

ECON 0150 | Economic Data Analysis

The economist's data analysis pipeline.

Part 1.6 | Transforming Data

Exercise 1.6 | Starbucks' Global Server 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 1.6 | Starbucks' Global Server Capacity

How many shops are open at any given time?

Looking at the data is a good place to start.

	country	open	close	timezone
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

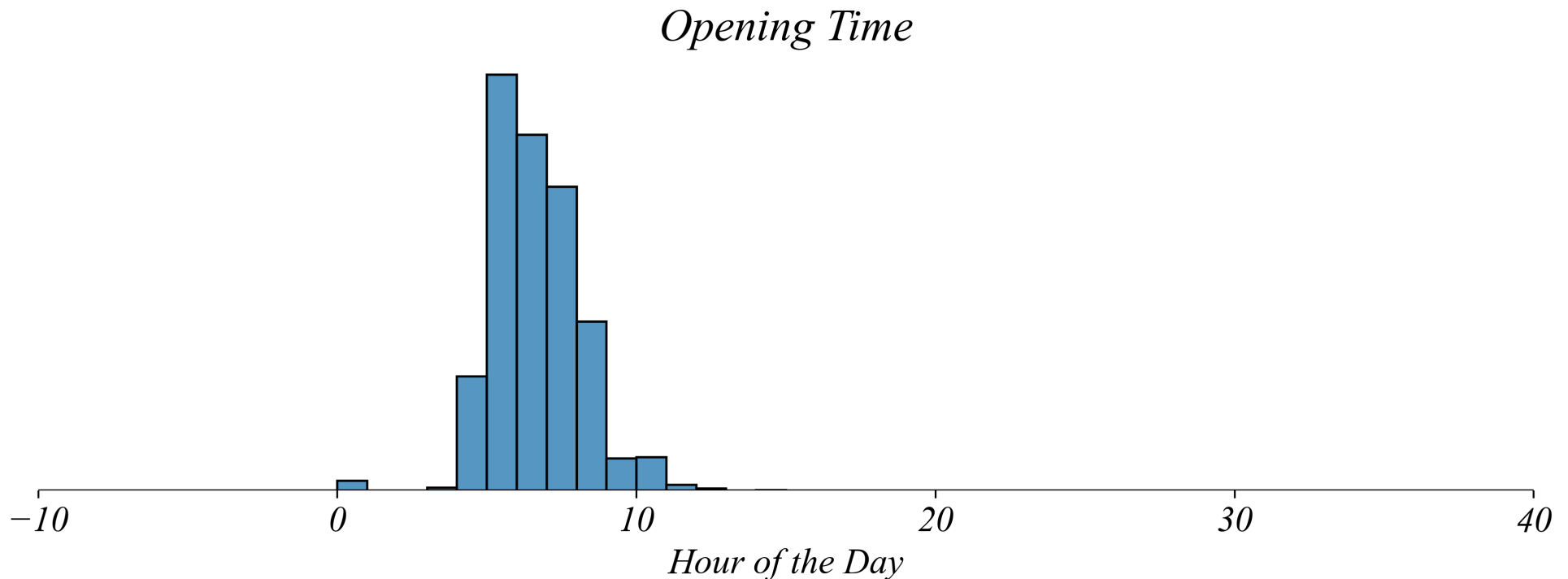
>as is common, it's difficult to understand the raw data on its own

Server Capacity: Local Opening Times

What times do shops open in their local times?

