DTP coverage

Tristan Verbeek, 2864049

2025-06-19

Set-up your environment

```
library(tidyverse)
```

```
## -- Attaching core tidyverse packages ------------------------ tidyverse 2.0.0 --
## v dplyr     1.1.4     v readr     2.1.5
## v forcats   1.0.0     v stringr   1.5.1
## v ggplot2   3.5.2     v tibble    3.2.1
## v lubridate 1.9.4     v tidyr     1.3.1
## v purrr     1.0.4
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors
```

```
library(readxl)
library(dplyr)
```

Title Page

By: G.J.R. Weydemuller (student number)

L. Waaijenberg (2862154)

P. Gandra Rodrigues (2870539)

Q. Faber (2861921)

R.A. Kali (2850680)

T. Adam (student number)

T.M.V. Verbeek (student number)

Tutor: J.F. Fitzgerald

Tutorial group: Tutorial group 1

Course: Programming for Economists

Institute: Vrije Universiteit Amsterdam

Part 1 - Identify a Social Problem

Use APA referencing throughout your document. Here's a link to some explanation.

1.1 Describe the Social Problem

Why is DTP vaccination a social problem?

DTP vaccination, protecting against diphtheria, tetanus, and pertussis, is a cornerstone of global public health. It is one of the most cost-effective interventions, preventing an estimated 4.4 million deaths annually (unicef, 2024). Despite its proven effectiveness and inclusion in nearly all national immunization schedules, millions of children worldwide remain under-vaccinated or entirely unvaccinated. In 2023 alone, approximately 21 million children were either unvaccinated or under-vaccinated, with 14.5 million receiving no vaccines at all (unicef, 2024). This gap in coverage leads to preventable illness, long-term disability, and death, particularly in low- and middle-income countries (Our World in Data, sd). The issue is not merely medical but deeply social, as it reflects and reinforces inequalities in access to healthcare, education, and infrastructure (unicef, 2024).

Vaccination coverage is also a proxy for the strength of a country's healthcare system. Low DTP coverage often signals broader systemic issues such as poverty, conflict, misinformation, and weak governance. These factors disproportionately affect vulnerable populations, making DTP vaccination a critical indicator of social justice and equity (unicef, 2024).

Research question: "How does government healthcare spending per capita correlate with national vaccination coverage rates for routine immunizations in low-, middle-, and high-income countries?"

Sources identifying DTP vaccination as a social problem

World Health Organization (WHO) and UNICEF have jointly warned that vaccine-preventable diseases, including diphtheria, are resurging due to declining immunization coverage, misinformation, and funding cuts. They emphasize that millions of children are at risk, and urgent investment in immunization is needed to prevent a reversal of decades of progress (WHO Media Team, 2025).

UNICEF outlines the severe consequences of DTP-preventable diseases. Diphtheria can cause heart and nerve damage, tetanus is often fatal even with treatment, and pertussis can lead to pneumonia and death in infants. These diseases disproportionately affect children in low-resource settings, highlighting the social dimension of under-vaccination (UNICEF, sd).

A UNICEF data report on DTP vaccine dropout rates shows that millions of children start but do not complete the DTP series. This dropout is especially prevalent in low-income countries and is linked to systemic barriers such as healthcare access, education, and infrastructure (Wang; et al., 2019).

What has not been fully researched

While many studies have examined DTP coverage and its determinants, there is limited comparative analysis across income groups that links government healthcare spending per capita directly to national DTP vaccination rates. Most existing research focuses on either micro-level determinants (e.g., parental education, rural access) or macro-level trends without integrating economic policy variables. UNICEF notes that while DTP3 coverage is widely used as a marker of how well countries provide routine immunization, disparities persist, and economic factors such as national healthcare investment are often underexplored in comparative frameworks (UNICEF, 2024).

Contribution of this report

This report aims to fill that gap by analyzing how government healthcare spending correlates with DTP vaccination coverage across low-, middle-, and high-income countries. This research will provide new insights into whether increased public investment in health translates into better immunization outcomes. This could inform both global health policy and economic development strategies, especially in resource-constrained settings.

