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 (CO2), methane (CH4), nitrous oxide (N2O), hydrofluorocarbons (HFCs), perfluorocarbons (PFCs), unspecified mix of HFCs and PFCs, sulphur hexafluoride (SF6) and nitrogen triflouride (NF3)) 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 (CO2) methane (CH4) nitrous oxide (N2O) hydrofluorocarbons (HFCs) perfluorocarbons (PFCs) unspecified mix of HFCs and PFCs sulphur hexafluoride (SF6) nitrogen triflouride (NF3)

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=',',stringsAsFactors = 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\_forestry\_lulucf\_in\_kilotonne\_co2\_equivalent  
## 2 carbon\_dioxide\_co2\_emissions\_without\_land\_use\_land\_use\_change\_and\_forestry\_lulucf\_in\_kilotonne\_co2\_equivalent  
## 3 carbon\_dioxide\_co2\_emissions\_without\_land\_use\_land\_use\_change\_and\_forestry\_lulucf\_in\_kilotonne\_co2\_equivalent  
## 4 carbon\_dioxide\_co2\_emissions\_without\_land\_use\_land\_use\_change\_and\_forestry\_lulucf\_in\_kilotonne\_co2\_equivalent  
## 5 carbon\_dioxide\_co2\_emissions\_without\_land\_use\_land\_use\_change\_and\_forestry\_lulucf\_in\_kilotonne\_co2\_equivalent  
## 6 carbon\_dioxide\_co2\_emissions\_without\_land\_use\_land\_use\_change\_and\_forestry\_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\_land\_use\_change\_and\_forestry\_lulucf\_in\_kilotonne\_co2\_equivalent" "carbon\_dioxide\_co2\_emissions\_without\_land\_use\_land\_use\_change\_and\_forestry\_lulucf\_in\_kilotonne\_co2\_equivalent" "carbon\_dioxide\_co2\_emissions\_without\_land\_use\_land\_use\_change\_and\_forestry\_lulucf\_in\_kilotonne\_co2\_equivalent" "carbon\_dioxide\_co2\_emissions\_without\_land\_use\_land\_use\_change\_and\_forestry\_lulucf\_in\_kilotonne\_co2\_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\_use\_land\_use\_change\_and\_forestry\_lulucf\_in\_kilotonne\_co2\_equivalent='CO2',  
 greenhouse\_gas\_ghgs\_emissions\_including\_indirect\_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\_kilotonne\_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\_kilotonne\_co2\_equivalent='HF3',  
 nitrous\_oxide\_n2o\_emissions\_without\_land\_use\_land\_use\_change\_and\_forestry\_lulucf\_in\_kilotonne\_co2\_equivalent='N2Os',  
 perfluorocarbons\_pfcs\_emissions\_in\_kilotonne\_co2\_equivalent='PFCs',  
 sulphur\_hexafluoride\_sf6\_emissions\_in\_kilotonne\_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 uniue 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 some of our input values are not formatted properly, because they contain commas (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" ...

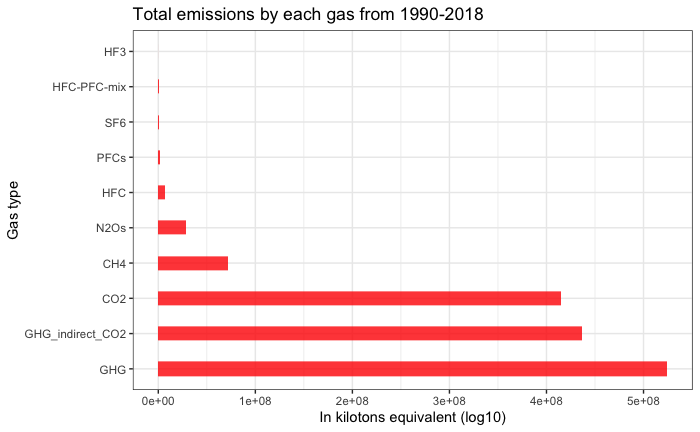
## 3.Data analysis

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$total\_value)  
#creating a variable 'each\_emission\_percentage' to store the percentage of contributions  
each\_gas\_percent <- mutate(total\_emissions, each\_emission\_percentage = (total\_emissions$total\_value/total\_gas\_sum)\*100)%>% mutate\_at(vars(each\_emission\_percentage), 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

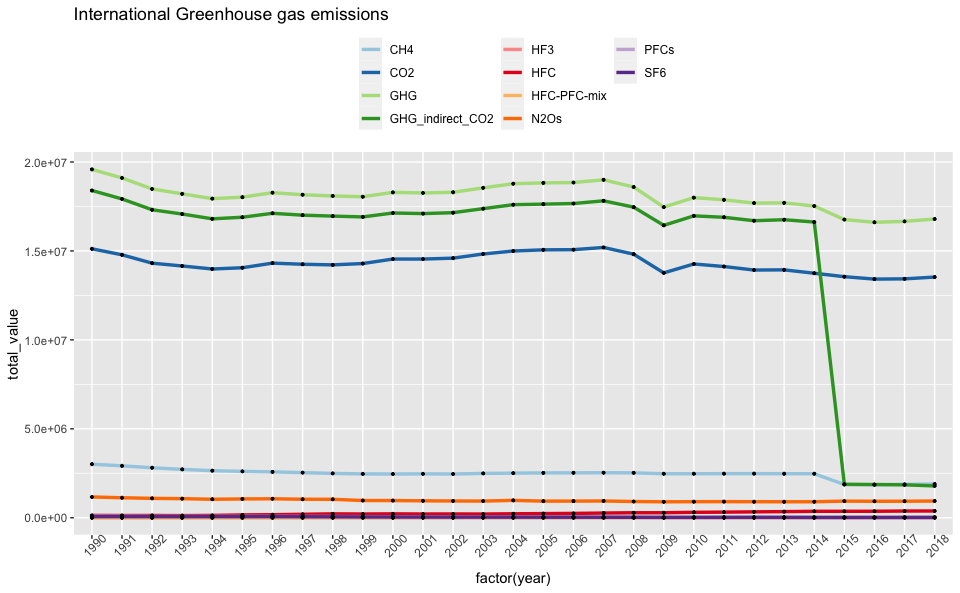
#Visualizing overall emissions of each gas from 1990-2018  
ggplot(total\_emissions, aes(x=reorder(category,-total\_value), y=total\_value)) +  
 geom\_bar(stat="identity", fill="red", alpha=.8, width=.4) +  
 xlab("Gas type") +  
 ylab("In kilotons equivalent (log10)")+  
 ggtitle("Total emissions by 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=category)) + 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))