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

```
1 # Histogram of opening times  
2 plt.hist(hours.open)
```



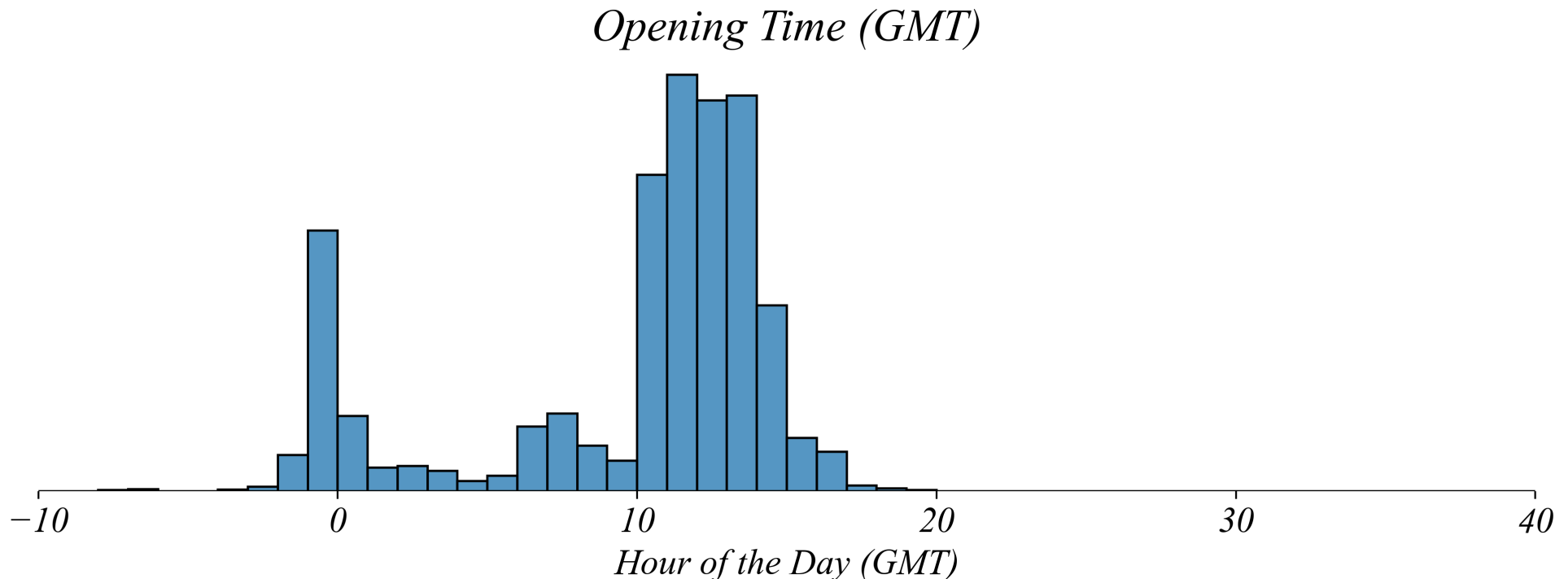
> *but does this tell us how many shops are open at one time?*

Server Capacity: Global Opening Times

What times do shops open (GMT)?

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

```
1 # Normalize to GMT
2 hours['open_GMT_simple'] = hours['open'] - hours['timezone']
3
4 # Histogram of opening times (GMT)
5 plt.hist(hours.open_GMT_simple)
```



