Computational Methods for Neuroimaging in R, Hemorrhagic Stroke and Neuroconductor

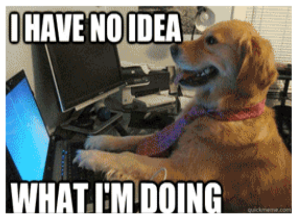
John Muschelli, PhD - Johns Hopkins Bloomberg School of Public Health
http://johnmuschelli.com/Rochester_2016.html
@StrictlyStat
https://github.com/muschellij2

October 13, 2016

BS - 2008

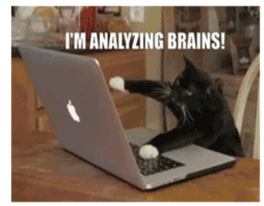
The Office

ScM - 2010



Consulting
2012

PhD
2016



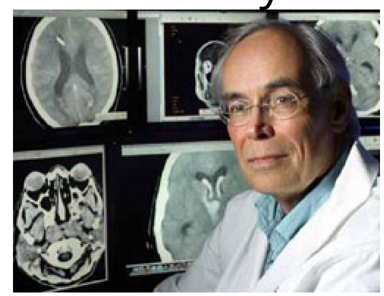
Assistant
Scientist

The MISTIE Stroke Trial

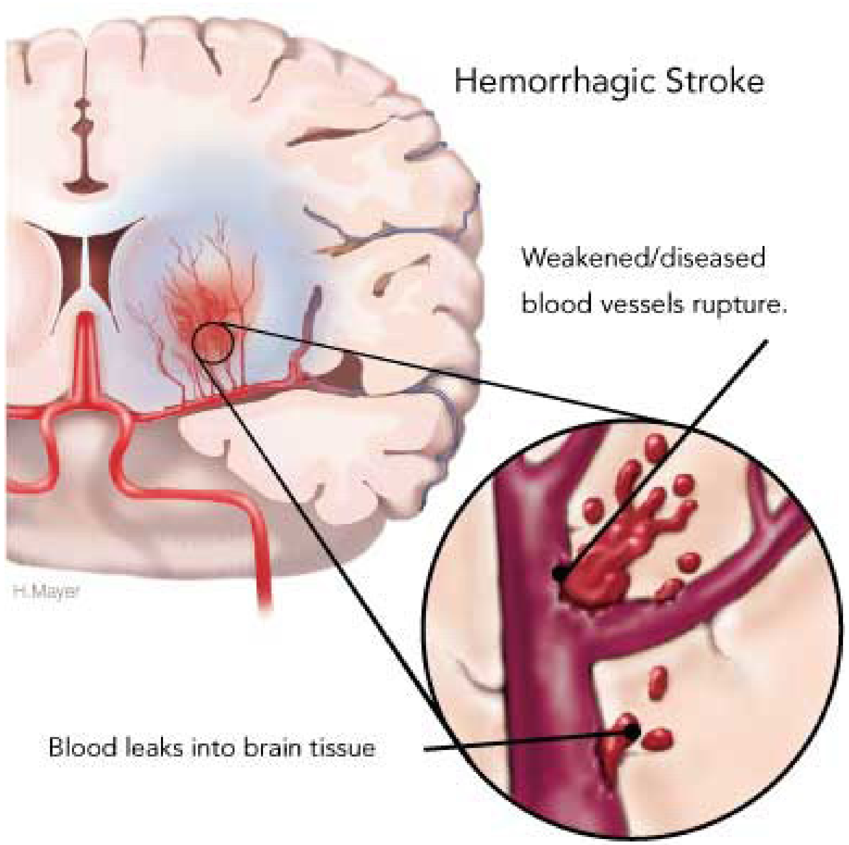
- Minimally Invasive Surgery plus r-tPA for Intracerebral Hemorrhage Evacuation (**MISTIE**)
  - Multi-center, multi-national Phase II clinical trial
  - Alteplase (donated by Genentech)
- MISTIE III

http://braininjuryoutcomes.com/mistie-about

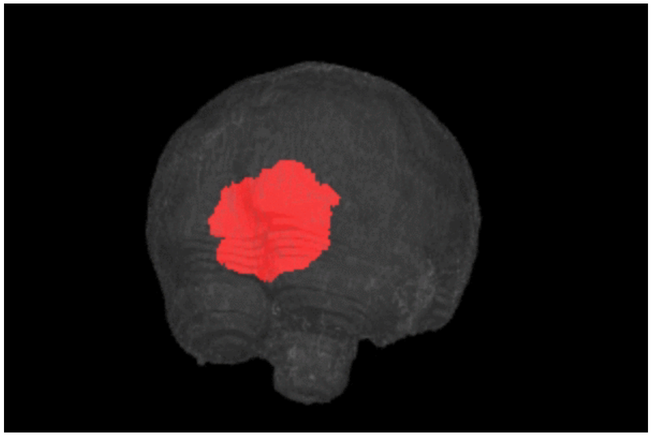
PI: Dr. Dan Hanley



Intracranial/Intracerebral Hemorrhage (ICH)



- When a blood vessel ruptures into tissue
- ≈ 13% of strokes



http://www.heartandstroke.com/site/c.ikIQLcMWJtE/b.3484153/k.7675/Stroke__Hemorrhagic_stroke.htm

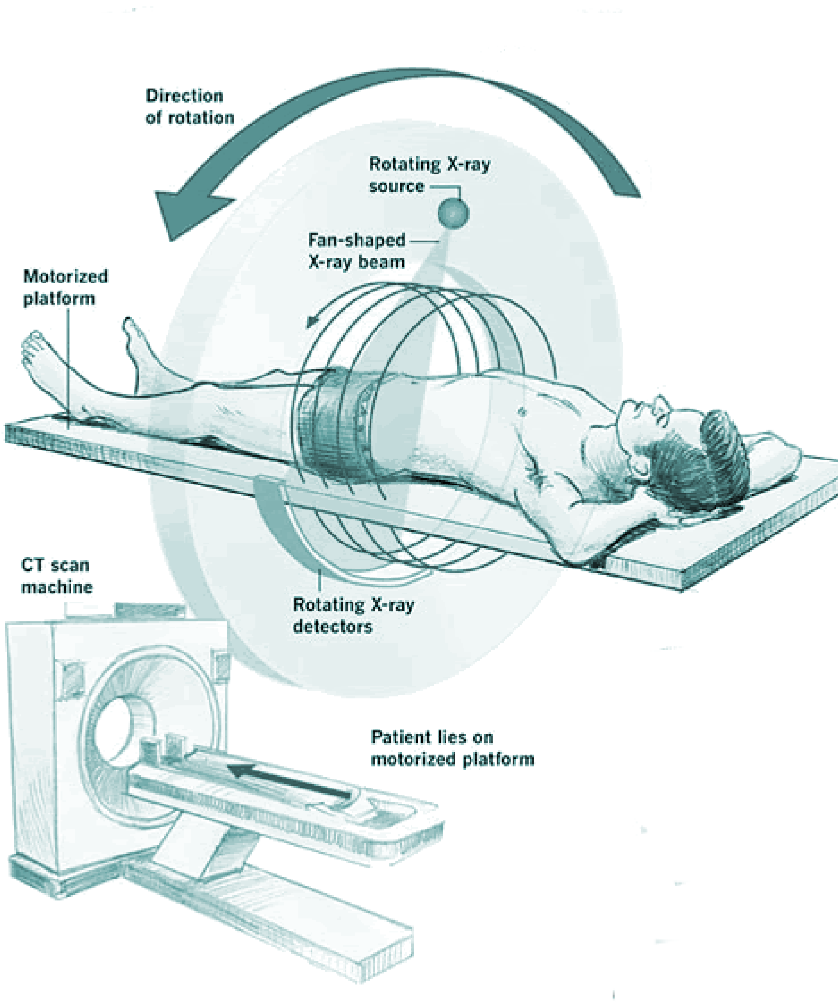
Larger Hemorrhage Volume $\Rightarrow$ Worse Outcome

Goal:
Automatic Hemorrhage Volume Estimation

How do we measure hemorrhages?

X-ray Computed Tomography (CT) Scans

Image from http://www.cyberphysics.co.uk/topics/medical/CTScanner.htm



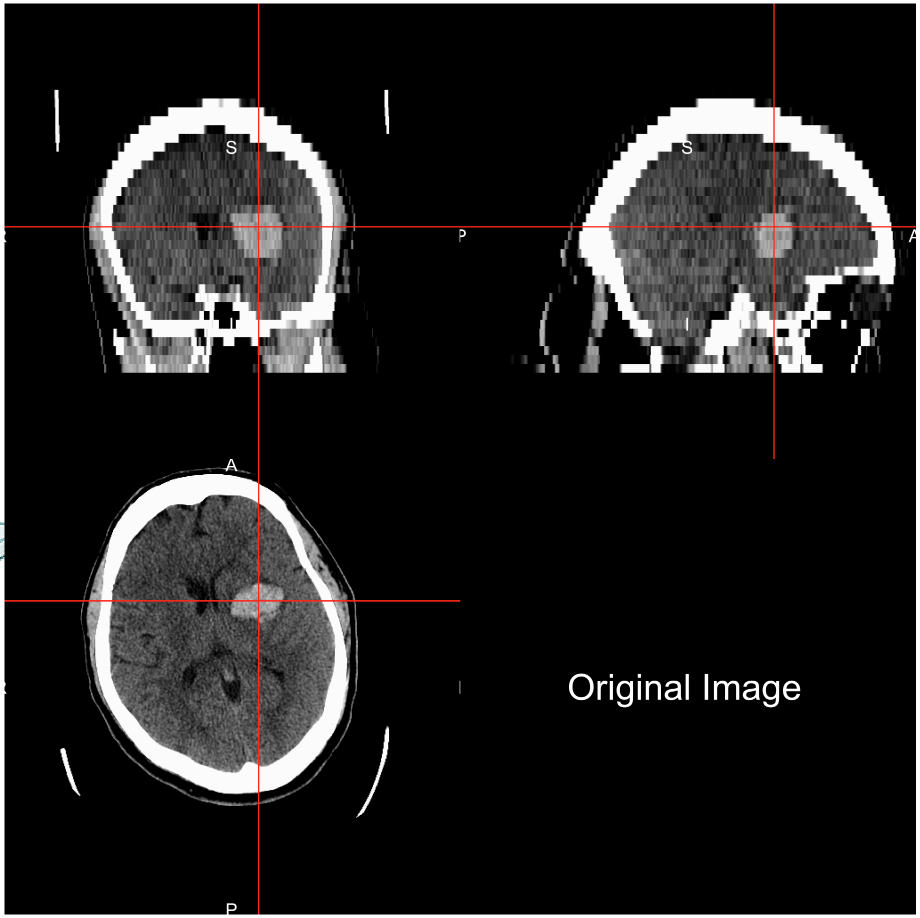
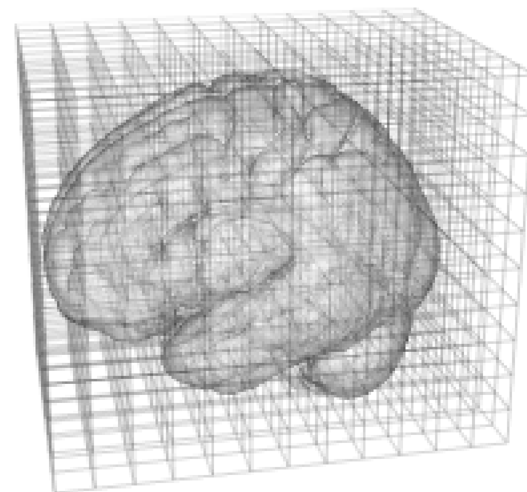


Image Representation: voxels (3D pixels)