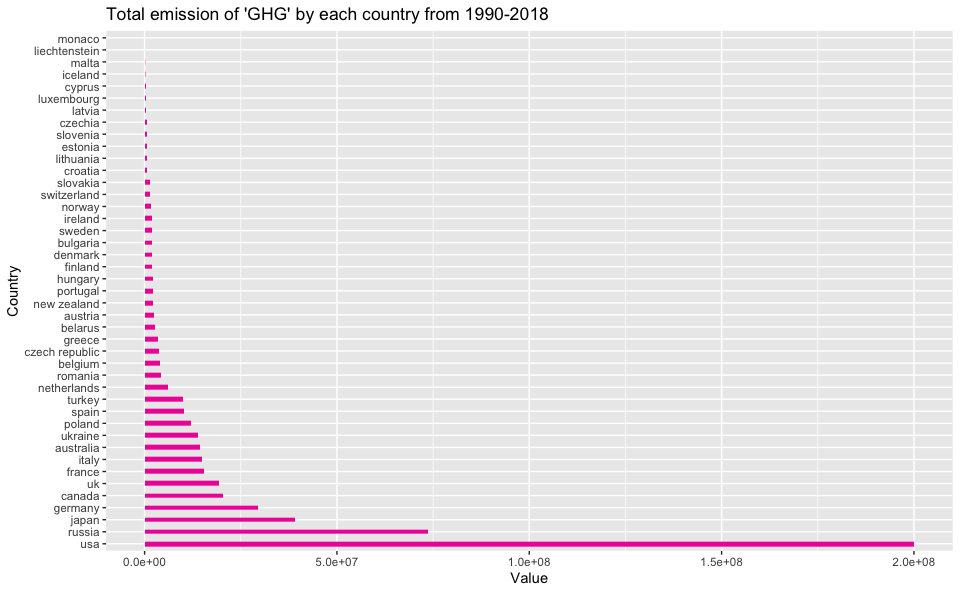
by looking at the above graph, ‘GHG’, ‘GHG\_indirect\_CO2’ and ‘CO2’ are the most emissioned gas over the past years and there is a substantial drop of ‘’GHG\_indirect\_gas’ from the year 2015 to 2018 and few gases have a little amount of ups and downs from where they started but they continued on the 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

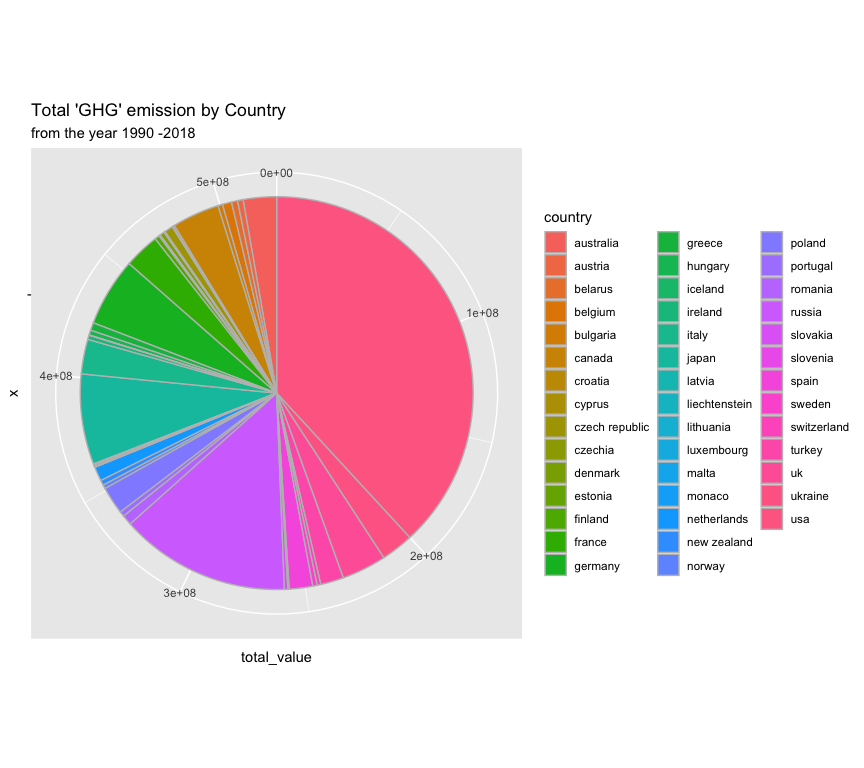
#selecting a subset of data that only contains "GHG" gas emission  
GHG\_data <- dataset[dataset$category=='GHG',]  
#grouping by each country to know which countries produced highest GHG  
GHG\_by\_country <- GHG\_data%>% group\_by(country) %>%summarise(sum(value))  
colnames(GHG\_by\_country)[2] <- "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.

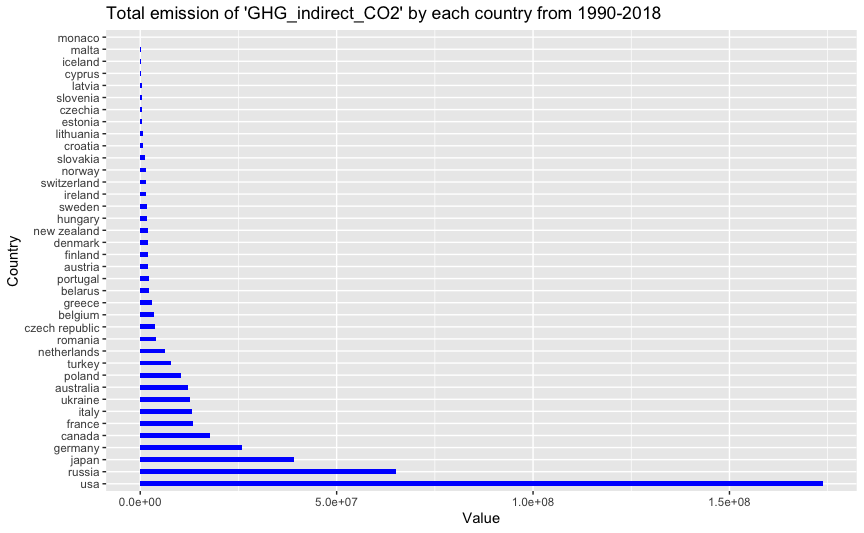
#visualizing % of contribution of each country  
pie <- ggplot(GHG\_by\_country,aes(x="",y=total\_value,fill = country))+  
 geom\_bar(stat = "identity",colour="gray")  
pie+coord\_polar("y", start=0,direction = 1)+  
 ggtitle("Total 'GHG' emission by Country",subtitle = "from the year 1990 -2018")



Top contributers from above graph

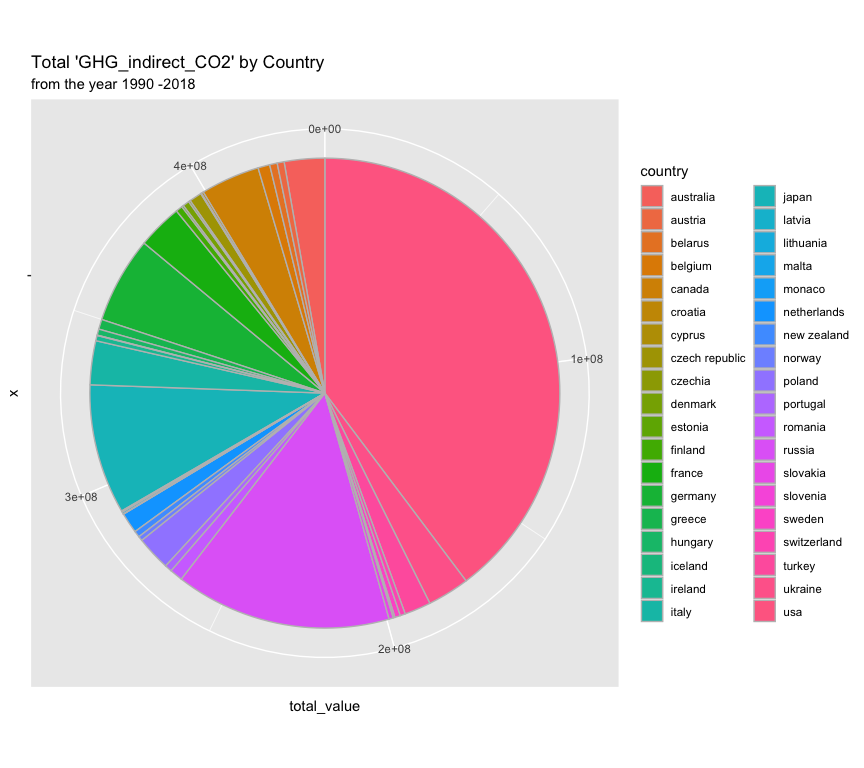
# 2. Analysis of “GHG\_indirect\_CO2” gas emission

