Wealth of Nations: Analysis of Income Disparity and Relationship to Poverty

Author: Nirmal Labh [*HUID: 80625529*]

Contact TF: Kela Roberts

Abstract

It is widely understood that income disparity has continued to increase. While world average GDP more than doubled from 191 billion in 2003 to 402 billion in 2013, and the average share of income for the higher 10% went up from 33% to 36%, the income of the lower 10% hardly increased from 2.26% to 2.5%. Average world poverty stood at 19%.

To attain better social stability, it is vital to improve the income of the less fortunate ones reflected in the 10% lower income population as well as reduce the headcount of people living under the poverty lines.

Using simple and multiple linear regressions on income data for all countries between 1974 and 2013 obtained from World Bank[1], we investigate below three explanatory variables as predictors of poverty level:

- GDP
- Income level of lower and higher 10%
- Literacy

The *model* as well as the *three explanatory variables* have extremely *small p-value.*

The paper therefore concludes that GDP growth, coupled with increase in literacy has important implications for improving income of low earners as well as reducing the level of poverty.

I. Introduction

Every country strives to increase GDP, a reflection of increased economic activity, and hence of increased productivity, trade and employment among many other things. It has direct consequences in improving income levels and poverty reduction that we look at in this paper.

We start off by looking at below variables in the WDI data[1,2]:

- Quantitative variables:
  - Income of lower 10%
  - Income of higher 10%
  - Wealth i.e. Net Income
  - Poverty Level
  - Population Count
- All of the data has below two dimensions:
  - Country: *While widely accepted count of countries is 195, it lists 214, more so as distinct data sources*
  - Year: *The data is for years starting 1974 to 2013*

In this paper we attempt to evaluate and validate correlation between:

- income share of the lower 10% and the higher 10%
- poverty and income share of higher 10%
- poverty and income share of lower 10%.
- poverty and literacy

Vizualization of the wealth and income disparity as well as the above four listed correlations is being done in R using *ggplot2*[3] CRAN module.

II. Methods

The data used in the analysis are retrieved using the *WDI CRAN module*[2]. This module provides direct API access to the World Bank[1] data from within R code. All of the numbers and charts appearing in this paper have been prepared using R markdown. Complete .Rmd file for this paper is available at github. R code for calculations and charts are included in the *Appendix* section.

Data Description:

- WDI Indicators used to lookup specific data points for the study:

Indicator	Description
BN.GSR.FCTY.CD	Net income (BoP, current US$)
NY.GDP.MKTP.CD	GDP (current US$)
SE.ADT.1524.LT.ZS	Literacy rate, youth total (% of people ages 15-24)
SE.ADT.LITR.ZS	Literacy rate, adult total (% of people ages 15 and above)
SI.DST.10TH.10	Income share held by highest 10%
SI.DST.FRST.10	Income share held by lowest 10%
SI.POV.NAHC	Poverty headcount ratio at national poverty line (% of population)
SP.POP.TOTL	Population, total

Table 1: Subset of WDI Indicators

- Sample Data:

	Country	Yr	GDP	Poverty	IncmLo10	IncmHi10	Literacy
215	Albania	2008	12.88	12.4	3.73	24.46	97.38
252	Armenia	2001	2.118	48.3	3.19	29.1	99.6
263	Armenia	2012	9.958	32.4	3.72	24.76	99.68
330	Argentina	1991	189.7	NA	1.75	36.6	97.22
340	Argentina	2001	325.5	NA	0.65	39.46	98.05
782	Burkina Faso	2003	4.206	51.1	2.46	33.97	26.53

Table 2: Sample data based on Indicators from Table 1

We start off by examining the simple correlation between the below pairs:

- Income of lower 10% share vs income of higher 10% share (section III.1.)
- Poverty vs Higher 10% Income Share (section III.2)
- Poverty vs Lower 10% Of Income Share (section III.3)
- Povery and Literacy (section III.4)

Establishing there is correlation, we further look at their combinations in the form of Multiple Regression Model. With iteratibve refinement, we com up to the model in section III.5, verify the distributions of residuals to be normal in section III.6, and continue on to do additional verification using ANOVA in section III.7. We see very good *p-values* that are nearly zero for this model.

