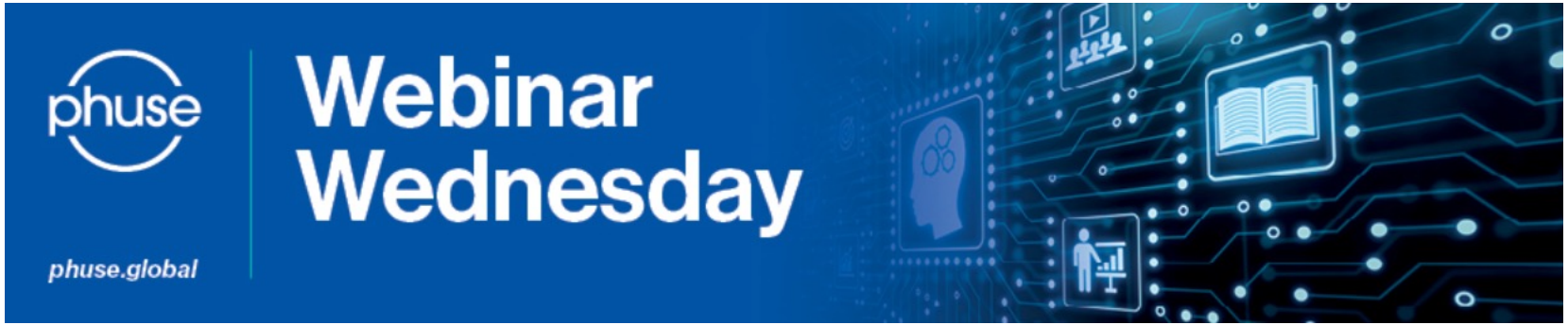
phuse·virtual
CSS 2021
The Computational Science Symposium
September 13th–15th
phuse.global

PHUSE State of the Union

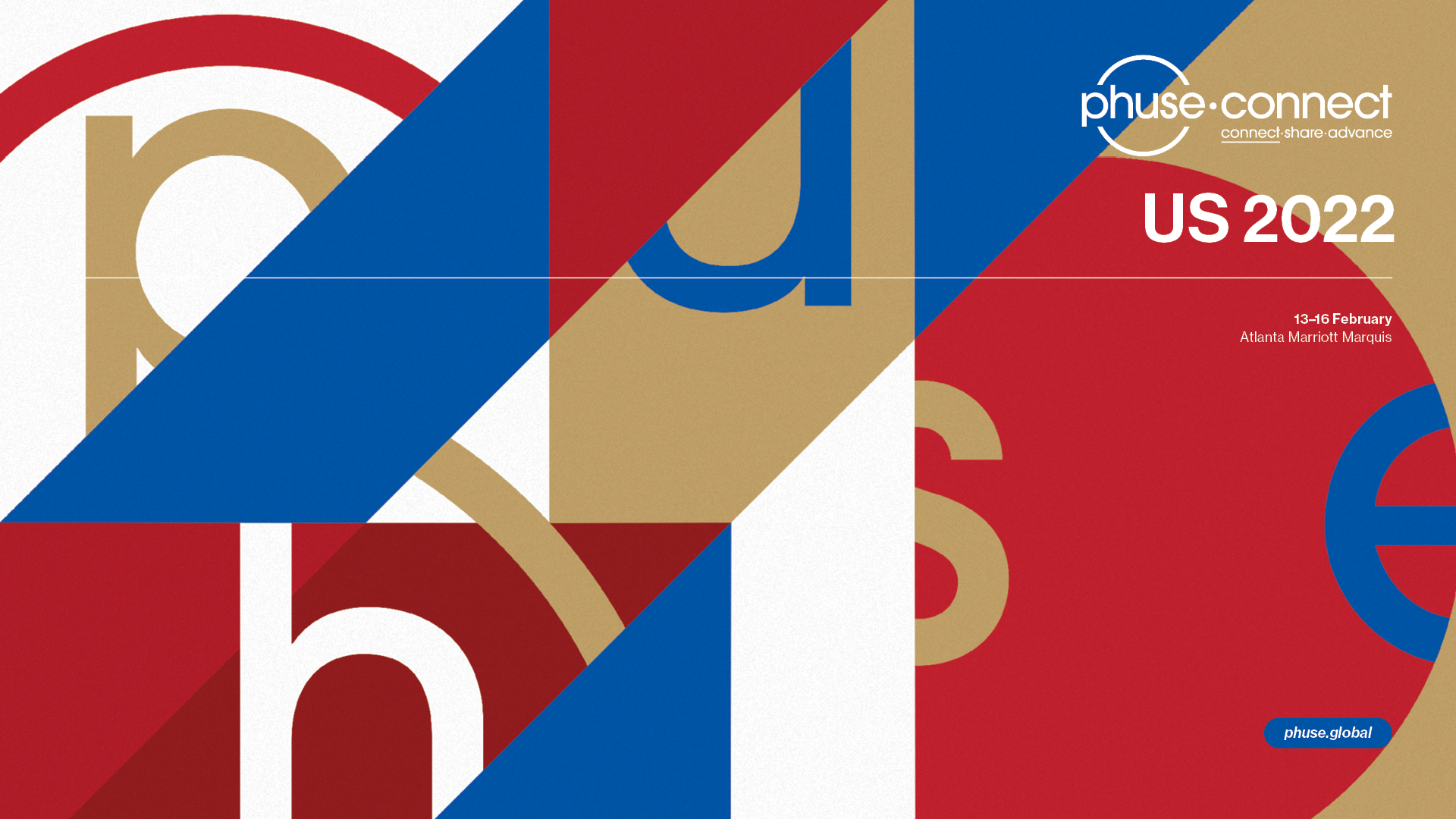
Wednesday 29 September
7am PDT / 10am EDT / 3pm BST / 4pm CET

