

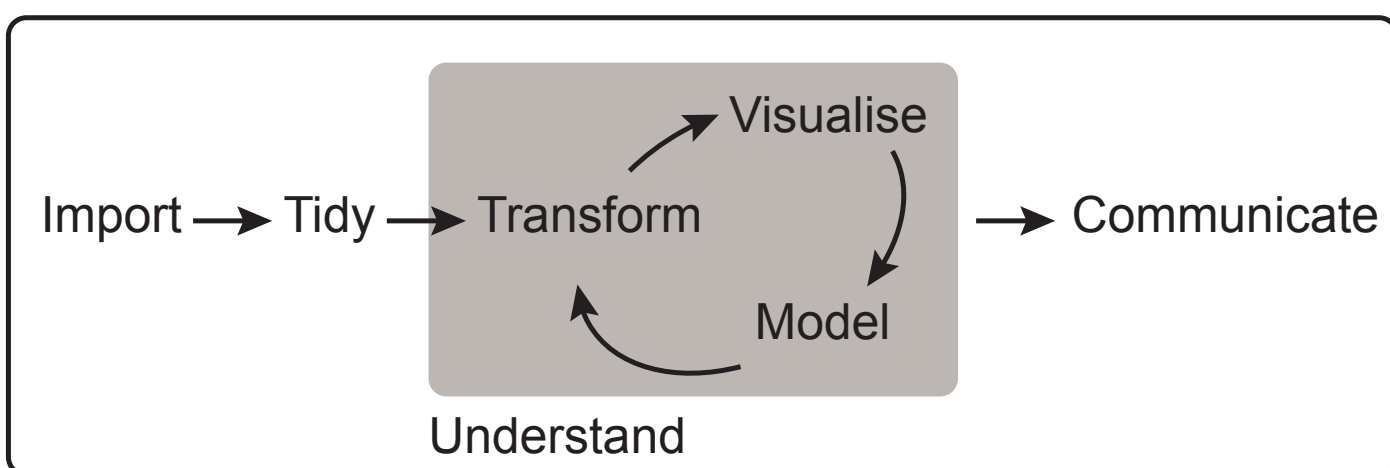
R for SDTM and ADaM Data

Yilong Zhang; Changhong Shi;
Lei Xu; Zhen Zeng; Gregory Golm

Merck Research Laboratories,
Merck & Co., Inc., Kenilworth, NJ, USA

BACKGROUND

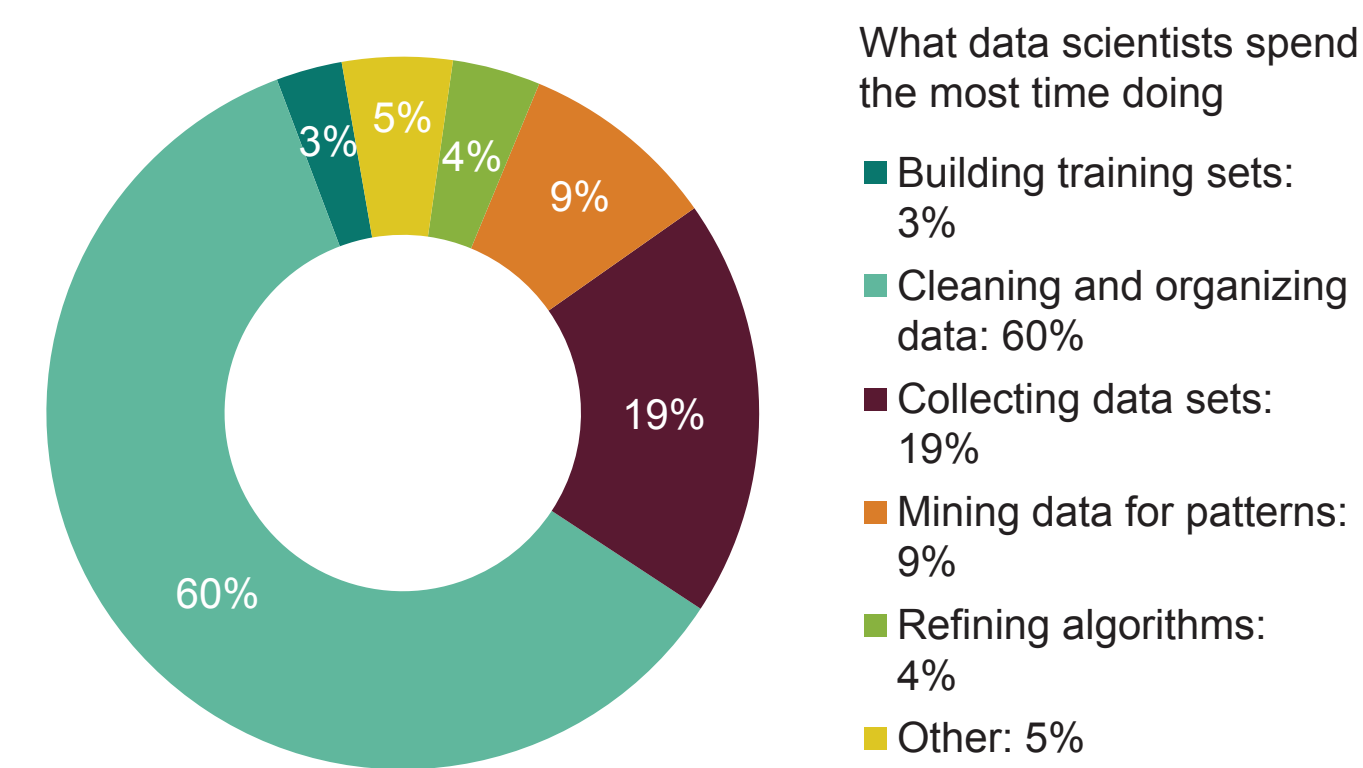
A General Data Analysis Project Structure



