

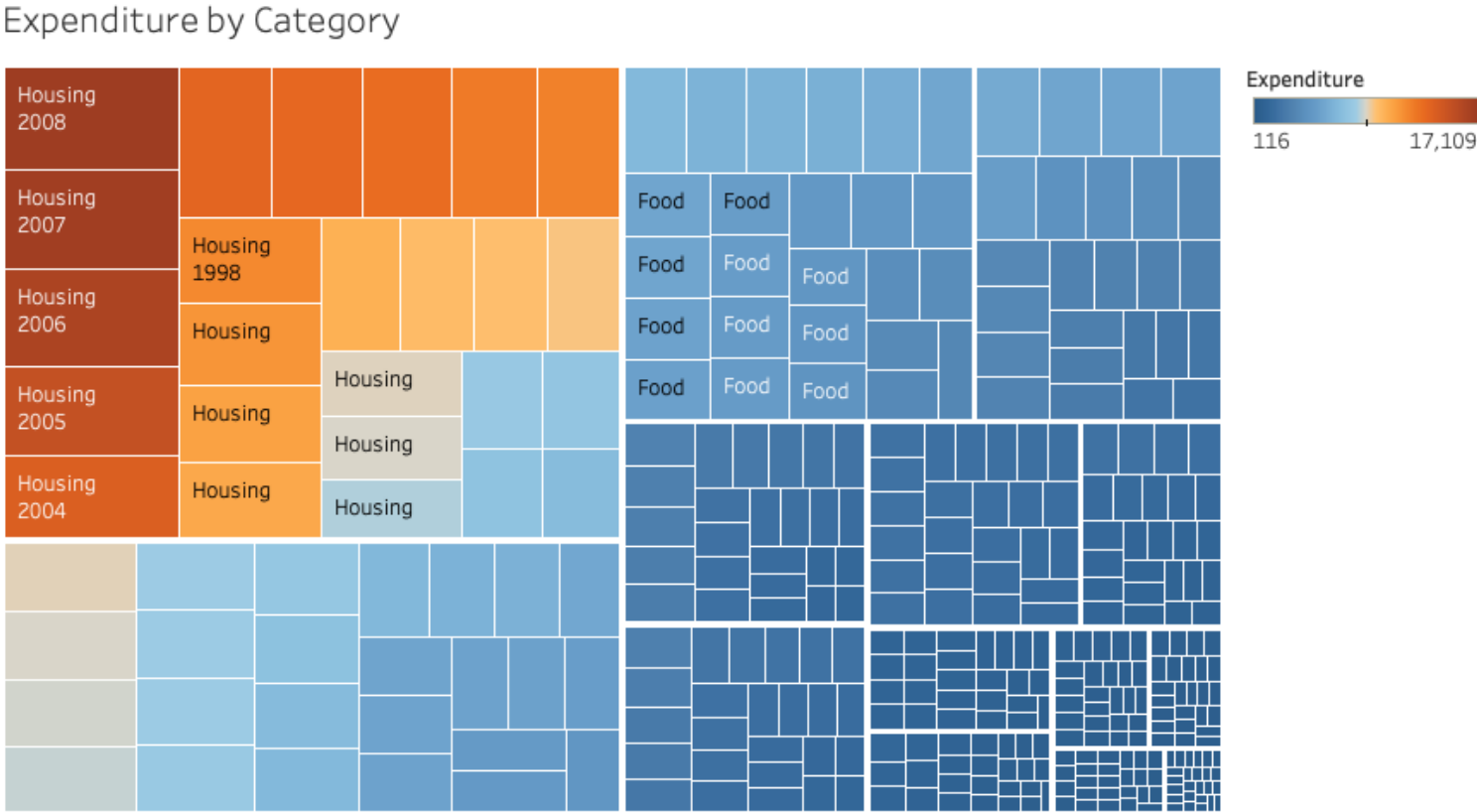
Gillian Tatreau

DSC 640

Assignment 3.2

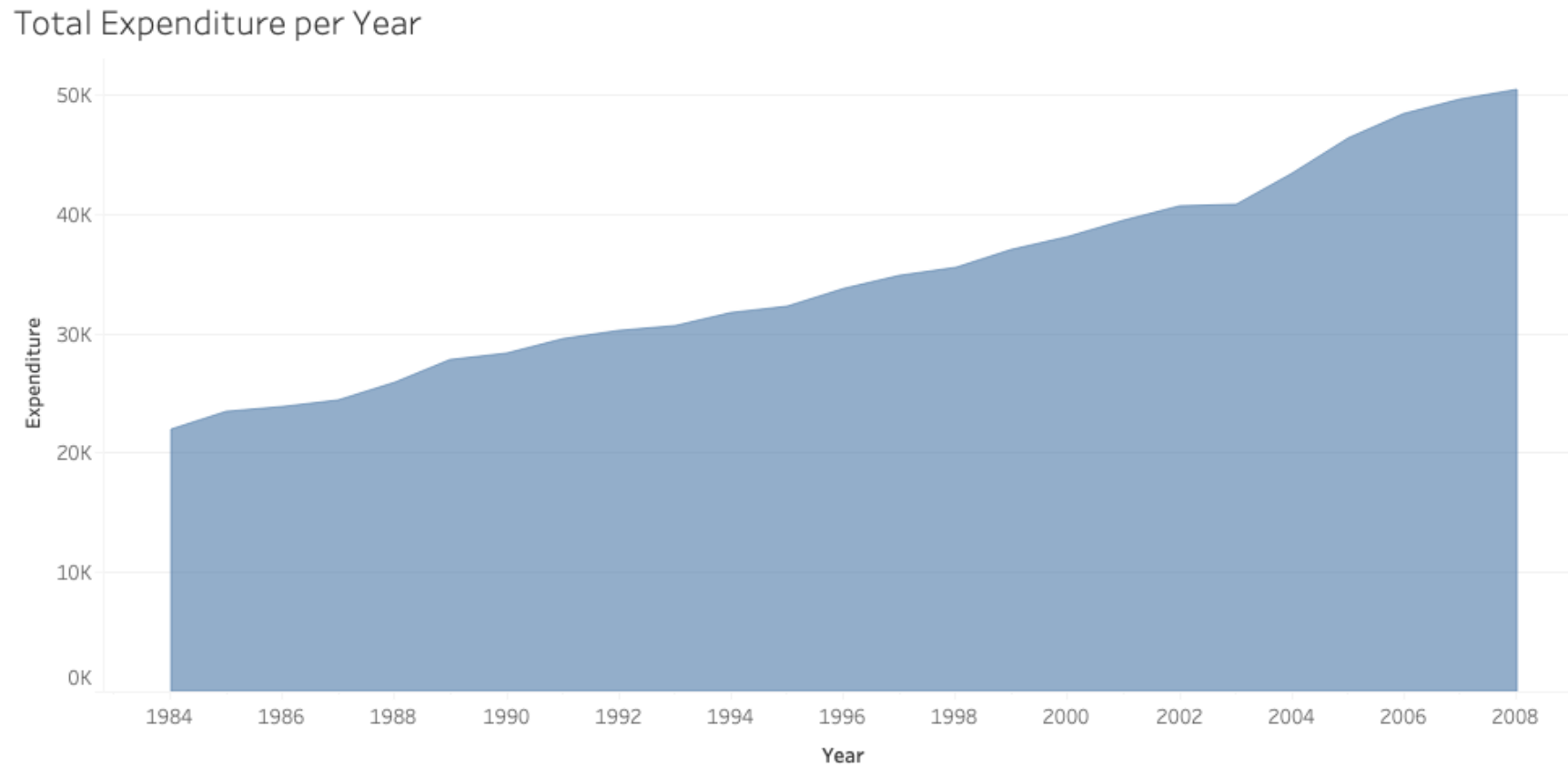
8 October 2023

Tableau- Tree Map Chart



Category and Year. Color shows sum of Expenditure. Size shows sum of