Green Houses Gases

Satish Gollu and Sameer Vaddhiparthy

12/04/2021

International Greenhouse Gas Emissions Data Analysis

The Greenhouse Gas (GHG) Inventory Data contains the most recently submitted information of all the countries, covering the period from 1990 to 2018. The GHG data contain information on anthropogenic emissions that includes the burning of fossil fuels, deforestation, land use changes, livestock, fertilization, etc., that result in a net increase in emissions by sources and removals by sinks of the following GHGs (carbon dioxide (CO_2), methane (CH_4), nitrous oxide (N_2O), hydrofluorocarbons (HFCs), perfluorocarbons (PFCs), unspecified mix of HFCs and PFCs, sulphur hexafluoride (SF_6) and nitrogen triflouride (NF_3)) that are not controlled by the Montreal Protocol. Montreal protocal is an international agreement made in 1987. It was designed to stop the production and import of ozone depleting substances and reduce their concentration in the atmosphere to help protect the earth's ozone layer.

1.Data Loading

The dataset here consists the values of different gas for different countries from 1990 to 2018.

In details, these are information on anthropogenic emissions by sources and removals by sinks of the following GHGs: carbon dioxide (CO_2) methane (CH_4) nitrous oxide (N_2O) hydrofluorocarbons (HFCs) perfluorocarbons (PFCs) unspecified mix of HFCs and PFCs sulphur hexafluoride (SF_6) nitrogen triflouride (NF_3)

Let's load the data and loading all the required packages for the analysis

```
#loading required libraries
library(dplyr)#for exploratory data analysis
library(ggplot2)#for data visualization
#loading the csv file
dataset <- read.csv("greenhouse_gas_inventory_data_data.csv",sep=',',stringsA
sFactors = FALSE)
head(dataset)#observing first 6 observations of the dataset
```

```
##   country_or_area year       value
## 1       Australia 2014  393126.947
## 2       Australia 2013 396913.9365
## 3       Australia 2012 406462.8477
## 4       Australia 2011 403705.5283
## 5       Australia 2010 406200.9932
## 6       Australia 2009  408448.479
```

```
##
category
## 1 carbon_dioxide_co2_emissions_without_land_use_land_use_change_and_forest
ry_lulucf_in_kilotonne_co2_equivalent
## 2 carbon_dioxide_co2_emissions_without_land_use_land_use_change_and_forest
ry_lulucf_in_kilotonne_co2_equivalent
## 3 carbon_dioxide_co2_emissions_without_land_use_land_use_change_and_forest
ry_lulucf_in_kilotonne_co2_equivalent
## 4 carbon_dioxide_co2_emissions_without_land_use_land_use_change_and_forest
ry_lulucf_in_kilotonne_co2_equivalent
## 5 carbon_dioxide_co2_emissions_without_land_use_land_use_change_and_forest
ry_lulucf_in_kilotonne_co2_equivalent
## 6 carbon_dioxide_co2_emissions_without_land_use_land_use_change_and_forest
ry_lulucf_in_kilotonne_co2_equivalent
```

2.Data Cleaning

```
#checking the dimensions of the dataset
dim(dataset)

## [1] 9688    4

#checking the internal structure of dataset
str(dataset)

## 'data.frame':    9688 obs. of  4 variables:
##  $ country_or_area: chr  "Australia" "Australia" "Australia" "Australia" .
..
##  $ year           : int  2014 2013 2012 2011 2010 2009 2008 2007 2006 2005
...
##  $ value          : chr  "393126.947" "396913.9365" "406462.8477" "403705.
5283" ...
##  $ category       : chr  "carbon_dioxide_co2_emissions_without_land_use_la
nd_use_change_and_forestry_lulucf_in_kilotonne_co2_equivalent" "carbon_dioxid
e_co2_emissions_without_land_use_land_use_change_and_forestry_lulucf_in_kilot
onne_co2_equivalent" "carbon_dioxide_co2_emissions_without_land_use_land_use_
change_and_forestry_lulucf_in_kilotonne_co2_equivalent" "carbon_dioxide_co2_e
missions_without_land_use_land_use_change_and_forestry_lulucf_in_kilotonne_co
2_equivalent" ...
```

there are 9688 observations(rows) and 4 parameters (columns) in the dataset and in columns it have 3 char type variables

```
#checking is there any missing values in the dataset
any(is.na(dataset))

## [1] FALSE
```

it seems there are no missing values in the dataset if we look at the dataset the ‘category’ variable. it has very large value names which quiet messy to do analysis so let’s rename the values without losing it’s original meaning and making easier to interpret in plots also

```
#renaming the large data points by using mutate and recode
dataset<-dataset %>% mutate(category=recode(category,
                                  carbon_dioxide_co2_emissions_without_land_u
se_land_use_change_and_forestry_lulucf_in_kilotonne_co2_equivalent='CO2',
                                  greenhouse_gas_ghgs_emissions_including_ind
irect_co2_without_lulucf_in_kilotonne_co2_equivalent='GHG_indirect_CO2',
                                  greenhouse_gas_ghgs_emissions_without_land_
use_land_use_change_and_forestry_lulucf_in_kilotonne_co2_equivalent='GHG',
                                  hydrofluorocarbons_hfcs_emissions_in_kiloto
nne_co2_equivalent='HFC',
                                  methane_ch4_emissions_without_land_use_land
_use_change_and_forestry_lulucf_in_kilotonne_co2_equivalent='CH4',
                                  nitrogen_trifluoride_nf3_emissions_in_kilot
onne_co2_equivalent='HF3',
                                  nitrous_oxide_n2o_emissions_without_land_us
e_land_use_change_and_forestry_lulucf_in_kilotonne_co2_equivalent='N2Os',
                                  perfluorocarbons_pfcs_emissions_in_kilotonn
e_co2_equivalent='PFCs',
                                  sulphur_hexafluoride_sf6_emissions_in_kilot
onne_co2_equivalent='SF6',
                                  unspecified_mix_of_hydrofluorocarbons_hfcs_
and_perfluorocarbons_pfcs_emissions_in_kilotonne_co2_equivalent='HFC-PFC-mix'
))

head(dataset)

##   country_or_area year       value category
## 1       Australia 2014  393126.947      CO2
## 2       Australia 2013 396913.9365      CO2
## 3       Australia 2012 406462.8477      CO2
## 4       Australia 2011 403705.5283      CO2
## 5       Australia 2010 406200.9932      CO2
## 6       Australia 2009  408448.479      CO2
```

checking Number of countries that are contributing the gases emission

```
n_distinct(dataset$country_or_area)#gives unie values

## [1] 44
```

there are total 44 unique countries are there in the dataset

```
#renaming the column 'country_or_area_ to 'country' for ease of use

dataset <- dataset %>% rename( country = country_or_area)
```

```
#changing all the values in the country column to the lower case for ease of
use
dataset$country <- tolower(dataset$country)

#changing few longer country names to shorter form
dataset[dataset$country=='russian federation','country'] <-'russia'
dataset[dataset$country=='united kingdom','country'] <-'uk'
dataset[dataset$country=='united states of america','country'] <-'usa'

unique(dataset$country)
```

```
##  [1] "australia"      "austria"        "belarus"        "belgium"
##  [5] "bulgaria"       "canada"         "croatia"        "cyprus"
##  [9] "czech republic" "denmark"        "estonia"        "european union"
## [13] "finland"        "france"         "germany"        "greece"
## [17] "hungary"        "iceland"        "ireland"        "italy"
## [21] "japan"          "latvia"         "liechtenstein"  "lithuania"
## [25] "luxembourg"     "malta"          "monaco"         "netherlands"
## [29] "new zealand"    "norway"         "poland"         "portugal"
## [33] "romania"        "russia"         "slovakia"       "slovenia"
## [37] "spain"          "sweden"         "switzerland"    "turkey"
## [41] "ukraine"        "uk"             "usa"            "czechia"
```

let's remove the value european union which is mix of countries, we are alanysing country wise so it's better to remove and keep it in a single layer level analysis.

```
#removing euopean union using filter funtion
dataset <- dataset %>% filter(country!='european union')
unique(dataset$country)
```

```
##  [1] "australia"      "austria"        "belarus"        "belgium"
##  [5] "bulgaria"       "canada"         "croatia"        "cyprus"
##  [9] "czech republic" "denmark"        "estonia"        "finland"
## [13] "france"         "germany"        "greece"         "hungary"
## [17] "iceland"        "ireland"        "italy"          "japan"
## [21] "latvia"         "liechtenstein"  "lithuania"      "luxembourg"
## [25] "malta"          "monaco"         "netherlands"    "new zealand"
## [29] "norway"         "poland"         "portugal"       "romania"
## [33] "russia"         "slovakia"       "slovenia"       "spain"
## [37] "sweden"         "switzerland"    "turkey"         "ukraine"
## [41] "uk"             "usa"            "czechia"
```

finally let's change the value column to numeric for the analysis

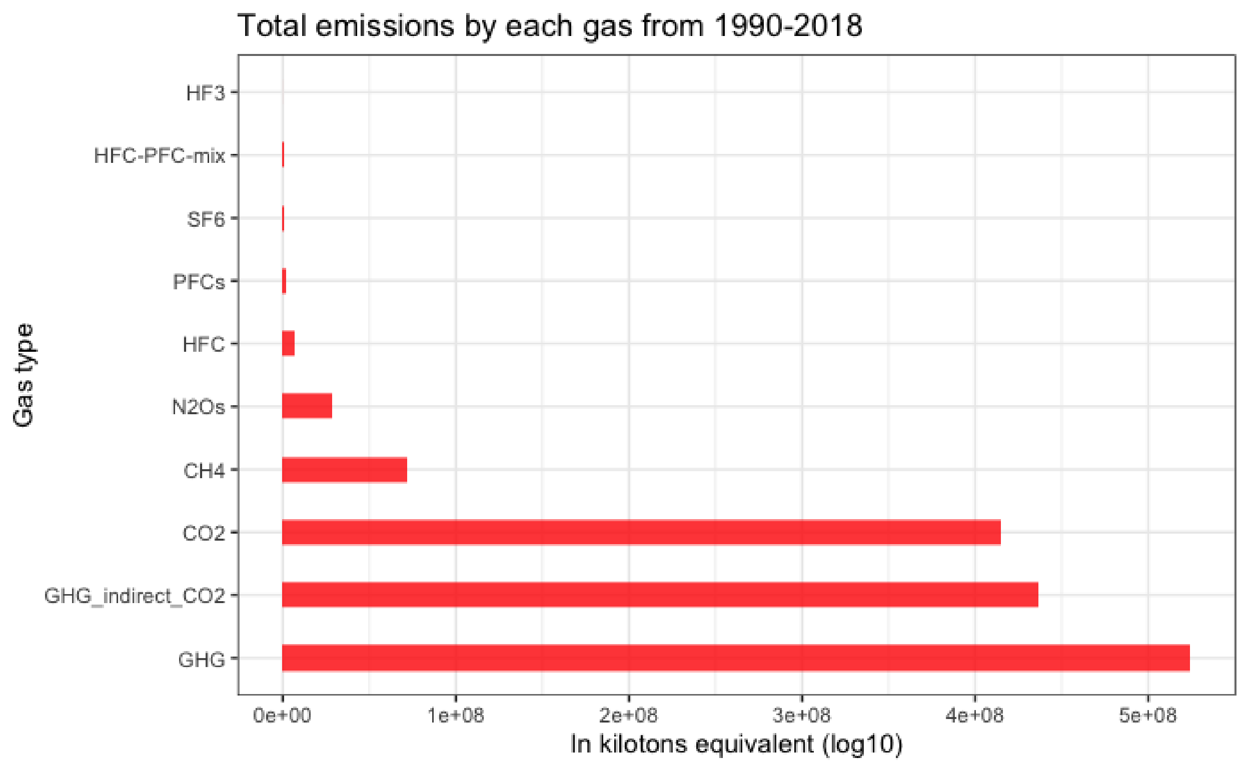
```
#while converting data from char to numeric we might get NA values because so
me of our input values are not formatted properly, because they contain comma
s (i.e. ,) between the numbers. We can remove these commas by using the gsub
function.
dataset$value <- gsub(",","",dataset$value)
dataset$value <- as.numeric(dataset$value)
str(dataset)
```

```
## 'data.frame':    9398 obs. of  4 variables:
##  $ country : chr  "australia" "australia" "australia" "australia" ...
##  $ year    : int  2014 2013 2012 2011 2010 2009 2008 2007 2006 2005 ...
##  $ value   : num  393127 396914 406463 403706 406201 ...
##  $ category: chr  "CO2" "CO2" "CO2" "CO2" ...
```

