MODELING & STATISTICS FROM A NON-STATISTICIAN

ROIS, VOXEL/VERTEXWISE, MATRIX ANALYSES, MULTIPLE COMPARISONS, DEATH

LAUREN HOPKINS

IOWA NEUROIMAGING CONSORTIUM

03/27/2022

MODELING STARTS DURING EXPERIMENTAL DESIGN

You decided to do an experiment because you were interested in something

You know your hypotheses

You know the stimuli you put in your experiment

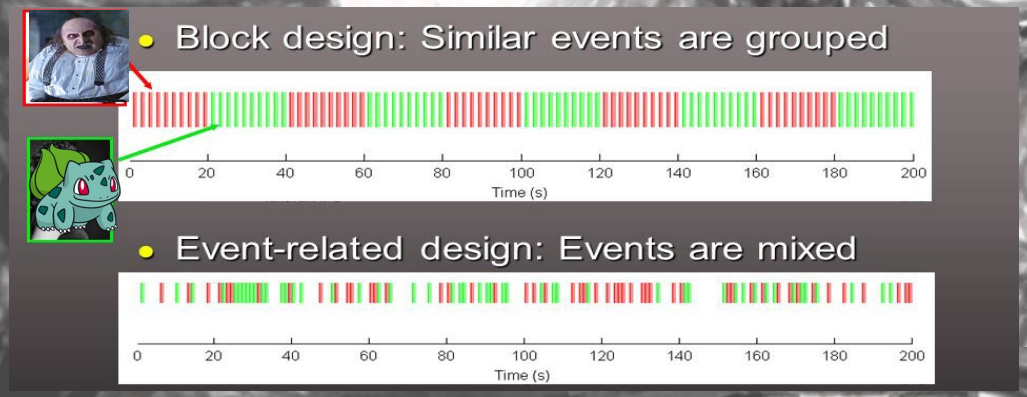
MODELING STARTS WITH YOUR STIMULI AND YOUR HYPOTHESES ABOUT THEM

You decided to do an experiment because you were interested in something

You know your hypotheses

You know the stimuli you put in your experiment

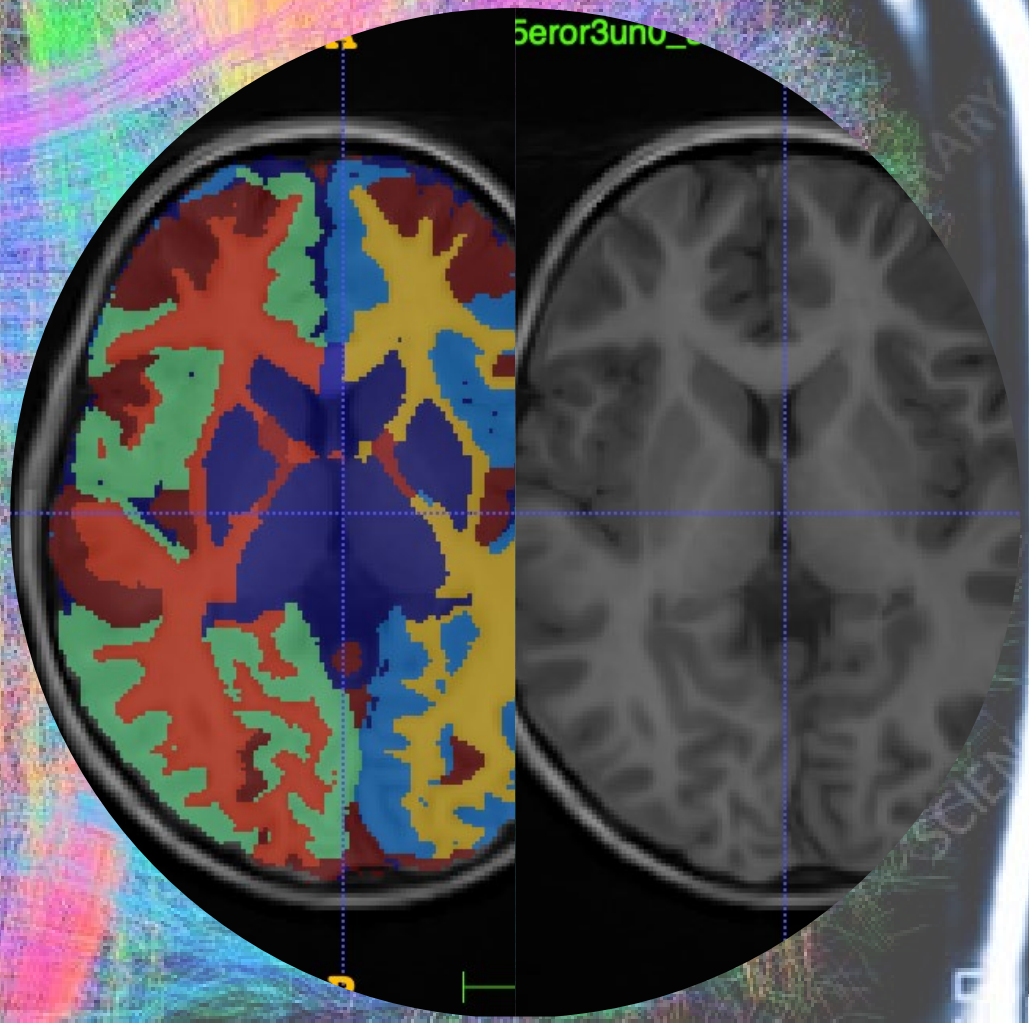
MODELING STARTS WITH WHAT HOW?



- Stimuli
  - Penguin & Bulbasaur
- (Potential) Hypotheses
  - Penguin (villain) > Bulbasaur (inherently neutral) in **amygdala**
  - Penguin (person) > Bulbasaur (animal?) in **FFA**
  - Penguin & Bulbasaur show no differential activation in **V1**

JUST REMEMBER TO ASK YOURSELF: WHY DID I GET THESE SCANS?

- Don't need explicit stimulus presentations to start modeling
- T1w: Volume differences between patients and controls
- DWI: Stronger WM integrity of cingulum in people who score higher on STAI



IF YOU HAVE THIS SCAN:

- T1w
- FLAIR
- DWI
- fMRI

YOUR HYPOTHESIS WILL PROBABLY INVESTIGATE:

- Volume (partial or whole)
- WM hyperintensities
- WM tract integrity (whole or partial)
- Regional brain activity

YOUR EXPERIMENTAL DESIGN AFFECTS THE SCANS YOU GET AND YOUR SCANS AFFECT THE MODELS YOU CREATE

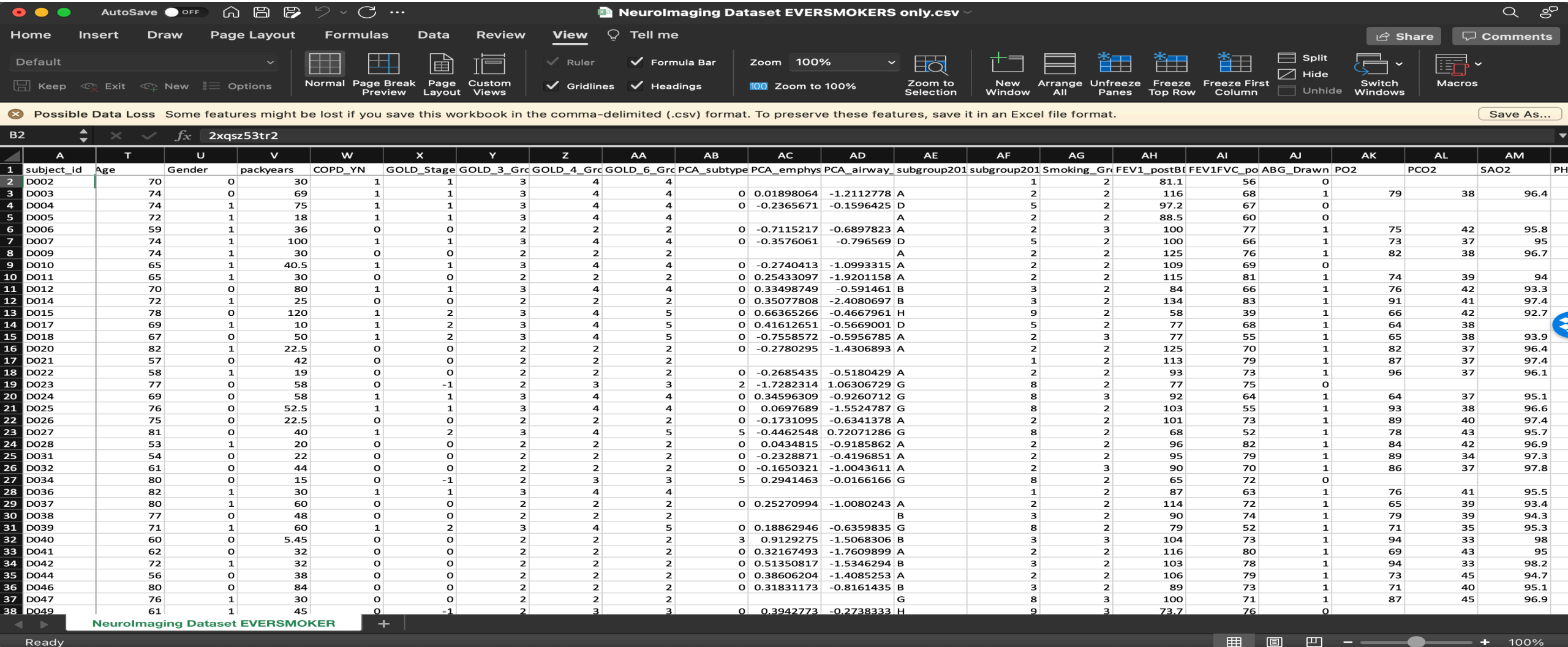
COOL, SO ALL I NEED ARE MY VARIABLES OF INTEREST?

LOL I wish, dude

Confounding variables

- Variables you think may affect your results in a patterned – but unplanned – way
  1. Age
  2. Years in school
  3. Years smoked
  4. Years taking a certain medication
  5. Score on a depression test
  6. Serum albumin levels (blood protein)

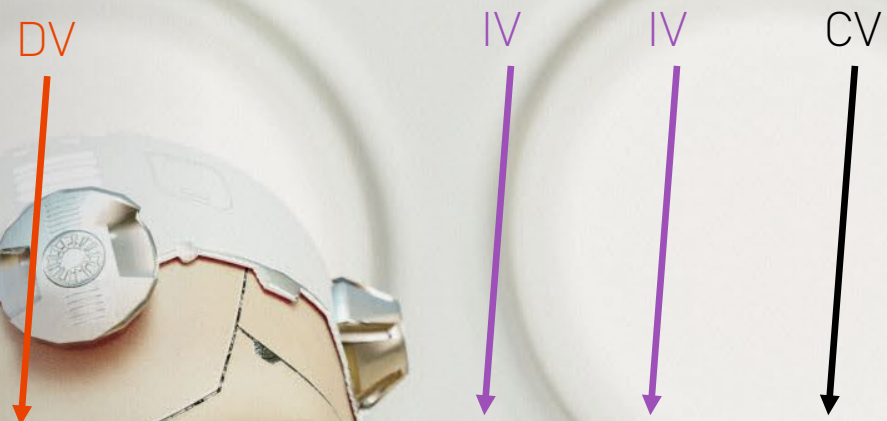
WHERE DOES THAT INFO COME FROM?



