

# Data wrangling with Tableau and Excel

—

October 11 2016

JRNL 520H

# What is data wrangling?

Data wrangling is the process of preparing raw data for use in a data analysis or visualization software.

# What are the causes of dirty data?

- Data entry error

# What are the causes of dirty data?

- Data entry error
- Incompatible tables

# What are the causes of dirty data?

- Data entry error
- Incompatible tables
- Incompatible table format

# What should we look out for when cleaning data?

- Table formatting

# What should we look out for when cleaning data?

- Table formatting
- Variable type

# What should we look out for when cleaning data?

- Table formatting
- Variable type
- Invalid character values



# What should we look out for when cleaning data?

- Table formatting
- Variable type
- Invalid character values
- Invalid numeric values

# What should we look out for when cleaning data?

- Table formatting
- Variable type
- Invalid character values
- Invalid numeric values
- Grouping data

# What should we look out for when cleaning data?

- Table formatting
- Variable type
- Invalid character values
- Invalid numeric values
- Grouping data
- Missing values

# Ideal format of data in Tableau

1. Start your data in cell A1. Remove all introductory information and footnotes.
2. Have the first row be the column headers/variable names
3. Have every subsequent row be one observation. No cross-tabulation!

# Ideal format of data in Tableau

Before

After

Market rental pricing for July 2016				
	1 Bedroom		2 Bedrooms	
City	Price	M/M %	Price	M/M %
Vancouver	1740	0.024	2750	-0.011
Toronto	1350	0.023	1720	0.042
Calgary	1110	0.037	1340	0.031
Victoria	1075	-0.023	1370	-0.021
Ottawa	1040	0.03	1280	-0.015
Edmonton	1010	0.031	1250	0.016
Regina	980	0.021	1180	0
Montreal	960	0.011	1250	0.042
Kingston	940	0.011	1070	-0.027
Kelowna	920	0.022	1400	0.029
Barrie	910	0.034	1300	0.008
Halifax	910	0.022	1100	0.019
St.Catherines	905	-0.005	1130	0
Oshawa	890	-0.011	1180	-0.017
Hamilton	880	-0.011	1050	-0.037
Winnipeg	880	0.023	1100	0.028
Saskatoon	870	0.048	1100	0
Kitchener	860	0	1000	-0.02
Windsor	840	0.024	1020	0.02
Abbotsford	800	0.013	1150	0.045
Quebec City	800	0.013	1070	0.029
London	790	0	950	-0.01
St. John's	775	-0.006	860	0
Sherbrooke	560	-0.018	630	-0.031
Saguenay	540	-0.018	620	-0.016

Source: <http://blog.padmapper.com/category/rental-data/>

# Ideal format of data in Tableau

Before

Market rental pricing for July 2016				
	1 Bedroom		2 Bedrooms	
City	Price	M/M %	Price	M/M %
Vancouver	1740	0.024	2750	-0.011
Toronto	1350	0.023	1720	0.042
Calgary	1110	0.037	1340	0.031
Victoria	1075	-0.023	1370	-0.021
Ottawa	1040	0.03	1280	-0.015
Edmonton	1010	0.031	1250	0.016
Regina	980	0.021	1180	0
Montreal	960	0.011	1250	0.042
Kingston	940	0.011	1070	-0.027
Kelowna	920	0.022	1400	0.029
Barrie	910	0.034	1300	0.008
Halifax	910	0.022	1100	0.019
St.Catherines	905	-0.005	1130	0
Oshawa	890	-0.011	1180	-0.017
Hamilton	880	-0.011	1050	-0.037
Winnipeg	880	0.023	1100	0.028
Saskatoon	870	0.048	1100	0
Kitchener	860	0	1000	-0.02
Windsor	840	0.024	1020	0.02
Abbotsford	800	0.013	1150	0.045
Quebec City	800	0.013	1070	0.029
London	790	0	950	-0.01
St. John's	775	-0.006	860	0
Sherbrooke	560	-0.018	630	-0.031
Saguenay	540	-0.018	620	-0.016

