Are the seniors a burden?

Influences of a larger senior population on a country

Course: EPA1315 Data analytics and visualization

Master program: Engineering Policy Analysis

University: TU Delft

Authors: Mees Hoff (4963059), Siemon Keij (4285654), Philip Seijger (4272803), Jason R Wang (4788281)

# Summary

With the increase of senior citizens compared to the working force of the world, it is important to figure out what the impact is of this increase on countries. Will the increase stimulate demand or will the burden of the increase in healthcare make countries poorer? These question lay at the root of this research. To simplify the goal of this paper, the following research question has been asked:

*"Will the increase in the age dependency ratio increase the stress on a society?"*

Stress can be viewed in many different ways, but in this paper stress will focus on the GDP, healthcare costs, and household consumption in a country. These indicators might be influenced by, or influence the number of senior people in a country. Furthermore, the future won't be forgotten, the age dependency ratio young is also taken into account.

With the help of a fork model a hypothesis is tested that age dependency old does influence the GDP per capita, the health expenditure, and the age dependency young in a country. The fork used 6 betas and 3 sigma's which had a prior of a normal distribution between 0 and 0.01 for the different betas, and a uniform distribution between 0.01 and 0.99 for the variances, which when squared gave the sigma's.

From this model it became clear a correlation between the different variables is possible, but not a linear connection. Because the model was only build to find a linear correlation, this to make the model not too complex, a definitive conclusion can't be reached. This research should be continued by developing a model which will check a possible second degree correlation between the aforementioned variables.

The code can be found on github via the following link: https://github.com/jasonrwang/oecd\_insights

Table of Contents

[Summary 2](#__RefHeading___Toc1513_2060011200)

[1. Introduction 4](#__RefHeading___Toc761_2060011200)

[2. Business understanding section 5](#__RefHeading___Toc763_2060011200)

[2.1. Issues 5](#__RefHeading___Toc1381_2060011200)

[2.2. Limitations 5](#__RefHeading___Toc1383_2060011200)

[2.3. Trends 5](#__RefHeading___Toc1385_2060011200)

[2.4. Policy Actualities 8](#__RefHeading___Toc1387_2060011200)

[2.5. Relevance for Organisation for Economic Co-operation and Development (OECD) 9](#__RefHeading___Toc1389_2060011200)

[3. Data collection 10](#__RefHeading___Toc765_2060011200)

[4. Data description 11](#__RefHeading___Toc767_2060011200)

[4.1. Parameters 11](#__RefHeading___Toc1068_2060011200)

[4.2. Structure 13](#__RefHeading___Toc1092_2060011200)

[5. Data quality 14](#__RefHeading___Toc769_2060011200)

[6. Data exploration 15](#__RefHeading___Toc771_2060011200)

[6.1. Correlations 16](#__RefHeading___Toc1094_2060011200)

[6.2. Causal diagram 17](#__RefHeading___Toc1108_2060011200)

[6.3. Relevance 17](#__RefHeading___Toc1391_2060011200)

[7. Data preparation 19](#__RefHeading___Toc773_2060011200)

[7.1. Cleaning procedure 20](#__RefHeading___Toc1393_2060011200)

[8. Test Design 21](#__RefHeading___Toc775_2060011200)

[9. Models 22](#__RefHeading___Toc777_2060011200)

[9.1. Models 22](#__RefHeading___Toc1395_2060011200)

[9.2. Chosen model 23](#__RefHeading___Toc1397_2060011200)

[10. Parameter settings 24](#__RefHeading___Toc779_2060011200)

[11. Model description 25](#__RefHeading___Toc781_2060011200)

[12. Assessment 27](#__RefHeading___Toc785_2060011200)

[References 28](#__RefHeading___Toc787_2060011200)

[Appendix 29](#__RefHeading___Toc1517_2060011200)

[A: Model diagnostics plots 29](#__RefHeading___Toc1519_2060011200)

[B: Output plots 33](#__RefHeading___Toc1531_2060011200)

# Introduction

In many countries in the western world, the percentage of the population that is the working force, is going down, and is expected to fall further. The reason for the current downward slope is that generation 'baby-boom' is now going into retirement. This results in more seniors per worker. Furthermore, the birthrate is also slowing down. This means less babies, but the same amount of adults. This can result in the same burden befalling the children of the 'baby-boomers' as they themselves are now experiencing.

The growing percentage of seniors has an influence on a country, this is no question. A question might be what the influence of the growing senior population is on different parts of the country. To further research this scope, data from the world bank was used. The world bank has free available data from most countries on a wide range of different subjects. As extension of age dependency data was gathered from different fields, which might influence of might be influenced by age dependency. This will be discussed later in the report in chapter X.

The research question which guides this paper is

*"Will the increase in the age dependency ratio increase the stress on a society?"*

The stress in a society can be measured with different parameters. An example might be the decrease in GDP which might be a consequence because of the increase of the percentage of seniors. This will be explored further in this paper.

The report follows the CRISP-DM process. This starts by doing business understanding in chapter 2. After collecting the data in chapter 3, exploring the data in chapter 4 and 5, and do an exploratory analysis in chapter 6. This leads to data preparation in chapter 7. After the data has been gathered, there will be some chapters building up to the final model created, namely test design in chapter 8, model options in chapter 9, parameter settings in chapter 10, and the final model description in chapter 11 After which an assessment will be given in chapter 12.

# Business understanding section

The age dependency ratio is the ratio between the number of people in the labour force, and the number of people that are not. For further analysis, this ratio can be divided in old-age dependency ratio and young-age dependency ratio. The old-age ratio refers to the number of people that have reached the age of retirement (65+). The young-age dependency ratio refers to the number of people, that are too younger than 15. (United Nations, 2005).

## 2.1. Issues

Age dependency ratios are used to indicate potential effects of change in the population age structures. These effects are mainly related to social and economic development, that point out broad trends in social support needs. In general, a higher dependency rate tends to increase financial stress between working people and dependents. It can cause serious problems if large proportions of governmental spending are related to social security, health and education, since these are most used by the dependents (both old and young). The fewer people in the working age, the fewer can support the costs related to the dependents.

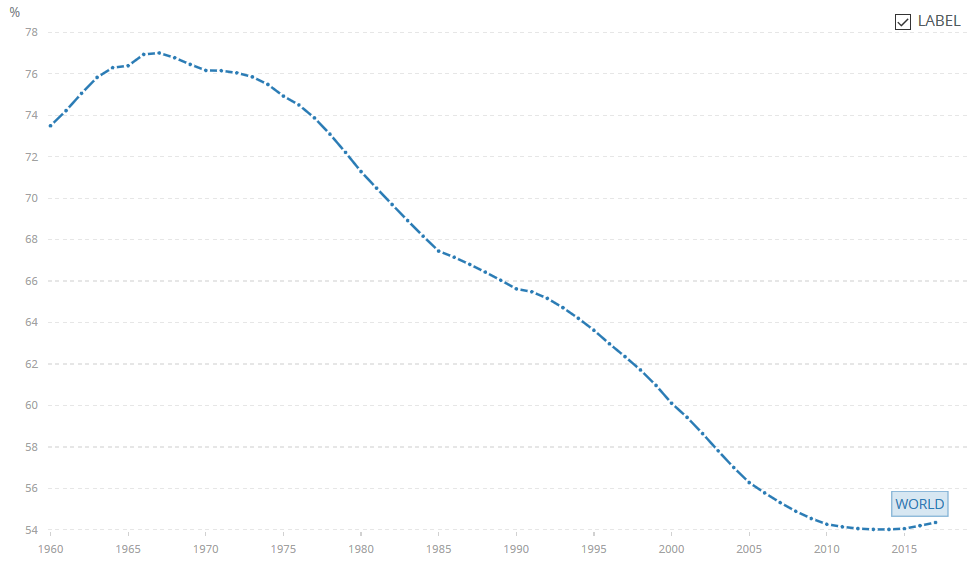
## 2.2. Limitations

The definition of age dependency ratio ignores some significant facts. Firstly, people above the age of 65 are not necessarily dependent, since an increasing part of them is working nowadays. Also, people in the working age are not all working. People stay out of work for a number of reasons. First and foremost, some people are simply unemployed. Also, people that follow education for a longer amount of time have a disturbing impact on the reliability of this definition.