> *what do the negative values mean?*

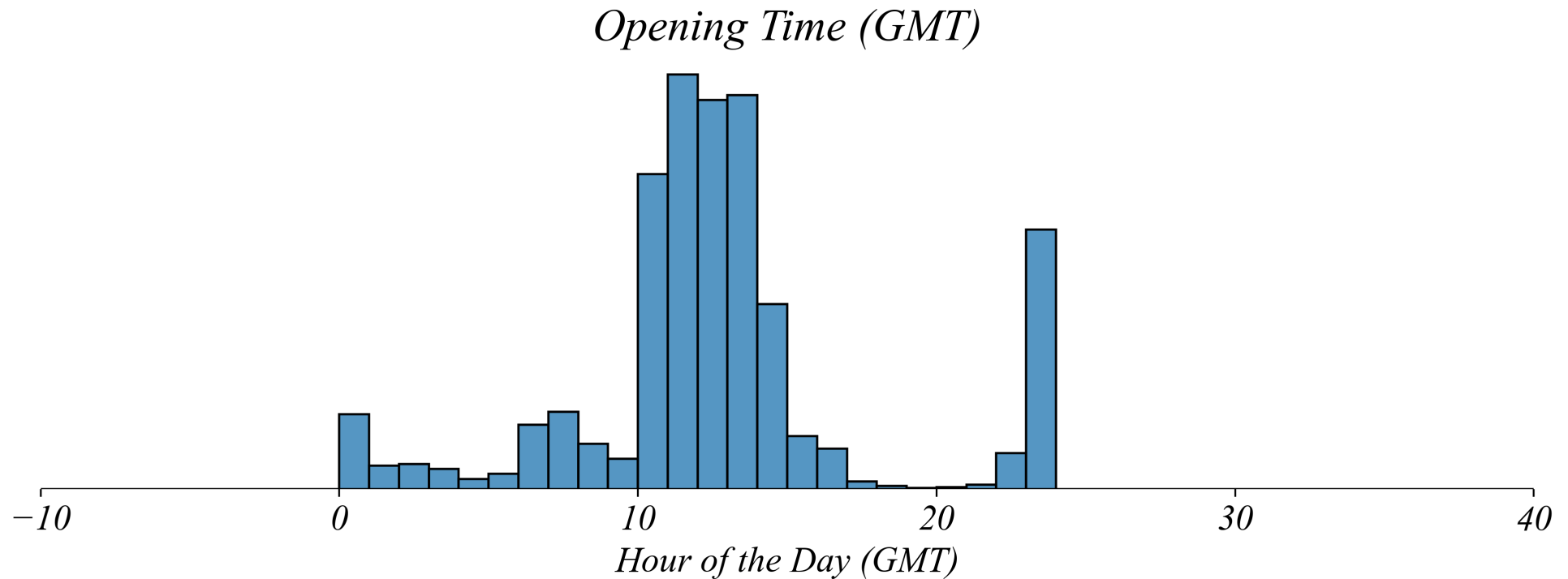
> *hour -1 (1 hour before GMT midnight) is the same as opening at hour 23*

Server Capacity: Standardizing Hours

Normalize the negative values to 24 hours.

Lets add 24 if the number is negative.

```
1 # Normalize to 24 hours  
2 hours['open_GMT'] = hours['open_GMT_simple'].mod(24)
```