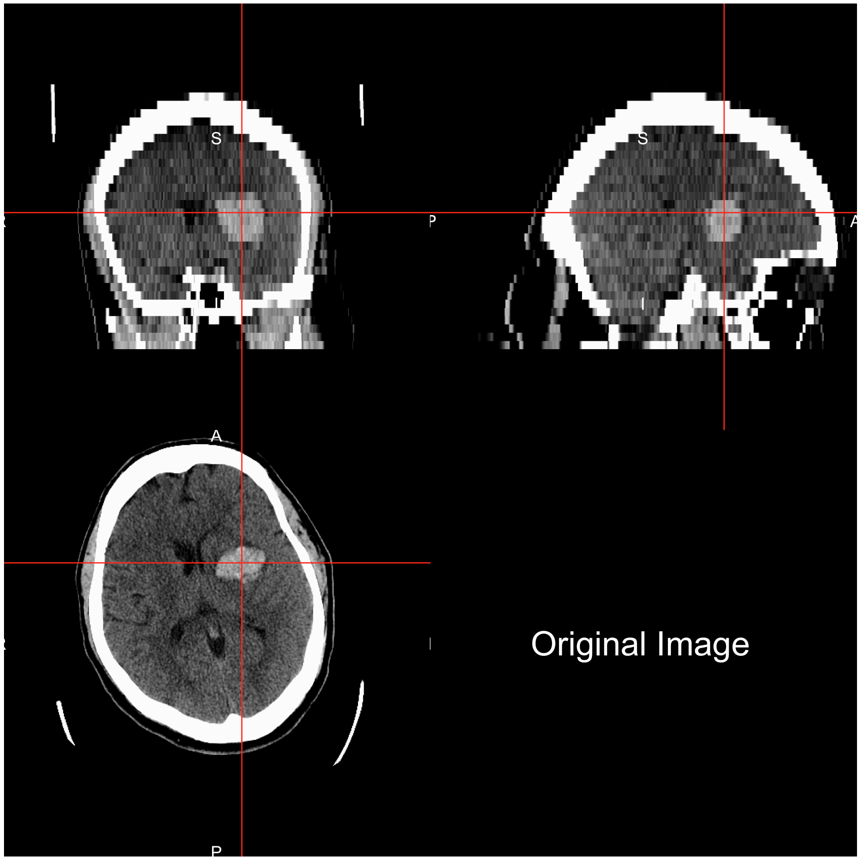
Muschelli, John, Elizabeth Sweeney, and Ciprian Crainiceanu. "brainR: Interactive 3 and 4D Images of High Resolution Neuroimage Data." R JOURNAL 6.1 (2014): 42-48.

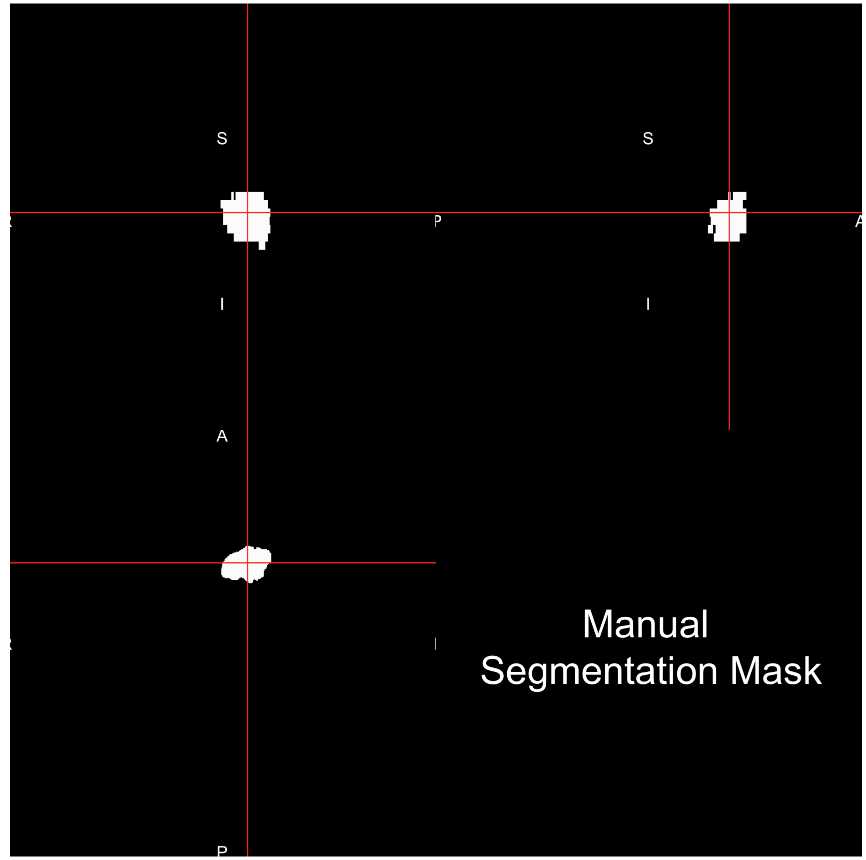
Terminology: Neuroimaging to Data/Statistics

- Segmentation ⇔ classification
- Image ⇔ 3-dimensional array
    - composed of voxels
    - a typical CT has $512 \times 512 \times 30 \approx 7.8$ million voxels
- Mask/Region of Interest ⇔ binary (0/1) image

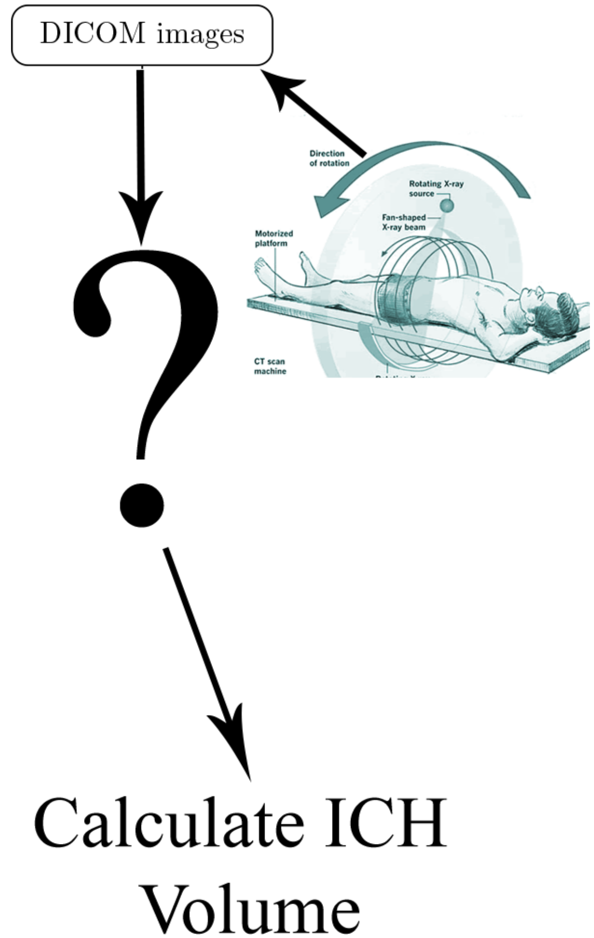
How do we measure volume (currently)?