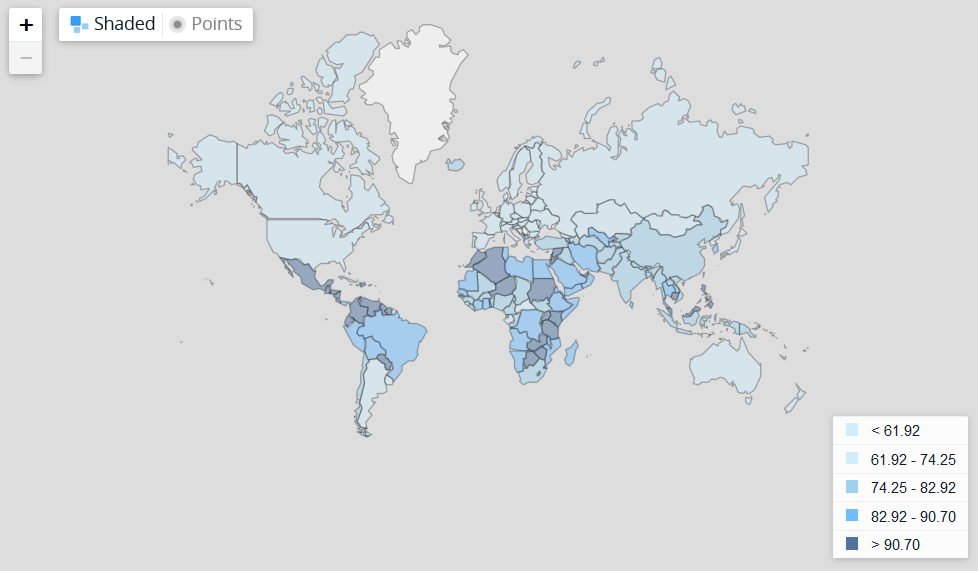
## 2.3. Trends

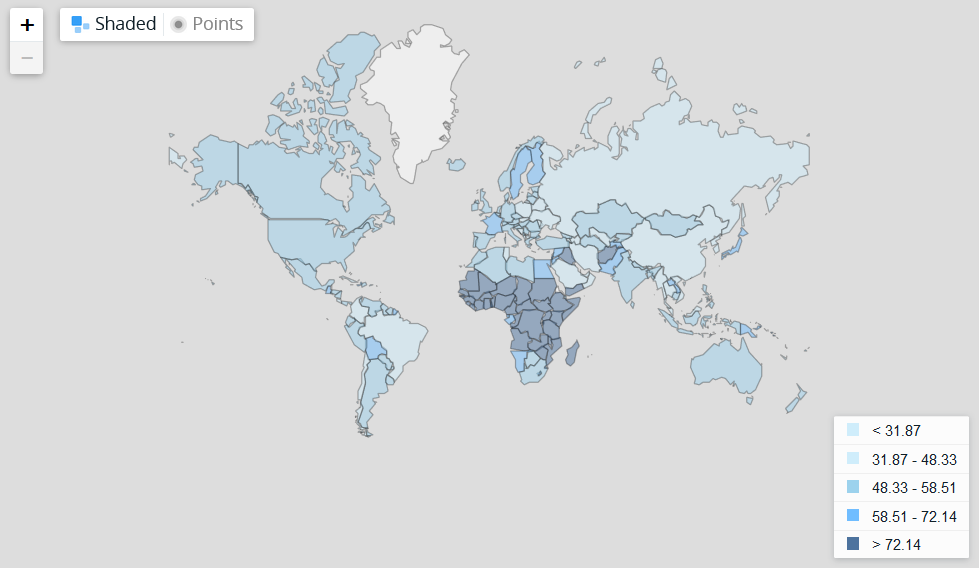
The life expectancy of people all over the world has significantly increased over the past decades. This increase can be explained by a number of reasons, but most importantly it is due three reasons. These reasons are better food production and distribution, improvement in public health and better medical technology [1]. According to a study in 188 countries, in the last two decades the life expectancy is especially increased due to lower death rates caused by infectious and cardiovascular diseases [2].

Unlike the life expectancy, age of retirement did not change much in most countries. This has as direct consequence that the age dependency ratio has decreased over the past decades. In figure 1 the age dependency of the world is depicted over the period from 1960 until 2017 (World Data Bank, 2017). It can be seen that global age dependency decreased from approximately 75% towards 54%.

*Figure 1: The change of age dependency old in the world from 1960 until 2017*

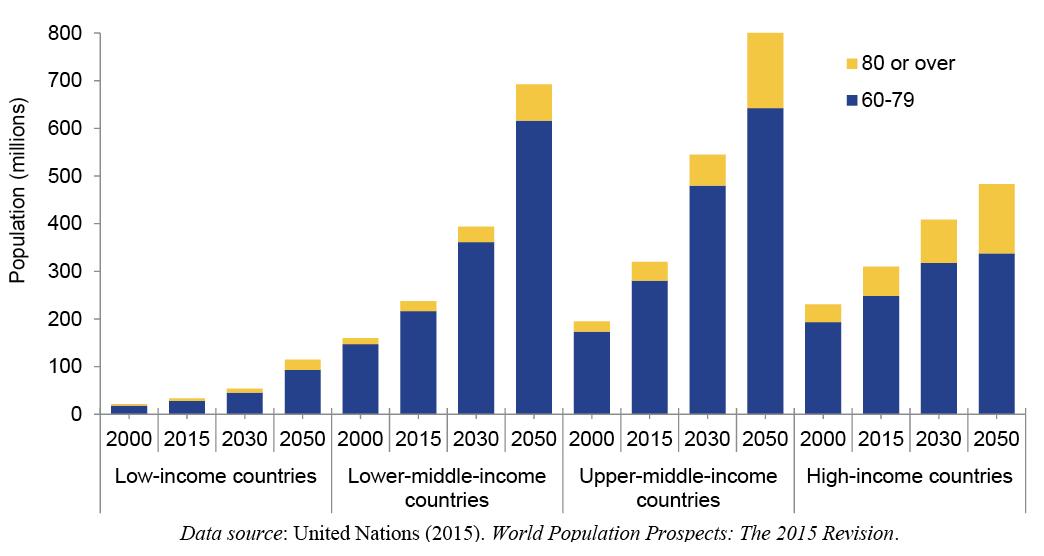
To further analyse the ongoing trend of age dependency over the past years, in figure 2 a map of the age dependency per country in 1960 is given. In figure 3 these data can be compared to the numbers of age dependency in 2017. The difference becomes most evident by comparing the legends of both maps. (World Data Bank, 2017)

*Figure 2: Age dependency old in 1960 across the world*

*Figure 3: Age dependency old in 2017 across the world*

\*Note: The colours of both pictures relate to different ratios, since the global distribution of age dependency was completely different in 1960, compared to 2017.

