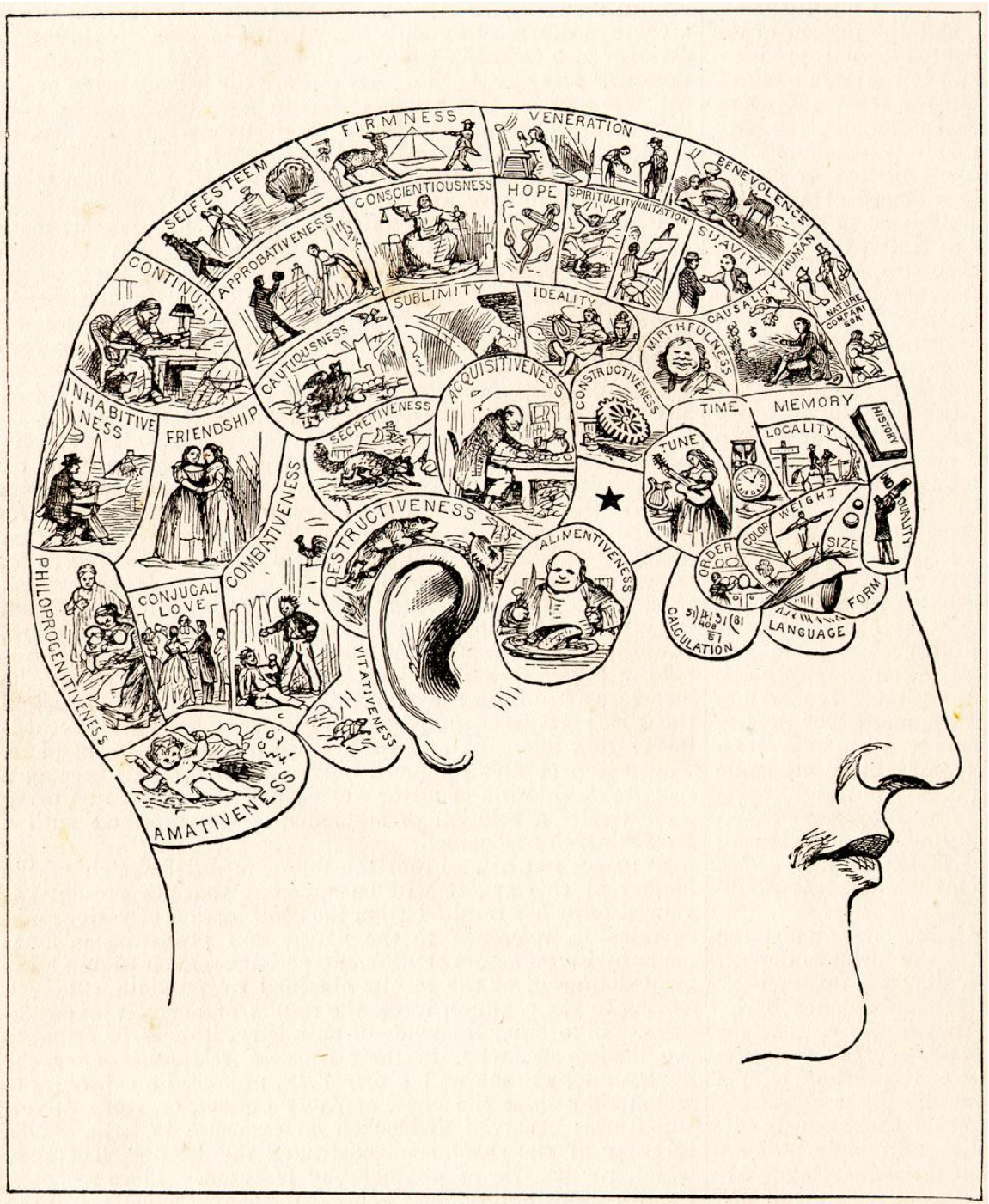
Prediction of Brain Age Using MRI data

Albert C. Yang, M.D., Ph.D.

Institutes of Brain Science/Digital Medicine Center
National Yang-Ming University

