

Public Expenditure and Poverty

Team 1: Le Van Minh, Callista Stephine Yu, Muhammad Imtiyaaz B M A M , Thum Jun Long Issac

2022-11-17

1. Introduction

The project aim is to explore different models using the same dataset to evaluate which method will be the best to build a model for the dependent variable. The dataset we are using for this project has 1901 data values from 23 different variables. Our plan for this project is to first build a model with all the variables in the dataset. We will then follow it up by building a model without the missing values and build a ridge regression and Lasso regression model. Our second approach will be to selectively remove variables from the model and to build the best model using regularisation and regression. Our next approach was to add the variables into the model one at a time to see if a different model can be achieved. Lastly, we performed imputation on our model to tackle the missing values present in our dataset. We then compared all of this approaches to come to a conclusion.

Github repo: <https://github.com/IntrovertHedgehog/PublicExpenditureAndPoverty>

2. Data Characteristic

```
library(needs)

needs("dplyr", "MASS", "readr", "tidyr", "ggplot2",
      "e1071", "moments", "corrplot", "Hmisc", "PerformanceAnalytics",
      "mice", "car", "glmnet", "ggforce", "lmtest")

prioritize(dplyr)
```

2.1. Nature of Data

The data set is collection The World Bank Data, the variables of interest are extracted from the raw data files and combined into a single data frame for analysis. The final data set includes:

1. **country.code**: Country code
2. **country.name**: Country name
3. **year**: Year
4. **income**: Income class

- Low income (L)
- Lower middle income (LM)
- Upper middle income (UM)
- High income (H)

5. **reg**: Region

6. **pov**: Poverty headcount ratio based on cut-off value of \$2.15 per day

7. **mpi**: Multidimensional Poverty Index

8. **edu.total**: Total expenditure on education (% of GDP)

9. **edu.pri**: Total expenditure on primary education (% of total education expenditure)

10. **edu.sec**: Total expenditure on secondary education (% of total education expenditure)

11. **edu.ter**: Total expenditure on tertiary education (% of total education expenditure)

12. **hlth**: Total expenditure on health (% of GDP)

13. **mil**: Total expenditure on military (% of GDP)

14. **fdi**: Foreign Direct Investment

15. **lbr.part**: Labour force participation (% of population ages 15+)

16. **unemp**: Unemployment rate

17. **pop.gwth.total**: Total population growth rate

18. **pop.gwth.rural**: Total rural population growth rate

19. **pop.gwth.urban**: Total urban population growth rate

20. **gdp.dflt**: GDP deflator

21. **gdr.eq1**: Gender equality rating

22. **gcf**: Gross Capital Formation

23. **trade**: Trade = import + export (% of GDP)

24. **gdp.pc**: GDP per capita (current US\$)

Data imports and combining:

```
# helper functions
importWDI <- function(filepath, value_name) {
  df <- read_csv(filepath, skip = 4)
```

```

colnames(df) <- tolower(gsub(" ", ".", colnames(df)))

df <- df %>%
  pivot_longer(5:ncol(.), names_to = "year",
  values_to = "value") %>%
  filter(!is.null(value) & !is.na(value)) %>%
  mutate(country.code = factor(country.code),
    country.name = factor(country.name), year = as.numeric(year)) %>%
  select(country.code, country.name, year, value)

colnames(df)[4] <- value_name

df
}

importRegionClass <- function(filepath) {
  df <- read_csv(filepath, skip = 4)

  colnames(df) <- tolower(gsub(" ", ".", colnames(df)))

  df %>%
    mutate(country.name = factor(country.name),
      region = factor(region)) %>%
    select(country.name, reg = region)
}

importIncomeClass <- function(filepath) {
  df <- read_csv(filepath, skip = 4)

  colnames(df) <- tolower(gsub(" ", ".", colnames(df)))

  df %>%
    pivot_longer(3:ncol(.), names_to = "year",
    values_to = "income") %>%
    filter(!is.null(income) & !is.na(income)) %>%
    mutate(country.code = factor(country.code),
      country.name = factor(country.name), year = as.numeric(year),
      income = factor(income)) %>%
    select(country.code, country.name, year, income)
}

# import data
setwd("../data")

poverty.headcount <- importWDI("poverty.headcount.215dollar.csv",
  "pov")
mpi <- importWDI("mpi.csv", "mpi")
education.expenditure.total <- importWDI("total.education.expenditure.csv",
  "edu.total")
education.expenditure.primary <- importWDI("primary.education.expenditure.csv",
  "edu.pri")
education.expenditure.secondary <- importWDI("secondary.education.expenditure.csv",
  "edu.sec")

```

```

education.expenditure.tertiary <- importWDI("tertiary.education.expenditure.csv",
  "edu.ter")
health.expenditure <- importWDI("health.expenditure.csv",
  "hlth")
military.expenditure <- importWDI("military.expenditure.csv",
  "mil")
fdi <- importWDI("fdi.csv", "fdi")
labour.force.participation <- importWDI("labour.force.participation.csv",
  "lbr.part")
unemployment.rate <- importWDI("unemployment.csv",
  "unemp")
population.growth <- importWDI("population.growth.csv",
  "pop.gwth.total")
rural.population.growth <- importWDI("rural.population.growth.csv",
  "pop.gwth.rural")
urban.population.growth <- importWDI("urban.population.growth.csv",
  "pop.gwth.urban")
gdp.deflator <- importWDI("gdp.deflator.csv", "gdp.dflt")
gender.equality <- importWDI("gender.equality.csv",
  "gdr.eql")
gross.capital.formation <- importWDI("gross.capital.formation.csv",
  "gcf")
trade <- importWDI("trade.csv", "trade")
region.class <- importRegionClass("region.class.csv")
income.class <- importIncomeClass("income.class.csv")
gdp.pc <- importWDI("gdp.pc.csv", "gdp.pc")

setwd("../src")

```

We found that the data sets collected from [World Bank's Data helpdesk](#) and [The World Bank's Data](#) have different naming convention for certain countries (e.g. “Czechia” vs. “Czechia Republic”). So we need to rename these countries to avoid some error when joining.

Furthermore, WDI’s data sets rate also account for non-country (e.g. country.name = “Low income” or “South Asia”). These special groups are not in our scope of interest, which is national, so we eliminate them.

```

# using poverty.headcount as a naming standard
# (as other data from WDI also use this
# convention) join a subset of data to process
# the names
d <- poverty.headcount %>%
  select(country.name) %>%
  mutate(inPov = T) %>%
  full_join(income.class %>%
    select(country.name) %>%
    mutate(inIncome = T), by = "country.name") %>%
  full_join(region.class %>%
    select(country.name) %>%
    mutate(inReg = T), by = "country.name") %>%
  mutate(inPov = !is.na(inPov), inIncome = !is.na(inIncome),
    inReg = !is.na(inReg))

d

```

```

## # A tibble: 62,759 x 4
##   country.name inPov inIncome inReg
##   <fct>        <lgl> <lgl>   <lgl>
## 1 Angola       TRUE  TRUE    TRUE
## 2 Angola       TRUE  TRUE    TRUE
## 3 Angola       TRUE  TRUE    TRUE
## 4 Angola       TRUE  TRUE    TRUE
## 5 Angola       TRUE  TRUE    TRUE
## 6 Angola       TRUE  TRUE    TRUE
## 7 Angola       TRUE  TRUE    TRUE
## 8 Angola       TRUE  TRUE    TRUE
## 9 Angola       TRUE  TRUE    TRUE
## 10 Angola      TRUE  TRUE    TRUE
## # ... with 62,749 more rows
## # i Use 'print(n = ...)' to see more rows

```

First, remove special economic groups from `poverty.headcount`. We figured these regions will not appear in `income.class` or `region.class`, so we might find something from looking at the countries **only** appear in `poverty.headcount`.

```

d %>%
  filter(inPov & (!inIncome | !inReg)) %>%
  distinct(country.name)

```

```

## # A tibble: 18 x 1
##   country.name
##   <fct>
## 1 Cote d'Ivoire
## 2 Czechia
## 3 East Asia & Pacific
## 4 Europe & Central Asia
## 5 Fragile and conflict affected situations
## 6 High income
## 7 IDA total
## 8 Latin America & Caribbean
## 9 Low income
## 10 Lower middle income
## 11 Low & middle income
## 12 Middle East & North Africa
## 13 South Asia
## 14 Sub-Saharan Africa
## 15 Sao Tome and Principe
## 16 Turkiye
## 17 Upper middle income
## 18 World

```

Lucky! We can look through these 18 results and compose a list of special regions.

```

spec.reg <- c("Fragile and conflict affected situations",
  "IDA total", "World", "East Asia & Pacific", "Europe & Central Asia",
  "Latin America & Caribbean", "Middle East & North Africa",
  "South Asia", "Sub-Saharan Africa", "Low income",
  "Low & middle income", "Lower middle income", "Upper middle income",
  "High income")

```

Then, we rename those countries with inconsistent naming convention. Since we should only care about countries whose poverty headcount is available, reusing the list generated above, we can identify:

1. Cote d'Ivoire (also Côte d'Ivoire)
2. Czechia (also Czechoslovakia or Czech Republic)
3. Curacao (also Curaçao)
4. Turkiye (formerly known as Turkey, also Türkiye)
5. Sao Tome and Principe (also São Tomé and Príncipe)

```
# mapping standard name and variation
nameMap <- tibble(standard = c("Cote d'Ivoire", "Czechia",
  "Czechia", "Curacao", "Turkiye", "Turkiye", "Sao Tome and Principe"),
  variation = c("Côte d'Ivoire", "Czechoslovakia",
  "Czech Republic", "Curaçao", "Turkey", "Türkiye",
  "São Tomé and Príncipe"))

correctName <- function(name) {
  tibble(name = name) %>%
    left_join(nameMap, by = c(name = "variation")) %>%
    mutate(standard = ifelse(is.na(standard), name,
      standard)) %>%
    select(standard) %>%
    pull()
}

orig.name <- c("Vietnam", "China", "Turkey", "Czechia Republic")
correctName(orig.name)
```

```
## [1] "Vietnam"          "China"           "Turkiye"         "Czechia Republic"
```

Let's test this out!

```
d <- poverty.headcount %>%
  # correct name here
  mutate(country.name = correctName(country.name)) %>%
  select(country.name) %>%
  mutate(inPov = T) %>%
  full_join(income.class %>%
    # correct name here
    mutate(country.name = correctName(country.name)) %>%
    select(country.name) %>%
    mutate(inIncome = T), by = "country.name") %>%
  full_join(region.class %>%
    # correct name here
    mutate(country.name = correctName(country.name)) %>%
    select(country.name) %>%
    mutate(inReg = T), by = "country.name") %>%
  filter(!(country.name %in% spec.reg)) %>%
  mutate(inPov = !is.na(inPov), inIncome = !is.na(inIncome), inReg = !is.na(inReg))

# countries not in region list, but is in Pov list
d %>%
  filter(!inReg & inPov) %>%
```

```

distinct(country.name) %>%
nrow()

## [1] 0

# countries not in income list, but is in Pov list
d %>%
filter(!inIncome & inPov) %>%
distinct(country.name) %>%
nrow()

## [1] 0

```

We are *pretty* confident that there's no inconsistent naming left unprocessed in the data sets.

```

# Rename countries in all data sets.
poverty.headcount <- poverty.headcount %>%
  mutate(country.name = correctName(country.name))
mpi <- mpi %>%
  mutate(country.name = correctName(country.name))
education.expenditure.total <- education.expenditure.total %>%
  mutate(country.name = correctName(country.name))
education.expenditure.primary <- education.expenditure.primary %>%
  mutate(country.name = correctName(country.name))
education.expenditure.secondary <- education.expenditure.secondary %>%
  mutate(country.name = correctName(country.name))
education.expenditure.tertiary <- education.expenditure.tertiary %>%
  mutate(country.name = correctName(country.name))
health.expenditure <- health.expenditure %>%
  mutate(country.name = correctName(country.name))
military.expenditure <- military.expenditure %>%
  mutate(country.name = correctName(country.name))
fdi <- fdi %>%
  mutate(country.name = correctName(country.name))
labour.force.participation <- labour.force.participation %>%
  mutate(country.name = correctName(country.name))
unemployment.rate <- unemployment.rate %>%
  mutate(country.name = correctName(country.name))
population.growth <- population.growth %>%
  mutate(country.name = correctName(country.name))
rural.population.growth <- rural.population.growth %>%
  mutate(country.name = correctName(country.name))
urban.population.growth <- urban.population.growth %>%
  mutate(country.name = correctName(country.name))
gdp.deflator <- gdp.deflator %>%
  mutate(country.name = correctName(country.name))
gender.equality <- gender.equality %>%
  mutate(country.name = correctName(country.name))
gross.capital.formation <- gross.capital.formation %>%
  mutate(country.name = correctName(country.name))
trade <- trade %>%
  mutate(country.name = correctName(country.name))

```

```

region.class <- region.class %>%
  mutate(country.name = correctName(country.name))
income.class <- income.class %>%
  mutate(country.name = correctName(country.name))
gdp.pc <- gdp.pc %>%
  mutate(country.name = correctName(country.name))

```

Join the data

```

countries <- poverty.headcount %>%
  # We used a full join here so we can conduct
  # a separate analysis on mpi later
full_join(mpi, by = c("country.name", "country.code",
  "year")) %>%
  left_join(income.class, c("country.name", "country.code",
    "year")) %>%
  left_join(region.class, by = "country.name") %>%
  left_join(education.expenditure.total, by = c("country.name",
    "country.code", "year")) %>%
  left_join(education.expenditure.primary, by = c("country.name",
    "country.code", "year")) %>%
  left_join(education.expenditure.secondary, by = c("country.name",
    "country.code", "year")) %>%
  left_join(education.expenditure.tertiary, by = c("country.name",
    "country.code", "year")) %>%
  left_join(health.expenditure, by = c("country.name",
    "country.code", "year")) %>%
  left_join(military.expenditure, by = c("country.name",
    "country.code", "year")) %>%
  left_join(fdi, by = c("country.name", "country.code",
    "year")) %>%
  left_join/labour.force.participation, by = c("country.name",
    "country.code", "year")) %>%
  left_join(unemployment.rate, by = c("country.name",
    "country.code", "year")) %>%
  left_join/population.growth, by = c("country.name",
    "country.code", "year")) %>%
  left_join(rural.population.growth, by = c("country.name",
    "country.code", "year")) %>%
  left_join(urban.population.growth, by = c("country.name",
    "country.code", "year")) %>%
  left_join(gdp.deflator, by = c("country.name",
    "country.code", "year")) %>%
  left_join(gender.equality, by = c("country.name",
    "country.code", "year")) %>%
  left_join(gross.capital.formation, by = c("country.name",
    "country.code", "year")) %>%
  left_join(trade, by = c("country.name", "country.code",
    "year")) %>%
  left_join(gdp.pc, by = c("country.name", "country.code",
    "year")) %>%
  # filter special groups
filter(!(country.name %in% spec.reg))

```

Data preview

```
head(countries)
```

```
## # A tibble: 6 x 24
##   count~1 count~2 year   pov   mpi income reg   edu.t~3 edu.pri edu.sec edu.ter
##   <fct>  <chr>   <dbl> <dbl> <dbl> <fct>  <fct>  <dbl> <dbl> <dbl> <dbl>
## 1 AGO    Angola  2000  21.4  NA L    Sub--  2.61   NA    NA    NA
## 2 AGO    Angola  2008  14.6  NA LM   Sub--  2.69   NA    NA    NA
## 3 AGO    Angola  2018  31.1  NA LM   Sub--  2.04   NA    NA    NA
## 4 ALB    Albania 1996   0.5  NA LM   Euro-- 3.08   NA    NA    NA
## 5 ALB    Albania 2002   1.1  NA LM   Euro-- 3.12   NA    NA    NA
## 6 ALB    Albania 2005   0.6  NA LM   Euro-- 3.28   NA    NA    NA
## # ... with 13 more variables: hlth <dbl>, mil <dbl>, fdi <dbl>, lbr.part <dbl>,
## #   unemp <dbl>, pop.gwth.total <dbl>, pop.gwth.rural <dbl>,
## #   pop.gwth.urban <dbl>, gdp.dflt <dbl>, gdr.eql <dbl>, gcf <dbl>,
## #   trade <dbl>, gdp.pc <dbl>, and abbreviated variable names 1: country.code,
## #   2: country.name, 3: edu.total
## # i Use 'colnames()' to see all variable names
```

```
str(countries)
```

```
## # tibble [1,901 x 24] (S3: tbl_df/tbl/data.frame)
## # $ country.code : Factor w/ 272 levels "AGO","ALB","ARE",...: 1 1 1 2 2 2 2 2 2 ...
## # $ country.name : chr [1:1901] "Angola" "Angola" "Angola" "Albania" ...
## # $ year         : num [1:1901] 2000 2008 2018 1996 2002 ...
## # $ pov          : num [1:1901] 21.4 14.6 31.1 0.5 1.1 0.6 0.2 0.6 1 0.1 ...
## # $ mpi          : num [1:1901] NA NA NA NA NA NA NA NA NA ...
## # $ income       : Factor w/ 4 levels "H","L","LM","UM": 2 3 3 3 3 3 3 4 4 4 ...
## # $ reg          : Factor w/ 7 levels "East Asia & Pacific",...: 7 7 7 2 2 2 2 2 2 ...
## # $ edu.total    : num [1:1901] 2.61 2.69 2.04 3.08 3.12 ...
## # $ edu.pri      : num [1:1901] NA NA NA NA NA ...
## # $ edu.sec      : num [1:1901] NA NA NA NA NA ...
## # $ edu.ter      : num [1:1901] NA NA NA NA NA ...
## # $ hlth         : num [1:1901] 1.91 3.32 2.54 NA 6.91 ...
## # $ mil          : num [1:1901] 6.39 3.57 1.87 1.38 1.32 ...
## # $ fdi          : num [1:1901] 8.79e+08 1.68e+09 -6.46e+09 9.01e+07 1.35e+08 ...
## # $ lbr.part     : num [1:1901] NA NA NA 38.8 59.6 ...
## # $ unemp        : num [1:1901] NA NA NA 12.3 15.8 ...
## # $ pop.gwth.total: num [1:1901] 3.277 3.711 3.276 -0.622 -0.3 ...
## # $ pop.gwth.rural: num [1:1901] 0.921 1.91 1.338 -1.546 -2.169 ...
## # $ pop.gwth.urban: num [1:1901] 5.682 5.02 4.312 0.812 2.181 ...
## # $ gdp.dflt     : num [1:1901] 418.02 19.37 28.17 38.17 3.65 ...
## # $ gdr.eql      : num [1:1901] NA 3 NA NA NA 4 NA NA NA NA ...
## # $ gcf          : num [1:1901] 30.5 30.8 17.9 18.1 35.3 ...
## # $ trade        : num [1:1901] 152.5 121.4 66.4 44.9 68.5 ...
## # $ gdp.pc       : num [1:1901] 557 4081 2525 1010 1425 ...
```

```
summary(countries)
```

```
##   country.code country.name           year        pov        mpi
##   BRA      : 36 Length:1901      Min.   :1967   Min.   : 0.0   Min.   : 2
```

```

##   CRI    : 34  Class :character  1st Qu.:2002   1st Qu.: 0.2  1st Qu.:18
##   ARG    : 32  Mode  :character  Median :2009   Median : 1.5  Median :25
##   USA    : 32                           Mean   :2007   Mean   :10.0  Mean   :27
##   DEU    : 30                           3rd Qu.:2014   3rd Qu.:11.6 3rd Qu.:33
##   HND    : 30                           Max.   :2021   Max.   :91.5  Max.   :74
##   (Other):1707                         NA's   :58     NA's   :1446
##   income                                reg      edu.total   edu.pri
##   H  :644   East Asia & Pacific    :167   Min.   : 1   Min.   : 1
##   L  :253   Europe & Central Asia  :883   1st Qu.: 4   1st Qu.:24
##   LM :501   Latin America & Caribbean:416   Median : 5   Median :30
##   UM :438   Middle East & North Africa:107   Mean   : 5   Mean   :32
##   NA's: 65   North America        : 50   3rd Qu.: 5   3rd Qu.:38
##   South Asia                          : 61   Max.   :16   Max.   :70
##   Sub-Saharan Africa                 :217   NA's   :596  NA's   :1090
##   edu.sec     edu.ter      hlth      mil      fdi
##   Min.   : 3   Min.   : 0   Min.   : 2   Min.   : 0.0  Min.   :-3.44e+11
##   1st Qu.:30  1st Qu.:17  1st Qu.: 5  1st Qu.: 1.0  1st Qu.: 2.98e+08
##   Median :36  Median :21  Median : 7  Median : 1.5  Median : 1.71e+09
##   Mean   :36  Mean   :21  Mean   : 7  Mean   : 1.8  Mean   : 1.60e+10
##   3rd Qu.:41  3rd Qu.:25  3rd Qu.: 9  3rd Qu.: 2.1  3rd Qu.: 9.82e+09
##   Max.   :72   Max.   :50   Max.   :18   Max.   :19.4  Max.   : 7.34e+11
##   NA's   :1094 NA's   :963  NA's   :464  NA's   :105  NA's   :17
##   lbr.part   unemp      pop.gwth.total pop.gwth.rural pop.gwth.urban
##   Min.   :30   Min.   : 0.2  Min.   :-3.63  Min.   :-8.56  Min.   :-4.08
##   1st Qu.:56  1st Qu.: 4.5  1st Qu.: 0.27  1st Qu.: -0.86 1st Qu.: 0.51
##   Median :61  Median : 6.9  Median : 1.03  Median : -0.02  Median : 1.48
##   Mean   :61   Mean   : 8.1  Mean   : 1.06  Mean   : 0.00  Mean   : 1.69
##   3rd Qu.:65  3rd Qu.:10.1 3rd Qu.: 1.78  3rd Qu.: 0.96  3rd Qu.: 2.66
##   Max.   :93   Max.   :49.7  Max.   : 5.61  Max.   : 4.60  Max.   :13.80
##   NA's   :320  NA's   :267   NA's   :14    NA's   :13
##   gdp.dflt   gdr.eql     gcf      trade     gdp.pc
##   Min.   :-26  Min.   : 2   Min.   : 0.0  Min.   : 1   Min.   : 120
##   1st Qu.: 2   1st Qu.: 3   1st Qu.:19.6  1st Qu.: 51  1st Qu.: 1928
##   Median : 4   Median : 4   Median :22.7  Median : 73  Median : 6032
##   Mean   : 16  Mean   : 4   Mean   :23.9  Mean   : 84  Mean   : 15220
##   3rd Qu.: 9   3rd Qu.: 4   3rd Qu.:26.8  3rd Qu.:105 3rd Qu.: 21490
##   Max.   :3334 Max.   : 5   Max.   :69.5  Max.   :380  Max.   :123679
##   NA's   :18   NA's   :1635 NA's   :72    NA's   :57  NA's   :8

```

There's no NA in `reg`, which is a sign that all naming in the data is remedied. There's some expected NAs in `income` and `pov`, as these data are collected by year. There's a substantial amount of missing data in `mpi`, as this is a relative new concept. We will address the nature, and processing of missing data in the next sections.

2.2. Missing values

As observed from the summary above, the data set contains a lot of missing values in some of the variables.

```
mean(is.na(countries))
```

```
## [1] 0.182
```

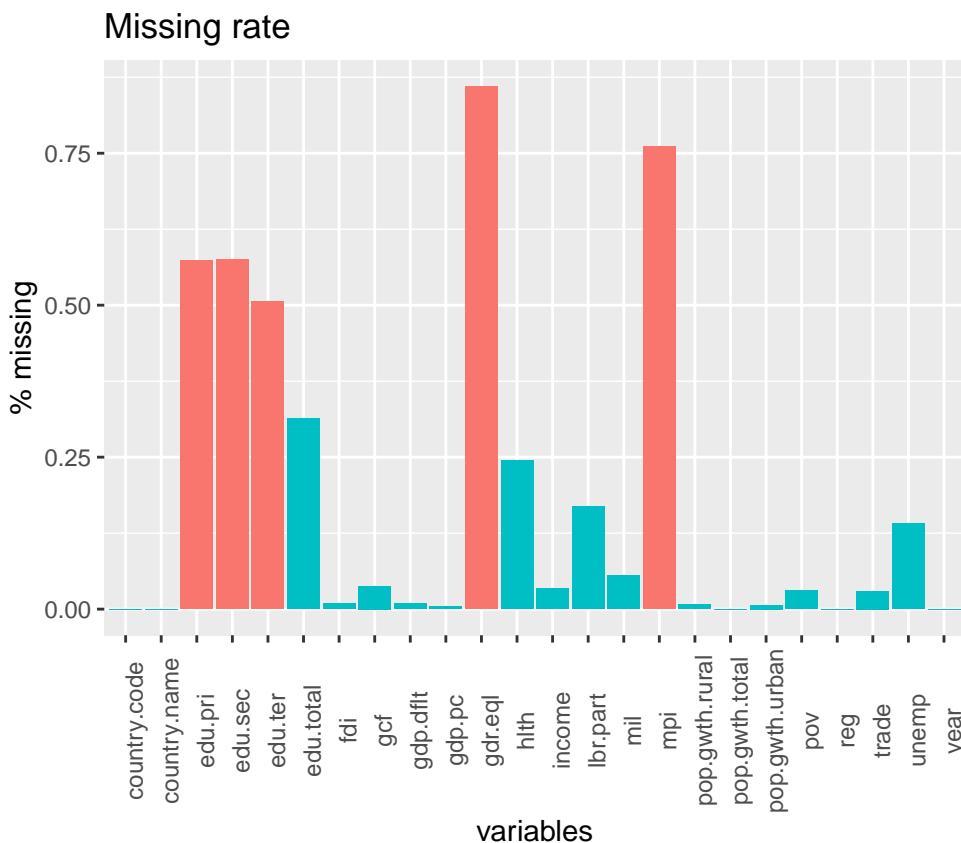
About 19% of the data set is missing.

```
nCompleteObs <- sum(complete.cases(countries))
print(paste("No. of complete cases:", nCompleteObs))
```

```
## [1] "No. of complete cases: 3"
```

There are only 3 complete cases where all the variable is available. This is nowhere near acceptable to conduct any meaningful analysis. Therefore, we need to eliminate some variables for a more balance data set.

```
missings <- colMeans(is.na(countries))
ggplot(mapping = aes(x = names(missings), y = missings,
  fill = missings < 0.35)) + geom_bar(stat = "identity") +
  ggtitle("Missing rate") + xlab("variables") + ylab("% missing") +
  theme(axis.text.x = element_text(size = 9, angle = 90))
```



```
missings[missings > 0.35]
```

```
##      mpi edu.pri edu.sec edu.ter gdr.eql
##  0.761   0.573   0.575   0.507   0.860
```

There are 5 variables with missing rate >35%.

expenditure in primary, secondary, and tertiary education can be very useful and relevant information to predict poverty reduction (Akbar et al. 2019). However, we would like to exclude these variables from some

first analyses to make use of the richer set of data. We can conduct a separate analysis with these variable to gain more insight.

```
# variables with high missing rate
hMiss <- names(missings[missings > 0.35])
# exclude these variables in countries1
countries1 <- countries %>%
  select(!hMiss) %>%
  filter(!is.na(pov))

## Note: Using an external vector in selections is ambiguous.
## i Use 'all_of(hMiss)' instead of 'hMiss' to silence this message.
## i See <https://tidyselect.r-lib.org/reference/faq-external-vector.html>.
## This message is displayed once per session.

str(countries1)

## tibble [1,843 x 19] (S3:tbl_df/tbl/data.frame)
## $ country.code : Factor w/ 272 levels "AGO","ALB","ARE",...: 1 1 1 2 2 2 2 2 2 ...
## $ country.name : chr [1:1843] "Angola" "Angola" "Angola" "Albania" ...
## $ year : num [1:1843] 2000 2008 2018 1996 2002 ...
## $ pov : num [1:1843] 21.4 14.6 31.1 0.5 1.1 0.6 0.2 0.6 1 0.1 ...
## $ income : Factor w/ 4 levels "H","L","LM","UM": 2 3 3 3 3 3 4 4 4 ...
## $ reg : Factor w/ 7 levels "East Asia & Pacific",...: 7 7 7 2 2 2 2 2 2 ...
## $ edu.total : num [1:1843] 2.61 2.69 2.04 3.08 3.12 ...
## $ hlth : num [1:1843] 1.91 3.32 2.54 NA 6.91 ...
## $ mil : num [1:1843] 6.39 3.57 1.87 1.38 1.32 ...
## $ fdi : num [1:1843] 8.79e+08 1.68e+09 -6.46e+09 9.01e+07 1.35e+08 ...
## $ lbr.part : num [1:1843] NA NA NA 38.8 59.6 ...
## $ unemp : num [1:1843] NA NA NA 12.3 15.8 ...
## $ pop.gwth.total: num [1:1843] 3.277 3.711 3.276 -0.622 -0.3 ...
## $ pop.gwth.rural: num [1:1843] 0.921 1.91 1.338 -1.546 -2.169 ...
## $ pop.gwth.urban: num [1:1843] 5.682 5.02 4.312 0.812 2.181 ...
## $ gdp.dflt : num [1:1843] 418.02 19.37 28.17 38.17 3.65 ...
## $ gcf : num [1:1843] 30.5 30.8 17.9 18.1 35.3 ...
## $ trade : num [1:1843] 152.5 121.4 66.4 44.9 68.5 ...
## $ gdp.pc : num [1:1843] 557 4081 2525 1010 1425 ...
```

Re-evaluate the `countries1` set.

```
mean(is.na(countries1))

## [1] 0.0547

sum(complete.cases(countries1))

## [1] 937

mean(complete.cases(countries1))

## [1] 0.508
```

On average, each column has 6% missing rate, results in 937 complete data point (i.e. 49%). This can be a sufficient number for the analysis. However, the missing data can induce loss of power due to the reduced sample size, and some other biases depending on which variables is missing.

```
# complete rate of data by regions
countries1 %>%
  mutate(isComplete = complete.cases(.)) %>%
  group_by(reg) %>%
  summarise(complete.rate = mean(isComplete)) %>%
  arrange(desc(complete.rate))

## # A tibble: 7 x 2
##   reg           complete.rate
##   <fct>          <dbl>
## 1 Europe & Central Asia      0.646
## 2 Latin America & Caribbean 0.505
## 3 Middle East & North Africa 0.462
## 4 East Asia & Pacific       0.437
## 5 South Asia                 0.245
## 6 Sub-Saharan Africa         0.192
## 7 North America               0.14
```

Countries from North America, Sub-Saharan Africa, and South Asia have the highest rate of missing data. We suspect that Sub-Saharan Africa, and South Asia are comparably less accessible regions. We also know that Americans don't like filling out forms, so their high rate of missing data is understandable as well.

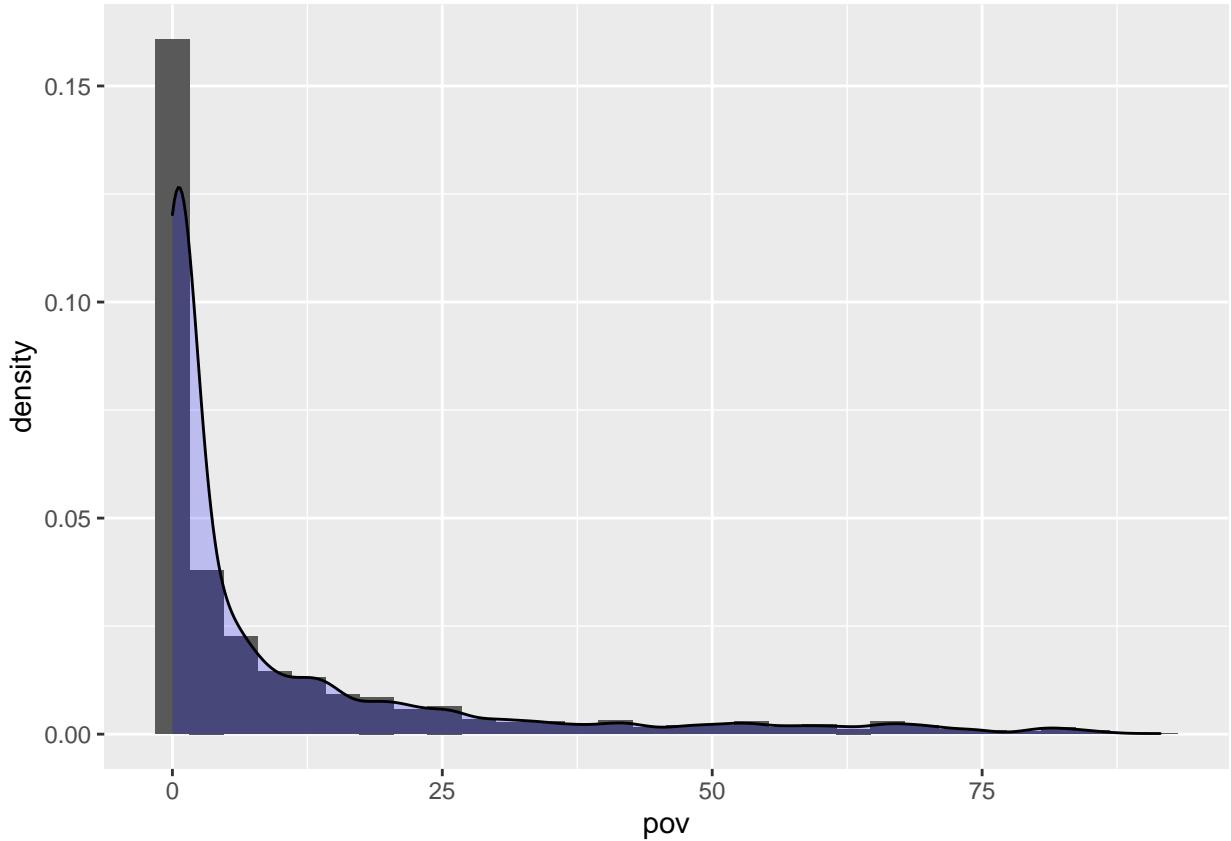
Still, we need to find a way to address this issue. we propose several approaches:

1. **Use complete cases:** Only use the complete cases for the analysis. This is a straightforward approach, but doesn't resolve the bias resulted from the massive loss of data.
2. **Selectively remove variables with high missing rate:** Start with all the variable in the data set, then step-by-step remove insignificant, or frequently missing variables to use as much data as possible.
3. **Forward variable selection:** Start with 0 variable, then selectively add potentially important variable to the data set to achieve optimal model.
4. **Imputation:** The idea is to replace the missing observations on the response or the predictors with artificial values that try to preserve the data set structure. This is a quite complex topic of its own. You can read more from Arel-Bundock and Pelc (2018).

2.3. Descriptive Analytics

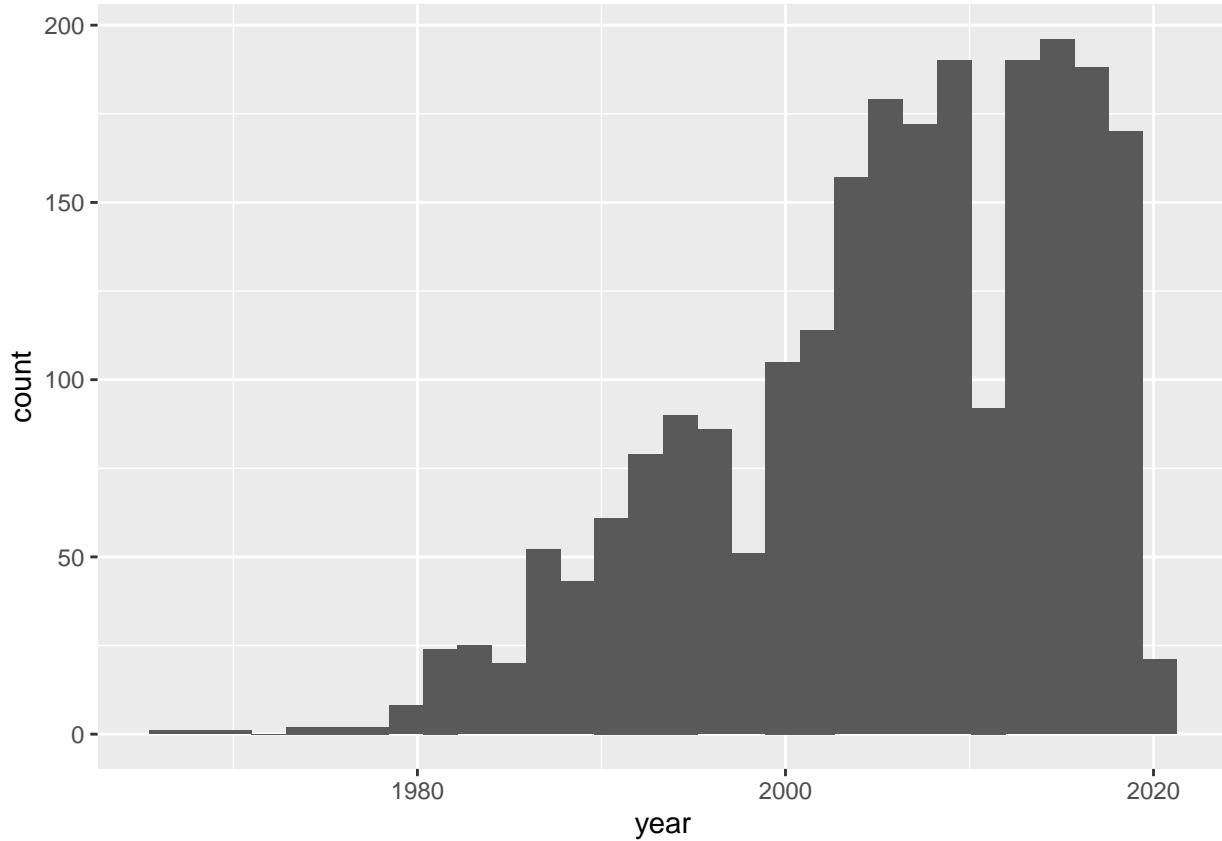
Distribution of the predicted variable pov

```
ggplot(countries, aes(x = pov)) + geom_histogram(aes(y = ..density..),
  bins = 30) + geom_density(alpha = 0.2, fill = "blue")
```



The graph displays a decreasing rate as poverty indicator increasing. This might not be representative of the current state of poverty in the world, but of the number presented in our data. For example, more recent data is likely to be more inclusive than ancient data, when poverty is more prevalent. We should look at data from the same period.

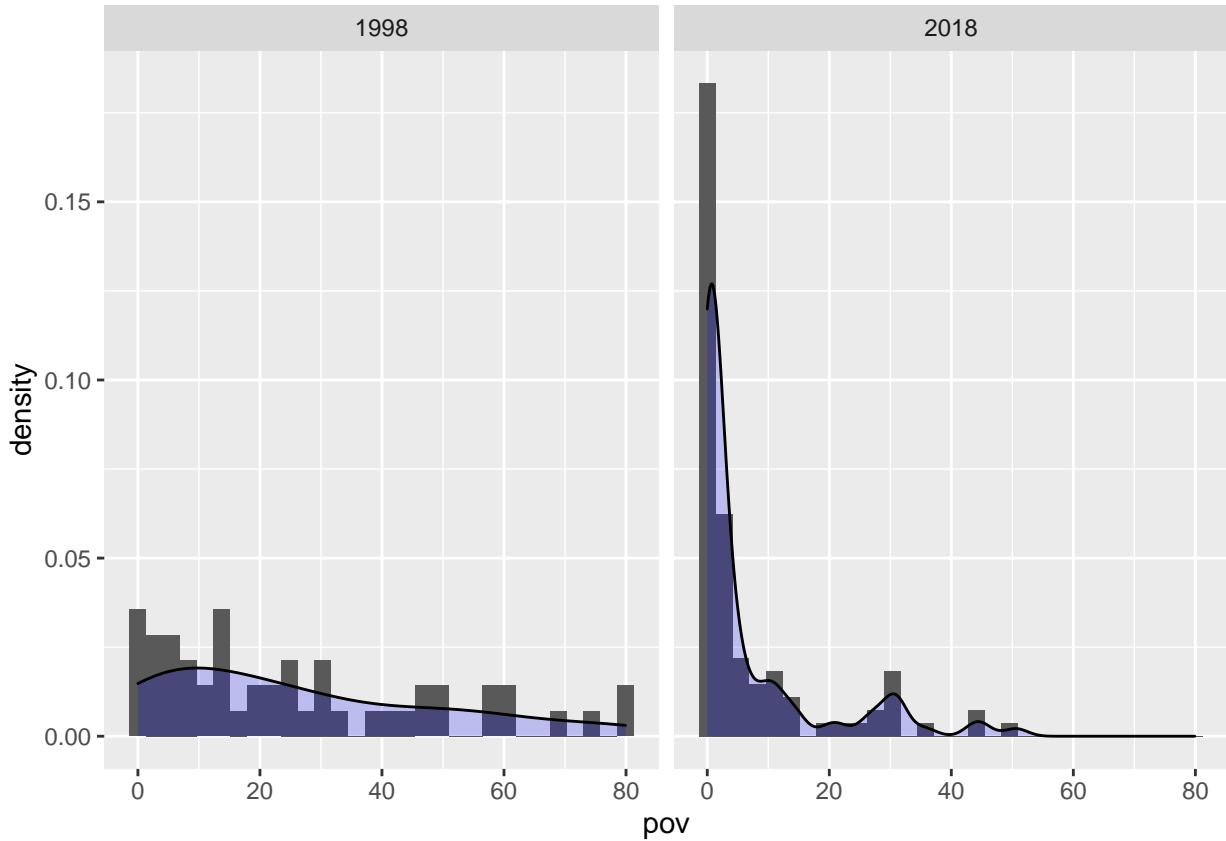
```
# number of data point available from 1967 to  
# 2021  
ggplot(poverty.headcount, aes(x = year)) + geom_histogram(bins = 30)
```



```
# pov data from 1998 and 2018
pov.98.18 <- poverty.headcount %>%
  filter(year == 1998 | year == 2018)
pov.98.18 %>%
  group_by(year) %>%
  summarise(sum = n())

## # A tibble: 2 x 2
##   year     sum
##   <dbl> <int>
## 1 1998     51
## 2 2018     99

ggplot(pov.98.18, aes(x = pov)) + geom_histogram(aes(y = ..density..),
  bins = 30) + geom_density(alpha = 0.2, fill = "blue") +
  facet_grid(cols = vars(year))
```



The graph for 1998 has a much gentler slope, meaning poverty was more popular during that time, as predicted from our intuition. What about the general progress of the world?

```
# Re-import pov and only take special regions
# geographic
geo.regs <- c("EAS", "ECS", "LCN", "MEA", "SAS", "SSF",
             "WLD")
# economics
eco.regs <- c("HIC", "LIC", "LMC", "LMY", "UMC")

pov.reg <- importWDI("../data/poverty.headcount.215dollar.csv",
                      "pov") %>%
  filter(country.code %in% c(geo.regs, eco.regs)) %>%
  mutate(type = ifelse(country.code %in% geo.regs,
                       "Geographic", "Economics"))

## New names:
## Rows: 266 Columns: 67
## -- Column specification
## ----- Delimiter: ","
## (4): Country Name, Country Code, Indicator Name, Indicator Code dbl (50): 1967,
## 1969, 1971, 1974, 1975, 1977, 1978, 1979, 1980, 1981, 1982, ... lgl (13): 1960,
## 1961, 1962, 1963, 1964, 1965, 1966, 1968, 1970, 1972, 1973, ...
## i Use 'spec()' to retrieve the full column specification for this data. i
## Specify the column types or set 'show_col_types = FALSE' to quiet this message.
## * ' -> '...67'
```

```

pov.reg %>%
  distinct(country.code, country.name, type) %>%
  arrange(type)

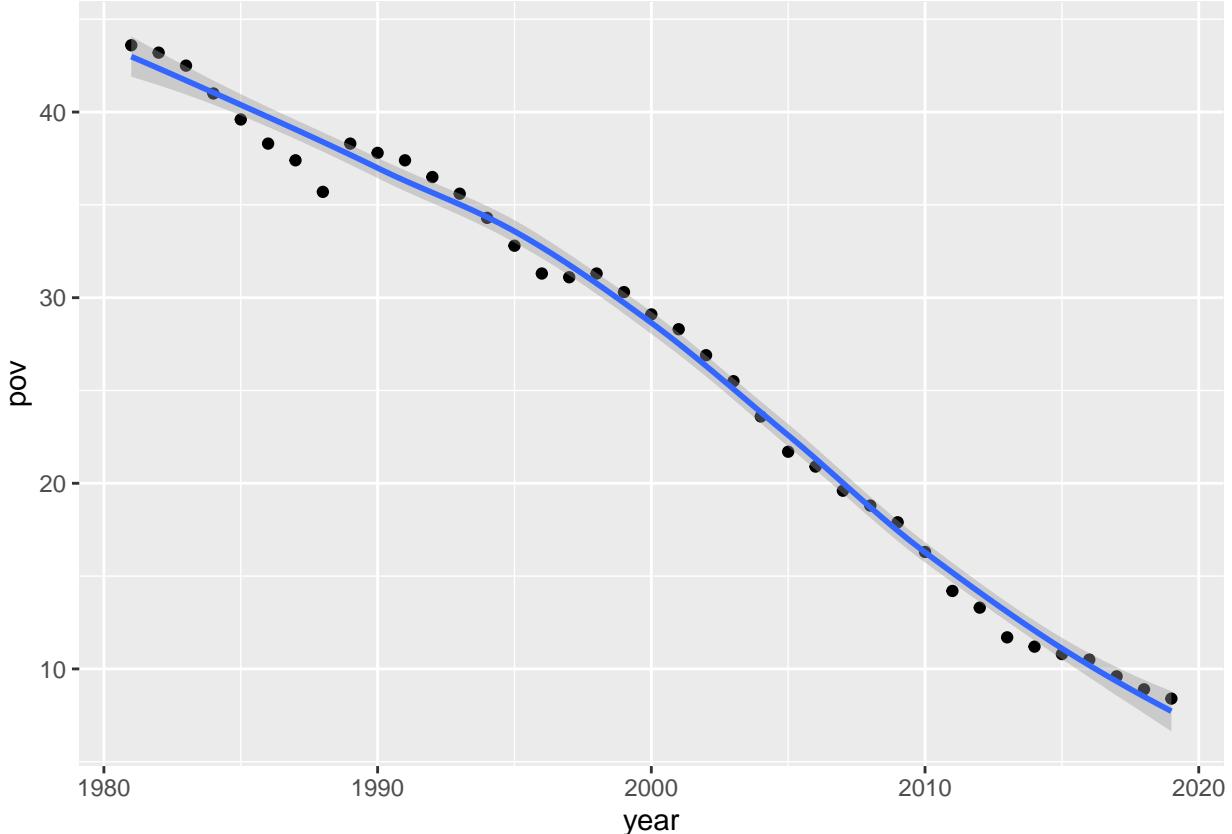
## # A tibble: 12 x 3
##   country.code country.name      type
##   <fct>        <fct>          <chr>
## 1 HIC          High income    Economics
## 2 LIC          Low income     Economics
## 3 LMC          Lower middle income Economics
## 4 LMY          Low & middle income Economics
## 5 UMC          Upper middle income Economics
## 6 EAS          East Asia & Pacific Geographic
## 7 ECS          Europe & Central Asia Geographic
## 8 LCN          Latin America & Caribbean Geographic
## 9 MEA          Middle East & North Africa Geographic
## 10 SAS         South Asia      Geographic
## 11 SSF         Sub-Saharan Africa Geographic
## 12 WLD         World          Geographic

```

```

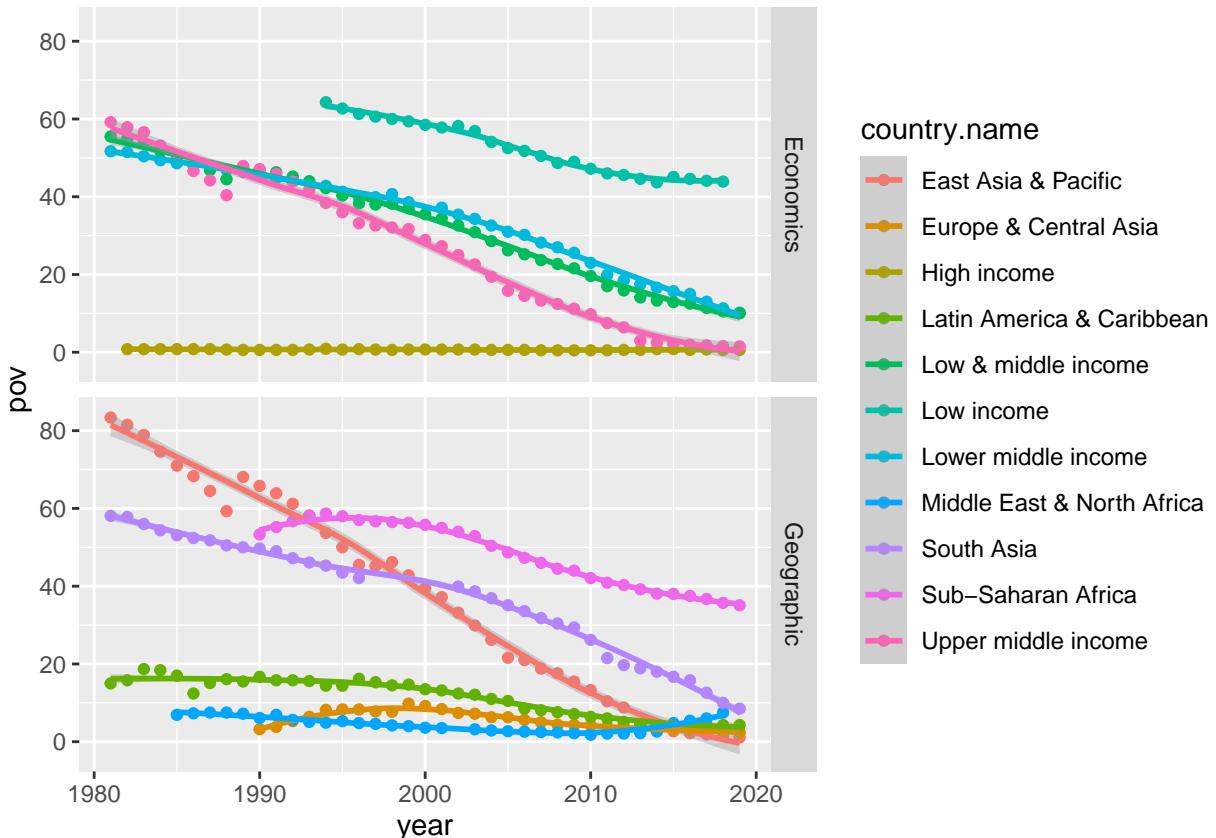
# World
ggplot(pov.reg %>%
  filter(country.code == "WLD"), aes(x = year, y = pov)) +
  geom_point() + geom_smooth(formula = y ~ x, method = loess)

```



An overall very steady decrease of poverty. How about each region?

```
ggplot(pov.reg %>%
  filter(country.code != "WLD"), aes(x = year, y = pov,
  color = country.name)) + geom_point() + geom_smooth(formula = y ~
  x, method = loess) + facet_grid(rows = vars(type))
```



There's a general steady, but distinct decline of poverty over time in each type region of respective type. Latin America & Caribbean, Europe & Central Asia, Middle East & North Africa, and High Income group has a more gradual decline as they are not very poor to begin with.

Among the income groups, Low & Middle Income, Lower & Middle Income, and Upper Middle Income have quite similar in term of poverty indicator and slope over the year. While these values vary greatly among different geographical regions.

