Visualisation Semester Project: Component 1

Global Environment Indicators

Ashutosh Maurya, MDS202110

The dataset used is the Global Environment Indicators dataset, supplied by the United Nations Statistics Division. The dateset has data on several environment indicators which help us to understand and analyze the health of the planet, and take actions based on these statistics. The indicators include Energy and Minerals, Forests, Waste, and many more. For this project, the data on Greenhouse Emissions from Air and Climate indicators, and data on Occurrences of Natural Disasters is taken.

The purpose of the project is to study and analyse the greenhouse emissions of the countries, and compare it with the occurrences of natural disasters around the globe. This can shed some light on the affect of harmful emissions on the planet's ecosystems and their stability, which in turn leads to all kinds of natural disasters, from droughts to floods. From the vast dataset, we take a handful of variables and do some exploratory analysis using visualisation.

Introduction to the dataset

From Air and Climate indicators, we take `GHG_Emissions.csv` and `GHG_Emissions_by_Sector.csv`, from which we take the variables `Country ID`, `Country`, `GHG total without LULUCF, latest year`, `% change since 1990` and `GHG emissions per capita, latest year`, `GHG from energy, as percentage to total`, `GHG from industrial processes and product use, as percentage to total`, `GHG from agriculture, as percentage to total`, `GHG from waste, as percentage to total`.

From Natural Disasters indicator, we take `Climatological disasters.csv`,`Geophysical disasters.csv`, `Hydrological disasters.csv` and `Meteorological disasters.csv`, from which we take the variables `CountryID`, `Countries or areas`, `Occurrence 1990-1999`, `Occurrence 1999-2009` and `Occurrence 2009-2019` from each file.

A brief description of the variables is given below.

Variable Name	**Type of Variable**	**Description**
`Country ID`	Nominal categorical	Unique number to identify country
Country	Nominal categorical	Name of country
`GHG total without LULUCF, latest year`	Continuous numerical	Total emissions of greenhouse gases in latest year (whose data is available), excluding emissions from LULUCEF (commercial uses, land-use change, and forestry) activities (in 1000 tonnes of CO_2 equivalent)
`% change since 1990`	Continuous numerical	Percentage change in total emissions from 1990 to latest year
`GHG emissions per capita, latest year`	Continuous numerical	Per capita emissions of greenhouse gases (in 1000 tonnes of CO_2 equivalent)

Table 1: Variables from `GHG_Emissions.csv`

Variable Name	Type of Variable	Description
`Country ID`	Nominal categorical	Unique number to identify country
`Country`	Nominal categorical	Name of country
`GHG from energy, as percentage to total`	Continuous numerical	Percentage of GHG gases emissions in energy production
`GHG from industrial processes and product use, as percentage to total`	Continuous numerical	Percentage of GHG gases emissions in industrial processes and product use
`GHG from agriculture, as percentage to total`	Continuous numerical	Percentage of GHG gases emissions in agriculture
`GHG from waste, as percentage to total`	Continuous numerical	Percentage of GHG gases emissions in waste disposal and management

Table 2: Variables from `GHG_Emissions_by_Sector.csv`

From the documentation, Natural Disasters are classified as:

- Climatological disasters: Hazards caused by long-lived, meso- to macro-scale atmospheric processes ranging from intra-seasonal to multi-decadal climate variability. They are sub-classified as Drought, Glacial Lake Outburst and Wildfires.
  Drought is an extended period of unusually low precipitation that produces a shortage of water for people, animals and plants. Glacial lake outburst is a flood that occurs when water dammed by a glacier is suddenly released.

- Geophysical disasters: Hazards originating from solid earth. They are classified as: Earthquakes, Mass Movements and Volcanic Activities.
  Earthquake is sudden movement of a block of the Earth's crust along a geological fault and associated ground shaking. Mass movement is any type of downslope movement of earth materials. Volcanic activity is a type of volcanic event near an opening/vent in the Earth's surface including volcanic eruptions of lava, ash, hot vapour, gas, and pyroclastic material.

- Hydrological disasters: Hazards caused by the occurrence, movement, and distribution of surface and subsurface freshwater and saltwater.They are further classified as: Flood, Landslide and Wave Action. Flood is a general term for the overflow of water from a stream channel onto normally dry land in the floodplain, higher-than-normal levels along the coast and in lakes or reservoirs, as well as ponding of water at or near the point where the rain fell. Landslide is the movement of soil or rock controlled by gravity and the speed of the movement usually ranges between slow and rapid, but not very slow.

- Meteorological disasters: Hazards caused by short-lived, micro- to meso-scale extreme weather and atmospheric conditions that last from minutes to days. They include extreme temperature, fog, and storm

Variable Name	Type of Variable	Description
`CountryID`	Nominal categorical	Unique number to identify country
`Countries or areas`	Nominal categorical	Name of country
`Occurrence 1990-1999`	Discrete numerical	Number of occurrences of disaster in the decade 1990-1999
`Occurrence 1999-2009`	Discrete numerical	Number of occurrences of disaster in the decade 2000-2009
`Occurrence 2009-2019`	Discrete numerical	Number of occurrences of disaster in the decade 2010-2019

Table 3: Variables from each of `Climatological disasters.csv`, `Hydrological disasters.csv`, `Meteorological disasters.csv` and `Geophysical disasters.csv`

Loading and cleaning the data

The required variables are extracted from the respective tables. In the disasters data, there are missing values present as "…" in the data, and since the variable is the number of occurrences of disasters, the missing values can be handled by assuming them as zero.

For the greenhouse gas data, the data with sector wise emissions is merged with the original data using Country ID as the key to facilitate visualisation.

Numerical variables are stored as characters in the dataset, so they are converted to integers or decimals in the cleaning process.

```
knit_print.data.frame = function(x, ...) {
  res = paste(c("", "", knitr::kable(x)), collapse = "\n")
  knitr::asis_output(res)
}

registerS3method(
  "knit_print", "data.frame", knit_print.data.frame,
  envir = asNamespace("knitr")
)
```

```
climate_dis = read.csv('Climatological disasters.csv')
climate_dis = climate_dis[,1:5]
climate_dis[climate_dis == '...'] = 0
climate_dis[,3:5] = apply(climate_dis[,3:5],2,as.integer)
head(climate_dis)
```

CountryID	Countries.or.areas	Occurrence.1990.1999	Occurrence.2000.2009	Occurrence.2010.2019
4	Afghanistan	3	3	2
8	Albania	1	1	0
12	Algeria	2	1	0
24	Angola	4	2	2
660	Anguilla	1	0	0
28	Antigua and Barbuda	1	0	0

```r
geophys_dis = read.csv('Geophysical disasters.csv')
geophys_dis = geophys_dis[,1:5]
geophys_dis[geophys_dis == '...'] = 0
geophys_dis[,3:5] = apply(geophys_dis[,3:5],2,as.integer)
head(geophys_dis)
```

CountryID	Countries.or.areas	Occurrence.1990.1999	Occurrence.2000.2009	Occurrence.2010.2019
4	Afghanistan	10	11	5
8	Albania	1	1	3
12	Algeria	3	5	1
16	American Samoa	0	1	0
32	Argentina	1	2	2
51	Armenia	1	0	0

```r
hydro_dis = read.csv('Hydrological disasters.csv')
hydro_dis = hydro_dis[,1:5]
hydro_dis[hydro_dis == '...'] = 0
hydro_dis[,3:5] = apply(hydro_dis[,3:5],2,as.integer)
head(hydro_dis)
```

CountryID	Countries.or.areas	Occurrence.1990.1999	Occurrence.2000.2009	Occurrence.2010.2019
4	Afghanistan	17	45	46
8	Albania	4	4	7
12	Algeria	8	25	6
16	American Samoa	0	1	0
24	Angola	0	21	19
32	Argentina	10	20	20

```r
meteo_dis = read.csv('Meteorological disasters.csv')
meteo_dis = meteo_dis[,1:5]
meteo_dis[meteo_dis == '...'] = 0
meteo_dis[,3:5] = apply(meteo_dis[,3:5],2,as.integer)
meteo_dis[is.na(meteo_dis)] = 0
head(meteo_dis)
```

CountryID	Countries.or.areas	Occurrence.1990.1999	Occurrence.2000.2009	Occurrence.2010.2019
4	Afghanistan	3	7	5
8	Albania	0	4	2
12	Algeria	0	4	2
16	American Samoa	0	2	0
660	Anguilla	1	0	1
28	Antigua and Barbuda	5	0	2

```r
ghg_emissions = read.csv('GHG_Emissions.csv')
ghg_emissions_sectorwise = read.csv('GHG_Emissions_by_Sector.csv')
```

```
merged_ghg =  merge(ghg_emissions[2:nrow(ghg_emissions),c(1,2,33,34,35)],
                    ghg_emissions_sectorwise[,c(1,2,6,10,12,14)],
                    by = c("Country.ID","Country"))
merged_ghg[is.na(merged_ghg)] = 0
merged_ghg[merged_ghg == '...'] = 0
merged_ghg[,3:9] = apply(merged_ghg[,3:9],2,as.numeric)
merged_ghg = merged_ghg[order(merged_ghg$GHG.total.without.LULUCF..latest.year,
                              decreasing = T),]
head(merged_ghg[,c(1,2,4,5)])
```

	Country.ID	Country	X..change.since.1990	GHG.emissions.per.capita…latest.year
14	156	China	0.00	8.79
182	840	United States of America	3.72	20.41
135	643	Russian Federation	-30.35	15.23
58	356	India	0.00	1.70
68	392	Japan	-2.50	9.74
162	76	Brazil	86.83	5.02

```
climate_dis$CTotal_Occurences = apply(climate_dis[,3:5], 1, sum)
hydro_dis$HTotal_Occurences = apply(hydro_dis[,3:5], 1, sum)
geophys_dis$GTotal_Occurences = apply(geophys_dis[,3:5], 1, sum)
meteo_dis$MTotal_Occurences = apply(meteo_dis[,3:5], 1, sum)

climate_dis = climate_dis[order(climate_dis$CTotal_Occurences, decreasing = T),]
hydro_dis = hydro_dis[order(hydro_dis$HTotal_Occurences, decreasing = T),]
geophys_dis = geophys_dis[order(geophys_dis$GTotal_Occurences, decreasing = T),]
meteo_dis = meteo_dis[order(meteo_dis$MTotal_Occurences, decreasing = T),]
```

Some univariate and mulitvariate visualisation follows, after which a proper analysis using them is done.

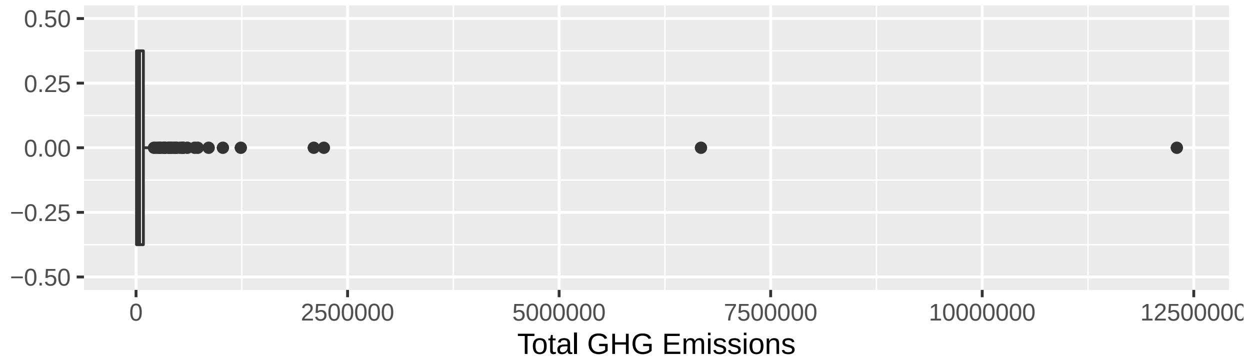
Univariate Analysis

Each variable is to be visualised and analysed individually to understand the structure and behaviour of the data. Bar plots and box plots are used for these visualisations, and careful use of colours is done to highlight important features of the variables rather than for mere beautification.

Greenhouse Gases Emissions

Since there are more than 190 countries and areas, most of the visualisations are limited to the top 10 contributors with respect to individual variables. Bar plots are used to compare that subset, while box plots provide insight into the distribution of the whole set.

```
options(repr.plot.width=14, repr.plot.height=4)
ggplot(merged_ghg) + geom_boxplot(aes(GHG.total.without.LULUCF..latest.year)) +
ylim(-0.5,0.5) +
xlab('Total GHG Emissions')
```

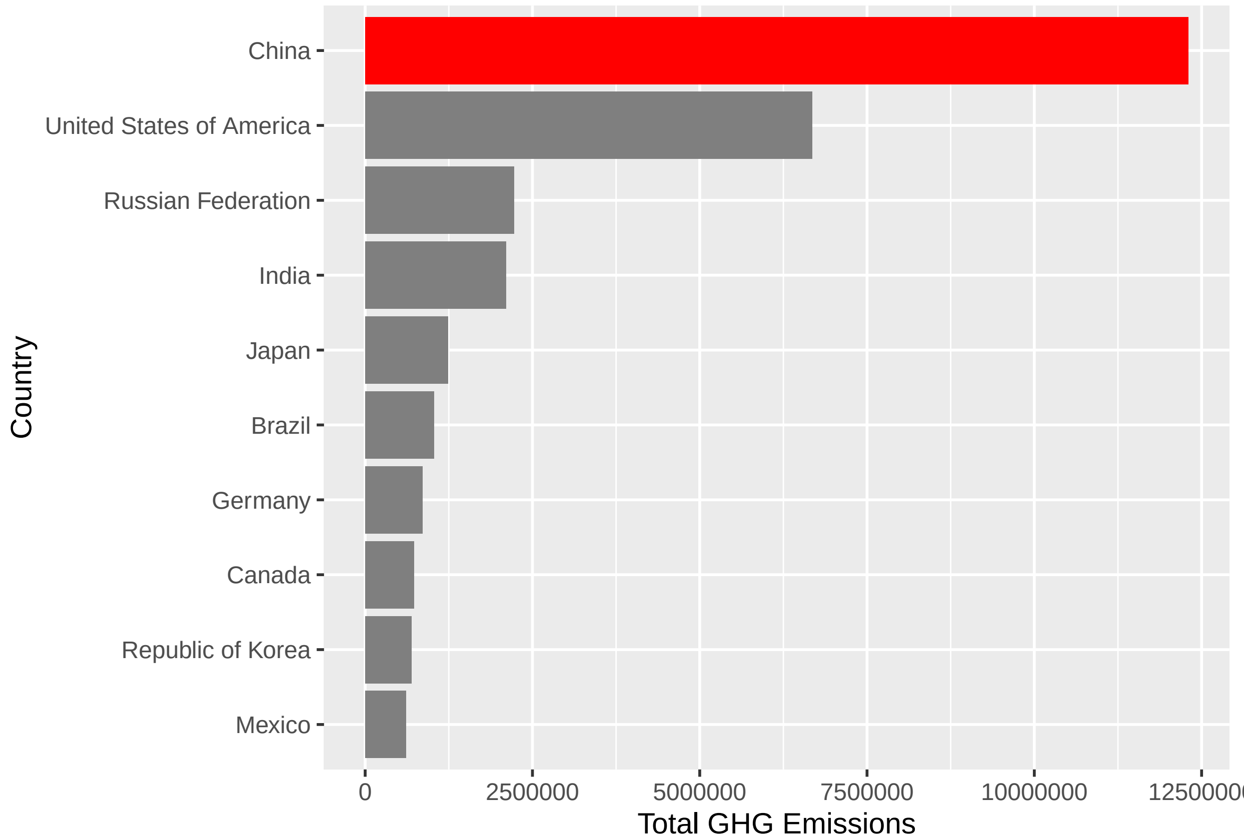


Amount of emissions of greenhouse gases has a lot of outliers and it is hard to analyse the variable statistics using a box plot. The 5-point summary can be directly checked.

```
summary(merged_ghg$GHG.total.without.LULUCF..latest.year)
```

