Healthy communities - Understanding determinants of HIV

12 December 2014

Contents

1 Executive Summary

2 Introduction

The expiry date of the Millennium Development Goals (MDGs) is just around the corner, meanwhile the post-2015 agenda is being discussed intensively. In this context, it is important to assess the achievement of the MDGs and try to understand why some goals have not been reached.

Reducing HIV incidence is an important aim of the MDGs. Target 6.A of the MDGs specifies that countries should "have halted by 2015 and begun to reverse the spread of HIV/AIDS" (United Nations 2014). In most regions of the world this goal has been fulfilled: new HIV infections declined and the overall number of new HIV/AIDS infections per 100 adults (15-49 years old) decreased by 44 per cent between 2001 and 2012 (United Nations 2014). However, this trend cannot be observed in all 189 member states of the United Nations. On the contrary, HIV/AIDS prevalence has even increased in some countries.

---

3 Aim, Research Question and Hypotheses

This paper aims to provide evidence to assess why some countries struggle to achieve MDG 6A. We believe that one possible explanation for the failure of some interventions in reducing HIV/AIDS may lie in the lack of a full understanding of the determinants of the disease, which can in turn lead to ill-specified interventions and wrongly targeted campaigns.

The literature reviewed for this paper identifies a myriad of determinants of health. When it comes to specific diseases however, the existing literature only provides a narrow selection of potential determinants. The aim of this paper is to test whether social and community networks (often neglected by the literature on disease-specific determinants of health) significantly explain HIV/AIDS incidence rates at the country level. The hypothesis of this paper is that female school enrollment is a strong predictor of HIV incidence rates. By identifying variables that help to explain HIV/AIDS incidence, this paper will help move forward the discussion of the determinants of HIV/AIDS .

---

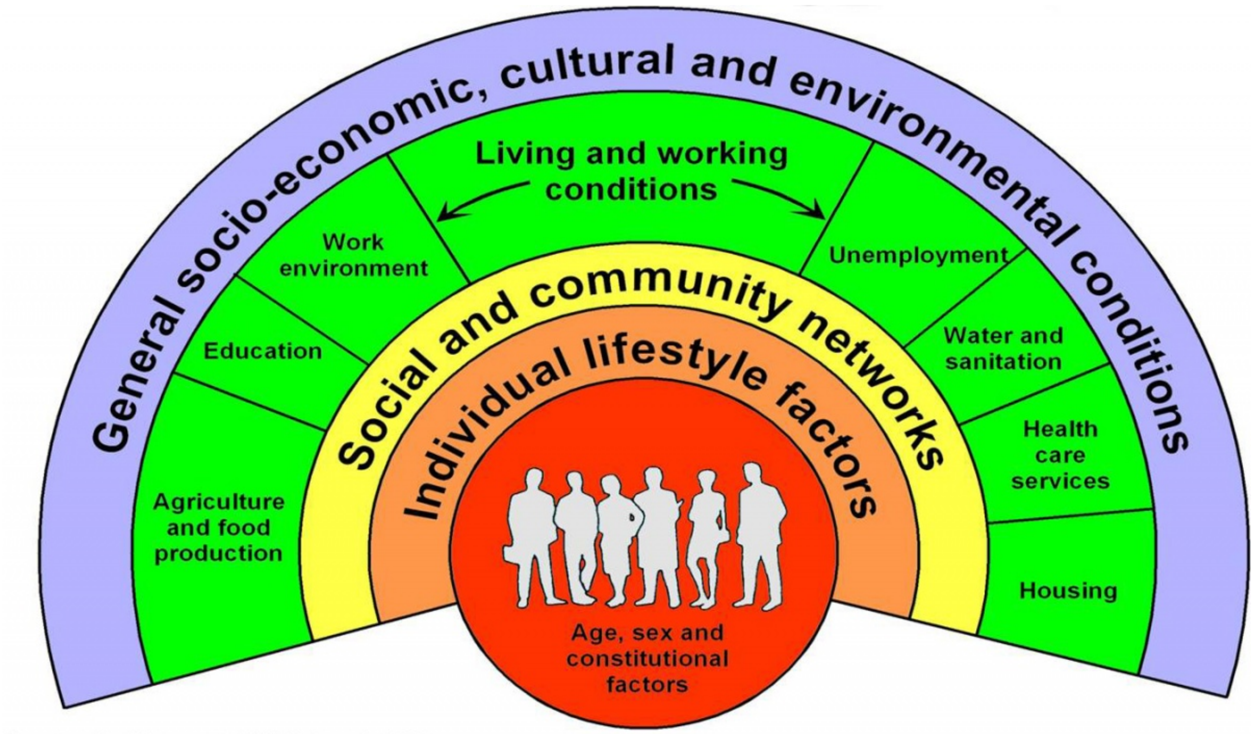
4 Literature Review

Hurrelmann (Hurrelmann 1989, 76) advocates an interdisciplinary framework for analysing what determines health outcomes. He considers it necessary to use a model that integrates all the aspects of the organism, individual and the environment.

One framework that shows the interaction between individual and environmental factors over time is the salutogenic model developed by Antonovsky in 1979. According to Hurrelmann, Antonovsky's model is a great contribution to interdisciplinary theory, but the downside is its complexity (Hurrelmann 1989).

A simpler and more common model on general determinants of health is the "rainbow model", developed by Dahlgren and Whitehead (Dahlgren and Whitehead 1991, 11). This model gives an overview of the main health determinants, reflecting the relationship between the individual, its environment and different health outcomes. Individuals are at the centre of the model with a set of fixed biological and genetical preconditions. Building upon these, four layers of influence on health can be identified: individual lifestyle factors, social and community networks, living and working conditions and general socio-economic, cultural and environmental conditions.

Figure 1: Main health determinants



The link between health, education and gender was extensively researched by Sachs and Dupas. While Sachs provides several examples of the interaction between these dimensions at the macro level, Dupas successfully tested the linkage between female education and health by conducting randomized control trials in Kenya. Among the findings of her research, it is worth mentioning the significant effect of risk reduction campaigns on HIV incidence. These campaigns were specifically targeting girls. (REF)

---

5 Data Sources

This paper utilises data from two different sources. For HIV/AIDS incidence rates at the country level, this research explored several databases such as UNAIDS, World Bank, Global Fund for AIDS, Tuberculosis and Malaria, WHO, the Institute for Health Metrics and Evaluation, PEPFA and the AIDS Data Hub. The database used for the data analysis was the one from UNAIDS given that it provided panel data for the period 2000-2012. All the other variables used in this research were obtained from the World Development Indicators (WDI) from the World Bank.

5.1 World Development Indicators (WDI)

The WDI database comprises 1342 indicators clustered in 10 thematic areas that range from health and education to infrastructure and public sector data. Information is available for 214 countries and dates back to 1960. All indicators are available for free at the World Bank website and can be downloaded as an Excel sheet, CSV, tabbed TXT or SDMX. In addition, there is a special R package WDI designed to download and use the data.

WDI have been used in a wide range of fields and HIV/AIDS research is not an exception. For examples of relevant literature that also make use of WDI please see (Haacker 2002), (Talbott 2007) and (Kalemli-Ozcan 2011). A list of all WDI indicators used for this research can be found in the Appendix.

5.2 UNAIDS Dataset

The UNAIDS dataset is available for free at the UNAIDS website (REF) and it has been put together by UNAIDS during the research for the UNAIDS GAP Report 2014 (REF). It provides data on HIV incidence, prevalence and HIV-related deaths for the period 1990-2013. The dataset aditionally provides uncertainty bounds (low and high estimated) for most of the variables. Further, there is information for all UN member states and there is also regionally aggregated data. Some of the indicators such as prevalence and incidence are disaggregated for the differents demographic groups, including data for adults, youth and children, male and female.

UNAIDS databases are often used in public health research. Some examples of papers using UNAIDS data on HIV incidence and prevalence are Letamo (REF), Bennell (REF) and Ferlay et al.().

The data is available as an Excel file.

6 Data Gathering and Cleaning

This section focuses on the process of gathering the data and cleaning the databases to prepare the variables for the data analysis.

The first step in this process was uploading the databases to R Studio. The first dataset consists of 29 World Development Indicators and it was downloaded from World Bank's website. These indicators represent the independent variables used for this research plus the population indicator that is used to filter small countries. Provided that the focus of this research is on country level data, all regionally aggregated data was dropped. Further, 169 rows that contained only NA values were deleted.

After dropping empty rows, the data frame was alphabetically (ascending) ordered, rows were grouped by iso2c code and variables were renamed.

The dataset was further cleaned preparing the data for imputation using the AMELIA package. This process requires that the panel is as balanced as possible, as it feeds from all variables to predict values for the missing observations. A more detailed explanation of the imputation process will be providid in the inferential statistics section of this paper. In order to improve the balance of the panel, the next step consisted of dropping variables for which more than 80% of the observations (552) were missing. In addition, countries with a population smaller than one million inhabitants were dropped from the database. 59 countries fell in that category: 46 islands, 5 European countries (Andorra, Liechtenstein, Luxemburg, Monaco and Montenegro), Bahrain, Bhutan, Belize, Djibouti, Equatorial Guinea, Guyana, Qatar and Suriname. Dropping these countries does not affect the research as the remaining database still contains a highly heterogeneous sample both in geographic and socio-economic terms. Furthermore, deleting these countries improves the dataset as most of these countries lacked information for most of the studied variables.

The second database used for this research was downloaded from UNAIDS' website and it provides information on HIV/AIDS incidence rates (as well as prevalence and deaths caused by HIV/AIDS). The data is publicly available. All columns except the country and the incidence rate were dropped. After renaming the variables, a unique identifier was created and missing values were recoded as NAs. Moreover, some observations in the database were not specific numbers; instead, it was indicated that for that year, prevalence was below a certain threshold (0.01%). In those cases, these observations were replaced by 0.009. The final step in the cleaning of the UNAIDS database consisted of deleting missing values for the dependent variable and deleting the regions with an iso2c equal to a country's iso2c (NA and ZA) to avoid problems in the merging process.

