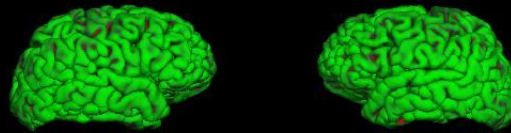
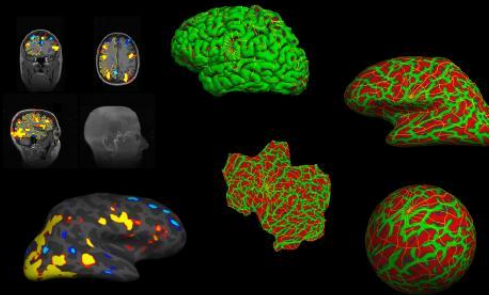


# Surface-based Group Analysis in FreeSurfer



**FreeSurfer**

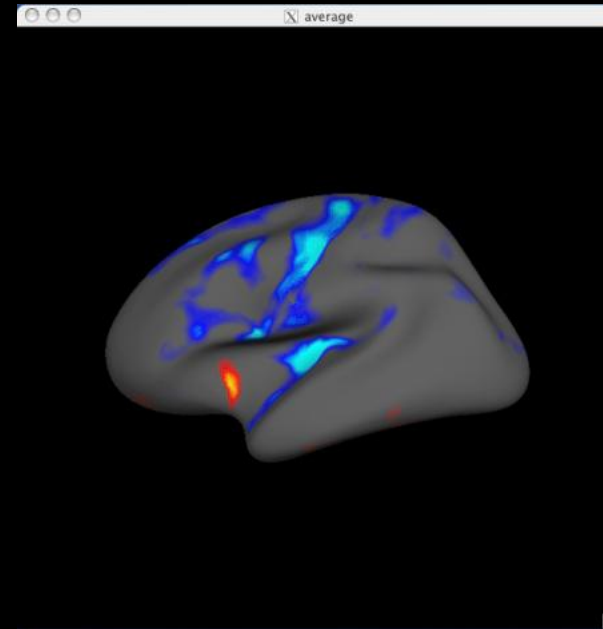
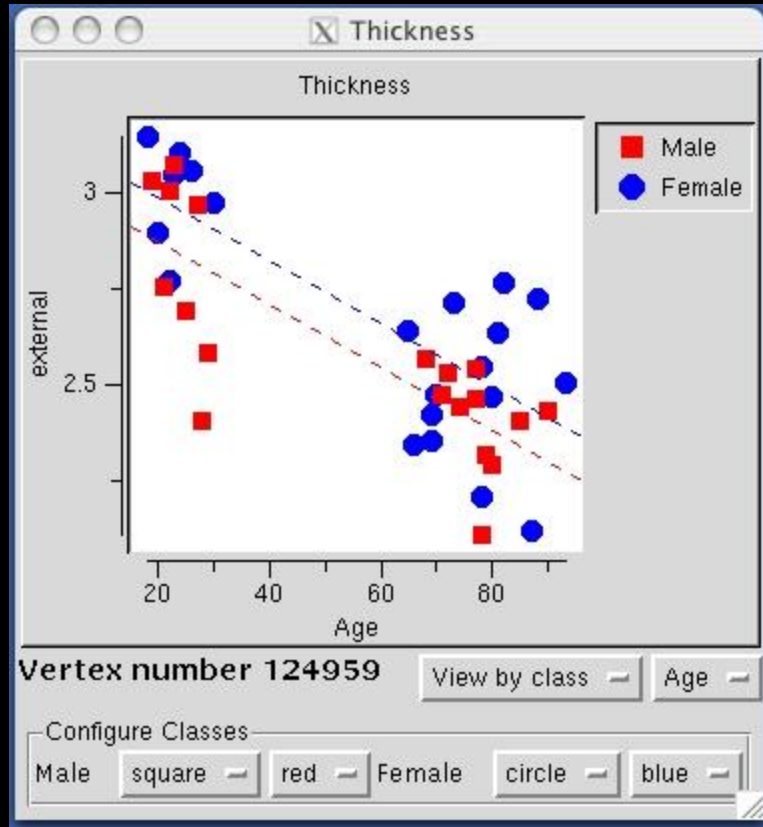


MASSACHUSETTS  
GENERAL HOSPITAL

# Outline

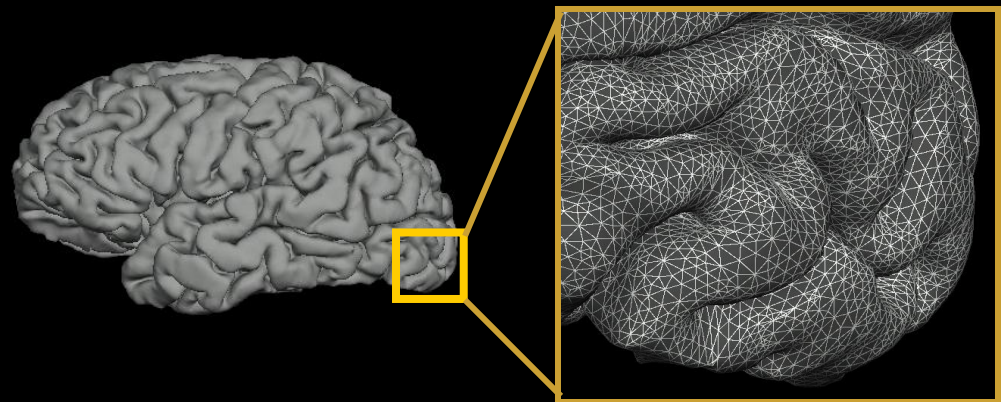
- GLM Theory
- FSGD File
- Processing Stages
- Vertex-wise Analysis
- Correction for multiple comparisons
- Interactive/Automated GUI (QDEC)

# Surface-based Study (Thickness)



# Surface-based Measures

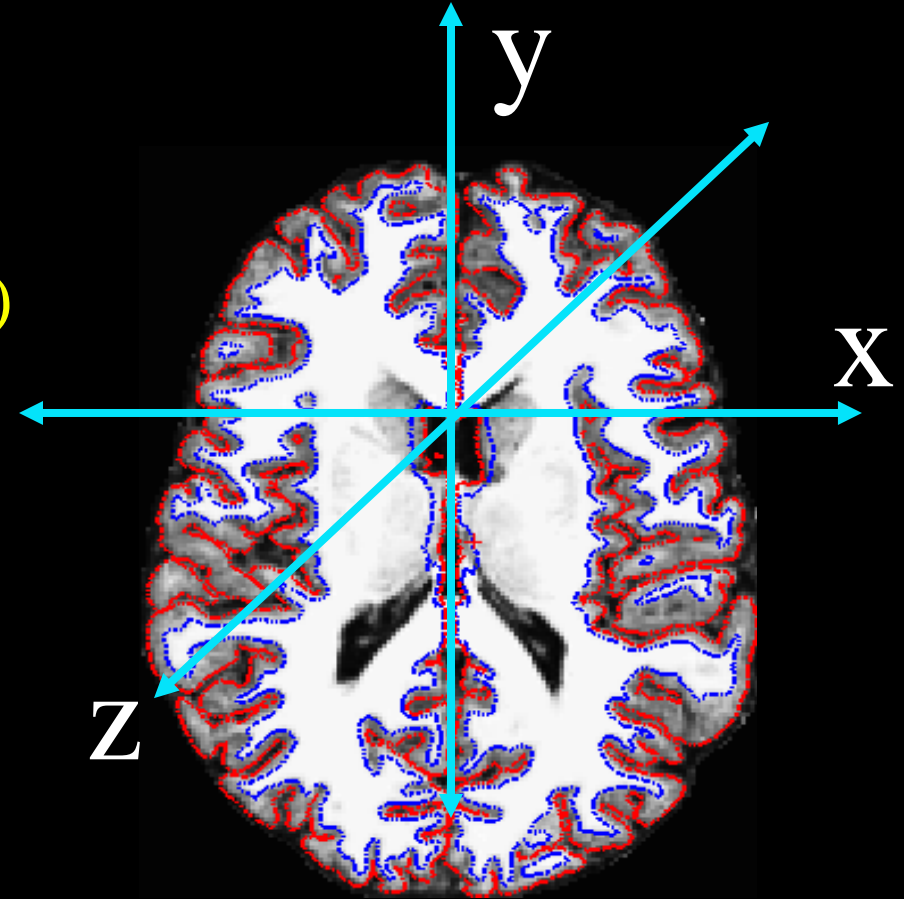
- Morphometric (eg, thickness)
- Functional
- PET
- MEG/EEG
- Diffusion (?) sampled just under the surface