- Clinical statisticians involved in every step, from importing data to communicating the results to clinicians
- Every step is critical to deliver a high-quality clinical trial report

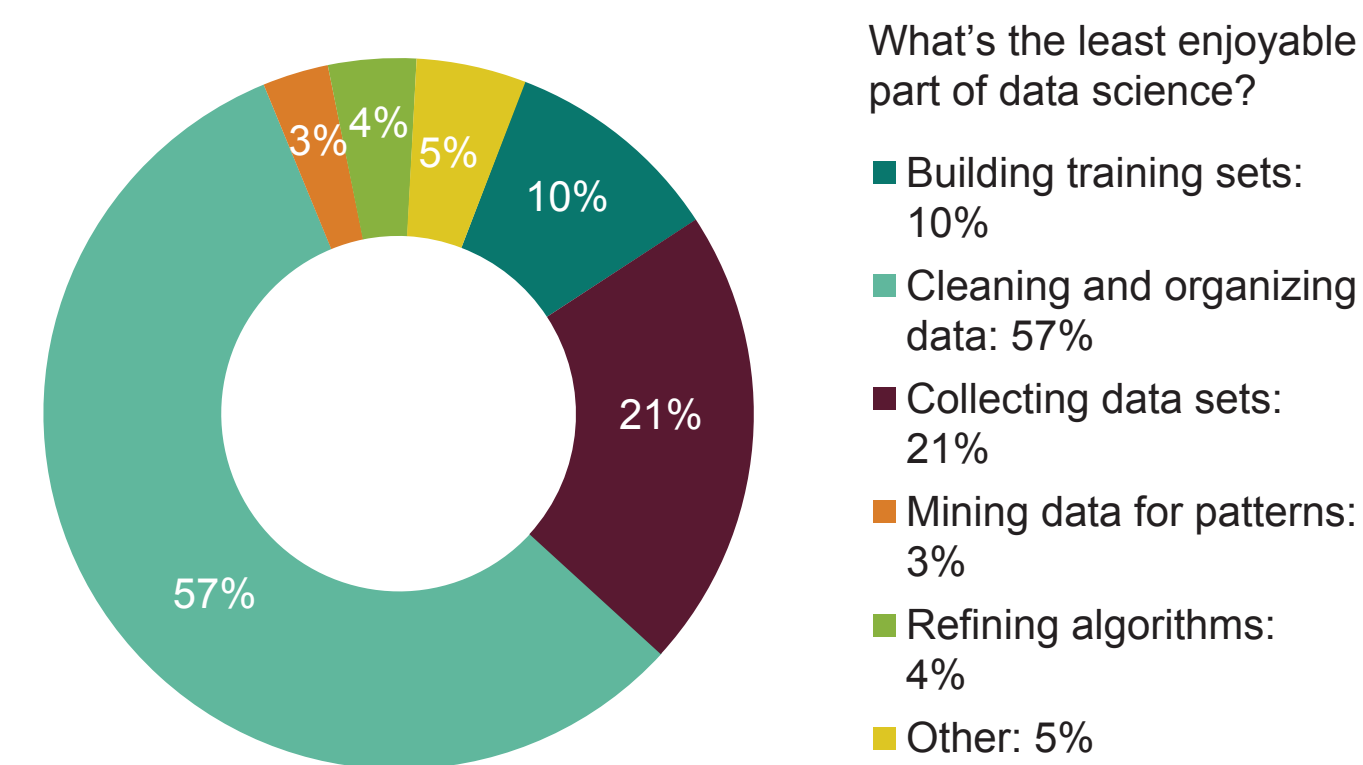
Adapted from: <http://r4ds.had.co.nz/introduction.html>

