openjmp

the data and code the behind the JMP WASH estimates

Linda Karani karanilinda@gmail.com

Lars Schöbitz lschoebitz@ethz.ch

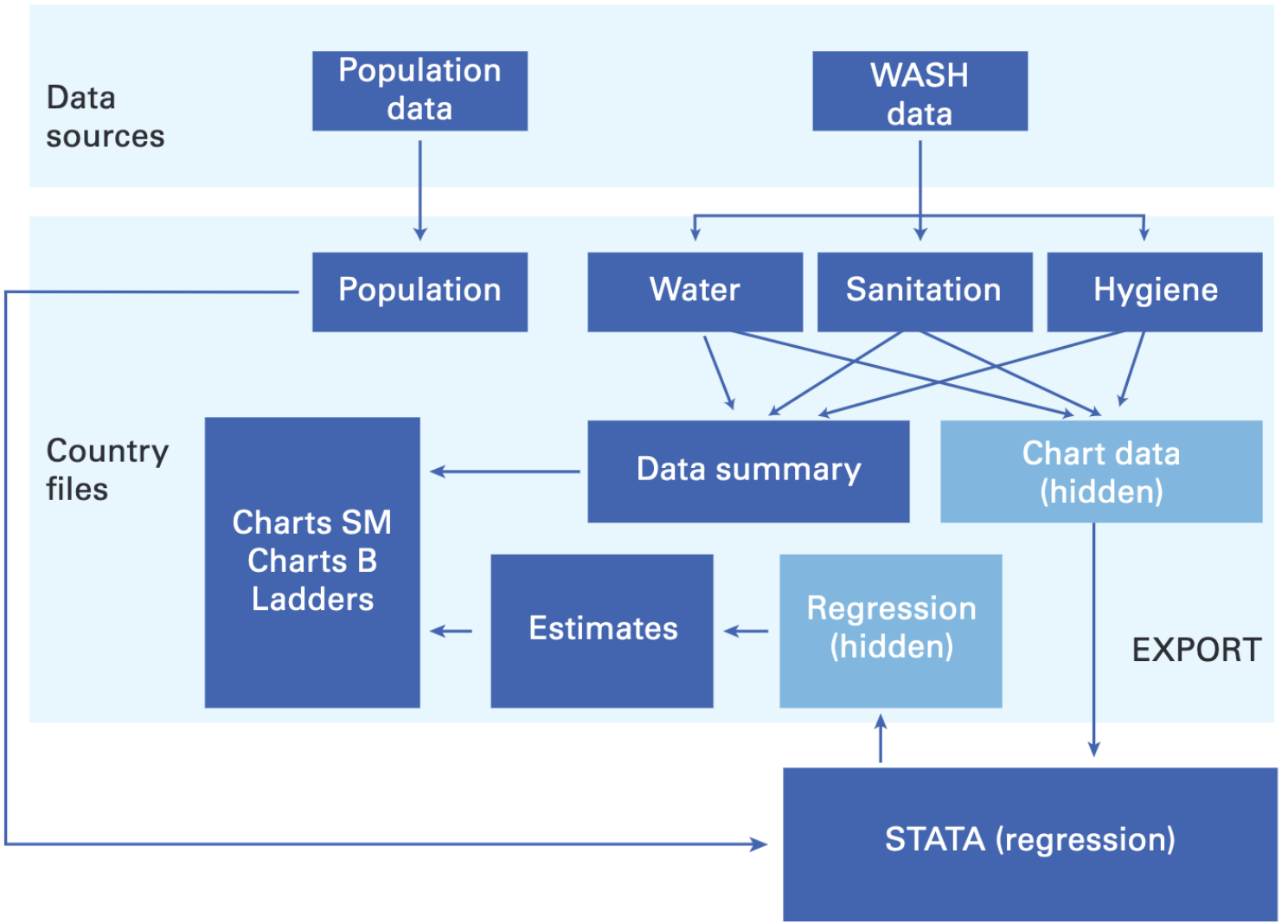
March 10, 2023

openjmp - why?

