Analysis on Factors affecting Life Expectancy

```
library(readr)
expect<-read_csv("D:/MITA 2019/Semester 2/Multivariate Analysis/lfe.csv")
```

```
## Parsed with column specification:
## cols(
##   .default = col_double(),
##   Country = col_character(),
##   Status = col_character()
## )
```

```
## See spec(...) for full column specifications.
```

```
head(expect)
```

```
## # A tibble: 6 x 22
##   Country  Year Status `Life expectanc~ `Adult Mortalit~ `infant deaths`
##   <chr>   <dbl> <chr>             <dbl>            <dbl>           <dbl>
## 1 Afghan~  2015 Devel~             65                263              62
## 2 Afghan~  2014 Devel~             59.9              271              64
## 3 Afghan~  2013 Devel~             59.9              268              66
## 4 Afghan~  2012 Devel~             59.5              272              69
## 5 Afghan~  2011 Devel~             59.2              275              71
## 6 Afghan~  2010 Devel~             58.8              279              74
## # ... with 16 more variables: Alcohol <dbl>, `percentage
## #   expenditure` <dbl>, `Hepatitis B` <dbl>, Measles <dbl>, BMI <dbl>,
## #   `under-five deaths` <dbl>, Polio <dbl>, `Total expenditure` <dbl>,
## #   Diphtheria <dbl>, `HIV/AIDS` <dbl>, GDP <dbl>, Population <dbl>,
## #   `thinness 1-19 years` <dbl>, `thinness 5-9 years` <dbl>, `Income
## #   composition of resources` <dbl>, Schooling <dbl>
```

```
library(tidyverse) # metapackage with lots of helpful functions
```

```
## -- Attaching packages ---------------------------------------------------
---------- tidyverse 1.2.1 --
```

```
## v ggplot2 3.2.1     v purrr   0.3.2
## v tibble  2.1.3     v dplyr   0.8.3
## v tidyr   1.0.0     v stringr 1.4.0
## v ggplot2 3.2.1     v forcats 0.4.0
```

```
## -- Conflicts ------------------------------------------------------------
---- tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
library(gridExtra) # grid.arrange to make quick subplots
```

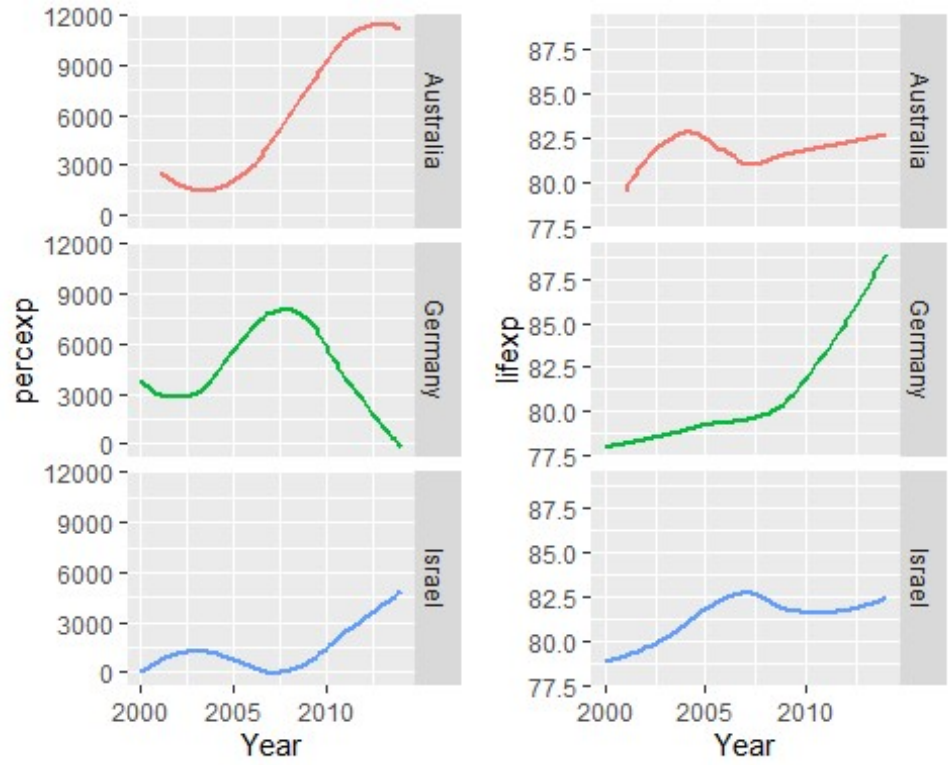
```
##
## Attaching package: 'gridExtra'

## The following object is masked from 'package:dplyr':
##
##     combine

library(reshape2)

##
## Attaching package: 'reshape2'

## The following object is masked from 'package:tidyr':
##
##     smiths

expect <- expect %>%
  drop_na() %>%
  rename('lifexp'='Life expectancy',
         'percexp'='percentage expenditure',
         'totexp'='Total expenditure',
         'admort'='Adult Mortality',
         'infmort'='infant deaths',
         'u5deaths'='under-five deaths',
         'hepb'='Hepatitis B',
         'HIV'='HIV/AIDS') %>%
  filter(!is.na(lifexp), !is.na(admort), !is.na(infmort))
head(expect)

## # A tibble: 6 x 22
##   Country  Year Status lifexp admort infmort Alcohol percexp  hepb Measles
##   <chr>   <dbl> <chr>   <dbl>  <dbl>   <dbl>   <dbl>   <dbl> <dbl>   <dbl>
## 1 Afghan~  2015 Devel~   65      263      62    0.01   71.3     65    1154
## 2 Afghan~  2014 Devel~   59.9    271      64    0.01   73.5     62     492
## 3 Afghan~  2013 Devel~   59.9    268      66    0.01   73.2     64     430
## 4 Afghan~  2012 Devel~   59.5    272      69    0.01   78.2     67    2787
## 5 Afghan~  2011 Devel~   59.2    275      71    0.01    7.10    68    3013
## 6 Afghan~  2010 Devel~   58.8    279      74    0.01   79.7     66    1989
## # ... with 12 more variables: BMI <dbl>, u5deaths <dbl>, Polio <dbl>,
## #   totexp <dbl>, Diphtheria <dbl>, HIV <dbl>, GDP <dbl>,
## #   Population <dbl>, `thinness 1-19 years` <dbl>, `thinness 5-9
## #   years` <dbl>, `Income composition of resources` <dbl>, Schooling <dbl>

#Descriptive statistics grouped by status of country
expect %>%
  group_by(Status) %>%
  summarize(count = n(),
            avg_lifexp = mean(lifexp, na.rm=TRUE),
            #avg_infmort = mean(infmort, na.rm=TRUE),
            #avg_admort = mean(admort, na.rm=TRUE),
            avg_GDP = mean(GDP, na.rm=TRUE),
```

```
            avg_Population = mean(Population, na.rm=TRUE),
            avg_totexp = mean(totexp,na.rm=TRUE))
```

