Exercise 3.2: Tree Maps, Area and Stacked Area Charts

Scott Breitbach

DSC640 - 01/13/2022

Plots Using **Python**

Load Data

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

In [2]:
```
# Load data
unempDF = pd.read_csv('unemployement-rate-1948-2010.csv')
expenDF = pd.read_csv('expenditures.csv')
```

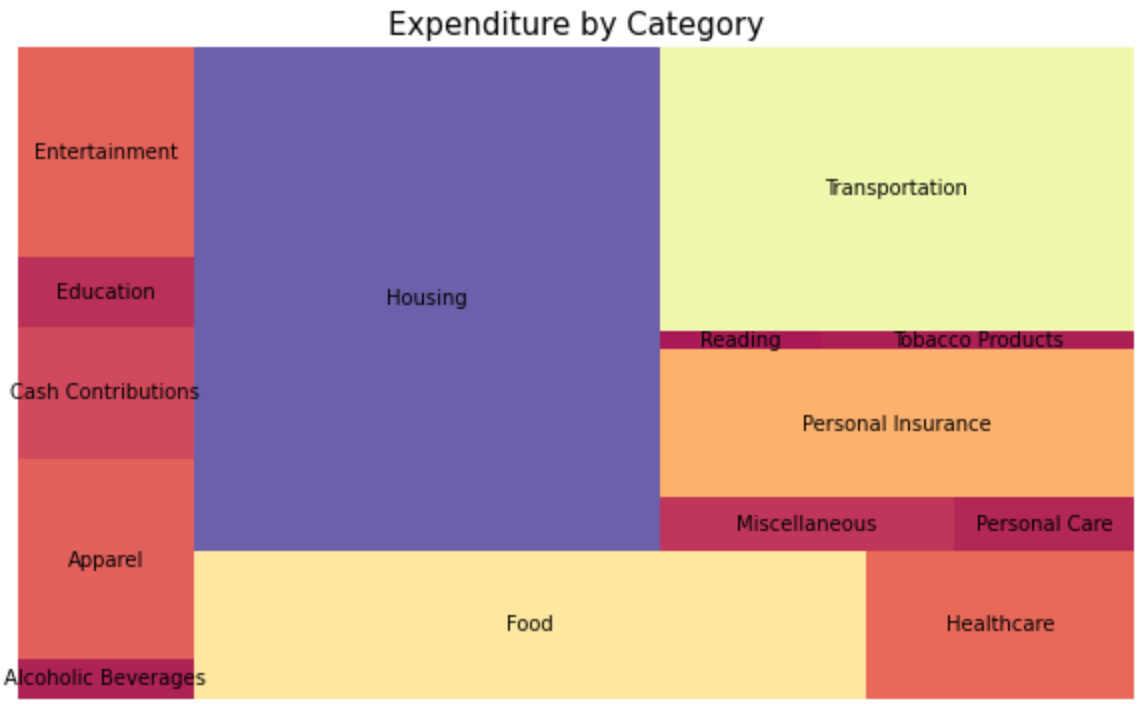
Tree Map

In [3]:
```
# Load libraries
import squarify
```

In [4]:
```
# Group expenses by category
expenCat = expenDF.groupby('category').sum()

# Set up values for chart
size = expenCat.expenditure
labels = expenCat.index
```

In [5]:
```
# Set up color palette, mapped to values
cmap = matplotlib.cm.Spectral # Blues is pretty good too
mini = min(size)
maxi = max(size)
norm = matplotlib.colors.Normalize(vmin=mini, vmax=maxi)
colors = [cmap(norm(value)) for value in size]
```

In [6]:
```
# Initialize the matplotlib figure
f, ax = plt.subplots(figsize=(10, 6))

# Plot tree map using squarify and matplotlib
squarify.plot(sizes = size,
              label = labels,
              alpha = 0.9,
              color = colors)
plt.axis('off')

# Add chart title
plt.title("Expenditure by Category", fontsize = 15)

plt.show()
```

