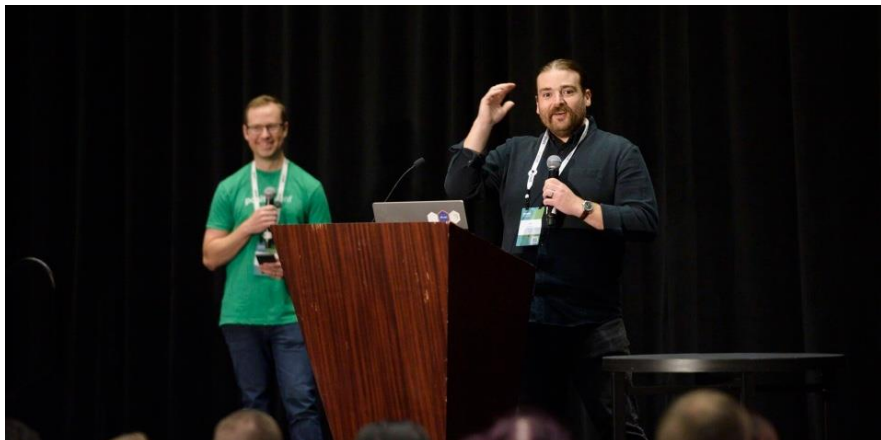




The Year of Yes! How Saying “Yes” Ignited GSK’s R Adoption

Becca Krouse and Ben Arancibia, GSK
PHUSE US Connect 2026

YES



Director of Data Science

**Scientific and Adoption Lead
for Open Source for GSK R&D**

**Passionate about open source
and teaching**



Associate Director of Data Science

**Enablement Lead for Open
Source for GSK R&D**

**Passionate about user-focus
and community**



INTRO	FIRST 3 MONTHS	SCALING UP	LOOKING FORWARD	FINAL JEOPARDY
\$200	\$200	\$200	\$200	\$200
\$400	\$400	\$400	\$400	\$400
\$600	\$600	\$600	\$600	\$600
\$800	\$800	\$800	\$800	\$800
\$1000	\$1000	\$1000	\$1000	\$1000

INTRO	FIRST 3 MONTHS	SCALING UP	LOOKING FORWARD	FINAL JEOPARDY
\$200	\$200	\$200	\$200	\$200
\$400	\$400	\$400	\$400	\$400
\$600	\$600	\$600	\$600	\$600
\$800	\$800	\$800	\$800	\$800
\$1000	\$1000	\$1000	\$1000	\$1000

