



Visualizing Crime inter-Alberta and within Calgary

CPSC 599 – DATA VISUALIZATION

DEBBIE TEE MACROHON

Table of Contents

Introduction	2
Data Sets	2
Alberta Crime Data Set	2
Cleansing, Augmenting, Profiling the Data	4
Calgary Communities Crime Data Set	9
Cleansing, Augmenting, Profiling the Data	11
Calgary Shape File	14
Decisions	15
Research Questions	15
Backup Questions	16

Introduction

I chose 3 different data sets, the first one outlining the different crimes in Alberta, and between cities in Alberta such as Calgary, Edmonton and Medicine Hat, from 1998 to around 2014. The second data set I chose is detailing the number of categorized crimes by community in Calgary from 2012 to 2016. The third data set is a shape file outlining the boundaries of each community, along with when structure details of it. I initially chose the second data set because I wanted to know if Forest Lawn, the community that I live in, was safe. However, this data set was not rich enough because it had a limited number of categories and spanned across only a few years. I decided to integrate the former data set because there are more datasets available for provinces, and not the varies communities in the cities that reside in that province. I hope to understand the trend of crime over the years for Alberta, Calgary, and various communities in Calgary and see if Calgary is becoming a more dangerous place.

Data Sets

Here are the following data sets that were selected.

Alberta Crime Data Set

The first data set was chosen from this website:

<https://open.canada.ca/data/en/dataset/5d203dae-26d2-4f21-85de-c69e33d9f254>

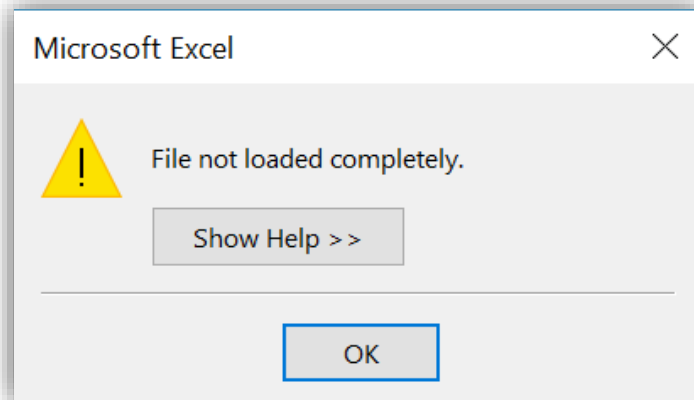


Figure 1 - File was too big to load into Excel

	A	B	C	D	E	F	G	H
1	Ref_Date	GEO	Geograph	VIOLATIO	STA	Vector	Coordinate	Value
2	1998	Alberta	48	Total, all v	Actual inc	v7182465	1.1.1	289665
3	1999	Alberta	48	Total, all v	Actual inc	v7182465	1.1.1	290612
4	2000	Alberta	48	Total, all v	Actual inc	v7182465	1.1.1	286787
5	2001	Alberta	48	Total, all v	Actual inc	v7182465	1.1.1	306773
6	2002	Alberta	48	Total, all v	Actual inc	v7182465	1.1.1	316476
7	2003	Alberta	48	Total, all v	Actual inc	v7182465	1.1.1	342201
8	2004	Alberta	48	Total, all v	Actual inc	v7182465	1.1.1	350224
9	2005	Alberta	48	Total, all v	Actual inc	v7182465	1.1.1	353060
10	2006	Alberta	48	Total, all v	Actual inc	v7182465	1.1.1	348036
11	2007	Alberta	48	Total, all v	Actual inc	v7182465	1.1.1	353358
12	2008	Alberta	48	Total, all v	Actual inc	v7182465	1.1.1	261212

Figure 1 - removing unnecessary columns

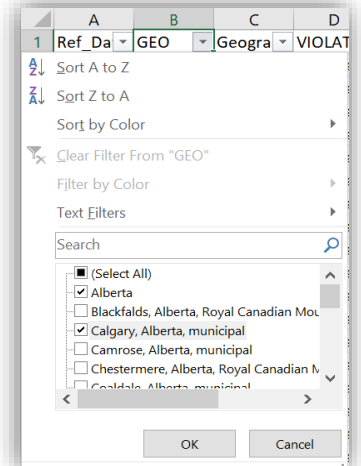


Figure 3 - Filtering out unnecessary locations

I opened the file, however it was not loaded completely because Excel can only handle around 2 million rows. The columns include the year, geography, type of violation, STA, and value, and includes data of Alberta as well as its municipalities. I also removed the Vector and Coordinate columns to save on space. I filtered the GEO column by Alberta, Medicine Hat, Calgary and Edmonton. We end up with around a million rows, and 5 columns:

Average: 2008.051404 Count: 1047474 Sum: 2103379628

Cleansing, Augmenting, Profiling the Data

I transformed the data into two different data sets, one pivoted by violation and the other pivoted by the column STA. I do this to calculate the correlations between violations or STA, which will allow for multiple views of the same data.

