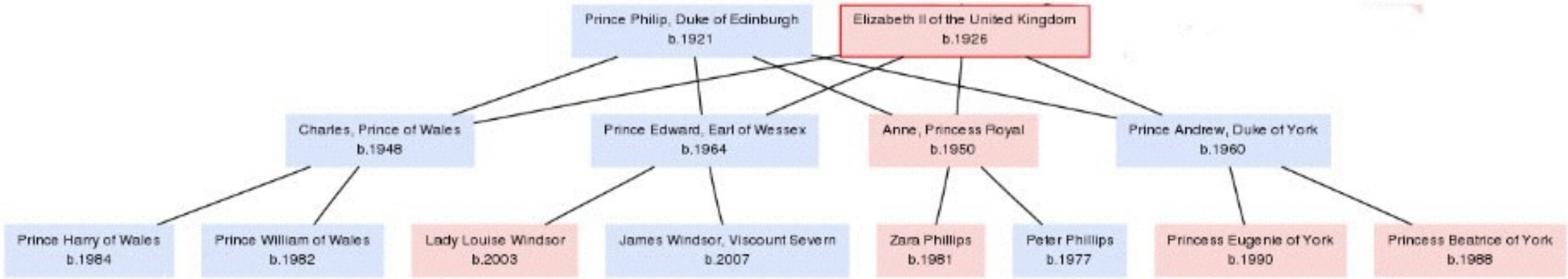
DESIGNING *WITH*

DATA

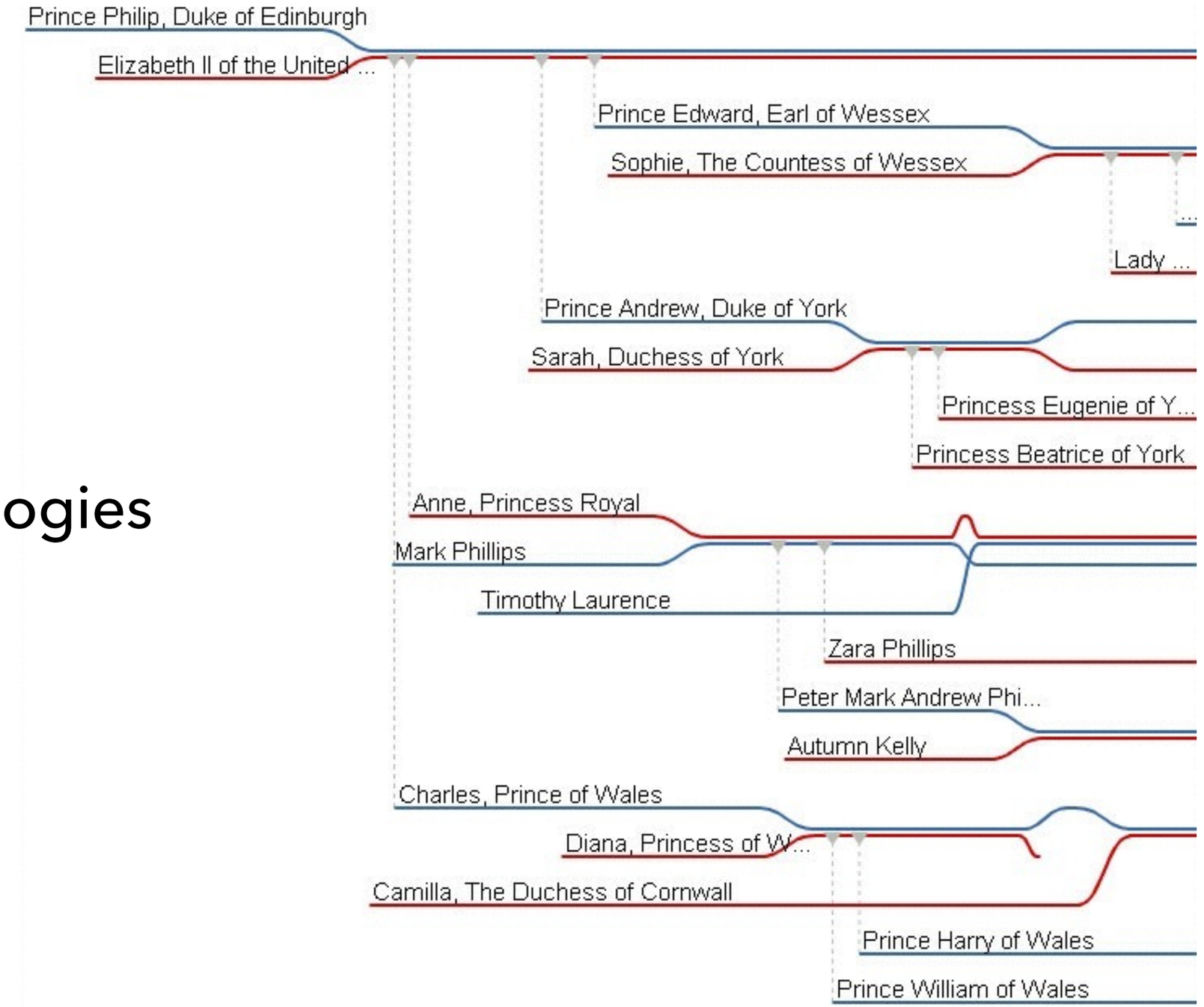
Jeffrey Heer @jeffrey_heer
Trifacta Inc. + U. Washington

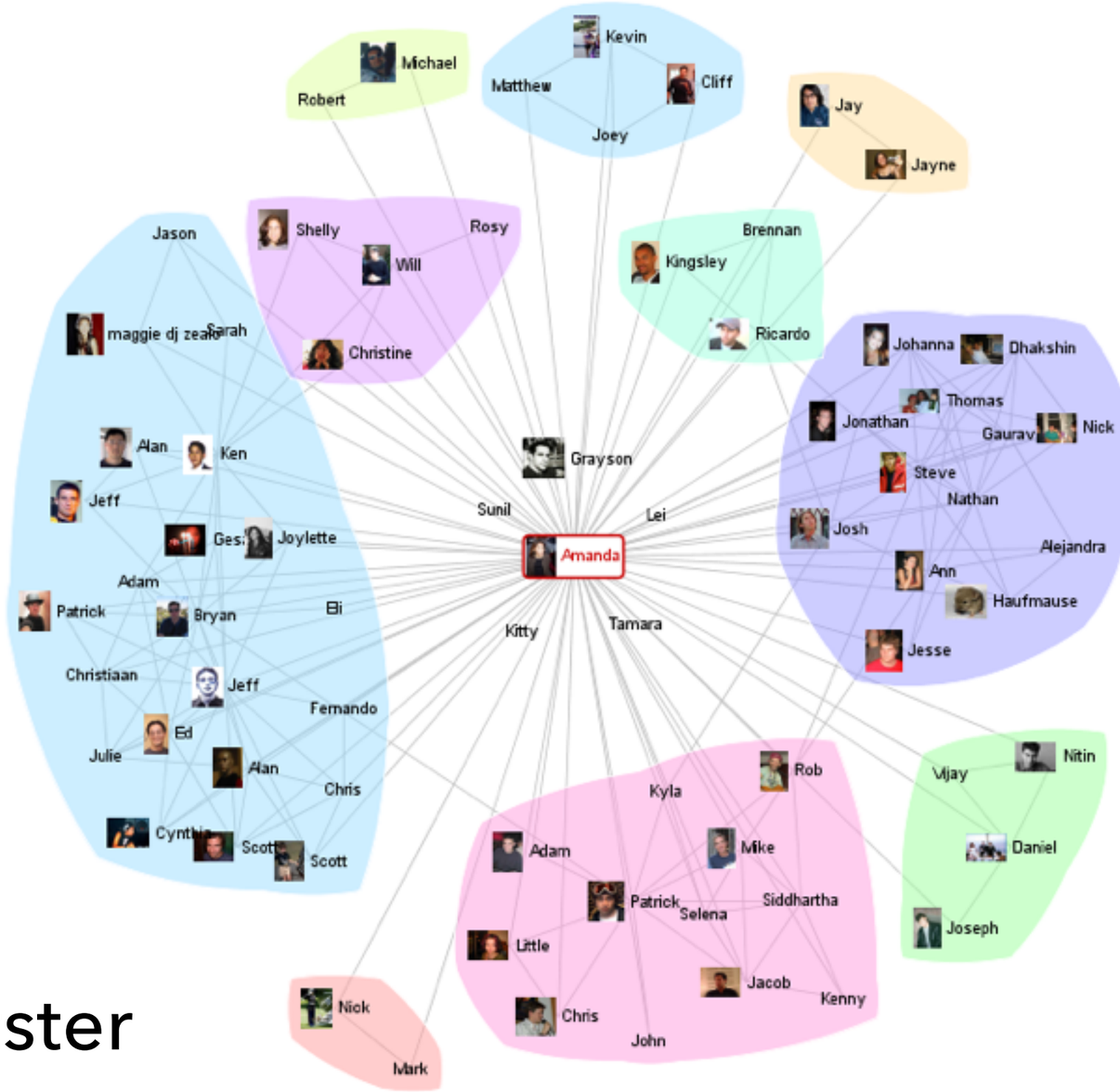
Data Sets I Have ***Known*** and ***Loved***

Genealogies

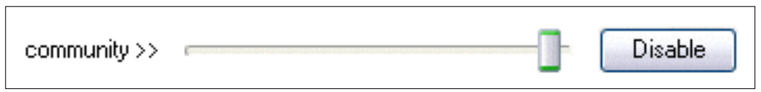


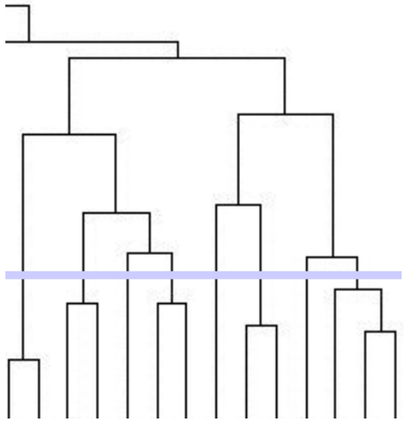
TimeNet Genealogies




Kevin
Michael
Robert
Matthew
Cliff
Joey
Jay
Jayne
Shelly
Rosy
Will
Christine
Jason
Brennan
Kingsley
Ricardo
maggie dj zeal
Sarah
Johanna
Dhakshin
Thomas
Jonathan
Gaurav
Nick
Steve
Nathan
Josh
Alejandra
Ann
Haufmause
Jesse
Grayson
Sunil
Lei
Amanda
Kitty
Tamara
Alan
Ken
Jeff
Ges
Joylette
Adam
Patrick
Bryan
Bi
Christiaan
Jeff
Fernando
Ed
Julie
Alan
Chris
Cynthia
Scott
Scott
Rob
Kyla
Adam
Mike
Patrick
Selena
Siddhartha
Little
Jacob
Kenny
Chris
John
Nick
Mark
Vijay
Nitin
Daniel
Joseph

Vizster


community >>
Disable



Enron Corpus Viewer

File Display Tools

connectivity >>

community >>

Enable

time >>

1/20/01 - 6/27/01

Feb Mar Apr May Jun

search >> california

search >> ferc

Jeff Dasovich

sarah.novosel@enron.com

Susan Mara

john.sheik@enron.com

dwatkiss@br

Tim Belden

Tom Heatson

joe.hartsoe@enron.com

Alan Co

Robert Frank

Linda Robertson

mary.hain@enr

miyung.buster@enron.com

james.steffes@enron.com

paul.kaufman@enron.com

Ray Alvarez

Richard Shapiro

Dan Leff

mark.palmer@enron.com

mona.petrochko@enron.com

ann.schmidt@enron.com

Tim Belden

2001-06-06 16:48:00.0 ISO's Response to BPA Rebuttal of Sh

2001-06-07 11:00:00.0 Legislative Update -- Two Track In The

2001-06-18 00:15:00.0 White House To Support FERC Action

2001-06-19 04:22:00.0 NEWS FLASH ON THIS MORNING'S

2001-06-20 10:37:00.0 Today's Senate Hearing

2001-06-21 02:15:00.0 More on FERC Refunds

0 0 0 0 0 0

that the FERC order

ing from the western

), for refunds

In essence, they are

ew order applied to

e order.

flavor is that Sen.

for deregulatioin

(by this point the

order applied only

t tried to

that Con Ed has

York and he could

wski said he had

from four western governors -- those from Arizona, North Dakota, Utah and Wyoming -- saying that since FERC has acted, there is no need for Congress to pursue price control legislation.

There were a series of questions and comments on details and technical aspects of the orders. I will do an e-mail on these items later today.

Please advise if you have any questions or comments.

Messages

Enron 'Mastermind' Pleads Guilty

SAN FRANCISCO, Oct. 17, 2002

Deputy Attorney General Larry Thompson, center, head of the Justice Departments Corporate Fraud Task Force, comments Thursday on the guilty plea by Timothy N. Belden, Enron's chief energy trader. (Photo: CBS/AP)

(AP) A former top energy trader, considered the mastermind of Enron Corp.'s scheme to drive up California's energy prices, pleaded guilty Thursday to a federal conspiracy charge.

Timothy Belden, the former head of trading in Enron's Portland, Ore., office, admitted to one count of conspiracy to commit wire fraud and promised to cooperate with state and federal prosecutors as well as any non-criminal effort to investigate the energy industry.

"I did it because I was trying to maximize profit for Enron," Belden told U.S. District Judge Martin Jenkins.

Enronic

Having real data as part of your project is as important as having real users look at your project. I have seen people … create mockups with fake data, show them to real users, get all sorts of feedback, and in the end it's not worth anything.

It's almost as if the **data is one of the stakeholders** in the project and you need its input from the beginning.

Martin Wattenberg [ACM Queue '09]

