

Assignment_3.2_Vayuvegula_Soma_Shekar_Python

```
In [1]: # Import Libraries
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import squarify
```

```
In [2]: #Read csv
df_unemp = pd.read_csv("unemployment-rate-1948-2010.csv")
df_unemp.head()
```

Out[2]:

	Series id	Year	Period	Value
0	LNS14000000	1948	M01	3.4
1	LNS14000000	1948	M02	3.8
2	LNS14000000	1948	M03	4.0
3	LNS14000000	1948	M04	3.9
4	LNS14000000	1948	M05	3.5

```
In [3]: unemp_df=df_unemp.groupby('Year')['Value'].sum()
unemp_df=unemp_df.to_frame().reset_index()
unemp_df.head()
```

Out[3]:

	Year	Value
0	1948	45.0
1	1949	72.6
2	1950	62.5
3	1951	39.4
4	1952	36.3

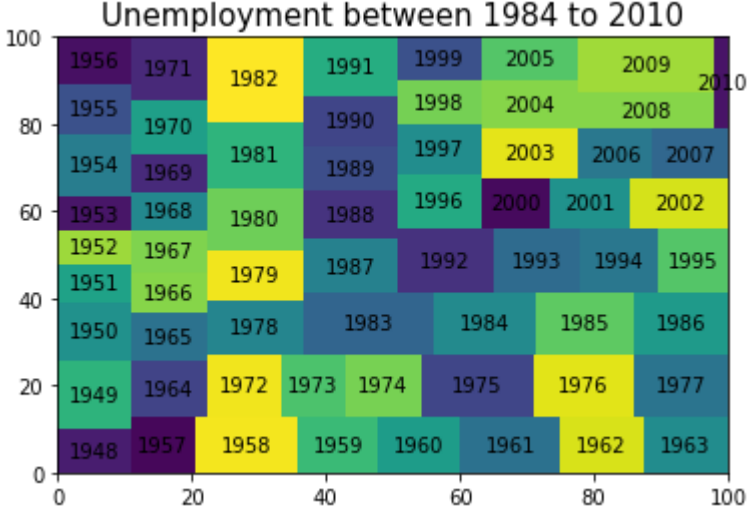
```
In [4]: #Read text file
df_exp = pd.read_csv("expenditures.txt",sep='\t',lineterminator='\r')
df_exp.head()
```

Out[4]:

	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

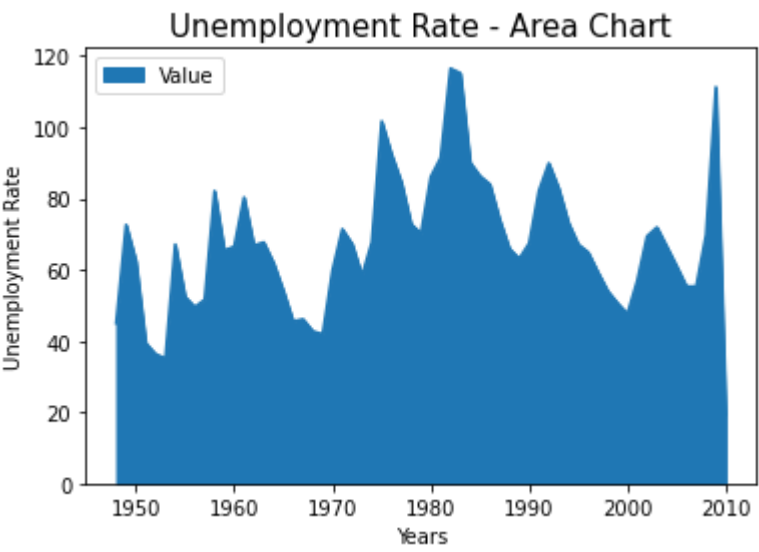
Tree Map

```
In [5]: squarify.plot(sizes=round(unemp_df['Value']), label=unemp_df['Year'].astype(str), alpha=1 )
plt.title("Unemployment between 1984 to 2010", size=15)
plt.savefig("Python_Tree_Map.png")
plt.show()
plt.close()
```



