

SECTION 5

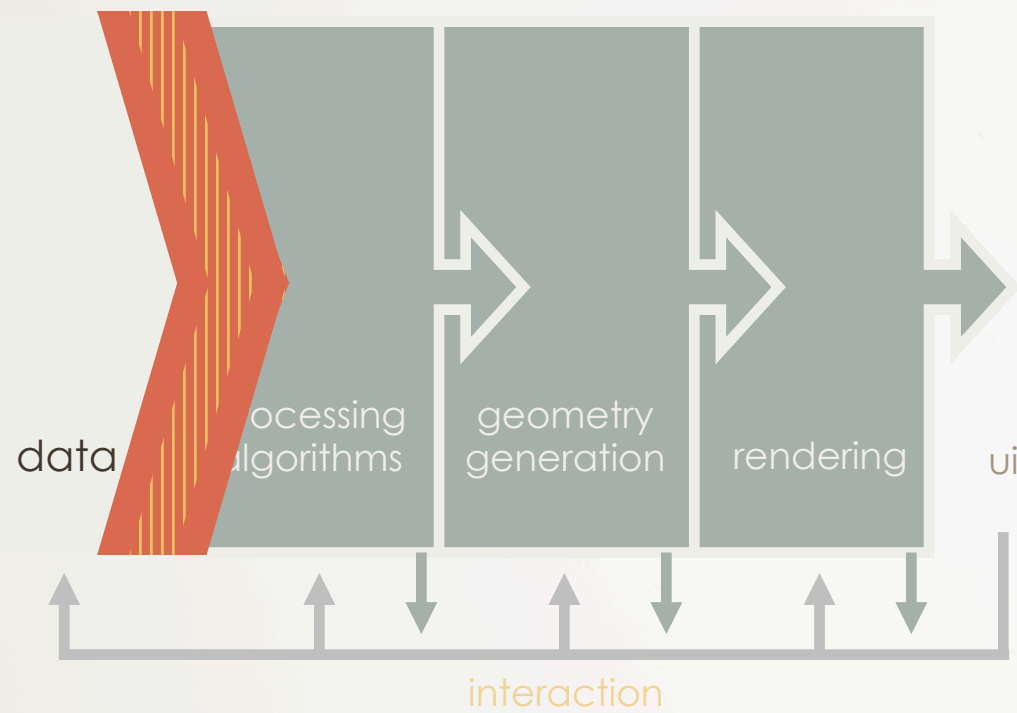
addressing bottlenecks



CENTER FOR DATA-DRIVEN DISCOVERY

BOTTLENECK

V



- data size
- data format

DATA WRANGLER

smarter text manipulation



CENTER FOR DATA-DRIVEN DISCOVERY

CONTACT US

ABOUT US

MOST WANTED

NEWS

STATS & SERVICES

SCAMS & SAFETY

JOBS

FUN & GAMES

Uniform Crime Reports

Select Language

Get FBI Updates

Criminal Justice Information Services Division

Feedback

Contact Us

Data Quality Guidelines

UCR Home

CIUS Home

Offenses Known to Law Enforcement

Violent Crime

Property Crime

Clearances

Persons Arrested

Police Employee Data

About CIUS

Table 4

Crime in the United States

by Region, Geographic Division, and State, 2010–2011

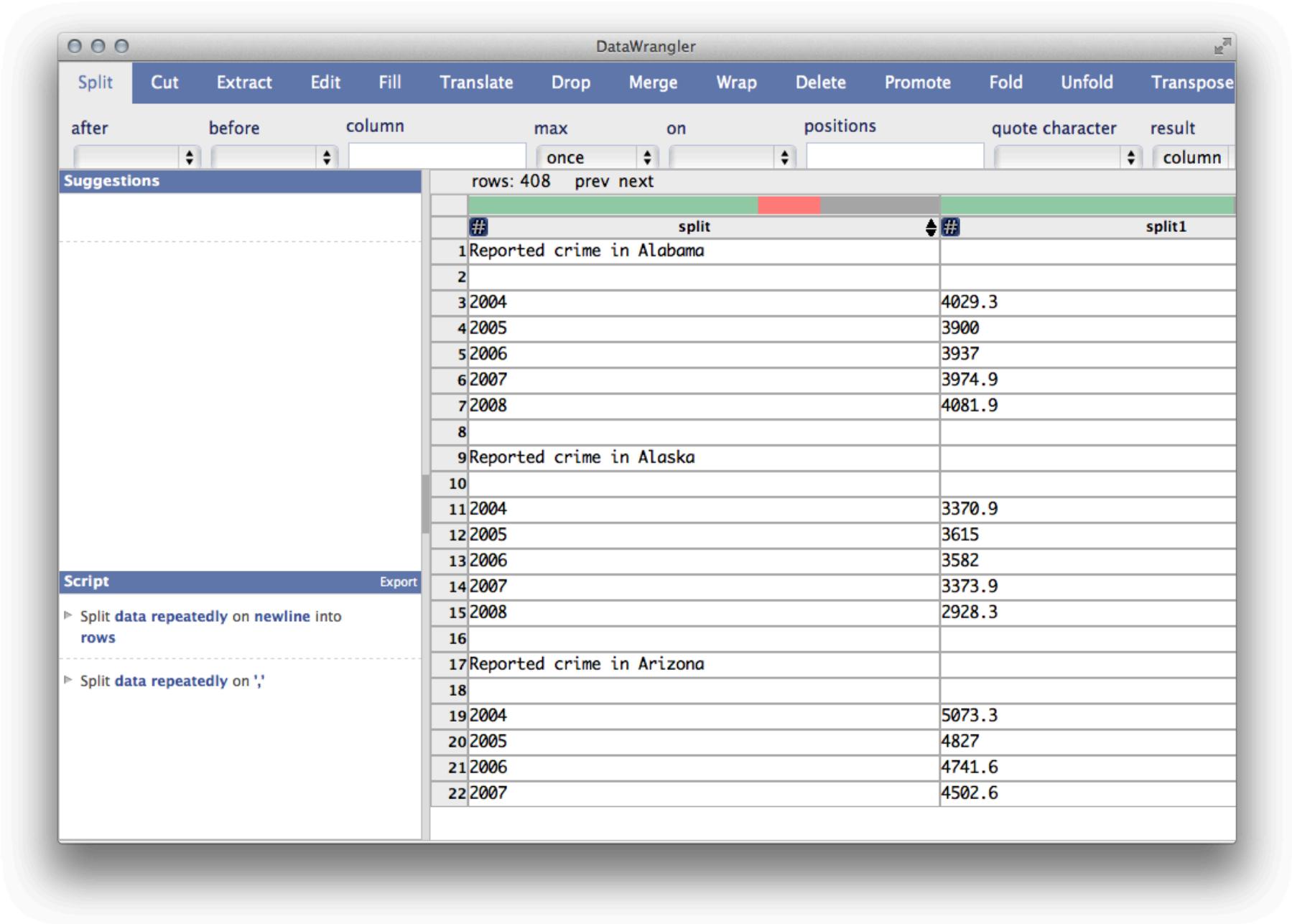
Overview

Data Declaration

Download Excel

Area	Year	Population ¹	Violent crime		Murder and nonnegligent manslaughter		Forcible rape		Robbery		Aggravated assault		Property	
			Number	Rate per 100,000	Number	Rate per 100,000	Number	Rate per 100,000	Number	Rate per 100,000	Number	Rate per 100,000	Number	Rate per 100,000
United States Total ^{2, 3, 4, 5}	2010	309,330,219	1,251,248	404.5	14,722	4.8	85,593	27.7	369,089	119.3	781,844	252.8	9,112,625	
	2011	311,591,917	1,203,564	386.3	14,612	4.7	83,425	26.8	354,396	113.7	751,131	241.1	9,063,173	
	Percent change		-3.8	-4.5	-0.7	-1.5	-2.5	-3.2	-4.0	-4.7	-3.9	-4.6	-0.5	
Northeast	2010	55,366,108	198,333	358.2	2,312	4.2	10,868	19.6	68,993	124.6	116,160	209.8	1,175,122	
	2011	55,521,598	195,482	352.1	2,169	3.9	10,641	19.2	68,855	124.0	113,817	205.0	1,178,072	
	Percent change		-1.4	-1.7	-6.2	-6.4	-2.1	-2.4	-0.2	-0.5	-2.0	-2.3	+0.3	
	2010	14,453,587	48,174	333.3	420	2.9	3,618	25.0	12,170	84.2	31,966	221.2	336,004	

DataWrangler		
Browse ► Crime Rates by State 2004-2008 http://www.fbi.gov/about-us/cjis/ucr/crime-in-the-u.s/2011/crime-in-the-u.s.-2011/tables/table-4/output.csv Wrangle		