Manual Segmentation

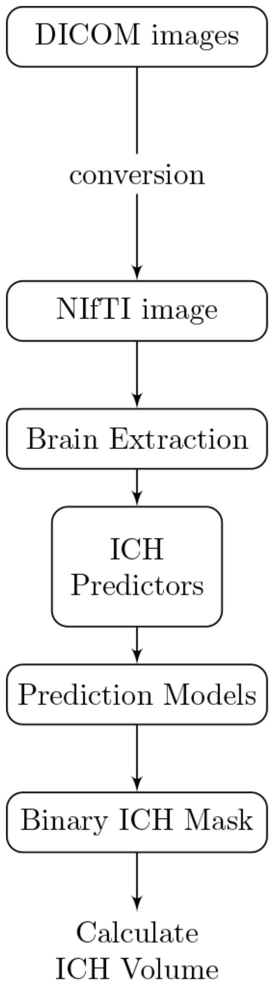




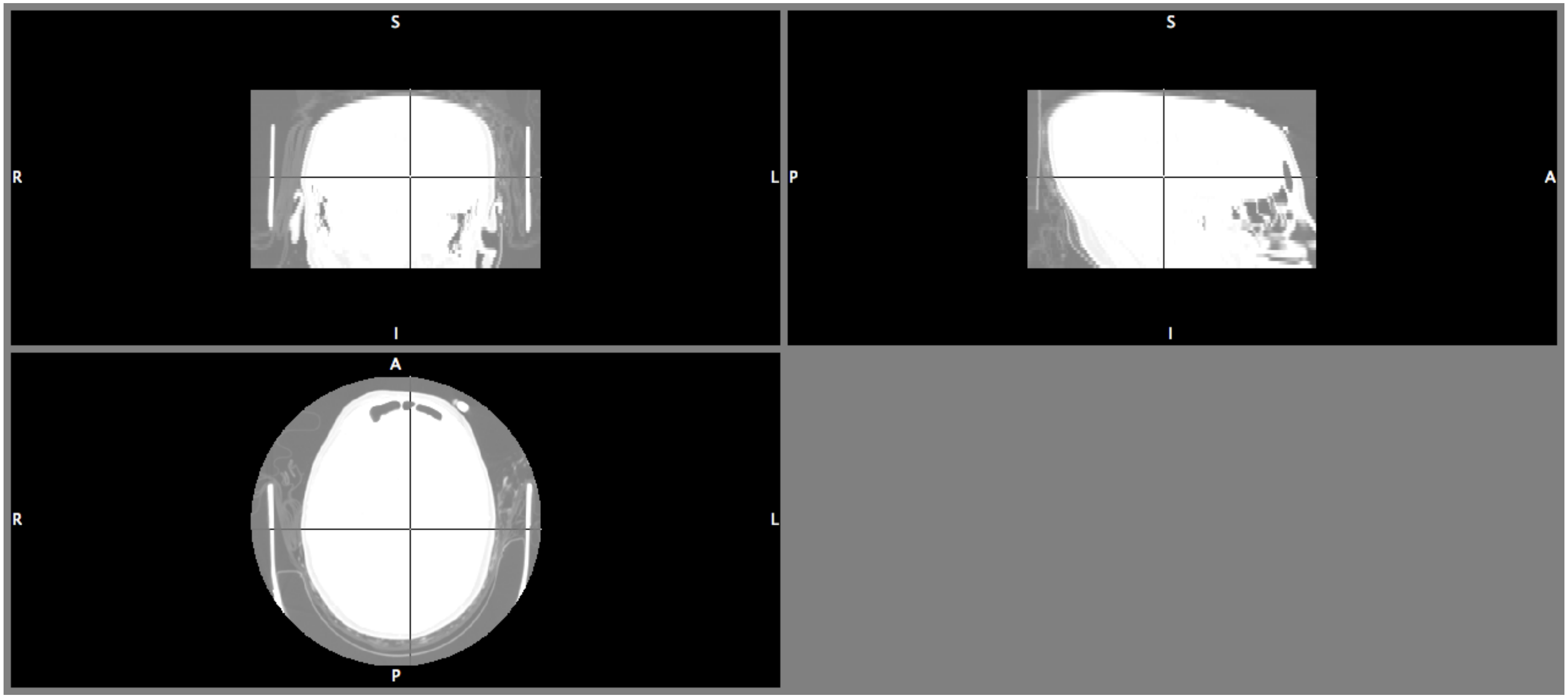
Workflow for the Analysis



Workflow for the Analysis

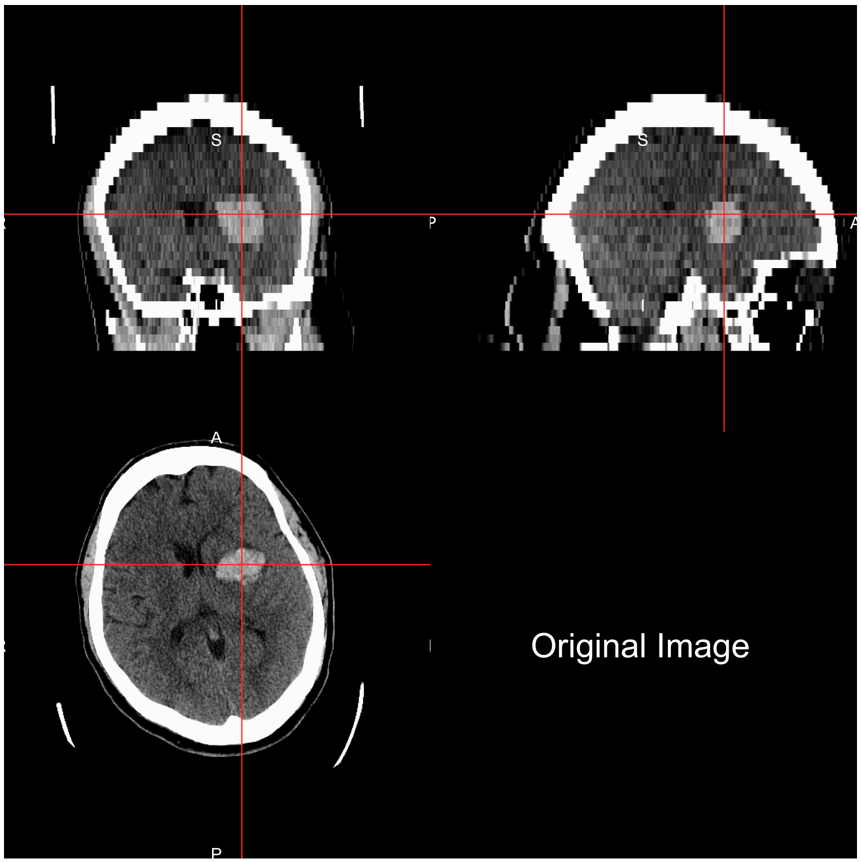


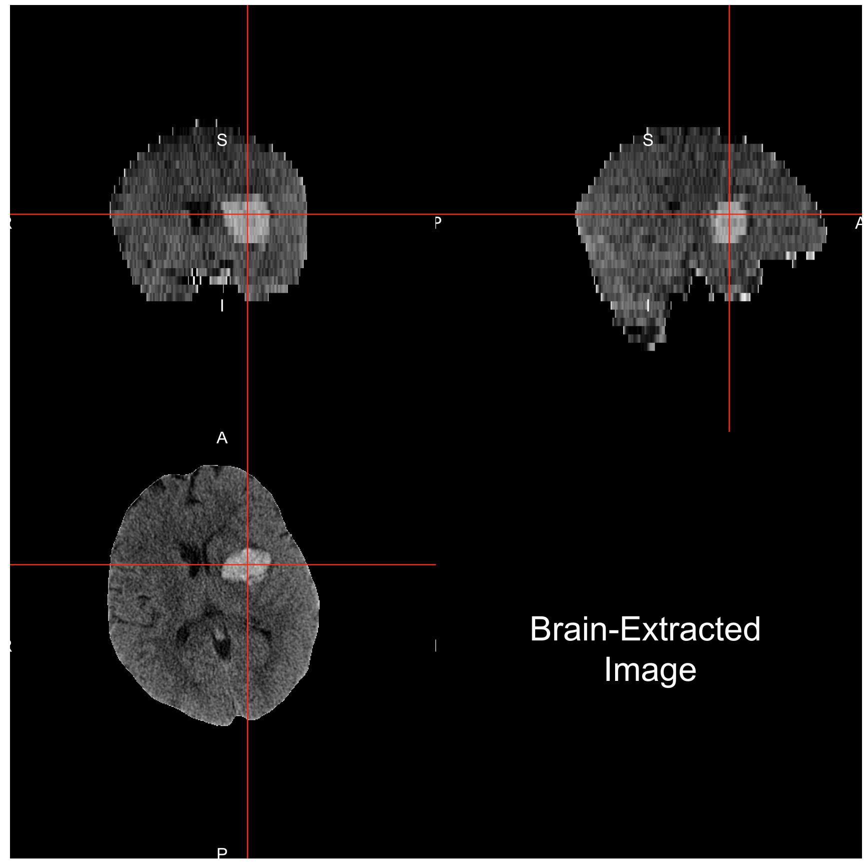
Problem: CT Scans Capture **Everything**



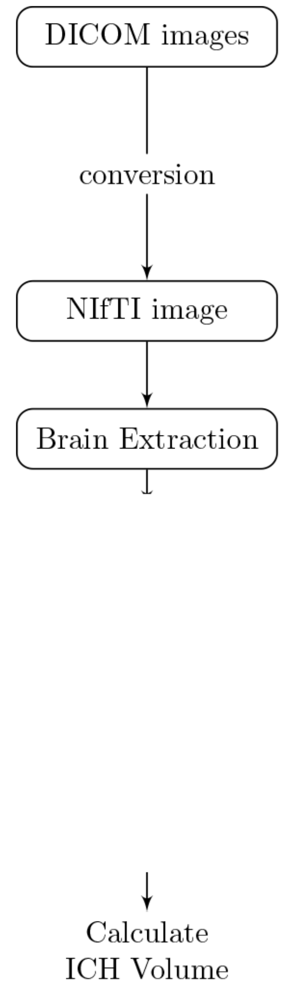
Brain Segmentation of CT Scans

- Muschelli, John, et al. "Validated automatic brain extraction of head CT images." NeuroImage 114 (2015): 379-385. http://bit.ly/CTBET_RCODE





Workflow for the Analysis



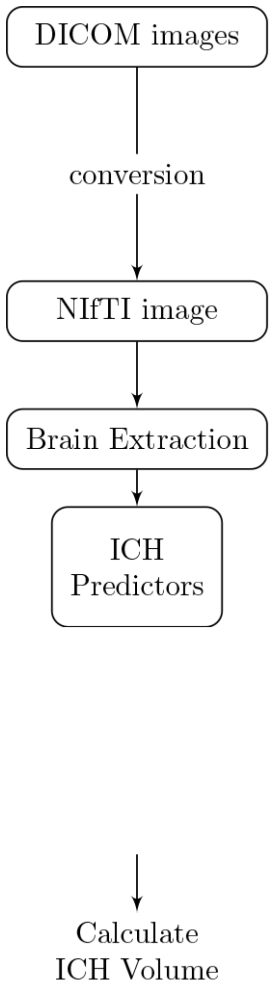
Hemorrhage Segmentation

Subject Data used: 111 scans (1 Per Patient)

	Overall
Age in Years: Mean (SD)	60.8 (11.2)
Male: N (%)	76 (68.5%)
Reader-Based Clot Location (%)	
Putamen	68 (61.3)
Lobar	33 (29.7)
Globus Pallidus	6 (5.4)
Thalamus	4 (3.6)
ICH Volume in mL: Mean (SD)	37.4 (20.1)

- Adults (inclusion criteria 18-80 years old)
- Mostly males
- Different sites, scanners, and locations in the brain

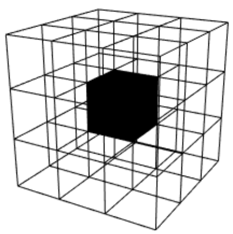
Workflow for the Analysis



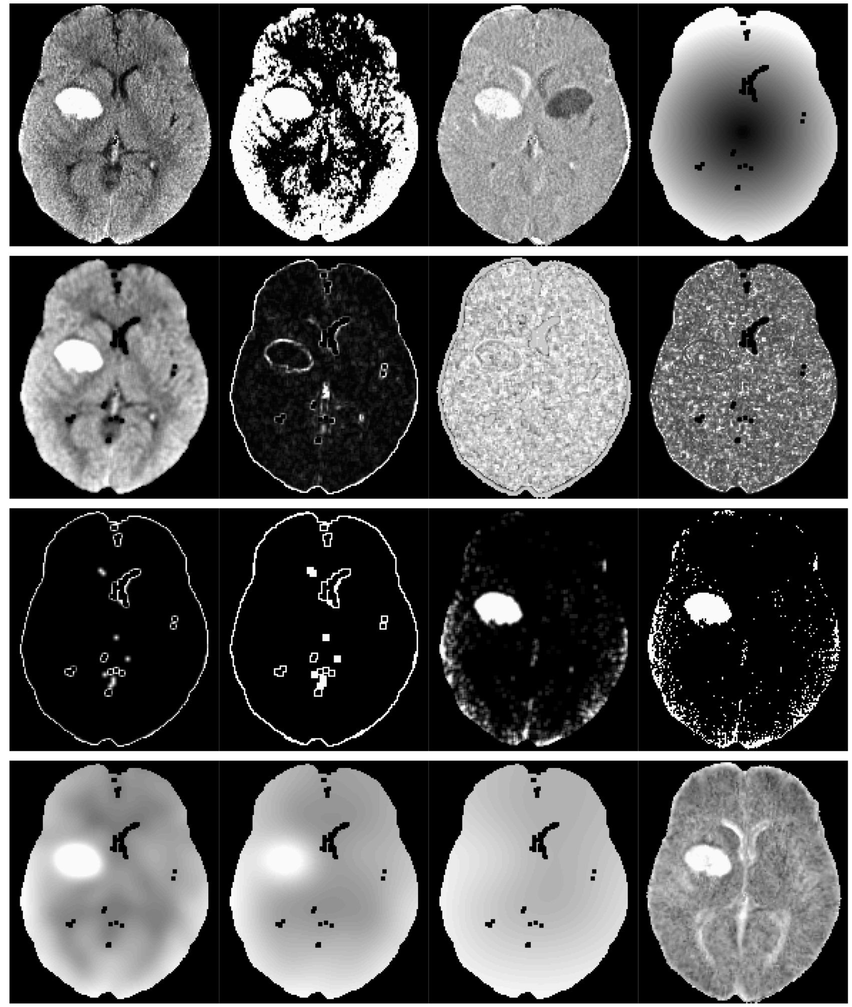
Predictors of ICH

- Local "neighborhood" moments (mean, sd, skew, kurtosis)

Local Mean: $\bar{x}(v) = \frac{1}{27} \sum_{k}^{27} x_k(v)$

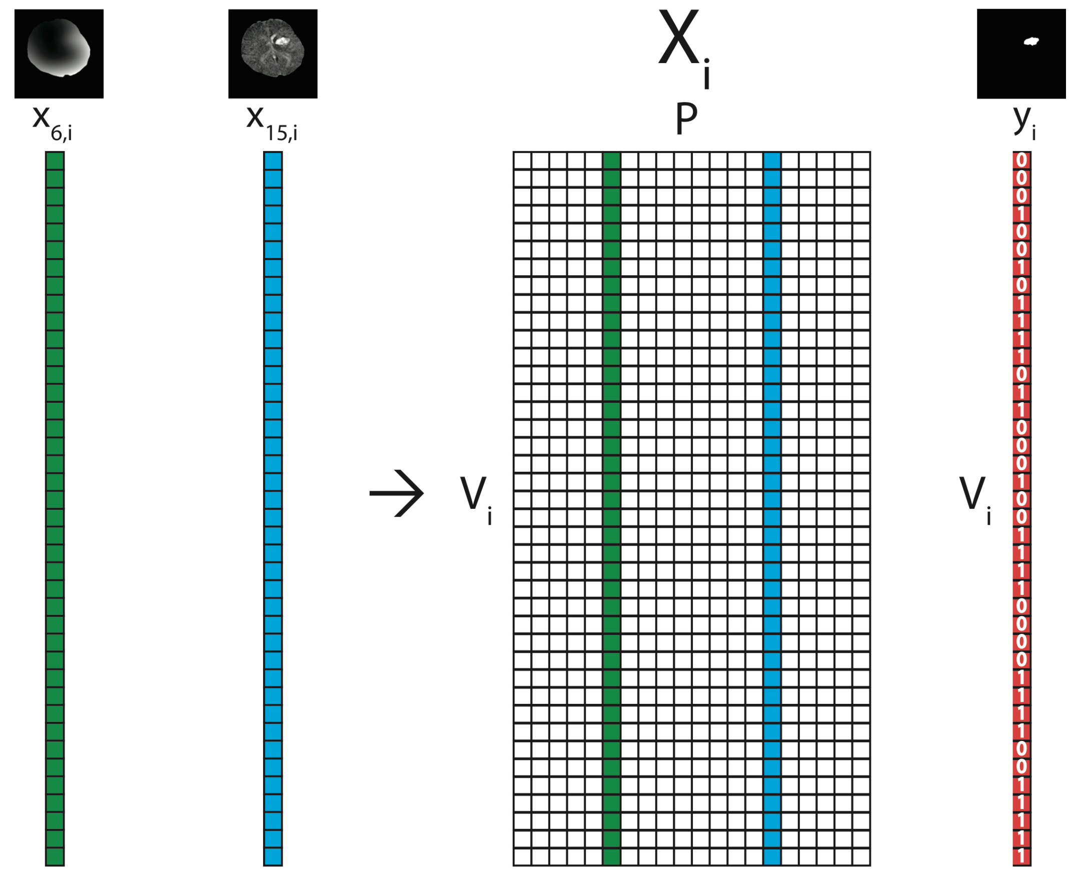


- Threshold intensity
- Gaussian Smoother (**big** neighborhood)



