CPC Assignment 2024

Aryan Gupta

May 11,2024

1 Data Exploration and Analysis Using R

This report presents an analysis of four datasets from the `openintro` package in R. The datasets used are `census`, `sowc_child_mortality`, `sowc_demographics`, and `sowc_maternal_newborn`. The analysis aims to answer specific questions and perform various data manipulation tasks.

1.1 Remove NA values from the personal income variable in the census data

Methodology

We observe from the summary of dataset `census` that the variable `total_personal_income` has 108 NAs. To remove NA values from the `total_personal_income` variable, we created a new dataset `census_no_na` by filtering out rows where `total_personal_income` is NA.

```
census_no_na <- census[!is.na(census$total_personal_income), ]
```

Result

The new dataset `census_no_na` has 392 rows, compared to the original 500 rows in the `census` dataset, as the rows with NA values in `total_personal_income` were removed.

1.2 Observe the 25th to 35th observations of the census data

Methodology

To view the 25th to 35th observations of the `census` data, we used the following code:

```
print(census[25:35, ])
```

Result

This code prints a subset of the `census` dataset containing rows 25 through 35.

	census_year	state_fips_code	total_family_income	age	sex	race_general	marital_status	total_personal_income
1	2000	Florida	90000	12	Male	White	Never married/single	NA
2	2000	Florida	38320	47	Male	White	Married/spouse present	34320
3	2000	Florida	103700	8	Female	White	Never married/single	NA
4	2000	Florida	0	67	Male	Black	Widowed	8400
5	2000	Florida	70700	17	Female	White	Never married/single	3700
6	2000	Florida	64800	69	Female	White	Divorced	4800
7	2000	Florida	60000	55	Male	White	Married/spouse present	53000
8	2000	Florida	118100	18	Female	White	Never married/single	7500
9	2000	Florida	21000	66	Female	White	Married/spouse present	6000
10	2000	Florida	40000	58	Female	White	Married/spouse present	0
11	2000	Florida	17300	21	Male	Black	Never married/single	4800

1.3 Find the mean total income for every person in the census dataset

Methodology

To find the mean total income for every person in the `census` dataset, we used the following code:

```
mean_total_income <- mean(census$total_personal_income, na.rm = TRUE)
cat("Mean total income for every person in the census dataset:",
mean_total_income, "\n")
```

The `na.rm = TRUE` argument ensures that NA values are removed before calculating the mean.

Result

The mean total income for the `census` dataset comes out to be **29081.72**.

1.4 Join `sowc_demographics` and `sowc_child_mortality` datasets

Methodology

We performed a merge between the `sowc_demographics` and `sowc_child_mortality` datasets based on the `countries_and_areas` column. The resulting dataset, `joined_data`, contains three variables: `countries_and_areas`, `life_expectancy_2018`, and `under5_mortality_2018`.

```
joined_data <- merge(sowc_demographics, sowc_child_mortality,
by = "countries_and_areas", suffixes = c("", ".drop"))

joined_data <- joined_data[, c("countries_and_areas",
"life_expectancy_2018", "under5_mortality_2018")]
```

Result

The `joined_data` dataset contains the following variables:

	countries_and_areas	life_expectancy_2018	under5_mortality_2018
1	Afghanistan	64	62
2	Albania	78	9
3	Algeria	77	23
4	Andorra	NA	3
5	Angola	61	77
6	Antigua and Barbuda	77	6

1.5 Left join `joined_data` with `sowc_maternal_newborn` dataset

Methodology

We performed a left join between the `joined_data` and `sowc_maternal_newborn` datasets based on the `countries_and_areas` column. This ensures that all rows from `joined_data` are retained, and missing values are introduced for countries not present in `sowc_maternal_newborn`.

```
joined_data <- merge(joined_data, sowc_maternal_newborn,
by = "countries_and_areas", all.x = TRUE)
```

Result

The resulting `joined_data` now includes additional variables from the `sowc_maternal_newborn` dataset, such as `life_expectancy_female`, `family_planning_1549`, `adolescent_birth_rate`, `antenatal_care_1`, and others.