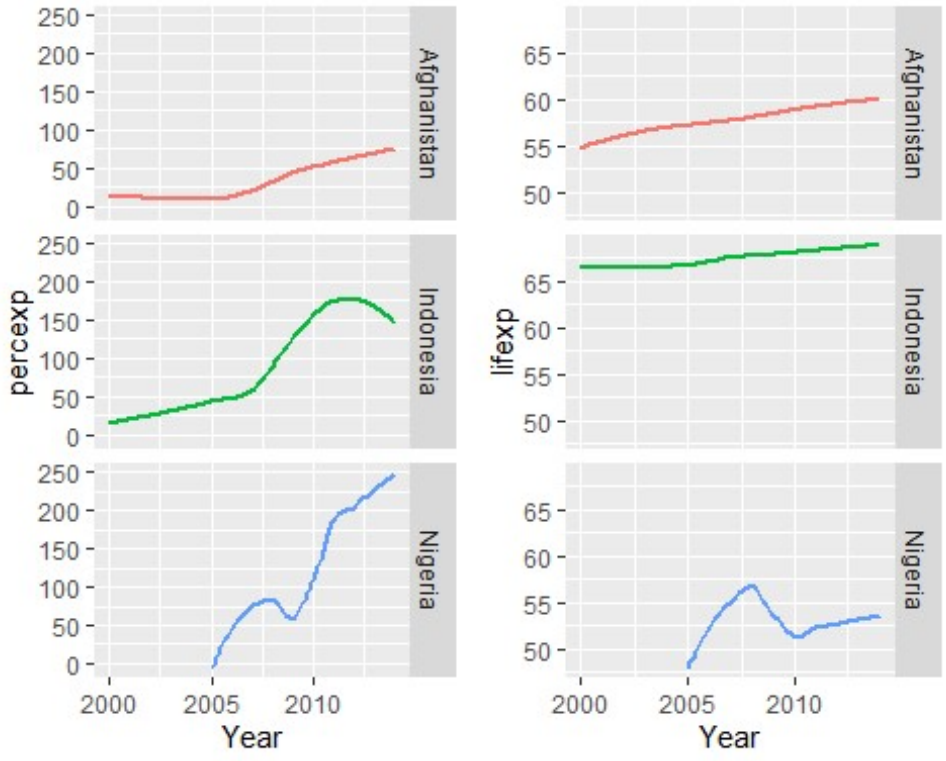
```
## # A tibble: 2 x 6
##   Status     count avg_lifexp avg_GDP avg_Population avg_totexp
##   <chr>      <int>      <dbl>   <dbl>          <dbl>      <dbl>
## 1 Developed    242       78.7  18977.       8744688.       7.02
## 2 Developing  1407       67.7   3259.      15669946.       5.77
```

```
devd <- expect %>%
  filter(Country %in% c('Australia', 'Germany', 'Israel'))
deving <- expect %>%
  filter(Country %in% c('Afghanistan', 'Nigeria', 'Indonesia'))

# removing 2015 data, has some NA values
devd <- devd %>%
  filter(Year != 2015)
deving <- deving %>%
  filter(Year != 2015)

#EDA for Developed countries against percentage expenditure and life
expectancy
devdp1 <- ggplot(devd, aes(Year, percexp))+
  geom_smooth(aes(color=Country), se=FALSE, show.legend=FALSE)+
  facet_grid(Country~.)

devdp2 <- ggplot(devd, aes(Year, lifexp))+
  geom_smooth(aes(color=Country), se=FALSE, show.legend=FALSE)+
  facet_grid(Country~.)
grid.arrange(devdp1, devdp2, nrow=1)
```