Let's see some important statistics

```
stats <- poverty.headcount %>%
  summarise(count = n(), skewness = skewness(pov),
  kurtosis = kurtosis(pov), std.deviation = sd(pov))

kable(stats)
```

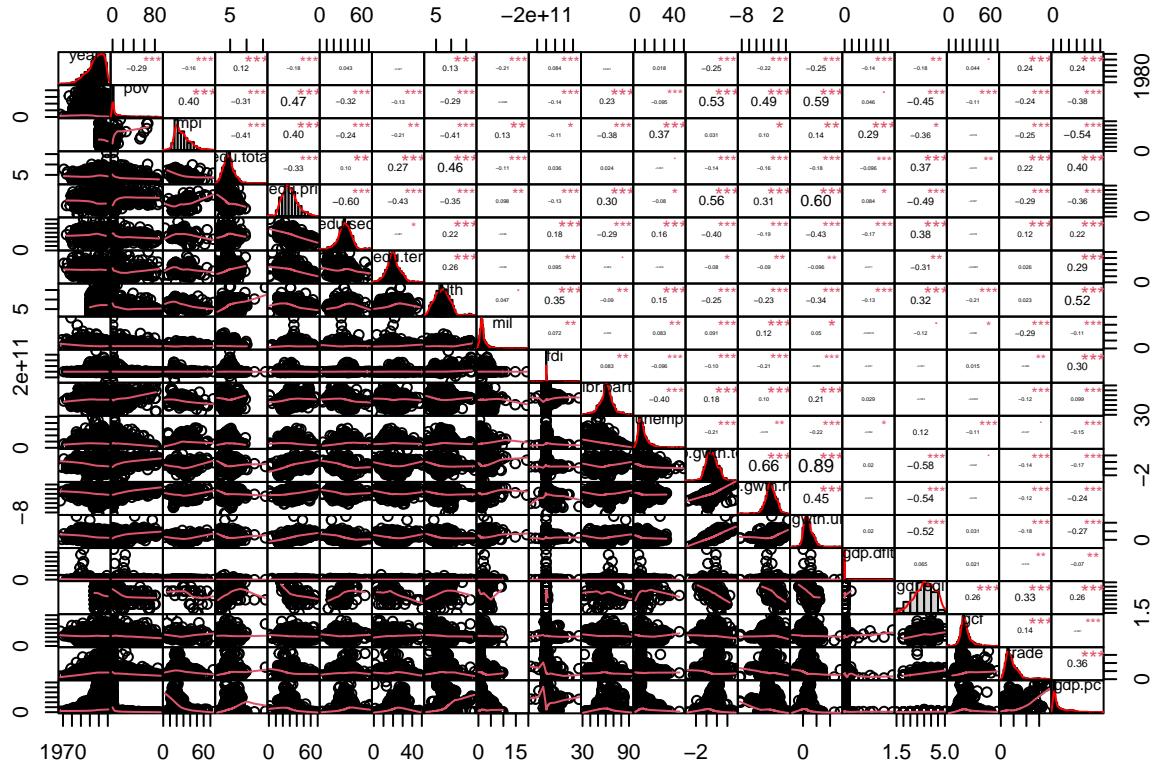
count	skewness	kurtosis	std.deviation
2322	1.6	1.66	19.5

2.4. Assumption Check

We can conduct some preliminary checks on linearity and correlation of predictors to have a better picture of the data. [Checklist 1. Linear relationship:](#)

Use correlation matrix to check linearity

```
# select only numeric data
countries.num <- countries %>%
  select(where(is.numeric))
chart.Correlation(countries.num, histogram = TRUE,
  pch = 19)
```



Which is utterly intelligible, but a majority of the fitted lines are linear at first glance. We should render separate graphs for the relationship between `pov` & other variables.

```
# lengthen data table to variable-value pairs,
# with pov as predicted variable and others as
# predictive
countries.num2 <- countries.num %>%
  pivot_longer(cols = 3:ncol(.), names_to = "variable",
               values_to = "value") %>%
```

```

filter(!is.na(value) & !is.na(pov))

drawGraph <- function(indvar, data) {
  ggplot(data %>%
    filter(variable == indvar), aes(x = value,
    y = pov)) + geom_point() + geom_smooth(method = lm) +
  labs(title = paste("Relationship pov ~", indvar),
       x = indvar, y = "pov")
}

ivs <- distinct(countries.num2, variable)[[1]]

# for (indvar in ivs) { print(
# ggplot(countries.num2 %>% filter(variable ==
# indvar), aes(x = value, y = pov)) +
# geom_point() + geom_smooth(formula = y~x,
# method = lm) + labs(title = paste('Relationship
# pov ~', indvar), x = indvar, y = 'pov') ) }

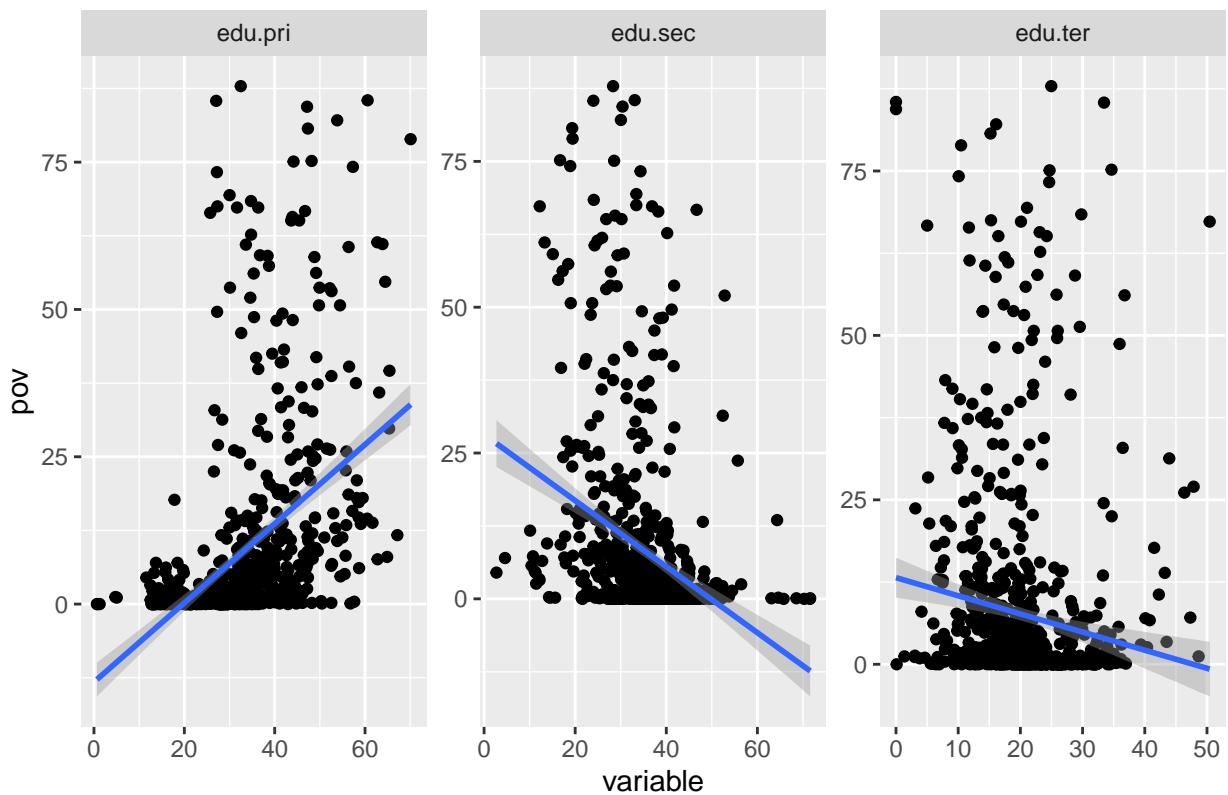
p <- ggplot(countries.num2, aes(x = value, y = pov)) +
  geom_point() + geom_smooth(formula = y ~ x, method = lm) +
  labs(title = paste("pov ~ variable"), x = "variable",
       y = "pov") + facet_wrap_paginate(~variable,
       ncol = 3, nrow = 1, scales = "free")

requiredPages <- n_pages(p)

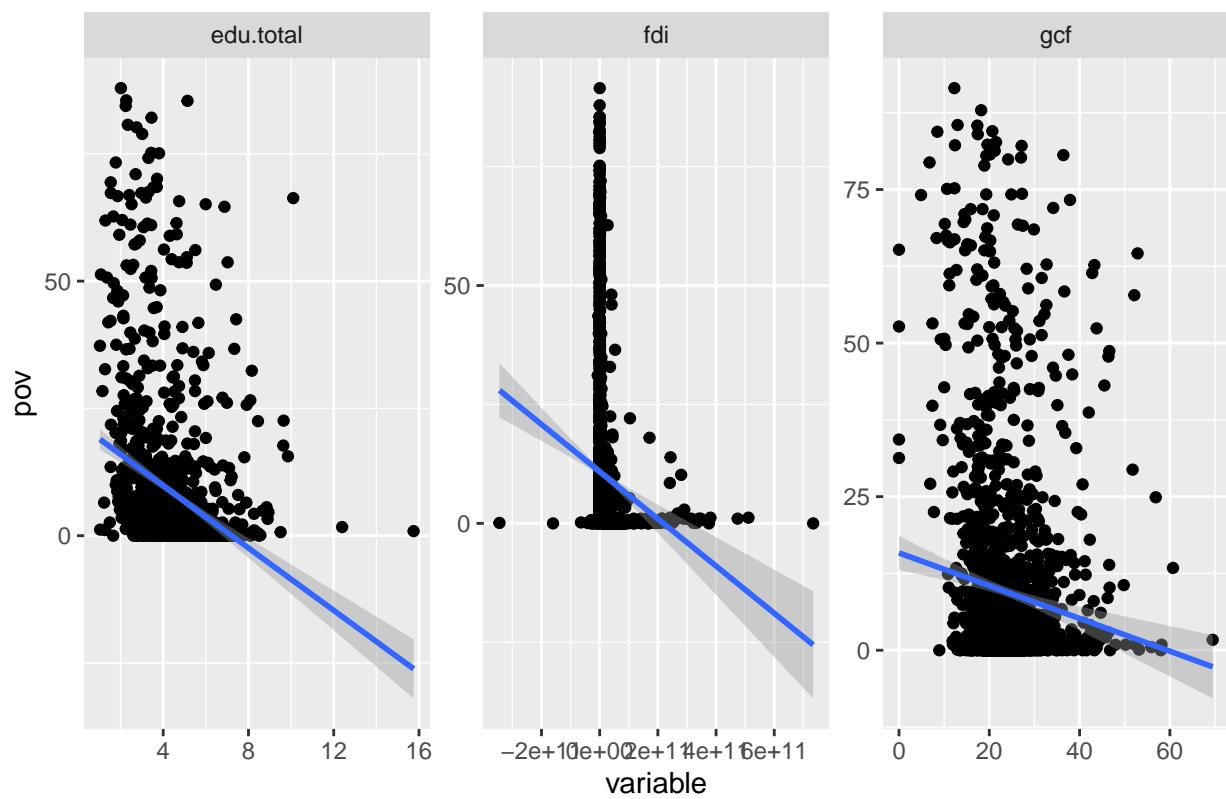
for (i in 1:requiredPages) {
  ggplot(countries.num2, aes(x = value, y = pov)) +
    geom_point() + geom_smooth(formula = y ~ x,
    method = lm) + labs(title = paste("pov ~ variable"),
    x = "variable", y = "pov") + facet_wrap_paginate(~variable,
    ncol = 3, nrow = 1, scales = "free", page = i) ->
  p
  print(p)
}

```

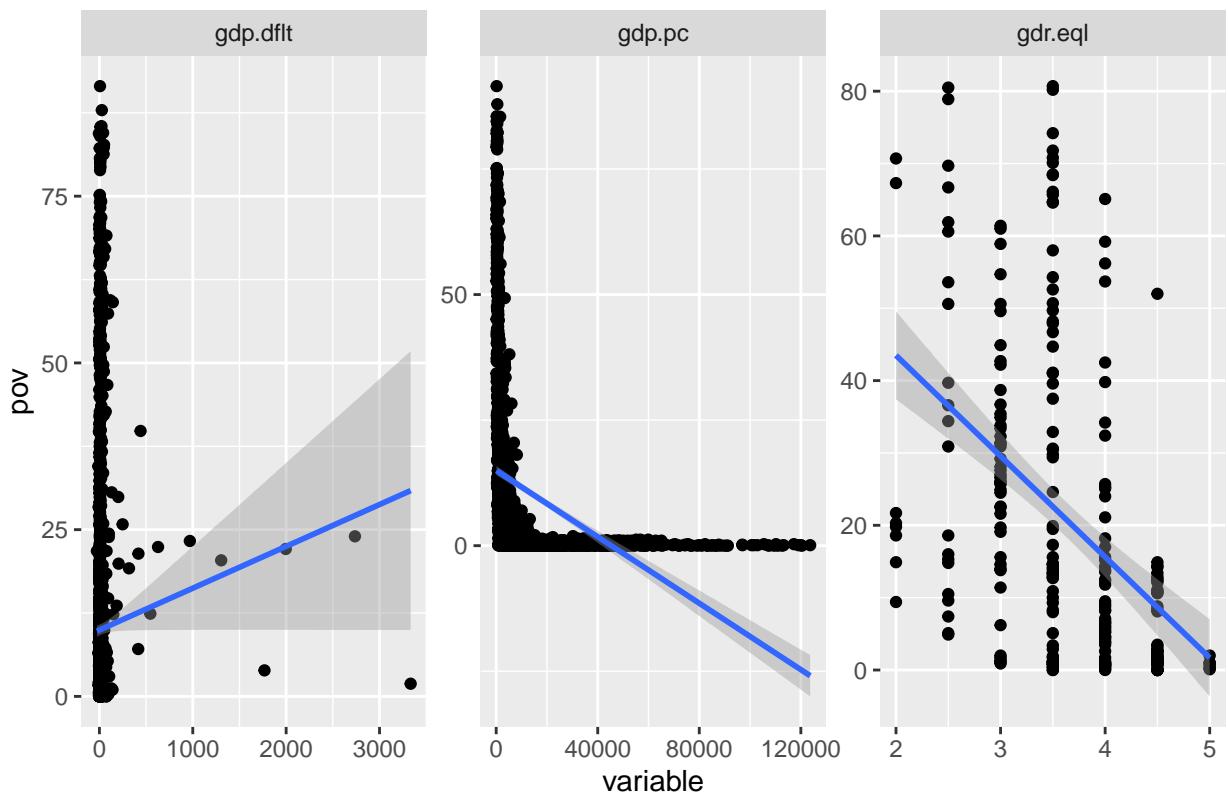
pov ~ variable



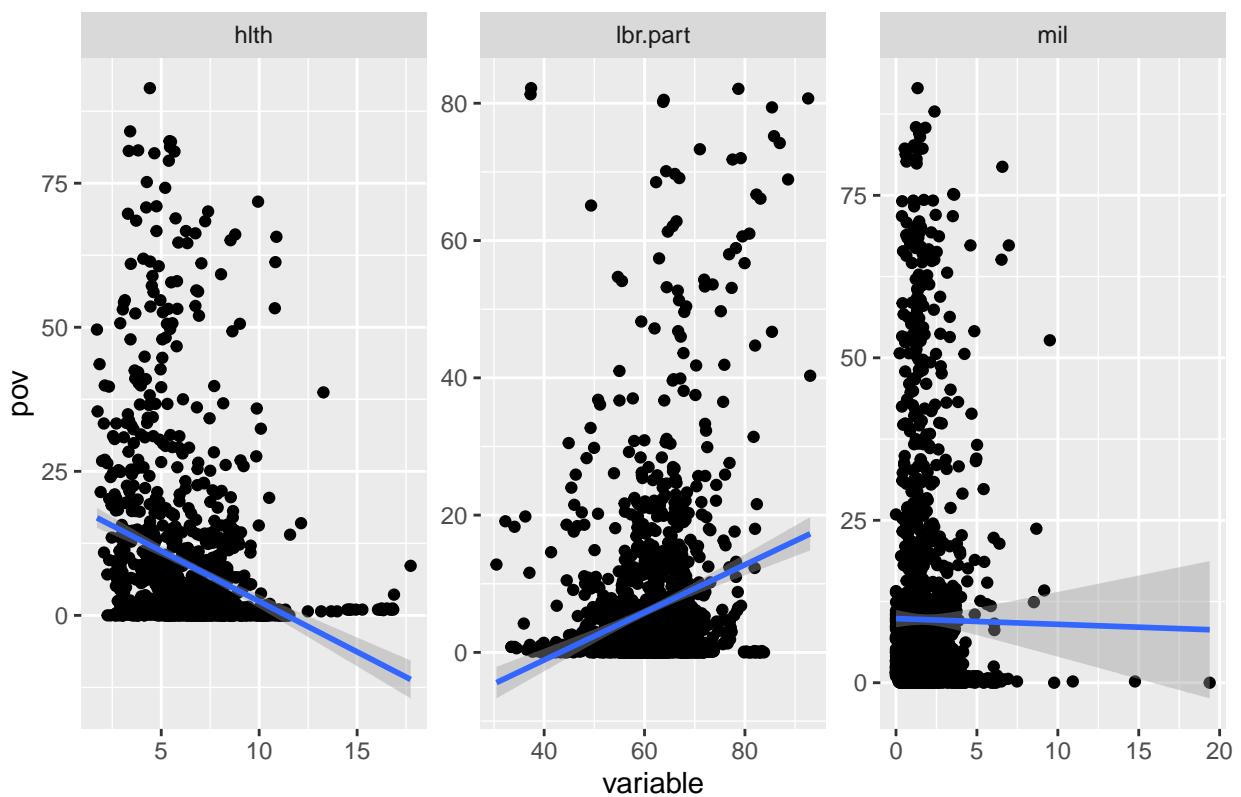
pov ~ variable



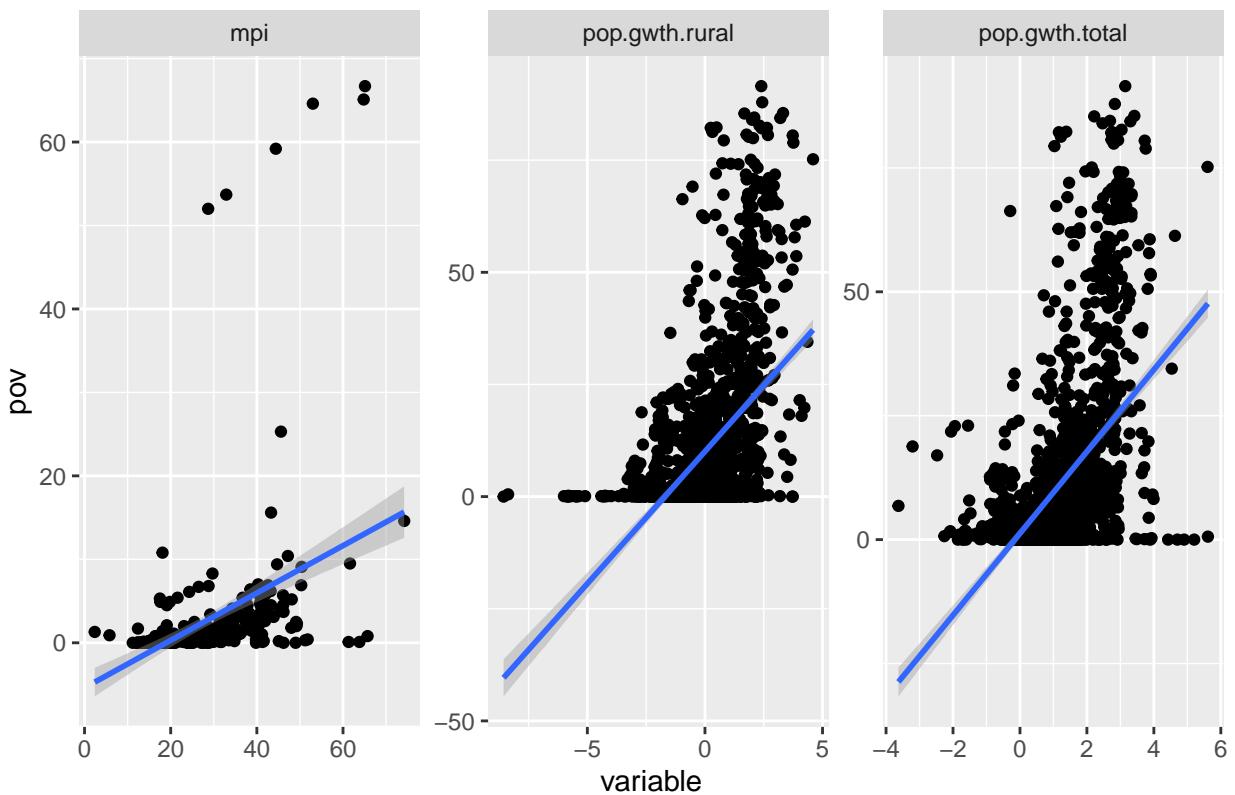
pov ~ variable



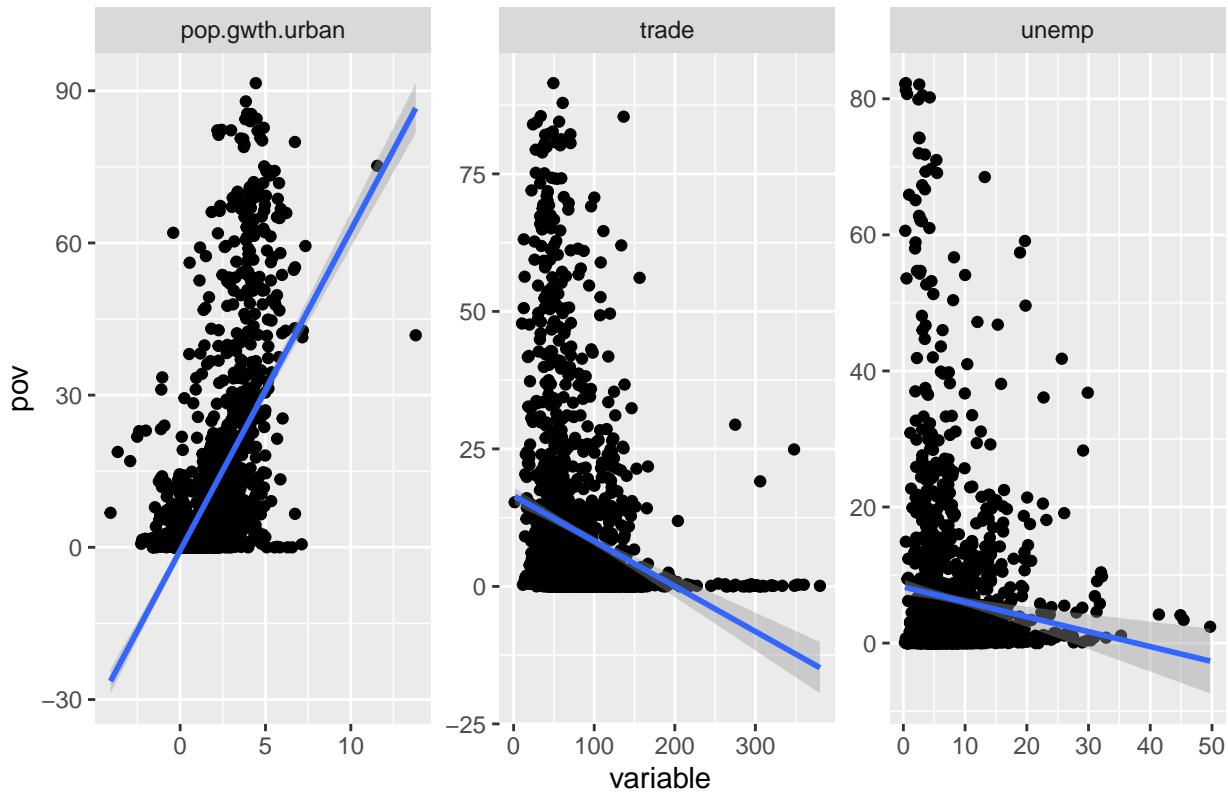
pov ~ variable



pov ~ variable



pov ~ variable

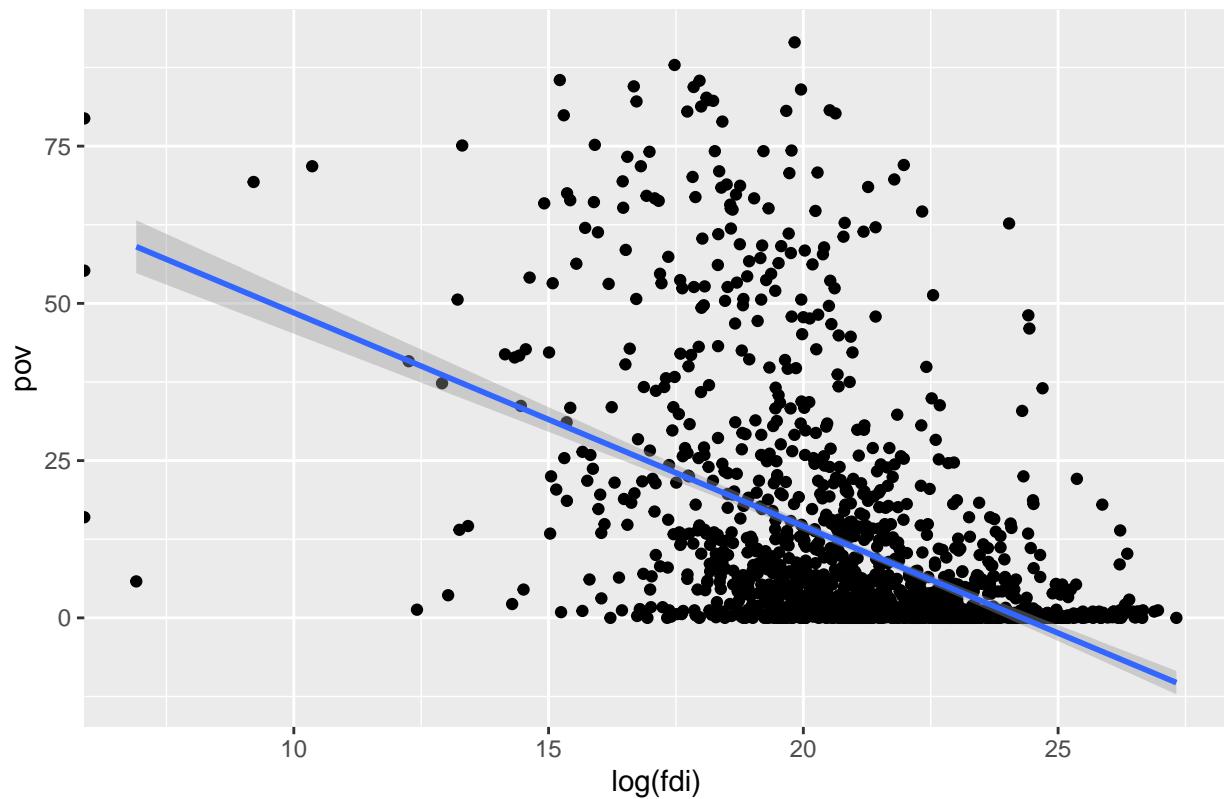


Most variables display linear relationship with some exceptions, which are more appropriate to assume a logarithmic relationship. Gender equality should be viewed as a categorical data.

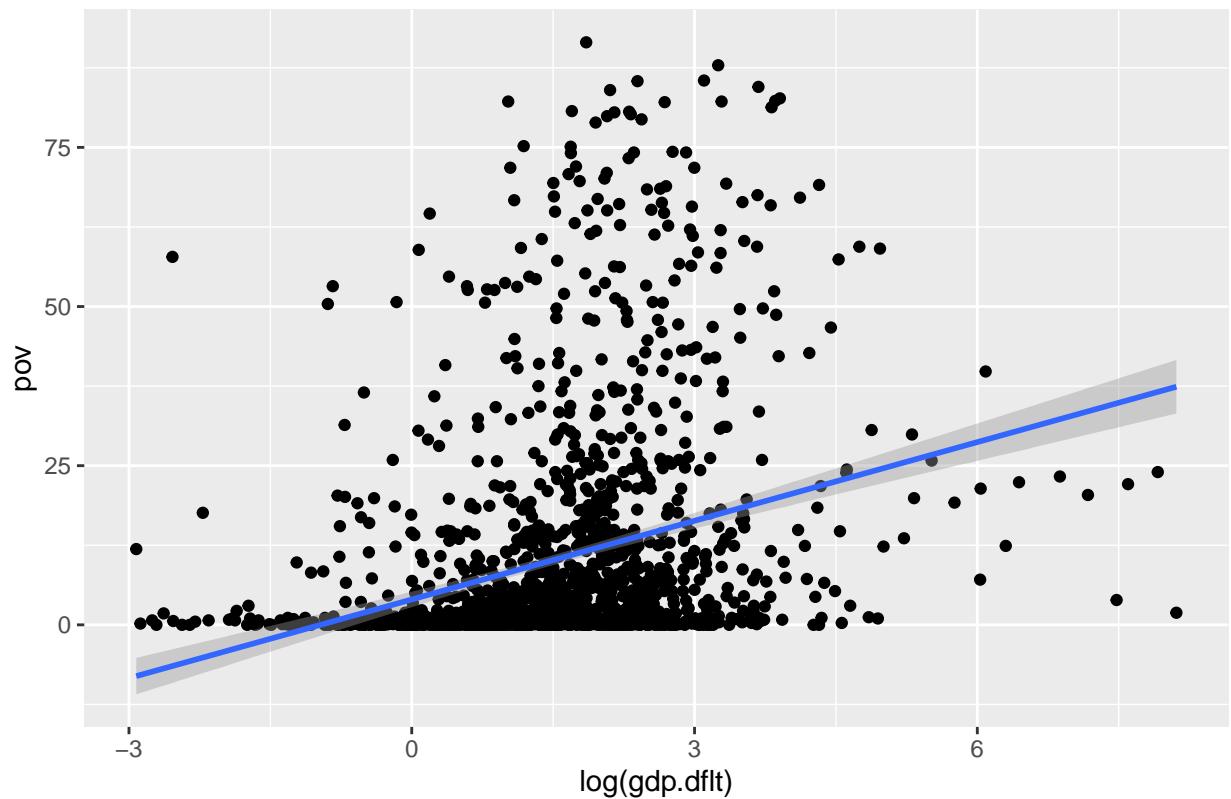
```
logIvs <- c("fdi", "gdp.dflt", "gdp.pc")

for (indvar in logIvs) {
  print(ggplot(countries.num2 %>%
    filter(variable == indvar), aes(x = log(value),
    y = pov)) + geom_point() + geom_smooth(formula = y ~
    x, method = lm) + labs(title = paste0("pov ~ log(", indvar, ")"),
    x = paste0("log(", indvar, ")"),
    y = "pov"))}
```

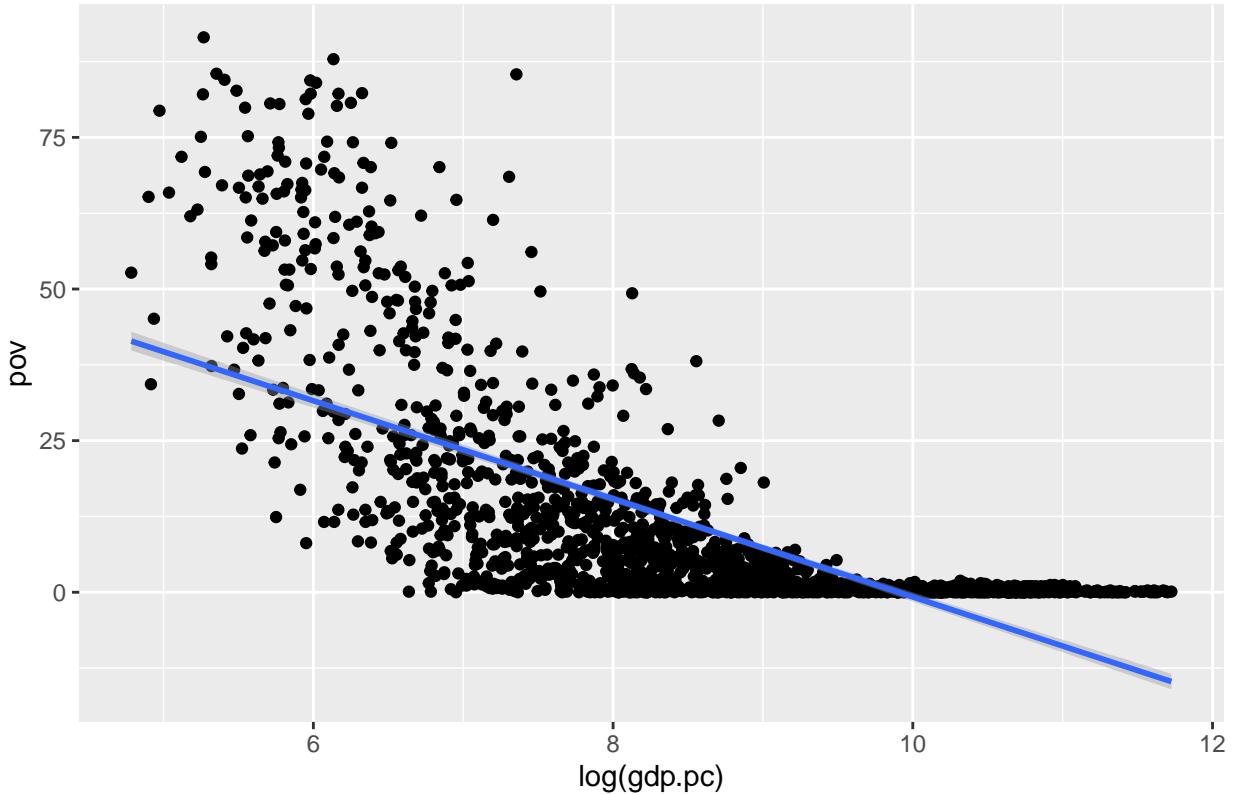
$pov \sim \log(fdi)$



$\text{pov} \sim \log(\text{gdp.dflt})$



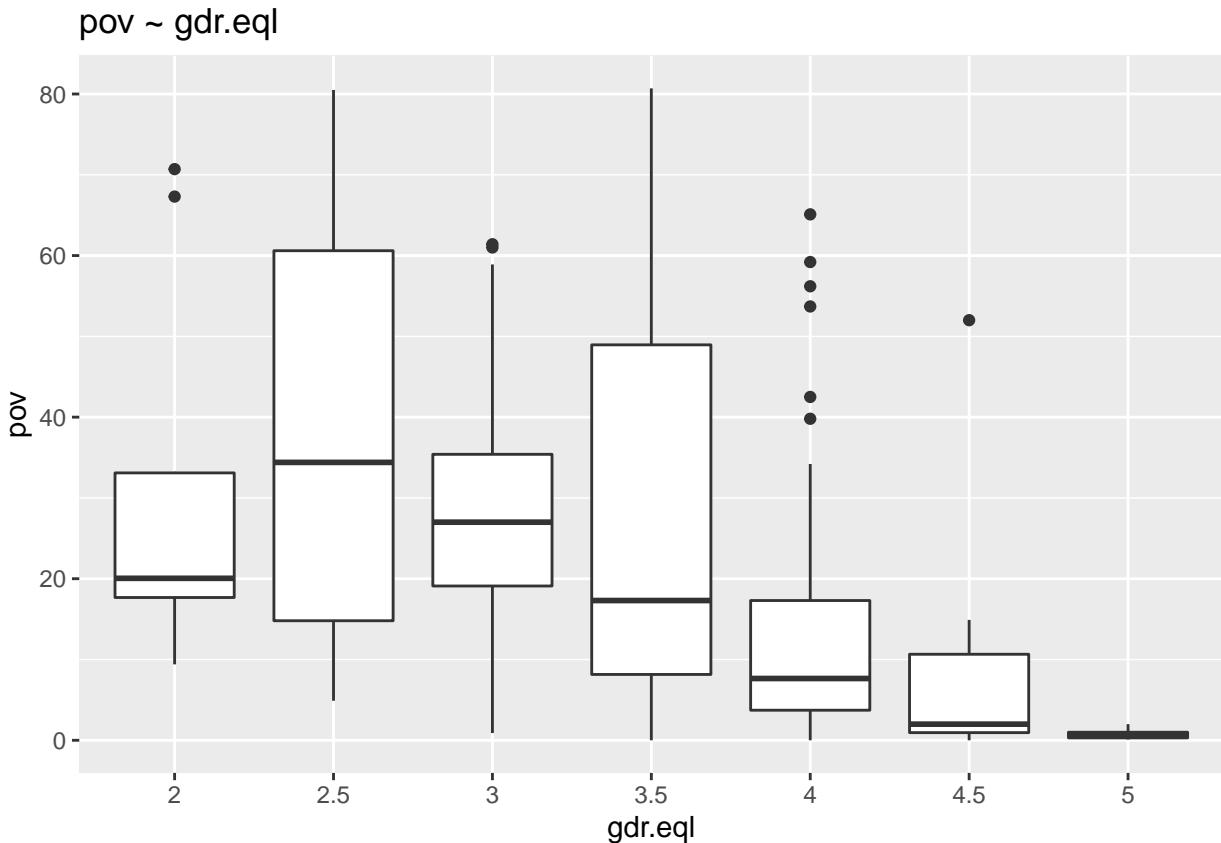
$pov \sim \log(gdp.pc)$



The relationship are more linear after the transformation.

```
gdr.eql <- countries.num2 %>%
  filter(variable == "gdr.eql") %>%
  mutate(value = factor(value))

ggplot(gdr.eql, aes(x = value, y = pov)) + geom_boxplot() +
  labs(title = paste("pov ~ gdr.eql"), x = "gdr.eql",
       y = "pov")
```



There's a significant correlation between gender equality and poverty.
Correlation coefficients between pov and predictors.

```
# transform some variables
countries.num3 <- countries.num %>%
  mutate(lfdi = log(fdi), lgdp.dflt = log(gdp.dflt),
    lgdp.pc = log(gdp.pc)) %>%
  mutate(lfdi = ifelse(is.nan(lfdi) | lfdi == -Inf,
    NA, lfdi))

# name of the independent variables
cn <- colnames(countries.num3)[-c(1, 2)]

# correlation of independent variable and pov
corWithPov <- function(indevar, data) {
  cor(data[[indevar]], data[["pov"]], use = "complete.obs")
}

cor.pov <- sapply(cn, corWithPov, data = countries.num3)
kable(cor.pov)
```

	x
mpi	0.404
edu.total	-0.310
edu.pri	0.471

	x
edu.sec	-0.317
edu.ter	-0.128
hlth	-0.294
mil	-0.007
fdi	-0.144
lbr.part	0.232
unemp	-0.095
pop.gwth.total	0.529
pop.gwth.rural	0.494
pop.gwth.urban	0.587
gdp.dflt	0.046
gdr.eql	-0.451
gcf	-0.111
trade	-0.239
gdp.pc	-0.377
lfdi	-0.488
lgdp.dflt	0.299
lgdp.pc	-0.705

edu.ter, mil, fdi, lbr.part, unemp, gdp.dflt, gcf have negligible correlation with pov. lgdp.pc, lfdi, and lgdp.dflt have stronger linear relationship with pov than their un-transformed counterparts.

```
countries1 <- countries1 %>%
  mutate(lfdi = log(fdi), lgdp.dflt = log(gdp.dflt),
        lgdp.pc = log(gdp.pc)) %>%
  mutate(lfdi = ifelse(is.nan(lfdi) | lfdi == -Inf,
                       NA, lfdi))
```

2. Predictors are independent There should be no correlation/multicollinearity between each pair of predictors.

```
countries.arr <- simplify2array(countries.num %>%
  select(-c("year", "pov")))

cor mtx <- rcorr(countries.arr, type = "spearman")

round(cor mtx$r, 2)
```

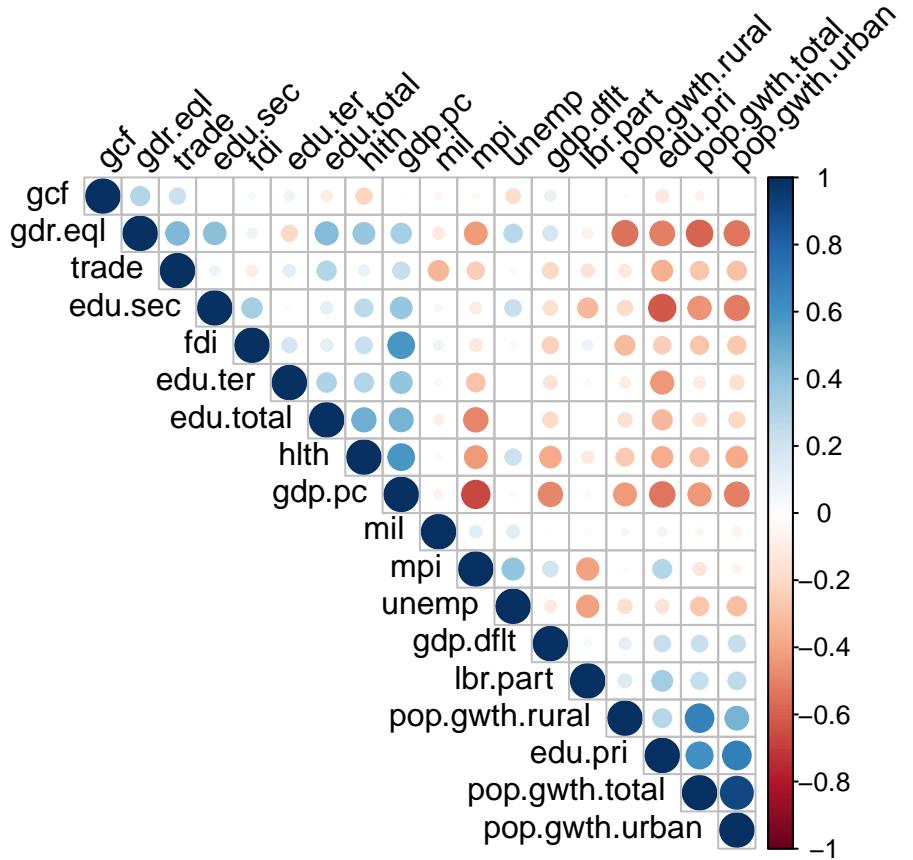
##	mpi	edu.total	edu.pri	edu.sec	edu.ter	hlth	mil	fdi
## mpi	1.00	-0.49	0.29	-0.10	-0.28	-0.43	0.13	-0.13
## edu.total	-0.49	1.00	-0.33	0.12	0.31	0.48	-0.08	0.12
## edu.pri	0.29	-0.33	1.00	-0.62	-0.44	-0.36	0.06	-0.24
## edu.sec	-0.10	0.12	-0.62	1.00	0.02	0.26	0.04	0.34
## edu.ter	-0.28	0.31	-0.44	0.02	1.00	0.30	0.04	0.19
## hlth	-0.43	0.48	-0.36	0.26	0.30	1.00	-0.03	0.22
## mil	0.13	-0.08	0.06	0.04	0.04	-0.03	1.00	0.07
## fdi	-0.13	0.12	-0.24	0.34	0.19	0.22	0.07	1.00
## lbr.part	-0.41	0.01	0.35	-0.33	-0.04	-0.11	0.02	0.08
## unemp	0.40	0.00	-0.14	0.22	0.00	0.22	0.14	-0.04
## pop.gwth.total	-0.14	-0.14	0.60	-0.45	-0.11	-0.29	-0.05	-0.27
## pop.gwth.rural	-0.02	-0.16	0.29	-0.18	-0.10	-0.26	0.04	-0.32

```

## pop.gwth.urban -0.06    -0.21    0.68   -0.51   -0.16  -0.37  -0.07  -0.27
## gdp.dflt      0.19     -0.19    0.22   -0.17   -0.15  -0.38   0.01  -0.22
## gdr.eql      -0.43     0.44   -0.50    0.41   -0.21   0.39  -0.11   0.08
## gcf        -0.04    -0.10   -0.12    0.01    0.08  -0.21  -0.05   0.04
## trade       -0.24     0.30   -0.35    0.08    0.12   0.09  -0.34  -0.09
## gdp.pc       -0.66     0.47   -0.53    0.38    0.39   0.59  -0.06   0.59
##          lbr.part unemp pop.gwth.total pop.gwth.rural pop.gwth.urban
## mpi           -0.41    0.40    -0.14   -0.02   -0.06
## edu.total     0.01    0.00    -0.14   -0.16   -0.21
## edu.pri       0.35   -0.14    0.60    0.29    0.68
## edu.sec      -0.33    0.22   -0.45   -0.18   -0.51
## edu.ter      -0.04    0.00   -0.11   -0.10   -0.16
## hlth         -0.11    0.22   -0.29   -0.26   -0.37
## mil          0.02    0.14   -0.05    0.04   -0.07
## fdi          0.08   -0.04   -0.27   -0.32   -0.27
## lbr.part      1.00   -0.41    0.24    0.15    0.25
## unemp        -0.41    1.00   -0.27   -0.16   -0.30
## pop.gwth.total 0.24   -0.27    1.00    0.68    0.91
## pop.gwth.rural 0.15   -0.16    0.68    1.00    0.46
## pop.gwth.urban 0.25   -0.30    0.91    0.46    1.00
## gdp.dflt      0.05   -0.11    0.22    0.11    0.23
## gdr.eql      -0.08    0.27   -0.58   -0.55   -0.53
## gcf          0.01   -0.17   -0.06    0.01    0.00
## trade        -0.14   -0.03   -0.28   -0.13   -0.29
## gdp.pc        -0.04    0.02   -0.44   -0.43   -0.50
##          gdp.dflt gdr.eql   gcf trade gdp.pc
## mpi           0.19   -0.43  -0.04  -0.24   -0.66
## edu.total     -0.19    0.44  -0.10  0.30    0.47
## edu.pri       0.22   -0.50  -0.12  -0.35   -0.53
## edu.sec      -0.17    0.41   0.01   0.08    0.38
## edu.ter      -0.15   -0.21   0.08   0.12    0.39
## hlth         -0.38    0.39  -0.21   0.09    0.59
## mil          0.01   -0.11  -0.05  -0.34   -0.06
## fdi          -0.22    0.08   0.04  -0.09    0.59
## lbr.part      0.05   -0.08   0.01  -0.14   -0.04
## unemp        -0.11    0.27  -0.17  -0.03    0.02
## pop.gwth.total 0.22   -0.58  -0.06  -0.28   -0.44
## pop.gwth.rural 0.11   -0.55  0.01  -0.13   -0.43
## pop.gwth.urban 0.23   -0.53  0.00  -0.29   -0.50
## gdp.dflt      1.00    0.18   0.09  -0.20   -0.49
## gdr.eql      0.18    1.00   0.30   0.45    0.33
## gcf          0.09    0.30   1.00   0.21   -0.01
## trade        -0.20    0.45   0.21   1.00    0.24
## gdp.pc       -0.49    0.33  -0.01   0.24    1.00

corrplot(corr mtx$r, type = "upper", order = "hclust",
         tl.col = "black", tl.srt = 45)

```



Some significant correlations can be found between `gdr.eql` with `edu.pri`, `pop.gwth.total`, and `fdi` with `mil`, etc. We expect these variables to be eliminated in the VIF test. This might be a good property for imputation (3.2.4)

2.5. Data Source

- `poverty.headcount`
- `mpi`
- `education.expenditure.primary`
- `education.expenditure.secondary`
- `education.expenditure.tertiary`
- `education.expenditure.total`
- `health.expenditure`
- `military.expenditure`
- `fdi`
- `unemployment.rate`
- `labour.force.participation`
- `gender.equality`
- `population.growth`
- `urban.population.growth`
- `rural.population.growth`
- `gdp.deflator`
- `gross capital formation`
- `trade`
- `region.class`
- `income.class`

- gross.capital.formation

3. Model Selection and Interpretation

We can now conduct linear regressions following the approaches mentioned above to address missing values issues.

