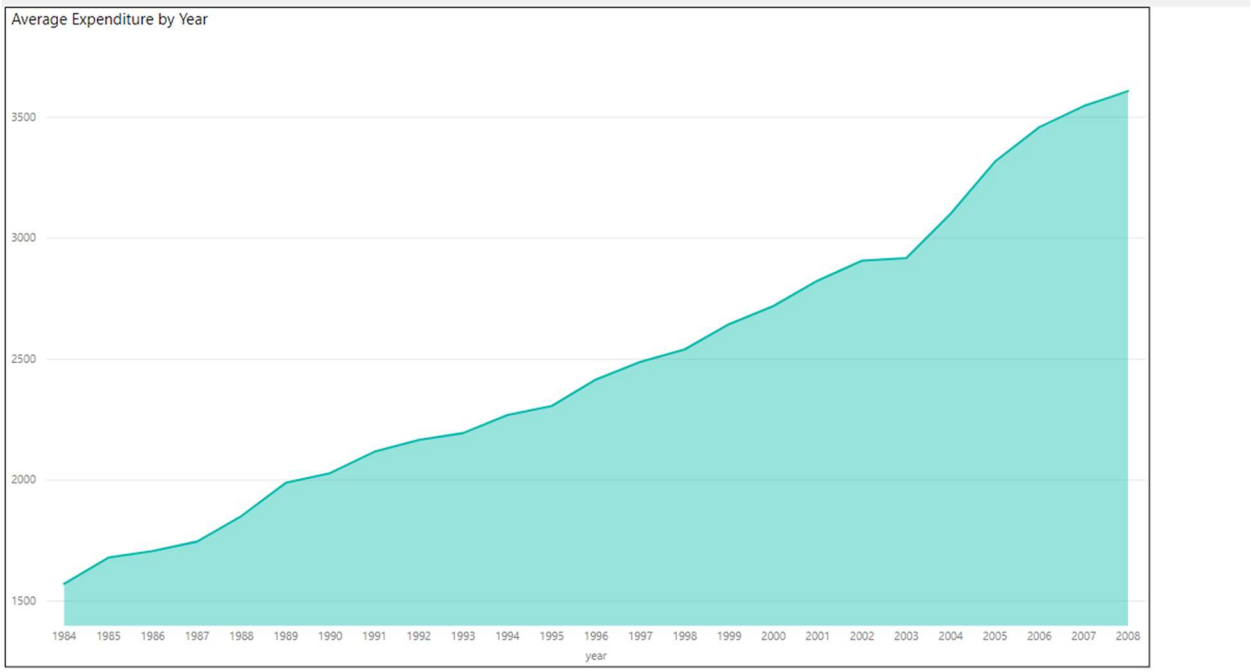
Sri R Sankaranarayanan – Week 5 & 6 - DSC640 – Data Visualization

