# Part 1.3 | Timeseries

Coffee Prices Through Time

The price of coffee beans is always on the mind of coffee shop owners. How might you use the historical data on coffee prices to help you plan for coffee prices two quarters from now?

Opening the dataset, we can see a column with dates and prices per pound in dollar recorded every workday. This is a numerical variable. So far we've used histograms and boxplots as the main summarization tools for cross-sectional numerical variables. A histogram of prices gives us a sense of the frequency of coffee prices.

Coffee Prices Between 1973 and 2025

Coffee Price ($)
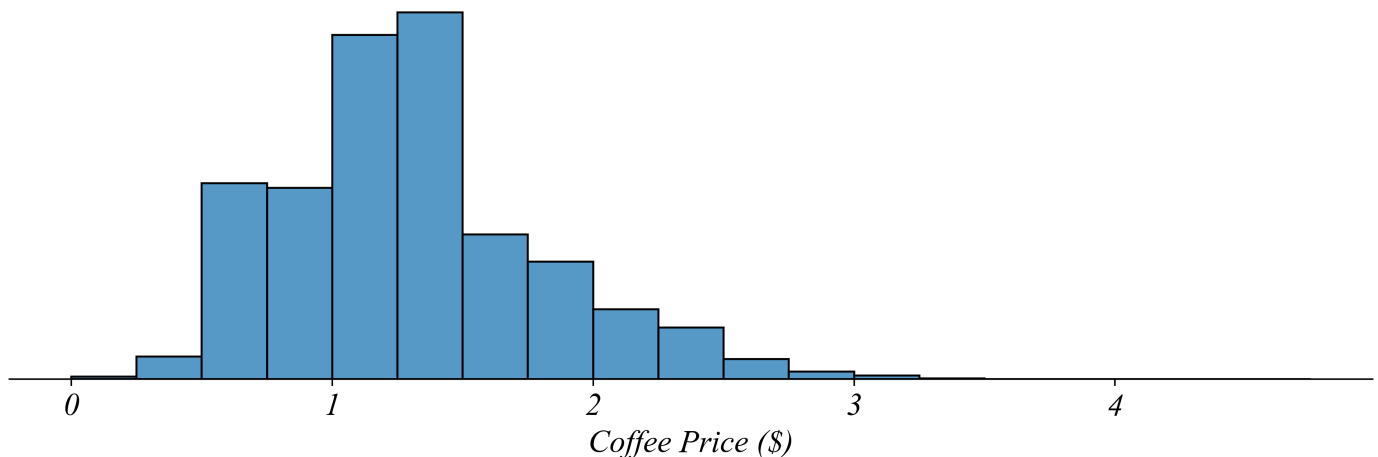
But this isn't great for our purposes for a couple of reasons. First, a dollar today isn't comparable to a dollar in 1980. Prices change through inflation. So as we'll do later, to compare a price today to a price in 1973 we would need to adjust for inflation. This is what we call a data transformation, what we'll cover in Part 1.5. But second, histograms do not show us the relationship between prices *through time*.

