

# RMINC Development

Chris Hammill

2016-04-17

# Roadmap

- RMINC history
- CRAN
- Software development and design
- RMINC under the hood
  - ▶ representations
  - ▶ abstractions
- Getting RMINC ready for CRAN
  - ▶ building
  - ▶ testing
  - ▶ documenting
  - ▶ standardizing
- Future Directions

# RMINC

- R interface to the world of MINC
- Statistics for minc volumes
- First commit in 2005!
- Version 1.3 released February 25th, 2016

# Inception

Mouse-Imaging-Centre / **RMINC**

Unwatch 12

Unstar 9

Fork 8

Code

Issues 32

Pull requests 0

Wiki

Pulse

Graphs

Statistics for MINC volumes: A library to integrate voxel-based statistics for MINC volumes into the R environment. Supports getting and writing of MINC volumes, running voxel-wise linear models, correlations, etc.; correcting for multiple comparisons using the False Discovery Rate, and more. With contributions from Jason Lerch, Jim Nikelski and... <https://wiki.phenogenomics.ca/display/MICePub/RMINC>

1 commit

7 branches

10 releases

8 contributors

Your recently pushed branches:

cfhammill:develop (about 1 hour ago)

Compare & pull request

Tree: b6297...

New pull request

New file

Upload files

Find file

HTTPS

<https://github.com/MouseImagingCentre/RMINC>



Download ZIP

author 1134408420 +0000 committed with committer 1134408420 +0000 First RMINC import

Latest commit b62971a on Dec 12, 2005

minc\_interface.R

First RMINC import

11 years ago

minc\_reader.c

First RMINC import


11 years ago


# Development

























# Most Recent Release

Branch: **master** [New pull request](#)

[New file](#) [Upload files](#) [Find file](#) [HTTPS](#) <https://github.com/Mouse>  [Download ZIP](#)

 **mcvaneede** Added URLs of the tutorials to the README

Latest commit 785ae51 on Feb 26

 <a href="#">R</a>	Changes to 3D tools, plus added tutorial	2 months ago
 <a href="#">data</a>	Convert RMINC to R project	6 months ago
 <a href="#">documentation</a>	added the documentation files, and tarballs of previous releases and ...	4 years ago
 <a href="#">inst</a>	Fix typos in RGL demo	2 months ago
 <a href="#">m4</a>	* added all the m4 files	7 years ago
 <a href="#">man</a>	Changes to 3D tools, plus added tutorial	2 months ago
 <a href="#">releases</a>	added the documentation files, and tarballs of previous releases and ...	4 years ago
 <a href="#">sample_data</a>	added the documentation files, and tarballs of previous releases and ...	4 years ago
 <a href="#">src</a>	Modified DESCRIPTION with updated URL	2 months ago
 <a href="#">.Rbuildignore</a>	Convert RMINC to R project	6 months ago
 <a href="#">.gitignore</a>	Ignore makevars and conf* files	4 months ago
 <a href="#">COPYING</a>	mincIO merge: add a copy of the BSD liscence.	6 years ago
 <a href="#">ChangeLog</a>	* added some news about the latest RMINC version, update the ChangeLo...	2 years ago
 <a href="#">DESCRIPTION</a>	Minor version bump in description, update news	2 months ago
 <a href="#">NAMESPACE</a>	Bundle overlay into default bld_obj plot method	2 months ago
 <a href="#">NEWS</a>	After installation the tutorial files end up in the doc/ directory (p...	2 months ago
 <a href="#">README</a>	Added URLs of the tutorials to the README	2 months ago
 <a href="#">RMINC-prototype-code.R</a>	* added RMINC-prototype code, sort of meant to be a self-documented	9 years ago
 <a href="#">RMINC.Rproj</a>	* added --no-test-load to allow it to build on OS X ei capitan	2 months ago
 <a href="#">autogen.sh</a>	Updated coding style	8 years ago
 <a href="#">configure</a>	Manually merged old master and develop into new develop	2 years ago
 <a href="#">configure.ac</a>	Corrected a few minor bits and pieces while ensuring a succesful bull...	6 years ago

# Most Recent Development Branch

Branch: **develop** New pull request

New file Upload files Find file HTTPS https://github.com/cfham Download ZIP

This branch is 28 commits ahead of Mouse-Imaging-Centre:develop. Pull request Compare

