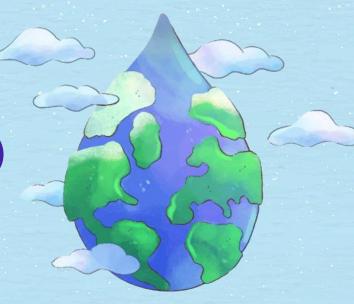
# Drinking Water, Sanitation, and Hygiene (WASH)

**Team Shabu Shabu** Tomoki Kyotani, Jessica Tanumihardja, Xinan Wang, Jiayue Han



## **Outline**

1 Outline

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 <u>Country Year Region ServiceLevel</u> granularity

#### For a world map:

- Converted wide to long
  - Each row shows percentages of all ServiceLevels at <u>Country Year Region</u> 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 NATIONAL											AC .																	
2						N	ATIONAL			RURAL			URBAN			Proportion of population using improved water supplies				Pro								
COUNTRY, AREA OR TERRITORY	▼	1503 	Year T-	Population (thousands)	% urban	At least basic	Limited (more than	Unimproved	Surface water	ast	Limited (more than	Unimproved	Surface water	Annual rate of	At least basic	Limited (more than	Unimproved	Surface water	Annual rate of	Safely managed	Accessible on	Available when	Free from amination	Annual rate of change in safely	Piped	Non-piped	Safely managed	Accessible on
4 Afghanistan			2020	38 928	26	75	1	15	9 2.35	66	2	20	12	2.25	>99	<1	<1	<1	2.39	28	48	1.	28	0.83	22	54	24	35
Albania Algeria		(A)	2020 2020	2 878 43 851	62 74	95 94	2 5	3 <1	<1 0.42 <1 0.23	94 90	2 9	4 <1	<1 <1	0.68	96 96	2	3 <1	<1 <1	0.00	71 72	82 79	71 72	96 85	1.08 0.12	81 72	16 28	69	88 71
7 American Samoa			2020	43 051	87	>99	<1	<1	<1 0.23	90	9	- 1	- 1	0.33		4	×1		0.09	98	99	12	98	1.03	>99	<1	09	
08 Andorra			2020	77	88	>99	<1	<1	<1 0.00	>99	<1	<1	<1	0.00	>99	<1	<1	<1	0.00	91	>99	>99	91	0.00	>99	<1		>99
Angola			2020	32 866	67	57	9	19	14 0.80	28	9	23	41	0.33	72	10	18	<1	0.53		39	31			42	24		7
50 Anguilla		AIA :	2020	15	100		-	-					-	-			-	-				-	- 2		-	-		
71 Antigua and Barbuda			2020	98	24		-	-				-	-	-		-	-	2				-	-		-	9		-
92 Argentina		6.070	2020	45 196	92			2		-	2	12	12	-	>99	<1	<1	<1	0.05		2	-	-	12	2	2		2
13 Armenia			2020	2 963	63	>99	<1	<1	<1 0.24	>99	<1	<1	<1	0.56	>99	<1	<1	<1	0.06	87	>99	92	87	0.35	>99	<1	•	>99
34 Aruba			2020	107	44		50.0	-			-	200	-	-					-		- 07	-	-		-	-		-
Australia Austria			2020 2020	25 500 9 006	86 59	>99	<1 <1	<1 <1	<1 0.01 <1 0.00	>99 >99	<1 <1	<1 <1	<1 <1	0.05	>99 >99	<1 <1	<1 <1	<1 <1	0.01	99	97 >99	96 99	>99	0.12	-		380	88 >99 :
97 Azerbaijan			2020	10 139	56	96	1	3	<1 1.12	91	2	7	<1	1.82	>99	<1	<1	<1	0.42	88	90	99	92	1.52	91	6	78	78
18 Bahamas			2020	393	83	-	1	-	- 1.12	- 31	-	-	-1	1.02	- 33		-	- 1	- 0.42		-	- 2	-	1.02	-	-	10	-
39 Bahrain			2020	1 702	90	>99	<1	<1	<1 0.00			-	2		-	-	2		1/20	99	99	>99	>99	0.01		2		
60 Bangladesh	В	BGD :	2020	164 689	38	98	1	<1	<1 0.15	98	<1	<1	<1	0.21	97	2	<1	<1	-0.05	59	82	96	59	0.16	15	83	62	81
Barbados	E	BRB :	2020	287	31	99	<1	1	<1 0.01				+	-		-	-		-		98	89			99	<1		-
02 Belarus	E	BLR :	2020	9 449	79	97	3	<1	<1 -0.14	99	<1	<1	<1	-0.01	96	4	<1	<1	-0.17	95	95	-	>99	0.70	96	4		94
