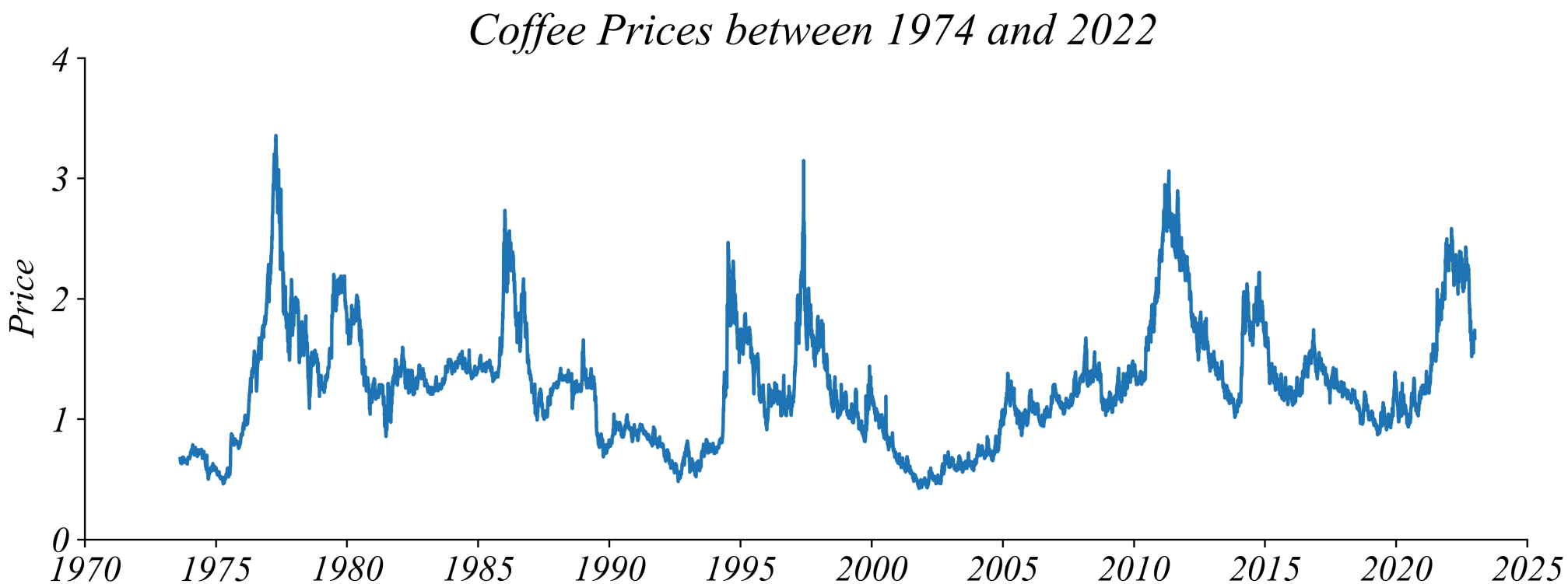
ECON 0150 | Economic Data Analysis

The economist's data analysis pipeline.

