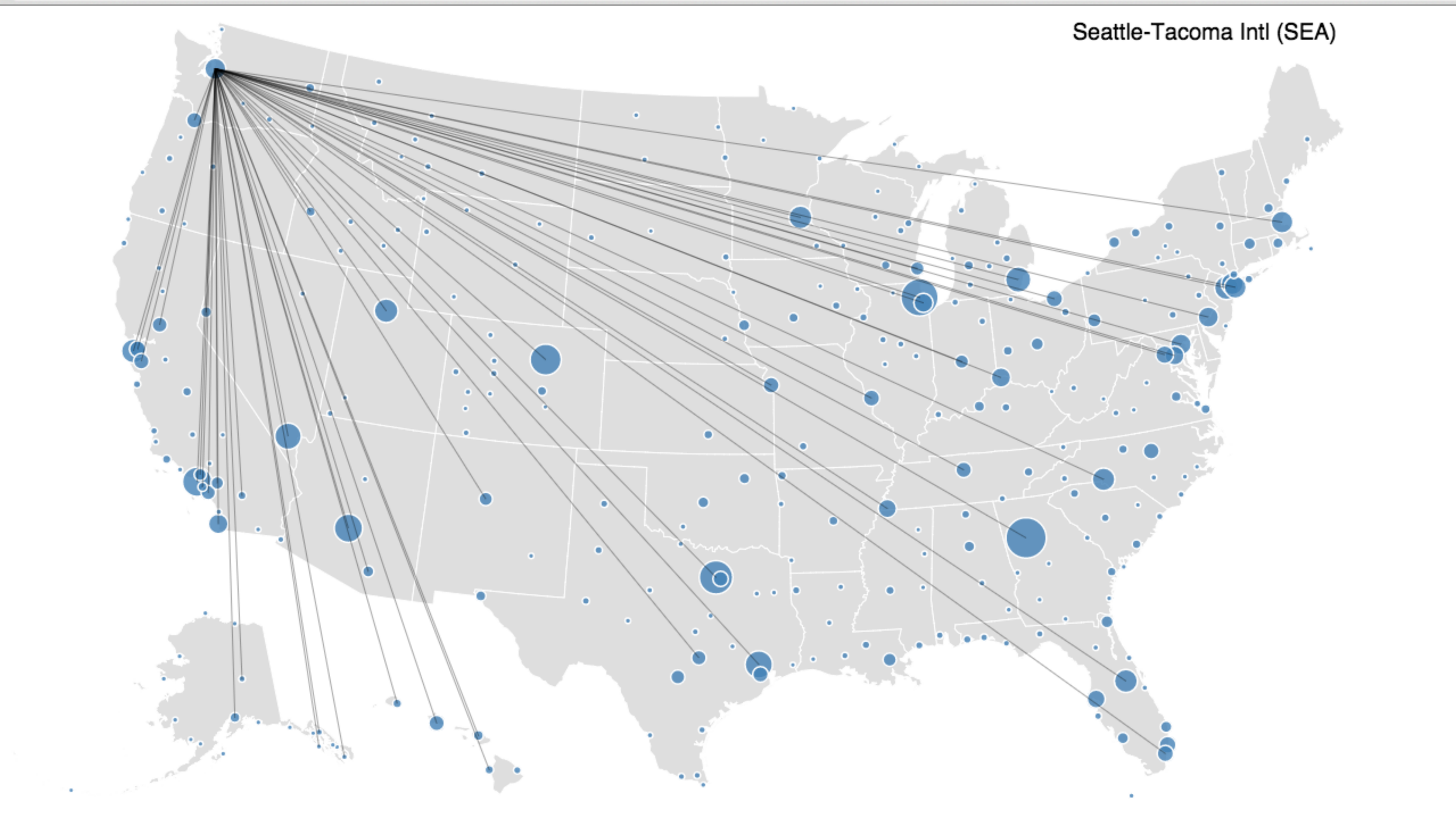
INTERACTIVE
Data Analysis

Jeffrey Heer @jeffrey_heer
U. Washington / Trifacta

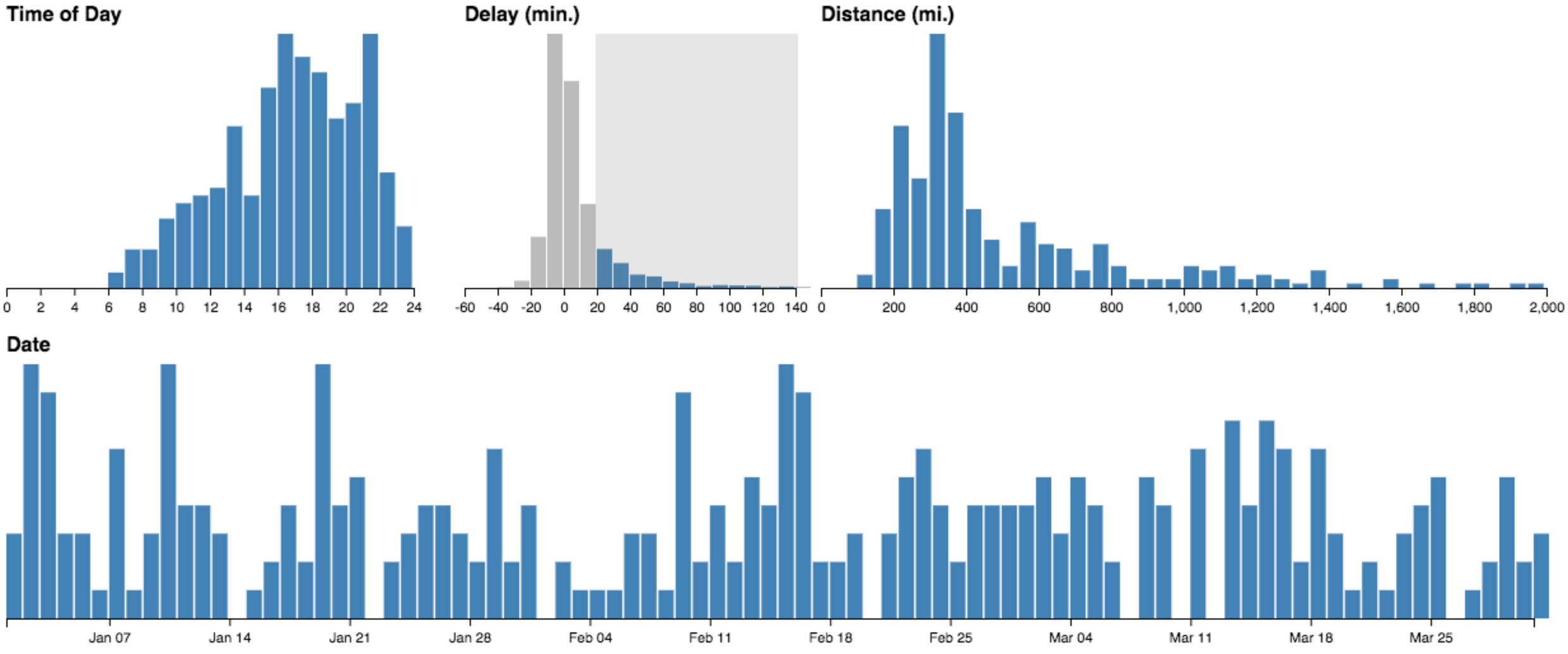





South Asia
Europe & Central Asia
Sub-Saharan Africa
America
East Asia & Pacific
Middle East & North Africa
Life Expectancy
Fertility
80
70
60
50
40
30
20
0
2
4
6
8
2005
2000
1995
Bangladesh
1990
1985
1980
1975
1970
1965
1960
1955
1990


Seattle-Tacoma Intl (SEA)

FAA Flight Records



prefuse

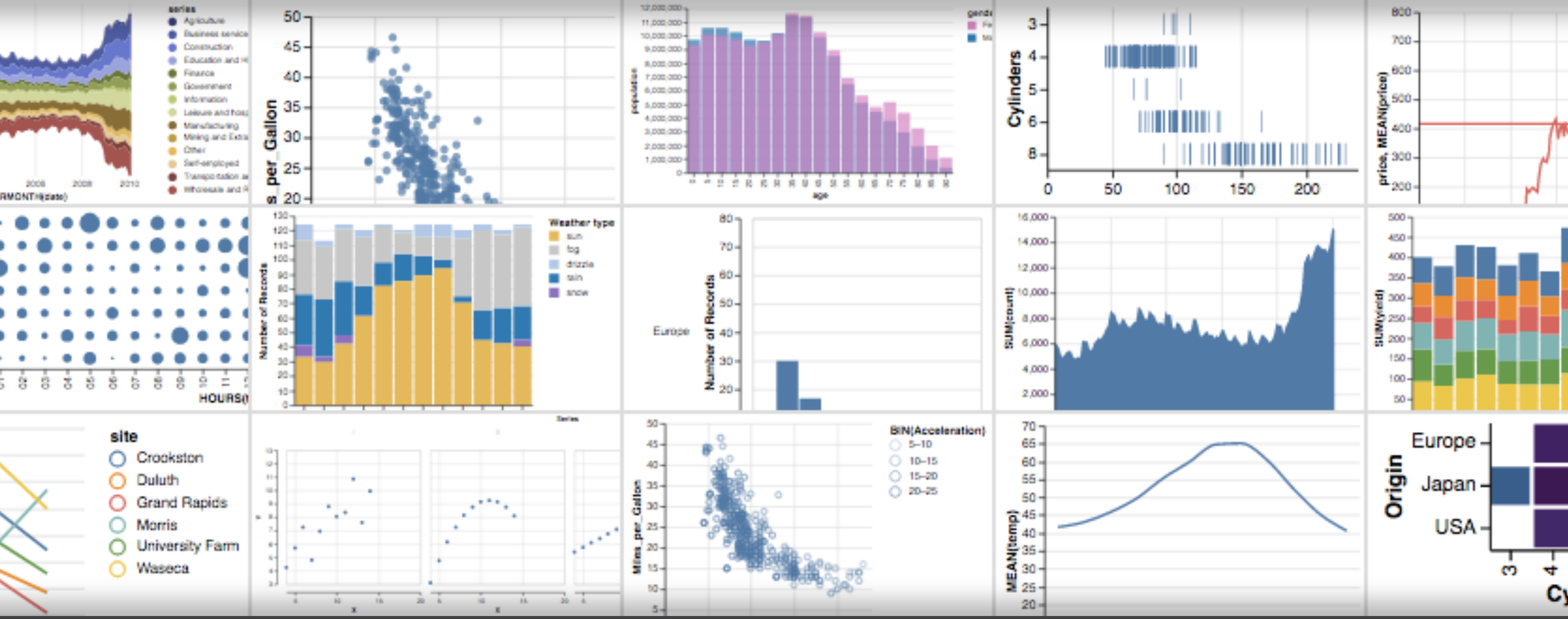
Protovis



vega

Data-Driven Documents

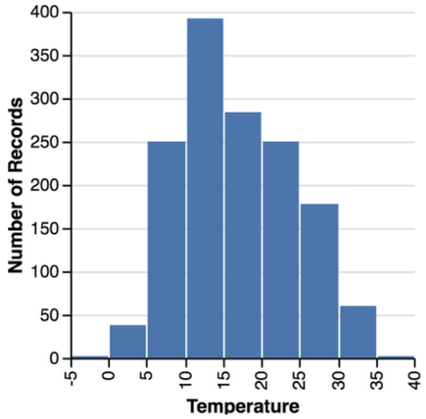
How might we author ***interactive*** **graphics** in the **midst of analysis**?



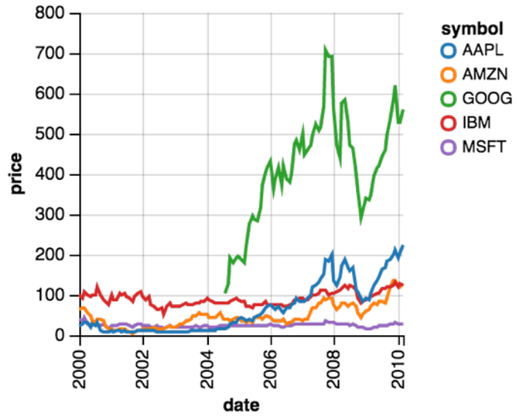
Vega-Lite: A Grammar of Interactive Graphics