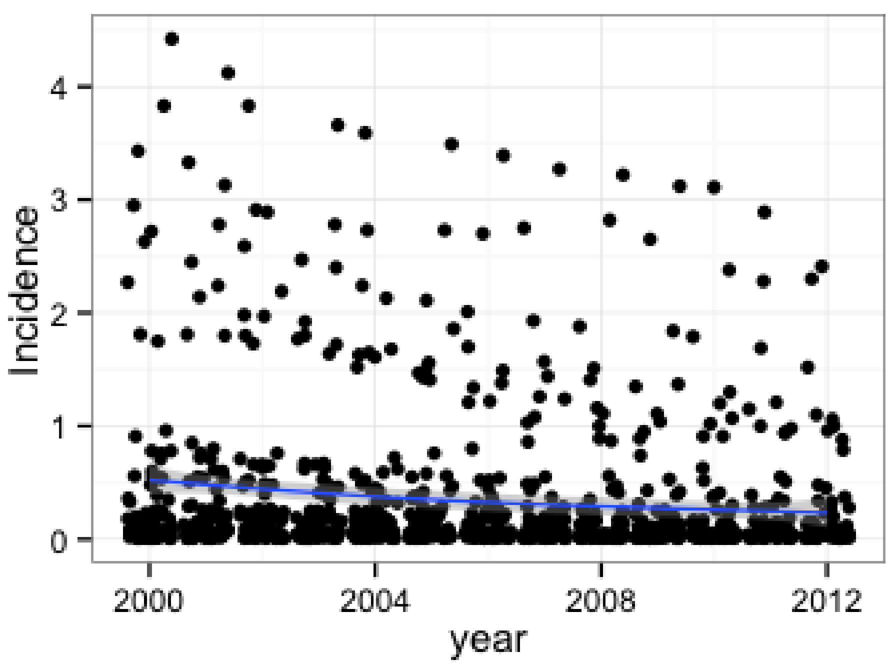
Once both databases were cleaned, the next step was to merge the datasets using the combination of iso2c and year as unique identifier. In the merging process, only observations that were present in both datasets were kept. It is worth noticing that UNAIDS' dataset included observations from 1990 to 2012 and WDI data covers the 2000-212 period so all observation from the UNAIDS dataset between 1990 and 1999 were dropped. Finally, unnecessary columns from the new database were eliminated.

7 Descriptive Statistics

The descriptive statistics part consists of analysing the main variables of interest and preparing them for the data analysis. Tables, plots and histograms are shown to understand the distribution of the variables.

Figure 2 shows that in most countries of our dataset HIV/AIDS incidence rates decreased between the period of 2000 to 2015 (see Figure 2). The blue line of Figure 2 is however only representing the general downwards sloping trend of HIV/AIDS incidence over time in our sample. The black dots in the above the blue line show that there are still some outliers to this trend.

Figure 2: HIV/AIDS Incidence Rate over Time

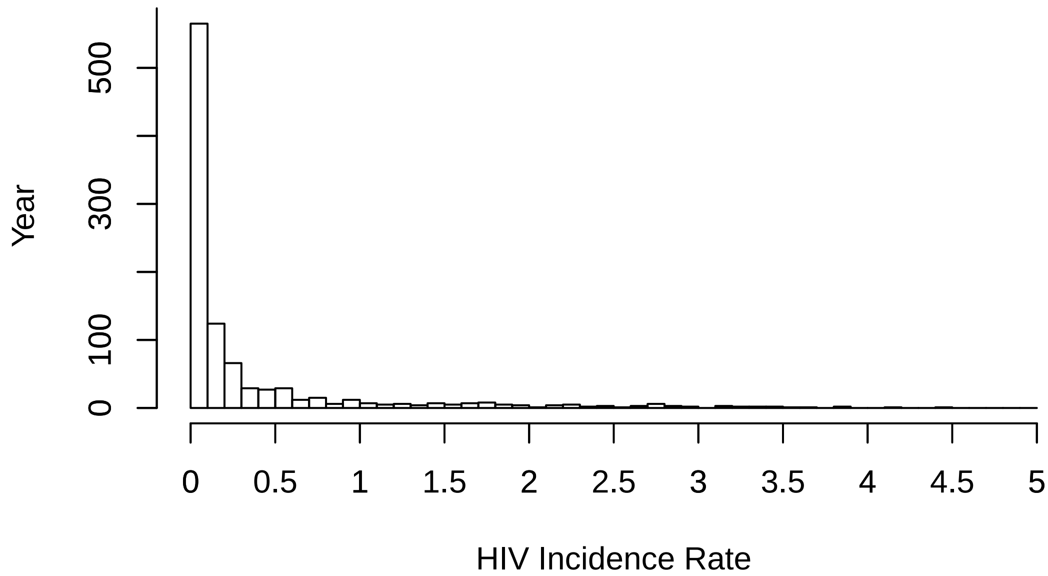


From Figures 3 and 4 (see Appendix) it can be seen which countries are not following this general downward sloping trend. Figure 3 plots the range of observations on HIV/AIDS incidence per country. The figure shows that most countries have low HIV/AIDS incidence rates slightly above 0, but some outliers (countries with high incidence rates) can be identified.

Figure 4 is showing the direction of the change of the HIV/AIDS incidence rate compared to the previous year by country. To analyse the change of the incidence rate from one year to the following year the incidence variable was lagged by one year and a new variable, calculating the difference between the lagged and the original incidence variable was created.

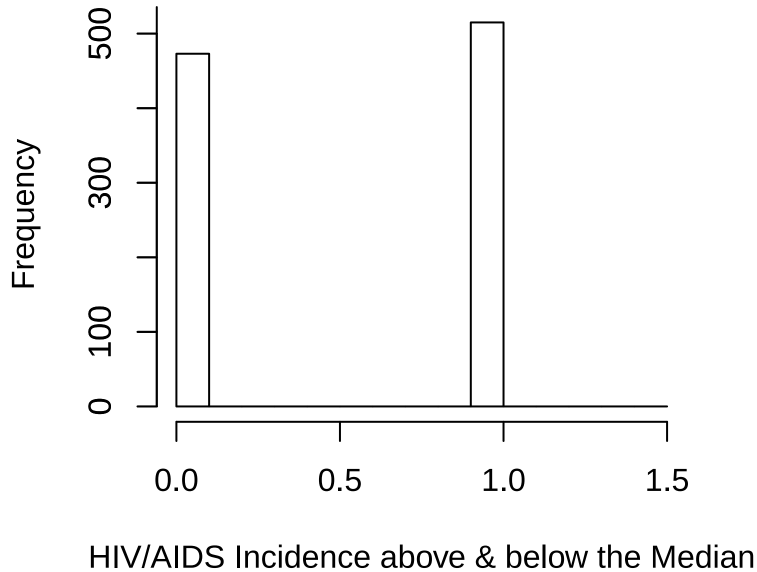
When looking at the distribution of the HIV/AIDS incidence variable in a histogram it becomes apparent that the incidence variable is indeed highly skewed to the left and only few incidence rates are higher than 1.

Figure 5: Distribution of HIV/AIDS Incidence Rate



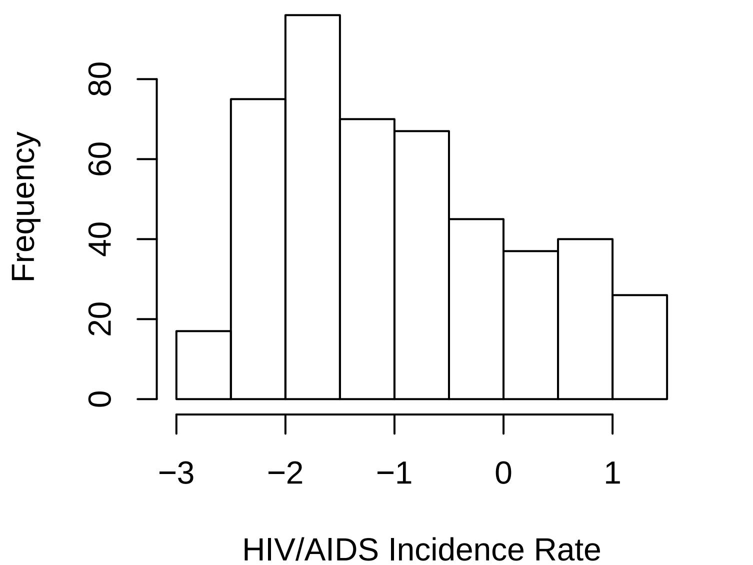
In order to deal with the skewness of this variable, a dummy variable, differentiating between countries with a high HIV incidence and countries with a low HIV incidence is created for the first empirical model. The most accurate measure of the central tendency of a skewed distribution is the median. Therefore, the dependent variable was coded as 1 if the HIV incidence rate is below the median and a value of 0 was assigned to countries with an HIV incidence rate above the median.

Figure 6: Dummy Variable for High and Low HIV/AIDS Incidence



For the second empirical model the sample was reduced and focuses only on countries with a high HIV/AIDS incidence in order to explore, whether the determinants identified in Model 1 hold true when zooming in on the more problematic cases. As the dependent variable was still skewed to the left after restricting the sample to those countries above the mean a log transformation was necessary to approach a normal distribution.

Figure 6: Logged HIV/AIDS Incidence Rate for Countries lying above the Median



Scatterplots were created for all independent variables by the categories of Dahlgrehn's model in order to see whether further relevant variables are skewed (see Appendix). Due to a high skewness in most of the variables and to ensure a better comparability of the variable units all but one of the independent variables in the sample were logged.

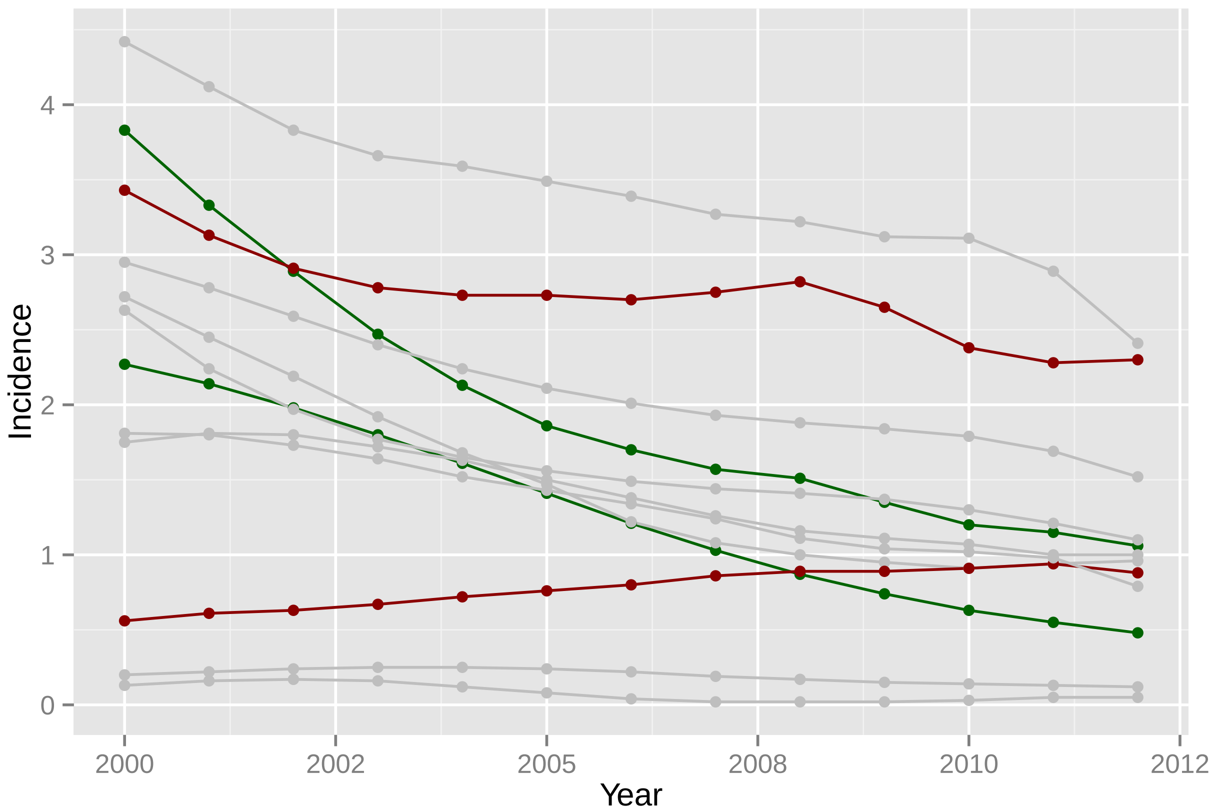
The only variable that was not logged is a variable comparing the share of female unemployment to total unemployment. The distribution of this variable comes already close to a normal distribution, so a log transformation is not urgently necessary. Further, it has some values of zero, which would make it difficult to log the variable.

