Class_3_Homework

Data Prep

Load the following Libraries and read in ProductInvoices.csv

```
library(tidyverse)
library(stringr)
library(lubridate)

SalesOrderDetail = read_csv(str_c(locationString,"SalesOrderDetail.csv"))
Product = read.csv(str_c(locationString,"Product.csv"))
```

The Product data needs work. The first part of the ProductID is really the ProductGroup, and the last part is the ProductID *(it's common for managers and accountants to combine keys so they can see groups better - NOT a best practice, but a reality you'll have to deal with).*

So, we'll want to use str_sub and str_locate *(don't forget str_locate returns a matrix with 2 columns)*, and str_length, to clean this all up.

We'll also need to covert the new ProductID to a numberic value, as that's what we'll find in the SalesOrderDetail. In the end, it should look like this:

ProductID	Name	Color	StandardCost	ListPrice	ProductGroup
864	Classic Vest, S	Blue	23.7490	63.50	VE
865	Classic Vest, M	Blue	23.7490	63.50	VE
866	Classic Vest, L	Blue	23.7490	63.50	VE
870	Water Bottle - 30 oz.	NULL	1.8663	4.99	WB
921	Mountain Tire Tube	NULL	1.8663	4.99	TT
922	Road Tire Tube	NULL	1.4923	3.99	TT
923	Touring Tire Tube	NULL	1.8663	4.99	TT

You'll then need to inner_join Product with SalesOrderDetail.

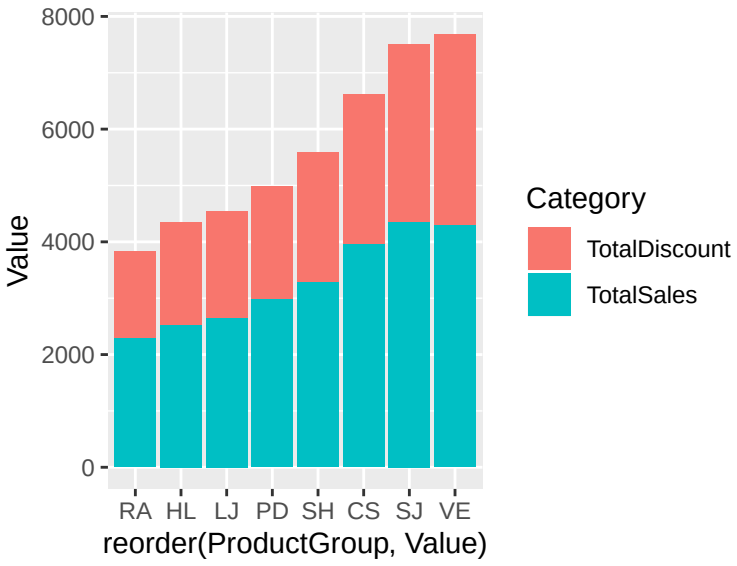
Then compute the Total Discount by multiplying the ListPrice * OrderQty *(which will give you the TotalList)*, then subtract LineTotal.

Then narrow down the columns to ProductGroup, LineTotal and Discount, and group_by ProductGroup, summarising TotalSales *(LineTotals)* and TotalDiscount.

Then filter out Product Groups with Sales over 10,000 and under 2,000. It should now look like this:

ProductGroup	TotalSales	TotalDiscount
CS	3968.868	2645.912
PD	2996.496	1997.664
SH	3299.805	2299.395
SJ	4354.288	3150.322
VE	4309.904	3373.596

Now, you'll have to pivot_longer the data so you can create a stacked bar graph. You'll want to gather up TotalSales and TotalDiscount and put that in "Category". Then, you should be ready to create a plot like the one below:



Now, you'll need to create a tabular analysis so we can more easily see the discounting and compare with TotalList. You'll need to use pivot_wider to spread the Category column and then mutate to create a TotalList column which is just TotalSales + TotalDiscount. When you're done, it should look like this:

ProductGroup	TotalSales	TotalDiscount	TotalList
CS	3968.868	2645.912	6614.78
PD	2996.496	1997.664	4994.16
SH	3299.805	2299.395	5599.20
SJ	4354.288	3150.322	7504.61
VE	4309.904	3373.596	7683.50