Report

This report focuses on the dataset 'The share of the population with access to electricity and clean fuels for cooking.' The dataset provides a comprehensive overview of global progress in access to modern energy services, capturing both the availability of electricity and clean cooking fuels across countries and over time. These indicators are essential for understanding development disparities and monitoring progress toward Sustainable Development Goal 7 (SDG7) - ensuring access to affordable, reliable, sustainable, and modern energy for all.

The dataset is organised by country and year, including global totals. It records the number of people with and without access to electricity and clean cooking fuels, allowing for analysis of regional differences and temporal trends in energy access and infrastructure development.

Table of variables

Variable name	Explanation
Entity	Country or region being observed, including the world as an aggregate.
Year	Calendar year of the observation, used to track changes over time.
Number.of.people.with.access.to.electricity	Total number of people with access to electricity, indicating energy availability and infrastructure development.
Number.of.people.without.access.to.electricity	Total number of people lacking access to electricity, reflecting levels of energy poverty.
Number_with_clean_fuels_cooking	Total number of people using clean cooking fuels, showing access to safer and more sustainable energy sources.
Number_without_clean_fuels_cooking	Total number of people relying on traditional or polluting cooking fuels, indicating limited access to clean household energy.

This dataset was chosen because access to electricity and clean cooking fuels is a fundamental indicator of social and economic development. From a social science perspective, energy access shapes multiple dimensions of human well-being - including health, education, and economic opportunity. Examining the share of the population with access to modern energy services allows for a deeper understanding of global inequality and the uneven pace of development across countries.

Data overview

Variable	Mean	Median	Min	Max	Standard Deviation	Missing Values
Number of people with access to electricity	199,300,000	5,392,000	0	6,912,000,000	679,348,400	0
Number of people without	50,200,000	64,780	0	1,628,000,000	173,187,100	0

Variable	Mean	Median	Min	Max	Standard Deviation	Missing Values
access to electricity						
Number of people with clean cooking fuels	158,500,000	5,124,000	2,032	4,406,000,000	475,491,600	2,661
Number of people without clean cooking fuels	153,700,000	1,856,000	0	3,094,000,000	475,064,600	2,661

The dataset spans from 1990 to 2019, covering 6,582 observations across various countries and including global aggregates. There are no missing values for electricity-related variables, while approximately 2,661 observations are missing for the clean cooking variables, possibly due to limited data availability for earlier years or smaller economies.

On average, around 199 million people per country-year have access to electricity, while 50 million lack access, showing significant progress toward electrification but also substantial variation (as reflected by the large standard deviations). Access to clean cooking fuels is notably lower on average (158 million with access versus 153 million without), revealing challenges in achieving universal clean energy for household use. The high variability indicates strong disparities between developed and developing countries, particularly in cooking fuel access.

Data Wrangling

The data wrangling process was conducted entirely using the Tidyverse packages. The objective was to clean, restructure, and compute meaningful indicators for analysis. Three main operations were performed:

```
data_selected ← data %>%
  select(entity = Entity,
         year = Year,
         with_elec = 'Number of people with access to electricity',
         without_elec = 'Number of people without access to electricity',
         with_cooking = number_with_clean_fuels_cooking,
         without_cooking = number_without_clean_fuels_cooking)
```

Variable selection and renaming to simplify column names and focus on relevant indicators.

```
data_clean ← data_selected %>%
  mutate(across(c(with_cooking, without_cooking),
                ~if_else(is.na(.), 0, .)))
```

