```
df=read.csv("data.csv")
attach(df)
```

```
library(readr)
library(tidyr)
library(dplyr)
```

```
##
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':
##
##     filter, lag
```

```
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
library(Hmisc)
```

```
## Loading required package: lattice
```

```
## Loading required package: survival
```

```
## Loading required package: Formula
```

```
## Loading required package: ggplot2
```

```
##
## Attaching package: 'Hmisc'
```

```
## The following objects are masked from 'package:dplyr':
##
##     src, summarize
```

```
## The following objects are masked from 'package:base':
##
##     format.pval, units
```

```
library(ggplot2)
library(ggpubr)
```

```
df$Country=as.factor(Country)
df$Status=as.factor(Status)
```

```
summary(df)
```

```
##       Year      Lifeexpectancy  AdultMortality   infantdeaths
##  Min.   :2000   Min.   :36.30   Min.   :  1.0   Min.   :   0.0
##  1st Qu.:2004   1st Qu.:63.20   1st Qu.: 74.0   1st Qu.:   0.0
##  Median :2008   Median :72.10   Median :144.0   Median :   3.0
##  Mean   :2008   Mean   :69.24   Mean   :164.6   Mean   :  30.3
##  3rd Qu.:2012   3rd Qu.:75.70   3rd Qu.:227.0   3rd Qu.:  22.0
##  Max.   :2015   Max.   :89.00   Max.   :723.0   Max.   :1800.0
##
##     Alcohol        percentageexpenditure   HepatitisB       Measles
##  Min.   : 0.0100   Min.   :    0.000     Min.   : 1.00   Min.   :     0.0
##  1st Qu.: 0.8925   1st Qu.:    4.685     1st Qu.:75.00   1st Qu.:     0.0
##  Median : 3.6850   Median :   64.913     Median :92.00   Median :    17.0
##  Mean   : 4.5460   Mean   :  738.251     Mean   :80.77   Mean   :  2419.6
##  3rd Qu.: 7.5500   3rd Qu.:  441.534     3rd Qu.:96.50   3rd Qu.:   360.2
##  Max.   :17.8700   Max.   :19479.912     Max.   :99.00   Max.   :212183.0
##
##       BMI        under.fivedeaths      Polio       Totalexpenditure
##  Min.   : 1.00   Min.   :   0.00   Min.   : 3.00   Min.   : 0.370
##  1st Qu.:19.20   1st Qu.:   0.00   1st Qu.:78.00   1st Qu.: 4.315
##  Median :43.05   Median :   4.00   Median :93.00   Median : 5.750
##  Mean   :38.18   Mean   :  42.04   Mean   :82.59   Mean   : 5.928
##  3rd Qu.:56.10   3rd Qu.:  28.00   3rd Qu.:97.00   3rd Qu.: 7.430
##  Max.   :87.30   Max.   :2500.00   Max.   :99.00   Max.   :17.600
##
##    Diphtheria       HIV.AIDS            GDP              Population
##  Min.   : 2.00   Min.   : 0.100   Min.   :     1.68   Min.   :3.400e+01
##  1st Qu.:78.00   1st Qu.: 0.100   1st Qu.:   429.15   1st Qu.:2.915e+05
##  Median :93.00   Median : 0.100   Median :  1537.81   Median :1.508e+06
##  Mean   :82.37   Mean   : 1.742   Mean   :  6765.71   Mean   :1.195e+07
##  3rd Qu.:97.00   3rd Qu.: 0.800   3rd Qu.:  5481.55   3rd Qu.:7.569e+06
##  Max.   :99.00   Max.   :50.600   Max.   :119172.74   Max.   :1.294e+09
##
##  thinness1.19years thinness5.9years Incomecompositionofresources
##  Min.   : 0.100    Min.   : 0.100   Min.   :0.0000
##  1st Qu.: 1.600    1st Qu.: 1.600   1st Qu.:0.4893
##  Median : 3.400    Median : 3.400   Median :0.6750
##  Mean   : 4.863    Mean   : 4.892   Mean   :0.6255
##  3rd Qu.: 7.200    3rd Qu.: 7.200   3rd Qu.:0.7780
##  Max.   :27.700    Max.   :28.600   Max.   :0.9480
##
##    Schooling                    Country            Status
##  Min.   : 0.00   Afghanistan        :  16   Developed : 512
##  1st Qu.:10.00   Albania            :  16   Developing:2426
##  Median :12.30   Algeria            :  16
##  Mean   :11.95   Angola             :  16
##  3rd Qu.:14.20   Antigua and Barbuda:  16
##  Max.   :20.70   Argentina          :  16
##                  (Other)            :2842
```

```
str(df)
```