Part 2.2 | Transforming Data

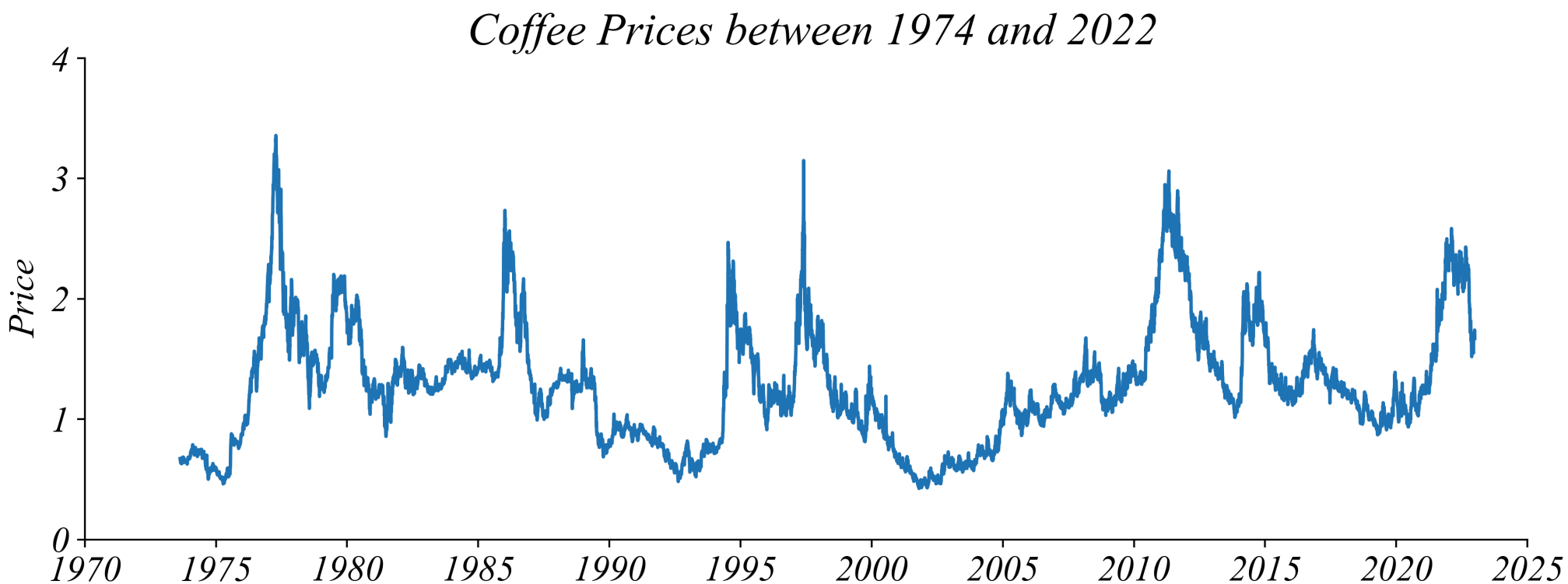
Coffee Prices

*Do you notice a **trend** in price?*



Coffee Prices

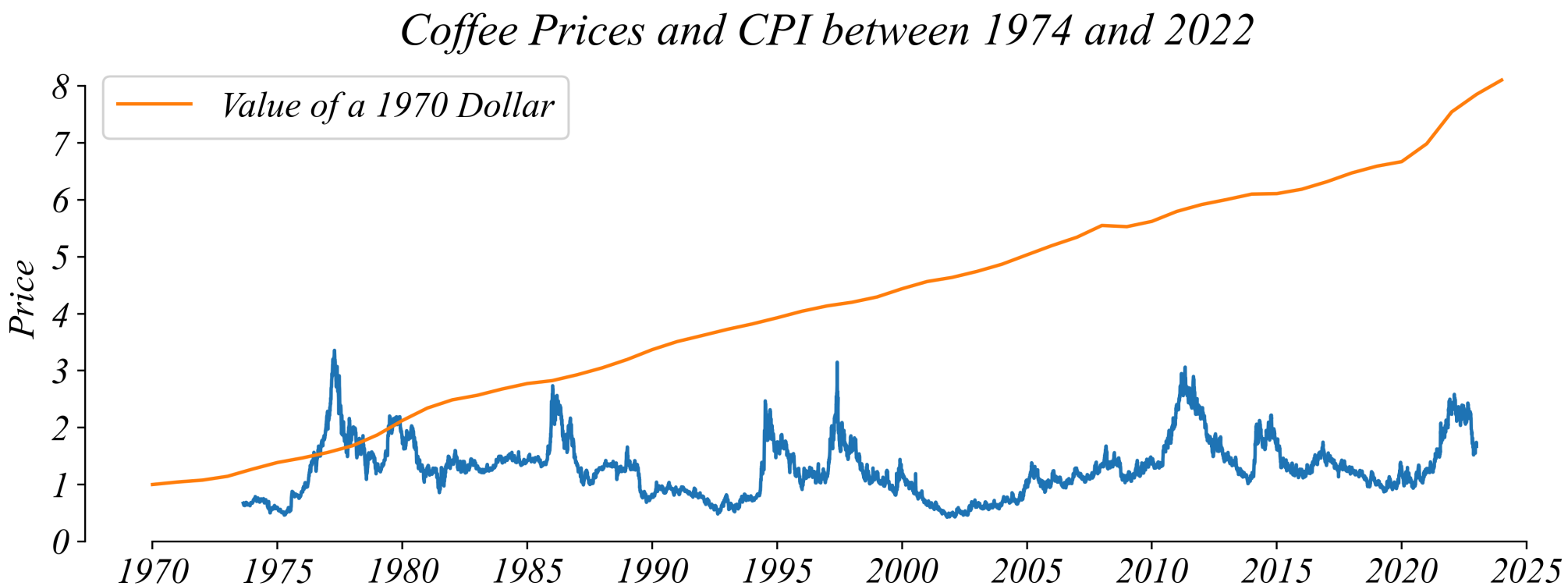
*Do you notice a **trend** in price?*



> *was coffee about as expensive in 1980 as it is today?*

Coffee Prices: Nominal vs Real Prices

*Do you notice a **trend** in price?*

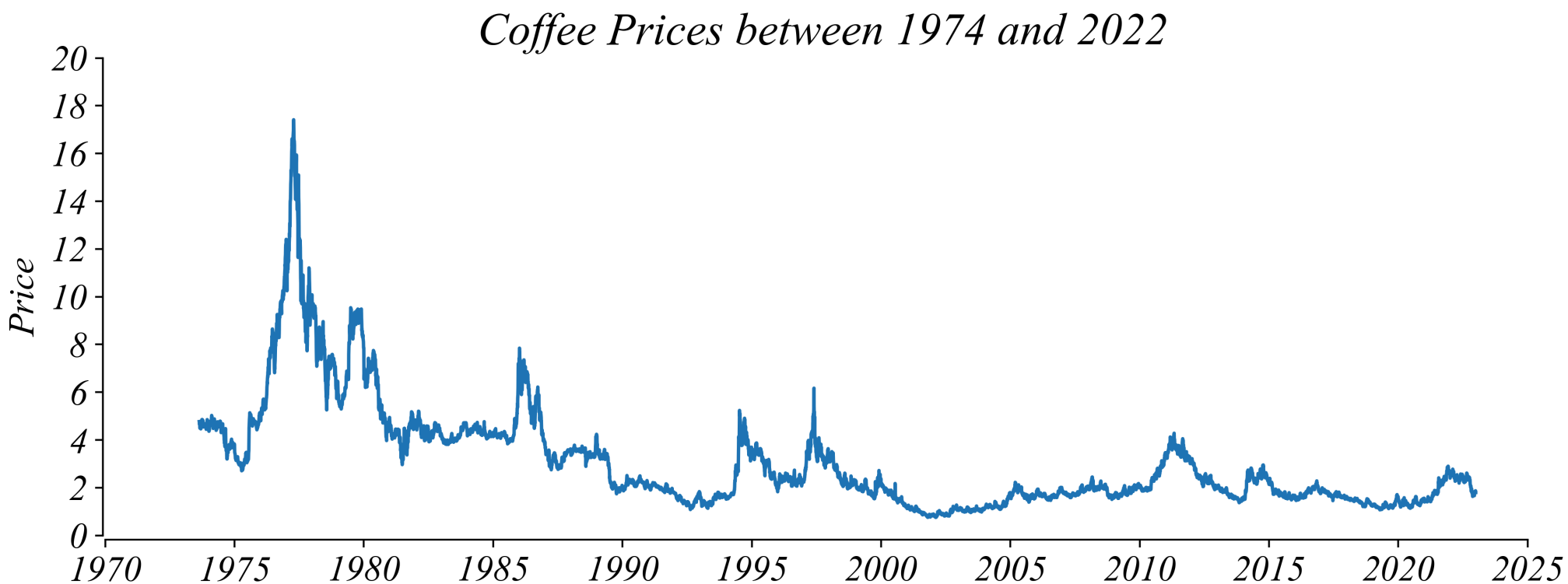


> *no! a dollar today is worth much less than in 1980!*

> *adjusting for inflation makes the picture clearer*

Real Coffee Prices: Adjust For Inflation

*Do you notice a **trend** in price?*



> prices have dropped a lot since 1970 and have been stable since 2000

Exercise 2.2 | Real Price Adjustment

Is there a trend in the real price of coffee?

Lets transform coffee prices from nominal dollars to real dollars.

- ***Data:*** `Coffee_Prices_CPI.csv`

Starbucks' Global Server Capacity

How many shops are opening at any given time?

- *Starbucks manages many shops globally and needs to maintain server capacity for all of them around the clock.*
- *Starbucks has a massive operation to make sure their shops are able to open every morning.*
- *Lets investigate how many coffee shops are opening at any hour of the day.*

Opening Times: Starbucks' Global Capacity

How many shops are opening at any given time?

Looking at the data is a good place to start.

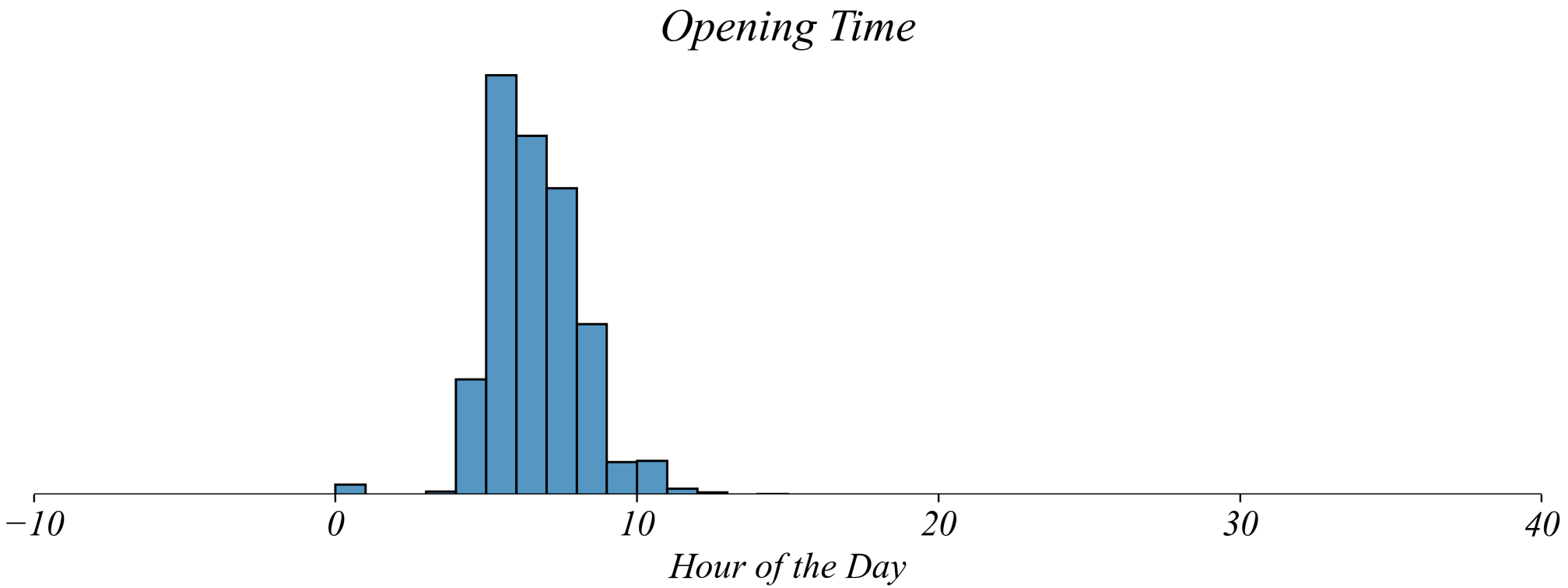
	country	open	close	GMT
0	HK	8	22	8
1	HK	7	22	8
2	HK	8	22	8
3	HK	8	22	8
4	HK	8	20	8

>as is common, it's difficult to understand the raw data on its own

Opening Times

What times do shops open in their local times?

Lets start by looking at what times shops open in local time.

