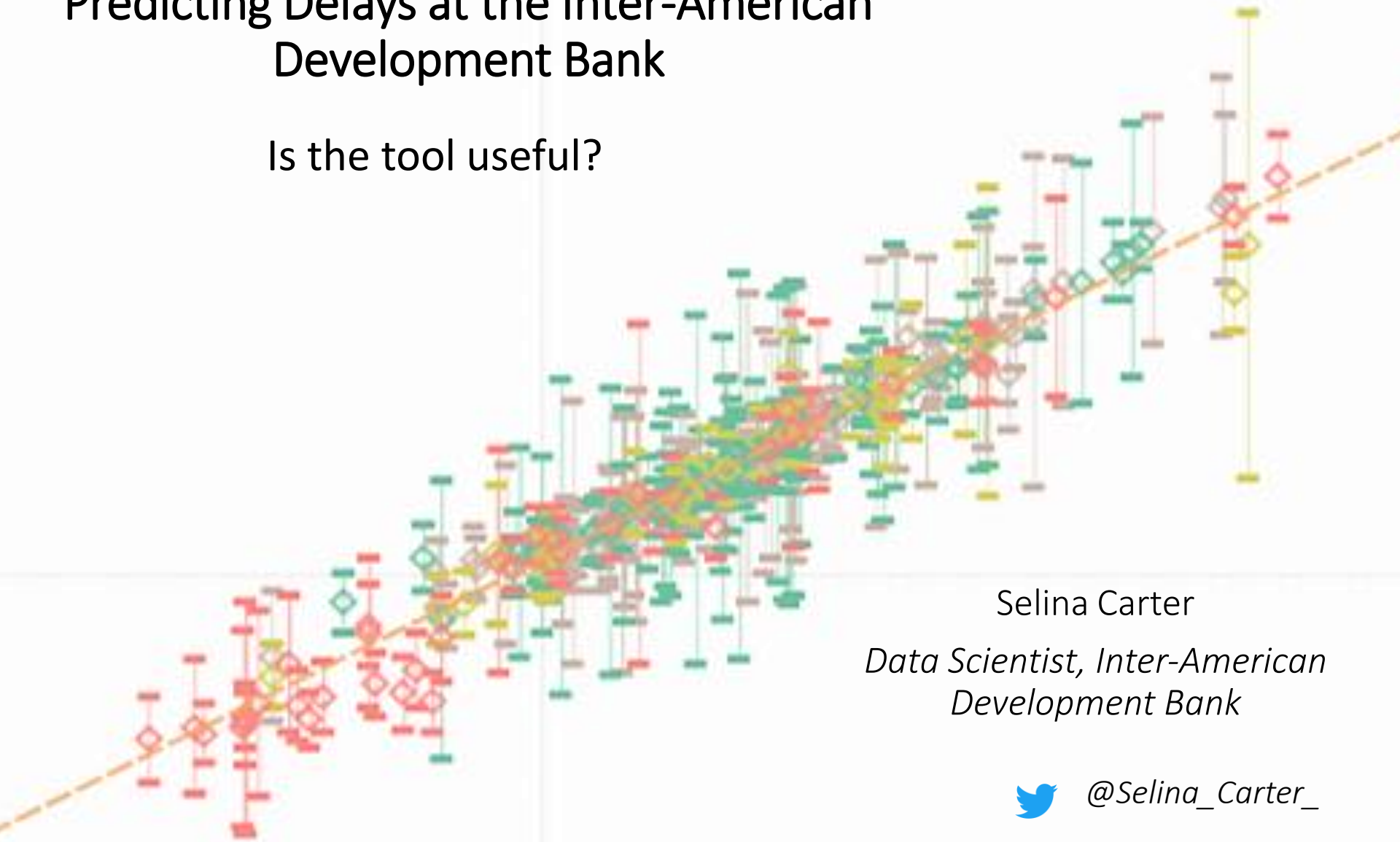
Predicting Delays at the Inter-American Development Bank

Is the tool useful?



Selina Carter

Data Scientist, Inter-American Development Bank

*@Selina_Carter_*

This presentation was prepared by the author in her personal capacity. The opinions expressed are the author's own and do not reflect the view of the Inter-American Development Bank (IDB).

R

R

R

R

Team

Selina Carter (IDB)

Marcelo Carrillo (IDB)

Alexis Estevez (IDB)

Pablo Riba (IDB)

Dr. Jonathan Hersh
(Chapman University)

Overview

I. What is the model?

- Business need
- Data
- Challenges
- Results

II. Do managers find the tool useful?

What's the **business need**?

What is the Inter-American Development Bank?

- Leading source of development financing for Latin America and the Caribbean
- Offers loans, grants, and technical assistance to governments
- Aims to reduce poverty and inequality
- Main instrument: sovereign guaranteed (SG) investment loans

SG investment loans	
Average loan size	$67 million
Loan approvals	≈ 90 new loans per year
Loans in execution	≈ 500 loans at any time point

IDB

Example Project: Salto Grande Hydroelectric Dam Modernization

RG-L1124 : Modernization of the Salto Grande Binational Hydropower Complex

Project Status: Implementation

The overall objective is to help ensure the availability of the Salto Grande Hydropower Complex(SGHC), enhancing the reliability and efficiency of the interconnection between Argentina and Uruguay. The specific objective is to assist in extending the useful life of the SGHC by modernizing its infrastructure and equipment

PROJECT DETAIL

PROJECT NUMBER	RG-L1124
APPROVAL DATE	November 28, 2018
PROJECT COUNTRY	Regional
PROJECT SECTOR	ENERGY
PROJECT SUBSECTOR	ENERGY INTEGRATION
PROJECT TYPE	Loan Operation
ENVIRONMENTAL AND SOCIAL IMPACT CATEGORY	B
PROJECT STATUS	Implementation
OPERATION NUMBER	4694/OC-RG
OPERATION NUMBER	4695/OC-RG