PHUSE Wednesday Webinar



Wednesday 27 October
7am PDT / 10am EDT / 3pm BST / 4pm CET

phuse·virtual
EU Connect 2021
The Clinical Data Science Conference
15th–19th November
phuse.global


phuse·connect
connect·share·advance
US 2022
13–16 February
Atlanta Marriott Marquis
phuse.global

phuse·advance
connect·share·advance

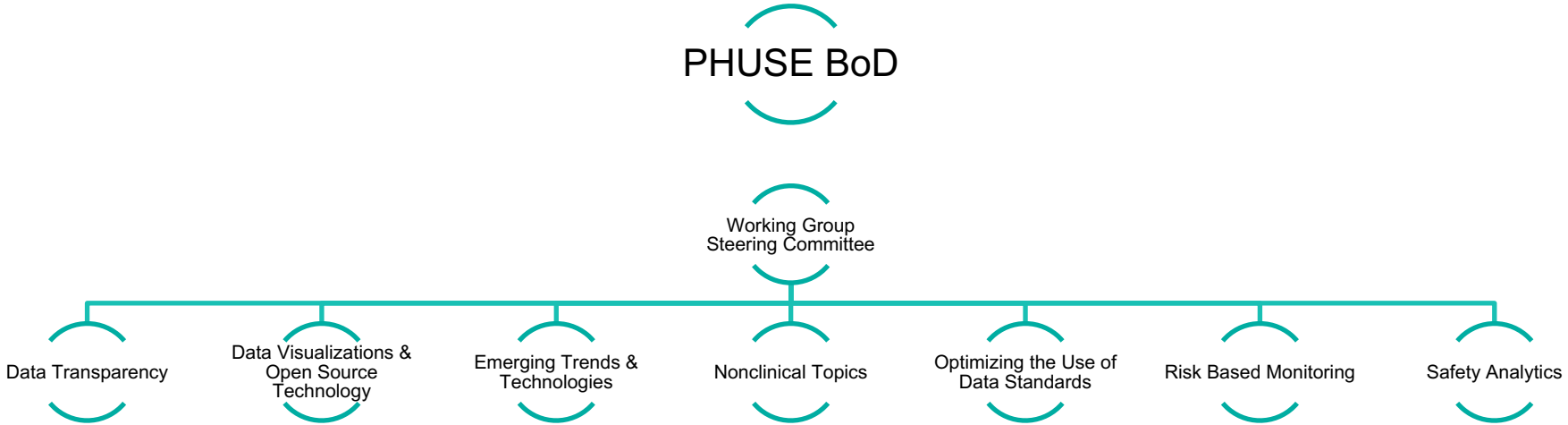
CSS 2022

The Computational Science Symposium

5–7 June
Silver Spring, MD

phuse.global

PHUSE Working Groups



Data Transparency

George Allan

Data Visualization & Open Source Technology

Ellis Hughes

Emerging Trends & Technology

Geoff Low

Non-Clinical Topics

Mark Carfagna

Optimizing the Use of Data Standards

Julie Maynard

Safety Analytics

William (Bill) Palo

Open Source in Clinical Data Science

James Black
Roche

Olivier Leconte
Janssen

Sam Hume
CDISC

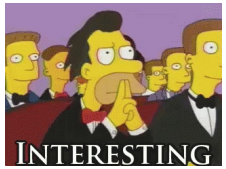
Andy Nicholls
GSK

Mike Stackhouse
Atorus

Mark Lambrecht
SAS

Consumers to contributors

Governance



Why the fuss on ‘ownership and governance’?