Apr 9, 2020

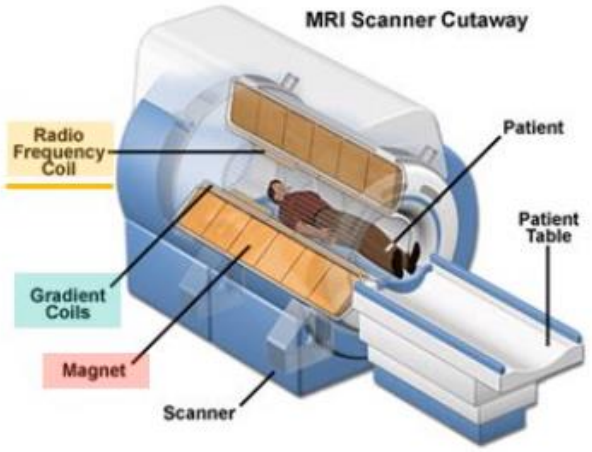
accyang@gmail.com



Phrenological Chart of the Faculties

Magnetic Resonance Imaging

Magnetic Resonance Imaging

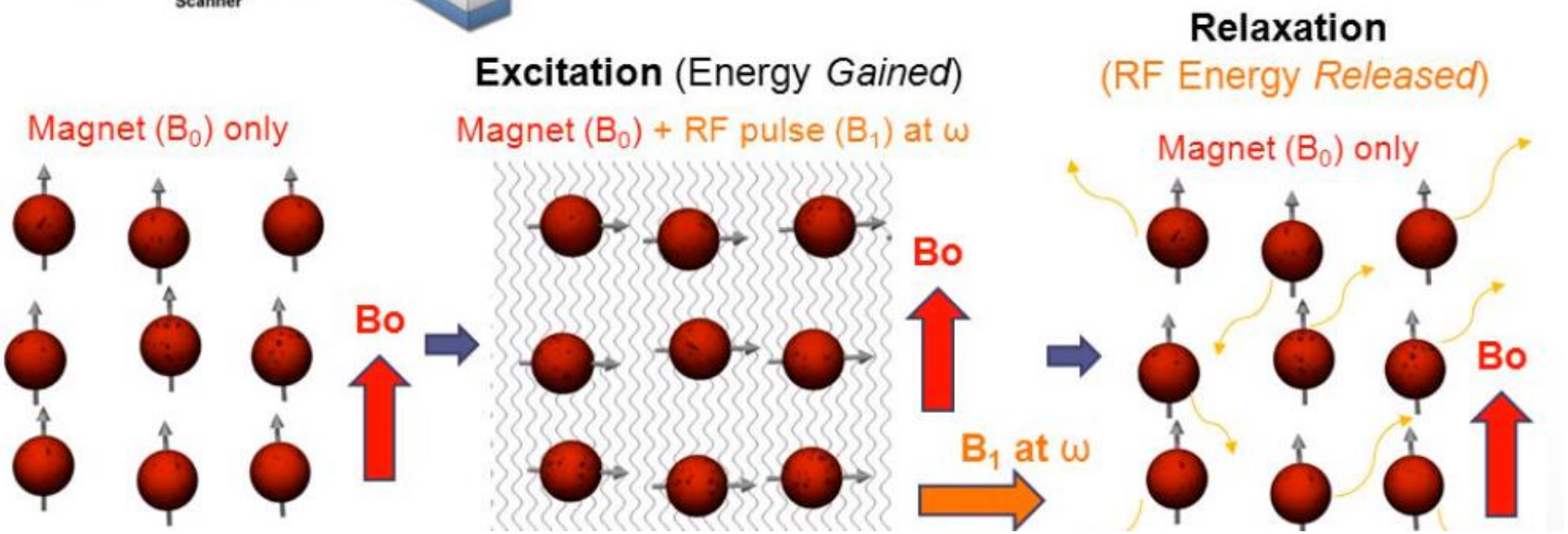


Precession frequency

$$\omega = \gamma B$$

Gyromagnetic ratio

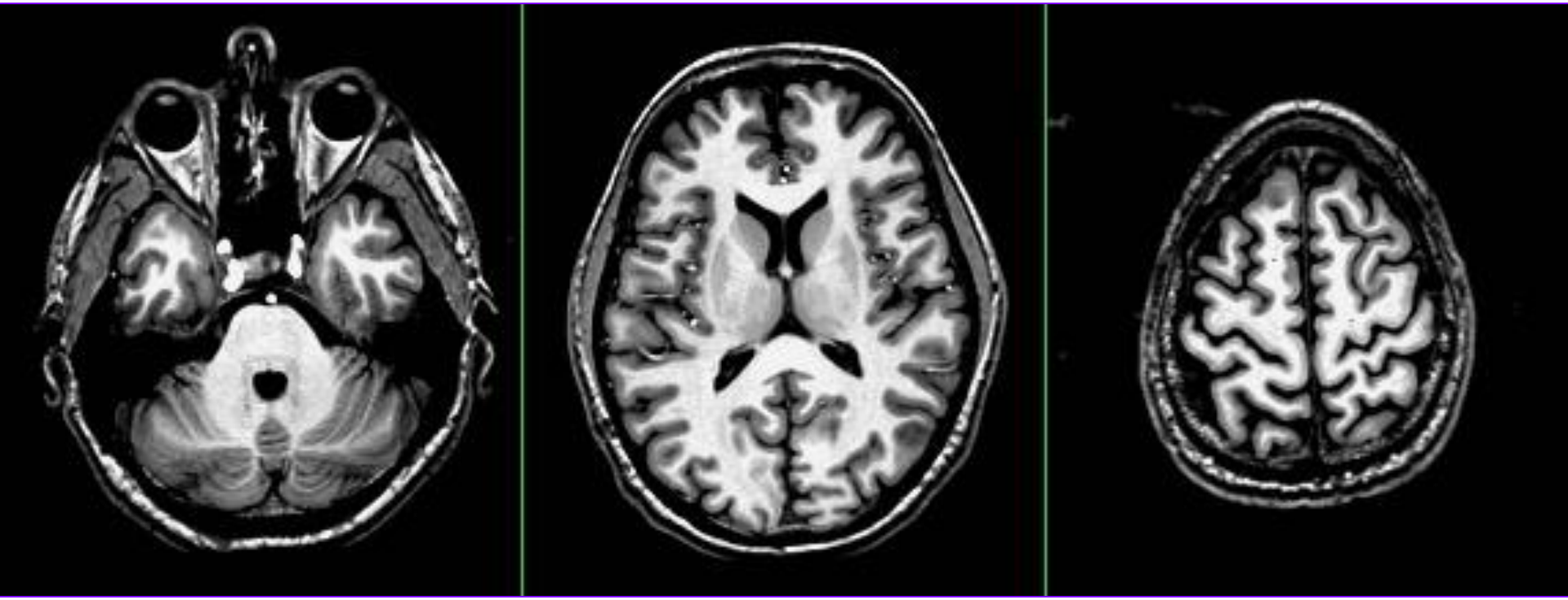
Magnetic field strength



Courtesy of Kevin C. Chan, PhD

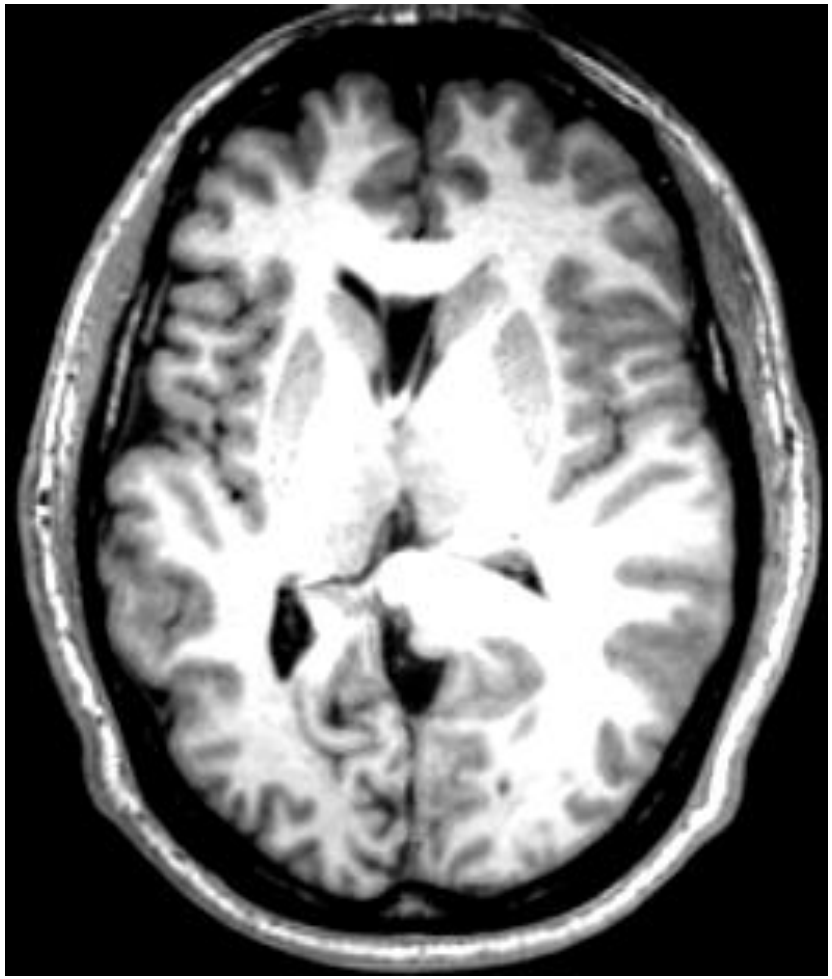
T1-Weighted Images

- Images whose design (timing of radio pulses and data readout) is to produce ***contrast*** between gray matter, white matter, and CSF



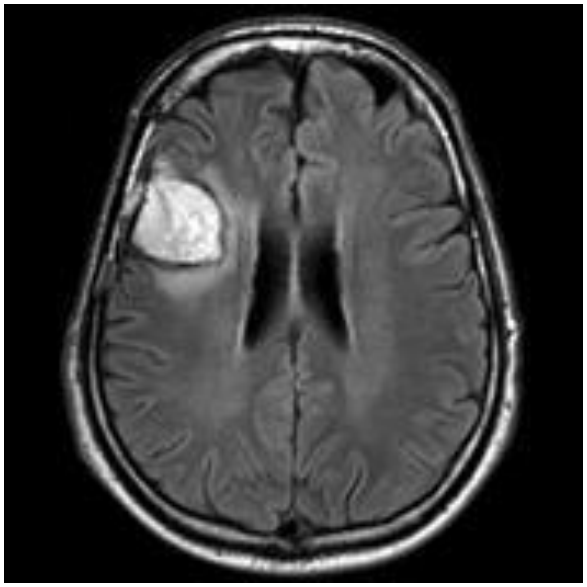
Spatial resolution is about 1 mm^3
Acquisition time for whole head is 5-10 minutes

Structural MRI

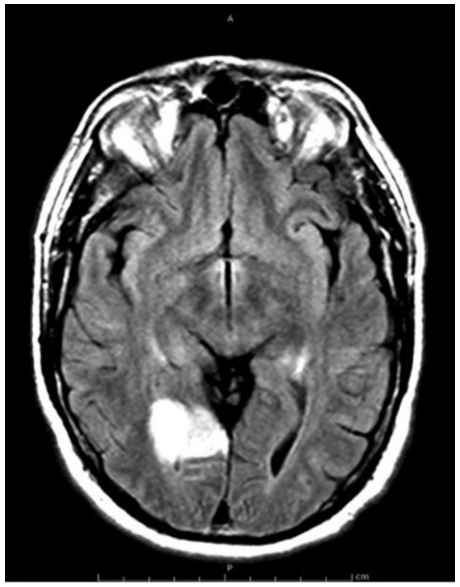


- Essential in clinical care.
- Radiologists perform qualitative “lightbox” reads.
- Most psychiatric and neurological disorders are invisible to reading radiologists.

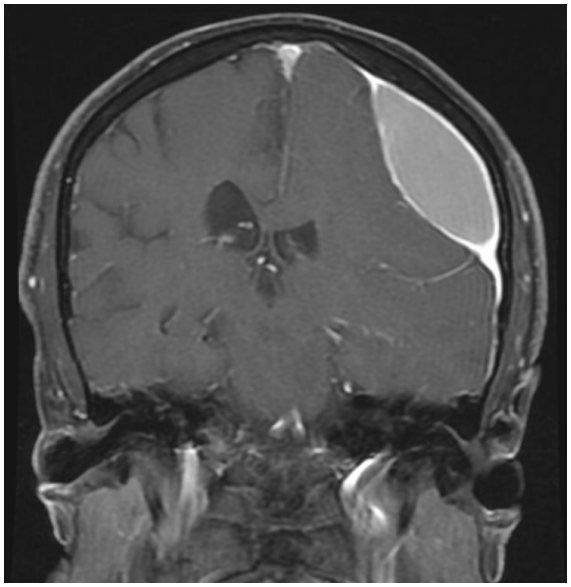
What Diseases Can be Detected with Visual Reading of MRI?