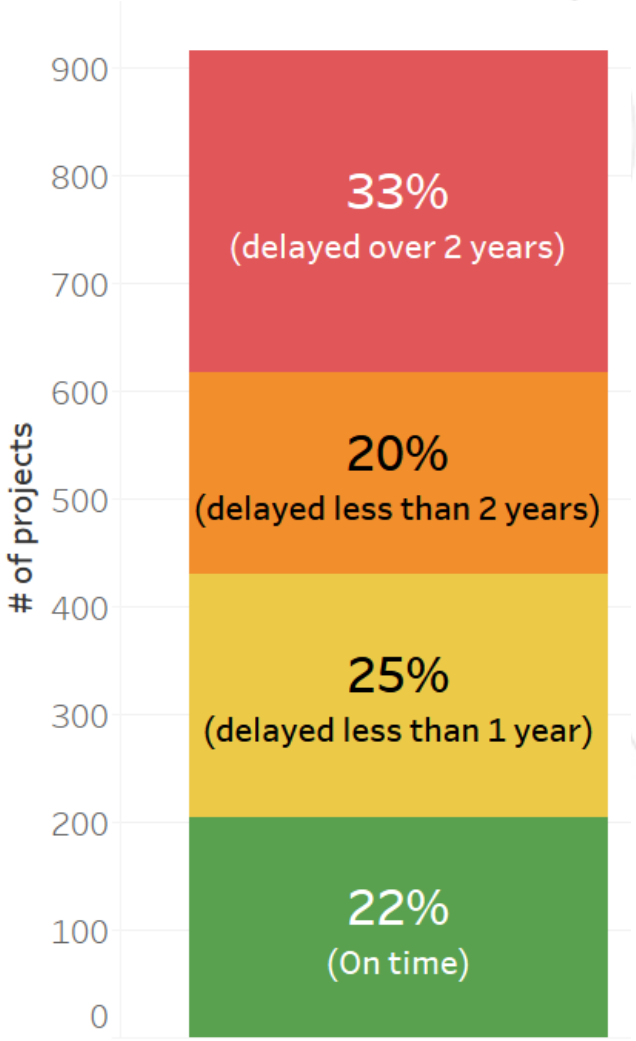
PROJECT INFORMATION

TOTAL COST	USD 80,000,000
COUNTRY COUNTERPART FINANCING	USD 0
AMOUNT	USD 80,000,000

≈ 5 years to disburse 100% of funds

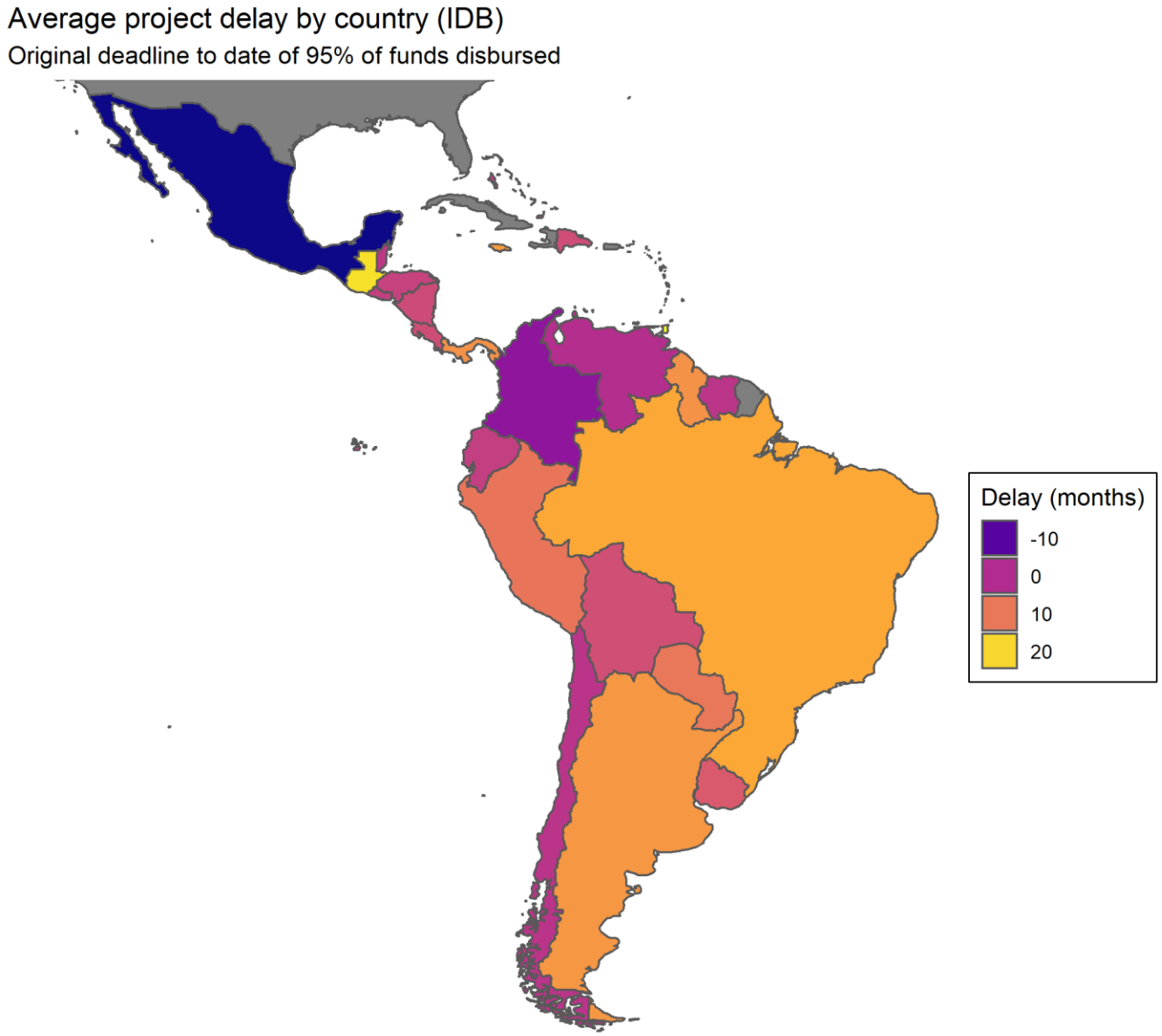
What's the **business need**?

- **"Delay" = IDB doesn't disburse 100% funds on time**
- **What percent of SG loans are delayed?**
  - **78%** (of SG loans since 2000)
  - Average delay = 14 months
  - **33%** had extension > 24 months.
- **What does this cost the Bank?**
  - IDB spent **$50 million** on the supervision of extended loans between 2010 and 2017.[1]



1: Quarterly Business Review (QBR-1 2018), SPD/SMO. Page 7.
2: Advisory Report September 2019, Office of the Executive Auditor, 2019

What's the **business need**?



What's the **business need**?

Objectives:

Predict delays

➡ Help managers develop a *proactive* approach to avoid delays.