Expenditure. The marks are labeled by Category and Year. The view is filtered on sum of Expenditure, which includes everything.

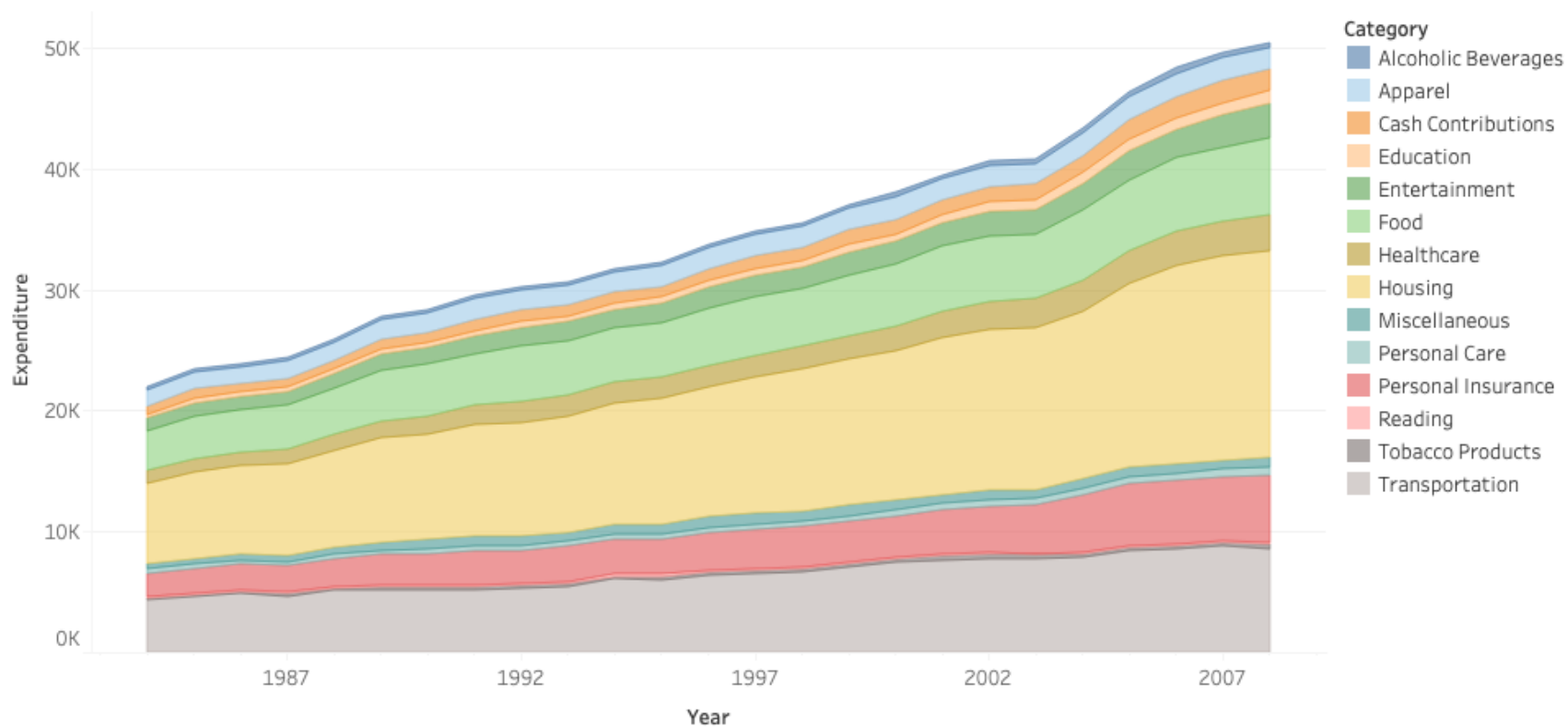
Tableau- Area Chart



The plot of sum of Expenditure for Year (copy) Year.