We care about the *sequence* of prices. Lets focus on 2021 for a second. Lets use the horizontal axis to show the sequence through time and the vertical axis to show the price at that point in the sequence.
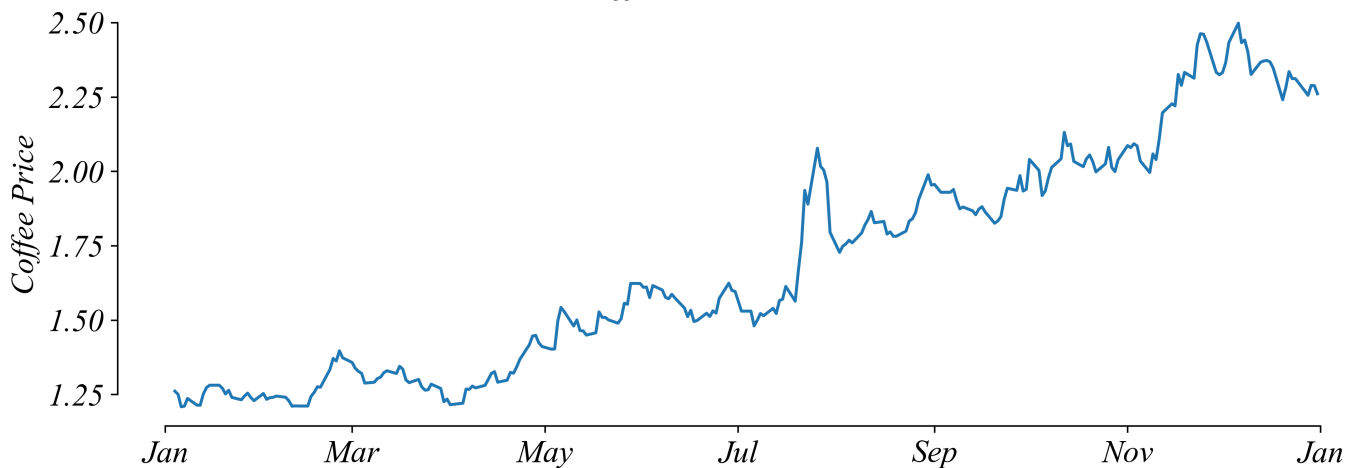
Coffee Prices in 2021
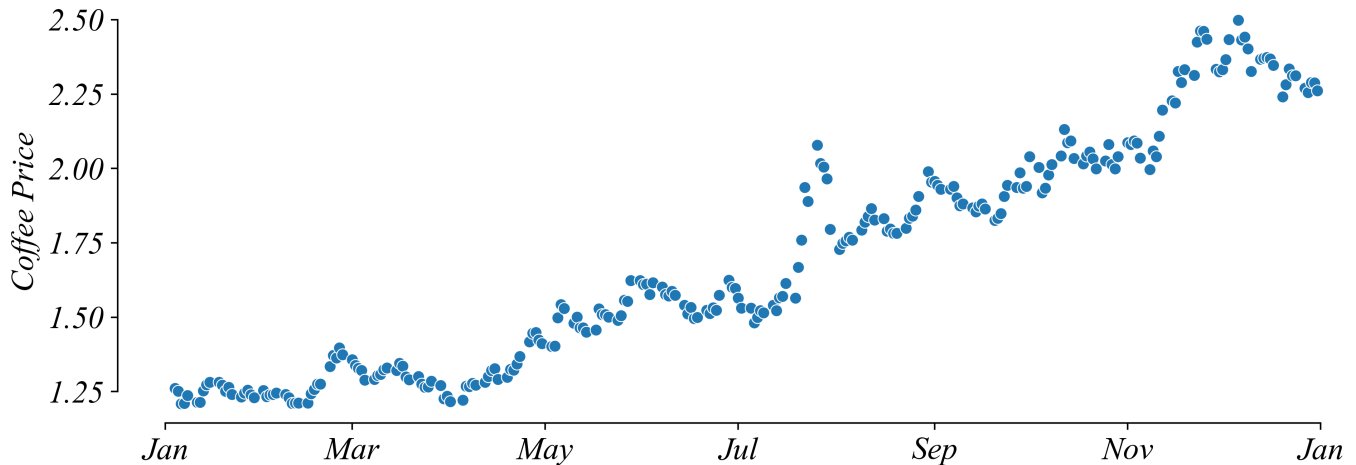
This gives us a better picture of prices. We can see that prices are higher later in the year. This is what we call a **scatterplot**, a tool we'll return to when we want to show the relationship between variables.

How might we improve this figure? Lets show the *sequence relationship* using a **line graph**, visualizing the relationship between related points.
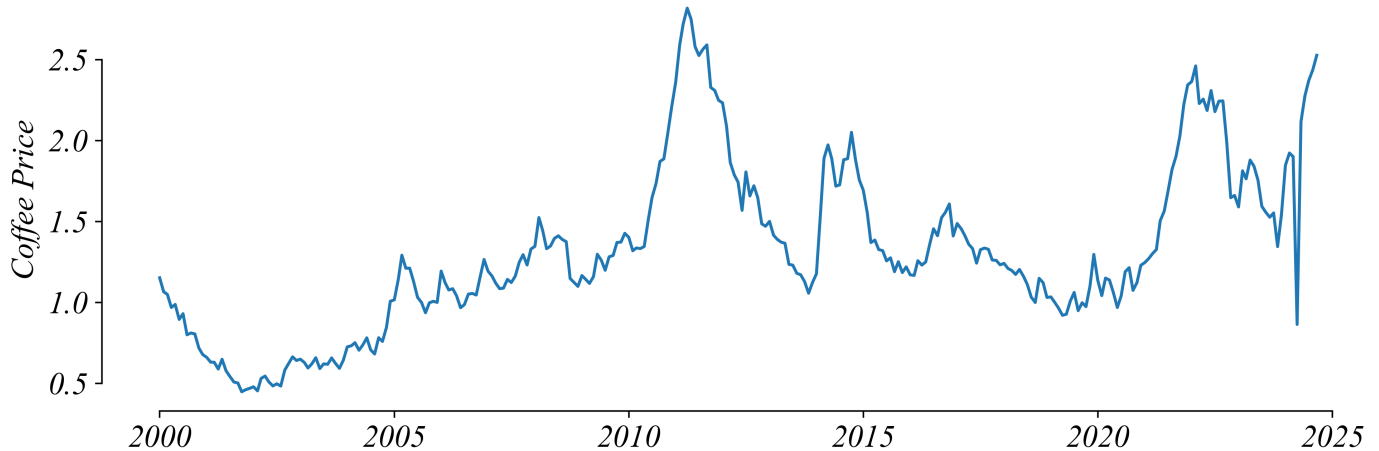
Coffee Prices in 2021

The resulting line graph illustrates how coffee prices changed over 2021. An ordered sequence of values taken through 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.

