

VizCommander

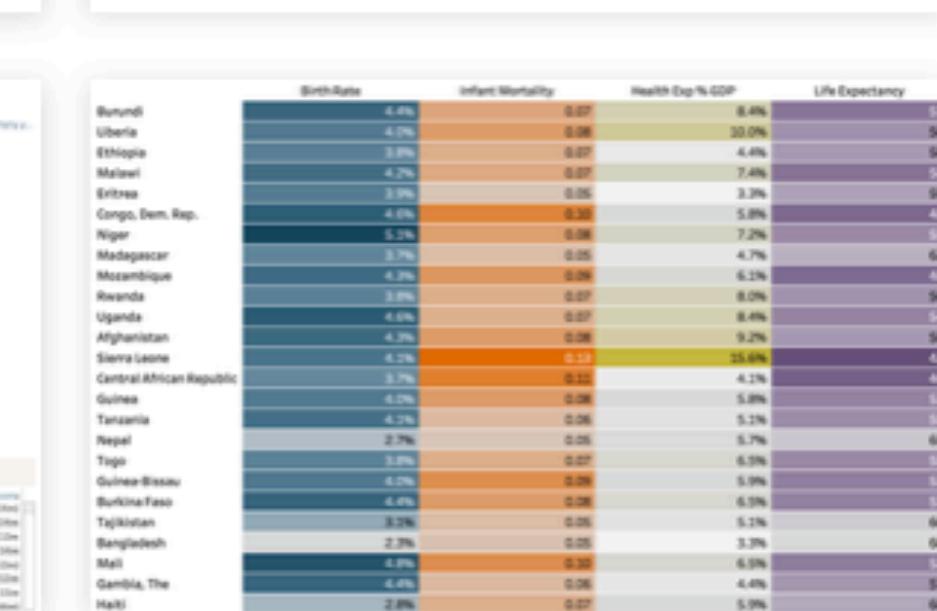
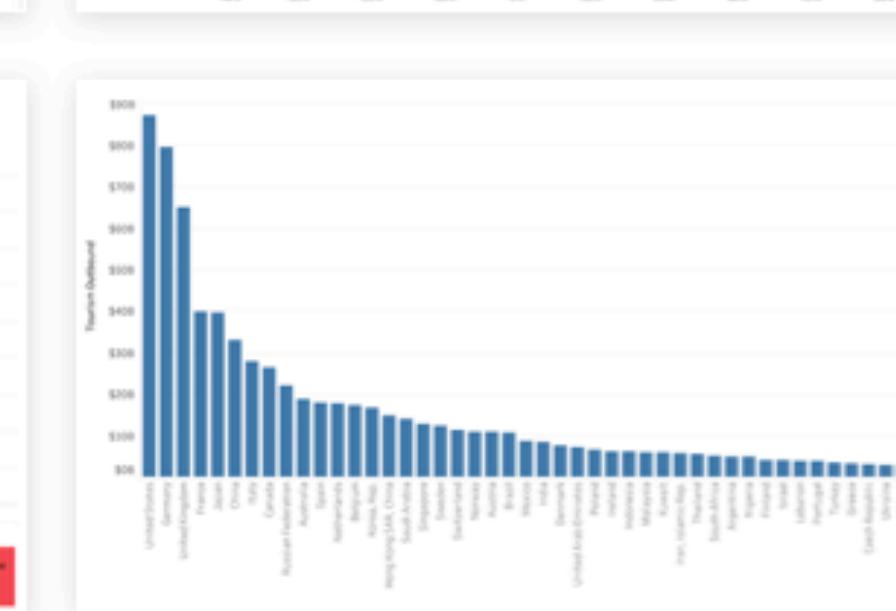
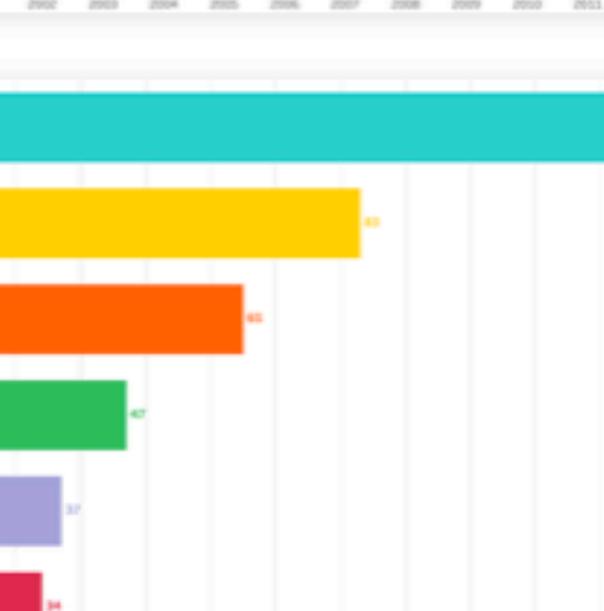
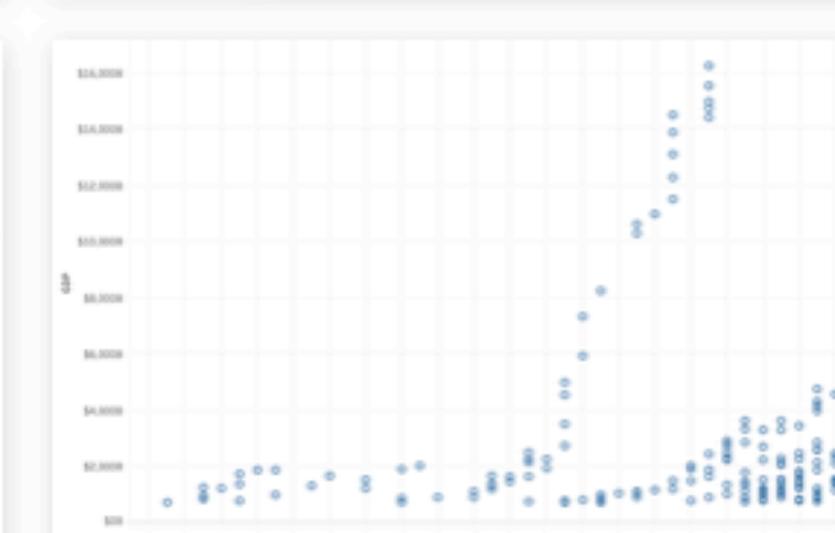
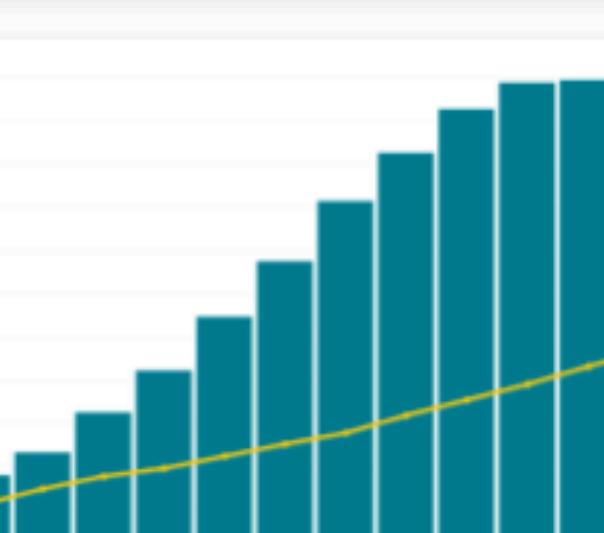
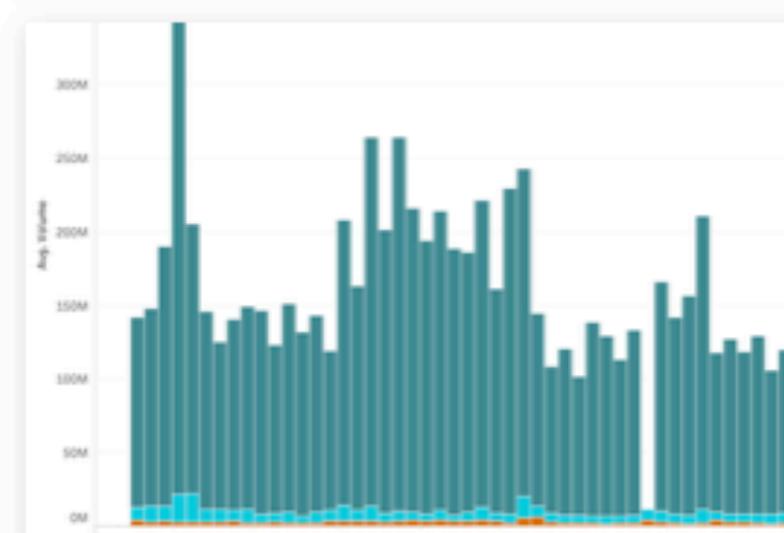
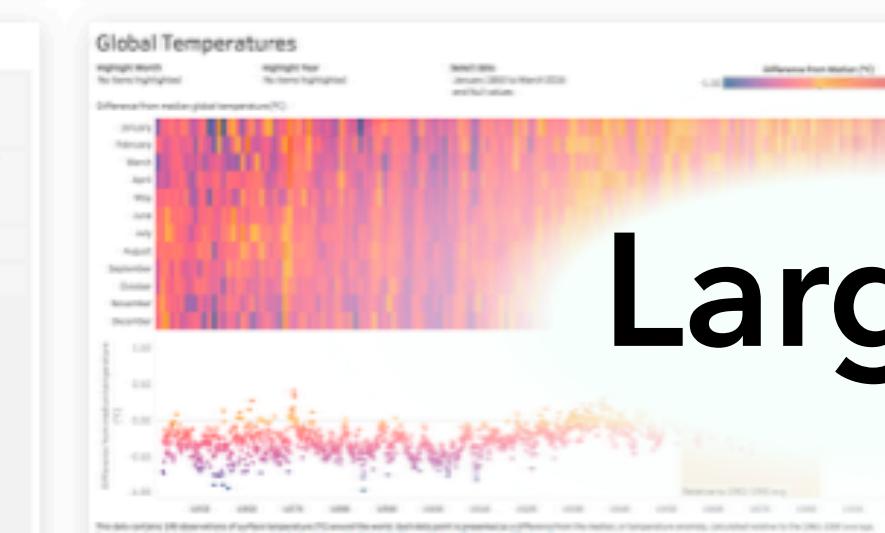
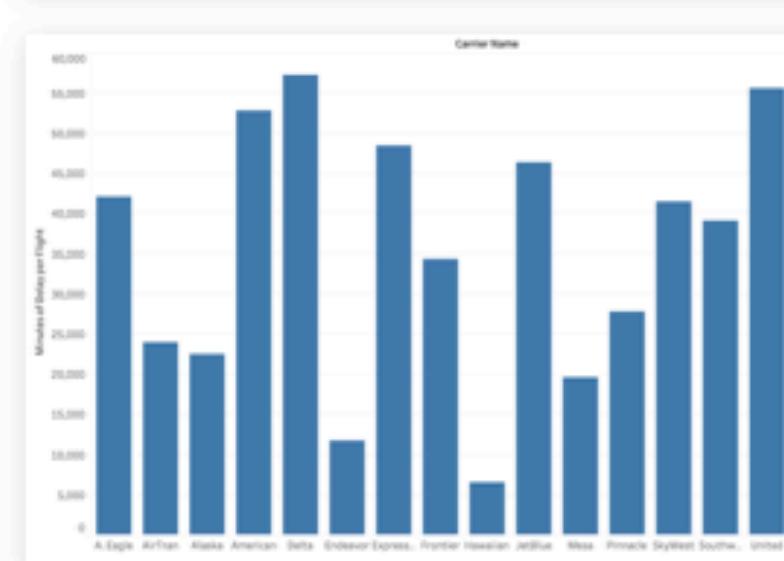
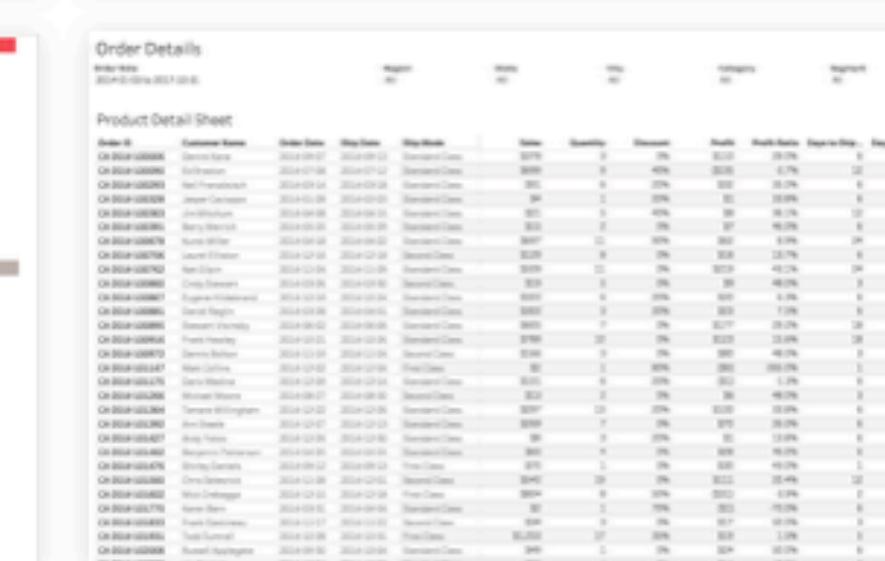
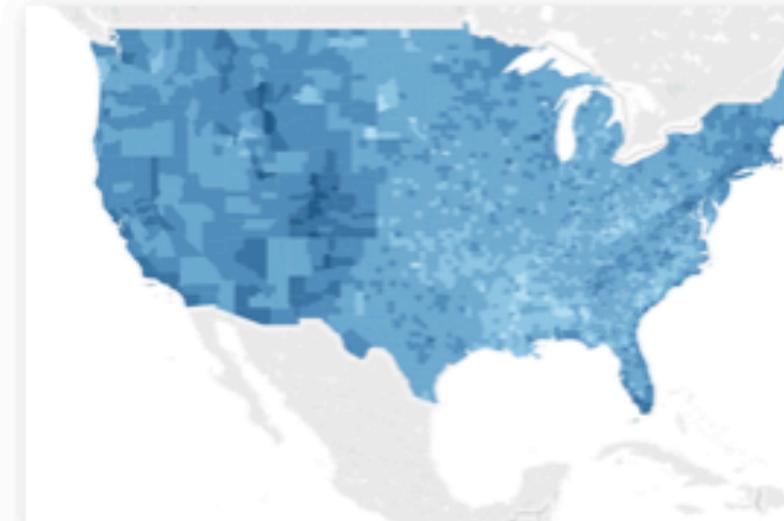
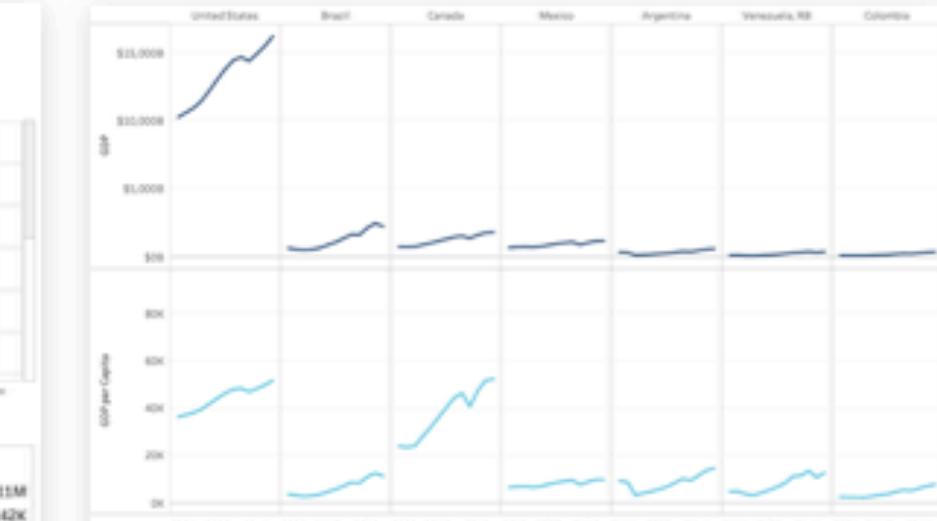
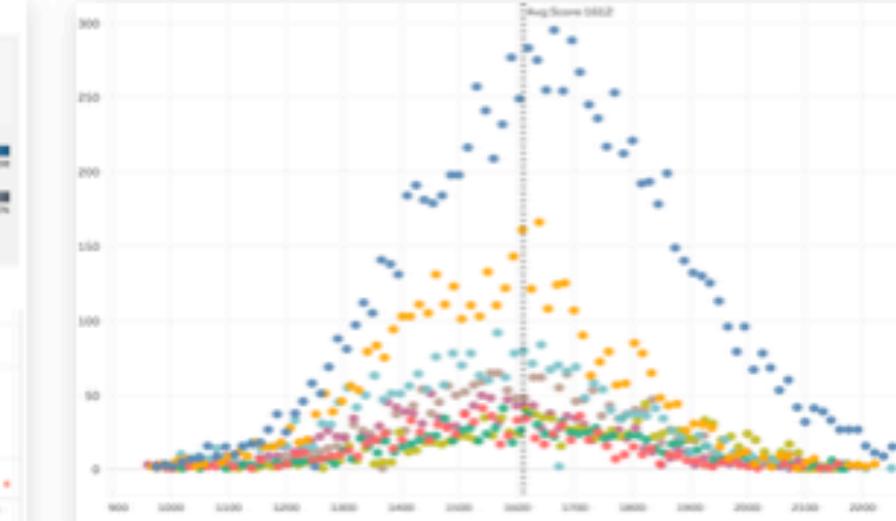
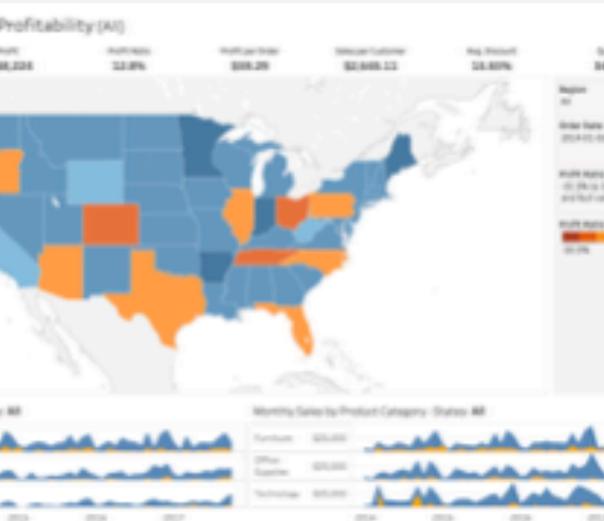
Computing Text-Based Similarity in Visualization Repositories for Content-Based Recommendations

Michael Oppermann, Robert Kincaid, and Tamara Munzner

Conference Talk, IEEE VIS 2020

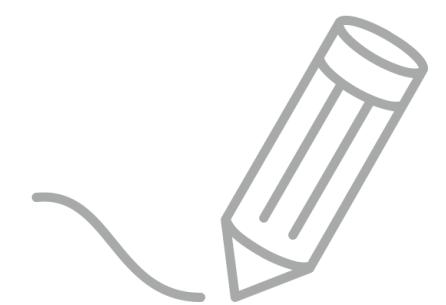
 michaeloppermann.com/work/viz-commander

Large-scale visualization repositories





**Users have difficulty
discovering relevant content.**



**Users often start from scratch
instead of reusing content.**

**Recommendation systems are increasingly used
to assist users by surfacing relevant content.**



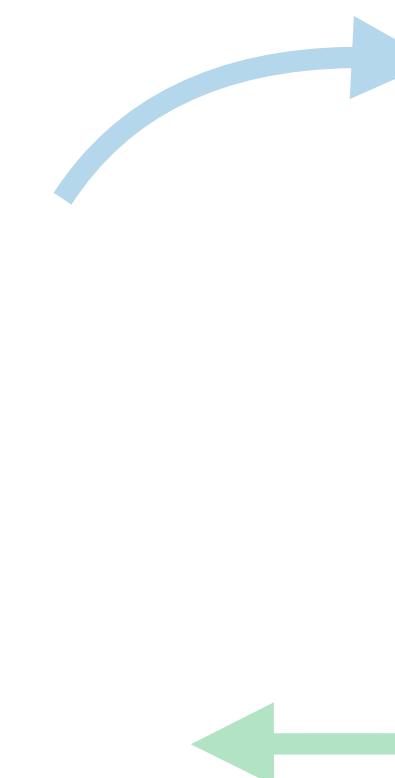
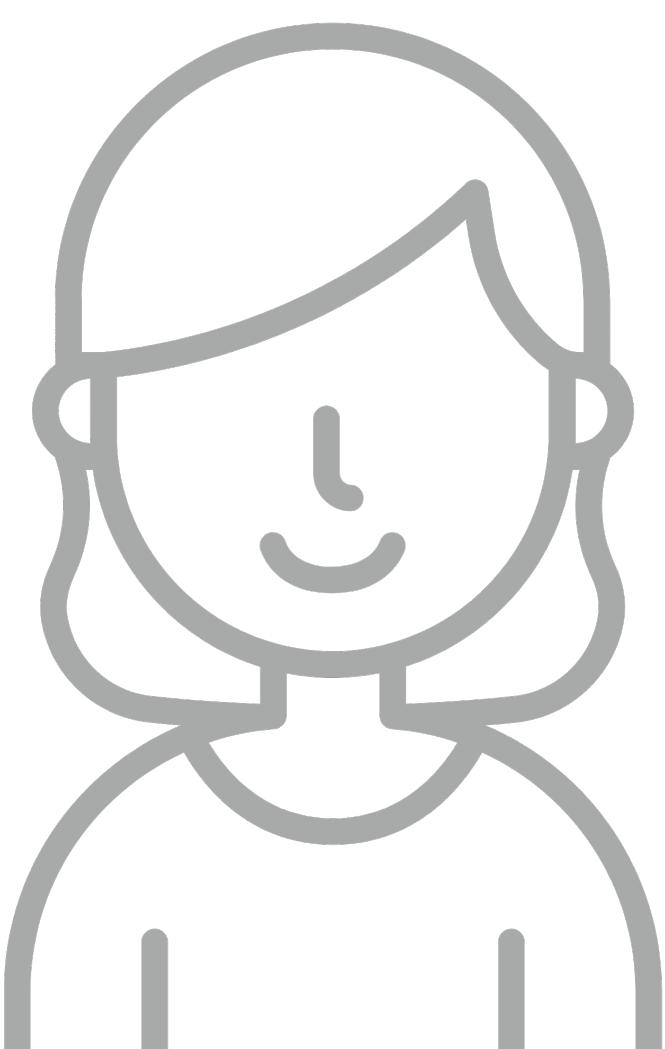
Visual encoding recommendation

Tableau ShowMe, Voyager, Draco, Data2Vis, ...