Part 2 - Data Sourcing

2.1 Load in the data

Preferably from a URL, but if not, make sure to download the data and store it in a shared location that you can load the data in from. Do not store the data in a folder you include in the Github repository!

```
# Load data on health spending  directly from Dropbox (public direct link)
url_spending <- "https://www.dropbox.com/scl/fi/01df0ly3ornjgcis9kwsk/CHE-per-region.csv?rlkey=nwvmxv33

# Load data on health coverage  directly from Dropbox (public direct link)
url_coverage <- "https://www.dropbox.com/scl/fi/08rf7e6sj3hcq3e5fnp81/datasetcoverage.xlsx?rlkey=x9iwci

# Download the file temporarily and read it in
temp_file <- tempfile(fileext = ".xlsx")
download.file(url_coverage, destfile = temp_file, mode = "wb")
raw_coverage_data <- read_excel(temp_file)

# Load data on child deaths directly from Dropbox (public direct link)
url_childdeath <- "https://www.dropbox.com/scl/fi/wuopfxxyux4dccc9rbibn/child-death.csv?rlkey=mfqw1b765

# Read the CSV into a dataframe
raw_healthspending_data <- read.csv(url_spending)

# Read the CSV into a dataframe
raw_childdeath_data <- read.csv(url_childdeath)


# Preview the first few rows to check the import
head(raw_healthspending_data)
```

```
##            IndicatorCode                                          Indicator
## 1 GHED_CHE_pc_US_SHA2011 Current health expenditure (CHE) per capita in US$
## 2 GHED_CHE_pc_US_SHA2011 Current health expenditure (CHE) per capita in US$
## 3 GHED_CHE_pc_US_SHA2011 Current health expenditure (CHE) per capita in US$
## 4 GHED_CHE_pc_US_SHA2011 Current health expenditure (CHE) per capita in US$
## 5 GHED_CHE_pc_US_SHA2011 Current health expenditure (CHE) per capita in US$
## 6 GHED_CHE_pc_US_SHA2011 Current health expenditure (CHE) per capita in US$
##   ValueType ParentLocationCode ParentLocation Location.type SpatialDimValueCode
## 1   numeric             GLOBAL         Global    WHO region                AMR
## 2   numeric             GLOBAL         Global    WHO region                AFR
## 3   numeric             GLOBAL         Global    WHO region                WPR
## 4   numeric             GLOBAL         Global    WHO region               SEAR
## 5   numeric             GLOBAL         Global    WHO region                EUR
```

```
## 6   numeric             GLOBAL         Global    WHO region                 EMR
##                Location Period.type Period IsLatestYear Dim1.type Dim1
## 1              Americas        Year   2022         true        NA   NA
## 2                Africa        Year   2022         true        NA   NA
## 3       Western Pacific        Year   2022         true        NA   NA
## 4       South-East Asia        Year   2022         true        NA   NA
## 5                Europe        Year   2022         true        NA   NA
## 6 Eastern Mediterranean        Year   2022         true        NA   NA
##   Dim1ValueCode Dim2.type Dim2 Dim2ValueCode Dim3.type Dim3 Dim3ValueCode
## 1            NA        NA   NA            NA        NA   NA            NA
## 2            NA        NA   NA            NA        NA   NA            NA
## 3            NA        NA   NA            NA        NA   NA            NA
## 4            NA        NA   NA            NA        NA   NA            NA
## 5            NA        NA   NA            NA        NA   NA            NA
## 6            NA        NA   NA            NA        NA   NA            NA
##   DataSourceDimValueCode DataSource FactValueNumericPrefix FactValueNumeric
## 1                     NA         NA                     NA           1256.0
## 2                     NA         NA                     NA            132.6
## 3                     NA         NA                     NA           1364.0
## 4                     NA         NA                     NA            241.0
## 5                     NA         NA                     NA           2938.0
## 6                     NA         NA                     NA            544.2
##   FactValueUoM FactValueNumericLowPrefix FactValueNumericLow
## 1           NA                        NA                  NA
## 2           NA                        NA                  NA
## 3           NA                        NA                  NA
## 4           NA                        NA                  NA
## 5           NA                        NA                  NA
## 6           NA                        NA                  NA
##   FactValueNumericHighPrefix FactValueNumericHigh   Value
## 1                         NA                   NA 1256.40
## 2                         NA                   NA  132.58
## 3                         NA                   NA 1363.92
## 4                         NA                   NA  240.99
## 5                         NA                   NA 2937.83
## 6                         NA                   NA  544.24
##   FactValueTranslationID FactComments Language             DateModified
## 1                     NA           NA       EN 2024-12-09T23:00:00.000Z
## 2                     NA           NA       EN 2024-12-09T23:00:00.000Z
## 3                     NA           NA       EN 2025-04-03T22:00:00.000Z
## 4                     NA           NA       EN 2025-04-03T22:00:00.000Z
## 5                     NA           NA       EN 2024-12-09T23:00:00.000Z
## 6                     NA           NA       EN 2024-12-09T23:00:00.000Z
```

```
head(raw_coverage_data)
```

```
