Descriptive statistics of Life Cycle Savings R Database

Jonas Markgraf

22 February 2016

This file contains descriptive analysis - measures of central tendency as well as measures of dispersion - of the LifeCycleSavings R dataset. Furthermore, it will be looked into correlations between different variables. A particular focus will be on the two variables *Aggregated Savings (sr)* and *Per-Capita Income (dpi)*, thus testing the hypothesis that aggregated savings increase with higher per-capita income.

Description of dataset and variables of interest

The LifeCycleSavings dataset is a data frame containing information on the savings ratio between 1960 and 1970 over 50 countries. It includes the following 5 variables (taken from R):

- ***sr***: aggregate personal savings
- ***pop15***: % of population under 15
- ***pop75***: % of population over 75
- ***dpi***: real per-capital disposable income
- ***ddpi***: % growth rate of dpi.

Guide to R code

The R code in this folder conducts a number of descriptive statistics, namely measures of central tendency and measures of dispersion. Furthermore, correlations between variables were analyzed as a first step to identify potential explanatory variables (per-capita GDP; per-capita GDP growth; demographic factors) for the aggregated personal savings.

Measures of central tendency: histograms, mean, median

Measures of dispersion: standard deviation, range, interquartile range, boxplots

Correlations: plots; significance tests

Descriptive Statistics

Measures of central tendency

First, we have a look at the key measures of central tendency starting with the means of the five variables:

```
for (i in 1:5) {
  LifeCycleSavings[, i] %>%
    mean() %>%
    paste(names(LifeCycleSavings)[i], ., "\n") %>%
    cat()
}
```

```
## sr 9.671
## pop15 35.0896
## pop75 2.293
## dpi 1106.7584
## ddpi 3.7576
```

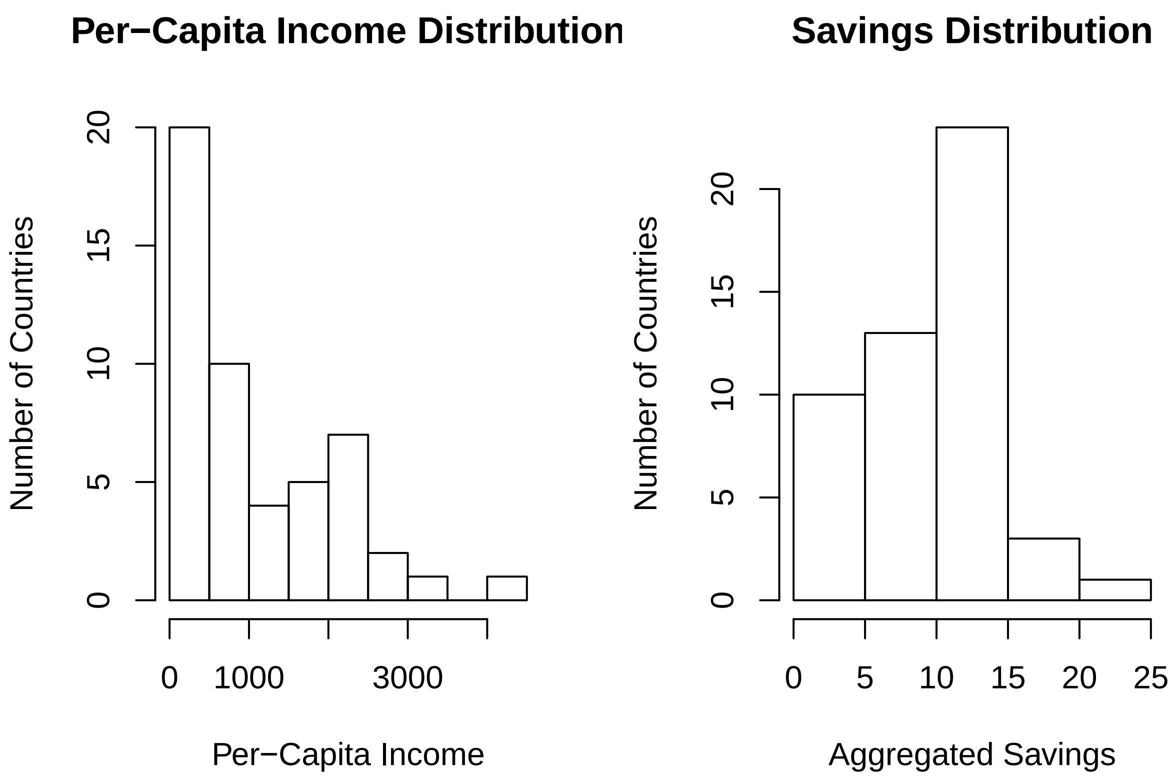
As a next step, the medians are calculated:

```
for (i in 1:5) {
  LifeCycleSavings[, i] %>%
    median() %>%
    paste(names(LifeCycleSavings)[i], ., "\n") %>%
    cat()
}
```

```
## sr 10.51
## pop15 32.575
## pop75 2.175
## dpi 695.665
## ddpi 3
```

