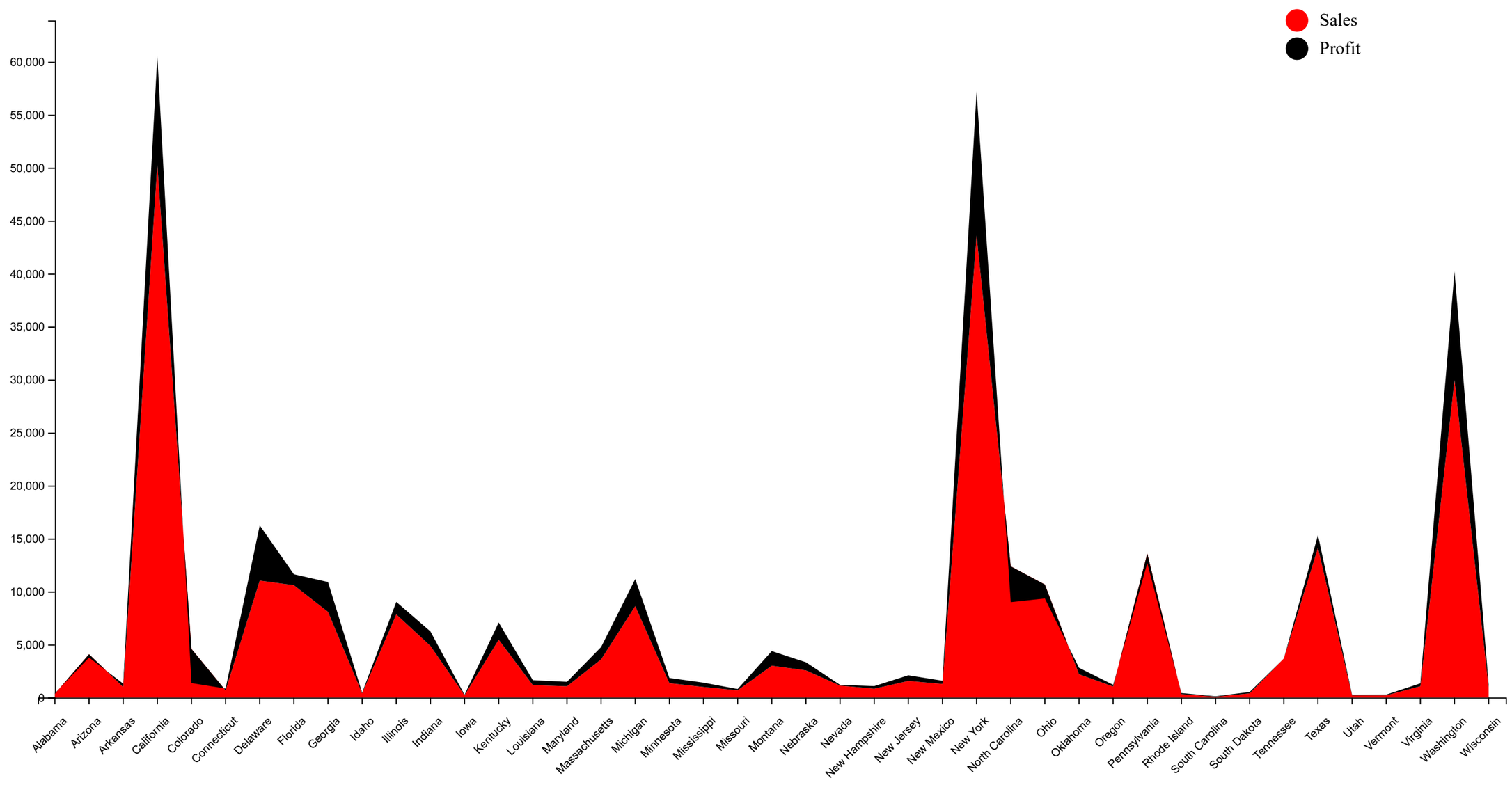
Assignment 7

Data visualization if often used to visualize trends in massive datasets. These visualizations do not provide accurate inference on the numbers but are instead used to infer a specific trend or behavior in the data. In this assignment, we are going to look at different types of visualizations that can be used to visualize trends. We will use an E-commerce dataset to the the same.

Question 1:

Data visualization is a very powerful tool to visualize trends in E-commerce datasets. Create a streamgraph to display profit and sales for the Technology Category after the year 2016 (>2016) for every state. Provide appropriate scale, axes, and legend to the plot. Do NOT hard code the scale/axes/legend in the html file. Associate different colors for profit and sales data points. For more details refer to expected_output.pdf.