Current Sample	
Full Dataset	87,618 rows
02/15/2018	

By Violation

Using Trifacta Wrangler, I pivoted the Violations Column so I could later do analysis on the different correlations between each type of crime.

< Recipe

Edit Step

×

Choose a transformation

pivot

Creates a new column for each unique value in a column and aggregates the result

Column

required

VIOLATIONS

×

Choose column

Functions

required

sum(value)

×

Edit formula

Group by

Ref_Date

×

GEO

×

Geographical_classification

×

Choose column

Max columns to create

600

⌵

Cancel

Add

Figure 4 - Pivoting the Violations column. Summing the value column is fine because it is grouped by the year, location and geographical classification.

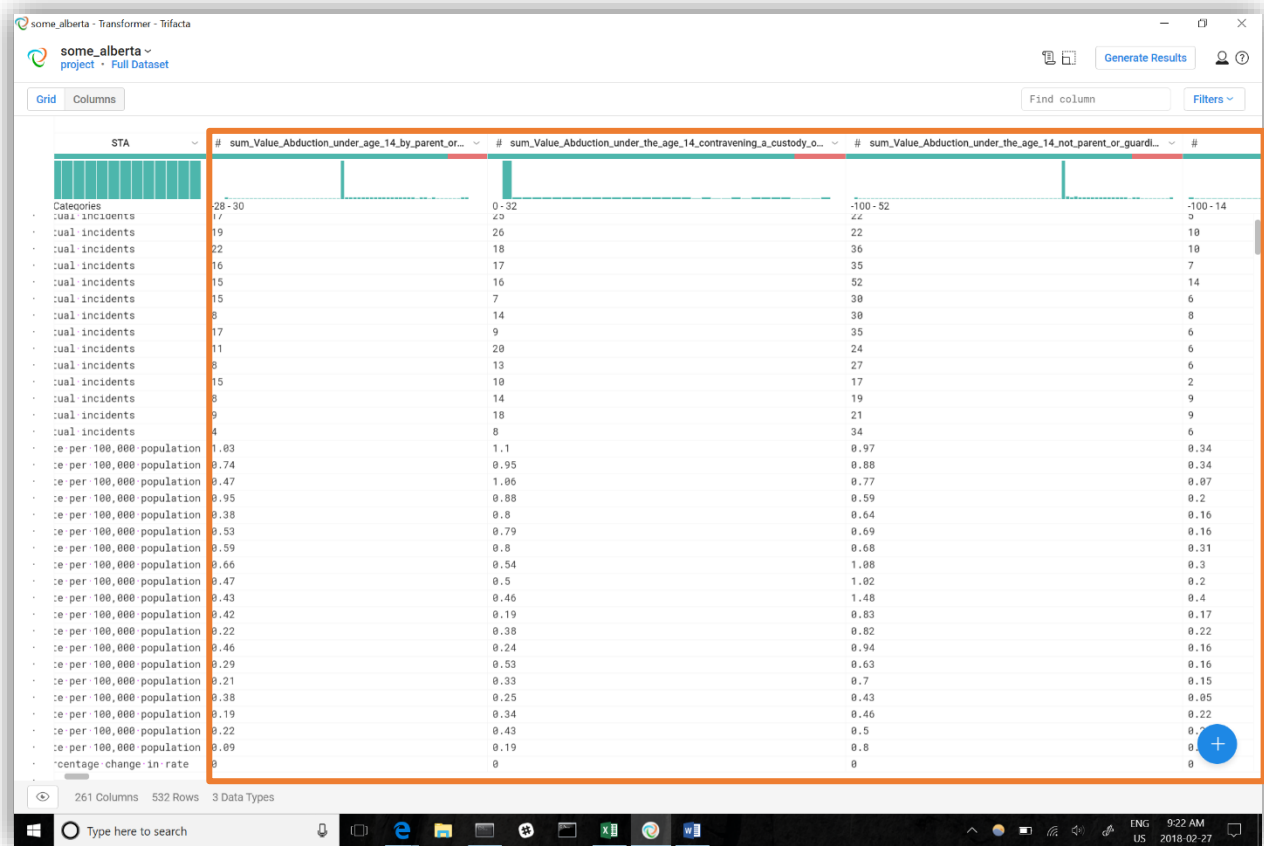


Figure 5 - new columns that have generated from the pivot

I didn't run into any problems here. It was straightforward and the only challenge was learning how the pivot worked. When profiling the data:

GridColumns

Find columnFilters

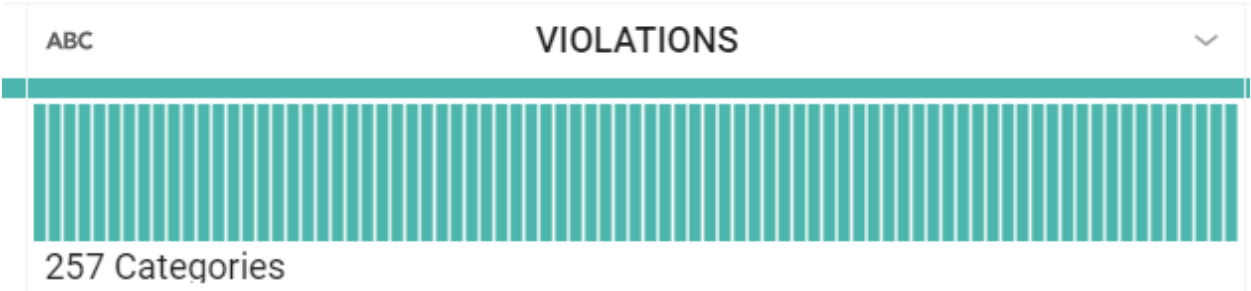
Preview

#	sum_Value_Assault_level_2_weapon_or_bodily_harm	#	sum_Value_Assault_level
0 - 7.47k		-100 - 703	
127.33		11.33	
134.74		11.75	
145.39		13.77	
140.49		12.91	
150.5		13.95	
162.96		14.85	
175.4		16.8	
192.61		16.1	
195.7		18.44	
195.79		18.74	
182.87		18.51	
167.61		17.2	
163.16		17.68	
165.1		18.12	
161.54		16.81	
161.97		16.58	
178.6		16.56	
175.74		14.7	
0		0	
-5.1		-9.62	
5.64		3.25	

There were many mismatched values but this is because the column is expecting integers, not float values. The total row count came to 532 as shown below:

261 Columns532 Rows3 Data Types

The number of violations were 257:



This initially didn't make sense because $87,618 \text{ rows} / 257 \text{ categories} = 340$, not 532. But analyzing the data again, I found that some crimes did not exist in later years. Thus, it's not an equal distribution of crimes for each year, since for example 1999 may have just 150 categories, compared to 2015 where we have 250. This means we can't evenly divide the data, but the statistic makes sense.

By STA

< Recipe Edit Step ×

Choose a transformation

pivot

Creates a new column for each unique value in a column and aggregates the result

Column required

STA

×

Choose column

Functions required

sum(Value)

×

Edit formula

Group by

Ref_Date

×

GEO

×

Geographical_classification

×

VIOLATIONS

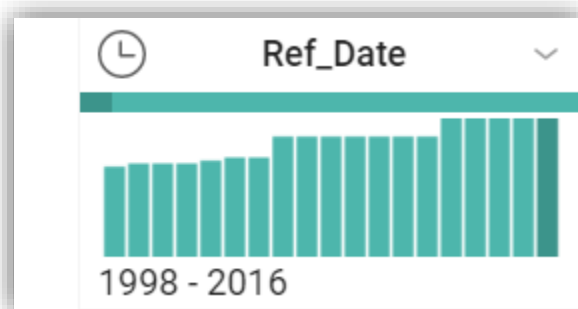
×

Choose column

Max columns to create

50

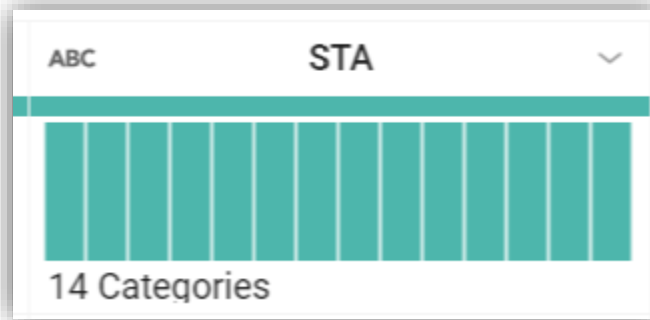
↕



I thought this statistic was strange at first, but it turns out that they add more types of crimes as years go by.