# Intersubject Registration

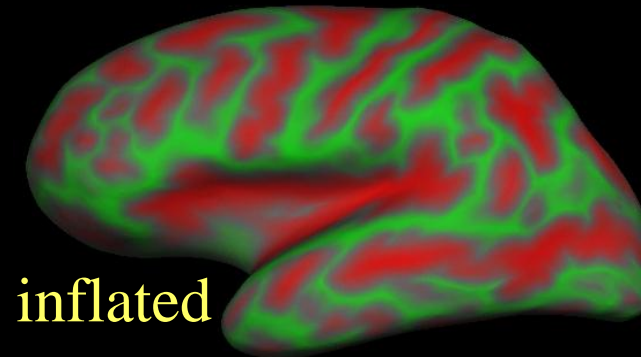
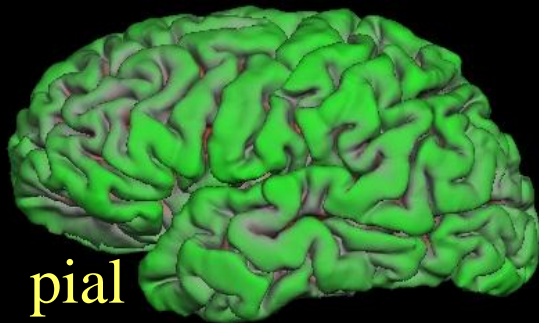
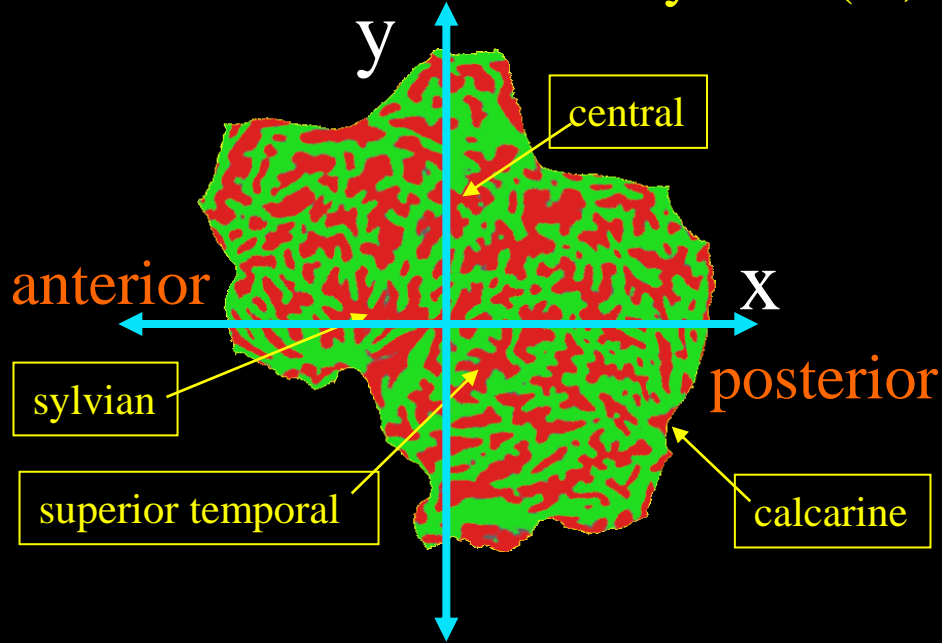
# Coordinate Systems: 3D (Volumetric)

- 3D Coordinate System
  - XYZ
  - RAS (Right-Anterior-Superior)
  - CRS (Column-Row-Slice)
  - Origin ( $XYZ=0$ , eg, AC)
  - MR Intensity at each XYZ



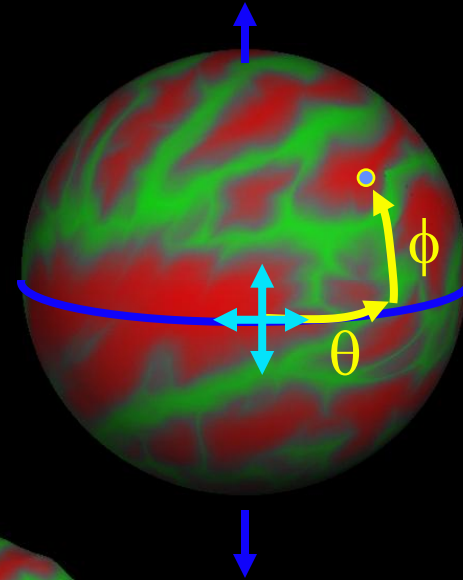
# Coordinate Systems: 2D (Surface)

Sheet: 2D Coordinate System (X,Y)



Sphere: 2D Coordinate System

- Latitude and Longitude ( $\theta$ ,  $\phi$ )
- Continuous, no cuts
- Value at each point (eg, thickness)



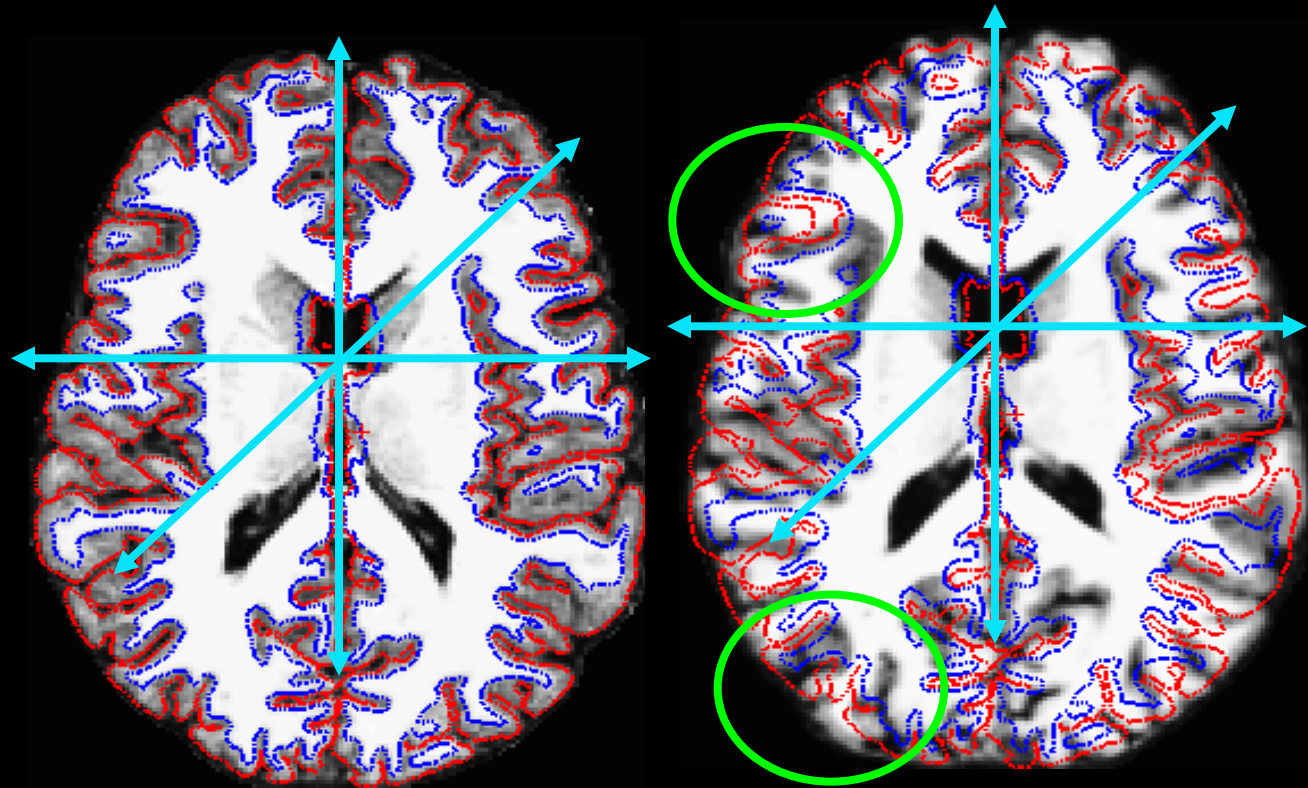
Curvature

- SULCUS (+)
- GYRUS (-)

