ECON 0150 | Economic Data Analysis

How economists analyze data.

Part 1.4 | Numerical Variables by Category

Example 1.4 | Coffee Shop Transaction

Examine Coffee_Sales_Receips.csv

```
# Load the data
sales = pd.read_csv(file_path + file_name)
```

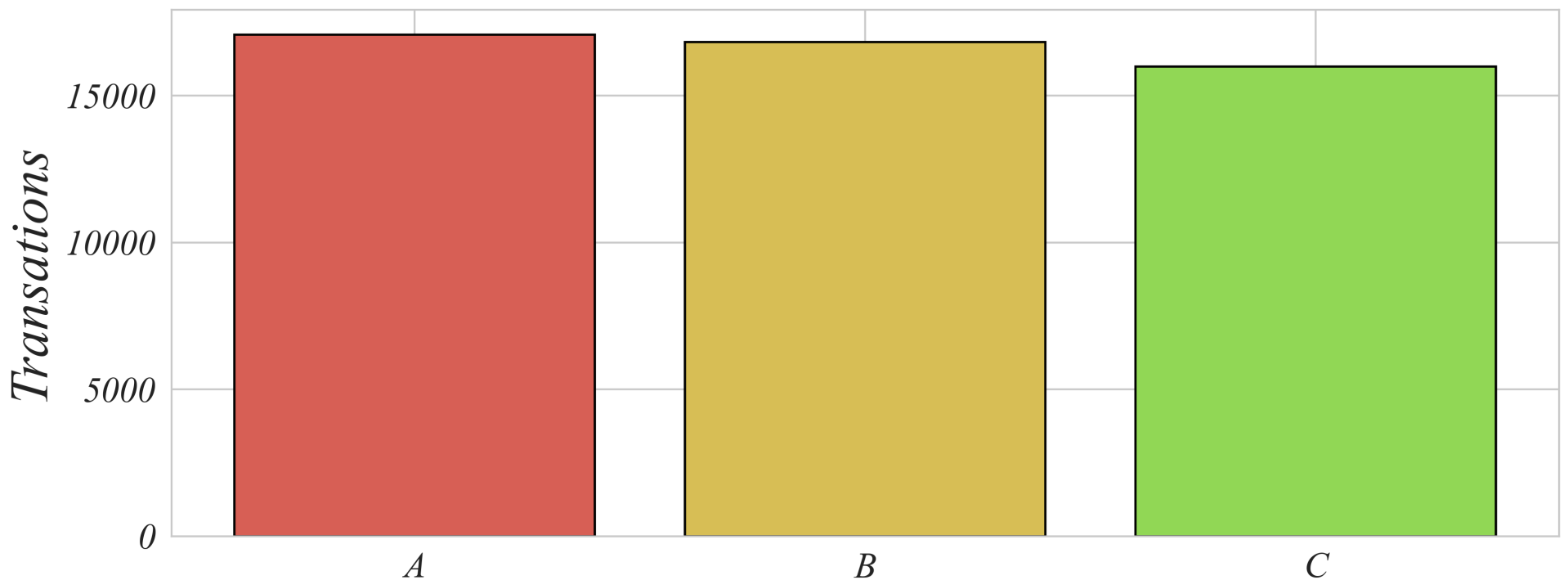
Part 1.4 | Numerical Variables by Category

