EDA Assignment

Abhi

July 11, 2019

Data Description

life = life expectancy income = gdp per capita year = the corresponding year for the data shown county = different countries region = location of the mentioned countries population = census data collected about every 10 years

Data summary

No. of observations = 41284 No. of variables = 7

Type of variables -: country - factor variable showing 197 different countries year - integer variable having a range from 1800 to 2015 life - numeric variable having a range from 1.0 year to 84.10 years population - factor variable with 15260 levels income - integer variable having a range from 1$ to 182668$ region - factor variable dividing the regions of the world into 6 commonly accepted geographical locations

Pre-processing the data

The data used is not normalized or standardized because only first order data visualization is our aim and to find for definite trends .

Questions :-

1. Which countries in Europe and Central Asia were impacted most by WW1 and WW2 in terms of Life Expectacy and drop in per capita income?
2. Was the impact on life more than the income?
3. Which countries showed the quickest recovery after the wars in terms of economic development?

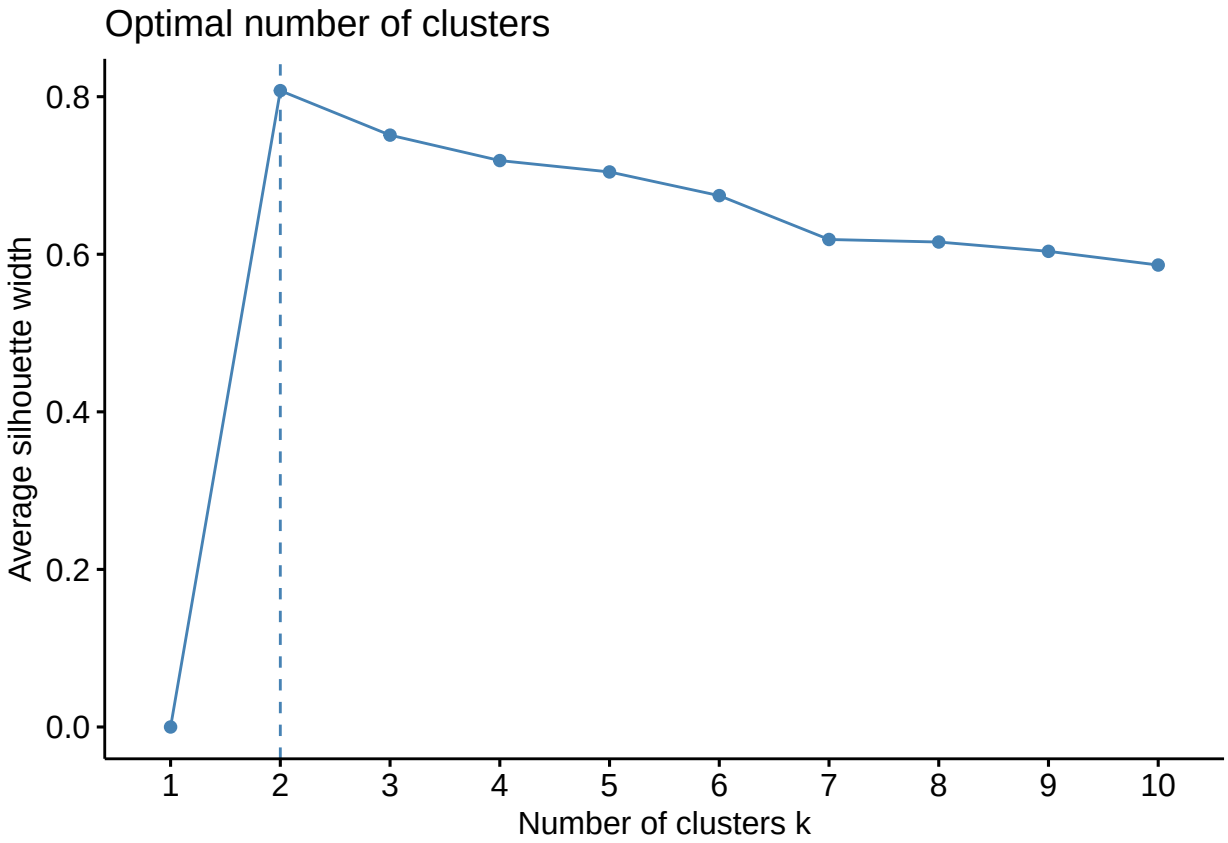
Data wrangling

We can see a lot of blank values in the population column. It is because earlier the census were done once in 10 years. For the ease of analysis here we assume the population to be constant till the next survey is done.