8 Case Studies - Botswana, Lesotho, Uganda and Malawi

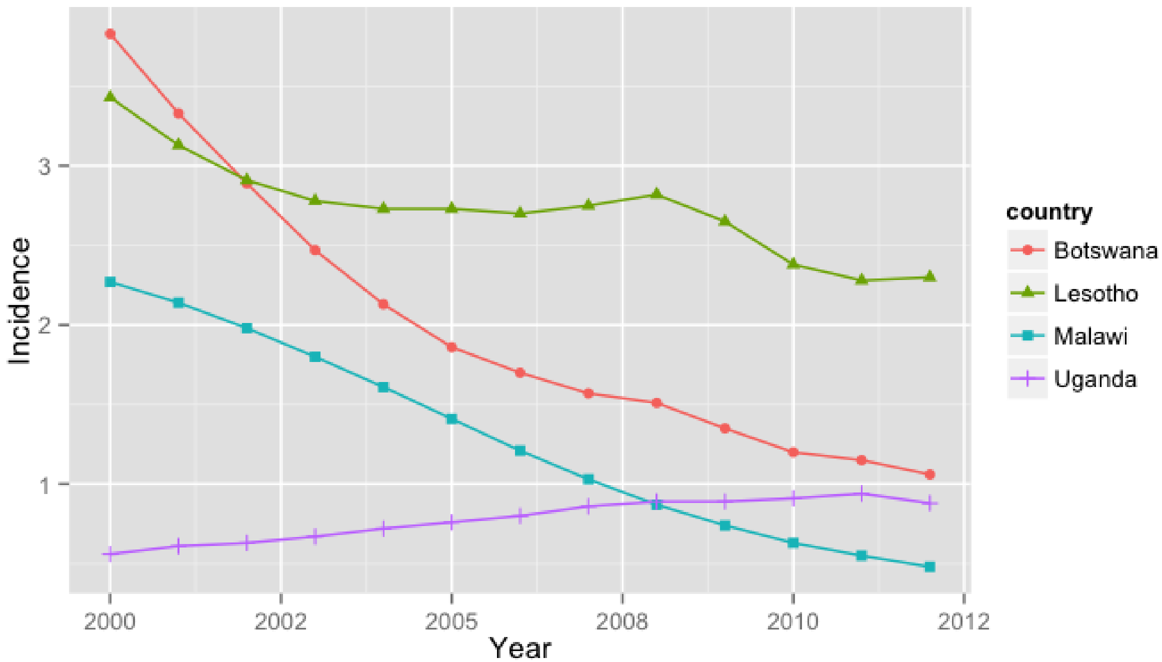
In the following, four case studies are presented to develop a better understanding of best practices and bad practices among countries with high changes in Incidence rates between 2000 and 2012. The cases were selected by the "extreme case selection method" on the basis of peculiar positive or negative changes in incidence rates over time (see Figure 4 in Appendix) (Gerring 2008).

Figure 8 shows the two best practices (green colour), namely Botswana and Malawi and the two bad practices (red colour), Uganda and Lesotho that were selected. In Figure 9 it

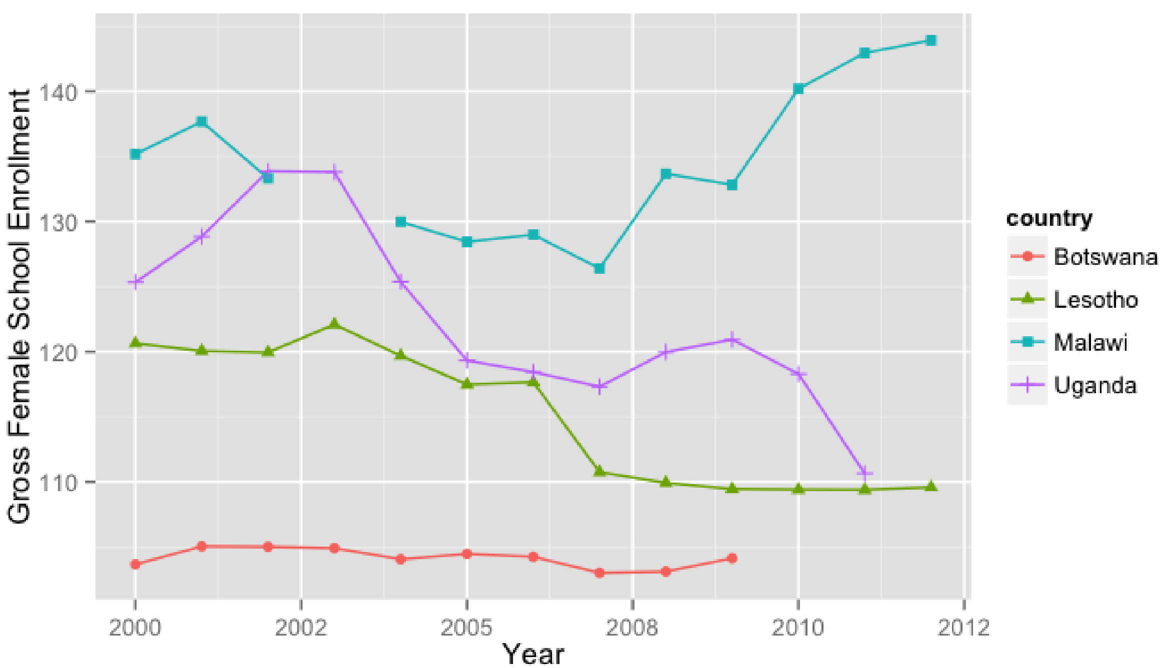
Figure 8: Interesting Cases for Change in HIV Incidence Rates



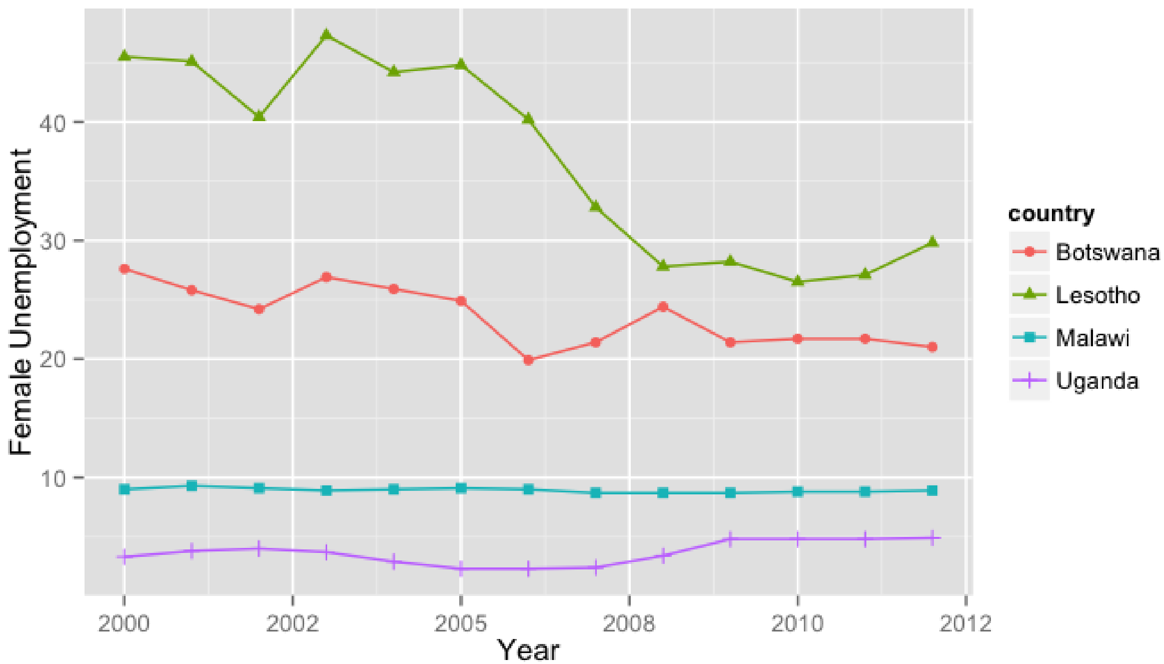
8.1 Case Studies



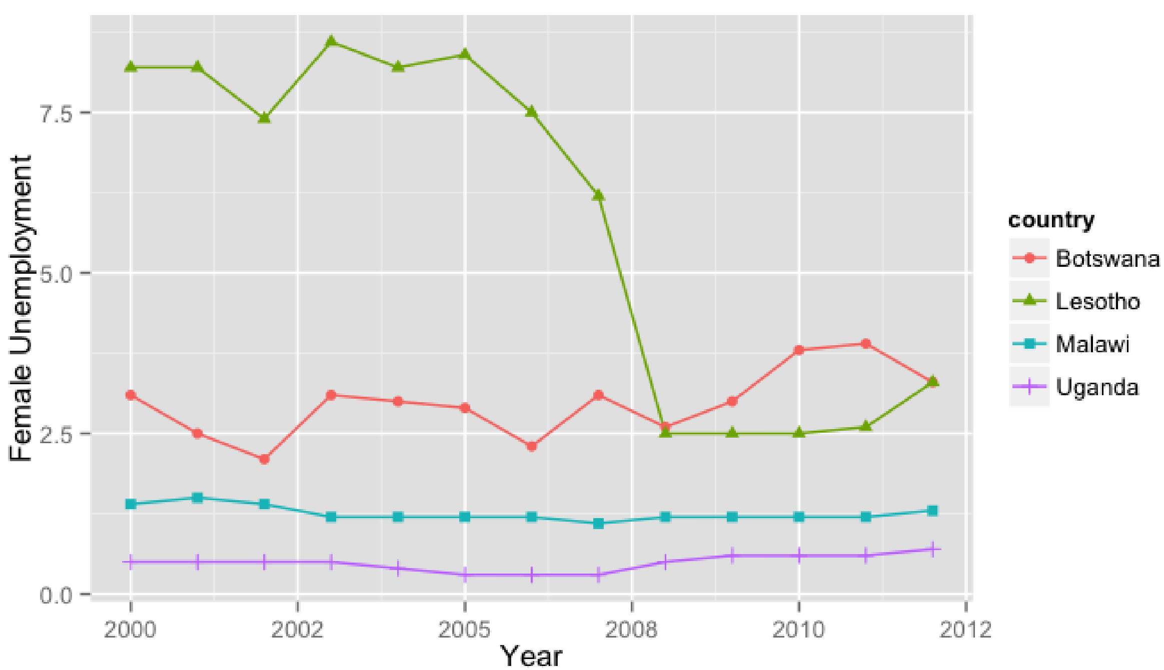
8.2 Female Schooling in Selected Countries