18 Columns 6,250 Rows 4 Data Types

The number of rows make sense, since there are 14 possible values in STA:



$87,618 \text{ rows} / 14 \text{ categories} = \sim 6250 \text{ rows}$

Calgary Communities Crime Data Set

The original file contains the category name, community, months and years, and was taken from this website:

<http://www.calgary.ca/cps/Documents/statistical-reports/2017%20Community%20Crime%20Statistics.xls>

FileHomeInsertDrawPage LayoutFormulasDataReviewViewTell me what you want to do

CutCopyFormat PainterClipboardFontAlignmentNumber

General\$ % & # 00 00 00

Conditional FormattingFormat as Table

NormalBadGoodNeutralCalculationCheck Cell

InsertDeleteFormatAutoSumFillClearSort & Find & FilterSelect

CellsEditing

B14Assault (Non-domestic)

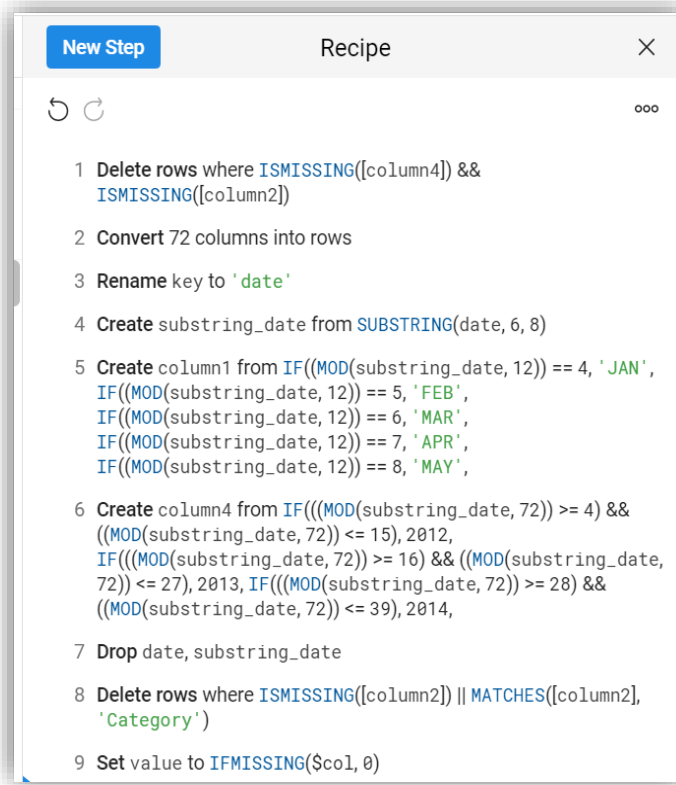
	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	AA	AB	AC	AD	AE	AF	AG	AH	AI	AJ	AK	AL	AM	
1	Category	CommunityName	2012*												2013												2014													
2			JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP	OCT	NOV	DEC	JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP	OCT	NOV	DEC	JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP	OCT	NOV	DEC	JAN	
7	ABBEYDALE	Residential Break & Enter	2	2	1	1	3	1	1	3	2	3	2	3	1	1	1	4	4	4	5	2	5	1	1	1	1	1	1	3	6	5	2	2	6	2	2	2	5	1
8	ABBEYDALE	Commercial Break & Enter																																						
9	ABBEYDALE	Theft Of Vehicle			1	7	1	5			7	2	3	4	2	3	1	4	3	4	1	3	1	2	1	2	1	2	1	1	2	3	6	3	6	3	2	2	1	1
10	ABBEYDALE	Theft From Vehicle	1	2	1	8	2		4	1	3	2	3	5	3	1	3	4	1	4	3	3	7	2	1	4	1	1	6	5	8	6	9	4	5	2	3			
11	ABBEYDALE	Social Disorder	29	22	21	29	28	33	47	51	50	22	10	18	26	37	29	28	46	43	43	62	41	33	28	24	30	29	30	31	47	36	70	54	34	33	20	29		
12	ABBEYDALE	Physical Disorder	4	1	5	3	3		2	3	5	1	4	2	3	2	6	3	3	3	1	5	2	5	3	3	4	2	2	3	5	2	3	5	3	3				
13																																								
14	ACADIA	Assault (Non-domestic)	3	2	3	2	4	3	1	3	4	5	4	1	1	2	4	1	8	2	2	3	2	2	2	3	5	2	1	3	3	1	3	2	1	3	2	1		
15	ACADIA	Commercial Robbery									2						1																							
16	ACADIA	Street Robbery						4				1						1					2	1		2			1									1	1	
17	ACADIA	Violence 'Other' (Non-domestic)	1	2	2				2	1	1	1	2		2	2		1	1	1		6	2	3		2		2	1		2		2	2	1	2				
18	ACADIA	Residential Break & Enter	1	1	1	2	2	10	6	3	1	3	4	5	1	2	4	4	1	3	2	2	4	8	6	5	4	4	5	3	6	4	2	2	3					
19	ACADIA	Commercial Break & Enter	2	2	2					1	1							1	2	2		1	3	2	2	1	1	2	2	1	3									
20	ACADIA	Theft Of Vehicle	1	2	3	3	1	1		5	3	4	1	2	3	6	4	2	4	4	1	5	1	2	2	3	3	5	6	5	4	4	2	2	3					
