R Data Packages of Multi-Reader Multi-Case Studies and Simulation Tools to Support the Development of Reader Performance Evaluation Methods

Si Wen, Qi Gong, Brandon Gallas

FDA/CDRH/OSEL Division of Imaging, Diagnostics, and Software Reliability

Outline

- MRMC studies and analysis
  - Multi-reader, Multi-case
- ROC: Receiver Operating Characteristic curves
- R Data packages
  - What and how
- Show examples of MRMC data analysis

- Why
  - Disseminate our research data, methods, and tools
  - Enables other investigators to develop methods and tools

Background

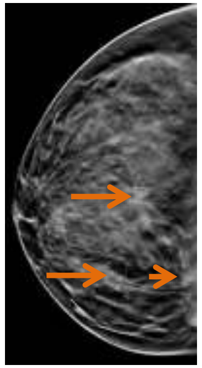
- Medical device evaluation
  - Assessing new device performance


(1)

V.S.

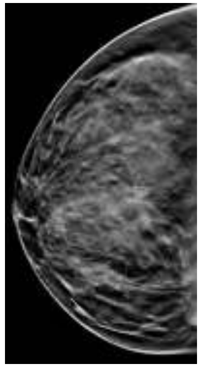

(2)

  - Assessing computer aided diagnosis devices


(3)

AI detected potential lesion

V.S.


(4)

Comparing two modalities in a medical task

Image source: (1) https://www.independentimaging.com/digital-x-rays-vs-traditional-x-rays/ (2) https://www.featurepics.com/online/Doctor-Holding-Ray-Picture298823.aspx
(3)(4) Morra, Lia, et al. "Breast cancer: computer-aided detection with digital breast tomosynthesis." *Radiology* 277.1 (2015): 56-63. https://www.ncbi.nlm.nih.gov/pubmed/25961633

Motivation

- Design a study to compare two modalities for a medical task
  - Ask clinicians to review cases under the two different modalities
- Random factors in the study:
  - **Cases**: randomly select samples from intended case population
  - **Readers**: randomly select samples from intended reader population

- **Multi-Reader Multi-Case (MRMC) Study**
  - Each case is reviewed by multiple clinicians/readers
  - Each reader review multiple cases
- **MRMC analysis**
  - Account for reader and case variability
  - Generalize results to the population of readers and cases

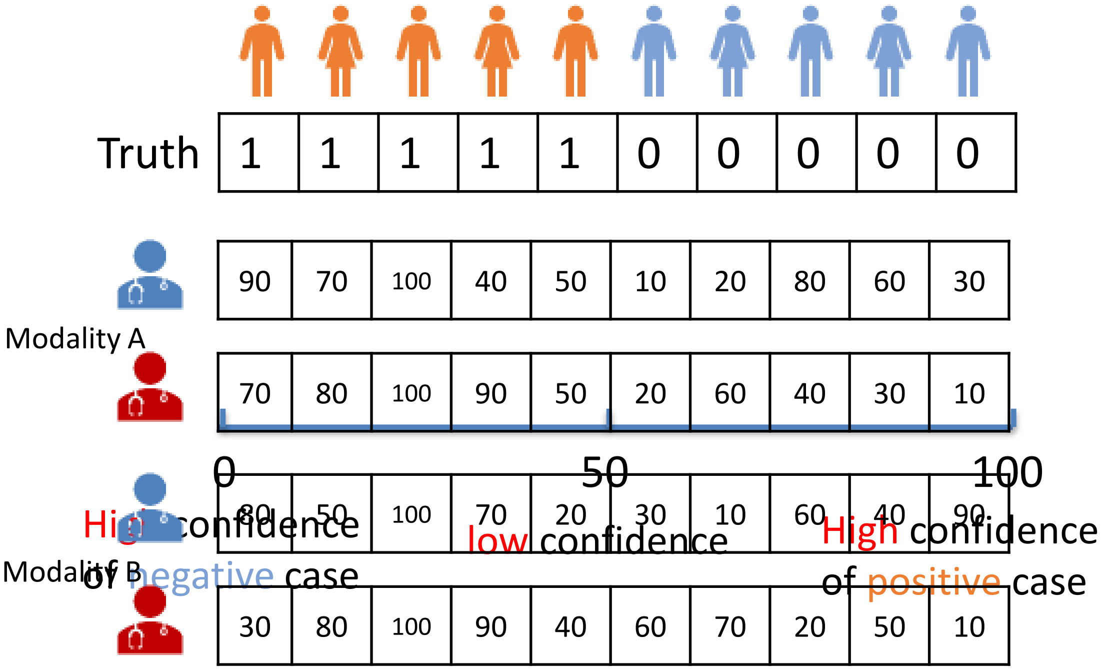
Multi-Reader Multi-Case Studies

- Application
  - Comparing manual scores from clinicians by using different medical device
  - Comparing machine learning algorithm result with manual scores
  - Comparing the results from different machine learning algorithms

Medical Imaging and other technologies

Multi-Reader Multi-Case Studies

- Data Structure



Truth	1	1	1	1	1	0	0	0	0	0
	90	70	100	40	50	10	20	80	60	30
	70	80	100	90	50	20	60	40	30	10
	80	50	100	70	20	30	10	60	40	90
	30	80	100	90	40	60	70	20	50	10

Reader study result

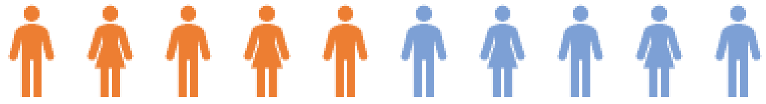
readerID	caseID	modalityID	confidenceScore
reader1	case1	A	90
reader1	case2	A	70
reader1	case3	A	100
⋮	⋮	⋮	⋮
reader2	case1	A	70
reader2	case2	A	80
reader2	case3	A	100
⋮	⋮	⋮	⋮
reader1	case1	B	80
reader1	case2	B	50
reader1	case3	B	100
⋮	⋮	⋮	⋮
reader2	case7	B	70
reader2	case8	B	20
reader2	case9	B	50
reader2	case10	B	10

Ground truth

caseID	Truth
case1	1
case2	1
case3	1
case4	1
case5	1
case6	0
case7	0
case8	0
case9	0
case10	0

Multi-Reader Multi-Case Studies

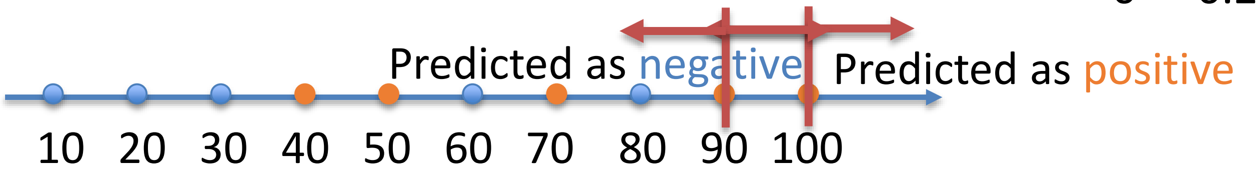
- Receiver Operating Characteristic (ROC) analysis

