SDS Project

```
## Loading required package: WDI

## Loading required package: MASS

## Loading required package: Matrix

## Loading required package: mvtnorm

## Loading required package: sandwich

## mediation: Causal Mediation Analysis
## Version: 4.5.0

## -- Attaching packages --------------------------------------- tidyverse 1.3.0 --

## v tibble  3.1.0     v dplyr   1.0.5
## v tidyr   1.1.3     v stringr 1.4.0
## v readr   1.4.0     v forcats 0.5.1
## v purrr   0.3.4

## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x tidyr::expand()     masks Matrix::expand()
## x dplyr::filter()     masks stats::filter()
## x dplyr::group_rows() masks kableExtra::group_rows()
## x dplyr::lag()        masks stats::lag()
## x tidyr::pack()       masks Matrix::pack()
## x dplyr::select()     masks MASS::select()
## x tidyr::unpack()     masks Matrix::unpack()
```

List of countries of interest

We consider all European countries: Austria, Belgium, Bulgaria, Switzerland, Cyprus, Czech Republic, Germany, Denmark, Spain, Estonia, Finland, France, United Kingdom, Greece, Croatia, Hungary, Iceland, Italy, Lithuania, Luxembourg, Latvia, Malta, Netherlands, Norway, Poland, Portugal, Romania, Slovakia, Slovenia and Sweden. The ISO country codes can be searched from https://www.iban.com/country-codes.

Time period of interest

In order to keep the time period consistent between the world bank and the LinkedIn data, we choose the time period of interest from 2015 to 2019.

Economic indicators of interest

For these indicators, we extract the data from the world bank interface. The entire list of indicators can be seen here: https://databank.worldbank.org/source/world-development-indicators.

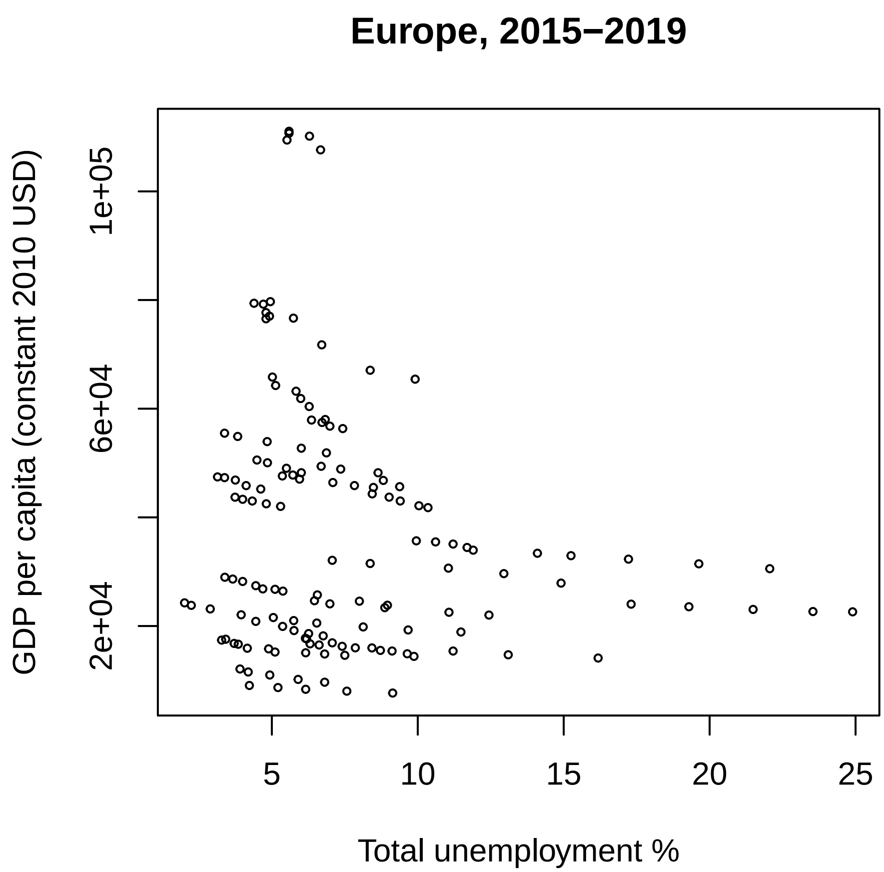
First, we need the GDP per capita. This is the main variable of interest. For the analyses, we consider this as the dependent variable, and seek to understand the effect of various unemployment related parameters on the GDP per capita. There are several indicators of GDP per capital: constant 2010 US dollars, current US dollars, constant LCU, current LCU, etc. We use the GDP per capita in constant 2010 US dollars for our analysis.