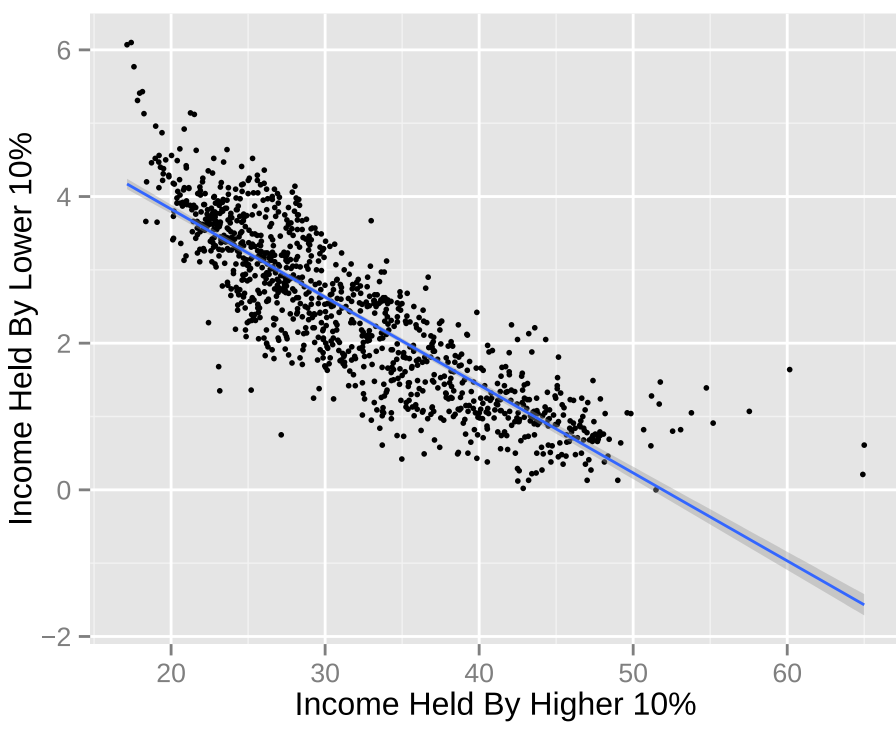
For completeness, one of the models that was rejected is listed in Appendix - section VI.12. It was rejected on the basis that most of the *p-values* associatd with the explanatory variables were rather big and hence not significant.

III. Results

1. Income Disparity

Based on the data, scatter plot with *income of lower 10% share on the y-axis* and *income of higher 10% share* on the x-axis shows *strong negative linear relationship* between the two variables.

It validates commonly held notion that *rich are getting richer and poor are getting poorer*.



We further validate this strong negative correlation, notice the nearly zero *p-value* and *negative co-efficient*:

```
##
## Call:
```

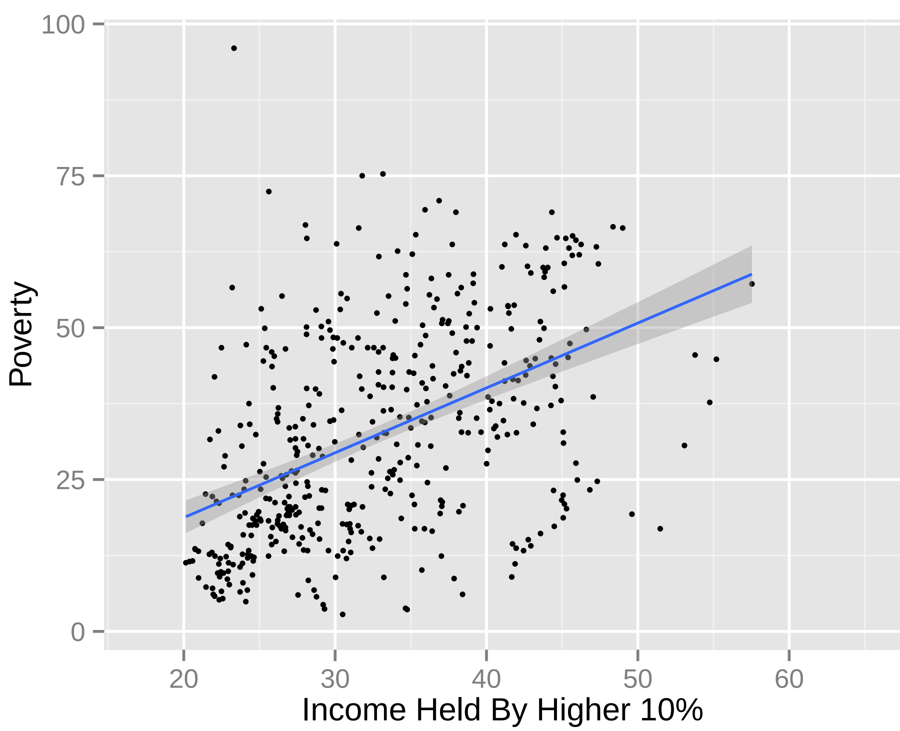
```
## lm(formula = income_low10 ~ income_high10, data = wdi_data)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -2.2208 -0.3710  0.0196  0.3511  2.6279
##
## Coefficients:
##               Estimate Std. Error t value            Pr(>|t|)
## (Intercept)    6.22675    0.07087    87.9 <0.0000000000000002 ***
## income_high10 -0.11992    0.00216   -55.6 <0.0000000000000002 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.582 on 1106 degrees of freedom
##   (8308 observations deleted due to missingness)
## Multiple R-squared:  0.736,  Adjusted R-squared:  0.736
## F-statistic: 3.09e+03 on 1 and 1106 DF,  p-value: <0.0000000000000002
```

Word of caution here is that the relationship seems to be very slightly curved, though mostly linear. We may need to consider additional variables as confounders that can then help explain this relationship better. Eliminating extreme outliers is yet another way we can deal with linearity.

