# DATA 607 Fall 2021 Final Project-Energy Forecast

## Mark Schmalfeld

## 10/31/2021

```
## -- Attaching packages --------------------------------------- tidyverse 1.3.1 --

## v ggplot2 3.3.5     v purrr   0.3.4
## v tibble  3.1.6     v dplyr   1.0.7
## v tidyr   1.1.4     v stringr 1.4.0
## v readr   2.0.1     v forcats 0.5.1


## -- Conflicts ------------------------------------------- tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()


## Loading required package: airports


## Loading required package: cherryblossom


## Loading required package: usdata


##
## Attaching package: 'RCurl'


## The following object is masked from 'package:tidyr':
##
##     complete
```

## Energy and Carbon Emissions data upload

**Use the US Energy Information Administration (EIA) at www.eia.gov.international to collect
following**

**Collect data on a world and regional level**

**Pair down the time series data for the review at the global level**

1) Population dataset
2) GDP
3) Carbon Emissions data
4) Energy production and demand
5) Energy use per capita

6) Energy use per GDP

Subset the data and rename to clean up the variable names. Create a data frame with import global varialbes in time series ( aprpox 40 year of annual data) Look at the glimpse of the dataset to see basic structure

```
URL<-"https://raw.githubusercontent.com/schmalmr/607Finalproject/main/EIAdata607FinalProject_Energy.csv"

df<-read.csv(URL)


dfglobal<-select(df,Year, World_Consumption_quadBTU, World_GDP_USDB, World_consume_Coa._quadBTU, World_c

dfglobal<-rename(dfglobal, gobal_zero_CO2_energy= World_consume_Nuclear..renewables_quadBTU)
dfglobal<-rename(dfglobal, global_population =World_pop)
dfglobal<-rename(dfglobal, global_CO2_fuels=Fossil_fuel_industry_emissions_GtCO2)
dfglobal<-rename(dfglobal, MBTU_per_GPD_Billions=World_1000BTU_GDPBUSD)
dfglobal<-rename(dfglobal, MBTU_per_person=World_MMBtu_person)
dfglobal<-rename(dfglobal, global_energy_use_quad_BTU= World_Consumption_quadBTU)
dfglobal<-rename(dfglobal, global_Coal_use_quad_BTU=World_consume_Coa._quadBTU)
dfglobal<-rename(dfglobal, global_NGAS_use_quad_BTU=World_consume_Natural.gas_quadBTU)
dfglobal<-rename(dfglobal, global_Oil_use_quad_BTU=World_consume_Petroleum_quadBTU)
dfglobal<-rename(dfglobal, global_fuelandlandland_use_GtCO2=Fossil.fuel_plus_landuse_emissions_GtCO2)
dfglobal<-rename(dfglobal, global_land_use_GtCO2=Land_use_emissions_GtCO2)

glimpse (dfglobal)
```

```
## Rows: 41
## Columns: 13
## $ Year                             <int> 1980, 1981, 1982, 1983, 1984, 1985, 1~
## $ global_energy_use_quad_BTU       <dbl> 292.8999, 289.4015, 289.6913, 293.731~
## $ World_GDP_USDB                   <dbl> 27745.48, 28640.20, 28776.47, 29593.0~
## $ global_Coal_use_quad_BTU         <dbl> 78.65628, 78.96407, 80.42250, 82.6310~
## $ global_NGAS_use_quad_BTU         <dbl> 53.86522, 53.89542, 54.05526, 55.1497~
## $ global_Oil_use_quad_BTU          <dbl> 132.0640, 126.8111, 123.9243, 122.215~
## $ gobal_zero_CO2_energy            <dbl> 28.31441, 29.73088, 31.28923, 33.7353~
## $ global_population                <dbl> 4298127, 4377060, 4456830, 4537794, 4~
## $ MBTU_per_person                  <dbl> 68.14595, 66.11779, 64.99940, 64.7300~
## $ global_fuelandlandland_use_GtCO2 <dbl> 23636499800, 23341806580, 23109445670~
## $ global_land_use_GtCO2            <dbl> 4143886286, 4314896401, 4233685392, 4~
## $ global_CO2_fuels                 <dbl> 19492613520, 19026910180, 18875760280~
## $ MBTU_per_GPD_Billions            <dbl> 10.556672, 10.104733, 10.066950, 9.92~
```

```
ggplot(data=df, aes(x=Year, y=World_Consumption_quadBTU))+geom_point(size=4)+geom_smooth(method=lm)+ggt
```

```
## `geom_smooth()` using formula 'y ~ x'
```

```
## Warning: Removed 1 rows containing non-finite values (stat_smooth).
```

```
## Warning: Removed 1 rows containing missing values (geom_point).
```

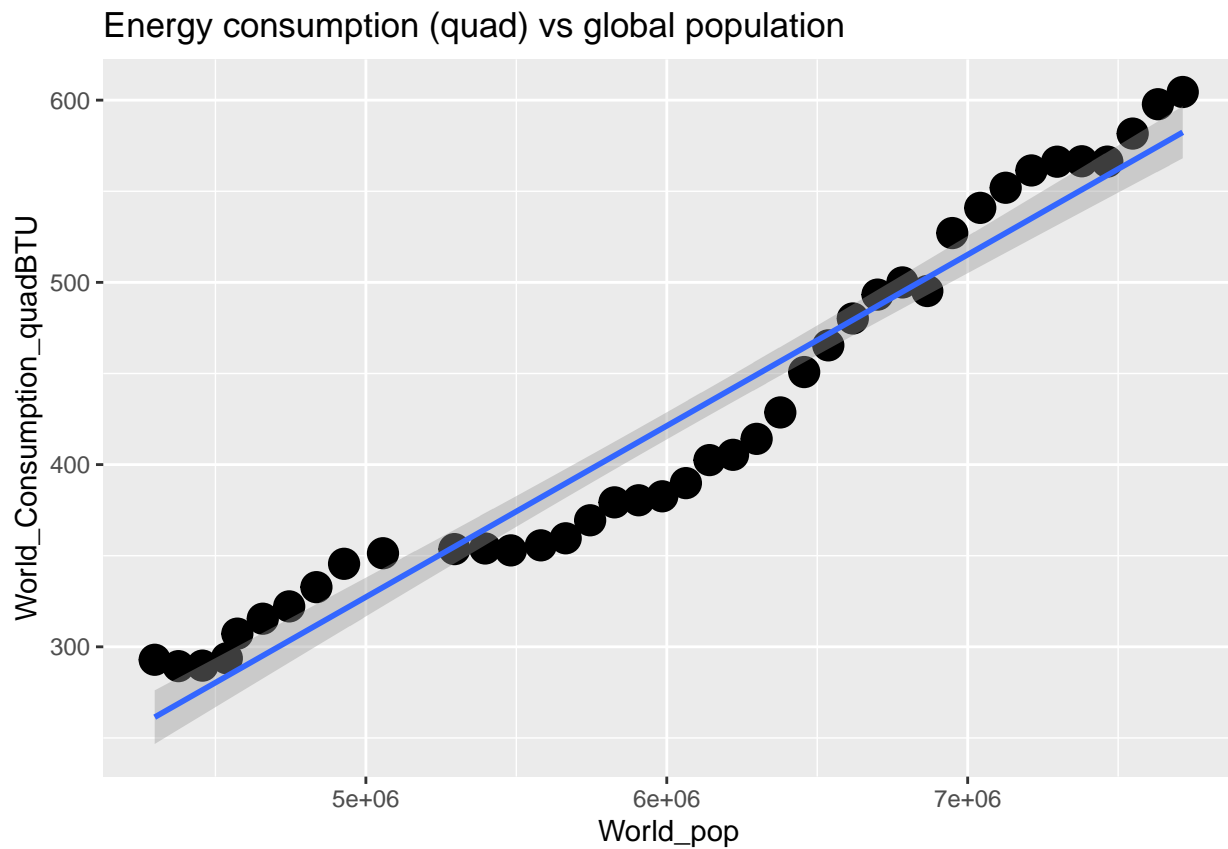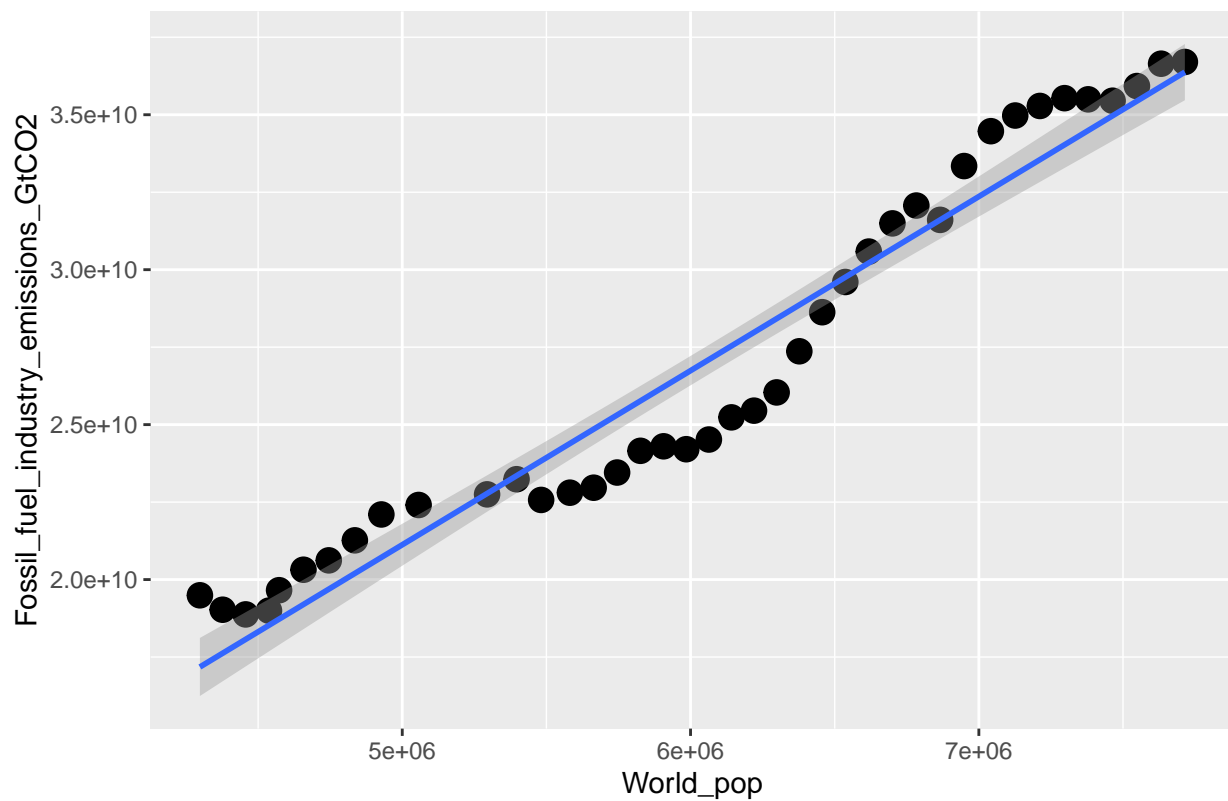## Energy consumption (quad) vs time
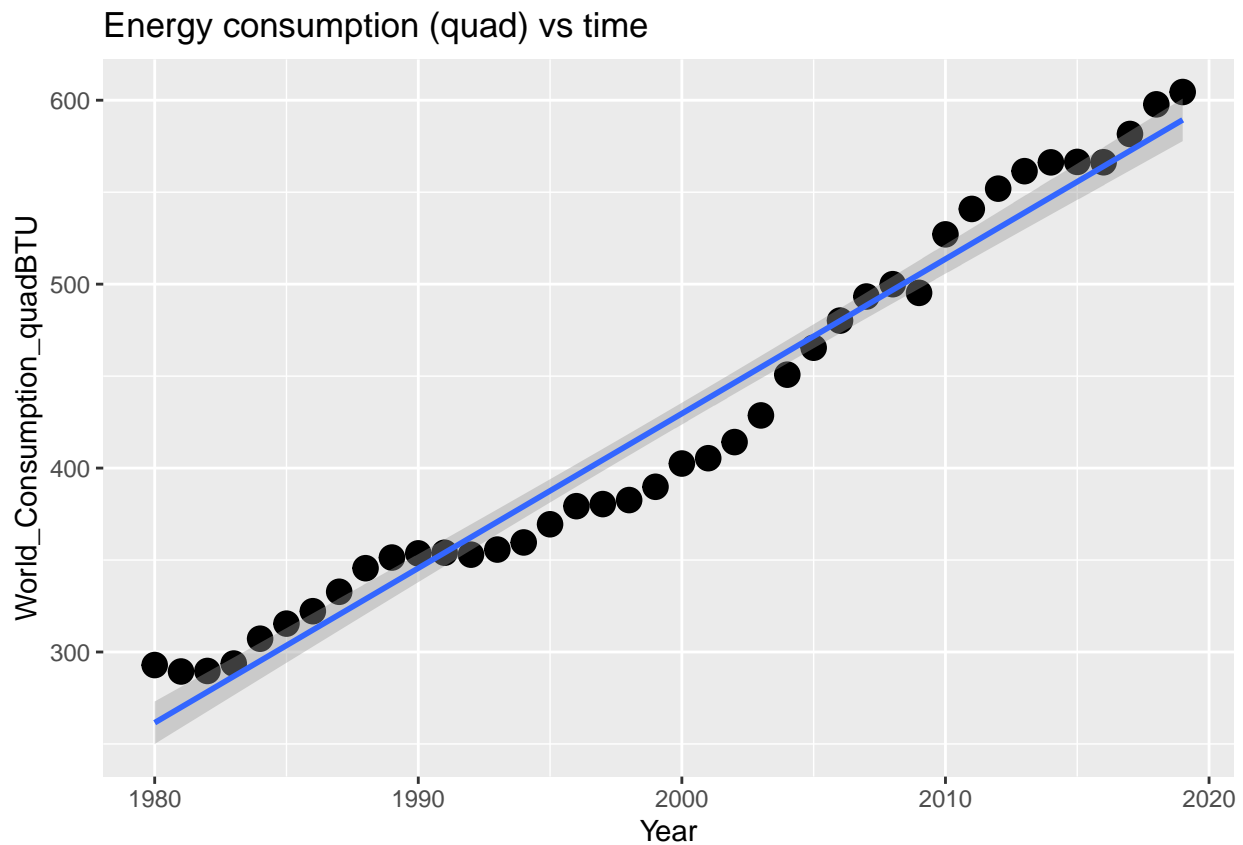


```
ggplot(data=df,aes(x=World_pop, y=World_Consumption_quadBTU))+geom_point(size=5)+geom_smooth(method=lm)
```

```
## `geom_smooth()` using formula 'y ~ x'
```

```
## Warning: Removed 1 rows containing non-finite values (stat_smooth).
```

```
## Warning: Removed 1 rows containing missing values (geom_point).
```