Brain Tumor

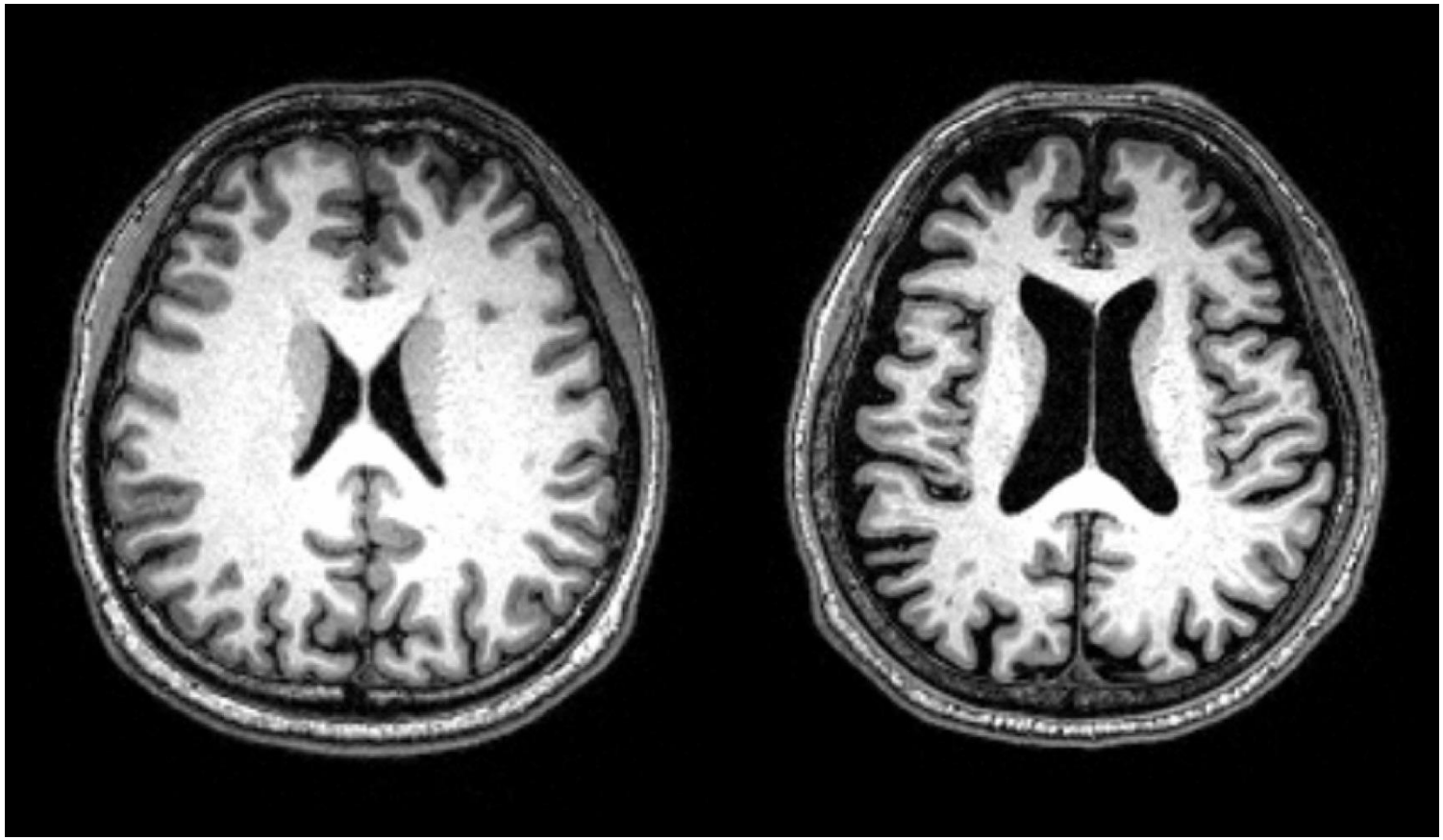


Stroke



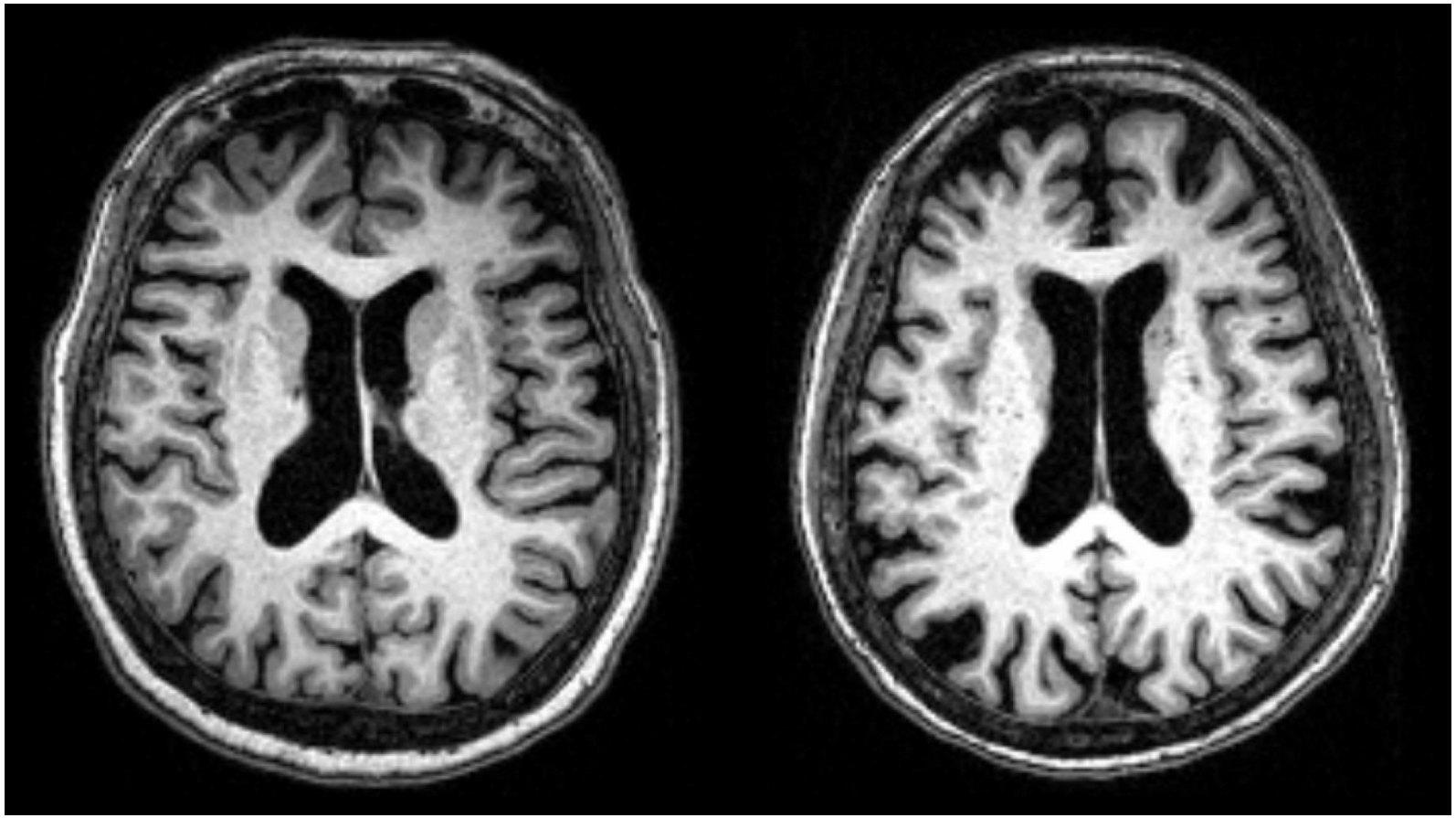
Hemorrhage

Which One is Older Brain?



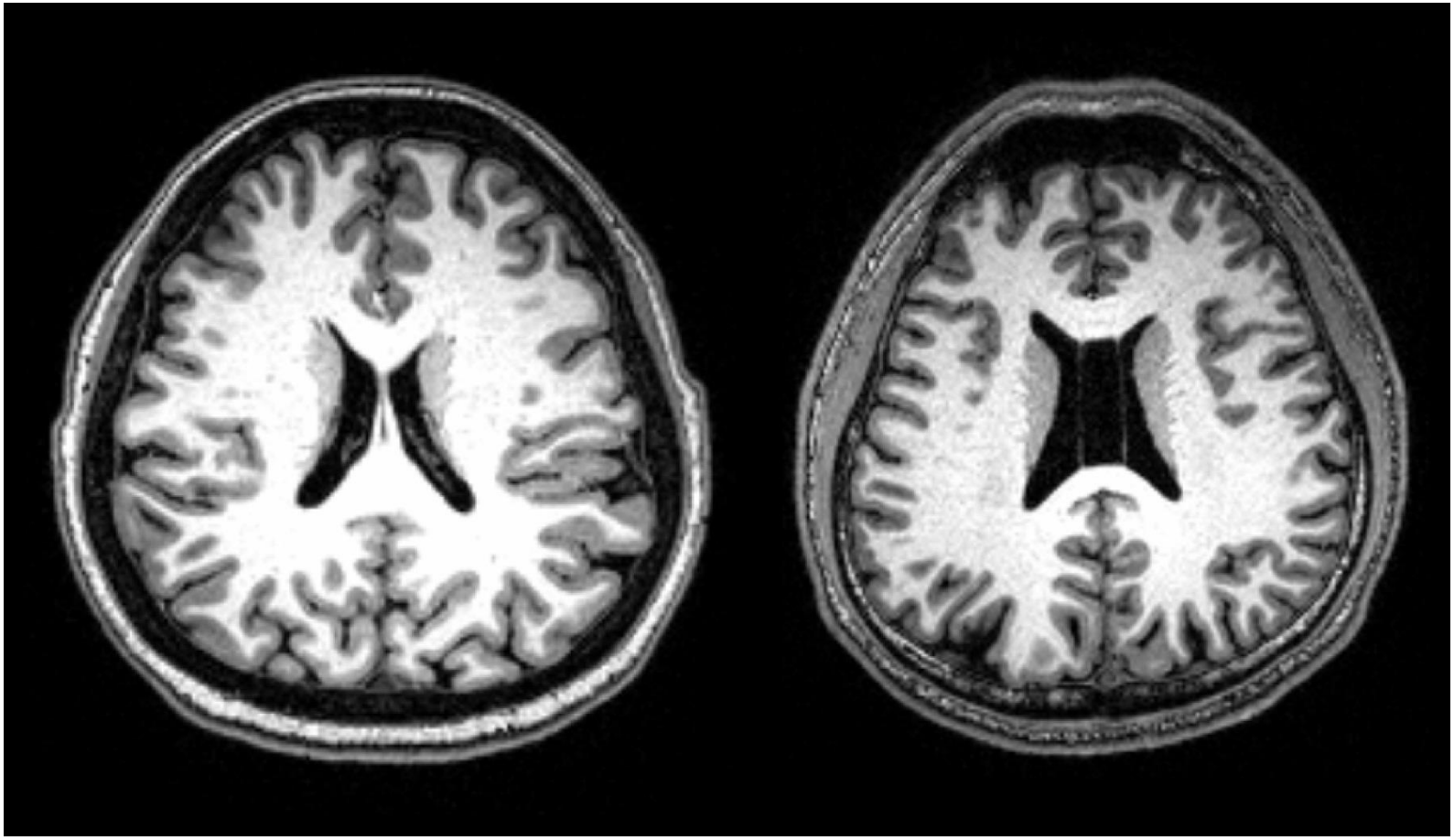
Young (Age 30) **Old (Age 80)**

Which Brain is Demented?