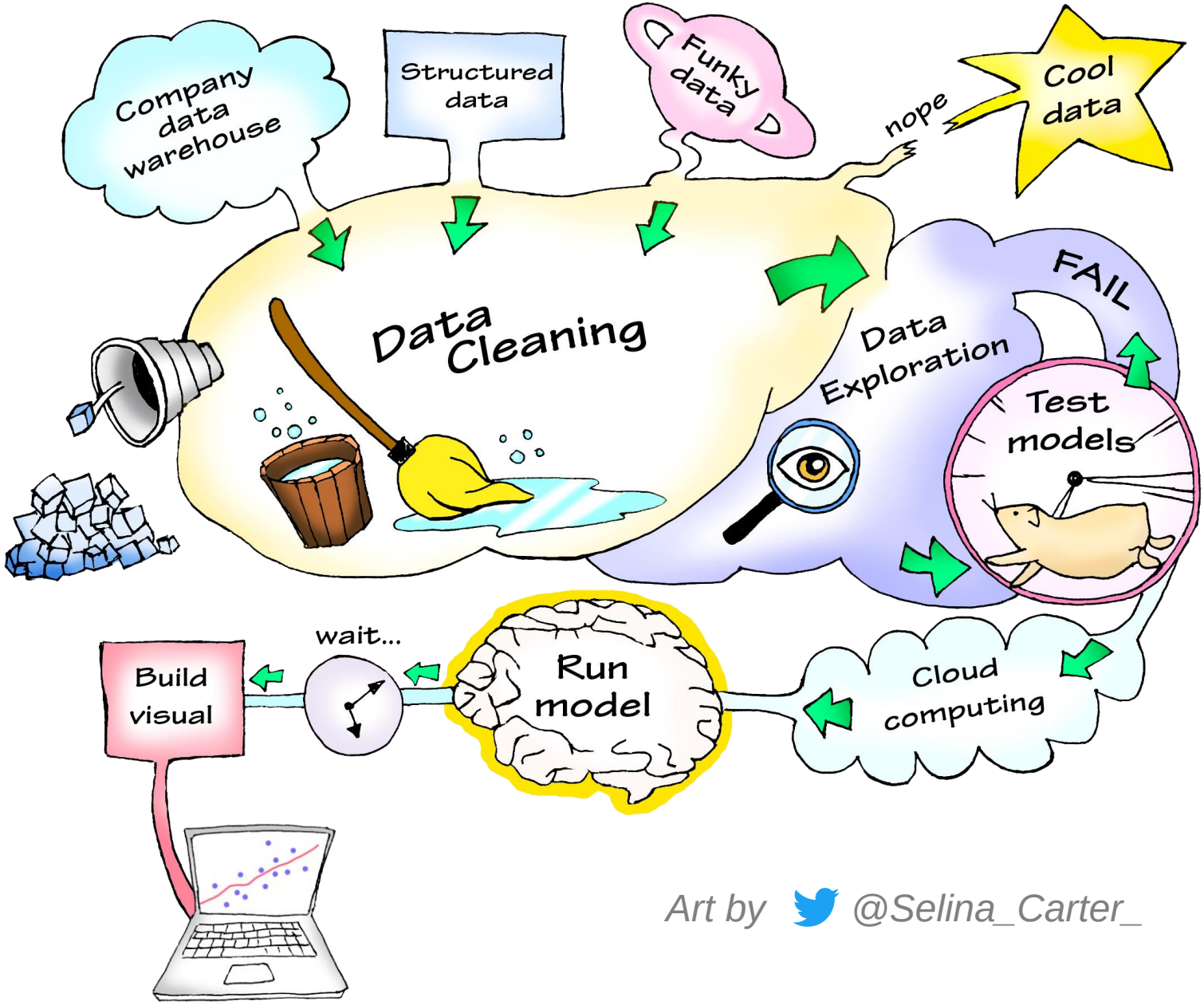
Real-time knowledge

➡ Know risk of delays *ex ante* (as soon as a project is approved)

➡ Updated monthly

Clear visualization

➡ Dashboard displays results.


Company data warehouse
Structured data
Funky data
nope
Cool data
Data Cleaning
Data Exploration
FAIL
Test models
Build visual
wait...
Run model
Cloud computing
Art by @Selina_Carter_

What's the **data**?

100+ features

(Fixed over time)

(Dynamic)

Fixed variables

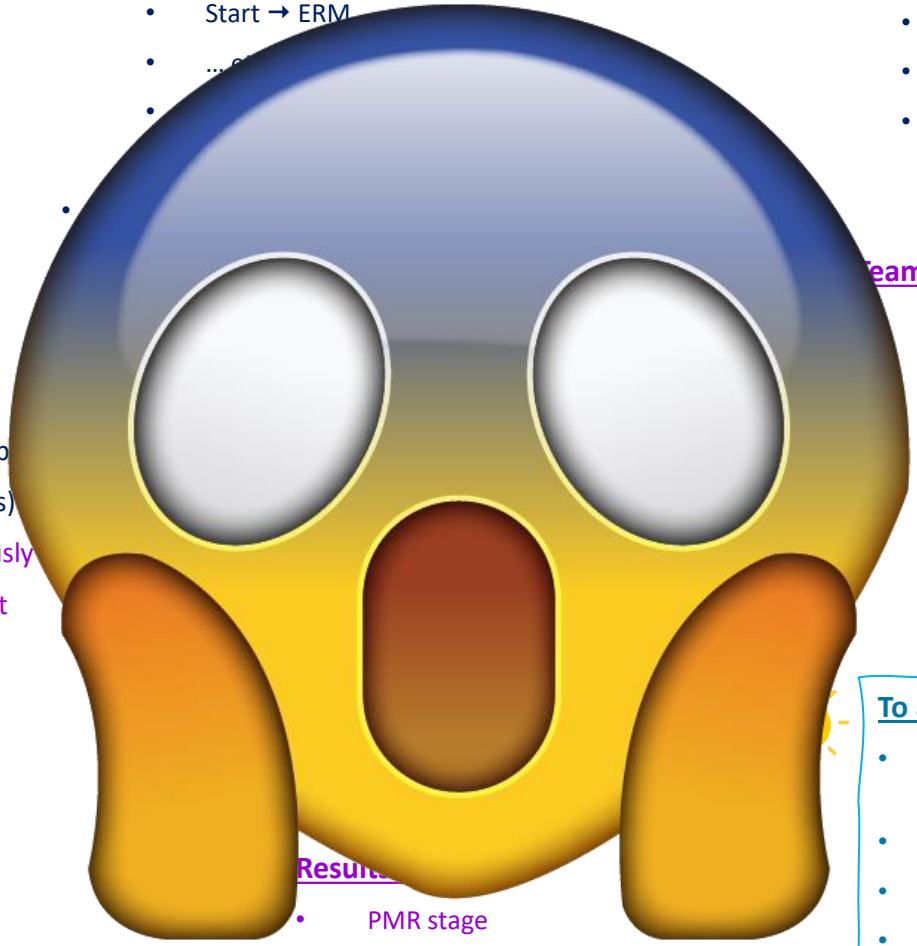
- Country
- Department
- Approval year
- Approval quarter
- Approved amount ($)
- Environmental/social risk rating
- Type of approval procedure
- Country requires ratification

Preparation data

- Time spans (months)
  - On Convergence → Start
  - Start → ERM
  - …
  - …
- …

Relationship data

- Has OPC TC
- Part of credit line
- Part of sequence
- # in sequence
- # of loan contracts



Team Leader data

- # changes in TL
- # projects TL managing
- # projects has managed in past
- Years been a TL
- *(no personal data)*

Executive agency experience

- Type of agency (ministry, municip
- Driving distance (km and minutes)
- # of projects agency simultaneously
- # of projects has managed in past
- Years been a client of IDB

Findings & recommendations

- language
- # of characters used
- # of fields entered
- Keyword search: "delay" and "disbursement"

Result

- PMR stage
- CPI & SPI
- Synthetic indicator (PMR)