## Energy consumption (quad) vs global population



```
ggplot(data=df,aes(x=World_pop, y=Fossil_fuel_industry_emissions_GtCO2))+geom_point(size=4)+geom_smooth
```

```
## `geom_smooth()` using formula 'y ~ x'
```

```
## Warning: Removed 1 rows containing non-finite values (stat_smooth).
```

```
## Warning: Removed 1 rows containing missing values (geom_point).
```

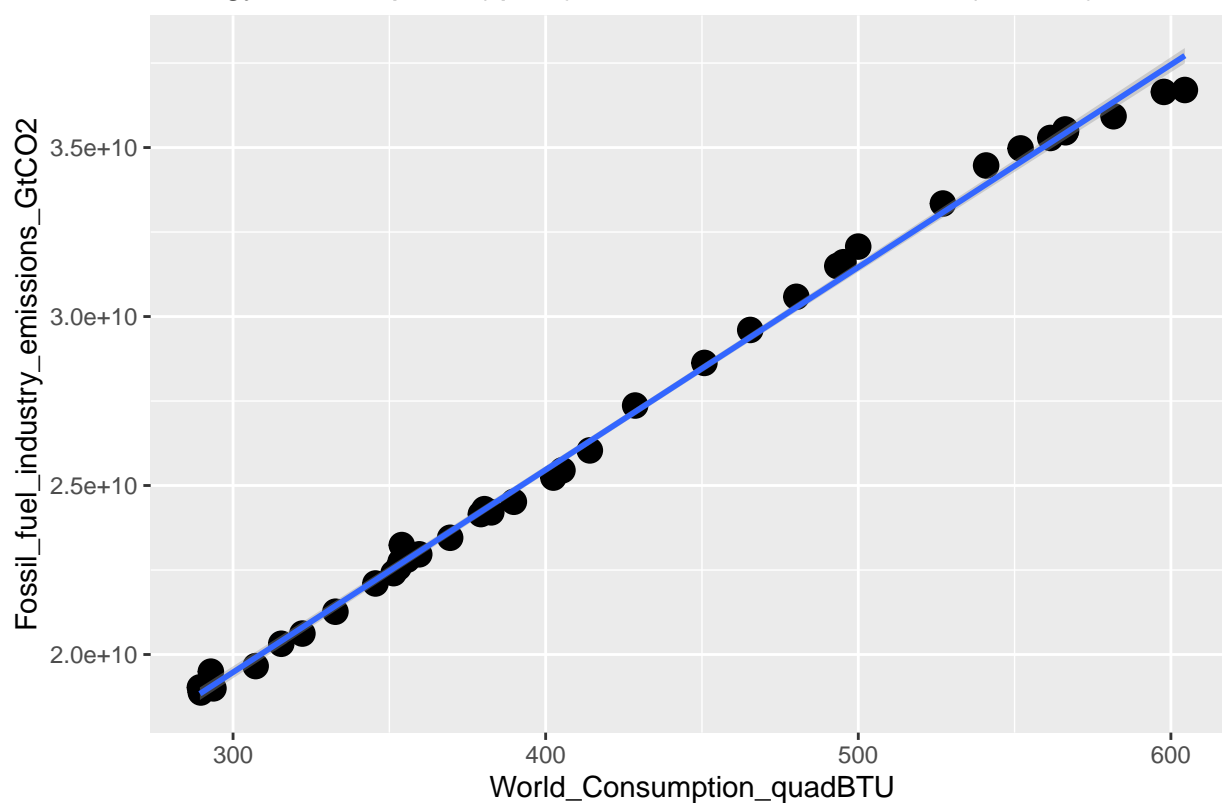## Fossile fuel carbon emission (GtCo2) vs world population



```
ggplot(data=df,aes(x=World_Consumption_quadBTU, y=Fossil_fuel_industry_emissions_GtCO2))+geom_point(size
```

```
## `geom_smooth()` using formula 'y ~ x'
```

```
## Warning: Removed 1 rows containing non-finite values (stat_smooth).
```

```
## Warning: Removed 1 rows containing missing values (geom_point).
```