Non-demented (Age 70) **Demented (Age 70)**

Which Brain has Mental Illness?

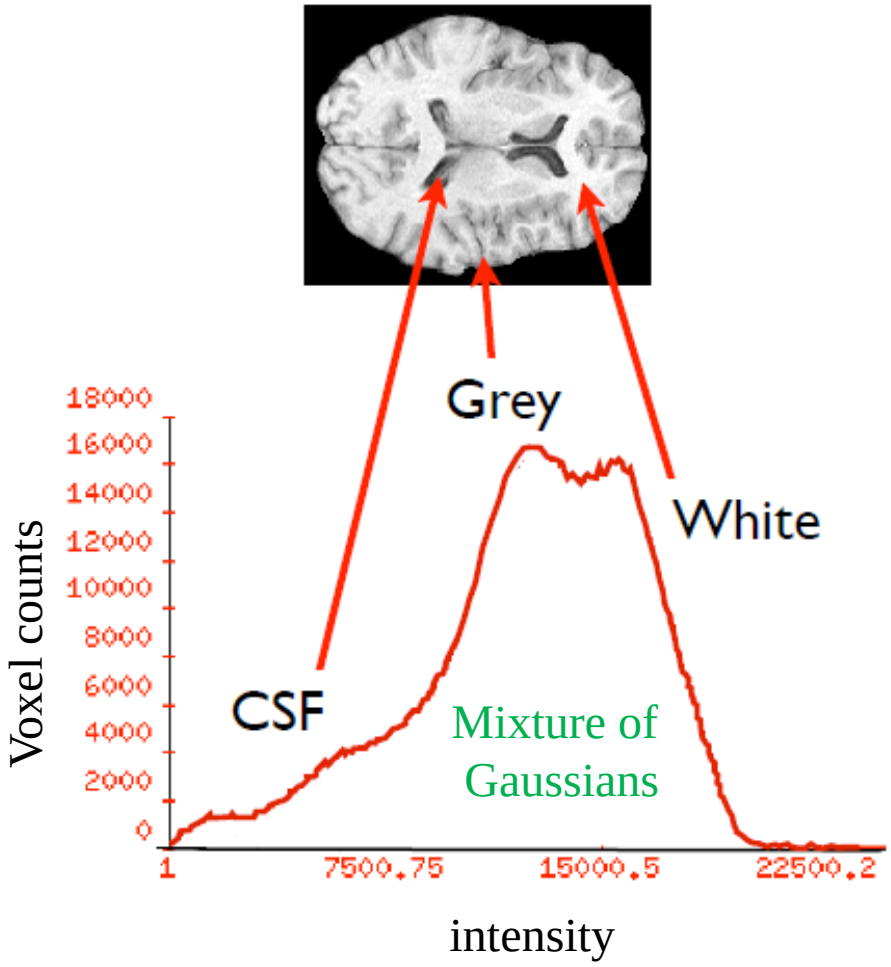


Schizophrenia (Age 30) **Normal (Age 30)**

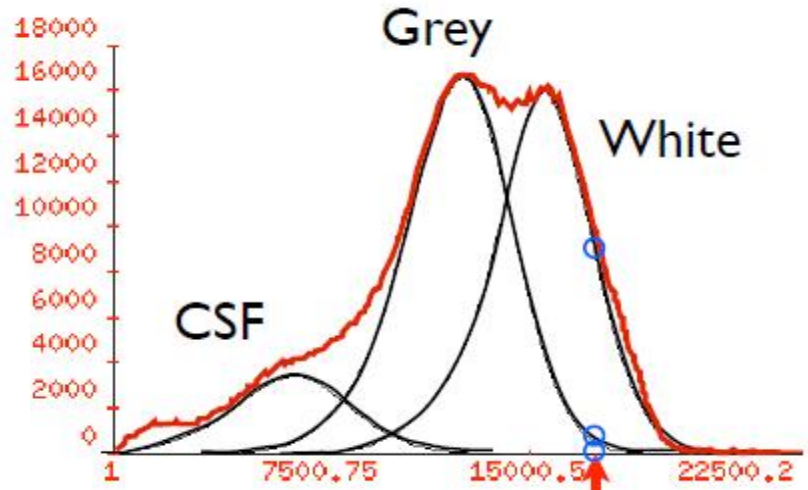
Department of Psychiatry, Taipei Veterans General Hospital

Brain Segmentation

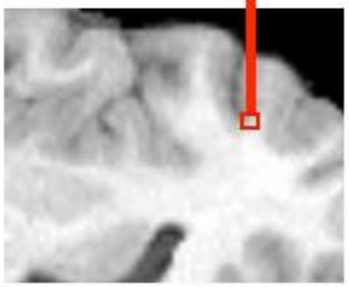
Tissue Intensity Distributions



Probability Distribution



Intensity = 17203



For example:
P (CSF) near zero
P (GM) low
P (WM) moderate

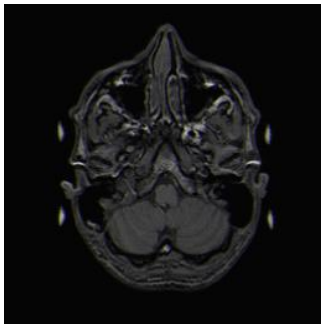
source: FSL lecture

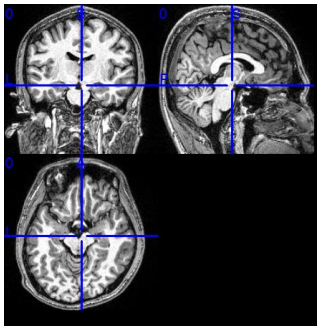
Structural Brain Image

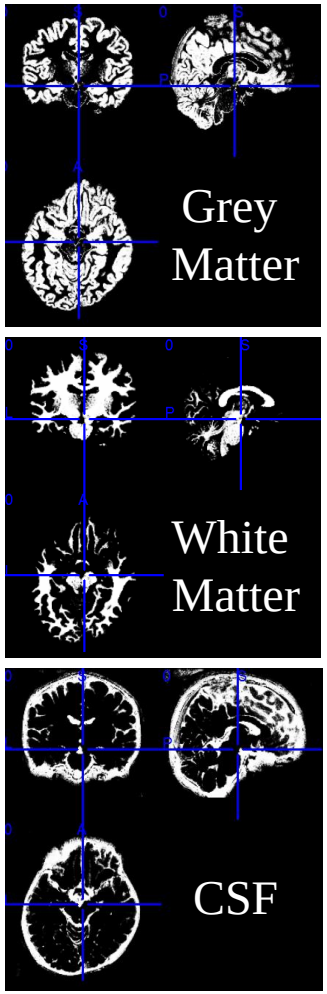
co-registered to MNI template

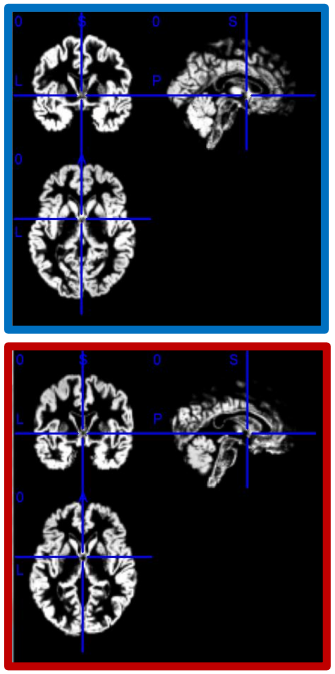
Segmentation

Normalizatoin (preprocessing)

