visulization and correlation

2023-08-27

```r
# import the csv to datframe
df1 <- read.csv('1519data.csv')

# check the dataframe structure
head(df1)
```

```
##   X   Country Code        Region  X2015_ghg  X2016_ghg  X2017_ghg
## X2018_ghg
## 1 0 Argentina  ARG  SouthAmerica  322554.61  322554.61  322554.61
## 322554.61
## 2 1 Australia  AUS       Oceania  544231.81  552354.83  559581.11
## 560827.41
## 3 2   Austria  AUT WesternEurope   78884.46   79821.26   82132.49
## 78854.37
## 4 3   Belarus  BLR EasternEurope   88307.46   87563.30   89247.60
## 92202.41
## 5 4   Belgium  BEL WesternEurope  118990.38  117419.57  116909.24
## 117584.92
## 6 5    Brazil  BRA  SouthAmerica 1041676.99 1014702.38 1014702.38
## 1014702.38
##    X2019_ghg    X2015_gdp    X2016_gdp    X2017_gdp    X2018_gdp
## X2019_gdp
## 1  322554.61 5.947493e+11 5.575314e+11 6.436287e+11 5.248197e+11
## 4.519324e+11
## 2  555244.93 1.350534e+12 1.206685e+12 1.326883e+12 1.428530e+12
## 1.391953e+12
## 3   79994.14 3.818176e+11 3.955686e+11 4.162305e+11 4.549459e+11
## 4.450119e+11
## 4   92291.63 5.645473e+10 4.772266e+10 5.472660e+10 6.003126e+10
## 6.440965e+10
## 5  116463.71 4.621497e+11 4.757396e+11 5.015229e+11 5.430085e+11
## 5.352887e+11
## 6 1014702.38 1.802212e+12 1.795693e+12 2.063515e+12 1.916934e+12
## 1.877824e+12
##       Status     ave_GHG      ave_GDP
## 1 Developing  322554.61 5.545323e+11
## 2  Developed  554448.02 1.340917e+12
## 3  Developed   79937.34 4.187149e+11
## 4 Developing   89922.48 5.666898e+10
## 5  Developed  117473.56 5.035419e+11
## 6 Developing 1020097.30 1.891236e+12
```

```r
## We only select the country, avg_ghg and avg_gdp
#The index column is 2, 16 ,17
```

```r
selected_df<-df1[c(2,16,17)]

# display the dataframe
selected_df
```

```
##                          Country      ave_GHG      ave_GDP
## 1                      Argentina   322554.610 5.545323e+11
## 2                      Australia   554448.018 1.340917e+12
## 3                        Austria    79937.344 4.187149e+11
## 4                        Belarus    89922.480 5.666898e+10
## 5                        Belgium   117473.564 5.035419e+11
## 6                         Brazil  1020097.302 1.891236e+12
## 7                       Bulgaria    57442.952 5.984284e+10
## 8                         Canada   717675.002 1.640223e+12
## 9                          Chile   107819.301 2.696702e+11
## 10 China (People's Republic of) 12300200.000 1.255600e+13
## 11                      Colombia   175751.182 3.091637e+11
## 12                    Costa Rica    14254.180 