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.

Α

| | A | В | С | 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 |

С

| | A | В | С | 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 |

В

| | Α | В | С | 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 | В6 | 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

| | Α | В | С | D | E | F |
|----|-----|----------|------------|------|---------------|--------------|
| 1 | | GTT date | GTT weight | time | glucose mg/dl | insulin ng/m |
| 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 |

| Total participants 292 1,058 1 465 1,653 1,515 1,516 1,710 1,730 1,77 | 2,4 | | | | | | | | | | | |
|--|----------|-------------|--|-------------|-------------|------------------|---------------------|-------------|--|------------------|------------------------|---|
| Table 5 Participation by Federal Electoral Division(a), Males and Age Gender apartheid | 1,5 | | | | | | | 7 | rvey, 201 | w Postal S | an Marriage La | 1800.0 Australi |
| Total participants | 1,5 | | | | | | | | | | ember 2017 | Released on 15 Nov |
| Total participants 18.19 years 29.24 years 25.29 years 30.34 years 30.34 years 30.34 years 50.49 years 50.49 years 50.59 y | 1, 2, | | | | | theid | nder apar | d Age Ger | (a), <mark>Males a</mark> | toral Division | on by Federal Elec | Table 5 Participation |
| Lingia C Eligible participants 572 2,910 3,789 3,996 3,607 3,506 3,645 3,331 2,9 | 2, | 55-59 years | 50-54 years | 45-49 years | 40-44 years | 35-39 years | 30-34 years | 25-29 years | 20-24 years | 18-19 years | NA | Yeah |
| Primary keyno Padicipation rate (%) 51.0 36.4 38.7 41.4 42.0 43.2 46.9 51.9 55.5 | | 1,753 | 1,730 | 1,710 | 1,516 | 1,515 | 1,653 | 1.465 | 1,058 | 292 | Total participants | |
| Merged cells Solomon | 6 | 2,960 | 3,331 | 3,645 | 3,506 | 3,607 | 3,996 | 3,789 | 2,910 | 572 | Eligible participants | Lingiar(c) |
| Merged cells Solomon | + ' | 59.2 | 51.9 | 46.9 | 43.2 | 42.0 | 41.4 | | | 51.0 | Participation rate (%) | rimary keyno |
| Solomon Bigible participants 750 2,991 3,994 4,155 3,634 3,398 3,427 3,066 2,994 4,155 3,634 3,398 3,427 3,066 2,994 4,155 3,634 3,398 3,427 3,066 2,994 3,994 4,155 3,634 3,398 3,427 3,066 2,994 3,994 4,155 3,634 3,398 3,427 3,066 2,994 3,994 3,994 4,155 3,634 3,398 3,427 3,066 2,994 3,994 | 4 | 2 124 | 2.100 | 2 224 | 2.057 | 2 100 | 2 257 | | | 440 | 1 | |
| Total participants 734 2,519 3,531 4,010 3,703 3,573 3,934 3,838 3,8 | | | | | -1 | | | | | | | - |
