# Preliminary Analysis - Effect of Economic Crisis

*Sneha Vasanth, Janani Kumar, Tigh Bradley, Gaurav Gada*

*February 21, 2017*

The financial crisis in 2007-2009 occurred as a result of risky financial dealings in the real estate market which spilled over into the world economy at large. The resulting economic downturn in the USA, particularly focusing on the economic changes before and after the collapse of Lehman Brothers in September-October 2008, provides an opportunity to compare impact of the crisis on national ecnomonies. Canada, China, India, and the USA experienced different economic changes during this period, including dramatic changes in unemployment rates.

This study explores what key economic indicators (GDP, CPI, and Unemployment rate) might have looked like in each of these countries if there had been no economic crisis in 2007-2009. Arima forecasts are utilized to project how ecnomic trends may have continued if the crisis had not occurred. The forecast models for the 'healthy' economic state will be compared to what actually occured in these econmies, providing a measure of the relative economic advantages and disadvantages each country experienced during this period.

Cleaning up the data!

```
cpi = read.csv("CPI_World_Bank.csv",check.names = FALSE)
#Get country and indicator column names and remove spaces using make.names
country_indicator_names <- make.names(names(cpi[c(1:4)]))
#country_indicator_names

#Get year column names
year_column_names <- names(cpi[c(5:ncol(cpi))])
#year_column_names

#Reassign the modified column names to actual column names
names(cpi) <- c(country_indicator_names,year_column_names)
names(cpi)
```

```
##  [1] "Country.Name"   "Country.Code"   "Indicator.Name" "Indicator.Code"
##  [5] "1960"           "1961"           "1962"           "1963"
##  [9] "1964"           "1965"           "1966"           "1967"
## [13] "1968"           "1969"           "1970"           "1971"
## [17] "1972"           "1973"           "1974"           "1975"
## [21] "1976"           "1977"           "1978"           "1979"
## [25] "1980"           "1981"           "1982"           "1983"
## [29] "1984"           "1985"           "1986"           "1987"
## [33] "1988"           "1989"           "1990"           "1991"
## [37] "1992"           "1993"           "1994"           "1995"
## [41] "1996"           "1997"           "1998"           "1999"
## [45] "2000"           "2001"           "2002"           "2003"
## [49] "2004"           "2005"           "2006"           "2007"
## [53] "2008"           "2009"           "2010"           "2011"
## [57] "2012"           "2013"           "2014"           "2015"
## [61] "2016"
```

```
#Gather all year column names into a new column called Year and assign its values to a column ca
lled CPI
clean_cpi <- gather(cpi,Year,CPI,-Country.Name,-Country.Code,-Indicator.Name,-Indicator.Code)
cpi_data <- clean_cpi[,!colnames(clean_cpi) %in% c("Indicator.Name","Indicator.Code")]
head(clean_cpi)
```

```
##   Country.Name Country.Code                    Indicator.Name
## 1        Aruba          ABW Consumer price index (2010 = 100)
## 2      Andorra          AND Consumer price index (2010 = 100)
## 3  Afghanistan          AFG Consumer price index (2010 = 100)
## 4       Angola          AGO Consumer price index (2010 = 100)
## 5      Albania          ALB Consumer price index (2010 = 100)
## 6   Arab World          ARB Consumer price index (2010 = 100)
##   Indicator.Code Year CPI
## 1    FP.CPI.TOTL 1960  NA
## 2    FP.CPI.TOTL 1960  NA
## 3    FP.CPI.TOTL 1960  NA
## 4    FP.CPI.TOTL 1960  NA
## 5    FP.CPI.TOTL 1960  NA
## 6    FP.CPI.TOTL 1960  NA
```

```
cpi = read.csv("CPI_World_Bank.csv",check.names = FALSE)
#Get country and indicator column names and remove spaces using make.names
country_indicator_names <- make.names(names(cpi[c(1:4)]))
#country_indicator_names

#Get year column names
year_column_names <- names(cpi[c(5:ncol(cpi))])
#year_column_names

#Reassign the modified column names to actual column names
names(cpi) <- c(country_indicator_names,year_column_names)
names(cpi)
```

```
##  [1] "Country.Name"   "Country.Code"   "Indicator.Name" "Indicator.Code"
##  [5] "1960"           "1961"           "1962"           "1963"
##  [9] "1964"           "1965"           "1966"           "1967"
## [13] "1968"           "1969"           "1970"           "1971"
## [17] "1972"           "1973"           "1974"           "1975"
## [21] "1976"           "1977"           "1978"           "1979"
## [25] "1980"           "1981"           "1982"           "1983"
## [29] "1984"           "1985"           "1986"           "1987"
## [33] "1988"           "1989"           "1990"           "1991"
## [37] "1992"           "1993"           "1994"           "1995"
## [41] "1996"           "1997"           "1998"           "1999"
## [45] "2000"           "2001"           "2002"           "2003"
## [49] "2004"           "2005"           "2006"           "2007"
## [53] "2008"           "2009"           "2010"           "2011"
## [57] "2012"           "2013"           "2014"           "2015"
## [61] "2016"
```

```
#Gather all year column names into a new column called Year and assign its values to a column ca
lled CPI
clean_cpi <- gather(cpi,Year,CPI,-Country.Name,-Country.Code,-Indicator.Name,-Indicator.Code)
cpi_data <- clean_cpi[,!colnames(clean_cpi) %in% c("Indicator.Name","Indicator.Code")]
head(cpi_data)
```

```
##   Country.Name Country.Code Year CPI
## 1        Aruba          ABW 1960  NA
## 2      Andorra          AND 1960  NA
## 3  Afghanistan          AFG 1960  NA
## 4       Angola          AGO 1960  NA
## 5      Albania          ALB 1960  NA
## 6   Arab World          ARB 1960  NA
```

```
unemp = read.csv("Unemployment_ILO.csv",check.names = FALSE)
country_indicator_names <- make.names(names(unemp[c(1:4)]))
#country_indicator_names

#Get year column names
year_column_names <- names(unemp[c(5:ncol(unemp))])
#year_column_names

#Reassign the modified column names to actual column names
names(unemp) <- c(country_indicator_names,year_column_names)
names(unemp)
```

```
##  [1] "Country.Name"   "Country.Code"   "Indicator.Name" "Indicator.Code"
##  [5] "1991"           "1992"           "1993"           "1994"
##  [9] "1995"           "1996"           "1997"           "1998"
## [13] "1999"           "2000"           "2001"           "2002"
## [17] "2003"           "2004"           "2005"           "2006"
## [21] "2007"           "2008"           "2009"           "2010"
## [25] "2011"           "2012"           "2013"           "2014"
```

```
#Gather all year column names into a new column called Year and assign its values to a column ca
lled unemp
clean_unemp <- gather(unemp,Year,unemp,-Country.Name,-Country.Code,-Indicator.Name,-Indicator.Co
de)
unemp_data <- clean_unemp[,!colnames(clean_unemp) %in% c("Indicator.Name","Indicator.Code")]
head(clean_unemp)
```

```
##   Country.Name Country.Code
## 1        Aruba          ABW
## 2      Andorra          AND
## 3  Afghanistan          AFG
## 4       Angola          AGO
## 5      Albania          ALB
## 6   Arab World          ARB
##                                                         Indicator.Name
## 1 Unemployment, total (% of total labor force) (modeled ILO estimate)
## 2 Unemployment, total (% of total labor force) (modeled ILO estimate)
## 3 Unemployment, total (% of total labor force) (modeled ILO estimate)
## 4 Unemployment, total (% of total labor force) (modeled ILO estimate)
## 5 Unemployment, total (% of total labor force) (modeled ILO estimate)
## 6 Unemployment, total (% of total labor force) (modeled ILO estimate)
##   Indicator.Code Year    unemp
## 1 SL.UEM.TOTL.ZS 1991       NA
## 2 SL.UEM.TOTL.ZS 1991       NA
## 3 SL.UEM.TOTL.ZS 1991  8.60000
## 4 SL.UEM.TOTL.ZS 1991  6.90000
## 5 SL.UEM.TOTL.ZS 1991 11.80000
## 6 SL.UEM.TOTL.ZS 1991 12.55804
```

```
#Get country and indicator column names and remove spaces using make.names
country_indicator_names <- make.names(names(unemp[c(1:4)]))
#country_indicator_names

#Get year column names
year_column_names <- names(unemp[c(5:ncol(unemp))])
#year_column_names

#Reassign the modified column names to actual column names
names(unemp) <- c(country_indicator_names,year_column_names)
names(unemp)
```

```
##  [1] "Country.Name"   "Country.Code"   "Indicator.Name" "Indicator.Code"
##  [5] "1991"           "1992"           "1993"           "1994"
##  [9] "1995"           "1996"           "1997"           "1998"
## [13] "1999"           "2000"           "2001"           "2002"
## [17] "2003"           "2004"           "2005"           "2006"
## [21] "2007"           "2008"           "2009"           "2010"
## [25] "2011"           "2012"           "2013"           "2014"
```

```
#Gather all year column names into a new column called Year and assign its values to a column ca
lled unemp
clean_unemp <- gather(unemp,Year,unemp,-Country.Name,-Country.Code,-Indicator.Name,-Indicator.Co
de)
unemp_data <- clean_unemp[,!colnames(clean_unemp) %in% c("Indicator.Name","Indicator.Code")]
head(unemp_data)
```

```
##   Country.Name Country.Code Year    unemp
## 1        Aruba          ABW 1991       NA
## 2      Andorra          AND 1991       NA
## 3  Afghanistan          AFG 1991  8.60000
## 4       Angola          AGO 1991  6.90000
## 5      Albania          ALB 1991 11.80000
## 6   Arab World          ARB 1991 12.55804
```

```
gdp = read.csv("gdp.csv",check.names = FALSE)
country_indicator_names <- make.names(names(gdp[c(1:4)]))
#country_indicator_names

#Get year column names
year_column_names <- names(gdp[c(5:ncol(gdp))])
#year_column_names

#Reassign the modified column names to actual column names
names(gdp) <- c(country_indicator_names,year_column_names)
names(gdp)
```

```
##  [1] "Country.Name"   "Country.Code"   "Indicator.Name" "Indicator.Code"
##  [5] "1960"           "1961"           "1962"           "1963"
##  [9] "1964"           "1965"           "1966"           "1967"
## [13] "1968"           "1969"           "1970"           "1971"
## [17] "1972"           "1973"           "1974"           "1975"
## [21] "1976"           "1977"           "1978"           "1979"
## [25] "1980"           "1981"           "1982"           "1983"
## [29] "1984"           "1985"           "1986"           "1987"
## [33] "1988"           "1989"           "1990"           "1991"
## [37] "1992"           "1993"           "1994"           "1995"
## [41] "1996"           "1997"           "1998"           "1999"
## [45] "2000"           "2001"           "2002"           "2003"
## [49] "2004"           "2005"           "2006"           "2007"
## [53] "2008"           "2009"           "2010"           "2011"
## [57] "2012"           "2013"           "2014"           "2015"
## [61] "2016"
```

```
#Gather all year column names into a new column called Year and assign its values to a column ca
lled gdp
clean_gdp <- gather(gdp,Year,gdp,-Country.Name,-Country.Code,-Indicator.Name,-Indicator.Code)
gdp_data <- clean_gdp[,!colnames(clean_gdp) %in% c("Indicator.Name","Indicator.Code")]
head(clean_gdp)
```

```
##   Country.Name Country.Code    Indicator.Name Indicator.Code Year
## 1        Aruba          ABW GDP (current US$) NY.GDP.MKTP.CD 1960
## 2      Andorra          AND GDP (current US$) NY.GDP.MKTP.CD 1960
## 3  Afghanistan          AFG GDP (current US$) NY.GDP.MKTP.CD 1960
## 4       Angola          AGO GDP (current US$) NY.GDP.MKTP.CD 1960
## 5      Albania          ALB GDP (current US$) NY.GDP.MKTP.CD 1960
## 6   Arab World          ARB GDP (current US$) NY.GDP.MKTP.CD 1960
##         gdp
## 1        NA
## 2        NA
## 3 537777811
## 4        NA
## 5        NA
## 6        NA
```

```
#Get country and indicator column names and remove spaces using make.names
country_indicator_names <- make.names(names(gdp[c(1:4)]))
#country_indicator_names

#Get year column names
year_column_names <- names(gdp[c(5:ncol(gdp))])
#year_column_names

#Reassign the modified column names to actual column names
names(gdp) <- c(country_indicator_names,year_column_names)
names(gdp)
```

```
##  [1] "Country.Name"   "Country.Code"   "Indicator.Name" "Indicator.Code"
##  [5] "1960"           "1961"           "1962"           "1963"
##  [9] "1964"           "1965"           "1966"           "1967"
## [13] "1968"           "1969"           "1970"           "1971"
## [17] "1972"           "1973"           "1974"           "1975"
## [21] "1976"           "1977"           "1978"           "1979"
## [25] "1980"           "1981"           "1982"           "1983"
## [29] "1984"           "1985"           "1986"           "1987"
## [33] "1988"           "1989"           "1990"           "1991"
## [37] "1992"           "1993"           "1994"           "1995"
## [41] "1996"           "1997"           "1998"           "1999"
## [45] "2000"           "2001"           "2002"           "2003"
## [49] "2004"           "2005"           "2006"           "2007"
## [53] "2008"           "2009"           "2010"           "2011"
## [57] "2012"           "2013"           "2014"           "2015"
## [61] "2016"
```

```
#Gather all year column names into a new column called Year and assign its values to a column ca
lled gdp
clean_gdp <- gather(gdp,Year,gdp,-Country.Name,-Country.Code,-Indicator.Name,-Indicator.Code)
gdp_data <- clean_gdp[,!colnames(clean_gdp) %in% c("Indicator.Name","Indicator.Code")]
head(gdp_data)
```

```
##   Country.Name Country.Code Year       gdp
## 1        Aruba          ABW 1960        NA
## 2      Andorra          AND 1960        NA
## 3  Afghanistan          AFG 1960 537777811
## 4       Angola          AGO 1960        NA
## 5      Albania          ALB 1960        NA
## 6   Arab World          ARB 1960        NA
```

```
# Join all data
gdp_cpi = merge(gdp_data, cpi_data, by=c("Country.Code","Year"), all = T) # NA's match

labor_data = merge(gdp_cpi, unemp_data, by=c("Country.Code","Year"), all = T) # NA's match

drops = c("Country.Name.y","Country.Name.x")
labor_data = labor_data[ , !(names(labor_data) %in% drops)]

names(labor_data)
```

```
## [1] "Country.Code" "Year"         "gdp"          "CPI"
## [5] "Country.Name" "unemp"
```

```
labor_data = labor_data[c("Country.Name", "Country.Code", "Year", "gdp", "unemp", "CPI")]

# Clean up data
labor_data = labor_data[!(is.na(labor_data$unemp) & is.na(labor_data$CPI) &
is.na(labor_data$gdp)),]

# Clean up the working space
rm(list = ls()[grep("labor", ls(), invert = T)])

country_codes = read.csv("country-codes.csv")

country_data = labor_data[labor_data$Country.Code %in% country_codes$ISO3166.1.Alpha.3,]

labor_data <- country_data

attach(labor_data)
```

# Exploratory Analysis

How many unique observations to you have?