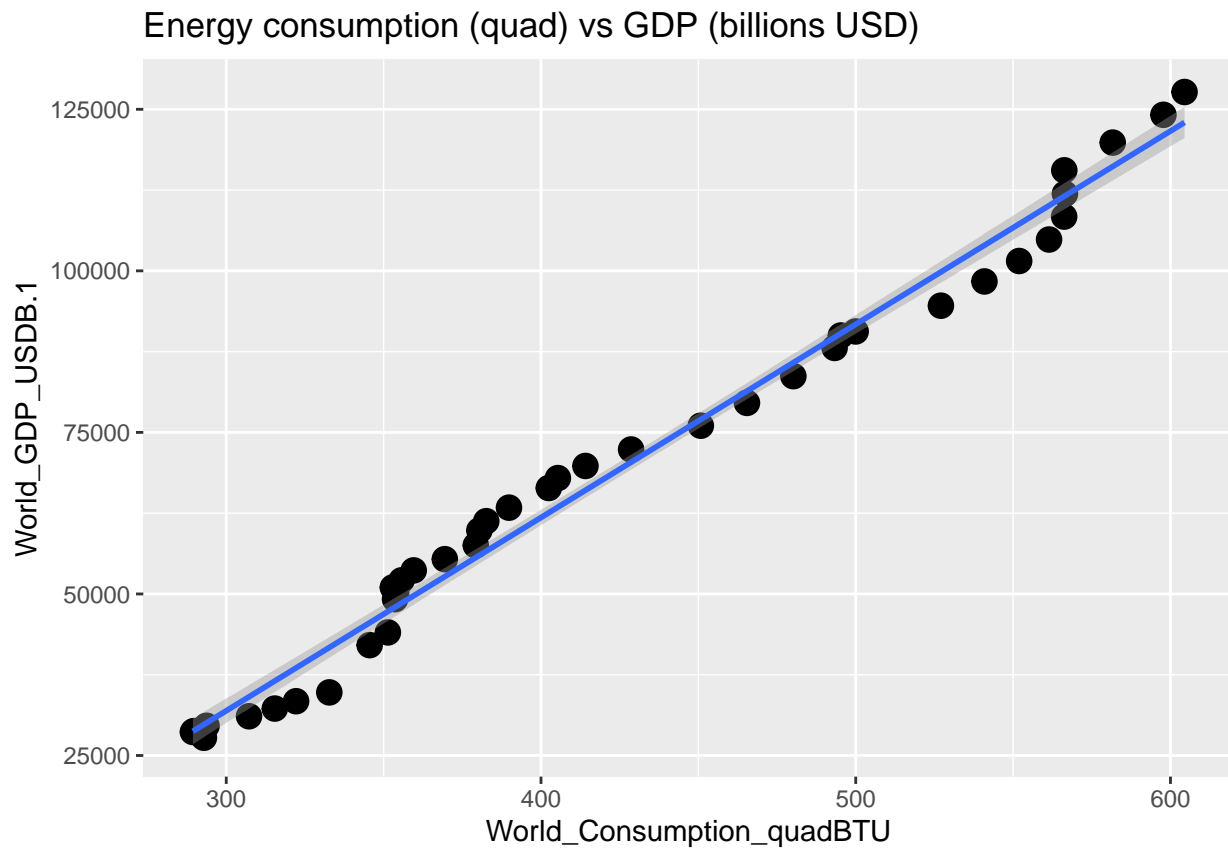Energy consumption (quad) vs fossile fuel emissions (GtCO2)

```
ggplot(data=df,aes(x=World_Consumption_quadBTU, y=World_GDP_USDB.1))+geom_point(size=4)+geom_smooth(metl
```

```
## `geom_smooth()` using formula 'y ~ x'
```

```
## Warning: Removed 1 rows containing non-finite values (stat_smooth).
```

```
## Warning: Removed 1 rows containing missing values (geom_point).
```

## Energy consumption (quad) vs GDP (billions USD)



```
ggplot(data=df,aes(x=Fossil_fuel_industry_emissions_GtCO2, y=World_GDP_USDB.1))+geom_point(size=4)+geom_
```

```
## `geom_smooth()` using formula 'y ~ x'
```

## Fossil fuel emissions (GtCO2) vs GPD (US $ Billions)
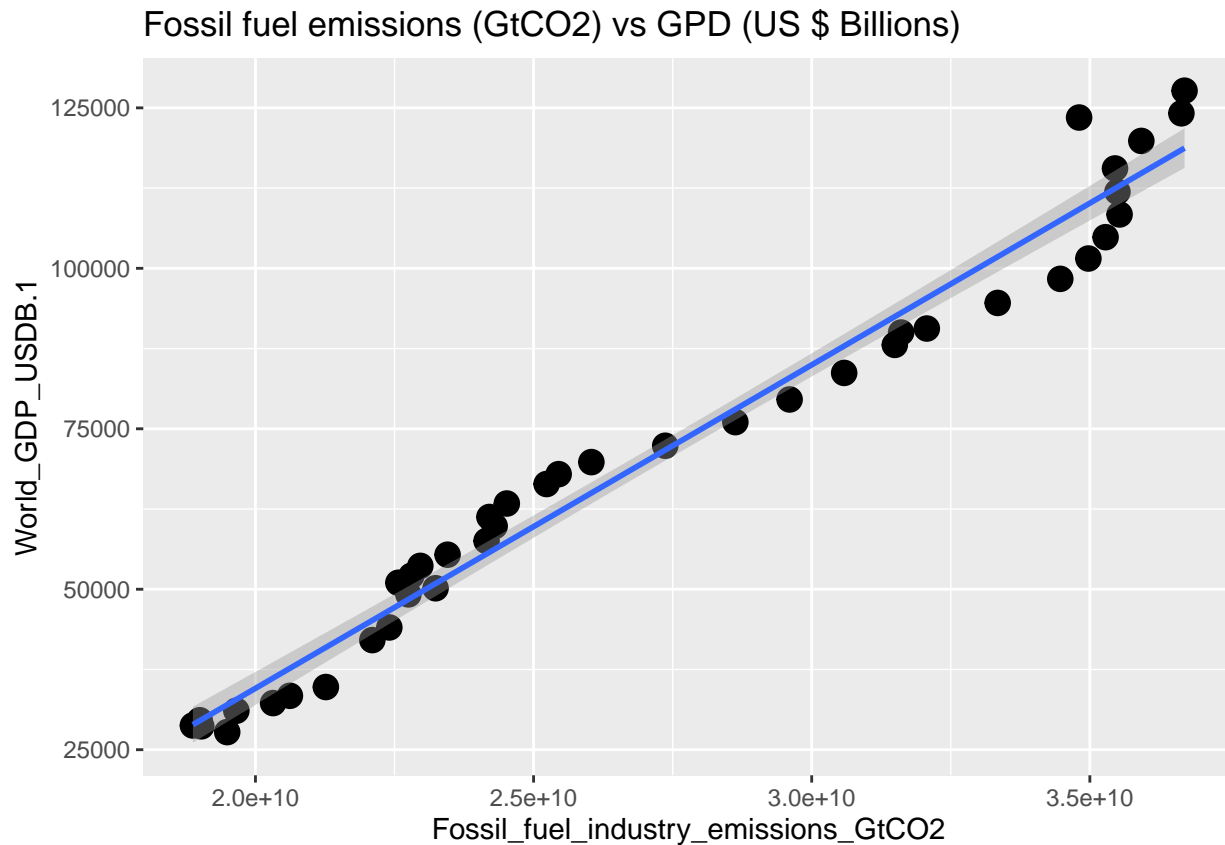


## Evaluate linear model of GPD vs year

```
m_gdp <- lm(World_GDP_USDB ~ Year, data = dfglobal)
summary(m_gdp)
```

```
##
## Call:
## lm(formula = World_GDP_USDB ~ Year, data = dfglobal)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -6004.0 -3035.4  -457.7  2302.6  8486.9
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept) -5.033e+06  1.098e+05  -45.85   <2e-16 ***
## Year         2.552e+03  5.489e+01   46.49   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 4008 on 38 degrees of freedom
##   (1 observation deleted due to missingness)
## Multiple R-squared:  0.9827, Adjusted R-squared:  0.9823
## F-statistic:  2161 on 1 and 38 DF,  p-value: < 2.2e-16
```

```
par(mfrow = c(2,2))
plot(m_gdp)
```
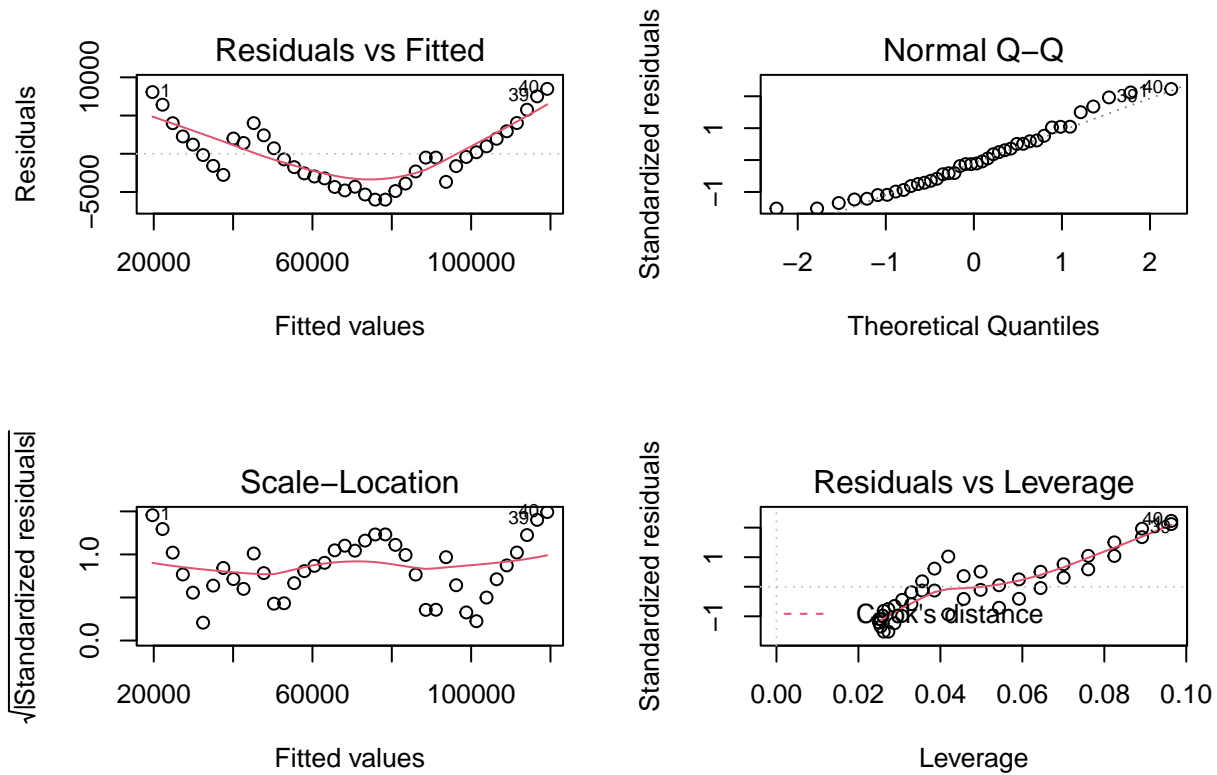


## Evaluate linear model of global population over time

```
m_population2 <- lm(global_population ~ Year, data = dfglobal)
summary(m_population2)
```

```
##
## Call:
## lm(formula = global_population ~ Year, data = dfglobal)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -106190  -41377  -18465   55885  111649
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept) -1.707e+08  1.827e+06  -93.39   <2e-16 ***
## Year         8.837e+04  9.138e+02   96.70   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 66720 on 38 degrees of freedom
##   (1 observation deleted due to missingness)
## Multiple R-squared:  0.996,  Adjusted R-squared:  0.9958
```

```
## F-statistic:  9351 on 1 and 38 DF,  p-value: < 2.2e-16
```

```
par(mfrow = c(2,2))
plot(m_population2)
```