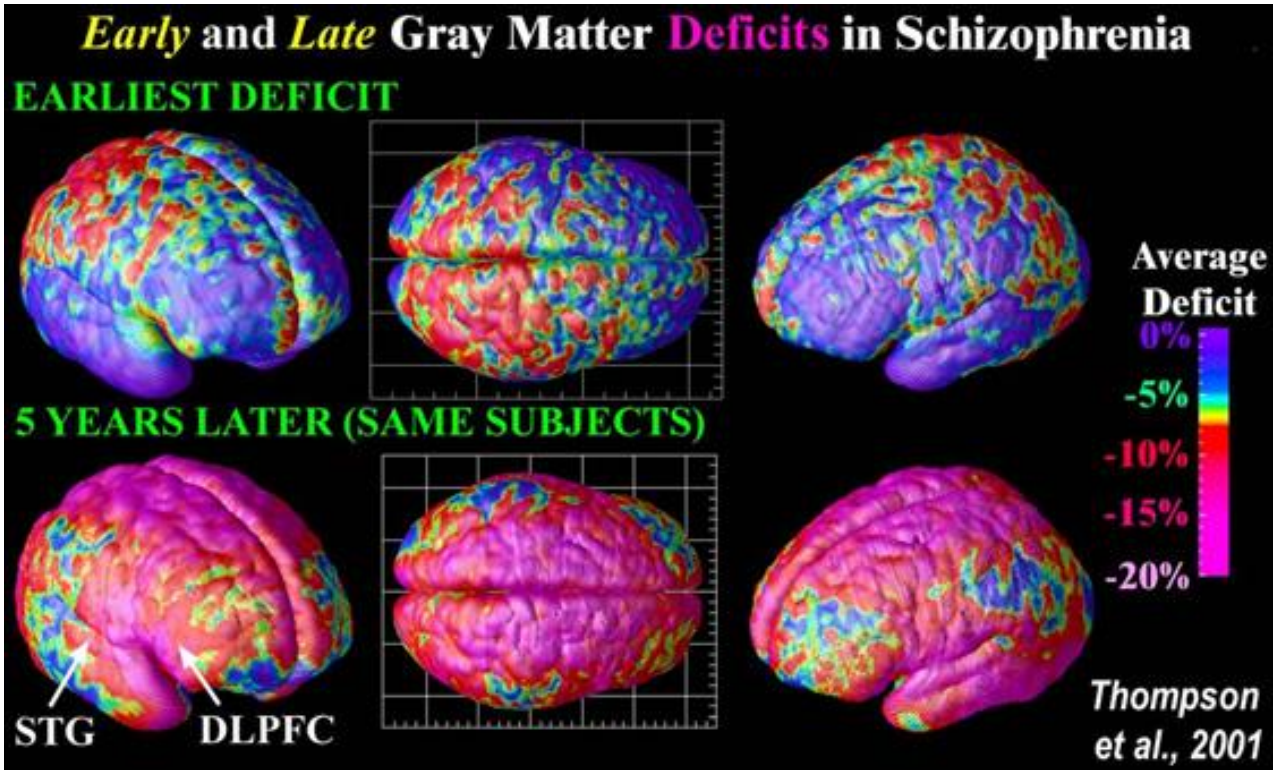








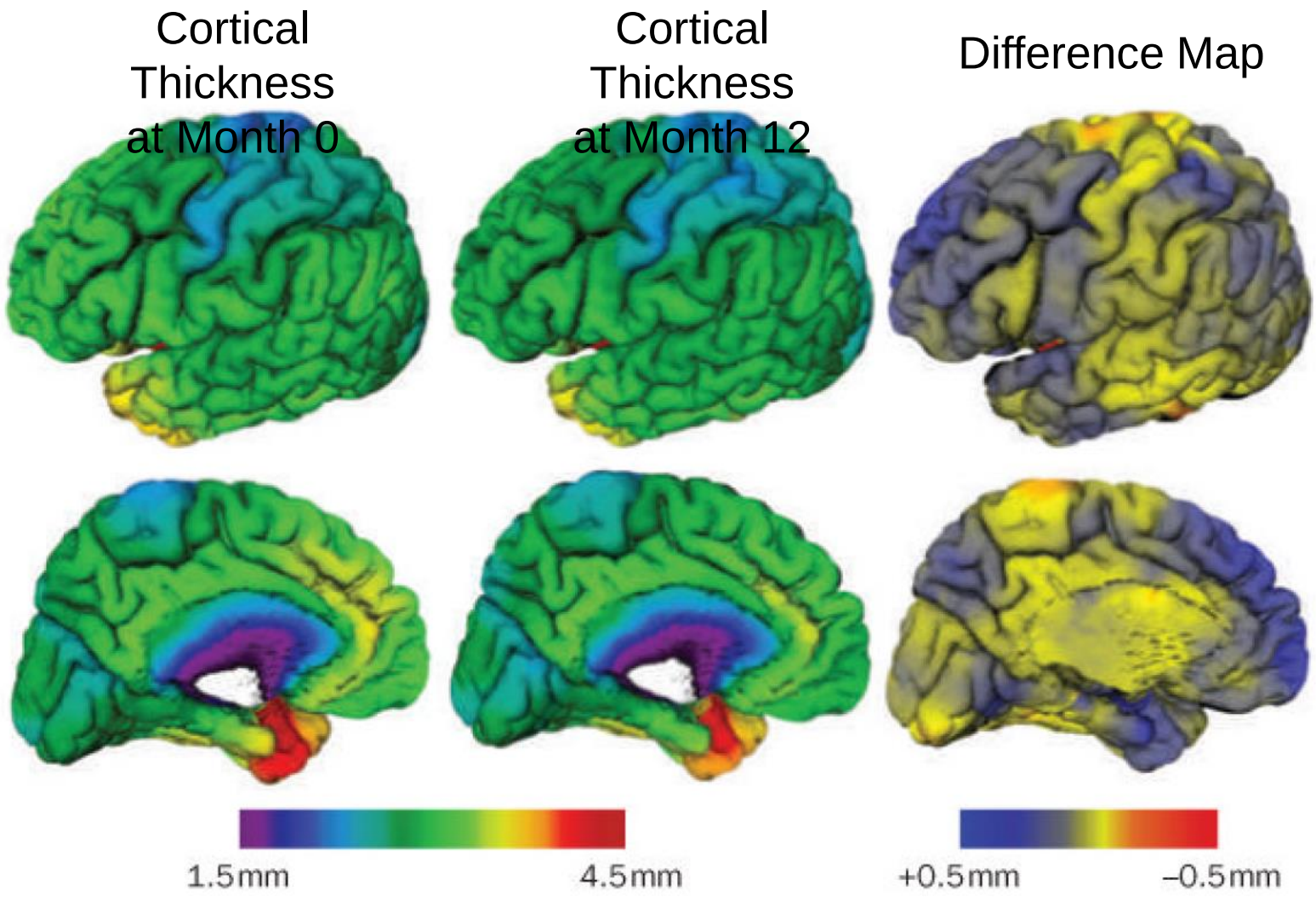
Structural Neuroimaging



Robust changes in brain volume in schizophrenia or bipolar disorder compared with healthy volunteers.

