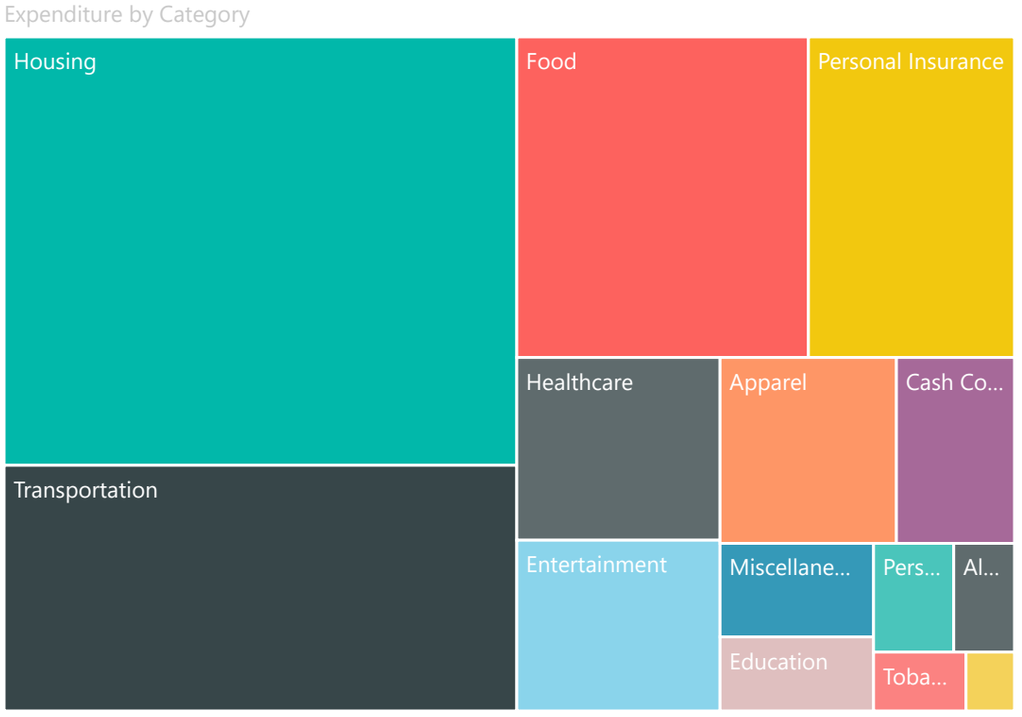
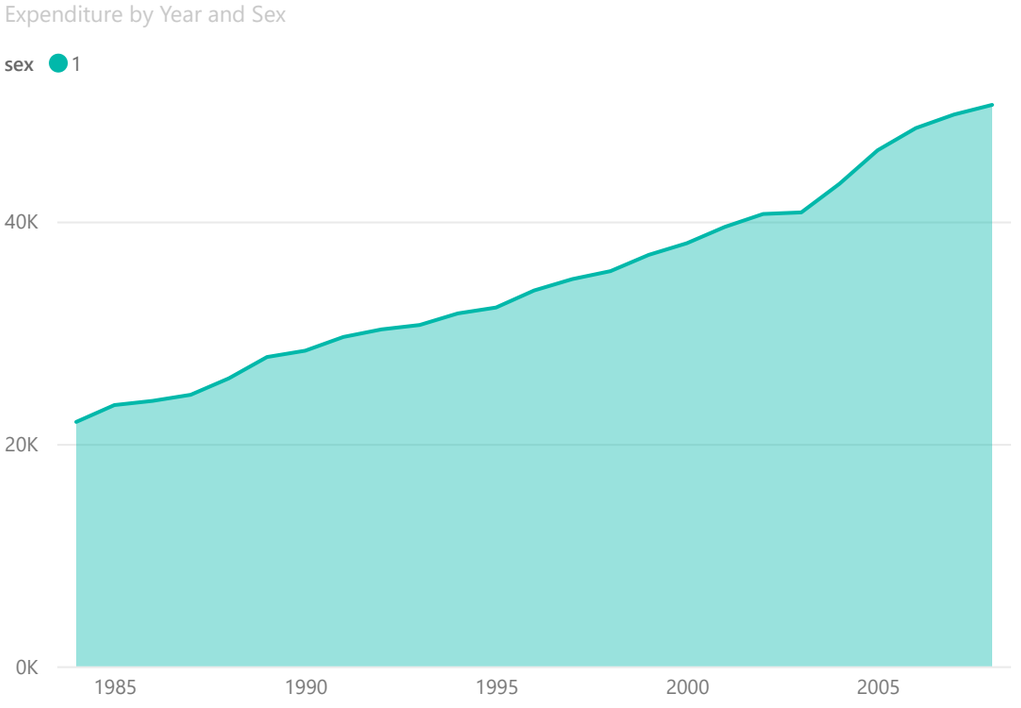
