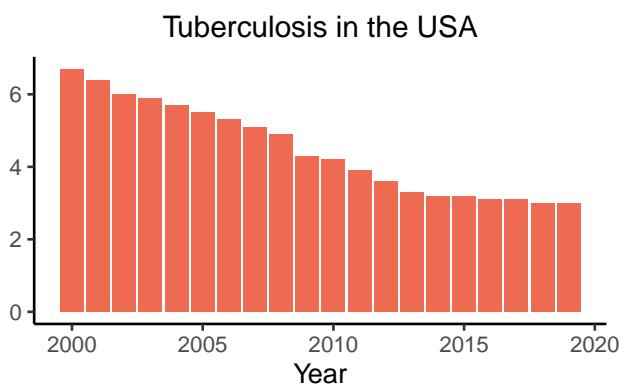
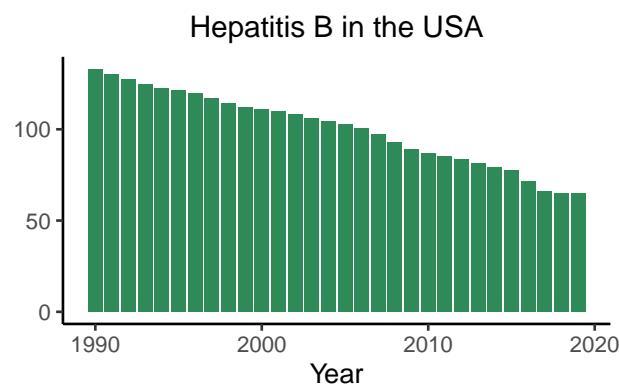
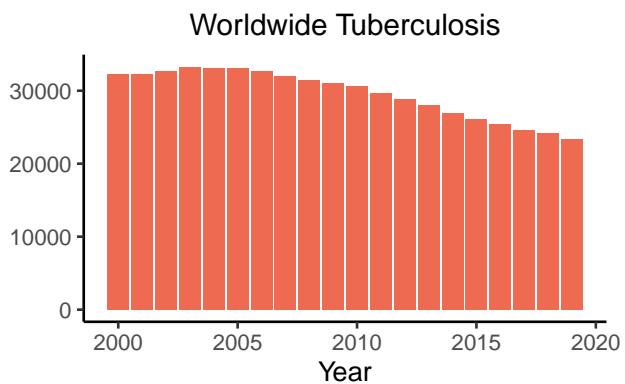
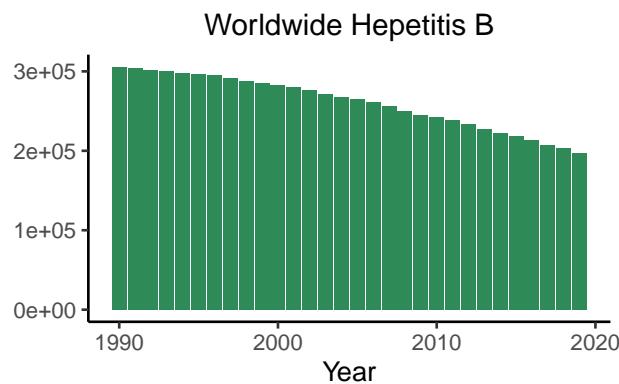


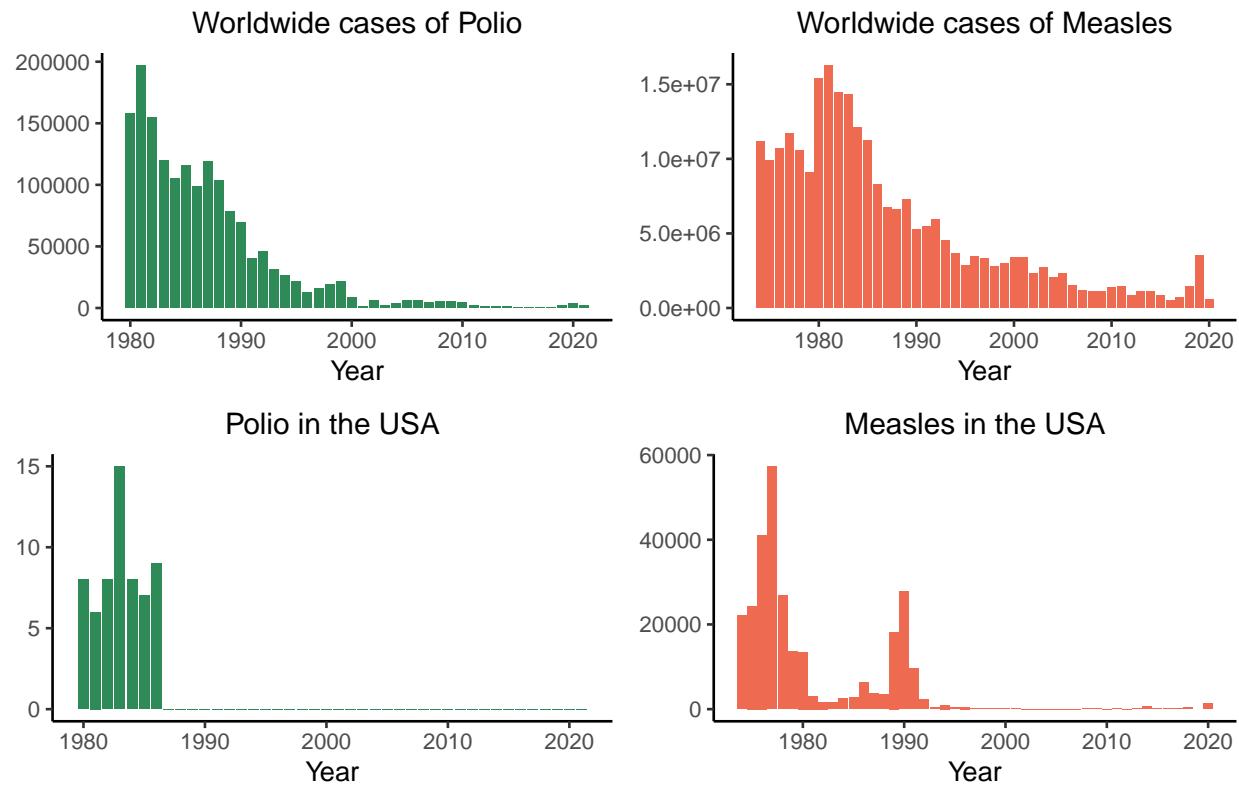
# ProjectDBM

Shika, Kevin, Laura

## Distribution of diseases

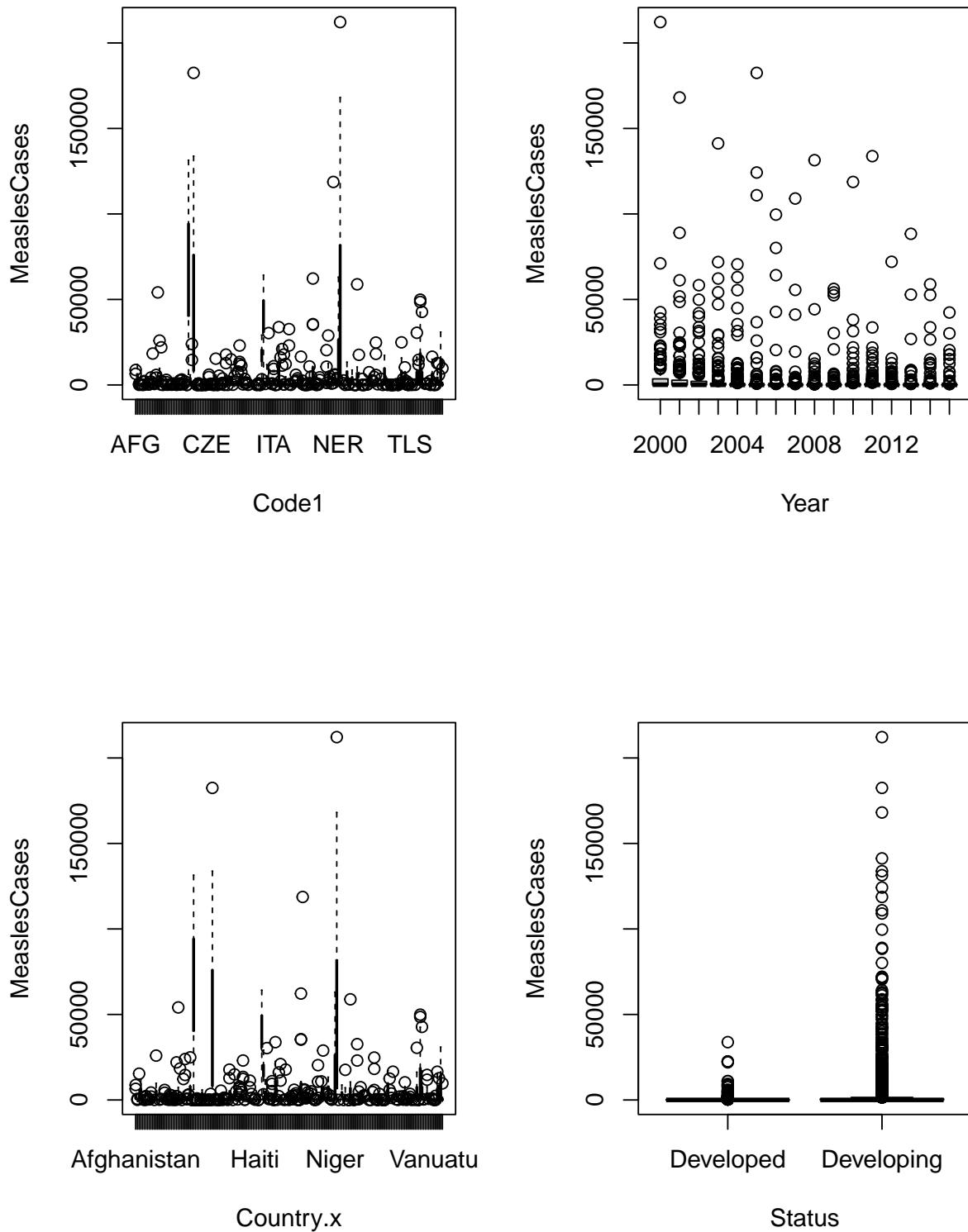
Cases per year across the world and Cases per year in US

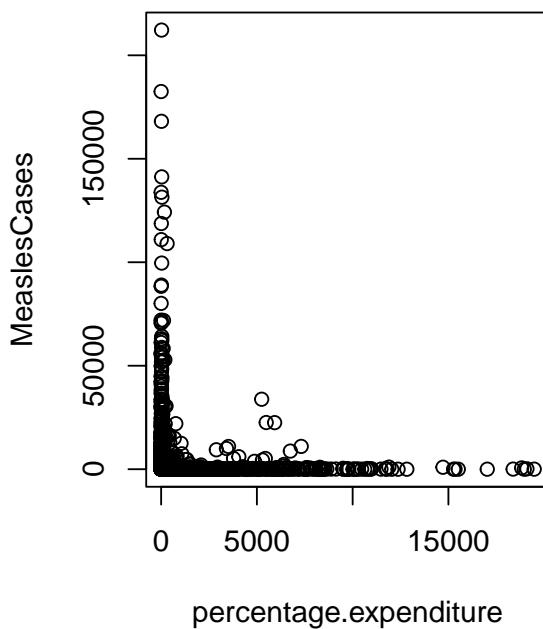
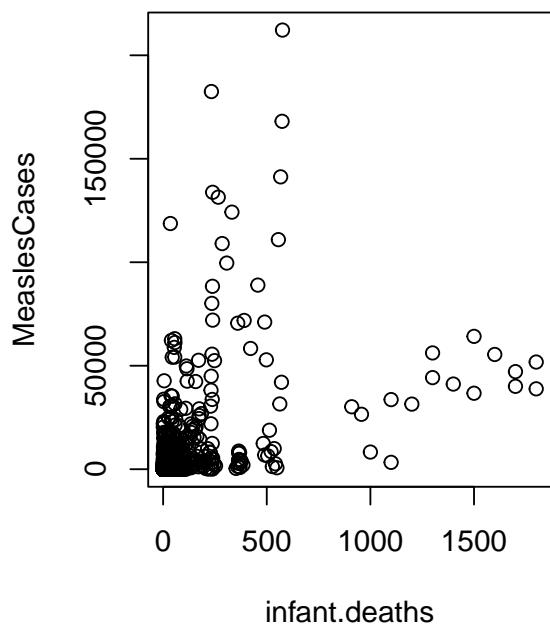
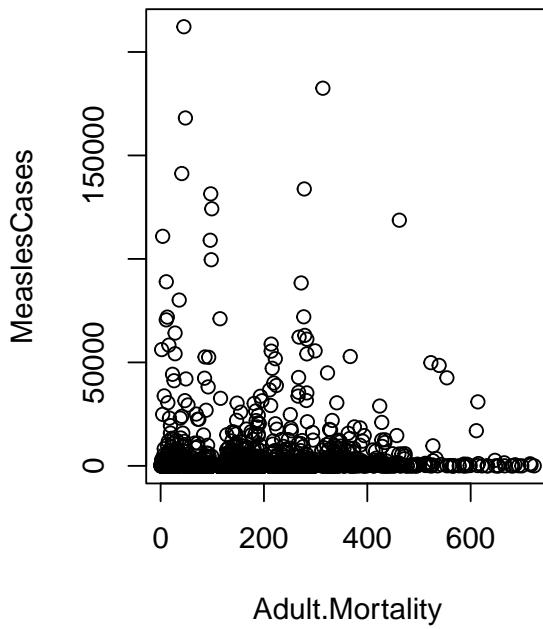
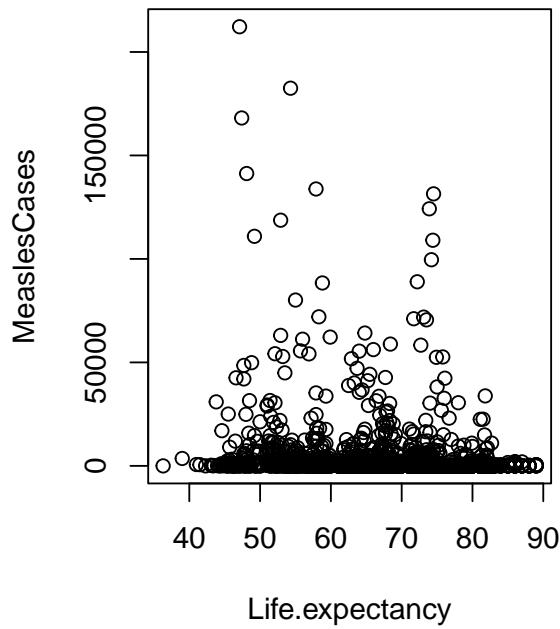