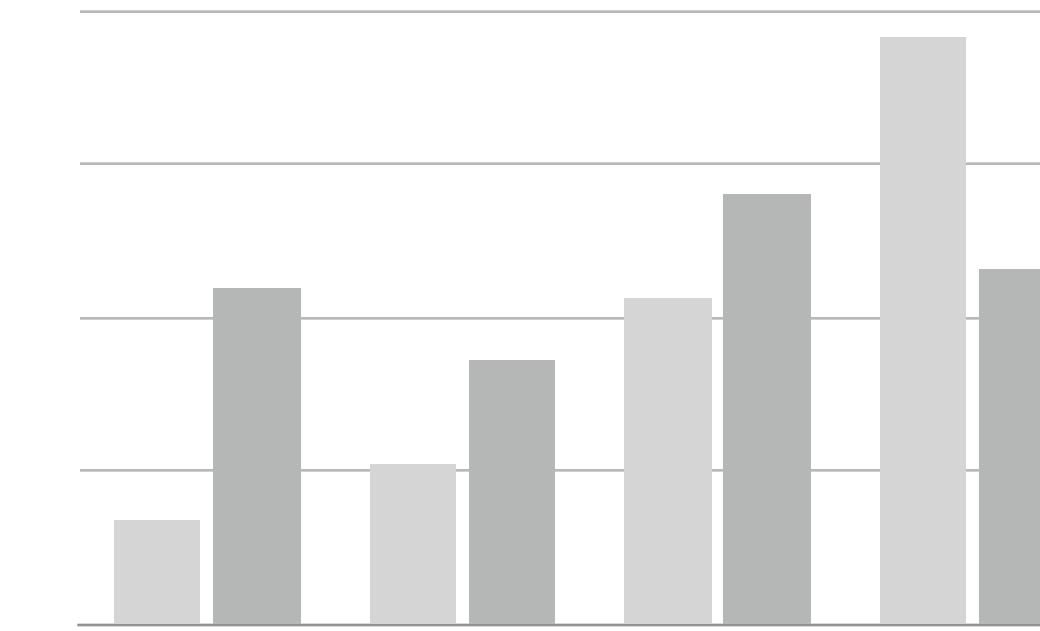


~~Visual encoding recommendation~~

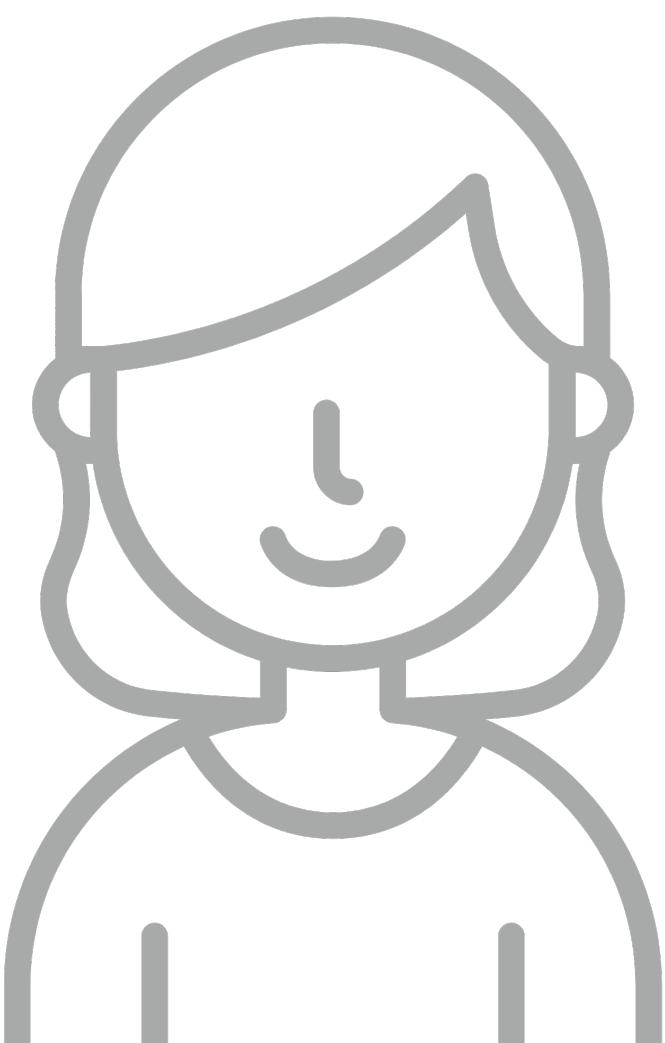
Tableau ShowMe, Voyager, Draco, Data2Vis, ...



uuu	uuuu	uu
uu	uuuuu	uuu
uuu	uuuu	uu
uu	uuuu	u
u	uuuu	uu



Visualization workbook recommendations based on content features



Recommendation Systems

Content-based filtering

Collaborative filtering

Recommendation Systems

Content-based filtering

- ▶ Focus of our work
- ▶ Finding relevant items based on their actual content
- ▶ Less diverse but more accurate recommendations
- ▶ Allows identification of *near-duplicate* items

Collaborative filtering

Recommendation Systems

Content-based filtering

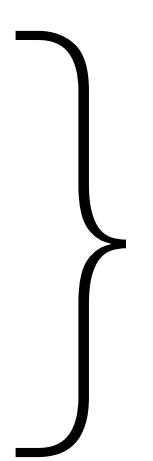
Collaborative filtering

- ▶ Recommendations based on user interactions
- ▶ Requires no domain knowledge, allows fast computation, serendipitous recommendations
- ▶ *Cold start* problem for new items or new users

Recommendation Systems

Content-based filtering

Collaborative filtering

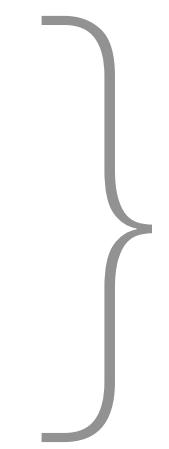


Hybrid system

Recommendation Systems

Content-based filtering

Collaborative filtering



Hybrid system

**Which content features are most informative
for comparisons?**

**What techniques can we use for comparing
and ranking viz specifications?**

Text-based similarity measure

- ▶ Content-based recommendations
- ▶ Facilitate information seeking



Overview

Close collaboration with the *Recommender Systems Group* at Tableau

Overview

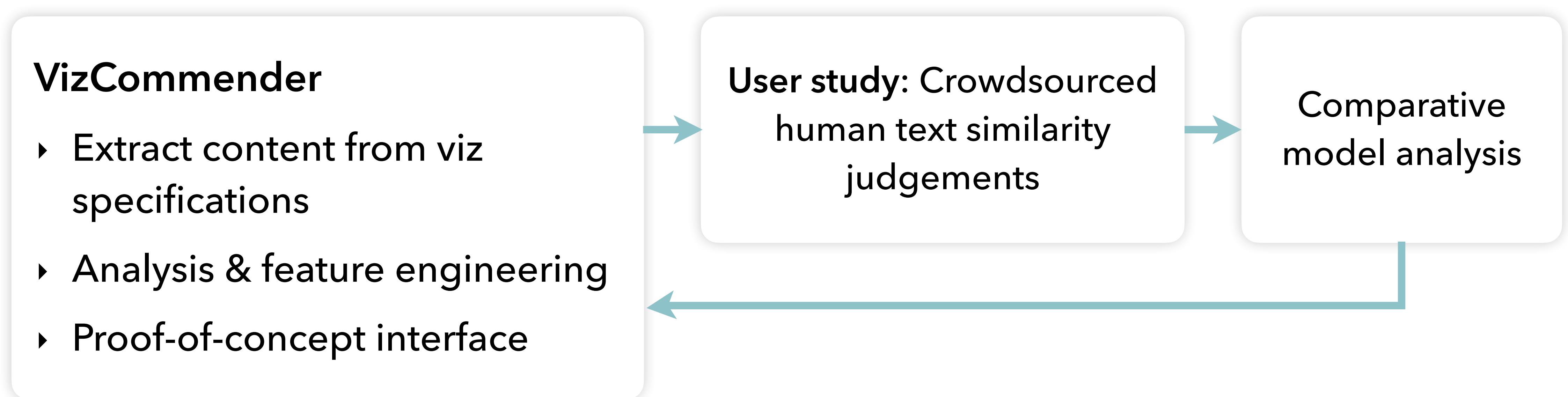
Close collaboration with the *Recommender Systems Group* at Tableau

VizCommander

- ▶ Extract content from viz specifications
- ▶ Analysis & feature engineering
- ▶ Proof-of-concept interface

Overview

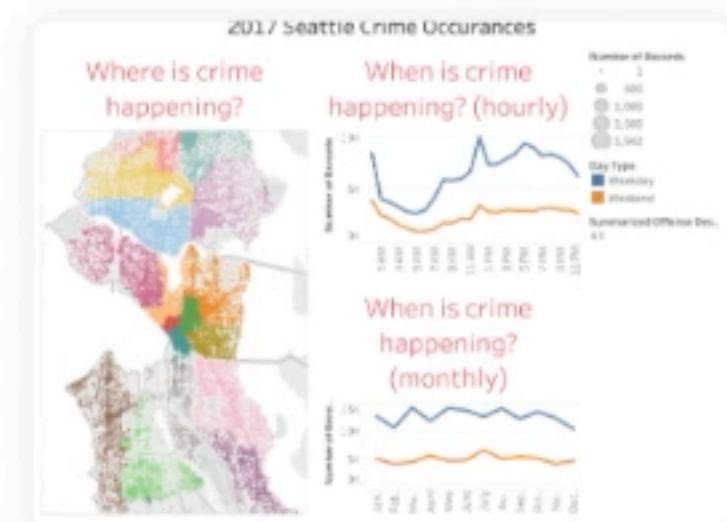
Close collaboration with the *Recommender Systems Group* at Tableau





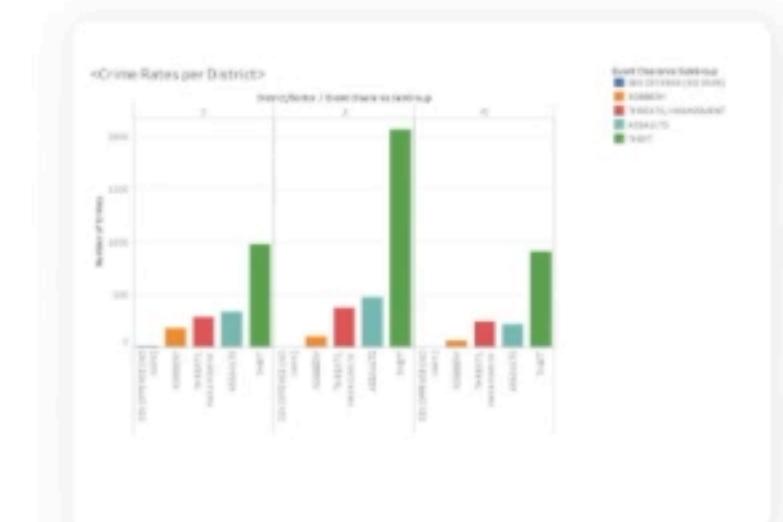
crime when car financial rate reported district finance weapon expense bar chart line chart geo map scatter plot table