	<pre> 000001 Reported crime in Alabama, 000002 . 000003 2004,4029.3 000004 2005,3900 000005 2006,3937 000006 2007,3974.9 000007 2008,4081.9 000008 . 000009 Reported crime in Alaska, 000010 . 000011 2004,3370.9 000012 2005,3615 000013 2006,3582 000014 2007,3373.9 000015 2008,2928.3 000016 . 000017 Reported crime in Arizona, 000018 . 000019 2004,5073.3 000020 2005,4827 000021 2006,4741.6 000022 2007,4502.6 000023 2008,4087.3 000024 . 000025 Reported crime in Arkansas, 000026 . 000027 2004,4033.1 000028 2005,4068 000029 2006,4021.6 000030 2007,3945.5 000031 2008,3843.7 000032 . 000033 Reported crime in California, 000034 . 000035 2004,3423.9 000036 2005,3321 000037 2006,3175.2 000038 2007,3032.6 000039 2008,2940.3 000040 . 000041 Reported crime in Colorado, 000042 . 000043 2004,3918.5 000044 2005,4041 000045 2006,3441.8 000046 2007,2991.3 000047 2008,2856.7 000048 . 000049 Reported crime in Connecticut, 000050 . 000051 2004,2684.9 000052 2005,2579 000053 2006,2575 000054 2007,2470.6 000055 2008,2490.8 ... </pre>	<pre> ... 000353 Reported crime in Utah, 000354 . 000355 2004,4038.6 000356 2005,3837 000357 2006,3505.3 000358 2007,3510.1 000359 2008,3374 000360 . 000361 Reported crime in Vermont, 000362 . 000363 2004,2343.6 000364 2005,2370 000365 2006,2368 000366 2007,2339 000367 2008,2559.8 000368 . 000369 Reported crime in Virginia, 000370 . 000371 2004,2678.2 000372 2005,2649 000373 2006,2479.6 000374 2007,2480 000375 2008,2531.8 000376 . 000377 Reported crime in Washington, 000378 . 000379 2004,4846.7 000380 2005,4890 000381 2006,4483.3 000382 2007,4026 000383 2008,3785 000384 . 000385 Reported crime in West Virginia, 000386 . 000387 2004,2555.8 000388 2005,2633 000389 2006,2639.9 000390 2007,2543.6 000391 2008,2554.4 000392 . 000393 Reported crime in Wisconsin, 000394 . 000395 2004,2665.7 000396 2005,2669 000397 2006,2820 000398 2007,2842.8 000399 2008,2761.1 000400 . 000401 Reported crime in Wyoming, 000402 . 000403 2004,3338.5 000404 2005,3158 000405 2006,2989.1 000406 2007,2883.2 000407 2008,2724.2 </pre>



