Energy Use and Economic Development Project: Preprocessing and Data Wrangling

Ben Holden & Lizzy MacIntosh

2022-02-10

Part 1: Importing and merging datasets:

```
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. This is measured in kilowatt-hours per
          energy_per_capita, # Primary energy consumption per capita, measured in kilowatt-hours per ye
          energy_cons_change_pct, #Annual percentage change in primary energy consumption
#       ------------- Fossil fuels: ------------------------------------------------
          fossil_cons_change_pct, #Annual percentage change in fossil fuel consumption
          fossil_share_energy, #Share of primary energy consumption that comes from fossil fuels
          biofuel_share_energy, #Share of primary energy consumption that comes from biofuels
          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 (renewables and nuclear): -------------------------------------------
          low_carbon_cons_change_pct, #Annual percentage change in low-carbon energy consumption
          low_carbon_share_energy, #Share of primary energy consumption that comes from low-carbon sour
          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
          nuclear_share_energy, #Share of primary energy consumption that comes from other renewables
          other_renewables_share_energy #Share of primary energy consumption that comes from other rene
        ) %>%
#       ------------- 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)%>%
#       ------------- Selecting years from 1991 and onwards: -----------------------------------------
    filter(year >= 1991) -> wec_tidy

glimpse(wec_tidy)
```

```
## Rows: 7,011
## Columns: 23
## $ iso_code                      <chr> "AFG", "AFG", "AFG", "AFG", "AFG", "AFG"~
```

```
## $ country                        <chr> "Afghanistan", "Afghanistan", "Afghanist~
## $ year                           <dbl> 1991, 1992, 1993, 1994, 1995, 1996, 1997~
## $ population                     <dbl> 13299016, 14485543, 15816601, 17075728, ~
## $ gdp                            <dbl> 24076392448, 24472467456, 18363240448, 1~
## $ energy_per_gdp                 <dbl> 0.567, 0.366, 0.487, 0.604, 0.338, 0.351~
## $ gdp_per_capita                 <dbl> 1810.3890, 1689.4408, 1161.0105, 835.758~
## $ gdp_per_capita_growth          <dbl> -0.1685820723, -0.0668078545, -0.3127841~
## $ energy_per_capita              <dbl> 1026.439, 618.620, 564.895, 504.612, 400~
## $ energy_cons_change_pct         <dbl> -56.839, -34.352, -0.294, -3.561, -15.90~
## $ fossil_cons_change_pct         <dbl> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, ~
## $ fossil_share_energy            <dbl> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, ~
## $ biofuel_share_energy           <dbl> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, ~
## $ coal_share_energy              <dbl> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, ~
## $ gas_share_energy               <dbl> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, ~
## $ oil_share_energy               <dbl> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, ~
## $ low_carbon_cons_change_pct     <dbl> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, ~
## $ low_carbon_share_energy        <dbl> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, ~
## $ hydro_share_energy             <dbl> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, ~
## $ solar_share_energy             <dbl> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, ~
## $ wind_share_energy              <dbl> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, ~
## $ nuclear_share_energy           <dbl> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, ~
## $ other_renewables_share_energy <dbl> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, ~
```

```
#      ------------- Pivoting years and selecting obsv. starting with year 1991: ----------------------
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)%>%
  filter(year >= 1991 ) -> co2_tidy

head(co2_tidy, 5)
```

```
## # A tibble: 5 x 4
##   `Country Code` `Country Name` year  co2_tons_per_capita
##   <chr>          <chr>          <chr>               <dbl>
## 1 ABW            Aruba          1991                   NA
## 2 ABW            Aruba          1992                   NA
## 3 ABW            Aruba          1993                   NA
## 4 ABW            Aruba          1994                   NA
## 5 ABW            Aruba          1995                   NA
```

```
#      ------------- Pivoting years and selecting obsv. starting with year 1991: ----------------------
glimpse(fertility)%>%
  pivot_longer(cols = `1960`:`2020`, names_to = "year", values_to = "fertility_rate")%>%
  select(`Country Code`, `Country Name`, year, fertility_rate)%>%
  filter(year >= 1991) -> fertility_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> "Fertility rate, total (births per woman)", "Fertilit~
## $ `Indicator Code` <chr> "SP.DYN.TFRT.IN", "SP.DYN.TFRT.IN", "SP.DYN.TFRT.IN",~
## $ `1960`           <dbl> 4.820000, 6.723308, 7.450000, 6.439002, 6.708000, 6.4~
## $ `1961`           <dbl> 4.655000, 6.738651, 7.450000, 6.455523, 6.790000, 6.4~
## $ `1962`           <dbl> 4.471000, 6.752818, 7.450000, 6.471399, 6.872000, 6.2~
## $ `1963`           <dbl> 4.271000, 6.765400, 7.450000, 6.487246, 6.954000, 6.1~
```

```
## $ `1964`          <dbl> 4.059000, 6.775406, 7.450000, 6.502619, 7.036000, 5.9~
## $ `1965`          <dbl> 3.842000, 6.783357, 7.450000, 6.519050, 7.116000, 5.7~
## $ `1966`          <dbl> 3.625000, 6.789885, 7.450000, 6.537615, 7.194000, 5.5~
## $ `1967`          <dbl> 3.417000, 6.796040, 7.450000, 6.560078, 7.267000, 5.3~
## $ `1968`          <dbl> 3.226000, 6.801478, 7.450000, 6.586665, 7.332000, 5.2~
## $ `1969`          <dbl> 3.054000, 6.806248, 7.450000, 6.617137, 7.388000, 5.0~
## $ `1970`          <dbl> 2.908000, 6.810063, 7.450000, 6.650564, 7.434000, 4.9~
## $ `1971`          <dbl> 2.788000, 6.811403, 7.450000, 6.685931, 7.467000, 4.7~
## $ `1972`          <dbl> 2.691000, 6.810668, 7.450000, 6.721698, 7.488000, 4.6~
## $ `1973`          <dbl> 2.613000, 6.806765, 7.450000, 6.755679, 7.498000, 4.5~
## $ `1974`          <dbl> 2.552000, 6.799593, 7.450000, 6.785903, 7.500000, 4.3~
## $ `1975`          <dbl> 2.506000, 6.789526, 7.450000, 6.811647, 7.494000, 4.2~
## $ `1976`          <dbl> 2.472000, 6.776643, 7.450000, 6.832240, 7.485000, 4.0~
## $ `1977`          <dbl> 2.446000, 6.763316, 7.449000, 6.847960, 7.475000, 3.9~
## $ `1978`          <dbl> 2.425000, 6.748156, 7.449000, 6.858350, 7.467000, 3.8~
## $ `1979`          <dbl> 2.408000, 6.731254, 7.449000, 6.863557, 7.461000, 3.7~
## $ `1980`          <dbl> 2.392000, 6.712571, 7.449000, 6.862295, 7.459000, 3.6~
## $ `1981`          <dbl> 2.377000, 6.690097, 7.449000, 6.854195, 7.459000, 3.5~
## $ `1982`          <dbl> 2.364000, 6.663042, 7.450000, 6.838568, 7.461000, 3.4~
## $ `1983`          <dbl> 2.353000, 6.633644, 7.452000, 6.816309, 7.462000, 3.3~
## $ `1984`          <dbl> 2.342000, 6.597501, 7.455000, 6.788058, 7.459000, 3.3~
## $ `1985`          <dbl> 2.332000, 6.555030, 7.458000, 6.754004, 7.451000, 3.2~
## $ `1986`          <dbl> 2.320000, 6.505686, 7.460000, 6.713787, 7.435000, 3.2~
## $ `1987`          <dbl> 2.307000, 6.447606, 7.461000, 6.668966, 7.409000, 3.1~
## $ `1988`          <dbl> 2.291000, 6.383627, 7.461000, 6.620663, 7.373000, 3.1~
## $ `1989`          <dbl> 2.272000, 6.311511, 7.461000, 6.570140, 7.328000, 3.0~
## $ `1990`          <dbl> 2.249000, 6.234901, 7.466000, 6.518146, 7.272000, 2.9~
## $ `1991`          <dbl> 2.221000, 6.153274, 7.479000, 6.465488, 7.208000, 2.9~
## $ `1992`          <dbl> 2.187000, 6.072896, 7.502000, 6.412092, 7.138000, 2.8~
## $ `1993`          <dbl> 2.149000, 5.997355, 7.535000, 6.358930, 7.065000, 2.7~
## $ `1994`          <dbl> 2.108000, 5.927071, 7.572000, 6.306824, 6.990000, 2.6~
## $ `1995`          <dbl> 2.064000, 5.861961, 7.606000, 6.256945, 6.918000, 2.5~
## $ `1996`          <dbl> 2.021000, 5.800563, 7.629000, 6.208205, 6.851000, 2.5~
## $ `1997`          <dbl> 1.978000, 5.741203, 7.632000, 6.161531, 6.789000, 2.4~
## $ `1998`          <dbl> 1.939000, 5.684033, 7.610000, 6.116602, 6.732000, 2.3~
## $ `1999`          <dbl> 1.903000, 5.628409, 7.561000, 6.073128, 6.683000, 2.2~
## $ `2000`          <dbl> 1.872000, 5.573421, 7.485000, 6.031044, 6.639000, 2.1~
## $ `2001`          <dbl> 1.846000, 5.519798, 7.387000, 5.989972, 6.601000, 2.0~
## $ `2002`          <dbl> 1.823000, 5.468439, 7.272000, 5.949391, 6.567000, 1.9~
## $ `2003`          <dbl> 1.803000, 5.418055, 7.148000, 5.908659, 6.533000, 1.8~
## $ `2004`          <dbl> 1.787000, 5.369067, 7.016000, 5.867800, 6.499000, 1.8~
## $ `2005`          <dbl> 1.774000, 5.319320, 6.875000, 5.826674, 6.461000, 1.7~
## $ `2006`          <dbl> 1.765000, 5.267324, 6.722000, 5.784452, 6.419000, 1.7~
## $ `2007`          <dbl> 1.760000, 5.213181, 6.555000, 5.741847, 6.372000, 1.6~
## $ `2008`          <dbl> 1.759000, 5.155493, 6.373000, 5.698842, 6.320000, 1.6~
## $ `2009`          <dbl> 1.761000, 5.093841, 6.180000, 5.654398, 6.260000, 1.6~
## $ `2010`          <dbl> 1.768000, 5.027698, 5.977000, 5.607672, 6.194000, 1.6~
## $ `2011`          <dbl> 1.779000, 4.956842, 5.770000, 5.557872, 6.120000, 1.6~
## $ `2012`          <dbl> 1.795000, 4.882058, 5.562000, 5.503781, 6.039000, 1.6~
## $ `2013`          <dbl> 1.813000, 4.804516, 5.359000, 5.446144, 5.953000, 1.6~
## $ `2014`          <dbl> 1.834000, 4.726220, 5.163000, 5.384336, 5.864000, 1.6~
## $ `2015`          <dbl> 1.854000, 4.647637, 4.976000, 5.319473, 5.774000, 1.6~
## $ `2016`          <dbl> 1.872000, 4.569675, 4.800000, 5.251674, 5.686000, 1.6~
## $ `2017`          <dbl> 1.886000, 4.493744, 4.633000, 5.182391, 5.600000, 1.6~
```

```
## $ `2018`           <dbl> 1.896000, 4.420264, 4.473000, 5.113003, 5.519000, 1.6~
## $ `2019`           <dbl> 1.901000, 4.349433, 4.321000, 5.044144, 5.442000, 1.5~
## $ `2020`           <lgl> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, N~
## $ ...66            <lgl> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, N~
```

```
  head(fertility_tidy, 5)
```

```
## # A tibble: 5 x 4
##   `Country Code` `Country Name` year  fertility_rate
##   <chr>          <chr>          <chr>          <dbl>
## 1 ABW            Aruba          1991            2.22
## 2 ABW            Aruba          1992            2.19
## 3 ABW            Aruba          1993            2.15
## 4 ABW            Aruba          1994            2.11
## 5 ABW            Aruba          1995            2.06
```