# Volumetric Intersubject Registration

- Affine/Linear

- Translate
- Rotate
- Stretch
- Shear
- (12 DOF)



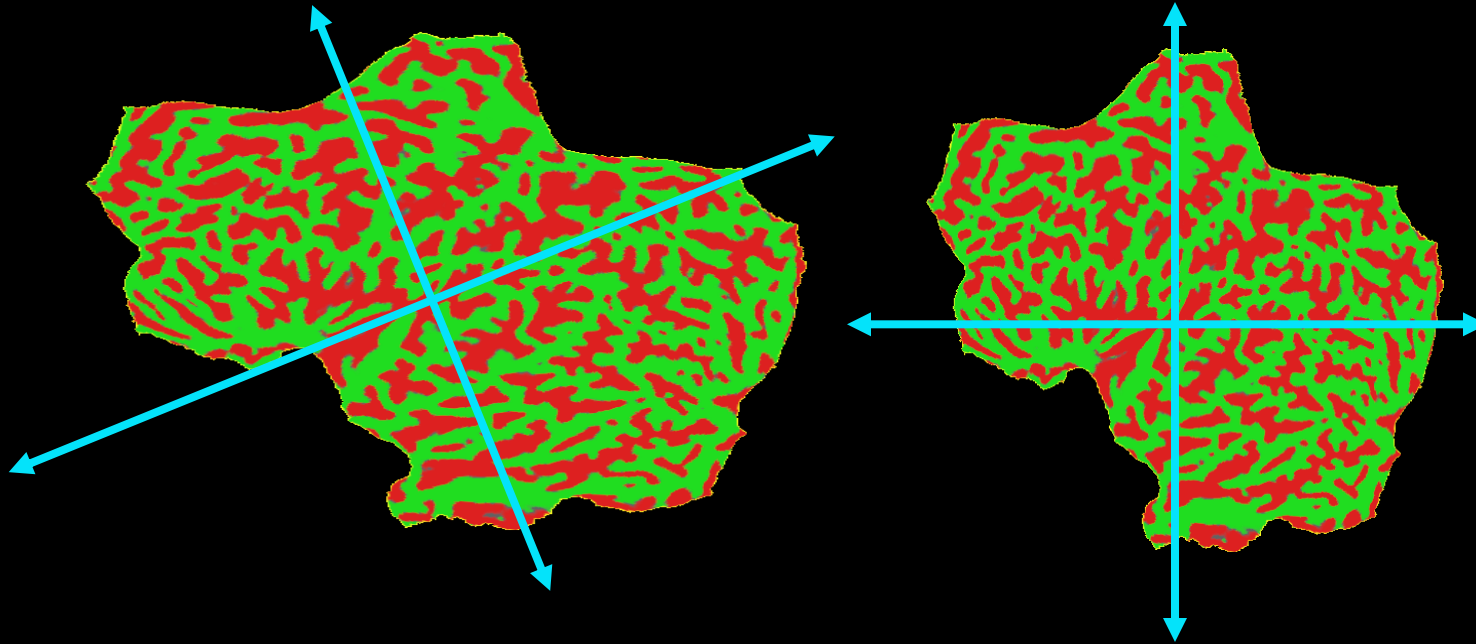
- Match Intensity, Voxel-by-Voxel
- Problems
- Can use nonlinear volumetric (cf CVS)



# Surface-based Intersubject Registration

Subject 1

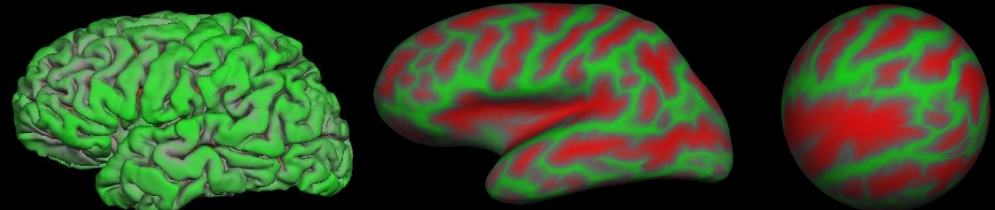
Subject 2



Curvature “Intensity”

- SULCUS (+)
- GYRUS (-)
- Codes folding pattern

- Translate, Rotate, Stretch, Shear (12 DOF)
- Match Curvature, Vertex-by-Vertex
- Nonlinear Stretching (“Morphing”) allowed (area regularization)
- Actually done on sphere
- “Spherical Morph”



# Surface-based Intersubject Registration

- Gray Matter-to-Gray Matter (it's all gray matter!)
- Gyrus-to-Gyrus and Sulcus-to-Sulcus
- Some minor folding patterns won't line up
- Fully automated, no landmarking needed
- Atlas registration is probabilistic, most variable regions get less weight.
- Done automatically in recon-all

# GLM Theory: Design Matrix

$$\begin{bmatrix} y_{11} \\ y_{12} \\ y_{13} \\ y_{21} \\ y_{22} \\ y_{23} \end{bmatrix} = \begin{matrix} & \begin{matrix} \text{G1} & \text{G2} & \text{G1*Age} & \text{G2*Age} \end{matrix} \\ \begin{bmatrix} 1 & 0 & 20 & 0 \\ 1 & 0 & 25 & 0 \\ 1 & 0 & 30 & 0 \\ 0 & 1 & 0 & 18 \\ 0 & 1 & 0 & 23 \\ 0 & 1 & 0 & 35 \end{bmatrix} \end{matrix} \bullet \begin{bmatrix} b_1 \\ b_2 \\ m_1 \\ m_2 \end{bmatrix}$$

Data

Design Matrix

Regression Coefficients

$$y = X * \beta$$

# GLM Theory: Contrasts

$$\begin{bmatrix} y_{11} \\ y_{12} \\ y_{13} \\ y_{21} \\ y_{22} \\ y_{23} \end{bmatrix} = \begin{array}{cc} & \begin{matrix} \text{G1} & \text{G2} & \text{G1*Age} & \text{G2*Age} \end{matrix} \\ \begin{bmatrix} 1 & 0 & 20 & 0 \\ 1 & 0 & 25 & 0 \\ 1 & 0 & 30 & 0 \\ 0 & 1 & 0 & 18 \\ 0 & 1 & 0 & 23 \\ 0 & 1 & 0 & 35 \end{bmatrix} & \bullet & \begin{bmatrix} b_1 \\ b_2 \\ m_1 \\ m_2 \end{bmatrix} \end{array}$$

$$C_{\text{Group}} = [1 \ -1 \ 0 \ 0] \quad \text{Group diff}=0?$$

$$C_{\text{Age}} = [0 \ 0 \ 0.5 \ 0.5] \quad \text{Mean slope}=0?$$

$$C_{\text{GroupXAge}} = [0 \ 0 \ +1 \ -1] \quad \text{Group/Age InterX?}$$

# Vertex-wise Analysis and Inference

$$\begin{aligned}y &= X * \beta + \varepsilon \\ \beta &= (X^T X)^{-1} X^T y \\ \gamma &= C * \beta\end{aligned}$$

$$t = \frac{C * \beta}{\sqrt{\sigma^2 C * (X^T X)^{-1} C^T}}$$

## p-value/significance

- value between 0 and 1
- closer to 0 means more significant

FreeSurfer stores p-values as  $-\log_{10}(p)$ :

- $0.1 = 10^{-1} \rightarrow \text{sig}=1$ ,  $0.01 = 10^{-2} \rightarrow \text{sig}=2$
- sig.mgh files
- Signed by sign of  $\gamma$
- p-value is for an unsigned test

## Need:

1.  $X$  (Design Matrix)
2.  $C$  (Contrast Matrix or Matrices)
3.  $y$  (Data)

# Specifying X and C

1. Manually create X – text editor
  - Manually create C
2. FSGD file – text editor
  - FreeSurfer creates X
  - Manually create C
3. QDEC – spread sheet, etc
  - FreeSurfer creates X and C

# FreeSurfer Group Descriptor (FSGD) File

- Simple text file (ascii, clear text)
- List of all subjects in the study
- Accompanying demographics