Lastly, figure 4 provides a prediction for the number of old-age dependent people over the world for the coming decades. A distinction is made for more developed regions, less developed regions and least developed countries (United Nations, 2015).



*Figure 4: Population aged 60-79 years and aged 80 years or over by development group, 2000, 2015, 2030 and 2050*

## 2.4. Policy Actualities

Nowadays approximately 20% of the people in Europe is aged over 65, which is a record high.

This increase in life expectancy and its impact on the age dependency ratio has caused discussions about the age of retirement. In the countries that are part of the Organisation for Economic Co-operation and Development (OECD), men retire at an average age of 65, for women this is 63.5 [3].

In 2013 the Dutch government, for instance, decided to start increasing the age of retirement periodically [4], but not only in Europe this discussion is ongoing. The prime minister of Canada decided in 2016 to keep the retirement age at 65, but the plan to increase retirement age is still on the agenda [5].

## 2.5. Relevance for Organisation for Economic Co-operation and Development (OECD)

This subject is of particular interest to the strategic counsellor of the OECD for a number of reasons. Firstly, the increase of life expectancy is especially relevant in the member states of this organization, since these are mostly more developed countries (figure 4). Also, the economic consequences are highly relevant towards the OECD, since its main interest is economic development. Lastly, the OECD has stated its interest in the 7 societal challenges set by the UN. This problem is part of the Health and Demographic Change & Wellbeing challenge. The fact that the topic of ageing is explicitly mentioned in the global issue report of the United Nations, strengthen this statement [6].

# Data collection

Open data can be collected from the world bank site. The World Bank collects data about all the countries in the world in many different fields. The World Bank has three different variances of age dependency. Old, young, and ratio. Age dependency old looks at the percentage of seniors (65+) compared to the working force. Age dependency young does this for the youth (15-). Age dependency ratio looks at the ratio between the working force and people that are younger than 15 or older than 65, so a sum of age dependency old and young. The data that has been collected, has to be cleaned and put into a template before it can be used, this will be further discussed in chapter 7, data preparation.

More data was collected than eventually used in the created model. Besides the earlier mentioned age dependency parameters, data was also imported about the GDP (Gross Domestic Product, how much a country produces in a year), healthcare, household consumption, life expectancy, total population, and GINI (coefficient that shows the inequality in a distribution). All the different data can be linked to age dependency. All the acquired data will be further discussed in the next chapter.

# Data description

The data acquired is divided into different parameters. With help of the explanation provided by the world bank, all the different possible parameters will be discussed here. After the parameters a further description of how the data is structured will be given.

## 4.1. Parameters

### Age dependency old

Age dependency ratio, old, is the ratio of older dependents – people older than 64 – to the working-age population – those ages 15-64. Data are shown as the proportion of dependents per 100 working-age population.

### Age dependency young

Age dependency ratio, young, is the ratio of younger dependents – people younger than 15 – to the working-age population – those ages 15-64. Data are shown as the proportion of dependents per 100 working-age population.

### Age dependency ratio

Age dependency ratio is the ratio of dependents – people younger than 15 or older than 64 – to the working-age population – those ages 15-64. Data are shown as the proportion of dependents per 100 working-age population.

### GDP per capita

GDP per capita is gross domestic product divided by midyear population. GDP is the sum of gross value added by all resident producers in the economy plus any product taxes and minus any subsidies not included in the value of the products. It is calculated without making deductions for depreciation of fabricated assets or for depletion and degradation of natural resources. Data are in current U.S. dollars.

### GDP

GDP at purchaser's prices is the sum of gross value added by all resident producers in the economy plus any product taxes and minus any subsidies not included in the value of the products. It is calculated without making deductions for depreciation of fabricated assets or for depletion and degradation of natural resources. Data are in current U.S. dollars. Dollar figures for GDP are converted from domestic currencies using single year official exchange rates. For a few countries where the official exchange rate does not reflect the rate effectively applied to actual foreign exchange transactions, an alternative conversion factor is used.

### Health expenditure per capita

Current expenditures on health per capita in current US dollars. Estimates of current health expenditures include healthcare goods and services consumed during each year.

### Health expenditure

Level of current health expenditure expressed as a percentage of GDP. Estimates of current health expenditures include healthcare goods and services consumed during each year. This indicator does not include capital health expenditures such as buildings, machinery, IT and stocks of vaccines for emergency or outbreaks.

### Household consumption

Household final consumption expenditure (formerly private consumption) is the market value of all goods and services, including durable products (such as cars, washing machines, and home computers), purchased by households. It excludes purchases of dwellings but includes imputed rent for owner-occupied dwellings. It also includes payments and fees to governments to obtain permits and licenses. Here, household consumption expenditure includes the expenditures of non-profit institutions serving households, even when reported separately by the country. Data are in current U.S. dollars.

### GINI

GINI measures the extent to which the distribution of income (or, in some cases, consumption expenditure) among individuals or households within an economy deviates from a perfectly equal distribution. A Lorenz curve plots the cumulative percentages of total income received against the cumulative number of recipients, starting with the poorest individual or household. GINI measures the area between the Lorenz curve and a hypothetical line of absolute equality, expressed as a percentage of the maximum area under the line. Thus a GINI of 0 represents perfect equality, while an index of 100 implies perfect inequality.

### Life expectancy at birth

Life expectancy at birth indicates the number of years a newborn infant would live if prevailing patterns of mortality at the time of its birth were to stay the same throughout its life.

### Total population

Total population is based on the de facto definition of population, which counts all residents regardless of legal status or citizenship. The values shown are midyear estimates.

## 4.2. Structure

The world bank collects data as of the 1960's. The amount of data becomes more complete as the years pass. In the 20th century there is a lot of missing data. After the turn of the century it becomes less and less. The Western world has, in general, quite complete data a long way back. But this also differs per country and per parameter. The data is structured per year per country, however, there are also some aggregated areas in the data. An example is the Arab World, which is composed of the members of the League of Arab States. The years go from 1960 to 2017, new data is added at least yearly.

# Data quality

The quality of the data depends on the number of missing data points. The amount of data depends on the year and the countries. More data is available in more recent years. In general Western countries are available as of the beginning of the data, poorer countries are joining the data much later. For some countries still not all data is available. Some parameters that mentioned earlier have not much data, for others some data goes back until the beginning.

The GINI data is very scarce, last year still had a lot of missing data. Other data, like the GDP or age dependency provides data for almost all countries in the last five years. The quality of the data will be of influence when the parameters for the model are chosen.

# Data exploration

To begin data exploration a correlation matrix of normalised variables was created. This makes it clear of any correlation might be found between different variables. In figure 5 such a matrix is shown for some, but not all, of the earlier mentioned variables. The correlation between two variables is high if the background is red, and low is the background is green.