We can use any interval of time we like. Let's zoom out all the way to 2000 to see if the trend over the last 25 years is also increasing.

Coffee Prices Between 2000 and 2025

This line graph represents daily coffee prices in US dollars per pound recorded from 2001 to 2025. This shows us some longer term trends. We can see that while there was an upward trend in 2021, the overall trend was much weaker. This longer interval makes it possible to see shorter trends within the larger trend. Not every trend was positive. Which intervals was the trend **negative**?

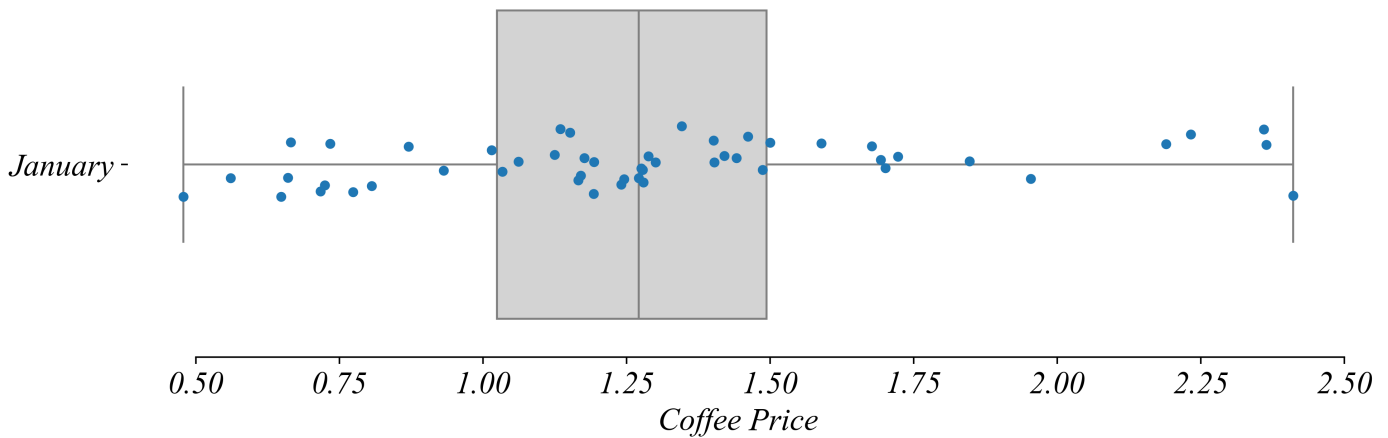## Coffee Prices Between 2000 and 2025
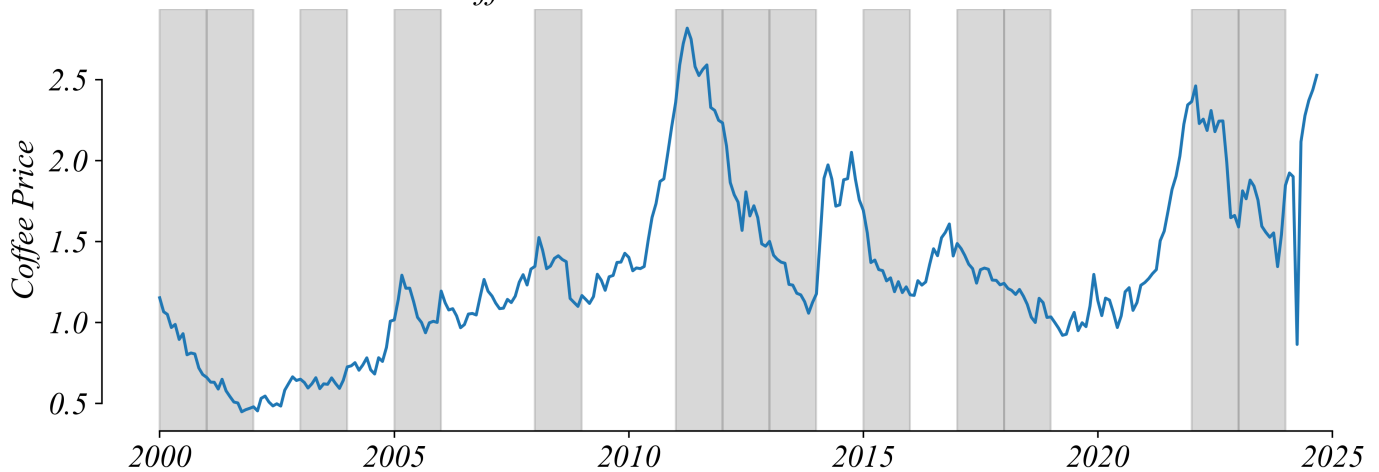
Seasonal Variation

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?

Just as a warmup lets focuss on January.

