Exercise

1. Please analyse using visual techniques for the economics dataset available in package ggplot2.
2. Visualize characteristics of this dataset.
3. Create a bubble chart for 3 variables of this dataset.

I. Data overview

1. Summary

This is a dataframe with 574 rows and 6 variables contains information about the economic situation in the US, observations are recorded monthly from July 1967 to April 2015

```
      date                 pce               pop        
 Min.   :1967-07-01   Min.   :  506.7   Min.   :198712  
 1st Qu.:1979-06-08   1st Qu.: 1578.3   1st Qu.:224896  
 Median :1991-05-16   Median : 3936.8   Median :253060  
 Mean   :1991-05-17   Mean   : 4820.1   Mean   :257160  
 3rd Qu.:2003-04-23   3rd Qu.: 7626.3   3rd Qu.:290291  
 Max.   :2015-04-01   Max.   :12193.8   Max.   :320402  
    psavert          uempmed          unemploy     
 Min.   : 2.200   Min.   : 4.000   Min.   : 2685  
 1st Qu.: 6.400   1st Qu.: 6.000   1st Qu.: 6284  
 Median : 8.400   Median : 7.500   Median : 7494  
 Mean   : 8.567   Mean   : 8.609   Mean   : 7771  
 3rd Qu.:11.100   3rd Qu.: 9.100   3rd Qu.: 8686  
 Max.   :17.300   Max.   :25.200   Max.   :15352  
```

2. Variables

- date: Month of data collection
- pce: personal consumption expenditures, in billions of dollars
- pop: total population, in thousands
- psavert: personal savings rate
- uempmed: median duration of unemployment, in weeks
- unemploy: number of unemployed in thousands

II. Data processing

1. Create “year” column

Create an additional column “Year” to serve the purpose of building overtime reports

2. Create a yearly summary table

Create an additional table contains average personal consumption expenditures, total population, personal savings rate, median duration of unemployment and number of unemployed in thousands group by Year

Year	avg_pce	avg_pop	avg_psavert	avg_uempmed	avg_unemploy
1967	514.4667	199200.3	12.433333	4.700000	3012.333
1968	556.8417	200663.8	11.333333	4.500000	2797.417
1969	603.6500	202648.7	10.900000	4.441667	2830.167
1970	646.7250	204982.3	12.800000	4.983333	4127.333
1971	699.9250	207589.3	13.475000	6.275000	5021.667
1972	768.1500	209837.6	12.350000	6.108333	4875.833
1973	849.5833	211857.2	13.450000	5.175000	4359.333
1974	930.1500	213814.8	13.300000	5.183333	5173.333
1975	1030.5583	215890.7	13.408333	8.466667	7939.583
1976	1147.6667	217999.2	11.625000	8.141667	7398.167
1977	1273.9750	220193.1	10.658333	7.066667	6966.917
1978	1422.2500	222525.4	10.725000	5.950000	6187.083
1979	1585.4250	225002.8	10.316667	5.583333	6135.333
1980	1750.6667	227621.9	11.058333	6.658333	7670.667

3. Create “pce_per_capita” column

Create an additional column:

pce_per_capita = pce / pop

to calculate average personal expenditure per capita to analyze personal expenditure relative to population size.

4. Create “unemployment_rate” column

Create an additional column:

unemployment_rate = unemploy / pop * 100

to calculate unemployment rate.

5. **Calculate correlation matrix between variables**

	pce	pop	psavert	uempmed	unemploy
pce	1.0000000	0.9872421	-0.7928546	0.7269616	0.6145176
pop	0.9872421	1.0000000	-0.8363147	0.6950085	0.6337165
psavert	-0.7928546	-0.8363147	1.0000000	-0.3251377	-0.3093769
uempmed	0.7269616	0.6950085	-0.3251377	1.0000000	0.8693097
unemploy	0.6145176	0.6337165	-0.3093769	0.8693097	1.0000000