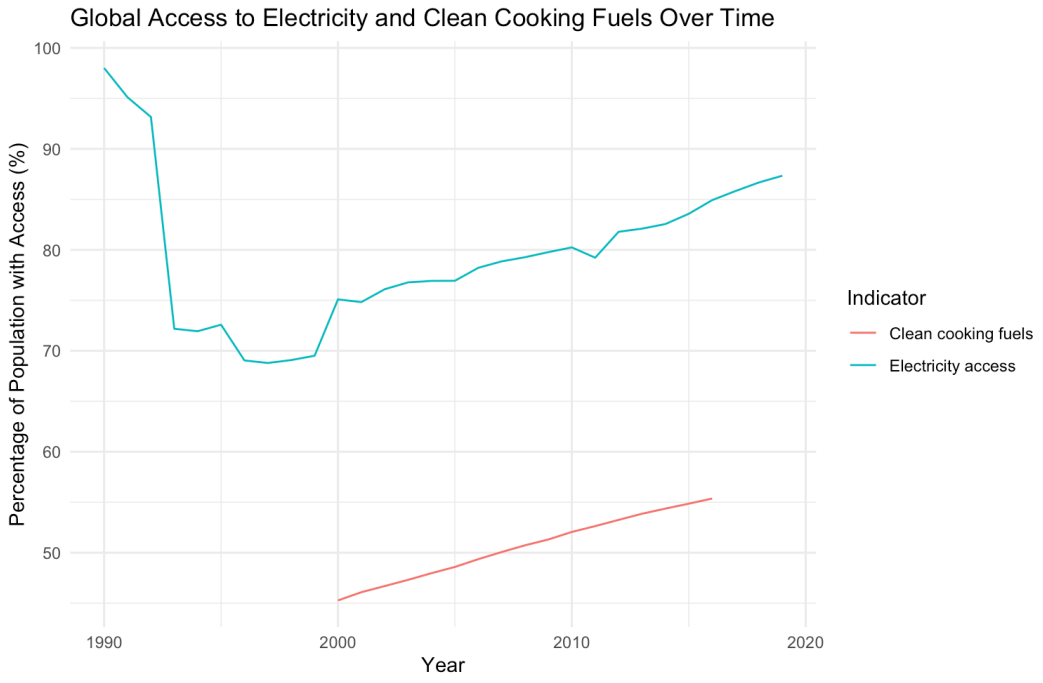
Missing data handling and therefore, replacing NA values with 0 to ensure consistent numerical computation.

```
data_summary ← data_clean %>%
  group_by(year) %>%
  summarise(
```

```
  total_with_elec = sum(with_elec, na.rm = TRUE),
  total_without_elec = sum(without_elec, na.rm = TRUE),
  total_with_cooking = sum(with_cooking, na.rm = TRUE),
  total_without_cooking = sum(without_cooking, na.rm = TRUE)
) %>%
mutate(
  pct_with_elec = 100 * total_with_elec / (total_with_elec + total_without_elec),
  pct_without_elec = 100 * total_without_elec / (total_with_elec + total_without_elec),
  pct_with_cooking = 100 * total_with_cooking / (total_with_cooking + total_without_cooking),
  pct_without_cooking = 100 * total_without_cooking / (total_with_cooking + total_without_cooking)
) %>%
```

Creating new variables - calculating yearly global percentages of the population with and without access to electricity and clean cooking fuels. This function was necessary because the dataset provides absolute population counts, but comparing countries and tracking progress over time is more meaningful when expressed as percentages. Since national populations vary widely, using percentages standardises comparisons and highlights proportional improvements in access rather than absolute numbers. The function takes the cleaned dataset as input, groups the data by year, and returns a summary table with total population counts and calculated global percentages for each indicator.

