Factors and Impacts of Greenhouse Gas Emissions in US from 1990-2017

https://github.com/fj26/DataAnalytics_Final

Vicky Jia

Contents

List of Tables

List of Figures

1 Rationale and Research Questions

Global warming resulted from greenhouse gas (GHG) emissions is one of the biggest challenges to human beings. While IPCC in its latest special report warns people the critial consequences of 1.5-degree-Celcius temperature increase from the pre-industrial level, the average global temperature on Earth has already increased by about 0.8° Celsius (1.4° Fahrenheit) since 1880.

In the United States, in the period of 1990-2017, the temperature increase and GHG emissions have been greatly contributed by the energy sector. This was attributed to the increase electricity consumptions in all sectors by more people, and electricity generation primarily relies on fossil-fuel combustions. In addition to temperature increase, the economic features also have significant changes along the increasing GHG emissions and electricity consumption in the meantime. Understanding environmental and socioeconomic impacts of GHG emissions is significant to policy makers to consider long-term decision making. Therefore, this project examines the impacts of energy consumptions and demographic changes on GHG emissions, as well as further investigates the environmental and economic effects of GHG emissions in the United States from 1990 to 2017.

Specifically, this project aims to study factors that might contribute to GHG emissions guided by the following questions:

1. What were impacts of US electricity consumption changes on GHG emissions from 1990 to 2017? What sectors of electricity econsumption were significant to total GHG emissions in these 28 years?
2. Was the population growth a signifant factor in affecting GHG emissisions in this period?

Following up with GHG emission causes, this project will then explore the impacts of GHG emissions in US from 1990 to 2017 guided by the following questions:

3. Which sectors were significant to temperature anomalies in this period?
4. Was the total GHG emissions significantly related to economic development in this period, namely gross domestic product (GDP)?

2 Dataset Information

2.1 Raw Database Information

There were 6 datasets used in this project retrieved from various websites. Detailed criteria to collect these raw data in csv file format are listed below.

2.1.1 BEA_GDP_raw.csv

US annual gross domestic production and personal consumption expenditures data from 1990-2017.Data were retrieved from US Bureau of Economic Analysis National Income and Product Accounts (NIPA) Interactive Data Tables tool. Data were transposed by the investigator.
(https://apps.bea.gov/iTable/iTable.cfm?reqid=19&step=2#reqid=19&step=2&isuri=1&1921=survey accessed on 2020-04-11.)

The following selections were made:
First Year: 1990-A&Q
Last Year: 2017-A&Q
Scale: billion
Series: Annual

2.1.2 EIA_electricity-consumption_sector_raw.csv

The annual retail sales of electricity to ultimate customers by sector, by state, by provider from 1990 to 2018 in US. The data was retrieved from Annual Electric Power Industry Report, Form EIA-861 detailed data files from US Energy Information Administration.

(https://www.eia.gov/electricity/data/state/sales_annual.xlsx accessed on 2020-04-11.)

2.1.3 EPA_GHG_Gas_raw.csv

US GHG emission data by gas types in all sectors from 1990-2017 provided by EPA's annual Inventory of U.S. Greenhouse Gas Emissions and Sinks. Data was retrieved from US EPA Greenhouse Gas Inventory Data Explorer. Data were transposed by the investigator.
(https://cfpub.epa.gov/ghgdata/inventoryexplorer/#allsectors/allgas/gas/all accessed on 2020-04-11.)

The following selections were made:
Sector: All sectors
Greenhouse gas: All gases
Break out by: Gas
Year: All years

2.1.4 EPA_GHG_Sector_raw.csv

GHG emission data by economic sectors for all greenhouse gases from 1990-2017 provided by EPA's annual Inventory of U.S. Greenhouse Gas Emissions and Sinks. Data was retrieved

from US EPA Greenhouse Gas Inventory Data Explorer. Data were transposed by the investigator.
(https://cfpub.epa.gov/ghgdata/inventoryexplorer/#allsectors/allgas/econsect/all accessed on 2020-04-11.)

The following selections were made:
Sector: All sectors
Greenhouse gas: All gases
Break out by: Economic sector
Year: All years

2.1.5 NOAA_temp_raw.csv

US mean annual temperature and temperature anomalies from 1990 to 2017. Data was retrieved from NOAA National Centers for Environmental Information Climate at a Glance. (https://www.ncdc.noaa.gov/cag/national/time-series/110/tavg/ann/12/1990-2017?base_prd=true&begbaseyear=1901&endbaseyear=2000, accessed on 2020-04-11.)

The following selections were made:
Parameter: Average Temperature
Time Scale: Annual
Start Year: 1990
End Year: 2017
*Display Base Period: Start: 1990, End: 2017

2.1.6 WB_pop_raw.csv

Total annual population data from 1960-2018 retrieved from World Bank dataset. Data were transposed by the investigator.
(https://data.worldbank.org/indicator/SP.POP.TOTL?locations=US accessed on 2020-04-11.)

The following selections were made in the searching window:
Population, total
United States

2.2 Data Structure

The raw data collected from some databases are beyond the scope of this project, so these datasets were wrangled to fit to the objective of this project. To provide of an overview of the data structure of each raw dataset, raw data were reorganized by gathering same types of data to one column together. Table 1 summarzies the data structure for all raw datasets used in this project after being gathered. These datasets are saved in “Processed Data” folder.

The specific content of each dataset and the detailed procedures of reframing each raw dataset to meet the objective of this project are introduced in Chapter 3 Exploratory Analysis.

Table 1: Summary of Data Structure

Dataset	Variables	Variable.Unit	Data.Source
BEA_GDP	Year	YYYY	US Bureau of Economic Analysis
	Type	Name	
	Value	Billions of dollars	
EIA_electricity-consumption_sector	Year	YYYY	US Energy Information Administration
	State	Name	
	Industry.Sector.Category	Name	
	Sector	Name	
	Electricity_Consumption	Megawatthours	
EPA_GHG_Gas	Year	YYYY	EPA Annual Inventory of U.S. Greenhouse Gas Emissions and Sinks.
	Gas	Name	
	Concentration	Million metric tons of carbon dioxide equivalents	
EPA_GHG_Sector	Year	YYYY	EPA Annual Inventory of U.S. Greenhouse Gas Emissions and Sinks.
	Sector	Name	
	Concentration	Million metric tons of carbon dioxide equivalents	
NOAA_Temp	Date	YYYYMM	NOAA National Centers for Environmental Information Climate at a Glance
	Value	Fahrenheit	
	Anomaly	Fahrenheit	
WB_pop	Year	YYYY	World Bank
	Country	Name	
	Population	person	

3 Exploratory Analysis

Some raw datasets are at global scale, or have longer temporal ranges, which are beyond the scope of this project. Also, some discrete variables were downloaded as characters, which could not be compatible by some statistical models. Therefore, thse datasets were first wrangled to focus on US region in the period between 1990 and 2017. This section demonsrates the process of data wrangling in details, as well as presents the overview of the content of each dataset.

3.1 US GDP

```
#Show the structure of the raw dataset
str(GDP)
```

```
## 'data.frame':    30 obs. of  27 variables:
##  $ Year                                         : int  1990 1991 1992 1993
##  $ Gross.domestic.product                       : num  5963 6158 6520 6859
##  $ Personal.consumption.expenditures            : num  3809 3943 4198 4452
##  $ Goods                                        : num  1491 1497 1563 1642
##  $ Durable.goods                                : num  497 477 508 552 607
##  $ Nondurable.goods                             : num  994 1020 1055 1091
##  $ Services                                     : num  2318 2446 2634 2810
##  $ Gross.private.domestic.investment            : num  993 944 1013 1107 1
##  $ Fixed.investment                             : num  979 945 997 1086 11
##  $ Nonresidential                               : num  739 724 742 799 869
##  $ Structures                                   : num  203 184 173 177 187
##  $ Equipment                                    : num  372 361 382 425 476
##  $ Intellectual.property.products               : num  164 179 188 197 206
##  $ Residential                                  : num  240 221 255 287 324
##  $ Change.in.private.inventories                : num  14.5 -0.4 16.3 20.8
##  $ Net.exports.of.goods.and.services            : num  -77.9 -28.6 -34.7 -
##  $ Exports                                      : num  552 595 633 655 721
```

```
## $ Goods.1                                                 : num  403 430 455 468 518
## $ Services.1                                              : num  149 165 178 187 203
## $ Imports                                                 : num  630 624 668 720 813
## $ Goods.2                                                 : num  508 501 545 593 677
## $ Services.2                                              : num  122 123 123 127 137
## $ Government.consumption.expenditures.and.gross.investment: num  1239 1299 1344 1365
## $ Federal                                                 : num  562 583 588 580 575
## $ National.defense                                        : num  405 414 406 392 382
## $ Nondefense                                              : num  157 169 182 188 193
## $ State.and.local                                         : num  676 716 756 785 828
```

```
#Rename column name for easier use
GDP <- GDP %>%
  rename('GDP' = 'Gross.domestic.product', 'Personal_consumption' = 'Personal.consumptio

#Reframe dataset to meet project objective
GDP.select<- GDP %>%
  filter(Year>1989 & Year < 2018) %>%
  select(Year, GDP, Personal_consumption)
dim(GDP.select)
```

```
## [1] 28  3
```

Figure 1 illustartes the trend of change of US GDP and personal consumptions over the years. Generally, GDP growth was increasing in most years, though slowing down at the beginning of the 21st century and decreasing in 2008-2009 when financial crisis ocurred. The changing trend of personal consumptions over the years was generally consistent to GDP growth, as the personal consumption is one of variables accounted to total GDP. Therefore, from this overview of the dataset, the variable of personal consumption would not be considered as an unique and significant variable to conduct additional analysis. Instead, in economic development analysis in Chapter 4, only total GDP would be studied.

```
## # A tibble: 1 x 1
##       n
##   <int>
## 1    56
```

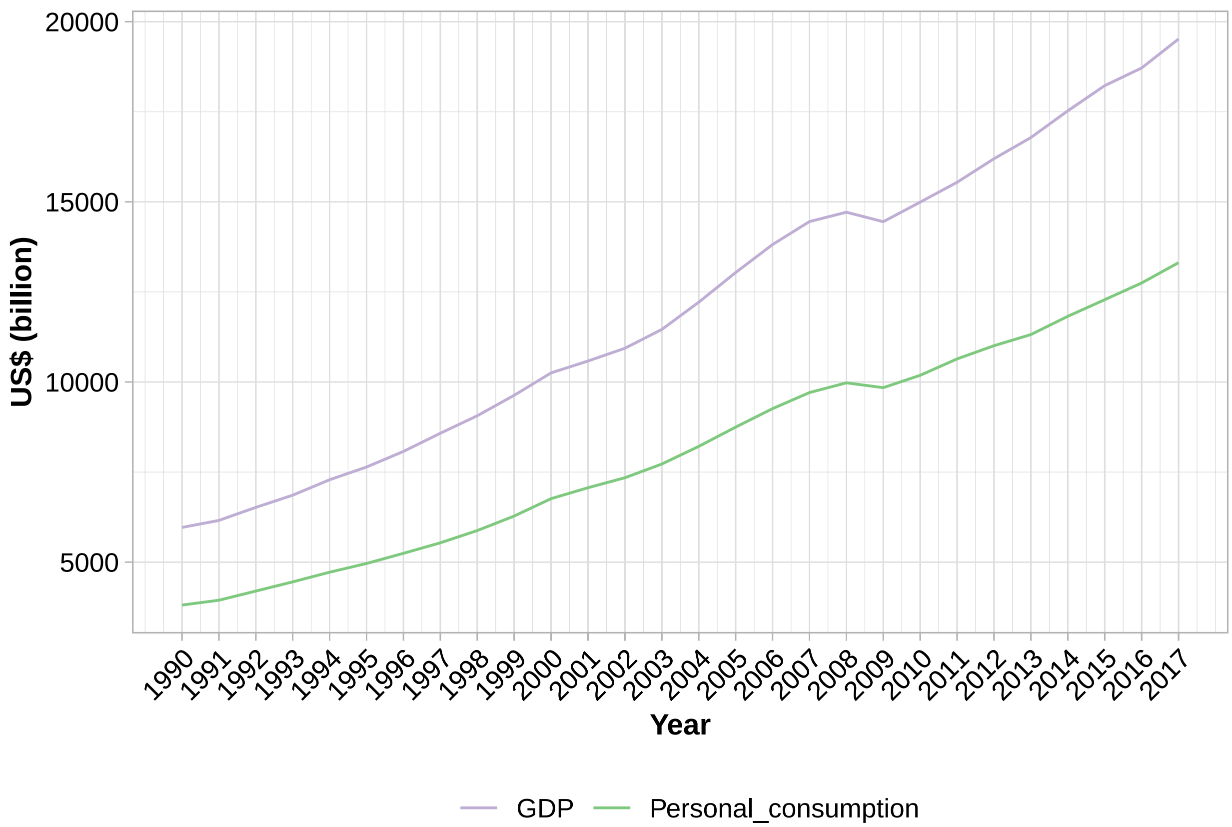

Figure 1: US Economic Features in 1990-2017

3.2 US Electricity Consumption

There were three types of industry sector in the raw dataset. This project chose to focus on "Total Electric Industry", which included all types of survices in electricity consumption in each sector. There were four sectors covered in the electricity consumption dataset, which were illustrated in Table 2. The values of consumption data are in factor format and separted by commas, so it was necessary to convert these values to numeric variables first. By visualizing electricity consumption data in Figure 2, we could observe that all four graphs are in different consumption scales. Notably, transportation sector had the lowest electricity consumption among all sectors, and also there was no data available for transportation sector before 2002. Therefore, transporation sector would be filtered out in the next step of statistic analysis.

```
#Show the structure of the raw dataset
str(electric)
```

```
## 'data.frame':    5015 obs. of  9 variables:
##  $ Year                    : int  2018 2018 2018 2018 2018 2018 2018 2018 2018 2018 .
##  $ State                   : Factor w/ 52 levels "AK","AL","AR",..: 1 2 3 4 5 6 7 8 9
##  $ Industry.Sector.Category: Factor w/ 3 levels "Energy-Only Providers",..: 3 3 3 3 3
##  $ Residential             : Factor w/ 3726 levels "0","1,008,481,682",..: 356 2277 1
##  $ Commercial              : Factor w/ 3893 levels "0","1","1,000,865,367",..: 1345 1
##  $ Industrial              : Factor w/ 3883 levels "-229,740","-298,775",..: 161 2341
##  $ Transportation          : Factor w/ 663 levels "-26,016","0",..: 2 2 432 592 583 6
##  $ Other                   : Factor w/ 1068 levels "-3,968","0","1,016,108",..: NA NA
##  $ Total                   : Factor w/ 3941 levels "0","1,009,966",..: 2723 3808 2576
```

```
#Rename column name for easier use
electric <- electric %>%
  rename('Industry_Sector' = Industry.Sector.Category, 'elec_residential' = Residential,
colnames(electric)
```

```
## [1] "Year"             "State"            "Industry_Sector"
## [4] "elec_residential" "elec_commercial"  "elec_industrial"
## [7] "elec_transport"   "Other"            "Total"
```

```
#Reframe dataset to meet project objective
elec.select <- electric %>%
  filter(State == 'US'& Year>1989 & Year < 2018 & Industry_Sector == 'Total Electric Inc
  select(Year, elec_residential, elec_commercial, elec_industrial, elec_transport)
dim(elec.select)
```

```
## [1] 28  5

## # A tibble: 1 x 1
##       n
##   <int>
## 1    56

## 'data.frame':    112 obs. of  3 variables:
```

Table 2: First 6 rows of electricity consumptions by sector in US (Mwh)

Year	elec_residential	elec_commercial	elec_industrial	elec_transport
2017	1,378,647,742	1,352,887,694	984,297,945	7,522,593
2016	1,411,058,153	1,367,191,386	976,715,181	7,496,910
2015	1,404,096,499	1,360,751,527	986,507,732	7,636,632
2014	1,407,208,311	1,352,158,263	997,576,138	7,757,555
2013	1,394,812,129	1,337,078,777	985,351,874	7,625,041
2012	1,374,514,708	1,327,101,196	985,713,854	7,320,028

```
##  $ Year                   : int  2017 2016 2015 2014 2013 2012 2011 2010 2009 2008 ..
##  $ Sector                 : chr  "elec_residential" "elec_residential" "elec_resident
##  $ Electricity_Consumption: num  1.38e+09 1.41e+09 1.40e+09 1.41e+09 1.39e+09 ...
```

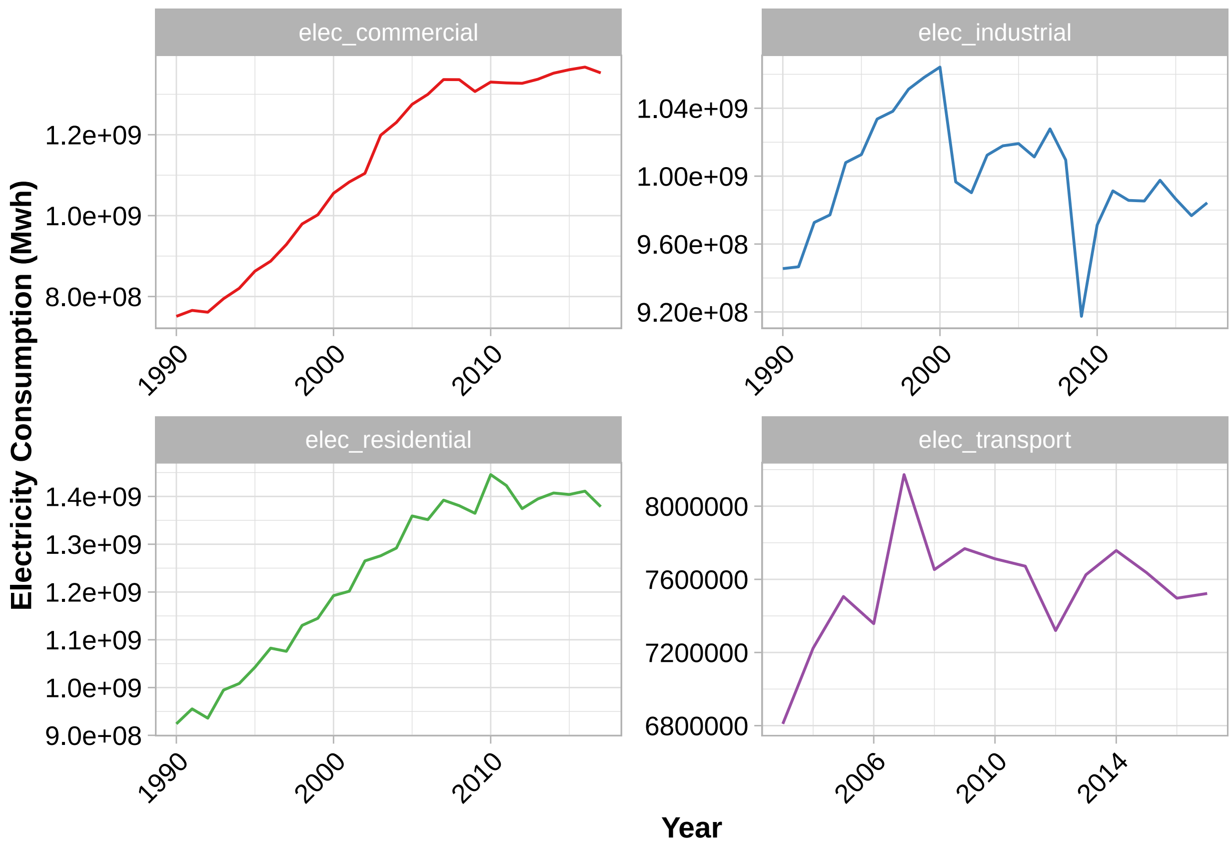

Figure 2: US Electricty Consumption by Sector in 1990-2017

3.3 US GHG Concentration by Gas

GHG concentration data refers to the concentration of different GHGs emitted to the atmosphere. The upper graph in Figure 3 displays the trends for all GHGs together, from which we can see CO2 was the major GHG emitted by human beings. To observe the trend of change more clearly, graphs were created below to show the concentration over years for each GHG. It was hard to summarize the general trend for all GHG emissions, so in the next step of analysis, the GHG type comparison would be first conducted

```
#Show the structure of the raw data
str(GHG.gas)
```

```
## 'data.frame':    28 obs. of  6 variables:
##  $ Year              : int  1990 1991 1992 1993 1994 1995 1996 1997 1998 1999 ...
##  $ Carbon.dioxide    : num  5121 5072 5175 5281 5375 ...
##  $ Methane           : num  780 784 783 770 775 ...
##  $ Nitrous.oxide     : num  370 369 372 385 377 ...
##  $ Fluorinated.gases: num  99.7 90.7 95.3 95 98.1 ...
##  $ Total             : num  6371 6316 6425 6532 6625 ...
```

```
#Rename column name for easier use
colnames(GHG.gas)
```

```
## [1] "Year"              "Carbon.dioxide"    "Methane"
## [4] "Nitrous.oxide"     "Fluorinated.gases" "Total"
```

```
GHG.gas <- GHG.gas %>%
  rename("CO2" = Carbon.dioxide, 'CH4' = Methane, 'N2O' = Nitrous.oxide, 'Fgas' = Fluori
colnames(GHG.gas)
```

```
## [1] "Year"  "CO2"   "CH4"   "N2O"   "Fgas"  "Total"
```

Table 3: First 6 years of GHG emissions by gas type (MMTCO2e)

Year	CO2	CH4	N2O	Fgas	Total
1990	5121.179	779.8456	370.3077	99.66786	6371.001
1991	5071.564	784.3849	368.9618	90.70467	6315.615
1992	5174.671	783.1766	371.7864	95.30071	6424.934
1993	5281.387	770.3084	385.3472	95.02735	6532.070
1994	5375.034	775.1607	376.5115	98.12981	6624.836
1995	5436.698	767.8453	388.5028	117.02114	6710.067

```
## # A tibble: 1 x 1
##       n
##   <int>
## 1   140
```

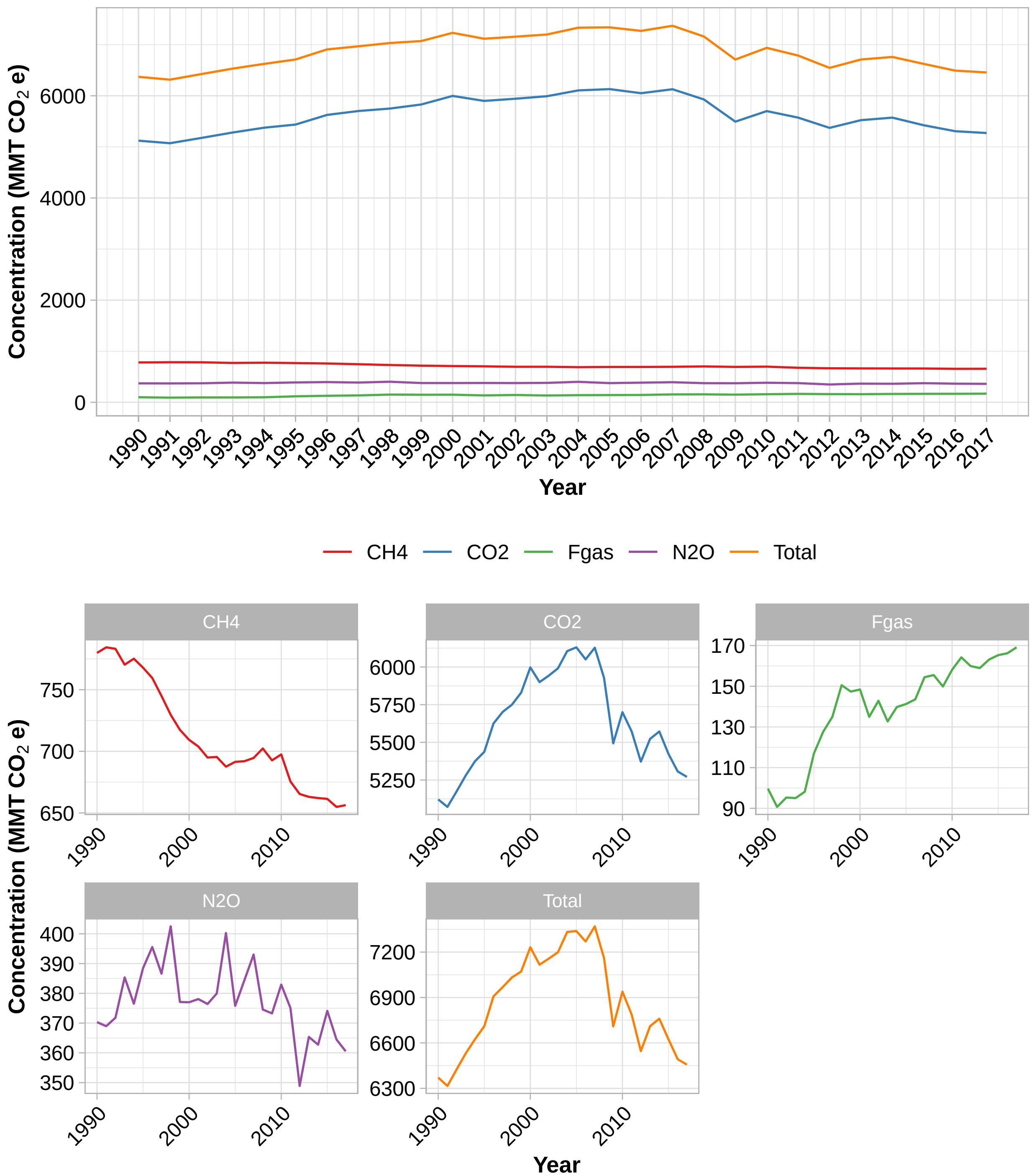

Figure 3: US GHG Concentration by Gas Type in 1990-2017 (MMTCO2e)

3.4 US GHG Emissions by Sector

Figure 4 illustrates the GHG emissions over the same period by sector. The ideology of Figure 4 was same as Figure 3, which internally compared the emissions by sectors. The total sectoral GHG emission was same as the sum of GHG concentration by gas type. Among all 6 sectors emitting GHGs, electricity generation, transporation, and industial sector had highest emissions. It was inspired to study how sectoral emissions had contributed to total GHG emissions, and how they were different from each other.

```
str(GHG.sector)
```

```
## 'data.frame':    28 obs. of  9 variables:
##  $ Year                    : int  1990 1991 1992 1993 1994 1995 1996 1997 1998 1999 ...
##  $ Transportation          : num  1527 1481 1541 1578 1632 ...
##  $ Electricity.generation: num  1876 1872 1887 1962 1987 ...
##  $ Industry                : num  1629 1601 1632 1605 1630 ...
##  $ Agriculture             : num  535 535 539 552 544 ...
##  $ Commercial              : num  427 434 429 425 427 ...
##  $ Residential             : num  345 354 361 372 363 ...
##  $ U.S..territories        : num  33.3 39.1 37.3 39.1 41.5 ...
##  $ Total                   : num  6371 6316 6425 6532 6625 ...
```

```
#Rename column name for easier use
colnames(GHG.sector)
```

```
## [1] "Year"                   "Transportation"
## [3] "Electricity.generation" "Industry"
## [5] "Agriculture"            "Commercial"
## [7] "Residential"            "U.S..territories"
## [9] "Total"
```

```
GHG.sector <- GHG.sector %>%
  rename('Electricity_generation' = Electricity.generation)
colnames(GHG.sector)
```

```
## [1] "Year"                   "Transportation"
## [3] "Electricity_generation" "Industry"
## [5] "Agriculture"            "Commercial"
## [7] "Residential"            "U.S..territories"
## [9] "Total"
```

```
#Reframe dataset to meet project objective
GHG.sector.select <- GHG.sector %>%
  select(Year, Transportation, Electricity_generation, Industry, Agriculture, Commercial
dim(GHG.sector.select)
```

```
## [1] 28  7
```

```
## # A tibble: 1 x 1
```

Table 4: First 6 years of GHG emissions by sector (MMTCO2e)

Year	Transportation	Electricity_generation	Industry	Agriculture	Commercial	Residential
1990	1527.077	1875.537	1628.556	534.8599	426.9285	344.7218
1991	1480.932	1871.568	1601.088	534.6403	433.9764	354.2868
1992	1540.536	1886.539	1631.550	538.7436	429.4007	360.8492
1993	1577.523	1962.302	1604.889	551.5106	424.5561	372.2031
1994	1632.154	1987.103	1629.785	543.9804	427.1918	363.1420
1995	1667.330	2003.827	1647.961	556.7418	426.3748	367.4087

```
##       n
##   <int>
## 1   168
```

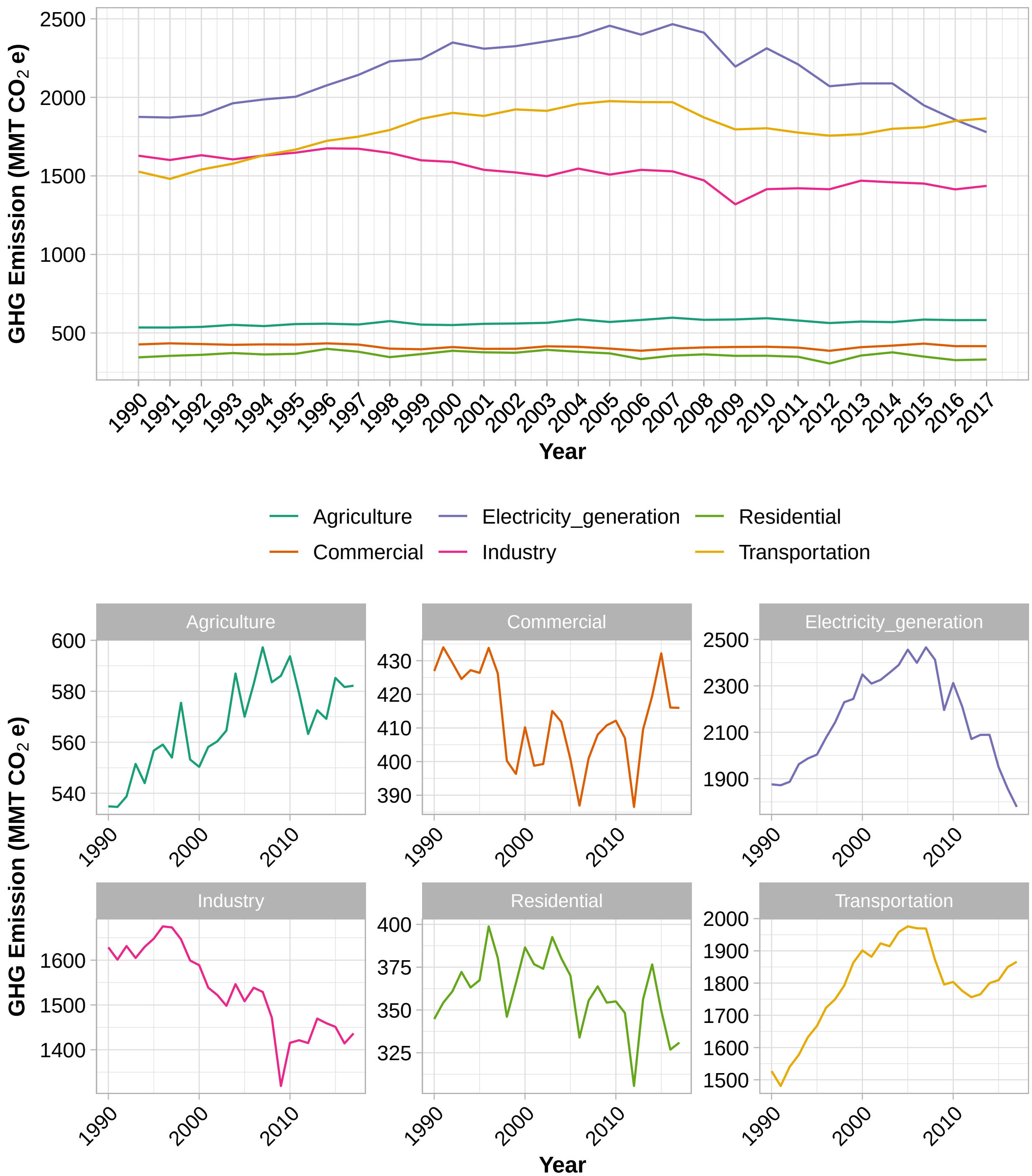

Figure 4: US GHG Emissions by Sector in 1990-2017 (MMTCO2e)

3.5 US Mean Annual Temperature

Mean annual temperature and temperature anomalies over 1990-2017 were visualized in Figure 5. Temperature anomaly refers to the difference between annual mean temperature and overall averaage temperature for that period. Therefore, the overall average temperature were also calculated and mutated as a new column in the table. Since overall average temperature is constant, temperature anomalies are in the same changing trends as annual mean temperature. As referene values, temperature anomalies can better and more easily demonstrate increasing and decreasing trends by positive and negative symbols. Therefore, temperature anomaly would be used in the future analyses.

```
str(temp)
```

```
## 'data.frame':    28 obs. of  3 variables:
##  $ Date   : int  199012 199112 199212 199312 199412 199512 199612 199712 199812 19991
##  $ Value  : num  53.5 53.2 52.6 51.3 52.9 ...
##  $ Anomaly: num  0.25 -0.1 -0.66 -2 -0.39 -0.61 -1.38 -1.06 0.97 0.62 ...
```

```
#Convert year characters to date format
colnames(temp)
```

```
## [1] "Date"    "Value"   "Anomaly"
```

```
temp <- mutate(temp, Year = year(as.Date(as.yearmon(as.character(temp$Date), format = "
str(temp)
```

```
## 'data.frame':    28 obs. of  4 variables:
##  $ Date   : int  199012 199112 199212 199312 199412 199512 199612 199712 199812 19991
##  $ Value  : num  53.5 53.2 52.6 51.3 52.9 ...
##  $ Anomaly: num  0.25 -0.1 -0.66 -2 -0.39 -0.61 -1.38 -1.06 0.97 0.62 ...
##  $ Year   : num  1990 1991 1992 1993 1994 ...
```

```
#Reframe dataset to meet project objective
temp.select <- temp %>%
  select(Year, Value, Anomaly) %>%
  mutate(Mean_all_temp =  mean(Value))
dim(temp.select)
```

```
## [1] 28  4
```

```
str(temp.select)
```

```
## 'data.frame':    28 obs. of  4 variables:
##  $ Year         : num  1990 1991 1992 1993 1994 ...
##  $ Value        : num  53.5 53.2 52.6 51.3 52.9 ...
##  $ Anomaly      : num  0.25 -0.1 -0.66 -2 -0.39 -0.61 -1.38 -1.06 0.97 0.62 ...
##  $ Mean_all_temp: num  53.3 53.3 53.3 53.3 53.3 ...
```

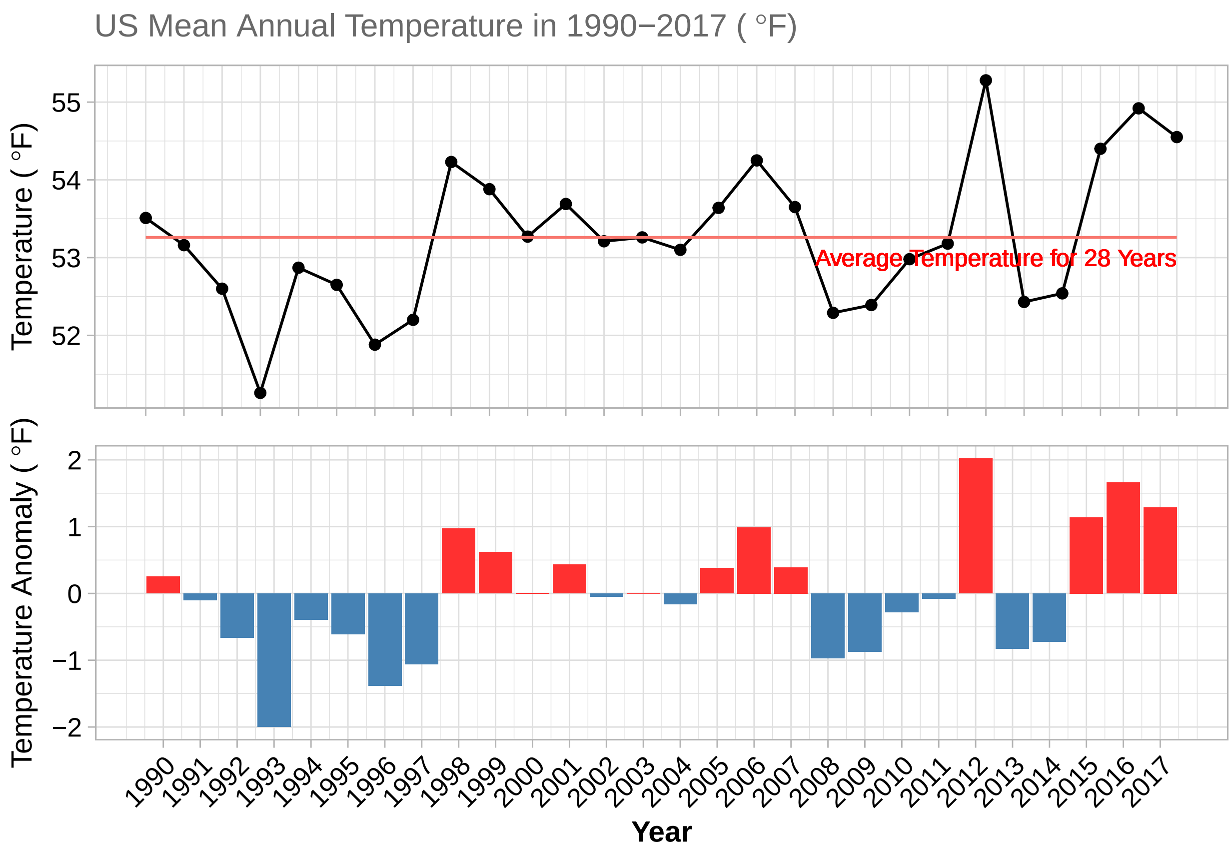

Figure 5: US Mean Annual Temperature Anomaly in 1990-2017 (F)

Table 5: First 6 years of US temperature data in 1990-2017 (F)

Year	Value	Anomaly	Mean_all_temp
1990	53.51	0.25	53.25964
1991	53.16	-0.10	53.25964
1992	52.60	-0.66	53.25964
1993	51.26	-2.00	53.25964
1994	52.87	-0.39	53.25964
1995	52.65	-0.61	53.25964

3.6 US Population

Population data retrieved from World Bank is at global scale covering 60 years since 1960. Therefore, both geograhic and temporal selection of the dataset have to be reframed. The sample of the wrangled dateset was shown as Table 6.

```r
str(ppl)
```

```
## 'data.frame':    60 obs. of  265 variables:
##  $ Year                                                              : int  1960 1961 1962 1963 196
##  $ Aruba                                                             : int  54211 55438 56225 56695
##  $ Afghanistan                                                       : int  8996973 9169410 9351441
##  $ Angola                                                            : int  5454933 5531472 5608539
##  $ Albania                                                           : int  1608800 1659800 1711319
##  $ Andorra                                                           : int  13411 14375 15370 16412
##  $ Arab.World                                                        : int  92197753 94724510 97334
##  $ United.Arab.Emirates                                              : int  92418 100796 112118 125
##  $ Argentina                                                         : int  20481779 20817266 21153
##  $ Armenia                                                           : int  1874121 1941492 2009526
##  $ American.Samoa                                                    : int  20123 20602 21253 22034
##  $ Antigua.and.Barbuda                                               : int  54131 55001 55841 56702
##  $ Australia                                                         : int  10276477 10483000 10742
##  $ Austria                                                           : int  7047539 7086299 7129864
##  $ Azerbaijan                                                        : int  3895397 4030322 4171426
##  $ Burundi                                                           : int  2797932 2852438 2907321
##  $ Belgium                                                           : int  9153489 9183948 9220578
##  $ Benin                                                             : int  2431622 2465867 2502896
##  $ Burkina.Faso                                                      : int  4829288 4894580 4960326
##  $ Bangladesh                                                        : int  48013504 49362843 50752
##  $ Bulgaria                                                          : int  7867374 7943118 8012946
##  $ Bahrain                                                           : int  162427 167894 173144 17
##  $ Bahamas..The                                                      : int  109534 115121 121091 12
##  $ Bosnia.and.Herzegovina                                            : int  3225668 3288603 3353226
##  $ Belarus                                                           : int  8198000 8271216 8351928
##  $ Belize                                                            : int  92064 94703 97384 10016
##  $ Bermuda                                                           : int  44400 45500 46600 47700
```

```
## $ Bolivia                                      : int  3656955 3728964 3802990
## $ Brazil                                       : int  72179226 74311343 76514
## $ Barbados                                     : int  230980 231718 232633 23
## $ Brunei.Darussalam                            : int  81702 85562 89481 93540
## $ Bhutan                                       : int  223288 228851 234554 24
## $ Botswana                                     : int  502745 512685 523778 53
## $ Central.African.Republic                     : int  1501668 1526066 1551910
## $ Canada                                       : int  17909009 18271000 18614
## $ Central.Europe.and.the.Baltics               : int  91401764 92232738 93009
## $ Switzerland                                  : int  5327827 5434294 5573815
## $ Channel.Islands                              : int  109420 110399 111457 11
## $ Chile                                        : int  8132990 8303811 8476897
## $ China                                        : int  667070000 660330000 665
## $ Cote.d.Ivoire                                : int  3503553 3631553 3770766
## $ Cameroon                                     : int  5176918 5285017 5398729
## $ Congo..Dem..Rep.                             : int  15248251 15637699 16041
## $ Congo..Rep.                                  : int  1018253 1043116 1069238
## $ Colombia                                     : int  16057724 16567811 17092
## $ Comoros                                      : int  191121 194139 197198 20
## $ Cabo.Verde                                   : int  201765 205327 210142 21
## $ Costa.Rica                                   : int  1330782 1381183 1433335
## $ Caribbean.small.states                       : int  4194710 4274060 4353628
## $ Cuba                                         : int  7141250 7291200 7453540
## $ Curacao                                      : int  124826 126125 128414 13
## $ Cayman.Islands                               : int  7865 8026 8146 8227 829
## $ Cyprus                                       : int  572930 576395 577691 57
## $ Czech.Republic                               : int  9602006 9586651 9624660
## $ Germany                                      : int  72814900 73377632 74025
## $ Djibouti                                     : int  83636 88498 94204 10062
## $ Dominica                                     : int  60011 61032 61982 62918
## $ Denmark                                      : int  4579603 4611687 4647727
## $ Dominican.Republic                           : int  3294224 3406280 3521018
## $ Algeria                                      : int  11057863 11336339 11619
## $ East.Asia...Pacific..excluding.high.income.  : int  894880101 894484115 906
## $ Early.demographic.dividend                   : num  9.80e+08 1.00e+09 1.03e
## $ East.Asia...Pacific                          : num  1.04e+09 1.04e+09 1.06e
## $ Europe...Central.Asia..excluding.high.income.  : int  275147578 279443876 283
## $ Europe...Central.Asia                        : int  667253650 674962981 682
## $ Ecuador                                      : int  4543666 4674172 4809201
## $ Egypt..Arab.Rep.                             : int  26632894 27366237 28112
## $ Euro.area                                    : int  265203934 267621101 270
## $ Eritrea                                      : int  1007590 1033328 1060486
## $ Spain                                        : int  30455000 30739250 31023
## $ Estonia                                      : int  1211537 1225077 1241623
## $ Ethiopia                                     : int  22151278 22671191 23221
```

```
##  $ European.Union                                     : int  356906076 359998418 363
##  $ Fragile.and.conflict.affected.situations           : int  174658134 178611702 182
##  $ Finland                                            : int  4429634 4461005 4491443
##  $ Fiji                                               : int  393481 407249 421665 43
##  $ France                                             : int  46621669 47240543 47904
##  $ Faroe.Islands                                      : int  34615 35076 35524 35969
##  $ Micronesia..Fed..Sts.                              : int  44514 45932 47367 48855
##  $ Gabon                                              : int  500928 505799 511287 51
##  $ United.Kingdom                                     : int  52400000 52800000 53250
##  $ Georgia                                            : int  3645600 3703600 3760300
##  $ Ghana                                              : int  6635230 6848295 7071971
##  $ Gibraltar                                          : int  23420 23813 24313 24889
##  $ Guinea                                             : int  3494162 3552065 3611429
##  $ Gambia..The                                        : int  365047 372445 379894 38
##  $ Guinea.Bissau                                      : int  616136 622761 628883 63
##  $ Equatorial.Guinea                                  : int  255333 258791 262219 26
##  $ Greece                                             : int  8331725 8398050 8448233
##  $ Grenada                                            : int  89932 91327 92484 93413
##  $ Greenland                                          : int  32500 33700 35000 36400
##  $ Guatemala                                          : int  4210747 4336143 4464249
##  $ Guam                                               : int  66742 68072 69604 71286
##  $ Guyana                                             : int  571819 589274 606285 62
##  $ High.income                                        : int  760193906 771546780 781
##  $ Hong.Kong.SAR..China                               : int  3075605 3168100 3305200
##  $ Honduras                                           : int  2038632 2096409 2155647
##  $ Heavily.indebted.poor.countries..HIPC.             : int  161734357 165573152 169
##  $ Croatia                                            : int  4140181 4167292 4196712
##   [list output truncated]
```

```r
#Reframe dataset to meet project objective
ppl.select <- ppl %>%
  select(Year, United.States) %>%
  filter(Year >1989 & Year < 2018)
dim(ppl.select)
```

```
## [1] 28  2
```

```r
#Rename column name for easier use
colnames(ppl.select)
```

```
## [1] "Year"          "United.States"
```

```r
ppl.select <- ppl.select %>%
  rename('US_pop' = United.States)
colnames(ppl.select)
```

```
## [1] "Year"   "US_pop"
```

Table 6: First 6 years of US population in 1990-2017

Year	US_pop
1990	249623000
1991	252981000
1992	256514000
1993	259919000
1994	263126000
1995	266278000

```
#Save all processed code to the Processed folder
write.csv(GDP.select, row.names = FALSE, file = './Processed Data/BEA_GDP_processed.csv'
write.csv(elec.select, row.names = FALSE, file = './Processed Data/EIA_electricity-consu
write.csv(GHG.gas, row.names = FALSE, file = './Processed Data/EPA_GHG_Gas_processed.csv
write.csv(GHG.sector.select, row.names = FALSE, file = './Processed Data/EPA_GHG_Sector_
write.csv(temp.select, row.names = FALSE, file = './Processed Data/NOAA_temp_processed.c
write.csv(ppl.select, row.names = FALSE, file = './Processed Data/WB_pop_processed.csv')
```

```
#Reload new datasets
newGDP <- read.csv('./Processed Data/BEA_GDP_processed.csv', head(T))
newelectricity <- read.csv('./Processed Data/EIA_electricity-consumption_sector_processe
newGHGgas <- read.csv('./Processed Data/EPA_GHG_Gas_processed.csv', head(T))
newGHGsector <- read.csv('./Processed Data/EPA_GHG_Sector_processed.csv', head(T))
newtemp <- read.csv('./Processed Data/NOAA_temp_processed.csv', head(T))
newpop <- read.csv('./Processed Data/WB_pop_processed.csv', head(T))

allGHG <- full_join(newGHGgas,newGHGsector, by = 'Year')
GHG_contribute <- full_join(newelectricity, newpop, by = 'Year')
GHG_impact <- full_join(newGDP, newtemp, by = 'Year')
```

4 Analysis

Figures in this sections are disordered - the float package was not compatible to my system. Texts in this section will specify which figures are referring to in certain analyses.

4.1 Question 1: What factors have contributed to GHG emissions in these 28 years?

4.1.1 Were emissions of different GHGs related to each other?

As discussed in Chapter 3, there was no general trend for the concentration of all GHGs. Therefore, before delving into external impacts on GHG concentrations, how different types of GHGs were varied from each other were first studied.

The distribution of all GHG data were assessed with shapiro.test. Most p-values were greater than 0.05, which implied that the distribution of these data were not significantly different from normal distribution. Therefore, these data were normally distributed. Other data, namely CH4 and Fgas, obtained a p-value less than 0.05 (even transforming these two variables by taking the log and square root), which indicated they were not normally distributed. However, from the normal Q-Q plot, the scatters lied close to the line and symmetry across the qq normal line, so these two data could still be seen as normally distributed. All GHG emissions by sector obtained p-values greater than 0.05 from the shapiro.test, so they were normally distributed. Parametric tests would be used for all GHG dataset.

Multiple Linear Regression was conducted between CO2 and other GHGs to see if there could be any relationships between CO2 and other GHGs, as there are always multiple GHGs emitted together by one source. The result illustarted that at least one of GHGs was significantly related to the CO2 emission (linear regression; R2=0.5562, dF = 24, p = 0.0001805). Specifically, changes in N2O emissions were significantly related to changes in CO2 emissions, while changes in CH4 and Fgas were not significantly related to CO2 emissions.

To step further, a correlation plot was created to reveal the possible relationships, as well as covariance among explanatory variables, as shown in Figure 6, in which blue indiates positively correlated and red indicate negatively correlated. The smaller ellipse indicates a higher correlationship between two variables, and the dark color implies a higher correlation cofficient. Therefore, both N2O and Fgas had positive correlationships with CO2, but N2O was more correlated and had a higher correlation coeffeicnt; on the other hand, CH4 had a negative correlationship with CO2 and such correlation was not significant as the ellipse was large. However, it was interesting to see CH4 and Fgas had a strong negative correlationship. Another format to observe the correlationship with CO2 concentration was shown in Figure 7.

Based on this reuslt, Akaike's Information Criterion (AIC) were computed to improve the model to best predict impacts on CO2 emission, from which only N2O and CH4 were remained in the improved model (Step: AIC=306.73). The improved model implied that both CH4 and N2O emissions were significantly related to CO2 emissions (linear regression; R2=0.5461,

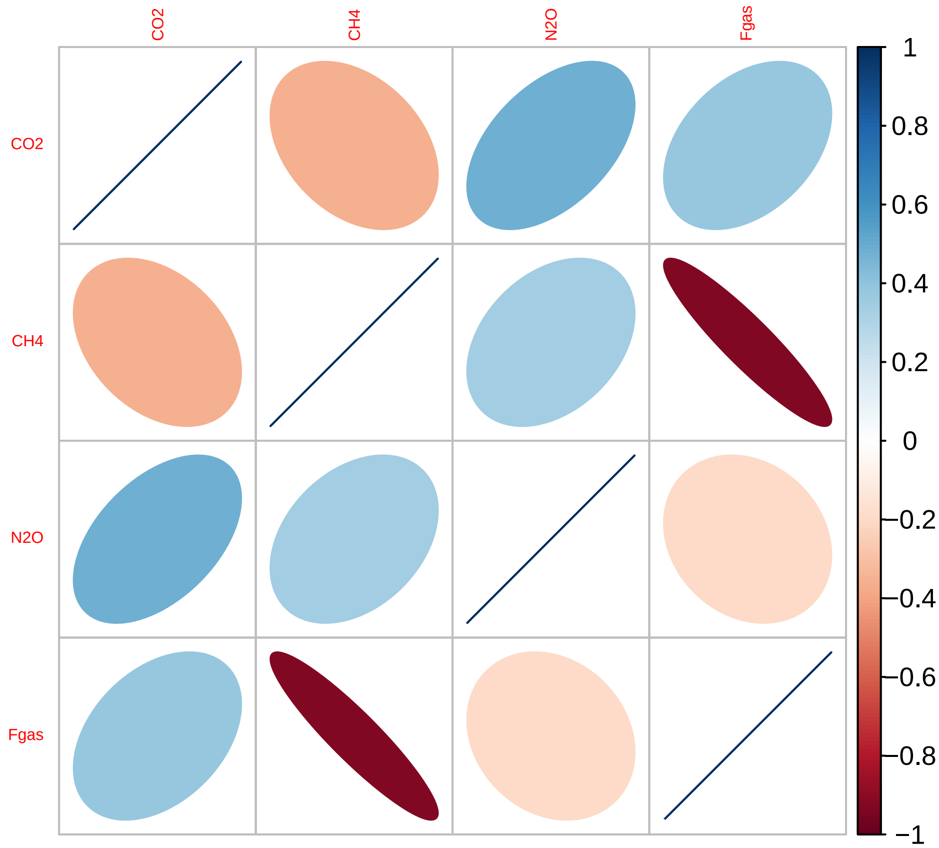

Figure 6: Correlation Plot of US GHG Concentration by Gas

dF = 25, $p < 0.0001$).

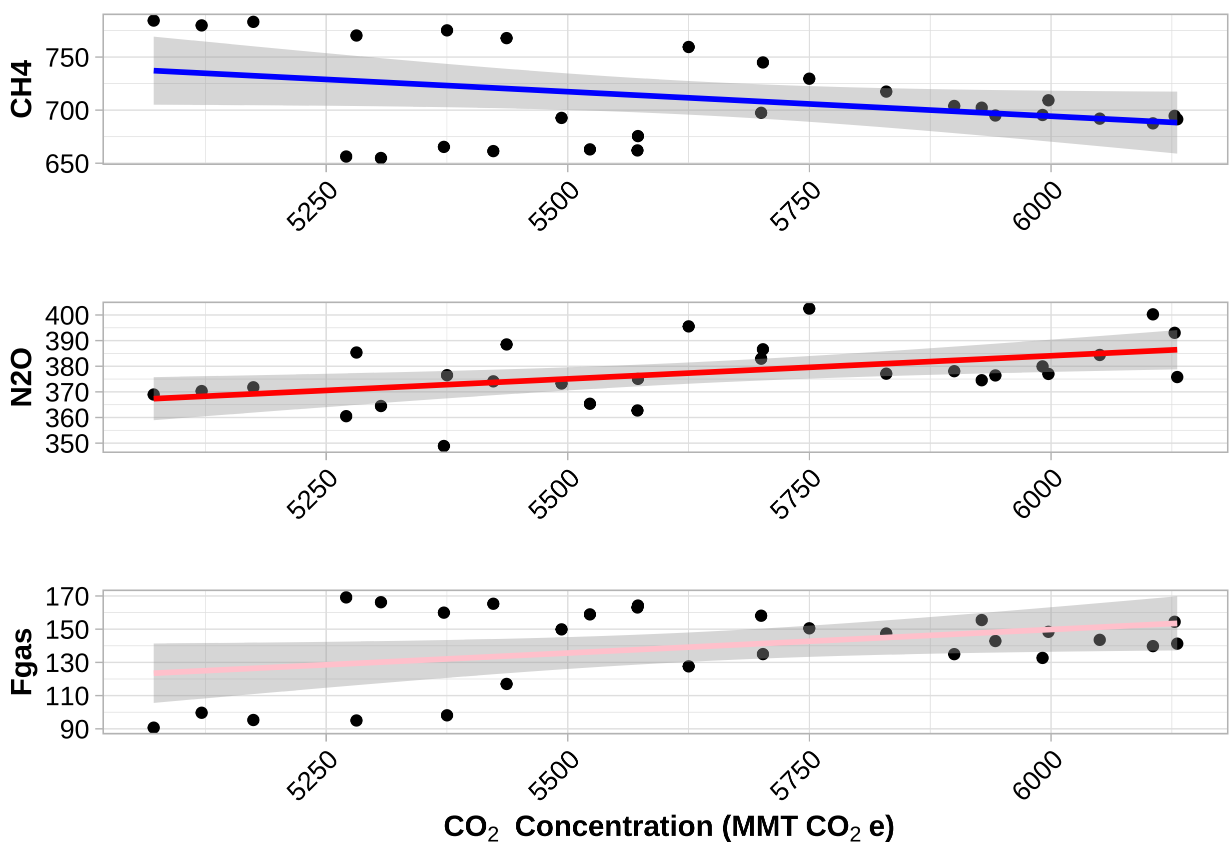

Figure 7: Correlationships between CO2 and other GHGs

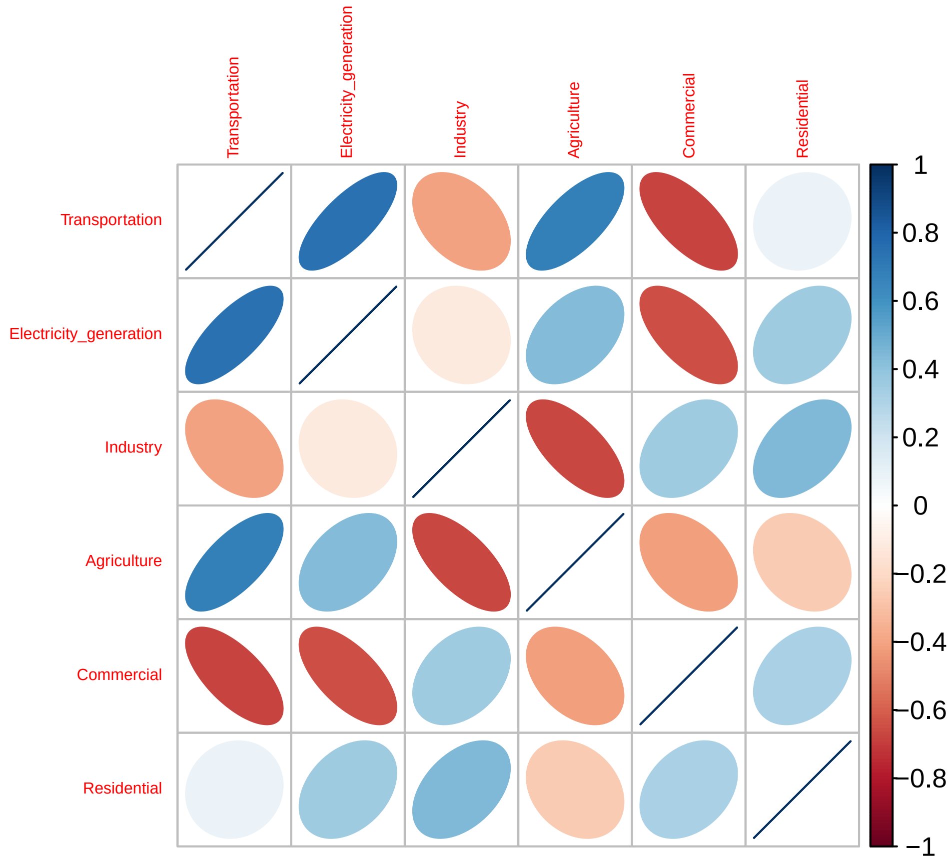

Figure 8: Correlation Plot of US GHG Emission by Sector

Similarly, multiple Linear Regression was also conducted among sectoral GHG emissions to see which sector was significant to electricity generation sector, which had the highest GHG emissions.

The result illustarted that multiple sectors had significant relationships with electricity generation sector (linear regression; R2=0.8566, dF = 22, $p < 0.0001$). Specifically, emission changes in agriculture, commercial, and residential sectors were significantly related to emission changes in electricity generation sector, while changes in transporation and industry sectors were not significantly related to electricity generation emissions. This result was also proved by AIC test that best model would be three significant explanatory variables.

The correlation plot in Figure 8 shows some interesting results of covariances among expanatory variables. Transporation sector was more positively correlated to electricity generation than agriculture and residential sectors. Industry sector had litte correlation with electricity generation while commercial was negatively correlated.

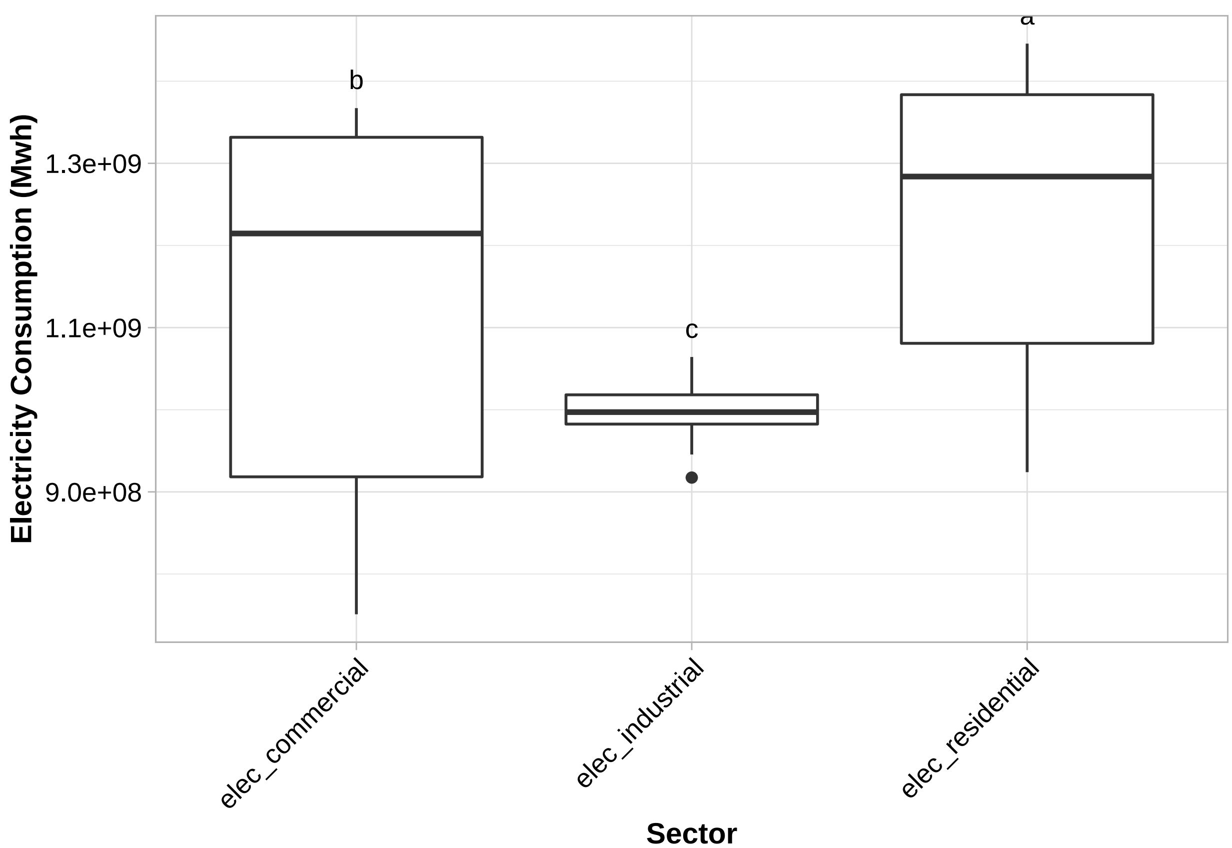

Figure 9: Post-hoc Test Plot - Pairwise Relationship between Emissions and Sectors

4.1.2 Did total electricity consumption differ among sectors?

The first impact factor on GHG emissions to study is the domestic electricity consumption. As illustrated in the previous section, there was no general consumption trend among sectors. Therefore, a detailed exploration on electricity consumption dataset was first conducted. Specifically, this section addresses on if electricity consumptions were significantly different among various sectors.

First, all data were gathered by sector and tested for the normality. The normal Q-Q plot shows that the emissions were normally distributed. Then, one-way Anova was tested, reflecting that electricity consumptions were significantly differed among various sectors (ANOVA, R2=0.2598, dF = 81, $p < 0.0001$). A post-hoc test was then run to look at the pairwise differences, and the result was shown in Figure 9, which illustrates that three sectors were significantly different from each other.

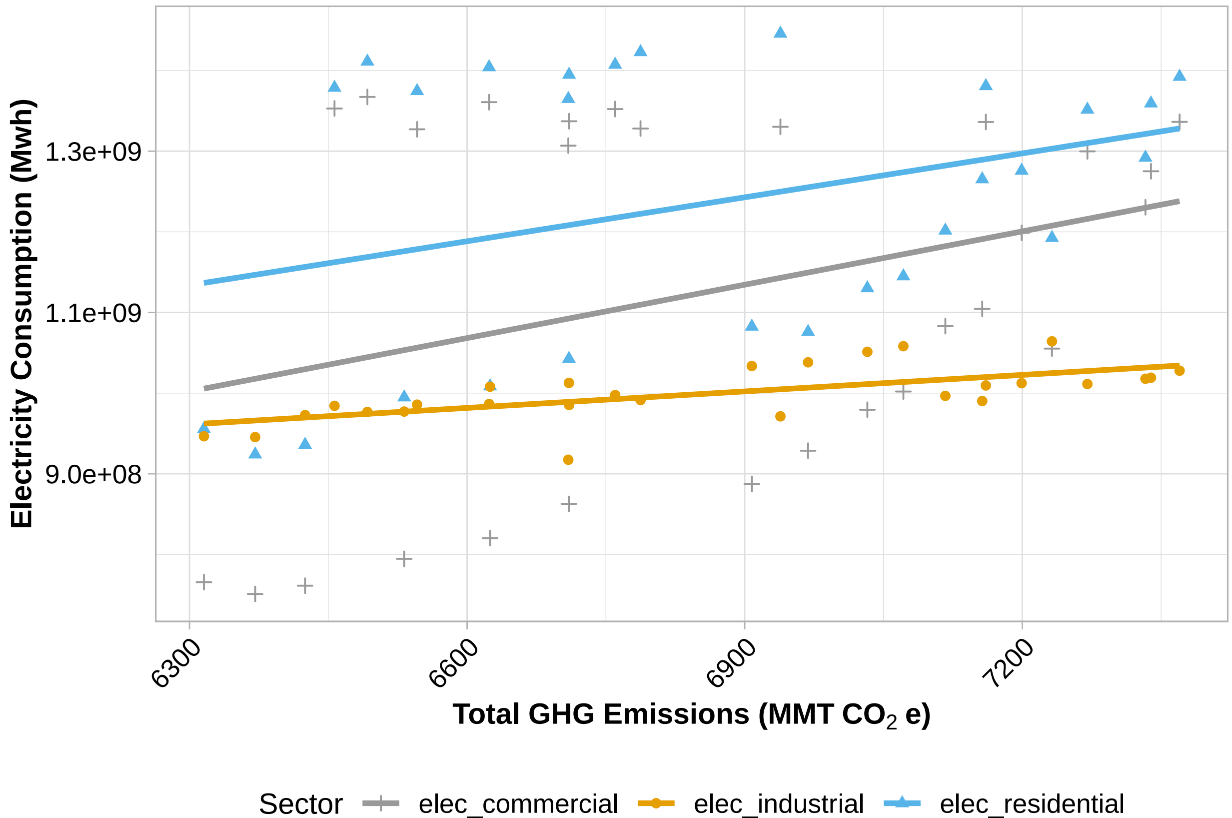

Figure 10: Relationships between Total GHG Emissions and Electricity Consumptions

4.1.3 Were electricity comsupmtions significant to total GHG emissions?

After looking at two datasets themselves, this section starts to answering the first question mentioned in the introduction - if the electricity consumption was significant to total GHG emissions. A multiple linear regression was run to see the relationship. The result showed that there was no significant relationship between electricity consumption and total GHG concentrations (linear regression, R2=0.2623, dF = 24, p = 0.05896), among which elec_industrial sector was the only significant variable. Since the p-value was close to 0.05 significane level, the AIC test was conducted to see the better model. The improved model was the test between the total GHG emission and electricity consumption in industrial sector (Step: AIC=321.5), and the result from the linear regression also indicated the significant relationship between them (linear regression, R2=0.1774, dF = 26, p = 0.0214).

The relationships of three explanatroy data with total GHG concentration were illustrated in Figure 10, in which commerncial electricity consumption data and residential electricity consumption data are away from the respective regression line, which results in the insignificant relationships with total GHG concentration.

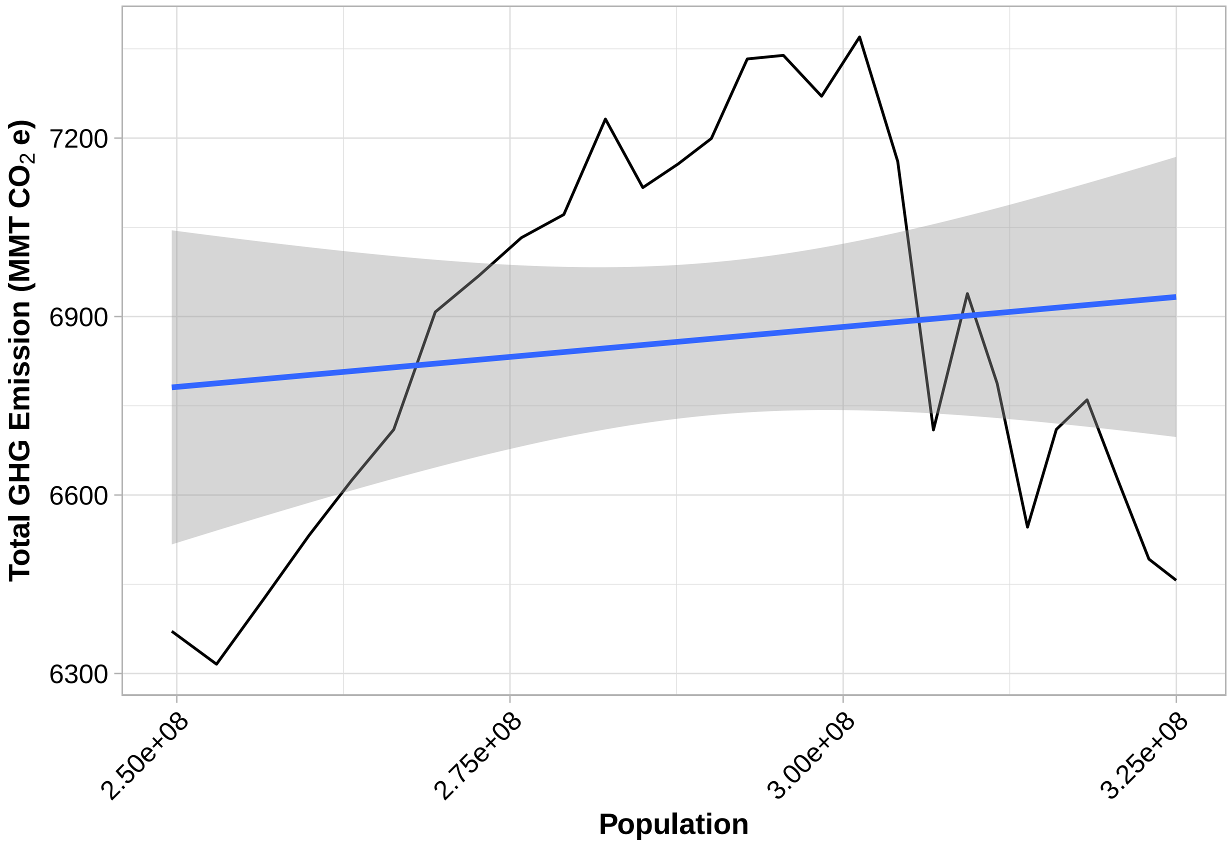

Figure 11: Relationship between Population and Total GHG Emissions

4.1.4 Is population growth significant to total GHG concentration?

Another factor that might affect GHG concentration attributed to human activities is population growth. This question was inspired by observing from the increasing trends of both GHG emissions and population. The shapiro test showed that the population data was normally distributed (W = 0.95398, p-value = 0.249). Then, a simple linear regression was conducted, which illustated there was no significant relationship between total GHG emissions and population growth (linear regression, R2 = 0.0007, dF = 26, p = 0.8948). The relationship between these two variables was displayed in Figure 11, which is not in a linear relationship.

4.2 Question 2: What were impacts of GHG emissions in these 28 years?

4.2.1 Which sectors of GHG emissions were significant to temperature anomaly?

Temperature is one of the most explicit impacts of GHG emissions. Instead of looking at sepcific types of GHGs, this project addresses on the impacts of sectoral emissions on the temperature change, and aims to figure out which sector was significant to the temperature.

The first step to achieve such goal was to study the chaning trend of temperature, specifically conducting a timer series analysis to see how the temperature had change over the 28-year period in US. The result showed that there was no significant trend of temperature change over the period of 1990-2017 in US (Mann-Kendall trend test z=1.758, p-value = 0.07869).

Then, the shapiro.test was conducted to confirm the normality of temperature anomaly data for the preparation of the linear regression (W = 0.98866, p-value = 0.9863). The result from the multiple linear regression illustarted that there was a significant relationship between temperature anomaly and sectoral emissions (linear regression, R2 = 0.8219, dF = 21, $p < 0.0001$). Specifically, emissions from transporation sector, agriculture sector, and residential sector were significant to such change. An AIC test were computed to improve the model to best illustrate explanatory variables in such a relationship, and the improved model included transportation, agiculture, commercial, and residential sectors (Step: AIC=-42.11). The multiple linear regression result proved that all of these sectoral emissions were significant to temperature anomaly (linear regression, R2 = 0.816, dF = 23, $p < 0.0001$).

A correlation plot was then created to visualize the correlationship and covariance between these variables. As shown in Figure 12, residential emissions and commercial emissions were mroe correlated to temperature anomaly changes in negative trends, while transportation emission and agriculture emissions were correlated to temperature anomaly in positive trends.

4.2.2 Was total GHG emissions significant to GDP growth?

As most GHGs were emitted by human activities, were such emissions significant to economic development? In this proejct, GDP was selected as the parameter to assess economic development. A similar time series analysis was conducted to assess if GDP had change over the study period in US. The result showed that there was a significant increasing trend of GDP change over the period of 1990-2017 in US (Mann-Kendall trend test z= 5.05, p-value < 0.0001).

GDP data was normally distributed according to shapio.test (W = 0.9486, p-value = 0.1827). However, there was no significant relationship between total GHG emissions and GDP growth over the time (linear regression, R2 = 0.0004, dF = 26, p = 0.7499). Such result was visualized in Figure 13, which illustrates that variables were dispersed away from the linear trend line.

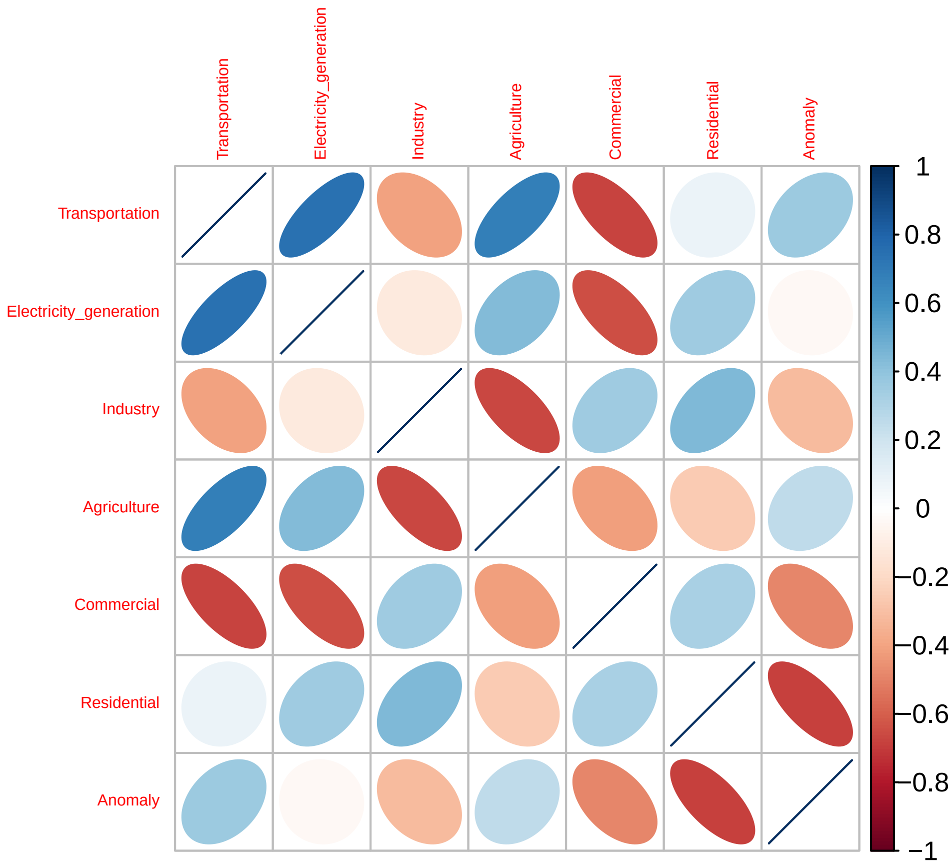

Figure 12: Correlation Plot between US Mean Annual Temperature Anomaly and Sectoral GHG Emissions

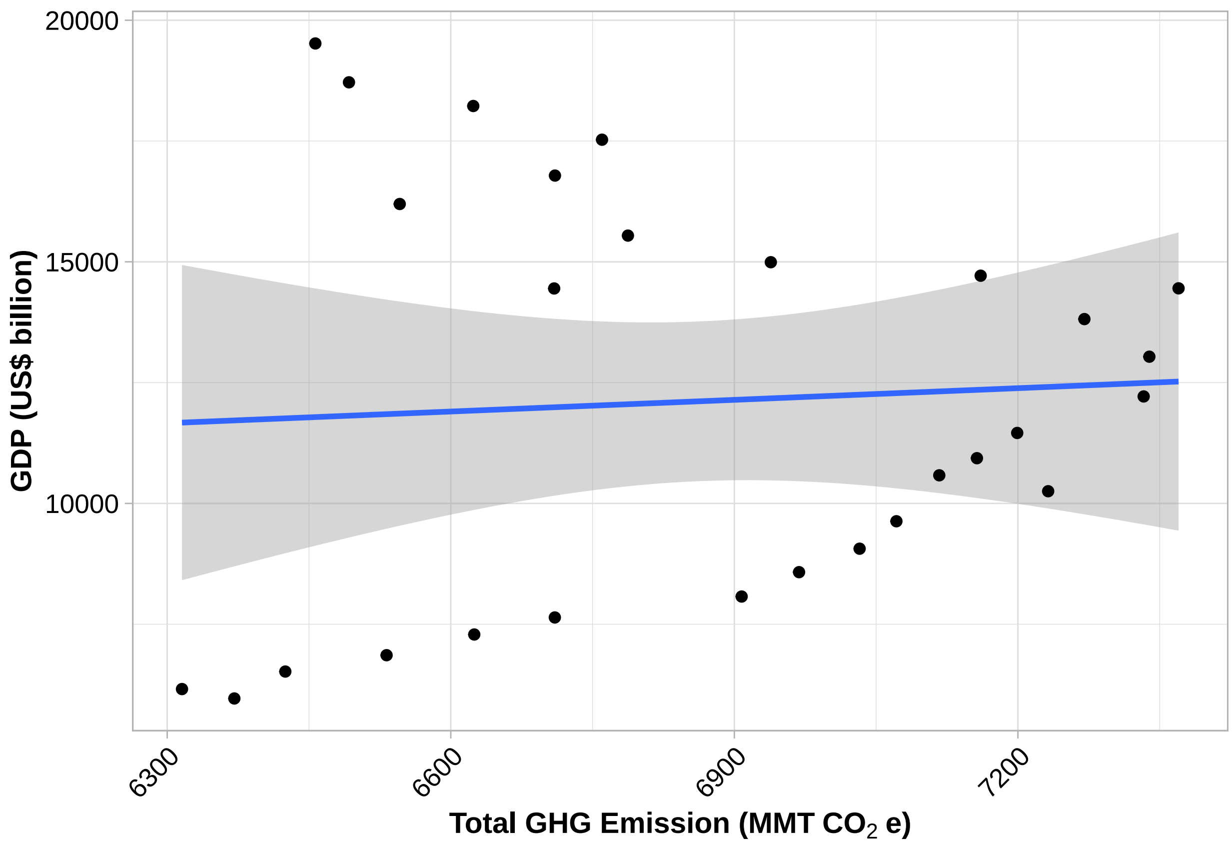

Figure 13: Relationship between US GDP and Total GHG Emissions

5 Summary and Conclusions

As Global Warming is a more server problem, GHG emissions as one of the most responsible factors are addressing more people's attention. This project studies the factors that might contribute to GHG emissions and also the potential impacts of GHG emissions in US from 1990-2017.

The first part of the project looked at the electricity consumption of which sector is significant to total GHG emissions and examined if population growth is significant to total GHG emission during this period. The tests from linear regression revealed that industrial electricity consumption was significant to total GHG emissions, and there was no significant relationship between population growth and total GHG emissions. In this section, the relationships among different types of GHGs as well as the relationships among sectoral emissions were also explored. By running different types of generalized linear models, the results illustrated that in 1990-2017:

1. CH4 and N2O emissions were significantly related to CO2 emissions;

2. CH4 and Fgas were strongly correlated;

3. Emissions in agriculture, commercial, and residential sectors were significantly related to emissions in electricity generation sector;

4. Electricity consumptions were significantly different among various sectors.

The second part analyzed the potential impacts of GHG emissions, specifically referring to temperature anomaly changes and GDP growth. Time series analysis was first conducted to evaluate how these two variables have changed over time, and the results demonstrated that there was no significant trend of temperature, and there was a significant increasing trend of GDP over 28 years. For temperature change, emissions in transportation, agiculture, commercial, and residential sectors were most significant to temperature anomalies; however, there was no significant relationship between total GHG emissions and GDP growth over the period.

This project is still a prelinary research on analyzing socioeconomic factors and impacts GHG emissions with many limitations. For example, the temporal range - 28 years - is short. Such factor was restricted by the data availability, while the GHG emissions always have a lag effects; that says, the impacts of GHG emissions, especially on temperature change, will not be revealed immediately. Also, the study on temperature change should be considered at a long-term scale, i.e. from pre-industrial period. Also, the geographic scale is also too broad. Different states in US have their only regulatories on GHG emissions, such as California as a leader on reducing GHG emissions over a long time. The analysis at the national level cannot illustrte the progress of policy inplementation. However, this project is still helpful to learn from the past about the causes and consequences of GHG emissions, as a reference as well as a warning to decision markers to think about the environmental impacts of emissions in the future.