# FSGD Format

GroupDescriptorFile 1

Class Male

Class Female

Variables

		Age	Weight	IQ
Input <b>bert</b>	Male	10	100	1000
Input <b>fred</b>	Male	15	150	1500
Input <b>jenny</b>	Female	20	200	2000
Input <b>margaret</b>	Female	25	250	2500

- One Discrete Factor (Gender) with Two Levels (M&F)
- Three Continuous Variables: Age, Weight, IQ
- \$SUBJECTS\_DIR/**bert**

# FSGD $\rightarrow$ X (Automatic)

$$\begin{array}{c}
 \text{Female Group} \quad \text{Male*Age} \\
 \text{Male Group} \quad \text{Female*Age}
 \end{array}
 \quad
 \mathbf{X} =
 \begin{bmatrix}
 1 & 0 & 10 & 0 & 100 & 0 & 1000 & 0 \\
 1 & 0 & 15 & 0 & 150 & 0 & 1500 & 0 \\
 0 & 1 & 0 & 20 & 0 & 200 & 0 & 2000 \\
 0 & 1 & 0 & 25 & 0 & 250 & 0 & 2500
 \end{bmatrix}$$

Age
Weight
IQ

Manual

$$\mathbf{C} = \begin{bmatrix} \underbrace{-1 \quad 1} & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

Tests for the difference in intercept/offset between groups

$$\mathbf{C} = \begin{bmatrix} 0 & 0 & \underbrace{-1 \quad 1} & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

Tests for the difference in age slope between groups

# FSGD Terminology

Continuous Variables/Factors: Age, IQ, Volume, etc

Discrete Variables/Factors: Gender, Handedness, Diagnosis

Levels of Discrete :

Handedness: Left and Right

Gender: Male and Female

Diagnosis: Normal, MCI, AD

Cannot specify Discrete Factors in FSGD

Group or Class: Specification of All Discrete Factors:

- Left-handed Male MCI
- Right-handed Female Normal

# Number of Regressors

Each Group/Class:

- Has its own Intercept
- Has its own Slope (for each continuous variable)

$$\text{NRegressors} = \text{NClasses} * (\text{NVariables} + 1)$$

# Data (y): Thickness Study

1. `$SUBJECTS_DIR/bert/surf/lh.thickness`
2. `$SUBJECTS_DIR/fred/surf/lh.thickness`
3. `$SUBJECTS_DIR/jenny/surf/lh.thickness`
4. `$SUBJECTS_DIR/margaret/surf/lh.thickness`
5. ...

# Assemble Data: mris\_preproc

**mris\_preproc --help**

- fsgd FSGDFile** : Specify subjects thru FSGD File
- hemi lh** : Process left hemisphere
- meas thickness** : \$SUBJECTS\_DIR/subjectid/surf/hemi.thickness
- target fsaverage** : common space is subject fsaverage
- o lh.thickness.mgh** : output “volume-encoded surface file”

Lots of other options!

lh.thickness.mgh – file with thickness maps for all subjects → Input to Smoother or GLM

See also: recon-all -qcache

# Spatial Smoothing

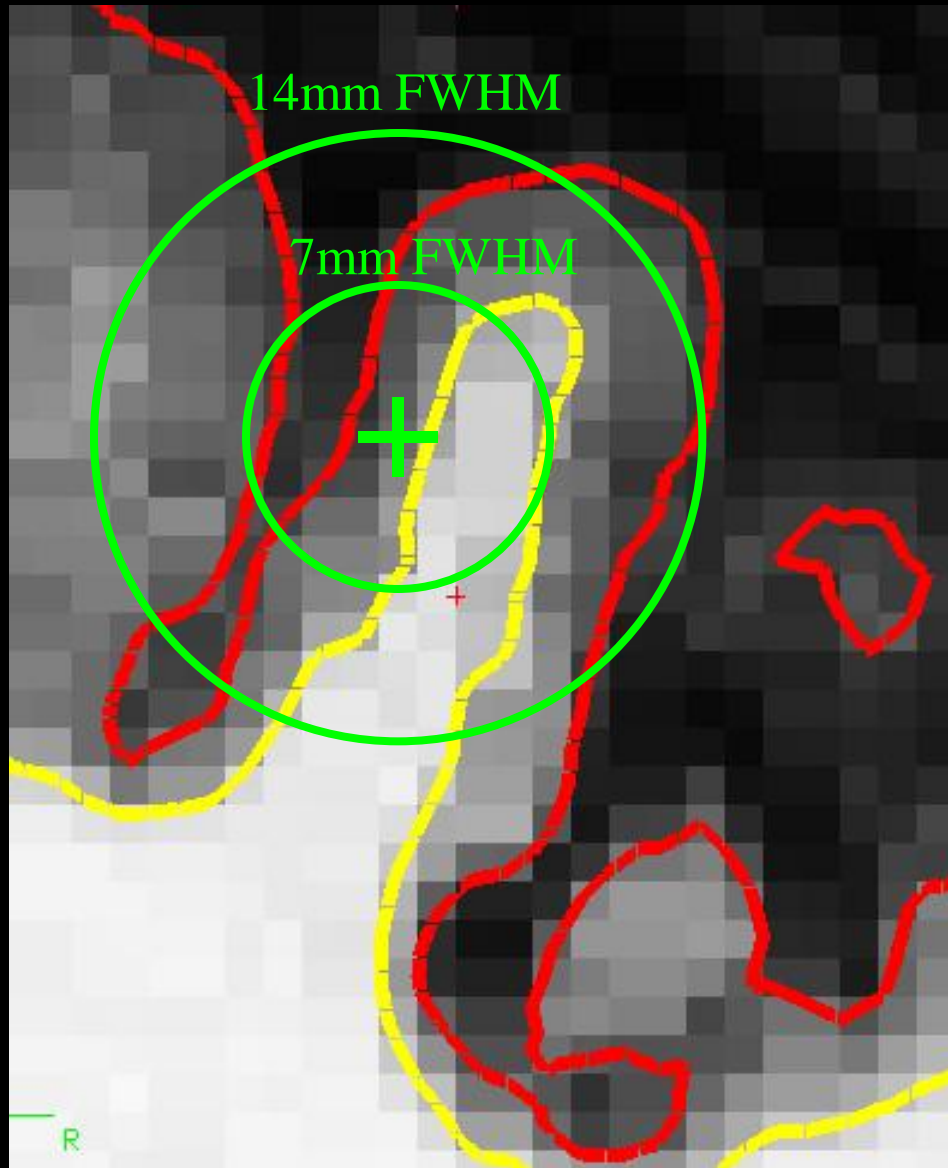
Why should you smooth?

- Might Improve CNR
- Improve intersubject registration (functional)

How much smoothing?

- Blob-size
- Typically 10-20 mm FWHM
- Surface smoothing more forgiving than volume-based