After the calculating mean and median, we have a look at the histograms of 2 key variables - Per-Capita Income and Aggregated Savings - in order to get an idea of the general distribution of those two variables:

```
par(mfcol = c(1, 2))
hist(LifeCycleSavings$dpi,
     main="Per-Capita Income Distribution",
     xlab="Per-Capita Income",
     ylab="Number of Countries")
hist(LifeCycleSavings$sr,
     main= "Savings Distribution",
     xlab="Aggregated Savings",
     ylab="Number of Countries")
```



Measures of dispersion

In a next step, we analyse the distribution for the variables in the *LifeCycleSavings* R Dataset. For that purpose, the standard deviation for all five variables is calculated.

```
## loop for standard deviation
for (i in 1:5) {
  LifeCycleSavings[, i] %>%
    sd() %>%
    paste(names(LifeCycleSavings)[i], ., "\n") %>%
    cat()
}
```

```
## sr 4.48040689205426
## pop15 9.15172716162454
## pop75 1.29077140359032
## dpi 990.868888965557
## ddpi 2.8698706221283
```

Furthermore, we look at the range and the interquartile range of the the key variables *aggregated savings* and *per-capita income*.

```
## range for 2 key variables
range(LifeCycleSavings$sr)
```

```
## [1]  0.6 21.1
```

```r
range(LifeCycleSavings$dpi)
```

```
## [1]   88.94 4001.89
```

```r
## interquartile range for 2 key variables
IQR(LifeCycleSavings$sr)
```

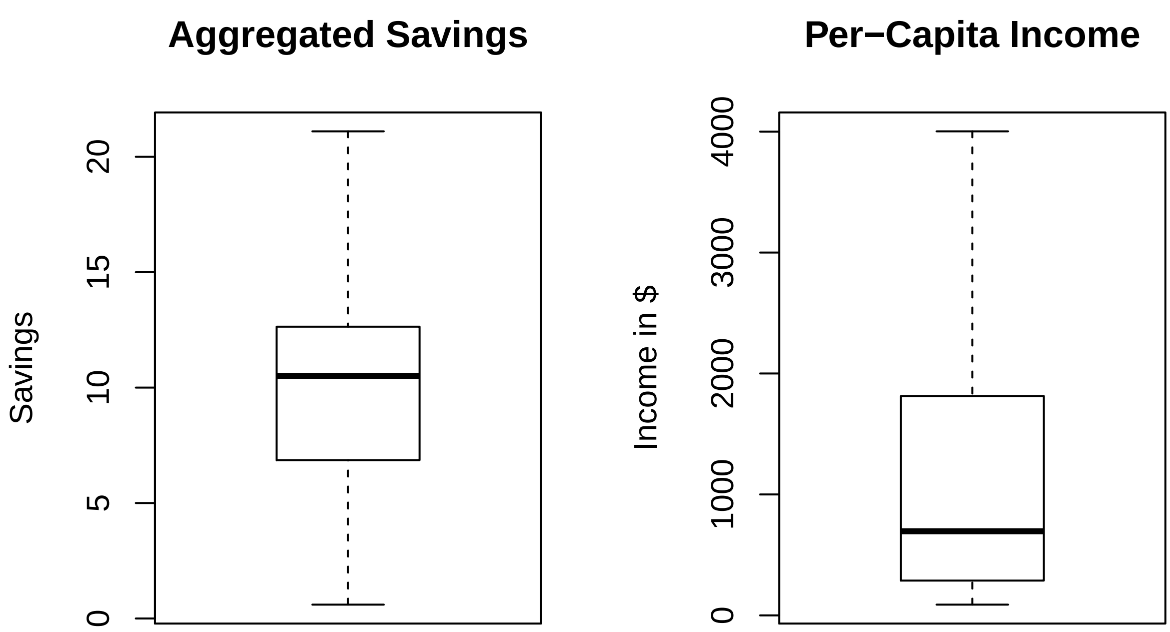
```
## [1] 5.6475
```

```r
IQR(LifeCycleSavings$dpi)
```

```
## [1] 1507.415
```

