

Raw to tidy data



Getting and Cleaning Data

Common Problems with Messy Data Sets

1. Column headers are values but should be variable names.
2. A single column has multiple variables.
3. Variables have been entered in both rows and columns.
4. Multiple "types" of data are in the same spreadsheet.
5. A single observation is stored across multiple spreadsheets.

A

	A	B	C	D	E	F
1						
2		101	102	103	104	105
3	sex	Male	Female	Male	Male	Male
4						
5		101	102	103	104	105
6	glucose	134.1	120.0	124.8	83.1	105.2
7						
8		101	102	103	104	105
9	insulin	0.60	1.18	1.23	1.16	0.73

C

	A	B	C	D	E	F	G
1							
2	Date	11/3/14					
3	Days on diet	126					
4	Mouse #	43					
5	sex	f					
6	experiment		values			mean	SD
7	control		0.186	0.191	1.081	0.49	0.52
8	treatment A		7.414	1.468	2.254	3.71	3.23
9	treatment B		9.811	9.259	11.296	10.12	1.05
10							
11	fold change		values			mean	SD
12	treatment A		15.26	3.02	4.64	7.64	6.65
13	treatment B		20.19	19.05	23.24	20.83	2.17

B

	A	B	C	D	E	F	G
1	1MIN						
2			Normal			Mutant	
3	B6	146.6	138.6	155.6	166	179.3	186.9
4	BTBR	245.7	240	243.1	177.8	171.6	188.1
5							
6	5MIN						
7			Normal			Mutant	
8	B6	333.6	353.6	408.8	450.6	474.4	423.8
9	BTBR	514.4	610.6	597.9	412.1	447.4	446.5

D

	A	B	C	D	E	F
1		GTT date	GTT weight	time	glucose mg/dl	insulin ng/ml
2	321	2/9/15	24.5	0	99.2	lo off curve
3				5	349.3	0.205
4				15	286.1	0.129
5				30	312	0.175
6				60	99.9	0.122
7				120	217.9	lo off curve
8	322	2/9/15	18.9	0	185.8	0.251
9				5	297.4	2.228
10				15	439	2.078
11				30	362.3	0.775
12				60	232.7	0.5
13				120	260.7	0.523
14	323	2/9/15	24.7	0	198.5	0.151
15				5	530.6	off curve lo

Australian Bureau of Statistics											
1800.0 Australian Marriage Law Postal Survey, 2017											
Released on 15 November 2017											
Table 5 Participation by Federal Electoral Division(a), Males and Age											
Year NA											
		18-19 years	20-24 years	25-29 years	30-34 years	35-39 years	40-44 years	45-49 years	50-54 years	55-59 years	60-64 years
Lingard(c)	Total participants	292	1,058	1,465	1,653	1,515	1,516	1,710	1,730	1,753	1,574
	Eligible participants	572	2,910	3,789	3,996	3,607	3,506	3,645	3,331	2,960	2,456
	Participation rate (%)	51.0	36.4	38.7	41.4	42.0	43.2	46.9	51.9	59.2	64.1
Solomon	Total participants	442	1,461	2,066	2,357	2,188	2,057	2,224	2,108	2,134	1,772
	Eligible participants	750	2,991	3,994	4,155	3,634	3,398	3,427	3,066	2,931	2,355
	Participation rate (%)	58.9	48.8	51.7	56.7	60.2	60.5	64.9	68.8	72.8	75.2
Northern Territory (Total)	Total participants	734	2,519	3,531	4,010	3,703	3,573	3,934	3,838	3,887	3,346
	Eligible participants	1,322	5,901	7,783	8,151	7,241	6,904	7,072	6,397	5,891	4,811
	Participation rate (%)	55.5	42.7	45.4	49.2	51.1	51.8	55.6	60.0	66.0	69.5
Australian Capital Territory Divisions											
Canberra(d)	Total participants	1,764	4,789	4,817	4,973	4,626	4,453	5,074	4,826	5,169	4,394
	Eligible participants	2,260	6,471	6,448	6,509	5,983	5,805	6,302	5,902	6,044	5,057
	Participation rate (%)	78.1	74.0	74.7	76.4	77.3	76.7	80.5	81.8	85.5	86.9
Fenner(e)	Total participants	1,477	4,687	5,178	5,786	6,025	5,463	5,191	4,208	3,948	3,465
	Eligible participants	1,904	6,354	7,121	7,822	7,960	7,155	6,480	5,206	4,692	3,945
	Participation rate (%)	77.6	73.8	72.7	74.0	75.7	76.4	80.1	80.8	84.1	87.8
NA Year											
Australian Capital Territory (Total)	Total participants	3,241	9,476	9,395	10,739	10,651	9,916	10,265	9,634	9,117	7,639
	Eligible participants	4,164	12,825	13,569	14,331	13,943	12,960	12,782	11,108	10,736	9,002
	Participation rate (%)	77.8	73.9	73.7	75.1	76.4	76.5	80.3	81.3	84.9	87.3
Australia											
Total	Total participants	151,297	438,166	441,658	460,548	462,206	479,360	524,620	517,693	543,449	506,799
	Eligible participants	201,439	635,909	646,916	665,250	656,446	660,841	693,850	659,150	664,720	597,386
	Participation rate (%)	75.1	68.9	68.3	69.2	70.4	72.5	75.6	78.5	81.8	84.8
(a) The Federal Electoral Divisions are current as at 24 August 2017											
(b) Includes those whose age is unknown											
(c) Includes Christmas Island and the Cocos (Keeling) Islands											
(d) Includes Norfolk Island											
(e) Includes Jervis Bay											

