Data Collection & Entry

Creating a Good Data Gathering Sheet

- How easy is it to read?
- Are column and row definitions clear?
- Is there metadata?
- How similar is it to your digital data entry form?
- Can you use it at 4am?

John Griffin
@JNGriffy

Following

@jebyrnes Start of my PhD, I used the cardboard case of a hummus container to record field data. Lesson learned!

Will Pearse
@willpearse

Following

@JNGriffy @jebyrnes best: waterproof paper; also worst: it fell into the harbour and children threw bacon at me as I retrieved it :-(

The Dimensions of Data

- Number of complete observations

- Number of things measured per observation

Wide Data

Site	Year	Season	NPP_wet	NPP_dry	NPP_carbon	NPP_nitrogen
ABUR	2002	autumn	0.054890914	0.003432687	0.000753216	6.14024E-05
ABUR	2002	summer	0.132411298	0.010272234	0.002535237	0.000251664
ABUR	2003	autumn	0.02226895	0.001607975	0.000405482	4.12825E-05
ABUR	2003	spring	0.01772959	0.001408595	0.000424416	4.03849E-05
ABUR	2003	summer	0.014608974	0.00159993	0.000576714	2.89788E-05
ABUR	2003	winter	0.04025526	0.00385542	0.001127229	7.52017E-05
ABUR	2004	autumn	0.018839953	0.002007113	0.000577988	2.73293E-05
ABUR	2004	spring	0.000331871	2.16794E-05	6.19E-06	5.95E-07
ABUR	2004	summer	0.008816522	0.00098858	0.00028975	1.35229E-05
ABUR	2004	winter	0.012641869	0.000887127	0.000232808	2.16131E-05
ABUR	2005	autumn	0.020872071	0.001859936	0.000514242	2.60997E-05
ABUR	2005	spring	0.009064189	0.000873205	0.000240074	2.42647E-05
ABUR	2005	summer	0.010172798	0.001163216	0.00032975	9.69E-06
ABUR	2005	winter	0.000461258	8.71866E-05	3.32603E-05	1.89E-06
ABUR	2006	autumn	0.002572592	0.000258652	7.87096E-05	3.91E-06
ABUR	2006	spring	0.00713973	0.000669038	0.000194911	1.51552E-05
ABUR	2006	summer	0.003043863	0.00041494	0.000119678	2.51E-06

- **One Observation per row**
- **All data fields are columns**

8/05/2010

Diversity Grazing Lab Trial
(all weights in g)

After Grazing Weights
Trial 2

Tank	Treatment	Macro	Ptery	Chodro	Rhody	Thalam
1	Macro_Poly					
2	Thal_Poly					
3	Chondracanthus					
4	Macro_Poly					
5	Thalamoporella					
6	Even_Poly					
7	Pterygophora					
8	Even_Poly					
9	Macrocystis					
10	Chondracanthus					
11	Ptery_Poly					
12	Chon_Poly					
13	Pterygophora					
14	Rho_Poly					
15	Thalamoporella					
16	Rho_Poly					
17	Rhodymenia					
18	Macrocystis					
19	Ptery_Poly					
20	Thal_Poly					
21	Rhodymenia					
22	Chon_Poly					

Long Data

YEAR	MONTH	DATE	SITE	TRANSECT	SP_CODE	PERCENT_CC	NOTES	TAXON_GEN	TAXON_SPEC	GROUP
2008	7	7/30/08	SCDI	2	LS	0	-99999	Laurencia	spp.	ALGAE
2009	7	7/29/09	SCDI	2	LS	3.75	-99999	Laurencia	spp.	ALGAE
2010	7	7/29/10	SCDI	2	LS	0	-99999	Laurencia	spp.	ALGAE
2011	7	7/26/11	SCDI	2	LS	0	-99999	Laurencia	spp.	ALGAE
2012	7	7/23/12	SCDI	2	LS	0	-99999	Laurencia	spp.	ALGAE
2013	7	7/29/13	SCDI	2	LS	0	-99999	Laurencia	spp.	ALGAE
2014	7	7/21/14	SCDI	2	LS	3.75	-99999	Laurencia	spp.	ALGAE
2004	8	8/30/04	SCDI	3	LS	1.25	-99999	Laurencia	spp.	ALGAE
2005	7	7/27/05	SCDI	3	LS	0	-99999	Laurencia	spp.	ALGAE
2006	7	7/24/06	SCDI	3	LS	0	-99999	Laurencia	spp.	ALGAE
2007	7	7/26/07	SCDI	3	LS	0	-99999	Laurencia	spp.	ALGAE
2008	7	7/30/08	SCDI	3	LS	0	-99999	Laurencia	spp.	ALGAE
2009	7	7/29/09	SCDI	3	LS	0	-99999	Laurencia	spp.	ALGAE
2010	7	7/29/10	SCDI	3	LS	0	-99999	Laurencia	spp.	ALGAE
2011	7	7/26/11	SCDI	3	LS	1.25	-99999	Laurencia	spp.	ALGAE