Sort by relevance ▾



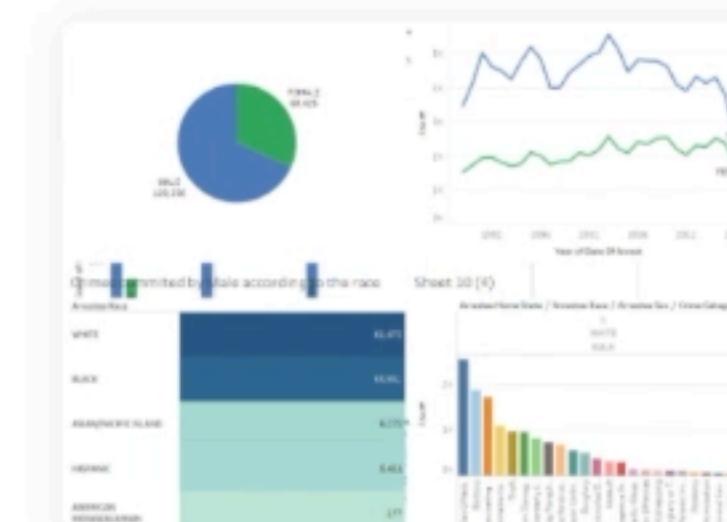
Seattle Crime

caitlin.streamer • 2018-04-09



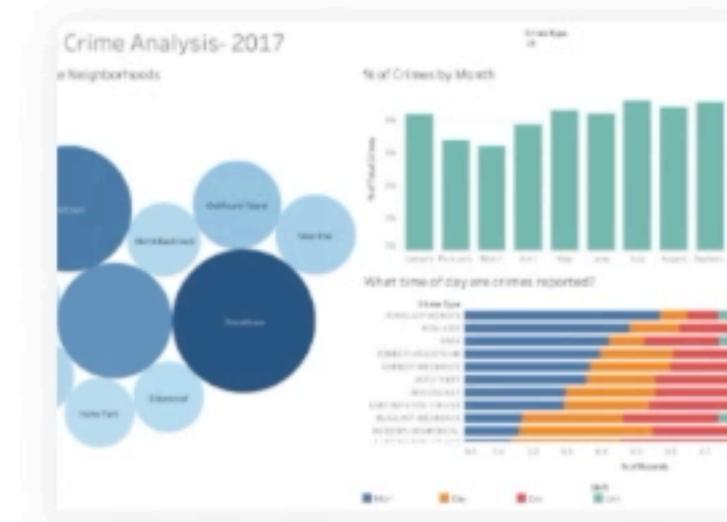
HCDE 210

arwa.mohammed6769 • 2016-11-23



urbana_crimes

ashish.khanal • 2018-03-11



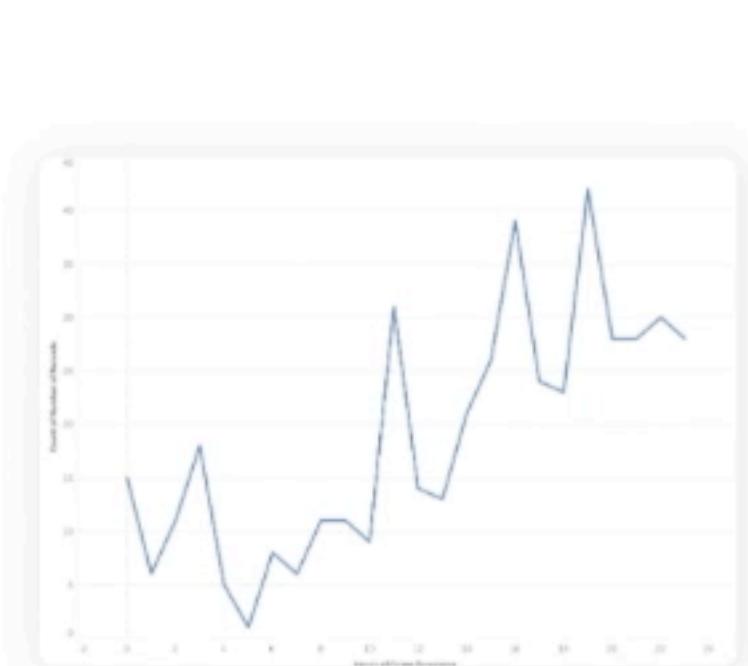
ATL Crime

megan2618 • 2018-06-07



Reported Cyber Crime

vishu.rahar • 2018-04-03



arwa.mohammed6769 • 2016-11-23



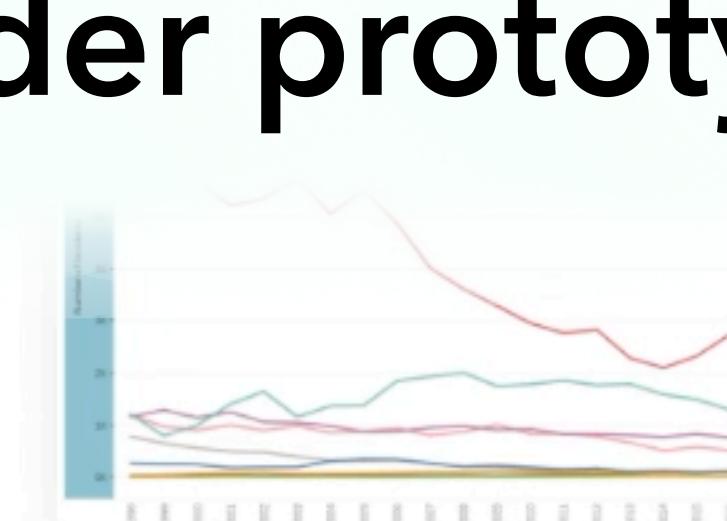
hekma • 2017-02-28



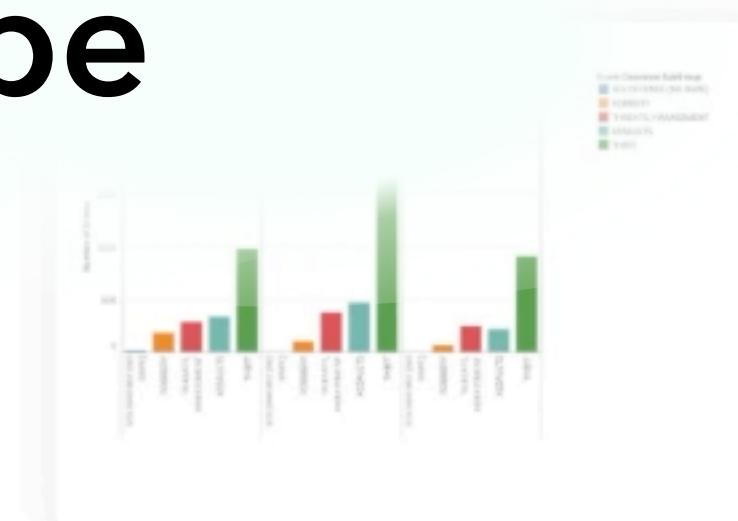
latauya • 2016-01-21



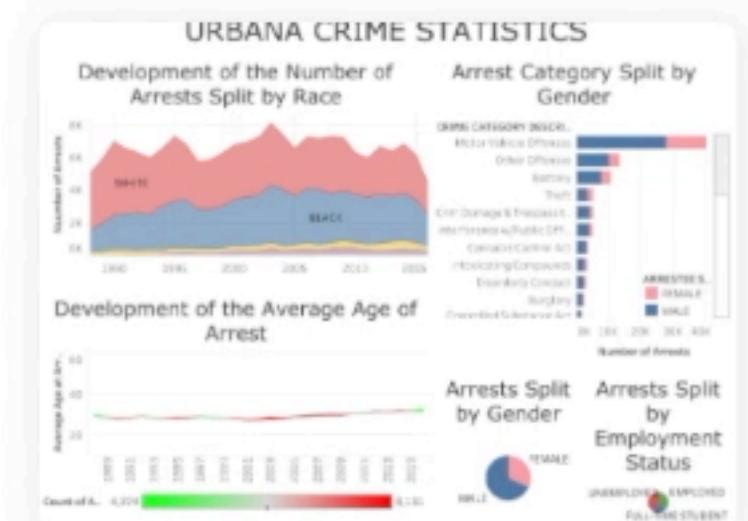
andrew1738 • 2018-05-21



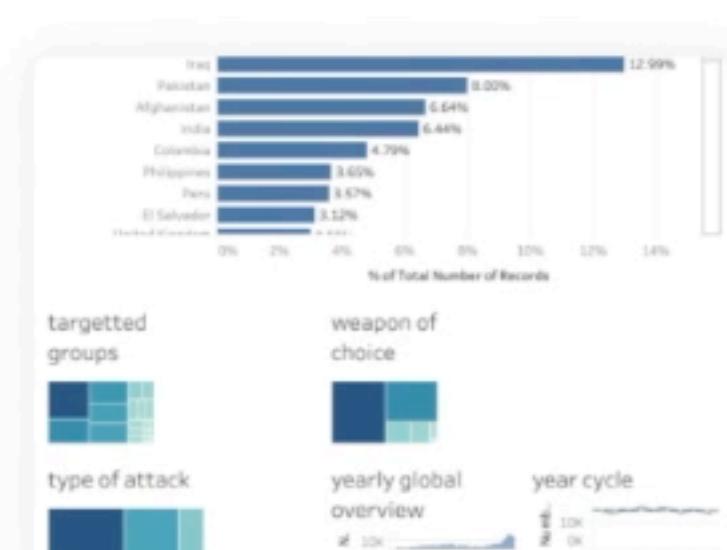
heather.keary • 2017-12-01



arwa.mohammed6769 • 2016-11-23

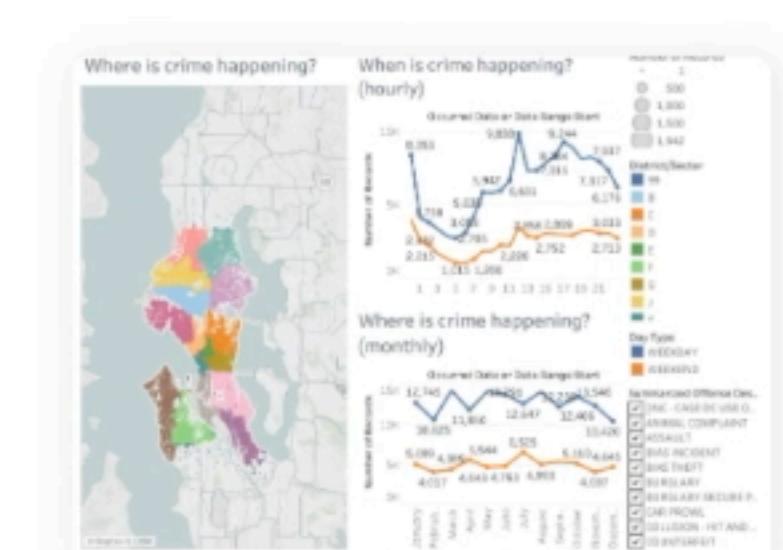


christoph.pressler • 2016-10-09



terrorism overview

olivier2575 • 2018-11-08



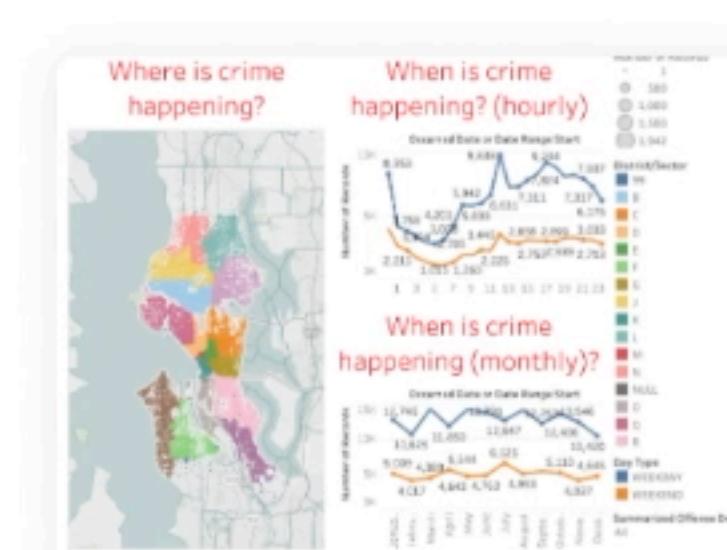
Bunch of Seattle Criminal

blair3220 • 2018-04-03



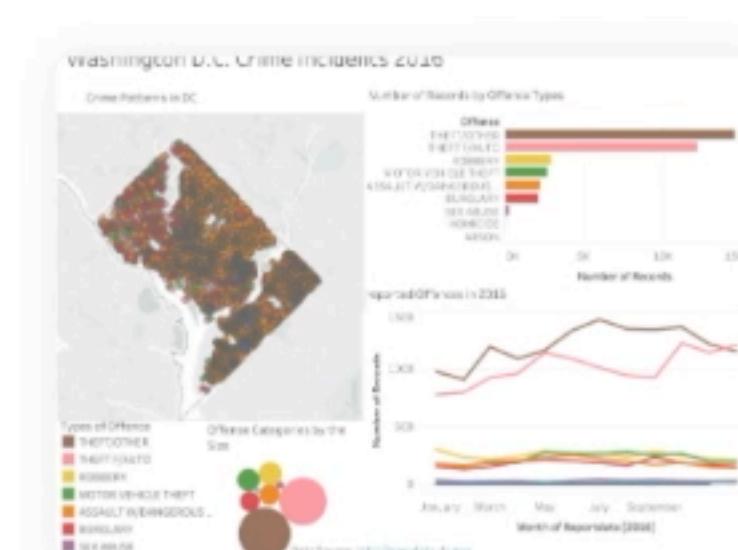
PSA Findings by Year

kevin4543 • 2017-12-15



#BLESSEDLIT

michael.valeri • 2018-04-03



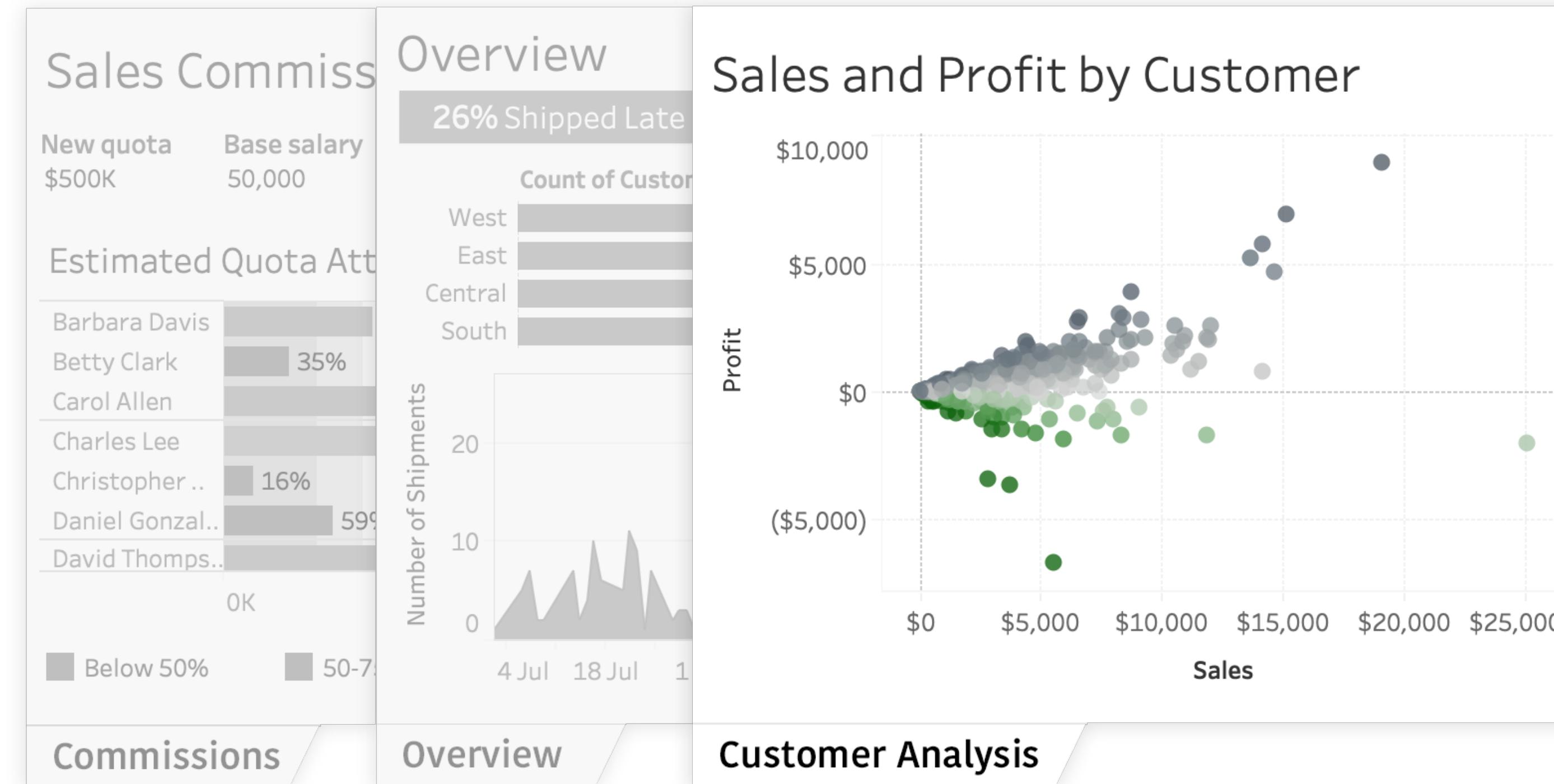
Data Visualization

danadaree • 2018-04-09

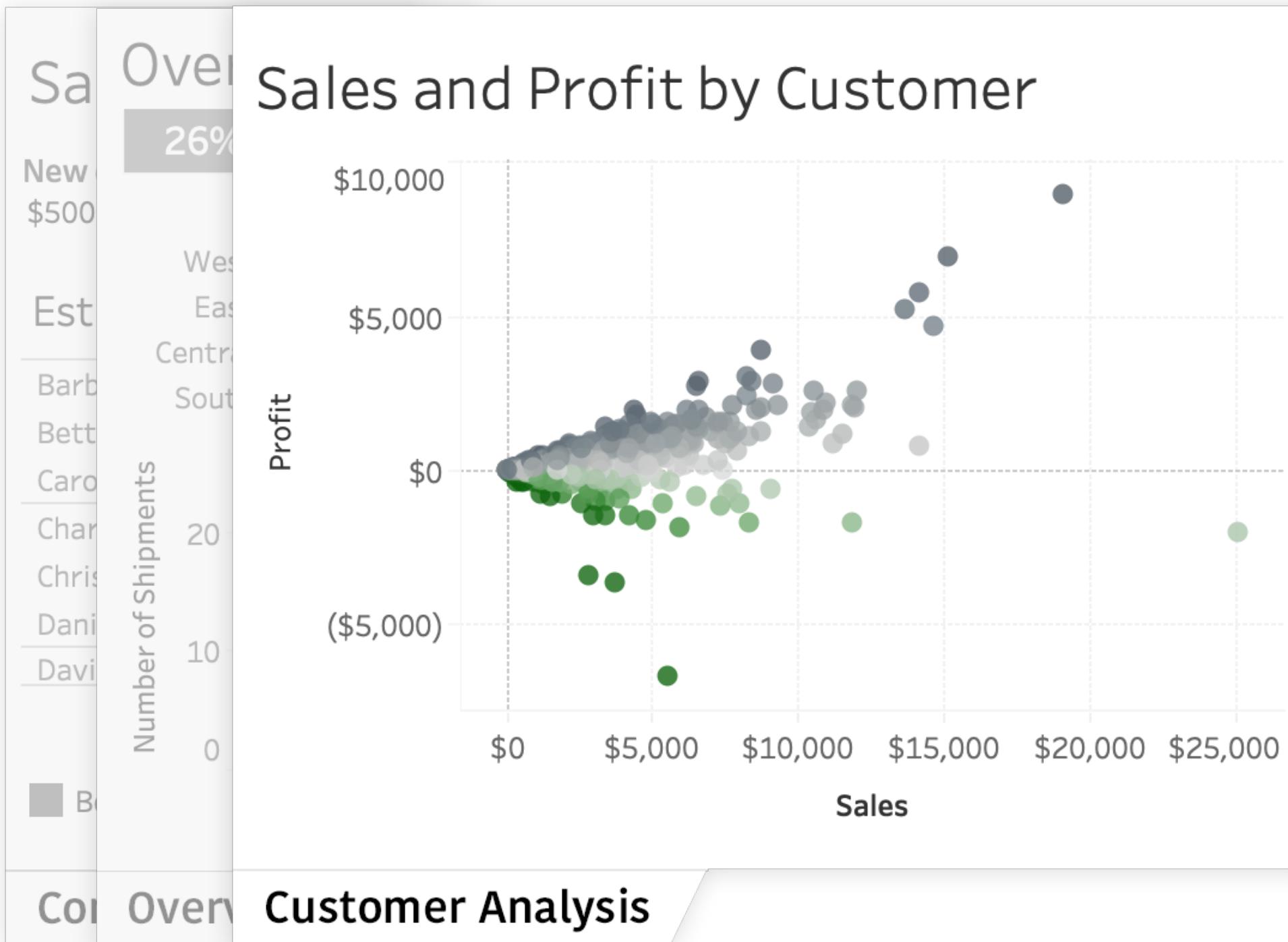


arwa.mohammed6769 • 2016-11-23

Tableau Visualization Workbook



Workbook

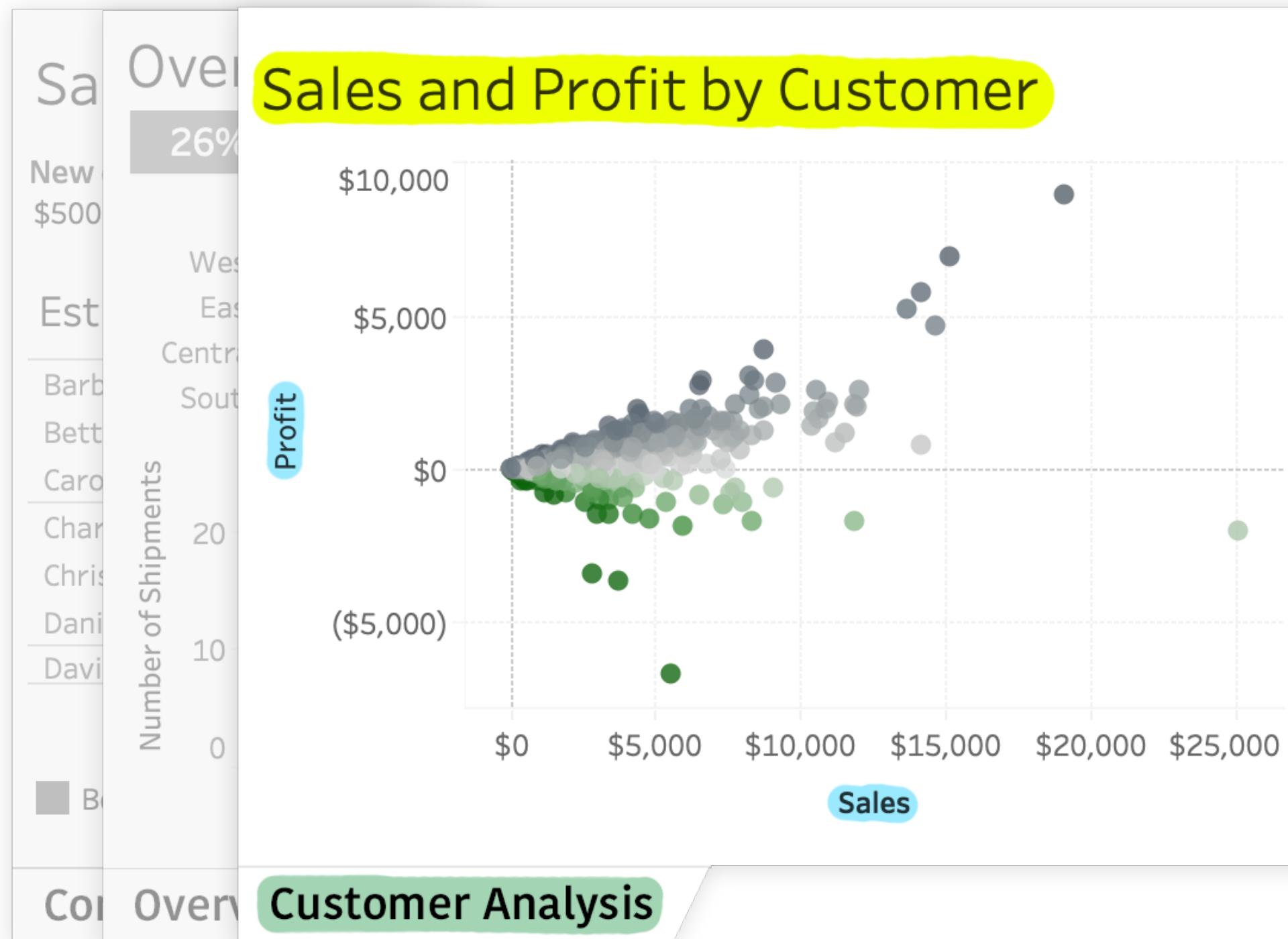


Visualization Specification

```
<worksheet name="Customer Analysis">
  <layout-options>
    <title>
      <formatted-text>
        <run>Sales and Profit by Customer</run>
      </formatted-text>
    </title>
  </layout-options>
  <table>
    <rows>...[sum:Profit:qk]</rows>
    <cols>...[sum:Sales:qk]</cols>
  </table>
  ...
</worksheet>
```



Workbook

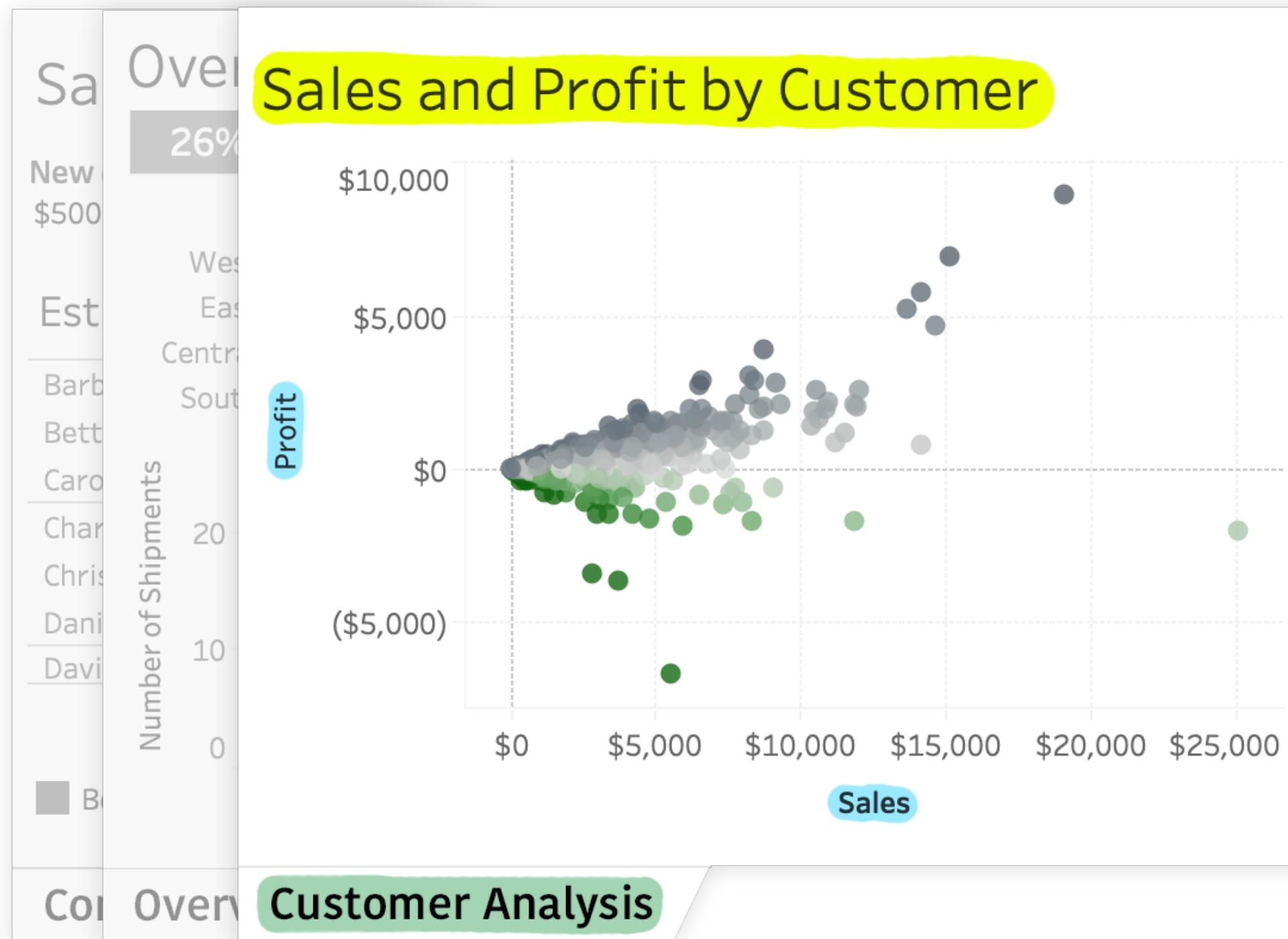


Visualization Specification

```
<worksheet name="Customer Analysis">
  <layout-options>
    <title>
      <formatted-text>
        <run>Sales and Profit by Customer</run>
      </formatted-text>
    </title>
  </layout-options>
  <table>
    <rows>...[sum:Profit:qk]</rows>
    <cols>...[sum:Sales:qk]</cols>
  </table>
  ...
</worksheet>
```