The number of unique observations for CPI, GDP and unemployment respectively are:

```
nrow(labor_data[!is.na(CPI),])
```

```
## [1] 7167
```

```
nrow(labor_data[!is.na(gdp),])
```

```
## [1] 8867
```

```
nrow(labor_data[!is.na(unemp),])
```

```
## [1] 4176
```

What information/features/characteristics do you have for each observation?

We modify the scope of our dataset from 1991 to 2014, since unemployment data is only available for this range.

```
labor_data <- labor_data[Year>=1991 & Year<=2014,]
head(labor_data)
```

```
##    Country.Name Country.Code Year        gdp unemp      CPI
## 32        Aruba          ABW 1991         NA    NA 52.03857
## 33        Aruba          ABW 1992         NA    NA 54.05422
## 34        Aruba          ABW 1993         NA    NA 56.87345
## 35        Aruba          ABW 1994 1330167598    NA 60.46277
## 36        Aruba          ABW 1995 1320670391    NA 62.49516
## 37        Aruba          ABW 1996 1379888268    NA 64.51081
```

What are the min/max/mean/median/sd values for each of these features?

```
#Summary values for CPI
summary(labor_data$CPI,na.rm = TRUE)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's
##    0.00   54.99   78.50   74.79   98.32  348.20     893
```

```
sd(labor_data$CPI, na.rm = TRUE)
```

```
## [1] 31.68783
```

```
#Summary values for GDP
summary(labor_data$gdp,na.rm = TRUE)
```

```
##      Min.   1st Qu.    Median      Mean   3rd Qu.      Max.      NA's
## 9.365e+06 2.653e+09 1.191e+10 2.334e+11 8.283e+10 1.739e+13       195
```

```
sd(labor_data$gdp, na.rm = TRUE)
```

```
## [1] 1.038937e+12
```

```
#Summary values for Unemployment
summary(labor_data$unemp,na.rm = TRUE)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's 
##   0.100   4.500   7.200   8.918  11.200  39.300     654
```

```
sd(labor_data$unemp, na.rm = TRUE)
```

```
## [1] 6.306749
```

We found the averages for CPI, gdp and unemployment for all the countries. Now, let’s manipulate the dataset to include a column to indicate if the time period for the data is before(1991-2006) or after(2007-2014) the economic crisis.

```
labor_data <- labor_data %>% mutate(Time_Period = ifelse(Year < 2007,"Before Crisis","After Cris
is"))
head(labor_data)
```

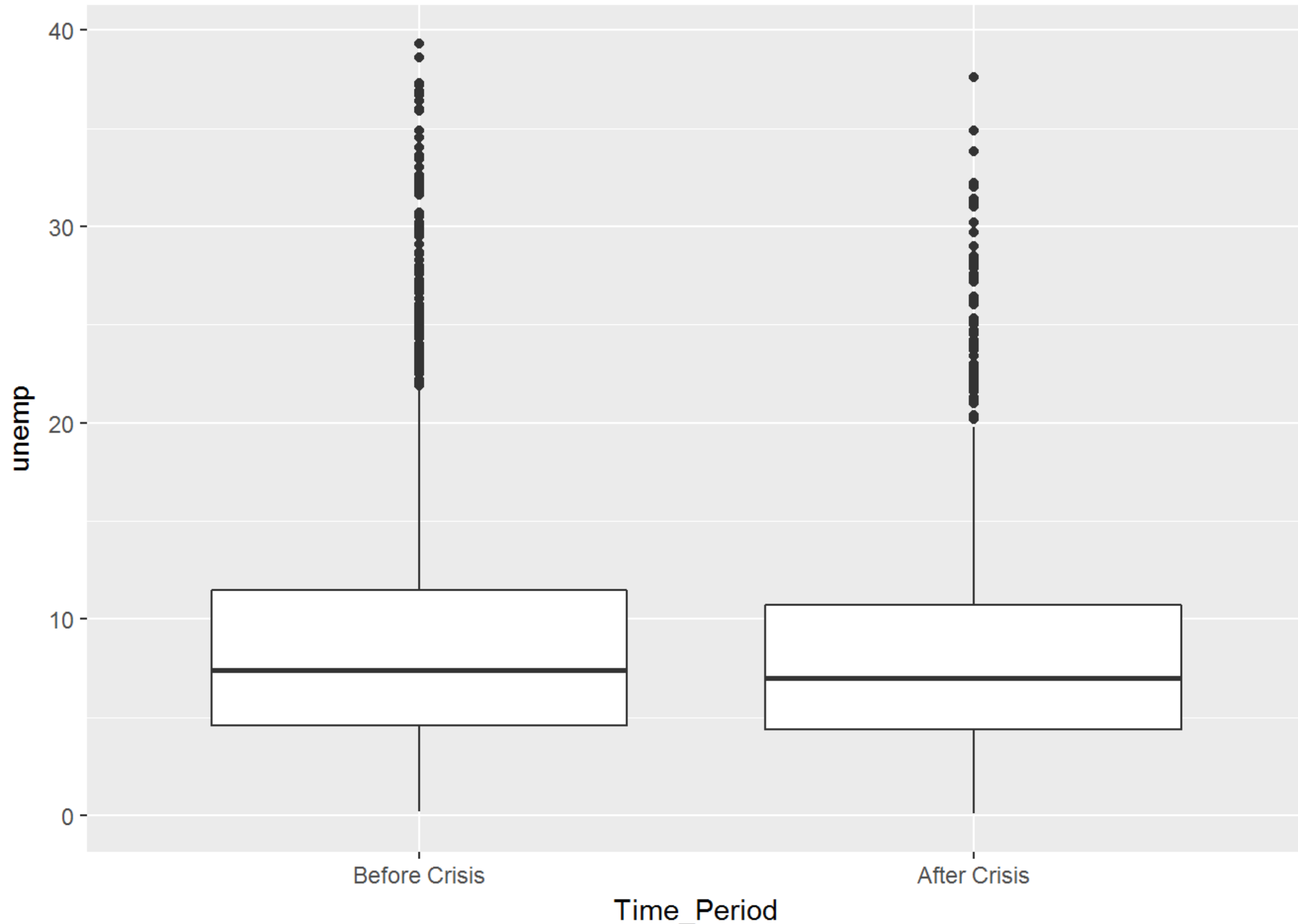
```
##   Country.Name Country.Code Year        gdp unemp      CPI   Time_Period
## 1        Aruba          ABW 1991         NA    NA 52.03857 Before Crisis
## 2        Aruba          ABW 1992         NA    NA 54.05422 Before Crisis
## 3        Aruba          ABW 1993         NA    NA 56.87345 Before Crisis
## 4        Aruba          ABW 1994 1330167598    NA 60.46277 Before Crisis
## 5        Aruba          ABW 1995 1320670391    NA 62.49516 Before Crisis
## 6        Aruba          ABW 1996 1379888268    NA 64.51081 Before Crisis
```

Converting Time_Period to a factor variable and releveling the labels.

```
labor_data$Time_Period <- as.character(labor_data$Time_Period)
labor_data$Time_Period <- factor(labor_data$Time_Period, levels = c("Before Crisis", "After Cris
is"))
levels(labor_data$Time_Period)
```

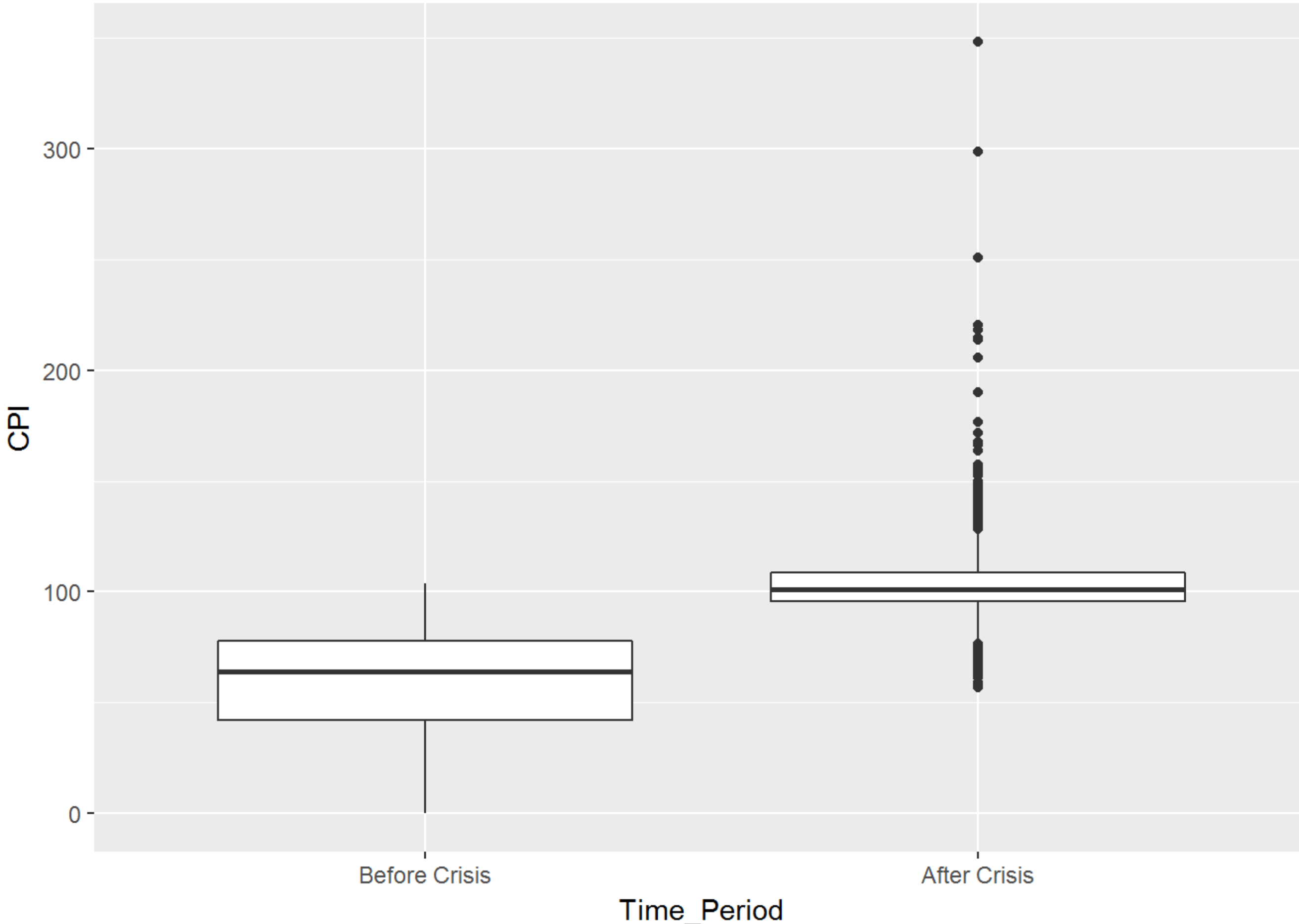
```
## [1] "Before Crisis" "After Crisis"
```

```
ggplot(labor_data, aes(x=Time_Period, y = unemp))+geom_boxplot(na.rm = T)
```



