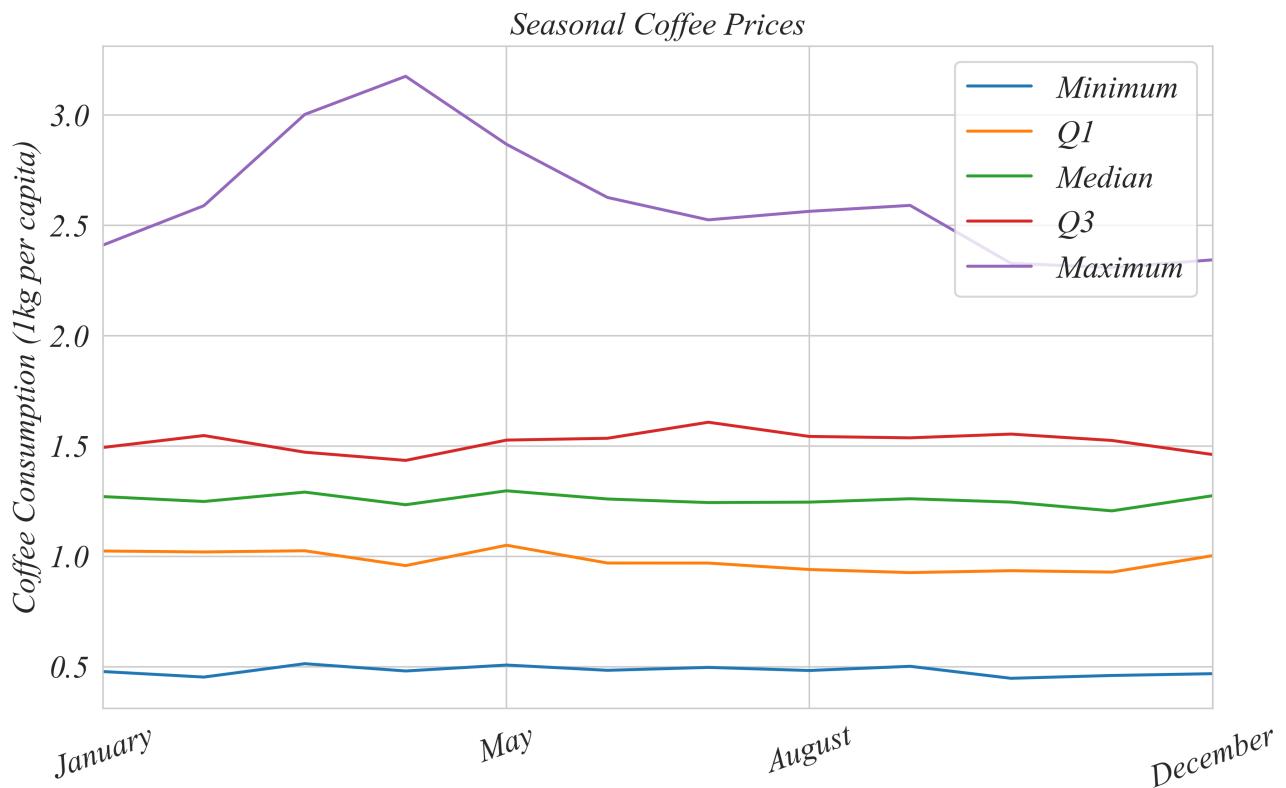
To assess the spread, we focus on the whiskers, ignoring the boxes. The minimum prices were similar across months, but spring months had the highest maximum prices. This makes the total lengths of whiskers — or the spread — the largest in spring months.

From boxplots, we learned that in spring months, coffee prices were most volatile.