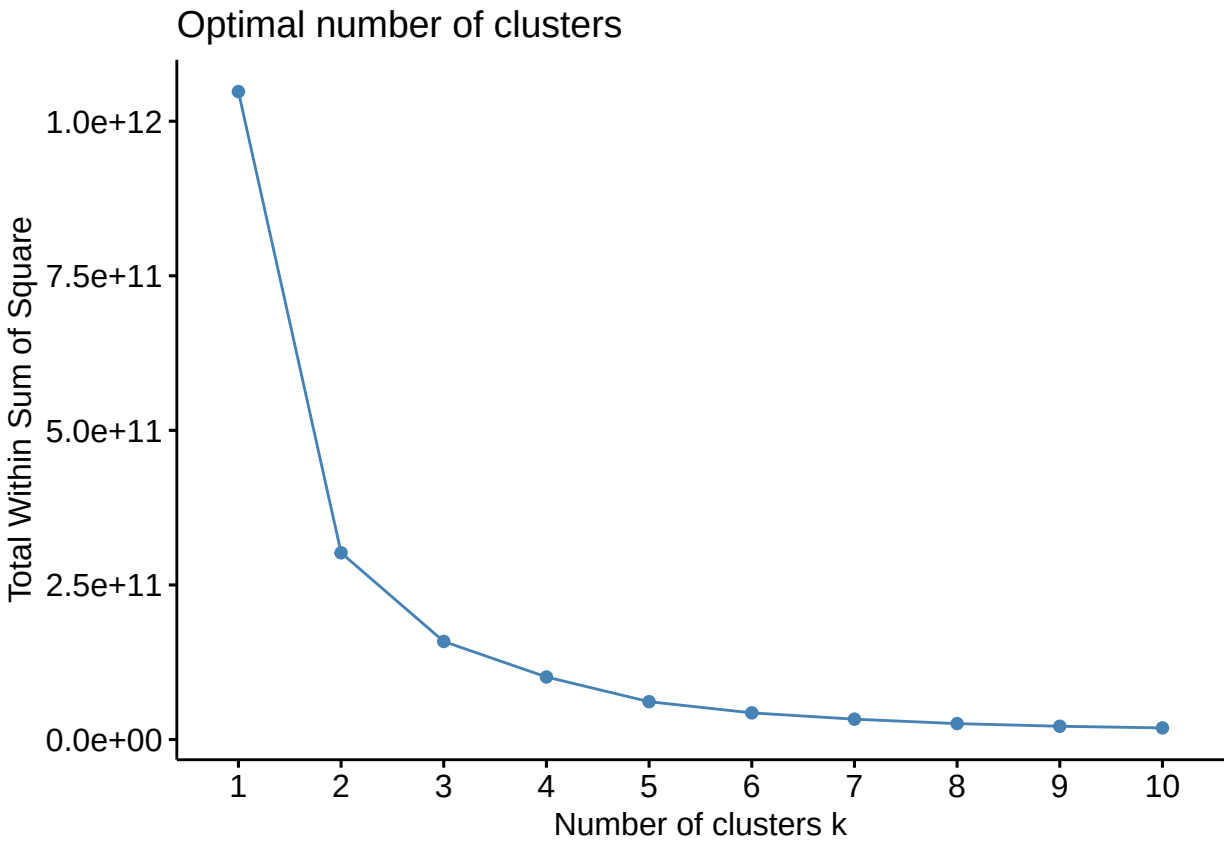
Cluster Analysis

Finding the optimum number of clusters by

A) Silhouette Method

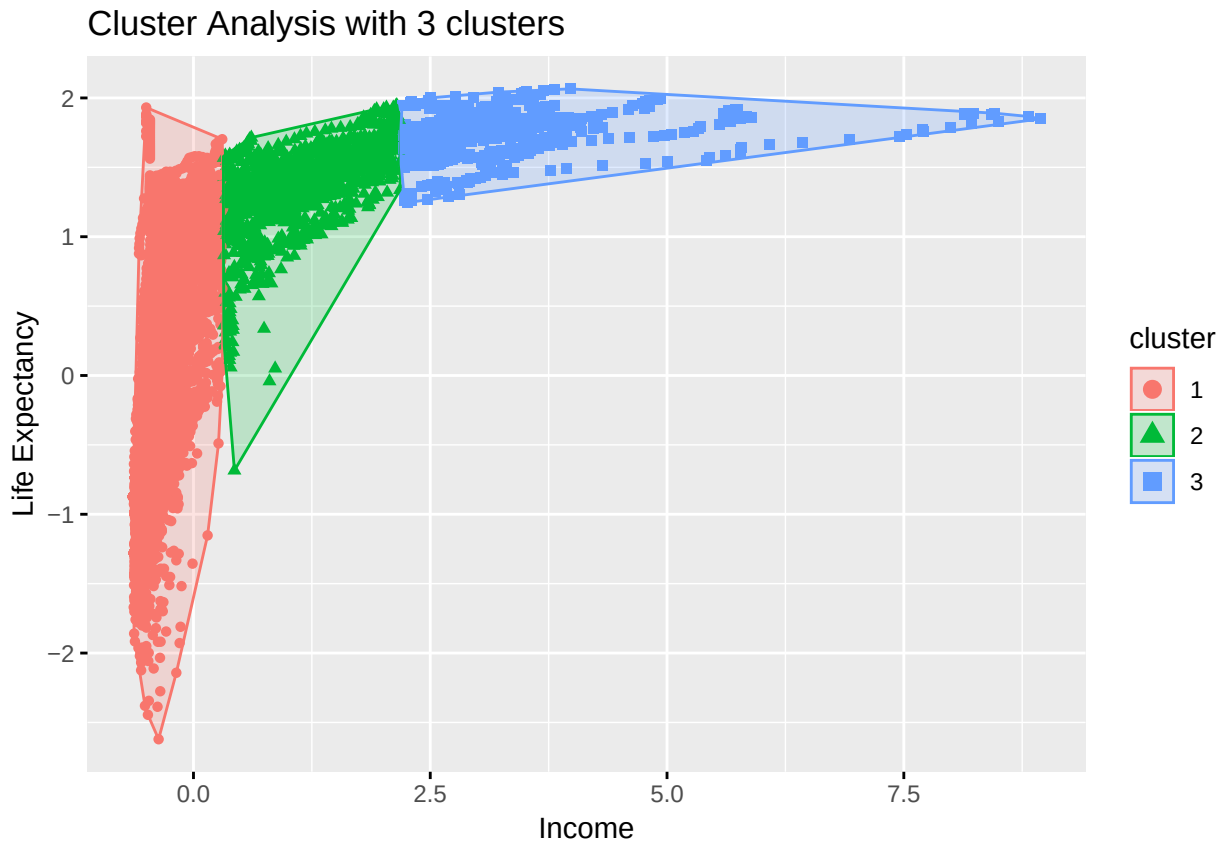


B) Elbow Method



By applying the above methods to determine the optimum number of