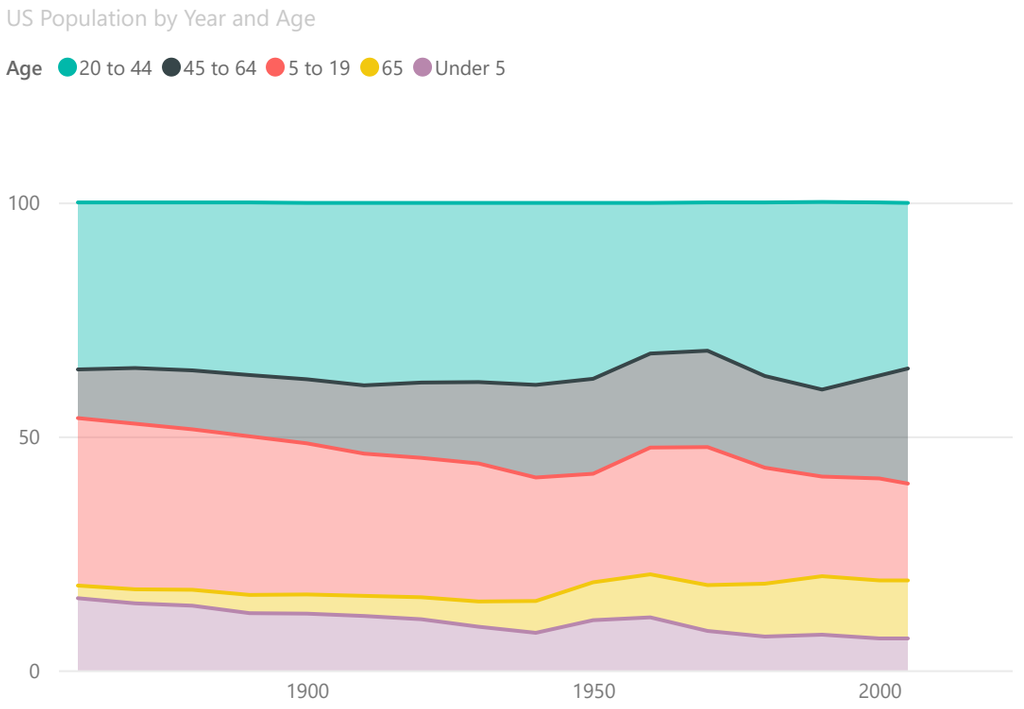
Power BI - Tree Map