8.3 Female Unemployment in Selected Countries



8.4 Female Unemployment Share in Selected Countries



9 Inferential Statistics

9.1 Data Imputation

The data loaded from the Worldbank contained a lot of missing data because some countries do not report information on specific variables on a yearly basis. A high fraction of missing values produces a variety of problems and limitations to the data analysis. The paper imputed the missing data by using the "Amelia" package to solve this problem. "Amelia" generates by default five multiple, complete datasets that contain estimations of missing data points.

Before the data imputation highly collinear variables had to be dropped. The level of the Variance Inflation Factors

As can be seen from the Scatterplots (see Appendix) most of the variables are not normally distributed. Further, the variables all have different scales. Therefore, the independent variables were logged for enabling comparisons in the data analysis part.

9.2 Model 1 - Comparison of countries with an HIV Incidence Rate below and above the Median

For Model 1 logistic regressions are used for predicting the likelihood that a country has a low HIV incidence rate (the dependent variable Y is equal to 1, rather than 0) given certain values of the HIV determinants (values of X) being hypothetically assumed.

As the incidence variable is highly skewed to the left, the most accurate measure of the central tendency of the distribution is the median. Therefore, the dependent variable is coded as a dummy variable, being 1 if the HIV incidence rate is below the median and 0 for countries with an HIV incidence rate above the median.

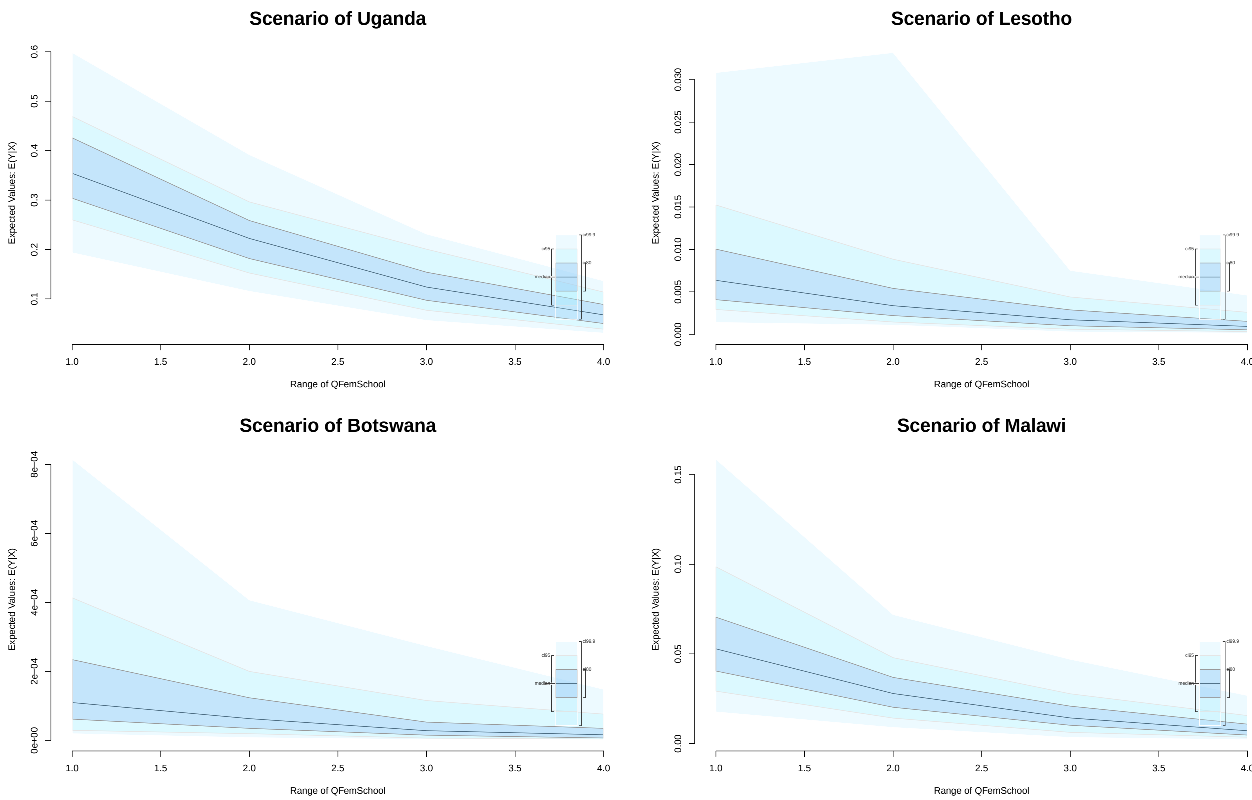
As Odds and Odds ratios are difficult to interpret and to present to a broad audience predicted probabilities are calculated after running the logistic regressions. The interpretation of the logistic regression results presented in table 1 will focus only on the significance of the variables.

A more detailed analysis of the predicted probabilities will be given for the main variables of interest Female School Enrollment and Share of Female Unemployment. Predicted probability scenarios are calculated for the four case studies that were introduced in the descriptive statistics part.

Table 1: Logistic Regression Results of Model 1

Variables	Coefficients	Std. Error	T-Statistic	P-Value
Constant	-105.86	10.20	-10.38	0.00
GDP per capita	-0.84	0.38	-2.23	0.03
Share of Rural Population	-1.12	0.42	-2.69	0.01
Share of CO2	-1.05	0.22	-4.66	0.00
Healthcare Expenditure	0.38	0.37	1.03	0.31
Access to Water	0.82	0.73	1.13	0.26
Access to Sanitation	0.16	0.25	0.63	0.53
Life Expectancy	31.14	2.55	12.19	0.00
Immunisation against DPT	-1.21	1.21	-1.00	0.32
Immunisation against Measles	1.69	1.27	1.32	0.19
Female School Enrollment	-4.09	0.53	-7.69	0.00
Share of Female Unemployment	-0.02	0.04	-0.56	0.58

Predicted Probabilities for Female School Enrollment



9.3 Model 2 - Focusing on Countries with an HIV Incidence Rate above the Median

Table 2: OLS Regression Results of Model 2

Variables	Coefficients	Std. Error	T-Statistic	P-Value
Constant	16.47	1.73	9.53	0.00
GDP per capita	0.16	0.09	1.82	0.08
Share of Rural Population	0.55	0.13	4.39	0.00
Share of CO2	0.13	0.05	2.56	0.02
Healthcare Expenditure	-0.10	0.10	-0.97	0.33
Access to Water	0.20	0.19	1.06	0.29
Access to Sanitation	0.01	0.06	0.09	0.93
Life Expectancy	-7.18	0.31	-23.06	0.00
Immunisation against DPT	0.26	0.27	0.95	0.34
Immunisation against Measles	-0.13	0.30	-0.42	0.68
Female School Enrollment	1.41	0.15	9.58	0.00
Share of Female Unemployment	0.13	0.02	7.12	0.00

Table 3: OLS Regression Results of Model 2 with robust standard errors

Variables	Coefficients	Std. Error	T-Statistic	P-Value
Constant	16.47	1.73	9.53	0.00
GDP per capita	0.16	0.09	1.82	0.08

Variables	Coefficients	Std. Error	T-Statistic	P-Value
Share of Rural Population	0.55	0.13	4.39	0.00
Share of CO2	0.13	0.05	2.56	0.02
Healthcare Expenditure	-0.10	0.10	-0.97	0.33
Access to Water	0.20	0.19	1.06	0.29
Access to Sanitation	0.01	0.06	0.09	0.93
Life Expectancy	-7.18	0.31	-23.06	0.00
Immunisation against DPT	0.26	0.27	0.95	0.34
Immunisation against Measles	-0.13	0.30	-0.42	0.68
Female School Enrollment	1.41	0.15	9.58	0.00
Share of Female Unemployment	0.13	0.02	7.12	0.00

Table 4: Fixed Effects Regression Results of Model 2

	Value	Std. Error	t-stat	p-value
(Intercept)	-2.7300	2.5149	-1.0855	0.2778
lGDPpc	-0.0203	0.1130	-0.1796	0.8576
lRural	3.1645	0.3597	8.7974	0.0000
lCO2	0.0483	0.0429	1.1253	0.2788
lHCexpend	-0.1592	0.0774	-2.0577	0.0411
lWater	-1.1317	0.3747	-3.0199	0.0040
lSanitation	-0.4475	0.2524	-1.7727	0.0781
lLifeExpect	-1.1275	0.3887	-2.9004	0.0043
lDPT	0.2734	0.1232	2.2197	0.0265
lMeasles	-0.4102	0.1315	-3.1186	0.0019
lFemSchool	-0.0409	0.0930	-0.4395	0.6609
ShFemUnempl	-0.0051	0.0153	-0.3300	0.7414
as.factor(country)Benin	-0.7993	0.4603	-1.7366	0.0848
as.factor(country)Botswana	3.8772	0.2709	14.3146	0.0000
as.factor(country)Burkina Faso	-1.9082	0.3952	-4.8284	0.0000
as.factor(country)Burundi	-0.7524	0.3303	-2.2776	0.0229
as.factor(country)Cambodia	-2.1633	0.3159	-6.8480	0.0000
as.factor(country)Cameroon	1.8094	0.1944	9.3081	0.0000
as.factor(country)Central African Republic	0.3542	0.3462	1.0230	0.3076
as.factor(country)Chad	-1.2226	0.3680	-3.3219	0.0010
as.factor(country)Congo, Dem. Rep.	-0.9227	0.2573	-3.5865	0.0003
as.factor(country)Congo, Rep.	1.4542	0.4873	2.9841	0.0033
as.factor(country)Cote d'Ivoire	0.6448	0.3799	1.6974	0.0935
as.factor(country)Dominican Republic	2.0847	0.2445	8.5252	0.0000
as.factor(country)El Salvador	1.2550	0.3037	4.1320	0.0000
as.factor(country)Eritrea	-2.0192	0.3929	-5.1395	0.0000
as.factor(country)Ethiopia	-2.4473	0.3581	-6.8344	0.0000
as.factor(country)Gabon	5.4720	0.5502	9.9464	0.0000
as.factor(country)Gambia, The	1.3601	0.2558	5.3171	0.0000
as.factor(country)Ghana	0.1653	0.4728	0.3496	0.7272
as.factor(country)Guatemala	0.6234	0.1842	3.3844	0.0007
as.factor(country)Guinea-Bissau	0.7392	0.3659	2.0201	0.0447
as.factor(country)Haiti	0.3572	0.2956	1.2087	0.2276
as.factor(country)Honduras	0.4446	0.2429	1.8303	0.0673
as.factor(country)Jamaica	1.7660	0.2004	8.8110	0.0000
as.factor(country)Kenya	0.2510	0.2213	1.1342	0.2570
as.factor(country)Lesotho	2.1828	0.3445	6.3361	0.0000