```
## 'data.frame':    2938 obs. of  22 variables:
##  $ Year                           : num  2015 2014 2013 2012 2011 ...
##  $ Lifeexpectancy                 : num  65 59.9 59.9 59.5 59.2 58.8 58.6 58.1 57.5 57.3 ...
##  $ AdultMortality                 : num  263 271 268 272 275 279 281 287 295 295 ...
##  $ infantdeaths                   : num  62 64 66 69 71 74 77 80 82 84 ...
##  $ Alcohol                        : num  0.01 0.01 0.01 0.01 0.01 0.01 0.01 0.03 0.02 0.03
...
##  $ percentageexpenditure          : num  71.3 73.5 73.2 78.2 7.1 ...
##  $ HepatitisB                     : num  65 62 64 67 68 66 63 64 63 64 ...
##  $ Measles                        : num  1154 492 430 2787 3013 ...
##  $ BMI                            : num  19.1 18.6 18.1 17.6 17.2 16.7 16.2 15.7 15.2 14.7
...
##  $ under.fivedeaths               : num  83 86 89 93 97 102 106 110 113 116 ...
##  $ Polio                          : num  6 58 62 67 68 66 63 64 63 58 ...
##  $ Totalexpenditure               : num  8.16 8.18 8.13 8.52 7.87 9.2 9.42 8.33 6.73 7.43 ...
##  $ Diphtheria                     : num  65 62 64 67 68 66 63 64 63 58 ...
##  $ HIV.AIDS                       : num  0.1 0.1 0.1 0.1 0.1 0.1 0.1 0.1 0.1 0.1 ...
##  $ GDP                            : num  584.3 612.7 631.7 670 63.5 ...
##  $ Population                     : num  33736494 327582 31731688 3696958 2978599 ...
##  $ thinness1.19years              : num  17.2 17.5 17.7 17.9 18.2 18.4 18.6 18.8 19 19.2 ...
##  $ thinness5.9years               : num  17.3 17.5 17.7 18 18.2 18.4 18.7 18.9 19.1 19.3 ...
##  $ Incomecompositionofresources: num  0.479 0.476 0.47 0.463 0.454 0.448 0.434 0.433 0.415
0.405 ...
##  $ Schooling                      : num  10.1 10 9.9 9.8 9.5 9.2 8.9 8.7 8.4 8.1 ...
##  $ Country                        : Factor w/ 193 levels "Afghanistan",..: 1 1 1 1 1 1 1 1 1
1 ...