3.1. Use Complete Cases

```
needs("tidyverse", "ggpubr", "gridExtra", "e1071",
  "ggplot2", "dplyr", "tidyr", "readxl", "corrplot",
  "car", "DescTools", "glmnet")
```

First we will write the function “eval_results(true,predict) to evaluate the models

```
eval_results <- function(true, predict) {
  actual <- data.matrix(true)
  SSE <- sum((predict - actual)^2)
  SST <- sum((actual - mean(actual))^2)
  R_square <- 1 - SSE/SST
  data.frame(MSE = MSE(predict, true), MAE = MAE(predict,
    true), RMSE = RMSE(predict, true), Rsquare = R_square)
}
```

First we read the csv file for the countries containing the dataset. We will then put the variables we need in to subset1.

```
df <- read.csv("../data/countries1.csv")
head(df)
```

```
##   X country.code country.name year pov income          reg edu.total
## 1 1      AGO      Angola 2000 21.4     L Sub-Saharan Africa  2.61
## 2 2      AGO      Angola 2008 14.6    LM Sub-Saharan Africa  2.69
## 3 3      AGO      Angola 2018 31.1    LM Sub-Saharan Africa  2.04
## 4 4      ALB      Albania 1996  0.5    LM Europe & Central Asia 3.08
## 5 5      ALB      Albania 2002  1.1    LM Europe & Central Asia 3.12
## 6 6      ALB      Albania 2005  0.6    LM Europe & Central Asia 3.28
##   hlth mil      fdi lbr.part unemp pop.gwth.total pop.gwth.rural
## 1 1.91 6.39 8.79e+08      NA     NA      3.277      0.921
## 2 3.32 3.57 1.68e+09      NA     NA      3.711      1.910
## 3 2.54 1.87 -6.46e+09     NA     NA      3.276      1.338
## 4  NA 1.38 9.01e+07    38.8  12.3     -0.622     -1.546
## 5 6.91 1.32 1.35e+08   59.6  15.8     -0.300     -2.169
## 6 6.34 1.35 2.62e+08   34.5  14.1     -0.512     -2.519
##   pop.gwth.urban gdp.dflt  gcf trade gdp.pc lfdi lgdp.dflt lgdp.pc
## 1      5.682  418.02 30.5 152.5    557 20.6     6.04    6.32
## 2      5.020  19.37 30.8 121.4   4081 21.2     2.96    8.31
## 3      4.312  28.17 17.9  66.4   2525  NA     3.34    7.83
## 4      0.812  38.17 18.1  44.9   1010 18.3     3.64    6.92
## 5      2.181  3.65 35.3  68.5   1425 18.7     1.29    7.26
## 6      1.826  3.31 36.9  70.9   2674 19.4     1.20    7.89
```

```

subset1 <- df %>%
  select("pov", "country.code", "year", "income",
         "edu.total", "hlth", "mil", "fdi", "lbr.part",
         "unemp", "pop.gwth.total", "pop.gwth.rural",
         "gdp.dflt", "gcf", "trade", "gdp.pc", "lfdi",
         "lgdp.dflt", "lgdp.pc")
head(subset1)

##   pov country.code year income edu.total hlth      mil      fdi lbr.part unemp
## 1 21.4        AGO 2000     L    2.61 1.91 6.39 8.79e+08      NA     NA
## 2 14.6        AGO 2008    LM    2.69 3.32 3.57 1.68e+09      NA     NA
## 3 31.1        AGO 2018    LM    2.04 2.54 1.87 -6.46e+09      NA     NA
## 4  0.5        ALB 1996    LM    3.08   NA 1.38 9.01e+07    38.8 12.3
## 5  1.1        ALB 2002    LM    3.12 6.91 1.32 1.35e+08    59.6 15.8
## 6  0.6        ALB 2005    LM    3.28 6.34 1.35 2.62e+08    34.5 14.1
##   pop.gwth.total pop.gwth.rural gdp.dflt      gcf      trade      gdp.pc      lfdi lgdp.dflt
## 1        3.277       0.921 418.02 30.5 152.5      557 20.6      6.04
## 2        3.711       1.910 19.37 30.8 121.4     4081 21.2      2.96
## 3        3.276       1.338 28.17 17.9  66.4     2525  NA      3.34
## 4       -0.622      -1.546 38.17 18.1  44.9     1010 18.3      3.64
## 5       -0.300      -2.169  3.65 35.3  68.5     1425 18.7      1.29
## 6       -0.512      -2.519  3.31 36.9  70.9     2674 19.4      1.20
##   lgdp.pc
## 1  6.32
## 2  8.31
## 3  7.83
## 4  6.92
## 5  7.26
## 6  7.89

```

Next we will build a SLR model. We will put 80% of dataset to train and the rest to test.

```

set.seed(10)
index <- sort(sample(x = nrow(subset1), size = nrow(subset1) *
  0.8))
train <- df[index, ]
test <- df[-index, ]

```

The model output will be placed in lm.slr and can be viewed in summary(lm.slr).

```

lm.slr <- lm(pov ~ country.code + year + income + edu.total +
  hlth + mil + fdi + lbr.part + unemp + pop.gwth.total +
  pop.gwth.rural + gdp.dflt + gcf + trade + gdp.pc +
  lfdi + lgdp.dflt + lgdp.pc, data = train)
summary(lm.slr)

##
## Call:
## lm(formula = pov ~ country.code + year + income + edu.total +
##     hlth + mil + fdi + lbr.part + unemp + pop.gwth.total + pop.gwth.rural +
##     gdp.dflt + gcf + trade + gdp.pc + lfdi + lgdp.dflt + lgdp.pc,
##     data = train)

```

```

##
## Residuals:
##      Min     1Q Median     3Q    Max
## -8.820 -0.731  0.000  0.754 14.036
##
## Coefficients:
##                               Estimate Std. Error t value Pr(>|t|)
## (Intercept)           4.01e+02   6.16e+01   6.51  1.7e-10 ***
## country.codeARG       2.20e+00   1.42e+00   1.56  0.12022
## country.codeARM      -7.48e-01   1.48e+00  -0.50  0.61434
## country.codeAUS       1.26e+00   2.09e+00   0.60  0.54581
## country.codeAUT      2.12e+00   1.94e+00   1.09  0.27557
## country.codeAZE      -8.72e+00   2.52e+00  -3.46  0.00059 ***
## country.codeBEL       4.25e+00   2.07e+00   2.06  0.04025 *
## country.codeBEN      1.36e+01   2.65e+00   5.13  3.9e-07 ***
## country.codeBFA      2.26e+01   2.76e+00   8.20  1.7e-15 ***
## country.codeBGD      1.27e+01   2.05e+00   6.21  1.0e-09 ***
## country.codeBGR      4.15e+00   1.40e+00   2.96  0.00316 **
## country.codeBLR      3.13e+00   1.39e+00   2.24  0.02521 *
## country.codeBOL      3.63e+00   1.87e+00   1.94  0.05334 .
## country.codeBRA      7.23e+00   1.46e+00   4.95  9.8e-07 ***
## country.codeBWA      2.08e+01   2.49e+00   8.33  6.2e-16 ***
## country.codeCAN      1.33e+00   2.66e+00   0.50  0.61716
## country.codeCHE     -5.63e-01   2.36e+00  -0.24  0.81180
## country.codeCHL      3.37e+00   1.49e+00   2.27  0.02383 *
## country.codeCHN      1.72e+01   1.61e+00  10.71 < 2e-16 ***
## country.codeCMR      2.11e+01   2.10e+00  10.08 < 2e-16 ***
## country.codeCOG      4.29e+01   2.53e+00  16.97 < 2e-16 ***
## country.codeCOL      7.15e+00   1.37e+00   5.22  2.5e-07 ***
## country.codeCPV      4.42e+00   2.44e+00   1.81  0.07127 .
## country.codeCRI      5.33e+00   1.36e+00   3.93  9.5e-05 ***
## country.codeCYP      1.90e+00   1.91e+00   0.99  0.32106
## country.codeCZE      4.23e+00   1.54e+00   2.75  0.00609 **
## country.codeDEU      1.58e+00   1.91e+00   0.83  0.40825
## country.codeDNK      7.38e-01   2.10e+00   0.35  0.72561
## country.codeDOM      4.49e+00   1.30e+00   3.46  0.00058 ***
## country.codeECU      5.42e+00   1.47e+00   3.68  0.00026 ***
## country.codeEGY     -2.46e+00   1.79e+00  -1.37  0.17110
## country.codeESP      2.89e+00   1.53e+00   1.89  0.05953 .
## country.codeEST      4.63e+00   1.58e+00   2.94  0.00347 **
## country.codeFIN      1.61e+00   1.77e+00   0.91  0.36375
## country.codeFJI      1.48e+00   2.38e+00   0.62  0.53393
## country.codeFRA      3.63e+00   2.11e+00   1.72  0.08583 .
## country.codeGBR     -2.28e-02   1.82e+00  -0.01  0.99004
## country.codeGEO      7.44e+00   1.26e+00   5.88  7.1e-09 ***
## country.codeGIN      3.39e+01   2.60e+00  13.08 < 2e-16 ***
## country.codeGNB      1.15e+01   2.63e+00   4.37  1.5e-05 ***
## country.codeGRC      1.06e+00   1.71e+00   0.62  0.53441
## country.codeGTM      7.33e+00   1.97e+00   3.71  0.00023 ***
## country.codeHND      1.30e+01   1.50e+00   8.68 < 2e-16 ***
## country.codeHRV      2.58e+00   1.60e+00   1.61  0.10697
## country.codeHUN      4.70e+00   1.59e+00   2.95  0.00326 **
## country.codeIDN      1.32e+01   1.39e+00   9.52 < 2e-16 ***
## country.codeIND      9.62e+00   2.09e+00   4.60  5.1e-06 ***

```

```

## country.codeIRL 2.89e+00 2.06e+00 1.40 0.16113
## country.codeIRN 3.53e+00 1.41e+00 2.50 0.01260 *
## country.codeISR 1.71e-01 1.84e+00 0.09 0.92622
## country.codeITA 2.16e+00 1.70e+00 1.27 0.20457
## country.codeJAM -6.86e-01 1.85e+00 -0.37 0.71140
## country.codeJOR 1.82e+00 2.73e+00 0.67 0.50546
## country.codeKAZ 2.25e+00 1.39e+00 1.62 0.10608
## country.codeKEN 2.27e+01 2.71e+00 8.39 3.9e-16 ***
## country.codeKGZ -1.21e+00 1.60e+00 -0.75 0.45131
## country.codeKOR 3.79e+00 2.46e+00 1.54 0.12481
## country.codeLKA 6.20e-01 1.86e+00 0.33 0.73848
## country.codeLTU 3.68e+00 1.54e+00 2.39 0.01723 *
## country.codeLUX -4.30e+00 3.39e+00 -1.27 0.20457
## country.codeLVA 4.20e+00 1.52e+00 2.76 0.00605 **
## country.codeMDA 1.19e+00 1.59e+00 0.75 0.45279
## country.codeMDG 6.74e+01 2.34e+00 28.80 < 2e-16 ***
## country.codeMEX 7.52e+00 1.50e+00 5.03 6.7e-07 ***
## country.codeMLI 5.55e+00 2.73e+00 2.03 0.04285 *
## country.codeMLT 7.22e+00 2.39e+00 3.02 0.00264 **
## country.codeMMR -1.13e+00 2.07e+00 -0.55 0.58407
## country.codeMNG 4.05e+00 1.51e+00 2.68 0.00752 **
## country.codeMUS 3.29e+00 1.92e+00 1.71 0.08694 .
## country.codeMYS 4.43e+00 1.80e+00 2.46 0.01428 *
## country.codeNER 4.94e+01 2.90e+00 17.07 < 2e-16 ***
## country.codeNLD 3.41e+00 2.23e+00 1.53 0.12705
## country.codeNOR -3.28e+00 2.29e+00 -1.43 0.15214
## country.codePAK 7.37e-01 2.01e+00 0.37 0.71368
## country.codePAN 1.03e+01 1.36e+00 7.55 1.8e-13 ***
## country.codePER 7.79e+00 1.41e+00 5.53 5.0e-08 ***
## country.codePHL 4.56e+00 1.68e+00 2.71 0.00688 **
## country.codePOL 1.99e+00 1.45e+00 1.37 0.17007
## country.codePRT 2.21e+00 1.53e+00 1.44 0.15018
## country.codePRY 2.35e+00 1.36e+00 1.73 0.08457 .
## country.codeROU 7.13e+00 1.34e+00 5.33 1.4e-07 ***
## country.codeRUS -2.66e+00 1.51e+00 -1.76 0.07864 .
## country.codeSDN 1.14e+01 2.52e+00 4.52 7.7e-06 ***
## country.codeSLE 1.55e+01 2.77e+00 5.61 3.2e-08 ***
## country.codeSLV 3.39e+00 1.30e+00 2.61 0.00942 **
## country.codeSRB 6.32e+00 1.43e+00 4.41 1.3e-05 ***
## country.codeSVK 3.65e+00 1.65e+00 2.21 0.02776 *
## country.codeSVN 3.79e+00 1.64e+00 2.31 0.02150 *
## country.codeSWE 1.46e+00 1.89e+00 0.77 0.44099
## country.codeSWZ 3.27e+01 2.40e+00 13.63 < 2e-16 ***
## country.codeSYC 6.39e+00 2.53e+00 2.52 0.01203 *
## country.codeTCD 2.06e+01 2.73e+00 7.54 2.0e-13 ***
## country.codeTGO 4.79e+01 2.27e+00 21.11 < 2e-16 ***
## country.codeTHA 1.76e+00 1.42e+00 1.24 0.21590
## country.codeTJK 5.66e+00 1.99e+00 2.84 0.00472 **
## country.codeTUN 6.59e-01 1.89e+00 0.35 0.72738
## country.codeTUR 1.38e+00 1.56e+00 0.88 0.37855
## country.codeTZA 3.61e+01 2.73e+00 13.22 < 2e-16 ***
## country.codeUGA 3.00e+01 2.44e+00 12.32 < 2e-16 ***
## country.codeUKR -3.45e+00 1.50e+00 -2.30 0.02173 *
## country.codeURY 1.52e+00 1.45e+00 1.05 0.29434

```

```

## country.codeUSA 1.19e+00 2.73e+00 0.44 0.66154
## country.codeVEN 6.94e+00 2.37e+00 2.93 0.00351 **
## country.codeZAF 2.36e+01 1.53e+00 15.44 < 2e-16 ***
## year -1.76e-01 3.17e-02 -5.54 4.8e-08 ***
## incomeL 6.10e-01 1.14e+00 0.54 0.59118
## incomeLM -1.40e+00 6.94e-01 -2.02 0.04396 *
## incomeUM -1.58e+00 4.62e-01 -3.42 0.00066 ***
## edu.total -3.10e-01 1.68e-01 -1.84 0.06565 .
## hlth -1.62e-01 1.41e-01 -1.16 0.24825
## mil 4.52e-01 1.83e-01 2.47 0.01390 *
## fdi 2.37e-13 3.09e-12 0.08 0.93884
## lbr.part 5.25e-02 3.18e-02 1.65 0.09893 .
## unemp 4.81e-02 4.20e-02 1.15 0.25211
## pop.gwth.total -5.72e-01 3.80e-01 -1.50 0.13310
## pop.gwth.rural 3.09e-01 2.49e-01 1.24 0.21554
## gdp.dflt 3.32e-02 2.19e-02 1.52 0.12921
## gcf -2.06e-01 2.84e-02 -7.23 1.6e-12 ***
## trade -1.62e-02 7.30e-03 -2.22 0.02691 *
## gdp.pc 2.14e-04 2.23e-05 9.61 < 2e-16 ***
## lfdi -1.27e-01 1.24e-01 -1.02 0.30693
## lgdp.dflt -4.52e-02 1.49e-01 -0.30 0.76114
## lgdp.pc -4.74e+00 4.94e-01 -9.59 < 2e-16 ***
## ---
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ',' 1
##
## Residual standard error: 1.97 on 555 degrees of freedom
##   (796 observations deleted due to missingness)
## Multiple R-squared: 0.962, Adjusted R-squared: 0.953
## F-statistic: 115 on 122 and 555 DF, p-value: <2e-16

```

Next we will be performing standardisation. However not all the variables are numerical. Therefore to overcome it we build the function below. This will allow us to perform standardisation to the dataset.

```

subset1$income <- factor(subset1$income)
subset1$country.code <- factor(subset1$country.code)

sd0 <- function(vct) {
  if (!is.numeric(vct)) {
    return(NA)
  }

  return(sd(vct, na.rm = T))
}

std0 <- function(vct, scl) {
  if (!is.numeric(vct)) {
    return(vct)
  }

  return(vct/scl)
}

```

```

scale <- function(data) {
  unlist(lapply(data, sd0))
}

standardise <- function(data) {
  scaler <- scale(data)

  numCols <- which(unlist(lapply(data, is.numeric)))
  num <- as.data.frame(mapply(std0, vct = data[, numCols], scl = scaler[numCols], SIMPLIFY = T))
  fct <- data[, -numCols]
  return(cbind(fct, num))
}

```

Next we will perform standardisation to the dataset.

```
str(subset1)
```

```

## 'data.frame': 1843 obs. of 19 variables:
## $ pov : num 21.4 14.6 31.1 0.5 1.1 0.6 0.2 0.6 1 0.1 ...
## $ country.code : Factor w/ 168 levels "AGO","ALB","ARE",...: 1 1 1 2 2 2 2 2 2 ...
## $ year : int 2000 2008 2018 1996 2002 2005 2008 2012 2014 2015 ...
## $ income : Factor w/ 4 levels "H","L","LM","UM": 2 3 3 3 3 3 3 4 4 4 ...
## $ edu.total : num 2.61 2.69 2.04 3.08 3.12 ...
## $ hlth : num 1.91 3.32 2.54 NA 6.91 ...
## $ mil : num 6.39 3.57 1.87 1.38 1.32 ...
## $ fdi : num 8.79e+08 1.68e+09 -6.46e+09 9.01e+07 1.35e+08 ...
## $ lbr.part : num NA NA NA 38.8 59.6 ...
## $ unemp : num NA NA NA 12.3 15.8 ...
## $ pop.gwth.total: num 3.277 3.711 3.276 -0.622 -0.3 ...
## $ pop.gwth.rural: num 0.921 1.91 1.338 -1.546 -2.169 ...
## $ gdp.dflt : num 418.02 19.37 28.17 38.17 3.65 ...
## $ gcf : num 30.5 30.8 17.9 18.1 35.3 ...
## $ trade : num 152.5 121.4 66.4 44.9 68.5 ...
## $ gdp.pc : num 557 4081 2525 1010 1425 ...
## $ lfdi : num 20.6 21.2 NA 18.3 18.7 ...
## $ lgdp.dflt : num 6.04 2.96 3.34 3.64 1.29 ...
## $ lgdp.pc : num 6.32 8.31 7.83 6.92 7.26 ...

```

```
df0 <- subset1
head(subset1)
```

```

##   pov country.code year income edu.total hlth      fdi lbr.part unemp
## 1 21.4        AGO 2000     L    2.61 1.91 6.39 8.79e+08      NA    NA
## 2 14.6        AGO 2008    LM    2.69 3.32 3.57 1.68e+09      NA    NA
## 3 31.1        AGO 2018    LM    2.04 2.54 1.87 -6.46e+09      NA    NA
## 4  0.5        ALB 1996    LM    3.08  NA 1.38 9.01e+07    38.8 12.3
## 5  1.1        ALB 2002    LM    3.12 6.91 1.32 1.35e+08    59.6 15.8
## 6  0.6        ALB 2005    LM    3.28 6.34 1.35 2.62e+08    34.5 14.1
##   pop.gwth.total pop.gwth.rural gdp.dflt  gcf trade gdp.pc lfdi lgdp.dflt
## 1          3.277           0.921  418.02 30.5 152.5   557 20.6    6.04
## 2          3.711           1.910  19.37 30.8 121.4  4081 21.2    2.96

```

```

## 3      3.276      1.338    28.17 17.9  66.4   2525   NA     3.34
## 4     -0.622     -1.546    38.17 18.1  44.9   1010  18.3    3.64
## 5     -0.300     -2.169    3.65 35.3  68.5   1425  18.7    1.29
## 6     -0.512     -2.519    3.31 36.9  70.9   2674 19.4    1.20
##   lgdp.pc
## 1    6.32
## 2    8.31
## 3    7.83
## 4    6.92
## 5    7.26
## 6    7.89

```

```
apply(subset1, 2, sd)
```

	pov	country.code	year	income	edu.total
##	17.54	NA	9.26	NA	NA
##	hlth	mil	fdi	lbr.part	unemp
##	NA	NA	NA	NA	NA
##	pop.gwth.total	pop.gwth.rural	gdp.dflt	gcf	trade
##	1.12	NA	NA	NA	NA
##	gdp.pc	lfdi	lgdp.dflt	lgdp.pc	
##	NA	NA	NA	NA	

Next we will plot a pov vs income scatter plot to see the effect of income on pov. From the scatter plot, we can deduce that the low income countries have a higher pov while the higher income countries have low pov. This is consistent with our intuition that countries that are richer do not have as much poverty issue as the poorer countries.

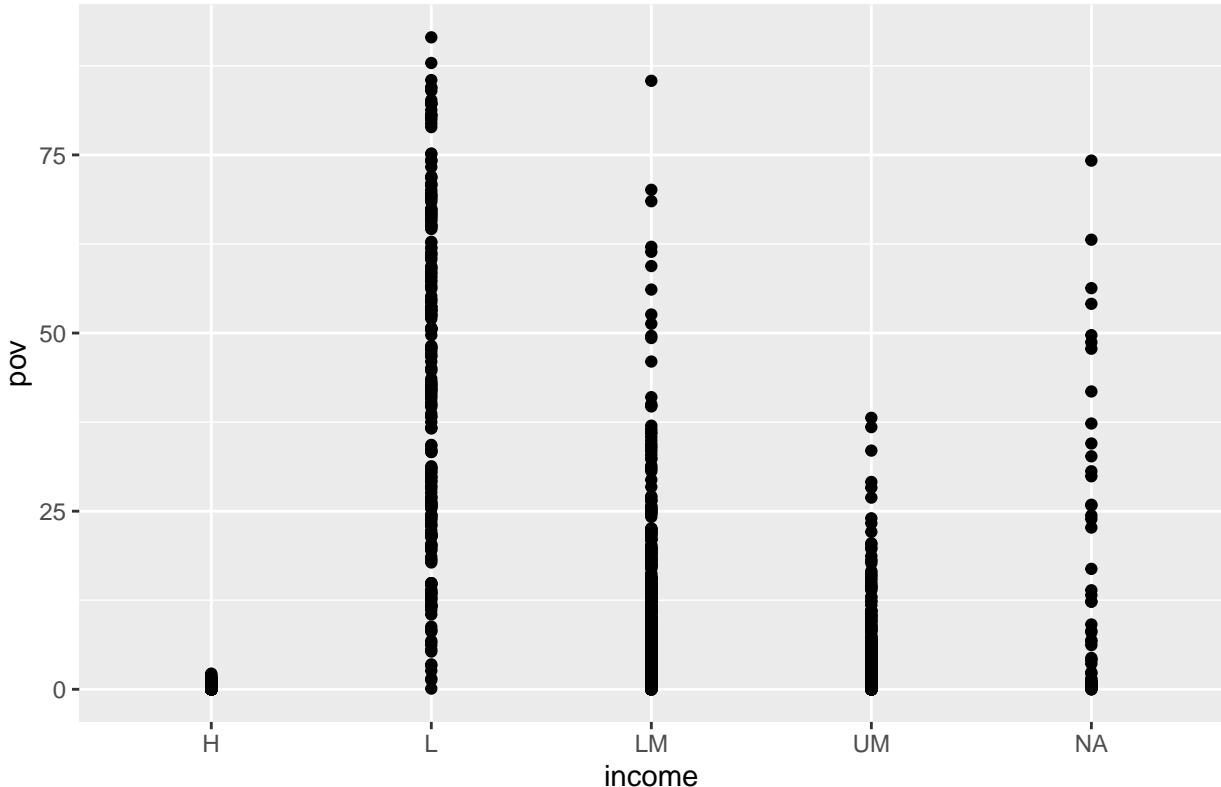
```

ggplot(data = subset1, aes(x = income, y = pov)) +
  geom_point() + geom_smooth(method = "lm") + labs(title = "Pov vs. Income",
  x = "income", y = "pov")

## `geom_smooth()` using formula 'y ~ x'

```

Pov vs. Income



Next we will build a MLR model with data from subset 1 to see the relationship between dependent variable:pov and the other variables

```
model1 <- lm(pov ~ ., data = subset1)
summary(model1)
```

```
##
## Call:
## lm(formula = pov ~ ., data = subset1)
##
## Residuals:
##    Min     1Q Median     3Q    Max 
## -9.411 -0.723 -0.025  0.702 15.150 
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)    
## (Intercept) 3.41e+02   5.43e+01   6.28  6.0e-10 ***
## country.codeARG 3.26e+00   1.27e+00   2.56  0.01056 *  
## country.codeARM -5.55e-01   1.31e+00  -0.42  0.67290  
## country.codeAUS  2.51e+00   1.86e+00   1.35  0.17685  
## country.codeAUT  3.39e+00   1.73e+00   1.96  0.04981 *  
## country.codeAZE -5.55e+00   1.56e+00  -3.55  0.00041 *** 
## country.codeBEL  5.28e+00   1.82e+00   2.90  0.00391 **  
## country.codeBEN  1.46e+01   2.51e+00   5.82  9.1e-09 *** 
## country.codeBFA  2.35e+01   2.59e+00   9.05 < 2e-16 *** 
## country.codeBGD  1.36e+01   1.94e+00   7.01  5.5e-12 ***
```

## country.codeBGR	4.87e+00	1.22e+00	4.01	6.8e-05	***
## country.codeBLR	3.43e+00	1.30e+00	2.65	0.00818	**
## country.codeBOL	4.89e+00	1.67e+00	2.93	0.00355	**
## country.codeBRA	8.40e+00	1.32e+00	6.38	3.3e-10	***
## country.codeBWA	2.13e+01	2.42e+00	8.80	< 2e-16	***
## country.codeCAN	2.67e+00	2.06e+00	1.29	0.19657	
## country.codeCHE	8.59e-01	2.02e+00	0.42	0.67157	
## country.codeCHL	4.26e+00	1.30e+00	3.27	0.00111	**
## country.codeCHN	1.50e+01	1.43e+00	10.49	< 2e-16	***
## country.codeCMR	2.23e+01	1.98e+00	11.26	< 2e-16	***
## country.codeCOG	4.29e+01	2.44e+00	17.60	< 2e-16	***
## country.codeCOL	7.96e+00	1.23e+00	6.47	1.8e-10	***
## country.codeCPV	4.22e+00	2.36e+00	1.79	0.07378	.
## country.codeCRI	6.10e+00	1.24e+00	4.91	1.1e-06	***
## country.codeCYP	3.61e+00	1.65e+00	2.18	0.02951	*
## country.codeCZE	4.57e+00	1.38e+00	3.32	0.00093	***
## country.codeDEU	2.96e+00	1.70e+00	1.74	0.08250	.
## country.codeDNK	2.72e+00	1.81e+00	1.50	0.13349	
## country.codeDOM	5.11e+00	1.15e+00	4.44	1.0e-05	***
## country.codeECU	6.06e+00	1.31e+00	4.65	4.0e-06	***
## country.codeEGY	-2.06e+00	1.64e+00	-1.25	0.21000	
## country.codeESP	3.70e+00	1.39e+00	2.66	0.00806	**
## country.codeEST	4.84e+00	1.41e+00	3.44	0.00061	***
## country.codeFIN	2.99e+00	1.58e+00	1.89	0.05918	.
## country.codeFJI	1.36e+00	2.30e+00	0.59	0.55576	
## country.codeFRA	4.59e+00	1.98e+00	2.32	0.02075	*
## country.codeGBR	1.68e+00	1.63e+00	1.03	0.30413	
## country.codeGEO	7.38e+00	1.15e+00	6.40	2.9e-10	***
## country.codeGIN	3.59e+01	2.47e+00	14.55	< 2e-16	***
## country.codeGNB	1.25e+01	2.47e+00	5.06	5.2e-07	***
## country.codeGRC	1.87e+00	1.58e+00	1.19	0.23530	
## country.codeGTM	8.20e+00	1.86e+00	4.42	1.1e-05	***
## country.codeHND	1.35e+01	1.37e+00	9.82	< 2e-16	***
## country.codeHRV	2.90e+00	1.38e+00	2.10	0.03641	*
## country.codeHUN	4.93e+00	1.43e+00	3.46	0.00058	***
## country.codeIDN	1.40e+01	1.27e+00	11.05	< 2e-16	***
## country.codeIND	9.52e+00	1.97e+00	4.83	1.7e-06	***
## country.codeIRL	3.55e+00	1.83e+00	1.94	0.05310	.
## country.codeIRN	2.99e+00	1.22e+00	2.45	0.01459	*
## country.codeISR	1.38e+00	1.65e+00	0.84	0.40300	
## country.codeITA	3.33e+00	1.48e+00	2.25	0.02467	*
## country.codeJAM	1.05e-02	1.77e+00	0.01	0.99526	
## country.codeJOR	9.68e-01	2.37e+00	0.41	0.68351	
## country.codeKAZ	2.79e+00	1.29e+00	2.17	0.03045	*
## country.codeKEN	2.46e+01	2.56e+00	9.63	< 2e-16	***
## country.codeKGZ	-5.64e-01	1.44e+00	-0.39	0.69488	
## country.codeKOR	4.42e+00	1.88e+00	2.35	0.01894	*
## country.codeLKA	-2.25e-01	1.59e+00	-0.14	0.88756	
## country.codeLTU	4.25e+00	1.38e+00	3.09	0.00210	**
## country.codeLUX	-4.08e+00	3.08e+00	-1.33	0.18475	
## country.codeLVA	4.80e+00	1.33e+00	3.60	0.00034	***
## country.codeMDA	1.49e+00	1.38e+00	1.08	0.27978	
## country.codeMDG	6.53e+01	2.02e+00	32.33	< 2e-16	***
## country.codeMEX	8.39e+00	1.31e+00	6.42	2.5e-10	***

```

## country.codeMLI 6.73e+00 2.58e+00 2.61 0.00921 **
## country.codeMLT 7.26e+00 2.12e+00 3.42 0.00066 ***
## country.codeMMR -1.07e+00 1.97e+00 -0.54 0.58673
## country.codeMNG 4.09e+00 1.38e+00 2.95 0.00325 **
## country.codeMUS 2.90e+00 1.63e+00 1.77 0.07671 .
## country.codeMYS 4.65e+00 1.64e+00 2.84 0.00464 **
## country.codeNER 5.07e+01 2.70e+00 18.76 < 2e-16 ***
## country.codeNIC 2.81e+00 2.30e+00 1.22 0.22173
## country.codeNLD 4.36e+00 1.96e+00 2.22 0.02664 *
## country.codeNOR -9.63e-01 2.02e+00 -0.48 0.63329
## country.codePAK 9.39e-01 1.74e+00 0.54 0.58952
## country.codePAN 1.02e+01 1.25e+00 8.12 2.1e-15 ***
## country.codePER 8.66e+00 1.28e+00 6.75 3.1e-11 ***
## country.codePHL 5.26e+00 1.55e+00 3.39 0.00073 ***
## country.codePOL 2.66e+00 1.29e+00 2.06 0.04005 *
## country.codePRT 3.25e+00 1.39e+00 2.34 0.01959 *
## country.codePRY 3.21e+00 1.26e+00 2.55 0.01091 *
## country.codeROU 7.72e+00 1.20e+00 6.43 2.4e-10 ***
## country.codeRUS -1.01e+00 1.37e+00 -0.74 0.46204
## country.codeSDN 1.13e+01 2.42e+00 4.69 3.3e-06 ***
## country.codeSEN 3.58e+01 2.38e+00 15.05 < 2e-16 ***
## country.codeSLE 1.70e+01 2.60e+00 6.53 1.3e-10 ***
## country.codeSLV 4.60e+00 1.17e+00 3.93 9.1e-05 ***
## country.codeSRB 6.55e+00 1.29e+00 5.08 4.9e-07 ***
## country.codeSVK 3.87e+00 1.49e+00 2.60 0.00952 **
## country.codeSVN 4.47e+00 1.47e+00 3.04 0.00245 **
## country.codeSWE 3.18e+00 1.68e+00 1.89 0.05867 .
## country.codeSWZ 3.32e+01 2.31e+00 14.41 < 2e-16 ***
## country.codeSYC 6.49e+00 2.44e+00 2.65 0.00813 **
## country.codeTCD 2.14e+01 2.57e+00 8.33 4.0e-16 ***
## country.codeTGO 4.88e+01 2.11e+00 23.16 < 2e-16 ***
## country.codeTHA 1.83e+00 1.32e+00 1.39 0.16550
## country.codeTJK 6.78e+00 1.84e+00 3.69 0.00024 ***
## country.codeTUN 1.64e+00 1.57e+00 1.05 0.29465
## country.codeTUR 1.60e+00 1.38e+00 1.16 0.24836
## country.codeTZA 3.72e+01 2.57e+00 14.47 < 2e-16 ***
## country.codeUGA 3.16e+01 2.23e+00 14.17 < 2e-16 ***
## country.codeUKR -2.44e+00 1.33e+00 -1.83 0.06752 .
## country.codeURY 2.45e+00 1.33e+00 1.84 0.06572 .
## country.codeUSA 3.25e+00 2.42e+00 1.34 0.17986
## country.codeVEN 7.76e+00 2.30e+00 3.38 0.00076 ***
## country.codeZAF 2.41e+01 1.43e+00 16.83 < 2e-16 ***
## country.codeZMB 6.40e+01 2.37e+00 26.98 < 2e-16 ***
## year -1.46e-01 2.80e-02 -5.21 2.5e-07 ***
## incomeL -1.22e-02 9.74e-01 -0.01 0.99000
## incomeLM -1.42e+00 6.15e-01 -2.30 0.02157 *
## incomeUM -1.54e+00 4.17e-01 -3.70 0.00023 ***
## edu.total -4.16e-01 1.47e-01 -2.83 0.00478 **
## hlth -1.63e-01 1.26e-01 -1.30 0.19503
## mil 4.44e-01 1.70e-01 2.60 0.00938 **
## fdi -1.04e-12 2.55e-12 -0.41 0.68494
## lbr.part 2.90e-02 2.79e-02 1.04 0.29968
## unemp 6.97e-02 3.73e-02 1.87 0.06228 .
## pop.gwth.total -4.40e-01 3.31e-01 -1.33 0.18313

```

```

## pop.gwth.rural 2.30e-01 2.17e-01 1.06 0.28932
## gdp.dflt      2.19e-02 1.96e-02 1.11 0.26580
## gcf          -1.69e-01 2.38e-02 -7.12 2.7e-12 ***
## trade        -1.10e-02 6.58e-03 -1.68 0.09371 .
## gdp.pc       2.05e-04 1.99e-05 10.30 < 2e-16 ***
## lfdi         -8.85e-02 1.11e-01 -0.79 0.42741
## lgdp.dflt    5.58e-02 1.34e-01 0.42 0.67743
## lgdp.pc     -4.81e+00 4.43e-01 -10.87 < 2e-16 ***
## ---
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.97 on 709 degrees of freedom
##   (1008 observations deleted due to missingness)
## Multiple R-squared: 0.961, Adjusted R-squared: 0.954
## F-statistic: 138 on 125 and 709 DF, p-value: <2e-16

vif(model1)

```

```

##              GVIF Df GVIF^(1/(2*Df))
## country.code 4.67e+12 106           1.15
## year        4.59e+00  1            2.14
## income      1.47e+02  3            2.30
## edu.total   8.90e+00  1            2.98
## hlth        1.65e+01  1            4.07
## mil         8.44e+00  1            2.90
## fdi         4.80e+00  1            2.19
## lbr.part    9.43e+00  1            3.07
## unemp       5.79e+00  1            2.41
## pop.gwth.total 2.15e+01  1            4.64
## pop.gwth.rural 1.80e+01  1            4.24
## gdp.dflt    4.02e+00  1            2.00
## gcf         4.89e+00  1            2.21
## trade       2.06e+01  1            4.54
## gdp.pc      3.16e+01  1            5.62
## lfdi        9.51e+00  1            3.08
## lgdp.dflt   4.59e+00  1            2.14
## lgdp.pc     6.50e+01  1            8.06

```

Next we will perform regularisation. We will split the data to predictor variable “train.x” and response variables “train.y”

```

subset2 <- subset(subset1, select = c("pov", "country.code",
  "year", "income", "edu.total", "hlth", "mil", "fdi",
  "lbr.part", "unemp", "pop.gwth.total", "pop.gwth.rural",
  "gdp.dflt", "gcf", "trade", "gdp.pc", "lfdi", "lgdp.dflt",
  "lgdp.pc"))
subset2 <- na.omit(subset2)
head(subset2)

```

```

##   pov country.code year income edu.total hlth mil      fdi lbr.part unemp
## 5  1.1        ALB 2002     LM      3.12 6.91 1.32 1.35e+08      59.6 15.8
## 6  0.6        ALB 2005     LM      3.28 6.34 1.35 2.62e+08      34.5 14.1
## 8  0.6        ALB 2012     UM      2.93 5.06 1.49 9.18e+08      57.0 13.4

```

```

## 9 1.0      ALB 2014      UM      3.05 5.50 1.35 1.15e+09      53.4 18.0
## 10 0.1     ALB 2015      UM      3.44 4.90 1.16 9.90e+08      55.5 17.2
## 12 0.4     ALB 2017      UM      3.61 5.01 1.11 1.02e+09      58.1 13.6
##   pop.gwth.total pop.gwth.rural gdp.dflt gcf trade gdp.pc lfdi lgdp.dflt
## 5      -0.300      -2.17    3.647 35.3 68.5 1425 18.7    1.2940
## 6      -0.512      -2.52    3.307 36.9 70.9 2674 19.4    1.1960
## 8      -0.165      -2.51    1.043 28.3 76.5 4248 20.6    0.0418
## 9      -0.207      -2.56    1.550 25.7 75.4 4579 20.9    0.4382
## 10     -0.291      -2.64    0.564 25.8 71.8 3953 20.7   -0.5727
## 12     -0.092      -2.43    1.451 25.1 78.2 4531 20.7    0.3723
##   lgdp.pc
## 5      7.26
## 6      7.89
## 8      8.35
## 9      8.43
## 10     8.28
## 12     8.42

```

```

train.x <- data.matrix(subset(subset2, select = -pov))
train.y = data.matrix(subset2$pov)

```

Next we will perform splitting of data to train and test.

```

split_train_test <- function(data) {
  set.seed(1984)
  # split data
  seeds <- data %>%
    group_by(country.code) %>%
    filter(row_number() == 1)

  plants <- data %>%
    group_by(country.code) %>%
    filter(row_number() != 1)

  isComplete <- which(complete.cases(plants))
  idx <- sample(isComplete, replace = F, 0.2 * nrow(plants))

  training <- plants[-idx, ] %>%
    rbind(seeds)
  test <- plants[idx, ]

  return(list(training, test))
}

training <- split_train_test(subset2)[[1]]
test <- split_train_test(subset2)[[2]]

ntrain <- 1:nrow(training)
combine <- rbind(training, test)
combine.mtrx <- model.matrix(pov ~ ., data = combine)
train.mtrx <- combine.mtrx[ntrain, ]
test.mtrx <- combine.mtrx[-ntrain, ]

```

```

train.x <- train.mtrx[, -1]
train.y <- training$pov

test.x <- test.mtrx[, -1]
test.y <- test$pov

```

We then build a MLR model with the training model to train it.

```

model3 <- lm(pov ~ ., data = training)
summary(model3)

```

```

##
## Call:
## lm(formula = pov ~ ., data = training)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -9.213 -0.704 -0.018  0.621 14.165
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 3.38e+02  6.39e+01   5.30  1.7e-07 ***
## country.codeARG 3.92e+00  1.39e+00   2.83  0.00485 **
## country.codeARM -1.77e-01 1.41e+00  -0.13  0.90019
## country.codeAUS 2.95e+00  1.99e+00   1.48  0.14016
## country.codeAUT 3.76e+00  1.94e+00   1.94  0.05276 .
## country.codeAZE -6.13e+00 1.72e+00  -3.57  0.00039 ***
## country.codeBEL 5.77e+00  1.98e+00   2.91  0.00371 **
## country.codeBEN 1.42e+01  2.69e+00   5.29  1.8e-07 ***
## country.codeBFA 2.32e+01  2.78e+00   8.36  4.9e-16 ***
## country.codeBGD 1.34e+01  2.06e+00   6.51  1.6e-10 ***
## country.codeBGR 4.70e+00  1.30e+00   3.62  0.00032 ***
## country.codeBLR 3.00e+00  1.44e+00   2.08  0.03801 *
## country.codeBOL 4.91e+00  1.91e+00   2.57  0.01037 *
## country.codeBRA 8.80e+00  1.41e+00   6.23  9.0e-10 ***
## country.codeBWA 2.10e+01  2.53e+00   8.31  7.0e-16 ***
## country.codeCAN 3.22e+00  2.20e+00   1.46  0.14398
## country.codeCHE 1.24e+00  2.21e+00   0.56  0.57567
## country.codeCHL 4.47e+00  1.39e+00   3.21  0.00141 **
## country.codeCHN 1.50e+01  1.55e+00   9.71  < 2e-16 ***
## country.codeCMR 2.24e+01  2.11e+00  10.59  < 2e-16 ***
## country.codeCOG 4.32e+01  2.56e+00  16.86  < 2e-16 ***
## country.codeCOL 8.15e+00  1.33e+00   6.14  1.5e-09 ***
## country.codeCPV 4.59e+00  2.47e+00   1.86  0.06307 .
## country.codeCRI 6.21e+00  1.35e+00   4.61  5.0e-06 ***
## country.codeCYP 3.51e+00  1.84e+00   1.90  0.05738 .
## country.codeCZE 4.37e+00  1.52e+00   2.87  0.00420 **
## country.codeDEU 3.36e+00  1.90e+00   1.77  0.07810 .
## country.codeDNK 2.83e+00  2.05e+00   1.38  0.16909
## country.codeDOM 4.99e+00  1.24e+00   4.02  6.6e-05 ***
## country.codeECU 5.90e+00  1.49e+00   3.96  8.5e-05 ***
## country.codeEGY -1.62e+00 1.76e+00  -0.92  0.35915
## country.codeESP 4.31e+00  1.52e+00   2.83  0.00480 **

```

```

## country.codeEST 4.65e+00 1.53e+00 3.04 0.00247 **
## country.codeFIN 3.20e+00 1.74e+00 1.84 0.06700 .
## country.codeFJI 1.07e+00 2.40e+00 0.45 0.65477
## country.codeFRA 5.12e+00 2.10e+00 2.44 0.01489 *
## country.codeGBR 2.12e+00 1.82e+00 1.16 0.24454
## country.codeGEO 7.28e+00 1.23e+00 5.92 5.6e-09 ***
## country.codeGIN 3.55e+01 2.60e+00 13.69 < 2e-16 ***
## country.codeGNB 1.26e+01 2.61e+00 4.83 1.8e-06 ***
## country.codeGRC 2.05e+00 1.77e+00 1.16 0.24662
## country.codeGTM 8.61e+00 1.96e+00 4.39 1.3e-05 ***
## country.codeHND 1.35e+01 1.61e+00 8.37 4.5e-16 ***
## country.codeHRV 3.01e+00 1.47e+00 2.05 0.04053 *
## country.codeHUN 5.06e+00 1.53e+00 3.31 0.00099 ***
## country.codeIDN 1.38e+01 1.41e+00 9.84 < 2e-16 ***
## country.codeIND 9.53e+00 2.10e+00 4.53 7.1e-06 ***
## country.codeIRL 3.97e+00 2.03e+00 1.96 0.05035 .
## country.codeIRN 3.12e+00 1.32e+00 2.37 0.01816 *
## country.codeISR 1.35e+00 1.78e+00 0.76 0.45013
## country.codeITA 3.66e+00 1.63e+00 2.24 0.02543 *
## country.codeJAM 2.41e-01 1.85e+00 0.13 0.89641
## country.codeJOR 1.35e+00 2.56e+00 0.53 0.59791
## country.codeKAZ 2.97e+00 1.39e+00 2.14 0.03252 *
## country.codeKEN 2.44e+01 2.71e+00 9.00 < 2e-16 ***
## country.codeKGZ 3.40e-02 1.59e+00 0.02 0.98293
## country.codeKOR 4.43e+00 1.97e+00 2.24 0.02521 *
## country.codeLKA -2.50e-01 1.67e+00 -0.15 0.88155
## country.codeLTU 4.11e+00 1.50e+00 2.73 0.00646 **
## country.codeLUX -4.07e+00 3.35e+00 -1.22 0.22476
## country.codeLVA 4.64e+00 1.46e+00 3.18 0.00158 **
## country.codeMDA 1.58e+00 1.49e+00 1.06 0.29058
## country.codeMDG 6.52e+01 2.18e+00 29.89 < 2e-16 ***
## country.codeMEX 8.55e+00 1.48e+00 5.78 1.2e-08 ***
## country.codeMLI 6.38e+00 2.76e+00 2.31 0.02105 *
## country.codeMLT 7.39e+00 2.32e+00 3.19 0.00151 **
## country.codeMMR -1.14e+00 2.09e+00 -0.54 0.58654
## country.codeMNG 3.64e+00 1.51e+00 2.41 0.01613 *
## country.codeMUS 2.73e+00 1.72e+00 1.59 0.11347
## country.codeMYS 4.26e+00 1.75e+00 2.43 0.01541 *
## country.codeNER 5.06e+01 2.89e+00 17.49 < 2e-16 ***
## country.codeNIC 3.20e+00 2.40e+00 1.33 0.18259
## country.codeNLD 4.58e+00 2.21e+00 2.07 0.03899 *
## country.codeNOR -7.14e-01 2.24e+00 -0.32 0.74962
## country.codePAK 2.00e+00 2.00e+00 1.00 0.31788
## country.codePAN 1.04e+01 1.33e+00 7.83 2.5e-14 ***
## country.codePER 9.09e+00 1.39e+00 6.56 1.2e-10 ***
## country.codePHL 4.85e+00 1.72e+00 2.82 0.00496 **
## country.codePOL 2.76e+00 1.40e+00 1.96 0.05019 .
## country.codePRT 3.61e+00 1.52e+00 2.38 0.01767 *
## country.codePRY 3.17e+00 1.36e+00 2.33 0.02039 *
## country.codeROU 8.39e+00 1.36e+00 6.19 1.2e-09 ***
## country.codeRUS -1.12e+00 1.46e+00 -0.77 0.44166
## country.codeSDN 1.16e+01 2.53e+00 4.58 5.6e-06 ***
## country.codeSEN 3.60e+01 2.51e+00 14.36 < 2e-16 ***
## country.codeSLE 1.70e+01 2.76e+00 6.17 1.3e-09 ***

```

```

## country.codeSLV 4.50e+00 1.29e+00 3.49 0.00053 ***
## country.codeSRB 6.79e+00 1.36e+00 4.98 8.6e-07 ***
## country.codeSVK 3.89e+00 1.61e+00 2.41 0.01622 *
## country.codeSVN 4.48e+00 1.61e+00 2.78 0.00560 **
## country.codeSWE 3.38e+00 1.87e+00 1.81 0.07077 .
## country.codeSWZ 3.36e+01 2.41e+00 13.95 < 2e-16 ***
## country.codeSYC 6.21e+00 2.56e+00 2.43 0.01547 *
## country.codeTCD 2.13e+01 2.76e+00 7.71 5.6e-14 ***
## country.codeTGO 4.85e+01 2.28e+00 21.25 < 2e-16 ***
## country.codeTHA 1.64e+00 1.41e+00 1.17 0.24446
## country.codeTJK 6.66e+00 1.98e+00 3.36 0.00083 ***
## country.codeTUN 1.81e+00 1.64e+00 1.10 0.27090
## country.codeTUR 1.07e+00 1.63e+00 0.66 0.51214
## country.codeTZA 3.72e+01 2.72e+00 13.66 < 2e-16 ***
## country.codeUGA 3.16e+01 2.40e+00 13.17 < 2e-16 ***
## country.codeUKR -2.49e+00 1.45e+00 -1.72 0.08649 .
## country.codeURY 2.63e+00 1.47e+00 1.79 0.07380 .
## country.codeUSA 4.63e+00 2.77e+00 1.67 0.09454 .
## country.codeVEN 7.94e+00 2.38e+00 3.33 0.00092 ***
## country.codeZAF 2.63e+01 1.86e+00 14.14 < 2e-16 ***
## country.codeZMB 6.44e+01 2.50e+00 25.79 < 2e-16 ***
## year -1.44e-01 3.29e-02 -4.38 1.4e-05 ***
## incomeL 7.52e-02 1.09e+00 0.07 0.94494
## incomeLM -1.60e+00 6.92e-01 -2.31 0.02131 *
## incomeUM -1.45e+00 4.50e-01 -3.23 0.00133 **
## edu.total -3.72e-01 1.67e-01 -2.23 0.02629 *
## hlth -2.48e-01 1.41e-01 -1.75 0.08019 .
## mil 5.20e-01 1.84e-01 2.82 0.00499 **
## fdi -1.54e-12 2.87e-12 -0.54 0.59207
## lbr.part 3.07e-02 3.03e-02 1.01 0.31114
## unemp 5.28e-02 4.20e-02 1.26 0.20974
## pop.gwth.total -5.21e-01 3.75e-01 -1.39 0.16464
## pop.gwth.rural 2.20e-01 2.45e-01 0.90 0.36877
## gdp.dflt 2.87e-02 2.23e-02 1.29 0.19864
## gcf -1.64e-01 2.66e-02 -6.17 1.3e-09 ***
## trade -1.00e-02 7.23e-03 -1.38 0.16727
## gdp.pc 2.06e-04 2.17e-05 9.51 < 2e-16 ***
## lfdi -1.27e-01 1.21e-01 -1.05 0.29306
## lgdp.dflt -1.29e-02 1.54e-01 -0.08 0.93334
## lgdp.pc -4.85e+00 5.01e-01 -9.68 < 2e-16 ***
## ---
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ',' 1
##
## Residual standard error: 2.02 on 564 degrees of freedom
## Multiple R-squared: 0.965, Adjusted R-squared: 0.957
## F-statistic: 124 on 125 and 564 DF, p-value: <2e-16

fit <- model3$fitted.values
true <- training$pov
summary_MLR3_train <- eval_results(true, fit)

```

We then evaluate the model between train and test

```

fit <- predict(model3, newdata = test)
true <- test$pov
summary_MLR3_test <- eval_results(true, fit)

```

We can see the evaluation of the model from the function below. From the train model, the R² value will be 0.965 and for the test model it will be 0.881.

```

summary_MLR <- rbind(summary_MLR3_train[-5], summary_MLR3_test[-5])
rownames(summary_MLR) <- c("Baseline MLR_Train", "Baseline MLR_Test")
knitr::kable(summary_MLR, digits = 3)

```

	MSE	MAE	RMSE	Rsquare
Baseline MLR_Train	3.34	1.14	1.83	0.965
Baseline MLR_Test	3.60	1.29	1.90	0.881

Next we will be performing ridge and lasso regression model using the following code. We will split the data to train and test.

```

# Training Data
data <- training
train.x <- data.matrix(data[, -ncol(data)])
train.y <- data.matrix(data[, ncol(data)])

# Test Data
data <- test
test.x <- data.matrix(data[, -ncol(data)])
test.y <- data.matrix(data[, ncol(data)])

```

Next we will perform the Ridge regression model.

```

# Fit Ridge Regression Model
set.seed(123)
cv_ridge <- cv.glmnet(train.x, train.y, alpha = 0)
glm_ridge <- glmnet(train.x, train.y, alpha = 0, lambda = cv_ridge$lambda.min)

# Evaluate Ridge Regression Model on Training Set
fit <- predict(glm_ridge, newx = train.x, type = "response")
true <- train.y
summary_ridge_train <- eval_results(true, fit)

# Evaluate Ridge Regression Model on Test Set
fit <- predict(glm_ridge, newx = test.x, type = "response")
true <- test.y
summary_ridge_test <- eval_results(true, fit)

# Summarise Ridge Regression Model Performance
# into Table
summary_ridge <- rbind(summary_ridge_train, summary_ridge_test)
rownames(summary_ridge) <- c("Ridge Model_Train", "Ridge Model_Test")

```

Next we will perform Lasso Regression model

```
# Fit LASSO Model
set.seed(123)
cv_lasso <- cv.glmnet(train.x, train.y, alpha = 1)
glm_lasso <- glmnet(train.x, train.y, alpha = 1, lambda = cv_lasso$lambda.min)

# Evaluate LASSO Model on Training Set
fit <- predict(glm_lasso, newx = train.x)
true <- train.y
summary_lasso_train <- eval_results(true, fit)

# Evaluate LASSO Model on Test Set
fit <- predict(glm_lasso, newx = test.x)
true <- test.y
summary_lasso_test <- eval_results(true, fit)

# Summarise LASSO Model Performance into Table
summary_lasso <- rbind(summary_lasso_train, summary_lasso_test)
rownames(summary_lasso) <- c("LASSO Model_Train", "LASSO Model_Test")
```

We will compare the 3 models that we have build, Multi linear ,Ridge and Lasso regression models. We evaluate the 3 models together and compare both train and test models of each of the 3 regression model.

```
summary_3models <- rbind(summary_MLR3_train, summary_ridge_train,
  summary_lasso_train, summary_MLR3_test, summary_ridge_test,
  summary_lasso_test)

rownames(summary_3models) <- c("Baseline MLR_Train",
  "Ridge Model_Train", "LASSO Model_Train", "Baseline MLR_Test",
  "Ridge Model_Test", "LASSO Model_Test")

knitr::kable(summary_3models[1:3, -5], digits = 3)
```

	MSE	MAE	RMSE	Rsquare
Baseline MLR_Train	3.336	1.137	1.826	0.965
Ridge Model_Train	0.221	0.356	0.470	0.862
LASSO Model_Train	0.216	0.352	0.465	0.865

```
knitr::kable(summary_3models[4:6, -5], digits = 3)
```

	MSE	MAE	RMSE	Rsquare
Baseline MLR_Test	3.598	1.293	1.897	0.881
Ridge Model_Test	0.171	0.318	0.414	0.863
LASSO Model_Test	0.159	0.311	0.398	0.873

From the 3 models we can see that the original multi linear regression model is the best one with the highest R^squared value. However the Lasso Regression Model has the lowest MSE, MAE and RMSE. We can conclude that the reason for the first model having the highest R^squared value might be due to overfitting as many values will get removed. Next we will build other models to see if they work better.

3.2. Selectively remove variables with high missing rate

```
needs("knitr", "readr", "tidyverse", "dplyr", "ggplot2",
  "stats", "e1071", "car", "glmnet",
  "corrplot", "plotmo")
```

```
countries <- read.csv("../data/countries.csv")
```

3.2.1. Model Fitting

A. Ordinary Linear Regression Previous analysis showed that of all the variables in the countries.csv dataset, many of them had a high missing rate of above 35%, an arbitrary percentage that our group has chosen to filter the predictor variables. This section attempts to build a model on variables that have missing rates of below 35%.