Among the unemployment indicators, we consider (i) the total unemployment percentage, (ii) the unemployment percentage in youths (ages 15-24), (iii) the unemployment percentage among males, and (iv) the unemployment percentage among females. In the world bank interface, for all these indicators, we have two estimates: (i) a national estimate and (ii) a modeled ILO estimate. For our analysis, we consider the national estimate.

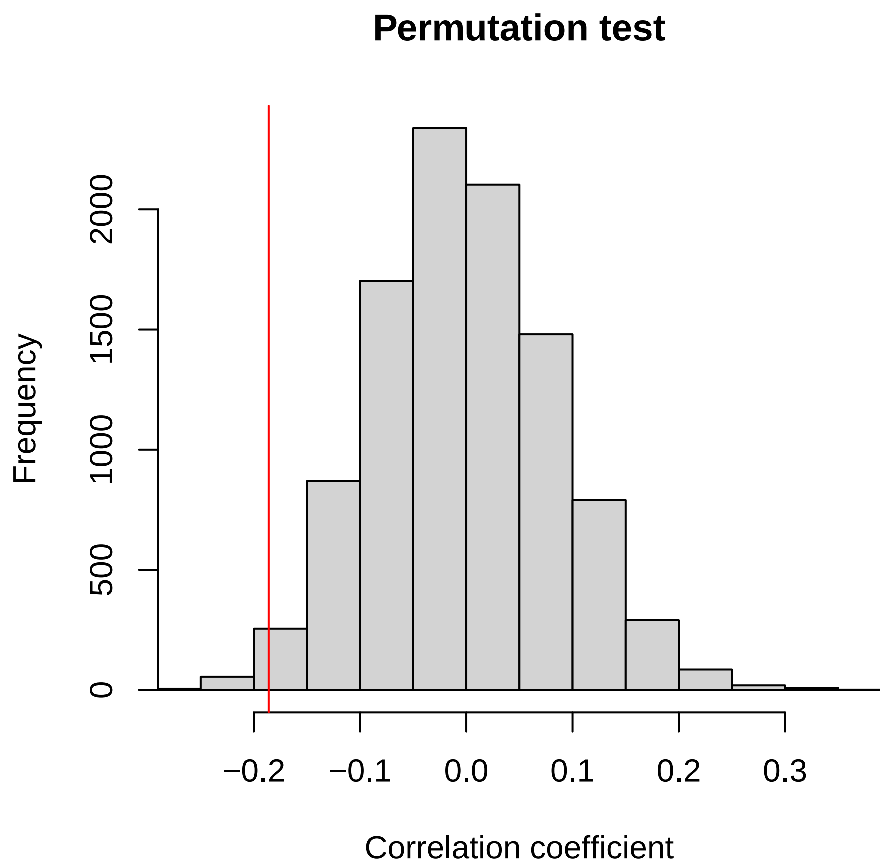
Finally, we consider the division of the total employment in different industry sectors: namely (i) agriculture (consisting of activities in agriculture, hunting, forestry and fishing), (ii) industry (consisting of mining and quarrying, manufacturing, construction, and public utilities (electricity, gas, and water) and (iii) services (consisting of wholesale and retail trade and restaurants and hotels; transport, storage, and communications; financing, insurance, real estate, and business services; and community, social, and personal services). For these indicators, only one estimate (the modeled ILO estimate) is available, and we use the same.