Split

Cut

Extract

Edit

Fill

Translate

Drop

Merge

Wrap

Delete

Promote

Fold

Unfold

Transpose

after

before

column

max

on

positions

quote character

result

in

[a-zA-Z]+

split

once

.*

column

Suggestions

Split **split** between positions 17, 18

Split **split** between 'in' and 'Alabama'

Split **split** between 'in' and 'any word'

Split **split** between 'in' and 'Alabama'

Split **split** between 'in' and 'Alabama'

Split **split** between 'in' and 'any word'

Extract from **split** between positions 17, 18

Script

Split data repeatedly on newline into rows

Split data repeatedly on ','

Split **split** between 'in' and 'any word'

Export

rows: 408

prev

next

#	split	split2	split3	#
1	Reported crime in Alabama	Reported crime in	Alabama	
2				
3	2004	2004		4029.3
4	2005	2005		3900
5	2006	2006		3937
6	2007	2007		3974.9
7	2008	2008		4081.9
8				
9	Reported crime in Alaska	Reported crime in	Alaska	
10				
11	2004	2004		3370.9
12	2005	2005		3615
13	2006	2006		3582
14	2007	2007		3373.9
15	2008	2008		2928.3
16				
17	Reported crime in Arizona	Reported crime in	Arizona	
18				
19	2004	2004		5073.3
20	2005	2005		4827
21	2006	2006		4741.6
22	2007	2007		4502.6

row

row is empty

Suggestions

Delete row 2

Delete empty rows

Delete rows where split2 is null

Delete rows where split3 is null

Delete rows where split1 is null

Fold using 2 as a key

Fold split2 using 2 as a key

Script

Export

Split data repeatedly on newline into rows

Split data repeatedly on ','

Split split between 'in' and ' any word '

Split split between 'in' and ' any word '

rows: 408 prev next

#	split2	split3	#	split3
1	Reported crime in	Alabama		
2				
3	2004			4029.3
4	2005			3900
5	2006			3937
6	2007			3974.9
7	2008			4081.9
8				
9	Reported crime in	Alaska		
10				
11	2004			3370.9
12	2005			3615
13	2006			3582
14	2007			3373.9
15	2008			2928.3
16				
17	Reported crime in	Arizona		
18				
19	2004			5073.3
20	2005			4827
21	2006			4741.6
22	2007			4502.6

Split

Cut

Extract

Edit

Fill

Translate

Drop

Merge

Wrap

Delete

Promote

Fold

Unfold

Transpose

column

direction

row

split3

above

Suggestions

Fill **split3** with values from **above**

Fill **split3** with values from **below**

Fill **split3** with values from **the left**

Drop **split3**

Fold **split3** using **header** as a key

Fold **split3** using **1** as a key

Fold **split3** using **1, 2** as keys

Script

rows

Split **data** repeatedly on **'**

Split **split** between **'in'** and **' any word '**

Split **split** between **'in'** and **' any word '**

Delete **empty rows**

Export

rows: 306 prev next

#	split2	split3	#	split
1	Reported crime in	Alabama		
2	2004	Alabama	4029.3	
3	2005	Alabama	3900	
4	2006	Alabama	3937	
5	2007	Alabama	3974.9	
6	2008	Alabama	4081.9	
7	Reported crime in	Alaska		
8	2004	Alaska	3370.9	
9	2005	Alaska	3615	
10	2006	Alaska	3582	
11	2007	Alaska	3373.9	
12	2008	Alaska	2928.3	
13	Reported crime in	Arizona		
14	2004	Arizona	5073.3	
15	2005	Arizona	4827	
16	2006	Arizona	4741.6	
17	2007	Arizona	4502.6	
18	2008	Arizona	4087.3	
19	Reported crime in	Arkansas		
20	2004	Arkansas	4033.1	
21	2005	Arkansas	4068	
22	2006	Arkansas	4021.6	

row

split2 = 'Reported cri

Suggestions

Delete row 1

Delete rows where split2 = 'Reported crime in'

Delete rows where split3 = 'Alabama'

Delete rows where split1 is null

Fill row 1 with values from the left

Wrap row 1

Fill rows where split2 = 'Reported

Script

Split data repeatedly on ','

Split split between 'in' and ' any word '