A. Satyanarayan, D. Moritz, K. Wongsuphasawat & J. Heer. *TVCG 2017*

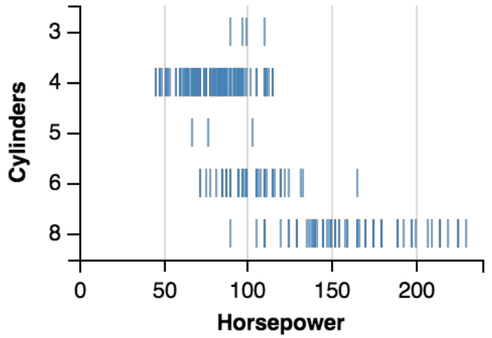
Histogram



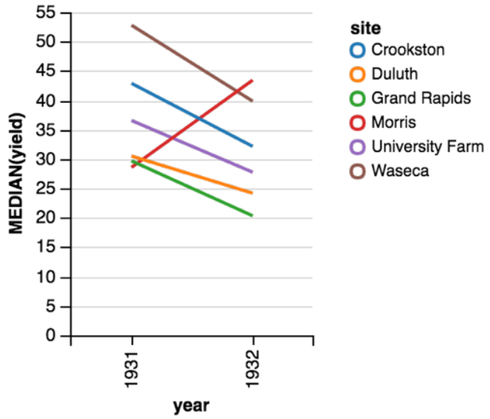
Line Chart



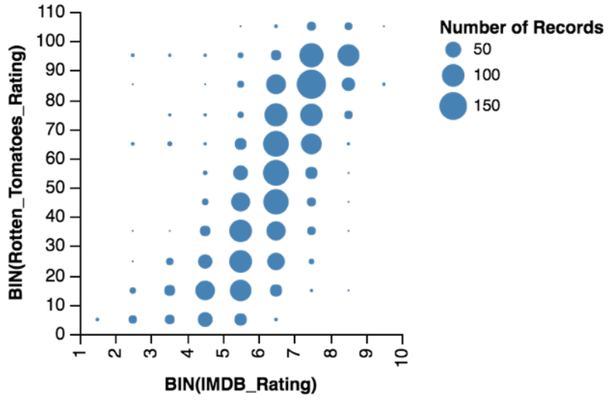
Strip Plot



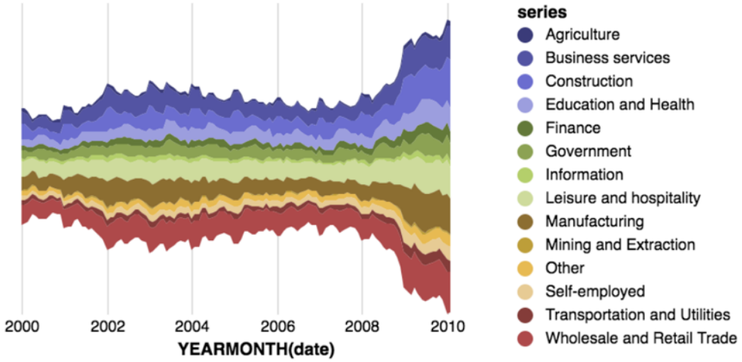
Slope Graph



Binned Scatter Plot

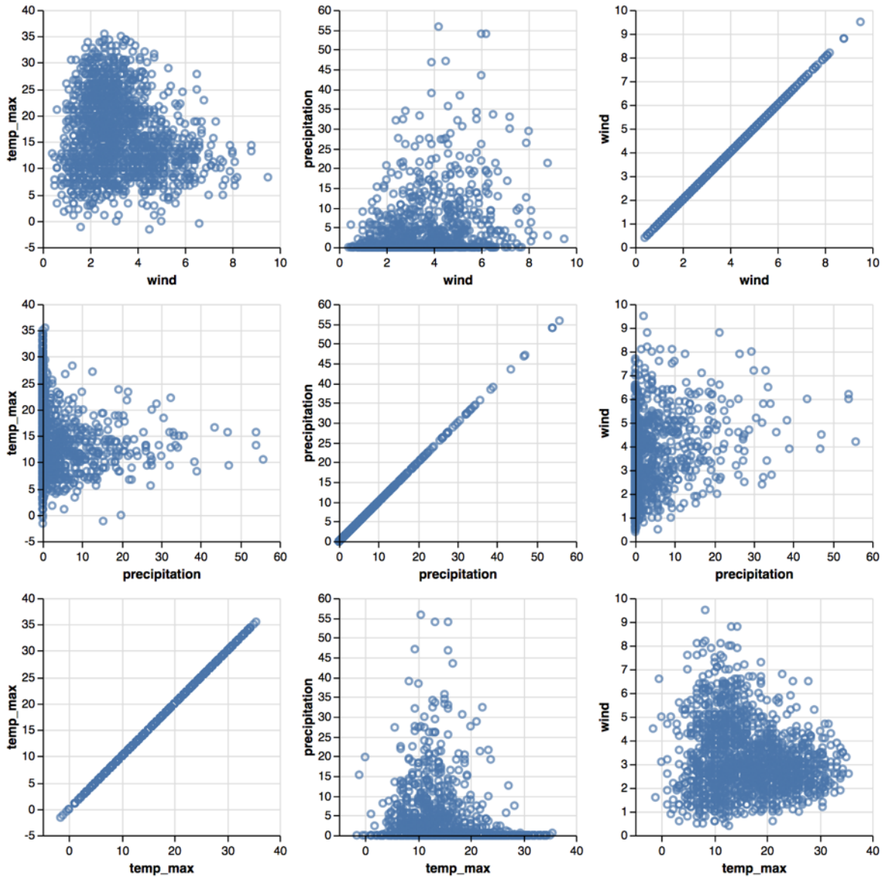


Area Chart

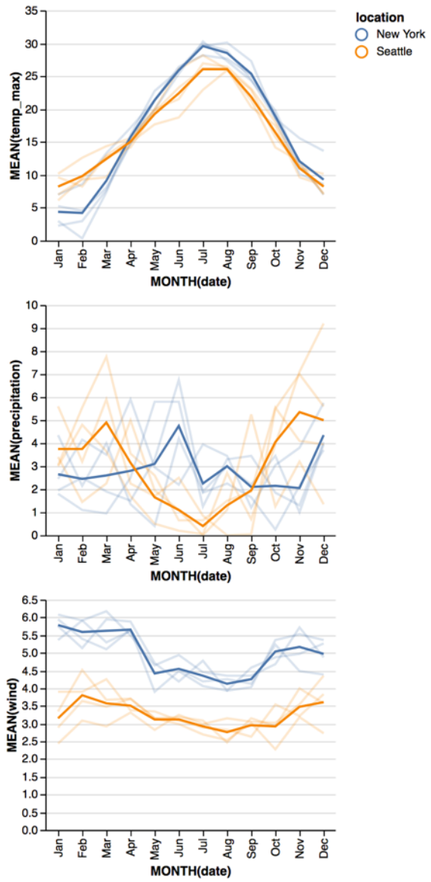


Vega-Lite: A Grammar of Graphics

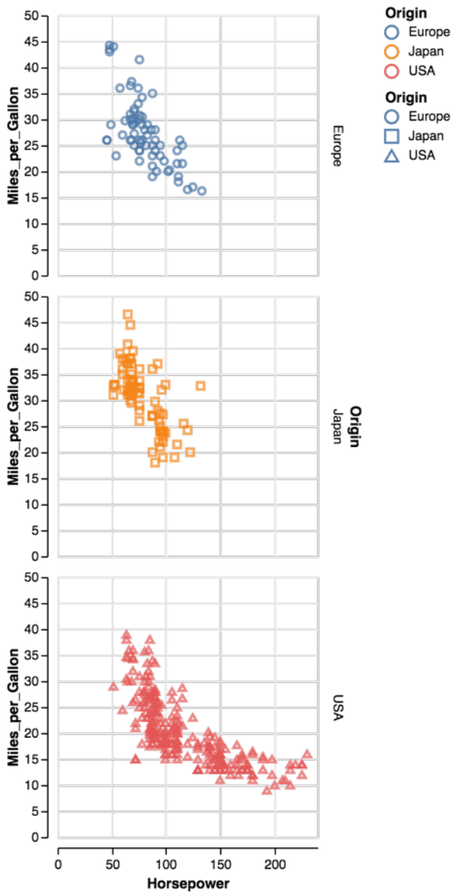
Scatter Plot Matrix



Concatenated & Layered View

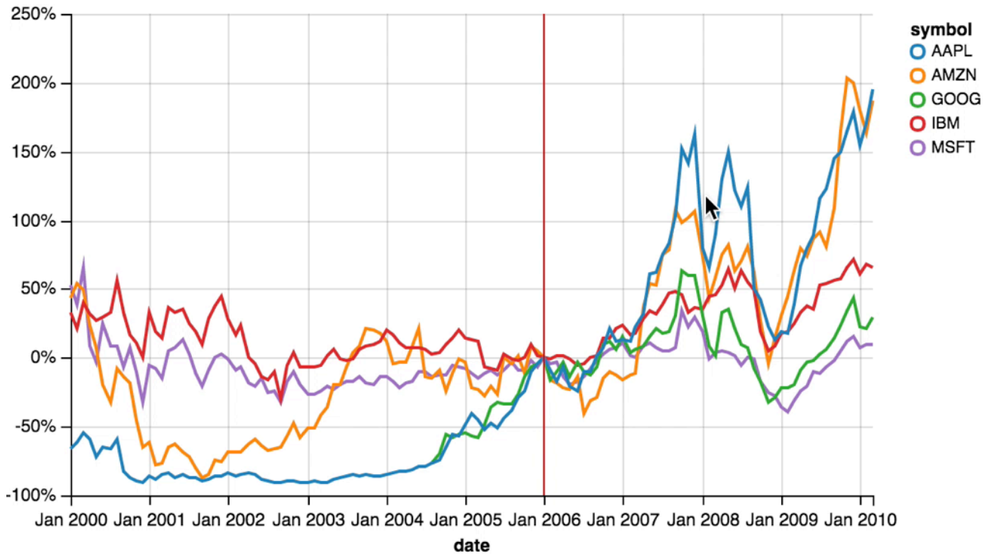


Faceted View

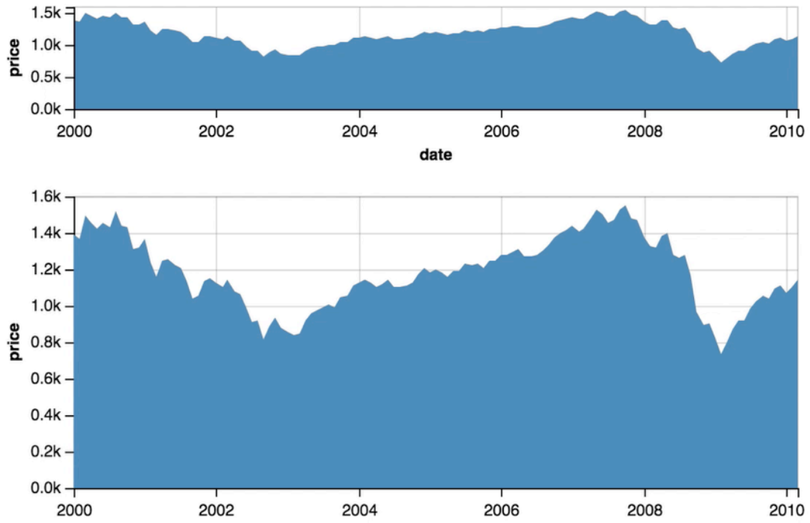


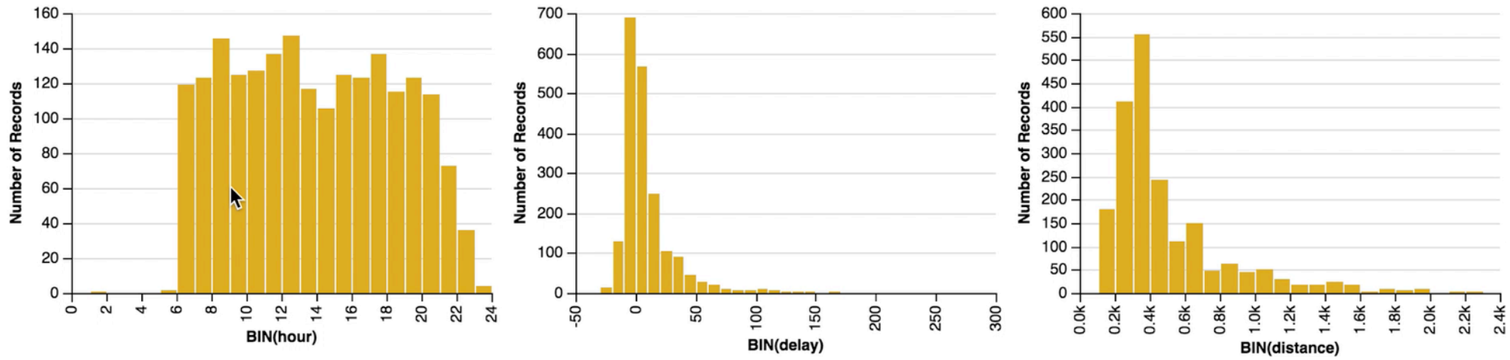
Vega-Lite: A Grammar of Multi-View Graphics

Indexed Chart



Focus + Context





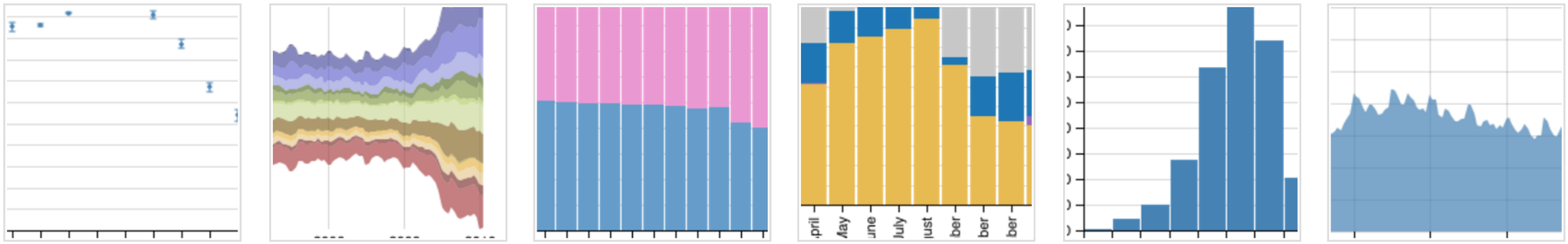
Cross-Filtering

Vega-Lite: A Grammar of Interactive Multi-View Graphics

Altair

DECLARATIVE VISUALIZATION IN PYTHON

Declarative Visualization in Python



Altair is a declarative statistical visualization library for Python, based on **Vega-Lite**.

With Altair, you can spend more time understanding your data and its meaning. Altair's API is simple, friendly and consistent and built on top of the powerful **Vega-Lite** visualization grammar. This elegant simplicity produces beautiful and effective visualizations with a minimal amount of code.

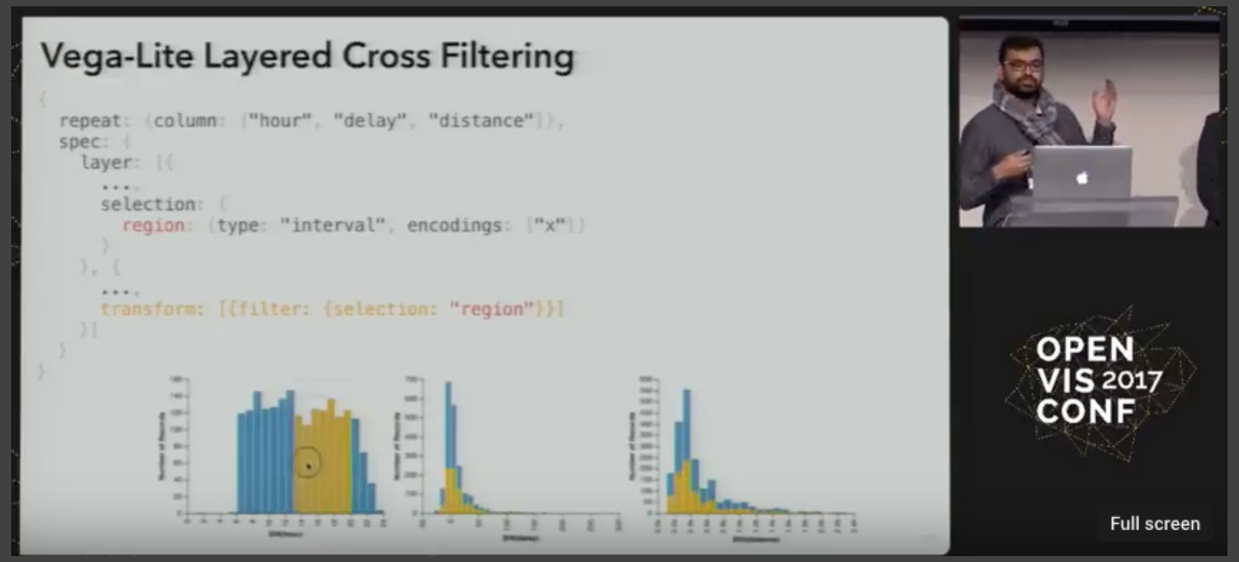
Altair: Vega-Lite in Python

Led by Jake VanderPlas & Brian Granger





To Learn More…



Vega-Lite: A Grammar of Interactive Graphics, *OpenVis Conf 2017*

youtu.be/9uaHRWj04D4



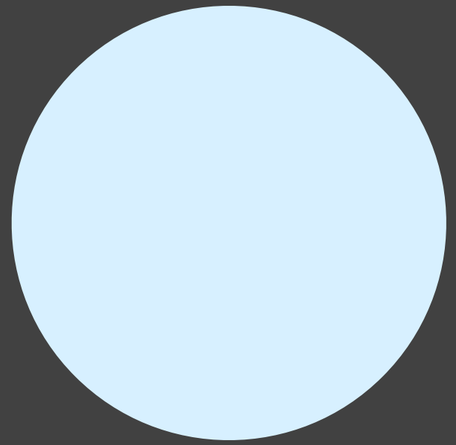
Altair: Declarative Visualization for Python, *PyData SF 2016*

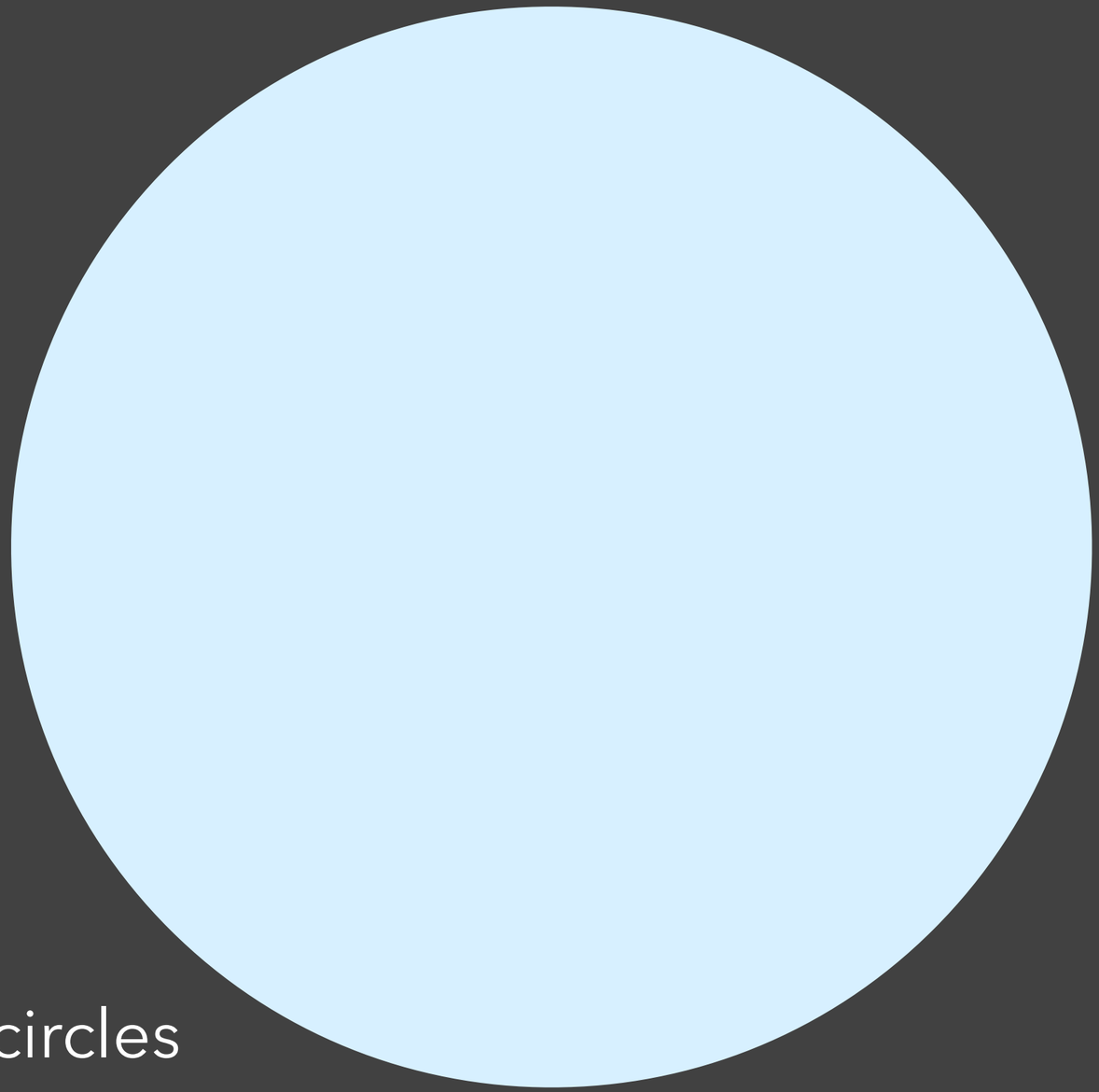
youtu.be/aRxahWy-ul8

How might our **tools** help us to become **better analysts**?

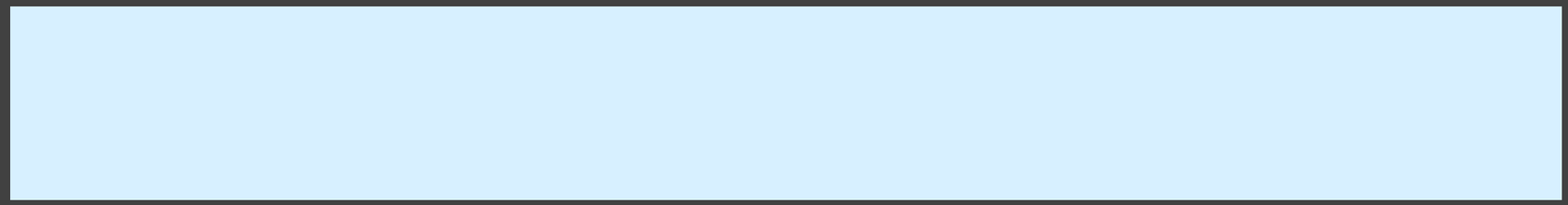
What makes a visualization **“good”**?

