

# Transforming Data

---



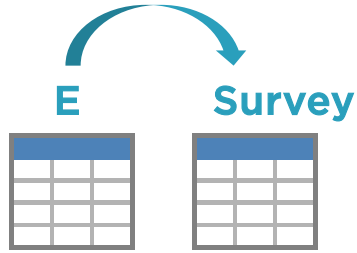
**Stacia Varga**

PRINCIPAL CONSULTANT

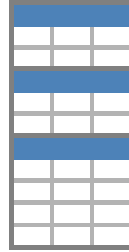
@\_StaciaV\_ [blog.datainspirations.com](http://blog.datainspirations.com)



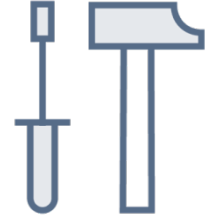
# Transforming Data



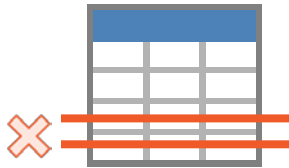
Rename objects



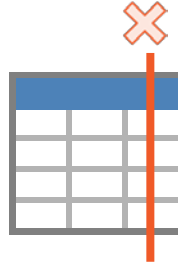
Combine queries



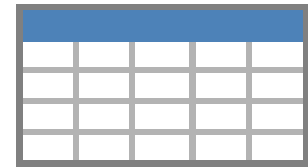
Fixing metadata



Filtering rows



Eliminating columns



Adding columns

And many more transformations!



# Renaming Queries

e2014wa...		

g20141wa		

E		

geography		



# Renaming Queries

WA Survey		

OR Survey		

CA Survey		

WA Geo		

OR Geo		

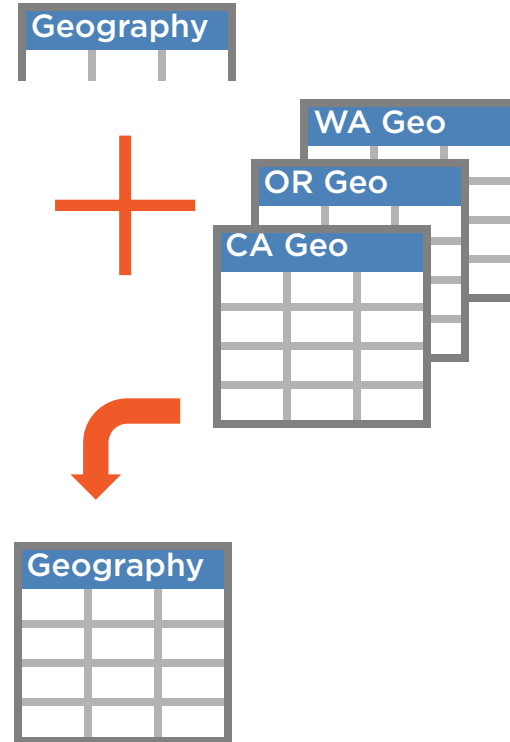
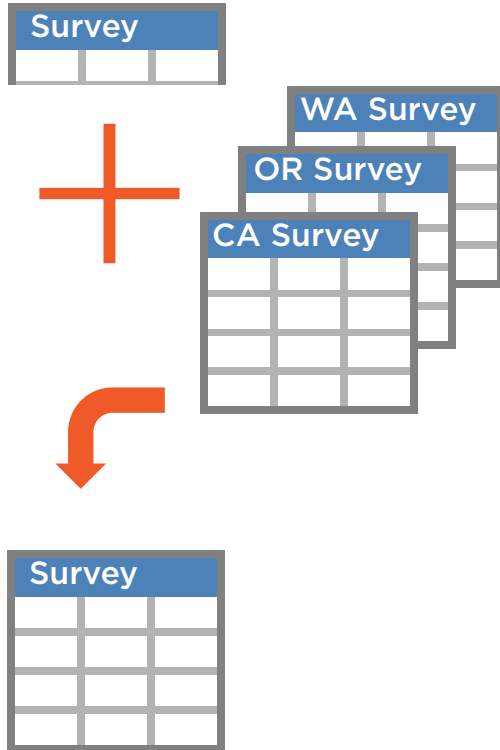
CA Geo		

Survey		

Geography		

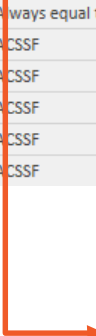


# Combining Queries: Append



# Fixing Metadata

	Column1	Column2	Column3
1	FILEID	STUSAB	SUMLEVEL
2	Always equal to ACS Summary File identification	State Postal Abbreviation	Summary Level
3	ACSSF	CA	40
4	ACSSF	CA	40
5	ACSSF	CA	40
6	ACSSF	CA	40
7	ACSSF	CA	40



	FILEID	STUSAB	SUMLEVEL
1	Always equal to A...	State Postal Abbr...	Summary Level
2	ACSSF	CA	40
3	ACSSF	CA	40
4	ACSSF	CA	40



# Fixing Metadata

	FILEID	STUSAB	SUMLEVEL
1	Always equal to A...	State Postal Abbr...	Summary Level
2	ACSSF	CA	40
3	ACSSF	CA	40
4	ACSSF	CA	40