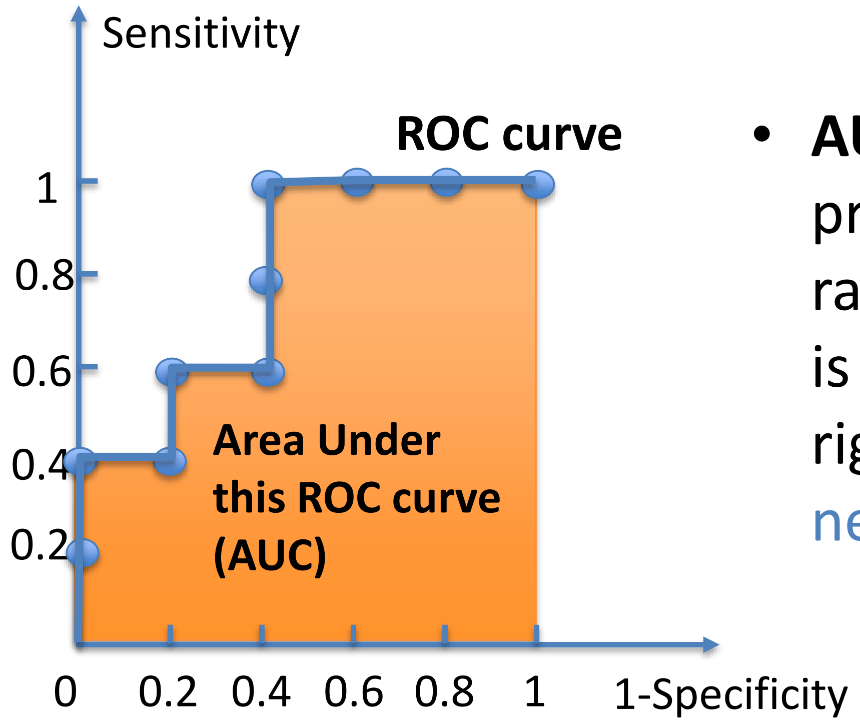


Truth	1	1	1	1	1	0	0	0	0	0
	90	70	100	40	50	10	20	80	60	30

t=90	Truth=1	Truth=0
Score≥90	2	0
Score<90	3	5

Sensitivity = 2/5 = 0.4 Specificity = 5/5=1



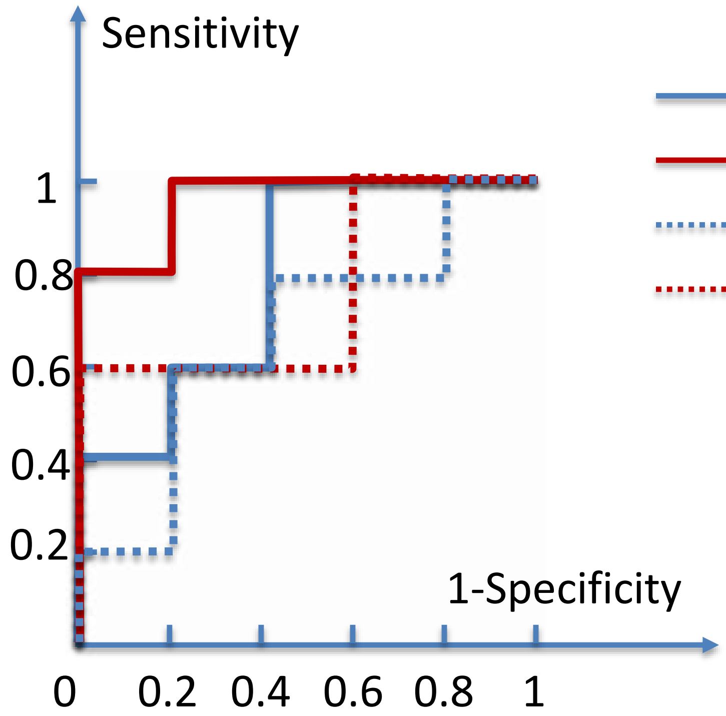


- **AUC** represents the probability that a random positive case is positioned to the right(>) of a random negative case

Multi-Reader Multi-Case Studies

- MRMC ROC analysis for two modalities

Truth	1	1	1	1	1	0	0	0	0	0
Modality A	90	70	100	40	50	10	20	80	60	30
	70	80	100	90	50	20	60	40	30	10
Modality B	80	50	100	70	20	30	10	60	40	90
	30	80	100	90	40	60	70	20	50	10



Reader-averaged AUC in modality A

$$= \frac{1}{2}(0.80 + 0.96) = \mathbf{0.88}$$

Reader-averaged AUC in modality B

$$= \frac{1}{2}(0.68 + 0.76) = \mathbf{0.72}$$

- **Difference of reader-averaged AUC =** $\mathbf{0.88 - 0.72 = 0.16}$
- **Variance of difference of reader-averaged AUC - hard**

Multi-Reader Multi-Case Studies

Difficulties	Solutions
Time consuming and labor intensive not much data is publicly available	**MRMC data storage and distribution** **R data packages**
Complicated correlation structure when computing variance of reader-averaged AUC	**MRMC performance analysis** **iMRMC R package**

R data package for MRMC study

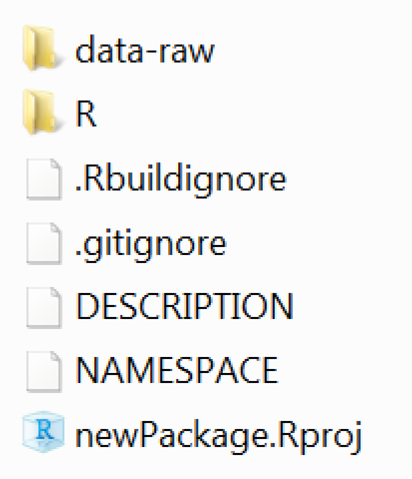
- To store and **distribute** data along with its **documentation**
  - Well-organized documentation: description, format, source, etc.
  - Public available: CRAN(The Comprehensive R Archive Network https://cran.r-project.org/) or GitHub (https://github.com/)

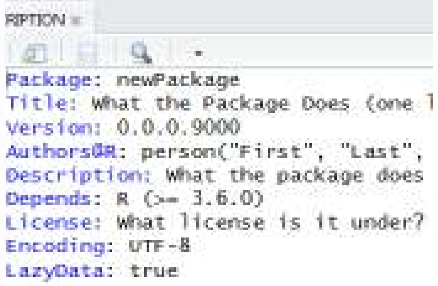
- To provide **use case** for existing and future functions/methods
  - Pre-processed reproducible data that can be easily accessed in R

- To help establishing **future studies**
  - Acting as a baseline for similar study
  - Providing information for sizing future study in this area

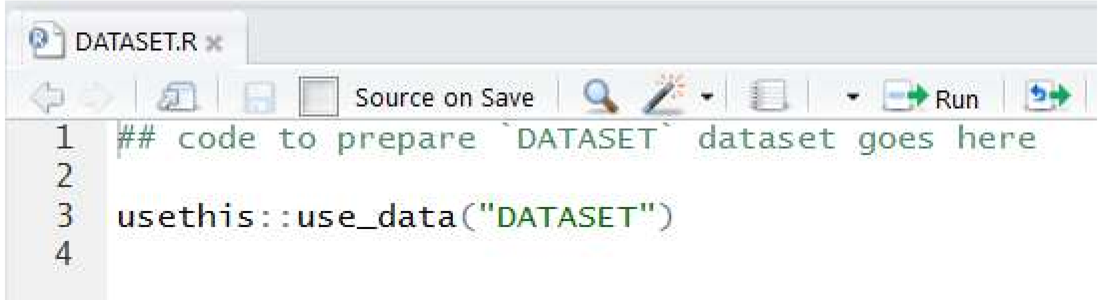
Building up an R data package

- Tool: devtools (https://www.rstudio.com/products/rpackages/devtools/)
- Steps:
  - devtools::create() create the package skeleton
  - devtools::use_data_raw() create data-raw/


data-raw
R
.Rbuildignore
.gitignore
DESCRIPTION
NAMESPACE
newPackage.Rproj



```
Package: newPackage
Title: What the Package Does (one
Version: 0.0.0.9000
Authors@R: person("First", "Last",
Description: What the package does
Depends: R (>= 3.6.0)
License: What license is it under?
Encoding: UTF-8
LazyData: true
```