21	ACADIA	Theft From Vehicle	2	4	9	8	7	19	11	12	6	9	5	12	4	5	5	2	9	10	20	5	12	5	3	7	6	8	2	3	6	6	5	11	2					
22	ACADIA	Social Disorder	65	78	91	66	87	86	97	90	88	78	73	78	52	53	52	68	80	101	71	98	73	70	57	71	77	46	52	84	79	91	115	91	120	107	98	68		
23	ACADIA	Physical Disorder	5	8	5	3	8	3	6	4	8	2	6	7	8	8	4	8	4	12	7	8	3	8	5	6	8	3	6	4	5	6	4	9	7	7	1	5		
24																																								
25	ALBERT PARK/RADISSON HEIGHTS	Assault (Non-domestic)			5	12	5	4	4	3	9	5	3		2	1	6	3	7	5	5	7	5	3	2	5	6	5	4	7	1	5	7	1	6	2	2			
26	ALBERT PARK/RADISSON HEIGHTS	Commercial Robbery																																						
27	ALBERT PARK/RADISSON HEIGHTS	Street Robbery	1	1	1			1	1	1							1	1	1	1	1	1	1	1	1	2		1	1	1		1	1		1					
28	ALBERT PARK/RADISSON HEIGHTS	Violence 'Other' (Non-domestic)	3	1	2	1		2	4	3					2	1	2	1	2	2	1	1	1			1		1	3	1					1	1	2			
29	ALBERT PARK/RADISSON HEIGHTS	Residential Break & Enter	2	3	16	3	1	1	4	2	3	1	1	3	3	1	1	4	2	4	5	3	6	5		1	3	3	3	4	1	2	3	3	3	1				
30	ALBERT PARK/RADISSON HEIGHTS	Commercial Break & Enter	1	2	3						2	1					1							1	2	2	3	2	3	1	1	2					1	2		
31	ALBERT PARK/RADISSON HEIGHTS	Theft Of Vehicle	4	3	8	5	5	2	3	1	4	2	4	2	1	3	2	6	3	6	2	1	1	6	4	3	4	4	4	7	6	12	6	8	4	4				
32	ALBERT PARK/RADISSON HEIGHTS	Theft From Vehicle	5	23	9	5	5	2	12	15	6	8	5	2	7	16	2	6	3	7	6	8	3	2	6	4	2	1	6	12	3	7	8	8	2	7				
33	ALBERT PARK/RADISSON HEIGHTS	Social Disorder	60	59	63	63	99	85	106	104	94	63	60	46	47	57	89	78	82	138	114	87	81	67	60	67	66	55	74	103	94	117	95	88	92	61	59			
34	ALBERT PARK/RADISSON HEIGHTS	Physical Disorder	1	7	16	13	10	4	6	7	9	8	2	5	6	3	6	4	8	6	7	6	4	11	1	5	9	7	4	9	8	3	4	7	10	6	6			
35																																								
36	ALTADORE	Assault (Non-domestic)			1	2	2	2	1	1	3	3	1		1	1						2		1												1	3			
37	ALTADORE	Commercial Robbery																																						
38	ALTADORE	Street Robbery	1																										1		1		1							
39	ALTADORE	Violence 'Other' (Non-domestic)							1							2		1				1	1	1		1						1	1	1		1	1			
40	ALTADORE	Residential Break & Enter	1	6	11	3	1	8		2	2	7	12	6		6	4	4	6	4	5	4	8	4	2	2	3	1	4	5	6	6	5	5	2	5	5			
41	ALTADORE	Commercial Break & Enter	2			2	1		1	8	6	2	2	1	1	2	1					2	3			3		3		1	1	1	1			1	5			
42	ALTADORE	Theft Of Vehicle	3	3	4	1	2	6	3	2	3	2		3	2	3		1	2	1	2	2	1	2	1		2	1	2	2	1	1	4	1		2	1			
43	ALTADORE	Theft From Vehicle	1	7	8	5	14	9	2	8	14	8	8	1	19	12	5	4	3	9	6	12	6	7	8	15	3	6	2	4	2	6	14	10	3	7	4			
44	ALTADORE	Physical Disorder	14	18	17	33	31	31	35	30	22	18	16	25	19	32	20	26	39	45	20	44	35	29	19	24	19	12	18	22	28	39	41	24	19	39	29	16		

