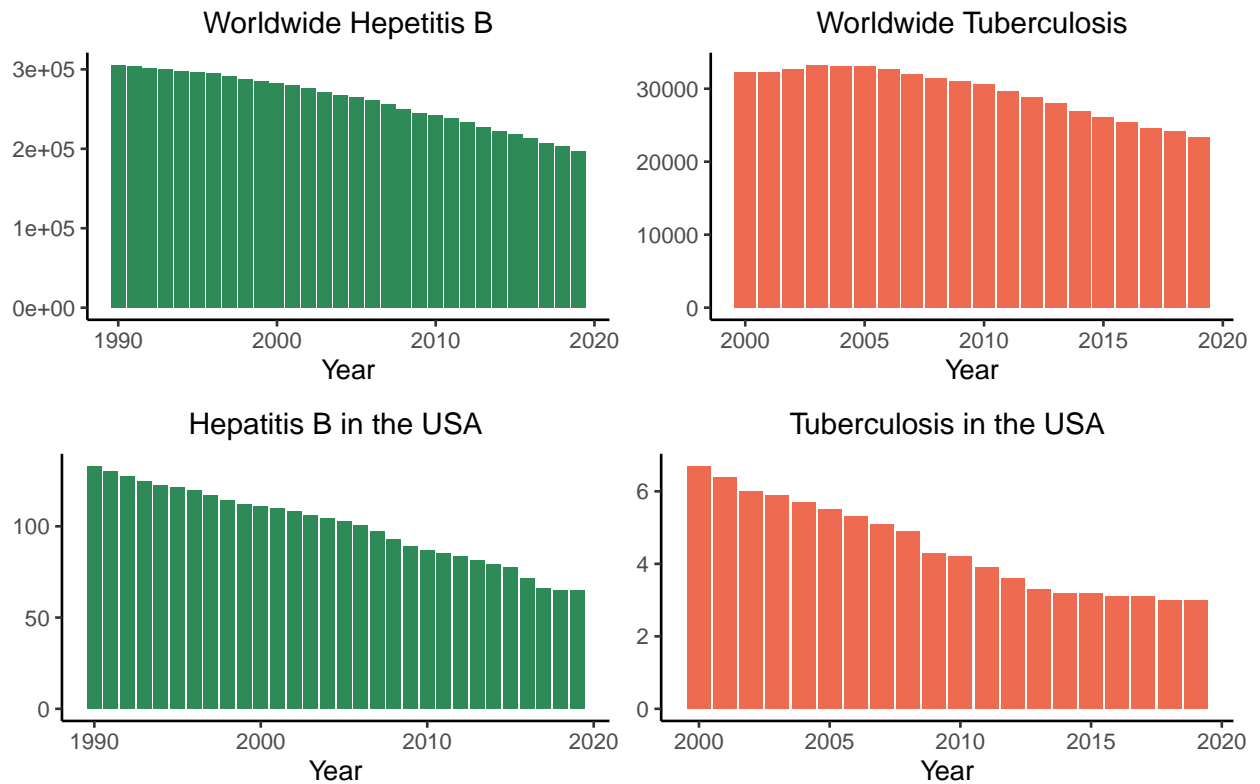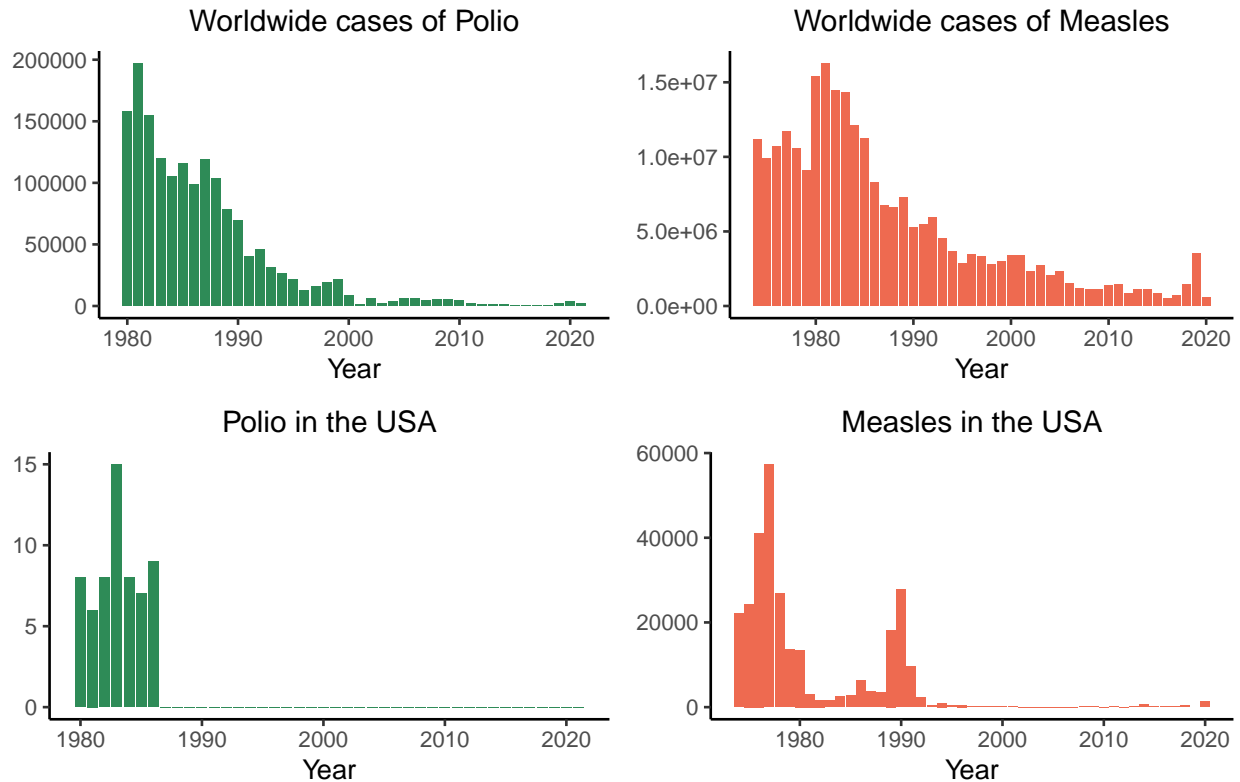# ProjectDBM