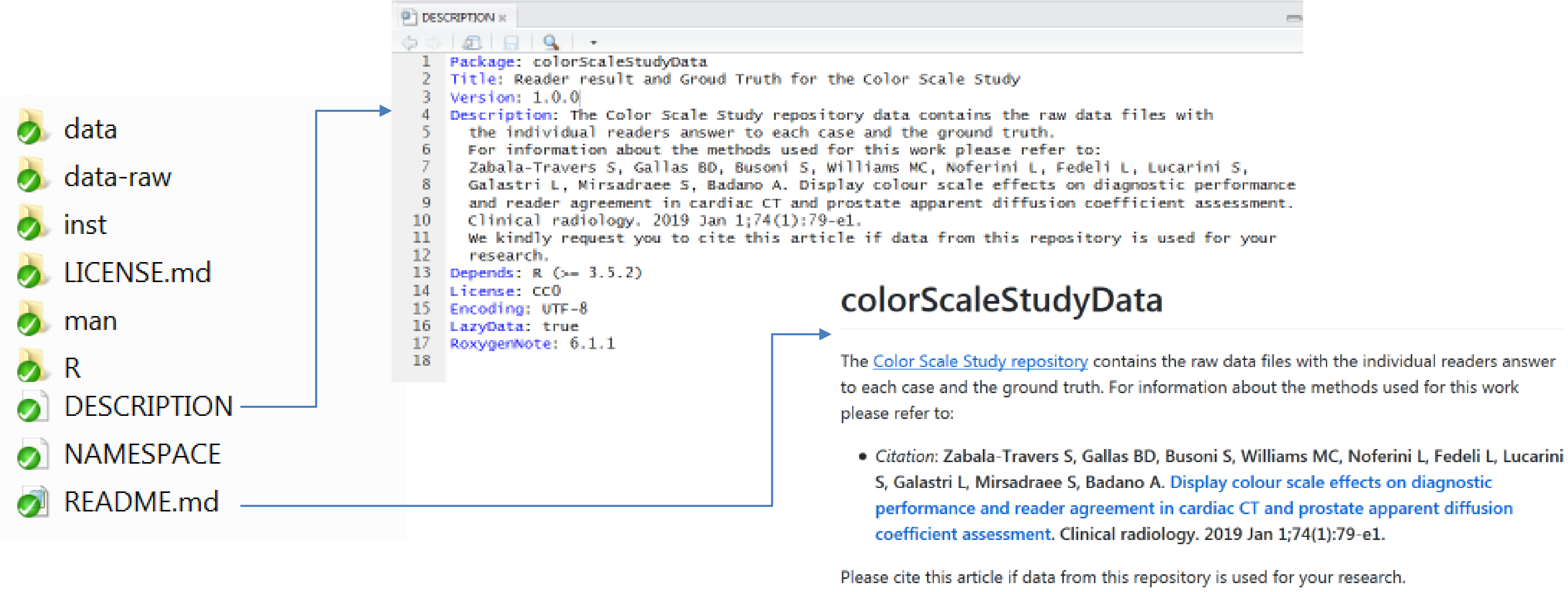


```
## code to prepare `DATASET` dataset goes here

usethis::use_data("DATASET")

```

  - Put data preprocessing script into .R file, which reads in raw data and at calls usethis::use_data() to save .rda formatted data files in data/

MRMC R data packages

- Structure
  - colorScaleStudyData (https://github.com/DIDSR/colorScaleStudyData)



```
Package: colorScaleStudyData
Title: Reader result and Groud Truth for the Color Scale Study
Version: 1.0.0
Description: The Color Scale Study repository data contains the raw data files with
  the individual readers answer to each case and the ground truth.
  For information about the methods used for this work please refer to:
  Zabala-Travers S, Gallas BD, Busoni S, Williams MC, Noferini L, Fedeli L, Lucarini S,
  Galastri L, Mirsadraee S, Badano A. Display colour scale effects on diagnostic performance
  and reader agreement in cardiac CT and prostate apparent diffusion coefficient assessment.
  Clinical radiology. 2019 Jan 1;74(1):79-e1.
  We kindly request you to cite this article if data from this repository is used for your
  research.
Depends: R (>= 3.5.2)
License: CC0
Encoding: UTF-8
LazyData: true
RoxygenNote: 6.1.1
```

colorScaleStudyData

The Color Scale Study repository contains the raw data files with the individual readers answer to each case and the ground truth. For information about the methods used for this work please refer to:

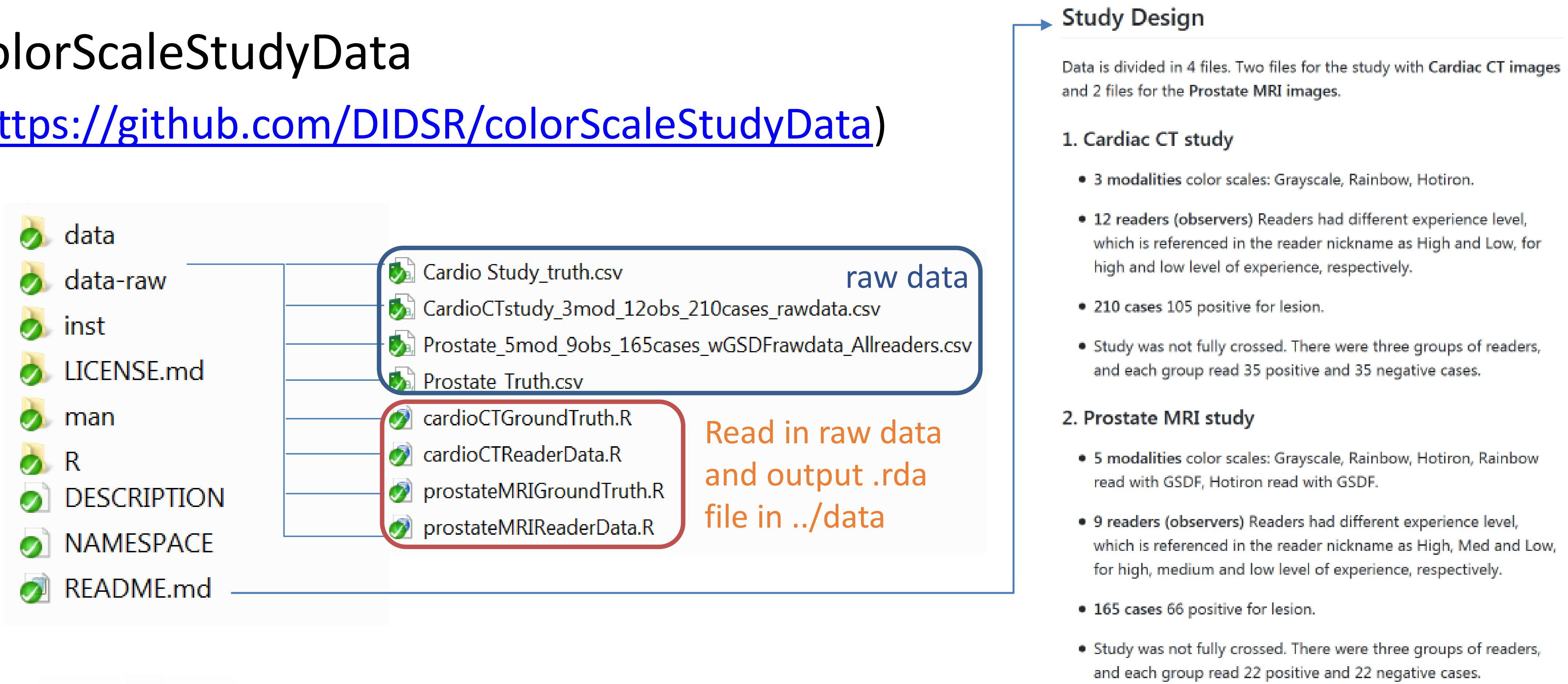
- *Citation*: Zabala-Travers S, Gallas BD, Busoni S, Williams MC, Noferini L, Fedeli L, Lucarini S, Galastri L, Mirsadraee S, Badano A. Display colour scale effects on diagnostic performance and reader agreement in cardiac CT and prostate apparent diffusion coefficient assessment. Clinical radiology. 2019 Jan 1;74(1):79-e1.

