

PROBLEM

Preparing to share a dataset that was not well documented during collection and analysis can be a daunting task for a researcher. Here we outline the concept of creating a discipline-specific template that can be followed to easily prepare an existing dataset for sharing and reuse. The template provides a workflow and checklist for researchers to go through their data files and analysis pipeline step by step and document their dataset in a way that makes it appropriate for reuse by others. While best practices for data sharing certainly exist, this template lowers the barrier for the researcher further by providing questions and prompts that are tailored to the specific discipline and by incorporating important methodological concerns that will need to be addressed in documentation.

Template Creation

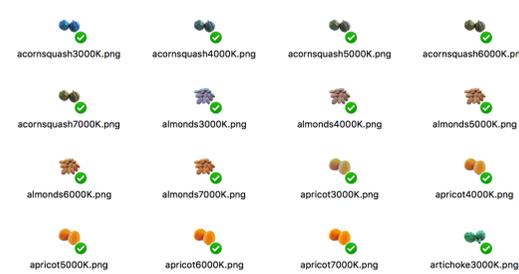
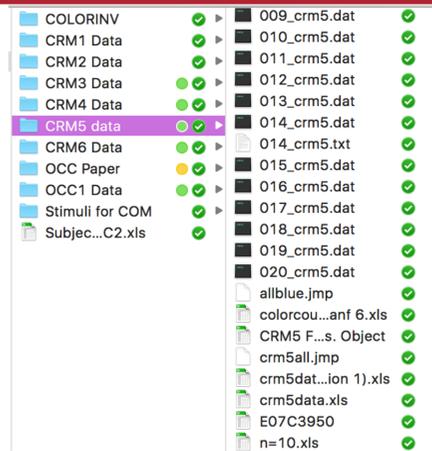
This template was created specifically with behavioral studies in experimental psychology in mind. It highlights both the best practices for data documentation and metadata that are domain-general as well as asking other questions that are tailored to psychology.

For example, psychology studies will need to pay special attention to anonymizing data from human subjects, presenting sample sizes and effect sizes, and detailing the experimental protocol and stimuli used in the study.

CONCLUSIONS

- Providing an easy to use discipline-specific template should make it easier for researchers to prepare existing data for archiving and reuse.
- Hopefully this will encourage more and better sharing of data that is useful to others in the same field and will reduce publications not supported by open data
- This concept could be applied to any discipline from the sciences to the humanities, or could be adapted to focus on particular types of studies, or adapted for specific journals and repositories.

Undocumented

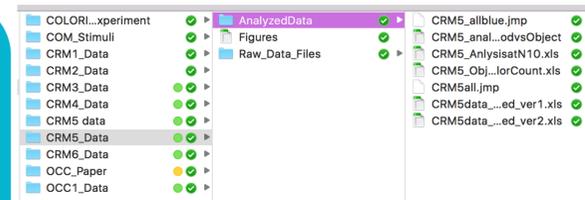


```
runcolorrecmem5.m
By Michael Tarr and Anabeth Van Gulick :: November 2007
%
% Matlab Script to run the color invariance faces experiment (or really any
% recognition memory study, Version 5 = aa (-10), a (-5), b (0), c(5) and
% cc(10) objects changed in LAB space. Study one object, test on that
% object and object 10 away.
%
% RUNNING: Matlab Version 7.4.0.287 (R2007a)
% Psychtoolbox 05 X Release 3.0.8 - Flavor: beta - Corresponds to SWN
% Revision 703
%
% This is a pretty basic recognition memory experiment. There is a list of
% lots of images. We pick 1/2 of them for study. Study objects can appear
% in one of four color temperatures: 3000k, 4000k, 6000k, 7000k with a
% mondrain background of the same temperature and background objects
%
% For test we show the full complement of images and we divide those among
% four conditions with no background (just white):
% Old/New x Same Color Temp/Diff Color Temp (always 5000k - or "normal")
% For New items 50% will be at 5000k and the other 50% will be roughly
% divided into equal numbers at 3000k, 4000k, 6000k, 7000k
%
% USAGE
% runcolorrecmem('001') where this is the subject id
% if no name is passed, subid is set to "test"
% script will not run if datafile for subid exists
```

A	B	C	D	E	F	G	H
1	subject #	trial type	object name	object type	object color	study color	chosen color RT
2	1	1t	whale	L	B	2	1 344
3	1	1t	mailbox	N	B	2	1 394
4	1	1t	handicapped	N	B	2	1 429
5	1	1t	bluebird	L	B	2	3 328
6	1	1t	dolphin	L	B	2	3 345
7	1	1t	bluebells	L	B	2	3 363
8	1	1t	jeans	N	B	2	3 428
9	1	1t	lavender	L	B	2	3 579
10	1	1t	blackberries	L	B	4	3 312
11	1	1t	lapislaz	N	B	4	3 418
12	1	1t	recyclingbin	N	B	4	5 319
13	1	1t	iris	L	B	4	5 324
14	1	1t	blueberries	L	B	4	5 358
15	1	1t	bluejay	L	B	4	5 377
16	1	1t	wave	N	B	4	5 520
17	1	1t	peacock	L	B	4	5 776
18	1	1t	cucumber	L	G	2	1 276
19	1	1t	tank	N	G	2	1 280
20	1	1t	iguana	L	G	2	1 321
21	1	1t	broccoli	L	G	2	1 383
22	1	1t	kiwi	L	G	2	1 396
23	1	1t	avocado	L	G	2	3 254
24	1	1t	honeydew	L	G	2	3 289
25	1	1t	lettuce	L	G	2	3 352
26	1	1t	celery	L	G	2	3 366
27	1	1t	greenbeans	L	G	2	3 430
28	1	1t	wreath	N	G	4	3 250