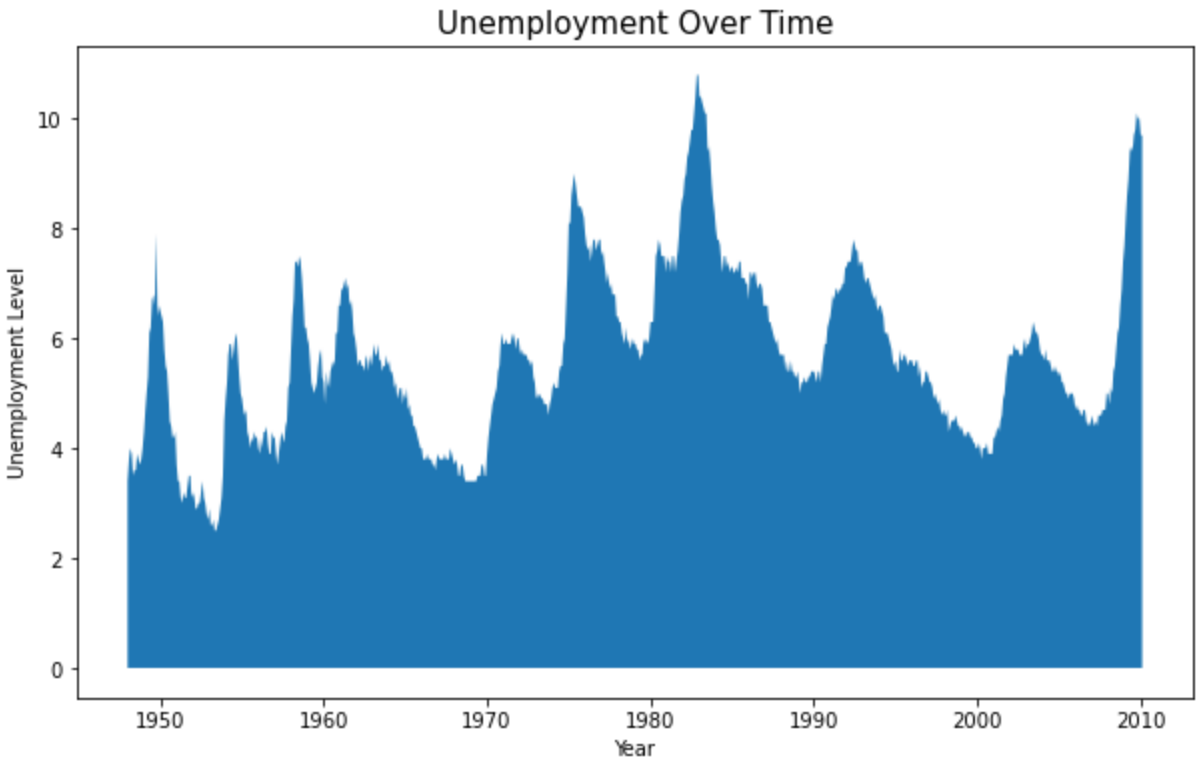


Area Chart

In [7]:
```
# Convert Month numbers to Month names
monthDict = {"M01": "Jan", "M02": "Feb", "M03": "Mar",
             "M04": "Apr", "M05": "May", "M06": "Jun",
             "M07": "Jul", "M08": "Aug", "M09": "Sep",
             "M10": "Oct", "M11": "Nov", "M12": "Dec"}
unempDF.replace({"Period": monthDict}, inplace=True)

# Combine Month and Year columns to a single datetime column
unempDF["MoYr"] = pd.to_datetime(unempDF.Period + ' ' + unempDF.Year.astype(str))
```

In [8]:
```
# Select data
x = unempDF.MoYr
y = unempDF.Value
```

In [9]:
```
# Initialize the matplotlib figure
f, ax = plt.subplots(figsize=(10, 6))

# Plot Area Chart
plt.fill_between(x, y)

# Add chart title and labels
plt.title("Unemployment Over Time", fontsize = 15)
plt.xlabel("Year")
plt.ylabel("Unemployment Level")

plt.show()
```



