DSC640 WEEKS 5 & 6

Cindy Herrera

1/16/2020

3.2 Exercise

You need to submit 3 tree maps, 3 area charts and 3 stacked area charts using Tableau or PowerBI, Python and R using the data below (or your own datasets). You can also submit using D3. You can choose which library to use in Python or R, documentation is provided to help you decide and as you start to play around in the libraries, you will decide which you prefer.

```
library(readxl)
library(ggplot2)
library(dplyr)
```

```
##
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':
##
##     filter, lag
```

```
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
library(plotly)
```

```
##
## Attaching package: 'plotly'
```

```
## The following object is masked from 'package:ggplot2':
##
##     last_plot
```

```
## The following object is masked from 'package:stats':
##
##     filter
```

```
## The following object is masked from 'package:graphics':
##
##     layout
```

```
library(treemap)
library(treemapify)
```

```
df1 <- read.delim("~/Desktop/00 data640/ex3-3/expenditures.txt")
```

View the data structure for Expenditures

```
str(df1)
```

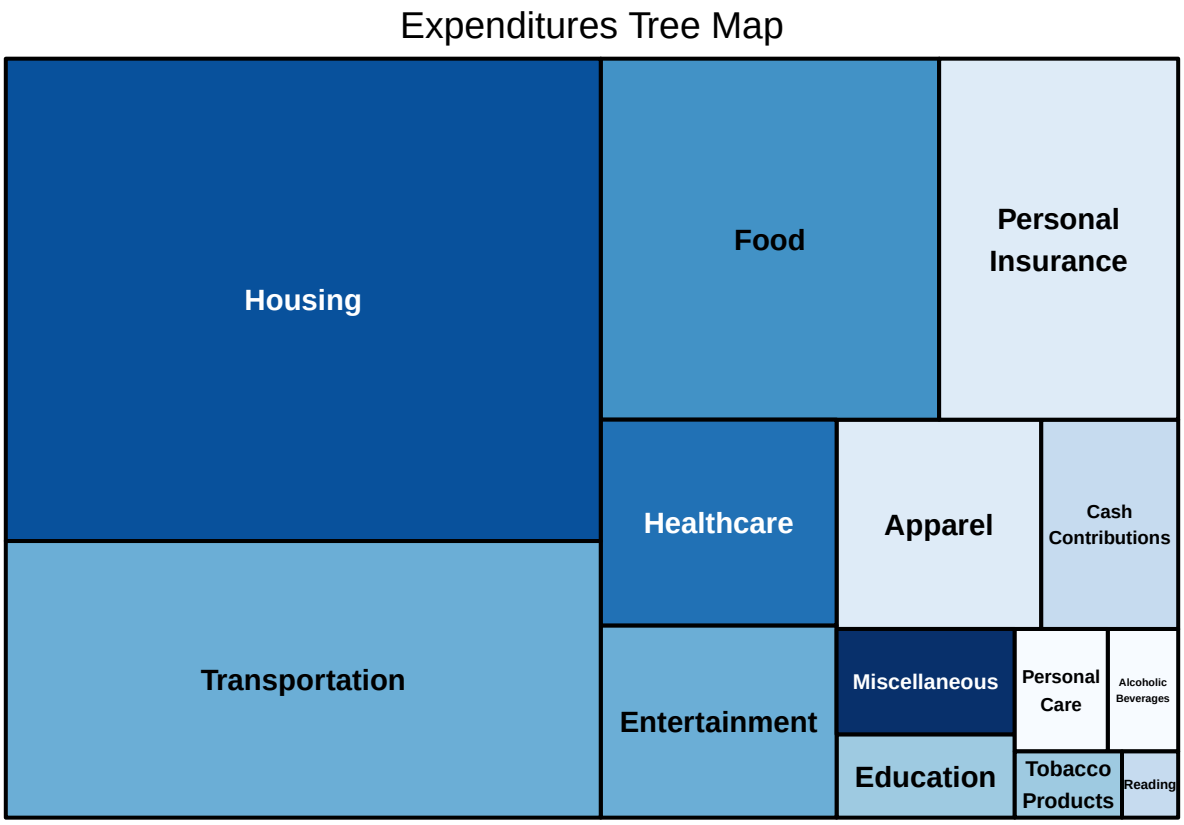
```
## 'data.frame':    350 obs. of  4 variables:
##  $ year       : int  2008 2008 2008 2008 2008 2008 2008 2008 2008 2008 ...
##  $ category   : Factor w/ 14 levels "Alcoholic Beverages",..: 6 1 8 2 14 7 5 10 12 4 ...
```

```
##  $ expenditure: int  6443 444 17109 1801 8604 2976 2835 616 116 1046 ...
##  $ sex        : int  1 1 1 1 1 1 1 1 1 1 ...
```