3.Data analysis

let's find out the total emission per gas that has been released in the total years from 1990 to 2018

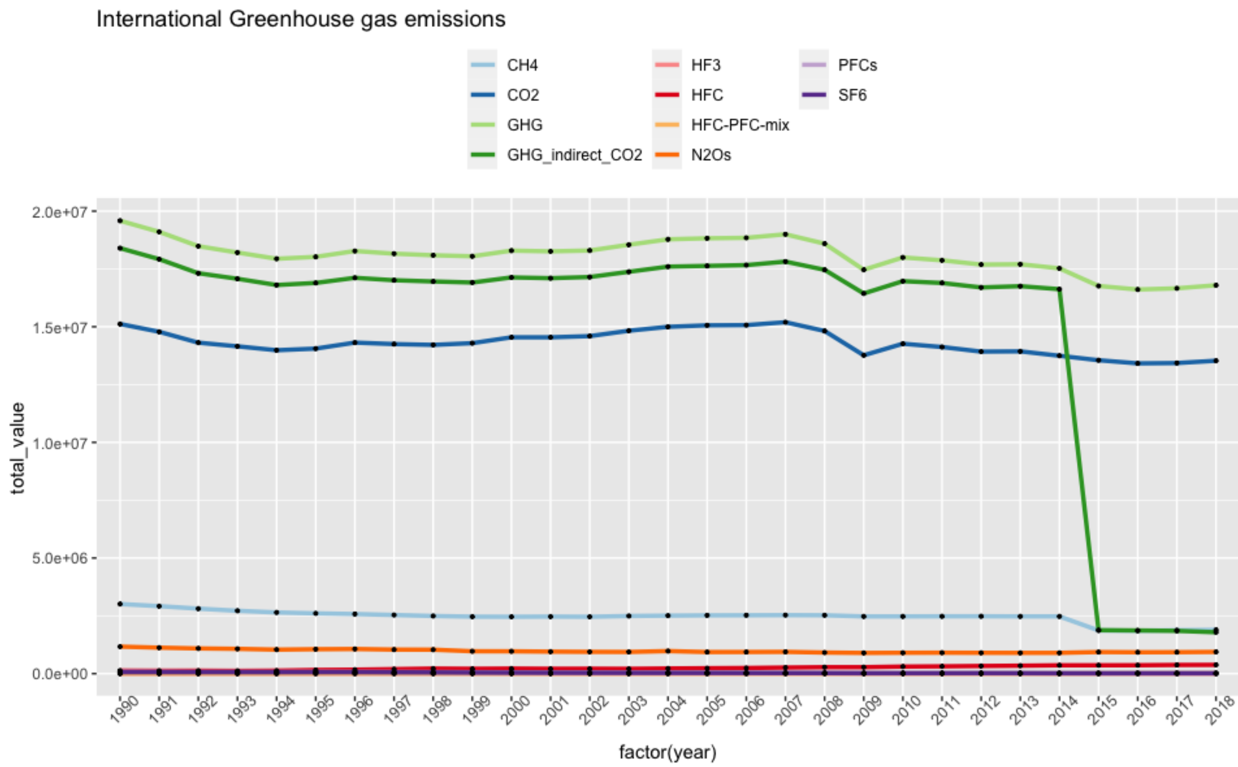
```
total_emissions <- dataset %>% group_by(category) %>%summarise(sum(value))
#arrange the data in descending order
colnames(total_emissions)[2] <- "total_value"#changing name of column
#ordering the data in descending order
total_emissions <- total_emissions %>% arrange(desc(`total_value`))
# getting sum of a target variable
total_gas_sum <- sum(total_emissions$total_value)
#creating a variable 'each_emission_percentage' to store the percentage of co
ntributions
each_gas_percent <- mutate(total_emissions, each_emission_percentage = (total
_emissions$total_value/total_gas_sum)*100)%>% mutate_at(vars(each_emission_pe
rcentage), funs(round(., 3)))#to round the values for 3 decimal places
each_gas_percent

## # A tibble: 10 x 3
##    category         total_value each_emission_percentage
##    <chr>                  <dbl>                    <dbl>
##  1 GHG               524444808.                   35.3
##  2 GHG_indirect_CO2  437114066.                   29.4
##  3 CO2               414828411.                   27.9
##  4 CH4                71608587.                    4.82
##  5 N2Os               28275799.                    1.90
##  6 HFC                 7084020.                    0.477
##  7 PFCs                1301123.                    0.088
##  8 SF6                 1068301.                    0.072
##  9 HFC-PFC-mix          248136.                    0.017
## 10 HF3                   30432.                    0.002

#Visualizing overall emissions of each gas from 1990-2018
ggplot(total_emissions, aes(x=reorder(category,-total_value), y=total_value))
+
    geom_bar(stat="identity", fill="red", alpha=.8, width=.4) +
    xlab("Gas type") +
    ylab("In kilotons equivalent (log10)")+
    ggtitle("Total emissions by each gas from 1990-2018")+
    theme_bw()+
  coord_flip(expand = TRUE)#rotate the axis to avoid data overlapping
```



As in the above plot it's clearly showing that "GHG" is the highest emission around the world, followed by "GHG_indirect_CO2" as second highest and next order is in "CO2", "CH4" and "N2Os". These are the top five emissions from the year 1990 -2018 around the world.

#now let's look at the total contribution of all gases over the years.

```
#group the data by category and year
Total_contribution_per_year <- dataset %>% group_by(category,year) %>%
  summarize(total_value = sum(value),.groups = 'drop')
#now visualizing the grouped data
ggplot(Total_contribution_per_year, aes(x=factor(year),y=total_value,group=ca
tegory)) + geom_line(aes(color=category),size=1.2) + #for line plot
  geom_point(size=.7,color='black') +#points in the graph and size of points
  scale_color_brewer(name='',palette='Paired') + #coloring the lining
  theme(legend.position='top') + #placing the legend position
  theme(axis.text.x = element_text(size=9,angle=45),#size of x axis text
        legend.text=element_text(size=9)) +
  labs(title='International Greenhouse gas emissions')+
  guides(color=guide_legend(ncol=3))
```



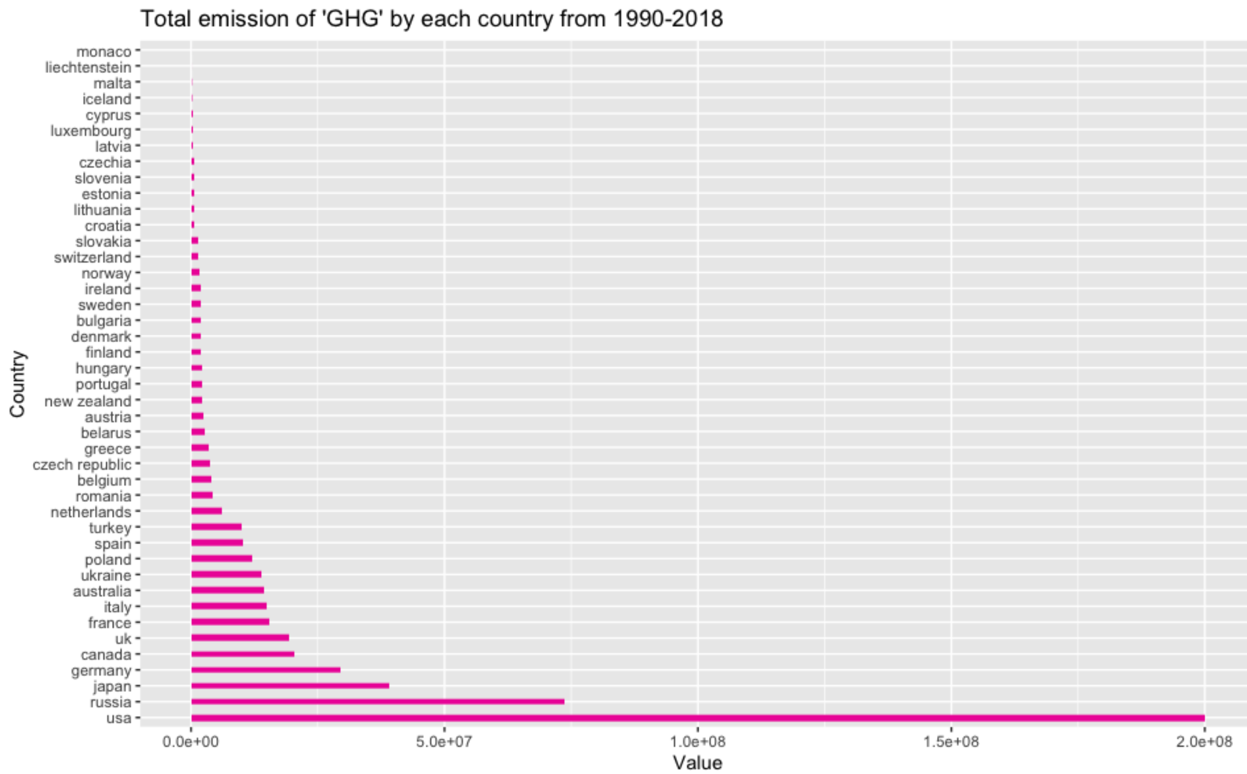
by looking at the above graph, ‘GHG’, ‘GHG_indirect_CO2’ and ‘CO2’ are the most emissioned gas over the past years and there is a substantial drop of ‘’GHG_indirect_gas’ from the year 2015 to 2018 and few gases have a little amount of ups and downs from where they started but they continued on the same level so far.

#let’s visualize which countries caused highest emissions per each gas and for deeper analysis,let’s visualise the top 4 gases that are produced by entire world over the years. **the top 4 are** 1.GHG 2.GHG_indirect_CO2 3.CO2 4.CH4

1. Analysis of “GHG” gas emission

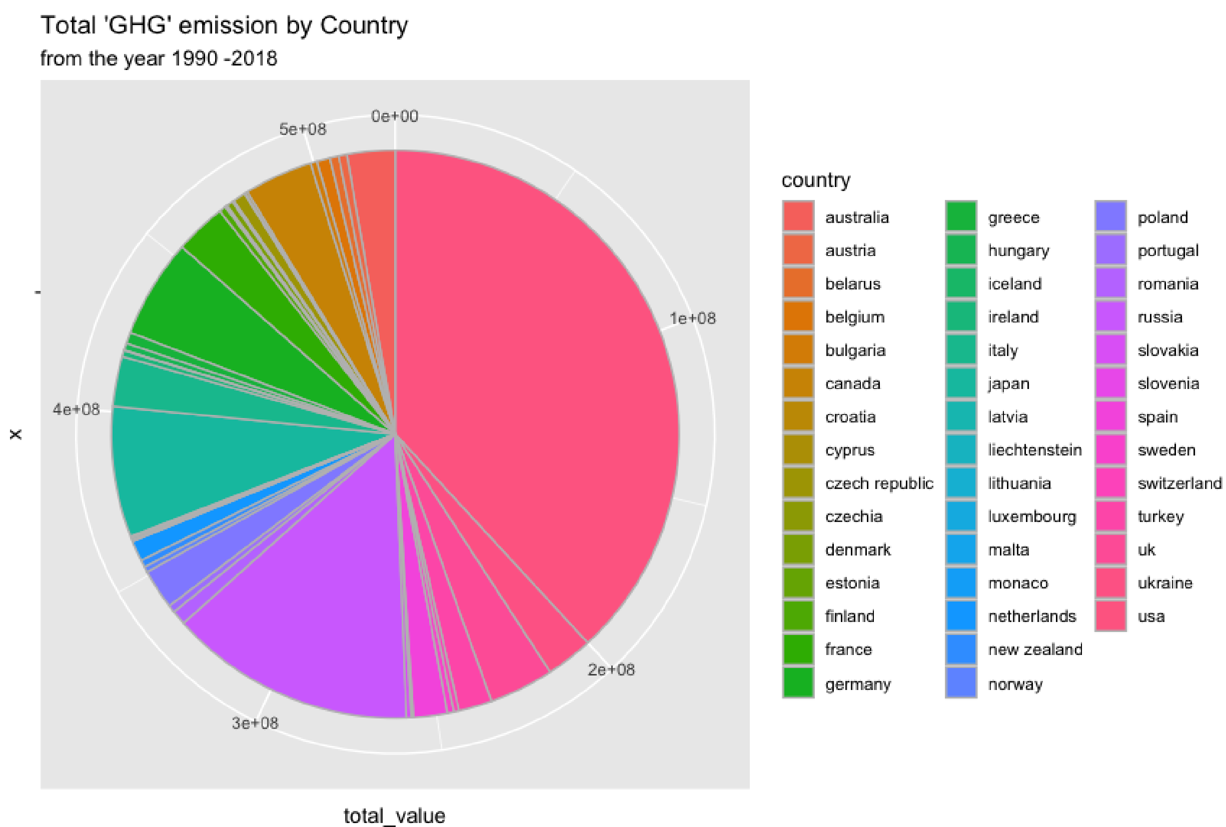
```
#selecting a subset of data that only contains "GHG" gas emission
GHG_data <- dataset[dataset$category=='GHG',]
#grouping by each country to know which countries produced highest GHG
GHG_by_country <- GHG_data%>% group_by(country) %>%summarise(sum(value))
colnames(GHG_by_country)[2] <- "total_value"#changing name of column
GHG_by_country <- GHG_by_country %>% arrange(desc(total_value))

ggplot(GHG_by_country, aes(x= reorder(country, -total_value), y=total_value))
+
    geom_bar(stat="identity", fill="maroon2", alpha=1, width=.4) +
    xlab("Country") +
```

```
    ylab("Value")+
    ggtitle("Total emission of 'GHG' by each country from 1990-2018")+
  coord_flip(expand = TRUE)
```