Data Cleaning Is Time Consuming



Adapted from: <https://www.forbes.com/sites/gilpress/2016/03/23/data-preparation-most-time-consuming-least-enjoyable-data-science-task-survey-says/#360ccea66f63>

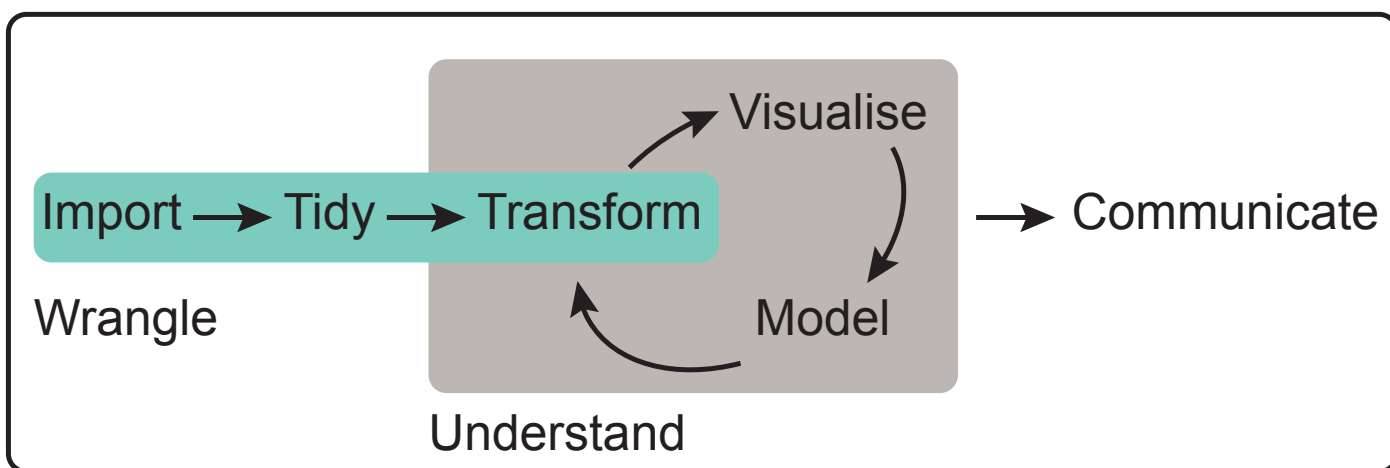
Data Cleaning Is Least Enjoyable



Adapted from: <https://www.forbes.com/sites/gilpress/2016/03/23/data-preparation-most-time-consuming-least-enjoyable-data-science-task-survey-says/#360ccea66f63>