**cfhamill** Implement changes suggested in issue #63 Latest commit 1cf4e08 3 minutes ago

<b>R</b>	Regenerate documentation after file moves	2 days ago
<b>inst</b>	Implement changes suggested in issue #63	3 minutes ago
<b>man</b>	Regenerate documentation after file moves	2 days ago
<b>src</b>	This should really be many commits	5 days ago
<b>.Rbuildignore</b>	reduce NOTES	2 days ago
<b>.gitignore</b>	Ignore makevars and conf* files	4 months ago
<b>.travis.yml</b>	Add automatic build testing on ubuntu trusty	4 hours ago
<b>COPYING</b>	minciO merge: add a copy of the BSD liscence.	6 years ago
<b>ChangeLog</b>	* added some news about the latest RMINC version, update the ChangeLo...	2 years ago
<b>DESCRIPTION</b>	Add automatic build testing on ubuntu trusty	4 hours ago
<b>NAMESPACE</b>	Regenerate documentation after file moves	2 days ago
<b>NEWS</b>	After installation the tutorial files end up in the doc/ directory (p...	2 months ago
<b>README</b>	Added URLs of the tutorials to the README	2 months ago
<b>RMINC.Rproj</b>	Move lmerTest to suggest, update Namespace	a month ago
<b>cleanup</b>	This should really be many commits	5 days ago
<b>configure</b>	Enable build minc-toolkit by default	an hour ago
<b>configure.ac</b>	Enable build minc-toolkit by default	an hour ago

# Improving accessibility

- Visibility (be on CRAN)



# Improving accessibility

- Visibility (be on CRAN)
- Availability (be on CRAN)

# Improving accessibility

- Visibility (be on CRAN)
- Availability (be on CRAN)
- Ease of install (be on CRAN)

# Improving accessibility

- Visibility (be on CRAN)
- Availability (be on CRAN)
- Ease of install (be on CRAN)
- Submitted Today!

# CRAN

- The Comprehensive R Archive Network
- 8287 Packages
- Holds all major R packages
- Run by the R-project
- Pretty high standards

# Package Development

- **Scripting vs. Software Development**

# Package Development

- **Scripting vs. Software Development**
- getting the job done vs. building a system

# Package Development

- **Scripting vs. Software Development**
- getting the job done vs. building a system
- comes down to design

# Symptoms of design problems

- **Rigid:**  
tendency for code to resist modification



# Symptoms of design problems

- **Rigid:**  
tendency for code to resist modification
- **Fragile:**  
tendency for modifications to break code in unexpected places

# Symptoms of design problems

- **Rigid:**  
tendency for code to resist modification
- **Fragile:**  
tendency for modifications to break code in unexpected places
- **Immobile:**  
tendency for supposedly modular code to not work in new situations

# Symptoms of design problems

- **Rigid:**  
tendency for code to resist modification
- **Fragile:**  
tendency for modifications to break code in unexpected places
- **Immobile:**  
tendency for supposedly modular code to not work in new situations
- **Viscous:**  
tendency for hacks to be easier to implement than robust solutions<sup>1</sup>

---

<sup>1</sup>Metaphors from Robert Martin

# Design focuses

- Design revolves around data
- **Representation**
- **Abstraction**

# Barriers to good design

- Cruft
- Maintaining compatibility
- Complexity
- Managing dependencies and interfacing with external code

# Refactoring

- Work to make code more understandable
- Work to make code easier to maintain
- Focus on writing small reliable functions
- Focus on composability
- Minimize global state

Representations for minc files

Abstractions for performing statistics

# Representations

- `mincSingleDim`: flat vector of intensities for single minc files and single value statistics tagged with metadata
- `mincMultiDim`: matrices with columns corresponding typically to different statistics, tagged with metadata
- `mincArray`: 3D array of intensities for a single minc file
- `mincMultiArray?`: Dimension respecting representations for results modelling



# Abstraction Goals

- Facilitate movement of imaging data to and from the file system
- Facilitate the generation of useful representations
- Facilitate the fitting of statistical models
- Reduce iteration time for data exploration

# Abstraction Goals

- Facilitate movement of imaging data to and from the file system
- Facilitate the generation of useful representations
- **Facilitate the fitting of statistical models**
- **Reduce iteration time for data exploration**

# Easy Model Fitting

- Modelling in R is typically direct

```
glm(binomial_response ~ covariate1 + ... + covariateN,  
    data = data_source,  
    family = binomial )
```

# Easy Model Fitting

- Modelling in R is typically direct

```
glm(binomial_response ~ covariate1 + ... + covariateN,  
    data = data_source,  
    family = binomial )
```

- **Challenge:** a full experiment will not fit in memory

# Easy Model Fitting

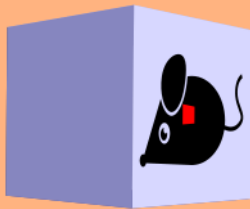
- Modelling in R is typically direct

```
glm(binomial_response ~ covariate1 + ... + covariateN,  
    data = data_source,  
    family = binomial )
```

- **Challenge:** a full experiment will not fit in memory
- **Solution:** iterate through files extracting vectors of voxel values those voxel values are then fit against covariates

