Final Project Code and Data

- Due Tuesday 2016/5/24 (last day of exam period) at 9am.
- You must fork this repository as described in the HW Submission Exercise.
- Ensure all data files and CSV files are in the folder `data`. So for example, say you want to load the the file `example.csv` in the `data` folder. You need to run `read.csv("data/example.csv", header=TRUE)` and not `read.csv("example.csv", header=TRUE)`.
- Fill in the project details below.
- Delete everything in this file before the section "Project Details".
- Push/commit **all files** necessary for me to reproduce your final report.

Repository README Cover Page

Any good repository on GitHub will have an informative cover page. It is an advertisement of your work and should give basic instructions on how to use it/replicate your work. The contents of the file `README.md` end up being what's on the repository cover page. For example:

- The repo page for `dplyr`.
- The contents of its `README.md` file (click on RAW).

The way you will create your `README.md` is via `README.Rmd`:

- **DO NOT EDIT** `README.md`
- Rather, edit `README.Rmd` and then Knit it. An updated `README.md` file will be outputted.
- When you push/commit `README.md`, it will show up as your repo's cover page.

Project Details

- Name: Paul Suberu
- Project title: Health in sub-Saharan Africa
- Abstract: The goal of my project was to see what the effect of spending on health had on certain health measurements in sub-Saharan African countries.

```
## Warning: package 'ggplot2' was built under R version 3.2.4
```

Intorduction

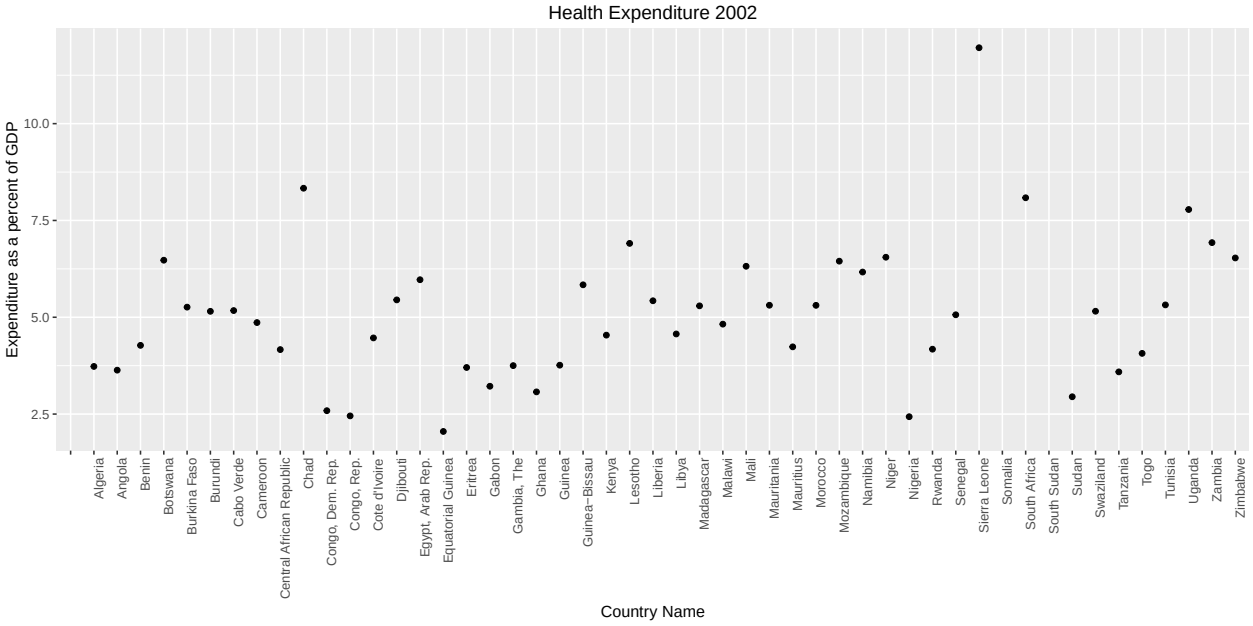
I did my data science project on health in sub-Saharan Africa. Health has been one of the biggest issues in sub-Saharan Africa with a lot of diseases plaguing the continent, examples being the HIV/AIDS epidemic in the 1990's and 2000's, and the recent ebola epidemic in 2014. Health issues have affected African countries especially It can be argued that poverty is the main cause for a lot of the health issues in sub-Saharan Africa, because a lot of African governments lack the necessary resources to and infrastructure develop the health sectors, diseases that should and could be easily controlled spread and affect millions of people.

My thesis is "What has been the effect of increased or decreased expenditures on Health Sectors in Developing countries"? Using mortality rates and increase of immunization as variables of interest to study. I chose mortality rate because it is a snapshot of how many people are dying in a country and thought it would be a good measure. Next I chose immunization rates because I felt this would be an adequate marker of how the potential spread of diseases is being controlled as more immunization leads to less exposure to curable diseases in the future

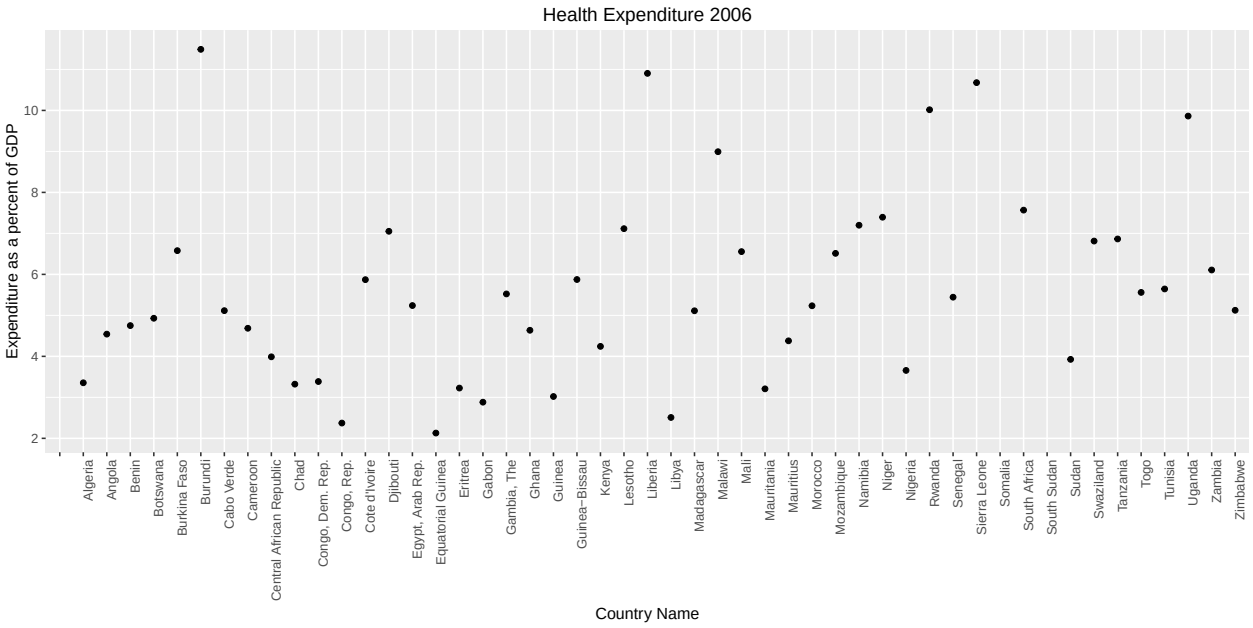
I primarily used the World Bank databases for my data source. The primary variables were about 55 sub-Saharan countries, and I looked at the years 2002 to 2014. I looked at the data through snapshots in time looking at years 2002, 2006, 2010 and 2014.

Expenditure Graphs

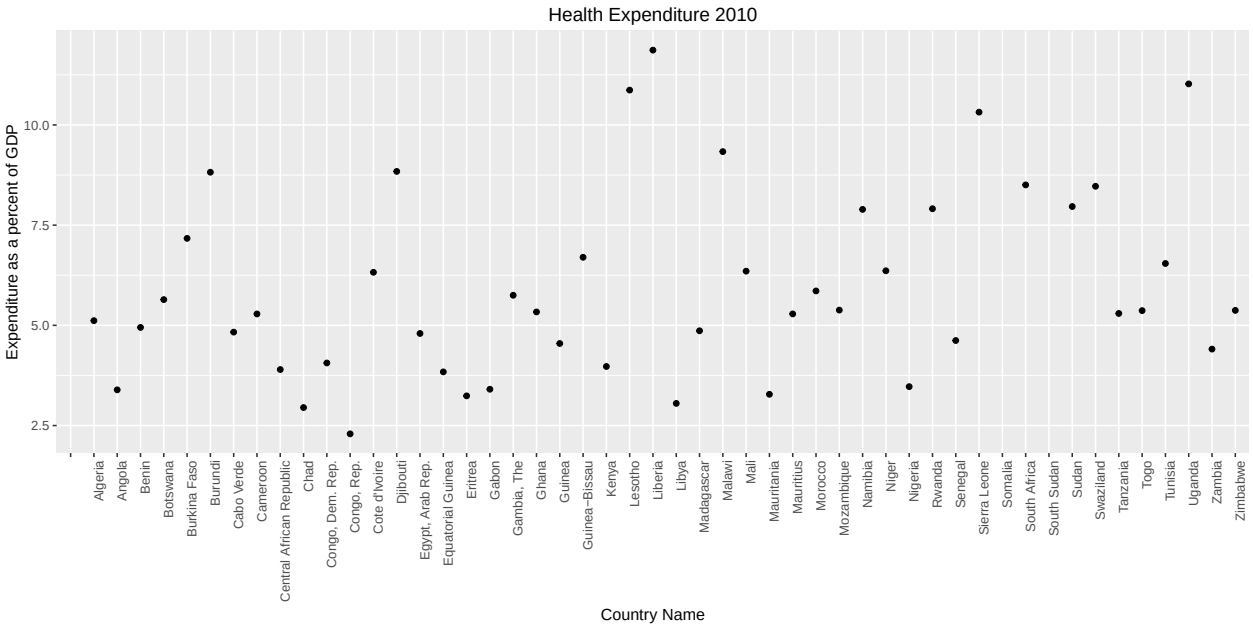
```
## Warning: Removed 7 rows containing missing values (geom_point).
```