Workbook

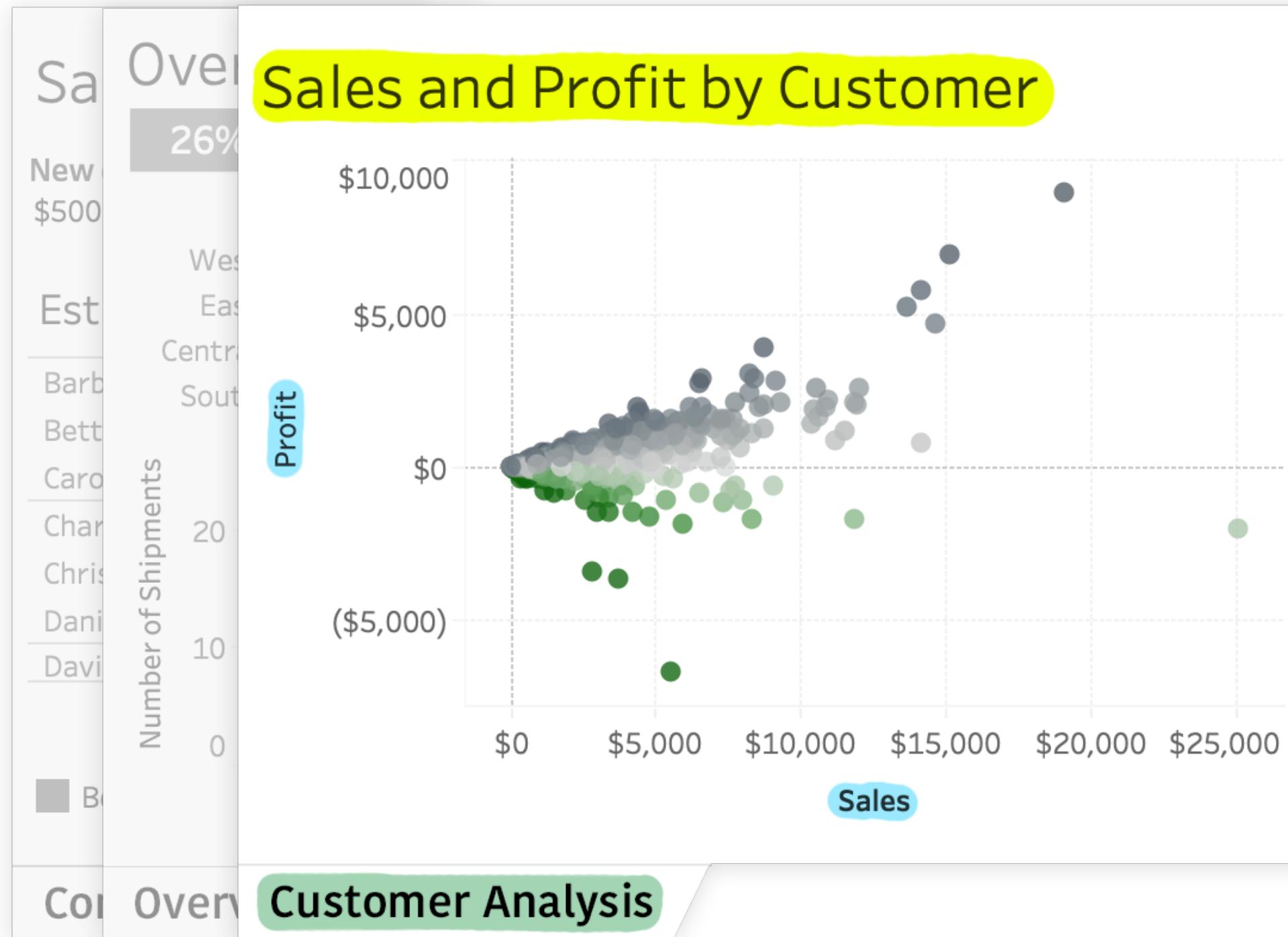


Visualization Specification

```
<worksheet name="Customer Analysis">
  <layout-options>
    <title>
      <formatted-text>
        <run>Sales and Profit by Customer</run>
      </formatted-text>
    </title>
  </layout-options>
  <table>
    <rows>...[sum:Profit:qk]</rows>
    <cols>...[sum:Sales:qk]</cols>
  </table>
  ...
</worksheet>
```

City	Customer Name	Sales	Discount	...

Workbook



City	Customer Name	Sales	Discount	...

Visualization Specification

```
<worksheet name="Customer Analysis">
  <layout-options>
    <title>
      <formatted-text>
        <run>Sales and Profit by Customer</run>
      </formatted-text>
    </title>
  </layout-options>
  <table>
    <rows>...[sum:Profit:qk]</rows>
    <cols>...[sum:Sales:qk]</cols>
  </table>
  ...
</worksheet>
```

```
<datasource>
  <metadata-records>
    <metadata-record class="column">
      <remote-name>City</remote-name>
      ...
    </metadata-record>
    <metadata-record class="column">
      <remote-name>Customer Name</remote-name>
      ...
    </metadata-record>
    <remote-name>Sales</remote-name>
```

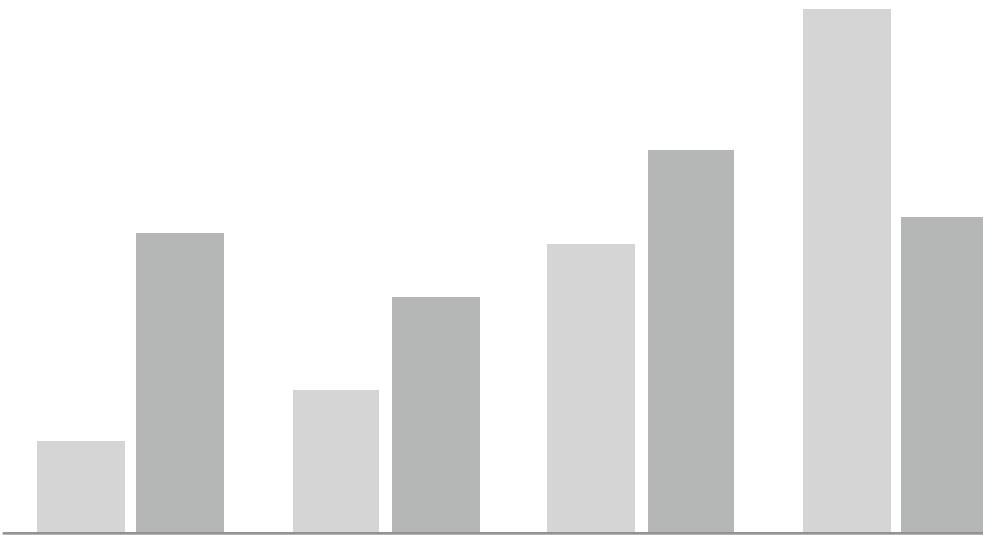


Initial Experiments

Extracted text

uuuu uuuu uuu uuu uuu
uuuuuuuu uuuuuu uuu uuu uuu
uu uuuu uuu uuu uuu uuu uuu
uu uuu uuuu uuu uuuuu
uuuuuuuu uuuuu uuu uuu
uuuuuuuu uuuuu uuu uuu
uuuu uuu uuuu uuuuu uuu uuu
uu uuu uuuu

Visual encodings

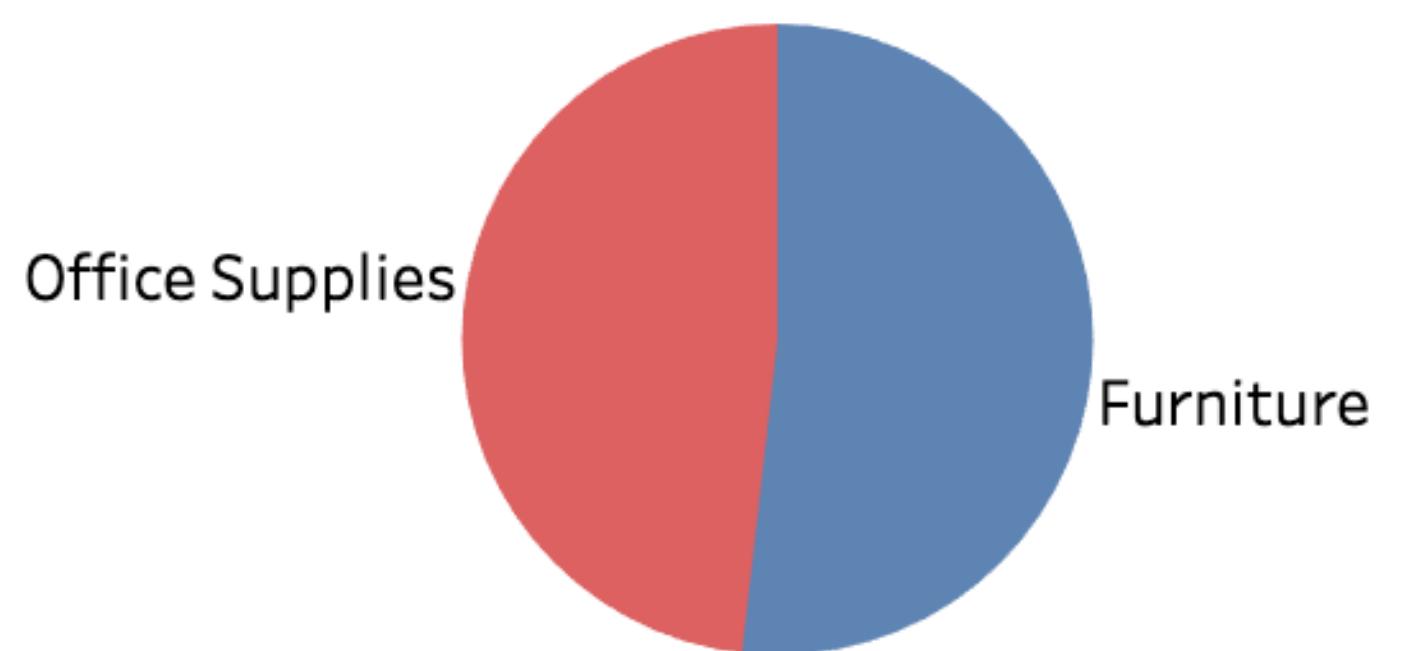


Different data, same encoding

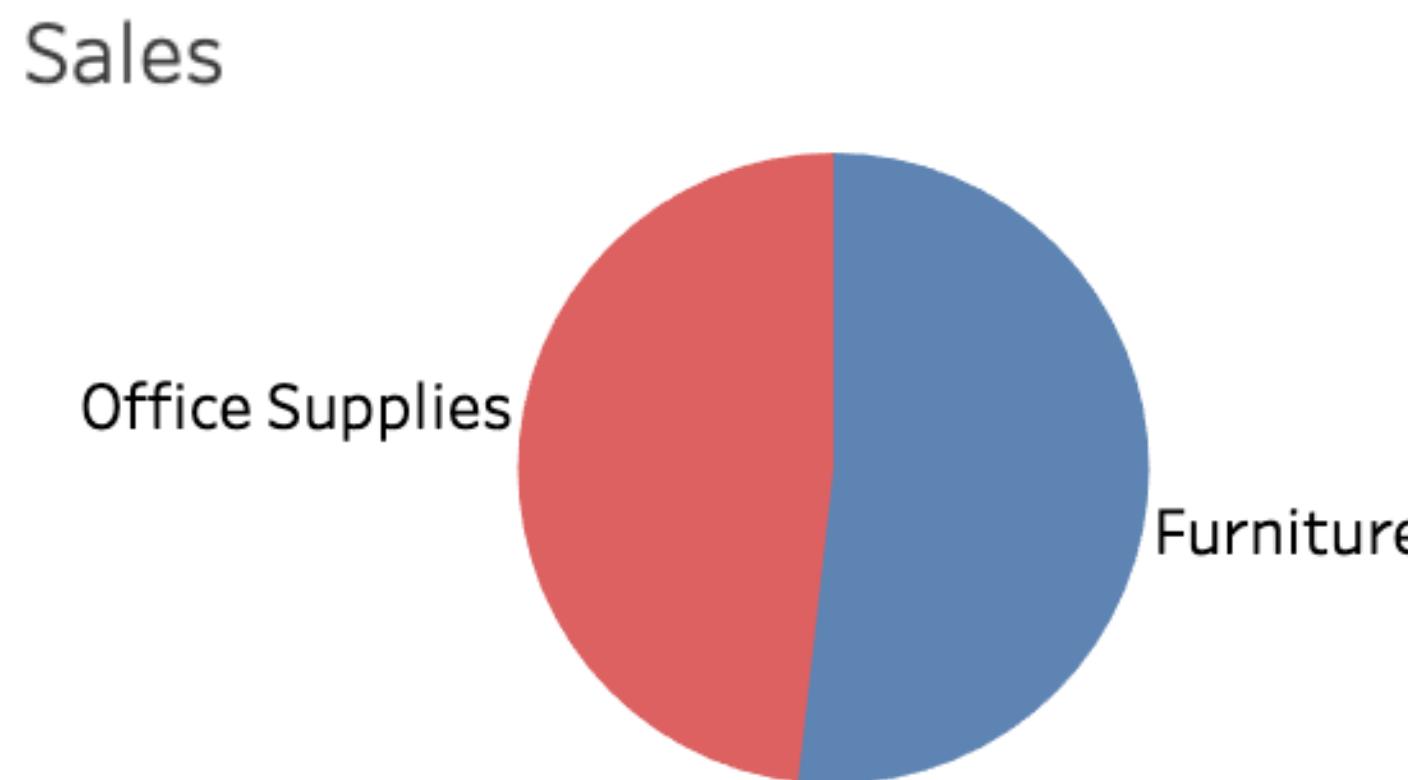
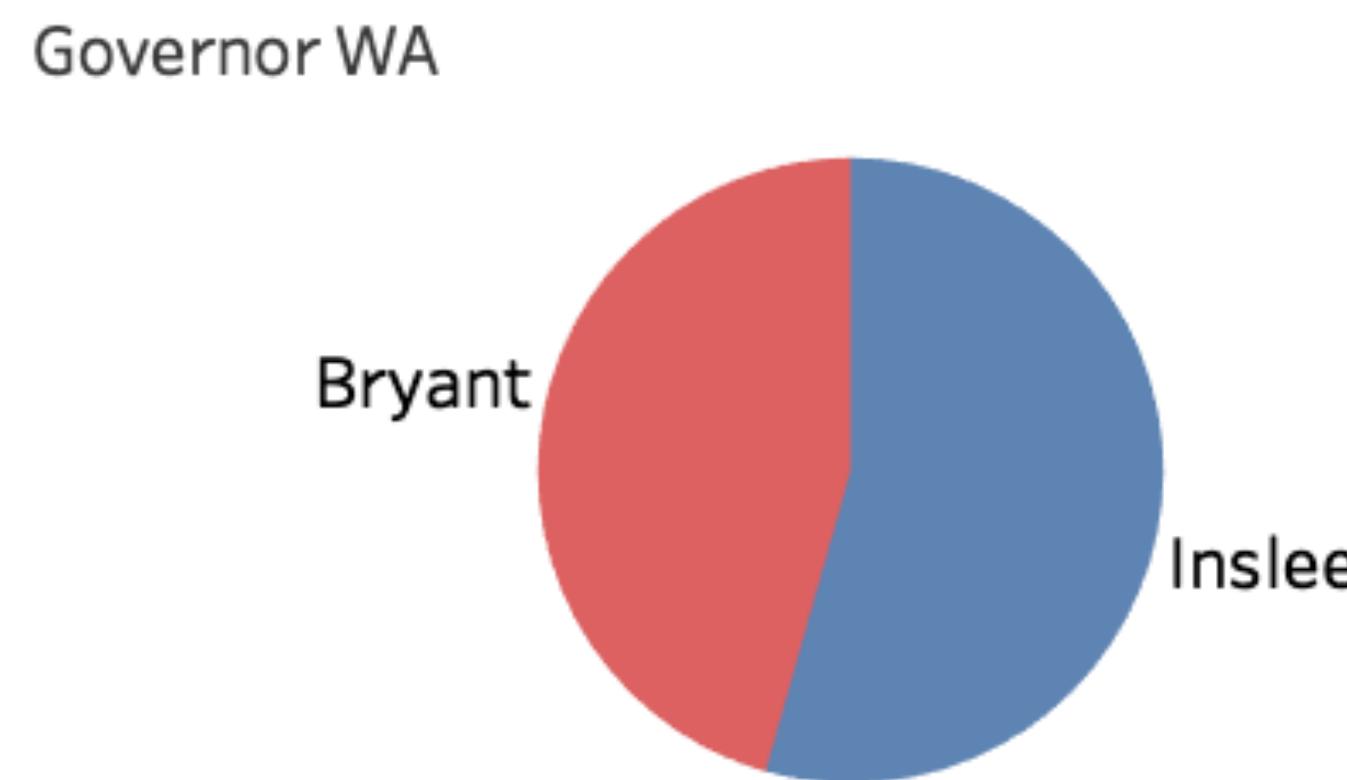
Governor WA



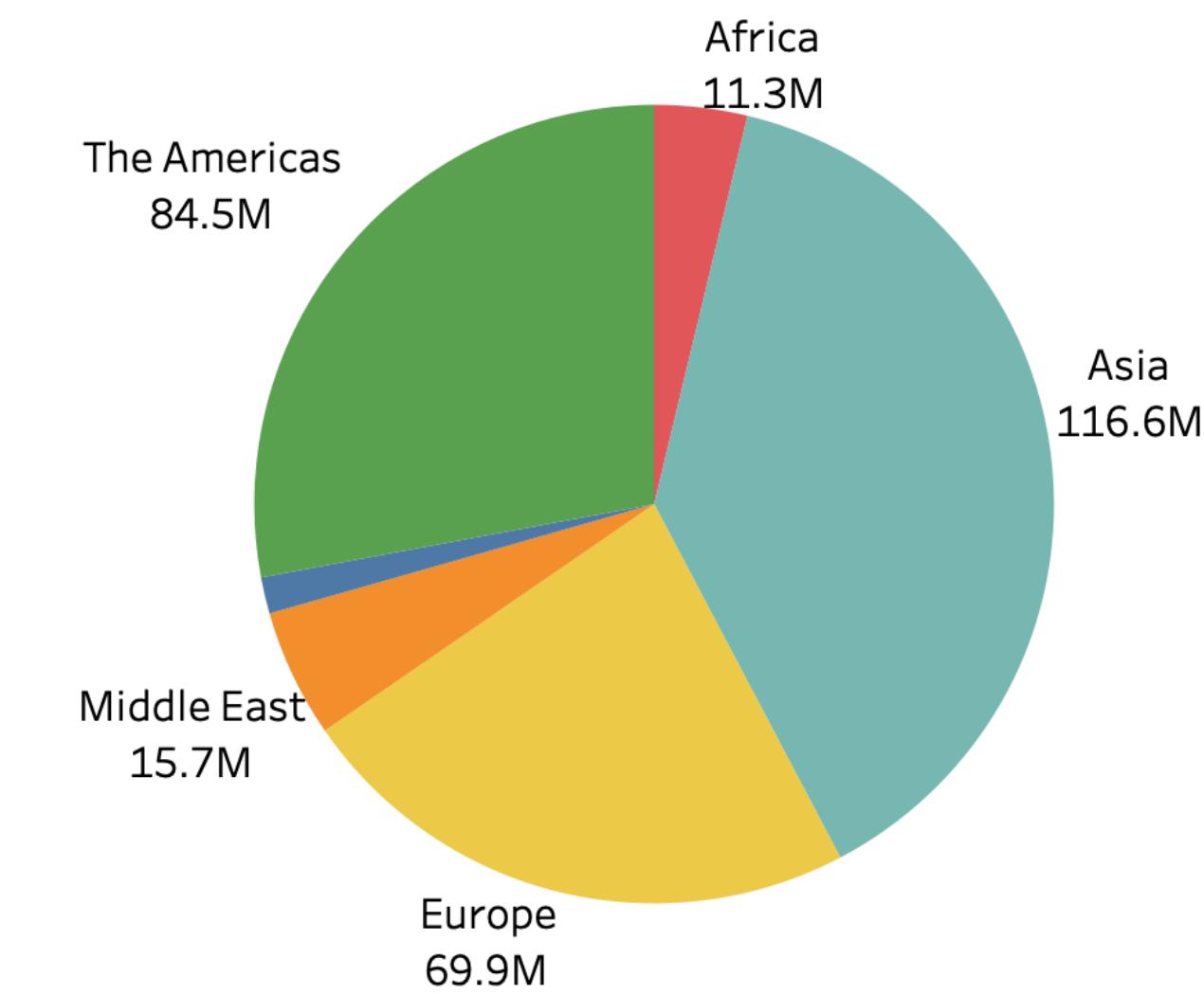
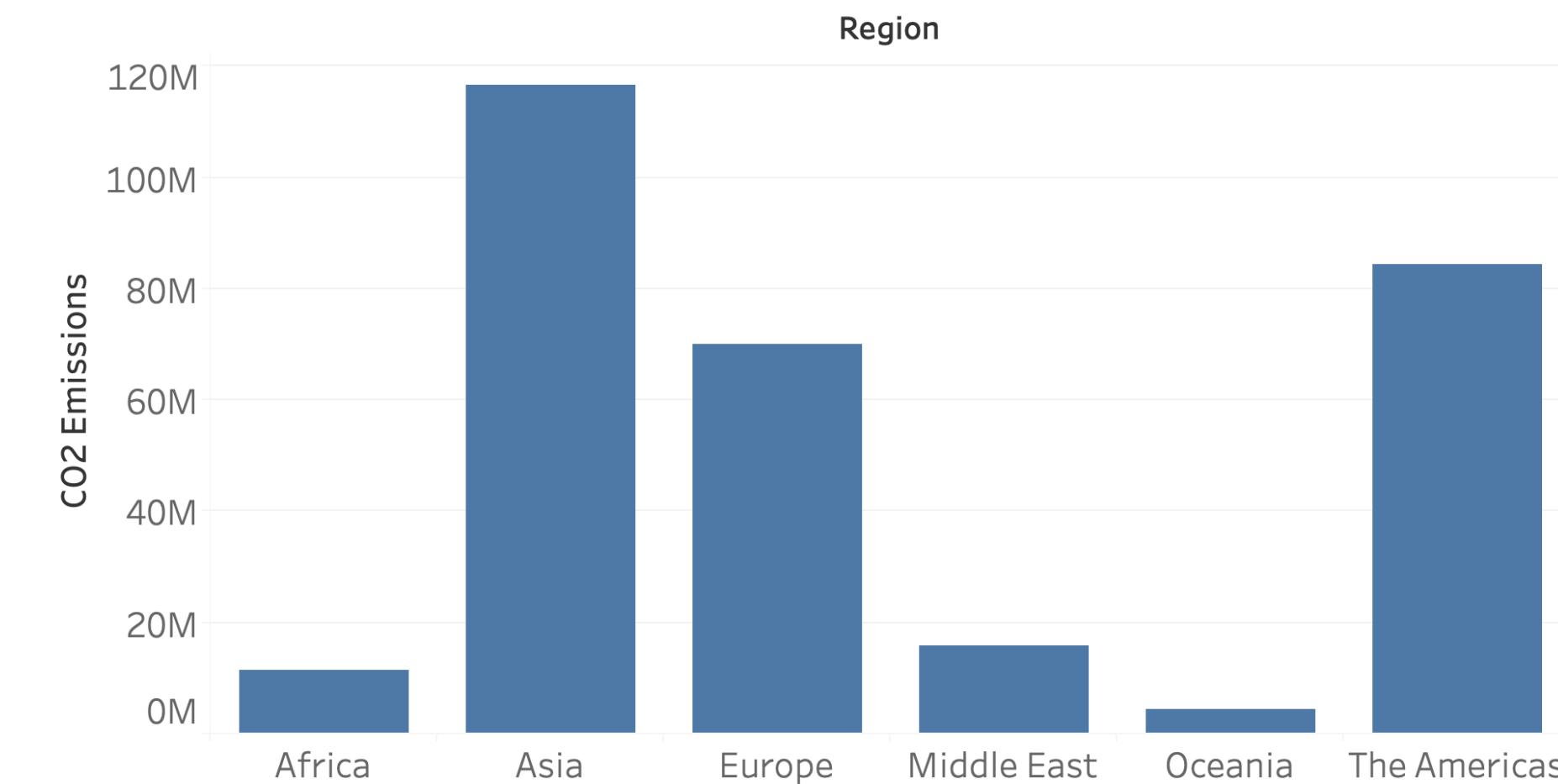
Sales



Different data, same encoding



Same data, different visual encodings



Features: Leaving Out Visual Encodings

Features: Leaving Out Visual Encodings

For design inspiration and learning

- Need visual style

Features: Leaving Out Visual Encodings

For design inspiration and learning

- Need visual style

Our primary task: Information seeking

- Core enterprise task
- Subject matter of a workbook
- **Do not need** visual style (marks, colors, layout properties, ...)