	A	T	U	V	W	X	Y	Z	AA	AB	AC	AD	AE	AF	AG	AH	AI	AJ	AK	AL	AM
1	subject_id	Age	Gender	packyears	COPD_YN	GOLD_Stage	GOLD_3_Gro	GOLD_4_Gro	GOLD_6_Gro	PCA_subtype	PCA_emphys	PCA_airway_	subgroup201	subgroup201	Smoking_Gr	FEV1_postBI	FEV1FVC_po	ABG_Drawn	PO2	PCO2	SAO2
2	D002	70	0	30	1	1	3	4	4					1	2	81.1	56	0			
3	D003	74	0	69	1	1	3	4	4	0	0.01898064	-1.2112778	A	2	2	116	68	1	79	38	96.4
4	D004	74	1	75	1	1	3	4	4	0	-0.2365671	-0.1596425	D	5	2	97.2	67	0			
5	D005	72	1	18	1	1	3	4	4				A	2	2	88.5	60	0			
6	D006	59	1	36	0	0	2	2	2	0	-0.7115217	-0.6897823	A	2	3	100	77	1	75	42	95.8
7	D007	74	1	100	1	1	3	4	4	0	-0.3576061	-0.796569	D	5	2	100	66	1	73	37	95
8	D009	74	1	30	0	0	2	2	2				A	2	2	125	76	1	82	38	96.7
9	D010	65	1	40.5	1	1	3	4	4	0	-0.2740413	-1.0993315	A	2	2	109	69	0			
10	D011	65	1	30	0	0	2	2	2	0	0.25433097	-1.9201158	A	2	2	115	81	1	74	39	94
11	D012	70	0	80	1	1	3	4	4	0	0.33498749	-0.591461	B	3	2	84	66	1	76	42	93.3
12	D014	72	1	25	0	0	2	2	2	0	0.35077808	-2.4080697	B	3	2	134	83	1	91	41	97.4
13	D015	78	0	120	1	2	3	4	5	0	0.66365266	-0.4667961	H	9	2	58	39	1	66	42	92.7
14	D017	69	1	10	1	2	3	4	5	0	0.41612651	-0.5669001	D	5	2	77	68	1	64	38	
15	D018	67	0	50	1	2	3	4	5	0	-0.7558572	-0.5956785	A	2	3	77	55	1	65	38	93.9
16	D020	82	1	22.5	0	0	2	2	2	0	-0.2780295	-1.4306893	A	2	2	125	70	1	82	37	96.4
17	D021	57	0	42	0	0	2	2	2					1	2	113	79	1	87	37	97.4
18	D022	58	1	19	0	0	2	2	2	0	-0.2685435	-0.5180429	A	2	2	93	73	1	96	37	96.1
19	D023	77	0	58	0	-1	2	3	3	2	-1.7282314	1.06306729	G	8	2	77	75	0			
20	D024	69	0	58	1	1	3	4	4	0	0.34596309	-0.9260712	G	8	3	92	64	1	64	37	95.1
21	D025	76	0	52.5	1	1	3	4	4	0	0.0697689	-1.5524787	G	8	2	103	55	1	93	38	96.6
22	D026	75	0	22.5	0	0	2	2	2	0	-0.1731095	-0.6341378	A	2	2	101	73	1	89	40	97.4
23	D027	81	0	40	1	2	3	4	5	5	-0.4462548	0.72071286	G	8	2	68	52	1	78	43	95.7
24	D028	53	1	20	0	0	2	2	2	0	0.0434815	-0.9185862	A	2	2	96	82	1	84	42	96.9
25	D031	54	0	22	0	0	2	2	2	0	-0.2328871	-0.4196851	A	2	2	95	79	1	89	34	97.3
26	D032	61	0	44	0	0	2	2	2	0	-0.1650321	-1.0043611	A	2	3	90	70	1	86	37	97.8
27	D034	80	0	15	0	-1	2	3	3	5	0.2941463	-0.0166166	G	8	2	65	72	0			
28	D036	82	1	30	1	1	3	4	4					1	2	87	63	1	76	41	95.5
29	D037	80	1	60	0	0	2	2	2	0	0.25270994	-1.0080243	A	2	2	114	72	1	65	39	93.4
30	D038	77	0	48	0	0	2	2	2				B	3	2	90	74	1	79	39	94.3
31	D039	71	1	60	1	2	3	4	5	0	0.18862946	-0.6359835	G	8	2	79	52	1	71	35	95.3
32	D040	60	0	5.45	0	0	2	2	2	3	0.9129275	-1.5068306	B	3	3	104	73	1	94	33	98
33	D041	62	0	32	0	0	2	2	2	0	0.32167493	-1.7609899	A	2	2	116	80	1	69	43	95
34	D042	72	1	32	0	0	2	2	2	0	0.51350817	-1.5346294	B	3	2	103	78	1	94	33	98.2
35	D044	56	0	38	0	0	2	2	2	0	0.38606204	-1.4085253	A	2	2	106	79	1	73	45	94.7
36	D046	80	0	84	0	0	2	2	2	0	0.31831173	-0.8161435	B	3	2	89	73	1	71	40	95.1
37	D047	76	1	30	0	0	2	2	2				G	8	3	100	71	1	87	45	96.9
38	D049	61	1	45	0	-1	2	3	3	0	0.3942773	-0.2738333	H	9	3	73.7	76	0			

LET'S SEE SOME MODEL EXAMPLES

(IN R CONVENTION BECAUSE I AM ULTRA TRASH AT PYTHON AND SPSS IS GAUCHE)



HIPP.V ~ GROUP

HIPP.V ~ GROUP*DRUG

HIPP.V ~ GROUP*DRUG + AGE

GLOBAL.FA ~ GROUP*AGE

NFL~ CRBLM.GM*TYPE + AGE

VOXEL ~ FEV1 + AGE + PACKYEARS

KNOW 👏 WHAT 👏 YOU 👏 WANT 👏 TO 👏 MODEL 👏

(OR AT LEAST HAVE A GOOD IDEA OF WHAT YOU WANT TO LOOK AT)

AN EXAMPLE OF HOW VARIABLES CAN TRICK YOU

STARRING THE SIMPSONS

"*SIMPSON'S PARADOX* OCCURS WHEN GROUPS OF DATA SHOW ONE PARTICULAR TREND, BUT THIS TREND IS REVERSED WHEN THE GROUPS ARE COMBINED TOGETHER."

Homer and Barney are in a competition to see who can identify the most different Duff Beers from Moe's. Barney identifies a higher proportion of beers correctly than Homer on each of 2 days. Did Barney answer a higher proportion correctly than Homer?

Simpsons Paradox

MONDAY: Homer = 7/8 beers. Barney = 2/2 beers.

Barney (100%) more accurate than Homer (87.5%)

TUESDAY: Homer = 1/2 beers. Barney = 5/8 beers.

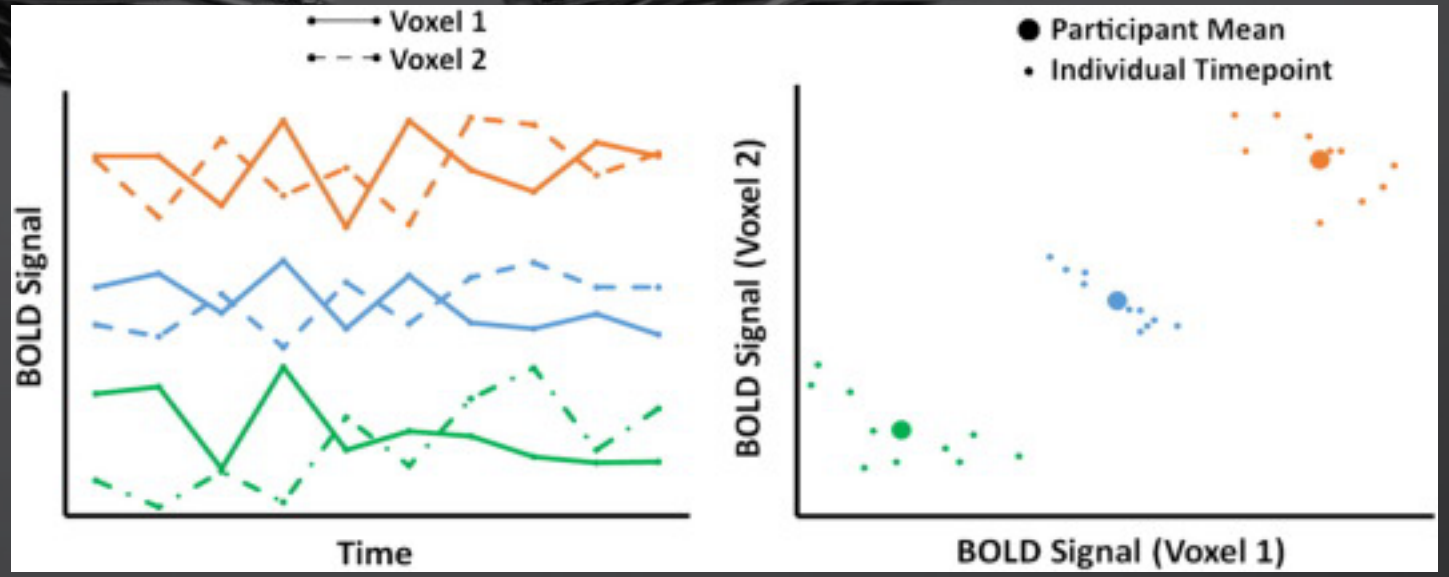
Barney (62.5%) more accurate than Homer (50%)