##  $ Status                         : Factor w/ 2 levels "Developed","Developing": 2 2 2 2 2 2
2 2 2 2 ...
```

anova test

```
data2 <- df %>% group_by (Country) %>% summarise(Average_life=mean(Lifeexpectancy),Average_sc
hooling=mean(Schooling))

attach(data2)
```

```
## The following object is masked from df:
##
##     Country
```

```
x<- data2 %>% filter(Average_schooling<=8.0)
y<-data2 %>% filter(Average_schooling>8.0 & Average_schooling<=12.0)
z<-data2 %>% filter(Average_schooling>12.0)
y1<-data.frame(Average_life = x$Average_life)
y1$Education ='Low'
y2<-data.frame(Average_life= y$Average_life)
y2$Education ='Middle'
y3<-data.frame(Average_life= z$Average_life)
y3$Education ='High'
combined_g <-data.frame(rbind(y1,y2,y3))
```

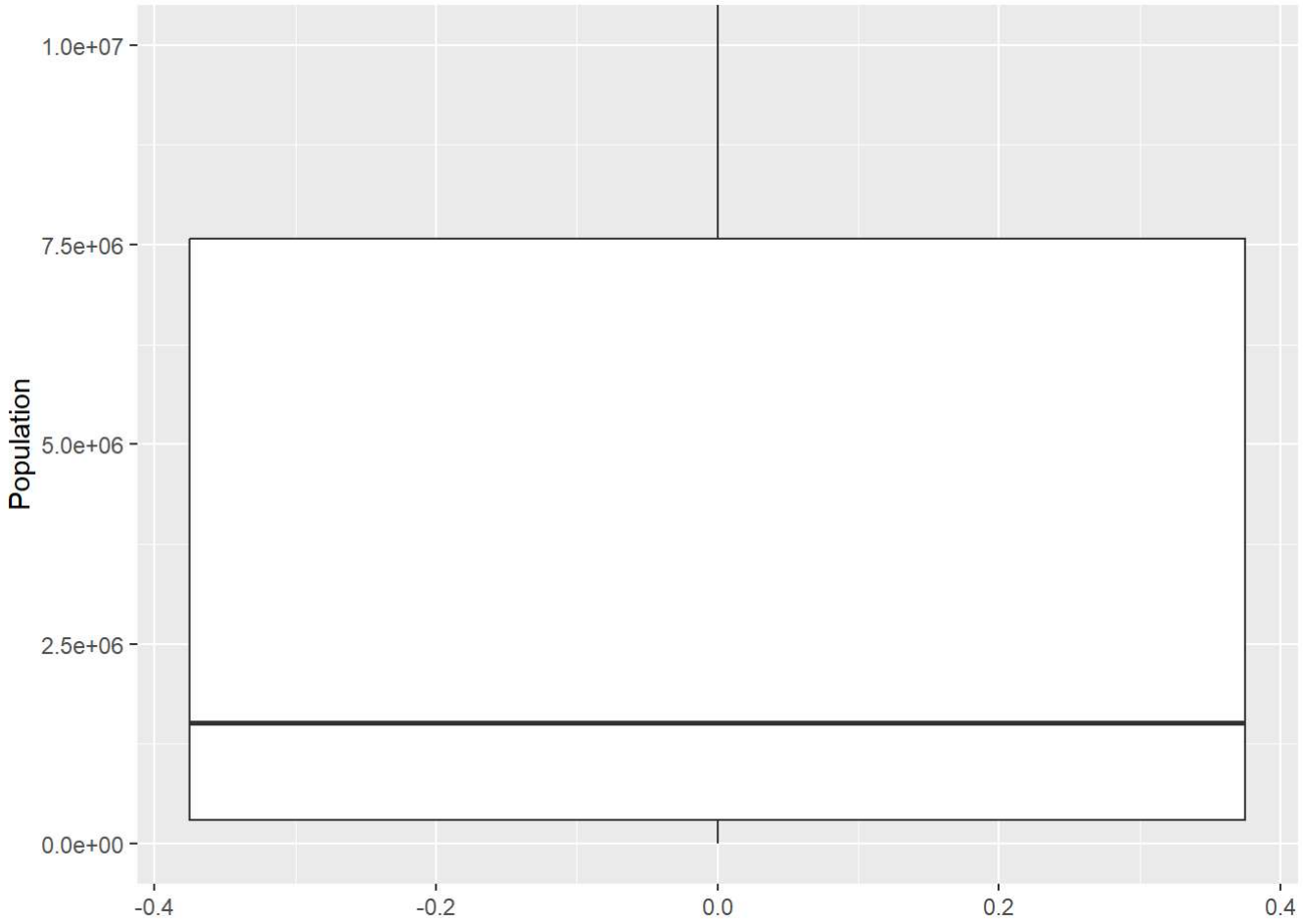
schooling using annova

```
Anova_Results <- aov(Average_life ~ Education, data= combined_g)
summary(Anova_Results)
```

```
##              Df Sum Sq Mean Sq F value Pr(>F)
## Education     2   8627    4313   114.4 <2e-16 ***
## Residuals   190   7166      38
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
box_plot_crop<-ggplot(df,aes(
  y=Population))
box_plot_crop+geom_boxplot(outliers.shape=NA)+coord_cartesian(ylim=c(5,9999999))
```

```
## Warning in geom_boxplot(outliers.shape = NA): Ignoring unknown parameters:
## `outliers.shape`
```




```
data3 <- df %>% group_by (Country) %>% summarise(Average_population=max(Population),Average_l
ife=mean(Lifeexpectancy))


attach(data3)
```

```
## The following objects are masked from data2:
##
##     Average_life, Country
```

```
## The following object is masked from df:
##
##     Country
```

```
mean(data3$Average_population)
```

```
## [1] 33996065
```

```
quantile(data3$Average_population,probs = c(.25,.5,.75),type=1)
```

```
##      25%      50%      75%
##  2287955  9897985 28656282
```

```