Area Chart

```
In [6]: unemp_df.plot.area(x="Year",y="Value")
plt.ylabel("Unemployment Rate")
plt.xlabel("Years")
plt.title("Unemployment Rate - Area Chart",size=15)
plt.savefig("Python_Area_chart.png")
plt.show()
plt.close()
```



Stacked Area Chart

```
In [7]: #Pivoting expenditure df
df_exp_pivot=pd.pivot_table(df_exp,values='expenditure',index=['year'],columns='category').reset_index()
index=df_exp_pivot['year']
df_exp_pivot.head()
```

Out[7]:

	category	year	Alcoholic Beverages	Apparel	Cash Contributions	Education	Entertainment	Food	Healthcare	Housing	Miscellaneous	Personal Care	Personal Insurance
	0	1984	275	1319	706	303	1055	3290	1049	6674	451	289	
	1	1985	306	1420	805	321	1170	3477	1108	7087	529	303	
	2	1986	271	1346	746	314	1149	3448	1135	7292	522	303	
	3	1987	289	1446	741	337	1193	3664	1135	7569	562	330	
	4	1988	269	1489	693	342	1329	3748	1298	8079	578	334	

```
In [8]: data_columns = list(df_exp_pivot.columns)
data_columns.remove('year')
data_columns
```

Out[8]:

```
['Alcoholic Beverages',
 'Apparel',
 'Cash Contributions',
 'Education',
 'Entertainment',
 'Food',
 'Healthcare',
 'Housing',
 'Miscellaneous',
 'Personal Care',
 'Personal Insurance',
 'Reading',
 'Tobacco Products',
 'Transportation']
```

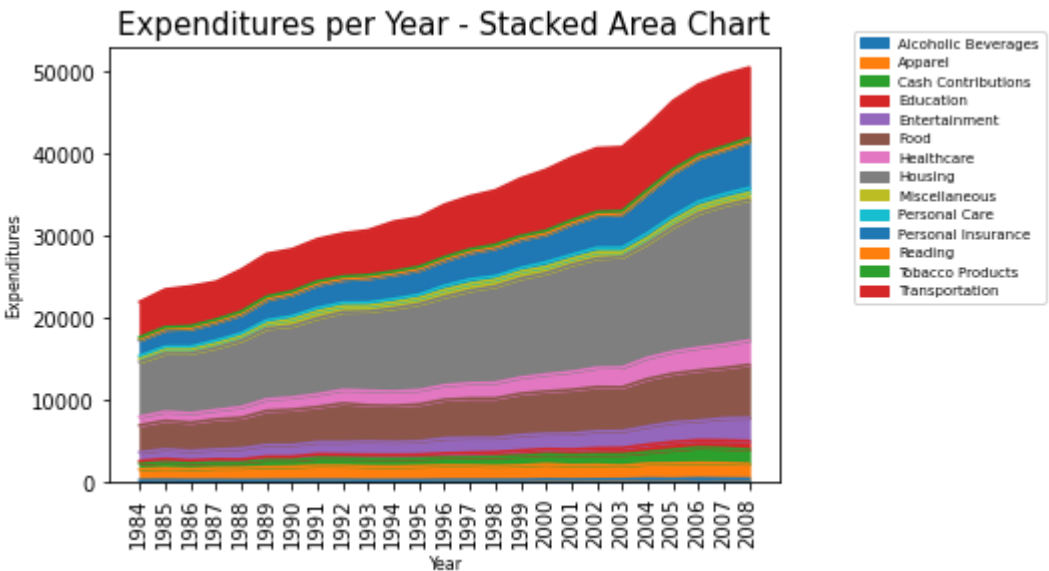
```
In [9]: data_temp = df_exp_pivot[data_columns].copy()

rec = data_temp.to_records(index=False)
data = list(rec)

temp_df = pd.DataFrame(data_temp, columns=data_columns)

col = list(data_temp.columns.values)
```

```
In [10]: pos=(0,1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,17,18,19,20,21,22,23,24)
axs=temp_df.plot.area(stacked=True,xticks=pos,rot=90)
axs.set_xticklabels(index)
plt.ylabel("Expenditures",size=8)
plt.xlabel("Year",size=8)
plt.title("Expenditures per Year - Stacked Area Chart",size=15)
axs.legend(bbox_to_anchor=(1.1,1.05),prop={"size":7})
plt.savefig("Python_Stacked_Area.png")
plt.show()
plt.close()
```



```
In [ ]:
```