	Value	Std. Error	t-stat	p-value
as.factor(country)Liberia	-0.0382	0.4446	-0.0859	0.9316
as.factor(country)Madagascar	-2.2479	0.3733	-6.0214	0.0000
as.factor(country)Malawi	0.4923	0.5099	0.9654	0.3363
as.factor(country)Mali	-1.2837	0.2701	-4.7531	0.0000
as.factor(country)Mauritius	1.1855	0.2279	5.2031	0.0000
as.factor(country)Mozambique	0.8860	0.3003	2.9506	0.0032
as.factor(country)Myanmar	-0.7971	0.2941	-2.7106	0.0067
as.factor(country)Namibia	2.3479	0.3017	7.7822	0.0000
as.factor(country)Nepal	-1.7712	0.3718	-4.7642	0.0000
as.factor(country)Niger	-2.6991	0.4578	-5.8964	0.0000
as.factor(country)Nigeria	0.5441	0.2181	2.4951	0.0142
as.factor(country)Papua New Guinea	-2.3027	0.2741	-8.3995	0.0000
as.factor(country)Rwanda	-0.2369	0.2642	-0.8969	0.3698
as.factor(country)Senegal	-0.5047	0.2134	-2.3656	0.0182
as.factor(country)Sierra Leone	-0.7559	0.4100	-1.8438	0.0672
as.factor(country)Somalia	-2.5336	0.3034	-8.3510	0.0000
as.factor(country)South Africa	4.4981	0.2584	17.4048	0.0000
as.factor(country)Swaziland	2.2975	0.1722	13.3423	0.0000
as.factor(country)Tanzania	-0.1928	0.4101	-0.4701	0.6386
as.factor(country)Thailand	-0.1477	0.2728	-0.5416	0.5881
as.factor(country)Togo	-0.4751	0.4150	-1.1449	0.2536
as.factor(country)Trinidad and Tobago	-0.4794	0.3178	-1.5085	0.1317
as.factor(country)Uganda	0.5005	0.2795	1.7908	0.0740
as.factor(country)Ukraine	2.6185	0.2686	9.7491	0.0000
as.factor(country)Vietnam	-0.7631	0.2862	-2.6659	0.0077
as.factor(country)Zambia	1.8621	0.1670	11.1532	0.0000
as.factor(country)Zimbabwe	2.1614	0.2794	7.7357	0.0000

The test for variance inflation factors showed that in our first logistic regression model six variables showed high multicollinearity and had a higher variance inflation than the threshold of 10. We tested the multicollinearity between the variables and found that there was high multicollinearity between the GDP and GDP per capital, Unemployment and Female unemployment, Primary education and female schooling. Therefore, we excluded one of these multicollinear variables for each group based on their explanatory strength for our research question, namely unemployment, primary education and GDP.

---

10 Conclusions

(Sachs and Malaney 2002) (Dupas 2009)

The IncidenceDif variable would actually be the most accurate variable in terms of measuring whether Target 6.A of the MDGs ("have halted by 2015 and begun to reverse the spread of HIV/AIDS" (United Nations 2014) was fulfilled until 2012. A dummy variable on whether MDG6a was fulfilled or not is however not very suitable for identifying the specific determinants of HIV/AIDS at country level as it is not differentiating between the relative size of the effect (an increase from 0,001 to 0,002 would be as bad as a change from 0,1 to 0,3).

Model 2 - variable is not perfect normal distribution - better beta regression

---

11 Limitations

The paper had to make some compromises regarding its original aim as outlined in the first research proposal. Due to the significant amount of missing values and the presence of multicollinearity, a considerable number of variables had to be dropped and could ultimately not be integrated in the logistic regression models.

The selection of these variables was not arbitrary but followed instead the theoretical framework guiding this research, i.e. Dahlgren's model. Two levels of Dahlgren's model (Social and Community Networks and Individual Lifestyle Factors) ended up underrepresented after dropping these variables. To deal with this limitation, the research will only use the theoretical framework as an instrument to guide the selection of variables but will not utilise the findings to test the validity of the model.

In terms of the data used to run the regressions, the relative high number of countries that have already halted or reversed the spread of HIV/AIDS in our sample can lead to biased results. In the next stage of the research, the effect of excluding those countries that only halted the spread will be explored.

Another shortcoming faced at this stage was the integration of figures from the descriptive statistics into the final report. A transitory solution was to save those pictures in a subfolder of the repository.

---

12 Appendix

Descriptive Statistics

Table 5: Descriptive statistics

Statistic	N	Mean	St. Dev.	Min	Max
X	988	494.5	285.4	1	988
year	988	2,006.0	3.7	2,000	2,012
GDPpc	945	5,296.0	5,153.0	441.2	30,875.0
Rural	988	57.7	19.5	8.7	91.8
CO2	829	1.7	3.5	0.004	38.2
Water	982	75.5	17.9	23.5	99.8
Sanitation	983	50.4	30.1	6.6	98.9
Unemploym	988	8.7	6.2	0.6	38.7
HCexpendpc	962	124.3	165.4	2.4	1,103.0
FemSchool	828	98.1	20.8	20.8	162.4
LifeExpect	988	61.4	9.9	38.1	79.6
DPT	988	79.9	18.0	19	99
Measles	988	79.4	17.5	16	99
Population	988	40,890,589.0	133,140,255.0	1,063,715	1,236,686,732
Incidence	988	0.3	0.7	0.01	4.4
lGDPpc	945	8.1	1.0	6.1	10.3
lRural	988	4.0	0.4	2.2	4.5
lCO2	829	−0.6	1.6	−5.6	3.6
lHCexpend	962	1.7	0.4	0.6	2.9
lWater	982	4.3	0.3	3.2	4.6
lSanitation	983	3.7	0.8	1.9	4.6
lUnemploym	988	1.9	0.7	−0.5	3.7
lHCexpendpc	962	4.1	1.2	0.9	7.0
lFemSchool	828	4.6	0.3	3.0	5.1
lLifeExpect	988	4.1	0.2	3.6	4.4
lDPT	988	4.3	0.3	2.9	4.6
lMeasles	988	4.3	0.3	2.8	4.6
lIncidence	988	−2.4	1.6	−4.7	1.5
QFemSchool	828	2.5	1.1	1	4
ShFemUnempl	988	1.5	2.9	−3.4	23.9
Dummy	988	0.5	0.5	0	1