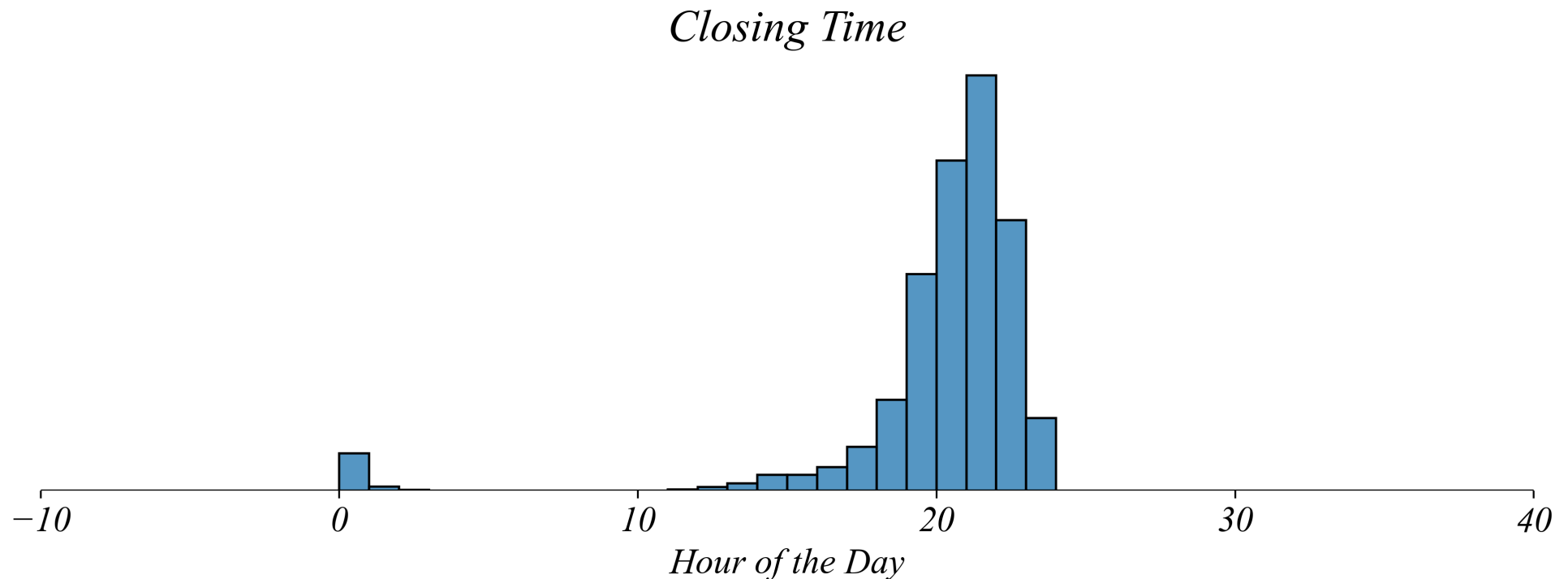


Server Capacity: Local Closing Times

What times do shops close in their local times?

Next lets examine what times shops close in their local time.

```
1 # Histogram of opening times  
2 plt.hist(hours.close)
```

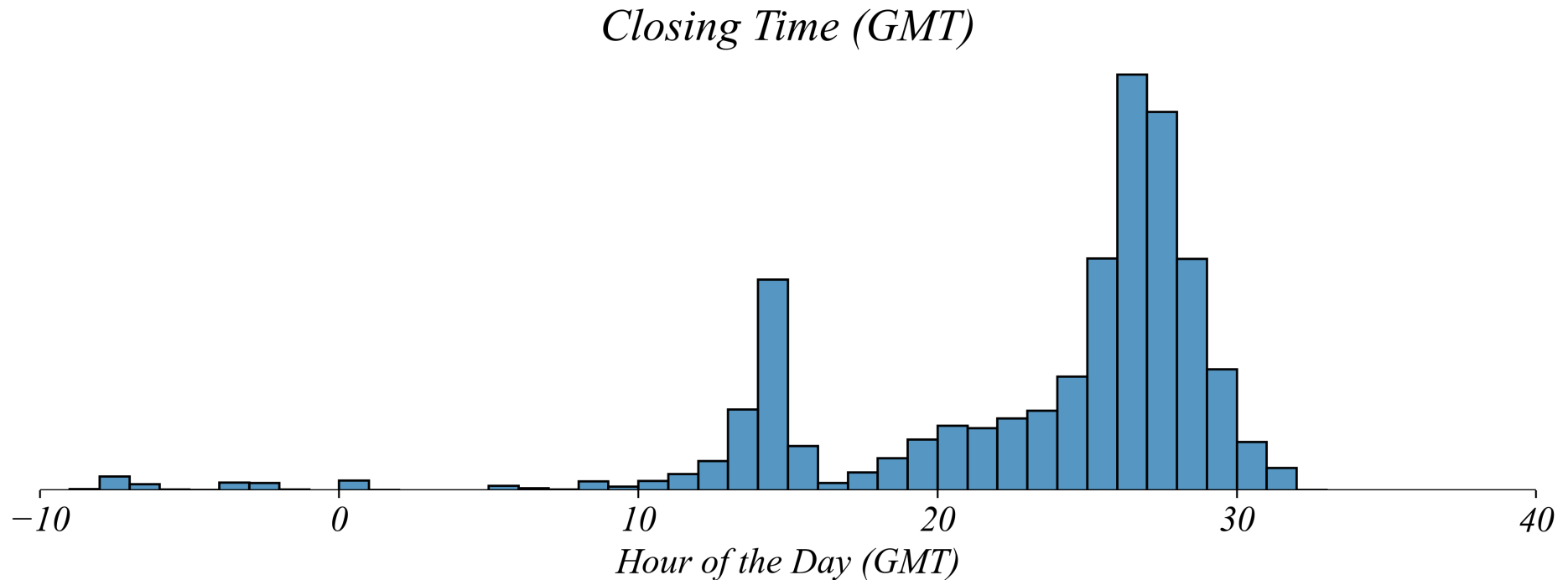


> *but lets standardize this too*

Location Hours: Global Closing Times

What times do shops close (GMT)?

```
1 # Normalize to GMT
2 hours['close_GMT_simple'] = hours['close'] - hours['timezone']
3
4 # Histogram of opening times (GMT)
5 plt.hist(hours.close_GMT_simple)
```