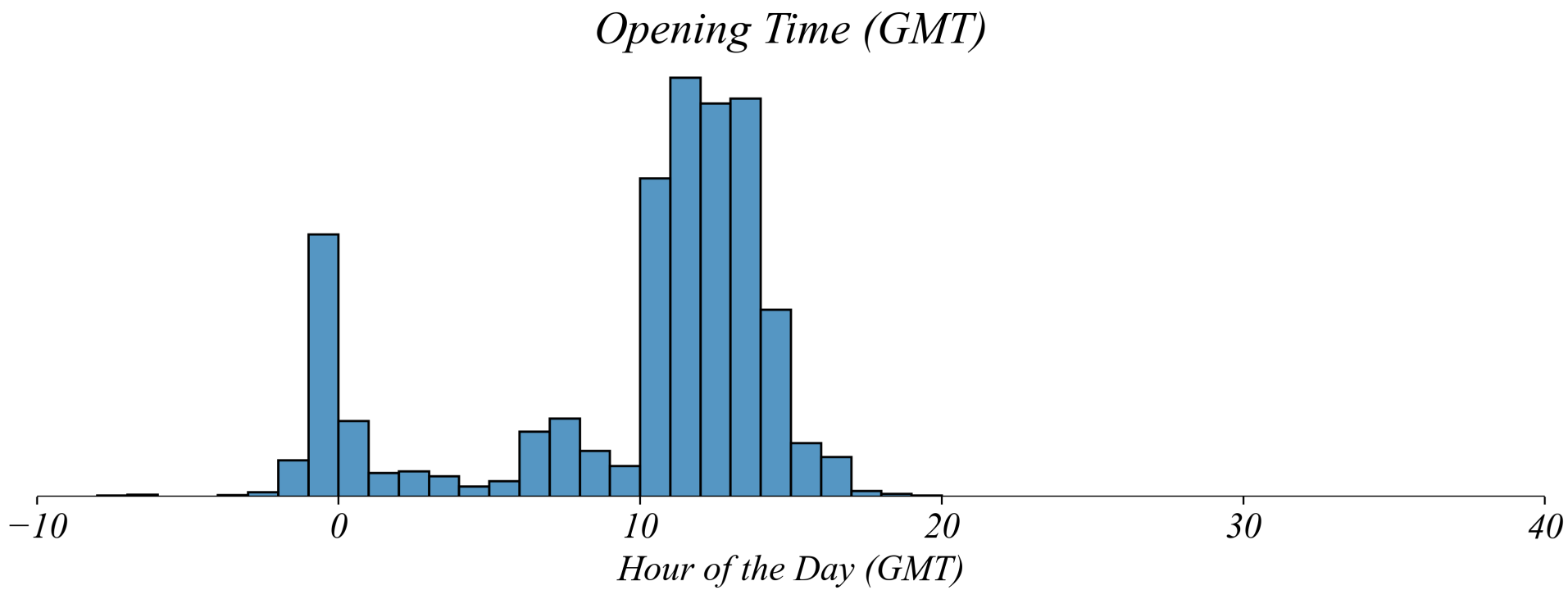


> but does this tell us how many shops are opening at one time?

Opening Times: Standardize by GMT

What times do shops open in GMT?

Lets standardize all times in *Greenwich Mean Time* (GMT).



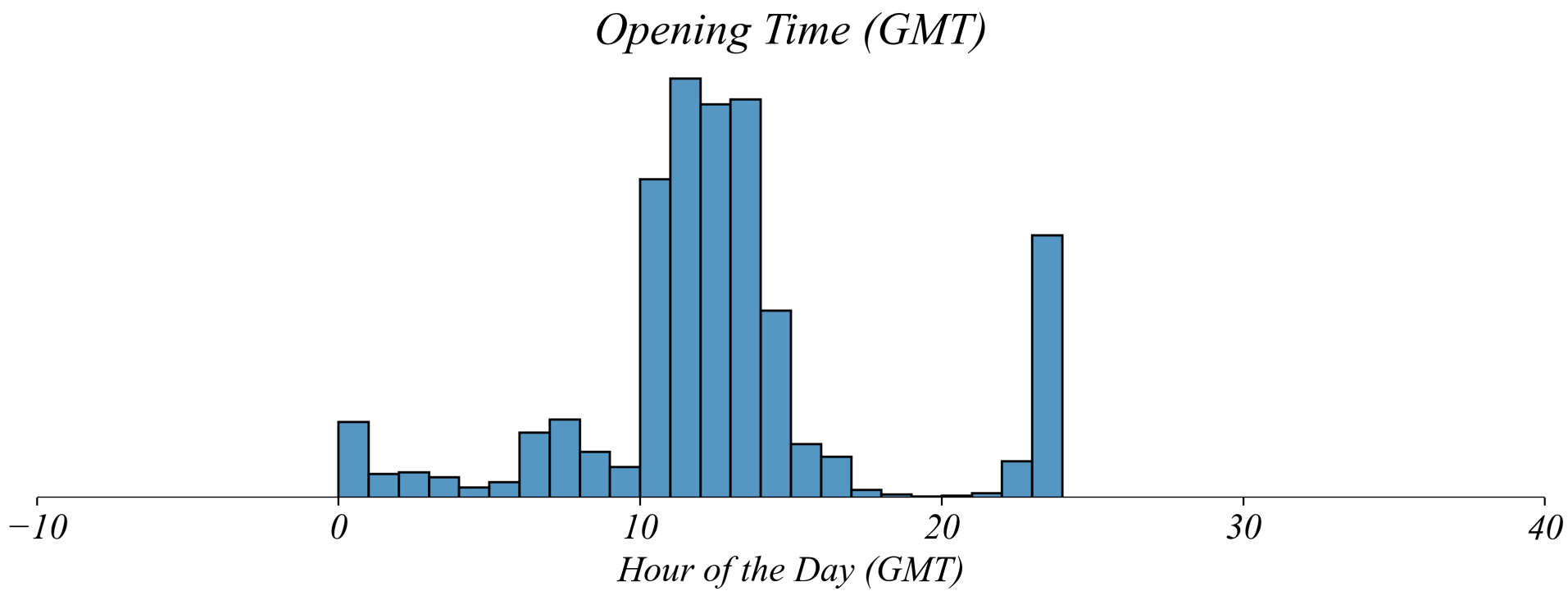
> what do the negative values mean?

> hour -1 (1 hour before GMT midnight) is the same as opening at hour 23

Opening Times: Normalize to 24 Hours

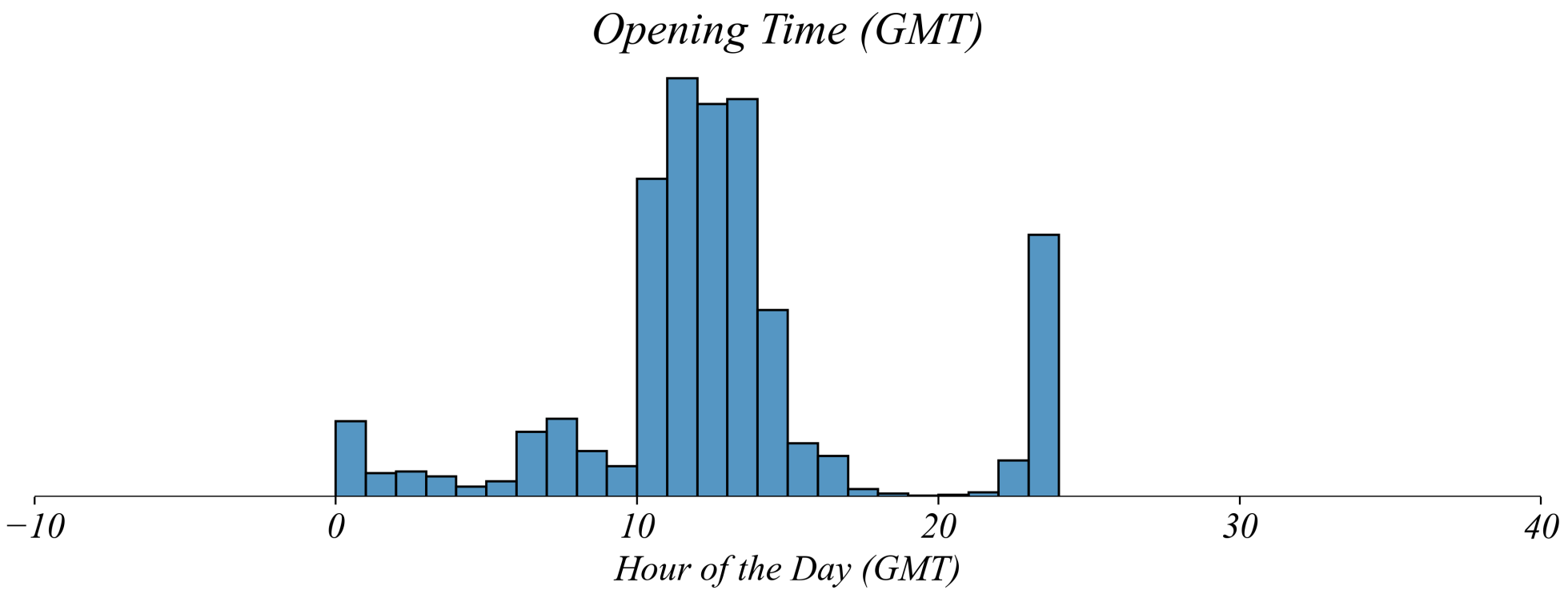
Normalize the negative values to 24 hours.

Lets add 24 if the number is negative.



Opening Times: Standardizing Hours

How many shops are opening at any given time?



> *a small bump during morning in Europe*

> *a huge spike during morning in the Americas*

> *a smaller spike during morning in Asia*

Exercise 2.2 | Starbucks' Global Capacity

How many shops are open at any given time?

- *Starbucks manages many shops globally and needs to maintain server capacity for all of them around the clock.*
- *We want to investigate how many coffee shops are open at any given hour to better understand server loads and Starbucks' global capacity needs.*
- *It's also just pretty interesting.*

Exercise 2.2 | Starbucks' Global Capacity

How many shops are opening at any given time?