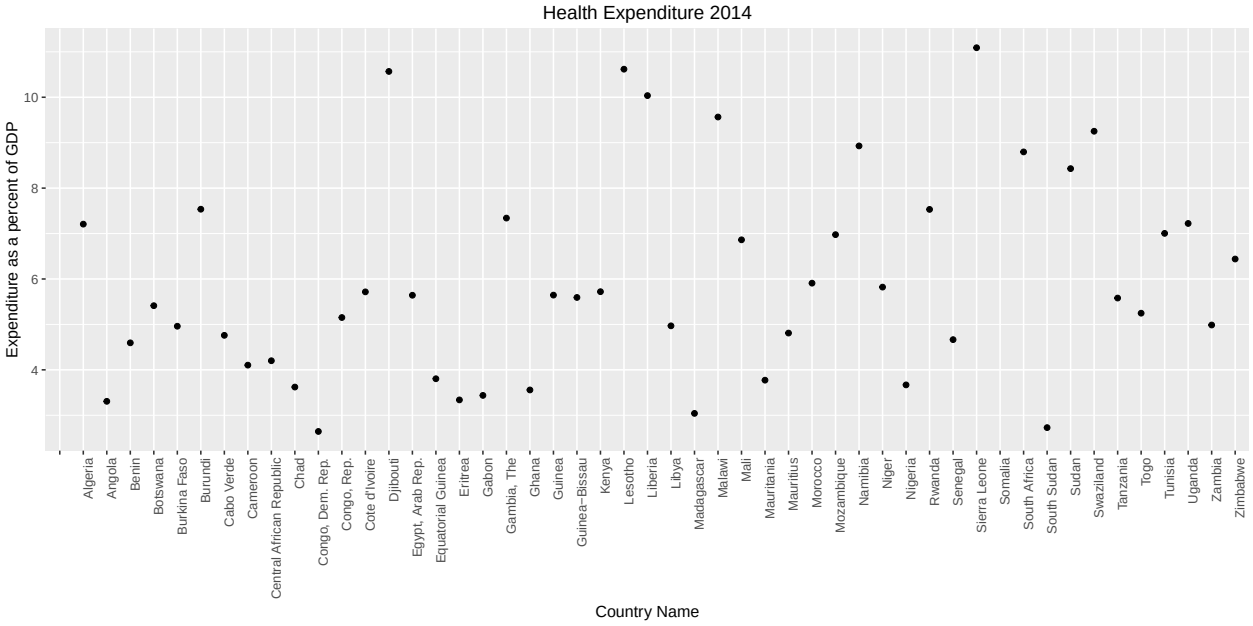

```
## Warning: Removed 7 rows containing missing values (geom_point).
```




```
## Warning: Removed 7 rows containing missing values (geom_point).
```



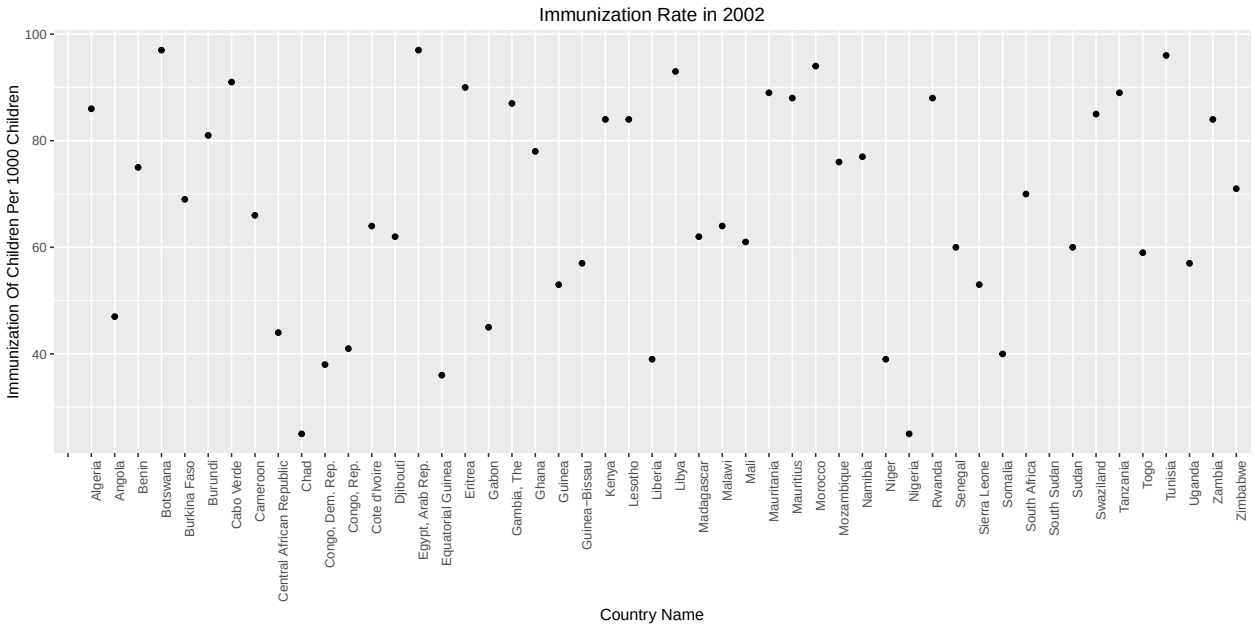

```
## Warning: Removed 6 rows containing missing values (geom_point).
```



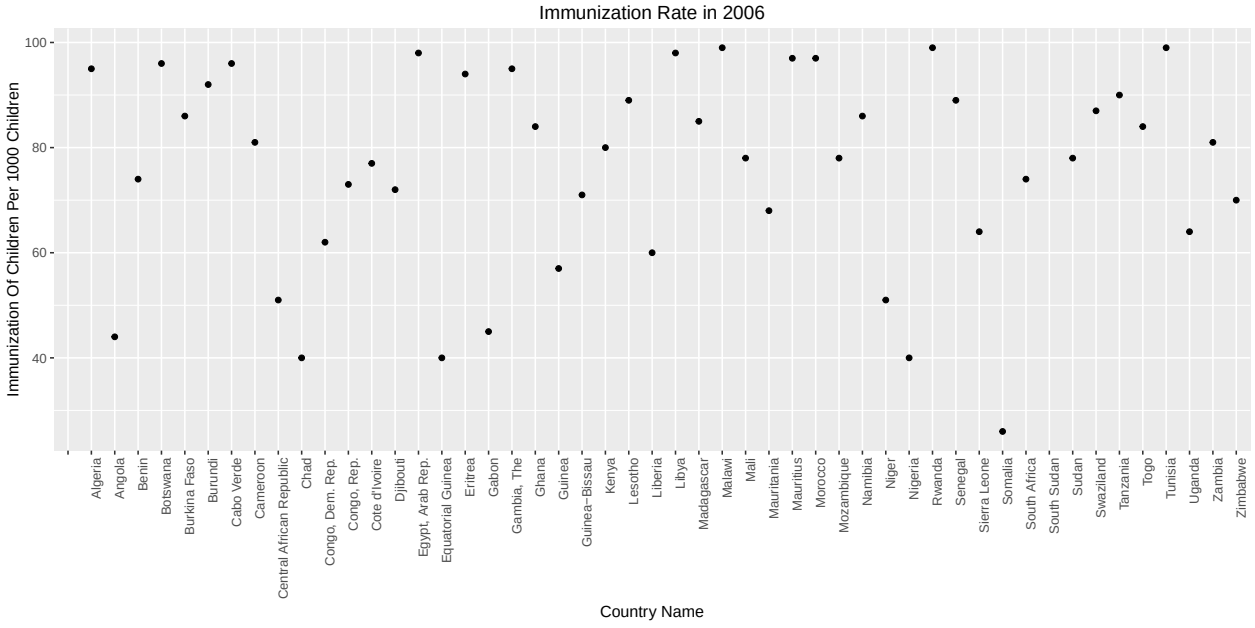
The first four graphs showe expenditures on health as a percent of GDP per capita for each country, in the year 2002, 2006, 2010 and 2014. I chose expenditue as a percent of GDP per capita because I wanted to decrease as much variance as I could between expenditures on health between countries. If I had chosen variables such as total GDP or total health numbers as bigger counries would have bigger expenditures which would skew the results I was looking for.