2. Poverty and Higher 10% Of Income Share

Below we see *positive correlation between poverty and income share of higher 10%.*

This is indicating that *poverty level only rises when the higher 10% income holders have even more income.*



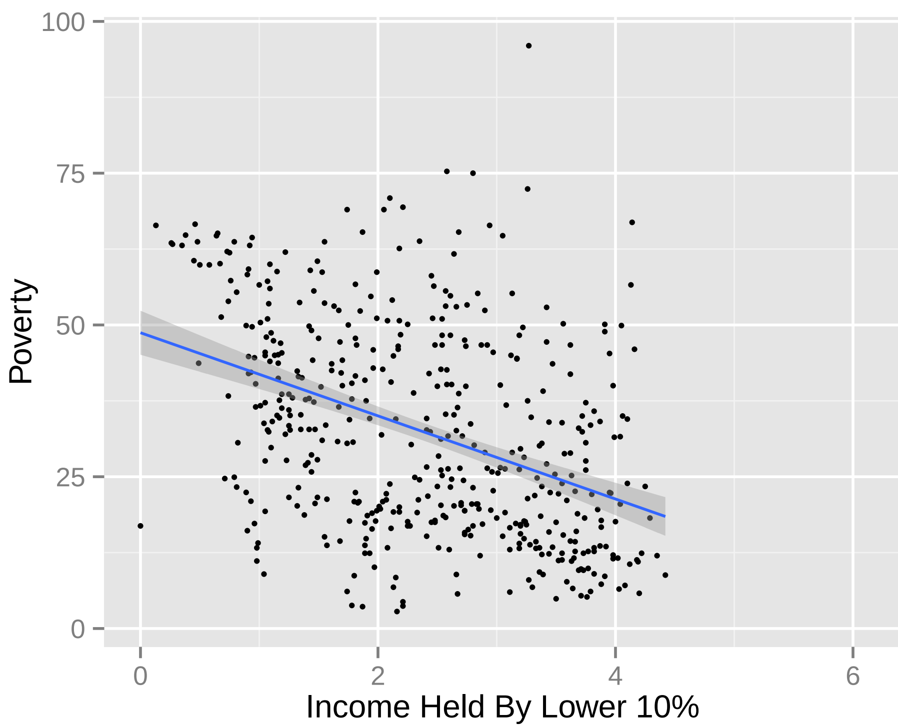
Looking at the simple regression model, we find fairly positive correlation, notice the nearly zero *p-value* with positive co-efficient below.

```
##
## Call:
## lm(formula = poverty ~ income_high10, data = wdi_data)
##
## Residuals:
##    Min     1Q Median     3Q    Max
##  -35.4  -10.4   -2.1   11.2   73.7
##
## Coefficients:
##               Estimate Std. Error t value            Pr(>|t|)
## (Intercept)    -2.6262     3.1245   -0.84                 0.4
## income_high10   1.0676     0.0928   11.51 <0.0000000000000002 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 15.3 on 460 degrees of freedom
##   (8954 observations deleted due to missingness)
## Multiple R-squared:  0.224,  Adjusted R-squared:  0.222
## F-statistic:  132 on 1 and 460 DF,  p-value: <0.0000000000000002
```

3. Poverty and Lower 10% Of Income Share

Below we see *negative correlation between poverty and income share of lower 10%.*

This is indicating that *poverty level goes down as the income goes up for the ones holding lower 10% of the income.*

