DAT565/DIT407 Assignment 1

Connell Hagen
connellh@student.chalmers.se

Måns Redin
mansre@student.chalmers.se

2024-09-09

1 Problem: Dependency Ratio

1.1 Dependency Ratio

In this part, the given dataset was loaded and cleaned. The dataset was divided into subgroups to separate the different age groups that we needed to calculate the dependency ratio. Equation 1

$$\text{dependency ratio} = 100 \cdot \frac{\text{People aged 0-14} + \text{People aged 65+}}{\text{People aged 15-64}} \tag{1}$$

Figure 1 shows the calculated dependency ratio over the desired period collected from the dataset.

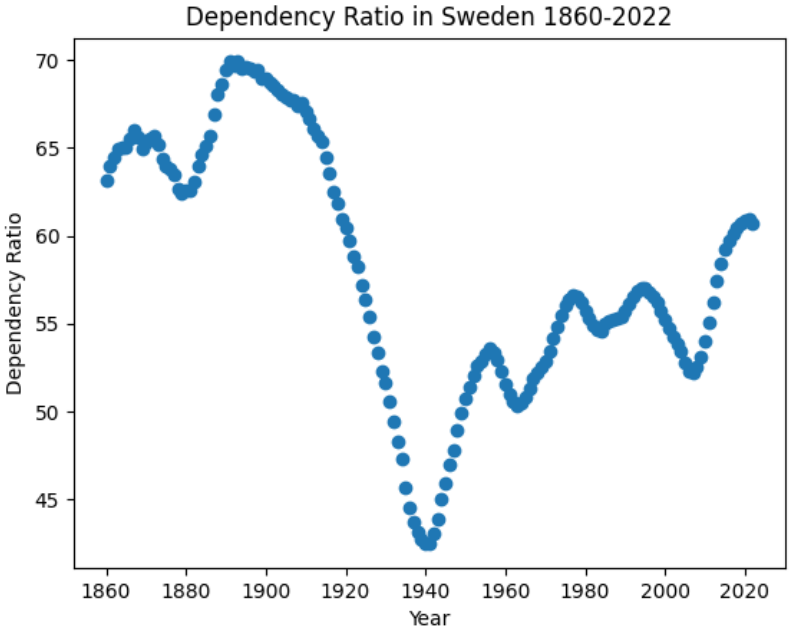

Figure 1: Dependency ratio in Sweden over time.

1.2 Age Population Fractions

In the second part of the problem the fractions of the population of the subgroups, compared to the total population were requested. Since the subgroups were already created, the fractions could easily be calculated with Equation 2.

$$\text{Age group population fraction} = \frac{\text{Age group population}}{\text{Total population}} \tag{2}$$

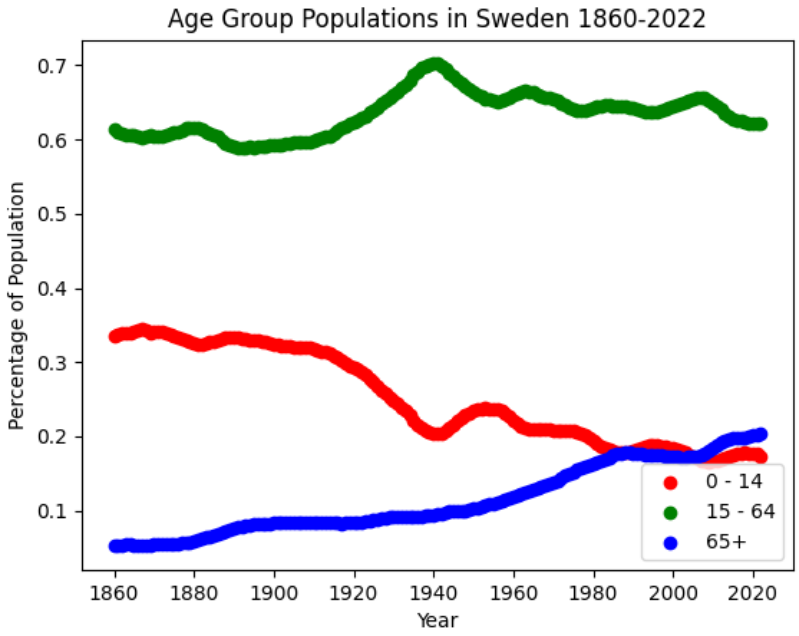

Figure 2: Fraction of population of each age group in Sweden over time.