FTE / code provenance imbalance

Pharma A

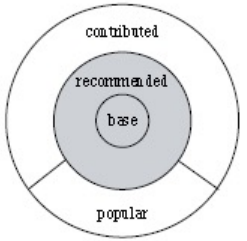
Vs.

Pharma B

```
#' Copy and Change Variable Labels of a \code{data.frame}
#'
#' Relabel a subset of the variables
#'
#' @inheritParams var_labels<-
#' @param ... name-value pairs, where name corresponds to a variable name in
#'   \code{x} and the value to the new variable label
#'
#' @return a copy of \code{x} with changed labels according to \code{...}
#'
#' @export
#'
#' @examples
#' x <- var_relabel(iris, Sepal.Length = "Sepal Length of iris flower")
#' var_labels(x)
#'
var_relabel <- function(x, ...) {
  # todo: make this function more readable / code easier
  stopifnot(is.data.frame(x))
  if (missing(...)) {return(x)}
  dots <- list(...)
  varnames <- names(dots)
  stopifnot(!is.null(varnames))

  map_varnames <- match(varnames, colnames(x))

  if (any(is.na(map_varnames))) {
    stop("variables: ", paste(varnames[is.na(map_varnames)], collapse = ", "), " not found")
  }

  if (any(vapply(dots, Negate(is.character), logical(1)))) {
    stop("all variable labels must be of type character")
  }

  for (i in seq_along(map_varnames)) {
    attr(x[[map_varnames[[i]]]], "label") <- dots[[i]]
  }

  x
}
```

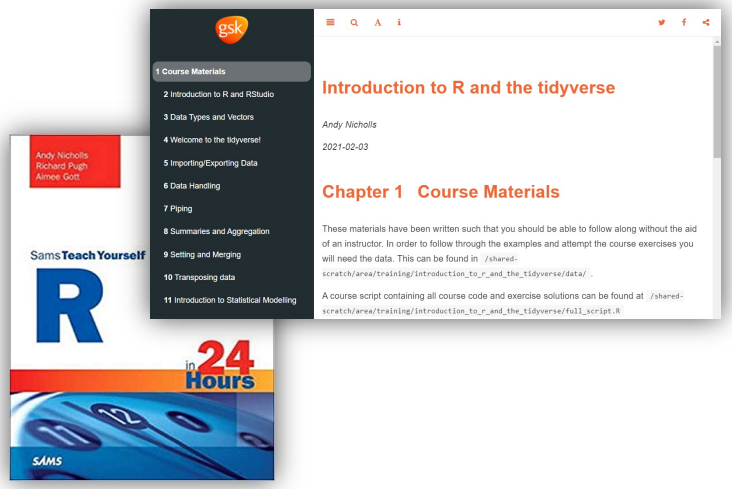
R/summary.R

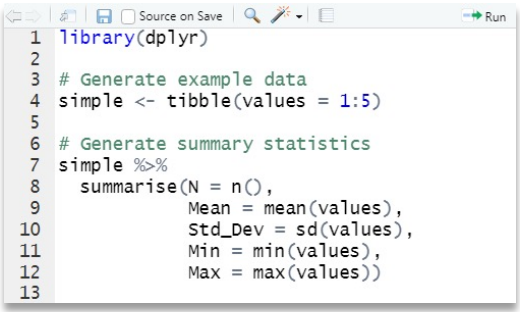
```
@@ -114,7 +114,7 @@ col_paths_summary <- function(x) {
      t(apply(pagdf, 1, function(xi) {
        c(
          indent_string(xi$label, floor(length(xi$path)/2 - 1)),
-         paste(xi$path, collapse=", ")
+         paste(xi$path, collapse = ", ")
        )
      }))
    )
```



*"The computer software used for data management and statistical analysis should be **reliable**, and documentation of **appropriate** software testing procedures should be available"*

ICH E9






```
# A tibble: 1 x 5
      N  Mean Std_Dev   Min   Max
  <int> <dbl>   <dbl> <int> <int>
1     5     3    1.58     1     5
```

Is this reliable?

German, D.M. & Adams, Bram & Hassan, Ahmed E.. (2013). The Evolution of the R Software Ecosystem. Proceedings of the Euromicro Conference on Software Maintenance and Reengineering, CSMR. 243-252. 10.1109/CSMR.2013.33.

Reliable Software

Why do we trust the summary?