Correlation between total unemployment and GDP per capita (2015 - 2019)

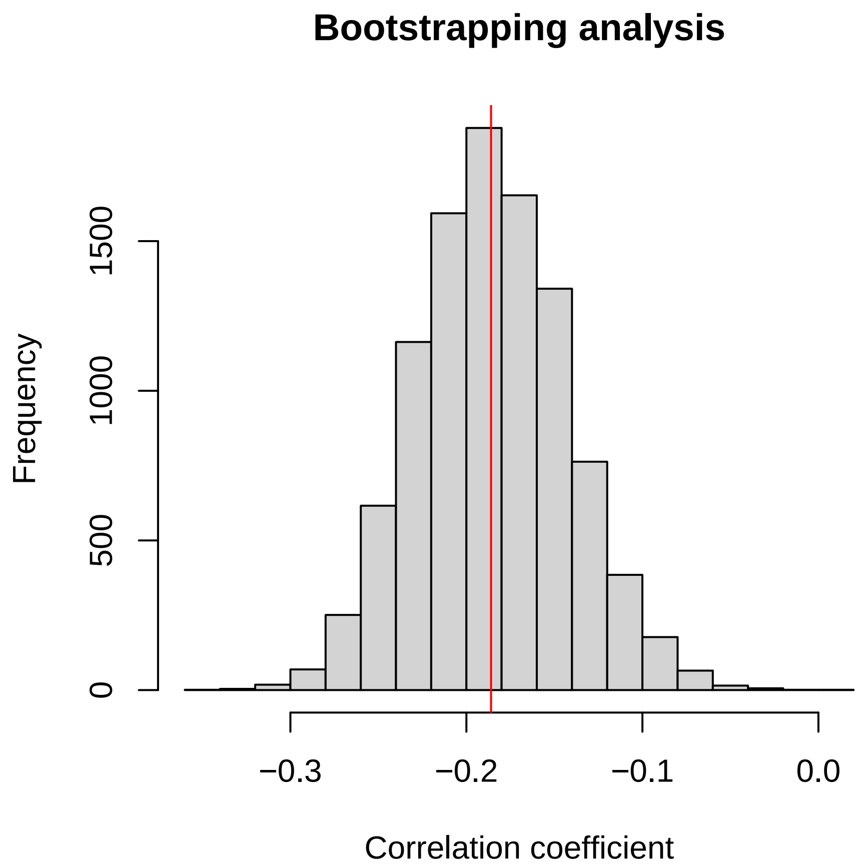
The Pearson correlation coefficient between the total unemployment rate and the per capita GDP between 2015 and 2019 in European countries is -0.186.



In order to check the statistical significance of the computed correlation, let's do a permutation analysis. Taking 10^4 permutations, we find the p-value of the measured correlation to be 0.0103. As the p-value is smaller than 0.05, the measured correlation is statistically significant.



Further, let's do a bootstrap analysis to quantify the uncertainty in the measured correlation value. Repeating the bootstrap analysis 10^4 times, we find the mean correlation to be -0.185, with a standard deviation of 0.0425. Thus, although the measured coefficient is unbiased, there seems to relatively high uncertainty in the measurement. This shows the unreliability of the measured correlation, and indicates that the real correlation value for a larger data-set may be remarkably different.



