Energy Use and Economic Development Project: Preprocessing and Data Wrangling

Ben Holden & Lizzy MacIntosh

2022-02-10

Intoduction:

It is becoming increasingly important to draw connections between the environment and human activities as the impact of the climate crisis is progressively more recognizable in our to day to day lives. In this analysis, we will focus on the connection between social and economic indicators and energy usage across the world.

The first dataset we will use is the World Energy Consumption dataset. This dataset is a collection of key metrics maintained by Our World in Data and includes data on energy consumption (primary energy, per capita, and growth rates), energy mix, electricity mix and other relevant metrics.

We will focus on the following key variables in our analysis:

Part 1: Importing and merging datasets:

This section loads, tidies, and merges the datasets used in this analysis.

First, tidy and select variables from the main dataset used – World Energy Consumption from Our World in Data

```
#glimpse(wec)

wec%>%
  select( #----- Key variables:  --------------------------------------------
          iso_code, #ISO country code
          country, #country
          year,    #year of observation
          population, #population total
          gdp, #  inflation-adjusted real GDP
          energy_per_gdp, # energy consumption per unit of GDP, kwh, 2011 dollars
          energy_per_capita, # Primary energy consumption per capita, kwh/year
          energy_cons_change_pct, #Annual pctage change in primary energy consumption
#      ------------- Energy  ---------------------------------------------------
          fossil_cons_change_pct, #Annual percentage change in fossil fuel consumption
          fossil_share_energy, #Share of primary energy consumption that comes from fossil fuels
          coal_share_energy, #Share of primary energy consumption that comes from coal
          gas_share_energy, #Share of primary energy consumption that comes from gas
          oil_share_energy, #   Share of primary energy consumption that comes from oil
          low_carbon_cons_change_pct, #Annual percentage change in low-carbon energy consumption
          hydro_share_energy,   #Share of primary energy consumption that comes from hydropower
          solar_share_energy, #Share of electricity consumption that comes from solar
          wind_share_energy, # Share of primary energy consumption that comes from wind
```

```r
         nuclear_share_energy, #Share of primary energy consumption that comes from other renewables
         biofuel_share_energy, #Share of primary energy consumption that comes from biofuels
         other_renewables_share_energy, #Share of primary energy consumption that comes from other ren
         low_carbon_share_energy, #Share of primary energy consumption that comes from low-carbon sour
#       ------------- Electricity -----------------------------------------------
         coal_share_elec, #Share of electricity consumption that comes from coal
         gas_share_elec, #Share of electricity consumption that comes from gas
         oil_share_elec, #Share of electricity consumption that comes from oil
         fossil_share_elec, #Share of electricity consumption that comes from fossil fuels (coal, oil
         hydro_share_elec, #Share of electricity consumption that comes from hydropower
         solar_share_elec, #Share of electricity consumption that comes from solar
         wind_share_elec, #Share of electricity consumption that comes from wind
         nuclear_share_elec, #Share of electricity consumption that comes from nuclear power
         biofuel_share_elec, #Share of electricity consumption that comes from biofuels
         other_renewables_share_elec_exc_biofuel,  #Share of electricity consumption that comes from o
         low_carbon_share_elec #Share of electricity consumption that comes from low-carbon sources. T
       ) %>%
#       ------------- Computing additional economic variables: ------------------------------------
  mutate(gdp_per_capita = (gdp/population),
         gdp_per_capita_growth = (gdp_per_capita-lag(gdp_per_capita))/lag(gdp_per_capita))%>% #creating
  relocate(gdp_per_capita, .before = energy_per_capita)%>% #relocating variables
  relocate(gdp_per_capita_growth, .after = gdp_per_capita)%>%
  filter(year >= 1960) -> wec_tidy

glimpse(wec_tidy)
```

```
## Rows: 11,827
## Columns: 34
## $ iso_code                                <chr> "AFG", "AFG", "AFG", "AFG", "A~
## $ country                                 <chr> "Afghanistan", "Afghanistan", ~
## $ year                                    <dbl> 1960, 1961, 1962, 1963, 1964, ~
## $ population                              <dbl> 8996967, 9169406, 9351442, 954~
## $ gdp                                     <dbl> 26970775552, 27196444672, 2766~
## $ energy_per_gdp                          <dbl> NA, NA, NA, NA, NA, NA, NA, NA~
## $ gdp_per_capita                          <dbl> 2997.763, 2965.999, 2958.855, ~
## $ gdp_per_capita_growth                   <dbl> 0.0171478201, -0.0105960876, -~
## $ energy_per_capita                       <dbl> NA, NA, NA, NA, NA, NA, NA, NA~
## $ energy_cons_change_pct                  <dbl> NA, NA, NA, NA, NA, NA, NA, NA~
## $ fossil_cons_change_pct                  <dbl> NA, NA, NA, NA, NA, NA, NA, NA~
## $ fossil_share_energy                     <dbl> NA, NA, NA, NA, NA, NA, NA, NA~
## $ coal_share_energy                       <dbl> NA, NA, NA, NA, NA, NA, NA, NA~
## $ gas_share_energy                        <dbl> NA, NA, NA, NA, NA, NA, NA, NA~
## $ oil_share_energy                        <dbl> NA, NA, NA, NA, NA, NA, NA, NA~
## $ low_carbon_cons_change_pct              <dbl> NA, NA, NA, NA, NA, NA, NA, NA~
## $ hydro_share_energy                      <dbl> NA, NA, NA, NA, NA, NA, NA, NA~
## $ solar_share_energy                      <dbl> NA, NA, NA, NA, NA, NA, NA, NA~
## $ wind_share_energy                       <dbl> NA, NA, NA, NA, NA, NA, NA, NA~
## $ nuclear_share_energy                    <dbl> NA, NA, NA, NA, NA, NA, NA, NA~
## $ biofuel_share_energy                    <dbl> NA, NA, NA, NA, NA, NA, NA, NA~
## $ other_renewables_share_energy           <dbl> NA, NA, NA, NA, NA, NA, NA, NA~
## $ low_carbon_share_energy                 <dbl> NA, NA, NA, NA, NA, NA, NA, NA~
## $ coal_share_elec                         <dbl> NA, NA, NA, NA, NA, NA, NA, NA~
## $ gas_share_elec                          <dbl> NA, NA, NA, NA, NA, NA, NA, NA~
## $ oil_share_elec                          <dbl> NA, NA, NA, NA, NA, NA, NA, NA~
```

```
## $ fossil_share_elec                        <dbl> NA, NA, NA, NA, NA, NA, NA, NA~
## $ hydro_share_elec                         <dbl> NA, NA, NA, NA, NA, NA, NA, NA~
## $ solar_share_elec                         <dbl> NA, NA, NA, NA, NA, NA, NA, NA~
## $ wind_share_elec                          <dbl> NA, NA, NA, NA, NA, NA, NA, NA~
## $ nuclear_share_elec                       <dbl> NA, NA, NA, NA, NA, NA, NA, NA~
## $ biofuel_share_elec                       <dbl> NA, NA, NA, NA, NA, NA, NA, NA~
## $ other_renewables_share_elec_exc_biofuel <dbl> NA, NA, NA, NA, NA, NA, NA, NA~
## $ low_carbon_share_elec                    <dbl> NA, NA, NA, NA, NA, NA, NA, NA~
```

Next we will tidy and conform the other datasets to prepare for joins:

CO2 Emissions: This dataset is from The World Bank and provides us with CO2 emissions in metric tons per capita from 1960-2020.

```
#glimpse(co2_emissions)

co2_emissions%>%
  pivot_longer(cols = `1960`:`2020`, names_to = "year", values_to = "co2_tons_per_capita")%>%
  select(`Country Code`, `Country Name`, year, co2_tons_per_capita)%>%
  rename(iso_code = `Country Code`,
         country = `Country Name`)%>%
  mutate(year = as.numeric(year)) -> co2_tidy

head(co2_tidy, 5)
```

```
## # A tibble: 5 x 4
##   iso_code country  year co2_tons_per_capita
##   <chr>    <chr>   <dbl>               <dbl>
## 1 ABW      Aruba    1960                205.
## 2 ABW      Aruba    1961                209.
## 3 ABW      Aruba    1962                226.
## 4 ABW      Aruba    1963                215.
## 5 ABW      Aruba    1964                208.
```

Fertility: This dataset is from The World Bank and provides us with the number of births per woman emissions from 1960-2019. Fertility is a useful measure of welfare becuase it indicates access to birth control and education on family planning.

```
#glimpse(fertility)

fertility%>%
  pivot_longer(cols = `1960`:`2020`, names_to = "year", values_to = "fertility_rate")%>%
  select(`Country Code`, `Country Name`, year, fertility_rate)%>%
  rename(iso_code = `Country Code`,
         country = `Country Name`)%>%
    mutate(year = as.numeric(year)) -> fertility_tidy

  head(fertility_tidy, 5)
```

```
## # A tibble: 5 x 4
##   iso_code country  year fertility_rate
##   <chr>    <chr>   <dbl>          <dbl>
## 1 ABW      Aruba    1960           4.82
## 2 ABW      Aruba    1961           4.66
## 3 ABW      Aruba    1962           4.47
## 4 ABW      Aruba    1963           4.27
```

```
## 5 ABW      Aruba    1964           4.06
```

Infant mortality: This dataset is from The World Bank and provides us with the number of infant deaths per 1000 births from 1960-2019. Infant Mortality Rate is a useful measure of welfare becuase indicates access to healthcare.

```
#glimpse(infant_mort)

infant_mort%>%
  pivot_longer(cols = `1960`:`2020`, names_to = "year", values_to = "infant_mortality_rate")%>%
  select(`Country Code`, `Country Name`, year, infant_mortality_rate)%>%
  rename(iso_code = `Country Code`,
         country = `Country Name`)%>%
    mutate(year = as.numeric(year)) -> infant_mort_tidy

head(infant_mort_tidy, 5)
```

```
## # A tibble: 5 x 4
##   iso_code country  year infant_mortality_rate
##   <chr>    <chr>   <dbl>                 <dbl>
## 1 ABW      Aruba    1960                    NA
## 2 ABW      Aruba    1961                    NA
## 3 ABW      Aruba    1962                    NA
## 4 ABW      Aruba    1963                    NA
## 5 ABW      Aruba    1964                    NA
```