6.045960e+10
## 13                       Croatia    24910.936 5.663337e+10
## 14                        Cyprus     8762.456 2.298954e+10
## 15                Czech Republic   128439.336 2.208764e+11
## 16                       Denmark    50436.474 3.304625e+11
## 17                       Estonia    18607.956 2.706323e+10
## 18                       Finland    55309.810 2.548306e+11
## 19                        France   454580.824 2.603340e+12
## 20                       Germany   863521.064 3.673828e+12
## 21                        Greece    92672.104 2.010125e+11
## 22                       Hungary    63955.100 1.442192e+11
## 23                       Iceland     4776.452 2.283269e+10
## 24                         India  2816004.988 2.524295e+12
## 25                     Indonesia   888046.000 9.939426e+11
## 26                          Iran   483669.170 3.601765e+11
## 27                       Ireland    62810.614 3.417532e+11
## 28                        Israel    78528.360 3.491912e+11
## 29                         Italy   436389.198 1.953788e+12
## 30                         Japan  1271035.472 4.913024e+12
## 31                    Kazakhstan   359359.694 1.698959e+11
## 32                         Korea   705030.186 1.593211e+12
## 33                        Latvia    10933.440 3.088923e+10
## 34                     Lithuania    20195.640 4.809995e+10
## 35                    Luxembourg    10390.940 6.584357e+10
## 36                         Malta     2038.446 1.322294e+10
## 37                        Mexico   758131.188 1.180223e+12
## 38                        Monaco       85.530 6.748740e+09
## 39                   Netherlands   189669.614 8.408789e+11
## 40                   New Zealand    78648.940 1.998151e+11
## 41                        Norway    52829.970 3.989927e+11
## 42                          Peru    95864.461 2.087508e+11
## 43                        Poland   396420.818 5.323287e+11
## 44                      Portugal    67563.016 2.217185e+11
```

```
## 45                   Romania   117153.636 2.137785e+11
## 46                    Russia  2086341.394 1.511849e+12
## 47              Saudi Arabia   548262.550 7.134561e+11
## 48            Slovak Republic   41367.938 9.684377e+10
## 49                  Slovenia    17493.378 4.892232e+10
## 50              South Africa   379837.160 3.689042e+11
## 51                     Spain   325739.178 1.309968e+12
## 52                    Sweden    52395.908 5.302224e+11
## 53               Switzerland    47914.446 7.139071e+11
## 54                   Türkiye   507295.054 8.264964e+11
## 55                   Ukraine   330798.534 1.162503e+11
## 56            United Kingdom   480033.462 2.831582e+12
## 57             United States  6650072.794 1.971429e+13
```