The previous analysis showed that the relations between the predicted variable and gdp.pc, gdp.dflt, and fdi are better fitted to logarithmic models. Hence, we transformed the data with the mutate() function before building the machine learning model.

```
subset0 <- subset(countries, select = c("country.code",
  "year", "edu.total", "fdi", "gcf", "gdp.dflt",
  "hlth", "income", "lbr.part", "mil", "pop.gwth.rural",
  "pop.gwth.urban", "pov", "trade", "unemp", "year",
  "gdp.pc"))

subset1_0 <- subset0 %>%
  mutate(lfdi = log(subset0$fdi)) %>%
  mutate(lgdp.pc = log(subset0$gdp.pc)) %>%
  mutate(lgdp.dflt = log(subset0$gdp.dflt))
```

subset1 contains the variables below 35% of missing values.

```
subset1 <- subset(subset1_0, select = c("edu.total",
  "lfdi", "gcf", "lgdp.dflt", "hlth", "income", "lbr.part",
  "mil", "pop.gwth.rural", "pop.gwth.urban", "pov",
  "trade", "unemp", "year", "lgdp.pc"))

subset1 <- na.omit(subset1)
```

We also performed standardisation to reduce the variance of our machine learning models. Below are the functions used to aid our standardisation step.

```
sd0 <- function(vct) {
  if (!is.numeric(vct)) {
    return(NA)
  }

  return(sd(vct, na.rm = T))
}
```

```

std0 <- function(vct, scl) {
  if (!is.numeric(vct)) {
    return(vct)
  }

  return(vct/scl)
}

scale <- function(data) {
  unlist(lapply(data, std0))
}

standardise <- function(data) {
  scaler <- scale(data)

  numCols <- which(unlist(lapply(data, is.numeric)))
  num <- as.data.frame(mapply(std0, vct = data[, numCols], scl = scaler[numCols], SIMPLIFY = T))
  fct <- data[, -numCols]
  return(list(cbind(fct, num), scaler))
}

df.countries0 <- standardise(subset1)[[1]]
scaler2 <- standardise(subset1)[[2]]
df.countries <- df.countries0

# Divide each feature/target by its standard
# deviation
head(df.countries)

```

	fct	edu.total	lfdi	gcf	lgdp.dflt	hlth	lbr.part	mil	pop.gwth.rural	
## 1	LM	2.25	9.91	5.55	1.1870	3.13	7.94	1.135		-1.62
## 2	LM	2.37	10.26	5.80	1.0971	2.87	4.59	1.161		-1.89
## 3	UM	2.11	10.92	4.46	0.0384	2.29	7.59	1.279		-1.88
## 4	UM	2.20	11.04	4.04	0.4020	2.49	7.11	1.158		-1.91
## 5	UM	2.48	10.96	4.06	-0.5254	2.22	7.39	0.999		-1.98
## 6	UM	2.60	10.98	3.94	0.3415	2.27	7.73	0.953		-1.82
##	pop.gwth.urban	pov	trade	unemp	year	lgdp.pc				
## 1		1.70	0.1202	1.45	3.59	383	5.84			
## 2		1.43	0.0656	1.50	3.20	384	6.34			
## 3		1.44	0.0656	1.62	3.04	385	6.72			
## 4		1.28	0.1093	1.60	4.10	385	6.78			
## 5		1.16	0.0109	1.52	3.91	386	6.66			
## 6		1.20	0.0437	1.66	3.10	386	6.77			

First, we split the data into training and testing data.

```

set.seed(123)
index <- sort(sample(x = nrow(df.countries), size = nrow(df.countries) *
  0.8))
train_1 <- subset1[index, ]
test_1 <- subset1[-index, ]

```

Our first attempt

```
# Everything
lm_1 <- lm(data = train_1, formula = pov ~ .)
summary(lm_1)

##
## Call:
## lm(formula = pov ~ ., data = train_1)
##
## Residuals:
##    Min     1Q Median     3Q    Max 
## -24.10  -2.65  -0.36   2.09  50.94 
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)    
## (Intercept) 205.02815  92.92122   2.21  0.02770 *  
## edu.total   -0.80558   0.19166  -4.20  3.0e-05 *** 
## lfdi        0.13301   0.17075   0.78  0.43626    
## gcf         -0.10351   0.03776  -2.74  0.00628 **  
## lgdp.df1t  -0.06451   0.24743  -0.26  0.79438    
## hlth        0.22709   0.14684   1.55  0.12246    
## incomeL     6.84962   2.36679   2.89  0.00393 **  
## incomeLM    -3.45936   1.48782  -2.33  0.02037 *  
## incomeUM    -3.45134   1.00494  -3.43  0.00063 *** 
## lbr.part    0.22187   0.03333   6.66  5.9e-11 *** 
## mil         -0.93851   0.21070  -4.45  9.9e-06 *** 
## pop.gwth.rural 0.63592   0.19567   3.25  0.00121 ** 
## pop.gwth.urban 1.92366   0.20369   9.44 < 2e-16 *** 
## trade       -0.01476   0.00553  -2.67  0.00776 **  
## unemp        0.28941   0.05994   4.83  1.7e-06 *** 
## year        -0.08979   0.04615  -1.95  0.05213 .  
## lgdp.pc      -3.51559   0.60014  -5.86  7.4e-09 *** 
## ---      
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 5.86 on 651 degrees of freedom
## Multiple R-squared:  0.588, Adjusted R-squared:  0.578 
## F-statistic: 58.1 on 16 and 651 DF, p-value: <2e-16
```

The summary of `lm_1` suggests that the model still contains predictor variables with $\text{pr}(>|t|) > 0.05$. We can imply that these variables do not provide a high significance to our model and are thus able to be removed. To avoid removing significant variables, we shall conduct the removal of the variables deliberately. In `lm_1`, we notice that the variable `lgdp.df1t` has a high $\text{Pr}(>|t|)$. Hence, we remove `lgdp.df1t`.

```
# remove hlth
lm_2 <- lm(data = train_1, formula = pov ~ . - lfdi -
            lgdp.df1t)
summary(lm_2)

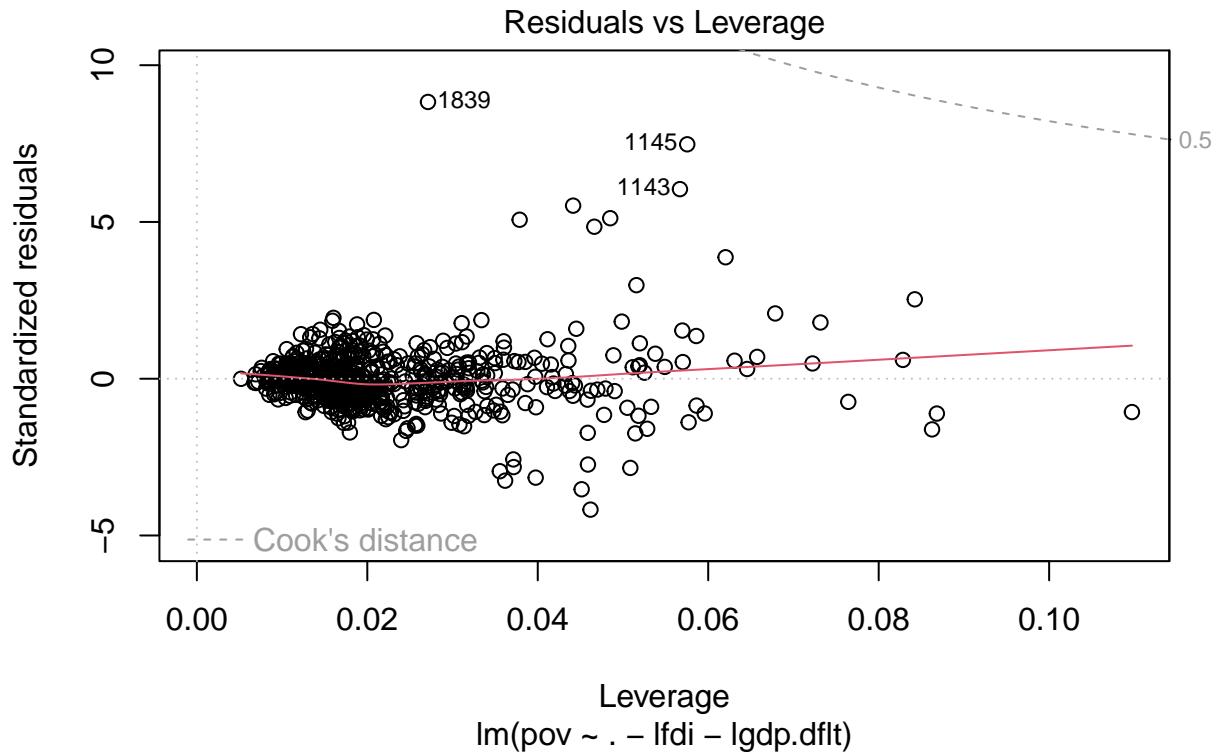
##
## Call:
## lm(formula = pov ~ . - lfdi - lgdp.df1t, data = train_1)
```

```

## 
## Residuals:
##   Min     1Q Median     3Q    Max 
## -23.88  -2.65  -0.37  2.10  50.98 
## 
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)    
## (Intercept) 206.1259   90.5962   2.28  0.02322 *  
## edu.total   -0.8282    0.1892  -4.38  1.4e-05 *** 
## gcf        -0.1015    0.0375  -2.71  0.00697 **  
## hlth       0.2235    0.1466   1.52  0.12777    
## incomeL     6.8791    2.3514   2.93  0.00356 **  
## incomeLM   -3.4336    1.4653  -2.34  0.01942 *  
## incomeUM   -3.4478    0.9778  -3.53  0.00045 *** 
## lbr.part    0.2239    0.0331   6.77  2.8e-11 *** 
## mil         -0.9289   0.2099  -4.42  1.1e-05 *** 
## pop.gwth.rural 0.6157   0.1938   3.18  0.00156 **  
## pop.gwth.urban 1.9411   0.2022   9.60 < 2e-16 *** 
## trade      -0.0156    0.0054  -2.89  0.00394 **  
## unemp       0.2851    0.0591   4.82  1.8e-06 *** 
## year        -0.0897   0.0451  -1.99  0.04716 *  
## lgdp.pc     -3.3417   0.5584  -5.98  3.6e-09 *** 
## --- 
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ',' 1 
## 
## Residual standard error: 5.86 on 653 degrees of freedom 
## Multiple R-squared:  0.588, Adjusted R-squared:  0.579 
## F-statistic: 66.5 on 14 and 653 DF, p-value: <2e-16 

plot(lm_2, which = 5)

```



```
# Remove lgdp.dflt
lm_3 <- lm(data = train_1, formula = pov ~ . - lgdp.dflt)
summary(lm_3)
```

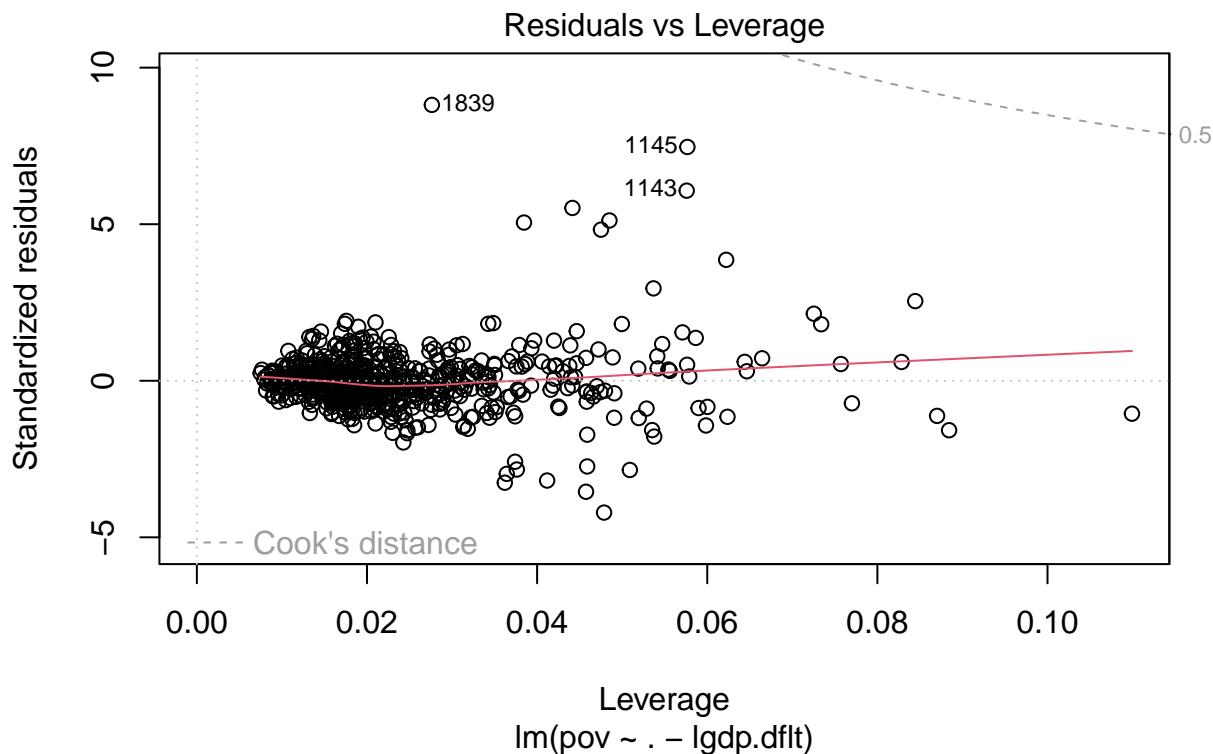
```
##
## Call:
## lm(formula = pov ~ . - lgdp.dflt, data = train_1)
##
## Residuals:
##      Min       1Q   Median       3Q      Max 
## -24.06    -2.71   -0.37    2.10   50.88 
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)    
## (Intercept) 200.16324  90.96351   2.20  0.02812 *  
## edu.total   -0.80991   0.19080  -4.24  2.5e-05 *** 
## lfdi        0.12942   0.17007   0.76  0.44694    
## gcf         -0.10409   0.03767  -2.76  0.00588 **  
## hlth        0.22762   0.14672   1.55  0.12129    
## incomeL     6.79249   2.35494   2.88  0.00405 **  
## incomeLM   -3.51750   1.46997  -2.39  0.01700 *  
## incomeUM   -3.50707   0.98125  -3.57  0.00038 *** 
## lbr.part    0.22282   0.03310   6.73  3.7e-11 *** 
## mil        -0.93973   0.21050  -4.46  9.5e-06 *** 
## pop.gwth.rural  0.63483  0.19548   3.25  0.00122 ** 
## pop.gwth.urban  1.92396  0.20354   9.45  < 2e-16 ***
```

```

## trade          -0.01474   0.00552   -2.67  0.00781  **
## unemp         0.29089   0.05962    4.88  1.3e-06 ***
## year          -0.08742   0.04521   -1.93  0.05361 .
## lgdp.pc       -3.50487   0.59830   -5.86  7.4e-09 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 5.86 on 652 degrees of freedom
## Multiple R-squared:  0.588, Adjusted R-squared:  0.579
## F-statistic:  62 on 15 and 652 DF, p-value: <2e-16

```

```
plot(lm_3, which = 5)
```



```

# remove lfdi
lm_4 <- lm(data = train_1, formula = pov ~ . - hlth -
lgdp.dflt - lfdi)
summary(lm_4)

##
## Call:
## lm(formula = pov ~ . - hlth - lgdp.dflt - lfdi, data = train_1)
##
## Residuals:
##     Min      1Q  Median      3Q     Max 
## -24.36   -2.61   -0.36    1.99   50.76

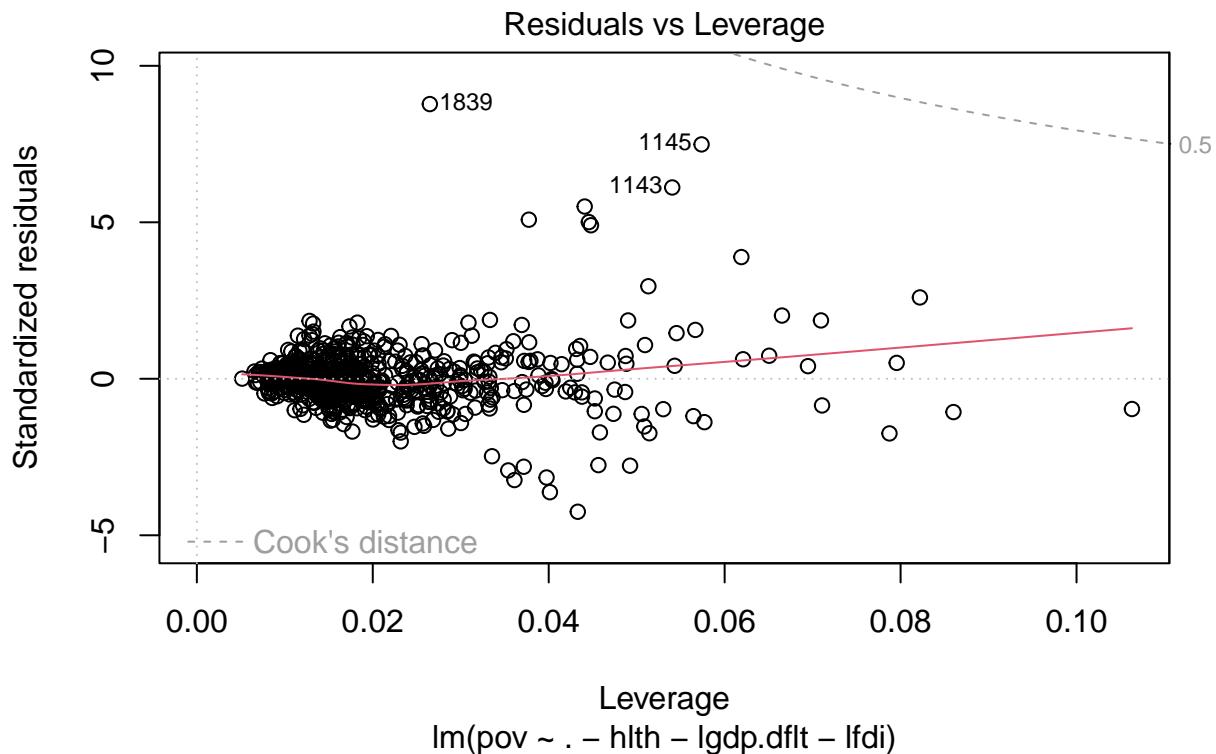
```

```

## 
## Coefficients:
##                               Estimate Std. Error t value Pr(>|t|)    
## (Intercept)      186.36637   89.75554   2.08  0.03825 *  
## edu.total       -0.73952    0.18026  -4.10  4.6e-05 *** 
## gcf            -0.11112    0.03701  -3.00  0.00278 **  
## incomeL          7.31879   2.33605   3.13  0.00181 **  
## incomeLM         -3.33101   1.46529  -2.27  0.02333 *  
## incomeUM         -3.57242   0.97541  -3.66  0.00027 *** 
## lbr.part          0.21730   0.03280   6.62  7.3e-11 *** 
## mil             -0.93100   0.21016  -4.43  1.1e-05 *** 
## pop.gwth.rural   0.57354   0.19201   2.99  0.00292 **  
## pop.gwth.urban   1.90680   0.20117   9.48 < 2e-16 *** 
## trade           -0.01696   0.00533  -3.18  0.00153 **  
## unemp            0.30422   0.05783   5.26  1.9e-07 *** 
## year            -0.07992   0.04469  -1.79  0.07416 .  
## lgdp.pc          -3.13114   0.54157  -5.78  1.1e-08 *** 
## --- 
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1 
## 
## Residual standard error: 5.86 on 654 degrees of freedom 
## Multiple R-squared:  0.586, Adjusted R-squared:  0.578 
## F-statistic: 71.3 on 13 and 654 DF, p-value: <2e-16

plot(lm_4, which = 5)

```



Our model shows that among all the 14 predictor variables that we started with, only 11 of them showed

significance in building the model.

We then evaluated the results of the model.

```
# evaluating lm_4
eval.metrics.linreg <- function(actual, predicted) {
  residual <- actual - predicted
  mse <- mean(residual^2)
  mae <- mean(abs(residual))
  rmse <- sqrt(mse)
  mape <- mean(abs(residual/actual)) * 100

  data.frame(MSE = mse, MAE = mae, RMSE = rmse, MAPE = mape)
}

actual <- train_1$pov
predicted <- predict(lm_4, newdata = train_1)
eval.metrics.linreg(actual, predicted)

##      MSE    MAE   RMSE   MAPE
## 1 33.6 3.56  5.8   Inf
```

The MAPE shows `Inf` due to the presence of 0 values in the Y variable.

B. Regularization Manual multiple regression models are prone to standard error. Hence, Ridge regression and LASSO regression are also used to build our model.

Next, we will perform Ridge regression and LASSO regression in a dataset that contains `country.code` and `year` in addition to the previous dataset that we used for the previous linear regression model.

```
# Subsetting the data

data0 <- subset(subset1_0, select = c("country.code",
  "year", "edu.total", "lfdi", "gcf", "lgdp.dflt",
  "hlth", "income", "lbr.part", "mil", "pop.gwth.rural",
  "pop.gwth.urban", "pov", "trade", "unemp", "year",
  "lgdp.pc"))
data <- na.omit(data0)
train.x <- data.matrix(subset(data, select = -pov))
train.y <- data.matrix(data$pov)
```

Splitting the aforementioned dataset into training and testing data.

```
split_train_test <- function(data) {
  set.seed(1984)
  # split data
  seeds <- data %>%
    group_by(country.code) %>%
    filter(row_number() == 1)

  plants <- data %>%
    group_by(country.code) %>%
    filter(row_number() != 1)
```

```

isComplete <- which(complete.cases(plants))
idx <- sample(isComplete, replace = F, 0.2 * nrow(plants))

training <- plants[-idx, ] %>%
  rbind(seeds)
test <- plants[idx, ]

return(list(training, test))
}

training <- split_train_test(data)[[1]]
test <- split_train_test(data)[[2]]

```

Further splitting the training and testing data into `train.x` and `train.y`.

```

ntrain <- 1:nrow(training)
combine <- rbind(training, test)
combine.mtrx <- model.matrix(pov ~ ., data = combine)
train.mtrx <- combine.mtrx[ntrain, ]
test.mtrx <- combine.mtrx[-ntrain, ]

train.x <- train.mtrx[, -1]
train.y <- training$pov

test.x <- test.mtrx[, -1]
test.y <- test$pov

```

Building a model with the lambda that outputs the smallest MSE.

```

cv_ridge <- cv.glmnet(train.x, train.y, alpha = 0)
glm_Ridge <- glmnet(train.x, train.y, alpha = 0, lambda = cv_ridge$lambda.min)

t(coef(glm_Ridge))

## 1 x 124 sparse Matrix of class "dgCMatrix"

##     [[ suppressing 124 column names '(Intercept)', 'country.codeARG', 'country.codeARM' ... ]]

##
## s0 321 -1.51 -3.41 -0.797 -0.142 -9.69 1.76 2.02 11.9 6.16 1.25 -1.41 -1.41
##
## s0 2.44 11.5 -2.1 -1.1 -1.15 6.28 12.5 30.2 2.46 -3.11 -1.42 -2.58 -1.26 0.328
##
## s0 0.411 -1.77 -0.63 -6.11 -1.3 -0.178 -0.297 -2.88 1.56 -1.16 3.21 26.5 2.65
##
## s0 -1.94 0.424 5.41 -0.831 0.303 5.48 4.93 -0.823 -1.58 -2.52 -0.13 -5.76 -5.15
##
## s0 -3.91 13.3 -5.24 -0.61 -2.78 0.744 0.085 0.694 -1.19 47.5 1.15 -3.21 0.00292
##
## s0 -2.52 -3.3 -2.78 -2.71 34.2 -2.74 1.73 -0.432 -1.86 0.693 2.56 -0.0818 -1.05
##
```

```

## s0 -1.23 -3.52 3.83 -3.12 7.75 25.4 8.68 -1.19 3.06 -1.49 -1.31 -0.139 23.9
##
## s0 -1.7 10.4 33.7 -4.52 1.77 -2.79 -3.6 20.8 16.7 -3.76 -1.91 2.53 0.494 17.4
##
## s0 48.6 -0.0739 -0.452 -0.26 -0.113 0.0536 -0.143 7.12 0.747 -0.899 0.0895
##
## s0 -0.105 0.449 0.839 -0.00776 0.0675 -0.075 -1.31

```

The ridge regression model shows that some countries have a higher influence on the model than others. At first glance, it seems like the Y variable, `pov`, depends heavily on whether or not it belongs to certain countries. To further illustrate the significance of `country.code`, we will plot the ridge regression graph for different lambda values.

```

# generate sequence of lambda values
lambda <- 10^seq(-6, 2, length = 100)
ridge_model = glmnet(train.x, train.y, alpha = 0, lambda = lambda)

head(t(coef(ridge_model)))

```

```

## 6 x 124 sparse Matrix of class "dgCMatrix"

## [[ suppressing 124 column names '(Intercept)', 'country.codeARG', 'country.codeARM' ... ]]

##
## s0  89 -0.166 -0.311 -0.292 -0.269 -0.505 -0.204 1.02 2.01 1.35 -0.1018 -0.357
## s1 101 -0.195 -0.370 -0.335 -0.301 -0.637 -0.212 1.15 2.32 1.54 -0.1052 -0.415
## s2 113 -0.229 -0.438 -0.382 -0.335 -0.801 -0.213 1.28 2.67 1.75 -0.1048 -0.479
## s3 127 -0.268 -0.517 -0.432 -0.369 -1.002 -0.206 1.40 3.05 1.97 -0.0989 -0.549
## s4 140 -0.312 -0.607 -0.486 -0.401 -1.246 -0.187 1.53 3.46 2.20 -0.0858 -0.625
## s5 155 -0.362 -0.708 -0.543 -0.432 -1.537 -0.155 1.63 3.90 2.44 -0.0639 -0.705
##
## s0 -0.00386 0.237 1.16 -0.322 -0.319 -0.180 0.568 1.89 3.73 0.373 -0.0710
## s1 -0.01595 0.284 1.37 -0.372 -0.365 -0.210 0.667 2.20 4.37 0.434 -0.0982
## s2 -0.03295 0.340 1.61 -0.429 -0.415 -0.245 0.781 2.54 5.10 0.503 -0.1340
## s3 -0.05587 0.404 1.89 -0.492 -0.469 -0.285 0.911 2.92 5.92 0.580 -0.1804
## s4 -0.08573 0.478 2.21 -0.561 -0.527 -0.329 1.059 3.33 6.84 0.665 -0.2397
## s5 -0.12339 0.562 2.57 -0.638 -0.587 -0.377 1.227 3.79 7.86 0.757 -0.3142
##
## s0 -0.155 -0.374 -0.340 -0.254 -0.250 -0.0975 -0.0537 -0.387 -0.263 -0.260
## s1 -0.185 -0.439 -0.392 -0.277 -0.275 -0.1252 -0.0706 -0.476 -0.299 -0.295
## s2 -0.219 -0.512 -0.449 -0.299 -0.298 -0.1599 -0.0917 -0.583 -0.340 -0.332
## s3 -0.260 -0.595 -0.510 -0.317 -0.318 -0.2027 -0.1178 -0.712 -0.383 -0.369
## s4 -0.306 -0.688 -0.575 -0.330 -0.333 -0.2548 -0.1492 -0.865 -0.429 -0.407
## s5 -0.360 -0.792 -0.644 -0.337 -0.344 -0.3172 -0.1861 -1.044 -0.479 -0.442
##
## s0 -0.279 -0.330 -0.214 -0.286 0.543 3.91 1.17 -0.290 0.390 0.745 -0.257 -0.250
## s1 -0.312 -0.393 -0.225 -0.323 0.630 4.55 1.33 -0.335 0.437 0.869 -0.291 -0.275
## s2 -0.345 -0.466 -0.229 -0.363 0.726 5.27 1.48 -0.384 0.483 1.008 -0.327 -0.300
## s3 -0.378 -0.550 -0.226 -0.404 0.834 6.08 1.65 -0.438 0.529 1.165 -0.365 -0.321
## s4 -0.409 -0.645 -0.212 -0.447 0.951 6.96 1.80 -0.498 0.571 1.339 -0.403 -0.338
## s5 -0.439 -0.752 -0.187 -0.492 1.079 7.93 1.95 -0.563 0.609 1.530 -0.442 -0.349
##
## s0 0.847 0.496 -0.305 -0.317 -0.335 -0.211 -0.369 -0.449 -0.360 2.46 -0.0967

```

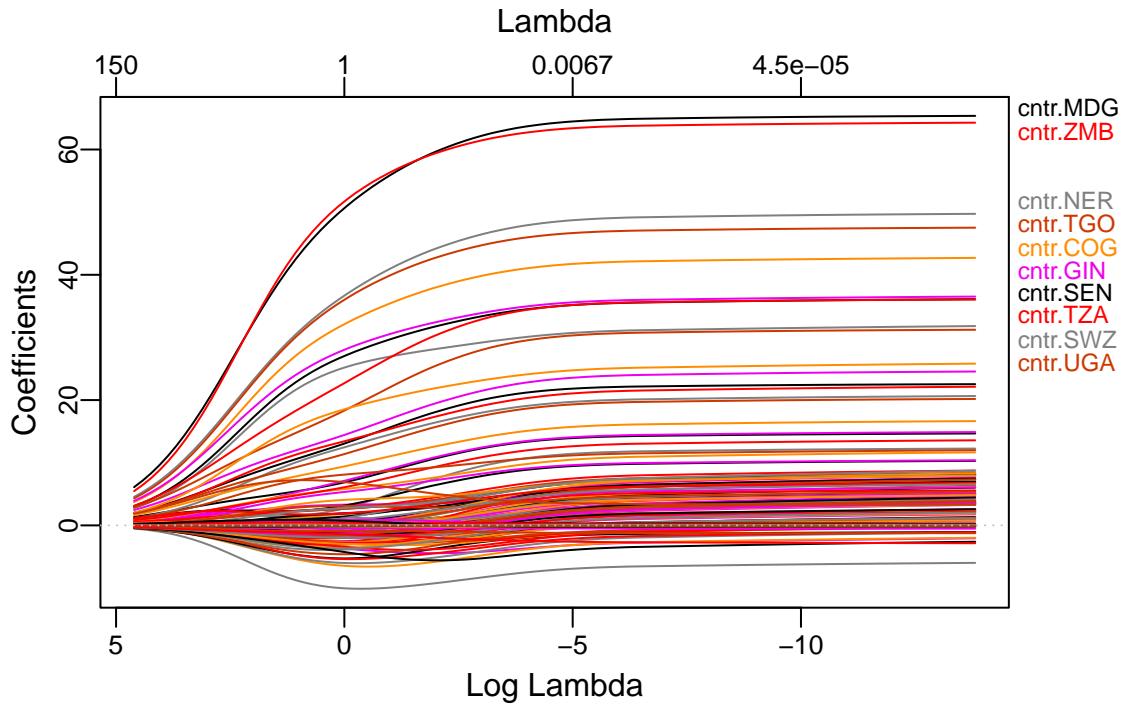
```

## s1 0.976 0.582 -0.349 -0.372 -0.391 -0.235 -0.455 -0.539 -0.438 2.83 -0.1469
## s2 1.119 0.680 -0.396 -0.436 -0.455 -0.259 -0.559 -0.645 -0.531 3.25 -0.2143
## s3 1.276 0.792 -0.447 -0.506 -0.527 -0.282 -0.684 -0.770 -0.641 3.71 -0.3027
## s4 1.446 0.920 -0.500 -0.584 -0.606 -0.303 -0.833 -0.914 -0.769 4.20 -0.4162
## s5 1.628 1.065 -0.556 -0.670 -0.695 -0.322 -1.007 -1.079 -0.917 4.72 -0.5585
##
## s0 -0.312 -0.168 -0.187 -0.270 -0.221 -0.157 6.10 0.0732 0.559 -0.261 -0.144
## s1 -0.360 -0.216 -0.204 -0.300 -0.247 -0.185 7.14 0.0852 0.600 -0.288 -0.186
## s2 -0.411 -0.274 -0.218 -0.330 -0.271 -0.216 8.31 0.0989 0.629 -0.315 -0.239
## s3 -0.466 -0.345 -0.228 -0.359 -0.294 -0.252 9.63 0.1146 0.641 -0.338 -0.302
## s4 -0.525 -0.429 -0.232 -0.385 -0.313 -0.291 11.11 0.1325 0.628 -0.358 -0.379
## s5 -0.585 -0.527 -0.229 -0.408 -0.326 -0.333 12.74 0.1530 0.585 -0.373 -0.468
##
## s0 -0.278 -0.375 -0.354 4.55 -0.139 -0.192 -0.302 0.473 0.012416 0.508 0.374
## s1 -0.346 -0.443 -0.417 5.31 -0.178 -0.195 -0.343 0.512 0.011038 0.584 0.415
## s2 -0.429 -0.520 -0.490 6.16 -0.227 -0.190 -0.387 0.545 0.008602 0.666 0.455
## s3 -0.529 -0.609 -0.573 7.12 -0.286 -0.175 -0.434 0.566 0.004916 0.754 0.493
## s4 -0.646 -0.708 -0.667 8.19 -0.359 -0.148 -0.482 0.573 -0.000151 0.849 0.526
## s5 -0.783 -0.819 -0.771 9.36 -0.445 -0.106 -0.530 0.561 -0.006624 0.949 0.553
##
## s0 -0.283 -0.275 -0.079 0.184 -0.359 1.11 3.06 1.65 0.0622 0.151 -0.307 -0.326
## s1 -0.322 -0.312 -0.116 0.227 -0.425 1.30 3.58 1.90 0.0639 0.194 -0.351 -0.373
## s2 -0.363 -0.351 -0.165 0.280 -0.501 1.50 4.19 2.18 0.0628 0.248 -0.400 -0.425
## s3 -0.407 -0.392 -0.229 0.344 -0.588 1.73 4.87 2.48 0.0582 0.315 -0.453 -0.480
## s4 -0.452 -0.435 -0.310 0.422 -0.685 1.99 5.64 2.81 0.0488 0.396 -0.509 -0.540
## s5 -0.497 -0.480 -0.410 0.517 -0.793 2.27 6.50 3.15 0.0336 0.493 -0.568 -0.602
##
## s0 -0.250 2.74 -0.333 1.97 4.31 -0.415 1.15 -0.284 -0.338 3.14 2.87 -0.377
## s1 -0.279 3.22 -0.390 2.27 5.04 -0.499 1.30 -0.337 -0.408 3.64 3.31 -0.446
## s2 -0.308 3.79 -0.454 2.60 5.87 -0.598 1.45 -0.398 -0.491 4.19 3.80 -0.524
## s3 -0.337 4.43 -0.527 2.95 6.80 -0.715 1.61 -0.468 -0.588 4.81 4.34 -0.613
## s4 -0.364 5.15 -0.607 3.33 7.84 -0.850 1.76 -0.548 -0.701 5.48 4.92 -0.714
## s5 -0.389 5.97 -0.695 3.72 8.98 -1.005 1.90 -0.638 -0.830 6.21 5.53 -0.826
##
## s0 -0.327 -0.0967 0.145 2.02 5.47 -0.0198 -0.141 -0.109 -0.00787 0.0953
## s1 -0.378 -0.0835 0.158 2.38 6.43 -0.0225 -0.160 -0.124 -0.00965 0.1052
## s2 -0.435 -0.0606 0.170 2.80 7.54 -0.0255 -0.180 -0.139 -0.01178 0.1149
## s3 -0.497 -0.0254 0.180 3.27 8.81 -0.0286 -0.202 -0.155 -0.01429 0.1237
## s4 -0.565 0.0243 0.187 3.81 10.23 -0.0318 -0.224 -0.171 -0.01722 0.1314
## s5 -0.638 0.0904 0.190 4.40 11.84 -0.0351 -0.247 -0.187 -0.02058 0.1376
##
## s0 -0.0827 1.86 0.243 -0.171 0.0223 -0.0277 0.153 0.247 -0.00299 -6.79e-05
## s1 -0.0925 2.14 0.270 -0.199 0.0256 -0.0332 0.173 0.283 -0.00342 5.60e-04
## s2 -0.1024 2.45 0.297 -0.232 0.0292 -0.0396 0.196 0.323 -0.00387 1.47e-03
## s3 -0.1122 2.79 0.323 -0.267 0.0332 -0.0470 0.219 0.365 -0.00436 2.71e-03
## s4 -0.1217 3.14 0.348 -0.307 0.0374 -0.0554 0.243 0.410 -0.00486 4.36e-03
## s5 -0.1305 3.51 0.370 -0.349 0.0418 -0.0648 0.267 0.457 -0.00537 6.46e-03
##
## s0 -0.0198 -0.309
## s1 -0.0225 -0.352
## s2 -0.0255 -0.399
## s3 -0.0286 -0.448
## s4 -0.0318 -0.499
## s5 -0.0351 -0.552

```

Plotting the coefficients against different lambda values.

```
plot_glmnet(ridge_model)
```



Interesting! For a low lambda value, the model exhibits `cntr.MDG`, which represents Madagascar, followed by `cntr.ZMB`, which represents Zimbabwe, as key predictor variables. This means that the model depends highly on whether the country is Madagascar or Zimbabwe, followed by some other significant countries. The other variables that previously defined the model is now undermined by the presence of the `country.code` variable and does not show high influence.

We then proceeded to evaluate the results

```
eval_results <- function(fit, true) {
  actual <- data.matrix(true)
  SSE <- sum((actual - fit)^2)
  SST <- sum((actual - mean(actual))^2)
  R_square <- 1 - SSE/SST
  data.frame(MSE = MSE(fit, true), MAE = MAE(fit,
    true), RMSE = RMSE(fit, true), MAPE = MAPE(fit,
    true), R2 = R_square)
}

fit <- predict(glm_Ridge, train.x)
true <- train.y
summary_Ridge_train <- eval_results(fit, true)
summary_Ridge_train
```

```
##      MSE  MAE RMSE MAPE      R2
## 1 5.37 1.47 2.32  Inf 0.943
```

The training dataset produced a much higher value compared to the regular multiple regression model shown above.

Fitting the model to the test dataset,

```
fit <- predict(glm_Ridge, test.x)
true <- test.y
summary_Ridge_test <- eval_results(fit, true)
summary_Ridge_test
```

```
##      MSE  MAE RMSE MAPE      R2
## 1 4.42 1.46 2.1  Inf 0.854
```

A slightly lower R^2 value, but still proves that the model is good enough.

Below is the summary of the results of the training and testing dataset.

```
summary_Ridge <- rbind(summary_Ridge_train, summary_Ridge_test)
rownames(summary_Ridge) <- c("Ridge_train", "Ridge_test")
knitr::kable(summary_Ridge, digits = 3)
```

	MSE	MAE	RMSE	MAPE	R2
Ridge_train	5.37	1.47	2.32	Inf	0.943
Ridge_test	4.42	1.46	2.10	Inf	0.854

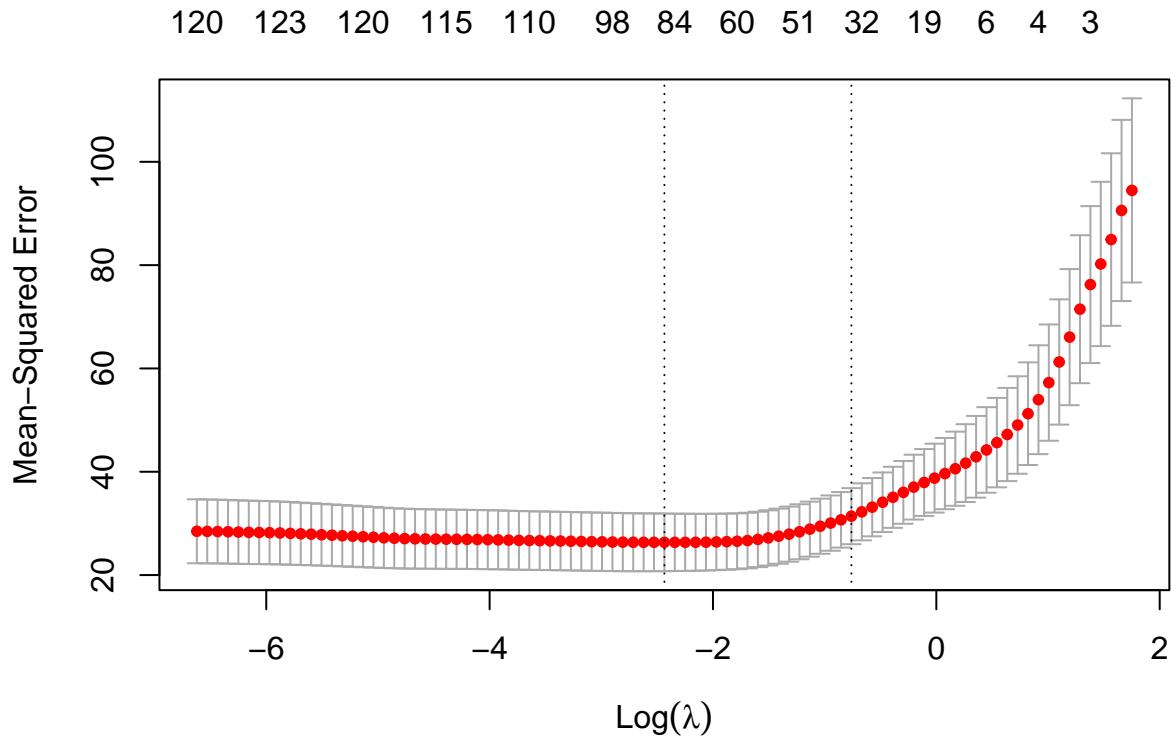
Next, we attempted LASSO regression as well. The steps are essentially similar to that of Ridge regression.

```
# Train a LASSO regression model
set.seed(123)
cv_lasso <- cv.glmnet(train.x, train.y, alpha = 1,
type.measure = "mse")
```

```
cv_lasso
```

```
##
## Call: cv.glmnet(x = train.x, y = train.y, type.measure = "mse", alpha = 1)
##
## Measure: Mean-Squared Error
##
##      Lambda Index Measure      SE Nonzero
## min  0.088     46    26.3 5.58       87
## 1se  0.467     28    31.4 5.43       35
```

```
plot(cv_lasso)
```



```
glm_Lasso <- glmnet(train.x, train.y, alpha = 1, lambda = cv_lasso$lambda.min)
t(coef(glm_Lasso))
```

```
## 1 x 124 sparse Matrix of class "dgCMatrix"

##      [[ suppressing 124 column names '(Intercept)', 'country.codeARG', 'country.codeARM' ... ]]

##
## s0 334 -0.186 -1.63 . . -8.04 0.257 3.91 14.4 8.92 0.288 -0.546 . 2.76 13.3
##
## s0 -0.0285 . . 7.76 16.7 37.8 3.78 . . -0.508 -0.043 . 0.0245 . 0.436 -3.59 . .
##
## s0 . . 0.0808 -0.35 4.28 31 4.01 . 1.46 6.74 . . 8.45 5.37 . . -0.762 . -2.42
##
## s0 -3.59 -1.25 16.8 -3.69 . . 0.15 . 0.668 -1.21 57.8 1.54 . . -0.0884 -0.749
##
## s0 -0.21 -1.3 41.3 . . 0.137 . 2.47 4.32 1.04 -0.338 -0.451 -0.535 4.25 -2.44
##
## s0 9.04 30.7 9.08 . 2.88 -0.195 . 0.0872 28.8 . 13 40.1 -3.46 2.34 -0.605
##
## s0 -0.878 26.2 22.5 -3.94 -0.723 0.778 1.03 20.9 58.4 -0.156 -0.299 . -0.135 .
##
## s0 . 6.27 . -1.06 0.0597 . 0.138 0.445 -0.000445 . -5.56e-05 -1.89
```

It is evident that the number of variables included in the model is less than that of ridge regression. This is in accordance with the nature of the LASSO regression model that reduces the coefficients of certain predictor variables to zero for model interpretability.

```

lambda <- 10^seq(-6, 2, length = 100)
lasso_model = glmnet(train.x, train.y, alpha = 1, lambda = lambda)

t(coef(lasso_model))

## 100 x 124 sparse Matrix of class "dgCMatrix"

##      [[ suppressing 124 column names '(Intercept)', 'country.codeARG', 'country.codeARM' ... ]]

###
## s0    4.48 .
## s1    4.48 .
## s2    4.48 .
## s3    4.48 .
## s4    4.48 .
## s5    4.48 .
## s6    4.48 .
## s7    4.48 .
## s8    4.48 .
## s9    4.48 .
## s10   4.48 .
## s11   4.48 .
## s12   4.48 .
## s13   4.48 .
## s14   4.48 .
## s15   4.48 .
## s16   5.82 .
## s17   9.75 .
## s18  12.64 .
## s19  14.35 .
## s20  15.78 .
## s21  16.86 .
## s22  17.76 .
## s23  18.45 .
## s24  18.94 .
## s25  18.26 .
## s26  17.81 .          -1.07 .
## s27  48.81 .         -3.07 .
## s28  89.47 .         -4.39 .
## s29 125.59 .         -5.43 .
## s30 157.00 .          -6.20 .     1.08 1.04 .
## s31 188.61 .        -0.0747 .    -6.75 .     3.61 2.57 .
## s32 217.31 .        -0.4939 .    -7.28 .     5.81 3.83 .
## s33 242.62 .        -0.8358 .    -7.69 .     7.70 4.93 .
## s34 264.06 .        -1.0871 .    -7.94 .     9.28 5.91 .
## s35 282.61 .        -1.3053 .    -8.14 .    0.384 10.74 6.76 .
## s36 303.23 .        -1.4421 .    -8.15 .    1.823 12.26 7.64 .
## s37 320.55 -0.0743 -1.5368 .    -8.13 0.0972 2.986 13.41 8.35 0.137
## s38 336.22 -0.2050 -1.6366 .    -8.04 0.2830 4.041 14.50 8.99 0.334

```

```

## s39 347.61 -0.3174 -1.8276 0.0571 .      -8.07 0.4608 4.889 15.42 9.47 0.424
## s40 355.96 -0.4230 -2.0245 0.2995 .      -8.12 0.6023 5.580 16.15 9.86 0.490
## s41 362.66 -0.5229 -2.2000 0.5324 0.0345 -8.13 0.7361 6.186 16.77 10.20 0.519
## s42 366.02 -0.6016 -2.4441 0.7958 0.0900 -8.14 0.9107 6.682 17.30 10.48 0.486
## s43 367.25 -0.6534 -2.6729 1.0241 0.1697 -8.17 1.1239 7.035 17.69 10.67 0.501
## s44 368.49 -0.6965 -2.8522 1.2297 0.2436 -8.18 1.2986 7.364 18.04 10.85 0.507
## s45 370.13 -0.7201 -2.9879 1.4143 0.3215 -8.17 1.4495 7.692 18.39 11.04 0.521
## s46 366.54 -0.6839 -3.1393 1.6930 0.4901 -8.18 1.6861 7.955 18.66 11.19 0.542
## s47 361.91 -0.5871 -3.4167 1.9229 0.6784 -8.25 1.9428 8.233 18.89 11.34 0.567
## s48 353.64 -0.4247 -3.6102 2.2473 0.9214 -8.26 2.2847 8.432 18.98 11.49 0.655
## s49 352.08 -0.3852 -3.6806 2.3765 1.0384 -8.26 2.4153 8.549 19.04 11.56 0.705
## s50 349.97 -0.3281 -3.7281 2.5483 1.1778 -8.23 2.5365 8.738 19.16 11.67 0.756
## s51 348.54 -0.1882 -3.7822 2.7703 1.3296 -8.18 2.6599 9.009 19.39 11.88 0.803
## s52 346.05 -0.0114 -3.8465 3.0518 1.5009 -8.10 2.8384 9.366 19.76 12.16 0.858
## s53 343.34  0.0241 -3.8957 3.2599 1.6245 -8.04 3.0017 9.625 20.01 12.36 0.914
## s54 339.37  0.1944 -3.9396 3.4985 1.7852 -7.99 3.2057 9.849 20.25 12.54 0.986
## s55 335.84  0.3816 -3.9706 3.7546 1.9649 -7.94 3.4091 10.069 20.48 12.73 1.050
## s56 334.96  0.4849 -3.9664 3.9015 2.0642 -7.86 3.5199 10.216 20.63 12.85 1.097
## s57 331.77  0.6280 -3.9699 4.1038 2.2163 -7.82 3.6801 10.380 20.81 12.99 1.157
## s58 328.75  0.8000 -3.9547 4.3135 2.3973 -7.75 3.8684 10.551 20.98 13.14 1.249
## s59 314.70  1.2831 -3.7448 4.9617 3.0484 -7.59 4.5312 10.774 21.20 13.35 1.589
## s60 314.74  1.3029 -3.7354 4.9868 3.0715 -7.58 4.5548 10.805 21.22 13.37 1.608
## s61 313.04  1.4398 -3.6634 5.1661 3.2397 -7.45 4.7257 10.971 21.38 13.51 1.724
## s62 310.97  1.5795 -3.5876 5.3518 3.4176 -7.34 4.9071 11.110 21.52 13.63 1.837
## s63 309.16  1.6982 -3.5188 5.5079 3.5704 -7.25 5.0579 11.222 21.64 13.73 1.931
## s64 307.33  1.8216 -3.4456 5.6667 3.7265 -7.16 5.2087 11.329 21.75 13.82 2.027
## s65 305.41  1.9471 -3.3664 5.8261 3.8858 -7.07 5.3609 11.430 21.85 13.92 2.126
## s66 303.45  2.0714 -3.2816 5.9846 4.0473 -6.98 5.5135 11.525 21.94 14.00 2.227
## s67 301.86  2.1694 -3.2124 6.1097 4.1768 -6.92 5.6350 11.596 22.01 14.07 2.308
## s68 300.68  2.2429 -3.1597 6.2034 4.2742 -6.86 5.7264 11.648 22.06 14.12 2.369
## s69 299.48  2.3165 -3.1061 6.2971 4.3722 -6.81 5.8180 11.700 22.10 14.17 2.430
## s70 298.29  2.3896 -3.0519 6.3901 4.4701 -6.76 5.9091 11.750 22.15 14.21 2.492
## s71 297.50  2.4378 -3.0158 6.4518 4.5350 -6.73 5.9698 11.783 22.18 14.24 2.532
## s72 297.10  2.4620 -2.9978 6.4828 4.5676 -6.71 6.0002 11.799 22.20 14.26 2.553
## s73 296.29  2.5106 -2.9611 6.5447 4.6332 -6.68 6.0612 11.831 22.23 14.29 2.594
## s74 295.90  2.5348 -2.9429 6.5756 4.6659 -6.66 6.0915 11.847 22.24 14.30 2.614
## s75 295.08  2.5828 -2.9064 6.6368 4.7311 -6.63 6.1519 11.879 22.27 14.33 2.655
## s76 294.69  2.6066 -2.8882 6.6671 4.7634 -6.61 6.1818 11.894 22.28 14.34 2.675
## s77 294.28  2.6302 -2.8701 6.6973 4.7957 -6.60 6.2116 11.909 22.30 14.36 2.695
## s78 293.88  2.6537 -2.8519 6.7273 4.8279 -6.58 6.2413 11.924 22.31 14.37 2.716
## s79 293.48  2.6770 -2.8339 6.7571 4.8598 -6.56 6.2708 11.939 22.33 14.39 2.736
## s80 293.08  2.7002 -2.8158 6.7866 4.8916 -6.55 6.3001 11.954 22.34 14.40 2.755
## s81 292.68  2.7230 -2.7978 6.8158 4.9231 -6.53 6.3292 11.968 22.35 14.41 2.775
## s82 292.29  2.7457 -2.7799 6.8448 4.9544 -6.52 6.3580 11.982 22.36 14.43 2.795
## s83 291.90  2.7681 -2.7620 6.8734 4.9854 -6.50 6.3865 11.995 22.38 14.44 2.814
## s84 291.51  2.7902 -2.7443 6.9017 5.0161 -6.48 6.4147 12.008 22.39 14.45 2.833
## s85 291.12  2.8121 -2.7266 6.9296 5.0466 -6.47 6.4427 12.021 22.40 14.46 2.852
## s86 290.74  2.8337 -2.7091 6.9572 5.0767 -6.45 6.4703 12.034 22.41 14.48 2.871
## s87 290.36  2.8549 -2.6917 6.9845 5.1065 -6.44 6.4977 12.046 22.42 14.49 2.890
## s88 289.98  2.8759 -2.6744 7.0114 5.1360 -6.42 6.5247 12.058 22.43 14.50 2.908
## s89 289.61  2.8966 -2.6573 7.0379 5.1652 -6.41 6.5514 12.069 22.44 14.51 2.926
## s90 289.24  2.9170 -2.6403 7.0641 5.1941 -6.39 6.5778 12.080 22.45 14.52 2.944
## s91 288.87  2.9372 -2.6234 7.0900 5.2227 -6.38 6.6039 12.091 22.46 14.53 2.962
## s92 288.51  2.9570 -2.6068 7.1155 5.2510 -6.37 6.6297 12.102 22.47 14.54 2.980

```

```

## s93 288.15 2.9765 -2.5903 7.1407 5.2789 -6.35 6.6552 12.112 22.48 14.55 2.997
## s94 287.79 2.9958 -2.5740 7.1655 5.3066 -6.34 6.6804 12.122 22.49 14.56 3.014
## s95 287.43 3.0148 -2.5578 7.1901 5.3339 -6.33 6.7053 12.132 22.50 14.57 3.031
## s96 287.08 3.0335 -2.5418 7.2142 5.3610 -6.31 6.7299 12.142 22.50 14.58 3.048
## s97 286.73 3.0520 -2.5260 7.2381 5.3877 -6.30 6.7542 12.151 22.51 14.59 3.065
## s98 286.38 3.0702 -2.5104 7.2617 5.4141 -6.29 6.7782 12.160 22.52 14.60 3.081
## s99 286.04 3.0881 -2.4949 7.2849 5.4402 -6.28 6.8020 12.169 22.53 14.60 3.097
##
## s0   .
## s1   .
## s2   .
## s3   .
## s4   .
## s5   .
## s6   .
## s7   .
## s8   .
## s9   .
## s10  .
## s11  .
## s12  .
## s13  .
## s14  .
## s15  .
## s16  .
## s17  .
## s18  .
## s19  .
## s20  .
## s21  .
## s22  .
## s23  .
## s24  .
## s25  .
## s26  .
## s27  .
## s28  .
## s29  .
## s30  .
## s31  .
## s32  .
## s33  .
## s34 -0.0467 .
## s35 -0.2615 .
## s36 -0.4141 .
## s37 -0.5064 .
## s38 -0.5539 .
## s39 -0.6430 .
## s40 -0.7119 .
## s41 -0.7963 .
## s42 -0.9299 .
## s43 -0.9908 .
## s44 -1.0479 .
## s45 -1.0950 .

