Hello, Quarto

# R Package jmpwashdata

For this analysis we will use the **jmpwashdata** R Package by (Dickinson 2021). The package contains all data compiled by the [WHO/UNICEF Joint Monitoring Programme (JMP)](https://washdata.org).

library(jmpwashdata)  
library(tidyverse)  
library(gt)  
library(ggthemes)

# World Bank income groups

We will also use the World Bank income classification for 218 countries. This data was downloaded and stored as an XLSX file using an R script in src.

income\_groups\_df %>%   
 count(income\_group) %>%   
 gt()

| income\_group | n |
| --- | --- |
| High income | 81 |
| Upper middle income | 54 |
| Lower middle income | 54 |
| Low income | 28 |
| NA | 1 |

# Basic Sanitation & GDP

Data for the most recent year, basic sanitation in urban areas, calculate urban population, and join income groups.

# Perform data manipulation operations on the jmp\_wld\_sanitation data frame  
jmp\_wld\_sanitation\_gdp\_income <- jmp\_wld\_sanitation |>   
 # Filter the rows where the year column is equal to the maximum year value  
 filter(year == max(year)) |>   
 # Select the columns from name to prop\_u and the san\_bas\_u column  
 select(name:prop\_u, san\_bas\_u) |>   
 # Create a new column named pop\_u  
 mutate(pop\_u = pop\_n \* 1000 \* prop\_u / 100) |>   
 # Drop the pop\_n and prop\_u columns  
 select(-pop\_n, -prop\_u) |>   
 # Perform a left join with the income\_groups\_df data frame  
 left\_join(income\_groups\_df) |>   
 # Drop the rows that have missing values in the san\_bas\_u & income\_group cols  
 drop\_na(san\_bas\_u, income\_group)

Joining with `by = join\_by(iso3)`

# Basic Sanitation Uganda

# Create a vector of color codes  
color\_scale\_sanitation <- c("#8cce8f", "#fff381", "#ffda5a", "#ffbc02")  
  
# Create a vector of sanitation indicators  
fct\_sanitation <- c("basic", "limited", "unimproved", "open defecation")  
  
# Perform data manipulation operations on the jmp\_wld\_sanitation data frame  
jmp\_uga\_sanitation <- jmp\_wld\_sanitation |>   
 # Filter the rows where the iso3 column is equal to "UGA" and the year column   
 # is equal to 2000 or 2020  
 filter(iso3 == "UGA") |>   
 filter(year == 2000 | year == 2020) |>   
 # Select the name, iso3, year, and columns from san\_bas\_n to san\_od\_n  
 select(name, iso3, year, san\_bas\_n:san\_od\_n) |>   
 # Reshape the data frame from wide to long format  
 pivot\_longer(cols = san\_bas\_n:san\_od\_n,   
 names\_to = "indicator",   
 values\_to = "percent") |>   
 # Rename the indicator column based on the values of the san\_bas\_n to san\_od\_n  
 # columns  
 mutate(indicator = case\_when(  
 indicator == "san\_bas\_n" ~ "basic",  
 indicator == "san\_lim\_n" ~ "limited",  
 indicator == "san\_unimp\_n" ~ "unimproved",  
 indicator == "san\_od\_n" ~ "open defecation"  
 )) |>   
 # Convert the indicator column to a factor with levels specified by the   
 # fct\_sanitation vector  
 mutate(indicator = factor(indicator, level = fct\_sanitation))

# Income

[Figure 1](#fig-san-bas-u-income) is a box- and jitterplot of countries with percentages of populations with access to basic sanitation in 2020 grouped by income classifications.

ggplot(data = jmp\_wld\_sanitation\_gdp\_income,  
 mapping = aes(x = income\_group,   
 y = san\_bas\_u,   
 color = income\_group)) +  
 geom\_boxplot(outlier.shape = NA) +  
 geom\_jitter(width = 0.1, size = 4, alpha = 0.5) +  
 labs(x = NULL, y = "percent") +  
 scale\_color\_colorblind() +  
 theme\_minimal(base\_size = 16) +  
 theme(legend.position = "none")

|  |
| --- |
| Figure 1: Access to sanitation (urban) in 2020 by income classifications. |

# Regions

**?@tbl-reg-sanitation** shows urban sanitation indicators for global regions in 2020.

jmp\_reg\_sanitation |>   
 filter(year == max(year)) |>   
 filter(!str\_detect(region, "income")) |>   
 select(region, san\_bas\_u:san\_od\_u) |>   
 drop\_na() |>   
 gt(rowname\_col = "region") |>   
 cols\_label(  
 san\_bas\_u = md("\*\*basic\*\*"),  
 san\_lim\_u = md("\*\*limited\*\*"),  
 san\_unimp\_u = md("\*\*unimproved\*\*"),  
 san\_od\_u = md("\*\*open defecation\*\*")  
 ) |>   
 fmt\_percent(columns = san\_bas\_u:san\_od\_u,  
 decimals = 0,   
 scale\_values = FALSE)

|  | **basic** | **limited** | **unimproved** | **open defecation** |
| --- | --- | --- | --- | --- |
| Central and Southern Asia | 79% | 17% | 3% | 1% |
| Eastern and South-Eastern Asia | 95% | 3% | 2% | 1% |
| Europe and Northern America | 99% | 1% | 1% | 0% |
| Latin America and the Caribbean | 93% | 4% | 3% | 0% |
| Northern Africa and Western Asia | 95% | 2% | 2% | 0% |
| Oceania | 71% | 9% | 17% | 3% |
| Sub-Saharan Africa | 46% | 32% | 17% | 5% |
| Fragile or Extremely Fragile | 62% | 22% | 13% | 3% |
| Least Developed Countries | 48% | 29% | 20% | 4% |
| Landlocked Developing Countries | 62% | 22% | 14% | 2% |
| Small Island Developing States | 83% | 10% | 5% | 2% |
| World | 88% | 8% | 3% | 1% |

# Uganda

[Figure 2](#fig-san-uga) shows the sanitation ladder for Uganda.

ggplot(data = jmp\_uga\_sanitation,   
 mapping = aes(x = year,   
 y = percent,   
 fill = indicator)) +  
 geom\_area() +  
 labs(title = "Uganda: sanitation ladder (national)",  
 x = NULL, y = "percent", fill = "indicators") +  
 scale\_fill\_manual(values = color\_scale\_sanitation) +  
 scale\_x\_continuous(breaks = c(2000, 2020)) +  
 theme\_minimal(base\_size = 16) +  
 theme(panel.grid.minor = element\_blank())

|  |
| --- |
| Figure 2: Sanitation indicators for Uganda on a national level. |

# They told you so

Dickinson, Nicolas. 2021. “Jmpwashdata: WHO/UNICEF Joint Monitoring Programme Water and Sanitation Data.”