Data Visualisation

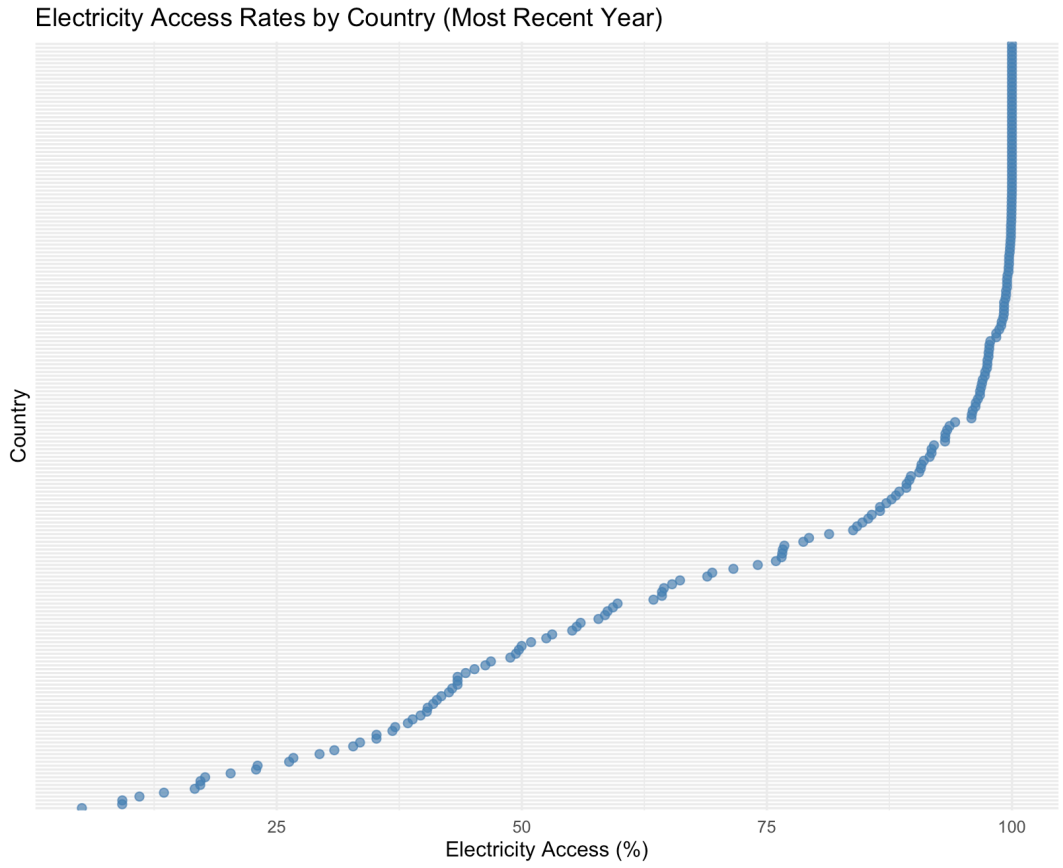



```
ggplot(data_summary, aes(x = year)) +
  geom_line(aes(y = pct_with_elec, colour = "Electricity access")) +
  geom_line(aes(y = pct_with_cooking, colour = "Clean cooking fuels")) +
  labs(
    title = "Global Access to Electricity and Clean Cooking Fuels Over Time",
```

```
  x = "Year",
  y = "Percentage of Population with Access (%)",
  colour = "Indicator"
) +
theme_minimal()
```

This line chart shows the global percentage of the population with access to electricity and clean cooking fuels between 1990 and 2019. Access to electricity has increased steadily since the early 2000s, reflecting substantial global progress in infrastructure development and rural electrification programs. However, the line for clean cooking fuels remains much lower and grows at a slower rate, indicating persistent inequality in access to clean household energy.

This visualisation was selected because it provides a temporal overview of progress toward Sustainable Development Goal 7 (SDG7). By focusing on global percentages, the chart captures the overall direction and pace of change, allowing readers to see whether access gaps are narrowing or persisting over time. It helps contextualise energy access as a long-term development process.