```

```

## s46 -1.1115 .      3.596 16.66 -0.6970 0.3416 .      10.51 18.054 39.5 4.429
## s47 -1.1113 .      3.739 16.85 -0.5441 0.5999 .      10.73 18.172 39.6 4.423
## s48 -1.0626 0.0190 3.968 17.01 -0.2579 0.9582 .      11.08 18.314 39.7 4.515
## s49 -1.0378 0.0752 4.051 17.05 -0.1917 1.1080 .      11.20 18.368 39.8 4.539
## s50 -1.0246 0.1644 4.167 17.08 -0.0940 1.2843 0.0347 11.36 18.497 39.8 4.590
## s51 -0.9824 0.3243 4.336 17.27 .      1.4675 0.1498 11.57 18.703 39.9 4.695
## s52 -0.9307 0.5373 4.554 17.56 0.0250 1.6872 0.3077 11.85 18.981 40.1 4.855
## s53 -0.8941 0.6668 4.711 17.73 0.2225 1.8636 0.4206 12.09 19.182 40.3 4.974
## s54 -0.8462 0.7907 4.904 17.92 0.4635 2.0752 0.5596 12.34 19.368 40.4 5.107
## s55 -0.7886 0.9486 5.107 18.14 0.7140 2.2977 0.7103 12.56 19.556 40.6 5.252
## s56 -0.7479 1.0462 5.227 18.26 0.8518 2.4181 0.8038 12.70 19.659 40.7 5.348
## s57 -0.7052 1.1680 5.382 18.41 1.0459 2.5977 0.9288 12.86 19.810 40.8 5.462
## s58 -0.6163 1.3080 5.557 18.56 1.2626 2.8038 1.0656 13.03 19.968 40.9 5.583
## s59 -0.3047 1.6848 6.043 18.92 1.9470 3.5232 1.5007 13.38 20.286 41.1 5.946
## s60 -0.2852 1.7112 6.064 18.94 1.9734 3.5486 1.5204 13.40 20.308 41.1 5.966
## s61 -0.1552 1.8816 6.213 19.08 2.1513 3.7333 1.6577 13.55 20.461 41.3 6.098
## s62 -0.0351 2.0467 6.364 19.22 2.3365 3.9250 1.8009 13.68 20.601 41.4 6.230
## s63  0.0441 2.1818 6.490 19.33 2.4922 4.0886 1.9216 13.79 20.721 41.5 6.339
## s64  0.1459 2.3161 6.618 19.44 2.6504 4.2540 2.0445 13.89 20.837 41.6 6.451
## s65  0.2506 2.4483 6.747 19.54 2.8101 4.4222 2.1690 13.99 20.950 41.7 6.564
## s66  0.3561 2.5801 6.875 19.64 2.9693 4.5923 2.2927 14.08 21.061 41.8 6.675
## s67  0.4405 2.6840 6.975 19.71 3.0956 4.7286 2.3902 14.15 21.147 41.8 6.763
## s68  0.5040 2.7614 7.051 19.77 3.1901 4.8309 2.4636 14.21 21.211 41.9 6.829
## s69  0.5678 2.8385 7.126 19.82 3.2846 4.9336 2.5370 14.26 21.274 41.9 6.894
## s70  0.6312 2.9150 7.200 19.88 3.3785 5.0360 2.6101 14.31 21.335 42.0 6.960
## s71  0.6731 2.9655 7.249 19.91 3.4408 5.1039 2.6585 14.34 21.376 42.0 7.003
## s72  0.6943 2.9909 7.274 19.93 3.4721 5.1381 2.6828 14.36 21.396 42.0 7.024
## s73  0.7359 3.0408 7.323 19.96 3.5345 5.2064 2.7313 14.39 21.436 42.1 7.067
## s74  0.7568 3.0658 7.347 19.98 3.5657 5.2405 2.7555 14.41 21.455 42.1 7.088
## s75  0.7985 3.1154 7.396 20.01 3.6274 5.3085 2.8036 14.44 21.495 42.1 7.131
## s76  0.8193 3.1401 7.420 20.03 3.6580 5.3422 2.8274 14.45 21.514 42.1 7.152
## s77  0.8400 3.1646 7.444 20.05 3.6885 5.3758 2.8511 14.47 21.533 42.1 7.173
## s78  0.8606 3.1891 7.467 20.06 3.7189 5.4093 2.8748 14.48 21.553 42.2 7.193
## s79  0.8812 3.2133 7.491 20.08 3.7490 5.4426 2.8982 14.50 21.571 42.2 7.214
## s80  0.9016 3.2374 7.514 20.09 3.7788 5.4757 2.9215 14.51 21.590 42.2 7.235
## s81  0.9218 3.2612 7.537 20.11 3.8084 5.5086 2.9445 14.53 21.609 42.2 7.255
## s82  0.9419 3.2848 7.560 20.12 3.8377 5.5411 2.9673 14.54 21.627 42.2 7.275
## s83  0.9618 3.3081 7.583 20.14 3.8667 5.5733 2.9899 14.56 21.644 42.2 7.295
## s84  0.9816 3.3312 7.605 20.15 3.8953 5.6052 3.0122 14.57 21.662 42.2 7.314
## s85  1.0011 3.3540 7.627 20.17 3.9237 5.6368 3.0343 14.58 21.679 42.3 7.334
## s86  1.0204 3.3765 7.649 20.18 3.9516 5.6680 3.0561 14.60 21.696 42.3 7.353
## s87  1.0395 3.3987 7.670 20.20 3.9793 5.6990 3.0776 14.61 21.712 42.3 7.372
## s88  1.0583 3.4205 7.691 20.21 4.0066 5.7295 3.0989 14.62 21.729 42.3 7.390
## s89  1.0770 3.4421 7.712 20.22 4.0335 5.7598 3.1199 14.63 21.744 42.3 7.409
## s90  1.0954 3.4634 7.732 20.24 4.0601 5.7896 3.1406 14.64 21.760 42.3 7.427
## s91  1.1136 3.4843 7.752 20.25 4.0864 5.8192 3.1610 14.66 21.775 42.3 7.444
## s92  1.1316 3.5050 7.772 20.26 4.1123 5.8484 3.1812 14.67 21.790 42.4 7.462
## s93  1.1493 3.5253 7.792 20.28 4.1379 5.8773 3.2011 14.68 21.805 42.4 7.479
## s94  1.1669 3.5453 7.811 20.29 4.1632 5.9058 3.2207 14.69 21.819 42.4 7.496
## s95  1.1842 3.5650 7.830 20.30 4.1881 5.9340 3.2401 14.70 21.833 42.4 7.513
## s96  1.2013 3.5844 7.848 20.31 4.2127 5.9619 3.2592 14.71 21.847 42.4 7.529
## s97  1.2181 3.6035 7.867 20.32 4.2370 5.9895 3.2781 14.72 21.860 42.4 7.546
## s98  1.2348 3.6224 7.885 20.34 4.2610 6.0167 3.2967 14.73 21.874 42.4 7.562
## s99  1.2512 3.6409 7.903 20.35 4.2847 6.0436 3.3150 14.74 21.887 42.4 7.577

```

```

## 
## s0 .
## s1 .
## s2 .
## s3 .
## s4 .
## s5 .
## s6 .
## s7 .
## s8 .
## s9 .
## s10 .
## s11 .
## s12 .
## s13 .
## s14 .
## s15 .
## s16 .
## s17 .
## s18 .
## s19 .
## s20 .
## s21 .
## s22 .
## s23 .
## s24 .
## s25 .
## s26 .
## s27 .
## s28 .
## s29 .
## s30 .
## s31 .           -0.682 .
## s32 .           -1.395 .
## s33 .           -1.973 .
## s34 .           -2.481 .
## s35 .           -2.857 .
## s36 .           -0.13965 .           -3.109 .
## s37 .           -0.0134 -0.36798 .           0.149 -3.381 .
## s38 .           .           -0.54504 -0.0558 .           0.0277 .           0.479 -3.626 .
## s39 .           .           -0.66000 -0.2087 .           0.1869 .           0.730 -3.870 .
## s40 .           .           -0.75240 -0.3299 .           0.3192 .           0.940 -4.103 .
## s41 .           .           -0.81104 -0.4230 .           0.4142 .           1.105 -4.294 .
## s42 .           .           -0.83886 -0.5214 .           0.5088 .           1.190 -4.476 .
## s43 .           .           -0.85774 -0.5806 .           0.6580 .           1.230 -4.682 .
## s44 .           .           -0.86095 -0.6244 .           0.7815 .           1.276 -4.841 .
## s45 .           .           -0.84632 -0.6446 0.0739 0.8803 .           1.337 -4.923 .
## s46 .           0.1184 -0.72309 -0.6045 0.2585 1.0966 .           1.414 -4.998 .
## s47 .           0.3960 -0.62126 -0.5610 0.4384 1.3161 0.0896 1.418 -5.047 0.109
## s48 .           0.7121 -0.48336 -0.4525 0.7281 1.5892 0.3296 1.420 -5.086 0.383
## s49 .           0.8009 -0.43423 -0.4069 0.8507 1.7074 0.3950 1.444 -5.122 0.473
## s50 0.0156 0.9029 -0.35624 -0.3434 1.0016 1.8537 0.4813 1.513 -5.127 0.599
## s51 0.1688 1.1142 -0.22317 -0.2736 1.1411 2.0289 0.6496 1.633 -5.037 0.766
## s52 0.3239 1.3941 -0.03068 -0.1814 1.3105 2.2428 0.8724 1.793 -4.883 0.978

```

```

## s53 0.4012  1.5914  0.00117 -0.1067  1.4567  2.3978  1.0427  1.890 -4.796 1.140
## s54 0.4675  1.8277  0.16214 -0.0137  1.6429  2.5833  1.2314  1.992 -4.696 1.336
## s55 0.5685  2.0641  0.35939  0.0272  1.8293  2.7925  1.4098  2.111 -4.564 1.539
## s56 0.6604  2.1765  0.48062  0.1052  1.9339  2.9052  1.5024  2.184 -4.476 1.662
## s57 0.7212  2.3474  0.64056  0.2079  2.0920  3.0740  1.6348  2.291 -4.367 1.825
## s58 0.8107  2.5471  0.81244  0.3373  2.2772  3.2664  1.7931  2.401 -4.248 2.003
## s59 0.9755  3.0374  1.39545  0.8018  2.9306  3.9594  2.1605  2.746 -3.952 2.542
## s60 0.9933  3.0588  1.42036  0.8236  2.9538  3.9856  2.1803  2.767 -3.942 2.560
## s61 1.0883  3.2106  1.60886  0.9722  3.1167  4.1722  2.3097  2.902 -3.816 2.705
## s62 1.1730  3.3543  1.80558  1.1244  3.2885  4.3647  2.4260  3.037 -3.687 2.857
## s63 1.2461  3.4715  1.97088  1.2528  3.4351  4.5258  2.5215  3.151 -3.576 2.986
## s64 1.3113  3.5898  2.13495  1.3812  3.5856  4.6893  2.6183  3.268 -3.464 3.116
## s65 1.3789  3.7075  2.29969  1.5124  3.7392  4.8554  2.7153  3.385 -3.354 3.247
## s66 1.4474  3.8214  2.46387  1.6448  3.8947  5.0226  2.8100  3.501 -3.244 3.378
## s67 1.5008  3.9096  2.59398  1.7507  4.0192  5.1562  2.8831  3.593 -3.159 3.480
## s68 1.5409  3.9751  2.69144  1.8304  4.1130  5.2565  2.9378  3.662 -3.096 3.558
## s69 1.5809  4.0403  2.78910  1.9106  4.2071  5.3570  2.9923  3.731 -3.031 3.635
## s70 1.6203  4.1045  2.88630  1.9905  4.3009  5.4572  3.0458  3.800 -2.967 3.711
## s71 1.6463  4.1465  2.95067  2.0434  4.3633  5.5237  3.0809  3.845 -2.925 3.762
## s72 1.6597  4.1676  2.98300  2.0701  4.3946  5.5571  3.0986  3.868 -2.904 3.787
## s73 1.6860  4.2096  3.04754  2.1233  4.4575  5.6239  3.1335  3.913 -2.862 3.839
## s74 1.6991  4.2304  3.07974  2.1499  4.4888  5.6573  3.1509  3.935 -2.841 3.864
## s75 1.7246  4.2718  3.14397  2.2030  4.5511  5.7236  3.1855  3.980 -2.799 3.914
## s76 1.7373  4.2922  3.17584  2.2293  4.5820  5.7566  3.2025  4.003 -2.778 3.939
## s77 1.7499  4.3125  3.20764  2.2556  4.6128  5.7895  3.2194  4.025 -2.758 3.964
## s78 1.7624  4.3326  3.23932  2.2818  4.6435  5.8222  3.2362  4.047 -2.737 3.989
## s79 1.7747  4.3525  3.27084  2.3078  4.6739  5.8548  3.2528  4.069 -2.717 4.013
## s80 1.7869  4.3722  3.30213  2.3337  4.7042  5.8871  3.2692  4.090 -2.697 4.038
## s81 1.7990  4.3916  3.33318  2.3594  4.7343  5.9192  3.2854  4.112 -2.677 4.062
## s82 1.8109  4.4108  3.36397  2.3849  4.7640  5.9511  3.3013  4.133 -2.657 4.085
## s83 1.8227  4.4296  3.39448  2.4101  4.7935  5.9827  3.3170  4.154 -2.637 4.109
## s84 1.8344  4.4482  3.42470  2.4352  4.8227  6.0140  3.3324  4.175 -2.618 4.132
## s85 1.8459  4.4664  3.45460  2.4599  4.8516  6.0449  3.3475  4.196 -2.599 4.155
## s86 1.8572  4.4843  3.48419  2.4844  4.8802  6.0756  3.3624  4.216 -2.580 4.178
## s87 1.8684  4.5019  3.51346  2.5087  4.9085  6.1060  3.3770  4.236 -2.561 4.200
## s88 1.8794  4.5192  3.54239  2.5327  4.9365  6.1360  3.3913  4.256 -2.543 4.222
## s89 1.8902  4.5362  3.57100  2.5565  4.9641  6.1657  3.4053  4.275 -2.525 4.244
## s90 1.9009  4.5529  3.59927  2.5800  4.9915  6.1951  3.4191  4.295 -2.507 4.265
## s91 1.9114  4.5693  3.62720  2.6032  5.0185  6.2242  3.4327  4.314 -2.490 4.287
## s92 1.9218  4.5854  3.65481  2.6261  5.0452  6.2529  3.4460  4.332 -2.472 4.308
## s93 1.9320  4.6012  3.68209  2.6488  5.0717  6.2813  3.4590  4.351 -2.455 4.328
## s94 1.9421  4.6168  3.70903  2.6713  5.0978  6.3094  3.4718  4.369 -2.439 4.349
## s95 1.9520  4.6321  3.73565  2.6935  5.1236  6.3371  3.4844  4.387 -2.422 4.369
## s96 1.9618  4.6471  3.76195  2.7154  5.1491  6.3645  3.4968  4.405 -2.406 4.389
## s97 1.9714  4.6618  3.78793  2.7371  5.1743  6.3917  3.5089  4.422 -2.390 4.408
## s98 1.9809  4.6764  3.81359  2.7585  5.1992  6.4185  3.5209  4.439 -2.375 4.427
## s99 1.9902  4.6906  3.83894  2.7797  5.2238  6.4450  3.5326  4.456 -2.359 4.447
##
## s0   .
## s1   .
## s2   .
## s3   .
## s4   .
## s5   .

```

```

## s6 .
## s7 .
## s8 .
## s9 .
## s10 .
## s11 .
## s12 .
## s13 .
## s14 .
## s15 .
## s16 .
## s17 .
## s18 .
## s19 .
## s20 .
## s21 .
## s22 .
## s23 .
## s24 . . . . . 2.16 .
## s25 . . . . . 6.94 .
## s26 . . . . . 10.90 .
## s27 . . . . . 14.11 .
## s28 . . . . . 0.675 16.79 .
## s29 . . . . . 1.383 19.17 .
## s30 . . . . . 1.974 21.24 .
## s31 . . . . . 2.466 23.15 .
## s32 . . . . . 2.816 24.75 .
## s33 . . . . . 3.107 26.11 .
## s34 . . . . . 3.392 27.25 .
## s35 . . . . . 3.615 28.30 0.776 .
## s36 . . . . . 3.870 29.42 2.122 .
## s37 . . . . -0.1529 4.115 30.30 3.195 .
## s38 . . . . 0.134 -0.3886 4.311 31.07 4.131 .
## s39 . . -0.24662 0.449 -0.5477 4.366 31.66 4.862 -0.2293 1.735 7.05
## s40 0.00886 . -0.54743 0.697 -0.6714 4.383 32.13 5.451 -0.4977 1.935 7.26
## s41 0.04371 . -0.81236 0.905 -0.7640 4.387 32.52 5.944 -0.7367 2.099 7.43
## s42 0.01639 0.027 -1.08470 1.114 -0.7802 4.320 32.79 6.278 -0.9674 2.187 7.56
## s43 0.03029 0.150 -1.31445 1.340 -0.7406 4.272 32.98 6.499 -1.1035 2.226 7.64
## s44 0.04393 0.254 -1.50457 1.529 -0.7059 4.237 33.17 6.716 -1.2166 2.272 7.71
## s45 0.05947 0.341 -1.63449 1.695 -0.6650 4.217 33.37 6.959 -1.3081 2.352 7.80
## s46 0.11364 0.520 -1.74088 1.913 -0.5037 4.174 33.52 7.122 -1.3265 2.448 7.89
## s47 0.13815 0.697 -1.77086 2.091 -0.3991 4.056 33.54 7.269 -1.3910 2.608 8.00
## s48 0.20873 0.953 -1.81640 2.357 -0.1503 4.030 33.52 7.322 -1.3324 2.775 8.09
## s49 0.26525 1.056 -1.88351 2.475 -0.0839 4.038 33.55 7.379 -1.3557 2.856 8.13
## s50 0.30334 1.202 -1.92956 2.630 . 4.051 33.66 7.470 -1.3407 2.981 8.19
## s51 0.35227 1.372 -1.88442 2.784 0.0268 4.056 33.79 7.641 -1.2856 3.170 8.31
## s52 0.40153 1.579 -1.76617 2.977 0.2445 4.051 33.99 7.877 -1.1942 3.408 8.50
## s53 0.43684 1.735 -1.69206 3.139 0.4286 4.059 34.15 8.035 -1.1189 3.572 8.62
## s54 0.47986 1.917 -1.60593 3.334 0.6465 4.070 34.29 8.192 -1.0151 3.746 8.77
## s55 0.54462 2.111 -1.50965 3.523 0.8691 4.082 34.41 8.345 -0.9084 3.926 8.92
## s56 0.60254 2.223 -1.46032 3.634 0.9945 4.115 34.49 8.436 -0.8524 4.023 8.99
## s57 0.65927 2.384 -1.37398 3.795 1.1764 4.132 34.61 8.558 -0.7505 4.169 9.11
## s58 0.75237 2.564 -1.27594 3.970 1.3701 4.183 34.71 8.696 -0.6339 4.338 9.25
## s59 1.14479 3.179 -1.00153 4.554 2.0296 4.408 34.93 8.920 -0.1782 4.744 9.61

```

```

## s60 1.16751 3.201 -0.99555 4.577 2.0537 4.426 34.96 8.947 -0.1680 4.768 9.63
## s61 1.30664 3.369 -0.87854 4.727 2.2283 4.519 35.10 9.073 -0.0549 4.918 9.78
## s62 1.44631 3.541 -0.76097 4.885 2.4111 4.606 35.22 9.181 0.0335 5.061 9.92
## s63 1.56368 3.686 -0.66322 5.018 2.5647 4.680 35.33 9.275 0.1410 5.185 10.03
## s64 1.68117 3.833 -0.56776 5.154 2.7220 4.757 35.43 9.371 0.2525 5.310 10.15
## s65 1.80298 3.982 -0.47250 5.292 2.8810 4.839 35.53 9.464 0.3677 5.433 10.27
## s66 1.92688 4.132 -0.38020 5.431 3.0399 4.924 35.63 9.552 0.4832 5.556 10.38
## s67 2.02643 4.251 -0.30940 5.541 3.1658 4.994 35.71 9.620 0.5745 5.653 10.47
## s68 2.10167 4.340 -0.25626 5.624 3.2603 5.046 35.76 9.670 0.6436 5.725 10.53
## s69 2.17745 4.430 -0.20275 5.707 3.3550 5.099 35.82 9.720 0.7129 5.797 10.60
## s70 2.25326 4.519 -0.14963 5.789 3.4491 5.152 35.87 9.769 0.7819 5.869 10.67
## s71 2.30359 4.578 -0.11512 5.844 3.5116 5.187 35.91 9.802 0.8278 5.916 10.71
## s72 2.32896 4.608 -0.09806 5.872 3.5430 5.205 35.93 9.818 0.8509 5.940 10.73
## s73 2.37951 4.667 -0.06293 5.927 3.6056 5.240 35.96 9.850 0.8972 5.986 10.77
## s74 2.40483 4.697 -0.04573 5.954 3.6368 5.258 35.98 9.866 0.9203 6.010 10.80
## s75 2.45538 4.756 -0.01060 6.009 3.6989 5.293 36.01 9.898 0.9660 6.056 10.84
## s76 2.48057 4.785 0.00272 6.036 3.7297 5.311 36.03 9.913 0.9887 6.079 10.86
## s77 2.50573 4.815 0.02043 6.062 3.7604 5.328 36.05 9.928 1.0113 6.102 10.88
## s78 2.53084 4.844 0.03805 6.089 3.7909 5.345 36.06 9.943 1.0338 6.125 10.90
## s79 2.55586 4.873 0.05550 6.115 3.8212 5.363 36.08 9.958 1.0561 6.147 10.92
## s80 2.58075 4.901 0.07271 6.142 3.8513 5.380 36.10 9.973 1.0782 6.170 10.94
## s81 2.60551 4.930 0.08971 6.168 3.8811 5.397 36.11 9.987 1.1002 6.192 10.96
## s82 2.63012 4.958 0.10649 6.193 3.9107 5.414 36.13 10.001 1.1219 6.214 10.98
## s83 2.65456 4.986 0.12303 6.219 3.9399 5.431 36.14 10.015 1.1434 6.235 11.00
## s84 2.67882 5.013 0.13932 6.244 3.9689 5.448 36.16 10.029 1.1648 6.256 11.02
## s85 2.70290 5.041 0.15537 6.269 3.9975 5.465 36.17 10.042 1.1858 6.277 11.04
## s86 2.72677 5.068 0.17117 6.293 4.0258 5.481 36.19 10.055 1.2067 6.297 11.06
## s87 2.75043 5.094 0.18671 6.317 4.0538 5.498 36.20 10.067 1.2273 6.318 11.08
## s88 2.77388 5.121 0.20202 6.341 4.0814 5.514 36.22 10.080 1.2477 6.337 11.10
## s89 2.79712 5.147 0.21708 6.365 4.1087 5.530 36.23 10.092 1.2679 6.357 11.11
## s90 2.82013 5.173 0.23190 6.388 4.1356 5.546 36.24 10.104 1.2878 6.376 11.13
## s91 2.84292 5.198 0.24649 6.411 4.1622 5.561 36.26 10.115 1.3075 6.395 11.15
## s92 2.86548 5.224 0.26086 6.434 4.1885 5.577 36.27 10.127 1.3269 6.414 11.17
## s93 2.88782 5.249 0.27500 6.457 4.2145 5.592 36.28 10.138 1.3462 6.432 11.18
## s94 2.90993 5.273 0.28893 6.479 4.2401 5.607 36.30 10.149 1.3652 6.450 11.20
## s95 2.93180 5.298 0.30265 6.501 4.2654 5.622 36.31 10.159 1.3840 6.468 11.21
## s96 2.95346 5.322 0.31616 6.523 4.2903 5.637 36.32 10.170 1.4025 6.485 11.23
## s97 2.97488 5.346 0.32947 6.544 4.3150 5.652 36.33 10.180 1.4208 6.502 11.25
## s98 2.99607 5.369 0.34259 6.565 4.3393 5.666 36.34 10.190 1.4390 6.519 11.26
## s99 3.01704 5.392 0.35551 6.586 4.3633 5.680 36.36 10.199 1.4569 6.536 11.28
##
## s0 .
## s1 .
## s2 .
## s3 .
## s4 .
## s5 .
## s6 .
## s7 .
## s8 .
## s9 .
## s10 .
## s11 .
## s12 .

```

```

## s13 .
## s14 .
## s15 .
## s16 .
## s17 .
## s18 .
## s19 .
## s20 .
## s21 .
## s22 .
## s23 .
## s24 .
## s25 .
## s26 . 1.08 .
## s27 . 2.10 .
## s28 . 3.07 .
## s29 . 3.91 .
## s30 . 4.70 .
## s31 . 5.44 .
## s32 . 6.01 0.721 .
## s33 . 6.53 1.830 .
## s34 . 7.02 2.766 .
## s35 . 7.41 3.559 .
## s36 . 7.82 4.320 .
## s37 . 8.16 4.908 .
## s38 . 8.48 5.444 .
## s39 . 8.70 5.856 .
## s40 . -0.0562 8.88 6.170 .
## s41 -0.0735 -0.1731 9.05 6.431 0.122 .
## s42 -0.3137 -0.2669 9.20 6.661 0.348 .
## s43 -0.4575 -0.2840 9.31 6.831 0.537 .
## s44 -0.5803 -0.3060 9.42 6.975 0.713 .
## s45 -0.6738 -0.3132 9.52 7.137 0.875 .
## s46 -0.7243 -0.2656 9.63 7.244 1.130 .
## s47 -0.7736 -0.1805 9.79 7.310 1.423 .
## s48 -0.7565 -0.0193 9.99 7.364 1.809 .
## s49 -0.7620 . 10.07 7.413 1.950 .
## s50 -0.7375 . 10.19 7.511 2.066 0.0186 -0.83942 0.9248 -4.267 -3.213
## s51 -0.6860 . 10.39 7.665 2.211 0.1148 -0.79955 1.0759 -4.173 -3.235
## s52 -0.6221 0.0789 10.65 7.891 2.418 0.2575 -0.66245 1.2688 -4.064 -3.083
## s53 -0.5778 0.1814 10.83 8.055 2.598 0.3496 -0.54743 1.4151 -4.016 -2.927
## s54 -0.5143 0.3122 11.01 8.204 2.811 0.4646 -0.42635 1.5958 -3.954 -2.733
## s55 -0.4424 0.4378 11.20 8.355 3.032 0.5788 -0.27954 1.7818 -3.856 -2.569
## s56 -0.3949 0.5147 11.32 8.470 3.157 0.6511 -0.18692 1.8864 -3.789 -2.510
## s57 -0.3189 0.6173 11.46 8.596 3.321 0.7484 -0.05569 2.0473 -3.717 -2.392
## s58 -0.2161 0.7537 11.62 8.715 3.518 0.8462 0.02603 2.2221 -3.605 -2.295
## s59 0.1328 1.2254 11.91 8.905 4.164 1.0965 0.54209 2.7925 -3.269 -1.992
## s60 0.1519 1.2472 11.93 8.926 4.189 1.1110 0.56396 2.8119 -3.256 -1.987
## s61 0.2719 1.3913 12.07 9.060 4.371 1.2041 0.72695 2.9593 -3.130 -1.898
## s62 0.3940 1.5391 12.21 9.191 4.559 1.2983 0.90244 3.1143 -3.009 -1.807
## s63 0.4994 1.6613 12.31 9.294 4.717 1.3759 1.04913 3.2475 -2.906 -1.736
## s64 0.6072 1.7833 12.42 9.394 4.872 1.4559 1.19806 3.3842 -2.803 -1.664
## s65 0.7187 1.9072 12.52 9.489 5.028 1.5358 1.34797 3.5230 -2.697 -1.597
## s66 0.8310 2.0314 12.62 9.580 5.185 1.6132 1.49692 3.6618 -2.589 -1.535

```

```

## s67 0.9202 2.1301 12.69 9.649 5.309 1.6725 1.61420 3.7714 -2.502 -1.491
## s68 0.9877 2.2045 12.75 9.701 5.402 1.7175 1.70244 3.8541 -2.437 -1.458
## s69 1.0555 2.2793 12.81 9.752 5.495 1.7624 1.79074 3.9370 -2.372 -1.426
## s70 1.1230 2.3536 12.86 9.802 5.588 1.8067 1.87879 4.0195 -2.307 -1.394
## s71 1.1678 2.4030 12.90 9.834 5.650 1.8360 1.93744 4.0742 -2.264 -1.373
## s72 1.1904 2.4278 12.91 9.851 5.681 1.8507 1.96679 4.1018 -2.242 -1.363
## s73 1.2357 2.4773 12.95 9.883 5.743 1.8802 2.02522 4.1571 -2.199 -1.342
## s74 1.2583 2.5020 12.97 9.899 5.774 1.8949 2.05437 4.1846 -2.177 -1.332
## s75 1.3032 2.5512 13.00 9.931 5.835 1.9237 2.11240 4.2393 -2.134 -1.313
## s76 1.3254 2.5756 13.02 9.947 5.866 1.9379 2.14125 4.2664 -2.113 -1.303
## s77 1.3476 2.5999 13.04 9.962 5.896 1.9519 2.17002 4.2934 -2.091 -1.294
## s78 1.3698 2.6241 13.05 9.977 5.926 1.9658 2.19866 4.3202 -2.070 -1.284
## s79 1.3917 2.6482 13.07 9.992 5.957 1.9796 2.22716 4.3469 -2.048 -1.275
## s80 1.4136 2.6721 13.09 10.007 5.987 1.9932 2.25547 4.3733 -2.027 -1.266
## s81 1.4352 2.6959 13.10 10.022 6.016 2.0066 2.28360 4.3995 -2.006 -1.257
## s82 1.4567 2.7194 13.12 10.036 6.046 2.0198 2.31151 4.4254 -1.985 -1.248
## s83 1.4779 2.7427 13.14 10.050 6.075 2.0328 2.33920 4.4510 -1.964 -1.240
## s84 1.4990 2.7658 13.15 10.064 6.103 2.0456 2.36663 4.4764 -1.944 -1.232
## s85 1.5199 2.7886 13.17 10.077 6.132 2.0582 2.39380 4.5015 -1.923 -1.224
## s86 1.5405 2.8112 13.18 10.091 6.160 2.0707 2.42070 4.5263 -1.903 -1.216
## s87 1.5609 2.8335 13.20 10.104 6.188 2.0829 2.44731 4.5508 -1.883 -1.208
## s88 1.5812 2.8555 13.21 10.116 6.215 2.0948 2.47363 4.5750 -1.864 -1.201
## s89 1.6012 2.8774 13.22 10.129 6.243 2.1066 2.49965 4.5989 -1.844 -1.193
## s90 1.6209 2.8989 13.24 10.141 6.269 2.1182 2.52537 4.6226 -1.825 -1.186
## s91 1.6405 2.9202 13.25 10.153 6.296 2.1296 2.55078 4.6459 -1.806 -1.180
## s92 1.6598 2.9413 13.26 10.164 6.322 2.1408 2.57589 4.6690 -1.787 -1.173
## s93 1.6790 2.9621 13.28 10.176 6.348 2.1518 2.60069 4.6917 -1.769 -1.167
## s94 1.6979 2.9826 13.29 10.187 6.374 2.1626 2.62519 4.7142 -1.751 -1.161
## s95 1.7166 3.0029 13.30 10.198 6.399 2.1733 2.64938 4.7364 -1.732 -1.155
## s96 1.7351 3.0230 13.31 10.208 6.424 2.1837 2.67327 4.7584 -1.715 -1.149
## s97 1.7533 3.0428 13.33 10.219 6.449 2.1940 2.69685 4.7801 -1.697 -1.143
## s98 1.7714 3.0624 13.34 10.229 6.473 2.2041 2.72014 4.8015 -1.680 -1.138
## s99 1.7893 3.0818 13.35 10.239 6.497 2.2140 2.74314 4.8226 -1.663 -1.132
##
## s0 .
## s1 .
## s2 .
## s3 .
## s4 .
## s5 .
## s6 .
## s7 .
## s8 .
## s9 .
## s10 .
## s11 .
## s12 .
## s13 .
## s14 .
## s15 .
## s16 .
## s17 .
## s18 .
## s19 .

```

## s20	23.43	.	
## s21	27.86	.	
## s22	31.86	.	
## s23	35.32	.	
## s24	38.39	.	
## s25	.	.	-0.44	40.75	.	
## s26	.	.	-1.69	42.77	.	
## s27	.	.	-2.68	44.51	.	
## s28	.	0.422	-3.30	46.15	.	
## s29	.	2.995	-3.73	47.74	.	
## s30	-0.258	5.331	-4.00	49.26	.	
## s31	-0.552	7.588	-4.08	50.79	.	
## s32	-0.827	9.500	-4.17	-0.207	52.07	.
## s33	-1.021	11.123	-4.23	-0.484	53.21	.
## s34	-1.111	12.462	-4.24	-0.758	54.23	0.276
## s35	-1.185	13.679	-4.23	0.145	-0.970	55.16	0.704	
## s36	-1.229	14.965	-4.01	.	.	0.0114	.	.	0.353	-1.044	56.21	1.055		
## s37	-1.240	15.930	-3.87	.	.	0.1220	.	.	0.550	-1.145	57.06	1.335		
## s38	-1.254	16.859	-3.67	.	.	0.1766	.	.	0.703	-1.211	57.85	1.572		
## s39	-1.330	17.673	-3.56	.	-0.272	0.1581	.	.	0.774	-1.304	58.49	1.767		
## s40	-1.383	18.294	-3.50	.	-0.524	0.1373	0.122	0.823	-1.409	59.00	1.939			
## s41	-1.403	18.821	-3.44	.	-0.761	0.0850	0.369	0.834	-1.567	59.48	2.086			
## s42	-1.422	19.265	-3.45	.	-1.074	.	0.639	0.759	-1.840	59.94	2.242			
## s43	-1.476	19.572	-3.49	0.016	-1.345	.	0.893	0.746	-2.028	60.24	2.374			
## s44	-1.505	19.856	-3.51	0.164	-1.564	.	1.121	0.733	-2.192	60.54	2.487			
## s45	-1.507	20.125	-3.49	0.304	-1.706	.	1.324	0.723	-2.295	60.82	2.604			
## s46	-1.466	20.357	-3.51	0.489	-1.863	.	1.644	0.738	-2.426	61.07	2.772			
## s47	-1.402	20.585	-3.52	0.577	-2.065	.	2.009	0.749	-2.452	61.32	3.013			
## s48	-1.266	20.668	-3.61	0.747	-2.230	.	2.500	0.803	-2.570	61.55	3.291			
## s49	-1.216	20.727	-3.62	0.847	-2.290	.	2.655	0.863	-2.566	61.65	3.403			
## s50	-1.149	20.830	-3.60	0.965	-2.317	0.0225	2.694	0.918	-2.543	61.79	3.544			
## s51	-1.011	21.067	-3.54	1.090	-2.309	0.0657	2.762	0.983	-2.509	62.03	3.754			
## s52	-0.840	21.401	-3.44	1.237	-2.287	0.0919	2.929	1.027	-2.495	62.37	4.023			
## s53	-0.731	21.603	-3.39	1.354	-2.287	0.1149	3.121	1.047	-2.520	62.61	4.210			
## s54	-0.612	21.799	-3.34	1.481	-2.291	0.1480	3.335	1.076	-2.525	62.85	4.427			
## s55	-0.468	22.024	-3.27	1.615	-2.281	0.1921	3.549	1.126	-2.511	63.08	4.653			
## s56	-0.366	22.165	-3.21	1.709	-2.256	0.2367	3.660	1.179	-2.494	63.23	4.784			
## s57	-0.260	22.331	-3.15	1.827	-2.228	0.2814	3.796	1.226	-2.462	63.40	4.958			
## s58	-0.125	22.496	-3.07	1.961	-2.181	0.3719	3.985	1.313	-2.385	63.59	5.152			
## s59	0.203	22.778	-2.86	2.424	-2.011	0.7560	4.684	1.666	-2.103	63.86	5.643			
## s60	0.227	22.806	-2.84	2.451	-2.001	0.7788	4.716	1.689	-2.087	63.88	5.667			
## s61	0.380	22.973	-2.72	2.603	-1.904	0.9092	4.919	1.817	-1.976	64.05	5.835			
## s62	0.525	23.127	-2.60	2.757	-1.813	1.0360	5.117	1.942	-1.865	64.20	6.000			
## s63	0.645	23.253	-2.49	2.886	-1.734	1.1435	5.283	2.047	-1.771	64.31	6.136			
## s64	0.764	23.375	-2.39	3.018	-1.652	1.2524	5.442	2.153	-1.672	64.43	6.273			
## s65	0.885	23.491	-2.30	3.152	-1.568	1.3671	5.603	2.265	-1.572	64.54	6.409			
## s66	1.006	23.600	-2.20	3.286	-1.481	1.4849	5.764	2.379	-1.468	64.64	6.542			
## s67	1.100	23.682	-2.12	3.393	-1.412	1.5803	5.893	2.472	-1.384	64.72	6.646			
## s68	1.171	23.743	-2.06	3.474	-1.360	1.6527	5.989	2.542	-1.320	64.77	6.724			
## s69	1.242	23.803	-2.01	3.555	-1.308	1.7257	6.087	2.612	-1.256	64.83	6.802			
## s70	1.313	23.861	-1.95	3.635	-1.255	1.7989	6.184	2.683	-1.191	64.88	6.878			
## s71	1.359	23.899	-1.91	3.688	-1.220	1.8474	6.249	2.730	-1.149	64.92	6.929			
## s72	1.383	23.918	-1.89	3.715	-1.203	1.8720	6.281	2.753	-1.127	64.93	6.954			
## s73	1.429	23.956	-1.86	3.769	-1.168	1.9209	6.346	2.801	-1.084	64.97	7.005			

```

## s74 1.452 23.975 -1.84 3.795 -1.150 1.9454 6.378 2.824 -1.062 64.99 7.030
## s75 1.498 24.012 -1.80 3.849 -1.115 1.9944 6.443 2.871 -1.019 65.02 7.080
## s76 1.521 24.030 -1.78 3.875 -1.097 2.0188 6.475 2.895 -0.997 65.04 7.105
## s77 1.544 24.048 -1.76 3.901 -1.080 2.0432 6.507 2.918 -0.976 65.05 7.129
## s78 1.567 24.066 -1.75 3.927 -1.063 2.0675 6.539 2.942 -0.955 65.07 7.154
## s79 1.589 24.084 -1.73 3.954 -1.045 2.0918 6.571 2.965 -0.933 65.08 7.178
## s80 1.612 24.101 -1.71 3.979 -1.028 2.1159 6.602 2.988 -0.912 65.10 7.202
## s81 1.634 24.118 -1.69 4.005 -1.011 2.1399 6.634 3.011 -0.891 65.11 7.226
## s82 1.656 24.134 -1.67 4.030 -0.994 2.1638 6.665 3.034 -0.871 65.13 7.249
## s83 1.678 24.151 -1.66 4.056 -0.977 2.1875 6.696 3.057 -0.850 65.14 7.272
## s84 1.699 24.167 -1.64 4.080 -0.961 2.2110 6.727 3.080 -0.829 65.16 7.295
## s85 1.720 24.182 -1.62 4.105 -0.944 2.2344 6.757 3.102 -0.809 65.17 7.317
## s86 1.741 24.198 -1.61 4.129 -0.928 2.2576 6.787 3.125 -0.789 65.18 7.339
## s87 1.762 24.213 -1.59 4.154 -0.912 2.2806 6.817 3.147 -0.769 65.20 7.361
## s88 1.782 24.228 -1.57 4.177 -0.896 2.3034 6.846 3.169 -0.749 65.21 7.382
## s89 1.802 24.242 -1.56 4.201 -0.880 2.3260 6.875 3.190 -0.729 65.22 7.403
## s90 1.822 24.256 -1.54 4.224 -0.864 2.3483 6.904 3.212 -0.710 65.23 7.424
## s91 1.842 24.270 -1.53 4.247 -0.849 2.3705 6.933 3.233 -0.691 65.25 7.445
## s92 1.861 24.284 -1.51 4.270 -0.834 2.3925 6.961 3.254 -0.672 65.26 7.465
## s93 1.880 24.297 -1.49 4.292 -0.819 2.4142 6.989 3.275 -0.653 65.27 7.485
## s94 1.898 24.310 -1.48 4.315 -0.804 2.4358 7.016 3.296 -0.634 65.28 7.504
## s95 1.917 24.323 -1.47 4.337 -0.790 2.4571 7.044 3.316 -0.616 65.29 7.523
## s96 1.935 24.335 -1.45 4.358 -0.775 2.4782 7.070 3.336 -0.598 65.30 7.542
## s97 1.953 24.347 -1.44 4.380 -0.761 2.4991 7.097 3.356 -0.580 65.31 7.561
## s98 1.970 24.359 -1.42 4.401 -0.747 2.5198 7.124 3.376 -0.562 65.32 7.579
## s99 1.988 24.370 -1.41 4.422 -0.733 2.5403 7.150 3.396 -0.545 65.33 7.597
##
## s0 .
## s1 .
## s2 .
## s3 .
## s4 .
## s5 .
## s6 .
## s7 .
## s8 .
## s9 .
## s10 .
## s11 .
## s12 .
## s13 .
## s14 .
## s15 .
## s16 .
## s17 .
## s18 .
## s19 .
## s20 .
## s21 .
## s22 .
## s23 .
## s24 .
## s25 .
## s26 .

```

## s27	20.86	.	.
## s28	23.91	.	.
## s29	.	.	.	-0.276597	.	.	26.67	.	.
## s30	.	.	.	-0.660704	.	.	29.16	.	.
## s31	.	.	.	-0.826840	.	-0.00166	31.47	.	.
## s32	.	.	.	-1.009547	.	-0.39088	33.38	.	.
## s33	.	.	.	-1.107970	.	-0.68511	35.05	.	.
## s34	.	.	.	-1.113013	.	-0.88011	36.53	.	.
## s35	.	.	.	-1.119494	.	-1.04746	37.85	.	.
## s36	.	.	.	-1.001691	.	-1.17177	39.25	.	.
## s37	.	.	.	-0.913514	.	-1.27362	40.39	.	.
## s38	.	.	-0.114	-0.741821	-0.23347	-1.31583	41.43	.	.
## s39	.	-0.00987	-0.339	-0.636551	-0.50745	-1.35937	42.25	.	.
## s40	.	-0.12969	-0.526	-0.574004	-0.73430	-1.38664	42.93	-0.0289	.
## s41	.	-0.20340	-0.687	-0.514162	-0.94569	-1.38827	43.55	-0.2805	.
## s42	0.0763	-0.27385	-0.882	-0.485171	-1.22295	-1.34864	44.11	-0.5512	0.180
## s43	0.4774	-0.29435	-1.062	-0.475152	-1.45797	-1.30380	44.47	-0.8038	0.494
## s44	0.8441	-0.30822	-1.206	-0.454806	-1.64696	-1.26336	44.81	-1.0043	0.739
## s45	1.2036	-0.30224	-1.288	-0.421767	-1.78293	-1.22021	45.15	-1.1400	0.923
## s46	1.5074	-0.21479	-1.398	-0.402280	-1.88253	-1.10299	45.43	-1.2677	1.227
## s47	1.7091	-0.05601	-1.644	-0.325703	-1.87924	-0.97820	45.65	-1.3197	1.530
## s48	1.8491	.	-1.833	-0.281957	-1.87101	-0.79135	45.81	-1.3720	2.046
## s49	1.9242	.	-1.905	-0.269197	-1.86199	-0.75606	45.93	-1.4013	2.202
## s50	2.0545	.	-1.934	-0.249015	-1.83524	-0.71948	46.11	-1.4045	2.354
## s51	2.2733	0.01087	-1.951	-0.136202	-1.75795	-0.62304	46.36	-1.3629	2.491
## s52	2.6202	0.08046	-1.930	.	-1.67948	-0.43042	46.67	-1.2969	2.707
## s53	2.8833	0.16386	-1.910	0.000592	-1.64986	-0.26820	46.88	-1.2706	2.944
## s54	3.1182	0.27861	-1.898	0.116964	-1.60794	-0.10059	47.09	-1.2199	3.221
## s55	3.3469	0.39769	-1.887	0.246909	-1.54110	.	47.29	-1.1561	3.458
## s56	3.4967	0.45575	-1.861	0.335283	-1.49333	0.09884	47.44	-1.1206	3.576
## s57	3.6685	0.54800	-1.830	0.429792	-1.43308	0.23618	47.60	-1.0610	3.754
## s58	3.8292	0.68750	-1.791	0.546779	-1.33088	0.38962	47.76	-0.9633	3.967
## s59	4.0619	1.26258	-1.671	0.783495	-0.99299	0.82514	48.00	-0.6775	4.673
## s60	4.0931	1.28748	-1.659	0.805150	-0.97730	0.85249	48.03	-0.6671	4.701
## s61	4.2632	1.45689	-1.561	0.936825	-0.84712	1.03080	48.20	-0.5518	4.878
## s62	4.4120	1.63484	-1.466	1.057896	-0.72327	1.20147	48.35	-0.4378	5.050
## s63	4.5310	1.78599	-1.388	1.154312	-0.61650	1.33810	48.48	-0.3401	5.188
## s64	4.6460	1.93247	-1.306	1.248000	-0.50920	1.46972	48.60	-0.2408	5.328
## s65	4.7537	2.08040	-1.225	1.339568	-0.39747	1.60000	48.72	-0.1406	5.470
## s66	4.8529	2.22979	-1.143	1.428647	-0.28409	1.72689	48.83	-0.0380	5.611
## s67	4.9274	2.34947	-1.079	1.497554	-0.19331	1.82611	48.91	0.0233	5.723
## s68	4.9819	2.43962	-1.030	1.548950	-0.12485	1.90016	48.98	0.0876	5.806
## s69	5.0355	2.53037	-0.982	1.600178	-0.05563	1.97386	49.04	0.1519	5.890
## s70	5.0873	2.62106	-0.934	1.650566	0.00689	2.04659	49.10	0.2155	5.973
## s71	5.1213	2.68126	-0.902	1.683683	0.05348	2.09458	49.14	0.2578	6.028
## s72	5.1386	2.71160	-0.886	1.700412	0.07721	2.11880	49.16	0.2792	6.055
## s73	5.1717	2.77197	-0.855	1.733282	0.12410	2.16604	49.20	0.3211	6.110
## s74	5.1882	2.80220	-0.840	1.749707	0.14767	2.18973	49.22	0.3423	6.137
## s75	5.2205	2.86254	-0.808	1.782172	0.19452	2.23684	49.25	0.3837	6.191
## s76	5.2362	2.89258	-0.793	1.798206	0.21785	2.26021	49.27	0.4044	6.218
## s77	5.2519	2.92256	-0.777	1.814136	0.24112	2.28349	49.29	0.4250	6.245
## s78	5.2673	2.95247	-0.762	1.829953	0.26432	2.30668	49.31	0.4454	6.272
## s79	5.2825	2.98226	-0.747	1.845623	0.28739	2.32970	49.33	0.4656	6.298
## s80	5.2973	3.01188	-0.731	1.861073	0.31029	2.35250	49.34	0.4856	6.324

```

## s81 5.3119 3.04133 -0.716 1.876335 0.33301 2.37506 49.36 0.5054 6.350
## s82 5.3261 3.07058 -0.701 1.891387 0.35554 2.39736 49.38 0.5250 6.376
## s83 5.3399 3.09960 -0.687 1.906218 0.37786 2.41938 49.40 0.5443 6.401
## s84 5.3535 3.12839 -0.672 1.920819 0.39997 2.44112 49.41 0.5634 6.426
## s85 5.3666 3.15693 -0.658 1.935186 0.42186 2.46255 49.43 0.5823 6.451
## s86 5.3795 3.18521 -0.643 1.949315 0.44352 2.48368 49.44 0.6009 6.475
## s87 5.3920 3.21322 -0.629 1.963206 0.46495 2.50451 49.46 0.6192 6.499
## s88 5.4041 3.24095 -0.616 1.976861 0.48615 2.52503 49.47 0.6373 6.523
## s89 5.4160 3.26841 -0.602 1.990282 0.50711 2.54525 49.49 0.6552 6.546
## s90 5.4275 3.29560 -0.589 2.003471 0.52784 2.56516 49.50 0.6728 6.569
## s91 5.4387 3.32250 -0.575 2.016431 0.54834 2.58476 49.52 0.6901 6.592
## s92 5.4496 3.34912 -0.562 2.029167 0.56861 2.60408 49.53 0.7072 6.615
## s93 5.4602 3.37547 -0.550 2.041683 0.58864 2.62310 49.54 0.7240 6.637
## s94 5.4704 3.40154 -0.537 2.053982 0.60845 2.64183 49.56 0.7406 6.659
## s95 5.4805 3.42733 -0.525 2.066069 0.62802 2.66028 49.57 0.7569 6.680
## s96 5.4902 3.45285 -0.513 2.077950 0.64737 2.67845 49.58 0.7730 6.702
## s97 5.4996 3.47810 -0.501 2.089627 0.66650 2.69635 49.59 0.7889 6.723
## s98 5.5089 3.50308 -0.489 2.101105 0.68540 2.71398 49.61 0.8045 6.743
## s99 5.5178 3.52780 -0.478 2.112390 0.70408 2.73136 49.62 0.8199 6.764
##
## s0 .
## s1 .
## s2 .
## s3 .
## s4 .
## s5 .
## s6 .
## s7 .
## s8 .
## s9 .
## s10 .
## s11 .
## s12 .
## s13 .
## s14 .
## s15 .
## s16 .
## s17 .
## s18 .
## s19 .
## s20 .
## s21 .
## s22 .
## s23 .
## s24 .
## s25 .
## s26 .
## s27 . . 0.583 .
## s28 . . 1.240 .
## s29 . . 1.848 .
## s30 . . 2.388 . . . 0.845 -0.112 0.776
## s31 . . 2.825 . . . 1.575 -0.585 2.520
## s32 . . 3.141 . . . 2.172 -1.014 3.946
## s33 . . 0.235 3.415 . . . 2.669 -1.378 5.171

```

## s34 . .	0.823	3.677	3.115	-1.670	6.212
## s35 . .	1.313	3.877	0.278	.	.	-0.185	3.494	-1.922	7.110	
## s36 . .	1.760	4.054	0.622	.	-0.0963	-0.329	3.807	-2.132	7.938	
## s37 . .	2.154	4.215	0.879	-0.1710	-0.3036	-0.452	4.094	-2.294	8.590	
## s38 0.140 .	2.515	4.336	1.073	-0.3548	-0.4807	-0.557	4.298	-2.457	9.126	
## s39 0.339 .	2.752	4.389	1.166	-0.5392	-0.6165	-0.695	4.394	-2.641	9.483	
## s40 0.519 .	2.945	4.445	1.240	-0.7042	-0.7397	-0.809	4.477	-2.791	9.741	
## s41 0.676 .	3.140	4.492	1.301	-0.8526	-0.8498	-0.908	4.523	-2.925	9.966	
## s42 0.854 .	3.303	4.514	1.302	-0.9979	-0.9295	-1.056	4.495	-3.056	10.113	
## s43 1.058 .	3.404	4.527	1.261	-1.0887	-0.9306	-1.207	4.490	-3.144	10.164	
## s44 1.236 0.0669	3.506	4.544	1.237	-1.1658	-0.9343	-1.320	4.486	-3.218	10.222	
## s45 1.379 0.1982	3.603	4.577	1.262	-1.2103	-0.9372	-1.391	4.505	-3.261	10.315	
## s46 1.656 0.2971	3.719	4.651	1.283	-1.2079	-0.8561	-1.427	4.530	-3.257	10.345	
## s47 1.907 0.2810	3.920	4.716	1.323	-1.2203	-0.7922	-1.412	4.565	-3.390	10.165	
## s48 2.256 0.2600	4.139	4.870	1.369	-1.1777	-0.6182	-1.365	4.645	-3.403	10.013	
## s49 2.390 0.2620	4.228	4.937	1.410	-1.1537	-0.5803	-1.363	4.709	-3.403	9.971	
## s50 2.542 0.3030	4.339	5.054	1.492	-1.0953	-0.5110	-1.327	4.789	-3.390	9.970	
## s51 2.751 0.4171	4.507	5.205	1.619	-1.0263	-0.4121	-1.215	4.884	-3.366	9.993	
## s52 3.012 0.6265	4.703	5.393	1.790	-0.9314	-0.2681	-1.066	4.980	-3.310	10.079	
## s53 3.199 0.7793	4.836	5.534	1.906	-0.8637	-0.1509	-0.980	5.045	-3.264	10.147	
## s54 3.423 0.9201	4.997	5.678	2.017	-0.7795	.	-0.876	5.122	-3.208	10.209	
## s55 3.672 1.0668	5.157	5.825	2.140	-0.6814	0.1023	-0.760	5.203	-3.136	10.266	
## s56 3.808 1.1718	5.248	5.920	2.225	-0.6058	0.1940	-0.698	5.271	-3.066	10.322	
## s57 3.998 1.2961	5.377	6.040	2.332	-0.5099	0.3350	-0.602	5.347	-3.000	10.397	
## s58 4.210 1.4125	5.530	6.183	2.456	-0.3945	0.4961	-0.476	5.451	-2.916	10.444	
## s59 4.927 1.6295	5.933	6.558	2.746	.	1.0387	-0.137	5.779	-2.576	10.605	
## s60 4.956 1.6521	5.952	6.582	2.767	0.0149	1.0608	-0.121	5.802	-2.555	10.621	
## s61 5.151 1.7931	6.087	6.732	2.909	0.1588	1.2054	.	5.928	-2.419	10.709	
## s62 5.347 1.9268	6.218	6.873	3.047	0.3086	1.3540	0.114	6.050	-2.283	10.806	
## s63 5.510 2.0375	6.330	6.991	3.165	0.4350	1.4798	0.222	6.152	-2.171	10.891	
## s64 5.677 2.1488	6.440	7.110	3.281	0.5616	1.6089	0.332	6.256	-2.055	10.976	
## s65 5.846 2.2561	6.551	7.230	3.395	0.6907	1.7406	0.442	6.363	-1.935	11.060	
## s66 6.016 2.3586	6.660	7.350	3.509	0.8201	1.8722	0.553	6.472	-1.814	11.143	
## s67 6.151 2.4366	6.746	7.444	3.599	0.9230	1.9765	0.640	6.558	-1.717	11.207	
## s68 6.252 2.4950	6.811	7.515	3.666	1.0006	2.0548	0.706	6.624	-1.645	11.256	
## s69 6.353 2.5531	6.875	7.585	3.733	1.0784	2.1332	0.771	6.690	-1.571	11.305	
## s70 6.454 2.6103	6.939	7.655	3.800	1.1558	2.2110	0.836	6.756	-1.498	11.353	
## s71 6.520 2.6478	6.981	7.701	3.843	1.2072	2.2627	0.879	6.799	-1.450	11.385	
## s72 6.554 2.6667	7.002	7.725	3.865	1.2332	2.2888	0.900	6.821	-1.425	11.402	
## s73 6.621 2.7039	7.044	7.770	3.909	1.2847	2.3406	0.943	6.865	-1.377	11.434	
## s74 6.654 2.7224	7.065	7.793	3.931	1.3105	2.3665	0.964	6.887	-1.353	11.450	
## s75 6.721 2.7591	7.107	7.838	3.974	1.3618	2.4178	1.007	6.930	-1.305	11.481	
## s76 6.754 2.7771	7.127	7.861	3.995	1.3873	2.4432	1.028	6.952	-1.281	11.497	
## s77 6.787 2.7951	7.148	7.883	4.016	1.4126	2.4686	1.048	6.974	-1.257	11.512	
## s78 6.820 2.8129	7.168	7.906	4.038	1.4379	2.4938	1.069	6.995	-1.233	11.527	
## s79 6.852 2.8305	7.188	7.928	4.059	1.4631	2.5188	1.090	7.016	-1.209	11.542	
## s80 6.884 2.8479	7.208	7.950	4.079	1.4880	2.5436	1.110	7.038	-1.185	11.557	
## s81 6.917 2.8650	7.228	7.972	4.100	1.5128	2.5682	1.131	7.059	-1.161	11.572	
## s82 6.948 2.8819	7.247	7.993	4.120	1.5373	2.5925	1.151	7.080	-1.138	11.586	
## s83 6.980 2.8985	7.267	8.015	4.141	1.5617	2.6167	1.171	7.100	-1.115	11.601	
## s84 7.011 2.9148	7.286	8.036	4.160	1.5858	2.6405	1.190	7.121	-1.092	11.615	
## s85 7.042 2.9308	7.304	8.056	4.180	1.6096	2.6641	1.209	7.141	-1.069	11.629	
## s86 7.072 2.9466	7.323	8.077	4.199	1.6332	2.6874	1.228	7.161	-1.046	11.643	
## s87 7.103 2.9620	7.341	8.097	4.218	1.6566	2.7104	1.247	7.181	-1.024	11.656	

```

## s88 7.132 2.9772 7.359 8.117 4.237 1.6797 2.7331 1.266 7.200 -1.002 11.670
## s89 7.162 2.9920 7.376 8.136 4.255 1.7025 2.7556 1.284 7.219 -0.980 11.683
## s90 7.191 3.0066 7.394 8.156 4.273 1.7251 2.7778 1.302 7.239 -0.958 11.696
## s91 7.220 3.0208 7.411 8.174 4.291 1.7474 2.7997 1.319 7.257 -0.937 11.708
## s92 7.248 3.0348 7.428 8.193 4.309 1.7695 2.8214 1.337 7.276 -0.916 11.721
## s93 7.276 3.0485 7.444 8.212 4.326 1.7913 2.8427 1.354 7.295 -0.895 11.733
## s94 7.304 3.0620 7.461 8.230 4.343 1.8129 2.8638 1.371 7.313 -0.874 11.745
## s95 7.332 3.0752 7.477 8.248 4.360 1.8341 2.8846 1.387 7.331 -0.854 11.757
## s96 7.359 3.0881 7.493 8.265 4.376 1.8552 2.9052 1.404 7.348 -0.834 11.769
## s97 7.386 3.1007 7.509 8.282 4.393 1.8760 2.9255 1.420 7.366 -0.814 11.780
## s98 7.412 3.1131 7.524 8.299 4.409 1.8965 2.9456 1.436 7.383 -0.794 11.792
## s99 7.438 3.1253 7.539 8.316 4.424 1.9168 2.9654 1.452 7.401 -0.775 11.803
##
## s0 .
## s1 .
## s2 .
## s3 .
## s4 .
## s5 .
## s6 .
## s7 .
## s8 .
## s9 .
## s10 .
## s11 .
## s12 .
## s13 .
## s14 .
## s15 .
## s16 .
## s17 .
## s18 .
## s19 .
## s20 .
## s21 .
## s22 .
## s23 .
## s24 1.13 .
## s25 6.47 .
## s26 10.88 .
## s27 14.54 .
## s28 17.47 .
## s29 19.95 .
## s30 22.06 .
## s31 23.92 .
## s32 25.45 1.58 .
## s33 26.75 3.27 .
## s34 27.82 4.64 .
## s35 28.74 5.90 .
## s36 29.60 7.24 .
## s37 30.24 8.25 .
## s38 30.80 9.19 .
## s39 31.26 10.00 .
## s40 31.60 10.64 .