clusters we see that we can chose both 2 and 3 as the number of clusters.

Visualizing with both 2 and 3 clusters



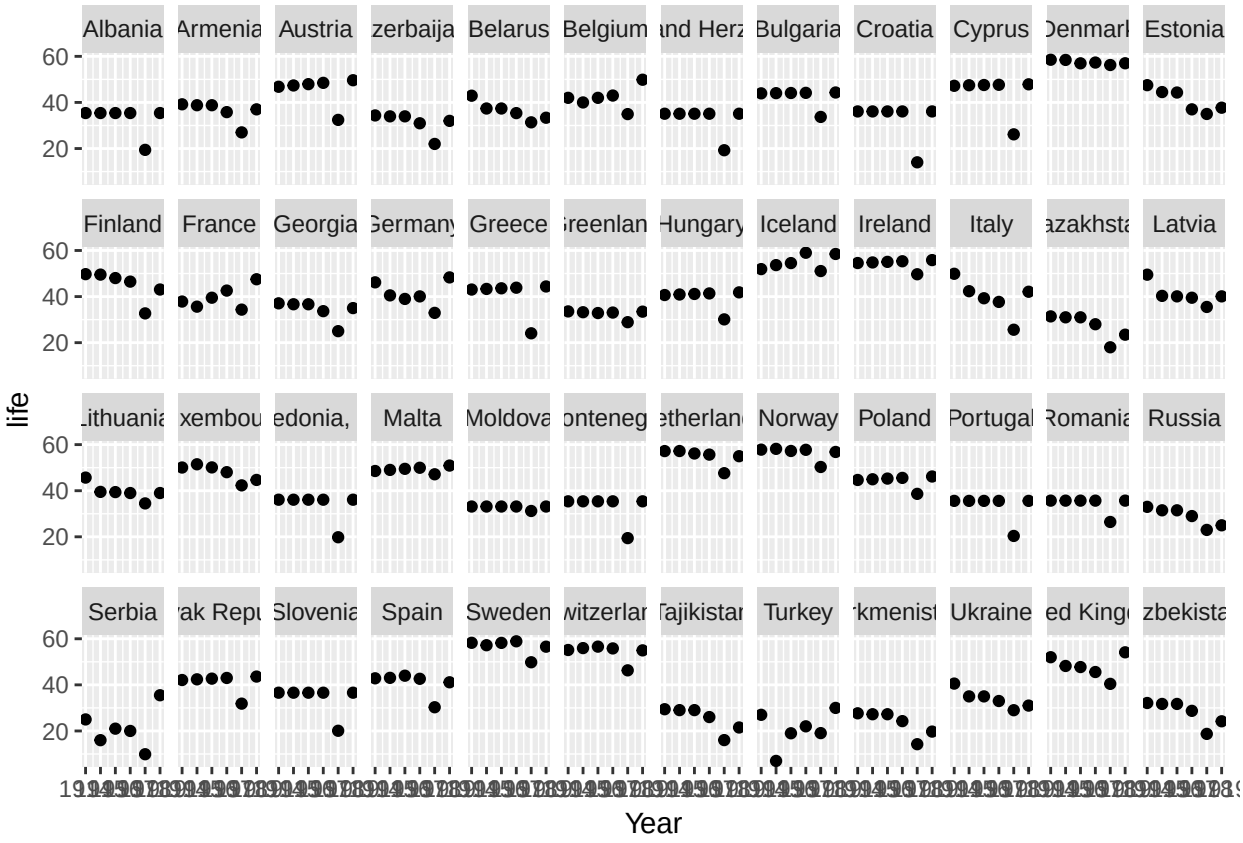
We can look at the plot with 3 clusters and say that it divides the data into 3 different regions -:

a) lowere income region (red)
b) middle income regions (green)
c) higher income regions(blue)

We can separate these regions and can do further analysis.

Impact on different European and Central Asian countries during War years

Impact on Life expectancy for WW1

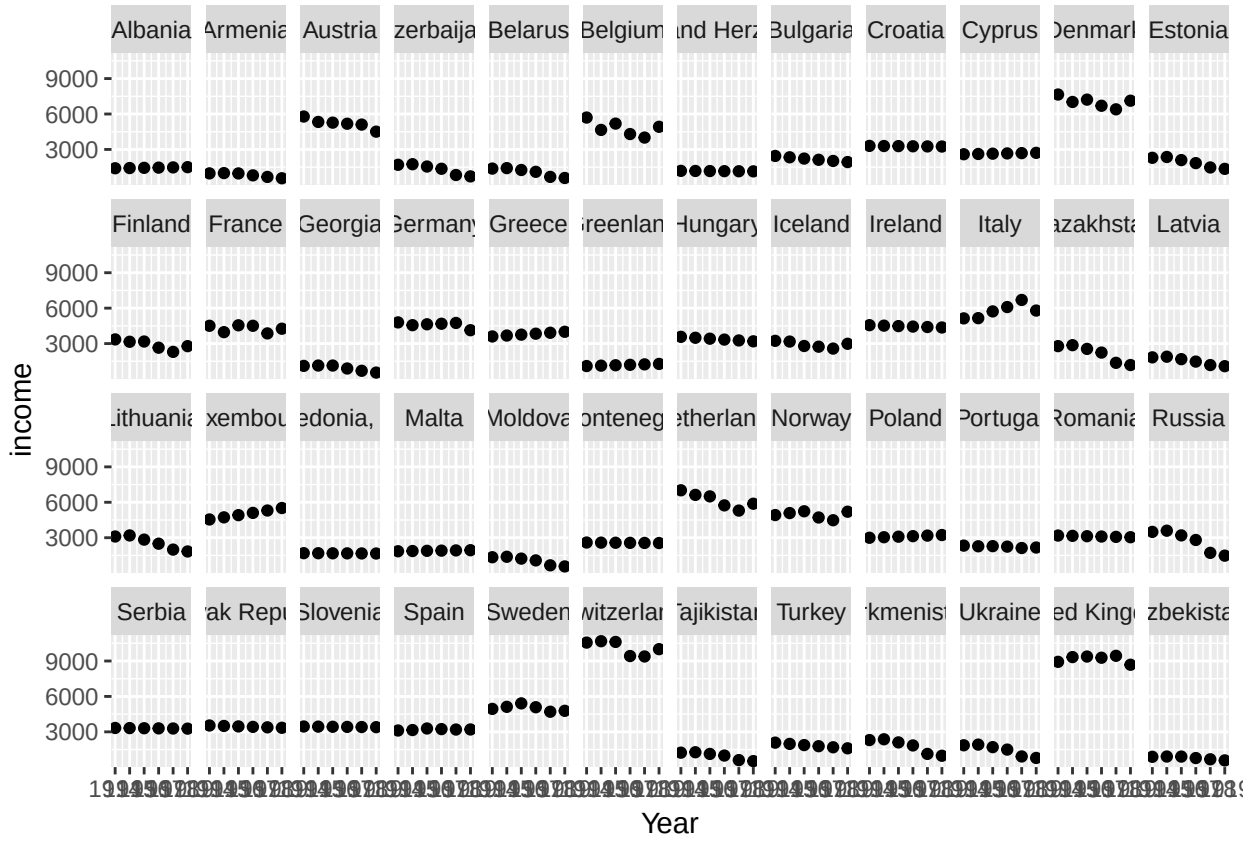


From the above plots of Life expectancy during the WW1 years for different European and Central Asian countries countries, we can see that :

1.Considerable drop of life expectancies for Armenia,Albania,Austria,Azerbaizan,Belarus,Belgium,Bosnia