x<- data3 %>% filter(Average_population<=2287955)
y<-data3 %>% filter(Average_population>2287955 & Average_population<=9897985)
z<-data3 %>% filter(Average_population>9897985 & Average_population<=28656282)
z1<-data3 %>% filter(Average_population>28656282)
```

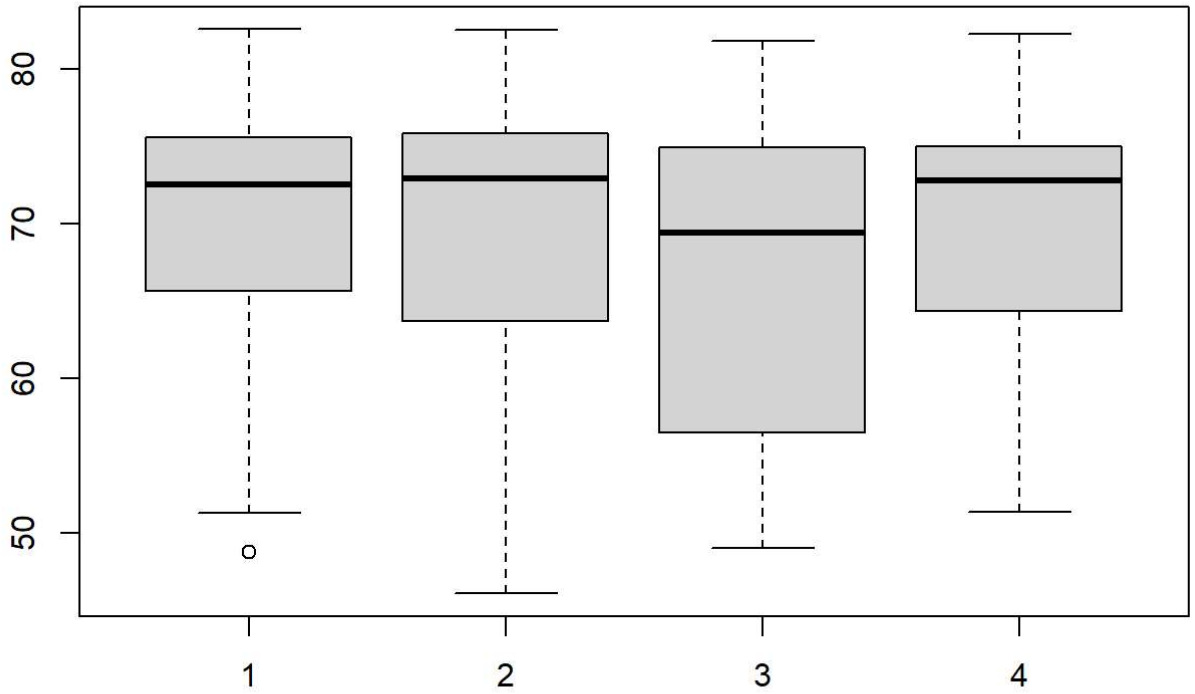
```
y1<-data.frame(Country=x$Country,Average_population = x$Average_population, Average_life=x$Av
erage_life)
y1$type_population ='Low'
y2<-data.frame(Country=y$Country,Average_population = y$Average_population, Average_life=y$Av
erage_life)
y2$type_population ='Middle'
y3<-data.frame(Country=z$Country,Average_population = z$Average_population, Average_life=z$Av
erage_life)
y3$type_population ='Dense'
y4<-data.frame(Country=z1$Country,Average_population = z1$Average_population, Average_life=z1
$Average_life)
y4$type_population='highly dense'
combined_g <-data.frame(rbind(y1,y2,y3,y4))
```

```
combined_g$type_population=as.factor(combined_g$type_population)
summary(combined_g)
```

```
##                 Country    Average_population   Average_life
##  Afghanistan        :  1   Min.   :2.920e+02   Min.   :46.11
##  Albania            :  1   1st Qu.:2.288e+06   1st Qu.:62.80
##  Algeria            :  1   Median :9.898e+06   Median :72.52
##  Angola             :  1   Mean   :3.400e+07   Mean   :69.46
##  Antigua and Barbuda:  1   3rd Qu.:2.866e+07   3rd Qu.:75.16
##  Argentina          :  1   Max.   :1.294e+09   Max.   :82.54
##  (Other)            :187
##      type_population
##  Dense       :48
##  highly dense:48
##  Low         :49
##  Middle      :48
##
##
##
```

As the population increases, we see that the median life expectancy decreases, as seen in the third boxplot. At first it seems counter intuitive as to why the expectancy increases in the fourth boxplot. But this is because the categories only refer to the population and not population density.