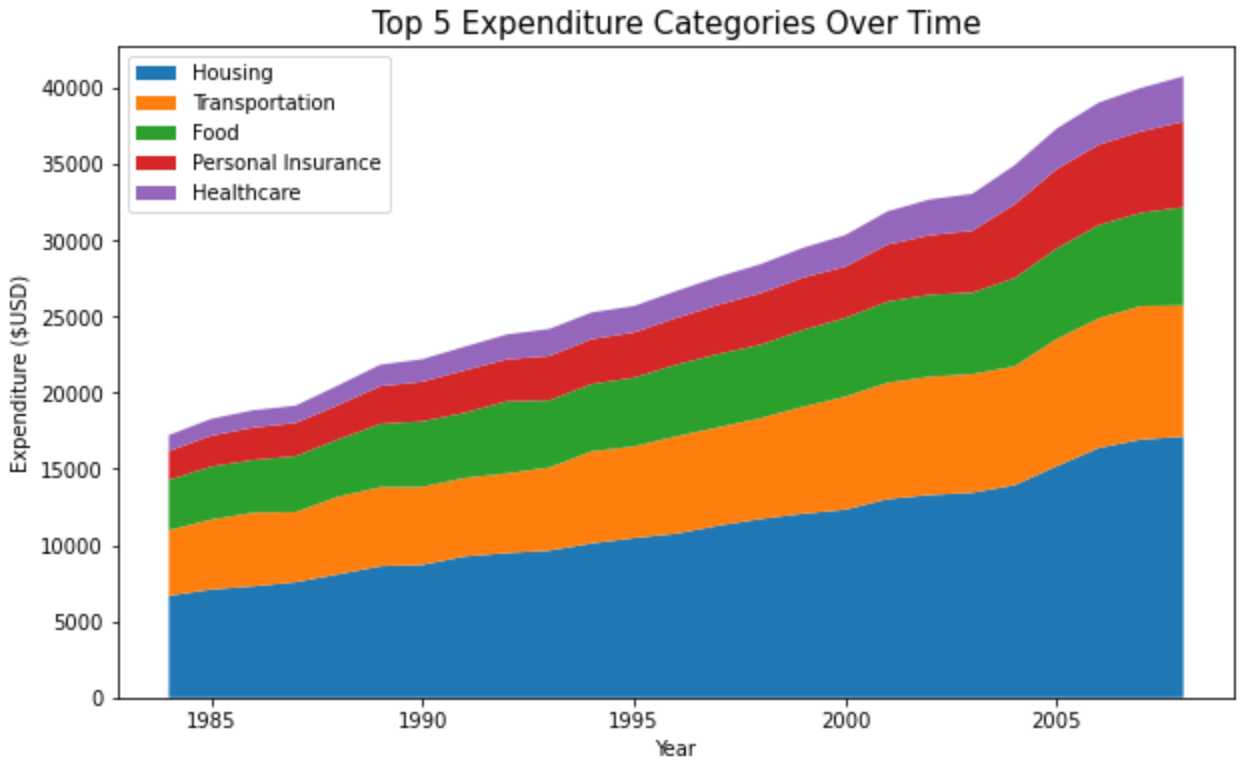
Stacked Area Chart

In [10]:
```
# Pivot dataframe into category expenditures by year
pivotDF = expenDF.pivot(index='year', columns='category')['expenditure']
```

In [11]:
```
# Get list of top 5 categories
sortedAvg = pivotDF.mean().sort_values(ascending=False)
topFiveCat = sortedAvg[:5].index.tolist()
```

In [12]:
```
# Select Data for chart
x = pivotDF.index
y0 = pivotDF[topFiveCat[0]]
y1 = pivotDF[topFiveCat[1]]
y2 = pivotDF[topFiveCat[2]]
y3 = pivotDF[topFiveCat[3]]
y4 = pivotDF[topFiveCat[4]]
```

In [13]:
```
# Initialize the matplotlib figure
f, ax = plt.subplots(figsize=(10, 6))

# Plot Stacked Area Chart
plt.stackplot(x, y0, y1, y2, y3, y4, labels=topFiveCat)
plt.legend(loc='upper left')

# Add chart title and labels
plt.title("Top 5 Expenditure Categories Over Time", fontsize = 15)
plt.xlabel("Year")
plt.ylabel("Expenditure ($USD)")

plt.show()
```



Exercise 3.2: TreeMaps, Area and Stacked Area Charts

Scott Breitbach

1/15/2022

Plots Using R

```
knitr::opts_chunk$set(echo = TRUE, warning = FALSE)

# Set Working Directory
setwd("C:/Users/micha/OneDrive/Documents/GitHub/DSC640/Weeks5-6/")

# Load libraries
library(treemap)
```

```
## Warning: package 'treemap' was built under R version 4.1.2
```