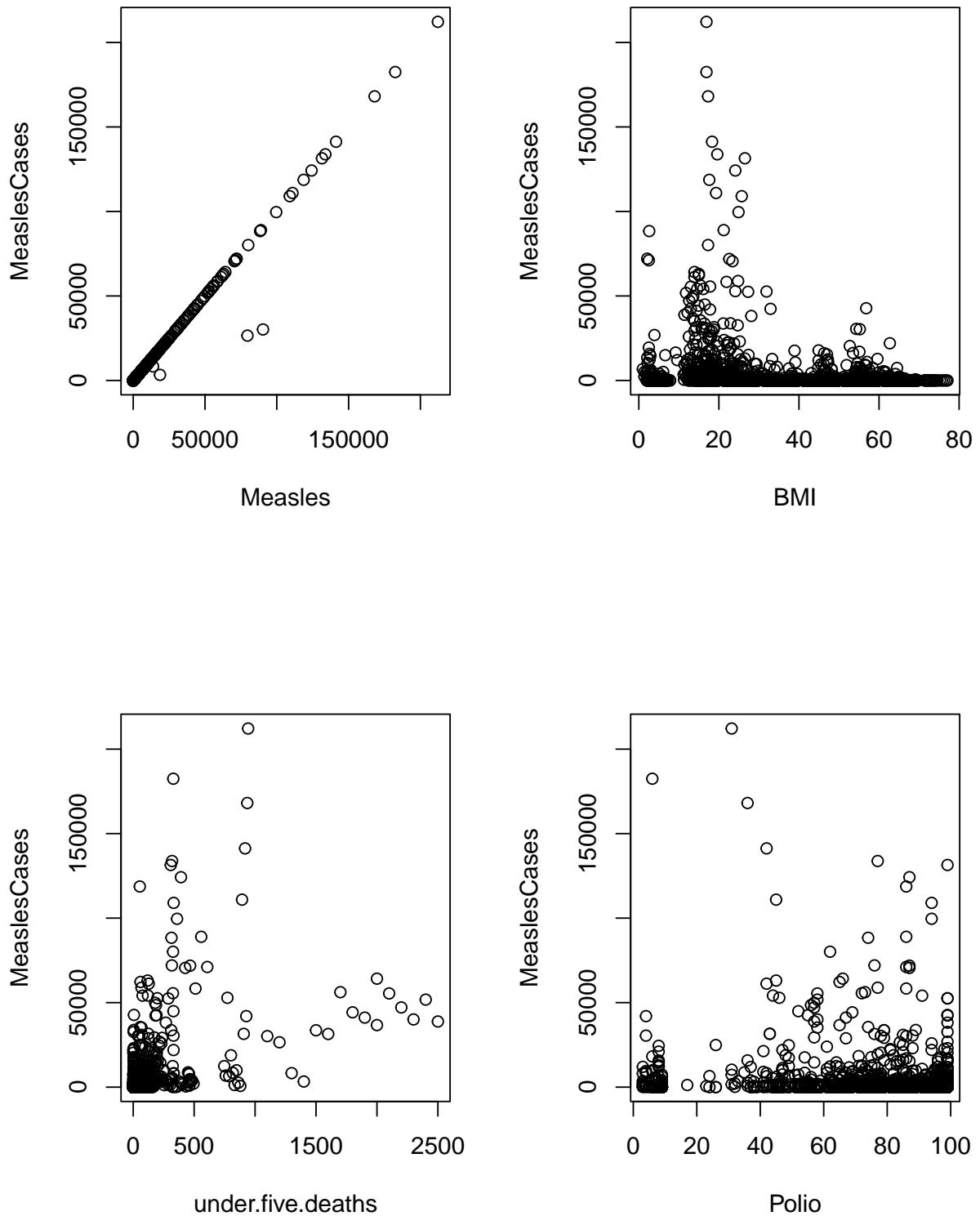


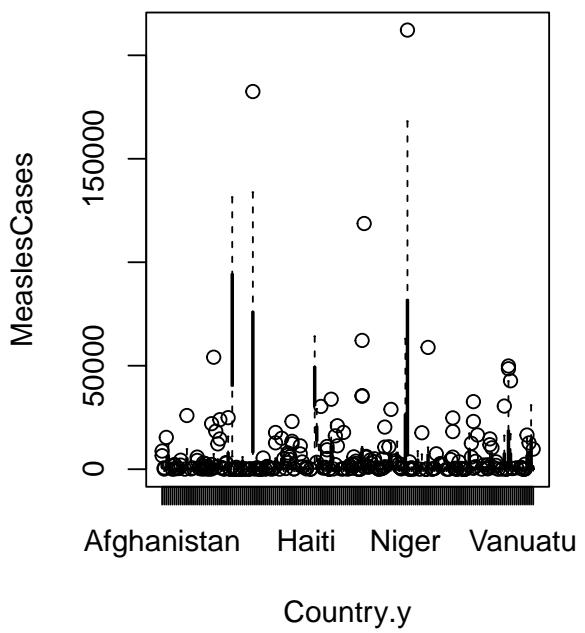
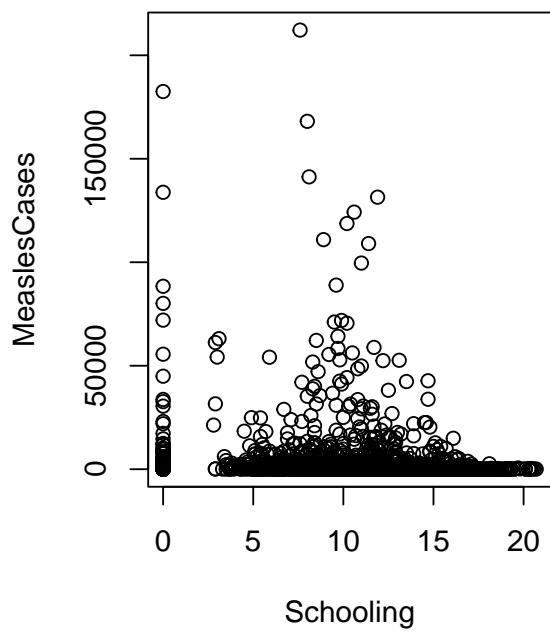
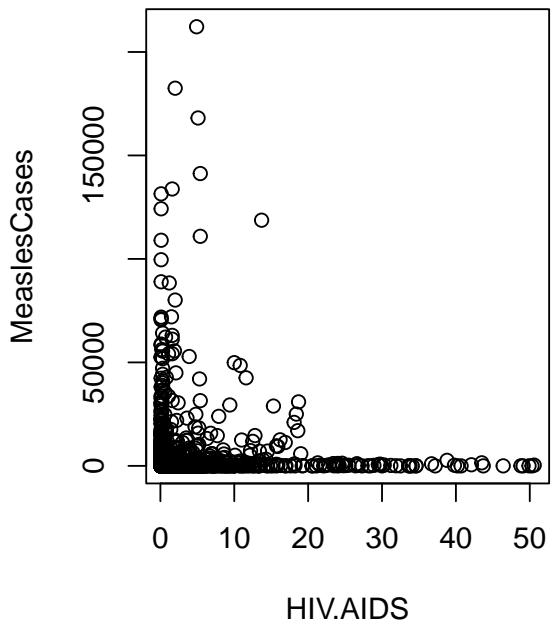
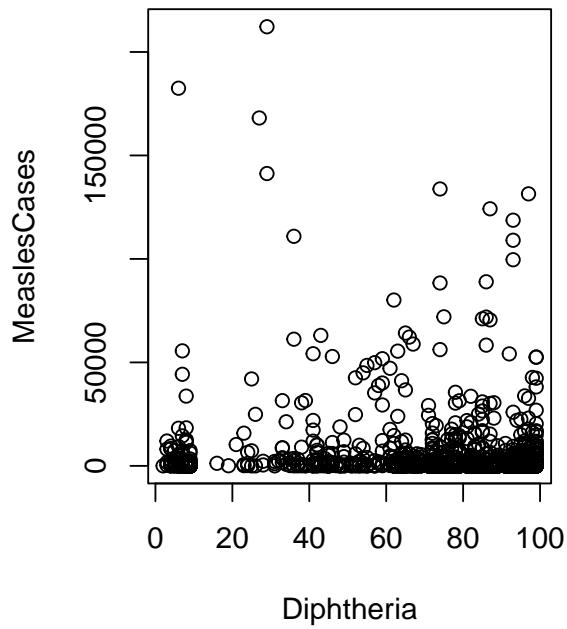
### Measles vs every other attribute

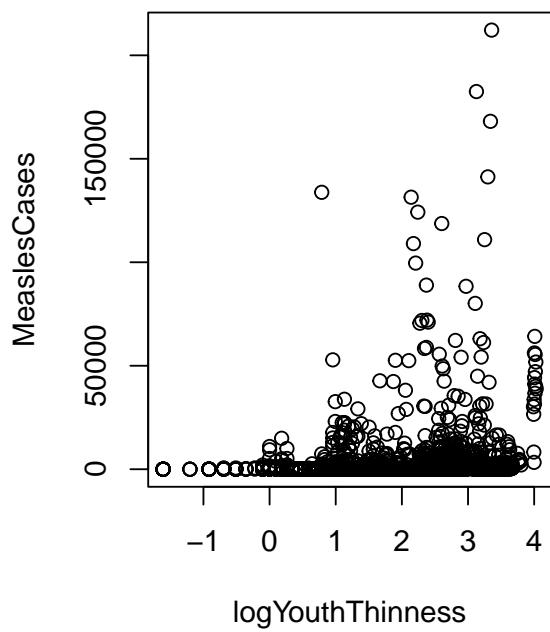
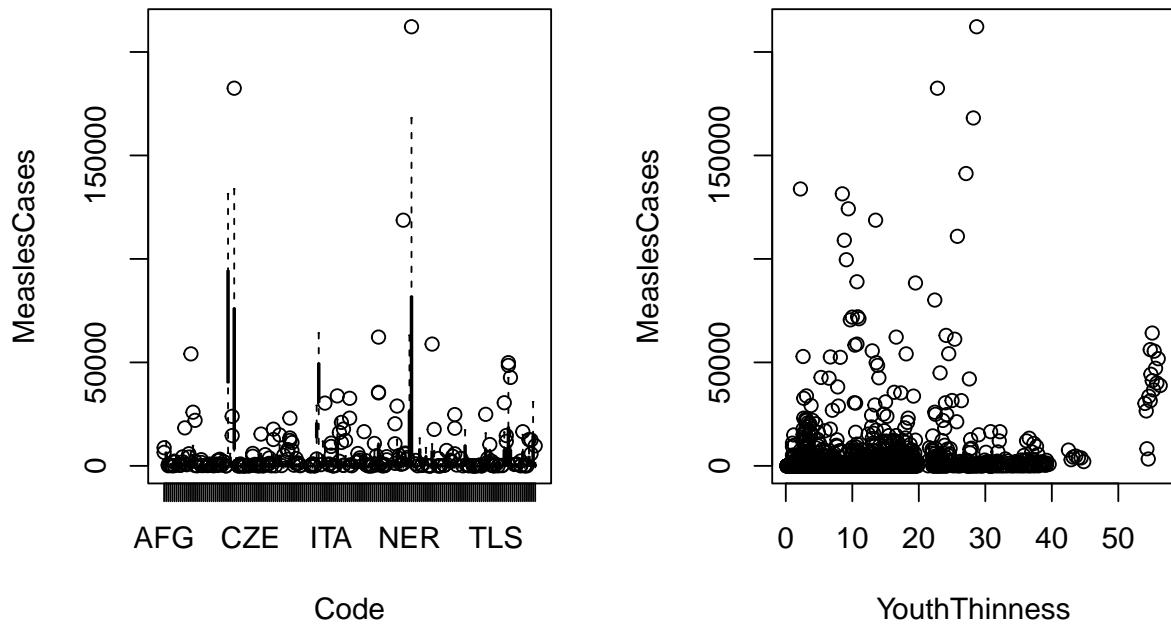
- Measles cases go down from 2000 to 20012
- There is definitely a significant difference of measles cases between developed and developing countries. more cases in developing countries.
- **Now we can do an anova to see if the difference is statistically significant.**
- life expectancy and adult mortality do not seem to have a pattern
- Infant deaths do seem to have a linear relationship with measles cases
- Polio immunization vs Measles cases have an interesting graph it does look like a somewhat negative relationship which would make sense.
- Diphtheria immunization vs Measles cases have an interesting graph it does look like a somewhat negative relationship as well
- **Do box plots of cases per country**
- YouthThinness vs Measles seem to have somewhat a linear relationship
- **What do you all think?**