## # A tibble: 6 x 11
##   GROUP       CODE  NAME      YEAR ANTIGEN ANTIGEN_DESCRIPTION COVERAGE_CATEGORY
##   <chr>       <chr> <chr>    <dbl> <chr>   <chr>               <chr>
## 1 WHO_REGIONS AFR   African~  2023 DTPCV3  DTP-containing vac~ WUENIC
## 2 WHO_REGIONS EMR   Eastern~  2023 DTPCV3  DTP-containing vac~ WUENIC
## 3 WHO_REGIONS EUR   Europea~  2023 DTPCV3  DTP-containing vac~ WUENIC
## 4 WHO_REGIONS AMR   Region ~  2023 DTPCV3  DTP-containing vac~ WUENIC
## 5 WHO_REGIONS SEAR  South-E~  2023 DTPCV3  DTP-containing vac~ WUENIC
```

```
## 6 WHO_REGIONS AFR   African~  2022 DTPCV3  DTP-containing vac~ WUENIC
## # i 4 more variables: COVERAGE_CATEGORY_DESCRIPTION <chr>, TARGET_NUMBER <dbl>,
## #   DOSES <dbl>, COVERAGE <dbl>
```

```
head(raw_healthspending_data)
```

```
##            IndicatorCode                                          Indicator
## 1 GHED_CHE_pc_US_SHA2011 Current health expenditure (CHE) per capita in US$
## 2 GHED_CHE_pc_US_SHA2011 Current health expenditure (CHE) per capita in US$
## 3 GHED_CHE_pc_US_SHA2011 Current health expenditure (CHE) per capita in US$
## 4 GHED_CHE_pc_US_SHA2011 Current health expenditure (CHE) per capita in US$
## 5 GHED_CHE_pc_US_SHA2011 Current health expenditure (CHE) per capita in US$
## 6 GHED_CHE_pc_US_SHA2011 Current health expenditure (CHE) per capita in US$
##   ValueType ParentLocationCode ParentLocation Location.type SpatialDimValueCode
## 1   numeric             GLOBAL         Global    WHO region                 AMR
## 2   numeric             GLOBAL         Global    WHO region                 AFR
## 3   numeric             GLOBAL         Global    WHO region                 WPR
## 4   numeric             GLOBAL         Global    WHO region                SEAR
## 5   numeric             GLOBAL         Global    WHO region                 EUR
## 6   numeric             GLOBAL         Global    WHO region                 EMR
##               Location Period.type Period IsLatestYear Dim1.type Dim1
## 1             Americas        Year   2022         true        NA   NA
## 2               Africa        Year   2022         true        NA   NA
## 3      Western Pacific        Year   2022         true        NA   NA
## 4      South-East Asia        Year   2022         true        NA   NA
## 5               Europe        Year   2022         true        NA   NA
## 6 Eastern Mediterranean        Year   2022         true        NA   NA
##   Dim1ValueCode Dim2.type Dim2 Dim2ValueCode Dim3.type Dim3 Dim3ValueCode
## 1            NA        NA   NA            NA        NA   NA            NA
## 2            NA        NA   NA            NA        NA   NA            NA
## 3            NA        NA   NA            NA        NA   NA            NA
## 4            NA        NA   NA            NA        NA   NA            NA
## 5            NA        NA   NA            NA        NA   NA            NA
## 6            NA        NA   NA            NA        NA   NA            NA
##   DataSourceDimValueCode DataSource FactValueNumericPrefix FactValueNumeric
## 1                     NA         NA                     NA           1256.0
## 2                     NA         NA                     NA            132.6
## 3                     NA         NA                     NA           1364.0
## 4                     NA         NA                     NA            241.0
## 5                     NA         NA                     NA           2938.0
## 6                     NA         NA                     NA            544.2
##   FactValueUoM FactValueNumericLowPrefix FactValueNumericLow
## 1           NA                        NA                  NA
## 2           NA                        NA                  NA
## 3           NA                        NA                  NA
## 4           NA                        NA                  NA
## 5           NA                        NA                  NA
## 6           NA                        NA                  NA
##   FactValueNumericHighPrefix FactValueNumericHigh   Value
## 1                         NA                   NA 1256.40
## 2                         NA                   NA  132.58
## 3                         NA                   NA 1363.92
## 4                         NA                   NA  240.99
## 5                         NA                   NA 2937.83
```

```
## 6                          NA                   NA  544.24
##   FactValueTranslationID FactComments Language             DateModified
## 1                     NA           NA       EN 2024-12-09T23:00:00.000Z
## 2                     NA           NA       EN 2024-12-09T23:00:00.000Z
## 3                     NA           NA       EN 2025-04-03T22:00:00.000Z
## 4                     NA           NA       EN 2025-04-03T22:00:00.000Z
## 5                     NA           NA       EN 2024-12-09T23:00:00.000Z
## 6                     NA           NA       EN 2024-12-09T23:00:00.000Z
```

```
head(raw_coverage_data)
```

```
## # A tibble: 6 x 11
##   GROUP       CODE  NAME      YEAR ANTIGEN ANTIGEN_DESCRIPTION COVERAGE_CATEGORY
##   <chr>       <chr> <chr>    <dbl> <chr>   <chr>               <chr>
## 1 WHO_REGIONS AFR   African~  2023 DTPCV3  DTP-containing vac~ WUENIC
## 2 WHO_REGIONS EMR   Eastern~  2023 DTPCV3  DTP-containing vac~ WUENIC
## 3 WHO_REGIONS EUR   Europea~  2023 DTPCV3  DTP-containing vac~ WUENIC
## 4 WHO_REGIONS AMR   Region ~  2023 DTPCV3  DTP-containing vac~ WUENIC
## 5 WHO_REGIONS SEAR  South-E~  2023 DTPCV3  DTP-containing vac~ WUENIC
## 6 WHO_REGIONS AFR   African~  2022 DTPCV3  DTP-containing vac~ WUENIC
## # i 4 more variables: COVERAGE_CATEGORY_DESCRIPTION <chr>, TARGET_NUMBER <dbl>,
## #   DOSES <dbl>, COVERAGE <dbl>
```

```
inline_code = TRUE
```