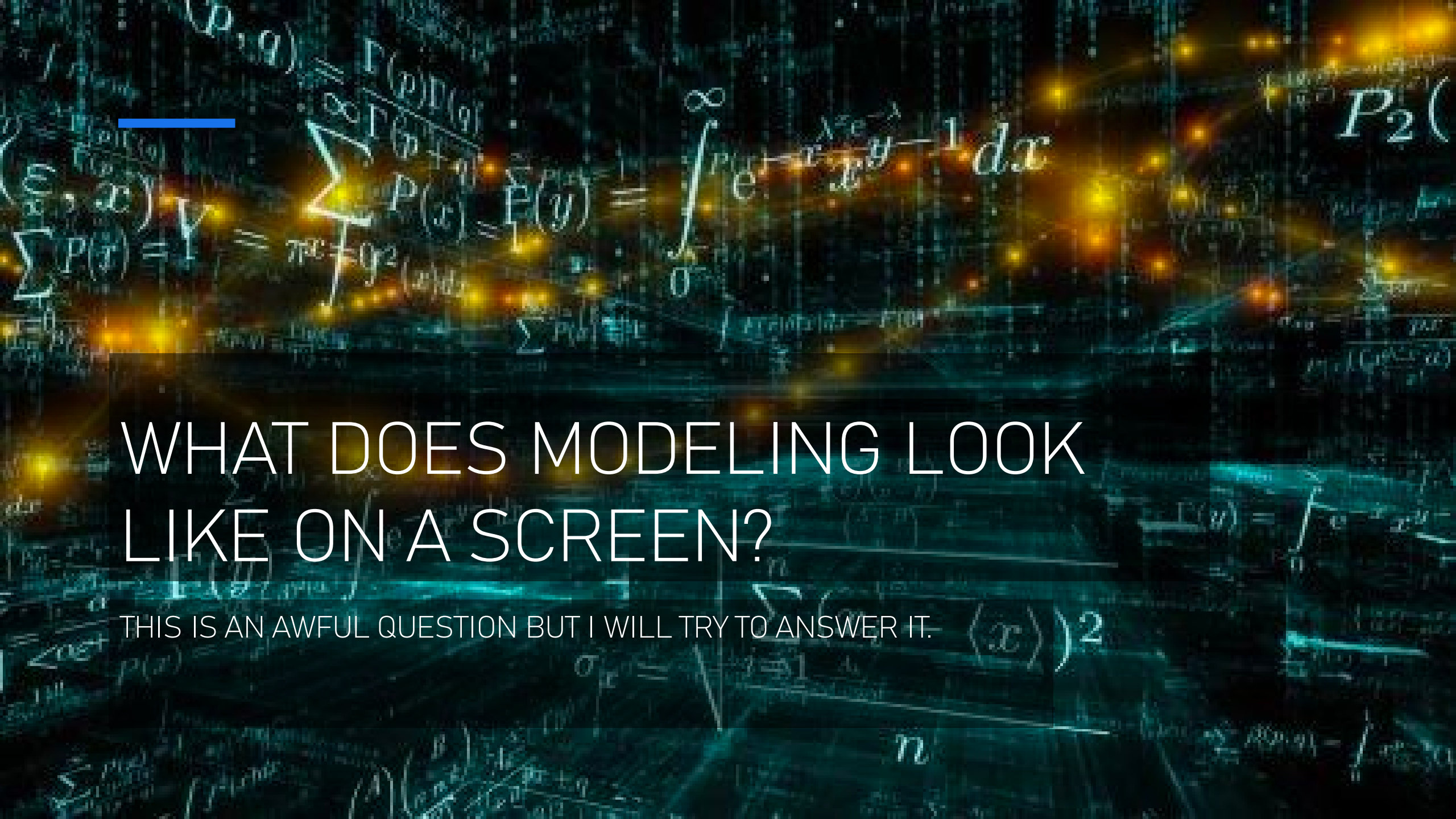
OVERALL: Homer = (80%) MORE ACCURATE than Barney (70%)

Simpsons Paradox

Simpsons Paradox & fMRI

- FMRI DATA CAN BE LOOKED AT IN MANY, MANY, MANY WAYS
- SIMPSONS PARADOX IS COMMON AND CAN HAPPEN IF DATA IS SUBSETTED IN CERTAIN WAYS
- KNOW YOUR DATA AND BE WARY!





WHAT DOES MODELING LOOK LIKE ON A SCREEN?

THIS IS AN AWFUL QUESTION BUT I WILL TRY TO ANSWER IT.

THE DATA STRUCTURE (E.G. IN R)

- ROWS = SINGLE OBSERVATION
- COLS = SINGLE MEASURE/VARIABLE
- THIS MEANS SESSIONS/TIMEPOINTS FOR THE SAME SUBJECT GO IN DIFFERENT ROWS

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U
ursi	aid	session_id	visit	family_id	date_of_visit	group_binary	group_predn	age	elapsed_byD	elapsed_byA	age_baseline	age_onset	disease_dura	epal	lnpal	gender	education	mirs	pegs.dom	tap.dom
119	1354	60153518	1	11350	1/10/18	DM1	DM1	61.417	0	0	61.417	NA	NA	NA	NA	2	16	2	81.5	34.57
119	1738	60771519	2	11350	2/22/19	DM1	DM1	62.5	1.11780822	1.083	61.417	NA	NA	NA	NA	2	16	1	83.66	26.57
119	2248	60555820	3	11350	2/7/20	DM1	DM1	63.417	2.07643536	2	61.417	63	0.49343536	NA	NA	2	16	2	81	35.
122	2497	63335916	1	99978	8/18/16	Unaffected	Unaffected	58.917	0	0	58.917	NA	NA	13	2.56494936	2	18	NA	86	45.
122	1123	60366018	2	99978	1/25/18	Unaffected	Unaffected	60.333	1.43733812	1.416	58.917	NA	NA	13	2.56494936	2	18	NA	76	38.
131	2323	64094516	1	11761	10/10/16	DM1	DM1	54.583	0	0	54.583	46	8.583	100	4.60517019	2	20	2	73	3
137	1744	65250814	1	99993	12/30/14	Unaffected	Unaffected	52.333	0	0	52.333	NA	NA	14	2.63905733	1	13	NA	70	58.
137	1786	64426718	3	99993	11/3/18	Unaffected	Unaffected	56	3.84383562	3.667	52.333	NA	NA	14	2.63905733	1	13	NA	66	46.
137	2449	61997416	2	99993	5/10/16	Unaffected	Unaffected	53.75	1.36067071	1.417	52.333	NA	NA	14	2.63905733	1	13	NA	71	4
139	2350	60512220	3	10078	2/4/20	DM1	DM1	43.917	2.22714275	2.25	41.667	25	18.8941428	126	4.83628191	1	13	2	59	3
139	1339	64573117	1	10078	11/13/17	DM1	DM1	41.667	0	0	41.667	25	16.667	126	4.83628191	1	14	2	63	3
139	2200	64616118	2	10078	11/16/18	DM1	DM1	42.667	1.00821918	1	41.667	25	17.6752192	126	4.83628191	1	13	2	58	5
152	1069	60167218	1	99956	1/11/18	Unaffected	Unaffected	46.75	0	0	46.75	NA	NA	13	2.56494936	2	18	NA	58	5
164	1597	61547518	1	99952	4/17/18	Unaffected	Unaffected	33.667	0	0	33.667	NA	NA	NA	NA	2	18	NA	54.47	50.28
164	1279	61347419	2	99952	4/3/19	Unaffected	Unaffected	34.667	0.96164384	1	33.667	NA	NA	NA	NA	2	18	NA	55.19	52.
166	1336	61894618	1	99947	5/11/18	Unaffected	Unaffected	59.25	0	0	59.25	NA	NA	12	2.48490665	2	16	NA	75.66	48.
172	2641	63720917	3	11317	9/15/17	DM1	DM1	40.167	2.96986301	2.917	37.25	16	24.219863	246	5.50533154	1	12	2	60	30.
172	2512	61118116	2	11317	3/17/16	DM1	DM1	38.667	1.4734037	1.417	37.25	16	22.7234037	246	5.50533154	1	12	2	78	31.
172	1021	63879514	1	11317	9/26/14	DM1	DM1	37.25	0	0	37.25	16	21.25	246	5.50533154	1	12	2	66	31.
188	2314	60638916	2	99996	2/13/16	Unaffected	Unaffected	49.917	1.17776031	1.167	48.75	NA	NA	12	2.48490665	2	14	NA	59.22	52.
188	1516	64961114	1	99996	12/10/14	Unaffected	Unaffected	48.75	0	0	48.75	NA	NA	12	2.48490665	2	14	NA	72	36.
188	2506	63495017	3	99996	8/30/17	Unaffected	Unaffected	51.5	2.72054795	2.75	48.75	NA	NA	12	2.48490665	2	14	NA	58	51.
199	1810	63705915	2	10231	9/14/15	DM1	DM1	50.417	1.02739726	1	49.417	24	26.4443973	131	4.87519732	2	12	4	117	11.
199	2599	63565914	1	10231	9/4/14	DM1	DM1	49.417	0	0	49.417	24	25.417	131	4.87519732	2	12	2	104	1
202	1636	60844717	1	10843	2/27/17	DM1	DM1	55	0	0	55	49	6	112	4.71849887	2	18	2	91	3
202	2524	64312818	2	10843	10/26/18	DM1	DM1	56.667	1.66027397	1.667	55	49	7.66027397	112	4.71849887	2	18	2	79	35.
206	1138	63134519	2	99938	8/5/19	Unaffected	Unaffected	57.083	1.54520548	1.045	56.038	NA	NA	5	1.60943791	2	14	NA	67	47.85
206	1126	60269018	1	9938	1/18/18	Unaffected	Unaffected	56.038	0	0	56.038	NA	NA	5	1.60943791	2	14	NA	70.47	5
209	2398	NA	1	11107	9/16/14	DM1	DM1	53.917	0	0	53.917	25	28.917	145	4.97673374	2	12	3	84	1
209	2755	NA	2	11107	3/10/16	DM1	DM1	55.333	1.48167528	1.416	53.917	25	30.3986753	145	4.97673374	2	12	3	107	22.
214	2593	60309716	2	99990	1/21/16	Unaffected	Unaffected	38.083	1.04094618	1	37.083	NA	NA	12	2.48490665	2	20	NA	56	58.
214	1543	60096015	1	99990	1/6/15	Unaffected	Unaffected	37.083	0	0	37.083	NA	NA	12	2.48490665	2	20	NA	63	53.
215	2335	61693018	2	99971	4/27/18	Unaffected	Unaffected	52.5	1.15068493	1.083	51.417	NA	NA	14	2.63905733	1	18	NA	63	4
215	1303	60902117	1	99971	3/3/17	Unaffected	Unaffected	51.417	0	0	51.417	NA	NA	14	2.63905733	1	18	NA	79	40.
217	1660	62050617	1	10819	5/22/17	Unaffected	Unaffected	57.917	0	0	57.917	NA	NA	5	1.60943791	1	18	1	67	3

df

Edit

100%

- fMRI experiments can include:
  - Multiple scans per subject
  - Multiple sessions per subject
  - Multiple timepoints
  - Scanner software change between scans