1.6 Create a variable `mean_life_exp` in `sowc_demographics`

Methodology

We created a new variable `mean_life_exp` in the `sowc_demographics` dataset, which calculates the mean life expectancy for each country using the life expectancy values from 1970, 2000, and 2018.

```
sowc_demographics$mean_life_exp <- rowMeans(sowc_demographics[,
c("life_expectancy_1970", "life_expectancy_2000", "life_expectancy_2018")],
na.rm = TRUE)
```

The `rowMeans()` function is used to calculate the mean across the specified columns, and `na.rm = TRUE` ensures that NA values are ignored when computing the mean.

Result

The `mean_life_exp` variable is added to the `sowc_demographics` dataset. Here are the first few rows:

```
  mean_life_exp
1      52.33333
2      73.00000
3      66.00000
4           NaN
5      49.66667
6           NaN
```

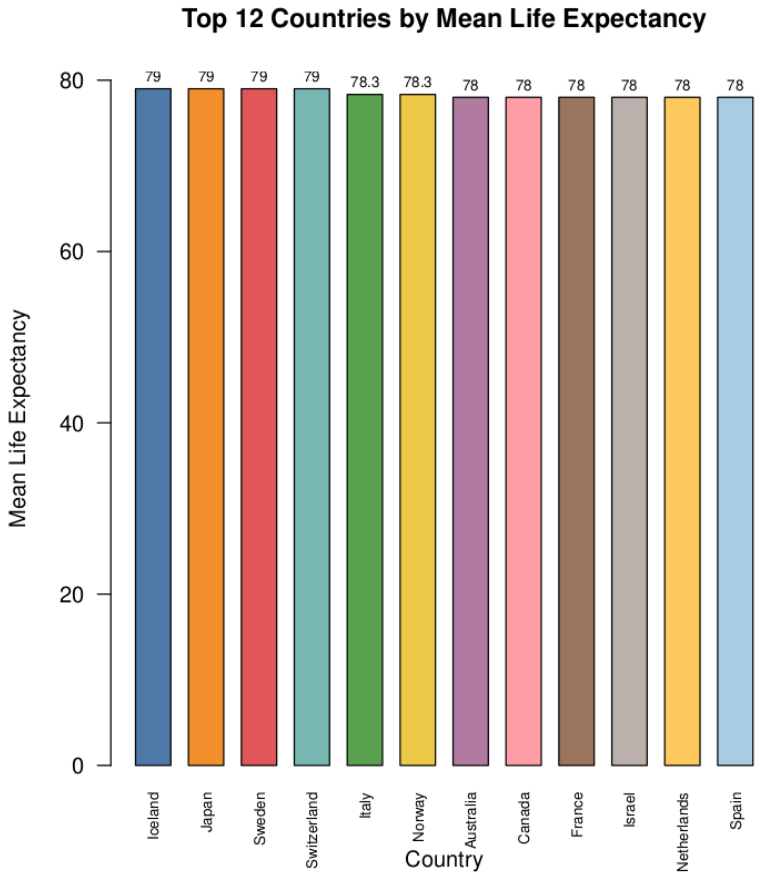
1.7 Bar chart of top 12 countries with the highest `mean_life_exp`

Methodology

We created a bar chart to visualize the top 12 countries with the highest `mean_life_exp`. The chart includes customized colors, labels, axis titles, and other formatting.

Result

The resulting bar chart provides a visual representation of the top 12 countries with the highest mean life expectancy, with labels displaying the exact values.



1.8 Countries present in `sowc_maternal_newborn` but not in `sowc_child_mortality`

Methodology

To find the countries present in `sowc_maternal_newborn` but not in `sowc_child_mortality`, we used the `setdiff()` function.

```
countries_diff <- setdiff(sowc_maternal_newborn$countries_and_areas,
                          sowc_child_mortality$countries_and_areas)
cat("Countries present in sowc_maternal_newborn but not in
sowc_child_mortality:", "\n", paste(countries_diff, collapse = ","), "\n")
```