Assignment_5_6_Vayuvegula_Soma_Shekar_R

Soma Shekar Vayuvegula

01/21/2023

```
##
## 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

## -- Attaching packages ----- tidyverse 1.3.2 --
## v tibble  3.1.7      v purrr  0.3.4
## v tidyr   1.2.0      v stringr 1.4.0
## v readr   2.1.2      v forcats 0.5.2
## -- Conflicts ----- tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
##
## Attaching package: 'reshape2'
##
##
## The following object is masked from 'package:tidyr':
##
##   smiths
##
##
## Attaching package: 'data.table'
##
##
## The following objects are masked from 'package:reshape2':
##
##   dcast, melt
##
##
## The following object is masked from 'package:purrr':
##
##   transpose
##
##
## The following objects are masked from 'package:dplyr':
##
##   between, first, last
```

```
##
##
##
## 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

df_unemp<-read.csv("unemployment-rate-1948-2010.csv")
head(df_unemp,5)

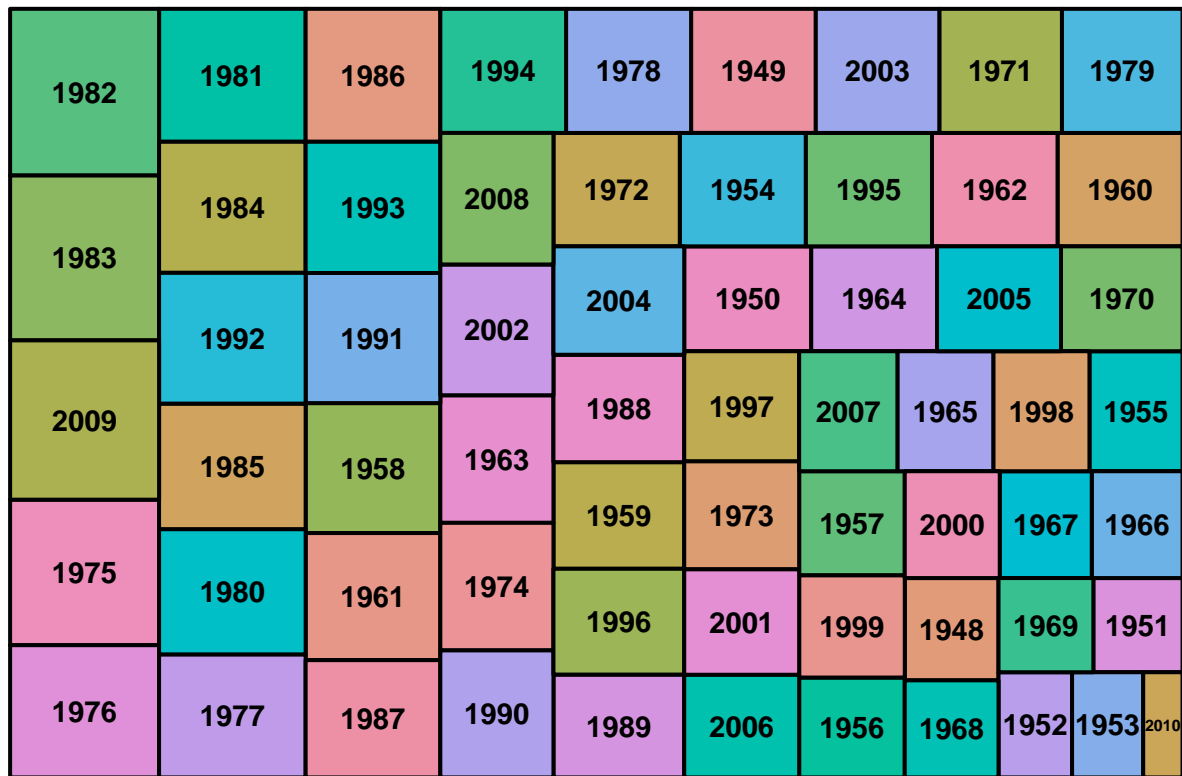
##      Series.id Year Period Value
## 1 LNS14000000 1948   M01   3.4
## 2 LNS14000000 1948   M02   3.8
## 3 LNS14000000 1948   M03   4.0
## 4 LNS14000000 1948   M04   3.9
## 5 LNS14000000 1948   M05   3.5

df_exp<-read.table("expenditures.txt",sep='\t',header=TRUE)
head(df_exp,5)

##   year      category expenditure sex
## 1 2008          Food          6443   1
## 2 2008 Alcoholic Beverages          444   1
## 3 2008          Housing         17109   1
## 4 2008          Apparel          1801   1
## 5 2008 Transportation          8604   1

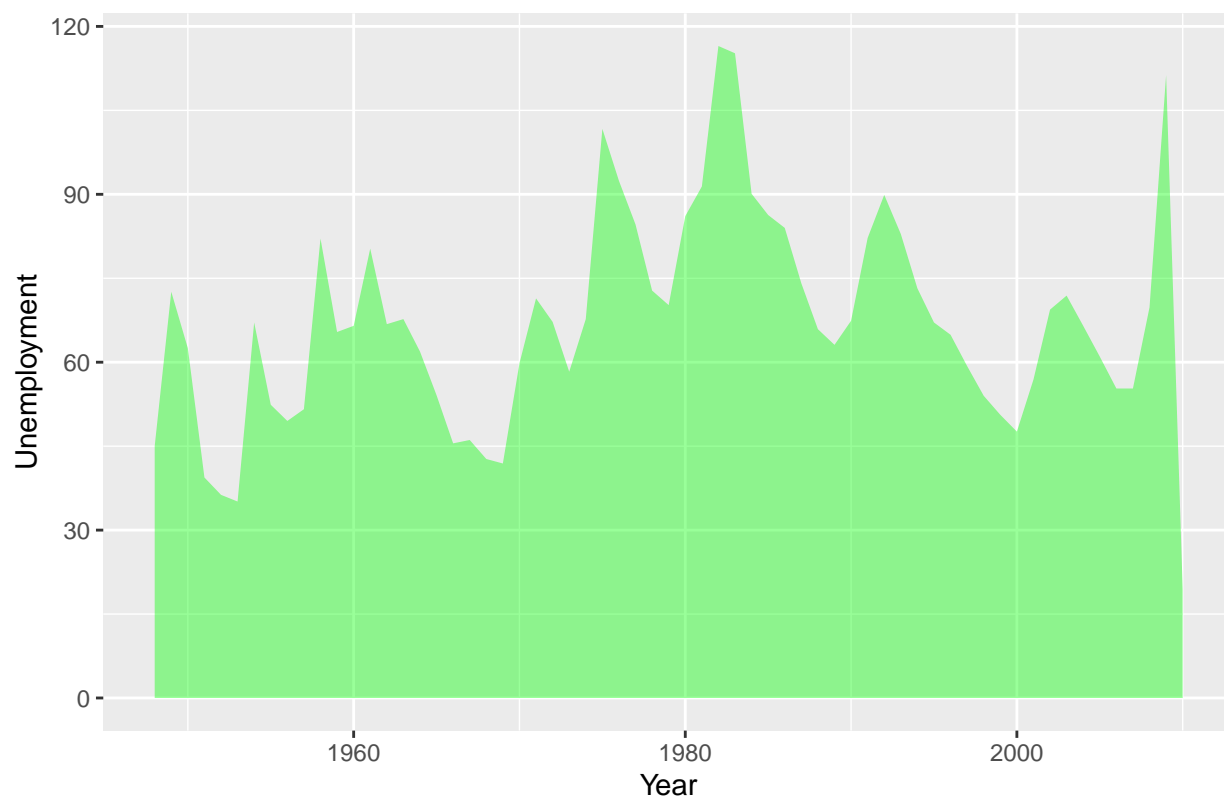
exp_agg_df<-aggregate(df_unemp$Value,by=list(Year=df_unemp$Year),FUN=sum)
group<-exp_agg_df$Year
value<-exp_agg_df$x
df_exp_agg<-data.frame(group,value)
treemap(df_exp_agg,index="group",vSize="value",type="index",title="Unemployment between 1948-2010",font
```