- **Each measurement 1 row**
- **All information about measurement in row**

Hybrid Data

YEAR	MONTH	DAY	DATE	SITE	TRANSECT	SP_CODE	0-20 IN	20-40 IN	40-20 OFF	20-0 OFF
2013	9	12	9/12/2013	CHANDLER_HOVEY	1	CABO	1	5	0	1
2013	9	12	9/12/2013	CHANDLER_HOVEY	1	CAIR	3	0	2	1
2013	9	12	9/12/2013	CHANDLER_HOVEY	1	HOAM	4	9	8	5
2013	9	12	9/12/2013	CHANDLER_HOVEY	1	ASFO	3	1	1	0
2013	9	12	9/12/2013	CHANDLER_HOVEY	1	HESA	2	1	0	1
2013	9	19	9/19/2013	CHANDLER_HOVEY	2	CABO	1	4	0	0
2013	9	19	9/19/2013	CHANDLER_HOVEY	2	CAMA	6	1	12	0
2013	9	19	9/19/2013	CHANDLER_HOVEY	2	HOAM	2	0	0	0
2013	9	19	9/19/2013	CHANDLER_HOVEY	2	CAIR	0	3	9	0

- **Neither long nor wide**
- **Each row has multiple measurements…but also some unique qualities**
- **Each row is not a complete observation**