and Herzegovina,Bulgaria,Croatia,Estonia,Finland,France,Georgia,Greece,Hungary,Italy,Kazakhastan,Latvia,Lithuania,Russia,Serbi Kingdom and Uzbekistan.

2. Also we can notice that the hardest hit countries were Belarus,Estonia,France,Germany,Italy,Russia,Serbia,Ukraine,United Kingdom and Uzbekistan.

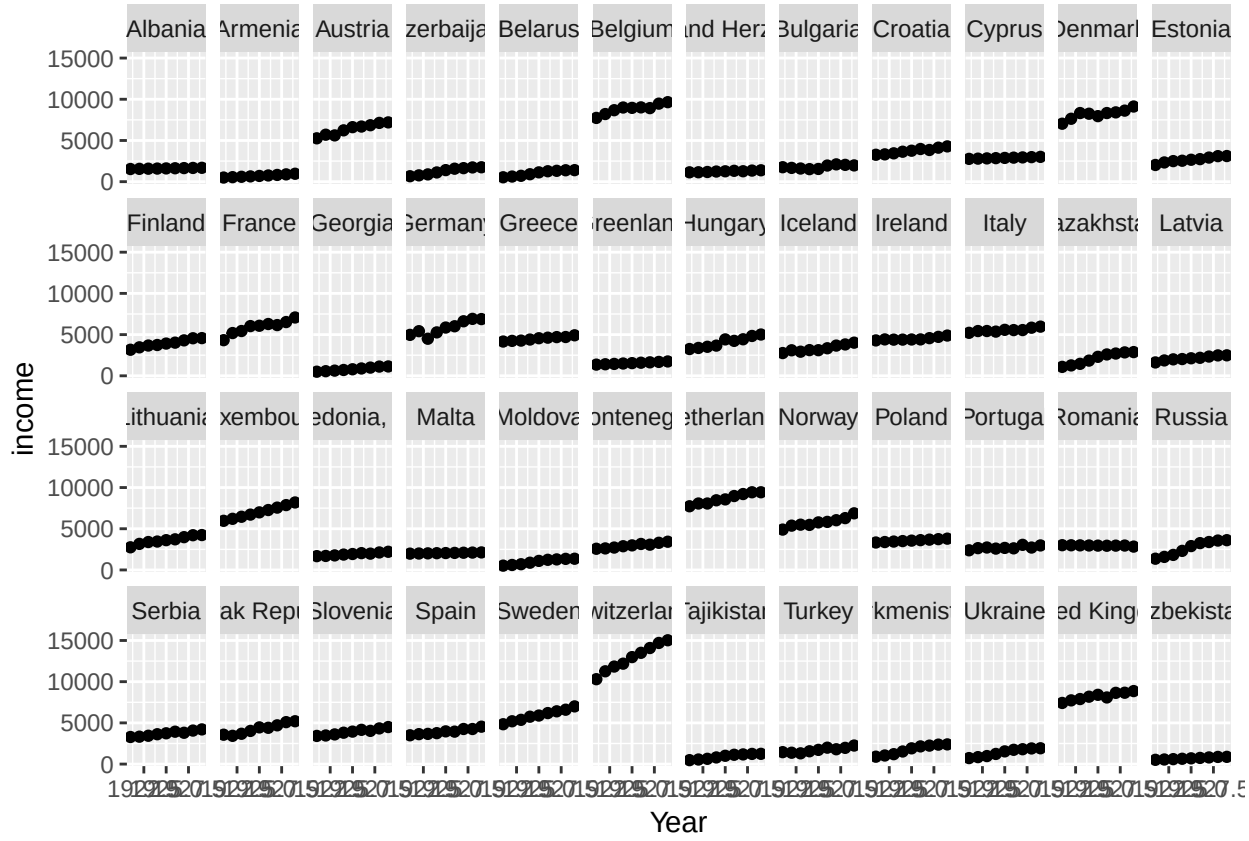
Impact on Per capita Income for WW1



From the above plots ofPer capita income during the WW1 years for different European and Central Asian countries , we can see that :