Data Structure for One Patient

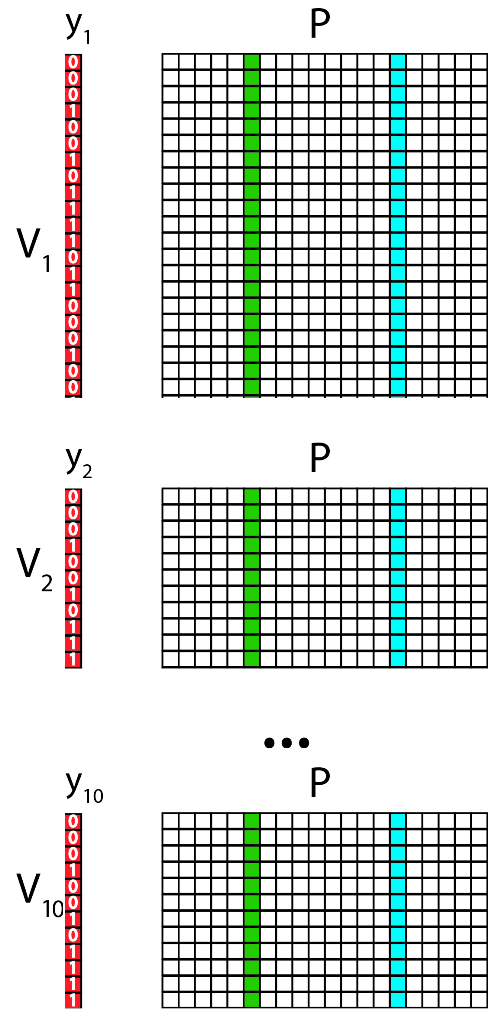
- Each subject has matrix X_i
- V_i - voxels per person
- P - number of predictors
- y_i - manual segmentation



Aggregate Data

Training Data Structure

- Stack together 10 randomly selected patients
- Train model/classifier on this design matrix



Fit Models / Classifier

Again, $y_i(v)$ is the presence / absence of ICH for voxel v from person i.

General model form:

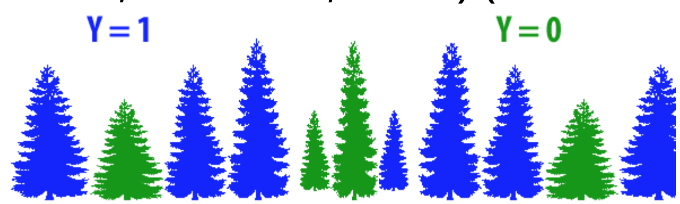
$$P\{y_i(v) = 1\} \propto f\{X_i(v)\}$$

Models Fit on the Training Data

- Logistic Regression: $f(X_i(v)) = \text{expit}\left\{\beta_0 + \sum_{k=1}^{p} x_{i,k}(v)\beta_k\right\}$ (`glm` function)
- Generalized Additive Model (Hastie, et al., 1990)
  - fit using penalized thin plate splines (`mgcv` package)
- LASSO (Tibshirani, 1996; Friedman, et al., 2010) (`glmnet` package):

$$\mathcal{L}\left[Y_i(v)|\, f\{X_i(v)\}\right] \propto \beta_0 + \sum_{k=1}^{p} x_{i,k}(v)\beta_k + \lambda \sum_{k=1}^{p} |\beta_k|$$

- Random Forests (Liaw, et al., 2002; Breiman, 2001) (`randomForest` package)

$f(X_i(v)) \propto$ 

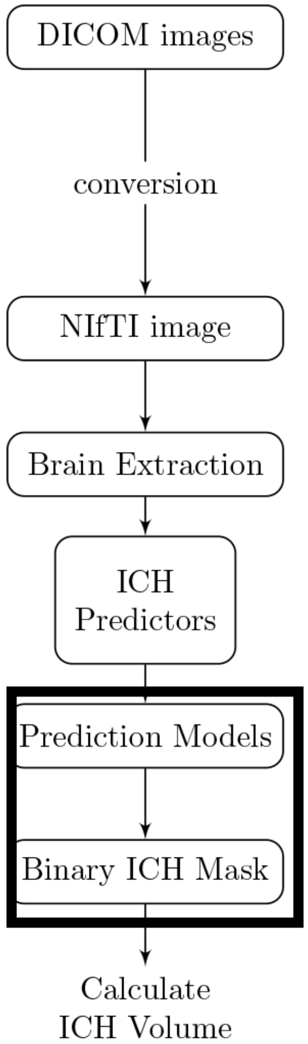
All results are on the 101 scans/patients
not in the training data

- Get prediction for each voxel: $\hat{p}_i(v) = \hat{P}(y_i(v) = 1)$
- Threshold to get a binary value: $\hat{y}_i = \hat{p}_i(v) > 0.5$

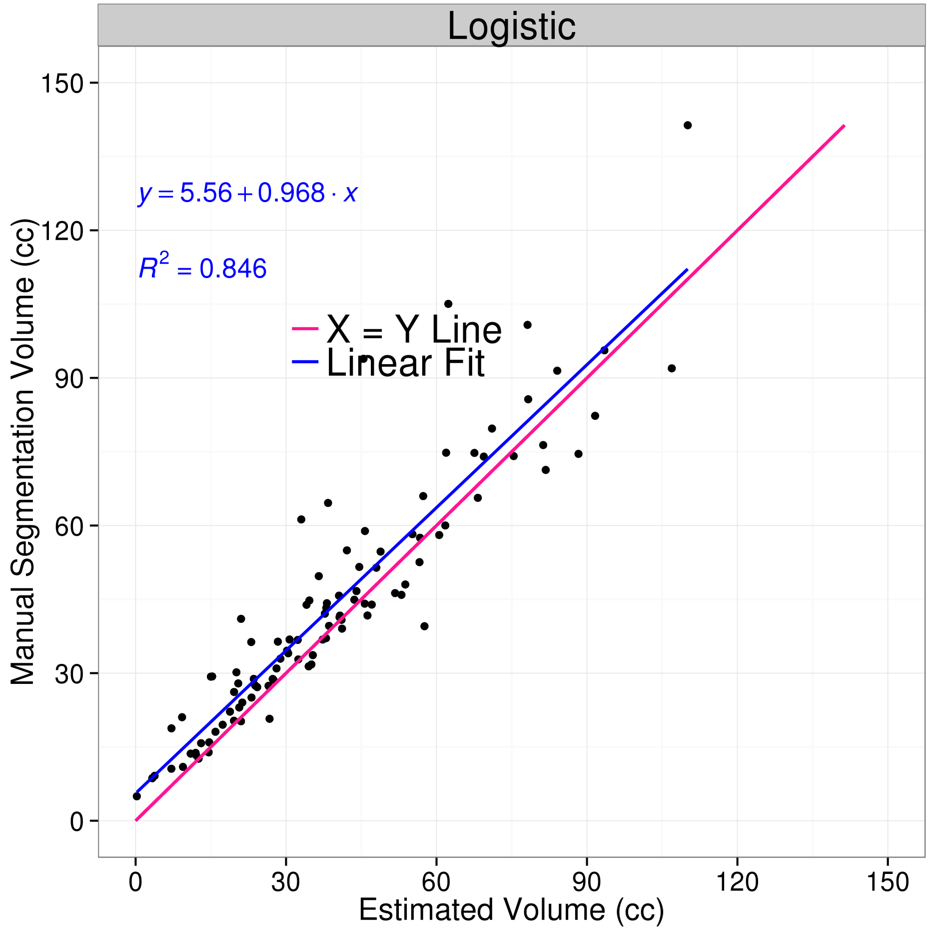
$$\text{Estimated Volume}_i = \text{cc per voxel} \times \sum_{v}^{V_i} \hat{y}_i(v)$$

$$\text{Volume}_i = \text{cc per voxel} \times \sum_{v}^{V_i} y_i(v)$$

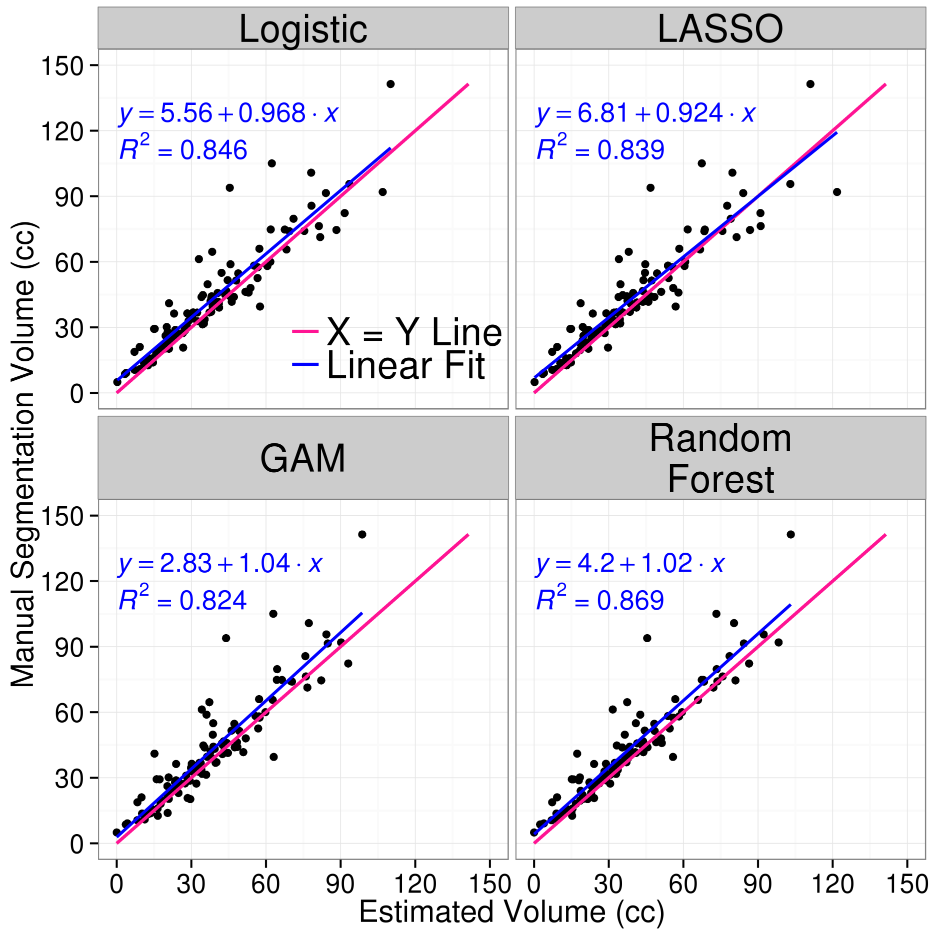
- cc = cubic centimeters
- one milliliter (mL) = 1 cc



