Leveraging GitHub Actions and Continuous Integration in R Package Qualification

A Case Study of the {gsm} R Package

Matt Roumaya

Mike Stackhouse

PHUSE US Connect 2024

atorus

Your Complete Data Solution

Agenda

- Introduction
- Qualification Workflow
- Automation via GitHub Actions
- Next Steps

atorus
Your Complete Data Solution

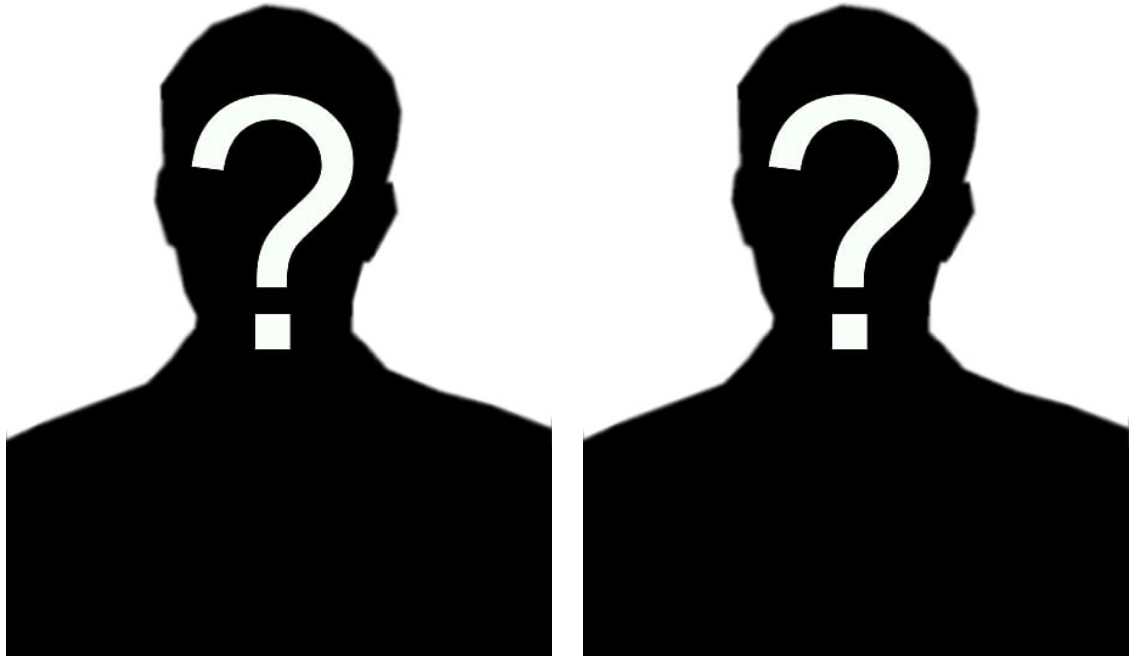
Introduction

atorus

Your Complete Data Solution

Introduction

- Who are we?



atorus
Your Complete Data Solution

Introduction

- What is {gsm}?