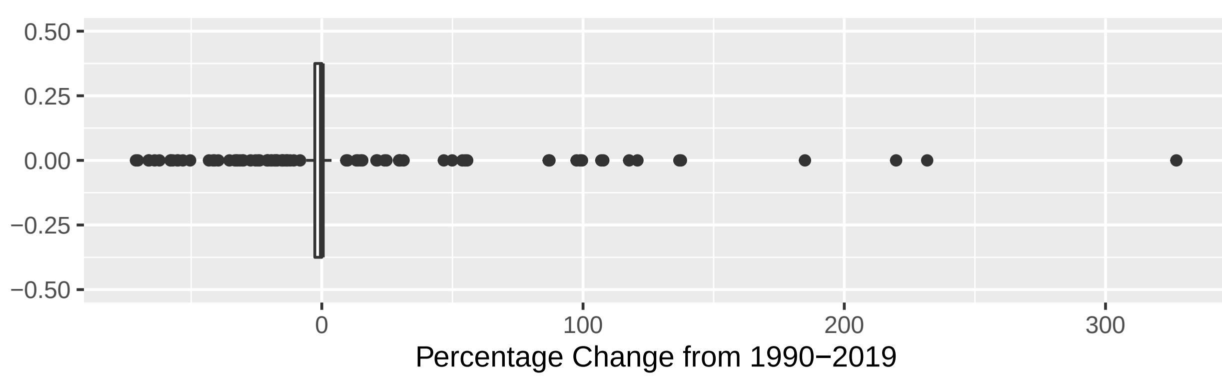
```
##     Min.  1st Qu.   Median     Mean  3rd Qu.     Max.
##       18     6421    25231   220084    86833 12300200
```

```
options(repr.plot.width=14, repr.plot.height=10)
ggplot(merged_ghg[1:10,],
       aes(x = reorder(Country, GHG.total.without.LULUCF..latest.year),
           y = GHG.total.without.LULUCF..latest.year,
           fill = factor(ifelse(Country == "China","Highlighted","Normal"))
           )
      ) +
geom_bar(stat = "identity") +
scale_fill_manual(name = "country", values = c("red","grey50")) +
coord_flip() + theme(legend.position="none") +
xlab('Country') + ylab('Total GHG Emissions')
```



China's emissions far surpasses emissions from any other country. Highly developed nations and fast growing nations are in the top 10 contributors.

```
options(repr.plot.width=14, repr.plot.height=4)
ggplot(merged_ghg) + geom_boxplot(aes(X..change.since.1990)) +
ylim(-0.5,0.5) + xlab('Percentage Change from 1990-2019')
```

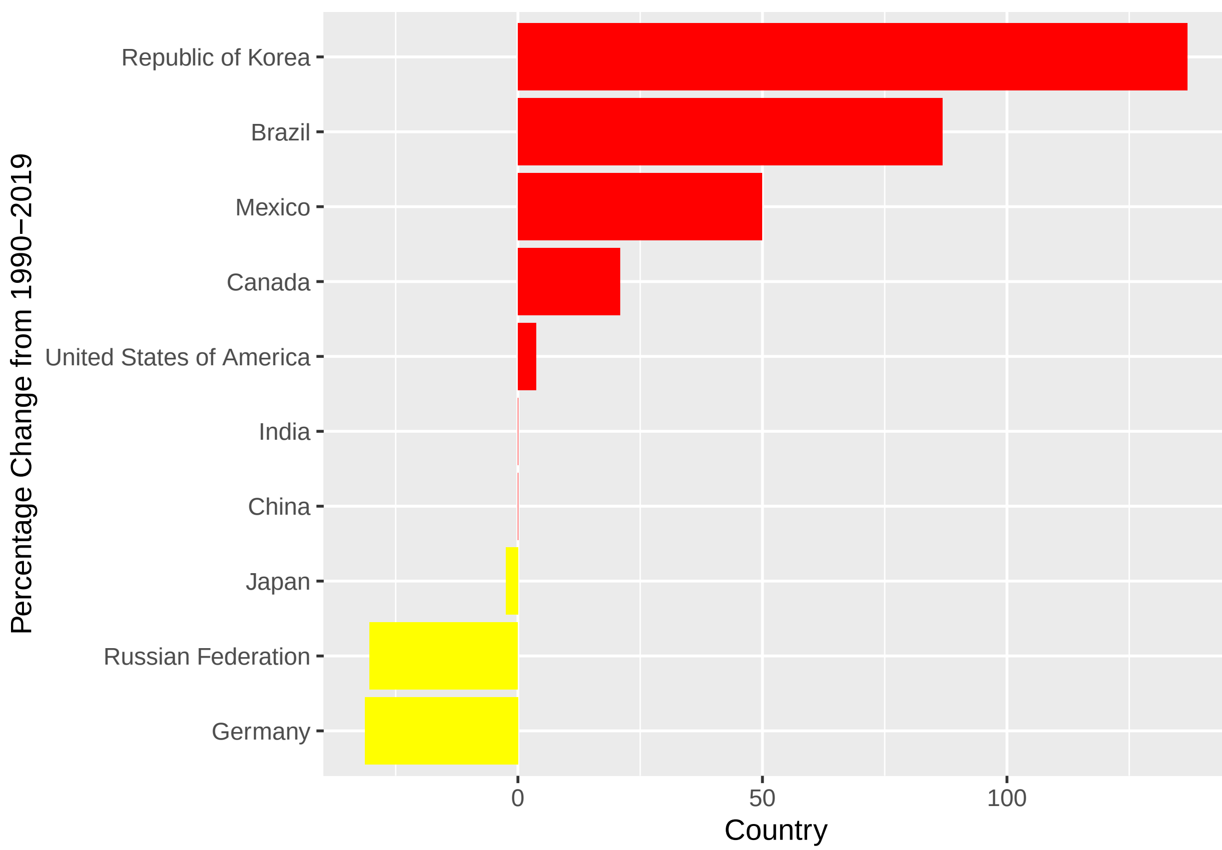


Again, the data has a lot of outliers and it is hard to analyse the variable statistics using a box plot. The 5-point summary can be directly checked.

```
summary(merged_ghg$X..change.since.1990)
```

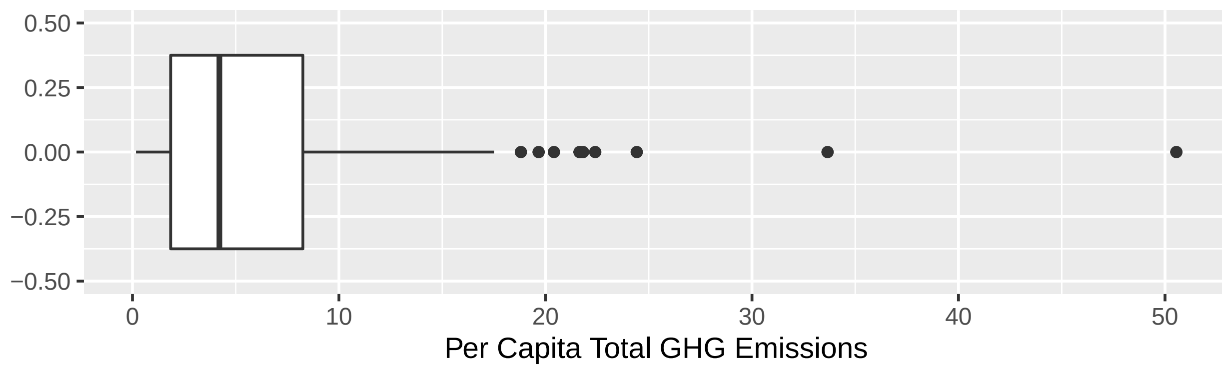
```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## -71.140  -2.670   0.000   5.955   0.000 327.120
```

```
inc_or_dec = function(column){
    colour = rep(NA,length(column))
    for(i in 1:length(column)){
        if(column[i]<0) colour[i] = 'Decreased'
        else colour[i] = 'Increased'
    }
    return(colour)
}
options(repr.plot.width=16, repr.plot.height=8)
ggplot(merged_ghg[1:10,],
       aes(x = reorder(Country, X..change.since.1990),
           y = X..change.since.1990,
           fill = inc_or_dec(merged_ghg$X..change.since.1990[1:10]))) +
geom_bar(stat = "identity") + coord_flip() +
scale_fill_manual(values=c("yellow","red")) +
theme(legend.position="none") +
xlab('Percentage Change from 1990-2019') + ylab('Country')
```



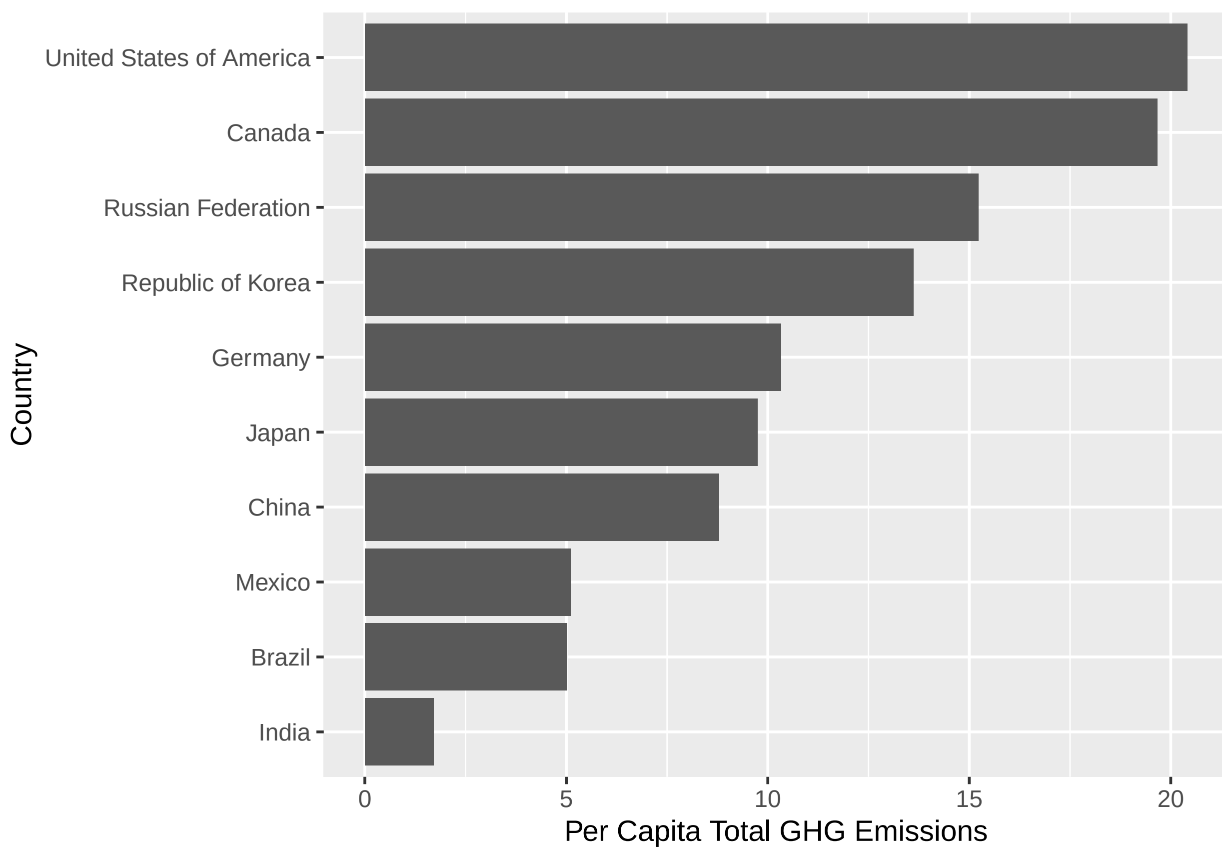
Developing nations seem to have a positive percentage changes, inferring that their emissions have greatly increased over the decades. Some developed countries have managed to get a negative percentage changes.

```
options(repr.plot.width=14, repr.plot.height=4)
ggplot(merged_ghg) + geom_boxplot(aes(GHG.emissions.per.capita...latest.year)) +
ylim(-0.5,0.5) + xlab('Per Capita Total GHG Emissions')
```



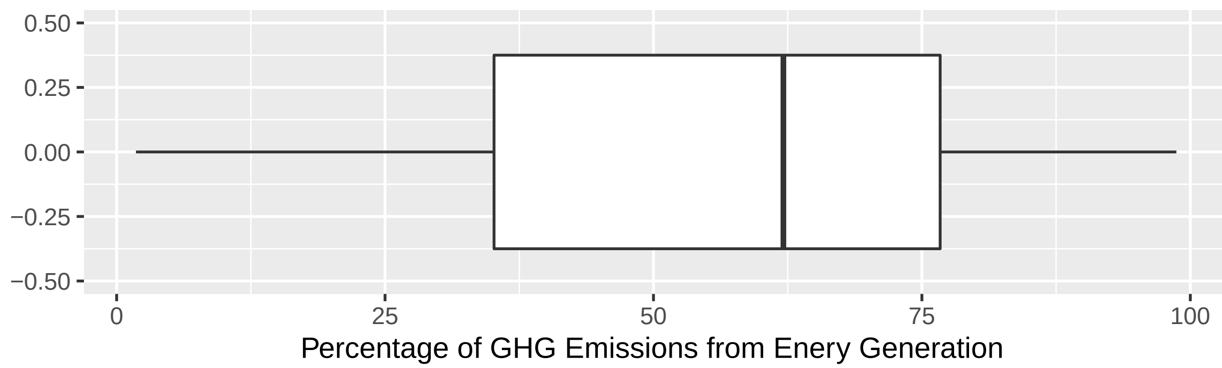
Per capita emissions seems to have a better distribution, indicating that total emissions may be a biased variable because the difference in population of countries is very large.

```
options(repr.plot.width=14, repr.plot.height=10)
ggplot(merged_ghg[1:10,],
       aes(x = reorder(Country, GHG.emissions.per.capita...latest.year),
                                   y = GHG.emissions.per.capita...latest.year)) +
geom_bar(stat = "identity") + coord_flip()+
  ylab('Per Capita Total GHG Emissions') + xlab('Country')
```



It is now clearly apparent that highly developed nations have the highest per capita emissions, while developing countries have lower per capita emissions.

```
options(repr.plot.width=14, repr.plot.height=4)
ggplot(merged_ghg) + geom_boxplot(
  aes(GHG.from.energy..as.percentage.to.total)) +
ylim(-0.5,0.5) + xlab('Percentage of GHG Emissions from Enery Generation ')
```

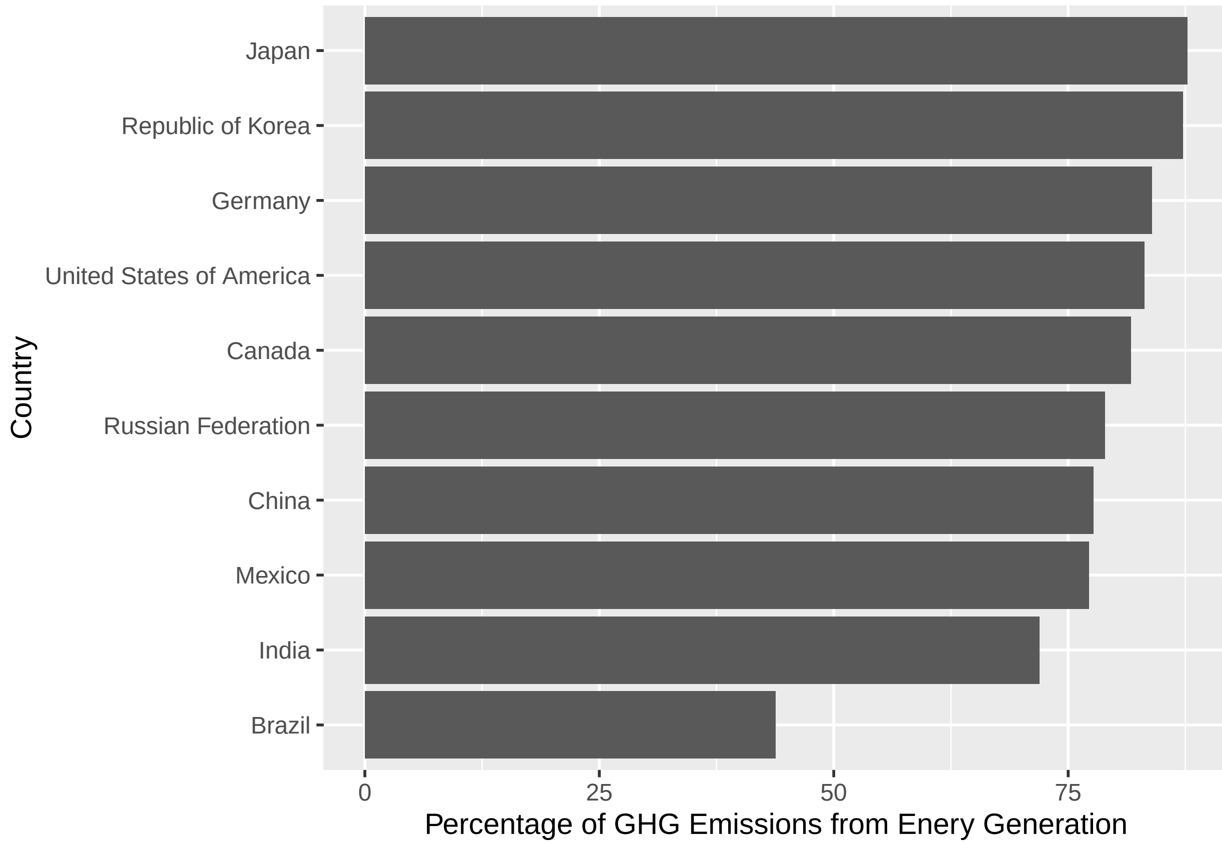



```
summary(merged_ghg$GHG.from.energy..as.percentage.to.total)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max. 
##    1.80   35.15   62.10   56.45   76.70   98.70
```