countries– USA, Russia, Japan, Germany, Canada, Uk and France are at top and all of these combines contributing a major chunk of 'GHG' emission.

```
#visualizing % of contribution of each country
pie <- ggplot(GHG_by_country,aes(x="",y=total_value,fill = country))+
  geom_bar(stat = "identity",colour="gray")
pie+coord_polar("y", start=0,direction = 1)+
  ggtitle("Total 'GHG' emission by Country",subtitle = "from the year 1990 -2
018")
```

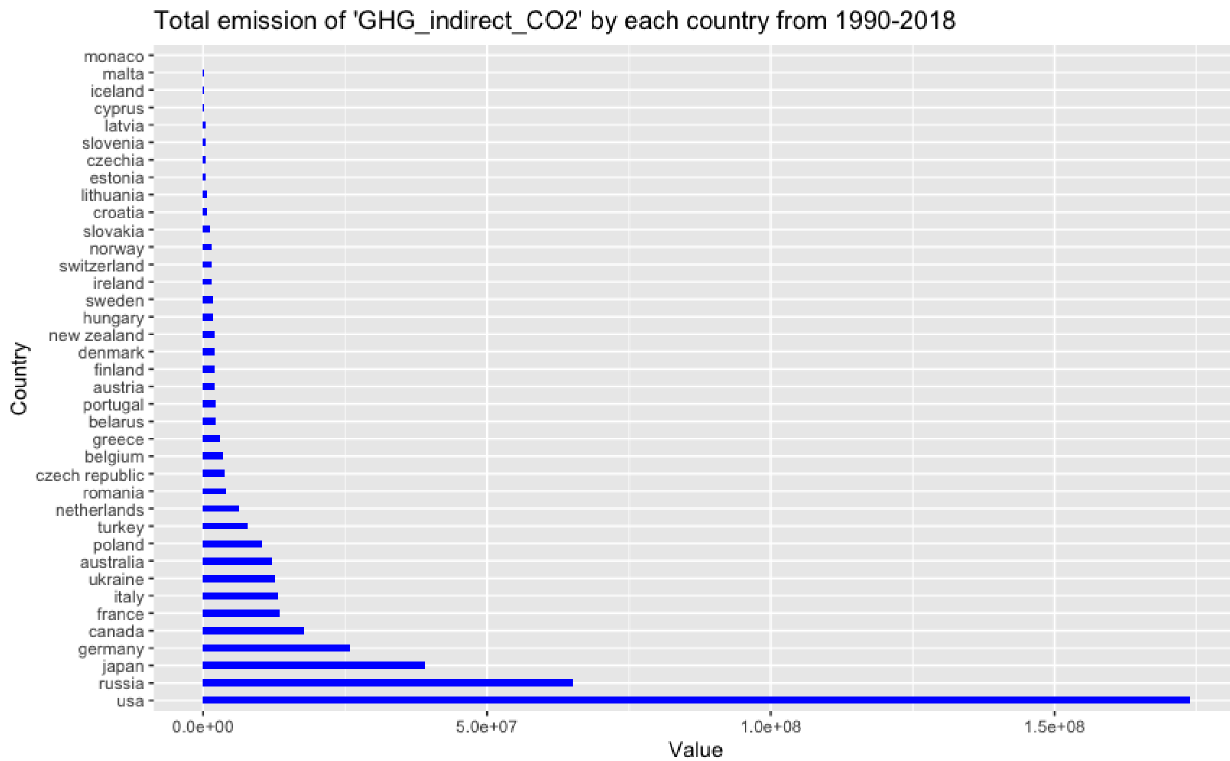


Top contributers from above graph

2. Analysis of "GHG_indirect_CO2" gas emission

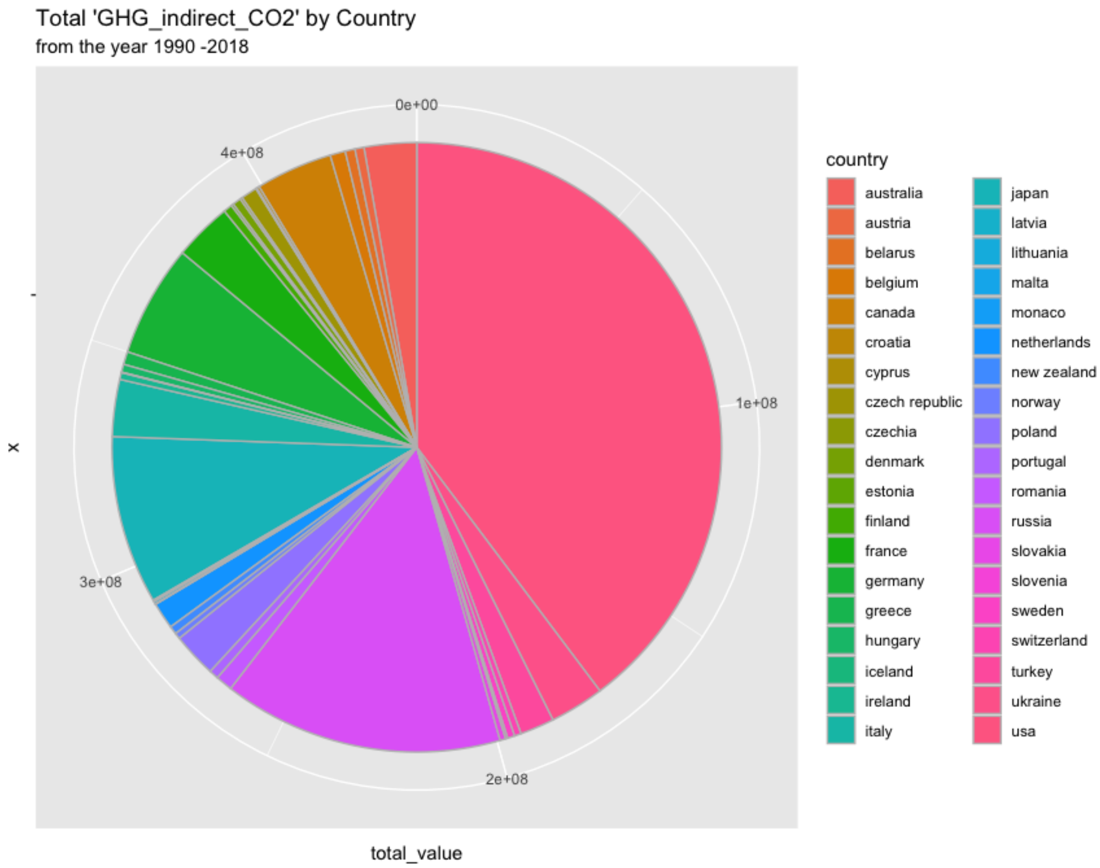
```
#selecting a subset of data that only contains "GHG" gas emission
GHG_indirect_CO2 <- dataset[dataset$category=='GHG_indirect_CO2',]
#grouping by each country to know which countries produced highest GHG
GHG_indirect_country <- GHG_indirect_CO2%>% group_by(country) %>%summarise(to
tal_value = sum(value),.groups = 'drop')
GHG_indirect_country <- GHG_indirect_country %>% arrange(desc(total_value))

#visualising the GHG_indirect_CO2
ggplot(GHG_indirect_country, aes(x= reorder(country, -total_value), y=total_v
alue)) +
    geom_bar(stat="identity", fill="blue", alpha=1, width=.4) +
    xlab("Country") +
```

```
    ylab("Value")+
    ggtitle("Total emission of 'GHG_indirect_CO2' by each country from 1990-2
018")+
  coord_flip(expand = TRUE)
```



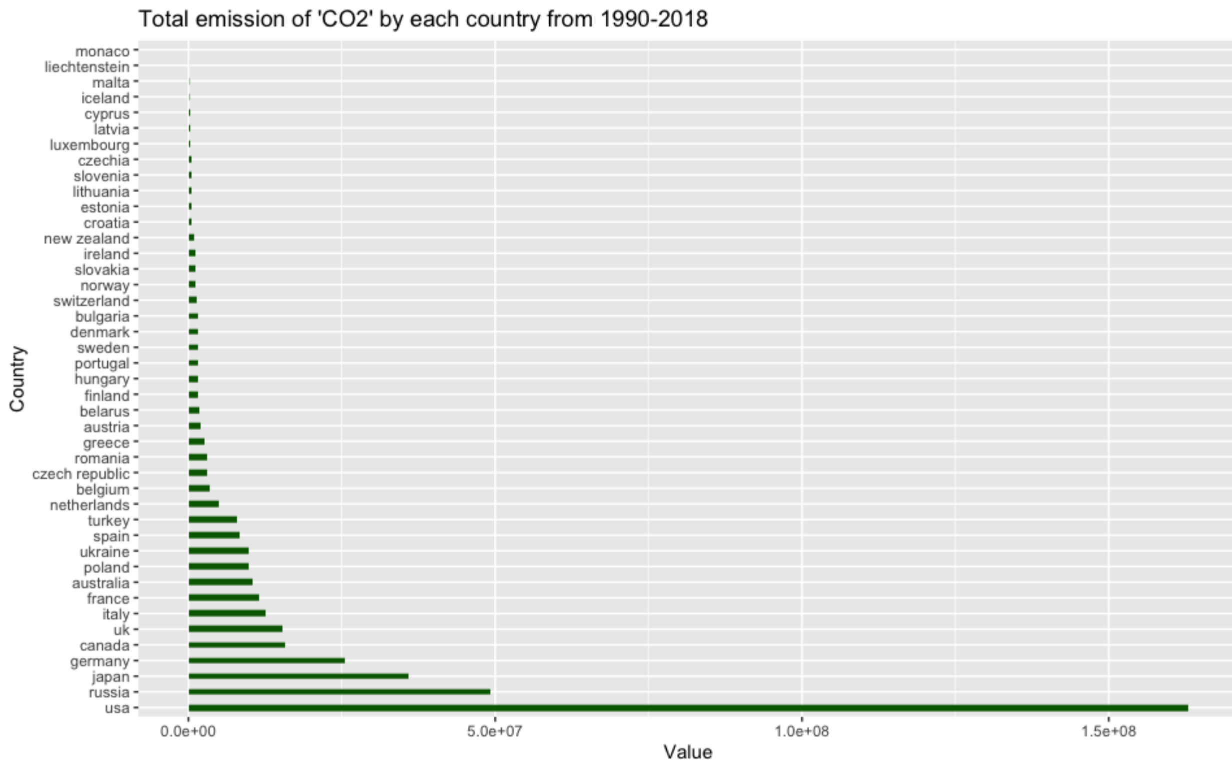
countries– usa,russia,japan,germany, canada,france and italy are at top and all of these combine contributing a major chunk of ‘GHG_indirect_CO2’ emission.

```
#visualizing % of contribution of each country
pie <- ggplot(GHG_indirect_country,aes(x="",y=total_value,fill = country))+
  geom_bar(stat = "identity",colour="gray")
pie+coord_polar("y", start=0,direction = 1)+
  ggtitle("Total 'GHG_indirect_CO2' by Country",subtitle = "from the year 199
0 -2018")
```