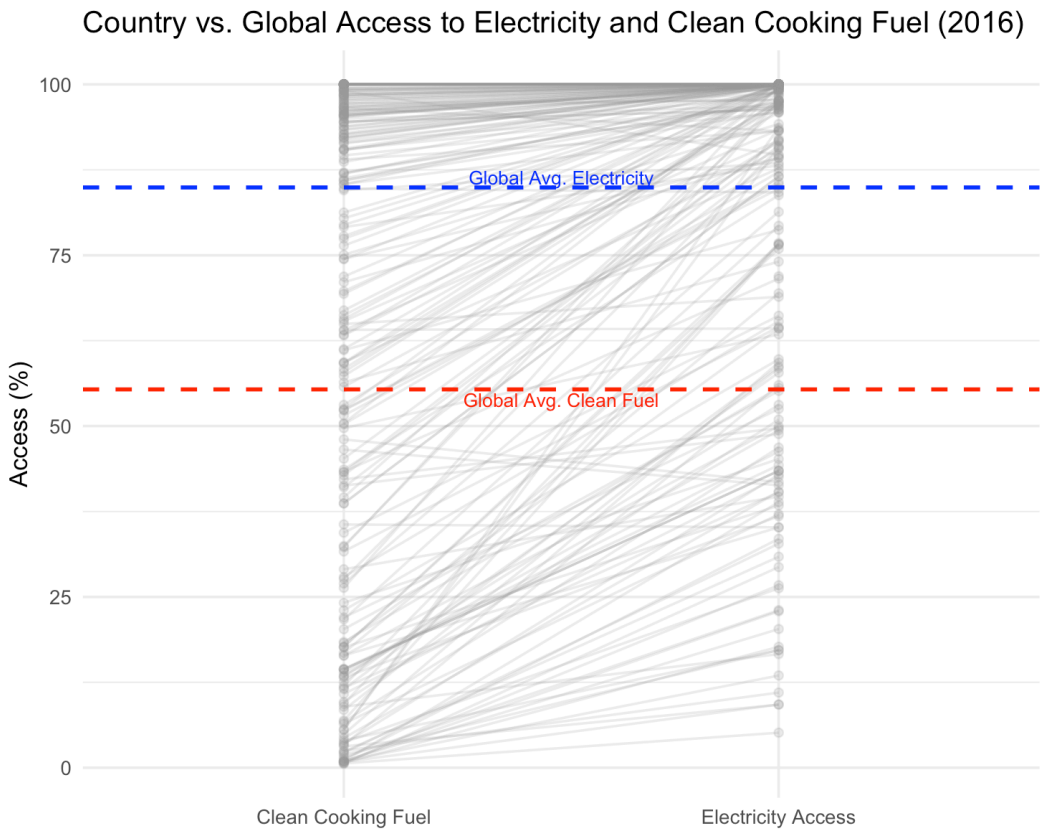



```
ggplot(country_percentages, aes(x = reorder(entity, elec_with_pct_country),
                                y = elec_with_pct_country)) +
  geom_point(colour = "steelblue", alpha = 0.7, size = 2) +
  coord_flip() +
  labs(
```

```
  title = "Electricity Access Rates by Country (Most Recent Year)",
  x = "Country",
  y = "Electricity Access (%)"
) +
theme_minimal(base_size = 12) +
theme(axis.text.y = element_blank(),  # hide country names
      axis.ticks.y = element_blank())
```

The scatter plot presents the share of the population with access to electricity across countries in the most recent available year. The x-axis shows the percentage of access, while each point represents a different country. The y-axis labels are hidden for readability due to the large number of countries, while coord_flip() improves visual clarity. The right-skewed distribution indicates that most countries now achieve near-universal electricity access (close to 100%), but a small number of countries still fall significantly below global averages, reflecting regional disparities - particularly across Sub-Saharan Africa and parts of South Asia.

This visualisation was chosen to reveal cross-country disparities that are hidden in global averages. It shows how progress is unevenly distributed, with some countries still struggling with basic energy access. This plot thus complements the first graph by shifting the focus from global trends to national inequalities.




```
ggplot(plot_data, aes(x = indicator, y = percentage, group = entity)) +
  geom_line(alpha = 0.2, colour = "grey60") +  # lines for each country
  geom_point(alpha = 0.2, colour = "grey60") +
```

```
  geom_hline(yintercept = global_2016$global_elec, colour = "blue",
             linewidth = 1, linetype = "dashed") +
  geom_hline(yintercept = global_2016$global_cooking, colour = "red",
             linewidth = 1, linetype = "dashed") +
  annotate("text", x = 1.5, y = global_2016$global_elec + 1.5,
           label = "Global Avg. Electricity", colour = "blue", size = 3.5) +
  annotate("text", x = 1.5, y = global_2016$global_cooking - 1.5,
           label = "Global Avg. Clean Fuel", colour = "red", size = 3.5) +
  labs(
    title = "Country vs. Global Access to Electricity and Clean Cooking Fuel (2016)",
    x = "",
    y = "Access (%)"
  ) +
  theme_minimal(base_size = 13)
```