To add

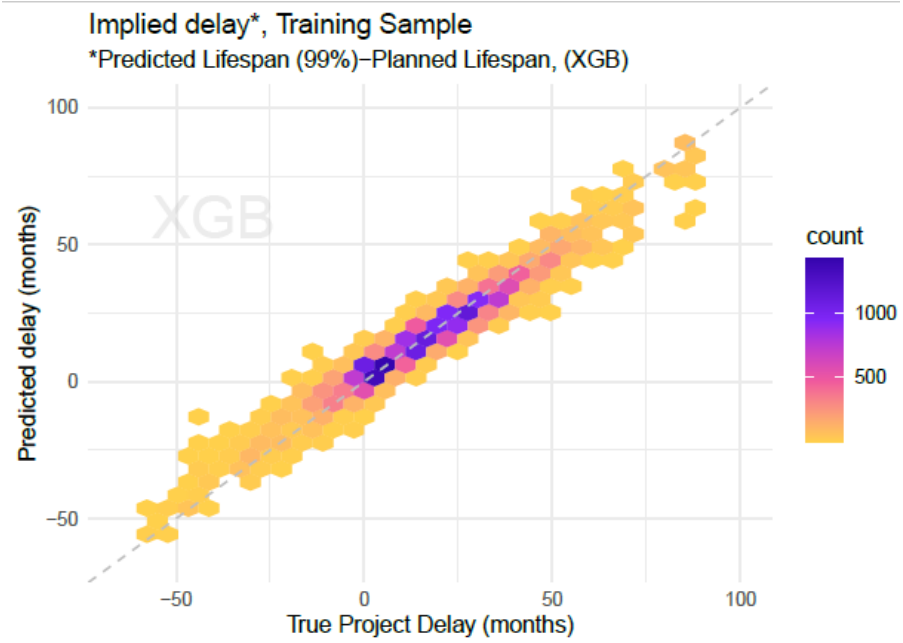
- External country-level variables (GDP growth, etc.)
- More time & labor info
- Expenses ($) data
- Team composition
- Ideas are welcome!

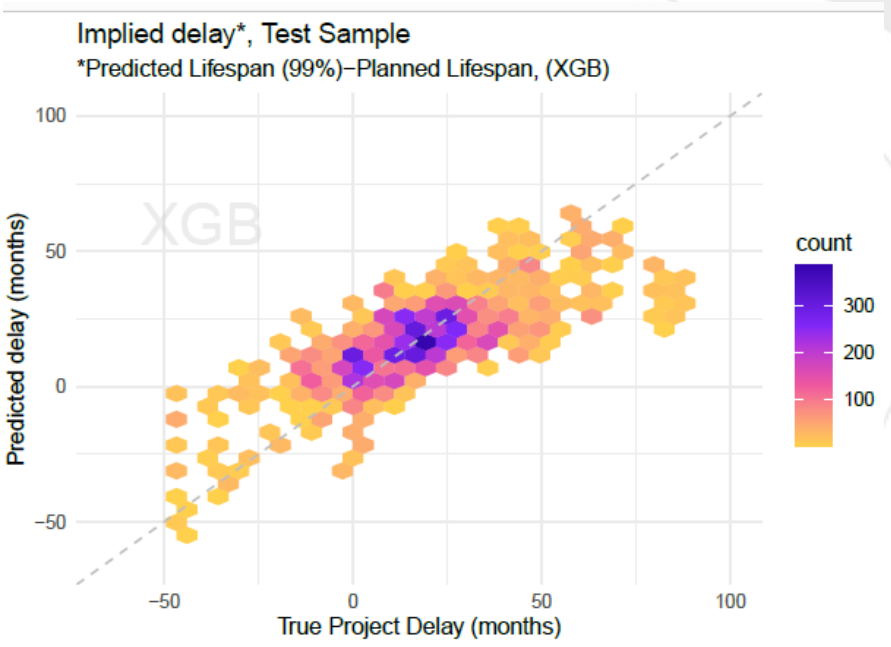
What are challenges?

Stay organized!

```
##########################################
# I. Prepare the data (~ 30 minutes to run)
##########################################


#Set password for EDW access
source(file = here::here("programs", "Initialize.R"))

#Prepare the feature data
source(file = here::here("programs", "Approvals.R"))
source(file = here::here("programs", "clean_HOPERMAS_fixed.variables.R"))
source(file = here::here("programs", "clean_HOPERMAS_temporal.variables.R"))
source(file = here::here("programs", "clean_relationships.R"))
source(file = here::here("programs", "clean_LoanContracts.R"))
source(file = here::here("programs", "clean_disbursement.R"))
source(file = here::here("programs", "clean_HOPERMAS_fixed.variables_IMPUTED.R"))
source(file = here::here("programs", "clean_outcomes.R"))
source(file = here::here("programs", "clean_preparation_expenses.R"))
source(file = here::here("programs", "clean_TimeLabor.R"))
source(file = here::here("programs", "clean_resultsIndicators.R"))
source(file = here::here("programs", "clean_ExecAgencies.R"))
source(file = here::here("programs", "clean_TeamLeaders.R"))
source(file = here::here("programs", "clean_findings_recomm.R"))
source(file = here::here("programs", "clean_CurrentDisbExprDate.R"))
source(file = here::here("programs", "clean_ExtData_econ.R"))
source(file = here::here("programs", "clean_combine_econ.R"))

#Merge the feature data to produce final dataset for modelling
source(file = here::here("programs", "IDBLoanDelays_build.R"))



##########################################
# II. Model (~10 hours to run)
##########################################

source(file = here::here("programs", "Model scripts (Jon)", "_LoanDelays_Exec_Master.R"))

```

What are the **results**?





Mean Absolute Error (MAE)	
Training set	Test set
3.6 months	11.1 months

What are the **results**?

1. Projects still disbursing | 2. Actual vs Predicted | 3. Breakdown

Outcome
- (•) delay
- ( ) lifespan

Confidence Interval: C.I. 99%

Color by: Modality

ESP
GCR
GOM
LBR

Region: (All)

Country: (All)

Sector: (All)

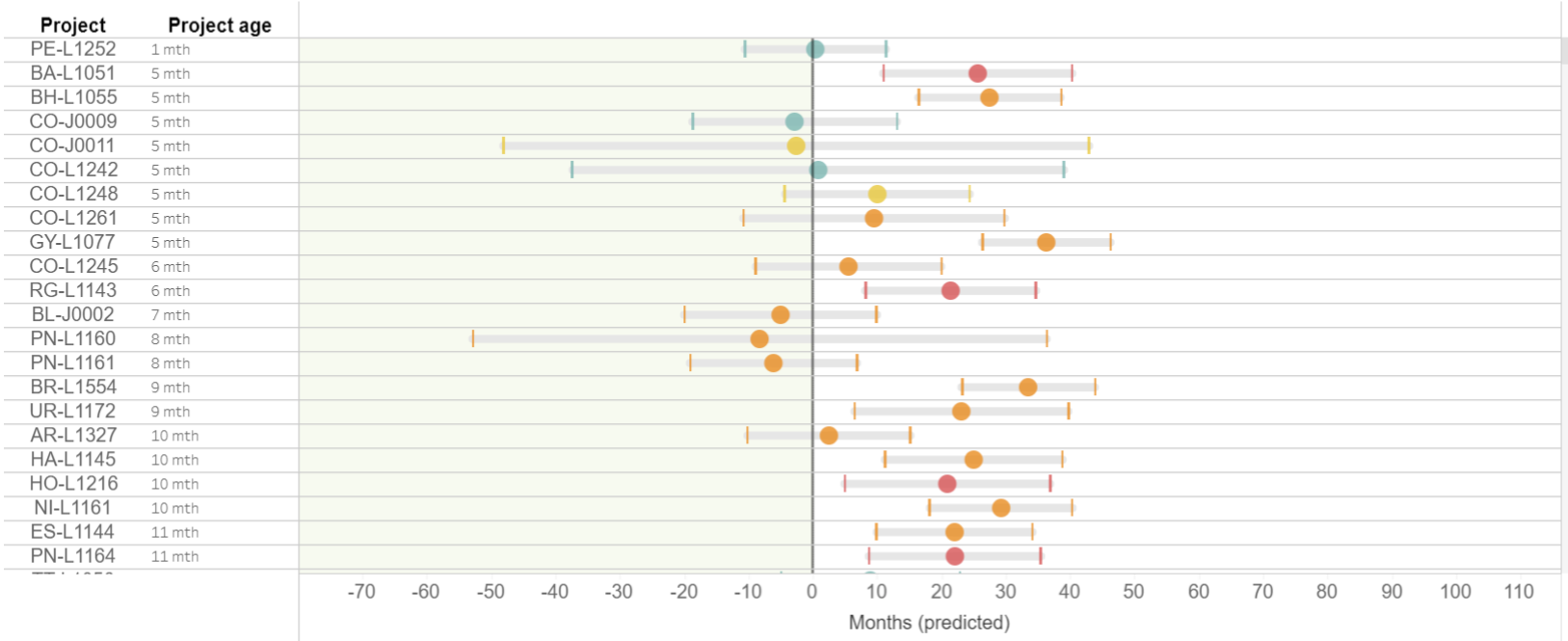
Division: (All)