By CommunityBy CategoryData Notes

Figure 6 - Calgary community crime data set

Cleansing, Augmenting, Profiling the Data

I initially had trouble figuring out how to wrangle the data, because I had to unpivot 5 large columns(2012 – 2017) and then further unpivot it for each month from January to December.



For Step 2 - I merged all the date columns into one. Each year has 12 months, and since the data spans from 2012 to 2017 we have $5 \times 12 = 72$ columns. When unpivoted, the values ranged from `column4` to `column75`.

ABC	column2	ABC	column3	2	date	#	substring_date	ABC	value
290 Categories	10 Categories		2012 - 2017		Jan 1970 - Dec 1970		9 Categories		
Category	CommunityName		column4	4				2012*	
Category	CommunityName		column5	5					
Category	CommunityName		column6	6					
Category	CommunityName		column7	7					
Category	CommunityName		column8	8					
Category	CommunityName		column9	9					
Category	CommunityName		column10	10					
Category	CommunityName		column11	11					
Category	CommunityName		column12	12					
Category	CommunityName		column13	13					
Category	CommunityName		column14	14					
Category	CommunityName		column15	15					
Category	CommunityName		column16	16				2013	
Category	CommunityName		column17	17					
Category	CommunityName		column18	18					
Category	CommunityName		column19	19					
Category	CommunityName		column20	20					
Category	CommunityName		column21	21					
Category	CommunityName		column22	22					
Category	CommunityName		column23	23					
Category	CommunityName		column24	24					
Category	CommunityName		column25	25					
Category	CommunityName		column26	26					
Category	CommunityName		column27	27					
Category	CommunityName		column28	28				2014	
Category	CommunityName		column29	29					
Category	CommunityName		column30	30					

Figure 8- Step 2 of the recipe, date column is the result of unpivoting columns 4 to 75

Step 5 - I decided to manually code it such that it would take the integer value from the newly made column1 (which I eventually renamed to date) created from unpivoting the columns into rows, and then take the mod value such that if an integer mod 12 is equal to a value, it would assign values from 'JAN' to 'DEC'.

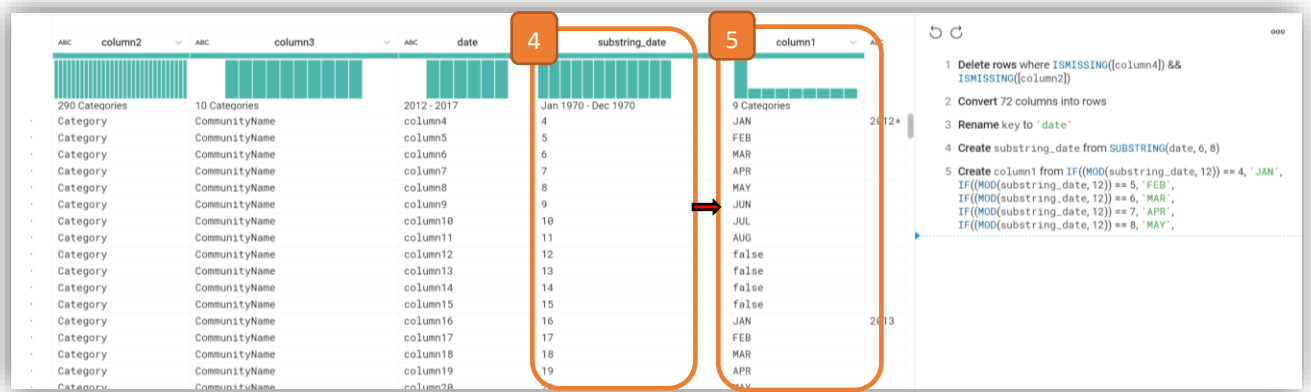


Figure 9 - Steps 4 and 5 from the recipe on the righthand panel

Step 6 - I made another column and set ranges for each year, for example for an integer mod 12, if the value is at a certain range then it will assign values from 2012 to 2017.

