Unemployment and GDP in Germany and France

Stephan.Huber@hs-fresenius.de

Date: 02.11.2022; Study Programme: THEM_bac; Semester: 4

Please answer all (!) questions in an R script. Normal text should be written as comments, using the '#' to comment out text. Make sure the script runs without errors before submitting it. Each task (starting with 1) is worth five points. You have a total of 120 minutes of editing time. Please do not forget to number your answers.

When you are done with your work, save the R script, export the script to pdf format and upload the pdf file.

Suppose you aim to empirically examine unemployment and GDP for Germany and France. The data set that we use in the following is 'forest.Rdata' and should already been known to you from the lecture.

(0) Write down your name, matriculation number, and date.

(1) Set your working directory.

(2) Clear your global environment.

(3) Install and load the following packages: 'tidyverse', 'sjPlot', and 'ggpubr'

(4) Download and load the data, respectively, with the following code:

```
load(url("https://github.com/hubchev/courses/raw/main/dta/forest.Rdata"))
```

If that is not working, you can also download the data from ILIAS, save it in your working directory and load it from there with:

```
load("forest.Rdata")
```

(5) Show the **first eight** observations of the dataset 'df'.

(6) Show the **last observation** of the dataset 'df'.

(7) Which type of data do we have here (Panel, cross-section,time series, ...)? Name the variable(s) that are necessary to identify the observations in the dataset.

(8) Explain what the **assignment operator** in R is and what it is good for.

(9) Write down the R code to store the number of observations and the number of variables that are in the dataset 'df'. Name the object in which you store these numbers 'observations_df'.

(10) In the dataset 'df', rename the variable 'country.x' to 'nation' and the variable 'date' to 'year'.

(11) Explain what the **pipe operator** in R is and what it is good for.

(12) For the upcoming analysis you are only interested the following **variables** that are part of the dataframe 'df': nation, year, gdp, pop, gdppc, and unemployment. Drop all other variables from the dataframe 'df'.

(13) Create a variable that indicates the GDP per capita ('gdp' divided by 'pop'). Name the variable 'gdp_pc'. (Hint: If you fail here, use the variable 'gdppc' which is already in the dataset as a replacement for 'gdp_pc' in the following tasks.)

(14) For the upcoming analysis you are only interested the following **countries** that are part of the dataframe 'df': Germany and France. Drop all other countries from the

dataframe 'df'.

(15) Create a table showing the **average** unemployment rate and GDP per capita for Germany and France in the given years. Use the pipe operator. (Hint: See below for how your results should look like.)

```
## # A tibble: 2 x 3
##   nation  `mean(unemployment)` `mean(gdppc)`
##   <chr>                  <dbl>         <dbl>
## 1 France                  9.75        34356.
## 2 Germany                 7.22        36739.
```

(16) Create a table showing the unemployment rate and GDP per capita for Germany and France in the **year 2020**. Use the pipe operator. (Hint: See below for how your results should look like.)

```
## # A tibble: 2 x 3
##   nation  `mean(unemployment)` `mean(gdppc)`
##   <chr>                  <dbl>         <dbl>
## 1 France                  8.01        35786.
## 2 Germany                 3.81        41315.
```

(17) Create a table showing the **highest** unemployment rate and the **highest** GDP per capita for Germany and France during the given period. Use the pipe operator. (Hint: See below for how your results should look like.)

```
## # A tibble: 2 x 3
##   nation  `max(unemployment)` `max(gdppc)`
##   <chr>                 <dbl>        <dbl>
## 1 France                 12.6       38912.
## 2 Germany                11.2       43329.
```

(18) Calculate the standard deviation of the unemployment rate and GDP per capita for Germany and France in the given years. (Hint: See below for how your result should look like.)

```
## # A tibble: 2 x 3
##   nation  `sd(gdppc)` `sd(unemployment)`
##   <chr>         <dbl>              <dbl>
## 1 France        2940.               1.58
## 2 Germany       4015.               2.37
```

(19) In statistics, the coefficient of variation (COV) is a standardized measure of dispersion. It is defined as the ratio of the standard deviation (σ) to the mean (μ): $COV = \frac{\sigma}{\mu}$. Write down the R code to calculate the coefficient of variation (COV) for the **unemployment rate** in Germany and France. (Hint: See below for what your result should should look like.)

```
## # A tibble: 2 x 4
##   nation  `sd(unemployment)` `mean(unemployment)`   cov
##   <chr>                <dbl>                <dbl> <dbl>
## 1 France                1.58                 9.75 0.162
## 2 Germany               2.37                 7.22 0.328
```

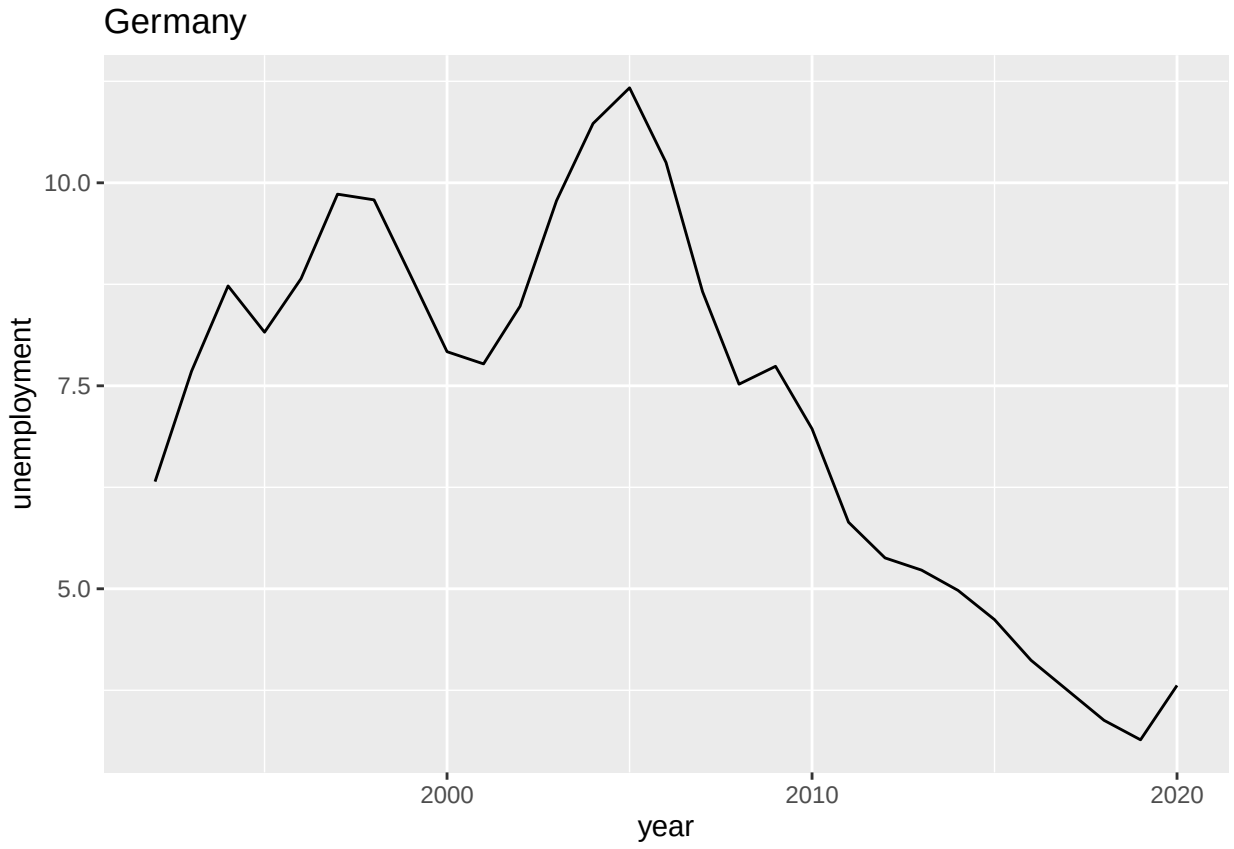
(20) Write down the R code to calculate the coefficient of variation (COV) for the **GDP per capita** in Germany and France. (Hint: See below for what your result should should look like.)

look like.)

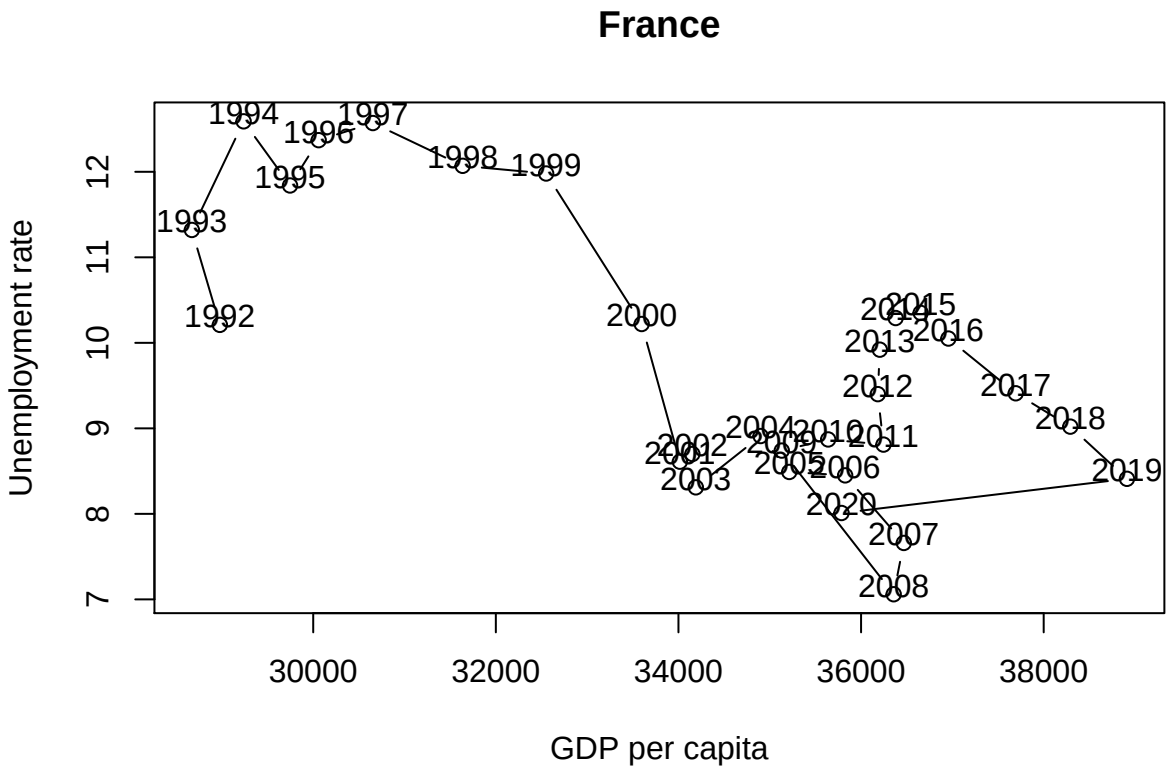
```
## # A tibble: 2 x 4
##   nation  `sd(gdppc)` `mean(gdppc)`    cov
##   <chr>         <dbl>         <dbl>  <dbl>
```

```
## 1 France        2940.        34356. 0.0856
## 2 Germany       4015.        36739. 0.109
```

(21) Create a chart (bar chart, line chart, or scatter plot) that shows the unemployment rate of **Germany** over the available years. Label the chart 'Germany' with 'ggtitle("Germany")'. Please note that you may choose any type of graphical representation. (Hint: Below you can see one of many possible examples of what your result may look like).



(22) and 23. *(This task is worth 10 points)* The following chart shows the simultaneous development of the unemployment rate and GDP per capita over time for **France**.