Figure 3:HIV Incidence Rates per Country over Time

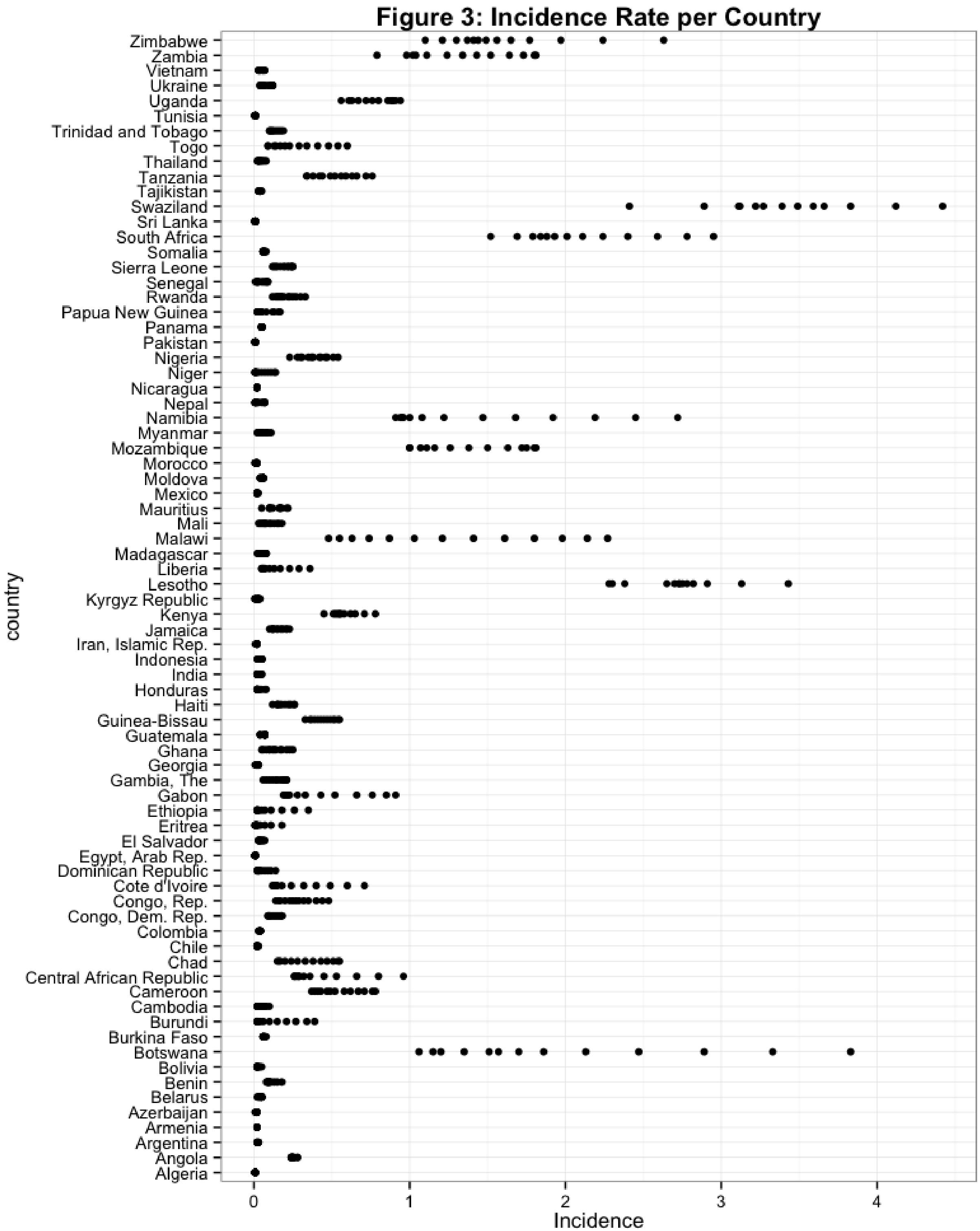


Figure 4:Change in HIV Incidence Rate compared to Previous Years per Country

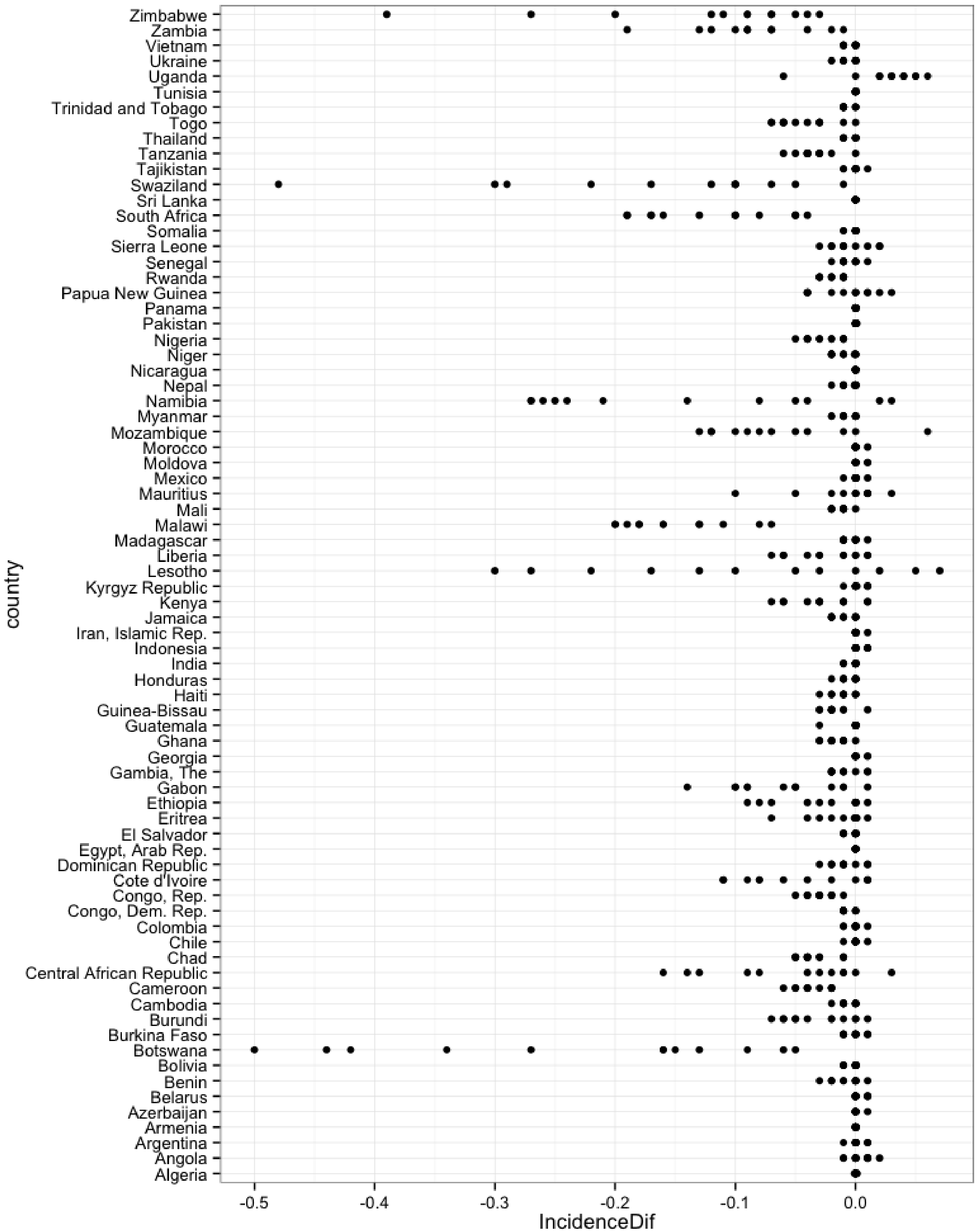


Figure 5:Scatterplot of variables for socio-economic, cultural and environmental conditions