Shika, Kevin, Laura

Distribution of diseases

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

Measles vs every other attribute

- 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.
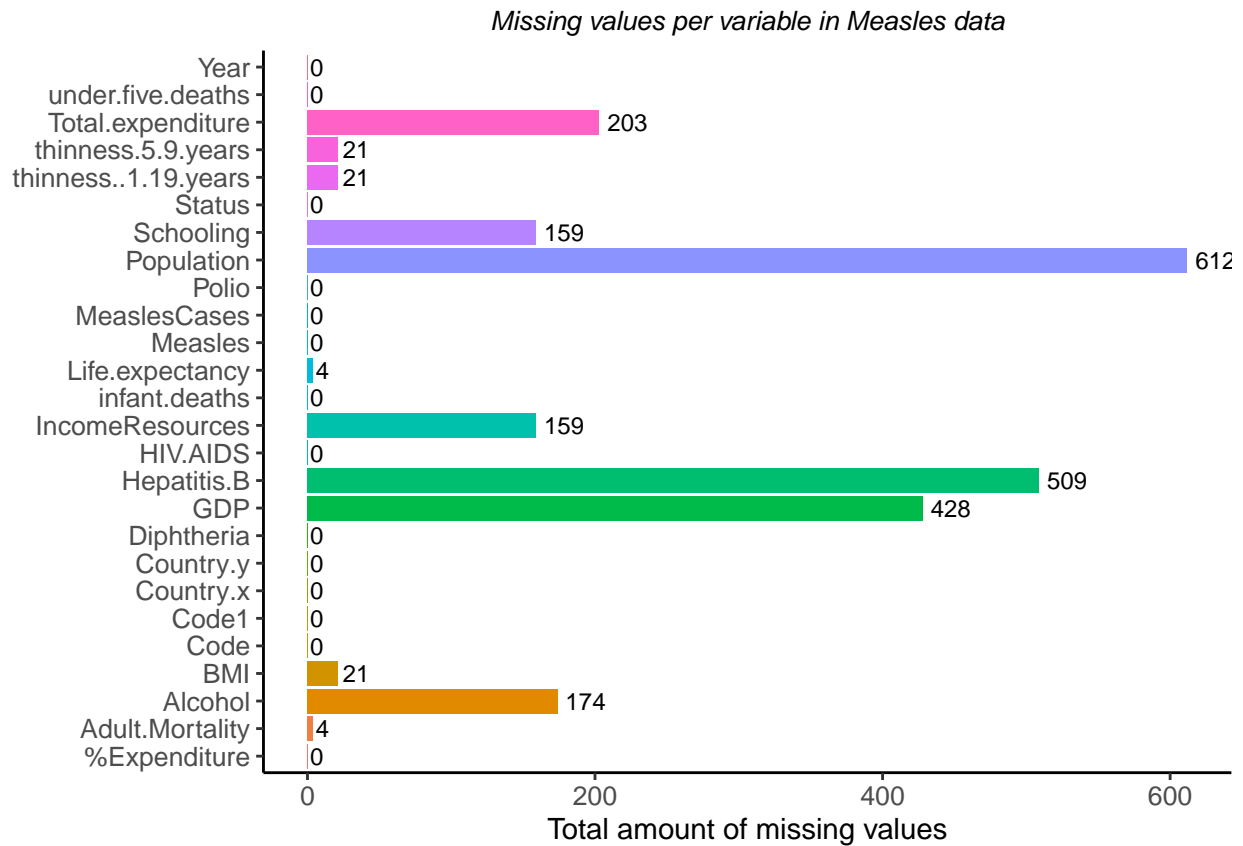- Youth Thinness vs Measles seem to have somewhat a linear relationship
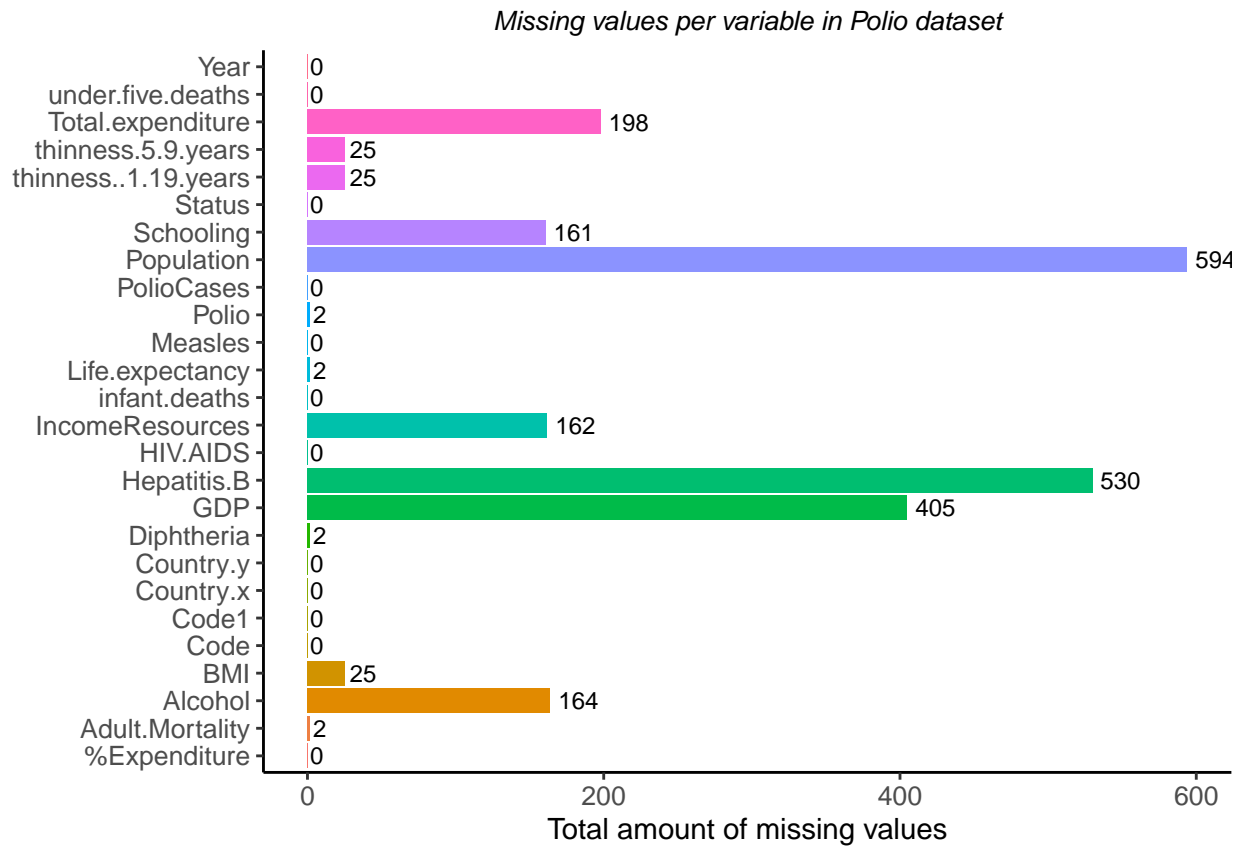
HepititisB vs all other attributes

