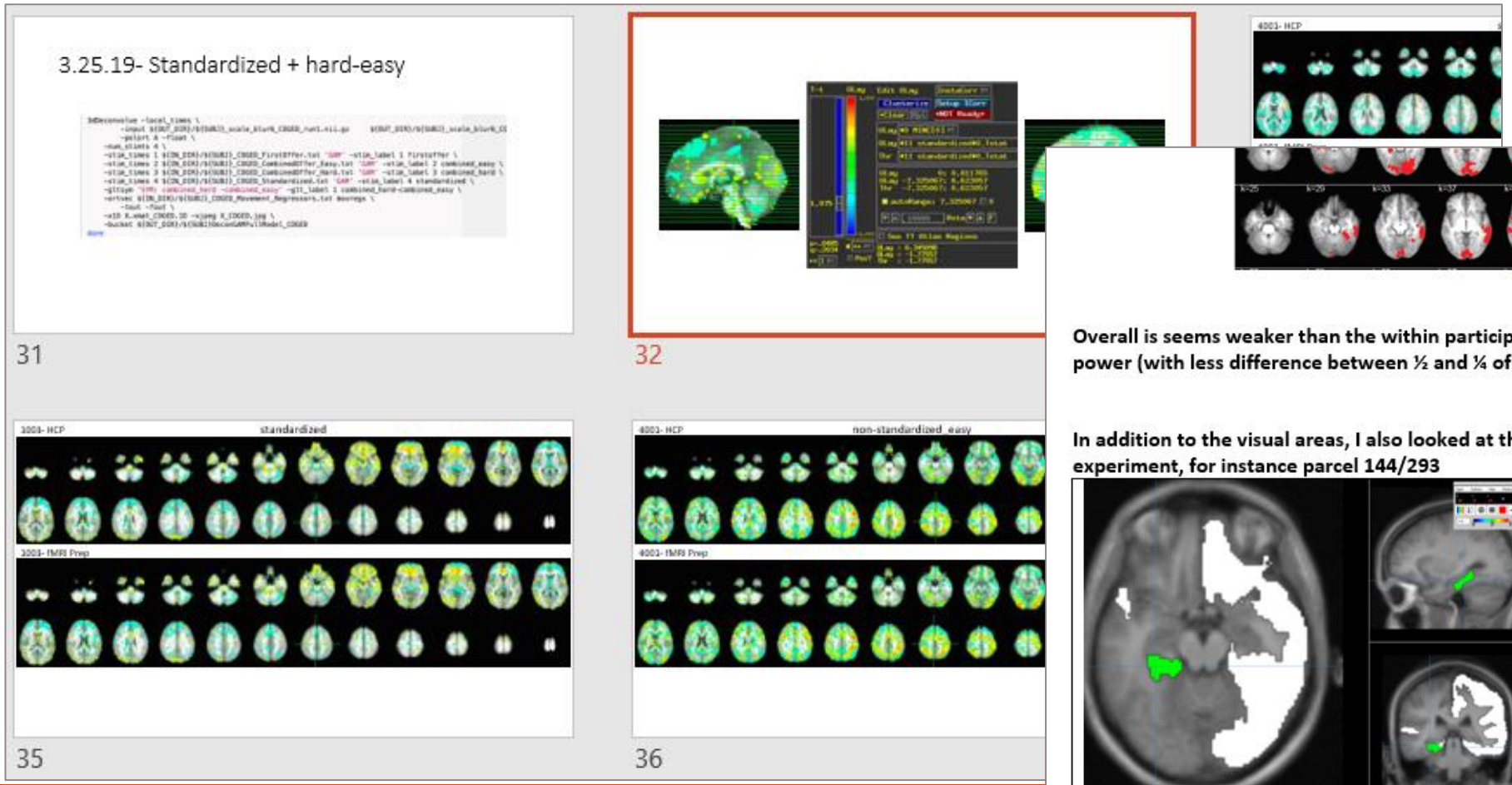
knitr for neuroimagers

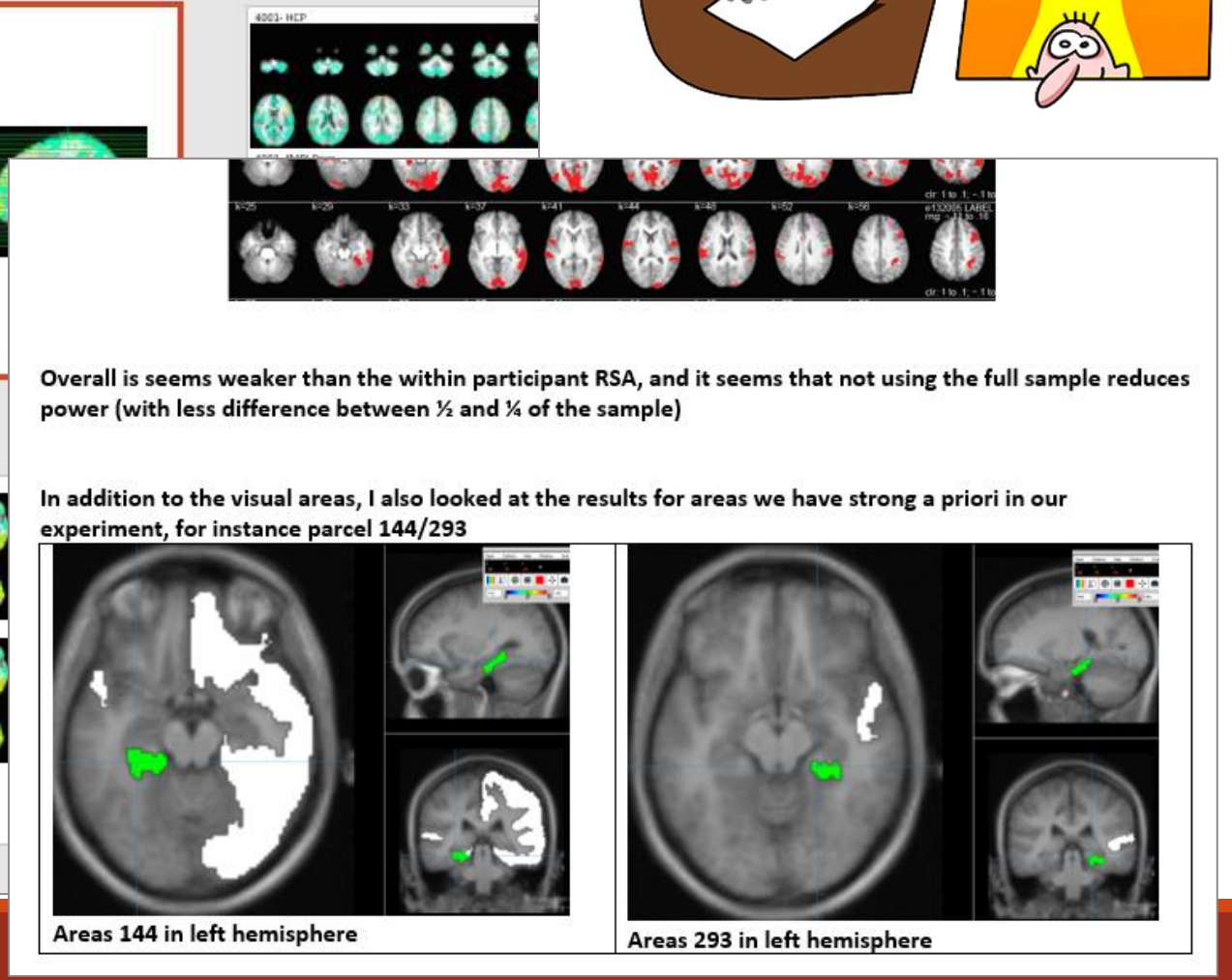
Joset A. Etzel, PhD
jetzel@wustl.edu | mvpa.blogspot.com | @JosetAEtzel
Cognitive Control and Psychopathology Lab
Washington University in St. Louis (USA)

I have some results! Time to share them with my colleagues.