Q. Which coffee shop is the busiest?

Part 1.4 | Numerical Variables by Category

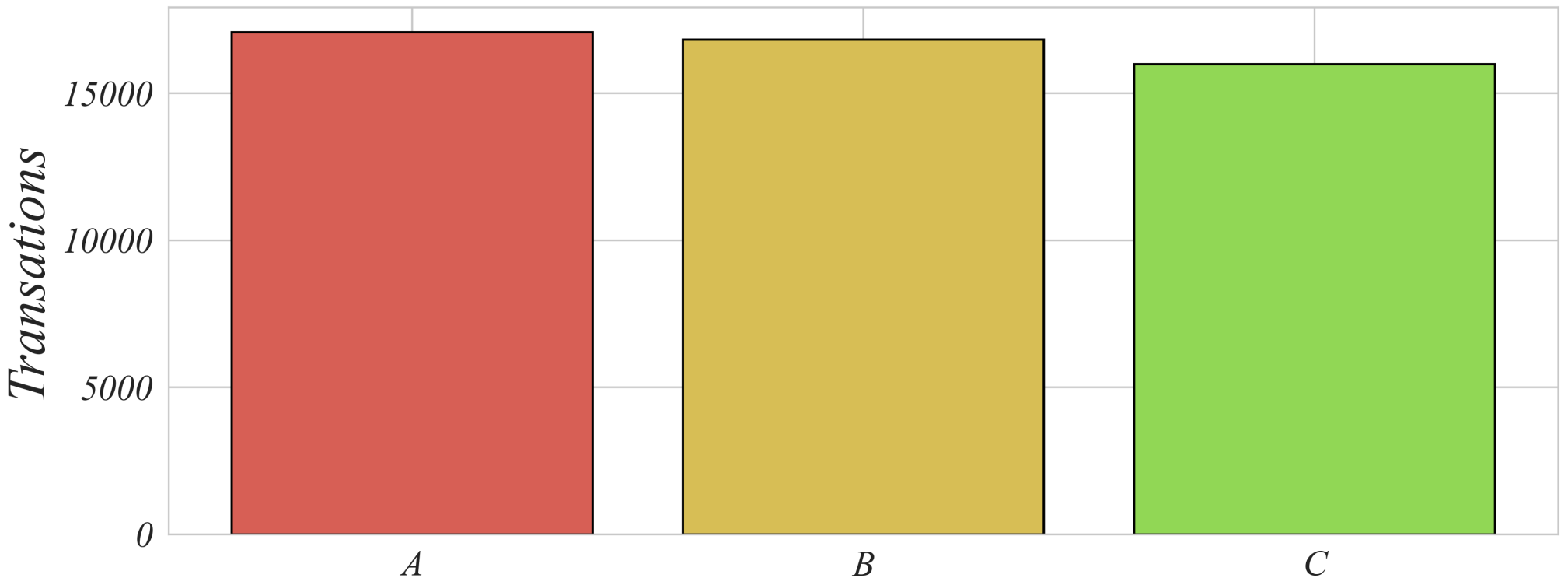
Q. Which coffee shop is the busiest?

```
# Count by category
sales_counts = sales['sales_outlet_id'].value_counts()

# Bar graph
plt.bar(['A','B','C'], sales_counts.values)
```