Immunization Rates Per Country

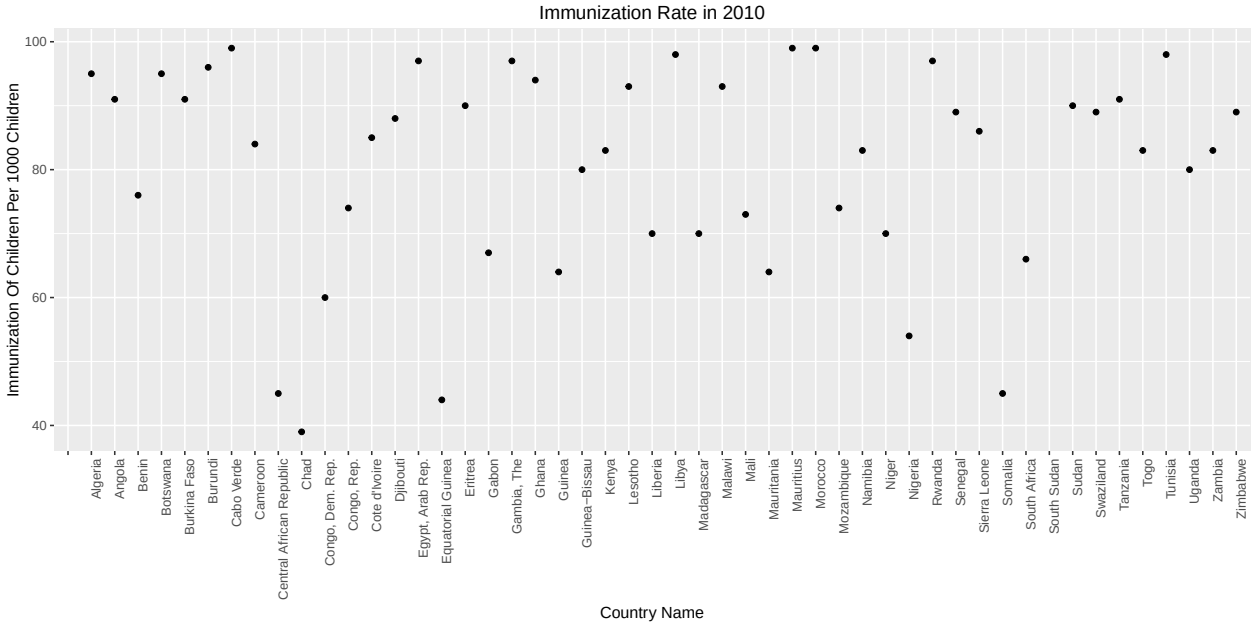
```
## Warning: Removed 6 rows containing missing values (geom_point).
```



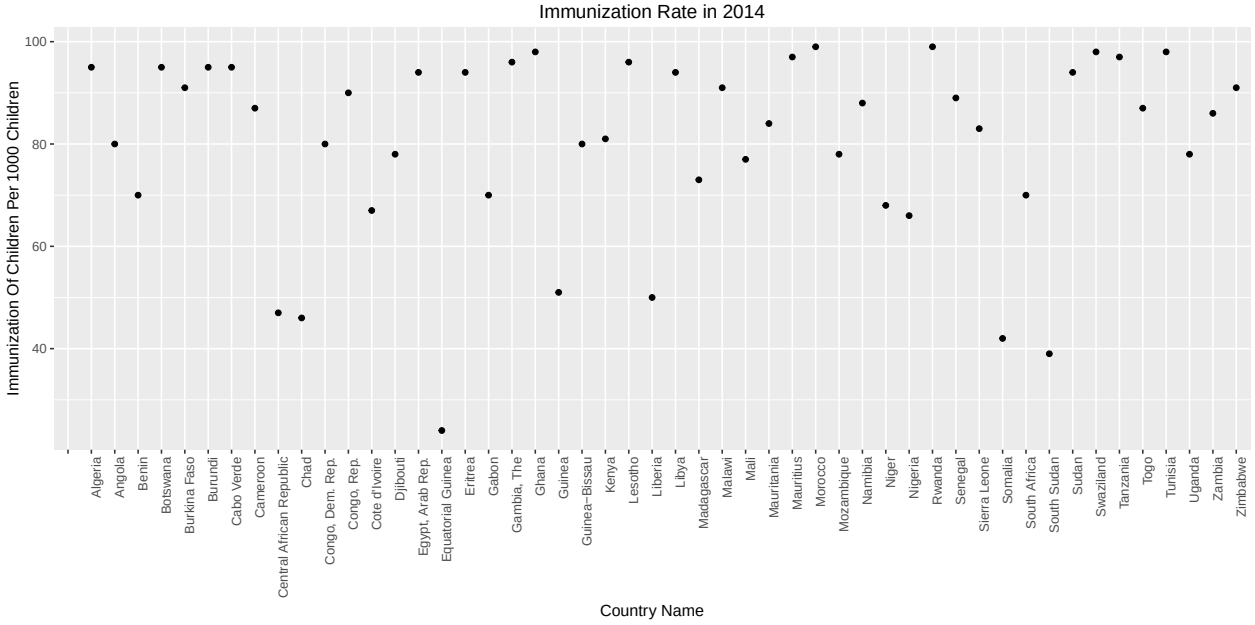

```
## Warning: Removed 6 rows containing missing values (geom_point).
```




```
## Warning: Removed 6 rows containing missing values (geom_point).
```



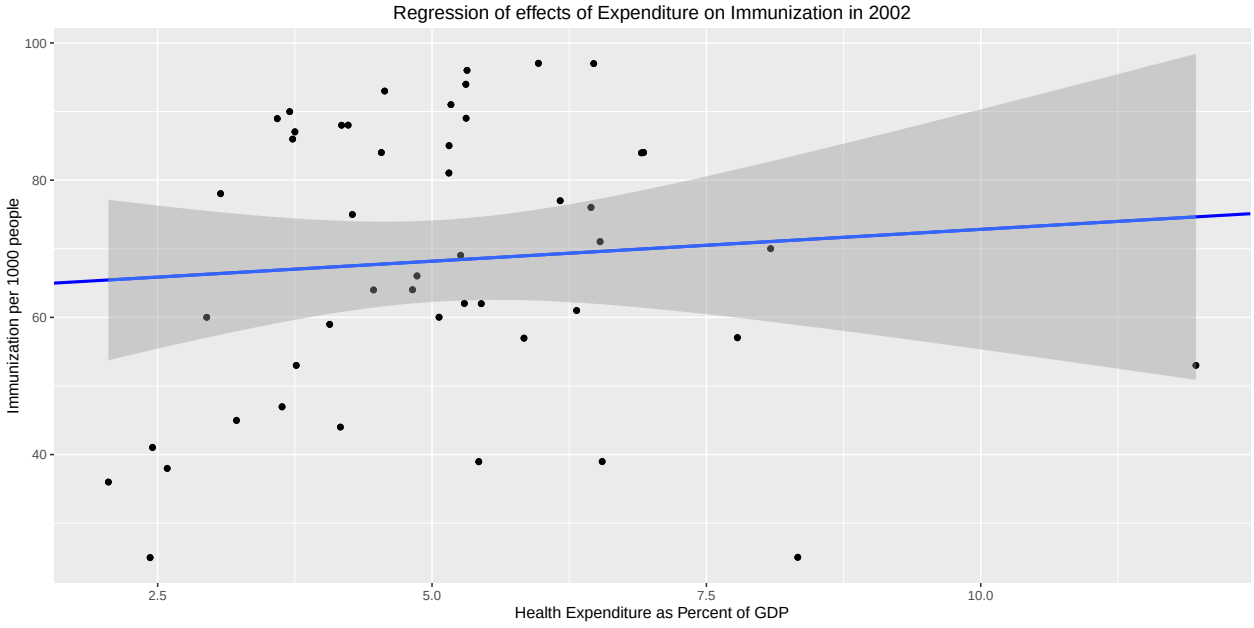

```
## Warning: Removed 5 rows containing missing values (geom_point).
```



The next four graphs show immunization rates for each country for the years 2002, 2006, 2010 and 2014.

```
##
## Call:
## lm(formula = Immunization..DPT....of.children.ages.12.23.months...SH.IMM.IDPT. ~
##     Health.expenditure..total....of.GDP...SH.XPD.TOTL.ZS., data = Worldbank2002A)
##
## Coefficients:
##                                          (Intercept)
##                                              63.5463
## Health.expenditure..total....of.GDP...SH.XPD.TOTL.ZS.
```

```
##                                         0.9279

##           1           2           3           4           5           6
##  18.9925299   7.4874071  12.6715507 -23.4110413 -24.8222557  -3.6910641
##           7           8           9          10          11          12
##  23.0173317 -21.5329501  11.6021964 -14.0376063  16.2425023  25.2146665
##          13          14          15          16          17          18
##  14.0436830  -6.4591102  -8.4079735  20.5230698   6.4698791   7.7308501
##          19          20          21          22          25          26
## -30.6244954 -40.8026783  20.5789365 -21.6444849  16.6697182  -8.3203201
##          27          28          29          30          31          32
## -13.7679008  14.0253645  27.5178305  22.1227700  -6.2807740  -1.0491050
##          33          34          35          36          37          38
##  -8.2454126  25.5280375  20.5258839  -4.0211835 -29.5812410 -11.9634057
##          39          40          41          42          43          44
##  19.9731143 -29.4494800  -6.6020499 -27.9469658 -46.2781002  -2.0588385
##          45          46          47          48          49          50
##   0.5715989  22.6536867  27.4468116 -19.9182584  27.9144573   1.3928187

## Warning: Removed 7 rows containing non-finite values (stat_smooth).

## Warning: Removed 7 rows containing missing values (geom_point).

## Warning: Removed 7 rows containing missing values (geom_point).
```