```

```

## s41 31.88 11.14 -0.0584 3.013 -0.8025 -0.3744 0.4705 29.471 .      15.07 42.131
## s42 32.08 11.55 -0.2631 2.950 -0.9704 -0.5155 0.5938 29.469 .      15.53 42.606
## s43 32.18 11.86 -0.4046 2.903 -1.0990 -0.5872 0.7469 29.419 .      15.84 42.947
## s44 32.28 12.14 -0.5227 2.863 -1.1984 -0.6432 0.8796 29.391 .      16.13 43.261
## s45 32.40 12.42 -0.5973 2.845 -1.2698 -0.6821 0.9873 29.387 .      16.44 43.566
## s46 32.49 12.63 -0.6440 2.811 -1.2920 -0.6482 1.2040 29.347 .      16.67 43.806
## s47 32.62 12.88 -0.6314 2.777 -1.3016 -0.5899 1.4230 29.322 .      16.82 44.014
## s48 32.68 12.95 -0.5639 2.788 -1.2370 -0.4944 1.7201 29.255 .      16.90 44.127
## s49 32.71 13.04 -0.5583 2.811 -1.2113 -0.4629 1.8391 29.247 0.0754 16.97 44.211
## s50 32.79 13.16 -0.5019 2.834 -1.1941 -0.4206 1.9915 29.231 0.1381 17.07 44.340
## s51 32.99 13.37 -0.4138 2.861 -1.1799 -0.3513 2.1760 29.315 0.2309 17.24 44.530
## s52 33.27 13.66 -0.2850 2.889 -1.1583 -0.2566 2.4027 29.421 0.3521 17.51 44.808
## s53 33.43 13.83 -0.1831 2.909 -1.1354 -0.1903 2.5688 29.455 0.4500 17.71 45.008
## s54 33.59 14.02 -0.0662 2.955 -1.0904 -0.0967 2.7688 29.495 0.5482 17.89 45.193
## s55 33.77 14.22 .      3.001 -1.0331 .      2.9841 29.573 0.6533 18.06 45.372
## s56 33.88 14.34 0.0371 3.046 -0.9833 0.0158 3.1056 29.635 0.7232 18.19 45.498
## s57 34.03 14.50 0.1409 3.096 -0.9263 0.1209 3.2794 29.697 0.8048 18.33 45.635
## s58 34.17 14.67 0.2726 3.174 -0.8319 0.2560 3.4745 29.777 0.9232 18.47 45.779
## s59 34.43 14.94 0.6485 3.477 -0.4120 0.7608 4.1381 30.019 1.3249 18.66 45.989
## s60 34.46 14.97 0.6692 3.496 -0.3954 0.7838 4.1646 30.041 1.3572 18.69 46.018
## s61 34.59 15.12 0.7951 3.594 -0.2666 0.9307 4.3449 30.142 1.5153 18.84 46.179
## s62 34.73 15.27 0.9113 3.694 -0.1324 1.0801 4.5291 30.241 1.6655 18.98 46.322
## s63 34.84 15.39 1.0110 3.783 -0.0161 1.2086 4.6839 30.331 1.7916 19.09 46.436
## s64 34.95 15.51 1.1131 3.873 0.0880 1.3381 4.8405 30.420 1.9151 19.20 46.545
## s65 35.05 15.62 1.2163 3.967 0.2104 1.4709 4.9996 30.512 2.0402 19.30 46.648
## s66 35.15 15.73 1.3211 4.064 0.3356 1.6050 5.1592 30.605 2.1652 19.40 46.744
## s67 35.23 15.81 1.4042 4.141 0.4366 1.7121 5.2865 30.678 2.2646 19.47 46.818
## s68 35.29 15.87 1.4667 4.200 0.5131 1.7929 5.3820 30.734 2.3394 19.53 46.872
## s69 35.34 15.93 1.5293 4.259 0.5901 1.8741 5.4776 30.790 2.4143 19.58 46.924
## s70 35.40 15.99 1.5913 4.319 0.6671 1.9550 5.5725 30.846 2.4886 19.63 46.976
## s71 35.44 16.03 1.6323 4.358 0.7184 2.0087 5.6357 30.883 2.5380 19.67 47.010
## s72 35.45 16.05 1.6529 4.378 0.7443 2.0359 5.6675 30.902 2.5632 19.68 47.027
## s73 35.49 16.09 1.6938 4.418 0.7959 2.0899 5.7308 30.940 2.6120 19.72 47.060
## s74 35.51 16.10 1.7142 4.438 0.8218 2.1170 5.7624 30.959 2.6367 19.74 47.076
## s75 35.54 16.14 1.7547 4.478 0.8733 2.1707 5.8252 30.995 2.6856 19.77 47.109
## s76 35.56 16.16 1.7748 4.497 0.8989 2.1975 5.8563 31.014 2.7100 19.79 47.124
## s77 35.58 16.18 1.7948 4.517 0.9245 2.2241 5.8874 31.032 2.7343 19.80 47.140
## s78 35.59 16.20 1.8147 4.536 0.9501 2.2507 5.9184 31.050 2.7586 19.82 47.156
## s79 35.61 16.21 1.8345 4.556 0.9755 2.2772 5.9491 31.068 2.7827 19.83 47.171
## s80 35.63 16.23 1.8540 4.575 1.0008 2.3034 5.9796 31.086 2.8066 19.85 47.186
## s81 35.64 16.25 1.8733 4.594 1.0260 2.3295 6.0099 31.104 2.8303 19.86 47.201
## s82 35.66 16.26 1.8924 4.613 1.0510 2.3553 6.0399 31.122 2.8538 19.88 47.215
## s83 35.68 16.28 1.9112 4.632 1.0758 2.3810 6.0696 31.139 2.8771 19.89 47.229
## s84 35.69 16.30 1.9298 4.650 1.1004 2.4064 6.0990 31.157 2.9002 19.91 47.243
## s85 35.71 16.31 1.9482 4.669 1.1248 2.4315 6.1281 31.174 2.9230 19.92 47.257
## s86 35.72 16.33 1.9663 4.687 1.1491 2.4564 6.1568 31.191 2.9456 19.94 47.270
## s87 35.74 16.34 1.9841 4.705 1.1731 2.4811 6.1853 31.207 2.9679 19.95 47.283
## s88 35.75 16.36 2.0017 4.723 1.1969 2.5055 6.2134 31.224 2.9900 19.96 47.295
## s89 35.76 16.37 2.0190 4.741 1.2204 2.5297 6.2412 31.240 3.0118 19.97 47.308
## s90 35.78 16.38 2.0361 4.758 1.2438 2.5536 6.2687 31.257 3.0334 19.99 47.320
## s91 35.79 16.40 2.0529 4.776 1.2669 2.5772 6.2959 31.273 3.0547 20.00 47.331
## s92 35.80 16.41 2.0695 4.793 1.2898 2.6006 6.3228 31.288 3.0758 20.01 47.343
## s93 35.82 16.43 2.0858 4.810 1.3125 2.6238 6.3493 31.304 3.0967 20.02 47.354
## s94 35.83 16.44 2.1019 4.827 1.3349 2.6466 6.3755 31.319 3.1173 20.03 47.365

```

```

## s95 35.84 16.45  2.1177 4.843  1.3571  2.6693 6.4014 31.335 3.1377 20.04 47.375
## s96 35.85 16.46  2.1333 4.860  1.3791  2.6916 6.4270 31.350 3.1578 20.06 47.386
## s97 35.87 16.47  2.1487 4.876  1.4009  2.7138 6.4523 31.364 3.1777 20.07 47.396
## s98 35.88 16.49  2.1638 4.892  1.4224  2.7357 6.4772 31.379 3.1973 20.08 47.406
## s99 35.89 16.50  2.1787 4.908  1.4437  2.7573 6.5019 31.393 3.2168 20.09 47.415
##
## s0   .
## s1   .
## s2   .
## s3   .
## s4   .
## s5   .
## s6   .
## s7   .
## s8   .
## s9   .
## s10  .
## s11  .
## s12  .
## s13  .
## s14  .
## s15  .
## s16  .
## s17  .
## s18  .
## s19  .
## s20  .
## s21  .
## s22  .
## s23  .
## s24  .
## s25  .          1.32  .
## s26 -0.46695  .          0.961 3.94  .
## s27 -1.34772  .          4.313 6.22  .
## s28 -2.02810  .          7.451 8.30  .
## s29 -2.54976  .          10.273 10.25 -0.532
## s30 -2.86285  .          12.914 12.13 -1.126
## s31 -3.05743  .          15.437 13.98 -1.634
## s32 -3.25311  .          17.532 15.55 -2.117
## s33 -3.38184  .          19.370 16.95 -2.551
## s34 -3.42530  .          -0.0184 21.022 18.23 -2.946
## s35 -3.48169 0.344  .      -0.3081 22.463 19.38 -3.300 -0.00732  .
## s36 -3.50020 1.190 -0.14579 -0.5345 23.953 20.62 -3.551 -0.30127 0.234  .
## s37 -3.49756 1.847 -0.42708 -0.7381 25.174 21.62 -3.778 -0.53801 0.555 0.572
## s38 -3.46460 2.415 -0.63707 -0.9023 26.337 22.55 -3.969 -0.75283 0.812 1.107
## s39 -3.47312 2.811 -0.82749 -1.0700 27.285 23.30 -4.170 -0.94357 1.066 1.507
## s40 -3.46213 3.116 -1.03146 -1.2274 28.054 23.91 -4.359 -1.10286 1.287 1.838
## s41 -3.43500 3.369 -1.22383 -1.3563 28.775 24.50 -4.553 -1.26242 1.490 2.110
## s42 -3.39161 3.515 -1.42809 -1.4637 29.473 25.09 -4.777 -1.42858 1.746 2.276
## s43 -3.34440 3.605 -1.60636 -1.5628 29.968 25.47 -4.930 -1.49967 2.001 2.374
## s44 -3.30091 3.699 -1.75192 -1.6374 30.422 25.84 -5.066 -1.56368 2.219 2.473
## s45 -3.24980 3.838 -1.85101 -1.6777 30.831 26.19 -5.158 -1.61566 2.418 2.570
## s46 -3.14593 3.911 -1.93924 -1.6856 31.199 26.51 -5.238 -1.59970 2.705 2.685
## s47 -3.02747 4.002 -1.97104 -1.7331 31.550 26.81 -5.340 -1.60191 2.850 2.770

```

```

## s48 -2.82806 4.002 -2.02622 -1.7040 31.869 27.11 -5.402 -1.49112 3.139 2.864
## s49 -2.77605 4.044 -2.06559 -1.7431 32.016 27.22 -5.403 -1.47039 3.259 2.932
## s50 -2.70461 4.084 -2.08485 -1.7777 32.195 27.36 -5.403 -1.43291 3.424 2.998
## s51 -2.58844 4.174 -2.02185 -1.7084 32.488 27.62 -5.397 -1.35224 3.551 3.138
## s52 -2.39317 4.313 -1.91037 -1.5641 32.876 28.00 -5.380 -1.24675 3.735 3.289
## s53 -2.22662 4.403 -1.86047 -1.4707 33.148 28.27 -5.376 -1.17104 3.904 3.362
## s54 -2.05689 4.497 -1.79051 -1.3501 33.423 28.56 -5.360 -1.05912 4.142 3.471
## s55 -1.89685 4.597 -1.69463 -1.2164 33.699 28.84 -5.324 -0.94424 4.360 3.600
## s56 -1.80368 4.677 -1.63106 -1.1353 33.877 29.01 -5.278 -0.87842 4.473 3.677
## s57 -1.68236 4.757 -1.54783 -1.0308 34.067 29.21 -5.240 -0.78056 4.669 3.785
## s58 -1.53751 4.870 -1.44844 -0.9196 34.269 29.41 -5.167 -0.64995 4.865 3.911
## s59 -1.15815 5.070 -1.14755 -0.6033 34.557 29.74 -4.869 -0.20601 5.609 4.312
## s60 -1.13550 5.096 -1.13571 -0.5903 34.591 29.76 -4.849 -0.18675 5.634 4.338
## s61 -0.97267 5.231 -1.01727 -0.4745 34.764 29.93 -4.721 -0.05833 5.789 4.463
## s62 -0.82198 5.354 -0.89620 -0.3601 34.914 30.07 -4.594 0.04561 5.964 4.588
## s63 -0.70084 5.459 -0.79430 -0.2653 35.034 30.19 -4.489 0.15478 6.115 4.699
## s64 -0.58261 5.562 -0.69312 -0.1676 35.149 30.31 -4.381 0.26978 6.274 4.814
## s65 -0.46516 5.663 -0.59126 -0.0692 35.259 30.42 -4.271 0.38860 6.436 4.932
## s66 -0.34989 5.761 -0.48970 0.0127 35.360 30.52 -4.157 0.50752 6.600 5.050
## s67 -0.25982 5.837 -0.41014 0.0890 35.436 30.60 -4.065 0.60201 6.731 5.144
## s68 -0.19285 5.893 -0.35024 0.1467 35.492 30.66 -3.996 0.67331 6.830 5.215
## s69 -0.12609 5.950 -0.28998 0.2044 35.546 30.71 -3.927 0.74470 6.929 5.286
## s70 -0.06024 6.005 -0.22999 0.2613 35.599 30.77 -3.857 0.81560 7.028 5.357
## s71 -0.01697 6.042 -0.19045 0.2990 35.633 30.80 -3.811 0.86269 7.094 5.404
## s72 0.00261 6.061 -0.17080 0.3181 35.651 30.82 -3.788 0.88646 7.127 5.428
## s73 0.04554 6.097 -0.13080 0.3556 35.684 30.85 -3.742 0.93354 7.193 5.475
## s74 0.06708 6.116 -0.11101 0.3744 35.701 30.87 -3.719 0.95715 7.226 5.499
## s75 0.10990 6.152 -0.07136 0.4114 35.734 30.91 -3.673 1.00380 7.291 5.545
## s76 0.13112 6.170 -0.05180 0.4298 35.749 30.92 -3.650 1.02704 7.323 5.569
## s77 0.15227 6.187 -0.03233 0.4480 35.765 30.94 -3.627 1.05018 7.356 5.592
## s78 0.17331 6.205 -0.01292 0.4661 35.781 30.95 -3.604 1.07319 7.388 5.615
## s79 0.19417 6.223 0.00514 0.4840 35.796 30.97 -3.581 1.09603 7.419 5.638
## s80 0.21483 6.240 0.02448 0.5017 35.811 30.98 -3.558 1.11869 7.451 5.660
## s81 0.23523 6.257 0.04365 0.5192 35.825 31.00 -3.536 1.14115 7.482 5.683
## s82 0.25537 6.274 0.06264 0.5365 35.840 31.01 -3.513 1.16339 7.513 5.705
## s83 0.27523 6.290 0.08142 0.5536 35.853 31.03 -3.491 1.18541 7.543 5.727
## s84 0.29479 6.307 0.09998 0.5704 35.867 31.04 -3.469 1.20720 7.573 5.749
## s85 0.31406 6.323 0.11831 0.5870 35.880 31.06 -3.447 1.22874 7.603 5.770
## s86 0.33302 6.338 0.13642 0.6033 35.893 31.07 -3.425 1.25004 7.633 5.791
## s87 0.35169 6.354 0.15429 0.6194 35.905 31.08 -3.404 1.27109 7.662 5.812
## s88 0.37005 6.369 0.17193 0.6352 35.917 31.10 -3.383 1.29189 7.691 5.833
## s89 0.38811 6.384 0.18933 0.6508 35.929 31.11 -3.362 1.31243 7.720 5.854
## s90 0.40588 6.399 0.20649 0.6661 35.940 31.12 -3.341 1.33272 7.748 5.874
## s91 0.42335 6.414 0.22342 0.6812 35.951 31.13 -3.320 1.35275 7.776 5.894
## s92 0.44053 6.428 0.24012 0.6961 35.962 31.14 -3.300 1.37253 7.803 5.914
## s93 0.45743 6.442 0.25658 0.7107 35.973 31.15 -3.280 1.39207 7.830 5.933
## s94 0.47405 6.456 0.27282 0.7250 35.983 31.16 -3.260 1.41135 7.857 5.953
## s95 0.49039 6.469 0.28884 0.7392 35.992 31.17 -3.240 1.43039 7.884 5.972
## s96 0.50647 6.483 0.30463 0.7531 36.002 31.18 -3.221 1.44918 7.910 5.990
## s97 0.52229 6.496 0.32020 0.7668 36.011 31.19 -3.202 1.46773 7.935 6.009
## s98 0.53784 6.508 0.33555 0.7803 36.020 31.20 -3.183 1.48605 7.961 6.027
## s99 0.55315 6.521 0.35070 0.7936 36.029 31.21 -3.164 1.50412 7.986 6.045
##
## s0 .

```

```

## s1   .
## s2   .
## s3   .
## s4   .
## s5   .
## s6   .
## s7   .
## s8   .
## s9   .
## s10  .
## s11  .
## s12  .
## s13  .
## s14  .
## s15  .
## s16  .           2.44
## s17  .           4.66
## s18  .           6.06
## s19  .           7.12
## s20  .           7.99
## s21  .     6.85  .
## s22  .     15.29 .
## s23  .     22.35 .
## s24  0.523 28.28 .     -0.00383 .
## s25  4.319 33.41 .     -0.08356 .
## s26  7.461 37.72 .     -0.11132 .
## s27  9.962 41.32 -0.0151 -0.13564 .
## s28  11.876 44.32 -0.0345 -0.15574 .
## s29  13.491 46.87 -0.0519 -0.17702 .
## s30  14.976 49.09 -0.0667 -0.19953 .
## s31  16.252 51.03 -0.0820 -0.23488 .
## s32  17.323 52.62 -0.0956 -0.26363 .
## s33  18.225 53.99 -0.1076 -0.28242 .
## s34  18.979 55.18 -0.1177 -0.28780 .
## s35  19.629 56.18 -0.1262 -0.28975 .
## s36  20.167 57.11 -0.1361 -0.29397 .
## s37  20.583 57.84 -0.1443 -0.28818 .
## s38  20.928 58.50 -0.1516 -0.29867 .
## s39  21.178 59.01 -0.1568 -0.32230 .
## s40  21.325 59.38 -0.1605 -0.33695 .
## s41  21.466 59.74 -0.1632 -0.34551 -0.00686 -0.15535 .
## s42  21.623 60.06 -0.1642 -0.35682 -0.03252 -0.16021 .
## s43  21.713 60.23 -0.1642 -0.36971 -0.05367 -0.16388 .
## s44  21.797 60.40 -0.1642 -0.37845 -0.07034 -0.16713 .
## s45  21.867 60.57 -0.1645 -0.38372 -0.08624 -0.16924 .
## s46  21.949 60.70 -0.1623 -0.39190 -0.10174 -0.16978 .
## s47  22.087 60.86 -0.1595 -0.40294 -0.11675 -0.16980 .
## s48  22.274 60.99 -0.1549 -0.39708 -0.14036 -0.16892 .
## s49  22.321 61.05 -0.1539 -0.39862 -0.14921 -0.16939 0.00252 0.000468 4.05
## s50  22.371 61.12 -0.1525 -0.40254 -0.15769 -0.16998 0.01676 0.000644 3.99
## s51  22.528 61.32 -0.1514 -0.41068 -0.16012 -0.17043 0.02184 0.006404 3.93
## s52  22.732 61.59 -0.1499 -0.42101 -0.16872 -0.16943 0.02919 0.013613 3.85
## s53  22.854 61.76 -0.1484 -0.42257 -0.18032 -0.16833 0.04120 0.017629 3.78
## s54  23.026 61.93 -0.1461 -0.42471 -0.19273 -0.16711 0.04651 0.016953 3.72

```

```

## s55 23.212 62.13 -0.1441 -0.43140 -0.20306 -0.16583 0.04632 0.018515 3.66
## s56 23.327 62.25 -0.1428 -0.43617 -0.21133 -0.16543 0.04622 0.022726 3.60
## s57 23.467 62.40 -0.1410 -0.44080 -0.21837 -0.16438 0.05041 0.022297 3.58
## s58 23.621 62.54 -0.1391 -0.44379 -0.22460 -0.16328 0.05080 0.021305 3.56
## s59 24.011 62.78 -0.1318 -0.45206 -0.23465 -0.15919 0.05297 0.006663 3.51
## s60 24.029 62.80 -0.1317 -0.45291 -0.23508 -0.15911 0.05295 0.007364 3.50
## s61 24.142 62.93 -0.1304 -0.45908 -0.23893 -0.15818 0.05405 0.010239 3.47
## s62 24.261 63.06 -0.1290 -0.46589 -0.24380 -0.15719 0.05527 0.011642 3.45
## s63 24.364 63.17 -0.1279 -0.47054 -0.24717 -0.15623 0.05626 0.012046 3.45
## s64 24.466 63.27 -0.1267 -0.47483 -0.24999 -0.15526 0.05626 0.011873 3.44
## s65 24.570 63.37 -0.1255 -0.47868 -0.25245 -0.15430 0.05616 0.011480 3.44
## s66 24.671 63.46 -0.1242 -0.48229 -0.25463 -0.15334 0.05613 0.010756 3.43
## s67 24.749 63.53 -0.1233 -0.48508 -0.25626 -0.15258 0.05608 0.009997 3.42
## s68 24.808 63.58 -0.1225 -0.48708 -0.25741 -0.15201 0.05597 0.009411 3.42
## s69 24.867 63.64 -0.1218 -0.48903 -0.25849 -0.15143 0.05580 0.008787 3.41
## s70 24.925 63.69 -0.1211 -0.49093 -0.25954 -0.15086 0.05556 0.008110 3.40
## s71 24.963 63.72 -0.1206 -0.49219 -0.26022 -0.15048 0.05543 0.007642 3.40
## s72 24.982 63.74 -0.1204 -0.49284 -0.26055 -0.15029 0.05538 0.007390 3.40
## s73 25.021 63.77 -0.1199 -0.49407 -0.26118 -0.14992 0.05520 0.006862 3.39
## s74 25.040 63.79 -0.1197 -0.49468 -0.26149 -0.14973 0.05512 0.006615 3.39
## s75 25.077 63.82 -0.1192 -0.49587 -0.26208 -0.14935 0.05492 0.006128 3.39
## s76 25.096 63.83 -0.1189 -0.49647 -0.26237 -0.14916 0.05482 0.005884 3.39
## s77 25.115 63.85 -0.1187 -0.49706 -0.26264 -0.14897 0.05472 0.005647 3.38
## s78 25.133 63.86 -0.1185 -0.49765 -0.26292 -0.14879 0.05461 0.005413 3.38
## s79 25.151 63.88 -0.1182 -0.49823 -0.26318 -0.14860 0.05451 0.005178 3.38
## s80 25.169 63.89 -0.1180 -0.49881 -0.26344 -0.14841 0.05439 0.004943 3.38
## s81 25.186 63.91 -0.1178 -0.49939 -0.26368 -0.14822 0.05427 0.004708 3.38
## s82 25.204 63.92 -0.1175 -0.49996 -0.26392 -0.14804 0.05414 0.004472 3.37
## s83 25.221 63.94 -0.1173 -0.50053 -0.26414 -0.14786 0.05400 0.004234 3.37
## s84 25.238 63.95 -0.1171 -0.50109 -0.26436 -0.14768 0.05386 0.003996 3.37
## s85 25.255 63.96 -0.1168 -0.50165 -0.26457 -0.14750 0.05372 0.003756 3.37
## s86 25.271 63.98 -0.1166 -0.50221 -0.26477 -0.14732 0.05357 0.003516 3.37
## s87 25.287 63.99 -0.1164 -0.50276 -0.26495 -0.14714 0.05343 0.003275 3.36
## s88 25.303 64.00 -0.1162 -0.50330 -0.26513 -0.14697 0.05328 0.003034 3.36
## s89 25.319 64.02 -0.1160 -0.50384 -0.26530 -0.14680 0.05312 0.002793 3.36
## s90 25.335 64.03 -0.1157 -0.50436 -0.26546 -0.14663 0.05297 0.002552 3.36
## s91 25.350 64.04 -0.1155 -0.50489 -0.26562 -0.14647 0.05282 0.002311 3.36
## s92 25.365 64.05 -0.1153 -0.50540 -0.26576 -0.14630 0.05267 0.002071 3.36
## s93 25.379 64.06 -0.1151 -0.50591 -0.26590 -0.14614 0.05251 0.001831 3.35
## s94 25.394 64.07 -0.1149 -0.50641 -0.26603 -0.14598 0.05236 0.001593 3.35
## s95 25.408 64.09 -0.1147 -0.50691 -0.26615 -0.14582 0.05221 0.001355 3.35
## s96 25.422 64.10 -0.1145 -0.50739 -0.26627 -0.14567 0.05206 0.001119 3.35
## s97 25.436 64.11 -0.1143 -0.50787 -0.26638 -0.14551 0.05191 0.000884 3.35
## s98 25.450 64.12 -0.1141 -0.50834 -0.26648 -0.14536 0.05176 0.000650 3.35
## s99 25.463 64.13 -0.1139 -0.50881 -0.26658 -0.14521 0.05161 0.000418 3.34
##
## s0 .
## s1 .
## s2 .
## s3 .
## s4 .
## s5 .
## s6 .
## s7 .

```

```

## s8 .
## s9 .
## s10 .
## s11 .
## s12 .
## s13 .
## s14 .
## s15 .
## s16 .
## s17 .
## s18 .
## s19 .
## s20 .
## s21 .
## s22 .
## s23 .
## s24 .
## s25 .
## s26 .
## s27 .
## s28 .
## s29 .
## s30 .
## s31 .
## s32 .
## s33 .
## s34 .
## s35 .
## s36 .
## s37 .
## s38 .
## s39 .
## s40 .
## s41 .
## s42 0.011512 -1.2283 0.057689 .
## s43 0.060319 -1.2149 0.056647 .
## s44 0.089533 -1.2089 0.055644 .
## s45 0.099736 -1.2079 0.055385 .
## s46 0.115018 -1.2040 0.054053 0.00252 0.2041 0.0860 .
## s47 0.107822 -1.2171 0.054115 0.05978 0.2129 0.0421 .
## s48 0.087933 -1.2273 0.051603 0.10516 0.2378 .
## s49 0.073033 -1.2360 0.050418 0.12543 0.2406 .
## s50 0.061427 -1.2370 0.049237 0.14904 0.2402 .
## s51 0.057390 -1.2401 0.048646 0.17949 0.2388 -0.0237 0.001310 0.02448
## s52 0.048349 -1.2453 0.048652 0.20356 0.2389 -0.0688 0.002034 0.02590
## s53 0.035750 -1.2484 0.048451 0.21582 0.2468 -0.1037 0.002344 0.02755
## s54 0.037217 -1.2489 0.048537 0.23285 0.2538 -0.1416 0.002695 0.02899
## s55 0.030528 -1.2522 0.048402 0.24919 0.2541 -0.1759 0.003105 0.02946
## s56 0.023277 -1.2546 0.048004 0.25626 0.2517 -0.1893 0.003338 0.02892
## s57 0.030343 -1.2490 0.048241 0.26716 0.2506 -0.2116 0.003681 0.03020
## s58 0.031488 -1.2450 0.048065 0.28546 0.2514 -0.2284 0.003903 0.03146
## s59 0.007871 -1.2439 0.046860 0.30051 0.2497 -0.2452 0.003855 0.03595
## s60 0.006227 -1.2448 0.046705 0.30098 0.2497 -0.2455 0.003852 0.03623
## s61 . -1.2445 0.046044 0.30354 0.2467 -0.2495 0.003817 0.03786

```

```

## s62 . -1.2409 0.045587 0.30500 0.2414 -0.2526 0.003763 0.03939
## s63 . -1.2372 0.045239 0.30655 0.2359 -0.2546 0.003726 0.04058
## s64 -0.000397 -1.2339 0.044893 0.30756 0.2310 -0.2559 0.003719 0.04198
## s65 -0.001781 -1.2313 0.044479 0.30846 0.2260 -0.2557 0.003701 0.04326
## s66 -0.006327 -1.2298 0.044000 0.30919 0.2209 -0.2541 0.003670 0.04452
## s67 -0.011242 -1.2293 0.043595 0.30965 0.2167 -0.2520 0.003637 0.04553
## s68 -0.015471 -1.2293 0.043307 0.30992 0.2136 -0.2503 0.003610 0.04628
## s69 -0.020136 -1.2294 0.043021 0.31014 0.2104 -0.2483 0.003581 0.04703
## s70 -0.025257 -1.2297 0.042733 0.31023 0.2071 -0.2462 0.003550 0.04778
## s71 -0.028731 -1.2299 0.042542 0.31023 0.2050 -0.2447 0.003528 0.04826
## s72 -0.030431 -1.2301 0.042454 0.31027 0.2039 -0.2439 0.003518 0.04850
## s73 -0.033929 -1.2303 0.042275 0.31035 0.2016 -0.2423 0.003497 0.04897
## s74 -0.035686 -1.2305 0.042185 0.31039 0.2005 -0.2415 0.003486 0.04922
## s75 -0.039319 -1.2308 0.042001 0.31042 0.1983 -0.2397 0.003462 0.04970
## s76 -0.041169 -1.2310 0.041905 0.31041 0.1972 -0.2388 0.003449 0.04994
## s77 -0.043025 -1.2313 0.041809 0.31040 0.1960 -0.2379 0.003436 0.05018
## s78 -0.044893 -1.2315 0.041714 0.31038 0.1949 -0.2370 0.003423 0.05042
## s79 -0.046770 -1.2317 0.041617 0.31035 0.1938 -0.2360 0.003410 0.05066
## s80 -0.048664 -1.2319 0.041520 0.31030 0.1927 -0.2350 0.003396 0.05089
## s81 -0.050577 -1.2322 0.041423 0.31023 0.1916 -0.2340 0.003382 0.05113
## s82 -0.052502 -1.2324 0.041326 0.31015 0.1905 -0.2330 0.003367 0.05135
## s83 -0.054435 -1.2327 0.041229 0.31005 0.1894 -0.2320 0.003352 0.05158
## s84 -0.056374 -1.2330 0.041133 0.30994 0.1883 -0.2309 0.003337 0.05180
## s85 -0.058316 -1.2332 0.041037 0.30981 0.1872 -0.2298 0.003322 0.05202
## s86 -0.060257 -1.2335 0.040941 0.30968 0.1862 -0.2287 0.003306 0.05223
## s87 -0.062196 -1.2338 0.040845 0.30953 0.1851 -0.2276 0.003290 0.05245
## s88 -0.064130 -1.2341 0.040751 0.30937 0.1840 -0.2264 0.003274 0.05266
## s89 -0.066057 -1.2344 0.040657 0.30921 0.1830 -0.2253 0.003258 0.05286
## s90 -0.067975 -1.2346 0.040564 0.30904 0.1819 -0.2241 0.003242 0.05306
## s91 -0.069883 -1.2349 0.040471 0.30886 0.1809 -0.2229 0.003226 0.05326
## s92 -0.071778 -1.2352 0.040380 0.30868 0.1798 -0.2217 0.003209 0.05346
## s93 -0.073661 -1.2355 0.040289 0.30850 0.1788 -0.2205 0.003193 0.05365
## s94 -0.075529 -1.2358 0.040200 0.30831 0.1778 -0.2193 0.003176 0.05384
## s95 -0.077382 -1.2361 0.040112 0.30813 0.1768 -0.2181 0.003159 0.05403
## s96 -0.079220 -1.2363 0.040025 0.30794 0.1757 -0.2168 0.003142 0.05421
## s97 -0.081041 -1.2366 0.039939 0.30775 0.1748 -0.2156 0.003126 0.05439
## s98 -0.082845 -1.2369 0.039854 0.30756 0.1738 -0.2144 0.003109 0.05457
## s99 -0.084633 -1.2372 0.039771 0.30736 0.1728 -0.2131 0.003092 0.05475
##
## s0 .
## s1 .
## s2 .
## s3 .
## s4 .
## s5 .
## s6 .
## s7 .
## s8 .
## s9 .
## s10 .
## s11 .
## s12 .
## s13 .
## s14 .

```

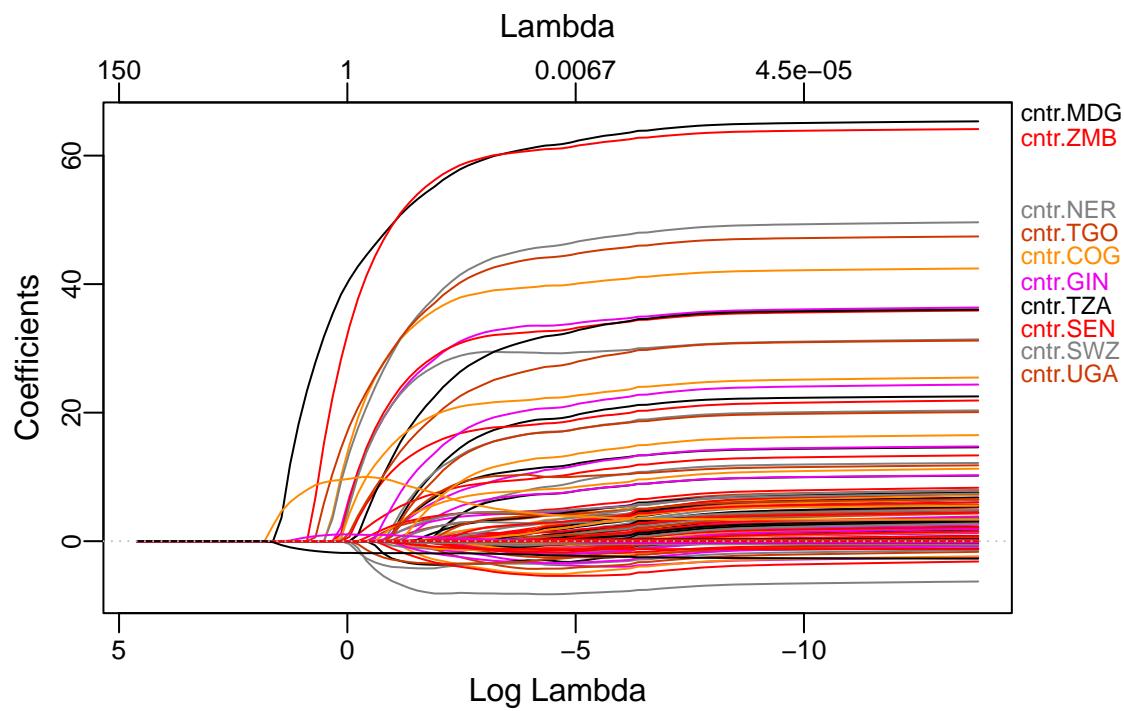
```

## s15   .
## s16   .      -0.164
## s17   .      -0.612
## s18   .      -0.955
## s19   .      -1.192
## s20   .      -1.388
## s21   .      -1.539
## s22   .      -1.662
## s23   .      -1.757
## s24   .      -1.826
## s25   .      -1.829
## s26   .      -1.848
## s27 -0.000459 -1.865
## s28 -0.001153 -1.863
## s29 -0.001542 -1.859
## s30 -0.002152 -1.861
## s31 -0.002372 -1.853
## s32 -0.002961 -1.864
## s33 -0.003293 -1.877
## s34 -0.003650 -1.890
## s35 -0.004237 -1.906
## s36 -0.004577 -1.897
## s37 -0.004902 -1.892
## s38 -0.005287 -1.884
## s39 -0.005513 -1.896
## s40 -0.005815 -1.914
## s41 -0.006134 -1.936
## s42 -0.006252 -1.964
## s43 -0.006392 -1.985
## s44 -0.006601 -2.010
## s45 -0.006779 -2.029
## s46 -0.006826 -2.077
## s47 -0.006905 -2.124
## s48 -0.006907 -2.195
## s49 -0.006877 -2.217
## s50 -0.007126 -2.232
## s51 -0.007461 -2.257
## s52 -0.007589 -2.286
## s53 -0.007641 -2.309
## s54 -0.007719 -2.334
## s55 -0.007837 -2.366
## s56 -0.008526 -2.382
## s57 -0.008736 -2.400
## s58 -0.009023 -2.419
## s59 -0.009001 -2.510
## s60 -0.009078 -2.513
## s61 -0.009480 -2.526
## s62 -0.009805 -2.538
## s63 -0.010061 -2.548
## s64 -0.010310 -2.560
## s65 -0.010540 -2.572
## s66 -0.010751 -2.585
## s67 -0.010905 -2.596
## s68 -0.011018 -2.604

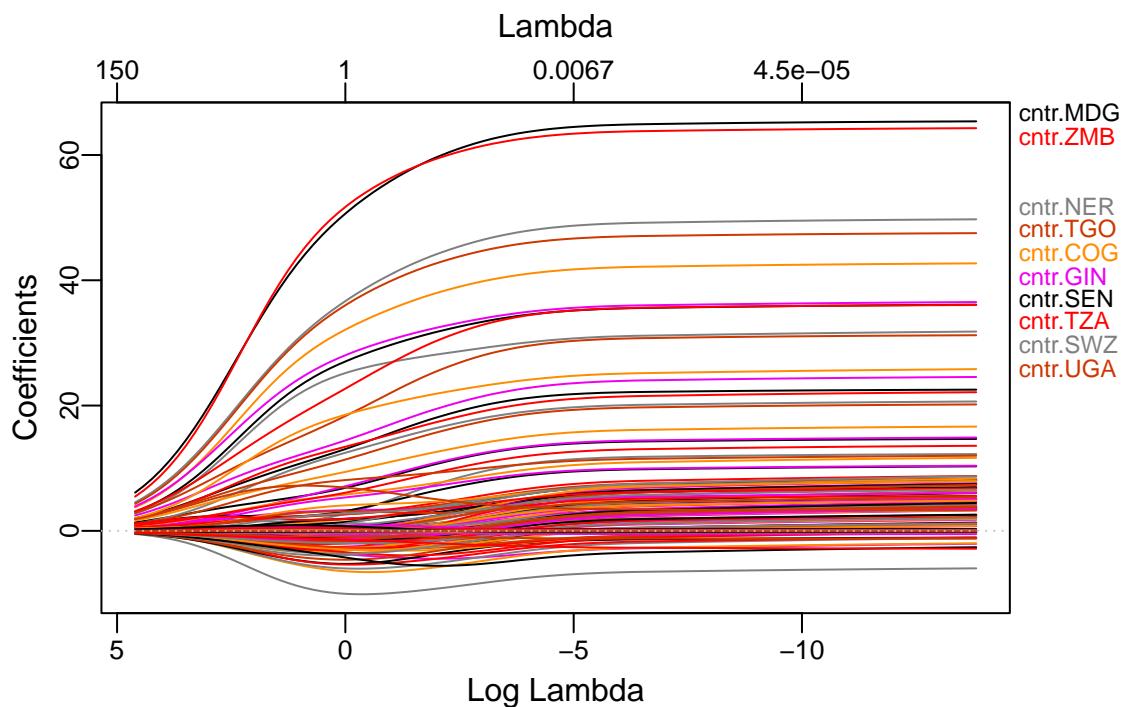
```

```
## s69 -0.011129 -2.613
## s70 -0.011238 -2.622
## s71 -0.011307 -2.628
## s72 -0.011343 -2.631
## s73 -0.011412 -2.637
## s74 -0.011446 -2.640
## s75 -0.011512 -2.646
## s76 -0.011544 -2.649
## s77 -0.011575 -2.652
## s78 -0.011606 -2.655
## s79 -0.011636 -2.658
## s80 -0.011665 -2.661
## s81 -0.011694 -2.664
## s82 -0.011722 -2.667
## s83 -0.011749 -2.670
## s84 -0.011775 -2.673
## s85 -0.011801 -2.676
## s86 -0.011826 -2.679
## s87 -0.011850 -2.682
## s88 -0.011873 -2.685
## s89 -0.011895 -2.688
## s90 -0.011917 -2.691
## s91 -0.011937 -2.694
## s92 -0.011957 -2.697
## s93 -0.011976 -2.700
## s94 -0.011995 -2.702
## s95 -0.012013 -2.705
## s96 -0.012029 -2.708
## s97 -0.012046 -2.711
## s98 -0.012061 -2.714
## s99 -0.012076 -2.716
```

```
plot_glmnet(lasso_model)
```



```
plot_glmnet(ridge_model)
```



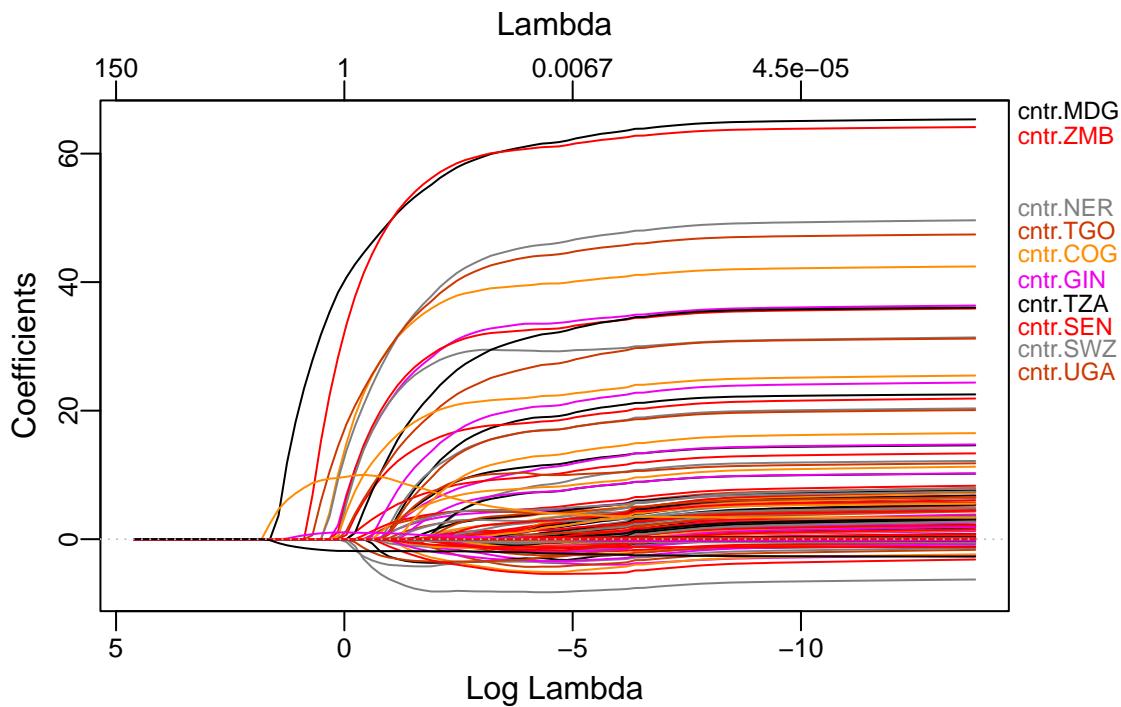
```

fit2 <- predict(glm_Lasso, train.x)
true2 <- train.y
summary_Lasso_train <- eval_results(fit2, true2)
summary_Lasso_train

##      MSE   MAE RMSE MAPE      R2
## 1 4.94 1.48 2.22  Inf 0.948

plot_glmnet(lasso_model)

```



```
fit3 <- predict(glm_Lasso, test.x)
true3 <- test.y
summary_Lasso_test <- eval_results(fit3, true3)
summary_Lasso_test
```

```
##      MSE MAE RMSE MAPE     R2
## 1 4.67 1.5 2.16 Inf 0.846
```

```
summary_lasso <- rbind(summary_Lasso_train, summary_Lasso_test)
rownames(summary_lasso) <- c("LASSO_train", "LASSO_test")
knitr::kable(summary_lasso, digits = 3)
```

	MSE	MAE	RMSE	MAPE	R2
LASSO_train	4.94	1.48	2.22	Inf	0.948
LASSO_test	4.67	1.50	2.16	Inf	0.846

3.2.2. Assessment

This section recapitulates the three models that we have generated so far through the method of gradually removing variables from the original dataset.

```

summary(lm_4)

##
## Call:
## lm(formula = pov ~ . - hlth - lgdp.dflt - lfdi, data = train_1)
##
## Residuals:
##    Min     1Q Median     3Q    Max 
## -24.36  -2.61  -0.36   1.99  50.76 
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)    
## (Intercept) 186.36637  89.75554   2.08  0.03825 *  
## edu.total   -0.73952   0.18026  -4.10  4.6e-05 *** 
## gcf        -0.111112  0.03701  -3.00  0.00278 **  
## incomeL      7.31879   2.33605   3.13  0.00181 **  
## incomeLM     -3.33101   1.46529  -2.27  0.02333 *  
## incomeUM     -3.57242   0.97541  -3.66  0.00027 *** 
## lbr.part      0.21730   0.03280   6.62  7.3e-11 *** 
## mil         -0.93100   0.21016  -4.43  1.1e-05 *** 
## pop.gwth.rural  0.57354   0.19201   2.99  0.00292 ** 
## pop.gwth.urban  1.90680   0.20117   9.48 < 2e-16 *** 
## trade        -0.01696   0.00533  -3.18  0.00153 **  
## unemp         0.30422   0.05783   5.26  1.9e-07 *** 
## year          -0.07992   0.04469  -1.79  0.07416 .  
## lgdp.pc       -3.13114   0.54157  -5.78  1.1e-08 *** 
## --- 
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 5.86 on 654 degrees of freedom
## Multiple R-squared:  0.586, Adjusted R-squared:  0.578 
## F-statistic: 71.3 on 13 and 654 DF, p-value: <2e-16

```

Our first model, `lm_4`, generated with regular multiple linear regression exhibits the lowest R^2 value. The model shows that `incomeL` is the most significant predictor variable with the highest coefficient. One interesting observation is that `incomeLM` and `incomeUM` has a negative correlation with the `pov` level, in contrast to `incomeL`. This shows that countries that have lower income tend to have high `pov` value. On the other hand, countries that are classified to have lower middle income and upper middle income seem to have lower `pov` values, as shown by their negative coefficients in the regression model.