Part 1.4 | Numerical Variables by Category

Q. Which coffee shop is the busiest?

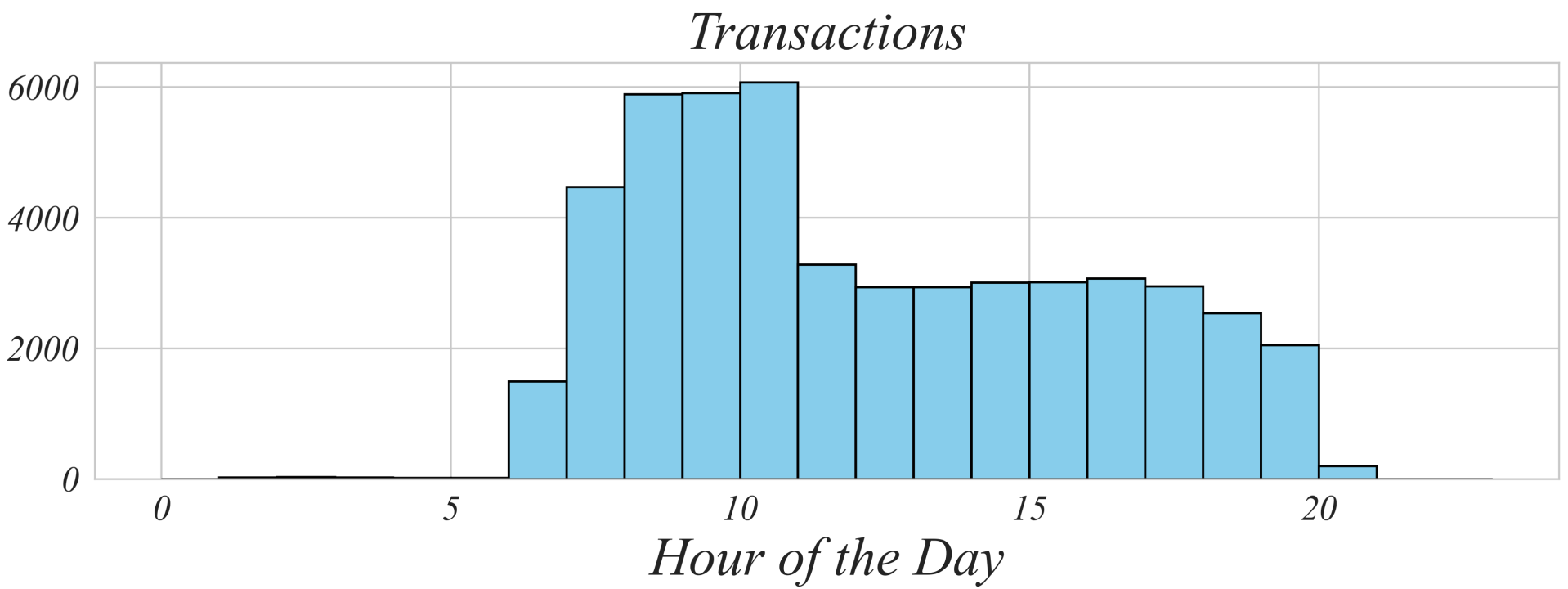


> a bar chart makes it easy to compare between categories

Part 1.4 | Numerical Variables by Category

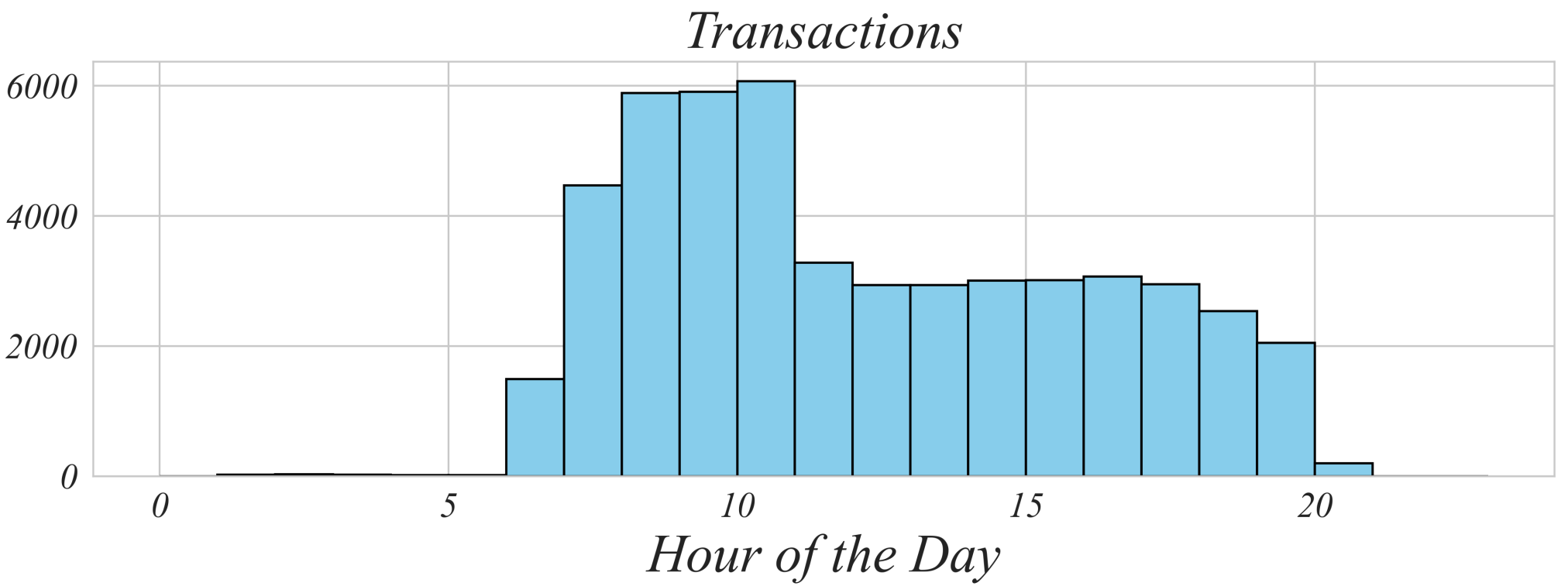
Q. What time of day is the busiest?

```
# Create bins from 0 to 24
bins = range(0, 24, 1)

# Create a histogram
plt.hist(sales['Hours'], bins=bins)
```