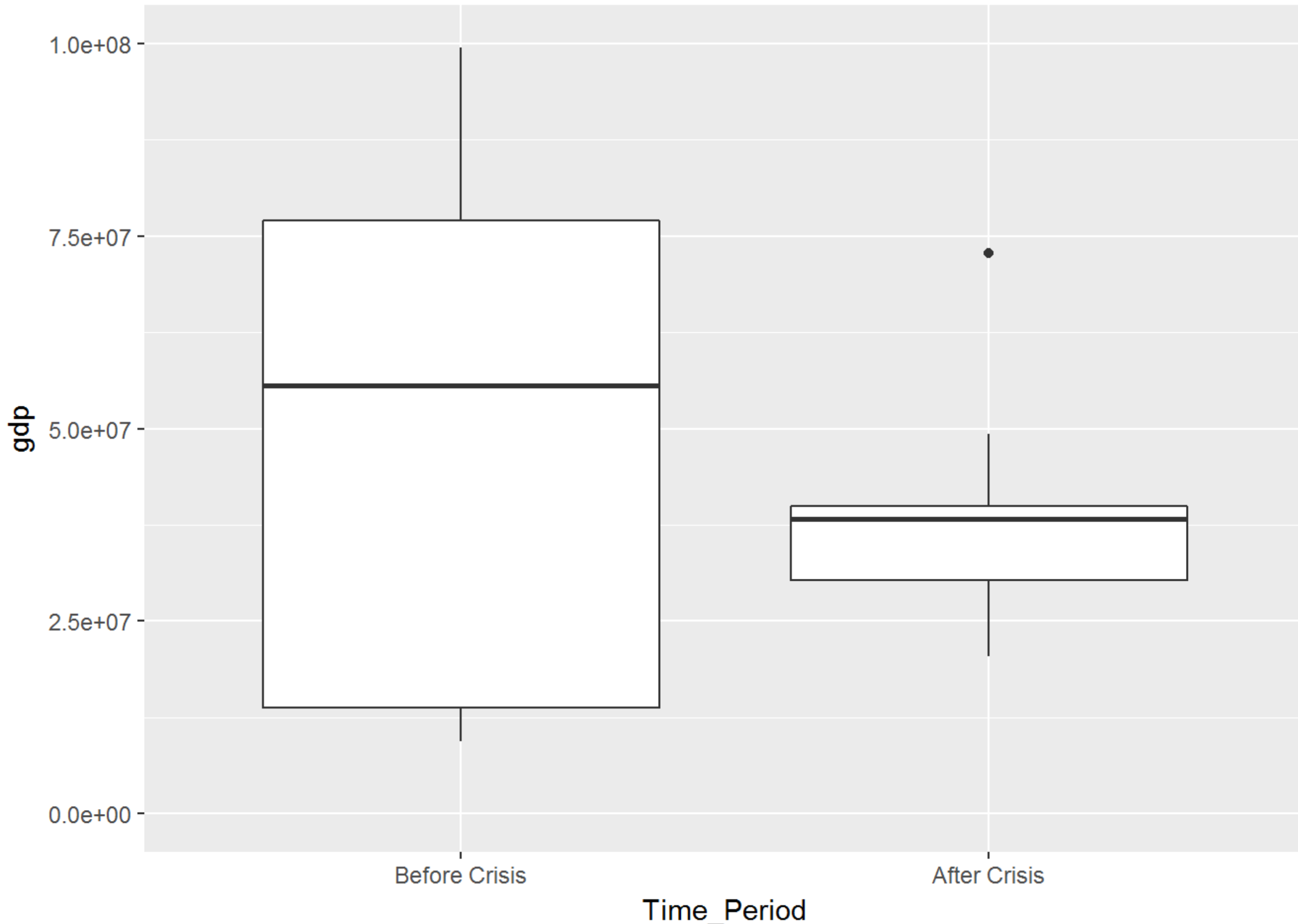

The boxplot confirms the finding the average unemployment is almost the same before and after crisis. This is an interesting finding since the 2007 economic crisis led to unemployment. We will try to analyze how the unemployment trends varied post recession by drilling down further.

```
ggplot(labor_data, aes(x=Time_Period, y = CPI))+geom_boxplot(na.rm = T)
```




The average value of CPI has increased increased post the crisis but the variability has got much less over time.

```
ggplot(labor_data, aes(x=Time_Period, y = gdp))+ylim(0,100000000)+geom_boxplot(na.rm = T)
```

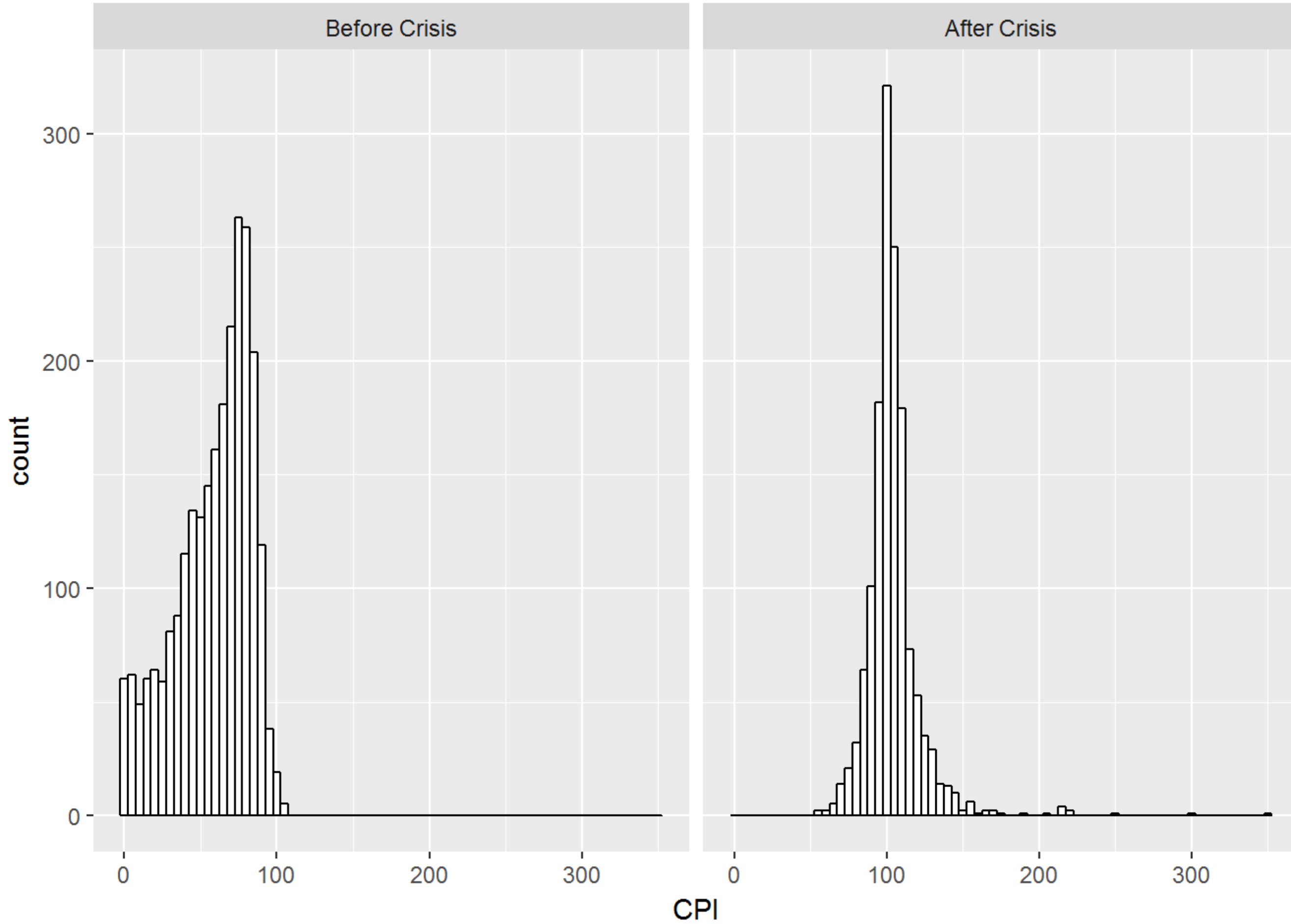



The average value of gdp has decreased post the crisis but the variability has got much less over time.

What is the distribution of the core features (show a histogram)?

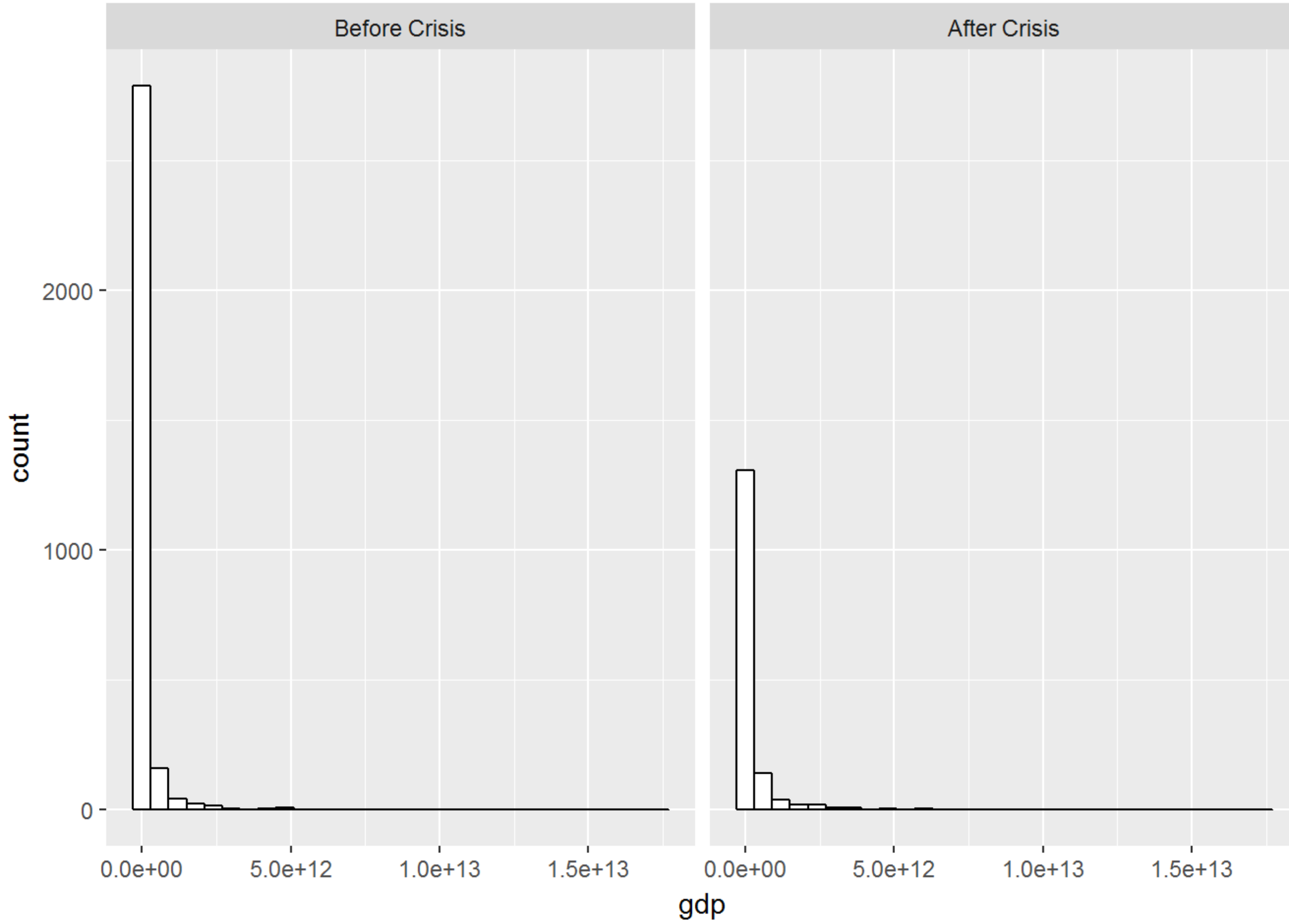
Below are the histograms of the distribution of the CPI, GDP and Unemployment before and after 2007 economic crisis.

```
ggplot(labor_data, aes(x = CPI)) +
  geom_histogram(fill = "white",
                 color = "black",
                 binwidth = 5,na.rm = T)+facet_grid(.~Time_Period)
```