| Northern Territory (Total) Total participants 1,322 5,901 7,783 8,151 7,241 6,904 7,072 6,397 5,8 7,783 8,151 7,241 6,904 7,072 6,397 5,8 7,783 8,151 7,241 6,904 7,072 6,397 5,8 7,783 8,151 7,241 6,904 7,072 6,397 5,8 7,783 8,151 7,241 6,904 7,072 6,397 5,8 7,80 7,072 6,397 5,8 7,000 | | 72.8 | The second secon | | | | | | | | | SOUTION |
| Covariate as Subheading Covariate as Sub | | 12.0 | 00.0 | 64.9 | 60.5 | 60.2 | 30.7 | 51.7 | 40.0 | 50.9 | articipation rate (%) | |
| Australian Capital Fernitory (Total) Eligible participants 1,322 5,901 1,765 42.7 45.4 49.2 51.1 51.8 55.6 60.0 66 | 3, | 3,887 | 3,838 | 3,934 | 3,573 | 3,703 | 4,010 | 3,531 | 2,519 | 734 | Total participants | |
| Participation rate (%) 55.5 42.7 45.4 49.2 51.1 51.8 55.6 60.0 66 | 4, | 5,891 | 6,397 | 7,072 | 6,904 | 7,241 | 8,151 | 7,783 | 5,901 | 1,322 | Eligible participants | |
| Total participants 1,764 4,789 4,817 4,973 4,626 4,453 5,074 4,826 5,18 | 6 | 66.0 | 60.0 | 55.6 | 51.8 | 51.1 | 49.2 | 45.4 | 42.7 | 55.5 | Participation rate (%) | rottay |
| Canberra(d) Eligible participants 2,260 6,471 6,448 6,509 5,983 5,805 6,302 5,902 6,0 Participation rate (%) 78.1 74.0 74.7 76.4 77.3 76.7 80.5 81.8 85 Fenner(e) Eligible participants 1,477 4,687 5,178 5,786 6,025 5,463 5,191 4,208 3,9 Fenner(e) Eligible participants 1,904 6,354 7,121 7,822 7,960 7,155 6,480 5,206 4,6 Participation rate (%) 77.6 73.8 72.7 74.0 75.7 76.4 80.1 80.8 84 NA Yeah | | | | | | 7.47745.00 | | | | | | Territory Divisions |
| Participation rate (%) 78.1 74.0 74.7 76.4 77.3 76.7 80.5 81.8 85 | | 5,169 | | | | | | | - Contraction of the Contraction | | | |
| Total participants 1,477 4,687 5,178 5,786 6,025 5,463 5,191 4,208 3,9 | | 6,044 | | | | - Annual Control | | | | | | |
| Fenner(e) Eligible participants 1,904 6,354 7,121 7,822 7,960 7,155 6,480 5,206 4,66 Participants 77.6 73.8 72.7 74.0 75.7 76.4 80.1 80.8 84 Section 80.1 80.8 Section 80.1 80.8 Section 80.8 84 Section 80.1 80.8 Section 80.1 80.8 Section 80.1 80.8 Section 80.1 80.8 Section 80 | 8 | 85.5 | 81.8 | 80.5 | 76.7 | 77.3 | 76.4 | 74.7 | 74.0 | 78.1 | Participation rate (%) | |
| Fenner(e) Eligible participants 1,904 6,354 7,121 7,822 7,960 7,155 6,480 5,206 4,66 Participants 77.6 73.8 72.7 74.0 75.7 76.4 80.1 80.8 84 Section 10,203 10,755 10,001 10,203 10,755 Australian Capital Eligible participants 4,164 12,825 13,569 14,331 13,943 12,960 12,782 11,108 10,77 Participation rate (%) 77.8 73.9 73.7 75.1 76.4 76.5 80.3 81.3 84 Australia Total participants 151,297 438,166 441,658 460,548 462,206 479,360 524,620 517,693 543,44 Fenner(e) Eligible participants 1,904 6,354 7,121 7,822 7,960 7,155 6,480 5,206 4,66 Rotal participants 1,904 6,354 7,121 7,822 7,960 7,155 6,480 5,206 4,66 Rotal