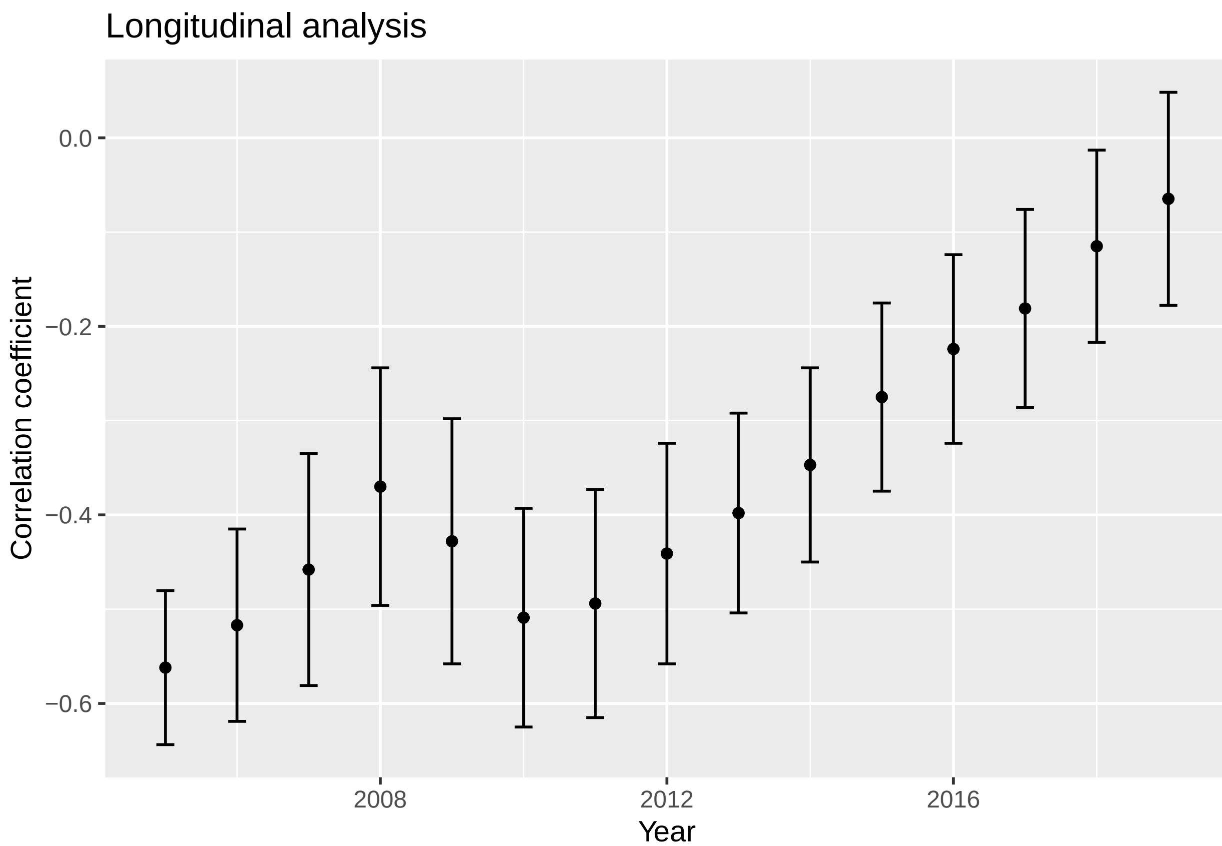
Yearwise analysis (2005 - 2019)