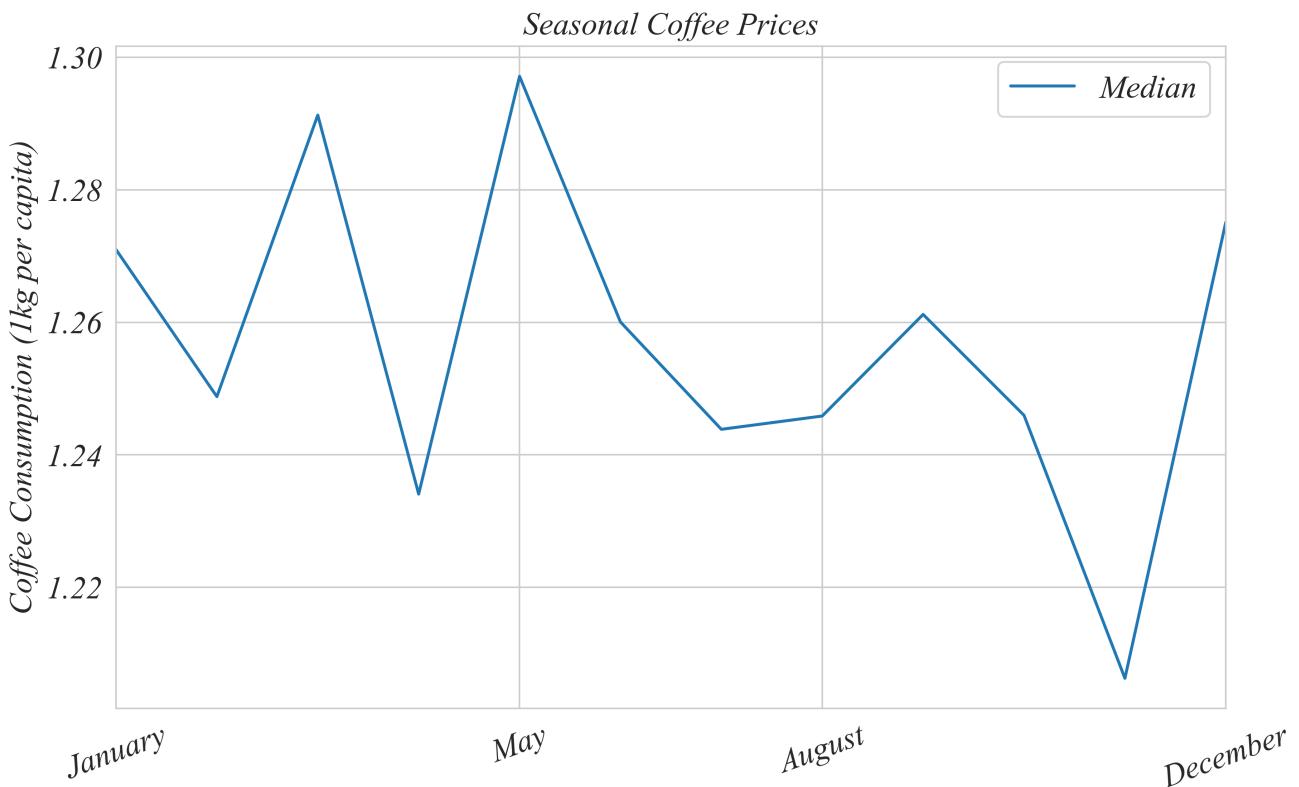
What can you say about the medians?

It's hard to compare the medians, as they don't change much compared to the full range of the data.

Instead of using boxplots, each with their own box and whisker, we can take the quantiles they represent and plot them on a line graph. This shows the minimum, median, maximum and quartiles for each month.



We can remove the other statistics from the plot to see the variation in the median more clearly.



With only the median displayed, we can see its variation more easily. What is the difference between the largest and the smallest median price per pound?

- About \$0.09
- About \$0.9
- About \$9

The largest and the smallest median prices appeared in May and November, respectively. The approximate difference in dollars between these two prices is

$$1.30 - 1.21 = 0.09$$

Notice that we don't need to know the exact medians to eliminate the other two answer options — only 0.9 has the right order of magnitude.

Boxplots helped us compare the spread of the prices in different months. To zoom in on the typical prices, we plotted just the medians.

The most appropriate visualization is a plot that displays only the necessary information.