```
boxplot(y1$Average_life,y2$Average_life,y3$Average_life,y4$Average_life)
```



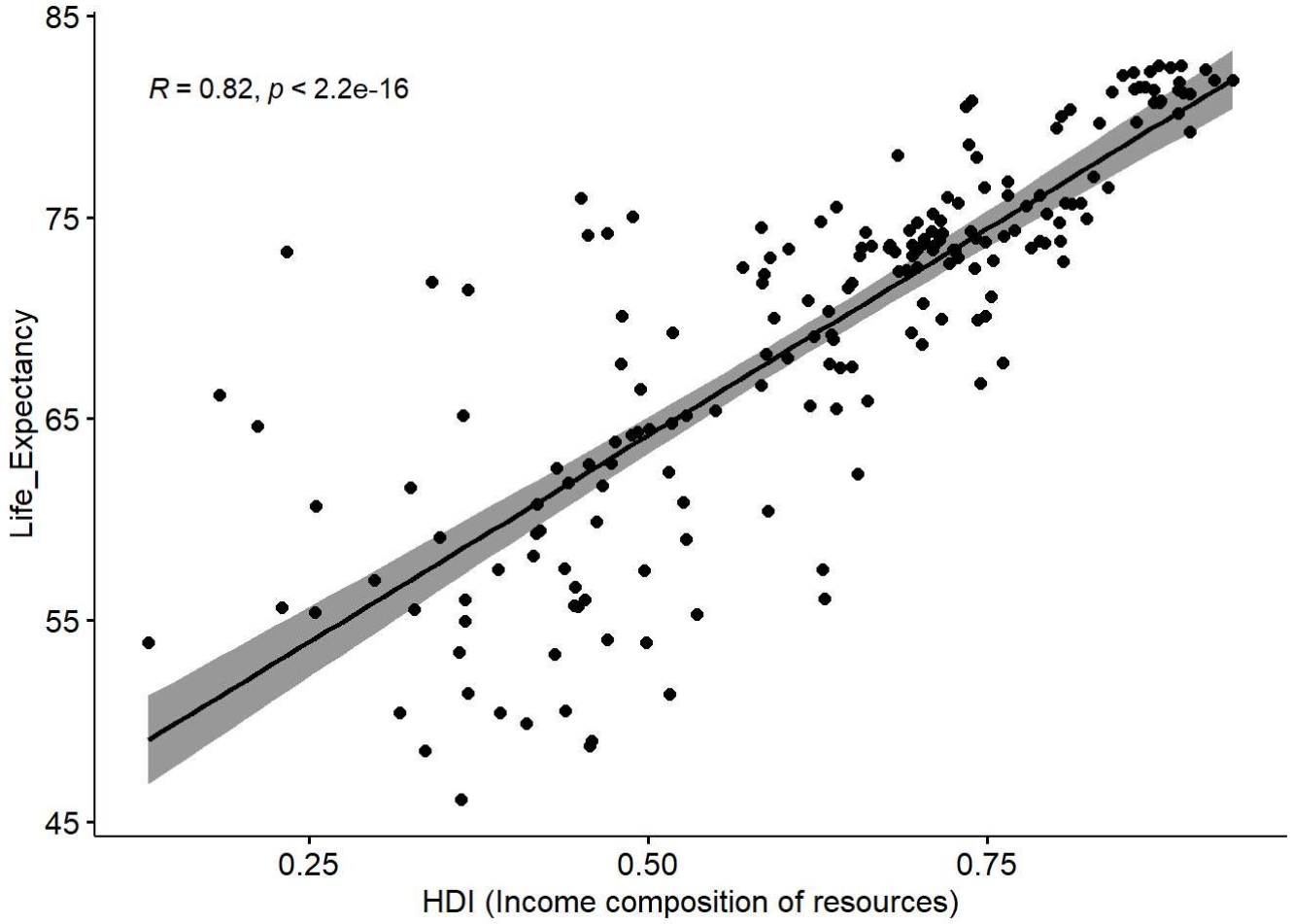
```
data1 <- df %>%  group_by (Country) %>% summarise(Average_HDI=mean(Incomecompositionofresources),Average_life=mean(Lifeexpectancy))


data1
```

```
## # A tibble: 193 × 3
##    Country             Average_HDI Average_life
##    <fct>                     <dbl>        <dbl>
##  1 Afghanistan               0.415         58.2
##  2 Albania                   0.710         75.2
##  3 Algeria                   0.695         73.6
##  4 Angola                    0.458         49.0
##  5 Antigua and Barbuda       0.489         75.1
##  6 Argentina                 0.794         75.2
##  7 Armenia                   0.698         73.4
##  8 Australia                 0.918         81.8
##  9 Austria                   0.862         81.5
## 10 Azerbaijan                0.703         70.7
## # … with 183 more rows
```

```
ggscatter(data1, x = "Average_HDI", y = "Average_life", add = "reg.line",
          conf.int = TRUE,
          cor.coef = TRUE, cor.method = "pearson",
          xlab = "HDI (Income composition of resources)", ylab = "Life_Expectancy")
```

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




```
cor.test(data1$Average_HDI, data1$Average_life)
```

```
##
##  Pearson's product-moment correlation
##
## data:  data1$Average_HDI and data1$Average_life
## t = 19.515, df = 191, p-value < 2.2e-16
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  0.7627543 0.8583782
## sample estimates:
##       cor
## 0.8160772
```

```
str(df)
```

```