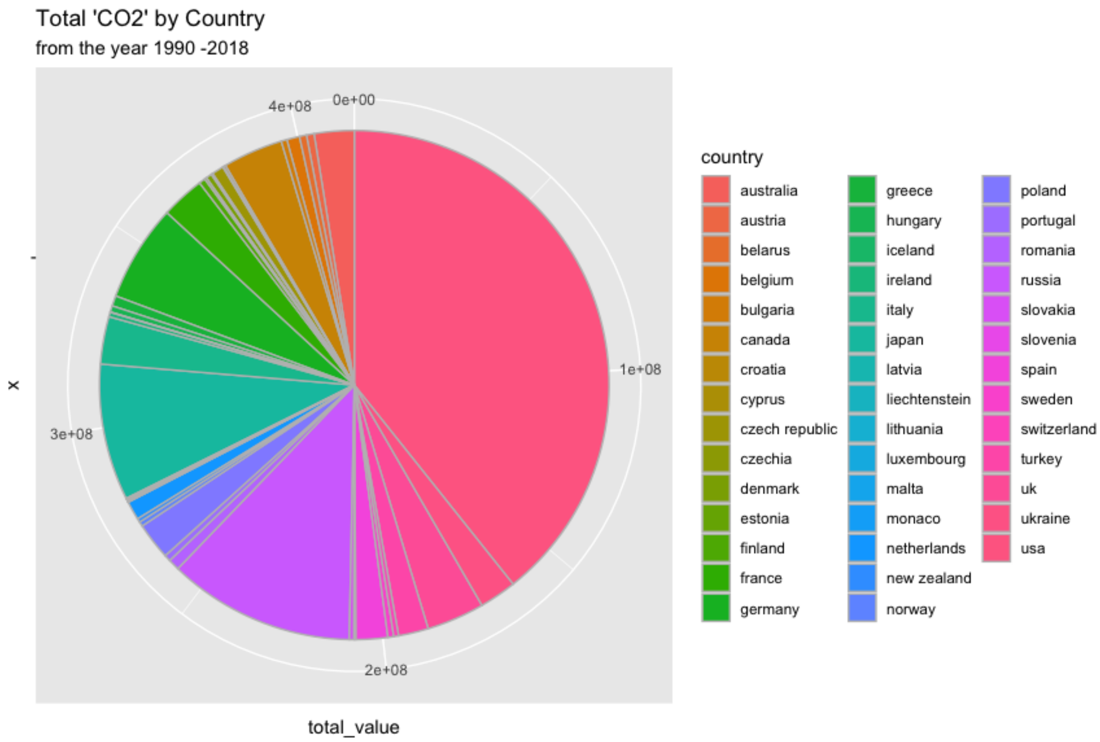
3. Analysis of "CO2" gas emission

```r
#selecting a subset of data that only contains "CO2" gas emission
CO2_data <- dataset[dataset$category=='CO2',]
#grouping by each country to know which countries produced highest GHG
CO2_by_country <- CO2_data%>% group_by(country) %>%summarise(total_value = su
m(value),.groups = 'drop')
CO2_by_country <- CO2_by_country %>% arrange(desc(total_value))

#visualising the CO2 emmission
ggplot(CO2_by_country, aes(x= reorder(country, -total_value), y=total_value))
+
    geom_bar(stat="identity", fill="darkgreen", alpha=1, width=.4) +
    xlab("Country") +
    ylab("Value")+
    ggtitle("Total emission of 'CO2' by each country from 1990-2018")+
  coord_flip(expand = TRUE)
```



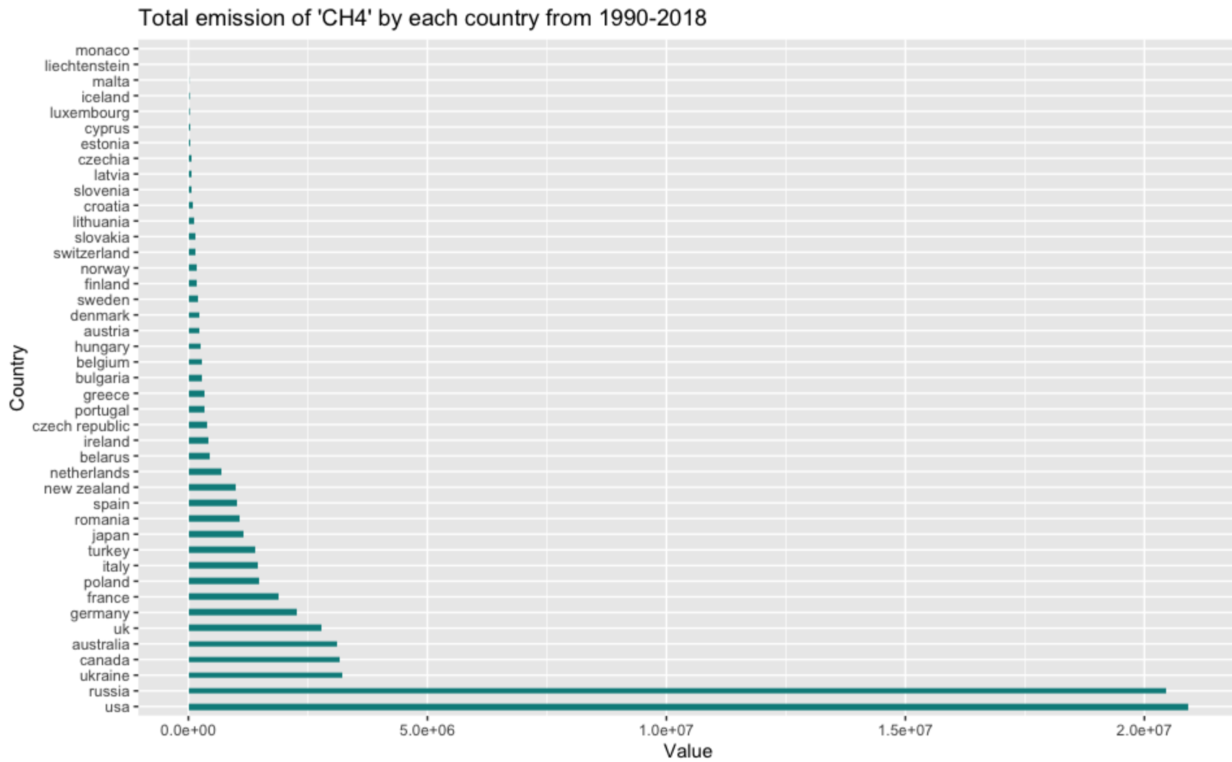
the Top 'CO2' emission countries are = usa, russia,japan,germany,canada,uk and Italy

```r
#visualizing % of contribution of each country
pie <- ggplot(CO2_by_country,aes(x="",y=total_value,fill = country))+
  geom_bar(stat = "identity",colour="gray")
pie+coord_polar("y", start=0,direction = 1)+
  ggtitle("Total 'CO2' by Country",subtitle = "from the year 1990 -2018")
```



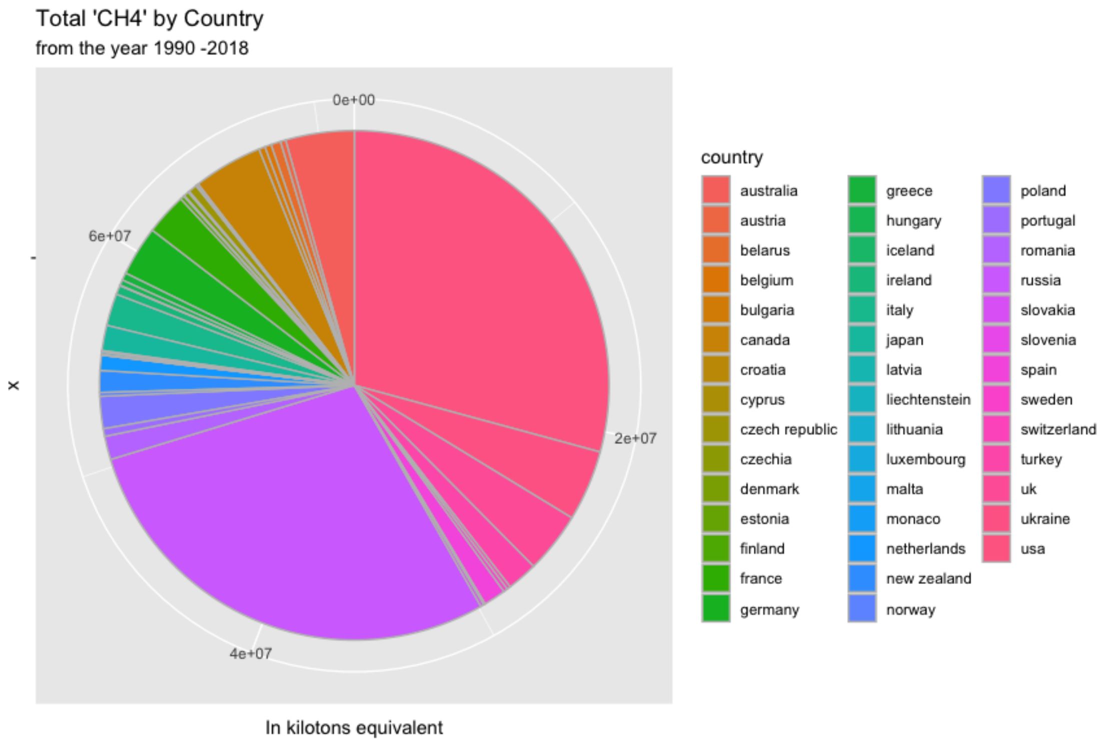
4. Analysis of “CH4” gas emission

```
#selecting a subset of data that only contains "CO2" gas emission
CH4_data <- dataset[dataset$category=='CH4',]
#grouping by each country to know which countries produced highest GHG
CH4_by_country <- CH4_data%>% group_by(country) %>%summarise(total_value = su
m(value),.groups = 'drop')
CH4_by_country <- CH4_by_country %>% arrange(desc(total_value))

#visualising the GHG_indirect_CO2
ggplot(CH4_by_country, aes(x= reorder(country, -total_value), y=total_value))
+
    geom_bar(stat="identity", fill="cyan4", alpha=1, width=.4) +
    xlab("Country") +
    ylab("Value")+
    ggtitle("Total emission of 'CH4' by each country from 1990-2018")+
  coord_flip(expand = TRUE)
```



the Top ‘CH4’ emission countries are = usa, russia,ukraine,canada,australia,uk and germany

```
#visualizing % of contribution of each country
pie <- ggplot(CH4_by_country,aes(x="",y=total_value,fill = country))+
  geom_bar(stat = "identity",colour="gray")
pie+coord_polar("y", start=0,direction = 1)+
  ggtitle("Total 'CH4' by Country",subtitle = "from the year 1990 -2018")+
  ylab("In kilotons equivalent")
```