**CAN WE DO 50% OF OUR
PROGRAMMING IN R?**

GSK Biostatistics' Commitment to Open-Source

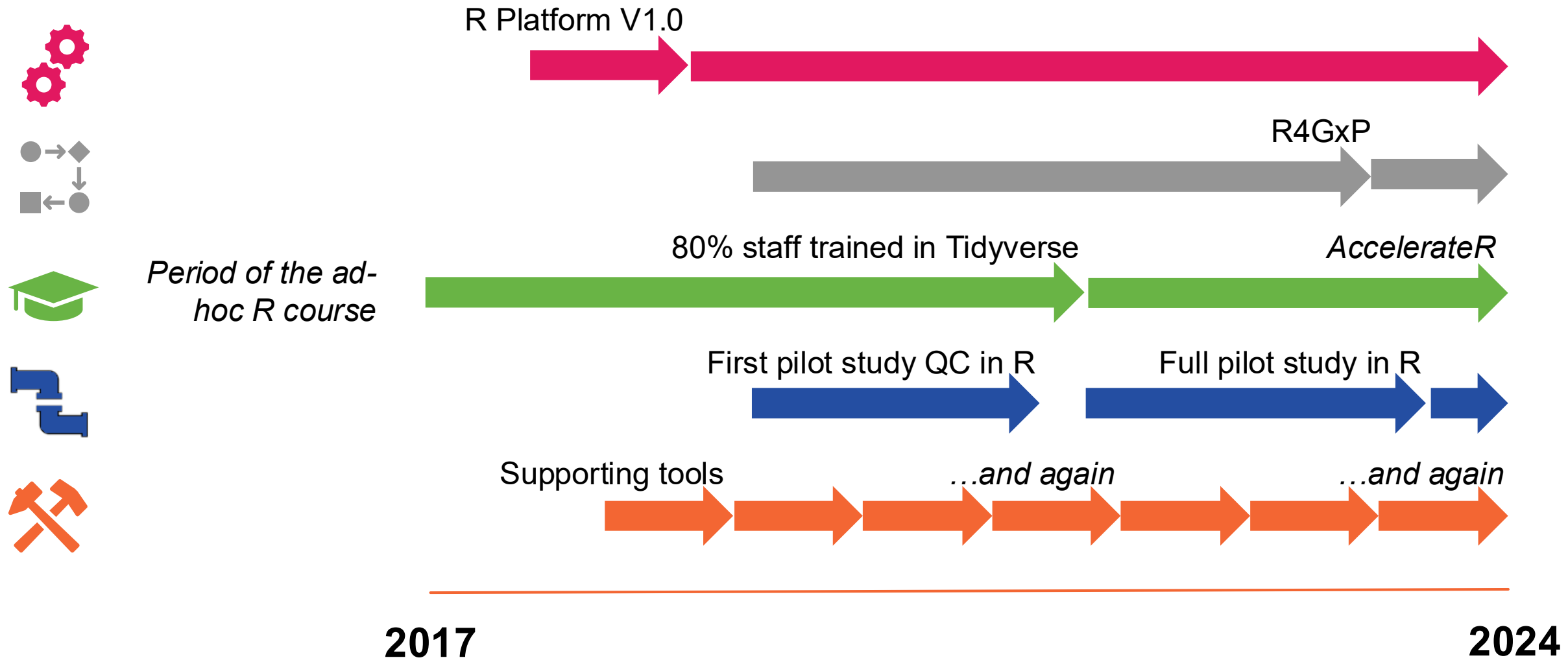


All central tools to be built
using open-source languages

50% of our code to be open-
source by end 2025

Photo by [Arturo Añez](#) on [Unsplash](#)

GSK's Open-source Journey Thus Far



STAR

Starting in 2025: GSK's R&D-Wide Open-Source Adoption Program

Photo by [Chantal & Ole on Unsplash](#)

27 March 2026

GSK

GSK Biostatistics' Commitment to Open-Source



All central tools to be built
using open-source languages

70%

~~50%~~ of our code to be open-
source by end 2025

stretch goal

Photo by [Arturo Añez](#) on [Unsplash](#)

Kicking off with a “burst” of support

GSK Jointly
launched STAR
and **STARburst**,
a support team
comprised of R
experts



Aims:

- Support STAR and the Biostatistics R Adoption Objective
- Enable skills development in the pipeline via *real-time* mentoring / support model
- Identify skills, tools, resource gaps

INTRO	FIRST 3 MONTHS	SCALING UP	LOOKING FORWARD	FINAL JEOPARDY
\$200	\$200	\$200	\$200	\$200
\$400	\$400	\$400	\$400	\$400
\$600	\$600	\$600	\$600	\$600
\$800	\$800	\$800	\$800	\$800
\$1000	\$1000	\$1000	\$1000	\$1000