WHAT DO YOU NEED FOR MODELING

- Preprocessed data (any modality)
- Variables datasheet
- Model to run

fMRI ~ group*CAG

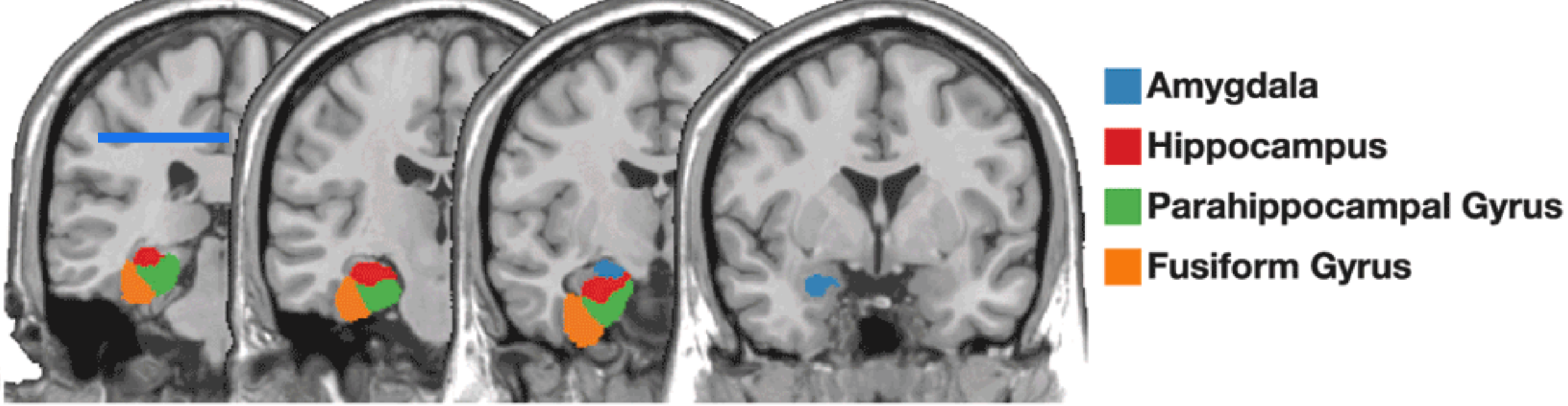
fMRI ~ group*CAG + age

fMRI ~ group*CAG + age + (1|family_numb/ursi)

Filter

	mrqid	ursi	assessment_id	age	sex	family_numb	group	CAG
1	60079215	420	97026065	18.333	Males	84	GNE	17
2	60080612	576	96392356	18.417	Females	140	GE	41
3	60137718	518	99994437	12.750	Females	284	GE	41
4	60138418	137	99997318	8.667	Females	284	GE	44
5	60151517	683	96611945	15.917	Females	271	GE	50
6	60165616	881	96715867	10.500	Females	123	GNE	22
7	60180915	982	97067273	11.833	Females	157	GNE	18
8	60181515	202	97242414	9.833	Females	157	GNE	19
9	60238515	553	97113783	18.417	Females	147	GE	47
10	60239215	750	96774599	12.250	Females	147	GE	46
11	60266813	130	96526002	15.083	Males	116	GNE	15
12	60266817	942	96645679	18.000	Males	246	GE	48
13	60267613	706	97296225	18.000	Females	116	GNE	17
14	60281418	642	99995885	18.000	Females	283	GNE	20
15	60282018	997	99994003	15.500	Females	283	GNE	20
16	60311015	106	97224532	7.833	Females	228	GE	43
17	60324016	842	97048176	16.917	Females	249	GNE	18
18	60339017	622	96163842	14.583	Males	269	GNE	18
19	60339617	640	96183111	13.583	Females	269	GNE	18
20	60569713	536	96599859	18.833	Females	161	GNE	17
21	60570313	363	96905838	15.167	Females	161	GNE	15
22	60598217	202	96281453	11.917	Females	157	GNE	19
23	60642415	467	96747872	16.167	Females	132	GNE	18

Showing 1 to 23 of 269 entries, 8 total columns



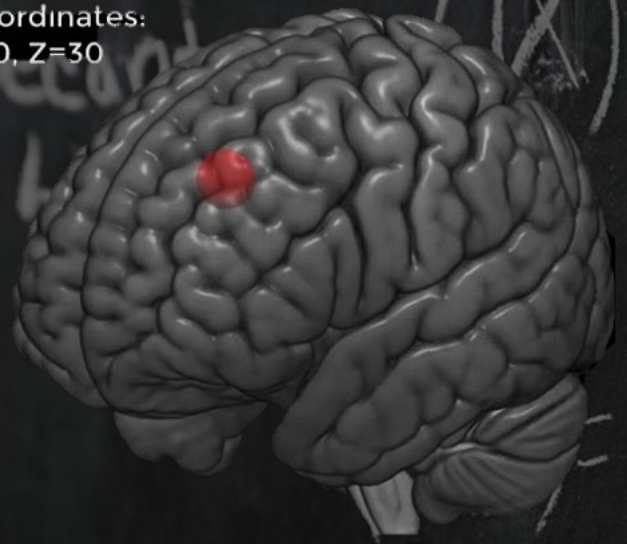
ROIs – REGION(S) OF INTEREST ANALYSES

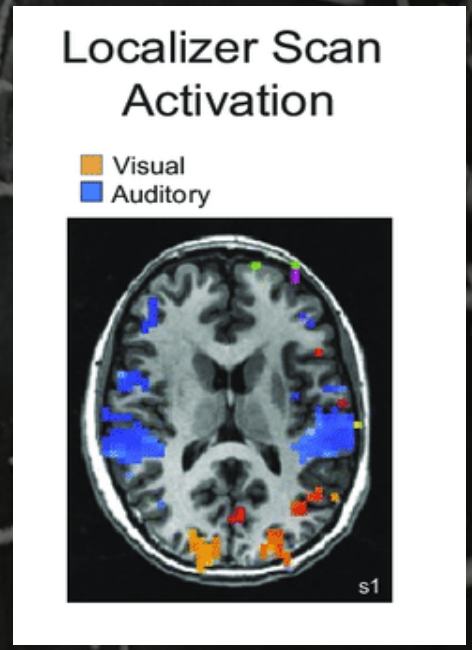
Looking at predefined anatomical areas

(Madan, 2015)

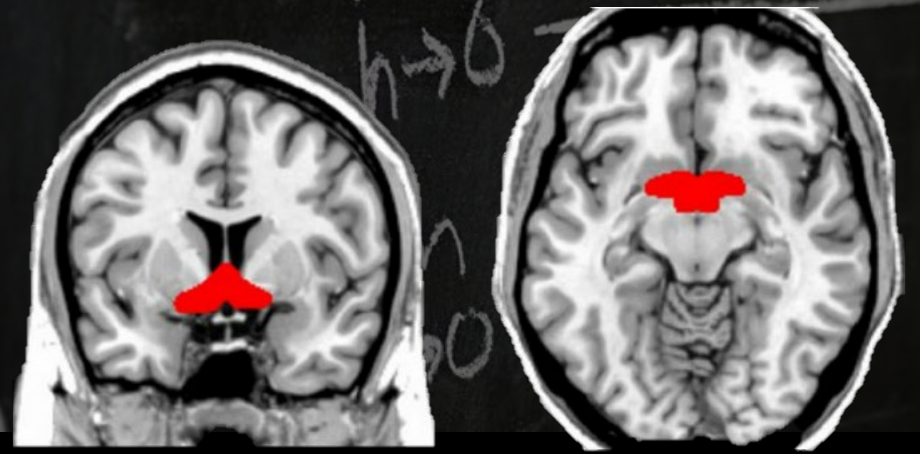
ROI Basics

e.g., MNI coordinates:
X=0, Y=30, Z=30

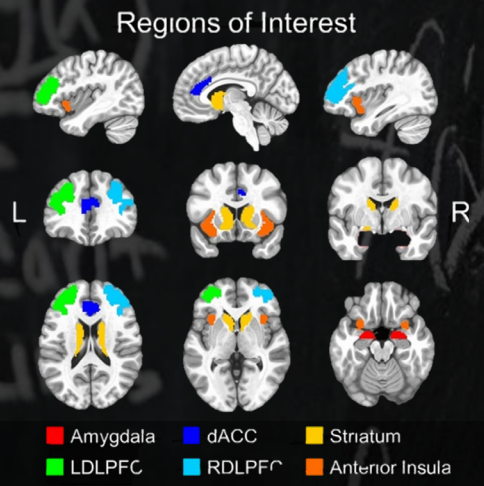


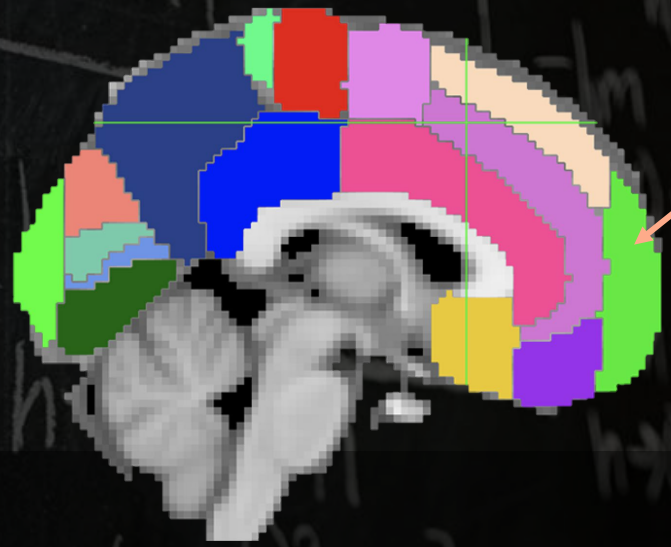


- THERE ARE MANY DIFFERENT WAYS OF DEFINING ROIS – 1) LOCALIZERS, 2) COORDINATE-BASED, 3) A PRIORI – THAT WE'RE NOT COVERING.
- SORRY.
- I AM BUT ONE MAN.
- I HAVE LET YOU DOWN.