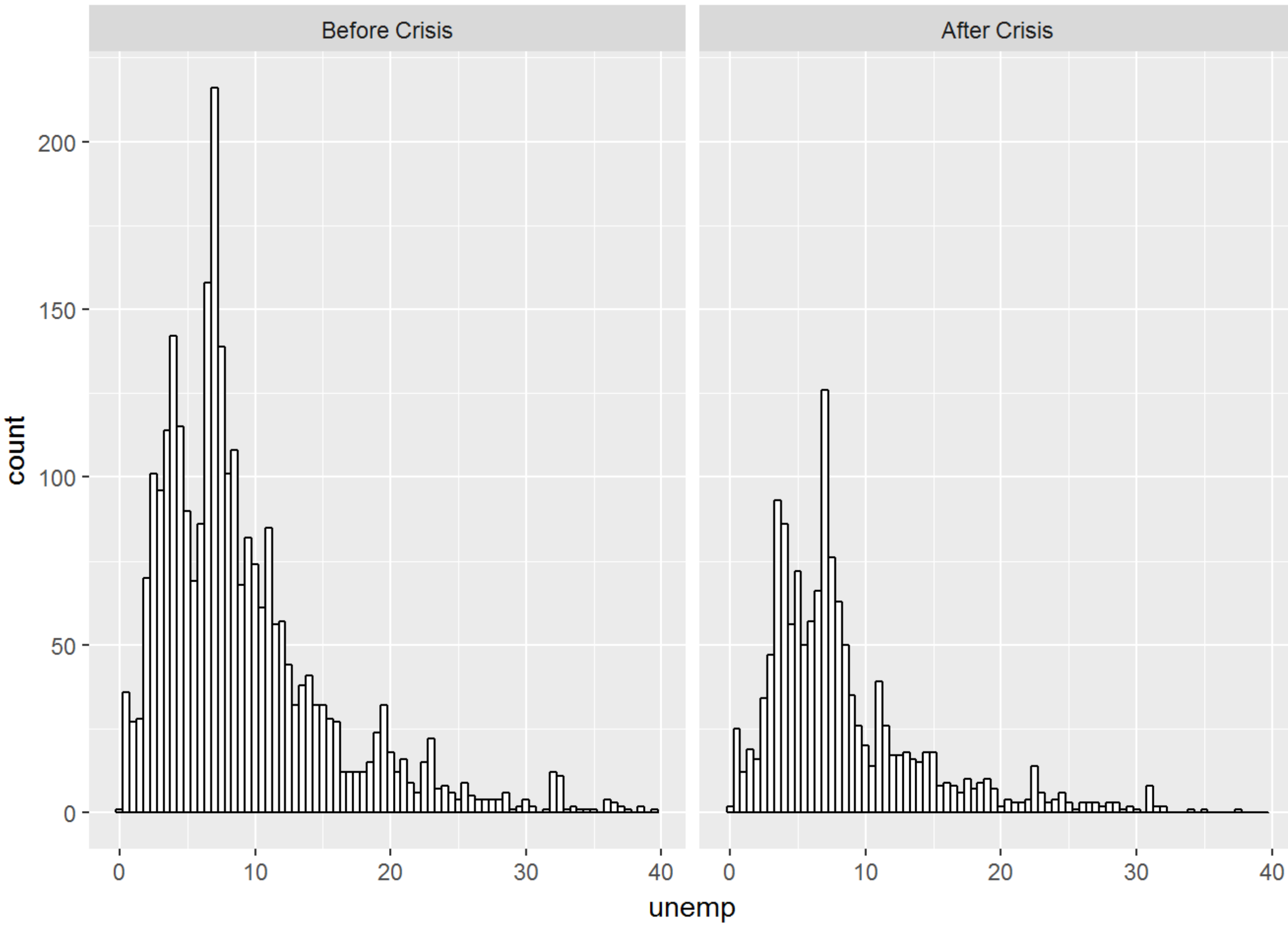



```
ggplot(labor_data, aes(x = gdp)) +
  geom_histogram(fill = "white",
                 color = "black",na.rm = T)+facet_grid(.~Time_Period)
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```




```
ggplot(labor_data, aes(x = unemp)) +
  geom_histogram(fill = "white",
                 color = "black",
                 binwidth = 0.5,na.rm = T)+facet_grid(.~Time_Period)
```

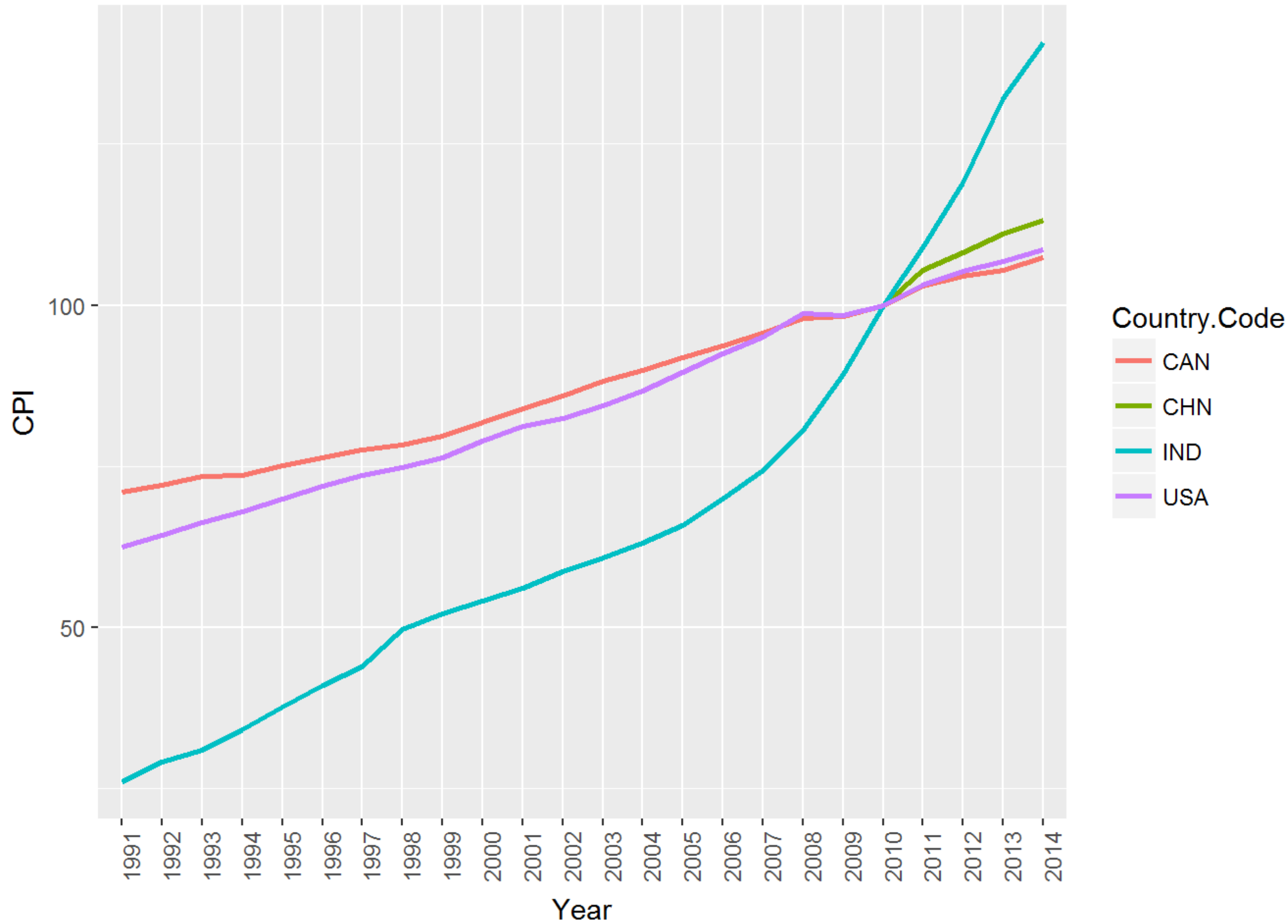



The distribution of CPI was significantly different before and after crisis. The count of the number of countries with higher inflation rates increased after crisis. The distribution for GDP and unemployment are not significantly different before and after crisis.

Let’s understand the effect of CPI,GDP and unemployment further by analyzing trend graph from 1991 to 2014 in USA,Canada,China and India.

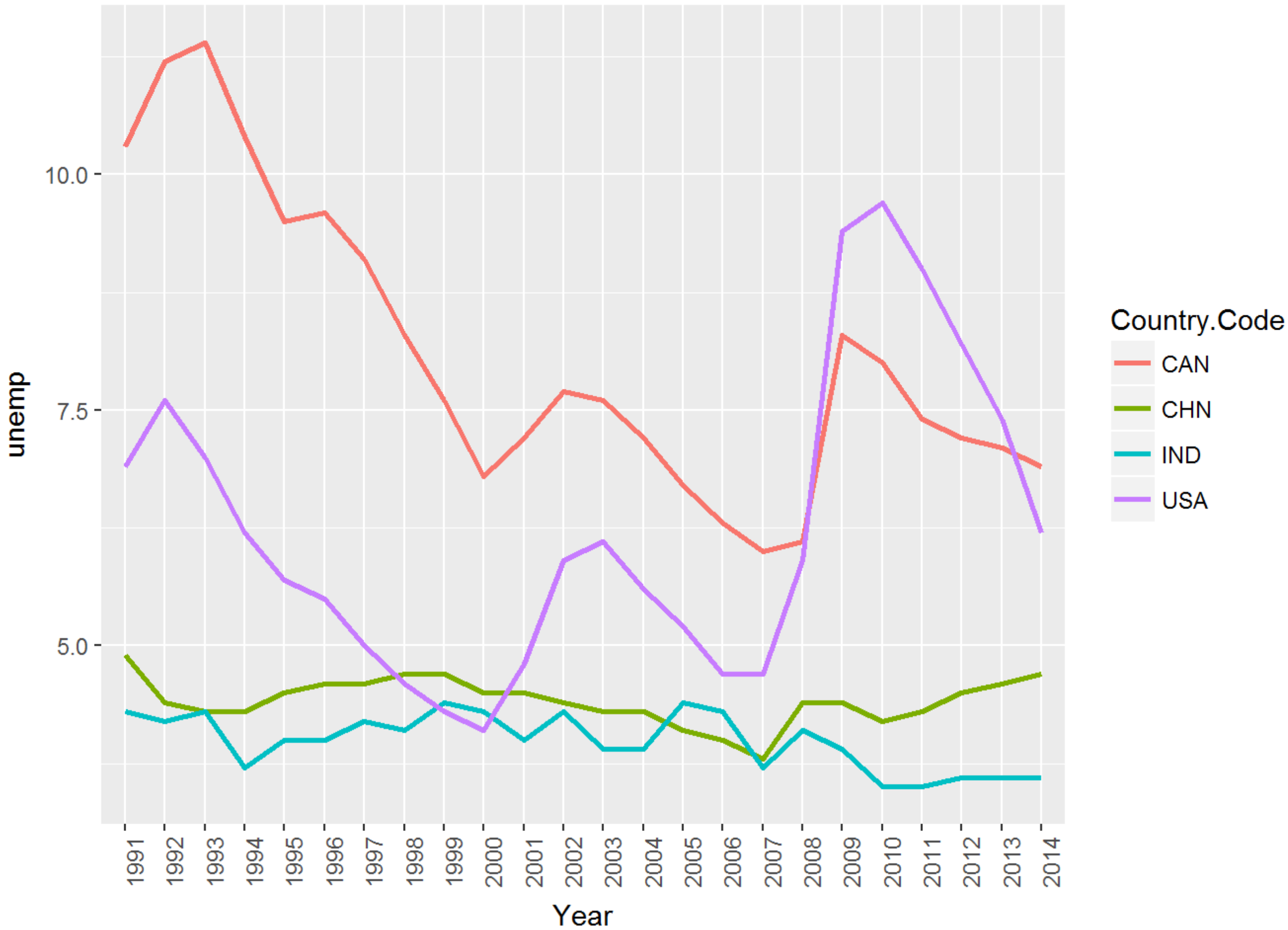
Are there obvious trends in the data (over time, across subgroups, etc.), and are the differences statistically significant?

```
cpi_data_1 <- labor_data %>%
  select(Country.Name,Country.Code,Year,CPI) %>%
  filter(Year >= 1991 & Year<= 2014 & (Country.Code =='IND' | Country.Code == 'CAN'| Country.Cod
e == 'CHN' | Country.Code == 'USA')) %>%
  group_by(Year)

ggplot(cpi_data_1,aes(x = Year, y = CPI, col = Country.Code, group = Country.Code)) +geom_line(n
a.rm = T,lwd = 1)+ theme(axis.text.x=element_text(angle=90, hjust=1))
```




There is a steep CPI increase for India and the CPI of USA and Canada are comparitively more stable post the 2008 recession.