SAS and R for Clinical Statisticians/Programmers

- SDTM and ADaM are standards for data submission to agencies
- SAS is the mainstream in preparing SDTM and ADaM data and generating deliverable table, figure and listing
- R can help to clean, explore, and visualize the data (internal usage)

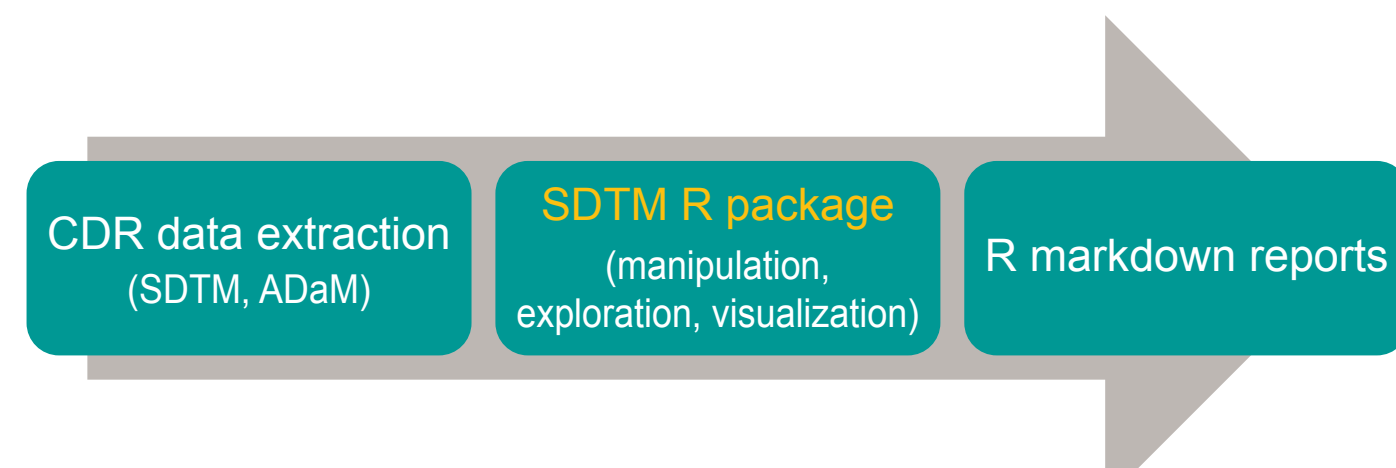


Program

Adapted from: <http://r4ds.had.co.nz/wrangle-intro.html>

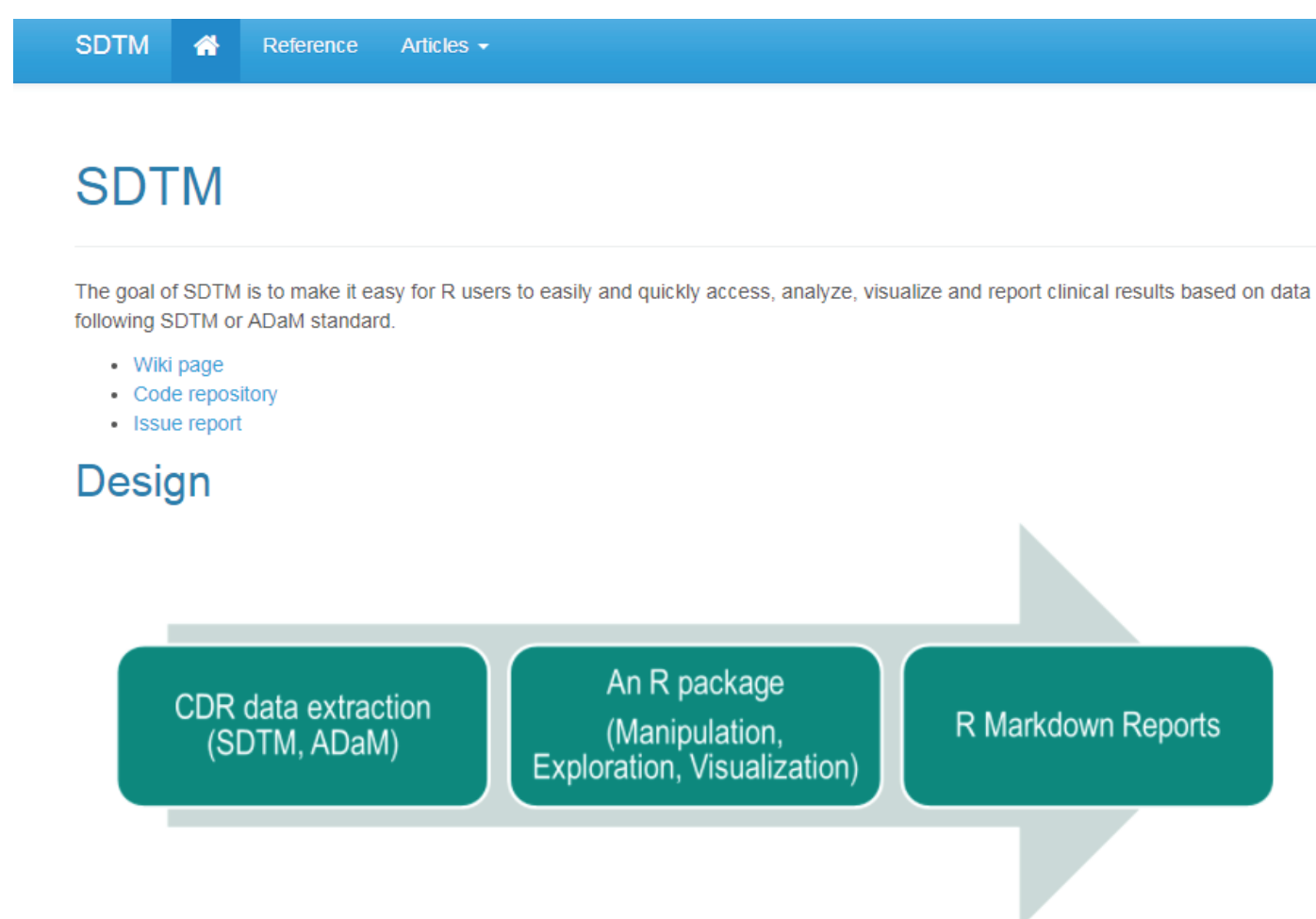
IMPLEMENTATION

Access SDTM and ADaM Data in R



SDTM R Package

- An R package to access SDTM and ADaM standard data:
 - 50 functions to input each data domain
 - 15 functions for data manipulation
 - 3 functions for data validation
 - 7 functions to set up working environment



Installation

- Download the SDTM package in the code repository
- Unzip the file
- Open SDTM.Rproj by Rstudio
- In Rstudio console, run: `devtools::install()`

RESULTS

Data Extraction Time Comparison

- Running time comparison between `SDTM::get_data` and `haven::read_sas`
- `SDTM::get_data` in our SDTM package is >35X faster in large data set
- `SDTM::get_data` only retrieves commonly used variables

Domain	Number of Rows	Number of Columns	get_data† (seconds)	read_sas† (seconds)
AE	583	118	0.3	0.7
LB	67,357	68	0.7	28.3
CF	53,113	127	1.1	49.7
EX	191,021	66	1.5	52.1

†Using the default setting of the function

Clinical Statisticians With R Experience

- Document review
 - Review and validate SAS code to create ADaM data and TFLs
 - Review CSR/eDMC TFLs
- SAP/sSAP development
 - Sample size calculation
 - Explore data sets
 - Post-hoc analyses
- Statistical methods evaluation