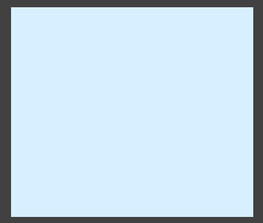
A Quick Experiment…



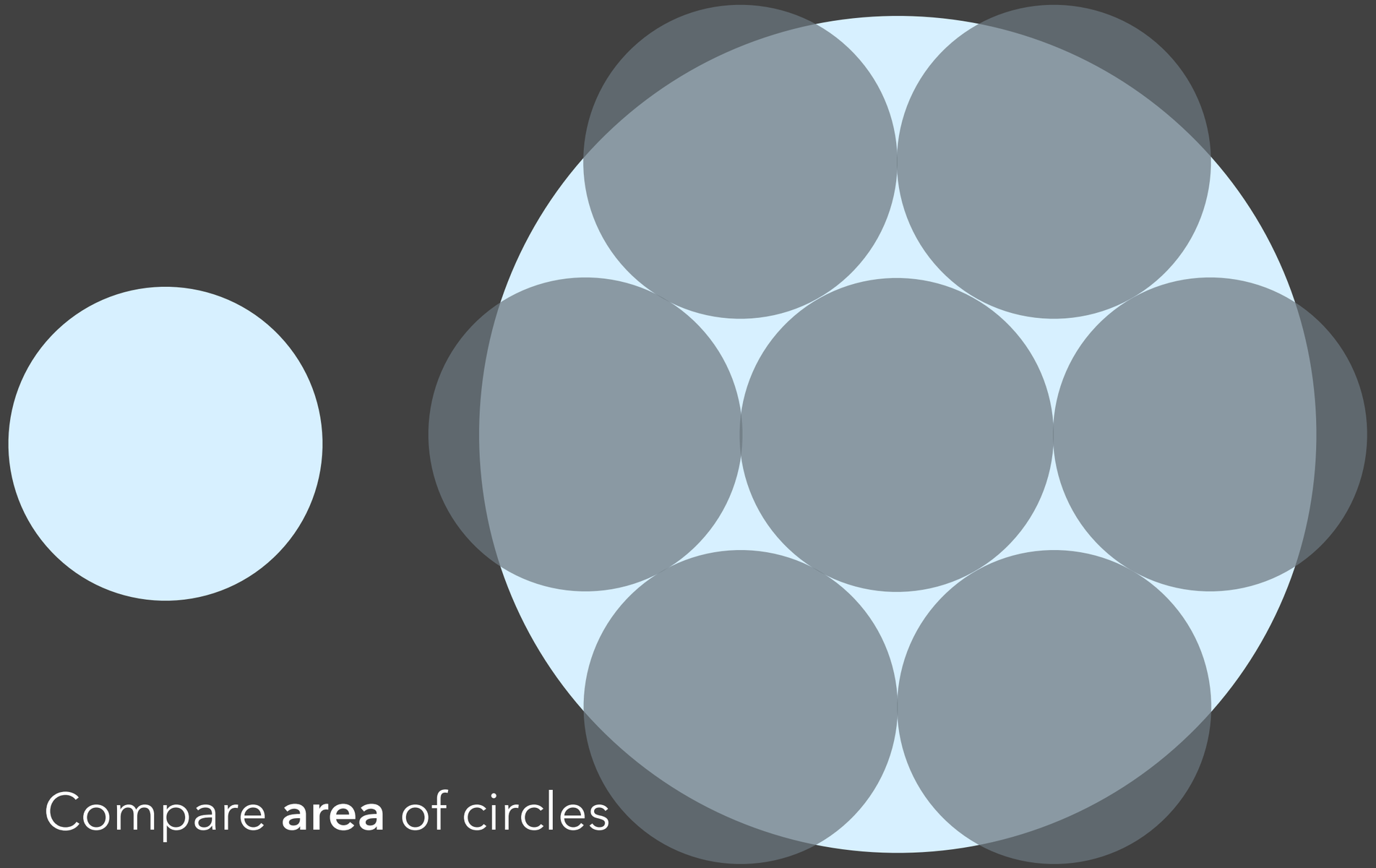


Compare **area** of circles





Compare **length** of bars

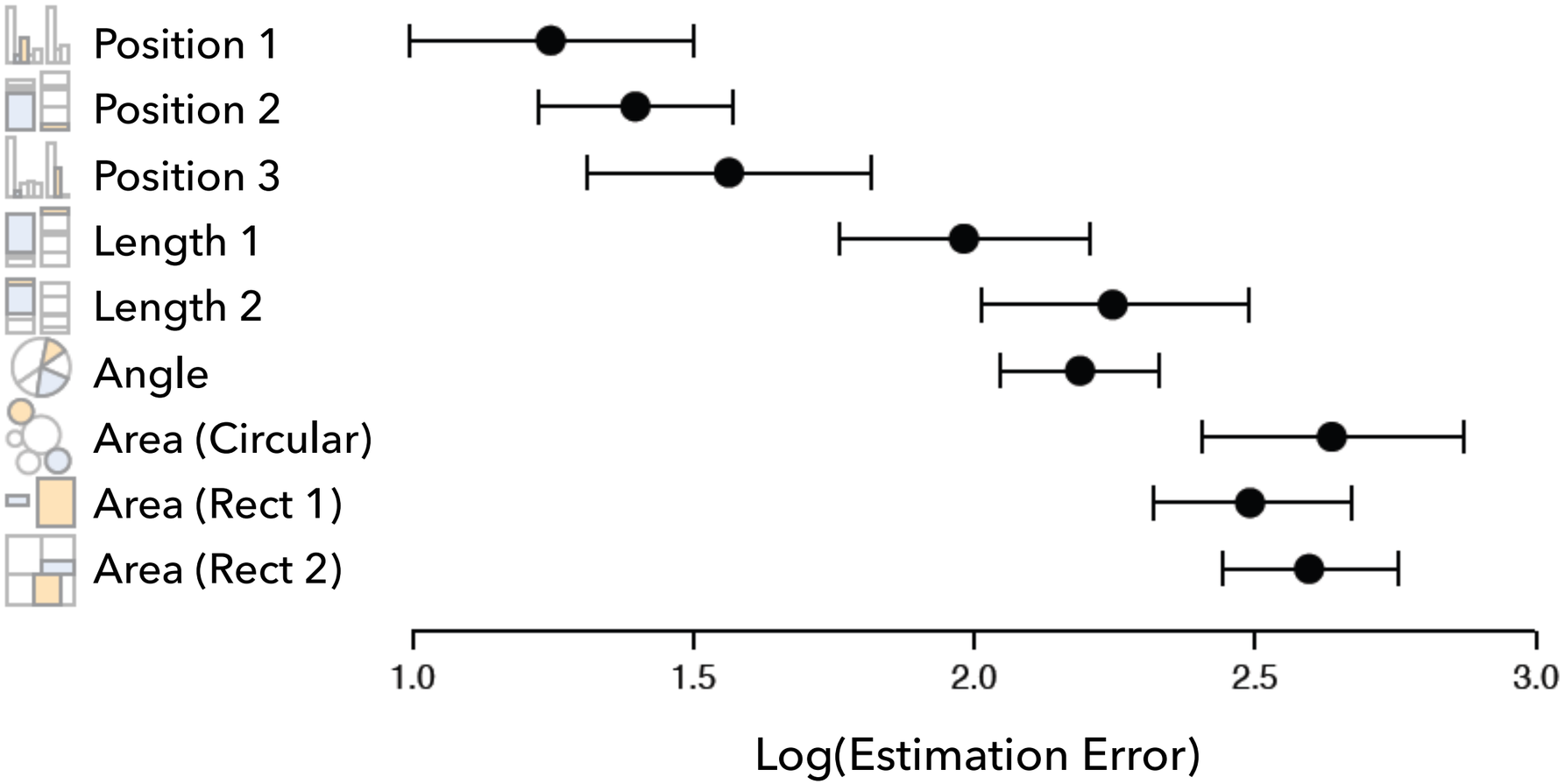


Compare **area** of circles



Compare **length** of bars

Accuracy of Visual Decoding



Ranking Visual Encodings

Most accurate

Comparing
Proportions

Least accurate



Position (common) scale
Position (non-aligned) scale

Length

Slope

Angle

Area

Volume

Color hue-saturation-density

21 Columns 1,537 Rows 6 Data Types Column Details

Sort: Default Edit

State

ABC City

WASHINGTON 40 2.60%

File_Number

ABC CommitteeID

ABC Occupation

ABC Memo_Text

ABC Report_Type

Trans_Date

Trans_Amount

ABC Employer

State

Ranking Visual Encodings

Most accurate

Comparing
Proportions

Least accurate



Position (common) scale
Position (non-aligned) scale

Length

Slope

Angle

Area

Volume

Color hue-saturation-density

Ranking Visual Encodings

Most accurate

Comparing
Proportions

Least accurate



Position (common) scale
Position (non-aligned) scale

Length

Slope

Angle

Area

Volume

Color hue-saturation-density

Ranking Visual Encodings

Most accurate

Comparing
Proportions

Least accurate



Position (common) scale
Position (non-aligned) scale

Length

Slope

Angle

Area

Volume

Color hue-saturation-density

21 Columns 1,537 Rows 6 Data Types Column Details

Sort: Default Edit

State

ABC City

WASHINGTON 40 2.60%

File_Number

ABC CommitteeID

ABC Occupation

ABC Memo_Text

ABC Report_Type

Trans_Date

Trans_Amount

ABC Employer

State

21 Columns 1,537 Rows 6 Data Types Column Details

Sort: Default Edit

State

ABC City

WASHINGTON 40 2.60%

File_Number

ABC CommitteeID

ABC Occupation

ABC Memo_Text

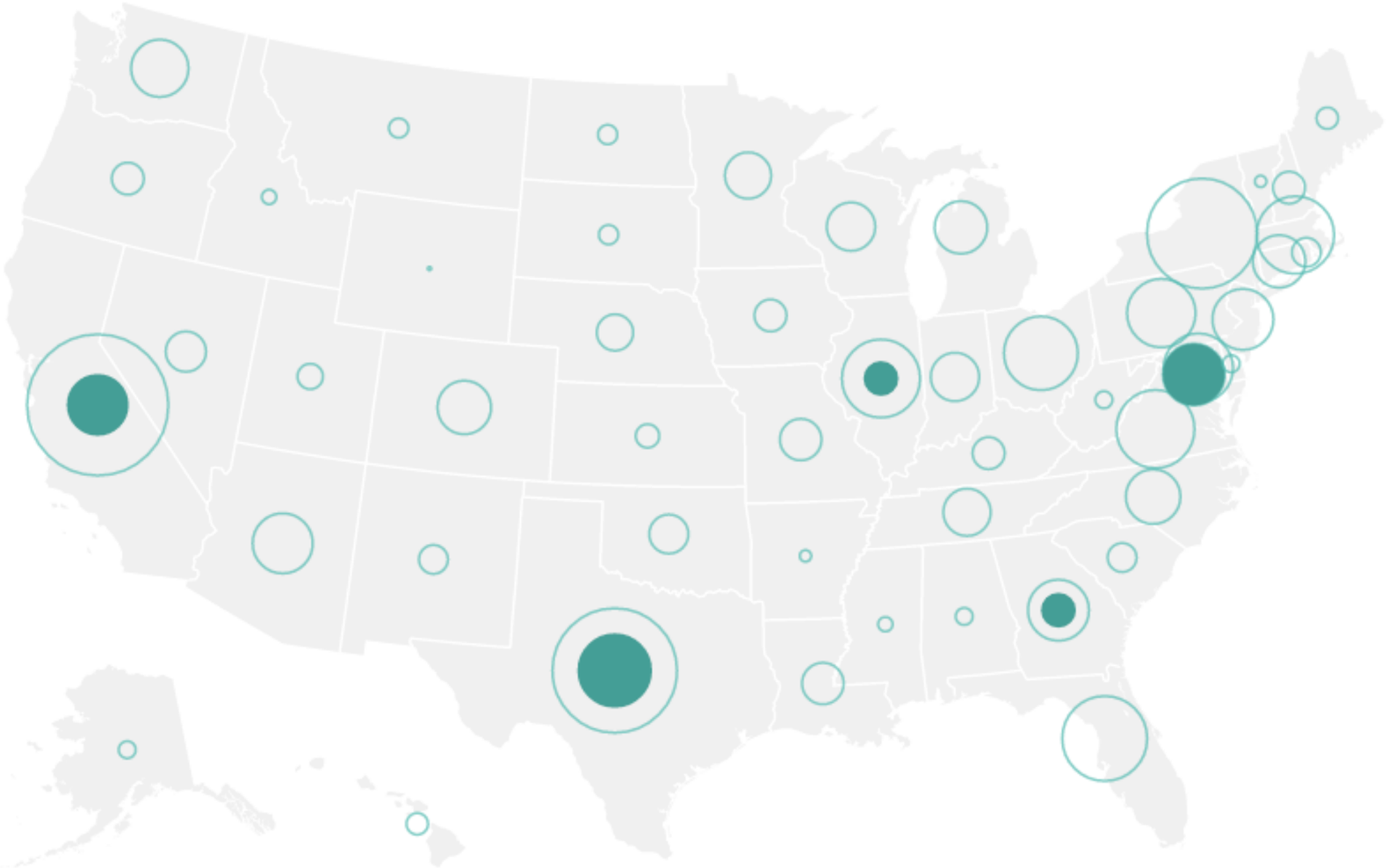
ABC Report_Type

Trans_Date

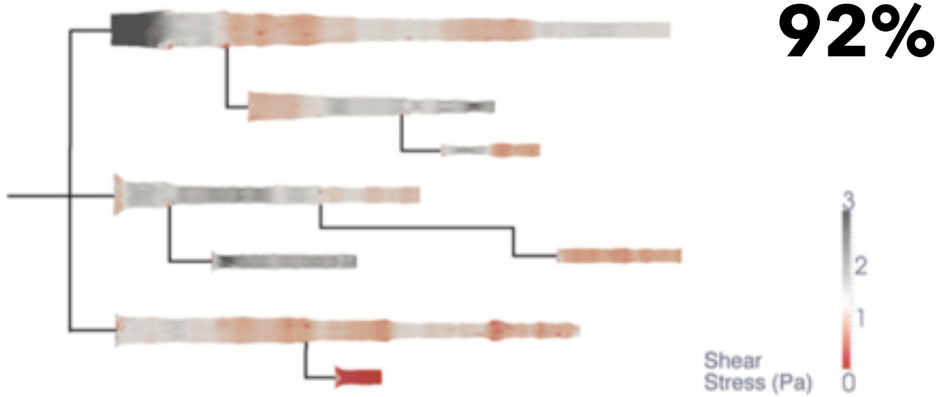
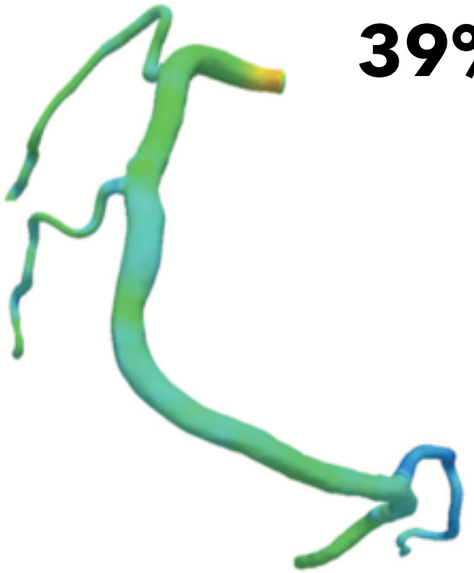
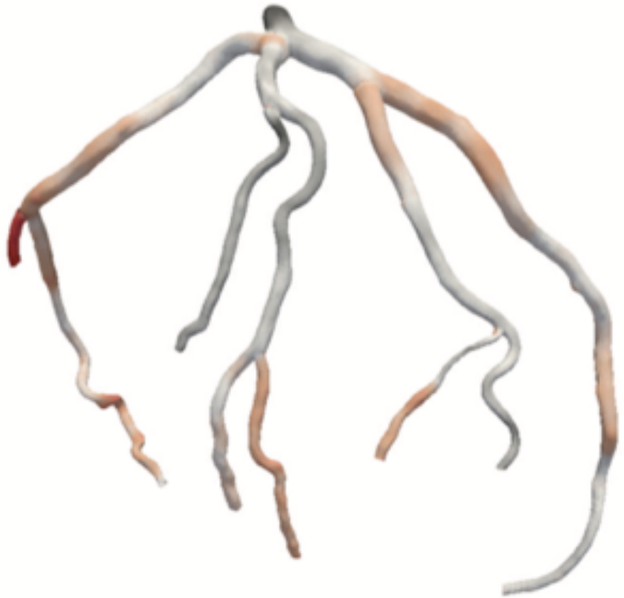
Trans_Amount