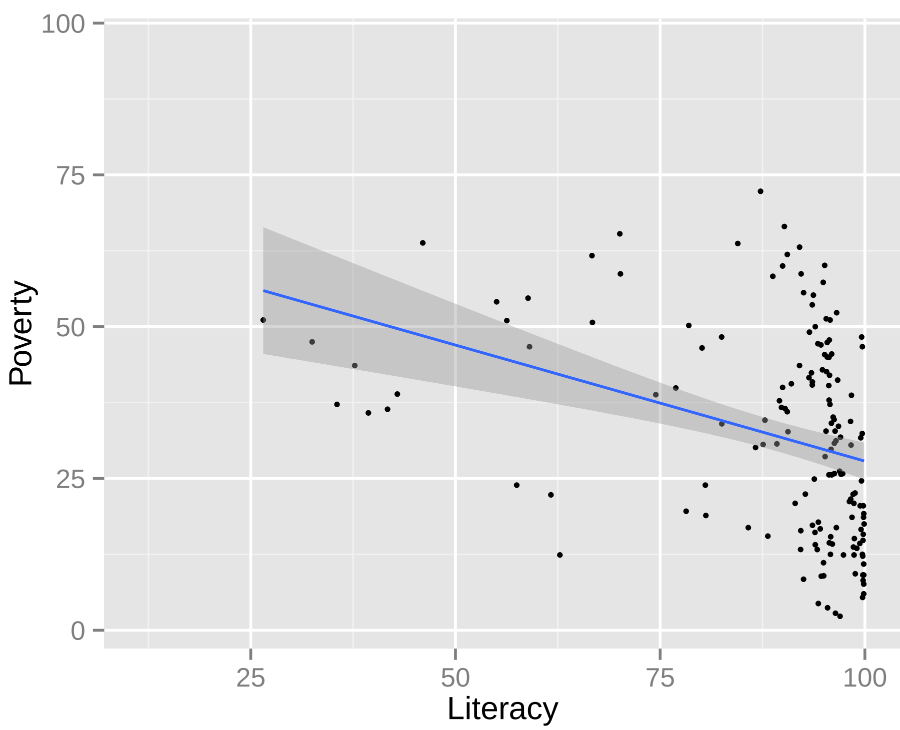


Looking at the simple regression model, we find fairly negative correlation, notice the nearly zero *p-value* with negative co-efficient below.

```
##
## Call:
## lm(formula = poverty ~ income_low10, data = wdi_data)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -32.73 -12.26  -2.07  11.43  69.67
##
## Coefficients:
##              Estimate Std. Error t value            Pr(>|t|)
## (Intercept)    48.720      1.847   26.38 <0.0000000000000002 ***
## income_low10   -6.847      0.709   -9.65 <0.0000000000000002 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 15.9 on 460 degrees of freedom
##   (8954 observations deleted due to missingness)
## Multiple R-squared:  0.168,  Adjusted R-squared:  0.167
## F-statistic: 93.2 on 1 and 460 DF,  p-value: <0.0000000000000002
```

4. Povery and Literacy

Below we see *negative correlation between poverty and literacy*. We would expect literacy can get people to get more and/or better employment. So, **bringing literacy to uneducated parts of the population can help reduce poverty**.



The trend is somewhat questionable in the statistical sense, in that the data seems to be fanning out. Further analysis using the regression model indicates low *p-value*, but not a very high R^2. So, we take into account the combination of all of the above variables to proceed with analysis using multiple regression.

```
##
## Call:
## lm(formula = poverty ~ income_low10, data = wdi_data)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -32.73 -12.26  -2.07  11.43  69.67
##
## Coefficients:
##              Estimate Std. Error t value            Pr(>|t|)
## (Intercept)    48.720      1.847   26.38 <0.0000000000000002 ***
## income_low10   -6.847      0.709   -9.65 <0.0000000000000002 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 15.9 on 460 degrees of freedom
##   (8954 observations deleted due to missingness)
## Multiple R-squared:  0.168,  Adjusted R-squared:  0.167
## F-statistic: 93.2 on 1 and 460 DF,  p-value: <0.0000000000000002
```

5. Multiple Regression Model

While we observe relationships above in four cases using simple regression model, we continue on to analyze them in a combined way using multiple regression.

We attempt to predict poverty using explanatoray varaibles as *gdp*, *income_low10* and *literacy*. Iterations of other combinations were dropped using stepwise refinment of the model.

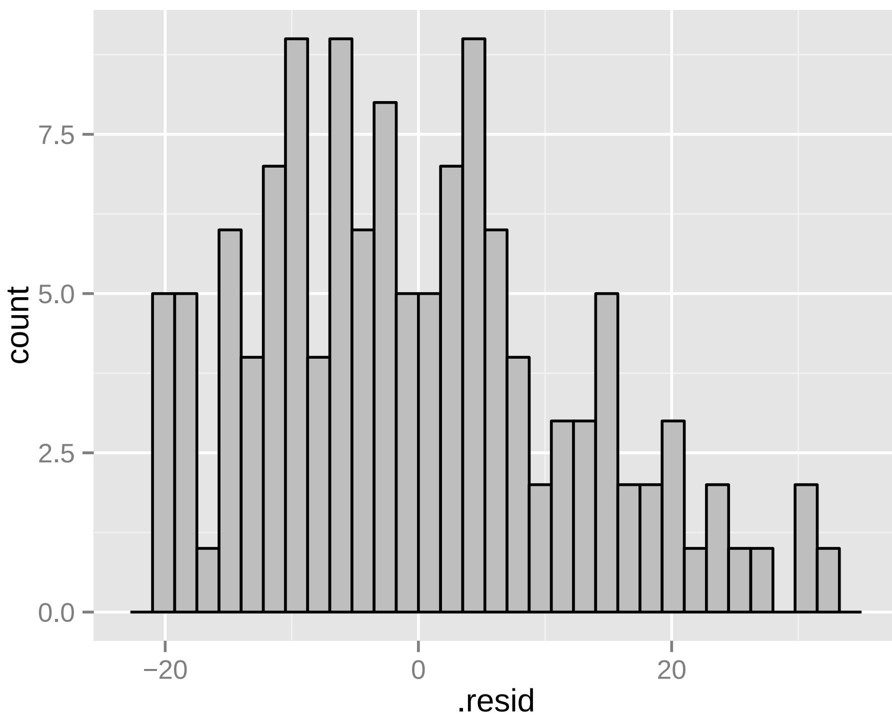
```
lmWDI=lm(poverty ~ gdp + income_low10 + literacy, data = wdi_data)
summary(lmWDI)
```

```
##
## Call:
## lm(formula = poverty ~ gdp + income_low10 + literacy, data = wdi_data)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -20.52  -9.81  -1.55   6.77  31.98
##
## Coefficients:
##                       Estimate        Std. Error t value
## (Intercept)  97.47569499340366  8.23393641012655   11.84
## gdp          -0.00000000001217  0.00000000000266   -4.58
## income_low10 -9.47925614894174  1.30126094186305   -7.28
## literacy     -0.46790060581192  0.07975164310094   -5.87
##                          Pr(>|t|)
## (Intercept)  < 0.0000000000000002 ***
## gdp                0.000011353307 ***
## income_low10       0.000000000033 ***
## literacy           0.000000037835 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
```

```
## Residual standard error: 12.6 on 124 degrees of freedom
##   (9288 observations deleted due to missingness)
## Multiple R-squared:  0.41,   Adjusted R-squared:  0.396
## F-statistic: 28.7 on 3 and 124 DF,  p-value: 0.0000000000000355
```

6. Distribution of Residuals

We check for residuals to see if they are fairly *normally distributided* below. We find that the residuals for the model more or less does follow normal distribution. However, we do see some extreme outliers, masking the symmetry of the distribution to some extent.



