LabBook for the LPS 2018/2 Final Project

Joao Pedro Oliveira

11/9/2018

My LabBook for the LPS 2018/2 Final Project.

This is an analysis of the “Millenium of Macroeconomic Data” dataset, gathered by the Bank of England.

First, only loading the necessary packages for this analysis. I chose to use readxl instead of the famous “xlsx” since it already comes with tidyverse, and so makes life a little easier.

I’ll try to awnser some questions with this data, but first let’s transform the messy data (very messy data) found in the xlsx archive and transform it into tidy data that’s good to analyse.

For this analysis i’ll only extract the “Headline series” sheet from the Excel file, since it’s the most relevant one and, as described in the documentation: “They are intended for users who wish a set of macroeconomic series without breaks for use in appropriate econometric work”. That is just what we’re trying to do here!

```
#Downloading the file, if it doesn't already exist
file = "millenniumofdata_v3_final.xlsx"
if(!file.exists(file)){
  download.file("https://www.bankofengland.co.uk/-/media/boe/files/statistics/research-datasets/a-millennium-of-macroeconomic-data-for-the-uk.xlsx?la=en&hash=73ABBFB603A709FEEB1FD349B1C61F11527F1DE4", destfile=file)
}

#Reading the xlsx file
uk_dataxl <-  read_excel(file, sheet="A1. Headline series")

#Removing useless rows
uk_dataxl_tidy <- uk_dataxl[-c(1,2,4,5,6),]

#Making the "Description" row, the header for the Dataframe
names(uk_dataxl_tidy) <- uk_dataxl_tidy[1,]

#Removing the first row beacuse it just turned into the header
uk_dataxl_tidy <- uk_dataxl_tidy[-c(1),]

#Removing NA's. This limits the data to all the years since 1929
uk_dataxl_tidy <- na.omit(uk_dataxl_tidy)

#Removing all the columns with no headers (or that only show changes in percentages from the past year). Since these columns appear in a random way through the dataset, I removed them mannualy.
uk_dataxl_tidy <- uk_dataxl_tidy[,-c(3,5,7,9,11,13, 27, 40, 55, 62, 64, 66, 68,69, 73,75,74,77)]
uk_dataxl_tidy <- uk_dataxl_tidy[,-c(26, 38, 52, 56, 58, 61, 63, 65)]
uk_dataxl_tidy <- uk_dataxl_tidy[,-c(17)]

#Transforming all the columns on the dataframe to Numeric values, as oposed to Chr
uk_dataxl_tidy[] <- lapply(uk_dataxl_tidy, function(x) {
    as.numeric(x)
})

#Renaming columns
uk_dataxl_tidy <- rename(uk_dataxl_tidy, c("Description" = "Year", "Population (GB+NI)" = "Population"))
uk_dataxl_tidy
```

```
## # A tibble: 88 x 56
##     Year `Real GDP of En~ `Real GDP of En~ `Real UK GDP at~
##    <dbl>            <dbl>            <dbl>            <dbl>
##  1  1929          215534.          182494.          245205.
##  2  1930          213608.          180903.          243254.
##  3  1931          202987.          171948.          231969.
##  4  1932          203872.          172808.          232128.
##  5  1933          210588.          178615.          239510.
##  6  1934          223656.          189820.          253801.
##  7  1935          231881.          196928.          263187.
##  8  1936          243283.          206743.          275737.
##  9  1937          251717.          214047.          285387.
## 10  1938          253383.          215602.          287602.
## # ... with 78 more rows, and 52 more variables: `Real UK GDP at factor
## #   cost, geographically-consistent estimate based on post-1922
## #   borders` <dbl>, `Index of real UK GDP at factor cost - based on
## #   changing political boundaries,` <dbl>, `Composite estimate of English
## #   and (geographically-consistent) UK real GDP at factor cost` <dbl>,
## #   `HP-filter of log of real composite estimate of English and UK real
## #   GDP at factor cost` <dbl>, `Real UK gross disposable national income
## #   at market prices, constant border estimate` <dbl>, `Real
## #   consumption` <dbl>, `Real investment` <dbl>, `Stockbuilding
## #   contribution` <dbl>, `Real government consumption of goods and
## #   services` <dbl>, `Export volumes` <dbl>, `Import volumes` <dbl>,
## #   `Nominal GDP of England at market prices` <dbl>, `Nominal UK GDP at
## #   market prices` <dbl>, Population <dbl>, `Population (England)` <dbl>,
## #   `Unemployment rate` <dbl>, `Average weekly hours worked` <dbl>,
## #   `Capital Services, whole economy` <dbl>, `TFP growth` <dbl>, `Labour
## #   productivity` <dbl>, `Labour share, whole economy excluding
## #   rents` <dbl>, `GDP deflator at market prices` <dbl>, `Export
## #   prices` <dbl>, `Import prices` <dbl>, `Terms of Trade` <dbl>, `$ Oil
## #   prices` <dbl>, `Consumer price index` <dbl>, `Consumer price
## #   inflation` <dbl>, `Real consumption wages` <dbl>, `Wholesale/producer
## #   price index` <dbl>, `Bank Rate` <dbl>, `10 year/medium-term government
## #   bond yields` <dbl>, `Consols / long-term government bond
## #   yields` <dbl>, `Mortgage rates` <dbl>, `Corporate borrowing rate from
## #   banks` <dbl>, `Corporate bond yields` <dbl>, `Share prices` <dbl>,
## #   `$/\u00a3 exchange rate` <dbl>, `Real $/\u00a3 exchange rate` <dbl>,
## #   `Real ERI` <dbl>, `House price index` <dbl>, Credit <dbl>, `Secured
## #   credit` <dbl>, `Bank of England Balance sheet` <dbl>, `Notes and coin
## #   in circulation` <dbl>, M1 <dbl>, `Public sector Total Managed
## #   Expenditure` <dbl>, `Public Sector Net Lending(+)/Borrowing(-)` <dbl>,
## #   `Central Government Gross Debt` <dbl>, `Trade deficit` <dbl>, `Current
## #   account` <dbl>, `Current account deficit including estimated
## #   non-monetary bullion flows` <dbl>
```

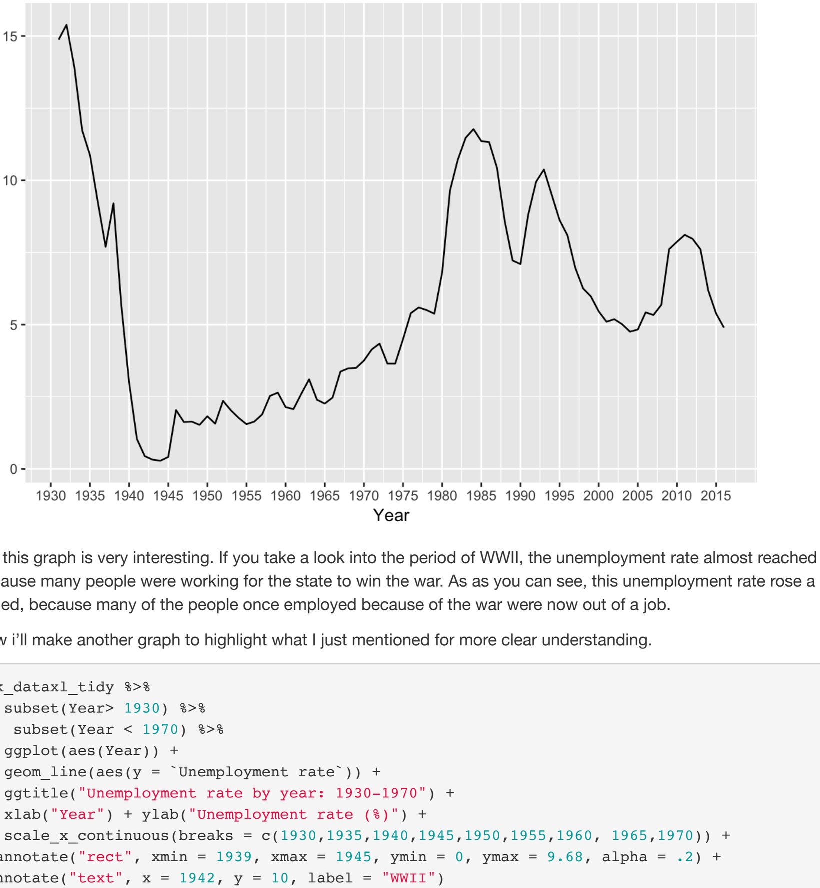
The question that I’m trying to awnser with this dataset is: Can we spot the effect of significant historical moments on the data? (Example: the Industrial Revolution, WWI, WWII, and the Great Recession)

To awnser that question, I figured we need to find and compare some indicators that might give us our awnser. For example, the Unemployment rate is a good indicator to spot a time of crisis.

So, I figured that there’s a lot of columns here (56!). Some of them really don’t matter to the things that I’m trying to figure out, but I’ll leave them there in the dataset by now so that I can have more options to analyse in the future if I need to.

Let’s look at the unemployment rate since 1930. This might be a good indicator to find some important historical moments.

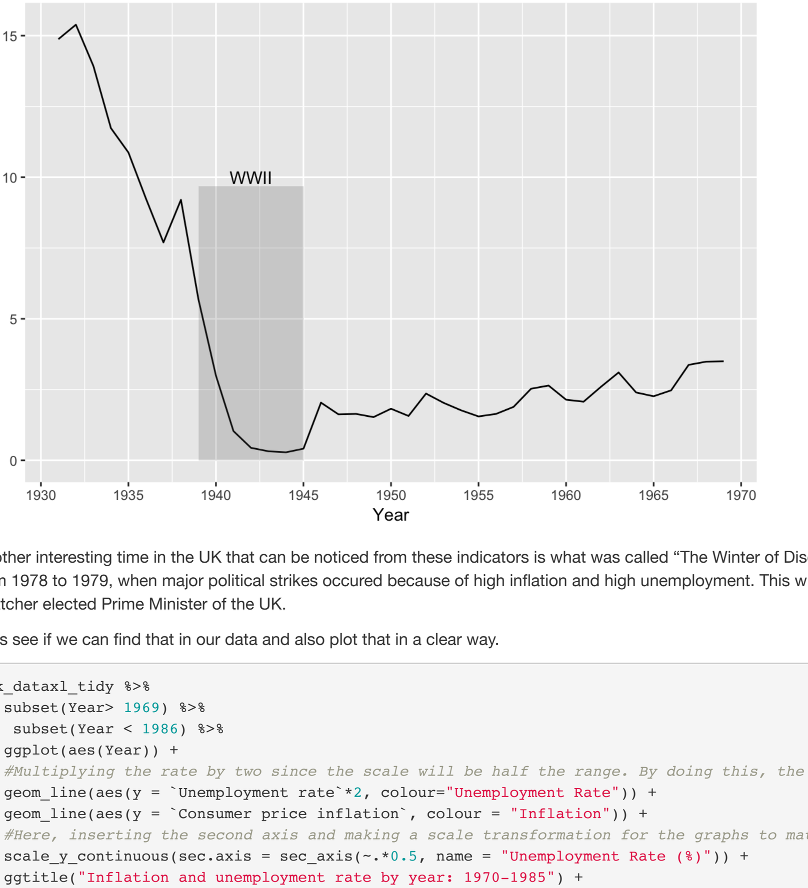
```
uk_dataxl_tidy %>%
  subset(Year> 1930) %>%
  ggplot(aes(Year)) +
  geom_line(aes(y = `Unemployment rate`)) +
  ggtitle("Unemployment rate by year: 1930-2016") +
  xlab("Year") + ylab("Unemployment rate (%)") +
  scale_x_continuous(breaks = c(1930,1935,1940,1945,1950,1955,1960,1965,1970, 1975, 1980, 1985, 1990, 1995, 2000, 2005, 2010, 2015))
```



So, this graph is very interesting. If you take a look into the period of WWII, the unemployment rate almost reached 0%. That might be explained because many people were working for the state to win the war. As as you can see, this unemployment rate rose a lot very quickly when the war ended, because many of the people once employed because of the war were now out of a job.

Now i’ll make another graph to highlight what I just mentioned for more clear understanding.

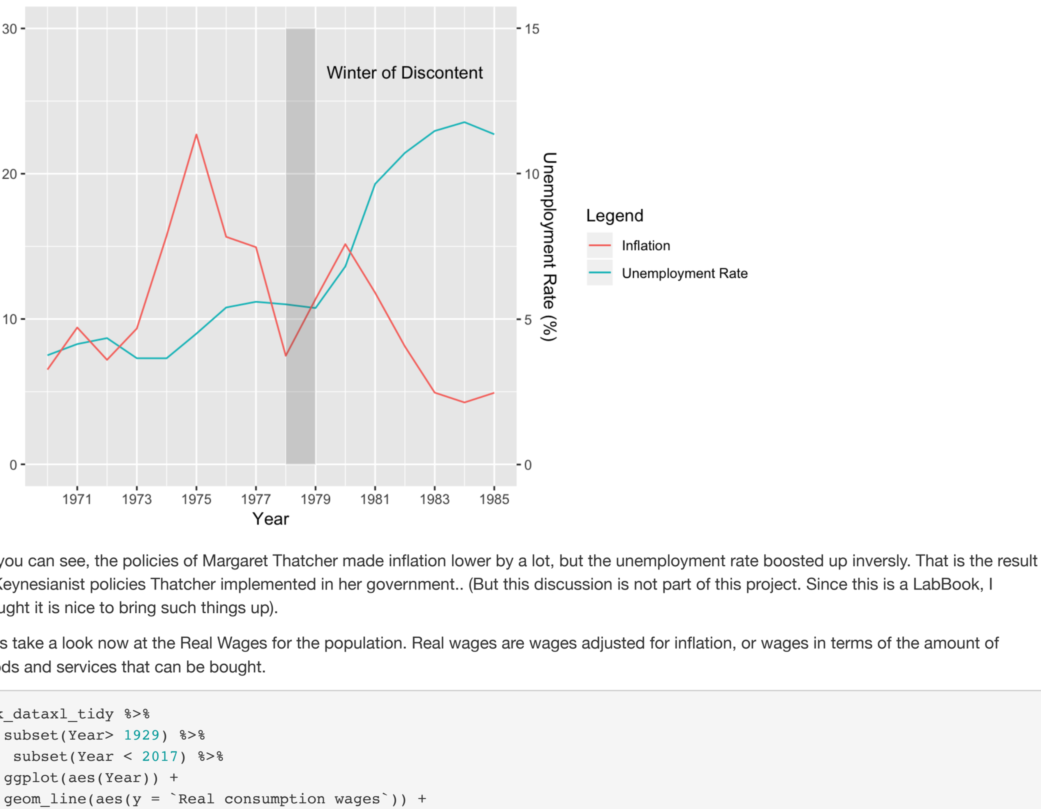
```
uk_dataxl_tidy %>%
  subset(Year> 1930) %>%
   subset(Year < 1970) %>%
  ggplot(aes(Year)) +
  geom_line(aes(y = `Unemployment rate`)) +
  ggtitle("Unemployment rate by year: 1930-1970") +
  xlab("Year") + ylab("Unemployment rate (%)") +
  scale_x_continuous(breaks = c(1930,1935,1940,1945,1950,1955,1960, 1965,1970)) +
 annotate("rect", xmin = 1939, xmax = 1945, ymin = 0, ymax = 9.68, alpha = .2) +
annotate("text", x = 1942, y = 10, label = "WWII")
```



Another interesting time in the UK that can be noticed from these indicators is what was called “The Winter of Discontent”. This was the winter from 1978 to 1979, when major political strikes occured because of high inflation and high unemployment. This winter helped get Margaret Thatcher elected Prime Minister of the UK.

Let’s see if we can find that in our data and also plot that in a clear way.

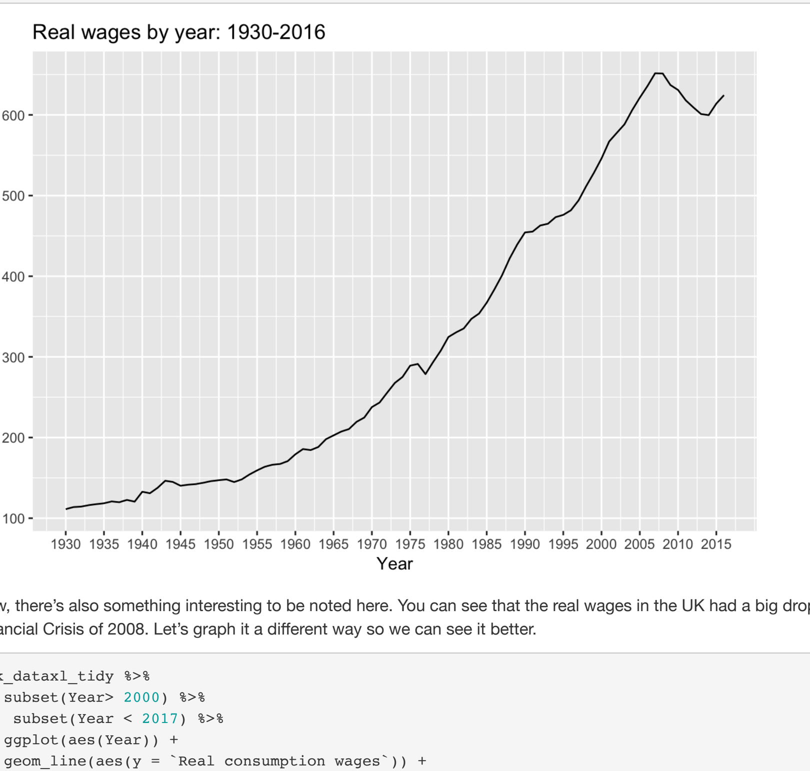
```
uk_dataxl_tidy %>%
  subset(Year> 1969) %>%
   subset(Year < 1986) %>%
  ggplot(aes(Year)) +
  #Multiplying the rate by two since the scale will be half the range. By doing this, the scale is correct
  geom_line(aes(y = `Unemployment rate`*2, colour="Unemployment Rate")) +
  geom_line(aes(y = `Consumer price inflation`, colour = "Inflation")) +
  #Here, inserting the second axis and making a scale transformation for the graphs to match the range
  scale_y_continuous(sec.axis = sec_axis(~.*0.5, name = "Unemployment Rate (%)")) +
  ggtitle("Inflation and unemployment rate by year: 1970-1985") +
  xlab("Year") + ylab("Inflation (%)") +
  scale_x_continuous(breaks = c(1971,1973,1975,1977,1979, 1981, 1983,1985)) +
 annotate("rect", xmin = 1978, xmax = 1979, ymin = 0, ymax = 30, alpha = .2)+
 annotate("text", x = 1982, y = 27, label = "Winter of Discontent")+
  guides(colour = guide_legend(title = "Legend"))
```



As you can see, the policies of Margaret Thatcher made inflation lower by a lot, but the unemployment rate boosted up inversly. That is the result of Keynesianist policies Thatcher implemented in her government.. (But this discussion is not part of this project. Since this is a LabBook, I thought it is nice to bring such things up).

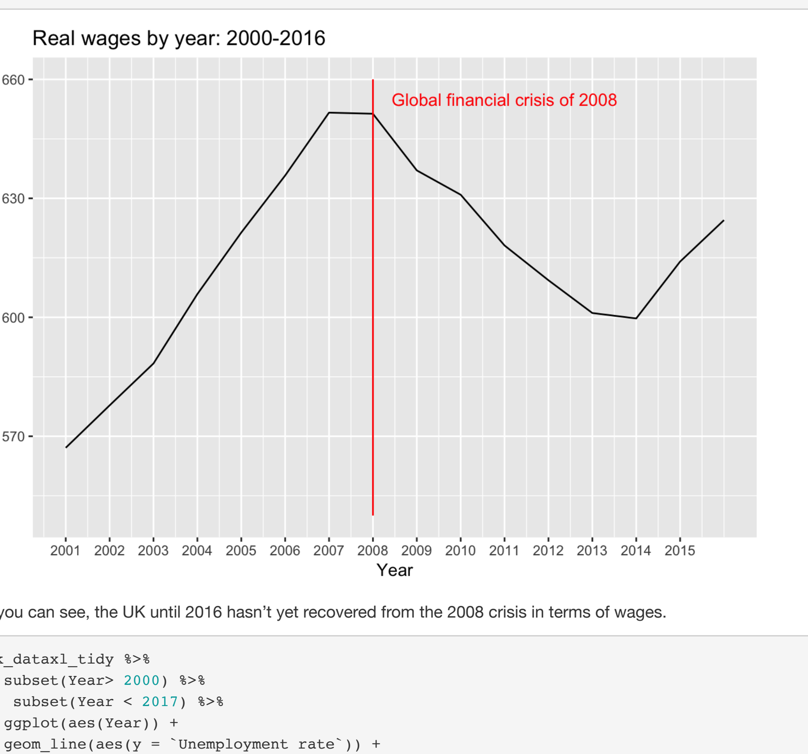
Let’s take a look now at the Real Wages for the population. Real wages are wages adjusted for inflation, or wages in terms of the amount of goods and services that can be bought.

```
uk_dataxl_tidy %>%
  subset(Year> 1929) %>%
   subset(Year < 2017) %>%
  ggplot(aes(Year)) +
  geom_line(aes(y = `Real consumption wages`)) +
  ggtitle("Real wages by year: 1930-2016") +
  xlab("Year") + ylab("Real consumption wages") +
  scale_x_continuous(breaks = c(1930,1935,1940,1945,1950,1955,1960,1965,1970, 1975, 1980, 1985, 1990, 1995, 2000, 2005, 2010, 2015))
```



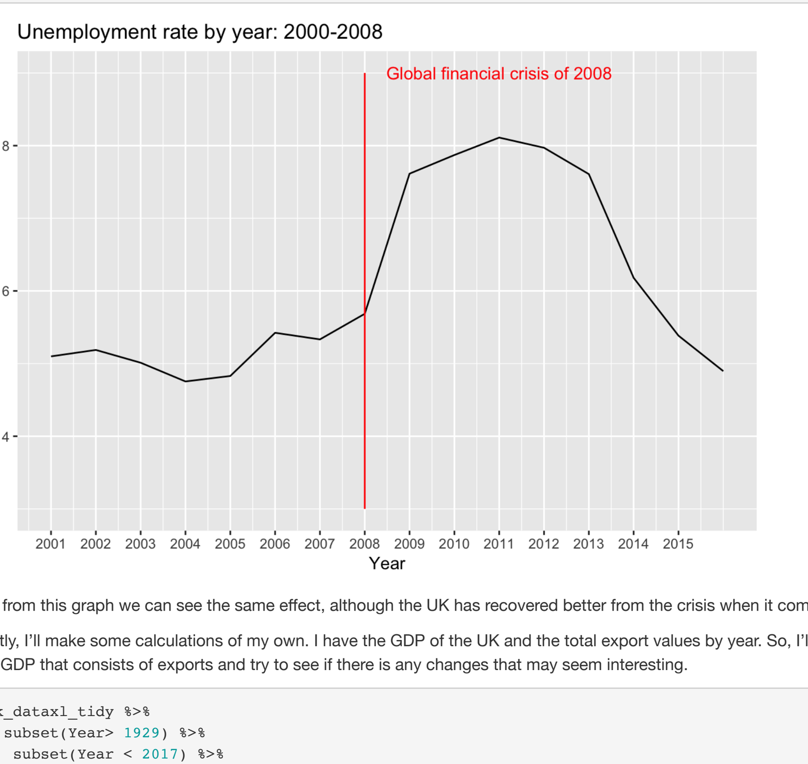
Now, there’s also something interesting to be noted here. You can see that the real wages in the UK had a big drop in 2008. It was the Global Financial Crisis of 2008. Let’s graph it a different way so we can see it better.

```
uk_dataxl_tidy %>%
  subset(Year> 2000) %>%
   subset(Year < 2017) %>%
  ggplot(aes(Year)) +
  geom_line(aes(y = `Real consumption wages`)) +
  ggtitle("Real wages by year: 2000-2016") +
  xlab("Year") + ylab("Real consumption wages") +
  scale_x_continuous(breaks = c(2000,2001,2002,2003,2004, 2005,2006,2007, 2008,2009,2010, 2011,2012,2013,2014,2015)) +
  annotate("segment", x = 2008, xend = 2008, y = 550, yend = 660,colour = "red")+
 annotate("text", x = 2011, y = 655, label = "Global financial crisis of 2008", colour="red")
```



As you can see, the UK until 2016 hasn’t yet recovered from the 2008 crisis in terms of wages.

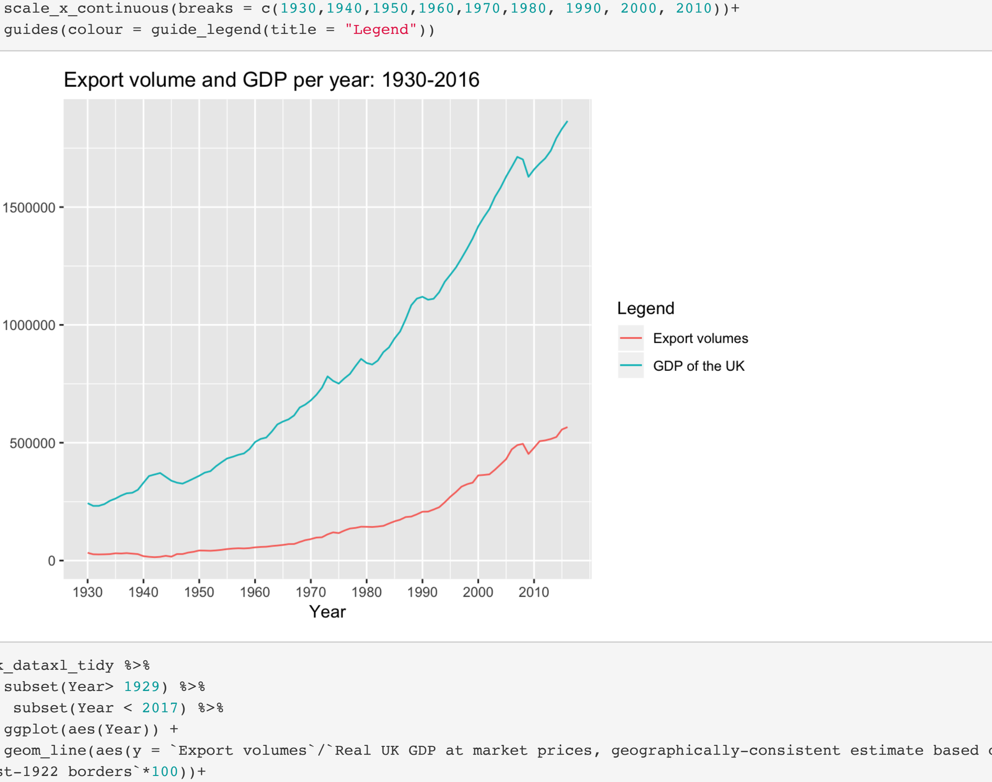
```
uk_dataxl_tidy %>%
  subset(Year> 2000) %>%
   subset(Year < 2017) %>%
  ggplot(aes(Year)) +
  geom_line(aes(y = `Unemployment rate`)) +
  ggtitle("Unemployment rate by year: 2000-2008") +
  xlab("Year") + ylab("Unemployment rate (%)") +
  scale_x_continuous(breaks = c(2000,2001,2002,2003,2004, 2005,2006,2007, 2008,2009,2010, 2011,2012,2013,2014,2015)) +
   annotate("segment", x = 2008, xend = 2008, y = 3, yend = 9,colour = "red")+
 annotate("text", x = 2011, y = 9, label = "Global financial crisis of 2008", colour="red")
```



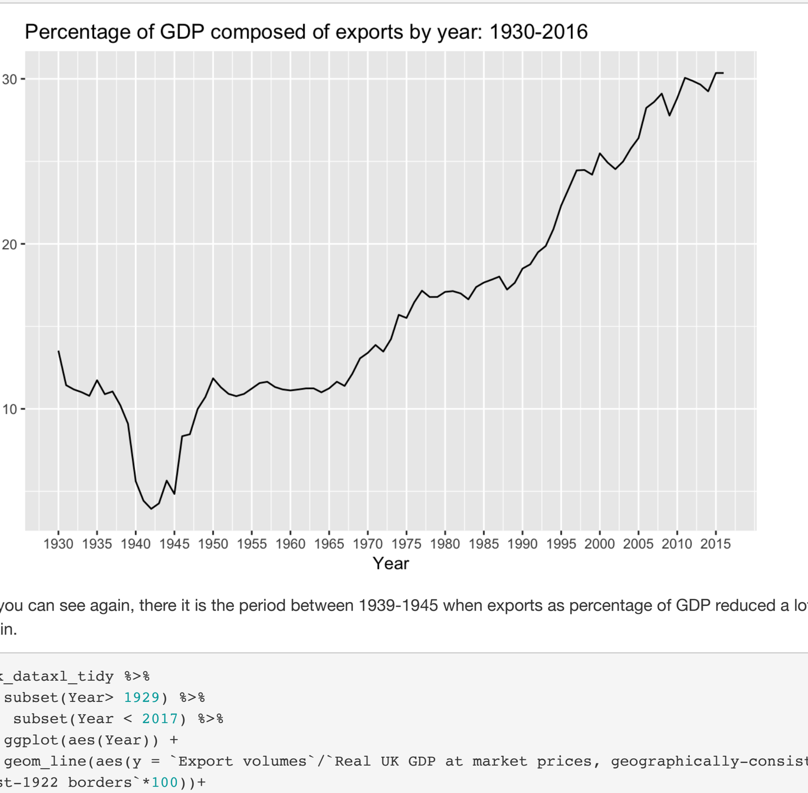
So, from this graph we can see the same effect, although the UK has recovered better from the crisis when it comes to unemployment.

Lastly, I’ll make some calculations of my own. I have the GDP of the UK and the total export values by year. So, I’ll calculate the percentage of the UK GDP that consists of exports and try to see if there is any changes that may seem interesting.

```
uk_dataxl_tidy %>%
  subset(Year> 1929) %>%
   subset(Year < 2017) %>%
  ggplot(aes(Year)) +
  geom_line(aes(y = `Export volumes`, colour="Export volumes")) +
  geom_line(aes(y = `Real UK GDP at market prices, geographically-consistent estimate based on post-1922 borders`, colour="GDP of the UK")) +
  ggtitle("Export volume and GDP per year: 1930-2016") +
  xlab("Year") + ylab("Millions of Pounds") +
  scale_x_continuous(breaks = c(1930,1940,1950,1960,1970,1980, 1990, 2000, 2010))+
  guides(colour = guide_legend(title = "Legend"))
```

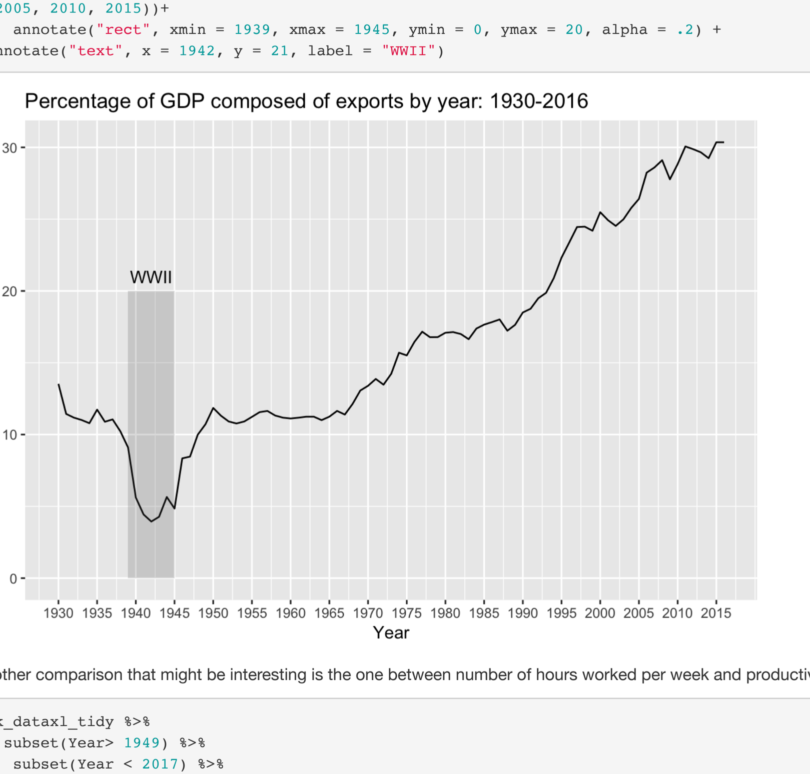



```
uk_dataxl_tidy %>%
  subset(Year> 1929) %>%
   subset(Year < 2017) %>%
  ggplot(aes(Year)) +
  geom_line(aes(y = `Export volumes`/`Real UK GDP at market prices, geographically-consistent estimate based on post-1922 borders`*100))+
  ggtitle("Percentage of GDP composed of exports by year: 1930-2016") +
  xlab("Year") + ylab("Percentage of GDP composed of exports (%)") +
  scale_x_continuous(breaks = c(1930,1935,1940,1945,1950,1955,1960,1965,1970, 1975, 1980, 1985, 1990, 1995, 2000, 2005, 2010, 2015))
```



As you can see again, there it is the period between 1939-1945 when exports as percentage of GDP reduced a lot. I’ll highlight that in the graph again.

```
uk_dataxl_tidy %>%
  subset(Year> 1929) %>%
   subset(Year < 2017) %>%
  ggplot(aes(Year)) +
  geom_line(aes(y = `Export volumes`/`Real UK GDP at market prices, geographically-consistent estimate based on post-1922 borders`*100))+
  ggtitle("Percentage of GDP composed of exports by year: 1930-2016") +
  xlab("Year") + ylab("Percentage of GDP composed of exports (%)") +
  scale_x_continuous(breaks = c(1930,1935,1940,1945,1950,1955,1960,1965,1970, 1975, 1980, 1985, 1990, 1995, 2000, 2005, 2010, 2015))+
   annotate("rect", xmin = 1939, xmax = 1945, ymin = 0, ymax = 20, alpha = .2) +
annotate("text", x = 1942, y = 21, label = "WWII")
```



Another comparison that might be interesting is the one between number of hours worked per week and productivity growth.

```
uk_dataxl_tidy %>%
  subset(Year> 1949) %>%
   subset(Year < 2017) %>%
  ggplot(aes(Year)) +
  geom_line(aes(y = `Labour productivity`, colour="Labour productivity")) +
  geom_line(aes(y = `Average weekly hours worked`*2, colour="Average weekly hours worked")) +
  ggtitle("Average weekly hours worked and labour productivity: 1950-2016") +
  scale_y_continuous(sec.axis = sec_axis(~.*0.5, name = "Hours")) +
  xlab("Year") + ylab("Real GDP per head") +
  scale_x_continuous(breaks = c(1950,1960,1970, 1980, 1990, 2000, 2010)) +
  guides(colour = guide_legend(title = "Legend"))
```

So, it’s interesting to see that somewhere between 1980 there is an interssection (maybe when computers entered the job market) when the productivity growth boosted up, but the hours of work actually dropped a lot.

That ends my analysis for now.

I could conclude, then, that major historical and political factor directly affected the macroeconomic indicators present in this dataset, those events ranging from war to political discontent, financial crisis and technological development.