ABC Employer

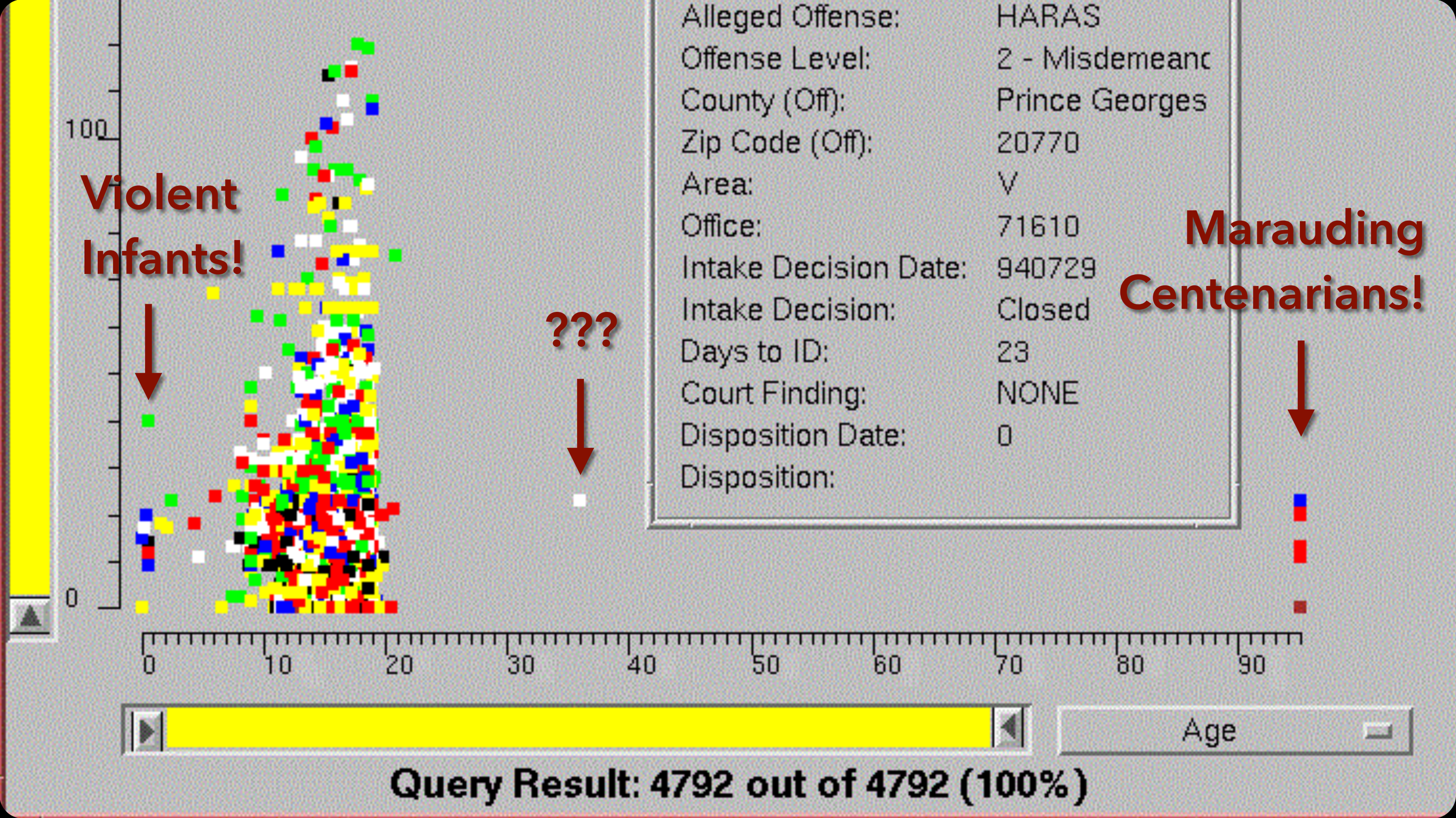
State

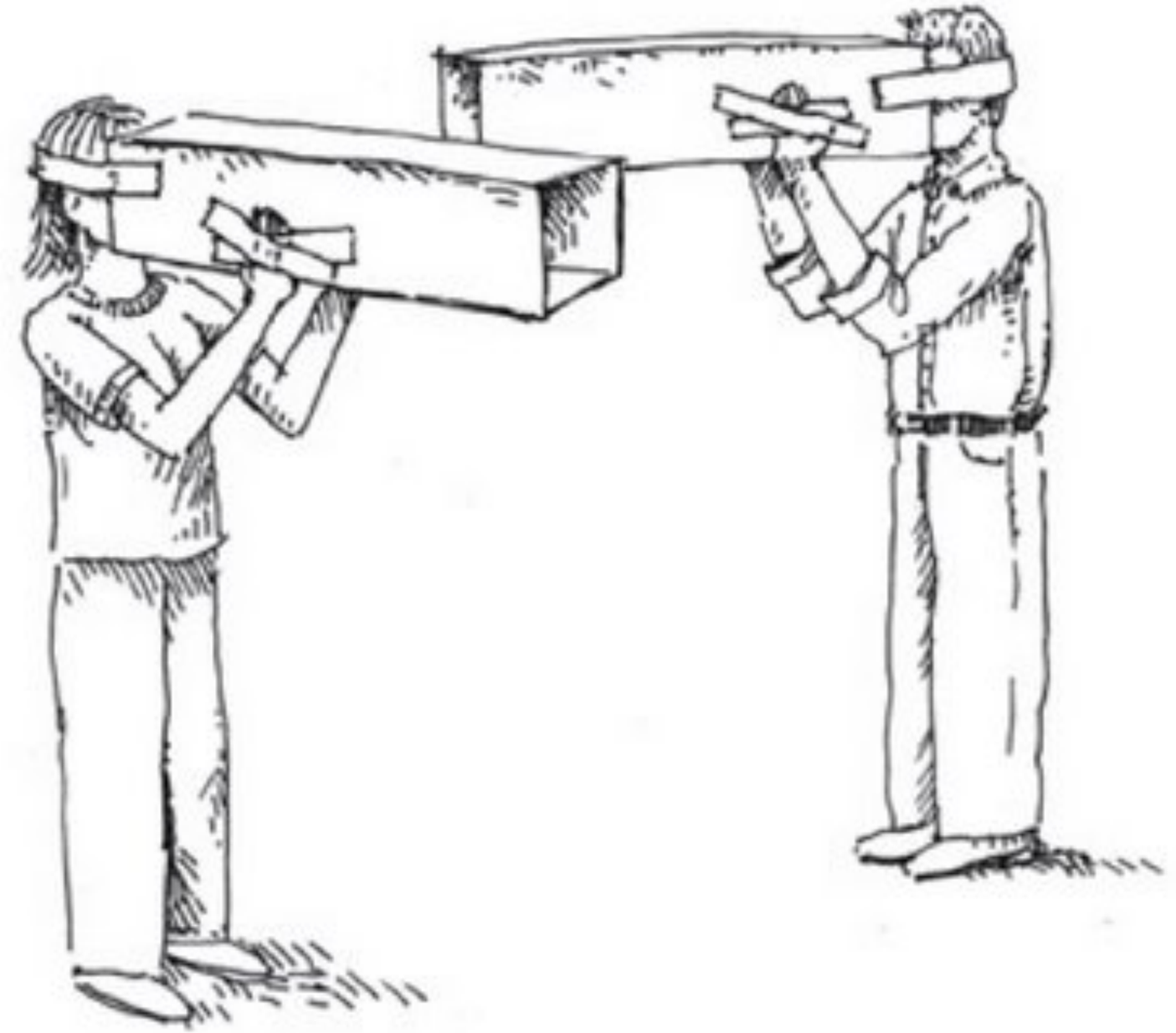


Artery Visualization [Borkin et al. 2011]

	Rainbow Palette	Diverging Palette
2D	**62%** 	**92%** 
3D	**39%** 	**71%** 

How might we support more **effective data exploration?**


Violent Infants!
???
Marauding Centenarians!
Alleged Offense: HARAS
Offense Level: 2 - Misdemeanc
County (Off): Prince Georges
Zip Code (Off): 20770
Area: V
Office: 71610
Intake Decision Date: 940729
Intake Decision: Closed
Days to ID: 23
Court Finding: NONE
Disposition Date: 0
Disposition:
100
0
0 10 20 30 40 50 60 70 80 90
Age
Query Result: 4792 out of 4792 (100%)



Common analysis pitfalls:

Common analysis pitfalls:

Overlook data quality issues

Common analysis pitfalls:

Overlook data quality issues

Fixate on specific relationships

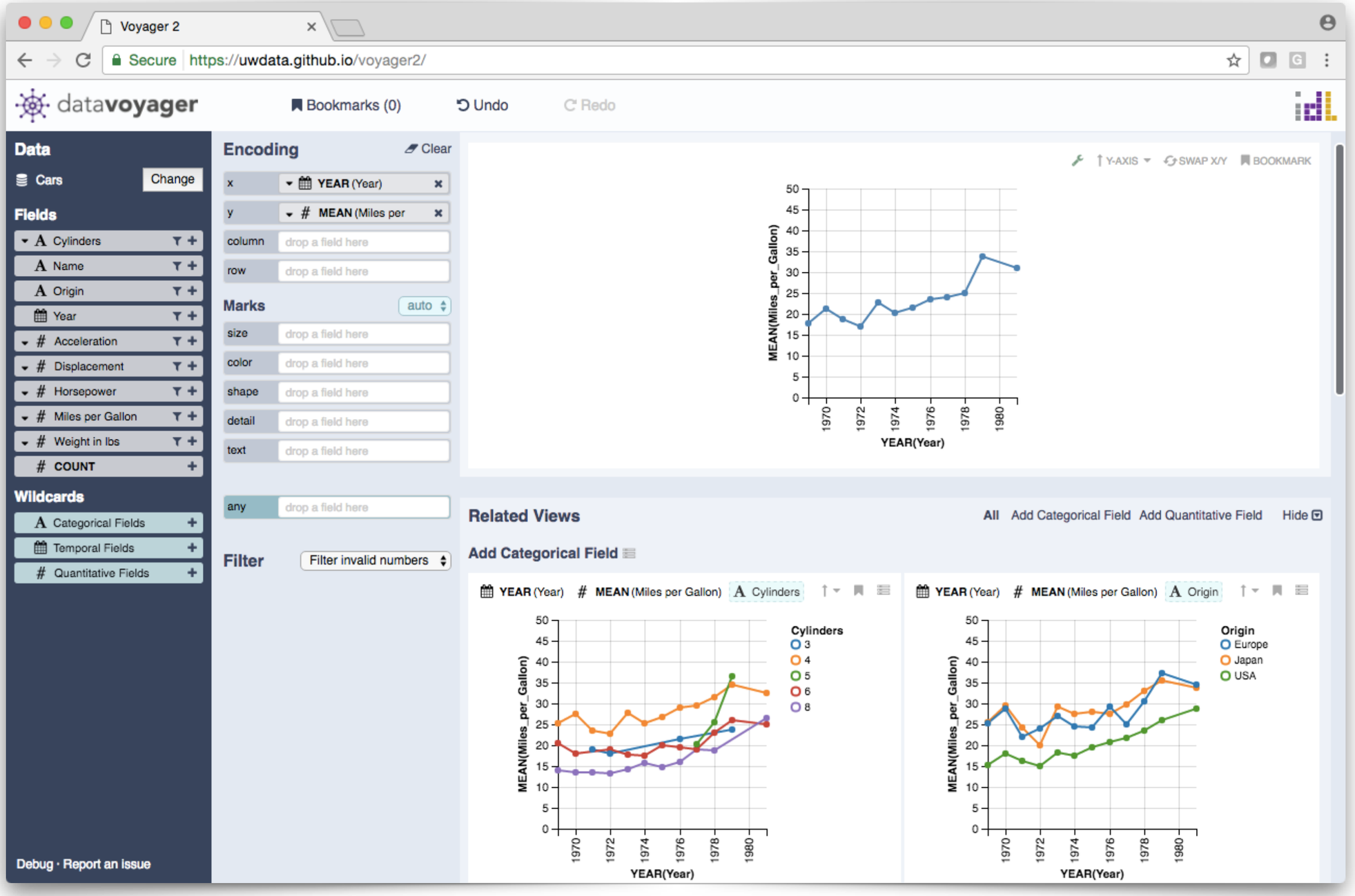
Common analysis pitfalls:

Overlook data quality issues

Fixate on specific relationships

Plus many other cognitive biases

[Heuer 1999, Kahneman 2011, …]



Voyager: Combine Manual Specification with Visualization Recommenders

Compared to existing tools, leads to **over 4x more variable sets seen**, and **over 2x more variable sets interacted with**.

“The related view suggestion accelerates exploration a lot.”

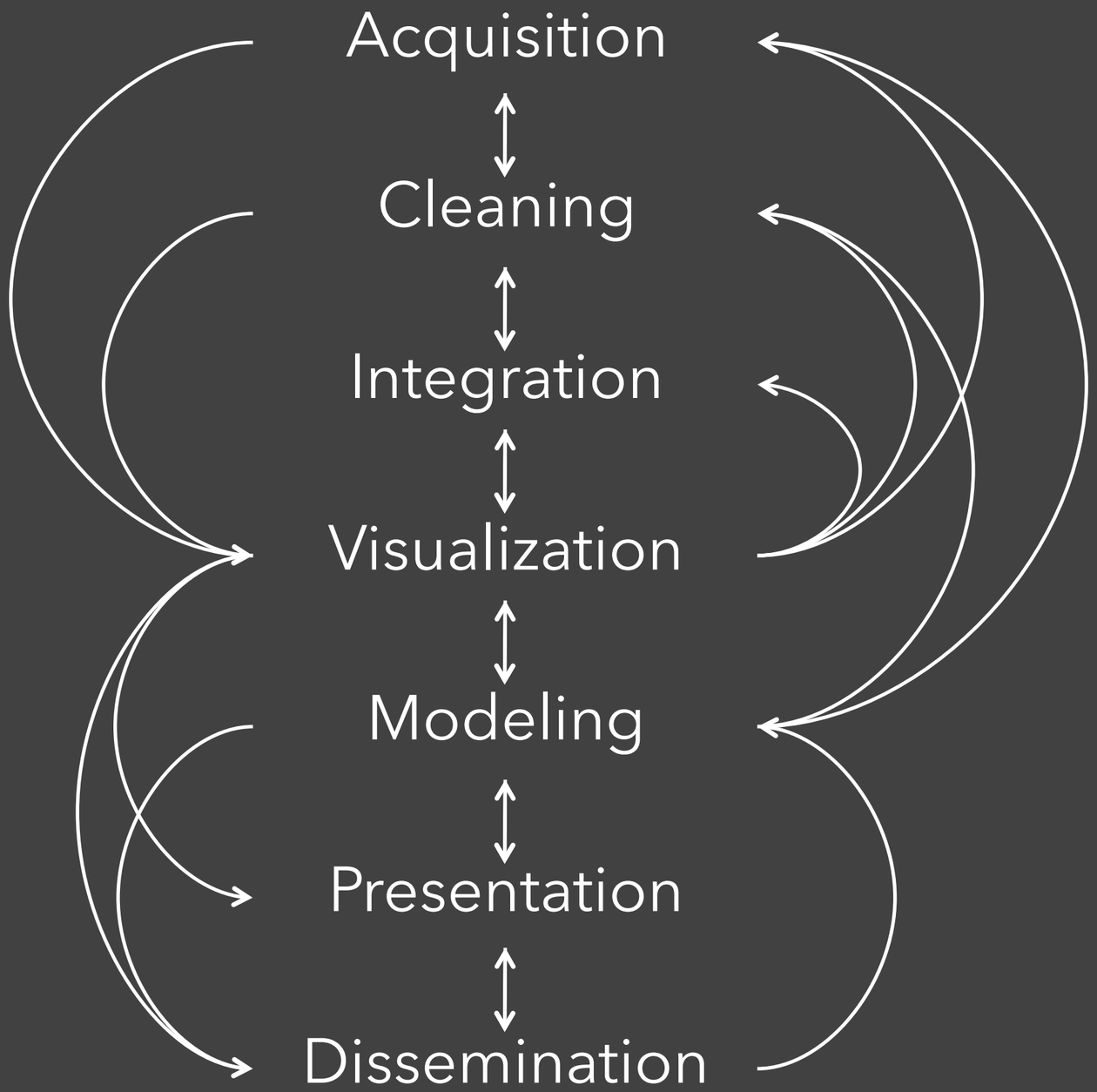
“I like that it shows me what fields to include in order to see a specific graph. Otherwise, I have to do a lot of trial and error and can't express what I wanted to see.”

“These related views are so good but it’s also spoiling that I start thinking less. I’m not sure if that’s really a good thing.”



WORK IN PROGRESS...

Voyager integration with JupyterLab!


Acquisition
Cleaning
Integration
Visualization
Modeling
Presentation
Dissemination

I spend more than half of my time integrating, cleansing and transforming data without doing any actual analysis. Most of the time I'm lucky if I get to do any "analysis" at all.

Anonymous Data Scientist
from our 2012 interview study





Big Data Borat

@BigDataBorat

Following

In Data Science, 80% of time spent prepare data, 20% of time spent complain about need for prepare data.

How might we support
interactive data wrangling?