Good Statistical Monitoring

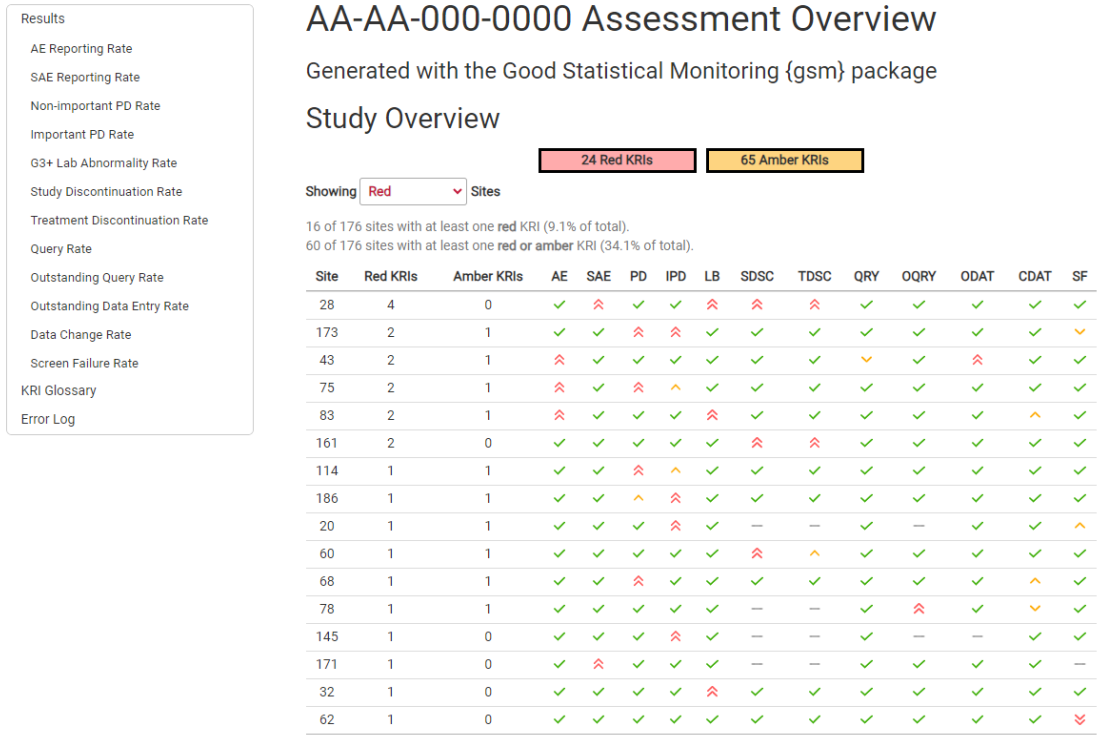
https://github.com/Gilead-BioStats/gsm

atorus

Your Complete Data Solution

Introduction

- The Good Statistical Monitoring or “gsm” package is a free and open-source framework for Risk Based Quality Monitoring (RBQM) analytics.



Results

- AE Reporting Rate
- SAE Reporting Rate
- Non-important PD Rate
- Important PD Rate
- G3+ Lab Abnormality Rate
- Study Discontinuation Rate
- Treatment Discontinuation Rate
- Query Rate
- Outstanding Query Rate
- Outstanding Data Entry Rate
- Data Change Rate
- Screen Failure Rate

KRI Glossary

Error Log

AA-AA-000-0000 Assessment Overview

Generated with the Good Statistical Monitoring {gsm} package

Study Overview

24 Red KRIs | 65 Amber KRIs

Showing Red Sites

16 of 176 sites with at least one **red** KRI (9.1% of total).
60 of 176 sites with at least one **red or amber** KRI (34.1% of total).

Site	Red KRIs	Amber KRIs	AE	SAE	PD	IPD	LB	SDSC	TDSC	QRY	OQRY	ODAT	CDAT	SF
28	4	0	✓	⌃⌃	✓	✓	⌃⌃	⌃⌃	⌃⌃	✓	✓	✓	✓	✓
173	2	1	✓	✓	⌃⌃	⌃⌃	✓	✓	✓	✓	✓	✓	✓	⌄
43	2	1	⌃⌃	✓	✓	✓	✓	✓	✓	⌄	✓	⌃⌃	✓	✓
75	2	1	⌃⌃	✓	⌃⌃	⌃	✓	✓	✓	✓	✓	✓	✓	✓
83	2	1	⌃⌃	✓	✓	✓	⌃⌃	✓	✓	✓	✓	✓	⌃	✓
161	2	0	✓	✓	✓	✓	✓	⌃⌃	⌃⌃	✓	✓	✓	✓	✓
114	1	1	✓	✓	⌃⌃	⌃	✓	✓	✓	✓	✓	✓	✓	✓
186	1	1	✓	✓	⌃	⌃⌃	✓	✓	✓	✓	✓	✓	✓	✓
20	1	1	✓	✓	✓	⌃⌃	✓	–	–	✓	–	✓	✓	⌃
60	1	1	✓	✓	✓	✓	✓	⌃⌃	⌃	✓	✓	✓	✓	✓
68	1	1	✓	✓	⌃⌃	✓	✓	✓	✓	✓	✓	✓	⌃	✓
78	1	1	✓	✓	✓	✓	✓	–	–	✓	⌃⌃	✓	⌄	✓
145	1	0	✓	✓	✓	⌃⌃	✓	–	–	✓	–	–	✓	✓
171	1	0	✓	⌃⌃	✓	✓	✓	–	–	✓	✓	✓	✓	–
32	1	0	✓	✓	✓	✓	⌃⌃	✓	✓	✓	✓	✓	✓	✓
62	1	0	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	⌄⌄

atorus
Your Complete Data Solution

Introduction

- The package provides a standardized RBQM framework for clinical trials that pairs a flexible data pipeline with interactive reporting.