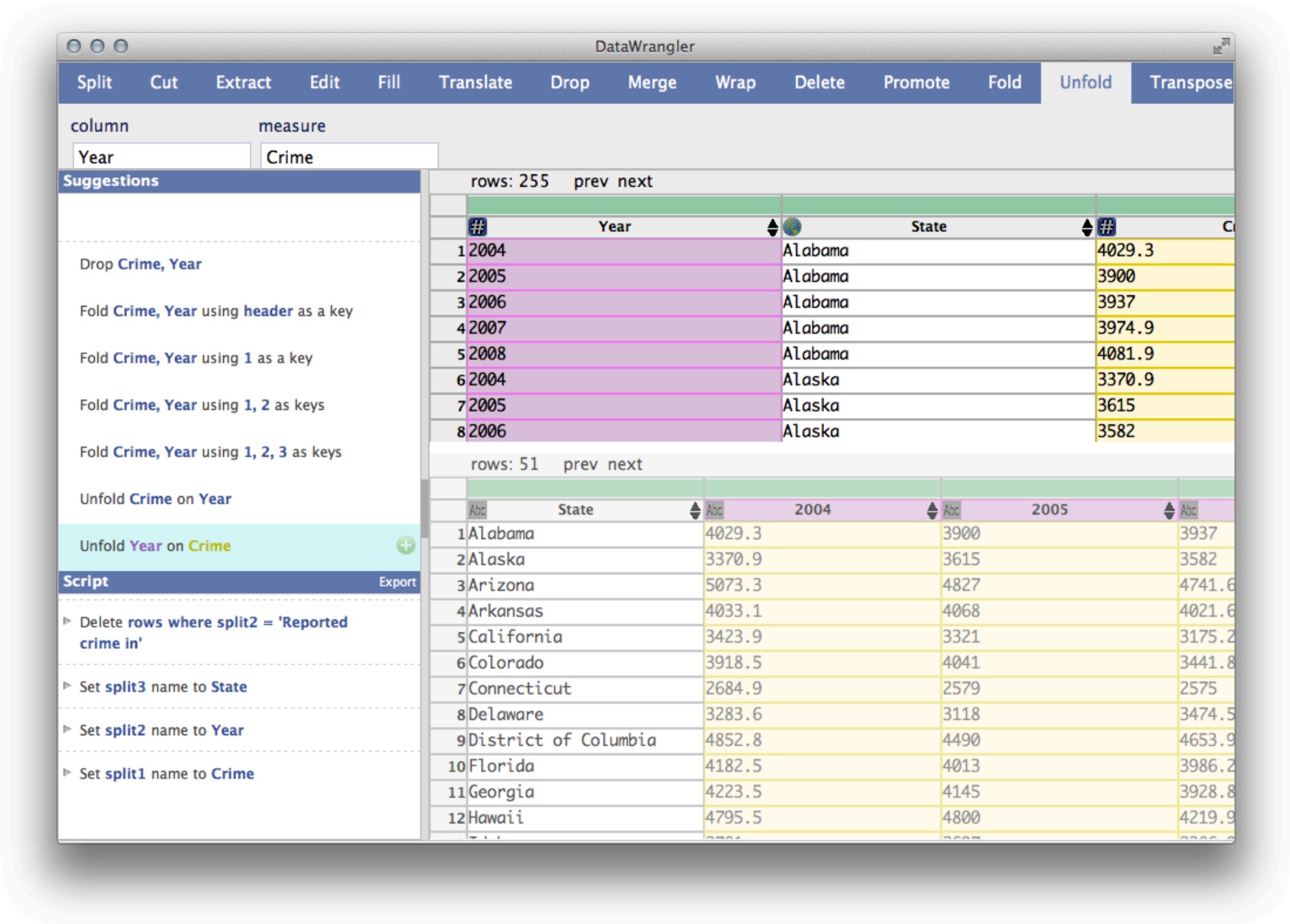
Split split between 'in' and ' any word '