Explore SAS Code With R

- Quickly identify SAS code:
 - Check_sas_all: A project-specified function used to search all SAS codes by blinded line "study <- "P002"

```
> library(SDTM)
> study <- "P002"
> check_sas_all(study, "DISCONTINUED")
[1] "-----"
[1] "File name: calllog0disposition.sas"
[1] "...line 36 :.dsdecod=tranwrd(dsdecod,\"DISCONTINUED\", \"DISCONTINUED\");"
[1] "...line 37 :.dsterm=tranwrd(dsterm,\"DISCONTINUED\", \"DISCONTINUED\");"
[1] "-----"
[1] "File name: ldamk20.sas"
[1] "...line 236 :.dsdecod=tranwrd(dsdecod,\"DISCONTINUED\", \"DISCONTINUED\");"
[1] "...line 237 :.dsterm=tranwrd(dsterm,\"DISCONTINUED\", \"DISCONTINUED\");"
```

Explore SDTM and ADaM Data With R Markdown

```
1 title: "USUBJID and SUBJID"
2 date: "r Sys.Date()"
3 output:
4   html_document:
5     toc: yes
6     toc_depth: 3
7     toc_float:
8       collapsed: yes
9   ---
10
11
12 [r setup, eval = TRUE, message = FALSE, warning=FALSE, include = FALSE]
13 source(paste0(devtools::as.package(.))$path, "/.Rprofile") # Load Rprofile for this
14
15
16 # P001 3.1.1 data
17
18 [r]
19 setwd_sdtm("P001")
20 root_path <- getwd()
21 extraction_date <- as.Date(get_data_extraction_date())
22
23
24 ## Analysis Data Information
25 * Data extraction date: r extraction_date
26 * Data Path: r root_path
27
28 [r]
29 setwd_sdtm("P001")
30 de <- get_data()
31
32
33 ## Explore SUBJID
34
35 [r]
36 de %>% select(-COUNTRY) %>% show_db()
37
38
39
```

Explore SDTM and ADaM Data With R Markdown

MK9999-P001 Home articles Reference

USUBJID and SUBJID

2017-10-25

P001 3.1.1 data

Analysis Data Information

- Data extraction date: 2017-10-10
- Data Path: I:\barbar-test\mk9431-diabetes\p845\data\sdmplus