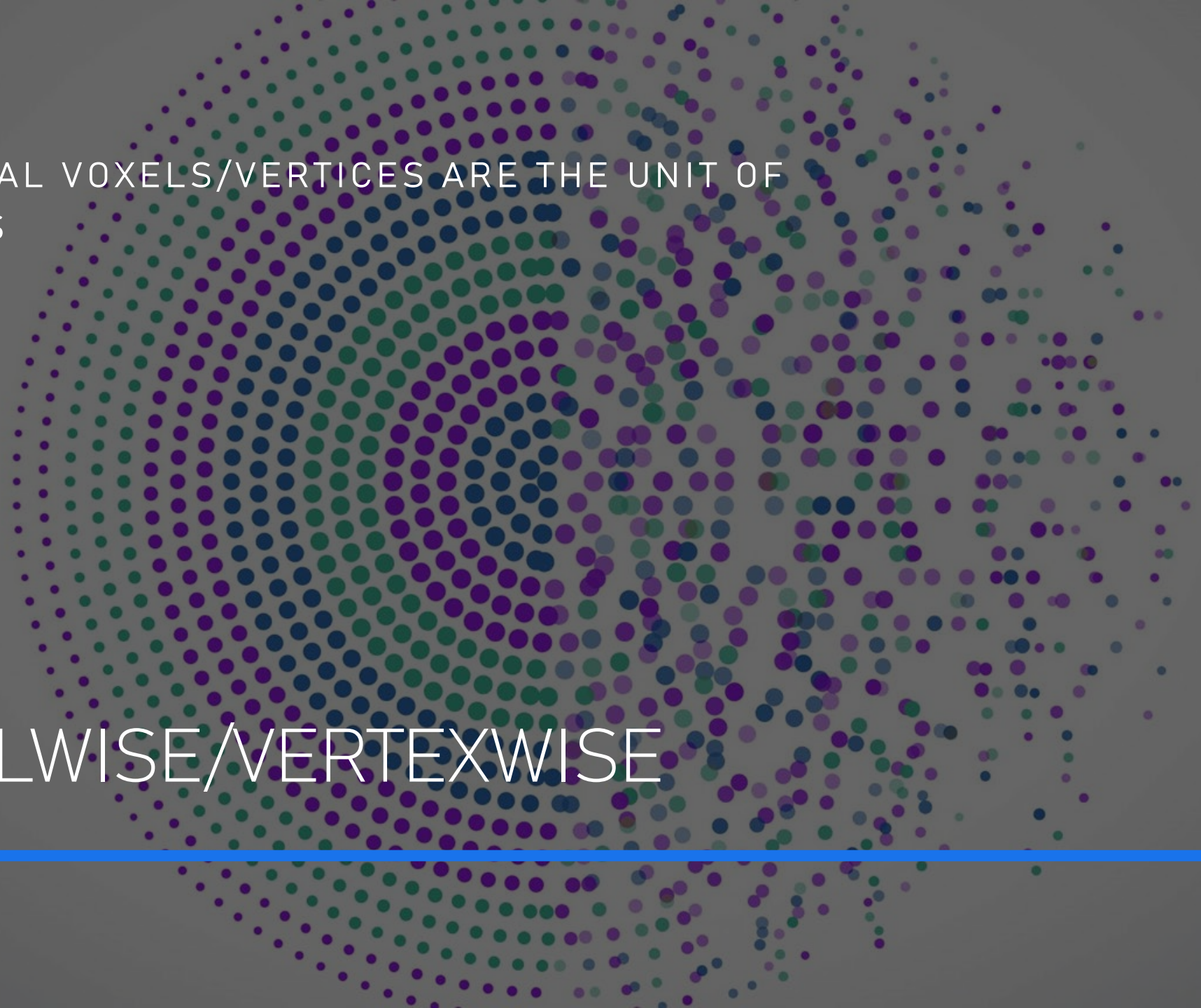


ROI Basics





- YOU PARCELLATE EITHER A) SOME OR, B) ALL OF THE BRAIN INTO ANATOMICAL **"REGIONS,"** FOCUSING ON AREAS THAT **"INTEREST"** YOU
  - THESE ANATOMICAL AREAS ARE OFTEN PREDEFINED BY A TEMPLATE
- ONCE YOU HAVE YOUR ROI(S) YOU USUALLY CONSOLIDATE ALL THE VOXELS IN THE REGION IN SOME STATISTICAL WAY
  - E.G. MEAN, MEDIAN, NON-ZERO MEAN/MEDIAN, SUM (WEIRD FLEX), SD, MIN & MAX
- CAN THEN COMPARE ROI STATS BETWEEN GROUPS OR SEE HOW THEY CHANGE OVER TIME (OR OTHER STUFF, GET CREATIVE)
- CAN DO ROI ANALYSES ON FMRI, DWI, VOLUMETRICS, T1RHO, ANYTHING THAT CAN BE PARCELLATED

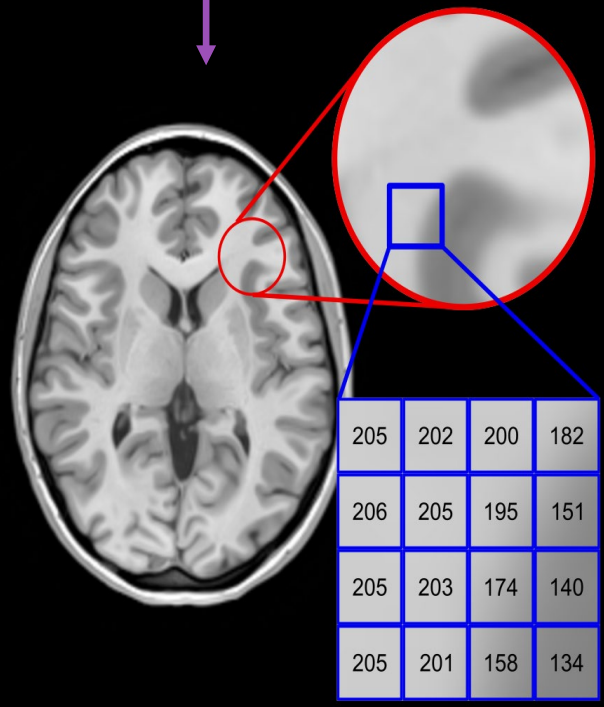


INDIVIDUAL VOXELS/VERTICES ARE THE UNIT OF ANALYSIS

VOXELWISE/VERTEXWISE

VOXELWISE BASICS

- REMEMBER: A BRAIN IMAGE IS MADE UP OF MANY THOUSANDS OF VOXELS WHERE THE VALUE OF THE VOXEL ITS INTENSITY*
- CAN PERFORM ANY REGRESSION ON ALL VOXELS
  - REGRESSION TEST IS PERFORMED ON EACH VOXEL INDIVIDUALLY
- RESULTS IN WHOLE BRAIN MAPS WITH A MAP FOR EACH EFFECT
  - E.G. VOX ~ GROUP*AGE WILL HAVE ONE MAP FOR GROUP EFFECT, ONE FOR AGE, ONE FOR THE INTERACTION

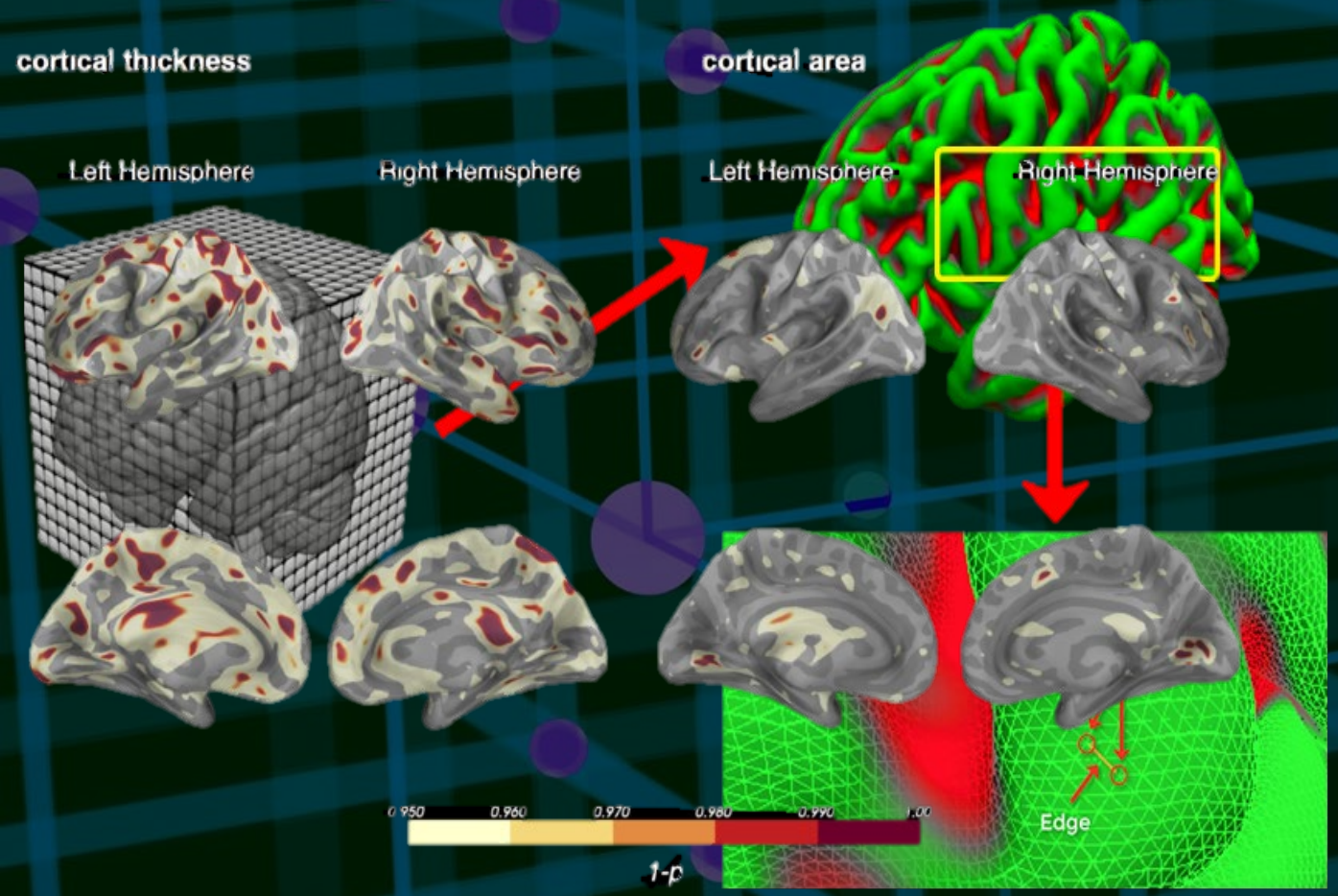


*voxel values can be something other than intensity – like a beta value – but think of it as intensity for this example

VOXELWISE REGRESSION MAP

- REMEMBER: CAN DO VOXELWISE ANALYSES ON ANY DATA
  - FRACTIONAL ANISOTROPY
- CAN GENERATE ANY STATISTICAL MAP THAT YOU WANT
  - T-MAP
  - SPECIFY IN CODE
- AT FIRST YOUR MAP HAS VALUES FOR 'ALL' VOXELS
  - CAN THRESHOLD TO GET A SENSE OF CLUSTERS AND THEIR STATISTICS