atorus
Your Complete Data Solution

Introduction

- Learn more here! https://gilead-biostats.github.io/gsm/
- Consider attending talk AR04 in Salon C @ 1:30pm 😎

atorus
Your Complete Data Solution

Qualification Framework

Specifications, Test Cases, and Reporting

atorus

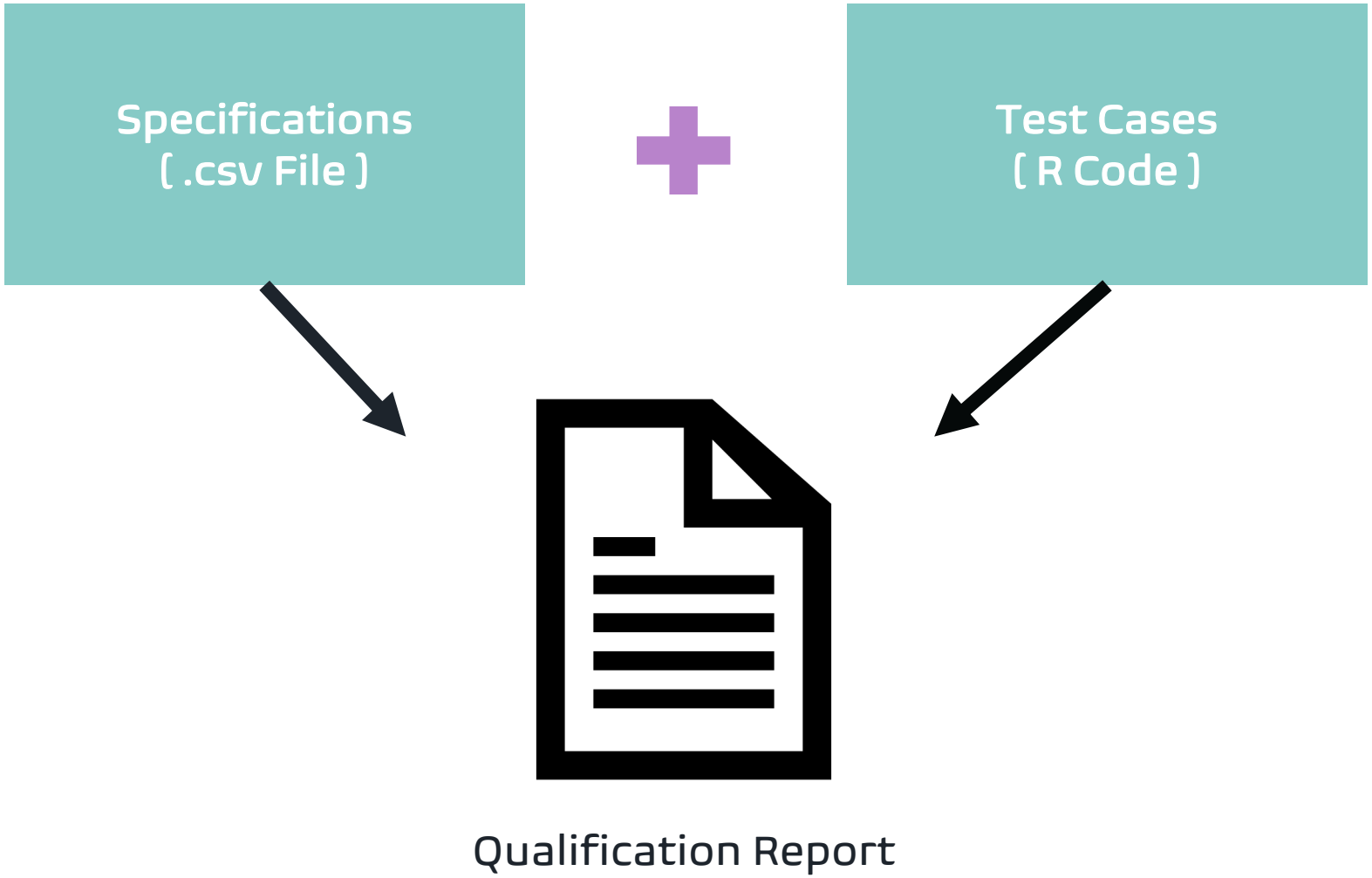
Your Complete Data Solution

Qualification vs. Unit Testing vs. Validation

- **Qualification**: *Can we demonstrate that the full package both functions and executes as expected?*
- **Unit Testing**: *Do these small pieces function like I intended?*
- **Validation**: *Can it be clearly demonstrated that the requirements of my intended users have been met?*

atorus
Your Complete Data Solution

Qualification Workflow



atorus
Your Complete Data Solution

Qualification Developer

- Use {testthat} R package framework
- Intentionally separate from dev team
- Double-program and confirm expected results

atorus
Your Complete Data Solution

Specifications: Example

- Specifications: Capture the most important uses cases for a given function.