WHO/UNICEF Joint Monitoring Programme (JMP)

- **JMP mandate:** internationally-comparable information on WASH since 1990
- **JMP data input:** raw database is updated every 2 years
- **JMP methods:** linear regression model with Stata 14.0
- **JMP data output:** 26 indicators for 232 countries, areas, and globally
- **JMP country files:** compile raw data input and data output for 26 indicators in spreadhsheet-based proprietary software

Current JMP workflow



Goals of openjmp project

- Document and publish R data package: `jmpinput`
- Document and publish R software package: `jmpmodel`
- Host half day online workshops to teach usage of developed packages in R
- Publish lessons as Open Educational Resources

```
jmpinput
```

`jmpinput` R data package - benefits

- Data accessible as a single table for any data analysis tool
- Data can be imported to R using one command
- Public website with detailed documentation _ e.g. `washdata` R Package https://katilingban.io/washdata/index.html

`jmpinput` R data package - benefits

- Data accessible as a single table for any data analysis tool
- Data can be imported to R using one command
- Public website with detailed documentation _ e.g. `washdata` R Package https://katilingban.io/washdata/index.html

washdata: Urban Water and Sanitation Survey Dataset

cran v0.1.3 | license CC0 | downloads 177/month | downloads 21K | build passing

build passing

This package contains four datasets from an urban water and sanitation survey in Dhaka, Bangladesh conducted by Water and Sanitation for the Urban Poor with technical support from Valid International in March 2017.

- `popBGD` : Dataset on estimated population of each primary sampling unit (PSU) that were surveyed. This dataset is a mix of data from WorldPop for the non-slum areas and from the 2014 Bangladesh Census of Slum Areas and Floating Population.

`jmpinput` - sanitation

- Data in long format (19,528 rows)
- 9 variables

iso3	source	type	year	var_short	var_long	residence	san_service_chain	value
AFG	MICS	Survey	2003	s_imp_u	Improved	urban	user interface	44.2
AFG	NRVS	Survey with microdata	2005	s_imp_u	Improved	urban	user interface	62.3
AFG	NVRA	Survey with microdata	2008	s_imp_u	Improved	urban	user interface	58.3
AFG	MICS	Survey with microdata	2011	s_imp_u	Improved	urban	user interface	70.9

`jmpinput` - new variables

- `residence`: urban/rural/national
- `san_service_chain`: sanitation service chain

san_service_chain	n
open defecation	2770
sharing	1553
user interface	12638
containment	195
emptying	1356
transport	10
FS treatment	85
WW treatment	921

`jmpinput` - use cases

1. Using JMP methods to reproduce estimates and apply different models - Linda Karani - MSc Data Science
2. Writing the `jmpmodel` R software package with a function to produce estimates (and a function to produce service ladder plots)

`jmpinput` - use cases

1. Using JMP methods to reproduce estimates and apply different models - Linda Karani - MSc Data Science
2. Writing the `jmpmodel` R software package with a function to produce estimates (and a function to produce service ladder plots)

```
estimate(iso3 = "AFG",              # default: all iso3 codes
         year = 2010:2030,          # Single year or range of years
         var_short = NULL,          # default: all variables (NULL)
         residence = "national") # default: national
```

`jmpinput` - use cases

1. Using JMP methods to reproduce estimates and apply different models - Linda Karani - MSc Data Science
2. Writing the `jmpmodel` R software package with a function to produce estimates (and a function to produce service ladder plots)

```
estimate(iso3 = "AFG",              # default: all iso3 codes
         year = 2010:2030,          # Single year or range of years
         var_short = NULL,          # default: all variables (NULL)
         residence = "national") # default: national
```