… but how? They're brain pictures and statistical tables.

One solution: copy-paste screenshots of the images into Word or PowerPoint, add explanatory text, and send those documents.





A few big problems with this copy-paste workflow:

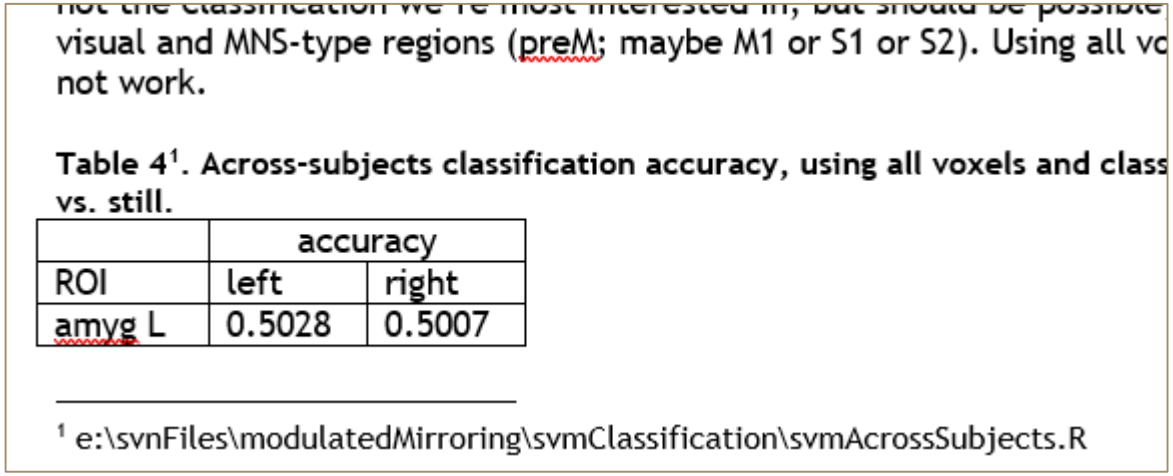
1: It takes a long time to do once … and **just as long** to do twice. (… or three times.)

2: It's easy to make a mistake (e.g., copy in the wrong picture) and almost impossible to **check** for accuracy without repeating everything over yet again.

3: It's **hard to reproduce** because it breaks the link between the description of the result and its source.

Footnotes! … but that was three computers ago, and many of the captions point to the same R input file.

… even if I found the code, I wouldn't be certain that the **version** of the file I found is the one that made the result.



visual and MNS-type regions (preM; maybe M1 or S1 or S2). Using all vo
not work.

Table 4[1]. Across-subjects classification accuracy, using all voxels and class vs. still.

	accuracy	
ROI	left	right
amyg L	0.5028	0.5007

[1] e:\svnFiles\modulatedMirroring\svmClassification\svmAcrossSubjects.R

The Solution: "**dynamic report generation**"

I use knitr, R, and LaTeX to make pdfs; other language options exist for the code (e.g., python), text (e.g., markdown), and output (e.g., html).

https://yihui.name/knitr/

Dynamic report generation?

Combines the code and text in one file (a bit like Jupyter notebooks), which is then compiled to produce a static **document**.

LaTeX

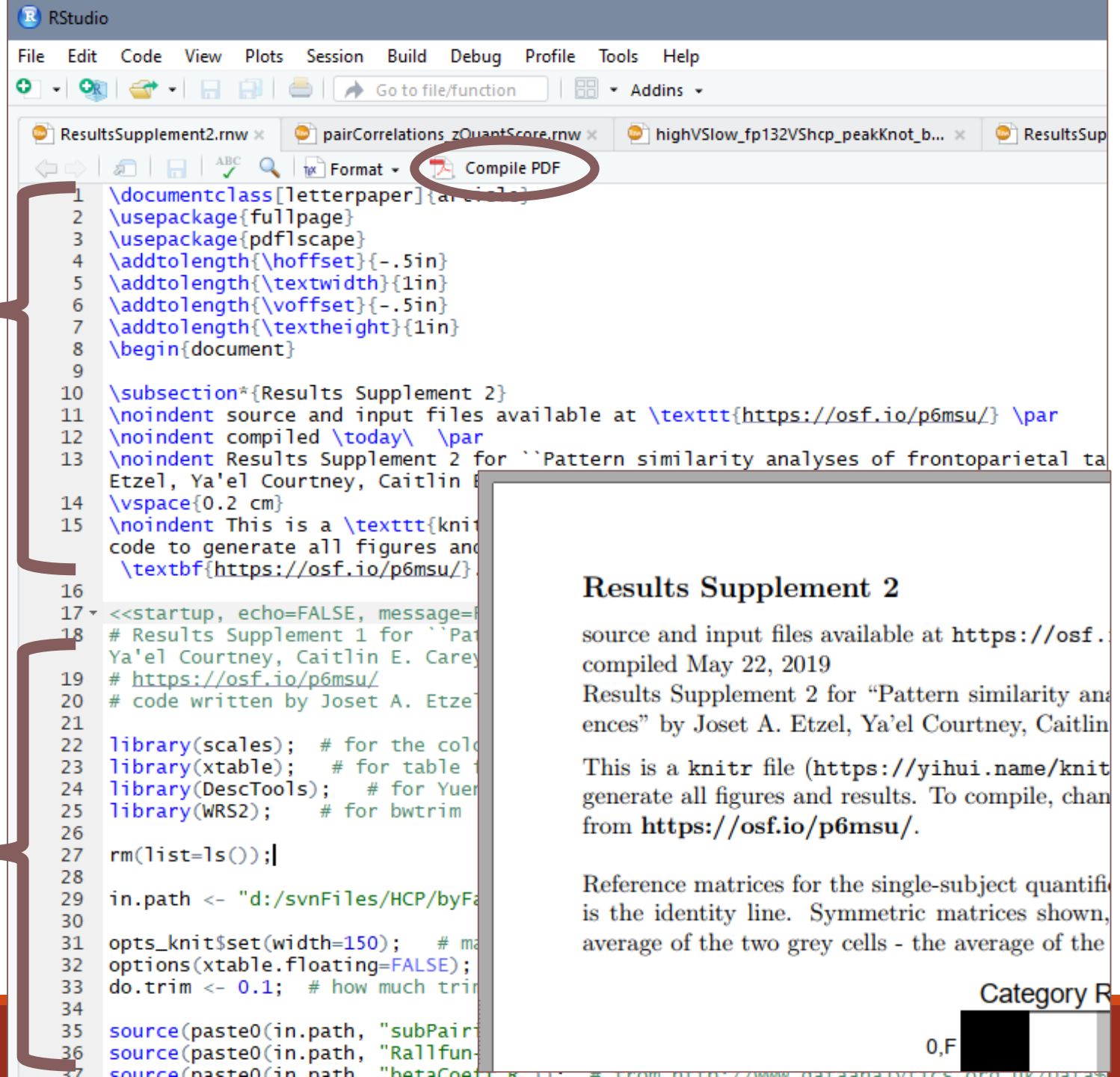
knitr documents are easiest to make with **RStudio**: it connects R, LaTeX, and the pdf viewer, so all you do is click Compile PDF.

LaTeX can be very annoying, but I've found it worth the trouble.

(RStudio helps with a menu of simple LaTeX formatting syntax; I didn't know LaTeX before, and still don't use it other than with knitr.)