Spec	Test ID	Tests	Function Name	Description
1	1	T1_1, T1_2, T1_3, T1_10	AE_Assess	Given appropriate input data, an Adverse Event Assessment can be done using the Normal Approximation method.

atorus
Your Complete Data Solution

Test Case: Example

```
test_that("Given an appropriate subset of Query Rate data, the assessment function correctly performs a
Query Rate Assessment grouped by the Site variable using the Poisson method and correctly assigns Flag
variable values when given a custom threshold.", {
  # gsm analysis
  dfInput <- gsm::QueryRate_Map_Raw(dfs = list(
    dfQUERY = clindata::edc_queries %>% filter(visit == "Week 120"),
    dfSUBJ = clindata::rawplus_dm,
    dfDATACHG = clindata::edc_data_points
  ))

  test10_1 <- QueryRate_Assess(
    dfInput = dfInput,
    strMethod = "Poisson",
    strGroup = "Site",
    vThreshold = c(-6, -4, 4, 6)
  )

  # double programming
  t10_1_input <- dfInput

  t10_1_transformed <- dfInput %>%
    qualification_transform_counts(
      countCol = "Count",
      exposureCol = "DataPoint"
    )

  # more code here...

  # compare results
  expect_equal(test10_1$lData[!names(test10_1$lData) %in% c("dfBounds", "dfConfig")], t10_1)
})
```

Specification

Test For Equality

atorus

Your Complete Data Solution

{gsm} Specifications + Test Cases

- 89 Specifications across 24 functions.
- 170+ test cases.
- This is in addition to roughly 1,500 unit tests, standard R package documentation, and a thorough release and QC process.
- Continually adding as development continues.