*Figure 5: Correlation matrix of the different variables*

From the figure above it can be stated that there is expected to be a correlation between age dependency old and the variables health expenditure, GDP per capita, age dependency young and age dependency ratio. The hypothesis about the other variables, GDP and household consumption, are that they don't have a correlation with age dependency old. There is expected to be a small correlation between age dependency old and health expenditure per capita. The numerical values of the correlation are given in table 1.

## 6.1. Correlations

After discussing the possible correlations between age dependency old and the other variables, now a short explanation will be provided if a correlation is expected.

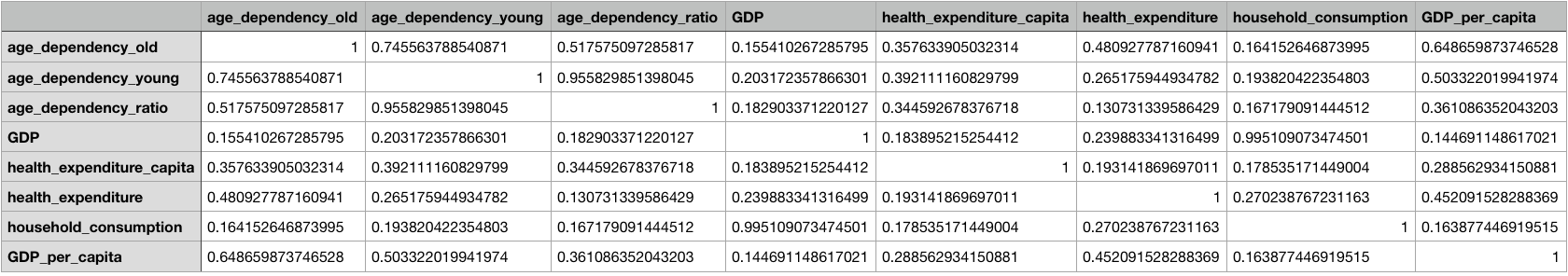
### Health expenditure

An explanation for the possible correlation with age dependency old is that cost of healthcare is mostly increasing with the amount of elderly people. Elderly people tend to need more care and therefore in the case of a high age dependency ratio (old) the cost of healthcare is generally higher. For the overall age dependency ratio this effect is not strong enough.

### Health expenditure per capita

The age dependency old shows a small correlation with health expenditure per capita. This correlation is similar for the young and old age dependency. An important note for this conclusion is the fact that in health expenditure per capita the effect of age dependency not is included. By calculating this expenditure, it is assumed that all people pay for the costs related to health. In practice these costs are covered by the working part. In order to really test the stress that is caused by health expenses, this should be further analysed.

*Table 1: Correlation values of the different variables*



### Household consumption

A correlation with age dependency old is unlikely, based on the correlation matrix. The observed correlation is very weak; therefore, it is uncertain whether the two are related. It is a possibility that age dependency old has an impact on the household consumption, but this impact is not strong enough for a stronger correlation, the other possibility is that this correlation is coincidental, or caused by other factors.

### GDP

The age dependency old is weakly correlated with the GDP of a country. However, is not possible to exclude the possibility of a relation between both variables solely on these data. Reason for this is that countries might have a higher age dependency for other reasons, for instance a high level of welfare. This might cause an increase in life expectancy of the inhibitors of this specific country. Now both age dependency and GDP (assumption based on high level of welfare) are high in this situation, but a direct link is missing.

### GDP per capita

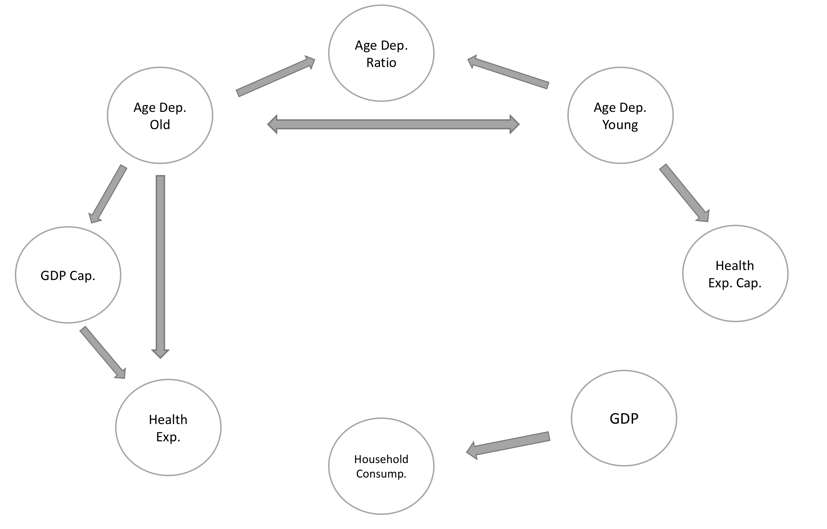
A correlation between age dependency old and GDP per capita clearly exist. However, for numerous reasons this does not have to be an indication for a causal relation. Considering the fact that age dependency old is the highest, a greater number of elderly people is expected to have a correlation with a higher GDP per capita. In order to really analyse the impact of age dependency a longer time frame should be analysed.

### Age dependency young

The correlation between age dependency old and age dependency young is quite strong. This is as most people expect, however, the correlation matrix shows only one year. It might be possible by external influences that a country has a lot of younger people but not a lot of older people. Examples might be war or disasters. A correlation over a longer period of time can be expected. If there are more young people now, there are expected to be more seniors later.

## 6.2. Causal diagram

With the help of the correlation matrix in figure 5, a number of causal models can be supported. In figure 6 these causal diagrams are graphic shown. Only the expected strong correlations are taken into account, the red background. The direction of the arrows is assumed after the exploration of the data. The figure gives a good overview of the possible relationships between the variables.



*Figure 6: Causal diagram of the different variables*

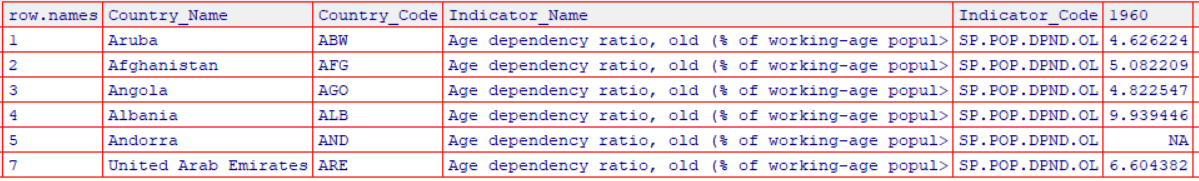
## 6.3. Relevance

From the causal diagram it becomes clear that not all data has a possible correlation with age dependency. Thus some data can be dropped for the research. Household consumption and GDP are possibly not even correlated to age dependency old, thus those are the first to go. Furthermore, health expenditure per capita isn't directly correlated to age dependency old, but only via age dependency young. Thus that one will also be dropped. Finally, age dependency ration will be dropped in favour of age dependency young. Age dependency ratio is age dependency young and old combined, if it was included it might give a wrong picture of the real correlations.

