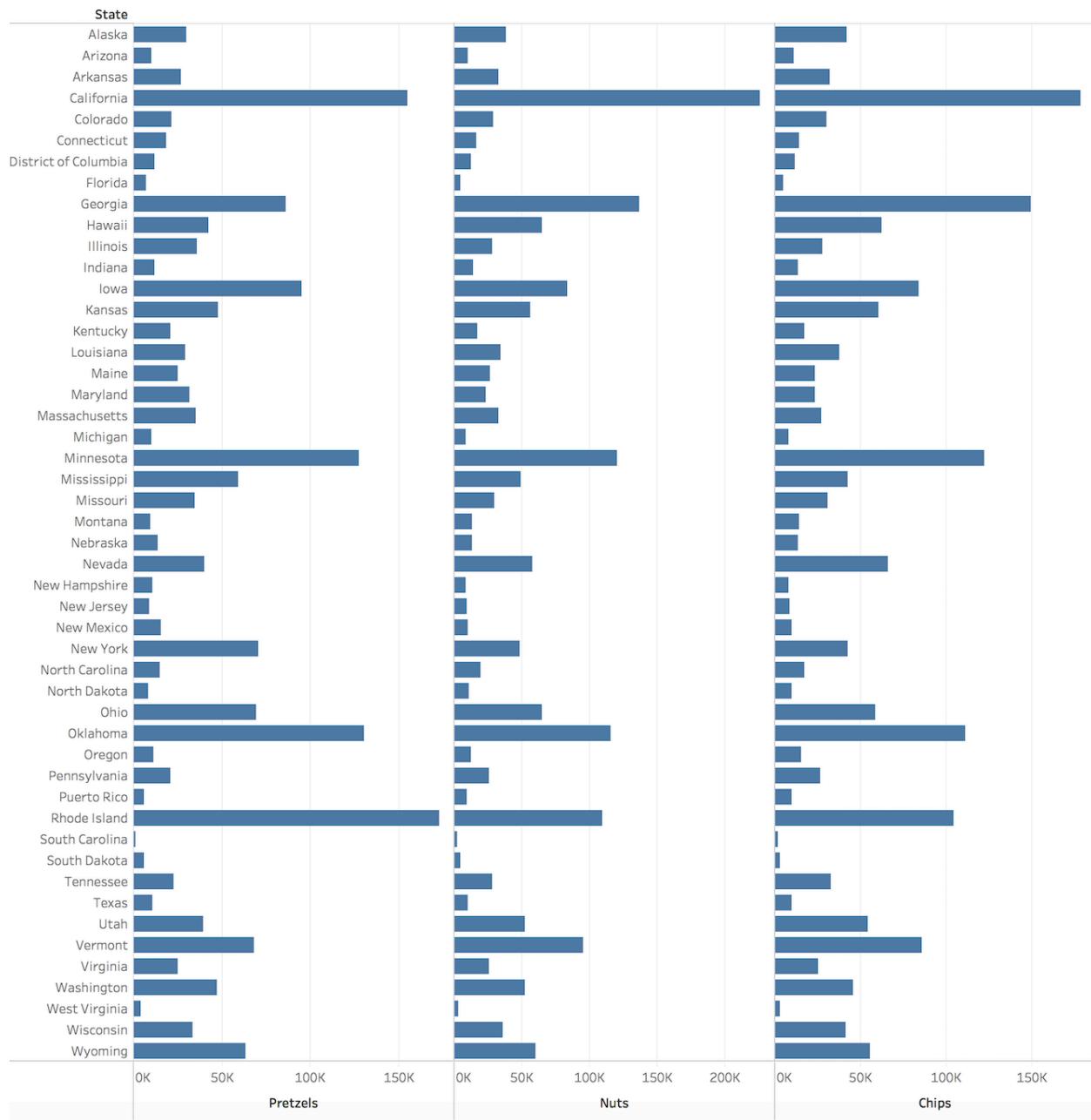


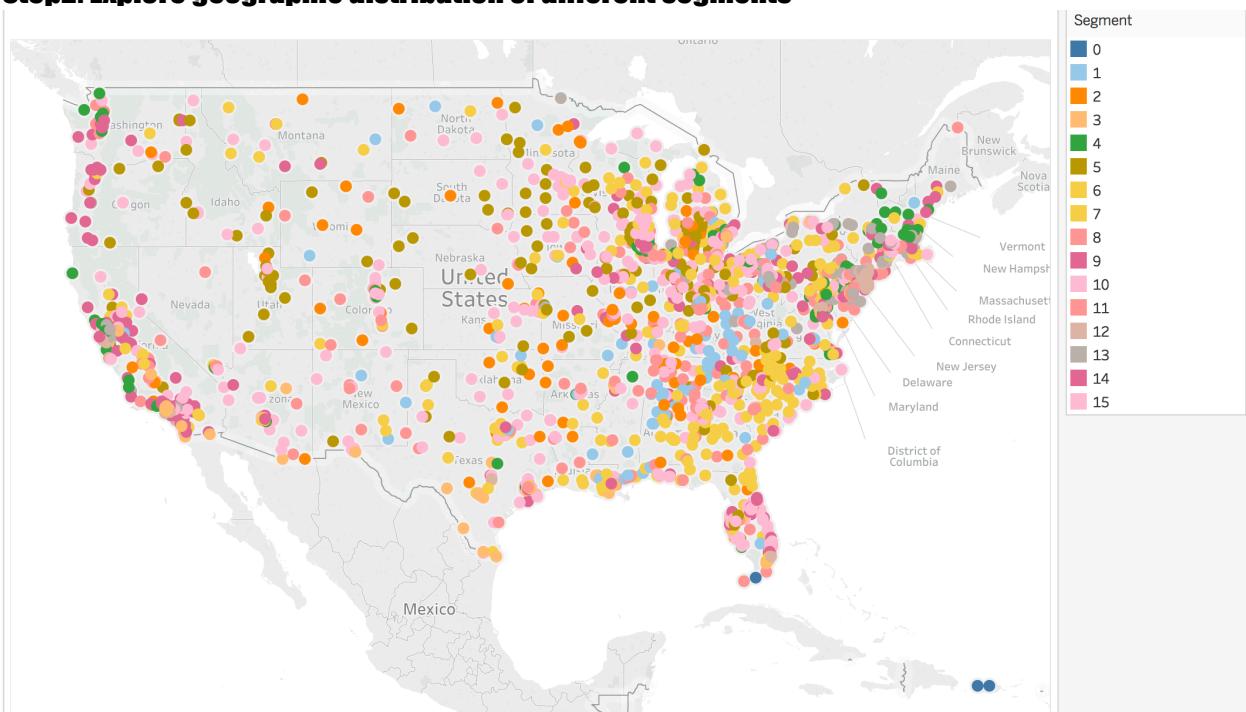
1. Exploratory Data Analysis

Step1: Explore sales of the three categories by state

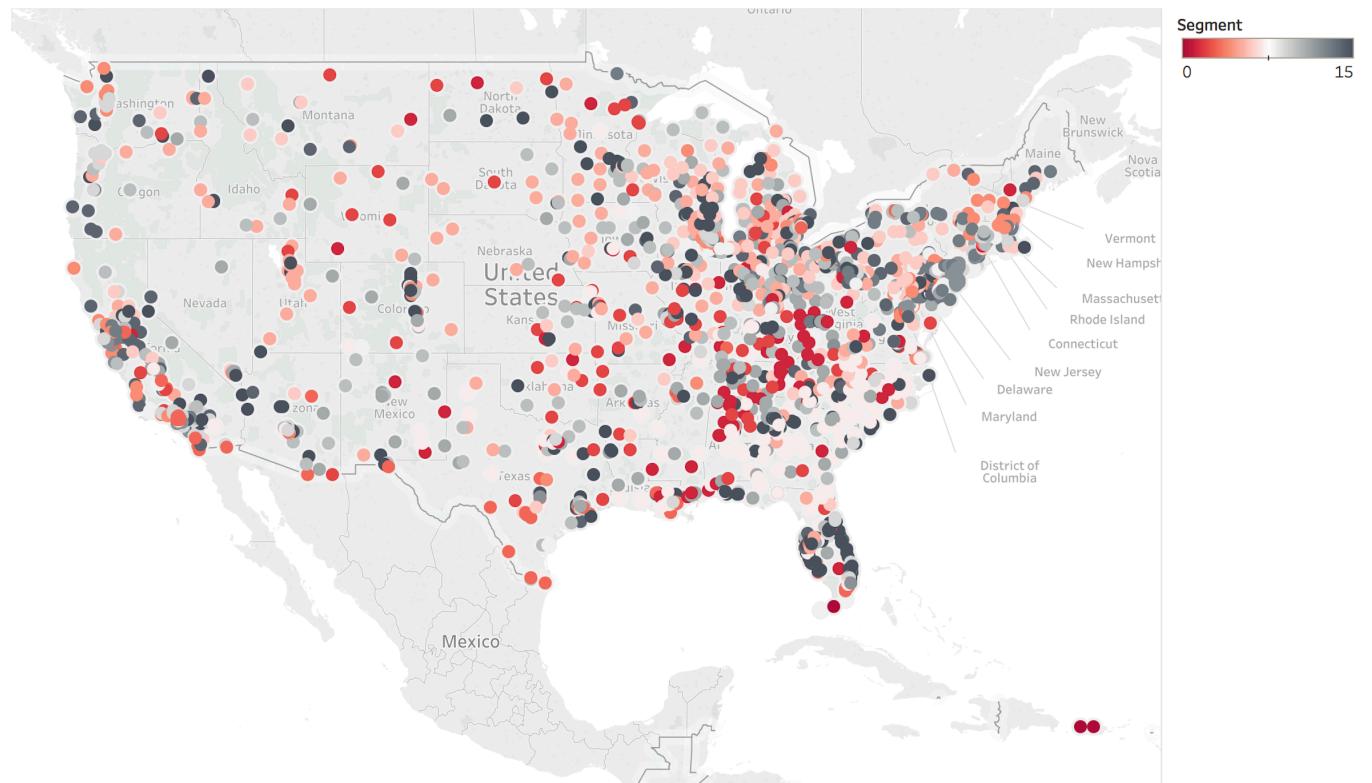


According to the bar plot, it seems the sales of chips, pretzels, and nuts do not show explicit complementary patterns. Rather, they are largely positively correlated, especially for those states with large sales volumes.