```
unemployment_data_1 <- labor_data %>%
  select(Country.Name,Country.Code,Year,unemp) %>%
  filter(Year >= 1991 & Year<= 2014 & (Country.Code =='IND' | Country.Code == 'CAN' | Country.Co
de == 'USA'|Country.Code == 'CHN')) %>%
  group_by(Year)

ggplot(unemployment_data_1, aes(x = Year, y = unemp,col = Country.Code, group = Country.Code))+g
eom_line(na.rm = T,lwd = 1)+ theme(axis.text.x=element_text(angle=90, hjust=1))
```




We see that the unemployment rates increases drastically in Canada and USA after 2007 crisis(2008 recession). The unemployment rates in China is stable wheras for India it decreases after the 2007 crisis.This can be due to increase in outsourcing.

What are the other salient aspects of the data (e.g. geospatial factors, text content, etc.)

The following plots are forecasts for each of the selected country's future unemployment rates based on the 2000-2009 data. This provides a projection of what the unemployment rates might have looked like in the absence of the financial

```
#canada
#install.packages(forecast)
library(forecast)
```

```
## Loading required package: zoo
```

```
##
## Attaching package: 'zoo'
```

```
## The following objects are masked from 'package:base':
##
##     as.Date, as.Date.numeric
```

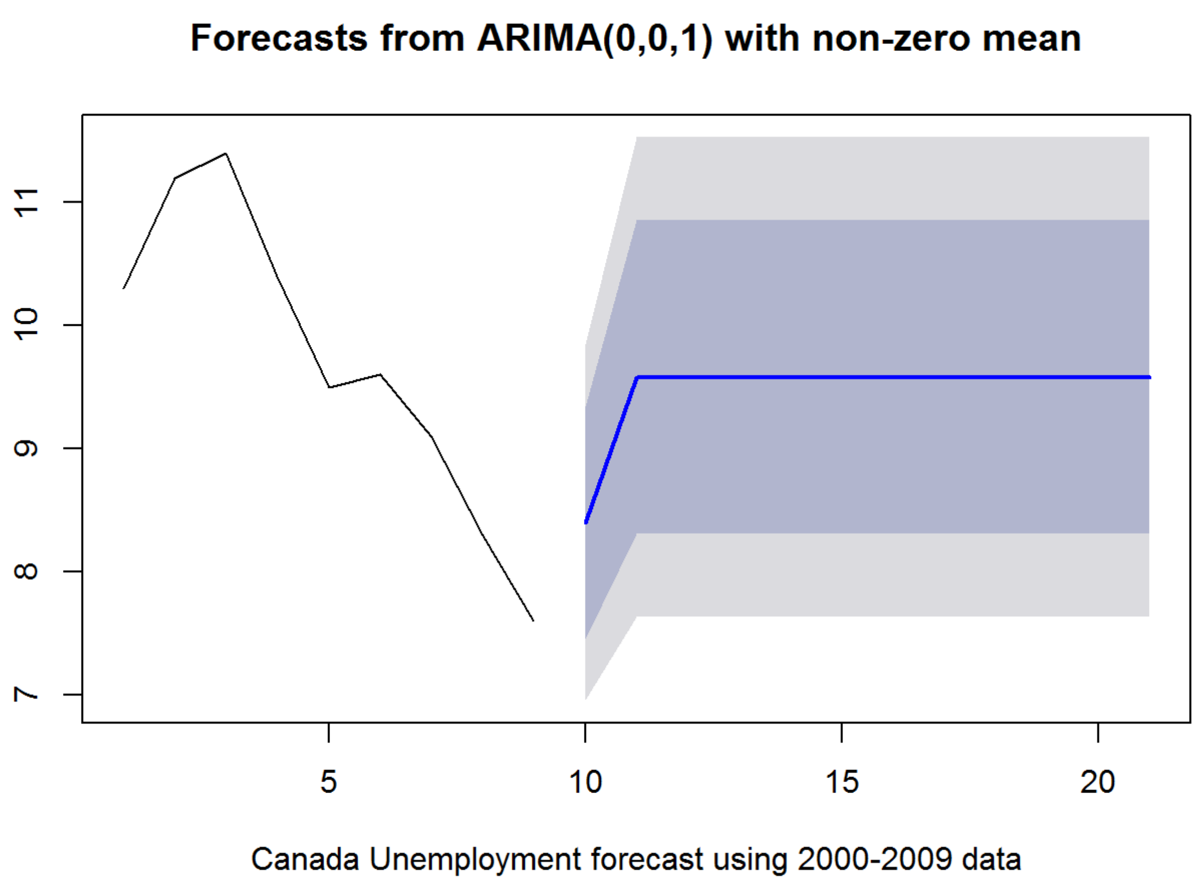
```
## Loading required package: timeDate
```

```
## This is forecast 7.3
```

```
unemployment_can <- unemployment_data_1[unemployment_data_1$Country.Code=="CAN",]
auto.arima(unemployment_can$unemp)
```

```
## Series: unemployment_can$unemp
## ARIMA(0,1,0)
##
## sigma^2 estimated as 0.4852:  log likelihood=-24.32
## AIC=50.64   AICc=50.83   BIC=51.77
```

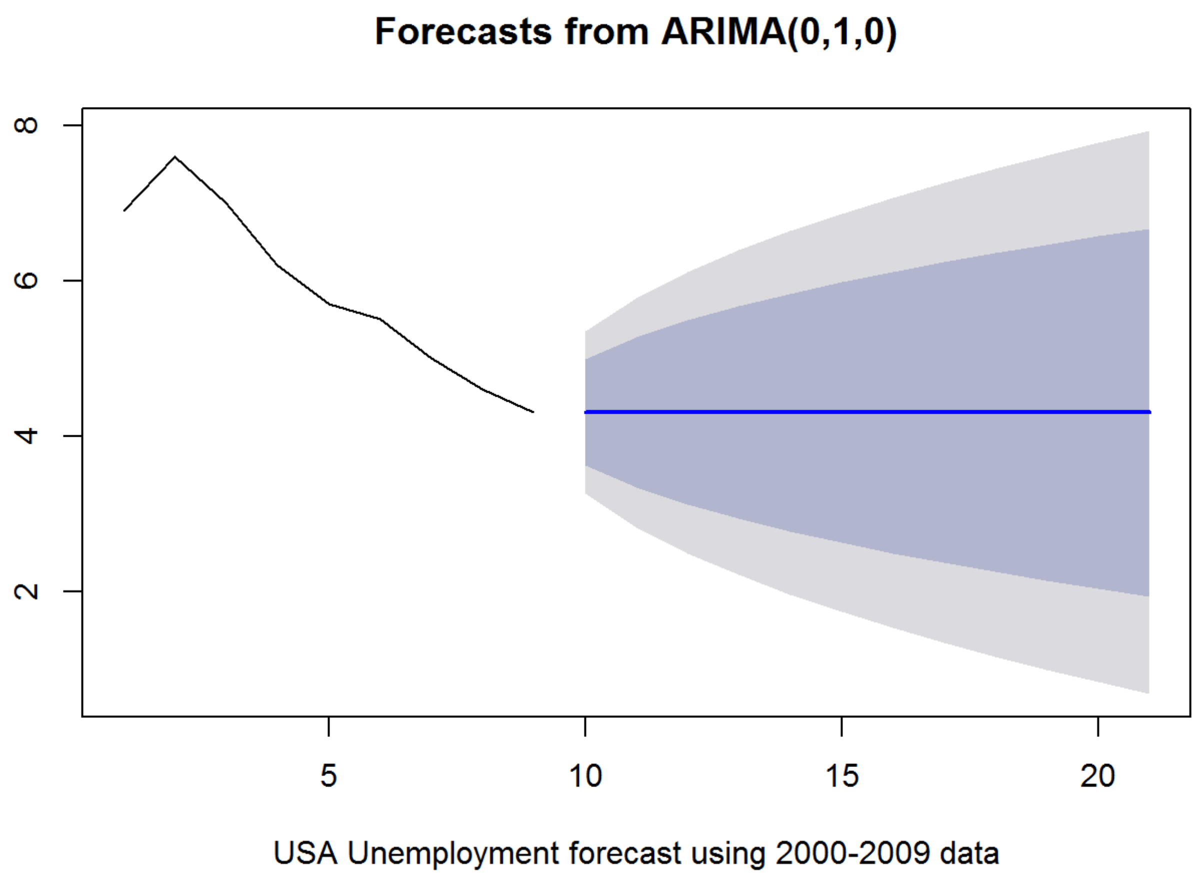
```
fit <-arima(x=unemployment_can$unemp[1:9], order=c(0,0,1))
preds <-forecast.Arima(fit, h=12)
plot.forecast(preds, xlab = "Canada Unemployment forecast using 2000-2009 data")
```




```
#us
unemployment_us <- unemployment_data_1[unemployment_data_1$Country.Code=="USA",]
auto.arima(unemployment_us$unemp)
```

```
## Series: unemployment_us$unemp
## ARIMA(1,0,1) with non-zero mean
##
## Coefficients:
##          ar1     ma1  intercept
##       0.6785  0.5644     6.2001
## s.e.  0.1466  0.1502     0.6903
##
## sigma^2 estimated as 0.6643:  log likelihood=-28.37
## AIC=64.74   AICc=66.84   BIC=69.45
```

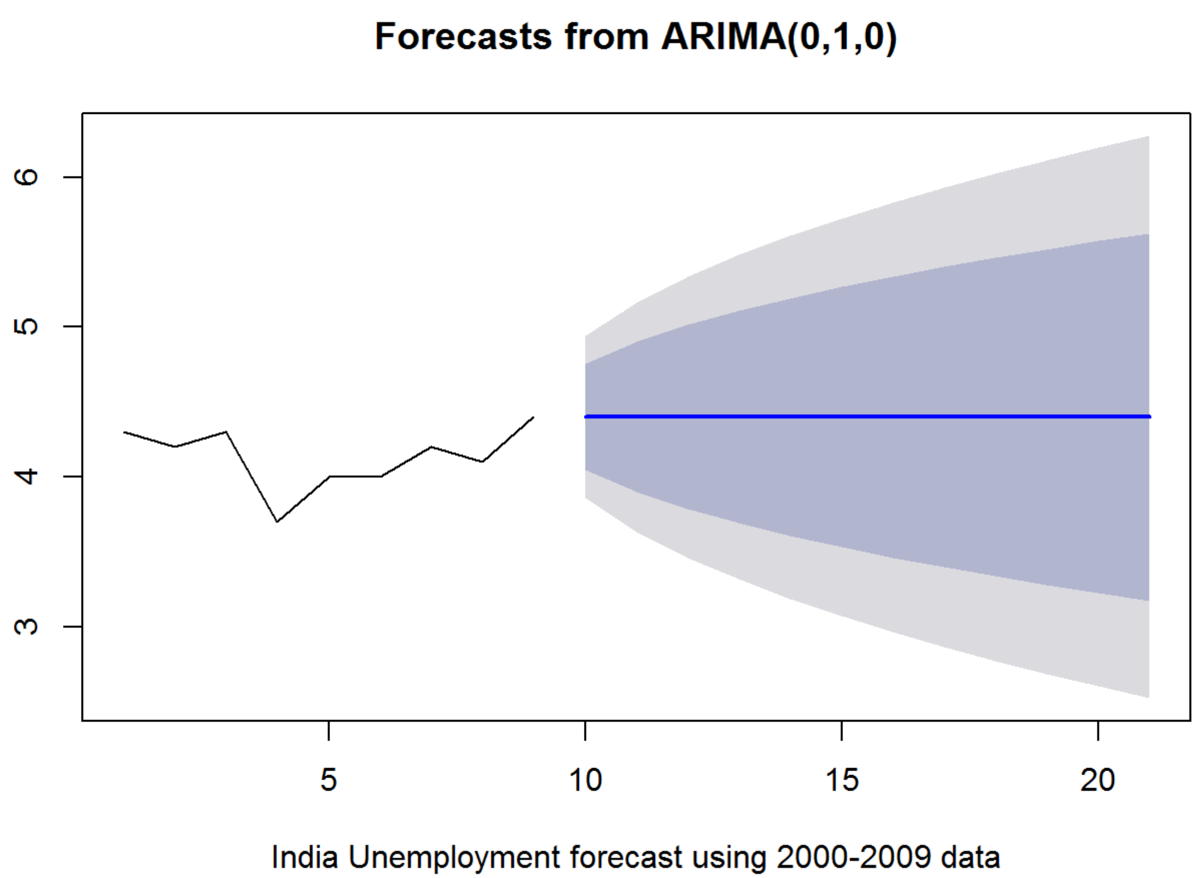
```
fit <-arima(x=unemployment_us$unemp[1:9], order=c(0,1,0))
preds <-forecast.Arima(fit, h=12)
plot.forecast(preds, xlab = "USA Unemployment forecast using 2000-2009 data")
```




```
#India
unemployment_ind <- unemployment_data_1[unemployment_data_1$Country.Code=="IND",]
auto.arima(unemployment_ind$unemp)
```

```
## Series: unemployment_ind$unemp
## ARIMA(0,1,1)
##
## Coefficients:
##           ma1
##       -0.5583
## s.e.   0.1779
##
## sigma^2 estimated as 0.06733:  log likelihood=-1.28
## AIC=6.56   AICc=7.16   BIC=8.84
```

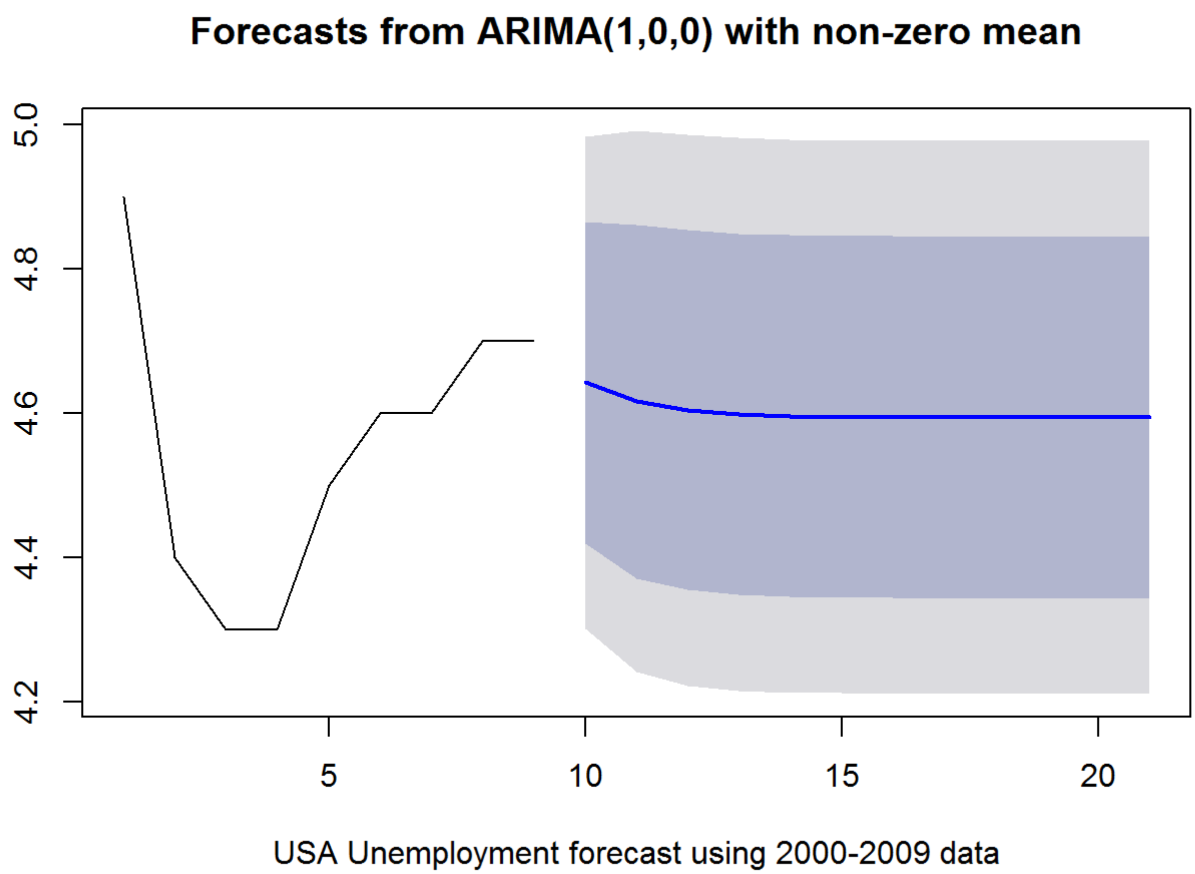
```
fit <-arima(x=unemployment_ind$unemp[1:9], order=c(0,1,0))
preds <-forecast.Arima(fit, h=12)
plot.forecast(preds, xlab = "India Unemployment forecast using 2000-2009 data")
```