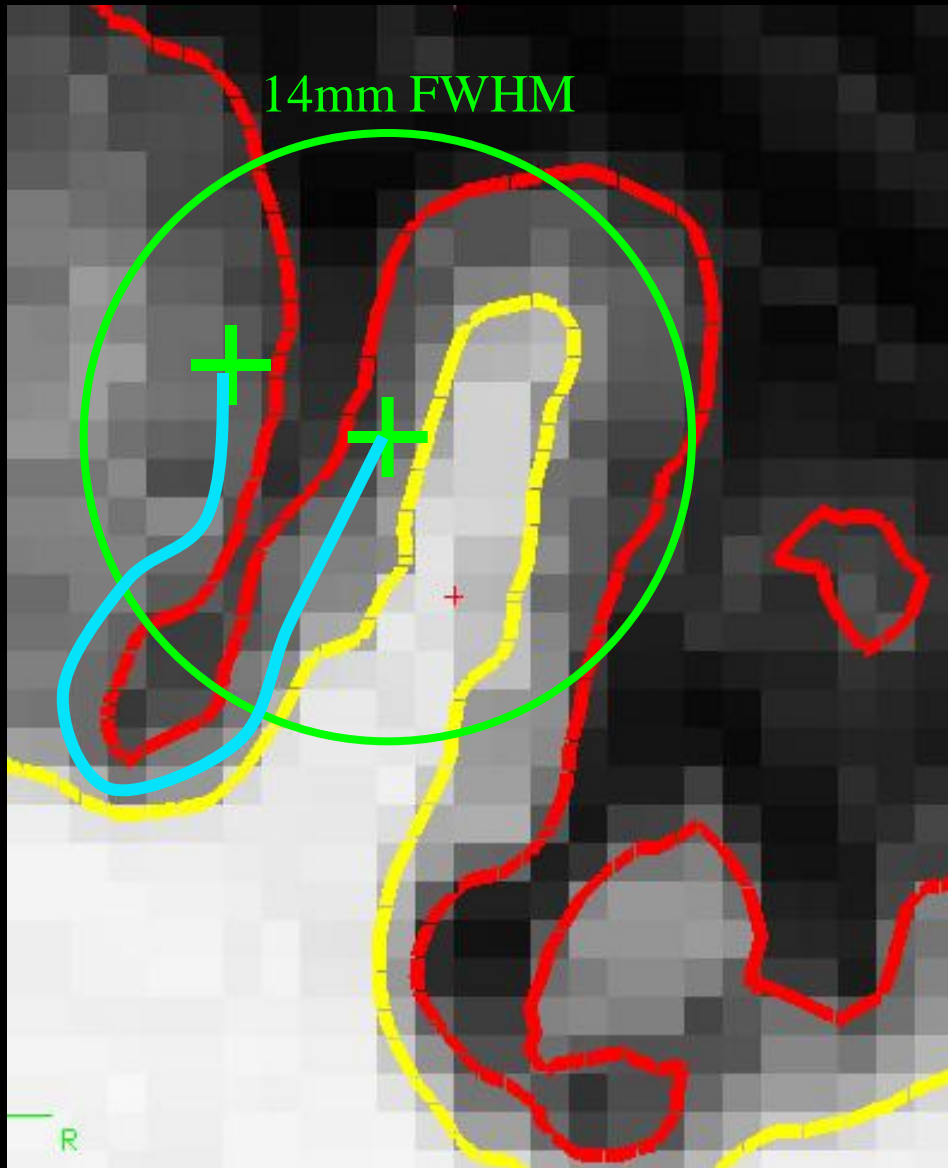
# Volume-based Smoothing



- Smoothing is averaging of nearby voxels



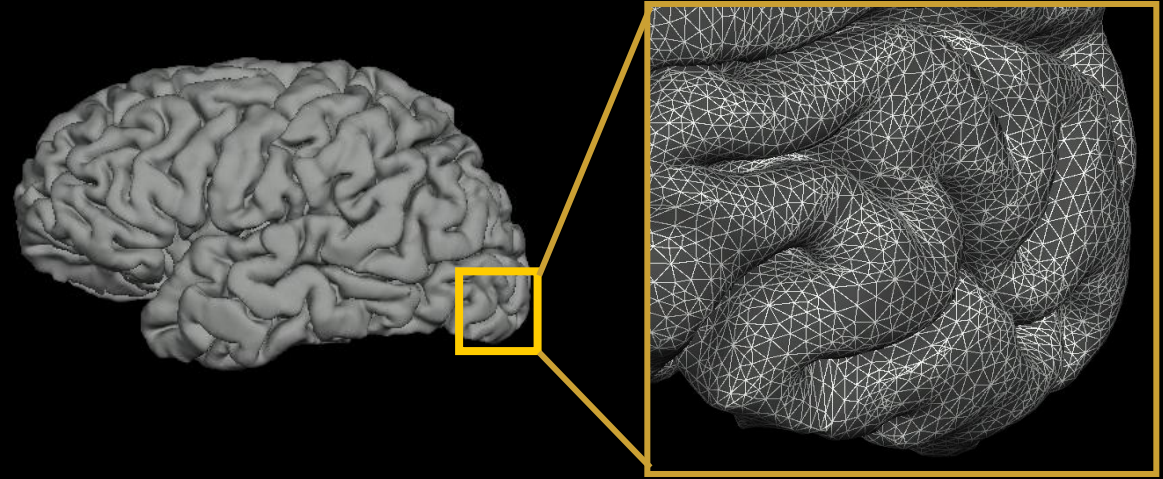
# Volume-based Smoothing



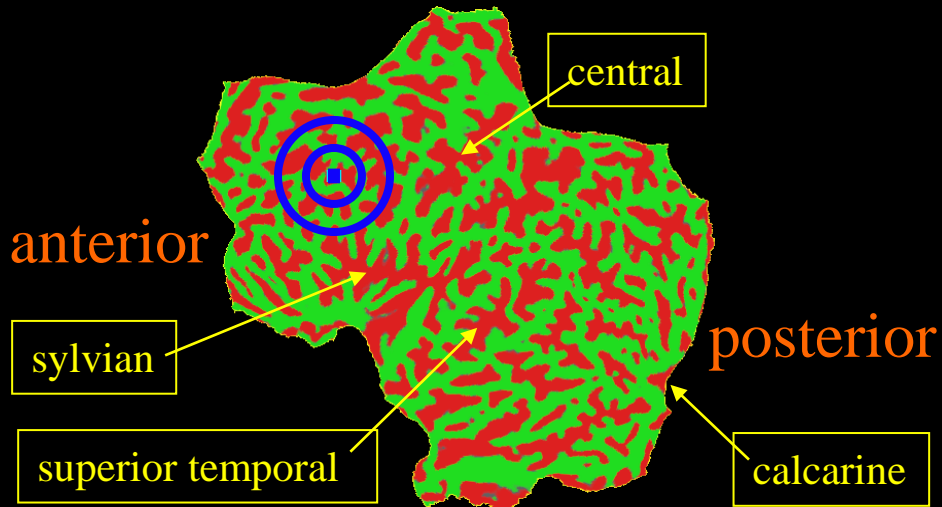
- 5 mm apart in 3D
- 25 mm apart on surface!
- Kernel much larger
- Averaging with other tissue types (WM, CSF)
- Averaging with other functional areas

# Surface-based Smoothing

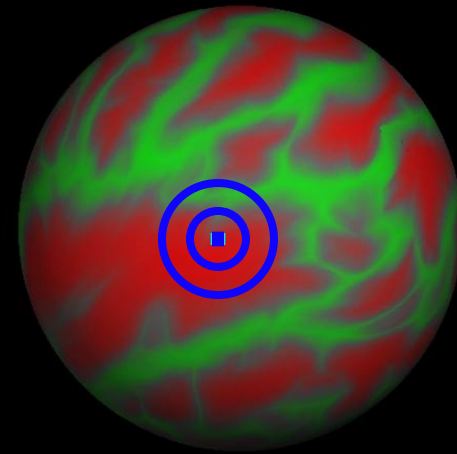
- Smoothing is averaging of nearby vertices



Sheet: 2D Coordinate System (X,Y)



Sphere: 2D Coordinate System ( $\theta, \phi$ )



# Surface Smoothing

- `mri_surf2surf --help`
- Loads `lh.thickness.mgh`
- 2D surface-based smoothing
- Specify FWHM (eg, `fwhm = 10 mm`)
- Saves `lh.thickness.sm10.mgh`
- Can be slow (~10-60min)
- `recon-all -qcache`

# Estimation and Inference: mri\_glmfit

- Reads in FSGD File and constructs  $X$
- Reads in your contrasts ( $C1, C2$ , etc)
- Loads data ( $y=lh.thickness.sm10.mgh$ )
- Fits GLM (ie, computes  $\beta$ )
- Computes contrasts ( $\gamma=C*\beta$ )
- t or F ratios, significances
- Significance  $-\log_{10}(p)$  ( $.01 \rightarrow 2, .001 \rightarrow 3$ )

# mri\_glmfit

mri\_glmfit

--y lh.thickness.sm10.mgh  
--fsgd gender\_age.txt  
--C age.mat --C gender.mat  
--surf fsaverage lh  
--cortex  
--glmdir lh.gender\_age.glmdir

mri\_glmfit --help

Creates: lh.gender\_age.glmdir/  
beta.mgh – parameter estimates  
rvar.mgh – residual error variance  
etc ...