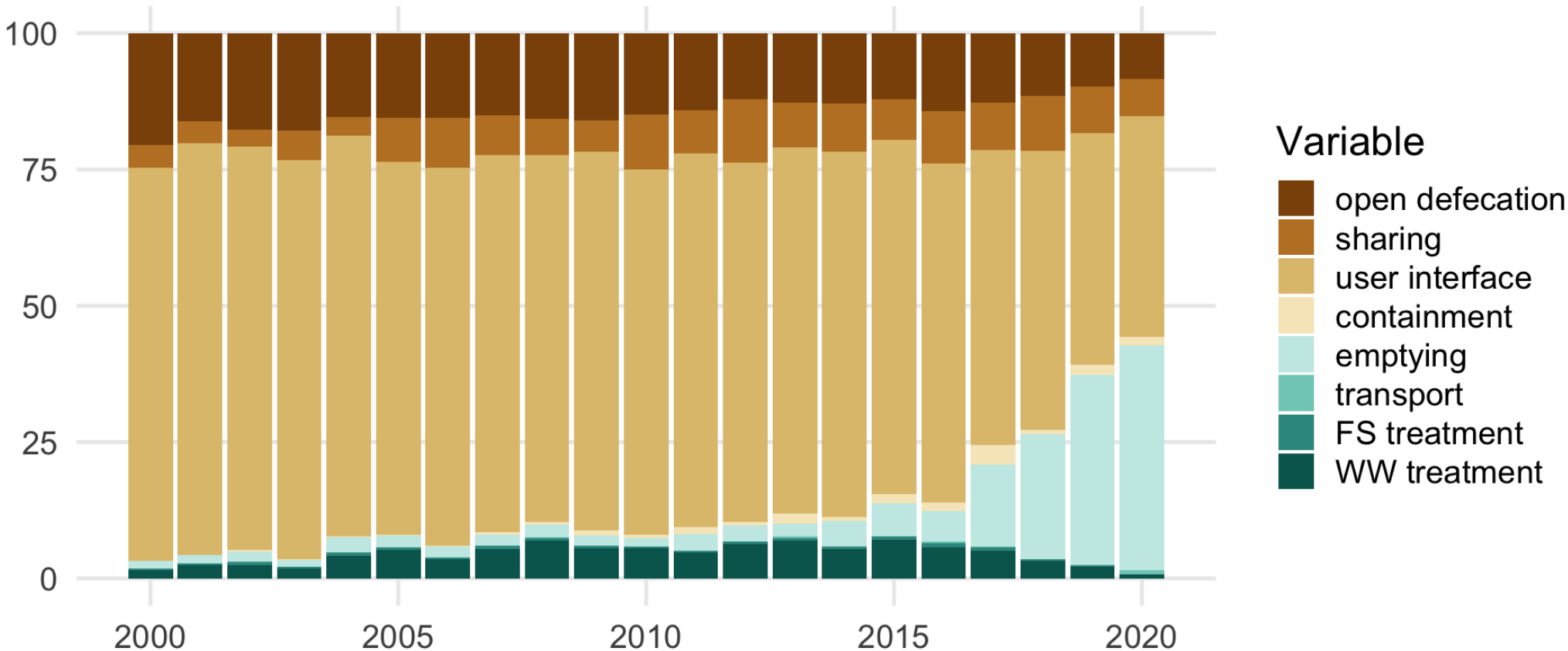
3. Great potential for unforeseen use cases enabled by making the data readily accessible (research, teaching, joining with other data, etc.)