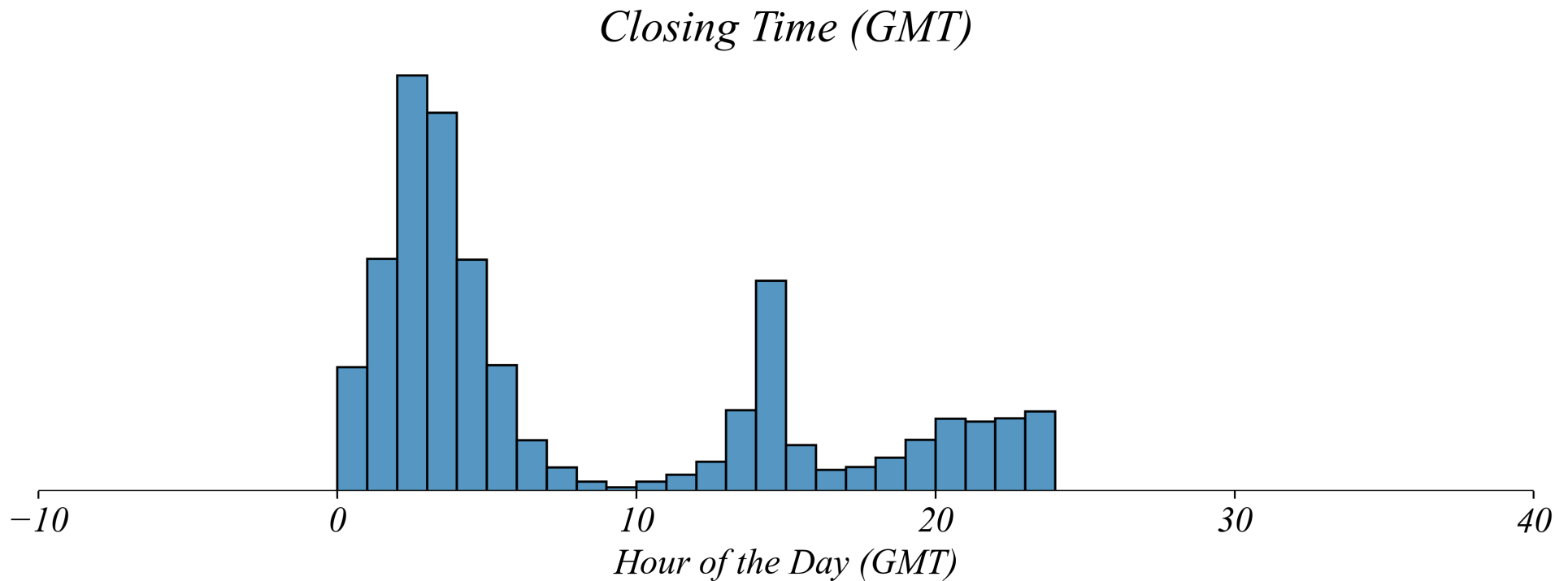
> *here we have another issue, that some values are greater than 24*

Location Hours: Global Closing Times

Normalize values to 24 hours.