Belgium			2020	11 590	98	>99	<1	<1	<1 0.00	>99	<1	<1	<1	0.00	>99	<1	<1	<1	0.00	>99	>99	-	>99	0.02	>99	<1		>99
44 Belize			2020	398	46	98	1	<1	<1 0.55	98	1	<1	<1	0.69	99	1	<1	<1	0.38		97	-	18		87	13	*	96
Benin .			2020	12 123	48	65	9	22	3 0.16	58	13	24	5	0.27	73	6	20	1	-0.21		25	58			37	38	*	10
86_Bermuda			2020	62	100	>99	<1	<1	<1 0.00		-		1.	- 000	>99	<1	<1	<1	0.00	-	>99 88	87	37	0.40	>99	<1	-	-
07 Bhutan 28 Bolivia (Plurinational State of)			2020 2020	772 11 673	42 70	97 93	2 <1	<1 2	<1 0.70 5 0.67	97 80	3 <1	<1	<1 16	0.88 1.24	98 >99	1 <1	<1 <1	<1 <1	0.12 0.19	37	88 86	87	3/	0.40	>99 69	<1 25	28	83 62
49 Bosnia and Herzegovina			2020	3 281	49	96	4	<1	<1 -0.06	97	3	4 <1	<1	0.05	95	5	<1	<1	-0.19	89	91	89	90	0.08	09	25		92
70 Botswana			2020	2 352	71	92	5	2	1 0.85	79	15	2	4	1.27	98	<1	2	<1	0.17		82	-		0.00	91	6		50
91 Brazil			2020	212 559	87	>99	<1	<1	- 0.29	96	<1	3	7.	1.13	>99	<1	<1	<1	0.08	86	>99	91	86	0.49	98	1	72	96
12 British Virgin Islands			2020	30	49	>99	<1	<1	<1 0.23		10.61		-	-				1000	-		98	-	-	-	-	-		
33 Brunei Darussalam			2020	437	78	>99	<1	<1	<1 0.00				-		>99	<1	<1	<1	0.00			-	-		>99	<1		
54 Bulgaria	В	BGR :	2020	6 948	76	>99	<1	<1	<1 -0.04	97	<1	3	<1	-0.12	>99	<1	<1	<1	-0.02	98	98	98	99	0.40	>99	<1		96
75 Burkina Faso			2020	20 903	31	47	31	21	<1 -0.54	33	39	28	<1	-1.04	80	15	5	<1	0.05		20	56	-	250	28	50		3
95 Burundi			2020	11 891	14	62	19	15	4 0.58	58	21	17	4	0.49	91	8	1	<1	0.43		11	-	- 5	-	38	43		3
17 Cabo Verde			2020	556	67	89	8	3	<1 0.48	80	10	10	<1	0.50	93	7	<1	<1	0.31		88	-	-		90	7	0	80
38 Cambodia			2020	16 719	24	71												<1	0.62	28	48	40	28	0.54	32	53	18	42
59 Cameroon 80 Canada			2020 2020	26 546 37 742	58 82	66 >99			•	•						1		<1	-0.02 -0.01	>99	35 >99	49	>99	0.04	36 >99	43		8 99
01 Caribbean Netherlands			2020	26	75	>99		(	Orig	un	<b>2</b> 1	1 ):	2Ť	၁င	$\Delta$ 1	<b>T</b>		`.'	-0.01	-33	- 55	- 1	- 55	0.04	-99	- 1	-	30
22 Cayman Islands			2020	66	100	100		-	ノロし	)	aı		at	uo	J	L			100	-		ē	2		2	i.		
43 Central African Republic			2020	4 830	42	37			_									<1	-1.68	6	6	38	29	-0.12	14	49	2	2
64 Chad			2020	16 426	24	46			- \	A /:			•			1		<1	-0.04	6	9	48	13	0.00	20	41	2	2
Channel Islands	(	CHI :	2020	174	31				_ \	/\/ L	മ	· <b>-</b>	$\cap$	m	<b>1</b>	<b>T</b>			1200		2	-	-		-	200		-
06 Chile			2020	19 116	88	>99			_ ,	/ V I'	uC	•	U	111	a	L		<1	0.04	99	>99	>99	99	0.33	>99	<1	*	98
China China			2020	1 463 141	62	94												<1	-0.07		93	91	-	-	80	15		89
48 China, Hong Kong SAR			2020	7 497	100	>99	7.	`!	0.00				-		- 33	7.	` .	<1	0.06	>99	>99	-	>99	0.09	>99	<1		-
69 China, Macao SAR			2020	649	100	>99	<1	<1	<1 0.00	-		:	5		>99	<1	<1	<1	0.00	>99	>99	>99	>99	0.00	>99	<1		-
Colombia			2020 2020	50 883 870	81	97	<1	<1	1 0.30	87	<1 -	5	7	0.70	>99	<1	<1	<1	0.10	73	96	76	82	0.21	89	9	40	83
Comoros  Introduction Water		Sanitation		Hygiene	29 Me	enstrua	l health	+						•									-	3.00				•