### Residuals vs Fitted

### Normal Q–Q

### Scale–Location

### Residuals vs Leverage
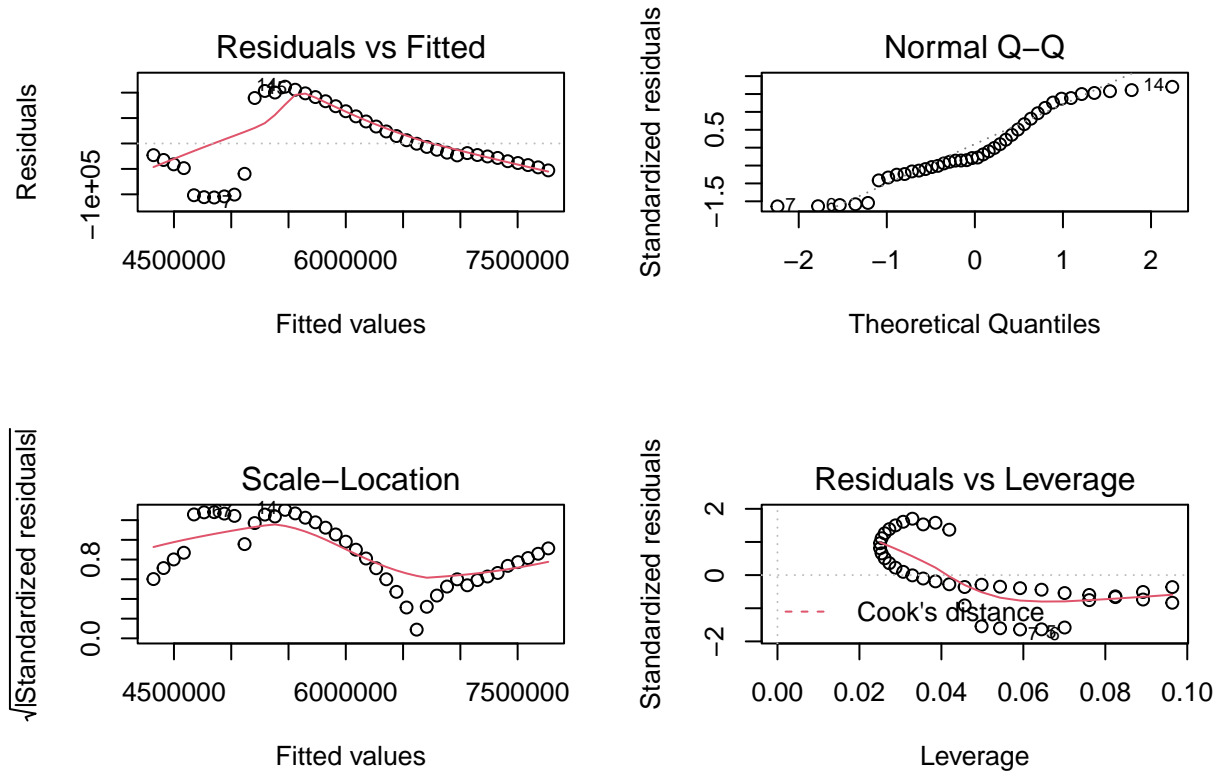
## Evalue Energy consumption total

**Model is good with the GPD and the Population.**

**Population has a p value of 0.06 (just a bit higher than desired)**

```
 m_energy <- lm(global_energy_use_quad_BTU ~World_GDP_USDB+global_population, data = dfglobal)
summary(m_energy)
```

**Population retained as a variable in the model**

```
##
## Call:
## lm(formula = global_energy_use_quad_BTU ~ World_GDP_USDB + global_population,
##     data = dfglobal)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -19.366  -9.582  -2.528   8.010  20.918
```

```
##
## Coefficients:
##                    Estimate Std. Error t value Pr(>|t|)
## (Intercept)        2.735e+02  4.034e+01    6.78 5.58e-08 ***
## World_GDP_USDB     4.004e-03  3.736e-04   10.72 6.71e-13 ***
## global_population -2.086e-05  1.086e-05   -1.92   0.0626 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 11.62 on 37 degrees of freedom
##   (1 observation deleted due to missingness)
## Multiple R-squared:  0.9872, Adjusted R-squared:  0.9865
## F-statistic:  1424 on 2 and 37 DF,  p-value: < 2.2e-16
```

```r
par(mfrow = c(2,2))
plot(m_energy)
```
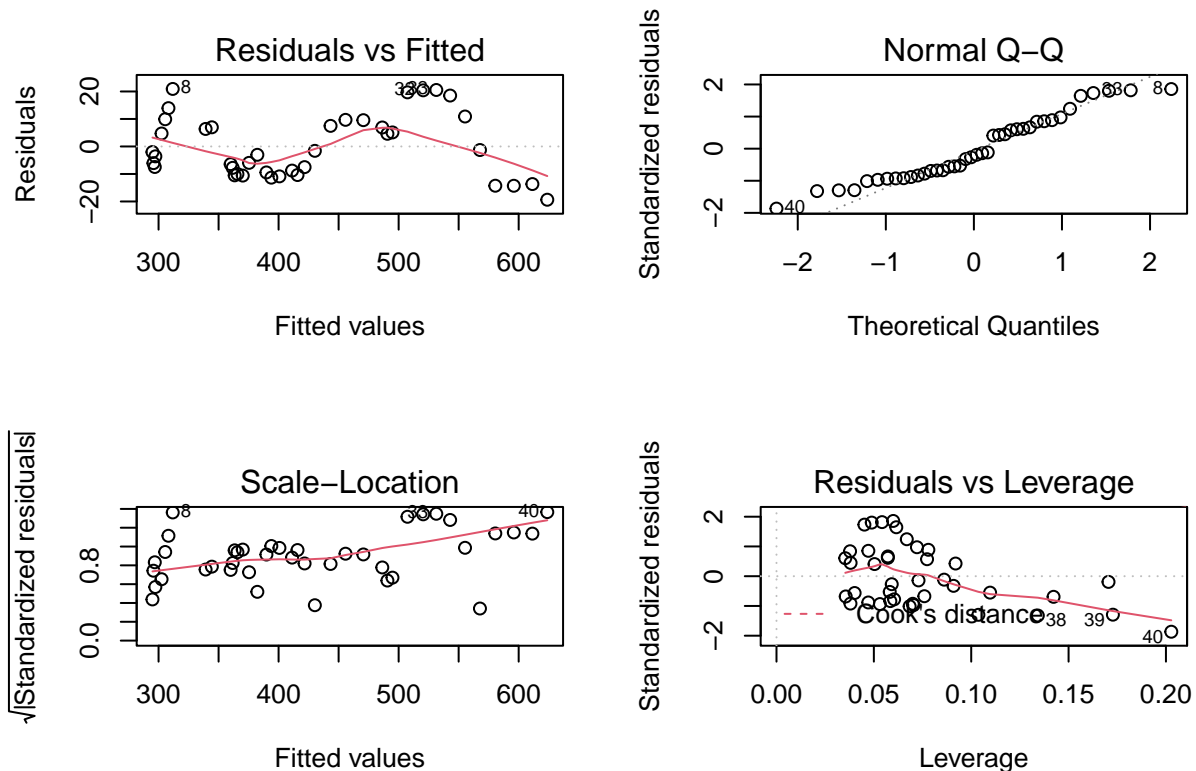


## Evaluate the Energy consumption not generating CO2

**Correlation is most strong with the Global GDP**

**GDP calculated from model above can be used to estimate model Energy**

consumption that is not generating CO2.

**Assumption for ease of model: Nuclear, renewables and water**

are net zero CO2 energy sources in this model All CO2 emissions are then attributed to fuel use.

```
 m_energyzero <- lm(gobal_zero_CO2_energy ~ World_GDP_USDB, data = dfglobal)
summary(m_energyzero)
```

```
##
## Call:
## lm(formula = gobal_zero_CO2_energy ~ World_GDP_USDB, data = dfglobal)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -6.2244 -1.3619  0.2599  1.8427  3.5481
##
## Coefficients:
##                 Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.902e+01  9.718e-01   19.58   <2e-16 ***
## World_GDP_USDB 5.592e-04  1.287e-05   43.46   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.418 on 38 degrees of freedom
##   (1 observation deleted due to missingness)
## Multiple R-squared:  0.9803, Adjusted R-squared:  0.9798
## F-statistic:  1889 on 1 and 38 DF,  p-value: < 2.2e-16
```

```
par(mfrow = c(2,2))
plot(m_energyzero)
```
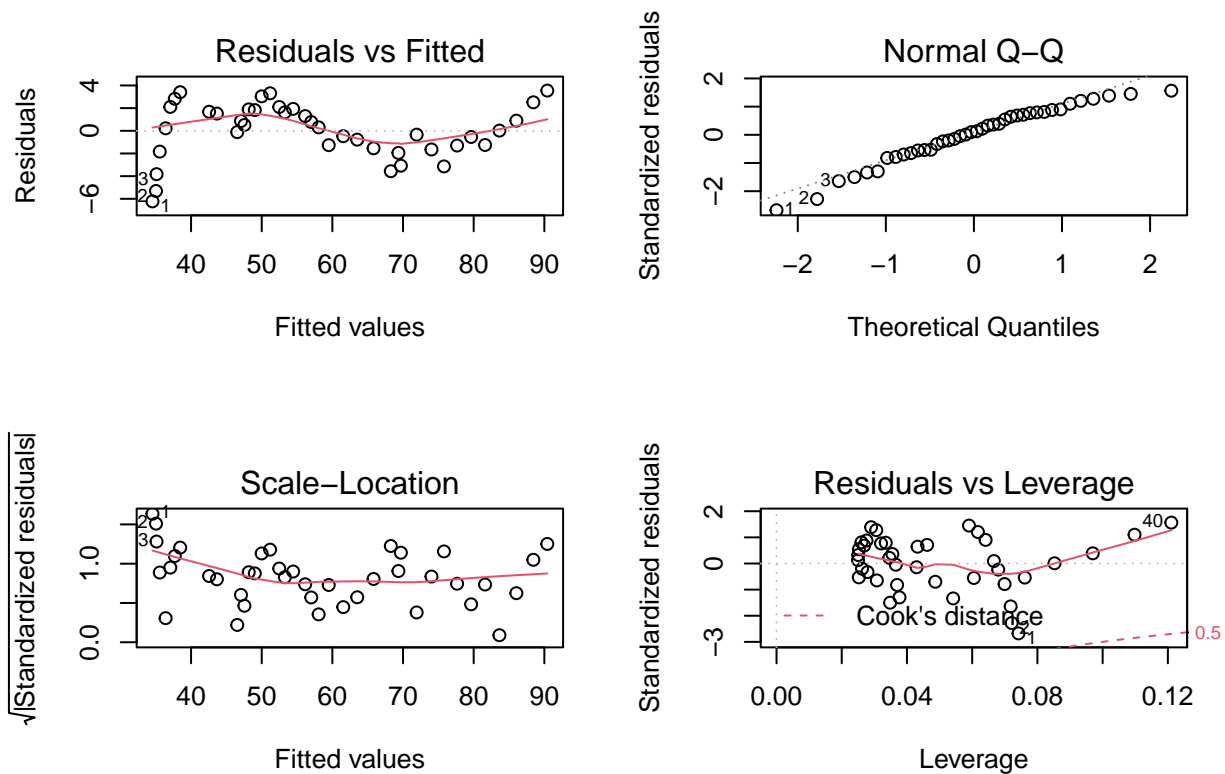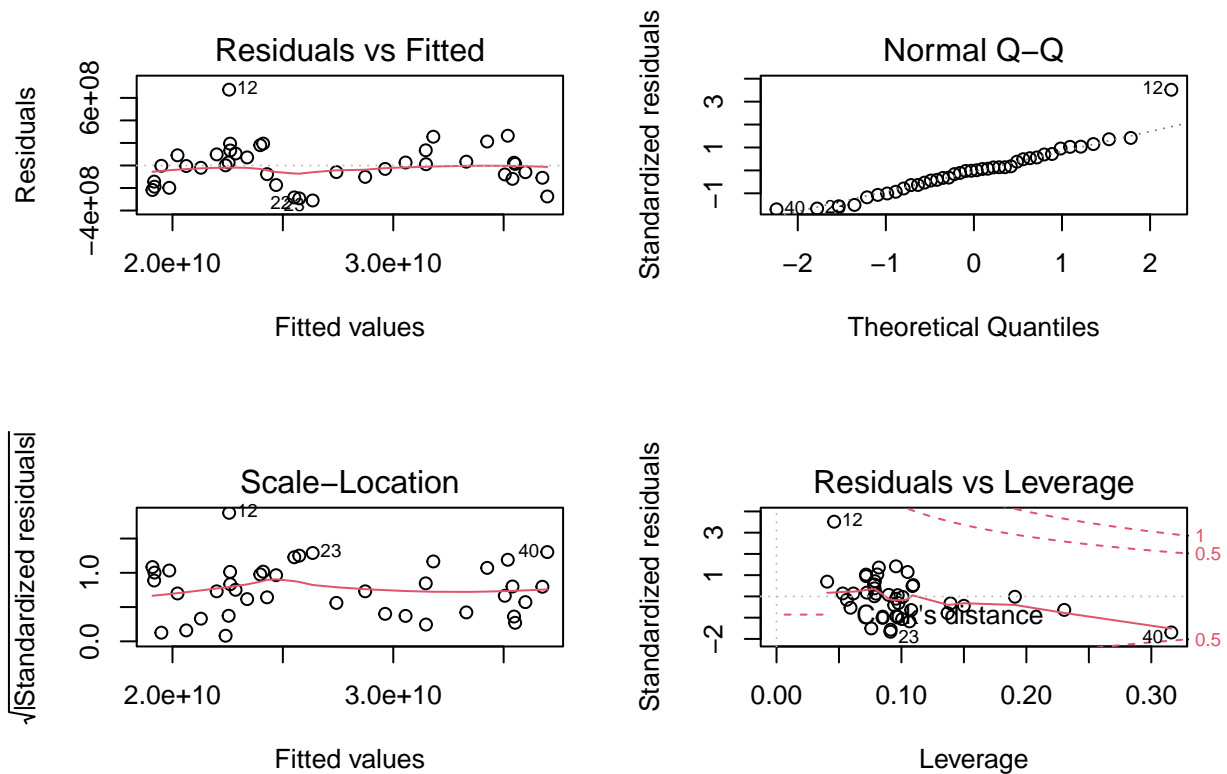


## Model of Carbon Emissions from Fuel over time

12

## Key variables for the model are the global energy use, global zero CO2 energy and the global population

**Note: For simplification the use of carbon capture technology is ignored** and it is assumed later as a simple application of the technology to remove carbon from the atmosphere. It is actually a very challenging and difficult problem to solve economically and with efficent use of energy.

```
m_co2 <- lm(global_CO2_fuels ~ global_energy_use_quad_BTU+gobal_zero_CO2_energy+global_population, data
summary(m_co2)
```

```
##
## Call:
## lm(formula = global_CO2_fuels ~ global_energy_use_quad_BTU +
##     gobal_zero_CO2_energy + global_population, data = dfglobal)
##
## Residuals:
##        Min         1Q     Median         3Q        Max
## -310518436 -111468218    -777360  100461163  672995576
##
## Coefficients:
##                             Estimate Std. Error t value Pr(>|t|)
## (Intercept)               -1.130e+09  4.345e+08  -2.601   0.0134 *
## global_energy_use_quad_BTU 6.910e+07  1.591e+06  43.443  < 2e-16 ***
## gobal_zero_CO2_energy     -1.009e+08  1.151e+07  -8.766 1.86e-10 ***
## global_population          7.547e+02  1.716e+02   4.398 9.29e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 195900000 on 36 degrees of freedom
##   (1 observation deleted due to missingness)
## Multiple R-squared:  0.999,  Adjusted R-squared:  0.9989
## F-statistic: 1.216e+04 on 3 and 36 DF,  p-value: < 2.2e-16
```

```
par(mfrow = c(2,2))
plot(m_co2)
```

Residuals vs Fitted

Residuals

6e+08
−4e+08

12
22 23

2.0e+10        3.0e+10

Fitted values

Normal Q–Q

Standardized residuals

3
1
−1

12

40 23

−2    −1    0    1    2

Theoretical Quantiles

Scale–Location

√|Standardized residuals|

1.0
0.0

12
23        40

2.0e+10        3.0e+10

Fitted values

Residuals vs Leverage

Standardized residuals

3
1
−2

12

Cook's distance
23                40

1
0.5

0.5

0.00   0.10   0.20   0.30

Leverage

## Modeling the future

**The past 40 years provide a dataset used to establish a frame set of models**

1) Models are noted to have some weaknesses in the models; however all of the models are very good at explaining the past. The R-squared values for the series of models is above 0.95 for all models (amazingly strong)

2) There are some indications of the residuals have a few outlyers in the models

- particularly noted in the Carbon Emissions model (some trends in each of models)

3) The Q-Q plots are pretty good for models except the population model over time which has more drift than others. Others do have some trends at the start and the tails of the plots.

4) Residuals are show good random variability for models except the

- population model seems to have poor randomness - with a distinct line
- GPD have drift in residuals up and down across the model

5) Major assumption for next step is to proceed to extrapolate the future with the current model.

- Extrapolation out until 2050
- Note the use of an extrapolation is not recommended and error expected
- Extrapolation judged a risk to consider based the very strong R-squared for each of the individual models.

6) Create an incremental estimate to demonstrate the magnitude of the incremental renewable or zero carbon generating energy development is needed.
Placeholder globally without specific constraints on population or GDP