```
#china
unemployment_china <- unemployment_data_1[unemployment_data_1$Country.Code=="CHN",]
auto.arima(unemployment_china$unemp)
```

```
## Series: unemployment_china$unemp
## ARIMA(1,0,0) with non-zero mean
##
## Coefficients:
##          ar1  intercept
##       0.6740     4.4730
## s.e.  0.1679     0.1142
##
## sigma^2 estimated as 0.03814:  log likelihood=5.89
## AIC=-5.77   AICc=-4.57   BIC=-2.24
```

```
fit <-arima(x=unemployment_china$unemp[1:9], order=c(1,0,0))
preds <-forecast.Arima(fit, h=12)
plot.forecast(preds, xlab = "USA Unemployment forecast using 2000-2009 data")
```

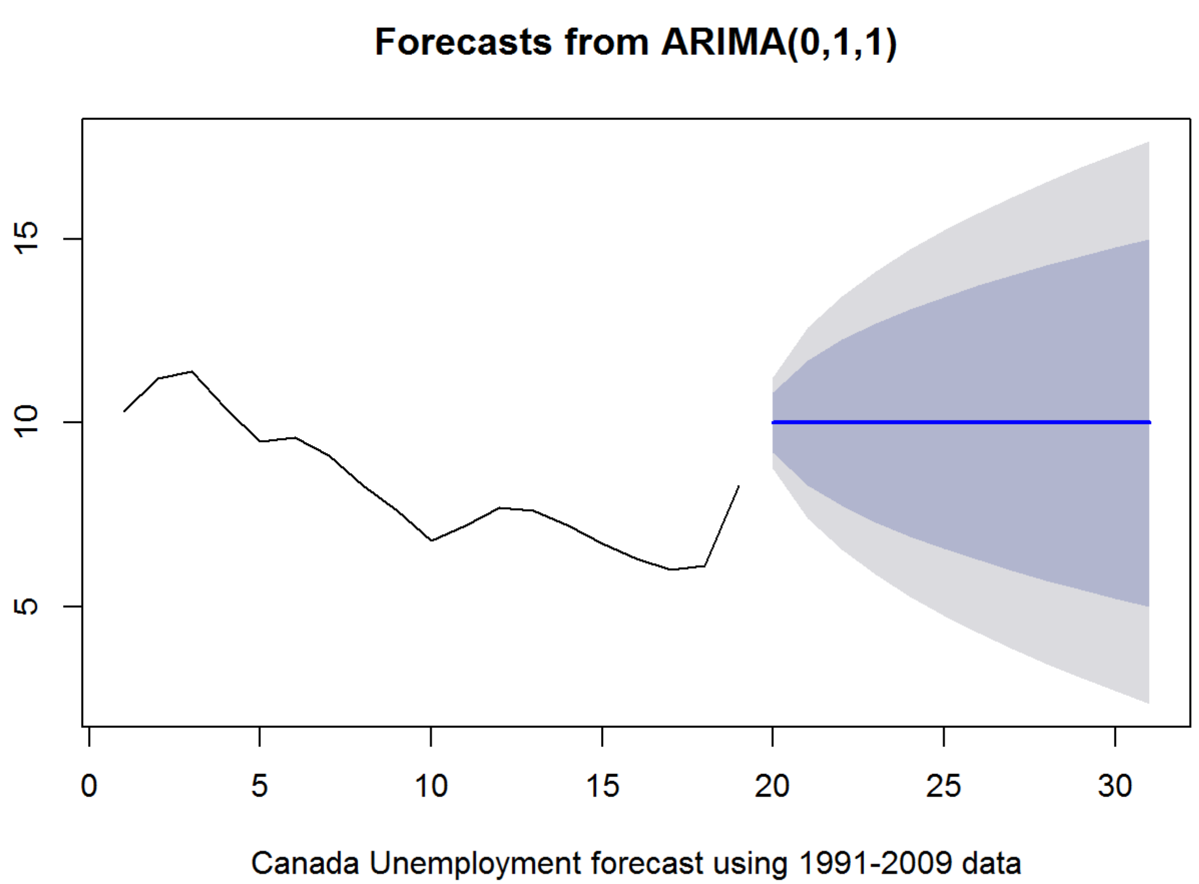



Given the different models auto.arima recommended above, a second attempt follows utilizing the unemployment data from 1991-2009:

```
unemployment_data_2 <- country_data %>%
  select(Country.Name,Country.Code,Year,unemp) %>%
  filter((Country.Code =='IND' | Country.Code == 'CAN' | Country.Code == 'USA'|Country.Code ==
'CHN')) %>%
  group_by(Year)
#canada
unemployment_can <- unemployment_data_2[unemployment_data_2$Country.Code=="CAN",]
auto.arima(unemployment_can$unemp[32:50])
```

```
## Series: unemployment_can$unemp[32:50]
## ARIMA(0,1,1)
##
## Coefficients:
##          ma1
##       0.8315
## s.e.  0.2080
##
## sigma^2 estimated as 0.427:  log likelihood=-17.96
## AIC=39.91   AICc=40.71   BIC=41.69
```

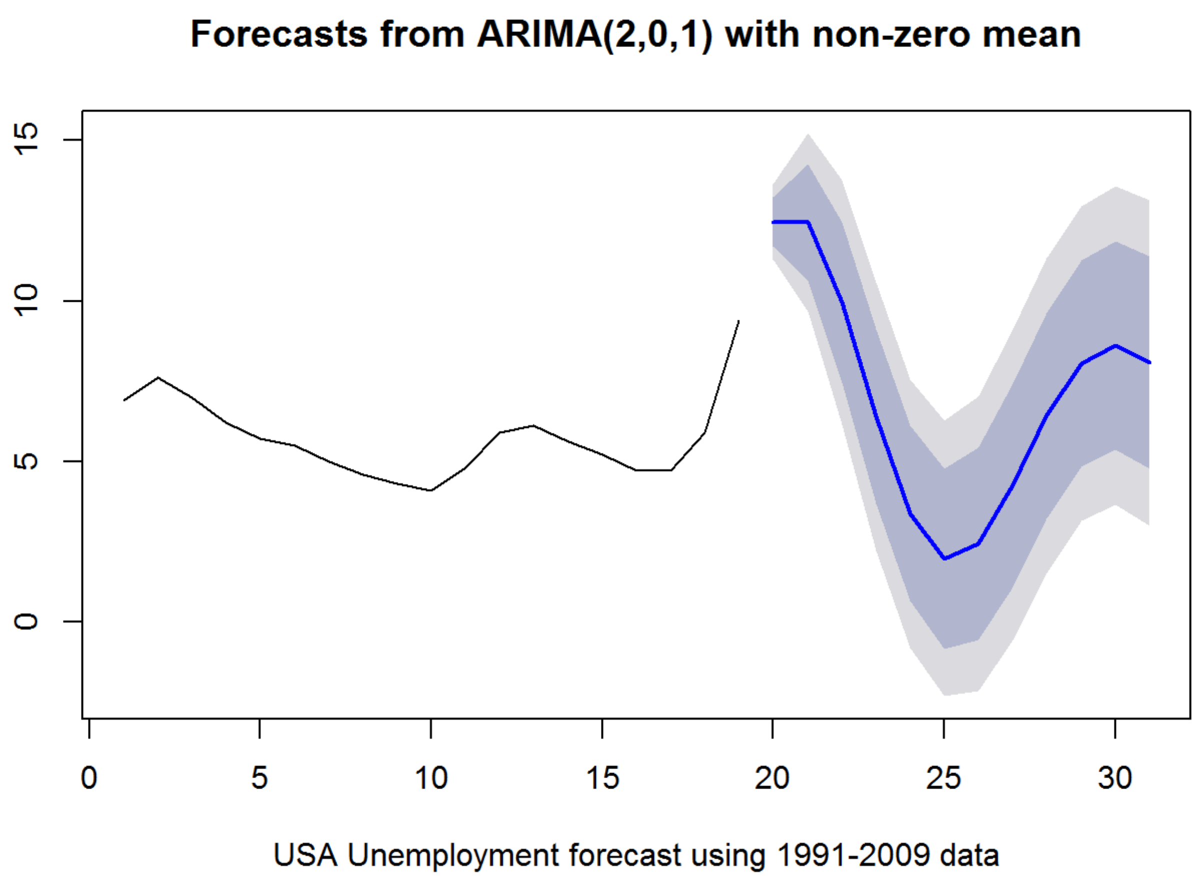
```
fit <-arima(x=unemployment_can$unemp[32:50], order=c(0,1,1))
preds <-forecast.Arima(fit, h=12)
plot.forecast(preds, xlab = "Canada Unemployment forecast using 1991-2009 data")
```



```
#us
unemployment_us <- unemployment_data_2[unemployment_data_2$Country.Code=="USA",]
auto.arima(unemployment_us$unemp[32:50])
```

```
## Series: unemployment_us$unemp[32:50]
## ARIMA(2,0,1) with non-zero mean
##
## Coefficients:
##          ar1      ar2     ma1  intercept
##       1.4230  -0.8141  0.7209     6.0716
## s.e.  0.2525   0.1796  0.3183     0.6800
##
## sigma^2 estimated as 0.4525:  log likelihood=-20.01
## AIC=50.02   AICc=54.63   BIC=54.74
```

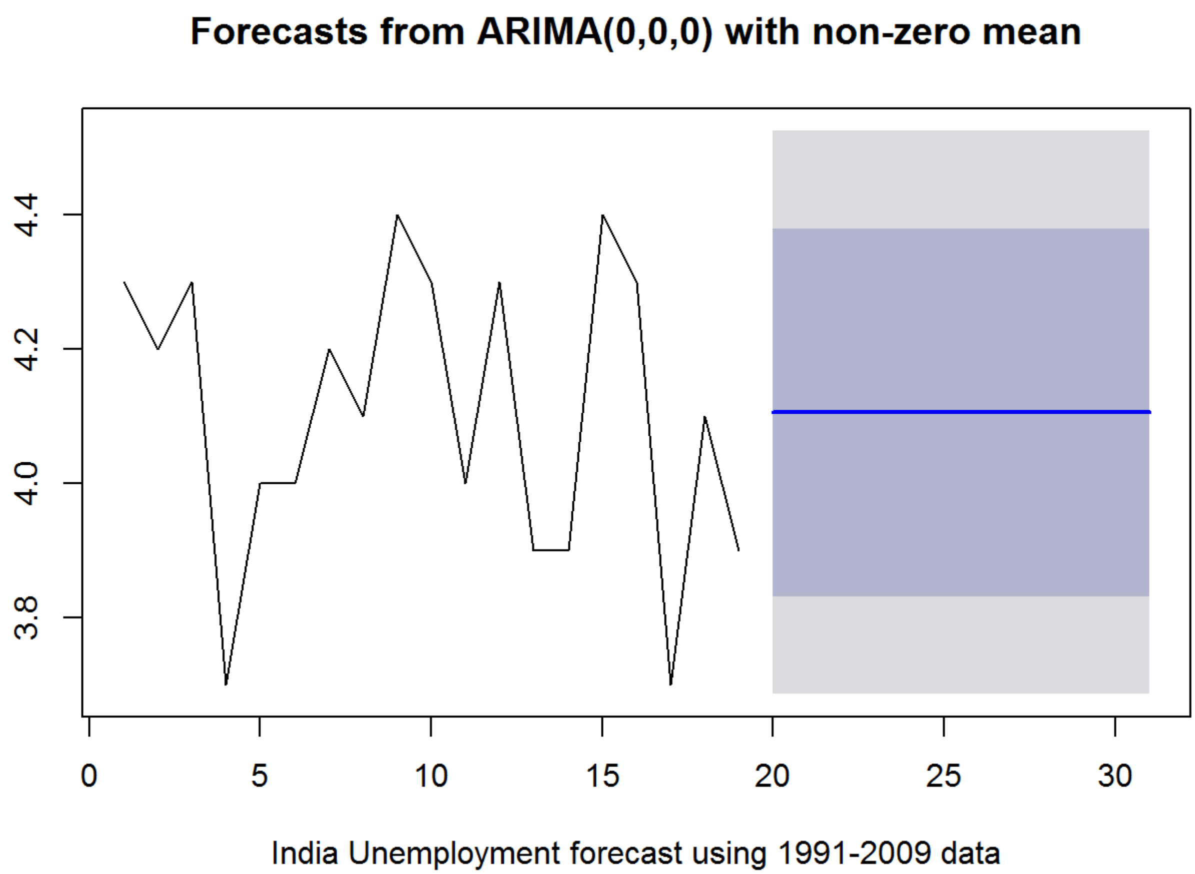
```
fit <-arima(x=unemployment_us$unemp[32:50], order=c(2,0,1))
preds <-forecast.Arima(fit, h=12)
plot.forecast(preds, xlab = "USA Unemployment forecast using 1991-2009 data")
```