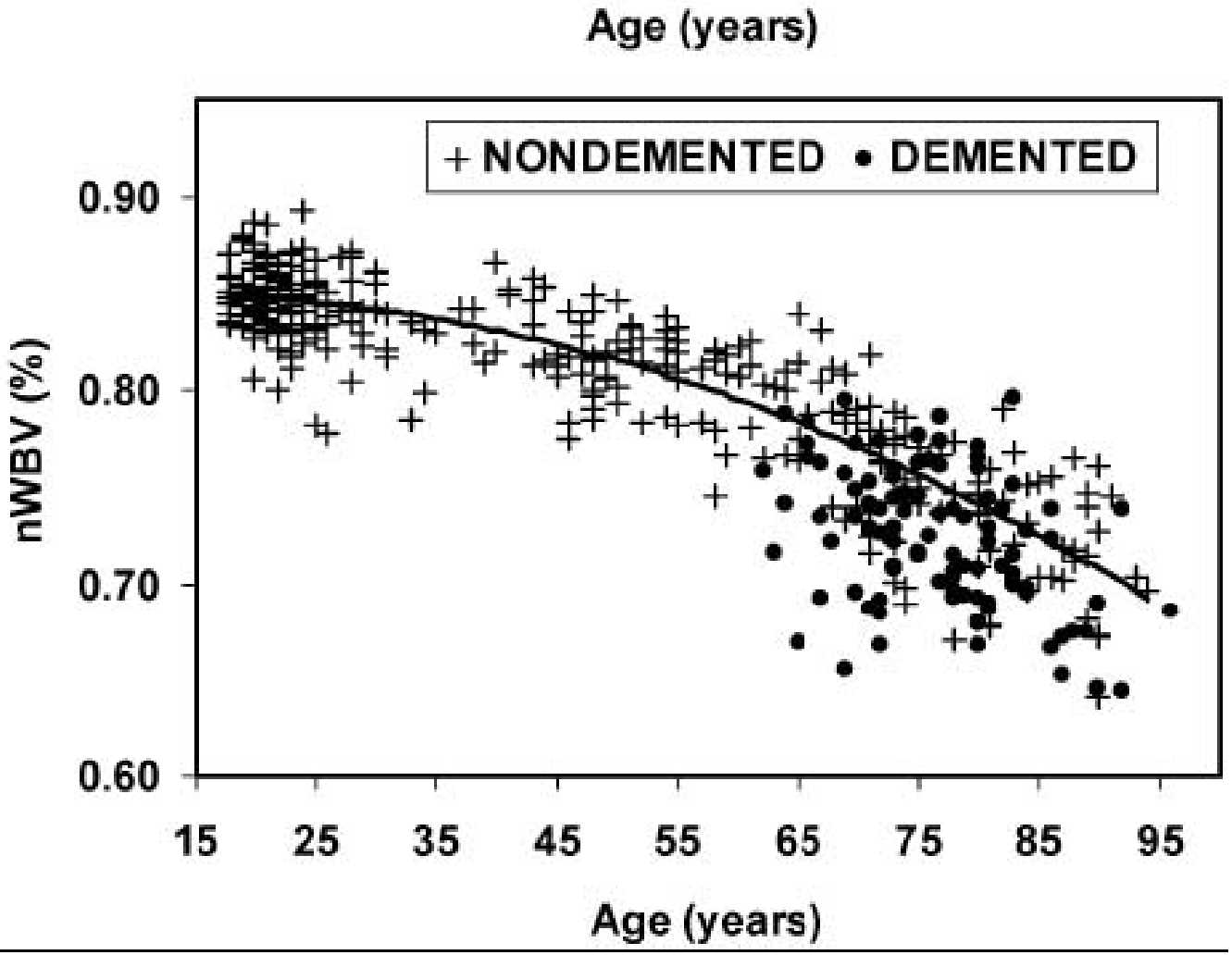
Structural Neuroimaging

Grey Matter Loss



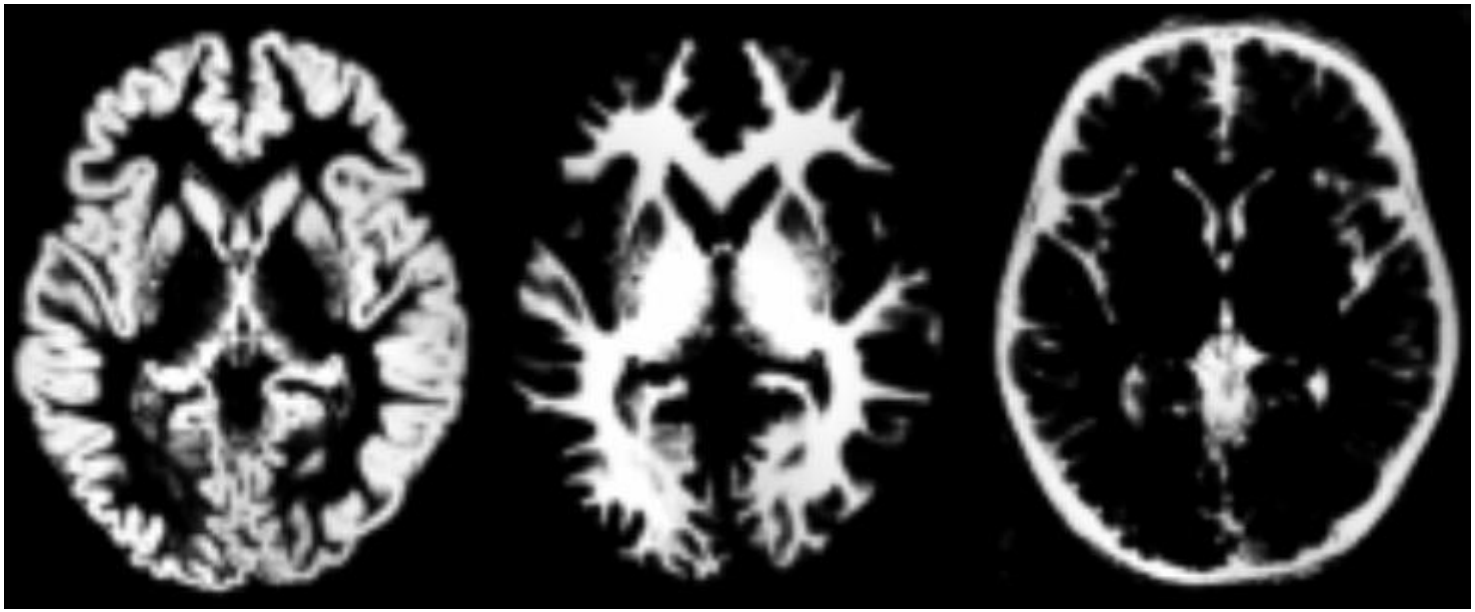
Lerch JP et al. Cereb Cortex 2005;15:995–10