This parallel coordinate plot compares each country's 2016 access rates for electricity and clean cooking fuels against the global averages (indicated by dashed lines). Most lines slope upward, showing that electricity access generally exceeds clean fuel access within the same country. The gap between the two indicators illustrates unequal progress across energy dimensions - many countries provide near-universal electricity while lagging behind in household clean energy adoption.

This visualisation was selected because it provides comparative insight - it allows the reader to assess how each country performs relative to the world average for both indicators simultaneously. This approach highlights multidimensional inequality and demonstrates that improvements in electricity access do not necessarily translate to similar progress in clean cooking energy.

This function was written to combine multiple ggplot layers - lines, points, reference lines, and annotations - into one coherent visual that communicates comparative context effectively. It's especially useful for identifying whether countries perform above or below global averages.

This report has examined global and country-level disparities in access to electricity and clean cooking fuels using data spanning from 1990 to 2019. Through systematic data wrangling, missing-value handling, and percentage conversion, the dataset was transformed into a comparable and interpretable format. Expressing access in relative rather than absolute terms was crucial to account for large differences in national population sizes and to reveal proportional inequalities across time and space.

Three visualisations were constructed to uncover different dimensions of the data. The global trend line chart highlighted substantial progress in electricity access since 2000, alongside the slower improvement in access to clean cooking fuels. The cross-country scatter plot exposed stark national disparities, where most countries approach universal electricity access while a minority still lag behind. Finally, the parallel coordinate plot offered a multidimensional comparison, illustrating that gains in electricity access often outpace those in clean cooking fuel adoption, underlining the uneven nature of energy development.

Together, these visualisations provide a holistic understanding of progress toward Sustainable Development Goal 7 (SDG7). They show that while electrification has improved globally, clean household energy remains an ongoing challenge, particularly in low-income regions.

Copilot Reflection

Throughout this project, I utilised Copilot as a collaborative tool to assist in the data wrangling and visualisation stages. Copilot acted as a computational partner that enhanced efficiency and accuracy.

Copilot was particularly helpful during the data wrangling phase, where it suggested efficient ways to filter and summarise datasets in R using tidyverse functions such as group_by(), summarise(), and mutate(). When constructing my visualisations, Copilot refined my ggplot code structure by suggesting appropriate geoms, aesthetic mappings, and annotation syntax. For example, when I attempted to overlay global averages onto country-level plots, Copilot proposed using geom_hline() combined with annotate() to make the reference lines visually distinct - a suggestion that both improved interpretability and reduced manual trial-and-error. These contributions demonstrate Copilot's value in providing context-aware code completion.

The most valuable aspect of Copilot was its ability to streamline repetitive coding tasks while offering alternative implementations I might not have considered. For example, it proposed a more elegant method for flipping axes with coord_flip() and hiding labels using the theme() function, which was especially useful when handling visualisations with many countries on the y-axis.

Despite its advantages, Copilot was not without limitations. Its suggestions occasionally required significant manual adjustment to fit my specific dataset structure or research intent. For example, it sometimes proposed generic variable names or aesthetic mappings that were inconsistent with my data. Additionally, Copilot tended to focus on producing visually appealing plots, but not necessarily statistically or contextually appropriate ones - requiring my own critical evaluation to ensure analytical validity.

These challenges highlighted the importance of human oversight and domain knowledge in guiding Copilot's outputs. It reinforced that Copilot functions best as a collaborator, not as an autonomous analyst.

Integrating Copilot into my workflow significantly influenced the trajectory of my assessment. It accelerated the coding process, enabling me to spend more time on data interpretation and theoretical reflection. This shift from syntax to strategy deepened my understanding of how computational tools can aid analytical thinking.