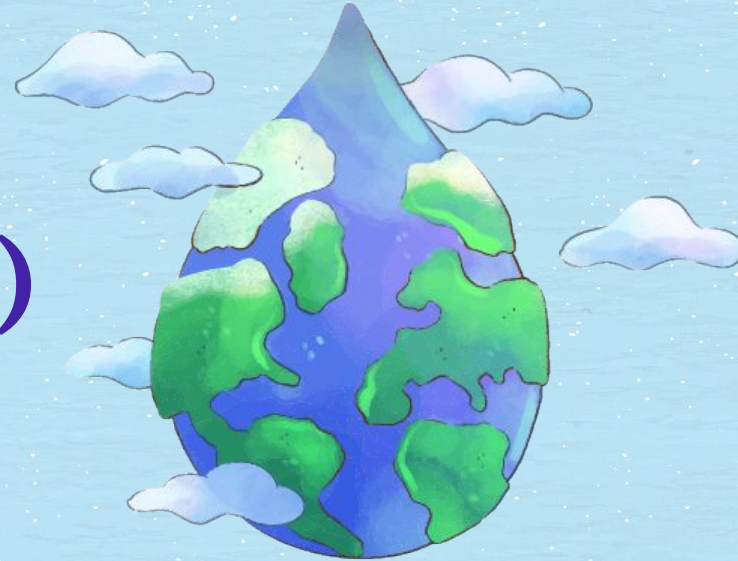


Drinking Water, Sanitation, and Hygiene (WASH)

Team Shabu Shabu

Tomoki Kyotani, Jessica Tanumihardja,
Xinan Wang, Jiayue Han



Outline

1

Outline

2

Introduction

3

Dataset

4

Method

5

Live Demo

6

Conclusion



Source: PROGRESS ON HOUSEHOLD DRINKING WATER, SANITATION AND HYGIENE, WHO/UNICEF JPM (2021)

Introduction

- Problem
 - Millions of people worldwide still lack access to reliable water resources.
- Approach
 - This project proposes a comprehensive analysis of world household access to safe drinking water, sanitation, and hygiene.
- Outcome
 - The findings from the comprehensive analysis will inform evidence-based recommendations for promoting access to safe drinking water, sanitation, and hygiene, with the aim of achieving sustainable development goals and helping more people get access to these necessities.





Dataset – Drinking Water

- **Country Information:** Country Name, Year, Population, % Urban
- **Water Supply Service Level for National / Rural / Urban:**
At Least Basic / Limited / Unimproved / Surface Water
- **Water Supply Quality for National / Rural / Urban:**
Accessible on premises / Available when needed / Free from contamination
Piped / Non - Piped
- **Region Information:**
SDG region / WHO region / UNICEF Programming region / UNICEF Reporting region



Dataset – Sanitation + Hygiene

- Country Information
- **Sanitation Service Level** for National / Rural / Urban
- **Sanitation Quality** for National / Rural / Urban
- **Hygiene Service Level** for National / Rural / Urban
- Region Information

Method – Data Cleaning + Map

The original dataset and its data cleaning steps:

- Column names are nested and contain spaces.
 - ⇒ Manually renamed columns
- Contained character values in numeric columns
 - ⇒ Replaced “-” with NA, “<1” with 0.5, and “>99” with 99.5
- Wide format
 - ⇒ Convert to two different long formats for plotting
 - Line, Donut, Bar, and Lollipop plots
 - World map

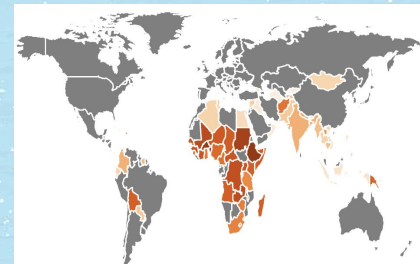
Method – Data Cleaning + Map

For Line, Donut, Bar, and Lollipop plots:

- Converted wide to long using `pivot_longer()`
 - Each row shows a percentage of its population in the respective ServiceLevel at Country - Year - Region - ServiceLevel granularity

For a world map:

- Converted wide to long
 - Each row shows percentages of all ServiceLevels at Country - Year - Region granularity
- Add longitude and latitude data to draw an interactive map using ***ggiraph::geom_polygon_interactive*** function
 - `maps::map_data("world")`
 - After joining with the map data, one file size is almost 2 GB.
- Filtered data by Year 2018 - 2020 for demo purpose



1		DRINKING WATER																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																															
---	--	----------------	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--