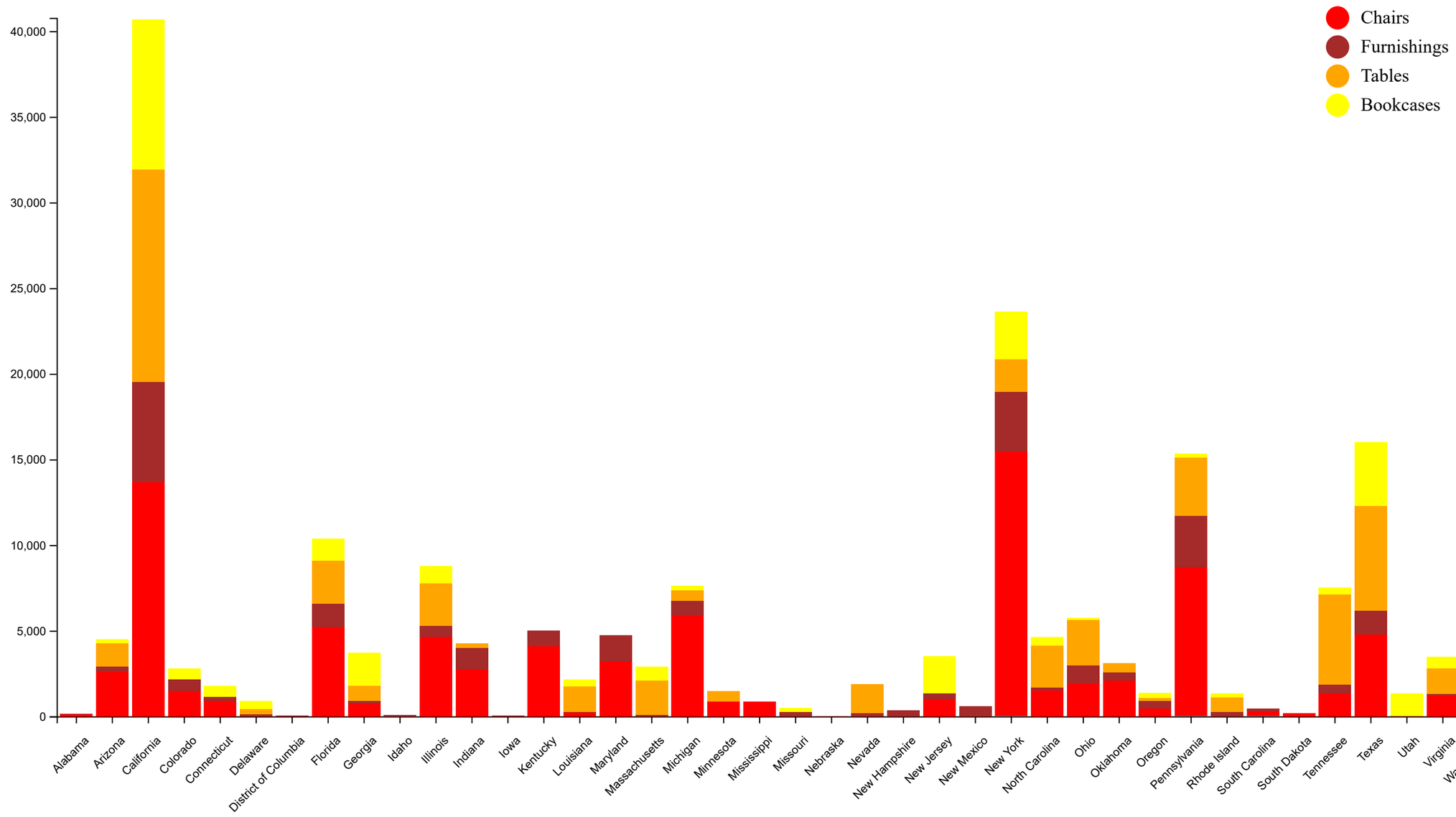
Plot:



Question 2:

For our next visualization we will see how sales are distributed between the sub-categories of the products. For this, you will create a stacked bar plot to display the sales of four sub-categories (Chairs, Furnishings, Tables, Bookcases) in the category of Furniture after the year 2016 (>2016). Provide appropriate scale, axes, and legend to the plot. Do NOT hard code the scale/axes/legend in the html file. Associate different colors for profit and sales data points. For more details refer to expected_output.pdf.

Plot:



Question 3:

Geographical location is a very important factor for many decisions in E-commerce. In this section we will see how the sales are distributed among the states by using a geomap. You have been provided with a symbol converter (stateSym) in app.js. You have also been provided with us-states.json (feature file to draw a geomap). You are required to to draw a geomap using the feature file provided. One observation you will make while creating the geomap is

that the variance between the sales values of different states is huge. Due to this the color coding is not consistent. One way to normalize the sales values is to use a log scale. For more details refer to expected_output.pdf. Note: The expected output follows a log scale.

Plot: