FinalProject for LPS course 2018/2

Joao Pedro Oliveira

11/29/2018

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

This analysis was created following the literate programming philosophy and made to be totally reproducible. All the data used for analysis was cleaned to become tidy data and all the graphs were created using the “checklist for good graphs” provided in class.

This part of the analysis is the “tip of the iceberg”, or the final report. The complete and in depht alaysis is avaible at the LabBook for this project.

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: WWI, WWII, 2008 crisis, etc).

These are the libraries required to reproduce the code for this analysis.

The code below is all the data gathering and cleaning needed for the project. The data cleaning was made in order to transform once a very messy dataset into a tidy data format, that’s easier and more precise to analyse.

```
#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-m
acroeconomic-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 (such as documentation)
uk_dataxl_tidy <- uk_dataxl[-c(1,2,4,5,6),]

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

#Removing the first row beacuse it just turned into the header (Since it was duplicated)
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 th
ese 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. This is needed for every step
 of the analysis.
uk_dataxl_tidy[] <- lapply(uk_dataxl_tidy, function(x) {
    as.numeric(x)
})

#Renaming columns (Year and Population, for simplicity)
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>
```

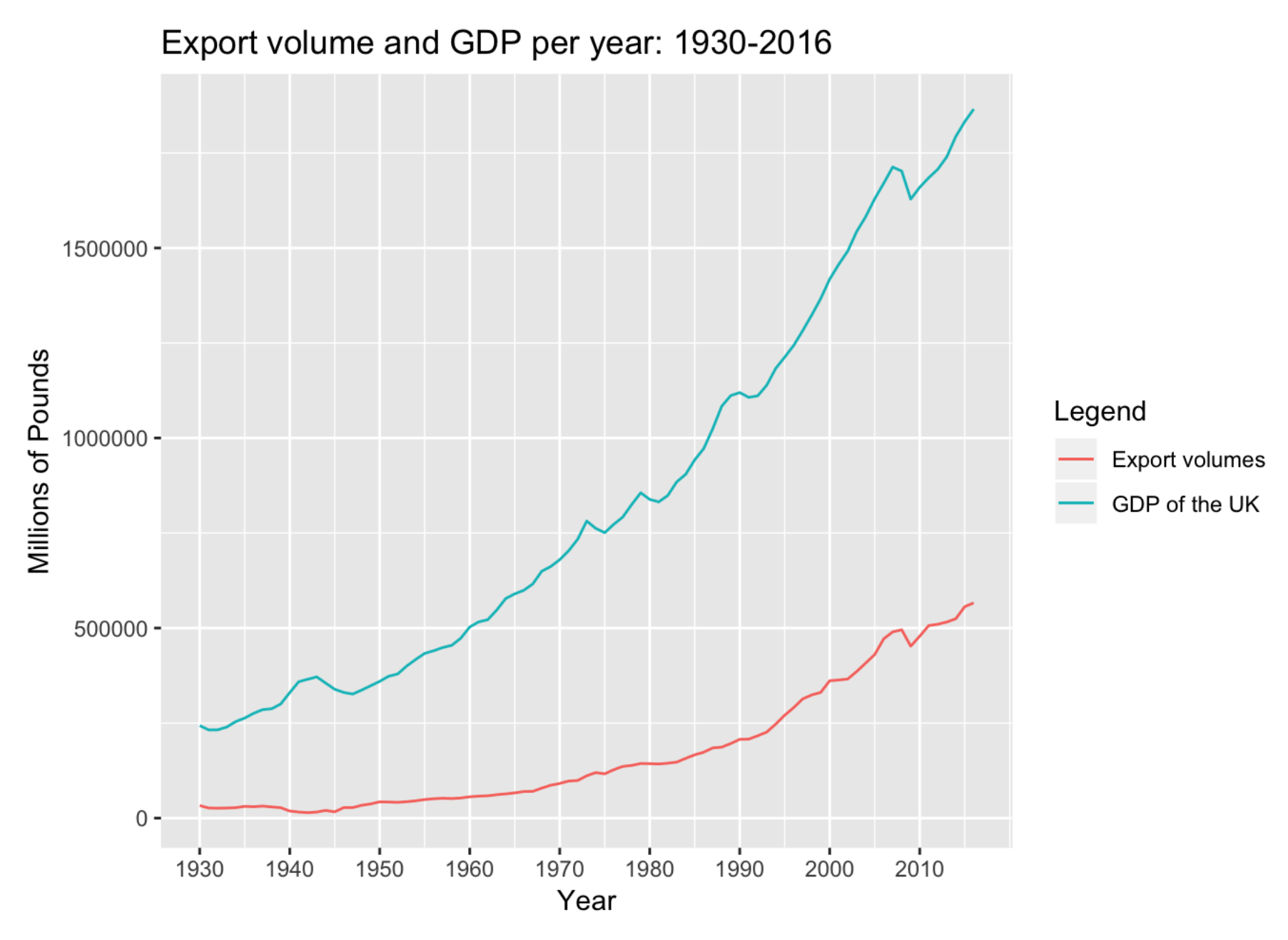
For the first graph: notice that if you take a look into the period of WWII, the unemployment rate almost reached 0%. That might be explained because many people were working for the government and military to win the war. As as you can see, this unemployment rate rose a lot very quickly when the war ended, because many of the people once employed because of the war were now out of a job.

```
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")
```

Now, I calculated the percentage of the UK GDP that consists of exports, and tried to look for some indicators of any historical moments.

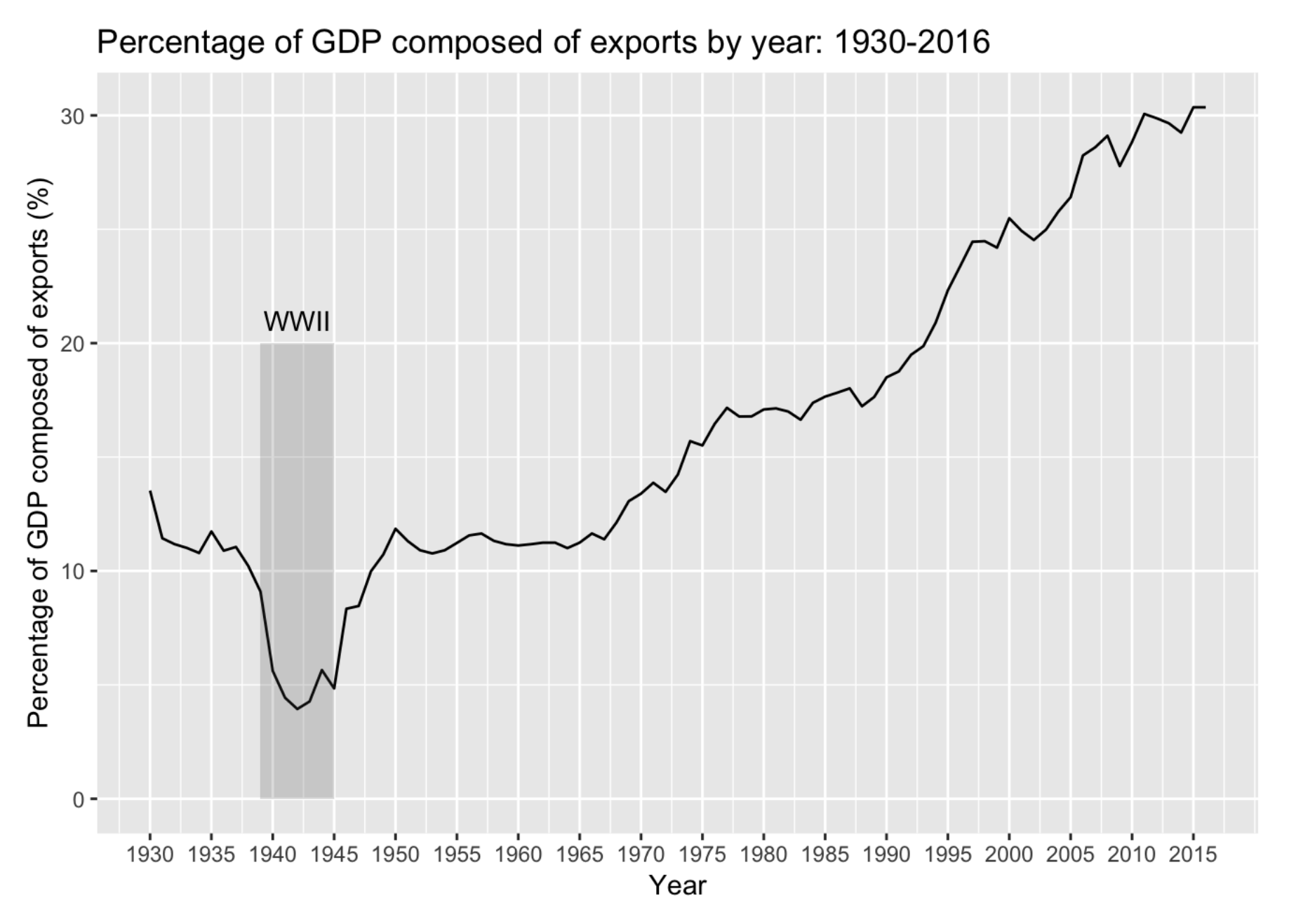
First, I plot two graphs togheter and then the percentage.

```
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"))
```