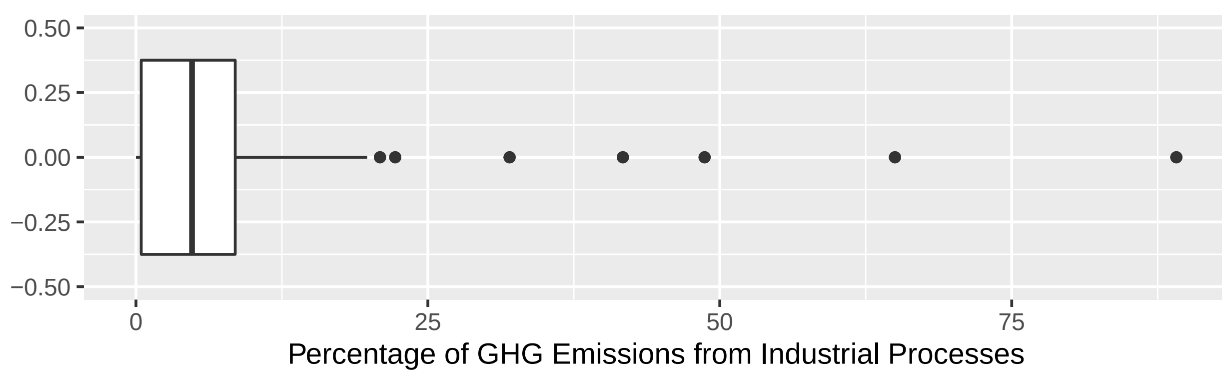
It is to be noted that on an average, more than half of the emissions are due to energy generation processes.

```
options(repr.plot.width=14, repr.plot.height=10)
ggplot(merged_ghg[1:10,],
       aes(x = reorder(Country, GHG.from.energy..as.percentage.to.total),
           y = GHG.from.energy..as.percentage.to.total)) +
geom_bar(stat = "identity") + coord_flip() +
ylab('Percentage of GHG Emissions from Enery Generation') +
xlab('Country')
```



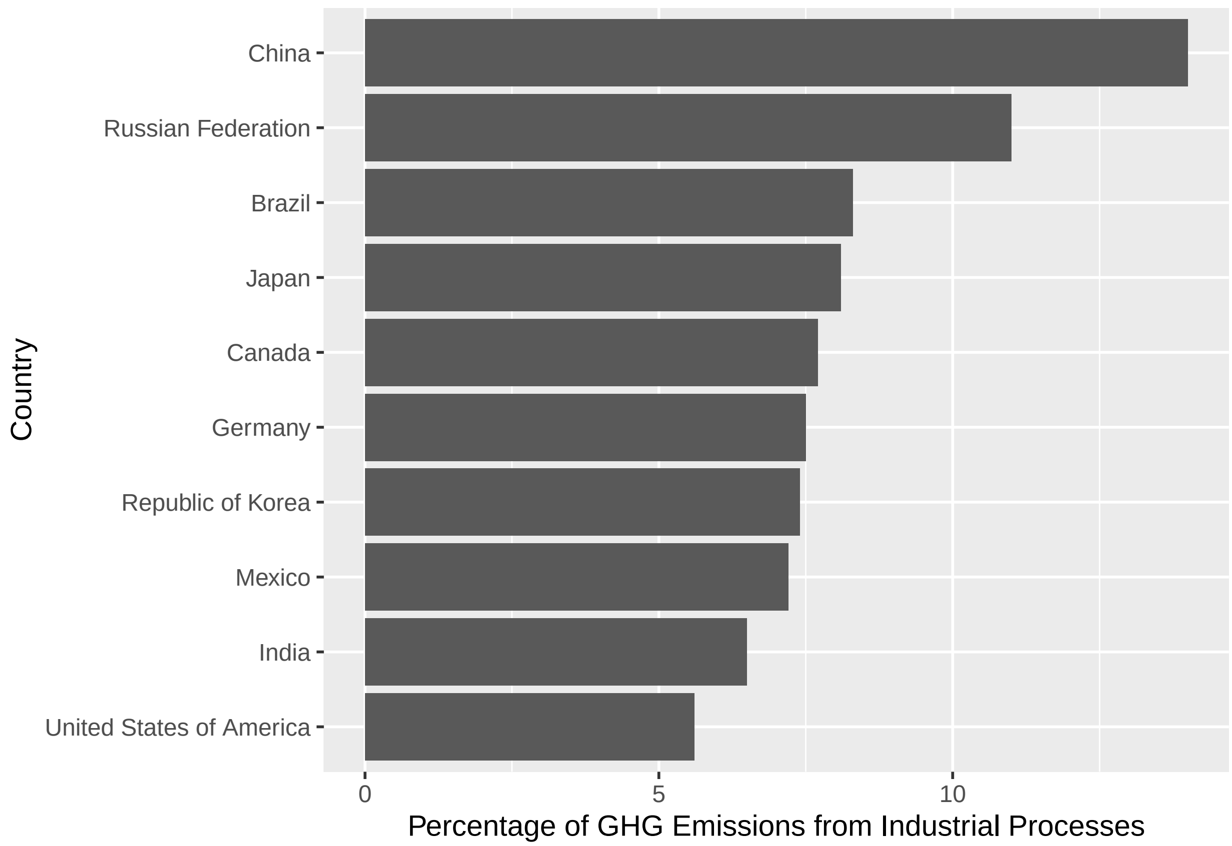
Highly developed countries and fast growing developing countries have more the 75% of their emissions due to energy generation processes.

```
options(repr.plot.width=14, repr.plot.height=4)
ggplot(merged_ghg) +
geom_boxplot(aes(
  GHG.from.industrial.processes.and.product.use..as.percentage.to.total)) +
ylim(-0.5,0.5) +
xlab('Percentage of GHG Emissions from Industrial Processes')
```

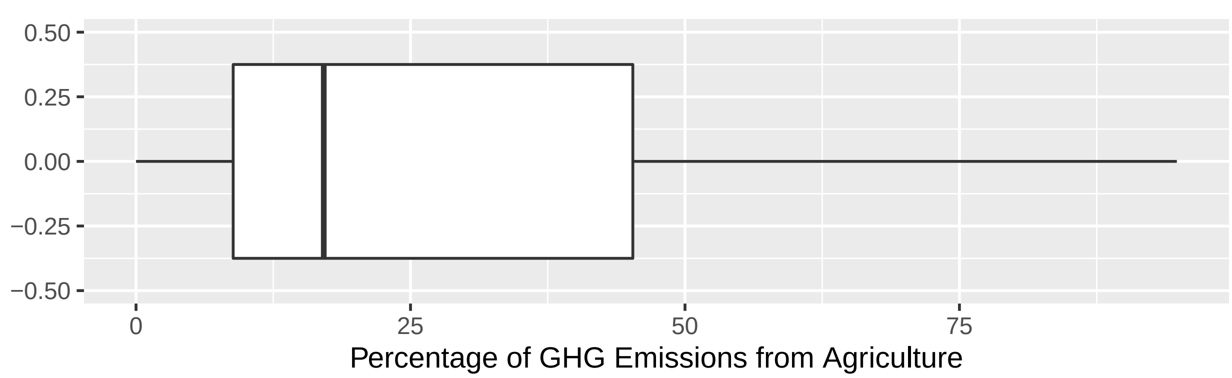



```
options(repr.plot.width=14, repr.plot.height=10)
ggplot(merged_ghg[1:10,],
  aes(x = reorder(Country,
```

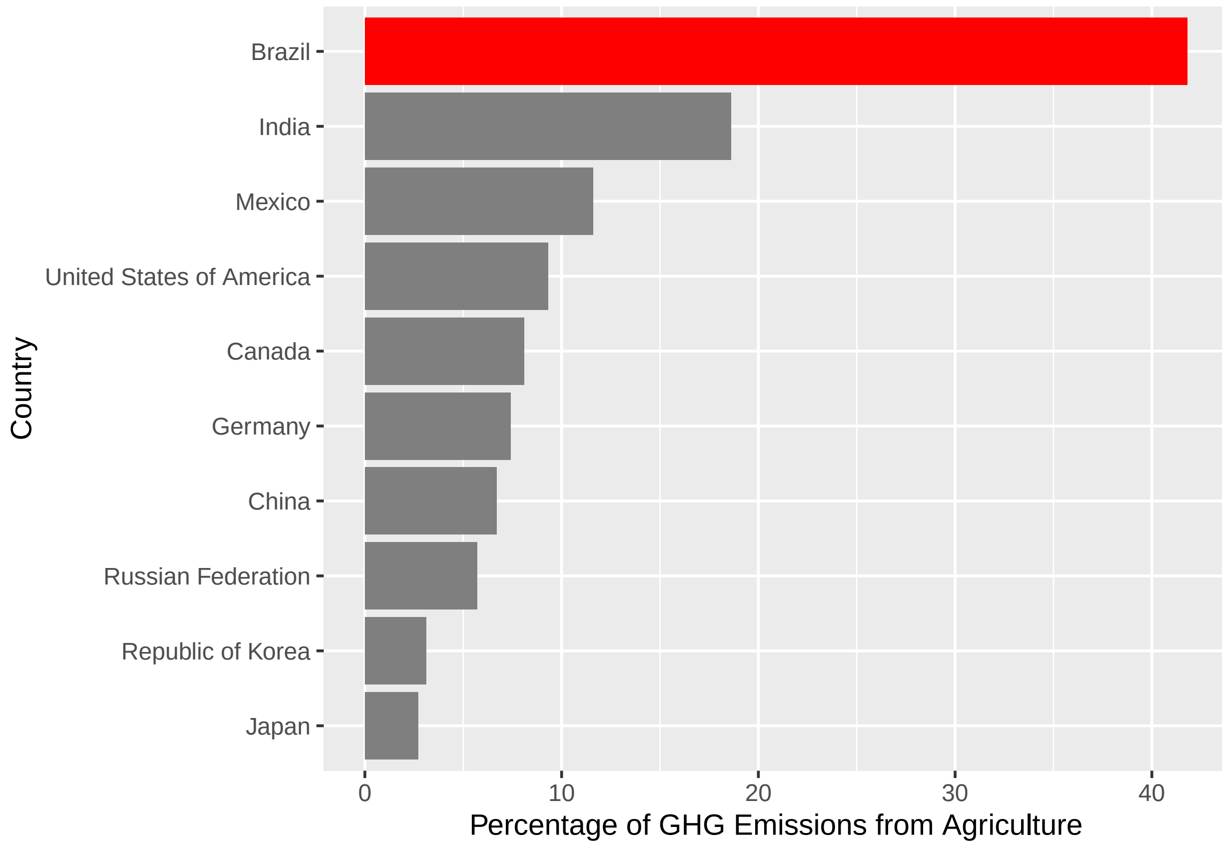
```
  GHG.from.industrial.processes.and.product.use..as.percentage.to.total),
  y = GHG.from.industrial.processes.and.product.use..as.percentage.to.total)) +
geom_bar(stat = "identity") + coord_flip() +
  xlab('Country') +
  ylab('Percentage of GHG Emissions from Industrial Processes')
```




```
options(repr.plot.width=14, repr.plot.height=4)
ggplot(merged_ghg) +
geom_boxplot(aes(GHG.from.agriculture..as.percentage.to.total)) +
ylim(-0.5,0.5) +
xlab('Percentage of GHG Emissions from Agriculture')
```

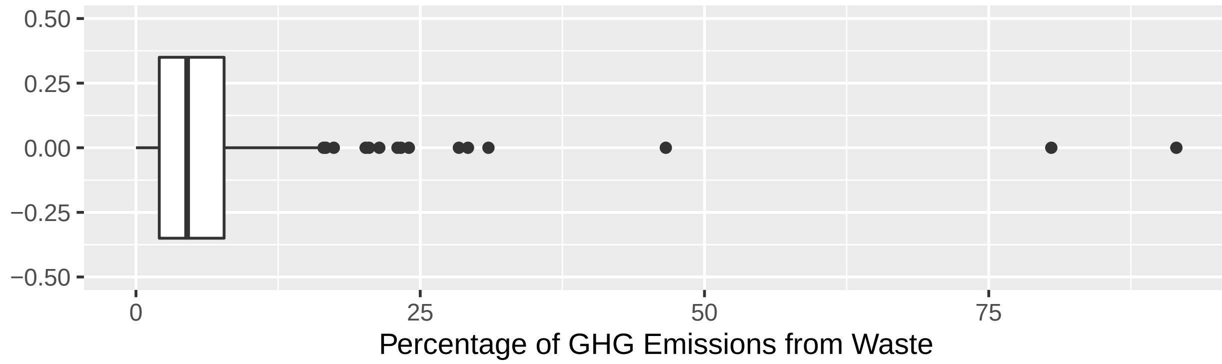


```r
options(repr.plot.width=14, repr.plot.height=10)
ggplot(merged_ghg[1:10,],
       aes(x = reorder(Country, GHG.from.agriculture..as.percentage.to.total),
           y = GHG.from.agriculture..as.percentage.to.total,
           fill = factor(ifelse(Country == "Brazil","Highlighted","Normal")))) +
geom_bar(stat = "identity") + coord_flip() +
scale_fill_manual(name = "country", values = c("red","grey50")) +
theme(legend.position="none") + xlab('Country') +
ylab('Percentage of GHG Emissions from Agriculture')
```

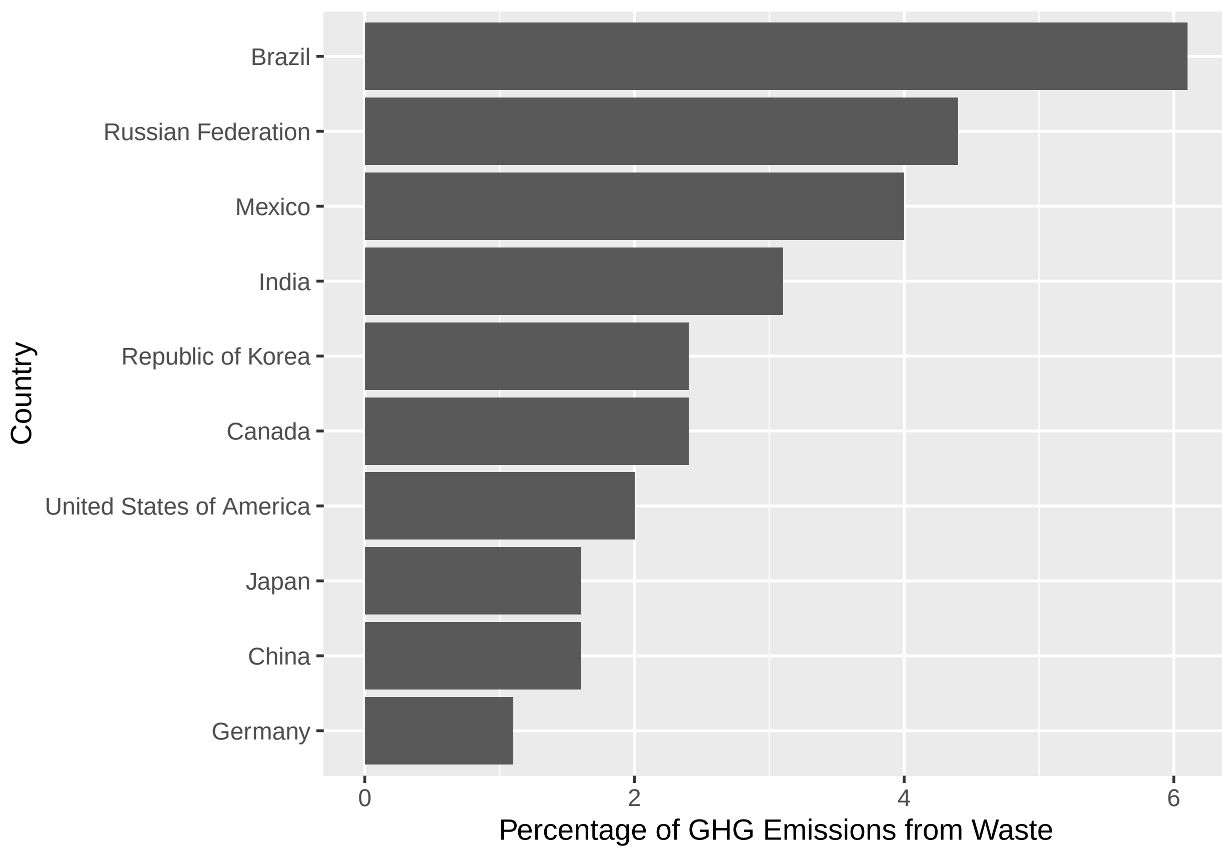


As expected, agricultural nations like Brazil and India have a high percentage of emissions due to agricultural processes.

```r
options(repr.plot.width=14, repr.plot.height=4)
ggplot(merged_ghg) +
geom_boxplot(aes(GHG.from.waste..as.percentage.to.total), width = 0.7) +
ylim(-0.5,0.5) + xlab('Percentage of GHG Emissions from Waste')
```



