

ECON 0150 | Economic Data Analysis

The economist's data analysis pipeline.

Part 1.4 | Numerical Variables by Category

Hiring a Barista

Use Coffee_Sales_Receipts.csv to help inform where to hire a barista.

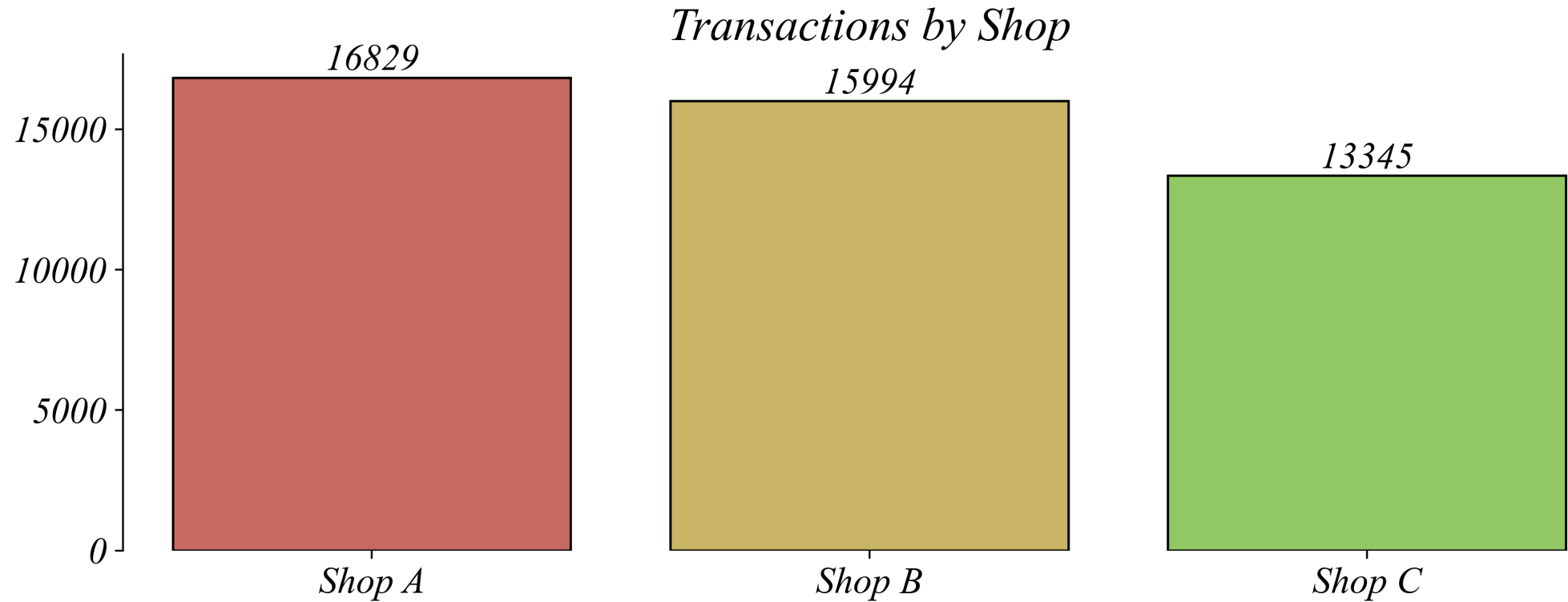
- *You manage three coffee shops and are considering where to hire a new barista.*
- *You have a dataset containing information about the transactions taking place at all three coffee shops throughout the day.*
- *Lets consider how to use this data to inform our decision.*