2.2 Provide a short summary of the dataset(s)

Source 1: DTP Vaccination Coverage

- Credibility: This data is published by the World Health Organization (WHO), the leading global authority on public health. WHO collaborates with national governments and UNICEF to compile and validate immunization data.

- Data Quality: The dataset includes WHO/UNICEF Estimates of National Immunization Coverage (WUENIC), which are based on a combination of administrative data, household surveys, and expert review. This triangulation improves accuracy and reliability.

- Relevance: It directly provides the DTP3 coverage rates, which are central to your research question. The data is also disaggregated by region and year, allowing for temporal and cross-regional analysis.

Source 2: Government Healthcare Spending

- Credibility: The World Bank is a globally respected institution that provides standardized economic and development data. It sources health expenditure data from the WHO Global Health Expenditure Database.

- Data Quality: The data is collected using internationally accepted methodologies and is updated annually. It includes consistent country-level indicators, making it ideal for cross-country comparisons.

- Relevance: This dataset provides the independent variable in your analysis—government health spending per capita—which you are correlating with vaccination coverage.

Source 3: Child Death Rates

- Credibility: This data is also from the World Health Organization, specifically its Global Health Observatory (GHO), which is the main repository for WHO's health-related statistics.

- Data Quality: The under-five mortality data is compiled from national vital registration systems, surveys, and statistical modeling, and is reviewed by global health experts.

- Relevance: While not part of your main correlation analysis, this dataset provides valuable contextual insight into the broader social impact of vaccination coverage and healthcare investment.

Complementarity of the Datasets

These datasets are complementary in structure and purpose:

· The DTP dataset provides the dependent variable (vaccination coverage).

· The health spending dataset provides the independent variable (government health expenditure per capita).

· The child death dataset provides the independent variable (child death)

Together, they allow for a cross-country analysis of how public investment in health and child death correlates with immunization outcomes.

Suitability for the Research Topic

The research question investigates the relationship between government healthcare spending and DTP vaccination coverage across income groups. These datasets are ideal because:

· They are comprehensive and global, enabling cross-country comparisons.

· They are updated annually, supporting time-series or cross-sectional analysis.

· They are standardized and well-documented, ensuring data quality and reproducibility.

Data Limitations

Despite their strengths, the datasets have limitations:

WHO DTP Coverage & WHO child death:

· Some countries have missing or estimated values, especially in conflict zones.

· Coverage estimates may be influenced by reporting biases or survey recall errors.

· The data may not fully reflect subnational disparities in vaccination access.

World Bank Health Spending:

· Spending data are in current US dollars, which may be affected by inflation and exchange rate fluctuations.

· The data do not indicate how funds are allocated (e.g., toward immunization vs. other services).

· Some countries have incomplete or outdated records for recent years.

These limitations mean that while the analysis can reveal correlations, it cannot establish causality or account for all contextual factors.

2.3 Describe the type of variables included

WHO DTP Vaccination Coverage:

- Antigen: DTP 1 and DTP3
- Coverage: vaccinated target population

- Units: percentage
- Target_Pop: target population for the given vaccine
- Years Covered: 1980–2023
- Region: WHO regions
- Frequency: Annual
- Source: WHO/UNICEF Joint Reporting Form and surveys (e.g., DHS, MICS)
- Estimation Method: Combination of administrative data, surveys, and expert review.

World Bank Health Spending:

- Indicator: Current health expenditure per capita
- Years Covered: 2000–2023
- Region: WHO regions
- Frequency: Annual
- Units: US dollars
- Source: WHO Global Health Expenditure Database[KR(1]

WHO child deaths among children under five:

- Indicator: Mortality rate under children under five
- Target_Pop: population of children aged 0–4 years
- Years covered: 1980-2023
- Region: WHO regions
- Frequency: Annual
- Source: WHO/UNICEF GHO, national civil registration, surveys (DHS, MICS)
- Estimation method: Combination of available national data (CRVS, surveys) and adjustments by statistical methods and expert review.