**CAN WE USE R END-TO-
END?**

▶ STARburst: 3-months of “yes” to every inquiry

*“Dear
STARburst
team...”*

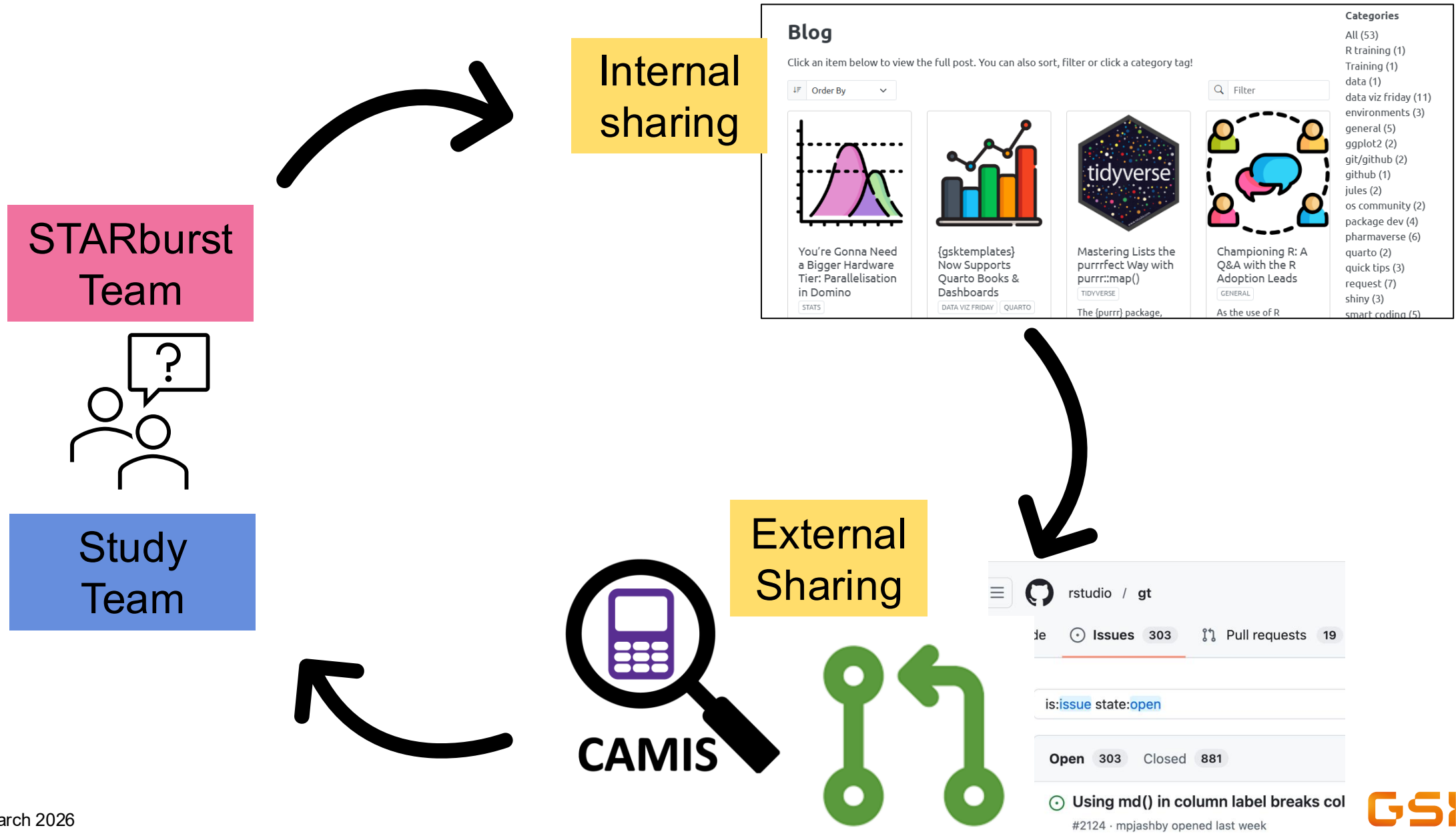


Unlocking and Upskilling

```
+ geom_boxplot()  
  
Error: object 'ADSL' not found
```

27 March 2026

Tackling requests with an open-source mentality





TFL Examples in R

Preface

Demographic Characteristics

DM-T01: Summary of Demographic and Baseline Characteristics

DM-T02: Summary of Age Groups

DM-T03: Summary of Number of Subjects by Country and Site Id.

DM-L01: Listing of Demographic Characteristics

DM-L02: Listing of Race

Adverse Events

TFL Examples in R

AUTHOR
STARburst Team

Preface

Important

This example library is intended solely as a reference for techniques and **should not be directly copied and pasted** for application in your studies. The code presented in this resource **has not undergone validation**. It is imperative to employ independent sources for quality control processes and to ensure a comprehensive understanding of any code utilized.