Number of data points for type of survey

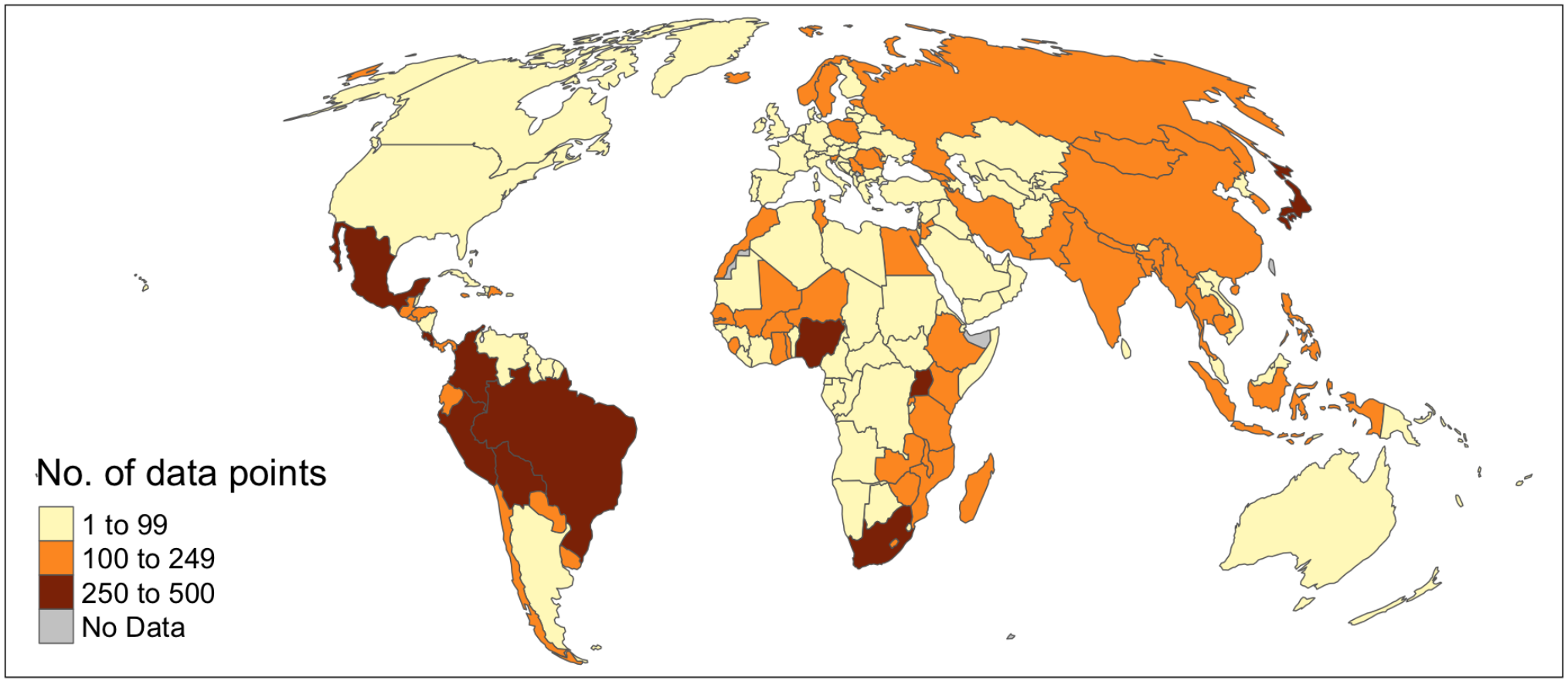
type	n
Survey with microdata	11149
Admin	3369
Survey	3124
Census	1732
Other	154