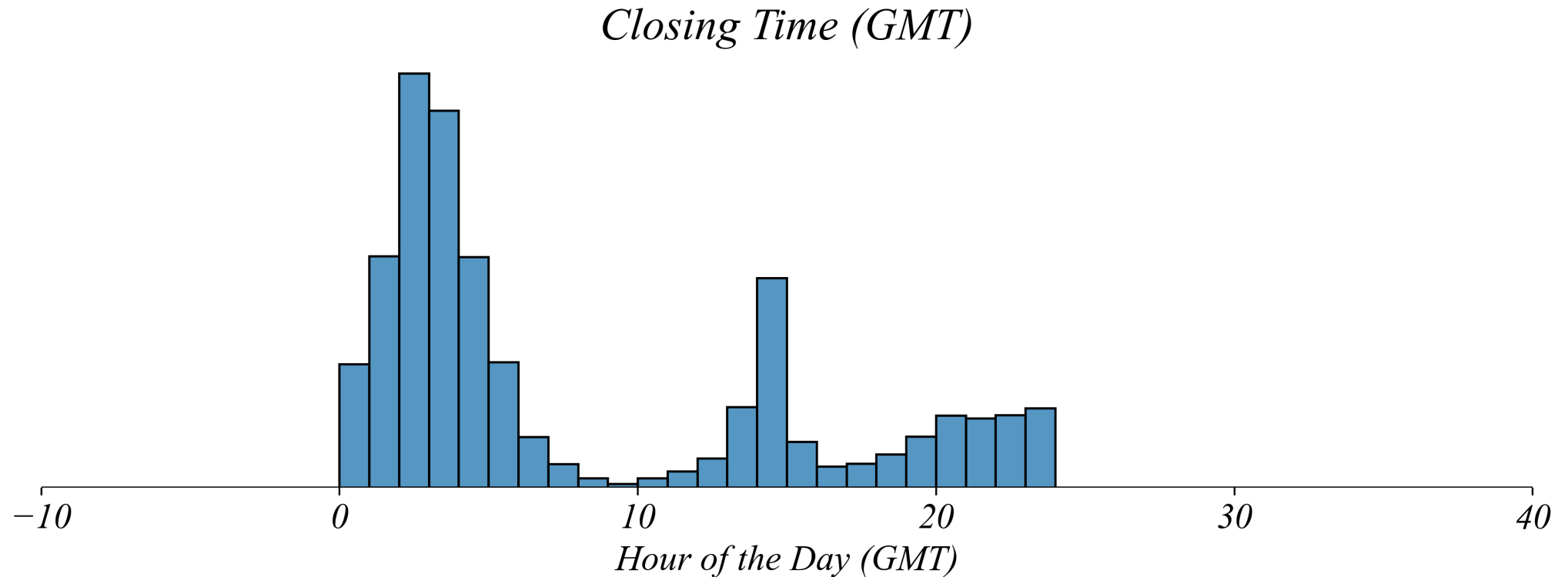
Lets add 24 if the number is negative and subtract 24 if the number is above 23.

```
1 # Normalize to 24 hours
2 hours['close_GMT'] = hours['close_GMT_simple'].mod(24)
```



Location Hours

So, how many locations are open at each hour of the day?



- > *this only tells us openings and closings at each hour, not total open*
- > *instead, lets sum up all the shops that have opened **that day***