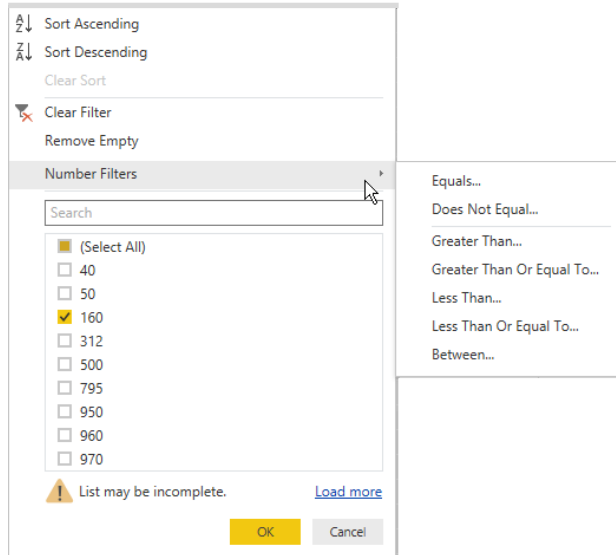
	FILEID	STUSAB	SUMLEVEL
1	ACSSF	CA	40
2	ACSSF	CA	40
3	ACSSF	CA	40
4	ACSSF	CA	40

# Fixing Metadata

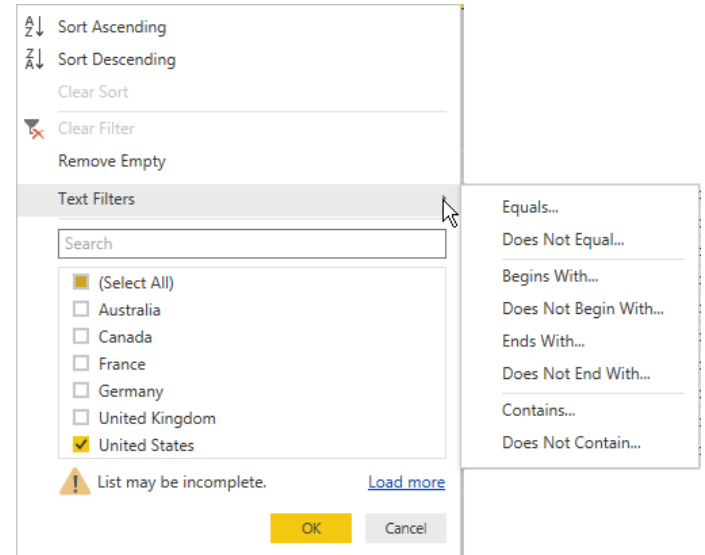




# Filtering Rows



**Number Filters**



**Text Filters**



# Eliminating Columns

## Choose columns to keep

III	Asc	Zip Code	123	STATE	123	PLACE	Asc	CLASSFP	123	GEOID	123	POPPT	123	HUPT
1		601		72		358	U1			7200358		4406		1968
2		602		72		616	U1			7200616		3212		1648
3		602		72		47873	U1			7247873		2886		1205
4		603		72		745	U1			7200745		13310		7479



III	Asc	Zip Code	123	STATE	123	PLACE
1		601		72		358
2		602		72		616
3		602		72		47873
4		603		72		745

## Choose columns to remove

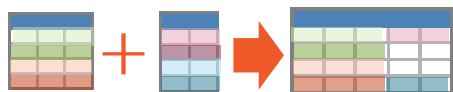
III	Asc	STATE	123	LOGRECNO	Asc	GEOID	Asc	GEOGRAPHY NAME
1		CA		1		04000US06		California
2		CA		2		04001US06		California -- Urban
3		CA		3		04043US06		California -- Rural
4		CA		4		040A0US06		California -- In metropolitan or micropolitan statistical area



III	Asc	STATE	123	LOGRECNO	Asc	GEOGRAPHY NAME
1		CA		1		California
2		CA		2		California -- Urban
3		CA		3		California -- Rural
4		CA		4		California -- In metropolitan or micropolitan statistical area



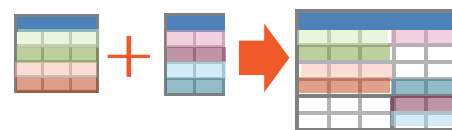
# Combining Queries: Merge



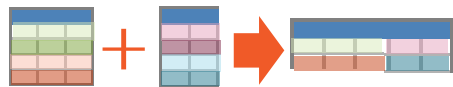
Left outer



Right outer



Full outer



Inner



Left anti



Right anti



# Adding Columns: Duplicate

	Code	STATE	PLACE	LOGRECNO	NAME
1		6	44000	110	Los Angeles city, California
2		6	44000	110	Los Angeles city, California
3		6	44000	110	Los Angeles city, California
4		6	44000	110	Los Angeles city, California

# Adding Columns: Duplicate

Duplicate a column



	p Code	123 STATE	123 PLACE	123 LOGRECNO	Asc NAME	Asc NAME - Copy
1		6	44000	110	Los Angeles city, California	Los Angeles city, California
2		6	44000	110	Los Angeles city, California	Los Angeles city, California
3		6	44000	110	Los Angeles city, California	Los Angeles city, California
4		6	44000	110	Los Angeles city, California	Los Angeles city, California

# Adding Columns: Split

Split by delimiter

	Asc Zip Code	123 STATE	123 PLACE	123 LOGRECNO	Asc NAME	Asc NAME - Copy.1	Asc NAME - Copy.2
1	90001	6	44000	110	Los Angeles city, California	Los Angeles city	California
2	90002	6	44000	110	Los Angeles city, California	Los Angeles city	California
3	90003	6	44000	110	Los Angeles city, California	Los Angeles city	California
4	90004	6	44000	110	Los Angeles city, California	Los Angeles city	California



# Summary



Renaming queries

Combining queries

Fixing metadata

Filtering rows

Eliminating columns

Adding columns