Eventually, we graphically plot the measures of central tendency and dispersion for our two key variables with two boxplots, which provides a graphical presentation of mean, range and interquartile range.

```r
par(mfcol = c(1, 2))
boxplot(LifeCycleSavings$sr, main = "Aggregated Savings", ylab = "Savings")
boxplot(LifeCycleSavings$dpi, main = "Per-Capita Income", ylab = "Income in $")
```



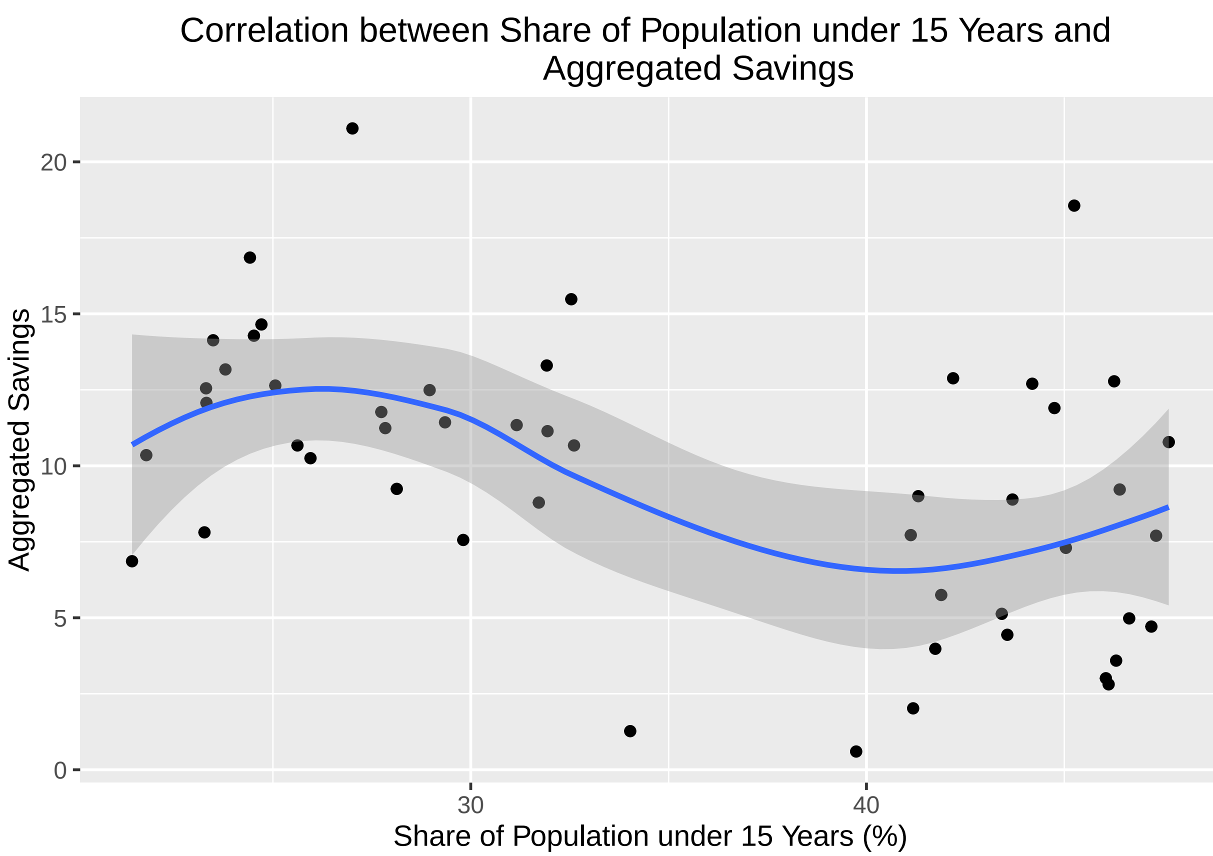
Joint distributions

In order to identify potential explanatory variables for the *aggregated savings* variable, we look into correlations between the the demographic variables (*pop15* and *pop75*) as well as different specifications of the income in the respective country (absolute (*dpi*) and relative (*ddpi*) measures of per-capita income).