#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(total\_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\_value)) +  
 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 1990-2018")+  
 coord\_flip(expand = TRUE)



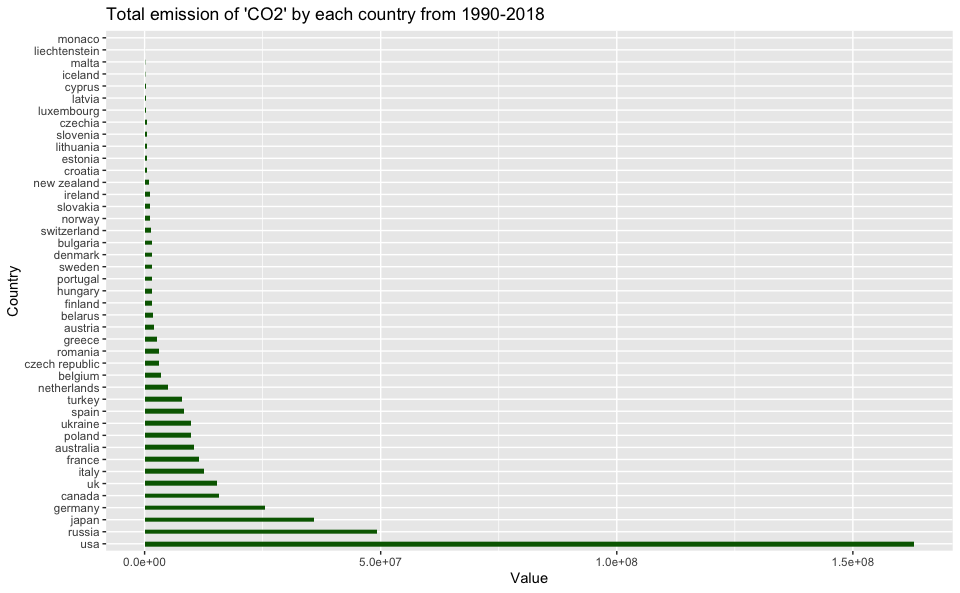
countries– usa,russia,japan,germany, canada,france and italy are at top and all of these combine contributing a major chunk of ‘GHG\_indirect\_CO2’ emission.

#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 1990 -2018")



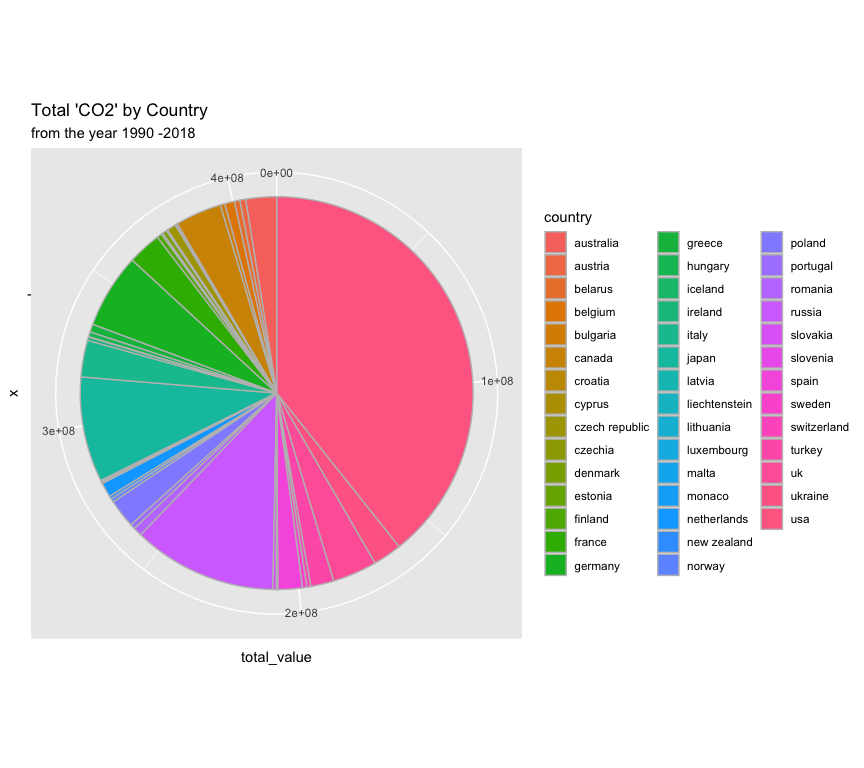
# 3. Analysis of “CO2” gas emission

#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 = sum(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)



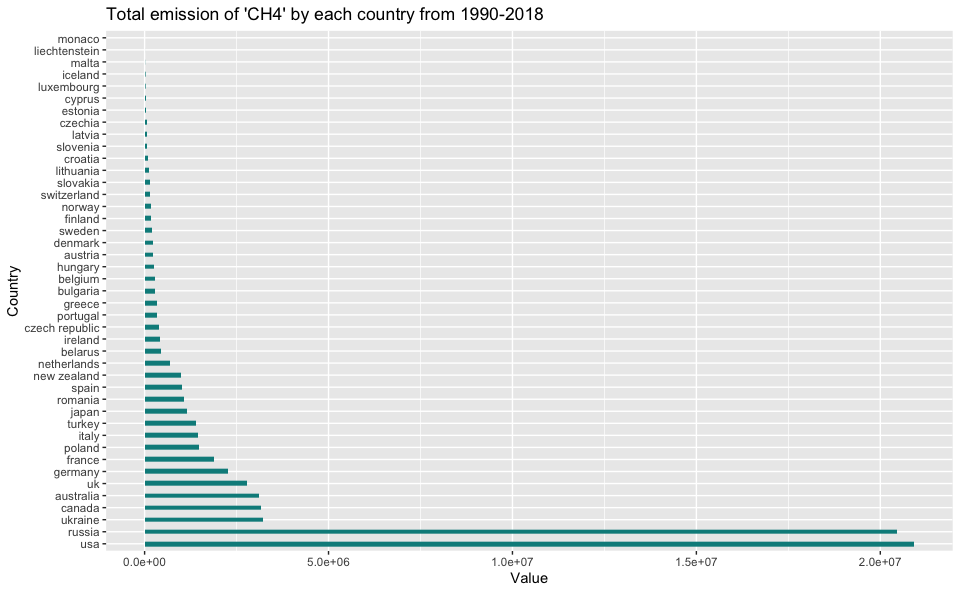
the Top ‘CO2’ emission countries are = usa, russia,japan,germany,canada,uk and Italy

#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")



