

Problem Set 6

Lilly Kirby-Rivera

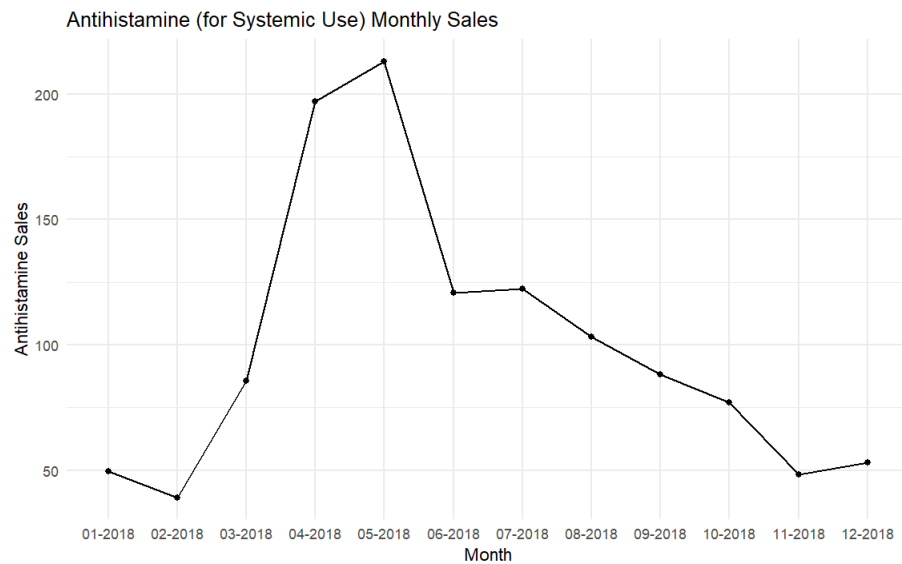
March 12, 2024

Q3

The data set I found from Kaggle on pharmaceutical monthly sales had been mostly cleaned already. They had already treated outliers and missing data, so all I did was change the date format to drop the day, since it should have been just year and month. Additionally, I narrowed the date range to just the 2018 calendar year since it was the most recent complete year in the dataset. This allows us to look at pharmaceutical sales by the type of drug for the year 2018 in real 2018 USD.

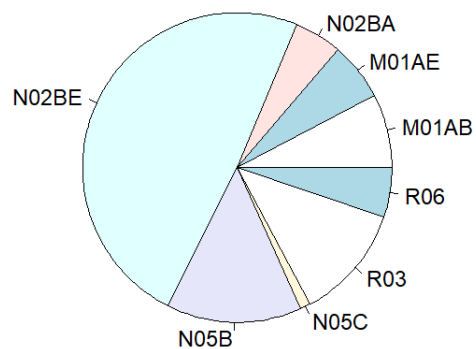
Q4

The first graph I generated was a simple line graph for antihistamine sales in 2018. Looking at the data, sales appear to peak dramatically in April to May. Sales continue to decline for the rest of the year and uptick slightly in December.



The next visual aid I used was a pie chart to see how annual sales could be broken down by product type. In the pie chart, we can see that the product type "N02BE" (other analgesics and antipyretics, pyrazolones and anilides) makes up about half of all pharmaceutical sales in this sample for 2018.

Total 2018 Sales by Product



The last graph I created was a monthly sales comparison of the top two pharmaceutical groups in terms of annual sales. In this bar graph, I compared the pharmaceutical groups N02BE and N05B, which we could see from the previous chart were the top selling product categories in 2018. In this chart, we can see that sales for N02BE pharmaceuticals fluctuate a bit, but remain consistently higher than sales for N05B pharmaceuticals. Meanwhile, sales for N05B pharmaceuticals remain fairly constant throughout the year and remain consistently lower.