Looking at the data is a good place to start.

	country	open	close	GMT
0	HK	8	22	8
1	HK	7	22	8
2	HK	8	22	8
3	HK	8	22	8
4	HK	8	20	8

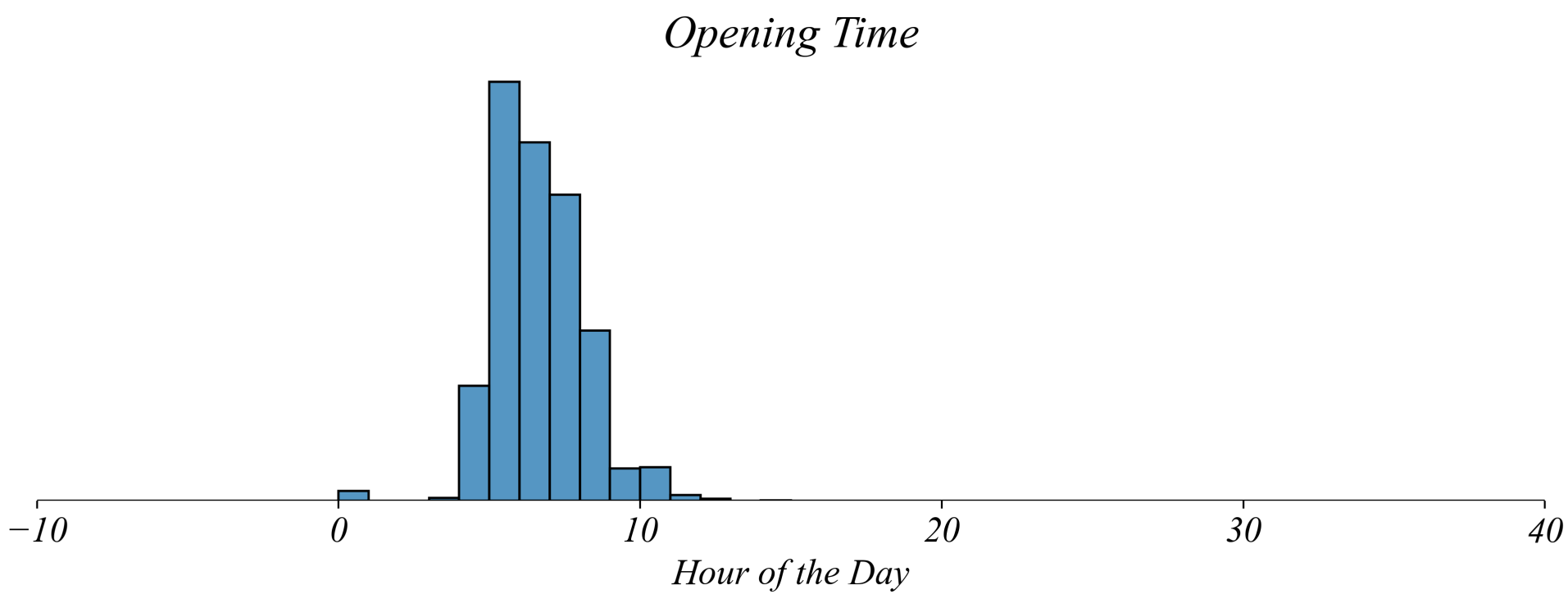
>as is common, it's difficult to understand the raw data on its own

Opening Times

What times do shops open in their local times?

Lets start by looking at what times shops open in local time.

```
# Histogram of opening times
sns.histplot(hours, x='open')
```



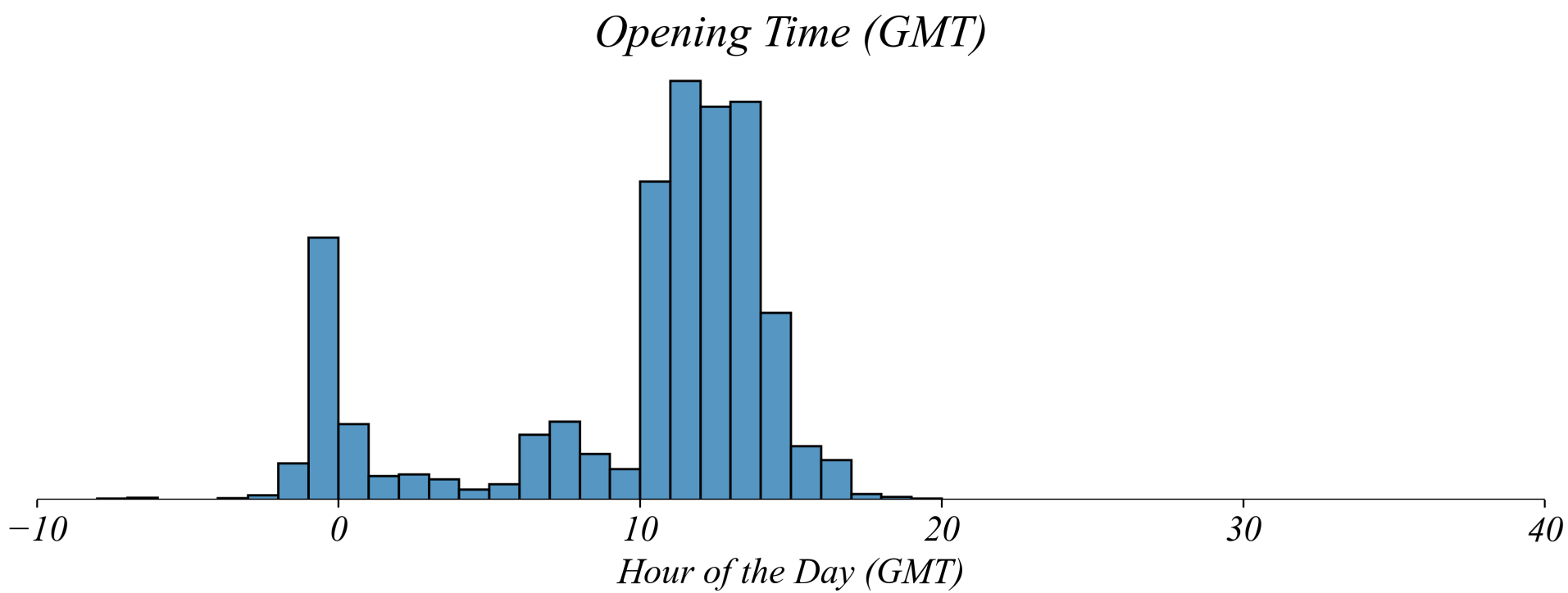
Opening Times: Standardize by GMT

What times do shops open (GMT)?

Lets standardize all times in *Greenwich Mean Time* (GMT).

```
# Normalize to GMT
hours['OpenGMT'] = hours['open'] - hours['GMT']

# Histogram of opening times (GMT)
sns.histplot(hours, x='OpenGMT')
```



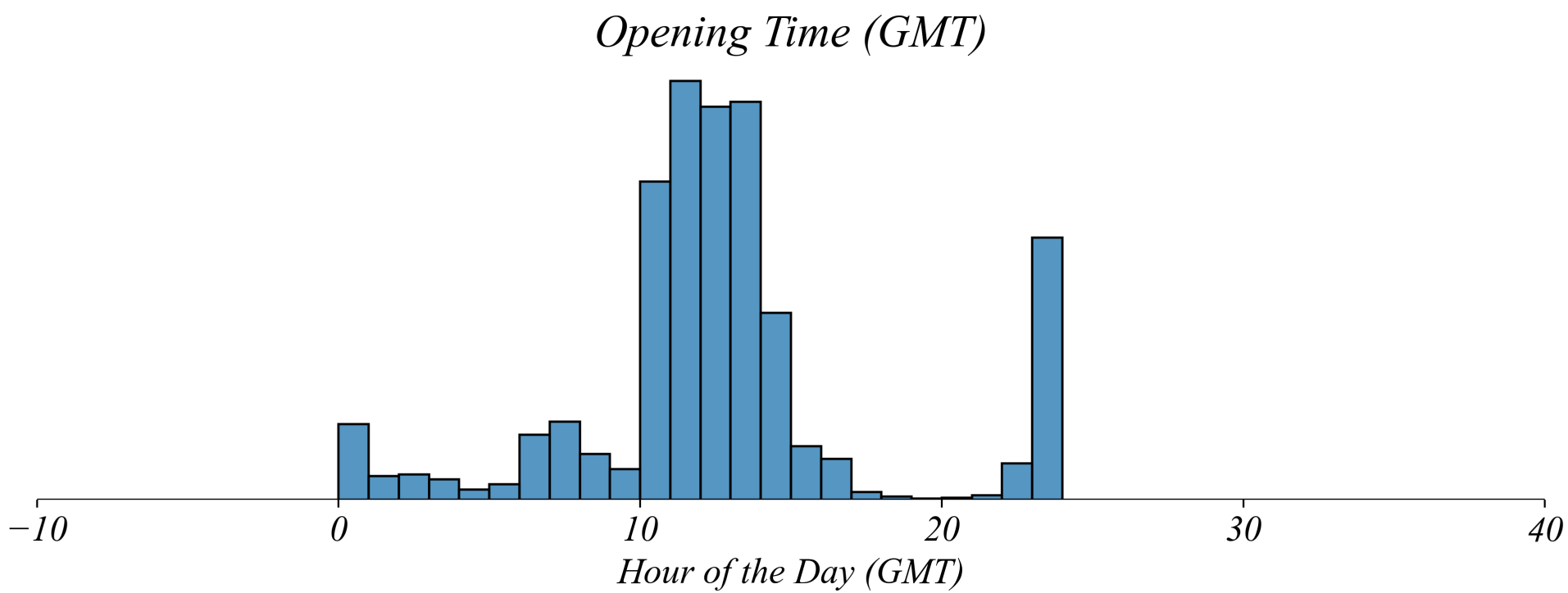
Opening Times: Standardizing Hours

Normalize the negative values to 24 hours.