Hiring a Barista

Q. Which coffee shop is the busiest?

Hiring a Barista: Bar Graphs Compare Shops

Q. Which coffee shop is the busiest?



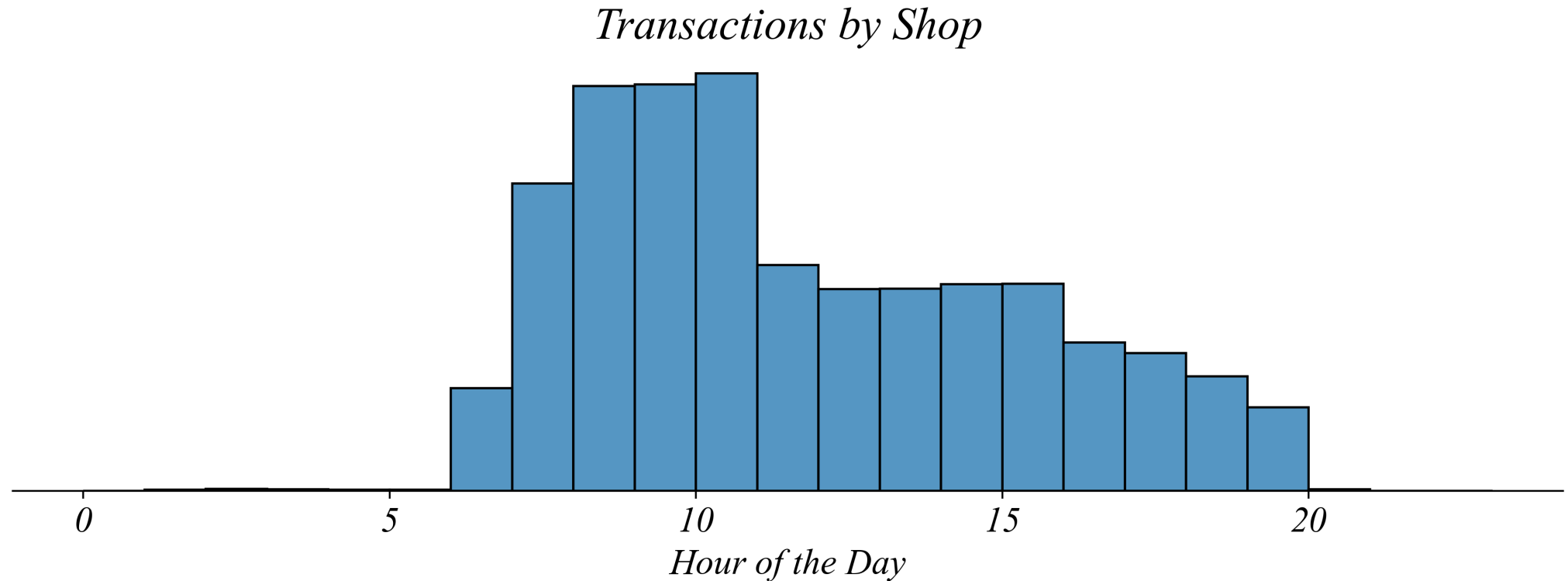
> a bar chart makes it easy to compare shops' busyness

Hiring a Barista

Q. What time of day is the busiest?

Hiring a Barista: Histograms Can Compare Times

Q. What time of day is the busiest?



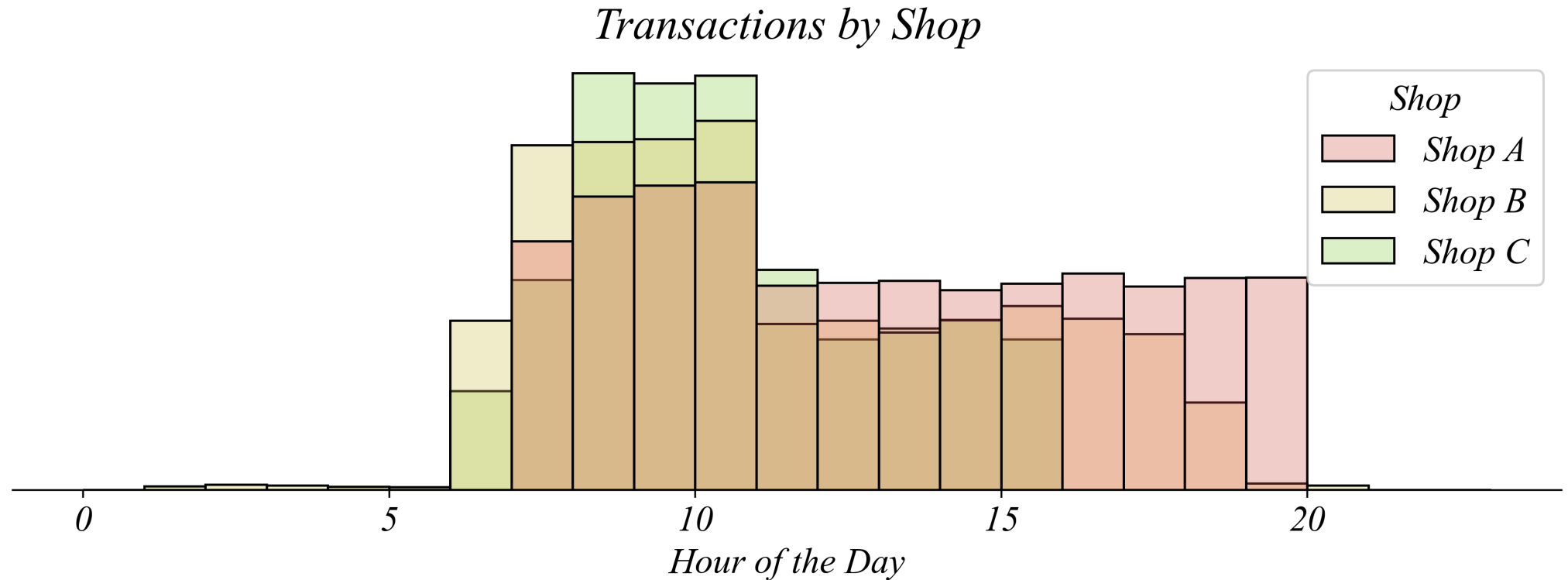
- > *a histogram makes it easy to compare transactions by time of day*
- > *does this mean the morning shift at Shop A is the busiest?*