Power BI - Area Chart



Power BI - Stacked Area Chart



Campbell640Week5-6

October 6, 2023

[39]:
```
# Load libraries
import pandas as pd
import numpy as np
# import data visualization libraries
import seaborn as sns
import matplotlib
import matplotlib.pyplot as plt
import squarify
```

[76]:
```
# Read csv files
expenditures = pd.read_csv('data sources/expenditures.txt', sep="\t")
population = pd.read_excel('data sources/us-population-by-age.xls')
```

[6]:
```
# Print shape
expenditures.shape
```

[6]:
```
(350, 4)
```

[77]:
```
# Print shape
population.shape
```

[77]:
```
(19, 6)
```

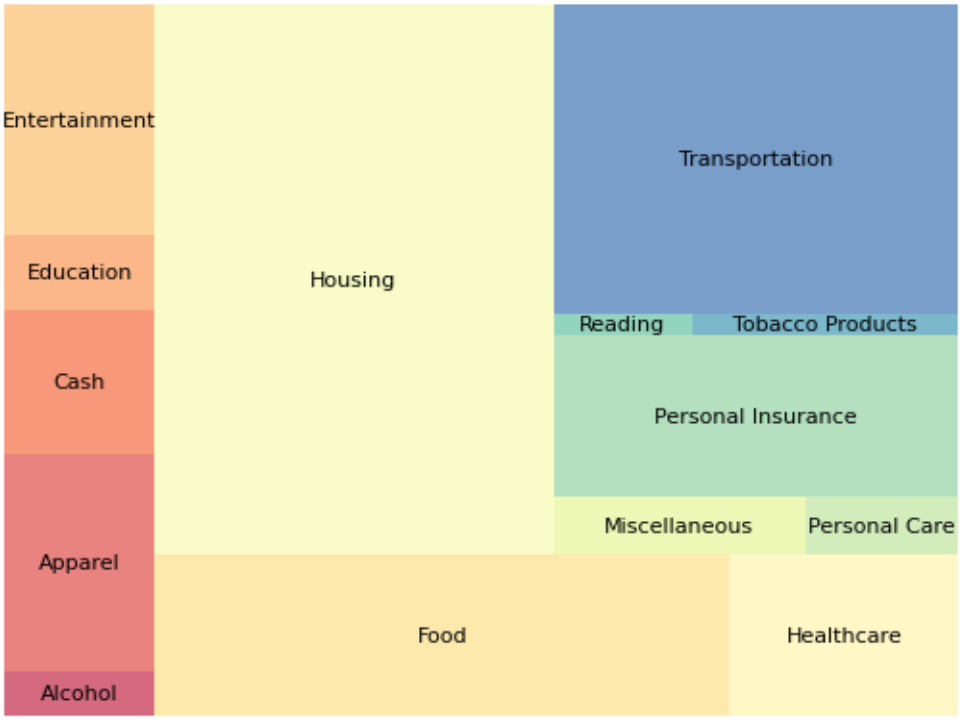
1 Python - Treemap

[63]:
```
# Group expenditures by category
cat_expend = expenditures.groupby('category', as_index=False)['expenditure'].
 ↪sum()
```

[64]:
```
# Shorten values
cat_expend.category = cat_expend.category.replace(['Cash Contributions',␣
 ↪'Alcoholic Beverages'],['Cash', 'Alcohol'])
```

[66]:
```
squarify.plot(sizes = cat_expend.expenditure, label = cat_expend.category,
              alpha = 0.7, color = sns.color_palette("Spectral", len(cat_expend.
 ↪expenditure)),
              text_kwargs={'fontsize':8, 'wrap':True})
```

