

In [21]: `from completejourney_py import get_data`

```

cj_data = get_data()

import pandas as pd
transactions = cj_data['transactions']
products = cj_data['products']
demographics = cj_data['demographics']

df1 = transactions.merge(demographics, how='left', on='household_id')

pizza_filter = products['product_type'].str.contains('pizza', case=False, na=False)
df2 = (products[pizza_filter].merge(df1, how='inner', on='product_id')
      .groupby(['marital_status'], as_index=False)
      .agg({'product_type': 'count', 'sales_value': 'sum'})
      )
df2 = df2.rename(columns={'product_type': 'pizza_product_purchased'})

df3 = (products[pizza_filter]
      .merge(transactions, how='inner', on='product_id')
      .merge(demographics, how='inner', on='household_id')
      )

df4 = df3[['transaction_timestamp', 'sales_value']]
sale_resample = df4.resample('D', on='transaction_timestamp').sum()

from bokeh.plotting import figure, output_file, show
from bokeh.resources import INLINE
import bokeh.io

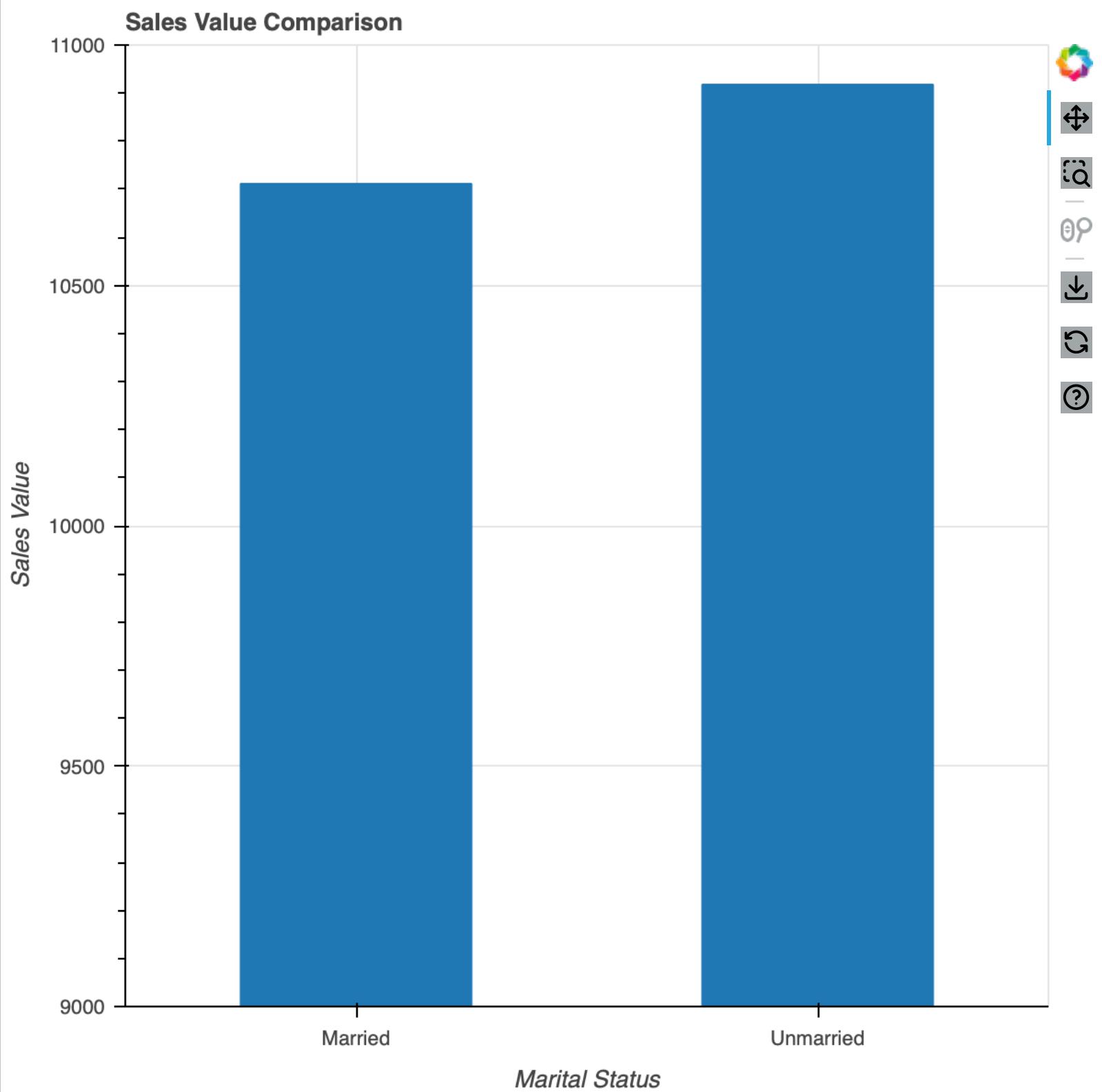
bokeh.io.output_notebook()
```



BokehJS 3.3.4 successfully loaded.