Bureau of Justice Statistics - Data Online
http://bjs.ojp.usdoj.gov/

Reported crime in Alabama

Year	Population	Property crime rate	Burglary rate	Larceny-theft rate	Motor vehicle theft rate
2004	4525375	4029.3	987	2732.4	309.9
2005	4548327	3900	955.8	2656	289
2006	4599030	3937	968.9	2645.1	322.9
2007	4627851	3974.9	980.2	2687	307.7
2008	4661900	4081.9	1080.7	2712.6	288.6

Reported crime in Alaska

Year	Population	Property crime rate	Burglary rate	Larceny-theft rate	Motor vehicle theft rate
2004	657755	3370.9	573.6	2456.7	340.6
2005	663253	3615	622.8	2601	391
2006	670053	3582	615.2	2588.5	378.3
2007	683478	3373.9	538.9	2480	355.1
2008	686293	2928.3	470.9	2219.9	237.5

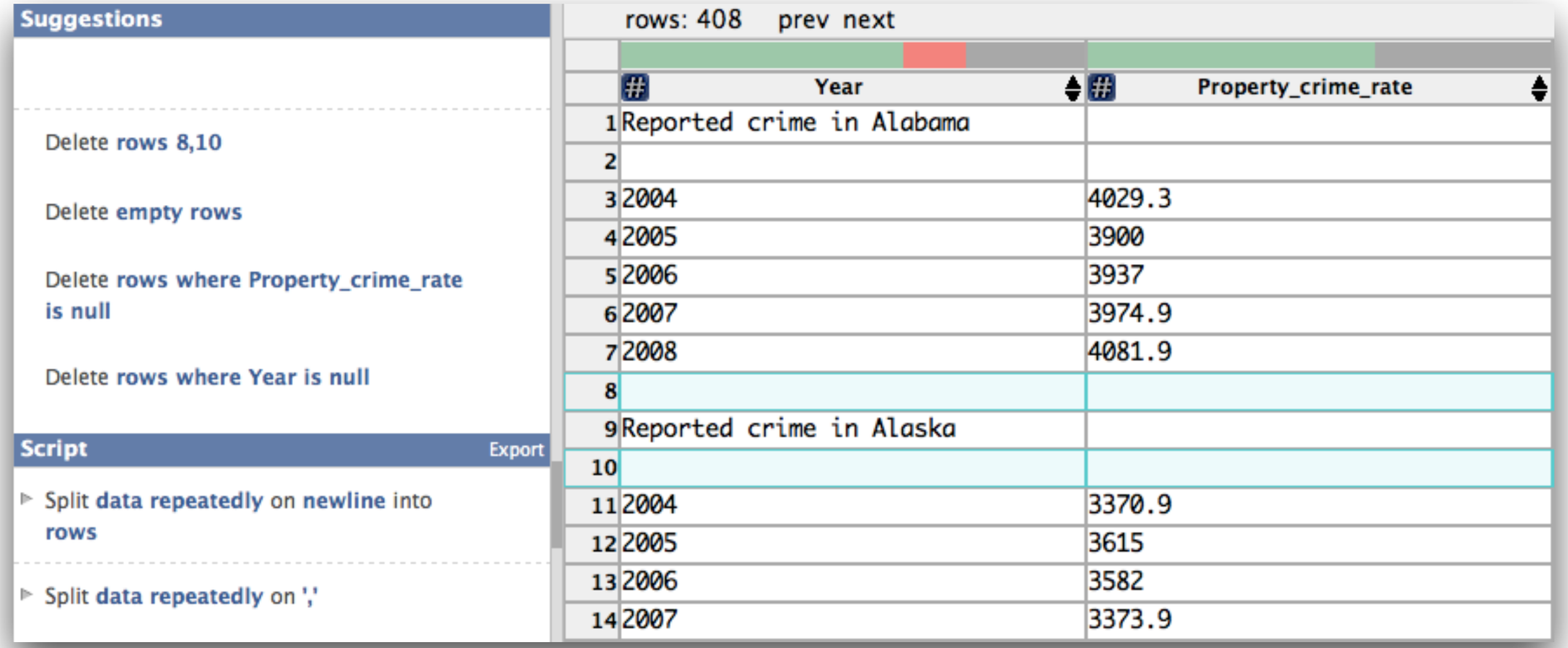
Reported crime in Arizona

Year	Population	Property crime rate	Burglary rate	Larceny-theft rate	Motor vehicle theft rate
2004	5739879	5073.3	991	3118.7	963.5
2005	5953007	4827	946.2	2958	922
2006	6166318	4741.6	953	2874.1	914.4
2007	6338755	4502.6	935.4	2780.5	786.7
2008	6500180	4087.3	894.2	2605.3	587.8

Reported crime in Arkansas

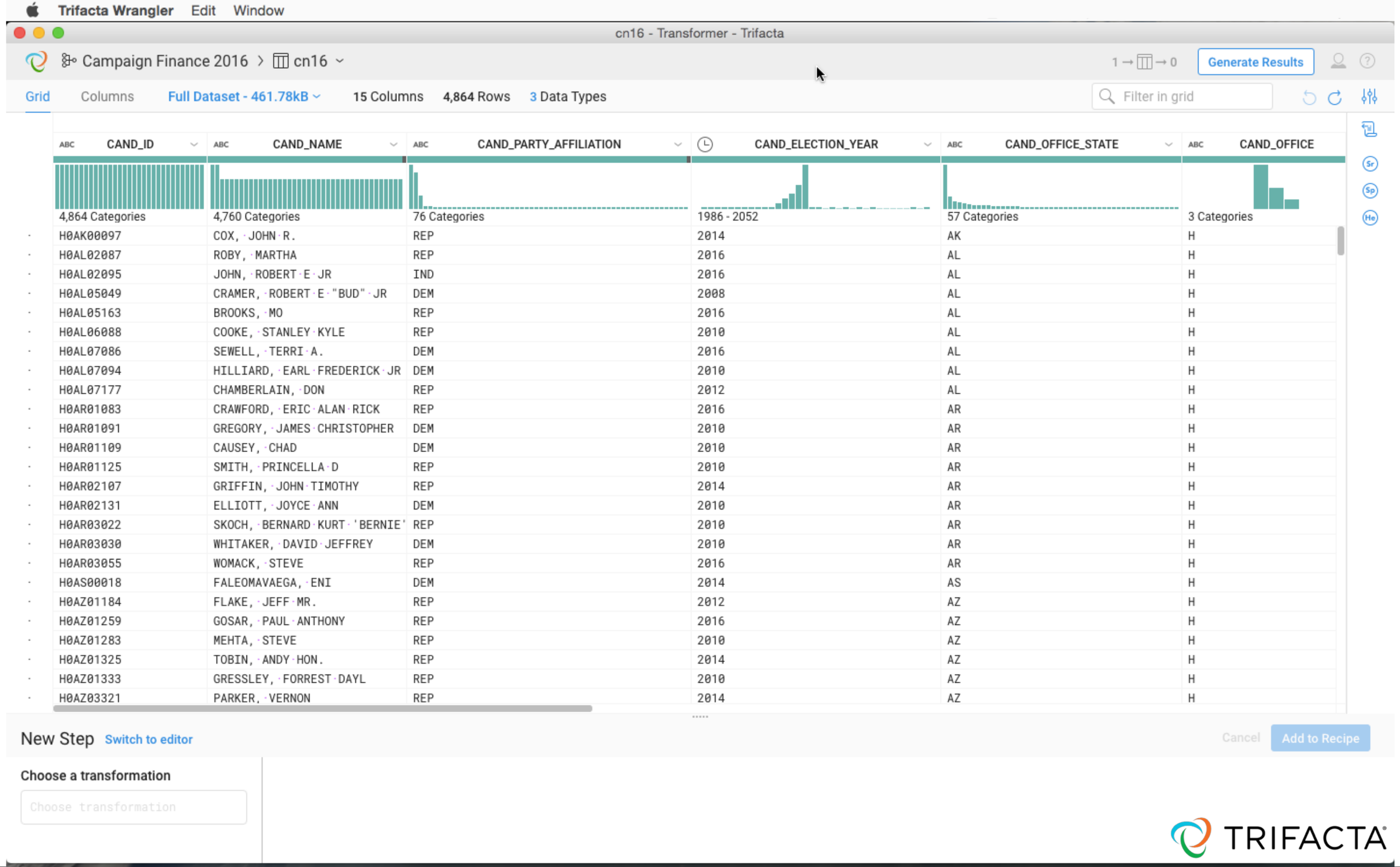
Year	Population	Property crime rate	Burglary rate	Larceny-theft rate	Motor vehicle theft rate
2004	2750000	4033.1	1096.4	2699.7	237
2005	2775708	4068	1085.1	2720	262
2006	2810872	4021.6	1154.4	2596.7	270.4
2007	2834797	3945.5	1124.4	2574.6	246.5
2008	2855390	3843.7	1182.7	2433.4	227.6

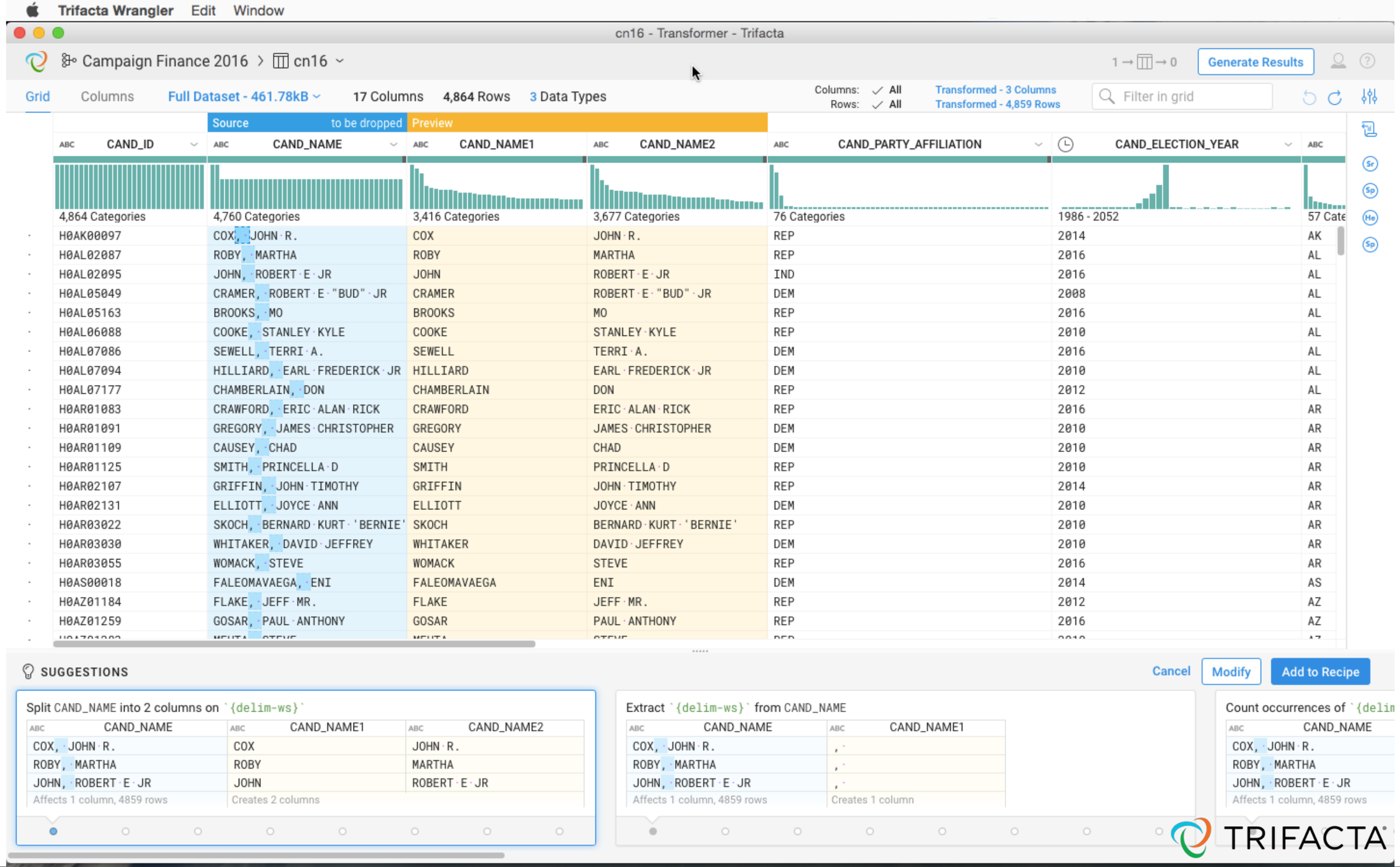
DataWrangler

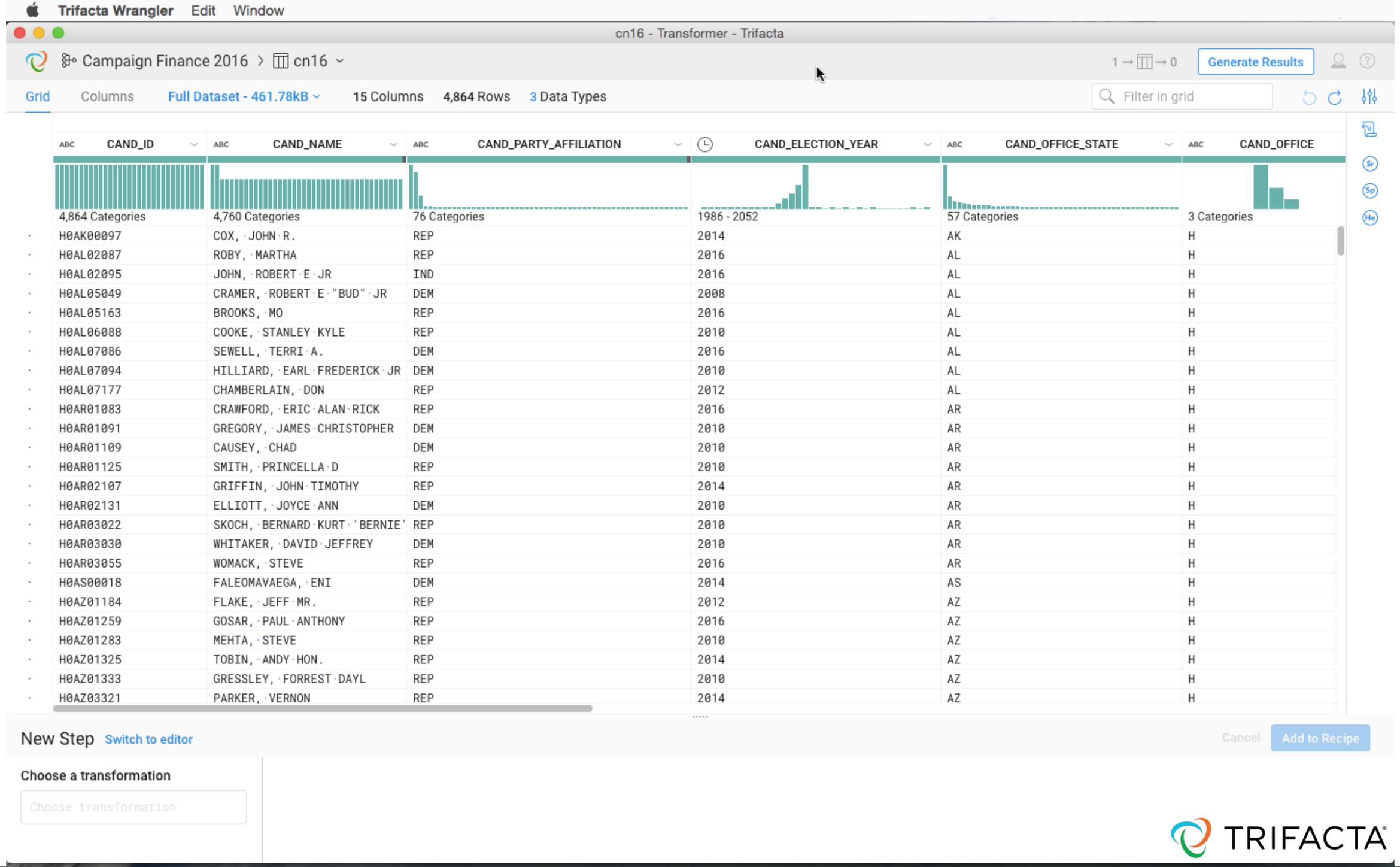


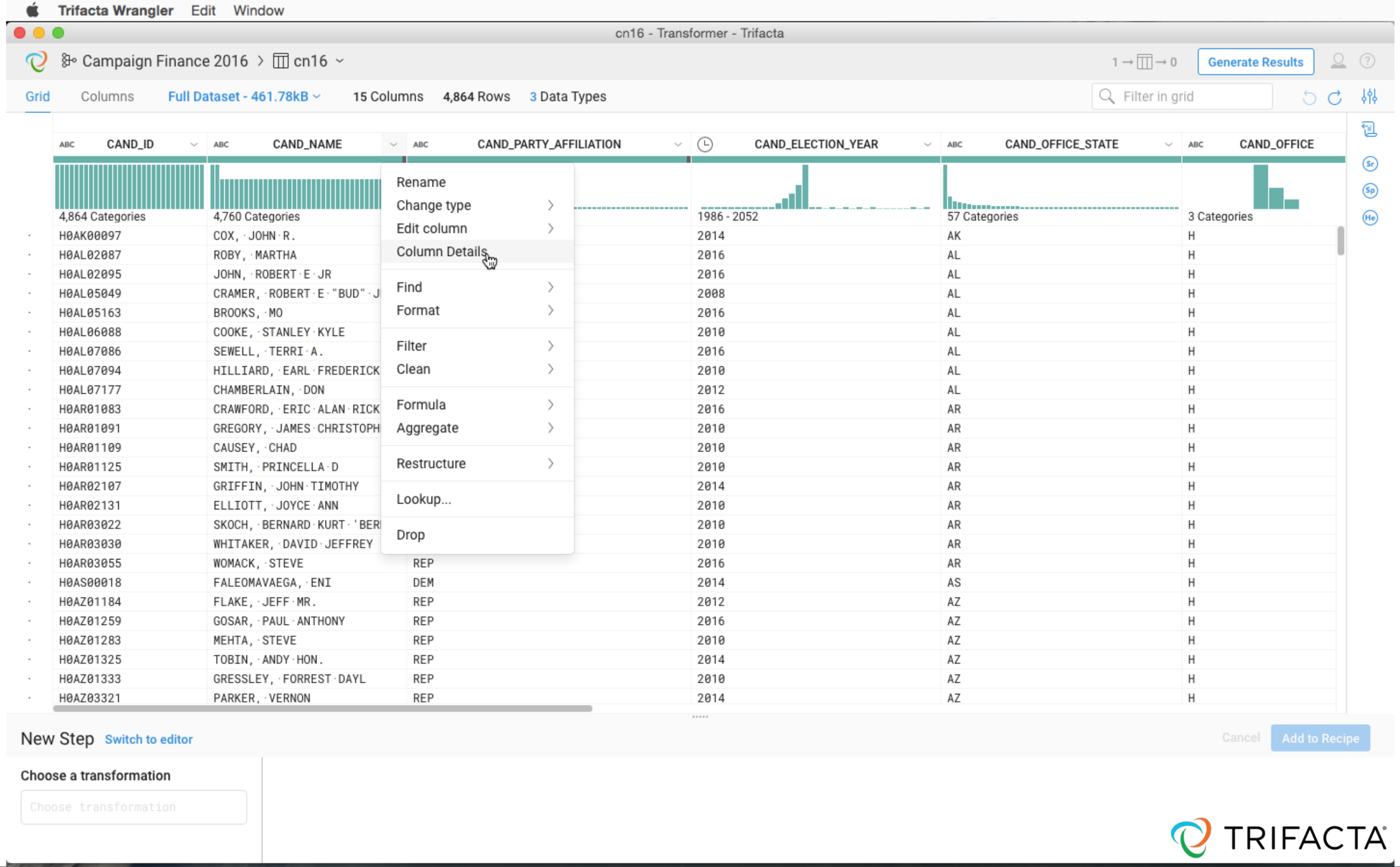
Wrangler: Interactive Visual Specification of Data Transformation Scripts