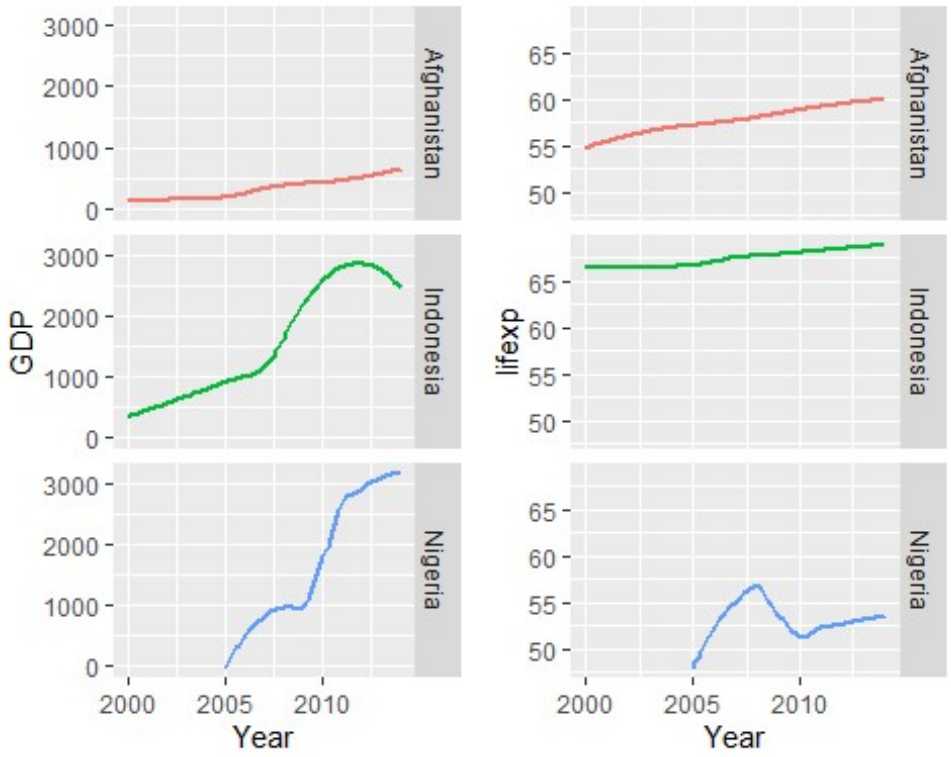
```
## `geom_smooth()` using method = 'loess' and formula 'y ~ x'
## `geom_smooth()` using method = 'loess' and formula 'y ~ x'
```




```
#EDA for developing countries against percentage expenditure and life
expectancy
devingp1 <- ggplot(deving, aes(Year, percexp))+
  geom_smooth(aes(color=Country), se=FALSE, show.legend=FALSE)+
  facet_grid(Country~.)

devingp2 <- ggplot(deving, aes(Year, lifexp))+
  geom_smooth(aes(color=Country), se=FALSE, show.legend=FALSE)+
  facet_grid(Country~.)
grid.arrange(devingp1, devingp2, nrow=1)

## `geom_smooth()` using method = 'loess' and formula 'y ~ x'
## `geom_smooth()` using method = 'loess' and formula 'y ~ x'
```




```
#EDA for developing countries against GDP and life expectancy
devingp1 <- ggplot(deving, aes(Year, GDP))+
  geom_smooth(aes(color=Country), se=FALSE, show.legend=FALSE)+
  facet_grid(Country~.)

devingp2 <- ggplot(deving, aes(Year, lifexp))+
  geom_smooth(aes(color=Country), se=FALSE, show.legend=FALSE)+
  facet_grid(Country~.)
grid.arrange(devingp1, devingp2, nrow=1)

## `geom_smooth()` using method = 'loess' and formula 'y ~ x'
## `geom_smooth()` using method = 'loess' and formula 'y ~ x'
```




```
#correlation between life expectancy and percentage expenditure, use cor() to
check
cor(expect$lifexp, expect$percexp)
```

```
## [1] 0.4096308
```

```
#correlation between life expectancy and GDP
cor(expect$lifexp, expect$GDP)
```

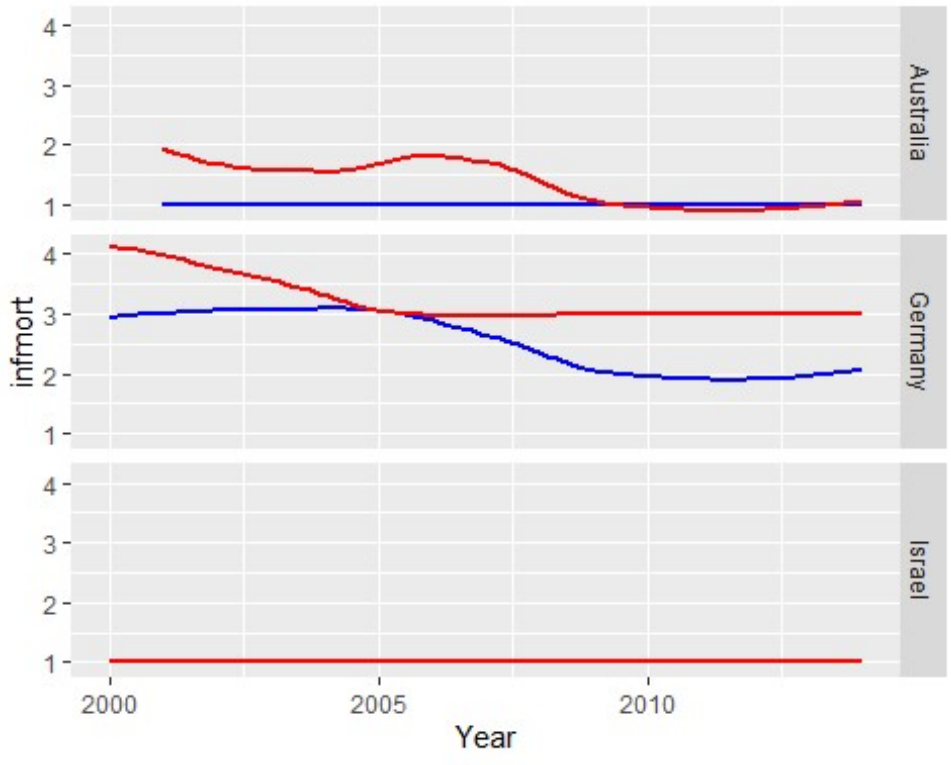
```
## [1] 0.4413218
```

```
#correlation between life expectancy and Alcohol
cor(expect$lifexp,expect$Alcohol)
```