```r
# Load the tidycerse package
library(tidyverse)
```

```
## Warning: package 'tidyverse' was built under R version 4.2.3

## Warning: package 'ggplot2' was built under R version 4.2.3

## Warning: package 'tibble' was built under R version 4.2.3

## Warning: package 'tidyr' was built under R version 4.2.3

## Warning: package 'readr' was built under R version 4.2.3

## Warning: package 'purrr' was built under R version 4.2.3

## Warning: package 'dplyr' was built under R version 4.2.3

## Warning: package 'stringr' was built under R version 4.2.3

## Warning: package 'forcats' was built under R version 4.2.3

## Warning: package 'lubridate' was built under R version 4.2.3

## ── Attaching core tidyverse packages ───────────────────── tidyverse
2.0.0 ──
## ✓ dplyr     1.1.2     ✓ readr     2.1.4
## ✓ forcats   1.0.0     ✓ stringr   1.5.0
## ✓ ggplot2   3.4.3     ✓ tibble    3.2.1
## ✓ lubridate 1.9.2     ✓ tidyr     1.3.0
## ✓ purrr     1.0.2
## ── Conflicts ──────────────────────────────────────
tidyverse_conflicts() ──
## ✗ dplyr::filter() masks stats::filter()
## ✗ dplyr::lag()    masks stats::lag()
## ℹ Use the conflicted package (<http://conflicted.r-lib.org/>) to force all
conflicts to become errors
```
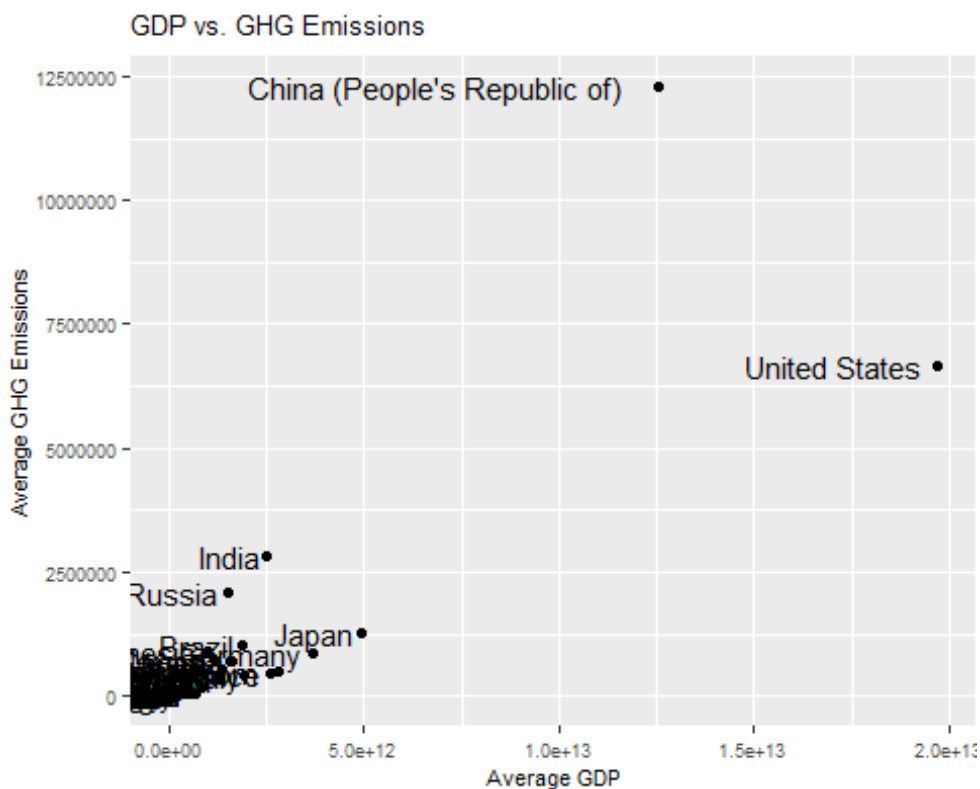
```
# we create the plot by using ggplot
ggplot(selected_df,
        #x data and y data
        aes(x = ave_GDP, y = ave_GHG)) +
  #scatter plot
  geom_point() +
  geom_text(aes(label = Country), hjust = 1.1) +
  labs(
    title = "GDP vs. GHG Emissions",
    x = "Average GDP",
    y = "Average GHG Emissions"
  ) +
  theme(
    plot.margin = unit(rep(0,4), 'cm'),
    text = element_text(size = 8)
  )
```


GDP vs. GHG Emissions

```
# import the region pivot table in long form (region)

region_df<- read.csv('pivot_region_long.csv')

#modify the data.
data_modify <- data.frame(
  region = region_df$Region,
  year = region_df$Year,
  #split one column to gdp and ghg
  GDP = region_df$Value[which(region_df$Category == 'gdp')],
```

```
    GHG = region_df$Value[which(region_df$Category == 'ghg')]
)

ggplot(data_modify, aes(x = GDP, y = GHG, color = region)) +
  geom_line(size = 1.2) +
  labs(title = "GDP vs. GHG by Region",
       x = "GDP",
       y = "GHG") +
  theme_minimal() +
  guides(color = guide_legend(override.aes = list(size = 2)))
```

```