The substantial uncertainty in the measured correlation could be due to multiple reasons. It may indicate that the relationship between the GDP per capita and the total unemployment is complicated, and perhaps dependent on several other variables. Further, the cumulative analysis over the 5 year time period (2015 - 2019) may be further complicating the relationship by introducing additional variations that may have occurred in different countries in different years within the considered time period. In order to alleviate this complication, we study the relationship between the GDP per capita and the total unemployment for each year separately. As there is no dependency of this yearwise analysis on the LinkedIN analysis, and to benefit from the additional data available via the WDI interface, we include an extended time period: from 2005 to 2019.

```
## [1] "Year | Correlation | BT mean | BT std | p-value"
```

```
## [1] "2005 -0.562 -0.569 0.0817 1e-04"
## [1] "2006 -0.517 -0.516 0.102 8e-04"
## [1] "2007 -0.458 -0.451 0.123 0.0035"
## [1] "2008 -0.37 -0.373 0.126 0.0209"
## [1] "2009 -0.428 -0.431 0.13 0.0056"
## [1] "2010 -0.509 -0.51 0.116 0.0012"
## [1] "2011 -0.494 -0.503 0.121 0.0019"
## [1] "2012 -0.441 -0.456 0.117 0.0031"
## [1] "2013 -0.398 -0.412 0.106 0.0083"
## [1] "2014 -0.347 -0.357 0.103 0.0161"
```

```
## [1] "2015 -0.275 -0.282 0.0998 0.0559"
## [1] "2016 -0.224 -0.227 0.1 0.1067"
## [1] "2017 -0.181 -0.177 0.105 0.1671"
## [1] "2018 -0.115 -0.106 0.102 0.2995"
## [1] "2019 -0.0647 -0.0498 0.113 0.3988"
```



We observe a clear trend in the yearwise analysis: the correlation between the total unemployment and the GDP per capita becomes consistently weaker from 2010 to 2019. The error bars in the plot above indicate the standard deviation values computed via bootstrapping. Further, permutation testing shows that the correlation values before 2016 are statistically significant, while the ones from 2016 are not.

Based on this analysis, we conclude that there may be several time-dependent factors affecting the relationship between the GDP per capita and the total unemployment. A multivariate analysis including such additional factors is outside the scope of this project. Accordingly, we decide to focus on a single year for the further analysis. As the correlation is statistically significant before 2016, and we have LinkedIn data available from 2015 onwards, 2015 is the only year which matches both our analysis criteria.

Group Analysis (2015)

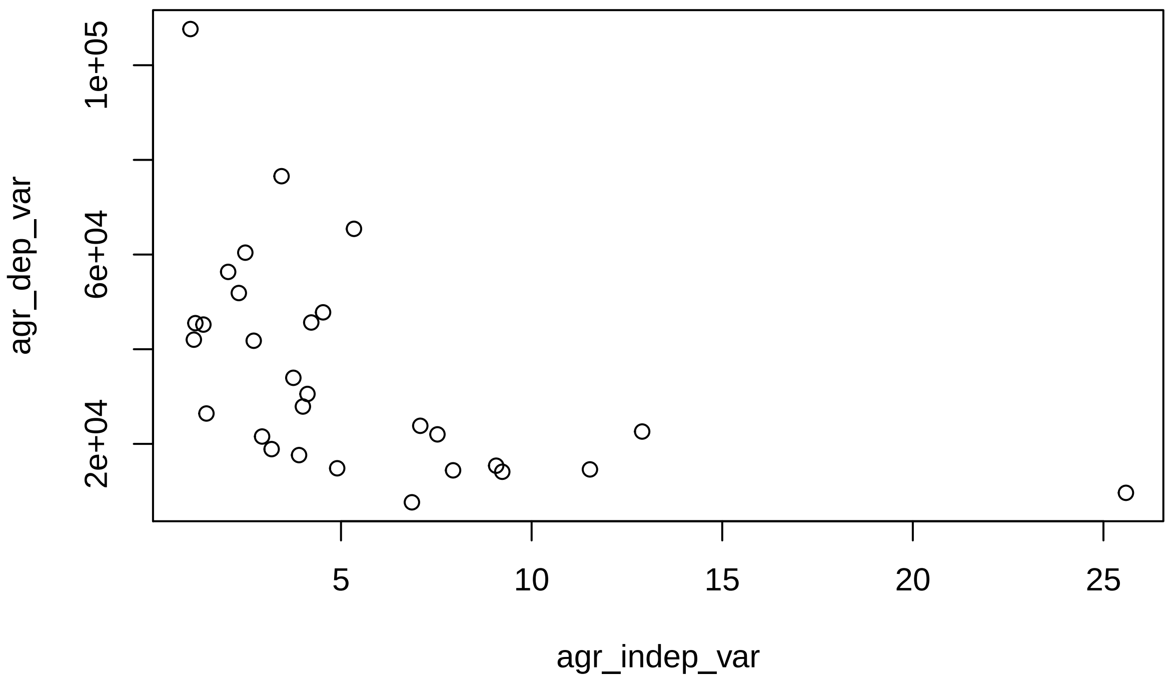
The table below shows the correlation values for different sub-groups in the population. It can be seen that the correlation values are quite similar for all the considered groups.