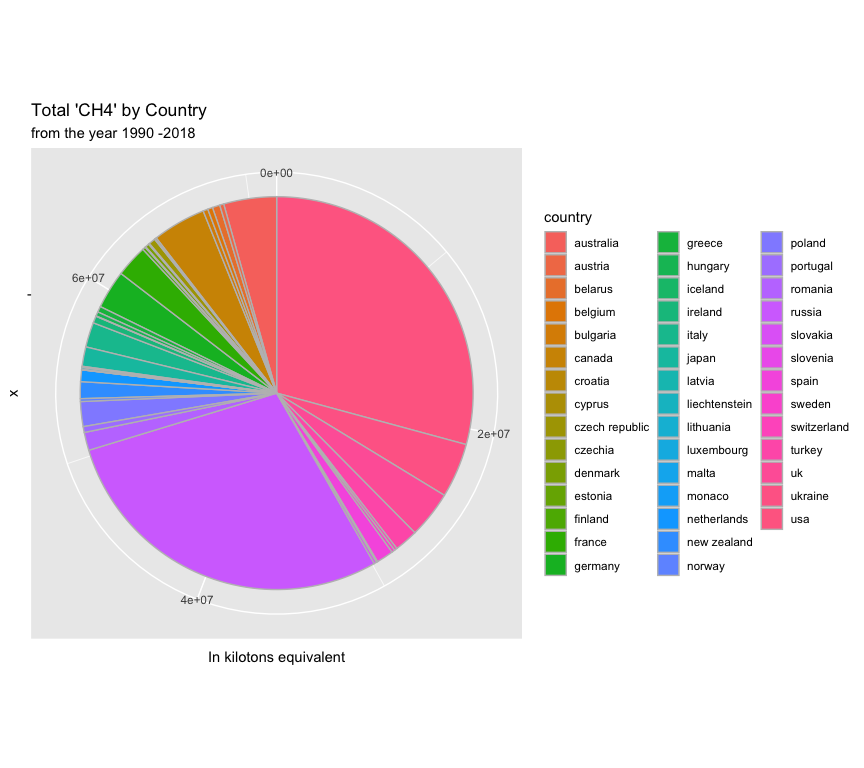
# 4. Analysis of “CH4” gas emission

#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 = sum(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)

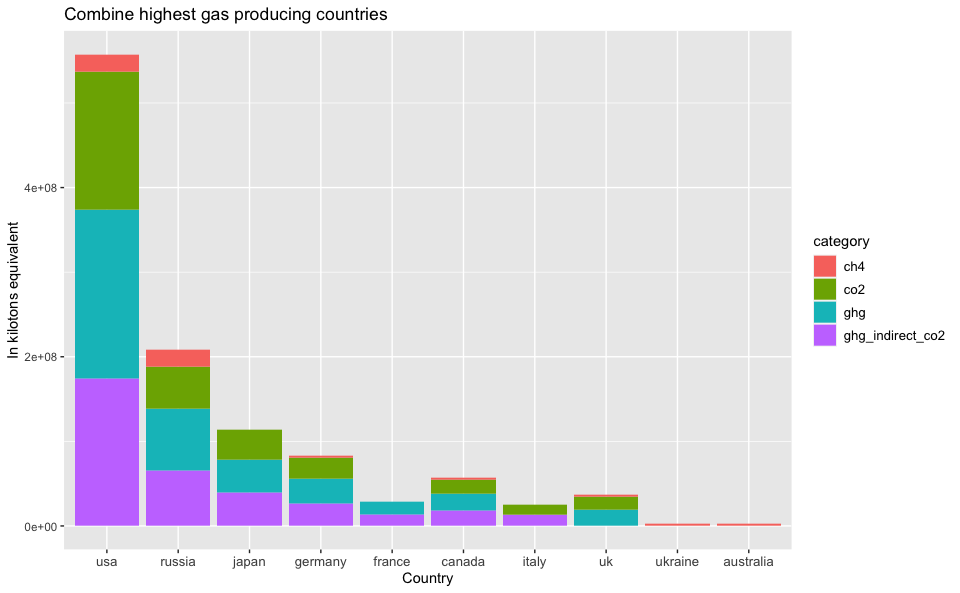
#adding a new column for each type of gas   
category <- c('ghg\_indirect\_co2','ghg\_indirect\_co2','ghg\_indirect\_co2','ghg\_indirect\_co2','ghg\_indirect\_co2','ghg\_indirect\_co2','ghg\_indirect\_co2')  
GHG\_indirect\_top <- data.frame(GHG\_indirect\_country[c(1:7),],category)

#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)

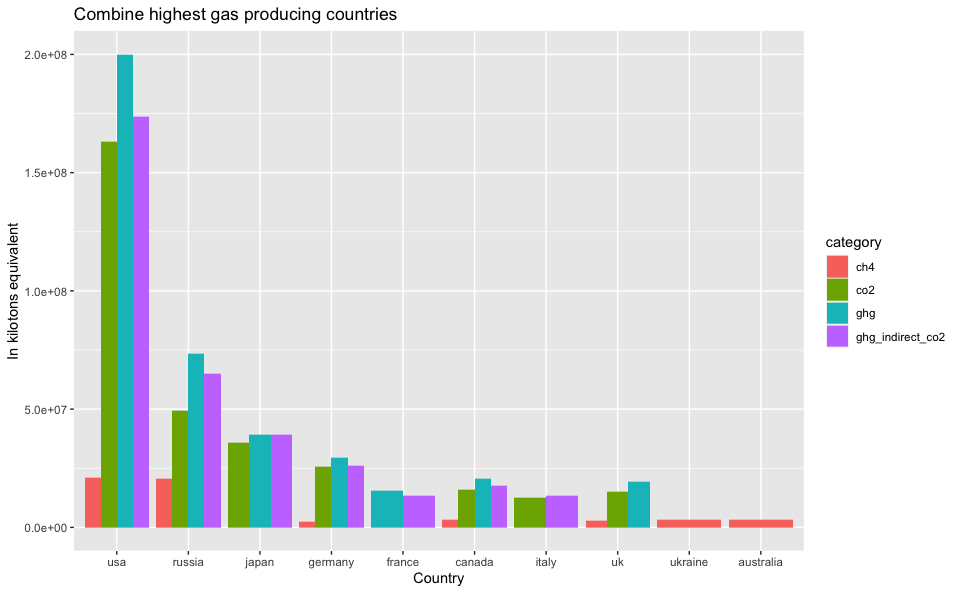
#adding a new column for each type of gas   
category <-c('ch4','ch4','ch4','ch4','ch4','ch4','ch4')  
CH4\_top\_countries <-data.frame(CH4\_by\_country[c(1:7),],category)

#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\_countries,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