age/

sig.mgh –  $-\log_{10}(p)$ , uncorrected

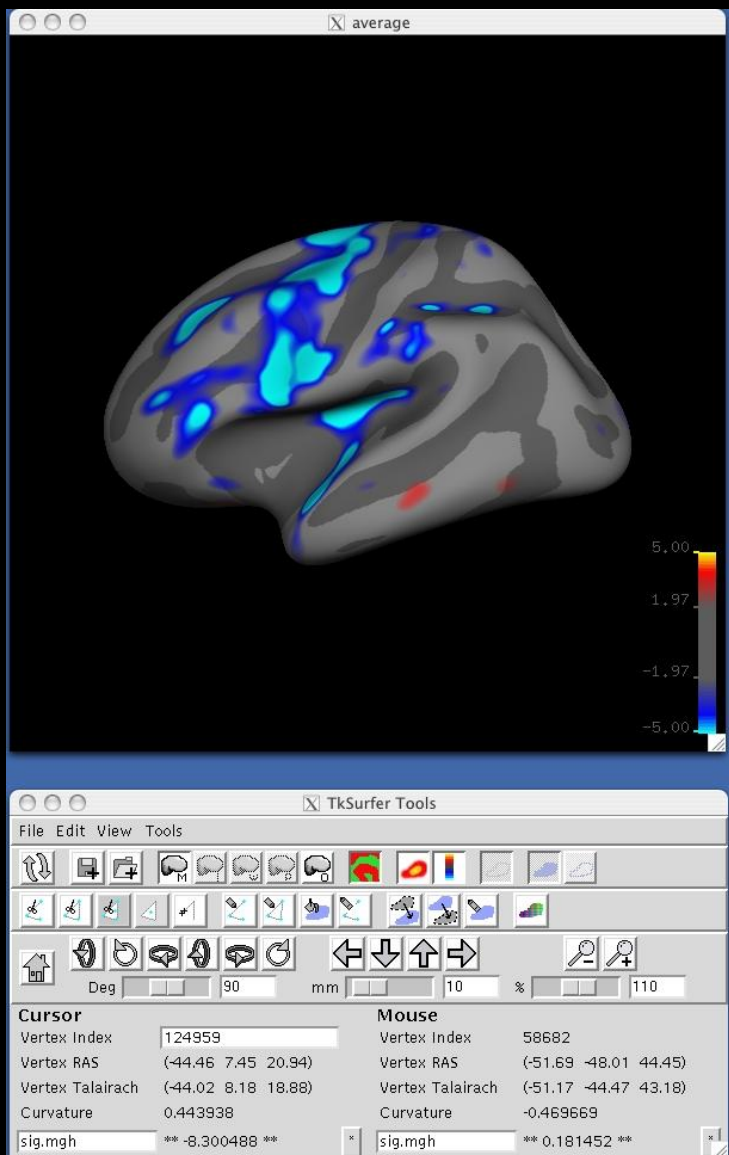
gamma.mgh, F.mgh

gender/

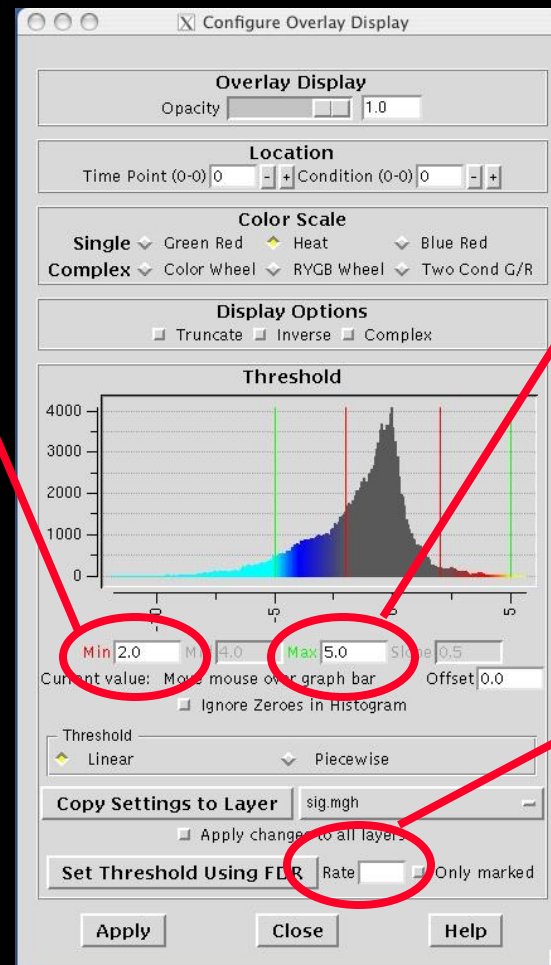
sig.mgh –  $-\log_{10}(p)$

gamma.mgh, F.mgh

# Visualization with tk-surfer



Threshold:  
 $-\log_{10}(p)$ ,  
Eg, 2=.01



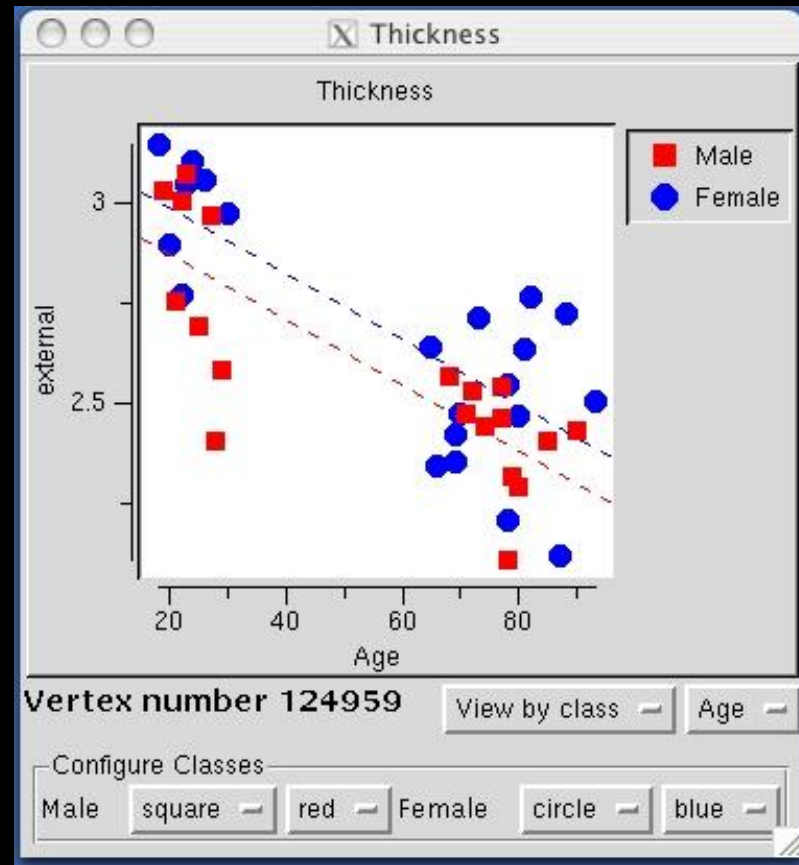
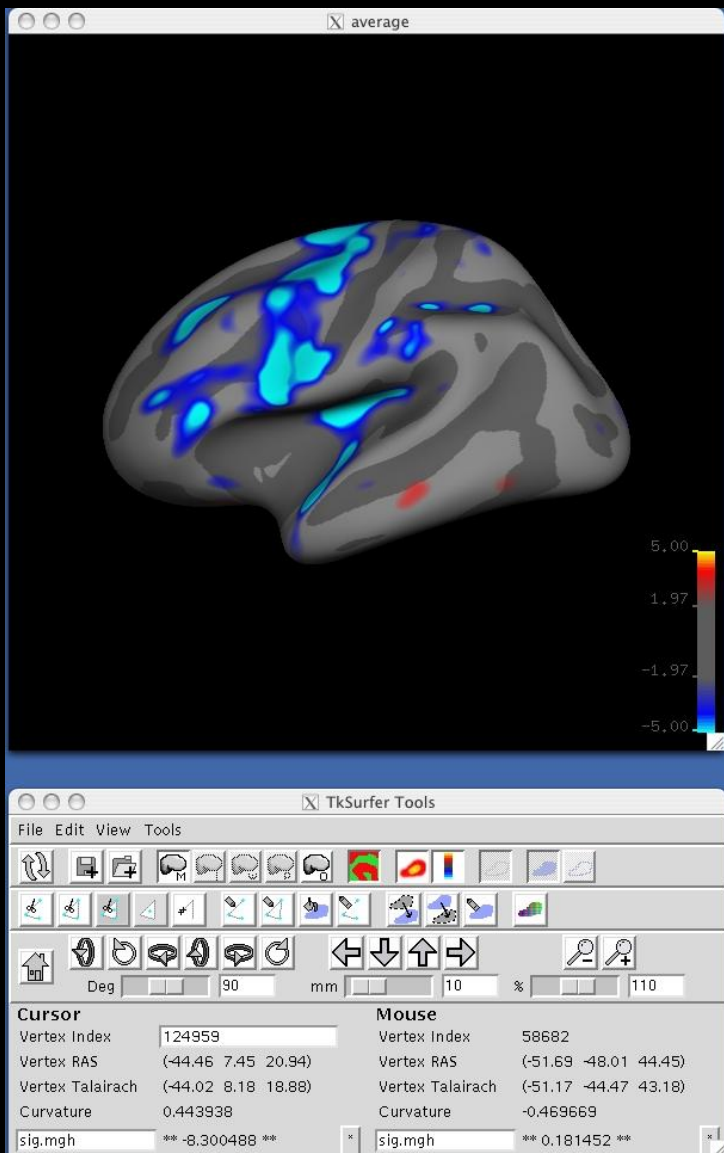
Saturation:  
 $-\log_{10}(p)$ ,  
Eg, 5=.00001