Hiring a Barista

Q. Which shift is the busiest?

Hiring a Barista: Transactions by Shop

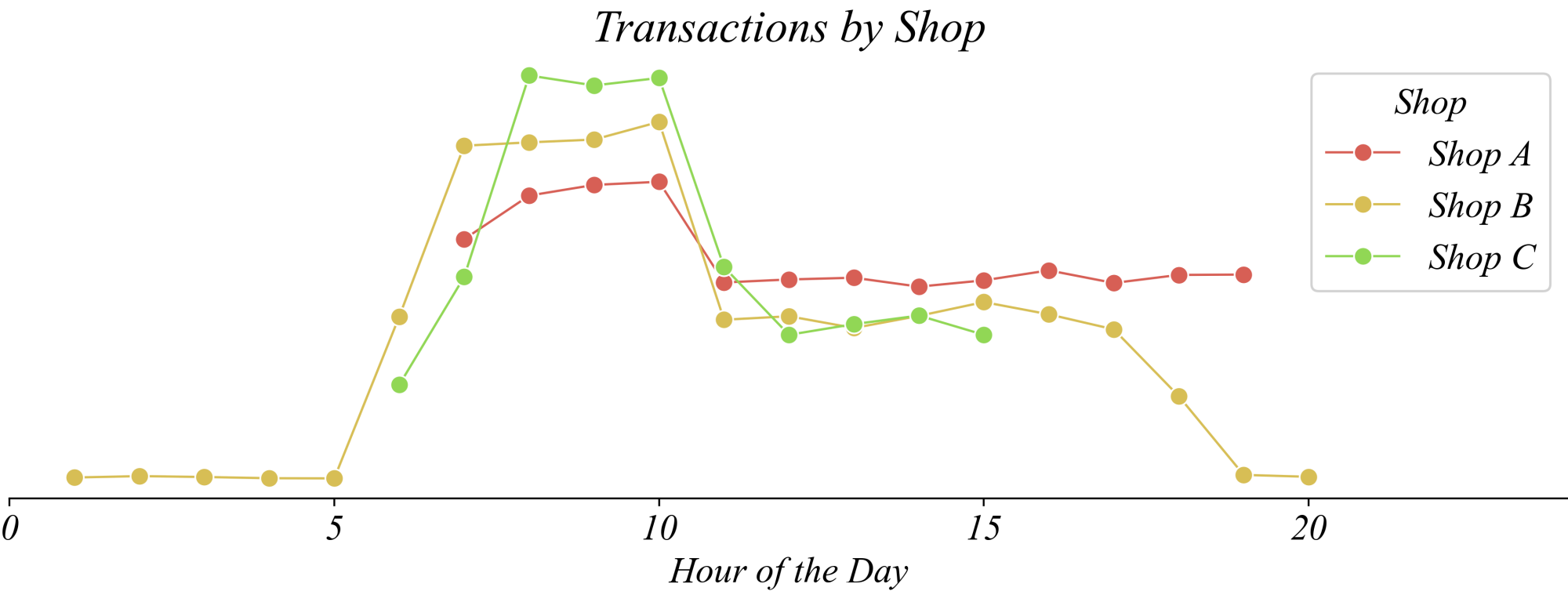
Q. Which shift is the busiest?