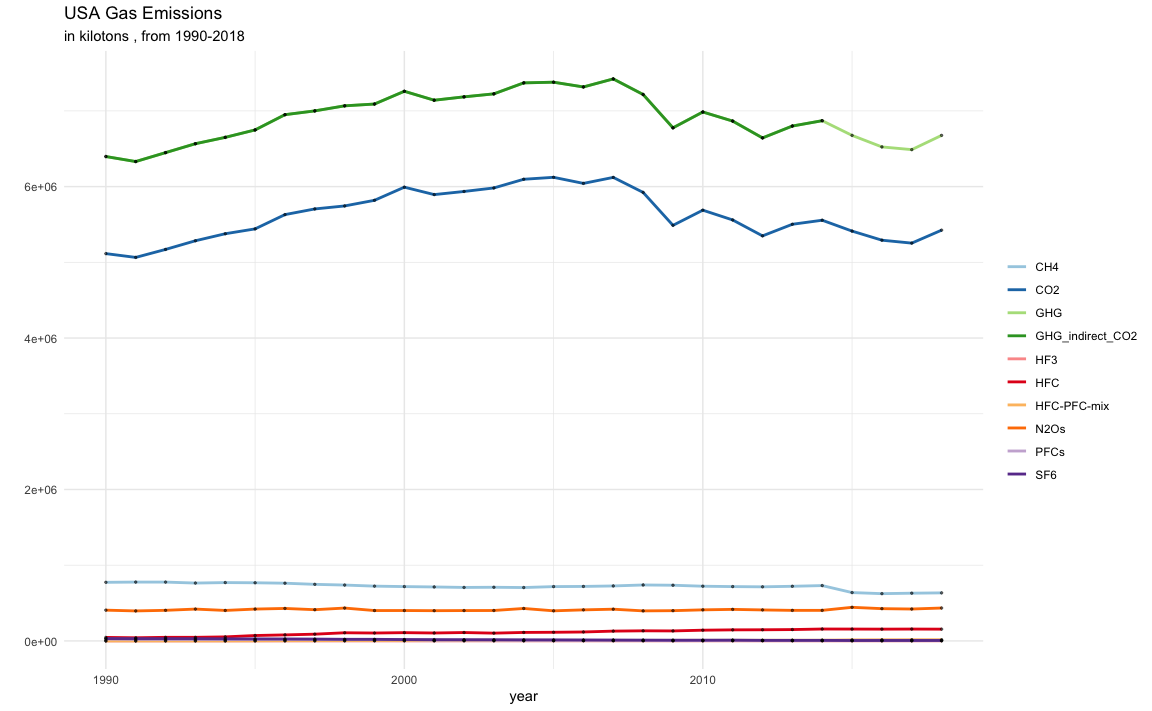
#changing the bars to individual  
ggplot(data = total\_top\_countries)+  
 geom\_col(mapping = aes(x = reorder(country, -total\_value), y = total\_value, fill = category),position = "dodge")+  
 xlab("Country")+  
 ylab("In kilotons equivalent")+  
 ggtitle("Combine highest gas producing countries")



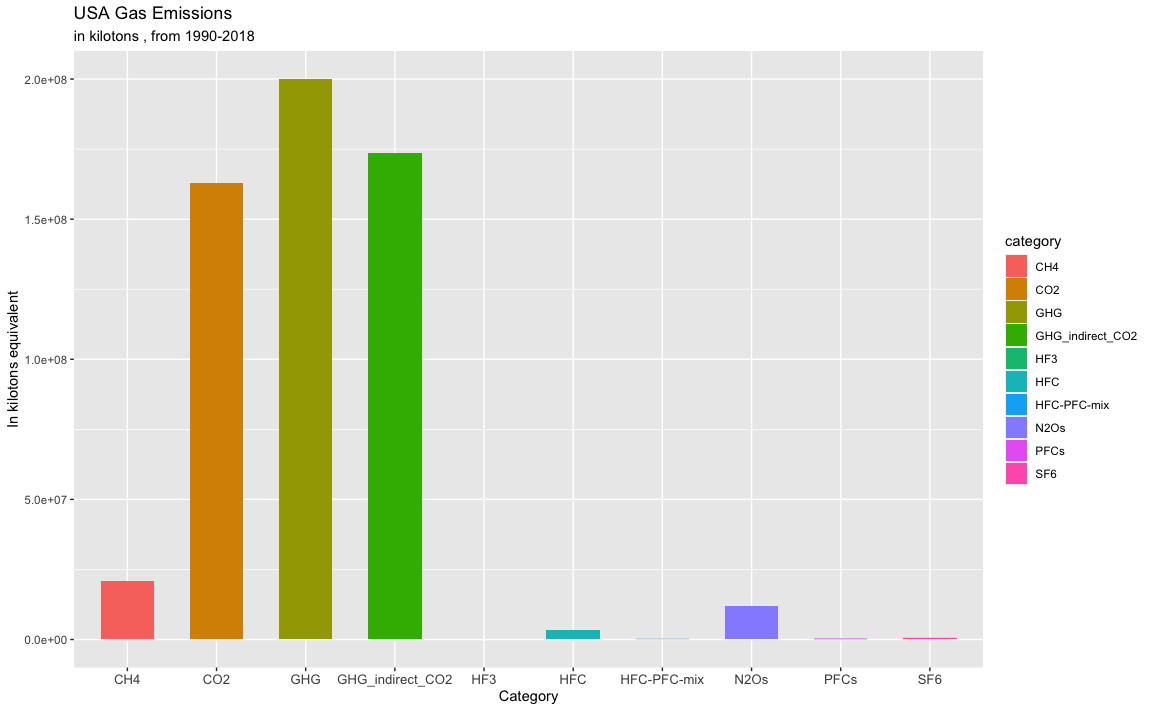
from above plot we can clearly see that countries “USA”,“Russia”,“Japan” and “Germany” contributing a high portion of these gases and keeping these observations in mind now let’s look at these four countries individual gas emissions over the years.

**USA emissions from 1990 to 2018**

#analysis of USA country emissions   
usa\_emissions <- dataset[dataset$country=='usa',] #selecting data for a single country  
usa\_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='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')