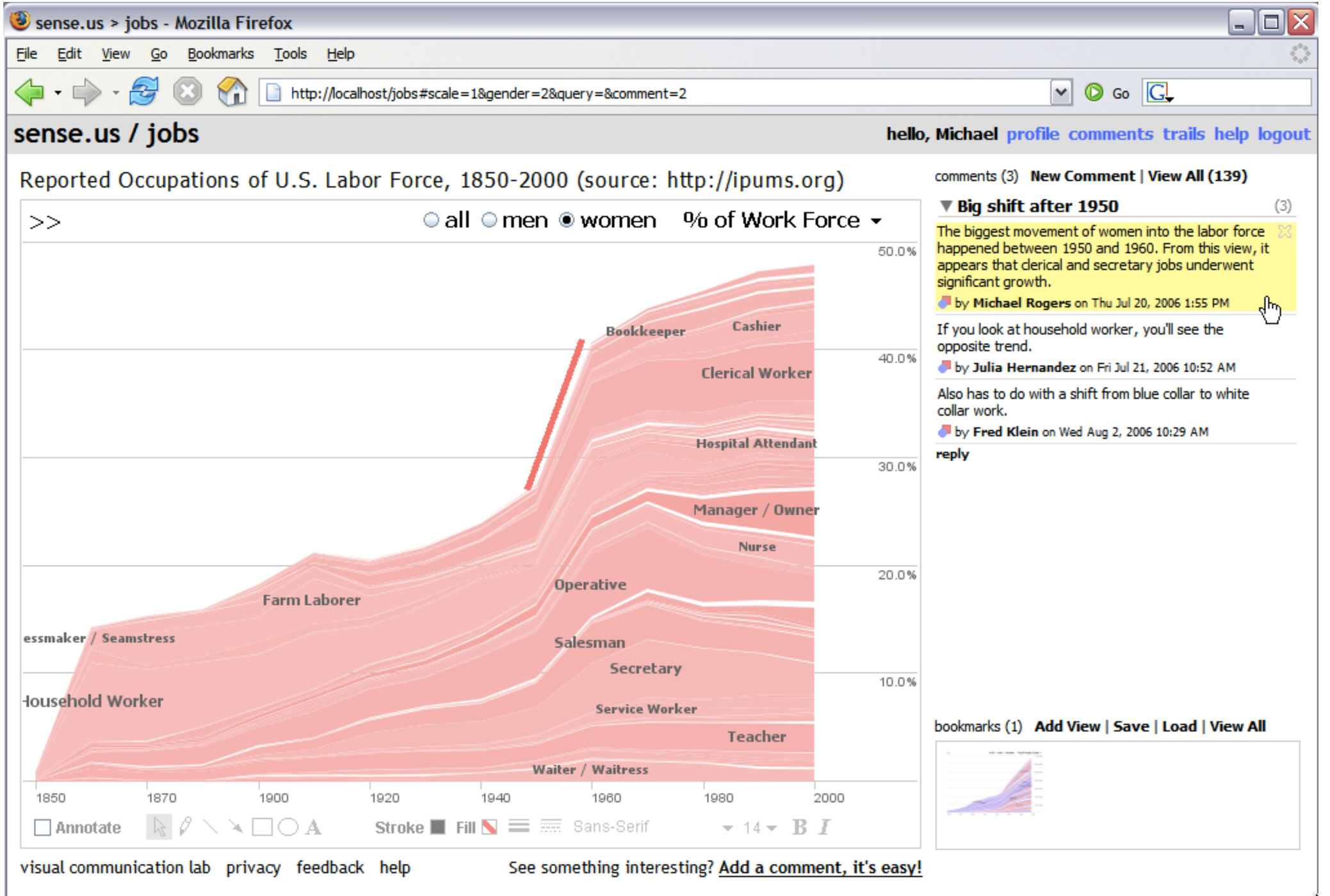
prefuse

Protovis

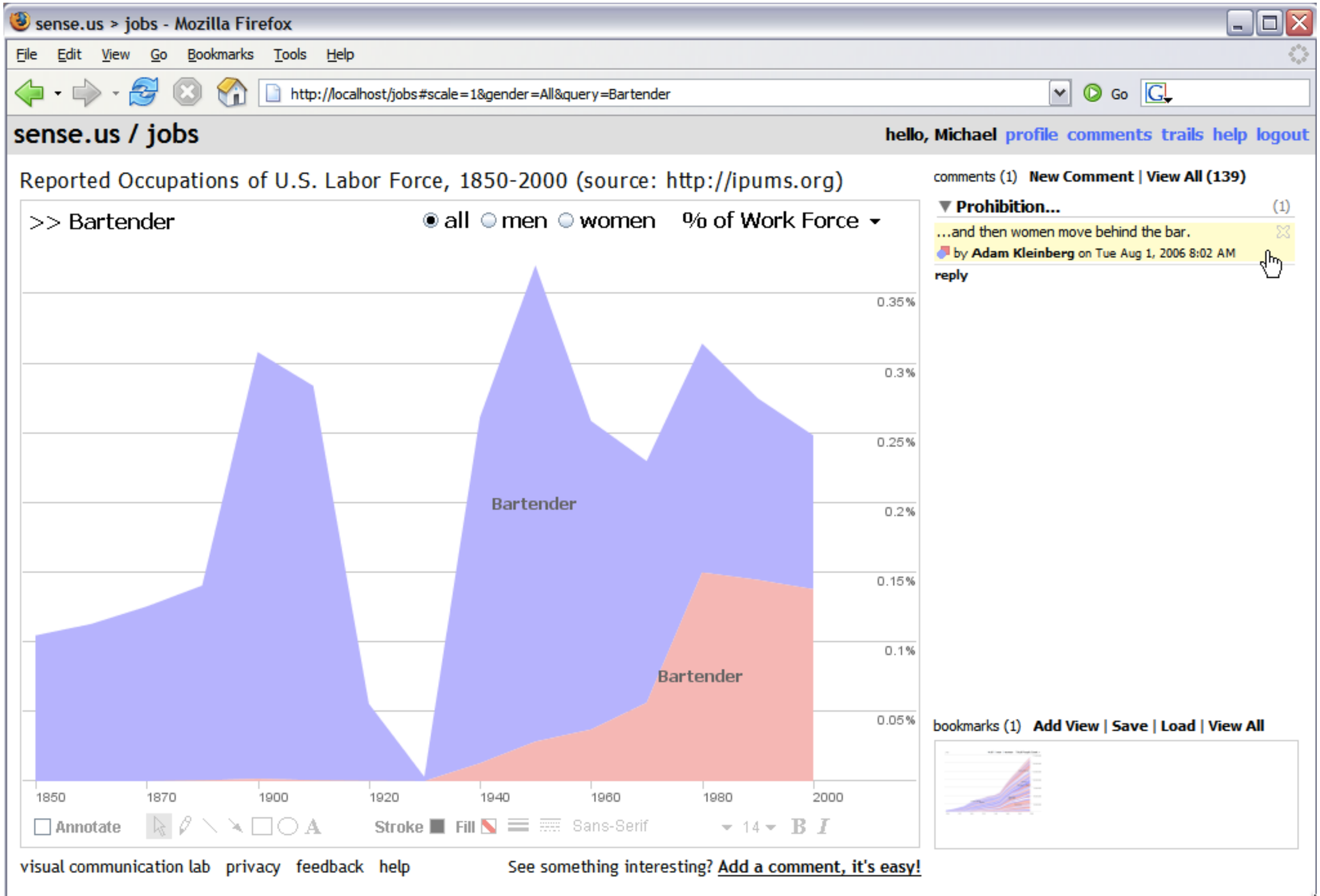
flare

Data-Driven Documents

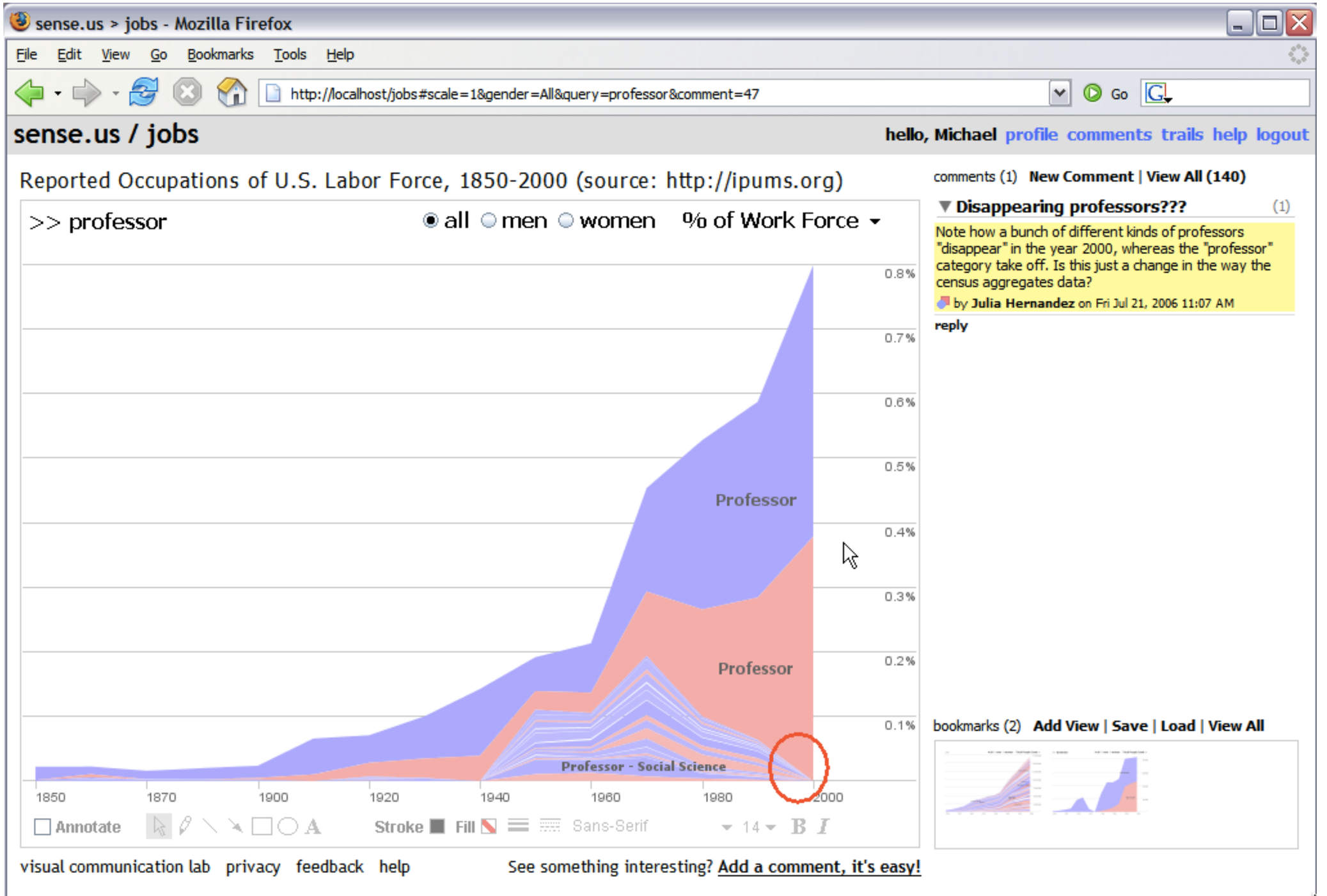
sense.us


sense.us > jobs - Mozilla Firefox
File Edit View Go Bookmarks Tools Help
http://localhost/jobs#scale=1&gender=2&query=&comment=2
Go
sense.us / jobs
hello, Michael profile comments trails help logout
Reported Occupations of U.S. Labor Force, 1850-2000 (source: http://ipums.org)
>>
all men women % of Work Force
50.0%
40.0%
30.0%
20.0%
10.0%
Bookkeeper
Cashier
Clerical Worker
Hospital Attendant
Manager / Owner
Nurse
Operative
Farm Laborer
essmaker / Seamstress
Salesman
Secretary
Household Worker
Service Worker
Teacher
Waiter / Waitress
1850
1870
1900
1920
1940
1960
1980
2000
Annotate
Stroke
Fill
Sans-Serif
14
comments (3) New Comment | View All (139)
Big shift after 1950 (3)
The biggest movement of women into the labor force happened between 1950 and 1960. From this view, it appears that clerical and secretary jobs underwent significant growth.
by Michael Rogers on Thu Jul 20, 2006 1:55 PM
If you look at household worker, you'll see the opposite trend.
by Julia Hernandez on Fri Jul 21, 2006 10:52 AM
Also has to do with a shift from blue collar to white collar work.
by Fred Klein on Wed Aug 2, 2006 10:29 AM
reply
bookmarks (1) Add View | Save | Load | View All
visual communication lab privacy feedback help
See something interesting? Add a comment, it's easy!

sense.us


sense.us > jobs - Mozilla Firefox
File Edit View Go Bookmarks Tools Help
http://localhost/jobs#scale=1&gender=All&query=Bartender
Go
sense.us / jobs
hello, Michael profile comments trails help logout
Reported Occupations of U.S. Labor Force, 1850-2000 (source: http://ipums.org)
>> Bartender
all men women
% of Work Force
0.35%
0.3%
0.25%
0.2%
0.15%
0.1%
0.05%
Bartender
Bartender
1850 1870 1900 1920 1940 1960 1980 2000
Annotate
Stroke
Fill
Sans-Serif
14
B I
comments (1) New Comment | View All (139)
Prohibition... (1)
...and then women move behind the bar.
by Adam Kleinberg on Tue Aug 1, 2006 8:02 AM
reply
bookmarks (1) Add View | Save | Load | View All
visual communication lab privacy feedback help
See something interesting? Add a comment, it's easy!

sense.us


sense.us > jobs - Mozilla Firefox
File Edit View Go Bookmarks Tools Help
http://localhost/jobs#scale=1&gender=All&query=professor&comment=47
Go
sense.us / jobs
hello, Michael profile comments trails help logout
Reported Occupations of U.S. Labor Force, 1850-2000 (source: http://ipums.org)
>> professor
all men women
% of Work Force
0.8%
0.7%
0.6%
0.5%
0.4%
0.3%
0.2%
0.1%
Professor
Professor
Professor - Social Science
1850 1870 1900 1920 1940 1960 1980 2000
Annotate
Stroke
Fill
Sans-Serif
14
B I
comments (1) New Comment | View All (140)
Disappearing professors??? (1)
Note how a bunch of different kinds of professors "disappear" in the year 2000, whereas the "professor" category take off. Is this just a change in the way the census aggregates data?
by Julia Hernandez on Fri Jul 21, 2006 11:07 AM
reply
bookmarks (2) Add View | Save | Load | View All
visual communication lab privacy feedback help
See something interesting? Add a comment, it's easy!

SENSE.US DESIGN PROCESS

SENSE.US DESIGN PROCESS

Collect data

SENSE.US DESIGN PROCESS

Collect data

Collect more data

SENSE.US DESIGN PROCESS

Collect data

Collect more data

Learn data is piecemeal & incompatible

SENSE.US DESIGN PROCESS

Collect data

Collect more data

Learn data is piecemeal & incompatible

Get frustrated!

SENSE.US DESIGN PROCESS

Collect data

Collect more data

Learn data is piecemeal & incompatible

Get frustrated!

Seek out subject matter experts

SENSE.US DESIGN PROCESS

Collect data

Collect more data

Learn data is piecemeal & incompatible

Get frustrated!

Seek out subject matter experts

Find better data source

SENSE.US DESIGN PROCESS (2)

SENSE.US DESIGN PROCESS (2)

Collect data

SENSE.US DESIGN PROCESS (2)

Collect data

Find data is too big for spreadsheets

SENSE.US DESIGN PROCESS (2)

Collect data

Find data is too big for spreadsheets

Peek at data through text editors

SENSE.US DESIGN PROCESS (2)

Collect data

Find data is too big for spreadsheets

Peek at data through text editors

Write code to parse data

SENSE.US DESIGN PROCESS (2)

Collect data

Find data is too big for spreadsheets

Peek at data through text editors

Write code to parse data

Run code; find errors; fix code; repeat

SENSE.US DESIGN PROCESS (2)

Collect data

Find data is too big for spreadsheets

Peek at data through text editors

Write code to parse data

Run code; find errors; fix code; repeat

Get frustrated!

SENSE.US DESIGN PROCESS (3)

SENSE.US DESIGN PROCESS (3)

Load data into relational database

SENSE.US DESIGN PROCESS (3)

Load data into relational database

Discover errors on import

SENSE.US DESIGN PROCESS (3)

Load data into relational database

Discover errors on import

More frustration!

SENSE.US DESIGN PROCESS (3)

Load data into relational database

Discover errors on import

More frustration!

Fix import script; load into database

SENSE.US DESIGN PROCESS (3)

Load data into relational database

Discover errors on import

More frustration!

Fix import script; load into database

Write queries against data; learn a little

SENSE.US DESIGN PROCESS (3)

Load data into relational database

Discover errors on import

More frustration!

Fix import script; load into database

Write queries against data; learn a little

Point Tableau at database to explore

SENSE.US DESIGN PROCESS (4)

SENSE.US DESIGN PROCESS (4)

Explore and profile data

SENSE.US DESIGN PROCESS (4)

Explore and profile data; *FUN!!*

SENSE.US DESIGN PROCESS (4)

Explore and profile data; *FUN!!*

Prototype visualizations (various tools)

SENSE.US DESIGN PROCESS (4)

Explore and profile data; *FUN!!*

Prototype visualizations (various tools)

Get crits & feedback from team

SENSE.US DESIGN PROCESS (4)

Explore and profile data; *FUN!!*

Prototype visualizations (various tools)

Get crits & feedback from team

Write custom visualization code

SENSE.US DESIGN PROCESS (4)

Explore and profile data; *FUN!!*

Prototype visualizations (various tools)

Get crits & feedback from team

Write custom visualization code

More team critique; user test

SENSE.US DESIGN PROCESS (4)

Explore and profile data; *FUN!!*

Prototype visualizations (various tools)

Get crits & feedback from team

Write custom visualization code

More team critique; user test

Deploy!

SENSE.US DESIGN PROCESS (4)

Explore and profile data; *FUN!!*

Prototype visualizations (various tools)

Get crits & feedback from team

Write custom visualization code

More team critique; user test

Deploy!

DATA

DATA

…is a stakeholder.

DATA

…is a stakeholder.