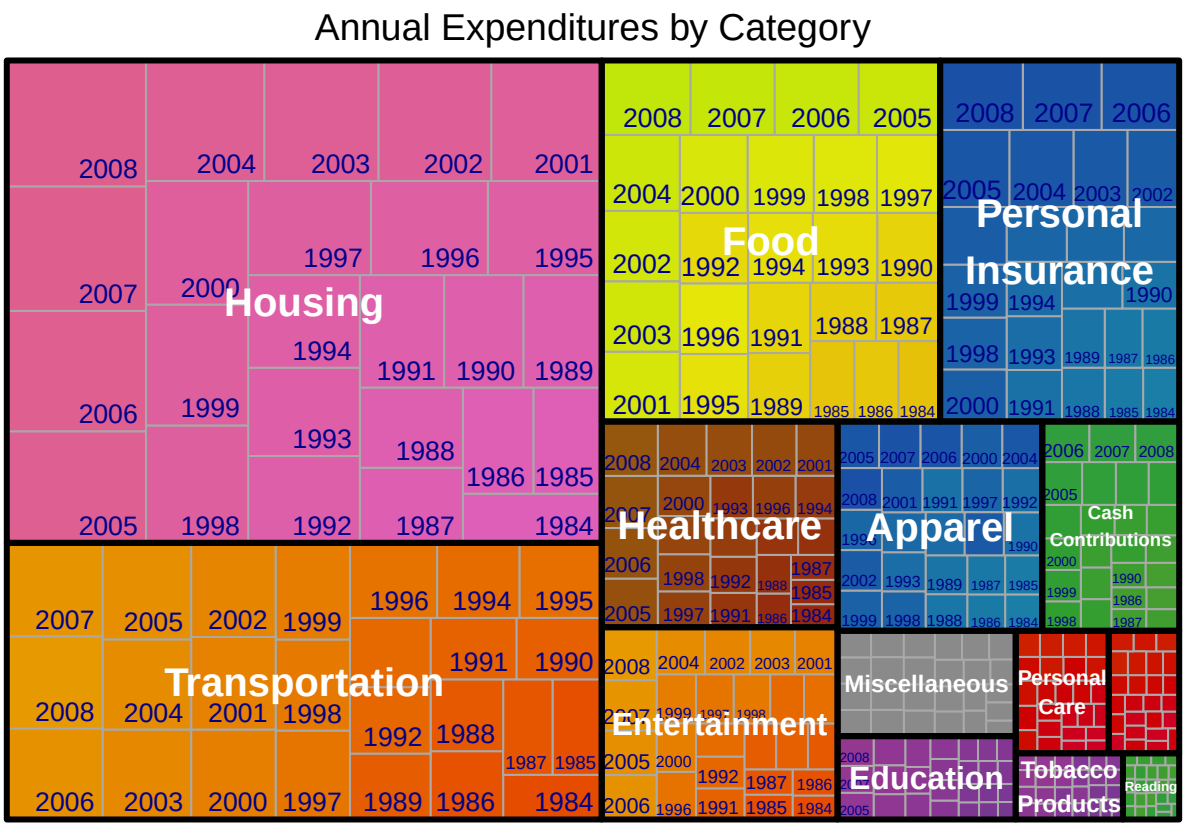
```
library(ggplot2)
```

Load Data

```
# Load data
unempDF <- read.csv("unemployment_cleaned.csv")
expenDF <- read.csv("expenditures.csv")
```