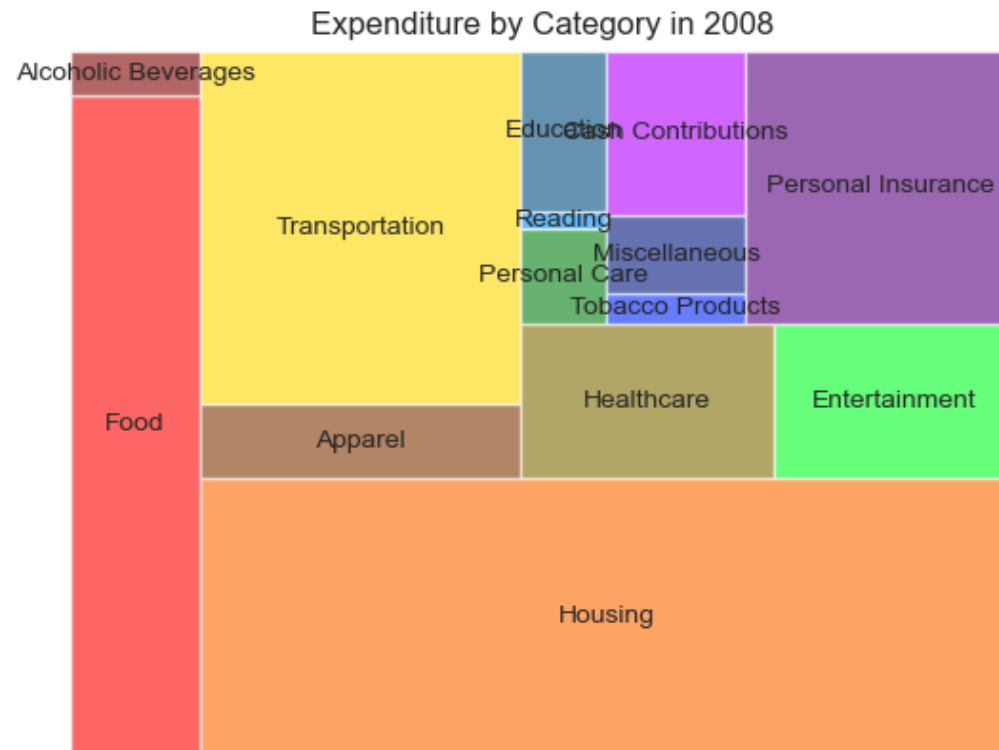
Tableau- Stacked Area Chart

Expenditure by Cateogry, from 1984-2008

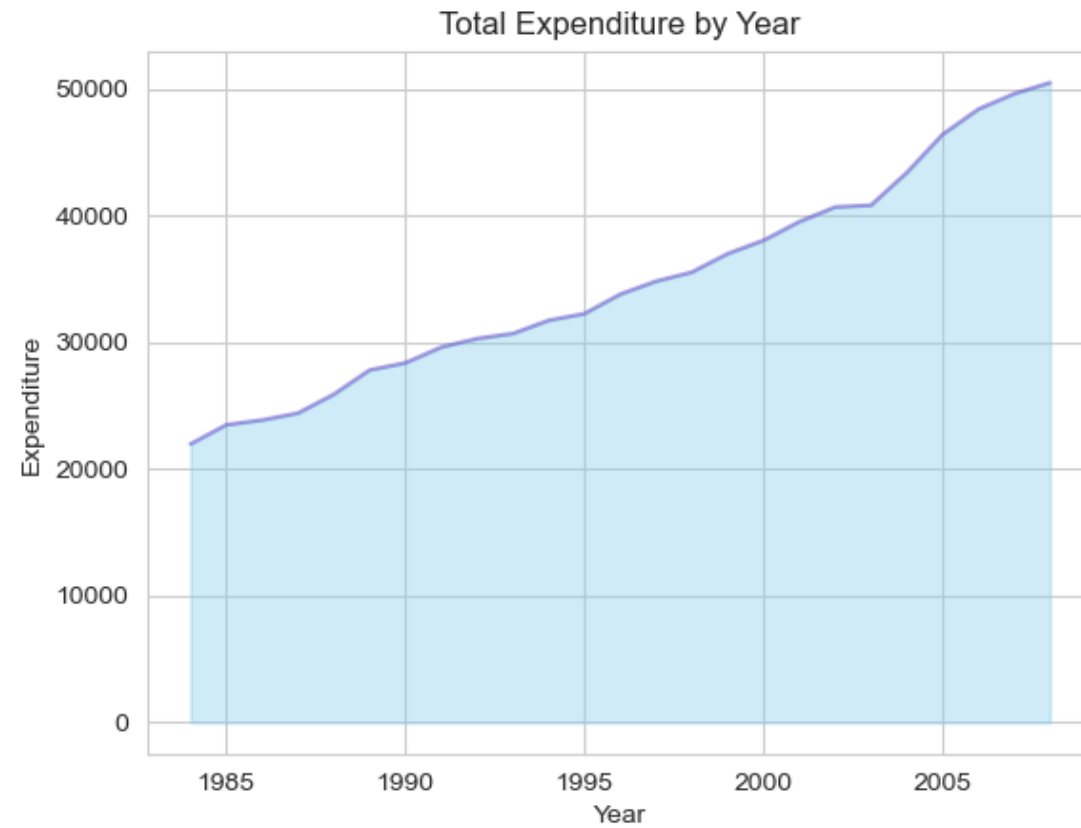


The plot of sum of Expenditure for Year (copy) Year. Color shows details about Category.