Overview

- This is a library of TFL examples separated by category. The examples are based on the end-to-end (e2e) display mock shells, which can be viewed [here](#).
- The examples include steps for preparing, filtering and processing the dataset using the **Tidyverse** and **{cards}**, and generating the output table using **{gt}** for listings or **{tfrmt}** for tables. The **{pharmaverseadam}** package is used for dummy data.
- The examples utilise some helper functions that have been developed specifically for the example

Adverse Events Plot

FOREST

Example: Common AEs sorted by RR (AE_F01)

Packages: pharmaverseadam, cards, patchwork, dcorator

Bar Chart

BAR

Example: Efficacy - Maximum percent reduction by subject

Box Plot

BOX

Example: Summary of PK parameters

Packages: pharmaverseadam, gskthemes

Categories

All (14)

bar (1)

box (3)

forest (3)

histogram (2)

Kaplan-Meier (2)

PK (1)

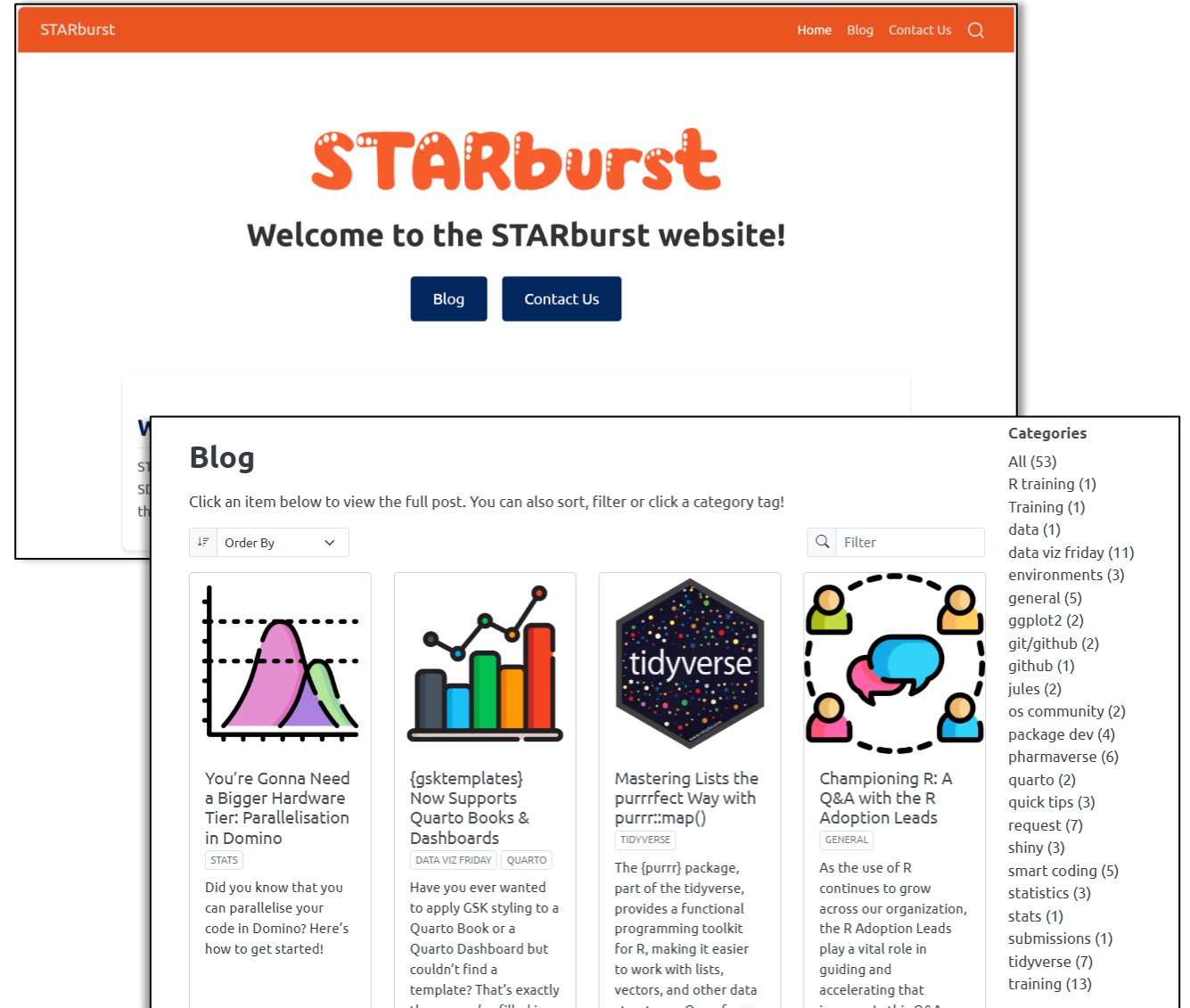
profile (4)

scatter (2)