Distribution of January Coffee Prices

What was the median coffee price per pound across all years in January? The vertical line in the box represents the median. So, the median January price for a pound of coffee was about $1.25.
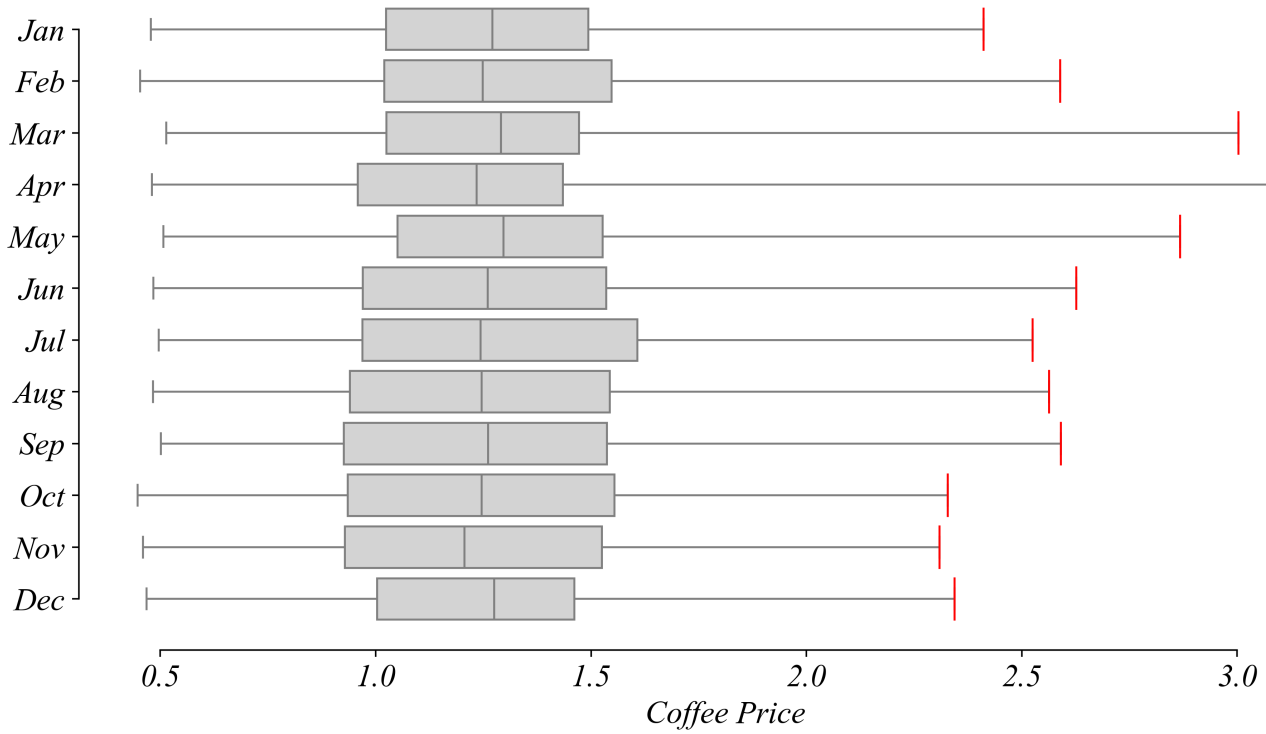
To compare the prices between months, let's look at boxplots for each month.