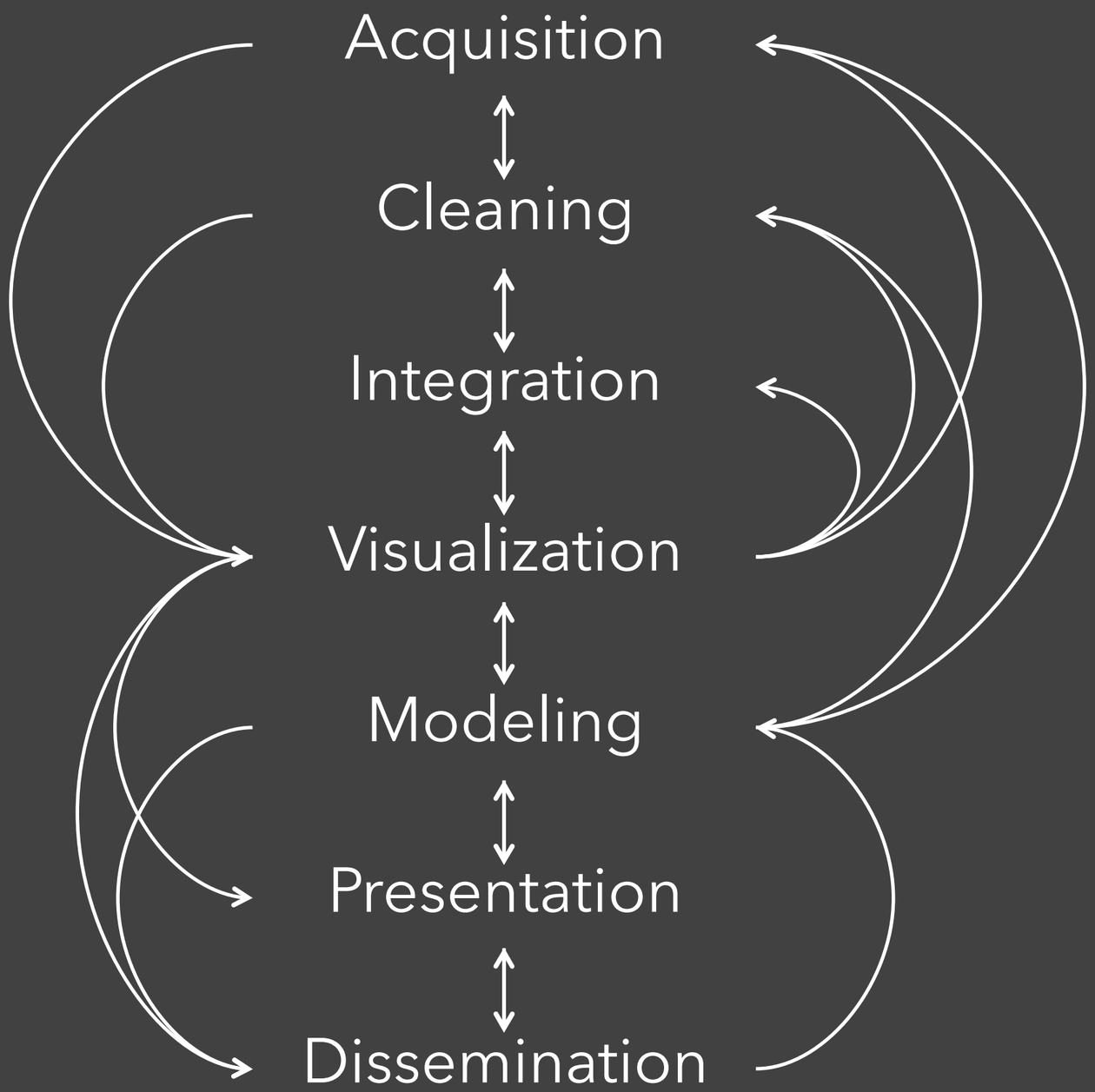
…comes in all shapes & sizes.

DATA

…is a stakeholder.

…comes in all shapes & sizes.

How to design tools for data unseen?


Acquisition
Cleaning
Integration
Visualization
Modeling
Presentation
Dissemination


TRIFACTA

Trifacta v1:
Enable Interactive Data Transformation

Trifacta v2:
Accelerate Data Discovery & Profiling

Trifacta v1:

Enable Interactive Data Transformation

Trifacta v2:

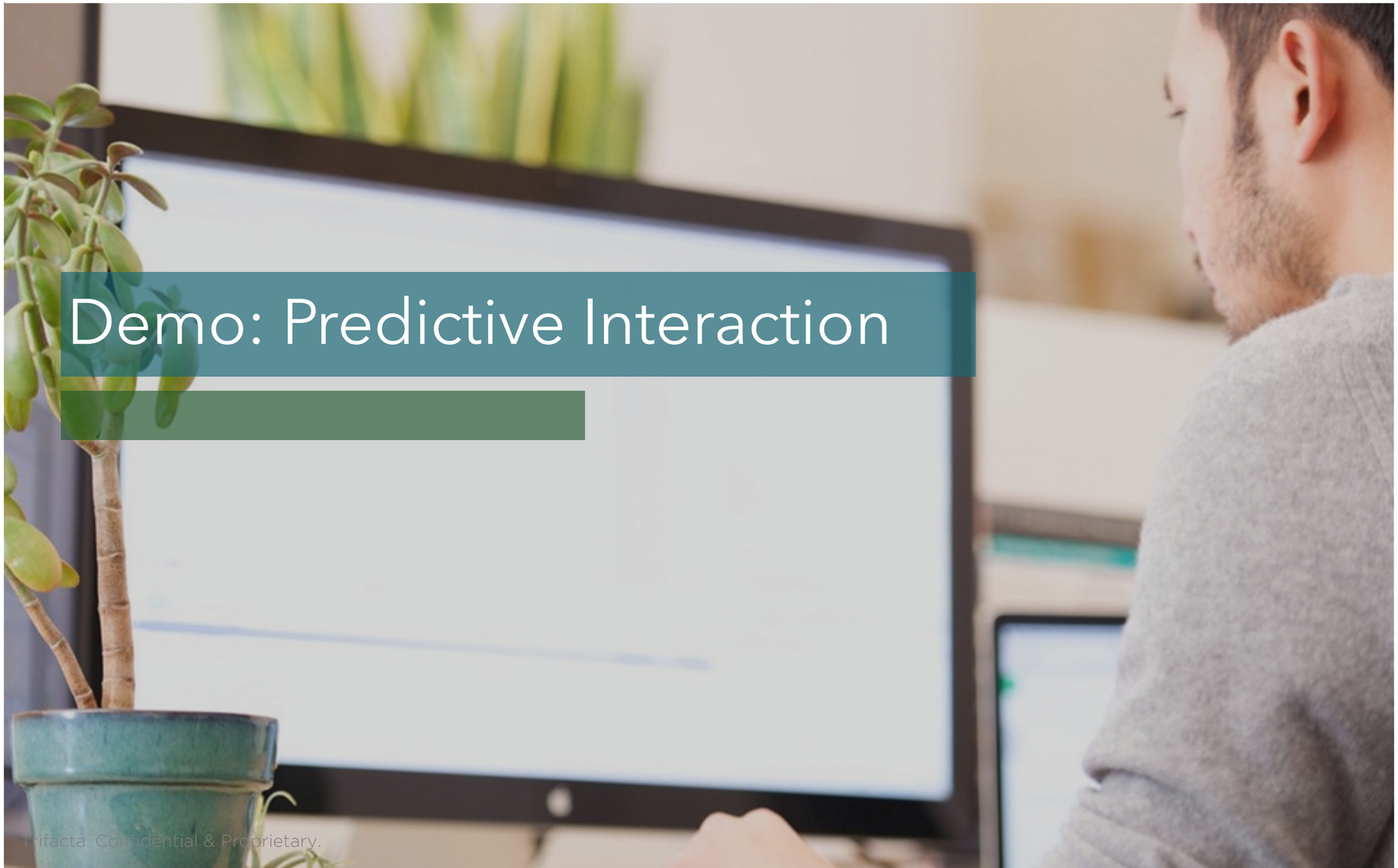
Accelerate Data Discovery & Profiling

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 interview study [IEEE VAST '12]

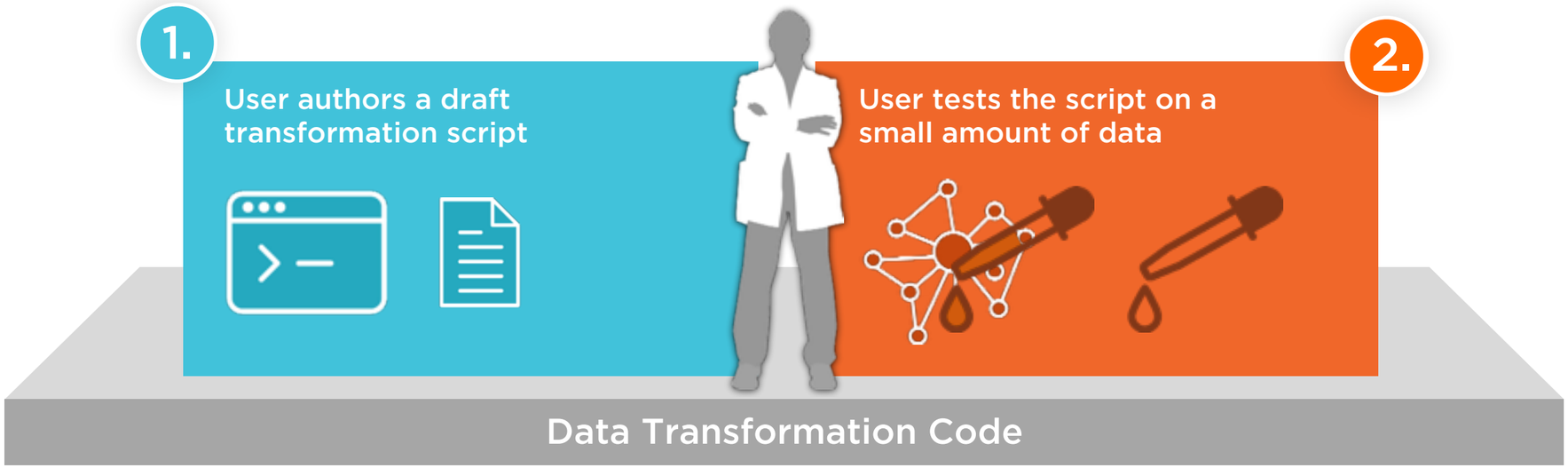
```
<MSIDN/IMSI/IMEI>  DATETTIME/DURATION/DISCONNECT REASON  MSWICENT:BASCENTCONT:BASTRASTA CALL_TYPE|CORRES_TYPE/CORRESP_IDN|CORRES2_TYPE/CORRESP2_ISDN
<600218408/208107376653317/35297700840926>  2013-12-19T8:49:19/52/11 MSC001:BSC001:BTS005     MOC1|SRV/600000231|/
<603159798/208106585213952/35643301870776>  2013-12-30T1:55:42/327/11      MSC001:BSC001:BTS011     MOC|ORG1/607971779|/
<603159798/208106585213952/35643301870776>  2013-12-18T23:29:32/284/11     MSC001:BSC002:BTS026     MOC|SFR/622153844|/
<603159798/208106585213952/35643301870776>  2013-12-20T19:52:25/326/11     MSC001:BSC002:BTS027     MOC|BOY/667808714|/
<603159798/208106585213952/35643301870776>  2013-12-30T10:02:48/281/11     MSC001:BSC002:BTS045     MOC|SFR/635393301|/
<603159798/208106585213952/35643301870776>  2013-12-24T7:50:56/302/11      MSC001:BSC002:BTS025     MOC|BOY/600316492|/
<603159798/208106585213952/35643301870776>  2013-12-18T5:57:43/328/11      MSC001:BSC001:BTS021     MOC|SFR/635580047|/
<603159798/208106585213952/35643301870776>  2013-12-28T22:42:20/328/11     MSC001:BSC002:BTS028     MOC|SFR/634262217|/
<603159798/208106585213952/35643301870776>  2013-12-27T0:41:24/274/11      MSC001:BSC001:BTS015     MOC|SFR/626481470|/
<603159798/208106585213952/35643301870776>  2013-12-18T5:33:57/290/11      MSC001:BSC001:BTS019     MOC|BOY/600383755|/
<603159798/208106585213952/35643301870776>  2013-12-24T13:40:03/307/11     MSC001:BSC001:BTS018     MOC|SFR/626633400|/
<603159798/208106585213952/35643301870776>  2013-12-21T23:36:20/303/11     MSC001:BSC001:BTS011     MOC|SFR/624221970|/
<603159798/208106585213952/35643301870776>  2013-12-30T5:50:25/292/11      MSC001:BSC002:BTS044     MOC|ORG1/640759322|/
<603159798/208106585213952/35643301870776>  2013-12-30T21:08:10/315/11     MSC001:BSC001:BTS004     MOC|SFR/625725535|/
<603159798/208106585213952/35643301870776>  2013-12-26T8:23:04/313/11      MSC001:BSC001:BTS002     MOC|ORG1/638260666|/
<603159798/208106585213952/35643301870776>  2013-12-25T12:40:24/304/11     MSC001:BSC001:BTS009     MOC|ORG1/608009941|/
<603159798/208106585213952/35643301870776>  2013-12-19T20:29:40/284/11     MSC001:BSC001:BTS024     MOC|ORG1/638180946|/
<603159798/208106585213952/35643301870776>  2013-12-18T11:08:13/324/11     MSC001:BSC001:BTS015     MOC|SFR/635024026|/
<603159798/208106585213952/35643301870776>  2013-12-24T19:59:25/285/11     MSC001:BSC002:BTS043     MOC|ORG1/640731421|/
<603159798/208106585213952/35643301870776>  2013-12-24T4:17:45/288/11      MSC001:BSC002:BTS039     MOC|SFR/634636870|/
<603159798/208106585213952/35643301870776>  2013-12-21T0:12:02/298/11      MSC001:BSC002:BTS044     MOC|BOY/658091748|/
<603159798/208106585213952/35643301870776>  2013-12-25T13:17:44/320/11     MSC001:BSC001:BTS015     MOC|BOY/600325625|/
<603159798/208106585213952/35643301870776>  2013-12-30T10:57:59/308/11     MSC001:BSC002:BTS044     MOC|BOY/669466150|/
<603159798/208106585213952/35643301870776>  2013-12-23T8:51:55/300/11      MSC001:BSC001:BTS014     MOC|BOY/600383355|/
<603159798/208106585213952/35643301870776>  2013-12-27T23:10:03/299/11     MSC001:BSC001:BTS009     MOC|SFR/646445986|/
<603159798/208106585213952/35643301870776>  2013-12-29T3:29:33/326/19      MSC001:BSC001:BTS001     MOC|SFR/634612331|/
<603159798/208106585213952/35643301870776>  2013-12-24T17:12:23/279/11     MSC001:BSC002:BTS044     MOC|ORG1/638801691|/
<603159798/208106585213952/35643301870776>  2013-12-30T2:08:09/302/11      MSC001:BSC002:BTS033     MOC|SFR/613232266|/
<603159798/208106585213952/35643301870776>  2013-12-31T7:06:42/326/11      MSC001:BSC002:BTS025     MOC|SFR/646903678|/
<603159798/208106585213952/35643301870776>  2013-12-25T7:07:12/290/11      MSC001:BSC001:BTS022     MOC|SFR/646420207|/
<603159798/208106585213952/35643301870776>  2013-12-22T19:17:05/299/11     MSC001:BSC001:BTS007     MOC|BOY/600371891|/
<603159798/208106585213952/35643301870776>  2013-12-19T1:41:46/313/11      MSC001:BSC001:BTS016     MOC|SFR/610384355|/
<603159798/208106585213952/35643301870776>  2013-12-23T7:52:29/285/11      MSC001:BSC001:BTS005     MOC|ORG1/638311751|/
<603159798/208106585213952/35643301870776>  2013-12-25T14:41:32/245/11     MSC001:BSC001:BTS004     MTC|ORG1/640379176|/
<603159798/208106585213952/35643301870776>  2014-01-01T11:51:09/251/19     MSC001:BSC001:BTS021     MTC|ORG1/638848686|/
<603159798/208106585213952/35643301870776>  2013-12-19T8:40:20/247/11      MSC001:BSC001:BTS005     MTC|SFR/621082531|/
<603159798/208106585213952/35643301870776>  2013-12-20T2:22:47/261/11      MSC001:BSC002:BTS033     MTC|ORG1/638706146|/
<603159798/208106585213952/35643301870776>  2013-12-21T13:36:53/233/11     MSC001:BSC001:BTS018     MTC|ORG1/638816545|/
<603159798/208106585213952/35643301870776>  2013-12-20T11:28:38/263/11     MSC001:BSC001:BTS006     MTC|FIX/154487892|/
<603159798/208106585213952/35643301870776>  2013-12-27T8:10:59/263/11      MSC001:BSC002:BTS048     MTC|SFR/655385658|/
<603159798/208106585213952/35643301870776>  2013-12-26T10:52:09/232/11     MSC001:BSC001:BTS015     MTC|ORG1/640515237|/
<603159798/208106585213952/35643301870776>  2013-12-27T1:34:52/262/11      MSC001:BSC002:BTS040     MTC|FIX/387534732|/
<603159798/208106585213952/35643301870776>  2014-01-01T0:06:25/237/11      MSC001:BSC002:BTS036     MTC|FIX/134505487|/
<603159798/208106585213952/35643301870776>  2013-12-22T9:53:36/242/11      MSC001:BSC001:BTS016     MTC|FIX/139043700|/
<603159798/208106585213952/35643301870776>  2013-12-22T14:58:57/240/11     MSC001:BSC002:BTS037     MTC|ORG1/638204881|/
<603159798/208106585213952/35643301870776>  2013-12-19T13:45:16/237/11     MSC001:BSC001:BTS019     MTC|ORG1/640755803|/
```

