Data 607 - Project 2

Shamecca Marshall

2023-10-08

The goal of this assignment is to give you practice in preparing different datasets for downstream analysis work. Your task is to: (1) Choose any three of the "wide" datasets identified in the Week 5 Discussion items. (You may use your own dataset; please don't use my Sample Post dataset, since that was used in your Week 6 assignment!) For each of the three chosen datasets: • Create a .CSV file (or optionally, a MySQL database!) that includes all of the information included in the dataset. You're encouraged to use a "wide" structure similar to how the information appears in the discussion item, so that you can practice tidying and transformations as described below. • Read the information from your .CSV file into R, and use tidyr and dplyr as needed to tidy and transform your data. [Most of your grade will be based on this step!] • Perform the analysis requested in the discussion item. • Your code should be in an R Markdown file, posted to rpubs.com, and should include narrative descriptions of your data cleanup work, analysis, and conclusions.

Load CVS file from github

```
theUrl <- "https://raw.githubusercontent.com/Meccamarshall/Data607/main/unemployment%20rate.csv"

df <- read.csv(file = theUrl, header = TRUE, sep = ",")
head(df)
```

```
##           State.Area Year Month
## 1            Alabama 1976   Jan
## 2             Alaska 1976   Jan
## 3            Arizona 1976   Jan
## 4           Arkansas 1976   Jan
## 5         California 1976   Jan
## 6 Los Angeles County 1976   Jan
##   Percent.....of.Labor.Force.Unemployed.in.State.Area
## 1                                                 6.6
## 2                                                 7.1
## 3                                                10.2
## 4                                                 7.3
## 5                                                 9.2
## 6                                                 8.9
```

Viewing column

```
colnames(df)
```

```
## [1] "State.Area"
## [2] "Year"
## [3] "Month"
## [4] "Percent.....of.Labor.Force.Unemployed.in.State.Area"
```

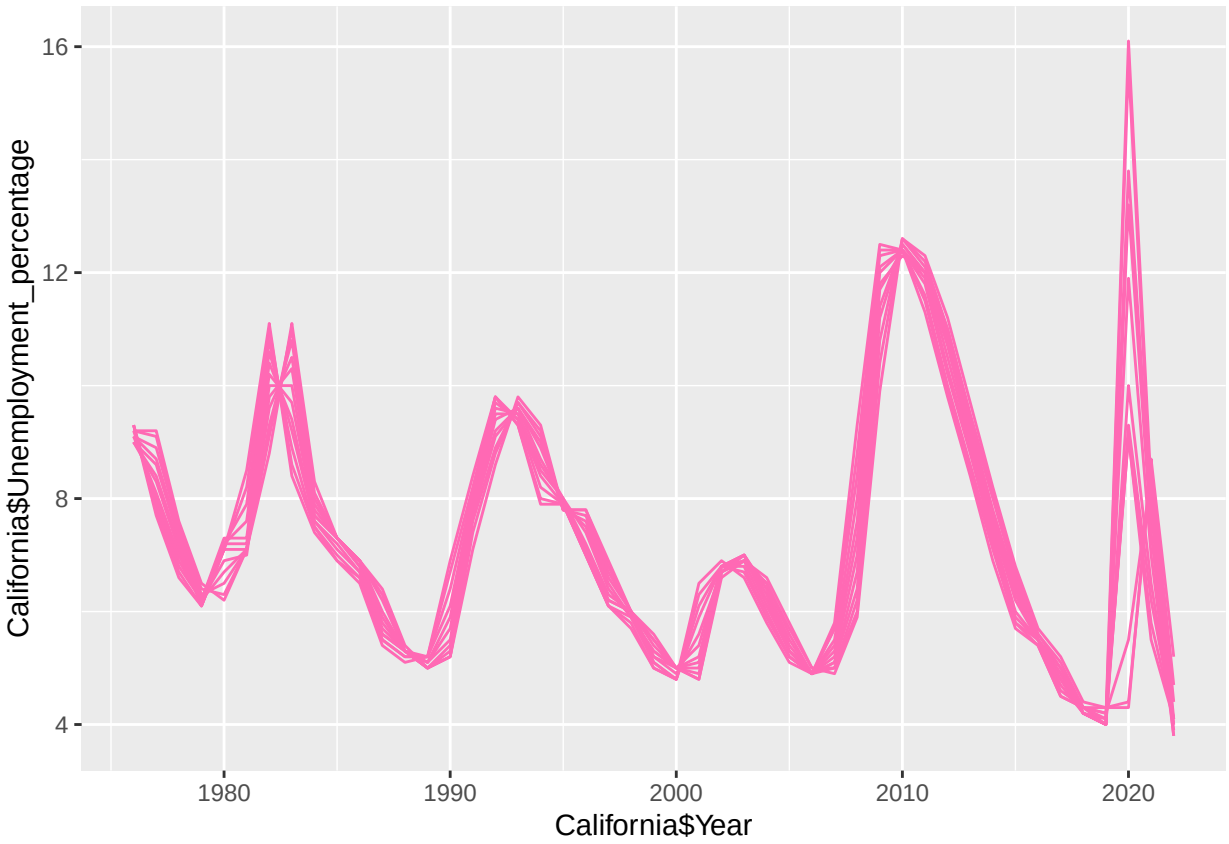
Renaming Column

```
library(tidyr)
df2<-rename(df,State=State.Area, Unemployment_percentage=Percent.....of.Labor.Force.Unemployed.in.State
head(df2)
```

```
##                State Year Month Unemployment_percentage
## 1            Alabama 1976   Jan                     6.6
## 2             Alaska 1976   Jan                     7.1
## 3            Arizona 1976   Jan                    10.2
## 4           Arkansas 1976   Jan                     7.3
## 5         California 1976   Jan                     9.2
## 6 Los Angeles County 1976   Jan                     8.9
```

Analysis of data will be performed by creating visualization of monthly unemployment rates of a single state for time from 1976-2020 The state chosen is California. Based on the chart, California has had a unemployment rate hovering between 3.8% and 16.1% between 1976 and 2020 The highest unemployment time for California was during the months of April 2020 and August 2020 during covid-19.

```
library(ggplot2)
California<-subset(df2,State=="California")
ggplot(California, aes(x=California$Year, y=California$Unemployment_percentage, fill=California$Month))
```


16
12
8
4
California$Unemployment_percentage
1980
1990
2000
2010
2020
California$Year