# Data preparation

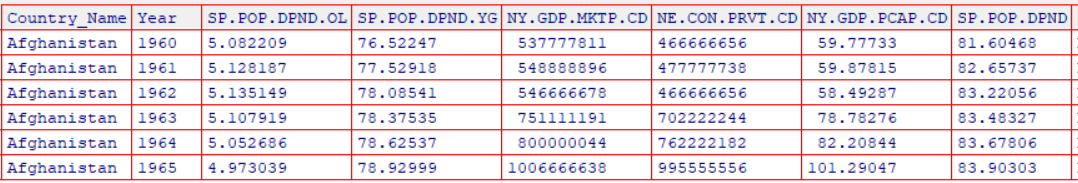
To prepare the data for further analysis, it should be cleaned and standardised. This has been done by renaming and reshaping the data to the same format. This is shown in table 2. With this format the data is structured the same in every data frame. This makes it easier to do certain tasks on it at a later stage.

*Table 2: Head of the initial data frame of the variable age dependency old*



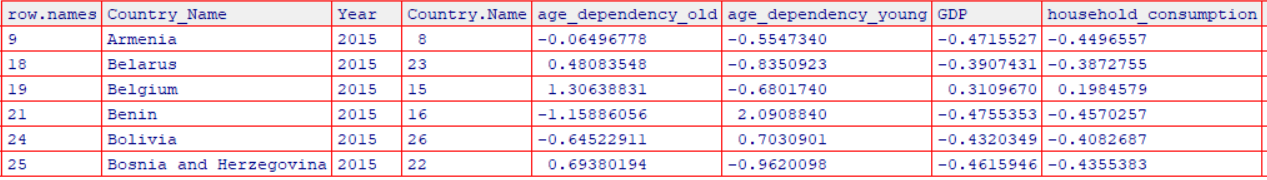
After all the data is structured the same it can be easily put together in one big data frame with the following structure as seen in table 3.

*Table 3: Head of the merged data frame of all the variables*



To finalize our data preparation, the data is normalized which is necessary at a later stage. This is visible in table 4.

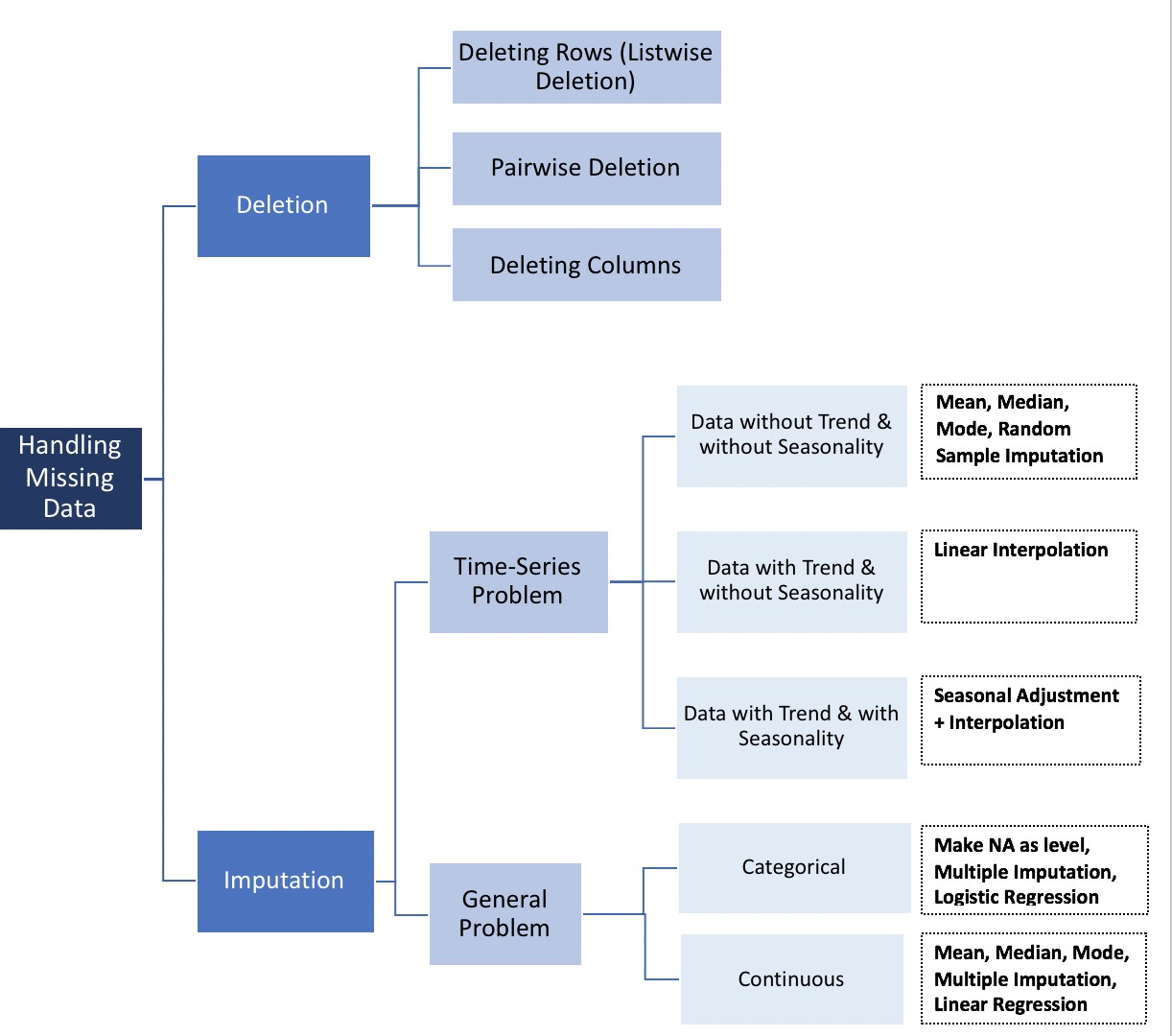
*Table 4: Overview of the final data frame with renamed headers and one year*



As is visible, the data shown still contains all variables, however, it does contain only 50 countries. That was set as the maximum number of countries to be researched. This number arbitrary. The model can easily accommodate all countries in the world, or only ten, depending on the kind of research. Another dataset was created with solely the data which will be used in the model.

## 7.1. Cleaning procedure

The general cleaning procedure which was used was based on the workflow as is visible in figure 7. This workflow provides a clear guideline what to do when missing data is encountered.

*Figure 7: Workflow how to deal with missing data*

# Test Design

To test if the model is correct, it is necessary to have a separate dataset for training and the results. The model will be tested with data from the year 2005, and the data from 2015 will be used to construct a conclusion. The model will be evaluated by looking at the different variables and the proposed correlations between the variables. As was visible in the correlation matrix, some correlation looked more like an exponential connection, and not a linear connection, which is the connection which will be looked for by the model. This might give a false positive or negative by the model on the correlation between variables. To combat this, special attention should be given to those combinations of variables.

# Models

There are three kinds of models that can be used to explore the data. These three options are a chain, a collider, and a fork. All three options will be discussed in the next section. These options will be viewed in a very simplified version. In the real world no variable is either in a chain or fork, or collider. These models are here to explain the basics, not for any other use.