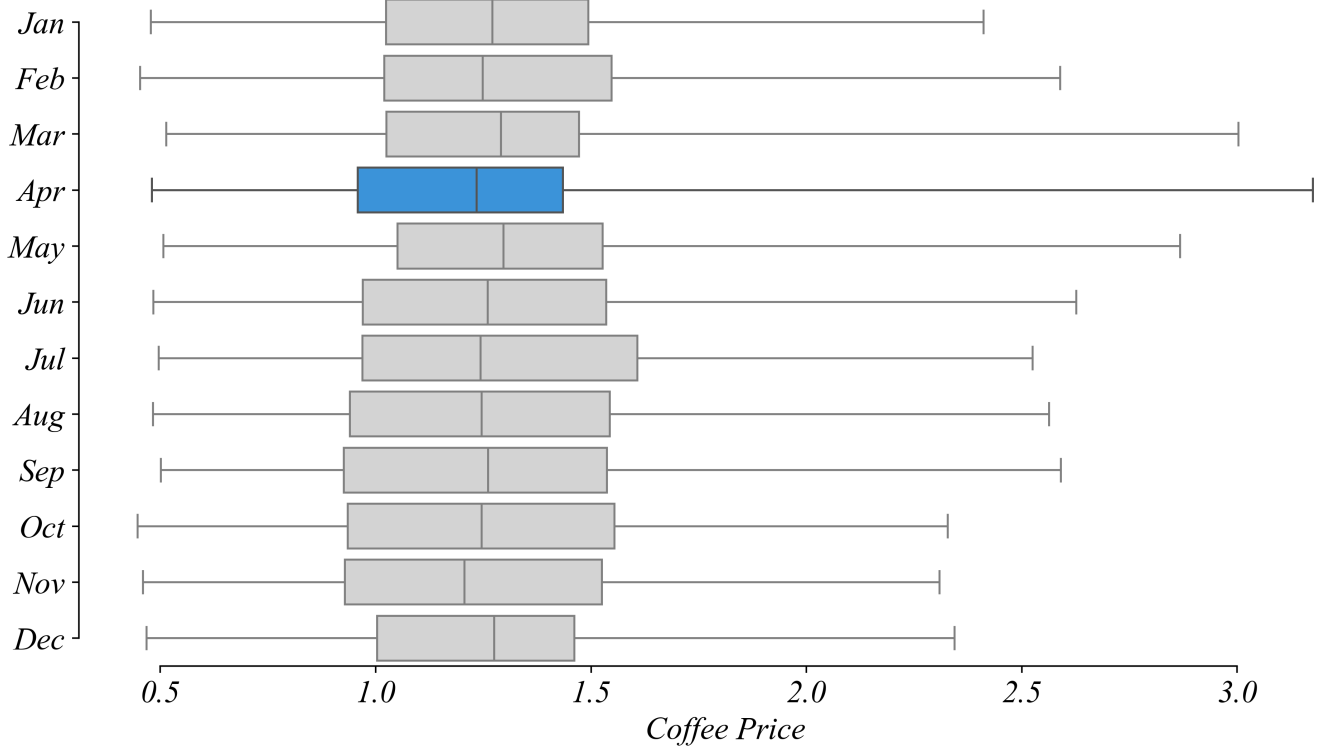## Coffee Price Distribution by Month
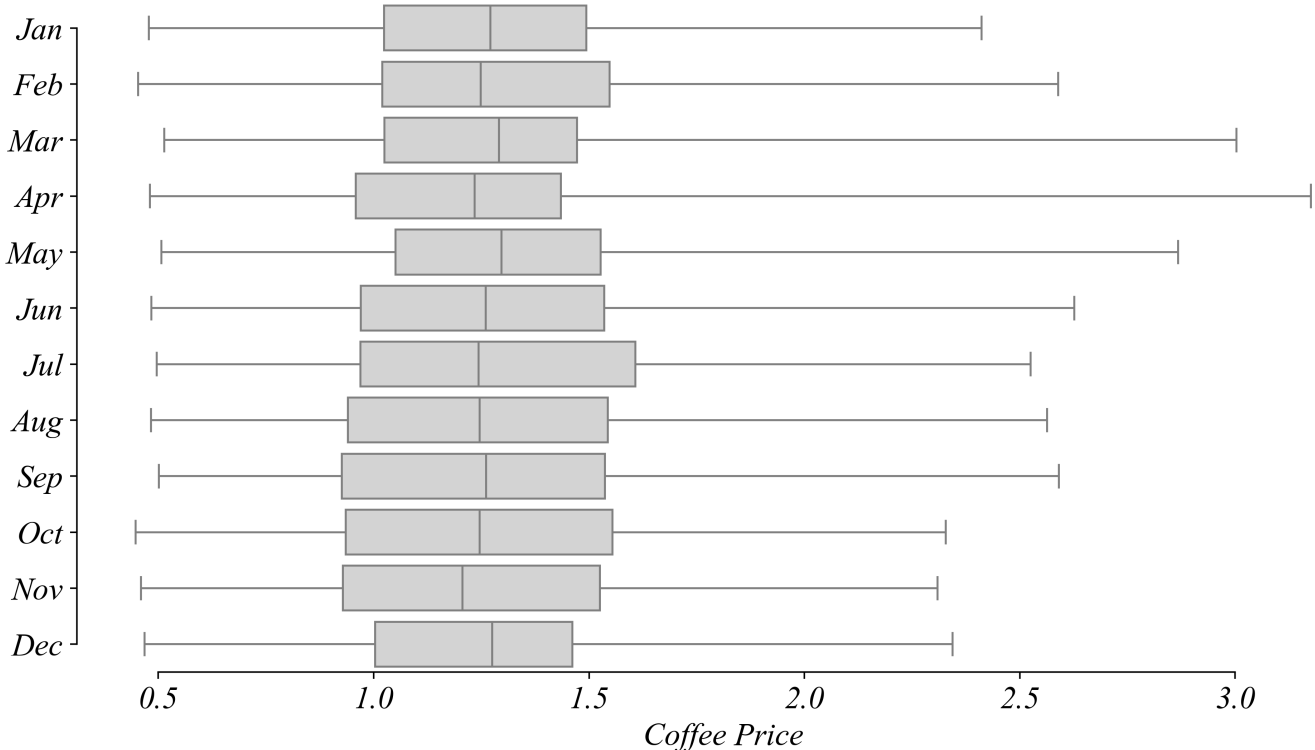
In which month was the record highest price