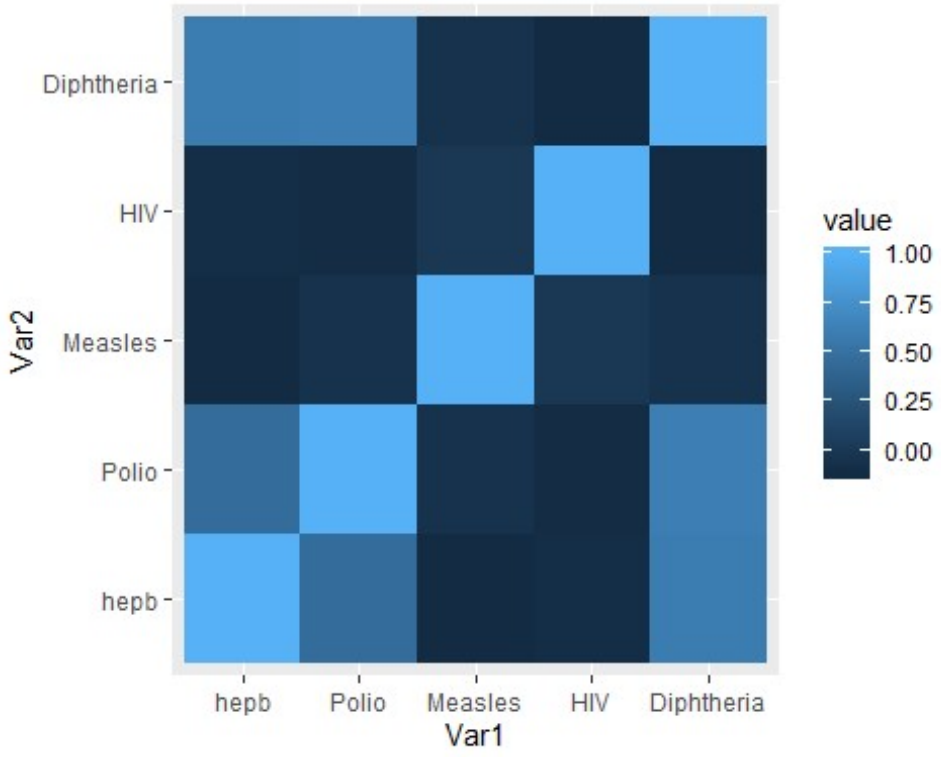
```
## [1] 0.4027183
```

```
# looking at infant deaths and under-five deaths for developed countries
ggplot(devd, show.legend=FALSE)+
  geom_smooth(aes(Year, infmort),color='blue', se=FALSE)+

  geom_smooth(aes(Year, u5deaths), color='red', se=FALSE)+
  facet_grid(Country~.)
```

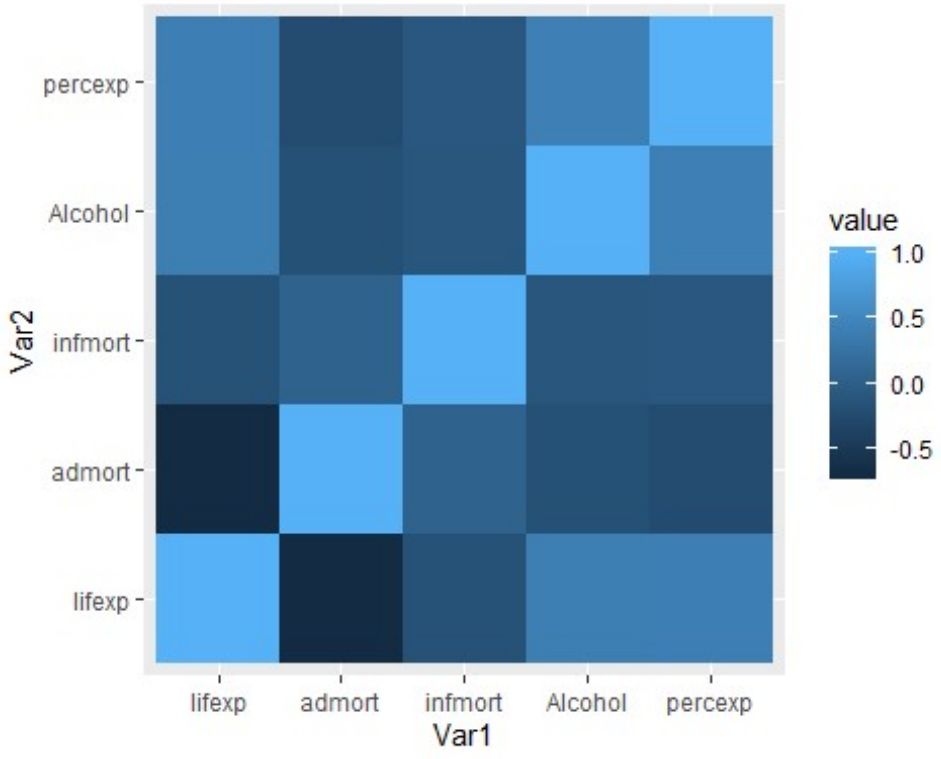
```
## `geom_smooth()` using method = 'loess' and formula 'y ~ x'
## `geom_smooth()` using method = 'loess' and formula 'y ~ x'
```




```
# any correlation between vaccinations?
cordf <- expect %>%
  drop_na() %>%
  select(hepb, Polio, Measles, HIV, Diphtheria)
cormat <- cor(cordf)
melted <- melt(cormat)
ggplot(melted)+
  geom_tile(aes(Var1, Var2, fill=value))
```