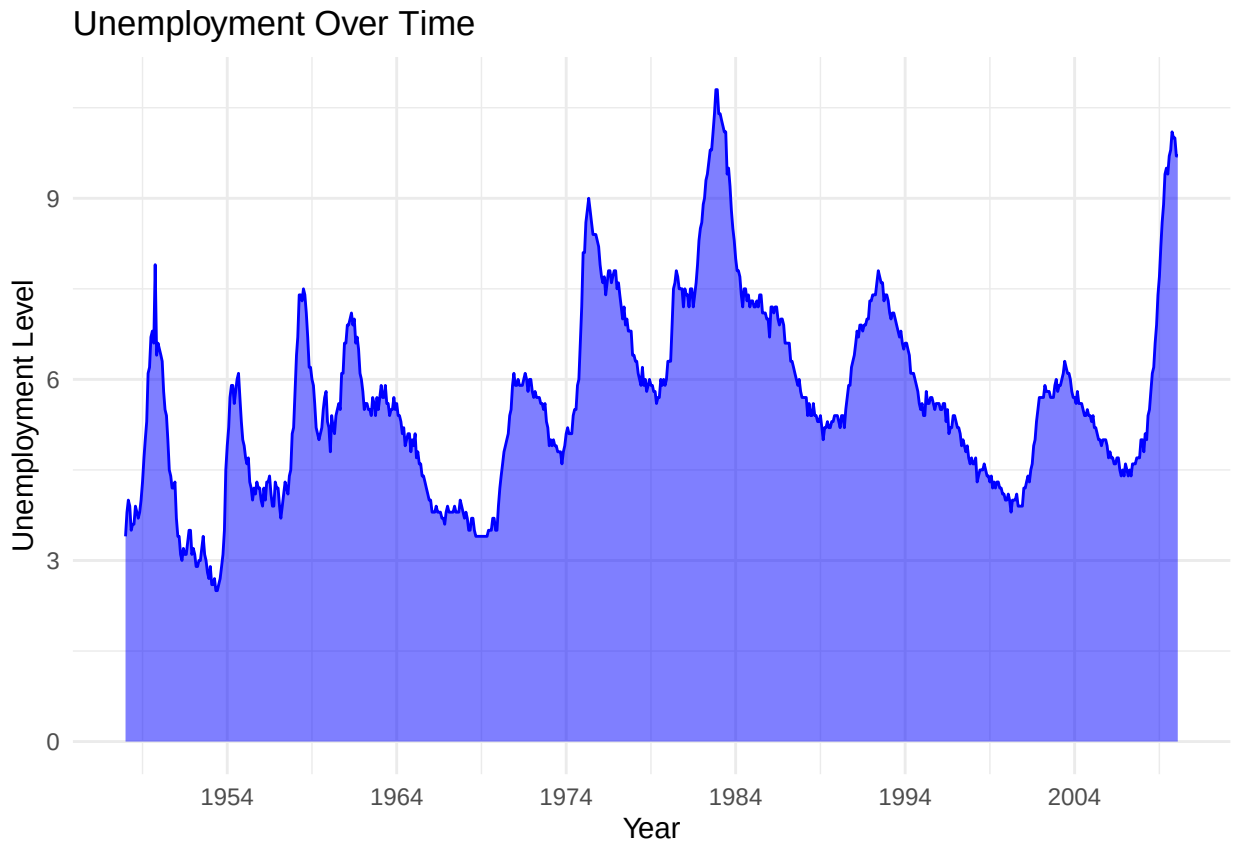
Tree Map

```
# Plot a TreeMap
treemap(expenDF,
        index = c("category", "year"),
        vSize = "expenditure",
        type = "index",
        palette = "Set1",
        title = "Annual Expenditures by Category",

        fontsize.labels = c(15, 10),    # Label size by level
        fontcolor.labels = c("white", "dark blue"),
        fontface.labels = c(2,1),       # 2: bold, 1: normal
        bg.labels = c("transparent"),   # Label background
        align.labels = list(            # Label locations
                c("center", "center"),
                c("right", "bottom")),
        overlap.labels = 0.5,           # Label overlap tolerance
        inflate.labels = FALSE,         # Increase label to size of box

        border.col = c("black", "dark gray"), # Color of borders
        border.lwds = c(3,1)            # Line width of borders
)
```