Apart from income, some other variables that contribute more significantly to the model are `lgdp.pc`, with a coefficient of -3.13114, exhibiting a stronger negative correlation, and `pop.gwth.urban`, exhibiting a stronger positive correlation with a coefficient of 1.90680. This is counter-intuitive, as the notion is that population growth in cities will help drive positive economic change due to the number of productive people living in the city. The model also shows that `pop.gwth.rural` gives a less significant impact on `pov` as compared to `pop.gwth.rural`.

```
knitr::kable(summary_lasso, digits = 3)
```

	MSE	MAE	RMSE	MAPE	R2
LASSO_train	4.94	1.48	2.22	Inf	0.948

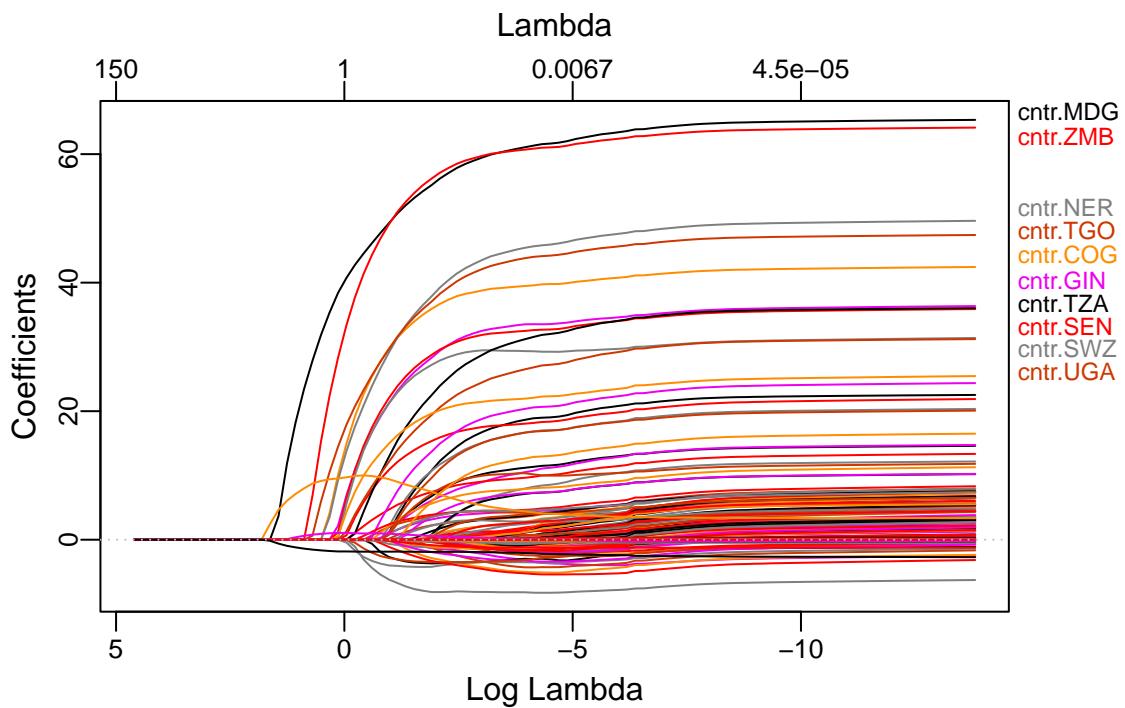
	MSE	MAE	RMSE	MAPE	R2
LASSO_test	4.67	1.50	2.16	Inf	0.846

```
knitr::kable(summary_Ridge, digits = 3)
```

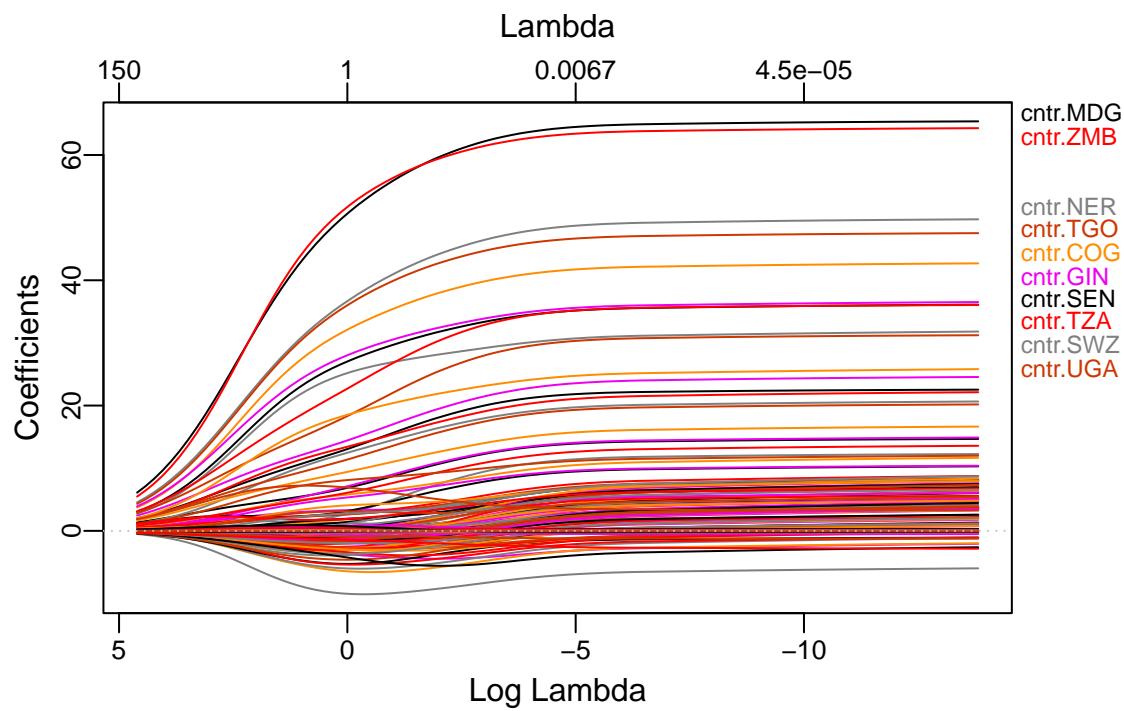
	MSE	MAE	RMSE	MAPE	R2
Ridge_train	5.37	1.47	2.32	Inf	0.943
Ridge_test	4.42	1.46	2.10	Inf	0.854

The MSE for the LASSO regression model is slightly higher than that of the Ridge regression model. This is to be expected, as the Ridge regression model aims for an improved accuracy while the LASSO model aims for an improved interpretability. Both models seem to exhibit very high R^2 value for both the training and testing dataset.

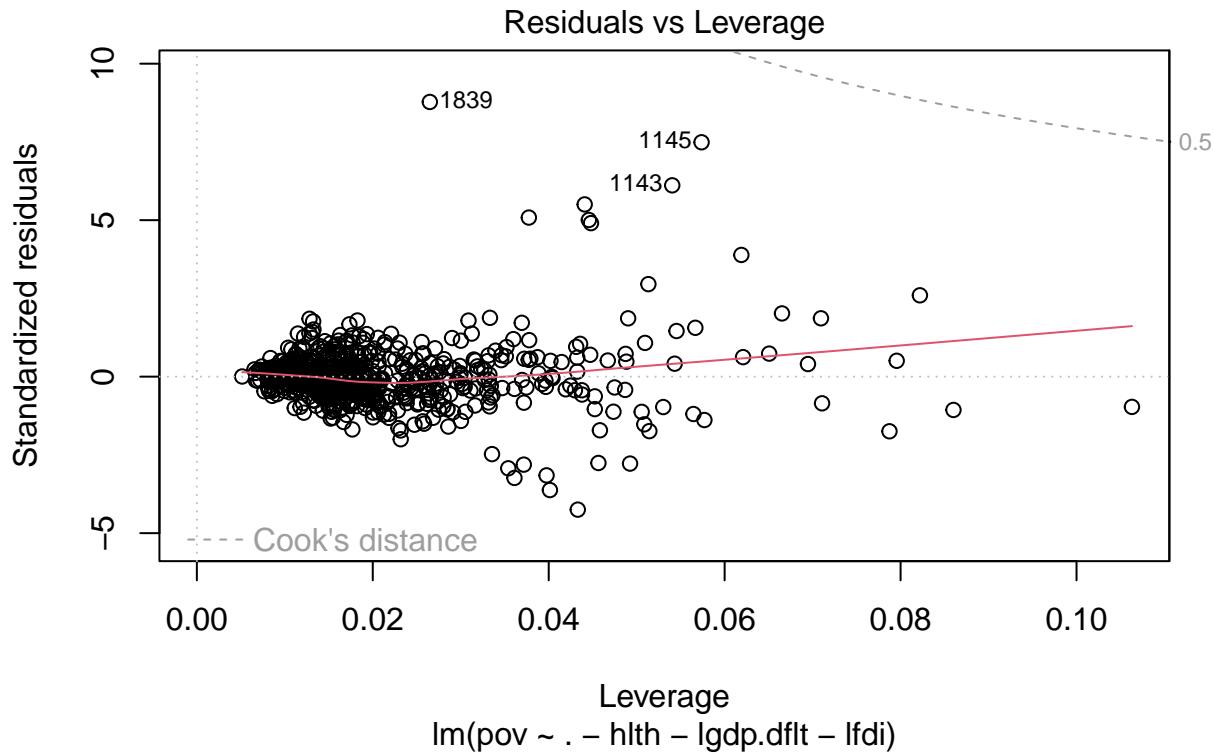
```
plot_glmnet(lasso_model)
```



```
plot_glmnet(ridge_model)
```



```
plot(lm_4, which = 5)
```



The Residual vs Leverage plot of `lm_4` suggests that heteroskedasticity might be present in the model. This means that the variance of the residuals is unequal over a range of observed values. This problem exists in the dataset rather than the models generated.

3.3. Forward variable selection

3.3.1. Model Fitting

```
df_count <- read.csv("../data/countries1.csv")
set.seed(10)
index <- sort(sample(x = nrow(df_count), size = nrow(df_count) *
  0.8))
train <- df_count[index, ]
test <- df_count[-index, ]
```

Through forward selection, we will do feature selection. We will stop the forward selection when the adjusted r-value > 0.75.

Through iteration of different models. A linear model consisting of income as a predictor has the highest adjusted r-square value = 0.6201, thus it is selected.

```
summary(lm(formula = pov ~ income, data = train)) #0.6201
```

```
##
```

```

## Call:
## lm(formula = pov ~ income, data = train)
##
## Residuals:
##    Min     1Q Median     3Q    Max 
## -41.90  -3.76  -0.25   0.75  74.84 
## 
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)    
## (Intercept)  0.352     0.498    0.71    0.48    
## incomeL     42.846    0.917   46.70   < 2e-16 *** 
## incomeLM    10.204    0.744   13.71   < 2e-16 *** 
## incomeUM     3.810    0.774    4.92   9.6e-07 *** 
## ---      
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1 
## 
## Residual standard error: 11 on 1422 degrees of freedom
##   (48 observations deleted due to missingness)
## Multiple R-squared:  0.621, Adjusted R-squared:  0.62 
## F-statistic: 776 on 3 and 1422 DF, p-value: <2e-16

```

Since income as a predictor brought the highest adjusted r-value, we add another predictor to income to find the predictor added to income which produced the highest adjusted r-value. reg as a predictor has the highest adjusted r-value=0.7181

```
summary(lm(formula = pov ~ reg + income, data = train)) #0.7181
```

```

## 
## Call:
## lm(formula = pov ~ reg + income, data = train)
##
## Residuals:
##    Min     1Q Median     3Q    Max 
## -39.80  -3.82   0.09   1.41  55.84 
## 
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)    
## (Intercept)  4.794     1.005    4.77  2.0e-06 *** 
## regEurope & Central Asia -4.787     0.957   -5.00  6.4e-07 *** 
## regLatin America & Caribbean  0.130     0.987    0.13   0.895  
## regMiddle East & North Africa -6.309     1.346   -4.69  3.0e-06 *** 
## regNorth America      -4.112     1.929   -2.13   0.033 *  
## regSouth Asia        0.608     1.800    0.34   0.736  
## regSub-Saharan Africa 18.682     1.218   15.34   < 2e-16 *** 
## incomeL            28.128     1.070   26.29   < 2e-16 *** 
## incomeLM           6.081     0.757    8.03  2.0e-15 *** 
## incomeUM           0.600     0.755    0.79   0.427  
## ---      
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1 
## 
## Residual standard error: 9.46 on 1416 degrees of freedom
##   (48 observations deleted due to missingness)
## Multiple R-squared:  0.72, Adjusted R-squared:  0.718 
## F-statistic: 404 on 9 and 1416 DF, p-value: <2e-16

```

The forward selection process continues and lgdp.pc is the new predictor added that produced highest adjusted r-square=0.7492

```
summary(lm(formula = pov ~ lgdp.pc + income + reg,
  data = train)) #0.7492

##
## Call:
## lm(formula = pov ~ lgdp.pc + income + reg, data = train)
##
## Residuals:
##   Min     1Q Median     3Q    Max
## -36.22  -3.73  -0.27   2.95  54.42
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 70.830     5.106   13.87 < 2e-16 ***
## lgdp.pc      -6.439     0.489  -13.16 < 2e-16 ***
## incomeL       2.658     2.185    1.22   0.22
## incomeLM     -10.440    1.444   -7.23 7.8e-13 ***
## incomeUM     -8.838     1.009   -8.76 < 2e-16 ***
## regEurope & Central Asia -3.941     0.903   -4.37 1.4e-05 ***
## regLatin America & Caribbean 0.178     0.929    0.19   0.85
## regMiddle East & North Africa -6.795     1.271   -5.35 1.0e-07 ***
## regNorth America      -1.918     1.822   -1.05   0.29
## regSouth Asia        -0.544     1.695   -0.32   0.75
## regSub-Saharan Africa 17.945     1.149   15.62 < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ',' 1
##
## Residual standard error: 8.89 on 1412 degrees of freedom
##   (51 observations deleted due to missingness)
## Multiple R-squared:  0.751, Adjusted R-squared:  0.749
## F-statistic: 426 on 10 and 1412 DF, p-value: <2e-16
```

The addition of predictor gdp.pc produces the highest adjusted r-square value above 0.75, adjusted r-square=0.7703

```
summary(lm(formula = pov ~ gdp.pc + income + reg +
lgdp.pc, data = train)) #0.7703

##
## Call:
## lm(formula = pov ~ gdp.pc + income + reg + lgdp.pc, data = train)
##
## Residuals:
##   Min     1Q Median     3Q    Max
## -34.77  -3.13   0.38   2.17  53.64
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.11e+02   6.01e+00   18.44 < 2e-16 ***
## gdp.pc      2.75e-04   2.40e-05   11.42 < 2e-16 ***
```

```

## incomeL           -6.94e+00  2.25e+00  -3.08   0.0021  **
## incomeLM          -1.39e+01  1.41e+00  -9.82  < 2e-16 ***
## incomeUM          -8.22e+00  9.67e-01  -8.49  < 2e-16 ***
## regEurope & Central Asia -4.06e+00  8.64e-01  -4.70  2.9e-06 ***
## regLatin America & Caribbean 4.25e-01  8.89e-01   0.48   0.6324
## regMiddle East & North Africa -6.47e+00  1.22e+00  -5.32  1.2e-07 ***
## regNorth America      -2.34e+00  1.74e+00  -1.34   0.1793
## regSouth Asia         -1.44e+00  1.62e+00  -0.89   0.3746
## regSub-Saharan Africa 1.73e+01  1.10e+00  15.71  < 2e-16 ***
## lgdp.pc              -1.13e+01  6.31e-01  -17.86 < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 8.51 on 1411 degrees of freedom
##   (51 observations deleted due to missingness)
## Multiple R-squared:  0.772, Adjusted R-squared:  0.77
## F-statistic:  435 on 11 and 1411 DF, p-value: <2e-16

```

Fitting of multiple linear regression with predictor gdp.pc, income, reg and lgd.pc.

```

lm1 <- lm(formula = pov ~ gdp.pc + income + reg + lgdp.pc,
           data = train)
summary(lm1)

```

```

##
## Call:
## lm(formula = pov ~ gdp.pc + income + reg + lgdp.pc, data = train)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -34.77  -3.13   0.38   2.17  53.64
##
## Coefficients:
##                               Estimate Std. Error t value Pr(>|t|)
## (Intercept)                1.11e+02  6.01e+00 18.44  < 2e-16 ***
## gdp.pc                     2.75e-04  2.40e-05 11.42  < 2e-16 ***
## incomeL                   -6.94e+00  2.25e+00 -3.08   0.0021  **
## incomeLM                  -1.39e+01  1.41e+00 -9.82  < 2e-16 ***
## incomeUM                  -8.22e+00  9.67e-01 -8.49  < 2e-16 ***
## regEurope & Central Asia -4.06e+00  8.64e-01 -4.70  2.9e-06 ***
## regLatin America & Caribbean 4.25e-01  8.89e-01   0.48   0.6324
## regMiddle East & North Africa -6.47e+00  1.22e+00 -5.32  1.2e-07 ***
## regNorth America          -2.34e+00  1.74e+00 -1.34   0.1793
## regSouth Asia             -1.44e+00  1.62e+00 -0.89   0.3746
## regSub-Saharan Africa     1.73e+01  1.10e+00 15.71  < 2e-16 ***
## lgdp.pc                   -1.13e+01  6.31e-01 -17.86 < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 8.51 on 1411 degrees of freedom
##   (51 observations deleted due to missingness)
## Multiple R-squared:  0.772, Adjusted R-squared:  0.77
## F-statistic:  435 on 11 and 1411 DF, p-value: <2e-16

```

3.3.2. Assessment

Assessment of linear model lm1: the predicted model has some NA values which are removed. The error rates are too high, therefore, this method of variable selection may not be suitable

```
eval.metrics.linreg <- function(actual, predicted) {  
  residual <- actual - predicted  
  mse <- mean(residual^2)  
  mae <- mean(abs(residual))  
  rmse <- sqrt(mse)  
  mape <- mean(abs(residual/actual)) * 100  
  
  data.frame(MSE = mse, MAE = mae, RMSE = rmse, MAPE = mape)  
}  
predicted <- predict(lm1, newdata = test)  
actual <- test$pov  
df23 <- na.omit(data.frame(actual, predicted))  
ac <- df23$actual  
pr <- df23$predicted  
eval.metrics.linreg(ac, pr)  
  
##      MSE   MAE  RMSE  MAPE  
## 1 56.8 4.76 7.53  Inf
```

3.3.3. Interpretation

gdp per capital, income class and region of the country can provide a model to predict the poverty headcount ratio in a country. The error rate of the model is high on the testing data set, showing it may not be a good model.

3.4. Imputation

What is imputation?

Imputation is a technique to handle missing values by replacing missing data with substitute values. In our study, we focus on “item imputation”, which means substituting for a component of a data point (i.e. a variable). The general idea is to take a value that preserve the property of the data (e.g., distribution, mean, standard deviation), and use other variables to predict the missing values (much like regression). Since we observed some correlations during descriptive analysis, we expect the imputation algorithm to work well.

We use **Multivariate Imputation By Chained Equations (MICE)** as primary tool for this technique, because of its wide acceptance in scientific studies (Alruhaymi and Kim (2021)). In order to use this technique, we have to assume that the “missingness” of a field can be explained by the values in other columns, (e.g., If the countries is in North America, it’s more likely to be missing as we have discovered in 2.2). The general ideas of the algorithm is to fill in the missing values, and improve it iteratively until predicted values converge to a stable point. The algorithmic details of MICE is very concisely (and enthrallingly) explained in Gopalan (2020).

We use relatively small parameters in the interest of time.

- `m = 3` imputed data sets
- `maxit = 20` iterations

The imputation method is `cart` (Classification and Regression Trees).

```

set.seed(1984)
# split data, make sure all the countries
seeds <- countries1 %>%
  group_by(country.code) %>%
  filter(row_number() == 1)

plants <- countries1 %>%
  group_by(country.code) %>%
  filter(row_number() != 1)

isComplete <- which(complete.cases(plants))
idx <- sample(isComplete, replace = F, 0.2 * nrow(plants))

countries.train.4 <- plants[-idx, ] %>%
  rbind(seeds)
countries.test.4 <- plants[idx, ]

meth <- c(rep("", 4), "cart", "", rep("cart", 16))
names(meth) <- colnames(countries1)
meth

##   country.code   country.name      year      pov      income
##           ""          ""        ""       ""      "cart"
##       reg     edu.total      hlth      mil       fdi
##           ""        "cart"      "cart"    "cart"    "cart"
##     lbr.part    unemp.pop.gwth.total pop.gwth.rural pop.gwth.urban
##       "cart"      "cart"      "cart"    "cart"    "cart"
##     gdp.dflt      gcf       trade    gdp.pc     lfdi
##       "cart"      "cart"      "cart"    "cart"    "cart"
##   lgdp.dflt    lgdp.pc
##       "cart"      "cart"

# run the algorithm, or read from the file we
# generated as this is a very time-consuming
# process
countries1.imputed <-
# mice(countries.train.4, m = 3, maxit = 20,
# method = meth) saveRDS(countries1.imputed,
# 'countries1.imputed.RData')
countries1.imputed <- readRDS("countries1.imputed.RData")
summary(countries1.imputed)

## Class: mids
## Number of multiple imputations: 3
## Imputation methods:
##   country.code   country.name      year      pov      income
##           ""          ""        ""       ""      "cart"
##       reg     edu.total      hlth      mil       fdi
##           ""        "cart"      "cart"    "cart"    "cart"
##     lbr.part    unemp.pop.gwth.total pop.gwth.rural pop.gwth.urban
##       "cart"      "cart"      ""      "cart"    "cart"
##     gdp.dflt      gcf       trade    gdp.pc     lfdi

```

```

##      "cart"      "cart"      "cart"      "cart"      "cart"
##    lgdp.dflt    lgdp.pc      "cart"      "cart"      "cart"
##      "cart"      "cart"
## PredictorMatrix:
##      country.code country.name year pov income reg edu.total hlth mil
## country.code          0          0   1   1     1   1       1   1   1
## country.name          1          0   1   1     1   1       1   1   1
## year                  1          0   0   1     1   1       1   1   1
## pov                   1          0   1   0     1   1       1   1   1
## income                1          0   1   1     0   1       1   1   1
## reg                   1          0   1   1     1   0       1   1   1
##      fdi lbr.part unemp pop.gwth.total pop.gwth.rural pop.gwth.urban
## country.code          1          1   1           1           1           1
## country.name          1          1   1           1           1           1
## year                  1          1   1           1           1           1
## pov                   1          1   1           1           1           1
## income                1          1   1           1           1           1
## reg                   1          1   1           1           1           1
##      gdp.dflt gcf trade gdp.pc lfdi lgdp.dflt lgdp.pc
## country.code          1   1   1     1   1       1   1
## country.name          1   1   1     1   1       1   1
## year                  1   1   1     1   1       1   1
## pov                   1   1   1     1   1       1   1
## income                1   1   1     1   1       1   1
## reg                   1   1   1     1   1       1   1
## Number of logged events: 961
##      it im      dep     meth
## 1  0 0      constant
## 2  1 1      income      cart
## 3  1 1  edu.total      cart
## 4  1 1      hlth      cart
## 5  1 1      mil       cart
## 6  1 1      fdi       cart
##
## 1
## 2
## 3 country.codeARE, country.codeBIH, country.codeCOD, country.codeDZA, country.codeEAS, country.codeE
## 4
## 5      country.codeBTN, country.codeCOM, country.codeDJI, country.codeEAS, country.codeECS, coun
## 6

```

We can take a look into one of the imputed data sets.

```

countries1.imputed.1 <- complete(countries1.imputed,
  1)
summary(countries1.imputed.1)

```

	country.code	country.name	year	pov	income
## BRA	: 32	Length:1508	Min. :1967	Min. : 0.0	H :484
## USA	: 30	Class :character	1st Qu.:1999	1st Qu.: 0.3	L :247
## HND	: 26	Mode :character	Median :2007	Median : 1.9	LM:429
## IDN	: 24		Mean :2006	Mean :11.3	UM:348
## ARG	: 23		3rd Qu.:2014	3rd Qu.:13.6	

```

##   CRI      : 23                         Max.    :2021   Max.    :91.5
## (Other):1350
##           reg      edu.total      hlth
## East Asia & Pacific    :145   Min.    : 1.03   Min.    : 1.72
## Europe & Central Asia  :664   1st Qu.: 3.41   1st Qu.: 4.90
## Latin America & Caribbean:322   Median  : 4.41   Median  : 6.72
## Middle East & North Africa: 90   Mean    : 4.51   Mean    : 6.78
## North America            : 46   3rd Qu.: 5.43   3rd Qu.: 8.37
## South Asia               : 46   Max.    :15.75   Max.    :17.73
## Sub-Saharan Africa       :195
##           mil      fdi      lbr.part      unemp
## Min.    : 0.00   Min.    :-3.44e+11   Min.    :30.5   Min.    : 0.2
## 1st Qu.: 1.05   1st Qu.: 2.11e+08   1st Qu.:56.0   1st Qu.: 4.2
## Median  : 1.49   Median  : 1.28e+09   Median  :61.4   Median  : 6.7
## Mean    : 1.83   Mean    : 1.45e+10   Mean    :61.3   Mean    : 8.2
## 3rd Qu.: 2.14   3rd Qu.: 8.06e+09   3rd Qu.:66.2   3rd Qu.:10.3
## Max.    :19.38   Max.    : 7.34e+11   Max.    :93.0   Max.    :49.7
##
##           pop.gwth.total  pop.gwth.rural  pop.gwth.urban  gdp.dflt      gcf
## Min.    :-3.63   Min.    :-8.56     Min.    :-4.08     Min.    :-26     Min.    : 0.0
## 1st Qu.: 0.31   1st Qu.: -0.79    1st Qu.: 0.55     1st Qu.:  2     1st Qu.:19.6
## Median  : 1.10   Median  : 0.03     Median  : 1.60     Median  : 4     Median :22.7
## Mean    : 1.13   Mean    : 0.09     Mean    : 1.79     Mean    : 21     Mean    :23.8
## 3rd Qu.: 1.90   3rd Qu.: 1.05     3rd Qu.: 2.85     3rd Qu.:  9     3rd Qu.:26.8
## Max.    : 5.61   Max.    : 4.60     Max.    :13.80     Max.    :3334    Max.    :69.5
##
##           trade      gdp.pc      lfdi      lgdp.dflt      lgdp.pc
## Min.    : 1     Min.    : 120     Min.    : 6.91     Min.    :-2.92    Min.    : 4.78
## 1st Qu.: 50    1st Qu.: 1579    1st Qu.:19.17    1st Qu.: 0.55    1st Qu.: 7.36
## Median  : 71    Median  : 5176    Median :20.97    Median  : 1.38    Median : 8.55
## Mean    : 83    Mean    : 14378   Mean   :20.39    Mean    : 1.25    Mean   : 8.56
## 3rd Qu.:104    3rd Qu.: 19598   3rd Qu.:22.80    3rd Qu.: 2.22    3rd Qu.: 9.88
## Max.    :380    Max.    :123679   Max.   :27.32    Max.    : 8.11    Max.   :11.73
##
sum(!complete.cases(countries1.imputed.1))

```

```
## [1] 0
```

We found no missing cases as expected.

3.4.1. Model Fitting

A. Ordinary Linear Regression We generated 3 sets of imputed (training) data. We can build separate models using each data set, and combine the estimates using [pooling rule](#). With all the generated data sets.

```

formula.str <- "pov ~ country.code+year+income+edu.total+hlth+mil+fdi+lbr.part+unemp+pop.gwth.total+pop

# Using single imputed data set
fit1 <- lapply(1:countries1.imputed$m, function(i) lm(as.formula(formula.str),
  data = complete(countries1.imputed, i)))

```

```

# Build models with all generated data set and
# pool the estimates
fit2 <- with(data = countries1.imputed, exp = lm(as.formula(formula.str)))
fit2.combined <- pool(fit2)
# dummy lm model
fit2.dummy <- lm(as.formula(formula.str), data = complete(countries1.imputed,
  1))
# replace coefficients of dummy model
name.coef <- names(fit2.dummy$coefficients)
fit2.dummy$coefficients <- fit2.combined$pooled$estimate
names(fit2.dummy$coefficients) <- name.coef

```

Helper function to calculate R-Squared

```

r2 <- function(pred, orig) {
  RSS <- sum((pred - orig)^2)
  TSS <- sum((orig - mean(orig))^2)
  R2 <- 1 - RSS/TSS
  return(R2)
}

adjR2 <- function(pred, orig, k) {
  R2 <- r2(pred, orig)
  n <- length(pred)
  adjr2 <- 1 - (1 - R2) * (n - 1)/(n - k - 1)
  return(adjr2)
}

```

R-Squared on train and test

```

fit1.summ <- lapply(fit1, function(f) {
  # number of variables
  k <- length(f$coefficients) - 1
  # predict value of test
  pred <- predict(f, countries.test.4)
  test.resid <- pred - countries.test.4$pov
  data.frame(`Train MSE` = mean(summary(f)$residuals^2),
    `Test MSE` = mean(test.resid^2), `Train MAE` = mean(abs(summary(f)$residuals)),
    `Test MAE` = mean(abs(test.resid)), `Train R2` = summary(f)$r.squared,
    `Test R2` = r2(pred, countries.test.4$pov))
})

# number of variables
k <- length(fit2.dummy$coefficients) - 1
# number of variables
fit2.pred <- predict(fit2.dummy, countries.test.4)
test.resid <- fit2.pred - countries.test.4$pov
fit2.summ <- data.frame(`Train MSE` = mean(summary(fit2.dummy)$residuals^2),
  `Test MSE` = mean(test.resid^2), `Train MAE` = mean(abs(summary(fit2.dummy)$residuals)),
  `Test MAE` = mean(abs(test.resid)), `Train R2` = pool.r.squared(fit2)[,
  "est"], `Test R2` = r2(fit2.pred, countries.test.4$pov))

res <- do.call(rbind.data.frame, fit1.summ)

```

```

res <- rbind(res, fit2.summ)
res <- cbind(data.frame(Set = c(1:3, "Pooled")), res)

kable(res)

```

Set	Train.MSE	Test.MSE	Train.MAE	Test.MAE	Train.R2	Test.R2
1	29.3	19.3	3.36	2.46	0.916	0.772
2	27.7	20.2	3.29	2.50	0.920	0.762
3	28.8	19.7	3.33	2.45	0.917	0.768
Pooled	29.3	19.6	3.36	2.45	0.918	0.769

Pooled model slightly improve performance. Test data fitting has better MSE and MAE than train data. However, R-Squared is lower in test data. There might be some over-fitting in our models.

```

# find coef of variable other than country.code
var <- c("year", "incomeL", "incomeLM", "incomeUM",
        "edu.total", "hlth", "mil", "fdi", "lbr.part",
        "unemp", "pop.gwth.total", "pop.gwth.rural", "pop.gwth.urban",
        "gdp.dflt", "gcf", "trade", "gdp.pc", "lfdi", "lgdp.dflt",
        "lgdp.pc")

res <- do.call(rbind, lapply(fit1, function(f) f$coefficients[var]))
res <- rbind(res, fit2.dummy$coefficients[var])
res <- cbind(data.frame(Set = c(1:3, "Pooled")), res)
kable(res)

```

Set	year	incomeL	incomeLM	incomeUM	edu.total	hlth	mil	fdi	lbr.part	unemp	pop.gwth.total	pop.gwth.rural	pop.gwth.urban	gdp.dflt	gcf	trade	gdp.pc	lfdi	lgdp.dflt	lgdp.pc
1	-	0.381	-	-	0.248	-	0	0.169	-	-	0.828	2.37	0.000	-	-	0	0.000	0.191	-	
	0.235	5.25	2.97	0.420	-	0.460	-	-	0.099	3.12	-	-	-	-	0.237	0.016	-	6.19	-	
2	-	1.086	-	-	-	0.416	-	0	0.287	0.010	-	0.789	2.14	0.001	-	0.002	0	-	0.116	-
	0.219	4.87	2.50	0.319	-	0.099	-	-	-	3.02	-	-	-	-	0.220	-	0.022	6.78	-	
3	-	-	-	-	-	0.305	-	0	0.158	-	-	0.843	2.31	0.000	-	-	0	-	0.035	-
	0.176	0.818	6.10	3.27	0.677	-	0.084	-	-	0.012	3.00	-	-	-	0.208	0.013	0.001	7.55	-	
Pooled	0.216	-	-	-	0.323	-	0	0.205	-	-	0.820	2.27	0.000	-	-	0	-	0.114	-	
	0.210	5.41	2.91	0.472	-	0.214	-	-	0.034	3.04	-	-	-	-	0.222	0.009	0.008	6.84	-	

In contradiction to expectation, `hlth` is positively correlated with `pov`, and lower-middle income countries (`incomeLM`) are less poor than high income countries (`incomeH`). `pop.gwth.total` is negatively related to `pov`, while `pop.gwth.rural` and `pop.gwth.urban` are positively related. They might be indications of over-fitting model.

```

# fit model on imputed set no. 1
summary(fit1[[1]])

```

```

## 
## Call:
## lm(formula = as.formula(formula.str), data = complete(countries1.imputed,
##           i))
## 
## 
```

```

## Residuals:
##      Min     1Q Median     3Q    Max
## -25.59  -2.00   0.05  1.92  35.96
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 5.41e+02  7.00e+01   7.73 2.1e-14 ***
## country.codeALB -1.75e+01  4.23e+00  -4.14 3.8e-05 ***
## country.codeARE -1.32e+01  5.78e+00  -2.28 0.02274 *
## country.codeARG -1.62e+01  3.98e+00  -4.06 5.2e-05 ***
## country.codeARM -1.50e+01  4.00e+00  -3.75 0.00018 ***
## country.codeAUS -1.79e+01  4.40e+00  -4.06 5.2e-05 ***
## country.codeAUT -1.75e+01  4.39e+00  -3.97 7.5e-05 ***
## country.codeAZE -2.56e+01  4.89e+00  -5.23 2.0e-07 ***
## country.codeBDI  1.98e+01  4.78e+00   4.14 3.7e-05 ***
## country.codeBEL -1.37e+01  4.39e+00  -3.12 0.00186 **
## country.codeBEN  1.72e+01  4.88e+00   3.52 0.00044 ***
## country.codeBFA  1.34e+01  4.35e+00   3.07 0.00216 **
## country.codeBGD -1.53e+01  4.28e+00  -3.58 0.00036 ***
## country.codeBGR -1.03e+01  4.29e+00  -2.39 0.01687 *
## country.codeBIH -1.47e+01  4.77e+00  -3.08 0.00214 **
## country.codeBLR -1.22e+01  4.01e+00  -3.04 0.00237 **
## country.codeBLZ  1.48e-01  4.33e+00   0.03 0.97272
## country.codeBOL -1.33e+01  3.79e+00  -3.51 0.00046 ***
## country.codeBRA -8.06e+00  3.85e+00  -2.09 0.03677 *
## country.codeBTN -1.36e+01  4.70e+00  -2.90 0.00384 **
## country.codeBWA  3.55e+00  4.44e+00   0.80 0.42399
## country.codeCAF  2.90e+01  5.44e+00   5.32 1.2e-07 ***
## country.codeCAN -1.89e+01  4.30e+00  -4.38 1.3e-05 ***
## country.codeCHE -2.35e+01  4.68e+00  -5.03 5.7e-07 ***
## country.codeCHL -1.17e+01  4.16e+00  -2.82 0.00487 **
## country.codeCHN  4.87e-01  4.08e+00   0.12 0.90495
## country.codeCIV -1.17e+01  3.97e+00  -2.96 0.00318 **
## country.codeCMR -5.74e-01  4.86e+00  -0.12 0.90604
## country.codeCOD  3.67e+01  5.47e+00   6.72 2.7e-11 ***
## country.codeCOG  2.39e+01  5.34e+00   4.47 8.3e-06 ***
## country.codeCOL -6.76e+00  3.91e+00  -1.73 0.08438 .
## country.codeCOM -1.14e+01  5.47e+00  -2.08 0.03787 *
## country.codeCPV -5.52e+00  4.90e+00  -1.13 0.25979
## country.codeCRI -1.55e+01  3.85e+00  -4.02 6.1e-05 ***
## country.codeCYP -1.43e+01  4.27e+00  -3.35 0.00084 ***
## country.codeCZE -1.46e+01  4.29e+00  -3.41 0.00066 ***
## country.codeDEU -1.84e+01  4.39e+00  -4.18 3.1e-05 ***
## country.codeDJI  1.06e+01  4.67e+00   2.26 0.02382 *
## country.codeDNK -1.90e+01  4.52e+00  -4.19 2.9e-05 ***
## country.codeDOM -1.38e+01  3.87e+00  -3.57 0.00037 ***
## country.codeDZA -1.49e+01  4.87e+00  -3.06 0.00228 **
## country.codeECU -9.32e+00  3.80e+00  -2.45 0.01439 *
## country.codeEGY -1.84e+01  4.11e+00  -4.48 8.1e-06 ***
## country.codeESP -1.49e+01  4.25e+00  -3.50 0.00048 ***
## country.codeEST -1.14e+01  4.32e+00  -2.65 0.00819 **
## country.codeETH -4.64e+00  4.49e+00  -1.03 0.30097
## country.codeFIN -1.72e+01  4.45e+00  -3.87 0.00011 ***
## country.codeFJI -1.38e+01  4.65e+00  -2.96 0.00310 **

```

```

## country.codeFRA -1.68e+01 4.32e+00 -3.89 0.00011 ***
## country.codeFSM -6.23e+00 5.64e+00 -1.11 0.26904
## country.codeGAB -8.68e+00 5.43e+00 -1.60 0.11001
## country.codeGBR -1.92e+01 4.30e+00 -4.46 8.8e-06 ***
## country.codeGEO -9.45e+00 4.04e+00 -2.34 0.01947 *
## country.codeGHA 1.16e+01 4.16e+00 2.78 0.00548 **
## country.codeGIN 5.83e+00 4.23e+00 1.38 0.16806
## country.codeGMB 2.33e+00 4.59e+00 0.51 0.61210
## country.codeGNB 1.12e+00 4.41e+00 0.25 0.79879
## country.codeGRC -1.36e+01 4.24e+00 -3.22 0.00132 **
## country.codeGTM -7.40e+00 4.25e+00 -1.74 0.08193 .
## country.codeGUY -2.52e+00 5.76e+00 -0.44 0.66194
## country.codeHND -4.52e+00 3.68e+00 -1.23 0.22001
## country.codeHRV -1.26e+01 4.38e+00 -2.88 0.00409 **
## country.codeHTI -1.24e+00 6.82e+00 -0.18 0.85533
## country.codeHUN -1.21e+01 4.22e+00 -2.86 0.00429 **
## country.codeIDN 3.78e+00 3.75e+00 1.01 0.31394
## country.codeIND 3.87e+00 3.97e+00 0.98 0.32954
## country.codeIRL -1.78e+01 4.38e+00 -4.06 5.1e-05 ***
## country.codeIRN -1.14e+01 3.94e+00 -2.90 0.00384 **
## country.codeIRQ -1.48e+01 5.39e+00 -2.75 0.00608 **
## country.codeISL -2.18e+01 4.44e+00 -4.91 1.1e-06 ***
## country.codeISR -1.53e+01 4.18e+00 -3.66 0.00026 ***
## country.codeITA -1.58e+01 4.31e+00 -3.67 0.00025 ***
## country.codeJAM -1.91e+01 4.49e+00 -4.24 2.4e-05 ***
## country.codeJOR -1.45e+01 4.30e+00 -3.36 0.00081 ***
## country.codeJPN -1.46e+01 5.62e+00 -2.59 0.00958 **
## country.codeKAZ -1.44e+01 3.96e+00 -3.64 0.00028 ***
## country.codeKEN -9.01e+00 4.40e+00 -2.05 0.04087 *
## country.codeKGZ -1.71e+01 3.84e+00 -4.44 9.7e-06 ***
## country.codeKIR -1.33e+01 5.52e+00 -2.41 0.01594 *
## country.codeKOR -1.27e+01 4.72e+00 -2.68 0.00743 **
## country.codeLAO -1.58e+01 4.27e+00 -3.69 0.00023 ***
## country.codeLBN -1.47e+01 6.82e+00 -2.15 0.03135 *
## country.codeLBR -2.63e+00 4.95e+00 -0.53 0.59500
## country.codeLCA 6.94e+00 5.59e+00 1.24 0.21449
## country.codeLKA -1.77e+01 4.27e+00 -4.16 3.4e-05 ***
## country.codeLSO 1.29e+01 4.70e+00 2.75 0.00599 **
## country.codeLTU -1.21e+01 4.38e+00 -2.76 0.00590 **
## country.codeLUX -2.37e+01 4.98e+00 -4.75 2.2e-06 ***
## country.codeLVA -1.09e+01 4.28e+00 -2.55 0.01103 *
## country.codeMAR -1.46e+01 4.27e+00 -3.42 0.00065 ***
## country.codeMDA -1.23e+01 4.08e+00 -3.01 0.00266 **
## country.codeMDG 2.63e+01 4.18e+00 6.29 4.2e-10 ***
## country.codeMDV -1.10e+01 4.63e+00 -2.37 0.01805 *
## country.codeMEX -1.05e+01 4.02e+00 -2.62 0.00879 **
## country.codeMHL -3.87e+00 7.23e+00 -0.53 0.59284
## country.codeMKD -5.39e+00 4.18e+00 -1.29 0.19800
## country.codeMLI 1.45e+01 4.61e+00 3.15 0.00166 **
## country.codeMLT -1.06e+01 4.64e+00 -2.29 0.02228 *
## country.codeMMR -1.55e+01 6.85e+00 -2.26 0.02368 *
## country.codeMNE -9.52e+00 4.33e+00 -2.20 0.02790 *
## country.codeMNG -1.17e+01 4.08e+00 -2.87 0.00416 **
## country.codeMOZ 3.65e+01 4.61e+00 7.91 5.2e-15 ***

```

```

## country.codeMRT -9.49e+00 4.07e+00 -2.33 0.01987 *
## country.codeMUS -1.36e+01 5.56e+00 -2.45 0.01431 *
## country.codeMWI 2.07e+01 4.53e+00 4.57 5.3e-06 ***
## country.codeMYS -1.76e+01 4.11e+00 -4.29 1.9e-05 ***
## country.codeNAM 6.98e+00 4.98e+00 1.40 0.16092
## country.codeNER 3.24e+01 4.33e+00 7.48 1.4e-13 ***
## country.codeNGA 9.25e+00 4.10e+00 2.26 0.02414 *
## country.codeNIC -1.15e+01 4.25e+00 -2.70 0.00712 **
## country.codeNLD -1.58e+01 4.53e+00 -3.48 0.00053 ***
## country.codeNOR -2.27e+01 4.62e+00 -4.91 1.0e-06 ***
## country.codeNPL -1.18e+01 5.03e+00 -2.35 0.01871 *
## country.codeNRU -1.16e+01 6.90e+00 -1.68 0.09363 .
## country.codePAK -5.89e+00 3.94e+00 -1.50 0.13490
## country.codePAN -6.63e+00 3.88e+00 -1.71 0.08777 .
## country.codePER -9.83e+00 3.92e+00 -2.51 0.01230 *
## country.codePHL -1.05e+01 4.38e+00 -2.41 0.01626 *
## country.codePNG 2.29e+01 5.46e+00 4.21 2.8e-05 ***
## country.codePOL -1.32e+01 4.35e+00 -3.04 0.00244 **
## country.codePRT -1.66e+01 4.30e+00 -3.86 0.00012 ***
## country.codePRY -1.69e+01 3.77e+00 -4.47 8.4e-06 ***
## country.codePSE -1.44e+01 4.10e+00 -3.52 0.00045 ***
## country.codeROU -7.85e+00 4.22e+00 -1.86 0.06300 .
## country.codeRUS -1.58e+01 3.97e+00 -3.97 7.6e-05 ***
## country.codeRWA 2.03e+01 4.44e+00 4.57 5.3e-06 ***
## country.codeSDN -2.42e+00 5.41e+00 -0.45 0.65496
## country.codeSEN 1.45e+01 4.24e+00 3.42 0.00066 ***
## country.codeSLB 4.26e+00 5.43e+00 0.78 0.43288
## country.codeSLE 6.81e+00 4.65e+00 1.46 0.14350
## country.codeSLV -1.32e+01 3.89e+00 -3.40 0.00070 ***
## country.codeSOM 4.13e+01 6.90e+00 5.98 2.9e-09 ***
## country.codeSRB -8.19e+00 4.39e+00 -1.87 0.06201 .
## country.codeSSD 2.95e+01 5.44e+00 5.42 7.2e-08 ***
## country.codeSTP -1.35e+00 4.88e+00 -0.28 0.78170
## country.codeSUR 4.59e+00 6.85e+00 0.67 0.50235
## country.codeSVK -1.20e+01 4.27e+00 -2.81 0.00506 **
## country.codeSVN -1.46e+01 4.35e+00 -3.36 0.00079 ***
## country.codeSWE -1.87e+01 4.37e+00 -4.29 1.9e-05 ***
## country.codeSWZ 4.12e+01 4.90e+00 8.42 < 2e-16 ***
## country.codeSYC -9.71e+00 5.55e+00 -1.75 0.08054 .
## country.codeSYR -1.14e+01 5.45e+00 -2.09 0.03702 *
## country.codeTCD 1.27e+01 4.87e+00 2.62 0.00898 **
## country.codeTGO 1.22e+01 4.90e+00 2.50 0.01270 *
## country.codeTHA -1.89e+01 3.85e+00 -4.92 9.9e-07 ***
## country.codeTJK -9.27e+00 4.43e+00 -2.09 0.03649 *
## country.codeTKM 1.11e+01 6.82e+00 1.62 0.10528
## country.codeTLS -1.23e+01 4.94e+00 -2.49 0.01286 *
## country.codeTON -1.48e+01 4.95e+00 -2.98 0.00289 **
## country.codeTTO -1.71e+01 5.62e+00 -3.05 0.00237 **
## country.codeTUN -1.22e+01 4.19e+00 -2.91 0.00372 **
## country.codeTUR -1.17e+01 3.81e+00 -3.08 0.00211 **
## country.codeTUV -9.63e+00 7.06e+00 -1.36 0.17268
## country.codeTZA 2.27e+01 4.61e+00 4.92 9.6e-07 ***
## country.codeUGA 7.75e+00 4.10e+00 1.89 0.05895 .
## country.codeUKR -1.90e+01 4.01e+00 -4.73 2.5e-06 ***

```

```

## country.codeURY -1.48e+01 4.14e+00 -3.57 0.00037 ***
## country.codeUSA -1.87e+01 4.46e+00 -4.19 3.0e-05 ***
## country.codeUZB 3.84e+01 4.64e+00 8.26 3.4e-16 ***
## country.codeVEN -1.05e+01 3.98e+00 -2.64 0.00841 **
## country.codeVNM -2.01e+01 4.02e+00 -5.01 6.1e-07 ***
## country.codeVUT -3.46e+00 5.37e+00 -0.64 0.51924
## country.codeWSM -1.26e+01 4.93e+00 -2.56 0.01050 *
## country.codeXKX -8.48e+00 4.20e+00 -2.02 0.04380 *
## country.codeYEM -1.51e+01 4.84e+00 -3.13 0.00181 **
## country.codeZAF 8.47e+00 4.97e+00 1.70 0.08877 .
## country.codeZMB 1.91e+01 4.07e+00 4.69 3.1e-06 ***
## country.codeZWE 1.30e+00 4.92e+00 0.26 0.79161
## year -2.35e-01 3.58e-02 -6.57 7.4e-11 ***
## incomeL 3.81e-01 1.66e+00 0.23 0.81848
## incomeLM -5.25e+00 1.22e+00 -4.32 1.7e-05 ***
## incomeUM -2.97e+00 9.38e-01 -3.16 0.00159 **
## edu.total -4.20e-01 1.48e-01 -2.83 0.00473 **
## hlth 2.48e-01 1.11e-01 2.22 0.02643 *
## mil -4.60e-01 2.11e-01 -2.18 0.02950 *
## fdi -8.27e-12 4.50e-12 -1.84 0.06627 .
## lbr.part 1.69e-01 2.88e-02 5.88 5.1e-09 ***
## unemp -9.86e-02 4.65e-02 -2.12 0.03422 *
## pop.gwth.total -3.12e+00 7.57e-01 -4.13 3.9e-05 ***
## pop.gwth.rural 8.28e-01 3.17e-01 2.61 0.00910 **
## pop.gwth.urban 2.37e+00 4.02e-01 5.89 4.8e-09 ***
## gdp.dflt -1.56e-04 1.04e-03 -0.15 0.88032
## gcf -2.37e-01 3.15e-02 -7.50 1.2e-13 ***
## trade -1.57e-02 9.03e-03 -1.74 0.08294 .
## gdp.pc 2.85e-04 2.76e-05 10.35 < 2e-16 ***
## lfdi -1.20e-04 5.61e-02 0.00 0.99829
## lgdp.dflt 1.91e-01 1.39e-01 1.37 0.17078
## lgdp.pc -6.19e+00 5.99e-01 -10.32 < 2e-16 ***
## ---
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ',' 1
##
## Residual standard error: 5.78 on 1320 degrees of freedom
## Multiple R-squared: 0.916, Adjusted R-squared: 0.904
## F-statistic: 76.9 on 187 and 1320 DF, p-value: <2e-16

```

There are also some weak variable, we should perform some variable selection.

B. VIF Our data set contains a lots of variables. We can perform some variable selection to reduce over-fitting.

```

gvif <- lapply(fit1, vif)
gvif

```

```

## [[1]]
##          GVIF Df GVIF^(1/(2*Df))
## country.code 2.63e+08 167      1.06
## year        5.47e+00  1       2.34
## income      1.16e+02  3       2.21
## edu.total   2.72e+00  1       1.65

```

```

## hlth          4.03e+00  1      2.01
## mil           3.95e+00  1      1.99
## fdi           2.34e+00  1      1.53
## lbr.part      3.29e+00  1      1.81
## unemp         3.50e+00  1      1.87
## pop.gwth.total 3.46e+01  1      5.88
## pop.gwth.rural 1.01e+01  1      3.18
## pop.gwth.urban 2.09e+01  1      4.57
## gdp.dflt     1.31e+00  1      1.15
## gcf            2.35e+00  1      1.53
## trade          9.59e+00  1      3.10
## gdp.pc         1.39e+01  1      3.73
## lfdi          1.96e+00  1      1.40
## lgdp.dflt    2.39e+00  1      1.55
## lgdp.pc        3.99e+01  1      6.32
##
## [[2]]
##                               GVIF  Df  GVIF^(1/(2*Df))
## country.code   4.42e+08 167      1.06
## year           6.02e+00  1      2.45
## income         1.13e+02  3      2.20
## edu.total     2.57e+00  1      1.60
## hlth           4.38e+00  1      2.09
## mil            3.81e+00  1      1.95
## fdi            2.34e+00  1      1.53
## lbr.part       3.53e+00  1      1.88
## unemp          3.63e+00  1      1.91
## pop.gwth.total 3.38e+01  1      5.82
## pop.gwth.rural 9.90e+00  1      3.15
## pop.gwth.urban 2.09e+01  1      4.57
## gdp.dflt      1.34e+00  1      1.16
## gcf            2.50e+00  1      1.58
## trade          1.09e+01  1      3.31
## gdp.pc         1.43e+01  1      3.78
## lfdi          2.01e+00  1      1.42
## lgdp.dflt    2.36e+00  1      1.54
## lgdp.pc        4.42e+01  1      6.65
##
## [[3]]
##                               GVIF  Df  GVIF^(1/(2*Df))
## country.code   5.47e+08 167      1.06
## year           5.88e+00  1      2.42
## income         1.11e+02  3      2.19
## edu.total     2.64e+00  1      1.63
## hlth           5.46e+00  1      2.34
## mil            3.85e+00  1      1.96
## fdi            2.33e+00  1      1.53
## lbr.part       3.54e+00  1      1.88
## unemp          3.71e+00  1      1.93
## pop.gwth.total 3.45e+01  1      5.87
## pop.gwth.rural 9.96e+00  1      3.16
## pop.gwth.urban 2.06e+01  1      4.54
## gdp.dflt      1.34e+00  1      1.16
## gcf            2.56e+00  1      1.60

```

```

## trade      1.16e+01   1      3.41
## gdp.pc    1.41e+01   1      3.76
## lfdi      2.07e+00   1      1.44
## lgdp.dflt 2.29e+00   1      1.51
## lgdp.pc   4.24e+01   1      6.51

```

Some forum suggested that we should use the standard $GVIF^{1/(2\cdot df)} < 2$ as equivalent to $GVIF < 4$ to account for high degree of freedom of some variables.

```

# adjusted gvif higher than 3
lapply(gvif, function(g) {
  g[g[, 3] > 3, 3]
})

## [[1]]
## pop.gwth.total pop.gwth.rural pop.gwth.urban      trade      gdp.pc
##      5.88          3.18          4.57      3.10      3.73
##      lgdp.pc
##          6.32
##
## [[2]]
## pop.gwth.total pop.gwth.rural pop.gwth.urban      trade      gdp.pc
##      5.82          3.15          4.57      3.31      3.78
##      lgdp.pc
##          6.65
##
## [[3]]
## pop.gwth.total pop.gwth.rural pop.gwth.urban      trade      gdp.pc
##      5.87          3.16          4.54      3.41      3.76
##      lgdp.pc
##          6.51

```

Remove the highest-ranked variables (pop.gwth.total, pop.gwth.urban, pop.gwth.rural, gdp.pc, lgdp.pc) and rebuild the models.

```

formula.str <- "pov ~ country.code+year+income+edu.total+hlth+mil+fdi+lbr.part+unemp+gdp.dflt+gcf+trade"

# Using single imputed data set
fit1 <- lapply(1:countries1.imputed$m, function(i) lm(as.formula(formula.str),
  data = complete(countries1.imputed, i)))

```

Re-run GVIF analysis

```

gvif <- lapply(fit1, vif)

lapply(gvif, function(g) {
  g[g[, 3] > 2, 3]
})

## [[1]]
## income  trade
##    2.05  2.98

```

```

## 
## [[2]]
## income   hlth   trade
##    2.05    2.07    3.18
##
## [[3]]
## income   hlth   trade
##    2.06    2.33    3.28

```

There're still some variable with adjusted GVIF > 2 . We are interested in the effects of `hlth` (expenditure in Healthcare), so we won't remove those variables. We can remove `trade`.

```

# remove trade
formula.str <- "pov ~ country.code+year+income+edu.total+hlth+mil+fdi+lbr.part+unemp+gdp.dflt+gcf+lfdi+lgdp.dfl

# Using single imputed data set
fit1 <- lapply(1:countries1.imputed$m, function(i) lm(as.formula(formula.str),
  data = complete(countries1.imputed, i)))

# gvif
gvif <- lapply(fit1, vif)

lapply(gvif, function(g) {
  g[g[, 3] > 2, 3]
})

## [[1]]
## [1] 2.05
##
## [[2]]
## income   hlth
##    2.04    2.07
##
## [[3]]
## income   hlth
##    2.05    2.32

```

The adjusted GVIF are acceptably low.