Result

The output lists the countries that are present in `sowc_maternal_newborn` but not in `sowc_child_mortality`. Countries present in sowc_maternal_newborn but not in sowc_child_mortality: **Anguilla, British Virgin Islands, Holy See, Liechtenstein, Montserrat, Tokelau, Turks and Caicos Islands**

1.9 Highest personal income for each race

Methodology

To find the highest personal income for each race in the `census` dataset, we used the `aggregate()` function.

```
highest_income_by_race <- aggregate(total_personal_income ~ race_general,
                                    data = census, max, na.rm = TRUE)
```

Result

This code groups the `census` data by `race_general` and applies the `max` function to the `total_personal_income` variable, effectively finding the highest personal income for each race group.

	race_general	total_personal_income
1	American Indian or Alaska Native	26700
2	Black	100000
3	Chinese	34600
4	Japanese	45000
5	Other	40600
6	Other Asian or Pacific Islander	317000
7	Two major races	69000
8	White	456000

1.10 Fill NA values in `under18_pop_2018` with 0

Methodology

In the `sowc_demographics` dataset, we filled the NA values in the `under18_pop_2018` column with 0 using the following code:

```
sowc_demographics$under18_pop_2018
            [is.na(sowc_demographics$under18_pop_2018)] <- 0
```

Result

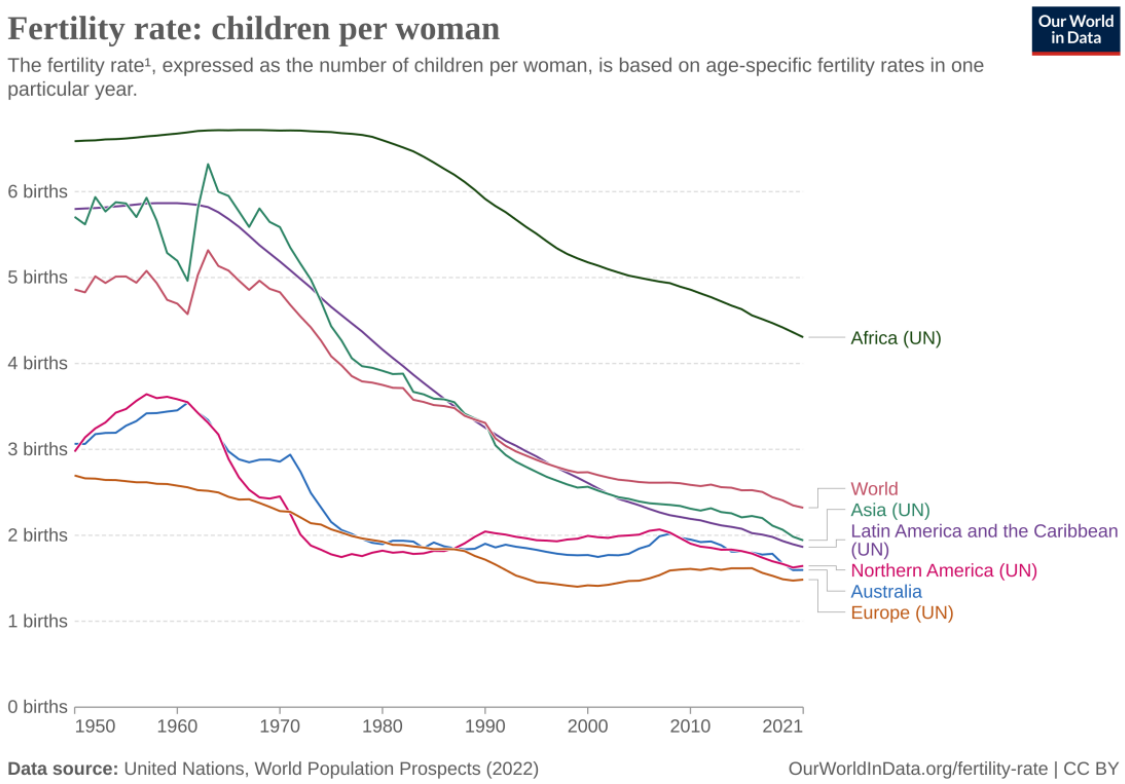
This code replaces NA values in the `under18_pop_2018` column with 0.