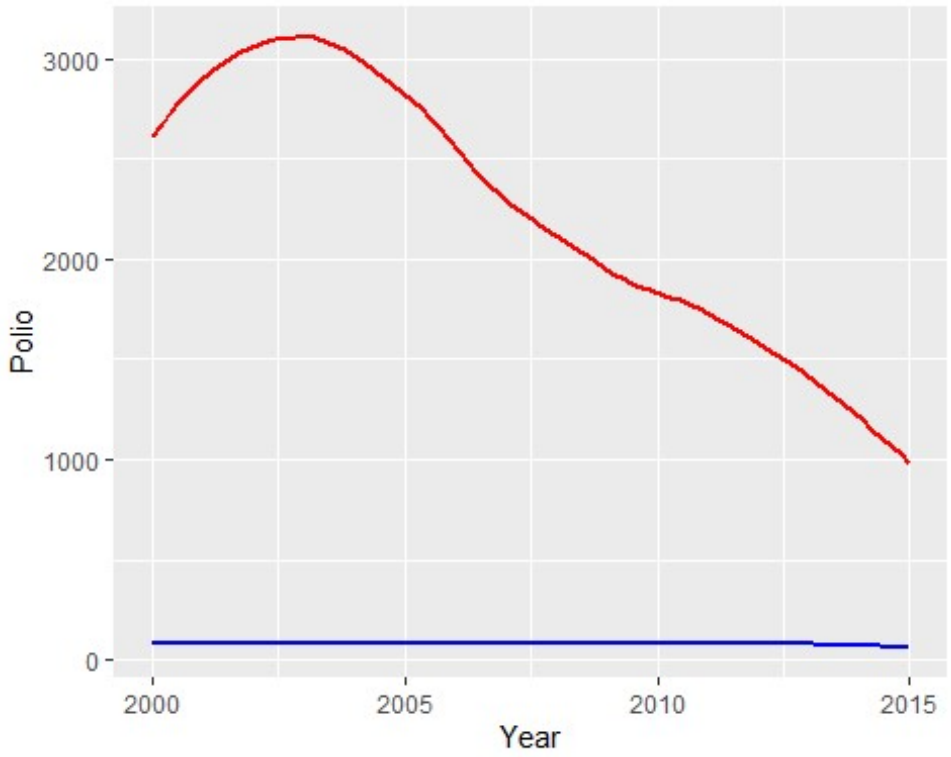



```
cordf <- expect %>%
  drop_na() %>%
  select(lifexp,admort,infmort,Alcohol,percexp)
cormat <- cor(cordf)
melted <- melt(cormat)
ggplot(melted)+
  geom_tile(aes(Var1, Var2, fill=value))
```



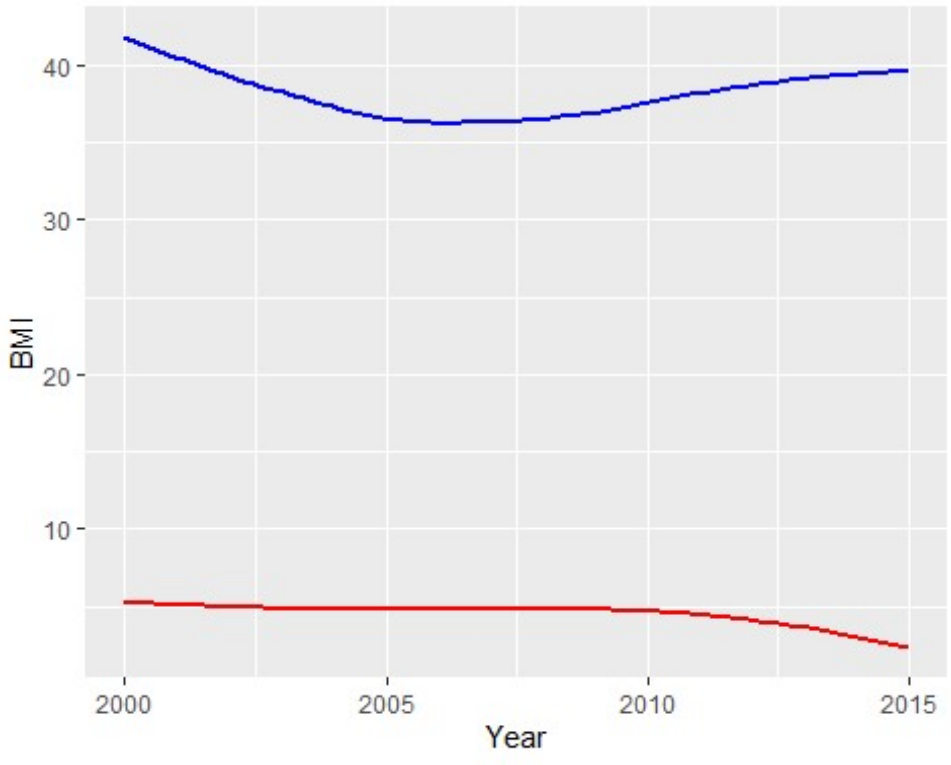

```
# how about adult mortality and infant mortality?
expect %>%
  group_by(Year) %>%
  summarize(Polio=mean(Polio),
            Measles=mean(Measles)) %>%
  ggplot()+
  geom_smooth(aes(Year, Polio), color='blue', se=FALSE)+
  geom_smooth(aes(Year, Measles), color='red', se=FALSE)
```

```
## `geom_smooth()` using method = 'loess' and formula 'y ~ x'
## `geom_smooth()` using method = 'loess' and formula 'y ~ x'
```

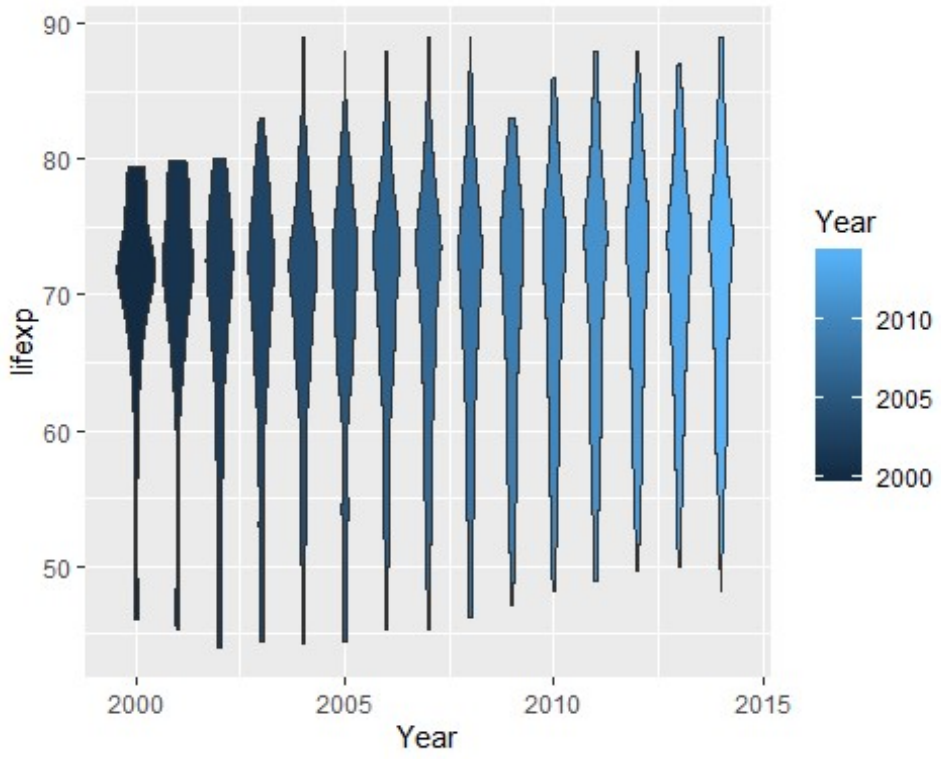



```
#geom_smooth(aes(Year, HIV), color='yellow', se=FALSE)+
#geom_smooth(aes(Year, Diphtheria), color='', se=FALSE)+

expect %>%
  group_by(Year) %>%
  summarize(BMI=mean(BMI),
            Alcohol=mean(Alcohol)) %>%
  ggplot()+
  geom_smooth(aes(Year, BMI), color='blue', se=FALSE)+
  geom_smooth(aes(Year, Alcohol), color='red', se=FALSE)
```

```
## `geom_smooth()` using method = 'loess' and formula 'y ~ x'
## `geom_smooth()` using method = 'loess' and formula 'y ~ x'
```