Modality: (All)

Project: (All)

What is the predicted delay for projects in execution?

Delay = time (in months) from the Original Disbursement Expiration Date to the date of 99% disbursed.



The algorithm uses historical data to estimate project 'delay': the time (in months) from the Original Disbursement Expiration Date to the date at which the project disburses at least 99% of its current approved amount.

- The universe includes SG investment loan projects except for the following modalities: CND, IRF, and TCR.
- Projects shown have disbursed less than 99% of current approved amount as of 2021-05-31.
- 'Project age' refers to the time (in months) from project approval to 2021-05-31.
- With questions, email predict_delays@iadb.org

What are the **results**?

1. Projects still disbursing | 2. Actual vs Predicted | 3. Breakdown

Outcome
- (•) delay
- ( ) lifespan

Project age (months): 0

Sample: test

Confidence Interval: C.I. 99%

Region: (All)

Country: (All)

Sector: (All)

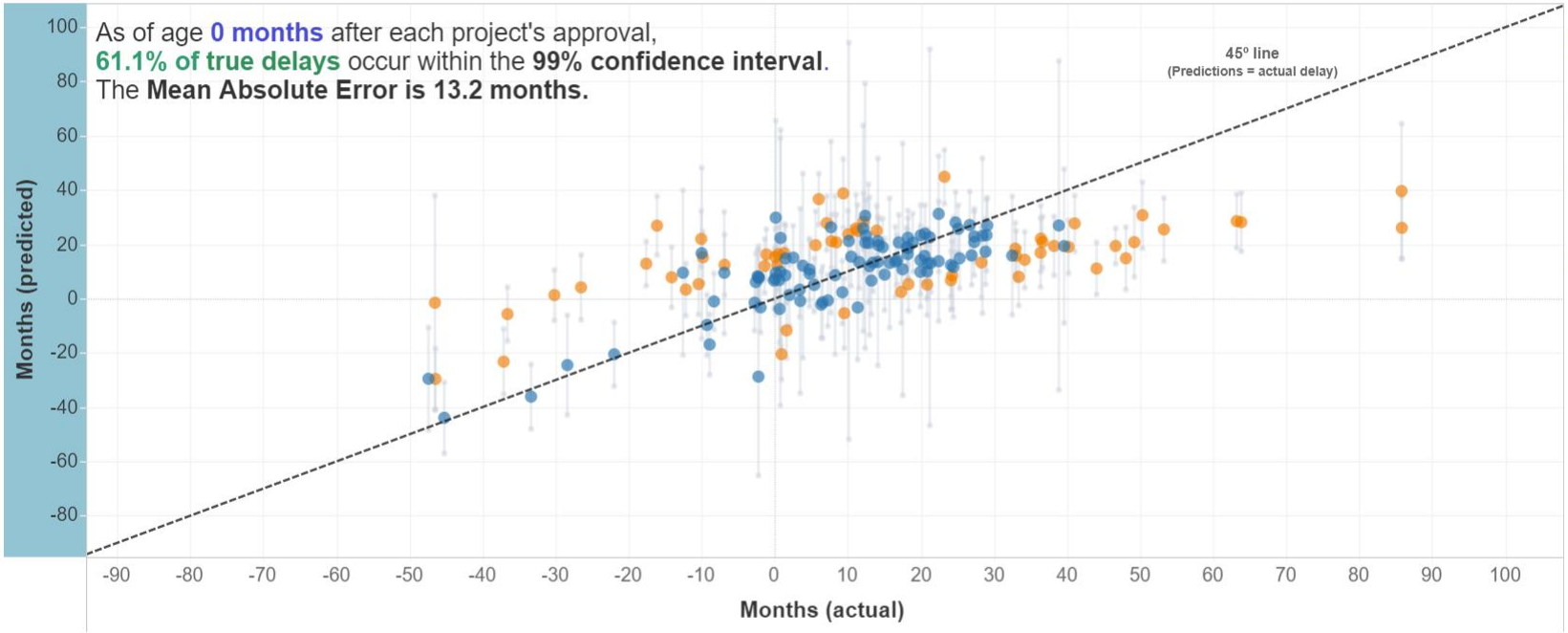
Division: (All)

Modality: (All)

Project: (All)

Color by: Accuracy

How accurate are the model's predictions?



Accurate
Not Accurate
C.I.

The algorithm uses historical data to estimate project 'delay': the time elapsed (months) from the Original Disbursement Expiration Date to the date at which the project disburses at least 99% of its current approved amount as of 2021-05-31.

- 'Project age' refers to the time elapsed from project approval to 2021-05-31.
- Training and test samples represent SG investment loans approved since 2008 that disbursed at least 99% of their current approved amount as of 2021-05-31. (We exclude the following modalities: CND, IRF, and TCR.)
- With questions, email predict_delays@iadb.org

What are the results?