Data Challenges

Data Challenges

Very limited text

Data Challenges

Very limited text

Additional challenges:

- Multi-sheet workbooks and nested visualizations

Data Challenges

Very limited text

Additional challenges:

- Multi-sheet workbooks and nested visualizations
- Incomplete workbooks

Data Challenges

Very limited text

Additional challenges:

- Multi-sheet workbooks and nested visualizations
- Incomplete workbooks
- Multiple versions

Data Challenges

Very limited text

Additional challenges:

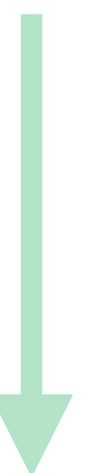
- Multi-sheet workbooks and nested visualizations
- Incomplete workbooks
- Multiple versions
- Out-of-vocabulary words

Extracted text

customer analysis sales profit discount
commission segment ratio ranking count ship
performance target furniture office home
supplies city drilldown late early product
category forecast order quantity target ...

Extracted text

```
customer analysis sales profit discount  
commission segment ratio ranking count ship  
performance target furniture office home  
supplies city drilldown late early product  
category forecast order quantity target ...
```



Transform?

Numeric document representation

0.37546	0.13540	0.01713	0.04225	0.01993	...
---------	---------	---------	---------	---------	-----

Extracted text

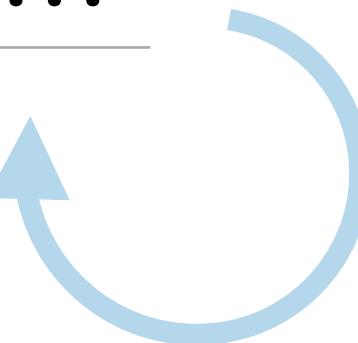
```
customer analysis sales profit discount  
commission segment ratio ranking count ship  
performance target furniture office home  
supplies city drilldown late early product  
category forecast order quantity target ...
```



Transform?

Numeric document representation

0.37546	0.13540	0.01713	0.04225	0.01993	...
---------	---------	---------	---------	---------	-----



Comparisons?

Extracted text

```
customer analysis sales profit discount  
commission segment ratio ranking count ship  
performance target furniture office home  
supplies city drilldown late early product  
category forecast order quantity target ...
```

Transform?

Numeric document representation

0.37546	0.13540	0.01713	0.04225	0.01993	...
---------	---------	---------	---------	---------	-----

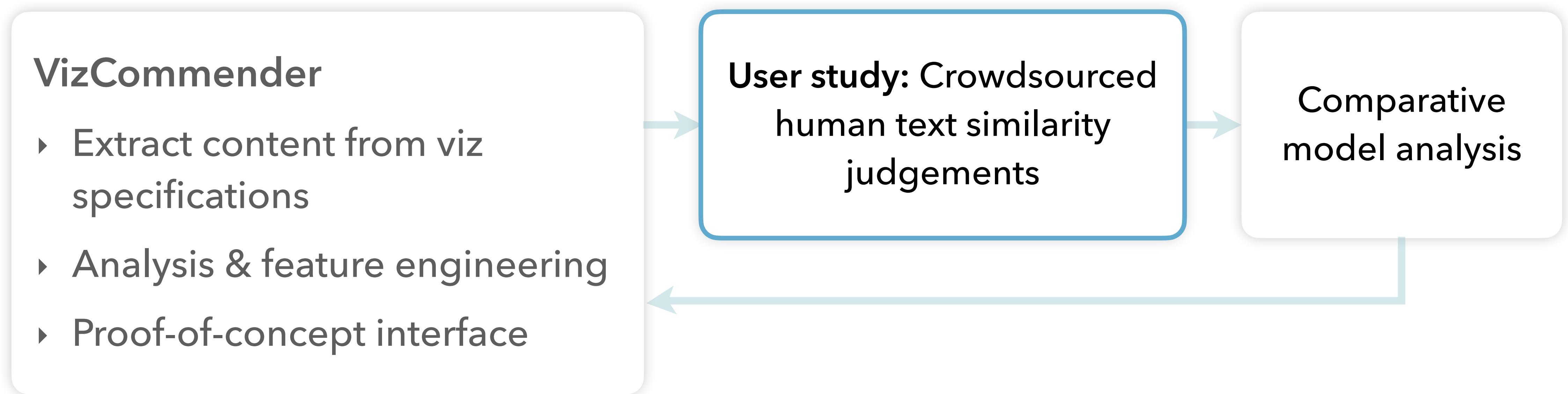
Recommendations

Comparisons?

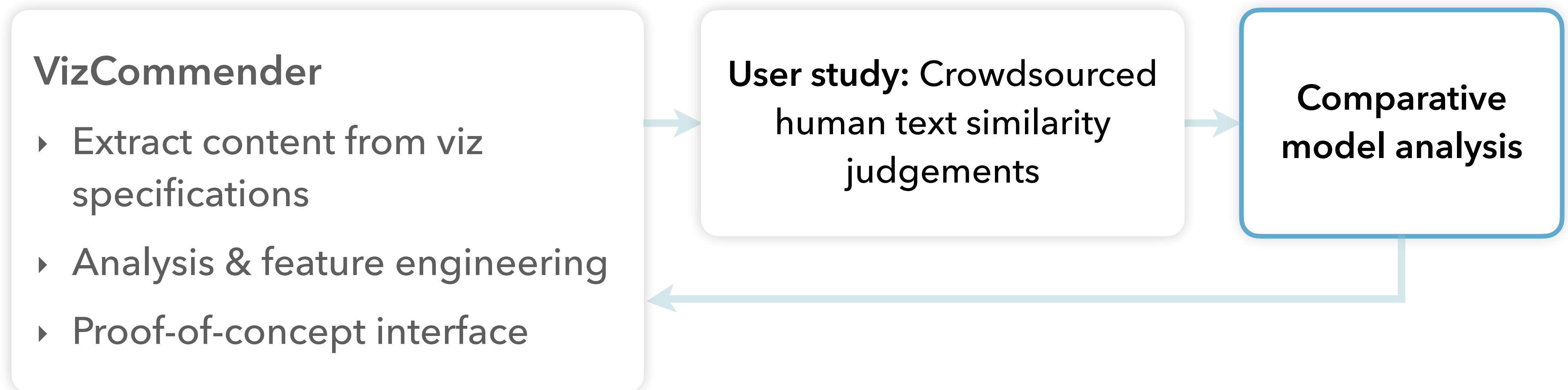
NLP Models

- TF-IDF & cosine similarity
- Latent semantic indexing (LSI) & cosine similarity
- Latent dirichlet allocation (LDA) & Jensen-Shannon divergence
- Word embeddings (Doc2Vec, GloVe) & cosine similarity

Overview



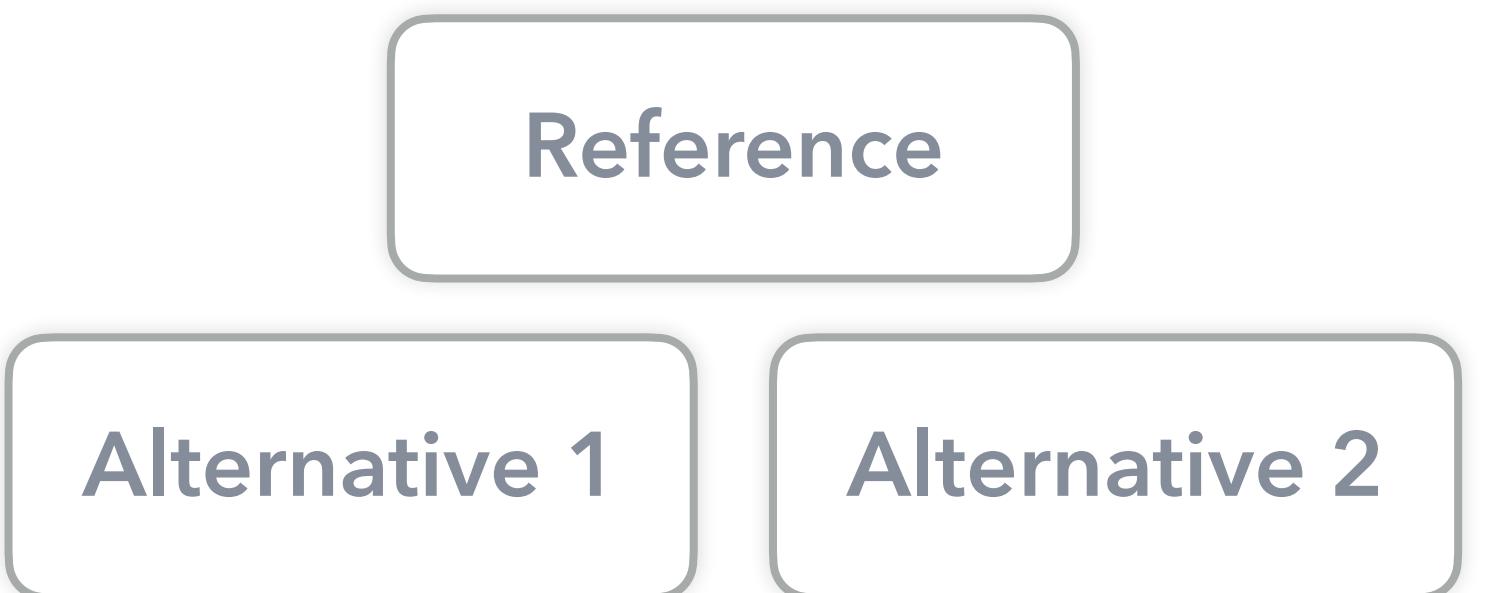
Overview



Crowdsourced human similarity judgements

2-Alternative Forced Choice Experiment

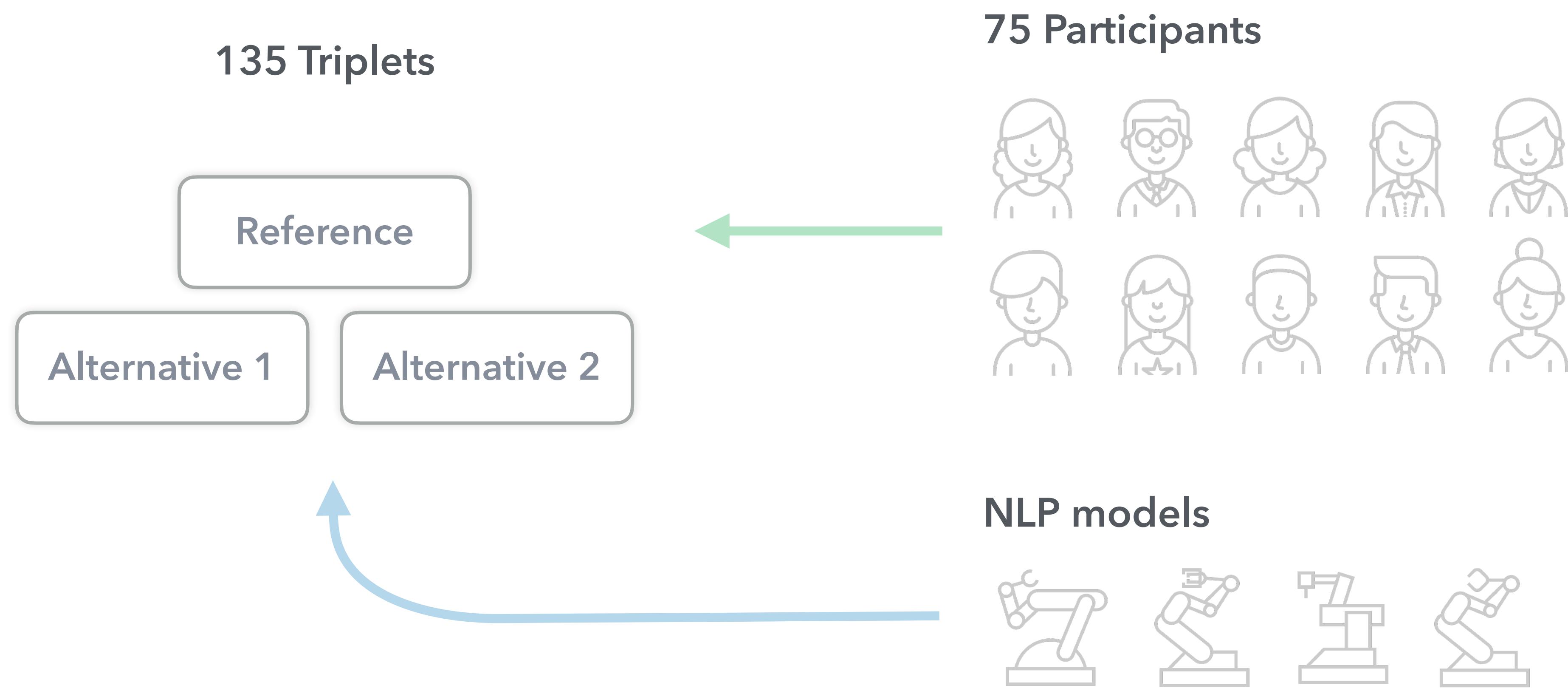
135 Triplets



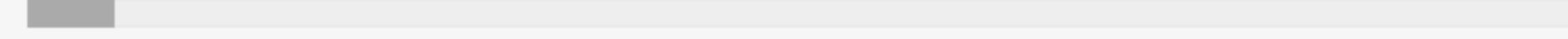
2-Alternative Forced Choice Experiment



2-Alternative Forced Choice Experiment



Experimental Stimulus

Progress 

Continue

Is A or B more similar to the Reference?

Reference

Sheet 1

airplane safety

AVG(Age) Sex Embarked

pclass survived name sex age sibsp parch ticket fare cabin embarked boat body home.dest

A

Flight incidents

Q2

SUM(Survived) Sex

pclass survived name sex age sibsp parch ticket fare cabin embarked boat body
home.dest

B

National Parks

Sheet 7

Park Name AVG(Recreation Visitors)

Park Year Recreation Visitors Non-Recreation Visitors Park Name Park Type State
Park Name (copy)

Experimental Stimulus

Progress 

Continue

Is A or B more similar to the Reference?

Reference

Baseball Story Final

Sheet 9
Height & Batting Average & Handedness

height AVG(avg) handedness

name,handedness,height,weight,avg,HR

A

Final Visualization

Weight

Baseball Player Weight distribution

Weight | Weight (lbs)

name handedness height weight avg HR

B

Olympics 2016 - Which Athlete and Sport are you

Event

SUM(Number of Athletes) Sport

id name nationality sex date_of_birth height weight sport gold silver bronze year of birth
Bronze (copy) Gold (copy) Height (copy) id1 Nationality (copy) Silver (copy) Weight (copy)
year of birth (copy) year of birth1

Agreement Scores

	LDA	TF-IDF	GloVe	Doc2Vec	LSI
TF-IDF	.978				
GloVe	.978	1			
Doc2Vec	.912	.889	.889		
LSI	.935	.956	.956	.848	
Human	.914	.892	.892	.871	.852

Agreement Scores

	LDA	TF-IDF	GloVe	Doc2Vec	LSI
TF-IDF	.978				
GloVe	.978	1			
Doc2Vec	.912	.889	.889		
LSI	.935	.956	.956	.848	
Human	.914	.892	.892	.871	.852

Agreement Scores

	LDA	TF-IDF	GloVe	Doc2Vec	LSI
TF-IDF	.978				
GloVe	.978	1			
Doc2Vec	.912	.889	.889		
LSI	.935	.956	.956	.848	
Human	.914	.892	.892	.871	.852

- ▶ Very good alignment between human similarity judgements and off-the-shelf model predictions
- ▶ LDA performed slightly better

News article

Gov. Gavin Newsom declared a state of emergency Tuesday in response to wildfires in California, as the state gave evacuation orders and battled the effects of a sweltering heat wave, rolling blackouts and the coronavirus pandemic.

By early Wednesday morning, the state fire authorities had ordered residents to evacuate in parts of Santa Cruz, San Mateo, Napa and Sonoma Counties, in Northern California, where thunderstorms brought lightning strikes this week.

The largest fire in the region, called the SCU Lightning Complex, had spread to 35,000 acres in several counties east of San Jose and was 4 percent contained. Another fire, called the LNU Lightning Complex fire, was quickly growing north of the Bay Area, with 32,000 acres burned by about 9:30 Tuesday night.

That fire forced evacuations in parts of Napa and Sonoma, with the authorities warning of an “immediate threat to life” in some places. Local news outlets showed structures consumed by flames in Vacaville, about 35 miles southwest of Sacramento, and fire overtaking a camera meant to help spot wildfires on Mount Vaca. Photos and videos on social media showed flames lapping at the road and, in the hours before dawn, some images showed a glowing red sky, as the fire lit up dense smoke.

To the south, residents in Oakland and San Francisco could smell smoke as they woke up on Wednesday morning. The authorities around Northern California warned of poor air quality in addition to the rising heat ...

News article

Gov. Gavin Newsom declared a state of emergency Tuesday in response to wildfires in California, as the state gave evacuation orders and battled the effects of a sweltering heat wave, rolling blackouts and the coronavirus pandemic.

By early Wednesday morning, the state fire authorities had ordered residents to evacuate in parts of Santa Cruz, San Mateo, Napa and Sonoma Counties, in Northern California, where thunderstorms brought lightning strikes this week.

The largest fire in the region, called the SCU Lightning Complex, had spread to 35,000 acres in several counties east of San Jose and was 4 percent contained. Another fire, called the LNU Lightning Complex fire, was quickly growing north of the Bay Area, with 32,000 acres burned by about 9:30 Tuesday night.

That fire forced evacuations in parts of Napa and Sonoma, with the authorities warning of an “immediate threat to life” in some places. Local news outlets showed structures consumed by flames in Vacaville, about 35 miles southwest of Sacramento, and fire overtaking a camera meant to help spot wildfires on Mount Vaca. Photos and videos on social media showed flames lapping at the road and, in the hours before dawn, some images showed a glowing red sky, as the fire lit up dense smoke.

To the south, residents in Oakland and San Francisco could smell smoke as they woke up on Wednesday morning. The authorities around Northern California warned of poor air quality in addition to the rising heat ...