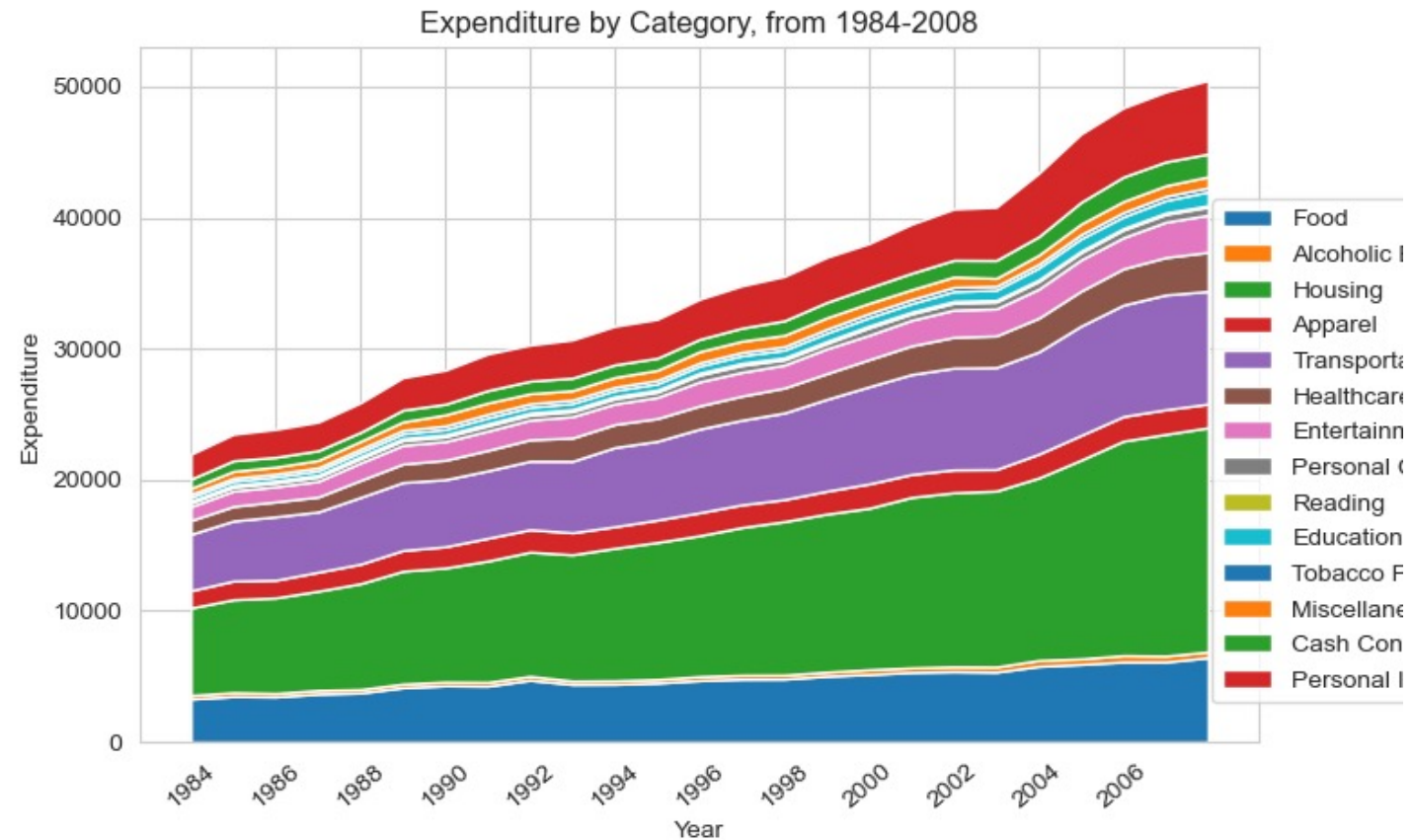
Python- Tree Map Chart



Python- Area Chart

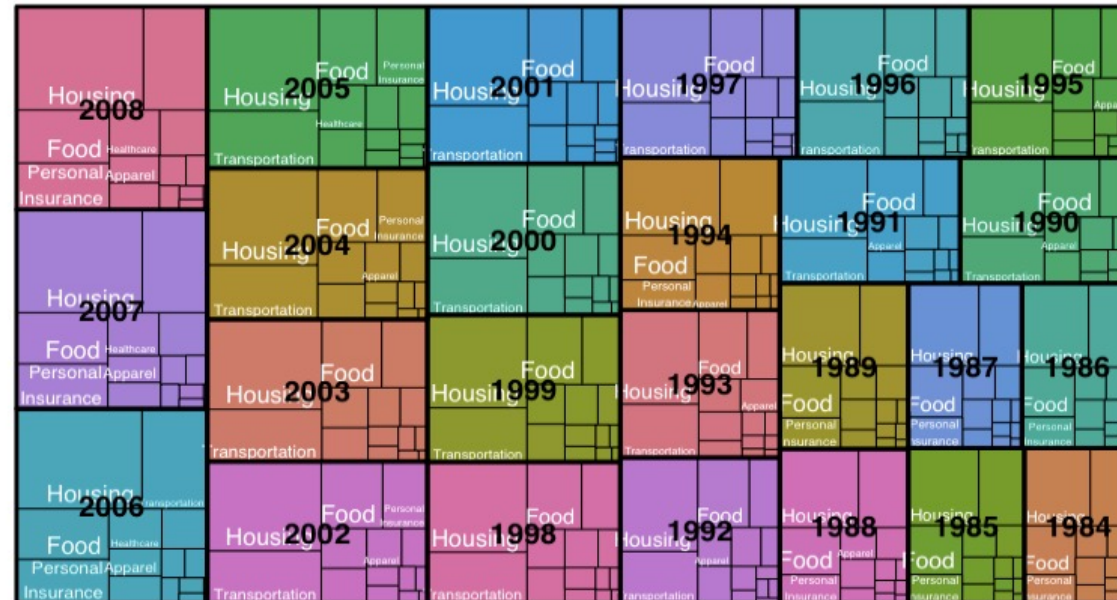


Python- Stacked Area Chart

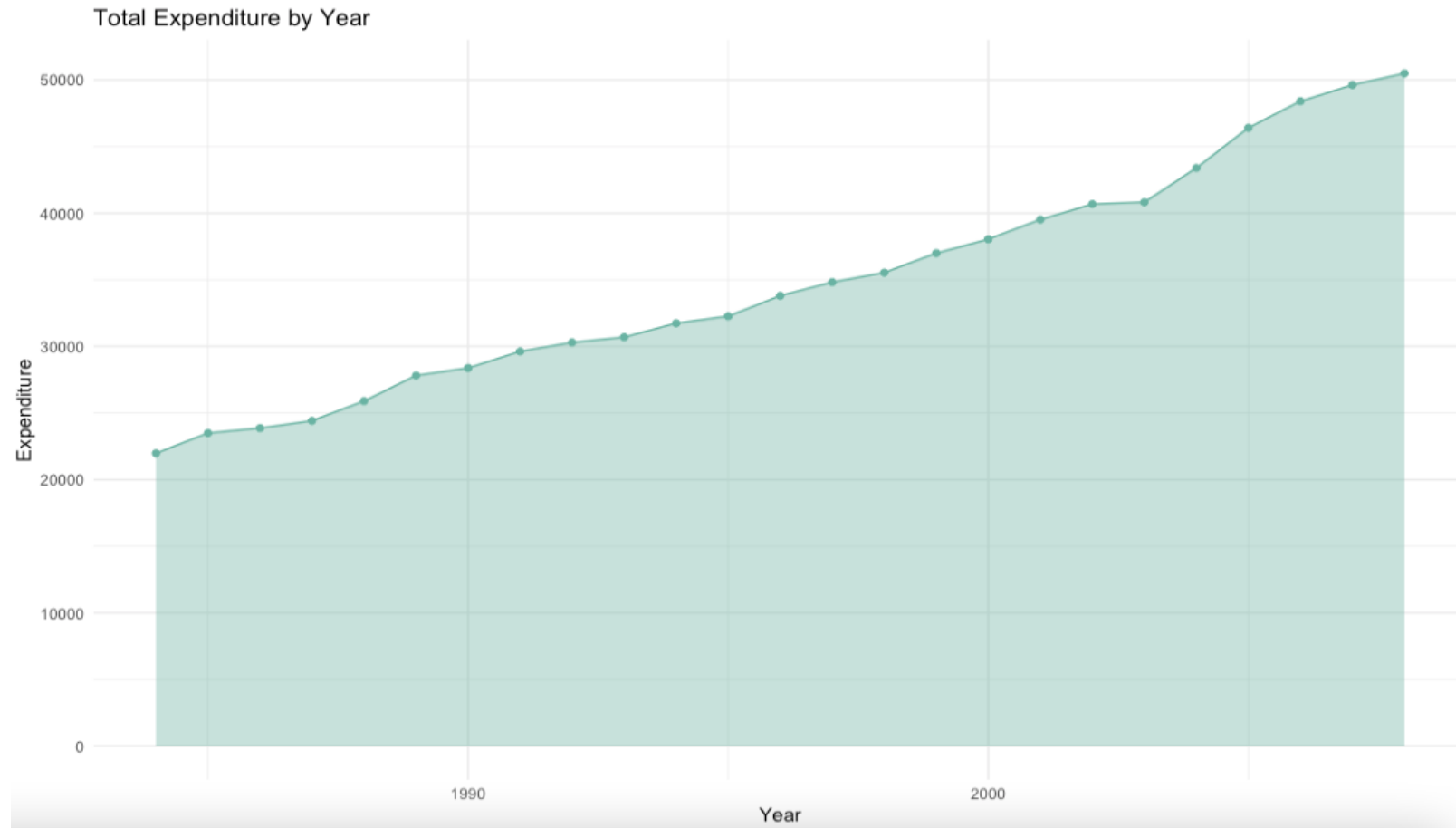


R- Tree Map Chart

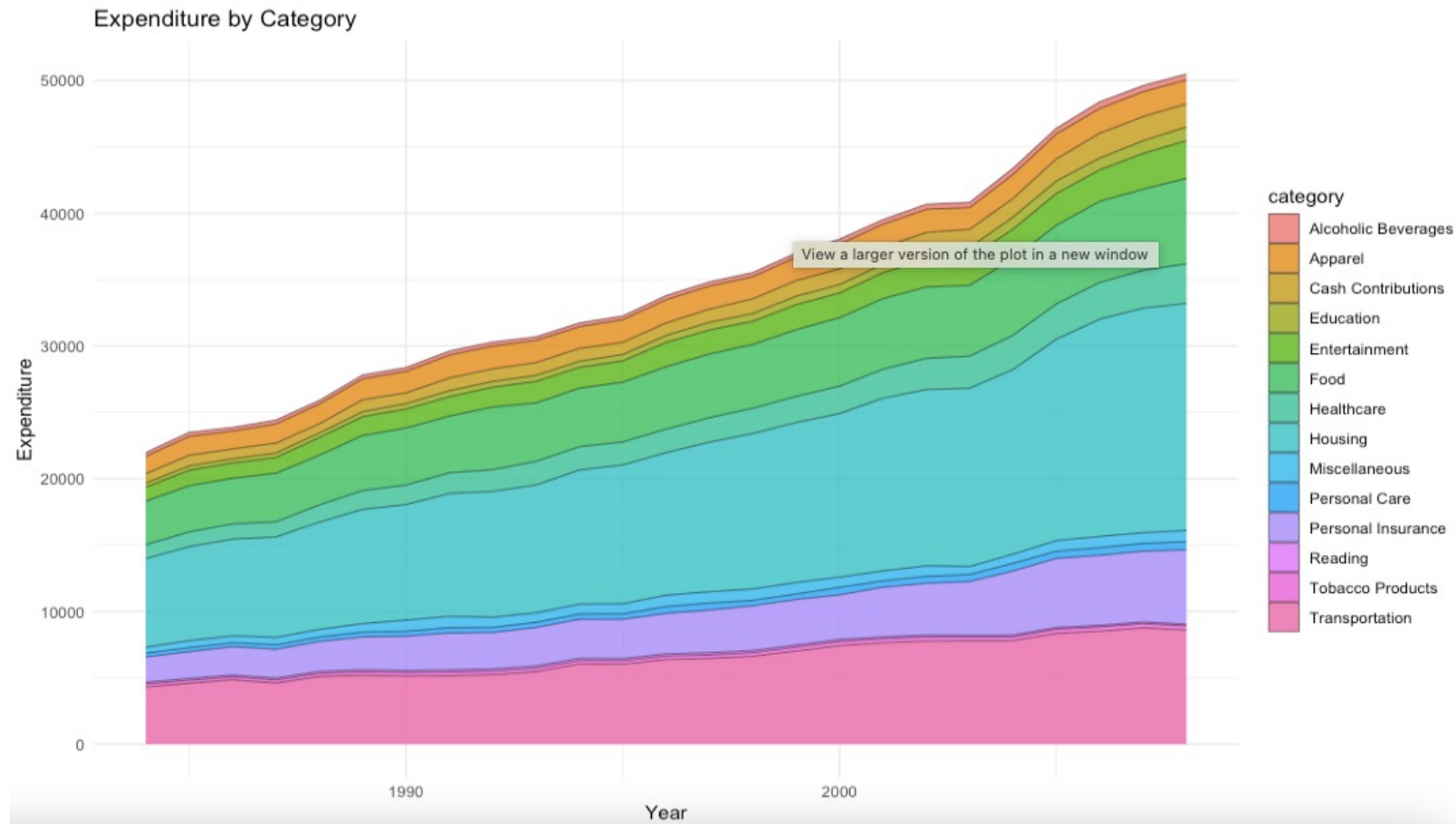
Expenditure by Year



R- Area Chart



R – Stacked Area Chart



Gillian Tatreau

DSC640

Weeks 5 and 6

8 October 2023

Python Code

```
In [1]: # load libraries