Proportion of data points along sanitation service chain

For all countries since 2000

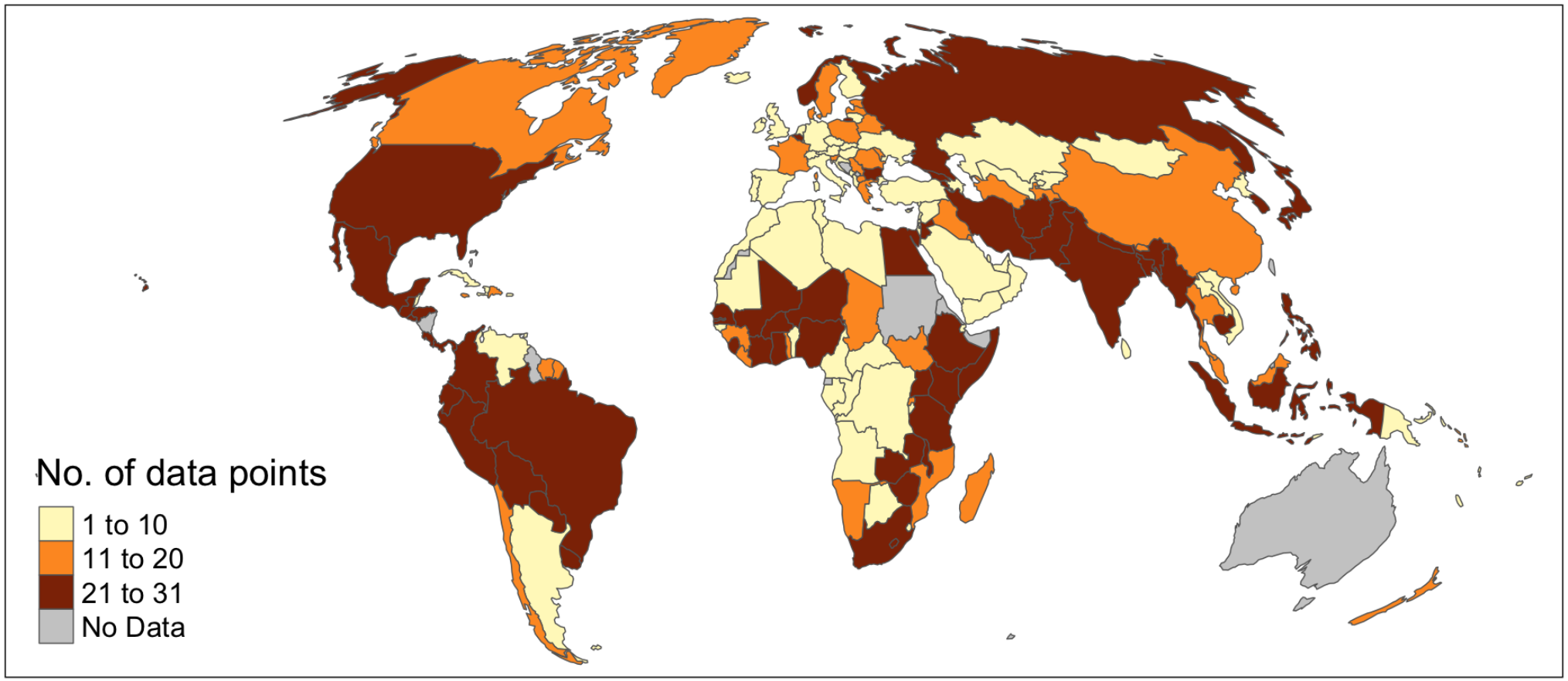


JMP raw data collection - Total number of data points since 2000