Part 3 - Quantifying

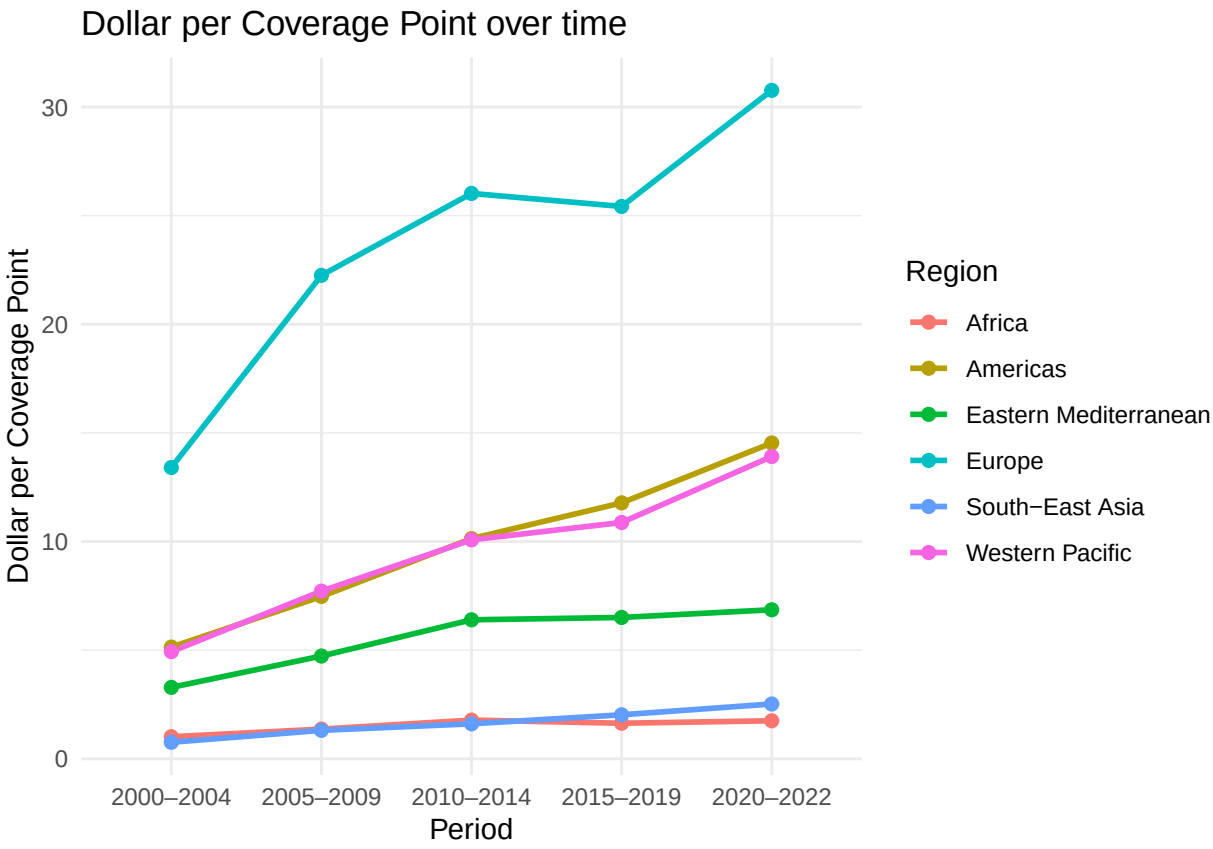
3.1 Data cleaning

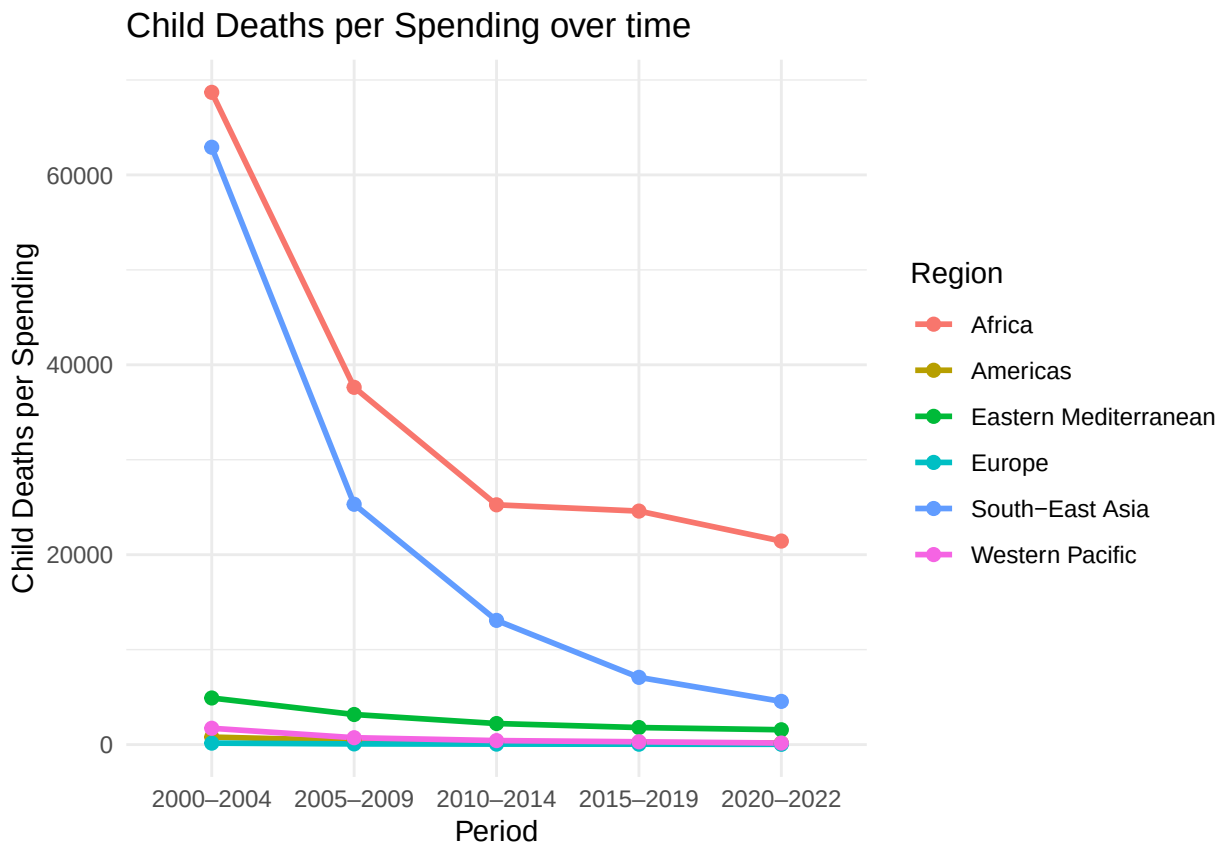
3.2 Generate necessary variables

3.25 merge datasets together

3.3 Visualize temporal variation

```
## Warning: Using ‘size‘ aesthetic for lines was deprecated in ggplot2 3.4.0.
## i Please use ‘linewidth‘ instead.
## This warning is displayed once every 8 hours.
## Call ‘lifecycle::last_lifecycle_warnings()‘ to see where this warning was
## generated.
```