set? The highest price in each month is represented by the upper end of the upper whisker.
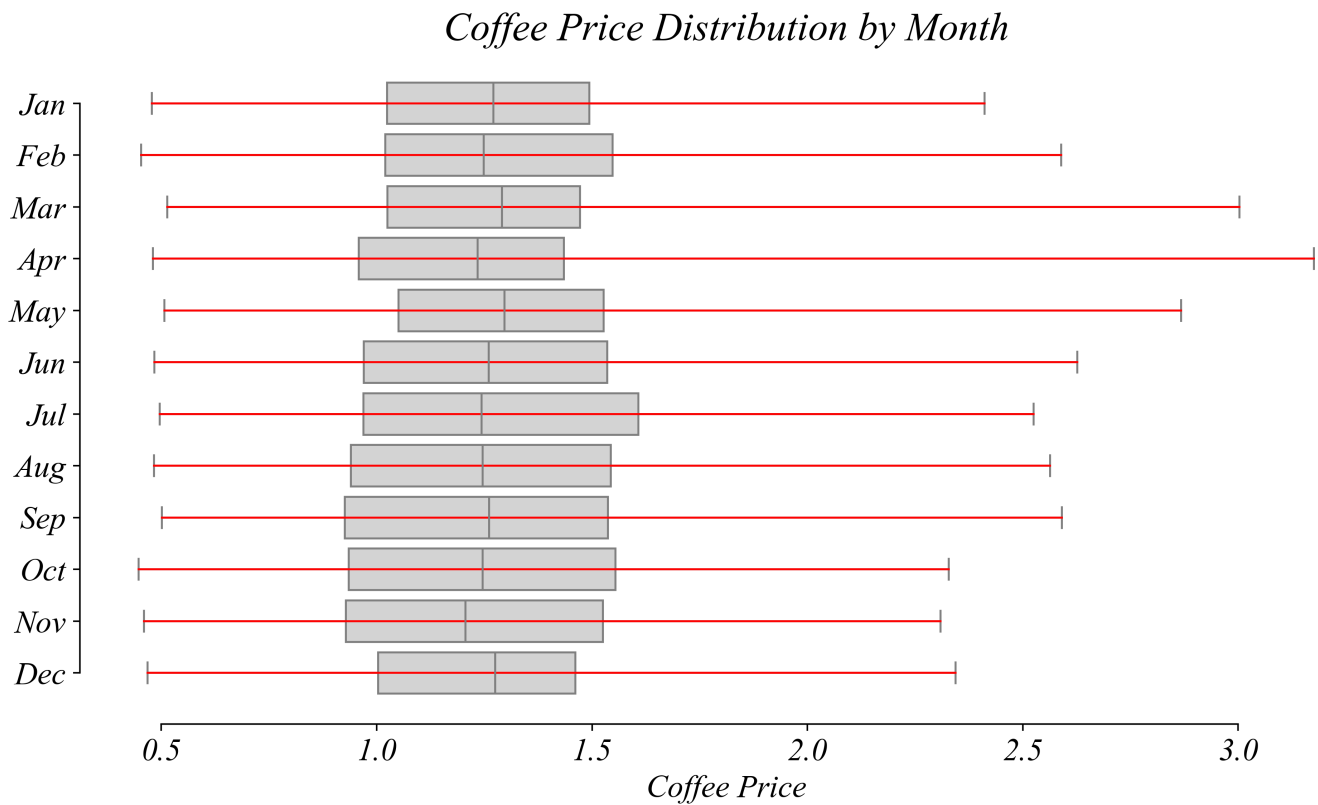
Coffee Price Distribution by Month