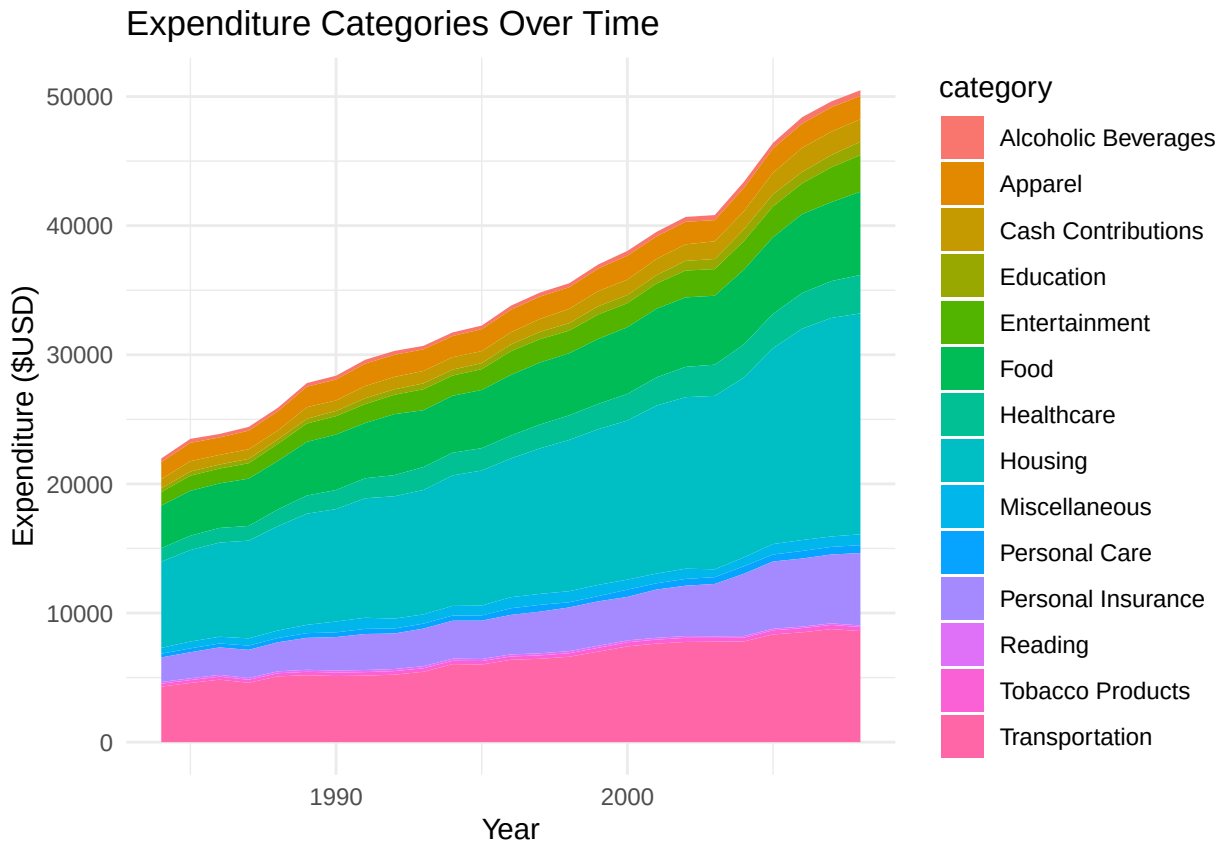
Area Chart

```
# Convert column to datetime data type
unempDF$MoYr <- as.Date(unempDF$MoYr, format = "%Y-%m-%d")

# Plot Area Chart
ggplot(unempDF, aes(x=MoYr, y=Value)) +
  geom_area(fill="blue", alpha=0.5) +
  geom_line(color="blue", size=0.5) +
  theme_minimal() +
  labs(y="Unemployment Level", x="Year") +
  ggtitle("Unemployment Over Time") +
  scale_x_date(date_labels="%Y", breaks = "10 years")
```