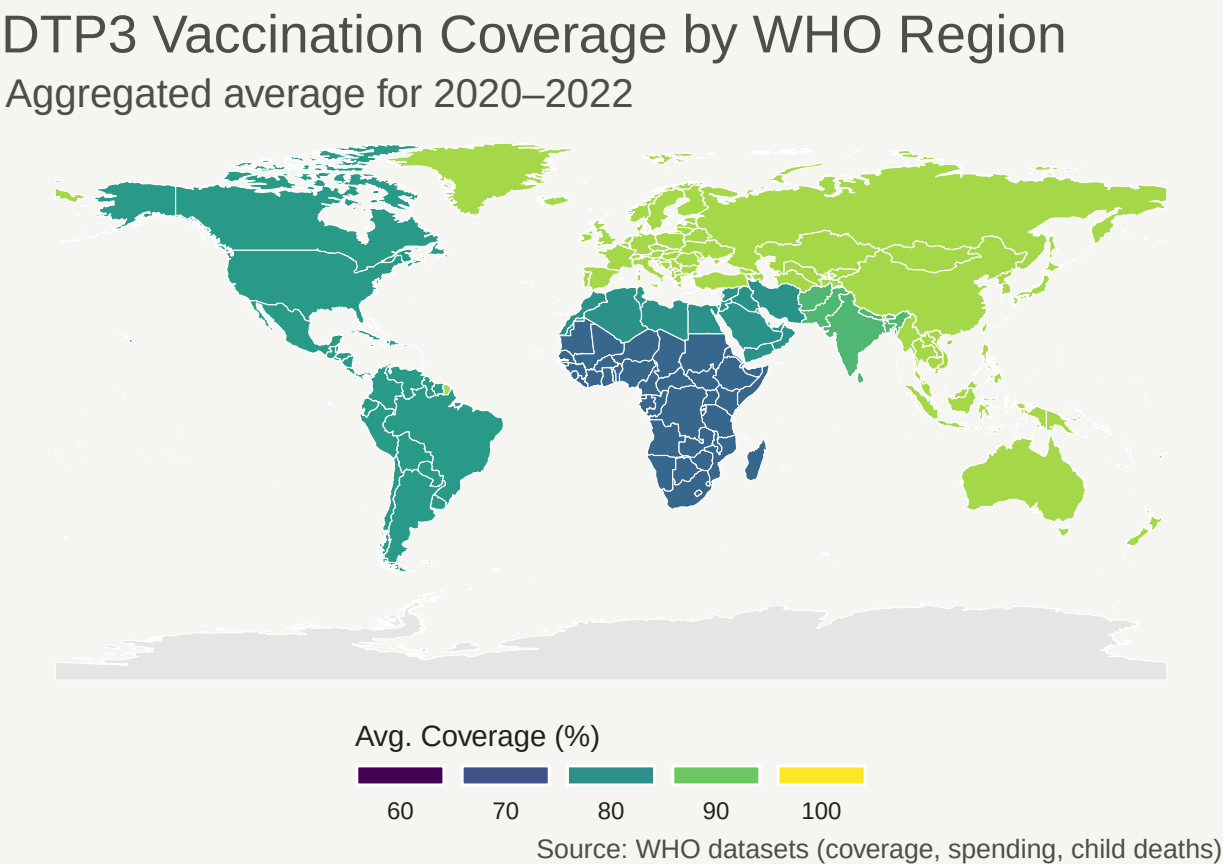

Dollar per Coverage Point over time
Dollar per Coverage Point
30
20
10
0
2000–2004
2005–2009
2010–2014
2015–2019
2020–2022
Period
Region
Africa
Americas
Eastern Mediterranean
Europe
South−East Asia
Western Pacific