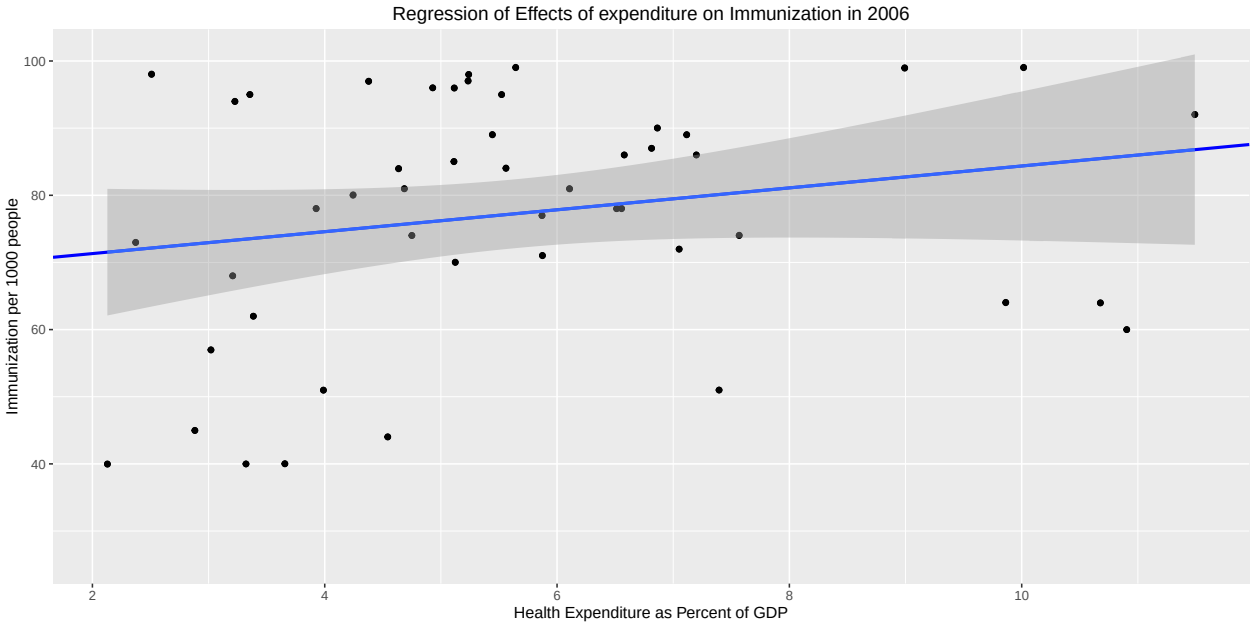

```
##
## Call:
## lm(formula = Immunization..DPT....of.children.ages.12.23.months...SH.IMM.IDPT. ~
##     Health.expenditure..total....of.GDP...SH.XPD.TOTL.ZS., data = Worldbank2002A)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
```

```
## -46.278 -13.835  -0.239  19.238  27.914
##
## Coefficients:
##                                                    Estimate Std. Error
## (Intercept)                                         63.5463     8.9080
## Health.expenditure..total....of.GDP...SH.XPD.TOTL.ZS.   0.9279     1.6583
##                                                    t value Pr(>|t|)
## (Intercept)                                          7.134 5.75e-09 ***
## Health.expenditure..total....of.GDP...SH.XPD.TOTL.ZS.   0.560    0.579
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 20.42 on 46 degrees of freedom
##   (7 observations deleted due to missingness)
## Multiple R-squared:  0.00676,    Adjusted R-squared:  -0.01483
## F-statistic: 0.3131 on 1 and 46 DF,  p-value: 0.5785
```

In the first regression that I ran, we can see that there is a positive effect on effects of health expenditure as a percent of GDP and immunization in 2002. Overall for countries that had higher expenditure percentage rates also had higher immunization rates. This is further shown in the regression model with a coefficient of .9279, and it is statistically significant. This means that for a one percent increase in expenditures as a percent of GDP, there is a .9279 percent increase of the population that is immunized, decreasing their chances of getting infectious diseases in the future.

```
##
## Call:
## lm(formula = Immunization..DPT....of.children.ages.12.23.months...SH.IMM.IDPT. ~
##     Health.expenditure..total....of.GDP...SH.XPD.TOTL.ZS., data = Worldbank2006A)
##
## Coefficients:
##                                             (Intercept)
##                                                   68.06
## Health.expenditure..total....of.GDP...SH.XPD.TOTL.ZS.
##                                                    1.63

##           1           2           3           4           5           6
##  21.4710884  -1.8039625   5.2073178 -23.5619849   1.0728635  -0.6309370
##           7           8           9          10          11          12
##  20.6808365 -27.7575400   8.3820466 -15.9829906   5.0224248  25.8504387
##          13          14          15          16          17          18
##   9.3399237   8.6067345  -0.7468876  21.8021764  -0.6773454   6.2033457
##          19          20          21          22          25          26
## -29.1148235 -34.0208879  14.6100121 -21.4683602   7.8328029   6.8757810
##          27          28          29          30          31          32
## -20.1389029   2.9841233  21.7386090  10.7500055   3.5391477  -6.3974873
##          33          34          35          36          37          38
##  12.0659849  20.4077097  -5.2883502  16.2790315 -25.8371628  -6.6369629
##          39          40          41          42          43          44
##  17.9368153 -31.5295149  -7.5552333 -11.5787427 -33.4744944   5.3019622
##          45          46          47          48          49          50
##   7.2148603  19.6018766  19.9040724 -31.4647813  21.3985761  -6.4132148

## Warning: Removed 7 rows containing non-finite values (stat_smooth).
```

```
## Warning: Removed 7 rows containing missing values (geom_point).

## Warning: Removed 7 rows containing missing values (geom_point).
```

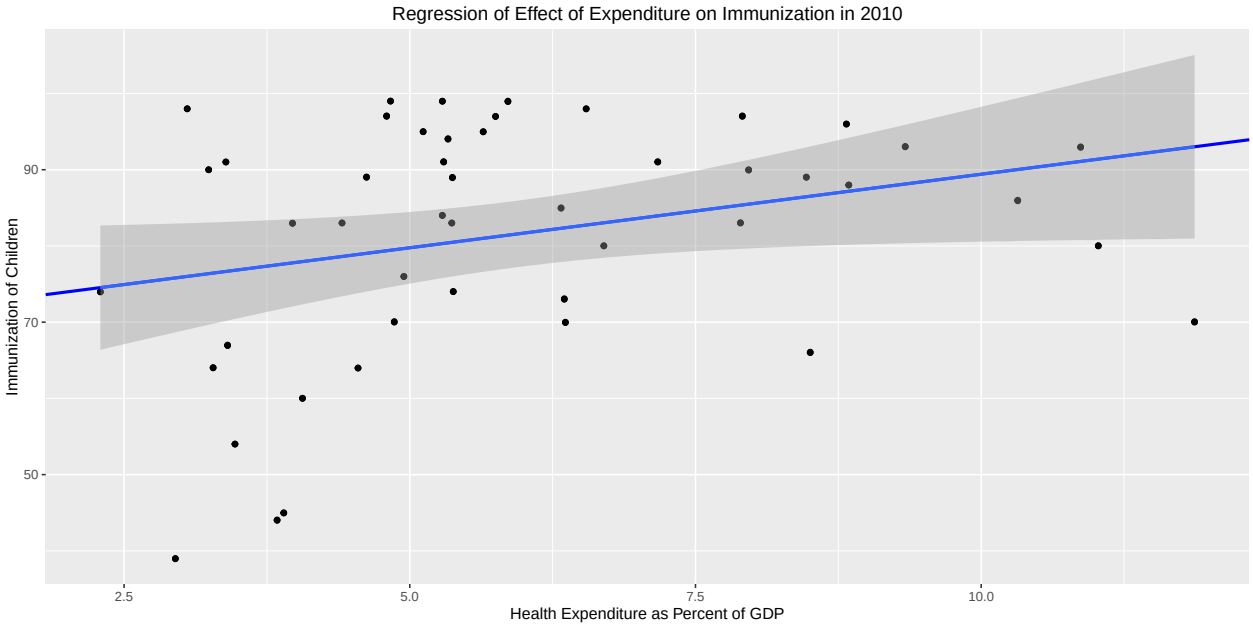



```
##
## Call:
## lm(formula = Immunization..DPT....of.children.ages.12.23.months...SH.IMM.IDPT. ~
##     Health.expenditure..total....of.GDP...SH.XPD.TOTL.ZS., data = Worldbank2006A)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -34.021  -8.561   4.281  12.702  25.850
##
## Coefficients:
##                                                    Estimate Std. Error
## (Intercept)                                          68.059      6.812
## Health.expenditure..total....of.GDP...SH.XPD.TOTL.ZS.    1.630      1.121
##                                                    t value Pr(>|t|)
## (Intercept)                                          9.992 4.18e-13 ***
## Health.expenditure..total....of.GDP...SH.XPD.TOTL.ZS.   1.455    0.152
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 17.65 on 46 degrees of freedom
##   (7 observations deleted due to missingness)
## Multiple R-squared:  0.044,  Adjusted R-squared:  0.02321
## F-statistic: 2.117 on 1 and 46 DF,  p-value: 0.1525
```