Brain Volume Declines with Age



Marcus et al., 2007

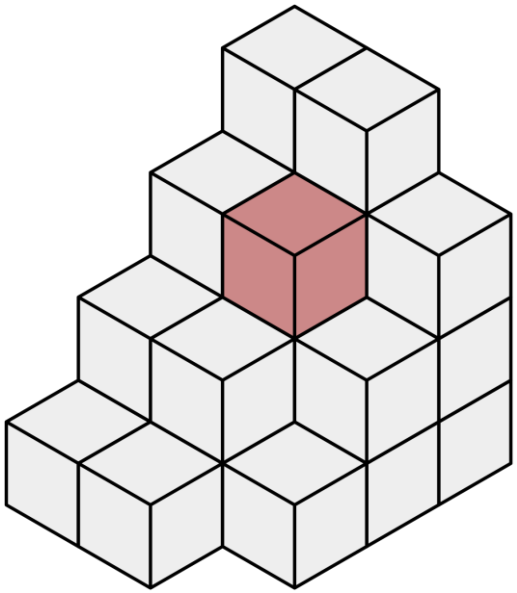
Structural Image Data

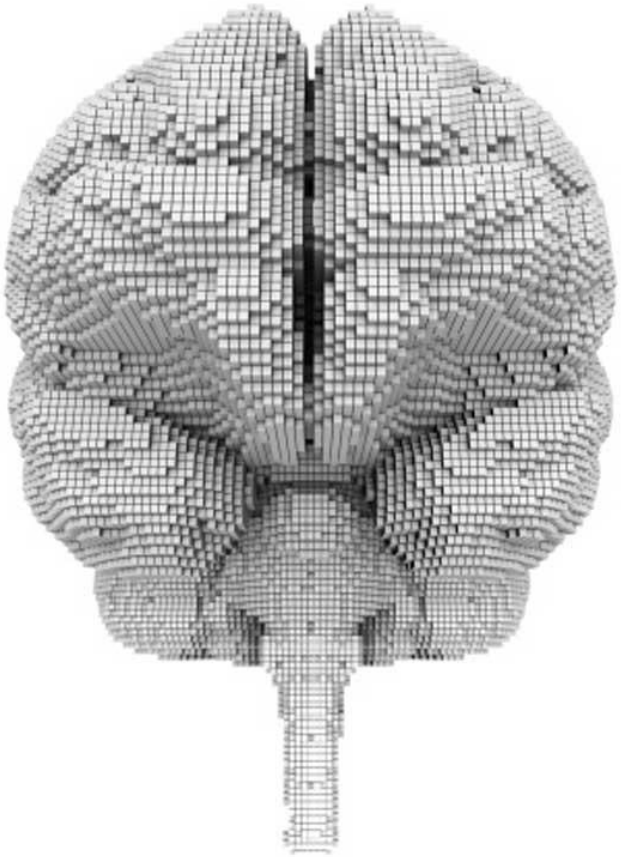


Grey Matter White Matter Cerebrospinal Fluid

902,629 (91 x 109 x 91) voxels

Volumetric Measures





Brain MRI Volume Dataset

Variables - data

data

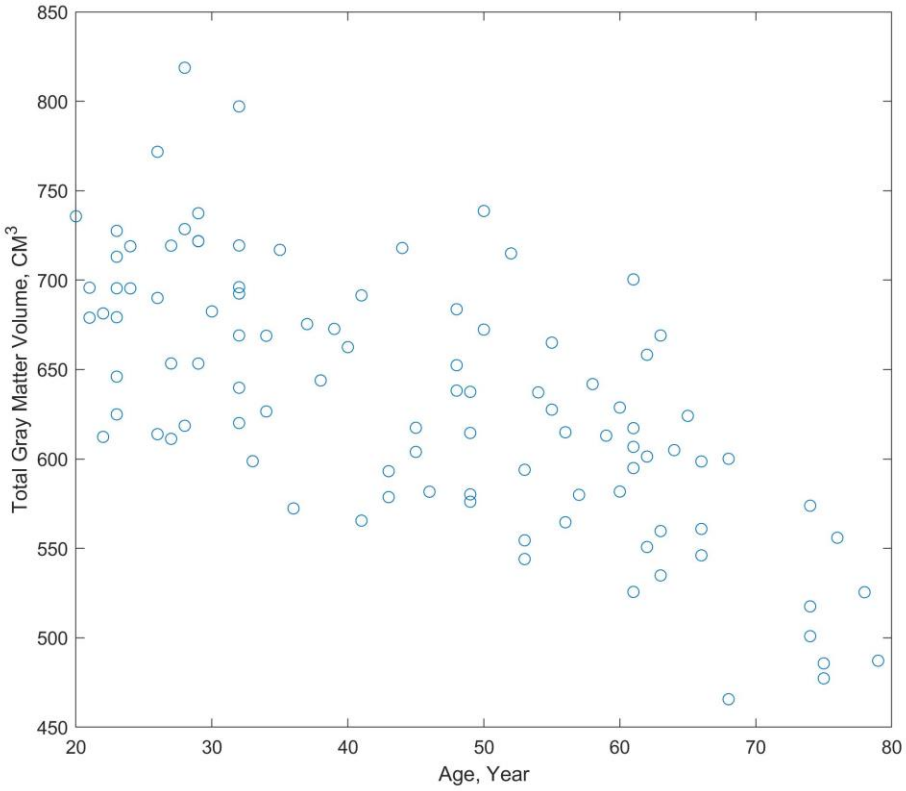
100x13 table

	1 age	2 precentral	3 postcentral	4 precuneus	5 acc	6 pcc	7 hippocampus	8 parahippocampus	9 insula	10 amygdala	11 thalamus	12 caudate	13 total_gray_volume
16	27	17.5899	19.5470	20.6115	9.1557	1.9313	7.5077	9.1782	15.6649	2.2511	4.5496	7.2177	653.3900
17	26	21.4060	25.5420	24.1940	11.5897	2.4352	8.6271	10.0453	15.0735	2.6491	6.1092	7.6653	771.7500
18	44	17.5474	20.5880	22.4920	10.3291	2.3315	8.5556	8.2774	16.0572	2.5229	4.5273	7.0791	717.9500
19	29	20.9110	21.7220	21.2680	10.3129	2.2671	8.3538	9.9375	15.3262	2.3059	5.0299	7.9006	721.7300
20	32	20.2210	23.4550	24.6470	9.6122	2.3891	8.6708	9.8314	16.1941	2.4998	5.4550	8.2280	719.3800
21	43	16.0321	18.2699	19.1615	7.9470	1.6957	7.4373	8.0384	12.6081	2.1004	3.9978	6.6893	593.1700
22	26	15.9008	18.0923	19.6363	9.2360	1.7474	6.9767	7.9961	12.3776	2.0601	4.4963	5.9797	613.8600
23	48	16.8592	18.5835	19.4226	10.1685	1.8224	7.6771	8.8067	15.7832	2.4749	4.2641	6.9144	652.4100
24	74	13.0699	14.9049	14.8380	7.1063	1.9611	6.2609	7.6923	10.3446	1.9756	5.1959	6.2096	500.9100
25	60	16.7262	18.5900	18.9921	8.9771	2.1409	7.3282	8.0500	11.9530	2.1007	5.3707	5.5111	628.7700
26	24	18.7963	19.2896	21.7730	11.3344	1.9282	7.9014	9.1262	16.4156	2.2462	4.7429	9.0899	718.9300
27	68	12.6193	14.0891	14.5775	6.3376	1.3371	6.0240	6.7006	9.7669	1.6293	4.0048	6.0954	465.6700
28	29	20.5290	23.4420	25.2650	10.2504	2.4192	8.2149	9.9839	15.8473	2.4094	5.1558	8.1896	737.3800
29	36	14.4570	17.4352	18.8009	7.4832	1.8637	6.7923	7.4485	12.2394	1.8508	4.2847	6.1735	572.3600
30	41	15.4622	16.3541	18.1379	7.3059	1.7580	7.0318	7.7080	12.3283	2.1908	3.3208	5.7975	565.5100
31	65	15.6366	17.7641	18.8618	8.9596	2.2766	7.9175	9.1094	12.7748	2.4556	6.3284	6.6824	624.0900
32	75	11.9840	12.8942	14.5222	7.8804	1.6511	5.9494	7.0322	9.9513	2.0034	5.1934	5.5290	485.7000
33	66	14.3238	16.2845	16.6410	8.3756	1.4963	6.5679	7.2022	11.8050	1.7956	4.9150	5.9942	546.1100
34	46	16.4590	18.6715	18.4810	7.6979	1.8800	7.7448	7.9234	13.4077	2.0473	4.9227	7.0460	581.6900
35	54	16.4806	18.3527	20.1259	9.5930	1.9759	8.3908	8.7922	15.2114	2.5049	4.7098	6.6024	637.2300
36	61	13.2905	14.0809	16.4829	8.4307	2.5791	5.7152	7.9009	8.2069	2.0576	4.4265	2.9059	525.6500
37	29	18.8721	20.2799	23.8680	11.7824	2.5513	8.4349	9.7644	15.7409	2.5386	5.2752	9.1470	721.8400
38	45	17.1595	20.2338	18.1130	9.4402	1.3560	7.3093	7.6598	12.5889	1.9722	4.3210	6.8141	617.4400
39	32	17.0372	20.2047	20.6155	10.2949	2.2564	8.4058	9.4740	15.2907	2.5611	4.5331	7.1849	669.0600
40	78	13.2881	14.7619	15.6737	7.9238	2.4616	5.2977	7.7142	8.6934	1.9108	4.6574	5.2943	525.4900
41	56	16.1493	18.1701	18.5214	9.2090	1.8230	7.7098	8.3926	14.3406	2.4055	4.5938	7.2445	614.9300

Taiwan Aging and Mental Illness Cohort

Visualize the Relationship Between Age and Total Gray Matter Volume

- load brain_age_prediction_data
- plot(data.age,data.total_gray_volume,'o');
- xlabel('Age, Year');
- ylabel('Total Gray Matter Volume, cm^3');



Generalized Machine Learning Workflow

- Divide data into training and testing subset
- Model training data
- Evaluate trained model in training data
- Use trained model to predict response in testing data
- Evaluate model performance in testing data

Apply Generalized Machin Learning Workflow to Brain Aging MRI Data

- **Divide data into training and testing subset**

```
test_index = zeros(length(data.age),1);
test_sample = randsample(length(data.age),fix(length(data.age)*0.3));
test_index(test_sample) = 1;
train_index = ~test_index;
train_data = data(train_index==1,:);
test_data = data(test_index==1,:);
```

- **Fit multiple linear regression to training data**

```
model = fitlm(train_data,'ResponseVar','age')
```

- **Predict Response in Testing Data**

```
ypred = predict(model,test_data);
```

- **Evaluate the model**

```
RMSE_test = sqrt(mean((ypred-test_data.age).^2))
RMSE_train = model.RMSE
```

```
RMSE_test =

   11.9643
```

```
RMSE_train =

    9.2163
```

Multiple Linear Regression Results

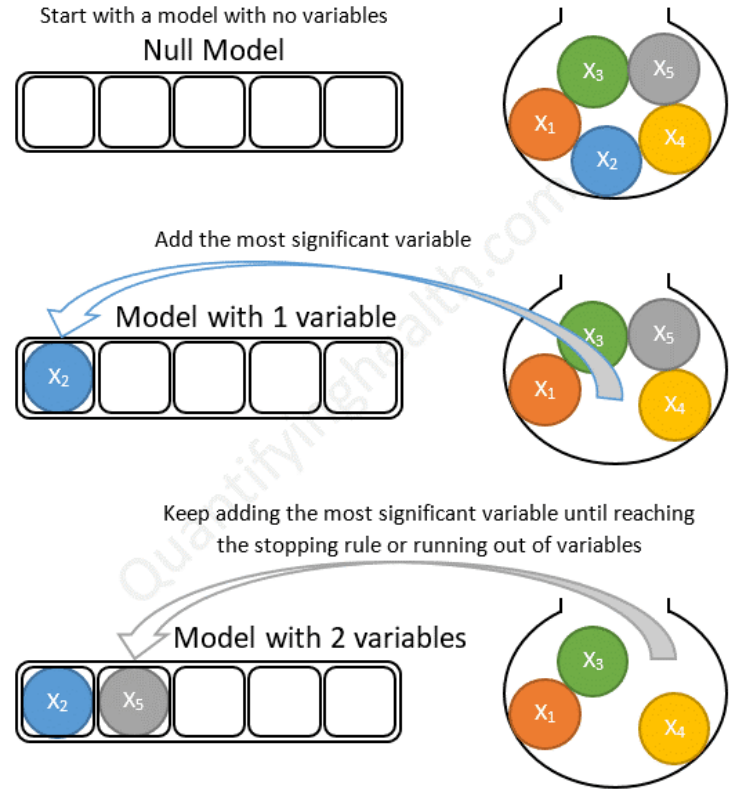
```
>> model = fitlm(train_data,'ResponseVar','age')

model =


Linear regression model:
    age ~ 1 + precentral + postcentral + precuneus + acc + pcc + hippocampus + parahippocampus + insula + amygdala + thalamus + caudate + total_gray_volume

Estimated Coefficients:
```

	Estimate	SE	tStat	pValue
(Intercept)	99.744	13.667	7.298	1.0138e-09
precentral	-0.49539	1.9895	-0.249	0.80426
postcentral	0.74475	1.5365	0.48469	0.62975
precuneus	-0.8219	1.3099	-0.62743	0.53288
acc	3.0926	1.858	1.6645	0.1015
pcc	11.397	6.6321	1.7184	0.091147
hippocampus	7.2695	3.9855	1.824	0.073395
parahippocampus	-12.269	3.15	-3.8949	0.00026045
insula	-2.3455	1.649	-1.4223	0.16038
amygdala	22.485	11.666	1.9274	0.058923
thalamus	9.3171	2.6046	3.5772	0.00071673
caudate	0.53712	2.1934	0.24489	0.80742
total_gray_volume	-0.17748	0.086991	-2.0402	0.045972

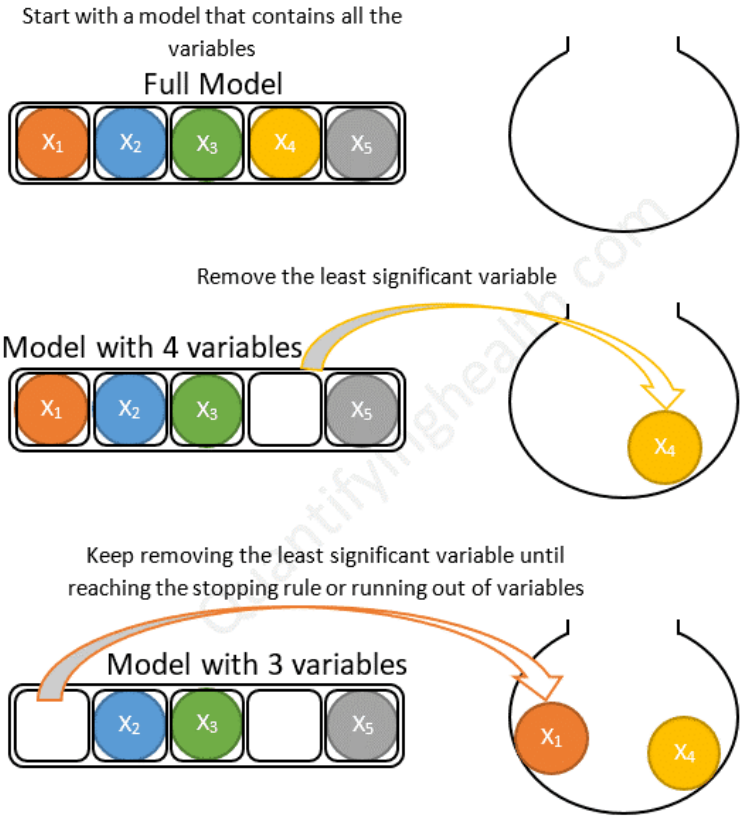
```
Number of observations: 70, Error degrees of freedom: 57
Root Mean Squared Error: 9.22
R-squared: 0.755,  Adjusted R-Squared 0.703
F-statistic vs. constant model: 14.6, p-value = 2.73e-13
```