## 9.1. Models

### Chain

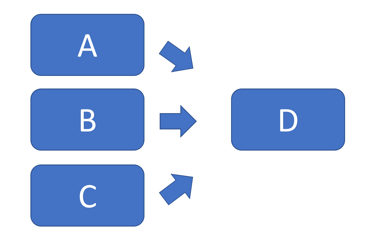
A chain is a model where three options influence each other. The first influences the second, and the second the third. A chain can be longer than three variable. A graphic display of a chain can be seen in figure 8. An example of a chain can be found in pool. The player hits a ball with the cue, this ball than hits another ball, which hits another ball. The first variable was the cue, but it ended with the third or even fourth ball to hit, going into the sack.



*Figure 8: Graphical view of a simplified chain model*

### Collider

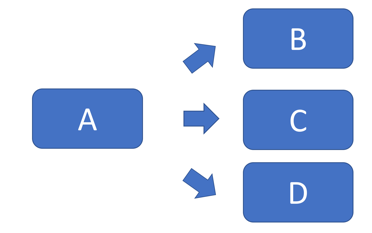
Another option for a model is a collider. In a collider three different variables together influence a fourth variable. A collider is different from a chain that more than one variable is necessary to influence another variable. A graphic display of a collider can be found in figure 9. An example of a collider is for instance a safe where more than one key is necessary to open it. If one of the keys doesn't turn, the door of the safe won't open, the variable won't change.



*Figure 9: Graphical view of a simplified collider model*

### Fork

The last option that will be discussed is a fork. A fork can be seen as a reversed collider. In stead of needing more than one variable for a single variable, a single variable will influence more than one variable. A graphical example can be found in figure 10. A textual example of a fork is how sea-level rising can create problems in the Netherlands, Britain, and other countries in multiple areas.



*Figure 10: Graphical view of a simplified fork model*

## 9.2. Chosen model

It has been decided to look further into a fork. As was visible in the causal diagram, age dependency old has possibly an influence on three other variables, namely age dependency young, GDP per capita, and health expenditure. Thus a fork seems the most logic option.

# Parameter settings

When a fork is used as a model some parameters need to be set. There are three types of parameters that have to be set; the prior for the beta, the prior for the inverse variance and the likelihood.

### Prior for beta

The prior for the beta will be taken from a normal distribution between 0 and 0.01. There are six betas used in the model, so there are six normal distributions necessary.

### Prior for the inverse variance

The inverse variance is used for set the sigma and the variable in the model. The variable is set by taking a number from a uniform distribution between 0.01 and 0.99. Thus a uniform distribution between 0 and 1, where 0 and 1 are not an option to be taken. The six variances are squared to create a sigma, they are also inverted to set the upper limit for the variables in the likelihood.

### Likelihood

The likelihood of the variables is calculated in two steps. The first step is to create a bottom number for the normal distribution, which is used in the second step. This number is calculated by taking a beta and add it to another beta times the main variable, which is in this case age dependency old. This low number will be used, together with the inverse variance created earlier, to form a normal distribution for all three variables, health expenditure, GDP per capita, and age dependency young. From these normal distribution numbers will be taken to be used in the model.

### Other parameters

Furthermore, there are also other parameters which can be changed in the model. For starters, the number of times the model run can be changed, this is set to four. The burn in period can also be changed if necessary, it is now set at 10000 samples. The last parameter to further change, is the number of samples that the model needs to generate (after the burn in period), that number has been set to 20000.

# Model description

A number of tests were used to evaluate model convergence. Although there are no established techniques for evaluating model convergence (Cunningham, 2018), the following methods were used:

* Trace plots: to assess the coverage and consistency through the sampling space for all parameters.
* Density plots: to assess the distribution of all parameter samples.
* Gelman plots: to assess the convergence around 1 for all parameters and determine that there is more variance within the chain than between chains.
* Autocorrelation plots: to assess whether each position in the parameter chain is independent of the previous and if the chains behave similarly over time (clumping).
* Effective sample sizes: to determine whether the sample sizes were large enough.

Using the methods above the model appears to have no major problems. Trace plots of all betas 1-5 show a similar range and consistency, beta 6 has a slightly lower upper bound but is equally consistent (figure A.1). In addition, all the beta density plots show a smooth distribution as expected (figure A.3). in contrast, sigma 1 is inconsistent with sigma’s 2 and 3, covering a larger range but showing a higher lower bound (figure A.2). The sigma density plots (figure A.4) show smooth distributions that tend to lean slightly to the left (sigma 2 and 3) or to the right (sigma 1). Furthermore, sigma 1 shows a much wider distribution as expected. Overall, it looks as though the model is successfully performing a random walk through the sampling spaces.

In order to test convergence, Gelman-Rubin statistics are plotted for all iterations. These plots (figure A.5), unilaterally show that the statistics converge around 1, indicating that there may be convergence in our model.

The autocorrelation plots in figure A.6 show that there is no autocorrelation for beta or sigma. Finally, we inspected the sample sizes for all parameters as can be seen in table 5, below. The sample sizes well exceed any number data observations and easily cross the threshold of 10,000 iterations.

*Table 5: Expected sample sizes for beta values 1-6 and sigma values 1-3*

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **beta1** | **beta2** | **beta3** | **beta4** | **beta5** | **beta6** | **sigma1** | **sigma2** | **sigma3** |
| 78090.87 | 80000.00 | 80000.00 | 79602.54 | 80431.92 | 81024.18 | 48717.56 | 45843.83 | 47655.60 |

Based on the model diagnostics discussed above, the fork requires no changes in burn-in period, the use of thinning, or additional runs/chains.

To check how the generated model fits with the original data, a simple posterior predictive check was run. That is to say, the mean beta values output from the model were coupled with the original age dependency (old) data to simulate data exactly in the form of a linear regression. The deviations (sigma) are ignored because they are large in magnitude. The original correlations are noisy and have many ‘outliers.’ If sigma was included, there would not have been an easily visually-comprehensible comparison.

Figures B.1 till B.3 show age dependency ratio (old) compared respectively to health expenditure, GDP per capita, and age dependency ratio (young). In each plot, the original data is shown in red and the simulated linear relationship in green. With health expenditure, the simulated data clearly follows a trend. With GDP, ??? (can’t finish writing this until we fix the data input). Lastly, with the age dependency (young), the simulated data indeed follows the trend, but the trend is also clearly non-linear. Rather, it looks to be inversely proportional, which matches with the definitions of both parameters.

As stated previously, the model can improve significantly by incorporating non-linear relationships between variables. Though these types of changes will increase the complexity of standardizing the data, it will offer a better model of how age dependency (old) actually affects societies. Another improvement can be the use of hidden variable analysis. Where there is noisy data, like in the comparison to health expenditure and to GDP, there are realistically likely to be intermediate parameters that link age dependency as an indicator to those parameters; age dependency, as stated previously is not causally linked to health expenditure and GDP. Rather, age dependency was hypothesized to be a parameter that should relate to many other societal indicators.

# Assessment

The performed research has several limitations. In this part these limitations are listed and also suggestions for further research are done, based on these limitations. The limitations are divided into 4 groups. Firstly, the limitations in relation to the concept of the age dependency ratio are discussed, secondly the impact of differences between different countries, thirdly the time-frame that was taken and lastly the quality of the data.