Lets add 24 if the number is negative.

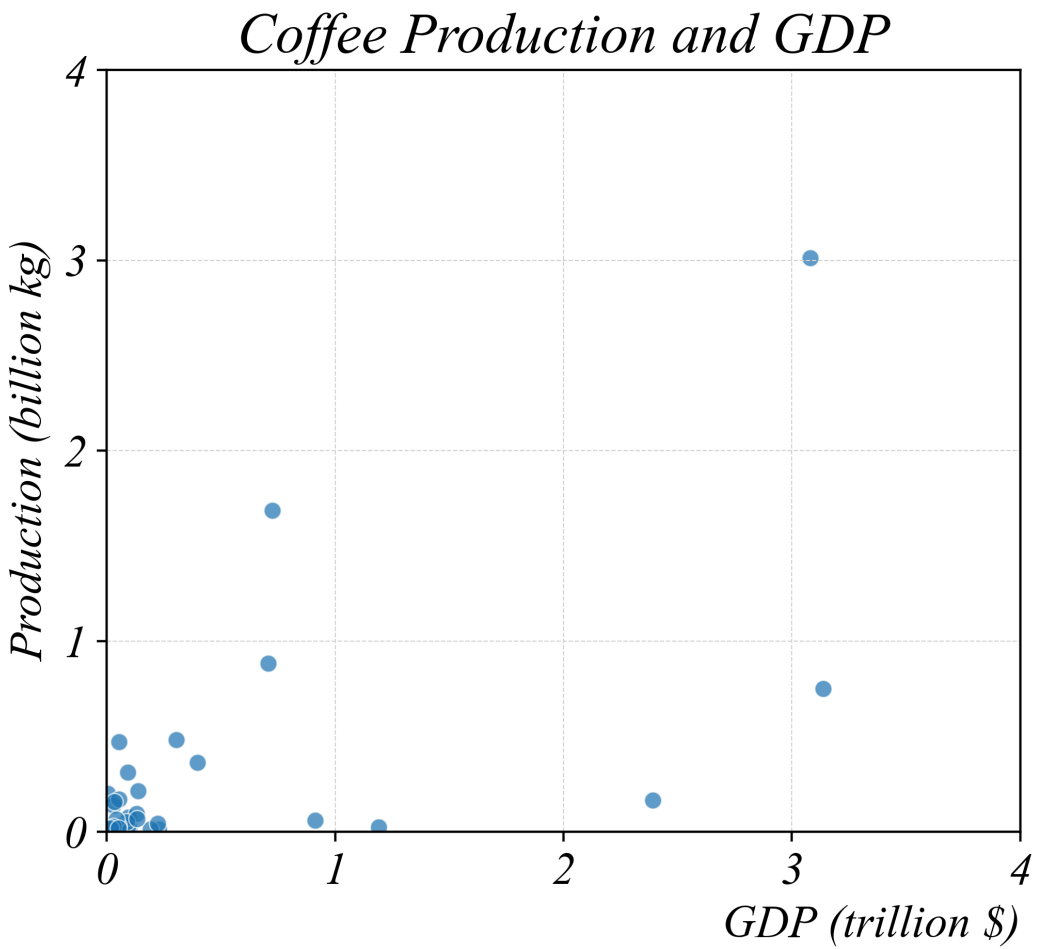
```
# Normalize to 24 hours
hours['OpenGMT24'] = hours['OpenGMT'].mod(24)

# Histogram of opening times (GMT, 24)
sns.histplot(hours, x='OpenGMT24')
```



Scatterplot: Linear Scale

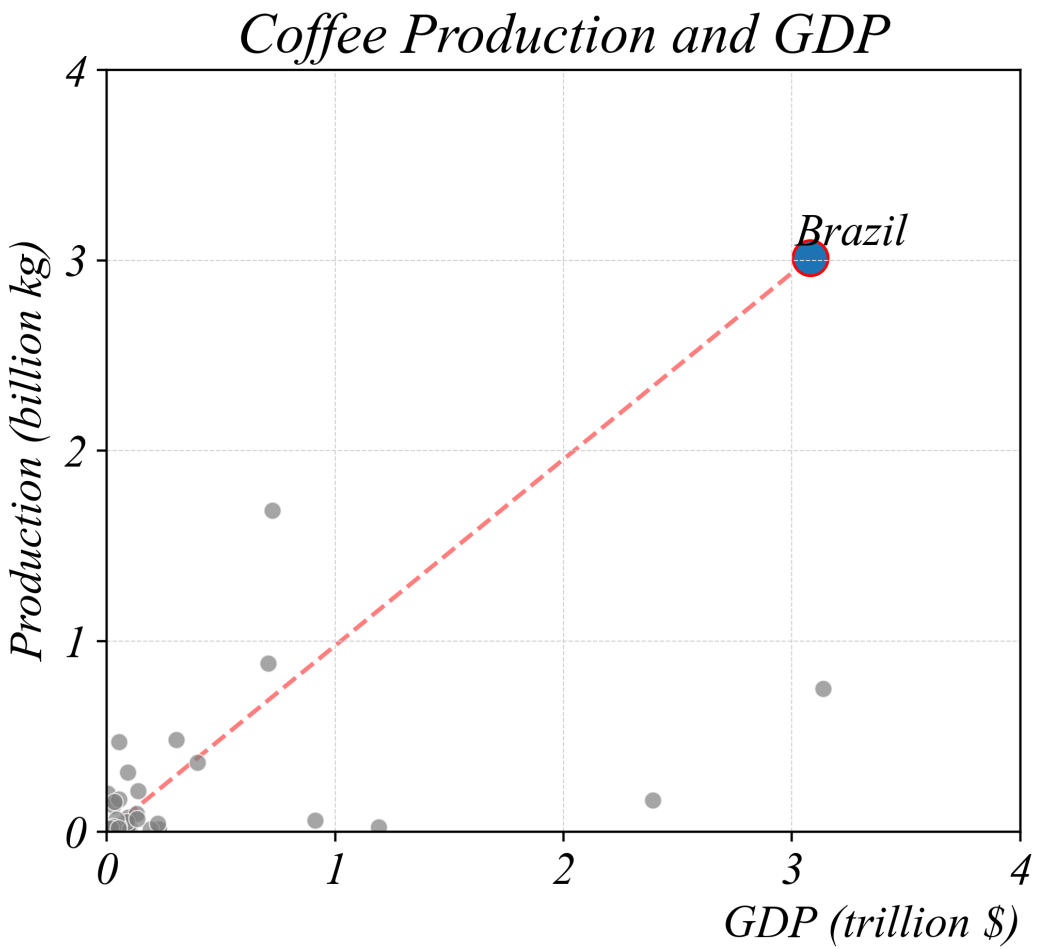
Q. Is there a relationship between GDP and coffee production?