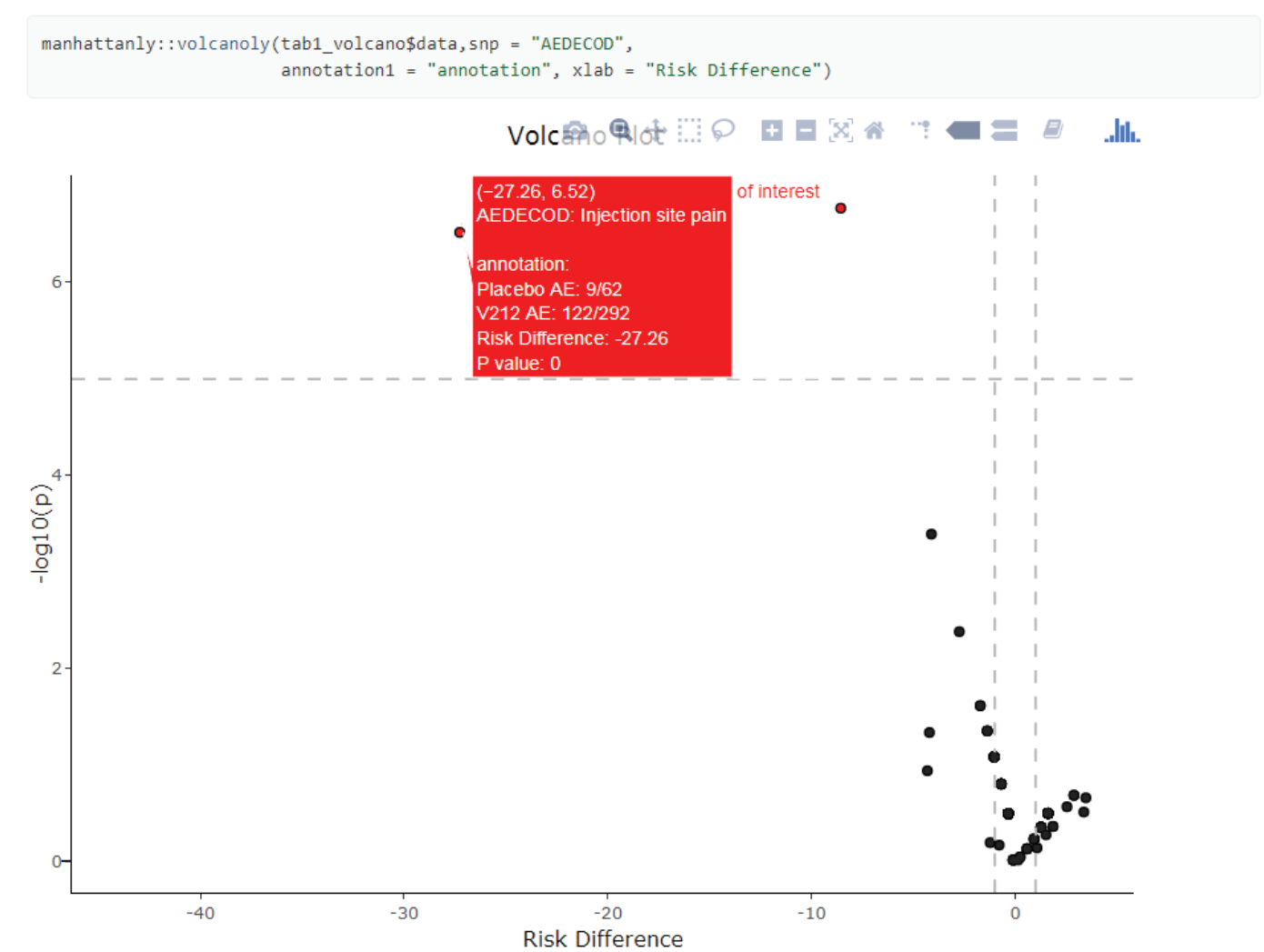
Explore SUBJID

Column visibility	Excel	Search:					
USUBJID	SUBJID	RFSTDTC	RFENDTC	AGE	SEX	RACE	
All						All	
1	MK9999_000100001	111111	2016-09-19	2017-04-12	63	F	MULTIPLE
2	MK9999_000100002	222222	2016-11-15	2017-06-07	56	M	WHITE
3	MK9999_000300001	333333	2016-08-04	2017-03-07	52	M	WHITE
4	MK9999_000300002	444444	2016-09-21	2017-05-02	57	M	WHITE
5	MK9999_000300003	555555	2016-09-22	2017-04-26	48	M	WHITE
6	MK9999_000400001	666666	2017-04-24		58	M	WHITE
7	MK9999_000400002	777777	2017-05-08		61	M	WHITE
8	MK9999_000400003	888888	2017-05-16		65	M	WHITE
9	MK9999_000500001	999999	2017-02-21	2017-03-27	55	M	WHITE
10	MK9999_010000003	111112	2017-01-31	2017-08-29	56	M	WHITE