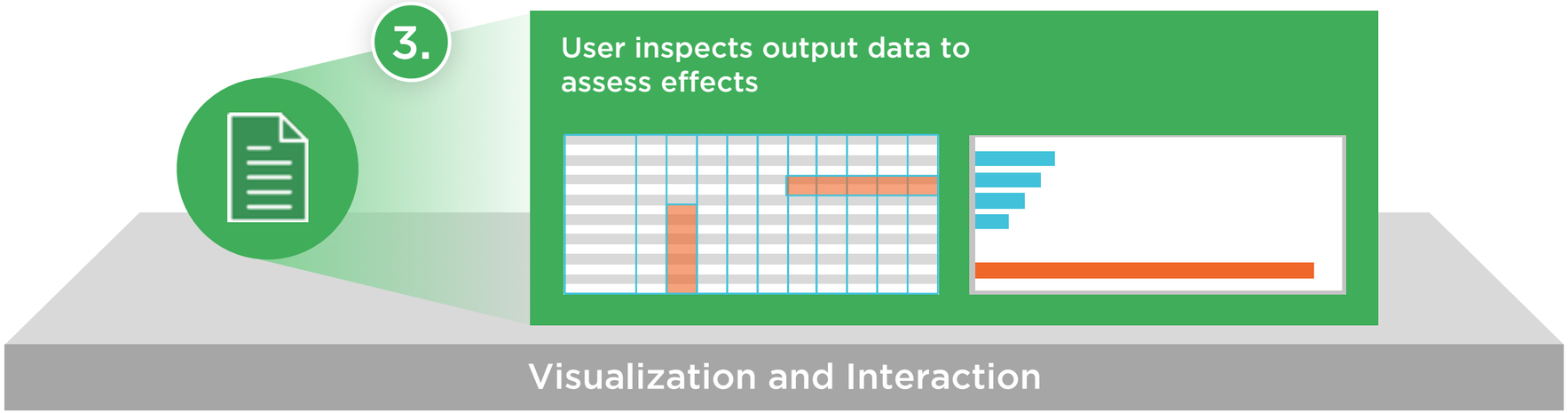
TRIFACTA



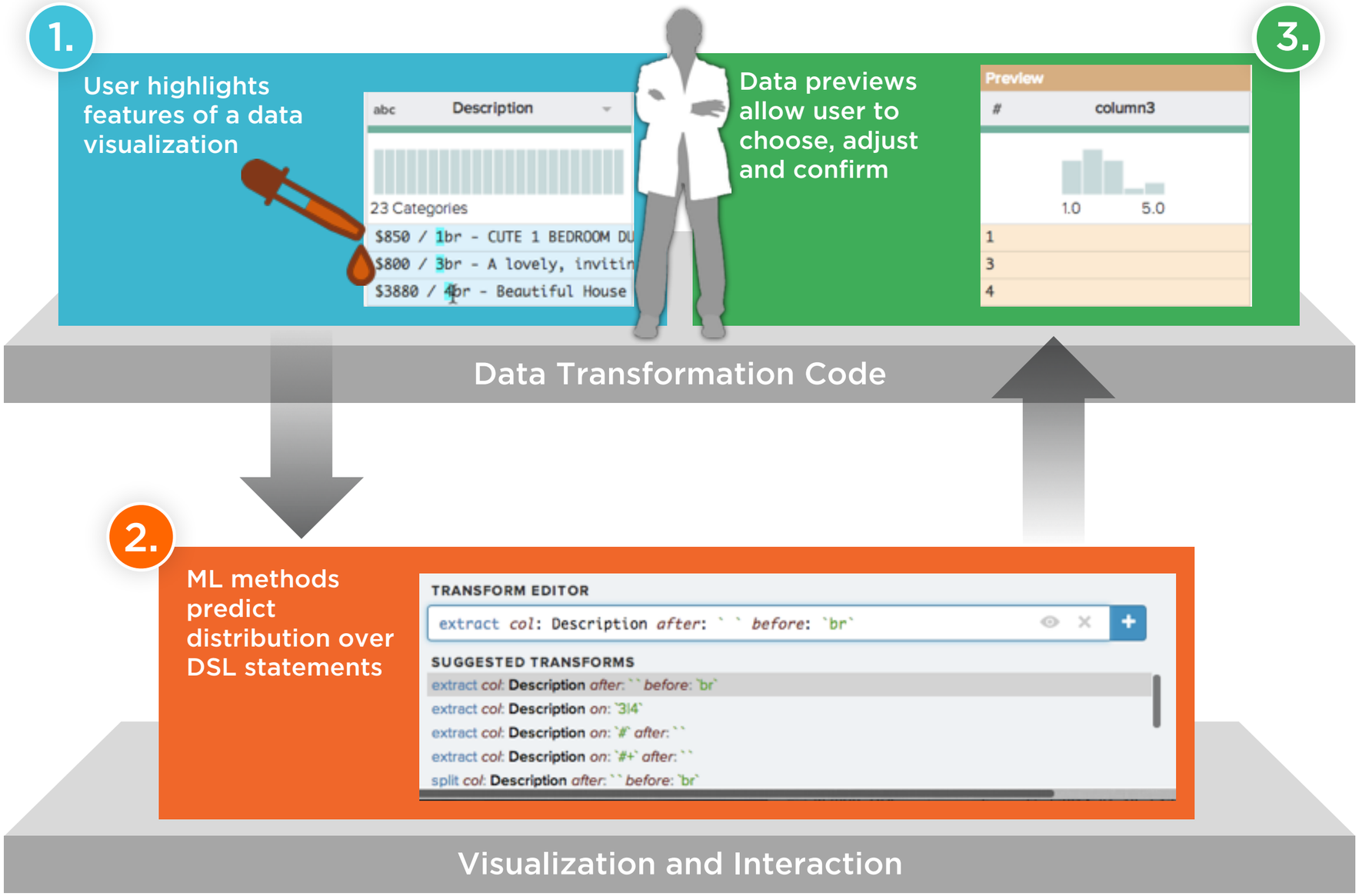
Demo: Predictive Interaction

Traditional Burden of Specification





Trifacta: Predictive Interaction™



DESIGN DESIDERATA

DESIGN DESIDERATA

Co-evolution of language & interface

DESIGN DESIDERATA

Co-evolution of language & interface

Embrace ambiguity

DESIGN DESIDERATA

Co-evolution of language & interface

Embrace ambiguity

Visualize transformation effects

DESIGN DESIDERATA

Co-evolution of language & interface

Embrace ambiguity

Visualize transformation effects

Interpretable transformations

Designing for Interpretability

SUGGESTED TRANSFORMS

extract *col:* **Screen_Detail** *on:* /(?<=adtam_source\=)[^\&]*(?=\&)/
extract *col:* **Screen_Detail** *on:* /(?<=\=)[^\&]*(?=\&)/ *limit:* 2
extract *col:* **Screen_Detail** *on:* /(?<=\=)[a-z]+/ *limit:* 2
extract *col:* **Screen_Detail** *on:* /[a-z]+/ *limit:* 4
countpattern *col:* **Screen_Detail** *on:* /[a-z]+/

Designing for Interpretability

SUGGESTED TRANSFORMS

extract *col:* **Screen_Detail** *on:* /(?<=adtam_source\=)[^\&]*(?=\&)/
extract *col:* **Screen_Detail** *on:* /(?<=\=)[^\&]*(?=\&)/ *limit:* 2
extract *col:* **Screen_Detail** *on:* /(?<=\=)[a-z]+/ *limit:* 2
extract *col:* **Screen_Detail** *on:* /[a-z]+/ *limit:* 4
countpattern *col:* **Screen_Detail** *on:* /[a-z]+/

SUGGESTED TRANSFORMS

extract *col:* **Screen_Detail** *after:* `adtam_source=` *before:* `&`
extract *col:* **Screen_Detail** *limit:* 2 *after:* `=` *before:* `&`
extract *col:* **Screen_Detail** *on:* `{lower}+` *limit:* 2
extract *col:* **Screen_Detail** *on:* `{lower}+` *limit:* 4
countpattern *col:* **Screen_Detail** *on:* `{lower}+`

DESIGN DESIDERATA

Co-evolution of language & interface

Embrace ambiguity

Visualize transformation effects

Interpretable transformations

…and lots of testing (data + users)!

Trifacta v1:

Enable Interactive Data Transformation

Trifacta v2:

Accelerate Data Discovery & Profiling

Trifacta v1:
Enable Interactive Data Transformation

Trifacta v2:
Accelerate Data Discovery & Profiling

What’s in your data?

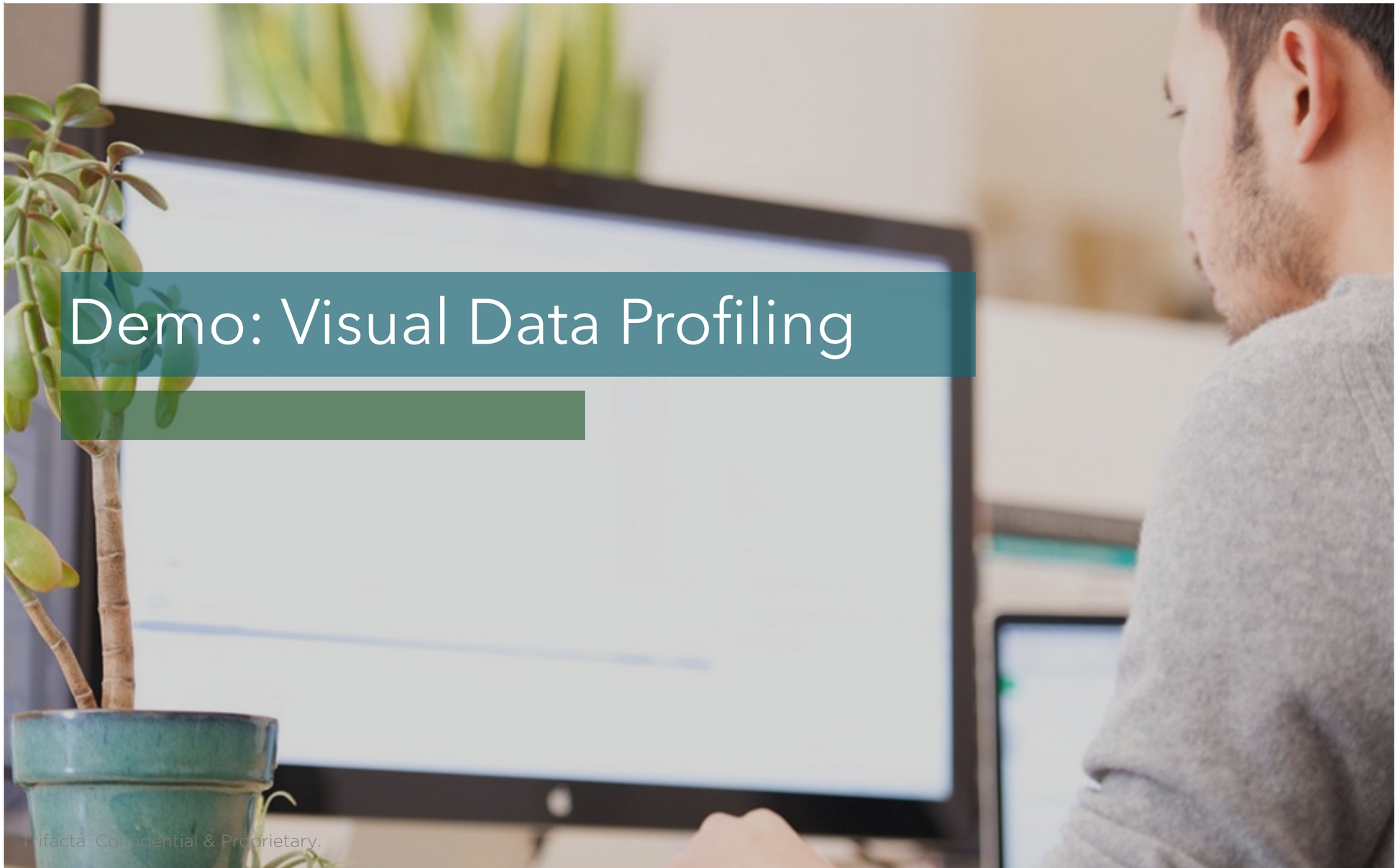

What’s in your data?

The first sign that a visualization is good is that it shows you a problem in your data.

Every successful visualization that I've been involved with has had this stage where you realize, "Oh my God, this data is not what I thought it would be!" So already, you've discovered something.