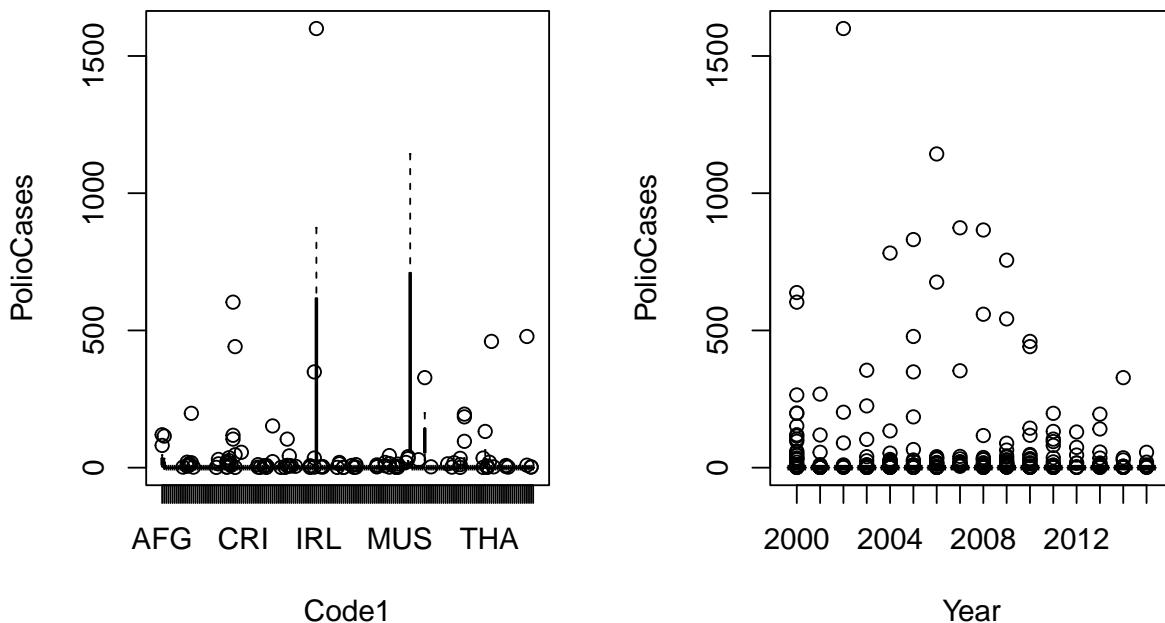


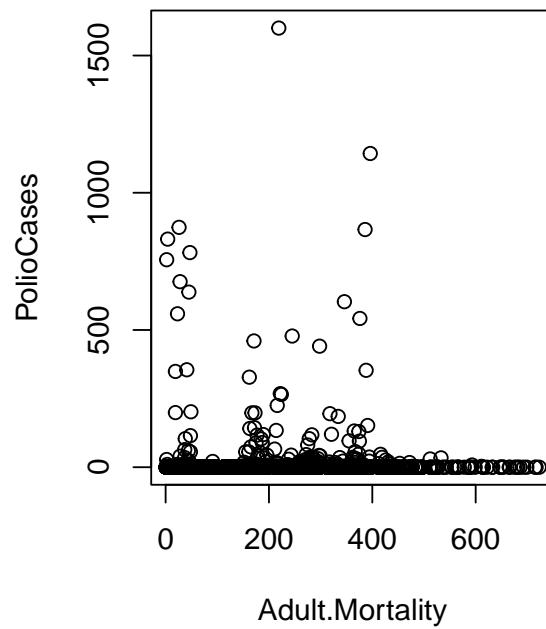
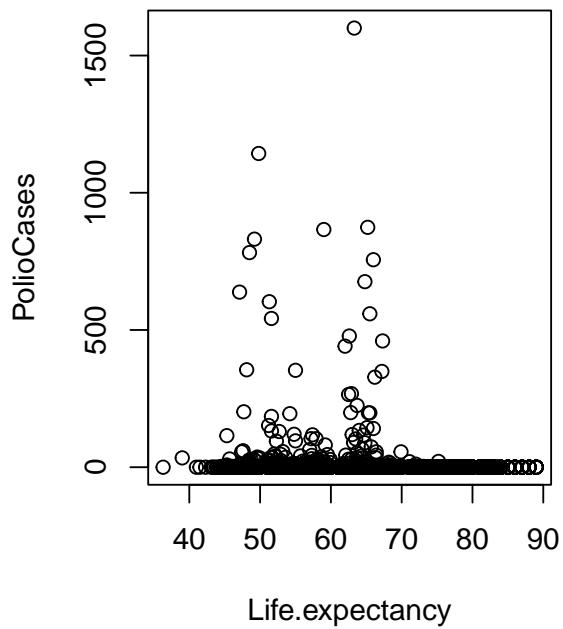
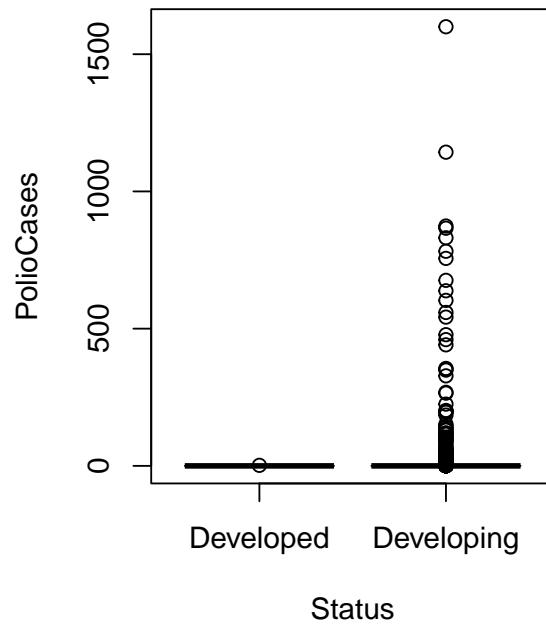
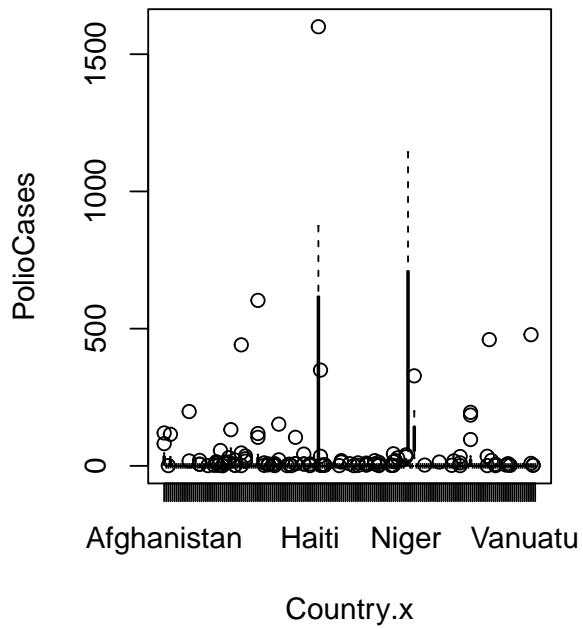


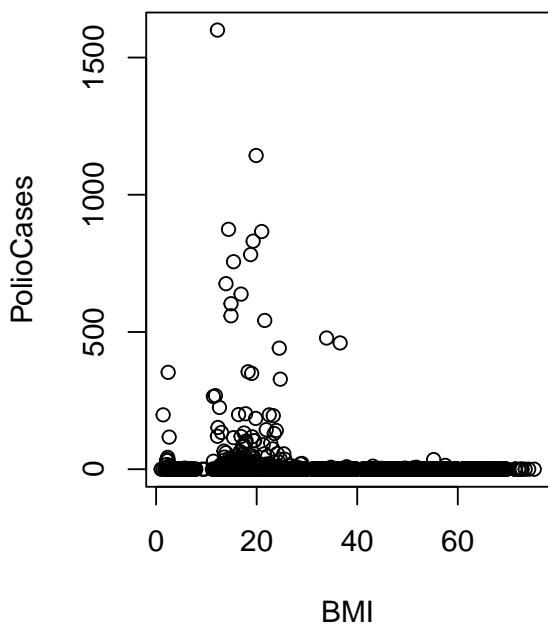
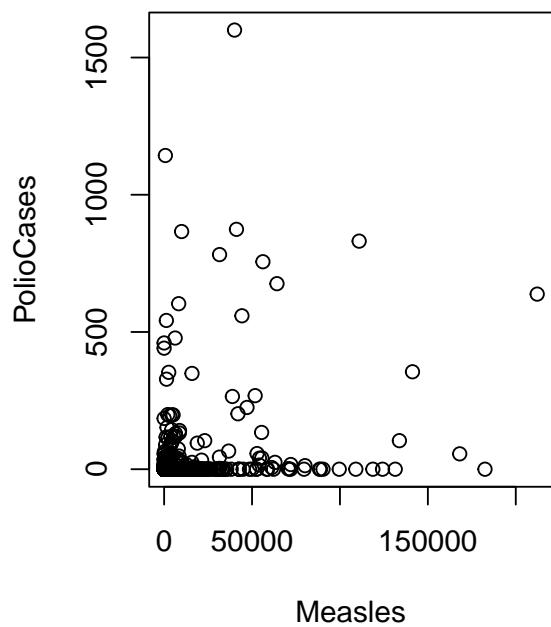
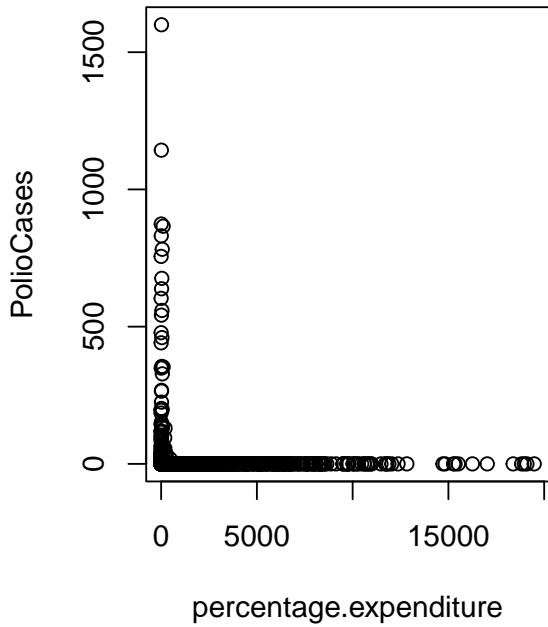
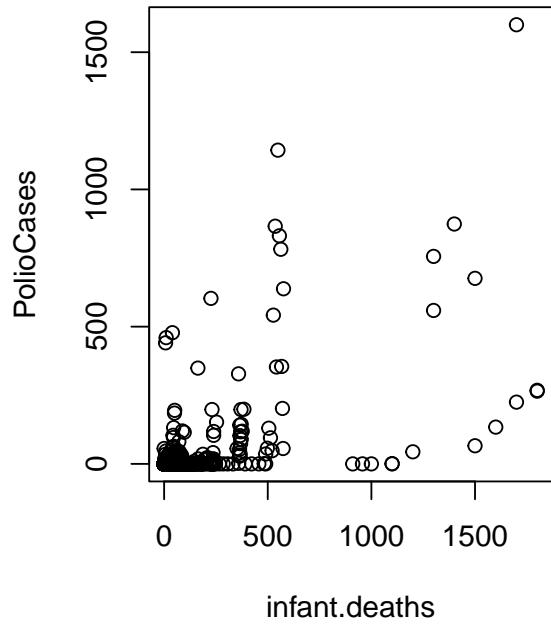