Please cite this article if data from this repository is used for your research.

- **Description** (required as a R package) contains the package name, title, and description
- **Readm.md** (for GitHub) contains the instruction information of the package

MRMC R data packages

- Structure
  - colorScaleStudyData (https://github.com/DIDSR/colorScaleStudyData)



data
data-raw
inst
LICENSE.md
man
R
DESCRIPTION
NAMESPACE
README.md

raw data

Cardio Study_truth.csv
CardioCTstudy_3mod_12obs_210cases_rawdata.csv
Prostate_5mod_9obs_165cases_wGSDFrawdata_Allreaders.csv
Prostate_Truth.csv

cardioCTGroundTruth.R
cardioCTReaderData.R
prostateMRIGroundTruth.R
prostateMRIReaderData.R

Read in raw data and output .rda file in ../data

Study Design

Data is divided in 4 files. Two files for the study with **Cardiac CT images** and 2 files for the **Prostate MRI images**.

1. Cardiac CT study

- **3 modalities** color scales: Grayscale, Rainbow, Hotiron.
- **12 readers (observers)** Readers had different experience level, which is referenced in the reader nickname as High and Low, for high and low level of experience, respectively.
- **210 cases** 105 positive for lesion.
- Study was not fully crossed. There were three groups of readers, and each group read 35 positive and 35 negative cases.

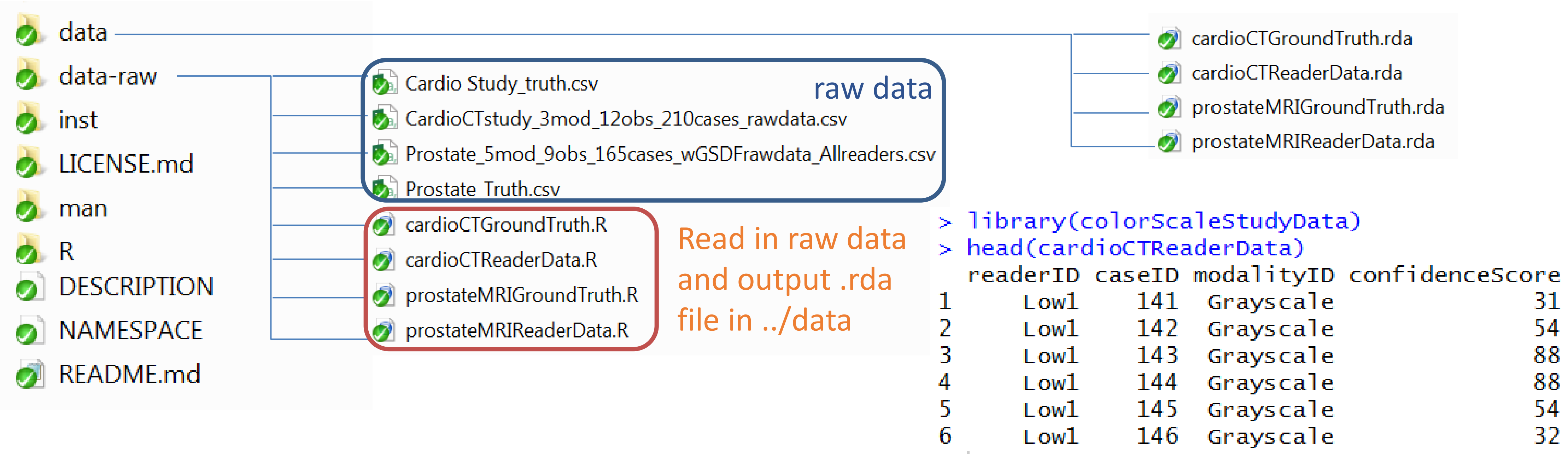
2. Prostate MRI study

- **5 modalities** color scales: Grayscale, Rainbow, Hotiron, Rainbow read with GSDF, Hotiron read with GSDF.
- **9 readers (observers)** Readers had different experience level, which is referenced in the reader nickname as High, Med and Low, for high, medium and low level of experience, respectively.
- **165 cases** 66 positive for lesion.
- Study was not fully crossed. There were three groups of readers, and each group read 22 positive and 22 negative cases.

- **Data-raw** contains the raw dataset and functions to output the R format data

MRMC R data packages

- Structure
  - colorScaleStudyData (https://github.com/DIDSR/colorScaleStudyData)



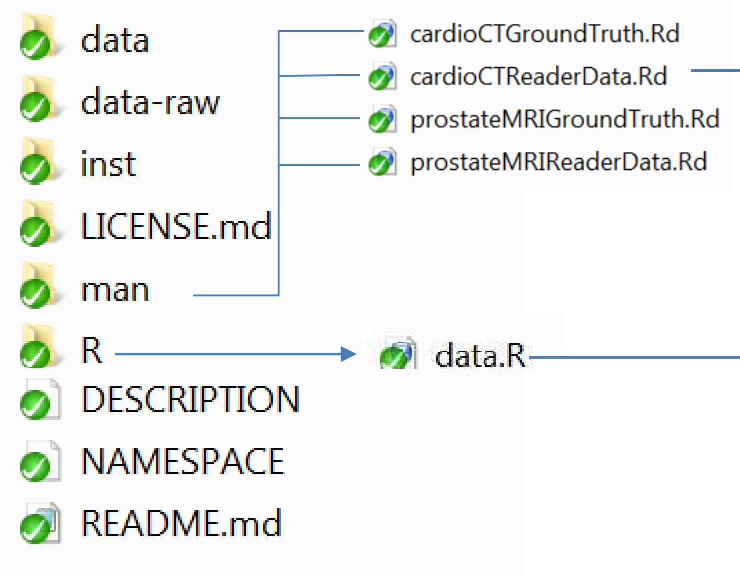
Read in raw data and output .rda file in ../data

```
> library(colorScaleStudyData)
> head(cardioCTReaderData)
  readerID caseID modalityID confidenceScore
1     Low1    141  Grayscale              31
2     Low1    142  Grayscale              54
3     Low1    143  Grayscale              88
4     Low1    144  Grayscale              88
5     Low1    145  Grayscale              54
6     Low1    146  Grayscale              32
```