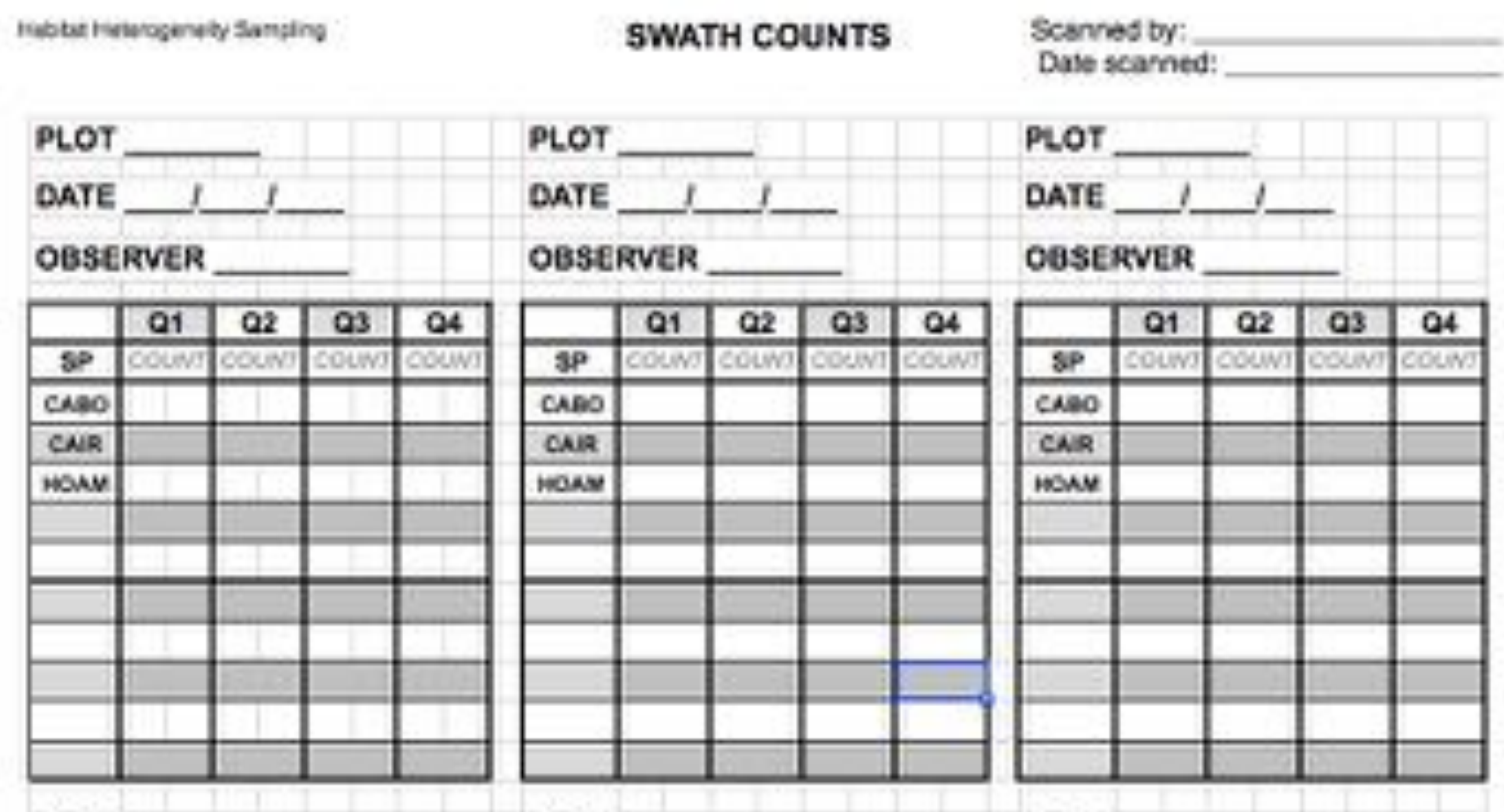
Efficiency May Drive Data Sheet Design

Initial Surveys: KELP/RED/BARE UPCs

Observer _____

	PLOT _____				PLOT _____				PLOT _____			
SP Code	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4
SL												
SLJ												
BLD												
LADI												
AGCL												
BARE												
RED												
EC												
OTHER												
SS												
S												
SH												
C												
BS												
BM												
BL												
B												

Efficiency May Drive Data Sheet Design


Habitat Heterogeneity Sampling

SWATH COUNTS

Scanned by: ____________
Date scanned: ____________

PLOT ________
DATE ___/___/___
OBSERVER ________

	Q1	Q2	Q3	Q4
SP	COUNT	COUNT	COUNT	COUNT
CABO				
CAIR				
HOAM				

PLOT ________
DATE ___/___/___
OBSERVER ________

	Q1	Q2	Q3	Q4
SP	COUNT	COUNT	COUNT	COUNT
CABO				
CAIR				
HOAM				

PLOT ________
DATE ___/___/___
OBSERVER ________

	Q1	Q2	Q3	Q4
SP	COUNT	COUNT	COUNT	COUNT
CABO				
CAIR				
HOAM				

Is this wide, long, other, or bad?

Source	Åsen (2006)								
Site	6.1 Kristiansand: **Dvergsøya** (semi-eksponert)								
Taxon	L. digitata – IL. hyperbore		L. saccharin	L. digitata – IL. hyperbore		L. saccharin	L. digitata – IL. hyperbore		L. saccharin
Year	1982	1982	1982	1988	1988	1988	2006	2006	2006
Month	June	June	June	June	June	June	June	June	June
Day	8	8	8	15	15	15	17	17	17
Date	8.6.	8.6.	8.6.	15.6.	15.6.	15.6.	17.8.	17.8.	17.8.
0	0	0	0	0	0	0	0	0	0
1	0	0	0	0	0	0	0	0	0
2	0	0	3	0	2	0	0	0	0
3	0	0	4	0	2	0	0	4	0
4	0	0	4	0	2	0	0	4	0
5	0	0	4	0	4	0	0	4	0
6	0	0	4	0	4	0	0	4	0
7	0	0	4	0	4	0	0	4	0
8	0	0	3	0	4	0	0	4	0
9	0	2	3	0	4	0	0	4	2
10	0	3	3	0	4	0	0	4	2
11	0	3	3	0	4	0	0	4	2
12	0	3	2	0	4	0	0	4	2
13	0	3	2	0	4	3	0	2	2
14	0	4	2	0	4	3	0	2	2
15	0	3	2	0	4	3	0	2	2
16	0	2	2	0	4	3	0	2	2
17	0	2	2	0	2	2	0	2	2
18	0	2	2	0	2	2	0	2	0
19	0	2	2	0	2	2	0	2	0

What is bad practice here?