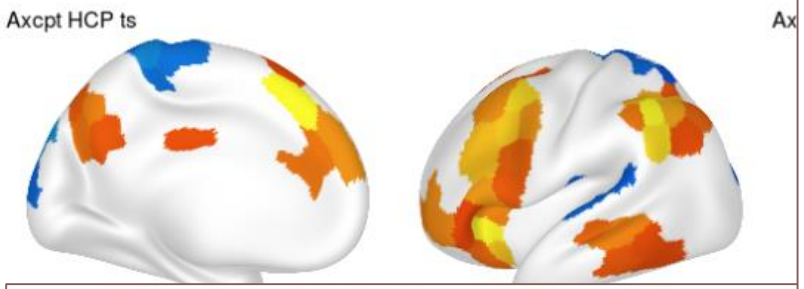
R

"If you want precise control of a document, LaTeX is a great way to go." - Karl Broman
https://kbroman.org/knitr_knutshell/pages/latex.html



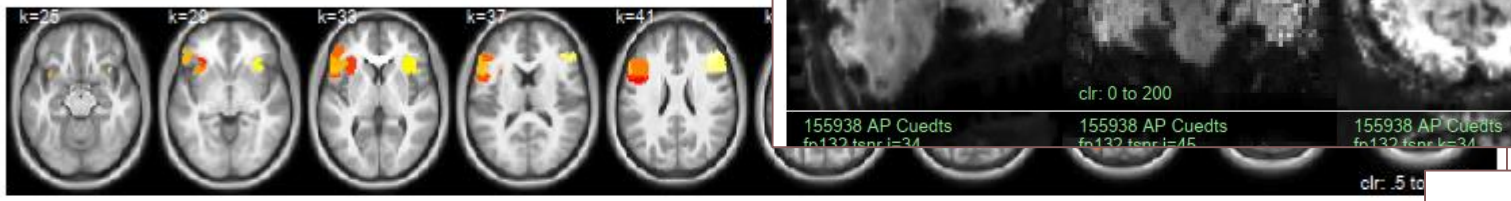
Can R & knitr make nice-enough looking brains to avoid copy-pasting? Yes.

Continuous t-values from the surface-only t-tests.

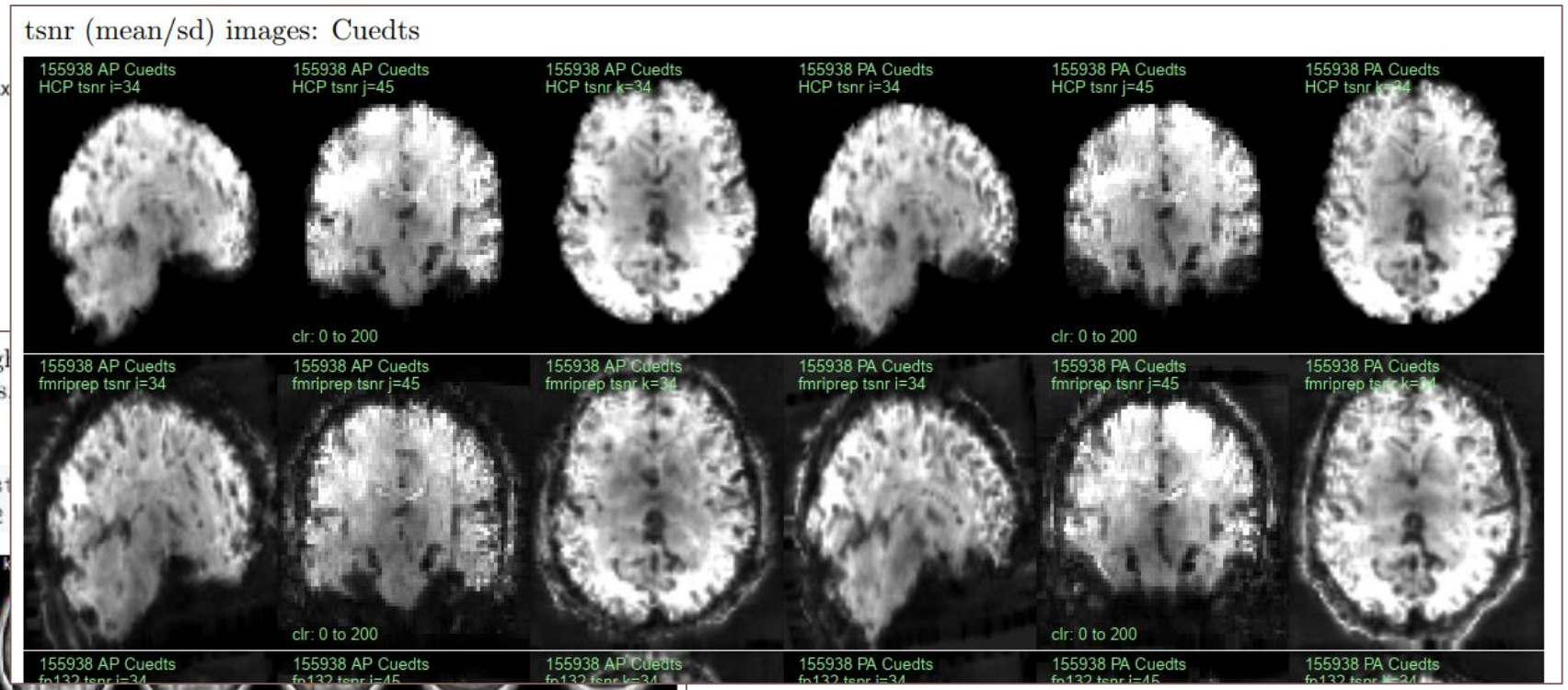


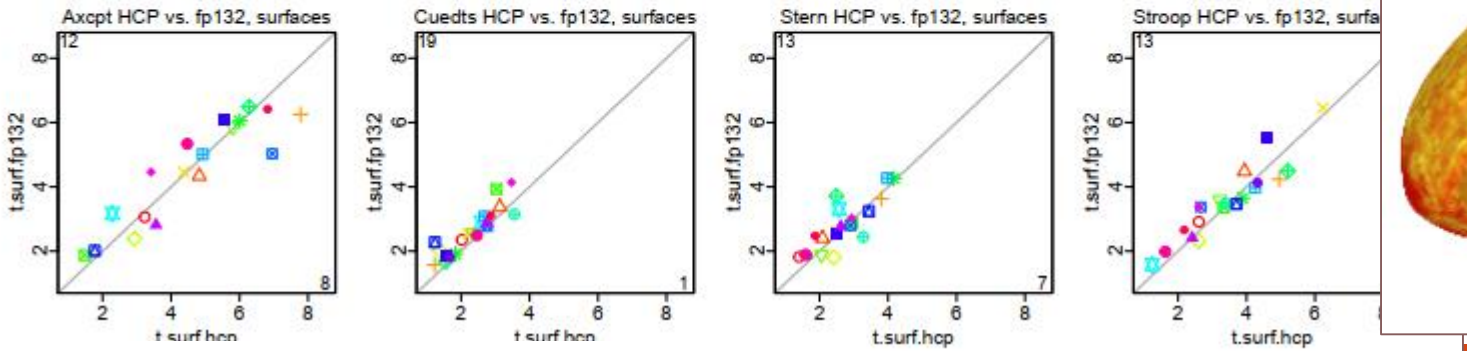
Scatterplots, only including the parcels with the t-test for high
HCP) and all four tasks. The first image shows these parcels
the number of parcels on each side of the x=y line.

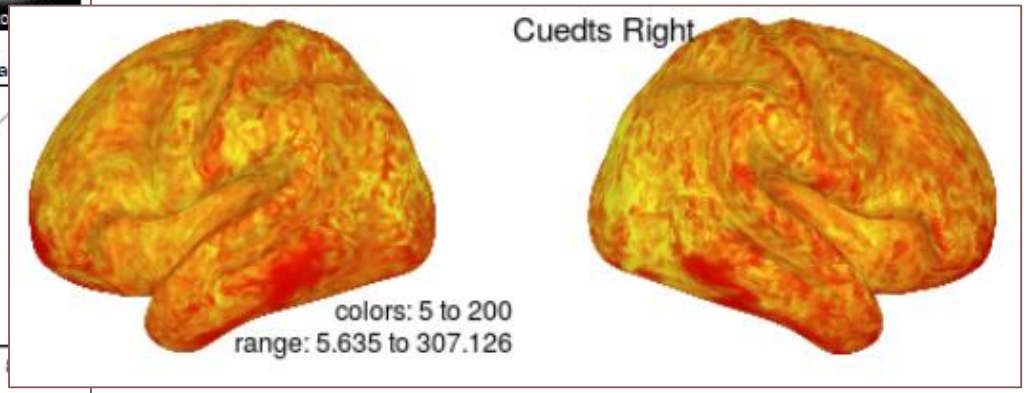
```
## [1] "the 20 parcels with t > 1 in all four t-test
##  [1]  90  91  99 101 105 130 136 140 143 148 172
```



Surface (gifti) parcels colored by statistical results.