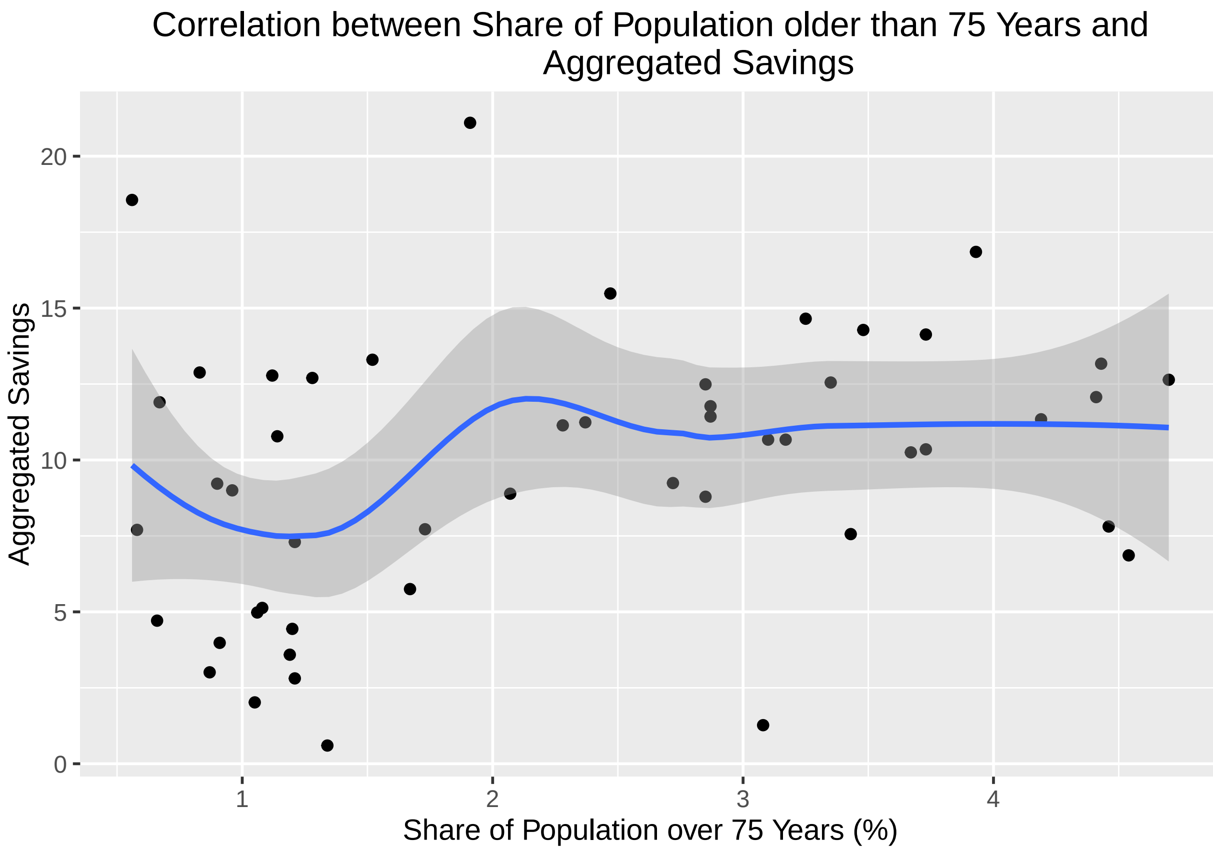
Demographic factors

First we start with the two demographic variables: The share of population older than 75 years and younger than 15 years.

```
ggplot(LifeCycleSavings, aes(pop15, sr)) +
  geom_point() +
  geom_smooth() +
  ggtitle("Correlation between Share of Population under 15 Years and
          Aggregated Savings") +
  xlab("Share of Population under 15 Years (%)") +
  ylab("Aggregated Savings")
```




```
ggplot(LifeCycleSavings, aes(pop75, sr)) +
  geom_point() +
  geom_smooth() +
  ggtitle("Correlation between Share of Population older than 75 Years and
          Aggregated Savings") +
  xlab("Share of Population over 75 Years (%)") +
  ylab("Aggregated Savings")
```