COUNTRY <sup>‡</sup>	YEAR <sup>‡</sup>	POPULATION(THOUSANDS)	REGION <sup>‡</sup>	ServiceLevel	Percentage <sup>‡</sup>						
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	Annual Rate Of Change In Basic	2.246896						
Afghanistan	2000	20779.96	URBAN	AtLeastBasic	52.105725						
Afghanistan	F	For Line, Donut, Bar, and Lollipop plots - Long Format									
Afghanistan											
Afghanistan	2000	20773.30	onson	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	or W	orld M	lap	3.500892	0.500000	2.394714
Afghanistan	2019	Urban	74	- Lor	ig For	mat	1.585383	0.500000	2.394714
Afghanistan	2020	Urban	74.89131	37.23164	99.50000	0.500000	0.500000	0.500000	2.394714

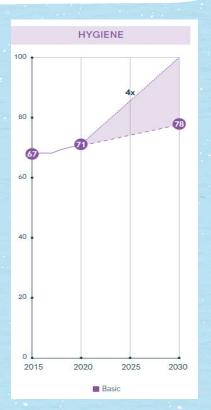
## **Method - Forecasting**

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



ggfortify

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



Source:

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

# Live Demo + Results















#### 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

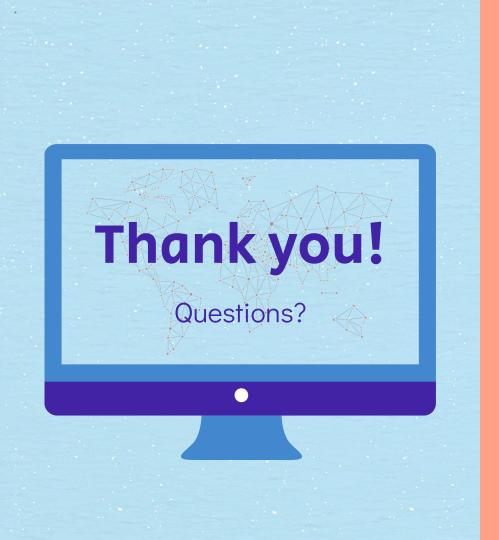
#### **Future Works:**

- 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











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





### 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/
- <a href="https://slidesgo.com/theme/happy-world-water-day-minitheme#search-water&position-4&results-128&rs=search&rs=search">https://slidesgo.com/theme/happy-world-water-day-minitheme#search-water&position-4&results-128&rs=search&rs=search</a>
- 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\_f1b99d6b7e9346c4b1488a174bab83
   9a.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)
- $\circ$  Dataset-Country varies: p =0/1/2, d = 1/2, q = 0/1. Indo: (1,2,1)
- o 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?
  - O 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
  - o Discussion and conclusions on each dataset
    - Drinking Water, Sanitation, Hygiene