III. Data Analysis

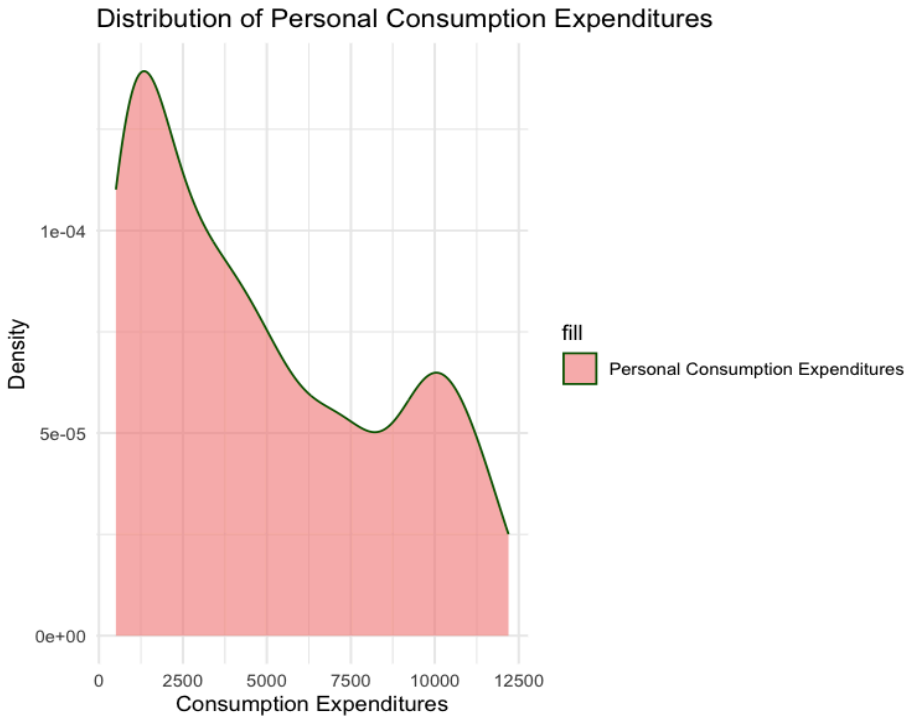
1. Distribution of continuous variables

Some variables in the dataset such as personal saving rates and consumption expenditures are continuous variables, the graphs below will help to understand the distribution of these continuous variables:

Personal savings rates

This variable is not normally distributed, specifically, it has a bimodal distribution, the first peak shows that the majority of Americans have a saving rate of 5.5% - 6%.

The second peak, lower than the first peak, shows that people with a saving rate of 10% - 11% also make up a large portion.

Personal Consumption Expenditures The pce distribution is not normal but slightly right skewed with the first peak being in the range of 0-2500 billion dollars. Interestingly, the chart shows another peak at around 9000-11000 billion dollars, but only half as high as the first peak	

Assumption

Personal Saving Rates:

- The existence of two peaks suggests that there are two distinct groups of saving behavior:
  - The larger group has a savings rate of 5.5%-6%: This may be the “stable” savings level common to the majority of the population, reflecting the minimum savings level that most Americans consider necessary to cope with future spending.
  - The smaller group has a savings rate of 10%-11%: These people tend to save more, which may be related to higher income levels, long-term planning, or economic anxiety.

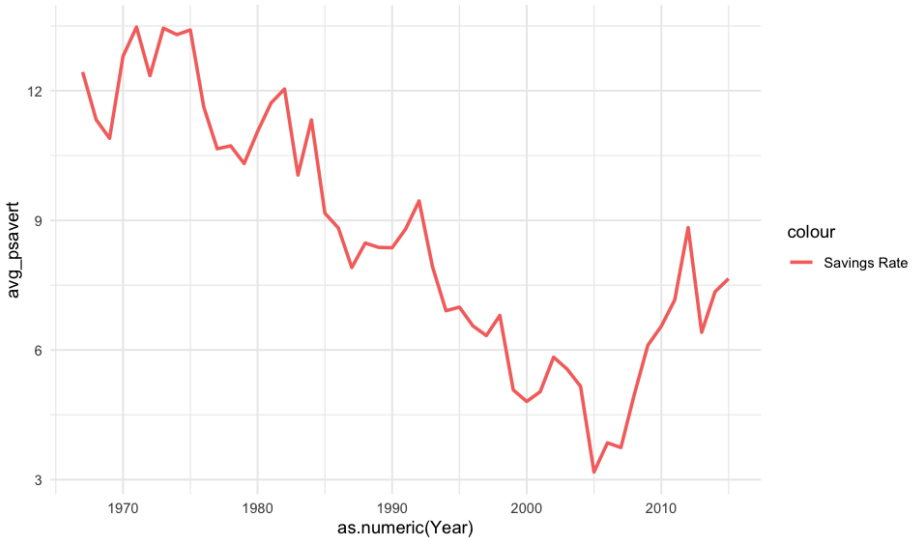
Personal Consumption Expenditures:

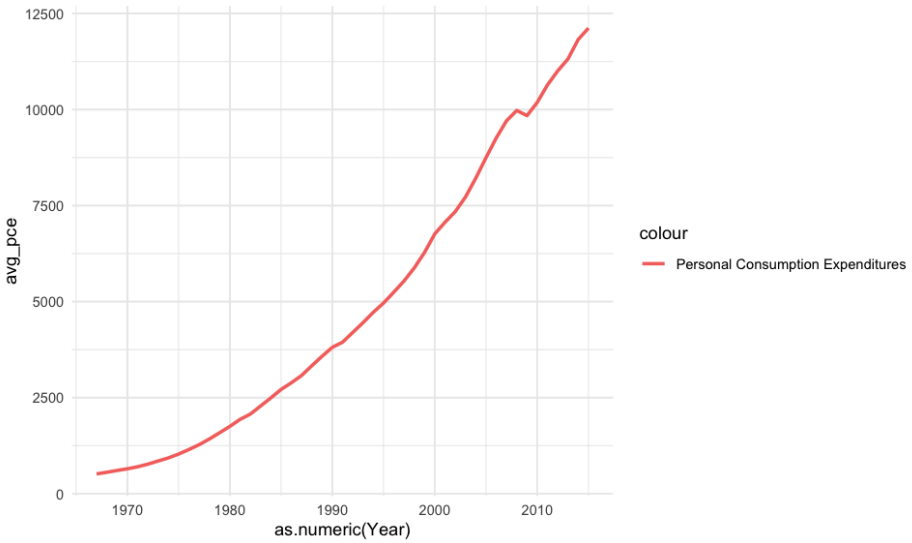
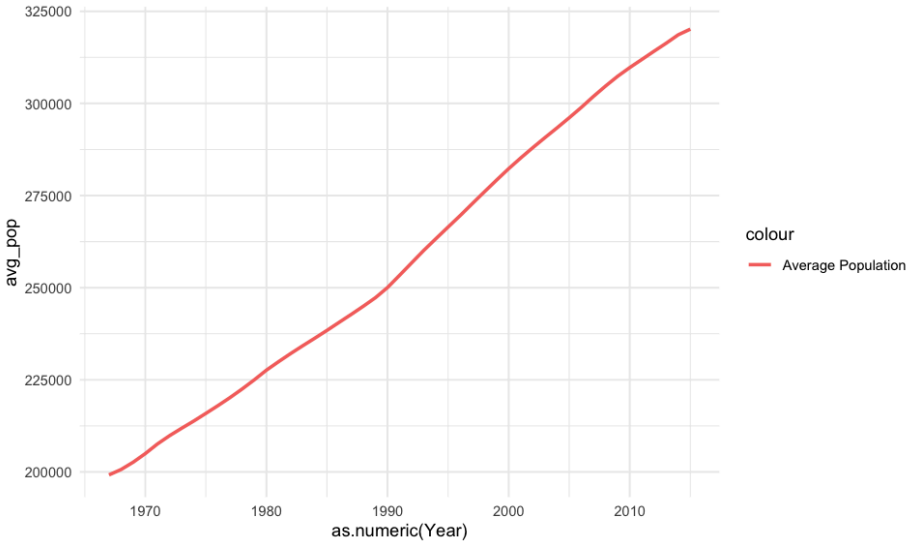
- Right-skewed distribution:
  - The majority of PCE values are in the range of $0-2.5 trillion, indicating that most personal spending is still focused on everyday and basic consumption expenditures.