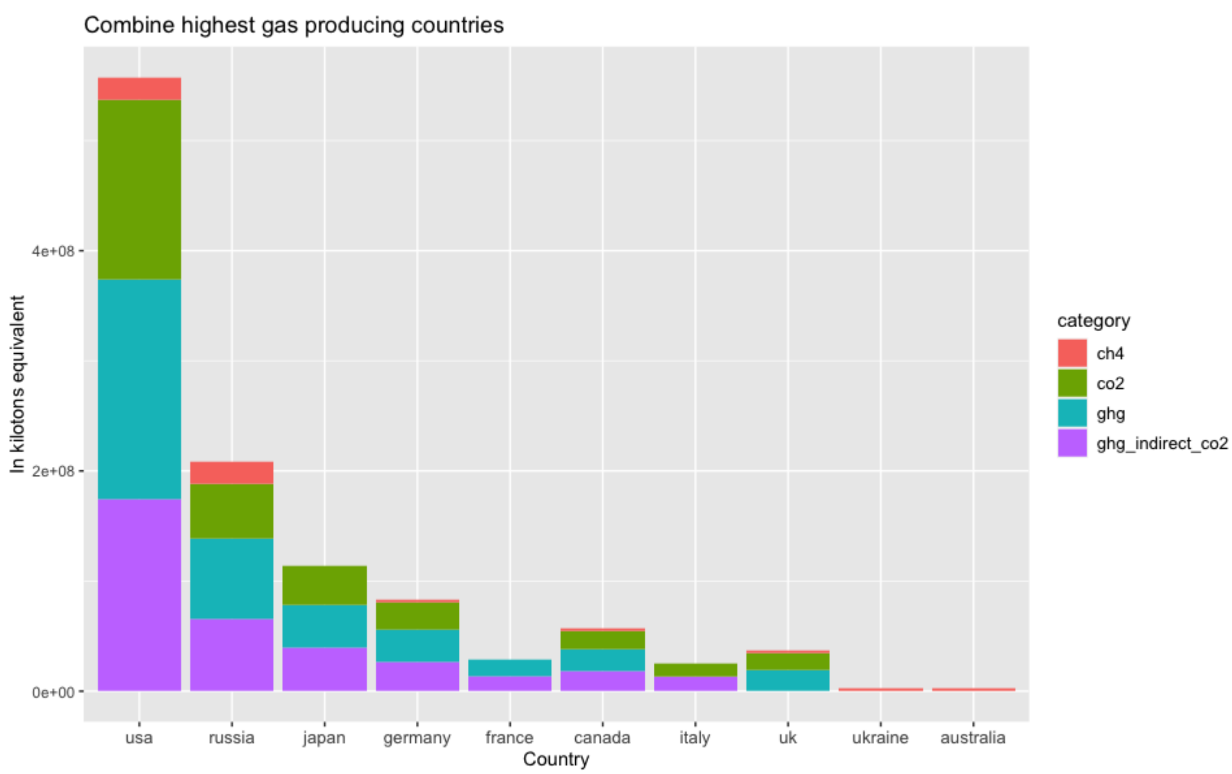


from the above graph the % of contribution is higher in USA, Russia, ukrine and canada

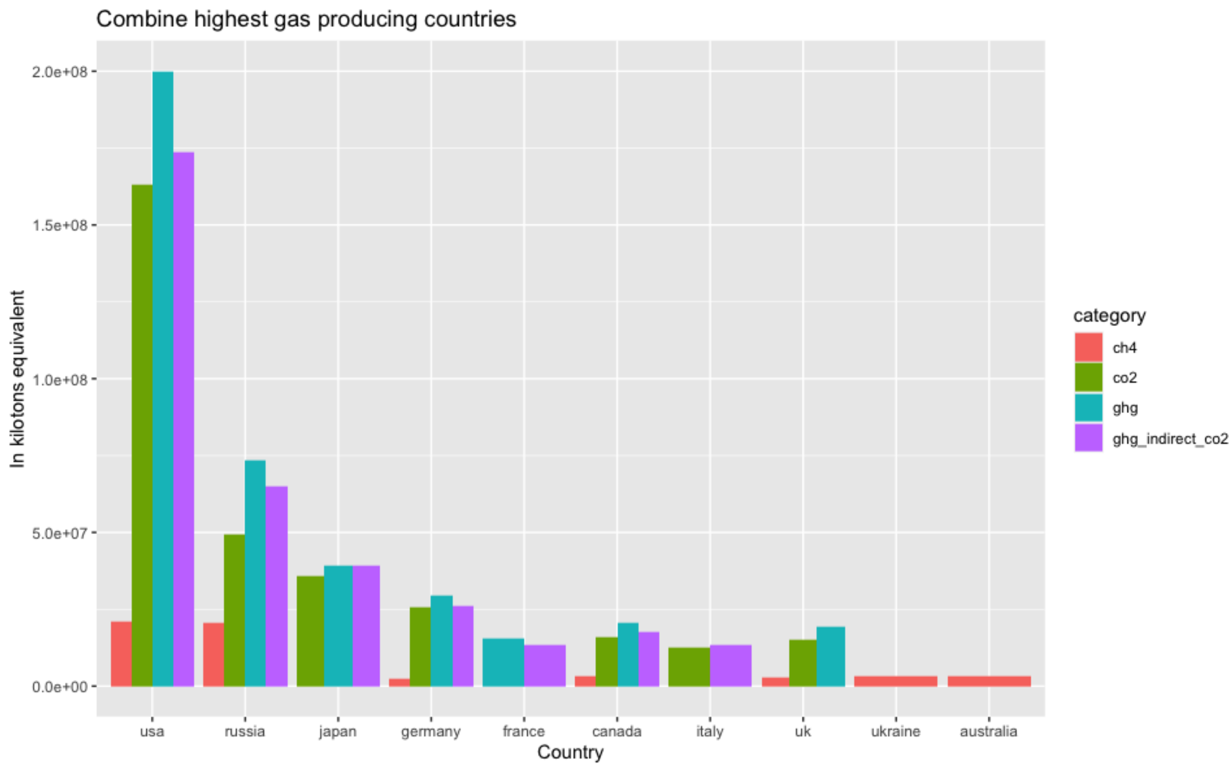
let's look at the top countries for the top 4 emissioned gases.

```
#subset the top seven 7 countries for each of high emissioned gas
category <- c('ghg','ghg','ghg','ghg','ghg','ghg','ghg')
GHG_top_countries <- data.frame(GHG_by_country[c(1:7),],category)

#adding a new column for each type of gas
category <- c('ghg_indirect_co2','ghg_indirect_co2','ghg_indirect_co2','ghg_i
ndirect_co2','ghg_indirect_co2','ghg_indirect_co2','ghg_indirect_co2')
GHG_indirect_top <- data.frame(GHG_indirect_country[c(1:7),],category)

#adding a new column for each type of gas
category <- c('co2','co2','co2','co2','co2','co2','co2')
CO2_top_countries <- data.frame(CO2_by_country[c(1:7),],category)

#adding a new column for each type of gas
category <-c('ch4','ch4','ch4','ch4','ch4','ch4','ch4')
CH4_top_countries <-data.frame(CH4_by_country[c(1:7),],category)
```

```
#using rbind we can combine all the data and make it visualize clearly
total_top_countries <- rbind(GHG_top_countries,GHG_indirect_top,CO2_top_count
ries,CH4_top_countries)

#plotting all the top gases and country's each contribution
col_chart<- ggplot(data = total_top_countries)+
  geom_col(mapping = aes(x = reorder(country, -total_value), y = total_value,
fill = category))+
  theme(axis.text.x = element_text(size=10),
        legend.text=element_text(size=10))+
  xlab("Country")+
  ylab("In kilotons equivalent")+
  ggtitle("Combine highest gas producing countries")
col_chart
```



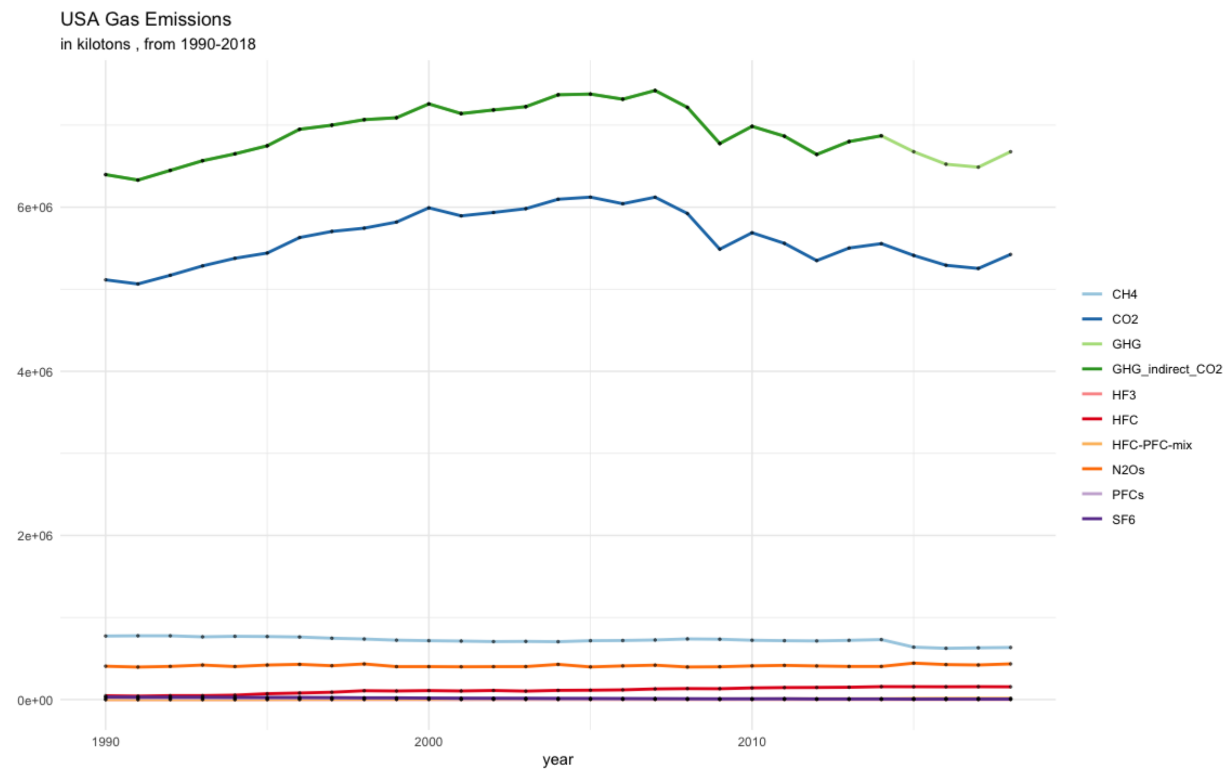

```
#changing the bars to individual
ggplot(data = total_top_countries)+
  geom_col(mapping = aes(x = reorder(country, -total_value), y = total_value,
fill = category),position = "dodge")+
  xlab("Country")+
  ylab("In kilotons equivalent")+
  ggtitle("Combine highest gas producing countries")
```



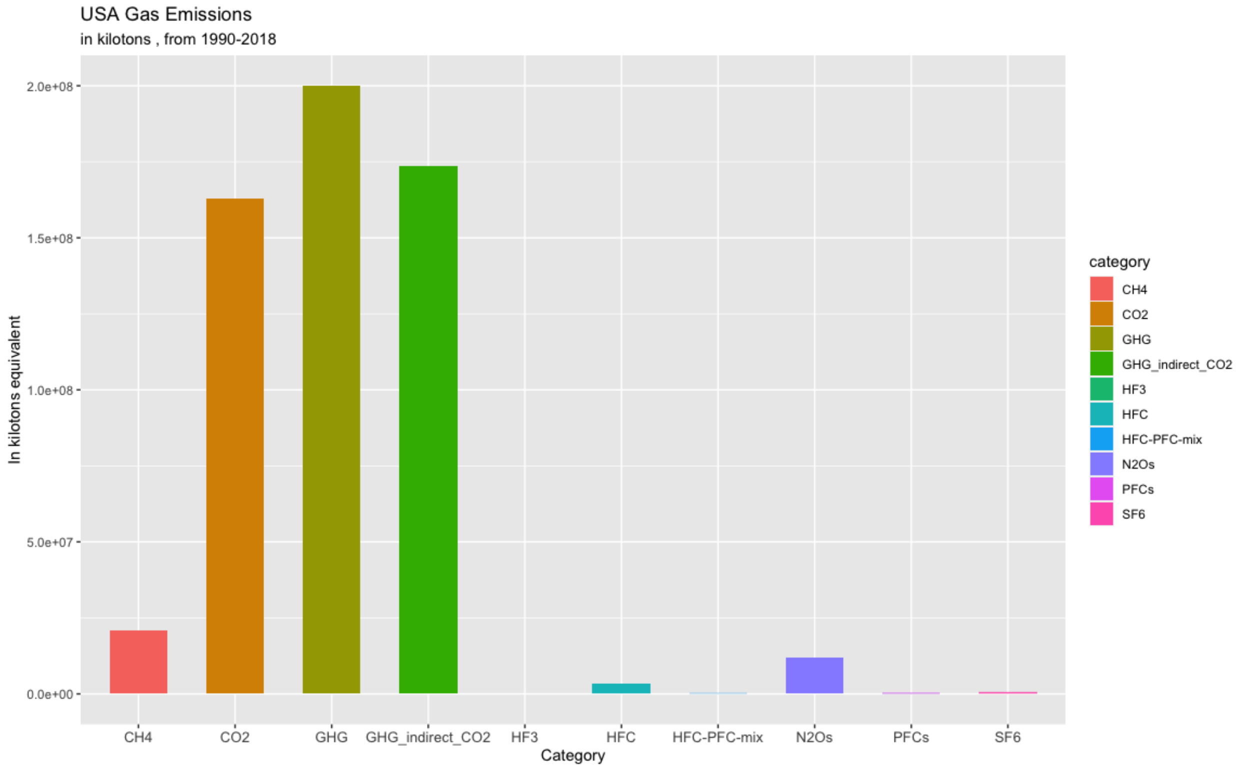
from above plot we can clearly see that countries “USA”,“Russia”,“Japan” and “Germany” contributing a high portion of these gases and keeping these observations in mind now let’s look at these four countries individual gas emissions over the years.

