Exercises3_2_PhillipsEmily

Emily Phillips

1/19/2022

#Reading in Data

```
library(readxl)
library(tidyverse)
```

```
## -- Attaching packages --------------------------------------- tidyverse 1.3.1 --

## v ggplot2 3.3.5     v purrr   0.3.4
## v tibble  3.1.3     v dplyr   1.0.7
## v tidyr   1.1.3     v stringr 1.4.0
## v readr   2.0.0     v forcats 0.5.1

## Warning: package 'ggplot2' was built under R version 4.1.2

## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
library(Hmisc)
```

```
## Loading required package: lattice

## Loading required package: survival

## Loading required package: Formula

##
## Attaching package: 'Hmisc'

## The following objects are masked from 'package:dplyr':
##
##     src, summarize

## The following objects are masked from 'package:base':
##
##     format.pval, units
```

```
#data for expenditures excel file
expend <- read_excel("ex3-3/expenditures.xlsx")
head(expend)
```

```
## # A tibble: 6 x 4
##    year category            expenditure   sex
##   <dbl> <chr>                     <dbl> <dbl>
## 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
```

```
#describing expenditures dataframe --> getting data types
summary(expend)
```

```
##       year        category          expenditure           sex
##  Min.   :1984   Length:350         Min.   :  116.0   Min.   :1
##  1st Qu.:1990   Class :character   1st Qu.:  401.5   1st Qu.:1
##  Median :1996   Mode  :character   Median : 1225.5   Median :1
##  Mean   :1996                      Mean   : 2482.7   Mean   :1
##  3rd Qu.:2002                      3rd Qu.: 3039.0   3rd Qu.:1
##  Max.   :2008                      Max.   :17109.0   Max.   :1
```

```
#data for unemployment rate excel file
unemploy <- read_excel("ex3-3/unemployement-rate-1948-2010.xlsx")
head(unemploy)
```

```
## # A tibble: 6 x 4
##   `Series id`  Year Period Value
##   <chr>       <dbl> <chr>  <dbl>
## 1 LNS14000000  1948 M01      3.4
## 2 LNS14000000  1948 M02      3.8
## 3 LNS14000000  1948 M03      4
## 4 LNS14000000  1948 M04      3.9
## 5 LNS14000000  1948 M05      3.5
## 6 LNS14000000  1948 M06      3.6
```

```
#describing unemployment dataframe --> getting data types
summary(unemploy)
```

```
##   Series id              Year         Period              Value
##  Length:746         Min.   :1948   Length:746         Min.   : 2.500
##  Class :character   1st Qu.:1963   Class :character   1st Qu.: 4.525
##  Mode  :character   Median :1979   Mode  :character   Median : 5.500
##                     Mean   :1979                      Mean   : 5.666
##                     3rd Qu.:1994                      3rd Qu.: 6.600
##                     Max.   :2010                      Max.   :10.800
```