The largest of these twelve maximum prices occurred in April.

# Coffee Price Distribution by Month

In which season are prices most spread out? To assess the spread, we focus on the whiskers, ignoring the boxes.

## Coffee Price Distribution by Month

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. So coffee prices were most volatile in spring months.

Coffee Price Distribution by Month
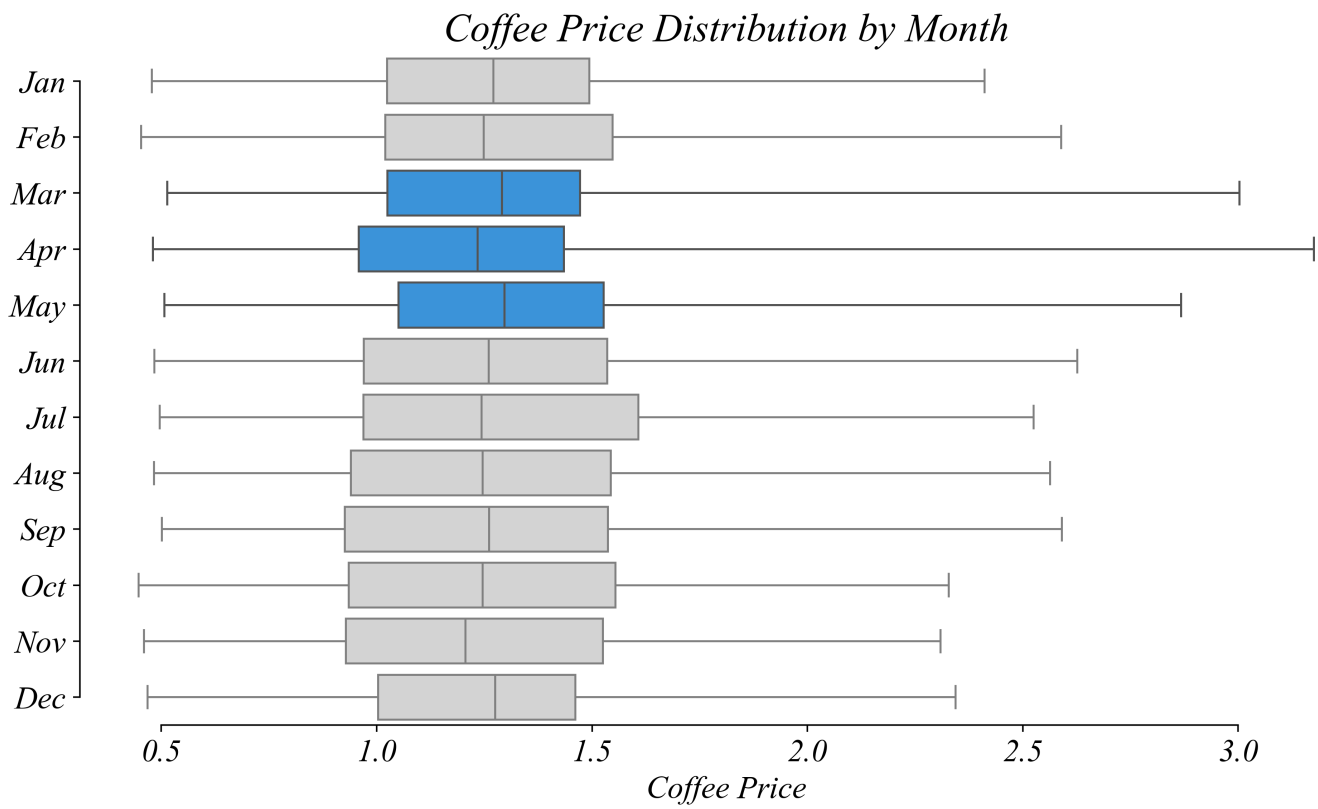
What can you say about the medians? It's hard to compare the medians, as they don't change much compared to the full range of the data. Instead of using boxplots, each with their own box and wisker, we can take some of the quantiles they represent and plot them on a line graph. This shows the minimum, median, and maximum for each month.

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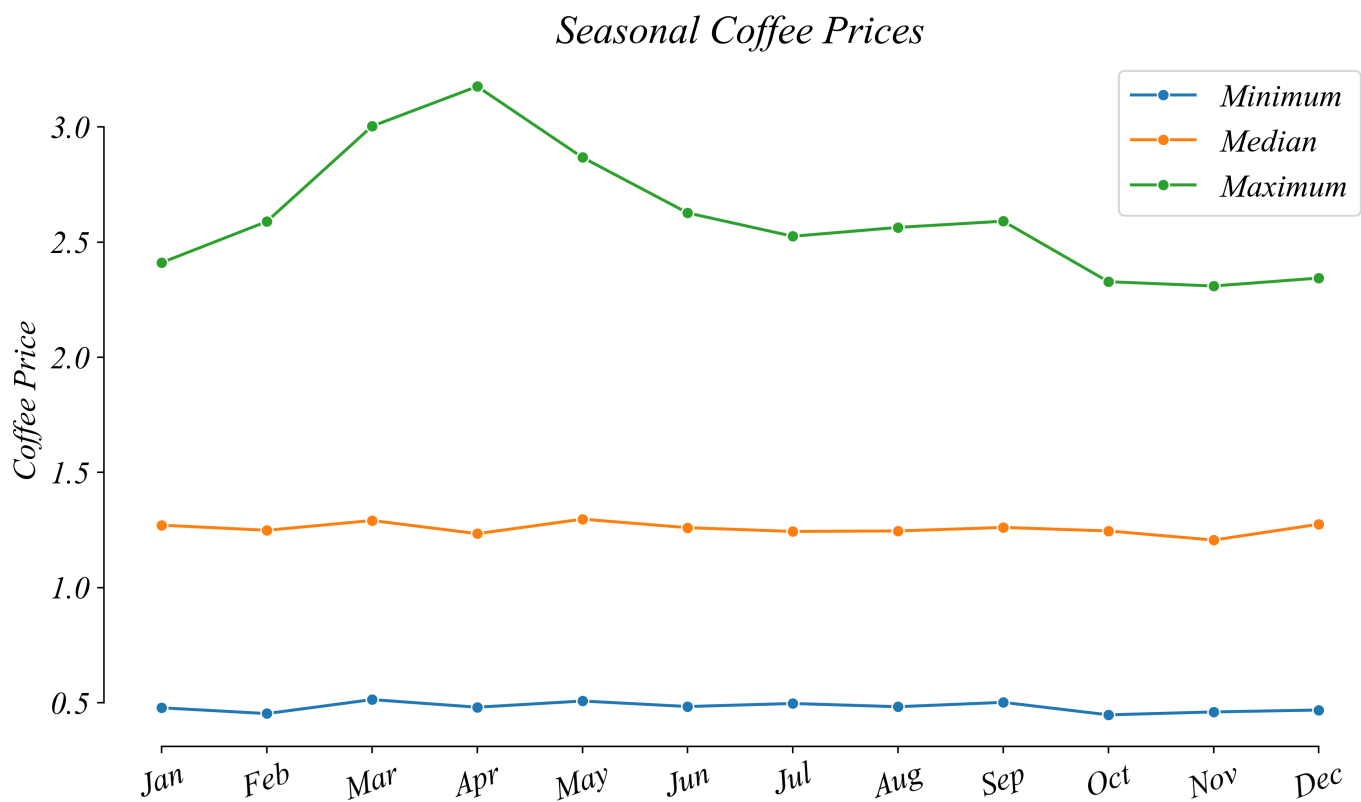
We can remove the other statistics from the plot to see the variation in the median more clearly.