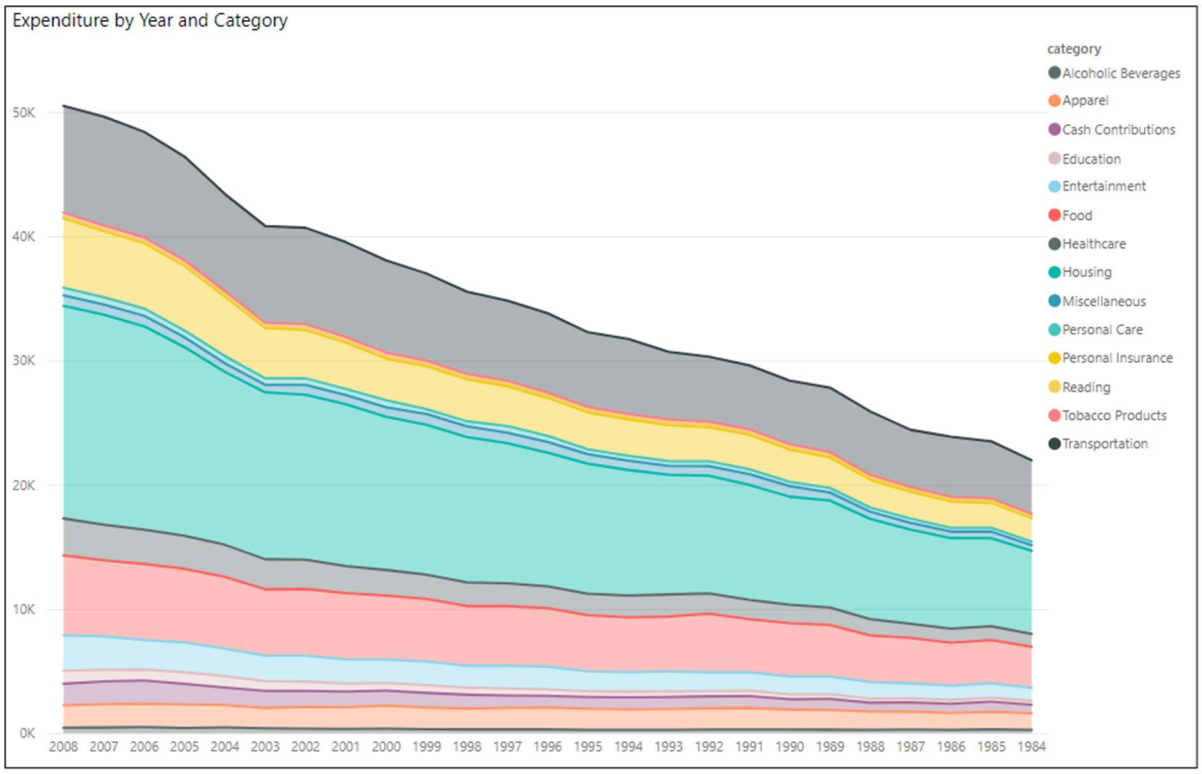
1. Power BI

TreeMap

Area chart



Stacked Area Chart



2. Python

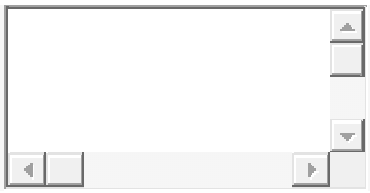
You need to submit 3 tree maps, 3 area charts and 3 stacked area charts using Tableau or PowerBI, Python and R using the data below (or your own datasets). You can also submit using D3. You can choose which library to use in Python or R, documentation is provided to help you decide and as you start to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

Python

Data Import step

```
# Import libraries
import csv
import pandas as pd
import matplotlib.pyplot as plt
import matplotlib as mpl
import squarify
from datetime import datetime as dt
```

In [102]:

```
11. 
12. # Read world population data
13. dirData = 'ex3-3'
14. file_expenditures = 'expenditures.txt'
15. file_unemployement = 'unemployement-rate-1948-2010.csv'
16.
17. dir_expenditures = dirData+'/'+file_expenditures
18. dir_unemployment = dirData+'/'+file_unemployement
19.
20. raw_expenditures = pd.read_csv(dir_expenditures, sep = '\t', header=0)
21. raw_unemployment = pd.read_csv(dir_unemployment)
22.
23. # Calculate total expenditure for categories
24. expenditures_cat = raw_expenditures.groupby(['category'])['expenditure'].sum().reset_index()
25.
26. # Calculate total expenditure by year
27. expenditures_year = raw_expenditures.groupby(['year'])['expenditure'].sum().reset_index()
28.
29. print(raw_expenditures.head())
30. print(expenditures_cat.head())
31. print(expenditures_year.head())
32. print(raw_unemployment.head())
33.     year             category  expenditure  sex
34.  0  2008                 Food         6443    1
35.  1  2008  Alcoholic Beverages          444    1
36.  2  2008              Housing        17109    1
37.  3  2008              Apparel         1801    1
38.  4  2008       Transportation         8604    1
39.                 category  expenditure
40.  0  Alcoholic Beverages         8424
41.  1              Apparel        41833
42.  2   Cash Contributions        27987
43.  3            Education        14498
44.  4        Entertainment        44273
45.     year  expenditure
46.  0  1984        21972
47.  1  1985        23489
48.  2  1986        23865
49.  3  1987        24415
50.  4  1988        25893
```

```
      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
```