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
import squarify
```

```
In [2]: # load data
df = pd.read_table("/Users/gillian/Documents/Bellevue Grad Program/Fall 2023/DSC640/Weeks5%266/DSC640_TatreauGillian_3.2PythonCode.ipynb?download=false")
df.head()
```

```
Out[2]:
```

	year	category	expenditure	sex
0	2008	Food	6443	1
1	2008	Alcoholic Beverages	444	1
2	2008	Housing	17109	1
3	2008	Apparel	1801	1
4	2008	Transportation	8604	1

Treemap chart

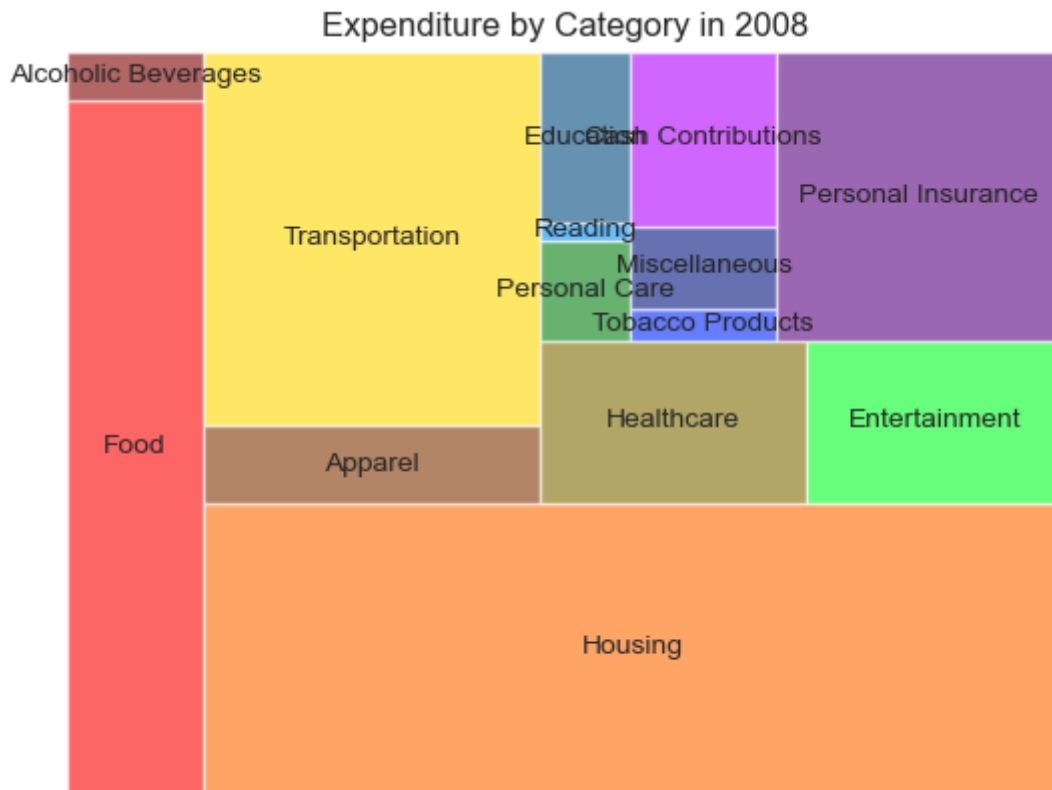
```
In [3]: # install squarify
# pip install squarify
```

```
In [4]: df2 = df.loc[df['year'] == 2008]
```

```
In [5]: colors = ['#fe0000', '#800001', '#fe6a00', '#803400', '#ffd800',
                 '#806b00', '#00fe21', '#007f0e', '#0094fe', '#00497e',
                 '#0026ff', '#001280', '#b100fe', '#590080']
```

```
In [6]: sns.set_style(style="whitegrid")
sizes = df2["expenditure"].values
label = df2["category"]
```