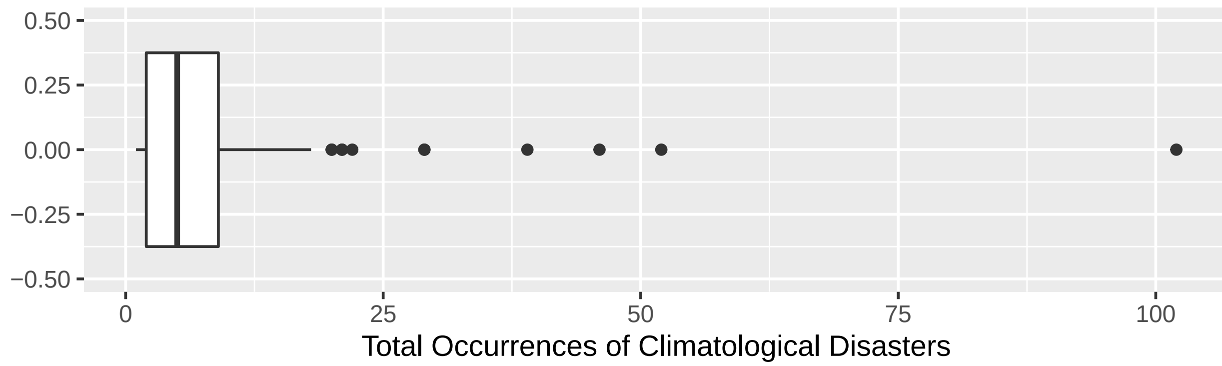

```r
options(repr.plot.width=14, repr.plot.height=10)
ggplot(merged_ghg[1:10,],
       aes(x = reorder(Country, GHG.from.waste..as.percentage.to.total),
           y = GHG.from.waste..as.percentage.to.total)) +
geom_bar(stat = "identity") + coord_flip() + xlab('Country')+
  ylab('Percentage of GHG Emissions from Waste')
```



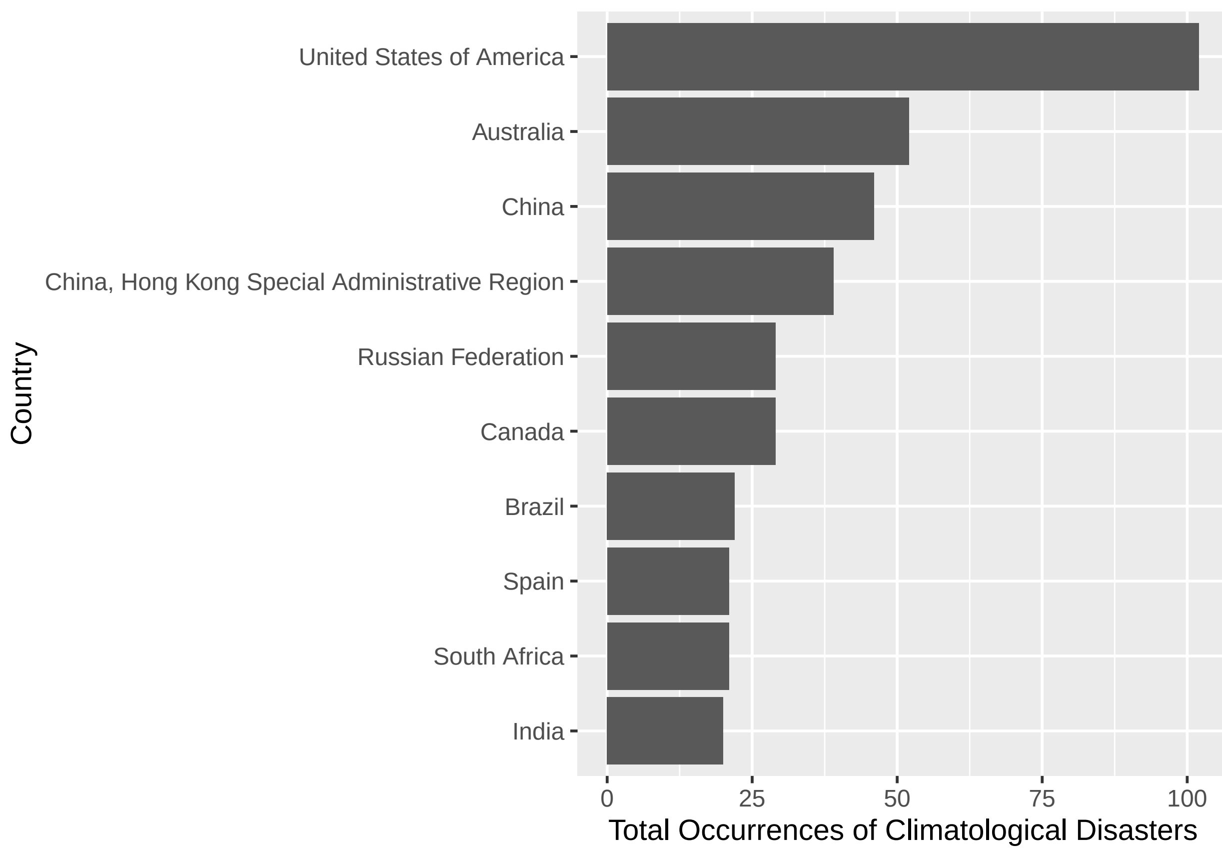
Natural Disasters

Again, most of the visualisations are limited to the top 10 disaster struck countries and areas with respect to individual variables. Note again that natural disasters are classified into four categories: Climatological, Geophysical, Hydrological and Meteorological disasters.

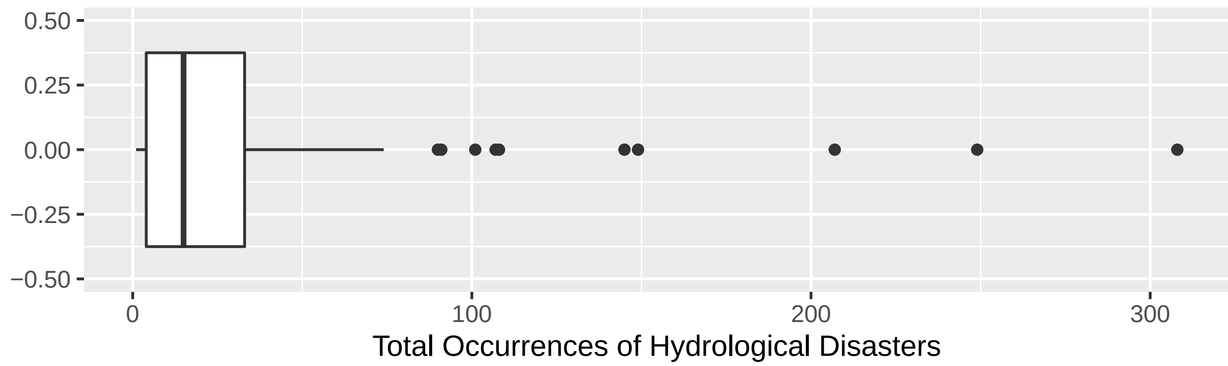
```r
options(repr.plot.width=14, repr.plot.height=4)
ggplot(climate_dis) +
geom_boxplot(aes(CTotal_Occurences)) +
ylim(-0.5,0.5) +
xlab('Total Occurrences of Climatological Disasters')
```



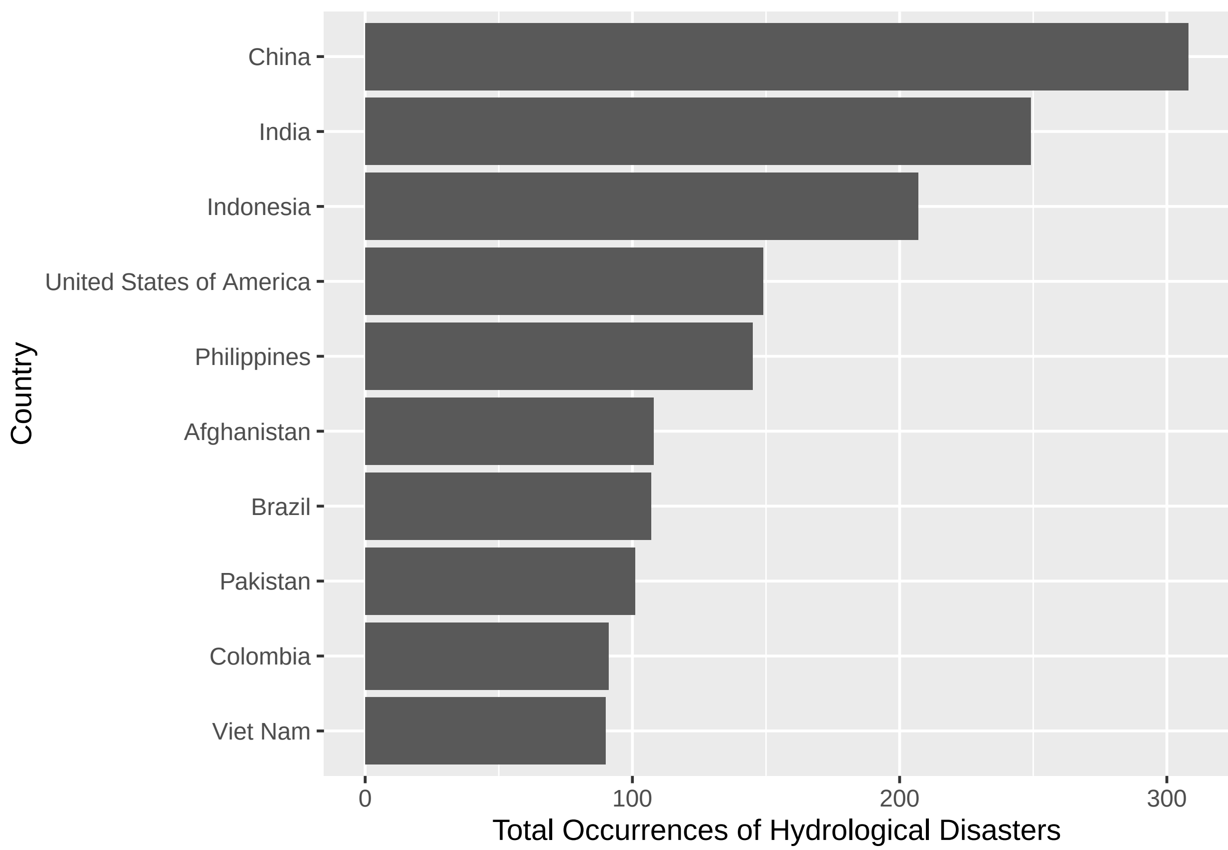

```r
options(repr.plot.width=14, repr.plot.height=10)
ggplot(climate_dis[1:10,], aes(x = reorder(Countries.or.areas, CTotal_Occurences),
                               y = CTotal_Occurences)) +
geom_bar(stat = "identity") + coord_flip() +
  ylab('Total Occurrences of Climatological Disasters') +
  xlab('Country')
```



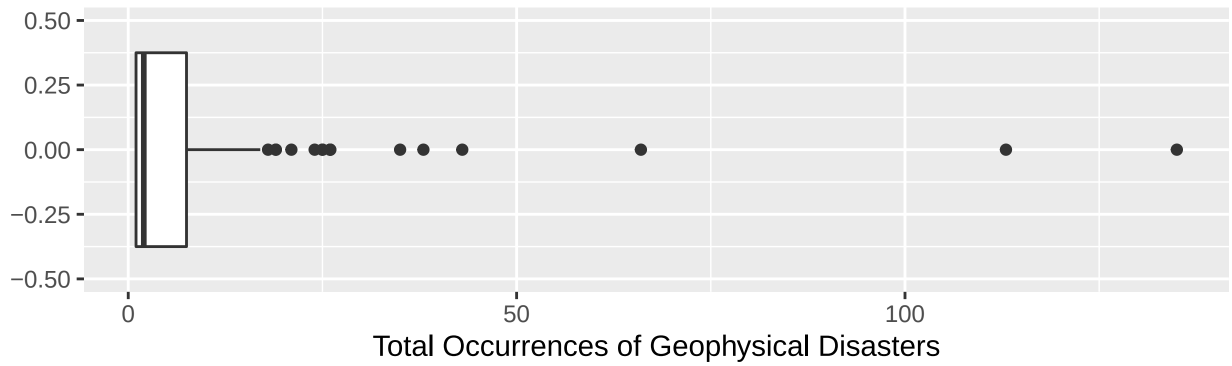
```
options(repr.plot.width=14, repr.plot.height=4)
ggplot(hydro_dis) +
geom_boxplot(aes(HTotal_Occurences)) +
ylim(-0.5,0.5) +
xlab('Total Occurrences of Hydrological Disasters')
```



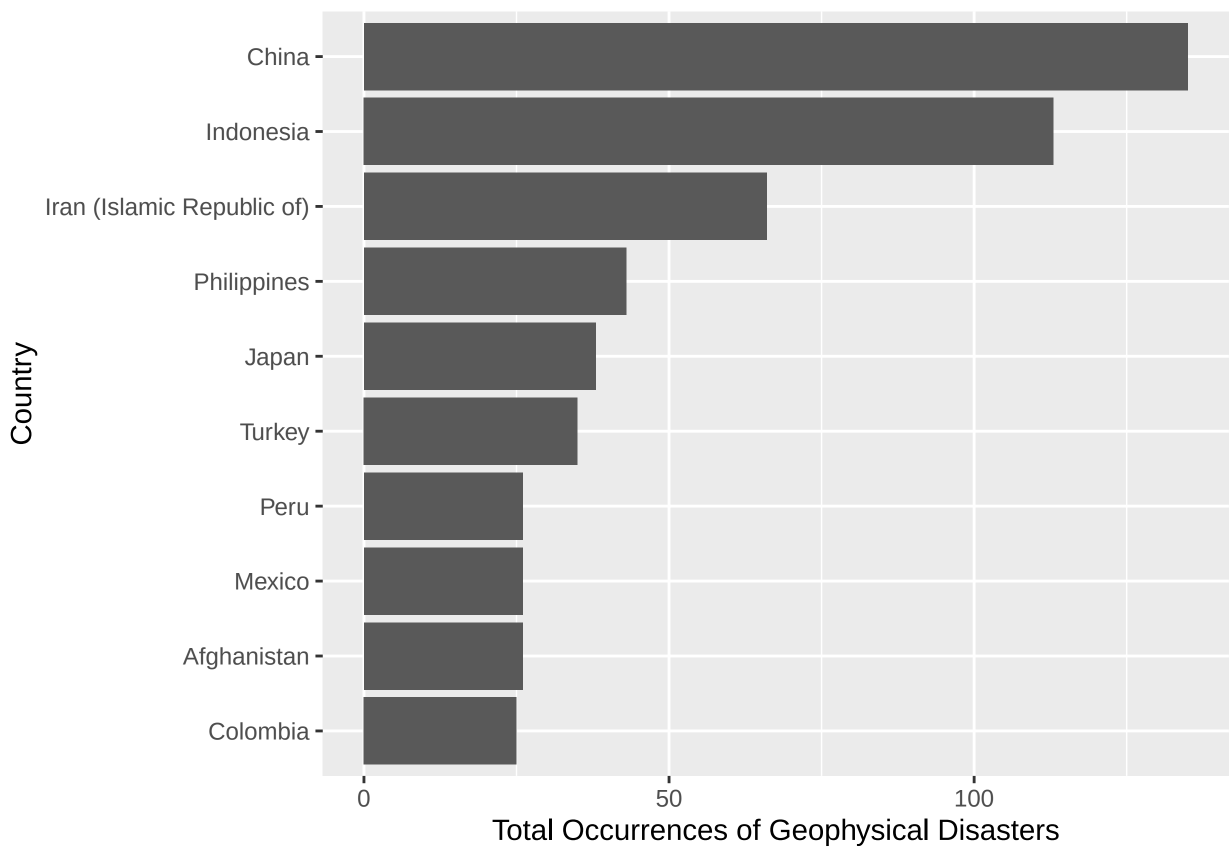

```
options(repr.plot.width=14, repr.plot.height=10)
ggplot(hydro_dis[1:10,], aes(x = reorder(Countries.or.areas, HTotal_Occurences),
                              y = HTotal_Occurences)) +
geom_bar(stat = "identity") + coord_flip()+
  ylab('Total Occurrences of Hydrological Disasters') +
  xlab('Country')
```



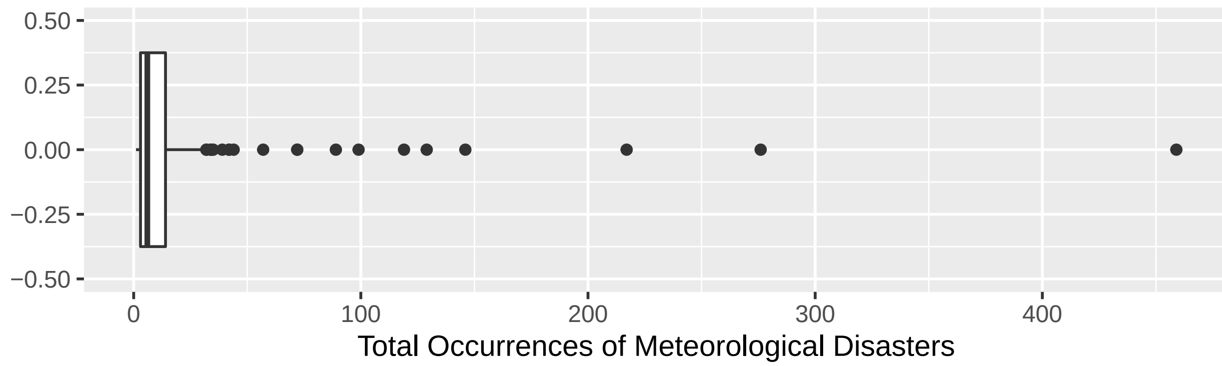
```
options(repr.plot.width=14, repr.plot.height=4)
ggplot(geophys_dis) +
geom_boxplot(aes(GTotal_Occurences)) +
ylim(-0.5,0.5) +
xlab('Total Occurrences of Geophysical Disasters')
```



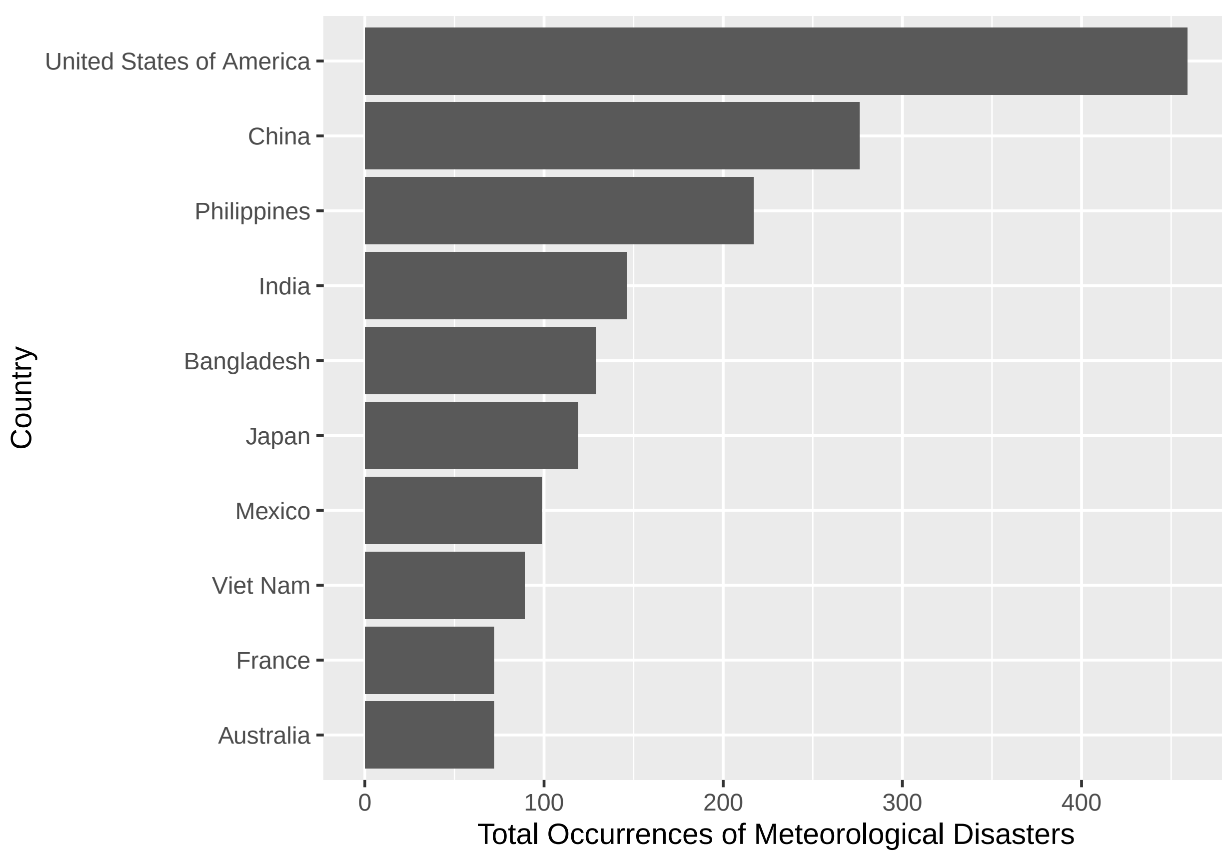

```
options(repr.plot.width=14, repr.plot.height=10)
ggplot(geophys_dis[1:10,], aes(x = reorder(Countries.or.areas, GTotal_Occurences),
                               y = GTotal_Occurences)) +
geom_bar(stat = "identity") + coord_flip()+
  ylab('Total Occurrences of Geophysical Disasters') +
  xlab('Country')
```



```
options(repr.plot.width=14, repr.plot.height=4)
ggplot(meteo_dis) +
geom_boxplot(aes(MTotal_Occurences)) +
ylim(-0.5,0.5) +
xlab('Total Occurrences of Meteorological Disasters')
```




```
options(repr.plot.width=14, repr.plot.height=10)
ggplot(meteo_dis[1:10,], aes(x = reorder(Countries.or.areas, MTotal_Occurences),
                                y = MTotal_Occurences)) +
geom_bar(stat = "identity") + coord_flip()+
  ylab('Total Occurrences of Meteorological Disasters') +
  xlab('Country')
```

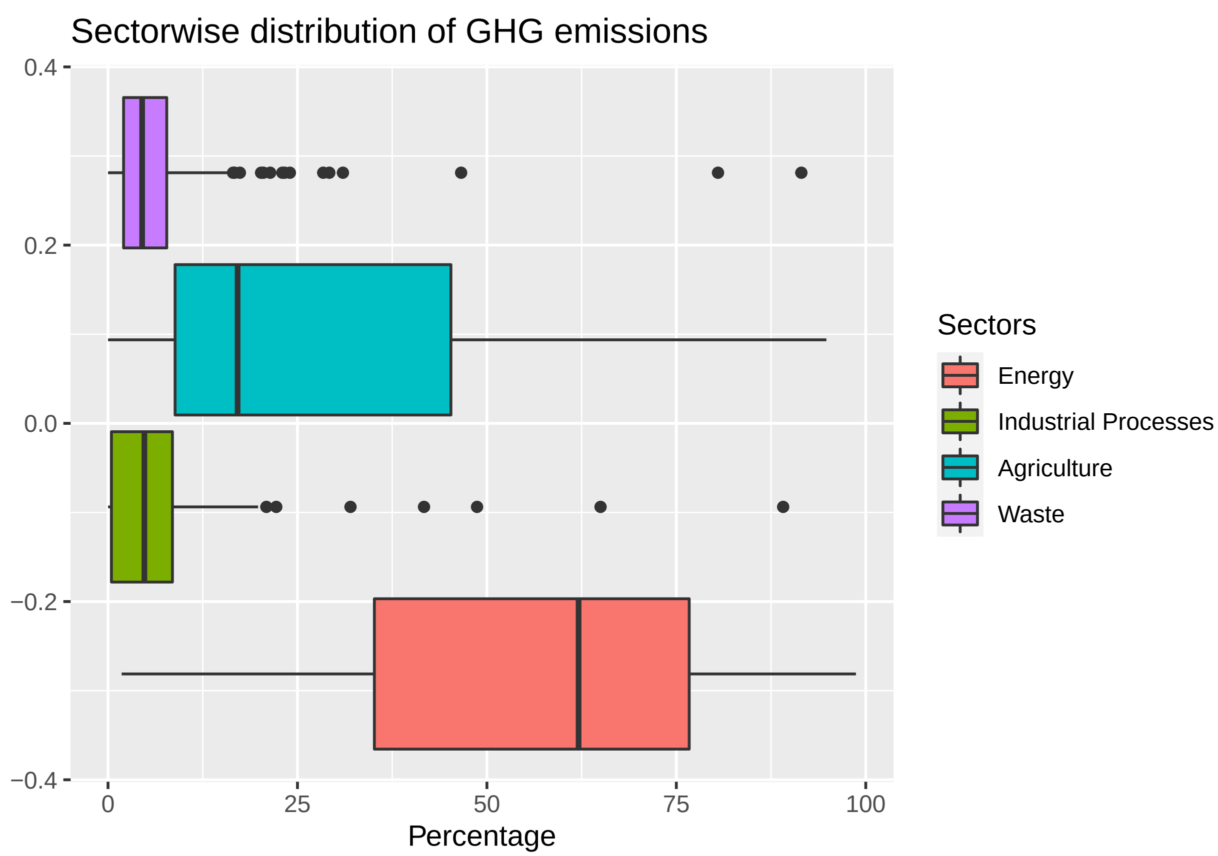


Total occurrences of natural disasters are fairly few among most countries, with few outliers, which include USA and some South Asian and South-East Asian countries.

Multivariate Analysis

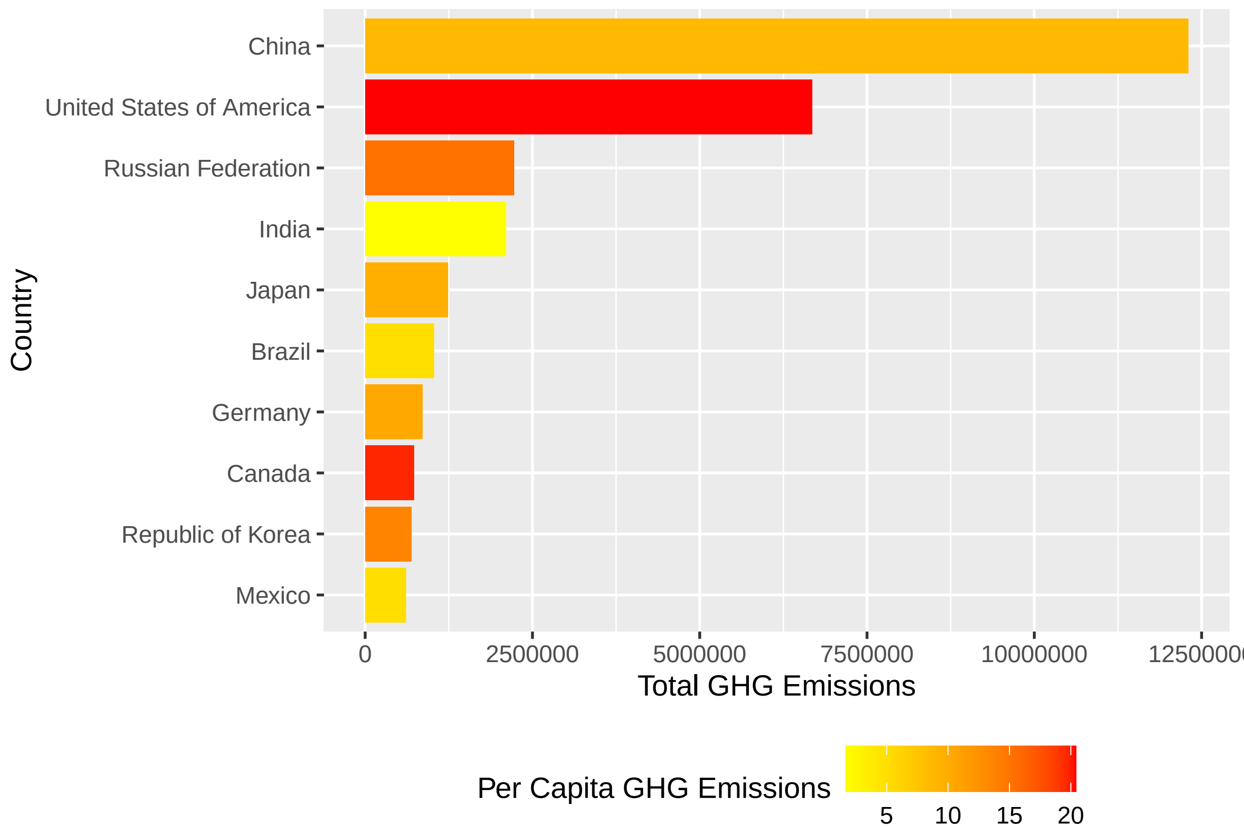
Multivariate visualisation must be carried out to understand the relationships, or independence, or different variables. Towards the purpose of trying to analyse the effect of emissions on the number of occurrences of natural disasters, bar plots and pie plots are used.

```
options(repr.plot.width=14, repr.plot.height=8)
long_ghg_percent = melt(merged_ghg[,c(2,6:9)], id.vars = 'Country')
ggplot(long_ghg_percent) + geom_boxplot(aes(value, fill = variable)) +
xlab('Percentage') +
labs(title = "Sectorwise distribution of GHG emissions") +
scale_fill_discrete(name="Sectors",
breaks=c("GHG.from.energy..as.percentage.to.total",
         "GHG.from.industrial.processes.and.product.use..as.percentage.to.total",
         "GHG.from.agriculture..as.percentage.to.total",
         "GHG.from.waste..as.percentage.to.total"),
labels=c("Energy", "Industrial Processes", "Agriculture", "Waste"))
```



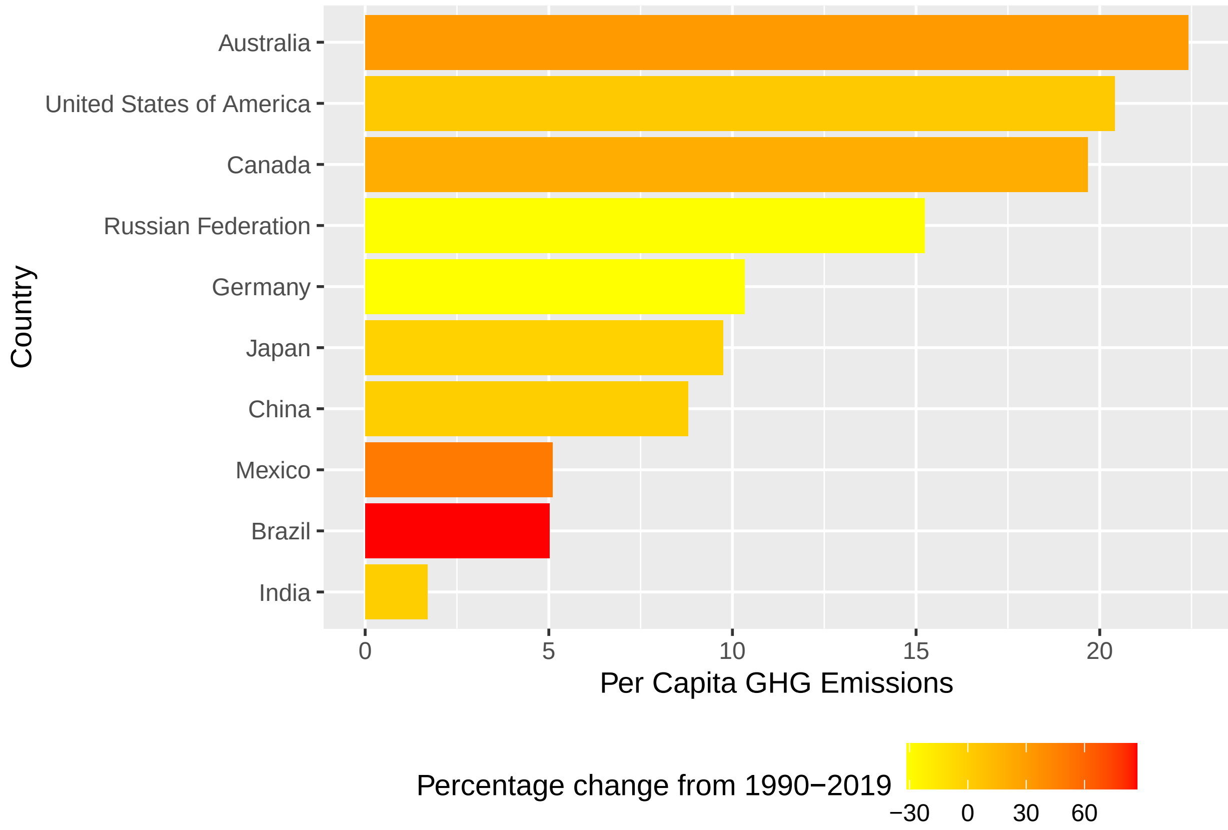
It can be noted that Energy generation is the biggest contributor across all countries, Agriculture seems to be the second, and Industrial and Waste follows next.

```r
options(repr.plot.width=20, repr.plot.height=10)
ggplot(merged_ghg[1:10,],
       aes(x = reorder(Country, GHG.total.without.LULUCF..latest.year),
           y = GHG.total.without.LULUCF..latest.year)) +
geom_bar(stat = "identity",
         aes(fill = GHG.emissions.per.capita...latest.year)) +
coord_flip() + scale_fill_gradient(low = 'yellow', high = 'red') +
  ylab('Total GHG Emissions') + xlab('Country') +
  labs(fill = "Per Capita GHG Emissions") +
  theme(legend.position="bottom")
```

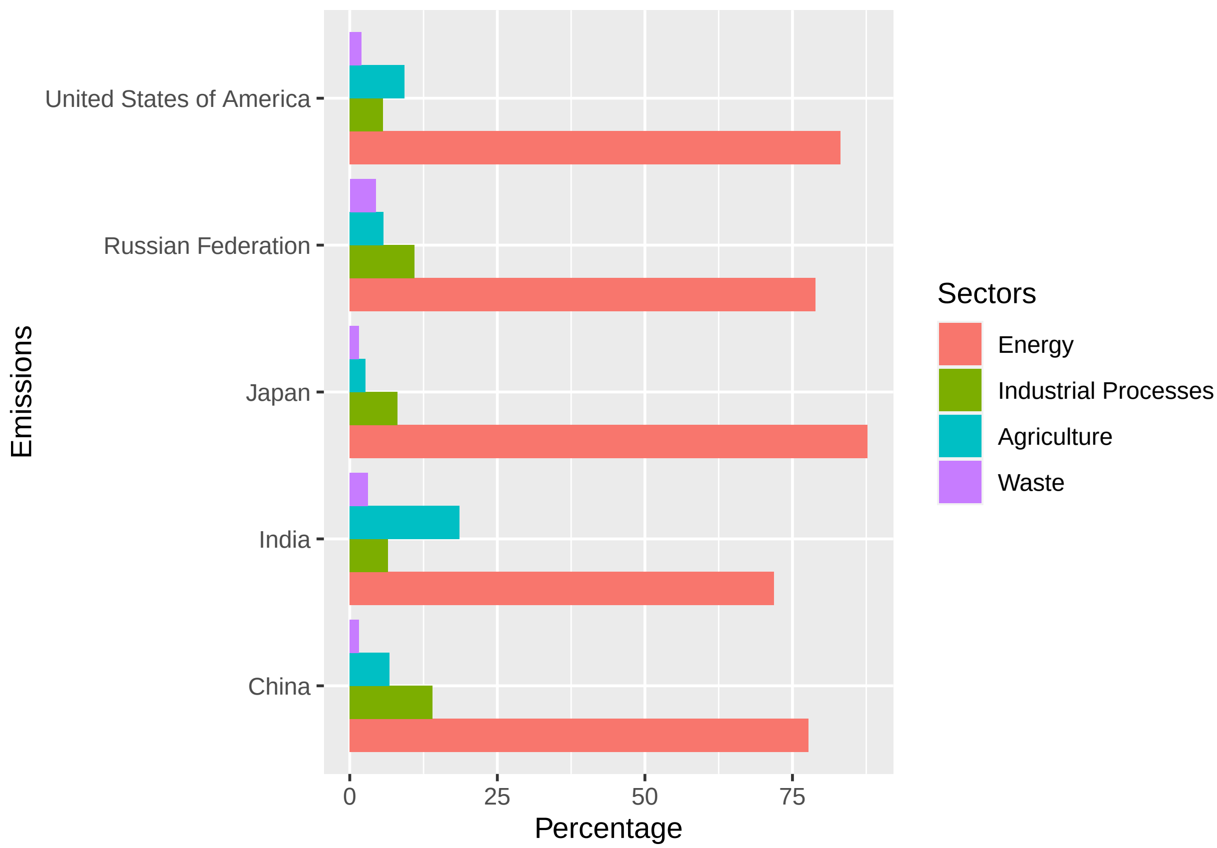


Adding to previous analysis, it can be noted here that developing countries have higher per capita emissions, while developing nations have lower capita emissions. China is an outlier since because of its large population, its total emissions are large too.

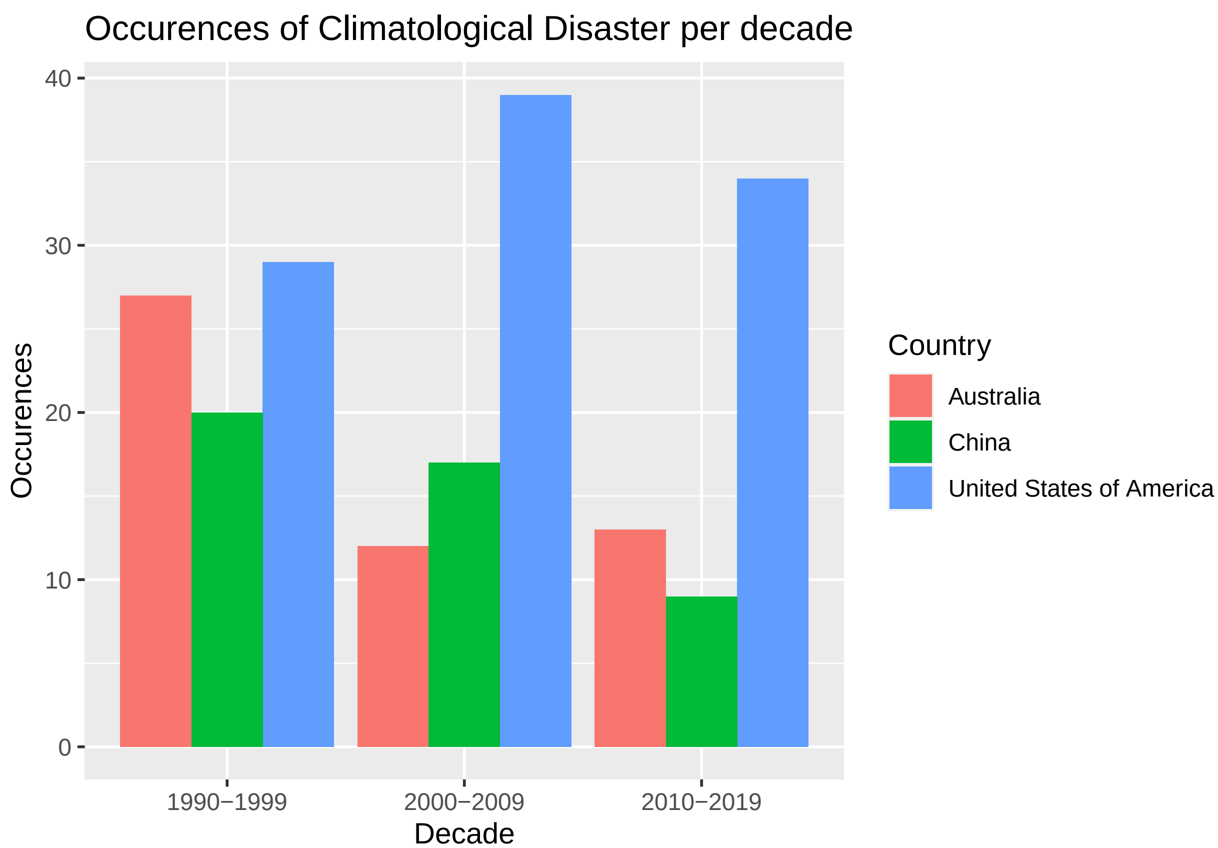
```r
options(repr.plot.width=20, repr.plot.height=10)
ggplot(merged_ghg[c(1:8,10,11),],
       aes(x = reorder(Country, GHG.emissions.per.capita...latest.year),
          y = GHG.emissions.per.capita...latest.year)) +
geom_bar(stat = "identity", aes(fill = X..change.since.1990)) +
coord_flip() + scale_fill_gradient(low = 'yellow', high = 'red')+
  ylab('Per Capita GHG Emissions') + xlab('Country') +
  labs(fill = "Percentage change from 1990-2019") +
  theme(legend.position="bottom")
```



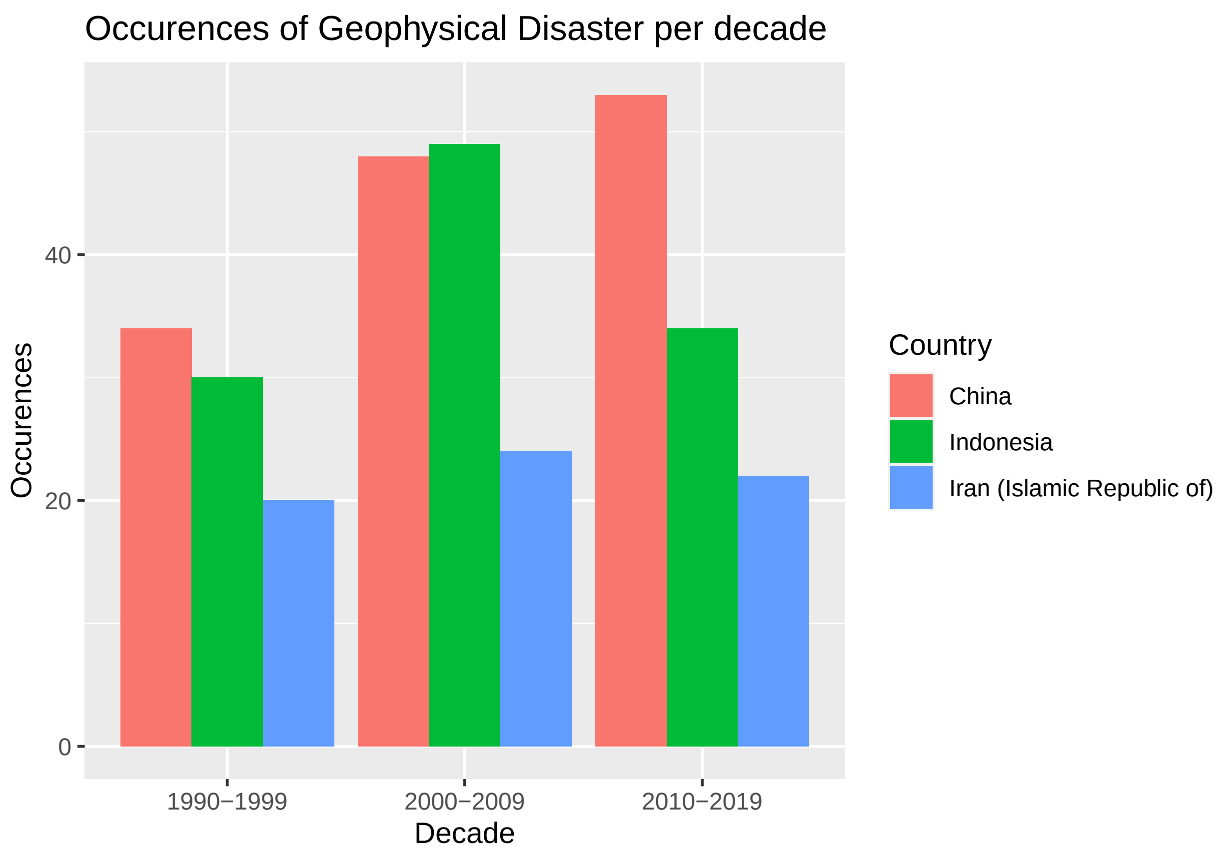

```
long_ghg = melt(merged_ghg[1:5,c(2,6:9)], id.vars = 'Country')

options(repr.plot.width=20, repr.plot.height=8)
ggplot(long_ghg, aes(x = Country, y = value, fill = variable)) +
geom_bar(stat = 'identity', position ='dodge')+
  coord_flip()+
  scale_fill_discrete(name="Sectors",
breaks=c("GHG.from.energy..as.percentage.to.total",
         "GHG.from.industrial.processes.and.product.use..as.percentage.to.total",
         "GHG.from.agriculture..as.percentage.to.total",
         "GHG.from.waste..as.percentage.to.total"),
labels=c("Energy", "Industrial Processes", "Agriculture", "Waste"))+
  xlab("Emissions") + ylab("Percentage") +
  labs(titel = "Sectorwise Percentage of Emissions
       of top 3 emitters")
```



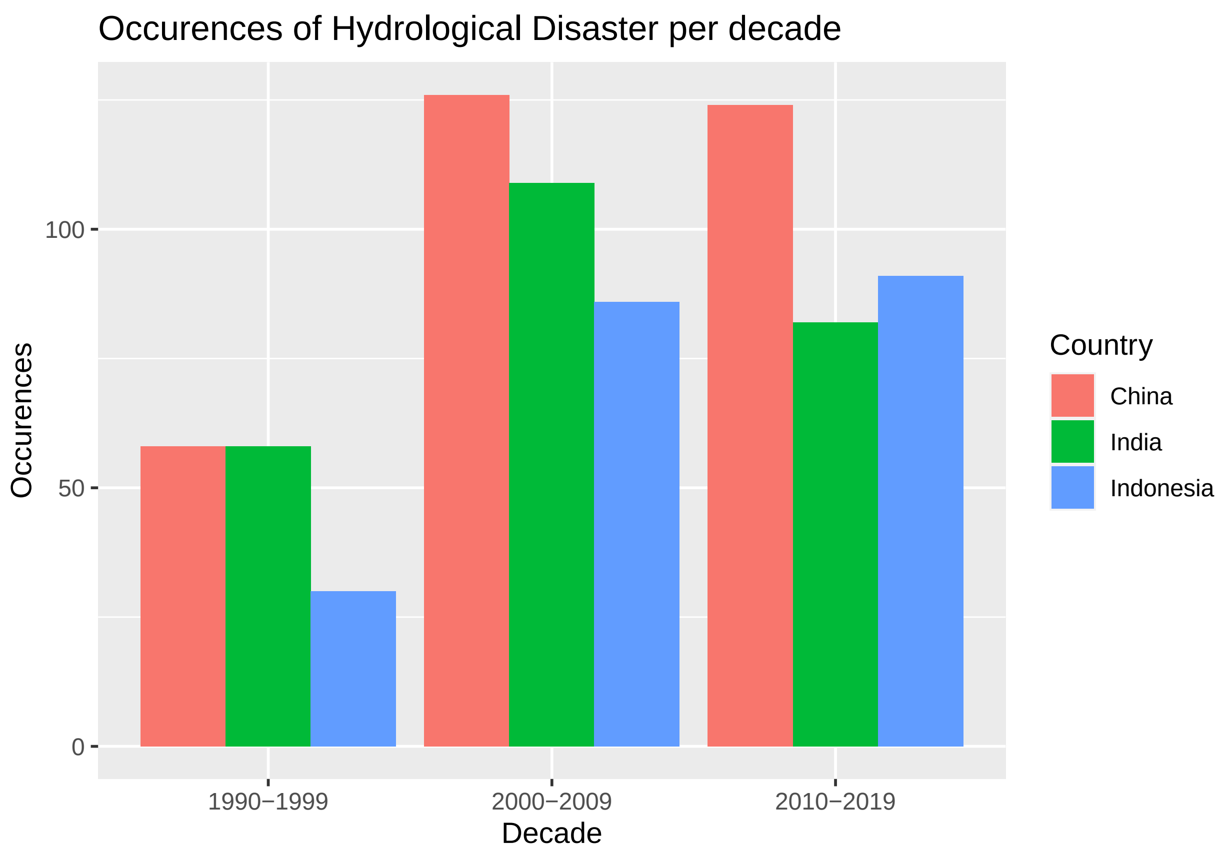

```
long_climate = melt(climate_dis[1:3,2:5], id.vars = 'Countries.or.areas')
long_geophys = melt(geophys_dis[1:3,2:5], id.vars = 'Countries.or.areas')
long_meteo = melt(meteo_dis[1:3,2:5], id.vars = 'Countries.or.areas')
long_hydro = melt(hydro_dis[1:3,2:5], id.vars = 'Countries.or.areas')

options(repr.plot.width=20, repr.plot.height=10)
ggplot(long_climate, aes(x = variable, y = value, fill = Countries.or.areas)) +
geom_bar(stat = 'identity', position ='dodge') +
  scale_fill_discrete(name = "Country")+
  scale_x_discrete(labels = c("1990-1999","2000-2009","2010-2019"))+
  ylab("Occurences") + xlab("Decade") +
  labs(title = "Occurences of Climatological Disaster per decade")
```



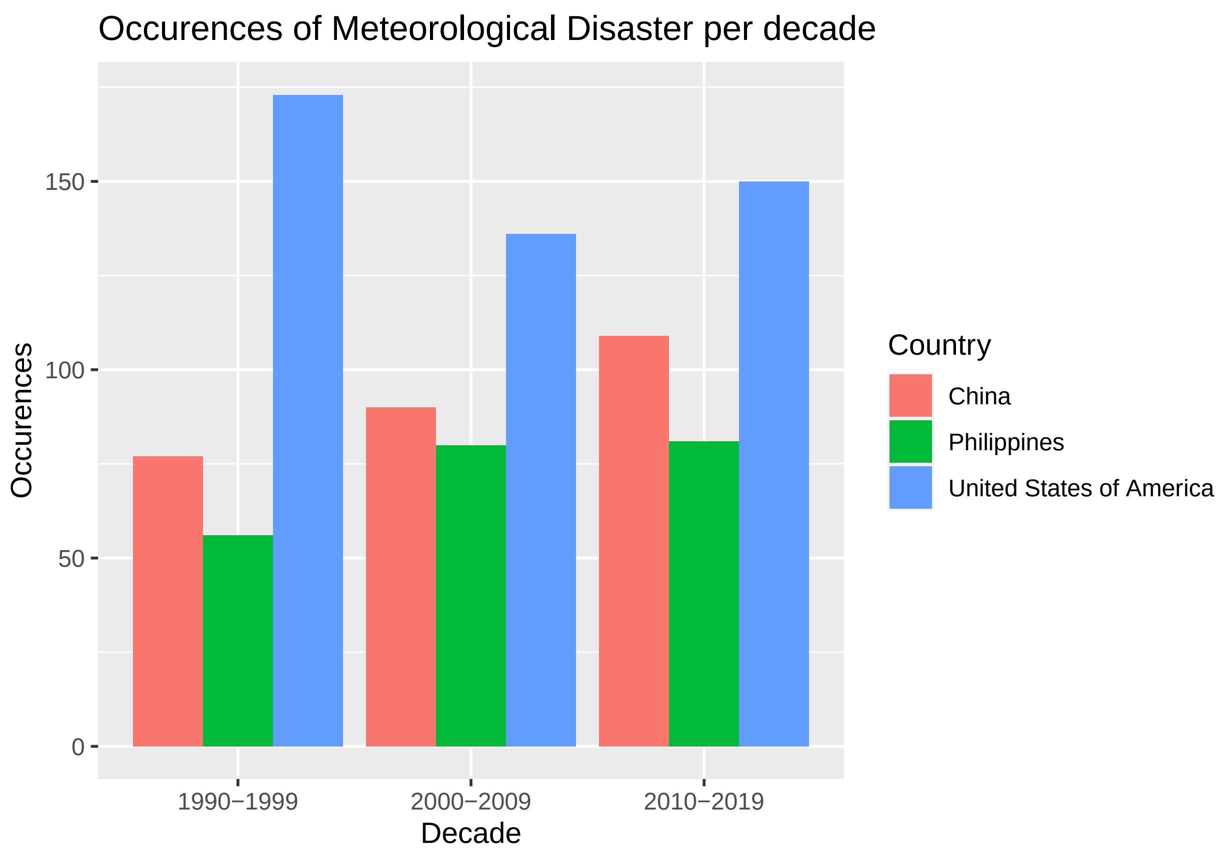

```
options(repr.plot.width=20, repr.plot.height=10)
ggplot(long_geophys, aes(x = variable, y = value, fill = Countries.or.areas)) +
geom_bar(stat = 'identity', position ='dodge')+
  scale_fill_discrete(name = "Country")+
  scale_x_discrete(labels = c("1990-1999","2000-2009","2010-2019"))+
  ylab("Occurences") + xlab("Decade") +
  labs(title = "Occurences of Geophysical Disaster per decade")
```



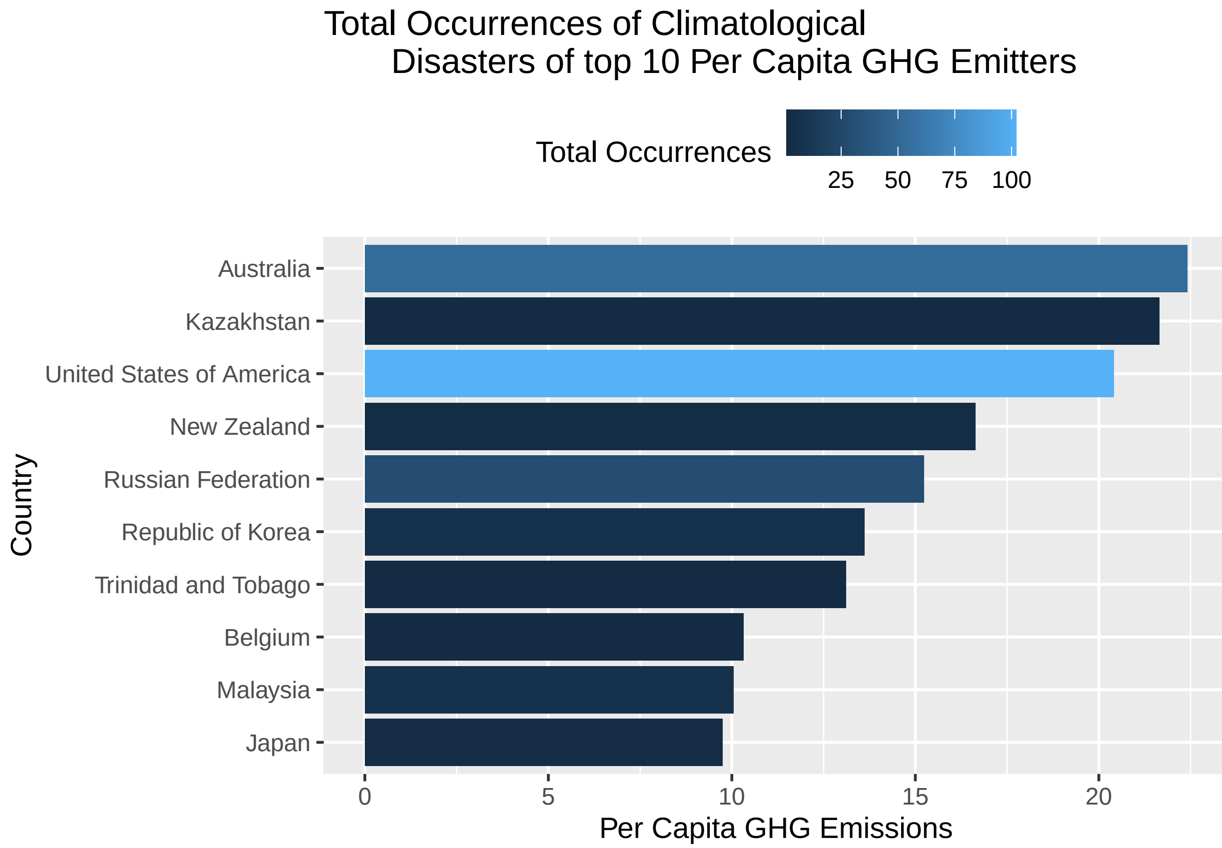

```
options(repr.plot.width=20, repr.plot.height=10)
ggplot(long_hydro, aes(x = variable, y = value, fill = Countries.or.areas)) +
geom_bar(stat = 'identity', position ='dodge')+
  scale_fill_discrete(name = "Country")+
  scale_x_discrete(labels = c("1990-1999","2000-2009","2010-2019"))+
  ylab("Occurences") + xlab("Decade") +
  labs(title = "Occurences of Hydrological Disaster per decade")
```



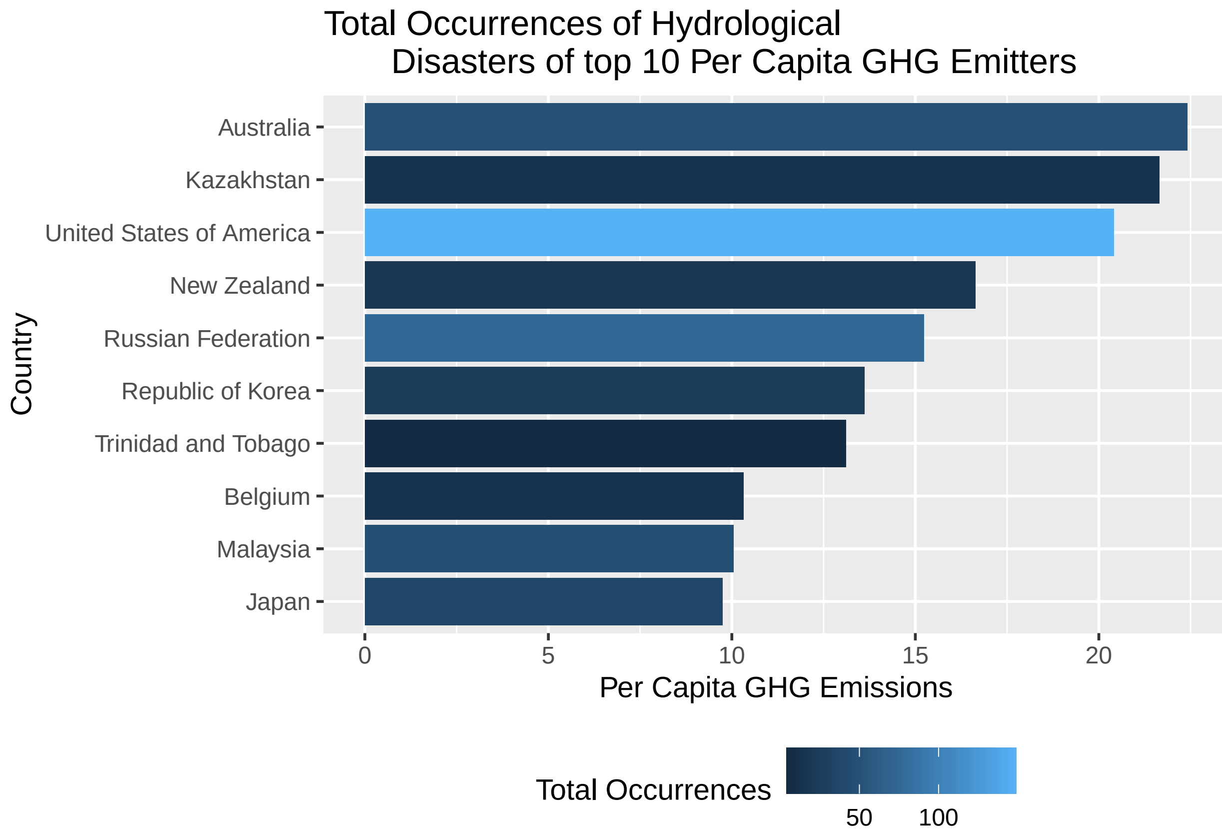

```
options(repr.plot.width=20, repr.plot.height=10)
ggplot(long_meteo, aes(x = variable, y = value, fill = Countries.or.areas)) +
geom_bar(stat = 'identity', position ='dodge')+
  scale_fill_discrete(name = "Country")+
  scale_x_discrete(labels = c("1990-1999","2000-2009","2010-2019"))+
  ylab("Occurences") + xlab("Decade") +
  labs(title = "Occurences of Meteorological Disaster per decade")
```



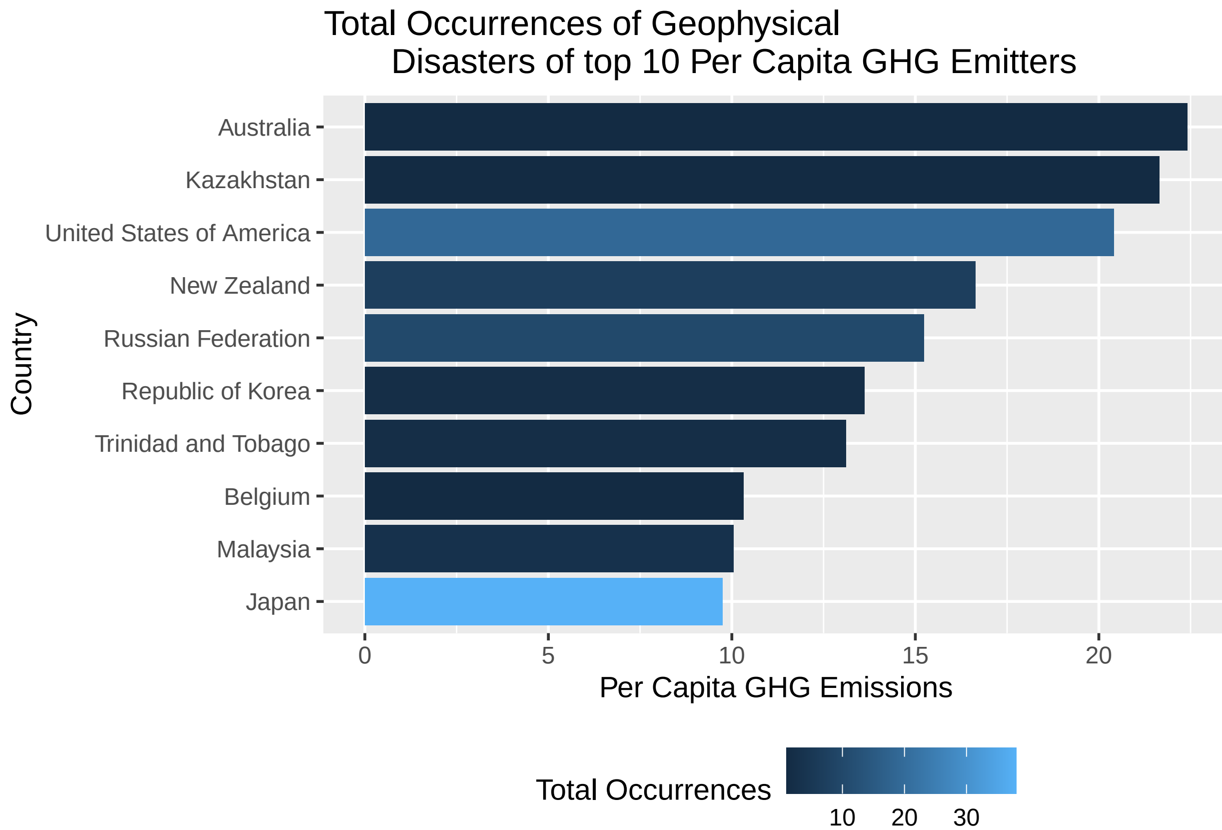

```
merged_ghg_dis =  merge(merged_ghg[,c(1:5)], climate_dis[,c(1,6)],
                        by.x = "Country.ID", by.y = 'CountryID')
merged_ghg_dis =  merge(merged_ghg_dis, hydro_dis[,c(1,6)],
                        by.x = "Country.ID", by.y = 'CountryID')
merged_ghg_dis =  merge(merged_ghg_dis, meteo_dis[,c(1,6)],
                        by.x = "Country.ID", by.y = 'CountryID')
merged_ghg_dis =  merge(merged_ghg_dis, geophys_dis[,c(1,6)],
                        by.x = "Country.ID", by.y = 'CountryID')

merged_ghg_dis = merged_ghg_dis[
  order(merged_ghg_dis$GHG.emissions.per.capita...latest.year, decreasing = T),]
merged_ghg_dis$Total_Occurences = apply(merged_ghg_dis[,c(6:9)], 1, sum)

options(repr.plot.width=20, repr.plot.height=10)
ggplot(merged_ghg_dis[1:10,],
       aes(x = reorder(Country, GHG.emissions.per.capita...latest.year),
          y = GHG.emissions.per.capita...latest.year)) +
geom_bar(stat = "identity", aes(fill = CTotal_Occurences))+
  coord_flip() +
  xlab("Country") + ylab("Per Capita GHG Emissions") +
  theme(legend.position = "top") +
  labs(title = "Total Occurrences of Climatological
       Disasters of top 10 Per Capita GHG Emitters",
       fill = "Total Occurrences")
```