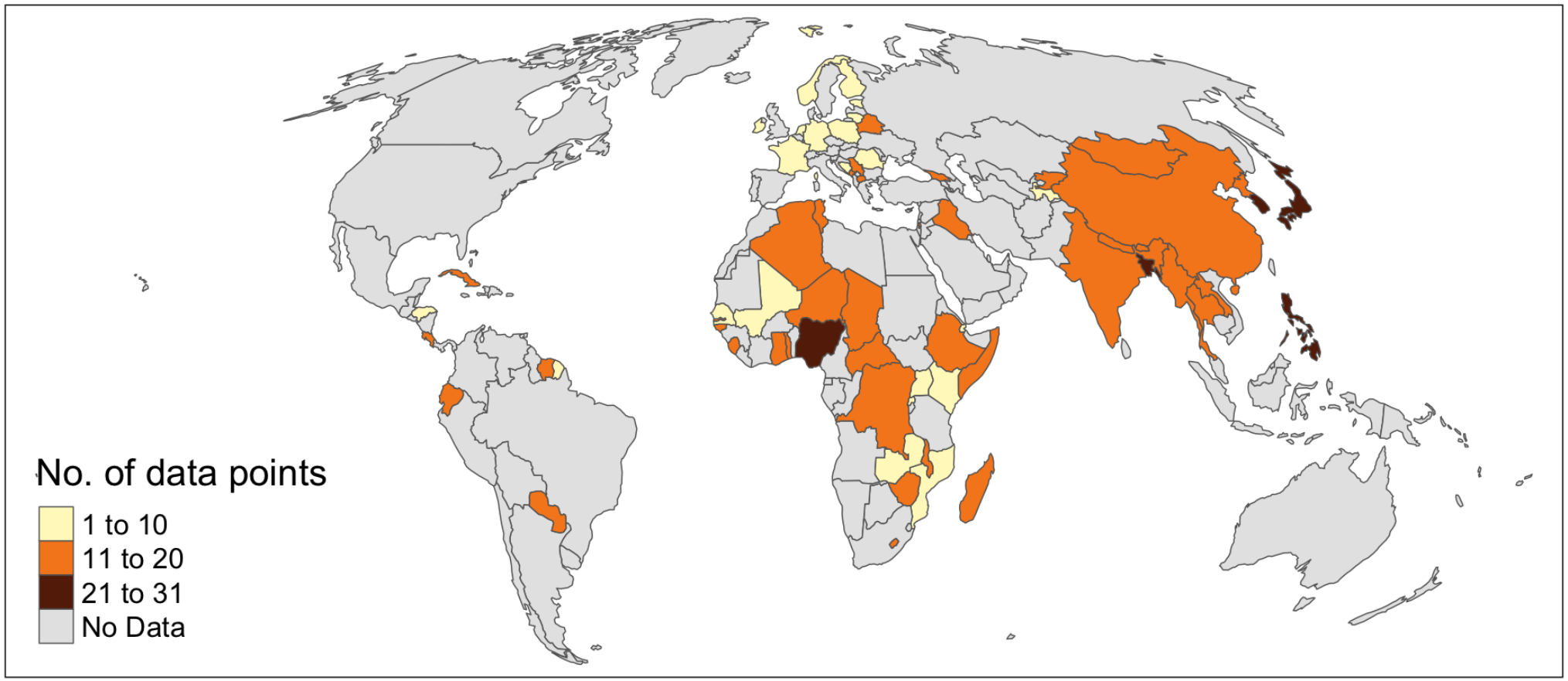
country	n
Peru	412
Mexico	392
Colombia	364
Nigeria	332
Brazil	304
Costa Rica	302
South Africa	301
Japan	278
Uganda	265
Bolivia	256