Variable Selection Methods

Forward stepwise selection example with 5 variables:



Backward stepwise selection example with 5 variables:



https://quantifyinghealth.com/stepwise-selection/

Stepwise Regression

- **Fit stepwise multiple linear regression to training data**

```
model1 = stepwiselm (train_data,'ResponseVar','age')
```

- **Predict Response in Testing Data**

```
ypred = predict(model1,test_data);
```

- **Evaluate the model**

```
RMSE_test = sqrt(mean((ypred-test_data.age).^2))
RMSE_train = model1.RMSE
```

```
RMSE_test =

   12.9883


RMSE_train =

    9.5631
```

Stepwise Regression Results

Command Window

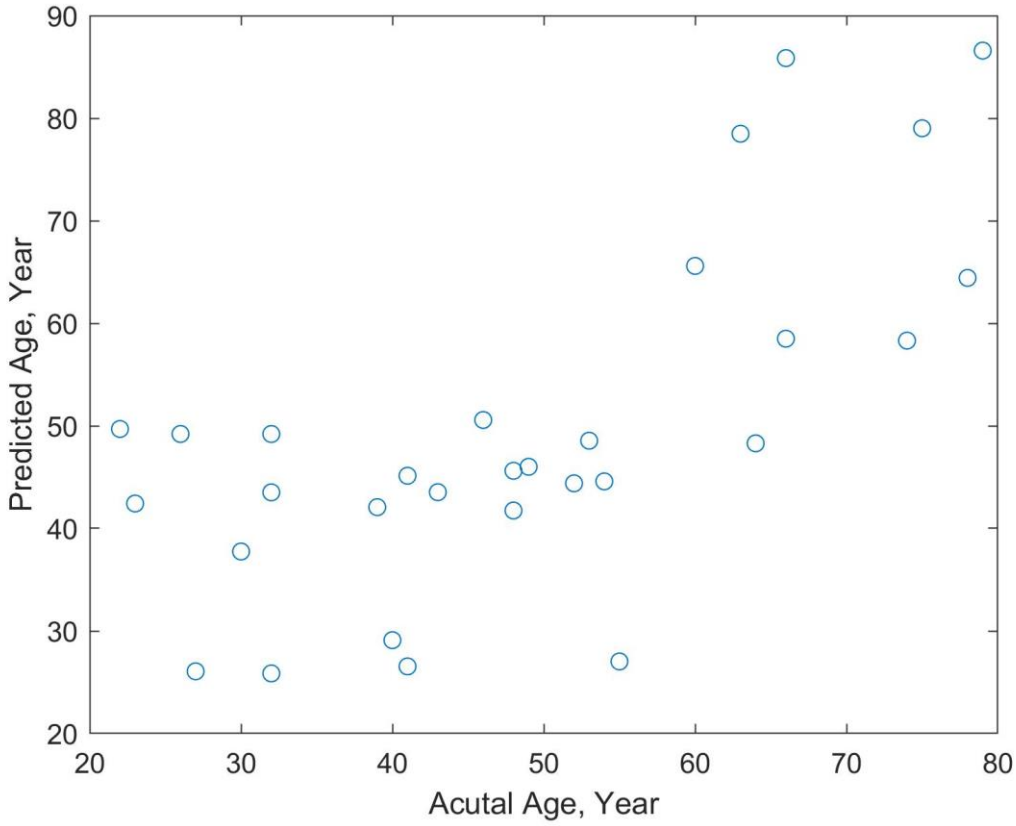
```
>> modell = stepwiselm (train_data,'ResponseVar','age')
1. Adding insula, FStat = 71.4548, pValue = 3.29711e-12
2. Adding thalamus, FStat = 7.8794, pValue = 0.0065421
3. Adding parahippocampus, FStat = 13.5219, pValue = 0.000475453
4. Adding amygdala, FStat = 13.1086, pValue = 0.000576218

modell =


Linear regression model:
    age ~ 1 + parahippocampus + insula + amygdala + thalamus

Estimated Coefficients:
```

	Estimate	SE	tStat	pValue
(Intercept)	92.298	13.07	7.0616	1.3501e-09
parahippocampus	-14.916	2.7831	-5.3596	1.1786e-06
insula	-3.9581	0.8891	-4.4518	3.4201e-05
amygdala	36.098	9.9702	3.6206	0.00057622
thalamus	11.301	2.3823	4.7436	1.1904e-05

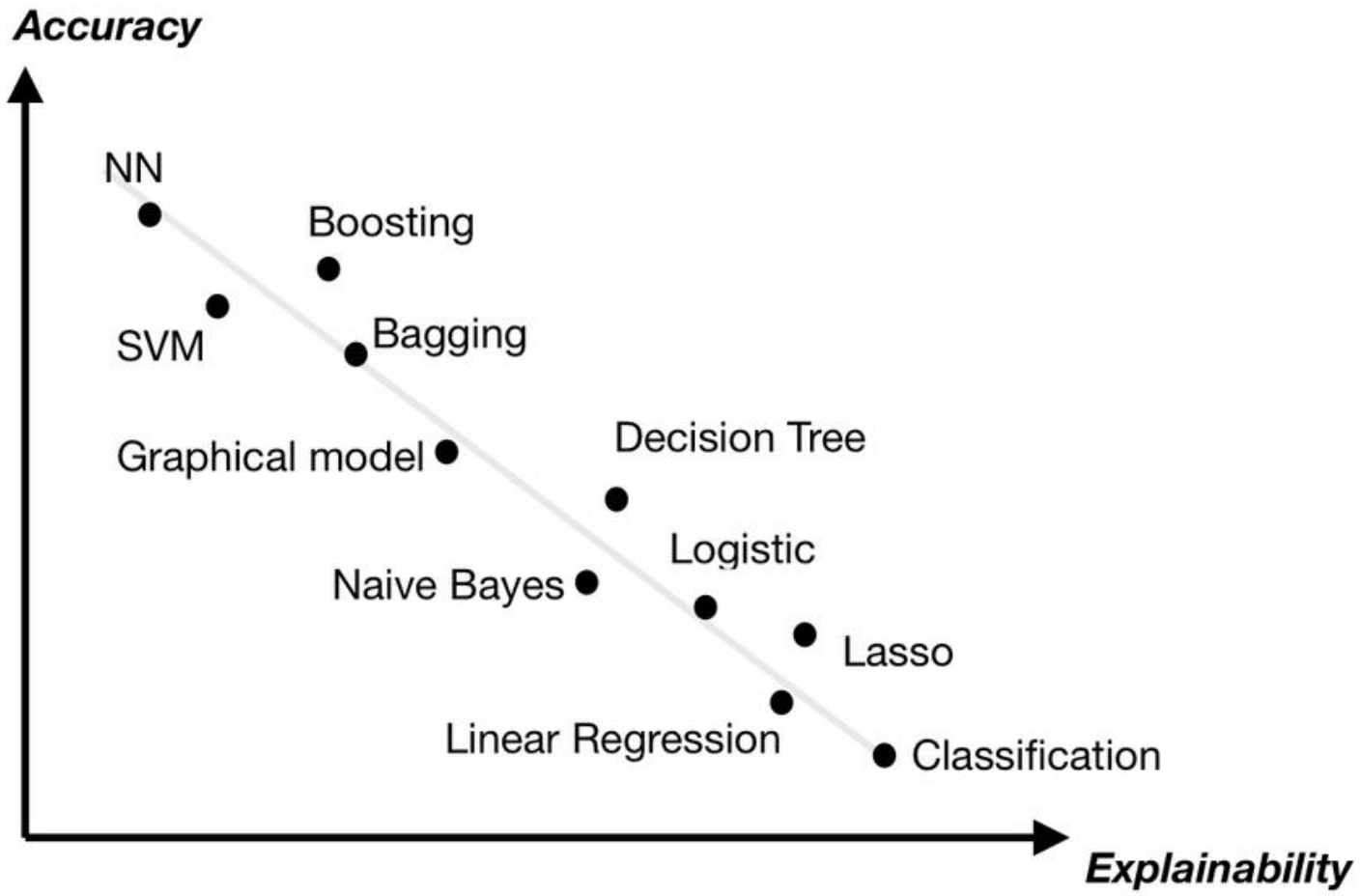
```
Number of observations: 70, Error degrees of freedom: 65
Root Mean Squared Error: 9.56
R-squared: 0.699,  Adjusted R-Squared 0.68
F-statistic vs. constant model: 37.7, p-value = 2.78e-16
```

Visualization of Predicted Brain Age

- plot(test_data.age,ypred,'o');
- xlabel('Acutal Age, Year');
- ylabel('Predicted Age, Year');



Explainability of Machine Learning Methods



https://www.researchgate.net/figure/Accuracy-vs-Explainability-of-the-main-machine-learning-algorithms_fig1_332209054