Sean Kandel et al. *ACM CHI 2011*


Trifacta Wrangler Edit Window
cn16 - Transformer - Trifacta
Campaign Finance 2016 › cn16
1 → 0
Generate Results
Grid Columns Full Dataset - 461.78kB 15 Columns 4,864 Rows 3 Data Types
Filter in grid
ABC CAND_ID
ABC CAND_NAME
ABC CAND_PARTY_AFFILIATION
CAND_ELECTION_YEAR
ABC CAND_OFFICE_STATE
ABC CAND_OFFICE
4,864 Categories
4,760 Categories
76 Categories
1986 - 2052
57 Categories
3 Categories
H0AK00097 COX, JOHN R. REP 2014 AK H
H0AL02087 ROBY, MARTHA REP 2016 AL H
H0AL02095 JOHN, ROBERT E JR IND 2016 AL H
H0AL05049 CRAMER, ROBERT E "BUD" JR DEM 2008 AL H
H0AL05163 BROOKS, MO REP 2016 AL H
H0AL06088 COOKE, STANLEY KYLE REP 2010 AL H
H0AL07086 SEWELL, TERRI A. DEM 2016 AL H
H0AL07094 HILLIARD, EARL FREDERICK JR DEM 2010 AL H
H0AL07177 CHAMBERLAIN, DON REP 2012 AL H
H0AR01083 CRAWFORD, ERIC ALAN RICK REP 2016 AR H
H0AR01091 GREGORY, JAMES CHRISTOPHER DEM 2010 AR H
H0AR01109 CAUSEY, CHAD DEM 2010 AR H
H0AR01125 SMITH, PRINCELLA D REP 2010 AR H
H0AR02107 GRIFFIN, JOHN TIMOTHY REP 2014 AR H
H0AR02131 ELLIOTT, JOYCE ANN DEM 2010 AR H
H0AR03022 SKOCH, BERNARD KURT 'BERNIE' REP 2010 AR H
H0AR03030 WHITAKER, DAVID JEFFREY DEM 2010 AR H
H0AR03055 WOMACK, STEVE REP 2016 AR H
H0AS00018 FALEOMAVAEGA, ENI DEM 2014 AS H
H0AZ01184 FLAKE, JEFF MR. REP 2012 AZ H
H0AZ01259 GOSAR, PAUL ANTHONY REP 2016 AZ H
H0AZ01283 MEHTA, STEVE REP 2010 AZ H
H0AZ01325 TOBIN, ANDY HON. REP 2014 AZ H
H0AZ01333 GRESSLEY, FORREST DAYL REP 2010 AZ H
H0AZ03321 PARKER, VERNON REP 2014 AZ H
New Step Switch to editor
Cancel
Add to Recipe
Choose a transformation
Choose transformation
TRIFACTA


Trifacta Wrangler Edit Window
cn16 - Transformer - Trifacta
Campaign Finance 2016 › cn16
Generate Results
Grid Columns Full Dataset - 461.78kB 17 Columns 4,864 Rows 3 Data Types
Columns: All Rows: All
Transformed - 3 Columns
Transformed - 4,859 Rows
Filter in grid
Source to be dropped
Preview
CAND_ID CAND_NAME CAND_NAME1 CAND_NAME2 CAND_PARTY_AFFILIATION CAND_ELECTION_YEAR
4,864 Categories 4,760 Categories 3,416 Categories 3,677 Categories 76 Categories 1986 - 2052
H0AK00097 COX, JOHN R. COX JOHN R. REP 2014 AK
H0AL02087 ROBY, MARTHA ROBY MARTHA REP 2016 AL
H0AL02095 JOHN, ROBERT E JR JOHN ROBERT E JR IND 2016 AL
H0AL05049 CRAMER, ROBERT E "BUD" JR CRAMER ROBERT E "BUD" JR DEM 2008 AL
H0AL05163 BROOKS, MO BROOKS MO REP 2016 AL
H0AL06088 COOKE, STANLEY KYLE COOKE STANLEY KYLE REP 2010 AL
H0AL07086 SEWELL, TERRI A. SEWELL TERRI A. DEM 2016 AL
H0AL07094 HILLIARD, EARL FREDERICK JR HILLIARD EARL FREDERICK JR DEM 2010 AL
H0AL07177 CHAMBERLAIN, DON CHAMBERLAIN DON REP 2012 AL
H0AR01083 CRAWFORD, ERIC ALAN RICK CRAWFORD ERIC ALAN RICK REP 2016 AR
H0AR01091 GREGORY, JAMES CHRISTOPHER GREGORY JAMES CHRISTOPHER DEM 2010 AR
H0AR01109 CAUSEY, CHAD CAUSEY CHAD DEM 2010 AR
H0AR01125 SMITH, PRINCELLA D SMITH PRINCELLA D REP 2010 AR
H0AR02107 GRIFFIN, JOHN TIMOTHY GRIFFIN JOHN TIMOTHY REP 2014 AR
H0AR02131 ELLIOTT, JOYCE ANN ELLIOTT JOYCE ANN DEM 2010 AR
H0AR03022 SKOCH, BERNARD KURT 'BERNIE' SKOCH BERNARD KURT 'BERNIE' REP 2010 AR
H0AR03030 WHITAKER, DAVID JEFFREY WHITAKER DAVID JEFFREY DEM 2010 AR
H0AR03055 WOMACK, STEVE WOMACK STEVE REP 2016 AR
H0AS00018 FALEOMAVAEGA, ENI FALEOMAVAEGA ENI DEM 2014 AS
H0AZ01184 FLAKE, JEFF MR. FLAKE JEFF MR. REP 2012 AZ
H0AZ01259 GOSAR, PAUL ANTHONY GOSAR PAUL ANTHONY REP 2016 AZ
SUGGESTIONS
Cancel
Modify
Add to Recipe
Split CAND_NAME into 2 columns on `{delim-ws}`
CAND_NAME CAND_NAME1 CAND_NAME2
COX, JOHN R. COX JOHN R.
ROBY, MARTHA ROBY MARTHA
JOHN, ROBERT E JR JOHN ROBERT E JR
Affects 1 column, 4859 rows
Creates 2 columns
Extract `{delim-ws}` from CAND_NAME
CAND_NAME CAND_NAME1
COX, JOHN R. ,
ROBY, MARTHA ,
JOHN, ROBERT E JR ,
Affects 1 column, 4859 rows
Creates 1 column
Count occurrences of `{delim
CAND_NAME
COX, JOHN R.
ROBY, MARTHA
JOHN, ROBERT E JR
Affects 1 column, 4859 rows
TRIFACTA


Trifacta Wrangler Edit Window
cn16 - Transformer - Trifacta
Campaign Finance 2016 › cn16
1 → 0
Generate Results
Grid Columns Full Dataset - 461.78kB 15 Columns 4,864 Rows 3 Data Types
Filter in grid
ABC CAND_ID
ABC CAND_NAME
ABC CAND_PARTY_AFFILIATION
CAND_ELECTION_YEAR
ABC CAND_OFFICE_STATE
ABC CAND_OFFICE
4,864 Categories
4,760 Categories
76 Categories
1986 - 2052
57 Categories
3 Categories
H0AK00097 COX, JOHN R. REP 2014 AK H
H0AL02087 ROBY, MARTHA REP 2016 AL H
H0AL02095 JOHN, ROBERT E JR IND 2016 AL H
H0AL05049 CRAMER, ROBERT E "BUD" JR DEM 2008 AL H
H0AL05163 BROOKS, MO REP 2016 AL H
H0AL06088 COOKE, STANLEY KYLE REP 2010 AL H
H0AL07086 SEWELL, TERRI A. DEM 2016 AL H
H0AL07094 HILLIARD, EARL FREDERICK JR DEM 2010 AL H
H0AL07177 CHAMBERLAIN, DON REP 2012 AL H
H0AR01083 CRAWFORD, ERIC ALAN RICK REP 2016 AR H
H0AR01091 GREGORY, JAMES CHRISTOPHER DEM 2010 AR H
H0AR01109 CAUSEY, CHAD DEM 2010 AR H
H0AR01125 SMITH, PRINCELLA D REP 2010 AR H
H0AR02107 GRIFFIN, JOHN TIMOTHY REP 2014 AR H
H0AR02131 ELLIOTT, JOYCE ANN DEM 2010 AR H
H0AR03022 SKOCH, BERNARD KURT 'BERNIE' REP 2010 AR H
H0AR03030 WHITAKER, DAVID JEFFREY DEM 2010 AR H
H0AR03055 WOMACK, STEVE REP 2016 AR H
H0AS00018 FALEOMAVAEGA, ENI DEM 2014 AS H
H0AZ01184 FLAKE, JEFF MR. REP 2012 AZ H
H0AZ01259 GOSAR, PAUL ANTHONY REP 2016 AZ H
H0AZ01283 MEHTA, STEVE REP 2010 AZ H
H0AZ01325 TOBIN, ANDY HON. REP 2014 AZ H
H0AZ01333 GRESSLEY, FORREST DAYL REP 2010 AZ H
H0AZ03321 PARKER, VERNON REP 2014 AZ H
New Step Switch to editor
Cancel
Add to Recipe
Choose a transformation
Choose transformation
TRIFACTA


Trifacta Wrangler Edit Window
cn16 - Transformer - Trifacta
Campaign Finance 2016 › cn16
1 → 0
Generate Results
Grid
Columns
Full Dataset - 461.78kB
15 Columns
4,864 Rows
3 Data Types
Filter in grid
CAND_ID
CAND_NAME
CAND_PARTY_AFFILIATION
CAND_ELECTION_YEAR
CAND_OFFICE_STATE
CAND_OFFICE
4,864 Categories
4,760 Categories
1986 - 2052
57 Categories
3 Categories
Rename
Change type
Edit column
Column Details
Find
Format
Filter
Clean
Formula
Aggregate
Restructure
Lookup...
Drop
H0AK00097 COX, JOHN R. 2014 AK H
H0AL02087 ROBY, MARTHA 2016 AL H
H0AL02095 JOHN, ROBERT E JR 2016 AL H
H0AL05049 CRAMER, ROBERT E "BUD" J 2008 AL H
H0AL05163 BROOKS, MO 2016 AL H
H0AL06088 COOKE, STANLEY KYLE 2010 AL H
H0AL07086 SEWELL, TERRI A. 2016 AL H
H0AL07094 HILLIARD, EARL FREDERICK 2010 AL H
H0AL07177 CHAMBERLAIN, DON 2012 AL H
H0AR01083 CRAWFORD, ERIC ALAN RICK 2016 AR H
H0AR01091 GREGORY, JAMES CHRISTOPH 2010 AR H
H0AR01109 CAUSEY, CHAD 2010 AR H
H0AR01125 SMITH, PRINCELLA D 2010 AR H
H0AR02107 GRIFFIN, JOHN TIMOTHY 2014 AR H
H0AR02131 ELLIOTT, JOYCE ANN 2010 AR H
H0AR03022 SKOCH, BERNARD KURT 'BER 2010 AR H
H0AR03030 WHITAKER, DAVID JEFFREY 2010 AR H
H0AR03055 WOMACK, STEVE REP 2016 AR H
H0AS00018 FALEOMAVAEGA, ENI DEM 2014 AS H
H0AZ01184 FLAKE, JEFF MR. REP 2012 AZ H
H0AZ01259 GOSAR, PAUL ANTHONY REP 2016 AZ H
H0AZ01283 MEHTA, STEVE REP 2010 AZ H
H0AZ01325 TOBIN, ANDY HON. REP 2014 AZ H
H0AZ01333 GRESSLEY, FORREST DAYL REP 2010 AZ H
H0AZ03321 PARKER, VERNON REP 2014 AZ H
New Step
Switch to editor
Cancel
Add to Recipe
Choose a transformation
Choose transformation
TRIFACTA

Trifacta Wrangler Edit Window

cn16 - Transformer - Trifacta

Campaign Finance 2016 › cn16

1 → → 0 Generate Results

Grid Columns Full Dataset - 461.78kB 15 Columns 4,864 Rows 3 Data Types

CAND_NAME

SUMMARY

Valid	4,863	100.0%
Unique	4,760	97.9%
Outliers	18	0.4%
Mismatched	0	0.0%
Missing	1	0.0%

STRING LENGTH STATISTICS

Minimum	4.00
Lower Quartile	14.00
Median	18.00
Upper Quartile	21.00
Maximum	70.00
Average	18.14
Standard Deviation	4.99

TOP VALUES

KALEMKARIAN, TIMOTHY CHARLES	3
MARTIN, ANDY	3
AGBEDE, AKINYEMI	2
AKIN, W TODD	2
ARMSTRONG, BRANDON CHRISTINA	2
BACHMANN, MICHELE	2
BALDWIN, TAMMY	2
BARR, BOB	2
BATES, DON JR	2
BELLIS, JOSEPH K III	2
BICKELMEYER, MICHAEL	2
BLASS, PIOTR DR	2
BLUNT, ROY	2
BOSS, JEFF	2

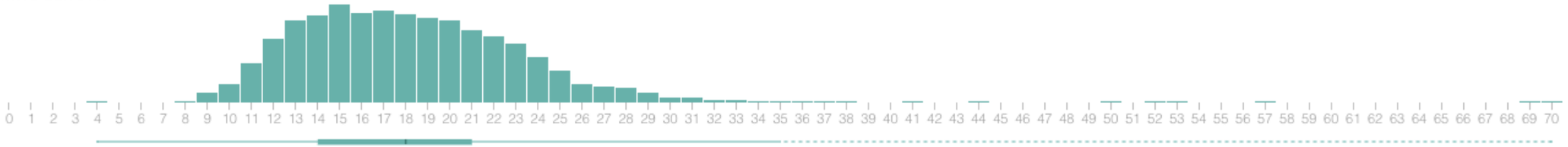
MISMATCHED VALUES

None

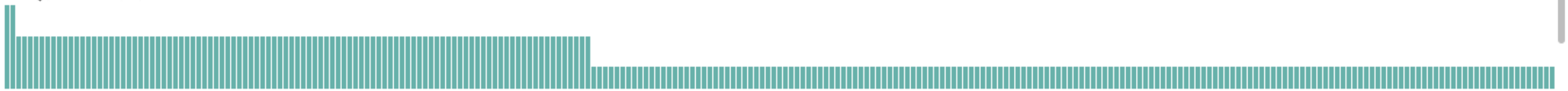
STRING LENGTH OUTLIERS

AAAAAAAAAAAAAAAAAAAAAAAAAAAAA…	1
AKA THE PROPHET AKA EARL, TRIP…	1
CLARKSON, JEREMY CHARLES ROBER…	1
CONNOLLY, MATTHEW DONALD (MATT…	1
DE BUONAPARTE, HRM CAESAR ST A…	1
EASTON, EARNEST LEE PROFESSOR …	1

STRING LENGTH



FREQUENT VALUES



New Step Switch to editor

Cancel Add to Recipe

Choose a transformation

Choose transformation

TRIFACTA

Trifacta Wrangler Edit Window

cn16 - Transformer - Trifacta

Campaign Finance 2016 › cn16

1 → → 0 Generate Results

Grid Columns Full Dataset - 461.78kB 15 Columns 4,864 Rows 3 Data Types

CAND_NAME

SUMMARY

Valid	4,863	100.0%
Unique	4,760	97.9%
Outliers	18	0.4%
Mismatched	0	0.0%
Missing	1	0.0%

STRING LENGTH STATISTICS

Minimum	4.00
Lower Quartile	14.00
Median	18.00
Upper Quartile	21.00
Maximum	70.00
Average	18.14
Standard Deviation	4.99

TOP VALUES

KALEMKARIAN, TIMOTHY CHARLES	3
MARTIN, ANDY	3
AGBEDE, AKINYEMI	2
AKIN, W TODD	2
ARMSTRONG, BRANDON CHRISTINA	2
BACHMANN, MICHELE	2
BALDWIN, TAMMY	2
BARR, BOB	2
BATES, DON JR	2
BELLIS, JOSEPH K III	2
BICKELMEYER, MICHAEL	2
BLASS, PIOTR DR	2
BLUNT, ROY	2
BOSS, JEFF	2

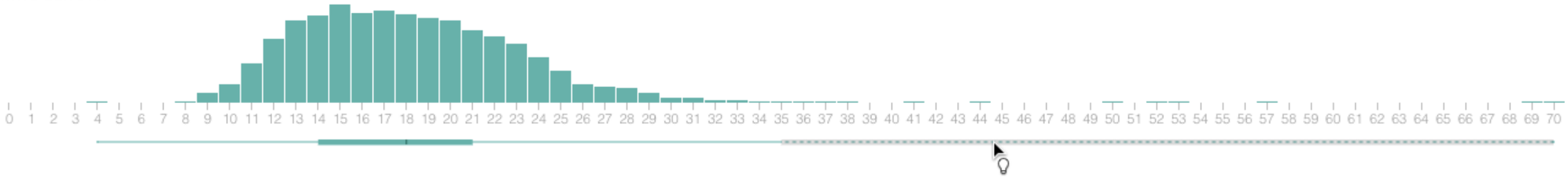
MISMATCHED VALUES

None

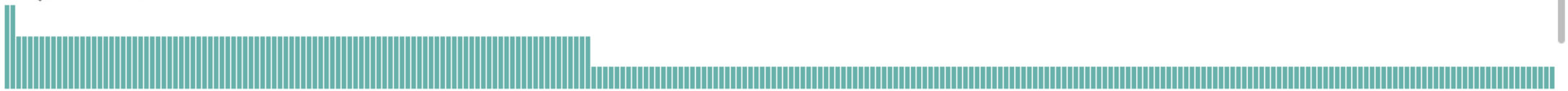
STRING LENGTH OUTLIERS

AAAAAAAAAAAAAAAAAAAAAAAAAAAAA…	1
AKA THE PROPHET AKA EARL, TRIP…	1
CLARKSON, JEREMY CHARLES ROBER…	1
CONNOLLY, MATTHEW DONALD (MATT…	1
DE BUONAPARTE, HRM CAESAR ST A…	1
EASTON, EARNEST LEE PROFESSOR …	1