USA emissions from 1990 to 2018

```
#analysis of USA country emissions
usa_emissions <- dataset[dataset$country=='usa',] #selecting data for a singl
e country
usa_emissions %>% group_by(category, year) %>% #grouping category and year to
gether
  summarise(total_value=sum(value),.groups = 'drop')%>%
  ggplot(aes(x=year,y=total_value, col=category ))+#plotting a line plot
  geom_line(na.rm=TRUE, size=1)+
  geom_point(size=.5,alpha = .5,color = "black")+
  scale_color_brewer(name='',palette='Paired')+
  theme_minimal()+
  labs(title='USA Gas Emissions',
       subtitle= 'in kilotons , from 1990-2018')+
  xlab('year')+ylab('')
```



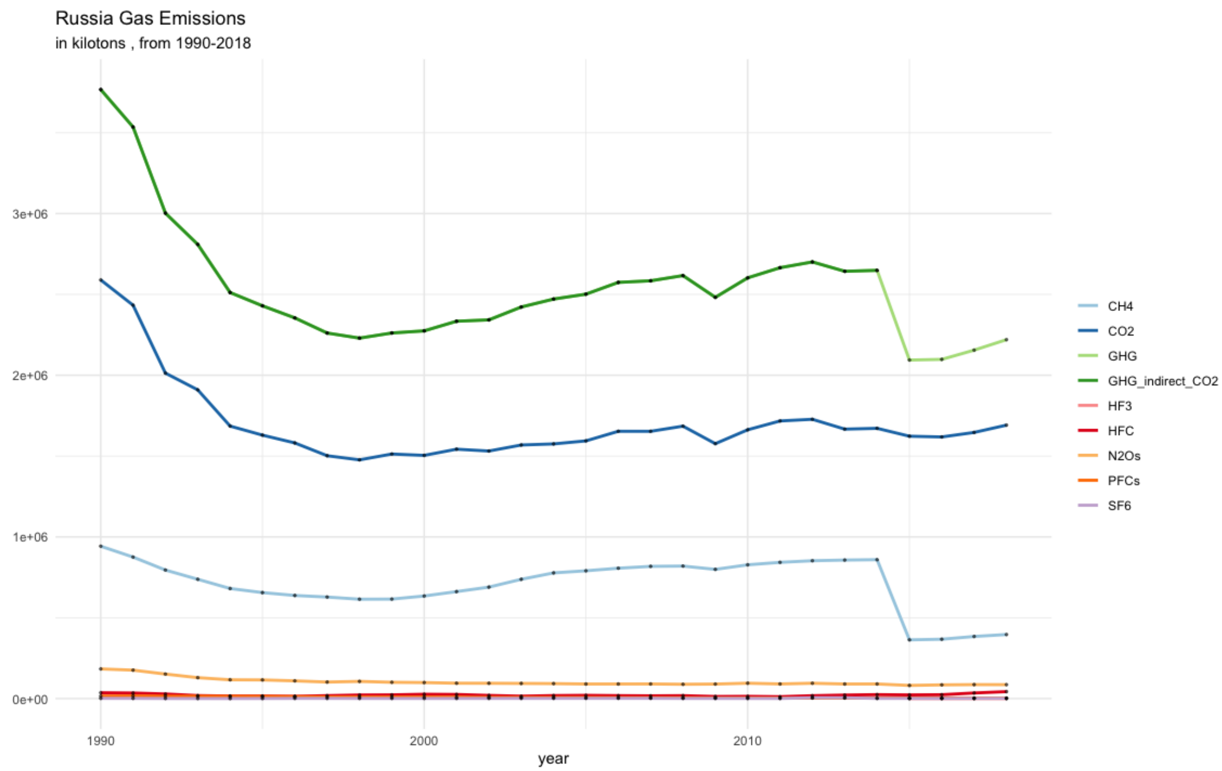

```
#visualizing bar plot to know the portion of each emission gases and compare with each other
 usa_emissions %>% group_by(category, year) %>%
  summarise(total_value=sum(value),.groups = 'drop')%>%
  ggplot(aes(x=category, y=total_value,fill = category))+
  geom_col(width = .6)+
  theme(axis.text.x = element_text(size=10))+#increasing the length of x axis text
  xlab("Category")+
  ylab("In kilotons equivalent")+
  labs(title='USA Gas Emissions',
       subtitle= 'in kilotons , from 1990-2018')
```




```
#lets find out the % of emission of each USA gases from 1990 to 2018
usa_gas_wise_total <- usa_emissions%>% group_by(category) %>%summarise(total_
value = sum(value),.groups = 'drop')
#find the sum of total country gas emission
usa_sum <- sum(usa_gas_wise_total$total_value)
#create a column name 'emission_percentage and calculate the % of contributio
n'
usa_percentage <- mutate(usa_gas_wise_total, usa_emission_percentage = (usa_g
as_wise_total$total_value/usa_sum)*100)%>% mutate_at(vars(usa_emission_percen
tage), funs(round(., 3)))
usa_percentage

## # A tibble: 10 x 3
##    category         total_value usa_emission_percentage
##  * <chr>                  <dbl>                   <dbl>
##  1 CH4                20922010.                   3.65
##  2 CO2               162992307.                  28.4
##  3 GHG               200068613.                  34.9
##  4 GHG_indirect_CO2  173703277.                  30.3
##  5 HF3                   11054.                   0.002
##  6 HFC                 3220566.                   0.561
##  7 HFC-PFC-mix          181693.                   0.032
##  8 N2Os               11949725.                   2.08
##  9 PFCs                 320123.                   0.056
## 10 SF6                  471137.                   0.082
```

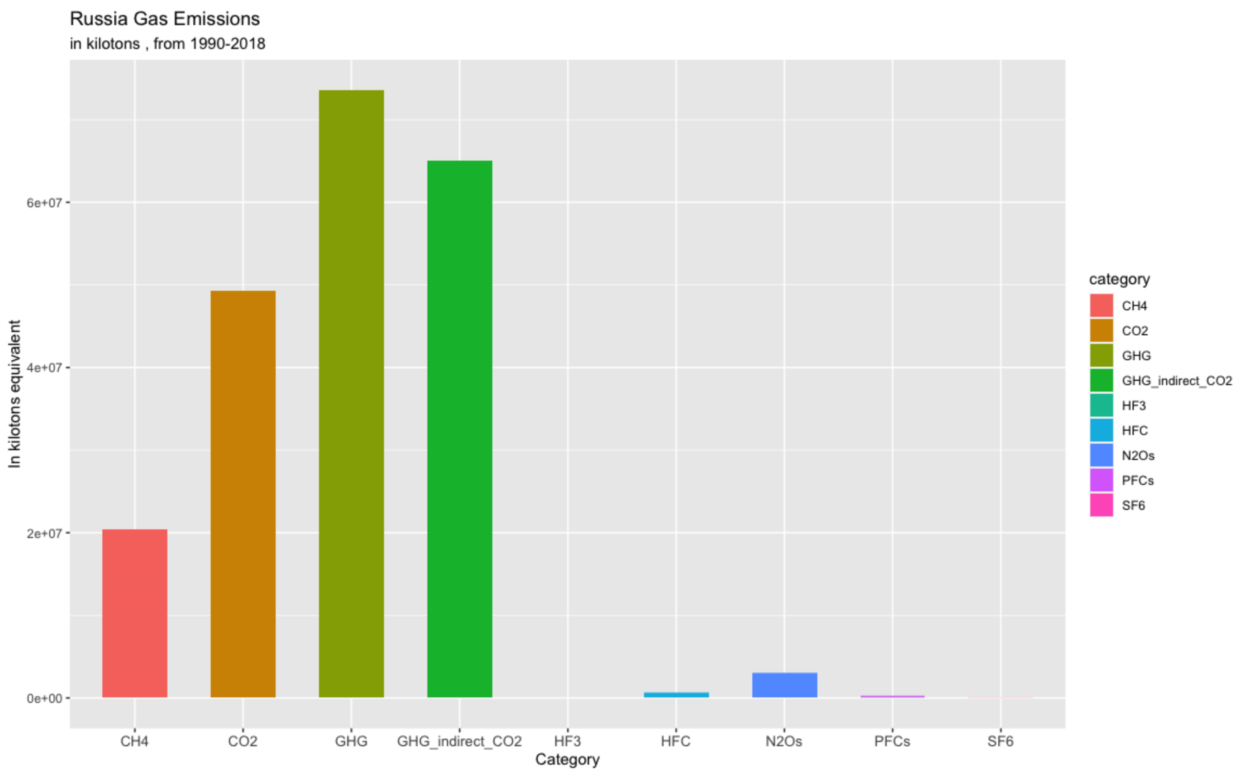
Russia emissions from 1990 to 2018

```
#analysis of Russia country emissions
russia_emissions <- dataset[dataset$country=='russia',] #selecting data for a single country
russia_emissions %>% group_by(category, year) %>% #grouping category and year together
  summarise(total_value=sum(value),.groups = 'drop')%>%
  ggplot(aes(x=year,y=total_value, col=category ))+#plotting a line plot
  geom_line(na.rm=TRUE, size=1)+
  geom_point(size=.5,alpha = .5,color = "black")+
  scale_color_brewer(name='',palette='Paired')+
  theme_minimal()+
  labs(title='Russia Gas Emissions',
       subtitle= 'in kilotons , from 1990-2018')+
  xlab('year')+ylab('')
```




```
#visualizing bar plot to know the portion of each emission gases and compare with each other
 russia_emissions %>% group_by(category, year) %>%
  summarise(total_value=sum(value),.groups = 'drop')%>%
  ggplot(aes(x=category, y=total_value,fill = category))+
  geom_col(width = .6)+
  theme(axis.text.x = element_text(size=10))+#increasing the length of x axis text
  xlab("Category")+
  ylab("In kilotons equivalent")+
```

```r
  labs(title='Russia Gas Emissions',
       subtitle= 'in kilotons , from 1990-2018')
```