## 'data.frame':    2938 obs. of  22 variables:
##  $ Year                           : num  2015 2014 2013 2012 2011 ...
##  $ Lifeexpectancy                 : num  65 59.9 59.9 59.5 59.2 58.8 58.6 58.1 57.5 57.3 ...
##  $ AdultMortality                 : num  263 271 268 272 275 279 281 287 295 295 ...
##  $ infantdeaths                   : num  62 64 66 69 71 74 77 80 82 84 ...
##  $ Alcohol                        : num  0.01 0.01 0.01 0.01 0.01 0.01 0.01 0.03 0.02 0.03
...
##  $ percentageexpenditure          : num  71.3 73.5 73.2 78.2 7.1 ...
##  $ HepatitisB                     : num  65 62 64 67 68 66 63 64 63 64 ...
##  $ Measles                        : num  1154 492 430 2787 3013 ...
##  $ BMI                            : num  19.1 18.6 18.1 17.6 17.2 16.7 16.2 15.7 15.2 14.7
...
##  $ under.fivedeaths               : num  83 86 89 93 97 102 106 110 113 116 ...
##  $ Polio                          : num  6 58 62 67 68 66 63 64 63 58 ...
##  $ Totalexpenditure               : num  8.16 8.18 8.13 8.52 7.87 9.2 9.42 8.33 6.73 7.43 ...
##  $ Diphtheria                     : num  65 62 64 67 68 66 63 64 63 58 ...
##  $ HIV.AIDS                       : num  0.1 0.1 0.1 0.1 0.1 0.1 0.1 0.1 0.1 0.1 ...
##  $ GDP                            : num  584.3 612.7 631.7 670 63.5 ...
##  $ Population                     : num  33736494 327582 31731688 3696958 2978599 ...
##  $ thinness1.19years              : num  17.2 17.5 17.7 17.9 18.2 18.4 18.6 18.8 19 19.2 ...
##  $ thinness5.9years               : num  17.3 17.5 17.7 18 18.2 18.4 18.7 18.9 19.1 19.3 ...
##  $ Incomecompositionofresources: num  0.479 0.476 0.47 0.463 0.454 0.448 0.434 0.433 0.415
0.405 ...
##  $ Schooling                      : num  10.1 10 9.9 9.8 9.5 9.2 8.9 8.7 8.4 8.1 ...
##  $ Country                        : Factor w/ 193 levels "Afghanistan",..: 1 1 1 1 1 1 1 1 1
1 ...
##  $ Status                         : Factor w/ 2 levels "Developed","Developing": 2 2 2 2 2 2
2 2 2 2 ...