1.3 Analysis

The result of the data visualization is that the dependency ratio reaches its minimum around 1940, which was likely a result of less children being born during the second world war. After the war, there was a baby boom which may be why there is a quick spike in the dependency ratio. Also, as shown by Figure 2, there was a much greater fraction of the population that was of an old age. This is probably due to better medical technology allowing people to live longer on average. Additionally, the graph shows that the numbers of babies/young children have been decreasing steadily over time, when compared to other age groups. This is because each person, on average, was having many more children before the world wars than they are in the modern day. This may be explained by there being significantly less infant deaths than there used to be with modern medicine, and also due to increased access to contraception within recent decades.

A Sample code

Here is a sample code listing. The options make keywords colored.

```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt

df = pd.read_csv("assignment_1/
    swedish_population_by_year_and_sex_1860-2022.csv")

df.loc[df["age"] == "110+", "age"] = "110"

years = [i for i in range(1860, 2023)]

age_0_to_14 =  df[df["age"].astype(int) <= 14].sum().
    drop(["age", "sex"], axis="index")
age_15_to_64 = df[(df["age"].astype(int) > 14) & (df["
    age"].astype(int) < 65)].sum().drop(["age", "sex"],
     axis="index")
age_65_plus =  df[df["age"].astype(int) >= 65].sum().
    drop(["age", "sex"], axis="index")

dep_ratio = np.array([[i, 0.0] for i in years])
dep_ratio = pd.DataFrame(dep_ratio)
dep_ratio.columns = ["year", "dep_ratio"]

total_pop = np.array([[i, 0.0] for i in years])
total_pop = pd.DataFrame(total_pop)
total_pop.columns = ["year", "population"]

frac_pop = np.array([[i, 0.0, 0.0, 0.0] for i in years
    ])
frac_pop = pd.DataFrame(frac_pop)
frac_pop.columns = ["year", "0_14", "15_64", "65+"]

for i in years:
    dep_ratio.loc[dep_ratio["year"] == i, "dep_ratio"]
        = 100 * (age_0_to_14.iloc[i - 1860] +
       age_65_plus.iloc[i - 1860]) / age_15_to_64.iloc
       [i - 1860]
    total_pop.loc[total_pop["year"] == i, "population"
       ] = age_0_to_14.iloc[i - 1860] + age_65_plus.
       iloc[i - 1860] + age_15_to_64.iloc[i - 1860]

    frac_pop.loc[frac_pop["year"] == i, "0_14"] =
       age_0_to_14.iloc[i - 1860] / total_pop.loc[
       total_pop["year"] == i, "population"]
    frac_pop.loc[frac_pop["year"] == i, "15_64"] =
       age_15_to_64.iloc[i - 1860] / total_pop.loc[
```

```
            total_pop["year"] == i, "population"]
33      frac_pop.loc[frac_pop["year"] == i, "65+"] =
            age_65_plus.iloc[i - 1860] / total_pop.loc[
            total_pop["year"] == i, "population"]
34
35  plt.scatter([i for i in years], [dep_ratio.loc[
        dep_ratio["year"] == i, "dep_ratio"] for i in years
        ])
36  plt.title("Dependency Ratio in Sweden 1860-2022")
37  plt.xlabel("Year")
38  plt.ylabel("Dependency Ratio")
39
40  plt.show()
41
42  plt.scatter([i for i in years], [frac_pop.loc[frac_pop
        ["year"] == i, "0_14"] for i in years], color = "
        red", label = "0 - 14")
43  plt.scatter([i for i in years], [frac_pop.loc[frac_pop
        ["year"] == i, "15_64"] for i in years], color = "
        green", label = "15 - 64")
44  plt.scatter([i for i in years], [frac_pop.loc[frac_pop
        ["year"] == i, "65+"] for i in years], color = "
        blue", label = "65+")
45  plt.legend()
46  plt.title("Age Group Populations in Sweden 1860-2022")
47  plt.xlabel("Year")
48  plt.ylabel("Percentage of Population")
49
50  plt.show()
```