Extracted text from viz workbook

**customer analysis sales profit discount commission segment ratio ranking
count ship performance target furniture office home supplies city drilldown
late early product category forecast order quantity ...**

Proof-of-concept implementation



park

national

visit

fewer

visitor

makeover

montime

country

zip

com

bar chart

area chart

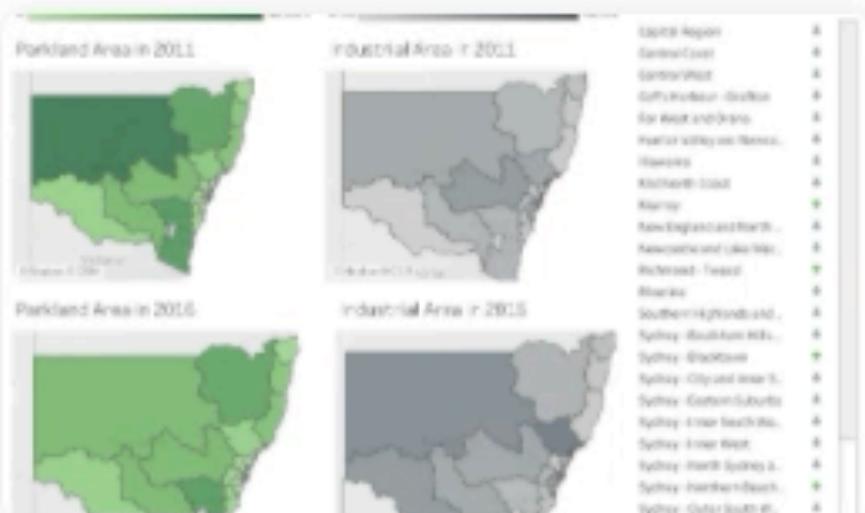
line chart

bubble chart

table

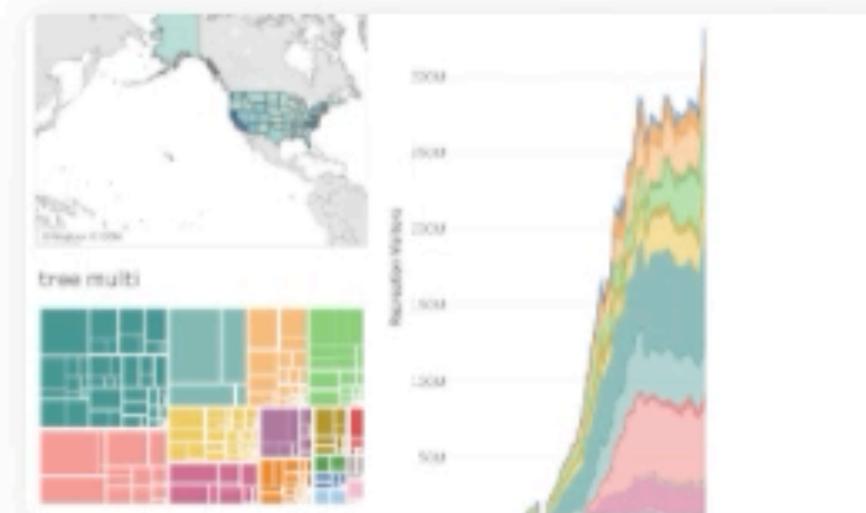


Sort by relevance ▾



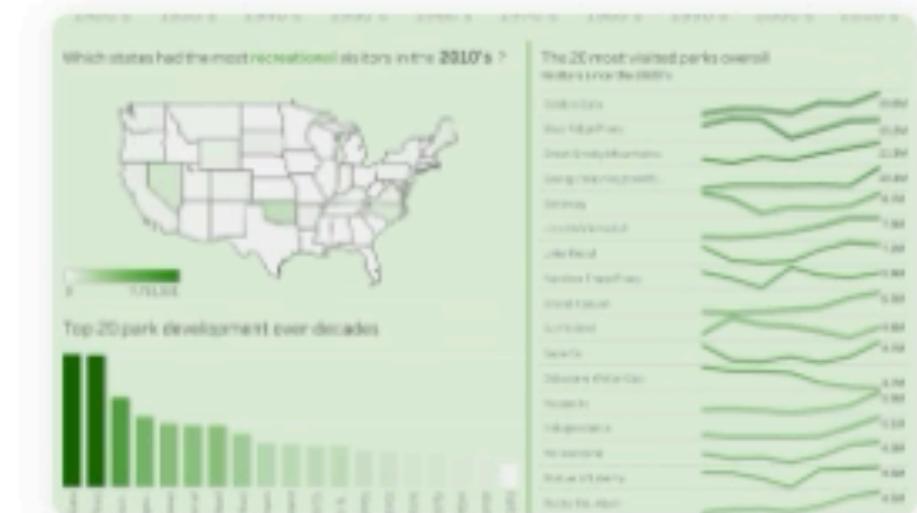
landuserv3

bing4493 • 2017-05-08



Makeover Monday Week 23 -

mak.gill • 2017-06-05



Makeover Monday Week 23

frans.rasmussen7740 • 2017-06-08

CA Match Expressions

Product Name	Product ID	Day of Express
Go Math, Middle School, Grade 6 © 2014	54406710	February 20, 2020
Go Math, Middle School, Grade 7 © 2014	544083050	February 20, 2020
Go Math, Middle School, Grade 7 © 2014	544063116	February 20, 2020
Go Math, Middle School, Grade 8 © 2014	544083075	February 20, 2020
Go Math, Middle School, Grade 8 © 2014	544083014	February 20, 2020
Go Math, NJ Middle School, Grade 6 © 2018	5790320061010	October 11, 2018

Asbury Park MyHRW Products

Product Name

Product Name	Product ID	Day of Express
Go Math, Middle School, Grade 6 © 2014	54406710	February 20, 2020
Go Math, Middle School, Grade 7 © 2014	544083050	February 20, 2020
Go Math, Middle School, Grade 7 © 2014	544063116	February 20, 2020
Go Math, Middle School, Grade 8 © 2014	544083075	February 20, 2020
Go Math, Middle School, Grade 8 © 2014	544083014	February 20, 2020
Go Math, NJ Middle School, Grade 6 © 2018	5790320061010	October 11, 2018

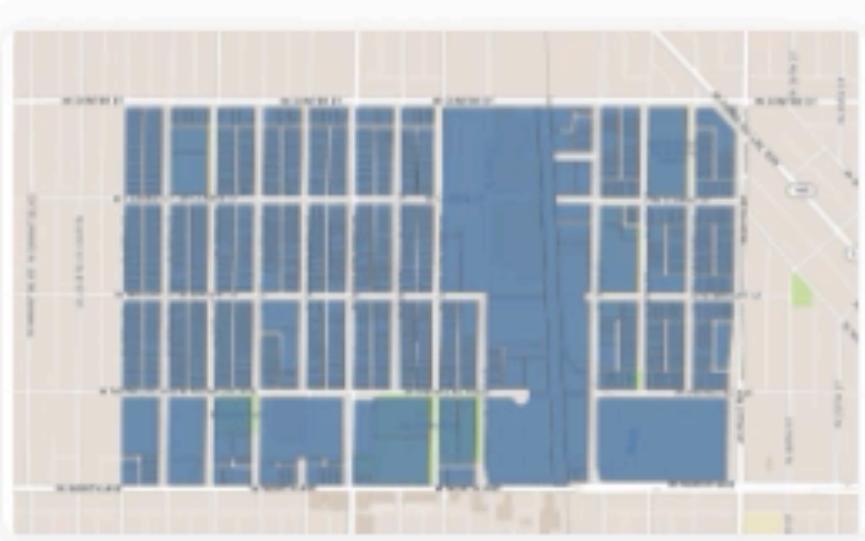
ASBURY Dashboard

brian.jones3491 • 2018-02-09



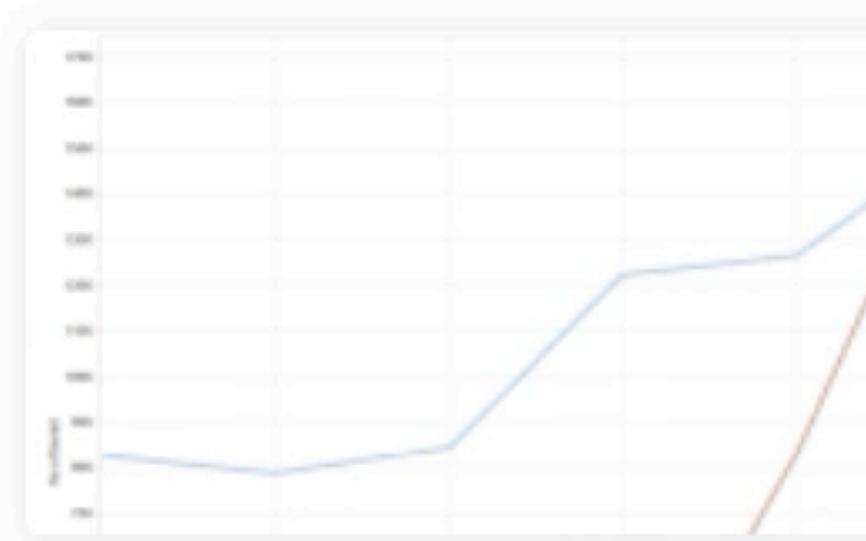
Makeover Monday Week 23

esben.michelsen • 2017-06-05



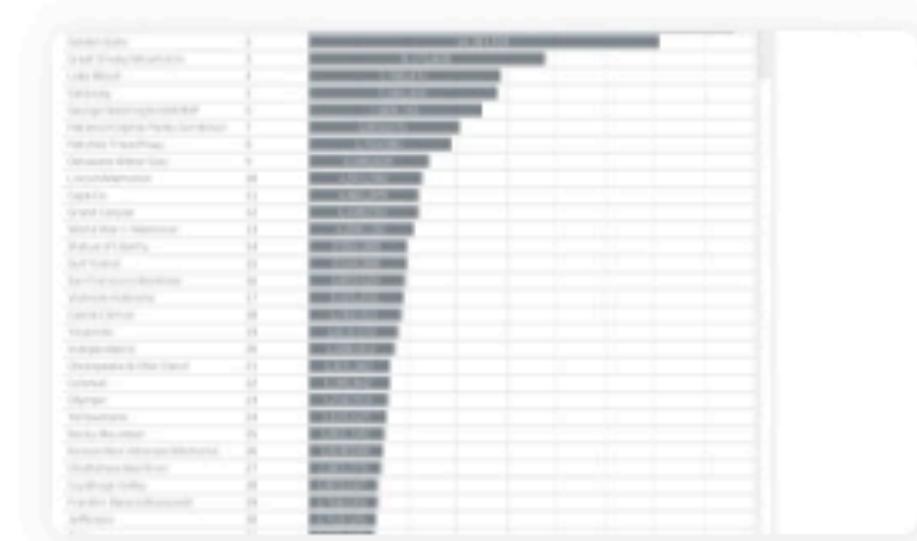
Metcalfe Park

safe.sound.mke • 2017-12-15



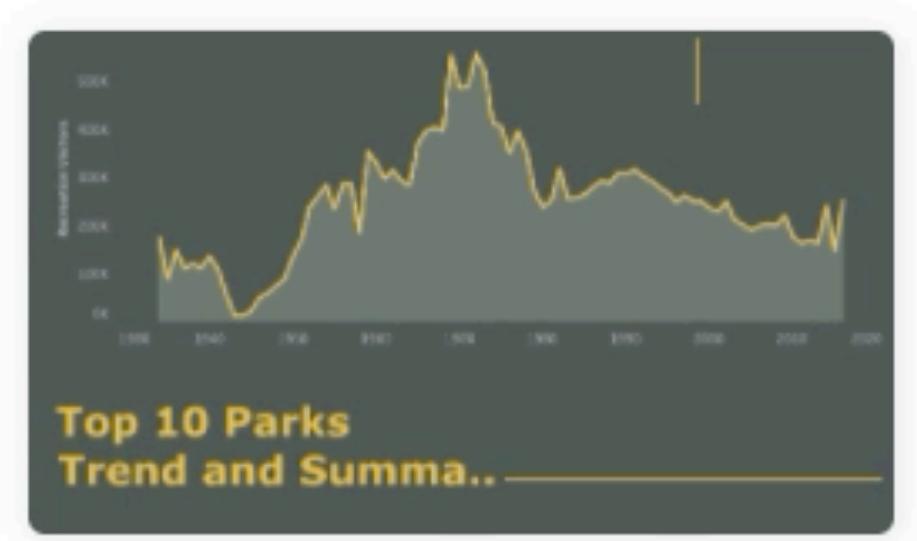
Grant's Nepalese Tourism

hollidaygg • 2015-03-08



Makeover Monday week 23

priyesh.singh • 2017-06-05



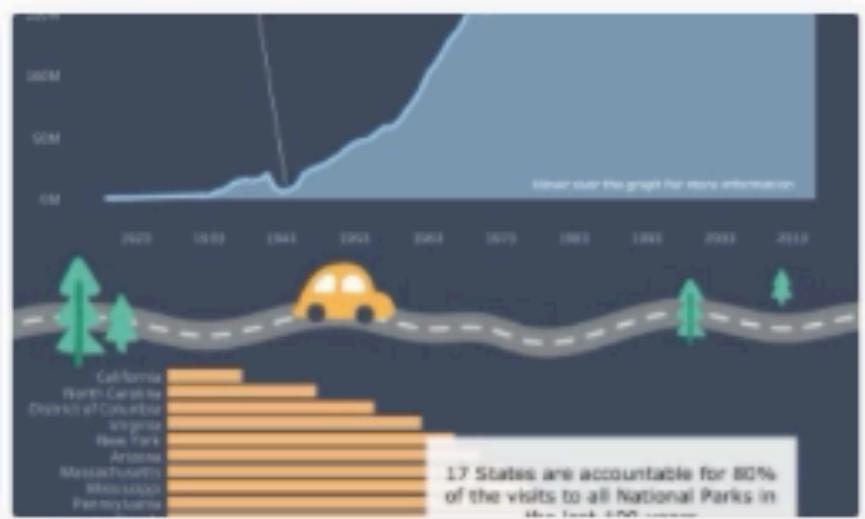
National Parks

poojagandhi • 2017-06-05



Top National Parks in Ame

sarah.bartlett • 2017-06-05



The National Parks Have N

pablolgomez • 2017-06-05



MakeOver Monday TOTC 2017

nick.bignell • 2017-06-05



Grant's Nepalese Tourism3

hollidaygg • 2015-03-08



MM Live

nai.louza • 2017-06-05



Windermere

lester.nare • 2018-06-08

Are people visiting the **national parks** of America?

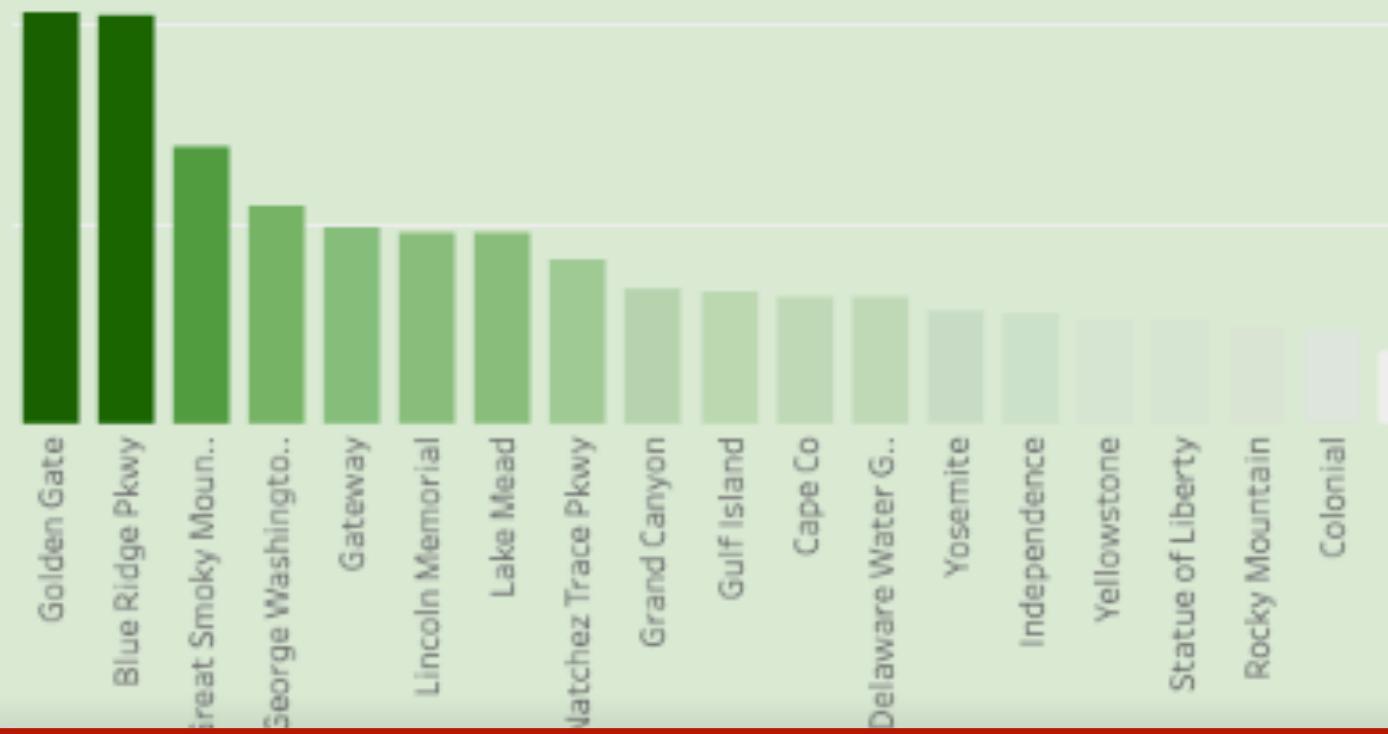
Click on a decade to filter

1920's 1930's 1940's 1950's 1960's 1970's 1980's 1990's 2000's 2010's

Which states had the most **recreational** visitors in the **2010's** ?



Top 20 park development over decades



The 20 most visited parks overall

Visitors since the 1920's

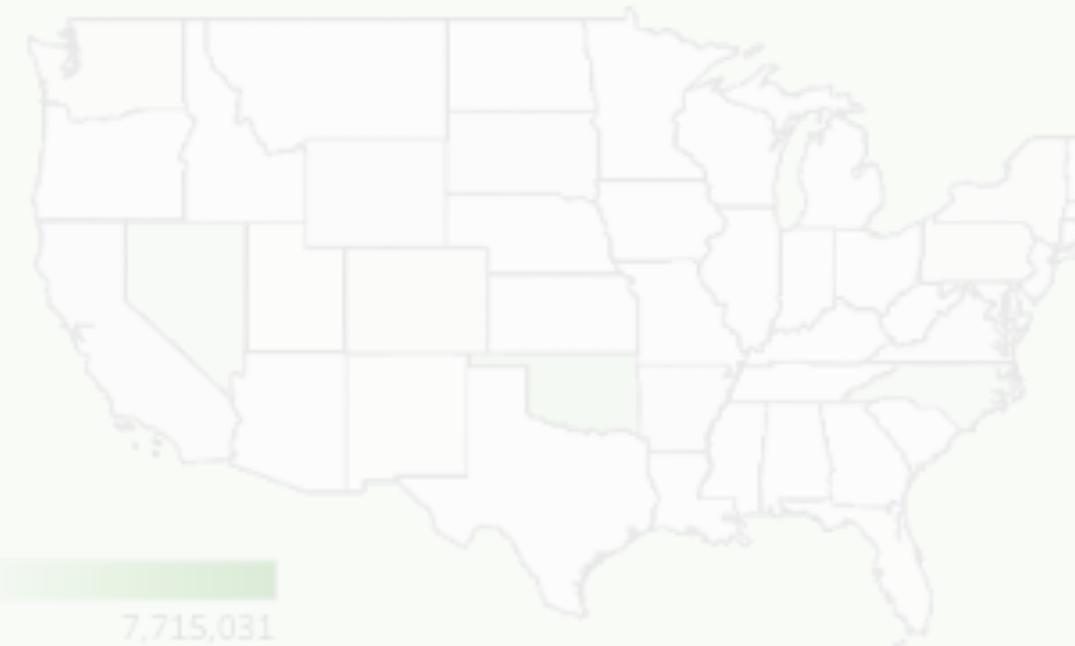


Interactive workbook

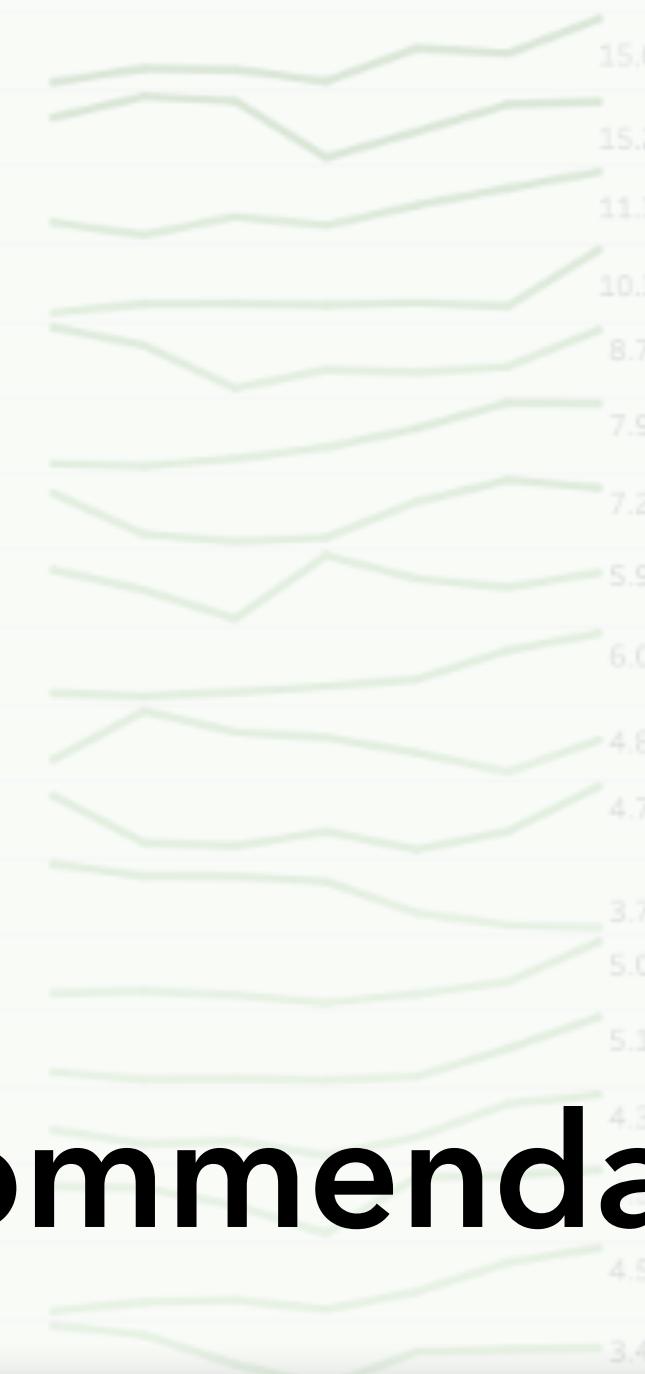
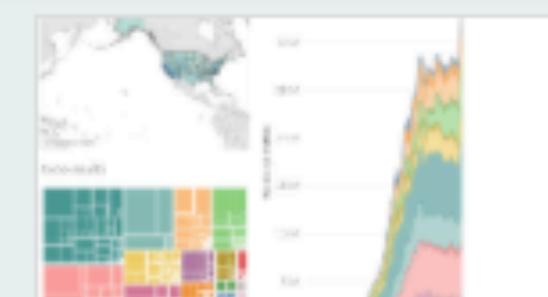
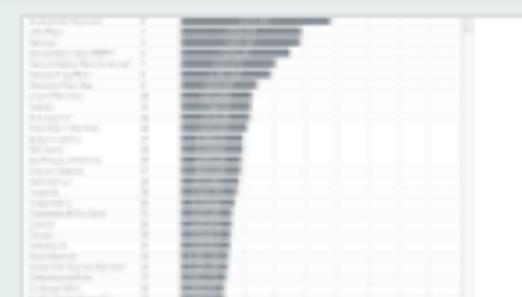
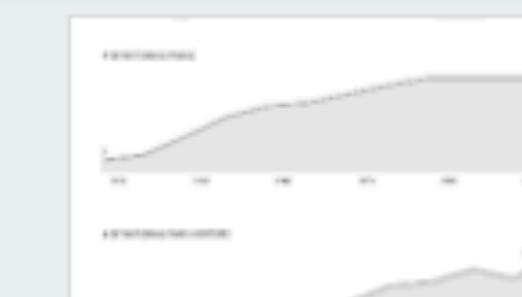
Are people visiting the **national parks** of America?

