R in Pharma

A tailored approach to converting programmers to R in an industry resistant to change

Kieran Martin and Craig Gower
Introduction

In the clinical setting, many programmers still use SAS

We wanted to change that!
But how?
Tailored Training

Tidyverse

Solving problems from day to day work and using internal data

Functional programming!

Unit testing and assertions
Tailored Training

- Tidyverse
- Solving problems from day to day work and using internal data
- Functional programming!
- Unit testing and assertions
R packages

Developing in house packages to meet specific goals

- Tables and graphs matching standards
- Standardised Shiny apps which have similar looks and feel
- Convenience packages (data access and use of HPC)
Available on CRAN

Created by Craig Gower and Kieran Martin

github: https://github.com/gowerc/diffdf

Aims to provide in depth information on differences on two datasets intended to be identical

Similar to "Proc Compare" in SAS
Available on CRAN

Created by Craig Gower and Kieran Martin

github: https://github.com/gowerc/diffdf

Aims to provide in depth information on differences on two datasets intended to be identical

Similar to "Proc Compare" in SAS
diffdf::diffdf - Introduction

Syntax

diffdf(
    base = ,           # Base data.frame
    compare = ,        # Comparison data.frame
    keys = ,           # Key variables that define a unique row
    ...                # More arguments here, not included for brevity!
)

suppressMessages(library(dplyr))
library(diffdf)
dm <- haven::read_sas("./data/dm.sas7bdat")
vs <- haven::read_sas("./data/vs.sas7bdat")

dm2 <- dm
dm2$AGE[5] <- NA

x <- diffdf(dm, dm2)

## Warning in diffdf(dm, dm2):
## Not all Values Compared Equal
diffdf::diffdf

Examples

Whenever diffdf finds differences, it saves them in the output object. Only the differences found will be stored

```r	names(x)
```

## [1] "NumDiff" "VarDiff_AGE"

```r
x$VarDiff_AGE
```

```r
# # A tibble: 1 x 4
# # VARIABLE ..ROWNUMBER.. BASECOMPARE
# * <chr> <int> <dbl> <dbl>
# 1 AGE 5 45 NA
**diffdf::diffdf**

**Attributes**

`diffdf()` can see difference in attributes.

```r
dm2 <- dm
attr(dm2$SEX, "label") <- "Gender"
diffdf(dm, dm2, suppress_warnings = TRUE)
```

```r
## Differences found between the objects!
##
## A summary is given below.
##
## There are columns in BASE and COMPARE with differing attributes !!
## All rows are shown in table below
##
## ================================================
## VARIABLE ATTR_NAME VALUES.BASE VALUES.COMP
## -----------------------------------------------
## SEX label Sex Gender
```
## Differences found between the objects!
##
## A summary is given below.
##
## Not all Values Compared Equal
## All rows are shown in table below
##
## ==-----------------------------==
## | Variable | No of Differences |
## ==-----------------------------==
## | AGE | 1 |
## ==-----------------------------==

## All rows are shown in table below
##
## ===============================
## | VARIABLE | USUBJID | BASE | COMPARE |
## ===============================
## | AGE | SHH4429G-S19914-16104 | 45 | <NA> |
Conclusions

Ongoing efforts of conversion

Tidyverse is a great way to convert people to R

Beauty of R: if there is a gap, you can make your own fixes!

Custom packages doing small jobs are great!
Doing now what patients need next