Semi quantitative: 1: single/rare, 2: Scattered, 3: common: 4: Dominating. +: occurence, not quantified, ?: occurence not known

a	b	c	d	e	f
6.1 Kristiansand: **Dvergsøya** (semi-eksponert)					
MK 449,411ED50 /MK 44790,40972WGS84	**1976**	**1982-83**	**1988**	**2005**	**2006**
Laminaria digitata – FINGERTARE	?	0	0	2	0
Laminaria hyperborea – STORTARE	+	4	4	4	4
Laminaria saccharina – SUKKERTARE	?	4	3	2	2
6.1 Kristiansand: **Bertesbukta** (semi-eksponert)					
MK 432,454ED50 /MK 42985,45338WGS84	**1976**	**1982-83**		**2005**	**2006**
Laminaria digitata – FINGERTARE	?	3		2	0
Laminaria hyperborea - STORTARE	?	4		3	3
Laminaria saccharina - SUKKERTARE	4	4		1	2
8 Sogne: **Høgsholmen** (semi-eksponert)					
MK 29696,36644WGS84		**1983**	**1989**		**2006**
Laminaria digitata – FINGERTARE		2	2		0
Laminaria hyperborea – STORTARE		2	2		1
Laminaria saccharina – SUKKERTARE		4	4		1
8 Sogne: **Holme** NV Skarpøya (semi-eksponert)					
MK 229749,35623WGS84		**1983**	**1989**		**2006**
Laminaria digitata – FINGERTARE		0	3		0
Laminaria hyperborea – STORTARE		3	3		1
Laminaria saccharina – SUKKERTARE		4	4		1
5.5-6 Sogne: **Lastad** (semi-eksponert)					

What is wrong here?

14/09/96		*Ecklonia*		DONKIN BAY			
MID SITE QUADRAT	DEPTH (m)	STIPE LENGTH (cm)	STIPE WEIGHT (g)	FROND WEIGHT (g)	TOTAL KELP WEIGHT (g)	WEIGHT TOTAL PER SQ metre (g)	PER SQ metre MEAN STIPE LENGTH (cm)
	2.5	200	1300	0	1300		
		190	800	1500	2300		
		195	1000	900	1900	5500	195.00
4	2.5	100	225	525	750	750	100.00
5	2.5	15	15	250	265		
		25	25	300	325		
		15	20	900	920	1510	18.33
6	2.5	22	25	150	175		
		10	10	70	80	255	16.00
7	2.5	115	400	1850	2250		
		165	475	750	1225		
		150	525	1200	1725		
		145	300	700	1000		
		142	450	1000	1450		
		100	225	180	405		
		122	325	550	875		
		32	50	180	230		
		7	10	50	60		
		25	50	275	325		
		16	30	200	230		
		26	25	100	125	9900	87.08
8	2.5	195	900	1600	2500	2500	195.00

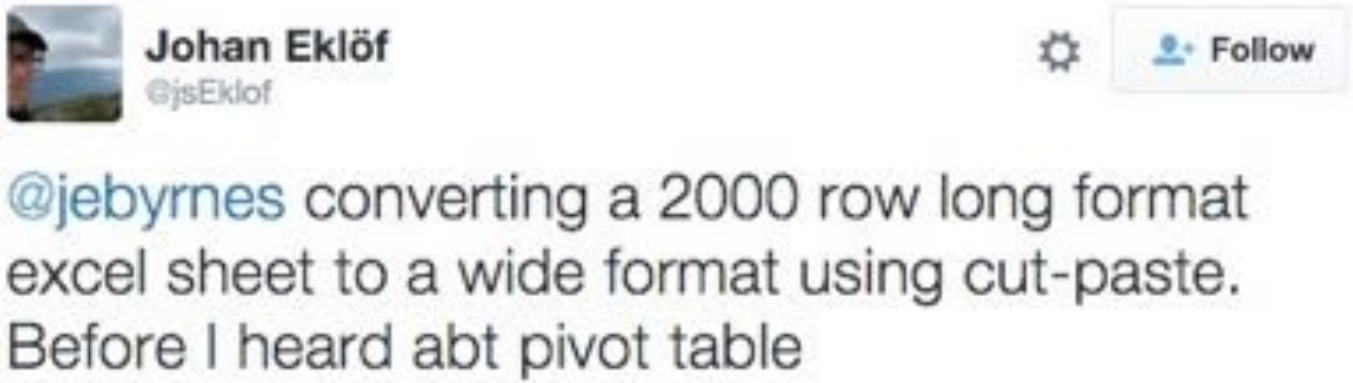
What is bad practice here?