EFFECT OF FEV1 ON FA :: nii ~ FEV1_postBD + Age + packyears

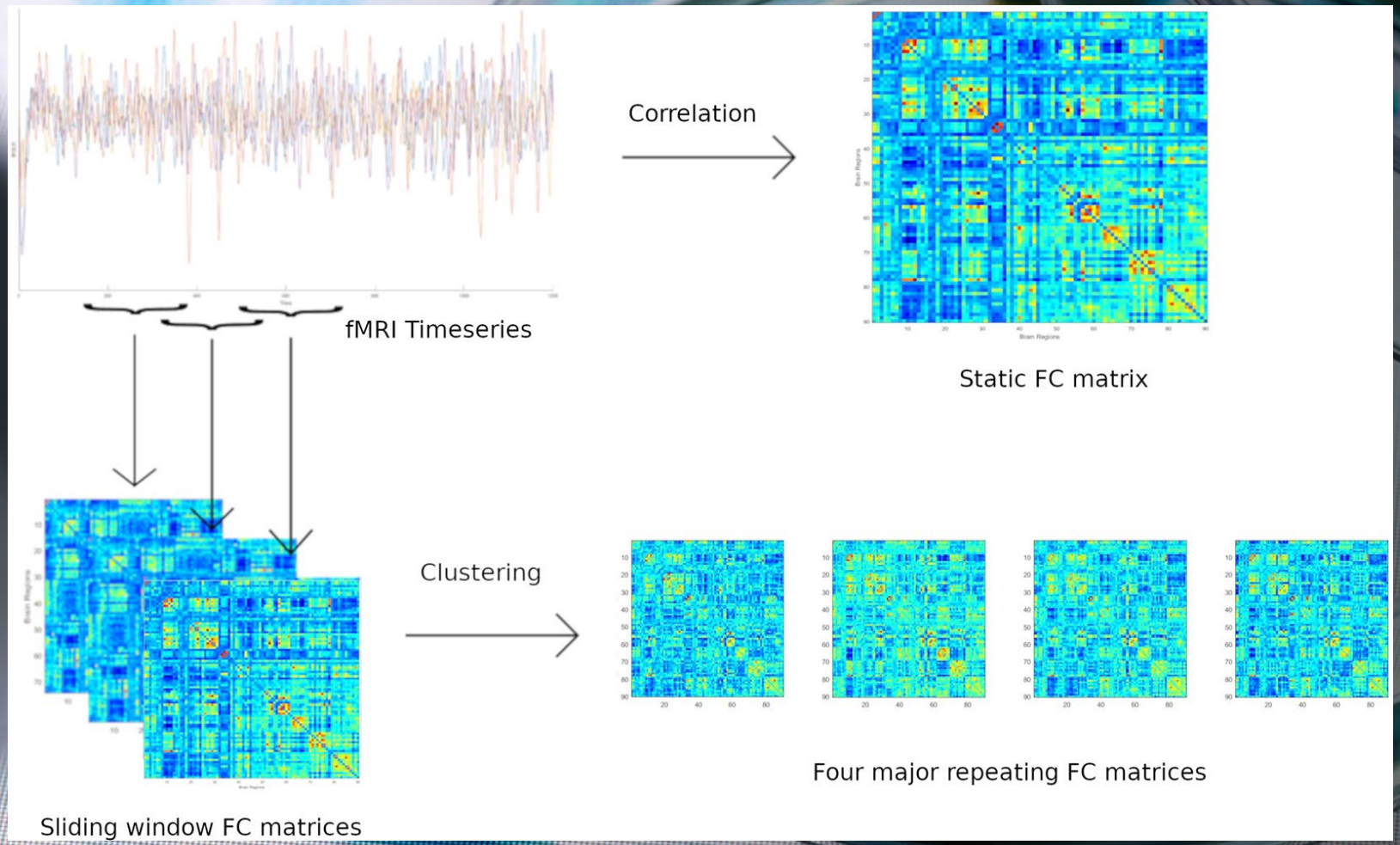


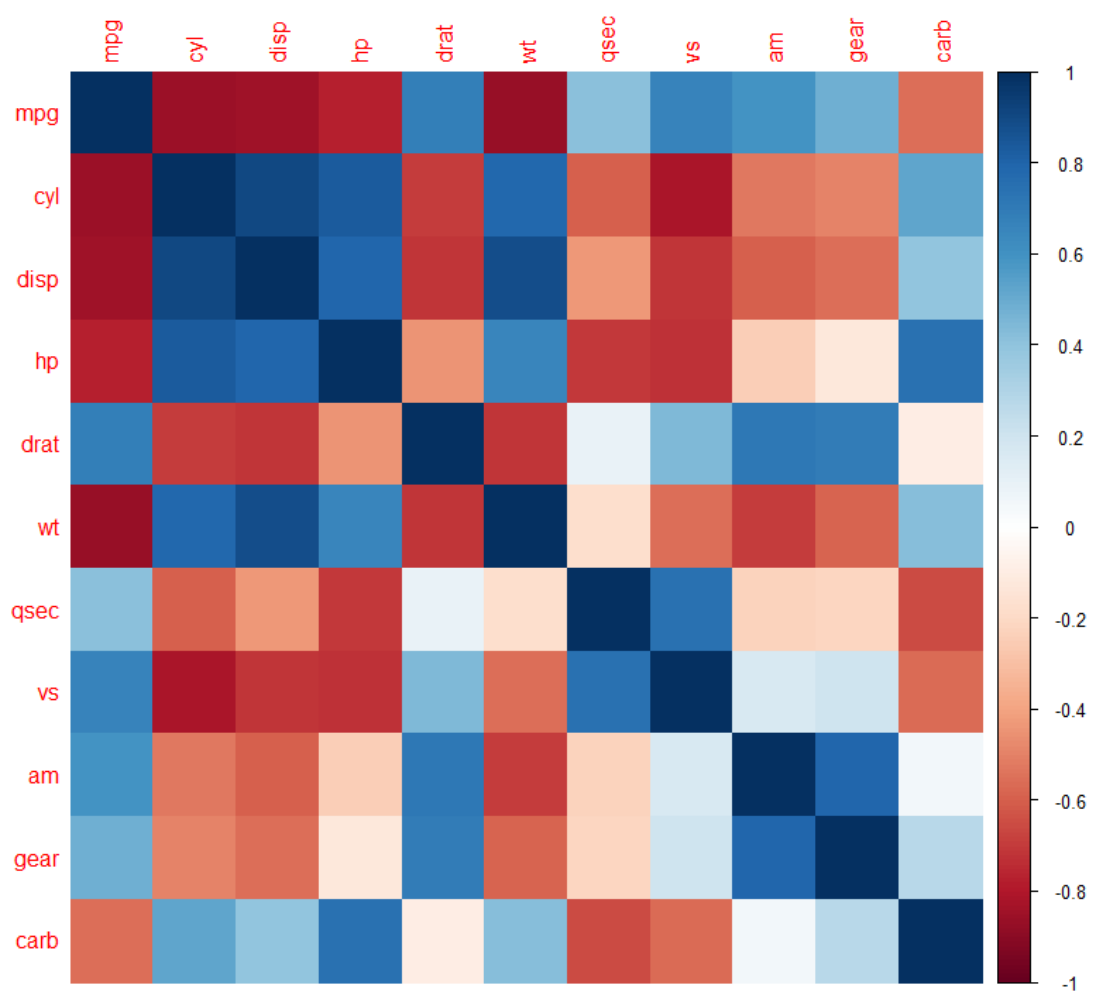
QUICK VERTEX

- BASICALLY SAME AS VOXELWISE ANALYSIS BUT ON SURFACE
- VERTEX IS POINT, EDGE IS LINE
- VERTEX REPLACES VOXEL AS VAR. OF ANALYSIS
- CALCULATE STRUCTURAL MEASUREMENTS AT EACH VERTEX
  - THICKNESS, VOLUME, SURFACE AREA
- FROM THERE CAN DO ANYTHING AN ROI ANALYSIS OR VOXELWISE ANALYSIS CAN DO*


MATRIX ANALYSES
THOSE COOL LOOKING SQUARES

MATRIX ANALYSIS AND FUNCTIONAL CONNECTIVITY



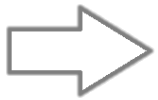


	mpg	cyl	disp	hp	drat	wt	qsec	vs	am	gear	carb
mpg	1	-0.85	-0.85	-0.78	0.68	-0.87	0.42	0.66	0.6	0.48	-0.55
cyl	-0.85	1	0.9	0.83	-0.7	0.78	-0.59	-0.81	-0.52	-0.49	0.53
disp	-0.85	0.9	1	0.79	-0.71	0.89	-0.43	-0.71	-0.59	-0.56	0.39
hp	-0.78	0.83	0.79	1	-0.45	0.66	-0.71	-0.72	-0.24	-0.13	0.75
drat	0.68	-0.7	-0.71	-0.45	1	-0.71	0.09	0.44	0.71	0.7	-0.09
wt	-0.87	0.78	0.89	0.66	-0.71	1	-0.17	-0.55	-0.69	-0.58	0.43
qsec	0.42	-0.59	-0.43	-0.71	0.09	-0.17	1	0.74	-0.23	-0.21	-0.66
vs	0.66	-0.81	-0.71	-0.72	0.44	-0.55	0.74	1	0.17	0.21	-0.57
am	0.6	-0.52	-0.59	-0.24	0.71	-0.69	-0.23	0.17	1	0.79	0.06
gear	0.48	-0.49	-0.56	-0.13	0.7	-0.58	-0.21	0.21	0.79	1	0.27
carb	-0.55	0.53	0.39	0.75	-0.09	0.43	-0.66	-0.57	0.06	0.27	1