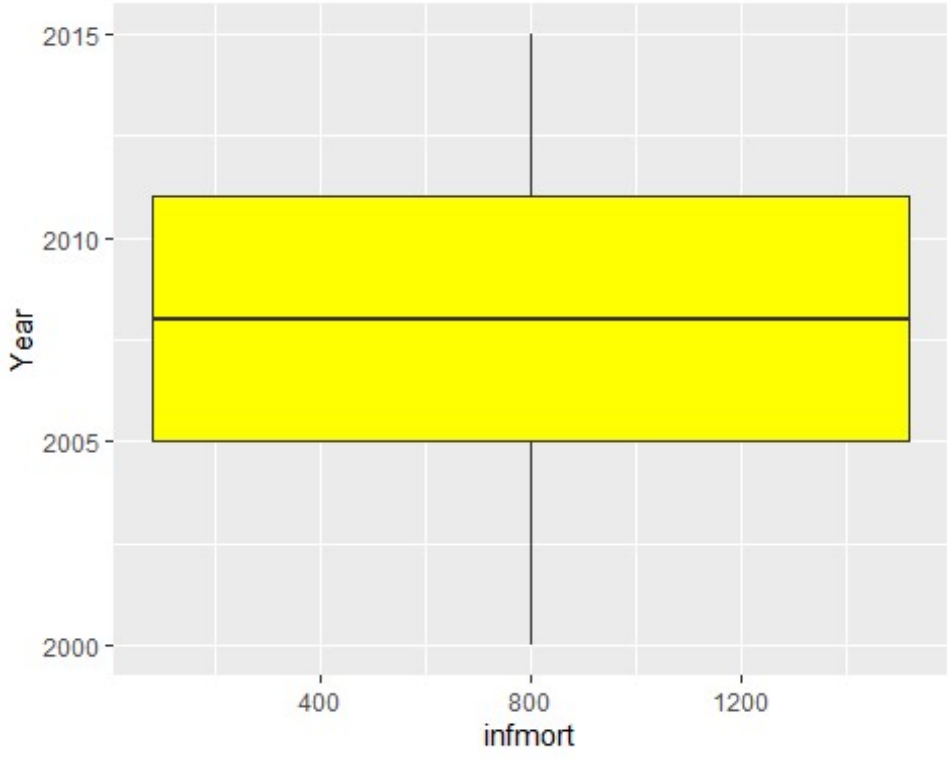



```r
#distribution of life expectancy by year
expect %>%
  ggplot()+
  geom_violin(aes(x=Year, y=lifexp, group=Year, fill=Year))
```

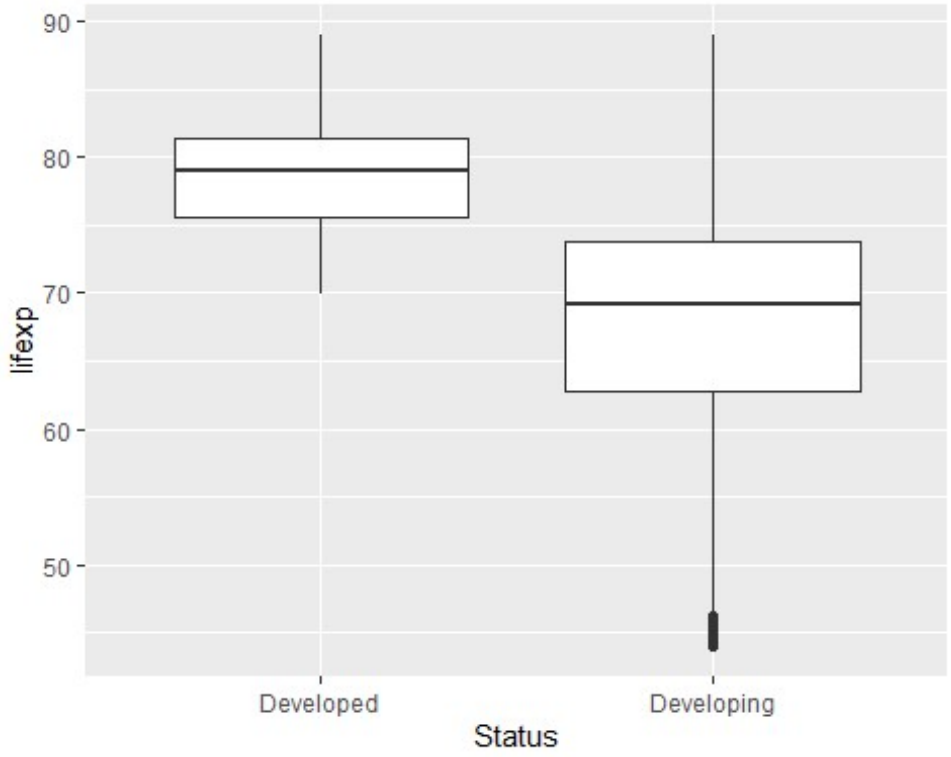



```
ggplot(expect,aes(y=Year,x=infmort))+geom_boxplot(fill='yellow')
```