- ✓ *"I know the actual answer"*
- ✓ *"It's in the ballpark of what I might expect to see"*

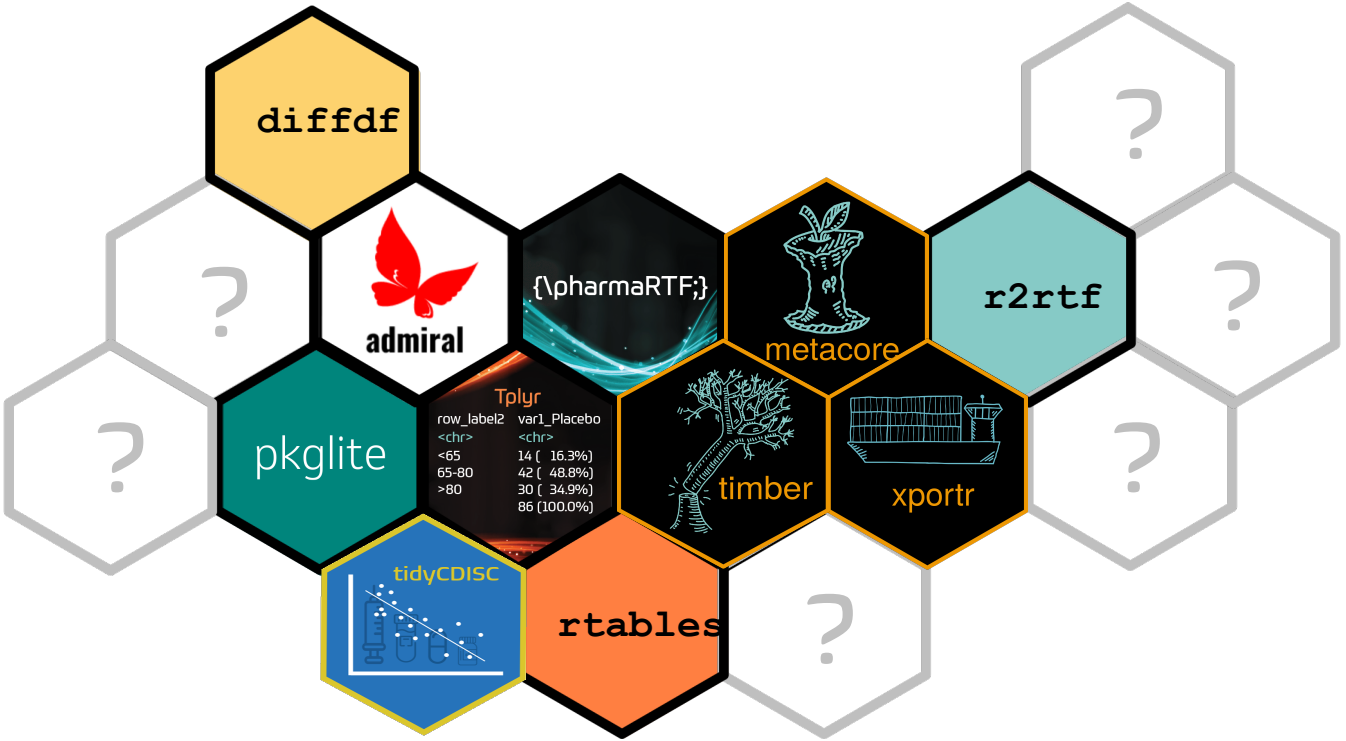
Intuition

- ✓ *"I've used the software before and it did what I expected"*
- ✓ *"When I learnt statistics, I was taught using the software"*
- ✓ *"Many others use the software and it does what they expect"*
- ✓ *"The software is used/cited in statistical literature"*

Community Exposure

- ✓ *"I trust that the software owner develops it using best practice"*
- ✓ *"The software owner provides tests that I can use to verify that it is working"*

Developer SDLC


diffdf
admiral
{\pharmaRTF;}
metacore
r2rtf
pkglite
Tplyr
row_label2 var1_Placebo
<chr> <chr>
<65 14 (16.3%)
65-80 42 (48.8%)
>80 30 (34.9%)
86 (100.0%)
timber
xportr
tidyCDISC
rtables

CDISC Open Source Alliance

Mission: The CDISC Open Source Alliance (COSA) supports, promotes, and sometimes sponsors open-source software development projects that create tools for implementing or developing CDISC standards to drive innovation in the CDISC community.

Additional Information:

- COSA Directory https://cdisc-org.github.io/cosa/
- Email cosa@cdisc.org

Initial Benefits
Directory for open-source projects
Conference sessions & webinars
Hackathons & workshops
Sandbox environment & test data
Community building
Information exchange
Open-source best practices
Licensing guidance

phuse·virtual

10 YEARS PHUSE/FDA

Q&A

PLATFORM CONTRIBUTOR

IBM

The Global Healthcare Data Science Community

Contact Channels

UK +44 1843 609600
office@phuse.global
phuse.global

Social Media

@phusetwitta
/phusebook
/phusetube
/company/phuse