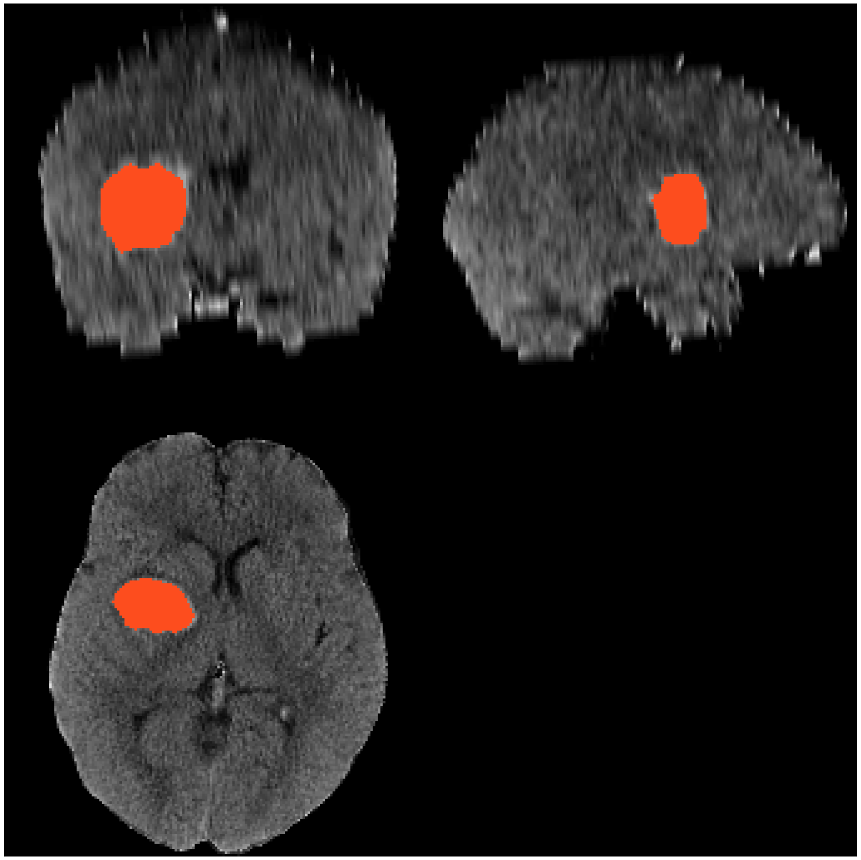
Predicted Volume Estimates True Volume

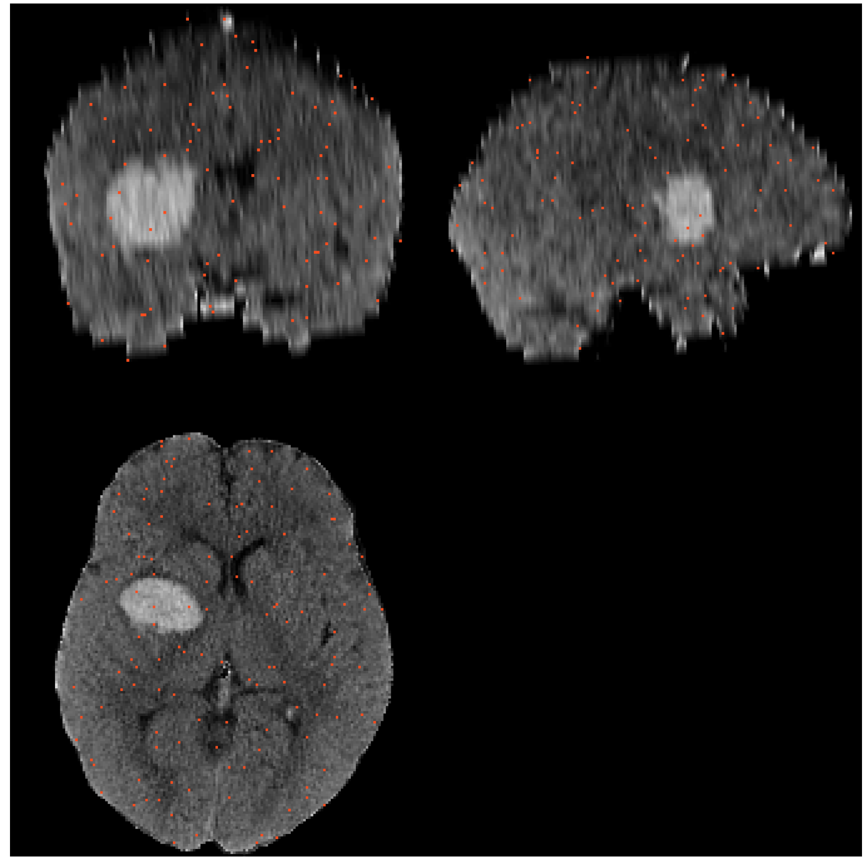


All models predicted volume pretty well

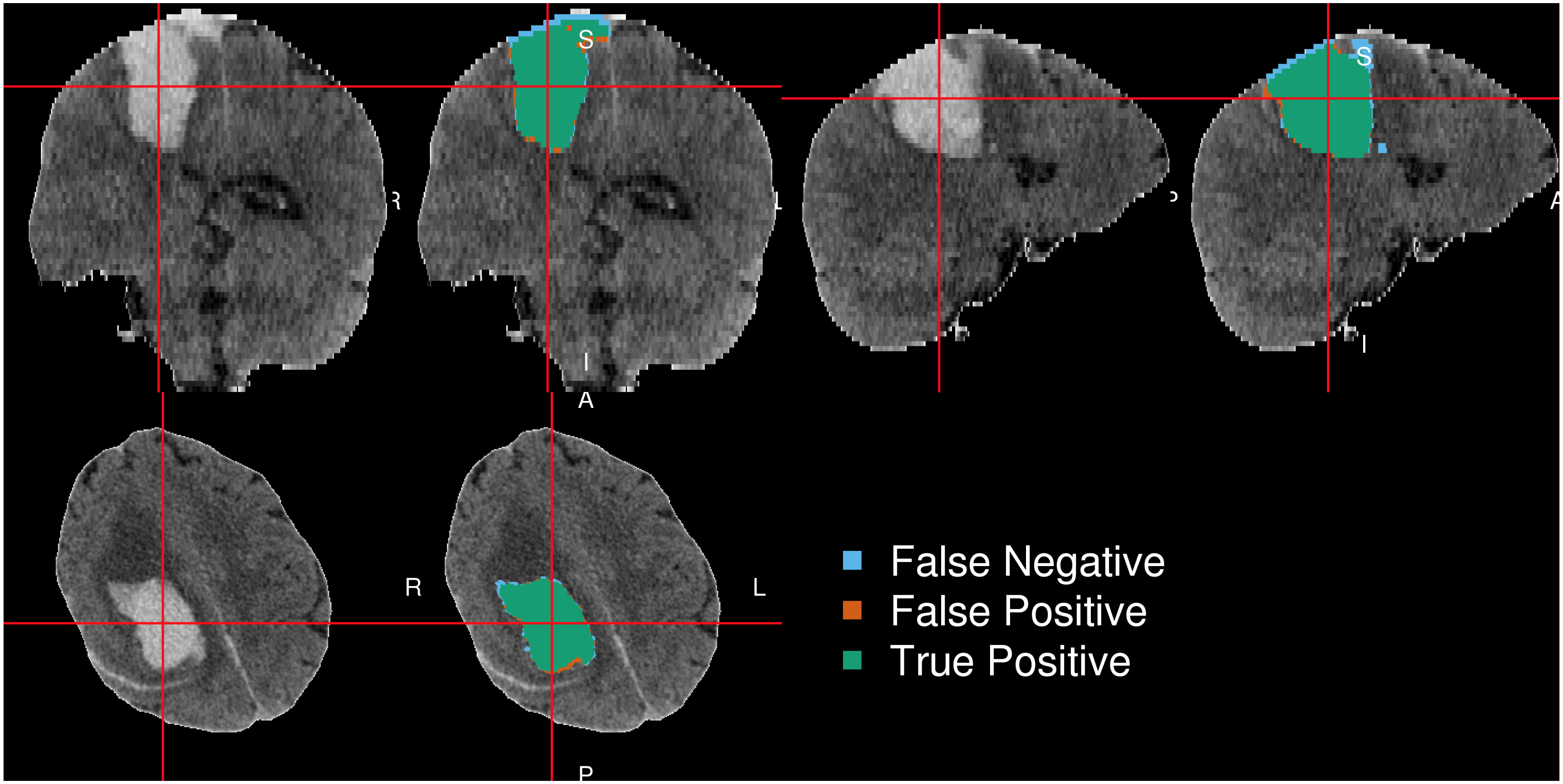


Two Segmentations: Same Volume





Patient with Median Overlap



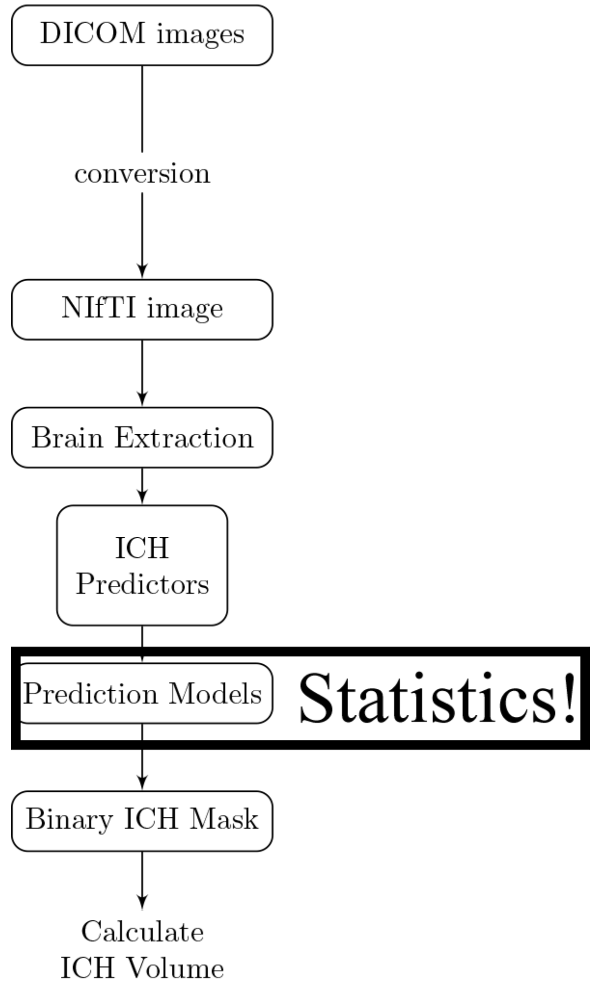
Conclusions of Stroke Analyses

- We can segment ICH volume from CT scans
- The predictors seemed to be the "secret sauce"
  - not the algorithm
- Chose the Random Forest
  - PItcHPERFeCT: Primary Intracranial Hemorrhage Probability Estimation using Random Forests on CT
- R Package: https://github.com/muschellij2/ichseg
- Shiny Application http://johnmuschelli.com/ich_segment.html