1.There was significant decline in income for Austria,Belgium,Denmark,Kazakhastan,Latvia,Netherlands,Russia and Ukraine 2. We dont see such a dramatic impact on income as we earlier saw as in life expectancies.

Economic recovery after WW1 for the hardest hit countries

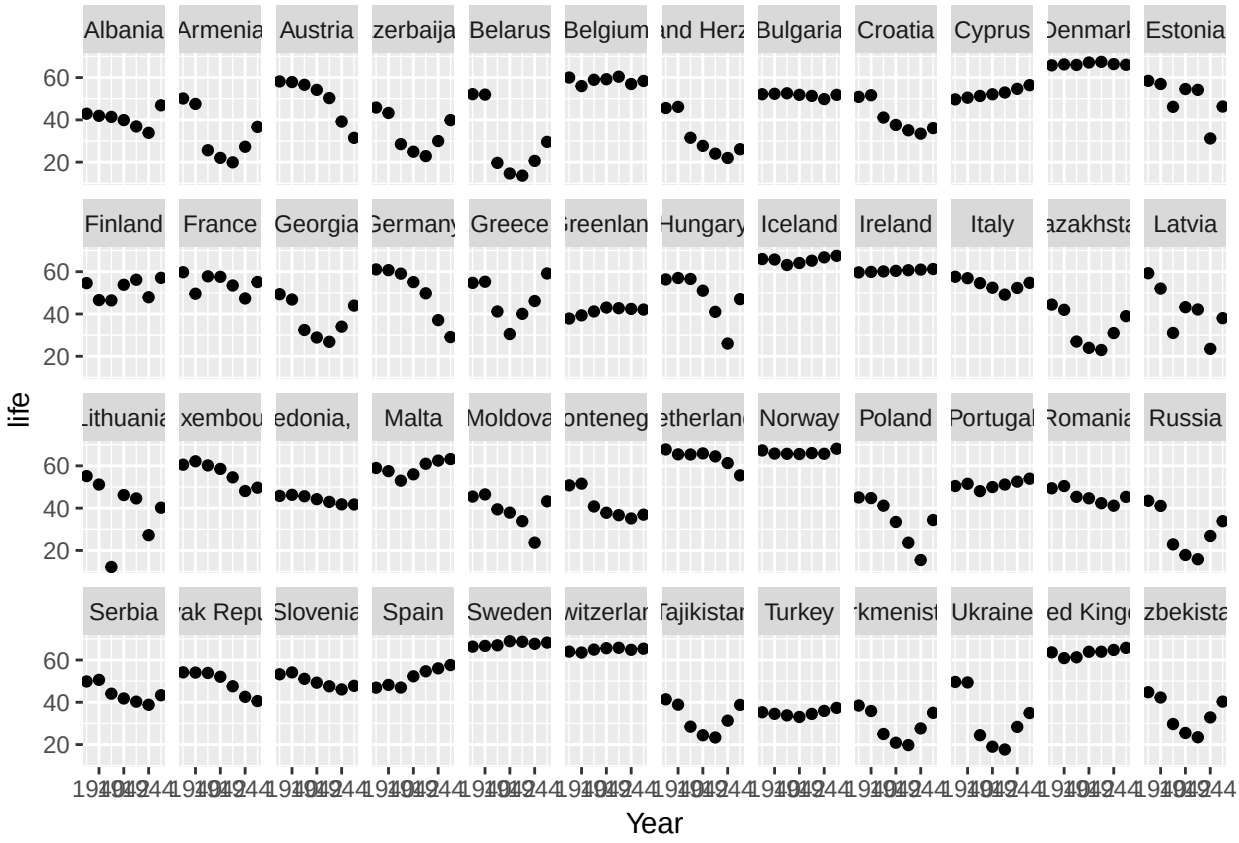


Looking at the Per capita income for the hardest hit countries in WW1 for the next decade (1920-30), we see:
1.Rapid increase in income for Austria,Belgium,Denmark and Russia.