2

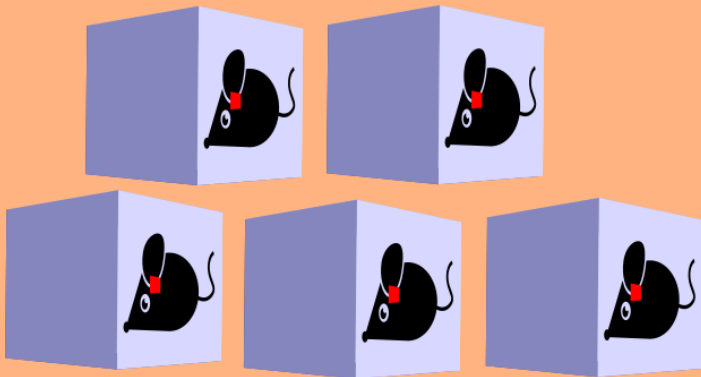
# Disk



---

<sup>2</sup>Mouse icon made by Freepik from [www.flaticon.com](http://www.flaticon.com)

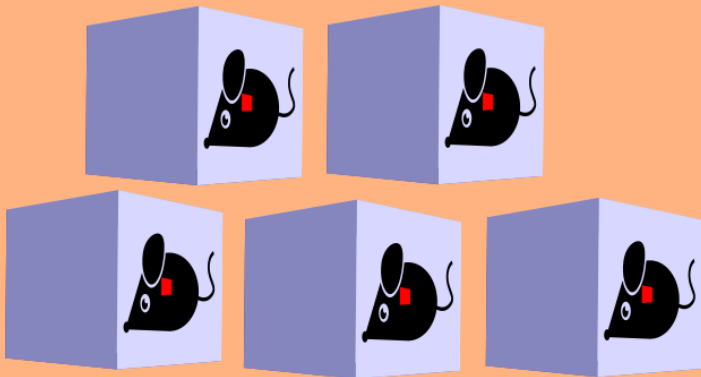
# Disk



# R

$$V_{ijk} = \{m_1, m_2, \dots, m_N\}$$

# Disk



# R

$$V_{ijk+1} = \{m_1, m_2, \dots, m_N\}$$



# Fitting models

- Abstractions:

- ▶ `mincLm`
- ▶ `mincLmer`
- ▶ `mincAnova`
- ▶ `mincTtest`
- ...
- ▶ `mincApply`
- ▶ `mincApplyRCP`

# mincApplyRCPP

- R fluent interface to applying over voxels

```
str(args(mincApplyRCPP))
```

```
## function (filenames, fun,  
##     ..., mask = NULL, maskval = NULL,  
##     filter_masked = FALSE,  
##     slab_sizes = c(1, 1, 1),  
##     return_indices = FALSE,  
##     collate = simplify2minc)
```

```
mincApplyRCPP(experiment_frame$filenames, sample, size = 5)
```

# Expedite fitting models

- **Challenge:** R is (generally) slow, fitting models can be slow
- **Solution:** Implement parallelism
- Abstractions:
  - ▶ pMincApply: General purpose parallelism
  - ▶ mcMincApply: Multicore parallelism
  - ▶ qMincApply: Grid computing parallelism
  - ▶ gMincApply: Open problem

# Dealing with clusters

- Implements flexible support for multiple parallelism backends
- Integration with 'BatchJobs' package
- So far works with HPF and our local cluster, *should* work on scinet but it hasn't been tested
- New abstractions allow computation to be performed in the background

# New Cluster Abstractions

- `qMincApply`: Turnkey function
- `qMincRegistry`: Create shared files for parallel jobs to be coordinated, once created, registries can be loaded from another R session or even another machine
- `qMincMap`: Split the calculation into many peices and prepare the scripts necessary for job submission
- `qMincReduce`: Retrieve the job results combining them into the object of your choice, defaults to minc statistic object like `mincMultiDim`
- Currently only performs voxel-wise calculations, but can be extended to more complex analyses or be used to parellelize hyper-parameter search

# Push for CRAN

- Must build automatically on two platforms
- Must perform as expected
- Must be well documented
- Must adhere to CRAN standards

# Building R Packages

- R uses a C-like build system
- Supports basic configure scripts and makevars
- Does not support cmake
- **Challenge:** Locating/Providing system dependencies
- **Solution(?):** Using autoconf to generate a configure script that can flexibly locate libminc and hdf5
- if they are not found it can build minc-toolkit-v2 from source using git and cmake (fraught for automatic builds)
- **Goal:**

```
install.packages("RMINC")
```

Should work without issue from all supported platforms

# Keeping things working

- Ensuring code performs as expected is a significant problem
- Strategies exist to ensure code performs as expected