- Adult mortality and hepatitis do have a linear relationship!
- There is also a negative relationship between life expectancy and a 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.**
- There is a negative relationship between diphtheria and Hepatitis
- Schooling vs Hepatitis b have a negative relationship
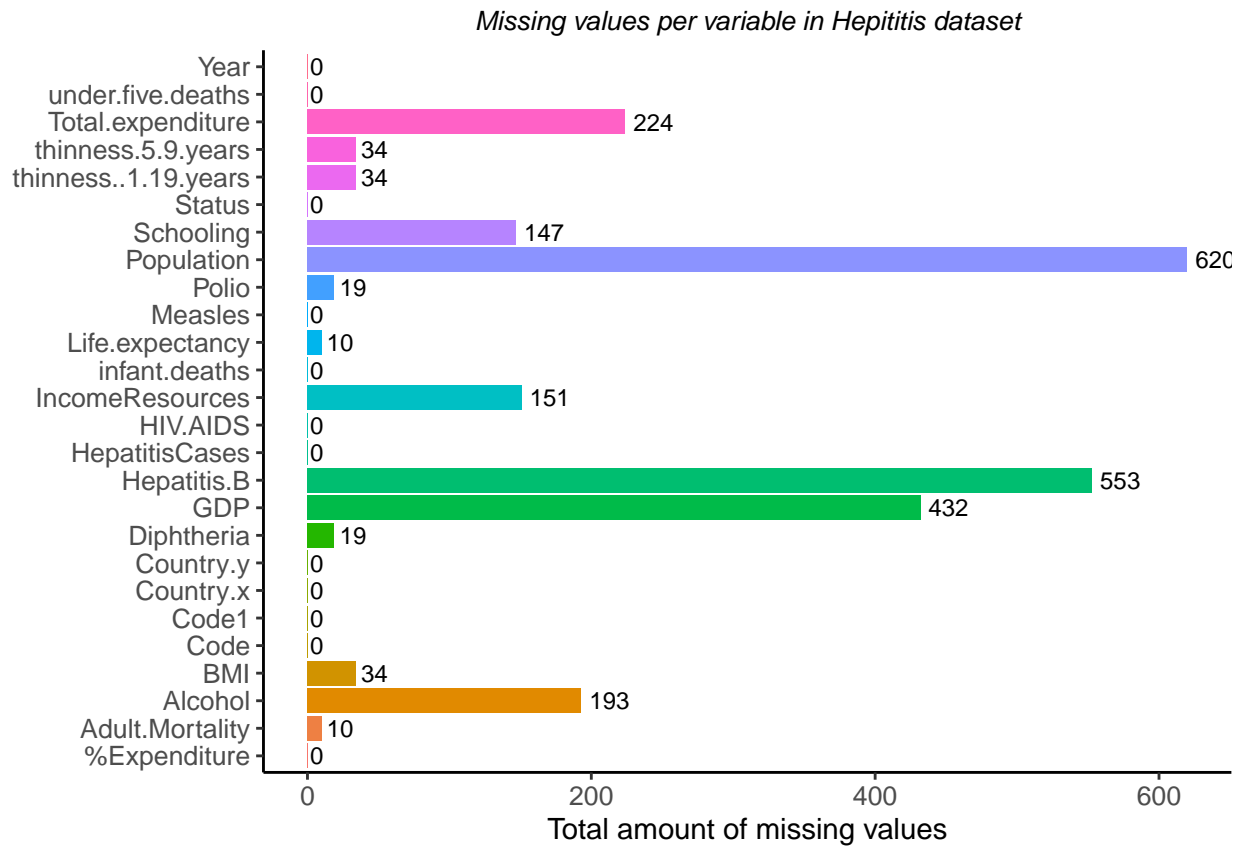- Youth thinness does have a positive relationship with hepatitis which makes sense