1. Projects still disbursing | 2. Actual vs Predicted | 3. Breakdown

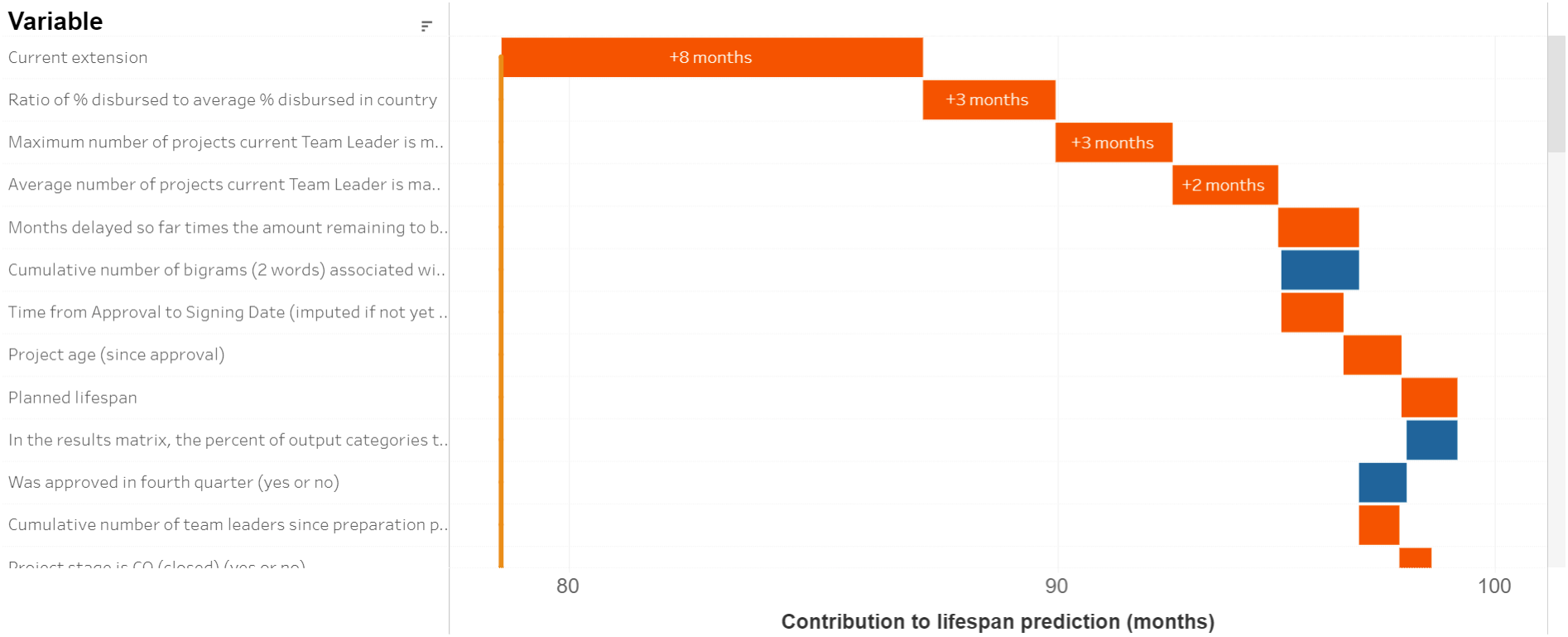
What influences predictions of project lifespan?

AR-L1067 (76% of $60M disbursed after 102 months)
as of 2021-05-31

Predicted lifespan: **99 months**
Predicted delay: **39 months**
Planned Lifespan: **60 months**
Current Extension: **54 months**