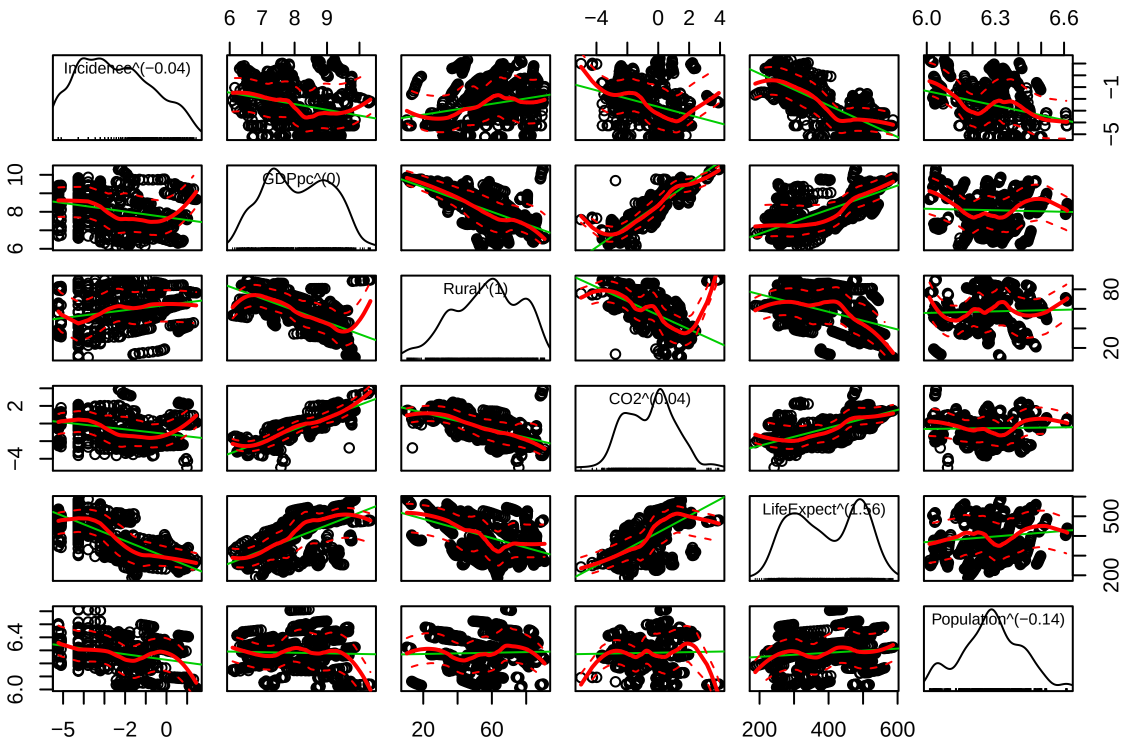


Figure 6:Scatterplot of variables for living and working conditions

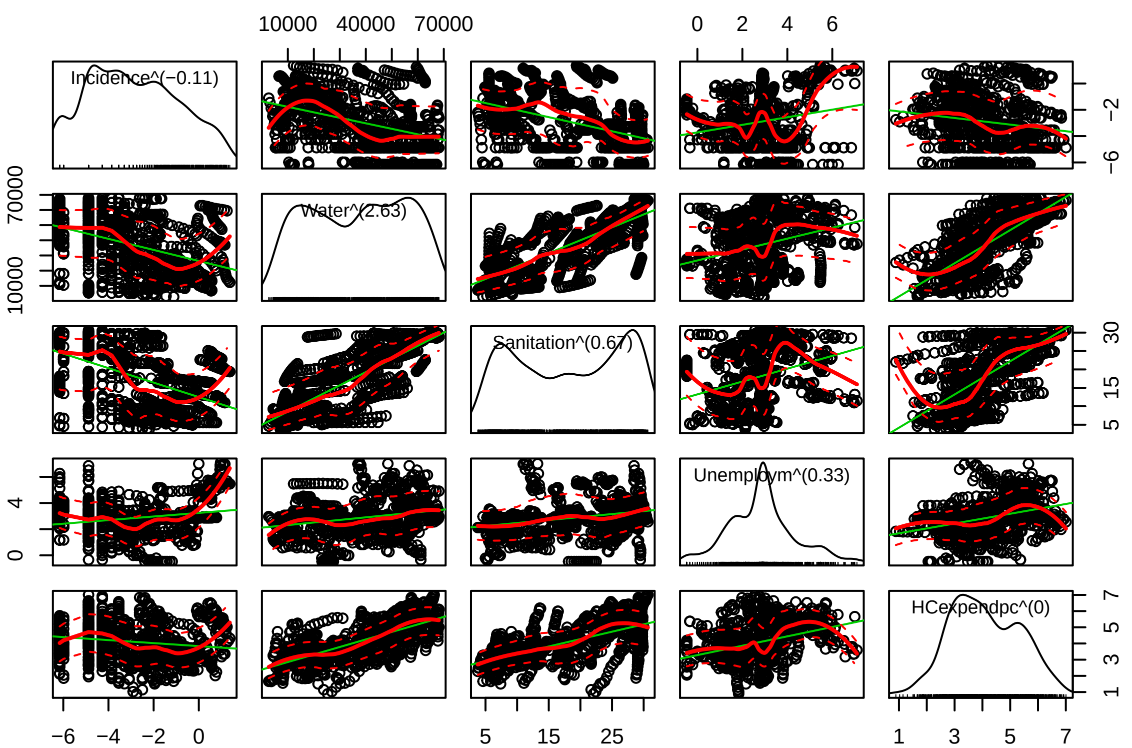
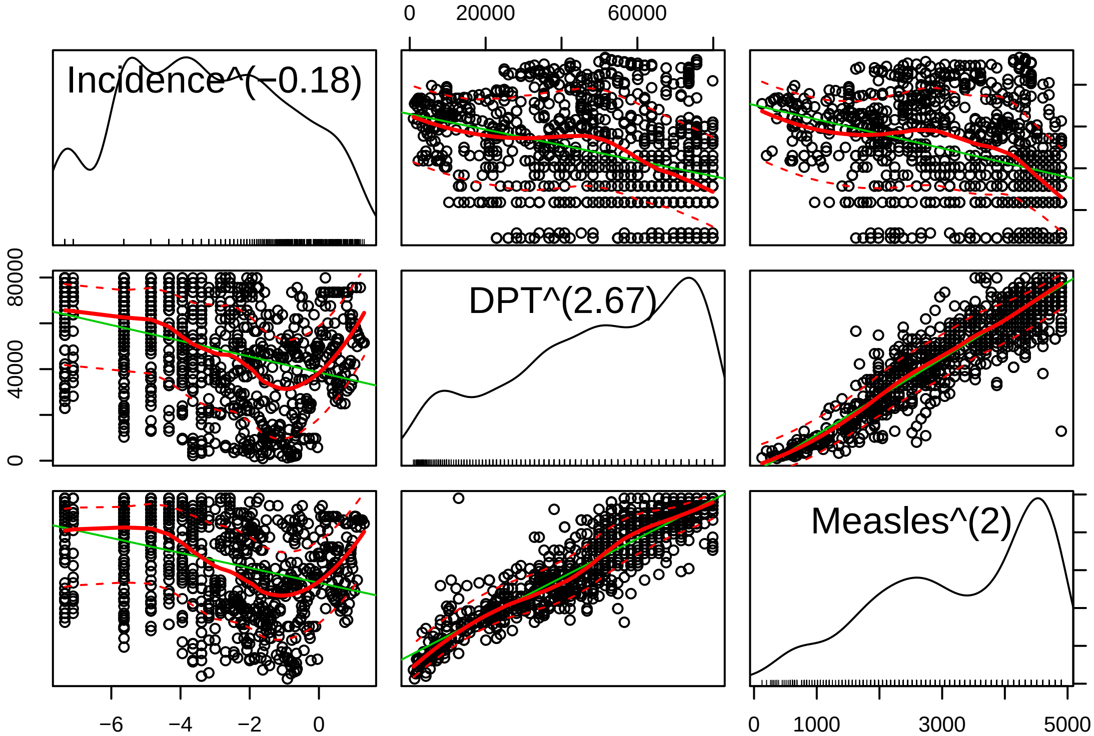
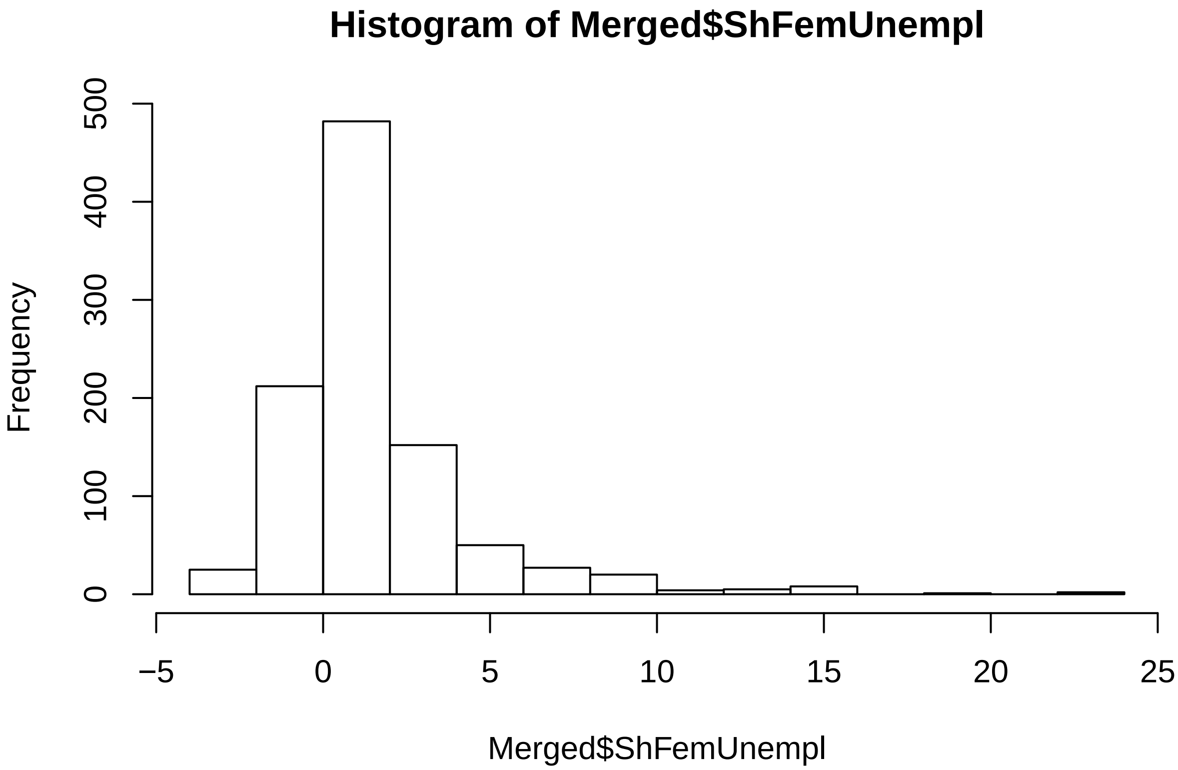


Figure 7:Scatterplot of variables for individual lifestyle factors



Histogram of Female Unemployment compared to Total Unemployment (not logged)



Testing for Multicollinearity of the Variables

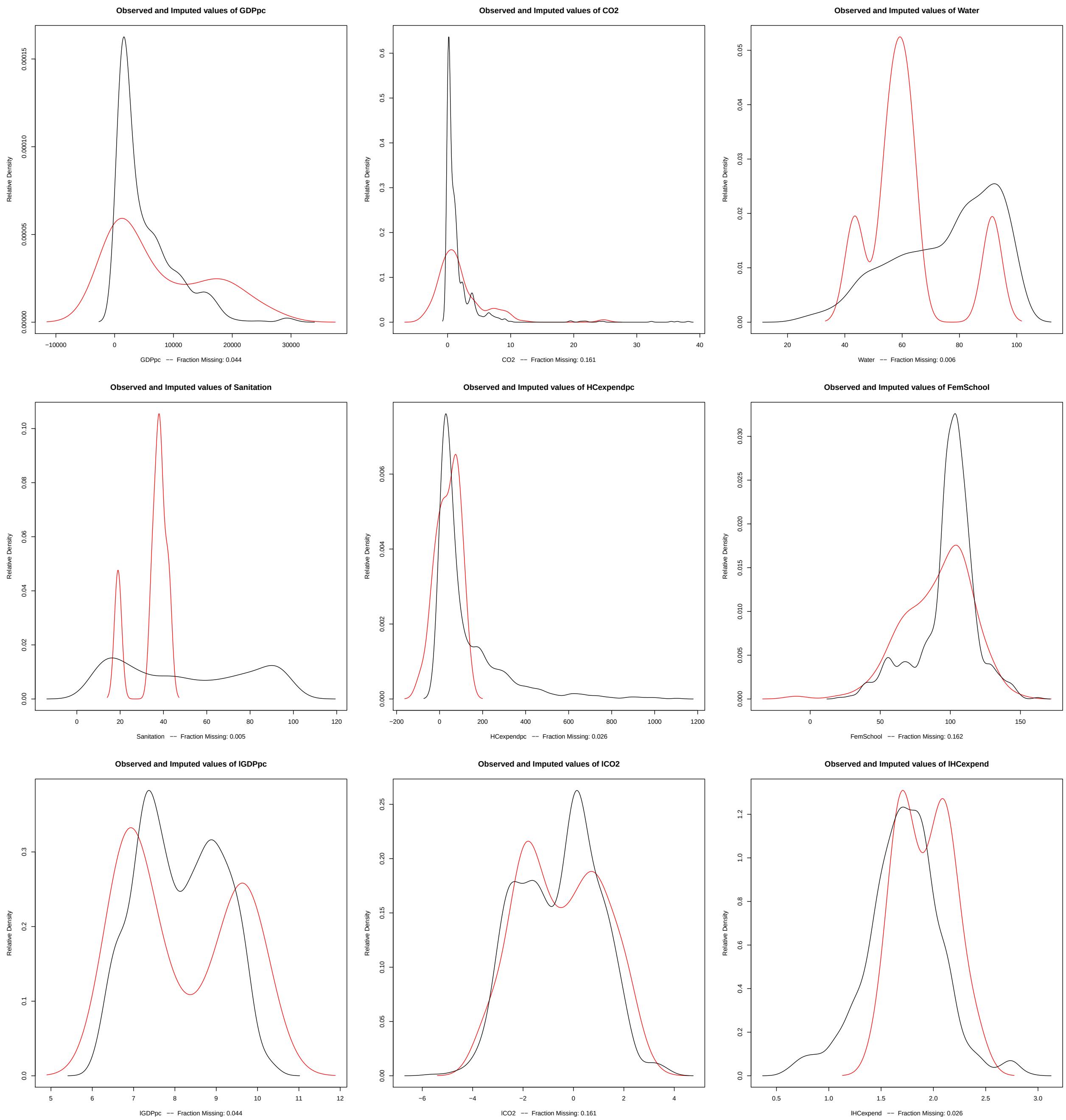
Table 6: Variance Inflation Factors - Table 1