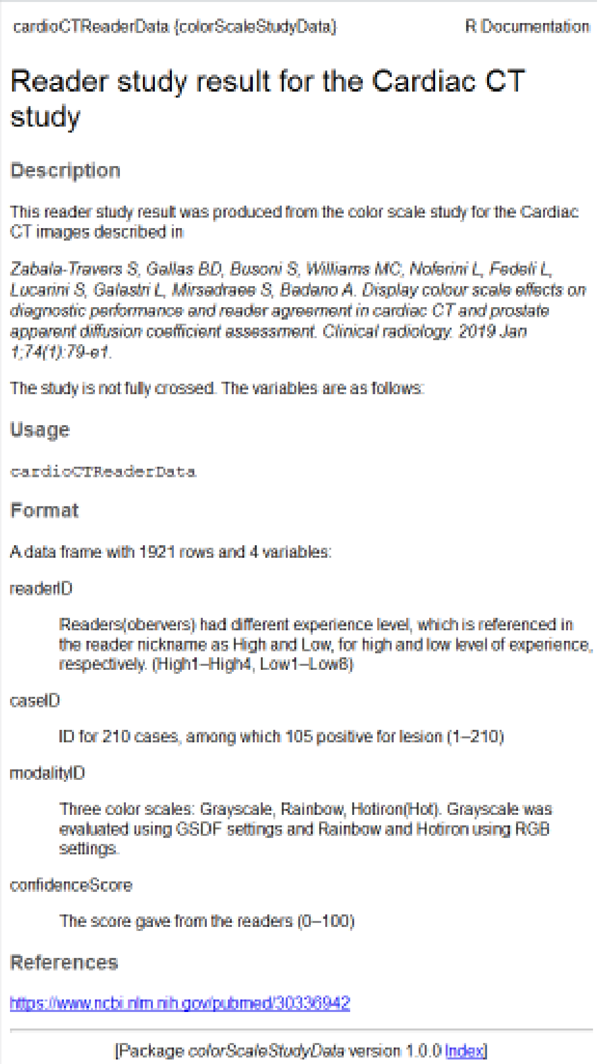
- **Data** contains the R format dataset can be read in R easily

MRMC R data packages

- Documentation
  - colorScaleStudyData (https://github.com/DIDSR/colorScaleStudyData)



```
#' @title{
#' Reader study result for the Cardiac CT study
#' }
#' @description{
#' This reader study result was produced from the color
#'  Cardiac CT images described in
#'
#'  \emph{Zabala-Travers S, Gallas BD, Busoni S, Willia
#'  Lucarini S, Galastri L, Mirsadraee S, Badano A. Dis
#'  effects on diagnostic performance and reader agreem
#'  prostate apparent diffusion coefficient assessment.
#'  2019 Jan 1;74(1):79-e1.}
#'
#'  The study is not fully crossed. The variables are a
#' }
#'
#' @format A data frame with 1921 rows and 4 variables:
#' \describe{
#'   \item{readerID}{Readers(obervers) had different ex
#'    is referenced in the reader nickname as High and
#'    level of experience, respectively. (High1--High4,
#'   \item{caseID}{ID for 210 cases, among which 105 po
#'   \item{modalityID}{Three color scales: Grayscale, R
#'    Grayscale was evaluated using GSDF settings and R
#'    RGB settings. }
#'   \item{confidenceScore}{The score gave from the rea
#' }
#' @references \url{https://www.ncbi.nlm.nih.gov/pubmed
"cardioCTReaderData"

#' @title{
#' Ground truth for the Cardiac CT study
#' }
```



cardioCTReaderData {colorScaleStudyData} R Documentation

Reader study result for the Cardiac CT study

Description

This reader study result was produced from the color scale study for the Cardiac CT images described in

Zabala-Travers S, Gallas BD, Busoni S, Williams MC, Noferini L, Fedeli L, Lucarini S, Galastri L, Mirsadraee S, Badano A. Display colour scale effects on diagnostic performance and reader agreement in cardiac CT and prostate apparent diffusion coefficient assessment. Clinical radiology. 2019 Jan 1;74(1):79-e1.

The study is not fully crossed. The variables are as follows:

Usage

`cardioCTReaderData`

Format

A data frame with 1921 rows and 4 variables:

readerID

Readers(obervers) had different experience level, which is referenced in the reader nickname as High and Low, for high and low level of experience, respectively. (High1–High4, Low1–Low8)

caseID

ID for 210 cases, among which 105 positive for lesion (1–210)

modalityID

Three color scales: Grayscale, Rainbow, Hotiron(Hot). Grayscale was evaluated using GSDF settings and Rainbow and Hotiron using RGB settings.

confidenceScore

The score gave from the readers (0–100)

References

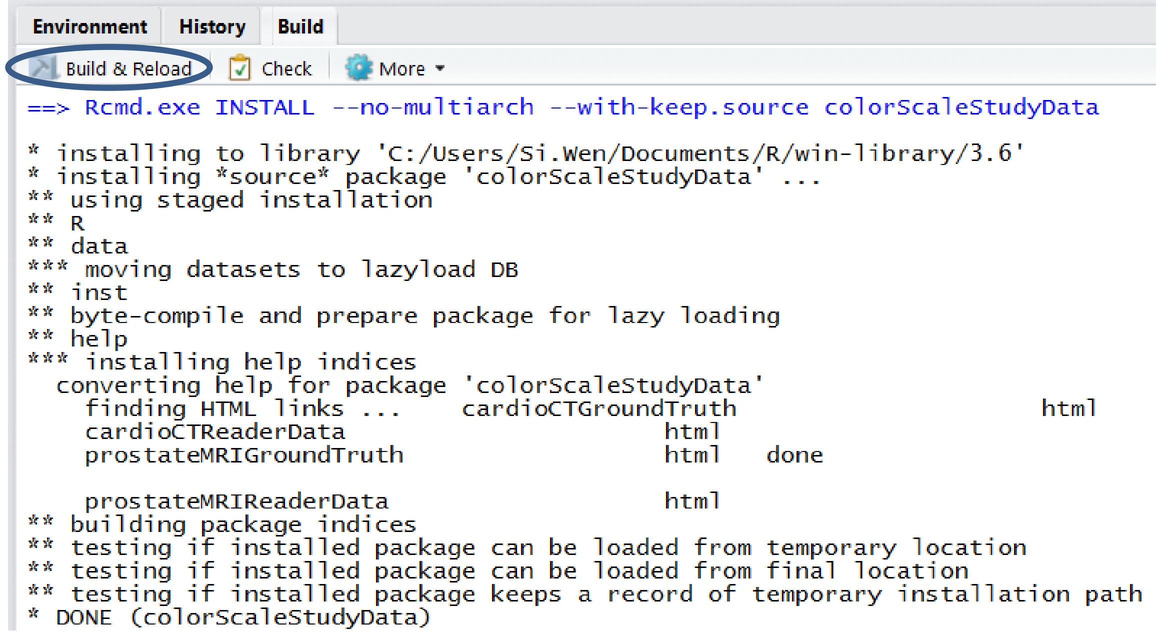
https://www.ncbi.nlm.nih.gov/pubmed/30336942