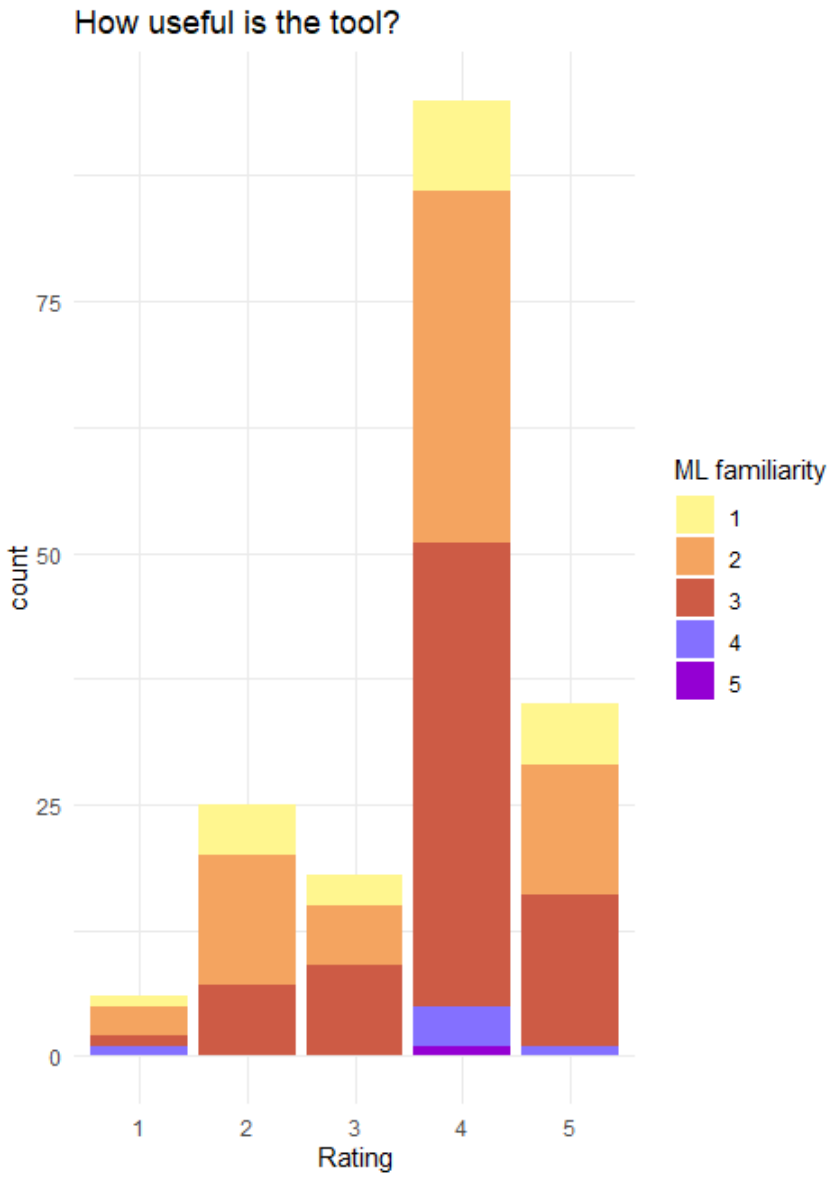
Choose a project
AR-L1067

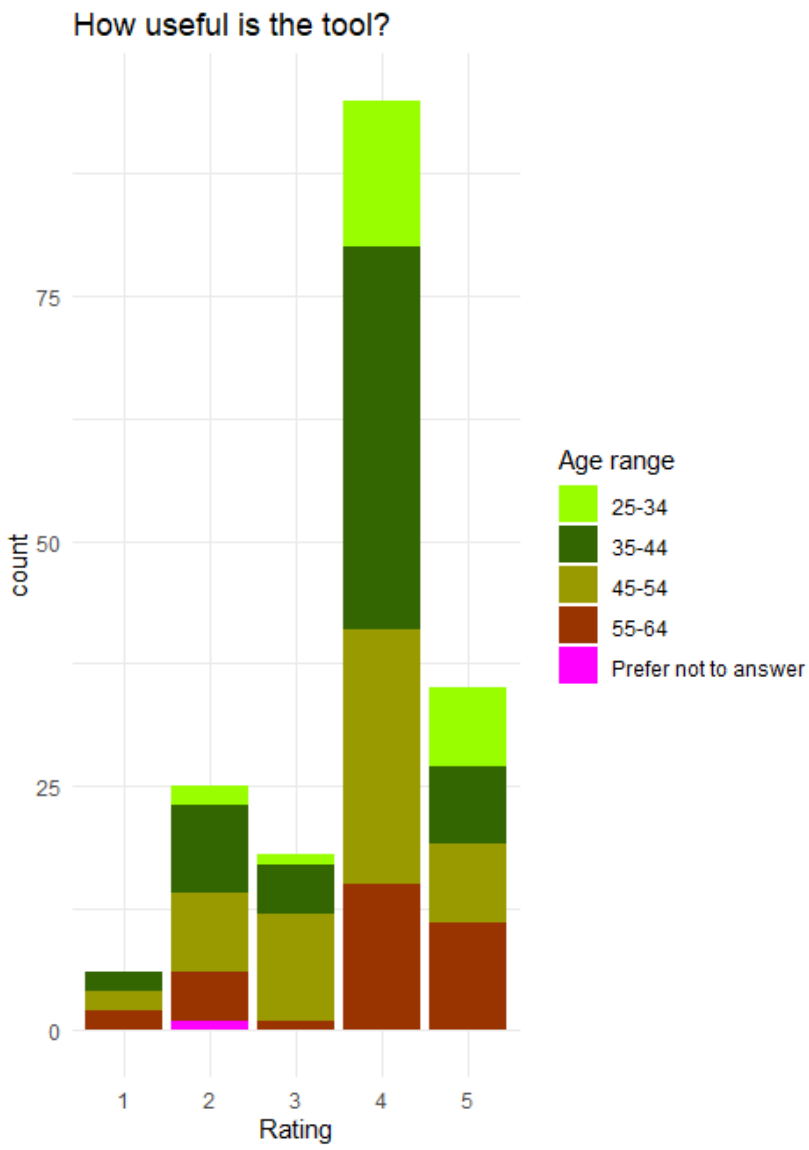
"Forest Sustainability and Competitiveness Program"
Approval date: **2012-11-28**
Country (Region): **AR (CSC)**
Sector/Division: **CSD/RND**
Modality: **ESP**



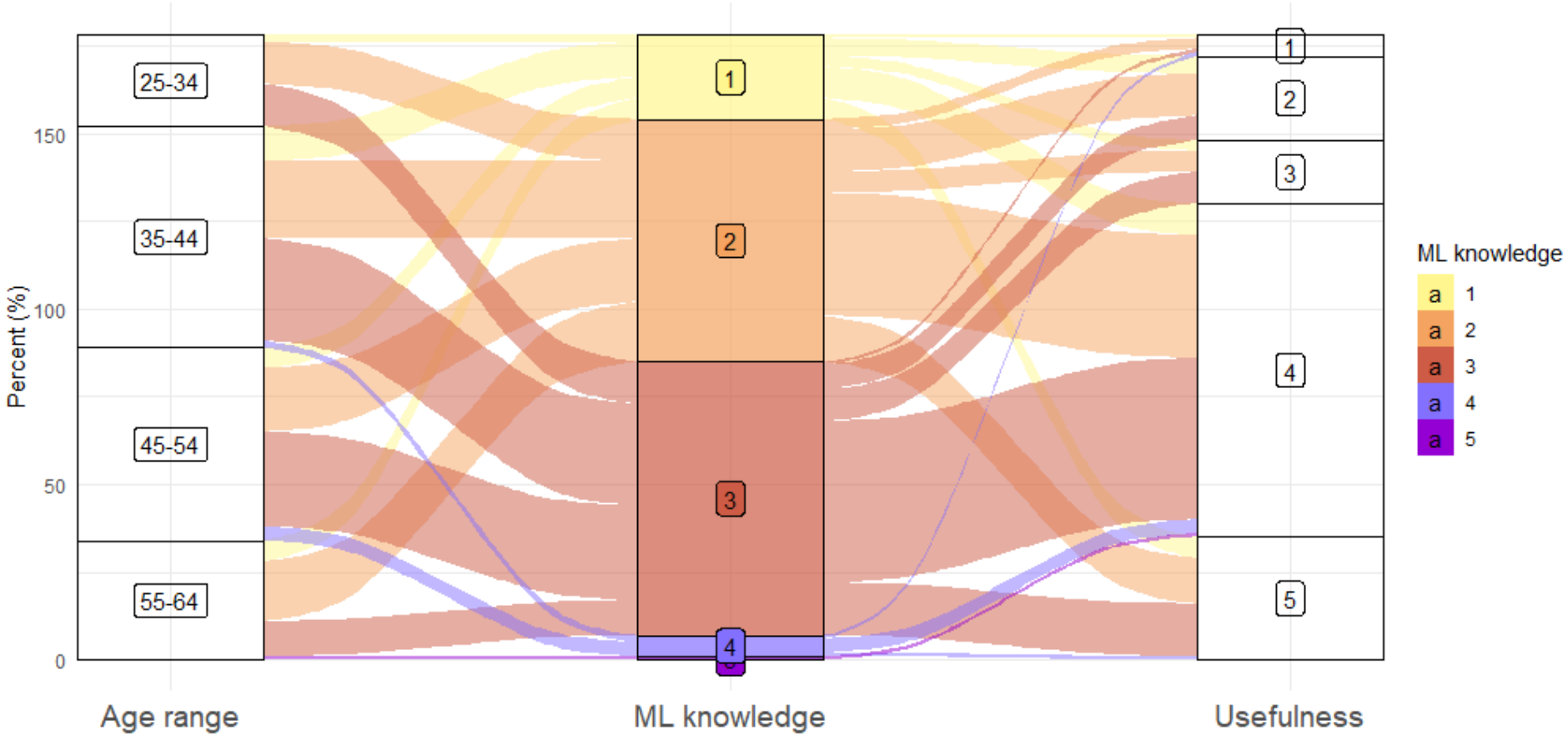
II. Do managers find the tool useful?