Click on a decade to filter

1920's 1930's 1940's 1950's 1960's 1970's 1980's 1990's 2000's 2010's

Which states had the most **recreational** visitors in the 2010's ?

Top 20 park development over decades

The 20 most visited parks overall
Visitors since the 1920's**Recommendation panel**Makeover Monday Week 23
- US Parks
mak.gillMakeover Monday week 23
priyesh.singhNational Park Growth
Trends
kruppNational Parks
Trend and Summa..
poojagandhiMakeover Monday Week 23
- The Popularity of US
National Parks
umar.hassanClassification of Ozone
Levels1
tihanaMakeOver Monday TOTC
2017
nick.bignellNick Twitter Report
nick.strohecker

Are people visiting the **national parks** of America?

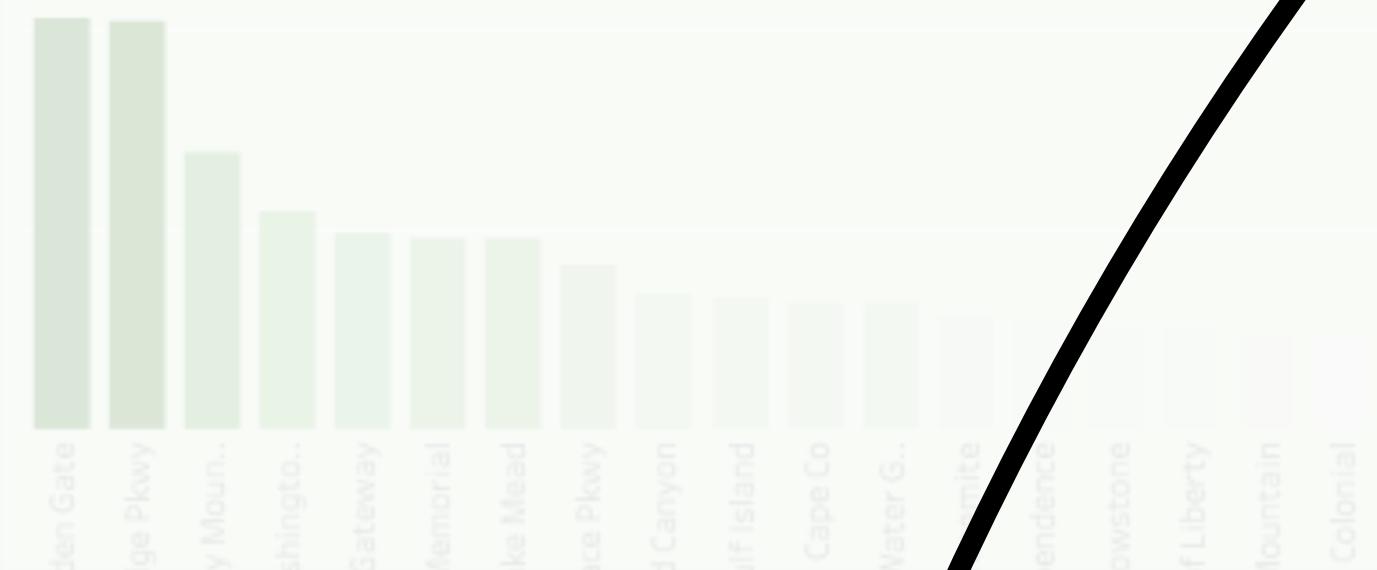
Click on a decade to filter

1920's 1930's 1940's 1950's 1960's 1970's 1980's 1990's 2000's 2010's

Which states had the most **recreational visitors** in the 2010's ?



Top 20 park development over decades



Makeover Monday Week 23
- US Parks
mak.gill



Makeover Monday week 23
priyeshsingh



National Park Growth
Trends
krupp



National Parks
poojagandhi



Makeover Monday Week 23
- The Popularity of US
National Parks
umar.hassan



Classification of Ozone
Levels1
tihana



MakeOver Monday TOTC
2017
nick.bignell

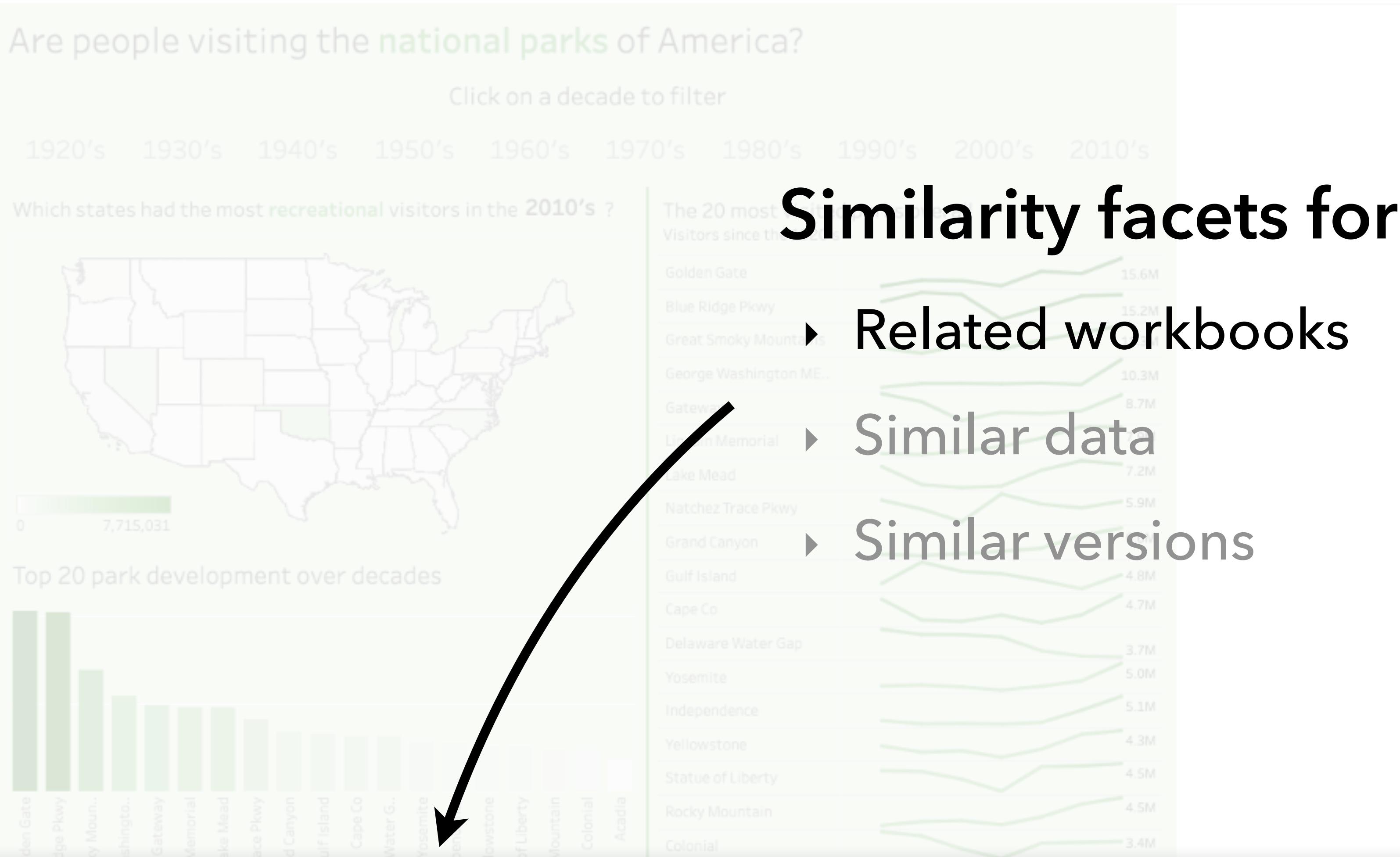


Nick Twitter Report
nick.strohecker

Similarity facets for different user tasks

- ▶ Related workbooks
- ▶ Similar data
- ▶ Similar versions





Similarity facets for different user tasks

- ▶ Related workbooks
 - ▶ Similar data
 - ▶ Similar versions



A treemap visualization showing the distribution of data across various categories. The categories are represented by colored rectangles of different sizes. The largest category, colored teal, occupies the top-left portion of the map. Other significant categories include a large grey area representing 'Other' or 'Unknown' data, a yellow area, and several smaller green, orange, red, and purple areas.

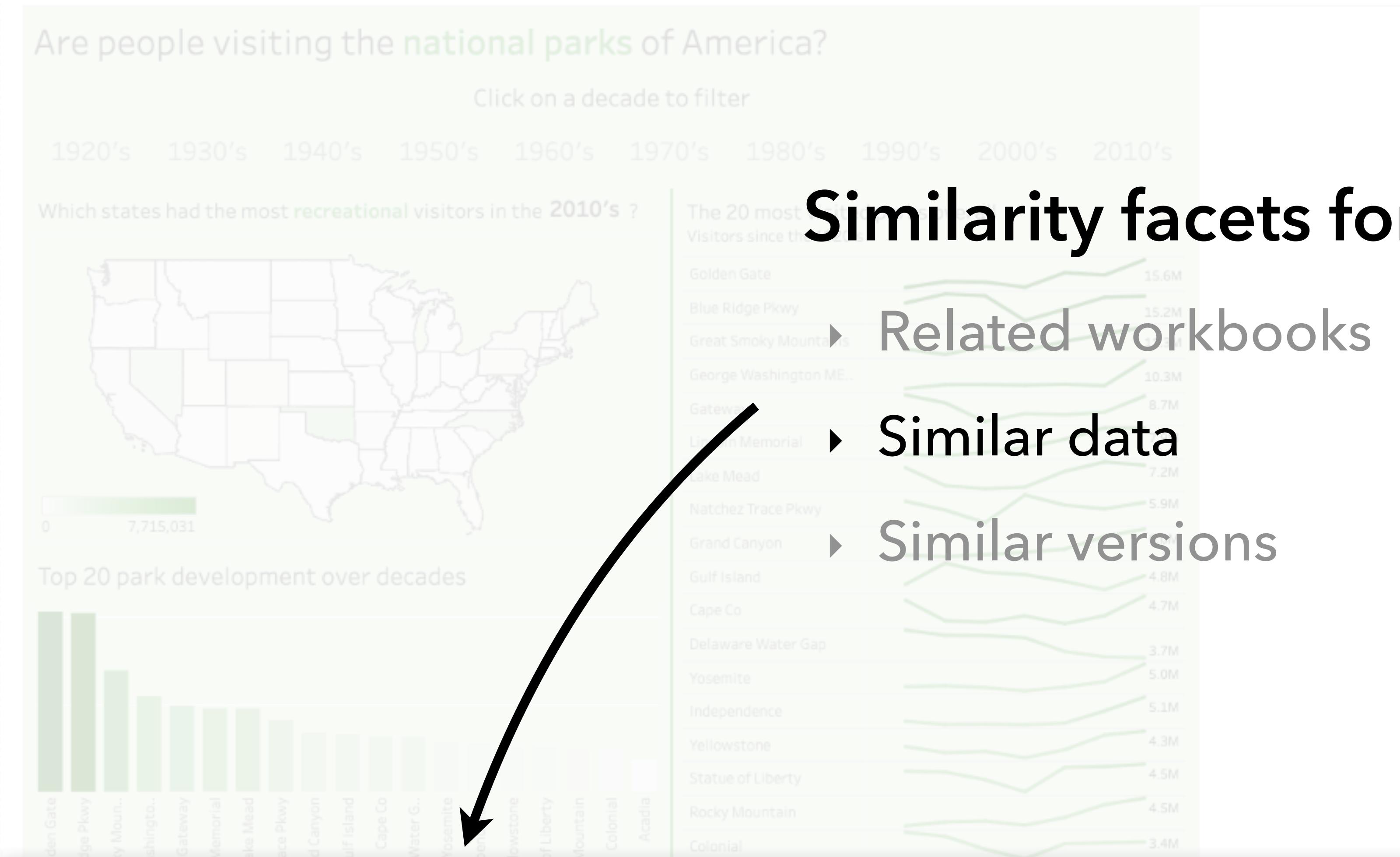
Term	Percentage (%)
Climate Change	95
Global Warming	85
Green Energy	78
Sustainable Development	75
Renewable Energy	72
Carbon Footprint	68
Eco-Friendly	65
Green Technology	62
Conservation	58
Organic	55
Recycling	52
Green Building	48
Energy Efficiency	45
Green Products	42
Green Living	38
Green Jobs	35
Green Politics	32
Green Labels	28
Green Labels	25
Green Labels	22
Green Labels	18
Green Labels	15
Green Labels	12
Green Labels	10

The graph displays a fluctuating line representing the performance of the top 10 parks over a decade. The y-axis is unlabeled with numerical ticks at 0, 10, 20, 30, and 40. The x-axis shows years from 2001 to 2011. The line starts around 20 in 2001, dips slightly, then rises steadily to about 35 by 2006. It fluctuates between 35 and 40 until 2009, where it peaks at approximately 42. After 2009, the line drops sharply to around 25 by 2011.

The figure consists of three separate heatmaps arranged horizontally. Each heatmap has 'Sheet 1' at the top, followed by a color-coded scale from light blue to dark red, and then 'Sheet 2' at the bottom. The first heatmap shows a light orange gradient. The second heatmap shows a dark blue gradient. The third heatmap shows a dark red gradient.

The dashboard displays a variety of data across several panels:

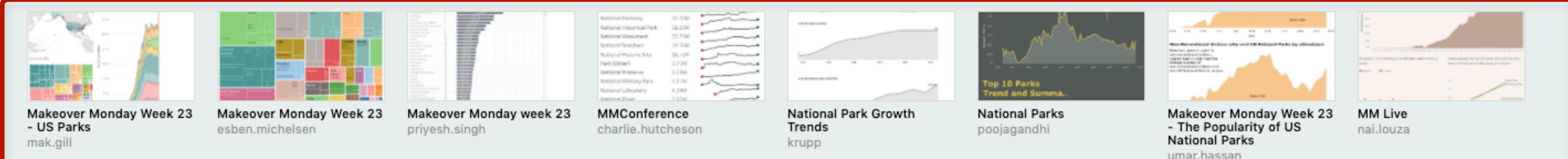
- Panel A:** A horizontal bar chart comparing two categories. The blue bar represents 'Type A' with a value of 200.00, and the orange bar represents 'Type B' with a value of 100.
- Panel B:** A vertical bar chart showing data for 'Category A' across four sub-categories. The values are approximately 100, 100, 100, and 100 respectively.
- Panel C:** A heatmap representing data for 'Category B' across four sub-categories. The colors range from light green (low values) to dark red (high values), with most cells showing high values (around 100).
- Panel D:** A detailed table with multiple rows and columns. The columns are labeled 'Category A', 'Category B', 'Category C', 'Category D', and 'Category E'. The rows are labeled 'Row 1' through 'Row 10'. The data values are mostly high, with some variations between rows.



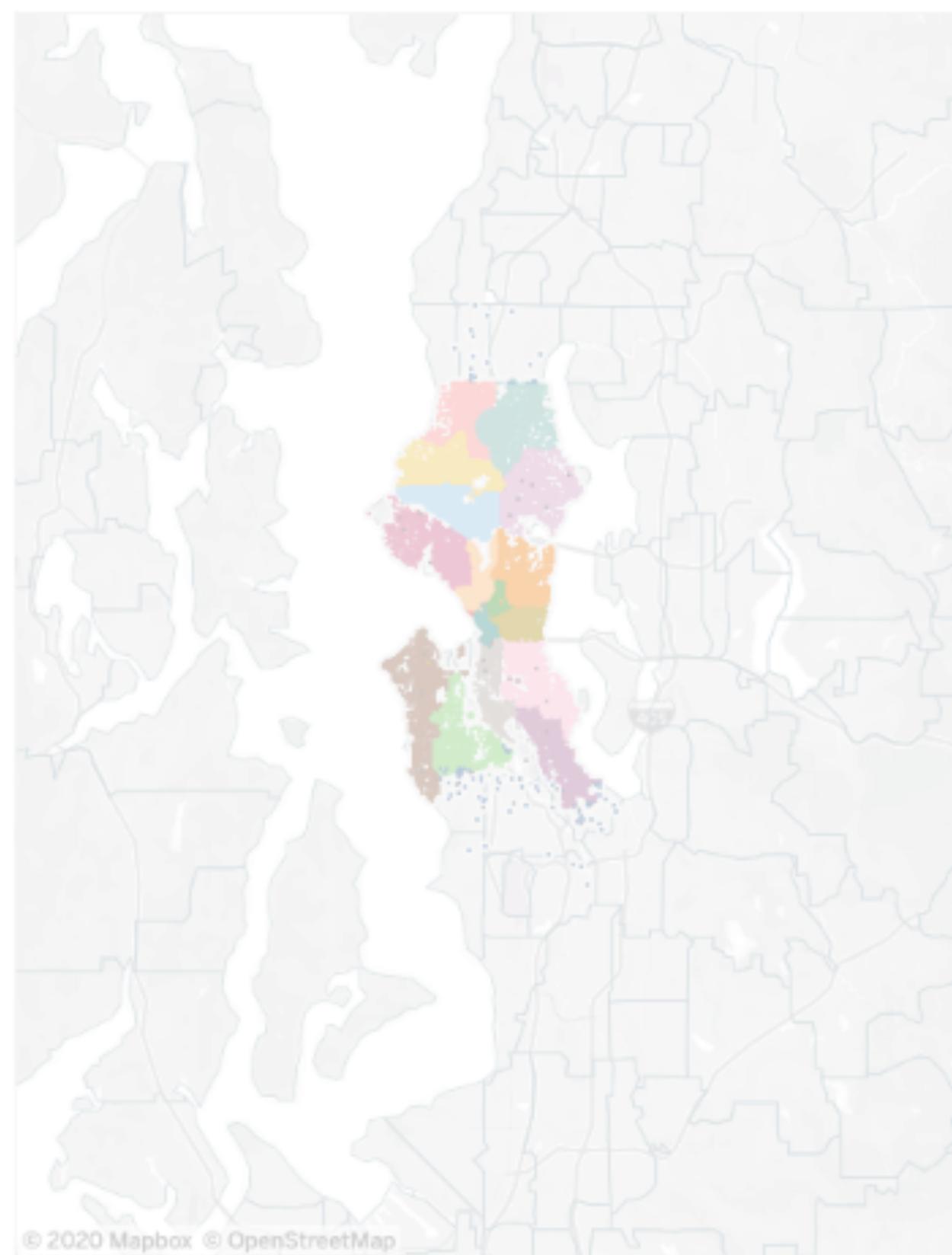
Similarity facets for different user tasks