Subsetting data for WW2(1939 - 1945)

```
mydata1_europeww2 <- mydata1[mydata$Year > 1938 , ]
mydata1_europeww2a <- mydata1_europeww2[mydata1_europeww2$Year < 1946 , ]
mydata1_europeww2b <- mydata1_europeww2a[mydata1_europeww2a$region == "Europe & Central Asia" , ]
```

Impact on Life expectancy

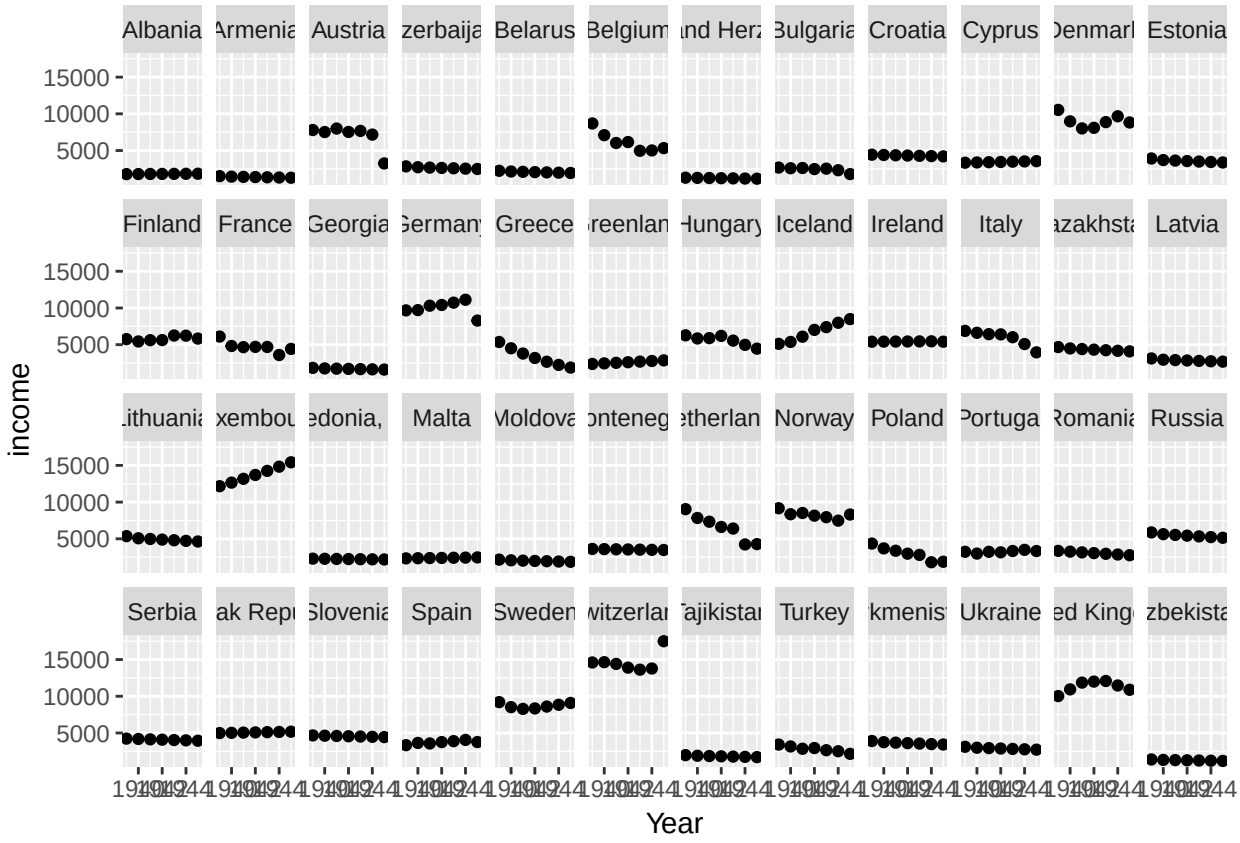


From the above plots of Life expectancy during the WW2 years for different European and Central Asian countries countries, we can see that :

1.Significant decline in Armenia,Austria,Azerbaizan,Bosnia and Herzegovina,Croatia,Estonia,Georgia,France,Germany,Greece,H Republic,Tazikistan,Ukraine and Uzbekistan

2. We see the hardest hit countries were Austria,Azerbaizan,Bosnia and Herzegovina,Georgia,Germany,Greece,Latvia,Poland, and Uzbekista