Unemployment between 1948–2010



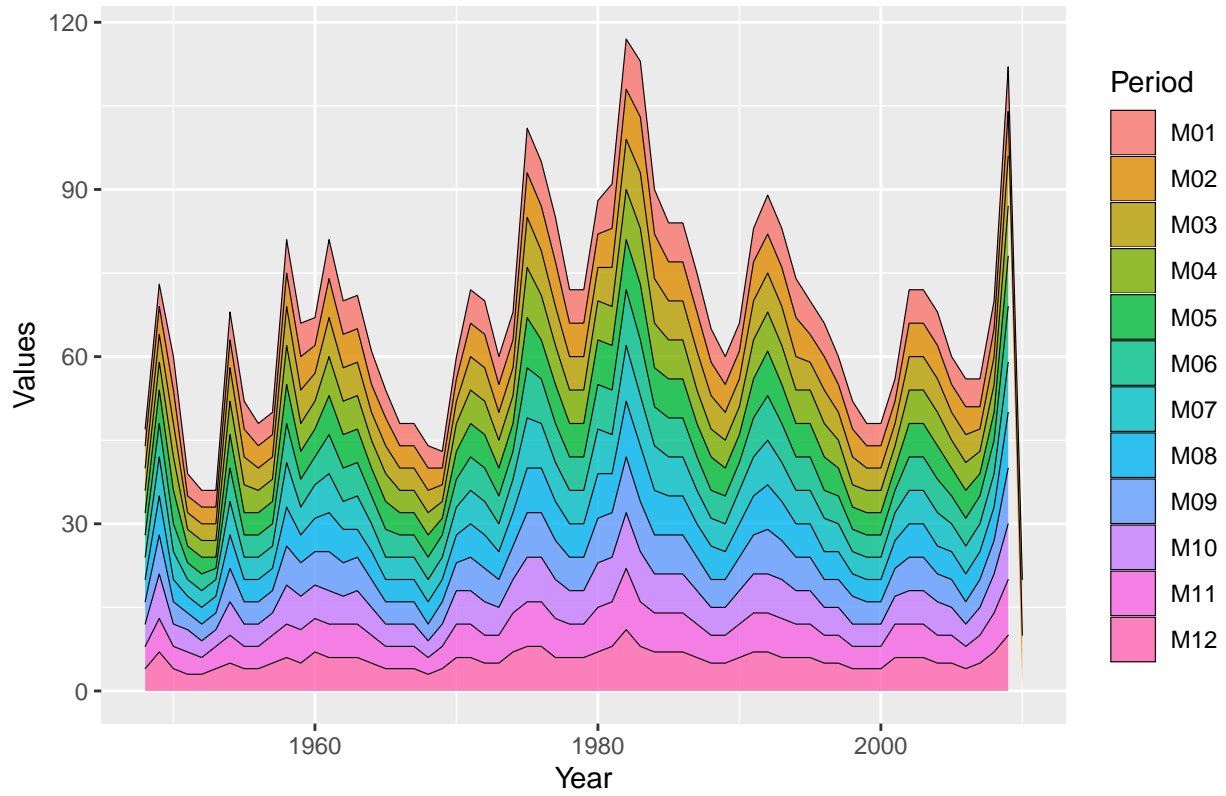
```
ggplot(df_exp_agg,aes(x=group,y=value))+geom_area(fill="green",alpha=0.4)+
  labs(x="Year",y="Unemployment")+ggtitle("Expenditures per year - Area Map")
```

Expenditures per year – Area Map



```
Values<-round(df_unemp$Value,0)
df1<-data.frame(Year=df_unemp$Year,Period=df_unemp$Period,Values=round(df_unemp$Value,0))
ggplot(df1,aes(x=Year,y=Values,fill=Period))+geom_area(colour="black",size=0.2,alpha=0.8)+ggtitle("Unemp
```