Mediation Analysis (2015)

```
## Running nonparametric bootstrap
##
## Running nonparametric bootstrap
```

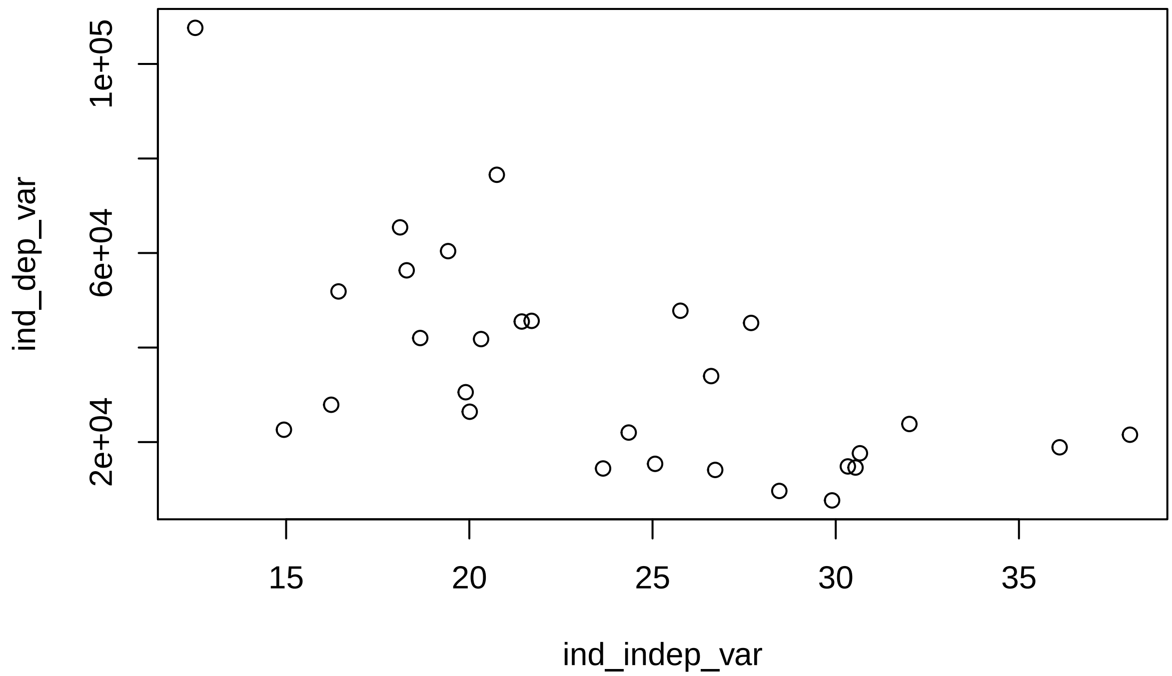
	Correlation	Bootstrapping mean	Bootstrapping std. dev.	P-value
Total	-0.275	-0.276	0.0398	0.0520
Youth	-0.287	-0.288	0.0420	0.0535
Male	-0.296	-0.296	0.0450	0.0477
Female	-0.251	-0.253	0.0379	0.0729

	I->D	p-value	I->M	p-value	ACME	p-value
Youth	-2330.969	0.0050255	0.6402022	0.1435720	-202.9785	0.306
Female	-2330.969	0.0050255	0.2482897	0.2463356	-149.0624	0.628




```
## Running nonparametric bootstrap
##
## Running nonparametric bootstrap
```