participants 1,904 1,904 1,905 1,905 1,905 1,905 Rotal participants 1,904 1,904 1,905 1,905 1,905 1,905 Rotal participants 1,904 1,905 1,905 1,905 1,905 Rotal participants 1,904 1,905 1,905 1,905 1,905 Rotal participants 1,904 1,905 1,905 1,905 Rotal participants 1,904 1,905 1,905 1,905 Rotal participants 1,904 1,905 Rotal participants 1,905 1,905 Rotal participants 1 | 3. | 3,948 | 4.208 | 5.191 | 5.463 | 6.025 | 5.786 | 5.178 | 4.687 | 1.477 | Total participants | |
| Participation rate (%) 77.6 73.8 72.7 74.0 75.7 76.4 80.1 80.8 84 | 3,9 | 4,692 | 5,206 | 6,480 | 7.155 | 7,960 | 7.822 | 7.121 | 6.354 | 1,904 | | |
| Australian Capital Territory (Total) Total participants 151,297 438,166 441,658 460,548 462,206 479,360 524,620 517,693 543,44 5,14 | | 84.1 | | | | | | 72.7 | | | | |
| Australia Capital Eligible participants 4,164 12,825 13,569 14,331 13,943 12,960 12,782 11,108 10,77 12,782 11,108 10,77 12,782 11,108 10,77 12,782 11,108 10,77 12,782 11,108 10,77 12,782 | | | | | | | | | | | | |
| Territory (Total) | 7,0 | 9,117 | | | | | | | | | | Australian Canital |
| Australia Total participants 151,297 438,166 441,658 460,548 462,206 479,360 524,620 517,693 543,4 | | 10,736 | | | | | | | | | | Territory (Total) |
| Total participants 151,297 438,166 441,658 460,548 462,206 479,360 524,620 517,693 543,4 | 8 | 84.9 | 81.3 | 80.3 | 76.5 | 76.4 | 75.1 | 73.7 | 73.9 | 77.8 | Participation rate (%) | 2 2 2 |
| The property of the property o | | | | | | | | | | | | Australia |
| Total Eligible participants 201,439 635,909 646,916 665,250 656,446 660,841 693,850 659,150 664,7 | 506, | 543,449 | 517,693 | 524,620 | 479,360 | 462,206 | 460,548 | 441,658 | 438,166 | 151,297 | Total participants | |
| | 597,3 | 664,720 | 659,150 | 693,850 | 660,841 | 656,446 | 665,250 | 646,916 | 635,909 | 201,439 | Eligible participants | |
| Participation rate (%) 75.1 68.9 68.3 69.2 70.4 72.5 75.6 78.5 81 | 8 | 81.8 | 78.5 | 75.6 | 72.5 | 70.4 | 69.2 | 68.3 | 68.9 | 75.1 | Participation rate (%) | |
| | | | | | | | | | 047 | | I Di I I | |
| a) The Federal Electoral Divisions are current as at 24 August 2017 | | | | | | n le | a desired a desired | | | as at 24 August | | |
| (b) Includes those whose age is unknown Return of the table junk | | | | | | nk | e table ju | urn of the | Ret | Marker Viete | | |
| (c) Includes Christmas Island and the Cocos (Keeling) Islands | - | | | | | | | | | Reeling) Islands | | AND RESIDENCE OF THE PROPERTY OF THE PERSON NAMED IN COLUMN TWO IS NOT THE PERSON NAMED IN COLUMN TWO IS |
| (d) Includes Norfolk Island (e) Includes Jervis Bay | _ | | | | | | | | | | | |