Source: <http://blog.padmapper.com/category/rental-data/>

After

City	1 Bdrm Price	1 Bdrm M/M %	2 Bdrm Price	2 Bdrm M/M %
Vancouver	1740	0.024	2750	-0.011
Toronto	1350	0.023	1720	0.042
Calgary	1110	0.037	1340	0.031
Victoria	1075	-0.023	1370	-0.021
Ottawa	1040	0.03	1280	-0.015
Edmonton	1010	0.031	1250	0.016
Regina	980	0.021	1180	0
Montreal	960	0.011	1250	0.042
Kingston	940	0.011	1070	-0.027
Kelowna	920	0.022	1400	0.029
Barrie	910	0.034	1300	0.008
Halifax	910	0.022	1100	0.019
St.Catherines	905	-0.005	1130	0
Oshawa	890	-0.011	1180	-0.017
Hamilton	880	-0.011	1050	-0.037
Winnipeg	880	0.023	1100	0.028
Saskatoon	870	0.048	1100	0
Kitchener	860	0	1000	-0.02
Windsor	840	0.024	1020	0.02
Abbotsford	800	0.013	1150	0.045
Quebec City	800	0.013	1070	0.029
London	790	0	950	-0.01
St. John's	775	-0.006	860	0
Sherbrooke	560	-0.018	630	-0.031
Saguenay	540	-0.018	620	-0.016



One header with descriptive variable names



Remove extra information above and below data

# Data Interpreter

Tableau's Data Interpreter feature draws out sub-tables and removes some of that extraneous information to help prepare your data source for analysis. Note: the data interpreter only works with Microsoft Excel files, not CSV or other file types.

# Data Interpreter

Tableau's Data Interpreter feature draws out sub-tables and removes some of that extraneous information to help prepare your data source for analysis. Note: the data interpreter only works with Microsoft Excel files, not CSV or other file types.

**Complete Tableau exercise**



# Joins

A JOIN is a means for combining columns from one or more tables by using values common to each. There are four main join types: inner, left, right and full outer.

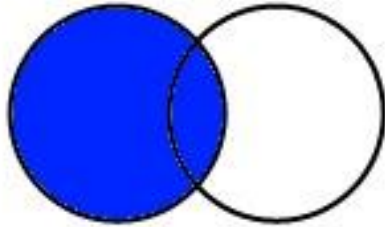
# Joins

Market rental pricing for July 2016				
City	1 Bedroom		2 Bedrooms	
	Price	M/M %	Price	M/M %
Vancouver	1740	0.024	2750	-0.011
Toronto	1350	0.023	1720	0.042
Calgary	1110	0.037	1340	0.031
Victoria	1075	-0.023	1370	-0.021
Ottawa	1040	0.03	1280	-0.015
Edmonton	1010	0.031	1250	0.016
Regina	980	0.021	1180	0
Montreal	960	0.011	1250	0.042
Kingston	940	0.011	1070	-0.027
Kelowna	920	0.022	1400	0.029
Barrie	910	0.034	1300	0.008
Halifax	910	0.022	1100	0.019

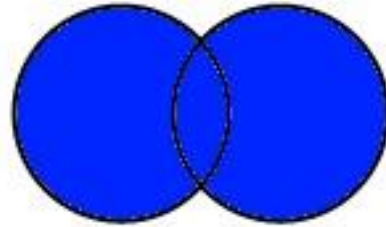
Market rental pricing for June 2016				
City	1 Bedroom		2 Bedrooms	
	Price	M/M %	Price	M/M %
Vancouver	\$1,700	-2.30%	\$2,780	-0.70%
Toronto	\$1,320	1.50%	\$1,650	-1.20%
Calgary	\$1,100	1.90%	\$1,400	2.90%
Victoria	\$1,070	-2.70%	\$1,300	-3.70%
Ottawa	\$1,010	-1.90%	\$1,300	-1.50%
Edmonton	\$980	3.20%	\$1,230	2.50%
Regina	\$960	-4.00%	\$1,180	0.00%
Montreal	\$950	2.20%	\$1,200	-4.00%
Kingston	\$930	0.00%	\$1,100	4.80%
Kelowna	\$910	-4.20%	\$1,130	1.80%
Barrie	\$900	-2.20%	\$1,200	1.70%
Halifax	\$900	1.10%	\$1,360	1.50%