- > *an overlaid histogram can show all three groups*
- > *does this show the data clearly?*

Hiring a Barista: Transactions by Shop

Q. Which shift is the busiest?



> *instead, lets use a line graph*

Part 1.4 | Numerical Variables by Category

Summary

- *Categorical variables and continuous variables can give us different views of the same data.*
- *We can visualize both views on the same graph.*
- *Line graphs help simplify the visualization of multiple categories.*

Exercise 1.4 | Coffee Shop Transactions

Use Coffee_Sales_Receipts.csv to help inform where to hire a barista.

```
1 # Load Dataset
2 sales = pd.read_csv(file_path + 'Coffee_Sales_Receipts.csv')
3 sales.head()
```

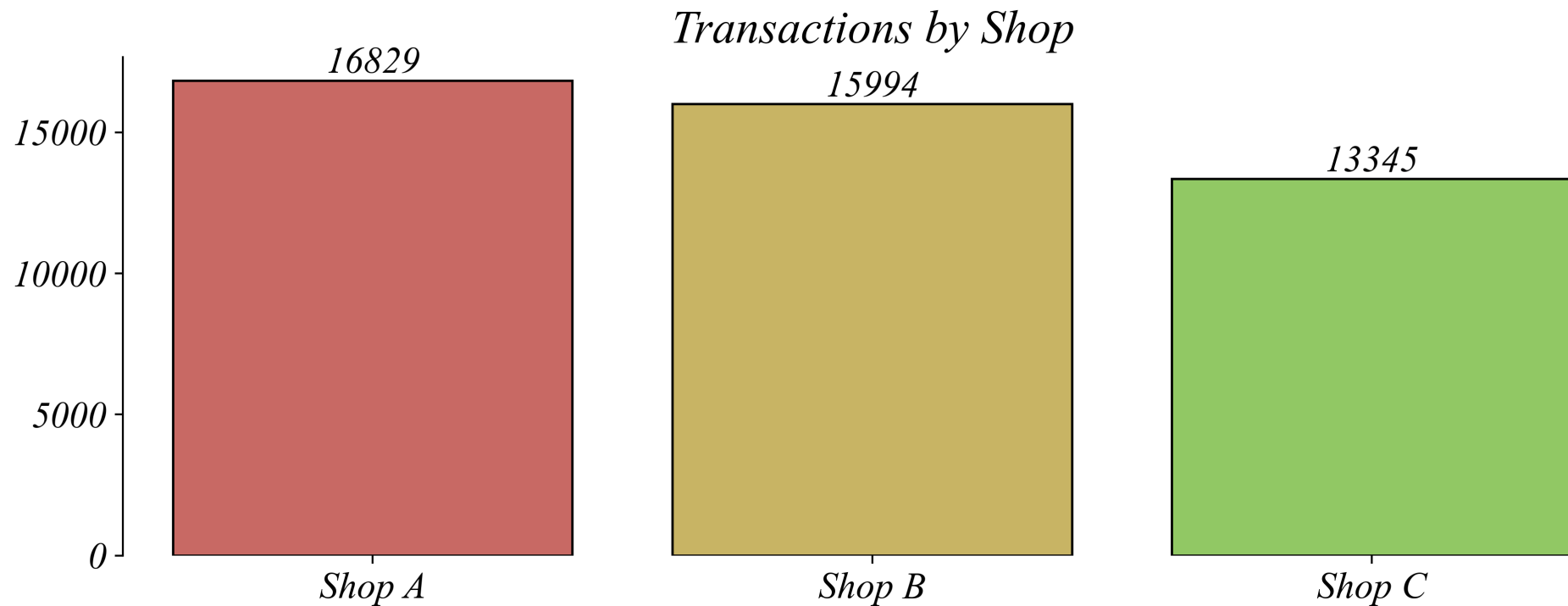
	Hours	Shop
0	12	Shop A
1	15	Shop A
2	14	Shop A
3	16	Shop A
4	19	Shop A

> you'll see a few more columns in your dataset

Exercise 1.4 | Bar Chart

Use Coffee_Sales_Receipts.csv to help inform where to hire a barista.