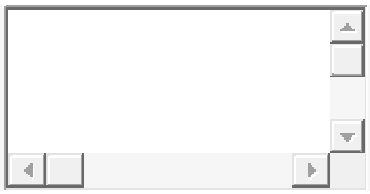
57. Treemap

58. Expenditure data

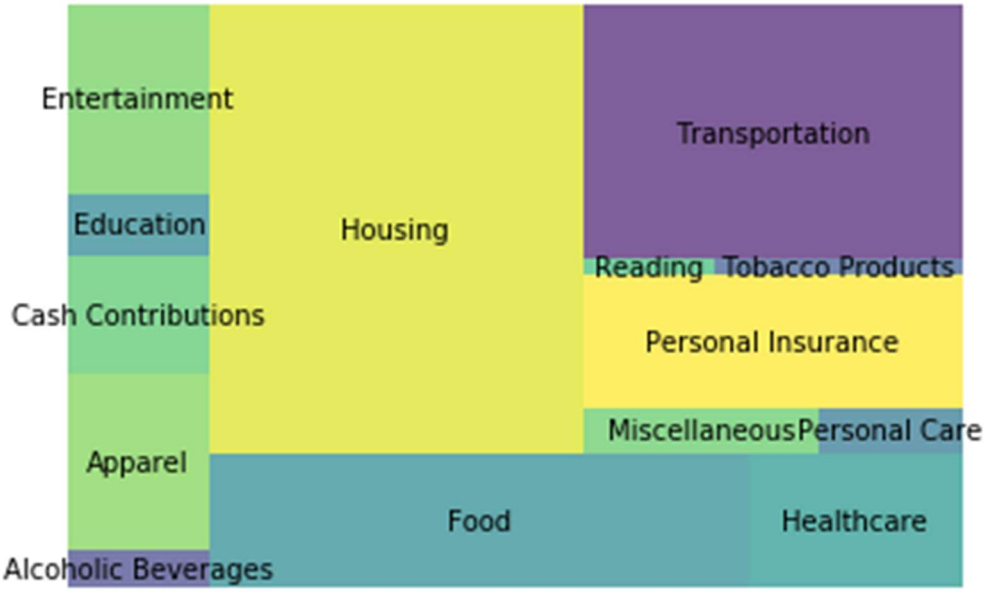
59. For this treemap, I would like to see how much each category has costed in total.

60. In [98]:

61.



```
62. # Create tree map
63. squarify.plot(sizes=expenditures_cat['expenditure'], label=expenditures_cat['category'], alpha=.7 )
64. plt.axis('off')
65. plt.show()
```



66.

67. Area Chart

68. Expenditure data

69. For this area chart, I would like to see how much was the total expenditure every year

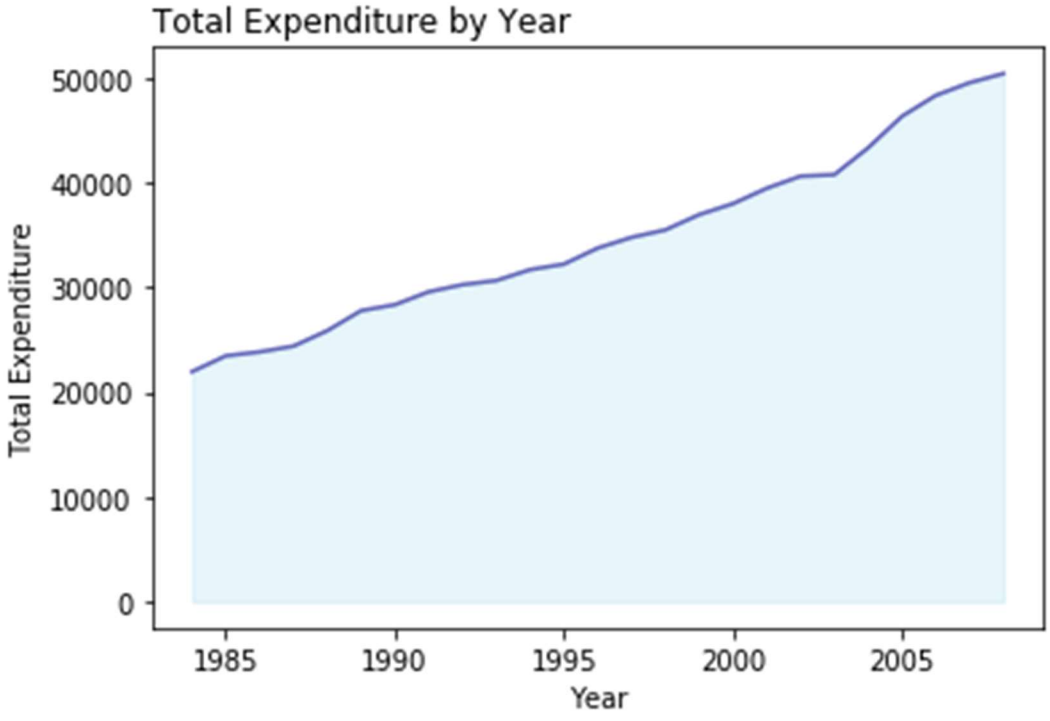
70. In [121]:

71.

```
72. # Create x and y values to plot
73. x = expenditures_year['year']
74. y = expenditures_year['expenditure']
75.
76. # Add a stronger line on top (edge)
77. plt.fill_between( x, y, color='skyblue', alpha=0.2)
78. plt.title('Total Expenditure by Year', loc='left')
79. plt.xlabel('Year')
80. plt.ylabel('Total Expenditure')
81. plt.plot(x, y, color='darkblue', alpha=0.6)
```

82. Out[121]:

83. `[<matplotlib.lines.Line2D at 0xd0ea9f0>]`



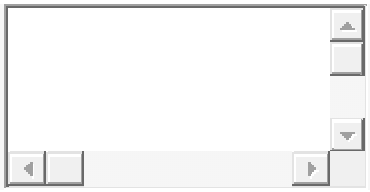
84.

85. Stacked Area Chart

86. Expenditure data

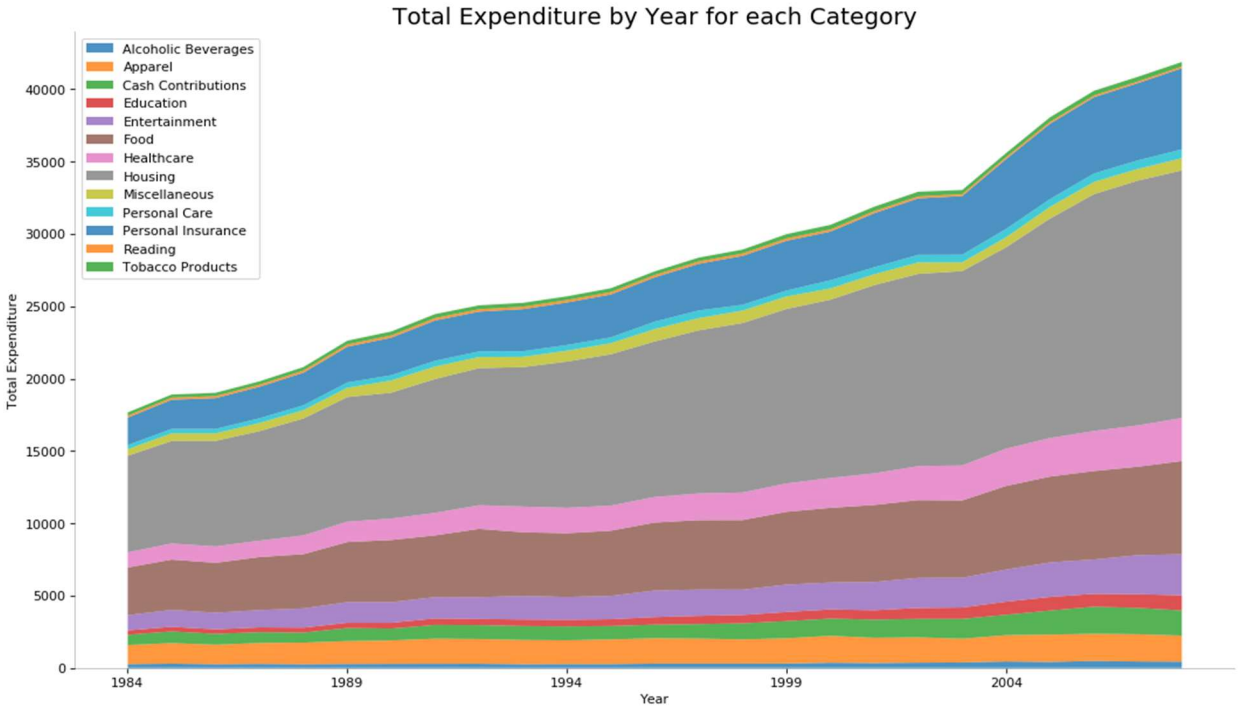
87. For stacked area chart, I would like to see how much was the total expenditure every year for each category