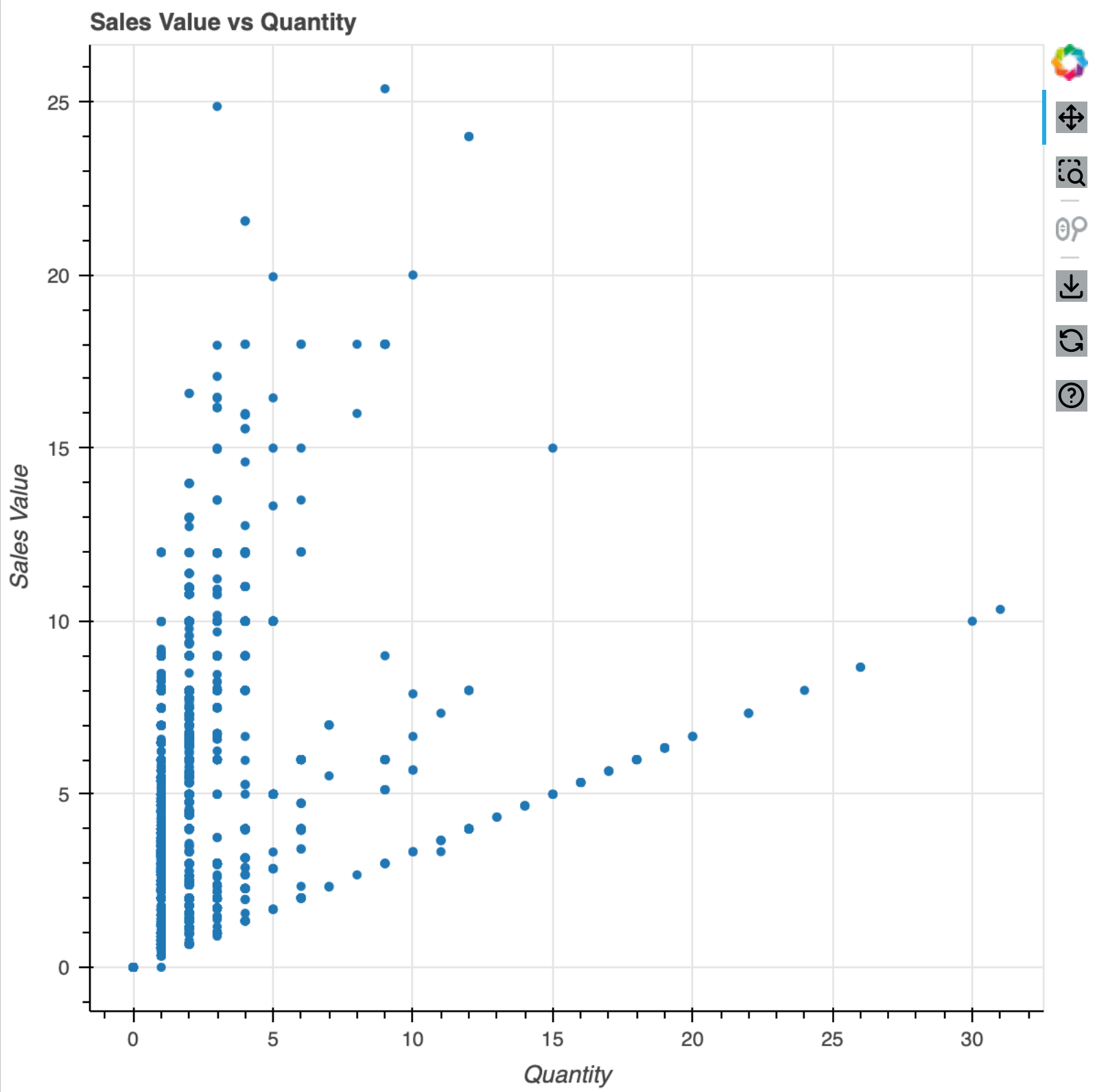
In [2]: `categories = df2['marital_status']
data = df2['sales_value']

p = figure(title='Sales Value Comparison', x_range=categories, x_axis_label='Marital Status', y_axis_label='Sales Value', y_range=(9000, 11000))
p.vbar(x=categories, top=data, width=.5)
show(p)`