## Warning: Using `size` aesthetic for lines was deprecated in ggplot2 3.4.0.
## i Please use `linewidth` instead.
## This warning is displayed once every 8 hours.
## Call `lifecycle::last_lifecycle_warnings()` to see where this warning was
## generated.
```
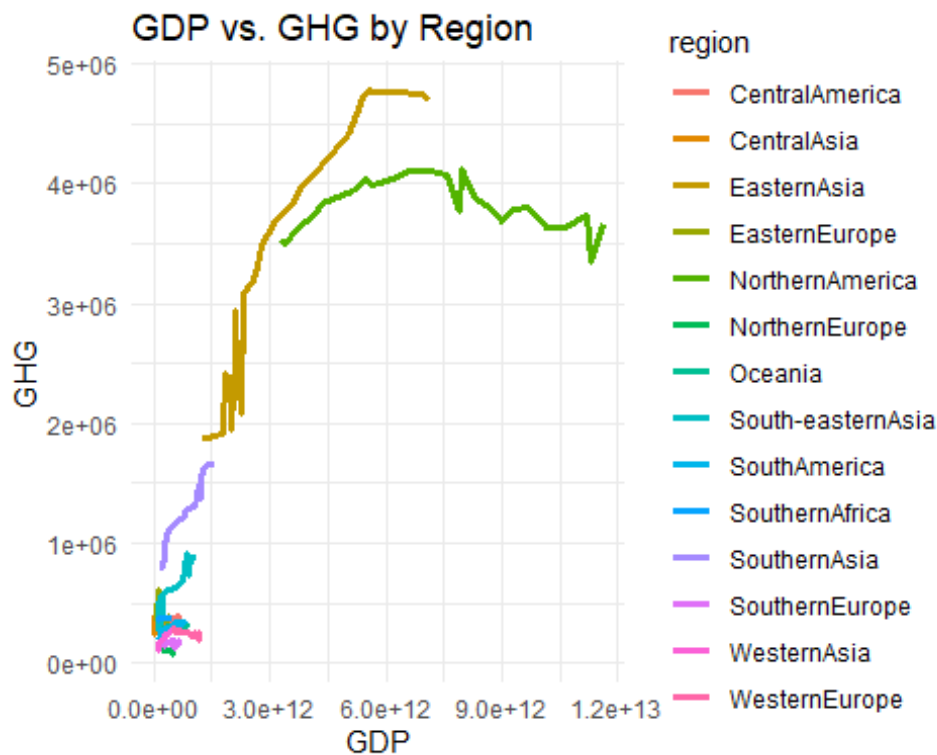


```
# import the pivot developing status data
#read the csv
deve.df<-read.csv('pivot_deve_long.csv')

#check the data structure
head(deve.df)
```

```
##   X    Status Year Category    Value
## 1 0  Developed 1990      ghg 428516.3
## 2 1 Developing 1990      ghg 555733.8
```

```
## 3 2  Developed 1991      ghg 422988.4
## 4 3 Developing 1991      ghg 550720.0
## 5 4  Developed 1992      ghg 422372.0
## 6 5 Developing 1992      ghg 532177.5

data_modify2 <- data.frame(
  Status = deve.df$Status,
  year = deve.df$Year,
    #split one column to gdp and ghg
  GDP = deve.df$Value[which(deve.df$Category == 'gdp')],
  GHG = deve.df$Value[which(deve.df$Category == 'ghg')]
)

ggplot(data_modify2, aes(x = GDP, y = GHG, color = Status)) +
  geom_line(size = 1.2) +
  labs(title = "GDP vs. GHG by Status",
       x = "GDP",
       y = "GHG") +
  theme_minimal() +
  guides(color = guide_legend(override.aes = list(size = 2)))
```
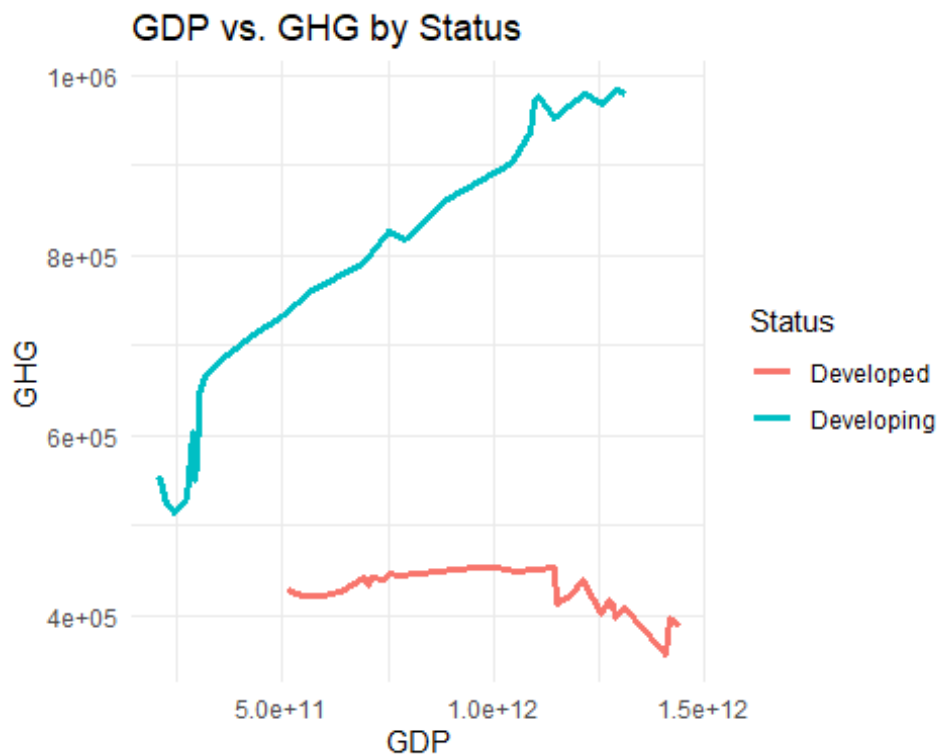


```
#spearman correlation

#read the csv
merged1<- read.csv('merged1.csv')
```

```r
#check the dataframe structure
head(merged1)
```

```
##   X   Country Code       Region X1990_ghg X1991_ghg X1992_ghg X1993_ghg
## 1 0 Argentina  ARG  SouthAmerica 220240.77 230413.76 234122.38 239560.14
## 2 1 Australia  AUS       Oceania 438056.76 438049.29 441752.18 442282.48
## 3 2   Austria  AUT WesternEurope  79047.23  82711.16  76142.81  76517.86
## 4 3   Belarus  BLR EasternEurope 145339.89 139713.44 126031.22 112296.43
## 5 4   Belgium  BEL WesternEurope 145844.47 148579.58 148077.75 146896.93
## 6 5    Brazil  BRA  SouthAmerica 552640.82 574601.49 583848.35 596099.40
##   X1994_ghg X1995_ghg X1996_ghg X1997_ghg X1998_ghg X1999_ghg X2000_ghg
## 1 243598.69 244499.37 256463.93 258303.62 262725.06 264637.84 280308.99
## 2 442610.56 451076.46 457574.99 469835.69 484037.12 489971.16 501588.99
## 3  76725.97  79953.24  83112.93  82719.63  82012.90  80455.15  80619.36
## 4  96343.45  87955.99  88760.16  89022.10  87757.93  84261.45  81145.92
## 5 151445.45 153579.82 157227.95 148820.67 154012.89 147732.82 148877.33
## 6 613905.50 638769.90 643708.82 668029.60 685872.35 702927.57 725162.14
##   X2001_ghg X2002_ghg X2003_ghg X2004_ghg X2005_ghg X2006_ghg X2007_ghg
## 1 276902.60 269315.64 286381.63  298167.4 