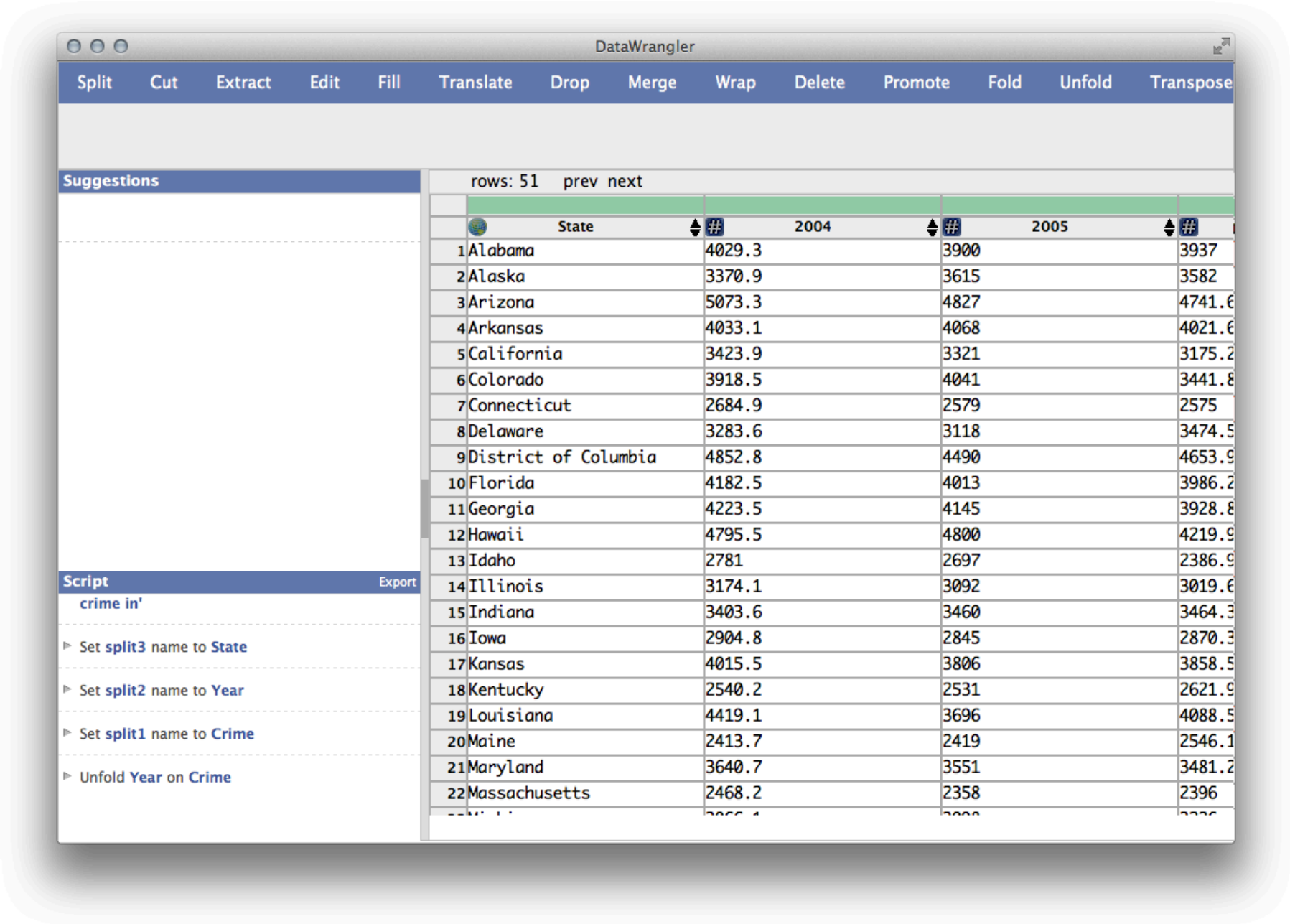
Delete empty rows

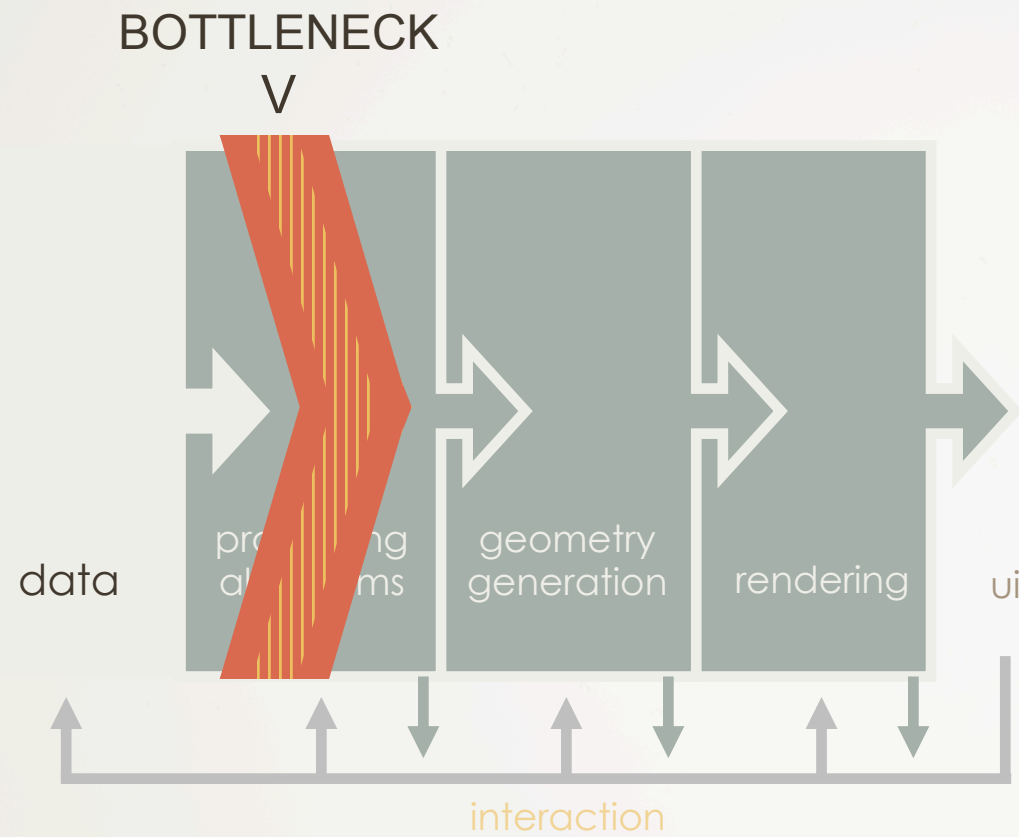
Fill split3 with values from above

rows: 306 prev next

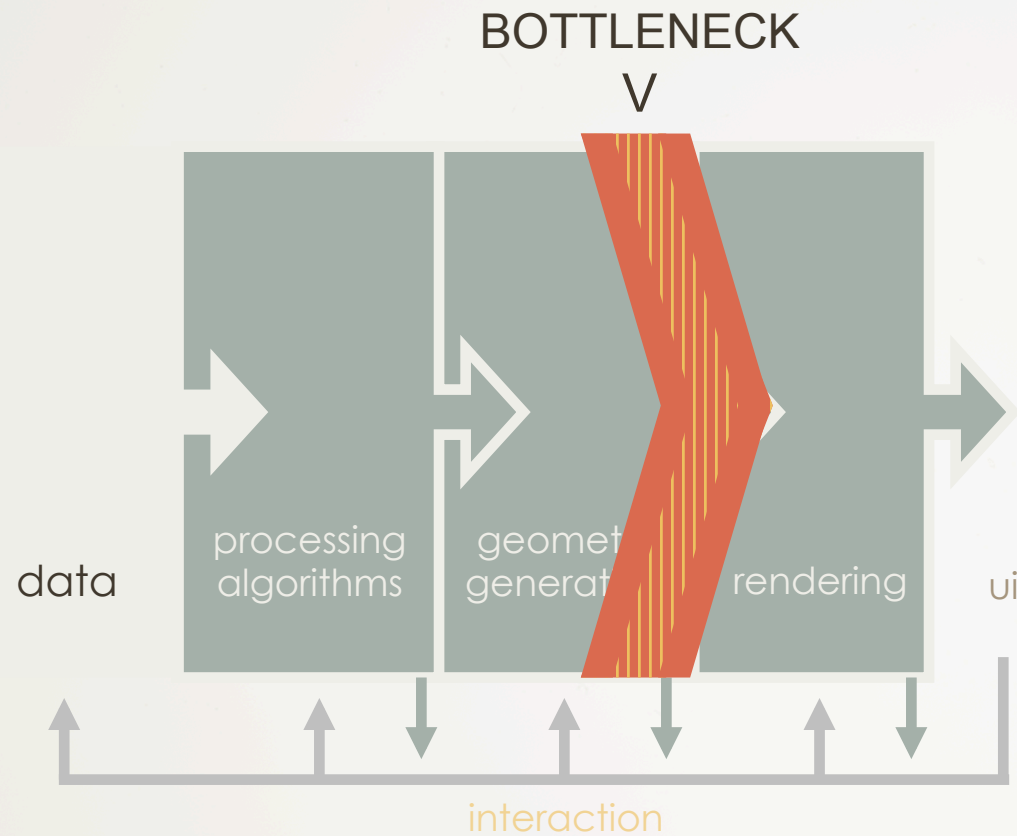
#	split2	split3	#	split1
1	Reported crime in	Alabama		
2	2004	Alabama	4029.3	
3	2005	Alabama	3900	
4	2006	Alabama	3937	
5	2007	Alabama	3974.9	
6	2008	Alabama	4081.9	
7	Reported crime in	Alaska		
8	2004	Alaska	3370.9	
9	2005	Alaska	3615	
0	2006	Alaska	3582	
1	2007	Alaska	3373.9	