STARburst by the numbers

A look at the first 3 months

- **64 blog posts (posted daily)**
 - Topics driven by the pipeline, featuring case studies and tricks/tips
 - 821 unique viewers (~90% returning)
 - 16004+ views
- **58 requests (~1 per day)**
 - Programming - 47%
 - Statistics - 48%



INTRO	FIRST 3 MONTHS	SCALING UP	LOOKING FORWARD	FINAL JEOPARDY
\$200	\$200	\$200	\$200	\$200
\$400	\$400	\$400	\$400	\$400
\$600	\$600	\$600	\$600	\$600
\$800	\$800	\$800	\$800	\$800
\$1000	\$1000	\$1000	\$1000	\$1000



DAILY DOUBLE

This Photo by Unknown Author is licensed under [CC BY-NC-ND](#)

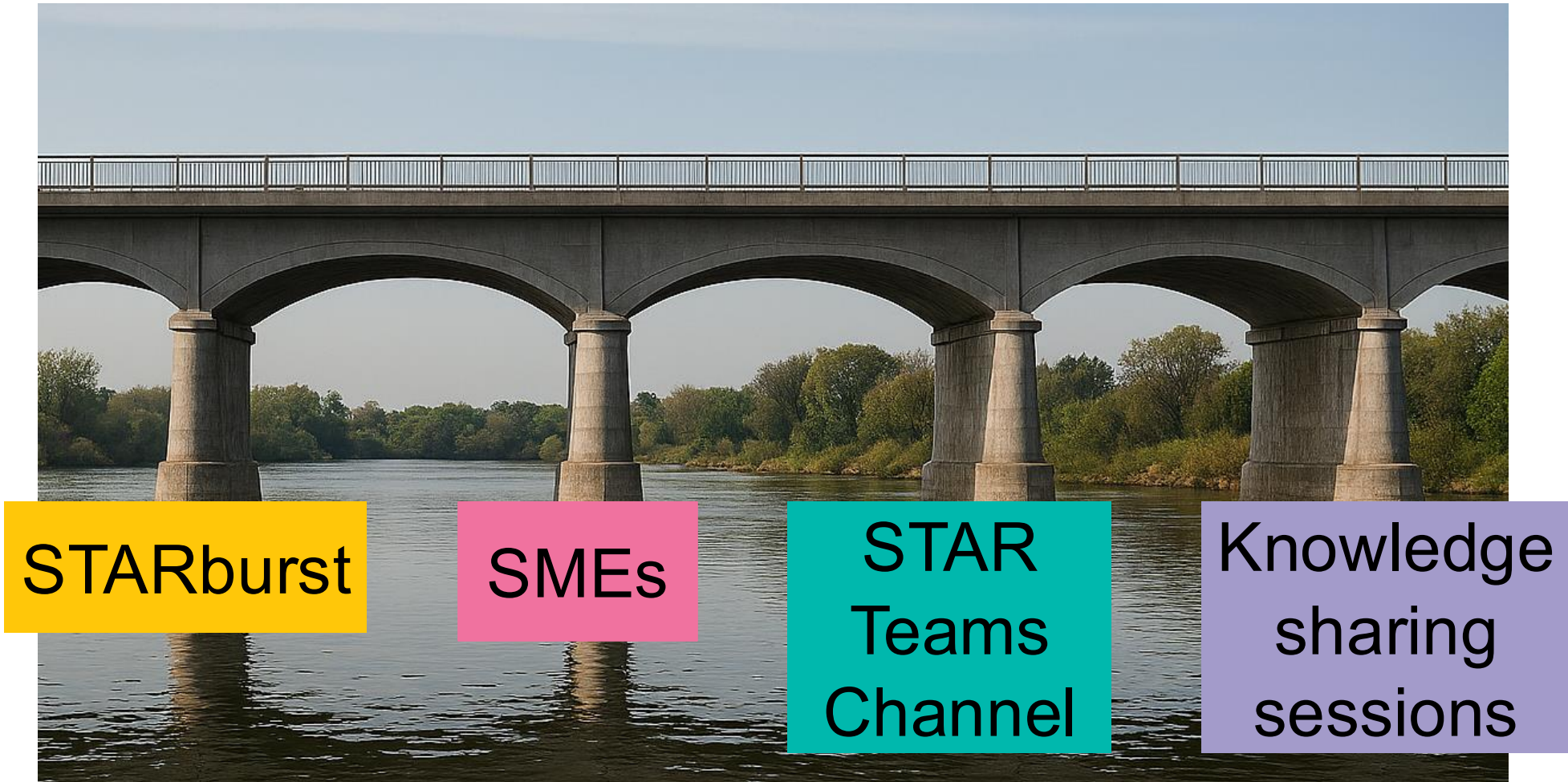
**HOW DO WE TURN OUR
LEARNINGS INTO SHARED
PRACTICE?**