307492.67 321462.03 335854.13
## 2 509456.08 513171.01 511714.21  528827.4 535589.76 540017.78 546739.52
## 3  84514.05  86267.54  91636.30   91427.9  92588.58  90158.67  87377.95
## 4  79825.86  79125.28  80871.08   84545.8  86538.40  89856.62  88250.97
## 5 147319.08 147415.07 147731.49  148523.9 145418.35 142715.60 139090.79
## 6 744084.15 760963.10 778785.71  815887.2 830176.87 837355.38 843618.61
##   X2008_ghg X2009_ghg X2010_ghg X2011_ghg X2012_ghg  X2013_ghg  X2014_ghg
## 1 331802.17 322554.23 300695.19 310314.53 310670.24  321011.89  322554.61
## 2 549128.82 552127.35 547172.76 549007.35 552305.55  543798.68  535451.44
## 3  86770.82  80137.03  84693.29  82505.76  79788.32   80228.52   76662.66
## 4  91672.82  89716.11  91884.27  91314.47  92432.61   94456.41   93371.44
## 5 139030.91 126438.52 133644.30 123192.99 120443.61  120523.72  114878.85
## 6 872592.78 858585.47 921055.10 954163.29 985944.61 1030985.67 1062897.04
##    X2015_ghg   X2016_ghg   X2017_ghg   X2018_ghg   X2019_ghg   X2020_ghg
## 1  322554.61   322554.61   322554.61   322554.61   322554.61   322554.61
## 2  544231.81   552354.83   559581.11   560827.41   555244.93   536739.72
## 3   78884.46    79821.26    82132.49    78854.37    79994.14    73910.84
## 4   88307.46    87563.30    89247.60    92202.41    92291.63    89940.28
## 5  118990.38   117419.57   116909.24   117584.92   116463.71   107272.65
## 6 1041676.99 1014702.38 1014702.38 1014702.38 1014702.38 1014702.38
##       X1990_gdp    X1991_gdp    X1992_gdp    X1993_gdp    X1994_gdp
## X1995_gdp
## 1 141352368715 189719984268 228788617202 236741715015 257440000000
## 258031750000
## 2 311326664102 325902990346 325480273783 312126194491 322807333103
## 367915800416
## 3 166463386179 173794177961 195078126722 190379720809 203535242742
## 241038283063
## 4  21650000000  20685750000  19721500000  18757250000  17793000000
## 13489222222
## 5 205331747948 210510999409 234781652447 224721795709 244884129491
## 288025588396
```

```
## 6 390725626003 342609227210 328187944301 368295777770 525369851354
769333330369
##         X1996_gdp     X1997_gdp     X1998_gdp     X1999_gdp     X2000_gdp
X2001_gdp
## 1 272149750000 292859000000 298948250000 283523000000 284203750000
268696750000
## 2 401089529590 435323994365 399404463135 389098884572 415576210513
379083932596
## 3 237250948791 212790348405 218259904402 217185787343 196799778883
197337879195
## 4  14756846154  14108846154  15264369565  12152867470  12738912134
12354820144
## 5 279201433225 252708051421 258528339631 258158533987 236204532891
236541297539