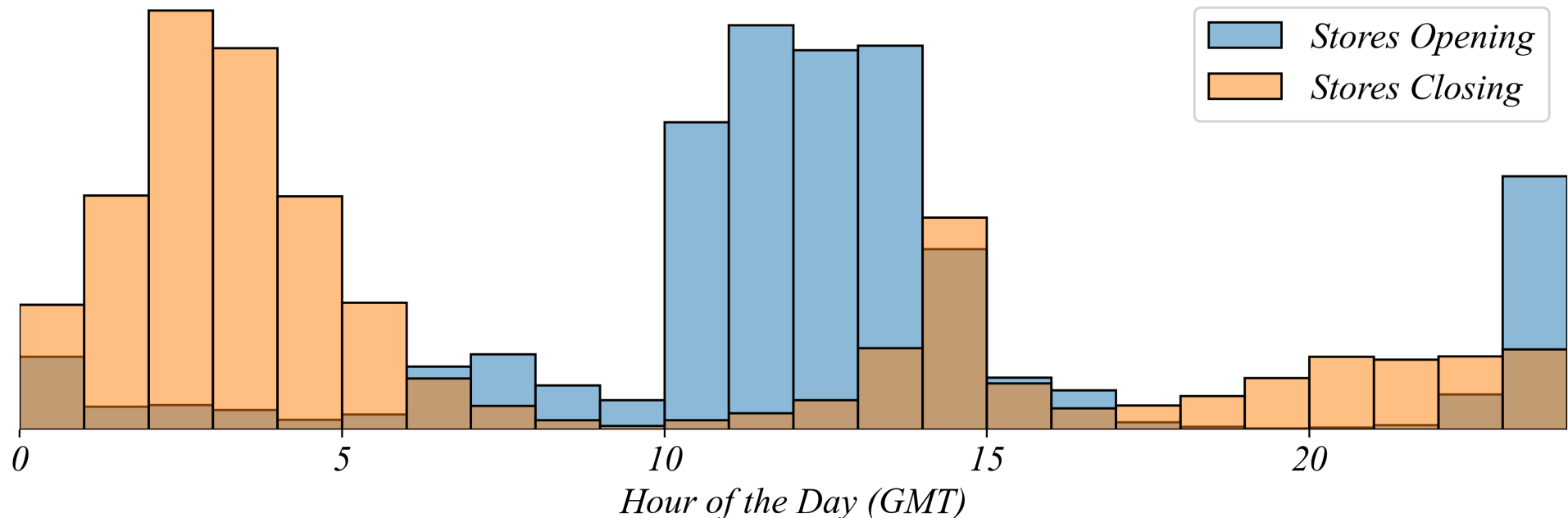
Location Hours

So, how many locations are open at each hour of the day?

> *lets sum up all the shops that have opened **that day***

```
1 # Construct values by bin
2 opened_values = hours['open_GMT'].value_counts().sort_index()
3
4 # Cumulative sum
5 total_opened = opened_values.cumsum()
```

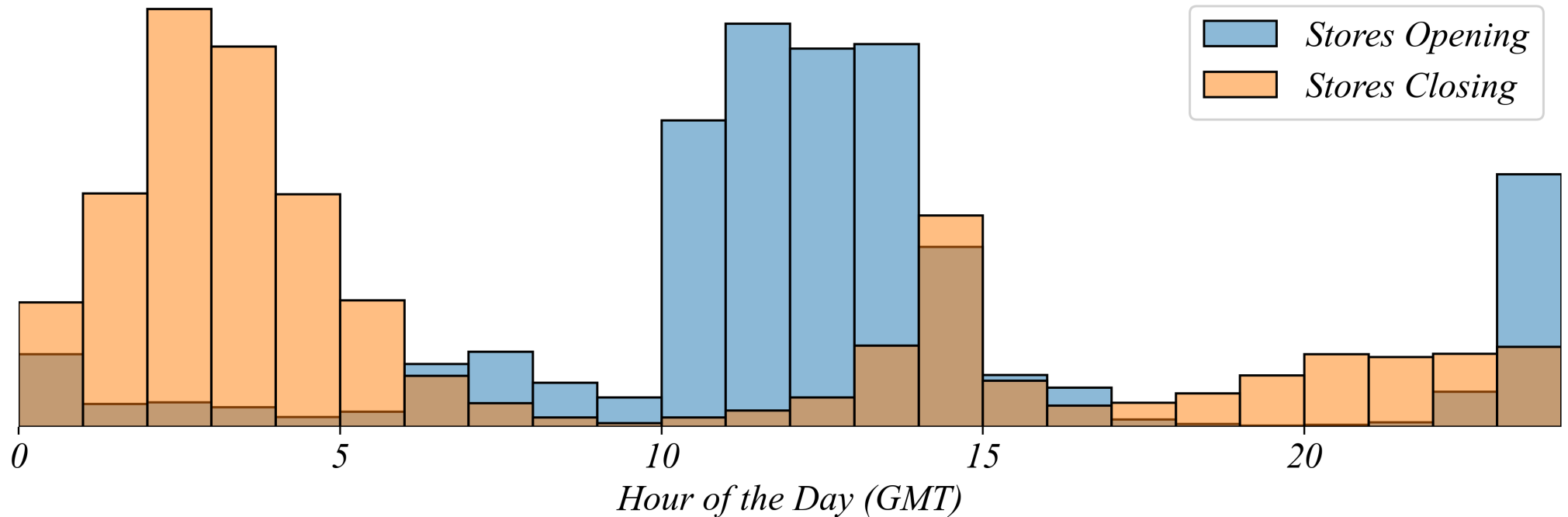
Opening and Closing Time (GMT)