### Limitations towards age dependency ratio

For our model the age dependency ratio was based upon the age of people. This means that all people under 15 and all people above 65 were considered to be the dependent group. In practice, a (significant) number of people above 65 still work. This factor is not accounted for in our research. Also, the impact of the employment ratio is not considered. In practice, unemployed people can be seen dependent too. For future research a suggestion might be to compare obtained numbers to the number of people that are actually working (include people that are 65+ and work, and include the employment rate of people within the working age)

### The chosen time-frame

For this research data from one year only were used. By using data from multiple years, the trend per country could be analysed. This might be very interesting and therefore this is also a suggestion for further research

### Different situations in different countries

Different situations in different countries were not accounted for. For instance, a lower GDP can be caused by many different reasons. In some cases, a wide spread in the scatter plots indicates huge differences between different countries. By grouping similar countries together possible disturbing factors might be recognized more easily. A developing country can have lower age dependency ratio, due to a lower life expectancy, whilst a more developed country can have a lower age dependency ratio due to a change in the retirement age. These causal relations are very different and could be further investigated by grouping countries and analysing these groups, before determining the overall correlation.

### Limitations towards the quality of data

In our research data from many different countries is used. For some countries only, limited amount of data was available. Also, outliers were present and not in all cases accounted for. For future research it could be an option to review the quality of data per country more thoroughly before deciding to use a certain country for the research.

# References

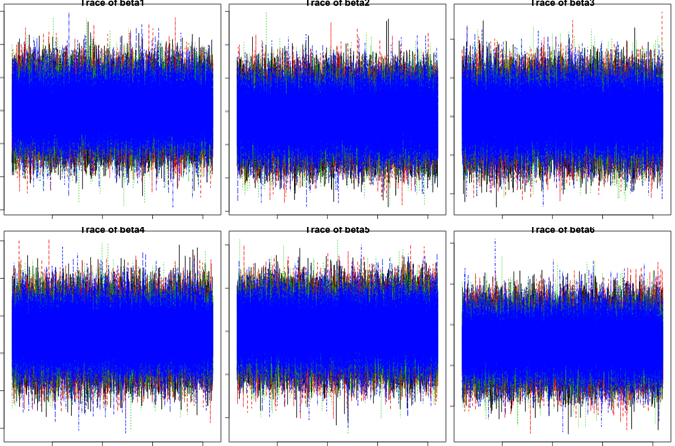
Cunningham, S. (2018). Lecture 19: Interpreting and Evaluating Your JAGS Model [Powerpoint slides]. Retrieved from https://www.dropbox.com/sh/xxj7j1bjwgj296d/AABNeAd9dhkeje6Jxi39\_rLpa/Module%203/Slides

# Appendix

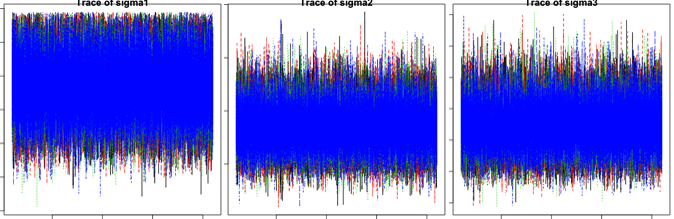
## A: Model diagnostics plots

### 4.1 Burn-in

The figures A.1 and A.2 show the trace plots of the beta and sigma where the burn-in period has been deleted.



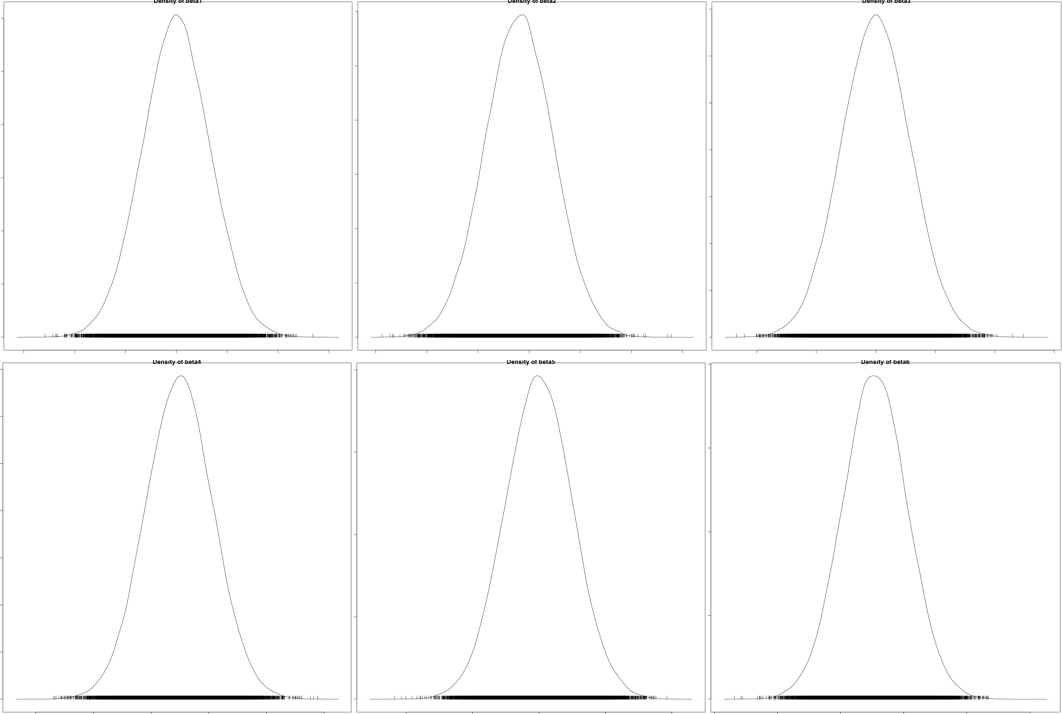
*Figure A.1: Trace plots of beta values 1-6*



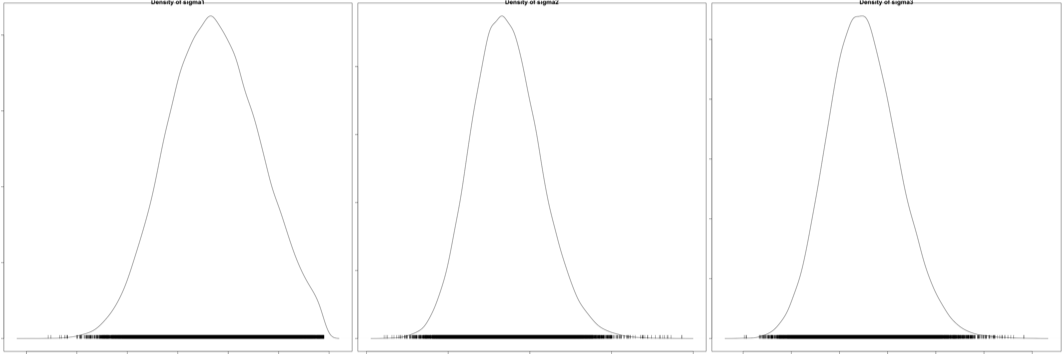
*Figure A.2: Trace plots of sigma values 1-3*

### 4.2 Density

Figures A.3 and A.4 show the different density plots of the betas and sigma's.