2 Worldwide Fertility Trends

2.1 Global Decline of Fertility Rate

One of the most striking demographic shifts of modern times has been the widespread, rapid decline in fertility rates worldwide. In 1950, the global total fertility rate (TFR) stood at a high of 4.84 children per woman. Over just seven decades, that number has halved to 2.41 children per woman as of 2021. This 49% drop represents an unprecedented fertility transition for humanity.
While the overall trajectory has been decisively downward, there is considerable variation in both current fertility levels and the pace of decline across regions [8]:

- Africa: Africa maintains the highest regional fertility rate at 4.3 children per woman in 2021, down from 6.6 in 1950 - a 35% reduction over 70 years.
- Europe/North America/Australia: These regions collectively had the lowest fertility rates of 1.4-1.6 children per woman in 2021, declining from 2.5-3.2 in 1950 - a 40-55% drop.
- Asia/Latin America: Perhaps the most dramatic regional shift occurred in Asia and Latin America, which plummeted from very high fertility of around 5.8 children per woman in 1950 to just 1.8-2.0 in 2021 - a staggering 66-69% decline.

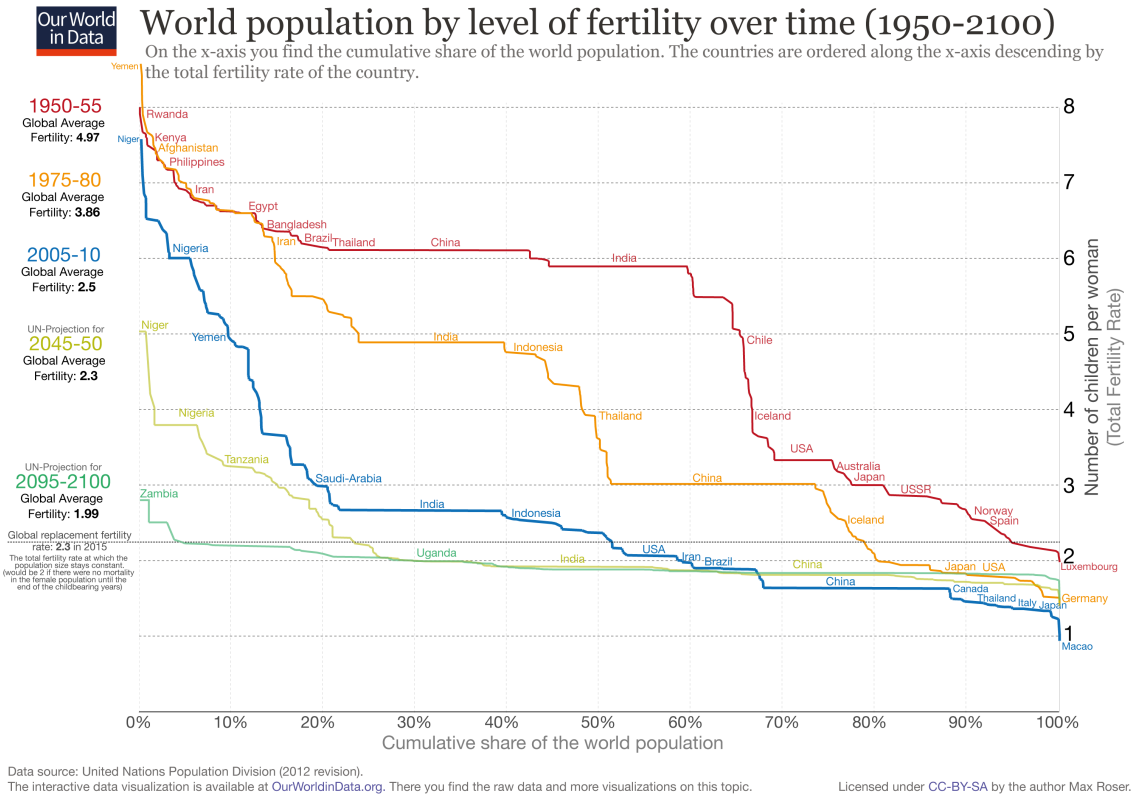


2.2 Pace of Transitions

A remarkable insight is the extremely rapid pace at which some nations transitioned from very high to low fertility over just a single decade. Iran took only 10 years to drop from over 6 children per woman to under 3 [1], while China made the same drastic transition in just 11 years largely before its one-child policy [7] . In contrast, it took 95 years for the United Kingdom and 82 years for the United States to achieve that level of reduction in the 19th century.