[Package *colorScaleStudyData* version 1.0.0 Index]

- The documentation is edited in the data.R file in **R**
- devtools::document() generate .Rd manuals in **man**

MRMC R data packages

- Build the package
  - **devtools:: build() build source package and generate .tar.gz file**
    - **No ./data-raw and .Rproj in the zip file**
  - **In Rstudio, you can also build your package by click Build & Reload**



```
==> Rcmd.exe INSTALL --no-multiarch --with-keep.source colorScaleStudyData

* installing to library 'C:/Users/Si.Wen/Documents/R/win-library/3.6'
* installing *source* package 'colorScaleStudyData' ...
** using staged installation
** R
** data
*** moving datasets to lazyload DB
** inst
** byte-compile and prepare package for lazy loading
** help
*** installing help indices
  converting help for package 'colorScaleStudyData'
    finding HTML links ...    cardioCTGroundTruth                     html
    cardioCTReaderData                          html
    prostateMRIGroundTruth                      html    done

    prostateMRIReaderData                       html
** building package indices
** testing if installed package can be loaded from temporary location
** testing if installed package can be loaded from final location
** testing if installed package keeps a record of temporary installation path
* DONE (colorScaleStudyData)
```

MRMC R data packages - Summary

- To store and distribute data along with its documentation
  - Structure
  - Documentation
  - How to create

- To provide use case for functions in existing and future packages
  - How to use : apply doIMRMC in iMRMC package to the data

MRMC Data Analysis

- iMRMC package (https://github.com/DIDSR/iMRMC)
  - A stand-alone, precompiled Java application developed by Dr. Brandon Gallas' lab
  - To analyze and size MRMC reader studies that compare the difference AUCs from two modalities, including the U-statistics estimation of variance of the difference of reader-averaged AUC
  - iMRMC R package: an R package that utilizes the core Java application and output the R list object that contains all the analysis result.

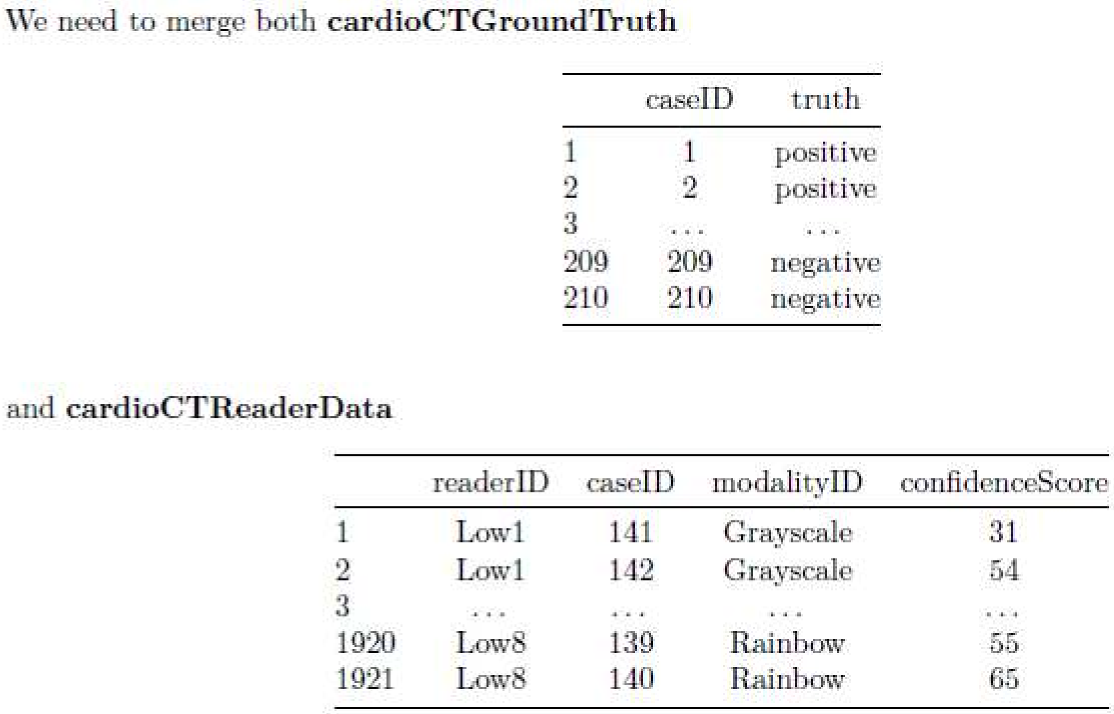
    (https://cran.r-project.org/web/packages/iMRMC/index.html)

MRMC Data Analysis

- iMRMC R package core functions:
  - doIMRMC:
    - Takes ROC data as a data frame and runs a MRMC analysis based on U-statistics
  - sim.gRoeMetz:
    - Simulate an MRMC data set of an ROC study comparing two modalities
    - Allow testing other new methods

MRMC Data Analysis

- Apply doIMRMC in iMRMC package to color scale study data

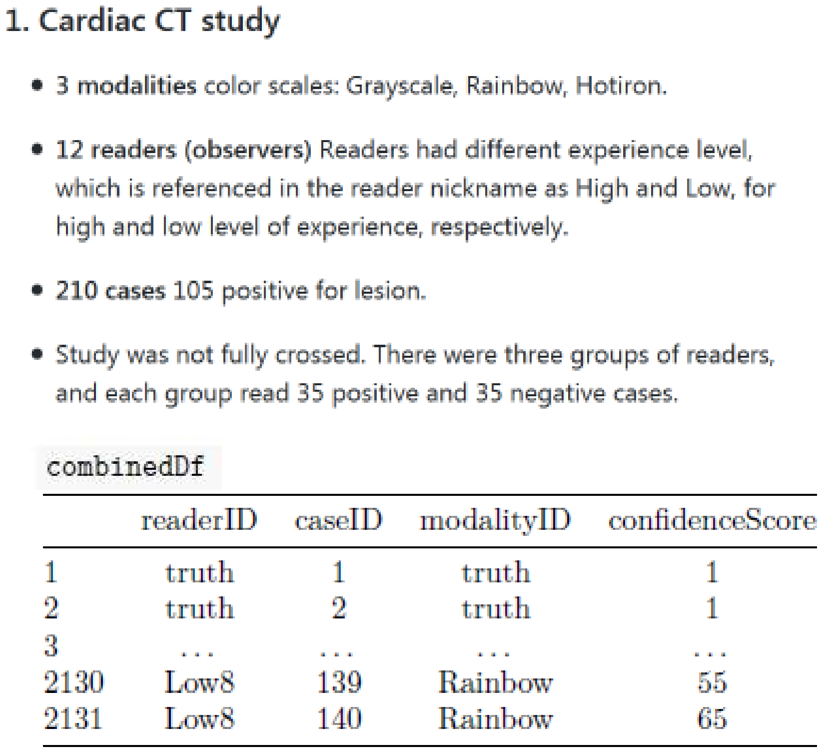


We need to merge both **cardioCTGroundTruth**

	caseID	truth
1	1	positive
2	2	positive
3	...	...
209	209	negative
210	210	negative

and **cardioCTReaderData**

	readerID	caseID	modalityID	confidenceScore
1	Low1	141	Grayscale	31
2	Low1	142	Grayscale	54
3	...	...	...	...
1920	Low8	139	Rainbow	55
1921	Low8	140	Rainbow	65



1. Cardiac CT study

- **3 modalities** color scales: Grayscale, Rainbow, Hotiron.
- **12 readers (observers)** Readers had different experience level, which is referenced in the reader nickname as High and Low, for high and low level of experience, respectively.
- **210 cases** 105 positive for lesion.
- Study was not fully crossed. There were three groups of readers, and each group read 35 positive and 35 negative cases.