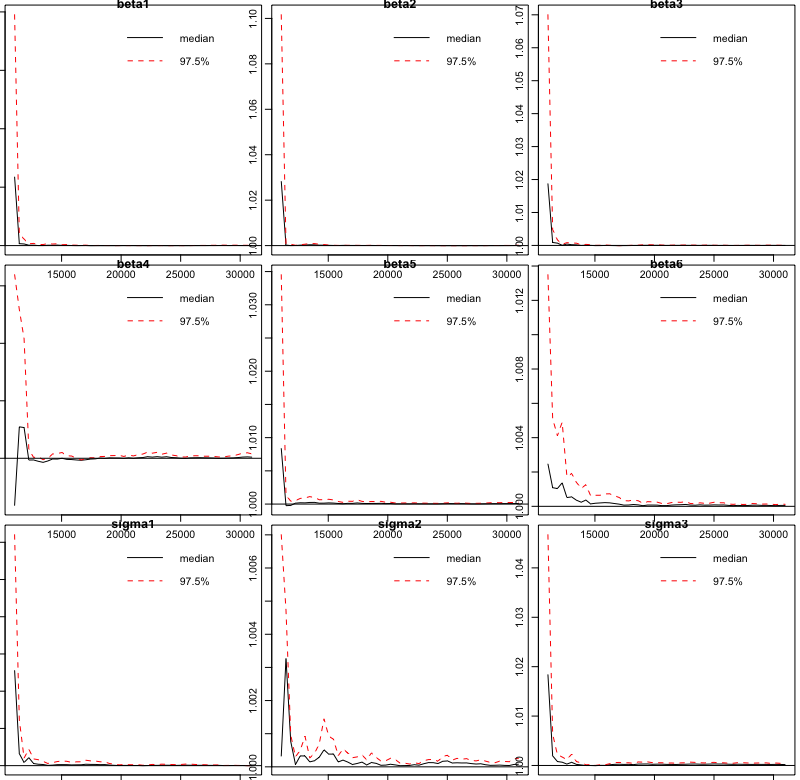
*Figure A.3: Density plots of beta values 1-6*



*Figure A.4: Density plots of sigma values 1-3*

### 4.3 Convergence

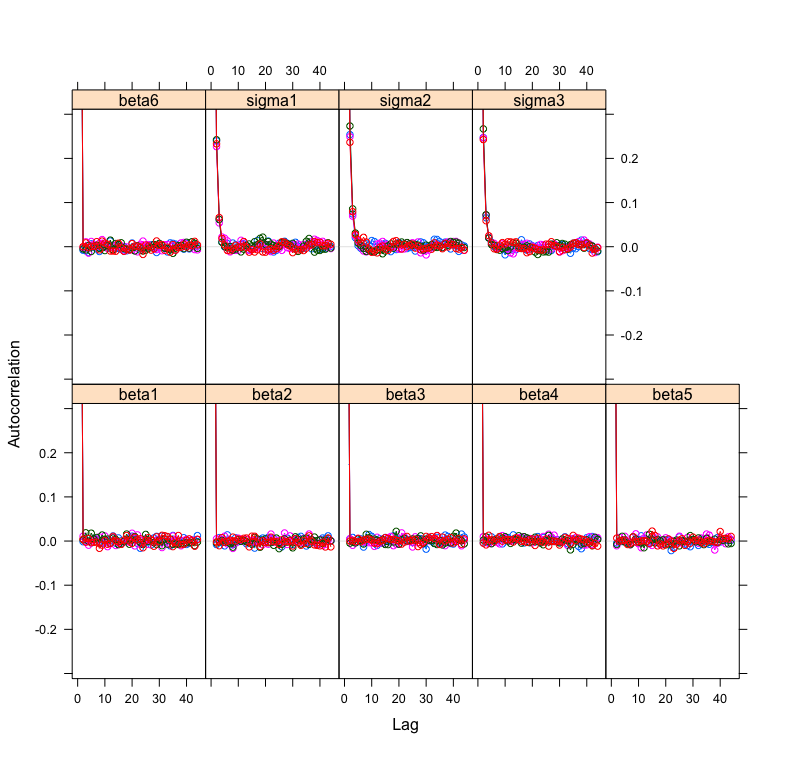
Figure A.5 shows the different convergence graphs for all sigma's and betas.



*Figure A.5: Gelman plots of beta values 1-6 and sigma values 1-3*

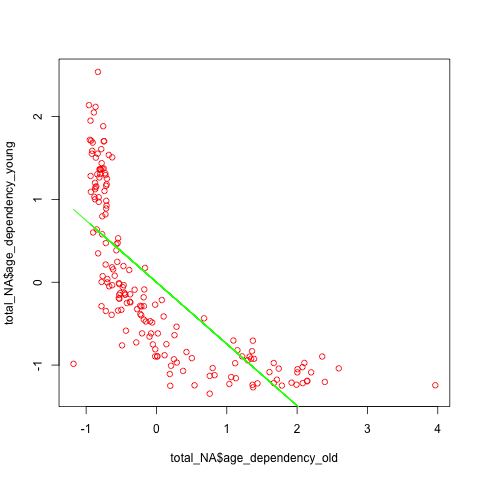
### 4.4 Autocorrelation

Figure A.6 show the autocorrelation of the different betas and sigma's.

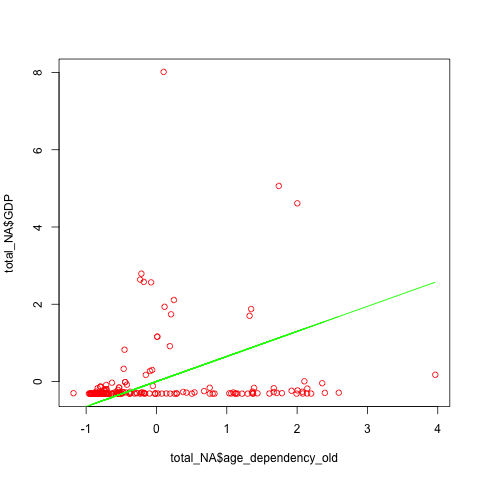


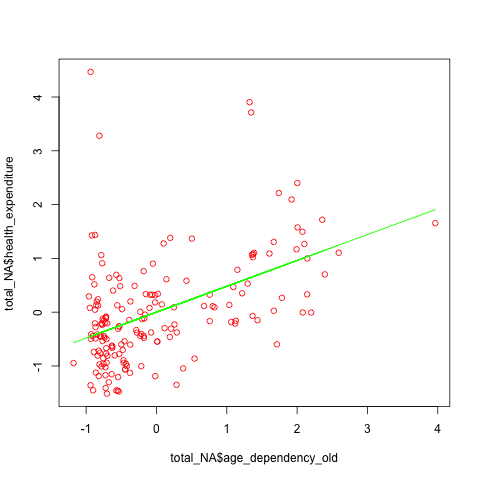
*Figure A.6: Autocorrelation plots for beta values 1-6 and sigma values 1-3*

## B: Output plots

In the figures B.1 untill B.3, the different variables are plot together in a graph. A possible correlation line was drawn as well.

*Figure B.1: Scatter plot with correlation between age dependency old and age dependency young*

*Figure B.2: Scatter plot with correlation between age dependency old and GDP*

*Figure B.3: Scatter plot with correlation between age dependency old and health expenditure*