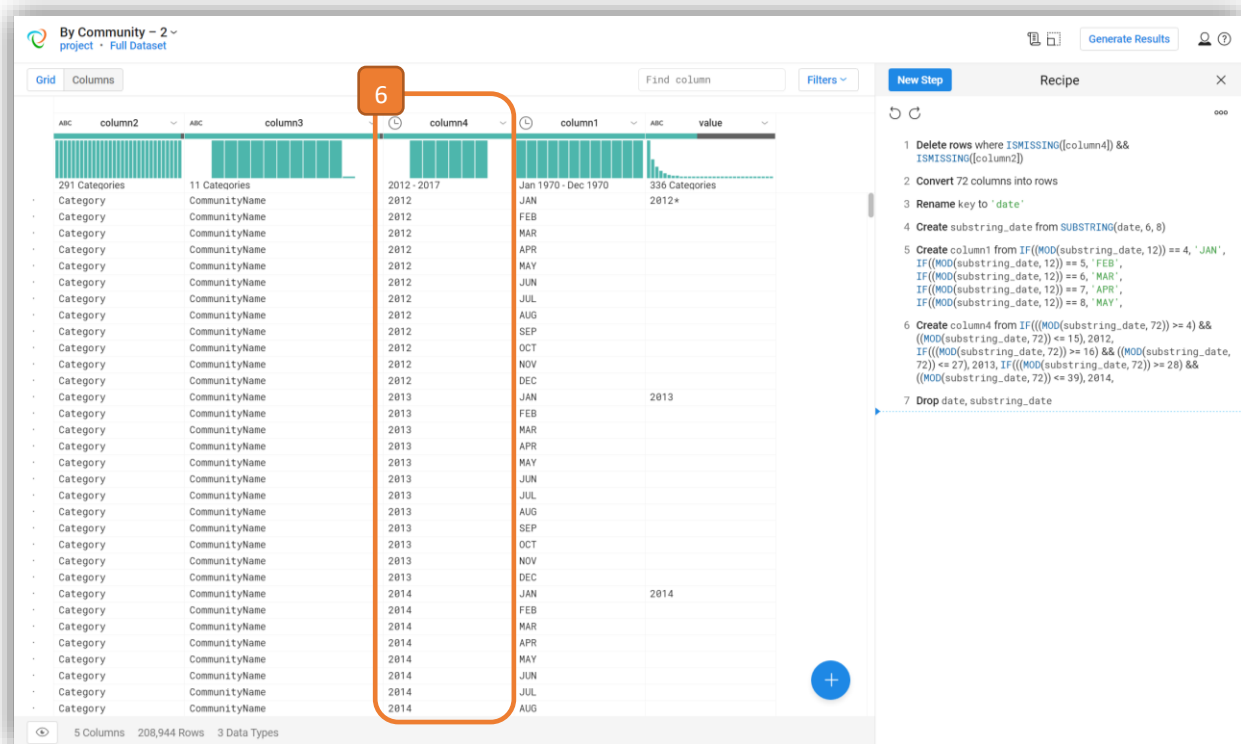


Figure 10 - Step 6 - based on a range, assign a year

Step 8 - I also had to delete some rows in the beginning because it was pivoted against essentially two nested columns being unpivoted.

After all those steps, this is the result:

By Community - 2
project - Full Dataset

Grid Columns Find column Filters

New Step Recipe

1 Delete rows where ISMISSING([column4]) && ISMISSING([column2])

2 Convert 72 columns into rows

3 Rename key to 'date'

4 Create substring_date from SUBSTRING(date, 6, 8)

5 Create column1 from IF((MOD(substring_date, 12)) == 4, 'JAN', IF((MOD(substring_date, 12)) == 5, 'FEB', IF((MOD(substring_date, 12)) == 6, 'MAR', IF((MOD(substring_date, 12)) == 7, 'APR', IF((MOD(substring_date, 12)) == 8, 'MAY', IF((MOD(substring_date, 12)) == 9, 'JUN', IF((MOD(substring_date, 12)) == 10, 'JUL', IF((MOD(substring_date, 12)) == 11, 'AUG', IF((MOD(substring_date, 12)) == 12, 'SEP', IF((MOD(substring_date, 12)) == 1, 'OCT', IF((MOD(substring_date, 12)) == 2, 'NOV', IF((MOD(substring_date, 12)) == 3, 'DEC', ''))))))))))