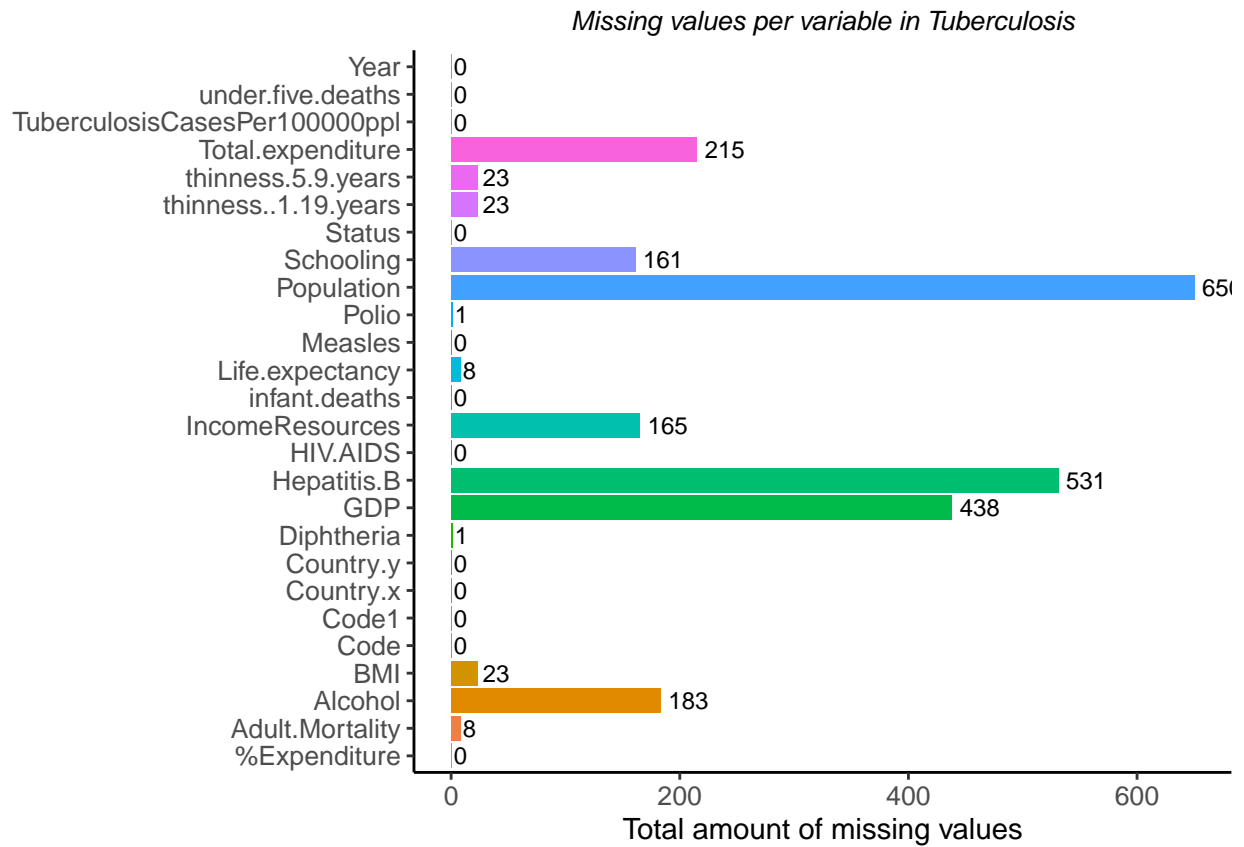## Tuberculosis vs all other attributes

- Life Expectancy has a negative relationship with Tuberculosis

- Adult mortality has a positive relationship with Tuberbulosis. The more cases the more changes of dying earlier in age.
- No relationship with BMI
- Immunization of polio vs tuberculosis seem to have a negative relationship
- Diphtheria also has a negative relationship with Tubercuosis.
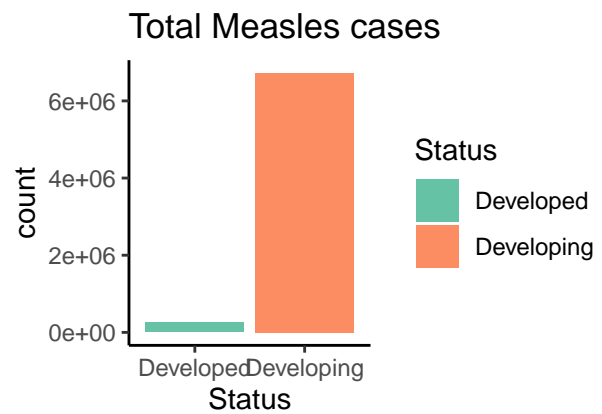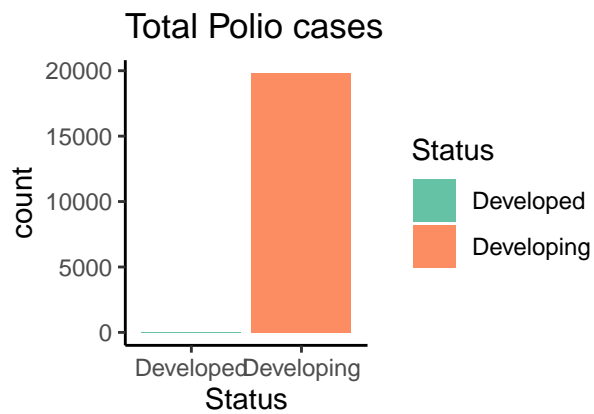- Tuberculosis has a positive relationship with HIV.AIDS

Missing values per variable in Measles data

Total amount of missing values

Missing values per variable in Polio dataset

Variable	Total amount of missing values
Year	0
under.five.deaths	0
Total.expenditure	198
thinness.5.9.years	25
thinness..1.19.years	25
Status	0
Schooling	161
Population	594
PolioCases	0
Polio	2
Measles	0
Life.expectancy	2
infant.deaths	0
IncomeResources	162
HIV.AIDS	0
Hepatitis.B	530
GDP	405
Diphtheria	2
Country.y	0
Country.x	0
Code1	0
Code	0
BMI	25
Alcohol	164
Adult.Mortality	2
%Expenditure	0