`combinedDf`

	readerID	caseID	modalityID	confidenceScore
1	truth	1	truth	1
2	truth	2	truth	1
3	...	...	...	...
2130	Low8	139	Rainbow	55
2131	Low8	140	Rainbow	65

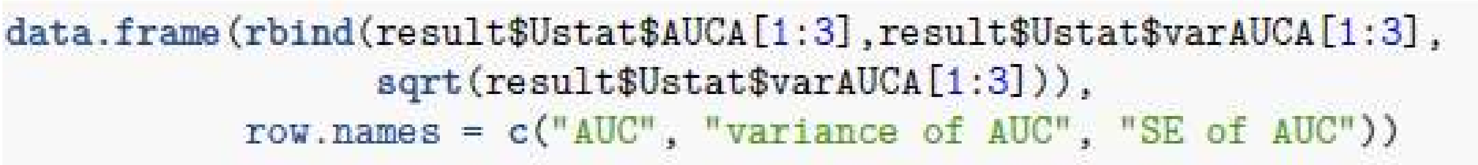
https://github.com/DIDSR/colorScaleStudyData/blob/master/inst/extra/applyIMRMC.pdf

MRMC Data Analysis

- Apply doIMRMC in iMRMC package to color scale study data

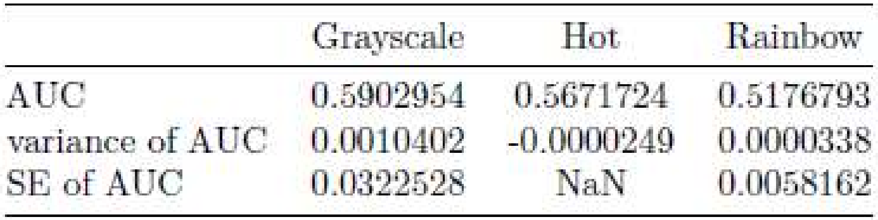
```
library(iMRMC)
result <- doIMRMC(data = combinedDf)
```

  - AUC and variance of AUC for each modality by U-statistics method



```
data.frame(rbind(result$Ustat$AUCA[1:3],result$Ustat$varAUCA[1:3],
                 sqrt(result$Ustat$varAUCA[1:3])),
           row.names = c("AUC", "variance of AUC", "SE of AUC"))
```

Table 6: AUC for different modalities : Ustat



	Grayscale	Hot	Rainbow
AUC	0.5902954	0.5671724	0.5176793
variance of AUC	0.0010402	-0.0000249	0.0000338
SE of AUC	0.0322528	NaN	0.0058162

https://github.com/DIDSR/colorScaleStudyData/blob/master/inst/extra/applyIMRMC.pdf

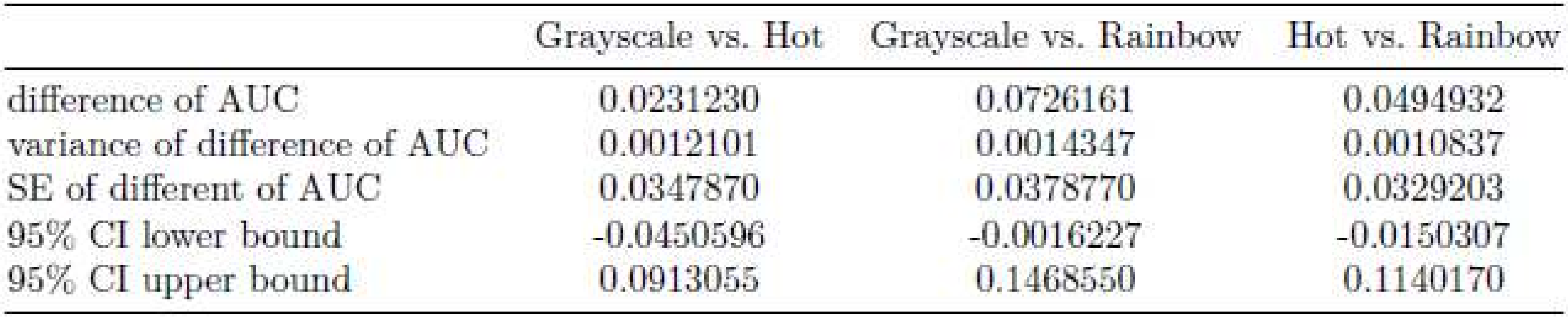
MRMC Data Analysis

- Apply doIMRMC in iMRMC package to color scale study data

```
library(iMRMC)
result <- doIMRMC(data = combinedDf)
```

  - Difference of AUC and variance of difference of AUC for pairwise comparison of different modalities by U-statistics method

Table 8: Difference of AUC among different modalities : Ustat



	Grayscale vs. Hot	Grayscale vs. Rainbow	Hot vs. Rainbow
difference of AUC	0.0231230	0.0726161	0.0494932
variance of difference of AUC	0.0012101	0.0014347	0.0010837
SE of different of AUC	0.0347870	0.0378770	0.0329203
95% CI lower bound	-0.0450596	-0.0016227	-0.0150307
95% CI upper bound	0.0913055	0.1468550	0.1140170

https://github.com/DIDSR/colorScaleStudyData/blob/master/inst/extra/applyIMRMC.pdf

MRMC Data Analysis

- R-markdown files
  - Demonstrate the use of functions and data
  - colorScaleStudyData (https://github.com/DIDSR/colorScaleStudyData/blob/master/inst/extra/applyIMRMC.Rmd)

- More package application
  - viperData R data package (https://didsr.github.io/viperData/)

MRMC Data Analysis

- iMRMC R package core functions:
  - doIMRMC:
    - Takes ROC data as a data frame and runs a MRMC analysis based on U-statistics
  - sim.gRoeMetz:
    - Simulate an MRMC data set of an ROC study comparing two modalities
    - Allow testing other new methods