This second model looking at 2006 statisitics shows a stronger correlation. In this regression, the coeffecient is 1.630 meaning that a one percent increase in health expenditures as a percent of GDP shows that there is 1.630 percent increase in those that were immunized, and is also statistically significant.

```
##
```

```
## Call:
## lm(formula = Immunization..DPT....of.children.ages.12.23.months...SH.IMM.IDPT. ~
##     Health.expenditure..total....of.GDP...SH.XPD.TOTL.ZS., data = Worldbank2010A)
##
## Coefficients:
##                                          (Intercept)
##                                                70.11
## Health.expenditure..total....of.GDP...SH.XPD.TOTL.ZS.
##                                                 1.93

##           1           2           3           4           5           6
##  15.0138373  -3.6601988   8.8669476 -32.6323961  -0.5359320   2.6852241
##           7           8           9          10          11          12
##  13.6356782  -9.6834857  13.5943724 -14.8861695   5.2203986  21.9980142
##          13          14          15          16          17          18
##   1.9122181  -9.4983108  -9.3692405  18.6894210  -6.4940149  -2.3434531
##          19          20          21          22          25          26
## -12.3876658 -22.8085655  11.6245496  -4.0256793   2.5437424   2.5302767
##          27          28          29          30          31          32
## -11.3864669   4.3837215  15.2611306  10.6673266   4.5180318 -20.5222672
##          33          34          35          36          37          38
##   9.9703479  17.5825351 -12.4396268   4.8736955 -23.0119732  -3.0372319
##          39          40          41          42          43          44
##  15.7916129 -33.5200061   0.8268159 -17.9484198 -36.8017091   3.6902502
##          45          46          47          48          49          50
##   7.0522436  19.5650907  13.9997538  14.3446181  17.6311604   8.5197984

## Warning: Removed 7 rows containing non-finite values (stat_smooth).

## Warning: Removed 7 rows containing missing values (geom_point).

## Warning: Removed 7 rows containing missing values (geom_point).
```



```
##
## Call:
## lm(formula = Immunization..DPT....of.children.ages.12.23.months...SH.IMM.IDPT. ~
##     Health.expenditure..total....of.GDP...SH.XPD.TOTL.ZS., data = Worldbank2010A)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -36.802  -9.545   2.614  12.117  21.998
##
## Coefficients:
##                                                    Estimate Std. Error
## (Intercept)                                         70.1113     5.9843
## Health.expenditure..total....of.GDP...SH.XPD.TOTL.ZS.   1.9298     0.9402
##                                                    t value Pr(>|t|)
## (Intercept)                                         11.716 2.09e-15 ***
## Health.expenditure..total....of.GDP...SH.XPD.TOTL.ZS.   2.053   0.0458 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 15 on 46 degrees of freedom
##   (7 observations deleted due to missingness)
## Multiple R-squared:  0.0839, Adjusted R-squared:  0.06399
## F-statistic: 4.213 on 1 and 46 DF,  p-value: 0.04583
```

This third model looking at 2010 statistics shows a stronger correlation. In this regression, the coeffecient is 1.9298 meaning that a one percent increase in health expenditures as a percent of GDP shows that there is 1.9298 percent increase in those that were immunized, and is also statistically significant.

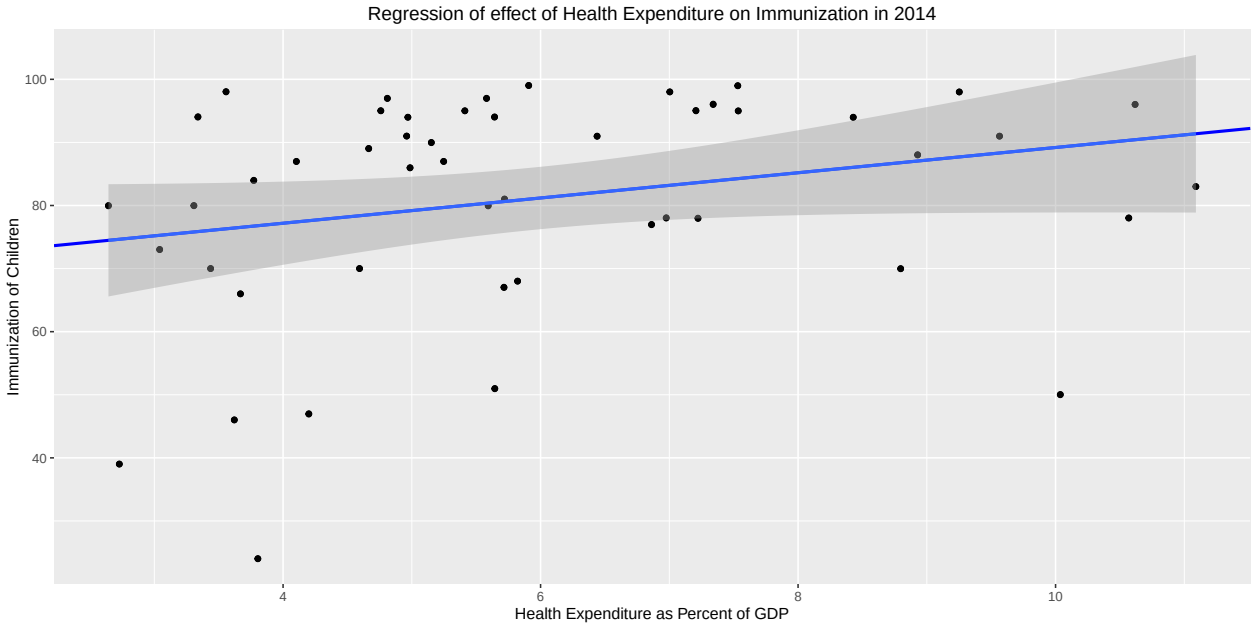
```
##
## Call:
## lm(formula = Immunization..DPT....of.children.ages.12.23.months...SH.IMM.IDPT. ~
##     Health.expenditure..total....of.GDP...SH.XPD.TOTL.ZS., data = Worldbank2014A)
##
## Coefficients:
##                                             (Intercept)
##                                                   69.19
## Health.expenditure..total....of.GDP...SH.XPD.TOTL.ZS.
##                                                    2.00

##           1           2           3           4           5           6
##  11.3937490  -8.3798573  10.7368951 -30.5916134  10.5057786 -13.6240822
##           7           8           9          10          11          12
##  18.1324526  -6.0659283  21.6944023 -29.4811136   0.3678815  14.8707753
##          13          14          15          16          17          18
##   5.5719285  -2.2735994  -5.9161446  18.1895945  -5.1430027   0.9503143
##          19          20          21          22          24          25
## -12.8341588 -10.5286377  14.7449246  -8.3726365 -35.6469394  10.3037325
##          26          27          28          29          30          31
##   7.3148248  -5.6361722   6.8351117  14.8004383  16.6482660   7.9502608
##          32          33          34          35          36          37
## -16.7862477  10.4787747  17.9928626   7.2635360   2.6776738 -39.2652443
##          38          39          40          41          42          43
##  -0.3777800  12.1272714 -52.7992316 -12.3292184   5.5213671 -30.4333581
```

```
##          44          45          46          47          48          49
##   9.6015110  11.8881210  16.2893039  14.9845332   4.1950806  13.5242528
##          50
##   8.9293480
```

```
## Warning: Removed 6 rows containing non-finite values (stat_smooth).
```

```
## Warning: Removed 6 rows containing missing values (geom_point).
```

```
## Warning: Removed 6 rows containing missing values (geom_point).
```