# Joins

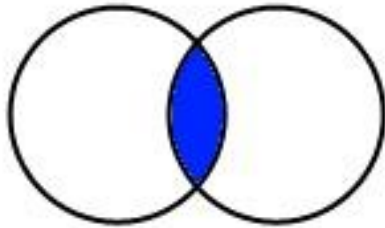
**LEFT JOIN**



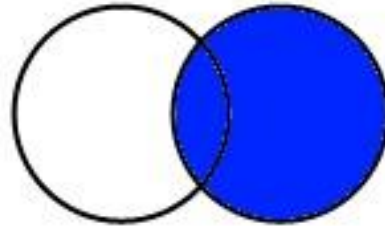
**FULL OUTER JOIN**



**INNER JOIN**



**RIGHT JOIN**



# Joins

Market rental pricing for July 2016					Market rental pricing for June 2016				
1 Bedroom			2 Bedrooms		1 Bedroom			2 Bedrooms	
City	Price	M/M %	Price	M/M %	City	Price	M/M %	Price	M/M %
Vancouver	1740	0.024	2750	-0.011	Vancouver	\$1,700	-2.30%	\$2,780	-0.70%
Toronto	1350	0.023	1720	0.042	Toronto	\$1,320	1.50%	\$1,650	-1.20%
Calgary	1110	0.037	1340	0.031	Calgary	\$1,100	1.90%	\$1,400	2.90%
Victoria	1075	-0.023	1370	-0.021	Victoria	\$1,070	-2.70%	\$1,300	-3.70%
Ottawa	1040	0.03	1280	-0.015	Ottawa	\$1,010	-1.90%	\$1,300	-1.50%
Edmonton	1010	0.031	1250	0.016	Edmonton	\$980	3.20%	\$1,230	2.50%
Regina	980	0.021	1180	0	Regina	\$960	-4.00%	\$1,180	0.00%
Montreal	960	0.011	1250	0.042	Montreal	\$950	2.20%	\$1,200	-4.00%
Kingston	940	0.011	1070	-0.027	Kingston	\$930	0.00%	\$1,100	4.80%
Kelowna	920	0.022	1400	0.029	Kelowna	\$910	-4.20%	\$1,130	1.80%
Barrie	910	0.034	1300	0.008	Barrie	\$900	-2.20%	\$1,200	1.70%
Halifax	910	0.022	1100	0.019	Halifax	\$900	1.10%	\$1,360	1.50%
					St.Catherines	\$890	-3.30%	\$1,090	3.80%

# Joins

Market rental pricing for July 2016					Market rental pricing for June 2016				
1 Bedroom			2 Bedrooms		1 Bedroom			2 Bedrooms	
City	Price	M/M %	Price	M/M %	City	Price	M/M %	Price	M/M %
Vancouver	1740	0.024	2750	-0.011	Vancouver	\$1,700	-2.30%	\$2,780	-0.70%
Toronto	1350	0.023	1720	0.042	Toronto	\$1,320	1.50%	\$1,650	-1.20%
Calgary	1110	0.037	1340	0.031	Calgary	\$1,100	1.90%	\$1,400	2.90%
Victoria	1075	-0.023	1370	-0.021	Victoria	\$1,070	-2.70%	\$1,300	-3.70%
Ottawa	1040	0.03	1280	-0.015	Ottawa	\$1,010	-1.90%	\$1,300	-1.50%
Edmonton	1010	0.031	1250	0.016	Edmonton	\$980	3.20%	\$1,230	2.50%
Regina	980	0.021	1180	0	Regina	\$960	-4.00%	\$1,180	0.00%
Montreal	960	0.011	1250	0.042	Montreal	\$950	2.20%	\$1,200	-4.00%
Kingston	940	0.011	1070	-0.027	Kingston	\$930	0.00%	\$1,100	4.80%
Kelowna	920	0.022	1400	0.029	Kelowna	\$910	-4.20%	\$1,130	1.80%
Barrie	910	0.034	1300	0.008	Barrie	\$900	-2.20%	\$1,200	1.70%
Halifax	910	0.022	1100	0.019	Halifax	\$900	1.10%	\$1,360	1.50%
					St.Catherines	\$890	-3.30%	\$1,090	3.80%