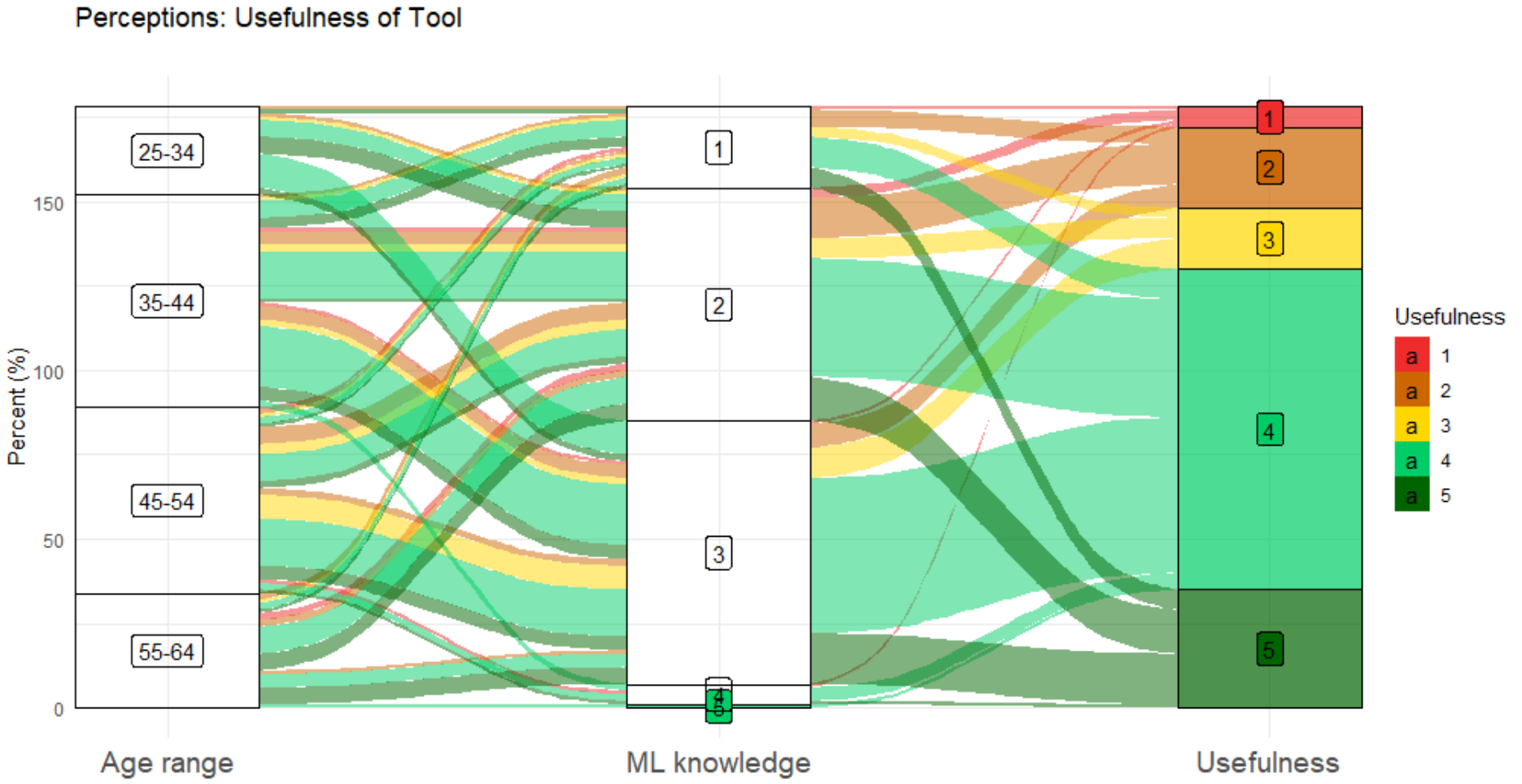
- Survey to managers and analysts (ongoing, started last week!)
  - 179 responses (≈40%)
- Main results:
  - **73%** find the tool useful
  - **85%** find the tool easy to understand
  - **66%** find the tool innovative

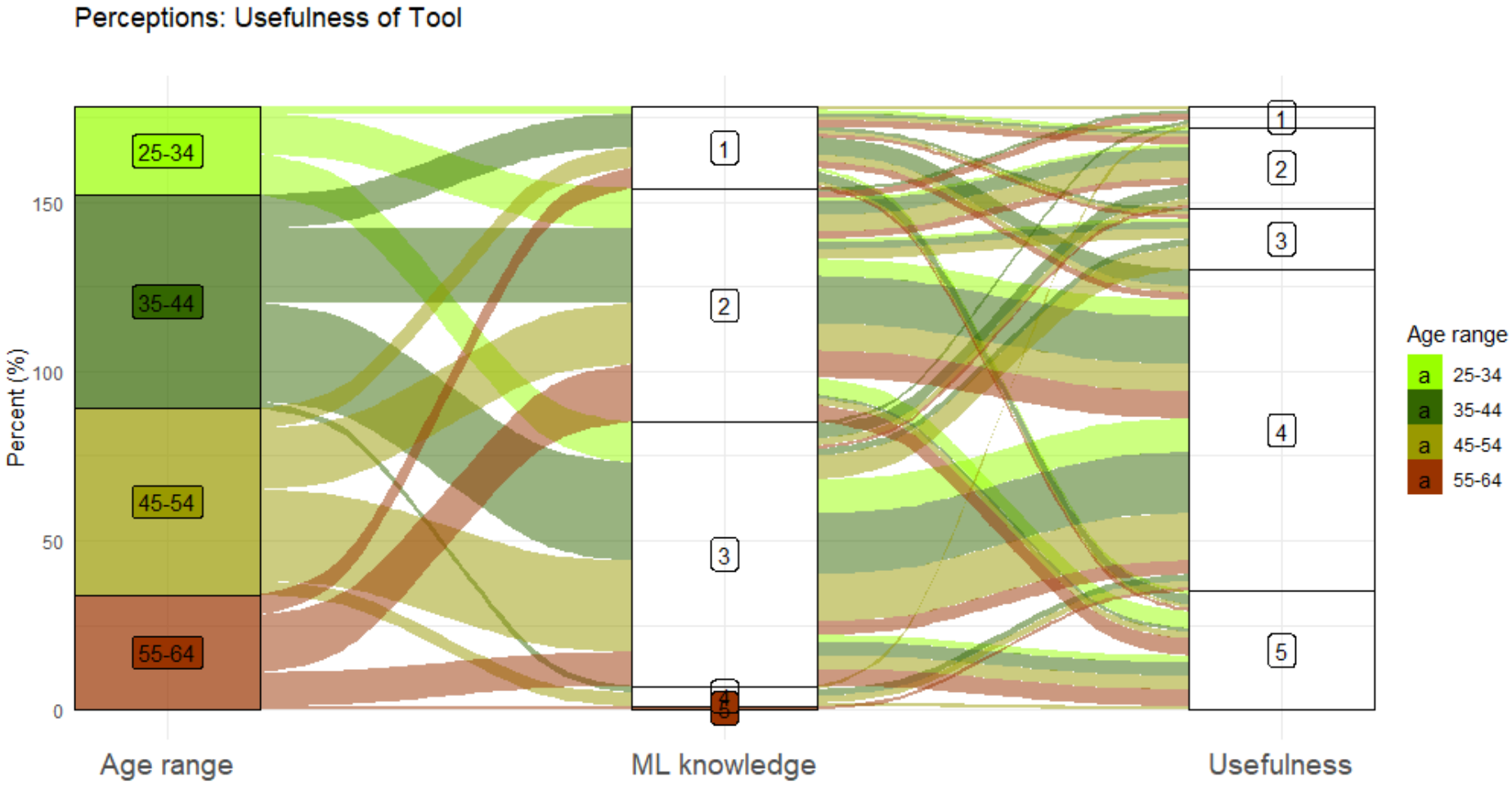

How useful is the tool?
count
75
50
25
0
1
2
3
4
5
Rating
ML familiarity
1
2
3
4
5


How useful is the tool?
count
75
50
25
0
1
2
3
4
5
Rating
Age range
25-34
35-44
45-54
55-64
Prefer not to answer

Perceptions: Usefulness of Tool




Perceptions: Usefulness of Tool
Percent (%)
150
100
50
0
25-34
35-44
45-54
55-64
1
2
3
4
5
Usefulness
a 1
a 2
a 3
a 4
a 5
Age range
ML knowledge
Usefulness


Perceptions: Usefulness of Tool
Percent (%)
150
100
50
0
25-34
35-44
45-54
55-64
1
2
3
4
5
1
2
3
4
5
Age range
ML knowledge
Usefulness
Age range
a 25-34
a 35-44
a 45-54
a 55-64