```

```
data5<- df %>% group_by(Country) %>% summarise(Average_life = mean(Lifeexpectancy), Average_P
olio = mean(Polio), Average_Diphtheria = mean(Diphtheria),Average_Hepatitis=mean(HepatitisB))
x1<- data5 %>% filter( Average_Polio <=85)
x2<- data5 %>% filter( Average_Polio >85)
y1<- data5 %>% filter( Average_Diphtheria <=85)
y2<- data5 %>% filter( Average_Diphtheria >85)
z1<- data5 %>% filter(Average_Hepatitis<=85)
z2<- data5 %>% filter(Average_Hepatitis>85)
a1<-data.frame(Average_life = x1$Average_life, Country  = x1$Country)
a1$Polio ='Low'
a2<-data.frame(Average_life= x2$Average_life, Country  = x2$Country)
a2$Polio ='High'
df1 <-data.frame(rbind(a1,a2))


b1<-data.frame(Average_life = y1$Average_life, Country  = y1$Country)
b1$Diphtheria ='Low'
b2<-data.frame(Average_life= y2$Average_life, Country  = y2$Country)
b2$Diphtheria ='High'

c1<-data.frame(Average_life=z1$Average_life,Country=z1$Country)
c1$Hepatitis="low"
c2<-data.frame(Average_life=z2$Average_life,Country=z2$Country)
c2$Hepatitis="high"

df2 <-data.frame(rbind(b1,b2))
df3<-data.frame(rbind(c1,c2))
```

```
df_final <- merge(df1,df2, by = "Country")
df_final <- merge(df_final,df3, by = "Country")
Anova_Results <- aov(Average_life.x ~ Polio + Diphtheria , data= df_final)
summary(Anova_Results)
```

```
##              Df Sum Sq Mean Sq F value   Pr(>F)
## Polio         1   4855    4855   98.59  < 2e-16 ***
## Diphtheria    1   1581    1581   32.10 5.35e-08 ***
## Residuals   190   9357      49
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
Anova_Results <- aov(Average_life ~ Hepatitis, data= df3)
summary(Anova_Results)
```

```
##              Df Sum Sq Mean Sq F value   Pr(>F)
## Hepatitis     1   2865  2865.0   42.33 6.61e-10 ***
## Residuals   191  12928    67.7
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
levels(df$Country)
```

```
##   [1] "Afghanistan"
##   [2] "Albania"
##   [3] "Algeria"
##   [4] "Angola"
##   [5] "Antigua and Barbuda"
##   [6] "Argentina"
##   [7] "Armenia"
##   [8] "Australia"
##   [9] "Austria"
##  [10] "Azerbaijan"
##  [11] "Bahamas"
##  [12] "Bahrain"
##  [13] "Bangladesh"
##  [14] "Barbados"
##  [15] "Belarus"
##  [16] "Belgium"
##  [17] "Belize"
##  [18] "Benin"
##  [19] "Bhutan"
##  [20] "Bolivia (Plurinational State of)"
##  [21] "Bosnia and Herzegovina"
##  [22] "Botswana"
##  [23] "Brazil"
##  [24] "Brunei Darussalam"
##  [25] "Bulgaria"
##  [26] "Burkina Faso"
##  [27] "Burundi"
##  [28] "Cabo Verde"
##  [29] "Cambodia"
##  [30] "Cameroon"
##  [31] "Canada"
##  [32] "Central African Republic"
##  [33] "Chad"
##  [34] "Chile"
##  [35] "China"
##  [36] "Colombia"
##  [37] "Comoros"
##  [38] "Congo"
##  [39] "Cook Islands"
##  [40] "Costa Rica"
##  [41] "Côte d'Ivoire"
##  [42] "Croatia"
##  [43] "Cuba"
##  [44] "Cyprus"
##  [45] "Czechia"
##  [46] "Democratic People's Republic of Korea"
##  [47] "Democratic Republic of the Congo"
##  [48] "Denmark"
##  [49] "Djibouti"
##  [50] "Dominica"
##  [51] "Dominican Republic"
##  [52] "Ecuador"
##  [53] "Egypt"
##  [54] "El Salvador"
##  [55] "Equatorial Guinea"
```