Literacy: This dataset is from The World Bank and provides us with the percent of people above the age of 15 in a given country who are literate from 1970-2020.

```
#glimpse(literacy)

glimpse(literacy)%>%
  pivot_longer(cols = `1960`:`2020`, names_to = "year", values_to = "literacy_rate")%>%
  select(`Country Code`, `Country Name`, year, literacy_rate)%>%
  rename(iso_code = `Country Code`,
         country = `Country Name`)%>%
    mutate(year = as.numeric(year)) -> literacy_tidy
```

```
## Rows: 266
## Columns: 66
## $ `Country Name`   <chr> "Aruba", "Africa Eastern and Southern", "Afghanistan"~
## $ `Country Code`   <chr> "ABW", "AFE", "AFG", "AFW", "AGO", "ALB", "AND", "ARB~
## $ `Indicator Name` <chr> "Literacy rate, adult total (% of people ages 15 and ~
## $ `Indicator Code` <chr> "SE.ADT.LITR.ZS", "SE.ADT.LITR.ZS", "SE.ADT.LITR.ZS",~
## $ `1960`           <lgl> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, N~
## $ `1961`           <lgl> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, N~
## $ `1962`           <lgl> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, N~
## $ `1963`           <lgl> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, N~
## $ `1964`           <lgl> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, N~
## $ `1965`           <lgl> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, N~
## $ `1966`           <lgl> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, N~
## $ `1967`           <lgl> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, N~
## $ `1968`           <lgl> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, N~
## $ `1969`           <lgl> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, N~
## $ `1970`           <dbl> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, N~
## $ `1971`           <dbl> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, N~
```

```
## $ `1972`           <dbl> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, N~
## $ `1973`           <dbl> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, N~
## $ `1974`           <dbl> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, N~
## $ `1975`           <dbl> NA, NA, NA, NA, NA, NA, NA, NA, 53.51488, NA, NA, NA,~
## $ `1976`           <dbl> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, N~
## $ `1977`           <dbl> NA, NA, NA, NA, NA, NA, NA, 44.90297, NA, NA, NA, NA,~
## $ `1978`           <dbl> NA, NA, NA, NA, NA, NA, NA, 45.26824, NA, NA, NA, NA,~
## $ `1979`           <dbl> NA, NA, 18.15768, NA, NA, NA, NA, 45.68688, NA, NA, N~
## $ `1980`           <dbl> NA, NA, NA, NA, NA, NA, NA, 46.20202, NA, 93.91286, N~
## $ `1981`           <dbl> NA, NA, NA, NA, NA, NA, NA, 46.76899, NA, NA, NA, NA,~
## $ `1982`           <dbl> NA, NA, NA, NA, NA, NA, NA, 47.41513, NA, NA, NA, NA,~
## $ `1983`           <dbl> NA, NA, NA, NA, NA, NA, NA, 47.62678, NA, NA, NA, NA,~
## $ `1984`           <dbl> NA, NA, NA, NA, NA, NA, NA, 48.47257, NA, NA, NA, NA,~
## $ `1985`           <dbl> NA, NA, NA, NA, NA, NA, NA, 49.32794, 71.23530, NA, N~
## $ `1986`           <dbl> NA, NA, NA, NA, NA, NA, NA, 50.30286, NA, NA, NA, NA,~
## $ `1987`           <dbl> NA, NA, NA, NA, NA, NA, NA, 50.89021, NA, NA, NA, NA,~
## $ `1988`           <dbl> NA, NA, NA, NA, NA, NA, NA, 51.87934, NA, NA, NA, NA,~
## $ `1989`           <dbl> NA, NA, NA, NA, NA, NA, NA, 52.78355, NA, NA, 98.7519~
## $ `1990`           <dbl> NA, NA, NA, NA, NA, NA, NA, 53.75409, NA, NA, NA, NA,~
## $ `1991`           <dbl> NA, NA, NA, NA, NA, NA, NA, 54.85839, NA, 96.04072, N~
## $ `1992`           <dbl> NA, NA, NA, NA, NA, NA, NA, 55.79609, NA, NA, NA, NA,~
## $ `1993`           <dbl> NA, NA, NA, NA, NA, NA, NA, 56.5178, NA, NA, NA, NA, ~
## $ `1994`           <dbl> NA, NA, NA, NA, NA, NA, NA, 57.35207, NA, NA, NA, NA,~
## $ `1995`           <dbl> NA, NA, NA, NA, NA, NA, NA, 58.33942, NA, NA, NA, NA,~
## $ `1996`           <dbl> NA, NA, NA, NA, NA, NA, NA, 59.80857, NA, NA, NA, NA,~
## $ `1997`           <dbl> NA, NA, NA, NA, NA, NA, NA, 60.54454, NA, NA, NA, NA,~
## $ `1998`           <dbl> NA, NA, NA, NA, NA, NA, NA, 61.72372, NA, NA, NA, NA,~
## $ `1999`           <dbl> NA, NA, NA, NA, NA, NA, NA, 62.75603, NA, NA, NA, NA,~
## $ `2000`           <dbl> 97.29125, NA, NA, NA, NA, NA, NA, 63.68453, NA, NA, N~
## $ `2001`           <dbl> NA, NA, NA, NA, 67.40542, 98.71298, NA, 64.87778, NA,~
## $ `2002`           <dbl> NA, NA, NA, NA, NA, NA, NA, 66.04557, NA, NA, NA, NA,~
## $ `2003`           <dbl> NA, NA, NA, NA, NA, NA, NA, 68.68774, NA, NA, NA, NA,~
## $ `2004`           <dbl> NA, NA, NA, NA, NA, NA, NA, 69.63666, NA, NA, NA, NA,~
## $ `2005`           <dbl> NA, NA, NA, NA, NA, NA, NA, 69.21840, 90.03385, NA, N~
## $ `2006`           <dbl> NA, NA, NA, NA, NA, NA, NA, 68.13701, NA, 98.61080, N~
## $ `2007`           <dbl> NA, NA, NA, NA, NA, NA, NA, 68.62272, NA, NA, NA, NA,~
## $ `2008`           <dbl> NA, NA, NA, NA, NA, 95.93864, NA, 69.62151, NA, NA, N~
## $ `2009`           <dbl> NA, NA, NA, NA, NA, NA, NA, 68.30745, NA, 98.98342, N~
## $ `2010`           <dbl> 96.82264, NA, NA, NA, NA, NA, NA, 69.24863, NA, 98.95~
## $ `2011`           <dbl> NA, NA, 31.44885, NA, NA, 96.84530, NA, 71.06061, NA,~
## $ `2012`           <dbl> NA, NA, NA, NA, NA, 97.24697, NA, 74.43569, NA, 99.10~
## $ `2013`           <dbl> NA, NA, NA, NA, NA, NA, NA, 74.04358, NA, 99.12195, N~
## $ `2014`           <dbl> NA, NA, NA, NA, 66.03011, NA, NA, 75.71389, NA, 98.99~
## $ `2015`           <dbl> NA, NA, NA, NA, NA, NA, NA, 73.49939, NA, 99.17996, N~
## $ `2016`           <dbl> NA, NA, NA, NA, NA, NA, NA, 74.84080, NA, 99.12501, N~
## $ `2017`           <dbl> NA, NA, NA, NA, NA, NA, NA, 75.76682, NA, NA, 99.7360~
## $ `2018`           <dbl> 97.80742, NA, NA, NA, NA, 98.14115, NA, 72.86909, NA,~
## $ `2019`           <dbl> NA, NA, NA, NA, NA, NA, NA, 73.11425, 97.55692, NA, N~
## $ `2020`           <dbl> NA, NA, NA, NA, NA, NA, NA, 73.36777, NA, NA, 99.7886~
## $ ...66            <lgl> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, N~
```

```r
head(literacy_tidy, 5)
```

```
## # A tibble: 5 x 4
##   iso_code country  year literacy_rate
```

```
##   <chr>    <chr>   <dbl>          <dbl>
## 1 ABW      Aruba    1960             NA
## 2 ABW      Aruba    1961             NA
## 3 ABW      Aruba    1962             NA
## 4 ABW      Aruba    1963             NA
## 5 ABW      Aruba    1964             NA
```

Life expectancy: This dataset is from The World Bank and provides us with the life expectancy at birth in a given country

```r
#glimpse(life_exp)

glimpse(life_exp)%>%
  pivot_longer(cols = `1960`:`2020`, names_to = "year", values_to = "life_expectancy")%>%
  select(`Country Code`, `Country Name`, year, life_expectancy)%>%
  rename(iso_code = `Country Code`,
         country = `Country Name`)%>%
    mutate(year = as.numeric(year)) -> life_exp_tidy
```

```
## Rows: 266
## Columns: 67
## $ ...1             <dbl> 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16~
## $ `Country Name`   <chr> "Aruba", "Africa Eastern and Southern", "Afghanistan"~
## $ `Country Code`   <chr> "ABW", "AFE", "AFG", "AFW", "AGO", "ALB", "AND", "ARB~
## $ `Indicator Name` <chr> "Life expectancy at birth, total (years)", "Life expe~
## $ `Indicator Code` <chr> "SP.DYN.LE00.IN", "SP.DYN.LE00.IN", "SP.DYN.LE00.IN",~
## $ `1960`           <dbl> 65.66200, 42.71605, 32.44600, 37.20538, 37.52400, 62.~
## $ `1961`           <dbl> 66.07400, 43.16694, 32.96200, 37.63255, 37.81100, 63.~
## $ `1962`           <dbl> 66.44400, 43.60399, 33.47100, 38.05261, 38.11300, 64.~
## $ `1963`           <dbl> 66.78700, 44.02562, 33.97100, 38.46375, 38.43000, 64.~
## $ `1964`           <dbl> 67.11300, 44.43272, 34.46300, 38.86707, 38.76000, 65.~
## $ `1965`           <dbl> 67.43500, 44.82692, 34.94800, 39.26484, 39.10200, 65.~
## $ `1966`           <dbl> 67.76200, 45.21305, 35.43000, 39.66276, 39.45400, 66.~
## $ `1967`           <dbl> 68.09500, 45.59429, 35.91400, 40.06641, 39.81300, 66.~
## $ `1968`           <dbl> 68.43600, 45.97406, 36.40300, 40.48283, 40.17800, 66.~
## $ `1969`           <dbl> 68.78400, 46.35240, 36.90000, 40.91422, 40.54600, 66.~
## $ `1970`           <dbl> 69.14000, 46.72880, 37.40900, 41.36512, 40.91400, 66.~
## $ `1971`           <dbl> 69.49800, 47.10287, 37.93000, 41.83721, 41.28200, 67.~
## $ `1972`           <dbl> 69.85100, 47.47117, 38.46100, 42.32704, 41.65000, 67.~
## $ `1973`           <dbl> 70.19100, 47.82940, 39.00300, 42.82908, 42.01600, 67.~
## $ `1974`           <dbl> 70.51900, 48.17500, 39.55800, 43.33922, 42.37400, 68.~
## $ `1975`           <dbl> 70.83300, 48.50331, 40.12800, 43.85504, 42.72100, 68.~
## $ `1976`           <dbl> 71.14000, 48.81084, 40.71500, 44.37397, 43.05300, 69.~
## $ `1977`           <dbl> 71.44100, 49.09785, 41.32000, 44.89223, 43.36700, 69.~
## $ `1978`           <dbl> 71.73600, 49.36633, 41.94400, 45.40268, 43.66000, 69.~
## $ `1979`           <dbl> 72.02300, 49.62000, 42.58500, 45.89748, 43.93100, 69.~
## $ `1980`           <dbl> 72.29300, 49.87069, 43.24400, 46.36610, 44.17800, 70.~
## $ `1981`           <dbl> 72.53800, 50.11589, 43.92300, 46.79829, 44.40400, 70.~
## $ `1982`           <dbl> 72.75100, 50.36346, 44.61700, 47.18862, 44.61100, 70.~
## $ `1983`           <dbl> 72.92900, 50.61038, 45.32400, 47.53399, 44.79900, 70.~
## $ `1984`           <dbl> 73.07100, 50.84833, 46.04000, 47.83054, 44.96600, 71.~
## $ `1985`           <dbl> 73.18100, 51.05806, 46.76100, 48.07938, 45.10700, 71.~
## $ `1986`           <dbl> 73.26200, 51.21428, 47.48600, 48.28458, 45.21300, 71.~
## $ `1987`           <dbl> 73.32500, 51.29978, 48.21100, 48.45413, 45.28300, 71.~
## $ `1988`           <dbl> 73.37800, 51.30860, 48.93000, 48.59742, 45.31700, 71.~
```

```
## $ `1989`          <dbl> 73.42500, 51.25176, 49.64000, 48.71923, 45.32400, 71.~
## $ `1990`          <dbl> 73.46800, 51.15411, 50.33100, 48.81700, 45.30600, 71.~
## $ `1991`          <dbl> 73.50900, 51.04841, 50.99900, 48.88593, 45.27100, 71.~
## $ `1992`          <dbl> 73.54400, 50.95726, 51.64100, 48.92342, 45.23000, 71.~
## $ `1993`          <dbl> 73.57300, 50.89025, 52.25600, 48.93371, 45.20100, 71.~
## $ `1994`          <dbl> 73.59800, 50.84186, 52.84200, 48.92551, 45.20100, 71.~
## $ `1995`          <dbl> 73.62200, 50.80848, 53.39800, 48.90994, 45.24600, 72.~
## $ `1996`          <dbl> 73.64600, 50.79616, 53.92400, 48.89988, 45.35000, 72.~
## $ `1997`          <dbl> 73.67100, 50.82061, 54.42400, 48.90913, 45.51900, 72.~
## $ `1998`          <dbl> 73.70000, 50.89761, 54.90600, 48.95534, 45.76300, 73.~
## $ `1999`          <dbl> 73.73800, 51.04419, 55.37600, 49.05264, 46.09300, 73.~
## $ `2000`          <dbl> 73.78700, 51.27613, 55.84100, 49.21971, 46.52200, 73.~
## $ `2001`          <dbl> 73.85300, 51.60646, 56.30800, 49.47514, 47.05900, 74.~
## $ `2002`          <dbl> 73.93700, 52.04315, 56.78400, 49.81693, 47.70200, 74.~
## $ `2003`          <dbl> 74.03800, 52.58585, 57.27100, 50.23943, 48.44000, 74.~
## $ `2004`          <dbl> 74.15600, 53.22891, 57.77200, 50.73342, 49.26300, 75.~
## $ `2005`          <dbl> 74.28700, 53.96655, 58.29000, 51.28349, 50.16500, 75.~
## $ `2006`          <dbl> 74.42900, 54.79171, 58.82600, 51.86820, 51.14300, 75.~
## $ `2007`          <dbl> 74.57600, 55.68234, 59.37500, 52.46396, 52.17700, 75.~
## $ `2008`          <dbl> 74.72500, 56.60980, 59.93000, 53.04950, 53.24300, 75.~
## $ `2009`          <dbl> 74.87200, 57.54877, 60.48400, 53.61209, 54.31100, 76.~
## $ `2010`          <dbl> 75.01700, 58.47070, 61.02800, 54.14431, 55.35000, 76.~
## $ `2011`          <dbl> 75.15800, 59.35359, 61.55300, 54.65000, 56.33000, 76.~
## $ `2012`          <dbl> 75.29900, 60.18556, 62.05400, 55.13894, 57.23600, 77.~
## $ `2013`          <dbl> 75.44100, 60.95336, 62.52500, 55.61899, 58.05400, 77.~
## $ `2014`          <dbl> 75.58300, 61.64737, 62.96600, 56.08827, 58.77600, 77.~
## $ `2015`          <dbl> 75.72500, 62.25929, 63.37700, 56.54201, 59.39800, 78.~
## $ `2016`          <dbl> 75.86800, 62.78768, 63.76300, 56.97476, 59.92500, 78.~
## $ `2017`          <dbl> 76.01000, 63.24626, 64.13000, 57.38236, 60.37900, 78.~
## $ `2018`          <dbl> 76.15200, 63.64899, 64.48600, 57.76235, 60.78200, 78.~
## $ `2019`          <dbl> 76.29300, 64.00520, 64.83300, 58.11572, 61.14700, 78.~
## $ `2020`          <lgl> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, N~
## $ ...67           <lgl> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, N~
```

```
  head(life_exp_tidy, 5)
```

```
## # A tibble: 5 x 4
##   iso_code country  year life_expectancy
##   <chr>    <chr>   <dbl>           <dbl>
## 1 ABW      Aruba    1960            65.7
## 2 ABW      Aruba    1961            66.1
## 3 ABW      Aruba    1962            66.4
## 4 ABW      Aruba    1963            66.8
## 5 ABW      Aruba    1964            67.1
```