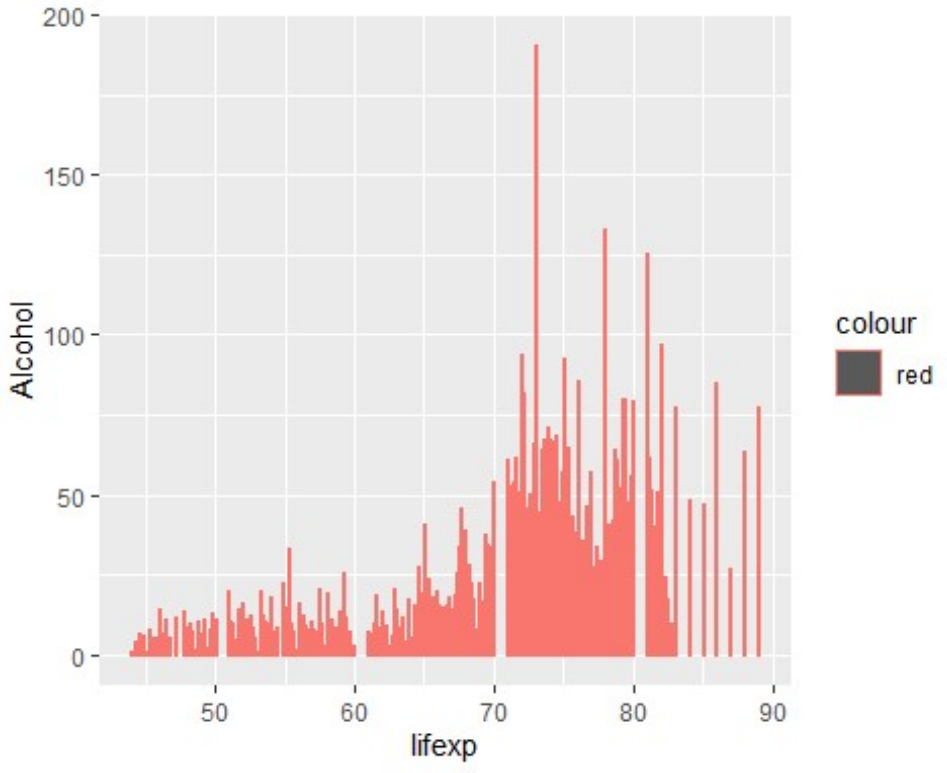
```
## Warning: Continuous x aesthetic -- did you forget aes(group=...)?
```



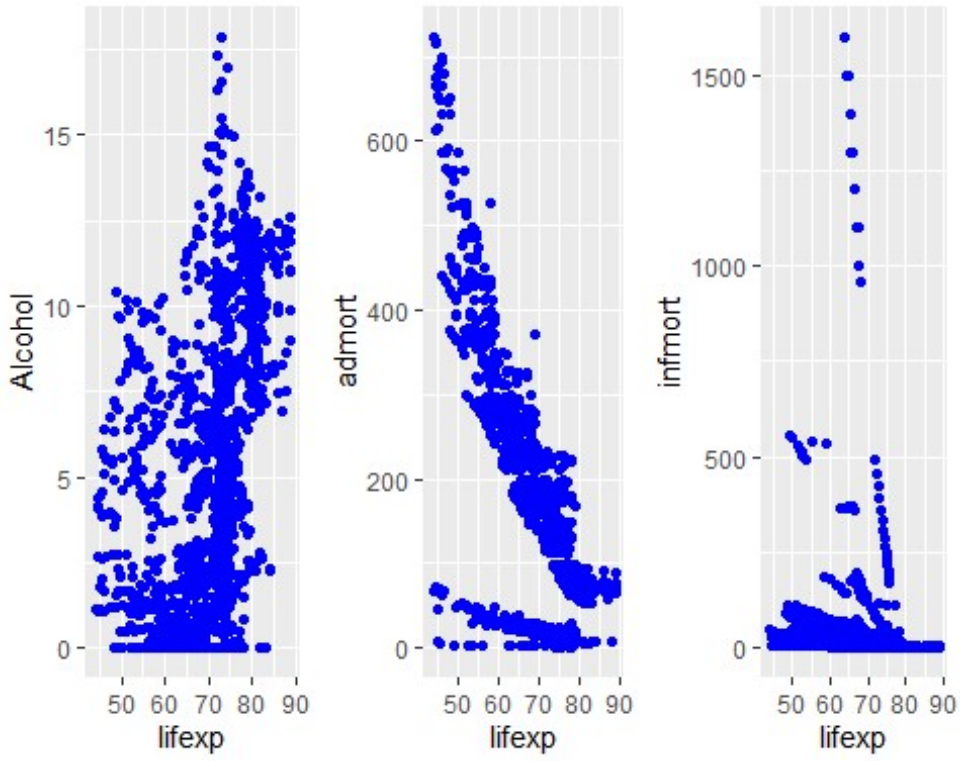
```r
#Converting status into categorical and plotting boxplot
expect$factor = as.factor(expect$Status)
ggplot(expect,aes(y=lifexp,x=Status))+geom_boxplot()
```



