

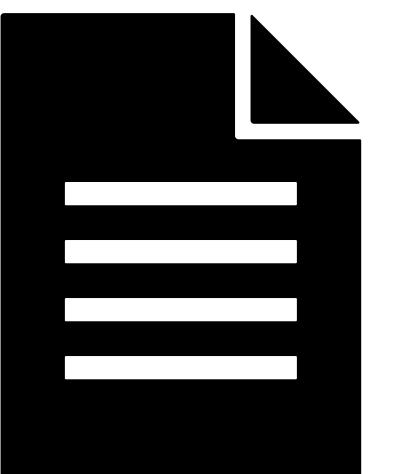
Data organization in spreadsheets

Section A02

Zhuji

Be consistent

- Use consistent codes for categorical variables
 - Male, M, male, m... (Stick to one of them!)
- Use a consistent fixed code for any missing values
 - N/A, NaN, Null ☒ -999, 999 ✗
- Use consistent variable names
 - grades_wk10, wk10Grades...
- Use consistent subject identifiers
 - studentA, aStudent, a...
- Use a consistent data layout in multiple files
 - -
- Use consistent file names
 - “quiz_010123.csv”, “010123q.csv”
- Use a consistent format for all dates
 - YYYY-MM-DD
- Use consistent phrases in your notes
 - -
- Be careful about extra spaces within cells
 - “Male” ≠ “ Male”



Choose good names for things

- don't use spaces
- glucose_6_weeks instead of glucose 6 weeks
- Careful about the spaces
 - Trailing Spaces
 - Leading Spaces
- Avoid special characters (#, \$, &, !)
- Short and meaningful



Dates

ISO 8601 standard

- YYYY-MM-DD
- YYYYMMDD

PUBLIC SERVICE ANNOUNCEMENT:

OUR DIFFERENT WAYS OF WRITING DATES AS NUMBERS CAN LEAD TO ONLINE CONFUSION. THAT'S WHY IN 1988 ISO SET A GLOBAL STANDARD NUMERIC DATE FORMAT.

THIS IS *THE* CORRECT WAY TO WRITE NUMERIC DATES:

2013-02-27


THE FOLLOWING FORMATS ARE THEREFORE DISCOURAGED:

02/27/2013 02/27/13 27/02/2013 27/02/13

20130227 2013.02.27 27.02.13 27-02-13

27.2.13 2013.II.27. $27\frac{1}{2}$ -13 2013.158904109

MMXIII-II-XXVII MMXIII $\frac{\text{LVII}}{\text{CCCLXV}}$ 1330300800

$((3+3) \times (111+1) - 1) \times 3 / 3 - 1 / 3^3$ ~~2013~~ 

10/11011/1101 02/27/20/13 $\begin{matrix} 2 & 3 & 1 & 4 \\ 0 & 1 & 2 & 3 & 7 \\ 5 & 6 & 7 & 8 \end{matrix}$

Cells

- No empty cells
 - N/A , NaN, Null
 - Blank cells could cause confusion
- Put just one thing in a cell
 - Don't combine multiple variables

NaN

Make it a rectangle

- First row should contain variable names
- Don't use more than one row for variable names

| | A | B | C | D | E |
|---|-----|--------|---------|---------|---------|
| 1 | id | sex | glucose | insulin | triglyc |
| 2 | 101 | Male | 134.1 | 0.60 | 273.4 |
| 3 | 102 | Female | 120.0 | 1.18 | 243.6 |
| 4 | 103 | Male | 124.8 | 1.23 | 297.6 |
| 5 | 104 | Male | 83.1 | 1.16 | 142.4 |
| 6 | 105 | Male | 105.2 | 0.73 | 215.7 |

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

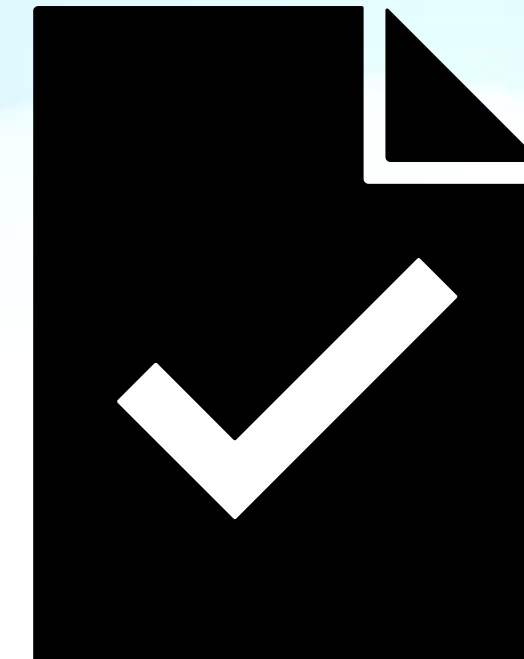
| | A | B | C | D | E | F | G |
|---|------|-------|--------|-------|-------|--------|-------|
| 1 | 1MIN | | | | | | |
| 2 | | | Normal | | | Mutant | |
| 3 | B6 | 146.6 | 138.6 | 155.6 | 166 | 179.3 | 186.9 |
| 4 | BTBR | 245.7 | 240 | 243.1 | 177.8 | 171.6 | 188.1 |
| 5 | | | | | | | |
| 6 | 5MIN | | | | | | |
| 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 |

Raw data files

- Primary data file should be a pristine store of data.
- Make a copy before making any calculations on data
- Do not write it on the original data files
- Make back-ups on your data in case of corruption/overwrite



Font color/Highlighting

Don't use them!

- Nice visually, but hard for later analysis
- Better encode highlight info in another column

| | A | B | C | D | E | F | G | H |
|----|--------------------------------|---------|---------|--------|-------|---------|---------|--------|
| 1 | | Mercury | Venus | Earth | Mars | Jupiter | Saturn | Uranus |
| 2 | Mass (10^{24} kg) | 0.33 | 4.87 | 5.97 | 0.642 | 1898 | 568 | 86.8 |
| 3 | Diameter (km) | 4879 | 12,104 | 12,756 | 6792 | 142,984 | 120,536 | 51,118 |
| 4 | Density (kg/m^3) | 5427 | 5243 | 5514 | 3933 | 1326 | 687 | 1271 |
| 5 | Gravity (m/s^2) | 3.7 | 8.9 | 9.8 | 3.7 | 23.1 | 9 | 8.7 |
| 6 | Escape Velocity (km/s) | 4.3 | 10.4 | 11.2 | 5 | 59.5 | 35.5 | 21.3 |
| 7 | Rotation Period (hours) | 1407.6 | -5832.5 | 23.9 | 24.6 | 9.9 | 10.7 | -17.2 |
| 8 | Length of Day (hours) | 4222.6 | 2802 | 24 | 24.7 | 9.9 | 10.7 | 17.2 |
| 9 | Distance from Sun (10^6 km) | 57.9 | 108.2 | 149.6 | 227.9 | 778.6 | 1433.5 | 2872.5 |
| 10 | Perihelion (10^5 km) | 46 | 107.5 | 147.1 | 206.6 | 740.5 | 1352.6 | 2741.3 |
| 11 | Aphelion (10^6 km) | 69.8 | 108.9 | 152.1 | 249.2 | 816.6 | 1514.5 | 3003.6 |
| 12 | Orbital Period (days) | 88 | 224.7 | 365.2 | 687 | 4331 | 10,747 | 30,589 |
| 13 | Orbital Velocity (km/s) | 47.4 | 35 | 29.8 | 24.1 | 13.1 | 9.7 | 6.8 |
| 14 | Orbital Inclination | 7 | 3.4 | 0 | 1.9 | 1.3 | 2.5 | 0.8 |
| 15 | Orbital Eccentricity | 0.205 | 0.007 | 0.017 | 0.094 | 0.049 | 0.057 | 0.046 |