Impact on Per capita Income



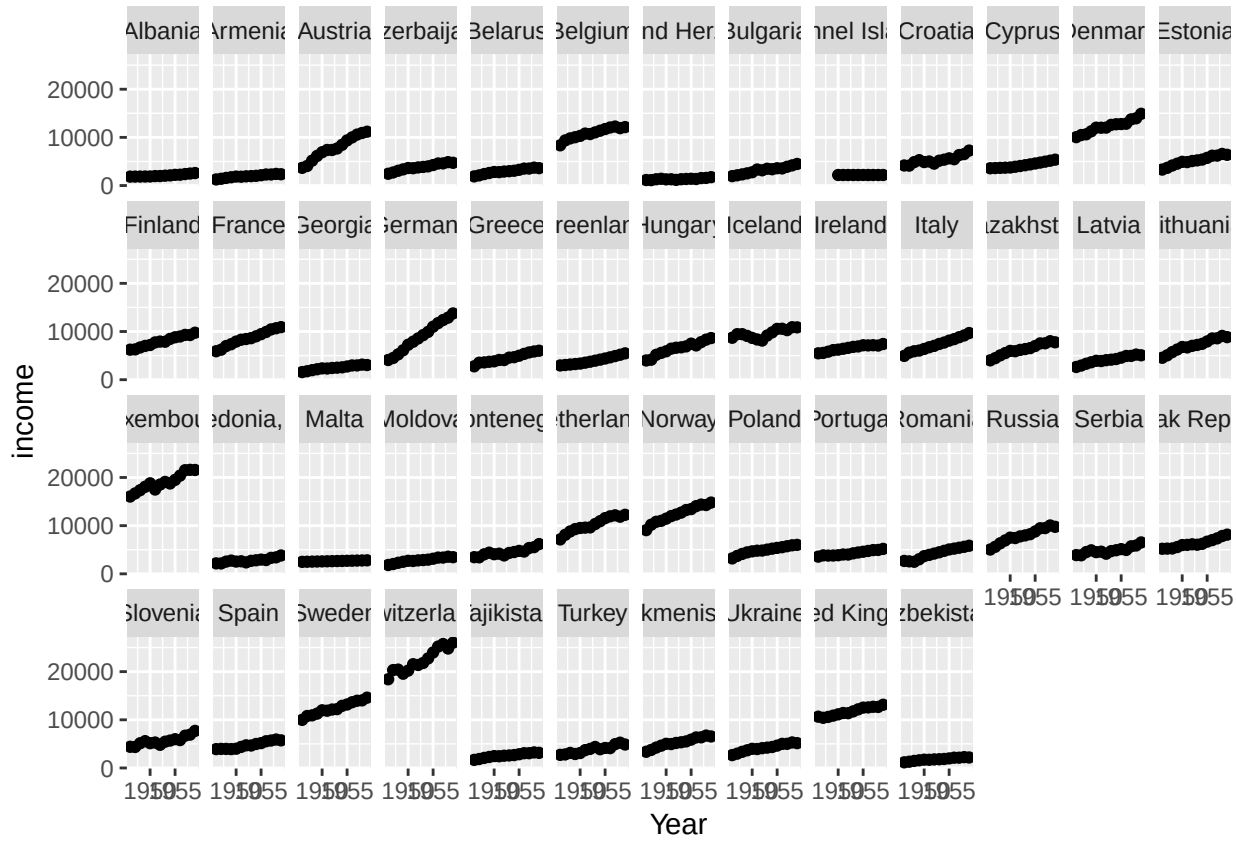
From the above plots ofPer capita income during the WW2 years for different European and Central Asian countries, we can see that :

1.There was significant decline in income for Belgium,Denmark,Greece,Italy,Netherlands and Poland.

2.Comparing these plots (on the considered scales of life expectancy and income) with the life expectanciy plots we again find that the impact on life was greater than on incomne.

Economic recovery after WW2 in Europe

To appreciate the economic recovery after WW2 we select a time period of 15 years after the war .(1945-59)



From the above plots we can see : 1. Rapid increase in Percapita income signalling an economic recovery in all of the European countries. 2.We see a steepest rise in income for Austria,Belgium,Denmark, Germany, Luxemburg,Netherlands,Sweden and Switzerland.

Summary

By looking our the data we can very well confirm the generally believed perception that the World Wars were the most catastrophic events in recent history. The impact on life was much greater than the economy during both the wars but also the resilient charecter of humans should be appreciated which can be seen in the economic recovery and increase in Life expectancy after the wars.