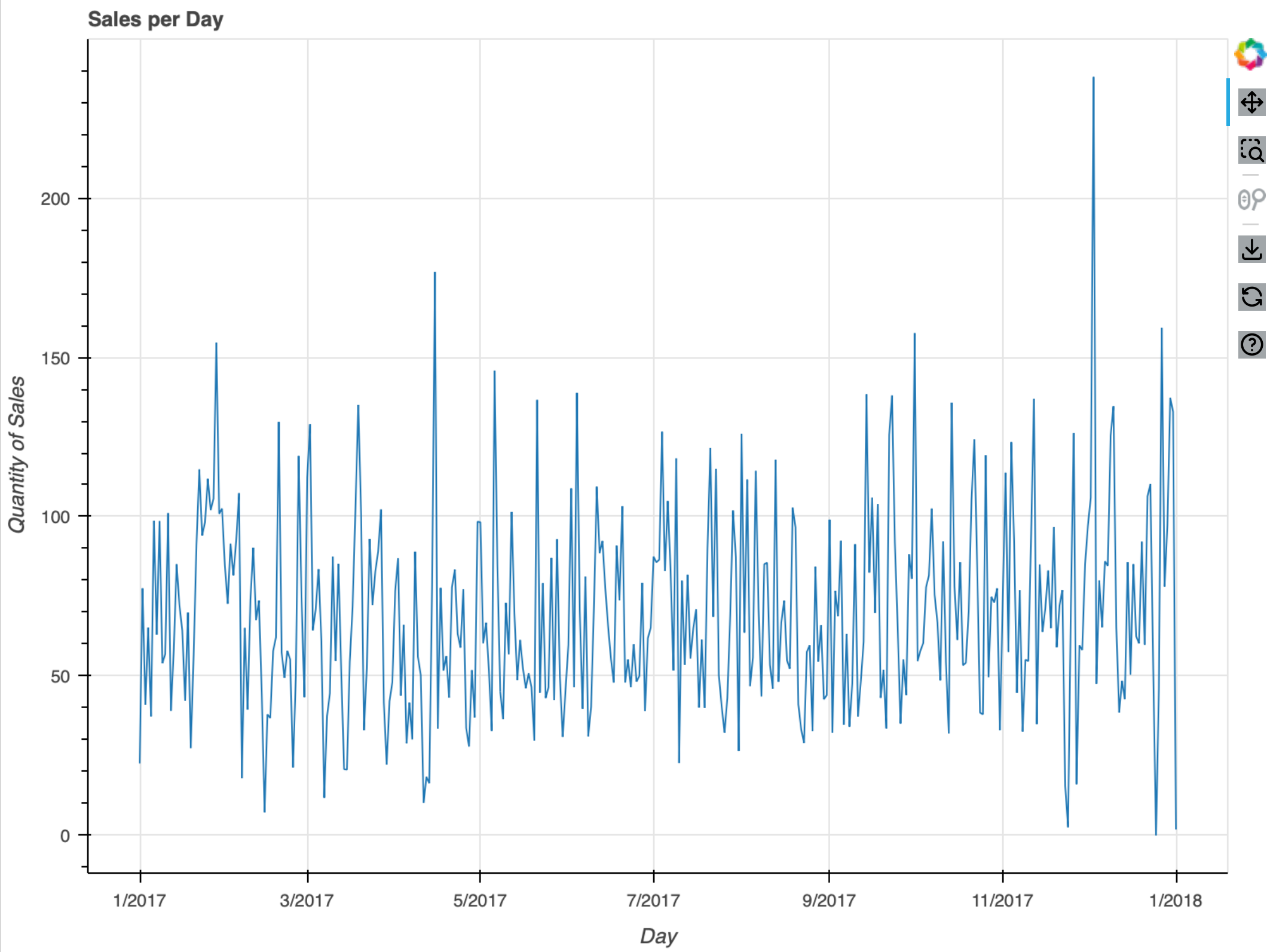


In [3]: `y = df3['sales_value']
x = df3['quantity']

p = figure(title='Sales Value vs Quantity', y_axis_label='Sales Value', x_axis_label='Quantity')
p.scatter(x, y)
show(p)`



In [31]: `p = figure(x_axis_type='datetime', title='Sales per Day', x_axis_label='Day', y_axis_label='Quantity of Sales', width=800)
p.line(x='transaction_timestamp', y='sales_value', source=sale_resample)
show(p)`



In []: