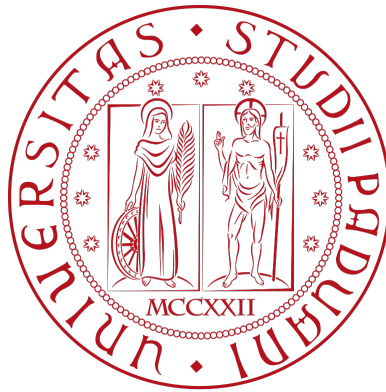University of Padua

Departments of Mathematics

Master Degree in Data Science

Low-Carbon Electricity Generation:
A Comprehensive Analysis

Valerio Rocca, 2094861

Giacomo Virginio, 2076681

Academic Year 2022/2023

Contents

Chapter 1

Abstract

This is the abstract

Chapter 2

Obtaining data

We could not find a single dataset containing all the information of interest. Thus, the project's first step is merging multiple datasets into one. The primary one is the *Energy dataset* by Our World in Data (from now onwards referred to as "OWID") [1], which contains various time series for each world country regarding energy and electricity production and consumption.

```
# Import the Energy dataset
main = read.csv("datasets//Total_energy_data.csv")
```

We then merge the following datasets into it:

- *GDP (constant 2015 US$)* by World Bank [2], which contains the time series of the GDP in each country from 1960 to 2021, measured in constant 2015 USA dollars.
- *Land Area* by OWID [3], which contains the time series of the land area of each country from 1961 to 2021, measured in squared kilometers;
- *Agricultural land* by OWID [4], which contains the time series of the share of land area used for agriculture in each country from 1961 to 2018;
- *Urbanization rate* by OWID [5], which contains the time series of the share of people living in urban areas in each country from 1960 to 2020;
- *Human Development Index* by OWID [6], which contains the time series of the HDI for each country from 1990 to 2021;
- *Death rate from air pollution* by OWID [7], which after filtering contains the time series of the number of deaths from outdoor particulate matter per 100,000 population in each country from 1990 to 2019;
- *Coal proved reserves* by OWID [8], which contains the reserves of coal in each country in 2021, measured in tonnes;
- *Oil proved reserves* by OWID [9], which contains the reserves of oil in each country in 2020, measured in tonnes;
- *Natural gas proved reserves* by OWID [10], which contains the time series of the reserves of natural gas in each country from 1980 to 2020 (CHECK BC THE LINK HAS ONLY 2020);
- *Uranium proved reserves* by OECD [11], which contains uranium reserves in each country in 2019, measured in tonnes.

Merging presents three critical issues, which are listed below, together with the implemented solutions.

1. Time series are recorded for different years. We tackled this problem by merging through left join: all the rows of the *Energy dataset* are included, while rows from the other datasets are included if there is a match; otherwise, a NA value is added. In order to perform correctly the left join, we need to remove the countries in *Energy dataset* without an ISO code. This is the case for some semi-autonomous territories inside of a country (e.g., Wake Island), countries that no longer exist (e.g., Yugoslavia), and country groupings (e.g., OPEC countries). We decide to remove those observations.
2. Coal, oil, and uranium reserves are stationary values, as time series for those variables are not publicly available. Therefore, we approached the issue by considering the reserves fixed through time, as it does not affect the quality of the analyses.
3. *GDP (constant 2015 US$)* dataset contains a column for each year, while the other time series datasets format the years using a specific variable. We modify the structure of *GDP* to fit the others'.

```
# Delete units from "main" without an ISO code
main = main[main$iso_code!='',]

# Creation of a function able to automatically join datasets from OWID
join_owid = function(main_data, secondary_dataset_link){
  place_data = read.csv(paste(secondary_dataset_link))
  main_data = left_join(main_data, select(place_data, -c("Entity")),
                by = c("iso_code" = "Code", "year" = "Year"))
  rm(place_data)
```

```r
    return(main_data)
}

# Import and merging of OWID variables, except of air pollution
main = join_owid(main, "datasets//country_areas.csv")
main = join_owid(main, "datasets//share_of_land_area_used_for_agriculture.csv")
main = join_owid(main, "datasets//urbanization_rate.csv")
main = join_owid(main, "datasets//human_development_index.csv")
main = join_owid(main, "datasets//oil_proved_reserves.csv")
main = join_owid(main, "datasets//natural_gas_proved_reserves.csv")

# Import, filtering and merging of coal proved reserves
coal_res = read.csv("datasets//coal_proved_reserves.csv")
coal_res = coal_res[coal_res$Year!="2020",]
colnames(coal_res) = c("Entity", "Code", "Year", "coal_reserves_2021")
main = left_join(main, select(coal_res, -c("Entity","Year")),
                 by = c("iso_code" = "Code"))
rm(coal_res)

# Import and merging of uranium reserves
uranium_res = read.csv("datasets//uranium_proved_reserves.txt", sep = "\t")
colnames(uranium_res) = c("V1","V2", "uranium_reserves_2019", "V4")
uranium_res$uranium_reserves_2019 = as.numeric(gsub(",","",uranium_res$uranium_reserves_2019))
uranium_res$V1 = sub(".","",uranium_res$V1)
main = left_join(main, select(uranium_res, -c("V2", "V4")),
                 by = c("country" = "V1"))
rm(uranium_res)

# Import the GDP dataset
gdp = read_excel("datasets//gdp_constant_2015_dollars.xlsx")[,4:66]
# Structure modification
colnames(gdp) = c("code", 1960:2021)
gdp = gather(gdp, key = "year", value = "gdp", -code)
gdp$year = as.integer(gdp$year)
# Merging
main = left_join(main, gdp, by = c("iso_code" = "code", "year" = "year"))
rm(gdp)

# Import, selection and merging of deaths from air pollution
 death_rates = read.csv("datasets//death_rates_from_air_pollution.csv")[, c(2,3,5)]
 colnames(death_rates) = c("Code", "Year", "particulate_pollution")
 main = left_join(main, death_rates,
                 by = c("iso_code" = "Code", "year" = "Year"))
rm(death_rates)
```

Chapter 3

Data pre-processing

In this section, we present the pre-processing activities performed.

1. **Units selection**, computed over the *Energy dataset*. Already partially computed and explained in the previous section, in this phase we also removed two regions with too many missing values: Anctartica and Western Sahara.
2. **Feature selection**, computed over the *Energy dataset*. From the original 129 variables, we kept only 36 relevant for the analyses.
3. **Feature renaming**, computed over *main*, as the features merged to the *Energy dataset* have inconvenient names.
4. **Feature addition**, computed over *main* in paragraph 5.3. The new categorical variable groups the world countries into six macroregions.

```
# 1. Units selection: remove Antarctica and Western Sahara
main = filter(main, iso_code != "ATA", iso_code != "ESH")

# 2. Feature selection
main = select(main, -c("gdp.x", "biofuel_cons_change_pct",
                       "biofuel_cons_change_twh", "biofuel_cons_per_capita",
                       "biofuel_elec_per_capita", "biofuel_consumption",
                       "biofuel_electricity", "biofuel_share_elec",
                       "biofuel_share_energy", "coal_cons_change_pct",
                       "coal_cons_change_twh", "coal_cons_per_capita",
                       "coal_elec_per_capita", "coal_prod_change_pct",
                       "coal_prod_change_twh", "coal_prod_per_capita",
                       "coal_consumption", "coal_share_energy",
                       "energy_cons_change_pct", "energy_per_capita",
                       "energy_per_gdp", "electricity_share_energy",
                       "fossil_cons_change_pct", "fossil_cons_change_twh",
                       "fossil_elec_per_capita", "fossil_fuel_consumption",
                       "fossil_share_energy", "gas_cons_change_pct",
                       "gas_cons_change_twh", "gas_elec_per_capita",
                       "gas_prod_change_pct", "gas_prod_change_twh",
                       "gas_prod_per_capita", "gas_consumption",
                       "gas_share_energy", "hydro_cons_change_pct",
                       "hydro_cons_change_twh", "hydro_elec_per_capita",
                       "fossil_energy_per_capita", "hydro_energy_per_capita",
                       "hydro_consumption", "hydro_share_energy",
                       "low_carbon_cons_change_pct", "low_carbon_cons_change_twh",
                       "low_carbon_elec_per_capita", "low_carbon_energy_per_capita",
                       "low_carbon_consumption", "low_carbon_share_energy",
                       "net_elec_imports_share_demand", "nuclear_cons_change_pct",
                       "nuclear_cons_change_twh", "nuclear_elec_per_capita",
                       "nuclear_energy_per_capita", "nuclear_consumption",
                       "nuclear_share_energy", "oil_prod_per_capita",
                       "gas_energy_per_capita", "oil_elec_per_capita",
                       "oil_prod_change_pct", "oil_prod_change_twh",
                       "oil_consumption", "oil_share_energy",
                       "other_renewable_exc_biofuel_electricity", "other_renewables_cons_change_pct",
                       "other_renewables_cons_change_twh", "other_renewables_elec_per_capita",
                       "other_renewables_elec_per_capita_exc_biofuel",
                       "other_renewables_energy_per_capita",
                       "other_renewables_share_elec_exc_biofuel", "other_renewable_consumption",
                       "other_renewables_share_energy", "per_capita_electricity",
```

```r
                        "renewables_cons_change_pct", "renewables_cons_change_twh",
                        "renewables_elec_per_capita", "renewables_energy_per_capita",
                        "renewables_consumption", "renewables_share_energy",
                        "solar_cons_change_pct", "solar_cons_change_twh",
                        "solar_elec_per_capita", "solar_consumption",
                        "solar_share_energy", "wind_cons_change_pct",
                        "wind_cons_change_twh", "wind_consumption",
                        "wind_share_energy", "solar_energy_per_capita",
                        "wind_elec_per_capita", "wind_energy_per_capita",
                        "oil_cons_change_pct", "oil_cons_change_twh",
                        "oil_energy_per_capita"))

# 3. Feature renaming
colnames(main) = c(colnames(main[,1:36]), "land_area", "agri_land_rate",
                   "urbaniz_rate", "hdi", "oil_reserves_2020", "gas_reserves",
                   "coal_reserves_2021", "uranium_reserves_2019", "gdp",
                   "particulate_pollution")
```

Chapter 4

Exploratory analyses

To fill. . . . .

Chapter 5

Descriptive analyses

5.1 Global analyses

We start the descriptive analyses by looking at the global low-carbon electricity generation trends. By **low-carbon** we refer to the electricity produced with substantially lower greenhouse gas emissions than conventional fossil fuel power generation [12]. In other words, the term low-carbon includes the renewable and nuclear sources. From now onward, we will refer to it with the acronym **LC**.

```r
# 1. World electricity generation from LC sources

# a. Area plot of the total generation
# Creation of the dataset
place = main[c("year", "low_carbon_electricity")] %>%
  mutate_all(~replace_na(.,0)) %>%
  group_by(year) %>%
  summarize(sum_lc = sum(low_carbon_electricity))

# Creation of the plot
gg1 = ggplot(place, aes(year, sum_lc)) +
  geom_area(alpha = 0.7, colour="white") +
  scale_x_continuous(limits = c(1965,2020), breaks = seq(1965, 2020, by = 5)) +
  labs(title = "World electricity generation from LC sources",
       x = "Year",
       y = "Terawatt-hours")

# b. Area plot of the ratio between LC generation and total electricity generation
# Creation of the dataset
place = main[c("year", "low_carbon_electricity", "electricity_generation")] %>%
  mutate_all(~replace_na(.,0)) %>%
  group_by(year) %>%
  summarize(sum_lc = sum(low_carbon_electricity)/sum(electricity_generation))

# Creation of the plot
gg2 = ggplot(place, aes(year, sum_lc)) +
  geom_area(alpha = 0.7, colour="white") +
  scale_x_continuous(limits = c(1985,2020), breaks = seq(1985, 2020, by = 5)) +
  scale_y_continuous(limits = c(0,0.5)) +
  labs(title = "% of world electricity generated from LC sources",
       x = "Year",
       y = "% of terawatt-hours")

# Visualization of gg1 and gg2
grid.arrange(gg1,gg2)
```

World electricity generation from LC sources

% of world electricity generated from LC sources
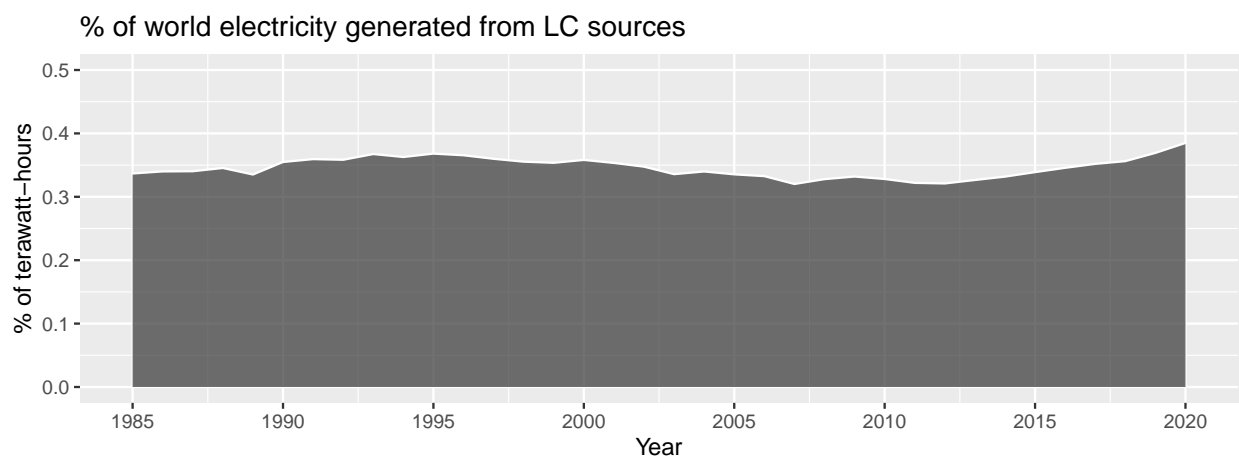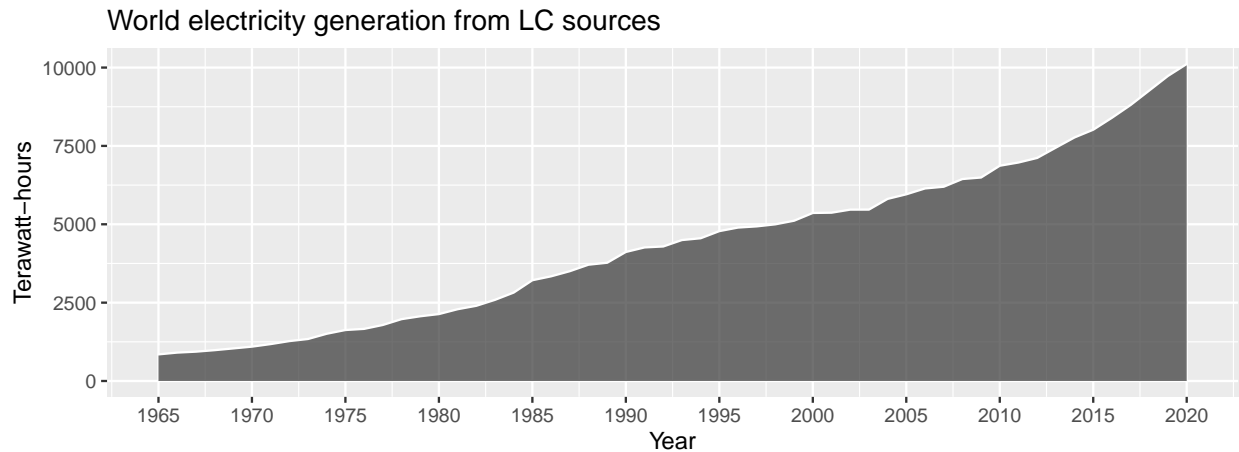
World's electricity production from LC sources constantly grew, going from a generation of less than 1000 TwH in 1965 to more than 10000 TwH in 2020, with an impressive average yearly growth of 21.6%.

Nonetheless, the second image clarifies an important point: even if the electricity generation from LC sources increased, the share of world electricity generated from LC sources has been stationary over the years, except for a timid increase from 2013 to 2020.

```r
# 2. World electricity generation from LC sources, grouped by countries

# a. Creation of a vector containing the ISO codes of the nine countries with the highest
# LC electricity production in 2020
place = main[c("year", "low_carbon_electricity")] %>% mutate_all(~replace_na(.,0))
place = cbind(iso_code = main$iso_code, place) %>%
  filter(year == 2020) %>%
  arrange(desc(low_carbon_electricity))
place_2 = place$iso_code[1:9]

# b. Group the other countries in a single class called "OTH" ("others")
place = main[c("year", "low_carbon_electricity")] %>% mutate_all(~replace_na(.,0))
place = cbind(iso_code = main$iso_code, place)
for(i in 1:nrow(place)){
  place$iso_code[i] = ifelse(place$iso_code[i] %in% place_2, place$iso_code[i], "OTH")
}
place = group_by(place, iso_code, year) %>%
  summarize(sum_lc = sum(low_carbon_electricity))
place$iso_code = factor(place$iso_code,
                        levels = c("CHN", "USA", "BRA", "CAN", "FRA", "RUS",
```

10

```
                              "IND", "DEU", "JPN", "OTH"))

# c. Plot the graph
gg1 = ggplot(place, aes(year, sum_lc, fill = iso_code)) +
  geom_area(alpha=0.6, colour="white") +
  scale_x_continuous(limits = c(1965,2020), breaks = seq(1965, 2020, by = 5)) +
  labs(title = "World electricity generation from LC sources",
       subtitle = "Grouped by country",
       x = "Year",
       y = "Terawatt-hours") +
  scale_fill_viridis_d(name = "Source", option = "E")

# d. Share of electricity production by country in 2020
place = filter(place, year == 2020)
place$sum_lc = round(place$sum_lc / sum(place$sum_lc),2)
place = place[,c(1,3)] %>% arrange(desc(sum_lc)) %>% as.data.frame()
colnames(place) = c("ISO code","% LC generation")

# Plot gg1 and table
gg1
```

World electricity generation from LC sources
Grouped by country
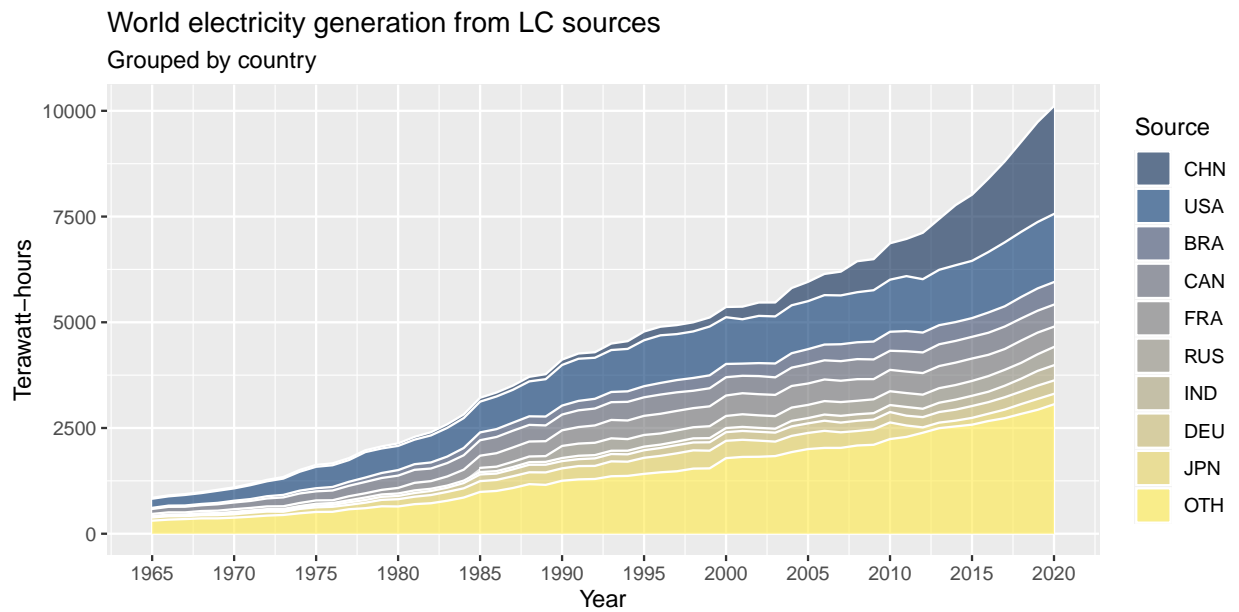

```
kable(cbind(place[1:5,], place[6:10,]),caption = "Share of LC electricity generation by country, 2020")
```

Table 1: Share of LC electricity generation by country, 2020

ISO code	% LC generation	ISO code	% LC generation
OTH	0.30	FRA	0.05
CHN	0.25	IND	0.04
USA	0.16	RUS	0.04
BRA	0.05	DEU	0.03
CAN	0.05	JPN	0.02

Not surprisingly, the generation of electricity from LC sources is not homogeneous between the countries: China and

11

the USA own 41% of the electricity generated in 2020; the nine nations with the highest electricity generation from LC sources account for 69%.

5.2 Analyses by source

```r
# 3. World electricity generation from LC sources, grouped by source

# a. Area plot of the total electricity generation
# Creation of the dataset
place = main[c("year", "hydro_electricity",
                         "nuclear_electricity", "solar_electricity", "wind_electricity",
                         "other_renewable_electricity")] %>%
  mutate_all(~replace_na(.,0)) %>%
  group_by(year) %>%
  summarize(Nuclear = sum(nuclear_electricity),
            Hydro = sum(hydro_electricity),
            Wind = sum(wind_electricity),
            Solar = sum(solar_electricity),
            Other = sum(other_renewable_electricity)) %>%
  gather(key = "Type",
         value = "elect",
         -year)
place$Type = factor(place$Type, levels = c("Nuclear", "Hydro", "Solar",
                                            "Wind", "Other"))

# Creation of the plot
gg1 = ggplot(place, aes(year, elect, fill = Type)) +
  geom_area(alpha=0.6 , size=.5, colour="white")+
  scale_x_continuous(limits = c(1965,2020), breaks = seq(1965, 2020, by = 5)) +
  scale_y_continuous(limits = c(0,10500)) +
  scale_fill_manual(values = c("#B2FF00", "#05B6FF", "#0060FA", "#141BDB", "#00296B")) +
  labs(title = "World electricity generation from LC sources",
       subtitle = "Grouped by source",
       x = "Year",
       y = "Terawatt-hours")

# b. Area plot of the total electricity generation, with sum = 100
# Creation of the dataset
place = main[c("year","hydro_electricity", "wind_electricity",
               "nuclear_electricity", "solar_electricity",
               "other_renewable_electricity")] %>%
  mutate_all(~replace_na(.,0)) %>%
  group_by(year) %>%
  summarize(Nuclear = sum(nuclear_electricity),
            Hydro = sum(hydro_electricity),
            Wind = sum(wind_electricity),
            Solar = sum(solar_electricity),
            Other = sum(other_renewable_electricity)) %>%
  select(Nuclear, Hydro, Wind, Solar,
         Other, Nuclear) %>%
  mutate(year = 1900:2022, row_total = rowSums(.)) %>%
  mutate(across(Nuclear:Other, ~ . / row_total * 100)) %>%
  select(-row_total) %>%
  gather(key = "Type",
         value = "elect",
         -year)
place$Type = factor(place$Type, levels = c("Nuclear", "Hydro", "Solar",
```

```r
                                          "Wind", "Other"))

# Creation of the plot
gg2 = ggplot(place, aes(year, elect, fill = Type)) +
  geom_area(alpha=0.6 , size=.5, colour="white")+
  scale_x_continuous(limits = c(1965,2020), breaks = seq(1965, 2020, by = 5)) +
    scale_fill_manual(values = c("#B2FF00", "#05B6FF", "#0060FA", "#141BDB", "#00296B")) +
      labs(title = "World electricity generation from LC sources",
        subtitle = "Sum = 100, grouped by source",
        x = "Year",
        y = "% of terawatt-hours")

# c. Line plot of the difference of generation with respect to 5 years in the past
# Creation of the dataset
place = main[c("year", "hydro_electricity", "nuclear_electricity",
              "solar_electricity", "wind_electricity",
              "other_renewable_electricity")] %>%
  mutate_all(~replace_na(.,0)) %>%
  filter(year <= 2020) %>%
  group_by(year) %>%
  summarize(Nuclear = sum(nuclear_electricity),
          Hydro = sum(hydro_electricity),
          Wind = sum(wind_electricity),
          Solar = sum(solar_electricity),
          Other = sum(other_renewable_electricity)) %>%
  mutate(Nuclear = (Nuclear - dplyr::lag(Nuclear,5)),
        Hydro = (Hydro - dplyr::lag(Hydro,5)),
        Wind = (Wind - dplyr::lag(Wind,5)),
        Solar = (Solar - dplyr::lag(Solar,5)),
        Other = (Other - dplyr::lag(Other,5))) %>%
  gather(key = "Type",
        value = "world_elect",
        -year)
place$Type = factor(place$Type, levels = c("Nuclear", "Hydro", "Solar",
                                            "Wind", "Other"))

# Creation of the plot
gg3 = ggplot(place, aes(year, world_elect, colour = Type)) +
  geom_line(size = 1) +
  scale_x_continuous(limits = c(1970,2020)) +
  scale_y_continuous() +
  scale_color_manual(values = c("#B2FF00", "#05B6FF", "#0060FA", "#141BDB", "#00296B")) +
    labs(title = "World electricity generation from LC sources",
      subtitle = "Difference with respect to 5 years in the past, grouped by source",
      x = "Year",
      y = "Terawatt-hours")

# Visualization of gg1, gg2 and gg3
grid.arrange(gg1, gg2, gg3)
```
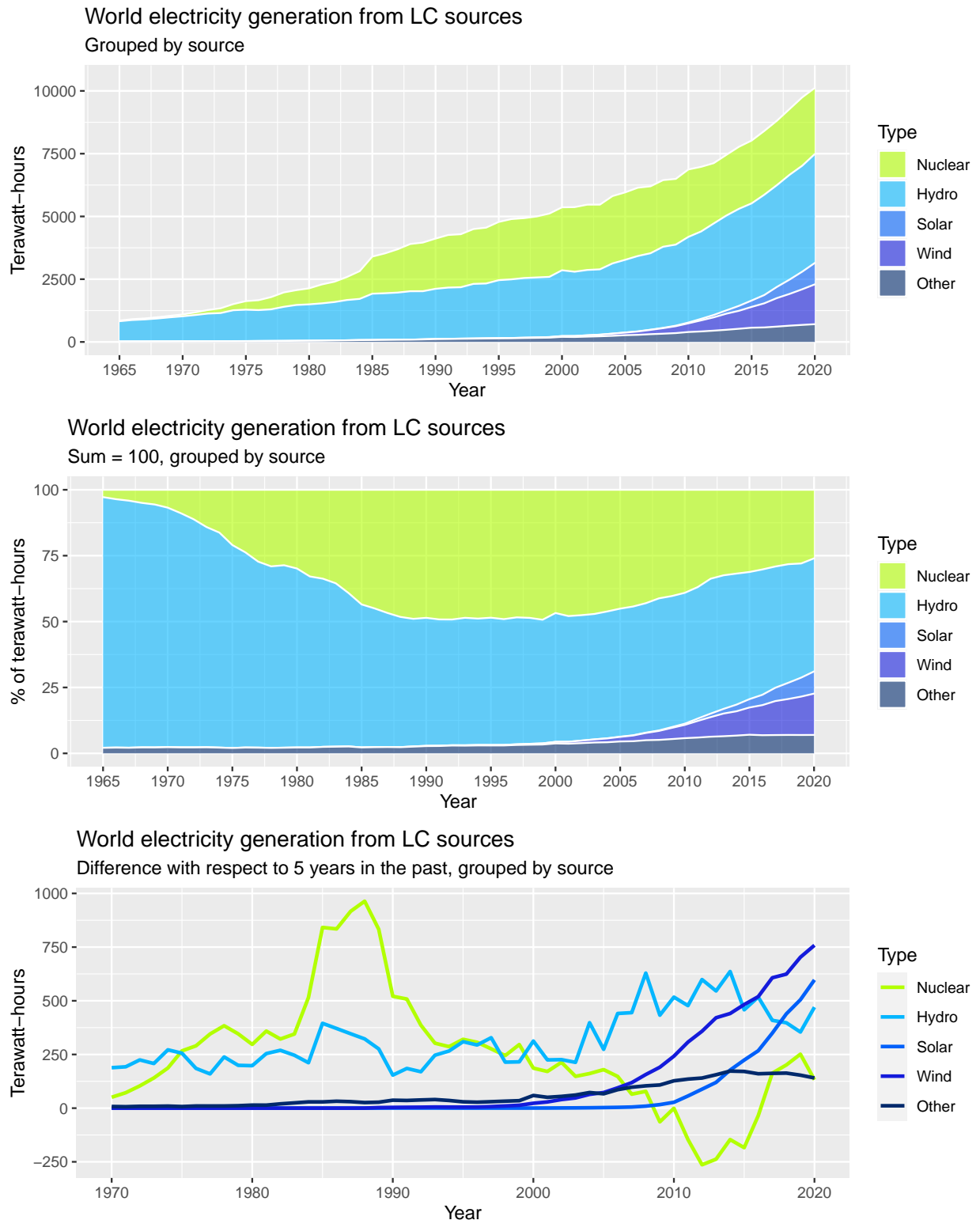
World electricity generation from LC sources
Grouped by source

World electricity generation from LC sources
Sum = 100, grouped by source

World electricity generation from LC sources
Difference with respect to 5 years in the past, grouped by source

The plots allow us to identify three different phases in the history of LC electricity.

1. **Dawn of LC electricity** (up to the mid-'80s). In this era, there are two different trends: on one side, the electricity generated by hydropower grows linearly with respect to the past generation; on the other side, the civil usage of nuclear power takes the first steps.

2. **Golden age of nuclear electricity** (from the mid-'80s to mid-'00s). The electricity generation from nuclear reaches its peak, while the production of hydropower plants continues to grow linearly.

3. **Golden age of renewables** (from the mid-'00s to nowadays). Nuclear power generation declines and gives way to renewables. In particular, the solar and wind generation skyrockets.

5.3 Analyses by macroregion

We are now interested in analizying the generation of electricity by LC sources in different areas. To do so, we aggregate world countries in six macroregions on the basis of geographical, economical and cultural factors.

1. **Developed countries**: Western Europe, Israel, USA, Canada, Australia, New Zealand, Japan, South Korea, Taiwan, Hong Kong and Macao.

2. **Latin America and Caribbean**: North and South America's countries, except of USA and Canada.

3. **Eastern Europe**: former members of the Warsaw Pact (excluding Kazakhstan, Turkmenistan, Uzbekistan, Tajikistan and Kyrgyzstan) and former Yugoslavia.

4. **Middle East and Northern Africa**: Morocco, Algeria, Tunisia, Libya, Egypt, Jordan, Palestine, Lebanon, Syria, Turkey, Iraq, Iran, Kuwait, Saudi Arabia, Yemen, Oman, United Arab Emirates, Bahrain and Qatar.

5. **Sub-Saharan Africa**: non-aforementioned African countries.

6. **Asia**: non-aforementioned Asian countries.

```r
# a. Creation of a vector for each macroregion, containing the ISO-codes
developed_countries = c("AUS", "AUT", "BEL", "CAN", "CYP", "DNK", "FRO", "FIN",
                        "FRA", "DEU", "GRC", "GRL", "HKG", "ISL", "IRL", "ISR",
                        "ITA", "JPN", "LUX", "MAC", "MLT", "NLD", "NZL", "NOR",
                        "PRT", "SPM", "KOR", "ESP", "SWE", "CHE", "TWN", "GBR",
                        "USA", "REU", "GIB")

latin_countries = c("ATG", "ARG", "ABW", "BHS", "BRB", "BLZ", "BMU", "BOL",
                    "BRA", "CYM", "CHL", "COL", "CRI", "CUB", "DMA", "DOM",
                    "ECU", "SLV", "FLK", "GUF", "GRD", "GLP", "GTM", "GUY",
                    "HTI", "HND", "JAM", "MTQ", "MEX", "MSR", "NIC", "PAN",
                    "PRY", "PER", "PRI", "KNA", "LCA", "VCT", "SUR", "TTO",
                    "TCA", "VIR", "URY", "VEN", "VGB", "ANT")

east_europe_countries = c("ALB", "ARM", "AZE", "BLR", "BIH", "BGR", "HRV",
                          "CZE", "EST", "GEO", "HUN", "LVA", "LTU", "MDA",
                          "MNE", "MKD", "POL", "ROU", "RUS", "SRB", "SVK",
                          "SVN", "UKR")

sub_african_countries = c("AGO", "BEN", "BWA", "BFA", "BDI", "CPV", "CMR",
                          "CAF", "TCD", "COM", "COG", "CIV", "COD", "DJI",
                          "GNQ", "ERI", "SWZ", "ETH", "GAB", "GMB", "GHA",
                          "GIN", "GNB", "KEN", "LSO", "LBR", "MDG", "MWI",
                          "MLI", "MRT", "MUS", "MOZ", "NAM", "NER", "NGA",
                          "RWA", "STP", "SEN", "SLE", "SOM", "ZAF", "SSD",
                          "SDN", "TZA", "TGO", "UGA", "ZMB", "ZWE", "SHN")

middle_east_countries = c("DZA", "BHR", "EGY", "IRN", "IRQ", "JOR", "KWT",
                          "LBN", "LBY", "MAR", "OMN", "PSE", "QAT", "SAU",
                          "SYR", "TUR", "ARE", "YEM", "TUN")

asian_countries = c("AFG", "ASM", "BGD", "BTN", "BRN", "KHM", "CHN", "COK",
                    "FJI", "PYF", "GUM", "IND", "IDN", "KAZ", "KIR", "KGZ",
                    "LAO", "MYS", "MDV", "FSM", "MNG", "MMR", "NRU", "NPL",
                    "NCL", "PRK", "MNP", "PAK", "PNG", "PHL", "WSM", "VNM",
                    "SYC", "SGP", "SLB", "LKA", "TJK", "THA", "TLS", "TON",
```

```
                   "TKM", "TUV", "UZB", "VUT", "NIU")

#b. Assign the grouping to each observation in "main"
tag = rep(0, nrow(main))

for(i in 1:length(tag)){
  if(main$iso_code[i] %in% developed_countries){
    tag[i] = "developed"
  }
  else{
    if(main$iso_code[i] %in% latin_countries){
      tag[i] = "latin"
    }
    else{
      if(main$iso_code[i] %in% east_europe_countries){
        tag[i] = "east_europe"
      }
      else{
        if(main$iso_code[i] %in% sub_african_countries){
          tag[i] = "sub_african"
        }
        else{
          if(main$iso_code[i] %in% middle_east_countries){
            tag[i] = "middle_east"
          }
          else{
            if(main$iso_code[i] %in% asian_countries){
              tag[i] = "asian"
            }
          }
        }
      }
    }
  }
}
main = cbind(main, tag)

# c.Plot of the world map
df_asian = data.frame(region = "Asia", tag = asian_countries)
df_east = data.frame(region = "Eastern Europe", tag = east_europe_countries)
df_middle = data.frame(region = "Mid. East & N. Africa", tag = middle_east_countries)
df_dev = data.frame(region = "Developed", tag = developed_countries)
df_africa = data.frame(region ="Sub-Sah. Africa", tag = sub_african_countries)
df_latin = data.frame(region = "Latin A. & Carr.", tag = latin_countries)
df_world = rbind(df_asian, df_east, df_middle, df_dev, df_africa, df_latin)


map = joinCountryData2Map(df_world, joinCode = "ISO3",
                          nameJoinColumn = "tag")
```

```
## 212 codes from your data successfully matched countries in the map
## 5 codes from your data failed to match with a country code in the map
## 31 codes from the map weren't represented in your data
```

```
mapCountryData(map, nameColumnToPlot = "region", catMethod = "categorical",
               missingCountryCol = gray(.8),
               colourPalette = c("#35B779", "#FDE725", "#21908C", "#440154",
```
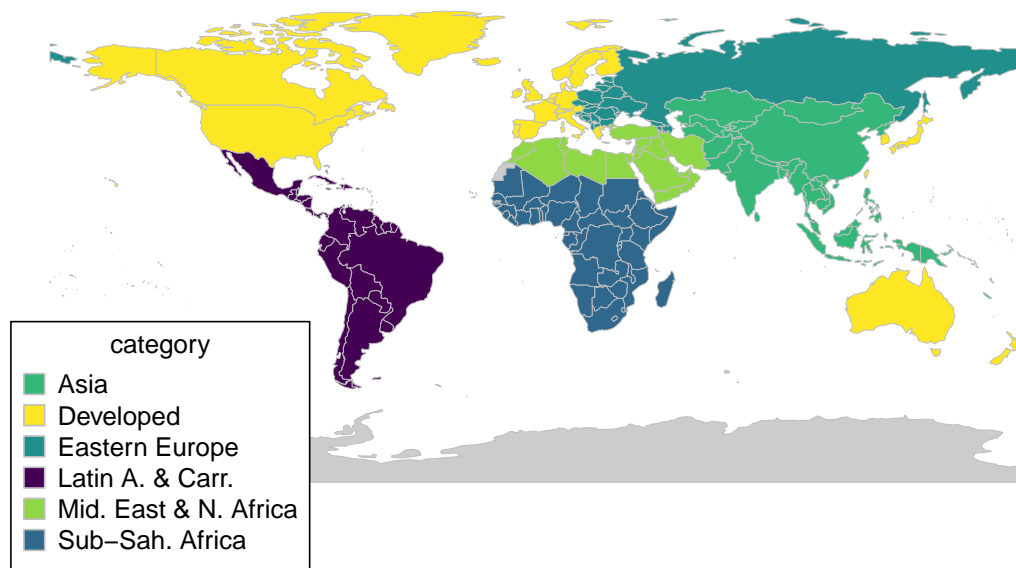
```
                        "#8FD744", "#30678D"),
              mapTitle = "World grouping in macroregions")
```

World grouping in macroregions


```
# 3. World electricity generation from LC sources, grouped by macroregion

# a. LC generation by macroregion
# Creation of the dataset
place = main[c("year","hydro_electricity", "wind_electricity",
                "nuclear_electricity", "solar_electricity",
                "other_renewable_electricity", "electricity_generation")] %>%
  mutate_all(~replace_na(.,0)) %>%
  cbind(., tag = main$tag) %>%
  group_by(year, tag) %>%
  summarize(Nuclear = sum(nuclear_electricity)/sum(electricity_generation),
            Hydro = sum(hydro_electricity)/sum(electricity_generation),
            Wind = sum(wind_electricity)/sum(electricity_generation),
            Solar = sum(solar_electricity)/sum(electricity_generation),
            Other = sum(other_renewable_electricity)/sum(electricity_generation)) %>%
  gather(key = "Type",
         value = "elect",
         -year, -tag)
place$Type = factor(place$Type, levels = c("Nuclear", "Hydro", "Solar",
                                            "Wind", "Other"))
# Here we remove data for Sub-Saharan countries from 2000, as data is not available
# for most of the countries
```

```r
for(i in 1:nrow(place)){
  if(place$year[i] < 2000 & place$tag[i] == "sub_african"){
    place$elect[i] = 0
  }
}

tag_modifier = function(place_data){
  place_data[place_data$tag  == "asian", "tag"] = "Asian"
  place_data[place_data$tag  == "developed", "tag"] = "Developed"
  place_data[place_data$tag  == "east_europe", "tag"] = "Eastern Europe"
  place_data[place_data$tag  == "latin", "tag"] = "Latin America & Caribbeans"
  place_data[place_data$tag  == "middle_east", "tag"] = "Middle East & Northern Africa"
  place_data[place_data$tag  == "sub_african", "tag"] = "Sub-Saharan Africa"
  return(place_data)
}

source_modifier = function(place_data){
  place_data[place_data$Source  == "nuclear_share_elec", "Source"] = "Nuclear"
  place_data[place_data$Source  == "hydro_share_elec", "Source"] = "Hydro"
  place_data[place_data$Source  == "solar_share_elec", "Source"] = "Solar"
  place_data[place_data$Source  == "wind_share_elec", "Source"] = "Wind"
  place_data[place_data$Source  == "other_renewables_share_elec", "Source"] = "Other"

  place_data$Source = factor(place_data$Source, levels = c("Nuclear", "Hydro", "Solar", "Wind",
                                          "Other"))
  return(place_data)
}

place = tag_modifier(place)


# Creation of the plot
gg1 = ggplot(place, aes(year, elect, fill = Type)) +
  geom_area(alpha=0.6 , size=.5, colour="white")+
  scale_x_continuous(limits = c(1985,2020)) +
  scale_fill_manual(values = c("#B2FF00", "#05B6FF", "#0060FA", "#141BDB", "#00296B")) +
  facet_wrap(~ tag, nrow = 2) +
  labs(title = "Share of electricity generation from LC sources",
       subtitle = "Grouped by source and macroregion",
       x = "Year",
       y = "% of terawatt-hours")

# b. Plot of renewables generation by macroregion, excluding hydropower
# Creation of the dataset
place = filter(place, Type == "Solar" | Type == "Wind" | Type == "Other")

# Creation of the plot
gg2 = ggplot(place, aes(year, elect, fill = Type)) +
  geom_area(alpha=0.6 , size=.5, colour="white")+
  scale_x_continuous(limits = c(1985,2020)) +
  scale_fill_manual(values = c("#0060FA", "#141BDB", "#00296B")) +
  facet_wrap(~ tag, nrow = 2) +
  labs(x = "Year",
       y = "% of terawatt-hours")

# Visualization of gg1, gg2 and gg3
grid.arrange(gg1, gg2, ncol=1)
```
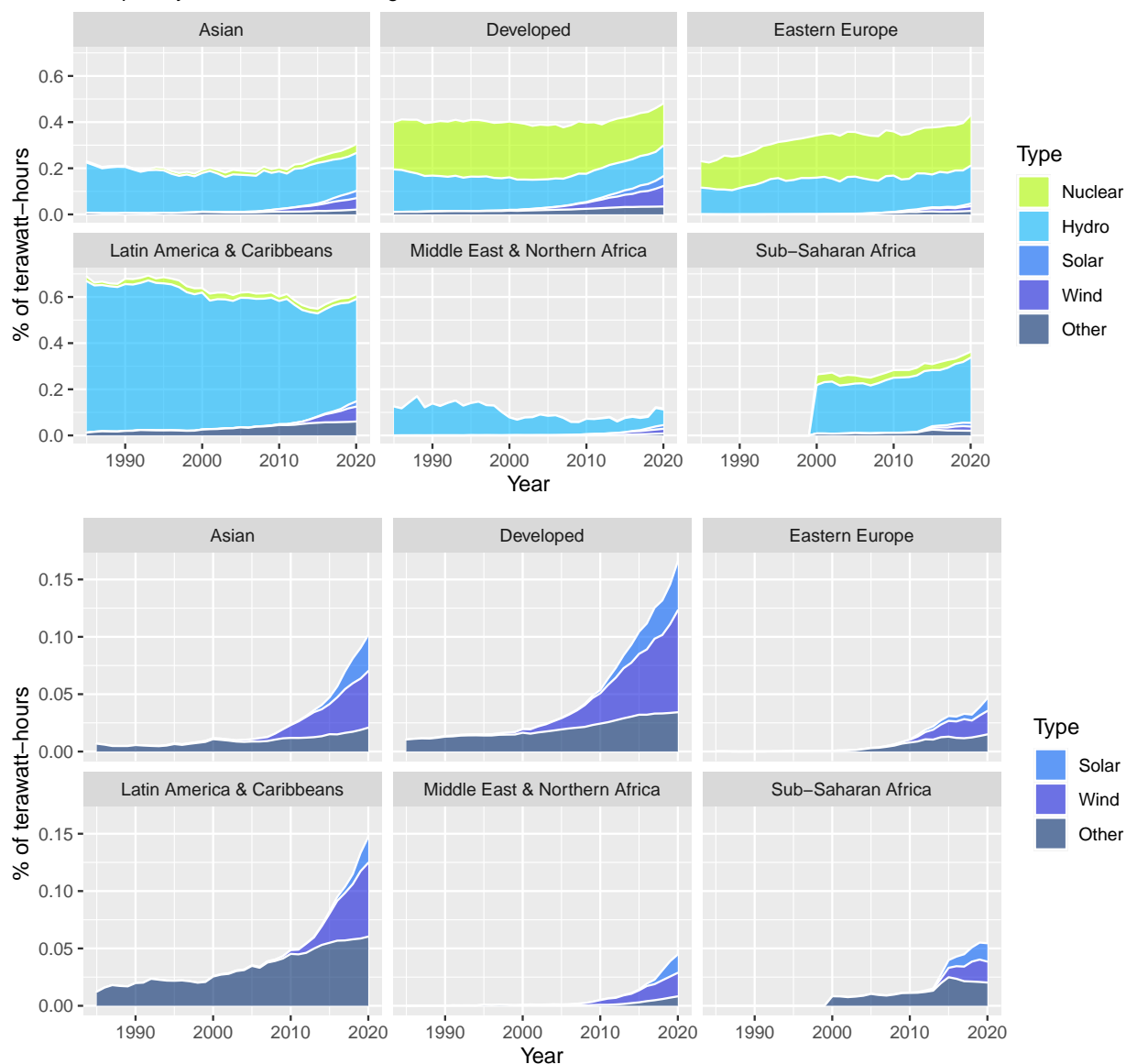
Share of electricity generation from LC sources
Grouped by source and macroregion

A variety of conclusions can be drawn from the previous graphs. Here we present the four main findings.

1. The different areas are **not homogeneous** in electricity generation from low-carbon sources: Latin American countries have a more significant share; developed, Eastern European and Sub-Saharan and Asian follow; Middle-East generation is negligible.

2. **Hydropower** is an essential source of electricity in all the considered macroregion.

3. **Nuclear electricity** is significant only in developed countries and Eastern Europe; in those macroregions, its importance is comparable to hydropower generation.

4. Developed countries drive **non-hydropower renewable** production. Nonetheless, it is also true that those sources are also rapidly becoming more relevant in Latin America and Asia.

It is important to highlight also that a lower total electricity generation heavily influences Asian, Latin American, and (especially) Sub-Saharan generation rates, as the graph below shows.

```
# Focus on electricity generation per capita
# Creation of the dataset
place = main[c("year","electricity_generation", "population")] %>%
  mutate_all(~replace_na(.,0)) %>%
  cbind(., tag = main$tag) %>%
  group_by(tag,year) %>%
  summarize(gen_per_capita = sum(electricity_generation)/sum(population))

place = tag_modifier(place)
place$tag = factor(place$tag, levels = c("Developed", "Eastern Europe",
                                         "Middle East & Northern Africa",
                                         "Asian", "Latin America & Caribbeans",
                                         "Sub-Saharan Africa"))
colnames(place) = c("Macroregion","year","gen_per_capita")


# Creation of the plot
ggplot(place, aes(year, gen_per_capita, color = Macroregion)) +
  geom_line() +
  geom_point() +
  scale_x_continuous(limits = c(1985,2020), breaks = seq(1985, 2020, by = 5)) +
  scale_color_manual(values = c("#EA5555", "#F39C3C", "#ECD03F", "#6EB35E", "#4996C8",
                                "#774ED8")) +
  labs(title = "Electricity generation per capita",
       subtitle = "Grouped by macroregion",
       x = "Year",
       y = "Per capita terawatt-hours")
```
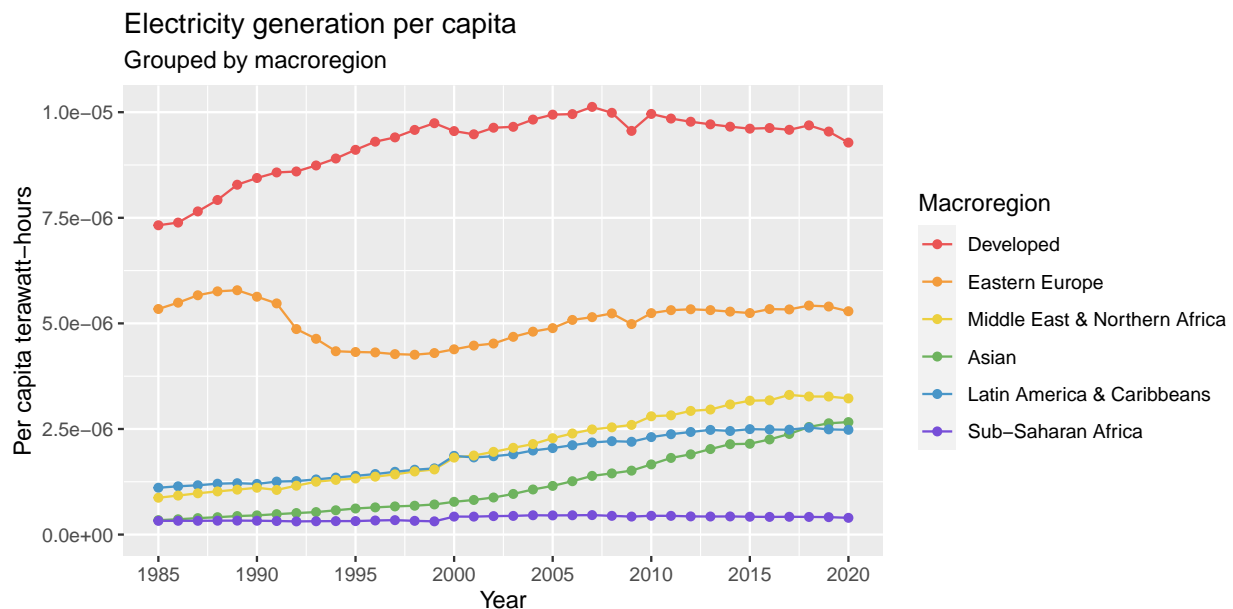

The electricity generation from non-hydro renewable sources in Sub-Saharan Africa increases rapidly between 2011 and 2015. Therefore, we studied the behavior of the five countries in the region with the highest non-hydro renewable electricity generation in 2015. As the plot shows, the steep increase is simply due to an exploding generation from solar and wind sources in South Africa.

```
suppressMessages({
# Extract the LC production grouped by source from 2011 to 2015 of the five countries
# with the highest renewables production in 2015
```

```r
#a. Extract the ISO of five countries with the highest non-hydro renewables generation
#   in 2015
place = select(main, iso_code, year, renewables_electricity, hydro_electricity) %>%
  filter(year == 2015 & tag == "sub_african") %>%
  mutate(non_hydro_elec = renewables_electricity - hydro_electricity) %>%
  arrange(desc(non_hydro_elec)) %>%
  select(iso_code) %>%
  top_n(5) %>%
  as.data.frame()

# b. Creation of the dataset
place = select(main, year, iso_code, renewables_electricity, hydro_electricity) %>%
  filter(year <= 2015 & year >= 2011 & iso_code %in% place$iso_code) %>%
  mutate(non_hydro_elec = renewables_electricity - hydro_electricity)
colnames(place) = c(colnames(place)[1], "ISO_code", colnames(place)[3:5])

# c. Creation the plot
ggplot(place, aes(year, non_hydro_elec, color = ISO_code)) +
  geom_line(size = 1) +
  geom_point(size = 1.5) +
  scale_x_continuous(limits = c(2011,2015), breaks = place$year) +
  scale_color_manual(values = c("#EA5555", "#F39C3C", "#ECD03F", "#6EB35E", "#4996C8")) +
    labs(title = "Share of electricity from non-hydro renewables",
      subtitle = "Selection of Sub-Saharan African countries",
      x = "Year",
      y = "% of terawatt-hours")
})
```


Share of electricity from non−hydro renewables
Selection of Sub−Saharan African countries
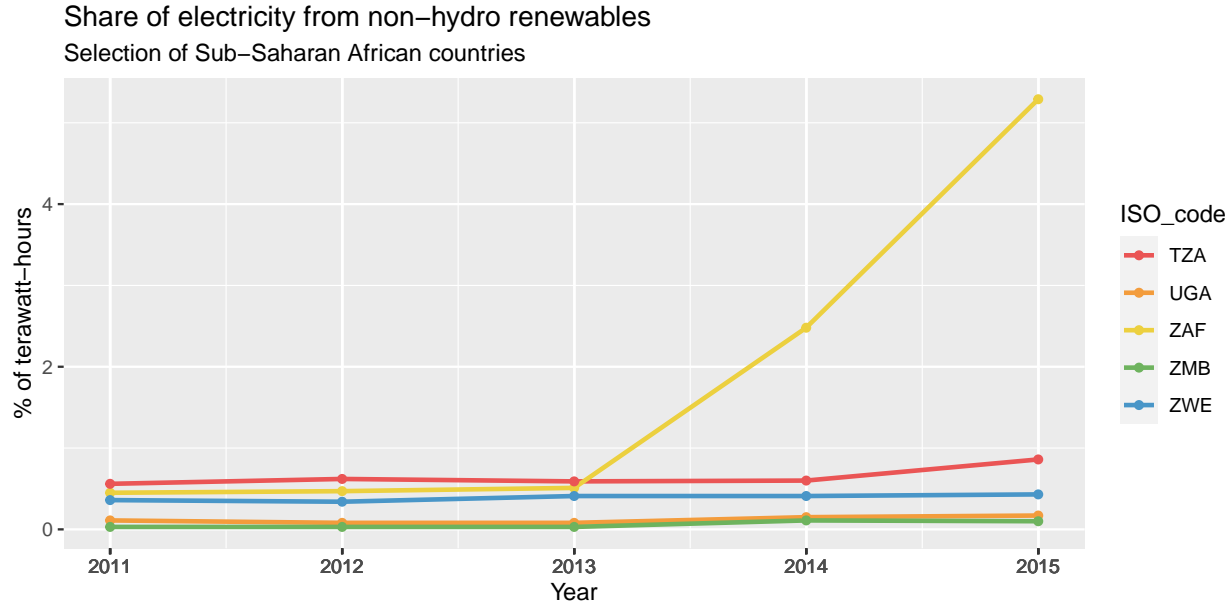
5.4 Green Score with focus on the sources

In this section we aim to study which countries are nearer to the full LC target (i.e., to produce from LC sources all the electricity they consume). To do so, we create a **Green Score**, defined as the following ratio:

$$\text{GS} = \frac{\text{electricity generated from LC sources}}{\text{electricity demand}}$$

Note that in paragraphs 5.4 and 5.5 analyses we removed the countries with a smaller population than 500,000. The reason is that we expect those countries to be too small to have a significant independent electricity policy with respect to their neighbors.

```
# Note. x is computed as the squared root of the electricity demand per capita, scaled in
# the interval [0,1]. We plot the electricity demand per capita to distinguish if the
# high Green Score is due to low electricity consumption or good green policies. We
# choose the described scaling because it allows for spacing more points in the graph.

# Scatterplot of the Green Score in 2020 in each country
# Creation of the dataset
place = filter(main, (year == 2020 & population > 500000 & iso_code != "REU")) %>%
  transform(elec_demand_per_capita = (sqrt(electricity_demand / population) - min(sqrt(electricity_demand / populat
         green_score_renew = renewables_electricity / electricity_demand,
         green_score_lc = low_carbon_electricity / electricity_demand) %>%
  select(iso_code, elec_demand_per_capita, green_score_renew, green_score_lc)

# Creation of the plot
ggplot()+
  geom_point(data = filter(place, iso_code != "LAO", iso_code != "BTN",
                           iso_code != "PRY", iso_code != "NOR"),
             mapping = aes(elec_demand_per_capita, green_score_lc,
                           color = green_score_lc)) +
  geom_text(data = filter(place, iso_code == "LAO" | iso_code == "BTN" |
                            iso_code == "PRY" | iso_code == "NOR"),
            mapping = aes(elec_demand_per_capita, green_score_lc,
                          label = iso_code, color = green_score_lc)) +
  geom_hline(yintercept = 1, linetype = "dashed", color = "red") +
  scale_color_gradient(low = "#ECD03F", high = "#59E80C", na.value = "#59E80C",
                       limits = c(0,1), guide = "none") +
  labs(title = "Green Score of world countries, 2020",
       x = "Scaled electricity demand per capita",
       y = "Green Score")
```


Green Score of world countries, 2020
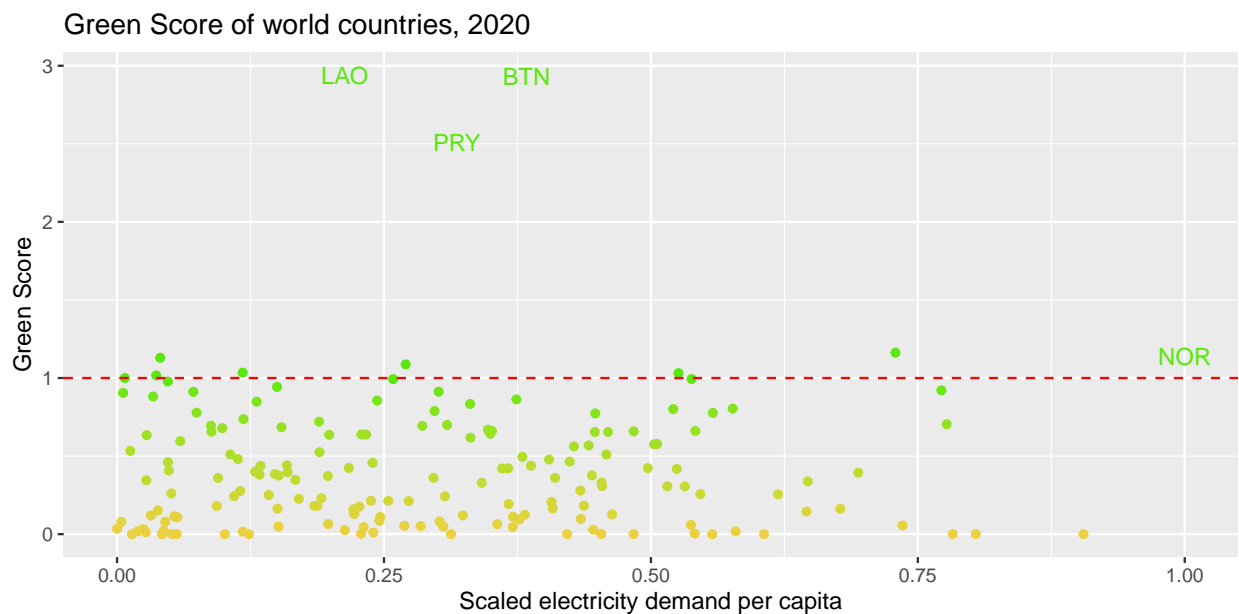
The plot highlights **four main outliers**: Laos, Bhutan, and Paraguay, with an overscaled Green Score and a low electricity demand per capita; Norway, with a positive Green Score while being the country with the highest electricity demand per capita. We want to understand if they have shared features that allow us to explain their

22

outperformance.

```r
# Barplot of the electricity generation mix in LAO, BTN, PRY and NOR
# Creation of the dataset
place = filter(main, (year == 2020 & (iso_code == "LAO" | iso_code == "BTN" |
                                      iso_code == "PRY" | iso_code == "NOR"))) %>%
  select(iso_code, solar_share_elec, wind_share_elec, hydro_share_elec,
         nuclear_share_elec, other_renewables_share_elec) %>%
  gather(key = "Source", value = "value", -iso_code)

# The following function modify the names of the sources to enchance the visualization
source_modifier = function(place_data){
  place_data[place_data$Source  == "nuclear_share_elec", "Source"] = "Nuclear"
  place_data[place_data$Source  == "hydro_share_elec", "Source"] = "Hydro"
  place_data[place_data$Source  == "solar_share_elec", "Source"] = "Solar"
  place_data[place_data$Source  == "wind_share_elec", "Source"] = "Wind"
  place_data[place_data$Source  == "other_renewables_share_elec", "Source"] = "Other"

  place_data$Source = factor(place_data$Source, levels = c("Nuclear", "Hydro", "Solar", "Wind",
                                                           "Other"))
  return(place_data)
}

place = source_modifier(place)


# Creation of the plot
ggplot(place, aes(x = iso_code, y = value, fill = Source)) +
  geom_bar(position = "stack", stat = "identity", alpha = 0.6, colour = "white") +
  scale_fill_manual(values = c("#B2FF00", "#05B6FF", "#0060FA", "#141BDB", "#00296B")) +
  labs(title = "Electricity generation mix in selected countries, 2020",
       x = "",
       y = "% gen. from LC sources")
```
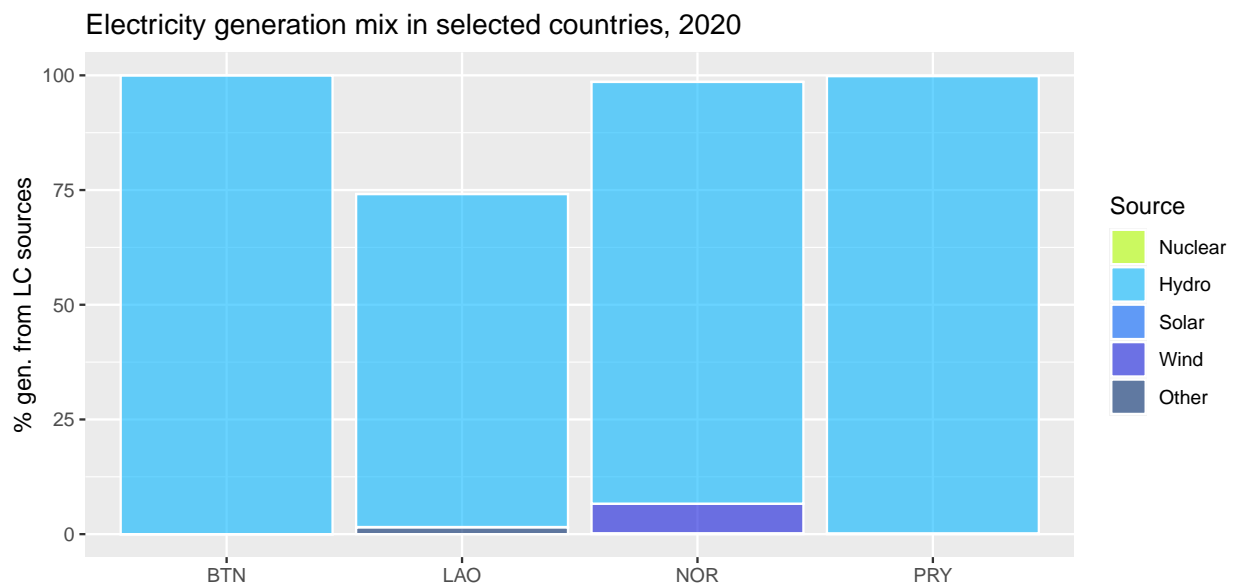

The high performance of the countries is due to a **dominant hydroelectric generation**. For example, further research on the topic highlights that Laos aims to become the "Battery of Southeast Asia" by further exploiting its impressive hydropower potential [13]. So those countries can achieve such an impressive result because of a resource

not available everywhere.

Nonetheless, some countries have great hydropower generation but do not exploit it: for instance, the Democratic Republic of Congo has a potential of 100.000 MW (more than four times the biggest hydropower plant in the world, the Three Gorges Dam [14]), but uses only 2.5% of it due to political instability and lack of investments [15].

We then analyze further the correlation between each LC source and the Green Score.

```
# Scatterplot of the green Score VS share of electricity produced from each LC source
# Creation of the dataset
place = filter(main, year == 2020) %>%
  transform(green_score_lc = (low_carbon_electricity / electricity_demand)) %>%
  select(iso_code, green_score_lc, nuclear_share_elec, hydro_share_elec,
         solar_share_elec, wind_share_elec, other_renewables_share_elec,
         green_score_lc) %>%
  filter(complete.cases(.)) %>%
  gather(key = "Source", value = "value", -iso_code, -green_score_lc) %>%
  # We excluded countries with an irrelevant production from each source
  filter(value > 1)

place = source_modifier(place)

# Creation of the plot
ggplot(filter(place, iso_code != "LAO", iso_code != "BTN", iso_code != "PRY"),
       aes(value, green_score_lc, color = value)) +
  geom_point() +
  geom_smooth(method = "loess", span = 2, se = FALSE, size = 1) +
  geom_hline(yintercept = 1, linetype = "dashed", color = "red") +
  scale_color_gradient(low = "grey30", high = "#59E80C", na.value = "#59E80C",
                       limits = c(0,40),
                       guide = "none") +
  facet_wrap(~Source, nrow = 2) +
  labs(title = "Green Score VS share of LC generation by source, 2020",
       x = "% gen. from LC source",
       y = "Green Score")
```

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


Green Score VS share of LC generation by source, 2020
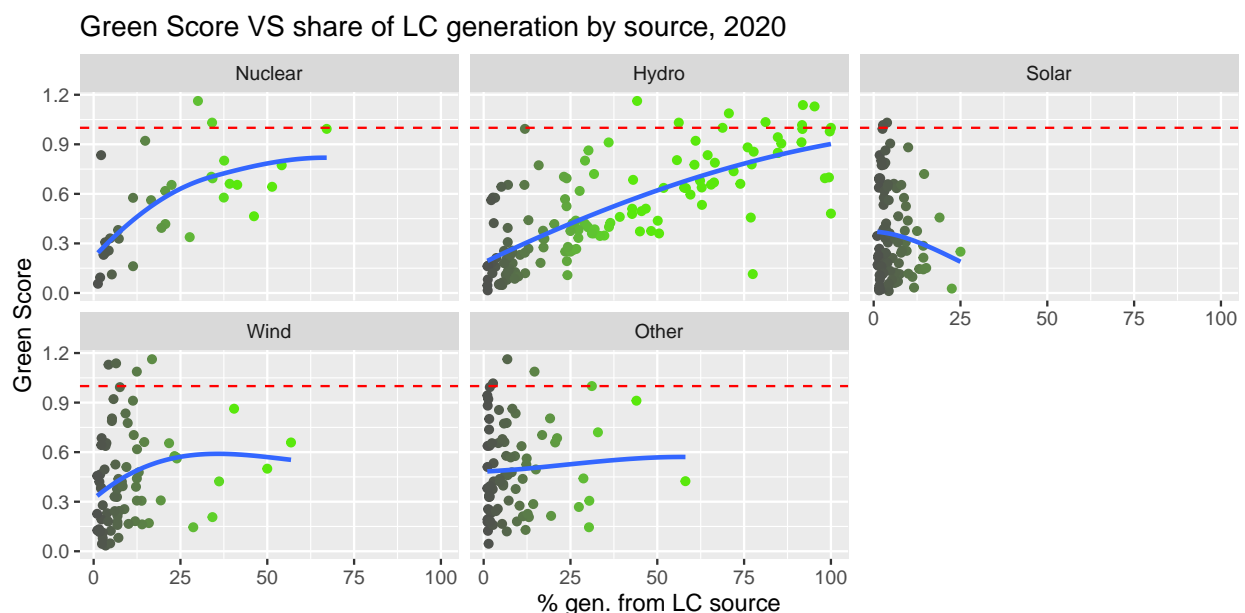
```
# Note. "loess" is a statistical technique used for estimating smooth curves in
# scatterplot data. It works by fitting multiple local regression models
# to different subsets of the data, allowing it to capture non-linear patterns and
# relationships between variables. It was introduced in the following plot only to
# highlight better the trends from a graphical point of view.
```

The plot confirms the correlation between the Green Score and the electricity generation from hydropower. It also shows an important link with nuclear power production but not with solar, wind, and other sources.

5.5 Green Score with focus on the macroregions

In this final section of descriptive analyses we focus on the analysis of the Green Score in each country.

```
# Boxplot of the Green Score
# Creation of the dataset
place = filter(main, year == 2020) %>%
  transform(green_score_lc = (low_carbon_electricity / electricity_demand)) %>%
  select(iso_code, tag, green_score_lc)

place[place$tag  == "asian", "tag"] = "Asian"
place[place$tag  == "developed", "tag"] = "Developed"
place[place$tag  == "east_europe", "tag"] = "East. Europe"
place[place$tag  == "latin", "tag"] = "L. Am. & C."
place[place$tag  == "middle_east", "tag"] = "M. East & N. Africa"
place[place$tag  == "sub_african", "tag"] = "S.S. Africa"

colnames(place) = c("iso_code", "Macroregion", "green_score_lc")

# Creation of the plot
ggplot(place, aes(x = Macroregion, y = green_score_lc)) +
  geom_boxplot(aes(fill = Macroregion)) +
  ylim(0,1.25) +
  scale_fill_manual(values = c("#EA5555", "#F39C3C", "#ECD03F", "#6EB35E", "#4996C8",
                               "#774ED8")) +
  geom_hline(yintercept = 1, linetype = "dashed", color = "red") +
    labs(title = "Boxplots of the Green Score, 2020",
         subtitle = "Grouped by macroregion",
       x = "",
       y = "Green Score")
```

Boxplots of the Green Score, 2020
Grouped by macroregion
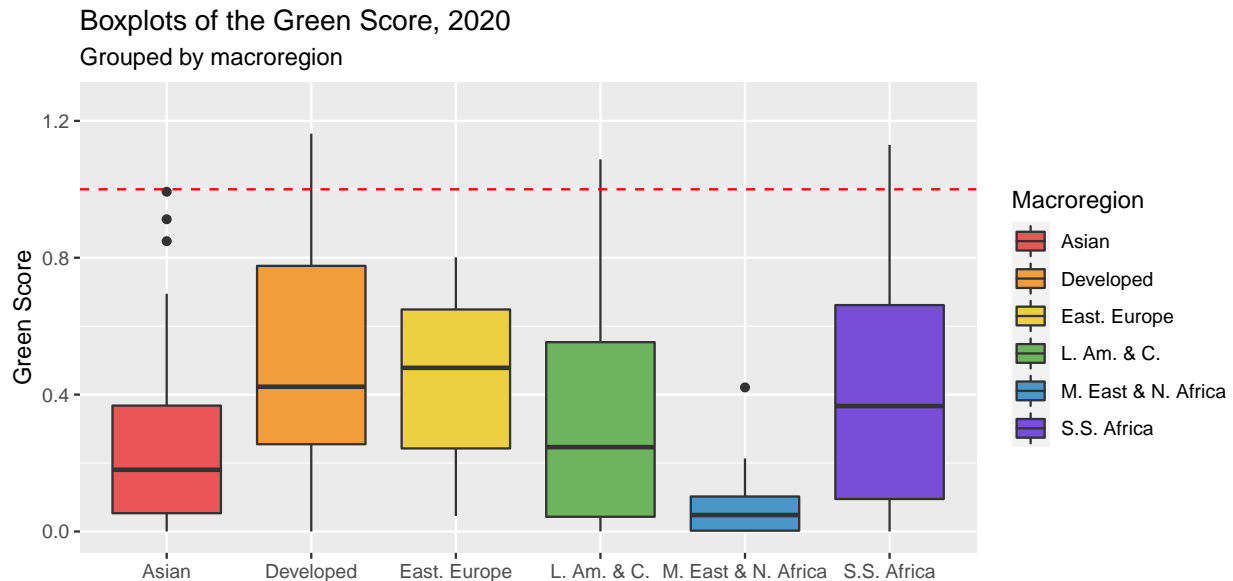
The leading macroregion by Green Score comprises the **developed countries**, followed by Eastern Europe, Sub-Saharian Africa, Asia, and Latin America & Caribbeans. The Middle East & Northern Africa has significantly lower score compared to other areas. Let us now explore each area separately.

```r
# Barplot of the Green Score
# Creation of the dataset
place = filter(main, year == 2020, !is.na(other_renewables_share_elec),
                population >= 500000) %>%
  transform(ratio = (low_carbon_electricity / (electricity_demand * low_carbon_share_elec))) %>%
  transform(solar_share_elec = ratio * solar_share_elec,
            wind_share_elec = ratio * wind_share_elec,
            hydro_share_elec = ratio * hydro_share_elec,
            nuclear_share_elec = ratio * nuclear_share_elec,
            other_renewables_share_elec = ratio * other_renewables_share_elec) %>%
  select(iso_code, tag, solar_share_elec, wind_share_elec,
         hydro_share_elec, nuclear_share_elec, other_renewables_share_elec) %>%
  # There are NaN values obtain because of division by zero. We want them to be 0
  mutate(across(where(is.numeric), ~ ifelse(is.nan(.), 0, .))) %>%
  # There are NA values. We want to remove them
  gather(key = "Source", value = "value", -iso_code, -tag)

place = source_modifier(place)

# Creation of the plot for the developed countries
ggplot(filter(place, tag == "developed"),
       aes(x = reorder(iso_code, value), y = value, fill = Source)) +
  geom_bar(position = "stack", stat = "identity", alpha = 0.6, colour = "white") +
  scale_fill_manual(values = c("#B2FF00", "#05B6FF", "#0060FA", "#141BDB", "#00296B")) +
  scale_y_continuous(breaks = seq(0, 1.2, by = 0.2)) +
  geom_hline(yintercept = 1, linetype = "dashed", color = "red") +
  labs(title = "Green Score in developed countries, 2020",
       subtitle = "Grouped by source",
       x = "",
       y = "Green Score") +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))
```

Green Score in developed countries, 2020
Grouped by source
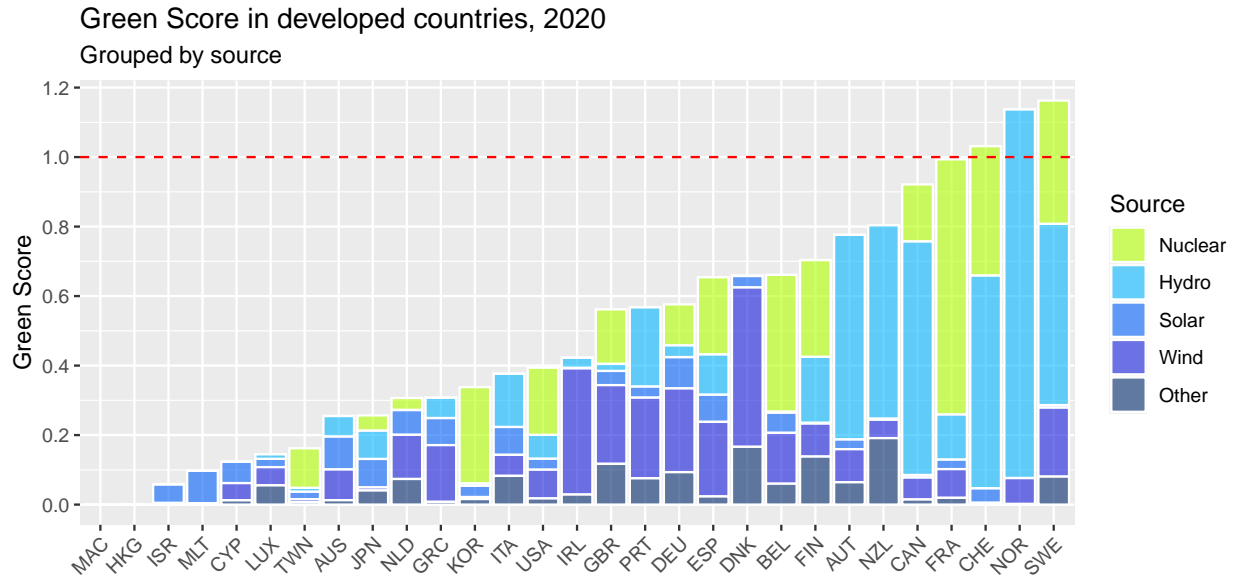
The developed countries with the highest Green Scores are **Sweden**, Norway, Switzerland, and France, while the ones with the lowest values are Macao, Hong Kong, Israel and Malta.

The LC sources are heterogeneous: while countries like Norway, Switzerland, and Canada are mainly driven by hydropower, others like France and Belgium mainly generate electricity from nuclear power, and others still, like Denmark and Ireland, are mainly driven by non-hydro renewables.

It is also interesting to highlight that the developed Asian countries have a lower score compared to the others: the second best performing country, Japan, outperforms only Australia, Luxembourg, Cyprus and Malta.

```
# Creation of the plot for Eastern Europe
ggplot(filter(place, tag == "east_europe"),
       aes(x = reorder(iso_code, value), y = value, fill = Source)) +
  geom_bar(position = "stack", stat = "identity", alpha = 0.6, colour = "white") +
  scale_fill_manual(values = c("#B2FF00", "#05B6FF", "#0060FA", "#141BDB", "#00296B")) +
    scale_y_continuous(breaks = seq(0, 1, by = 0.2)) +
  geom_hline(yintercept = 1, linetype = "dashed", color = "red") +
  labs(title = "Green Score in Eastern Europe, 2020",
       subtitle = "Grouped by source",
       x = "",
       y = "Green Score") +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))
```

Green Score in Eastern Europe, 2020
Grouped by source
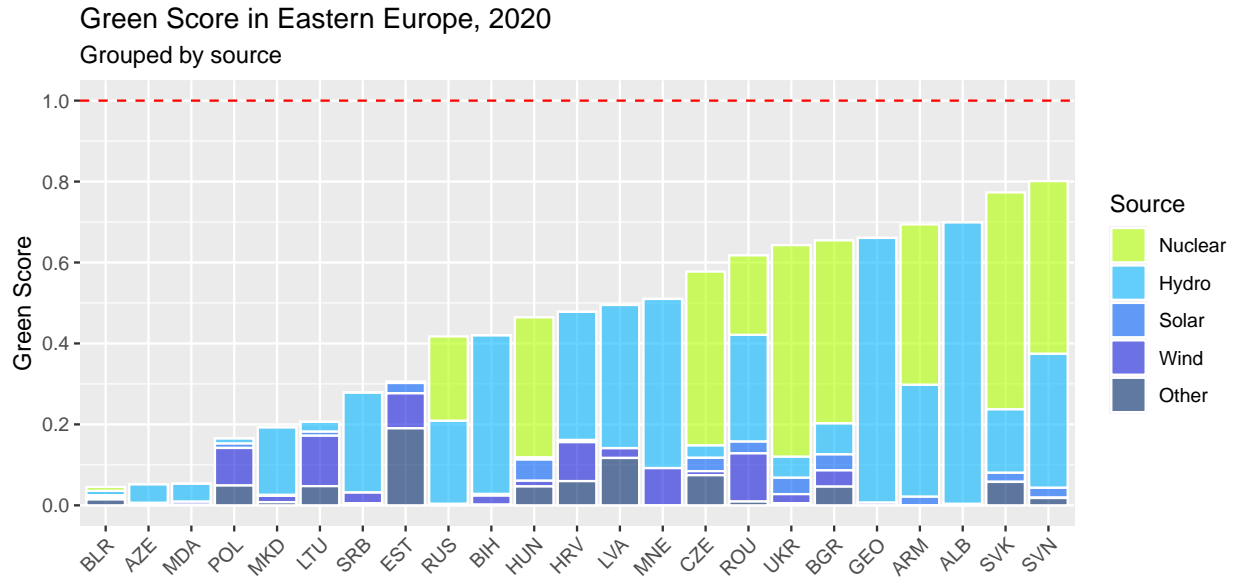
No country in Eastern Europe is near a Green Score equal to 1: the best-performing one is **Slovenia**, with a score of 0.72. Follow Slovakia, Albania and Armenia. In the region there are also three countries with a score near zero: Belarus, Azerbaijan and Moldova. As the barplot shows, Eastern European countries mainly produce LC electricity through hydro and nuclear sources.

```r
# Creation of the plot for Latin America & Caribbean
gg1 = ggplot(filter(place, tag == "latin", iso_code != "PRY"),
        aes(x = reorder(iso_code, value), y = value, fill = Source)) +
  geom_bar(position = "stack", stat = "identity", alpha = 0.6, colour = "white") +
  scale_fill_manual(values = c("#B2FF00", "#05B6FF", "#0060FA", "#141BDB", "#00296B")) +
    scale_y_continuous(breaks = seq(0, 1.2, by = 0.2)) +
  geom_hline(yintercept = 1, linetype = "dashed", color = "red") +
  labs(title = "Green Score in L. America & Caribbean, Asia, and Sub-Saharan countries, 2020",
      subtitle = "Grouped by source. Note: Paraguay, Laos and Bhutan removed being outliers",
      x = "",
      y = "Green Score") +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))

# Creation of the plot for the Asian countries
gg2 = ggplot(filter(place, tag == "asian", iso_code != "LAO", iso_code != "BTN"),
        aes(x = reorder(iso_code, value), y = value, fill = Source)) +
  geom_bar(position = "stack", stat = "identity", alpha = 0.6, colour = "white") +
  scale_fill_manual(values = c("#B2FF00", "#05B6FF", "#0060FA", "#141BDB", "#00296B")) +
    scale_y_continuous(breaks = seq(0, 1, by = 0.2)) +
  geom_hline(yintercept = 1, linetype = "dashed", color = "red") +
  labs(x = "",
      y = "Green Score") +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))

# creation of the plot for the Sub-Saharan Africa
gg3 = ggplot(filter(place, tag == "sub_african"),
        aes(x = reorder(iso_code, value), y = value, fill = Source)) +
  geom_bar(position = "stack", stat = "identity", alpha = 0.6, colour = "white") +
  scale_fill_manual(values = c("#B2FF00", "#05B6FF", "#0060FA", "#141BDB", "#00296B")) +
    scale_y_continuous(breaks = seq(0, 1.2, by = 0.2)) +
  geom_hline(yintercept = 1, linetype = "dashed", color = "red") +
  labs(x = "",
```

```r
        y = "Green Score") +
  theme(axis.text.x = element_text(angle = 90, hjust = 1))

# Visualization of the plots
grid.arrange(gg1, gg2, gg3, ncol=1)
```

Green Score in L. America & Caribbean, Asia, and Sub–Saharan countries, 2020

Grouped by source. Note: Paraguay, Laos and Bhutan removed being outliers
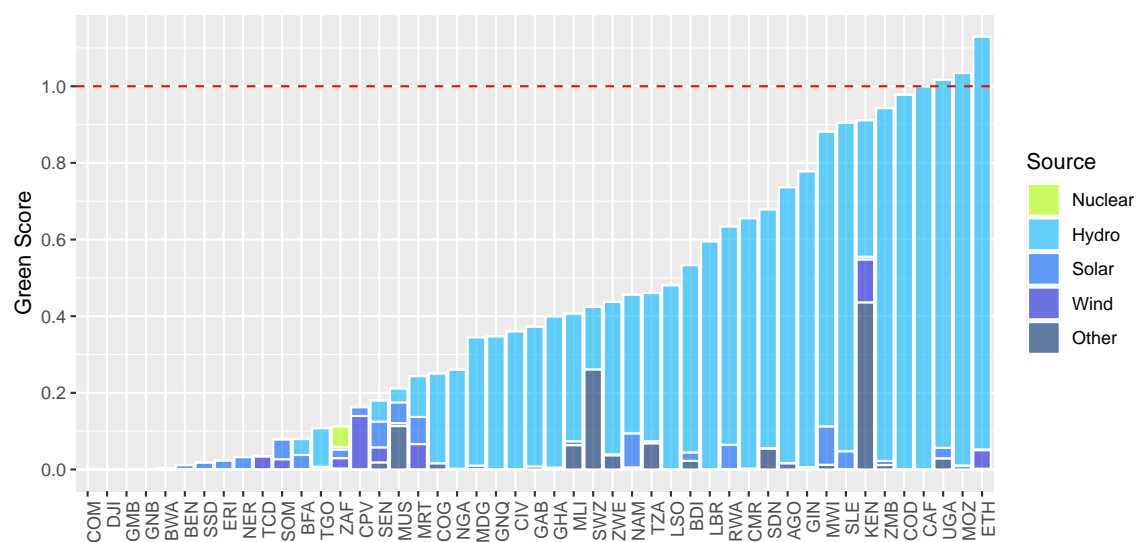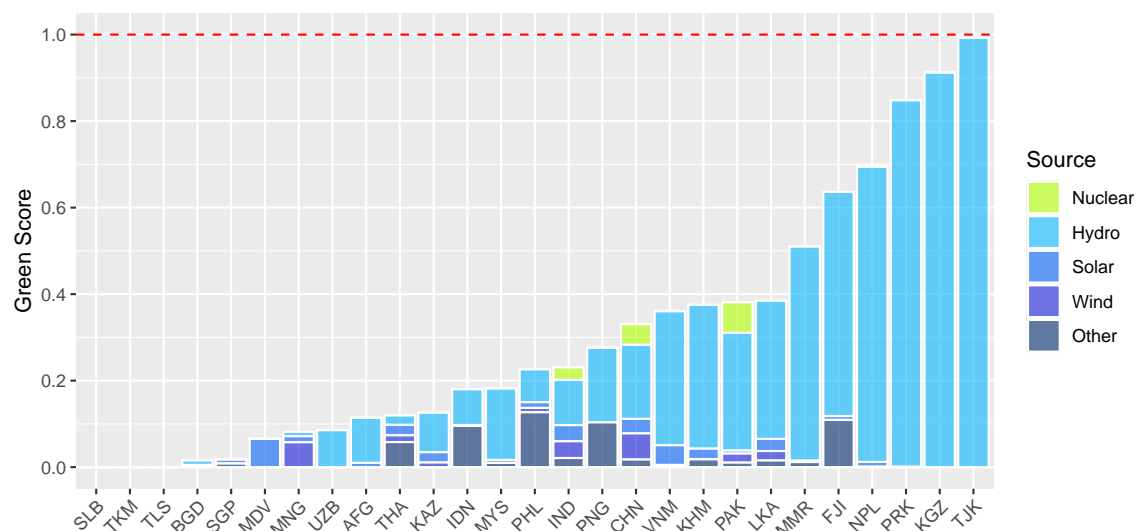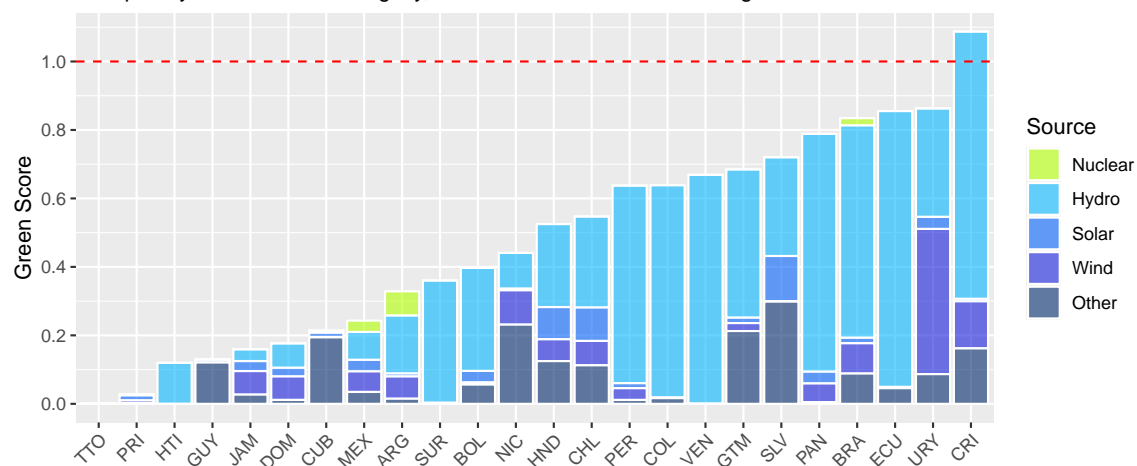
We

grouped the findings for Latin America & Caribbean, Asia, and Sub-Saharan countries because their electricity mixes are similar and mainly driven by hydropower.

The **best-performing** countries in each macroregion are Costa Rica, Uruguay, and Ecuador for Latin America & Caribbean; Tajikistan, Kyrgyzstan, and North Korea for Asia; Ethiopia, Mozambique, and Uganda for Sub-Saharan Africa. Instead, the countries with the lowest Green Score are respectively: Haiti, Porto Rico, and Trinidad & Tobago; Timor Est, Turkmenistan, and Solomon Islands; Botswana, Guinea-Bissau, Gambia, Djibouti, and Comoros (tied with totally fossil-dependent electricity generation).

Here are the other main observations:

1. some Latin American and Caribbean countries also have significant production from other sources, notably **Uruguay**, which mainly produces electricity through wind power;

2. **Kenya** is the only Sub-Saharan country with a good Green Score that generates significantly from non-hydropower sources, specifically geothermal [16]

```r
# Creation of the plot for the North African & Middle-East countries
ggplot(filter(place, tag == "middle_east"),
       aes(x = reorder(iso_code, value), y = value, fill = Source)) +
  geom_bar(position = "stack", stat = "identity", alpha = 0.6, colour = "white") +
  scale_fill_manual(values = c("#B2FF00", "#05B6FF", "#0060FA", "#141BDB", "#00296B")) +
    scale_y_continuous(breaks = seq(0, 1, by = 0.2)) +
  geom_hline(yintercept = 1, linetype = "dashed", color = "red") +
  labs(title = "Green Score in Middle East & North Africa 2020",
       subtitle = "Grouped by source",
       x = "",
       y = "Green Score") +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))
```
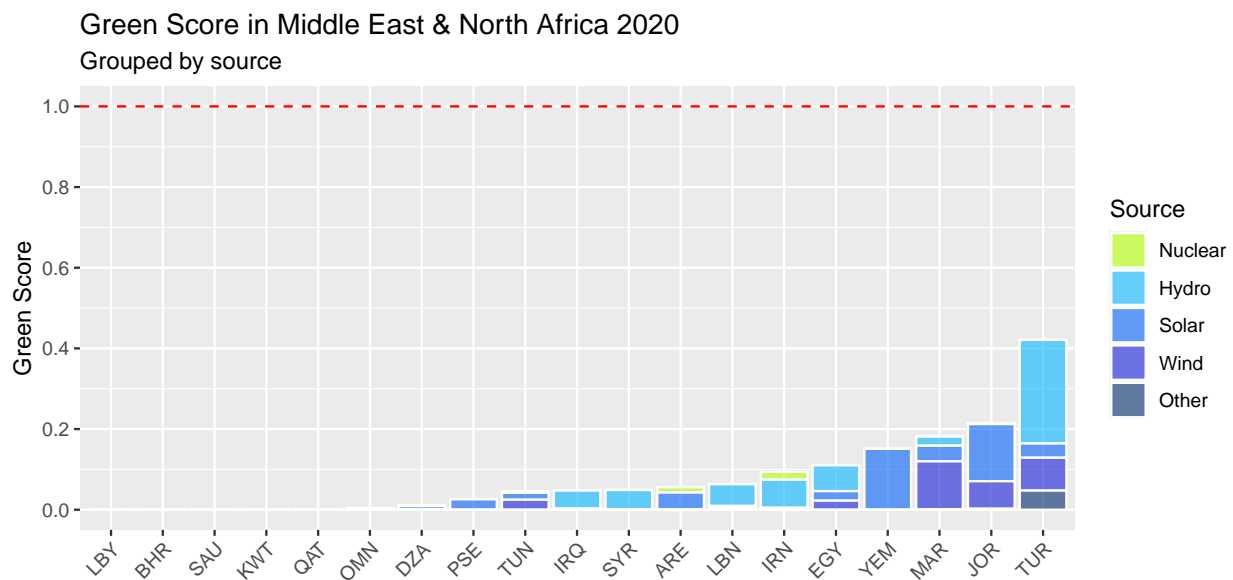

We conclude the descriptive analyses by looking at the performance of Northern Africa and Middle East countries. Consistently with the findings at paragraph 5.3, the barplot shows that all the countries fail to reach a good performance: the best performing one, **Turkey**, has a Green Score smaller than 0.5. As the following table shows, this is mainly due to an large availability of oil and gas.

```r
# Creation of a dataset containing per capita average fossil reserves in Middle-East
# and Northern African countries VS the rest of the world
place = filter(main, year == 2020, tag == "middle_east") %>%
```

```
  select(population, oil_reserves_2020, gas_reserves) %>%
  mutate_all(~replace_na(.,0)) %>%
  summarize(oil_reserves_2020 = round(mean(oil_reserves_2020 / population),2),
            gas_reserves = round(mean(gas_reserves / population),2))

place2 = filter(main, year == 2020, tag != "middle_east") %>%
  select(population, oil_reserves_2020, gas_reserves) %>%
  mutate_all(~replace_na(.,0)) %>%
  summarize(oil_reserves_2020 = round(mean(oil_reserves_2020 / population),2),
            gas_reserves = round(mean(gas_reserves / population),2))

place = rbind(place, place2)
place = cbind(c("Middle East & North Africa", "Other countries"), place)
colnames(place) = c("Macroregion", "Oil (2020)", "Gas (2020)")
kable(place, caption = "Per capita reserves of fossil electricity sources")
```

Table 2: Per capita reserves of fossil electricity sources

Macroregion	Oil (2020)	Gas (2020)
Middle East & North Africa	469.81	583180.3
Other countries	18.40	22920.9

Chapter 6

Modeling

Bibliografia e link

[1] https://github.com/owid/energy-data

[2] https://data.worldbank.org/indicator/NY.GDP.MKTP.KD

[3] https://ourworldindata.org/grapher/land-area-km

[4] https://ourworldindata.org/grapher/share-of-land-area-used-for-agriculture

[5] https://ourworldindata.org/grapher/share-of-population-urban

[6] https://ourworldindata.org/grapher/human-development-index

[7] https://ourworldindata.org/grapher/death-rates-from-air-pollution

[8] https://ourworldindata.org/grapher/coal-proved-reserves

[9] https://ourworldindata.org/grapher/oil-proved-reserves

[10] https://www.oecd.org/publications/uranium-20725310.htm

[11] https://ourworldindata.org/grapher/natural-gas-proved-reserves

[12] https://en.wikipedia.org/wiki/Low-carbon_power

[13] https://www.andritz.com/hydro-en/hydronews/hydro-news-asia/laos

[14] https://en.wikipedia.org/wiki/List_of_largest_hydroelectric_power_stations

[15] https://www.andritz.com/hydro-en/hydronews/hydropower-africa/democratic-rep-congo

[16] https://lowcarbonpower.org/region/Kenya