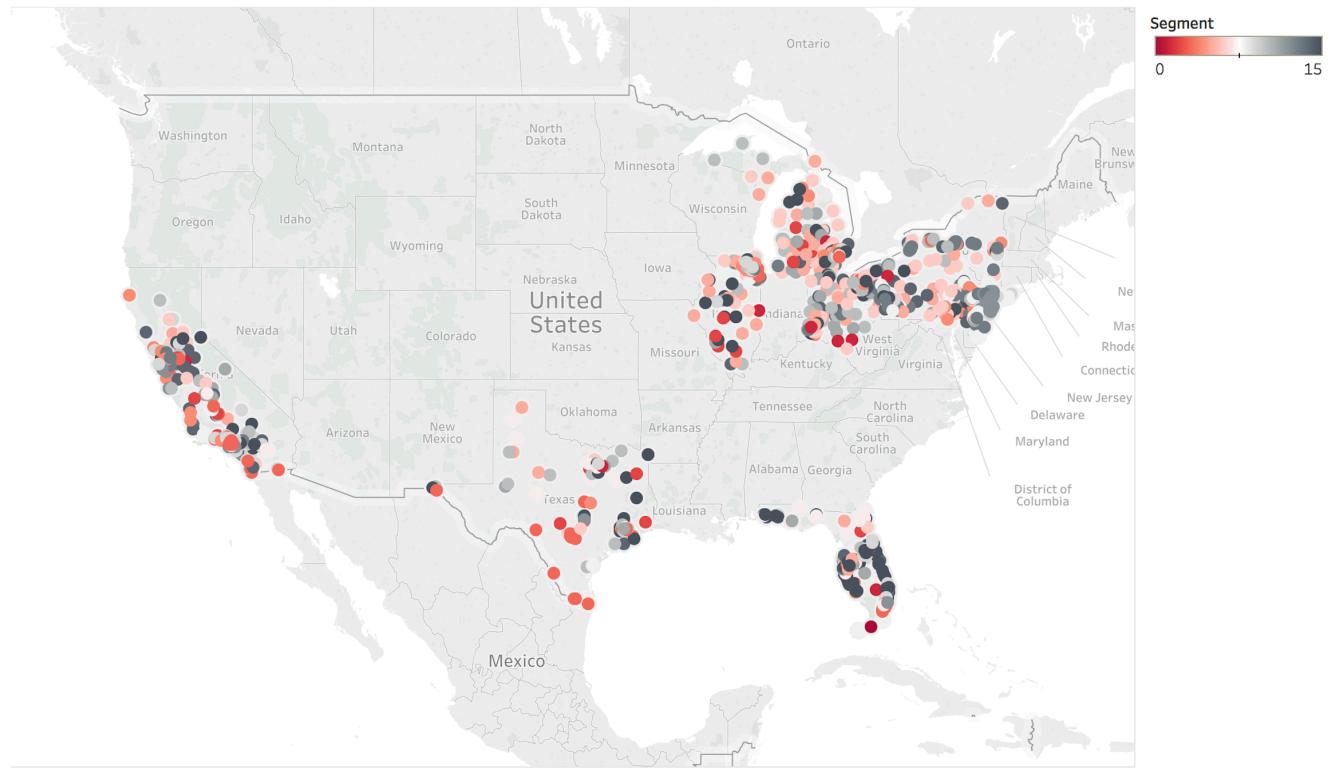
Step2: Explore geographic distribution of different segments



According to the discrete segment mapping by zip code, the clustering is not strictly based on geographic location.



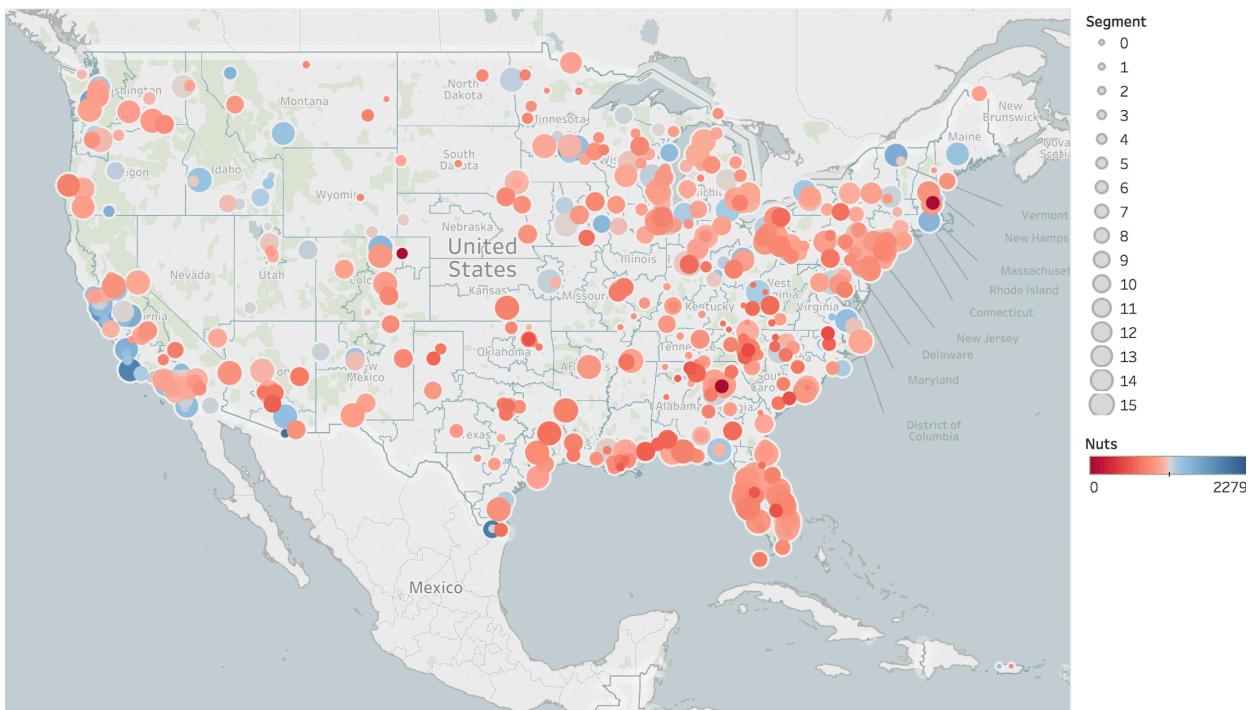
Since 15 segments are a little overwhelming to visually distinguish by color, I also mapped them as a continuous variable to see if any patterns as the segment number grows. It seems the black dots are more concentrated around California, Georgia, and northeast coast. To test the assumption that the larger segment number, the greater total sales, I filter the states with long bars in the first image. The filtered result is showed as below.