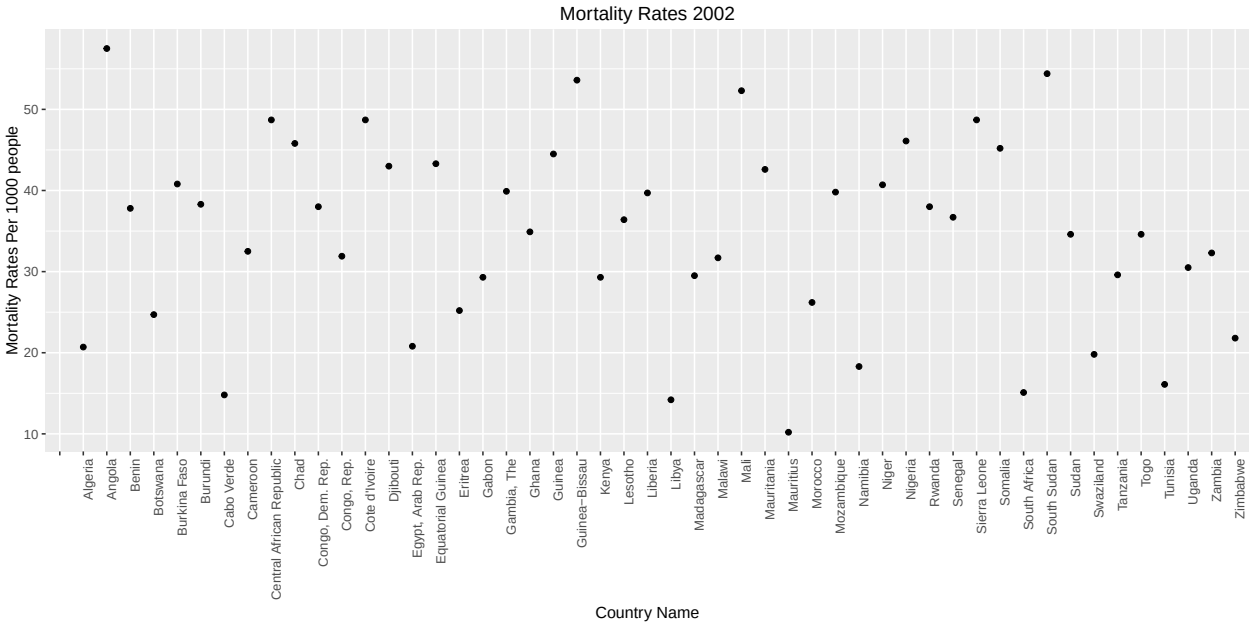

```
##
## Call:
## lm(formula = Immunization..DPT....of.children.ages.12.23.months...SH.IMM.IDPT. ~
##     Health.expenditure..total....of.GDP...SH.XPD.TOTL.ZS., data = Worldbank2014A)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -52.799  -8.373   5.572  11.888  21.694
##
## Coefficients:
##                                                  Estimate Std. Error
## (Intercept)                                        69.190      7.054
## Health.expenditure..total....of.GDP...SH.XPD.TOTL.ZS.     2.000      1.110
##                                                  t value Pr(>|t|)
## (Intercept)                                        9.808 5.96e-13 ***
## Health.expenditure..total....of.GDP...SH.XPD.TOTL.ZS.    1.801   0.0781 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 17.2 on 47 degrees of freedom
##   (6 observations deleted due to missingness)
```

```
## Multiple R-squared:  0.06458,    Adjusted R-squared:  0.04468
## F-statistic: 3.245 on 1 and 47 DF,  p-value: 0.07806
```

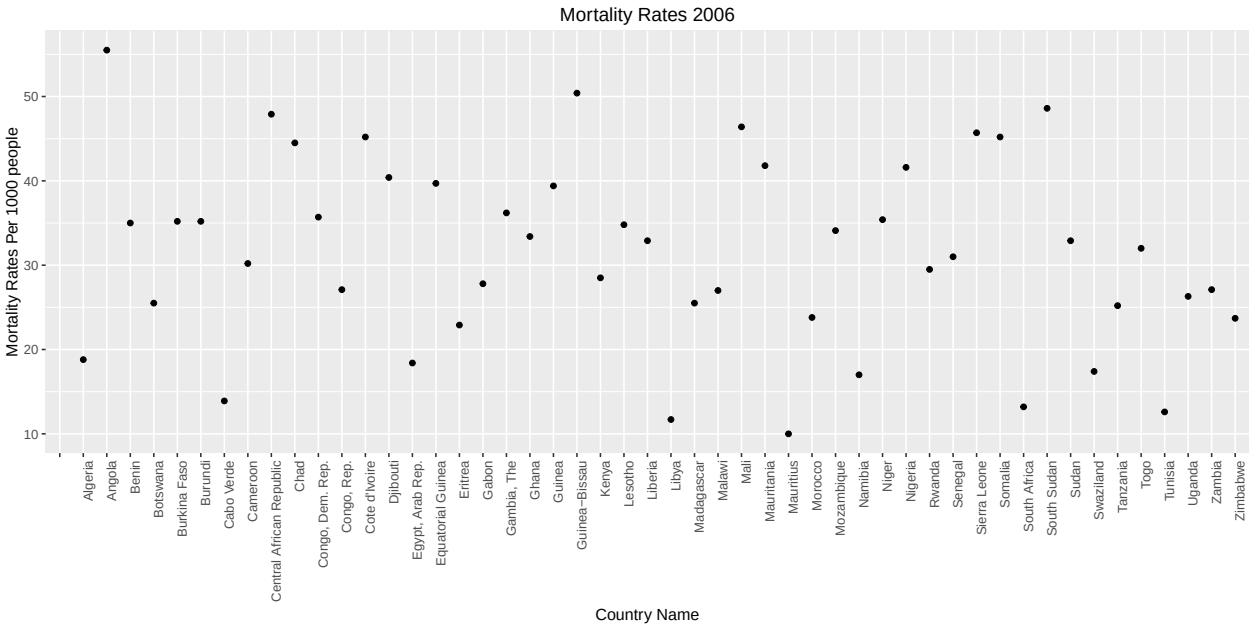
This fourth modellooking at 2014 statistics shows a stronger correlation. In this regression, the coeffecient is 2 meaning that a one percent increase in health expenditures as a percent of GDP shows that there is 2 percent increase in those that were immunized, and is also statistically significant.

Overall we see that there are also increases in the effeciency of expenditures as a percent of GDP to immunization rates in the population.However the models aren't good fits as the r and r-squared numbers are pretty low. This effect can be explained by some health related effects as infrastuructre development has overall increased making access to immunization, easier and more effective so making expenditures as a percentage of GDP have a larger effect on immunization rates.

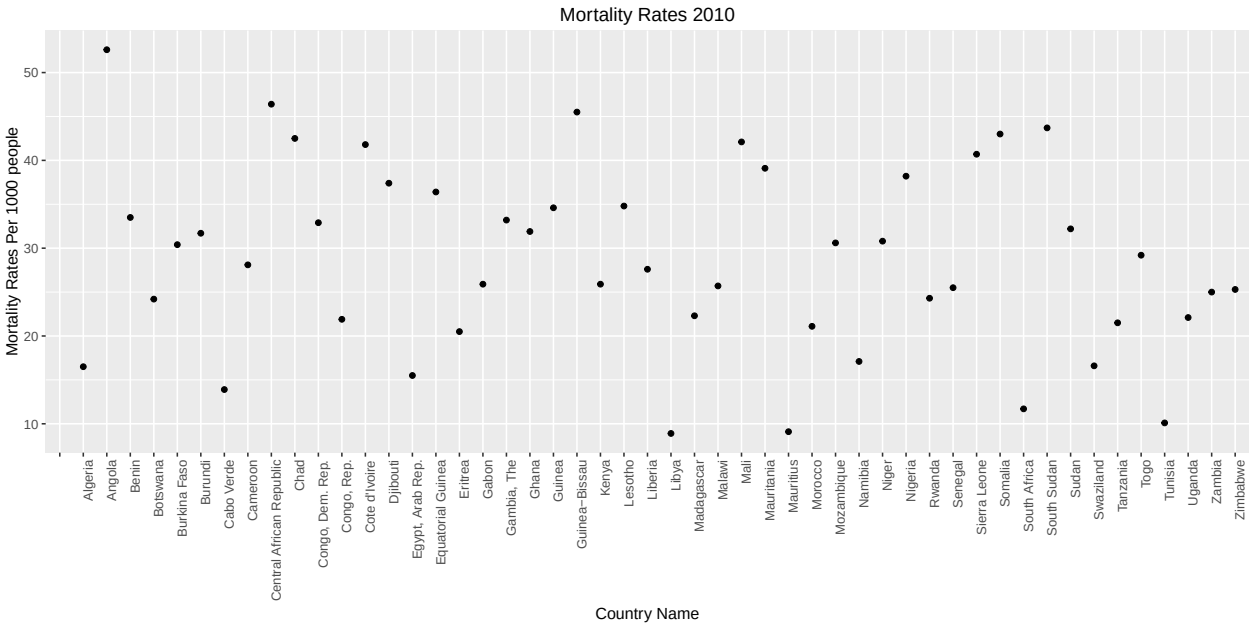
```
## Warning: Removed 5 rows containing missing values (geom_point).
```