Workflow for the Analysis

Statistical methods are just one part

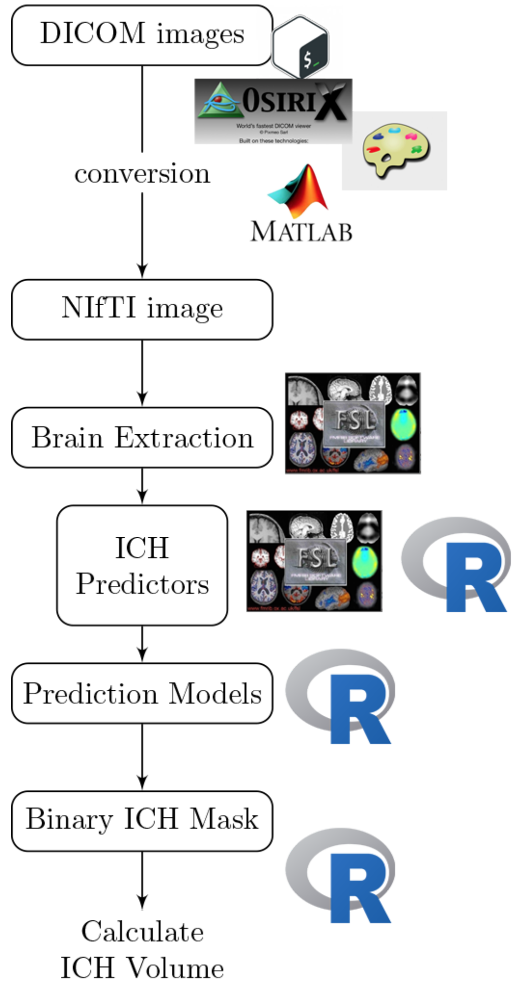
- Downstream from many steps



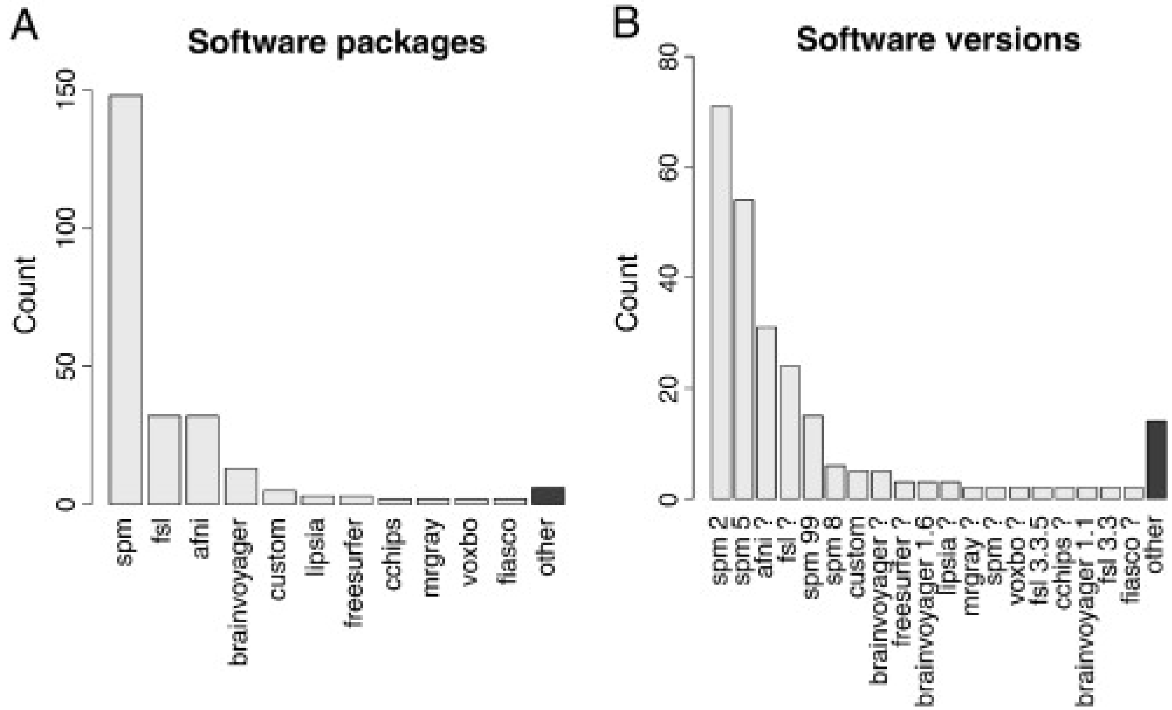
Workflow for the Analysis

Multiple pieces of software used

- all different syntax



This isn't atypical: lots of software choices



From Carp, Joshua. "The secret lives of experiments: methods reporting in the fMRI literature." Neuroimage 63.1 (2012): 289-300.

Neuroconductor:
A Framework for Reproducible Neuroimaging Analysis in R

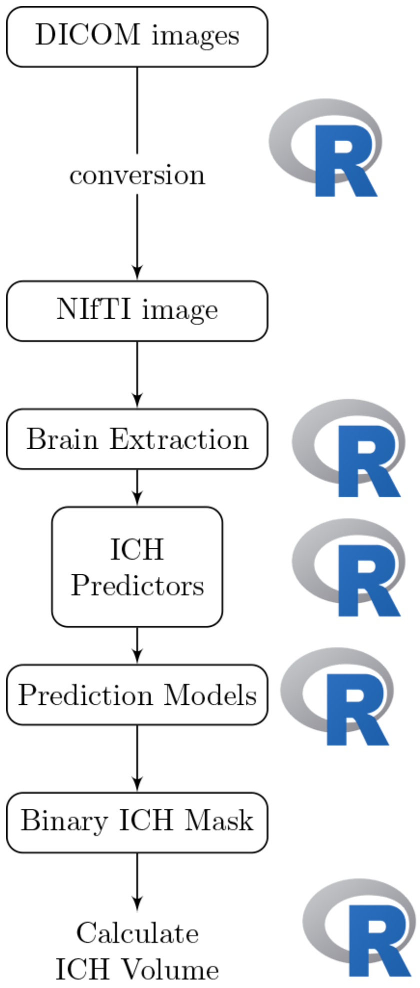
Neuroconductor Goal:

Lower the bar to entry

- all R code
  - pipeline tool
  - native R code

Statisticians involved with entire pipeline

- pipeline sensitivity analyses

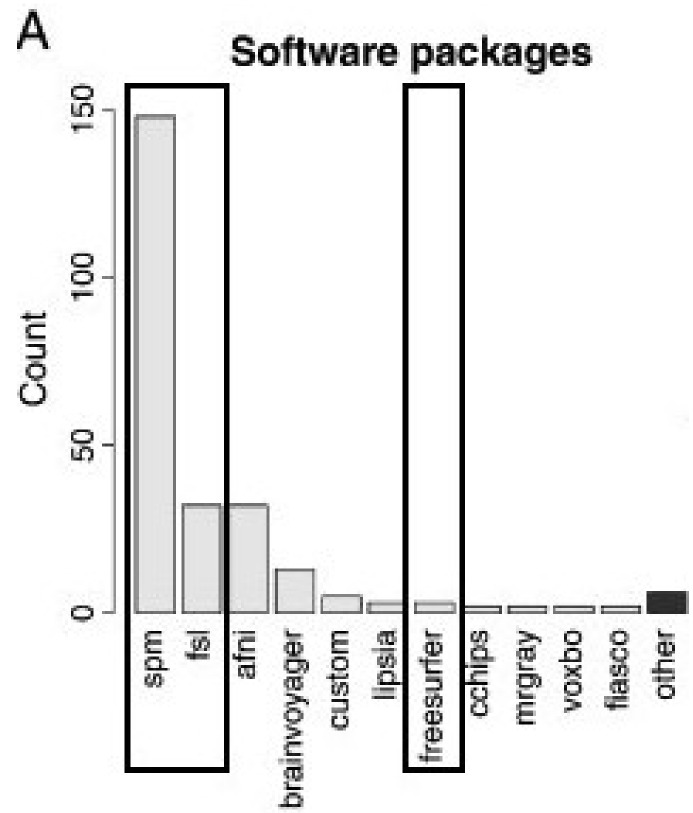


Package All Time Last Week

Current Progress (my R packages)

Number of downloads as of October 10, 2016 (`cranlogs` package):

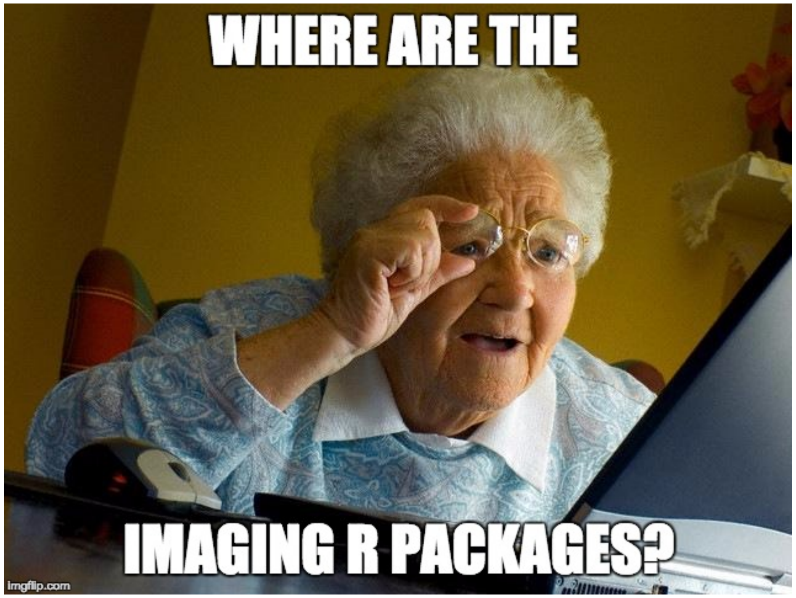
Package	All Time	Last Week
fslr	8930	139
oasis	1442	63
papayar	653	59
spm12r	1773	52
matlabr	3110	41
brainR	8764	37
freesurfer	161	35
WhiteStripe	4862	31



Neuroconductor Goal:

Centralize the packages

- [Medical Imaging Task View]
    - no tutorials
- [Neuroconductor Repository]

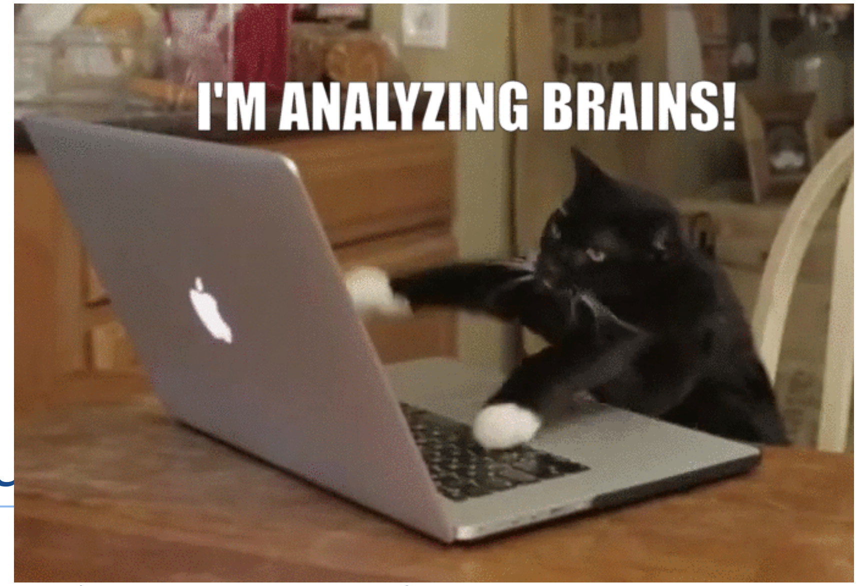


From https://imgflip.com/memegenerator/Grandma-Finds-The-Internet.

Neuroconductor Goal:

Detailed **tutorials** on how to actually perform an analysis

http://johnmuschelli.com/neu



From http://i.imgur.com/0Y1xISa.gifv.

Neuroconductor Goal:

Provide package checks / rules / **stability**

- check against other imaging software (e.g. FSL)

GitHub

Travis CI

Benefits of Neuroconductor:

Allow neuroimaging to use all the things R has to offer.

- Statistics
- Versioning and testing
- Reproducibile reports and analyses
- Shiny

Neuroconductor Downsides

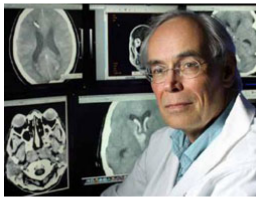
1. More control over the workflow = more work (**for us statisticians**!)
2. Users need external software (versions/installation)
3. No control over external software
    - if maintainer changes something, not much recourse
4. Need the content (buy-in from the community)

How do I get started?