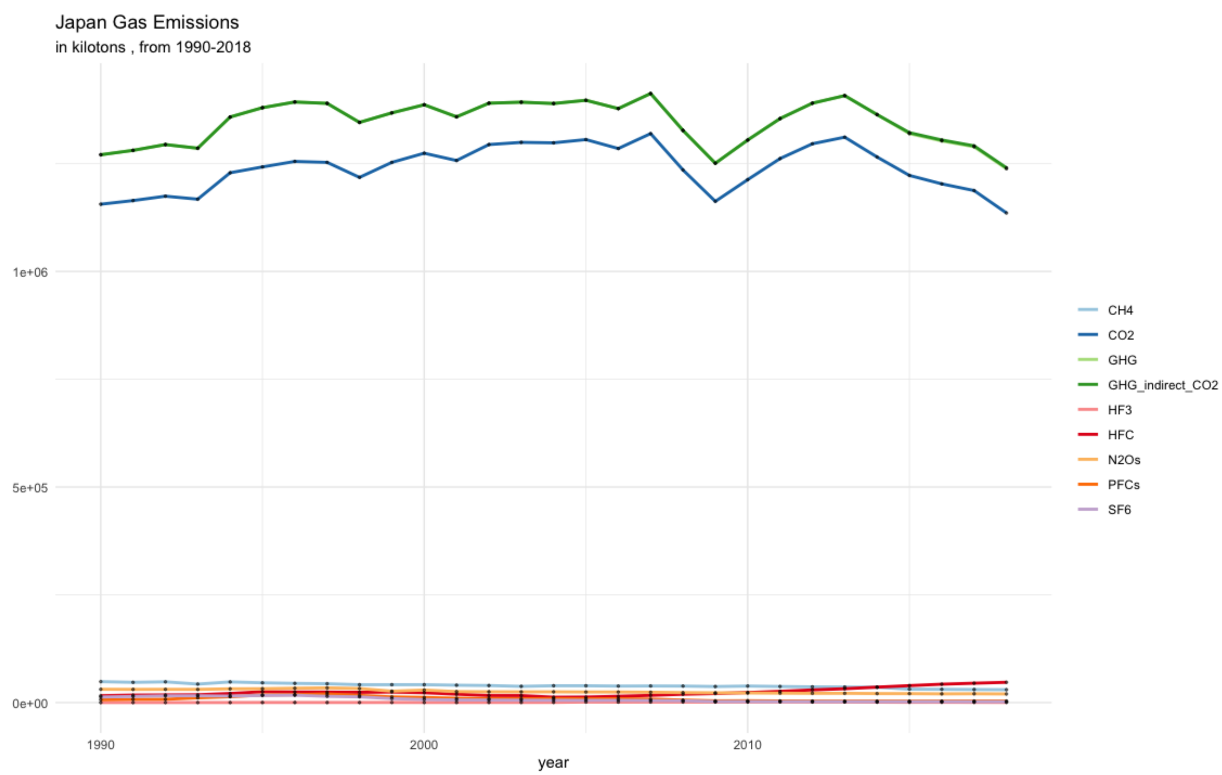



```r
#lets find out the % of emission of each Russia gases from 1990 to 2018
russia_gas_wise_total <- russia_emissions%>% group_by(category) %>%summarise(
total_value = sum(value),.groups = 'drop')
#find the sum of total country gas emission
russia_sum <- sum(russia_gas_wise_total$total_value)
#create a column name 'emission_percentage and calculate the % of contributio
n'
russia_percentage <- mutate(russia_gas_wise_total, russia_emission_percentage
= (russia_gas_wise_total$total_value/russia_sum)*100)%>% mutate_at(vars(russi
a_emission_percentage), funs(round(., 3)))
russia_percentage
```

```
## # A tibble: 9 x 3
##   category         total_value russia_emission_percentage
## * <chr>                  <dbl>                      <dbl>
## 1 CH4                 20462518.                      9.64
## 2 CO2                 49241132.                     23.2
## 3 GHG                 73594670.                     34.7
## 4 GHG_indirect_CO2    65027127.                     30.6
## 5 HF3                        9                       0
## 6 HFC                   624129.                      0.294
## 7 N2Os                 3009207.                      1.42
## 8 PFCs                  224047.                      0.106
## 9 SF6                    33627.                      0.016
```

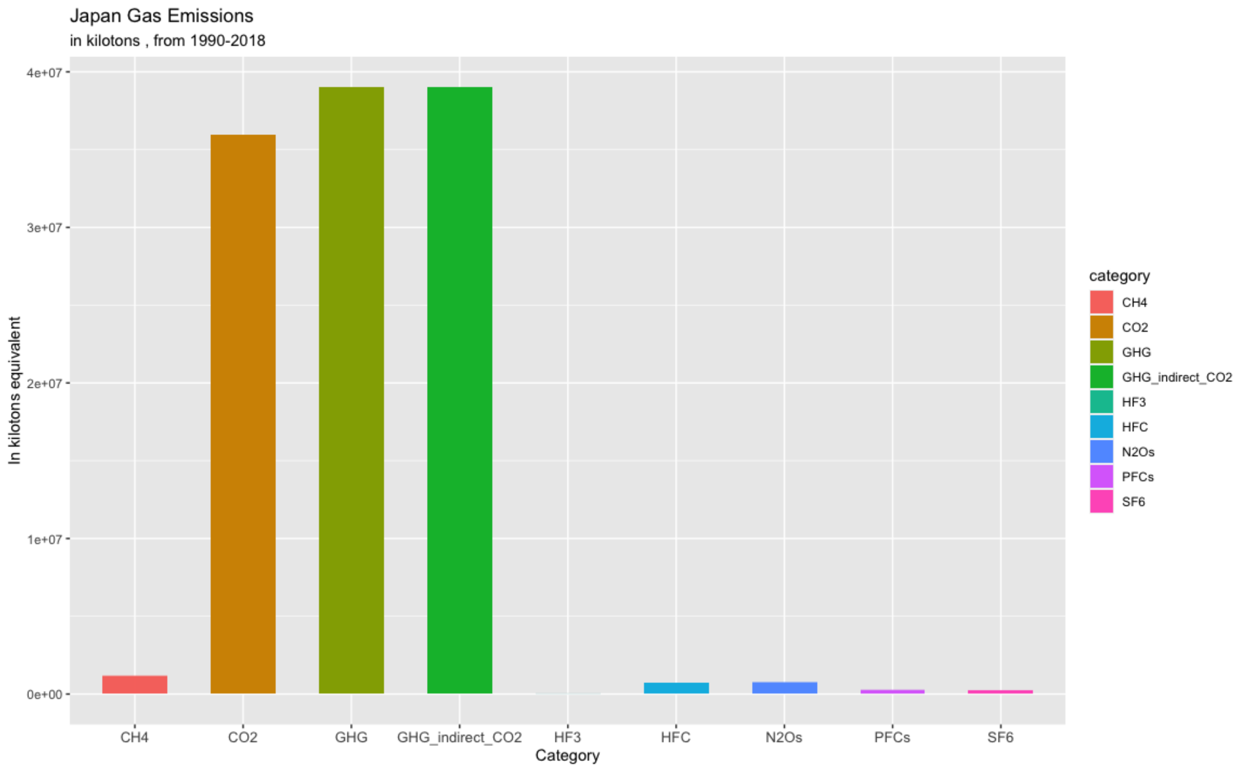
__Japan emissions from 1990 to 2018__

```
#analysis of Japan country emissions
japan_emissions <- dataset[dataset$country=='japan',] #selecting data for a s
ingle country
japan_emissions %>% group_by(category, year) %>% #grouping category and year
together
  summarise(total_value=sum(value),.groups = 'drop')%>%
  ggplot(aes(x=year,y=total_value, col=category ))+#plotting a line plot
  geom_line(na.rm=TRUE, size=1)+
  geom_point(size=.5,alpha = .5,color = "black")+
  scale_color_brewer(name='',palette='Paired')+
  theme_minimal()+
  labs(title='Japan Gas Emissions',
       subtitle= 'in kilotons , from 1990-2018')+
  xlab('year')+ylab('')
```




```
#visualizing bar plot to know the portion of each emission gases and compare
with each other
 japan_emissions %>% group_by(category, year) %>%
  summarise(total_value=sum(value),.groups = 'drop')%>%
  ggplot(aes(x=category, y=total_value,fill = category))+
  geom_col(width = .6)+
  theme(axis.text.x = element_text(size=10))+#increasing the length of x axis
text
```

```
  xlab("Category")+
  ylab("In kilotons equivalent")+
  labs(title='Japan Gas Emissions',
       subtitle= 'in kilotons , from 1990-2018')
```

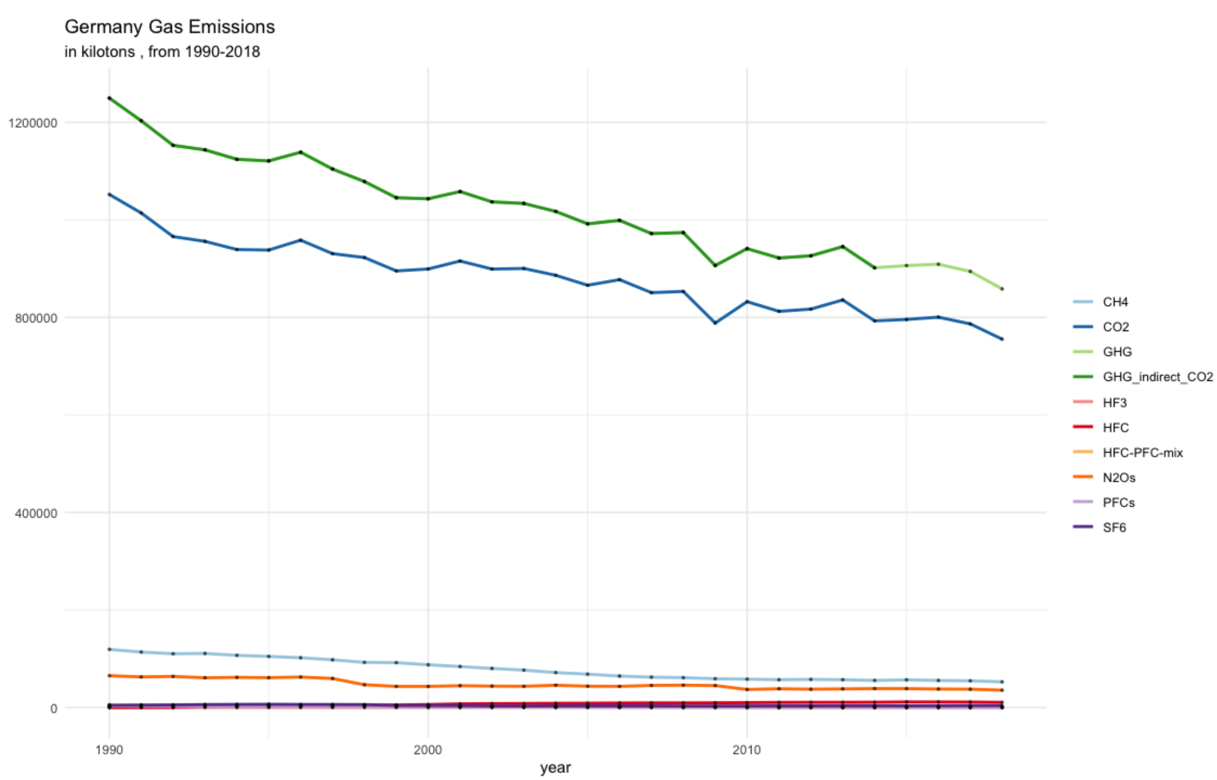



```
#lets find out the % of emission of each Japan gases from 1990 to 2018
japan_gas_wise_total <- japan_emissions%>% group_by(category) %>%summarise(to
tal_value = sum(value),.groups = 'drop')
#find the sum of total country gas emission
japan_sum <- sum(japan_gas_wise_total$total_value)
#create a column name 'emission_percentage and calculate the % of contributio
n'
japan_percentage <- mutate(japan_gas_wise_total, japan_emission_percentage =
(japan_gas_wise_total$total_value/japan_sum)*100)%>% mutate_at(vars(japan_emi
ssion_percentage), funs(round(., 3)))
japan_percentage

## # A tibble: 9 x 3
##   category         total_value japan_emission_percentage
## * <chr>                  <dbl>                     <dbl>
## 1 CH4                 1145197.                     0.978
## 2 CO2                35944195.                    30.7
## 3 GHG                39022712.                    33.3
## 4 GHG_indirect_CO2   39031111.                    33.3
## 5 HF3                   16772.                     0.014
## 6 HFC                  704438.                     0.602
## 7 N2Os                 750053.                     0.641
```

```
## 8 PFCs                       247705.                          0.212
## 9 SF6                        214351.                          0.183
```