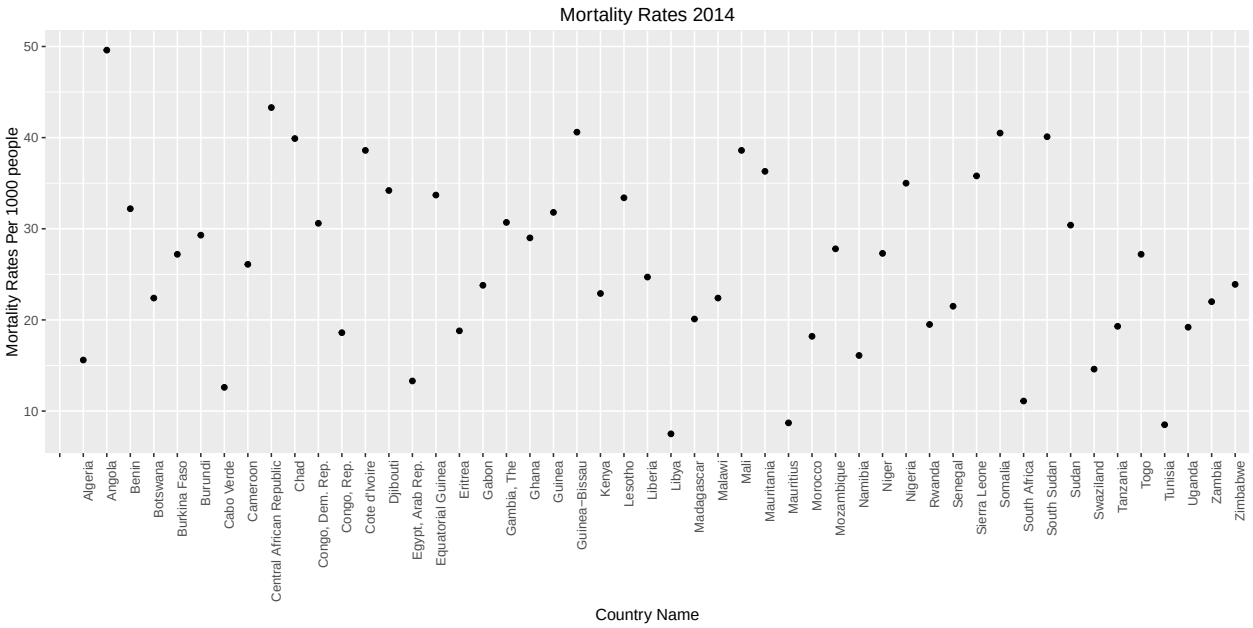

```
## Warning: Removed 5 rows containing missing values (geom_point).
```




```
## Warning: Removed 5 rows containing missing values (geom_point).
```




```
## Warning: Removed 5 rows containing missing values (geom_point).
```



These four graphs show neonatal mortality rates for african countries in the years 2002, 2006,2010 and 2014. Overall mortality rates have decreased and life expectancy has increased for many africans countries. Angola has some of the higest mortality rates of about 50 people per 1000 or about 5 percent while, Lybia has a mortality rate of under one percent.

```
##
## Call:
## lm(formula = Mortality.rate..neonatal..per.1.000.live.births...SH.DYN.NMRT. ~
##     Health.expenditure..total....of.GDP...SH.XPD.TOTL.ZS., data = Worldbank2002A)
##
## Coefficients:
##                                          (Intercept)
##                                              34.1132
## Health.expenditure..total....of.GDP...SH.XPD.TOTL.ZS.
##                                              -0.0326

##           1           2           3           4           5           6
## -13.2915765   3.8261724   4.3548399  14.7226041  -2.1332224  14.7324435
##           7           8           9          10          11          12
##  -8.7924480  -4.7082502   0.8870011  10.5094824  -4.6652222 -19.7642441
##          13          14          15          16          17          18
##   2.5120397  -4.4405690  18.3927721 -23.7750807   5.8970641 -15.6121059
##          19          20          21          22          25          26
##   6.8003802  12.0660896   4.0229563  14.9767719 -14.1450957   0.6194164
##          27          28          29          30          31          32
##  -3.3594430  -1.5873166 -17.8397587  -4.3961529   0.5828889 -18.7495621
##          33          34          35          36          37          38
##   2.7519221  -7.7401174   8.6599583  -2.2559568   5.7637224  19.6771509
##          39          40          41          42          43          44
##   5.9091057   9.2536790   9.0644536   3.9711596  11.9584843  -1.4546337
##          45          46          47          48          49          50
##   6.8583520 -19.1445324  -9.2021254  23.5052888 -13.1185575 -12.1002282

## Warning: Removed 7 rows containing non-finite values (stat_smooth).
```

```
## Warning: Removed 7 rows containing missing values (geom_point).

## Warning: Removed 7 rows containing missing values (geom_point).

## Warning: Removed 1 rows containing missing values (geom_abline).
```


```
##
## Call:
## lm(formula = Mortality.rate..neonatal..per.1.000.live.births...SH.DYN.NMRT. ~
##     Health.expenditure..total....of.GDP...SH.XPD.TOTL.ZS., data = Worldbank2002A)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -23.7751  -8.0032   0.7532   7.3088  23.5053
##
## Coefficients:
##                                                    Estimate Std. Error
## (Intercept)                                         34.1132     5.0300
## Health.expenditure..total....of.GDP...SH.XPD.TOTL.ZS.  -0.0326     0.9364
##                                                    t value Pr(>|t|)
## (Intercept)                                          6.782 1.94e-08 ***
## Health.expenditure..total....of.GDP...SH.XPD.TOTL.ZS.  -0.035    0.972
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 11.53 on 46 degrees of freedom
##   (7 observations deleted due to missingness)
## Multiple R-squared:  2.636e-05,  Adjusted R-squared:  -0.02171
## F-statistic: 0.001212 on 1 and 46 DF,  p-value: 0.9724
```

This regression shows that there has been a negative correlation between expenditures on GDP and mortality rates in 2002. For every one percent of health expenditures as a percent of GDP there is a decrease of the mortality rate by .03.

```
##
## Call:
## lm(formula = Mortality.rate..neonatal..per.1.000.live.births...SH.DYN.NMRT. ~
##     Health.expenditure..total....of.GDP...SH.XPD.TOTL.ZS., data = Worldbank2006A)
##
## Coefficients:
##                                          (Intercept)
##                                             30.39785
## Health.expenditure..total....of.GDP...SH.XPD.TOTL.ZS.
##                                              0.09721

##           1           2           3           4           5
## -11.92398579   4.14037797   3.68524157  17.11442342  -3.52848957
##           6           7           8           9          10
##  14.23145559  -7.81148082  -2.87799739   2.55146764   8.70856147
##          11          12          13          14          15
##  -2.31035365 -18.94175032   3.71048034  -5.39475568  15.36492374
##          16          17          18          19          20
## -20.82348464   3.06906977 -14.09766230   4.28336877  10.84668308
##          21          22          25          26          27
##  -1.87151222  14.26419631 -13.66013470   1.06166561  -5.05654278
##          28          29          30          31          32
##  -3.89149413 -18.34651245  -5.86507099   2.12045284 -17.93348336
##          33          34          35          36          37
##   0.07300537  -7.10662134  11.09035624  -4.27200696   1.44220867
##          38          39          40          41          42
##  19.43109633   5.26530441   9.09521622   9.31673098   4.97304333
##          43          44          45          46          47
##  13.77925851  -0.65330691   4.16264319 -16.99504531  -5.37702870
##          48          49          50
##  24.66059961 -12.50716588  -7.19594504

## Warning: Removed 7 rows containing non-finite values (stat_smooth).

## Warning: Removed 7 rows containing missing values (geom_point).

## Warning: Removed 7 rows containing missing values (geom_point).

## Warning: Removed 1 rows containing missing values (geom_abline).
```


```
##
## Call:
## lm(formula = Mortality.rate..neonatal..per.1.000.live.births...SH.DYN.NMRT. ~
##     Health.expenditure..total....of.GDP...SH.XPD.TOTL.ZS., data = Worldbank2006A)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -20.8235  -6.1755   0.5673   6.1261  24.6606
##
## Coefficients:
##                                                    Estimate Std. Error
## (Intercept)                                        30.39785    4.22800
## Health.expenditure..total....of.GDP...SH.XPD.TOTL.ZS.  0.09721    0.69557
##                                                    t value Pr(>|t|)
## (Intercept)                                           7.19 4.74e-09 ***
## Health.expenditure..total....of.GDP...SH.XPD.TOTL.ZS.    0.14    0.889
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 10.96 on 46 degrees of freedom
##   (7 observations deleted due to missingness)
## Multiple R-squared:  0.0004244,  Adjusted R-squared:  -0.02131
## F-statistic: 0.01953 on 1 and 46 DF,  p-value: 0.8895
```