Mixing text, brain images, and graphs.

I'm sold! How can I get started with knitr?

Can't actually show you in this lightning talk … but I wrote a few tutorials (which include the brain plotting functions):

http://mvpa.blogspot.com/2014/12/tutorial-knitr-for-neuroimagers.html
http://mvpa.blogspot.com/2018/06/tutorial-plotting-gifti-images-in-r.html

(also can click the **knitr** label on mvpa.blogspot.com)

more of my knitr examples:

osf projects: https://osf.io/p6msu/ & https://osf.io/c3s75/

https://github.com/ccplabwustl/dualmechanisms has knitr files (and example data) for analyzing behavioral performance during scanning the DMCC tasks; more will be added soon.

There **is** a learning curve, especially if you're new to programming and/or LaTeX, but it will quickly start paying off.

Start with my two tutorials: that should be enough R and LaTeX for everyday reports (with brains!).

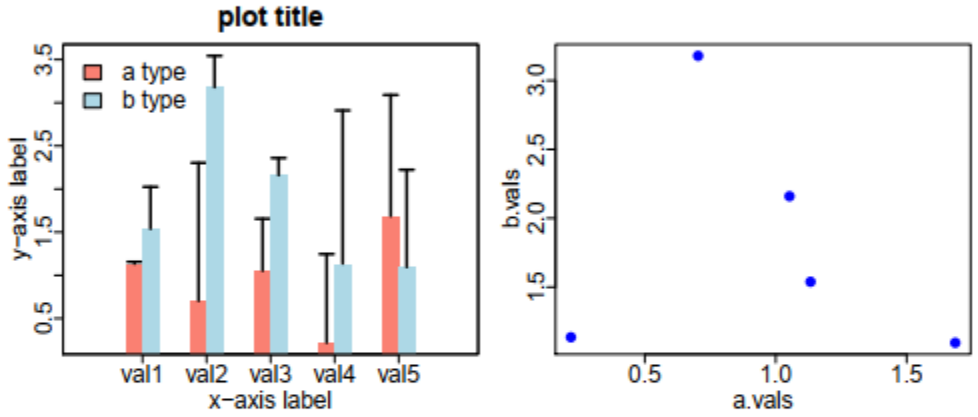
brainPlotsDemo.rnw

source file: `d:/svnFiles/demoCode/knitRdemo/knitr/brainPlotsDemo.rnw`
compiled December 17, 2014
This file was written by Jo Etzel (`jetzel@artsci.wustl.edu`) in December, 2014. It may be adapted for personal use, but should be cited rather than redistributed.

The purpose of this document is to demonstrate using knitr, particularly to display "brain blob" images, but also graphs, tables, and text. The key function for plotting NIfTI images is the `make.plots` function in the `startup` code block. I (Jo Etzel) wrote the code, functions, and text in this document but didn't create knitr! The `knitr` package itself is described

two non-fMRI plots, side-by-side

This is two plots, side-by-side, to demonstrate using `layout` and displaying plots.



Finally, here is a table showing the plotted values, with code `echo`-ed. Fancier tables can be made with `ascii`, but I often prefer simply printing tables like this.

```
# knitr has nice syntax coloring when showing code
tbl <- data.frame(paste0("val", 1:5), a.vals, b.vals)
colnames(tbl) <- c("case", " a.vals", " b.vals")
tbl
```

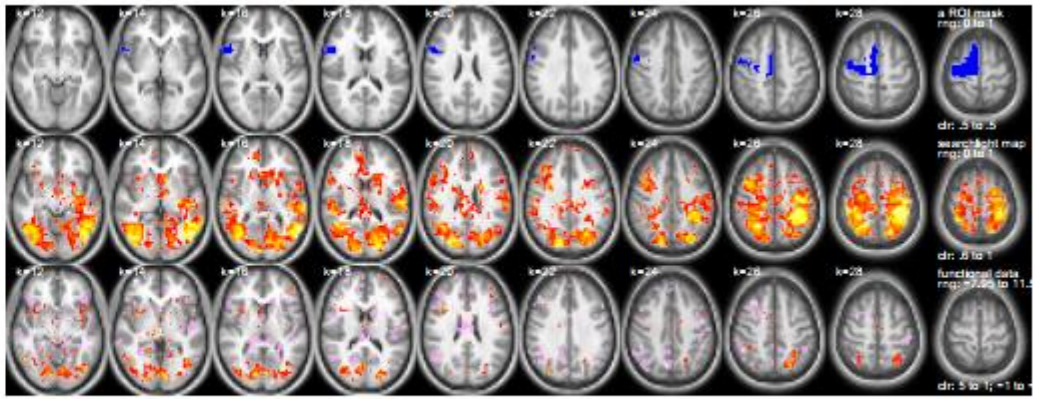
```
##   case  a.vals  b.vals
## 1 val1  1.1329   1.538
## 2 val2  0.7036   3.179
## 3 val3  1.0525   2.159
## 4 val4  0.2179   1.136
## 5 val5  1.6854   1.095
```

ROI masks and blob-type plots on an anatomical underlay

The color scaling in `make.plots` follows common neuroimaging conventions: positive values are hot, negative are cool. The function sets values more extreme than the values sent in `plot.lim` (and displayed as `clr` on the far-right slice) to the extreme value. For example, if the function is called with `plot.lim=c(0.5, 0.7)`, a voxel with a value of 0.8 will be colored bright yellow (most extreme hot color), while one with a value of 0.4 will *not* be plotted at all. Likewise, when the color scale includes both positive and negative values (e.g. `plot.lim=c(-1, 1)`), values more extreme than the first value will be given most extreme negative color (e.g. a voxel with a value of -2 will be colored cyan). Values around zero can be omitted (not plotted) by setting `neg.center` and `pos.center`. For example, if `make.plots` is called with `plot.lim=c(-1, 1)`, `neg.center=-0.5`, `pos.center=0.5`, voxels with values between -0.5 and 0.5 will not be plotted; ones with values -0.5 to -1 will be plotted with cool colors, and voxels with 0.5 to 1 will be plotted with hot values (more extreme values are always plotted, so a voxel of 1.5 will be plotted bright yellow).

The images are labeled the slice number (in the *i,j,k* sense: the voxel array, not anatomic space), with the displayed slices set in `plot.slices`. The right-most slice gives a title for the plot, along with the range (`rng`) of values in the overlay image as a whole, not just the displayed slices. The color scaling (`clr`) is also listed.

Note that with `cache=TRUE` (as it is in `code1`) you need to make a change in the code block for it to re-execute the code; otherwise it will just read from the cache. So, changing something in the startup code block only will not change the plotted image. Adding a blank line or space (meaningless change) is sufficent to trigger knitr to re-execute the code block.