```
#India
unemployment_ind <- unemployment_data_2[unemployment_data_2$Country.Code=="IND",]
auto.arima(unemployment_ind$unemp[32:50])
```

```
## Series: unemployment_ind$unemp[32:50]
## ARIMA(0,0,0) with non-zero mean
##
## Coefficients:
##       intercept
##          4.1053
## s.e.     0.0491
##
## sigma^2 estimated as 0.0483:  log likelihood=2.34
## AIC=-0.68   AICc=0.07   BIC=1.21
```

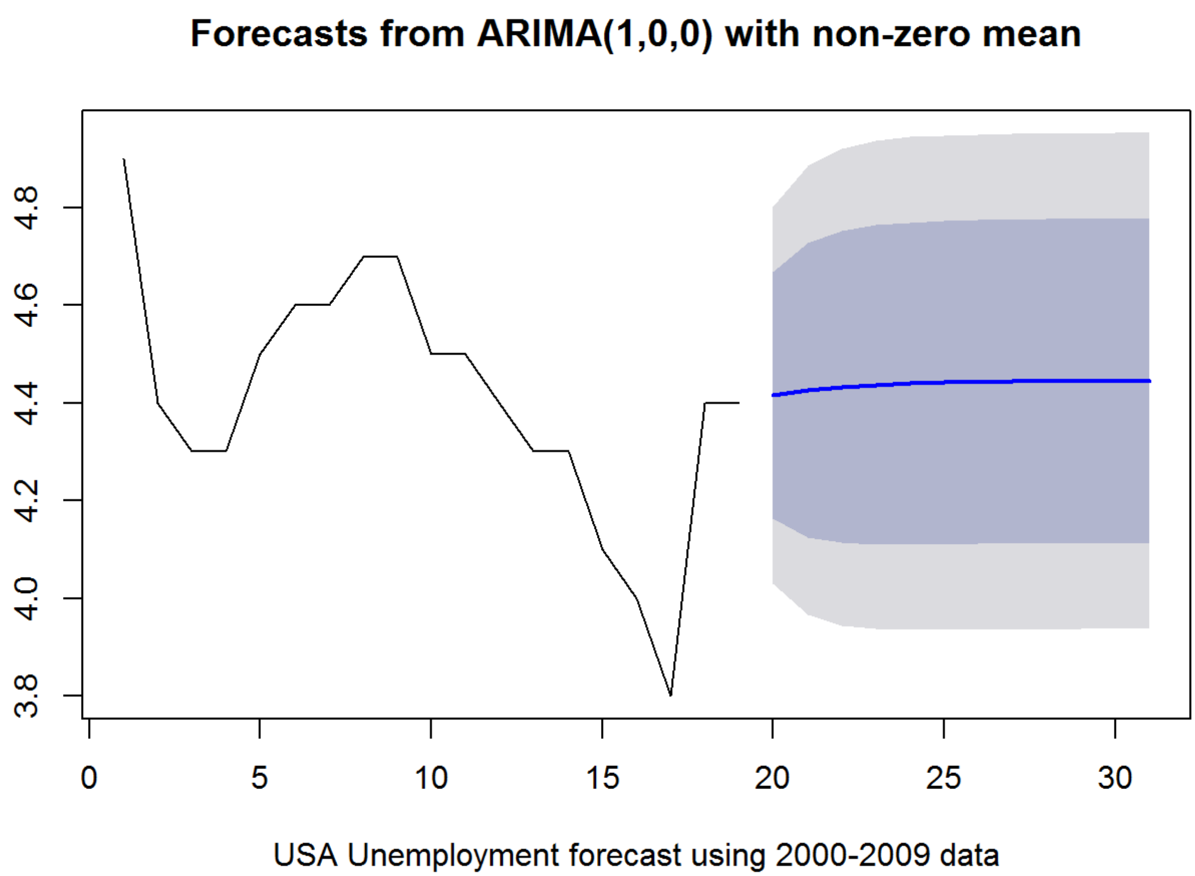
```
fit <-arima(x=unemployment_ind$unemp[32:50], order=c(0,0,0))
preds <-forecast.Arima(fit, h=12)
plot.forecast(preds, xlab = "India Unemployment forecast using 1991-2009 data")
```




```
#china
unemployment_china <- unemployment_data_2[unemployment_data_2$Country.Code=="CHN",]
auto.arima(unemployment_china$unemp[32:50])
```

```
## Series: unemployment_china$unemp[32:50]
## ARIMA(0,1,0)
##
## sigma^2 estimated as 0.04611:  log likelihood=2.15
## AIC=-2.3   AICc=-2.05   BIC=-1.41
```

```
fit <-arima(x=unemployment_china$unemp[32:50], order=c(1,0,0))
preds <-forecast.Arima(fit, h=12)
plot.forecast(preds, xlab = "USA Unemployment forecast using 2000-2009 data")
```




Attempt to find trends in the CPI data. This data set contains CPI for 265 countries over a period of 56 years. The variable CPI is time sensitive.

```
CPI_trend<-read.csv("CPI_World_Bank.csv")
#plot(CPI_trend$)
colnames(CPI_trend)
```

```
##  [1] "Country.Name"   "Country.Code"   "Indicator.Name" "Indicator.Code"
##  [5] "X1960"          "X1961"          "X1962"          "X1963"
##  [9] "X1964"          "X1965"          "X1966"          "X1967"
## [13] "X1968"          "X1969"          "X1970"          "X1971"
## [17] "X1972"          "X1973"          "X1974"          "X1975"
## [21] "X1976"          "X1977"          "X1978"          "X1979"
## [25] "X1980"          "X1981"          "X1982"          "X1983"
## [29] "X1984"          "X1985"          "X1986"          "X1987"
## [33] "X1988"          "X1989"          "X1990"          "X1991"
## [37] "X1992"          "X1993"          "X1994"          "X1995"
## [41] "X1996"          "X1997"          "X1998"          "X1999"
## [45] "X2000"          "X2001"          "X2002"          "X2003"
## [49] "X2004"          "X2005"          "X2006"          "X2007"
## [53] "X2008"          "X2009"          "X2010"          "X2011"
## [57] "X2012"          "X2013"          "X2014"          "X2015"
## [61] "X2016"
```

```
#goal here is time period on x and CPI on y.
trends<-read.csv("Timeseries.csv")
class(trends)
```

```
## [1] "data.frame"
```

```
#summary(lm(Time ~ CPI_USA, data=trends))
#plot(trends)
#abline(lsfit(x=trends$Time,y=trends$CPI_USA),col="red")
#install.packages('tseries')
require(tseries)
```

```
## Loading required package: tseries
```

```
#install.packages('xts')
require(xts)
```

```
## Loading required package: xts
```

```
##
## Attaching package: 'xts'
```

```
## The following objects are masked from 'package:dplyr':
##
##     first, last
```

```
#converting data into time series data
trends.ts<-ts(as.vector(trends), start = c(1960,1), end = c(2015),frequency = 1)
trends.ts
```

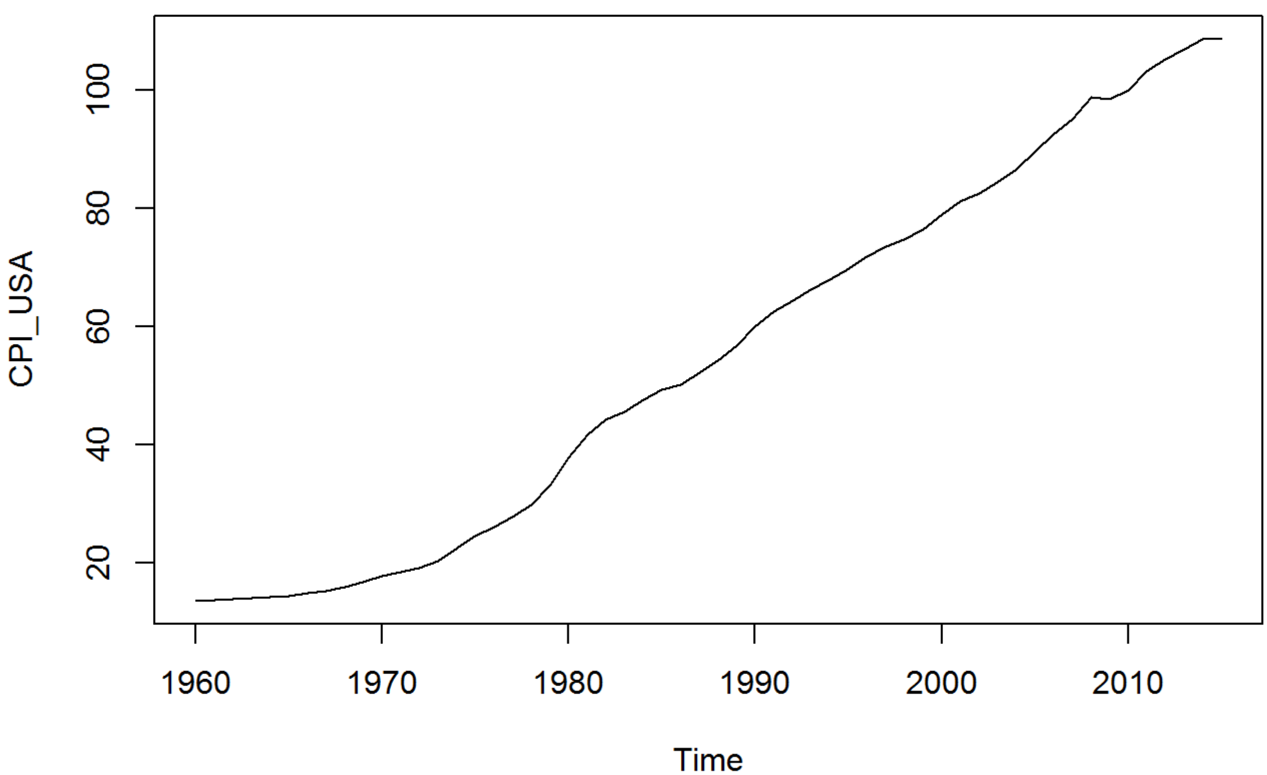
```
## Time Series:
## Start = 1960
## End = 2015
## Frequency = 1
##         CPI_USA
##  [1,]  13.57217
##  [2,]  13.71809
##  [3,]  13.87120
##  [4,]  14.03961
##  [5,]  14.22333
##  [6,]  14.46064
##  [7,]  14.89316
##  [8,]  15.30654
##  [9,]  15.95213
## [10,]  16.81589
## [11,]  17.80724
## [12,]  18.56510
## [13,]  19.17879
## [14,]  20.37172
## [15,]  22.61980
## [16,]  24.68542
## [17,]  26.10163
## [18,]  27.79470
## [19,]  29.92029
## [20,]  33.29112
## [21,]  37.78854
## [22,]  41.68663
## [23,]  44.25479
## [24,]  45.67644
## [25,]  47.64842
## [26,]  49.34524
## [27,]  50.26243
## [28,]  52.14269
## [29,]  54.23313
## [30,]  56.85097
## [31,]  59.91976
## [32,]  62.45734
## [33,]  64.34906
## [34,]  66.24842
## [35,]  67.97581
## [36,]  69.88282
## [37,]  71.93123
## [38,]  73.61276
## [39,]  74.75543
## [40,]  76.39110
## [41,]  78.97072
## [42,]  81.20257
## [43,]  82.49047
## [44,]  84.36308
## [45,]  86.62168
## [46,]  89.56053
## [47,]  92.44971
## [48,]  95.08699
```

```
## [49,]  98.73748
## [50,]  98.38642
## [51,] 100.00000
## [52,] 103.15684
## [53,] 105.29150
## [54,] 106.83385
## [55,] 108.56693
## [56,] 108.69572
```