# Scale of Testing

*Optimism is an occupational hazard of programming: feedback is the treatment - Kent Beck*

- ① **Optimism** - “this really looks like it should work”
- ② **Hand testing** - “I tried a few variants of this and they work”
- ③ **Unit testing** - “I broke the problem into small pieces, and had the computer check that they all work”
- ④ **Property testing** - “I specified rules about return values, the computer explores parameter space to see if rules hold”
- ⑤ **Formal proof** - “I mathematically demonstrated the program is correct”

# Scale of Testing

*Optimism is an occupational hazard of programming: feedback is the treatment - Kent Beck*

- ① **Optimism** - “this really looks like it should work”
- ② **Hand testing** - “I tried a few variants of this and they work”
- ③ **Unit testing** - “I broke the problem into small pieces, and had the computer check that they all work”
- ④ **Property testing** - “I specified rules about return values, and had the computer explore parameter space to see if rules hold”
- ⑤ **Formal proof** - “I mathematically demonstrated the program is correct”

# Unit testing

- Unit tests are a set of assertions about the value functions return.

```
library(testthat)
test_that("My Code Works",
  expect_equal(mean(1:5), 3))

test_that("My other code also works",
  expect_equal(sd(1:5), 1.5))
```

```
## Error: Test failed: 'My other code also works'
## Not expected: sd(1:5) not equal to 1.5
## 1.58 - 1.5 == 0.0811.
```

# RMINC Testing

- RMINC includes a tool for testing the package

```
runRMINCTestbed()
```

runs a suite of tests on our code to check that all is working as expected

- `runRMINCTestbed` calls out to 'test\_directory' which runs all tests in `inst/tests`
- Canonical approach is to have tests in `test` so that CRAN can test them automatically, but some thought is needed on how best to do the conversion



# Travis CI

- Services provide automatic testing on multiple platforms
- Runs whenever code is changed on github
- Runs whenever a pull request is issued

# R Documentation

- CRAN requires all user visible functions to be documented.  
This is huge benefit to users as all documentation can be quickly accessed with the standard ? call

# R Documentation

- CRAN requires all user visible functions to be documented.  
This is huge benefit to users as all documentation can be quickly accessed with the standard ? call
- **Challenge:** Documentation must be kept in sync with the code

# R Documentation

- CRAN requires all user visible functions to be documented.  
This is huge benefit to users as all documentation can be quickly accessed with the standard ? call
- **Challenge:** Documentation must be kept in sync with the code
- **Solution:** Documentation lives in source with the code and is autogenerated each time the package is rebuilt.



# R Documentation

- CRAN requires all user visible functions to be documented.  
This is huge benefit to users as all documentation can be quickly accessed with the standard ? call
- **Challenge:** Documentation must be kept in sync with the code
- **Solution:** Documentation lives in source with the code and is autogenerated each time the package is rebuilt.
- 'roxygen2' integration with rstudio supports this

# Roxygen Documentation

```
## Retrieve Voxel Values  
##  
## Return the intensity of a given voxel in a set of minc files  
##  
## @param filenames paths to the minc files  
## @param v1 Either a 3-element vector of voxel coordinates  
## or the first  
## @param v2 the second voxel coordinate if not NULL  
## @param v3 the third voxel coordinate if not NULL  
## @return Returns a \code{mincVoxel} object containing a vector  
## of intensities and attributes specify the voxel and world  
## coordinates of the values.  
## @export  
mincGetVoxel <- function(filenames, v1, v2=NULL, v3=NULL) {
```

# Retrieve Voxel Values

## Description

Return the intensity of a given voxel in a set of minc files

## Usage

```
mincGetVoxel(filenamees, v1, v2 = NULL, v3 = NULL)
```

## Arguments

- |                         |   |
|-------------------------|---|
| <code>filenamees</code> | paths to the minc files                                     |
| <code>v1</code>         | Either a 3-element vector of voxel coordinates or the first |
| <code>v2</code>         | the second voxel coordinate if not NULL                     |
| <code>v3</code>         | the third voxel coordinate if not NULL                      |

## Value

Returns a `mincVoxel` object containing a vector of intensities and attributes specify the voxel and world coordinates of the values.

---

[Package *RMINC* version 1.3.0.0 [Index](#)]

# Appeasing R CMD check

- CRAN is very particular about code quality
- Luckily R CMD check tests many requirements for you
- Highlights missed arguments in documentation, missing global variables (helpful for catching typos)
- Ensures dependencies can be loaded properly

# Future directions

- Almost ready for release
- Harden parallel code
- Convert testing scheme to canonical approach
- Add abstractions for multivariate statistics

Ever need R help?  
Come by my desk or drop me an email

Questions?