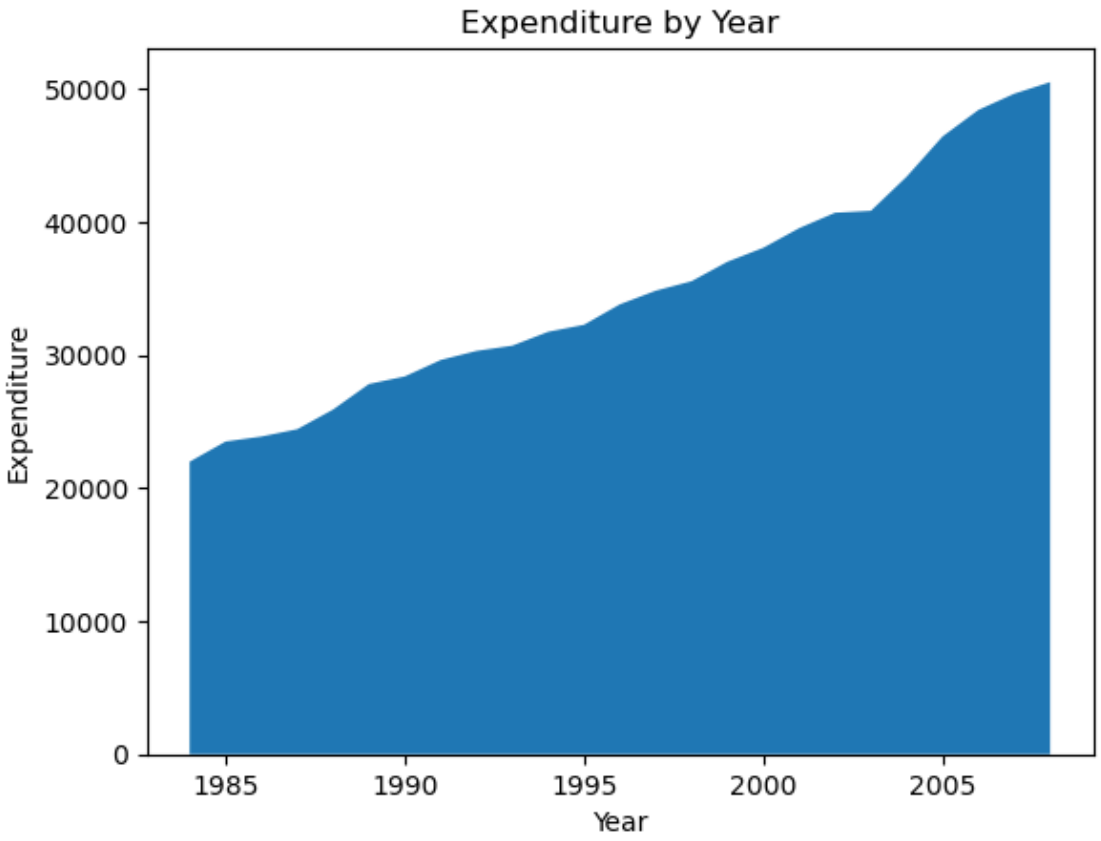
```
plt.axis('off')
plt.show()
```



2 Python - Area Chart

[67]:
```
# Group expenditures by year
year_expend = expenditures.groupby('year', as_index=False)['expenditure'].sum()
```

[72]:
```
plt.stackplot(year_expend.year, year_expend.expenditure)
plt.title("Expenditure by Year")
plt.xlabel("Year")
plt.ylabel("Expenditure")
plt.show()
```



3 Python - Stacked Area Chart

[101]:
```
# Rename year column
population = population.rename(columns={'Unnamed: 0': 'Year'})
```

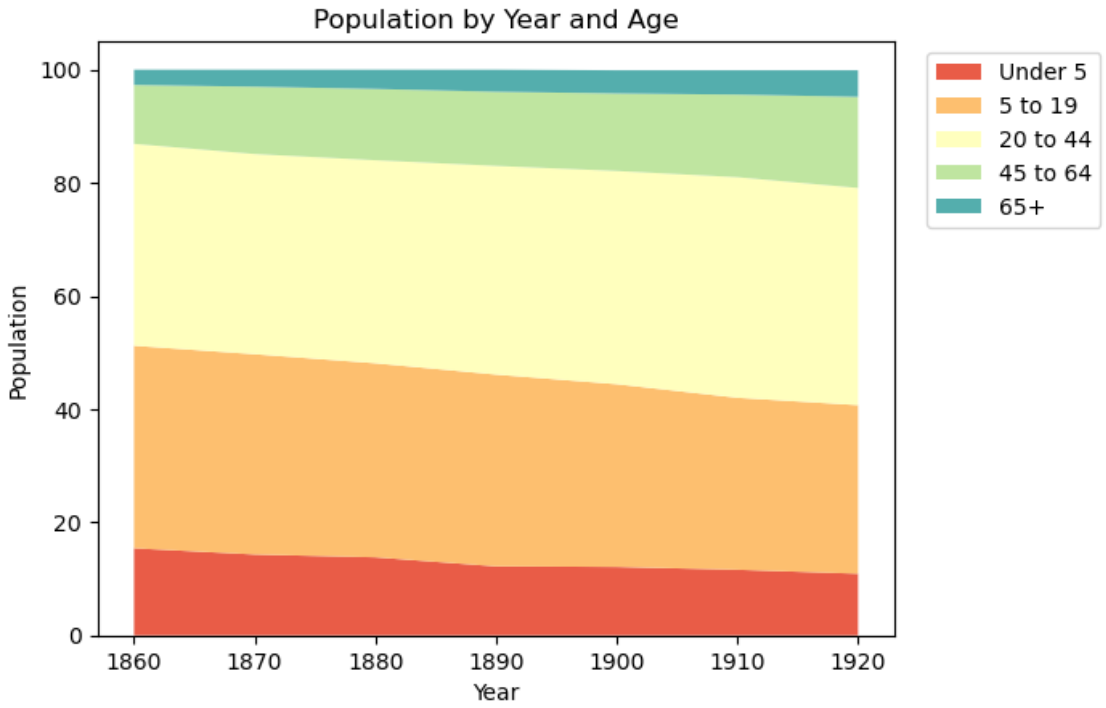
[102]:
```
# Remove extra rows
population.drop(population.tail(3).index,inplace=True)
```

[123]:
```
# Change Year to numeric
population.Year = pd.to_numeric(population.Year)
```