**Complete Tableau exercise**

# Wrangling in Excel

Sometimes the data interpreter in Tableau isn't able to detect all of the errors in the dataset. In cases like this, you will need to manually clean the data in Excel.

**Complete Tableau exercise**

# Pivot

## Tabular format

Abc CMHC Data Location	# CMHC Data 1992	# CMHC Data 1993	# CMHC Data 1994	# CMHC Data 1995	# CMHC Data 1996	# CMHC Data 1997	# CMHC Data 1998	# CMHC Data 1999	# CMHC Data 2000	# CMHC Data 2001	# CMHC Data 2002
Saskatchewan	454	458	459	465	477	494	507	522	529	546	554
Alberta	553	548	541	540	543	565	607	633	651	701	734
British Columbia	682	699	716	725	737	739	746	742	753	772	795
Metropolitan Areas	<i>null</i>	<i>null</i>	<i>null</i>	<i>null</i>	<i>null</i>	<i>null</i>	<i>null</i>	<i>null</i>	<i>null</i>	<i>null</i>	<i>null</i>
St. John's	566	554	559	565	570	567	513	517	552	575	589
Halifax	595	604	610	615	617	616	631	637	648	673	704
Moncton	469	479	495	504	513	523	531	538	560	561	578
Saint John	429	436	439	437	441	449	452	457	460	483	492
Saquenav	420	419	416	417	423	425	428	428	438	439	440

## Columnar format

Abc CMHC Data Location	Pivot Year	# Pivot Rent
Canada2	1992	568
Provinces	1992	<i>null</i>
Newfoundland and L...	1992	517
Prince Edward Island	1992	503
Nova Scotia	1992	562
New Brunswick	1992	461
Quebec	1992	473
Ontario	1992	672
Manitoba	1992	545
Saskatchewan	1992	454
Alberta	1992	553
British Columbia	1992	682
Metropolitan Areas	1992	<i>null</i>

# Pivot

## Tabular format

Abc CMHC Data Location	# CMHC Data 1992	# CMHC Data 1993	# CMHC Data 1994	# CMHC Data 1995	# CMHC Data 1996	# CMHC Data 1997	# CMHC Data 1998	# CMHC Data 1999	# CMHC Data 2000	# CMHC Data 2001	# CMHC Data 2002
Saskatchewan	454	458	459	465	477	494	507	522	529	546	554
Alberta	553	548	541	540	543	565	607	633	651	701	734
British Columbia	682	699	716	725	737	739	746	742	753	772	795
Metropolitan Areas	<i>null</i>	<i>null</i>	<i>null</i>	<i>null</i>	<i>null</i>	<i>null</i>	<i>null</i>	<i>null</i>	<i>null</i>	<i>null</i>	<i>null</i>
St. John's	566	554	559	565	570	567	513	517	552	575	589
Halifax	595	604	610	615	617	616	631	637	648	673	704
Moncton	469	479	495	504	513	523	531	538	560	561	578
Saint John	429	436	439	437	441	449	452	457	460	483	492
Saquenav	420	419	416	417	423	425	428	428	438	439	440

## Columnar format

Abc CMHC Data Location	Abc Pivot Year	# Pivot Rent
Canada2	1992	568
Provinces	1992	<i>null</i>
Newfoundland and L...	1992	517
Prince Edward Island	1992	503
Nova Scotia	1992	562
New Brunswick	1992	461
Quebec	1992	473
Ontario	1992	672
Manitoba	1992	545
Saskatchewan	1992	454
Alberta	1992	553
British Columbia	1992	682
Metropolitan Areas	1992	<i>null</i>

Complete Tableau exercise