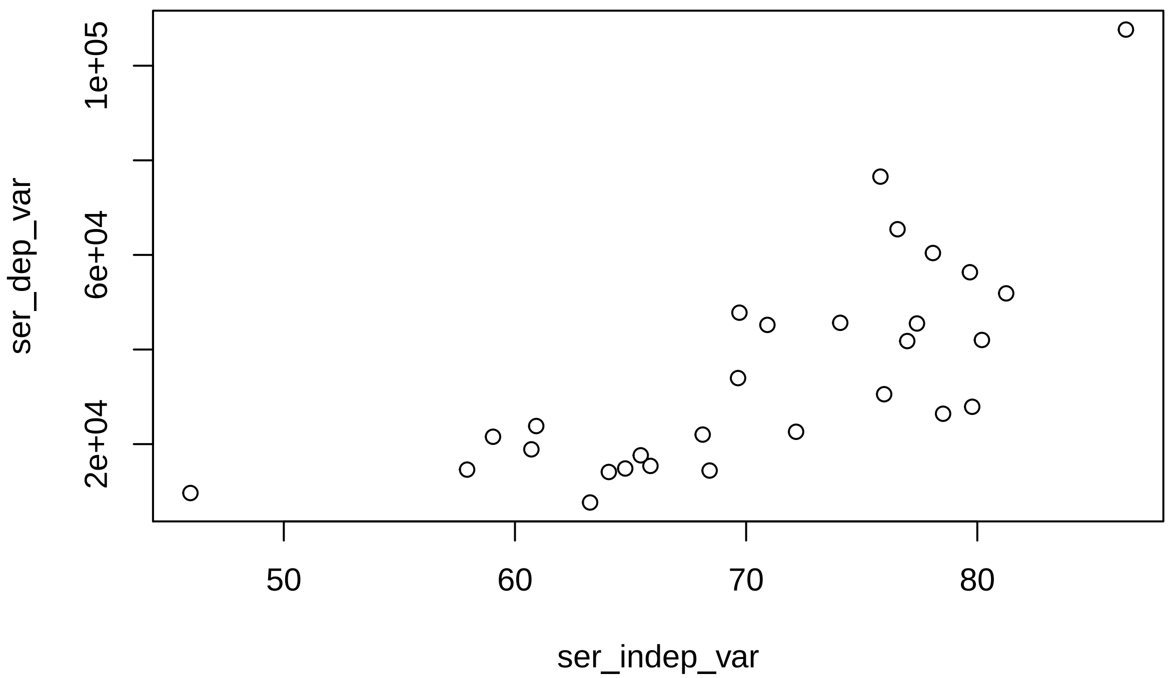
	I->D	p-value	I->M	p-value	ACME	p-value
Youth	-2238.58	0.0003038	-0.2893222	0.4036032	230.4218	0.374
Female	-2238.58	0.0003038	-0.1903536	0.2552074	318.8837	0.266




```
## Running nonparametric bootstrap
##
## Running nonparametric bootstrap
```

	I->D	p-value	I->M	p-value	ACME	p-value
Youth	1877.316	7.1e-06	-0.0519346	0.8352255	26.83183	0.692
Female	1877.316	7.1e-06	0.0199327	0.8694560	-22.56174	0.824

	Correlation	Bootstrapping mean	Bootstrapping std. dev.	P-value
Agriculture	-0.507	-0.512	0.0247	2e-04
Industry	-0.623	-0.623	0.0426	1e-04
Services	0.730	0.731	0.0249	1e+00



Sector Analysis with WDI data (2015)

In the remaining part of the project, we consider the effect of different employment sectors on the GDP. The table below shows the correlation values for different employment sectors. It can be seen that employment in agriculture and industry is negatively correlated with GDP per capita, while employment in services is positively correlated with the GDP per capita.

Sector Analysis with LinkedIN data (2015)