7. Analysis of Variance

Here, we observe very small p-values.

```
anova(lmWDI)
```

```
## Analysis of Variance Table
##
## Response: poverty
##               Df Sum Sq Mean Sq F value      Pr(>F)
## gdp            1   2515    2515    15.9     0.00011 ***
## income_low10   1   5664    5664    35.9 0.000000021 ***
## literacy       1   5436    5436    34.4 0.000000038 ***
## Residuals    124  19585     158
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

IV. Conclusion and Discussion

This paper demonstrates the poverty level and income share of lower 10% is significantly correlated to GDP, income of lower 10%, and literacy, with *p-values* for each of the three explanatory variables close to zero. With such *small p-values*, *R^2* of *41%* , *F-statistic* of *28.7* and a fairly *normal distribution of the residuals*, we have significant confidence in the model.

To further read into this model and reason out, we can say that increase in GDP resulting from many factors such as increased productivity, trade, etc causes increase in employment. More economic activities and increase in employment directly as well as indirectly increases income of wage earners. That in turn results in an uplift to the earnings of the lower income population and help reduce poverty.

One flaw in this study is that there are too many possibile confounding variables that can skew the model. Also, all of *the study is inherently observational in nature* and so *no conclusions of cause-and-effect should be drawn.* As frequently stated, *correlation is not causation.* So then, what good is this study for? Well, it is good in that it adds value by studying and observing different economic and social attributes that affect income and hence betterment for the population, in particular the ones that are in the lower income or living in poverty.

The wider implication of this study is that government bodies and organizations should recognize the critical importance of the *growing income disparity*, its *negative effect on poverty*, and *potentially positive effect of increase in literacy.*

This study is in no way conclusive and *can be improved with injection of additional confounding variables such as gender, region, race, geo-political factors* and many more. It may be worthwhile to investigate such additional factors and statistical methods to analyze and potentially improve income for low earners and poverty reduction.

V. References

1. World Bank: Data Catalog - Wealth Accounting

- http://data.worldbank.org/data-catalog/wealth-accounting

2. WDI: World Development Indicators (World Bank)

- *Search, extract and format data from the World Bank's World Development Indicators in R*
- http://cran.r-project.org/web/packages/WDI/WDI.pdf

3. ggplot2: An implementation of the Grammar of Graphics

- http://cran.r-project.org/web/packages/ggplot2/ggplot2.pdf

4. pander: An R Pandoc Writer

- http://cran.r-project.org/web/packages/pander/pander.pdf

VI. Appendix

1. Various CRAN Modules and R Options Used

```r
require(WDI)
library(ggplot2)
library(pander)


options(warn=-1)
options(scipen=999)
panderOptions('table.split.table', Inf)
panderOptions('table.split.cells', Inf)
panderOptions('table.alignment.default', 'left')
```

2. Acquiring WDI Data

```r
# initial extract of all countries using WDI(country = "all", indicators...) had all
# countries as well as additional rows for regional aggregations. Extracted the
# country codes, removed the region rows, then saved it to local CSV file
raw_country_code = read.csv('C:/STAT-100/Project/country_codes.csv',
                            skip=0,header = FALSE)
country_code=c(raw_country_code[1])
all_countries=as.character(country_code$V1)

# Zambia country code of 'NA' is read in as R NA
# Updating it to be character 'NA'
all_countries[which(is.na(all_countries))]='NA'

# Get WDI data using API and functions as part of CRAN WDI module
wdi_data=WDI(country = all_countries, indicator=c('BN.GSR.FCTY.CD','NY.GDP.MKTP.CD',
                                                  'SI.POV.NAHC','SI.DST.FRST.10',
                                                  'SI.DST.10TH.10','SE.ADT.LITR.ZS',
                                                  'SE.ADT.1524.LT.ZS','SP.POP.TOTL'),
             start=1970,end=2013,extra=FALSE)
names(wdi_data)=c('isoc2','country','year','net_income','gdp','poverty',
                  'income_low10','income_high10','literacy_a','literacy_y',
                  'population')

# 2 literacy pecentages - under 15 and above 15 - being averaged
# as proportion of population in each category i snot known
wdi_data$literacy=(wdi_data$literacy_a+wdi_data$literacy_y)/2
```

3. Various mean(s) used in the paper

```r
# average income share of top 10% for year 2003 and 2013
mean_hi10_2003=
  round(mean(wdi_data[which(!is.na(wdi_data$income_high10) & wdi_data$year=='2003'),
                      'income_high10']),0)
mean_hi10_2013=
  round(mean(wdi_data[which(!is.na(wdi_data$income_high10) & wdi_data$year=='2013'),
                      'income_high10']),0)
```