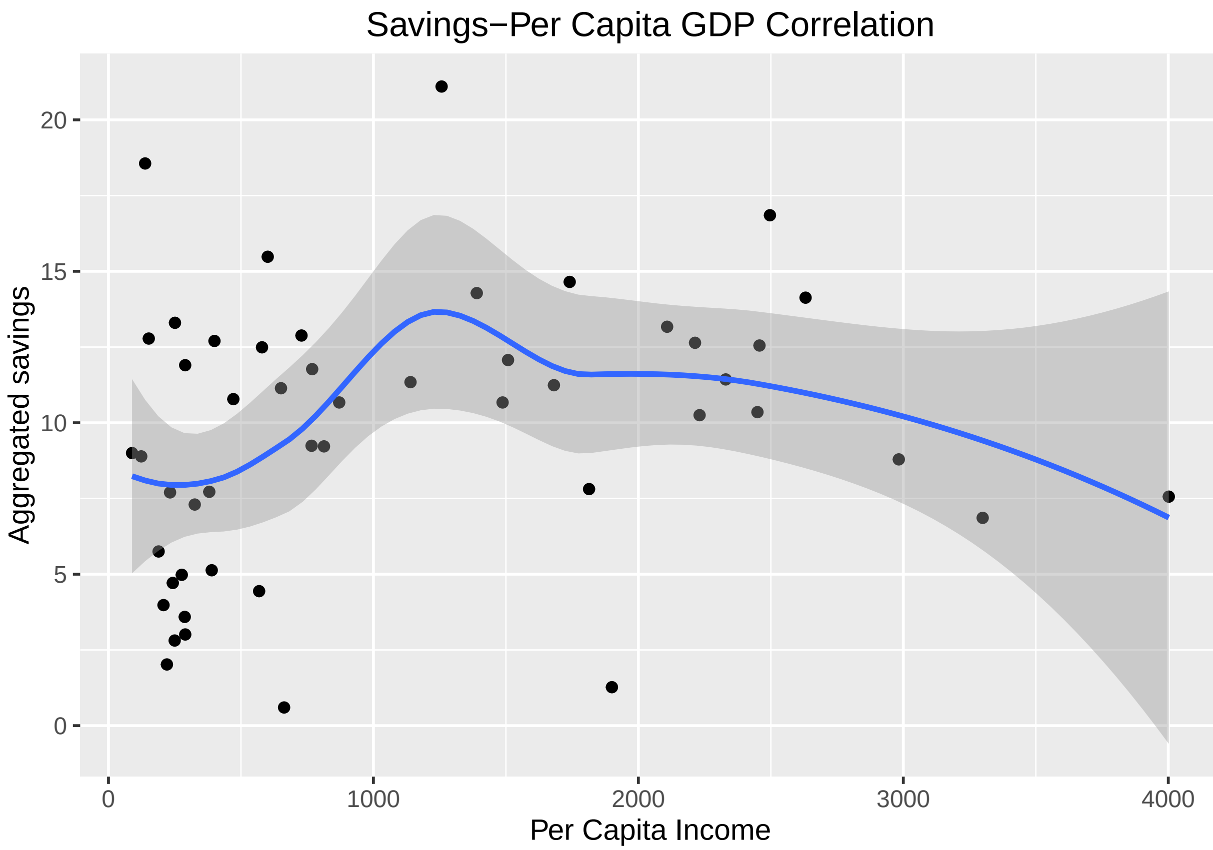


There seems to be a slightly negative correlation between *pop15* and *sr*, confirming the common wisdom that the younger the population, the lower the overall savings. At the same time, *pop75* and *sr* appear to be weakly positively correlated, indicating that aging populations tend to save more (this effect however seems to disappear for pop75>2%).