TREEMAP

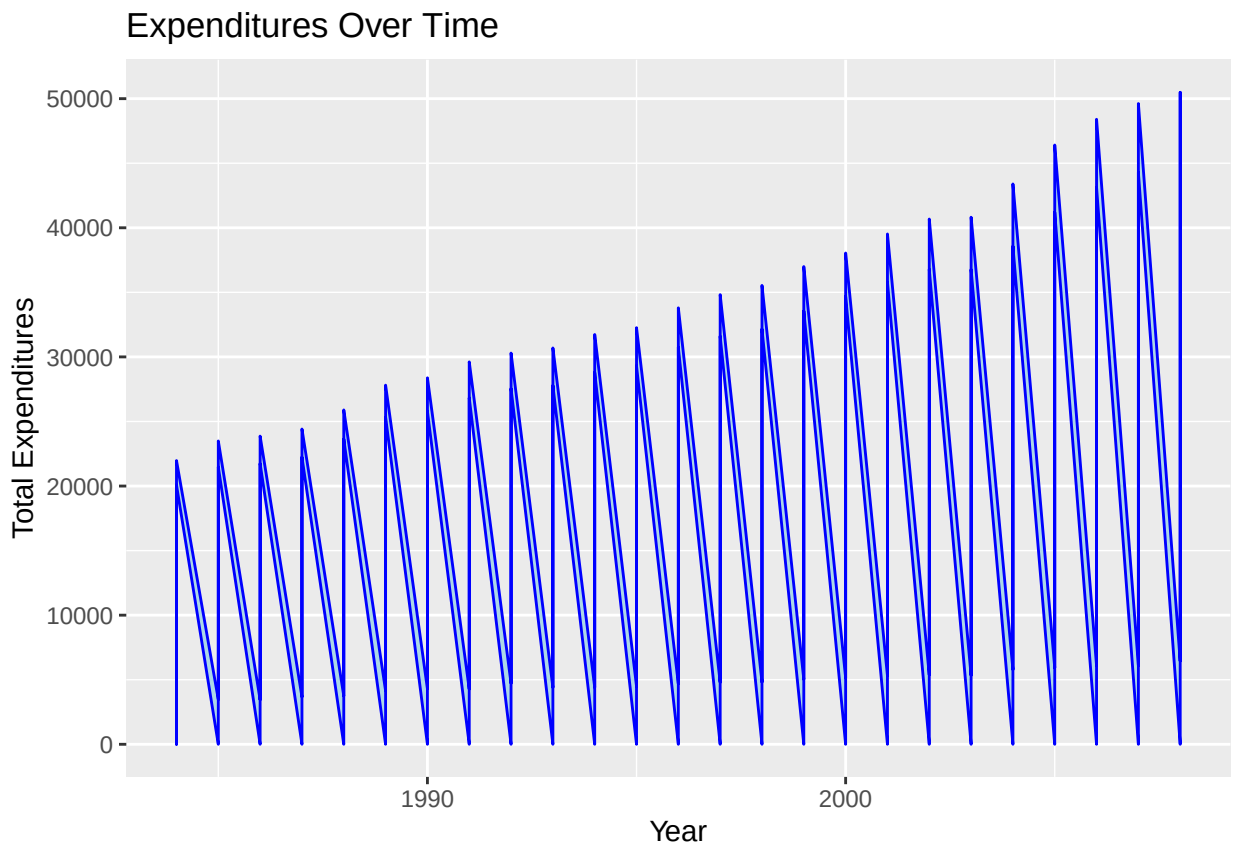
This is great for visualizing categorical data

```
treemap(df1, index = c("category"), vSize = "expenditure", type = "index", palette = "Blues",
        title = "Expenditures Tree Map", fontsize.title = 14)
```