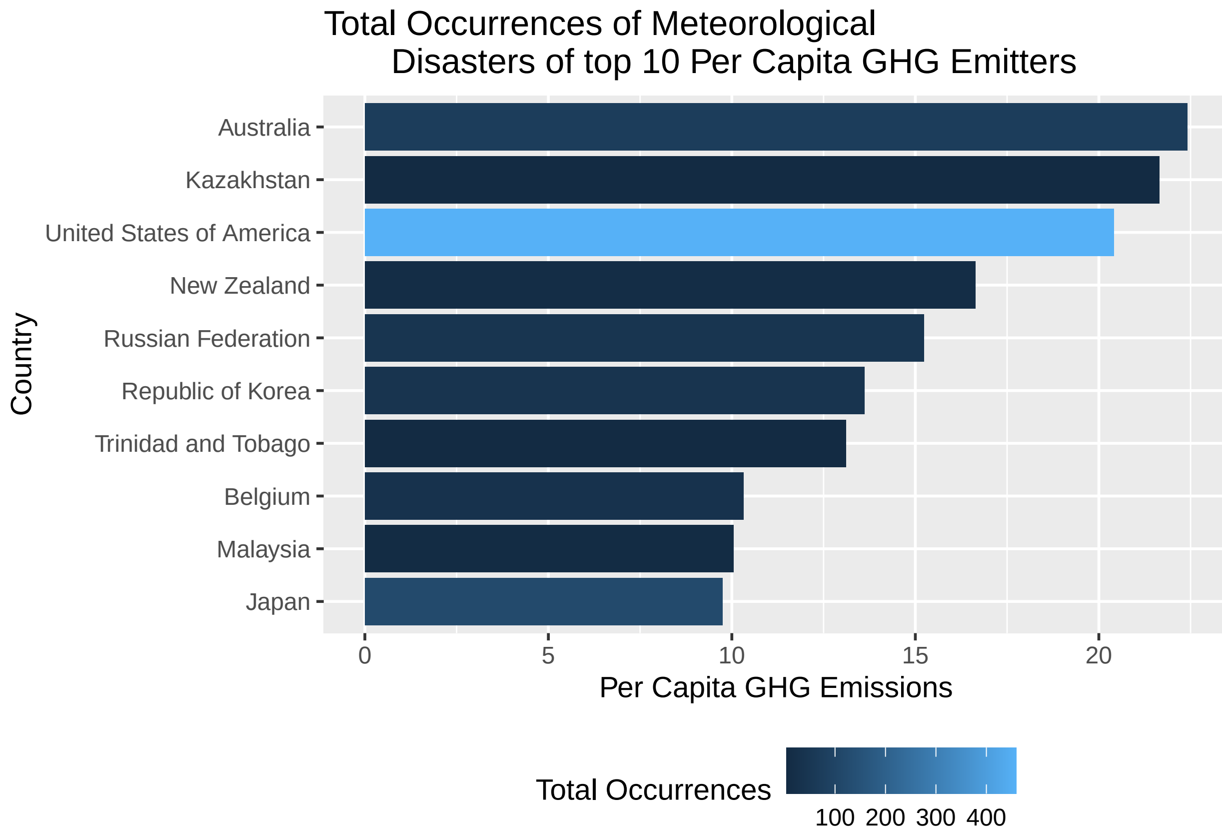

```
options(repr.plot.width=20, repr.plot.height=10)
ggplot(merged_ghg_dis[1:10,], aes(
  x = reorder(Country, GHG.emissions.per.capita...latest.year),
  y = GHG.emissions.per.capita...latest.year)) +
geom_bar(stat = "identity", aes(fill = HTotal_Occurences)) +
    coord_flip() +
  xlab("Country") + ylab("Per Capita GHG Emissions") +
  theme(legend.position = "bottom") +
  labs(title = "Total Occurrences of Hydrological
       Disasters of top 10 Per Capita GHG Emitters",
       fill = "Total Occurrences")
```



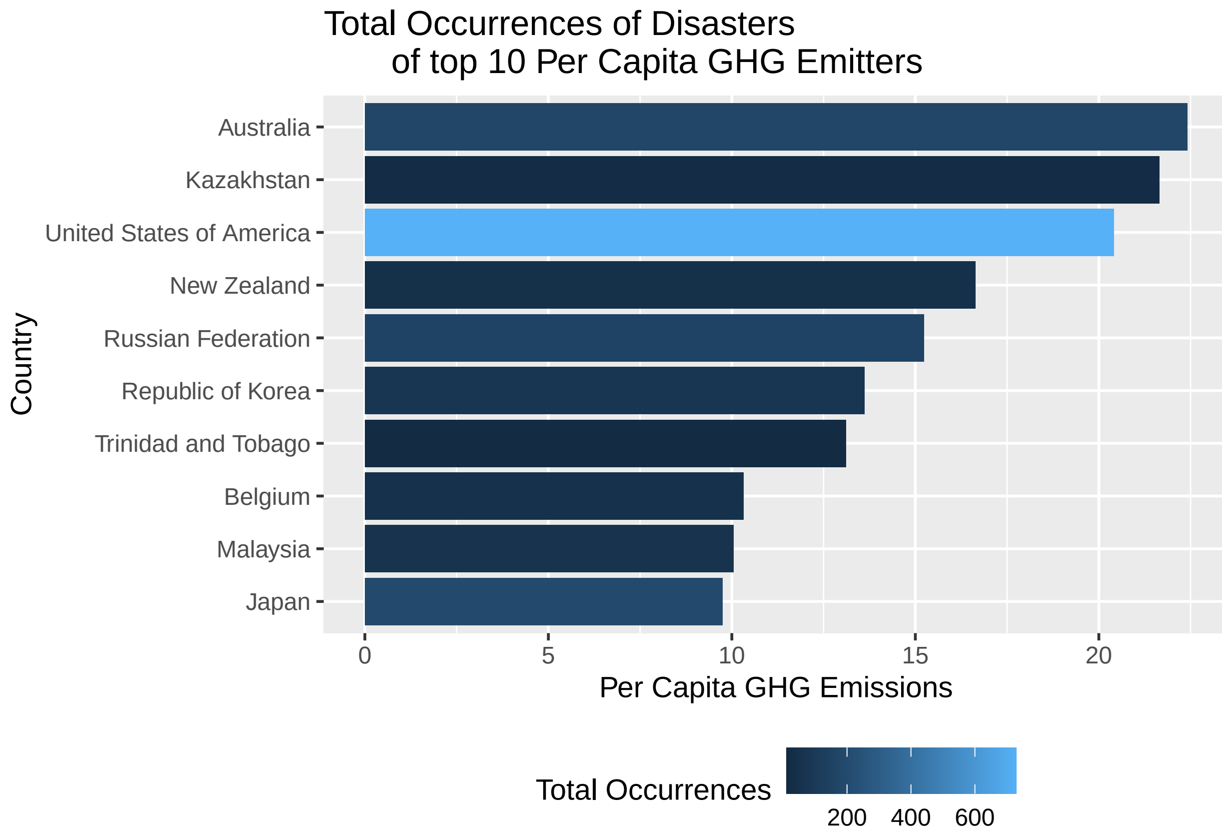

```
options(repr.plot.width=20, repr.plot.height=10)
ggplot(merged_ghg_dis[1:10,],
       aes(x = reorder(Country, GHG.emissions.per.capita...latest.year),
           y = GHG.emissions.per.capita...latest.year)) +
geom_bar(stat = "identity", aes(fill = GTotal_Occurences)) +
    coord_flip() +
  xlab("Country") + ylab("Per Capita GHG Emissions") +
  theme(legend.position = "bottom") +
  labs(title = "Total Occurrences of Geophysical
       Disasters of top 10 Per Capita GHG Emitters",
       fill = "Total Occurrences")
```




```
options(repr.plot.width=20, repr.plot.height=10)
ggplot(merged_ghg_dis[1:10,],
       aes(x = reorder(Country, GHG.emissions.per.capita...latest.year),
           y = GHG.emissions.per.capita...latest.year)) +
geom_bar(stat = "identity", aes(fill = MTotal_Occurences)) +
    coord_flip() +
  xlab("Country") + ylab("Per Capita GHG Emissions") +
  theme(legend.position = "bottom") +
  labs(title = "Total Occurrences of Meteorological
       Disasters of top 10 Per Capita GHG Emitters",
       fill = "Total Occurrences")
```



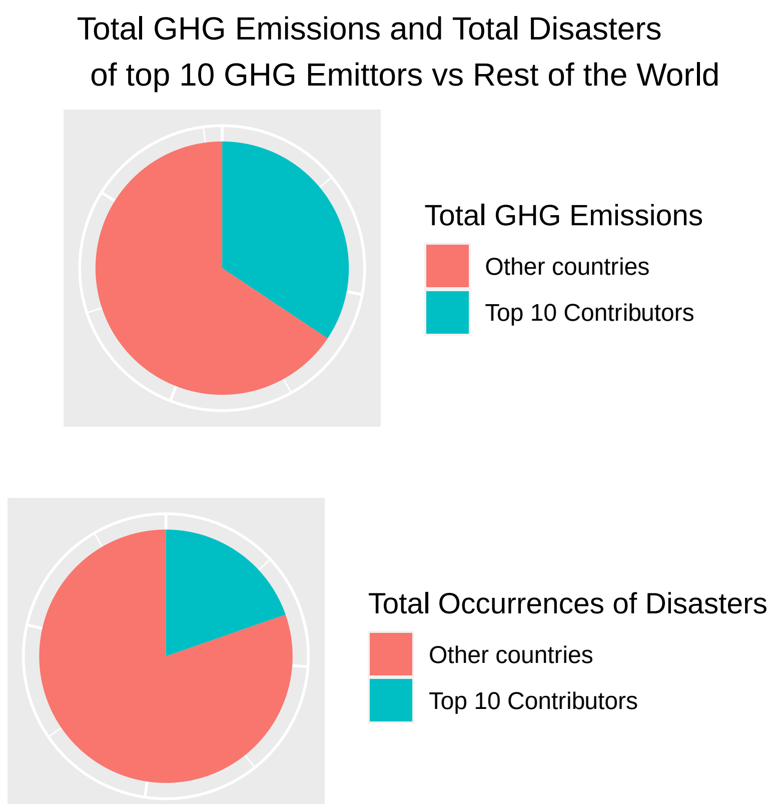

```
options(repr.plot.width=20, repr.plot.height=10)
ggplot(merged_ghg_dis[1:10,],
       aes(x = reorder(Country, GHG.emissions.per.capita...latest.year),
          y = GHG.emissions.per.capita...latest.year)) +
geom_bar(stat = "identity", aes(fill = Total_Occurences)) +
    coord_flip() +
  xlab("Country") + ylab("Per Capita GHG Emissions") +
  theme(legend.position = "bottom") +
  labs(title = "Total Occurrences of Disasters
       of top 10 Per Capita GHG Emitters",
       fill = "Total Occurrences")
```



It can be noted that countries with higher per capita emissions have one of the highest number of occurrences of disasters. Some countries have high disasters of a certain type due to their geographical location, like Japan has a high number of geophysical disasters as it is prone to earthquakes.

```
top10_ghg = data.frame(
    group = c("Top 10 Contributors", "Other countries"),
    values = c(sum(merged_ghg_dis$GHG.total.without.LULUCF..latest.year[1:10]),
              sum(merged_ghg_dis$GHG.total.without.LULUCF..latest.year[
                10:nrow(merged_ghg_dis)]))
)

top10_dis = data.frame(
    group = c("Other countries", "Top 10 Contributors"),
    values = c(sum(merged_ghg_dis$Total_Occurences[
      10:length(merged_ghg_dis$Total_Occurences)]),
               sum(merged_ghg_dis$Total_Occurences[1:10])
              ))

p1 = ggplot(top10_ghg, aes(x = "", y = values, fill = group)) +
geom_bar(stat = "identity", width = 1) +
coord_polar("y", start = 0) +
  labs(fill = "Total GHG Emissions ") +
xlab("") + ylab("") +
  theme(axis.text.x=element_blank(),
        axis.ticks.x=element_blank(),
        axis.text.y=element_blank(),
```

```
        axis.ticks.y=element_blank())

p2 = ggplot(top10_dis, aes(x = "", y = values, fill = group)) +
geom_bar(stat = "identity", width = 1) +
coord_polar("y", start = 0) +
  labs(fill = "Total Occurrences of Disasters") +
  xlab("") + ylab("") +
    theme(axis.text.x=element_blank(),
        axis.ticks.x=element_blank(),
        axis.text.y=element_blank(),
        axis.ticks.y=element_blank())

options(repr.plot.width=20, repr.plot.height=8)
grid.arrange(p1, p2, nrow=2,
             top = "Total GHG Emissions and Total Disasters
       of top 10 GHG Emittors vs Rest of the World")
```



It is to be noted that top 10 GHG emitters contribute to around one-third of the total GHG emissions, but only suffer around one-sixth of the total disasters.

Summary

From the visualisation of Greenhouse Gases Emissions, it can be seen clearly that the variables are heavily skewed and there are too many outliers. The top contributors for greenhouse emissions seem to be the developed nations and those developing nations whose growth is very high, like China, India and Brazil.

From the visualisation of Natural Disasters, it can be seen that different countries suffered from different disasters, but some countries are frequently hit by every kind of disasters, while some don't ever get struck by disasters. This may be due to a bias in data towards accurate reporting of developed nations compared to possible misrepresentation of poorly developed nations.

From the entire visualisation process, the most important findings are noted below.

- Top 10 contributors of greenhouse gases emissions emit around half the the amount emitted by the rest of 180 countries, but only suffer from less than one-fifth of the total disasters.
- Even though fast growing developing countries have high emissions, developed nations have higher per capita emissions. But developing nations are lowering their emissions, while developing nations have high positive percentage change.
- It is clear that energy generation creates the most emissions across all countries.
- There may be a bias in the disasters data in the form of misrepresentation of poorly developed countries, due to which the grave reality of their emaciation is not apparent.

From these findings, a few steps can be suggested for the improvement of the health of our environment.

- Developing countries have to be assisted to adopt improved technologies for energy generation, since those countries will continue to contribute heavily to GHG emissions in this aspect.
- Developing and under developed countries have to be better prepared for disasters, as due to poor infrastructure and laggy preventive measures, they will suffer more casualties even if number of occurrences is low compared to developed countries.
- Collection of accurate data from every single country has to be prioritised so that decisions can be made to assist under represented countries and regions.