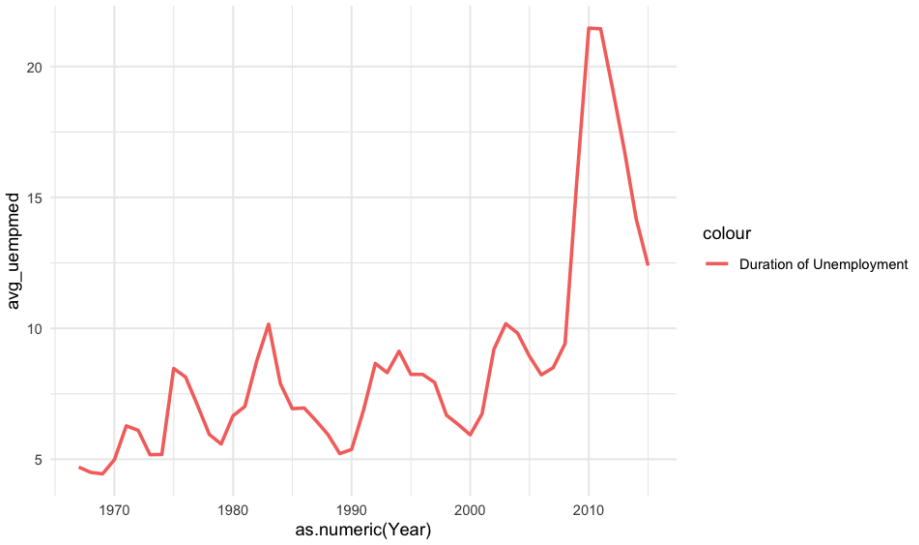
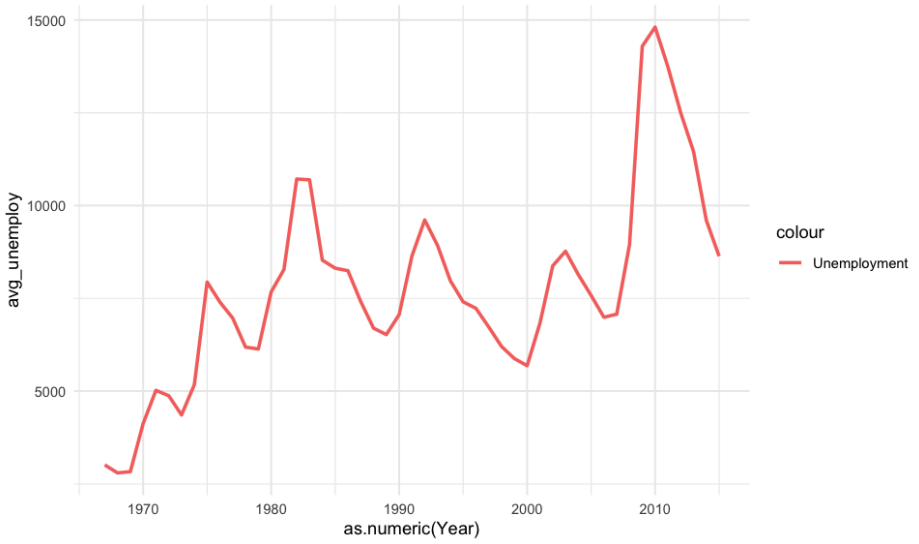
- Second peak ($9 trillion-11 trillion): Although less common, the presence of this peak indicates that a small group of the population has very high personal spending.

2. Trends of variables over time

The charts below show the changes and trends of the average of the variables over time.