7) The appraoch is to show how much more is needed to just hold Carbon emissions flat or near constant. The later adjustments show how difficult or significant the energy transition needed is to achieve reductions on Carbon emissions.

```
year <- c(1980,1990,2000,2010,2019, 2020, 2021,2022,2023,2024,2025,2026,2027,2028,2029,2030,2035,2040,20

df_model <- data.frame(year)

df_model<- df_model %>%mutate(GDP=(-5033000+2552*year),.after=year)
df_model<- df_model %>%mutate(population=(-170700000+88370*year),.after=GDP)
df_model<- df_model %>%mutate(energy_total=(273.5+0.004*(GDP)),.after=population)
df_model<- df_model %>%mutate(energy_zerocarbon=(19.02+0.0005592*(GDP)),.after=energy_total)
df_model<- df_model %>%mutate(globalcarbonfuels=(-1133000000+6.91E7*energy_total-100900000*energy_zeroca

df_model<- df_model %>%mutate(energywithcarbon=(energy_total-energy_zerocarbon),.after=energy_zerocarbon
df_model<- df_model %>%mutate(carbon_per_energy=(globalcarbonfuels/(energy_total-energy_zerocarbon)),.a

df_model$incremental_new_zero_carbonenergy <- c(0,0,0,0,0,20,35,55,75,95,115,150,170,180,190,195,250,500
df_model<-df_model %>% mutate(energy_new_zerocarbon=energy_zerocarbon+incremental_new_zero_carbonenergy)
df_model<-df_model %>% mutate(energy_new_withcarbon=(energy_total-(energy_new_zerocarbon)))
df_model<-df_model %>% mutate(new_global_carbonemissionsfuel=(globalcarbonfuels+72000000*(energy_zeroca
```
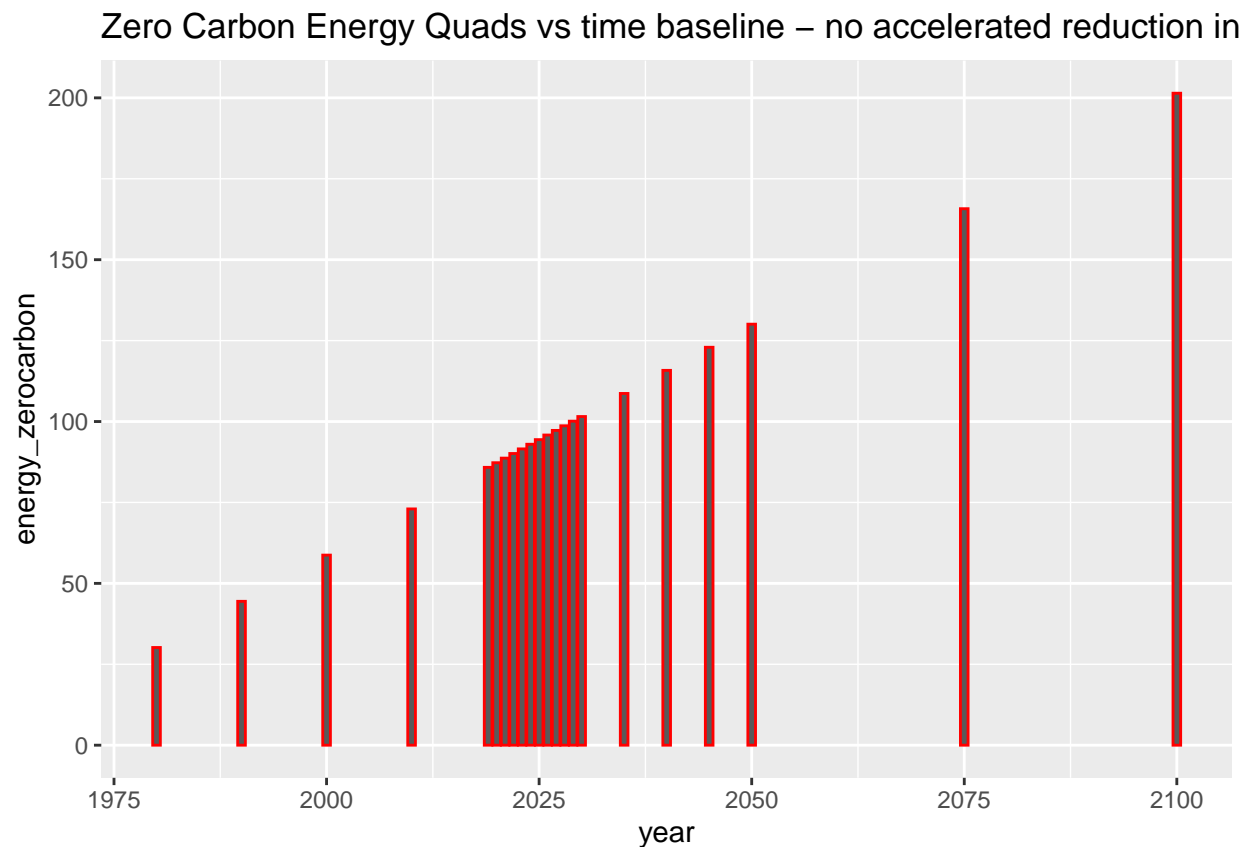
## Zero Carbon Energy Quads vs time baseline – no accelerated reduction in

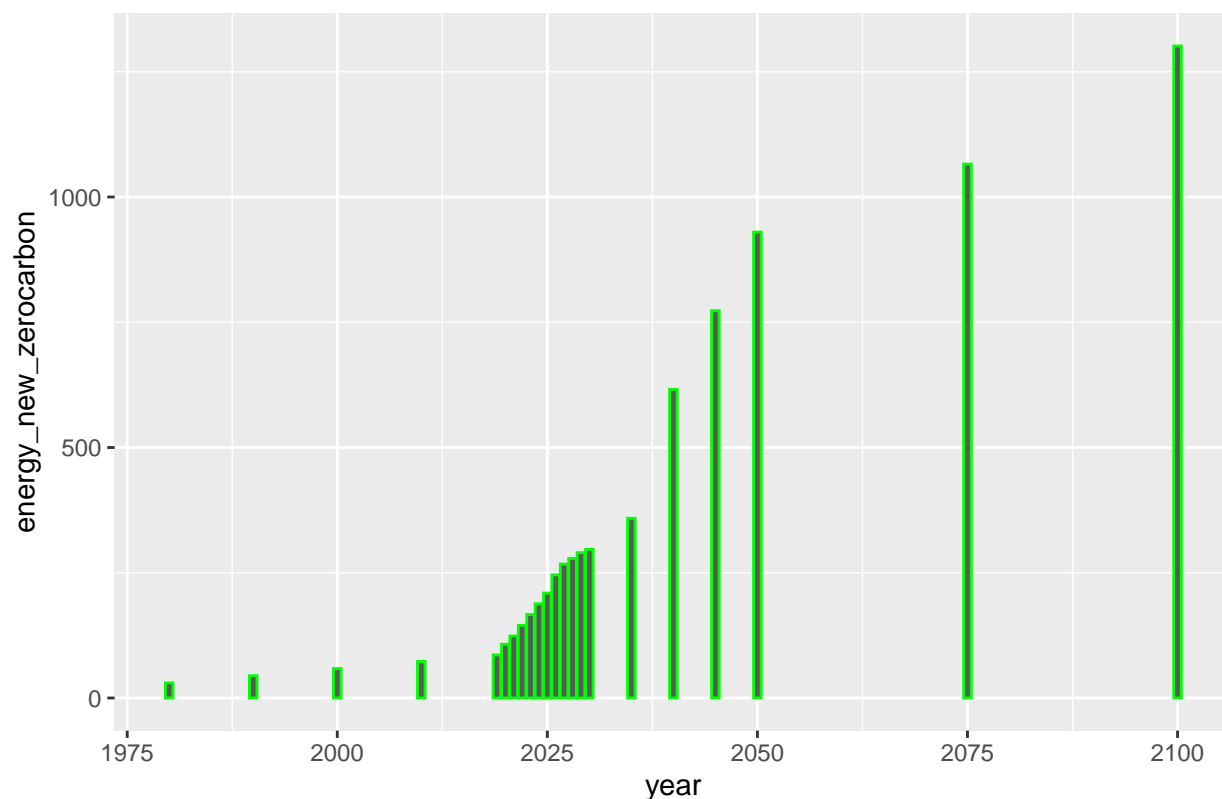## Energy demand forecast and associated carbon emission trend in progres
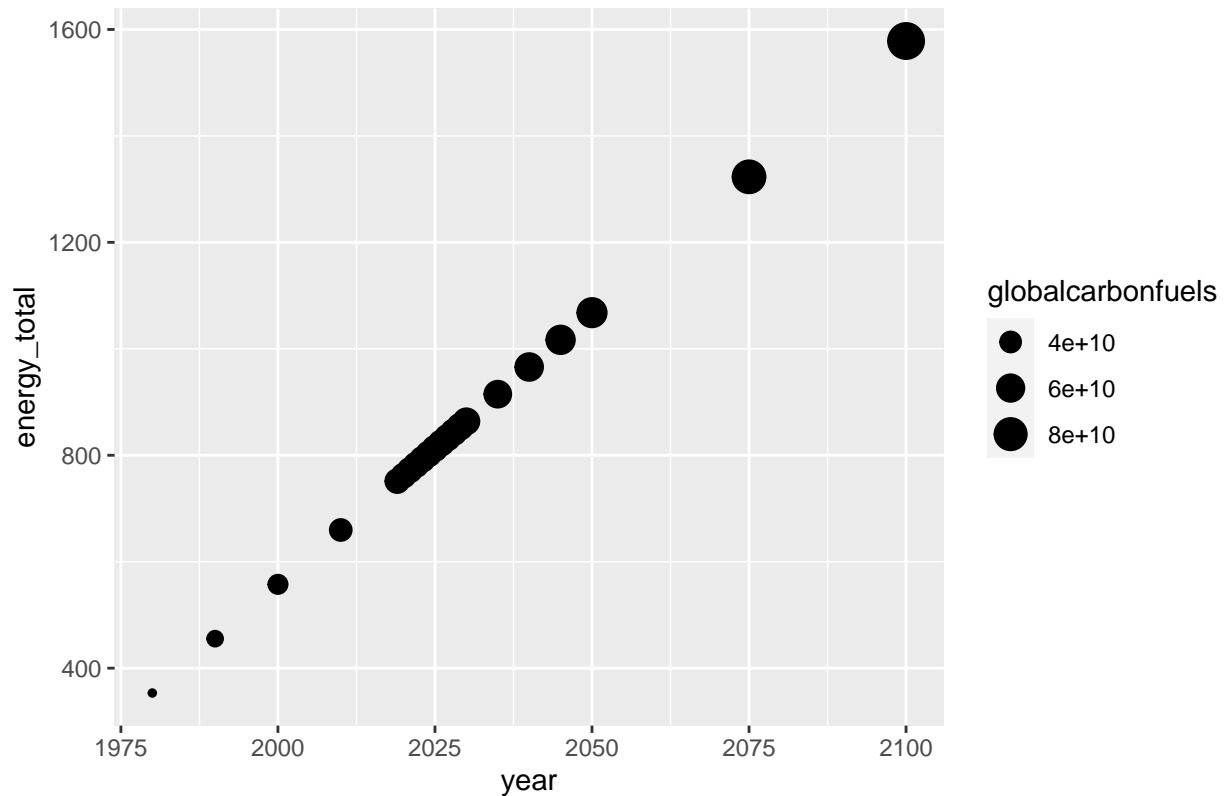


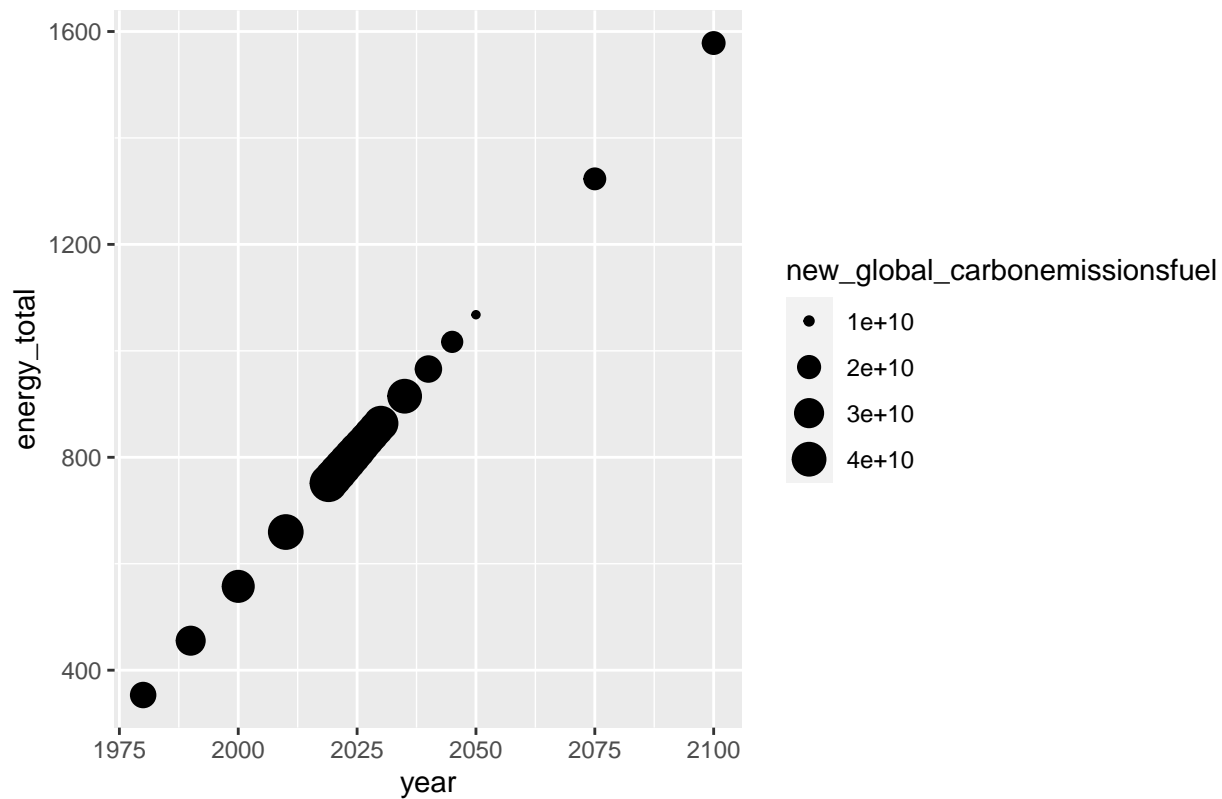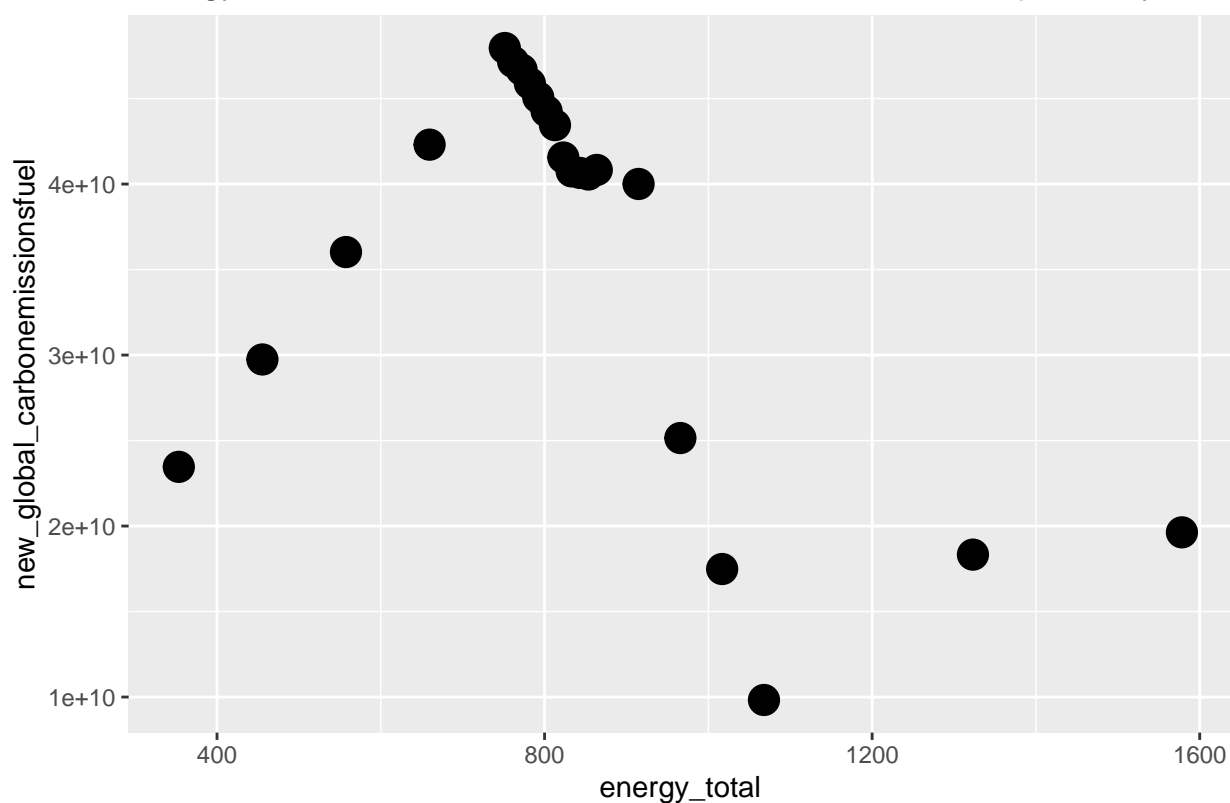## Energy demand with carbon emissions forecast based on past 40 yr trenc
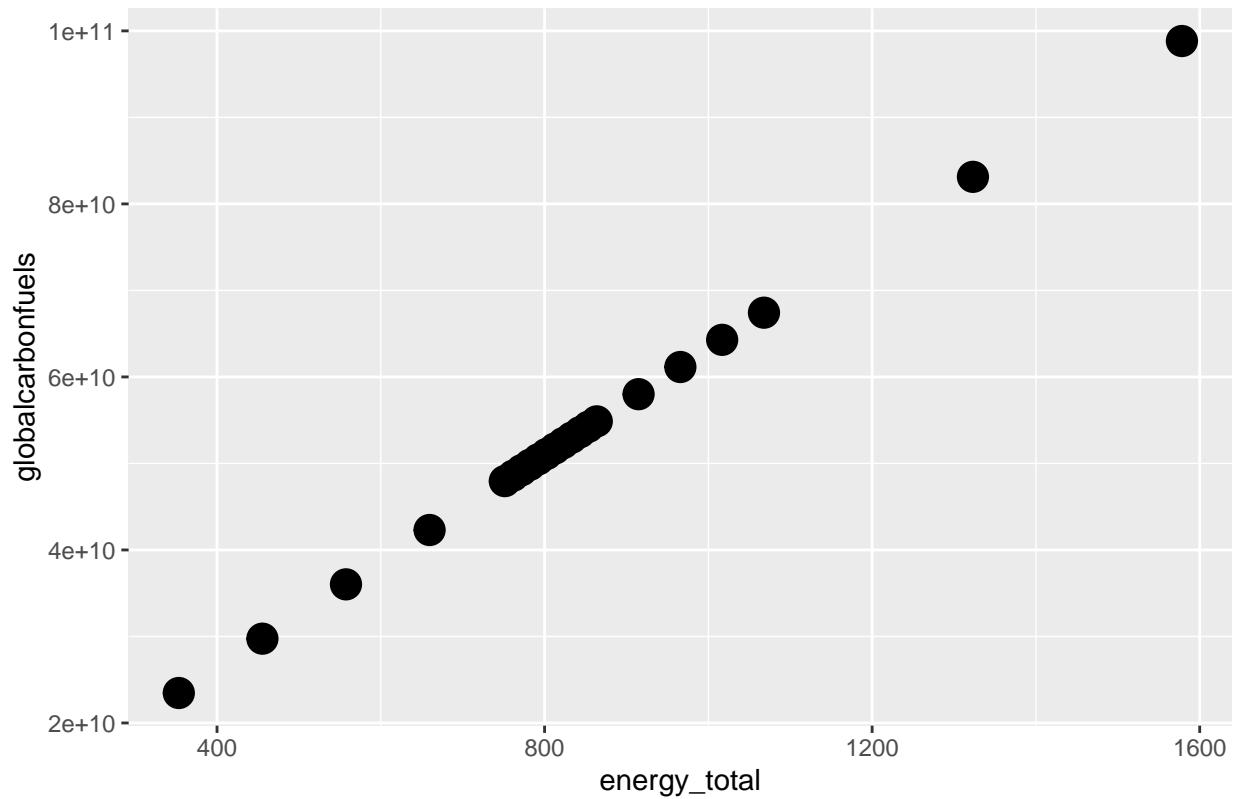
## Zero Carbon Energy quads with accelerated carbon emissions reduction



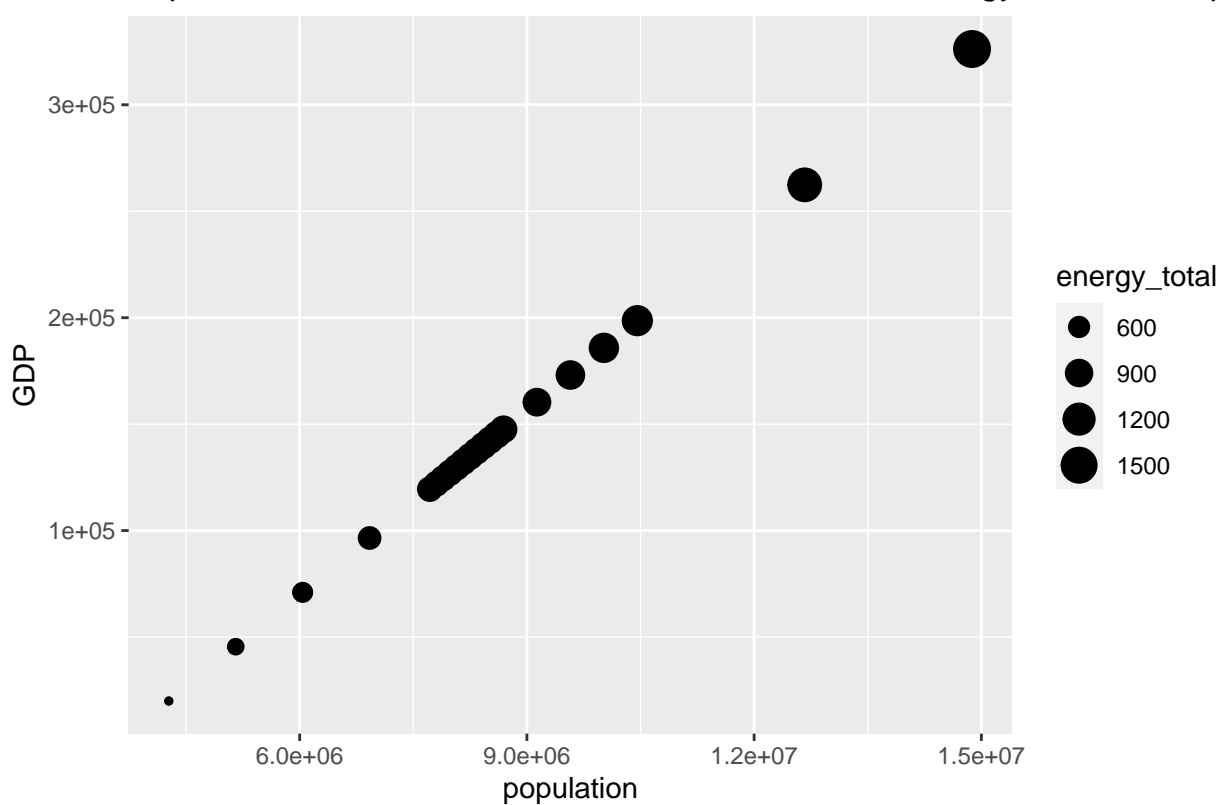## Energy demand forecast and associated carbon emission trend in progres