False  
Discovery  
Rate  
Eg, .01

View->Configure->Overlay

File->LoadOverlay

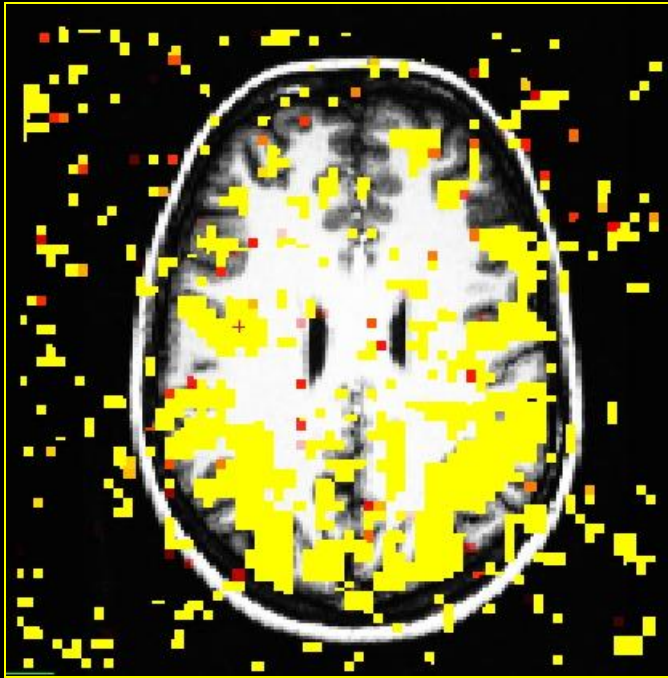
# Visualization with tksurfer



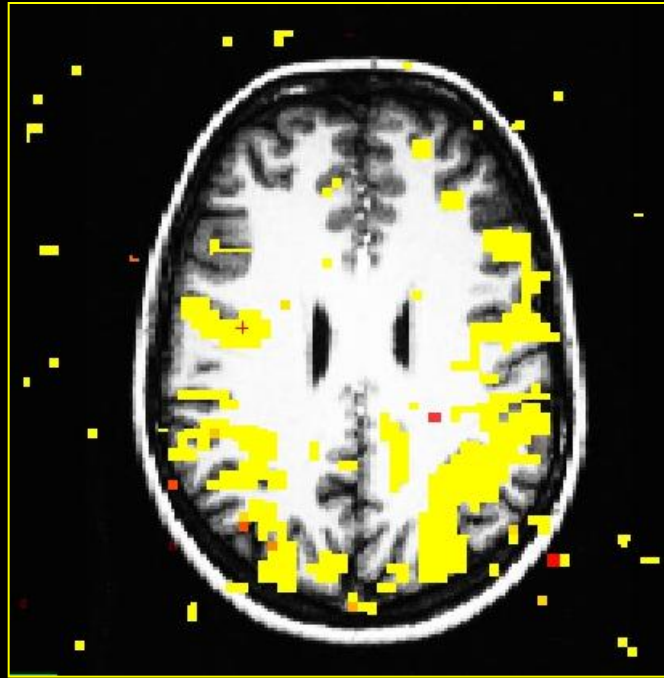
File->  
Load Group Descriptor File ...



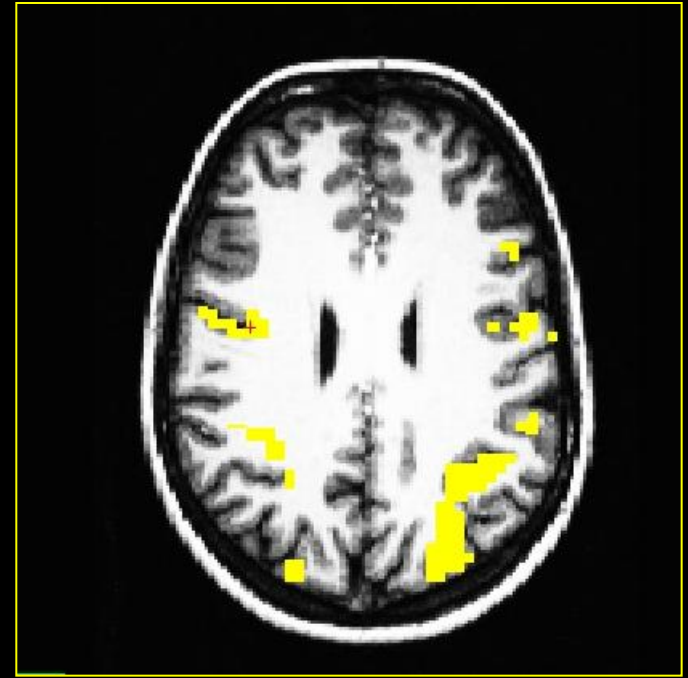
# Problem of Multiple Comparisons



$p < 0.10$



$p < 0.01$



$p < 10^{-7}$

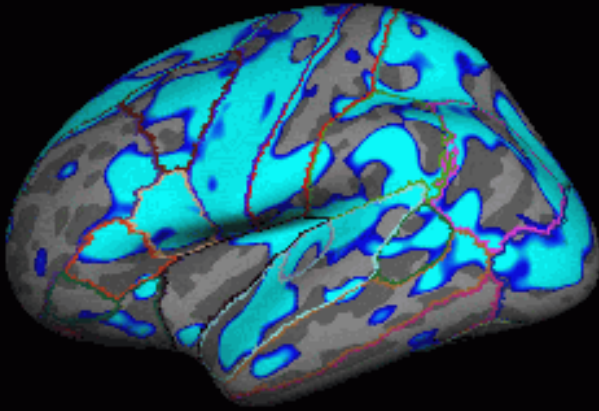


# Correction for Multiple Comparisons

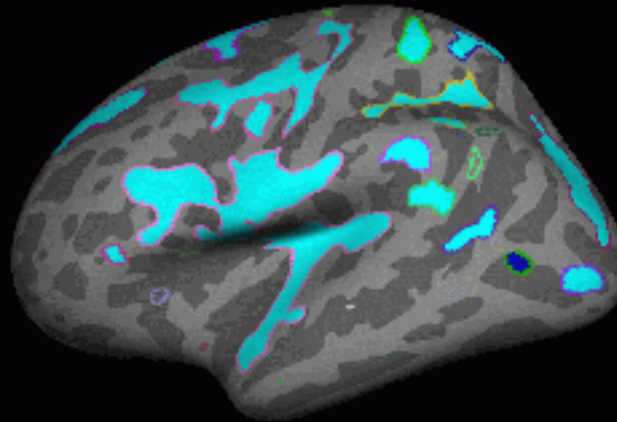
- Cluster-based
  - Monte Carlo simulation
  - Permutation Tests
  - Surface Gaussian Random Fields (GRF)
    - There but not fully tested
- False Discovery Rate (FDR) – built into tksurfer and QDEC. (Genovese, et al, NI 2002)