Martin Wattenberg [ACM Queue '09]

TRIFACTA

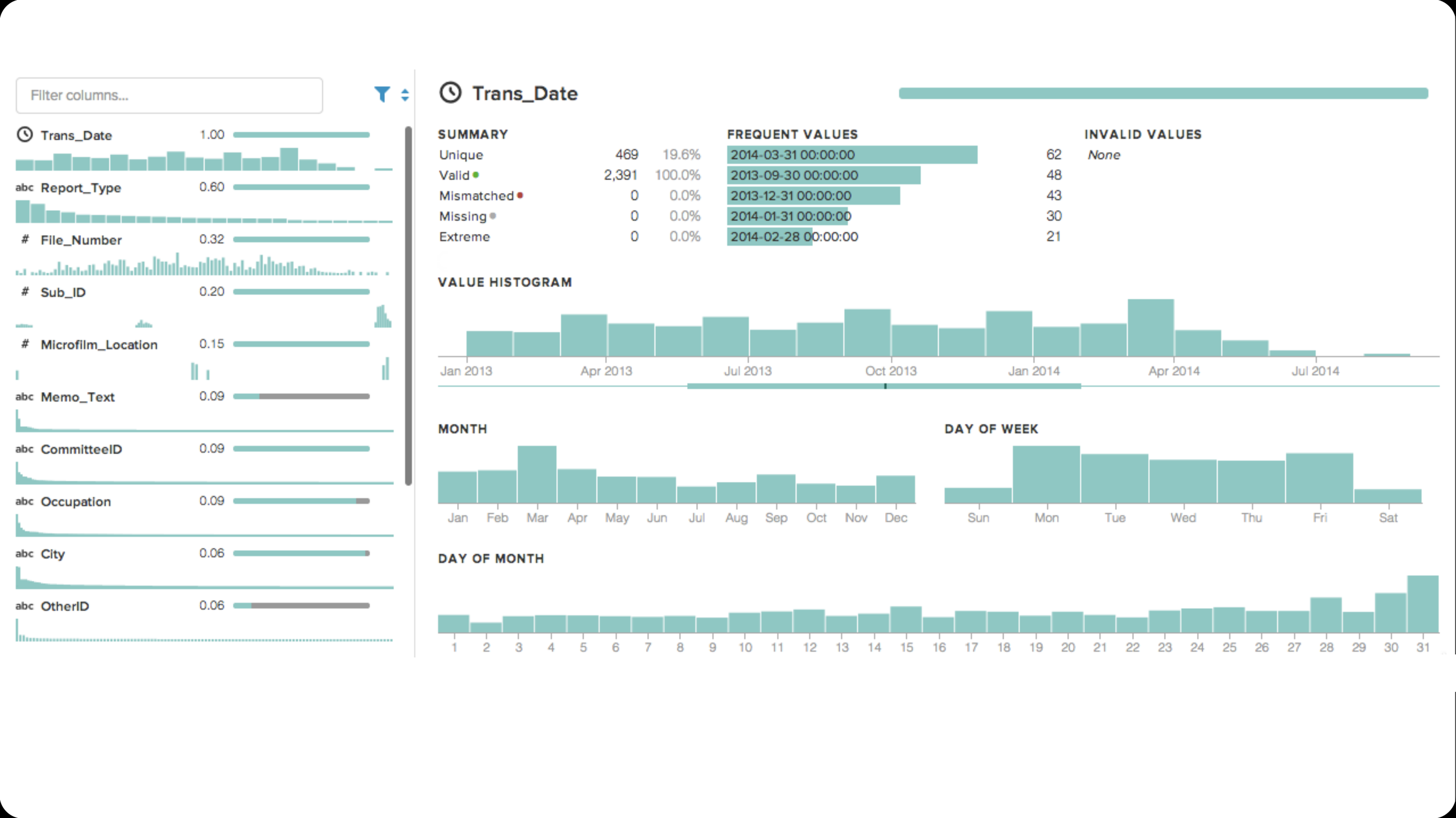


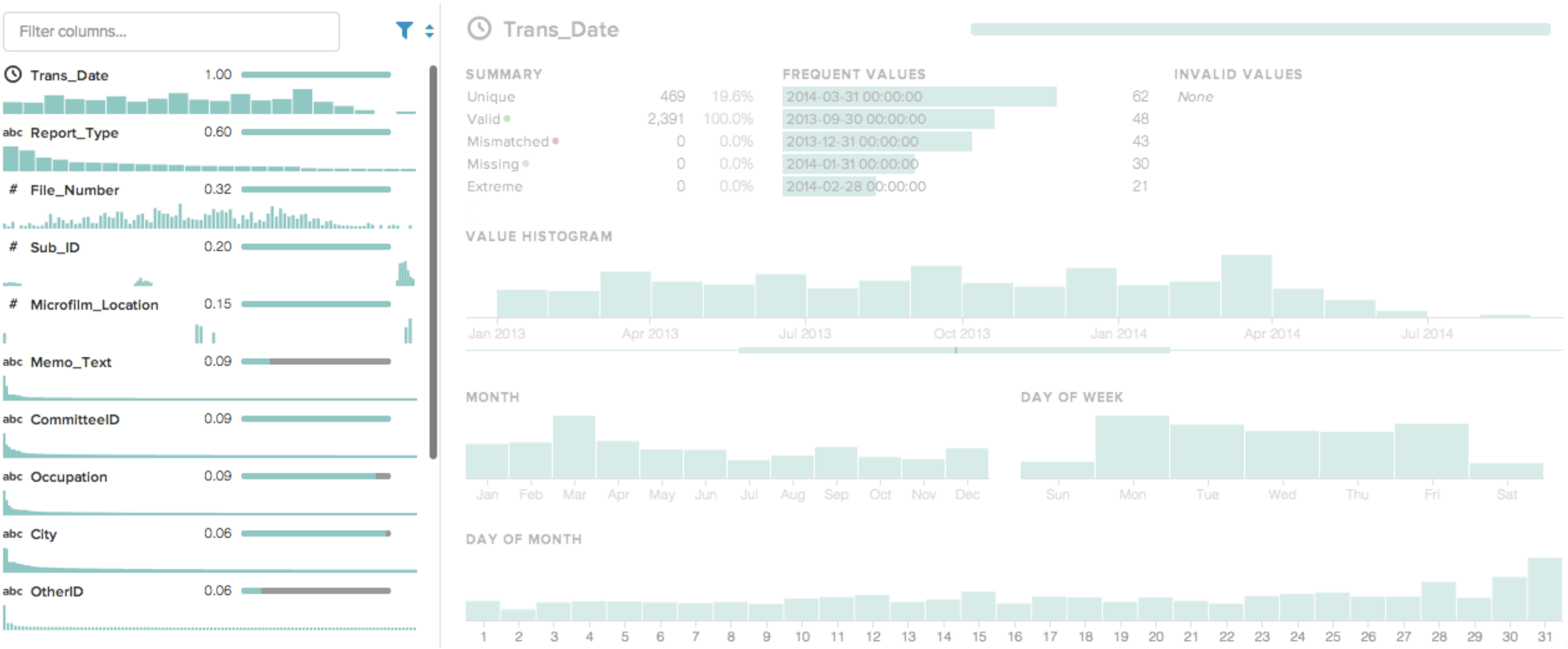
Demo: Visual Data Profiling

DESIGN DESIDERATA

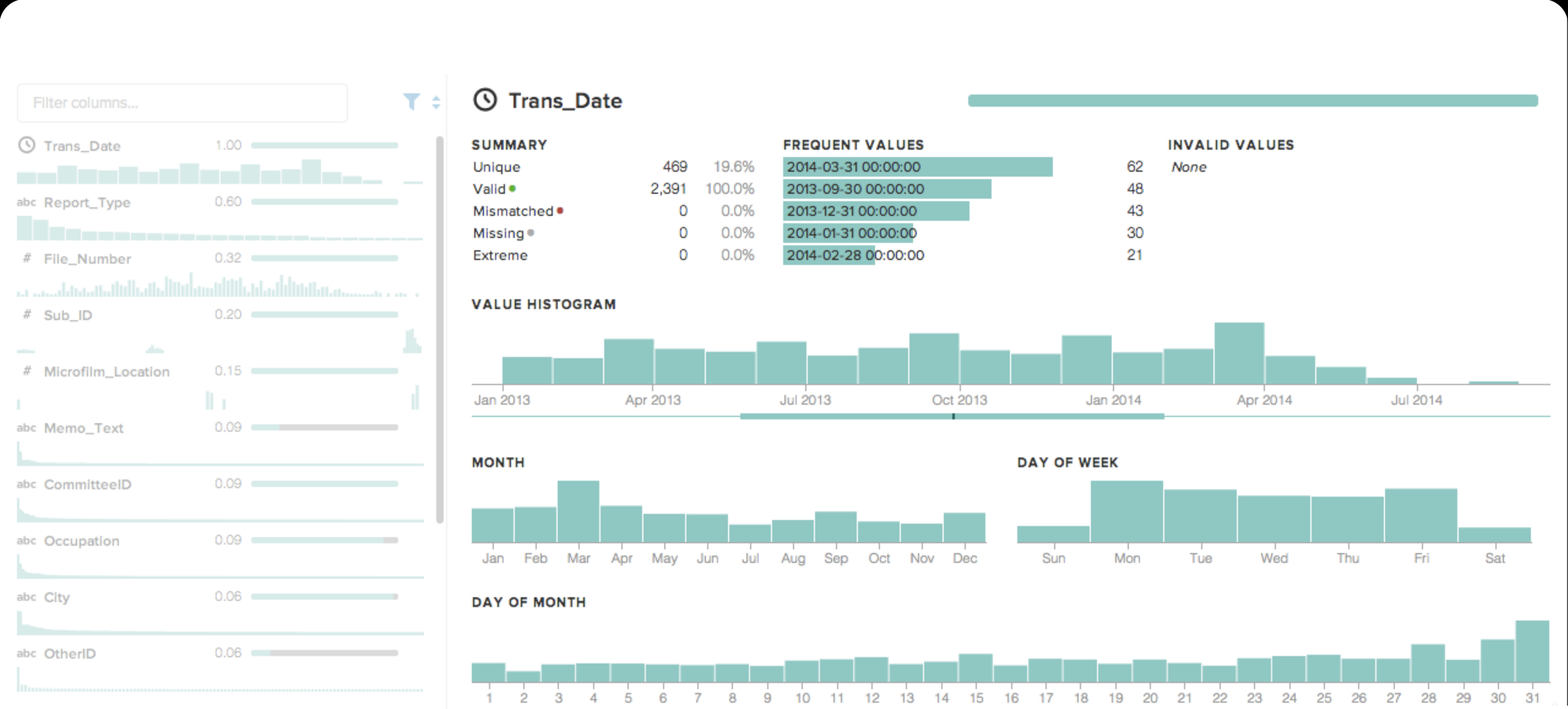
DESIGN DESIDERATA

Automatic visualization presentation


Filter columns...
Trans_Date 1.00
abc Report_Type 0.60
File_Number 0.32
Sub_ID 0.20
Microfilm_Location 0.15
abc Memo_Text 0.09
abc CommitteeID 0.09
abc Occupation 0.09
abc City 0.06
abc OtherID 0.06
Trans_Date
SUMMARY
Unique 469 19.6%
Valid 2,391 100.0%
Mismatched 0 0.0%
Missing 0 0.0%
Extreme 0 0.0%
FREQUENT VALUES
2014-03-31 00:00:00 62
2013-09-30 00:00:00 48
2013-12-31 00:00:00 43
2014-01-31 00:00:00 30
2014-02-28 00:00:00 21
INVALID VALUES
None
VALUE HISTOGRAM
Jan 2013
Apr 2013
Jul 2013
Oct 2013
Jan 2014
Apr 2014
Jul 2014
MONTH
Jan Feb Mar Apr May Jun Jul Aug Sep Oct Nov Dec
DAY OF WEEK
Sun Mon Tue Wed Thu Fri Sat
DAY OF MONTH
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