The year 2006 was an outlier, my regression shows that there was a positive correlation between expensiture on health as a percent of GDP and mortality rates, meaning that expenditures on health did lead to increases in mortalitiy rate.There are no clear reasons why this is the case other than a high number of deaths that year. In 2010, the correlation goes back to negative.

```
##
## Call:
## lm(formula = Mortality.rate..neonatal..per.1.000.live.births...SH.DYN.NMRT. ~
##     Health.expenditure..total....of.GDP...SH.XPD.TOTL.ZS., data = Worldbank2010A)
```

```
##
## Coefficients:
##                                          (Intercept)
##                                              29.4189
## Health.expenditure..total....of.GDP...SH.XPD.TOTL.ZS.
##                                              -0.2012
```

```
##           1           2           3           4           5           6
## -11.8893501   5.0766647   4.0557899  17.7652444  -7.0575952  13.6534328
##           7           8           9          10          11          12
##  -8.2669659  -2.8336897   3.5543836   6.0959638  -2.7194078 -19.9047433
##          13          14          15          16          17          18
##   7.5681130  -6.1402139  13.9591114 -19.2555262   2.2635989 -10.7307950
##          19          20          21          22          25          26
##   2.6610324   9.4793513  -3.5274590  13.3573774 -11.1147732   0.8610663
##          27          28          29          30          31          32
##  -5.1007459  -3.5321752 -18.0023509  -6.8532226   4.3836467 -16.0078910
##          33          34          35          36          37          38
##  -2.9890766  -7.1401215  10.3408854  -1.8406529   0.5687306  17.4287566
##          39          40          41          42          43          44
##   4.9380799   7.7535266   9.7599741   4.2981933  13.6743757  -0.2556127
##          45          46          47          48          49          50
##   2.4236886 -14.5468242  -4.0836210  23.8633802 -12.9537127  -3.0378412
```

```
## Warning: Removed 7 rows containing non-finite values (stat_smooth).
```

```
## Warning: Removed 7 rows containing missing values (geom_point).
```

```
## Warning: Removed 7 rows containing missing values (geom_point).
```

```
## Warning: Removed 1 rows containing missing values (geom_abline).
```

```
##
## Call:
## lm(formula = Mortality.rate..neonatal..per.1.000.live.births...SH.DYN.NMRT. ~
##     Health.expenditure..total....of.GDP...SH.XPD.TOTL.ZS., data = Worldbank2010A)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -19.9047  -6.9043   0.1566   6.4640  23.8634
##
## Coefficients:
##                                                    Estimate Std. Error
## (Intercept)                                         29.4189     4.1529
## Health.expenditure..total....of.GDP...SH.XPD.TOTL.ZS.  -0.2012     0.6524
##                                                    t value Pr(>|t|)
## (Intercept)                                          7.084 6.82e-09 ***
## Health.expenditure..total....of.GDP...SH.XPD.TOTL.ZS.  -0.308    0.759
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 10.41 on 46 degrees of freedom
##   (7 observations deleted due to missingness)
## Multiple R-squared:  0.002063,   Adjusted R-squared:  -0.01963
## F-statistic: 0.0951 on 1 and 46 DF,  p-value: 0.7592
```

This regression shows that there continues to be a negative correlation between expenditures on GDP and mortality rates in 2010 as the coeffecient is -.2012. A one percent increase in health expenditures as a percnet of GDP translates to a reduction of mortality rates by .2012. For every one percent of health expenditures as a percent of GDP there is a decrease of the mortality rate by .2012.

```
##
## Call:
## lm(formula = Mortality.rate..neonatal..per.1.000.live.births...SH.DYN.NMRT. ~
##     Health.expenditure..total....of.GDP...SH.XPD.TOTL.ZS., data = Worldbank2014A)
##
## Coefficients:
##                                          (Intercept)
##                                              30.3043
## Health.expenditure..total....of.GDP...SH.XPD.TOTL.ZS.
##                                              -0.7228

##           1           2           3           4           5           6
##  -9.4951163   5.2163527   4.4422346  16.0315247  -7.9809774  12.4272921
##           7           8           9          10          11          12
##  -9.0914578  -4.0197739   1.2668294   5.5756311  -3.2698041 -19.2128672
##          13          14          15          16          17          18
##  10.7699143  -8.0060780  13.2555173 -18.1280709   2.5374913  -7.7508484
##          19          20          21          22          24          25
##   1.2032022   7.3474238  -5.3606668  13.5112281  11.7674816  -9.0172094
##          26          27          28          29          30          31
##   0.6880227  -5.8843044  -4.6999803 -16.7420714  -6.9711196   6.1878259
##          32          33          34          35          36          37
## -12.8460404  -5.4325647  -7.8342929   8.7225208  -0.9909862   1.6497325
##          38          39          40          41          42          43
```

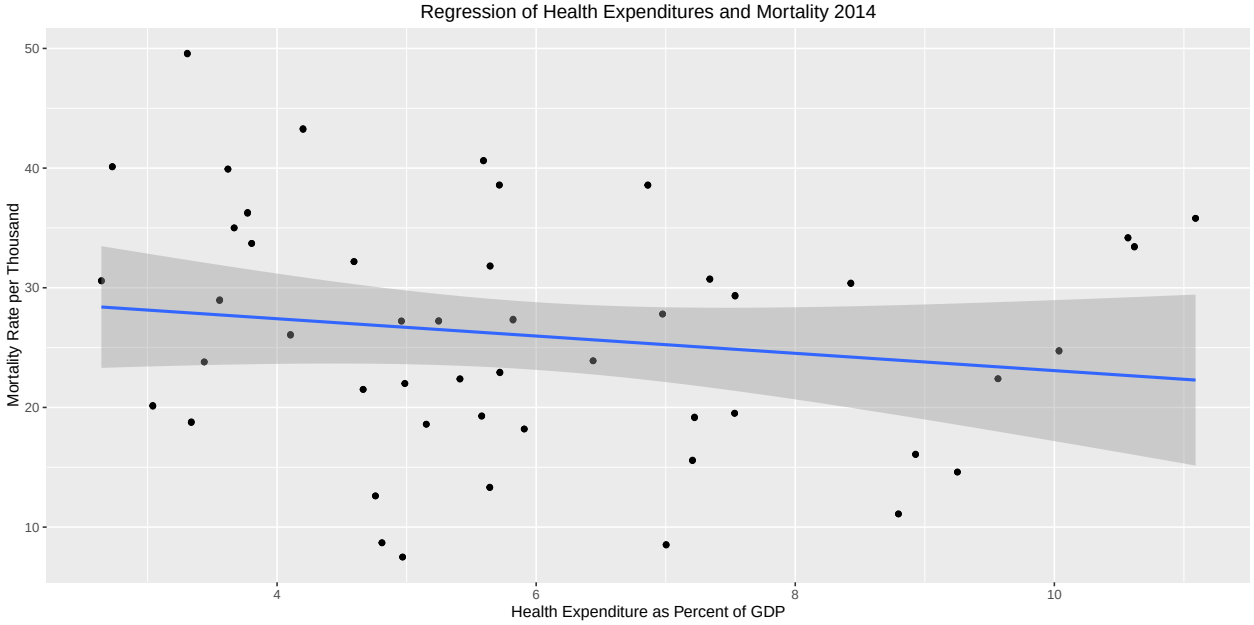
```
##  14.3382920   5.7011741   6.1452015  11.5341942   2.2066649  12.2129950
##          44          45          46          47          48          49
##  -1.2382598   0.4808651 -14.2641004  -3.9926280  21.6859118 -12.9263080
##          50
##  -1.7499972
```

```
## Warning: Removed 6 rows containing non-finite values (stat_smooth).
```

```
## Warning: Removed 6 rows containing missing values (geom_point).
```

```
## Warning: Removed 6 rows containing missing values (geom_point).
```

```
## Warning: Removed 1 rows containing missing values (geom_abline).
```




```
##
## Call:
## lm(formula = Mortality.rate..neonatal..per.1.000.live.births...SH.DYN.NMRT. ~
##     Health.expenditure..total....of.GDP...SH.XPD.TOTL.ZS., data = Worldbank2014A)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -19.2129  -7.7508   0.4809   6.1878  21.6859
##
## Coefficients:
##                                                  Estimate Std. Error
## (Intercept)                                       30.3043     4.0342
## Health.expenditure..total....of.GDP...SH.XPD.TOTL.ZS.  -0.7228     0.6350
##                                                  t value Pr(>|t|)
## (Intercept)                                        7.512 1.38e-09 ***
## Health.expenditure..total....of.GDP...SH.XPD.TOTL.ZS.  -1.138    0.261
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
##
## Residual standard error: 9.834 on 47 degrees of freedom
##   (6 observations deleted due to missingness)
## Multiple R-squared:  0.02682,    Adjusted R-squared:  0.006119
## F-statistic: 1.295 on 1 and 47 DF,  p-value: 0.2608
```

This regression shows that in 2014, there continued to be a negative correlation between expenditures on GDP and mortality rates, as the coefficient is -.7228. For every one percent of health expenditures as a percent of GDP there is a decrease of the mortality rate by .7228. However much like the previous regression models, the correlation is not strong with an adjusted r squared value of 0.006119, which is an extremely weak correlation.

Limitations

The biggest limitation with the analysis are the fit of the regression models. The r's and r-squared are very low signifying that the correlation may be very low as well. There are many reasons why correlations can't be established. One is variance of expenditures among countries. Expenditures as a percent of GDP can vary form country to country depending on necessity. A country that is going through an infectious disease crisis might need to spend a higher percent on GDP on health than a country with a relativly healthier population. In 2006 there was a positive correlation between increases in expenditures and mortality rates which was strange compared to the other measures. What would have helped my analysis would have been looking at individual countries and studying them over time to further detect patterns between increases and decreases in mortality rates. This would have been particularly helpful for countries with the most increases or decreases in health expenditures as a percent of GDP.

Conclusion

Although the regression fits are weak, there is are correlations between countries that spend higher on health and an increase in immmunization and disease prevention,and a decrease in mortality rates.Over time, the effects of health expenditures as a percent of GDP on the parameters, especially immunization increased over time.As mentioned before this is due to better infrastrucutre developments that improve the access of healthcare to people and the spread of immunization and medicines to people.

Sources Cited World Bank Databases