#Tree Maps

```r
#install.packages("treemapify")
library(treemapify)
```

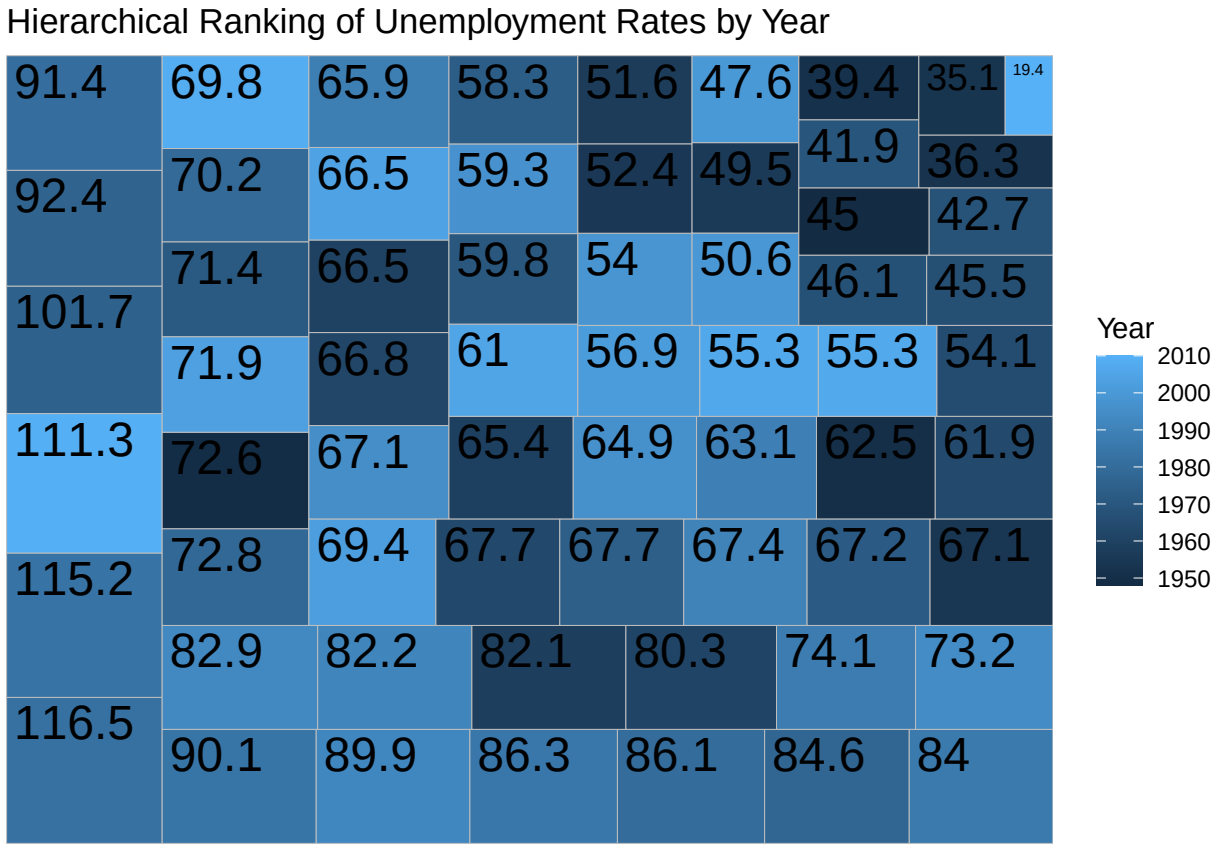
```
## Warning: package 'treemapify' was built under R version 4.1.2
```

```r
#install.packages("ggplot2")
library(ggplot2)

#aggregating values in the dataframe by sums per year
aggregated_df <- aggregate(unemploy$Value, list(unemploy$Year), FUN=sum)

#renaming columns
# Rename column where names is "Sepal.Length"
names(aggregated_df)[names(aggregated_df) == "Group.1"] <- "Year"
names(aggregated_df)[names(aggregated_df) == "x"] <- "unemployment_rate"
aggregated_df
```

```
##    Year unemployment_rate
## 1  1948              45.0
## 2  1949              72.6
## 3  1950              62.5
## 4  1951              39.4
## 5  1952              36.3
## 6  1953              35.1
## 7  1954              67.1
## 8  1955              52.4
## 9  1956              49.5
## 10 1957              51.6
## 11 1958              82.1
## 12 1959              65.4
## 13 1960              66.5
## 14 1961              80.3
## 15 1962              66.8
## 16 1963              67.7
## 17 1964              61.9
## 18 1965              54.1
## 19 1966              45.5
## 20 1967              46.1
## 21 1968              42.7
## 22 1969              41.9
## 23 1970              59.8
## 24 1971              71.4
## 25 1972              67.2
## 26 1973              58.3
## 27 1974              67.7
## 28 1975             101.7
## 29 1976              92.4
## 30 1977              84.6
## 31 1978              72.8
## 32 1979              70.2
## 33 1980              86.1
## 34 1981              91.4
```