# Clustering

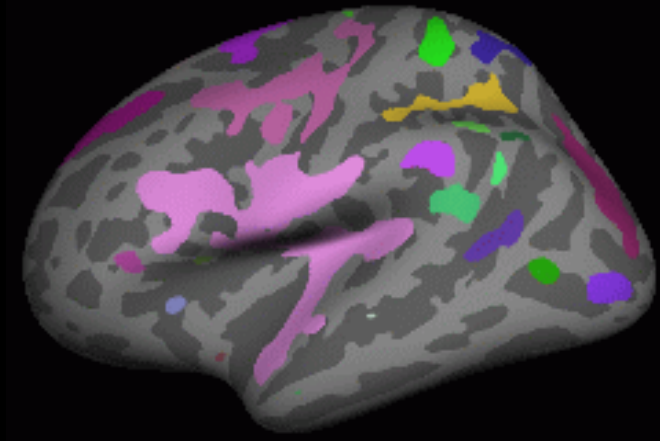
1. Choose a vertex-wise threshold
  - Eg, 2 ( $p < .01$ ), or 3 ( $p < .001$ )
  - Sign (pos, neg, abs)
2. A cluster is a group of connected (neighboring) vertices above threshold
3. Cluster has a size (area in  $\text{mm}^2$ )



$p < .01$  ( $-\log_{10}(p) = 2$ )  
Negative



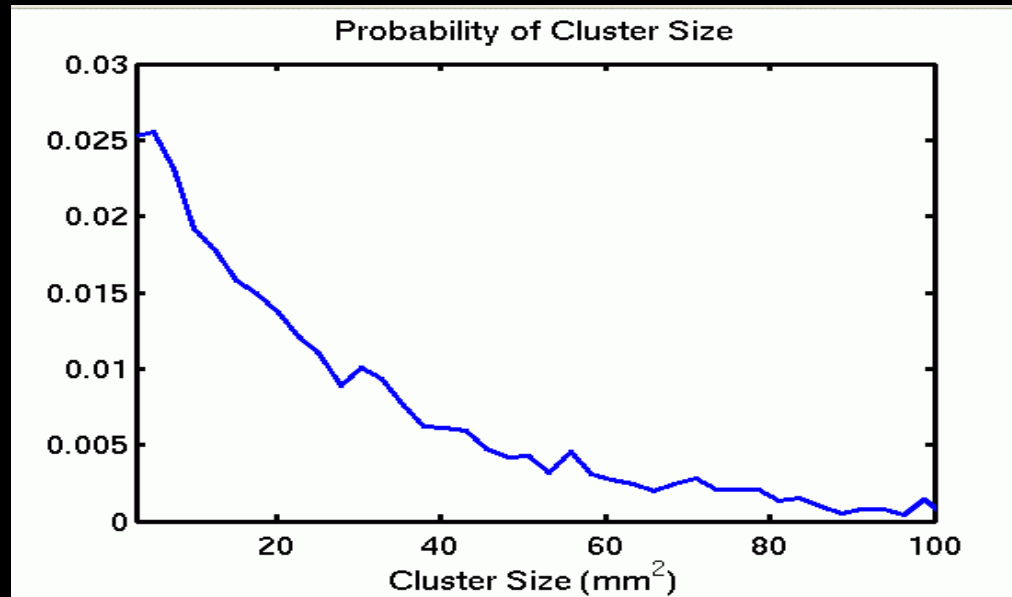
$p < .0001$  ( $-\log_{10}(p) = 4$ )  
Negative



# Cluster-based Correction for Multiple Comparisons

1. Simulate data under Null Hypothesis:
  - Synthesize Gaussian noise and then smooth (Monte Carlo)
  - Permute rows of design matrix (Permutation, orthog)
2. Analyze, threshold, cluster, max cluster size
3. Repeat 10,000 times
4. Analyze real data, get cluster sizes
5.  $P(\text{cluster}) = \# \text{MaxClusterSize} > \text{ClusterSize} / 10000$

mri\_glmfit-sim



# Command-line Processing Stages

- Model (X, FSGD) and Contrasts (C)
- Assemble Data (mris\_preproc)
  - Resample into Common Space
  - Concatenate into one file
- Smooth (mri\_surf2surf)
- Fit Model (Estimate) (mri\_glmfit)
- Correct for multiple comparisons (mri\_glmfit-sim)
- Visualize (tksurfer)

recon-all -qcache

# QDEC – An Interactive Statistical Engine GUI

Query – Select subjects based on Match Criteria

Design – Specify discrete and continuous factors

Estimate – Fit Model

Contrast – Automatically Generate Contrast Matrices

Interactive – Makes easy things easy (that used to be hard)

...a work in progress

- No Query yet
- Two Discrete Factors (Two Levels)
- Two Continuous Factors
- Surface only

# QDEC – Spreadsheet

qdec.table.dat – spreadsheet with subject information – spreadsheet can be huge!

fsid	gender	age	diagnosis	Left-Cerebral-White-Matter-Vol
011121_vc8048	Female	70	Demented	202291
021121_62313-2	Female	71	Demented	210188
010607_vc7017	Female	73	Nondemented	170653
021121_vc10557	Male	75	Demented	142029
020718_62545	Male	76	Demented	186087
020322_vc8817	Male	77	Nondemented	149810

gender.levels

Female  
Male

diagnosis.levels

Demented  
Nondemented

Discrete Factors need a  
factorname.level file

# Tutorial

## 1. Command-line Stream

- Create an FSGD File for a thickness study
- Age and Gender
- Run
  - mris\_preproc
  - mri\_surf2surf
  - mri\_glmfit
  - mri\_glmfit-sim
  - tksurfer

## 2. QDEC – same data set





# Another FSGD Example

- Two Discrete Factors
  - Gender: Two Levels (M&F)
  - Handedness: Two Levels (L&R)
- One Continuous Variable: Age

GroupDescriptorFile 1

Class MaleRight

Class MaleLeft

Class FemaleRight

Class FemaleLeft

Variables

Age

Input bert MaleLeft 10

Input fred MaleRight 15

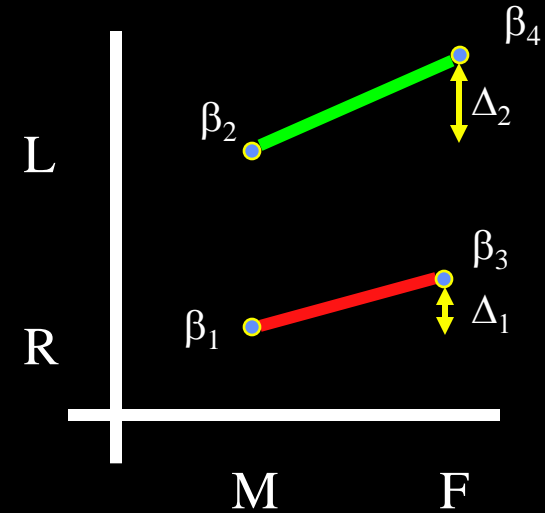
Input jenny FemaleRight 20

Input margaret FemaleLeft 25

Class = Group

# Interaction Contrast

- Two Discrete Factors (no continuous, for now)
  - Gender: Two Levels (M&F)
  - Handedness: Two Levels (L&R)
- Four Regressors (Offsets)
  - MR ( $\beta_1$ ), ML ( $\beta_2$ ), FR ( $\beta_3$ ), FL ( $\beta_4$ )



GroupDescriptorFile 1

Class MaleRight

Class MaleLeft

Class FemaleRight

Class FemaleLeft

Input bert MaleLeft

Input fred MaleRight

Input jenny FemaleRight

Input margaret FemaleLeft

$$\gamma = \Delta_1 - \Delta_2 \stackrel{?}{=} 0$$

$$\begin{aligned} \gamma &= (\beta_3 - \beta_1) - (\beta_4 - \beta_2) \\ &= -\beta_1 + \beta_2 + \beta_3 - \beta_4 \end{aligned}$$

$$C = [-1 \ +1 \ +1 \ -1]$$

# FSGDF $\rightarrow$ X

Input:

- y
- X
- C

DOSS – Different Offset, Same Slope

Female Class      Age for Males and Females  
Male Class

$$X = \begin{bmatrix} 1 & 0 & 10 & 100 & 1000 \\ 1 & 0 & 15 & 150 & 1500 \\ 0 & 1 & 20 & 200 & 2000 \\ 0 & 1 & 25 & 250 & 2500 \end{bmatrix}$$

$$C = \begin{bmatrix} -1 & 1 & 0 & 0 & 0 \end{bmatrix} \rightarrow \text{Same test, different vector}$$

#Regressors =  $N_v + N_c = 3 + 2 = 5 \rightarrow$  Fewer regressors than DODS

DOF = #Rows - #Regressors