Exposure of all variables



Show relevant perspectives

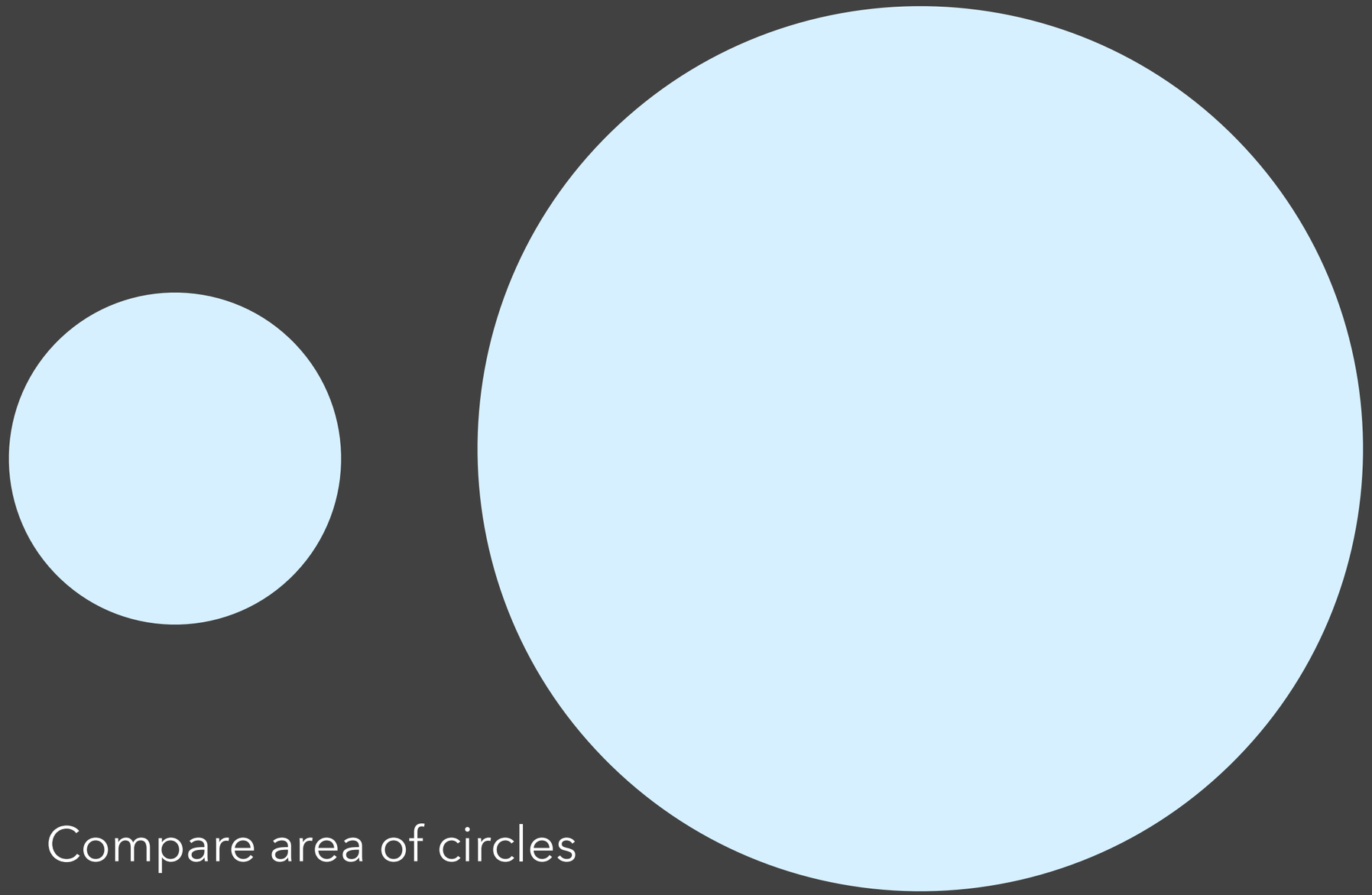
DESIGN DESIDERATA

Automatic visualization presentation

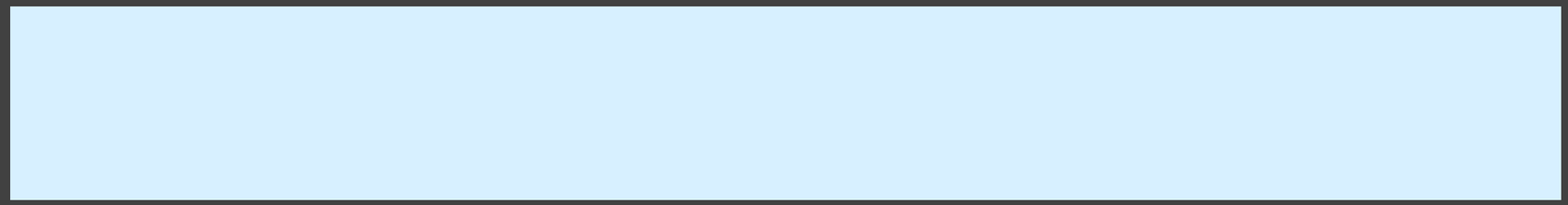
DESIGN DESIDERATA

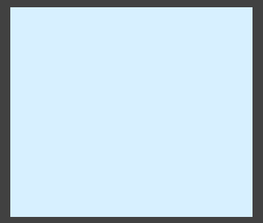
Automatic visualization presentation

Perceptually effective encodings



Compare area of circles





Compare length of bars

Perception of Quantitative Values

Most accurate

Least accurate

Position (common) scale

Position (non-aligned) scale

Length

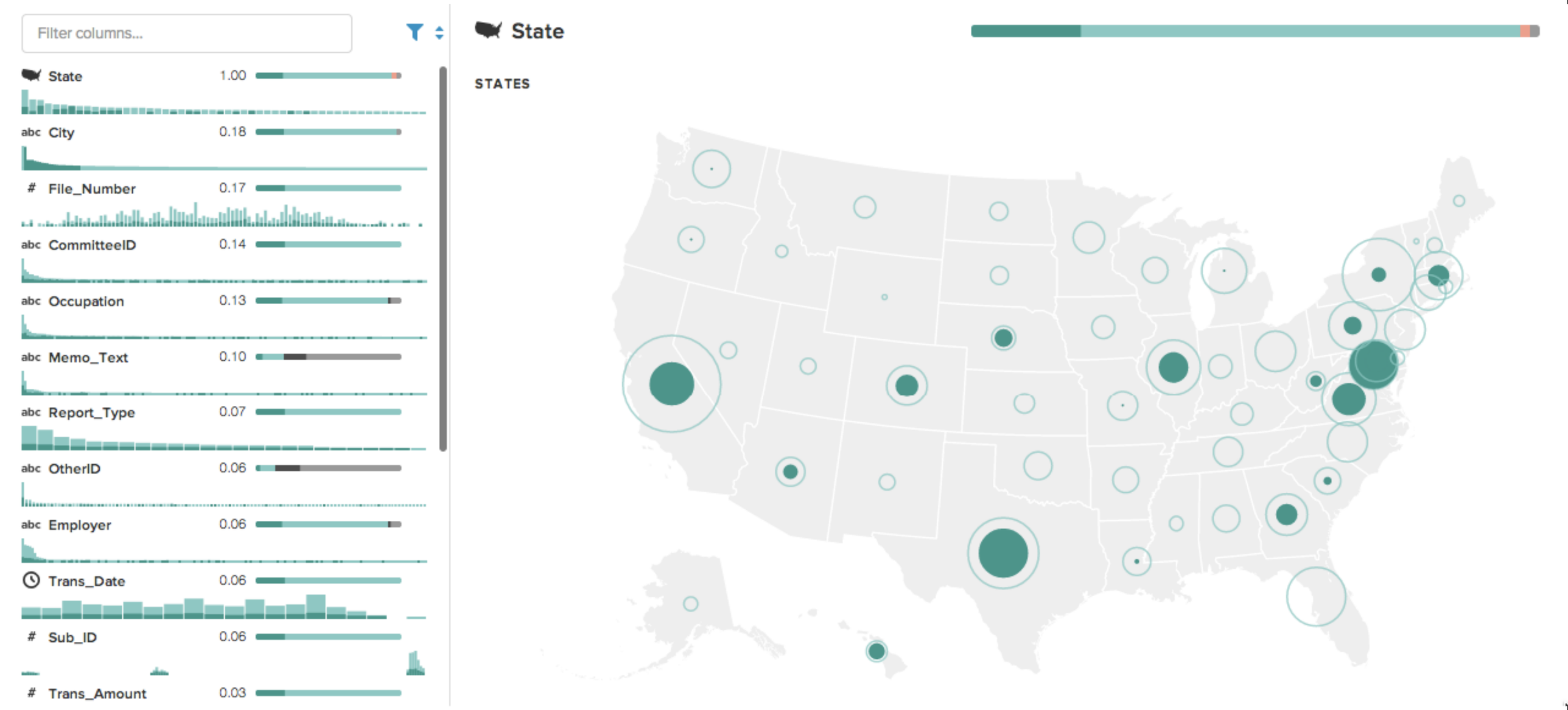
Slope

Angle

Area

Volume

Color hue-saturation-density



Area encoding >> Color encoding

DESIGN DESIDERATA

Automatic visualization presentation

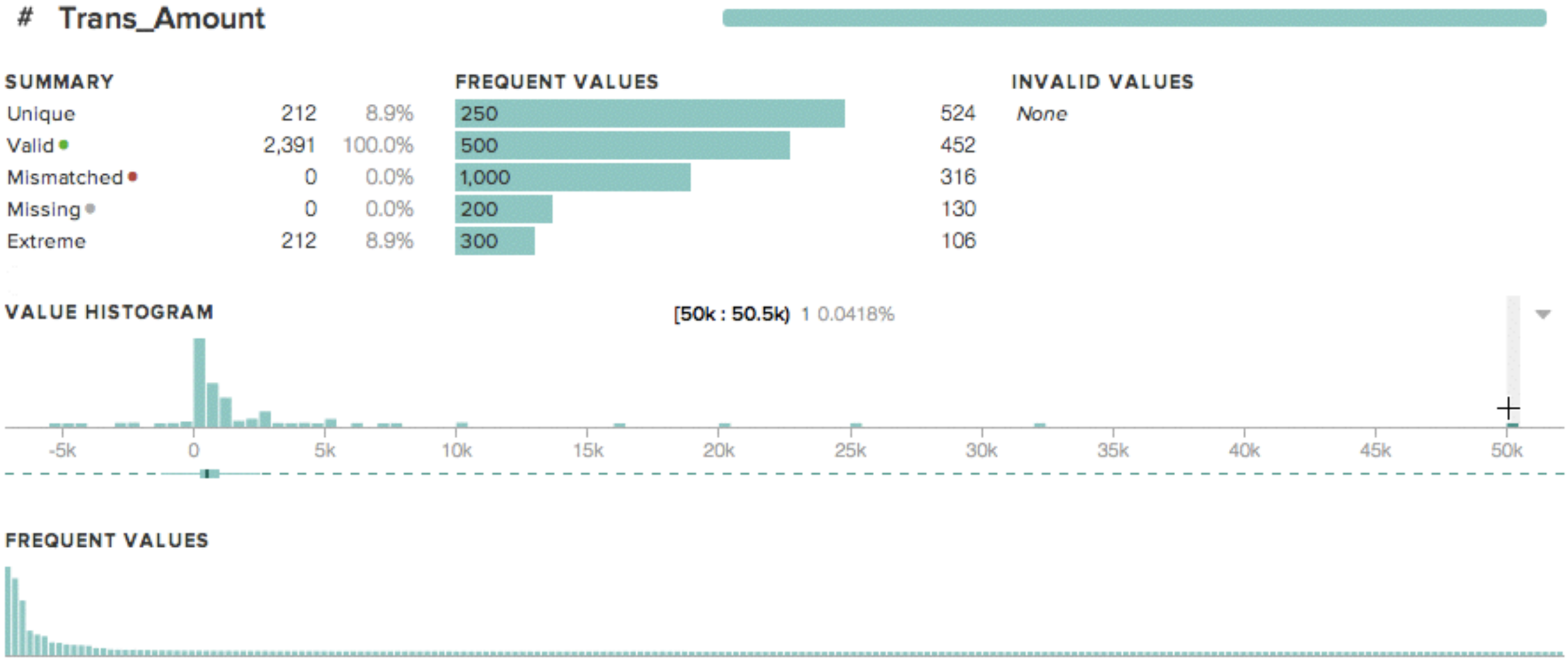
Perceptually effective encodings

DESIGN DESIDERATA

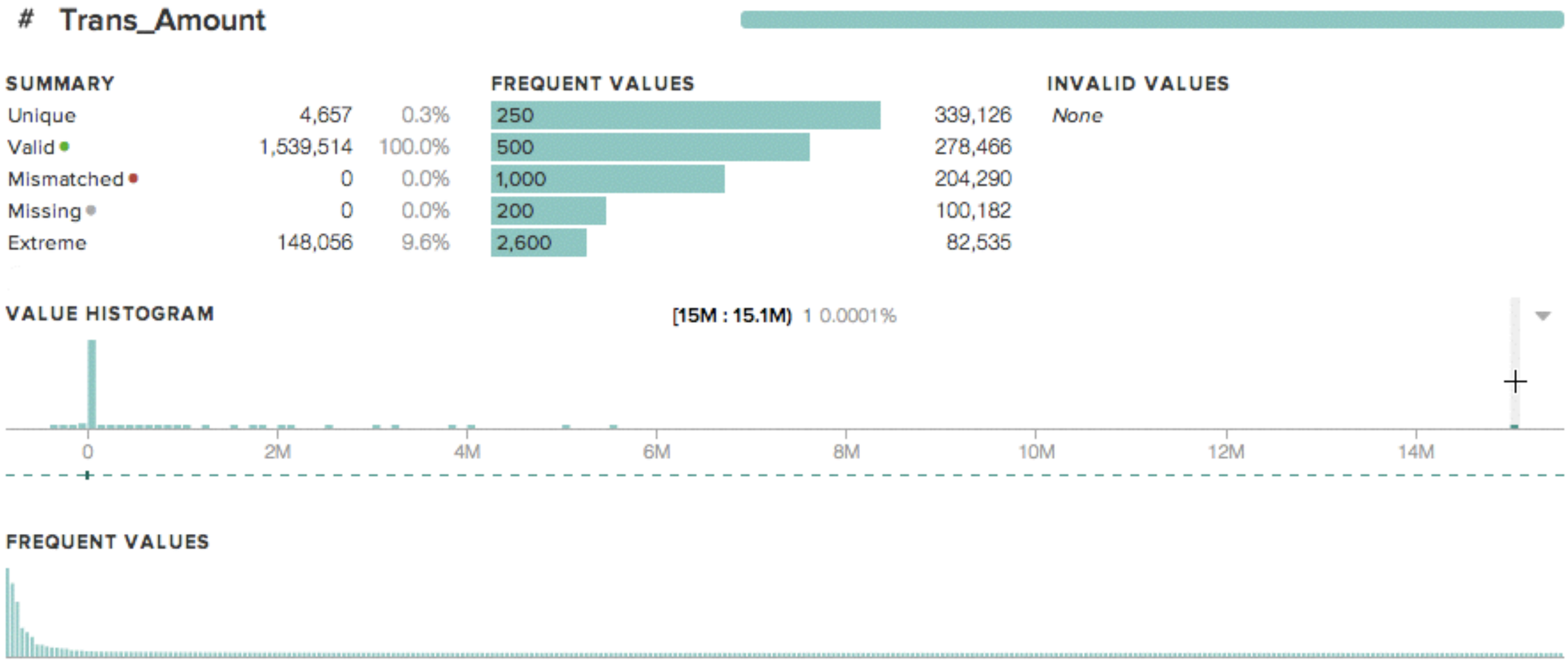
Automatic visualization presentation

Perceptually effective encodings

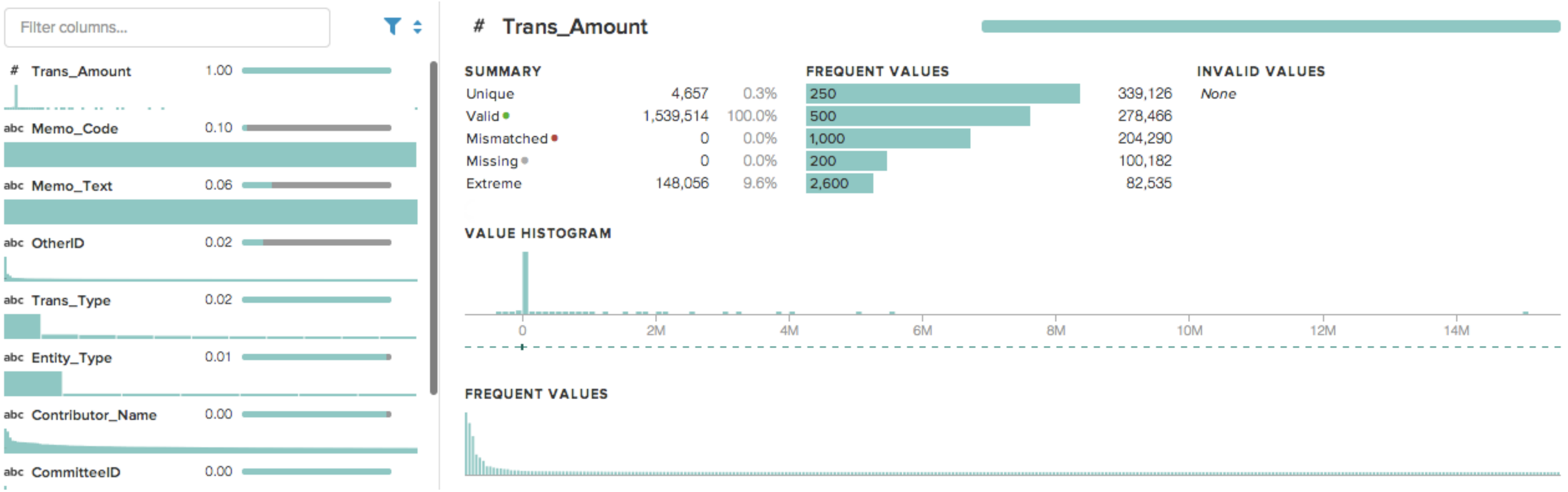
Visualize all data, big or small



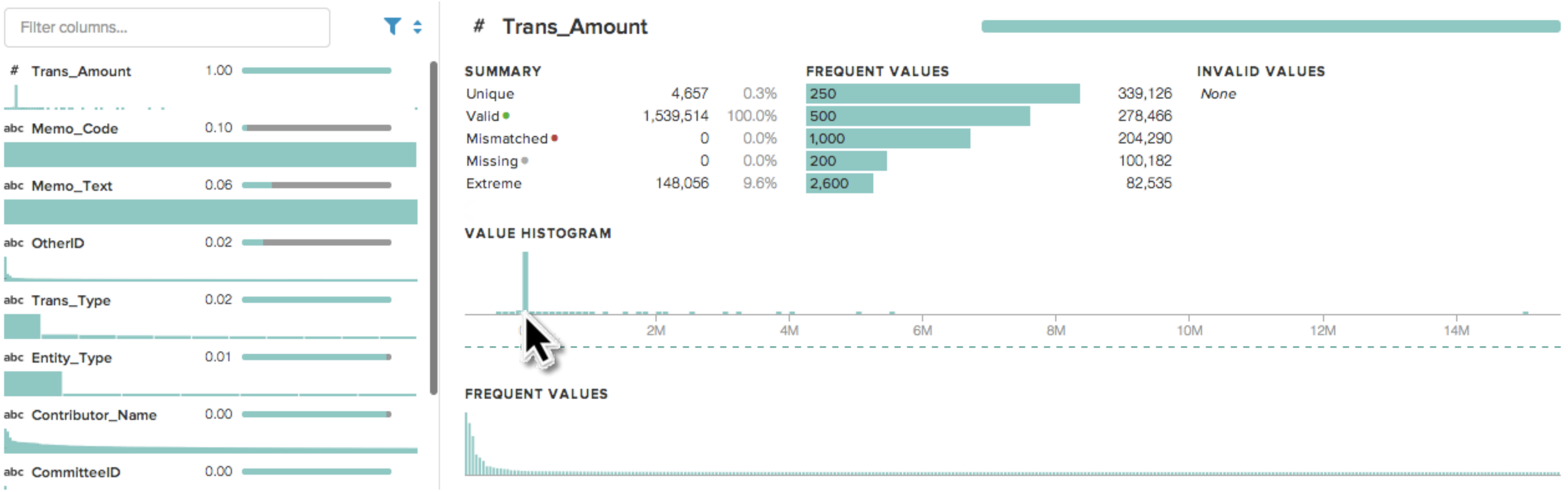
Summarize random sample



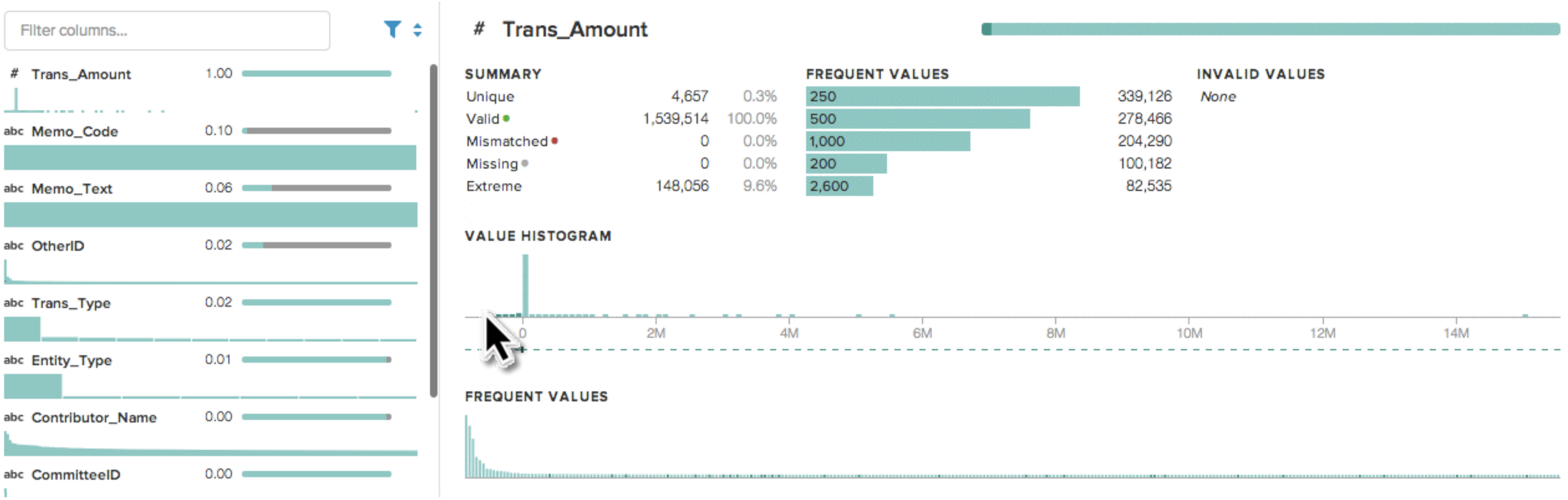
Summarize full data set



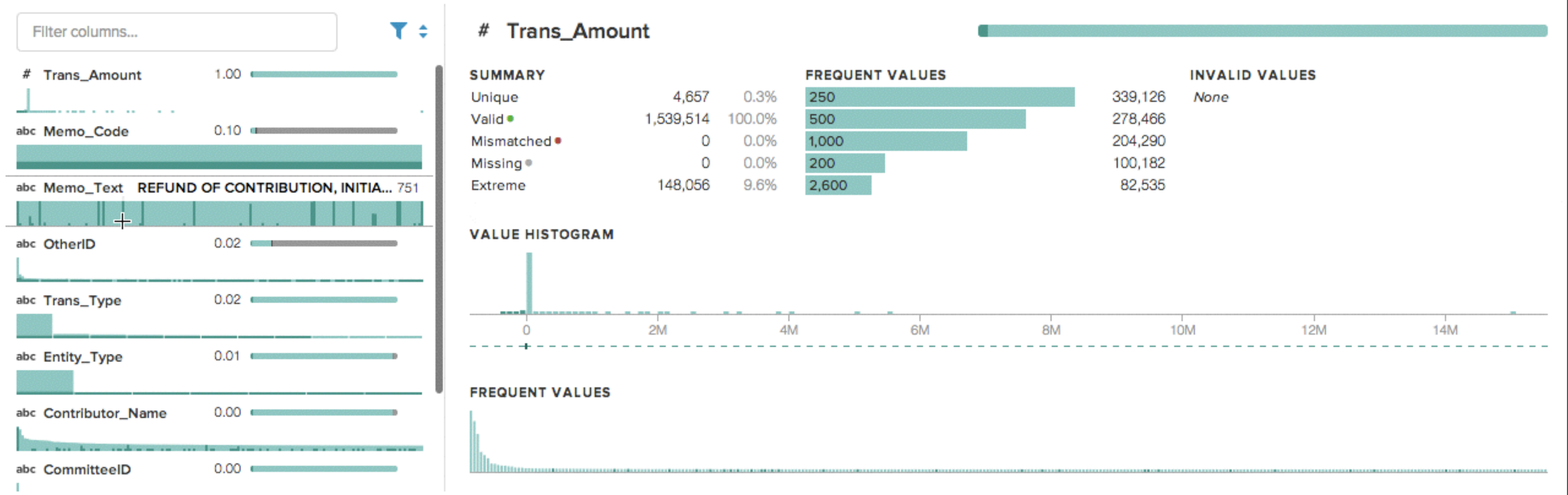
Immediate brushing & linking



Immediate brushing & linking



Immediate brushing & linking



Immediate brushing & linking

DESIGN DESIDERATA

Automatic visualization presentation

Perceptually effective encodings

Visualize all data, big or small

DESIGN DESIDERATA

Automatic visualization presentation

Perceptually effective encodings

Visualize all data, big or small

Springboard correction & refinement

TRANSFORM EDITOR

SCRIPT

```
splitrows col: column1 on: '\n'
split col: column1 on: '|' limit: 20