STRING LENGTH



FREQUENT VALUES



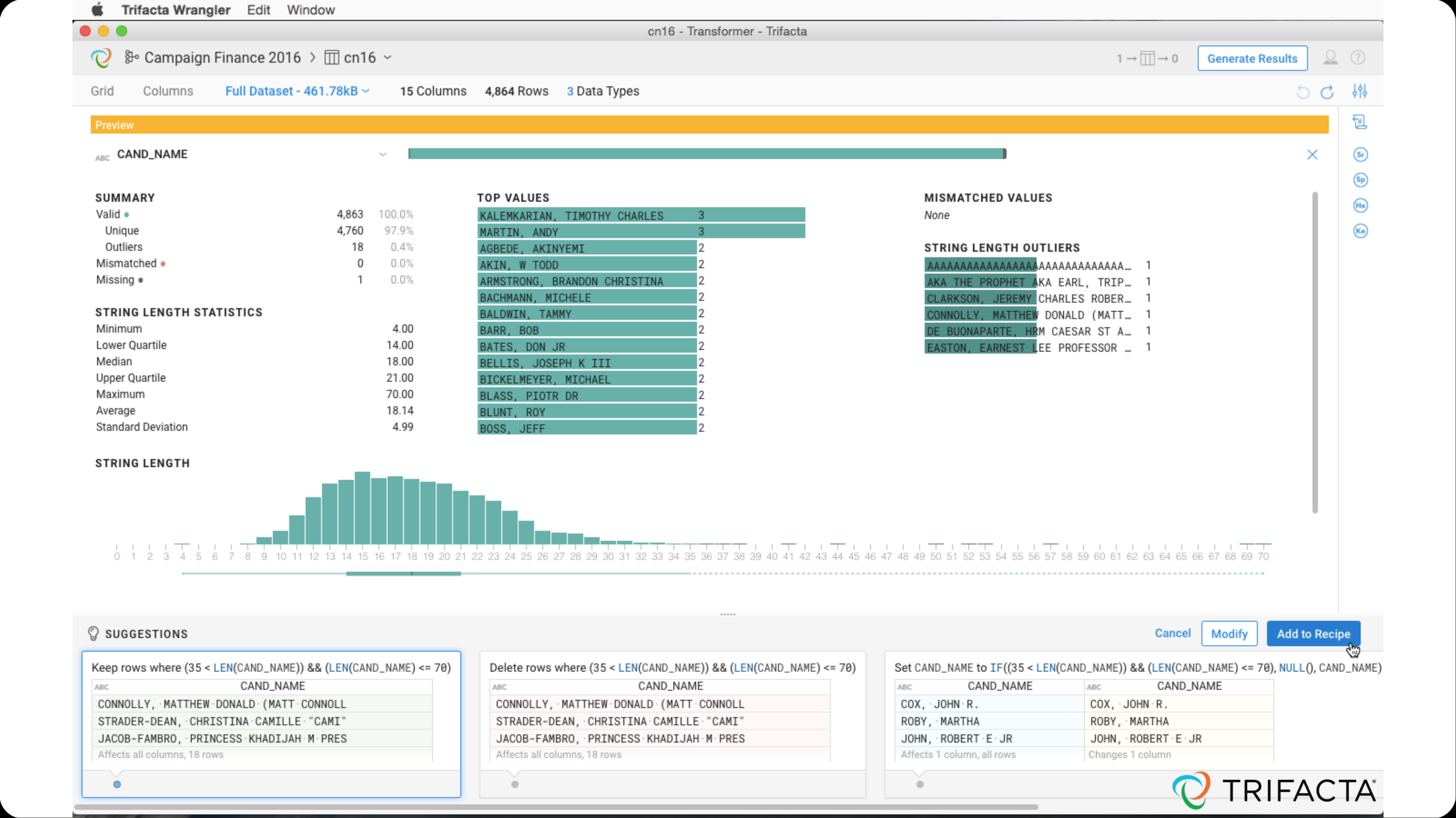
New Step Switch to editor

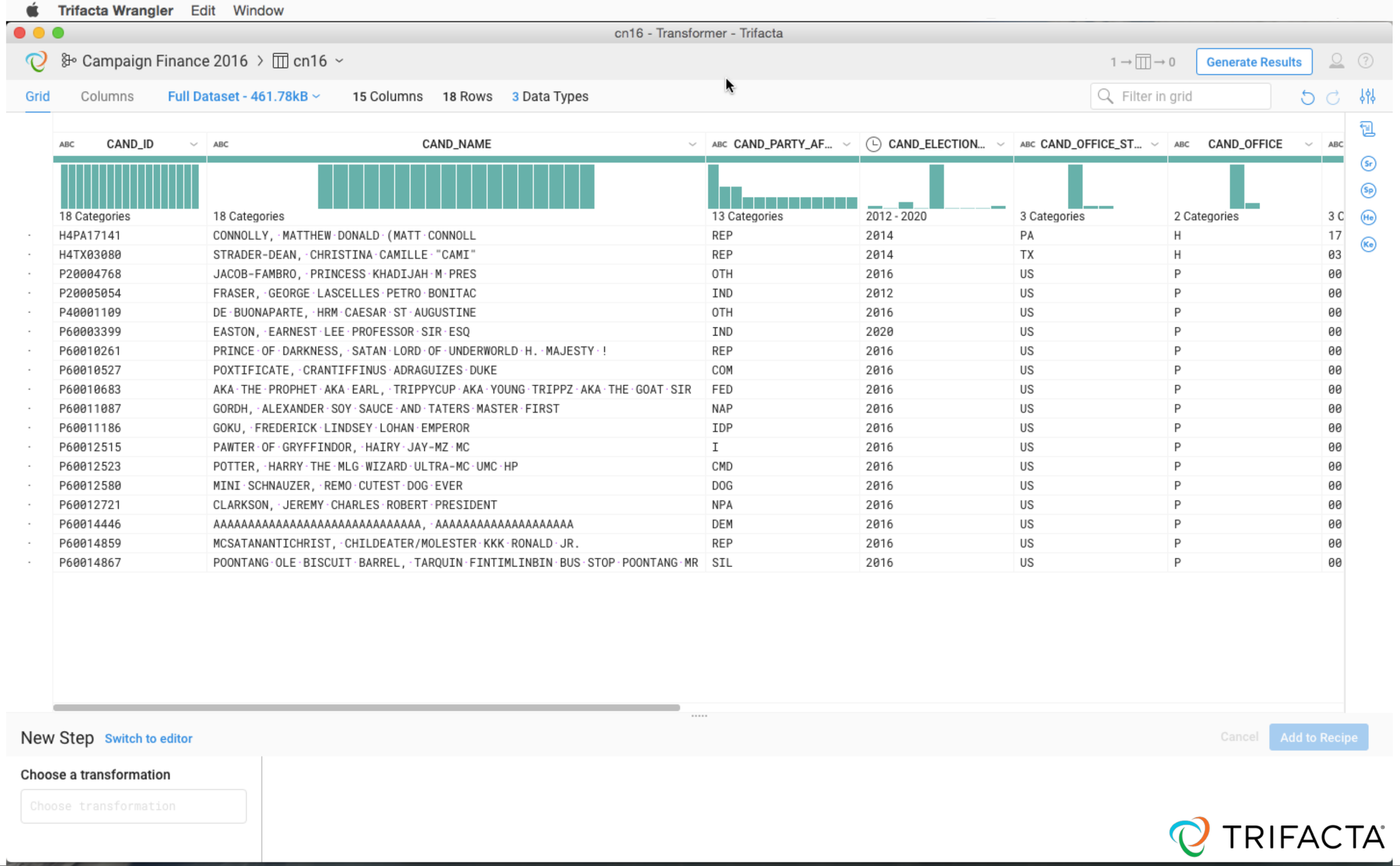
Cancel Add to Recipe

Choose a transformation

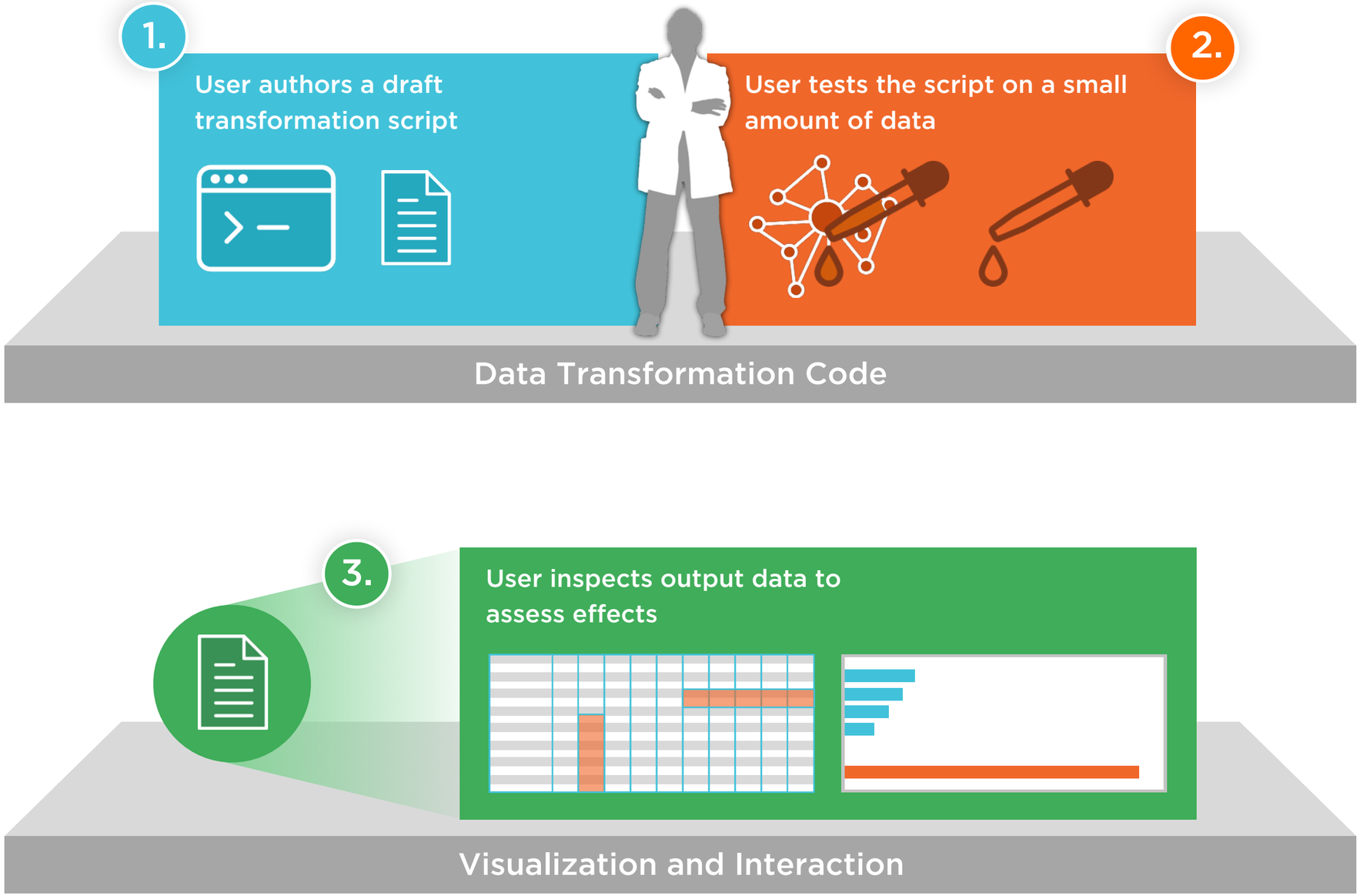
Choose transformation

TRIFACTA

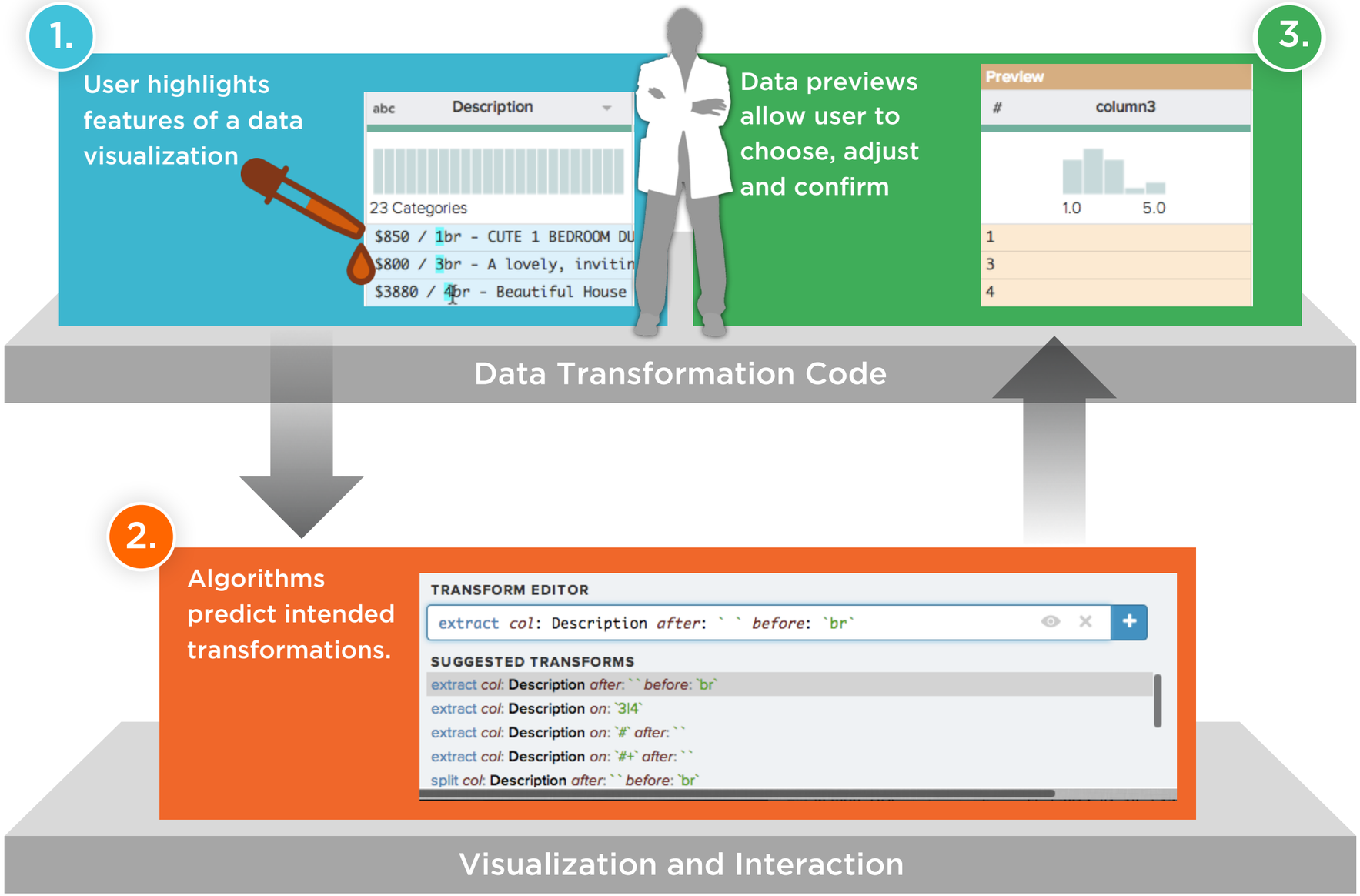

Trifacta Wrangler Edit Window
cn16 - Transformer - Trifacta
Campaign Finance 2016 › cn16
1 → → 0
Generate Results
Grid Columns Full Dataset - 461.78kB 15 Columns 4,864 Rows 3 Data Types
Preview
CAND_NAME
SUMMARY
Valid 4,863 100.0%
Unique 4,760 97.9%
Outliers 18 0.4%
Mismatched 0 0.0%
Missing 1 0.0%
STRING LENGTH STATISTICS
Minimum 4.00
Lower Quartile 14.00
Median 18.00
Upper Quartile 21.00
Maximum 70.00
Average 18.14
Standard Deviation 4.99
TOP VALUES
KALEMKARIAN, TIMOTHY CHARLES 3
MARTIN, ANDY 3
AGBEDE, AKINYEMI 2
AKIN, W TODD 2
ARMSTRONG, BRANDON CHRISTINA 2
BACHMANN, MICHELE 2
BALDWIN, TAMMY 2
BARR, BOB 2
BATES, DON JR 2
BELLIS, JOSEPH K III 2
BICKELMEYER, MICHAEL 2
BLASS, PIOTR DR 2
BLUNT, ROY 2
BOSS, JEFF 2
MISMATCHED VALUES
None
STRING LENGTH OUTLIERS
AAAAAAAAAAAAAAAAAAAAAAAAAAAAAA… 1
AKA THE PROPHET AKA EARL, TRIP… 1
CLARKSON, JEREMY CHARLES ROBER… 1
CONNOLLY, MATTHEW DONALD (MATT… 1
DE BUONAPARTE, HRM CAESAR ST A… 1
EASTON, EARNEST LEE PROFESSOR … 1
STRING LENGTH
0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60 61 62 63 64 65 66 67 68 69 70
SUGGESTIONS
Cancel
Modify
Add to Recipe
Keep rows where (35 < LEN(CAND_NAME)) && (LEN(CAND_NAME) <= 70)
CAND_NAME
CONNOLLY, MATTHEW DONALD (MATT CONNOLL
STRADER-DEAN, CHRISTINA CAMILLE "CAMI"
JACOB-FAMBRO, PRINCESS KHADIJAH M PRES
Affects all columns, 18 rows
Delete rows where (35 < LEN(CAND_NAME)) && (LEN(CAND_NAME) <= 70)
CAND_NAME
CONNOLLY, MATTHEW DONALD (MATT CONNOLL
STRADER-DEAN, CHRISTINA CAMILLE "CAMI"
JACOB-FAMBRO, PRINCESS KHADIJAH M PRES
Affects all columns, 18 rows
Set CAND_NAME to IF((35 < LEN(CAND_NAME)) && (LEN(CAND_NAME) <= 70), NULL(), CAND_NAME)
CAND_NAME
COX, JOHN R.
ROBY, MARTHA
JOHN, ROBERT E JR
Affects 1 column, all rows
CAND_NAME
COX, JOHN R.
ROBY, MARTHA
JOHN, ROBERT E JR
Changes 1 column
TRIFACTA