```
# average income share of top 10% for year 2003 and 2013
mean_low10_2003=
  round(mean(wdi_data[which(!is.na(wdi_data$income_low10) & wdi_data$year=='2003'),
                      'income_low10']),2)
mean_low10_2013=
  round(mean(wdi_data[which(!is.na(wdi_data$income_low10) & wdi_data$year=='2013'),
                      'income_low10']),2)

# average GDP for year 2003 and 2013
mean_gdp_2003=
  round(mean(wdi_data[which(!is.na(wdi_data$gdp) & wdi_data$year=='2003'),'gdp'])
        /1000000000,0)
mean_gdp_2013=
  round(mean(wdi_data[which(!is.na(wdi_data$gdp) & wdi_data$year=='2013'),'gdp'])
        /1000000000,0)

# average poverty for year 2013
mean_poverty_2013=
  round(mean(wdi_data[which(!is.na(wdi_data$poverty) & wdi_data$year=='2013'),
                      'poverty']),0)
```

4. Indicators Used

```
# WDI Indicators used to lookup specific data points for the study:
indicator_description=WDIsearch(string='BN.GSR.FCTY.CD$|NY.GDP.MKTP.CD$|SI.POV.NAHC$|
                                SI.DST.FRST.10$|SI.DST.10TH.10$|SE.ADT.LITR.ZS$|
                                SE.ADT.1524.LT.ZS$|SP.POP.TOTL$', field='indicator')
colnames(indicator_description)=c('Indicator','Description')

set.caption('Subset of WDI Indicators')
pander(indicator_description)
```

Indicator	Description
BN.GSR.FCTY.CD	Net income (BoP, current US$)
NY.GDP.MKTP.CD	GDP (current US$)
SE.ADT.LITR.ZS	Literacy rate, adult total (% of people ages 15 and above)
SI.DST.10TH.10	Income share held by highest 10%
SI.POV.NAHC	Poverty headcount ratio at national poverty line (% of population)
SP.POP.TOTL	Population, total

Table 3: Subset of WDI Indicators

5. Sample Data

```r
# Sample Data:
wdi_data_disp=wdi_data
wdi_data_disp$gdp=wdi_data_disp$gdp/1000000000
names(wdi_data_disp)=c('isoc2','Country','Yr','net_income','GDP','Poverty','IncmLo10',
                       'IncmHi10','literacy_a','literacy_y','population','Literacy')

set.caption('Sample data based on Indicators from Table 1')
pander(head(wdi_data_disp[which( !is.na(wdi_data_disp$Literacy)
                                 & !is.na(wdi_data_disp$IncmLo10) ),
                          c(2,3,5,6,7,8,12)]))
```

	Country	Yr	GDP	Poverty	IncmLo10	IncmHi10	Literacy
215	Albania	2008	12.88	12.4	3.73	24.46	97.38
252	Armenia	2001	2.118	48.3	3.19	29.1	99.6
263	Armenia	2012	9.958	32.4	3.72	24.76	99.68
330	Argentina	1991	189.7	NA	1.75	36.6	97.22
340	Argentina	2001	325.5	NA	0.65	39.46	98.05
782	Burkina Faso	2003	4.206	51.1	2.46	33.97	26.53

Table 4: Sample data based on Indicators from Table 1

6. Income Disparity with the Simple Regression Model

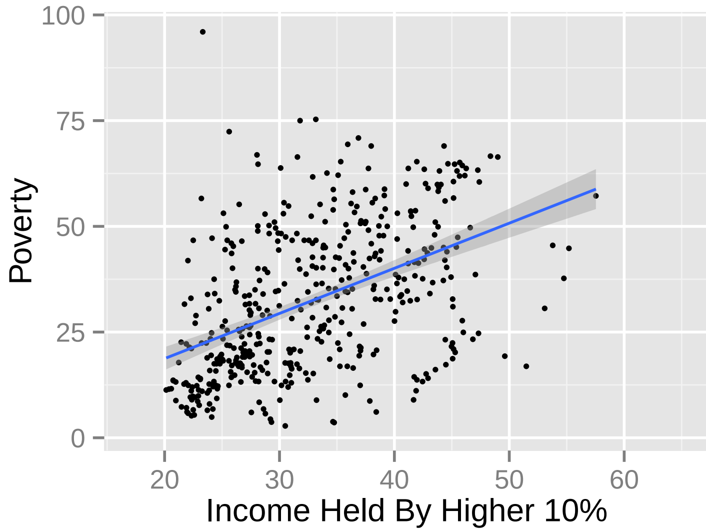
```r
# Income Disparity:
ggplot(wdi_data, aes(x=income_high10, y=income_low10)) +
  geom_point(size=1) +  geom_smooth(method="lm") +
  labs(x='Income Held By Higher 10%', y='Income Held By Lower 10%')
```


```
# model for income disparity
summary(lm(income_low10 ~ income_high10, data=wdi_data))
```

```
##
## Call:
## lm(formula = income_low10 ~ income_high10, data = wdi_data)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -2.2208 -0.3710  0.0196  0.3511  2.6279
##
## Coefficients:
##                Estimate Std. Error t value            Pr(>|t|)
## (Intercept)     6.22675    0.07087    87.9 <0.0000000000000002 ***
## income_high10 -0.11992    0.00216   -55.6 <0.0000000000000002 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.582 on 1106 degrees of freedom
##   (8308 observations deleted due to missingness)
## Multiple R-squared:  0.736,  Adjusted R-squared:  0.736
## F-statistic: 3.09e+03 on 1 and 1106 DF,  p-value: <0.0000000000000002
```

7. Povery and Higher 10% Of Income Share with the Simple Regression Model