88. In [217]:

89. 

```
90. # Reshape data to be used for stacked area chart
91. plt_expenditures = raw_expenditures.loc[:, raw_expenditures.columns != 'sex'].pivot(index='year', co
    lumns='category', values='expenditure')
92. plt_expenditures.reset_index(level=0, inplace=True)
93.
94. # Draw Plot and Annotate
95. fig, ax = plt.subplots(1,1,figsize=(16, 9), dpi= 80)
96. columns = plt_expenditures.columns[1:]
97. labs = plt_expenditures.values.tolist()
98.
99. # Prepare data
100.  x  = plt_expenditures['year'].values.tolist()
101.  y0 = plt_expenditures[columns[0]].values.tolist()
102.  y1 = plt_expenditures[columns[1]].values.tolist()
103.  y2 = plt_expenditures[columns[2]].values.tolist()
104.  y3 = plt_expenditures[columns[3]].values.tolist()
105.  y4 = plt_expenditures[columns[4]].values.tolist()
106.  y5 = plt_expenditures[columns[5]].values.tolist()
107.  y6 = plt_expenditures[columns[6]].values.tolist()
108.  y7 = plt_expenditures[columns[7]].values.tolist()
109.  y8 = plt_expenditures[columns[8]].values.tolist()
110.  y9 = plt_expenditures[columns[9]].values.tolist()
111.  y10 = plt_expenditures[columns[10]].values.tolist()
112.  y11 = plt_expenditures[columns[11]].values.tolist()
113.  y12 = plt_expenditures[columns[12]].values.tolist()
114.  y = np.vstack([y0, y1, y2, y3, y4, y5, y6, y7, y8, y9, y10, y11, y12])
115.
116.  # Plot for each column
117.  labs = columns.values.tolist()
118.  ax = plt.gca()
119.  ax.stackplot(x, y, labels=labs, alpha=0.8)
120.
121.  # Create title
122.  ax.set_title('Total Expenditure by Year for each Category', fontsize=18)
123.  plt.xlabel('Year')
124.  plt.ylabel('Total Expenditure')
```

```
125.
126.  # Show legend
127.  ax.legend(fontsize=10, ncol=1, loc = 'upper left')
128.  plt.xticks(x[::5], fontsize=10, horizontalalignment='center')
129.
130.  # Lighten borders
131.  plt.gca().spines["top"].set_alpha(0)
132.  plt.gca().spines["bottom"].set_alpha(.3)
133.  plt.gca().spines["right"].set_alpha(0)
134.  plt.gca().spines["left"].set_alpha(.3)
135.
136.  # Output graph
137.  plt.show()
138.
```



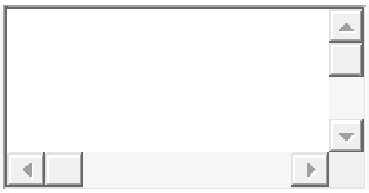
139.

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.

R

Data Import step

In [10]:



```
# Import required packages
library('magrittr')

# Import data to be used for visualization
dir = paste(getwd(),'ex3-3',sep = '/')
file_expenditures = 'expenditures.txt'
file_unemployement = 'unemployement-rate-1948-2010.csv'

raw_expenditures =  read.table(paste(dir,file_expenditures,sep='/'), header = TRUE, sep = '\t', dec = '.', fill = T
RUE)

raw_unemployement = read.csv2(paste(dir,file_unemployement,sep='/'), sep=',', stringsAsFactors = FALSE)
%>%
  dplyr::mutate(Value = as.numeric(Value)) %>%
  as.data.frame()

# Examine data
print(head(raw_expenditures))
print(head(raw_unemployement))
```

```
  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
6 2008          Healthcare        2976   1
    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
6 LNS14000000 1948    M06   3.6
```