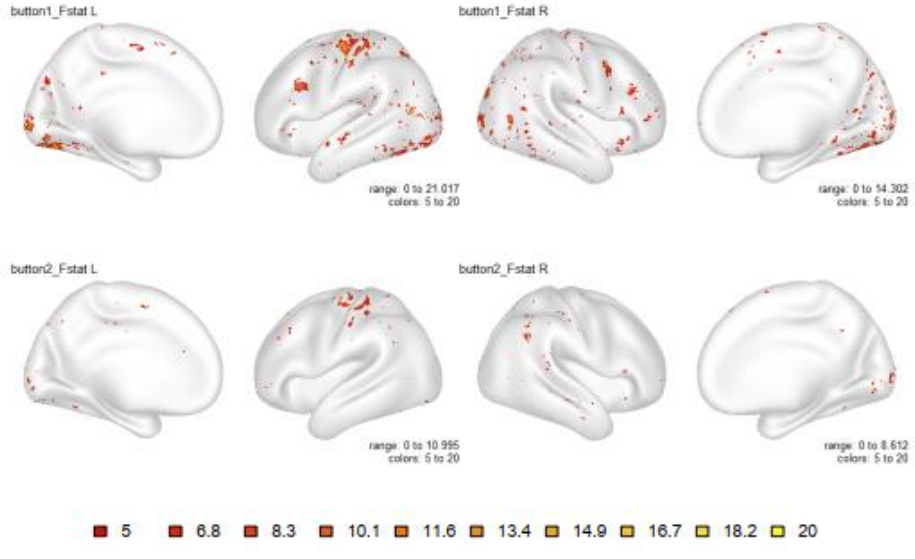
output from tutorial at http://mvpa.blogspot.com/2014/12/tutorial-knitr-for-neuroimagers.html

giftiPlottingDemo.rnw

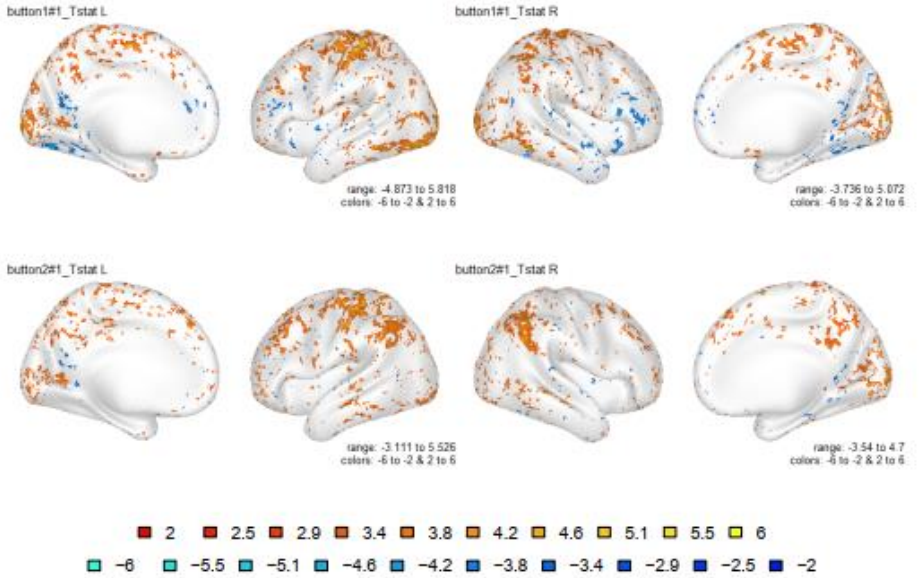
code written by Joset A. Etzel (jetzel@wustl.edu) on 1 June 2018 and released on mvpa.blogspot.com.
compiled April 10, 2019

This code demonstrates how to read a GIfTI brain image in to R and display in a standard format. The overlay images are t and F statistics from a single-subject afni GLM fitting button-pushes. Two buttons, each pressed with the right hand.

F-statistics: only positive (hot) values.



t-statistics: both positive (hot) and negative (cool) values.



Coloring MMP **parcels**, rather than vertices. This colors MMP parcels 1, 10, and 15 red (bilaterally), to match the demo at `http://mvpa.blogspot.com/2017/11/tutorial-assigning-arbitrary-values-to.html`.

output from tutorial at http://mvpa.blogspot.com/2018/06/tutorial-plotting-gifti-images-in-r.html

Remember:

the compiled .pdf file is not interactive (useful for archiving and sharing), but that means it's half of a pair: the source .rnw **must be kept along with the compiled file** to gain the reproducibility benefits of dynamic report generation.

Some of our project storage: **pairs** of pdf and .rnw files.

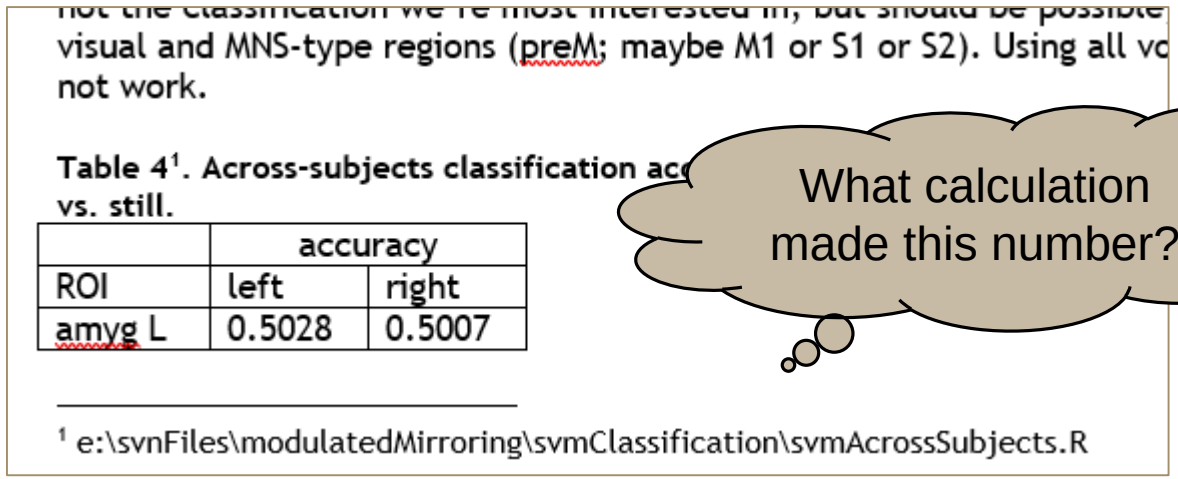

visual and MNS-type regions (preM; maybe M1 or S1 or S2). Using all vo
not work.

Table 4[1]. Across-subjects classification ac
vs. still.

	accuracy	
ROI	left	right
amyg L	0.5028	0.5007

[1] e:\svnFiles\modulatedMirroring\svmClassification\svmAcrossSubjects.R

What calculation made this number?

look at the .rnw code for Table 4.

knitr for neuroimagers

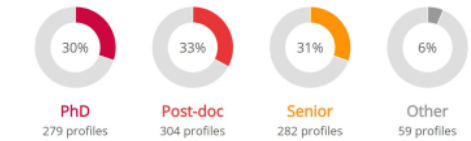
Joset A. Etzel, PhD
jetzel@wustl.edu | mvpa.blogspot.com | @JosetAEtzel
Cognitive Control and Psychopathology Lab
Washington University in St. Louis (USA)

This work was supported by the National Institutes of Health, grant number R37MH066078 to Todd Braver.