rename col: column2 to: 'CommitteeID'
rename col: column3 to: 'Amendment_Indicator'
```

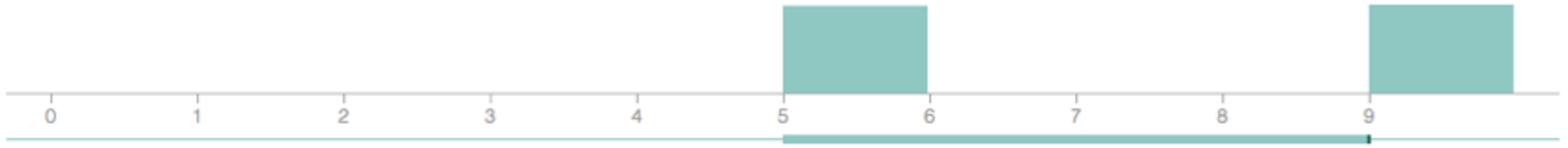
Grid | 1D Profile | 2D Profile | 21 Columns | 2,391 Rows | 4 Data Types | Filter data...

Filter columns...

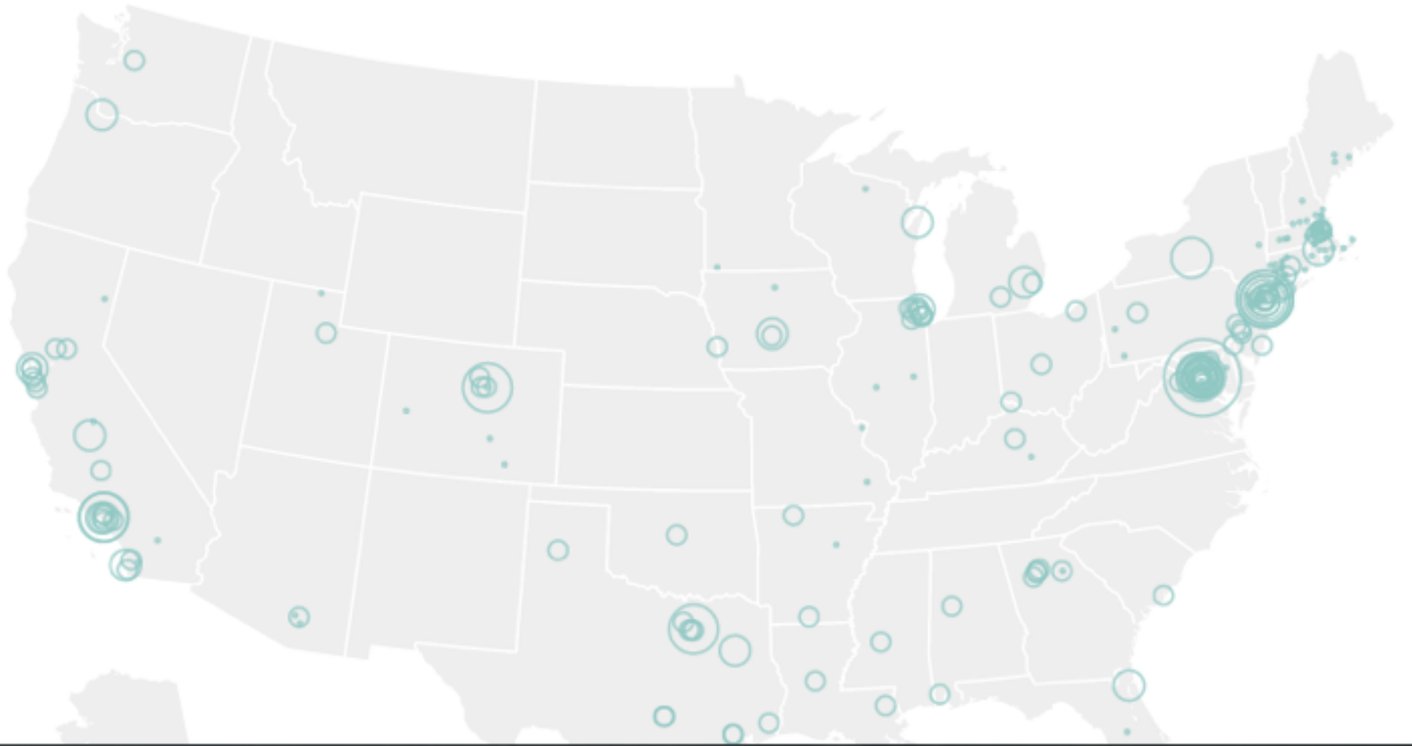
- abc CommitteeID
- abc Amendment_Indicator
- abc Report_Type
- abc Primary_General_Indicator
- # Microfilm_Location
- abc Trans_Type
- abc Entity_Type
- abc Contributor_Name
- abc City
- State
- abc Zip_Code
- abc Employer

abc Zip_Code

STRING LENGTHS



FREQUENT ZIP CODES



TRANSFORM EDITOR

SCRIPT

```
splitrows col: column1 on: '\n'
split col: column1 on: '|' limit: 20
rename col: column2 to: 'CommitteeID'
rename col: column3 to: 'Amendment_Indicator'
```

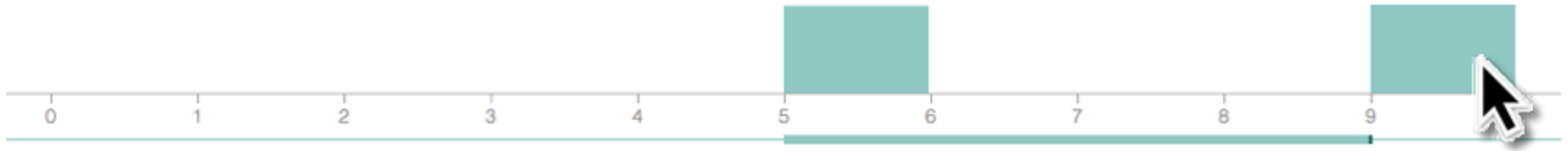
Grid | 1D Profile | 2D Profile | 21 Columns | 2,391 Rows | 4 Data Types | Filter data...

Filter columns...

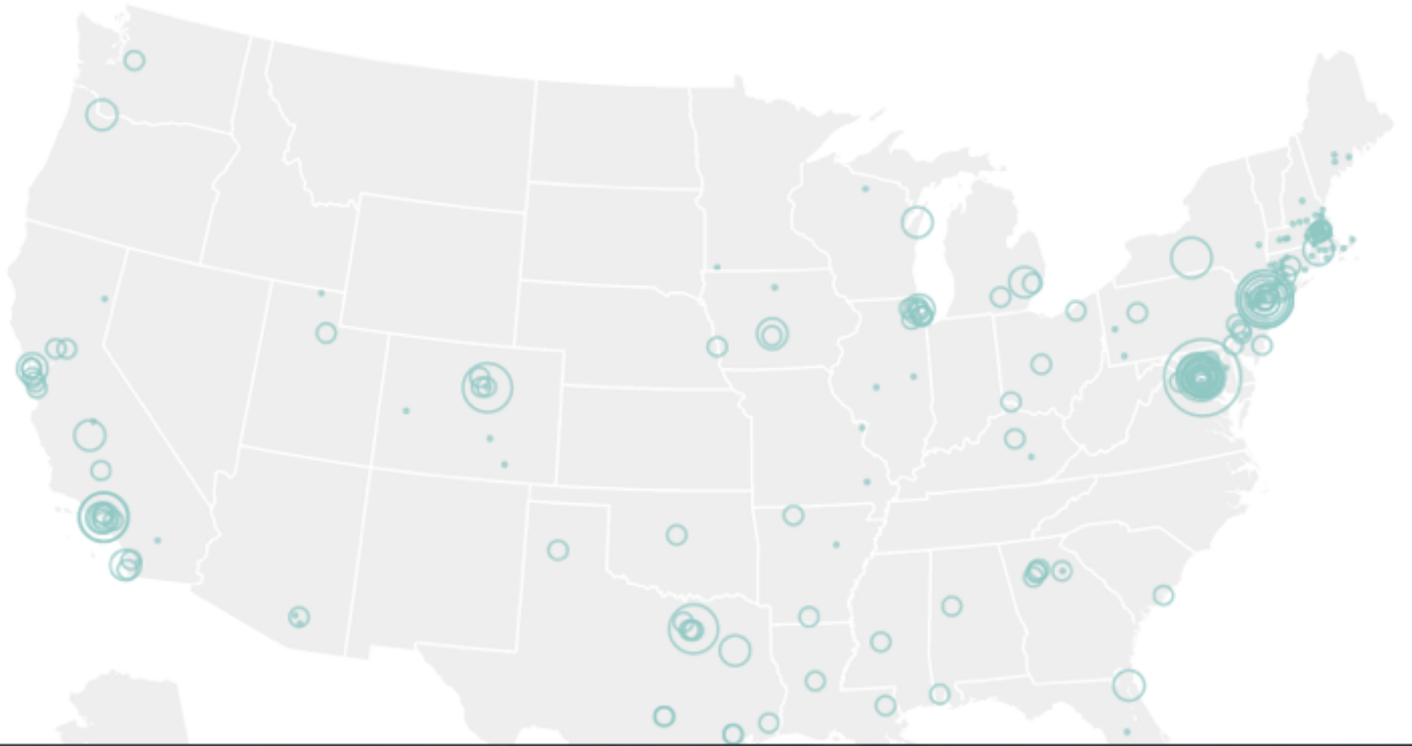
- abc CommitteeID
- abc Amendment_Indicator
- abc Report_Type
- abc Primary_General_Indicator
- # Microfilm_Location
- abc Trans_Type
- abc Entity_Type
- abc Contributor_Name
- abc City
- State
- abc Zip_Code
- abc Employer

abc Zip_Code

STRING LENGTHS



FREQUENT ZIP CODES



TRANSFORM EDITOR

```
set col: Zip_Code value: (length(Zip_Code) >= 9) ? left(Zip_Code,5) : Zip_Code
```

SCRIPT

```
splitrows col: column1 on: '\n'
split col: column1 on: '|' limit: 20
rename col: column2 to: 'CommitteeID'
rename col: column3 to: 'Amendment_Indicator'
```

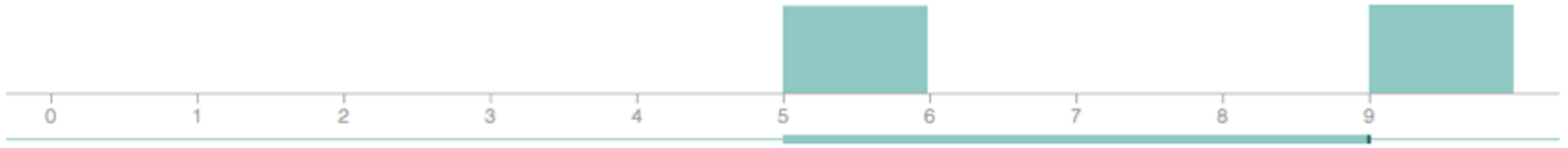
Grid | 1D Profile | 2D Profile | 21 Columns | 2,391 Rows | 4 Data Types | Filter data...

Filter columns...