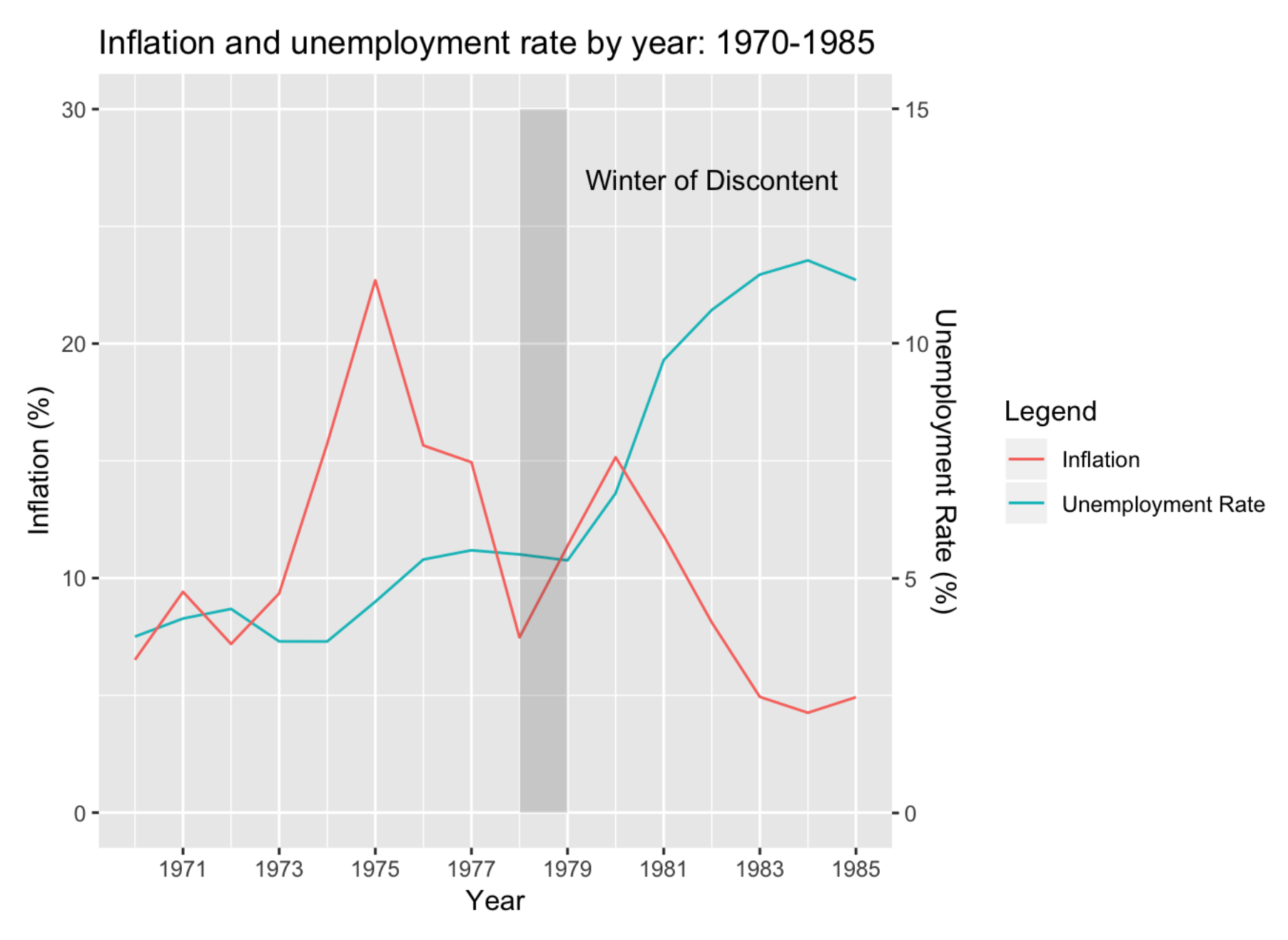
There we can see again a great change in the WWII period, similar to the previous Unemployment graph.

```
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 p
ost-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 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. Here is a look into the economic indicators of that period.

```
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)) +
  #Highlighting the period that i'm trying to visualize
 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, that were proven later to have this effect on the economy.

Another historical period that is very visible from the data is the 2008 Global Financial Crisis, and there are many ways to visualize it using plots. I chose again the unemployment rate and the real consumption wages. Real wages are wages adjusted for inflation, or wages in terms of the amount of goods and services that can be bought. It’s a good indicator to see the buying power of the population in a certain period of time.

```
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, 2011,2012,2013,2014,201
5)) +
  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 it is clear, the wages in the UK haven’t yet recovered from the 2008 financial crisis. Now, let’s look at unemployment.

```
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, 2011,2012,2013,2014,201
5)) +
   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")
```

One of the most interesting things that can be discovered through this dataset too is the comparison between hours worked per week and productivity growth. In this comparison it’s clear to see the introduction of the computer and automated labour, leading to people working less hours per day but having a lot more productivity.

```
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")) +
  annotate("segment", x = 1981, xend = 1981, y = 40, yend = 100,colour = "black") +
  annotate("text", x = 1970, y = 102, label = "MS-DOS release", colour="black")
```

Of course it wasn’t only the MS-DOS release that led to this sudden change of the economy, but it is a nice way to view how technology has changed the way we work.

And that concludes my Report for the final project of the Literate Programming and Statistics course.

-Joao Pedro Oliveira