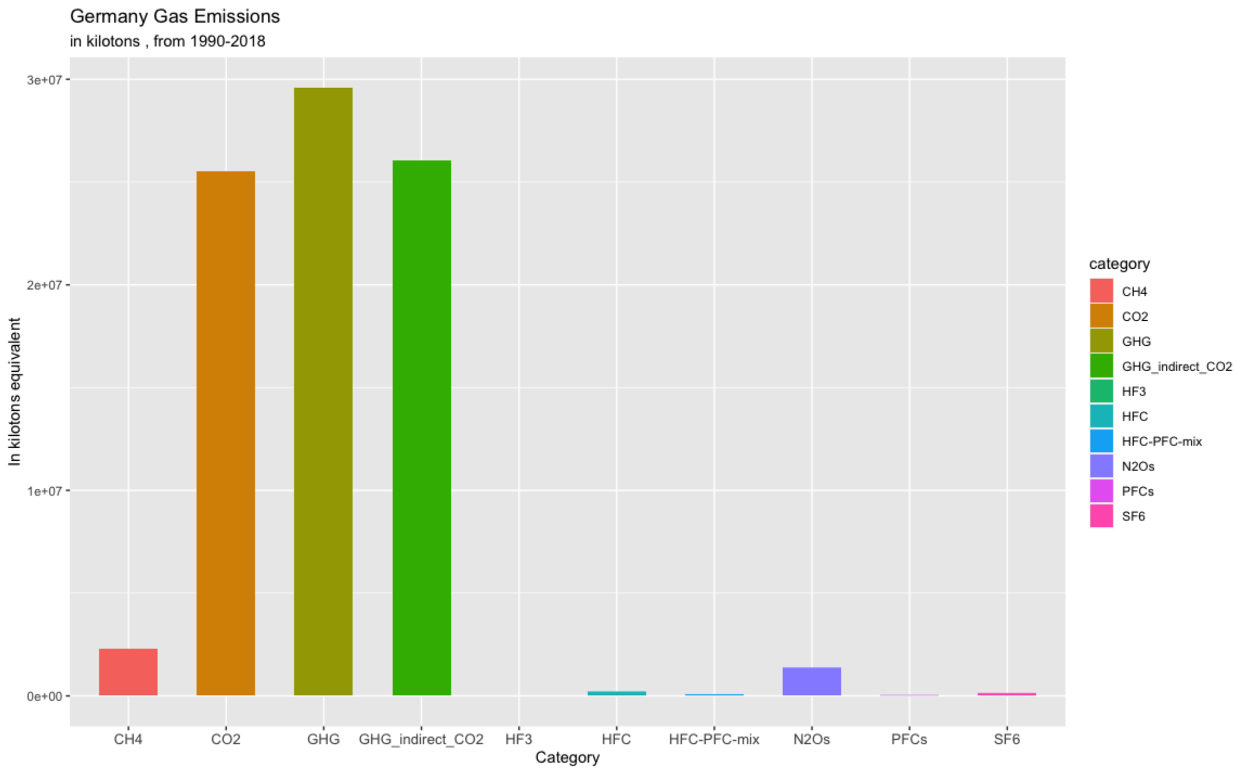
Germany emissions from 1990 to 2018

```
#analysis of Germany country emissions
germany_emissions <- dataset[dataset$country=='germany',] #selecting data for
a single country
germany_emissions %>% group_by(category, year) %>% #grouping category and yea
r together
  summarise(total_value=sum(value),.groups = 'drop')%>%
  ggplot(aes(x=year,y=total_value, col=category ))+#plotting a line plot
  geom_line(na.rm=TRUE, size=1)+
  geom_point(size=.5,alpha = .5,color = "black")+
  scale_color_brewer(name='',palette='Paired')+
  theme_minimal()+
  labs(title='Germany Gas Emissions',
       subtitle= 'in kilotons , from 1990-2018')+
  xlab('year')+ylab('')
```




```
#visualizing bar plot to know the portion of each emission gases and compare
with each other
 germany_emissions %>% group_by(category, year) %>%
  summarise(total_value=sum(value),.groups = 'drop')%>%
  ggplot(aes(x=category, y=total_value,fill = category))+
  geom_col(width = .6)+
  theme(axis.text.x = element_text(size=10))+#increasing the length of x axis
```

```
text
  xlab("Category")+
  ylab("In kilotons equivalent")+
  labs(title='Germany Gas Emissions',
       subtitle= 'in kilotons , from 1990-2018')
```




```
#lets find out the % of emission of each Japan gases from 1990 to 2018
germany_gas_wise_total <- germany_emissions%>% group_by(category) %>%summaris
e(total_value = sum(value),.groups = 'drop')
#find the sum of total country gas emission
germany_sum <- sum(germany_gas_wise_total$total_value)
#create a column name 'emission_percentage and calculate the % of contributio
n'
germany_percentage <- mutate(germany_gas_wise_total, germany_emission_percent
age = (germany_gas_wise_total$total_value/germany_sum)*100)%>% mutate_at(vars
(germany_emission_percentage), funs(round(., 3)))
germany_percentage

## # A tibble: 10 x 3
##    category         total_value germany_emission_percentage
##  * <chr>                  <dbl>                       <dbl>
##  1 CH4                 2270162.                       2.66
##  2 CO2                25538180.                      30.0
##  3 GHG                29598643.                      34.7
##  4 GHG_indirect_CO2   26030606.                      30.5
##  5 HF3                     514.                       0.001
##  6 HFC                  206257.                       0.242
```

```
## 7 HFC-PFC-mix            62914.                         0.074
## 8 N2Os                 1370390.                         1.61
## 9 PFCs                   31578.                         0.037
## 10 SF6                  118650.                         0.139
```

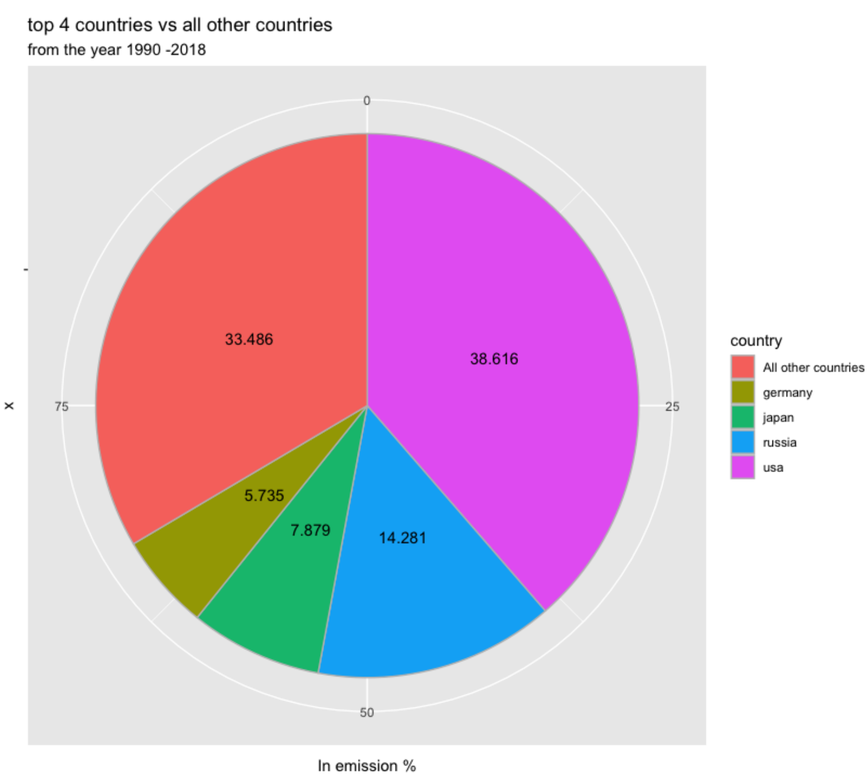
Total emission of gases around the world from 1990-2018 and each country's share %

```
total_gases_by_each_country <- dataset%>% group_by(country) %>%summarise(tota
l_value = sum(value),.groups = 'drop')
#find the sum of total country gas emission
total_sum <- sum(total_gases_by_each_country$total_value)
#create a column name 'emission_percentage and calculate the % of contributio
n'
each_country_percentage <- mutate(total_gases_by_each_country, country_emissi
on_percentage = (total_gases_by_each_country$total_value/total_sum)*100)%>% m
utate_at(vars(country_emission_percentage), funs(round(., 3)))
#sorting the data from highest to lowest using order
each_country_percentage<- each_country_percentage[order(each_country_percenta
ge$country_emission_percentage,decreasing = TRUE),]

each_country_percentage

## # A tibble: 43 x 3
##    country   total_value country_emission_percentage
##    <chr>           <dbl>                       <dbl>
##  1 usa        573840505.                       38.6
##  2 russia     212216466.                       14.3
##  3 japan      117076535.                        7.88
##  4 germany     85227893.                        5.74
##  5 canada      58744677.                        3.95
##  6 france      44216187.                        2.98
##  7 italy       43052403.                        2.90
##  8 australia   40841736.                        2.75
##  9 ukraine     40520112.                        2.73
## 10 uk          38679656.                        2.60
## # … with 33 more rows
```

#Top 4 countries vs all other countries contribution

```
#selecting top 4 countries data
top_4_countries <- each_country_percentage[1:4,]

#selcting remaining all other countries sum of 'total_value' and emission per
centage
remaining_39_countries <- each_country_percentage[5:43,]
total_value <- sum(remaining_39_countries$total_value)
country_emission_percentage <- sum(remaining_39_countries$country_emission_pe
rcentage)
country <- "All other countries"
rest_of_countries_data <- data.frame(country,total_value,country_emission_per
```

```
centage)
final_data <- rbind(top_4_countries,rest_of_countries_data)
final_data

## # A tibble: 5 x 3
##   country             total_value country_emission_percentage
##   <chr>                     <dbl>                       <dbl>
## 1 usa                  573840505.                       38.6
## 2 russia               212216466.                       14.3
## 3 japan                117076535.                        7.88
## 4 germany               85227893.                        5.74
## 5 All other countries  497642284.                       33.5

#graphical representation of top 4 countries VS all other countries
pie <- ggplot(final_data,aes(x="",y=country_emission_percentage,fill = countr
y))+
  geom_bar(stat = "identity",colour="gray")
pie+coord_polar("y", start=0,direction = 1)+
  geom_text(aes(label = country_emission_percentage), position = position_sta
ck(vjust = 0.5))+
  ggtitle("top 4 countries vs all other countries",subtitle = "from the year
1990 -2018")+
  ylab("In emission %")
```



key findings and conclusions

From the year 1990 - 2018, around the world the total amount of gas emission is '1486003683' kilotons. from these there are 4 gases which contributes a large amount of emission, they are i) GHG ii) GHG_indirect_CO2 iii) CO2 iv) CH4

"GHG"contributes 35.29% of total emission from the year 1990 - 2018, followed by 'GHG_indirect_CO2' which accounts for 29.41% of total emission and 'CO2' accounts for 27.91% of total emission and 'CH4' contributes to 4.81% of total emissions.

out of total 10 different gases, these 4 gases combined contributing 97.42% of gas emissions where rest of gases contributing 2.58%

GHG -(total contribution 35.29%)

the top GHG producing countries are–'USA','Russia','Japan' and 'Germany'

GHG_indirect_CO2 -(total contribution 29.42%)

the top indirect CO2 producing countries are – 'USA','Russia','Japan',and 'Germany'

CO2 (total contribution 27.91%)

the top CO2 producing countries are – 'USA','Russia','Japan',and 'Germany'

CH4 (total contribution 4.81%)

the top CH4 producing countries are – 'USA','Russia','Ukraine',and 'Canada'

USA, Russia, Japan and Germany these are the common countries that produce top high gases.

out of all countries, these 4 countries together producing 66.52% of total emissions which is huge and out of 66.52% of emissions, USA itself producing 38.62% of total emissions which has major impact to the environment followed by Russia which contributes 14.28% of emission.

in general The largest source of greenhouse gas emissions from human activities is from burning fossil fuels for electricity, heat, and transportation.

these 4 countries are major manufacturers and exports of industries includes automobiles, consumer electronics, computers,aerospace, semiconductors, and iron and steel.

By initiating controlling of gas emissions with these 4 countries will bring down the overall global impact.

if these 4 countries can adapt to renewable energy to store the power of wind,sun,water,tides and other planetary resources like geothermal heat, which comes from the Earth's core to produce electric power.

Agricultural "biomass" products also can be used to generate electricity and heat.Renewables generate electricity without producing greenhouse gases—or producing very little when compared to traditional energy sources.

Using recycled materials will also helps to minimize the green house gas emissions

This is the best way to work around to minimize the green house gas emissions.