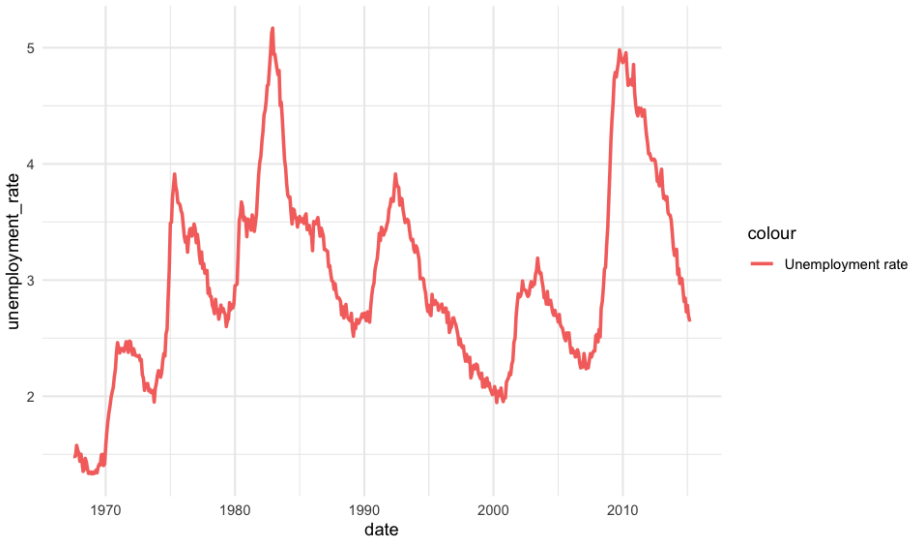
Average of Saving Rate In general, it can be seen that the average saving rate of people tends to go down, hitting bottom in 2005. After 2005, the average saving rate in the US increased again but was not as high as compared to the period of 1970-1980.	

Average of Personal Consumption Expenditures The trend of average consumption expenditures in the US during this period increased steadily over the years.	
Population Similarly, the average population in the US during this period also increased steadily over the years.	

Duration of unemployment In general, years of recorded unemployment in the US range from 5 to 10 weeks. Notably, 2010 recorded a record high average number of weeks of unemployment of more than 20 weeks.	
Number of unemployment The trend of this chart is quite similar to the chart of average unemployment duration. This is also relatively easy to understand.	

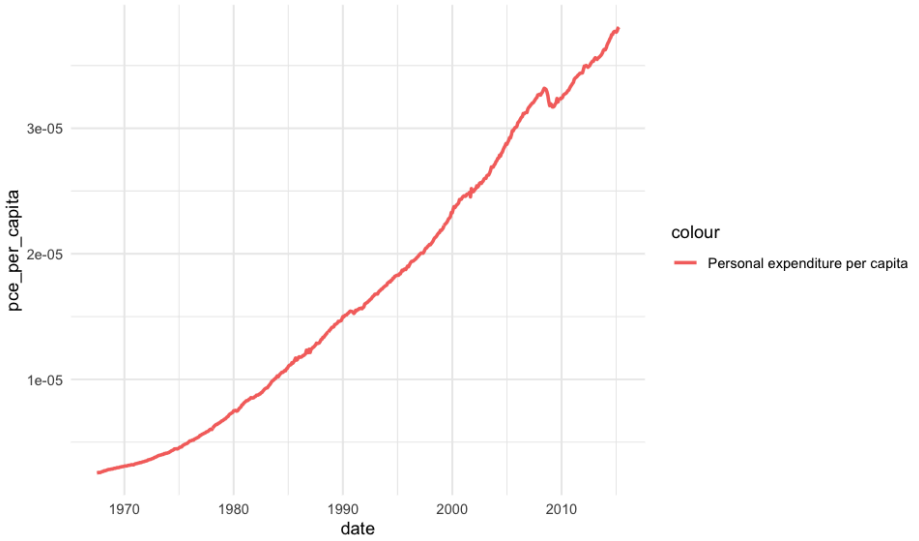
Unemployment Rate

The trend of this chart is quite similar to the chart of average unemployment duration. This is also easy to understand.



Pce per capita

Just like the charts on population and consumption of people in the US, the chart on consumption per capita also tends to increase steadily over time.