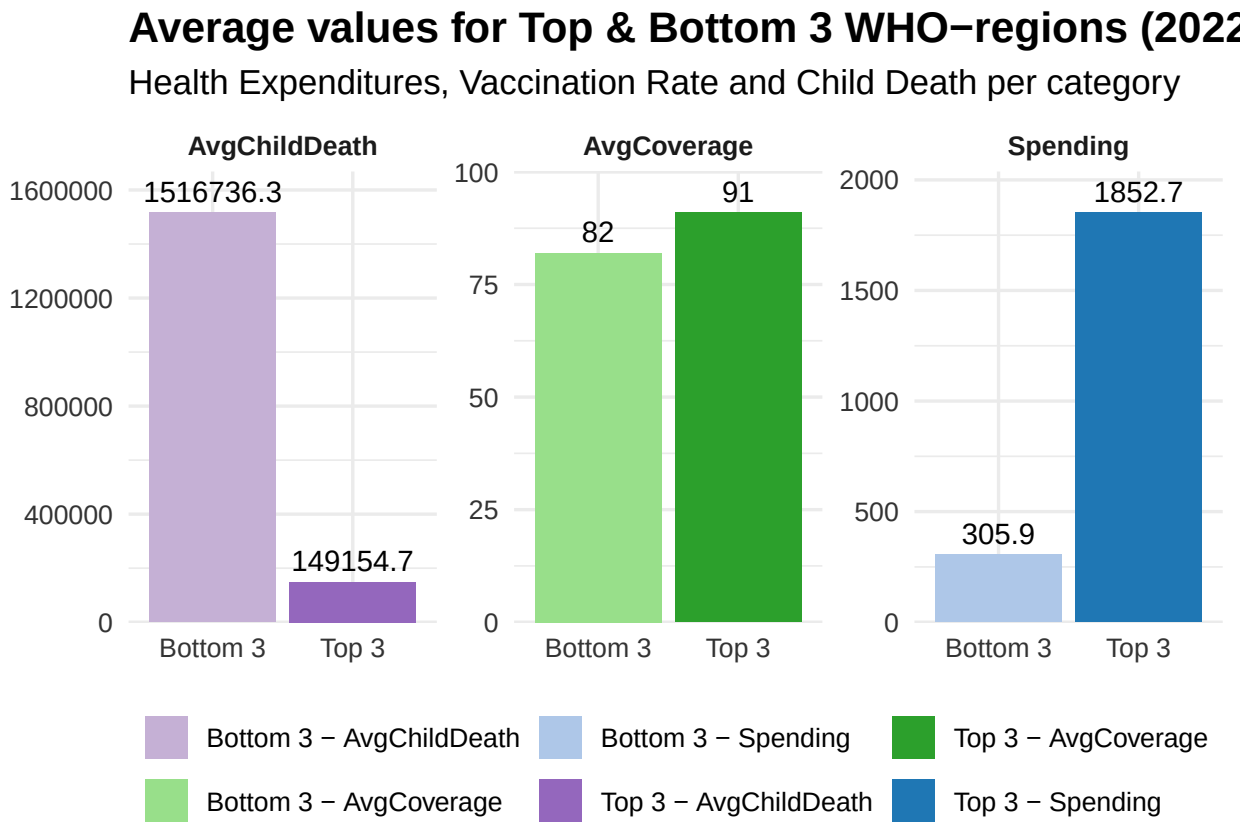
with linear trend

p2 <- ggplot(data, aes(x=my_x, y=my_y)) + geom_point() + geom_smooth(method=lm , color="red", se=FALSE) + theme_ipsum()