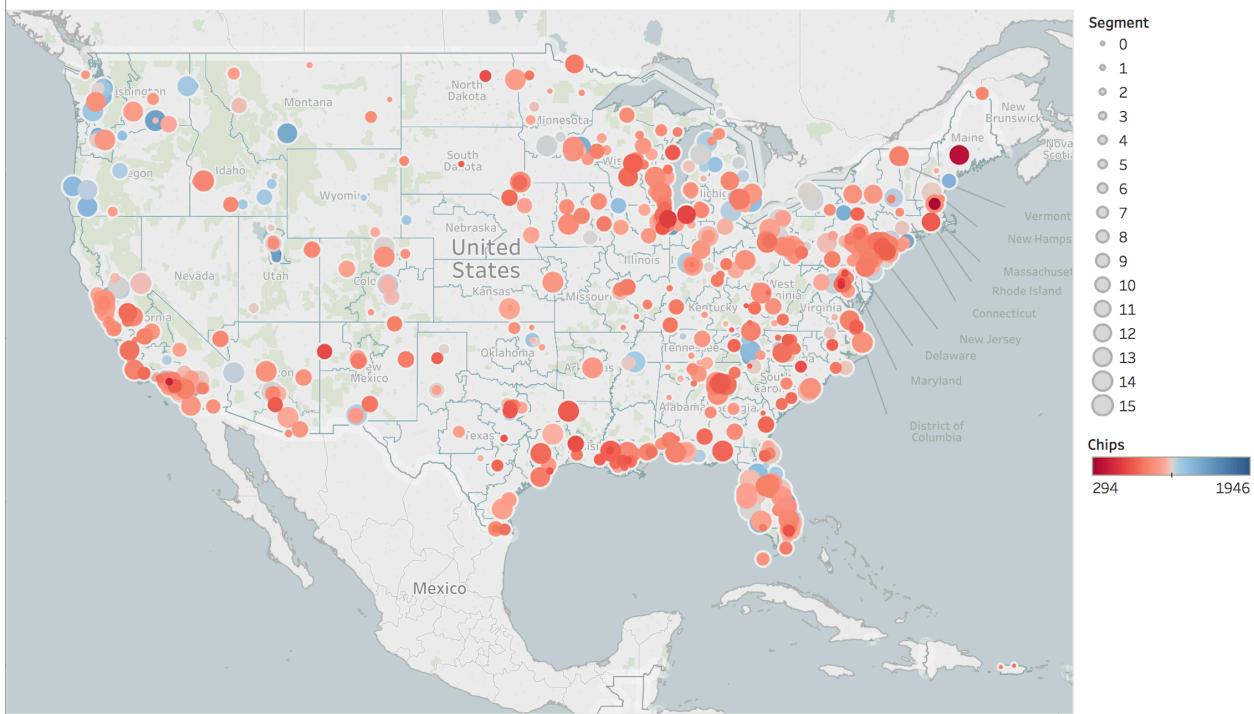
The output partially holds true because black ones are playing larger proportion now, and since the map is based on a smaller scale, zip code, it is possible that even the whole state generate great sales, some areas within the state still has low sales.