> a scatterplot effectively visualizes scross sectional data with two dimensions

Scatterplot: Linear Scale

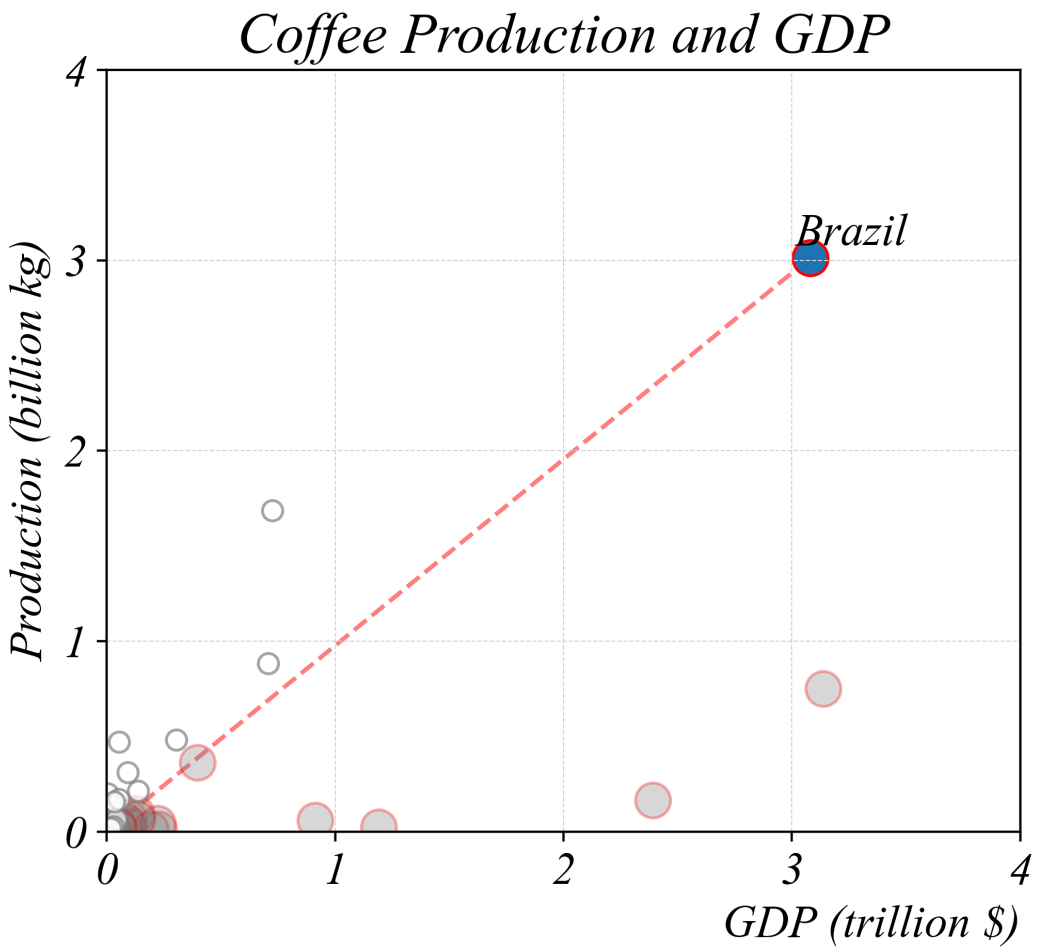
Which countries produce less coffee per dollar than Brazil?



> *separating lines can help make comparisons between ratios*

Scatterplot: Linear Scale

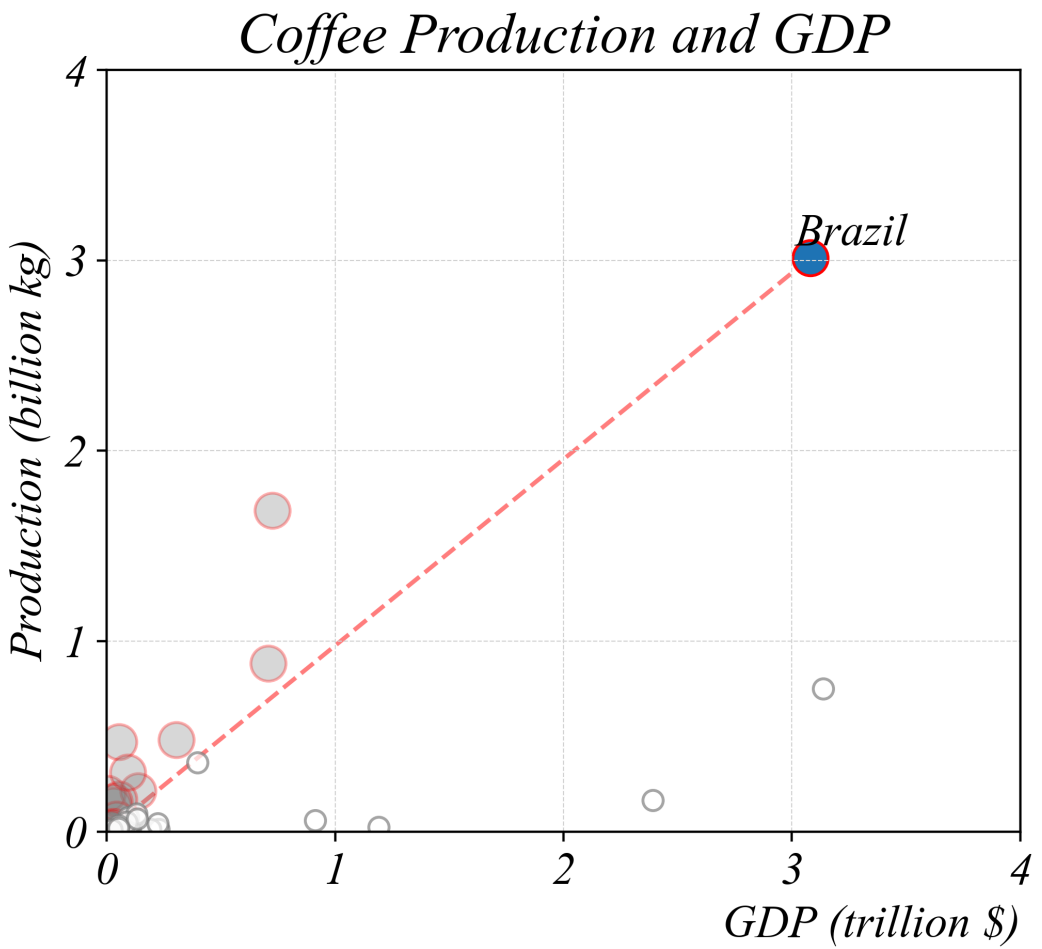
Which countries produce less coffee per dollar than Brazil?



> *separating lines can help make comparisons between ratios*

Scatterplot: Linear Scale

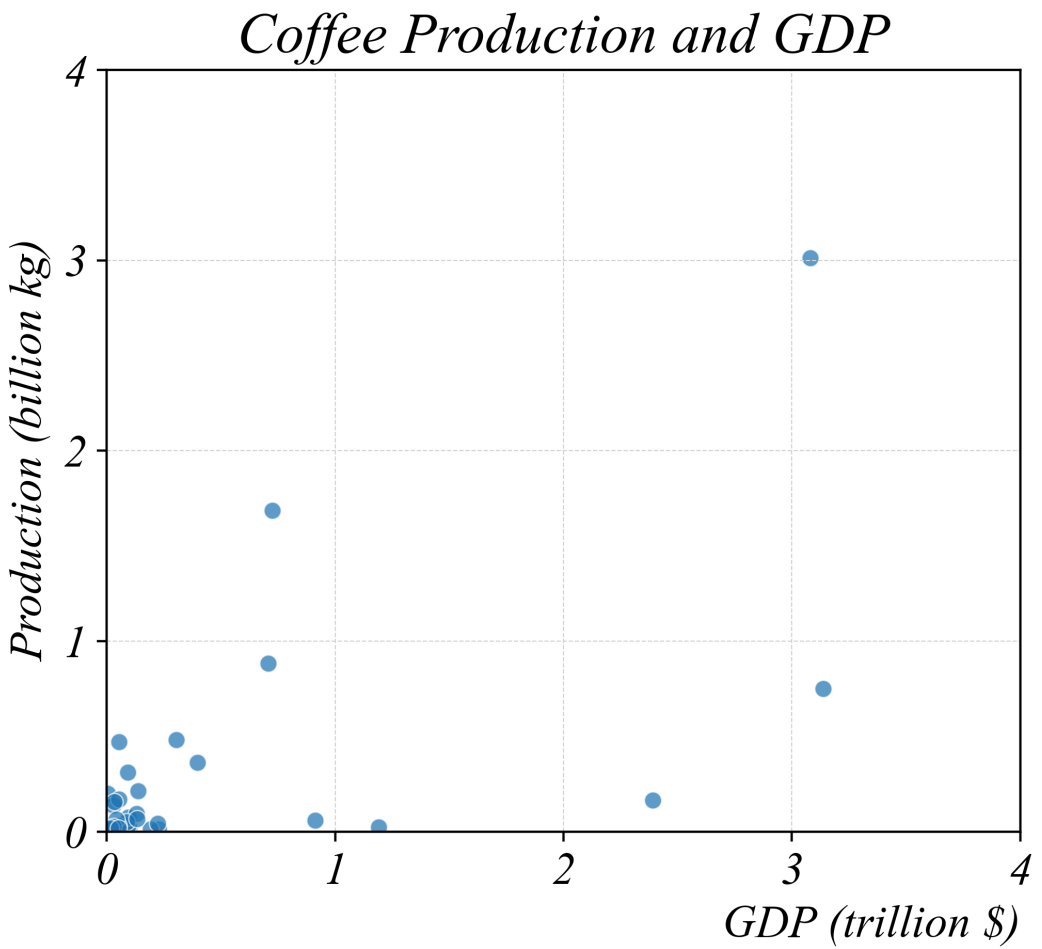
Which countries produce more coffee per dollar than Brazil?



> *separating lines can help make comparisons between ratios*

Scatterplot: Linear Scale

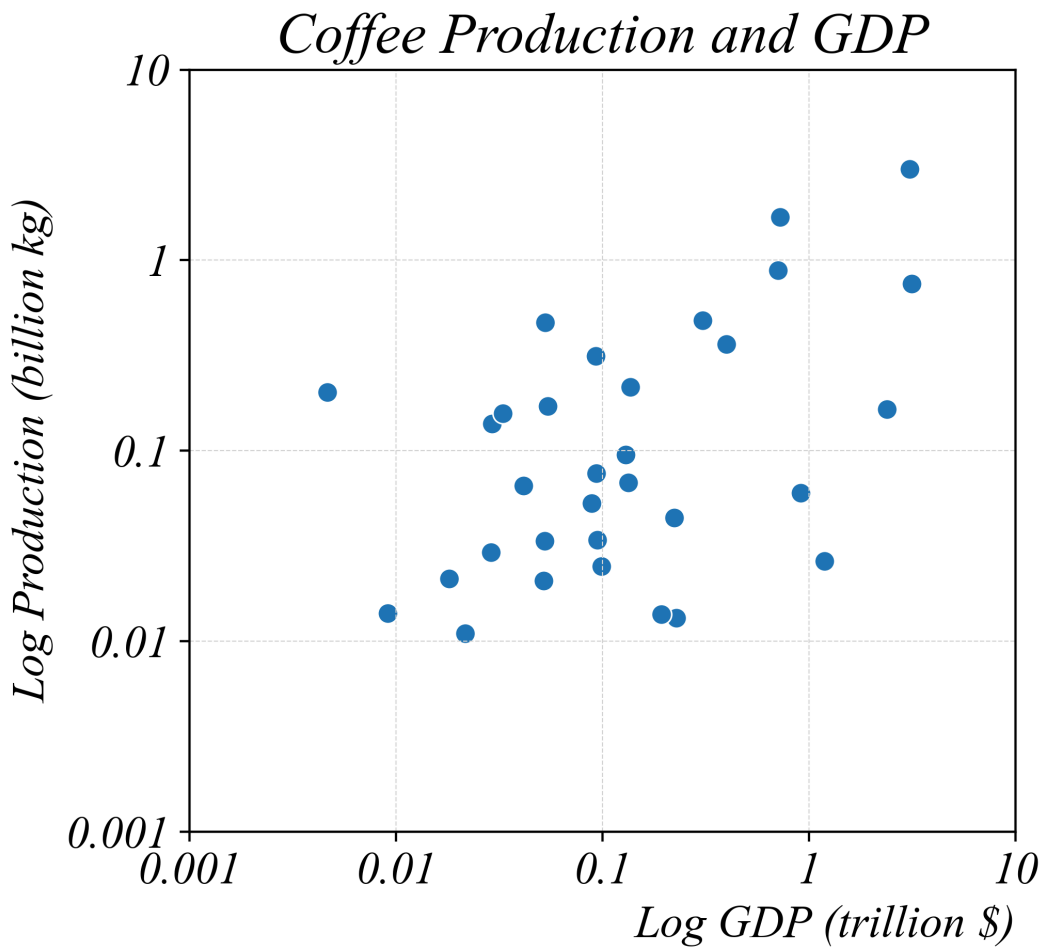
How does GDP relate to coffee production?



> *small values are bunched; large data is very separated*

Scatterplot: Log Scale

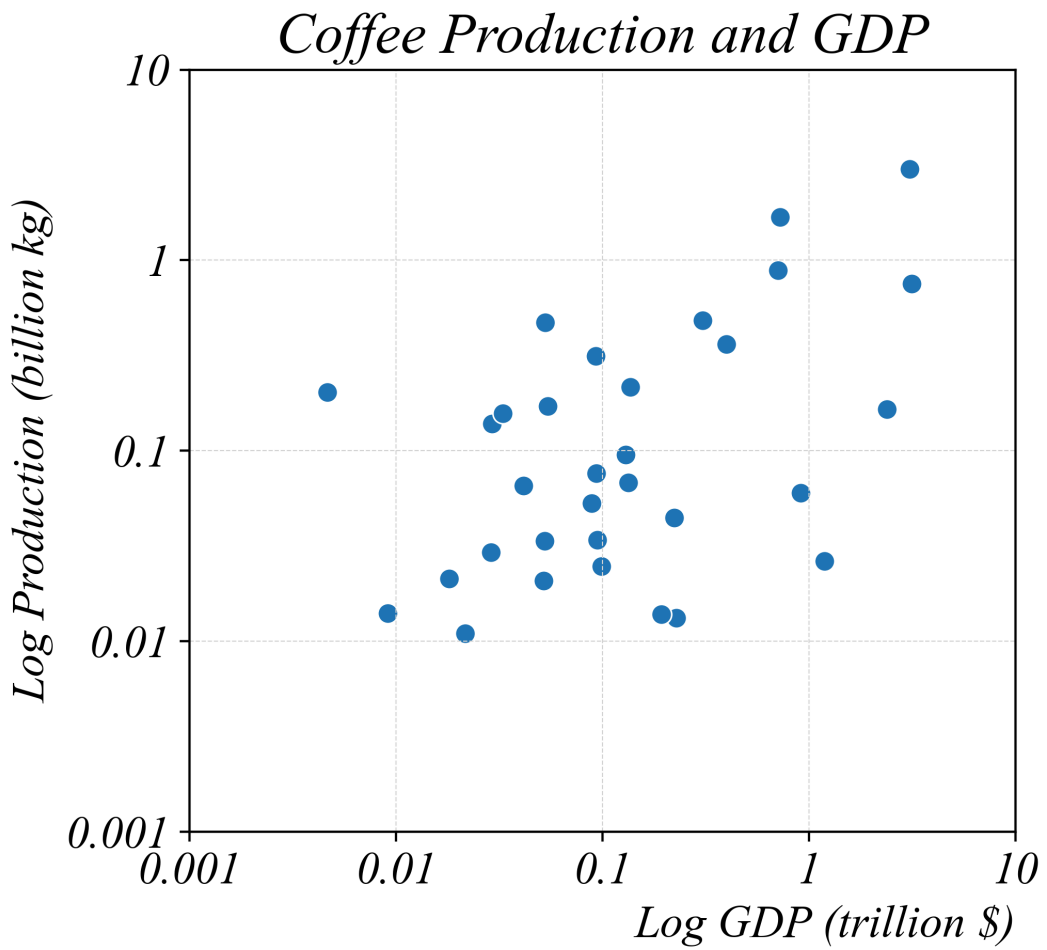
How does GDP relate to coffee production?



> *we can fix this by applying a log transformation*

Scatterplot: Log Scale

How does GDP relate to coffee production?



> *looks positive, but we'll formally test this in Part 4*

Exercise 2.2 | Log Transformation

How does GDP relate to coffee production?

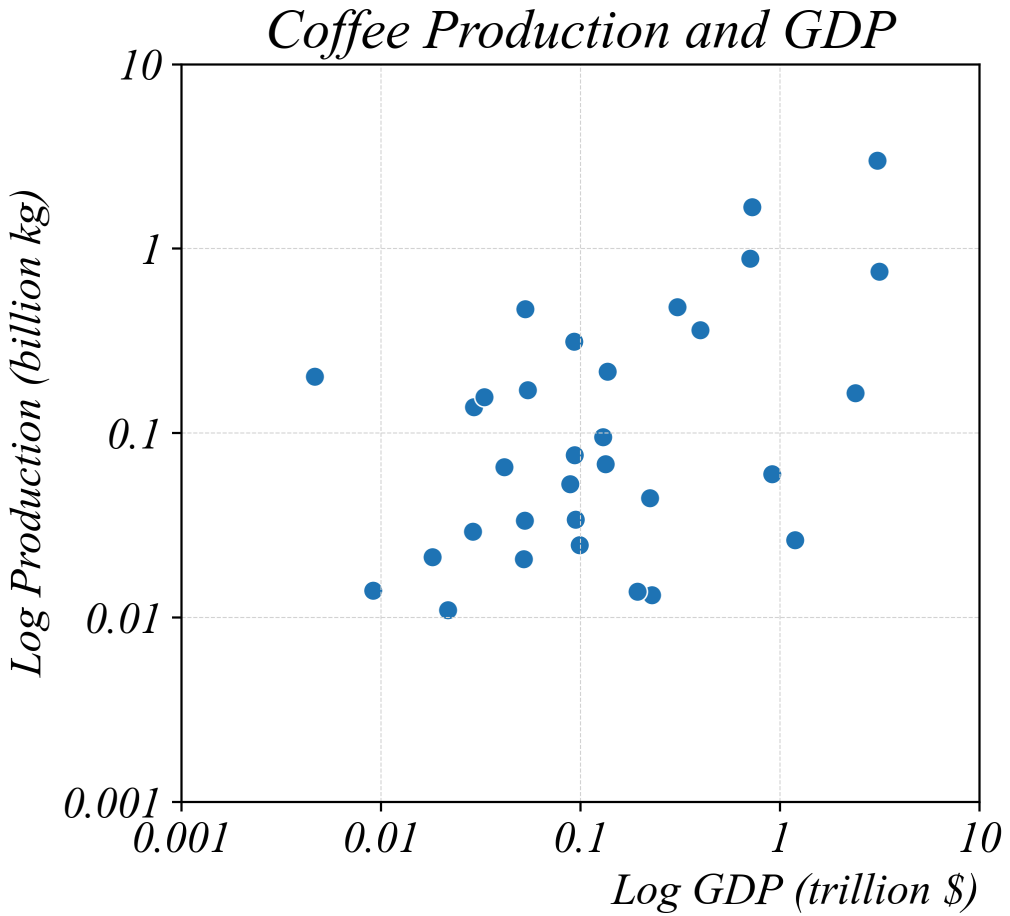
	Code	Year	coffee_prod	Entity	GDP
49	AGO	2019	0.013257	Angola	0.227856
189	BOL	2019	0.024841	Bolivia	0.098836
248	BRA	2019	3.011745	Brazil	3.080049
307	BDI	2019	0.014059	Burundi	0.009110
416	CMR	2019	0.034061	Cameroon	0.094488