Clinic Name	Clinic Location	Address	Contact Number	Drop-in	Operational Hours	By Appointment	Services
Birth Control and Sexual Health Centre	Dufferin St./Lawrence Ave. W.	Suite 403, 960 Lawrence Ave. W., Toronto, On M6A 3B5	416-789-4541			Monday: 2 pm - 5 pm Tuesday: 4 pm - 7 pm Wednesday: 12 noon - 5 pm Thursday: 5 pm - 8 pm Friday: 12:30 pm - 4:30 pm	Birth control counselling Low cost or free birth control Free condoms Plan B (emergency contraceptive pill) STI testing and free treatment HIV testing Pregnancy testing, counselling and referral Sexuality and relationship counselling Anonymous HIV testing (including the rapid HIV test)
Names/addresses across multiple lines			Merged column heading! Column headings on the second row!		Attributes spanning multiple rows, without the ID (clinic name) also spanning! Not tidy 🙄		
Black Creek Community Health Centre (Sheridan Mall Site)	Jane St./Wilson Ave.	North York Sheridan Mall, Unit 5 2202 Jane St., Toronto, On M2M 1A4	416-249-8000	Tuesday: 4 pm - 6 pm			Birth control counselling Low cost or free birth control Free condoms Plan B (emergency contraceptive pill) STI testing and free treatment HIV testing Pregnancy testing, counselling and referral Sexuality and relationship counselling
Black Creek Community Health Centre (Yorkgate mall Site)	Jane St./Finch Ave. W.	1 York Gate Blvd., Toronto, On M3N 3A1	416-246-2388	Wednesday: 1 pm - 4:30 pm			Birth control counselling Low cost or free birth control Free condoms Plan B (emergency contraceptive pill) STI testing and free treatment HIV testing Pregnancy testing, counselling and referral Sexuality and relationship counselling
Special Treatment Clinic	Bay St./College St.	8th Floor, 790 Bay St., Toronto, On M5G 1N8	416-351-3800 ext 2207	Tuesday: 5 pm - 7 pm Wednesday: 5 pm - 7 pm Thursday: 5 pm - 7 pm			Birth control counselling Low cost or free birth control Free condoms Plan B (emergency contraceptive pill) STI testing and free treatment HIV testing Pregnancy testing, counselling and referral Sexuality and relationship counselling
Taibu Community Health Center	Neilson Rd./McLevin Ave.	27 Tapscott Rd., Toronto, On M1B 4Y7	416-644-3536			Tuesday: 2:30 pm - 6:30 pm	Birth control counselling Low cost or free birth control Free condoms Plan B (emergency contraceptive pill) STI testing and free treatment HIV testing Pregnancy testing, counselling and referral Sexuality and relationship counselling
The Talk Shop	Yonge St./Empress Ave.	5110 Yonge St., Toronto, On M2N 6M1	416-338-7000	Monday: 2 pm - 6:30 pm Wednesday: 9:30 am - 11:30 am, 1 pm - 3:30 pm Thursday: 2 pm - 6:30 pm			Birth control counselling Low cost or free birth control Free condoms Plan B (emergency contraceptive pill) STI testing and free treatment HIV testing Pregnancy testing, counselling and referral Sexuality and relationship counselling Rapid HIV testing
NA NA NA NA			NA NA NA NA		NA NA NA NA		
					NA NA NA NA		

D

	A	B	C	D	E	F
1		GTT date	GTT weight	time	glucose mg/dl	insulin ng/ml
2	321	2/9/15	24.5	0	99.2	lo off curve
3				5	349.3	0.205
4				15	286.1	0.129
5				30	312	0.175
6				60	99.9	0.122
7				120	217.9	lo off curve
8	322	2/9/15	18.9	0	185.8	0.251
9				5	297.4	2.228
10				15	439	2.078
11				30	362.3	0.775
12				60	232.7	0.5
13				120	260.7	0.523
14	323	2/9/15	24.7	0	198.5	0.151
15				5	530.6	off curve lo

A

	A	B	C
1	id	GTT date	GTT weight
2	321	2/9/15	24.5
3	322	2/9/15	18.9
4	323	2/9/15	24.7

B

	A	B	C	D	E
1	id	GTT time	glucose mg/dl	insulin ng/ml	note
2	321	0	99.2	NA	insulin below curve
3	321	5	349.3	0.205	
4	321	15	286.1	0.129	
5	321	30	312	0.175	
6	321	60	99.9	0.122	
7	321	120	217.9	NA	insulin below curve
8	322	0	185.8	0.251	
9	322	5	297.4	2.228	
10	322	15	439	2.078	
11	322	30	362.3	0.775	
12	322	60	232.7	0.5	
13	322	120	260.7	0.523	
14	323	0	198.5	0.151	
15	323	5	530.6	NA	insulin below curve

Summarizing Raw to tidy data



Getting and Cleaning Data