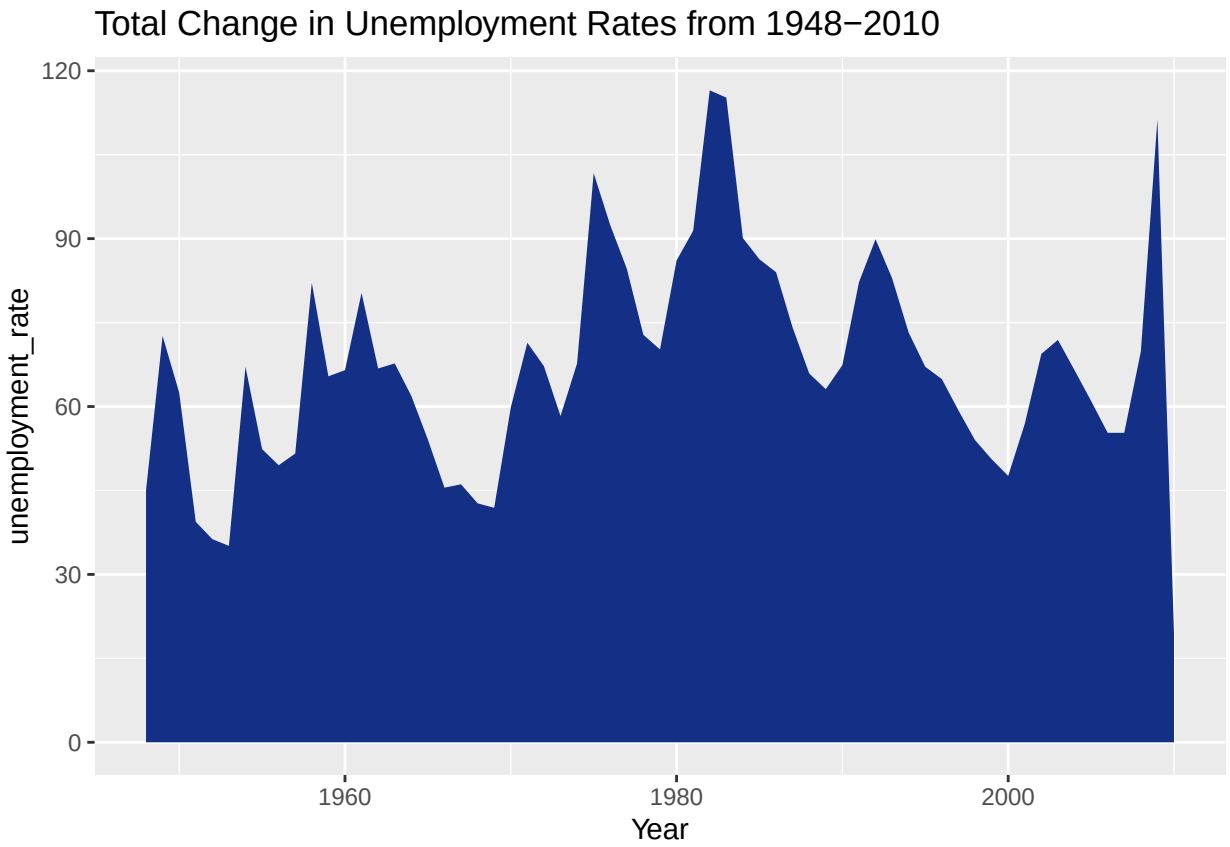
```
## 35 1982             116.5
## 36 1983             115.2
## 37 1984              90.1
## 38 1985              86.3
## 39 1986              84.0
## 40 1987              74.1
## 41 1988              65.9
## 42 1989              63.1
## 43 1990              67.4
## 44 1991              82.2
## 45 1992              89.9
## 46 1993              82.9
## 47 1994              73.2
## 48 1995              67.1
## 49 1996              64.9
## 50 1997              59.3
## 51 1998              54.0
## 52 1999              50.6
## 53 2000              47.6
## 54 2001              56.9
## 55 2002              69.4
## 56 2003              71.9
## 57 2004              66.5
## 58 2005              61.0
## 59 2006              55.3
## 60 2007              55.3
## 61 2008              69.8
## 62 2009             111.3
## 63 2010              19.4
```

```
#plotting tree map for total unemployment rates per year
#label on Year
ggplot(aggregated_df, aes(area =unemployment_rate, fill = Year,label=unemployment_rate)) +
  geom_treemap() + geom_treemap_text() + ggtitle("Hierarchical Ranking of Unemployment Rates by Year")
```