2	2008	Alaska	2928.3	
3	Reported crime in	Arizona		
4	2004	Arizona	5073.3	
5	2005	Arizona	4827	
6	2006	Arizona	4741.6	
7	2007	Arizona	4502.6	
8	2008	Arizona	4087.3	
9	Reported crime in	Arkansas		
0	2004	Arkansas	4033.1	
1	2005	Arkansas	4068	
2	2006	Arkansas	4021.6	







- processing power



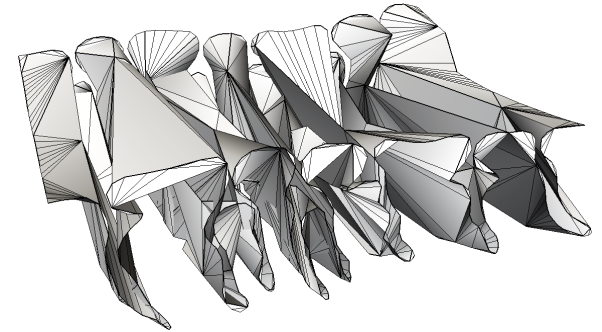
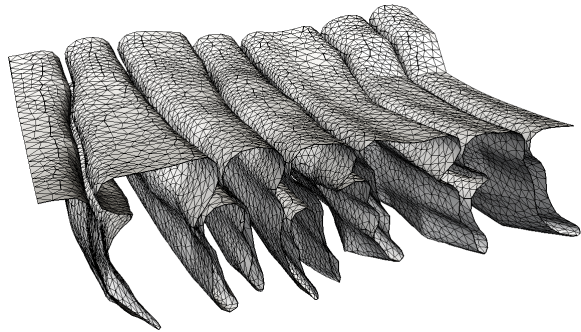
- number of records
- number of triangles
(base rendering units)
- number voxels