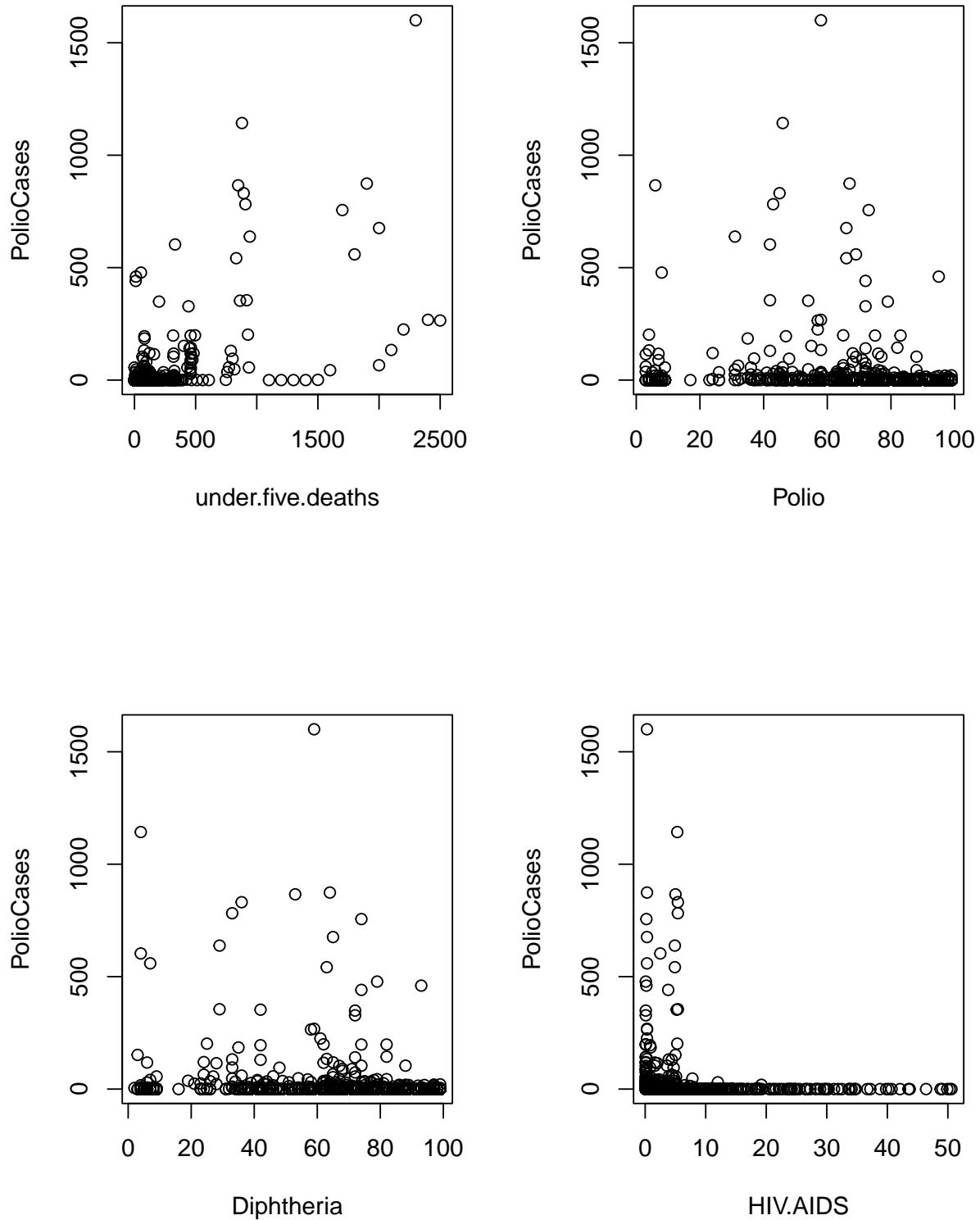
## Polio vs all other attributes

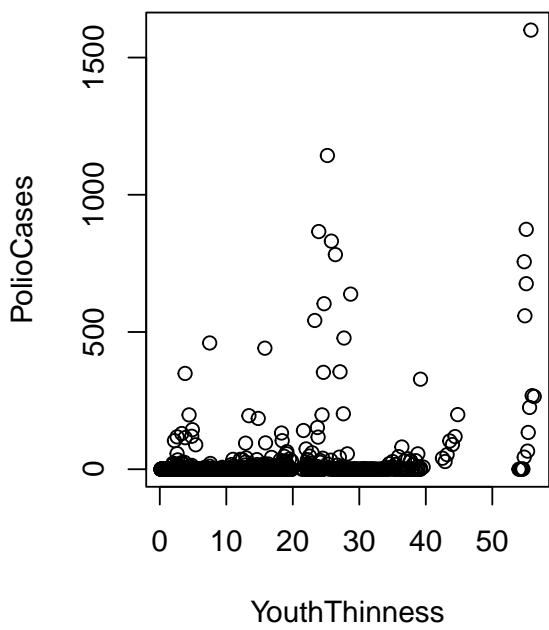
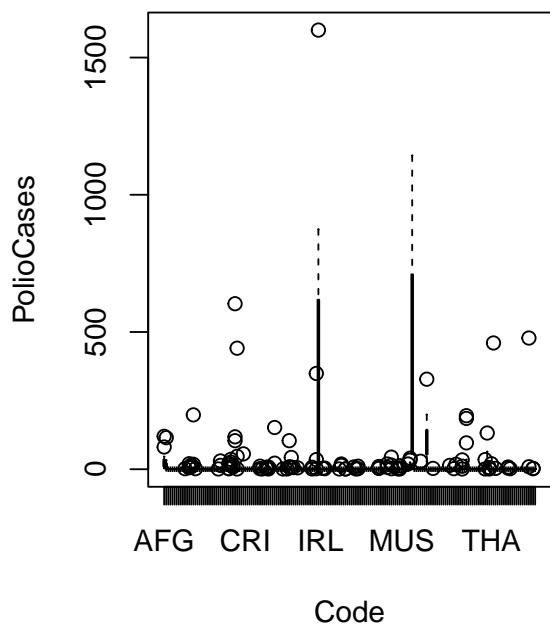
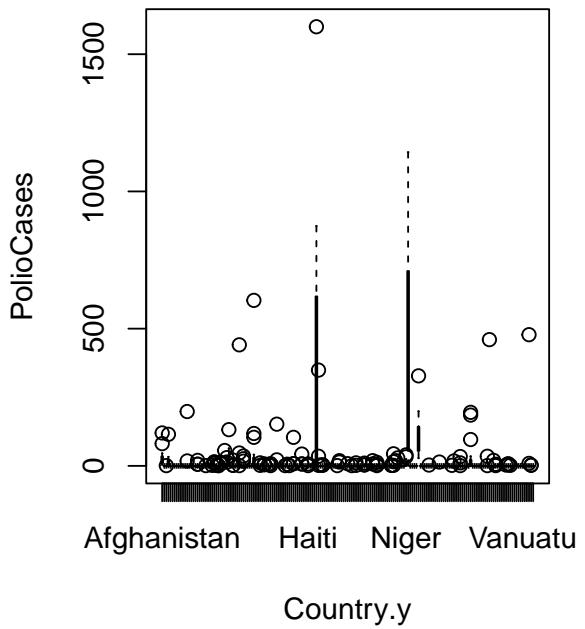
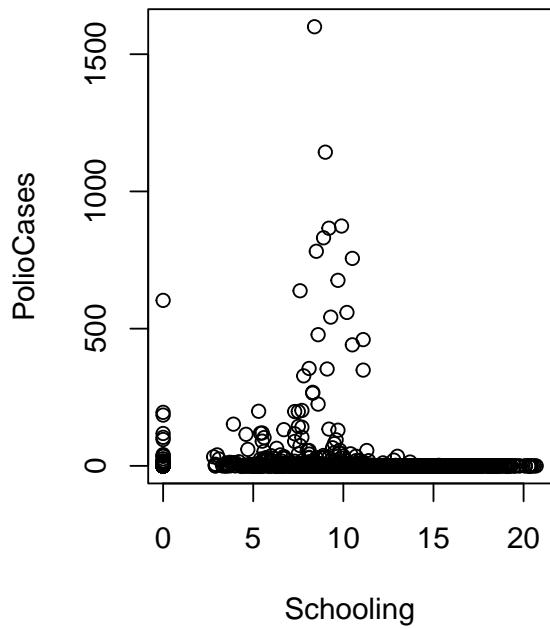
- Polio cases do seem to go down from 2000 to 2015 but slower decrease than that observed in measles
- Polio seems to have almost no cases in Developed vs developing!
- **create more graphs to get deeper insights**
- \*\* looking online “Polio is transmitted through contaminated water and food or contact with an infected person.” which would explain the developed vs developing graph
- Nigeria and Haiti seem to have the highest cases per Country y
- The graph for code has IRL and MUS as the highest cases
- **Definitely more graphs on countries for all diseases, as the graphs do noe include all countries**





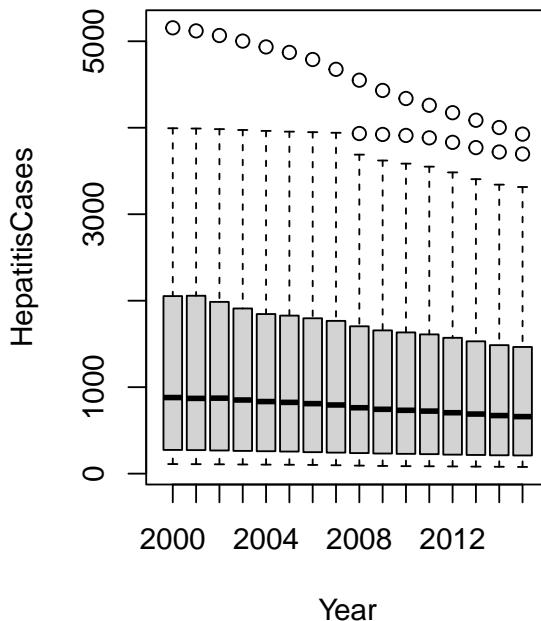
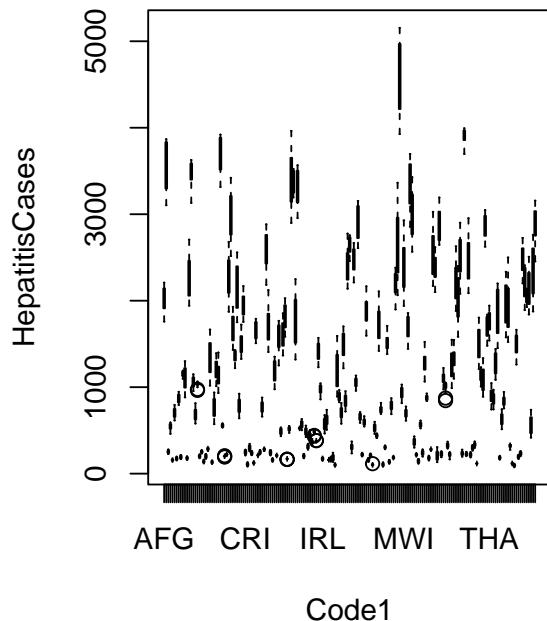


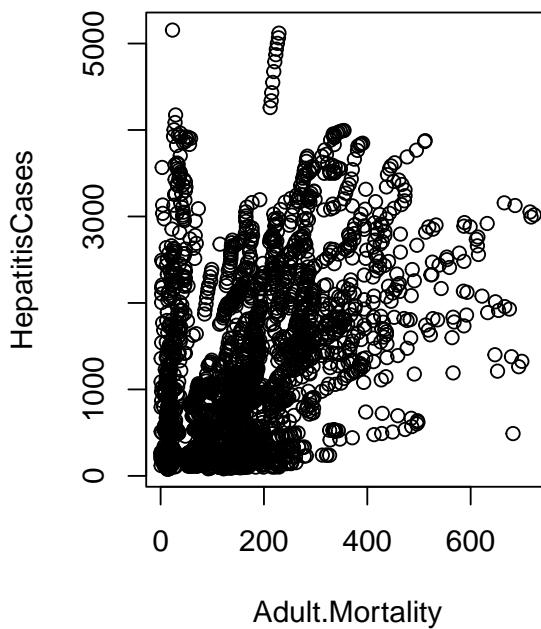
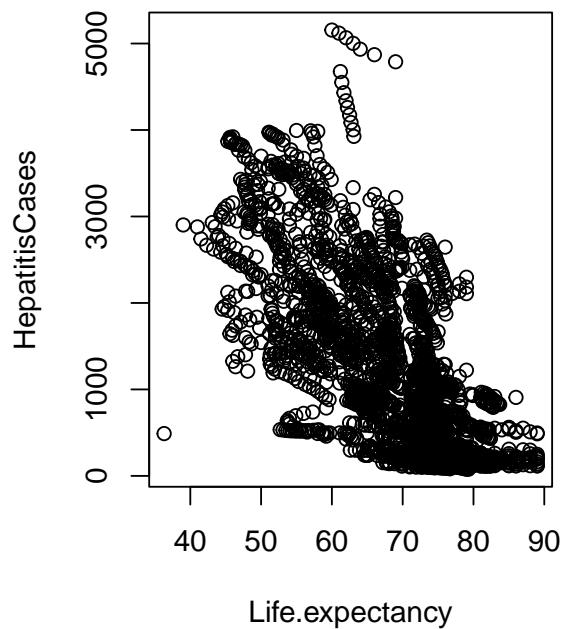
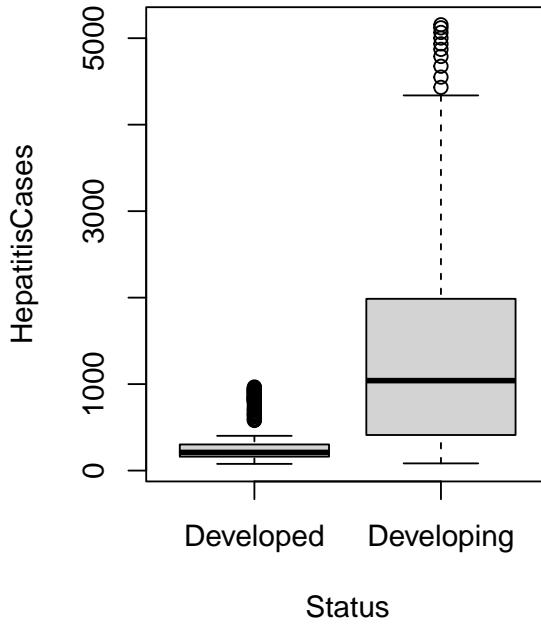
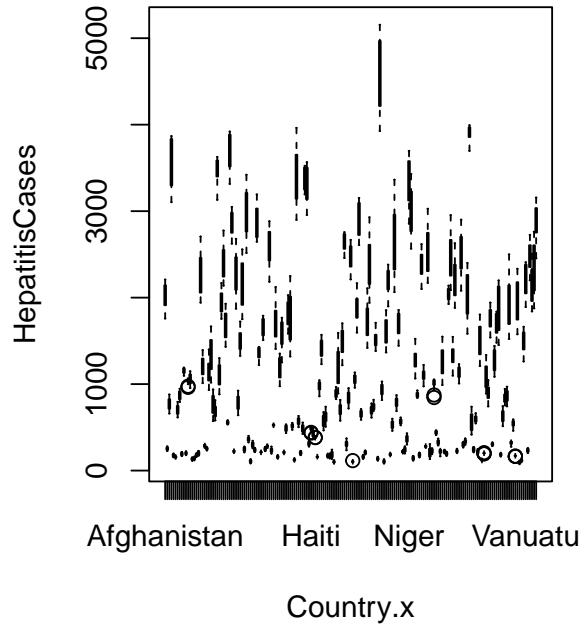


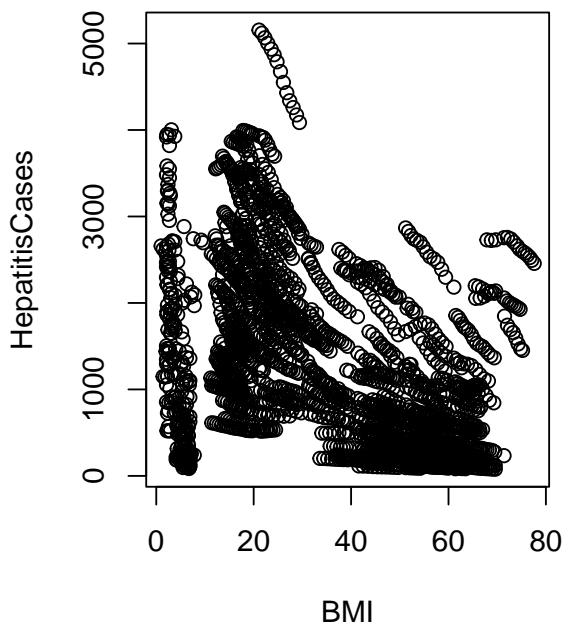
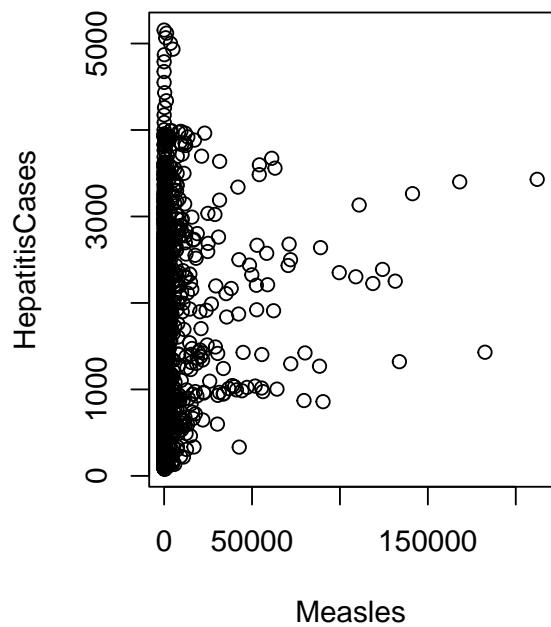
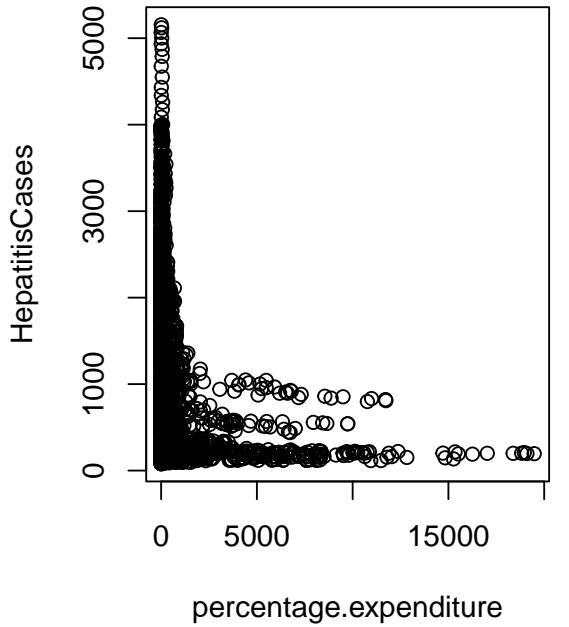
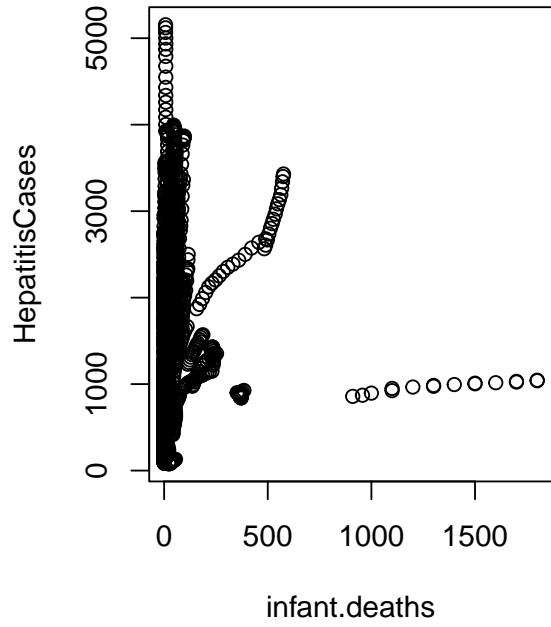