Coursera Course:
Introduction to
Neurohacking In R

https://www.coursera.org/learn/ne

Thanks

Dan Hanley 

Jean-Philippe Fortin 

Ciprian Crainiceanu 

Adi Gherman 

Brian Caffo 

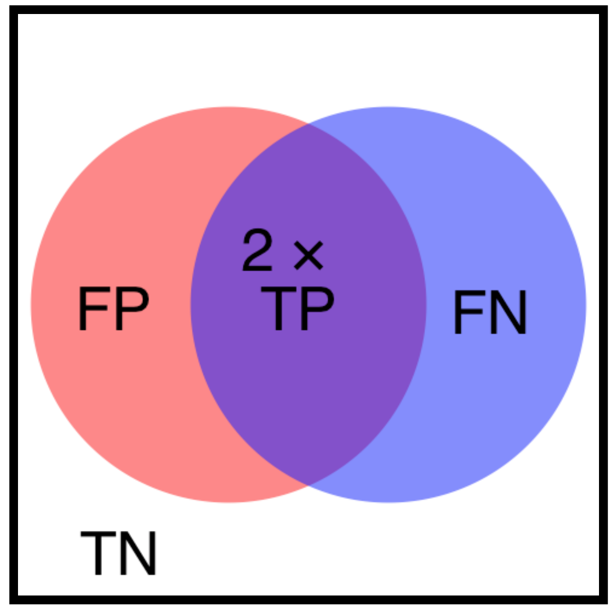
Elizabeth Sweeney 

Questions

Assessing Performance

For each manual and automated segmentation, we can calculate the following 2-by-2 table, where the cells represent number of voxels and a corresponding Venn diagram:

		Manual	
		0	1
PitCH	0	TN	FN
	1	FP	TP

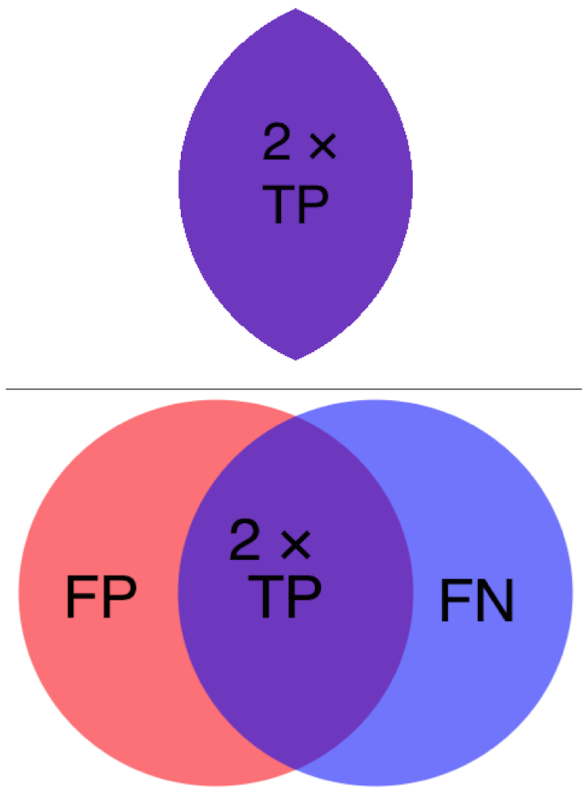


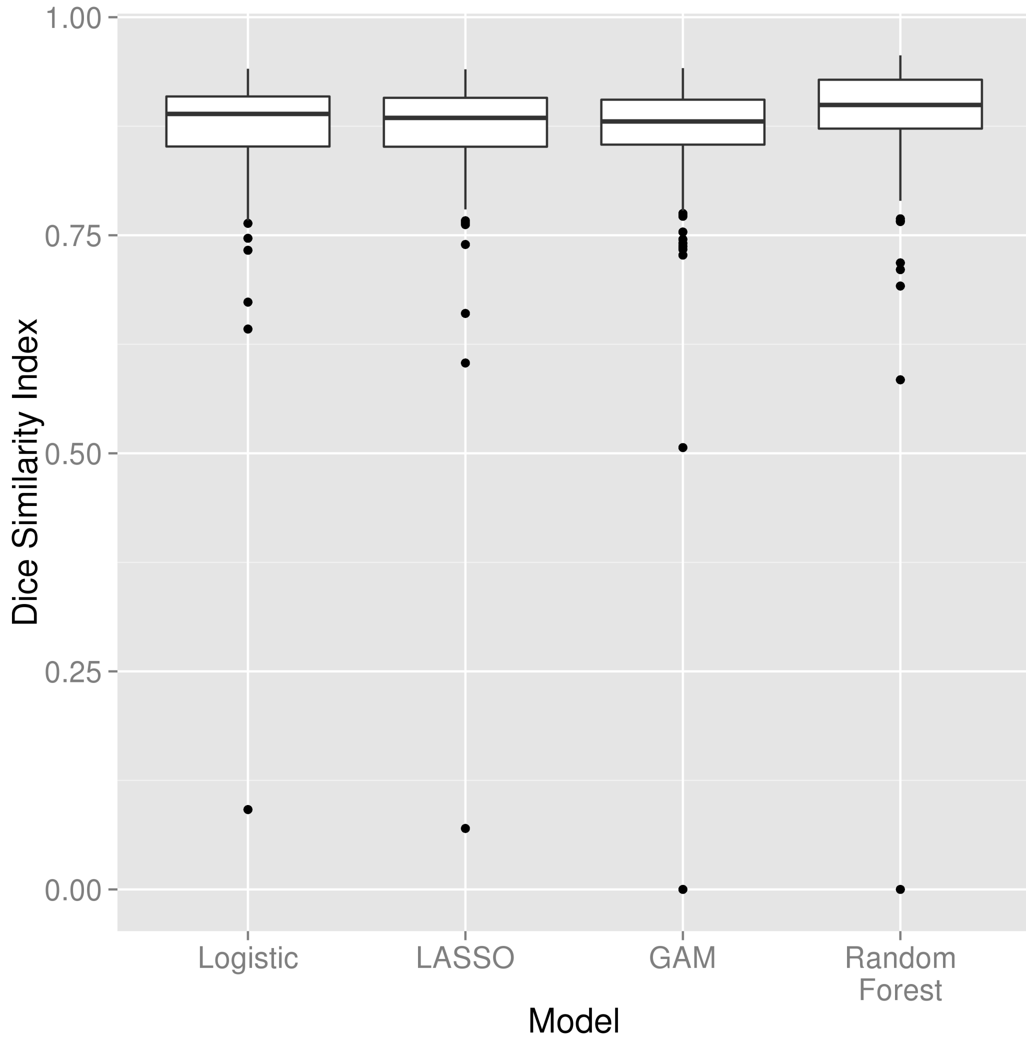
Dice Similarity

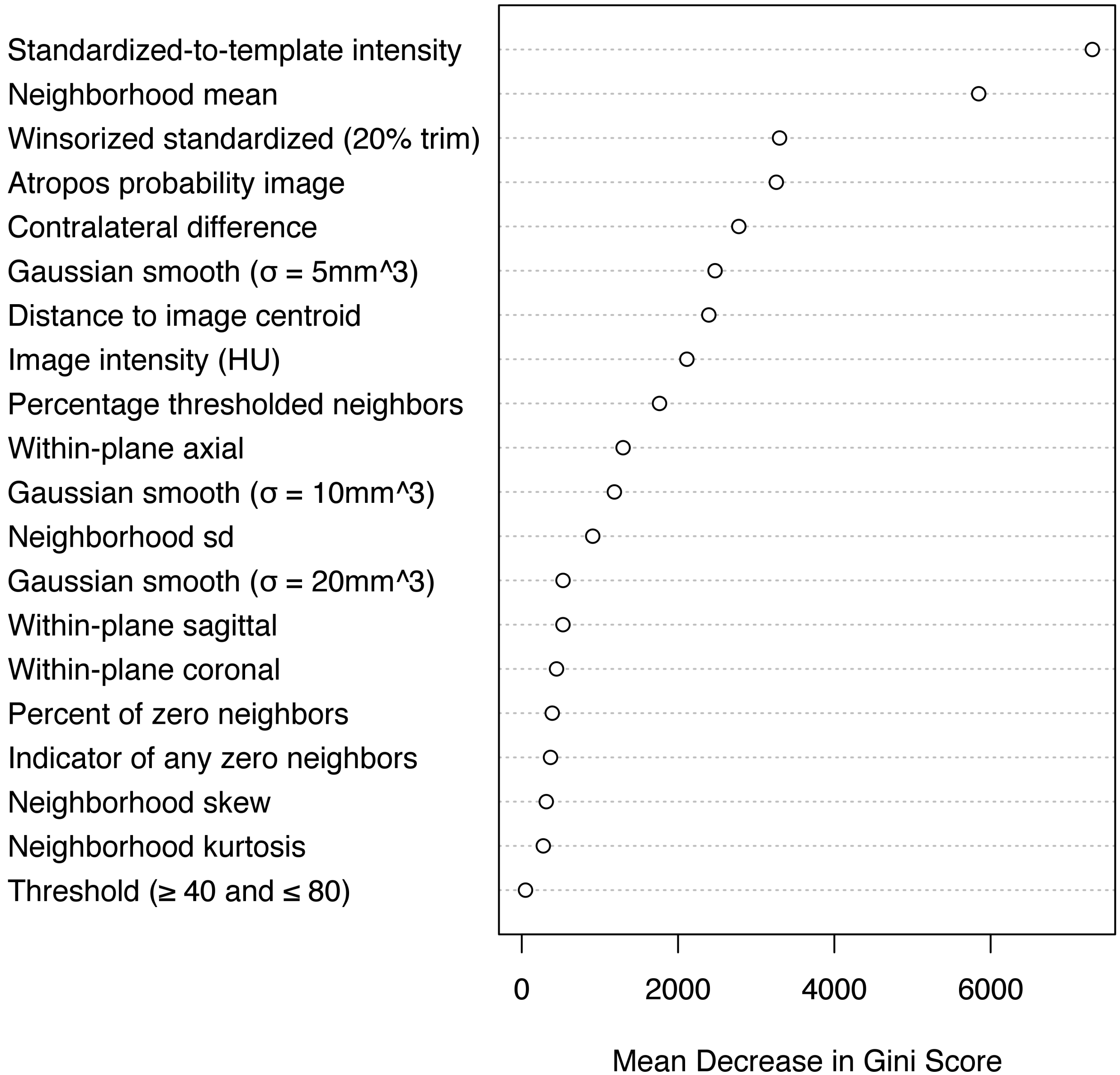
We calculate the Dice Similarity Index (DSI):

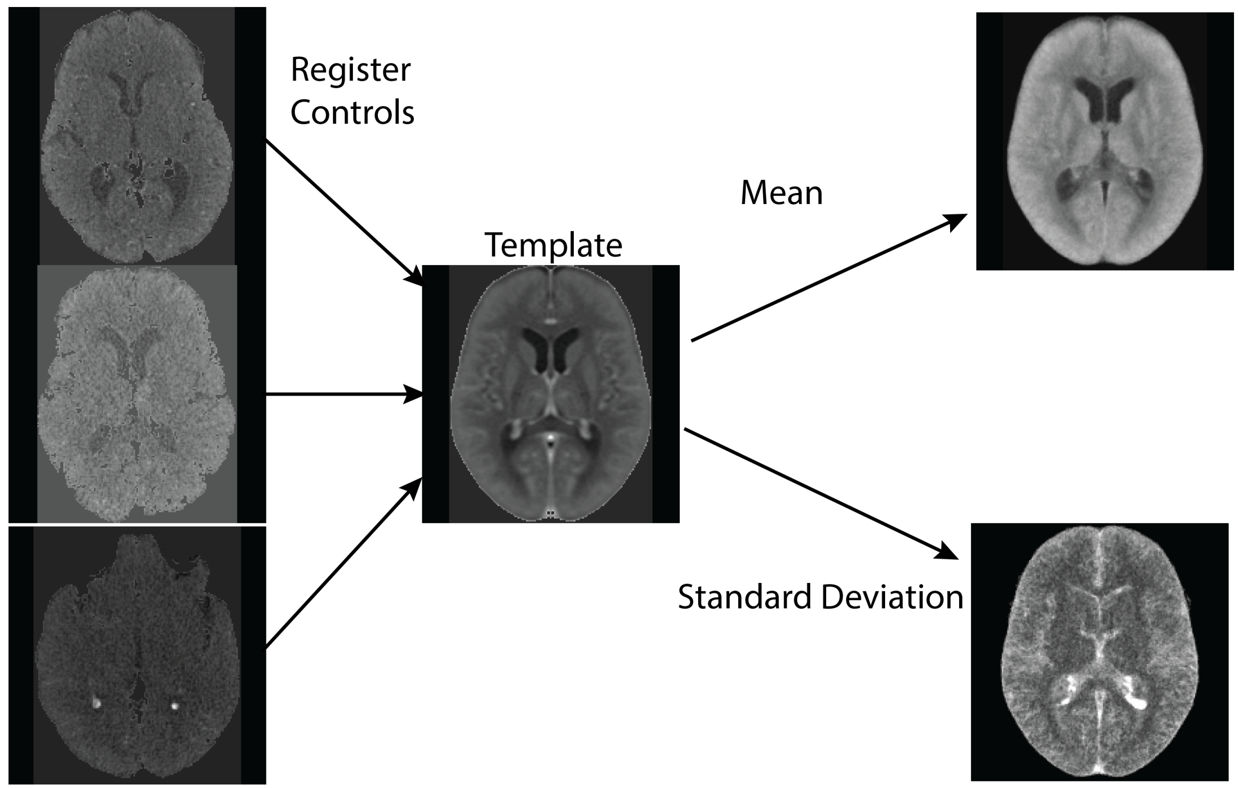
$$\frac{2 \times \#\mathrm{TP}}{2 \times \#\mathrm{TP} + \mathrm{FN} + \mathrm{FP}}$$