BIG DATA

*LOD **vs** Parallelism*



CENTER FOR DATA-DRIVEN DISCOVERY



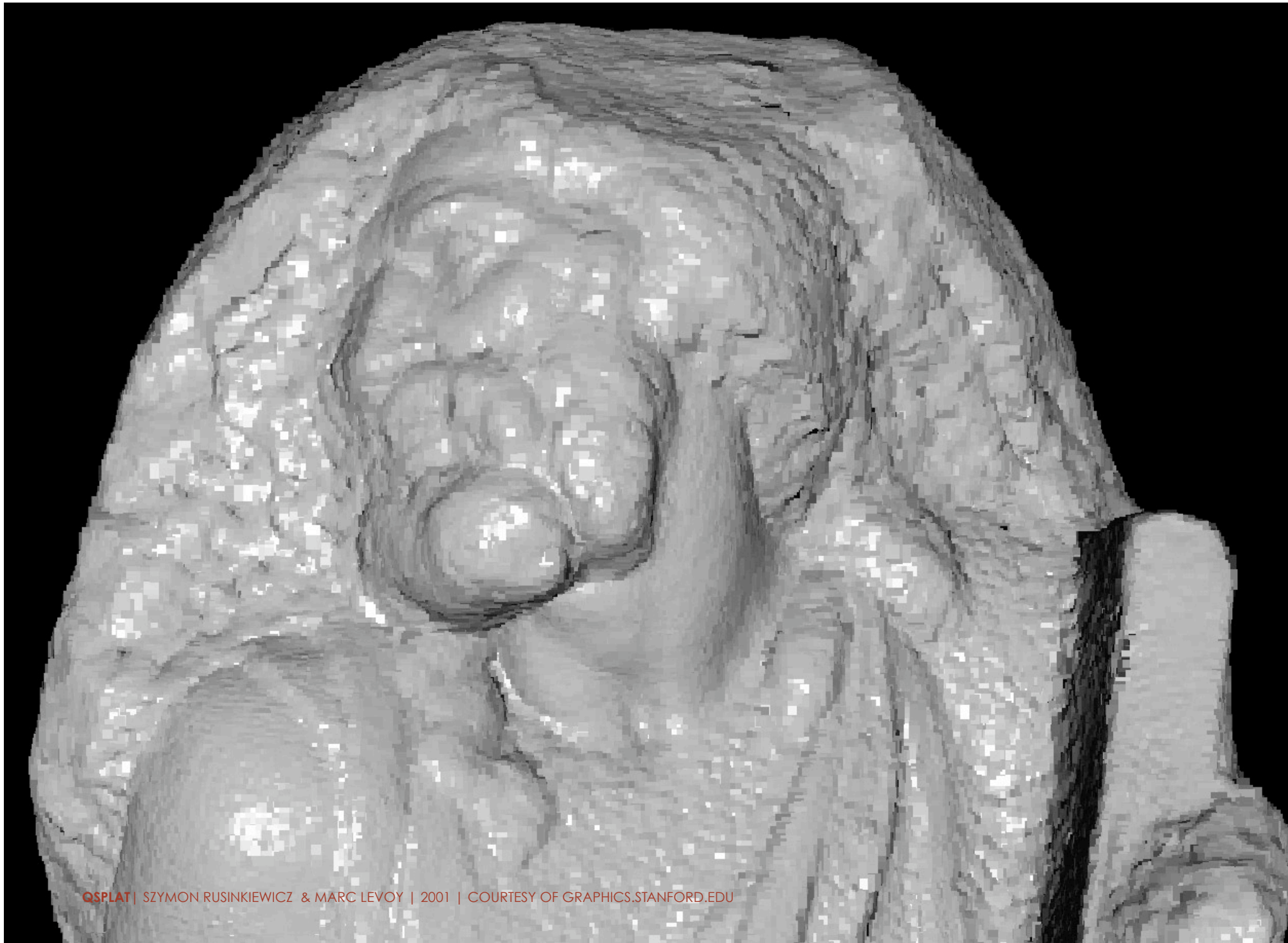
LEVELS OF DETAIL (LOD)

- requires pre-processing
- requires larger storage (*original+...*)



LEVELS OF DETAIL (LOD)

- quality
 - algorithm ↔ time
- when to use lod?
 - improve interactivity
 - smaller footprint

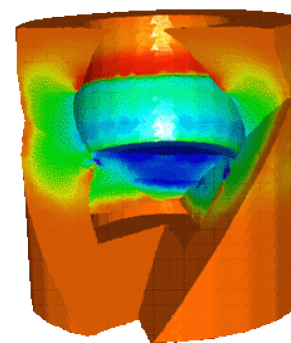
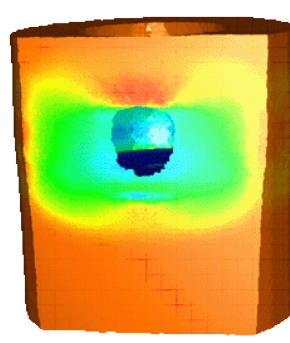
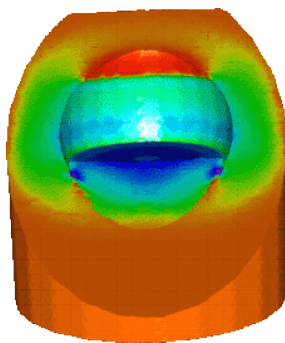
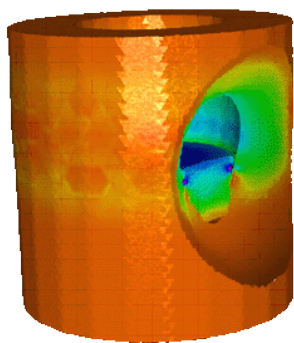