Part 1.4 | Numerical Variables by Category

Q. What time of day is the busiest?



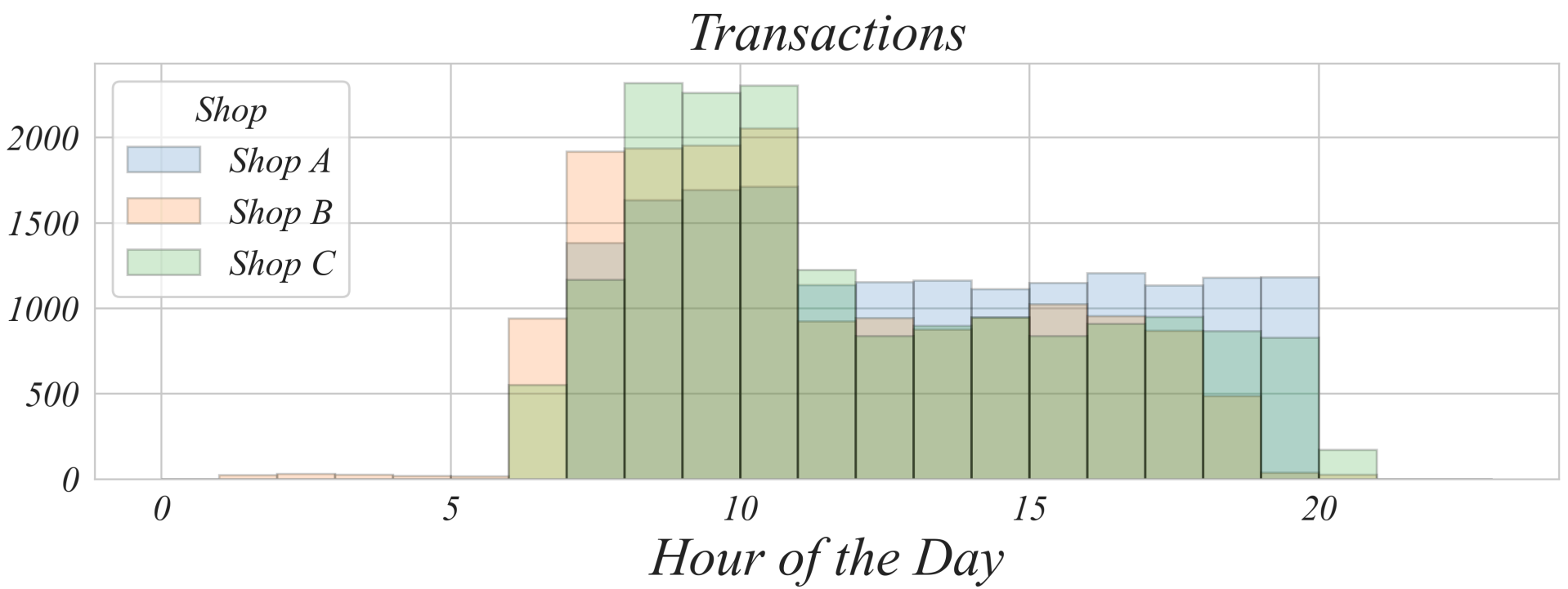
> a histogram makes it easy to compare time of day

> does this mean the morning shift at Shop A is the busiest?

Part 1.4 | Numerical Variables by Category

Q. Which shift is the busiest?

> an overlaid histogram can show all three groups



> does this show the data clearly?