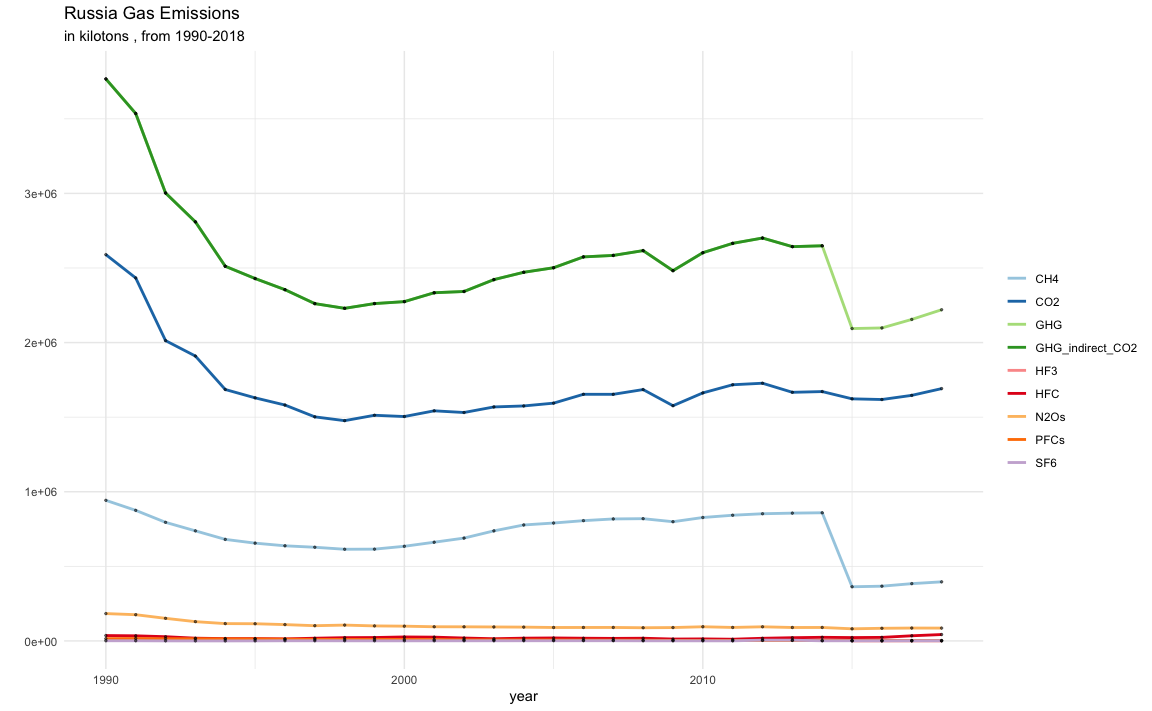


#lets find out the % of emission of each USA gases from 1990 to 2018  
usa\_gas\_wise\_total <- usa\_emissions%>% group\_by(category) %>%summarise(total\_value = sum(value),.groups = 'drop')  
#find the sum of total country gas emission  
usa\_sum <- sum(usa\_gas\_wise\_total$total\_value)  
#create a column name 'emission\_percentage and calculate the % of contribution'  
usa\_percentage <- mutate(usa\_gas\_wise\_total, usa\_emission\_percentage = (usa\_gas\_wise\_total$total\_value/usa\_sum)\*100)%>% mutate\_at(vars(usa\_emission\_percentage), funs(round(., 3)))  
usa\_percentage

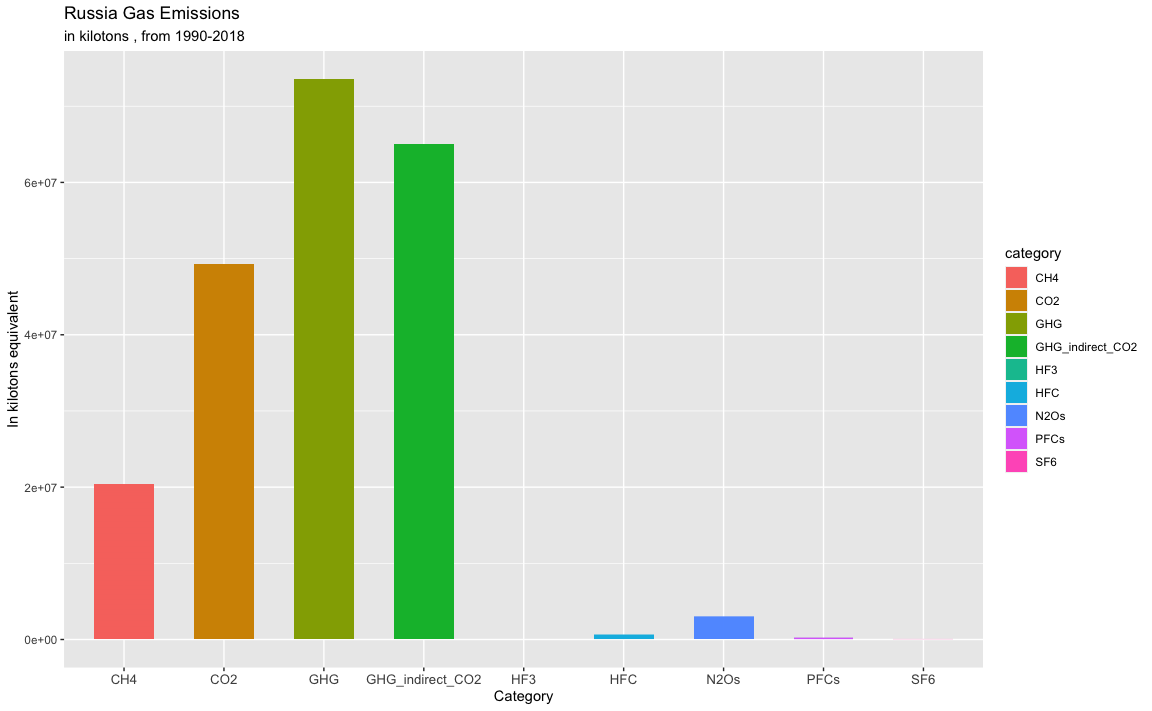
## # A tibble: 10 x 3  
## category total\_value usa\_emission\_percentage  
## \* <chr> <dbl> <dbl>  
## 1 CH4 20922010. 3.65   
## 2 CO2 162992307. 28.4   
## 3 GHG 200068613. 34.9   
## 4 GHG\_indirect\_CO2 173703277. 30.3   
## 5 HF3 11054. 0.002  
## 6 HFC 3220566. 0.561  
## 7 HFC-PFC-mix 181693. 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")+  
 labs(title='Russia Gas Emissions',  
 subtitle= 'in kilotons , from 1990-2018')

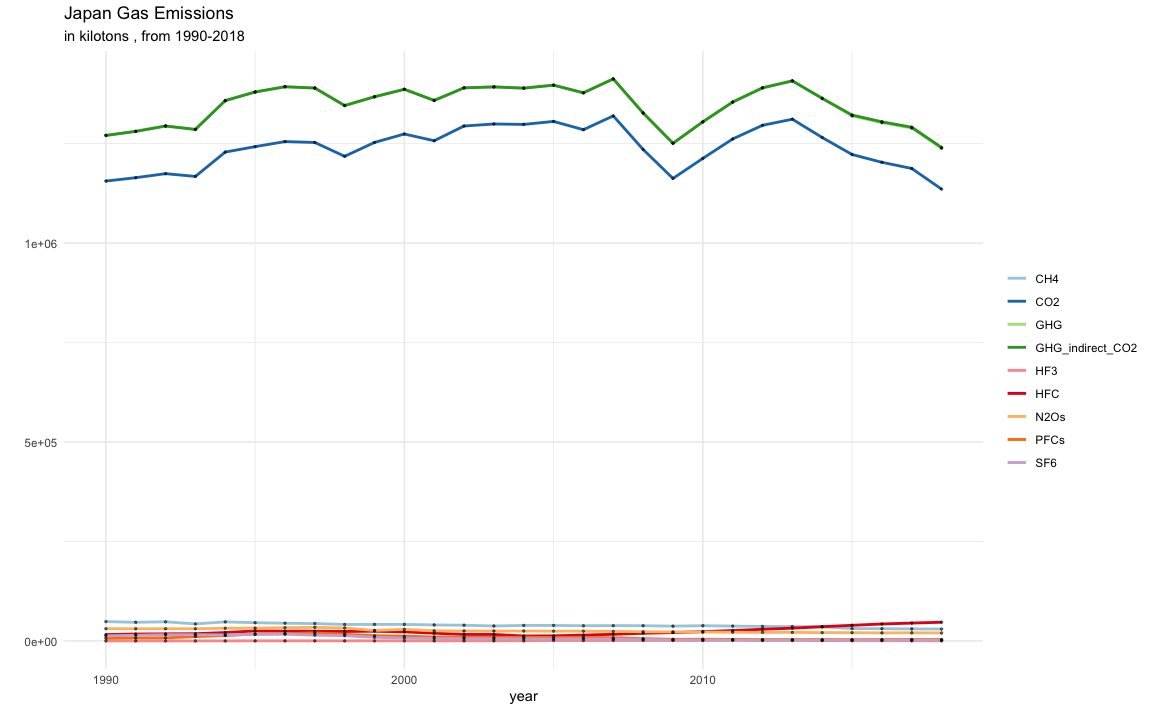


#lets find out the % of emission of each Russia gases from 1990 to 2018  
russia\_gas\_wise\_total <- russia\_emissions%>% group\_by(category) %>%summarise(total\_value = sum(value),.groups = 'drop')  
#find the sum of total country gas emission  
russia\_sum <- sum(russia\_gas\_wise\_total$total\_value)  
#create a column name 'emission\_percentage and calculate the % of contribution'  
russia\_percentage <- mutate(russia\_gas\_wise\_total, russia\_emission\_percentage = (russia\_gas\_wise\_total$total\_value/russia\_sum)\*100)%>% mutate\_at(vars(russia\_emission\_percentage), funs(round(., 3)))  
russia\_percentage