Simulation Tool for MRMC Data

- Roe and Metz model for MRMC ROC analysis

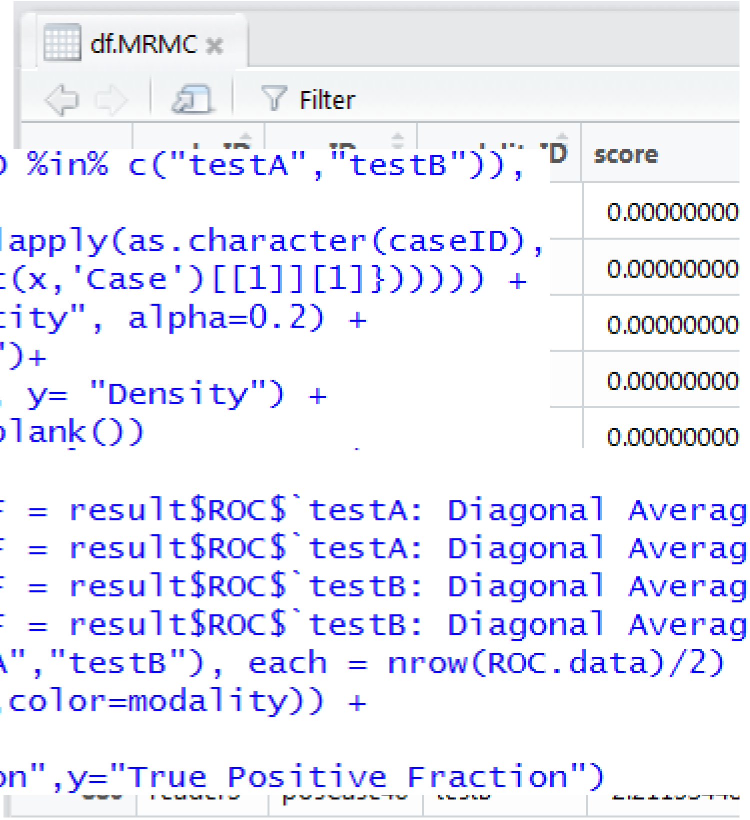
$$X_{ijkt}^{R\&M} = \mu_t + \tau_{it} + R_{jt} + C_{kt} + [RC]_{jkt} + [\tau R]_{ijt} + [\tau C]_{ikt} + [\tau RC]_{ijkt} + E_{ijkt}$$

  - where X_{ijkt} denotes the score for modality i, reader j, case k, and truth state t
  - modality and truth state are fixed factors (μ_t and τ_{it} are fixed effects)
  - reader and case are random factors(the remaining terms are independent zero-mean Gaussian random variables with its own variance)

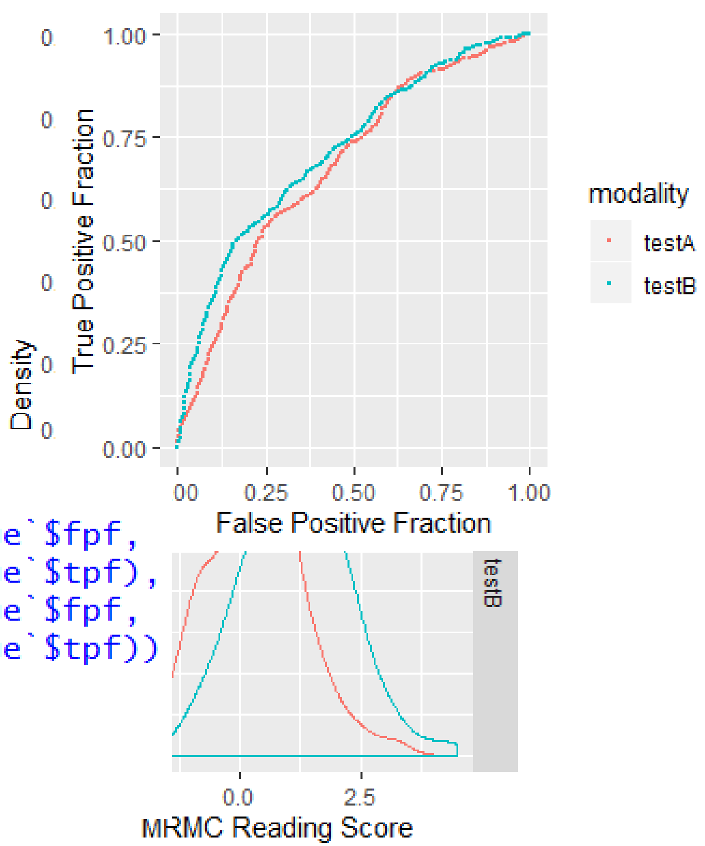
Simulation Tool for MRMC Data

- Apply sim.gRoeMetz to simulate MRMC data

```
> config<-sim.gRoeMetz.config()
> df.MRMC <- sim.gRoeMetz(config)
> View(df.MRMC)
> ggplot(subset(df.MRMC,modalityID %in% c("testA","testB")),
+         aes(x=score,
+             color=factor(unlist(lapply(as.character(caseID),
+              function(x){strsplit(x,'Case')[[1]][1]}))))) +
+   geom_density(position = "identity", alpha=0.2) +
+   facet_grid(rows = "modalityID")+
+   labs(x = "MRMC Reading Score", y= "Density") +
+   theme(legend.title = element_blank())
> result <- doIMRMC(df.MRMC)
> ROC.data <- rbind(data.frame(FPF = result$ROC$`testA: Diagonal Average`$fpf,
+                              TPF = result$ROC$`testA: Diagonal Average`$tpf),
+                   data.frame(FPF = result$ROC$`testB: Diagonal Average`$fpf,
+                              TPF = result$ROC$`testB: Diagonal Average`$tpf))
> ROC.data$modality = rep(c("testA","testB"), each = nrow(ROC.data)/2)
> ggplot(ROC.data,aes(x=FPF,y=TPF,color=modality)) +
+   geom_point(size=0.5) +
+   labs(x="False Positive Fraction",y="True Positive Fraction")
```



score
0.00000000
0.00000000
0.00000000
0.00000000
0.00000000



Summary

- Multi-Reader Multi-Case (MRMC) study
  - Data storage and distribution: **R data Package**
  - MRMC ROC analysis: **doIMRMC in iMRMC package**
  - MRMC simulation tool: **sim.gRoeMetz in iMRMC package**

Reference Links

- colorScaleStudyData R data package https://github.com/DIDSR/colorScaleStudyData
- viperData R data package https://didsr.github.io/viperData/
- iMRMC package https://github.com/DIDSR/iMRMC
- iMRMC R package https://cran.r-project.org/web/packages/iMRMC/index.html
- References:
  - Gallas, Brandon D., et al. "Impact of prevalence and case distribution in lab-based diagnostic imaging studies." *Journal of Medical Imaging* 6.1 (2019): 015501. https://www.spiedigitallibrary.org/journals/Journal-of-Medical-Imaging/volume-6/issue-1/015501/Impact-of-prevalence-and-case-distribution-in-lab-based-diagnostic/10.1117/1.JMI.6.1.015501.full
  - Gallas, Brandon D., and Stephen L. Hillis. "Generalized Roe and Metz receiver operating characteristic model: analytic link between simulated decision scores and empirical AUC variances and covariances." *Journal of Medical Imaging* 1.3 (2014): 031006. https://www.spiedigitallibrary.org/journals/journal-of-medical-imaging/volume-1/issue-3/031006/Generalized-Roe-and-Metz-receiver-operating-characteristic-model--analytic/10.1117/1.JMI.1.3.031006.full
  - Gallas, Brandon D., et al. "A framework for random-effects ROC analysis: biases with the bootstrap and other variance estimators." *Communications in Statistics—Theory and Methods* 38.15 (2009): 2586-2603. https://www.tandfonline.com/doi/abs/10.1080/03610920802610084
  - Gallas, Brandon D., et al. "Evaluating imaging and computer-aided detection and diagnosis devices at the FDA." *Academic radiology* 19.4 (2012): 463-477. https://www.sciencedirect.com/science/article/pii/S1076633211006246

Thank you!

Q&A