Coefficients overview.

```

# find coef of variable other than country.code
var <- c("year", "incomeL", "incomeLM", "incomeUM",
  "edu.total", "hlth", "mil", "fdi", "lbr.part",
  "unemp", "gdp.dflt", "gcf", "lfdi", "lgdp.dfl")

res <- do.call(rbind, lapply(fit1, function(f) f$coefficients[var]))
kable(res)

```

year	incomeL	incomeLM	incomeUM	edu.total	hlth	mil	fdi	lbr.part	unemp	gdp.dflt	gcf	lfdi	lgdp.dfl
-	8.88	-	-3.37	-0.624	0.337	-	0	0.212	-	0.001	-	-	0.188
0.350		1.298				0.526			0.121		0.287	0.110	

year	incomeL	incomeM	incomeU	edu.total	hlth	mil	fdi	lbr.part	unemp	gdp.dflt	gcf	ldfi	lgdp.dflt
-	10.61	-	-2.51	-0.443	0.481	-	0	0.335	-	0.003	-	-	0.149
0.355		0.365				0.251			0.018		0.287	0.115	
-	8.75	-	-3.33	-0.864	0.389	-	0	0.215	-	0.002	-	-	0.074
0.346		1.325				0.228			0.038		0.285	0.118	

There are several sign-switching in income coefficients, indicating the effect of removing some multicollinearity.

```
fit1.summ <- lapply(fit1, function(f) {
  k <- length(f$coefficients) - 1
  pred <- predict(f, countries.test.4)
  data.frame(`Train Adj.R2` = summary(f)$r.squared,
             `Test Adj.R2` = r2(pred, countries.test.4$pov))
})

res <- do.call(rbind.data.frame, fit1.summ)
res <- cbind(data.frame(Set = 1:3), res)
kable(res)
```

Set	Train.Adj.R2	Test.Adj.R2
1	0.902	0.717
2	0.907	0.712
3	0.902	0.719

Directly removing variables seems to be penalising our test results. As our interest is to investigate the effect of certain variables on poverty, the yielded R-Squared is within acceptable range. We can continue variable selection on the new models.

```
formula.str <- "pov ~ country.code+year+income+edu.total+hlth+mil+fdi+lbr.part+unemp+gdp.dflt+gcf+ldfi+
```

B. Step-wise AIC We can conduct a step-wise AIC variable selection. It is similar to the procedure we use in class, but based on a metric call AIC (Akaike Information Criterion), which is an estimator of prediction error and relative quality of statistical models. The lower AIC is, the better the model fits. \

```
# helper
fitAIC <- function(i) {
  set.seed(1984)
  fit <- lm(as.formula(formula.str), complete(countries1.imputed,
                                                 i))
  aic.fit <- stepAIC(fit, trace = F, direction = "backward")
  return(aic.fit)
}

# build models
aic.fits <- lapply(1:countries1.imputed$m, fitAIC)

# check final terms of each model
lapply(aic.fits, function(mod) formula(mod$terms))
```

```

## [[1]]
## pov ~ country.code + year + income + edu.total + hlth + mil +
##       lbr.part + unemp + gcf + lfdi + lgdp.dflt
## <environment: 0x559d17afe228>
##
## [[2]]
## pov ~ country.code + year + income + edu.total + hlth + lbr.part +
##       gdp.dflt + gcf + lfdi
## <environment: 0x559d21c1f088>
##
## [[3]]
## pov ~ country.code + year + income + edu.total + hlth + lbr.part +
##       gdp.dflt + gcf + lfdi
## <environment: 0x559d14144828>

```

The most commonly removed variables are:

- gdp.dflt
- ldgp.dflt
- unemp
- fdi

```

# predict with each model
aic.pred <- lapply(aic.fits, predict, newdata = countries.test.4)
# calculate adj r2
aic.train.r2 <- unlist(lapply(aic.fits, function(model) summary(model)$adj.r.squared))

k <- lapply(aic.fits, function(m) length(m$coefficients))
aic.test.r2 <- unlist(mapply(r2, aic.pred, MoreArgs = list(orig = countries.test.4$pov)))

res <- data.frame(`set no.` = 1:3, train = aic.train.r2,
                   test = aic.test.r2)

```

We can compare R-Squared to estimate relative over-fitting in the new models.

```
kable(res)
```

set.no.	train	test
1	0.889	0.719
2	0.895	0.709
3	0.888	0.717

Performance on train set and test set are slightly improved. However, we have simplified the model by removing some unnecessary terms. We can update our models.

```

formula.str <- "pov ~ country.code+year+income+edu.total+hlth+mil+lbr.part+gcf+lfdi"

fit3 <- lapply(1:countries1.imputed$m, function(i) lm(as.formula(formula.str),
                                                 data = complete(countries1.imputed, i)))

# Build models with all generated data set and

```

```

# pool the estimates
fit4 <- with(data = countries1.imputed, exp = lm(as.formula(formula.str)))
fit4.combined <- pool(fit4)
# dummy lm model
fit4.dummy <- lm(as.formula(formula.str), data = complete(countries1.imputed,
  1))
# replace coefficients of dummy model & predict
fit4.dummy$coefficients <- fit4.combined$pooled$estimate

fit3.summ <- lapply(fit3, function(f) {
  k <- length(f$coefficients) - 1
  pred <- predict(f, countries.test.4)
  test.resid <- pred - countries.test.4$pov
  data.frame(`Train MSE` = mean(summary(f)$residuals^2),
    `Test MSE` = mean(test.resid^2), `Train MAE` = mean(abs(summary(f)$residuals)),
    `Test MAE` = mean(abs(test.resid)), `Train R2` = summary(f)$r.squared,
    `Test R2` = r2(pred, countries.test.4$pov))
})

k <- length(fit4.dummy$coefficients) - 1
fit4.pred <- predict(fit4.dummy, countries.test.4)
test.resid <- fit4.pred - countries.test.4$pov
fit4.summ <- data.frame(`Train MSE` = mean(summary(fit4.dummy)$residuals^2),
  `Test MSE` = mean(test.resid^2), `Train MAE` = mean(abs(summary(fit4.dummy)$residuals)),
  `Test MAE` = mean(abs(test.resid)), `Train R2` = pool.r.squared(fit4)[,
    "est"], `Test R2` = r2(fit4.pred, countries.test.4$pov))

res <- do.call(rbind.data.frame, fit3.summ)
res <- rbind(res, fit4.summ)
res <- cbind(data.frame(Set = c(1:3, "Pooled")), res)

kable(res)

```

Set	Train.MSE	Test.MSE	Train.MAE	Test.MAE	Train.R2	Test.R2
1	34.2	23.9	3.73	2.80	0.902	0.719
2	32.5	24.5	3.69	2.85	0.907	0.711
3	34.4	23.9	3.76	2.79	0.901	0.719
Pooled	34.2	24.0	3.73	2.80	0.903	0.718

Coefficients overview.

```

var <- c("year", "incomeL", "incomeLM", "incomeUM",
  "edu.total", "hlth", "mil", "lbr.part", "gcf",
  "lfdi")

res <- do.call(rbind, lapply(fit3, function(f) f$coefficients[var]))
kable(res)

```

year	incomeL	incomeLM	incomeUM	edu.total	hlth	mil	lbr.part	gcf	lfdi
-0.359	8.70	-1.466	-3.47	-0.651	0.252	-0.541	0.223	-0.268	-0.131

year	incomeL	incomeLM	incomeUM	edu.total	hlth	mil	lbr.part	gcf	ldfi
-0.368	10.41	-0.534	-2.59	-0.450	0.452	-0.281	0.333	-0.274	-0.132
-0.353	8.56	-1.508	-3.45	-0.887	0.352	-0.242	0.217	-0.272	-0.130

We can further treat over-fitting with regularization.

C. Regularization Helper functions to standardise data.

```

sd0 <- function(vct) {
  if (!is.numeric(vct)) {
    return(NA)
  }

  return(sd(vct, na.rm = T))
}

std0 <- function(vct, scl) {
  if (!is.numeric(vct)) {
    return(vct)
  }

  return(vct/scl)
}

standardise <- function(train, test) {
  # save number of train and combine both data
  # set
  trainCases <- 1:nrow(train)
  data <- rbind(train, test)
  # calc scaler
  scaler <- unlist(lapply(data, sd0))
  # get the numeric columns
  numCols <- which(unlist(lapply(data, is.numeric)))
  # divide numeric columns by scalers, and
  # combine with factor columns
  num <- as.data.frame(mapply(std0, vct = data[, numCols], scl = scaler[numCols], SIMPLIFY = T))
  fct <- data[, -numCols]
  final <- cbind(fct, num)
  # split to train and test
  trainSet <- final[trainCases, ]
  testSet <- final[-trainCases, ]

  res <- list(final = final, train = trainSet, test = testSet,
             scaler = scaler)
  return(res)
}

```

Helper function to build the optimal model.

```

optimalModel <- function(formula, train, test) {
  set.seed(1984)
  # standise and save number of train cases (to
  # split later)
  std.data <- standardise(train, test)
  trainCases <- 1:nrow(train)
  # matrix-ise
  xs <- model.matrix(formula, std.data$final)[, -1]
  ys <- std.data$final$pov
  # split data
  train.x <- xs[trainCases, ]
  train.y <- ys[trainCases]
  test.x <- xs[-trainCases, ]
  test.y <- ys[-trainCases]
  # list of tested alpha, resolution = 0.1
  alphas <- seq(0, 1, 0.1)
  # build a bunch of models
  models <- lapply(alphas, function(a) cv.glmnet(train.x,
    train.y, type.measure = "mse", alpha = a))
  cv.error <- unlist(lapply(models, function(model) model$cvm[model$lambda ==
    model$lambda.min]))
  # best model
  best.model.idx <- which.min(cv.error)
  # optimal alpha and lambda
  alpha.opt <- alphas[best.model.idx]
  lambda.opt <- models[[best.model.idx]]$lambda.min
  best.model <- glmnet(train.x, train.y, alpha = alpha.opt,
    lambda = lambda.opt)
  # predict for test data
  train.fit <- predict(best.model, train.x)
  test.fit <- predict(best.model, test.x)
  # evaluation - train
  k <- best.model$df
  train.r2 <- best.model$dev.ratio
  train.adjR2 <- adjR2(train.fit, train.y, k)
  train.resid <- train.fit - train.y
  train.mse <- mean(train.resid^2)
  train.rmse <- sqrt(train.mse)
  train.mae <- mean(abs(train.resid))
  train.mape <- mean(abs((train.resid/train.y)))
  # evaluation - test
  test.r2 <- r2(test.fit, test.y)
  test.adjR2 <- adjR2(test.fit, test.y, k)
  test.resid <- test.fit - test.y
  test.mse <- mean(test.resid^2)
  test.rmse <- sqrt(test.mse)
  test.mae <- mean(abs(test.resid))
  test.mape <- mean(abs(test.resid/test.y))
  # final results
  results <- list(train = list(data = train, r2 = train.r2,
    adj.r2 = train.adjR2, fitted = train.fit, residuals = train.resid,
    mse = train.mse, rmse = train.rmse, mae = train.mae,
    mape = train.mape), test = list(data = test,

```

```

        r2 = test.r2, adj.r2 = test.adjR2, fitted = test.fit,
        residuals = test.resid, mse = test.mse, rmse = test.rmse,
        mae = test.mae, mape = test.mape), scaler = std.data$scaler,
        alpha = alpha.opt, lambda = lambda.opt, model = best.model)

    return(results)
}

```

Build the optimal models using each imputed data set.

```

# helper function to build from a specific
# imputed data
buildWith <- function(set, imputed, test, formula,
  fun) {
  train <- complete(imputed, set)
  model <- fun(formula, train, test)
  return(model)
}

set.seed(1984)

models <- lapply(1:countries1.imputed$m, buildWith,
  imputed = countries1.imputed, test = countries.test.4,
  formula = as.formula(formula.str), fun = optimalModel)

```

Display parameters and evaluation.

```

result.list <- lapply(models, function(mod) {
  data.frame(alpha = mod$alpha, lambda = mod$lambda,
    `Train MSE` = mod$train$mse, `Test MSE` = mod$test$mse,
    `Train MAE` = mod$train$mae, `Test MAE` = mod$test$mae,
    `Train R2` = mod$train$r2, `Test R2` = mod$test$r2)
})

result.df <- do.call(rbind.data.frame, result.list)
result.df <- cbind(data.frame(Set = 1:countries1.imputed$m),
  result.df)
kable(result.df)

```

Set	alpha	lambda	Train.MSE	Test.MSE	Train.MAE	Test.MAE	Train.R2	Test.R2
1	0.0	0.079	0.116	0.073	0.211	0.147	0.897	0.735
2	0.0	0.080	0.111	0.075	0.209	0.149	0.902	0.729
3	0.7	0.000	0.112	0.078	0.214	0.159	0.901	0.718

There are some improvement on the test data performance based on adjusted R-Squared. MSE and MAE remain low on both data sets.

Coefficients overview.

```

var <- c("year", "incomeL", "incomeLM", "incomeUM",
       "edu.total", "hlth", "mil", "lbr.part", "gcf",
       "lfdi")

res <- do.call(rbind, lapply(models, function(f) coefficients(f$model)[var,
      ]))
res <- cbind(data.frame(Set = 1:3), res)
kable(res)

```

Set	year	incomeL	incomeLM	incomeUM	edu.total	hlth	mil	lbr.part	gcf	lfdi
1	-0.142	0.763	0.076	-0.118	-0.071	-0.013	-0.030	0.123	-0.091	-0.040
2	-0.146	0.812	0.080	-0.091	-0.058	0.001	-0.017	0.155	-0.092	-0.043
3	-0.179	0.524	-0.061	-0.186	-0.083	0.041	-0.012	0.110	-0.109	-0.028

All non-country.code variables are non-zero. We can observe that the regularisation models just remove the effect of `country.code`, for countries that have similar “baseline”.

3.4.2. Assessment

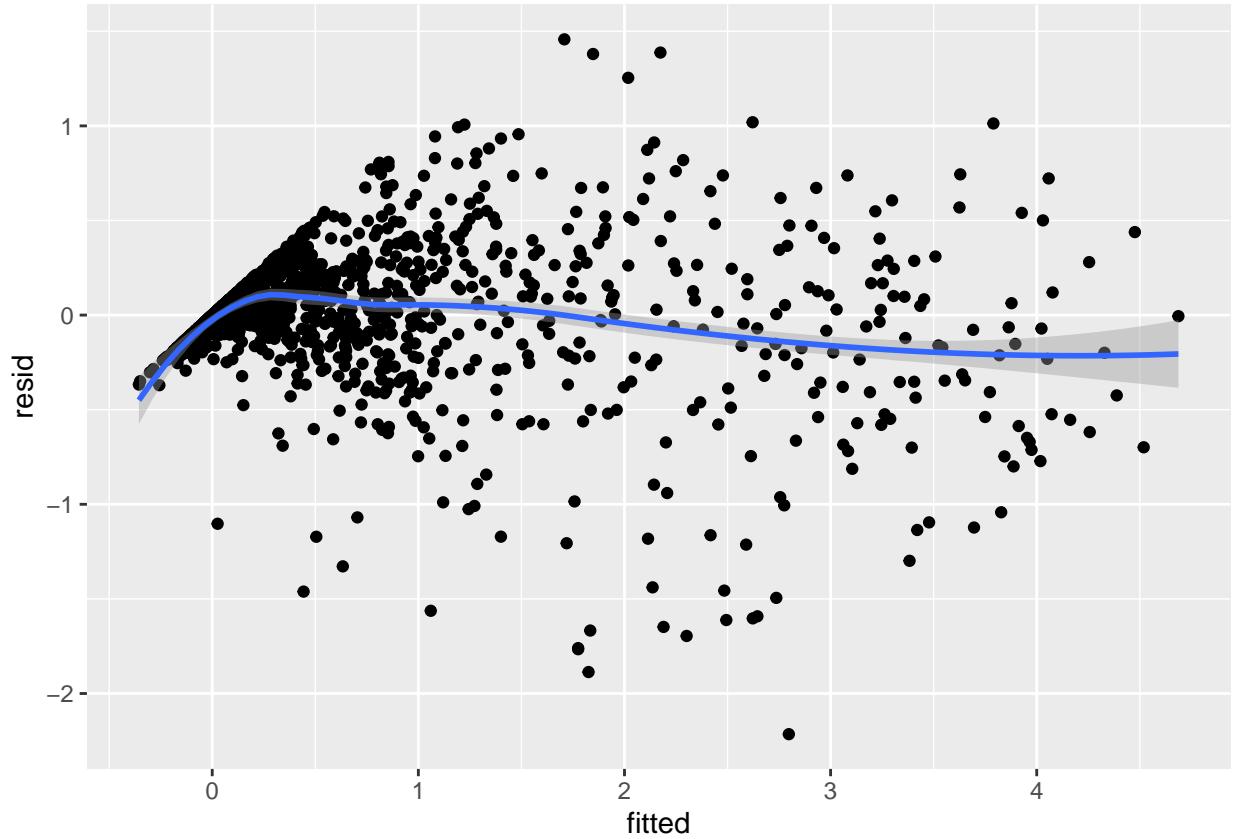
Now we are evaluating some performance metrics, checking assumptions, and remedy some potential problems.

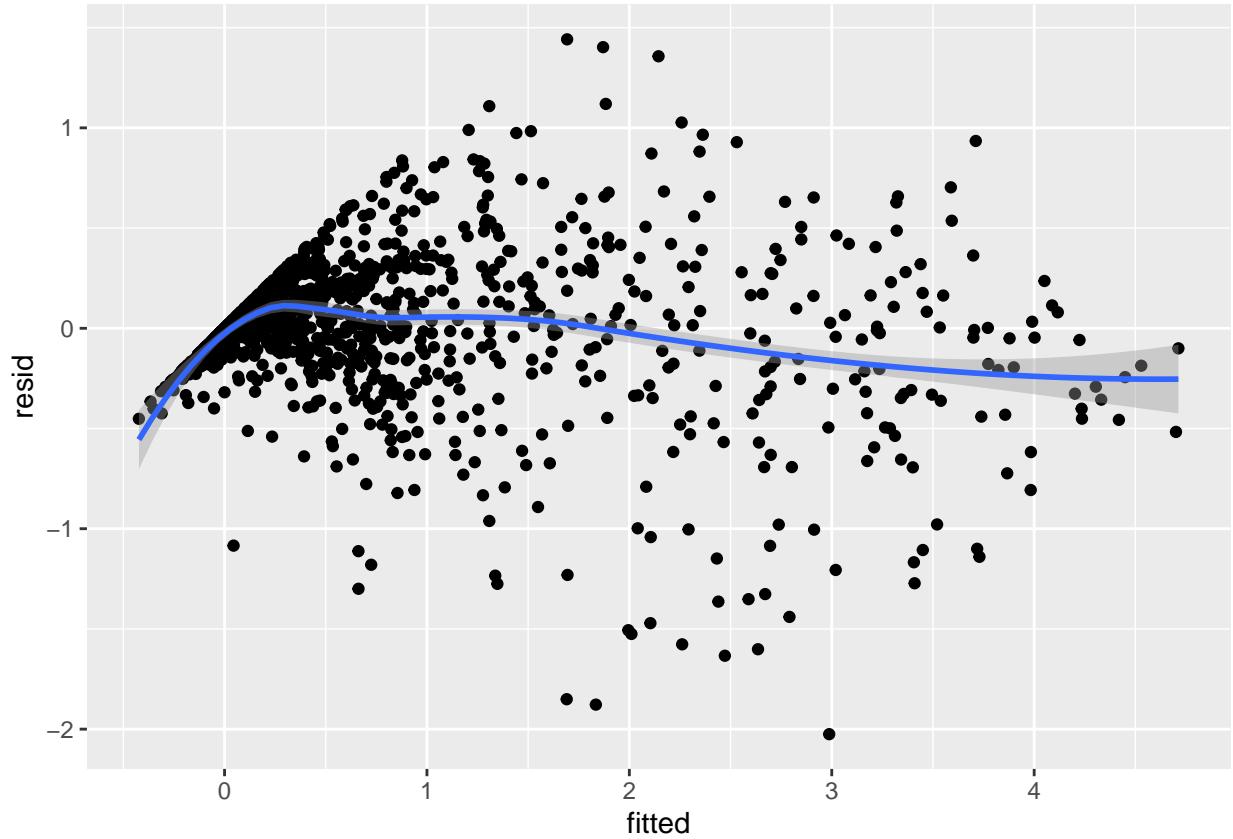
A. Homoscedasticity Plot the residuals ~ pov

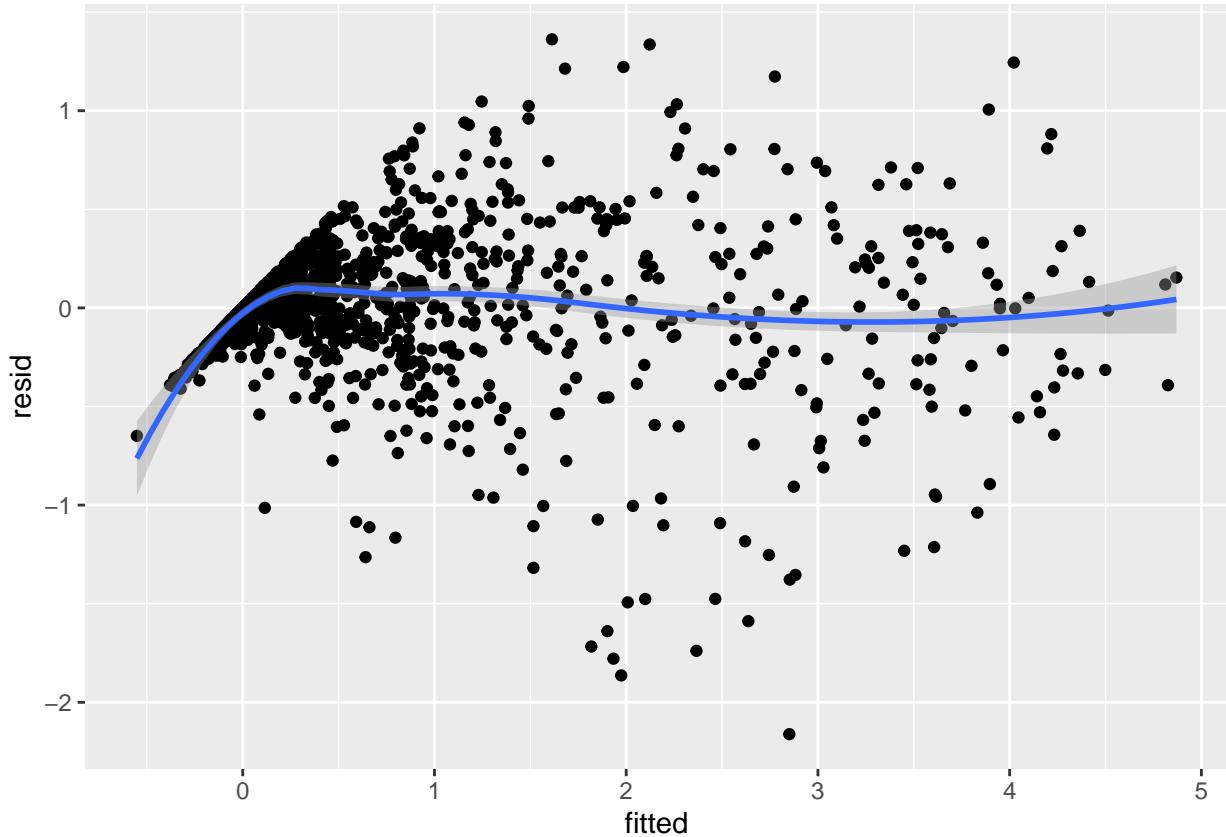
```

for (mod in models) {
  resid <- mod$train$residuals
  fitted <- mod$train$fitted
  # plot(resid ~ pov, main = paste('Model', i))
  print(ggplot(mapping = aes(y = resid, x = fitted)) +
    geom_point() + geom_smooth(formula = y ~ x,
      method = loess))
}

```







There seems to be some larger variance with higher values of `pov`. We can conduct a statistical test to confirm the present of heteroscedasticity.

The Goldfeld-Quandt test is performed by eliminating a certain number of observations from the dataset's center, then comparing the spread of residuals between the two datasets on either side of the central observations.

The Goldfeld-Quandt test examines two submodels' variances divided by a defined breakpoint and rejects if the variances disagree.

Under H0, the Goldfeld-Quandt test's test statistic follows an F distribution with degrees of freedom as specified in the parameter.

- **Null (H0):** Heteroscedasticity is not present.
- **Alternative (H1):** Heteroscedasticity is present.

```
lapply(models, function(m) {
  resid <- m$train$residuals
  # apply Goldfeld-Quandt test
  gqttest(formula = resid ~ 1, order.by = m$train$fitted)
})
```

```
## [[1]]
##
##  Goldfeld-Quandt test
##
## data:  resid ~ 1
```

```

## GQ = 17, df1 = 753, df2 = 753, p-value <2e-16
## alternative hypothesis: variance increases from segment 1 to 2
##
## 
## [[2]]
##
## Goldfeld-Quandt test
##
## data: resid ~ 1
## GQ = 14, df1 = 753, df2 = 753, p-value <2e-16
## alternative hypothesis: variance increases from segment 1 to 2
##
## 
## [[3]]
##
## Goldfeld-Quandt test
##
## data: resid ~ 1
## GQ = 13, df1 = 753, df2 = 753, p-value <2e-16
## alternative hypothesis: variance increases from segment 1 to 2

```

There is heteroscedasticity in our models. [Generalized Least Squares With Unknown For of Variance](#) is a possible remedy of this problem, but at the cost of model interpretation. The high variance in higher end of pov can be explained by the relative volatile economical and political climate in highly poor countries (one standard deviation from mean), leading to unstable effect of predictors.

B. Independence Some possible source of dependency are poverty measured on the same country, or in the same year. These are accounted in our models by controlling those variables.

Some countries have been engaging in international wars, being under the influences of foreign forces such as Iraq, and Afghanistan. They possess similar (and different) political climates. Some other groups of economical alliance, or enjoying similar natural resources: OPEC, EU, ASEAN, etc. might have similar characteristics that make them dependent.

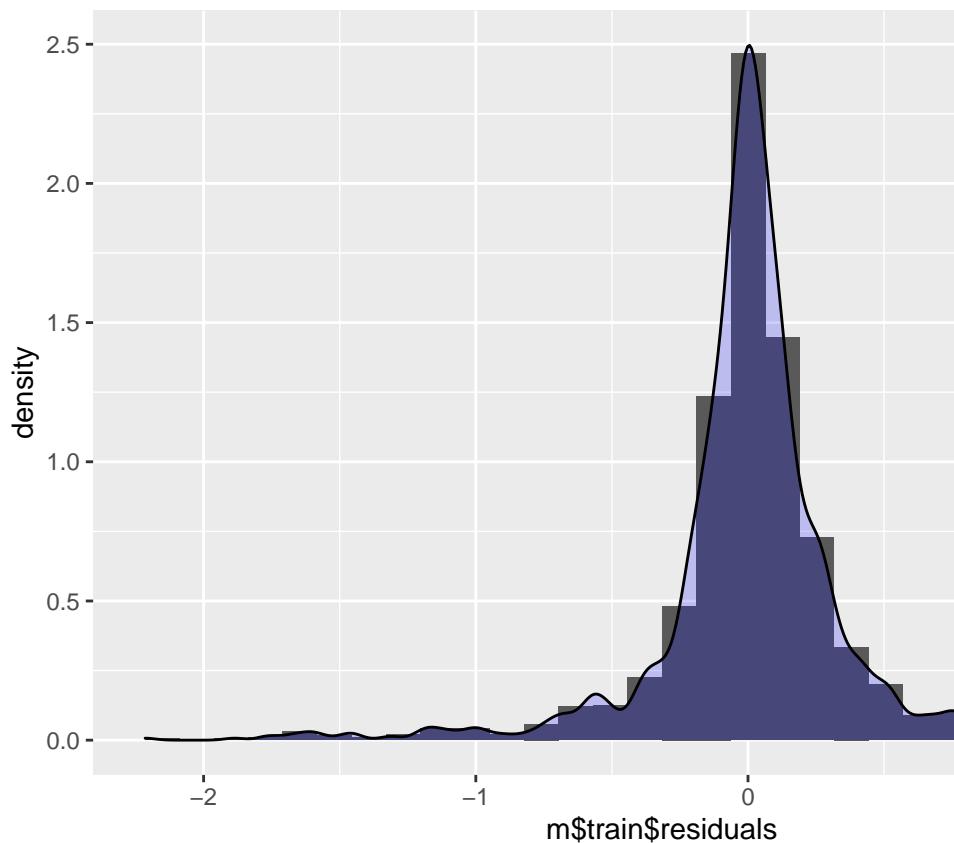
A solution to resolve this is to conduct the study on countries with similar geographical characteristics. In our study, however, we are interested in the general effects of certain variable across the globe. Although there are some dependence, the large number of entities with various features, we hypothesize, will cancel each other out, resulting in a net effect that gives an overview of the influences of variables.

```

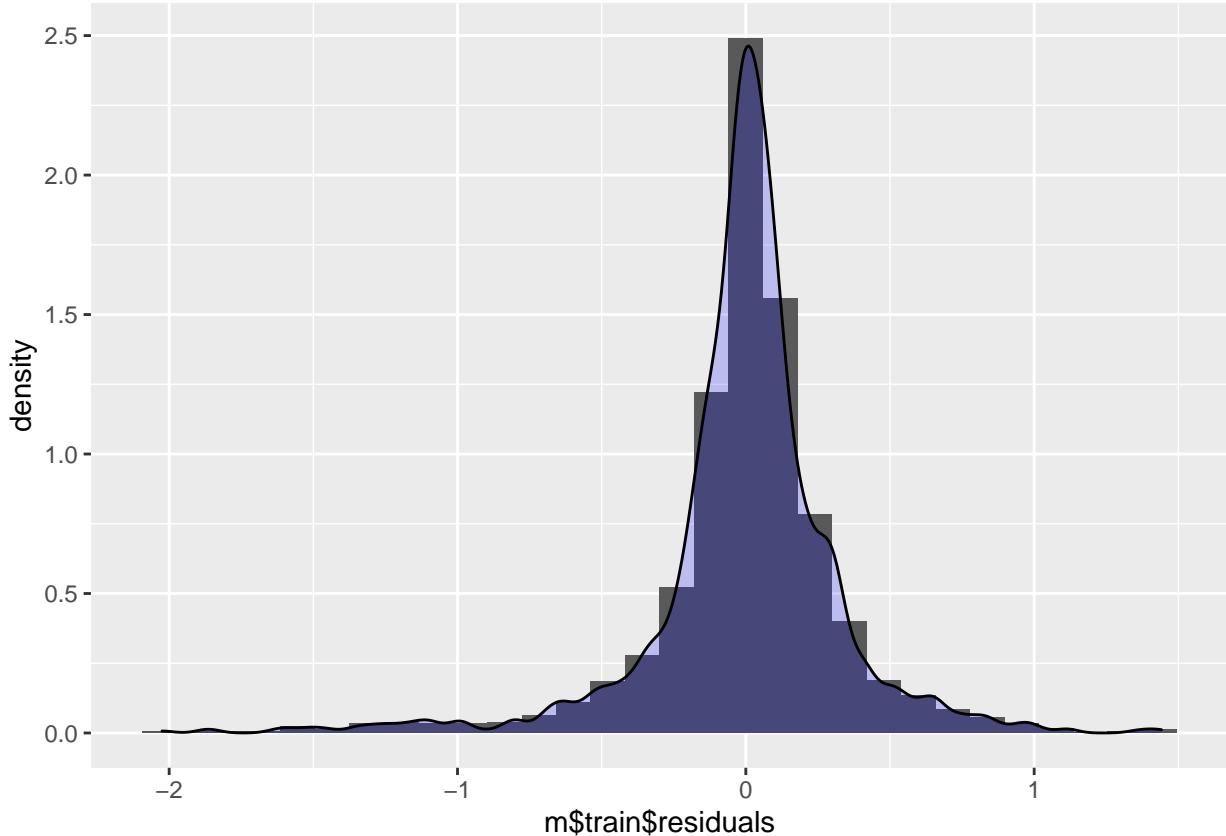
for (m in models) {
  p <- ggplot(mapping = aes(x = m$train$residuals)) +
    geom_histogram(aes(y = ..density..), bins = 30) +
    geom_density(alpha = 0.2, fill = "blue")

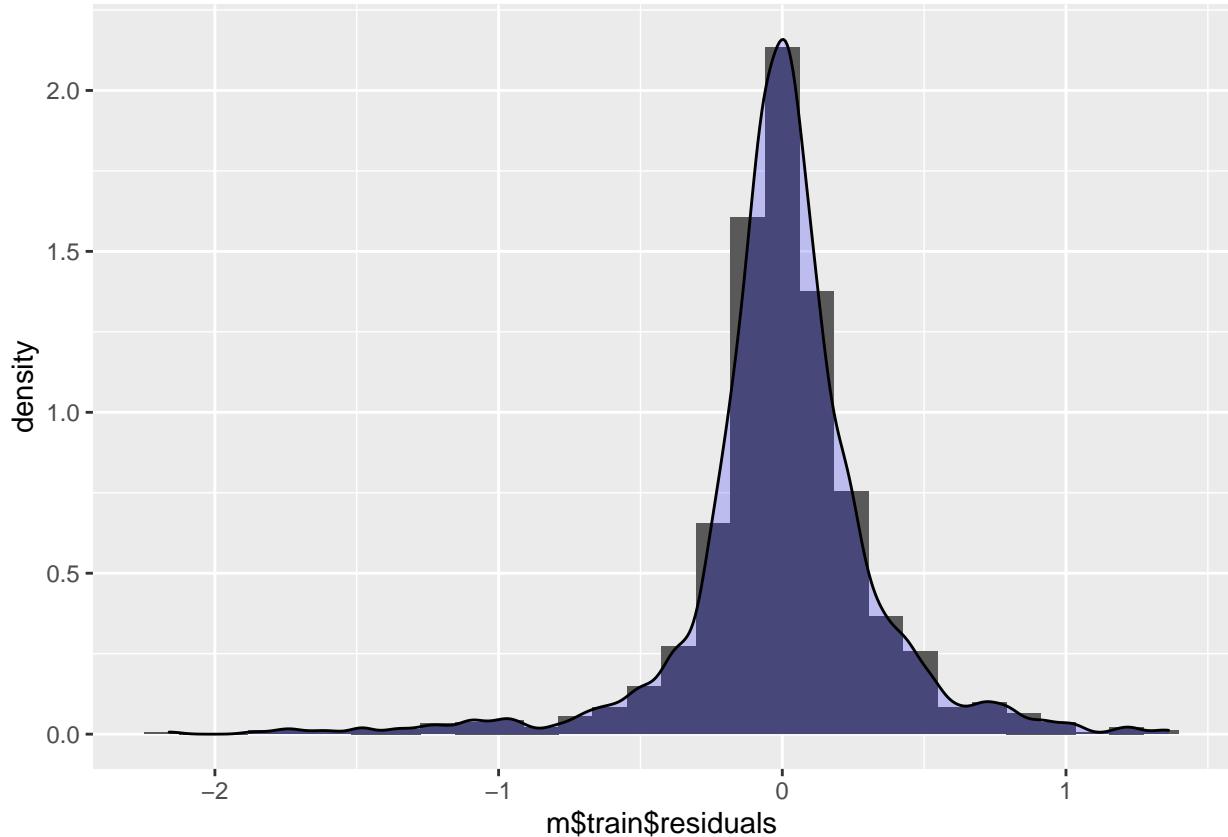
  print(p)
}

```



C. Normally Distributed Residuals

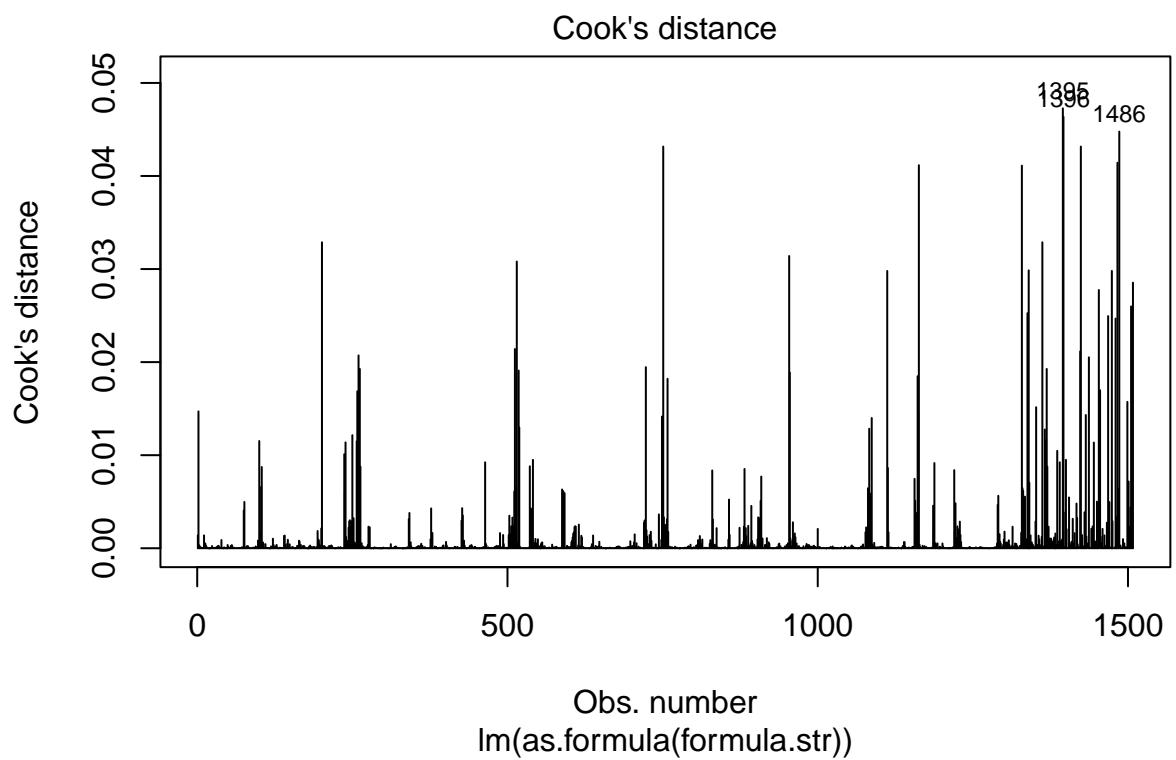


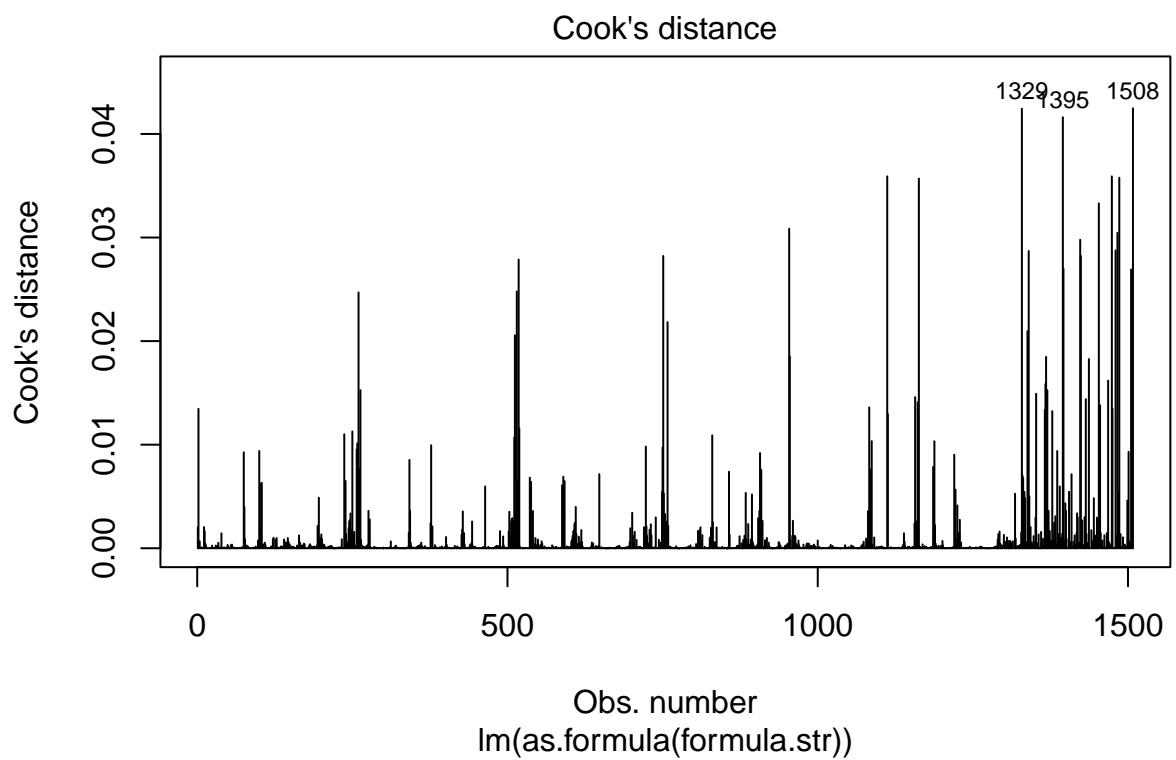


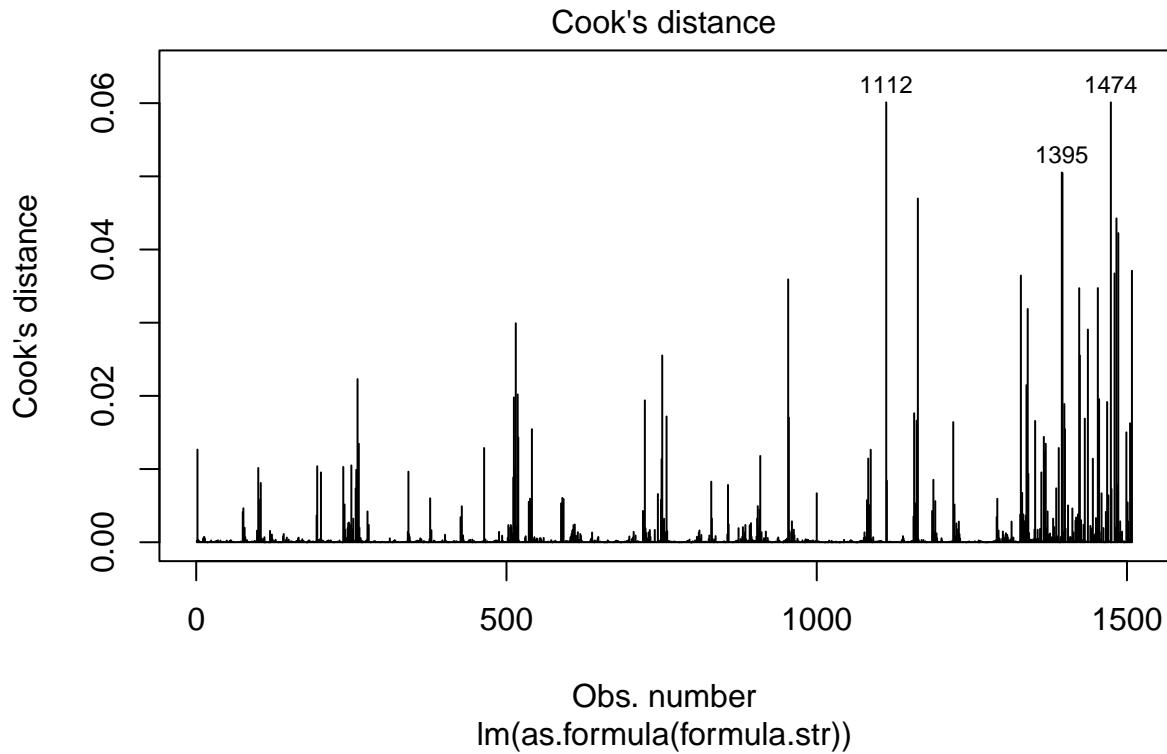
Although quite negatively skewed, the residuals are visibly normal.

D. Outliers As elastic net models produced in the last section are not very fit for detecting outliers using R-support function. As they just removed dummy variables in `country.code`, we can use OLS to detect some potential outliers. They are the same models we produced after the step-AIC selection (`fit3` and `fit4`)

```
for (mod in fit3) {
  plot(mod, which = 4)
}
```







Some obvious outliers are point 1112, 1329, 1395, 1396, 1424, 147, 1486, 1508.

```

out <- c(1112, 1329, 1395, 1396, 1424, 147, 1486, 1508)
var <- c("country.code", "country.name", "year", "pov",
       "income", "edu.total", "hlth", "mil", "lbr.part",
       "gcf", "lfdi")
# See the original data instead of imputed data
countries.train.4[out, var]

## # A tibble: 8 x 11
## # Groups:   country.code [8]
##   country.~1 count~2 year   pov income edu.t~3 hlth     mil lbr.p~4   gcf   lfdi
##   <fct>      <chr>  <dbl> <dbl> <fct>    <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>
## 1 SSD        South ~ 2016  67.3 L       1.54 NA    4.60   NA  NA   NA
## 2 YEM        Yemen,~ 2014  19.8 LM      NA    4.84  3.97  36.2 NA   NA
## 3 GIN        Guinea  1991  87.9 L       2.02 NA    2.38   NA  18.2 17.5
## 4 GMB        Gambia~ 1998  74.1 L       NA    NA    0.385  NA  4.88 17.0
## 5 LCA        St. Lu~ 1995  38.1 UM      NA    NA    NA    67.8 NA   17.3
## 6 BOL        Bolivia 2001  19   LM      NA    4.83  2.26  67.8 14.3 20.4
## 7 TJK        Tajiki~ 1999  62   L        2.08 NA    1.39   NA  17.3 15.7
## 8 ZWE        Zimbab~ 2011  21.6 L       NA    8.08  1.41  82.4 17.4 19.7
## # ... with abbreviated variable names 1: country.code, 2: country.name,
## #   3: edu.total, 4: lbr.part

```

These countries all appear very infrequently in our data set (only twice for South Sudan and St.Lucia). The gap between their appearance are quite large, too. Such reasons could have made the model struggle to

predict their poverty states.

Yemen's data were collected in 1998, 2005, and 2014, a huge gap in time, by when it has doubled its poverty rate.

Guinea is an interesting point. Because the outlier data (87.9% in 1991) is quite different to the next data point (40.8% in 1994), a huge jump in just 3 years. This might be a consequence of the political changes (new constitution, establishment of the Supreme Court, etc.) in this period. They are the changes not captured by the variables.

Gambia has the issue with Guinea, when it saw a huge reduction of poverty of 25% from 1998-2003.

Lucia only has two data points with large gap in time (1995-2016). It can be difficult to know much about someone if you only met them twice, with 21 years in between.

These data seems valid and there is no convincing reasons to remove them from the model (in the case of Guinea and Gambia, maybe there was some false reports because the jump were quite suspicious, but we need more evidence to conclude that they are false data).

3.4.3. Interpretation

Final model coefficients (exclude country.code):

```
var <- c("year", "incomeL", "incomeLM", "incomeUM",
       "edu.total", "hlth", "mil", "lbr.part", "gcf",
       "lfdi")

res <- do.call(rbind, lapply(models, function(f) coefficients(f$model)[var,
      ]))
res <- cbind(data.frame(Set = 1:3), res)
kable(res)
```

Set	year	incomeL	incomeLM	incomeUM	edu.total	hlth	mil	lbr.part	gcf	lfdi
1	-0.142	0.763	0.076	-0.118	-0.071	-0.013	-0.030	0.123	-0.091	-0.040
2	-0.146	0.812	0.080	-0.091	-0.058	0.001	-0.017	0.155	-0.092	-0.043
3	-0.179	0.524	-0.061	-0.186	-0.083	0.041	-0.012	0.110	-0.109	-0.028

We should convert them to un-standardised value for easier interpretation.

```
var <- c("year", "incomeL", "incomeLM", "incomeUM",
       "edu.total", "hlth", "mil", "lbr.part", "gcf",
       "lfdi")

res <- do.call(rbind, lapply(models, function(f) {
  # get scaler saved in the model
  scaler <- f$scaler
  # categorical has scaler = 1
  scaler[c("incomeL", "incomeLM", "incomeUM")] <- 1
  # beta = std.beta * sigma_y / sigma_x
  coefficients(f$model)[var, ] * scaler["pov"]/scaler[var]
}))
```

kable(res)

year	incomeL	incomeLM	incomeUM	edu.total	hlth	mil	lbr.part	gcf	ldfi
-0.268	13.38	1.32	-2.06	-0.775	-0.087	-0.394	0.240	-0.223	-0.197
-0.277	14.24	1.41	-1.60	-0.632	0.007	-0.217	0.306	-0.230	-0.214
-0.339	9.19	-1.07	-3.27	-0.919	0.289	-0.160	0.215	-0.269	-0.139

4. Conclusion

Approach 1, 2 and 4 seem to have the best performance.

First Approach:

```
knitr::kable(summary_3models[1:3, -5], digits = 3)
```

	MSE	MAE	RMSE	Rsquare
Baseline MLR_Train	3.336	1.137	1.826	0.965
Ridge Model_Train	0.221	0.356	0.470	0.862
LASSO Model_Train	0.216	0.352	0.465	0.865

```
knitr::kable(summary_3models[4:6, -5], digits = 3)
```

	MSE	MAE	RMSE	Rsquare
Baseline MLR_Test	3.598	1.293	1.897	0.881
Ridge Model_Test	0.171	0.318	0.414	0.863
LASSO Model_Test	0.159	0.311	0.398	0.873

Second Approach:

```
knitr::kable(summary_lasso, digits = 3)
```

	MSE	MAE	RMSE	MAPE	R2
LASSO_train	4.94	1.48	2.22	Inf	0.948
LASSO_test	4.67	1.50	2.16	Inf	0.846

```
knitr::kable(summary_Ridge, digits = 3)
```

	MSE	MAE	RMSE	MAPE	R2
Ridge_train	5.37	1.47	2.32	Inf	0.943
Ridge_test	4.42	1.46	2.10	Inf	0.854

Fourth Approach:

```
kable(result.df[c("alpha", "lambda", "Train.R2", "Test.R2")])
```

	alpha	lambda	Train.R2	Test.R2
	0.0	0.079	0.897	0.735
	0.0	0.080	0.902	0.729
	0.7	0.000	0.901	0.718

We can deduce our final interpretation from these models.

```
var <- c("incomeL", "incomeLM", "incomeUM", "edu.total",
       "hlth", "mil")
result1 <- coefficients(model3)[var]
# result1[names(scaler2)] <-
# result1[names(scaler2)] /
# scaler2[names(scaler2)] result1 <- result1 *
# scaler2['pov'] result1 <- result1[var]

result2 <- coefficients(glm_Ridge)[, "s0"]
result2[names(scaler2)] <- result2[names(scaler2)]/scaler2[names(scaler2)]
result2 <- result2 * scaler2["pov"]
result2 <- result2[var]

result4 <- res[2, var]

# combine
finalResults <- data.frame(Approach = c(1, 2, 4))
finalResults <- finalResults %>%
  cbind(rbind(result1, result2, result4))

kable(finalResults)
```

	Approach	incomeL	incomeLM	incomeUM	edu.total	hlth	mil
result1	1	0.075	-1.60	-1.45	-0.372	-0.248	0.520
result2	2	65.178	6.84	-8.22	-2.980	-0.591	-0.828
result4	4	14.241	1.41	-1.60	-0.632	0.007	-0.217

Some important interpretation:

- Low income countries generally have higher poverty index than High income countries.
- As suggested by 1st approach, Lower-Middle income countries have higher poverty index than High income countries. While 2nd and 4th approach suggest otherwise.
- Upper-Middle income countries in general have lower poverty index than High income countries.

As suggested by these models. Upper-Middle income countries have lowest poverty rate. This can be attributed to that these countries tend to have more even wealth distribution than high income countries, and have enough money to alleviate poor people.

- Spending on education is negatively correlated with poverty. On average, 1% (of GDP) more spending in education results in 3 point decrease in poverty index. This suggests that public spending on education can reduce poverty.

- Spending on Healthcare is negatively correlated with poverty. On average, 1% (of GDP) more spending in education results in 0.5 point decrease in poverty index.
- Spending on military has ambiguous effect on poverty. Different model results in different interpretation. This can be accounted to the difference in the data used in each approach. Approach 1 only use complete cases, which caused the model to predict positive relationship. However, as more data is utilised in approach 2&4 use more data, the effect tend to be neutral/negative.

5. References

- Akbar, Muhammad, Mukaram Khan, Haidar Farooqe, and Kaleemullah. 2019. “Public Spending, Education and Poverty: A Cross Country Analysis” 4 (April): 12–20.
- Alruhaymi, Abdullah Z, and Charles J Kim. 2021. “Why Can Multiple Imputations and How (MICE) Algorithm Work?” *Open J. Stat.* 11 (05): 759–77.
- Arel-Bundock, Vincent, and Krzysztof J. Pelc. 2018. “When Can Multiple Imputation Improve Regression Estimates?” *Political Analysis* 26 (2): 240–45. <https://doi.org/10.1017/pan.2017.43>.
- Gopalan, Bhuvaneswari. 2020. “Mice Algorithm to Impute Missing Values in a Dataset.” *Numpy Ninja*. Numpy Ninja. <https://www.numpyninja.com/post/mice-algorithm-to-impute-missing-values-in-a-dataset>.