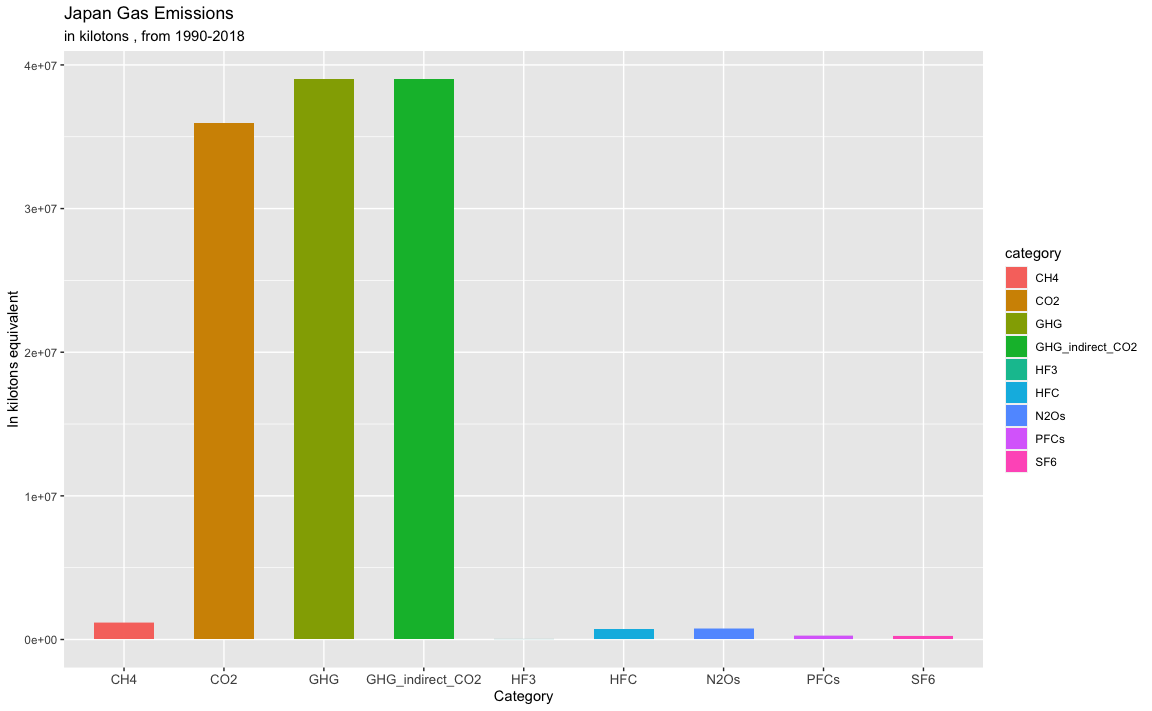
## # A tibble: 9 x 3  
## category total\_value russia\_emission\_percentage  
## \* <chr> <dbl> <dbl>  
## 1 CH4 20462518. 9.64   
## 2 CO2 49241132. 23.2   
## 3 GHG 73594670. 34.7   
## 4 GHG\_indirect\_CO2 65027127. 30.6   
## 5 HF3 9 0   
## 6 HFC 624129. 0.294  
## 7 N2Os 3009207. 1.42   
## 8 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 single 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')

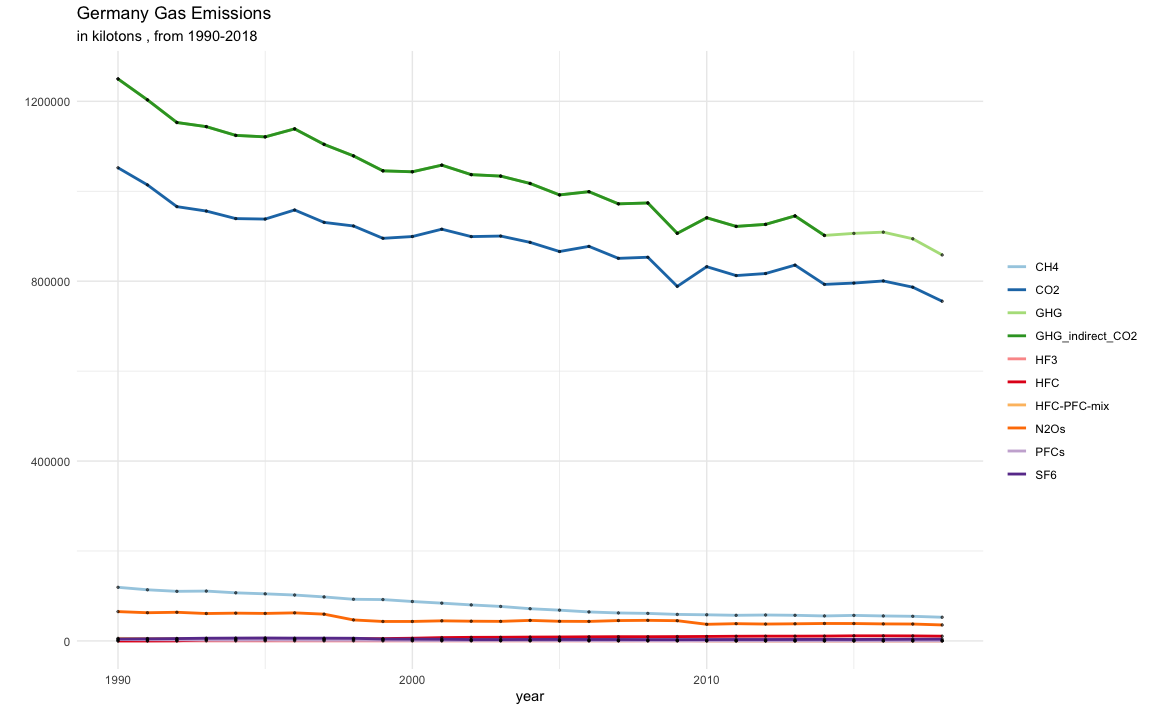


#lets find out the % of emission of each Japan gases from 1990 to 2018  
japan\_gas\_wise\_total <- japan\_emissions%>% group\_by(category) %>%summarise(total\_value = sum(value),.groups = 'drop')  
#find the sum of total country gas emission  
japan\_sum <- sum(japan\_gas\_wise\_total$total\_value)  
#create a column name 'emission\_percentage and calculate the % of contribution'  
japan\_percentage <- mutate(japan\_gas\_wise\_total, japan\_emission\_percentage = (japan\_gas\_wise\_total$total\_value/japan\_sum)\*100)%>% mutate\_at(vars(japan\_emission\_percentage), funs(round(., 3)))  
japan\_percentage