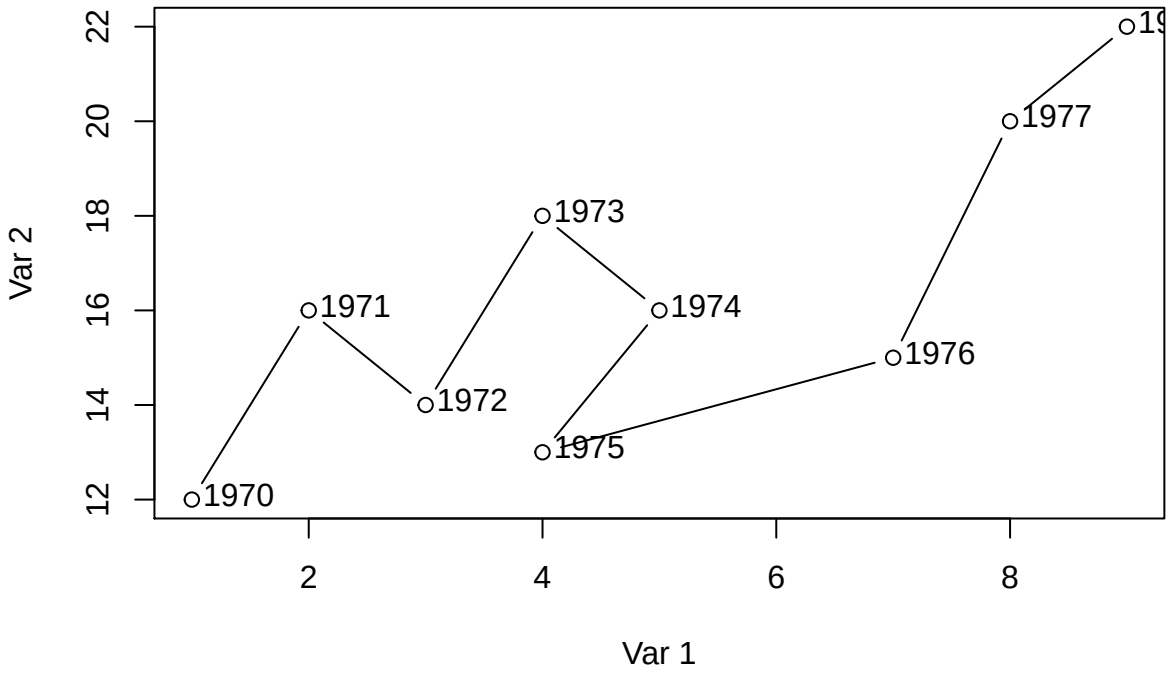


Suppose you want to visualize the simultaneous evolution of the unemployment rate and GDP per capita over time for Germany as well.

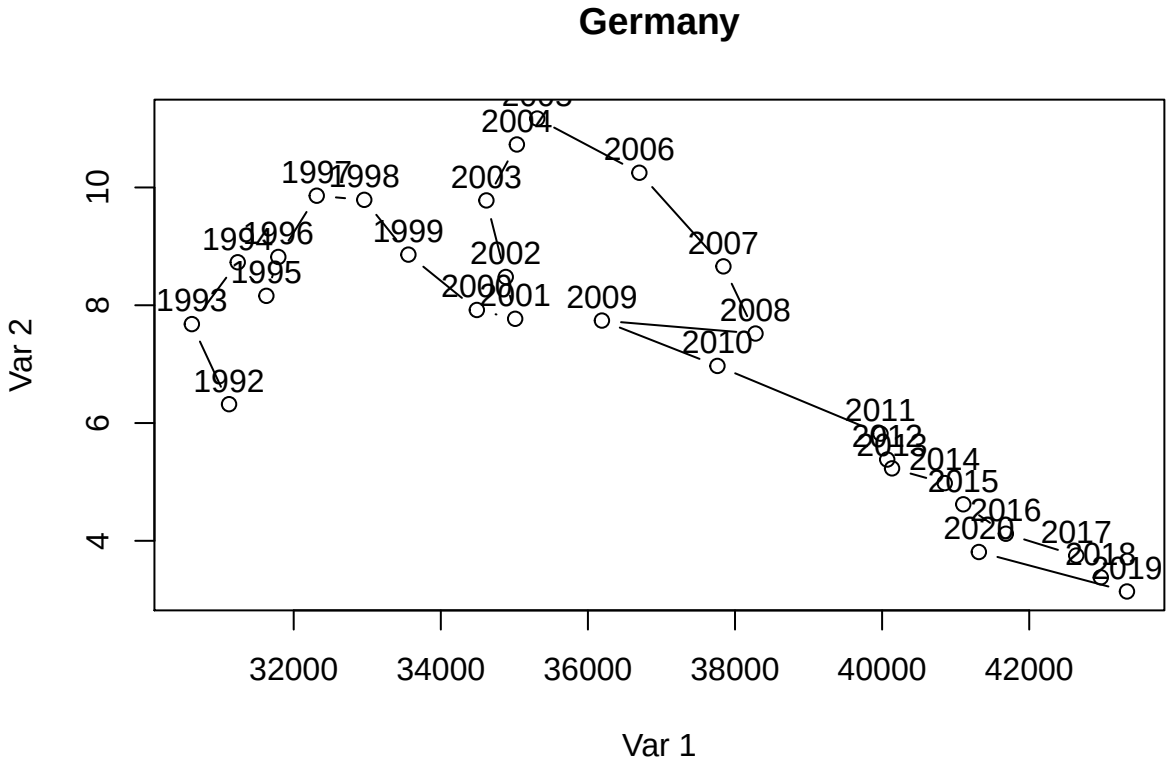
Suppose further that you have found the following lines of code that create the kind of chart you are looking for.

```
# Data
x <- c(1, 2, 3, 4, 5, 4, 7, 8, 9)
y <- c(12, 16, 14, 18, 16, 13, 15, 20, 22)
labels <- 1970:1978


# Connected scatter plot with text
plot(x, y, type = "b", xlab = "Var 1", ylab = "Var 2"); text(x + 0.4, y + 0.1, labels)
```



Use these lines of code and customize them to create the co-movement visualization for **Germany** using the available 'df' data. The result should look something like this:



(24) Interpret the two graphs above, which show the simultaneous evolution of the unem-

ployment rate and GDP per capita over time for Germany and France. What are your expectations regarding the correlation between the unemployment rate and GDP per capita variables? Can you see this expectation in the figures? Discuss.