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

1. Power BI

**TreeMap**

**Chart, treemap chart

Description automatically generated**

**Area chart**

**Chart, line chart

Description automatically generated**

**Stacked Area Chart**

**Chart, histogram

Description automatically generated**

1. **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

1. *# Import libraries*
2. **import** csv
3. **import** pandas **as** pd
4. **import** matplotlib.pyplot **as** plt
5. **import** matplotlib **as** mpl
6. **import** squarify
7. **from** datetime **import** datetime **as** dt
8. In [102]:
9. 
10. *# Read world population data*
11. dirData **=** 'ex3-3'
12. file\_expenditures **=** 'expenditures.txt'
13. file\_unemployement **=** 'unemployement-rate-1948-2010.csv'
14. ​
15. dir\_expenditures **=** dirData**+**'/'**+**file\_expenditures
16. dir\_unemployment **=** dirData**+**'/'**+**file\_unemployement
17. ​
18. raw\_expenditures **=** pd.read\_csv(dir\_expenditures, sep **=** '\t', header**=**0)
19. raw\_unemployment **=** pd.read\_csv(dir\_unemployment)
20. ​
21. *# Calculate total expenditure for categories*
22. expenditures\_cat **=** raw\_expenditures.groupby(['category'])['expenditure'].sum().reset\_index()
23. ​
24. *# Calculate total expenditure by year*
25. expenditures\_year **=** raw\_expenditures.groupby(['year'])['expenditure'].sum().reset\_index()
26. ​
27. print(raw\_expenditures.head())
28. print(expenditures\_cat.head())
29. print(expenditures\_year.head())
30. print(raw\_unemployment.head())
31. year category expenditure sex
32. 0 2008 Food 6443 1
33. 1 2008 Alcoholic Beverages 444 1
34. 2 2008 Housing 17109 1
35. 3 2008 Apparel 1801 1
36. 4 2008 Transportation 8604 1
37. category expenditure
38. 0 Alcoholic Beverages 8424
39. 1 Apparel 41833
40. 2 Cash Contributions 27987
41. 3 Education 14498
42. 4 Entertainment 44273
43. year expenditure
44. 0 1984 21972
45. 1 1985 23489
46. 2 1986 23865
47. 3 1987 24415
48. 4 1988 25893
49. Series id Year Period Value
50. 0 LNS14000000 1948 M01 3.4
51. 1 LNS14000000 1948 M02 3.8
52. 2 LNS14000000 1948 M03 4.0
53. 3 LNS14000000 1948 M04 3.9
54. 4 LNS14000000 1948 M05 3.5

### Treemap

#### Expenditure data

1. For this treemap, I would like to see how much each category has costed in total.
2. In [98]:
3. 
4. *# Create tree map*
5. squarify.plot(sizes**=**expenditures\_cat['expenditure'], label**=**expenditures\_cat['category'], alpha**=**.7 )
6. plt.axis('off')
7. plt.show()
8. Chart, treemap chart

   Description automatically generated

### Area Chart

#### Expenditure data

1. For this area chart, I would like to see how much was the total expenditure every year
2. In [121]:
3. 
4. *# Create x and y values to plot*
5. x **=** expenditures\_year['year']
6. y **=** expenditures\_year['expenditure']
8. *# Add a stronger line on top (edge)*
9. plt.fill\_between( x, y, color**=**'skyblue', alpha**=**0.2)
10. plt.title('Total Expenditure by Year', loc**=**'left')
11. plt.xlabel('Year')
12. plt.ylabel('Total Expenditure')
13. plt.plot(x, y, color**=**'darkblue', alpha**=**0.6)
14. Out[121]:
15. [<matplotlib.lines.Line2D at 0xd0ea9f0>]
16. Chart, line chart

    Description automatically generated

### Stacked Area Chart

#### Expenditure data

1. For stacked area chart, I would like to see how much was the total expenditure every year for each category
2. In [217]:
3. 
4. *# Reshape data to be used for stacked area chart*
5. plt\_expenditures **=** raw\_expenditures.loc[:, raw\_expenditures.columns **!=** 'sex'].pivot(index**=**'year', columns**=**'category', values**=**'expenditure')
6. plt\_expenditures.reset\_index(level**=**0, inplace**=True**)
7. ​
8. *# Draw Plot and Annotate*
9. fig, ax **=** plt.subplots(1,1,figsize**=**(16, 9), dpi**=** 80)
10. columns **=** plt\_expenditures.columns[1:]
11. labs **=** plt\_expenditures.values.tolist()
12. ​
13. *# Prepare data*
14. x **=** plt\_expenditures['year'].values.tolist()
15. y0 **=** plt\_expenditures[columns[0]].values.tolist()
16. y1 **=** plt\_expenditures[columns[1]].values.tolist()
17. y2 **=** plt\_expenditures[columns[2]].values.tolist()
18. y3 **=** plt\_expenditures[columns[3]].values.tolist()
19. y4 **=** plt\_expenditures[columns[4]].values.tolist()
20. y5 **=** plt\_expenditures[columns[5]].values.tolist()
21. y6 **=** plt\_expenditures[columns[6]].values.tolist()
22. y7 **=** plt\_expenditures[columns[7]].values.tolist()
23. y8 **=** plt\_expenditures[columns[8]].values.tolist()
24. y9 **=** plt\_expenditures[columns[9]].values.tolist()
25. y10 **=** plt\_expenditures[columns[10]].values.tolist()
26. y11 **=** plt\_expenditures[columns[11]].values.tolist()
27. y12 **=** plt\_expenditures[columns[12]].values.tolist()
28. y **=** np.vstack([y0, y1, y2, y3, y4, y5, y6, y7, y8, y9, y10, y11, y12])
29. ​
30. *# Plot for each column*
31. labs **=** columns.values.tolist()
32. ax **=** plt.gca()
33. ax.stackplot(x, y, labels**=**labs, alpha**=**0.8)
34. ​
35. *# Create title*
36. ax.set\_title('Total Expenditure by Year for each Category', fontsize**=**18)
37. plt.xlabel('Year')
38. plt.ylabel('Total Expenditure')
39. ​
40. *# Show legend*
41. ax.legend(fontsize**=**10, ncol**=**1, loc **=** 'upper left')
42. plt.xticks(x[::5], fontsize**=**10, horizontalalignment**=**'center')
43. ​
44. *# Lighten borders*
45. plt.gca().spines["top"].set\_alpha(0)
46. plt.gca().spines["bottom"].set\_alpha(.3)
47. plt.gca().spines["right"].set\_alpha(0)
48. plt.gca().spines["left"].set\_alpha(.3)
49. ​
50. *# Output graph*
51. plt.show()
52. ​
53. Chart, line chart

    Description automatically generated

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 **=** TRUE)

​

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')

**Chart, treemap chart

Description automatically generated**

### 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')

**Chart, histogram

Description automatically generated**

### 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()

**Chart

Description automatically generated**