Celebrating our impact, recognizing the challenge ahead

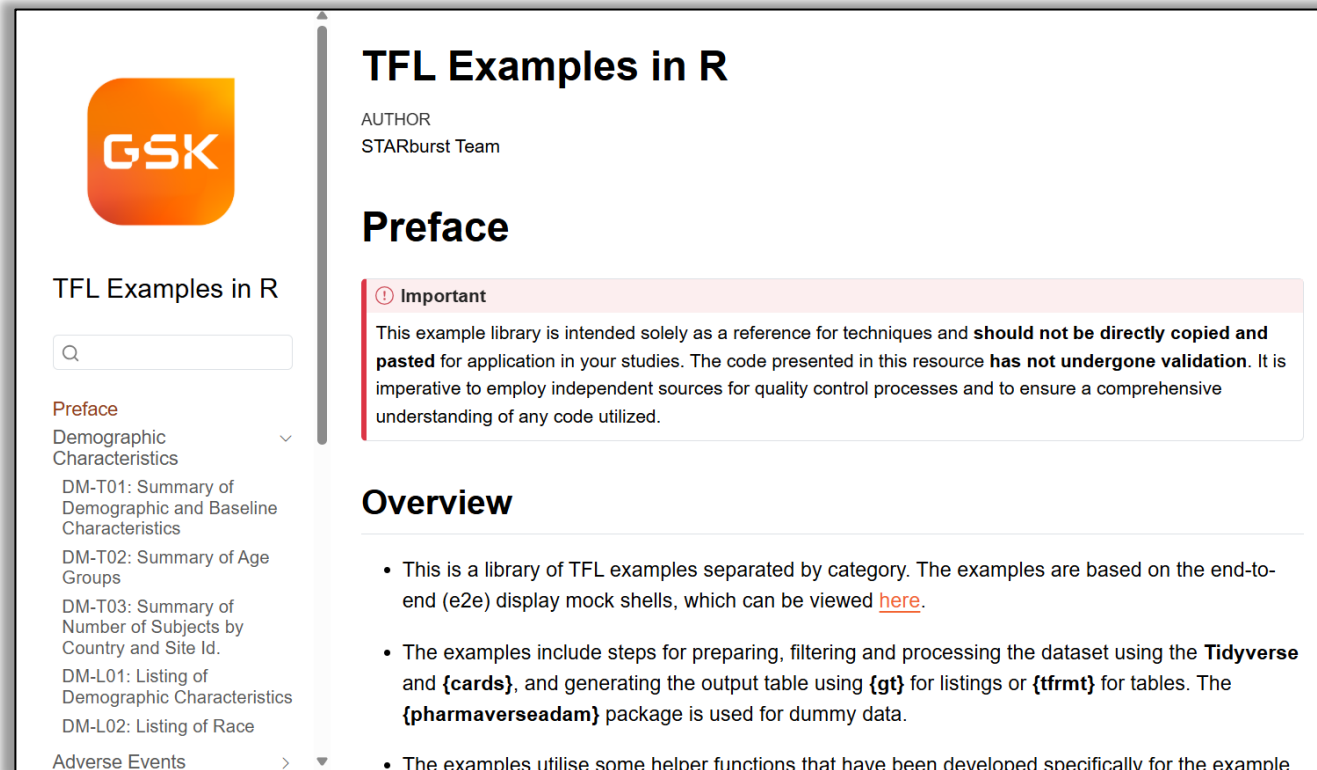
- Users want more than isolated examples, they want step-by-step workflows
- Each department, therapeutic area, study, regulatory agency, etc. presents unique scenarios
- We don't know what we don't know