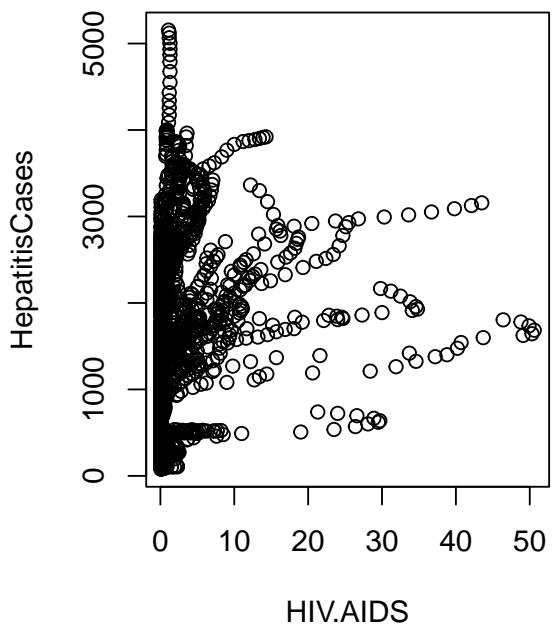
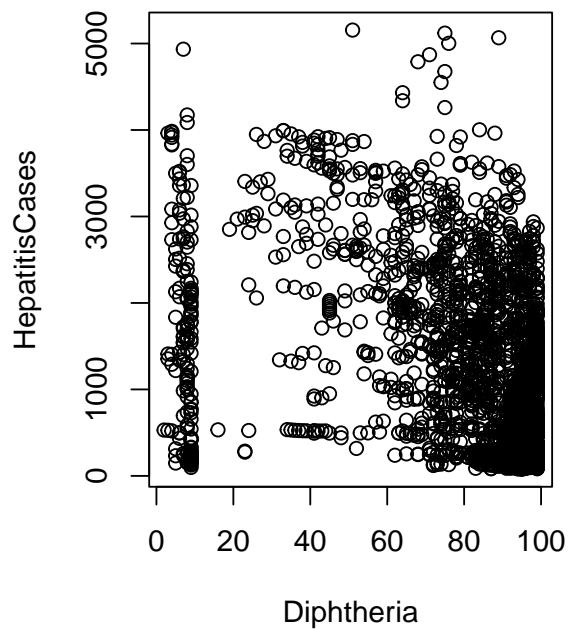
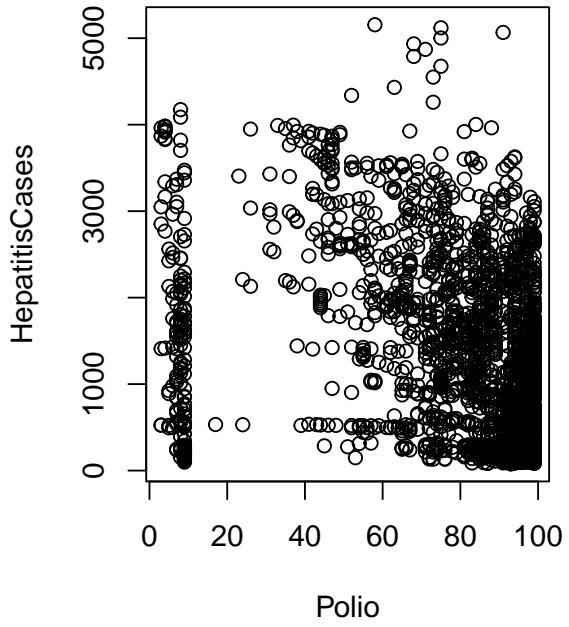
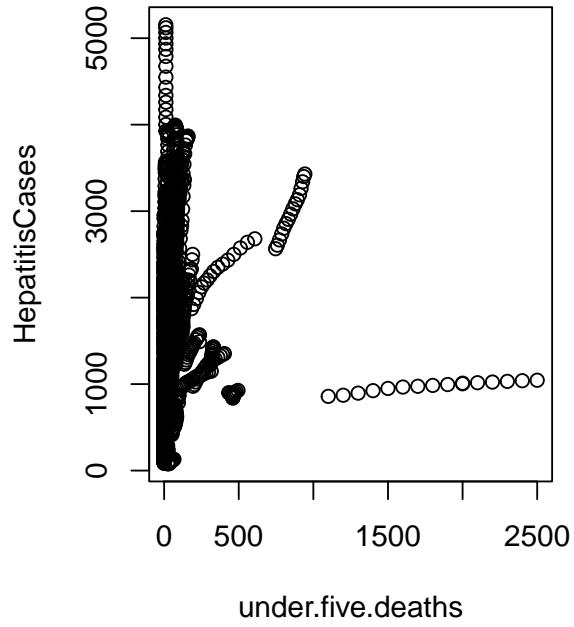
## HepatitisB vs all other attributes

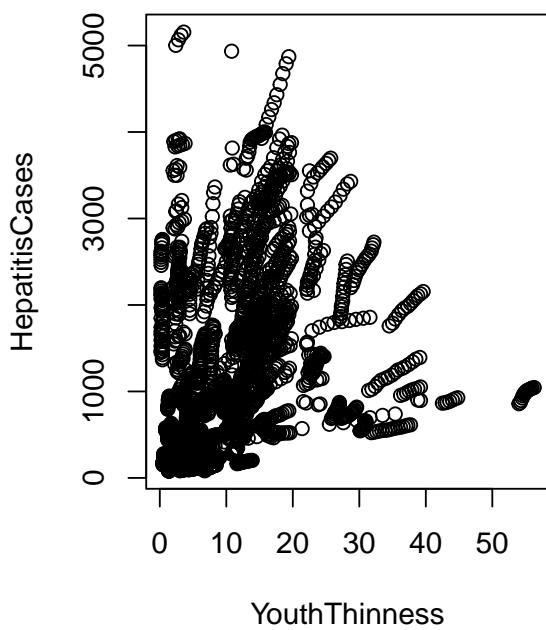
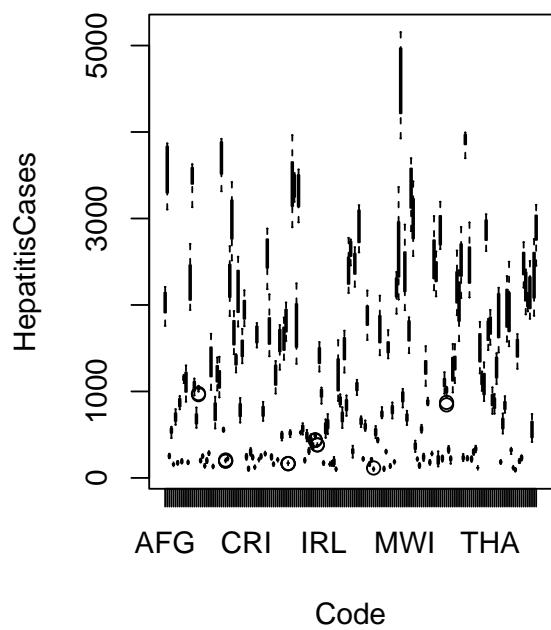
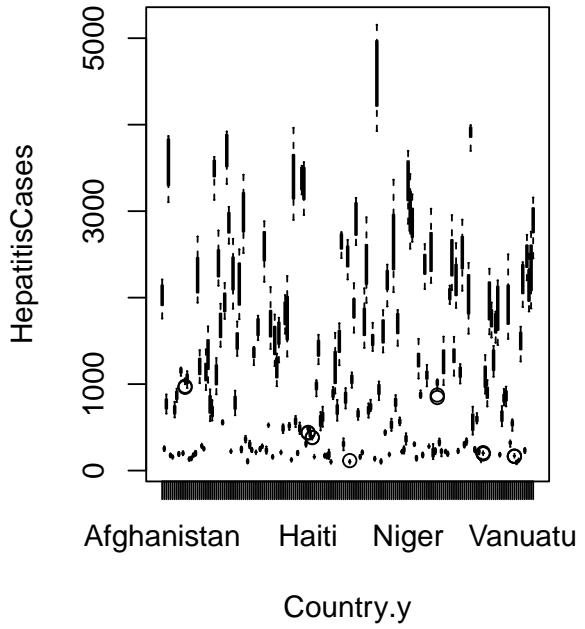
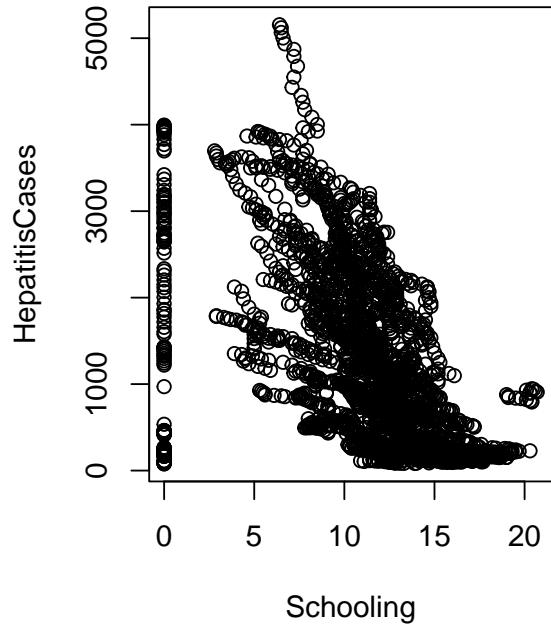
- There is definitely a decrease in hepatitis cases from 2000 to 2018. The mean decreases and so does the standard deviation
- We see again a different mean in developed vs developing. A mean of about 200 in developed to a mean of about 1000 in developing.
- **create a anova to see if it is also statistically significant?**
- Adult mortality and hepatitis do have a linear relationship!
- there is also a negative relationship between life expectancy and measles. The more measles cases the lower the life expectancy the lower or no measles cases lead to higher life expectancy.
- there is also a negative relationship with BMI.
- **there seems to be a relationship, positive, between HIV and hepatitis. we might need to look into it**
- there is a negative relationship between diphtheria and Hepatitis
- **we could look into it**
- schooling vs Hepatitis b have a negative relationship
- youth thinness does have a positive relationship with hepatitis which makes sense