Missing values per variable in Hepititis dataset

Missing values per variable in Tuberculosis

Year — 0
under.five.deaths — 0
TuberculosisCasesPer100000ppl — 0
Total.expenditure — 215
thinness.5.9.years — 23
thinness..1.19.years — 23
Status — 0
Schooling — 161
Population — 650
Polio — 1
Measles — 0
Life.expectancy — 8
infant.deaths — 0
IncomeResources — 165
HIV.AIDS — 0
Hepatitis.B — 531
GDP — 438
Diphtheria — 1
Country.y — 0
Country.x — 0
Code1 — 0
Code — 0
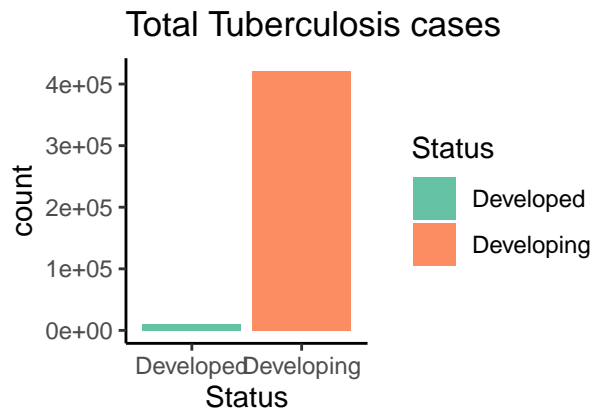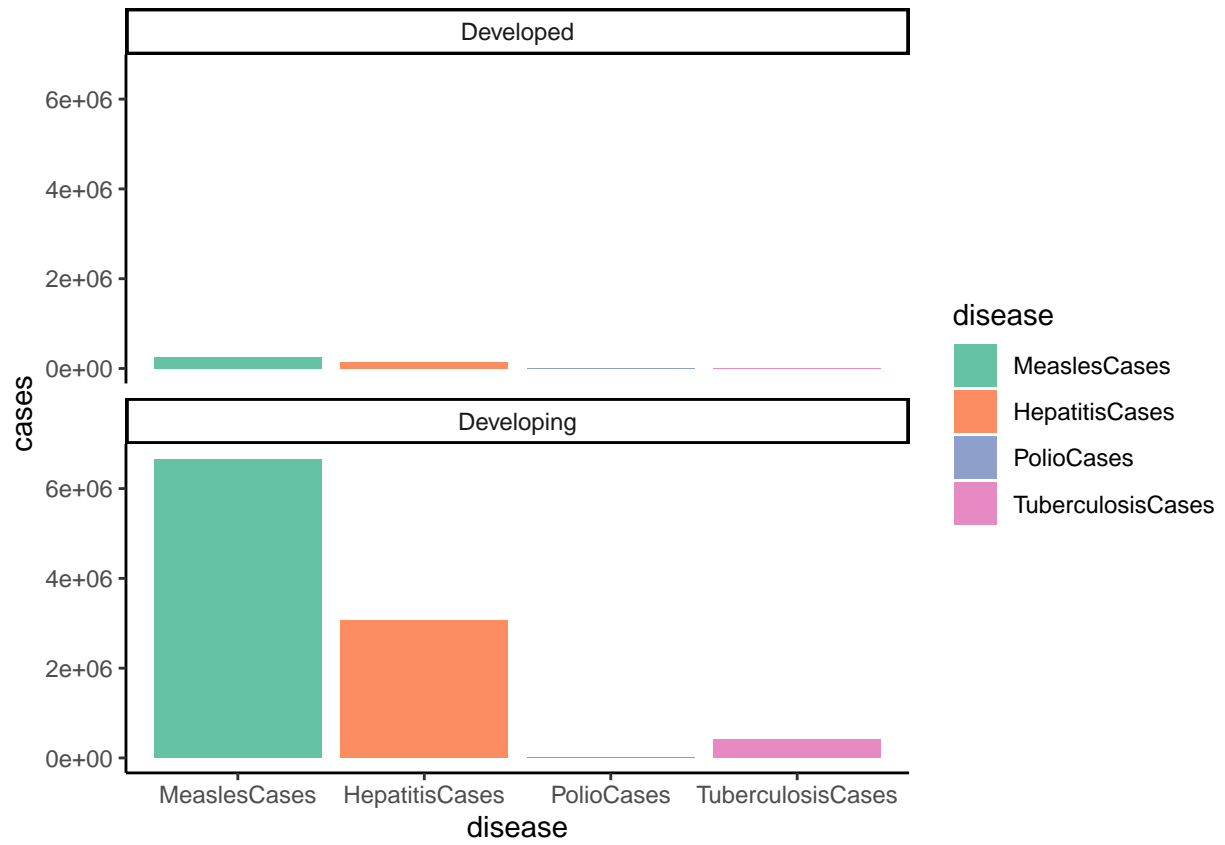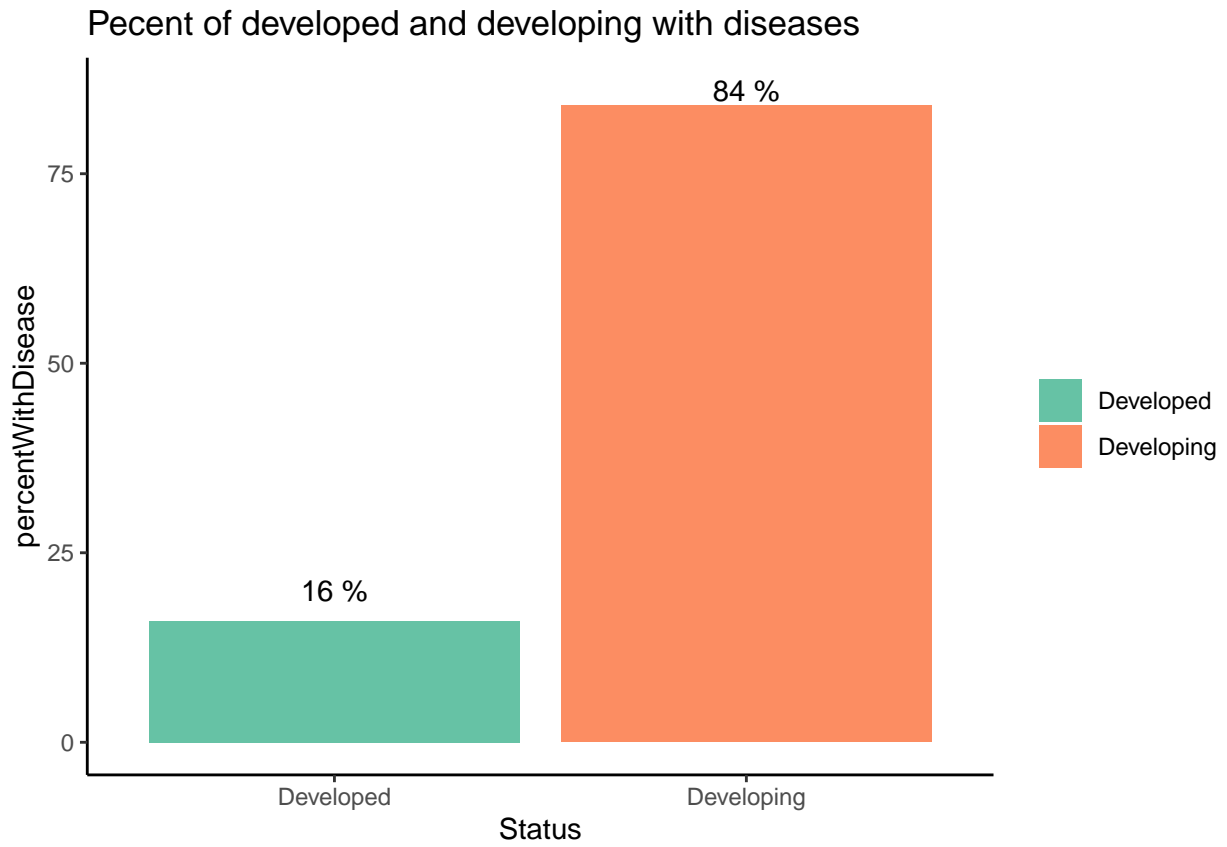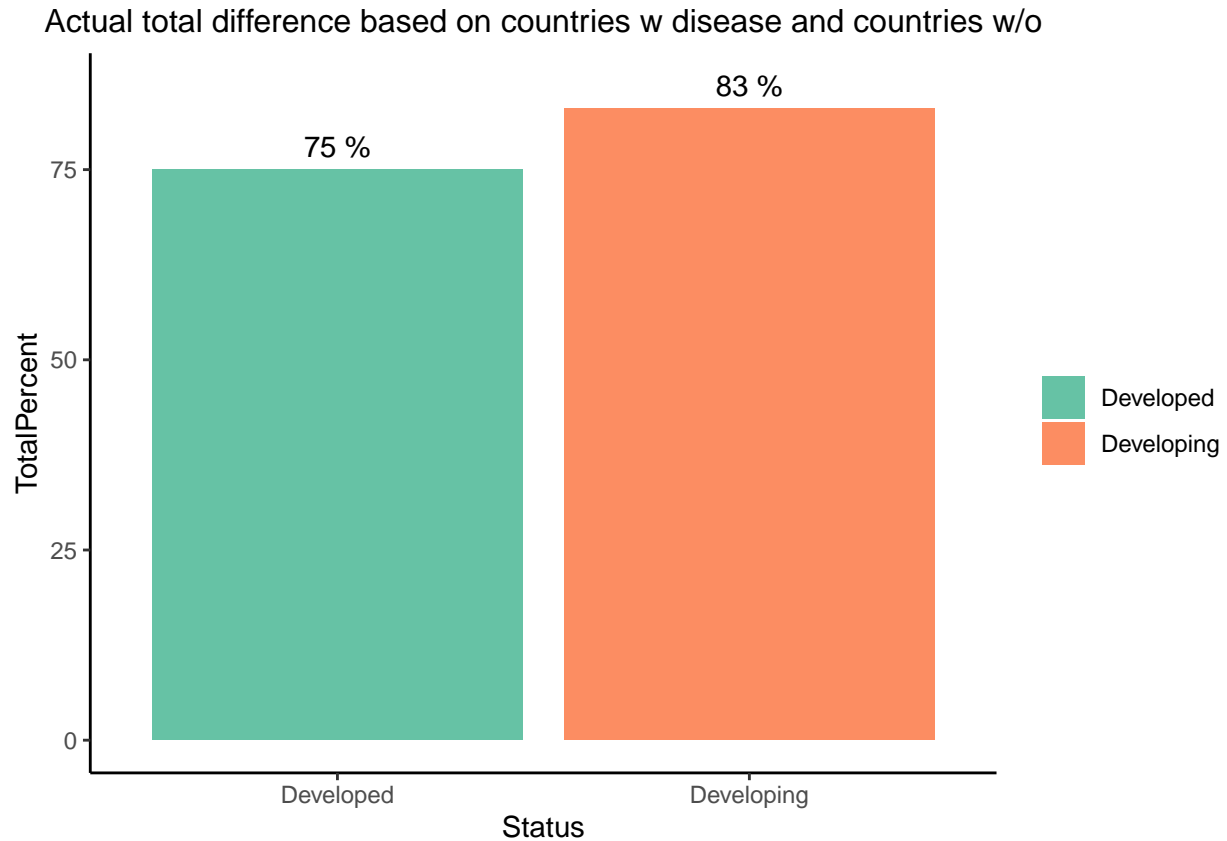BMI — 23
Alcohol — 183
Adult.Mortality — 8
%Expenditure — 0

