3.2 Exercises: Tree Maps, Area Charts, and Stacked Area Charts

# Tableau Charts:







****

# Python Chart:

# Week 5-6 - Assignment

**Prepare - Tree Map, Area Chart & Stacked Area Chart**

**By**

**Shani Kumar**

### Introduction: Assignment Details

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.

### Source Data

<https://content.bellevue.edu/cst/dsc/640/datasets/ex3-2.zip>

In [1]:

*# Impprt required libraries/packages*

**import** **numpy** **as** **np**

**import** **pandas** **as** **pd**

**import** **squarify**

**import** **matplotlib.pyplot** **as** **plt**

*# configure display of graph*

%**matplotlib** inline

### Load data into a dataframe

In [2]:

*# load the csv file as a data frame*

expenditures = pd.read\_csv('expenditures.txt', sep = '**\t**', header=0)

*# summarize the shape of the dataset*

print("Expenditures:**\n**")

print("Dataset Shape: ",expenditures.shape)

*# see the sample of the data*

print("Sample Data: ")

expenditures.head()

Expenditures:

Dataset Shape: (350, 4)

Sample Data:

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 |

In [3]:

*# load the csv file as a data frame*

unemployement\_rt = pd.read\_csv('unemployement-rate-1948-2010.csv')

*# summarize the shape of the dataset*

print("Unemployment Rate:**\n\n**Dataset Shape: ",unemployement\_rt.shape)

*# see the sample of the data*

print("Sample Data: ")

unemployement\_rt.head()

Unemployment Rate:

Dataset Shape: (746, 4)

Sample Data:

Out[3]:

|  | **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 |

## Tree Map

In [4]:

*# Calculate total expenditure for categories*

expenditures\_cat = expenditures.groupby(['category'])['expenditure'].sum().reset\_index()

*# Plot tree map now*

squarify.plot(sizes=expenditures\_cat['expenditure'],

label=expenditures\_cat['category'],

alpha=.8 )

plt.title('Expenditure by Category')

plt.axis('off')

plt.show()

Chart, treemap chart

Description automatically generated

## Area Chart

In [5]:

*# Calculate total expenditure by year*

expenditures\_year = expenditures.groupby(['year'])['expenditure'].sum().reset\_index()

*# Area Chart*

plt.fill\_between(expenditures\_year['year'], expenditures\_year['expenditure'])

plt.xlabel('Year')

plt.ylabel('Expenditure')

plt.title('Expenditure by Year')

plt.show()

Icon

Description automatically generated

## Stacked Area Chart

In [6]:

*# Reorge the shape of dataset structure for stacked area chart*

expenditures\_reorg = expenditures.loc[:, expenditures.columns != 'sex'].pivot(index='year', columns='category', values='expenditure')

expenditures\_reorg.reset\_index(level=0, inplace=**True**)

*#Get label name*

labs = expenditures\_reorg.columns[1:].values.tolist()

*# see the sample of the data*

print("After Reorg: ")

expenditures\_reorg.head()

After Reorg:

Out[6]:

| **category** | **year** | **Alcoholic Beverages** | **Apparel** | **Cash Contributions** | **Education** | **Entertainment** | **Food** | **Healthcare** | **Housing** | **Miscellaneous** | **Personal Care** | **Personal Insurance** | **Reading** | **Tobacco Products** | **Transportation** |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **0** | 1984 | 275 | 1319 | 706 | 303 | 1055 | 3290 | 1049 | 6674 | 451 | 289 | 1897 | 132 | 228 | 4304 |
| **1** | 1985 | 306 | 1420 | 805 | 321 | 1170 | 3477 | 1108 | 7087 | 529 | 303 | 2016 | 141 | 219 | 4587 |
| **2** | 1986 | 271 | 1346 | 746 | 314 | 1149 | 3448 | 1135 | 7292 | 522 | 303 | 2127 | 140 | 230 | 4842 |
| **3** | 1987 | 289 | 1446 | 741 | 337 | 1193 | 3664 | 1135 | 7569 | 562 | 330 | 2175 | 142 | 232 | 4600 |
| **4** | 1988 | 269 | 1489 | 693 | 342 | 1329 | 3748 | 1298 | 8079 | 578 | 334 | 2249 | 150 | 242 | 5093 |

In [7]:

*# Stacked Area Chart*

plt.figure(figsize=(20,10))

plt.stackplot(expenditures\_reorg['year'],

expenditures\_reorg['Alcoholic Beverages'],

expenditures\_reorg['Apparel'],

expenditures\_reorg['Cash Contributions'],

expenditures\_reorg['Education'],

expenditures\_reorg['Entertainment'],

expenditures\_reorg['Food'],

expenditures\_reorg['Healthcare'],

expenditures\_reorg['Housing'],

expenditures\_reorg['Miscellaneous'],

expenditures\_reorg['Personal Care'],

expenditures\_reorg['Personal Insurance'],

expenditures\_reorg['Reading'],

expenditures\_reorg['Tobacco Products'],

expenditures\_reorg['Transportation'],

labels=labs, alpha=0.7)

plt.title('Expenditure for each Category by Year', fontsize=20)

plt.xlabel('Year', fontsize=15)

plt.ylabel('Expenditure', fontsize=15)

plt.legend(title='Category', fontsize=10, ncol=1, loc = 'upper left')

Out[7]:

<matplotlib.legend.Legend at 0x7fdf4bfda208>

Chart, line chart

Description automatically generated

# R Programming Chart:

Week 5-6: Prepare - Tree Map, Area Chart & Stacked Area Chart

Shani Kumar

## Week 5-6: Exercises: Charts

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.

**Data source** We are using dataset from [Data Source URL](https://content.bellevue.edu/cst/dsc/640/datasets/ex3-2.zip) file.

## 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

### Data structure:

## '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 ...

## 'data.frame': 746 obs. of 4 variables:  
## $ Series.id: Factor w/ 1 level "LNS14000000": 1 1 1 1 1 1 1 1 1 1 ...  
## $ Year : int 1948 1948 1948 1948 1948 1948 1948 1948 1948 1948 ...  
## $ Period : Factor w/ 12 levels "M01","M02","M03",..: 1 2 3 4 5 6 7 8 9 10 ...  
## $ Value : num 3.4 3.8 4 3.9 3.5 3.6 3.6 3.9 3.8 3.7 ...

### Construct Charts:

**Treemap Chart**

Chart, treemap chart

Description automatically generated

**Area Chart**

Chart, line chart

Description automatically generated

**Stacked Area Chart**

Chart

Description automatically generated