## Energy demand with carbon emissions forecast based on past 40 yr trend



## Population vs GDP with bubble size based on total energy demand in qua



Note that the `echo = FALSE` parameter was added to the code chunk to prevent printing of the R code that

generated the plot.

## Illustration of the impact energy shift away from fossil fuels

Assumptions for solar: 1) One square meter of solar cells generates 400 Kw hr per year and efficiency is 70% sunshine at 8 hr per day. 2) Therefore: 1 Quad= 10^15 BTU of energy or will require 755 Km2 of solar panels 3) Also 1 quad is approx equavlent to 183 million barrels of oil, 38.5 Million tons of coal or 980 billion cubic feet of natural gas. 4) How do we conceptualize the scale of the task? a) Two refinery locations at 250,000 BPD of oil processing is 1 Quad in a year (combined largest complex in the world that would exist today) b) Or appox 500 quad of energy would require a land area the size of Japan, Germany or a bit smaller than Spain. or approx 100 quad of energy would require a land area the size of Pannama or Ireland. c) or 1.2 Million wind turbines to generate 100 quads of energy which could fit into an area the scale of Delaware.

```
df_model<- df_model %>%mutate (Solar_total_energy_kmsquared=(energy_total*755),.after=year)
df_model<- df_model %>% mutate (Solar_totalenergy_num_Irelands=(Solar_total_energy_kmsquared/75500))

df_model<-df_model %>% mutate(Solar_renewable_energy_kmsquared=(energy_new_zerocarbon*755))
df_model<-df_model %>% mutate(Solar_renewable_num_Irelands=(Solar_renewable_energy_kmsquared/75500))

df_model<-df_model %>% mutate(wind_turbines_totalenergy=(energy_total/100*1200000))
df_model <-df_model %>% mutate(wind_turbines_totalenergy_num_delawares=(energy_total/100))
```
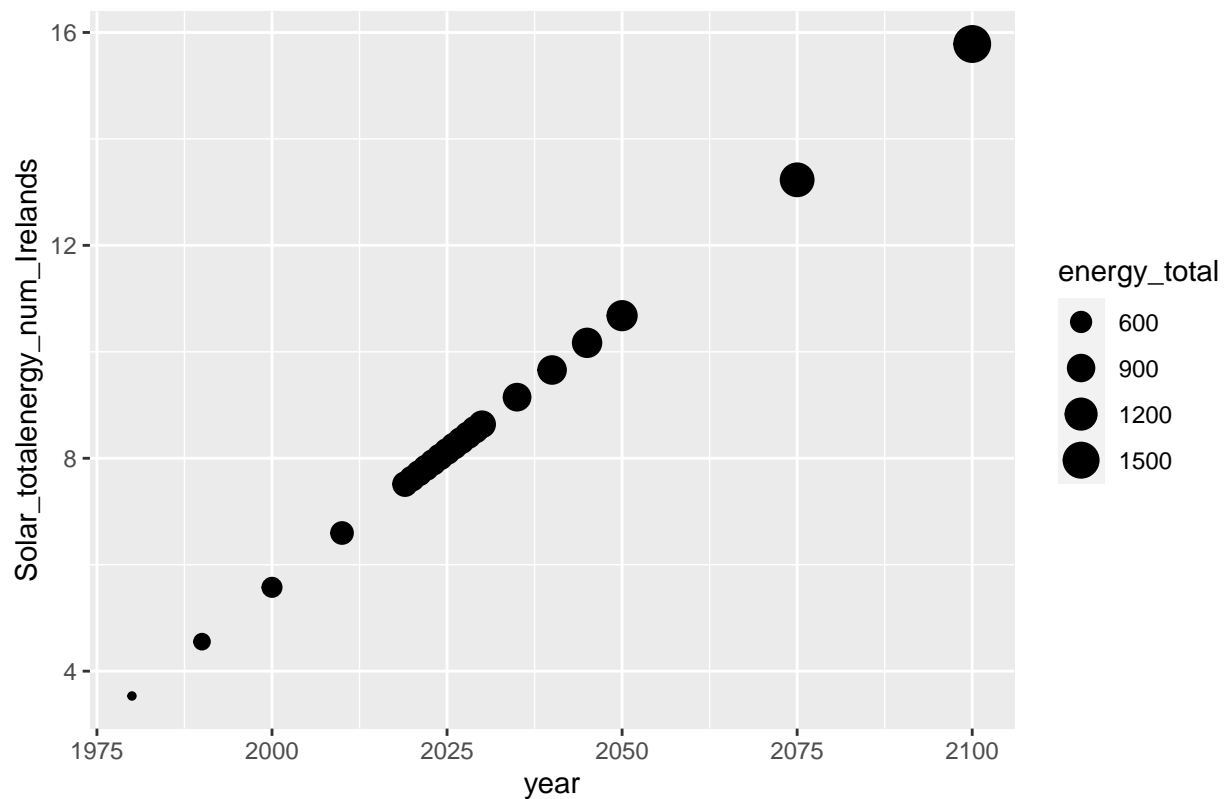
## Plot relative measures of tasks scale to reduce CO2 Emissions with renewables

Renewable energy as solar or as wind turbines to supply total energy demand

Equivalent areas estimated to provide benchmark of the scale of the requirements

## Solar: #Ireland sized areas to meet the global energy demand (energy Qua



## Wind Turb # Delaware sized areas required for global energy demand (ener
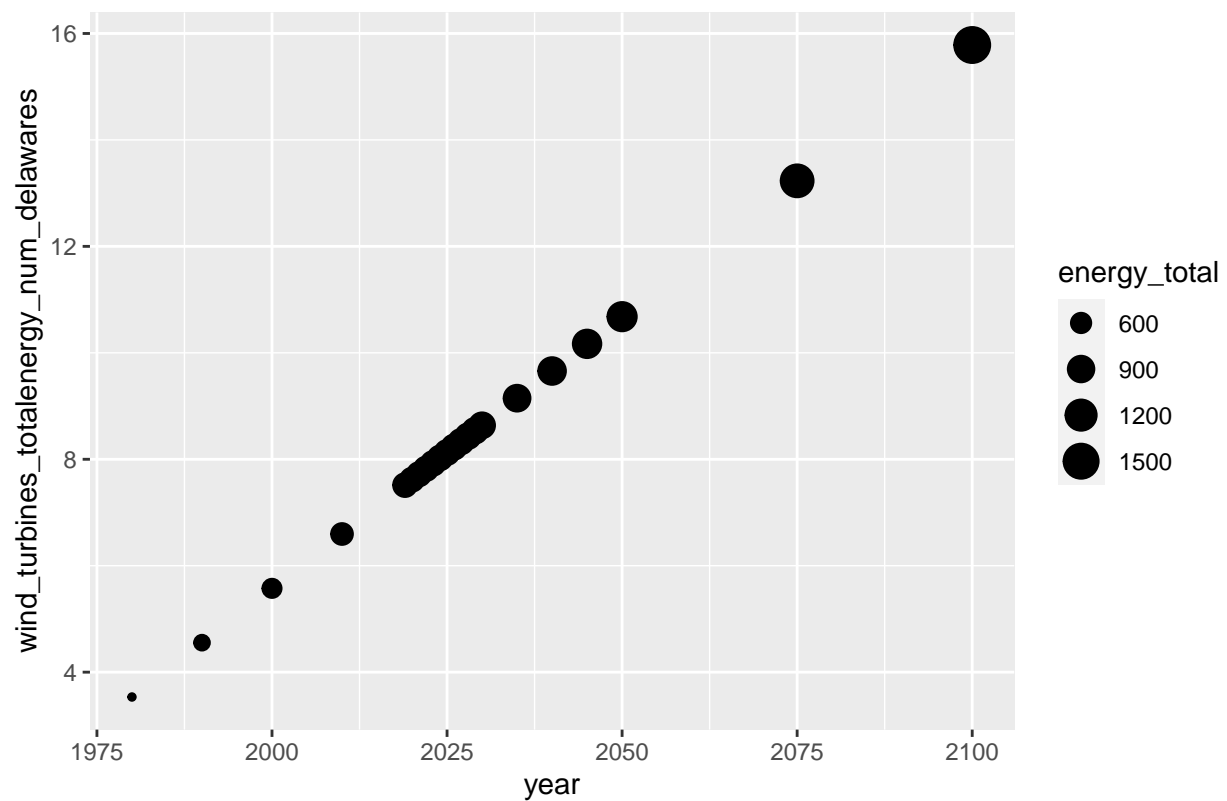
## References

1. US Energy Information Administration (EIA) at www.eia.gov.international

2. 2020 BP World Energy Outlook at www.bp.com

3. Jefferson W. Tester, et all, (2005) "Sustainable Energy Choosing Among Options" (MIT press)

4. Hadley Wickham, Garret Grolemund.(2017) "R for Data Science" (OReilly)

5. Winston Chang (2019), "R Graphics Cookbook" (OReilly)