```
# Poverty and Higher 10% Of Income Share
ggplot(wdi_data, aes(x=income_high10, y=poverty)) + geom_point(size=1) +
  geom_smooth(method="lm") + labs(x='Income Held By Higher 10%', y='Poverty')
```

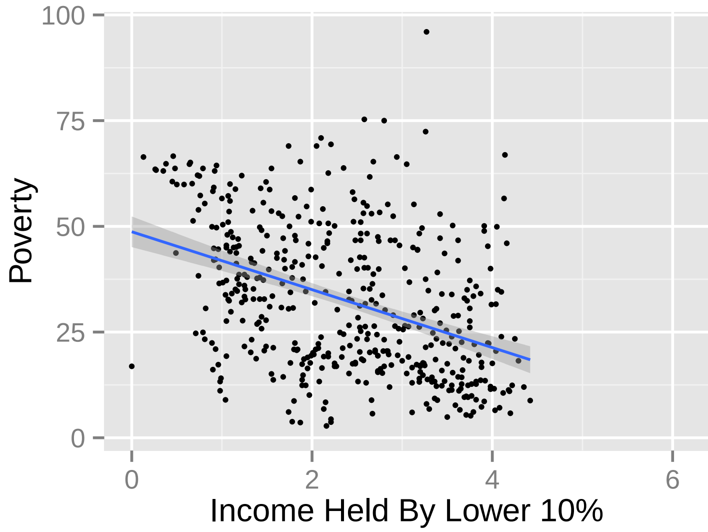



```r
summary(lm(poverty ~ income_high10, data=wdi_data))
```

```
##
## Call:
## lm(formula = poverty ~ income_high10, data = wdi_data)
##
## Residuals:
##    Min     1Q Median     3Q    Max
##  -35.4  -10.4   -2.1   11.2   73.7
##
## Coefficients:
##               Estimate Std. Error t value            Pr(>|t|)
## (Intercept)    -2.6262     3.1245   -0.84                 0.4
## income_high10   1.0676     0.0928   11.51 <0.0000000000000002 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 15.3 on 460 degrees of freedom
##   (8954 observations deleted due to missingness)
## Multiple R-squared:  0.224,  Adjusted R-squared:  0.222
## F-statistic:  132 on 1 and 460 DF,  p-value: <0.0000000000000002
```

8. Povery and Lower 10% Of Income Share with the Simple Regression Model

```r
# Poverty and Lower 10% Of Income Share
ggplot(wdi_data, aes(x=income_low10, y=poverty)) + geom_point(size=1) +
  geom_smooth(method="lm") + labs(x='Income Held By Lower 10%', y='Poverty')
```

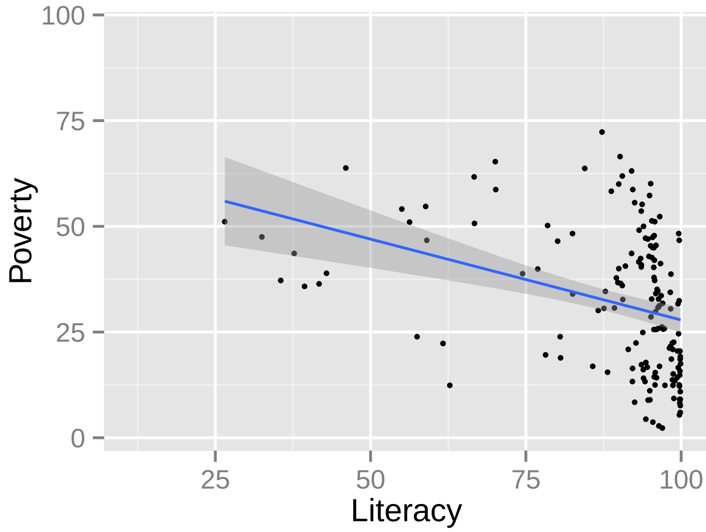



```
# simple regression model
summary(lm(poverty ~ income_low10, data=wdi_data))
```

```
##
## Call:
## lm(formula = poverty ~ income_low10, data = wdi_data)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -32.73 -12.26  -2.07  11.43  69.67
##
## Coefficients:
##              Estimate Std. Error t value            Pr(>|t|)
## (Intercept)    48.720      1.847   26.38 <0.0000000000000002 ***
## income_low10   -6.847      0.709   -9.65 <0.0000000000000002 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 15.9 on 460 degrees of freedom
##   (8954 observations deleted due to missingness)
## Multiple R-squared:  0.168,  Adjusted R-squared:  0.167
## F-statistic: 93.2 on 1 and 460 DF,  p-value: <0.0000000000000002
```

9. Povery and Literacy with the Simple Regression Model

```
# Povery and Literacy
ggplot(wdi_data, aes(x=literacy, y=poverty)) +
  geom_point(size=1) +  geom_smooth(method="lm") + labs(x='Literacy', y='Poverty')
```