Poverty Rate: This dataset is from The World Bank and provides us with the share of people living on less than $1.90 a day from 1981-2019.

```
#glimpse(poverty)

poverty%>%
  select(Entity, Code, Year, `$1.90 per day - share of population below poverty line`)%>%
  rename(poverty_rate_dollar_ninety = `$1.90 per day - share of population below poverty line`,
         year = Year,
         country = Entity,
         iso_code = Code) -> poverty_tidy
```

```r
head(poverty_tidy, 5)
```

```
## # A tibble: 5 x 4
##   country iso_code  year poverty_rate_dollar_ninety
##   <chr>   <chr>    <dbl>                      <dbl>
## 1 Albania ALB       1981                      0.465
## 2 Albania ALB       1982                      0.408
## 3 Albania ALB       1983                      0.465
## 4 Albania ALB       1984                      0.535
## 5 Albania ALB       1985                      0.535
```

```r
wec_tidy%>%
  left_join( co2_tidy, by = c("iso_code", "year", "country"))%>%
  left_join(fertility_tidy, by = c("iso_code", "year", "country"))%>%
   left_join(infant_mort_tidy, by = c("iso_code", "year", "country")) %>%
   left_join(life_exp_tidy, by = c("iso_code", "year", "country"))%>%
    left_join(literacy_tidy, by = c("iso_code", "year", "country")) -> master

glimpse(master)
```

Merging all the datasets together

```
## Rows: 11,827
## Columns: 39
## $ iso_code                      <chr> "AFG", "AFG", "AFG", "AFG", "A~
## $ country                       <chr> "Afghanistan", "Afghanistan", ~
## $ year                          <dbl> 1960, 1961, 1962, 1963, 1964, ~
## $ population                    <dbl> 8996967, 9169406, 9351442, 954~
## $ gdp                           <dbl> 26970775552, 27196444672, 2766~
## $ energy_per_gdp                <dbl> NA, NA, NA, NA, NA, NA, NA, NA~
## $ gdp_per_capita                <dbl> 2997.763, 2965.999, 2958.855, ~
## $ gdp_per_capita_growth         <dbl> 0.0171478201, -0.0105960876, -~
## $ energy_per_capita             <dbl> NA, NA, NA, NA, NA, NA, NA, NA~
## $ energy_cons_change_pct        <dbl> NA, NA, NA, NA, NA, NA, NA, NA~
## $ fossil_cons_change_pct        <dbl> NA, NA, NA, NA, NA, NA, NA, NA~
## $ fossil_share_energy           <dbl> NA, NA, NA, NA, NA, NA, NA, NA~
## $ coal_share_energy             <dbl> NA, NA, NA, NA, NA, NA, NA, NA~
## $ gas_share_energy              <dbl> NA, NA, NA, NA, NA, NA, NA, NA~
## $ oil_share_energy              <dbl> NA, NA, NA, NA, NA, NA, NA, NA~
## $ low_carbon_cons_change_pct    <dbl> NA, NA, NA, NA, NA, NA, NA, NA~
## $ hydro_share_energy            <dbl> NA, NA, NA, NA, NA, NA, NA, NA~
## $ solar_share_energy            <dbl> NA, NA, NA, NA, NA, NA, NA, NA~
## $ wind_share_energy             <dbl> NA, NA, NA, NA, NA, NA, NA, NA~
## $ nuclear_share_energy          <dbl> NA, NA, NA, NA, NA, NA, NA, NA~
## $ biofuel_share_energy          <dbl> NA, NA, NA, NA, NA, NA, NA, NA~
## $ other_renewables_share_energy <dbl> NA, NA, NA, NA, NA, NA, NA, NA~
## $ low_carbon_share_energy       <dbl> NA, NA, NA, NA, NA, NA, NA, NA~
## $ coal_share_elec               <dbl> NA, NA, NA, NA, NA, NA, NA, NA~
## $ gas_share_elec                <dbl> NA, NA, NA, NA, NA, NA, NA, NA~
## $ oil_share_elec                <dbl> NA, NA, NA, NA, NA, NA, NA, NA~
## $ fossil_share_elec             <dbl> NA, NA, NA, NA, NA, NA, NA, NA~
## $ hydro_share_elec              <dbl> NA, NA, NA, NA, NA, NA, NA, NA~
## $ solar_share_elec              <dbl> NA, NA, NA, NA, NA, NA, NA, NA~
```

```
## $ wind_share_elec                         <dbl> NA, NA, NA, NA, NA, NA, NA, NA~
## $ nuclear_share_elec                      <dbl> NA, NA, NA, NA, NA, NA, NA, NA~
## $ biofuel_share_elec                      <dbl> NA, NA, NA, NA, NA, NA, NA, NA~
## $ other_renewables_share_elec_exc_biofuel <dbl> NA, NA, NA, NA, NA, NA, NA, NA~
## $ low_carbon_share_elec                   <dbl> NA, NA, NA, NA, NA, NA, NA, NA~
## $ co2_tons_per_capita                     <dbl> 0.04605674, 0.05358886, 0.0737~
## $ fertility_rate                          <dbl> 7.450, 7.450, 7.450, 7.450, 7.~
## $ infant_mortality_rate                   <dbl> NA, NA, 232.0, 227.8, 223.9, 2~
## $ life_expectancy                         <dbl> 32.446, 32.962, 33.471, 33.971~
## $ literacy_rate                           <dbl> NA, NA, NA, NA, NA, NA, NA, NA~
```

Next, we need to tidy the data once more before analysis:

```
master%>%
  filter(year >= 1991)%>%
  inner_join(poverty_tidy, by = c("iso_code", "year", "country" ))%>%
  relocate((co2_tons_per_capita:poverty_rate_dollar_ninety), .before = gdp) -> World_Energy_Consumption_

write_csv(World_Energy_Consumption_Main, "data/World_Energy_Consumption_Main.csv")
```

Part 2: Analysis

Part 2: Analysis

Research Question: How does energy usage and sources relate to social and economic indicators?

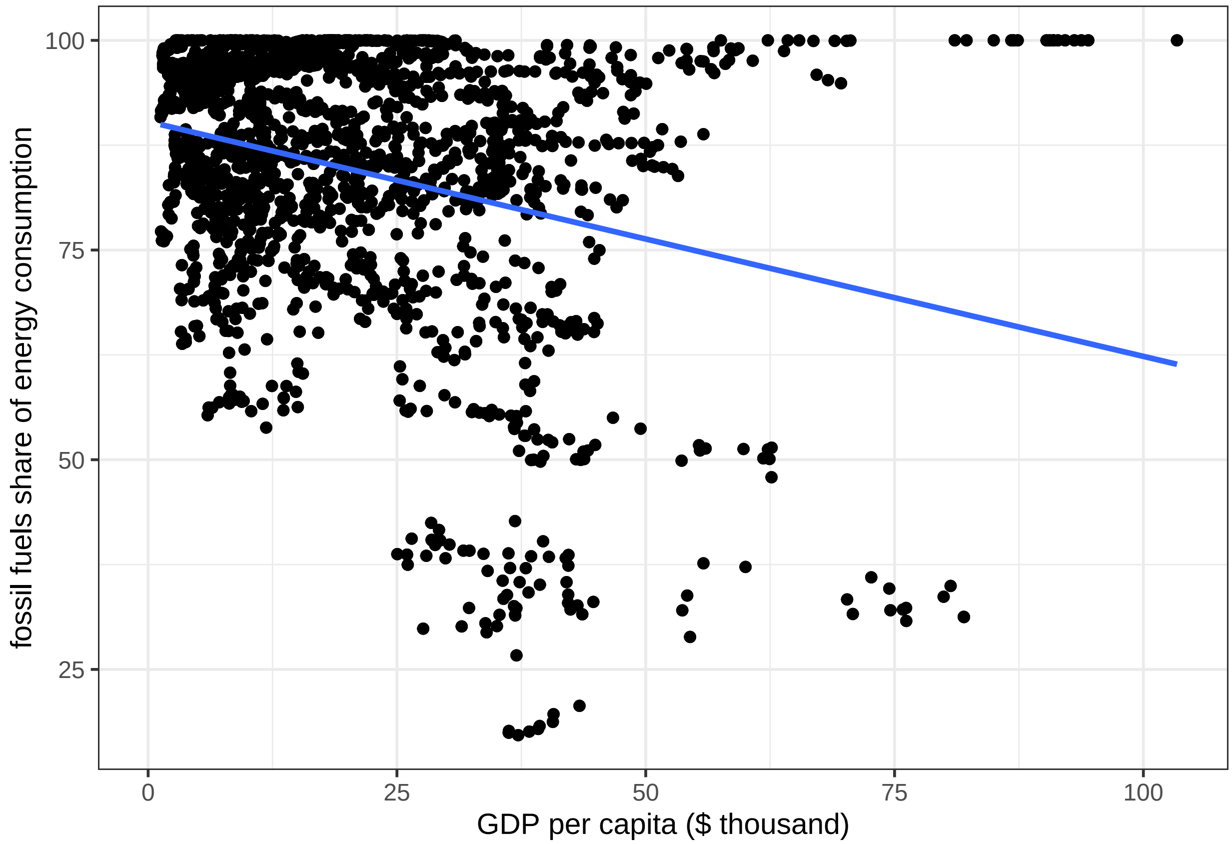
First we'll look at general correlation between GDP per capita and share of primary energy of fossil fuel.

```
ggplot(World_Energy_Consumption_Main, aes(gdp_per_capita/1000, fossil_share_energy))+
  geom_point()+
  geom_jitter()+
  geom_smooth(method = lm, se = FALSE)+
  theme_bw()+
  ylab("fossil fuels share of energy consumption")+
  xlab("GDP per capita ($ thousand)")
```

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

## Warning: Removed 2829 rows containing non-finite values (stat_smooth).

## Warning: Removed 2829 rows containing missing values (geom_point).
## Removed 2829 rows containing missing values (geom_point).
```

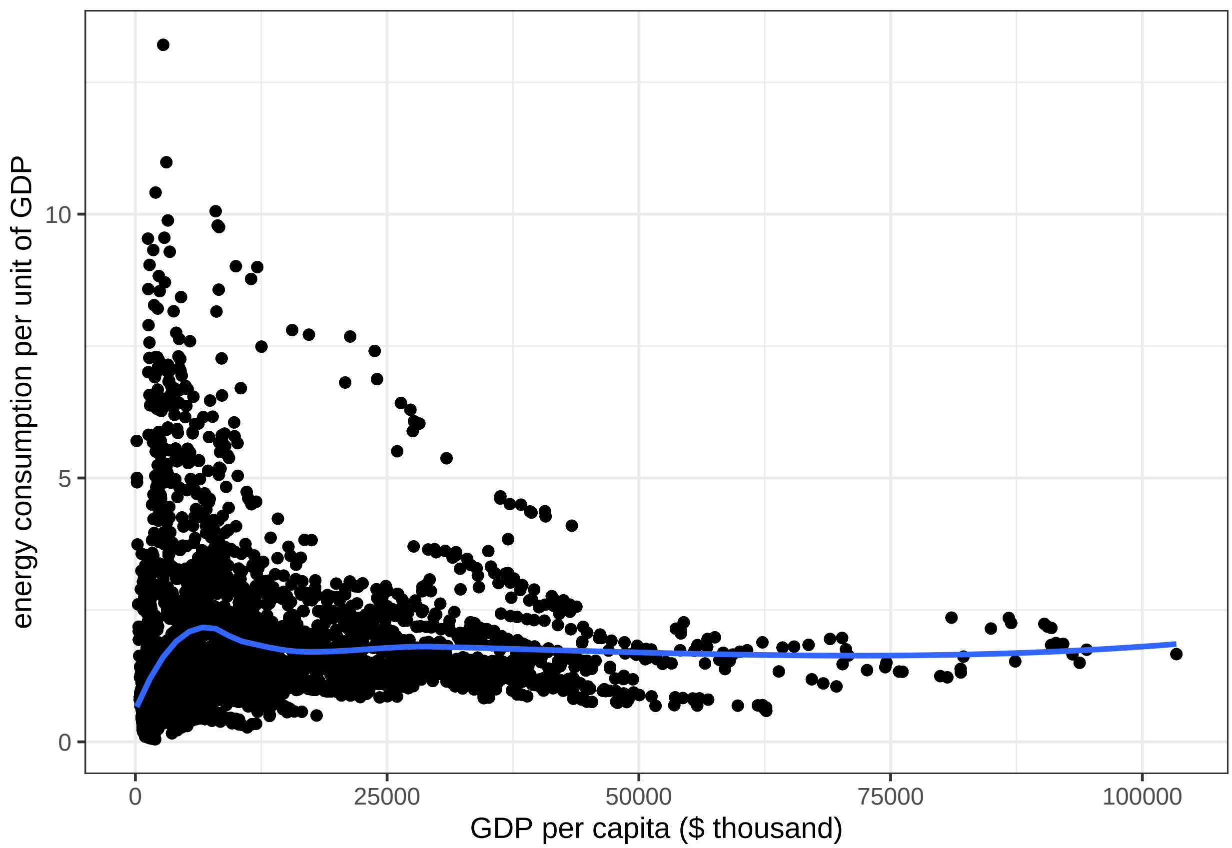



```
World_Energy_Consumption_Main%>%
  ggplot(aes(gdp_per_capita, energy_per_gdp))+
    geom_point()+
  geom_jitter()+
  geom_smooth(method = loess, se = FALSE)+
   theme_bw()+
  ylab("energy consumption per unit of GDP")+
  xlab("GDP per capita ($ thousand)")
```

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

## Warning: Removed 856 rows containing non-finite values (stat_smooth).

## Warning: Removed 856 rows containing missing values (geom_point).
## Removed 856 rows containing missing values (geom_point).
```



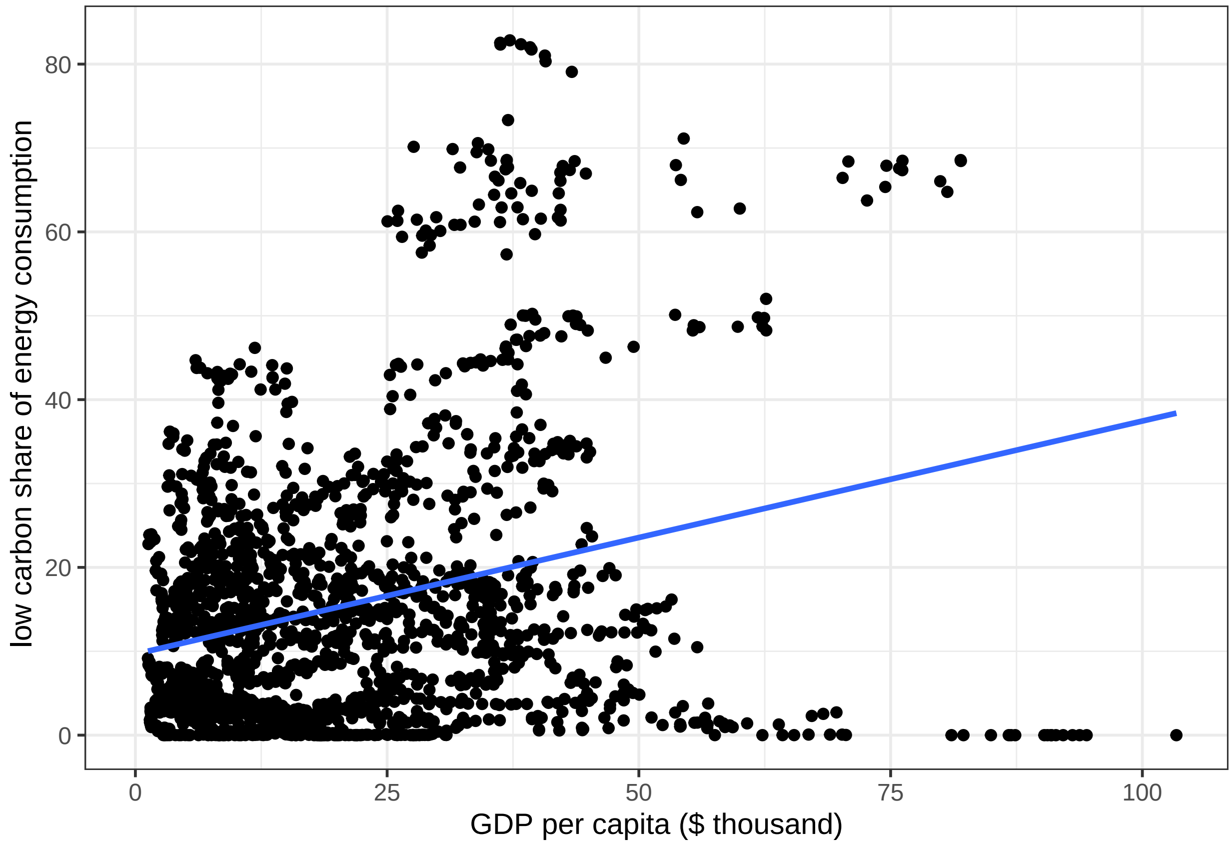
Now let's look at the complimentary plot. Here we'll be comparing GDP per capita and share of primary energy of "Low Carbon". *Note: Low Carbon is defined as the sum of primary energy from renewables and nuclear*

```
World_Energy_Consumption_Main%>%
  ggplot(aes(gdp_per_capita/1000, low_carbon_share_energy))+
  geom_point()+
    theme_bw()+
    geom_smooth(method = lm, se = FALSE)+
  ylab(" low carbon share of energy consumption")+
  xlab("GDP per capita ($ thousand)")
```

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

## Warning: Removed 2829 rows containing non-finite values (stat_smooth).

## Warning: Removed 2829 rows containing missing values (geom_point).
```



Now that we have seen that there is an association between energy source and GDP, we'll look at welfare measures in comparison to low carbon.

Infant Mortality rate:

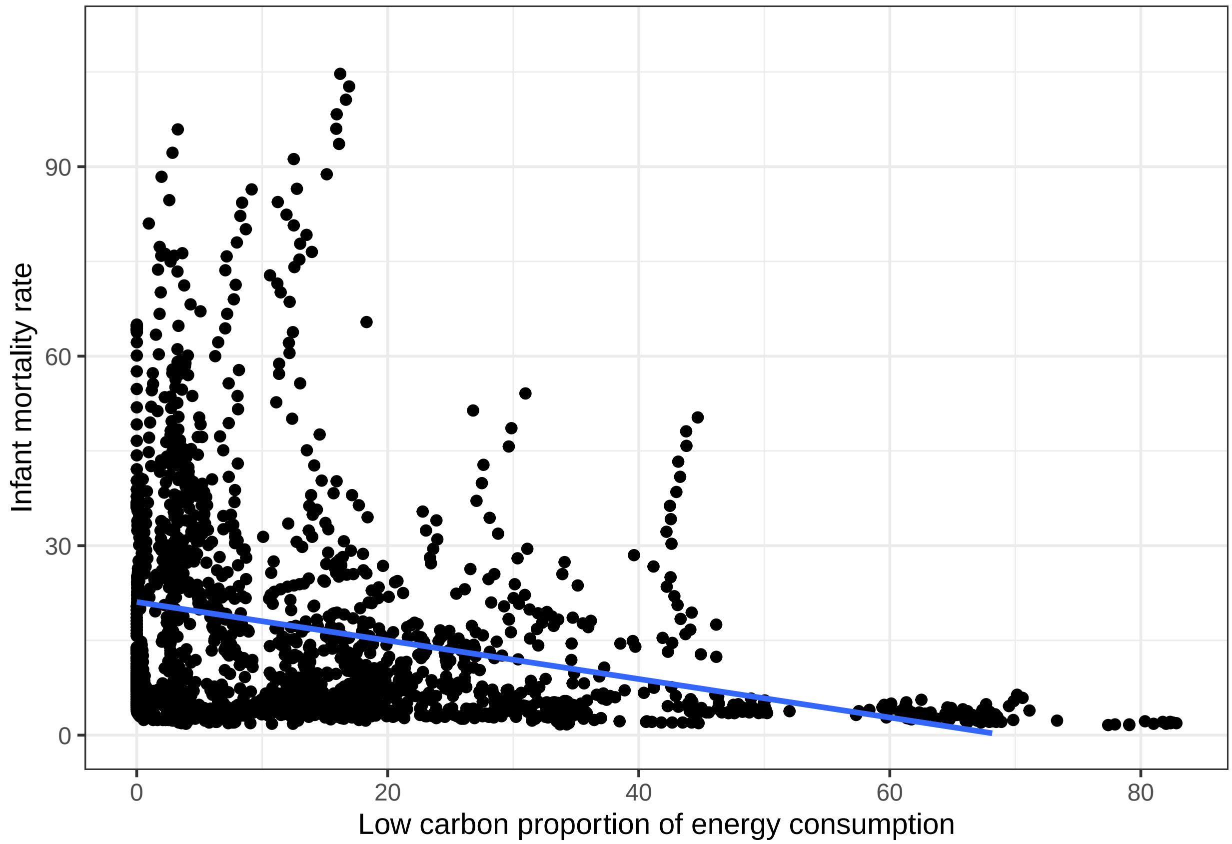
```
World_Energy_Consumption_Main%>%
  ggplot(aes(low_carbon_share_energy, infant_mortality_rate))+
  geom_point()+
  theme_bw()+
  geom_smooth(method = lm, se = FALSE)+
  ylim(0,110) +
  xlab("Low carbon proportion of energy consumption") +
  ylab("Infant mortality rate")
```

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

```
## Warning: Removed 2872 rows containing non-finite values (stat_smooth).
```

```
## Warning: Removed 2872 rows containing missing values (geom_point).
```

```
## Warning: Removed 14 rows containing missing values (geom_smooth).
```



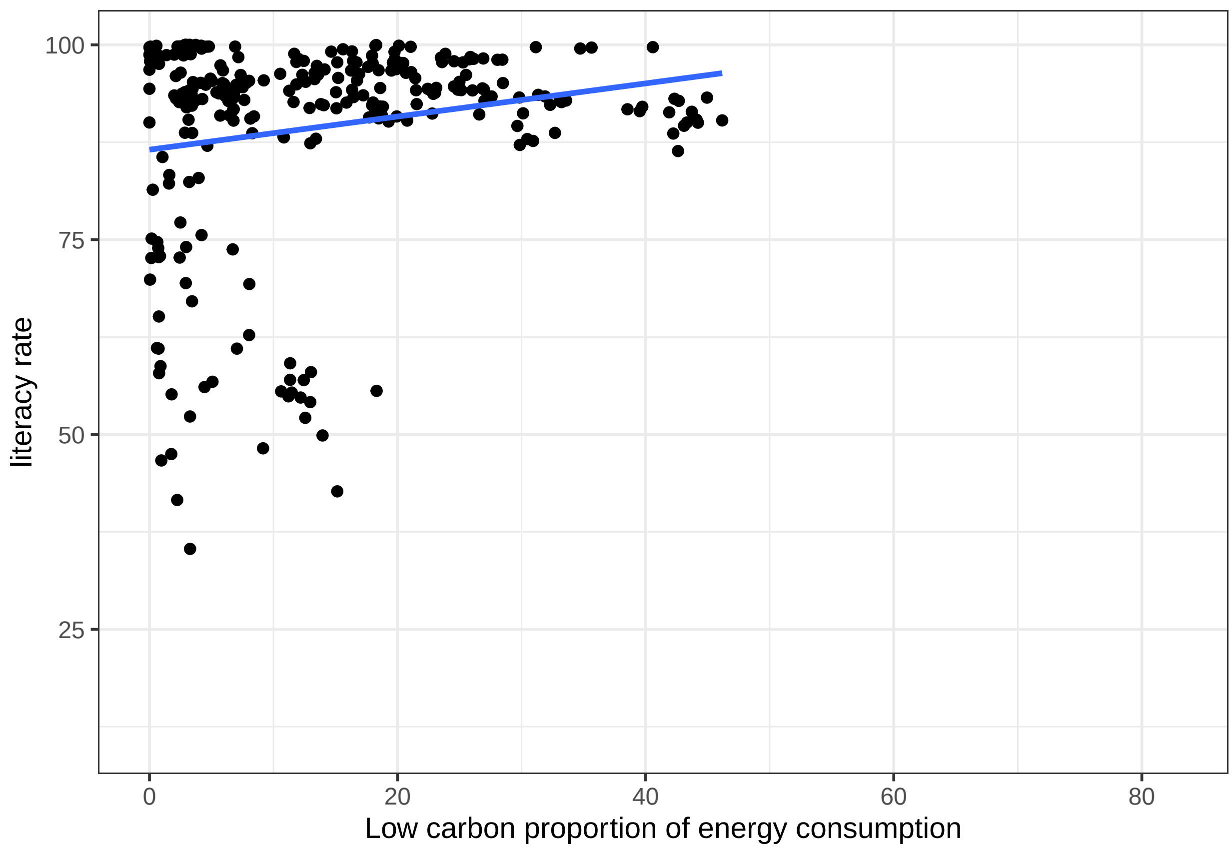
Literacy Rate:

```
World_Energy_Consumption_Main%>%
  ggplot(aes(low_carbon_share_energy, literacy_rate))+
  geom_point()+
  theme_bw()+
  xlab("Low carbon proportion of energy consumption") +
  ylab("literacy rate")+
  geom_smooth(method = lm, se = FALSE)
```

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

## Warning: Removed 4413 rows containing non-finite values (stat_smooth).

## Warning: Removed 4413 rows containing missing values (geom_point).
```



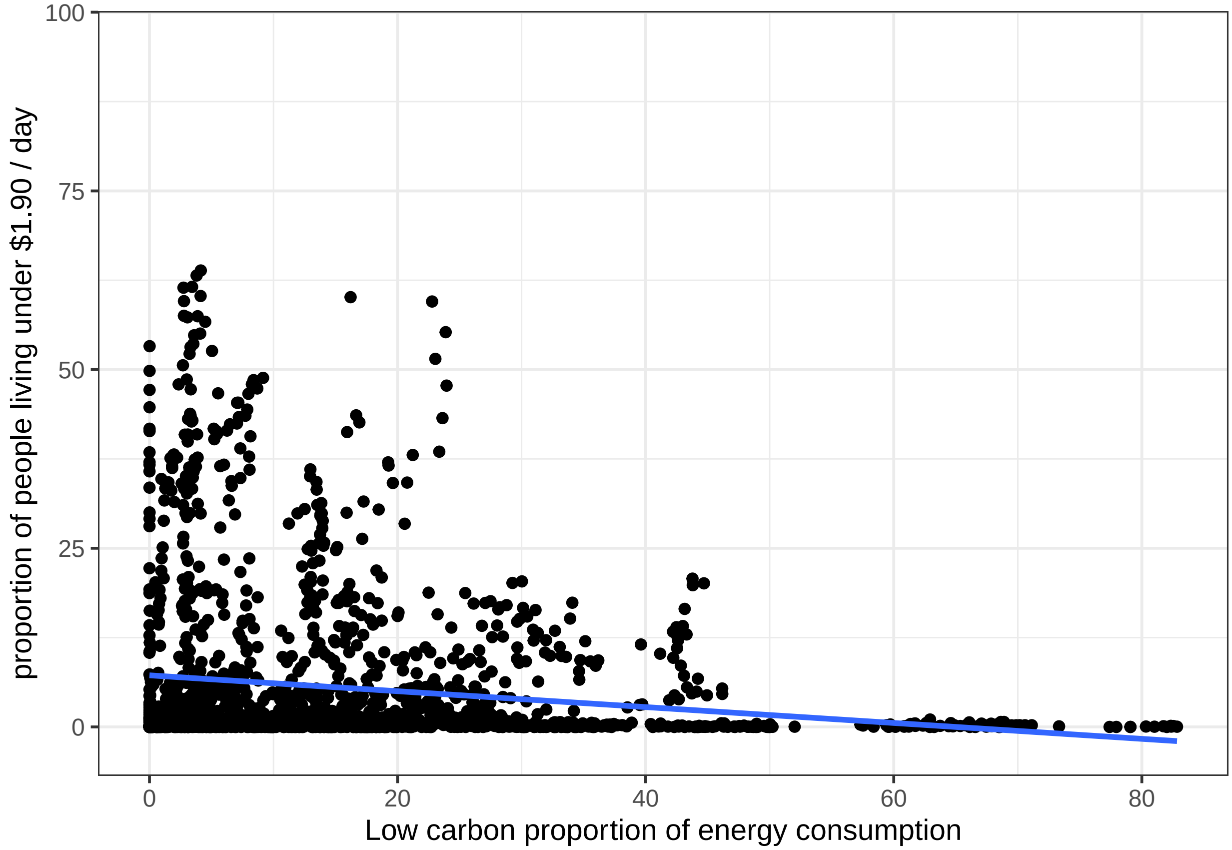
Poverty Rate:

```
World_Energy_Consumption_Main%>%
  ggplot(aes(low_carbon_share_energy, poverty_rate_dollar_ninety))+
  geom_point()+
 theme_bw()+
  xlab("Low carbon proportion of energy consumption") +
  ylab("proportion of people living under $1.90 / day")+
  geom_smooth(method = lm, se = FALSE)
```

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

```
## Warning: Removed 2618 rows containing non-finite values (stat_smooth).
```

```
## Warning: Removed 2618 rows containing missing values (geom_point).
```