Exercise 2.2 | Log Transformation

How does GDP relate to coffee production?

```
# Log both x and y variables
gdp['log_GDP'] = np.log(gdp['GDP'])
gdp['log_prod'] = np.log(gdp['coffee_prod'])
```

```
# Plot the log variables
sns.scatterplot(gdp, y='log_prod', x='log_GDP')
```


Coffee Production and GDP
Log Production (billion kg)
10
1
0.1
0.01
0.001
0.001
0.01
0.1
1
10
Log GDP (trillion $)

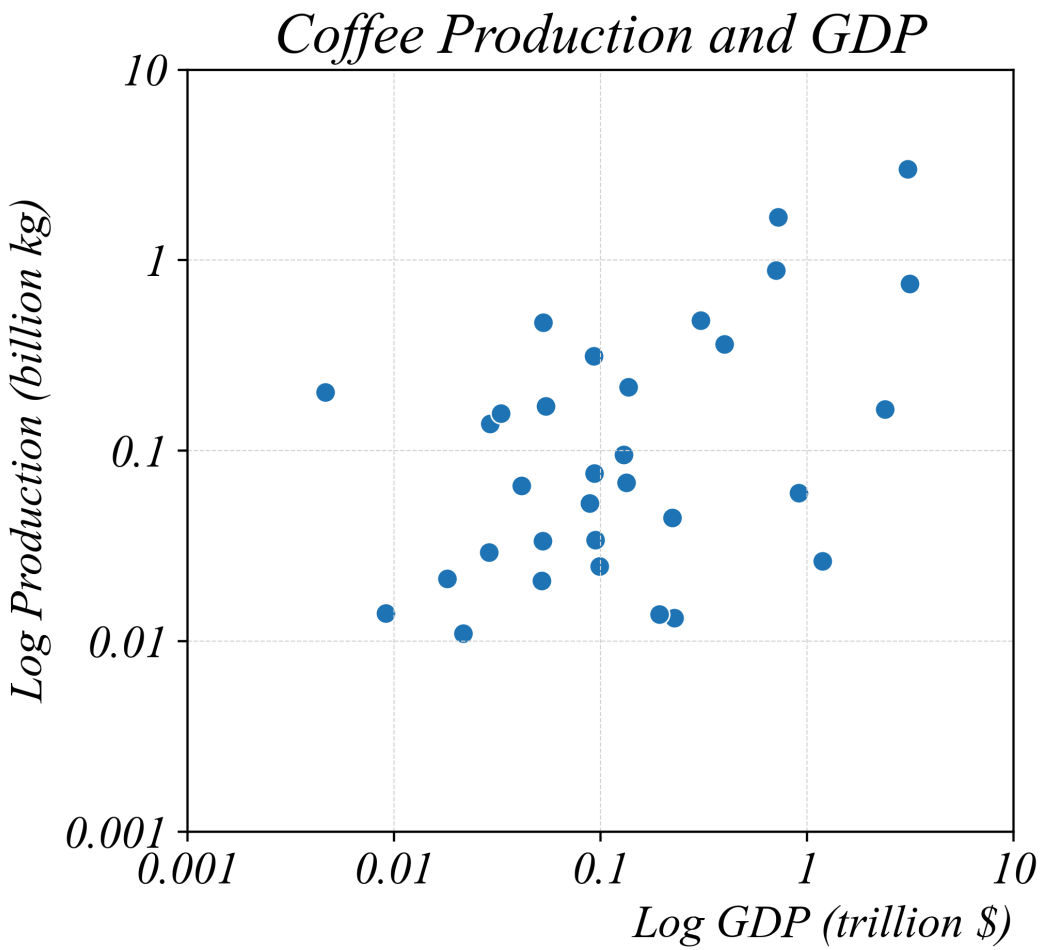
Exercise 2.2 | Log Transformation

How does GDP relate to coffee production?

```
# Use a log scale without transforming the variable
sns.scatterplot(gdp, y='coffee_prod', x='GDP')
plt.xscale('log')
plt.yscale('log')
```

Scatterplot: Log Scale

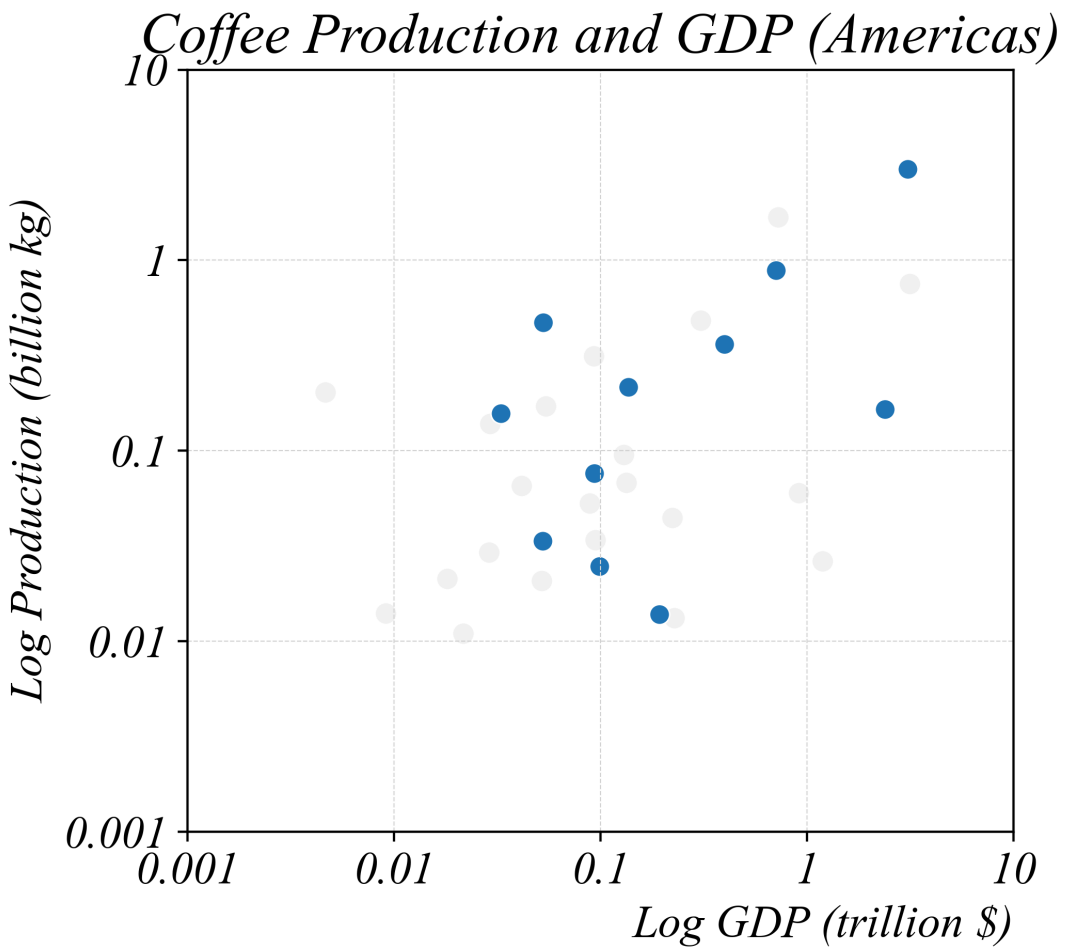
How does GDP relate to coffee production in the Americas?



> *lets use a filter with this data*

Filtering Data: next time!

How does GDP relate to coffee production in the Americas?



>