Trifacta Wrangler Edit Window
cn16 - Transformer - Trifacta
Campaign Finance 2016 › cn16
1 → 0
Generate Results
Grid
Columns
Full Dataset - 461.78kB
15 Columns
18 Rows
3 Data Types
Filter in grid
ABC CAND_ID
ABC CAND_NAME
ABC CAND_PARTY_AF...
CAND_ELECTION...
ABC CAND_OFFICE_ST...
ABC CAND_OFFICE
18 Categories
18 Categories
13 Categories
2012 - 2020
3 Categories
2 Categories
H4PA17141 CONNOLLY, MATTHEW DONALD (MATT CONNOLL REP 2014 PA H 17
H4TX03080 STRADER-DEAN, CHRISTINA CAMILLE "CAMI" REP 2014 TX H 03
P20004768 JACOB-FAMBRO, PRINCESS KHADIJAH M PRES OTH 2016 US P 00
P20005054 FRASER, GEORGE LASCELLES PETRO BONITAC IND 2012 US P 00
P40001109 DE BUONAPARTE, HRM CAESAR ST AUGUSTINE OTH 2016 US P 00
P60003399 EASTON, EARNEST LEE PROFESSOR SIR ESQ IND 2020 US P 00
P60010261 PRINCE OF DARKNESS, SATAN LORD OF UNDERWORLD H. MAJESTY ! REP 2016 US P 00
P60010527 POXTIFICATE, CRANTIFFINUS ADRAGUIZES DUKE COM 2016 US P 00
P60010683 AKA THE PROPHET AKA EARL, TRIPPYCUP AKA YOUNG TRIPPZ AKA THE GOAT SIR FED 2016 US P 00
P60011087 GORDH, ALEXANDER SOY SAUCE AND TATERS MASTER FIRST NAP 2016 US P 00
P60011186 GOKU, FREDERICK LINDSEY LOHAN EMPEROR IDP 2016 US P 00
P60012515 PAWTER OF GRYFFINDOR, HAIRY JAY-MZ MC I 2016 US P 00
P60012523 POTTER, HARRY THE MLG WIZARD ULTRA-MC UMC HP CMD 2016 US P 00
P60012580 MINI SCHNAUZER, REMO CUTEST DOG EVER DOG 2016 US P 00
P60012721 CLARKSON, JEREMY CHARLES ROBERT PRESIDENT NPA 2016 US P 00
P60014446 AAAAAAAAAAAAAAAAAAAAAAAAAAAA, AAAAAAAAAAAAAAAAAA DEM 2016 US P 00
P60014859 MCSATANANTICHRIST, CHILDEATER/MOLESTER KKK RONALD JR. REP 2016 US P 00
P60014867 POONTANG OLE BISCUIT BARREL, TARQUIN FINTIMLINBIN BUS STOP POONTANG MR SIL 2016 US P 00
New Step
Switch to editor
Cancel
Add to Recipe
Choose a transformation
Choose transformation
TRIFACTA

Traditional Specification



Predictive Interaction



Parting Thoughts

Interactive Data Analysis Tools

Interactive Data Analysis Tools

A careful balance of **automation** and **control**.

Interactive Data Analysis Tools

A careful balance of **automation** and **control**.

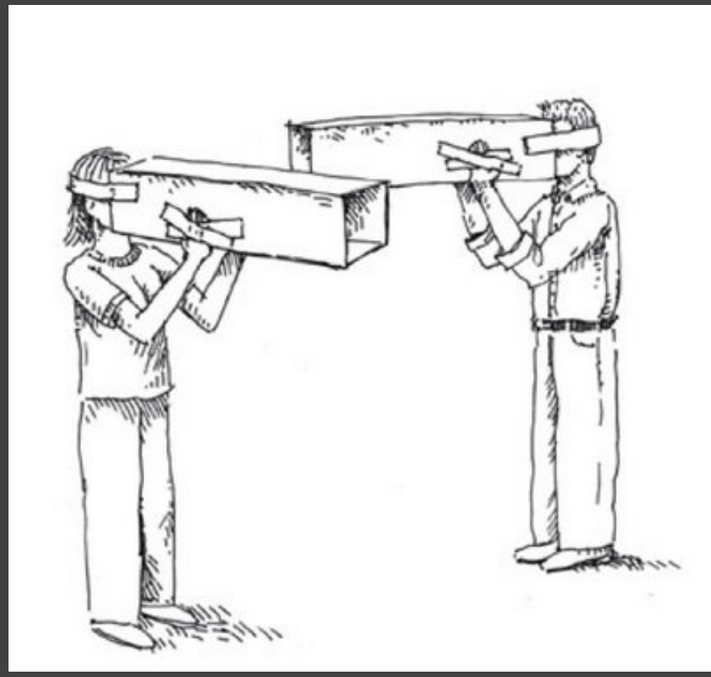
Pitfalls of automation:
Loss of agency, intuition & domain expertise.
Consequences of poor models let loose in the world.

Interactive Data Analysis Tools

A careful balance of **automation** and **control**.

Pitfalls of automation:
Loss of agency, intuition & domain expertise.
Consequences of poor models let loose in the world.

Pitfalls of user control:
Cognitive biases, blinder vision, mistakes.



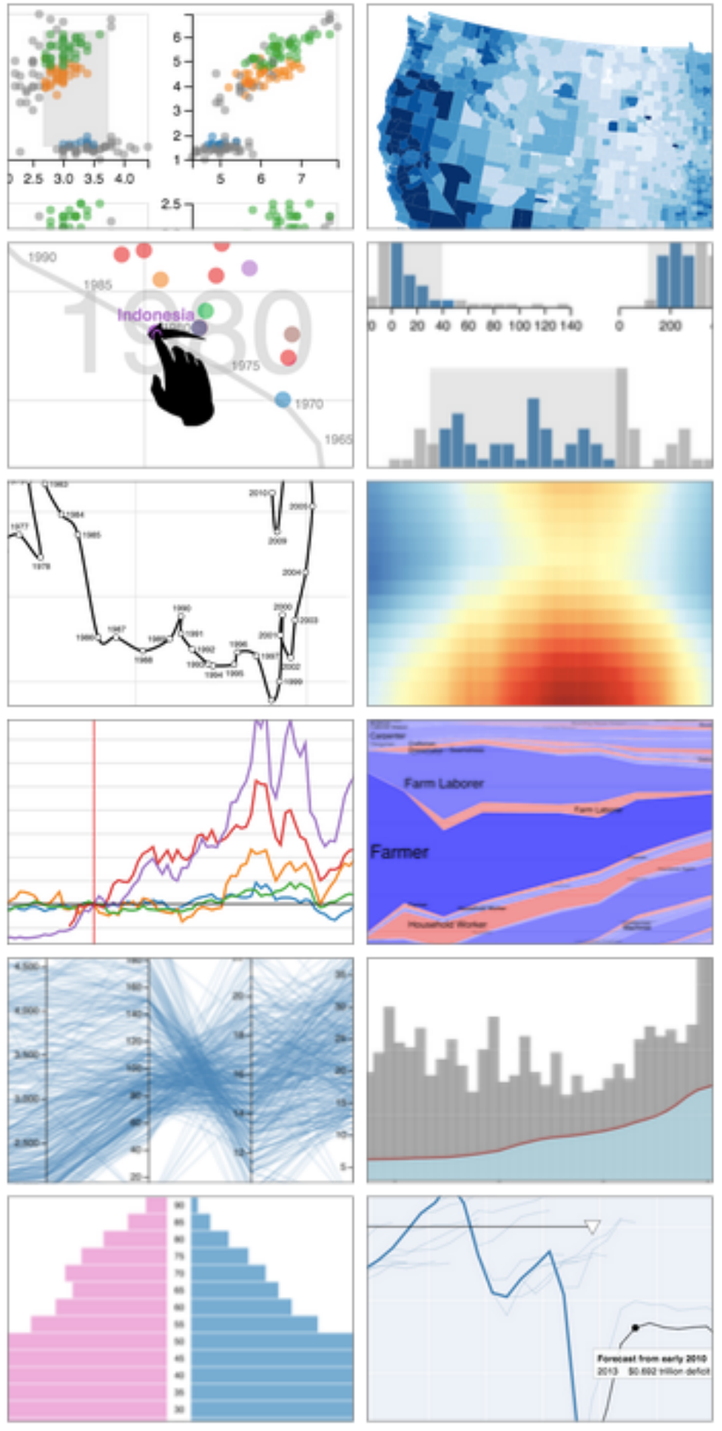
Interactive Data Analysis Tools

A careful balance of **automation** and **control**.

Pitfalls of automation:
Loss of agency, intuition & domain expertise.
Consequences of poor models let loose in the world.

Pitfalls of user control:
Cognitive biases, blinder vision, mistakes.

Enhance interfaces with **models of capabilities, actions & goals** to present relevant options.



Interactive Data Analysis Tools

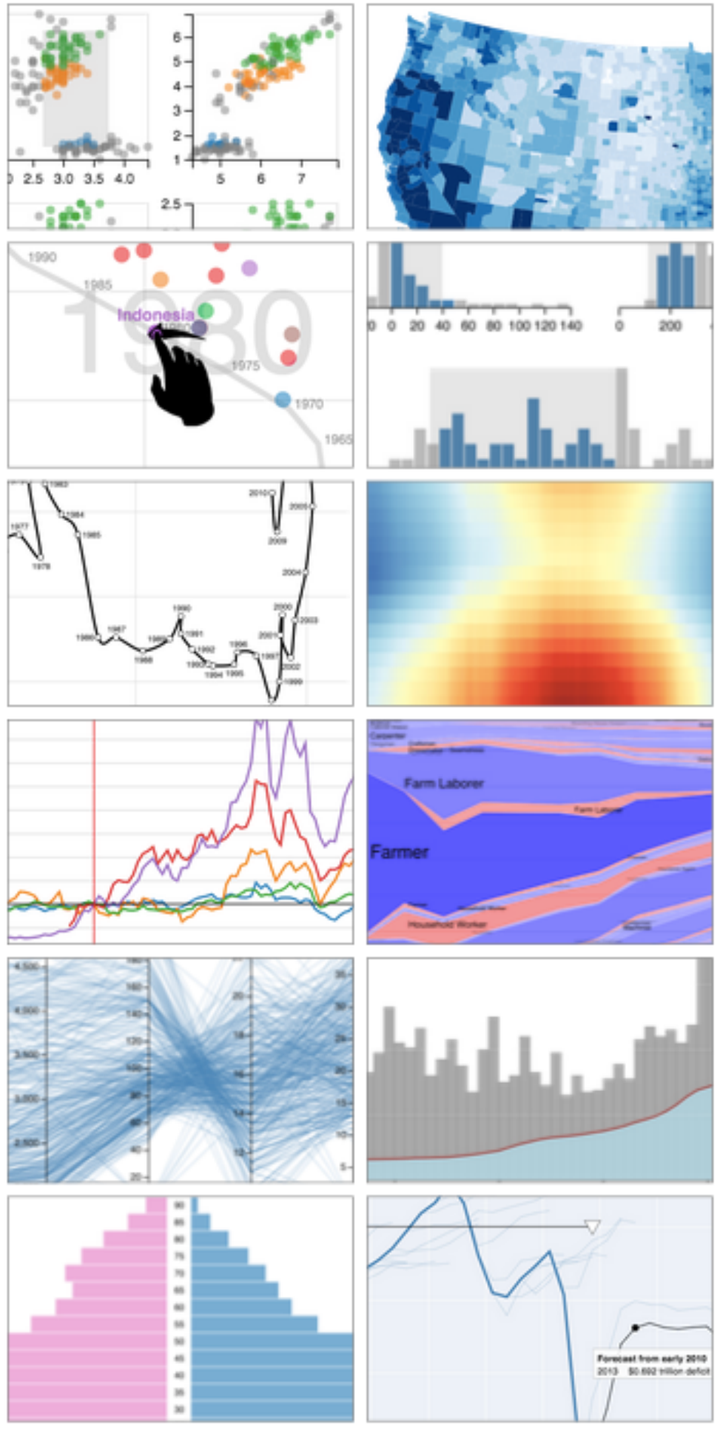
A careful balance of **automation** and **control**.

Pitfalls of automation:
Loss of agency, intuition & domain expertise.
Consequences of poor models let loose in the world.

Pitfalls of user control:
Cognitive biases, blinder vision, mistakes.

Enhance interfaces with **models of capabilities, actions & goals** to present relevant options.

Models adapt over time: **interfaces that learn**.



Interactive Data Analysis Tools

A careful balance of **automation** and **control**.

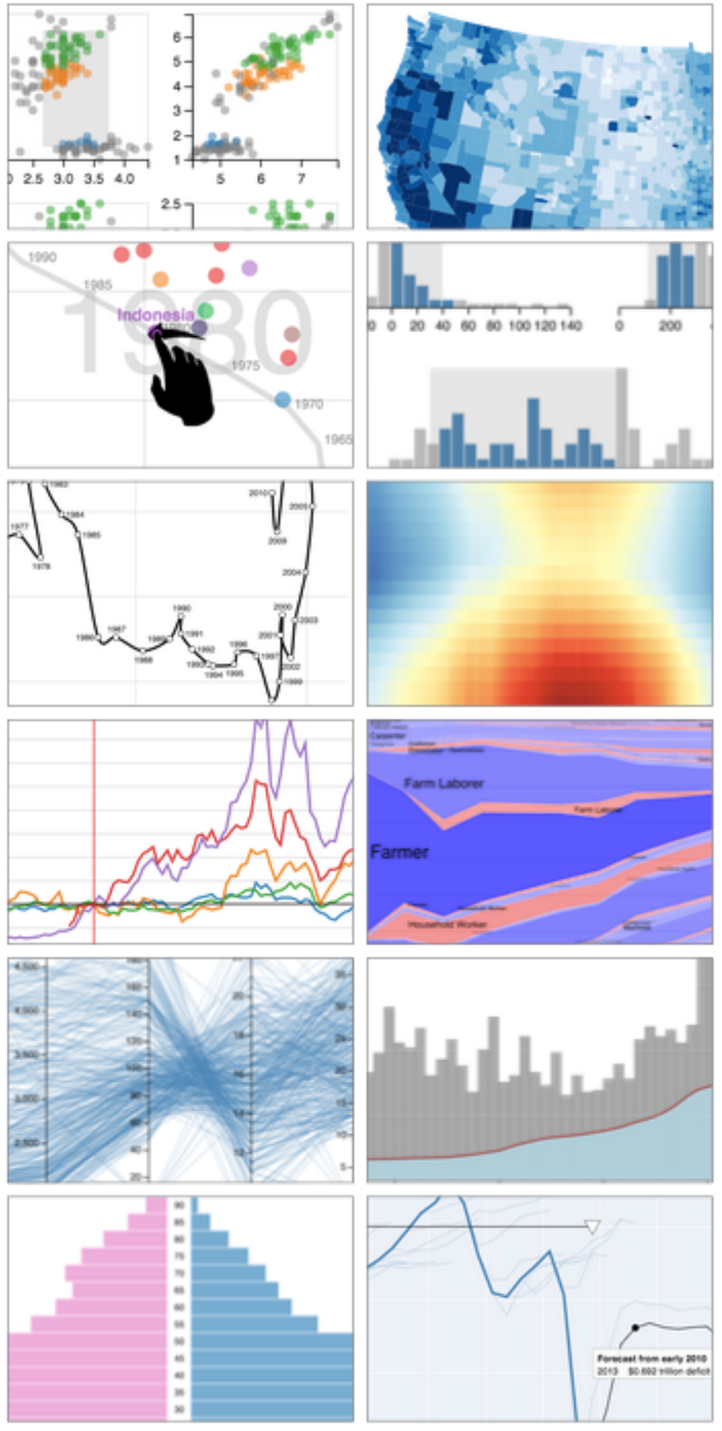
Pitfalls of automation:
Loss of agency, intuition & domain expertise.
Consequences of poor models let loose in the world.

Pitfalls of user control:
Cognitive biases, blinder vision, mistakes.

Enhance interfaces with **models of capabilities, actions & goals** to present relevant options.

Models adapt over time: **interfaces that learn**.

Requires means to **inspect, monitor & audit** models!





idL

ISTC BIG DATA

DARPA

GORDON AND BETTY MOORE FOUNDATION

THE PAUL G. ALLEN FAMILY FOUNDATION

INTERACTIVE

Data Analysis

Jeffrey Heer @jeffrey_heer
http://idl.cs.washington.edu