6 Create column4 from IF((MOD(substring_date, 72)) >= 4) && ((MOD(substring_date, 72)) <= 15), 2012, IF((MOD(substring_date, 72)) >= 16) && ((MOD(substring_date, 72)) <= 27), 2013, IF((MOD(substring_date, 72)) >= 28) && ((MOD(substring_date, 72)) <= 39), 2014, ''

7 Drop date, substring_date

8 Delete rows where ISMISSING([column2]) || MATCHES([column2], 'Category')

9 Set value to IFMISSING(\$col1, 0)

5 Columns 208,800 Rows 4 Data Types

Figure 11 - Final result after all the steps of the recipe

Calgary Shape File

This data set was taken from:

<https://data.calgary.ca/Base-Maps/Community-Boundaries/ab7m-fwn6>

Using Tableau, I joined together the community file and the shape file, to see what the shape file looks like:

By Community – 2 (By Community – 2)

By Community – 2



geo_export_226f4aa8-1601-4...

Figure 12 – Inner join with Community (from Community Crime Data Set) and Name column in shape file

By Community – 2	By Community – 2	By Community – 2	By Community – 2	By Community – 2	geo_export_226f4aa8-1601-4...	geo_export_226f4aa8-1601-4...	geo_export_226f4aa8-1601-4...	geo_export_226f4aa8-1601-4...	geo_export_226f4aa8-1601-4...	geo_export_226f4aa8-1601-4...	geo_export_226f4aa8-1601-4...	geo_export_226f4aa8-1601-4...
Community	Crime	Year	Month	Value	Comm Struc	Class	Class Code	Sector	Name	Comm Code	Srg	Geometry
SUNALTA	Physical Disor...	2017	DEC		4	INNER CITY	Residential	1.00	CENTRE	SUNALTA	SNA	BUILT-OUT POLYGON
SUNALTA	Physical Disor...	2017	NOV		3	INNER CITY	Residential	1.00	CENTRE	SUNALTA	SNA	BUILT-OUT POLYGON
SUNALTA	Physical Disor...	2017	OCT		8	INNER CITY	Residential	1.00	CENTRE	SUNALTA	SNA	BUILT-OUT POLYGON
SUNALTA	Physical Disor...	2017	SEP		1	INNER CITY	Residential	1.00	CENTRE	SUNALTA	SNA	BUILT-OUT POLYGON
SUNALTA	Physical Disor...	2017	AUG		5	INNER CITY	Residential	1.00	CENTRE	SUNALTA	SNA	BUILT-OUT POLYGON
SUNALTA	Physical Disor...	2017	JUL		5	INNER CITY	Residential	1.00	CENTRE	SUNALTA	SNA	BUILT-OUT POLYGON

Figure 13 - What the shape file contains

Using the shape file, I was able to retrieve a lot more attributes like the structure, sector, and class, dimensions which may later be useful for my analysis.

Decisions

A problem that I struggled with was deciding to join these data sets (Alberta Crime Data Set and Calgary Community Crime Data Set) or not. I opted not to join these because of the following reasons:

1. I only have community info for Calgary, since I'm not interested in other municipality communities. Thus it would be a left join with a lot of null values for other municipalities, and would be hard to use for visualization since Calgary values would be bloated.
2. Violations counts are different. We only have 10 different violation categories for the Calgary Community Crime Data Set, whereas the Alberta Crime Data set has 257. There would not be a clean mapping for this.

Research Questions

1. What qualities of communities more prone to crime in Calgary? (older communities, quadrants)
2. What crime is the most prevalent in each city? Are there any significant differences in crime between municipalities?
3. What is the ratio of adult vs youth committing crime, on average? Does it change drastically depending on the crime?

4. What are the top 2 crimes that are strongly correlated with one another? Which crimes can predict a violent crime such as first-degree murder?
5. Is there a month or year that crime rises in Calgary? (e.g. before Christmas, after New Year's where debt is high, summer where the weather is good, recessions)
6. Is Calgary becoming a more dangerous city? Based on the answers to the questions above, and based on the crime statistics alone, would you decide to live in Calgary and if so where?

Backup Questions

7. Is there a relationship between violent crime and non-violent crime? Which non-violent crimes can estimate violent crime?
8. Is there a big difference of actual crime incidents committed versus total charges of these that were cleared?
9. Which municipality is most lenient towards their youth committing crimes?
10. What crimes have experienced significant deviances over the years in Calgary? List at least three.