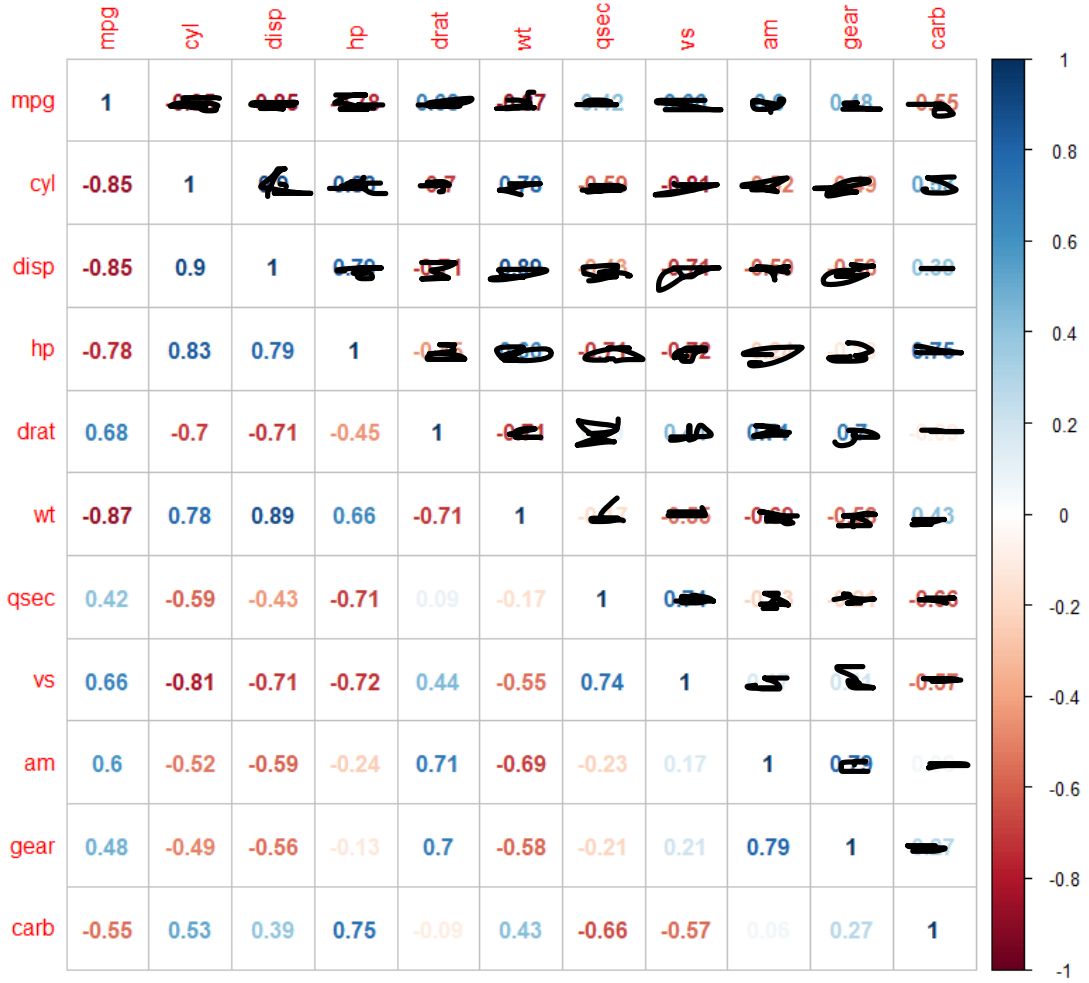
WHAT DOES A MATRIX LOOK LIKE BEHIND THE SCENES

- Can keep lower triangle of matrix and vectorize (read. unfurl) it into a single row and then perform modeling on it
- Matrix CAN BE a variable in models
  - RSA (Relational Similarity Analysis)

1	1	0
4	2	1
0	2	1



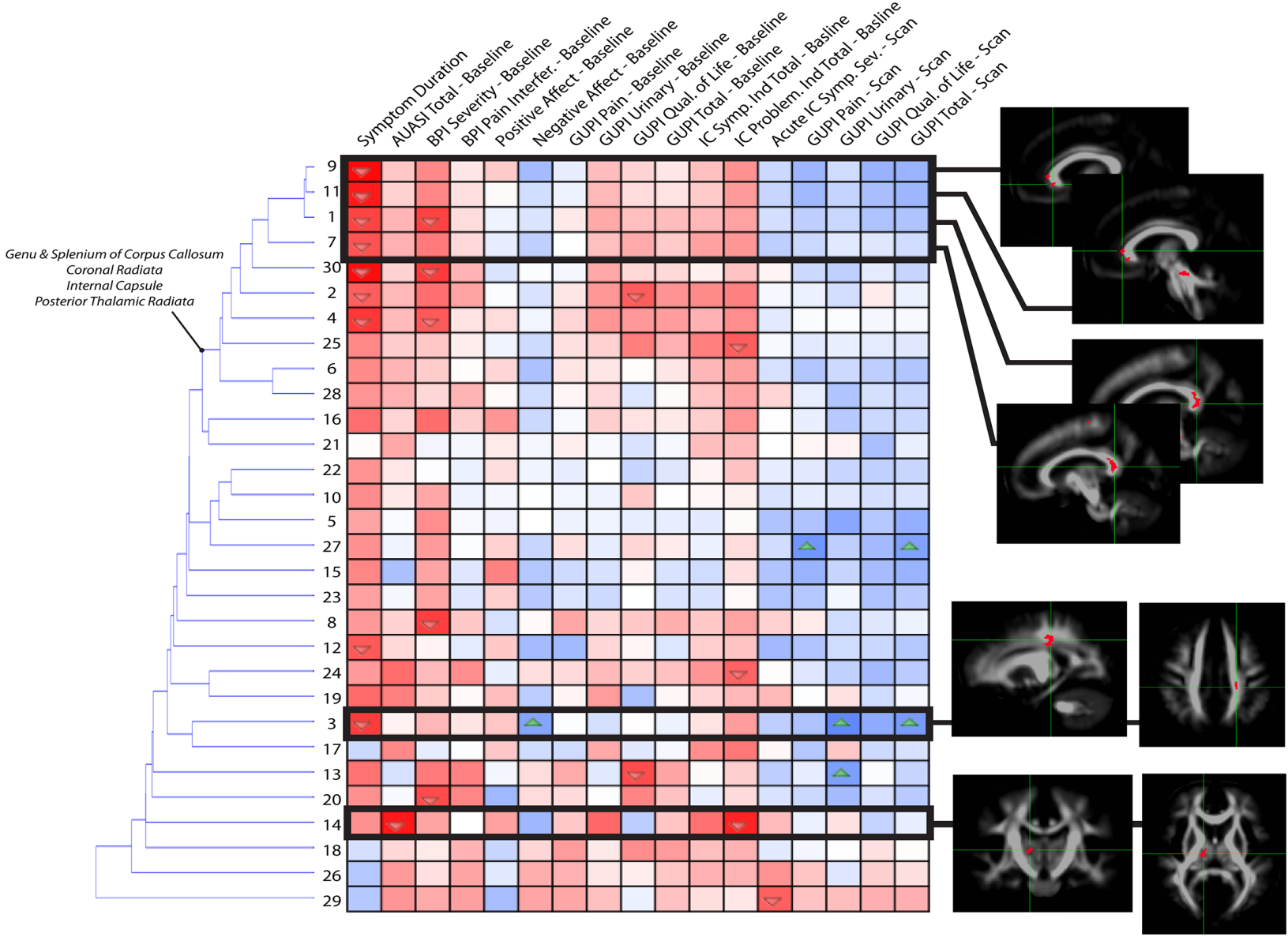
1
1
0
4
2
1
0
2
1



CAN MODEL WITH A MATRIX

What can be a matrix?

- IF YOU HAVE 2 MEASURES YOU CAN CORRELATE (OR COMPARE SOMEHOW) AND MANY PAIRS BEING COMPARED YOU CAN MAKE A MATRIX OUT OF IT
- YOU CAN MAKE SOME SICK STUFF



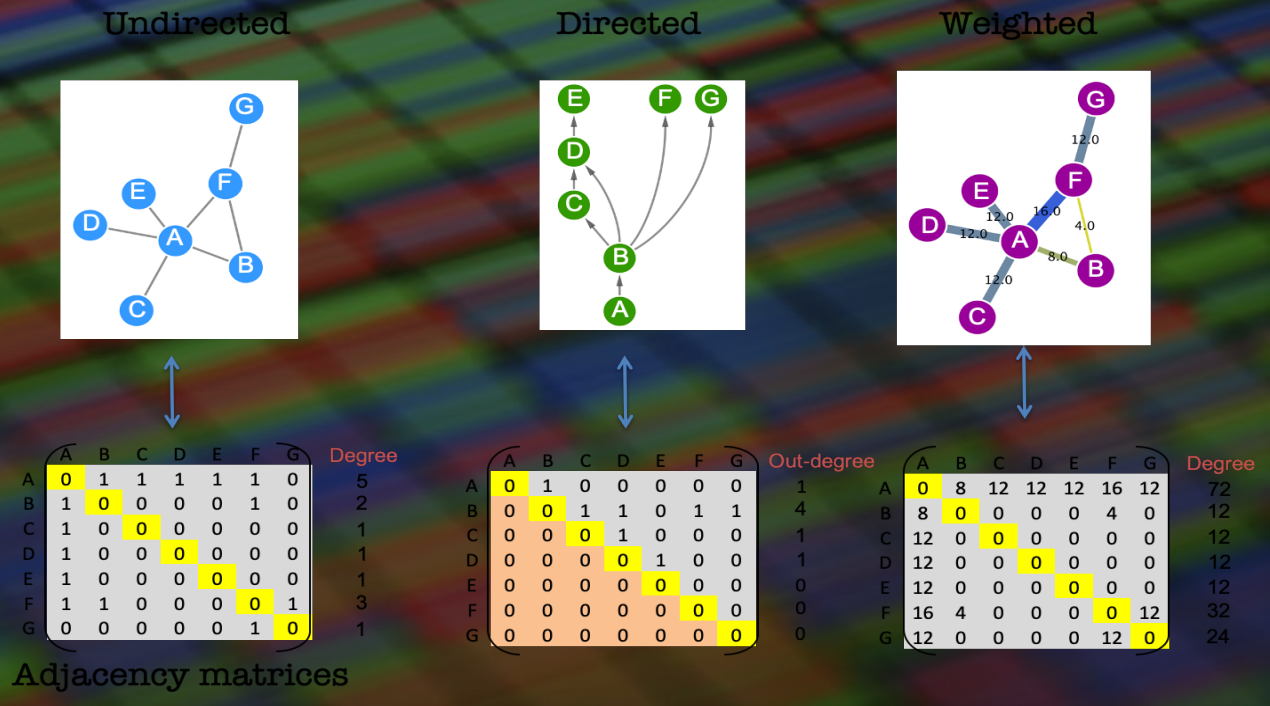
- CAN PERFORM A NETWORK ANALYSIS FROM A CORRELATION MATRIX

1. Transform corr. matrix to UNDIRECTED adjacency matrix
   - Transform every non-zero corr. to 1 OR
   - Threshold corr. so every value above a thresh. is set to 1 – THIS IS AN EDGE
2. Corr. mat can be seen as UNDIRECTED adjacency matrix of graph where the partial correlations represent edge weights