Women in Neuroscience

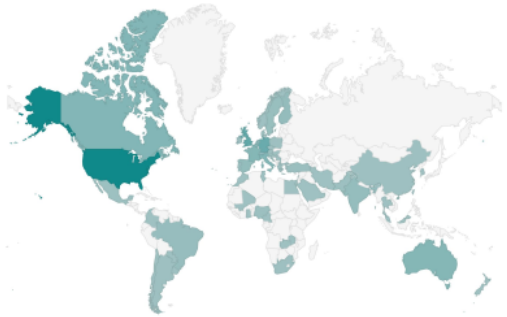
Repository for women in neuroscience

- www.winrepo.org
- over 900 profiles
- easy search
- recommendations

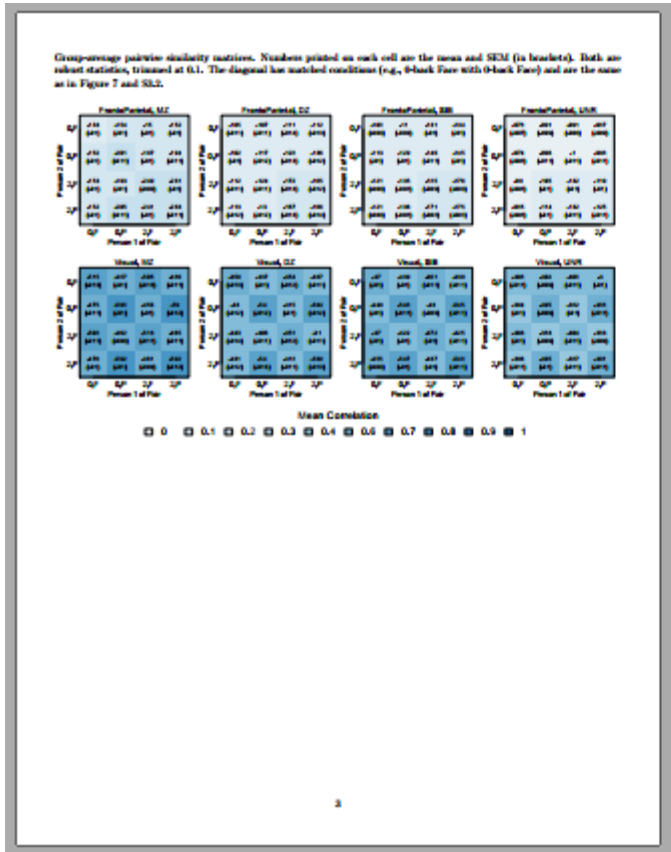


Support the project:

- sign up
- spread the word
- submit recommendations



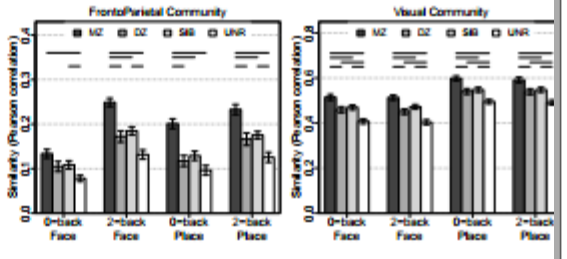
@WINRePo1
www.facebook.com/WiNRepository/



Group-average pairwise similarity matrices. Numbers printed on each cell are the mean and SEM (in brackets). Both are robust statistics, trimmed at 0.1. The diagonal has matched conditions (e.g., 0-back Face with 0-back Face) and are the same as in Figure 7 and S3.2.

S3.2 Pairwise similarity of matched conditions

Mean similarity of each stimulus type separately; error bars are standard error of the mean (SEM). Both are robust statistics, trimmed at 0.1. Horizontal lines indicate bars that significantly (p <.0083, Bonferroni correction of .05 for 6 comparisons) differ in a t-test; see below for t-test t and p values. Note that the y-axis scaling differs between the two plots.



Mean (SEM) of each stimulus type separately, as plotted above and Figure 7. Both are robust statistics, trimmed at 0.

Community	Condition	MZ	DZ	SIB	UNR
FrontoParietal	0-back,Face	.113 (.0096)	.036 (.0112)	.108 (.0092)	.078 (.0090)
FrontoParietal	2-back,Face	.219 (.0092)	.172 (.0120)	.185 (.0092)	.132 (.0093)
FrontoParietal	0-back,Place	.201 (.0109)	.117 (.0122)	.129 (.0101)	.096 (.0098)
FrontoParietal	2-back,Place	.233 (.0113)	.166 (.014)	.175 (.0094)	.125 (.0111)
Visual	0-back,Face	.535 (.0128)	.459 (.0132)	.47 (.0094)	.408 (.0098)
Visual	2-back,Face	.513 (.011)	.451 (.012)	.472 (.0105)	.404 (.0097)
Visual	0-back,Place	.598 (.0102)	.54 (.0123)	.548 (.0108)	.495 (.0094)
Visual	2-back,Place	.592 (.0118)	.539 (.013)	.548 (.0105)	.491 (.0095)


```
sth1 <- subset(mm.th1, mm.th1$community.id == "FrontoParietal" & mm.th1$condition.id == "2,F");
lme.out <- lme(fixed=similarity~pair.group, random="1|pair.id, data=sth1);
#anova(lme.out);
summary(glht(lme.out, mcp(pair.group="Tukey")));

##
##  Simultaneous Tests for General Linear Hypotheses
##
## Multiple Comparisons of Means: Tukey Contrasts
##
##
## Fit: lme.formula(fixed = similarity ~ pair.group, data = sth1, random = "1 |
##     pair.id)
##
## Linear Hypotheses:
##                Estimate Std. Error z value Pr(>|z|)
## MZ - DZ == 0   0.06673    0.01462   4.564  < 0.001 ***
## SIB - DZ == 0  0.01097    0.01481   0.741  0.88033
## UNR - DZ == 0 -0.04000    0.01478  -2.707  0.03455 *
## SIB - MZ == 0 -0.05576    0.01370  -4.069  < 0.001 ***
## UNR - MZ == 0 -0.10673    0.01367  -7.809  < 0.001 ***
## UNR - SIB == 0 -0.05097   0.01387  -3.676  0.00133 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## (Adjusted p values reported -- single-step method)
```

S3.4 Similarity of matched conditions: ACE modeling

For MZ and DZ.

Community	Condition	h2		c2		e2	
FrontoParietal	0-back,Face	0.05	[0,0.1]	0.08	[0.04,0.12]*	0.87	[0.85,0.89]*
FrontoParietal	2-back,Face	0.16	[0.1,0.21]*	0.09	[0.05,0.14]*	0.75	[0.73,0.77]*
FrontoParietal	0-back,Place	0.16	[0.1,0.22]*	0.04	[-0.01,0.09]	0.8	[0.78,0.82]*
FrontoParietal	2-back,Place	0.13	[0.07,0.19]*	0.1	[0.05,0.15]*	0.76	[0.74,0.78]*
Visual	0-back,Face	0.1	[0.05,0.15]*	0.41	[0.36,0.45]*	0.49	[0.48,0.51]*
Visual	2-back,Face	0.13	[0.07,0.18]*	0.38	[0.34,0.43]*	0.49	[0.47,0.51]*
Visual	0-back,Place	0.12	[0.08,0.16]*	0.48	[0.44,0.51]*	0.41	[0.39,0.42]*
Visual	2-back,Place	0.11	[0.06,0.16]*	0.48	[0.44,0.52]*	0.41	[0.4,0.43]*

For MZ and DZ+SIB.

Community	Condition	h2		c2		e2	
FrontoParietal	0-back,Face	0.04	[0,0.09]	0.09	[0.06,0.12]*	0.87	[0.85,0.89]*
FrontoParietal	2-back,Face	0.14	[0.09,0.18]*	0.11	[0.08,0.14]*	0.75	[0.73,0.77]*
FrontoParietal	0-back,Place	0.15	[0.1,0.2]*	0.05	[0.02,0.09]*	0.8	[0.78,0.82]*
FrontoParietal	2-back,Place	0.12	[0.08,0.17]*	0.11	[0.08,0.15]*	0.76	[0.75,0.78]*
Visual	0-back,Face	0.09	[0.05,0.14]*	0.41	[0.38,0.45]*	0.49	[0.48,0.51]*
Visual	2-back,Face	0.11	[0.06,0.15]*	0.4	[0.37,0.44]*	0.49	[0.47,0.51]*
Visual	0-back,Place	0.11	[0.07,0.15]*	0.48	[0.46,0.51]*	0.41	[0.39,0.42]*
Visual	2-back,Place	0.1	[0.06,0.13]*	0.49	[0.47,0.52]*	0.41	[0.4,0.43]*