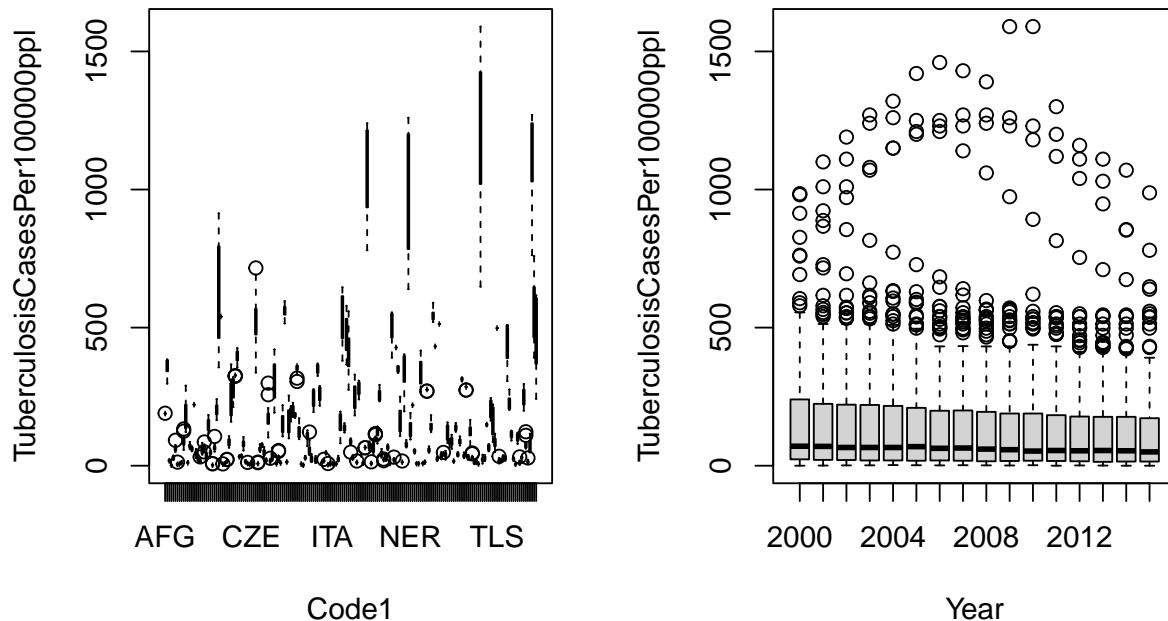


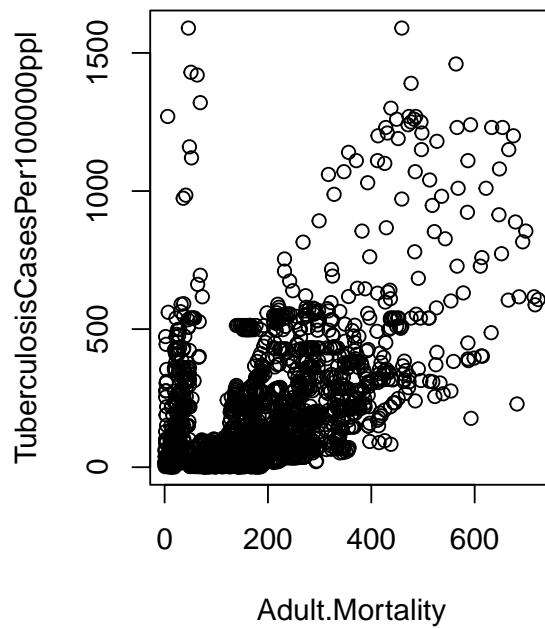
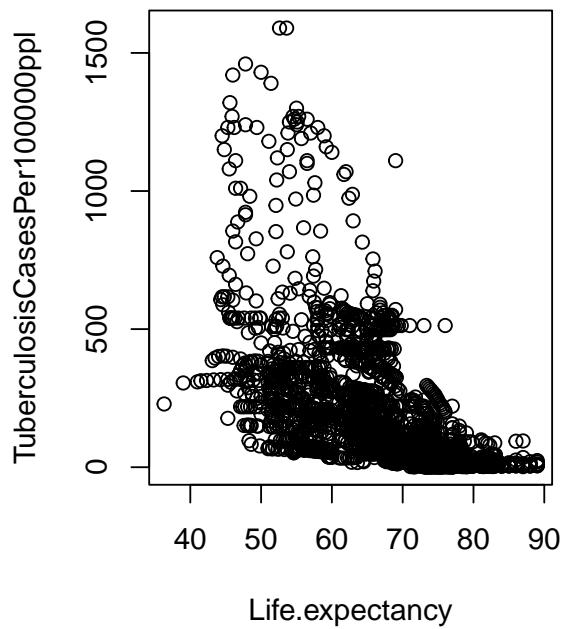
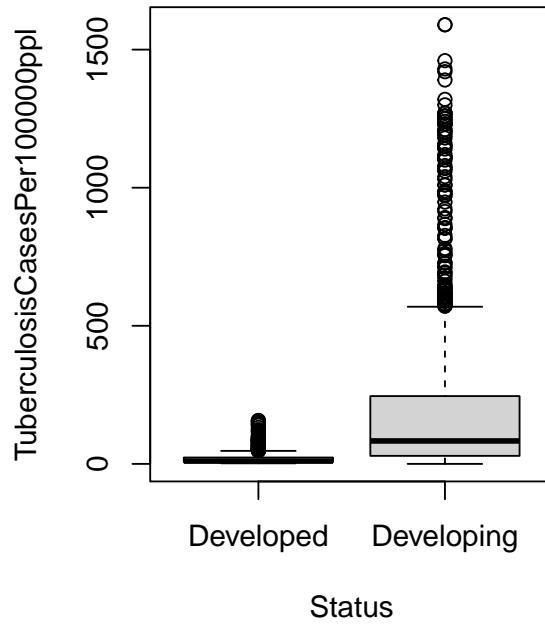
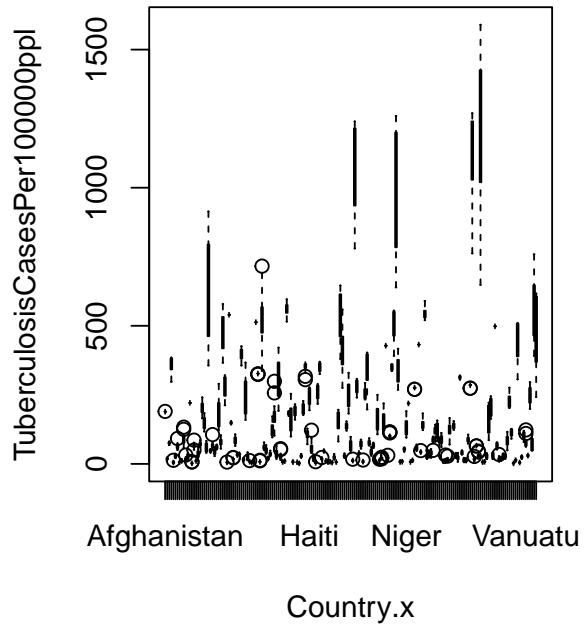


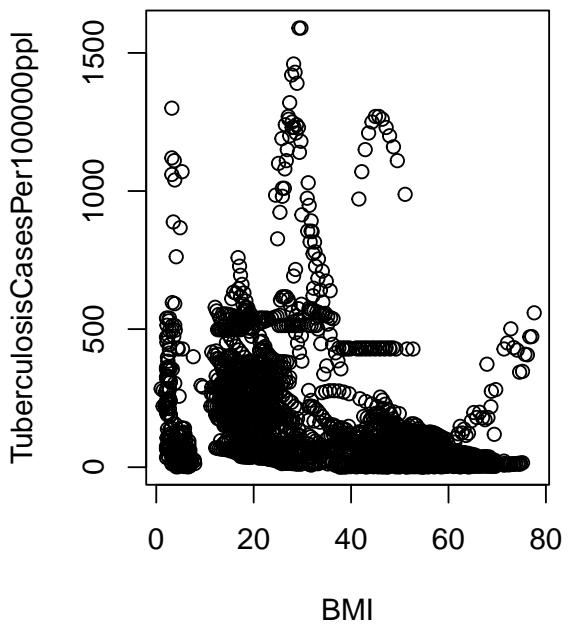
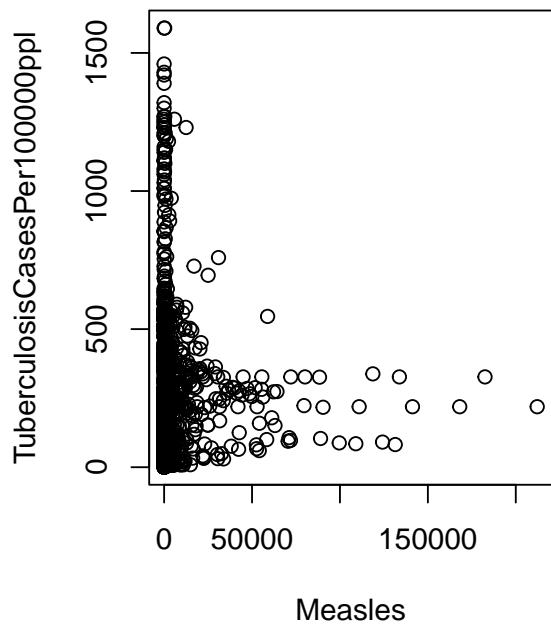
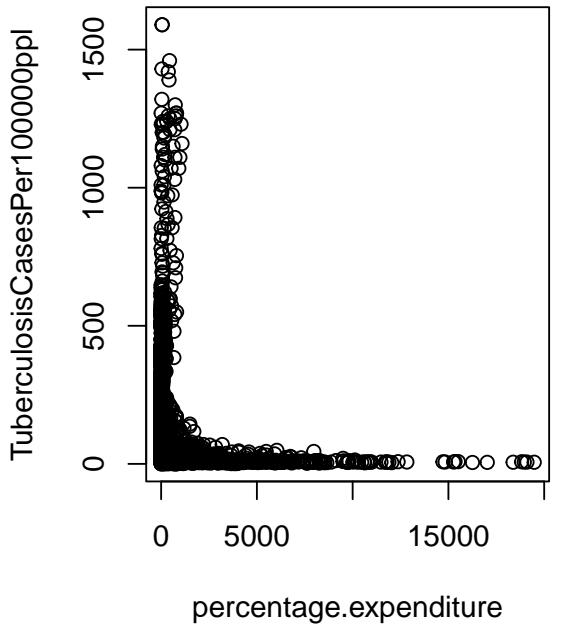
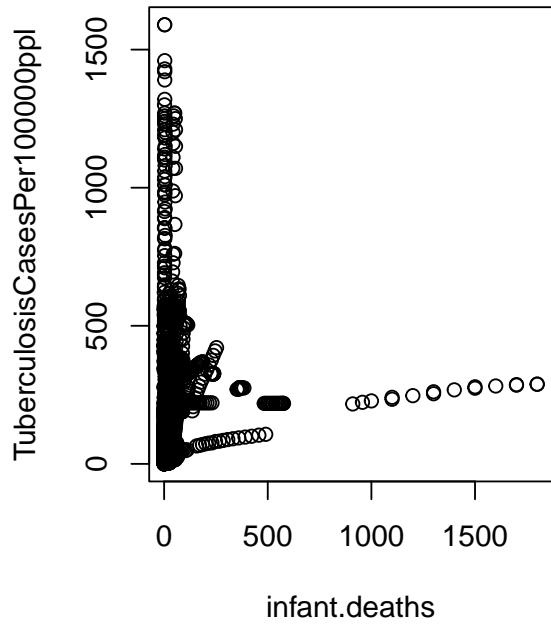


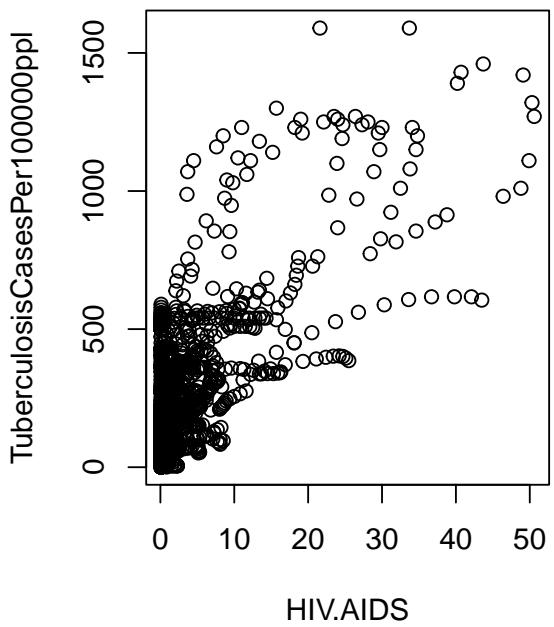
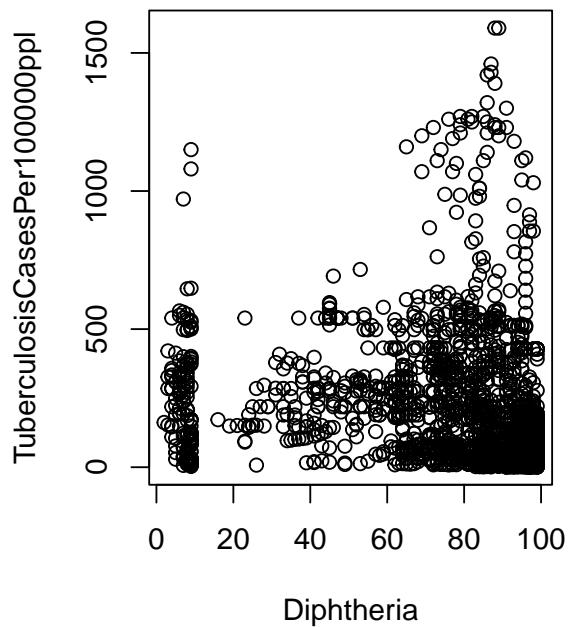
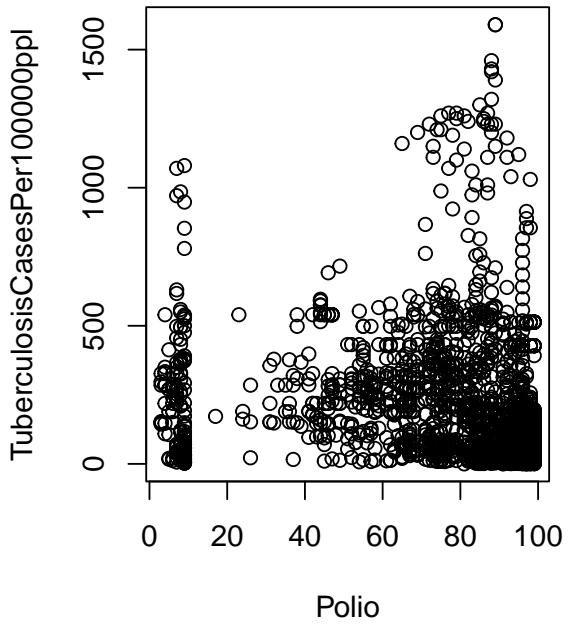
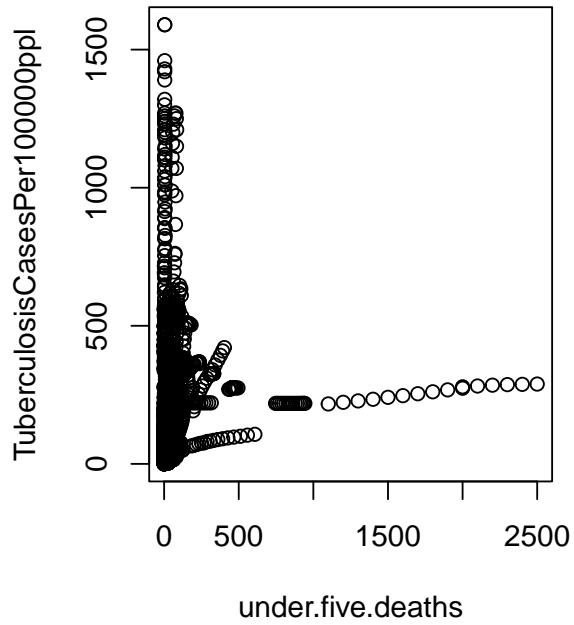
## Tuberculosis vs all other attributes