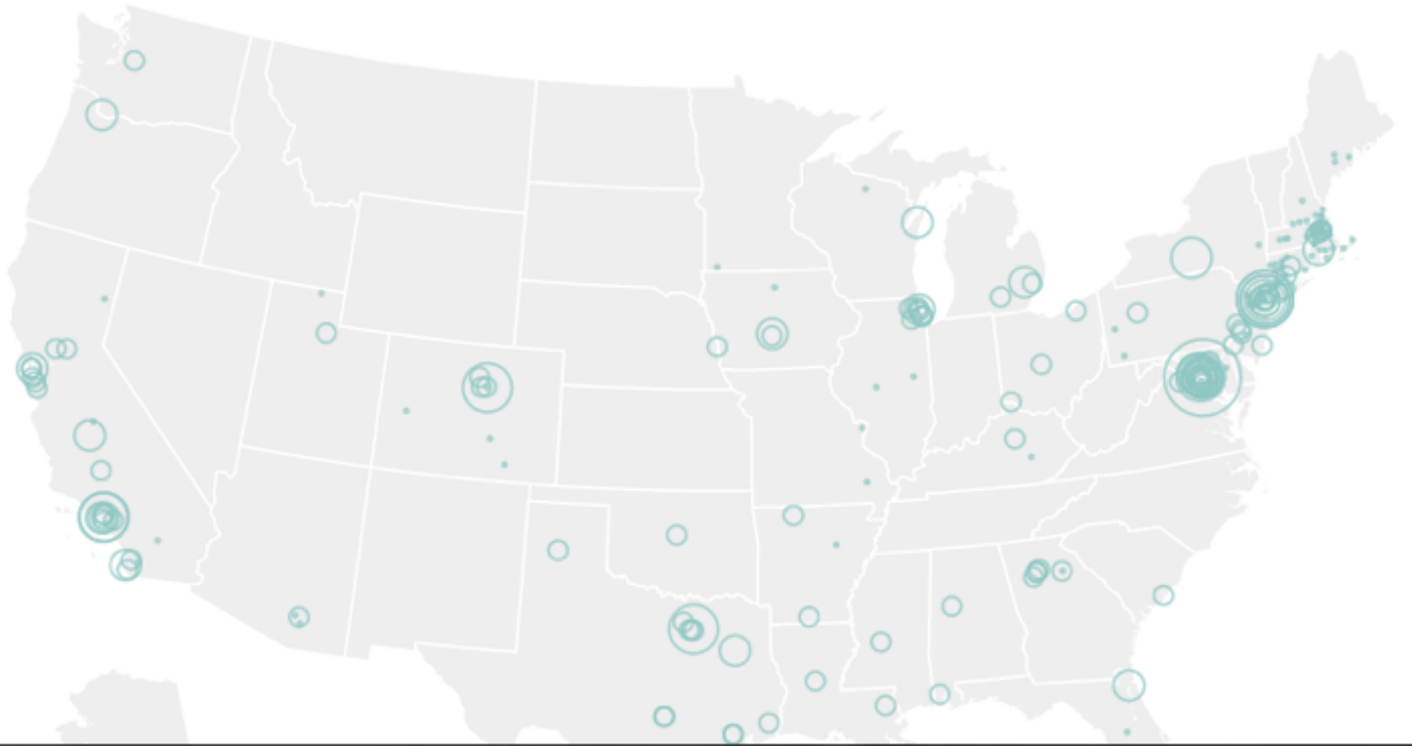
- abc CommitteeID
- abc Amendment_Indicator
- abc Report_Type
- abc Primary_General_Indicator
- # Microfilm_Location
- abc Trans_Type
- abc Entity_Type
- abc Contributor_Name
- abc City
- State
- abc Zip_Code
- abc Employer

abc Zip_Code

STRING LENGTHS



FREQUENT ZIP CODES



TRANSFORM EDITOR

```
set col: Zip_Code value: (length(Zip_Code) >= 9) ? left(Zip_Code,5) : Zip_Code
```

SCRIPT

```
splitrows col: column1 on: '\n'
split col: column1 on: '|' limit: 20
rename col: column2 to: 'CommitteeID'
rename col: column3 to: 'Amendment_Indicator'
```

Grid | 1D Profile | 2D Profile | 21 Columns | 2,391 Rows | 4 Data Types

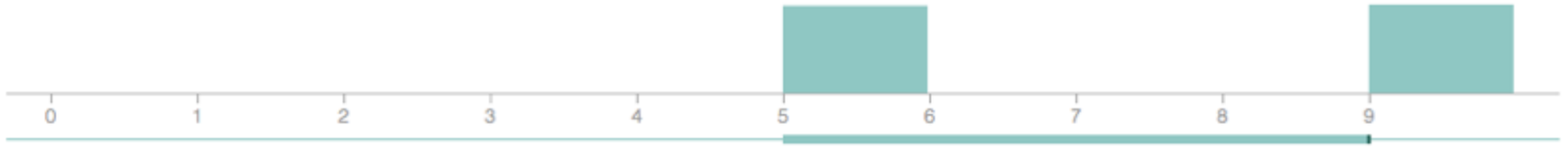
Filter data...

Filter columns...

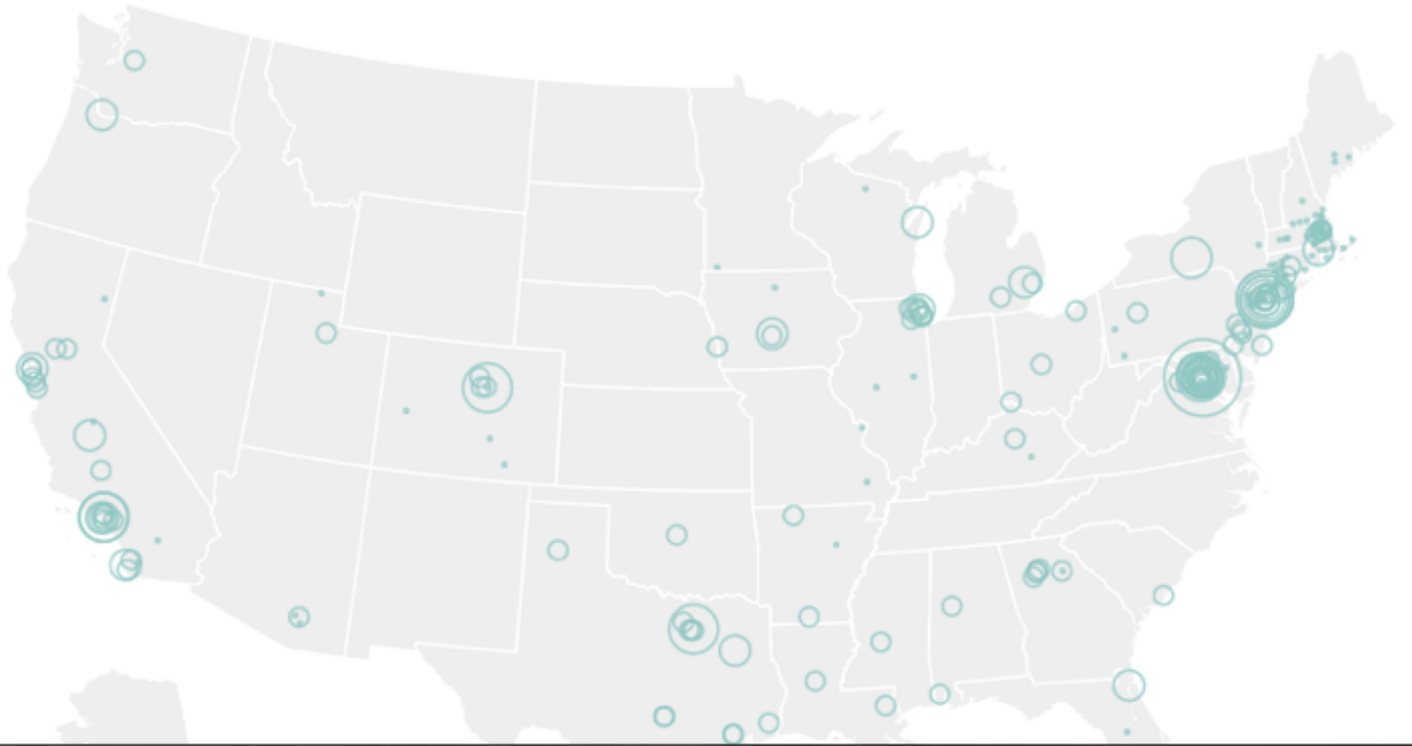
- abc CommitteeID
- abc Amendment_Indicator
- abc Report_Type
- abc Primary_General_Indicator
- # Microfilm_Location
- abc Trans_Type
- abc Entity_Type
- abc Contributor_Name
- abc City
- State
- abc Zip_Code
- abc Employer

abc Zip_Code

STRING LENGTHS



FREQUENT ZIP CODES



TRANSFORM EDITOR

Enter transform expression

SCRIPT

```
delete row: length(Zip_Code) == 8
set col: Zip_Code value: (length(Zip_Code) >= 9) ? left(Z
```

Grid | 1D Profile | 2D Profile | 21 Columns | 2,391 Rows | 4 Data Types

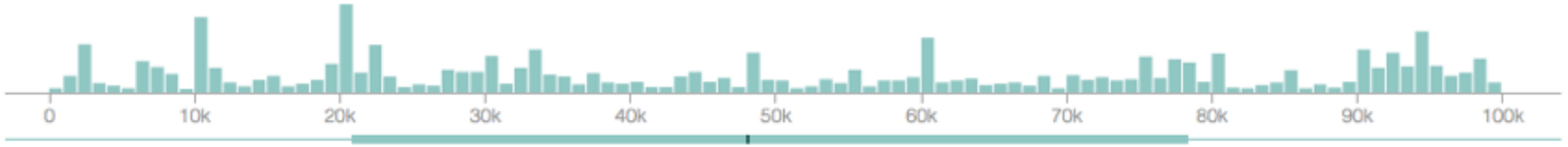
Filter data...

Filter columns...

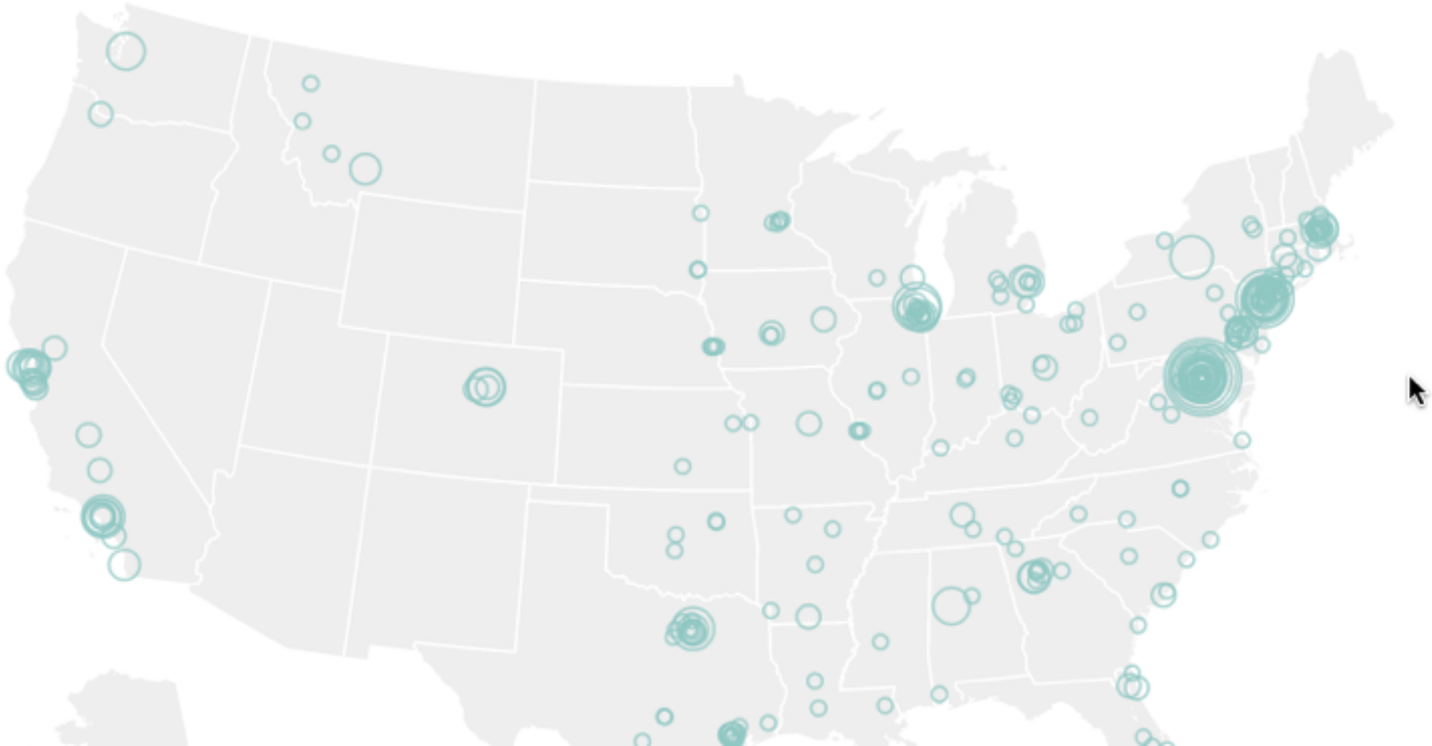
- abc CommitteeID
- abc Amendment_Indicator
- abc Report_Type
- abc Primary_General_Indicator
- # Microfilm_Location
- abc Trans_Type
- abc Entity_Type
- abc Contributor_Name
- abc City
- State
- # Zip_Code
- abc Employer

Zip_Code

VALUE HISTOGRAM



FREQUENT ZIP CODES



DESIGN DESIDERATA

Automatic visualization presentation

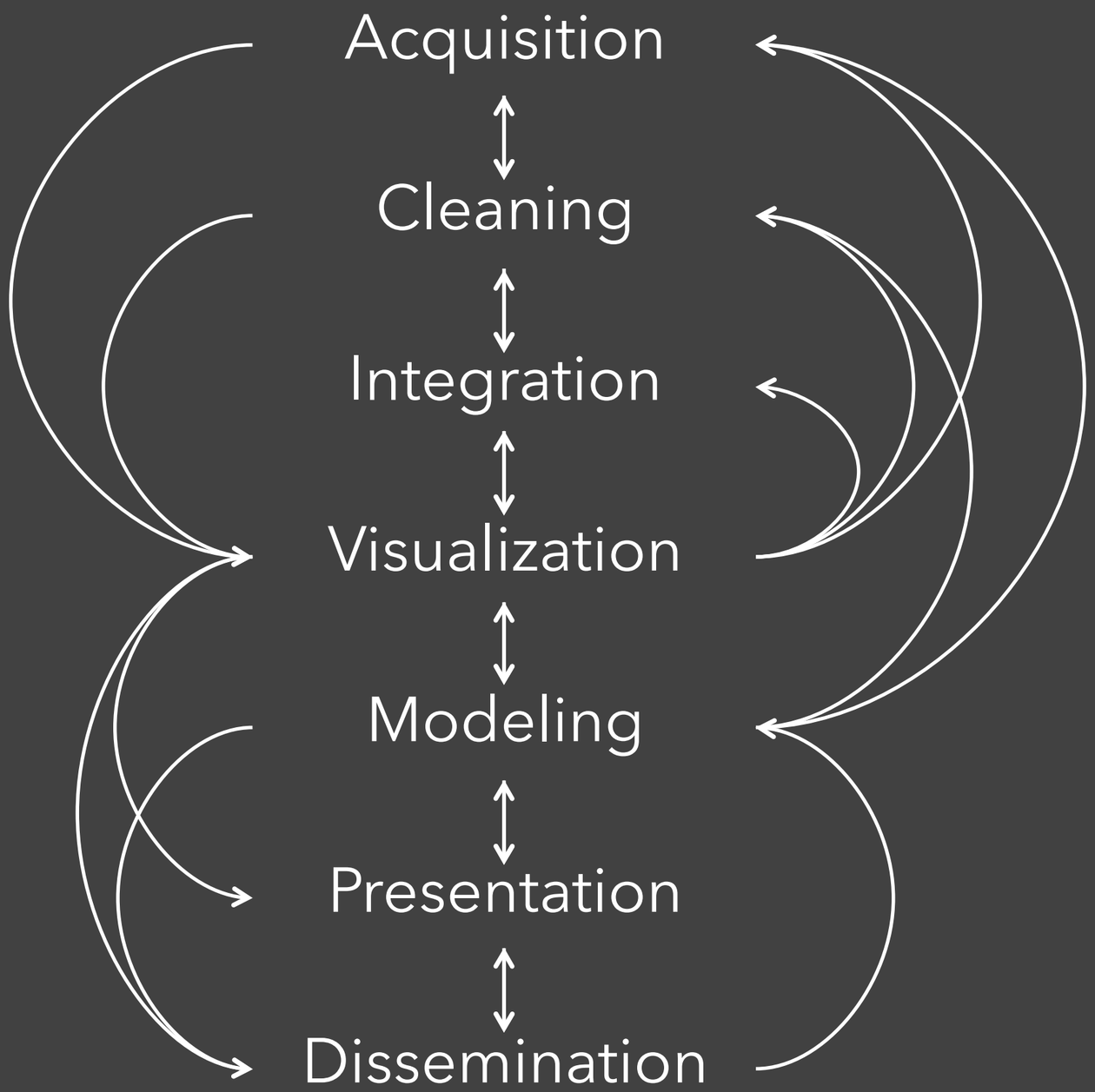
Perceptually effective encodings

Visualize all data, big or small

Springboard correction & refinement

Trifacta v1:
Enable Interactive Data Transformation

Trifacta v2:
Accelerate Data Discovery & Profiling


Acquisition
Cleaning
Integration
Visualization
Modeling
Presentation
Dissemination

DESIGNING *WITH*

DATA

Jeffrey Heer @jeffrey_heer
http://www.trifacta.com

TRIFACTA

Visit us at our booth or at

http://www.trifacta.com