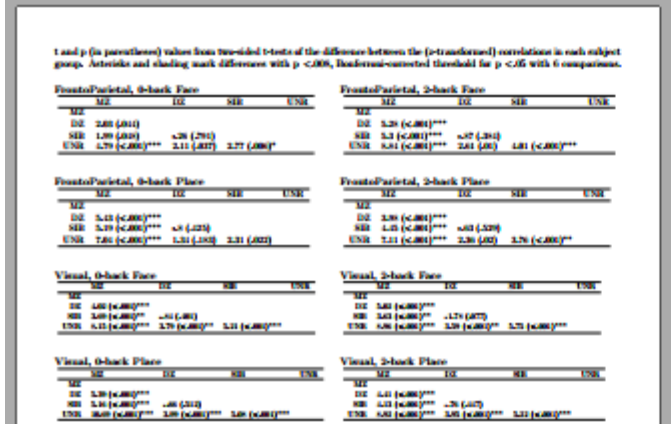
S3.5 Comparison of variance components with other studies

Reference	Mean Age	MZ... similarity (Npairs)	DZ... similarity (Npairs)	UNR similarity (Npairs)	a2	c2	e2	Reference table or figure
Etzel FPN	22-36 Avg 29.2	.13-.25 (105)	.11-.19 (78+99 SIB)	.08-.13 (100)	5-15%**	5-15%	75-87%	Figure 7, S3.2, S3.4
Etzel Visual	22-36 Avg 29.2	.35-.60 (105)	.26-.55 (78+99 SIB)	.46-.55 (100)	9-11%**	40-49%**	41-49%**	Figure 7, S3.2, S3.4
Polk (2007)	18-29	~.73 (13)	~.63 (11)	~.62 (22)	20%*	53%*	27%*	Figure 2 (faces)
Pinel (2015)	Avg 21.7	~.40 (16)	~.25 (13)	~.27#	30%*	10%*	60%*	Figure 6 (faces)
Pinel (2015)	Avg 21.7	.489 (16)	.302 (13)	-	55%** [9-82%]	0%** [0-47%]	45%** [18-100%]	Tables 3, 4 (OFAface, L-L)
Bloklund (2008)	21-27	.39-.42 (29)	~.26-.30 (31)	-	11-36.5%**	0-19.3%	63.5-81.4%	Table 2
Bloklund (2011)	20-30	Fig2b (75)	Fig2b (66)	-	33%** (average across regions)	-	67%	Figure 2b, 3
Bloklund (2017)	16-30	.39-.54 (110)	.05-.33 (138)	-	41%** (average across regions)	-	59%	Supplement Table 1

N.B. MZ and DZ similarity coefficients (typically, correlations) prefaced by a ~ represent approximations from Figures where a precise estimate of the correlation was not provided; similarly, the * in the estimate of additive genetic (a2), common environment (c2) and individual-specific environment (e2) denotes that those estimates were computed for the purpose of this table, based on: e2=1-rMZ; a2=2(rMZ-rDZ); and c2=rMZ-a2, and were not provided in the study either using such equations or via formal model-fitting (latter denoted by **); # study does not specify number of unrelated pairs.

The table above outlines variance components estimates for brain activation during a working memory task across multiple studies of MZ and DZ twins. The current study (Etzel) is among the largest. Based the table above, we see similarities and distinctions across the studies with regards to each variance component:

- Individual-specific environment: The estimate of individual-specific environment (e2) is roughly derived from subtracting the MZ correlation from unity; this estimate is typically estimated with reasonable power even in smaller samples and includes an estimate of measurement error. With the exception of Polk (2007), which includes the fewest MZ pairs and thus may have derived a higher rMZ (the feature selection procedure may also have increased the rMZ), estimates of e2 are >40%, and often >60%, although less so for Visual in the current study. The observation that e2 estimates are the highest for FPN also support our hypothesis of that this network's structure is more idiosyncratic (and so has additional sources of person-specific variance).
- Additive genetics, or heritability: Despite the larger sample size of the current study, estimates of heritability (a2) were lower for both FrontoParietal (FPN) and Visual in the current study, although when compared to Bloklund (2011), heritability of behavioral performance (accuracy and mean reaction time, see Table 2 in Bloklund and SI.6 in current study) were quite comparable especially for the 2-back in Bloklund (2011) suggesting that the lower heritability in our study may be attributed to our analytic approach and our communities of interest.
- Common environment: Importantly, unlike a majority of the other studies, we were able to parse familial effects (i.e., rMZ) into its heritable and common environmental sources, where the latter reflects those environments that are received or perceived equivalently by members of MZ and DZ pairs (and, in our case, non-twin siblings as our analyses did not reveal any evidence for special twin environment). The only other study to hint at common environmental influence is Polk (2007) although we arrive at this estimate via approximation based on the relative magnitude of their MZ and DZ correlations where the latter appears considerably greater than half the former. Interestingly, our choice to contrast the FPN and Visual communities further underscored the role of c2. For instance, while familial effects (i.e., MZ similarity) on Visual were greater than those on FPN, the greater familiality in Visual was primarily attributable to common environment. We might speculate that our estimates deviate from those reported by other studies due to



t and p (in parentheses) values from two-sided t-tests of the difference between the (z-transformed) correlations in each subject group. Asterisks and shading mark differences with p <.008, Bonferroni-corrected threshold for p <.05 with 6 comparisons.

FrontoParietal, 0-back Face

	MZ	DZ	SIB	UNR
MZ				
DZ	2.03 (.044)			
SIB	1.99 (.05)	+.26 (.794)		
UNR	5.79 (<.001)***	3.11 (.017)	2.77 (.006)*	

FrontoParietal, 2-back Face

	MZ	DZ	SIB	UNR
MZ				
DZ	5.26 (<.001)***			
SIB	5.3 (<.001)***	+.07 (.941)		
UNR	6.81 (<.001)***	2.61 (.03)	1.81 (<.001)***	

FrontoParietal, 0-back Place

	MZ	DZ	SIB	UNR
MZ				
DZ	5.13 (<.001)***			
SIB	5.19 (<.001)***	+.8 (.425)		
UNR	7.01 (<.001)***	1.31 (.193)	2.31 (.022)	

FrontoParietal, 2-back Place

	MZ	DZ	SIB	UNR
MZ				
DZ	3.99 (<.001)***			
SIB	4.25 (<.001)***	+.63 (.529)		
UNR	7.11 (<.001)***	2.36 (.02)	3.76 (<.001)**	

Visual, 0-back Face

	MZ	DZ	SIB	UNR
MZ				
DZ	4.66 (<.001)***			
SIB	5.03 (<.001)***	+.82 (.41)		
UNR	8.15 (<.001)***	3.76 (<.001)**	3.21 (<.001)***	

Visual, 2-back Face

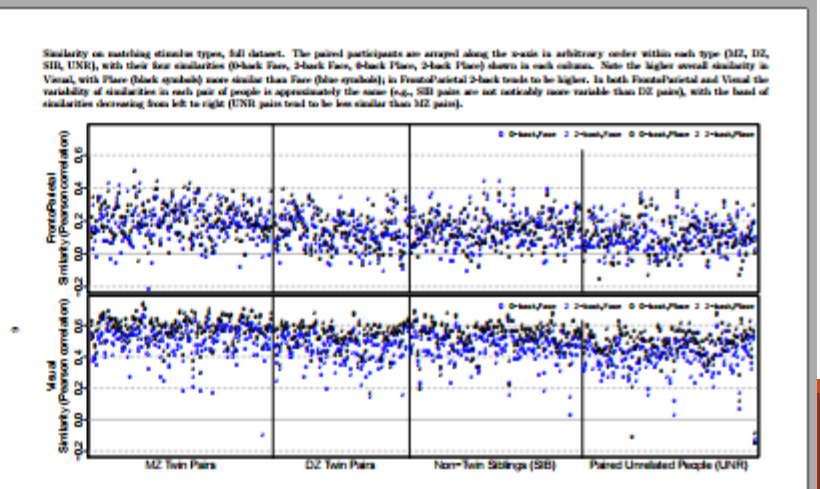
	MZ	DZ	SIB	UNR
MZ				
DZ	5.62 (<.001)***			
SIB	5.63 (<.001)***	+.179 (.857)		
UNR	8.56 (<.001)***	3.59 (<.001)**	3.75 (<.001)***	