Stacked Area Chart

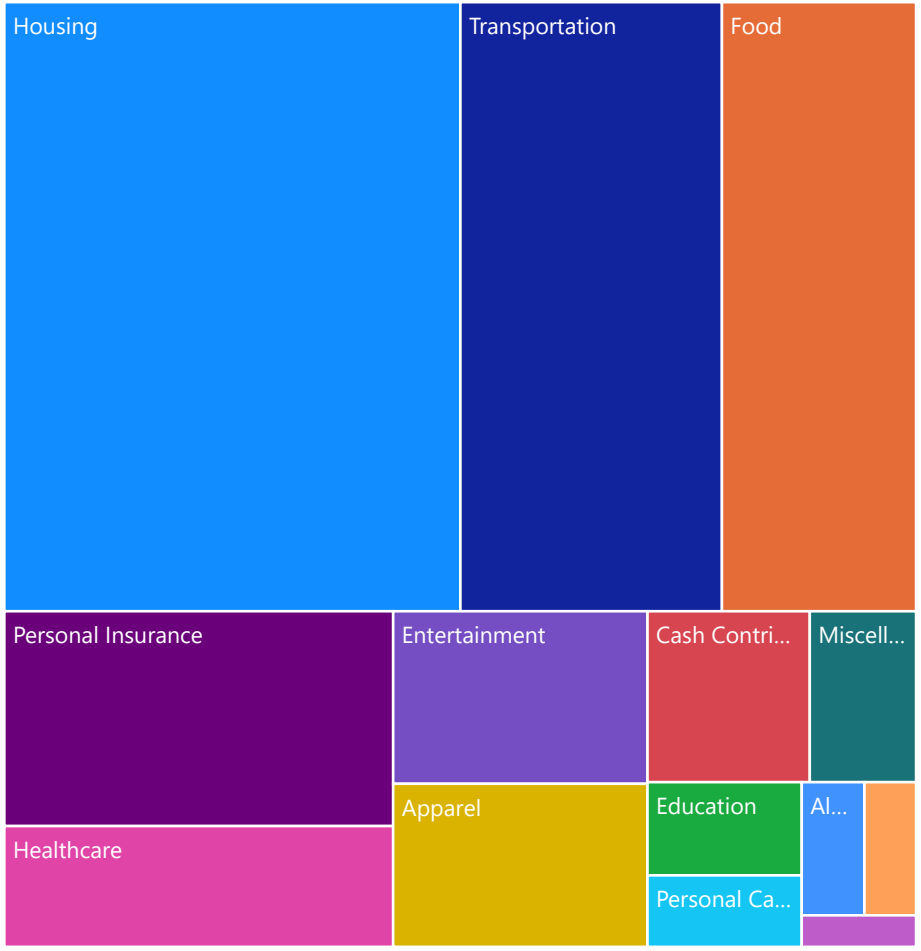
```
# Plot Stacked Area Chart
ggplot(expenDF, aes(x=year, y=expenditure, fill=category)) +
  geom_area() +
  theme_minimal() +
  labs(y="Expenditure ($USD)", x="Year") +
  ggtitle("Expenditure Categories Over Time")
```



Power BI

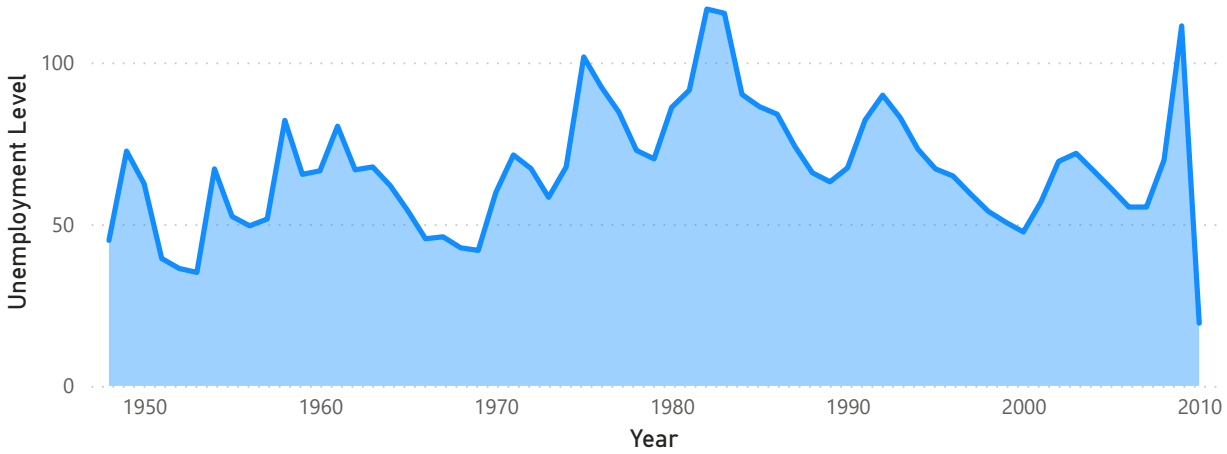
Tree Map

Expenditure by Category



Area Chart

Unemployment Over Time



Stacked Area Chart

Top 5 Expenditure Categories Over Time