Total amount of missing values

Developed vs developing

or

Total Tuberculosis cases

Total HPB2 Cases

Total Polio cases

Total Measles cases

Pecent of developed and developing with diseases

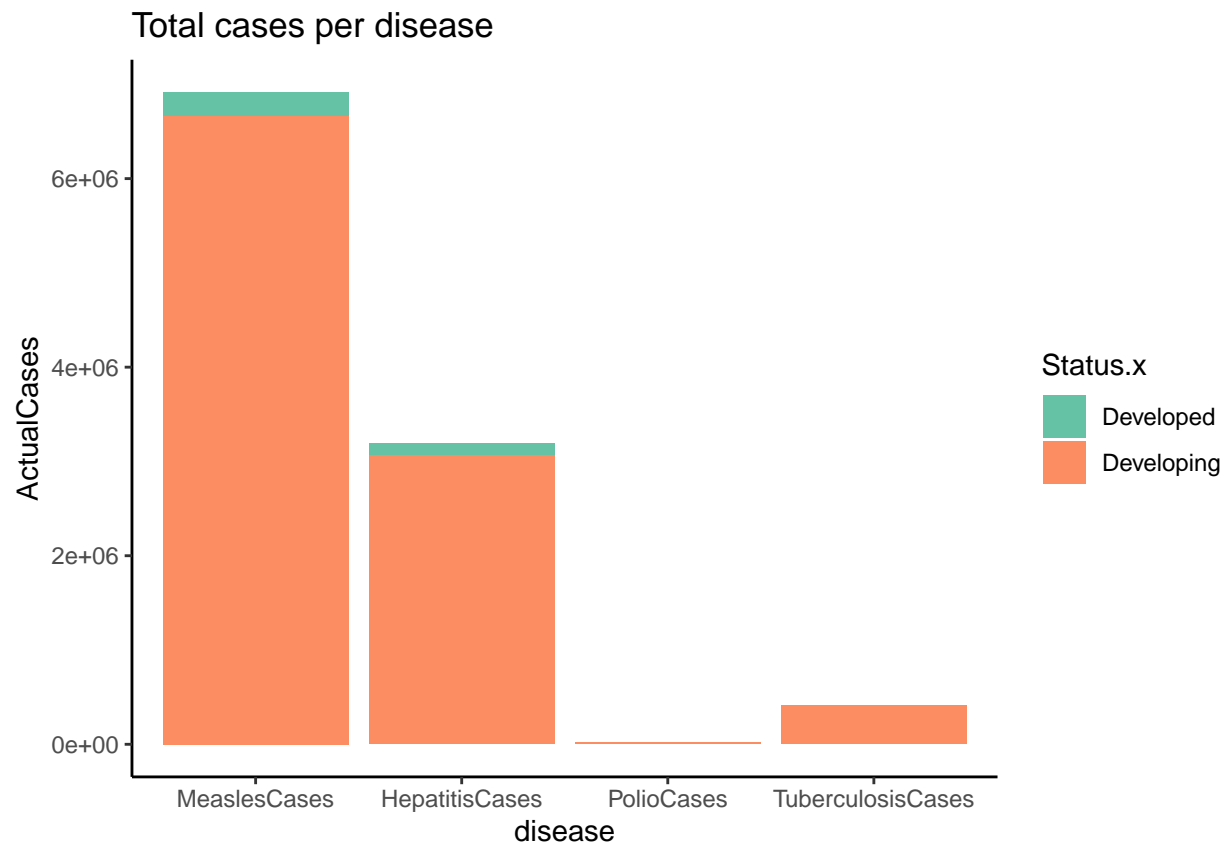## Actual total difference based on countries w disease and countries w/o
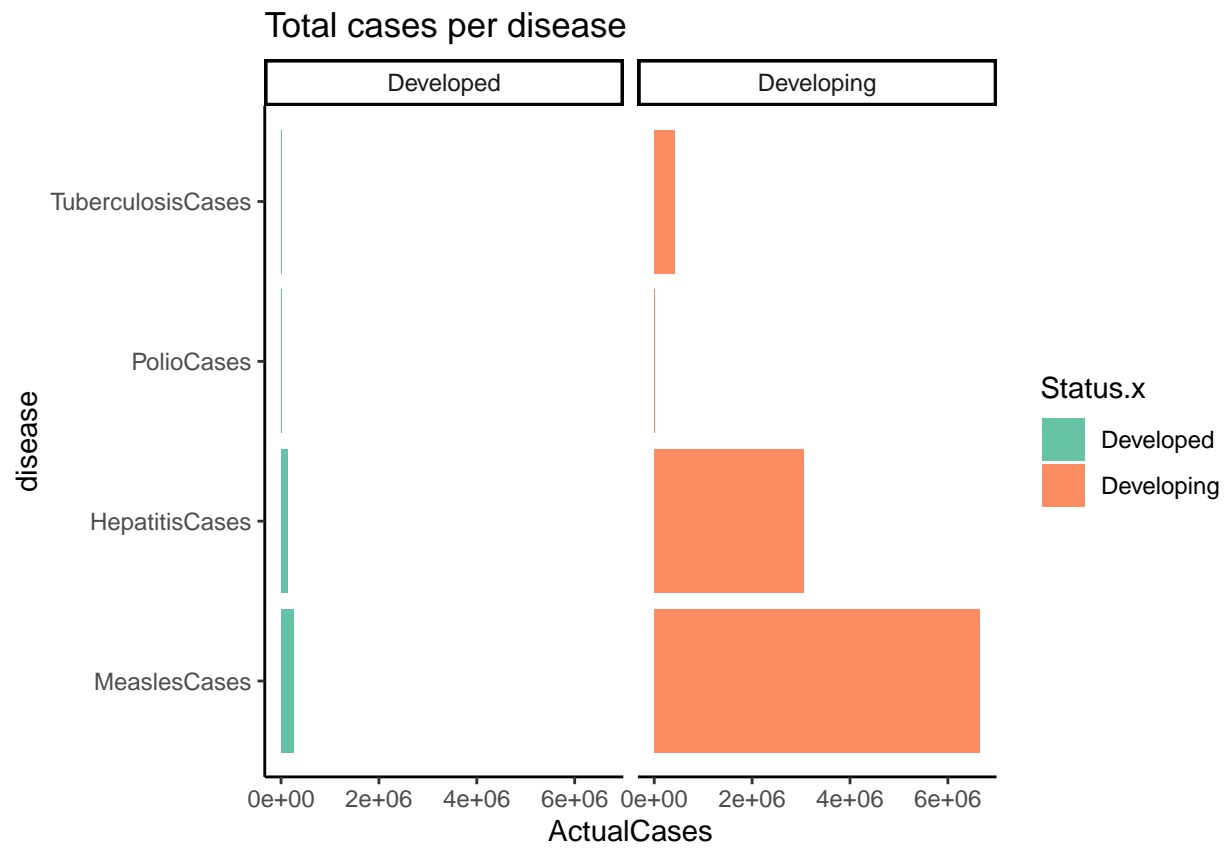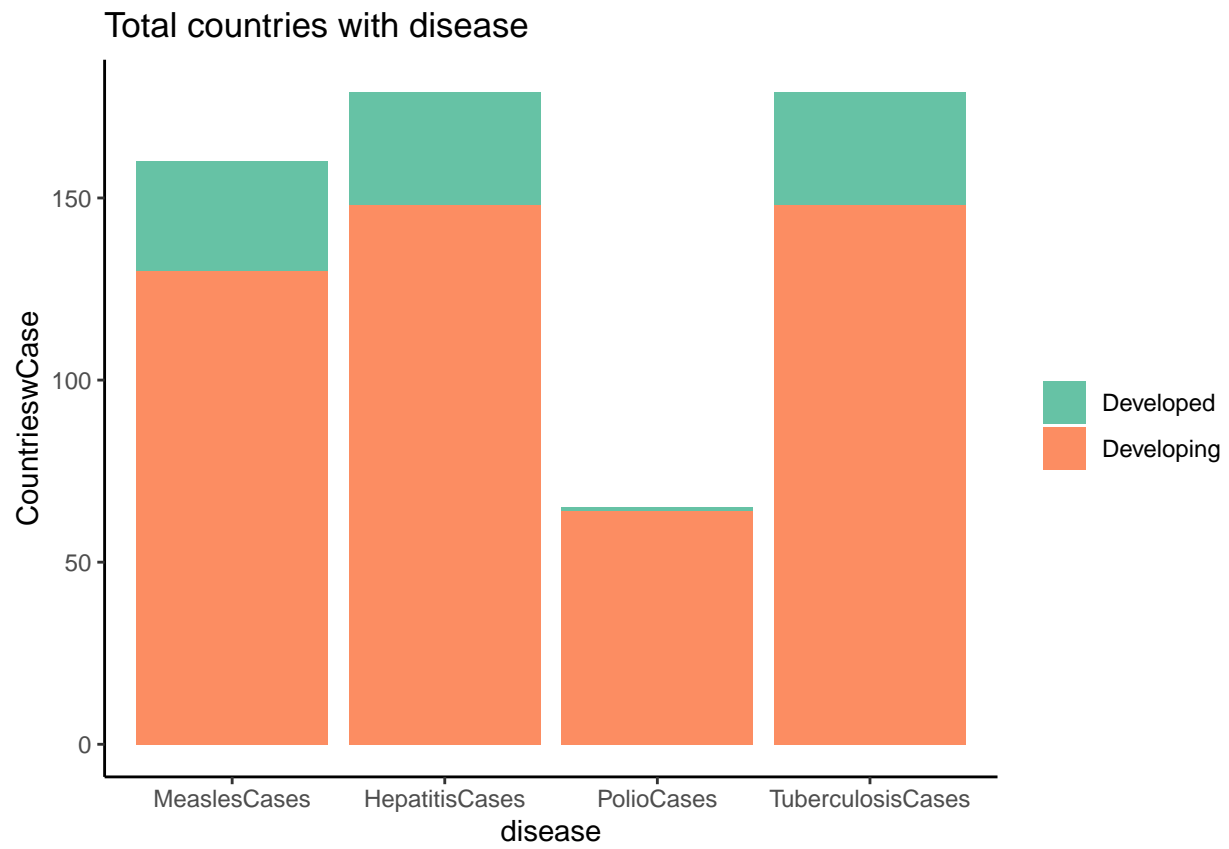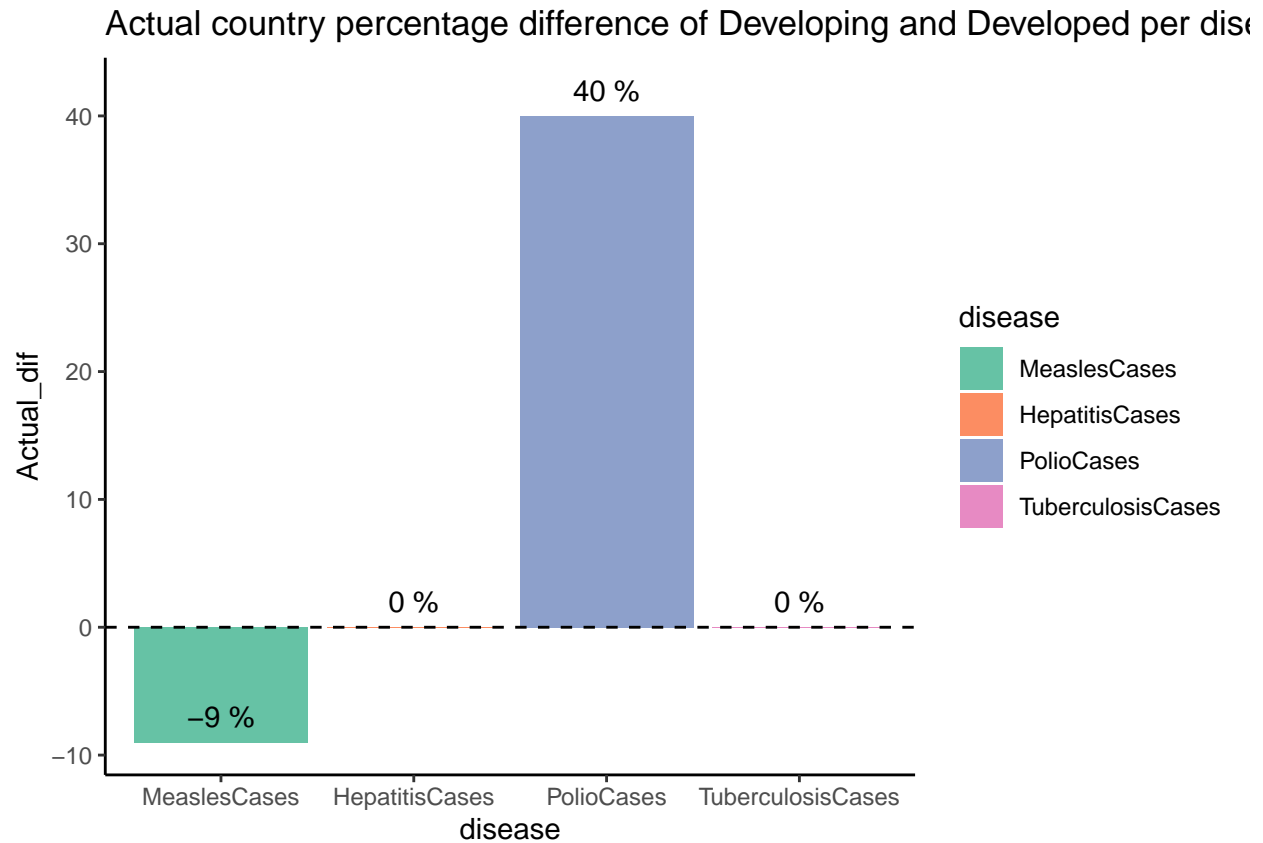

Developed vs Developing

- The data is composed of 17% developed countries and 83% developing

- When filter the data with country disease that is 0 it changes 2%. Changing the distribution to 15% developed and 85% developing

- All countries in the data had Tubeculosis

- **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**

- *Here is the table of that actual difference and percentage which is actually not that much*

- Polio developed vs developed almost no difference

Total cases per disease

Total cases per disease

Total countries with disease

Actual country percentage difference of Developing and Developed per disease

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
```