- NETWORKS CAN BE EXPRESSED BY ADJACENCY MATRICES

NOW YOU'RE A GRAPH ANALYST
GOOD JOB

Correlation Matrices are a Small Step from Graph Analyses



MULTIPLE COMPARISONS

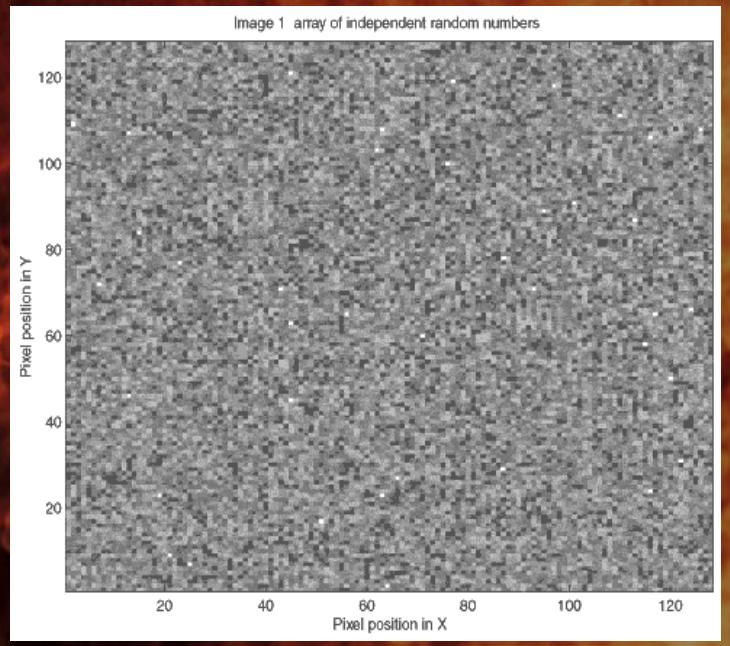
WELCOME TO HELL!

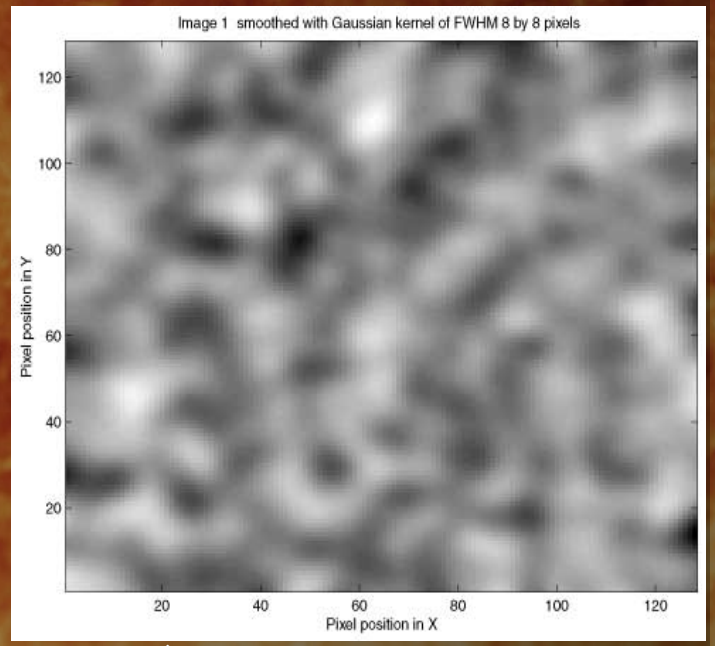
MULTIPLE COMPARISON CORRECTION BASICS

- EVERYONE AGREES MCC SHOULD BE PERFORMED
  - WHY?
    - WE HAVE DONE STATS OF THOUSANDS OF VOXELS
    - MAKES US MORE LIKELY TO MAKE TYPE 1 ERROR
    - HAVE TO CORRECT FOR THAT
- NO ONE AGREES ON HOW TO DO MCC
  - WAIT, REALLY?
    - YEAH, LITERALLY NO ONE
  - WHAT DO YOU DO?
    - YOU PICK WHATEVER METHOD YOU LIKE BEST
  - DON'T PEOPLE GET IN TROUBLE FOR THAT?
    - YEAH ALL THE TIME
  - ISNT THERE, LIKE, A MATHEMATICALLY CORRECT WAY?
    - NO
  - WOW
    - YEAH
  - THIS IS CRAZY
    - SORRY

WHY IS MULTIPLE COMPARISON CORRECTION SO HARD

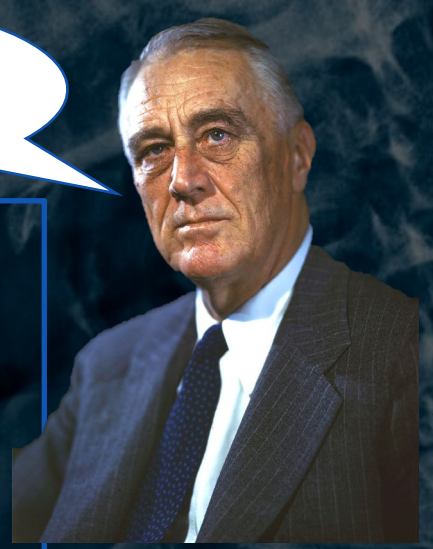
- "BONNFERONI" CORRECTION
  - First correction most think of
  - Correct by total number of voxels
    - Each voxel p-value * #voxels
    - **DO NOT DO – THIS IS INSANE**
- Voxels are not independent observations
  - **RESELS (RESolution ELEmentS)**





  - RESEL number is similar (not equal) to number of independent observations
- TAKE HOME POINT: Voxels around each other are smoothed together so are more similar to each other – each voxel is NOT an independent observation

(Worsley et al., 2019)



- DO SOME FALSE DISCOVERY RATE (FDR) CORRECTION, I GUESS
  - CONTROL THE PROPORTION OF POSITIVE RESULTS THAT ARE FALSE POSITIVES
- SELECT FDR THRESHOLD (Q)
  - IMPORTANT: **YOU** CONTROL THIS THRESHOLD (E.G. 0.05)
  - REPRESENTS 5% "SIGNIFICANT" VOXELS ARE FALSE
- GIVEN Q, A SINGLE-VOXEL THRESHOLD IS CALCULATED
  - THIS IS NUMBER YOU THRESH. VOXELS AT
- MORE SENSITIVE TO LARGE REGION ACTIVTY

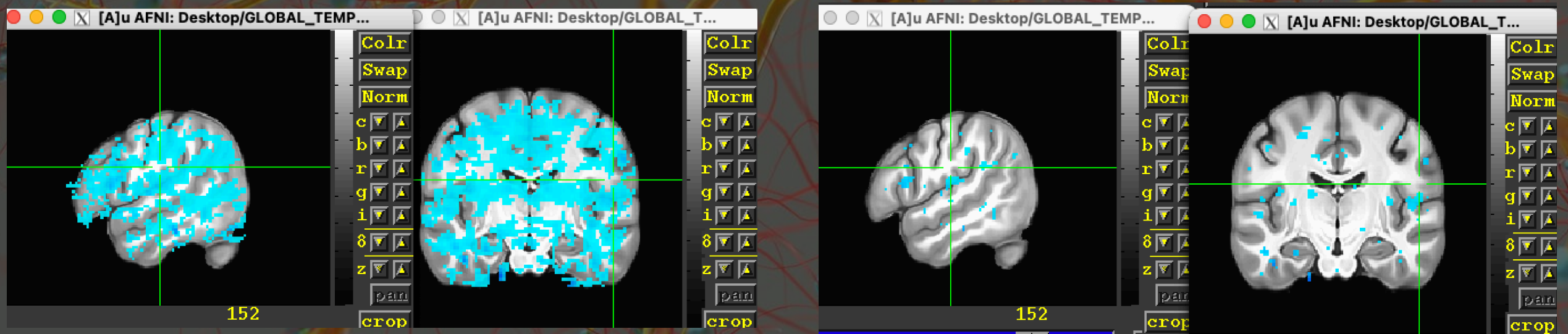
SO WHAT'S TO BE DONE?

FDR(data = NULL, sp.cols = NULL, var.cols = NULL, pvalues = NULL, model.type = NULL, family = "auto", correction = "fdr", q = 0.05)

FDR & CLUSTERING

ONCE YOU HAVE YOUR SINGLE-VOXEL THRESHOLD VALUE FROM FDR YOU CAN THRESHOLD FURTHER BY PERFORMING CLUSTER THRESHOLDING

- ONLY ALLOW CLUSTERS LARGER THAN A CERTAIN NUMBER
- IMPORTANT: IF YOU THRESHOLD BY CLUSTER YOU CANNOT MAKE CONCLUSIONS ABOUT INDIVIDUAL VOXELS



```r
args <- commandArgs(trailingOnly=TRUE)
which.run <- as.numeric(args[1])
num.cores <- as.numeric(args[2])

library(doParallel)
#library(lmerTest)
library(car)
library(nifti.io)
library(foreach)

dir.project <- "/Shared/hothlab/copd_bids"

FORM.ls <- rep("nii ~ FEV1_postBD + Age + packyears",1)
FUNC.ls <- rep("lm",1)
MOD.ls <- c("FA", "AD", "RD")
DATA.ls <- paste0(dir.project, "/derivatives/dwi/scalars_HCPICBM_1mm/", MOD.ls, "/unzip")
NAME.ls <- paste0("Regression_Hopkins-", MOD.ls)

dir.analysis <- paste0(dir.project, "/derivatives/dwi/analyses/DWI_Hopkins_20210506")

# load data frame for analysis
pid.var <- "subject_id"
sid.var <- "session_id"
ogf <- read.csv(sprintf("%s/NeuroImaging_Dataset_EVERSMOKERS_only.csv", dir.project))
ogf <- ogf[!is.na(ogf$FEV1_postBD),]


# set up data for run
dir.data <- DATA.ls[which.run]
dir.save <- paste0(dir.analysis, "/", NAME.ls[which.run])
# dir.create cannot create recursive directories
dir.create(dir.save, showWarnings = FALSE)
save.prefix <- "DWI_FEV1"
MODEL.NAME <- NAME.ls[which.run]
FORM <- FORM.ls[which.run]
FUNC <- FUNC.ls[which.run]

# match subjects to data
ogf$fls <- character(nrow(ogf))
for (i in 1:nrow(ogf)) {
  tname <- tname <- list.files(dir.data, pattern=paste0("sub-", ogf[i, pid.var], "_ses-", ogf[i, sid.var]), full.names = TRUE)
  if (length(tname) != 0) {
    ogf$fls[i] <- tname[1]
  }
}
#ogf
ogf <- ogf[ogf$fls != "", ]
#only keep good people and people with valid dti
ogf <- ogf[ogf$GOLD_4_Groups != 1, ]
```