



Introduction to R

Module 11-04, Tidy Data

Billy Fetzner

Tidy Data



- 3 rules to create tidy data

Variables

Observations

Values

Letter	Color	Number	Count
a	blue	548	67
a	green	857	78
b	blue	3689	9
b	green	8765	78
c	blue	2345	34
c	green	6768798	6778

Letter	Color	Number	Count
a	blue	548	67
a	green	857	78
b	blue	3689	9
b	green	8765	78
c	blue	2345	34
c	green	6768798	6778

Letter	Color	Number	Count
a	blue	6548	67
a	green	857	78
b	blue	3689	9
b	green	8765	78
c	blue	2345	34
c	green	6768798	6778



Why Tidy Data?



- Easier to learn tools that work with all of your datasets
- Variables in columns supports R's vectorized nature
- Other packages in tidyverse are designed to work with tidy data
 - dplyr
 - ggplot2

Why would anything be untidy?



- Data entry

	A	B	C	D	E	F	G	H
1	Block 1	Condition 1						
2		Var 1		Var 2		Var 3		
3		count	positive	count	positive	count	positive	
4	Day 1							
5	Day 2							
6	Day 3							
7	Day 4							
8	Day 5							
9	Day 6							
10	Day 7							
11	Day 8							
12	Day 9							
13	Day 10							
14	Day 11							
15	Day 12							
16	Day 13							
17	Day 14							
18	Day 15							
19	Day 16							
20	Day 17							
21	Day 18							
22	Day 19							
23	Day 20							
24	Day 21							
25	Day 22							
26								