JMP raw data collection - Number of data points for 'user interface' since 2015



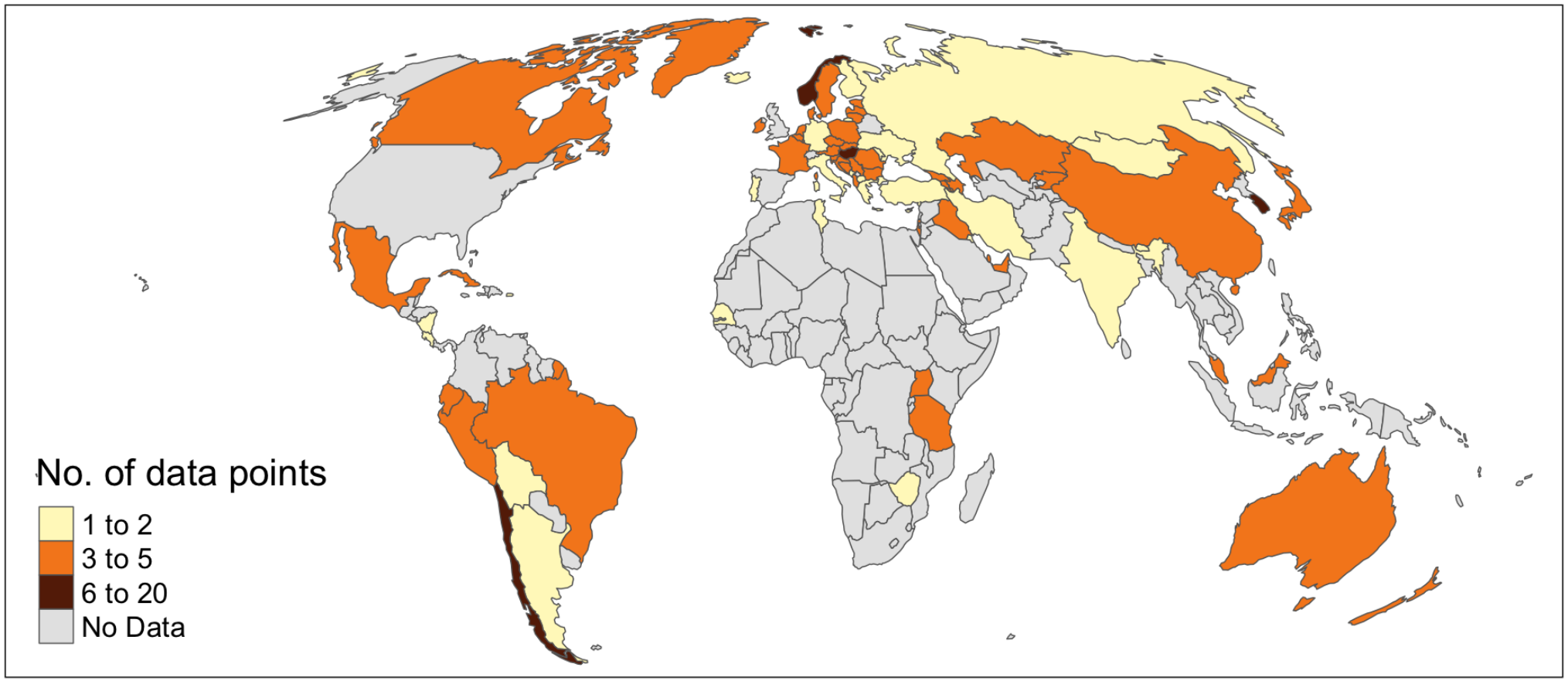
country	n
Peru	101
Colombia	96
Nigeria	80
Mexico	79
Ghana	62
Uganda	62
Costa Rica	56
Guatemala	56
South Africa	52
Bolivia	48

JMP raw data collection - Number of data points for 'emptying' since 2015



country	n
Philippines	62
Nigeria	48
Bangladesh	40
Japan	40
South Korea	32
Ethiopia	20
Niger	20
Belarus	16
China	16
Congo - Kinshasa	16

JMP raw data collection - Number of data points for 'wastewater treatment' since 2015



country	n
Chile	14
Hong Kong SAR China	10
South Korea	8
Macao SAR China	8
Mauritius	8
Norway	8
Hungary	6
Armenia	5
Belgium	5
Brazil	5

JMP raw data collection - Number of data points for 'faecal sludge treatment' since 2015

country	n
South Korea	8
Lithuania	5
Norway	5
Japan	3
Poland	3
Bhutan	2
Austria	1
Finland	1
Iceland	1

openjmp - what's next?

openjmp - what's next

- start of project: ~ July/August 2023
- timeline and activities for `jmpinput` & `jmpmodel`: Work Packages & Activities Google Sheet
- stay up to date on progress and add requests: openwashdata/jmpinput/issues
- join openwasdata community: openwashdata.org

Thanks! 

Thanks 🌻

This project was supported by the Open Research Data Program of the ETH Board.

The slides were created via revealjs and Quarto: https://quarto.org/docs/presentations/revealjs/

You can view source code of slides on GitHub

Or you can download slides in PDF format

References

"Urban Water and Sanitation Survey Dataset." n.d. https://katilingban.io/washdata/index.html.

WHO/UNICEF Joint Programme for Water Supply, Sanitsation and Hygiene (JMP). 2018. "JMP Methodology - 2017 Update & SDG Baselines," March. https://doi.org/https://washdata.org/report/jmp-methodology-2017-update.