

tut2

```
# A tibble: 6 x 29
  `Geographic area` Indicator Sex `Wealth Quintile` `Series Name`
  <chr> <chr> <chr> <chr> <chr>
1 Afghanistan Neonatal mortality ra~ Total Total Multiple Ind~
2 Afghanistan Neonatal mortality ra~ Total Total Multiple Ind~
3 Afghanistan Neonatal mortality ra~ Total Total Multiple Ind~
4 Afghanistan Neonatal mortality ra~ Total Total Multiple Ind~
5 Afghanistan Neonatal mortality ra~ Total Total Multiple Ind~
6 Afghanistan Neonatal mortality ra~ Total Total Afghanistan ~
# i 24 more variables: `Series Year` <chr>, `Regional group` <chr>,
# TIME_PERIOD <chr>, OBS_VALUE <dbl>, COUNTRY_NOTES <chr>, CONNECTION <lgl>,
# DEATH_CATEGORY <lgl>, CATEGORY <chr>, `Observation Status` <chr>,
# `Unit of measure` <chr>, `Series Category` <chr>, `Series Type` <chr>,
# STD_ERR <dbl>, REF_DATE <dbl>, `Age Group of Women` <chr>,
# `Time Since First Birth` <chr>, DEFINITION <chr>, INTERVAL <dbl>,
# `Series Method` <chr>, LOWER_BOUND <dbl>, UPPER_BOUND <dbl>, ...
```

```
names(raw_igme_data)
```

[1] "Geographic area"	"Indicator"	"Sex"
[4] "Wealth Quintile"	"Series Name"	"Series Year"
[7] "Regional group"	"TIME_PERIOD"	"OBS_VALUE"
[10] "COUNTRY_NOTES"	"CONNECTION"	"DEATH_CATEGORY"
[13] "CATEGORY"	"Observation Status"	"Unit of measure"
[16] "Series Category"	"Series Type"	"STD_ERR"
[19] "REF_DATE"	"Age Group of Women"	"Time Since First Birth"
[22] "DEFINITION"	"INTERVAL"	"Series Method"
[25] "LOWER_BOUND"	"UPPER_BOUND"	"STATUS"
[28] "YEAR_TO_ACHIEVE"	"Model Used"	

```

cleaned_igme_data <-
  clean_names(raw_igme_data) |>
  filter(
    sex == "Total",
    series_name == "UN IGME estimate",
    geographic_area %in% c("Argentina", "Australia", "Canada", "Kenya"),
    indicator == "Neonatal mortality rate") |>
  select(geographic_area, time_period, obs_value)

head(cleaned_igme_data)

```

```

# A tibble: 6 x 3
  geographic_area time_period obs_value
  <chr>           <chr>         <dbl>
1 Argentina      1970-06         24.9
2 Argentina      1971-06         24.7
3 Argentina      1972-06         24.6
4 Argentina      1973-06         24.6
5 Argentina      1974-06         24.5
6 Argentina      1975-06         24.1

```

```

cleaned_igme_data <-
  cleaned_igme_data |>
  mutate(
    time_period = str_remove(time_period, "-06"),
    time_period = as.integer(time_period)
  ) |>
  filter(time_period >= 1971) |>
  rename(nmr = obs_value, year = time_period, country = geographic_area)

head(cleaned_igme_data)

```

```

# A tibble: 6 x 3
  country    year  nmr
  <chr>    <int> <dbl>
1 Argentina  1971  24.7
2 Argentina  1972  24.6
3 Argentina  1973  24.6
4 Argentina  1974  24.5

```

```
5 Argentina 1975 24.1
6 Argentina 1976 23.3
```

```
write_csv(x = cleaned_igme_data, file = "cleaned_igme_data.csv")

#### Explore ####
cleaned_igme_data <-
  read_csv(
    file = "cleaned_igme_data.csv",
    show_col_types = FALSE
  )

head(cleaned_igme_data)
```

```
# A tibble: 6 x 3
  country    year  nmr
  <chr>      <dbl> <dbl>
1 Argentina 1971  24.7
2 Argentina 1972  24.6
3 Argentina 1973  24.6
4 Argentina 1974  24.5
5 Argentina 1975  24.1
6 Argentina 1976  23.3
```

```
cleaned_igme_data |>
  ggplot(aes(x = year, y = nmr, color = country)) +
  geom_point() +
  theme_minimal()
```