QSPLAT | SZYMON RUSINKIEWICZ & MARC LEVOY | 2001 | COURTESY OF GRAPHICS.STANFORD.EDU

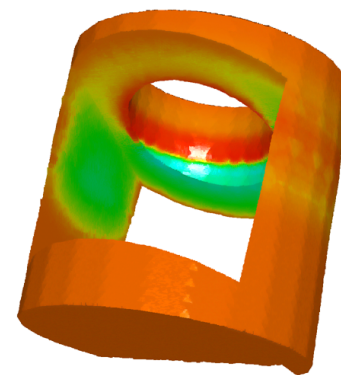
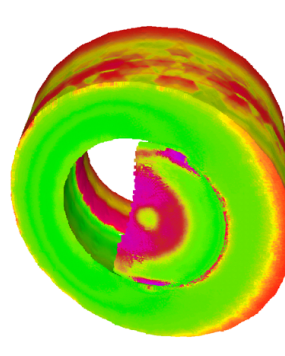
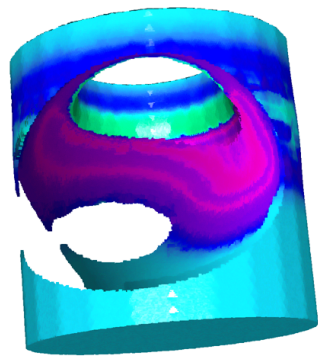
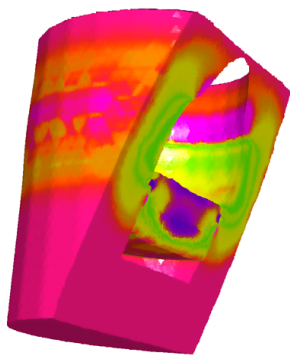


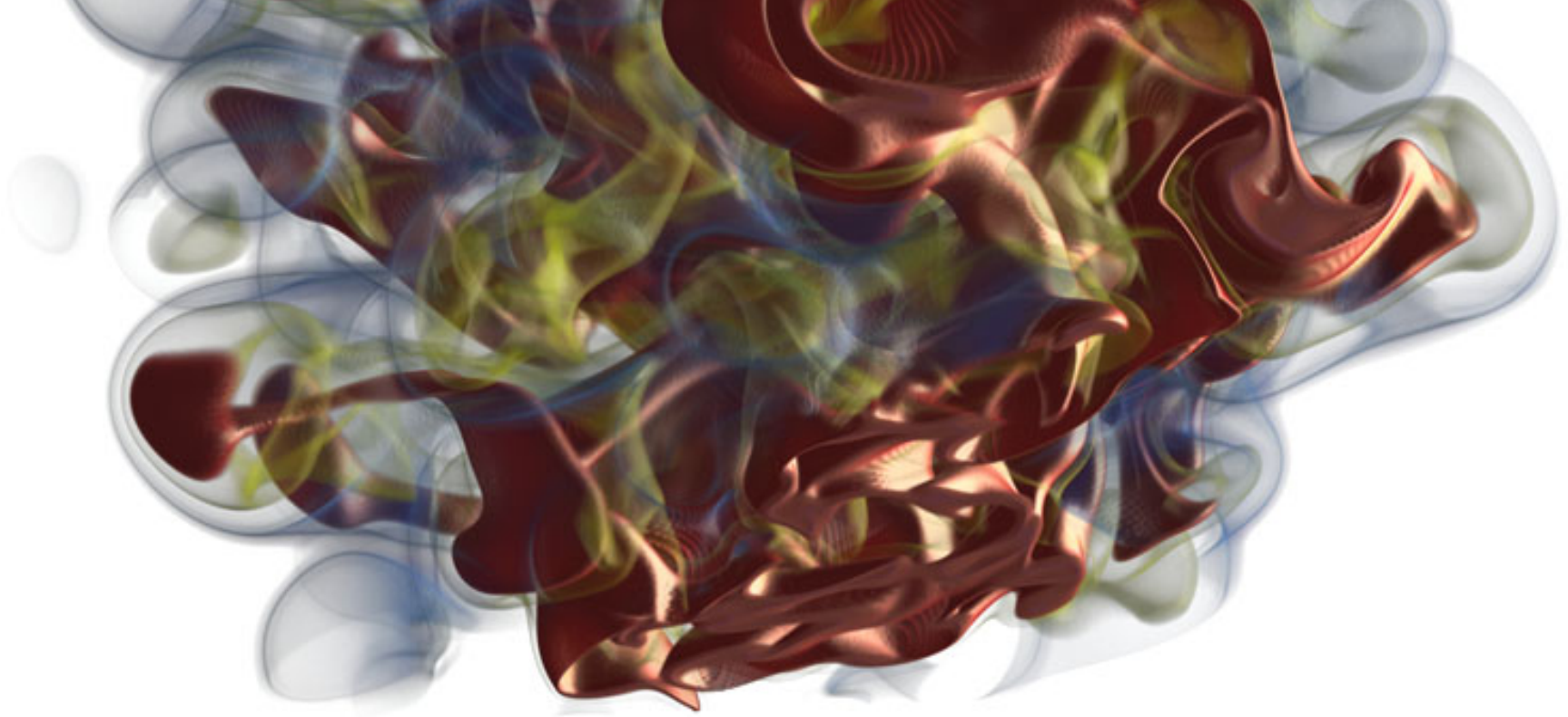






TETSPLAT | SANTIAGO LOMBEYDA & KEN MUSETH | CALTECH | 2004





THE VOLUMEPRO VOLUME RENDERING CLUSTER | SANTIAGO LOMBEYDA & JOHN MCCORQUODALE | CALTECH | 2003

PARALLEL PROCESSING

- requires a parallel system
- increases sw complexity
- less likely to be “portable”



CENTER FOR DATA-DRIVEN DISCOVERY