Assumption

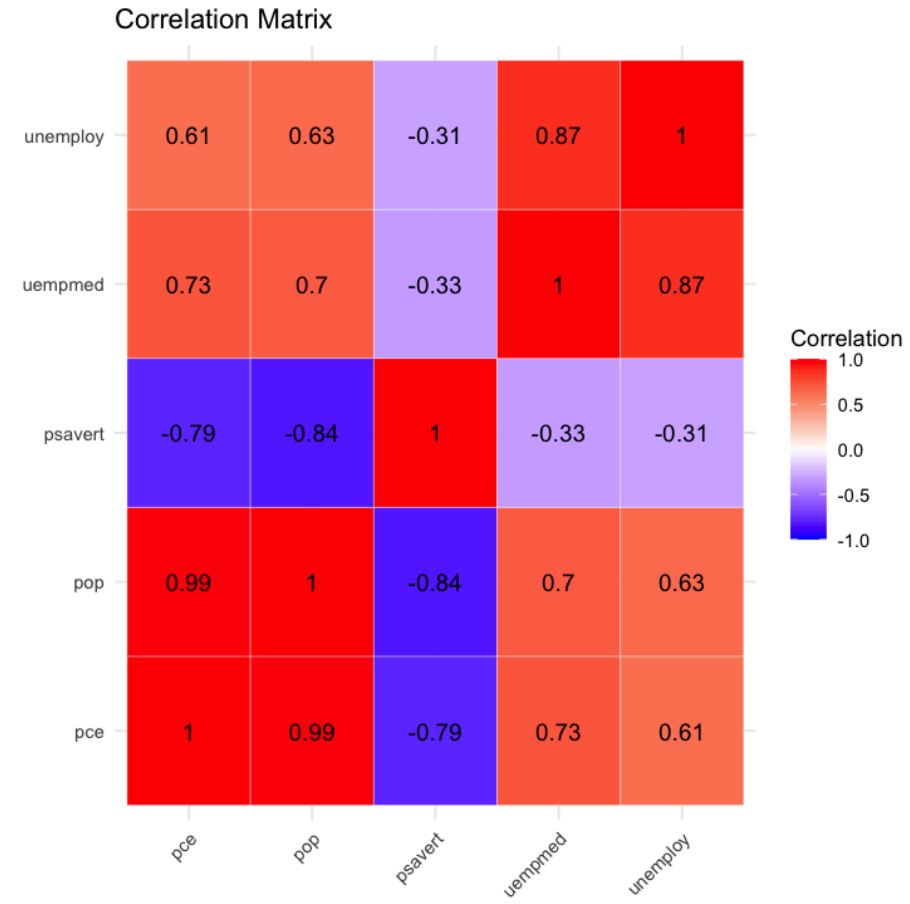
- **Average Saving Rate:** The average saving rate of Americans shows a downward trend, hitting a bottom in 2005. Afterward, it increased but did not reach the high levels observed during the 1970s–1980s. This could reflect:
    - Before 1980: Savings were prioritized, likely due to economic concerns or policies favoring savings.
    - 1980–2005: A decline, potentially due to increased consumer spending, easier access to credit, or confidence in financial markets.
    - Post-2005: A slight recovery, likely influenced by the 2008 financial crisis,

which heightened financial caution.