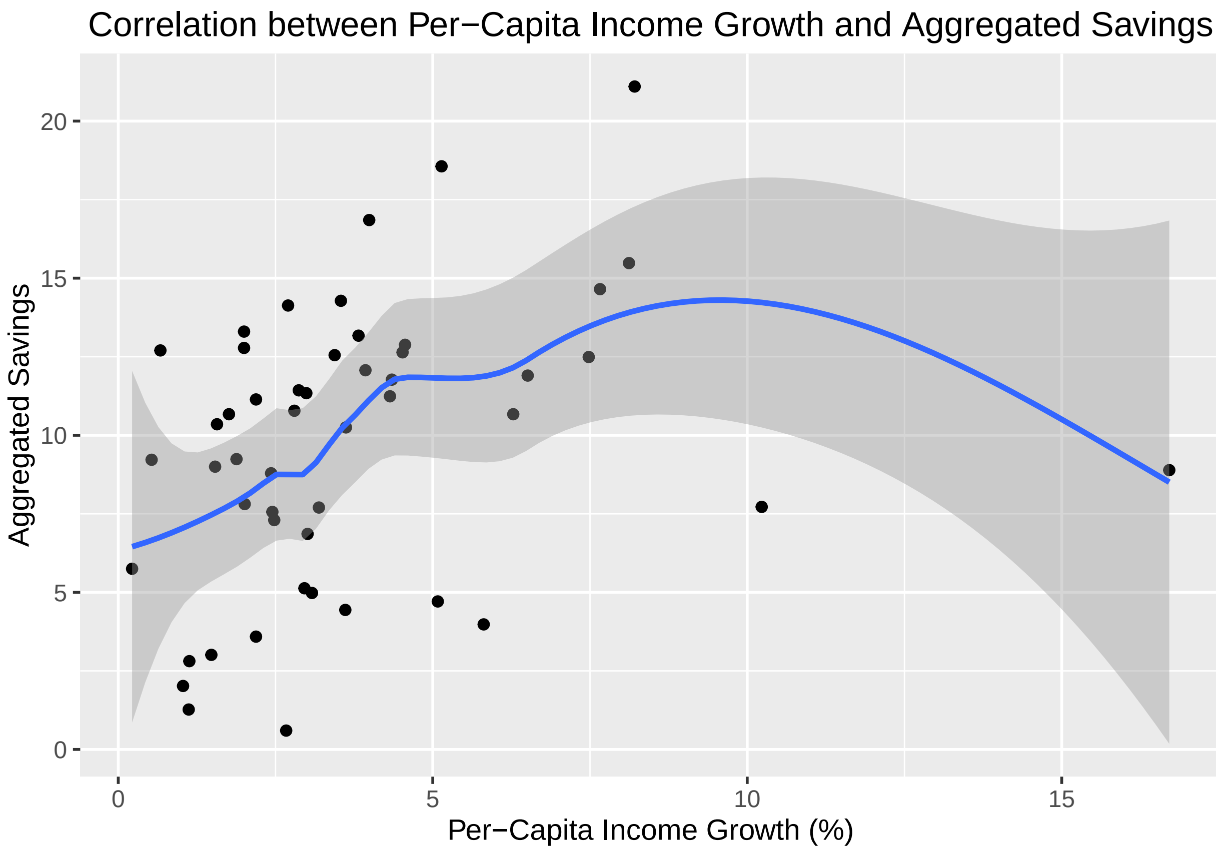
Income factors

In a next step, we attempt to better understand the relationship between savings and per-capita income by looking at the correlation of *aggregated savings* and absolute per-capita income level *dpi* and relative per-capita income growth *ddpi*.

```
ggplot(LifeCycleSavings, aes(dpi, sr)) +
  geom_point() +
  geom_smooth() +
  ggtitle("Savings-Per Capita GDP Correlation") +
  xlab("Per Capita Income") +
  ylab("Aggregated savings")
```




```r
ggplot(LifeCycleSavings, aes(ddpi, sr)) +
  geom_point() +
  geom_smooth() +
  ggtitle("Correlation between Per-Capita Income Growth and Aggregated Savings") +
  xlab("Per-Capita Income Growth (%)") +
  ylab("Aggregated Savings")
```



While absolute per-capita income level (*dpi*) seems to be positively correlated with aggregated savings for small values of *dpi* (less than $1250), this changes for larger values of per-capita income levels; this however seems to be largely driven by an outlier (the US). a similar finding provides the correlation plot between per-capita income growth (*ddpi*) and aggregate savings, although the correlation is more strongly positively correlated; again, this positive relationship disappears for very high growth rates of per-capita GDP (mainly caused by an outlier).