Example application of the documentation template for a data set from vision science on the categorical nature of object color memory and its relationship to object color diagnosticity (Van Gulick & Tarr, 2010, doi: 10.1167/10.7.407)

Documented



Template Workflow

1. Research Question

- a. What was the research question?
- b. How did this experiment seek to answer it?
- c. Was there an a priori effect size based on previous work?
- d. Is there a publication for this dataset you can link to?

2. Experiment Paradigm

- a. Stimuli – what were they? How were they created? Where can they be found?
- b. Equipment for experiment presentation and data collection
- c. Experiment Procedure (trials, order, timing, response recording). Was the procedure the same for all subjects? Was the trial order randomized? Was the study run in a single session or in several sessions?

3. Subjects

- a. Sample size – why was this selected?
- b. How were subjects recruited? Were they selected for specific traits or experience?
- c. Provide information about subjects - Age, sex, race, education level (if applicable)
- d. Have you de-identified all subject information in your data sets? i.e. subject numbers cannot be traced to names or any other personally identifying information (email, address)

4. Variables and measures

- a. Define all of the variables in the data set
- b. Create a codebook for all variable names, data descriptors, especially any qualitative data values
- c. What unit is reported for each variable?

5. Analysis Pipeline

- a. Where any subjects or data points omitted? If so, why?
- b. Where any variables collapsed? If so, why?
- c. What statistical tests did you perform on the data?
- d. Did you correct for multiple comparisons?
- e. How large are the effects you observed? What is the power of the study?

1. Research Question
 - a. What was the research question?
 - i. What is the nature of color object memory? And if color is encoded in visual object memory, is it exact or categorical? How does the color diagnosticity of an object affect the precision of color memory?
 - b. How did this experiment (set of experiments) seek to answer it?
 - i. Subjects were shown objects in a shifted or neutral color at study and then at test given a two-option forced choice between two different colors of the same object and asked to choose which one they had seen at study (often neither one). A preference for one of the shifted colors over another would indicate the nature of the visual color memory for that object.
 - c. Was there an a priori effect size based on previous work?
 - i. None was used.
2. Experiment Paradigm
 - a. Stimuli – what were they? How were they created? Where can they be found?
 - i. Stimuli were common objects that were a mix of living and non-living things that were predominately a single color (i.e. lemon, basketball, lobster) (food items such as fruits and vegetables were considered living). In some of the experiments in the series compared objects that have a predominant diagnostic color (tennis ball, tomato, mailbox) with those that are man-made objects that could come in any color (bowls, shirts). Shifted colors were created in Photoshop by shifting the colorspace of the image in L*a*b space +10 or -10 on the a-axis (red-green).
 - b. Equipment for experiment presentation and data collection
 - i. The experiment was administered on Apple iMacs using Matlab and Psychtoolbox software. Reaction times were recorded with button presses on an Apple keyboard.
3. Subjects
 - a. Sample size – why was this selected?
 - i. Sample size was based on previous experiments on object recognition.
 - b. How were subjects recruited? Were they selected for specific reasons?
 - i. Subjects were recruited from the Brown University community and received either course credit or payment for their time. Subjects had to have normal or corrected to normal visual acuity and normal color vision, which was verified with Ishihara color plates prior to the experiment.

	A	B	C
1	subno	age	sex
2	4	30	1
3	5	20	2
4	6	25	2
5	7	21	1
6	8	21	1
7	9	28	1
8	10	21	2
9	11	23	1
10	12	27	2
11	13	30	1
12	14	19	1
13	15	35	2
14	16	19	1
15	17	20	1
16	18	22	1
17	19	19	2
18	20	19	2

	A	B	C	D	E	F	G
1	CRM5 Variable	Codebook					
2							
3	Sub	Subject Number					
4	Trial Type	s = study, t=test					
5	Object	Name of the stimulus presented on that trial					
6	ObjType	N=non-living, L=living					
7	ObjectColor	B=Blue, G=Green, R=Red, Y=Yellow					
8	StudyColor	Shifted color at study, 2=Shifted green, 4=Shifted red					
9	ChosenColor	Image chosen at Test, 1=Far Shifted green, 3 = Neutral, 5=Far Shifted Red					
10	RT	Reaction Time in msec					
11							
12							

Analysis of Chosen Color by Trial based on Object Color and Study Color						
Total number of trials of that type, Count for each chosen color, and percent calculated						
Trials with RT Below 250ms or above 500ms excluded						
N = 20	Obj/Color	StudyColor	TotalTrials	ChosenColor Count		Percent
	Blue	2	160	1	76	47.50
	Blue	2		3	84	52.50
	Blue	4	153	3	97	63.40
	Blue	4		5	56	36.60
	Green	2	187	1	88	47.06
	Green	2		3	99	52.94
	Green	4	243	3	164	67.49
	Green	4		5	79	32.51