Part 1.4 | Numerical Variables by Category

Q. Which shift is the busiest?

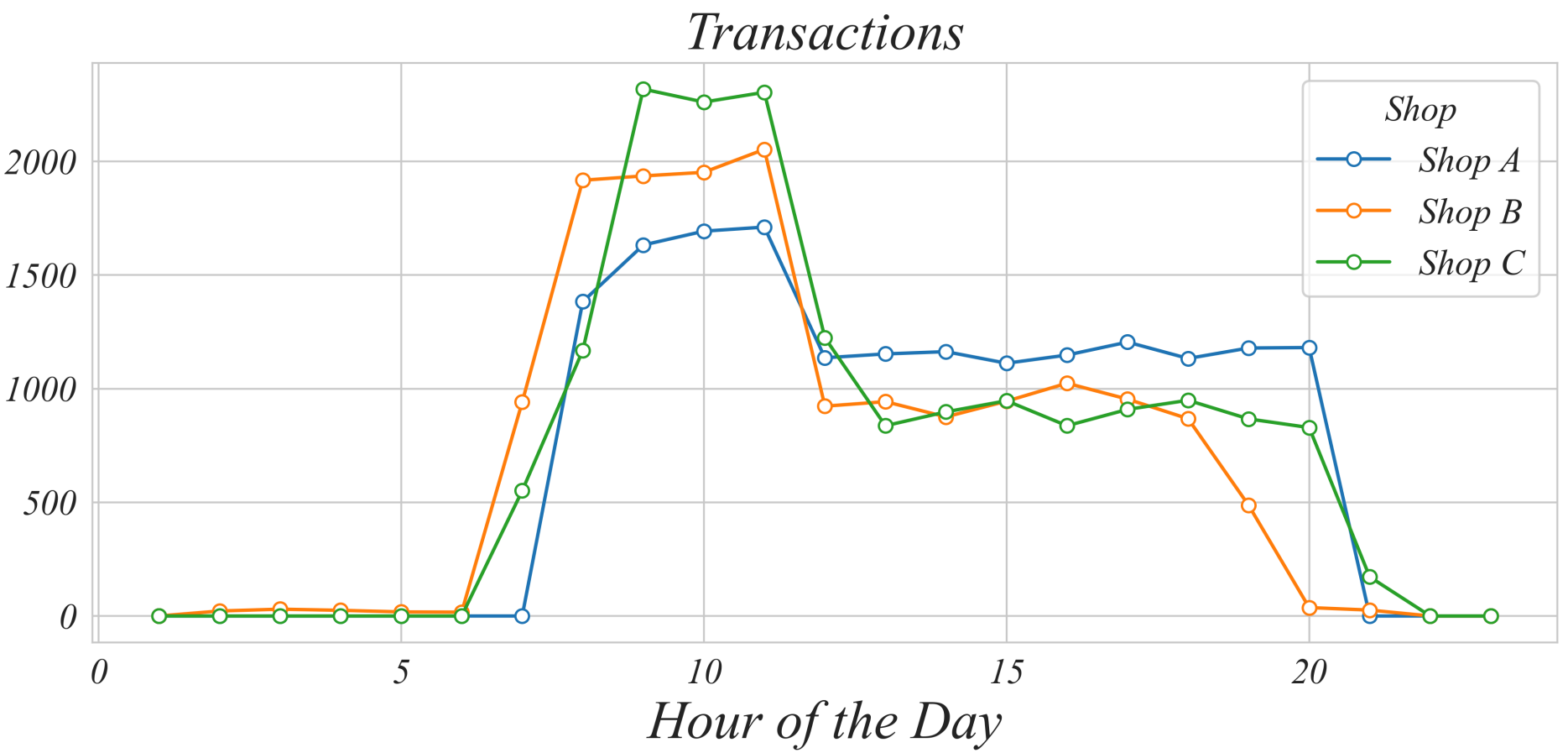
> instead, lets use a line graph

```
# Select Shop A data, summarize, and sort
shop_A = sales[sales.Shop == 'A'].Hours.value_counts().sort_index()

# Plot Shop A
plt.plot(shop_A, label='Shop A')

# Shop B
shop_B = sales[sales.Shop == 'B'].Hours.value_counts().sort_index()
plt.plot(shop_B, label='Shop B')

# Shop C
shop_C = sales[sales.Shop == 'C'].Hours.value_counts().sort_index()
plt.plot(shop_C, label='Shop C')
```

Part 1.4 | Numerical Variables by Category

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.
- Often we can visualize both views on the same graph using visualization techniques for continuous variables within the category.
- Line graphs help simplify the visualization of multiple categories.