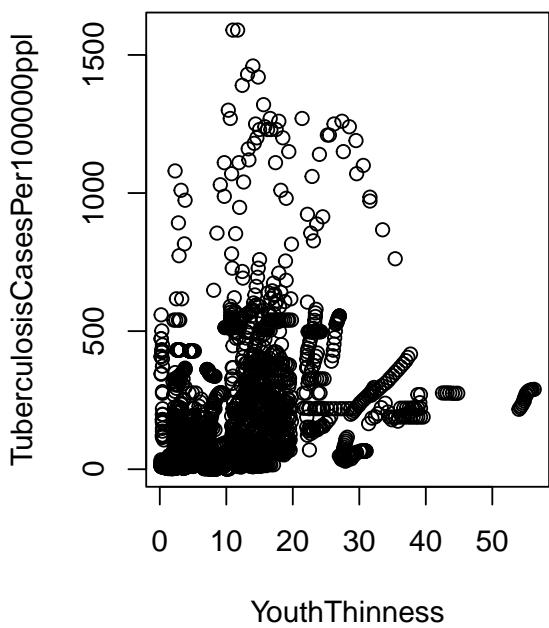
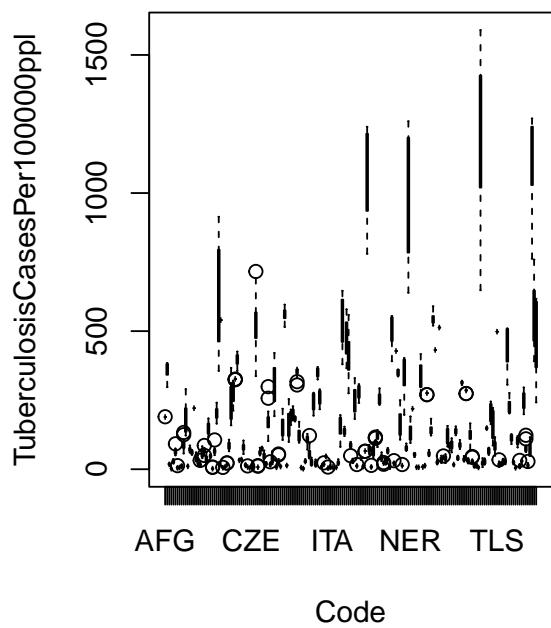
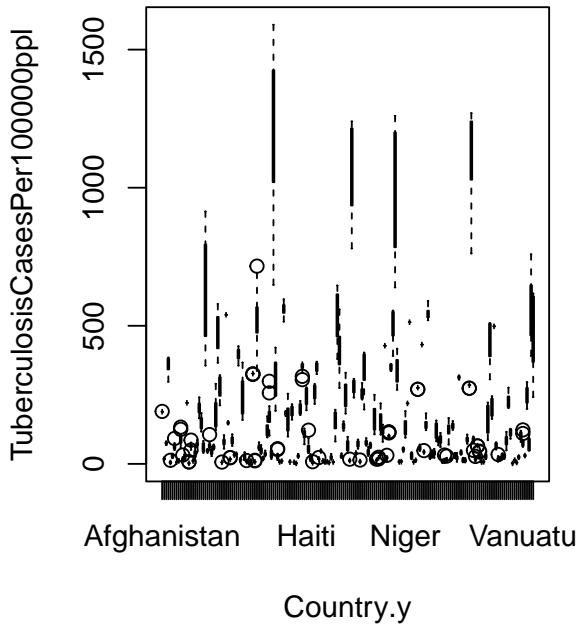
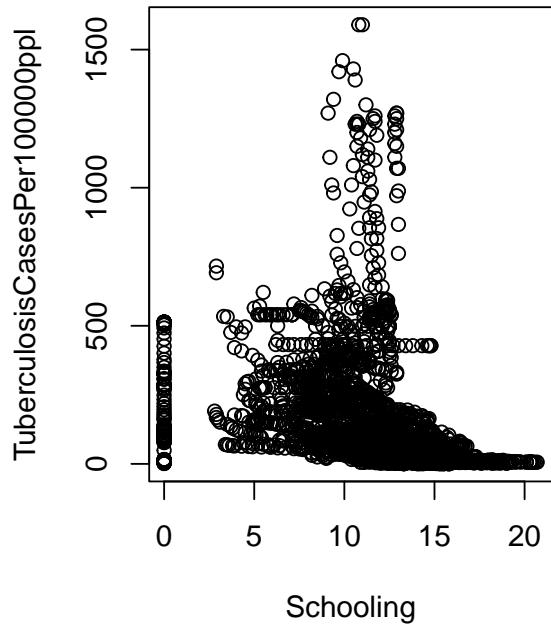
- There is no clear decline of mean deacreas on Tuberculosis per 100000 people crom 200 to 20014.however the range does decreases through the years slightly.
- **need to look if there is a vaccine for tuberculosis or more info on it**
- Developed vs developing has a huge difference the mean in developed is very small. Almos 5 in developed compared to about 20 in developing which is not that much but it has very large outlier and bigger standar devaiation.
- Life Expectancy has a negative relationship with tuberculosis
- adult mortality has a positive relationship with tuberblosis. the more cases he more changes of dying earlier in age.
- **what do you all think of the infant deaths? there seems to be somwhat a small positive relationshio..**
- no relationship with BMI
- inmunitzation of polio vs tuberculosis seem to have a negative reationship
- same for diphtheria nagative relationship with tubercuosis.
- **we might want to look into that**
- Also interesting positive relationship with HIV.AIDS
- **We should look into that.**



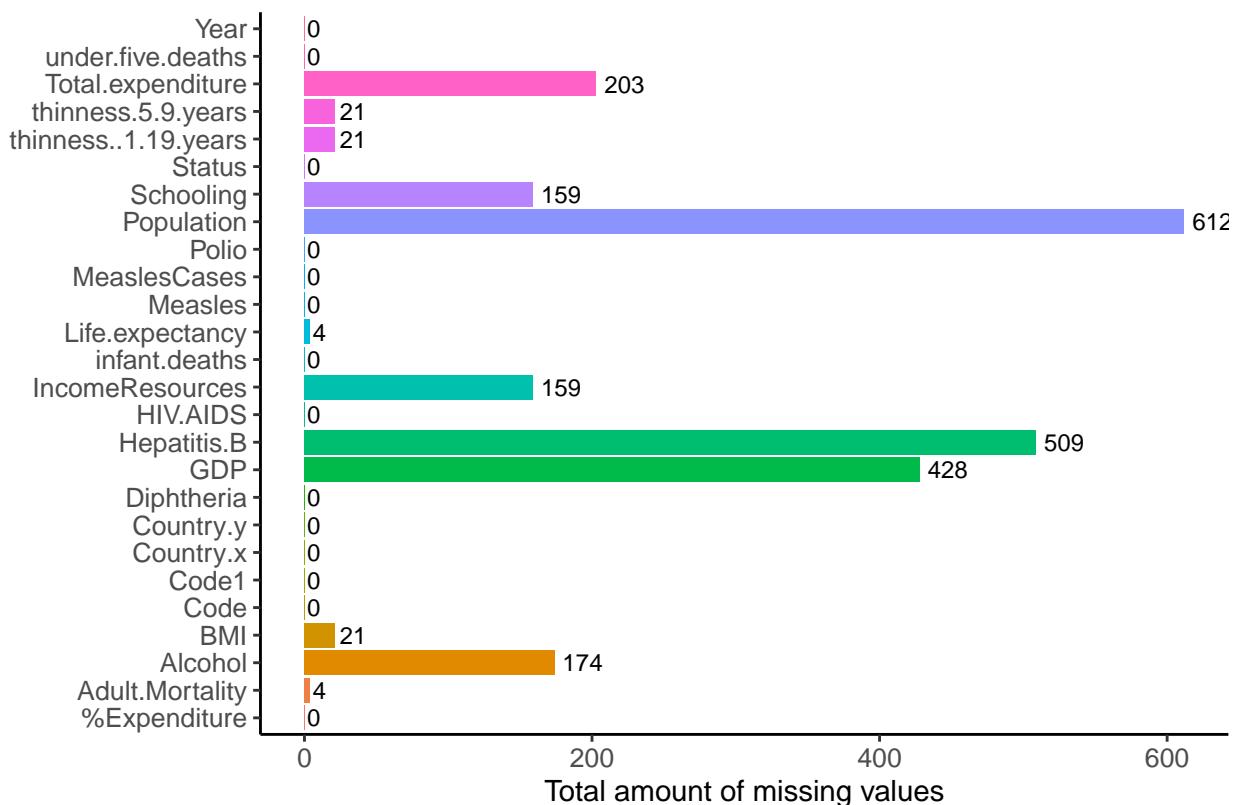




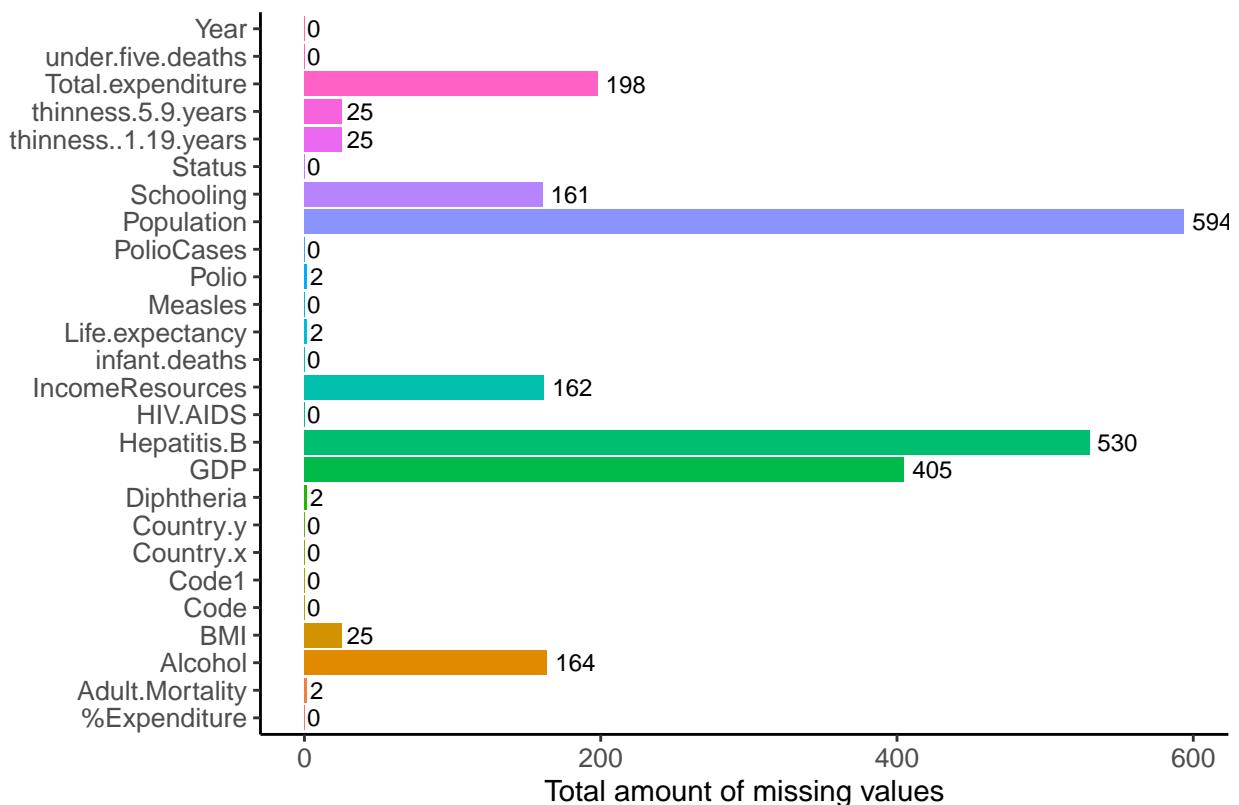




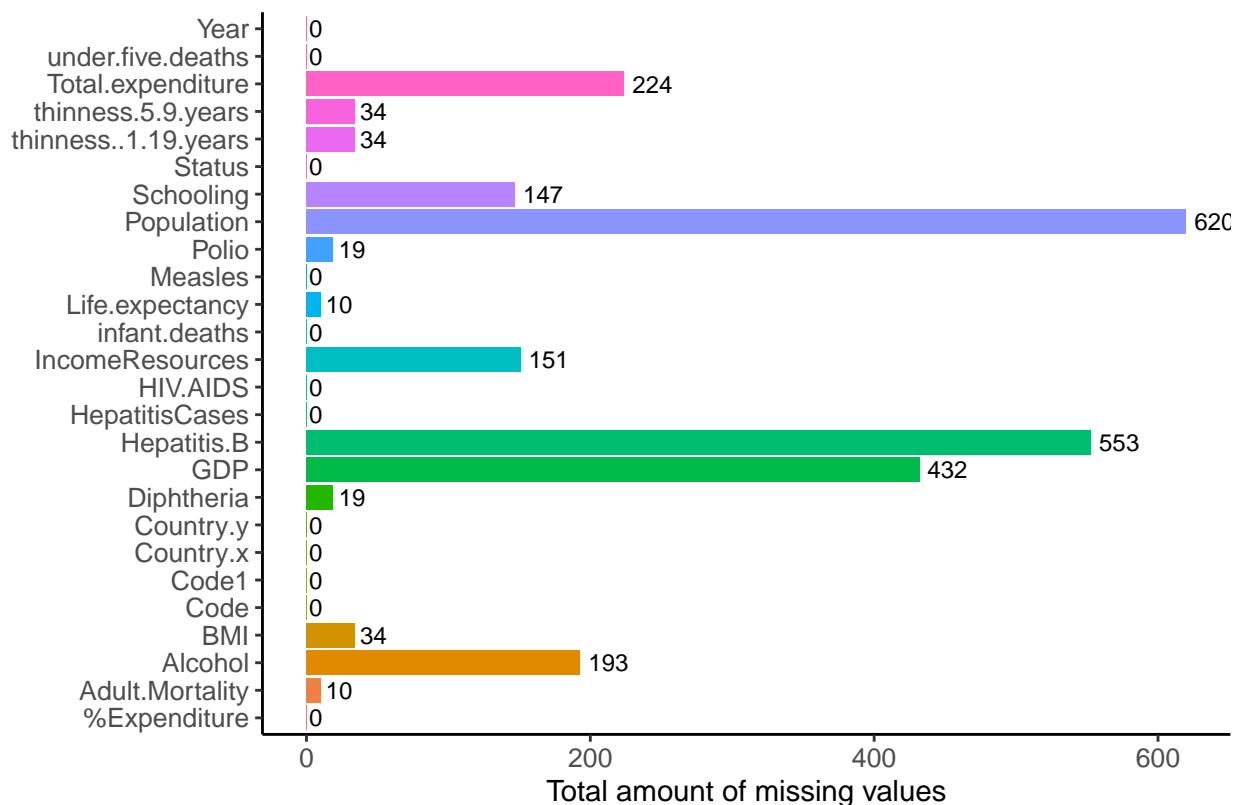
### Missing values per variable in Measles data



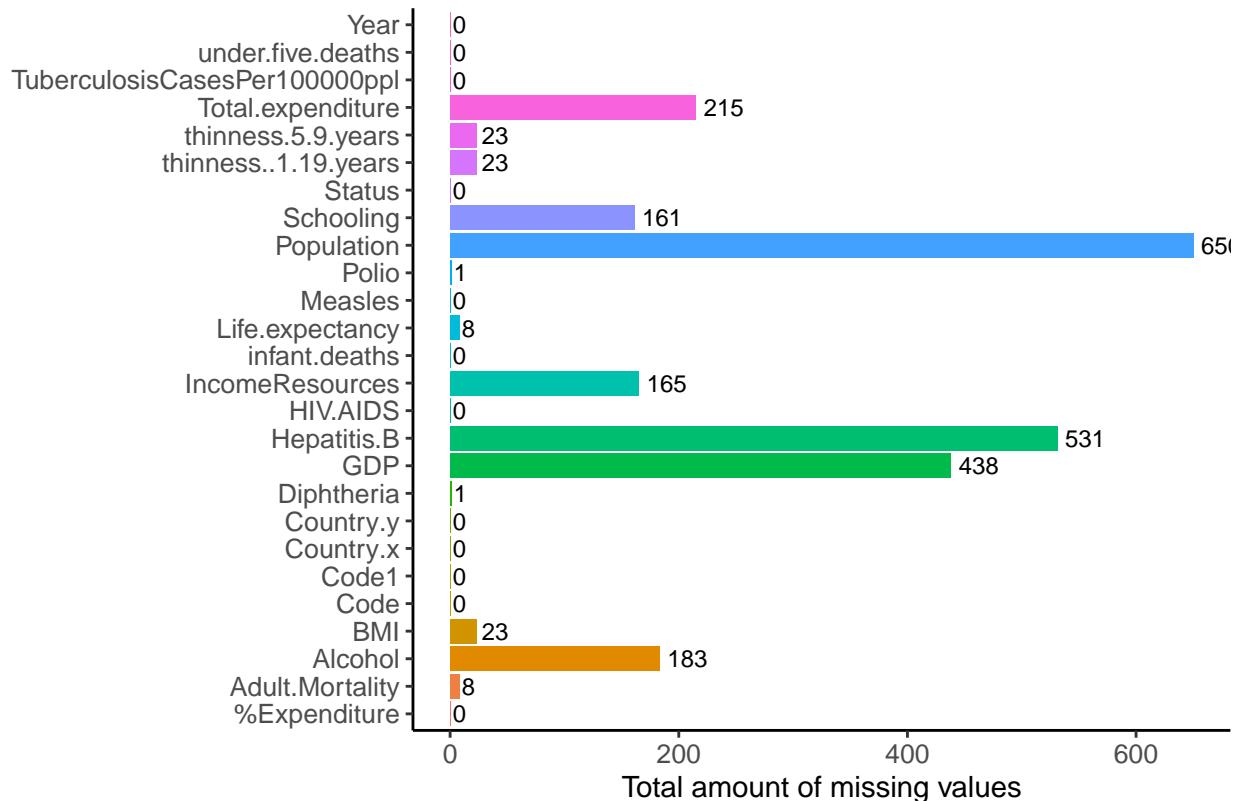
*Missing values per variable in Polio dataset*