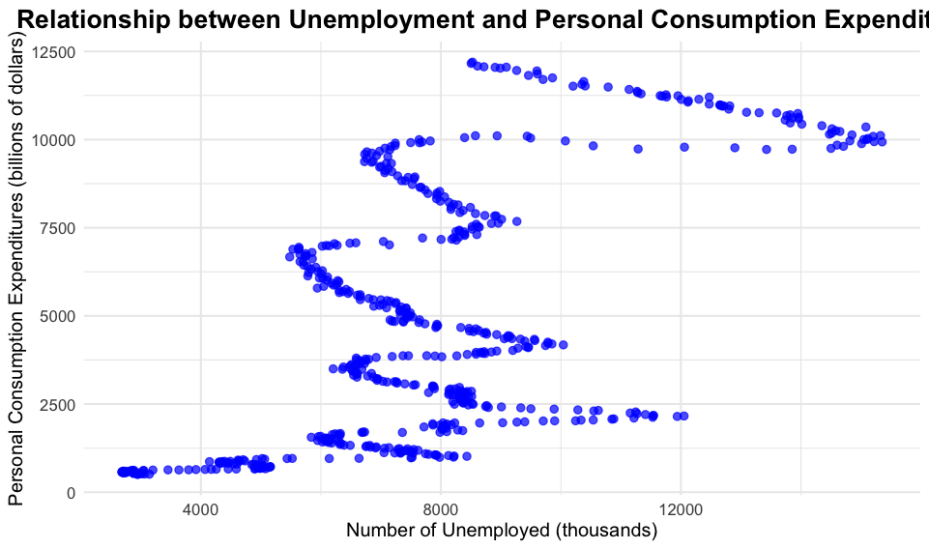
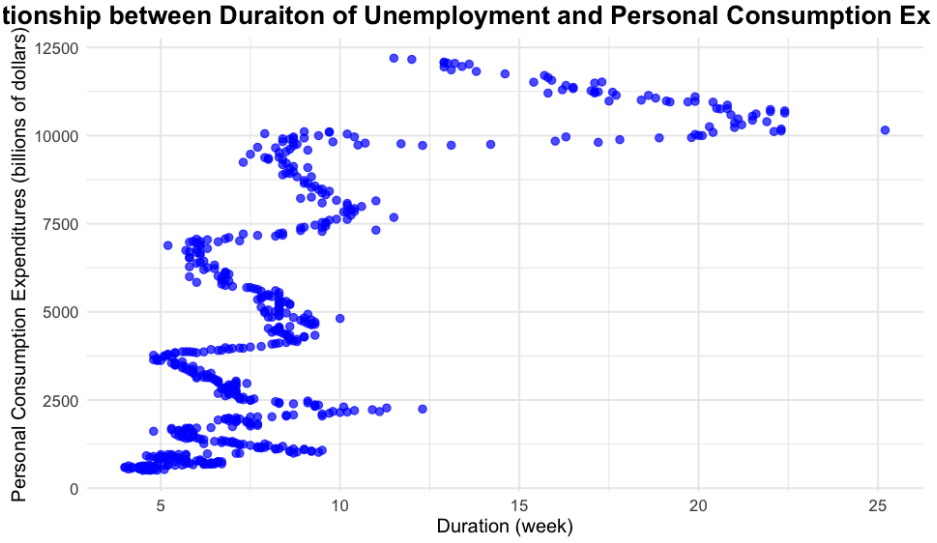
- **Average Personal Consumption Expenditures:** Personal consumption expenditures steadily increased over the years, indicating the increasing of purchasing power, which means people were able to spend more, likely due to rising incomes and stable inflation
- **Duration of Unemployment and number of Unemployment:** The peak of these charts are in 2010. This could reflects the 2008 financial crisis when severe effects caused prolonged job searches.

3. The relationship between variables

Based on these correlation coefficients, it can be seen that some variables are highly correlated with each other, some things are quite easy to explain. However, there are also some interesting points such as the positive correlation coefficient of 0.73 between unemployment duration and people's consumption, and the positive correlation coefficient of 0.61 between the number of unemployed people and consumption.



The scatter plots below will delve deeper into the relationships between the interesting features that can be seen above.

Unemployment vs PCE In general, it is true that the higher the number of unemployed people, the higher the PCE.	
Duration of Unemployment vs PCE This chart tends to be quite similar to the chart above.	

Assumption

- The positive correlation between unemployment and personal consumption may highlight the impact of government policies, essential needs, and inflationary pressures.

- Even as unemployment rises, consumption may remain stable or increase due to financial assistance programs, behavioral responses to uncertainty, and rising living costs.

The bubble chart below illustrates the relationship between unemployment, personal spending, and population.

This graph is similar to the scatter plot showing the relationship between PCE and unemployment above, but with the addition of colors representing population. Through the graph, we can see the trends of all three variables together.