| Clinic Name | Clinic Location | Address | Contact Number | Operational H | | Services |
|---------------------------------------|-------------------------------|--|--------------------|---|----------------------------------|--|
| | | | | Drop-in | By Appointment | |
| irth Control and Sexual Health Centre | Dufferin St./Lawrence Ave. W. | Suite 403, 960 Lawrence Ave. W., Toronto, On M6A 3B5 | | | Monday: 2 pm - 5 pm | Birth control counselling |
| | | | Merge | d column heading! | Tuesday: 4 pm - 7 pm | Low cost or free birth control |
| | | _ | | | Wednesday: 12 noon - 5 pm | Free condoms |
| | | C | olumn hea | dings on the second row! | Thursday: 5 pm - 8 pm | Plan B (emergency contraceptive pill) |
| | | | | 3- | Friday: 12:30 pm - 4:30 pm | STI testing and free treatment |
| | | | | | | HIV testing |
| | | | | Δ | ttributes spanning | Pregnancy testing, counselling and referral |
| Name of Jordanne | | ll | | ^ | tiributes sparifiling | Sexuality and relationship counselling |
| Names/address | es across multiple | # IInes | | mı | Iltiple rows, without | Anonymous HIV testing (including the rapid HIV test) |
| | | | | | | Perongrisous Fire teating (monuting the rapid Fire teat) |
| lack Creek Community Health Centre | Jane St./Wilson Ave. | North York Sheridan Mall, Unit 5 | 416-249-8000 | Tuesday: 4 pm - 6 pm th | e. LD _v (clinic name) | Birth control counselling |
| Sheridan Mall Site) | June Gerrmaon rive. | 2202 Jane St., Toronto, On M2M 1A4 | 410 240 0000 | | | Low cost or free birth control |
| Sheridan Mali Site) | | 2202 Jane St., Toronto, On Walkin 1A4 | | a | lso spanning! Not | Free condoms |
| | | | | | | |
| | | | | | tidy 🖀 | Plan B (emergency contraceptive pill) |
| | | | | | | STI testing and free treatment |
| | | | | | | HIV testing |
| | | | | | | Pregnancy testing, counselling and referral |
| | | | | | | Sexuality and relationship counselling |
| | | | | | | |
| lack Creek Community Health Centre | 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 |
| Yorkgate mall Site) | | | | | | Low cost or free birth control |
| | | | | | | Free condoms |
| | | | | | | Plan B (emergency contraceptive pill) |
| | | | | | | STI testing and free treatment |
| | | | | | | HIV testing and free treatment |
| | | | | | | |
| | | | | | | 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 2 | 207 | | Birth control counselling |
| | | | | | | Low cost or free birth control |
| | | | | Tuesday: 5 pm - 7 pm | | Free condoms |
| | | | | Wednesday: 5 pm - 7 pm NA NA | | Plan B (emergency contraceptive pill) |
| | | | | Thursday: 5 pm - 7 pm | | 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 | | | Birth control counselling |
| | | | | | | Low cost or free birth control |
| | | | | | | Free condoms |
| | | | | | | Plan B (emergency contraceptive pill) |
| | | | | | | STI testing and free treatment |
| | | | | | | HIV testing |
| | | | | <u> </u> | | Pregnancy testing, counselling and referral |
| | | | | NA NA NA NA | | Sexuality and relationship counselling |
| | | | | IVA IVA IVA IVA | 1 | |
| The Talk Shop | Yonge St./Empress Ave. | 5110 Yonge St., Toronto, On M2N 6M1 | 416-338-7000 | | | Birth control counselling |
| The Talk Office | ronge descripted nie. | orrorongo ot, fotolito, Oli Mari Oli I | 710 000-7000 | | | Low cost or free birth control |
| | | | | Monday: 2 pm - 6:30 pm | | Free condoms |
| | | | | | | |
| | | | | Wednesday: 9:30 am - 11:30 am, 1 pm - 3:30 pm | | Plan B (emergency contraceptive pill) |
| | | | | Thursday: 2 pm - 6:30 pm | | STI testing and free treatment |
| | | | | | | HIV testing |
| | | | | | | Pregnancy testing, counselling and referral |
| | | | | | | Sexuality and relationship counselling |
| | | | | | | Rapid HIV testing |
| | | | | | | |
| | | | | | | |
| | | | | | | |
| | N. | IA KIA KIA KI | ٨ | | | |
| | - 1 | IA NA NA NA | | | | |
| | | */ | 1 | | | |
| | | | | | | 1 N I N N I N |
| | | | | | | NA NA |
| | | | | | 1 4/ / 1 4/ | 1 1 4/ 1 1 4/ 1 |
| | | | | | | |
| | | | | | | |

D

| | A | В | С | 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 |

Α

| | Α | В | С |
|---|-----|----------|------------|
| 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 |

В

| | Α | В | С | 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