atorus
Your Complete Data Solution

Qualification Report: Example

Qualification Report

atorus
Your Complete Data Solution

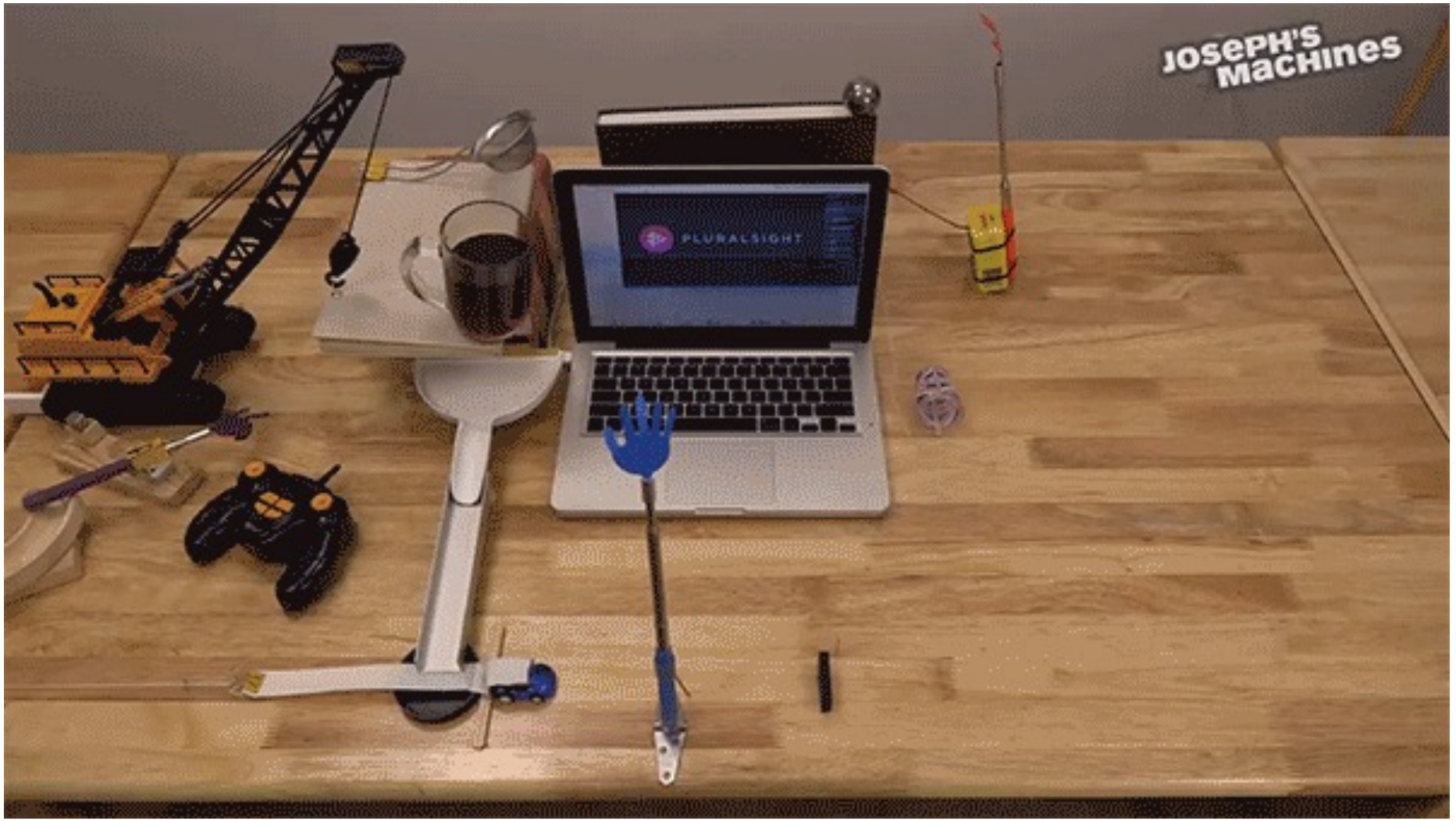
The Fun Part

Automation Via GitHub Actions

atorus

Your Complete Data Solution

GitHub Actions



atorus
Your Complete Data Solution

YAML File: Qualification Checks

```yaml
# Workflow derived from https://github.com/r-lib/actions/tree/master/examples
# Need help debugging build failures? Start at https://github.com/r-lib/actions#where-to-find-help
on:
  pull_request:
    branches: dev

name: R-CMD-check-dev

jobs:
  R-CMD-check:
    runs-on: ${{ matrix.config.os }}

    name: ${{ matrix.config.os }} (${{ matrix.config.r }})

    strategy:
      fail-fast: false
      matrix:
        config:
          - {os: ubuntu-latest, r: '4.0.5', repos:
"https://packagemanager.rstudio.com/cran/__linux__/focal/latest"}
          - {os: ubuntu-latest, r: '4.1.3', repos:
"https://packagemanager.rstudio.com/cran/__linux__/focal/latest"}




      - name: run qualification tests
        run: source(here::here("tests", "testqualification", "test_qualification.R"))
        shell: Rscript {0}
```

atorus

Your Complete Data Solution

YAML File: Qualification Checks

```yaml
on:
  release:
    types: [published]
  pull_request:
    branches: main

name: qualification-report

jobs:
  qualification-report:
    runs-on: ubuntu-latest
    env:
      GITHUB_PAT: ${{ secrets.GITHUB_TOKEN }}

      - name: build vignette
        run: tools::buildVignette("./vignettes/articles/Qualification.Rmd")
        shell: Rscript {0}

      # Upload the qulafication report to the release
      - name: Upload report to release
        if: success()
        uses: svenstaro/upload-release-action@v2
        with:
          file: ./Qualification.pdf
          asset_name: qualification-report.pdf
          repo_token: ${{ secrets.GITHUB_TOKEN }}
          tag: ${{ github.ref_name }}
          overwrite: false
```

atorus
Your Complete Data Solution

Next Steps

{qcthat}?

atorus

Your Complete Data Solution

Quietly Announcing {qcthat}

- We have created a small utility package that has borrowed heavily from {usethis} to help R package developers use a similar qualification framework in your own R package.
- In your R package, simply install and run:

```
devtools::install_github("Gilead-BioStats/qcthat")

qcthat::qcthat()
```

atorus
Your Complete Data Solution

{qcthat} Structure

- {qcthat} gives you a basic scaffold:
  - 📁./inst/qualification/qualification_specs.csv
  - 📁./tests/testqualification/qualification
  - 📁./vignettes/articles/Qualification.Rmd
  - 📁./.github/workflows/qualification_report.yaml
  - Modifies DESCRIPTION file
  - Adds test case placeholder

atorus
Your Complete Data Solution

https://gilead-biostats.github.io/qcthat/

atorus
Your Complete Data Solution

Thank You!