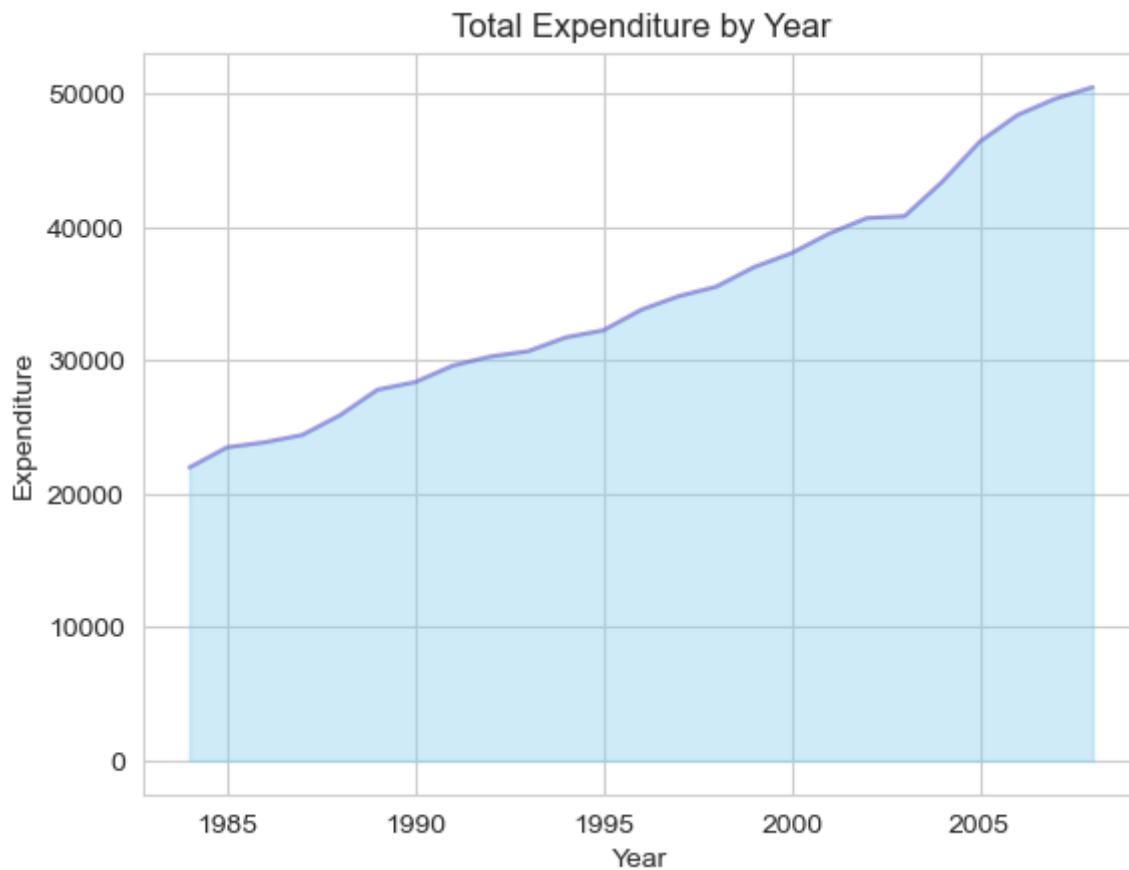
```
squarify.plot(sizes=sizes, label=label, alpha=0.6, color = colors).set(title="Expenditure by Category in 2008")
plt.axis('off')
plt.savefig('Python-TreemapChart.png')
plt.show()
```



Area Chart

```
In [7]: df3 = df.groupby('year')['expenditure'].sum().reset_index(name='total')
```

```
In [8]: plt.fill_between(df3['year'], df3['total'], color="skyblue", alpha=0.4)
plt.plot(df3['year'], df3['total'], color="Slateblue", alpha=0.6)
plt.title("Total Expenditure by Year")
plt.xlabel("Year")
plt.ylabel("Expenditure")
plt.savefig('Python-AreaChart.png')
plt.show()
```