Original Dataset
- Wide Format

COUNTRY	YEAR	POPULATION(THOUSANDS)	REGION	ServiceLevel	Percentage
Afghanistan	2000	20779.96	NATIONAL	AtLeastBasic	28.171415
Afghanistan	2000	20779.96	NATIONAL	Limited(> 30min)	3.660638
Afghanistan	2000	20779.96	NATIONAL	Unimproved	43.178306
Afghanistan	2000	20779.96	NATIONAL	SurfaceWater	24.989641
Afghanistan	2000	20779.96	NATIONAL	AnnualRateOfChangeInBasic	2.346000
Afghanistan	2000	20779.96	RURAL	AtLeastBasic	21.389997
Afghanistan	2000	20779.96	RURAL	Limited(> 30min)	3.441358
Afghanistan	2000	20779.96	RURAL	Unimproved	45.790160
Afghanistan	2000	20779.96	RURAL	SurfaceWater	29.378486
Afghanistan	2000	20779.96	RURAL	AnnualRateOfChangeInBasic	2.246896
Afghanistan	2000	20779.96	URBAN	AtLeastBasic	52.105725
Afghanistan	2000	20779.96	URBAN	Limited(> 30min)	4.434565
Afghanistan	2000	20779.96	URBAN	Unimproved	33.960040
Afghanistan	2000	20779.96	URBAN	SurfaceWater	9.499670
Afghanistan	2000	20779.96	URBAN	AnnualRateOfChangeInBasic	2.394714

COUNTRY, AREA OR TERRITORY	Year	Region	long	lat	At least basic	Limited (more than 30 mins)	Unimproved	Surface water	Annual rate of change in basic
Afghanistan	2018	National	74.89131	37.23164	69.60193	2.363431	17.604824	10.429814	2.346000
Afghanistan	2019	National	74.89131	37.23164	72.43422	1.912148	16.042290	9.611344	2.346000
Afghanistan	2020	National	74.89131	37.23164	75.09141	1.447542	14.560263	8.900782	2.346000
Afghanistan	2018	Rural	74.89131	37.23164	60.88440	2.826300	22.431076	13.858224	2.246896
Afghanistan	2019	Rural	74.89131	37.23164	63.58820	2.409521	21.057013	12.945268	2.246896
Afghanistan	2020	Rural	74.89131	37.23164	66.32792	1.956825	19.682949	12.032311	2.246896
Afghanistan	2018	Urban	74.89131	37.23164	99.50000	0.500000	3.500892	0.500000	2.394714
Afghanistan	2019	Urban	74.89131	37.23164	99.50000	0.500000	1.585383	0.500000	2.394714
Afghanistan	2020	Urban	74.89131	37.23164	99.50000	0.500000	0.500000	0.500000	2.394714

Method – Forecasting

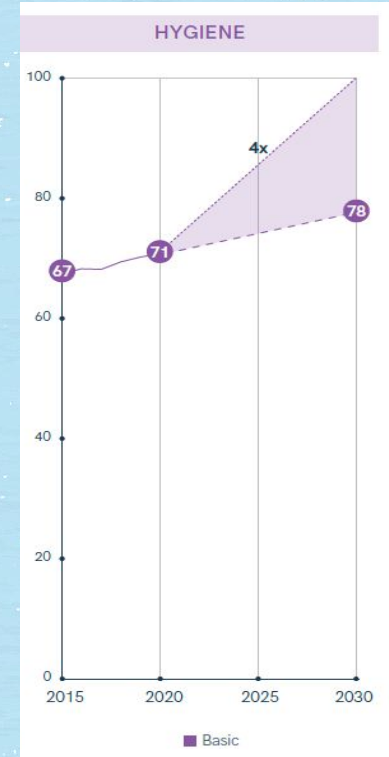
- 2030 goal:
 - 100% population have access to “at least basic” services
- Time series library:

tsbox

forecast

ggfortify

- Forecasting method:
 - Exponential smoothing (Holt)
 - Auto-ARIMA (p,d,q)




Source:

<https://washdata.org/reports/jmp-2021-wash-households>

Live Demo + Results

Water, Sanitation & Hygiene (WASH)

 Summary

 Drinking Water

 Sanitation

 Hygiene

 Forecast

 About



Discussion/Conclusion

Conclusion:

- Data visualization helps identify and compare status of access to household necessities.
- Each type of plot emphasizes different aspect of the dataset.
- Exponential smoothing forecast the 2030 values well; most are not achieving the goal

Future Works:

- Add other useful datasets (Education, Health Expenditure, etc..) and explore further
 - Focus on countries that are fairly doing well and find the reasons behind it



World Health
Organization

WHO
UNICEF



J M P





Thank you!