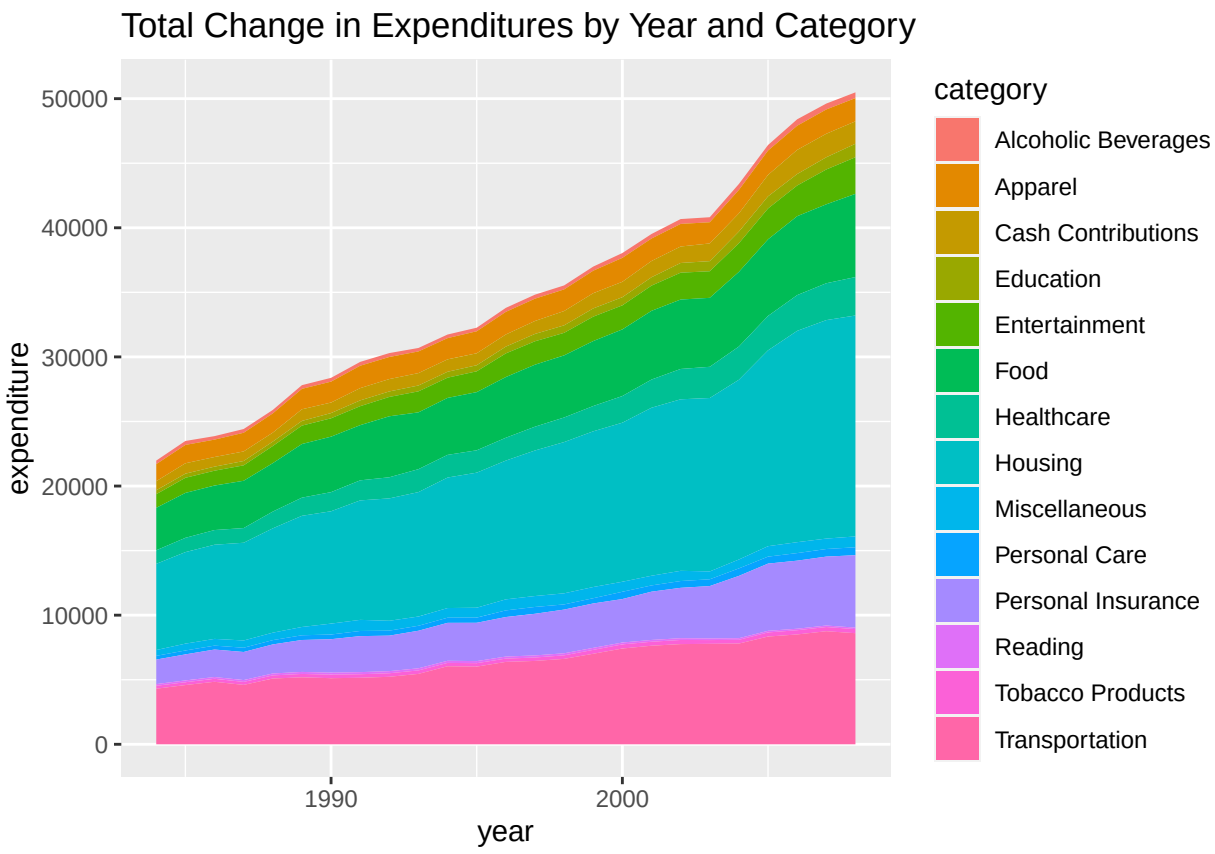
Area Chart

```
#plotting the area chart for unemployment rate vs. year
#label on Year
ggplot(aggregated_df, aes(x =Year, y = unemployment_rate)) +geom_area(fill='#142F86',alpha=2) + ggtitle
```



#Stacked Area Chart

```
#plotting the area chart for expenditure vs. year by category
#label on Year
ggplot(expend, aes(x =year, y = expenditure,fill=category)) +geom_area() + ggtitle("Total Change in Expe
```


Total Change in Expenditures by Year and Category
expenditure
50000
40000
30000
20000
10000
0
1990
2000
year
category
Alcoholic Beverages
Apparel
Cash Contributions
Education
Entertainment
Food
Healthcare
Housing
Miscellaneous
Personal Care
Personal Insurance
Reading
Tobacco Products
Transportation