```
class(trends.ts)
```

```
## [1] "ts"
```

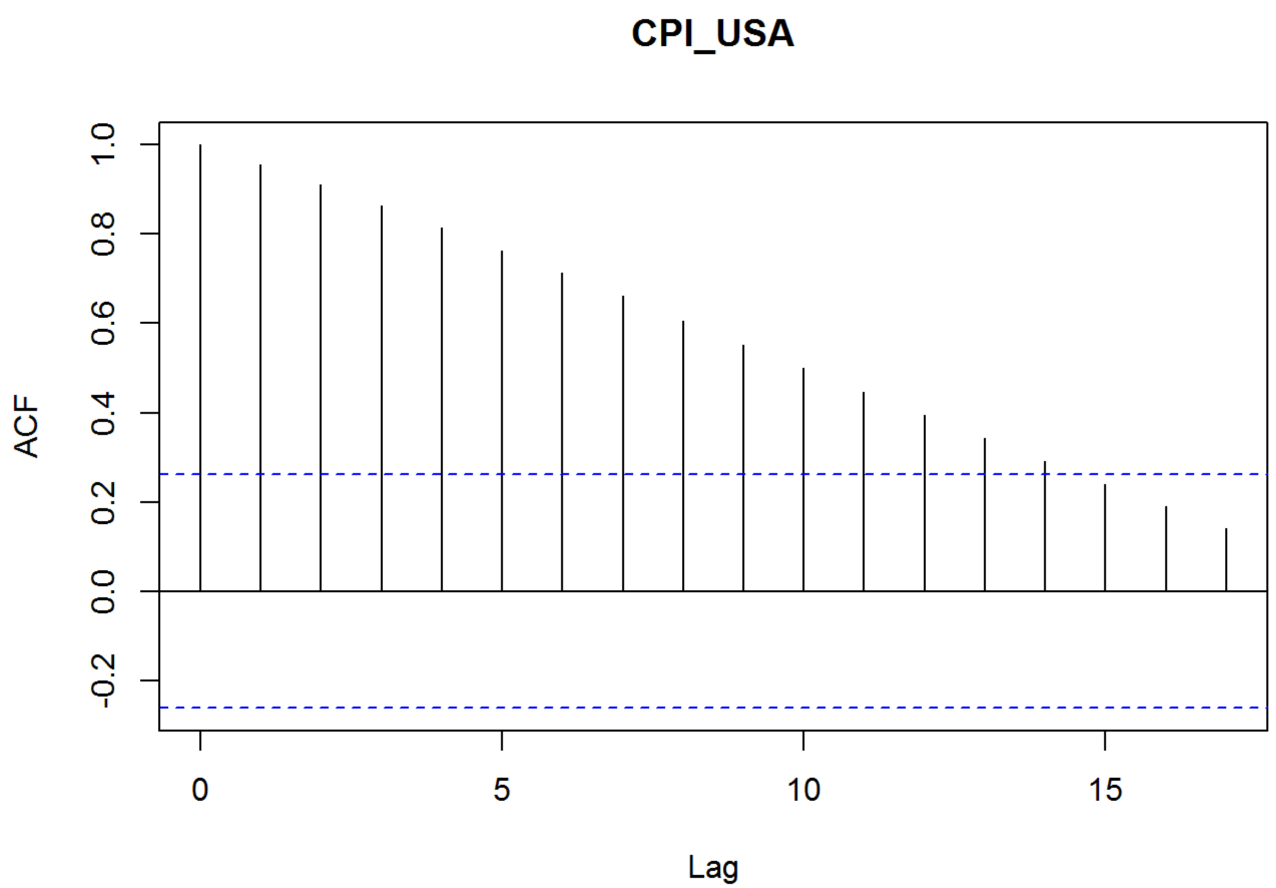
```
plot(trends.ts)
```



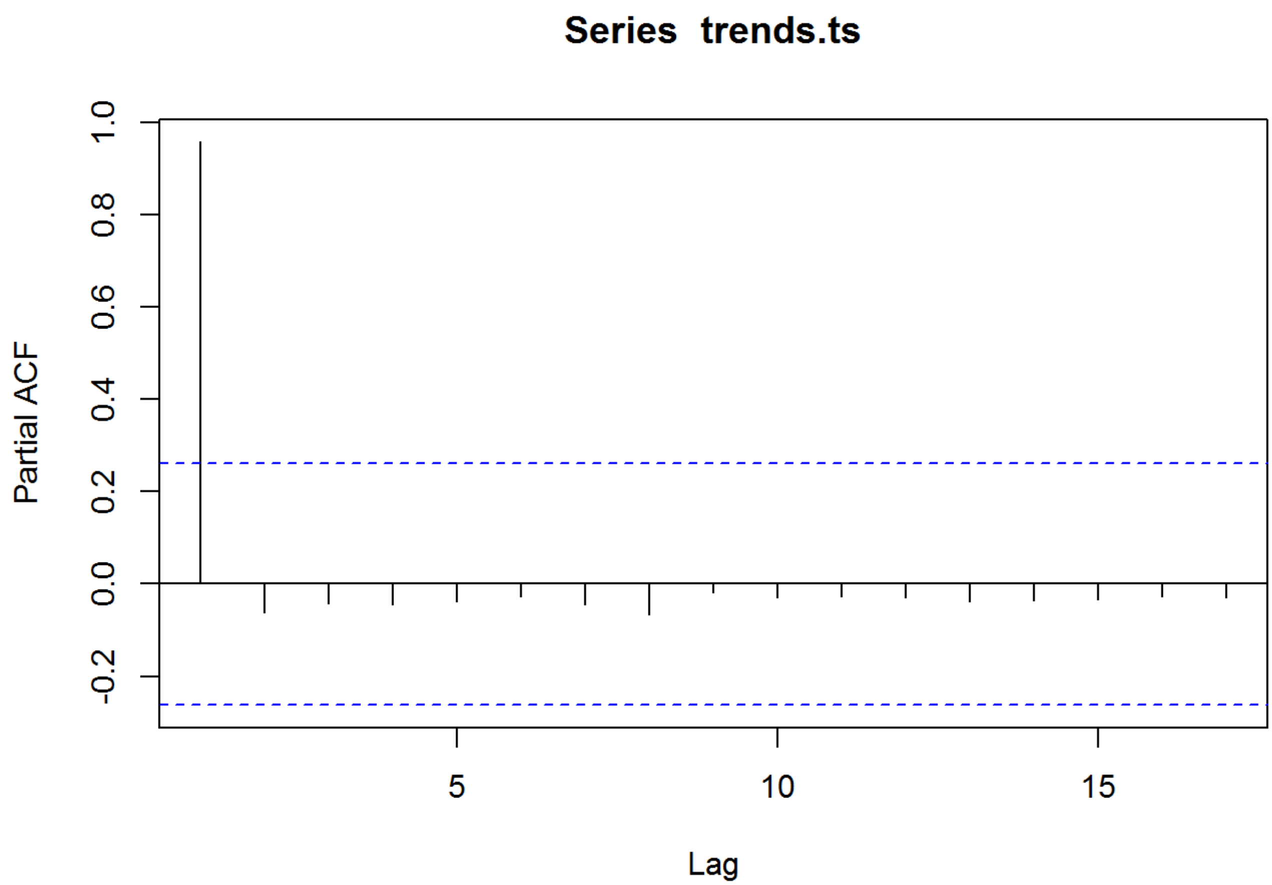

```
#Dickey-fuller test to see if the time series is stationary.
adf.test(trends.ts)
```

```
##
##  Augmented Dickey-Fuller Test
##
## data:  trends.ts
## Dickey-Fuller = -3.2846, Lag order = 3, p-value = 0.08307
## alternative hypothesis: stationary
```

```
#Checking for p and q values through auto correlation function and partial auto correlation func
tion.
acf(trends.ts)
```




```
pacf(trends.ts)
```



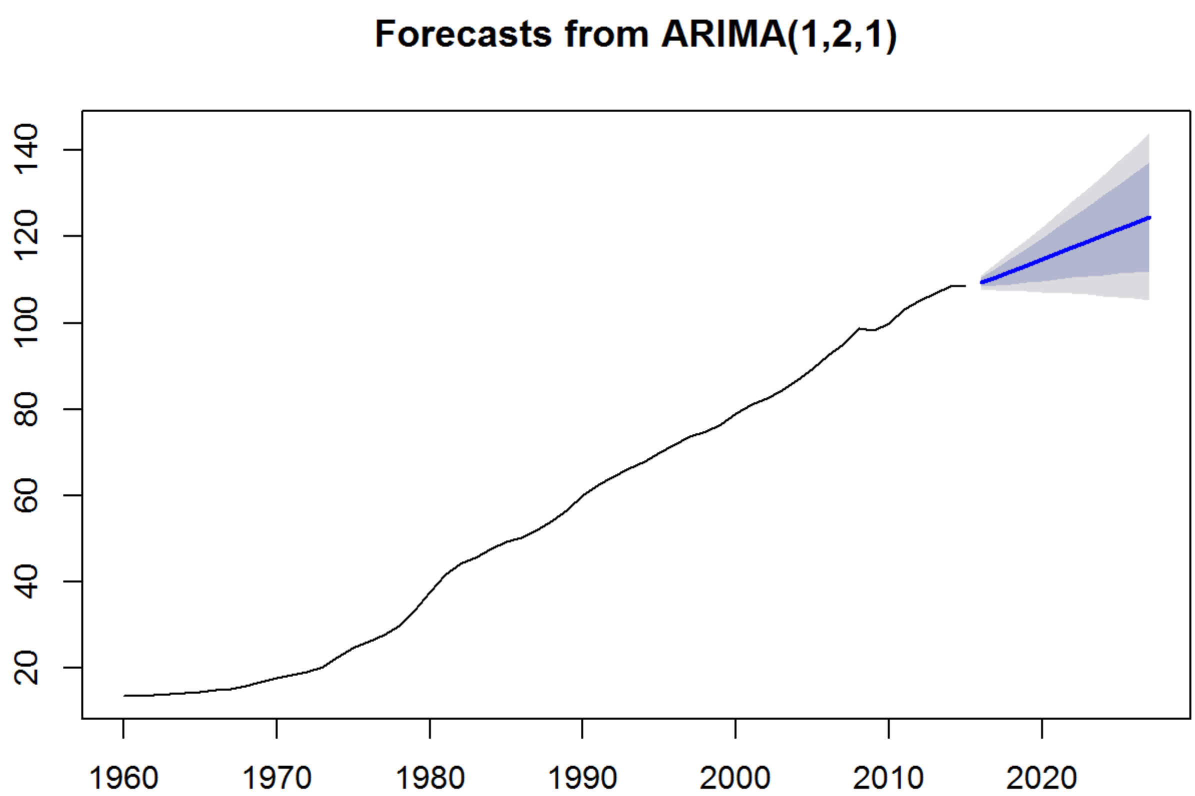

```r
#this is an AR model
#install.packages('forecast')
library('forecast')
auto.arima(trends.ts)
```

```
## Series: trends.ts
## ARIMA(1,2,1)
##
## Coefficients:
##          ar1      ma1
##       0.4603  -0.8296
## s.e.  0.2091   0.1391
##
## sigma^2 estimated as 0.7067:  log likelihood=-66.45
## AIC=138.9   AICc=139.38   BIC=144.87
```

```r
trends.arimausa<- arima(trends.ts, order=c(1,2,1))
trends.arimausa
```

```
##
## Call:
## arima(x = trends.ts, order = c(1, 2, 1))
##
## Coefficients:
##          ar1      ma1
##       0.4603  -0.8296
## s.e.  0.2091   0.1391
##
## sigma^2 estimated as 0.6805:  log likelihood = -66.45,  aic = 138.9
```

```
trend.preds<-forecast.Arima(trends.arimausa,h=12)
plot.forecast(trend.preds)
```




```
summary(trend.preds)
```

```
##
## Forecast method: ARIMA(1,2,1)
##
## Model Information:
##
## Call:
## arima(x = trends.ts, order = c(1, 2, 1))
##
## Coefficients:
##          ar1      ma1
##       0.4603  -0.8296
## s.e.  0.2091   0.1391
##
## sigma^2 estimated as 0.6805:  log likelihood = -66.45,  aic = 138.9
##
## Error measures:
##                      ME     RMSE       MAE       MPE     MAPE      MASE
## Training set 0.06899142 0.810064 0.5382748 0.4357933 1.112776 0.3089477
##                    ACF1
## Training set 0.02865192
##
## Forecasts:
##      Point Forecast    Lo 80    Hi 80    Lo 95    Hi 95
## 2016       109.5173 108.4601 110.5745 107.9005 111.1341
## 2017       110.6577 108.6354 112.6801 107.5648 113.7506
## 2018       111.9450 108.9485 114.9415 107.3623 116.5278
## 2019       113.2999 109.3225 117.2773 107.2169 119.3828
## 2020       114.6859 109.7148 119.6569 107.0833 122.2884
## 2021       116.0861 110.1025 122.0698 106.9349 125.2374
## 2022       117.4930 110.4728 124.5132 106.7566 128.2295
## 2023       118.9030 110.8190 126.9869 106.5396 131.2663
## 2024       120.3143 111.1373 129.4913 106.2793 134.3493
## 2025       121.7262 111.4257 132.0268 105.9729 137.4796
## 2026       123.1385 111.6832 134.5938 105.6191 140.6579
## 2027       124.5509 111.9094 137.1924 105.2174 143.8844
```

```
#Hypothesis testing for USA CPI data
#install.packages('lmtest')
library('lmtest')
coeftest(trends.arimausa)
```

```
##
## z test of coefficients:
##
##     Estimate Std. Error z value  Pr(>|z|)
## ar1  0.46035    0.20911  2.2014   0.02771 *
## ma1 -0.82959    0.13908 -5.9649 2.447e-09 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
#z value is 2.2014, p=0.02771
#another way to obtain p value to verify
(1-pnorm(abs(trends.arimausa$coef)/sqrt(diag(trends.arimausa$var.coef))))*2
```

```
##          ar1          ma1
## 2.770520e-02 2.447359e-09
```