Unemployment Per Year – Stacked Area Chart



Assignment_5_6_Vayuvegula _Soma_Shekar_Tableau

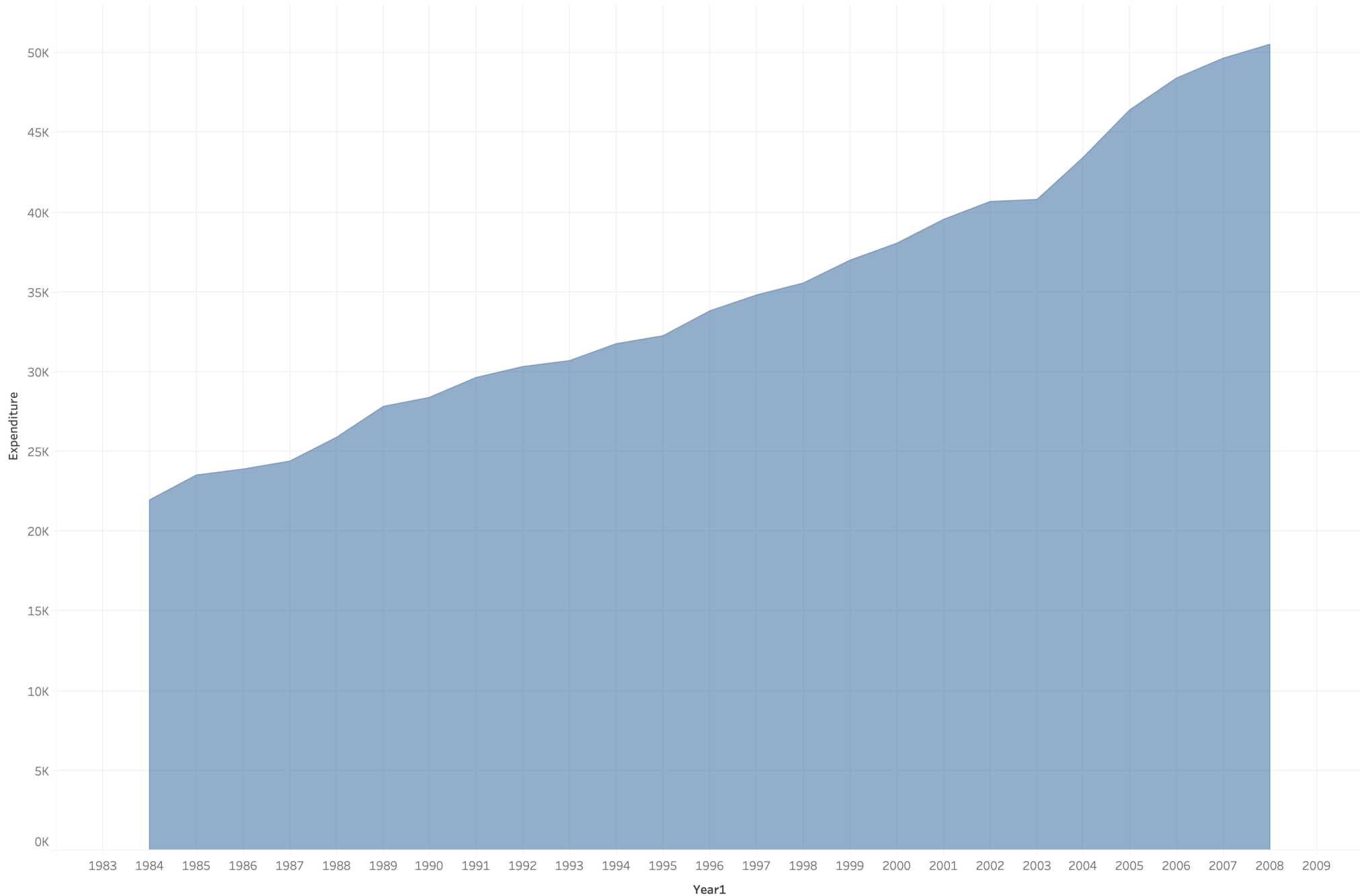
File created on: 1/22/23 8:53:36 AM CST

Expenditures - Tree Map



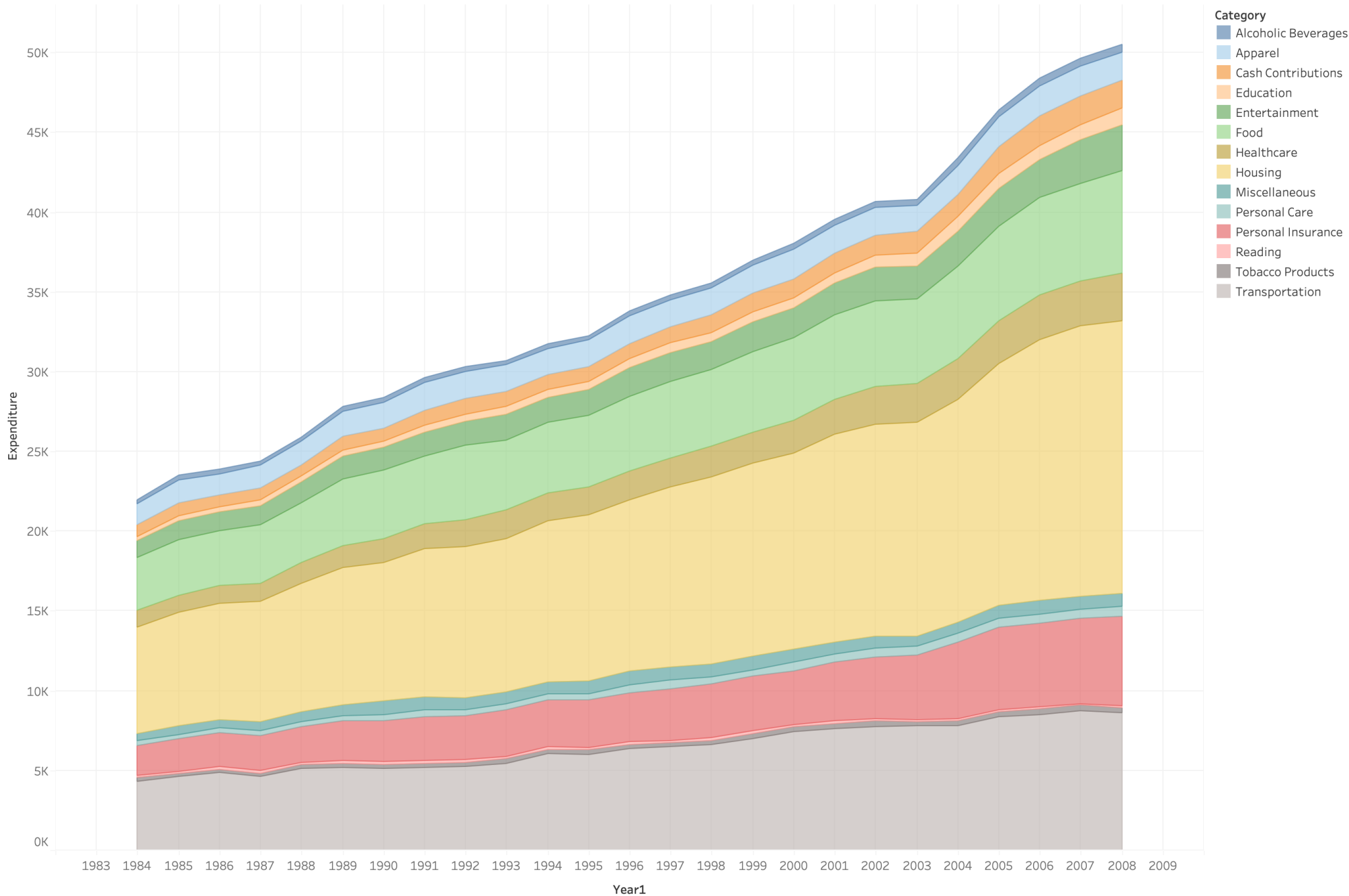
Category. Size shows sum of Expenditure. The marks are labeled by Category.

Expenditures - Area Chart



The plot of sum of Expenditure for Year1.

Expenditures - Stacked Area Chart



The plot of sum of Expenditure for Year1. Color shows details about Category.