Stacked Area Chart

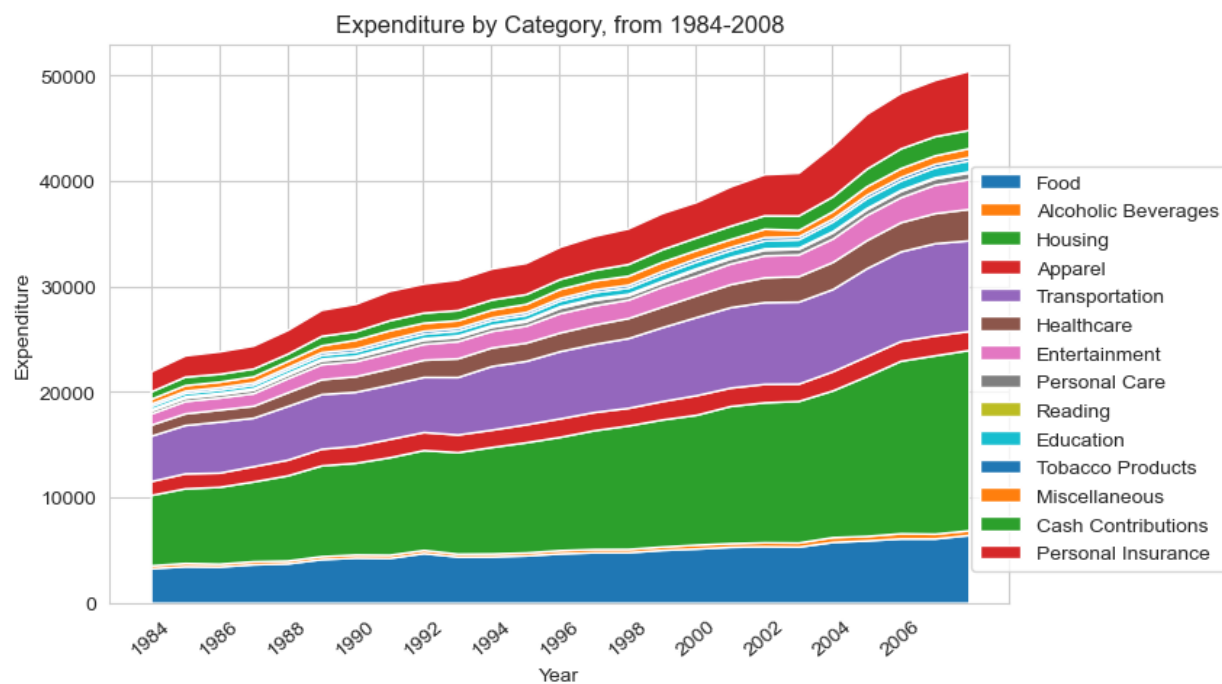
```
In [9]: year = np.unique(df['year'].values)
        year2 = np.flip(year)
```

```
In [10]: labels = df['category'].unique().tolist()
```

```
In [11]: food = (df[df['category']=="Food"]['expenditure'])
        alc = (df[df['category']=="Alcoholic Beverages"]['expenditure'])
        house = (df[df['category']=="Housing"]['expenditure'])
        apparel = (df[df['category']=="Apparel"]['expenditure'])
        transp = (df[df['category']=="Transportation"]['expenditure'])
        health = (df[df['category']=="Healthcare"]['expenditure'])
        ent = (df[df['category']=="Entertainment"]['expenditure'])
        pers = (df[df['category']=="Personal Care"]['expenditure'])
        read = (df[df['category']=="Reading"]['expenditure'])
        edu = (df[df['category']=="Education"]['expenditure'])
        tob = (df[df['category']=="Tobacco Products"]['expenditure'])
        misc = (df[df['category']=="Miscellaneous"]['expenditure'])
        cash = (df[df['category']=="Cash Contributions"]['expenditure'])
        insure = (df[df['category']=="Personal Insurance"]['expenditure'])
```

```
In [12]: fig = plt.figure(figsize=(8,5))
        plt.stackplot(year2, food, alc, house, apparel, transp, health, ent, pers, read, edu, tob, misc, cash, insure)
        plt.legend(loc='upper center', bbox_to_anchor=(1.1, 0.8), shadow=False, ncol=1)
        plt.xticks(np.arange(1984,2008,step=2), rotation=40)
        plt.title("Expenditure by Category, from 1984-2008")
        plt.xlabel("Year")
        plt.ylabel("Expenditure")
```

```
plt.savefig('Python-StackedAreaChart.png')  
plt.show()
```



```

# R Code

# install treemap
# install.packages("treemap")

# load packages
library(treemap)
library(dplyr)
library(ggplot2)

# load data
df <- read.delim("/Users/gillian/Documents/Bellevue Grad Program/Fall
2023/DSC640/Weeks5&6/expenditures.txt")

# treemap
t <- treemap(df, index=c("year","category"),
             vSize="expenditure", type="index",
             fontsize.labels=c(15,12),
             fontface.labels=c(2,1),
             fontcolor.labels=c("black","white"),
             bg.labels=c("transparent"),           # Background color of labels
             align.labels=list(
               c("center", "center"),
               c("right", "bottom")),
             title = 'Expenditure by Year', fontsize.title=12
)
t

# area chart
# Group year by sum of values
agg_tbl <- df %>% group_by(year) %>%
  summarise(spend_sum = sum(expenditure),
            .groups = 'drop')
# Convert tibble to df
df2 <- agg_tbl %>% as.data.frame()

a <- ggplot(df2, aes(x=year, y=spend_sum)) +
  geom_area( fill="#69b3a2", alpha=0.4) +
  geom_line(color="#69b3a2") +
  geom_point(color="#69b3a2") +
  ggtitle("Total Expenditure by Year") +
  xlab('Year') + ylab('Expenditure') +
  theme_minimal()
a

# stacked area chart
s <- ggplot(df, aes(x=year, y=expenditure, fill=category)) +
  geom_area(colour="black", size=.2, alpha=.8) +
  ggtitle("Expenditure by Category") +
  xlab('Year') + ylab('Expenditure') +
  theme_minimal()
s

```