```r
# simple regression model
summary(lm(poverty ~ income_low10, data=wdi_data))
```

```
##
## Call:
## lm(formula = poverty ~ income_low10, data = wdi_data)
##
## Residuals:
##    Min      1Q Median     3Q    Max
## -32.73 -12.26  -2.07  11.43  69.67
##
## Coefficients:
##              Estimate Std. Error t value            Pr(>|t|)
## (Intercept)    48.720      1.847   26.38 <0.0000000000000002 ***
## income_low10   -6.847      0.709   -9.65 <0.0000000000000002 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 15.9 on 460 degrees of freedom
##   (8954 observations deleted due to missingness)
## Multiple R-squared:  0.168,  Adjusted R-squared:  0.167
## F-statistic: 93.2 on 1 and 460 DF,  p-value: <0.0000000000000002
```

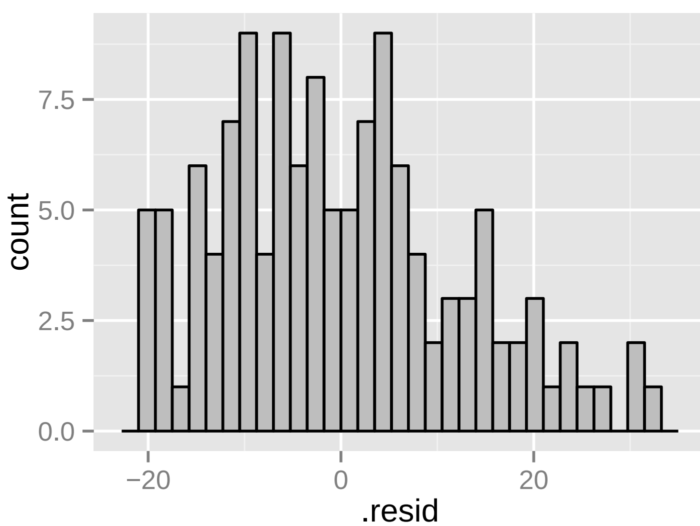
10. Multiple Regression Model and Residual Histogram

```r
# Multiple Regression Model
lmWDI=lm(poverty ~ gdp + income_low10 + literacy, data = wdi_data)
summary(lmWDI)
```

```
##
## Call:
## lm(formula = poverty ~ gdp + income_low10 + literacy, data = wdi_data)
```

```
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -20.52  -9.81  -1.55   6.77  31.98
##
## Coefficients:
##                       Estimate        Std. Error t value
## (Intercept)  97.47569499340366  8.23393641012655   11.84
## gdp          -0.00000000001217  0.00000000000266   -4.58
## income_low10 -9.47925614894174  1.30126094186305   -7.28
## literacy     -0.46790060581192  0.07975164310094   -5.87
##                          Pr(>|t|)
## (Intercept)  < 0.0000000000000002 ***
## gdp                0.000011353307 ***
## income_low10       0.000000000033 ***
## literacy           0.000000037835 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 12.6 on 124 degrees of freedom
##   (9288 observations deleted due to missingness)
## Multiple R-squared:  0.41,   Adjusted R-squared:  0.396
## F-statistic: 28.7 on 3 and 124 DF,  p-value: 0.0000000000000355
```

```
# histogram of residuals
flm=fortify(lmWDI)
ggplot(flm, aes(x=.resid)) + geom_histogram(fill="grey", color="black")
```



“‘

11. Analysis of Variance

```r
# Analysis of Variance
anova(lmWDI)
```

```
## Analysis of Variance Table
##
## Response: poverty
##               Df Sum Sq Mean Sq F value      Pr(>F)
## gdp            1   2515    2515    15.9     0.00011 ***
## income_low10   1   5664    5664    35.9 0.000000021 ***
## literacy       1   5436    5436    34.4 0.000000038 ***
## Residuals    124  19585     158
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

12. One of the alternative models among the few that were looked at

```r
# Alternative model among the many looked at
summary(lm(poverty ~ gdp + income_low10 + literacy + population +
             net_income + income_high10, data = wdi_data))
```

```
##
## Call:
## lm(formula = poverty ~ gdp + income_low10 + literacy + population +
##     net_income + income_high10, data = wdi_data)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -23.43  -7.91  -2.09   6.12  30.18
##
## Coefficients:
##                         Estimate        Std. Error t value   Pr(>|t|)
## (Intercept)    36.41583927655401 21.41571268661858    1.70    0.09254 .
## gdp            -0.00000000000045  0.00000000000741   -0.06    0.95171
## income_low10   -1.83595320335941  2.79798480277874   -0.66    0.51341
## literacy       -0.51463216195026  0.09463434894304   -5.44 0.00000047 ***
## population     -0.00000005761251  0.00000004132823   -1.39    0.16678
## net_income      0.00000000042783  0.00000000040106    1.07    0.28898
## income_high10   1.44271734476952  0.37661751309501    3.83    0.00024 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 10.8 on 89 degrees of freedom
##   (9320 observations deleted due to missingness)
## Multiple R-squared:  0.565,  Adjusted R-squared:  0.536
## F-statistic: 19.3 on 6 and 89 DF,  p-value: 0.0000000000000272
```