

# Tidy Data in R

## The Structure of Tidy Data

1. Each variable in a dataset should be in one column
2. Each observation of that variable should be in a different row
3. There should be only one table for each kind of variable \*\*\*

## Let's view examples

Example1 - Transforming this dataset in order to have each variable in a column and each observation in a row.

This dataset was created in R to demonstrate a common type of untidy data

### Before

	Product_ID	Country	Costco.01	Walmart.01	Costco.02	Walmart.02
1	XDS12	CAN	5423	3908	4614	8748
2	XDS19	CAN	6095	1835	1909	9152
3	XDS9	CAN	3891	4562	3292	4566
4	XDS7	CAN	4520	6221	5210	2078
5	XDS2	CAN	5275	7805	5599	7497
6	XDS17	CAN	4107	6717	7108	3360

### After

	Product_ID	Country	Transit	Costco	Walmart
1	XDS1	USA	01	3485	4559
2	XDS1	USA	02	3217	9754
3	XDS10	USA	01	8431	5235
4	XDS10	USA	02	9971	3042
5	XDS11	USA	01	6435	3749
6	XDS11	USA	02	2403	9691

Example2 - Transforming this dataset by splitting a single column into four real variables: LastName - FirstName - Title - PhoneNumber

This dataset was created in Excel to demonstrate another common problem with messy data: Many variables are stored in one column

Before

ID	Company	Contact.Person
1	1 Company A	Bedecs/Anna/Owner/((123)555-0100)
2	2 Company B	Gratacos Solsona/Antonio/Owner/((123)555-0100)
3	3 Company C	Axen/Thomas/Purchasing Representative/((123)555-0100)
4	4 Company D	Lee/Christina/Purchasing Manager/((123)555-0100)
5	5 Company E	O<U+0092>Donnell/Martin/Owner/((123)555-0100)
6	6 Company F	Pérez-Olaeta/Francisco/Purchasing Manager/((123)555-0100)

After

	ID	Company	LastName	FirstName	Title	PhoneNumber
1	1	Company A	Bedecs	Anna	Owner	123 555 0100
2	2	Company B	Gratacos Solsona	Antonio	Owner	123 555 0100
3	3	Company C	Axen	Thomas	Purchasing Representative	123 555 0100
4	4	Company D	Lee	Christina	Purchasing Manager	123 555 0100
5	5	Company E	ODonnell	Martin	Owner	123 555 0100
6	6	Company F	Pérez-Olaeta	Francisco	Purchasing Manager	123 555 0100