Scaling through community-wide support



Scaling through a wider variety of examples



TFL Examples in R

AUTHOR
STARburst Team

Preface

Important

This example library is intended solely as a reference for techniques and **should not be directly copied and pasted** for application in your studies. The code presented in this resource **has not undergone validation**. It is imperative to employ independent sources for quality control processes and to ensure a comprehensive understanding of any code utilized.

Overview

- This is a library of TFL examples separated by category. The examples are based on the end-to-end (e2e) display mock shells, which can be viewed [here](#).
- The examples include steps for preparing, filtering and processing the dataset using the **Tidyverse** and **{cards}**, and generating the output table using **{gt}** for listings or **{tfmrt}** for tables. The **{pharmaverseadam}** package is used for dummy data.
- The examples utilise some helper functions that have been developed specifically for the example

+ FAQs

**HOW DO WE FIGURE OUT
WHAT WE DON'T KNOW?**

Scaling through robust workflows



Photo by [Roger Starnes Sr](#) on [Unsplash](#)

- “Tiger Team” assembled to deep dive on and share detailed workflows
- Examples:
 - Risk-based QC with evolving tooling
 - Shift from sas7bdat files to language-agnostic Parquet for internal storage
- Outcome: foundational support and clear pathways to confident, day-to-day R adoption and integration

Scaling through interdisciplinary problem solving

- Formed interdisciplinary squads to tackle complex challenges
 - Statistical methodology
 - Code/environment sharing
 - Evolving coding/debugging best practices
- Shifted focus from quick fixes to long-term solutions
- Result: strengthened tools, practices, and shared ways of working



Photo by [Helena Lopes](#) on [Unsplash](#)

INTRO	FIRST 3 MONTHS	SCALING UP	LOOKING FORWARD	FINAL JEOPARDY
\$200	\$200	\$200	\$200	\$200
\$400	\$400	\$400	\$400	\$400
\$600	\$600	\$600	\$600	\$600
\$800	\$800	\$800	\$800	\$800
\$1000	\$1000	\$1000	\$1000	\$1000

**CAN WE CONTINUE TO
GROW IN A SELF-
SUSTAINING MANNER?**

Requirements for self-sustaining R adoption

- Proactively identifying and closing gaps
- Deepen technical proficiencies in advanced skills such as debugging and managing R environments
- Embrace continuous learning and adaptation
- Fostering an open-source culture



The ultimate goal

Working toward a
future where
**innovation is
shared and
collaboratively
maintained**



INTRO	FIRST 3 MONTHS	SCALING UP	LOOKING FORWARD	FINAL JEOPARDY
\$200	\$200	\$200	\$200	\$200
\$400	\$400	\$400	\$400	\$400
\$600	\$600	\$600	\$600	\$600
\$800	\$800	\$800	\$800	\$800
\$1000	\$1000	\$1000	\$1000	\$1000

FINAL JEOPARDY

CAN WE DO 50% OF OUR
PROGRAMMING IN R?

CAN WE USE R END-TO-END?

CAN WE SCALE TO THE FULL
ORGANIZATION?

CAN WE CONTINUE TO GROW IN A
SELF-SUSTAINING MANNER?

What is

Yes!



Questions?



Contact the authors:

Becca Krouse
becca.z.krouse@gsk.com

Ben Arancibia
benjamin.c.arancibia@gsk.com