## 6 850426433004 883206452795 863711007325 599642075004 655448188259
559983704094
##         X2002_gdp     X2003_gdp     X2004_gdp     X2005_gdp     X2006_gdp
X2007_gdp
## 1  97724004252 127586973492 164657930453 198737095012 2.325573e+11
2.875305e+11
## 2 395342716618 467390797903 614166310997 695075176665 7.475562e+11
8.539554e+11
## 3 213377771504 261695778781 300904221505 315974418605 3.359986e+11
3.886914e+11
## 4  14594249023  17827791321  23144351852  30207567317 3.695431e+10
4.527740e+10
## 5 257157820440 317381715576 368537000248 385570948887 4.079181e+11
4.703243e+11
## 6 509795270685 558233724165 669289321945 891633826603 1.107627e+12
1.397114e+12
##         X2008_gdp     X2009_gdp     X2010_gdp     X2011_gdp     X2012_gdp
X2013_gdp
## 1 3.615580e+11 3.329765e+11 4.236274e+11 5.301633e+11 5.459824e+11
5.520251e+11
## 2 1.055127e+12 9.280430e+11 1.147589e+12 1.397908e+12 1.546509e+12
1.576335e+12
## 3 4.302943e+11 4.001723e+11 3.918927e+11 4.311203e+11 4.094252e+11
4.300687e+11
## 4 6.076348e+10 5.087408e+10 5.722249e+10 6.175779e+10 6.568510e+10
7.552798e+10
## 5 5.152235e+11 4.813459e+11 4.809516e+11 5.226455e+11 4.961813e+11
5.216427e+11
## 6 1.695855e+12 1.666996e+12 2.208838e+12 2.616157e+12 2.465228e+12
2.472819e+12
##         X2014_gdp     X2015_gdp     X2016_gdp     X2017_gdp     X2018_gdp
X2019_gdp
## 1 5.263197e+11 5.947493e+11 5.575314e+11 6.436287e+11 5.248197e+11
4.519324e+11
## 2 1.467505e+12 1.350534e+12 1.206685e+12 1.326883e+12 1.428530e+12
1.391953e+12
```

```
## 3 4.419961e+11 3.818176e+11 3.955686e+11 4.162305e+11 4.549459e+11
4.450119e+11
## 4 7.881384e+10 5.645473e+10 4.772266e+10 5.472660e+10 6.003126e+10
6.440965e+10
## 5 5.346781e+11 4.621497e+11 4.757396e+11 5.015229e+11 5.430085e+11
5.352887e+11
## 6 2.456044e+12 1.802212e+12 1.795693e+12 2.063515e+12 1.916934e+12
1.877824e+12
##        X2020_gdp      Status
## 1 3.892881e+11 Developing
## 2 1.327836e+12  Developed
## 3 4.332585e+11  Developed
## 4 6.025824e+10 Developing
## 5 5.218613e+11  Developed
## 6 1.444733e+12 Developing
```

```r
# select the ghg data and gdp data
# to have a check
ghg_ex<-merged1[5:35]
gdp_ex<-merged1[36:66]



# get the mean by column direction
# by using sapply
ghg_ex<-sapply(merged1[5:35],mean)
gdp_ex<-sapply(merged1[36:66],mean)



cor.test(gdp_ex, ghg_ex, method="spearman")
```

```
##
##   Spearman's rank correlation rho
##
## data:  gdp_ex and ghg_ex
## S = 194, p-value < 2.2e-16
## alternative hypothesis: true rho is not equal to 0
## sample estimates:
##        rho
## 0.9608871
```