Showing 1 to 11 of 745 entries

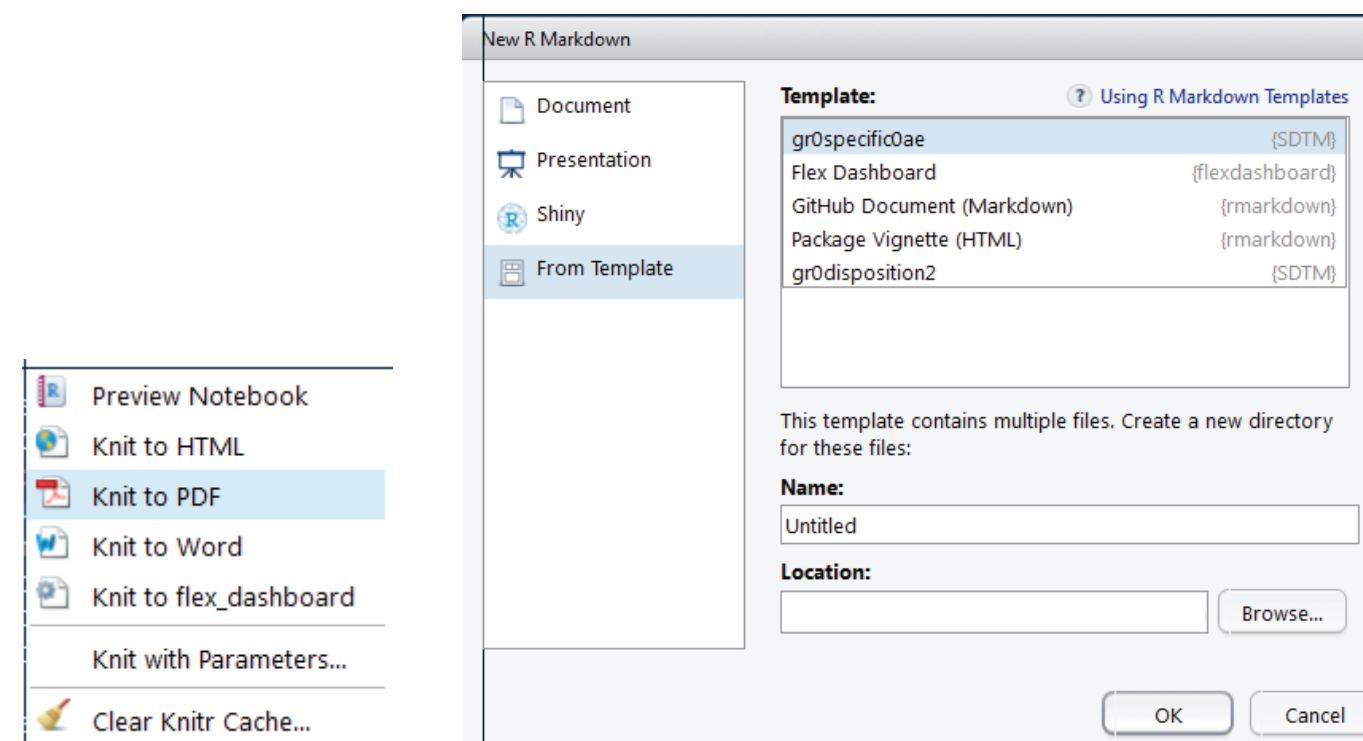
Showing 1 to 11 of 740 entries

SDTM:AE Data Visualization With R (Volcano Plot)



SDTM Package Features

- Easy to set up:** One line of code to install the R package
 - `devtools::install("path_to_package")`
- Easy to use:** The R markdown templates are ready to use when we create a new R markdown from templates
- Easy to manage:** One set of code creates a report in multiple formats



Build Statistical Monitoring Life Cycle With R and Beyond (pkgdown)

MK9999-001 Home articles Reference

Articles version 0.1.0

Data Issues

- EPOCH Validation
- EX Records Validation
- HA Records Validation with Linked AE and LBFSG Records
- retec Records Validation

Exploratory

- Efficacy Endpoints A1C and FPG Derivation
- Efficacy HA Endpoints
- Exploratory Subjects
- USUBJID and SUBJID

Developed by Yilong Zhang.

Site built with pkgdown.

CONCLUSION

- R works well with SAS to support clinical data review
- SDTM R package builds a foundation to help users access SDTM and ADaM data efficiently
- Data and code access: Same as SAS programs in the analysis and report folder