[145]:
```
colors = sns.color_palette("Spectral", 5)
labels=["Under 5", "5 to 19", "20 to 44", "45 to 64", "65+"]
```

[146]:
```
plt.stackplot(population.Year, population['Under 5'], population['5 to 19'],␣
↪population['20 to 44'], population['45 to 64'], population['65+'],␣
↪labels=labels, colors=colors)
plt.legend(loc = "upper center", bbox_to_anchor=(1.15, 1), ncol=1)
plt.title("Population by Year and Age")
```

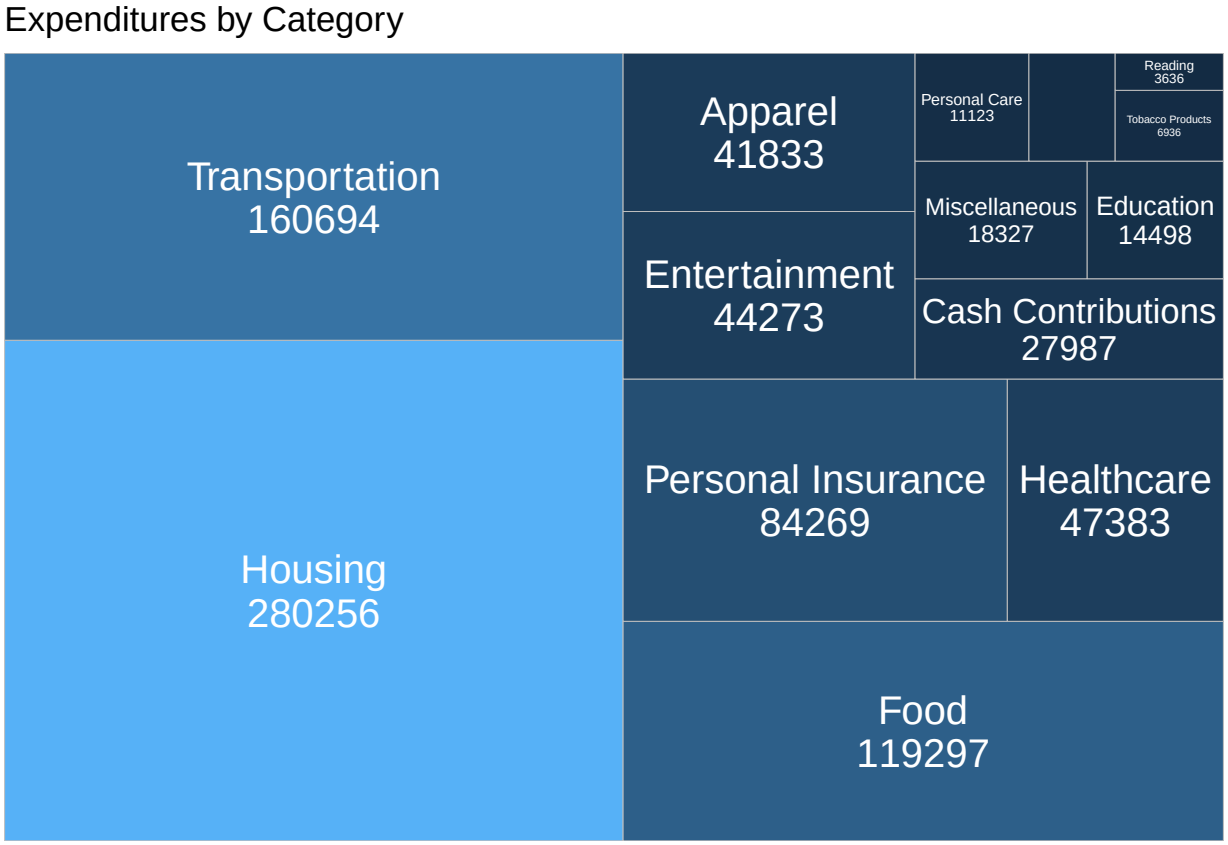
```
plt.xlabel("Year")
plt.ylabel("Population")
plt.show()
```