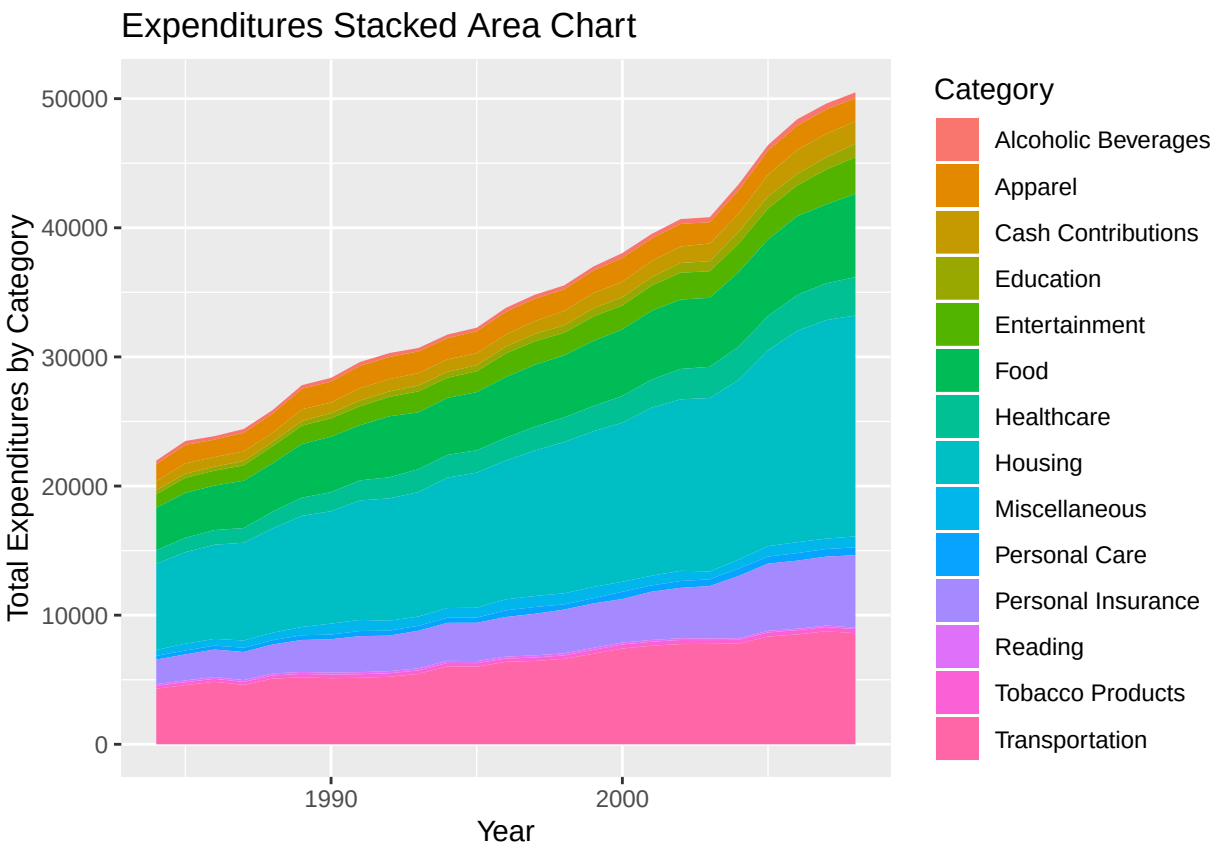
AREA CHART

```
ggplot(df1, aes(x = df1$year, y = df1$expenditure), fill = df1$expenditure) +
  geom_area(fill = 'lightblue', color = "blue") + ggtitle('Expenditures Over Time') +
  labs(x="Year", y="Total Expenditures")
```



STACKED AREA CHART

```
plot <- ggplot(df1, aes(x = df1$year, y = df1$expenditure, fill = df1$category)) +
  geom_area() + ggtitle('Expenditures Stacked Area Chart') +
  labs(x = "Year", y = "Total Expenditures by Category", fill = "Category")
plot
```


Expenditures Stacked Area Chart
Total Expenditures by Category
50000
40000
30000
20000
10000
0
1990
2000
Year
Category
Alcoholic Beverages
Apparel
Cash Contributions
Education
Entertainment
Food
Healthcare
Housing
Miscellaneous
Personal Care
Personal Insurance
Reading
Tobacco Products
Transportation