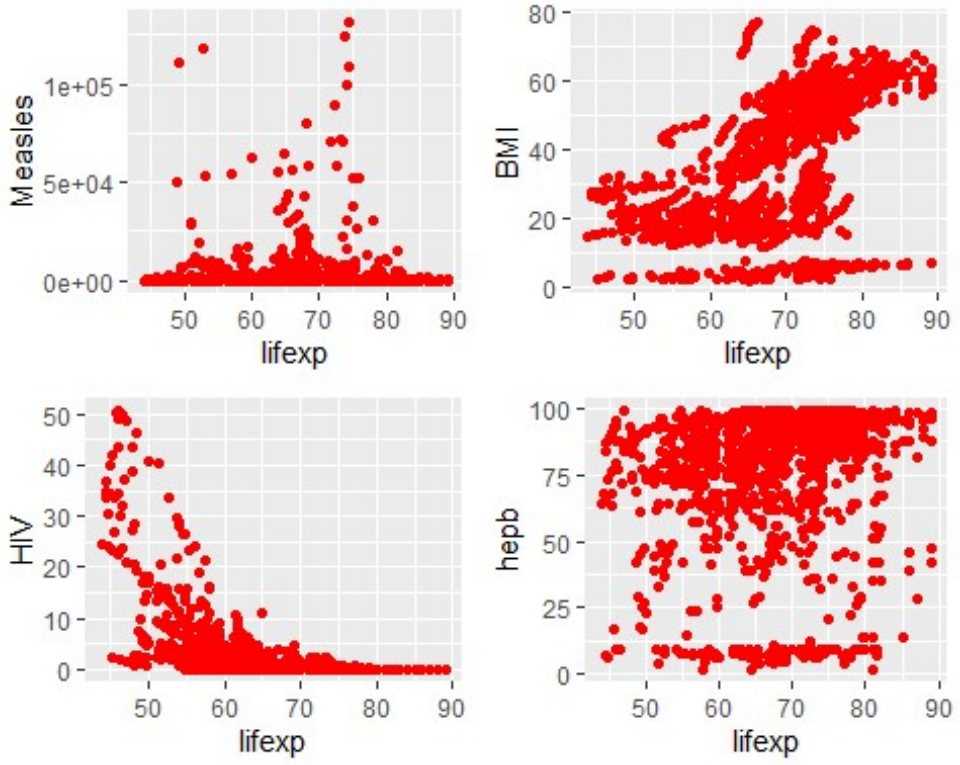

```r
#Plotting bar graph for life expectancy and Alcohol
x<-ggplot(data=expect, mapping=aes(x=lifexp,y=Alcohol))
x+geom_col(aes(color="red"))
```



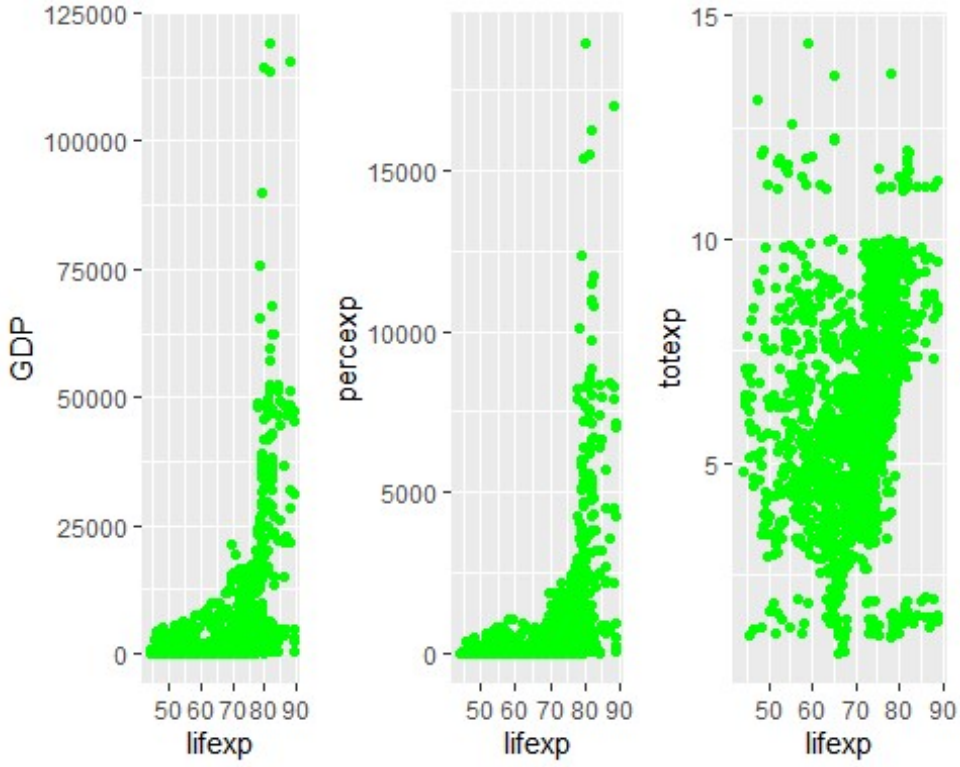

```
#ScatterPlot- Considering lifexp vs social, immunizations and economic
factors.
e1<-ggplot(expect,aes(x=lifexp,y=Alcohol))+geom_point(color='blue')
e2<-ggplot(expect,aes(x=lifexp,y=admort))+geom_point(color='blue')
e3<-ggplot(expect,aes(x=lifexp,y=infmort))+geom_point(color='blue')

e4<-ggplot(expect,aes(x=lifexp,y=Polio))+geom_point(color='red')
e5<-ggplot(expect,aes(x=lifexp,y=Measles))+geom_point(color='red')
e6<-ggplot(expect,aes(x=lifexp,y=BMI))+geom_point(color='red')
e7<-ggplot(expect,aes(x=lifexp,y=HIV))+geom_point(color='red')
e8<-ggplot(expect,aes(x=lifexp,y=hepb))+geom_point(color='red')

e9<-ggplot(expect,aes(x=lifexp,y=GDP))+geom_point(color='green')
e10<-ggplot(expect,aes(x=lifexp,y=percexp))+geom_point(color='green')
e11<-ggplot(expect,aes(x=lifexp,y=totexp))+geom_point(color='green')

grid.arrange(e1,e2,e3,ncol=3)
```




```
grid.arrange(e5,e6,e7,e8,ncol=2,nrow=2)
```




```
grid.arrange(e9,e10,e11,ncol=3)
```


GDP
125000
100000
75000
50000
25000
0
50 60 70 80 90
lifexp
percexp
15000
10000
5000
0
50 60 70 80 90
lifexp
totexp
15
10
5
50 60 70 80 90
lifexp