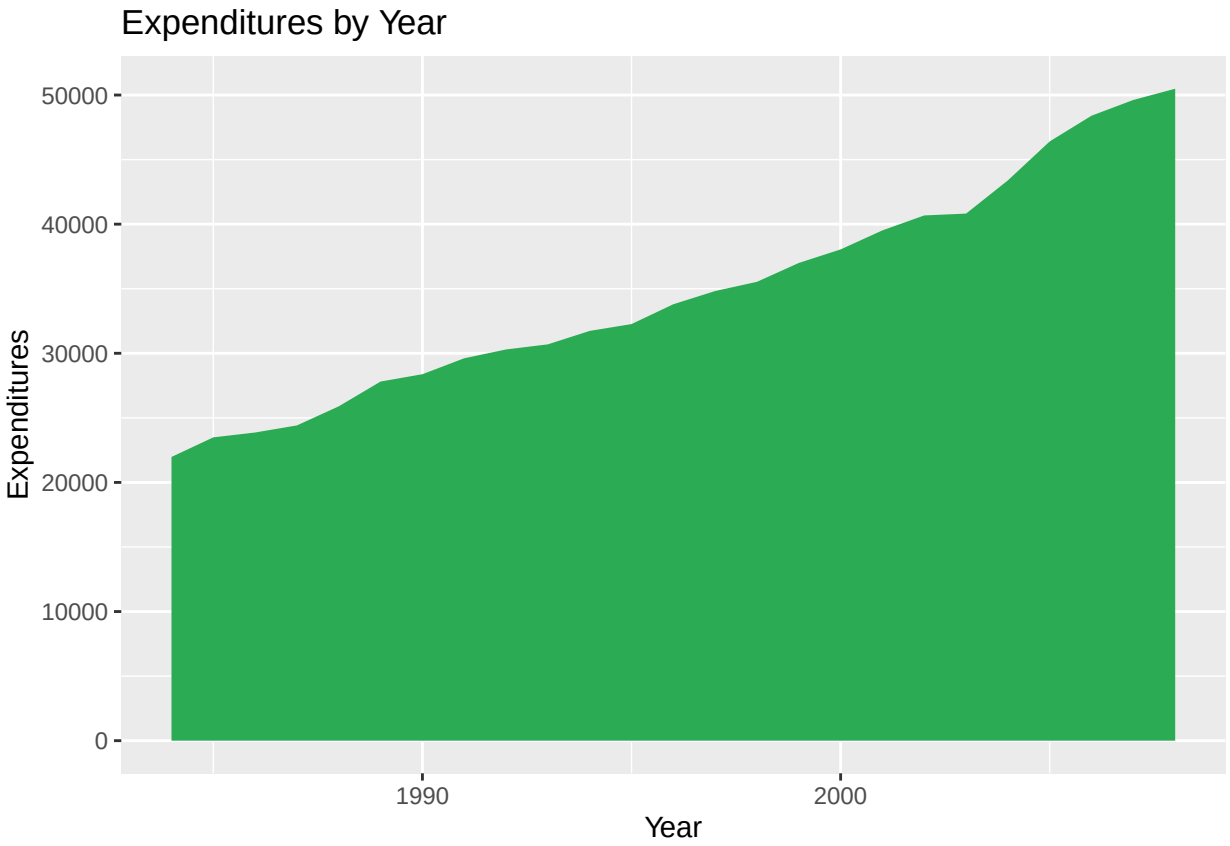
Campbell640Week5-6

2023-10-06

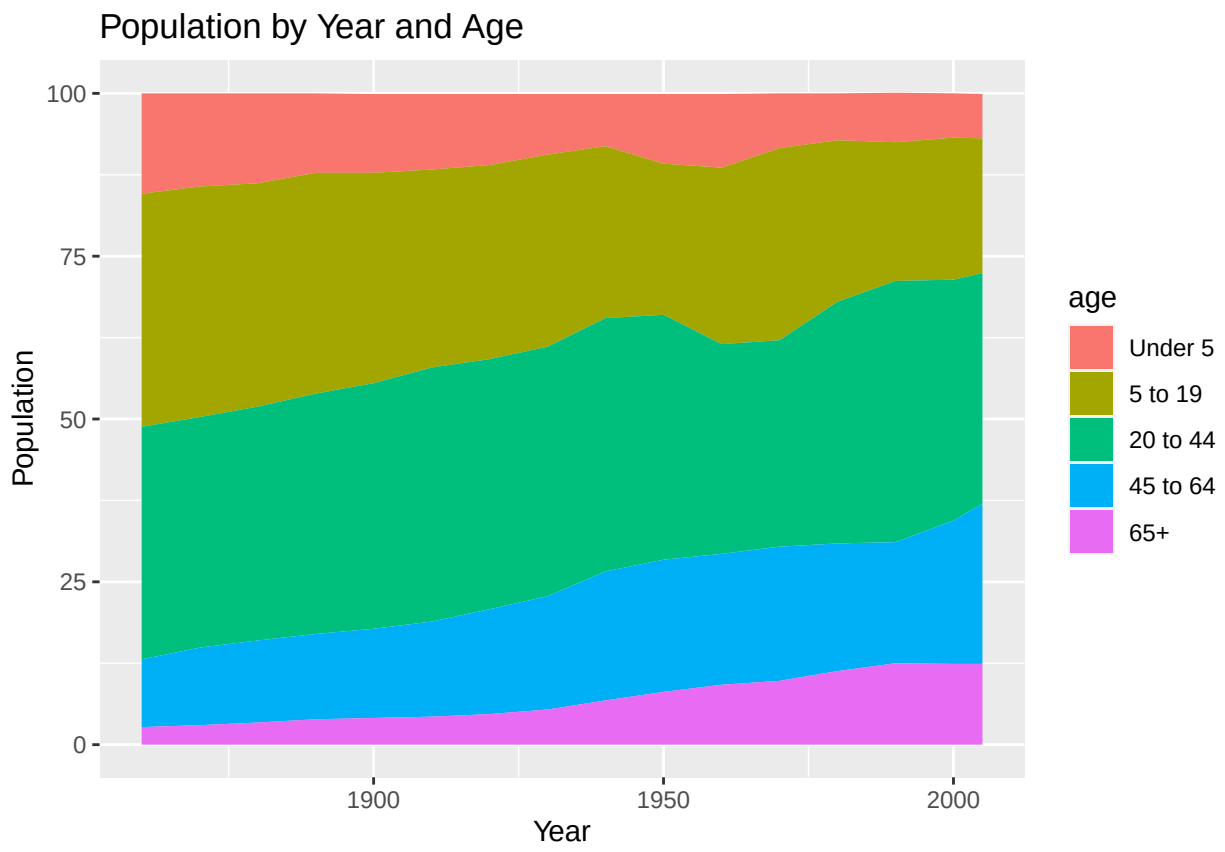
R - Treemap



R - Area Chart


Expenditures by Year
50000
40000
30000
20000
10000
0
Expenditures
1990
2000
Year

R - Stacked Area Chart



Code Repository

```
## Set the working directory to the root of your DSC 640 directory
setwd("C:/Users/jcamp/Documents/DSC640/Assignments/data sources")

# Load libraries
library(ggplot2)
library(treemapify)
library(dplyr)
library(readxl)
library(reshape2)
library(tidyr)

# Load files
expenditures_df <- read.table("expenditures.txt", header=T, sep="\t")
population_df <- read_excel("us-population-by-age.xls")
# Treemap
# Group expenditures by category
cat_expend = expenditures_df %>%
                group_by(category) %>%
                summarise(count = sum(expenditure))

# Create treemap
ggplot(cat_expend, aes(area = count, fill = count,
```

```
                      label = paste(category, count, sep = "\n"))) +
  geom_treemap() +
  geom_treemap_text(colour = "white",
                    place = "centre",
                    size = 15) +
  theme(legend.position = "none") +
  ggtitle("Expenditures by Category")
# Area Chart
# Group expenditures by year
year_expend = expenditures_df %>%
                group_by(year) %>%
                summarise(count = sum(expenditure))

# Plot Area Chart
ggplot(year_expend, aes(x=year, y=count)) +
  geom_area(fill = "#2bab53") +
  ggtitle("Expenditures by Year") +
  xlab("Year") +
  ylab("Expenditures")
# Stacked Area Chart
# Rename first column to year
population_df <- population_df %>% rename_at(1, ~'year')

# Convert year to numeric
population_df$year = suppressWarnings(as.numeric(population_df$year))

# Drop empty rows
population_df = drop_na(population_df)

# Melt to get rows with age category
age_pop = melt(population_df, id = c("year"))
age_pop <- age_pop %>% rename_at(2, ~'age')

# Plot stacked area chart
ggplot(age_pop, aes(x=year, y=value, fill=age)) +
  geom_area() +
  ggtitle("Population by Year and Age") +
  xlab("Year") +
  ylab("Population")
```