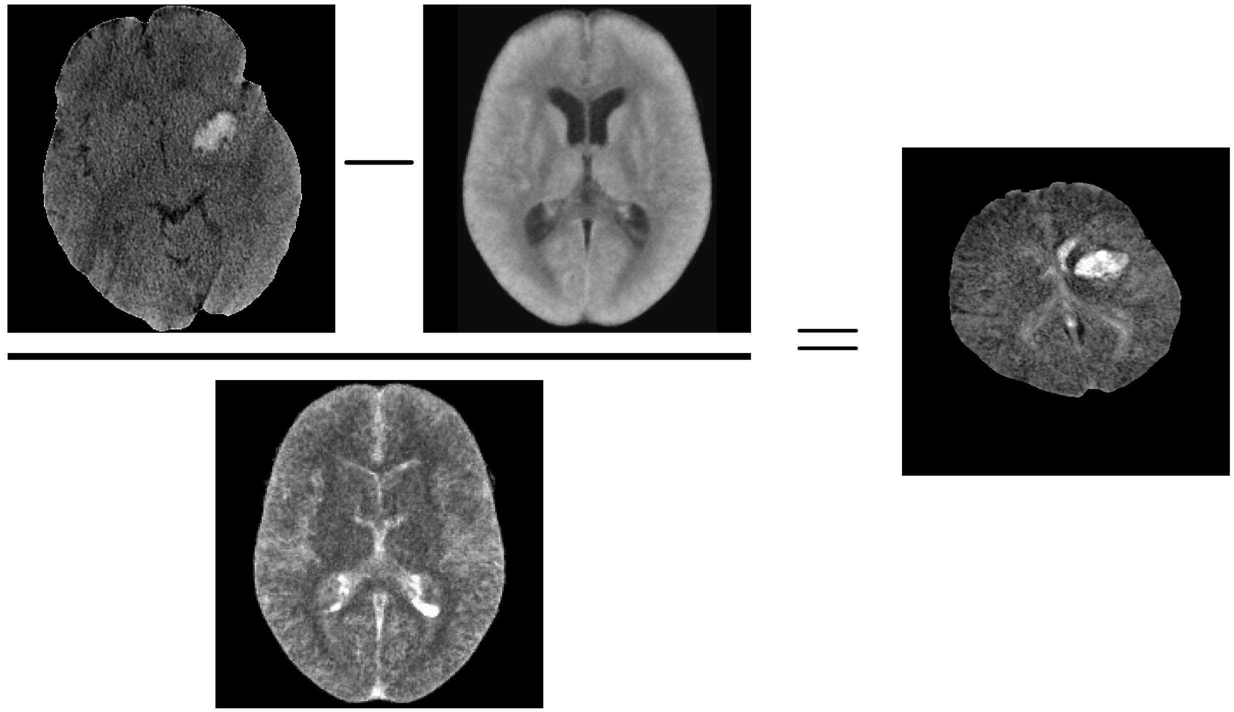
- 0 indicates no overlap
- 1 means perfect agreement




1.00
0.75
0.50
0.25
0.00
Dice Similarity Index
Logistic
LASSO
GAM
Random
Forest
Model


Standardized-to-template intensity
Neighborhood mean
Winsorized standardized (20% trim)
Atropos probability image
Contralateral difference
Gaussian smooth (σ = 5mm^3)
Distance to image centroid
Image intensity (HU)
Percentage thresholded neighbors
Within-plane axial
Gaussian smooth (σ = 10mm^3)
Neighborhood sd
Gaussian smooth (σ = 20mm^3)
Within-plane sagittal
Within-plane coronal
Percent of zero neighbors
Indicator of any zero neighbors
Neighborhood skew
Neighborhood kurtosis
Threshold (≥ 40 and ≤ 80)
0
2000
4000
6000
Mean Decrease in Gini Score


Register
Controls
Template
Mean
Standard Deviation



Under Development R Packages

1. nitrc - download data from the **NITRC repository**
2. MNITemplate* - data of a population-level "template" image
3. EveTemplate* - data of a different template image
4. kirby21 - data package with 2 subjects, 2 visits with multimodal imaging
5. rcamino - interface to analyze DTI data
6. msseg - MS lesion segmentation
7. extrantsr - pipelines for structural imaging analysis

Not started yet

1. hcp - interface with Human Connectome Project
2. afnir - R port of AFNI software (No. 2 on the chart)

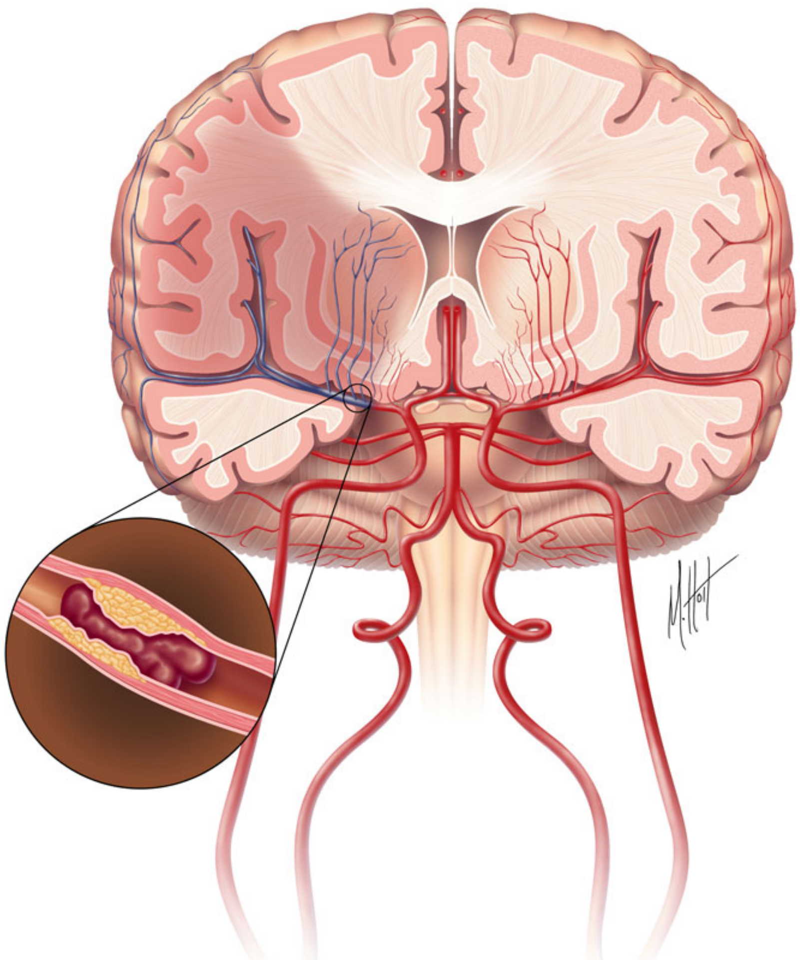
* - working with Jean-Philippe Fortin on these

The knitcitations, knitr, pander, plyr, RefManageR, rmarkdown, and tableone packages were used to create this presentation in RStudio.

```
setting  value
version  R version 3.3.1 (2016-06-21)
system   x86_64, darwin13.4.0
ui       X11
language (EN)
collate  en_US.UTF-8
tz       America/New_York
date     2016-10-12

package       * version  date
assertthat      0.1      2013-12-06
bibtex          0.4.0    2014-12-31
bitops          1.0-6    2013-08-17
class           7.3-14   2015-08-30
devtools        1.12.0   2016-06-24
digest          0.6.10   2016-08-02
DT              0.2      2016-08-09
e1071           1.6-7    2015-08-05
evaluate        0.9      2016-04-29
formatR         1.4      2016-05-09
htmltools       0.3.6    2016-09-26
htmlwidgets     0.7      2016-09-26
httr            1.2.1    2016-07-03
jsonlite        1.1      2016-09-14
knitcitations * 1.0.7.1  2016-02-01
knitr         * 1.14     2016-08-13
```

Ischemic Stroke



- Ischemic stroke - clot blocks oxygen/nutrients
- Tissue dies
- ≈87% of strokes

Image from http://www.strokecenter.org/patients/about-stroke/ischemic-stroke/

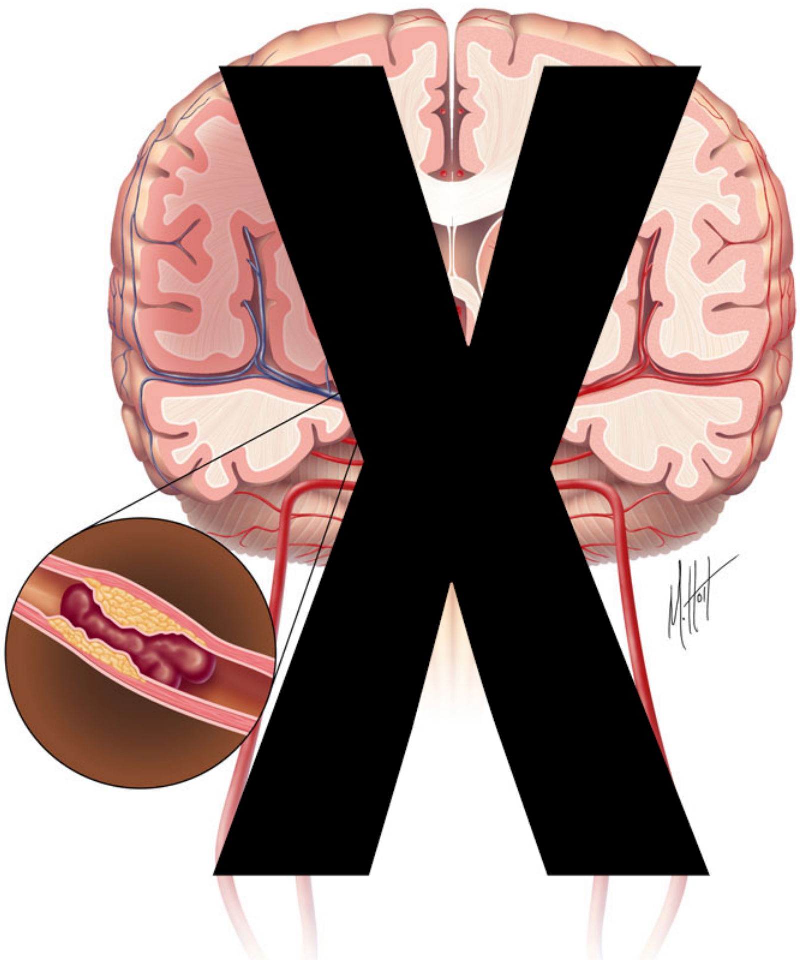
Neuroconductor Goal:

From https://memegenerator.net/Everywhere-Toy-Story.

R **Data** packages for testing

- getting data is a common problem
- developers can test packages
- Kirby21 Data:
  - https://github.com/muschellij2/kirby21.t1
  - https://github.com/muschellij2/kirby21.t2
  - https://github.com/muschellij2/kirby21.flair
  - https://github.com/muschellij2/kirby21.fmri
  - https://github.com/muschellij2/kirby21.dti

Not Ischemic Stroke



- Ischemic stroke - clot blocks oxygen/nutrients
- Tissue dies
- ≈87% of strokes

Image from http://www.strokecenter.org/patients/about-stroke/ischemic-stroke/