2.3 Convergence of Rates

Another vital trend identified in the data is the convergence of fertility rates across countries as the era has progressed. Whereas there was immense variance globally in the 1950s, with rates ranging from 1.6 to over 8 children per woman, by 2022 over 80% of the world's population resided in countries with TFRs tightly clustered between 1-3 [8]. Highly divergent outliers persist, with nations like Niger (6.9), Somalia (6.3), and DR Congo (6.1) contrasted by South Korea (0.9),Taiwan (1.1) and Singapore (1.1) [8]. However, UN projections indicate this divide will continue narrowing, with global fertility expected just below the replacement level of 2.1 by 2100 [8].



2.4 Drivers from Past Literature

So what factors identified in past literature can help explain these tectonic shifts in global fertility patterns? Extensive research points to three key drivers:

1. Increased empowerment and educational/economic opportunities for women, enabling greater autonomy over reproductive decisions and making high fertility less necessary or desirable. As women's labor force participation rose, the opportunity costs of having many children also increased [3, 9].
2. Declining infant and child mortality rates globally. When a smaller proportion of children die before age 5, parents require fewer births to achieve desired family sizes [3].
3. The diminishing economic value and increasing costs of large family sizes as societies transitioned from agricultural to industrialized economies that no longer relied on child labor [3].

2.5 Additional Factors

Beyond these pivotal socioeconomic transformations, additional qualitative factors like urbanization, personal fertility preferences shaped by gender norms/roles, and policy interventions like national family planning programs have also been theorized to impact fertility patterns [5, 2, 6].
Urbanization in particular appears tightly linked to lower fertility, as cities provide greater access to education, employment prospects for women, and contraceptive services - all facilitating smaller family sizes [4]. However, studies reveal nuanced variations based on socioeconomic status even within urban areas. Cultural fertility preferences rooted in traditional gender norms and perceived costs/benefits of children have proven highly influential as well [2].
In many contexts, pronatalist government policies promoting higher birth rates through financial incentives and childcare support have aimed to combat perceived threats of long-term population aging and labor shortages resulting from sub-replacement fertility levels [6]. Conversely, other nations prioritized voluntary family planning initiatives to empower reproductive choices.

References

[1] M. J. Abbasi-Shavazi, P. McDonald, and M. Hosseini-Chavoshi. The fertility transition in iran: Revolution and reproduction. 2009.

[2] C. A. Bachrach and S. P. Morgan. A cognitive-social model of fertility intentions. *Population and Development Review*, 39(3):459–485, 2013.

[3] J. Bongaarts. Completing the fertility transition in the developing world: The role of educational differences and fertility preferences. *Population Studies*, 57(3):321–335, 2003.

[4] I. Günther and K. Harttgen. Desired fertility and number of children born across time and space. *Demography*, 53(1):55–83, 2016.

[5] M. Lerch. Fertility decline in urban and rural areas of developing countries. *Population and Development Review*, 45(2):301–320, 2019.

[6] A. Luci-Greulich and O. Thévenon. The impact of family policies on fertility trends in developed countries. *European Journal of Population*, 29(4):387–416, 2013.

[7] X. Peng. China's demographic history and future challenges. *Science*, 333(6042):581–587, 2011.

[8] United Nations, Department of Economic and Social Affairs, Population Division. World population prospects 2022. 2022.

[9] U. D. Upadhyay, J. D. Gipson, M. Withers, S. Lewis, E. J. Ciaraldi, A. Fraser, and N. Prata. Women's empowerment and fertility: A review of the literature. *Social Science & Medicine*, 115:111–120, 2014.