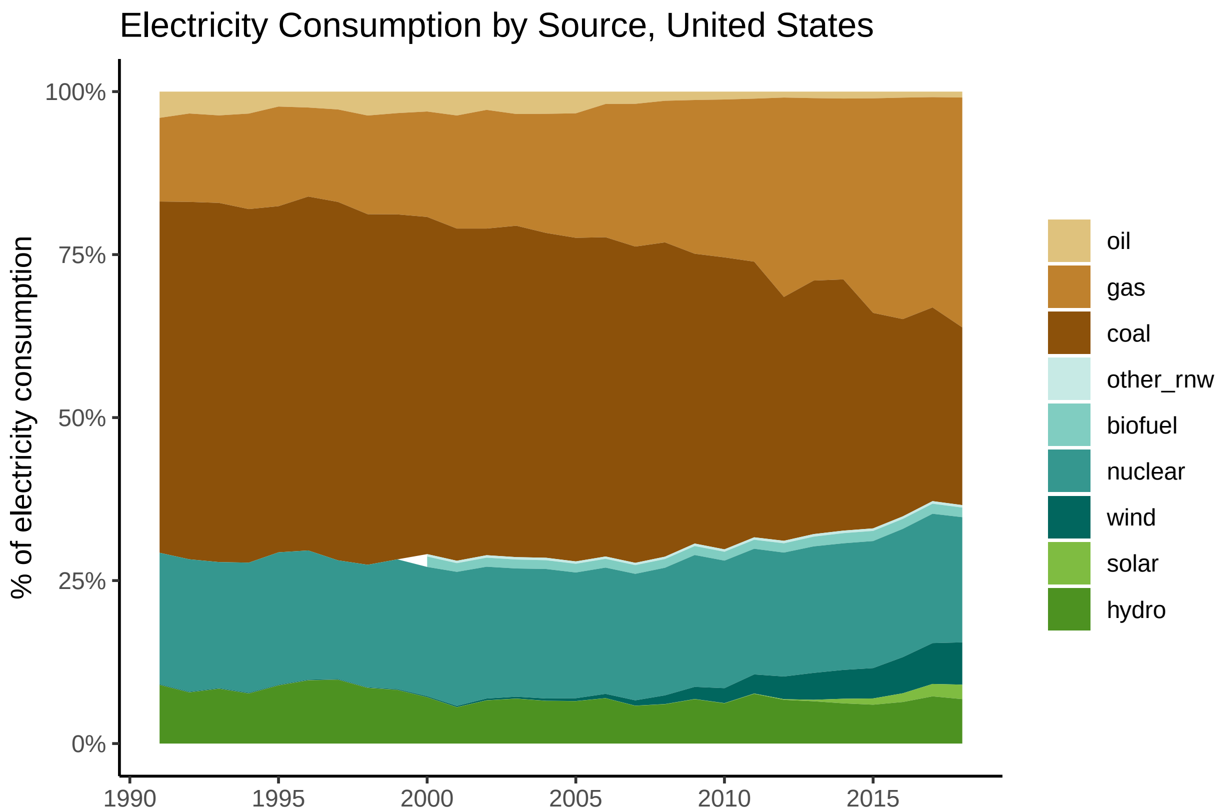


We'll also look at the breakdown of several country's energy usage. Let's focus on the 3 countries with the highest and lowest nominal GDP rankings.

```
World_Energy_Consumption_Main%>%
  filter(year < 2019)%>%
  filter(country == "United States")%>%
  select(year, coal_share_elec , gas_share_elec, oil_share_elec, hydro_share_elec, solar_share_elec, wi
  biofuel_share_elec,other_renewables_share_elec_exc_biofuel)%>%
  rename(coal = coal_share_elec,
         gas = gas_share_elec,
         oil = oil_share_elec,
         hydro = hydro_share_elec,
         solar = solar_share_elec,
         wind = wind_share_elec,
         nuclear = nuclear_share_elec,
         biofuel = biofuel_share_elec,
         other_rnw = other_renewables_share_elec_exc_biofuel)%>%
pivot_longer(cols = coal:other_rnw, names_to = "elec_source", values_to = "percent_of_elec")%>%
ggplot(aes(year, percent_of_elec, fill = factor(elec_source, levels = rev(y))))+
  geom_area(position = "fill")+
  theme_classic()+
  theme(legend.position = "right",
        legend.title=element_blank())+
  guides(color = guide_legend(nrow = 1),
         shape = guide_legend(override.aes = list(size = 0.5)))+
  labs(title = "Electricity Consumption by Source, United States")+
  ylab("% of electricity consumption")+
```

```
  xlab("")+
  scale_fill_manual(values = rev(yco),
                    labels = rev(y))+
  scale_y_continuous(labels = scales::percent)+
  scale_x_continuous(breaks = c(seq(from = 1990, to = 2020, by= 5)))
```

```
## Warning: Removed 18 rows containing missing values (position_stack).
```

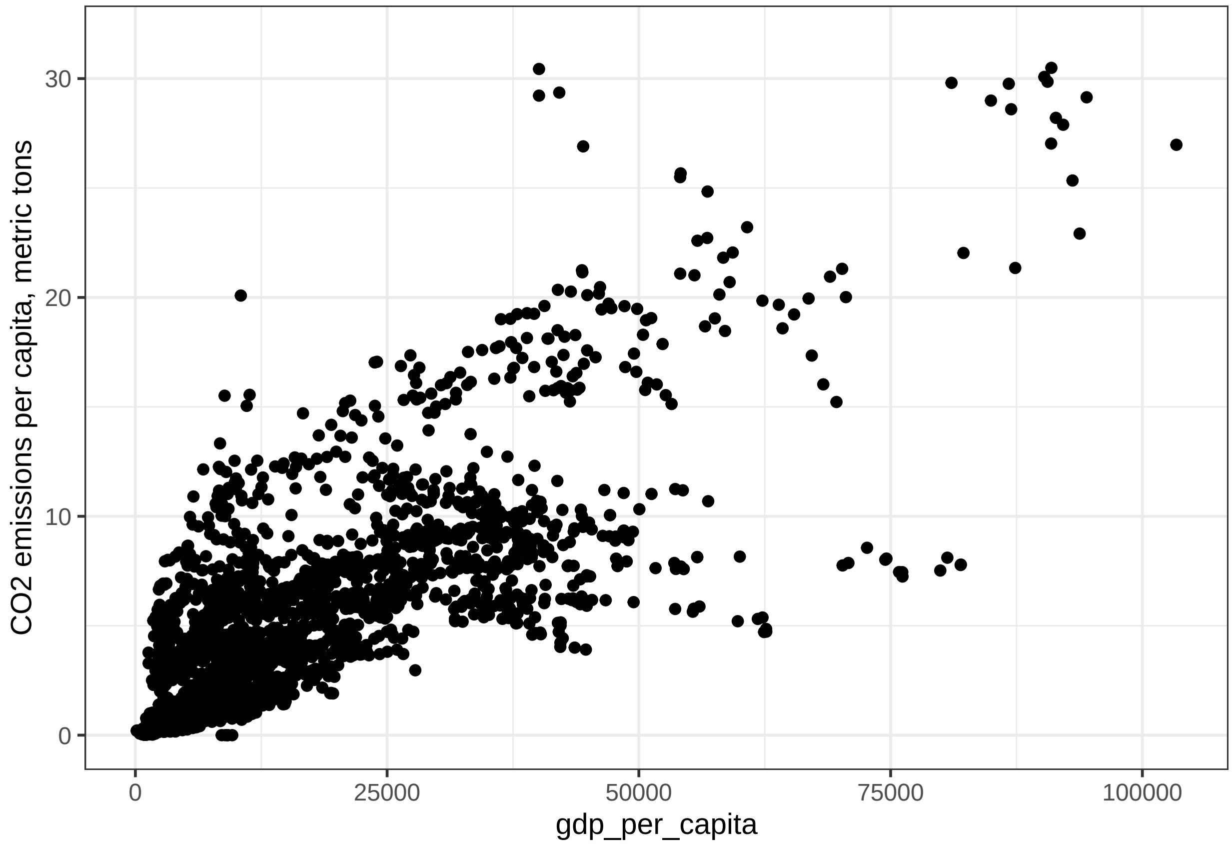


As seen above, the United States' electricity generation profile is moving away from dirtier fossil fuels. Do per capita carbon emissions reflect this trend?

```
World_Energy_Consumption_Main%>%
  filter(year < 2019)%>%
  filter(country == "United States")%>%
  select(year, co2_tons_per_capita)%>%
  ggplot(aes(year, co2_tons_per_capita))+
  geom_line()+
  theme_bw()+
  xlab("")+
  ylab("CO2 emissions per capita, metric tons")
```


```
World_Energy_Consumption_Main%>%
  filter(year < 2019)%>%
  select(gdp_per_capita, co2_tons_per_capita)%>%
  ggplot(aes(gdp_per_capita, co2_tons_per_capita))+
  geom_point()+
  theme_bw()+
  xlab("gdp_per_capita")+
  ylab("CO2 emissions per capita, metric tons")
```

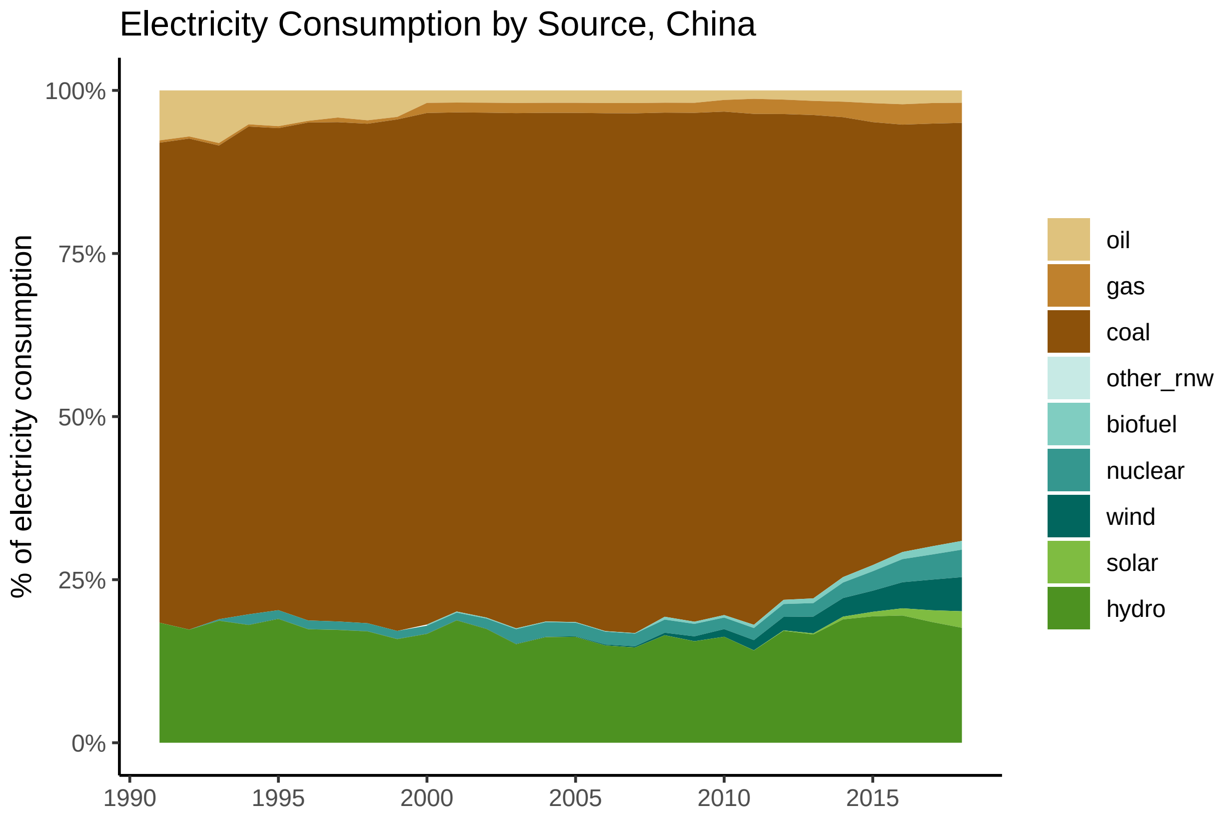
```
## Warning: Removed 1176 rows containing missing values (geom_point).
```




```
World_Energy_Consumption_Main%>%
  filter(year < 2019)%>%
  filter(country == "China")%>%
  select(year, coal_share_elec , gas_share_elec, oil_share_elec, hydro_share_elec, solar_share_elec, wi
  biofuel_share_elec,other_renewables_share_elec_exc_biofuel)%>%
  rename(coal = coal_share_elec,
         gas = gas_share_elec,
         oil = oil_share_elec,
         hydro = hydro_share_elec,
         solar = solar_share_elec,
         wind = wind_share_elec,
         nuclear = nuclear_share_elec,
         biofuel = biofuel_share_elec,
         other_rnw = other_renewables_share_elec_exc_biofuel)%>%
pivot_longer(cols = coal:other_rnw, names_to = "elec_source", values_to = "percent_of_elec")%>%
ggplot(aes(year, percent_of_elec, fill = factor(elec_source, levels = rev(y))))+
  geom_area(position = "fill")+
  theme_classic()+
  theme(legend.position = "right",
        legend.title=element_blank())+
  guides(color = guide_legend(nrow = 1),
         shape = guide_legend(override.aes = list(size = 0.5)))+
  labs(title = "Electricity Consumption by Source, China")+
  ylab("% of electricity consumption")+
  xlab("")+
  scale_fill_manual(values = rev(yco),
```

```
                     labels = rev(y))+
  scale_y_continuous(labels = scales::percent)+
  scale_x_continuous(breaks = c(seq(from = 1990, to = 2020, by= 5)))
```

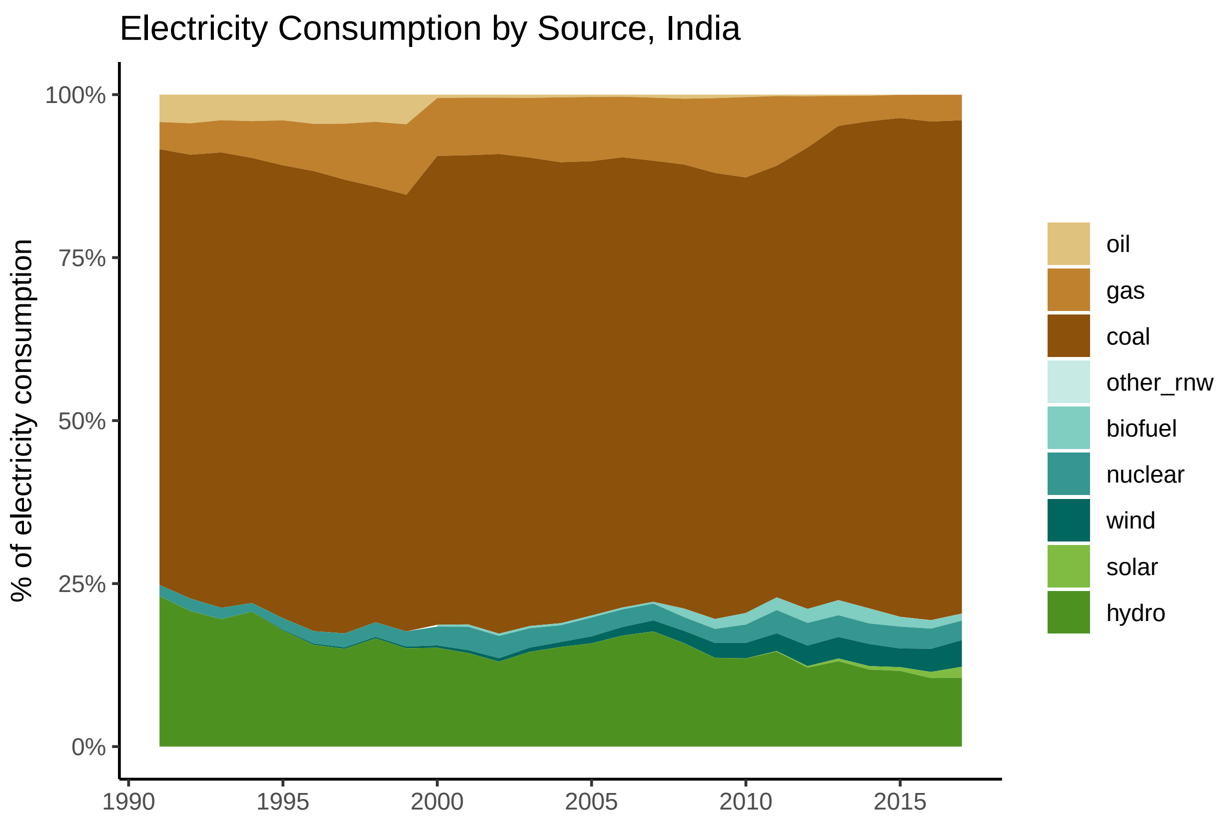
```
## Warning: Removed 18 rows containing missing values (position_stack).
```




```
World_Energy_Consumption_Main%>%
  filter(year < 2019)%>%
  filter(country == "India")%>%
  select(year, coal_share_elec , gas_share_elec, oil_share_elec, hydro_share_elec, solar_share_elec, wi
  biofuel_share_elec,other_renewables_share_elec_exc_biofuel)%>%
  rename(coal = coal_share_elec,
         gas = gas_share_elec,
         oil = oil_share_elec,
         hydro = hydro_share_elec,
         solar = solar_share_elec,
         wind = wind_share_elec,
         nuclear = nuclear_share_elec,
         biofuel = biofuel_share_elec,
         other_rnw = other_renewables_share_elec_exc_biofuel)%>%
pivot_longer(cols = coal:other_rnw, names_to = "elec_source", values_to = "percent_of_elec")%>%
ggplot(aes(year, percent_of_elec, fill = factor(elec_source, levels = rev(y))))+
  geom_area(position = "fill")+
  theme_classic()+
  theme(legend.position = "right",
        legend.title=element_blank())+
```

```
  guides(color = guide_legend(nrow = 1),
         shape = guide_legend(override.aes = list(size = 0.5)))+
  labs(title = "Electricity Consumption by Source, India")+
  ylab("% of electricity consumption")+
  xlab("")+
  scale_fill_manual(values = rev(yco),
                    labels = rev(y))+
  scale_y_continuous(labels = scales::percent)+
  scale_x_continuous(breaks = c(seq(from = 1990, to = 2020, by= 5)))
```

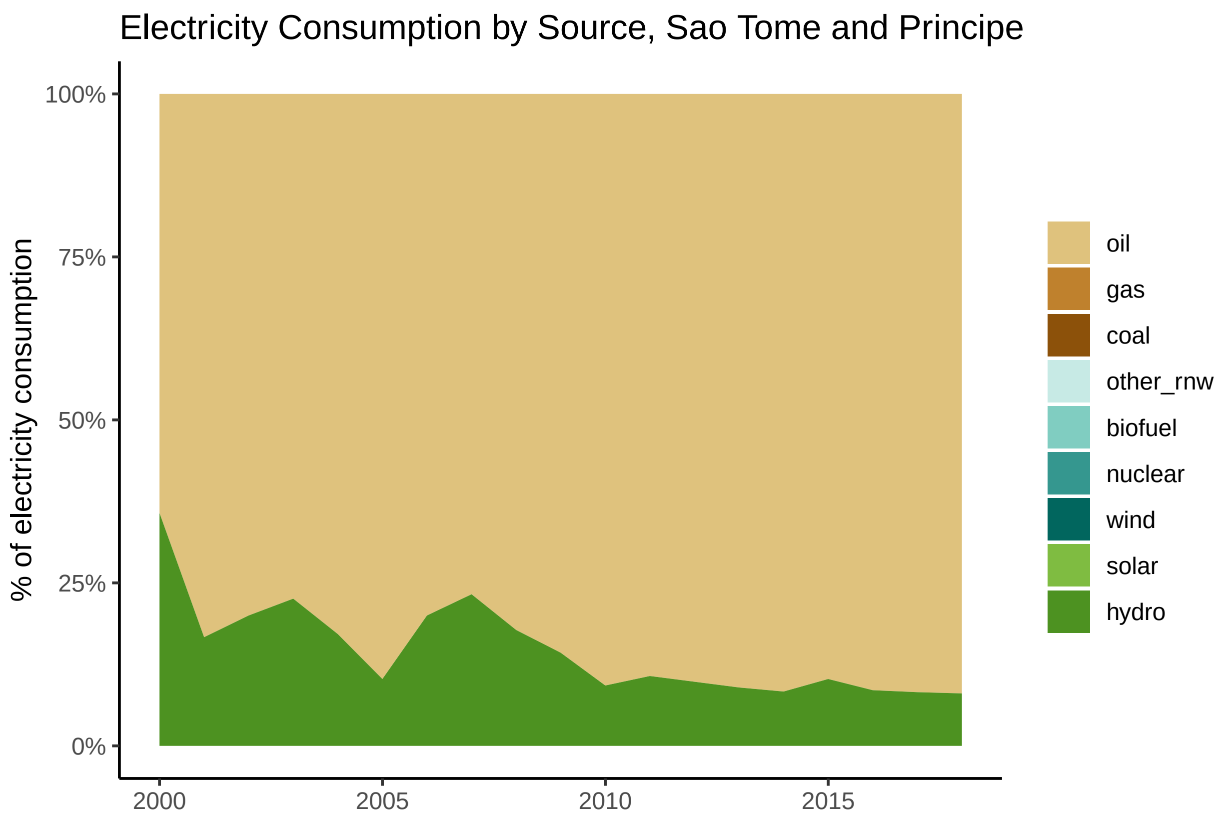
```
## Warning: Removed 18 rows containing missing values (position_stack).
```




```
World_Energy_Consumption_Main%>%
  filter(year < 2019)%>%
  filter(country == "Sao Tome and Principe")%>%
  select(year, coal_share_elec , gas_share_elec, oil_share_elec, hydro_share_elec, solar_share_elec, wi
  biofuel_share_elec,other_renewables_share_elec_exc_biofuel)%>%
  rename(coal = coal_share_elec,
         gas = gas_share_elec,
         oil = oil_share_elec,
         hydro = hydro_share_elec,
         solar = solar_share_elec,
         wind = wind_share_elec,
         nuclear = nuclear_share_elec,
         biofuel = biofuel_share_elec,
         other_rnw = other_renewables_share_elec_exc_biofuel)%>%
```

```
pivot_longer(cols = coal:other_rnw, names_to = "elec_source", values_to = "percent_of_elec")%>%
ggplot(aes(year, percent_of_elec, fill = factor(elec_source, levels = rev(y))))+
  geom_area(position = "fill")+
  theme_classic()+
  theme(legend.position = "right",
        legend.title=element_blank())+
  guides(color = guide_legend(nrow = 1),
         shape = guide_legend(override.aes = list(size = 0.5)))+
  labs(title = "Electricity Consumption by Source, Sao Tome and Principe")+
  ylab("% of electricity consumption")+
  xlab("")+
  scale_fill_manual(values = rev(yco),
                    labels = rev(y))+
  scale_y_continuous(labels = scales::percent)+
  scale_x_continuous(breaks = c(seq(from = 1990, to = 2020, by= 5)))
```

```
## Warning: Removed 81 rows containing missing values (position_stack).
```

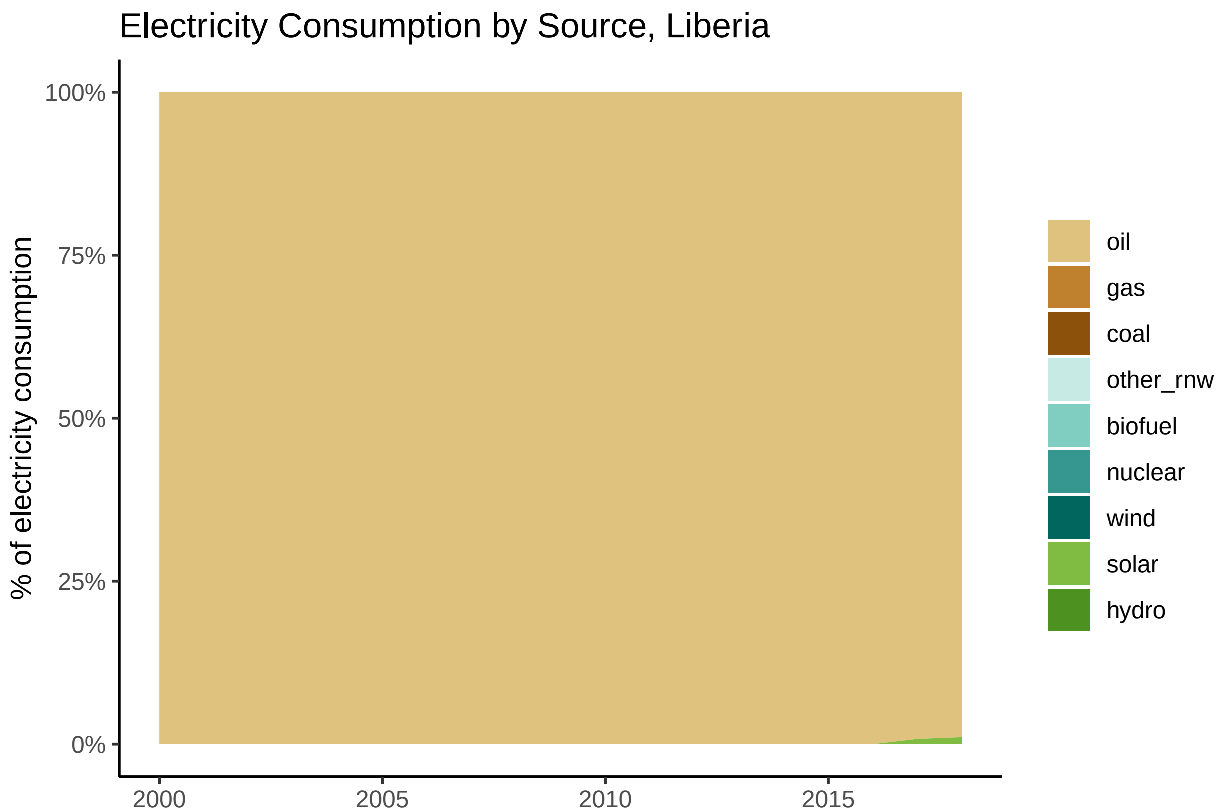



```
#######

World_Energy_Consumption_Main%>%
  filter(year < 2019)%>%
  filter(country == "Liberia")%>%
  select(year, coal_share_elec , gas_share_elec, oil_share_elec, hydro_share_elec, solar_share_elec, wi
  biofuel_share_elec,other_renewables_share_elec_exc_biofuel)%>%
  rename(coal = coal_share_elec,
```

```r
         gas = gas_share_elec,
         oil = oil_share_elec,
         hydro = hydro_share_elec,
         solar = solar_share_elec,
         wind = wind_share_elec,
         nuclear = nuclear_share_elec,
         biofuel = biofuel_share_elec,
         other_rnw = other_renewables_share_elec_exc_biofuel)%>%
pivot_longer(cols = coal:other_rnw, names_to = "elec_source", values_to = "percent_of_elec")%>%
ggplot(aes(year, percent_of_elec, fill = factor(elec_source, levels = rev(y))))+
  geom_area(position = "fill")+
  theme_classic()+
  theme(legend.position = "right",
        legend.title=element_blank())+
  guides(color = guide_legend(nrow = 1),
         shape = guide_legend(override.aes = list(size = 0.5)))+
  labs(title = "Electricity Consumption by Source, Liberia")+
  ylab("% of electricity consumption")+
  xlab("")+
  scale_fill_manual(values = rev(yco),
                    labels = rev(y))+
  scale_y_continuous(labels = scales::percent)+
  scale_x_continuous(breaks = c(seq(from = 1990, to = 2020, by= 5)))
```

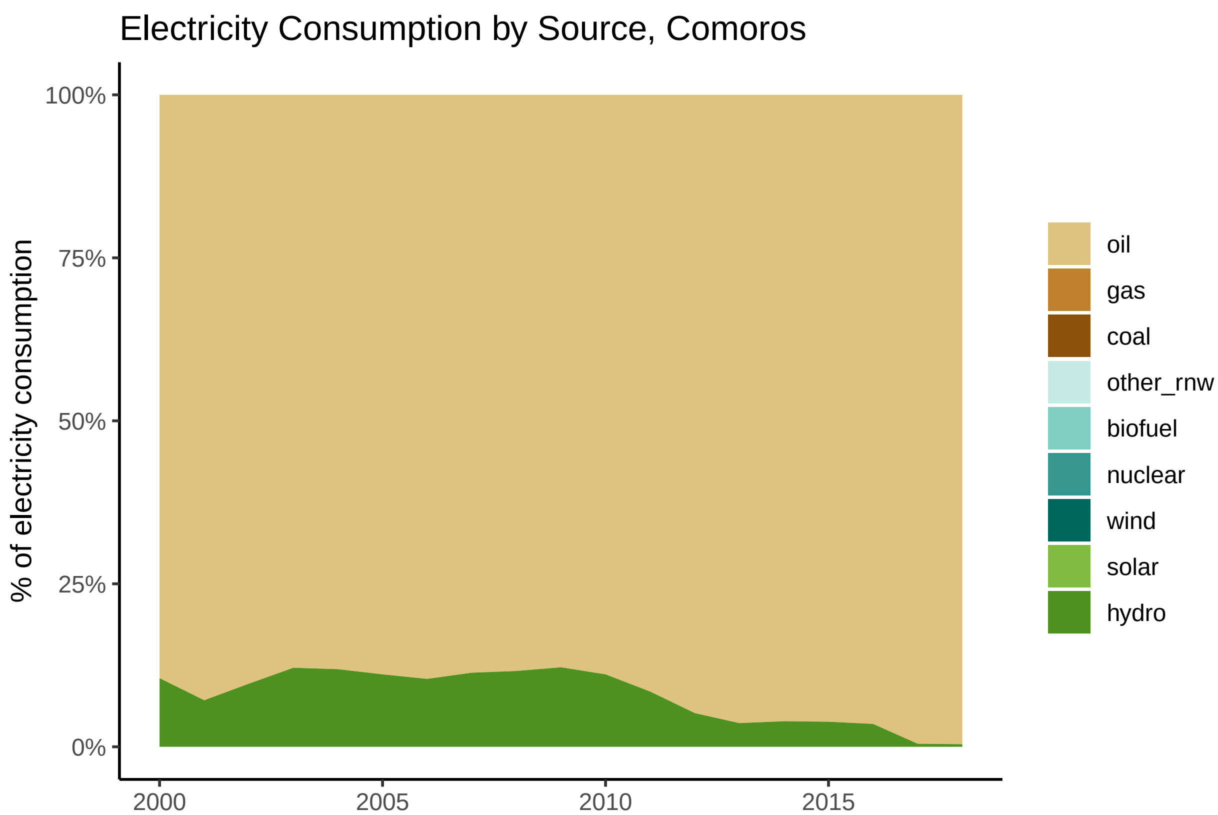
```
## Warning: Removed 81 rows containing missing values (position_stack).
```



```
#######

World_Energy_Consumption_Main%>%
  filter(year < 2019)%>%
  filter(country == "Comoros")%>%
  select(year, coal_share_elec , gas_share_elec, oil_share_elec, hydro_share_elec, solar_share_elec, wi
  biofuel_share_elec,other_renewables_share_elec_exc_biofuel)%>%
  rename(coal = coal_share_elec,
         gas = gas_share_elec,
         oil = oil_share_elec,
         hydro = hydro_share_elec,
         solar = solar_share_elec,
         wind = wind_share_elec,
         nuclear = nuclear_share_elec,
         biofuel = biofuel_share_elec,
         other_rnw = other_renewables_share_elec_exc_biofuel)%>%
pivot_longer(cols = coal:other_rnw, names_to = "elec_source", values_to = "percent_of_elec")%>%
ggplot(aes(year, percent_of_elec, fill = factor(elec_source, levels = rev(y))))+
  geom_area(position = "fill")+
  theme_classic()+
  theme(legend.position = "right",
        legend.title=element_blank())+
  guides(color = guide_legend(nrow = 1),
         shape = guide_legend(override.aes = list(size = 0.5)))+
  labs(title = "Electricity Consumption by Source, Comoros")+
  ylab("% of electricity consumption")+
  xlab("")+
  scale_fill_manual(values = rev(yco),
                    labels = rev(y))+
  scale_y_continuous(labels = scales::percent)+
  scale_x_continuous(breaks = c(seq(from = 1990, to = 2020, by= 5)))
```

```
## Warning: Removed 81 rows containing missing values (position_stack).
```



Conclusion

- Share Fossil Fuels and GDP per capita: There is clear correlation between the primary share of fossil fuels in a country and GDP. As shown in the table above, countries with lower GDP rely more heavily on fossil fuels as an energy source than countries with higher GDP. This highlights the challenge that developing countries face of transitioning from dirtier forms of energy at an earlier stage of industrialization than their western counterparts had to.
- Share Low Carbon and GDP per capita: This graph should be thought of as the compliment to the one before. There is clear correlation between the primary share of low carbon in a country and GDP. As shown in the table above, countries with lower GDP rely less heavily on fossil fuels as an energy source than countries with higher GDP.
- Welfare indicators and renewable energy:
  - Infant Mortality Rate: Countries with a lower share of energy from low carbon have a higher rate of infant mortality.
  - Literacy Rate: Countries with a lower share of energy from low carbon have a lower literacy rate than countries that rely more heavily on low carbon energy.
  - Poverty Rate: Countries with a lower share of energy from low carbon have a higher poverty rate than countries that rely more heavily on low carbon energy.
- GDP and Electricity Usage:
  - Top 3 GDP Countries: A mixture of several different energy sources. Heavy use of coal as well as hydro, nuclear, oil and other renewable energy sources. -The United States has undergone

- Bottom 3 GDP Countries: Stark contrast from the previous example. Theses countries are very reliant on a single electricity source - oil. Sao Tome and Principe and Comoros use a little bit of hydro power as well.

There is definitely a relationship between energy usage and social and economic indicators. From this analysis it seems that there are significant benefits both in terms of social and economic well being for countries who diversify their energy usage and invest in renewable energy. With that being said, the countries with the highest GDP rankings do not exclusively use renewable energy sources and therefore, it is fair to assume that there are benefits to using cheaper, more polluting energy sources as well.

Sources of bias

We believe that climate change is one of the most pressing issues of our time. We recognize that this is a potential source for bias and I am likely to favor renewable energy sources and other measures that we believe will help slow this crisis. Another potential bias is the expectation from being familiar with economics, that wealthier countries tend to diversify their energy sources and might therefore appear cleaner, even though the total level of energy consumption is significantly higher than that of developing nations.

System info

```
sessionInfo()
```

```
## R version 4.1.1 (2021-08-10)
## Platform: x86_64-apple-darwin17.0 (64-bit)
## Running under: macOS Big Sur 10.16
##
## Matrix products: default
## BLAS:   /Library/Frameworks/R.framework/Versions/4.1/Resources/lib/libRblas.0.dylib
## LAPACK: /Library/Frameworks/R.framework/Versions/4.1/Resources/lib/libRlapack.dylib
##
## locale:
## [1] en_US.UTF-8/en_US.UTF-8/en_US.UTF-8/C/en_US.UTF-8/en_US.UTF-8
##
## attached base packages:
## [1] stats     graphics  grDevices utils     datasets  methods   base
##
## other attached packages:
## [1] forcats_0.5.1   stringr_1.4.0   dplyr_1.0.7     purrr_0.3.4
## [5] readr_2.1.2     tidyr_1.2.0     tibble_3.1.6    ggplot2_3.3.5
## [9] tidyverse_1.3.1
##
## loaded via a namespace (and not attached):
##  [1] Rcpp_1.0.8       lattice_0.20-45  lubridate_1.8.0  assertthat_0.2.1
##  [5] digest_0.6.29    utf8_1.2.2       R6_2.5.1         cellranger_1.1.0
##  [9] backports_1.4.1  reprex_2.0.1     evaluate_0.14    highr_0.9
## [13] httr_1.4.2       pillar_1.7.0     rlang_1.0.1      curl_4.3.2
## [17] readxl_1.3.1     rstudioapi_0.13  Matrix_1.4-0     rmarkdown_2.11
## [21] labeling_0.4.2   splines_4.1.1    bit_4.0.4        munsell_0.5.0
## [25] broom_0.7.12     compiler_4.1.1   modelr_0.1.8     xfun_0.29
## [29] pkgconfig_2.0.3  mgcv_1.8-38      htmltools_0.5.2  tidyselect_1.1.1
## [33] fansi_1.0.2      crayon_1.4.2     tzdb_0.2.0       dbplyr_2.1.1
## [37] withr_2.4.3      grid_4.1.1       nlme_3.1-155     jsonlite_1.7.3
## [41] gtable_0.3.0     lifecycle_1.0.1  DBI_1.1.2        magrittr_2.0.2
## [45] scales_1.1.1     cli_3.1.1        stringi_1.7.6    vroom_1.5.7
## [49] farver_2.1.0     fs_1.5.2         xml2_1.3.3       ellipsis_0.3.2
```

```
## [53] generics_0.1.2   vctrs_0.3.8      tools_4.1.1      bit64_4.0.5
## [57] glue_1.6.1       hms_1.1.1        parallel_4.1.1   fastmap_1.1.0
## [61] yaml_2.2.2       colorspace_2.0-2 rvest_1.0.2      knitr_1.37
## [65] haven_2.4.3
```