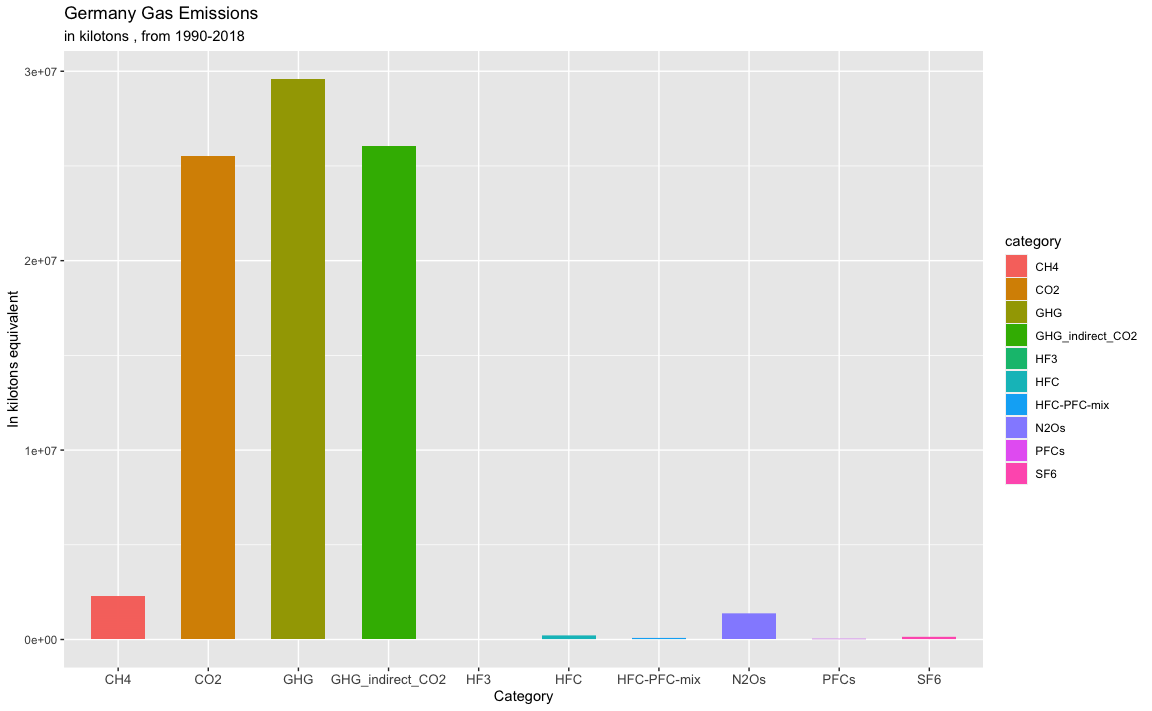
## # A tibble: 9 x 3  
## category total\_value japan\_emission\_percentage  
## \* <chr> <dbl> <dbl>  
## 1 CH4 1145197. 0.978  
## 2 CO2 35944195. 30.7   
## 3 GHG 39022712. 33.3   
## 4 GHG\_indirect\_CO2 39031111. 33.3   
## 5 HF3 16772. 0.014  
## 6 HFC 704438. 0.602  
## 7 N2Os 750053. 0.641  
## 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 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='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) %>%summarise(total\_value = sum(value),.groups = 'drop')  
#find the sum of total country gas emission  
germany\_sum <- sum(germany\_gas\_wise\_total$total\_value)  
#create a column name 'emission\_percentage and calculate the % of contribution'  
germany\_percentage <- mutate(germany\_gas\_wise\_total, germany\_emission\_percentage = (germany\_gas\_wise\_total$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(total\_value = sum(value),.groups = 'drop')  
#find the sum of total country gas emission  
total\_sum <- sum(total\_gases\_by\_each\_country$total\_value)  
#create a column name 'emission\_percentage and calculate the % of contribution'  
each\_country\_percentage <- mutate(total\_gases\_by\_each\_country, country\_emission\_percentage = (total\_gases\_by\_each\_country$total\_value/total\_sum)\*100)%>% mutate\_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\_percentage$country\_emission\_percentage,decreasing = TRUE),]  
  
each\_country\_percentage

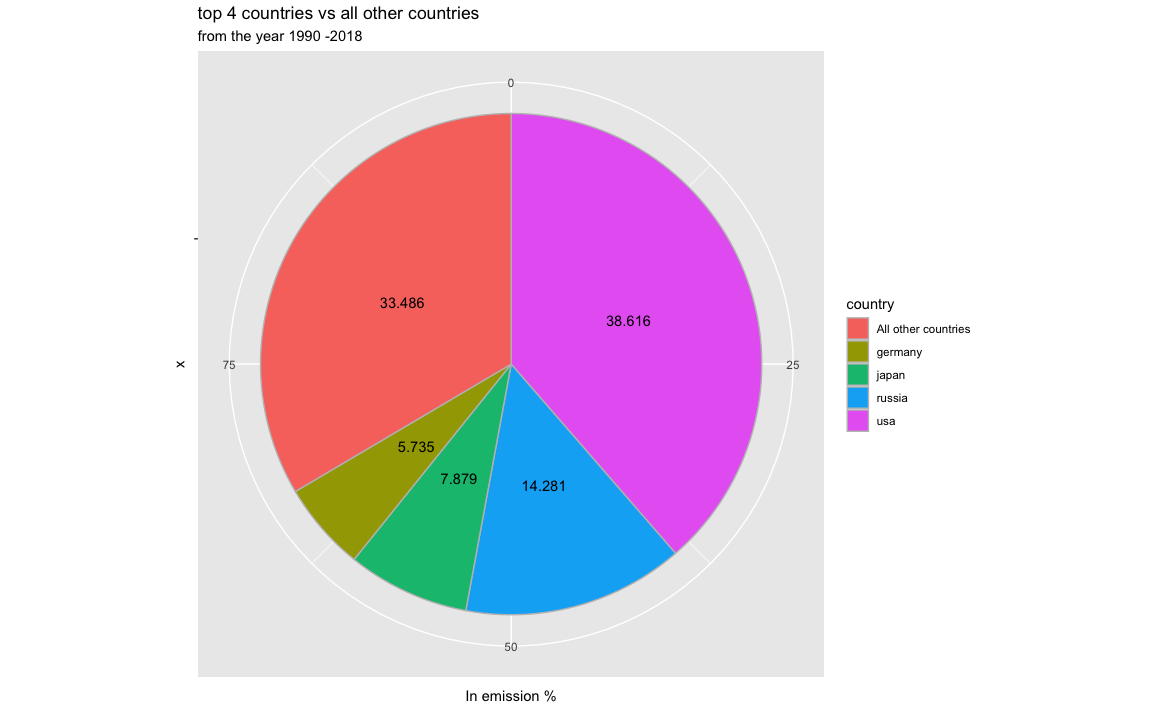
## # A tibble: 43 x 3  
## country total\_value country\_emission\_percentage  
## <chr> <dbl> <dbl>  
## 1 usa 573840505. 38.6   
## 2 russia 212216466. 14.3   
## 3 japan 117076535. 7.88  
## 4 germany 85227893. 5.74  
## 5 canada 58744677. 3.95  
## 6 france 44216187. 2.98  
## 7 italy 43052403. 2.90  
## 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 percentage  
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\_percentage)  
country <- "All other countries"  
rest\_of\_countries\_data <- data.frame(country,total\_value,country\_emission\_percentage)  
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 = country))+  
 geom\_bar(stat = "identity",colour="gray")  
pie+coord\_polar("y", start=0,direction = 1)+  
 geom\_text(aes(label = country\_emission\_percentage), position = position\_stack(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.