3.4 Visualize spatial variation

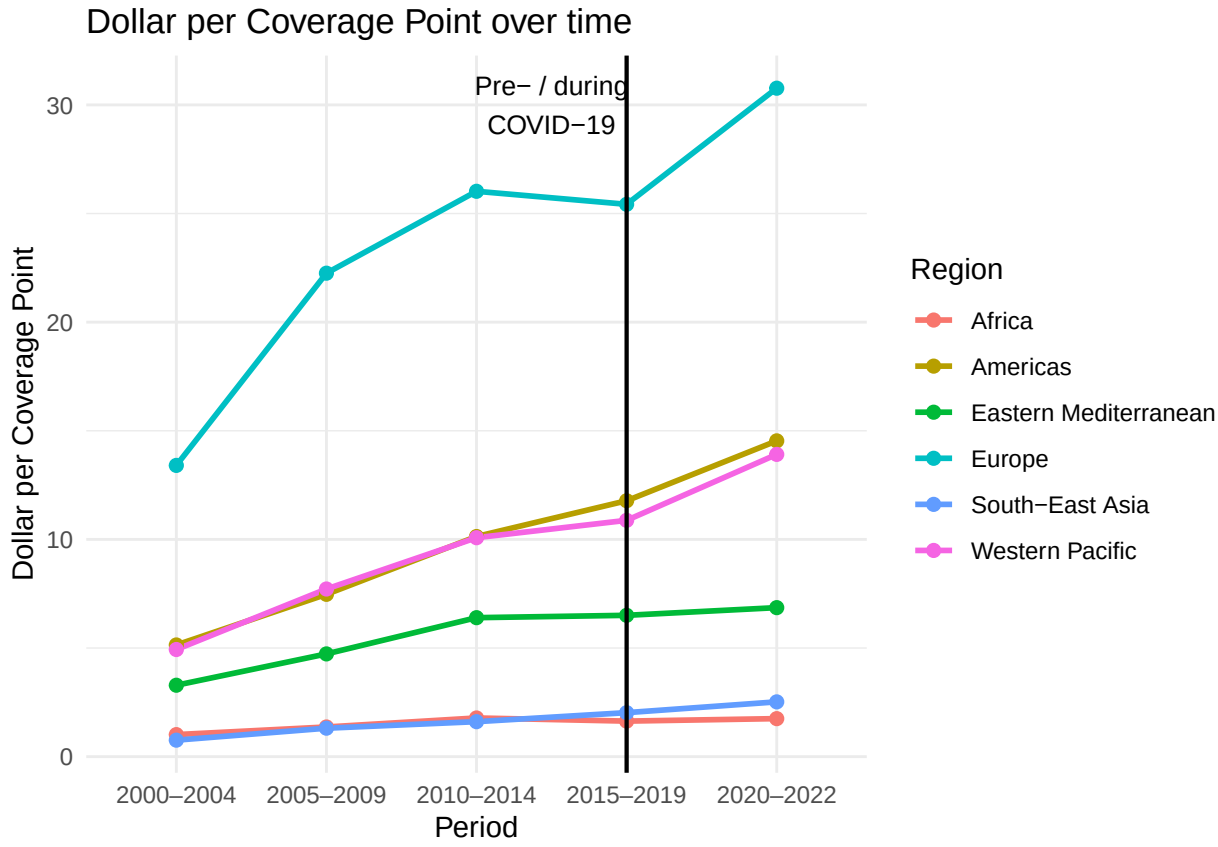


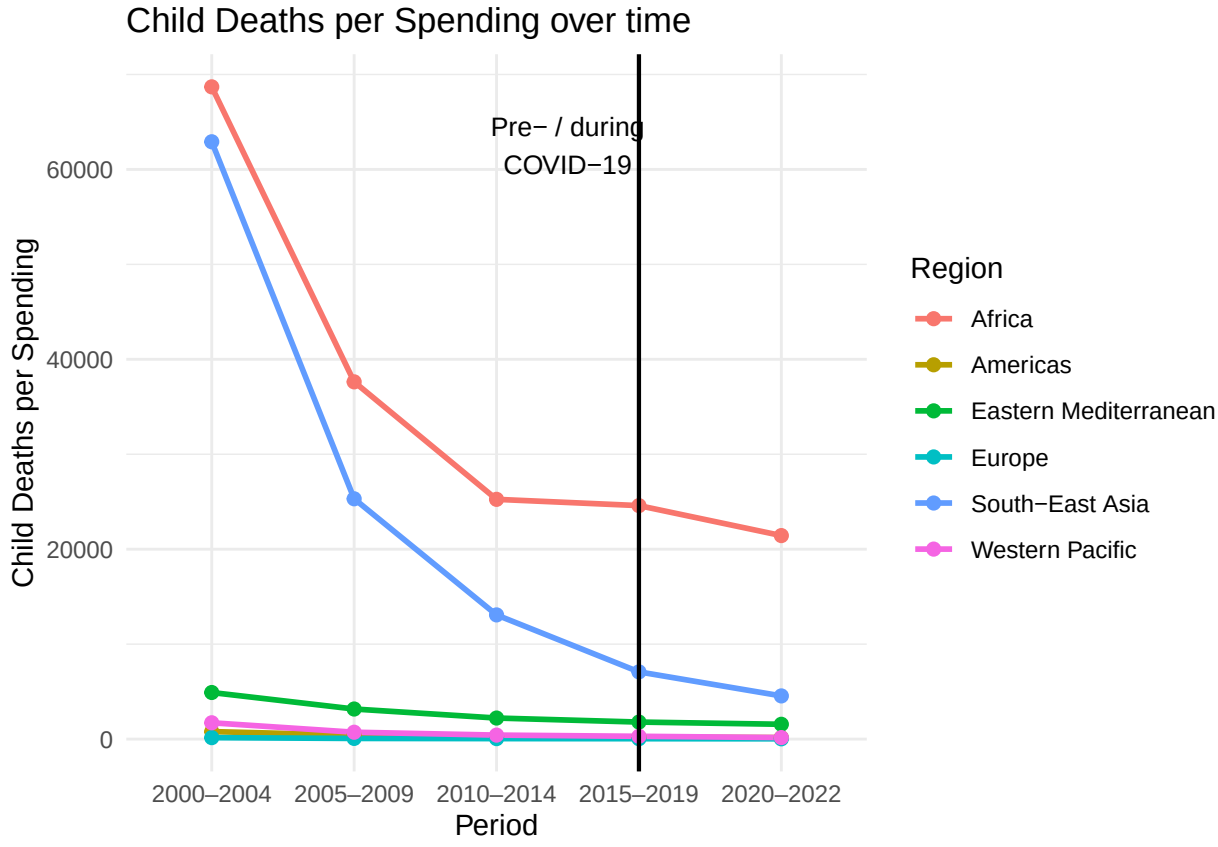
3.5 Visualize sub-population variation



Here you provide a description of why the plot above is relevant to your specific social problem.

3.6 Event analysis





Here you provide a description of why the plot above is relevant to your specific social problem.

Part 4 - Discussion

4.1 Discuss your findings

Part 5 - Reproducibility

5.1 Github repository link

Provide the link to your PUBLIC repository here: ...

5.2 Reference list

References

Our World in Data. (n.d.). *Deaths caused by vaccine-preventable diseases, World.* Retrieved from Our World in Data: https://ourworldindata.org/grapher/deaths-caused-by-vaccine-preventable-diseases-over-time

unicef. (2024, July). *Immunization.* Retrieved from unicef: https://data.unicef.org/topic/child-health/immunization/

UNICEF. (2024, March). *Vaccination & Immunization Statistics - UNICEF Data.* Retrieved from UNICEF: https://knowledge.unicef.org/resource/vaccination-immunization-statistics-unicef-data

UNICEF. (n.d.). *Vaccines and the diseases they prevent.* Retrieved from UNICEF: https://www.unicef.org/parenting/health/vaccines-and-diseases-they-prevent

WHO. (2023). *global.* Retrieved from immunizationdata: https://immunizationdata.who.int/global/wiise-detail-page/diphtheria-tetanus-toxoid-and-pertussis-(dtp)-vaccination-coverage?CODE=AFR+EMR+EUR+AMR+SEAR+WPR&ANTIGEN=&YEAR=

WHO. (2023). *Indicators.* Retrieved from who: https://www.who.int/data/gho/data/indicators/indicator-details/GHO/number-of-under-five-deaths

WHO. (2025). *Health expenditure.* Retrieved from who: https://www.who.int/data/gho/data/indicators/indicator-details/GHO/current-health-expenditure-(che)-per-capita-in-us-dollar

WHO Media Team. (2025, April 24). *Increases in vaccine-preventable disease outbreaks threaten years of progress, warn WHO, UNICEF, Gavi* . Retrieved from World Health Organization: https://www.who.int/news/item/24-04-2025-increases-in-vaccine-preventable-disease-outbreaks-threaten-years-of-progress--warn-who--unicef--gavi

Xinhu Wang; Mamadou Diallo. (2019, September 24). *UNICEF.* Retrieved from How tracking DTP vaccine dropout rates helps protect children from preventable diseases: https://data.unicef.org/data-for-action/tracking-dtp-vaccine-dropout-rates-protects-children-from-preventable-diseases/