Location Hours

So, how many locations are open at each hour of the day?

Opening and Closing Time (GMT)



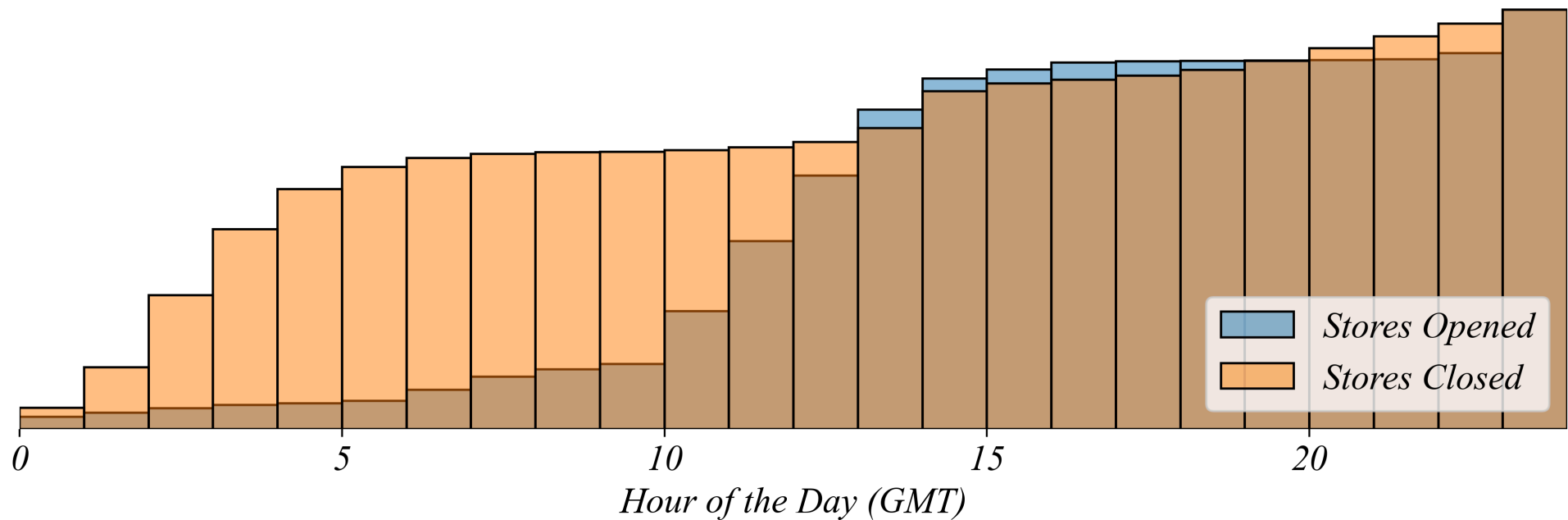
> from here, to find the total that have opened/closed, we take the difference

Location Hours

So, how many locations are open at each hour of the day?

```
1 # Take the difference  
2 net_increase = total_opened - total_closed
```

Opening and Closing Time (GMT)



> *why is the green line negative?*

> *lets add the number open at midnight (GMT).*

Location Hours

So, how many locations are open at each hour of the day?

```
1 # Add those open at midnight
2 count_open_after_close = len(hours[hours['open_GMT'] >= hours['close_GMT']])
3 cumulative_open = net_increase + count_open_after_close
```

Change in Starbucks Open Globally