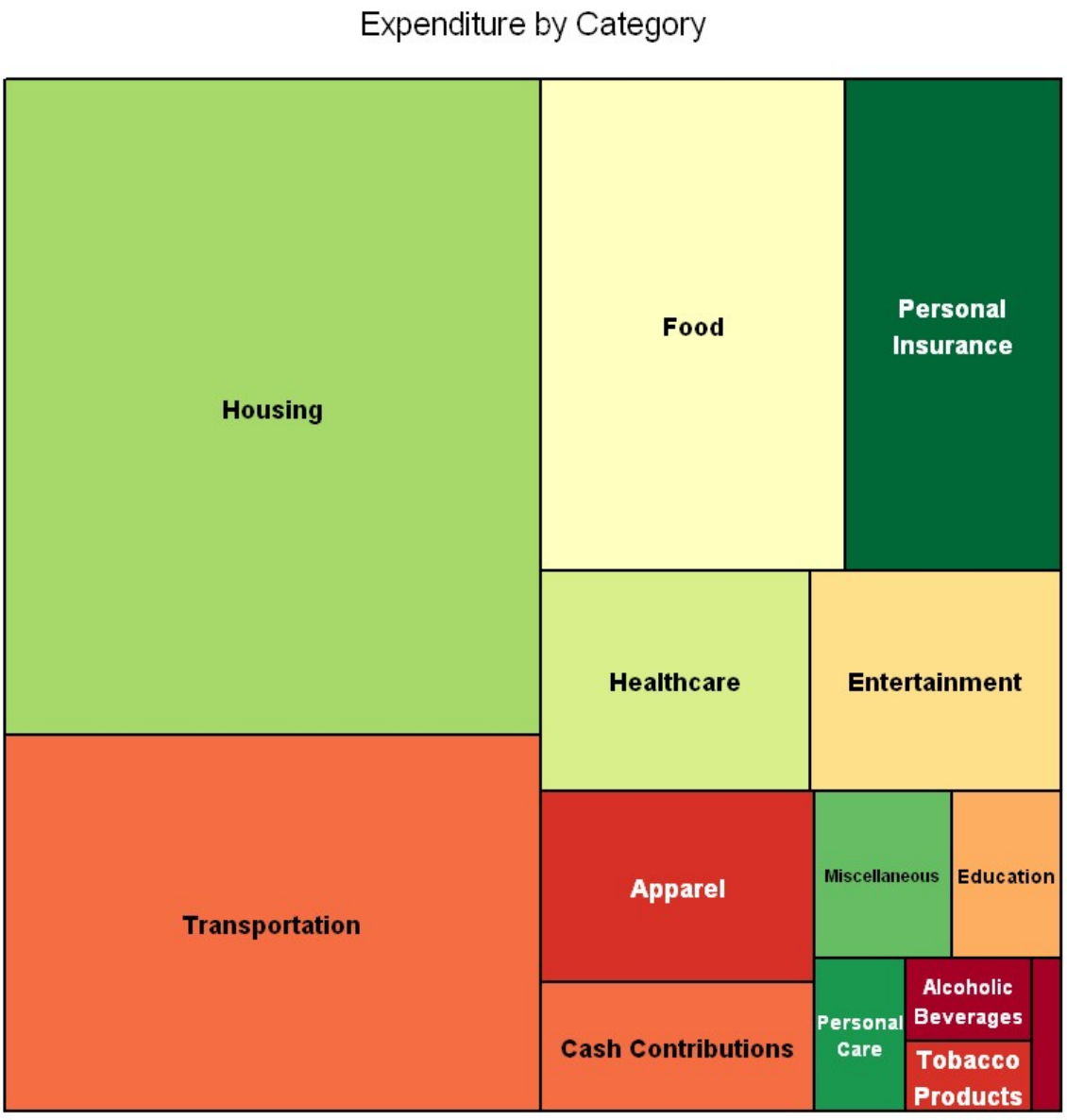
Treemap

Expenditure data

For this treemap, I would like to see how much each category has costed

In [2]:

```
treemap::treemap(raw_expenditures,index = c('category'),
                 vSize = 'expenditure',
                 title = 'Expenditure by Category',
                 palette = 'RdYlGn')
```

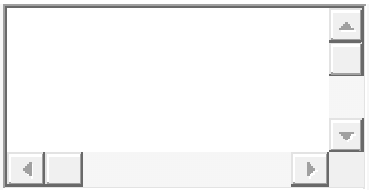


Area Chart

Unemployment data

For this graph, I would like to see the average value over the years, from the unemployment dataset.

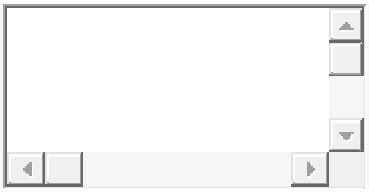
In [12]:



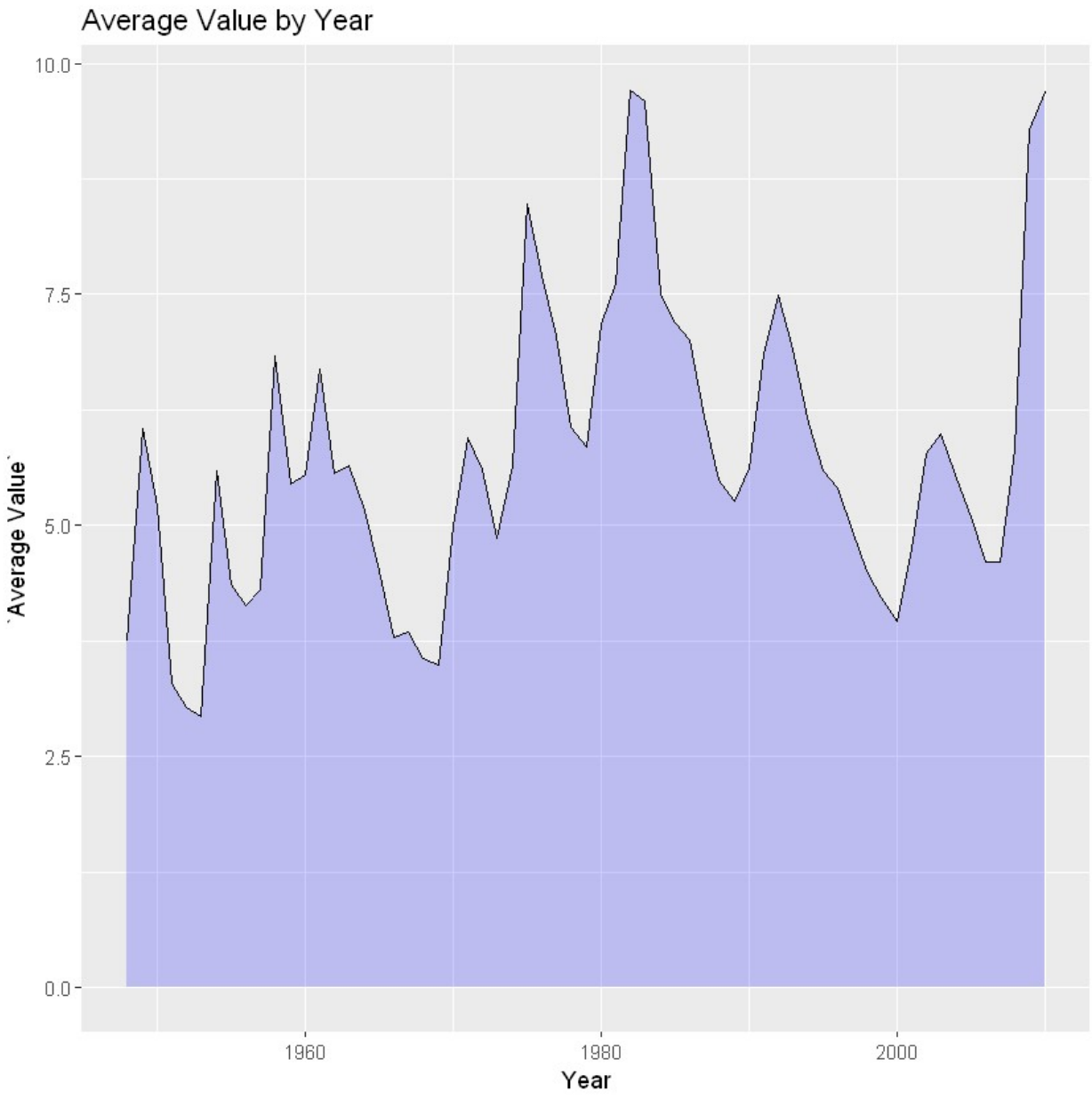
```
avg_unemployement = raw_unemployement %>%
  dplyr::group_by(Year) %>%
  dplyr::summarize('Average Value' = mean(Value))

head(avg_unemployement)
```

Year	Average Value
1948	3.750000
1949	6.050000
1950	5.208333
1951	3.283333
1952	3.025000
1953	2.925000

In [17]:



```
ggplot2::ggplot(avg_unemployement, ggplot2::aes(x=Year , y=`Average Value`)) +
ggplot2::geom_area( fill='blue', alpha=.2) +
ggplot2::geom_line() +
ggplot2::ggtitle('Average Value by Year')
```



Stacked Area Chart

Expenditure data

For this graph, I would like to see what is the trend of the expenditure of each category over the years.

In [8]:

```
ggplot2::ggplot(raw_expenditures, ggplot2::aes(x=year, y=expenditure, fill=category)) +
 ggplot2::geom_area()
```