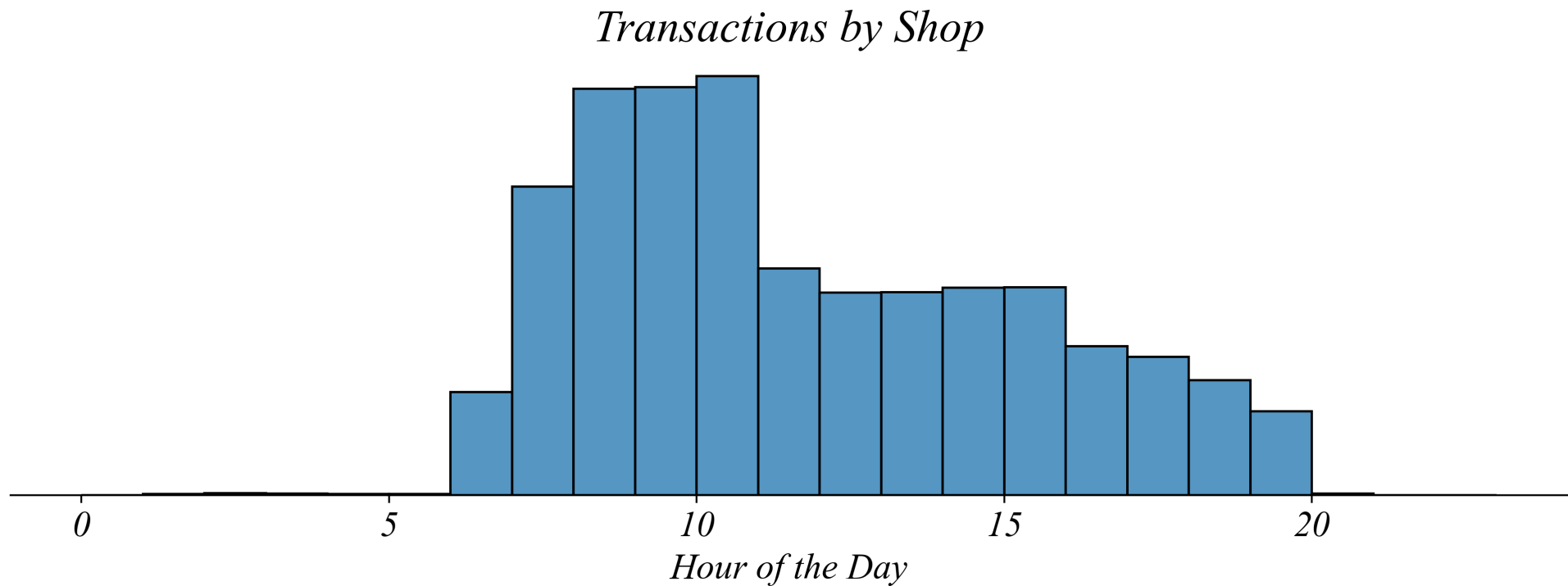
```
1 # Bar graph  
2 sns.countplot(data=sales, x='Shop', hue='Shop')
```



Exercise 1.4 | Histogram

Use Coffee_Sales_Receipts.csv to help inform where to hire a barista.

```
1 # Create a histogram
2 sns.histplot(data=sales, x='Hours', bins=range(0,24,1))
```



Exercise 1.4 | Multiple Line Graph

Q. Which shift is the busiest?

```
1 # Create hourly counts by shop
2 hourly_counts = sales.groupby(['Shop', 'Hours']).size().reset_index(name='Count')
3
4 # Figure
5 sns.lineplot(data=hourly_counts, x='Hours', y='Count', hue='Shop')
```

	Shop	Hours	Count
0	Shop A	7	1383
1	Shop A	8	1632
2	Shop A	9	1693
3	Shop A	10	1711
4	Shop A	11	1136

Exercise 1.4 | Multiple Line Graph

Q. Which shift is the busiest?

```
1 # Create hourly counts by shop
2 hourly_counts = sales.groupby(['Shop', 'Hours']).size().reset_index(name='Count')
3
4 # Figure
5 sns.lineplot(data=hourly_counts, x='Hours', y='Count', hue='Shop')
```