```
##  [56] "Eritrea"
##  [57] "Estonia"
##  [58] "Ethiopia"
##  [59] "Fiji"
##  [60] "Finland"
##  [61] "France"
##  [62] "Gabon"
##  [63] "Gambia"
##  [64] "Georgia"
##  [65] "Germany"
##  [66] "Ghana"
##  [67] "Greece"
##  [68] "Grenada"
##  [69] "Guatemala"
##  [70] "Guinea"
##  [71] "Guinea-Bissau"
##  [72] "Guyana"
##  [73] "Haiti"
##  [74] "Honduras"
##  [75] "Hungary"
##  [76] "Iceland"
##  [77] "India"
##  [78] "Indonesia"
##  [79] "Iran (Islamic Republic of)"
##  [80] "Iraq"
##  [81] "Ireland"
##  [82] "Israel"
##  [83] "Italy"
##  [84] "Jamaica"
##  [85] "Japan"
##  [86] "Jordan"
##  [87] "Kazakhstan"
##  [88] "Kenya"
##  [89] "Kiribati"
##  [90] "Kuwait"
##  [91] "Kyrgyzstan"
##  [92] "Lao People's Democratic Republic"
##  [93] "Latvia"
##  [94] "Lebanon"
##  [95] "Lesotho"
##  [96] "Liberia"
##  [97] "Libya"
##  [98] "Lithuania"
##  [99] "Luxembourg"
## [100] "Madagascar"
## [101] "Malawi"
## [102] "Malaysia"
## [103] "Maldives"
## [104] "Mali"
## [105] "Malta"
## [106] "Marshall Islands"
## [107] "Mauritania"
## [108] "Mauritius"
## [109] "Mexico"
## [110] "Micronesia (Federated States of)"
## [111] "Monaco"
```

```
## [112] "Mongolia"
## [113] "Montenegro"
## [114] "Morocco"
## [115] "Mozambique"
## [116] "Myanmar"
## [117] "Namibia"
## [118] "Nauru"
## [119] "Nepal"
## [120] "Netherlands"
## [121] "New Zealand"
## [122] "Nicaragua"
## [123] "Niger"
## [124] "Nigeria"
## [125] "Niue"
## [126] "Norway"
## [127] "Oman"
## [128] "Pakistan"
## [129] "Palau"
## [130] "Panama"
## [131] "Papua New Guinea"
## [132] "Paraguay"
## [133] "Peru"
## [134] "Philippines"
## [135] "Poland"
## [136] "Portugal"
## [137] "Qatar"
## [138] "Republic of Korea"
## [139] "Republic of Moldova"
## [140] "Romania"
## [141] "Russian Federation"
## [142] "Rwanda"
## [143] "Saint Kitts and Nevis"
## [144] "Saint Lucia"
## [145] "Saint Vincent and the Grenadines"
## [146] "Samoa"
## [147] "San Marino"
## [148] "Sao Tome and Principe"
## [149] "Saudi Arabia"
## [150] "Senegal"
## [151] "Serbia"
## [152] "Seychelles"
## [153] "Sierra Leone"
## [154] "Singapore"
## [155] "Slovakia"
## [156] "Slovenia"
## [157] "Solomon Islands"
## [158] "Somalia"
## [159] "South Africa"
## [160] "South Sudan"
## [161] "Spain"
## [162] "Sri Lanka"
## [163] "Sudan"
## [164] "Suriname"
## [165] "Swaziland"
## [166] "Sweden"
## [167] "Switzerland"
```

```
## [168] "Syrian Arab Republic"
## [169] "Tajikistan"
## [170] "Thailand"
## [171] "The former Yugoslav republic of Macedonia"
## [172] "Timor-Leste"
## [173] "Togo"
## [174] "Tonga"
## [175] "Trinidad and Tobago"
## [176] "Tunisia"
## [177] "Turkey"
## [178] "Turkmenistan"
## [179] "Tuvalu"
## [180] "Uganda"
## [181] "Ukraine"
## [182] "United Arab Emirates"
## [183] "United Kingdom of Great Britain and Northern Ireland"
## [184] "United Republic of Tanzania"
## [185] "United States of America"
## [186] "Uruguay"
## [187] "Uzbekistan"
## [188] "Vanuatu"
## [189] "Venezuela (Bolivarian Republic of)"
## [190] "Viet Nam"
## [191] "Yemen"
## [192] "Zambia"
## [193] "Zimbabwe"
```