*Missing values per variable in Hepatitis dataset*



*Missing values per variable in Tuberculosis*



## Developed vs developing

### Single data with total cases of each disease per country

- 133 countries without a specific disease
- 114 Polio and 19 Measles

Countries without measles cases Status n missing 19 0

Value Developed Developing Frequency 1 18 Proportion 0.053 0.947

Countries without polio cases Status.x n missing  
114 0

Value Developed Developing Frequency 30 84 Proportion 0.263 0.737

```
## `summarise()` has grouped output by 'Country.y'. You can override using the
## `.` argument.
## `summarise()` has grouped output by 'Country.y'. You can override using the
## `.` argument.
## `summarise()` has grouped output by 'Country.y'. You can override using the
## `.` argument.
## `summarise()` has grouped output by 'Country.y'. You can override using the
## `.` argument.
```

```

##          Country.y    Status.x      disease cases
## 1      Afghanistan Developing MeaslesCases 37796
## 2          Albania Developing MeaslesCases   854
## 3          Algeria Developing MeaslesCases 31102
## 4          Angola Developing MeaslesCases 56982
## 5 Antigua and Barbuda Developing MeaslesCases     0
## 6      Argentina Developing MeaslesCases    32

```

## Developed vs Developing

- The data is composed of 17% developed countries and 83% developing
- When filter the data with country disease that is 0 it changes 2%. It changes 15% developed and 85% developing
- All countries in the data had Tuberculosis
- **Out of the Developed countries 97% of the developed countries had Measles cases**
- **Out of the Developing countries 88% of the developing countries had Measles cases**
- **Out of the Developed countries 3% of the developed countries had Polio cases**
- **Out of the Developing countries 43% of the developing countries had Polio cases**
- **Both Developing and Developed Countries had Tuberculosis cases and HepatitisB**
- Now when it come to actual cases in developed vs developing
- Here is the table of tha actual difference and percentage which is actually not that much
- **tried to graph but having issue... need help**

```

## # A tibble: 4 x 4
## # Groups:  disease [4]
##   disease      Developed Developing percent
##   <fct>       <dbl>     <dbl>    <dbl>
## 1 MeaslesCases  254111    6660847   3.81
## 2 HepatitisCases 134293.   3064074.  4.38
## 3 PolioCases        2       19813   0.0101
## 4 TuberculosisCases 9992.   409470.  2.44

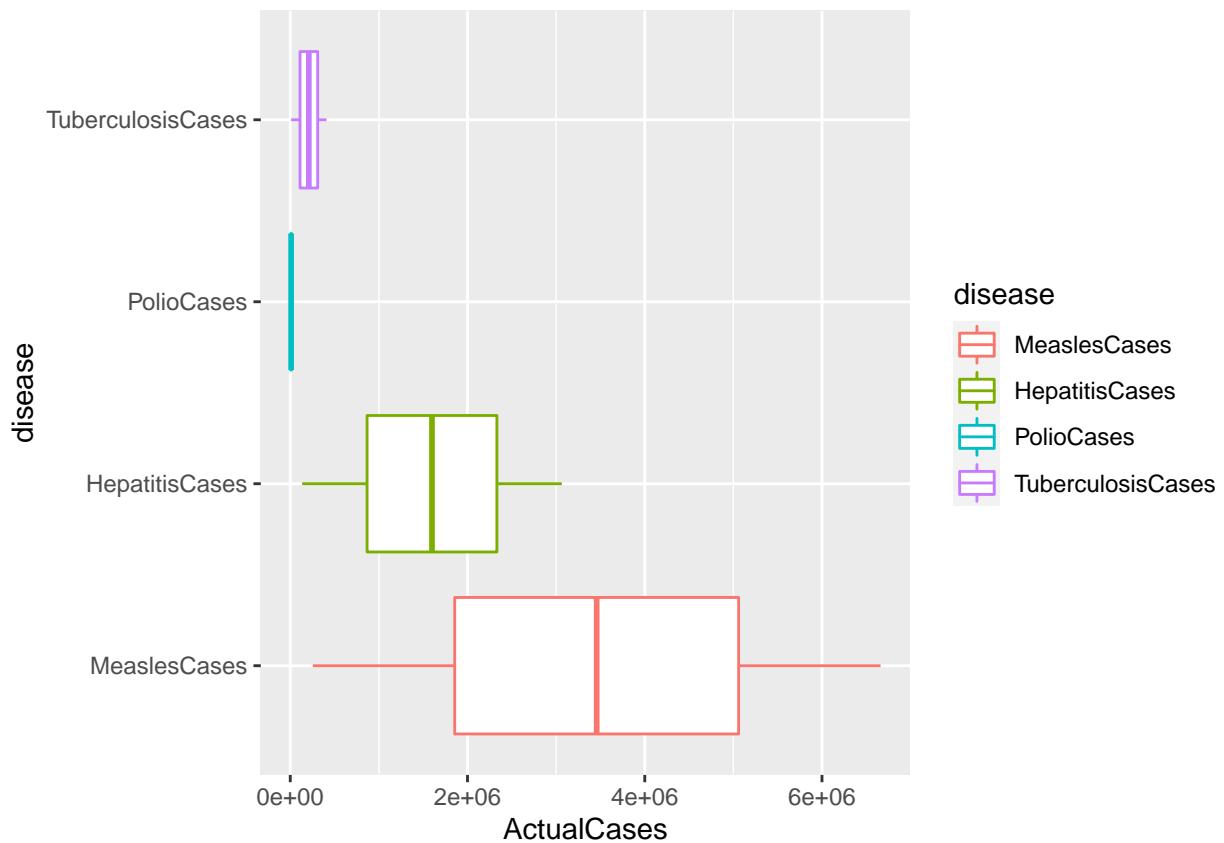
```

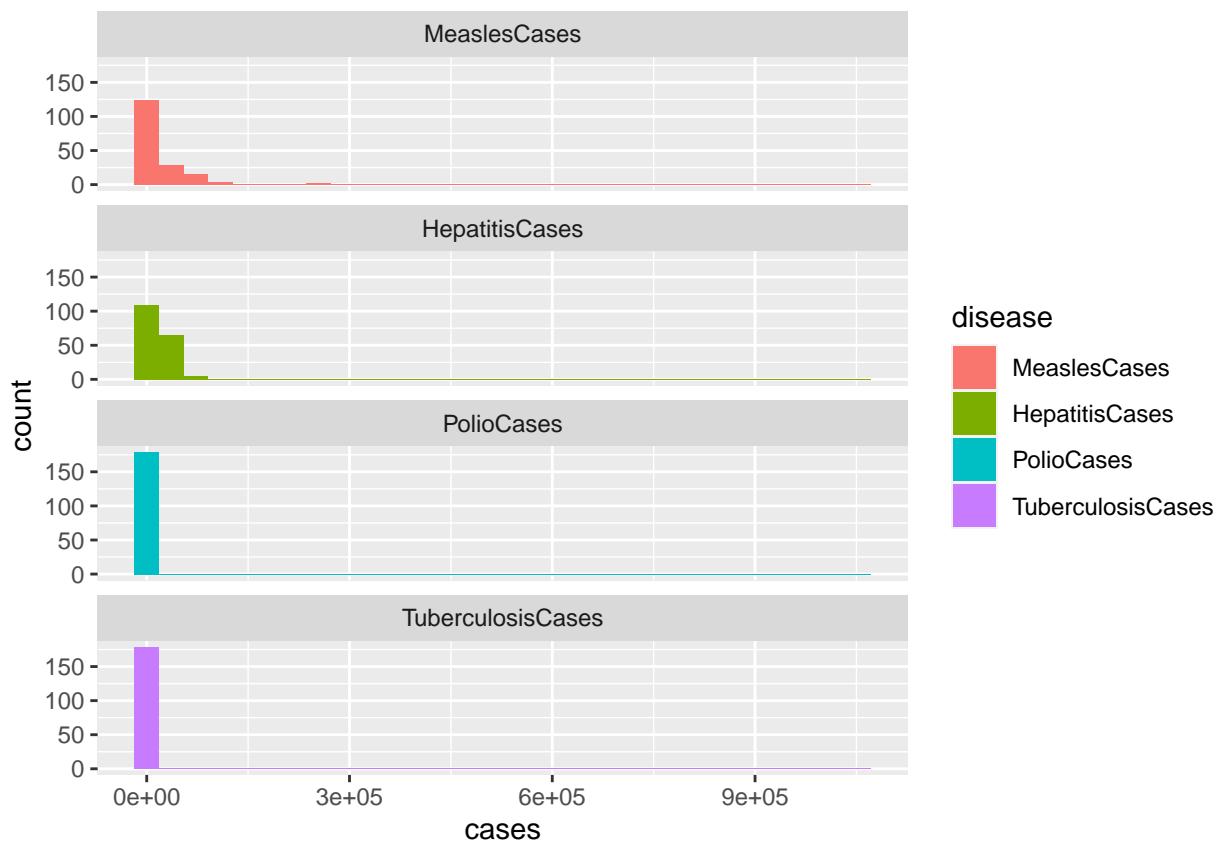
## Filter data on Cases vs Status vs Disease and some graphs on it

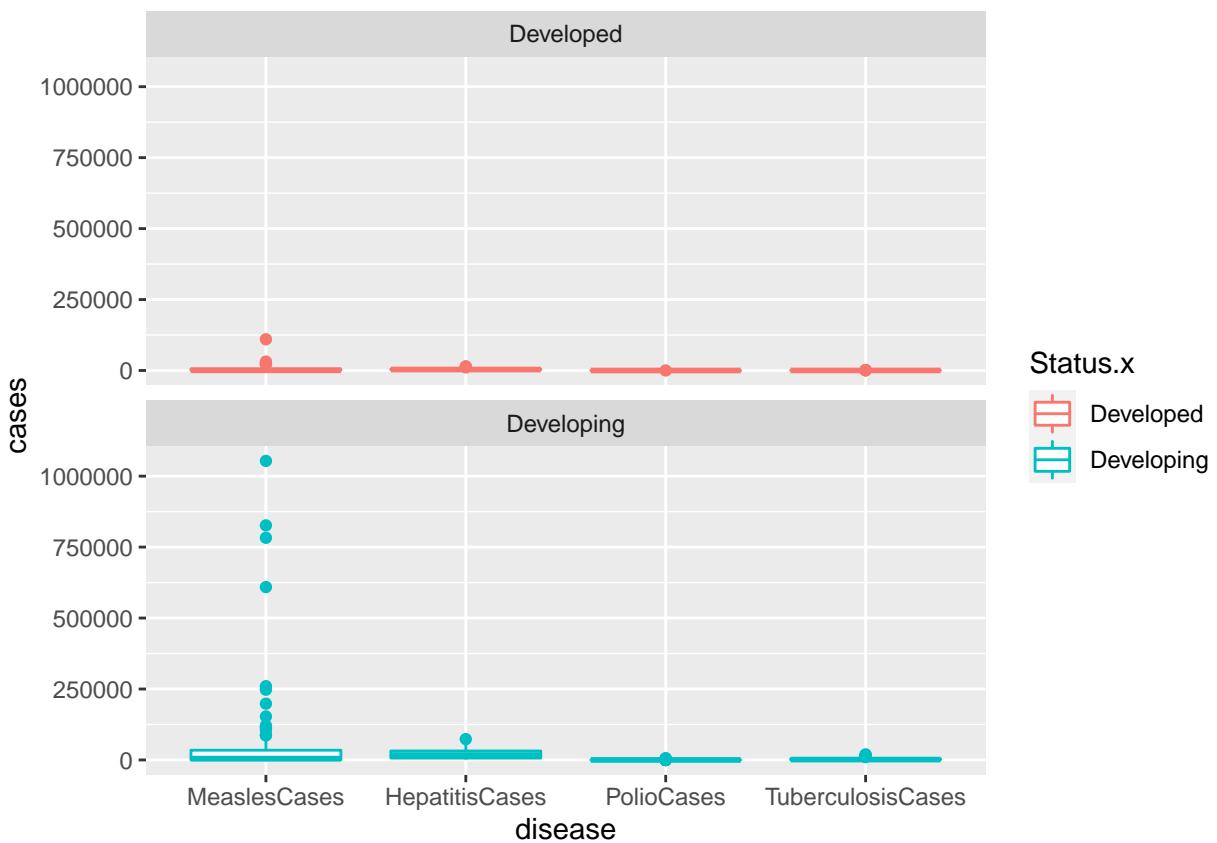
```

## # A tibble: 8 x 7
## # Groups:  disease [4]
##   disease Status.x TotalCountries CountrieswCase ActualCases NoCases PercentDif
##   <fct>   <chr>      <int>        <int>      <dbl>    <int>      <dbl>
## 1 Measles~ Develop~       31           30     254111      1      97
## 2 Measles~ Develop~      148          130     6660847     18      88
## 3 Hepatit~ Develop~       31           31     134293.     0      100
## 4 Hepatit~ Develop~      148          148     3064074.     0      100
## 5 PolioCa~ Develop~       31            1       2         30       3
## 6 PolioCa~ Develop~      148           64     19813       84      43
## 7 Tubercu~ Develop~       31           31     9992.       0      100
## 8 Tubercu~ Develop~      148          148     409470.     0      100

```







Percent of developed and developing with diseases