Questions?



Source: PROGRESS ON HOUSEHOLD DRINKING WATER, SANITATION AND HYGIENE,
WHO/UNICEF JPM (2021)

JPM forecast result

Source:
<https://washdata.org/reports/jmp-2021-wash-households>



FIGURE 1 Global coverage of WASH services, 2015-2020 (%), and acceleration required to meet targets by 2030

Resources



- <https://washdata.org/data/downloads#WLD>
- <https://washdata.org/reports/jmp-2021-wash-households>
- <https://sdgs.un.org/goals/goal6>
- <https://data.unicef.org/resources/wash-2030-agenda/>
- <https://slidesgo.com/theme/happy-world-water-day-minitheme#search-water&position-4&results-128&rs=search&rs=search>
- <https://www.capitalone.com/tech/machine-learning/understanding-arima-models/>
- <https://people.duke.edu/~rnau/411arim.htm#arima010>
- https://rstudio-pubs-static.s3.amazonaws.com/303786_f1b99d6b7e9346c4b1488a174bab839a.html
-

Tomoki – Live Demo Notes – Summary

- Map
 - Shows the “At least basic” level access
 - Filter by Classification Area and Year (2018-2020)
 - Find out which country is doing well or not relative to other countries
- Lollipop
 - Shows the top 10 countries with the highest % for the worst kind of ServiceLevel
- These plots are useful to decide countries that need immediate action to improve their environments, and you can look deeper into the data about such countries in our different tabs

Jiayue – Live Demo Notes – Line

- Line
 - Drinking Water
 - For world national, we can see more and more population get access to drinking water with at least basic service level from 2000 to 2020. Also the population get access to Unimproved drinking water is also decreasing.
 - We can see the same trending for population in rural and urban area.

Xinan – Live Demo Notes – Bar

- Bar Plot (Select a specific country: Canada, India)
 - Drinking Water
 - For world national, population that could get piped water is increasing with the growth of the world's population
 - Developed Country (Canada): all piped population; Developing Country (India): half non-piped population, half piped population
 - For rural region, the proportion of non-piped population will be higher than urban.
 - Sanitation
 - For world national, population that use sewer connections as sanitation facility takes the largest proportion. And population with no facility is decreasing with the growth of the year.
 - Developed Country (Canada): Sewer connections largest, few septic tanks; Developing Country (India): Latrines, septic tanks increasing by year, no facility takes a large proportion but decreasing by year
 - Rural: Latrines and septic tanks largest; Urban: Sewer connection largest

Xinan – Live Demo Notes – Bar

- Bar Plot
 - Hygiene
 - Total population that has no or limited facility is increasing by year
 - Compared to rural, urban population of no or limited facility is very less. However, both of rural and urban population that has no or limited facility is increasing by year.

Jessica – Live Demo Notes

- Donut
 - Choose a geographical region: "Eastern and South-Eastern Asia"
 - Snapshot of the line plot at a given year (only select the end range for Year)
 - Summary of distribution per region
 - Hygiene: Hover over to show number & zoom in if area is too small
 - Use median % so that it is less sensitive to outlier
- Forecast
 - Decomposition: trend
 - Subtitle shows values
 - ARIMA: world (0,1,0). Dataset-SDG region varies from (1,1,0) (2,1,0) to (0,2,0)
 - Dataset-Country varies: $p = 0/1/2$, $d = 1/2$, $q = 0/1$. Indo: (1,2,1)
 - Therefore ES holt is good

Notes: 15mins for each team to present their final project included Q&A session.

- Introduction - Jiayue
 - What is the problem?
 - Why is it worth further research?
 - How could you solve this problem?
- Data Set - Nancy
 - Introduce the background, and structure of your data set
- Method - Jessica-> Forecasting, Data cleaning + Map -> Tomoki
 - Select and introduce two interesting implementations from your app
 - Forecasting
 - Data cleaning + Map
- Results - Tomoki ->Summary, Jessica->Donut, Jiayue->Line, Nancy->Bar, Forecast -> Jessica
 - Run your Shiny app demo (5-8 min)
 - Present the results which approach your goal/s
 - Interpretations of each plot, how useful it is, etc..
 - Talk about results while doing the live demo
- Conclusions/Discussion - Tomoki
 - Discussion and conclusions on each dataset
 - Drinking Water, Sanitation, Hygiene