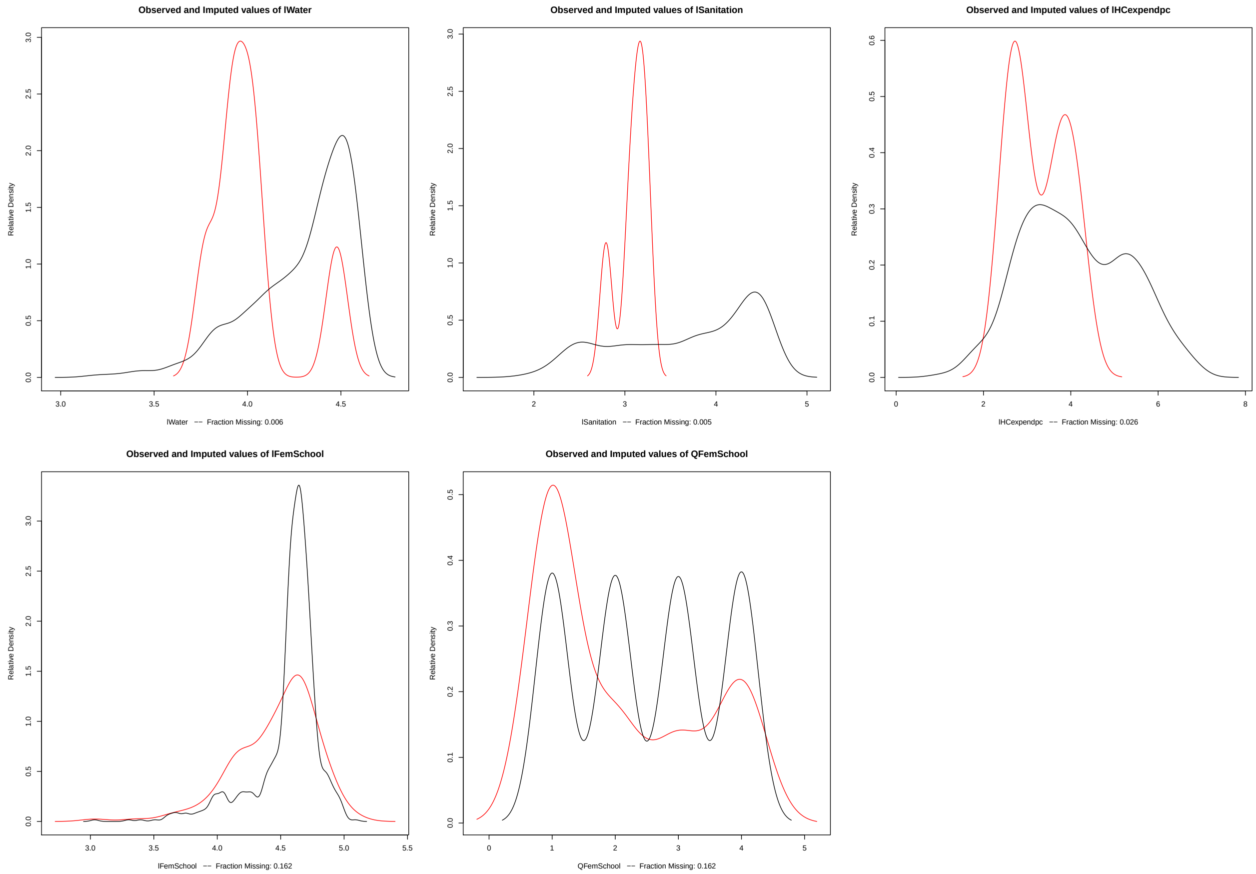
	vif
GDP	3.09
GDPpc	11.78
Rural	2.49
CO2	3.20
HCexpend	1.61
Primary	47.65
Water	3.88
Sanitation	4.41
Unemploym	1.98
HCexpendpc	7.00
ShFemUnempl	1.74
FemSchool	53.08
LifeExpect	3.54
DPT	8.82
Measles	9.17
Population	2.93

Table 7: Variance Inflation Factors - Table 2

	vif
GDPpc	9.28
Rural	2.34
CO2	3.13
Water	3.72
Sanitation	4.28
HCexpendpc	4.92
ShFemUnempl	1.67
Unemploym	1.77
FemSchool	1.33
LifeExpect	3.32
DPT	8.57
Measles	8.95
Population	1.16

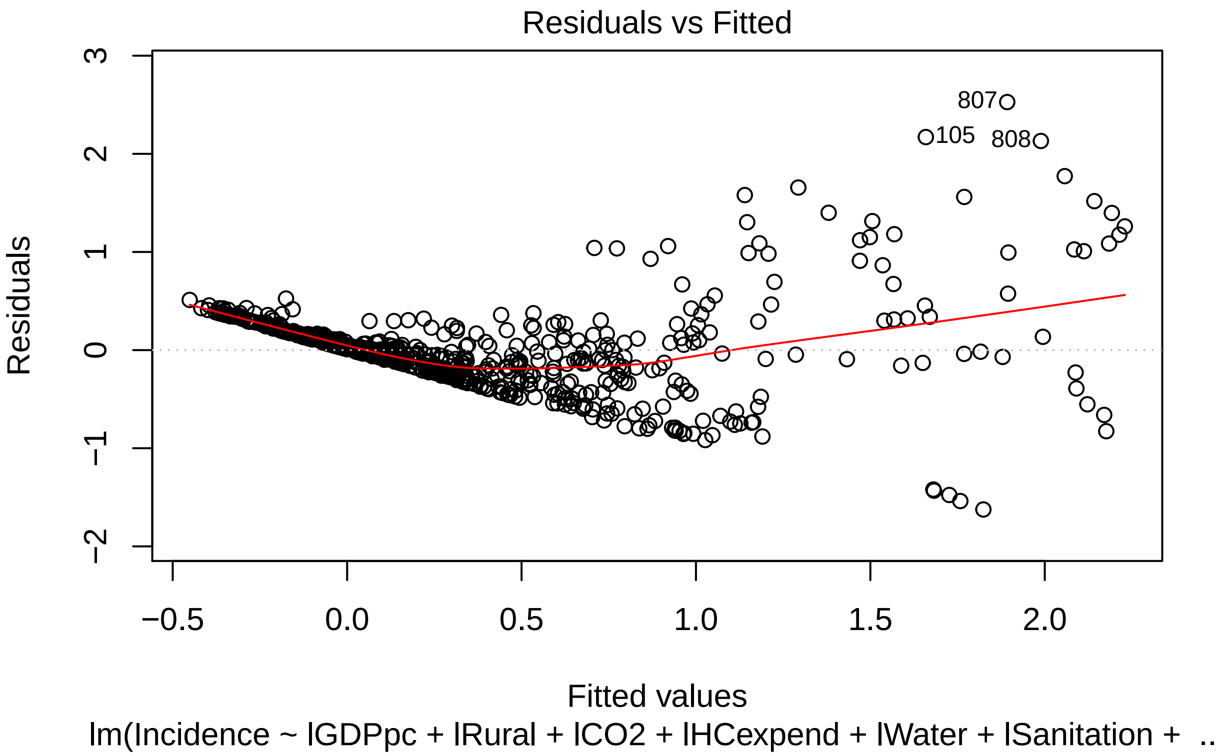
Data Imputation Matrix



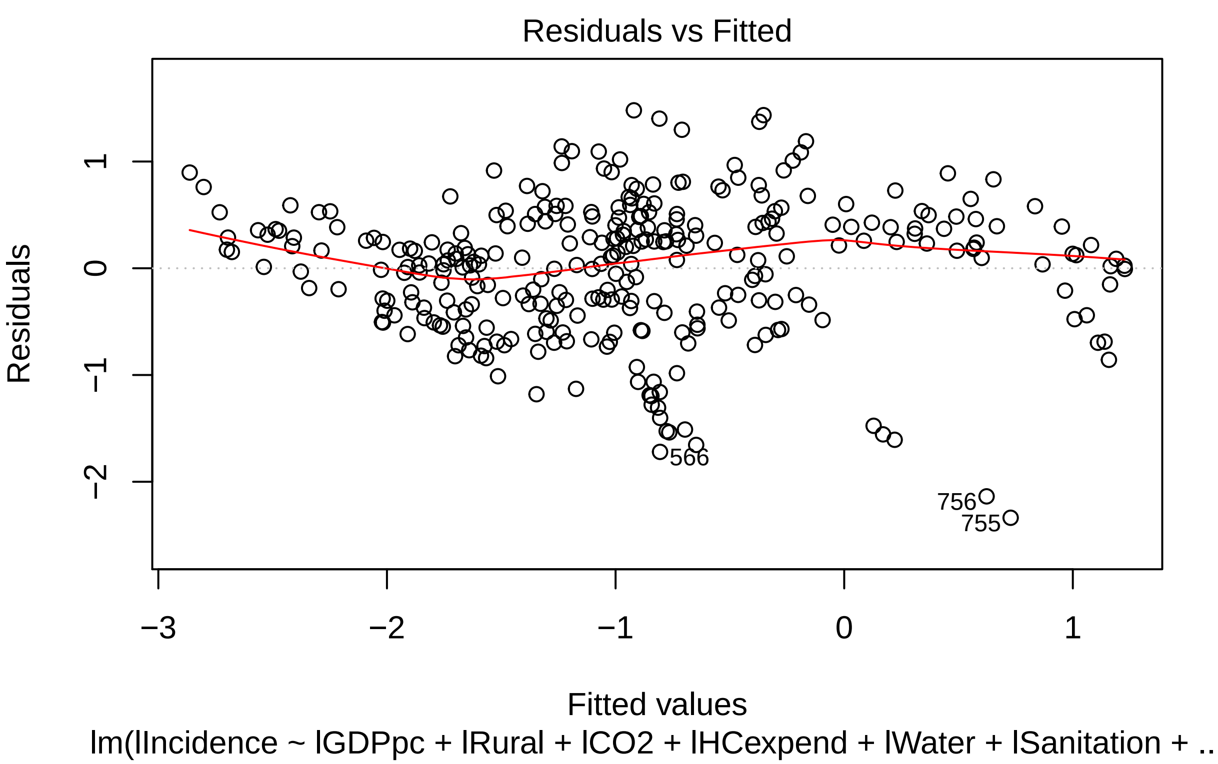


Observed and Imputed values of lWater
Relative Density
lWater -- Fraction Missing: 0.006
Observed and Imputed values of lSanitation
Relative Density
lSanitation -- Fraction Missing: 0.005
Observed and Imputed values of lHCexpendpc
Relative Density
lHCexpendpc -- Fraction Missing: 0.026
Observed and Imputed values of lFemSchool
Relative Density
lFemSchool -- Fraction Missing: 0.162
Observed and Imputed values of QFemSchool
Relative Density
QFemSchool -- Fraction Missing: 0.162

Testing for Heteroscedasticity - Model 1



Testing for Heteroscedasticity - Model 2



Dahlgren, Göran, and Margaret Whitehead. 1991. “Policies and Strategies to Promote Social Equity in Health.” *Institute for Futures Studies*.

Dupas, Pascaline. 2009. *Do Teenagers Respond to HIV Risk Information? Evidence from a Field Experiment in Kenya*. National Bureau of Economic Research.

Gerring, John. 2008. “Case Selection for Case-Study Analysis: Qualitative and Quantitative Techniques.” *J. Box-Steffensmeier, HE Brady, & D. Collier (Reds.), Oxford Handbook of Political Methodology, Oxford University Press, Oxford*, 645–84.

Haacker, Markus. 2002. “The Economic Consequences of HIV/AIDS in Southern Africa.” *IMF Working Paper* 02 (38).

Hurrelmann, Klaus. 1989. “Human Development and Health.” *Springer-Verlag*.

Kalemli-Ozcan, Sebnem. 2011. “AIDS, "Reversal” of the Demographic Transition and Economic Development: Evidence from Africa.” *Springer-Verlag*.

Sachs, Jeffrey, and Pia Malaney. 2002. “The Economic and Social Burden of Malaria.” *Nature* 415 (6872): 680–85.

Talbott, John R. 2007. “Size Matters: The Number of Prostitutes and the Global HIV/AIDS Pandemic.” *PLoS ONE*.

United Nations. 2014. “Millennium Development Goals.” http://www.un.org/millenniumgoals/aids.shtml.