Visual, 0-back Place

	MZ	DZ	SIB	UNR
MZ				
DZ	5.39 (<.001)***			
SIB	5.56 (<.001)***	+.46 (.651)		
UNR	10.69 (<.001)***	5.99 (<.001)***	5.69 (<.001)***	

Visual, 2-back Place

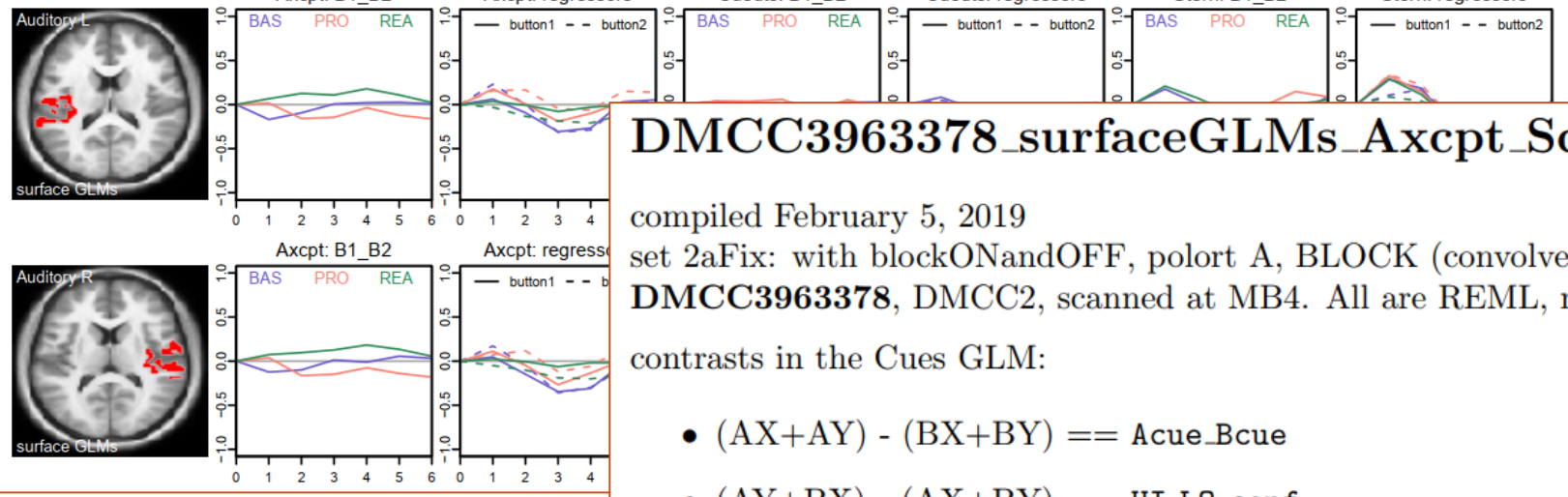
	MZ	DZ	SIB	UNR
MZ				
DZ	6.31 (<.001)***			
SIB	6.53 (<.001)***	+.76 (.457)		
UNR	9.82 (<.001)***	3.95 (<.001)**	3.31 (<.001)***	



Similarity on matching stimulus types, full dataset. The paired participants are arrayed along the x-axis in arbitrary order within each type (MZ, DZ, SIB, UNR), with their four similarities (0-back Face, 2-back Face, 0-back Place, 2-back Place) shown in each column. Note the higher overall similarity in Visual, with Place (black symbols) more similar than Face (blue symbols); in FrontoParietal 2-back tends to be higher. In both FrontoParietal and Visual the variability of similarities in each pair of people is approximately the same (e.g., SIB pairs are not noticably more variable than DZ pairs), with the band of similarities decreasing from left to right (UNR pairs tend to be less similar than MZ pairs).

more example knitrs: syntax coloring, formatted tables, rotated pages for longer graphs, itemized lists of text, captions, …

Gordon communities



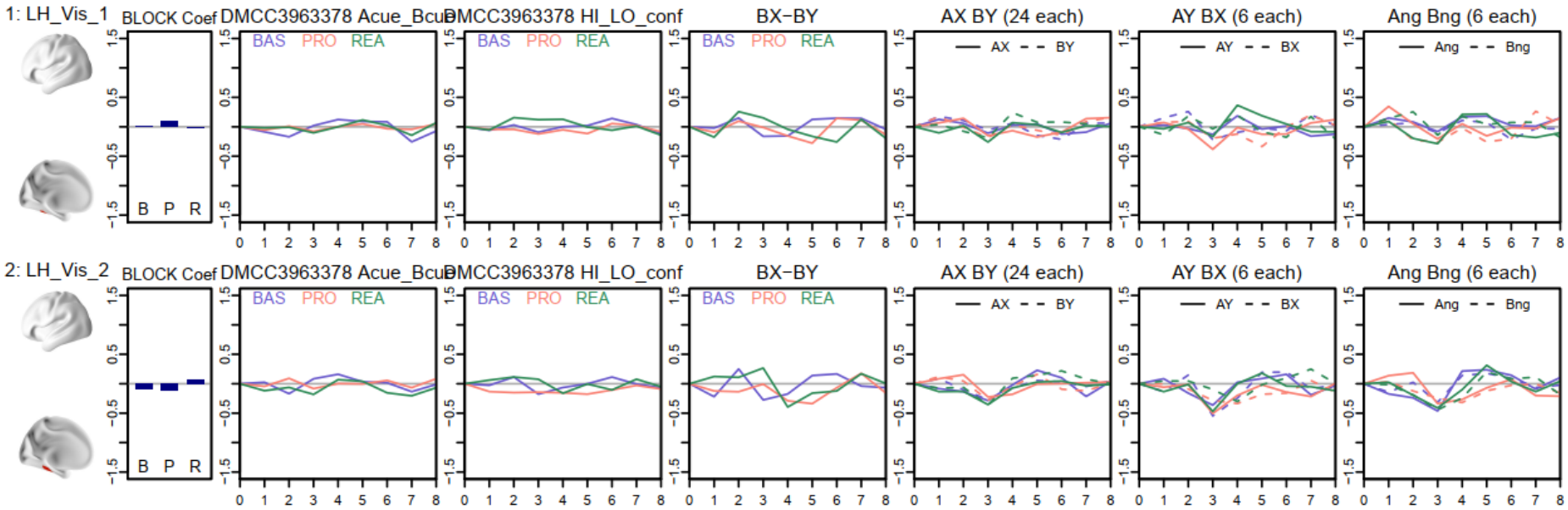
more example knitrs: parcel-average TENT GLM results, with brain images to show parcel locations.

DMCC3963378_surfaceGLMs_Axcpt_Schaefer_censored.rnw

compiled February 5, 2019
set 2aFix: with blockONandOFF, polort A, BLOCK (convolved block); durations fixed.
DMCC3963378, DMCC2, scanned at MB4. All are REML, not ICA-FIX, 2 TRs for each TENT knot for MB4.

contrasts in the Cues GLM:

- (AX+AY) - (BX+BY) == `Acue_Bcue`
- (AY+BX) - (AX+BY) == `HI_LO_conf`
- (Ang+Bng) - (AX+AY+BX+BY) == `Nogo_Go` (calculated, but replaced here with BX - BY)



thresholded statistic images from different afni group GLMs, overlaid on anatomic images.

Key: I didn't list (or copy-paste) every image to be plotted, but just wrote one loop – very easy to update with different contrasts, subjects, or thresholds.

LaTeX

one R loop