Step3: Explore sales of the each category by zip code and segment

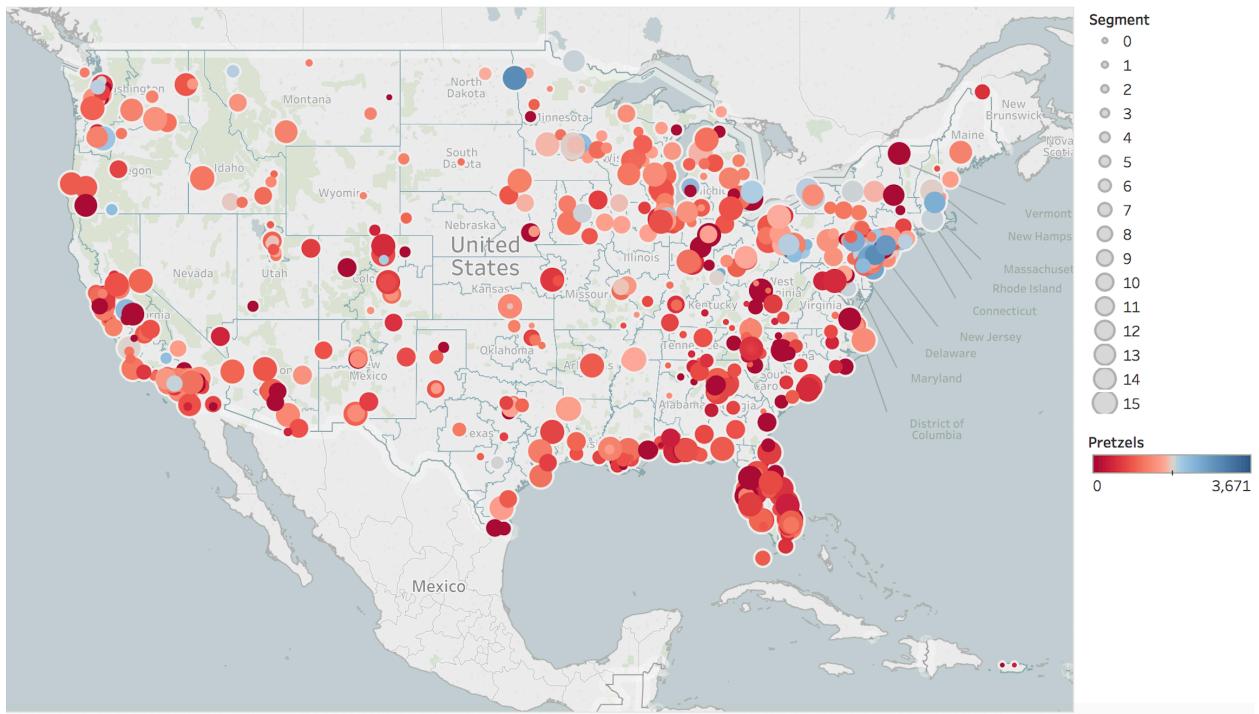
nuts



chips

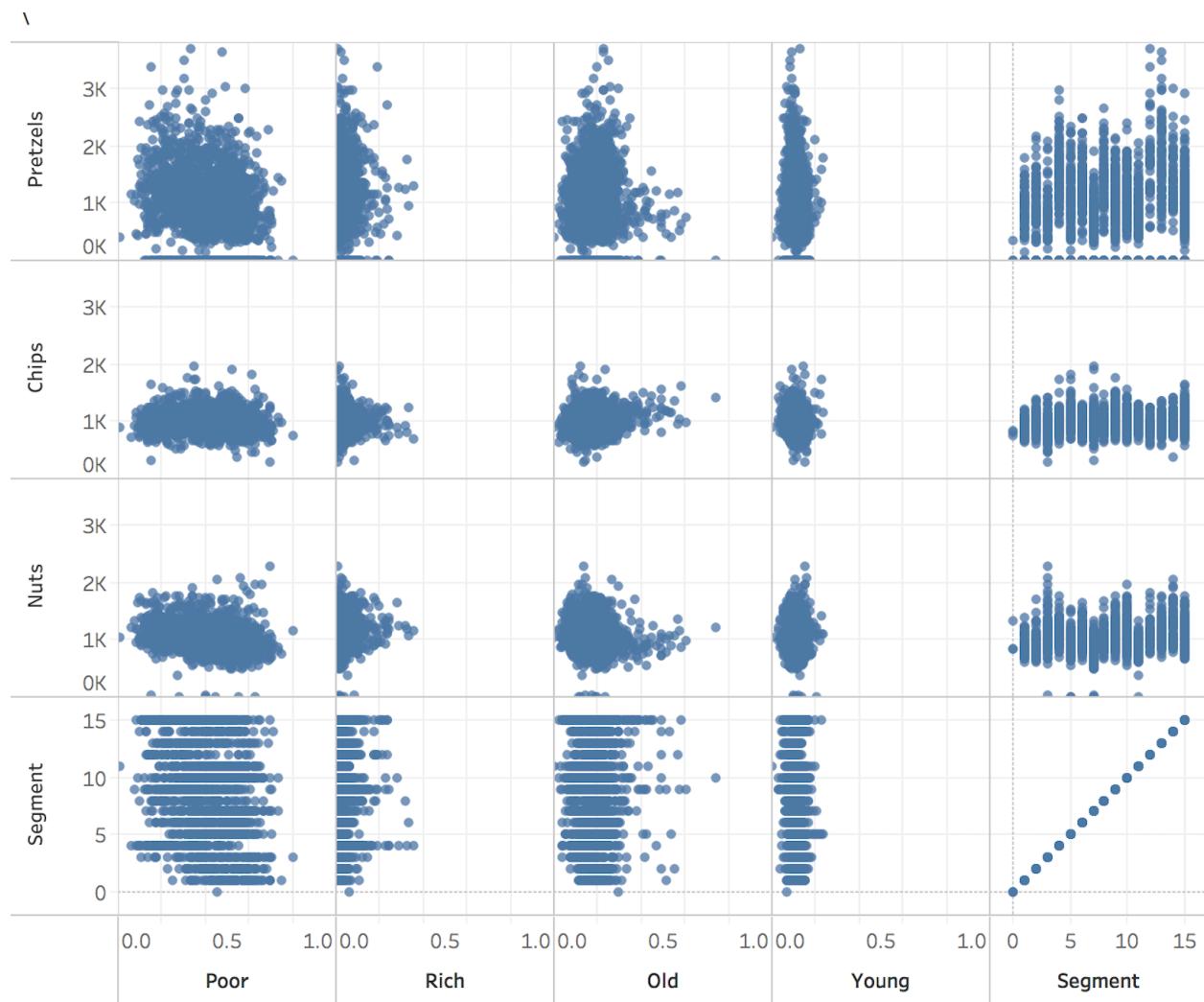


pretzels



The result shows nuts are more popular around southwest, while chips sale better in northwest. Pretzels are typically concentrated on northeast and show a division into two opposing extremes. The result can be an outcome of current resources distribution such as marketing , personnel and transportation if it is uneven. Or it can be potential opportunities, depending on further analysis of marginal profit for these regions.

Step4: Explore the correlation between segment and other possible related attributes



According to the matrix, as the proportion of rich people grows, the sales of all three categories decrease and shrinks towards average significantly.

Overall, I think the segment clustering is a blend of overall sales and location, with some specific regions picked out due to its high sales volume of a certain kind of product.