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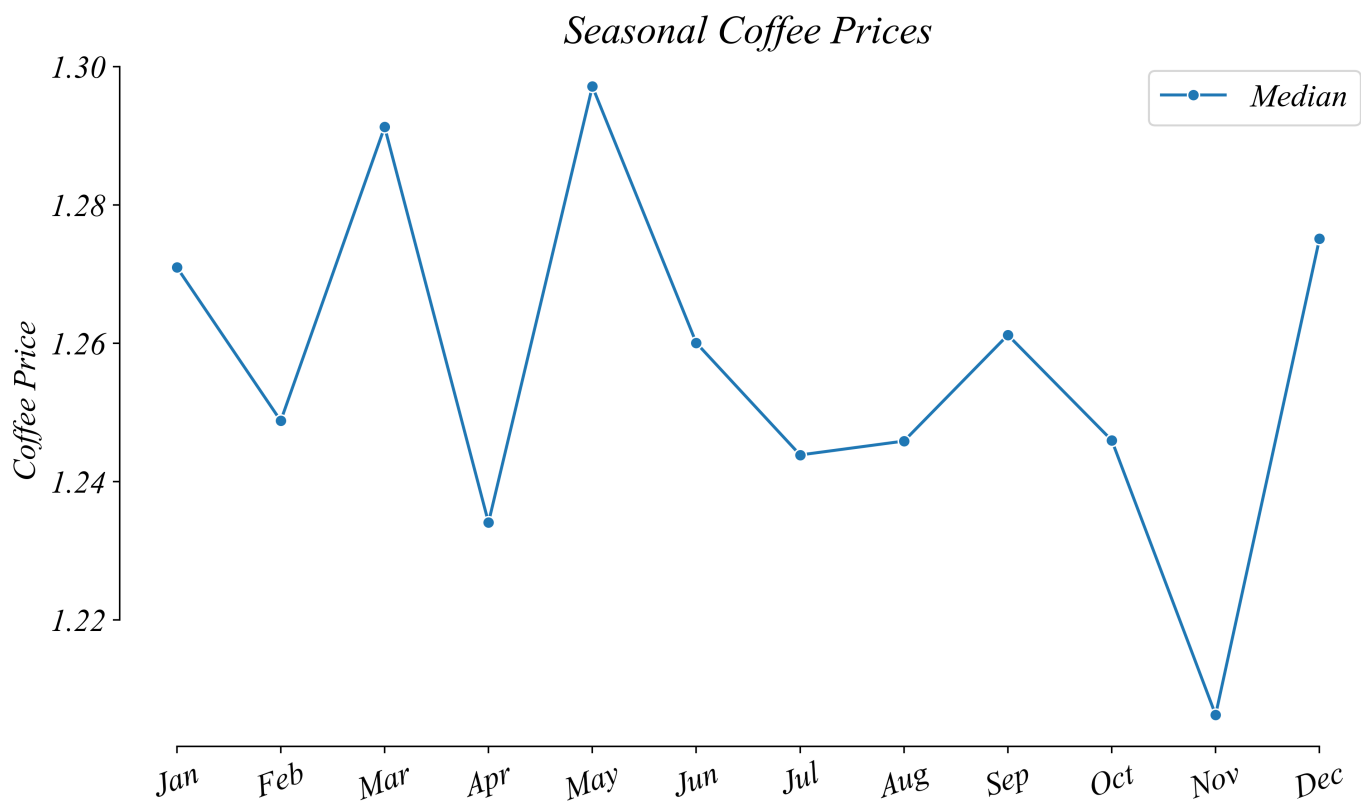
With only the median displayed, we can see its variation more easily. What is the difference between the largest and the smallest median price per pound?

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`.

The most appropriate visualization is a plot that displays only the necessary information. Boxplots helped us compare the spread of the prices in different months. To zoom in on the typical prices, we plotted just the medians.

Excel Exercise