sea duck							
Percent Cover	Algal Species						
Hild	Lith	Clath	Phym	C.off	S. derm	L. sacharina	L. digitata
0	50	0	0	2	0	0	0
0	20	0	0	0	5	0	0
10	5	0	10	0	10	15	0
15	10	0	15	0	5	0	0
5	15	0	0	2	0	0	0
10	20	0	0	0	2	50	0
0	2	0	10	0	2	0	0
5	0	15	20	0	0	0	0
10	10	20	30	0	5	0	0
15	5	20	25	0	0	12	0
15	0	10	5	0	0	0	0
15	0	10	5	0	0	0	0
15	0	5	5	0	0	0	0
20	15	15	25	0	0	0	0
40	0	5	40	0	0	0	0
30	25	10	30	0	0	0	0
15	10	5	25	0	0	0	0
10	10	5	15	0	0	2	0
5	0	0	5	0	0	0	0
10	10	0	35	0	0	0	0
ram island							
Percent Cover	Algal Species						
Hild	Lith	Clath	Phym	C.off	S. derm	L. sacharina	L. digitata
20	0	60	18	0	0	0	0
4	4	70	20	0	0	0	0
8	4	70	15	0	0	0	0

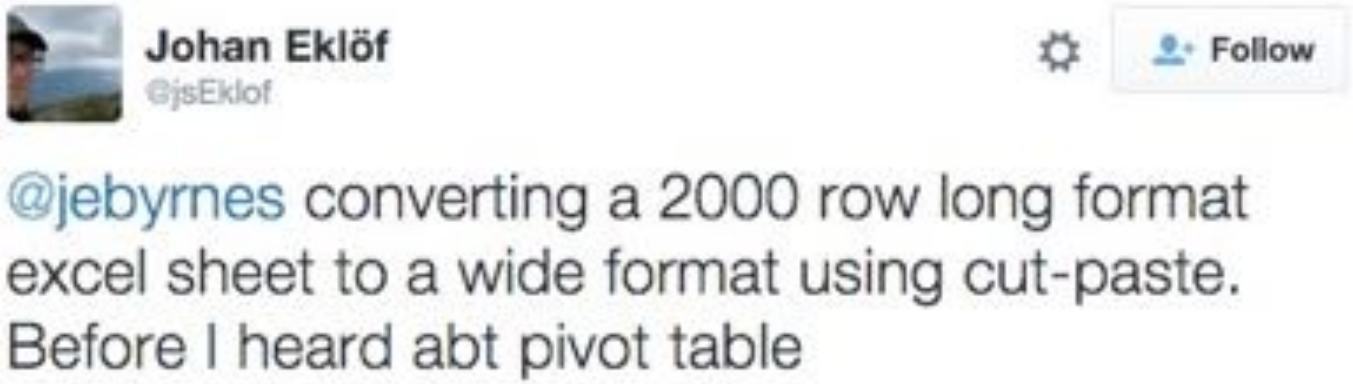
General Rules for Good Data Creation

- Columns should have only 1 type of data
- Keep metadata separate or repeated
- No bare space
- Consistent NA character
- Others?

EXCEL TIME!


Johan Eklöf
@jsEklof
Follow
@jebyrnes converting a 2000 row long format excel sheet to a wide format using cut-paste. Before I heard abt pivot table

EXCEL TIME!



Entry
Fills
Basic Functions
Functions for Error Checking
Pivot tables
Controlled Vocabularies