Related workbooks

- ▶ Similar data
 - ▶ Similar versions



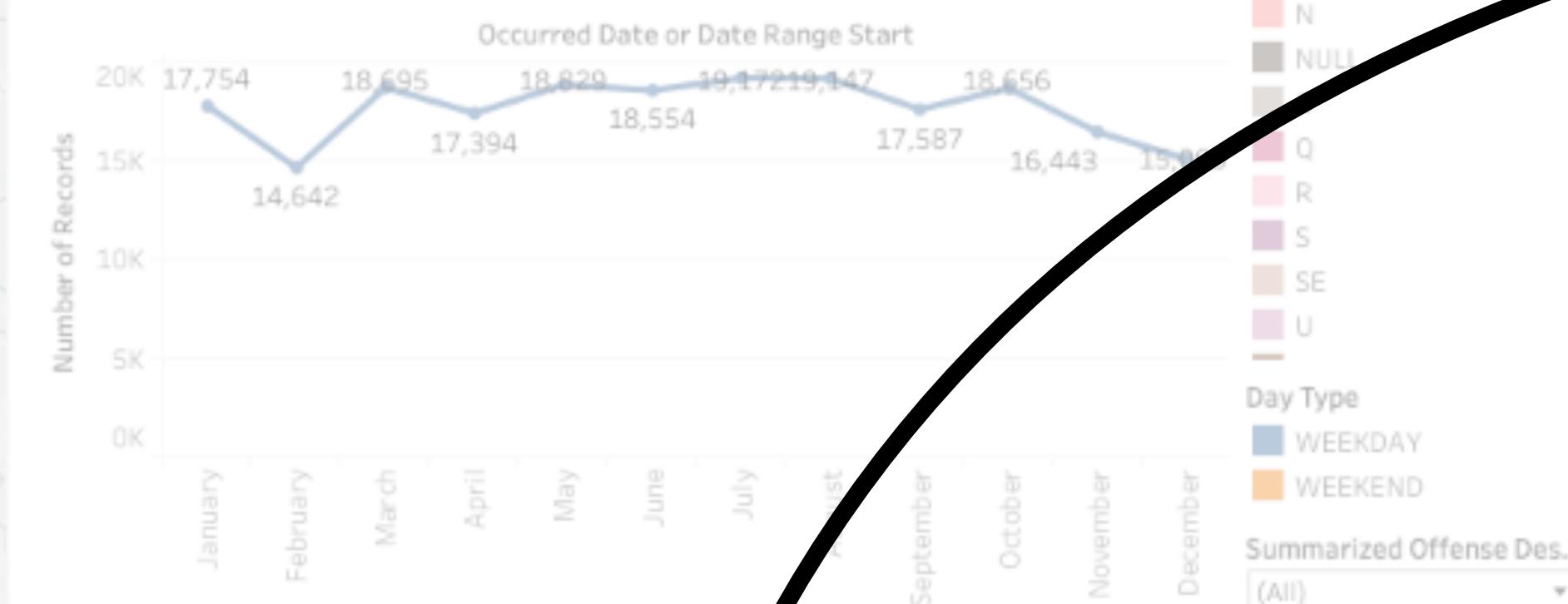
Where is crime happening



When is crime happening? (hourly)

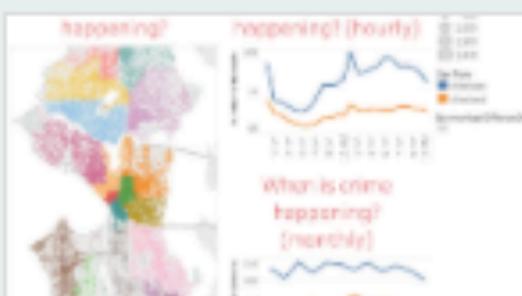


When is crime happening? (monthly)

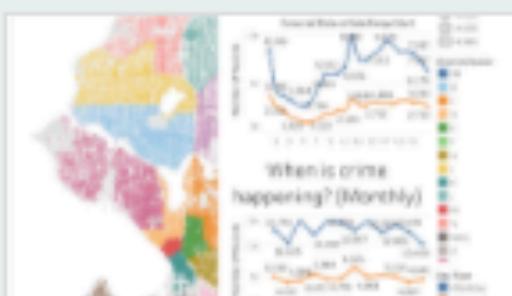


Similarity facets for different user tasks

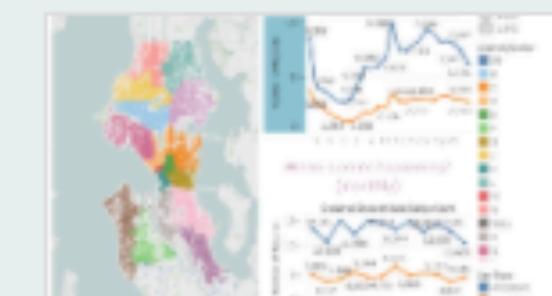
- ▶ Related workbooks
- ▶ Similar data
- ▶ Similar versions



Seattle Crime
caitlin.streamer



Seattle Crime Dashboard
kelsey.hofmann



Crime in Seattle
danya.setiawan

Generalizable to other viz specifications



Generalizable to other viz specifications



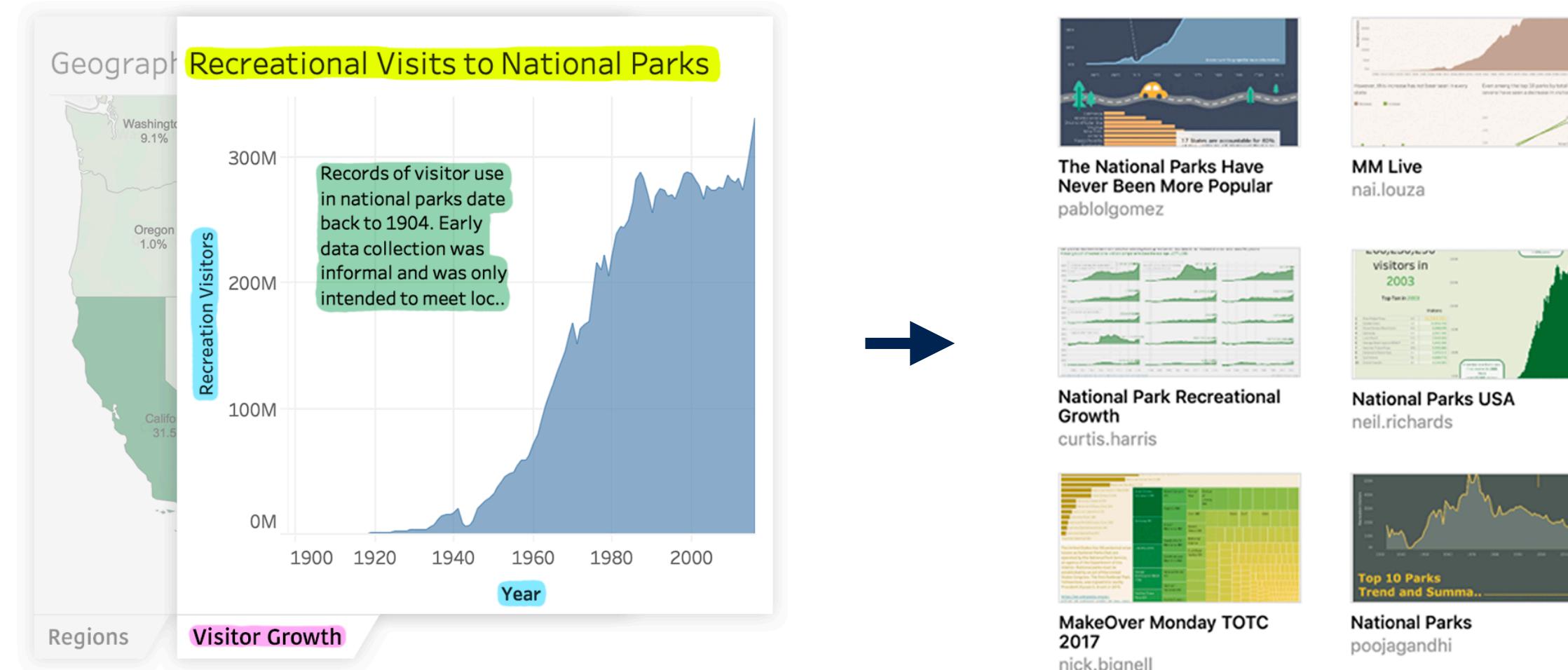
...

VizCommender

Computing Text-Based Similarity in Visualization Repositories for Content-Based Recommendations

Michael Oppermann, Robert Kincaid, and Tamara Munzner

► michaeloppermann.com/work/viz-commender



Contributions

- ▶ Challenges for content-based visualization recommendations
- ▶ Design and implementation of a proof-of-concept pipeline
- ▶ Analysis of applicable NLP techniques and a user study assessing the alignment with human judgements of similarity



THE UNIVERSITY
OF BRITISH COLUMBIA



Interactive Visual Analysis Tool

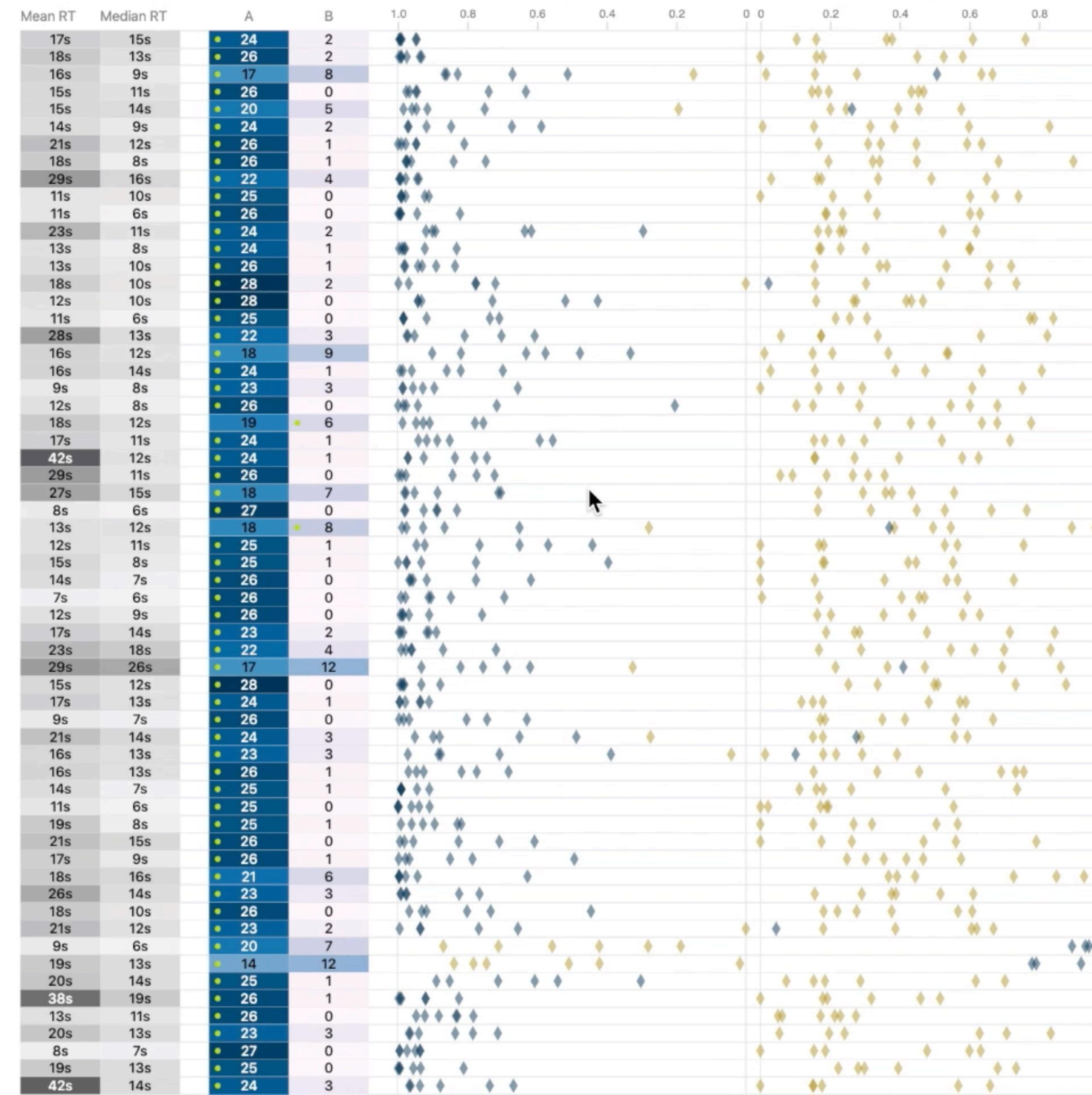
Participants (75)												Filter triplets	Disagree with majority	Temporal alignment	
ID	TIMING	TOTAL RT	AVG. RT	MEDIAN RT	MIN. RT	MAX. RT	POS A	POS B	INTRA-RATER AGREEMENT	INTER-RATER AGREEMENT	EXPERT AGREEMENT	DIFFICULTY	BOTH RELEVANT	NONE RELEVANT	DEMOGRAPHICS
187		17min (19)	22s	16s	4s	2min	21	26	100%	96%	97.87%	3/5	2/5	2/5	54, female, Training
184		12min (15)	16s	13s	7s	37s	21	26	100%	91%	93.62%	2/5	2/5	2/5	57, female, Training
183		10min (10)	12s	7s	2s	2min	22	25	50%	85%	82.98%	4/5	4/5	2/5	30, male, College, no degree
182		17min (21)	22s	18s	8s	60s	27	20	100%	91%	89.36%	3/5	4/5	2/5	34, female, Bachelor's degree
181		5min (6)	6s	5s	3s	32s	24	23	100%	83%	85.11%	2/5	4/5	3/5	40, male, Bachelor's degree
180		7min (10)	9s	7s	3s	34s	21	26	100%	98%	95.74%	1/5	3/5	1/5	26, female, Bachelor's degree
179		7min (11)	9s	8s	4s	24s	22	25	50%	87%	85.11%	5/5	3/5	3/5	53, male, Bachelor's degree
176		6min (8)	8s	7s	3s	25s	24	23	50%	91%	89.36%	2/5	2/5	1/5	31, male, College, no degree
175		7min (9)	10s	6s	3s	32s	26	21	100%	98%	95.74%	2/5	3/5	2/5	32, female, Bachelor's degree
174		10min (13)	13s	10s	5s	37s	25	22	100%	96%	93.62%	1/5	2/5	1/5	35, female, Bachelor's degree
173		11min (13)	15s	12s	4s	55s	20	27	100%	98%	95.74%	5/5	2/5	1/5	28, male, College, no degree
172		8min (10)	11s	8s	3s	1min	17	30	50%	87%	85.11%	2/5	5/5	1/5	30, male, College, no degree
171		6min (7)	8s	7s	3s	29s	27	20	100%	98%	95.74%	2/5	1/5	2/5	23, female, Bachelor's degree
170		9min (11)	11s	9s	3s	31s	24	23	100%	91%	89.36%	3/5	4/5	2/5	55, female, Bachelor's degree
169		7min (9)	10s	9s	4s	20s	22	25	100%	94%	91.49%	2/5	2/5	2/5	47, female, Bachelor's degree
168		14min (16)	18s	7s	3s	3min	27	20	100%	74%	76.6%	4/5	2/5	5/5	39, male, Bachelor's degree
167		13min (17)	16s	14s	4s	1min	22	25	100%	87%	89.36%	2/5	3/5	2/5	40, female, Bachelor's degree
166		20min (28)	26s	12s	4s	3min	23	24	100%	81%	78.72%	2/5	3/5	2/5	40, male, Associate degree
165		14min (17)	18s	7s	3s	4min	22	25	100%	91%	89.36%	2/5	2/5	2/5	30, female, Bachelor's degree
164		4min (7)	5s	5s	2s	10s	25	22	100%	89%	91.49%	1/5	2/5	2/5	36, male, Bachelor's degree
163		14min (15)	18s	15s	4s	44s	22	25	50%	91%	89.36%	4/5	2/5	3/5	36, male, College, no degree
161		7min (9)	9s	8s	4s	20s	28	19	50%	98%	97.87%	3/5	2/5	2/5	37, male, Bachelor's degree
160		30min (32)	38s	35s	13s	1min	29	18	50%	91%	89.36%	5/5	5/5	2/5	69, female, College, no degree
159		8min (9)	10s	7s	3s	37s	24	23	0%	87%	85.11%	2/5	2/5	1/5	33, female, Bachelor's degree
158		7min (9)	9s	8s	4s	27s	24	23	100%	85%	87.23%	1/5	2/5	2/5	40, female, High school
157		18min (20)	23s	17s	6s	2min	23	24	100%	94%	95.74%	3/5	2/5	3/5	39, female, Bachelor's degree
155		10min	12s	9s	4s	1min	29	18	100%	70%	76.6%	2/5	1/5	1/5	41, male, College, no

Triplets

 tfidf lda_150 lsi_150 glove_pre glove_tf d2v_new

Filter triplets

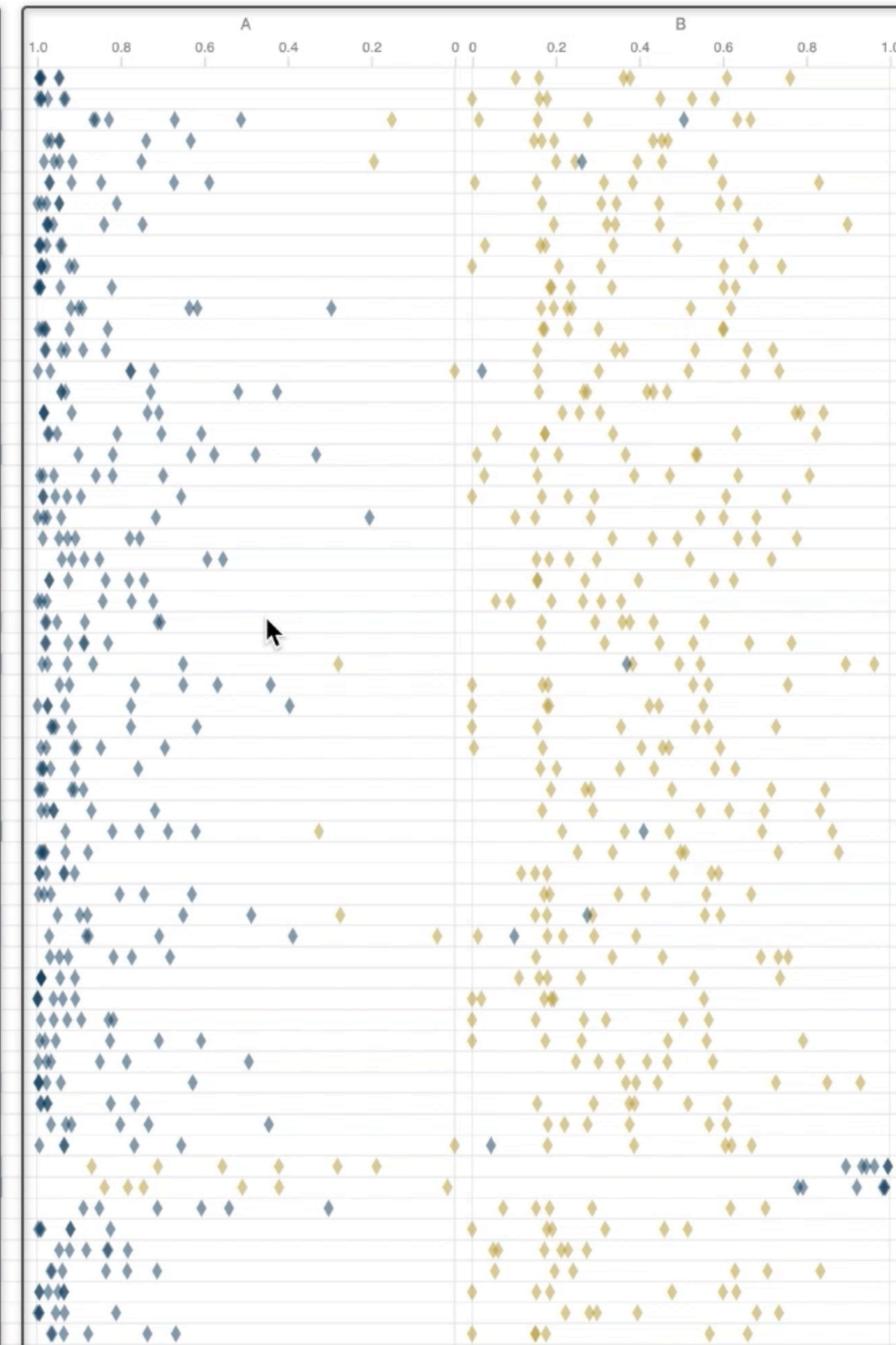
Settings



Human Judgements**Model Predictions****Triplets**

Mean RT Median RT

A	B
24	2
26	2
17	8
26	0
20	5
24	2
26	1
26	1
22	4
25	0
26	0
24	2
24	1
26	1
28	2
28	0
25	0
22	3
18	9
24	1
23	3
26	0
19	6
24	1
24	1
26	0
42s	12s
29s	11s
27s	15s
8s	6s
23s	12s
13s	11s
15s	8s
14s	7s
7s	6s
12s	9s
17s	14s
23s	18s
29s	26s
15s	12s
17s	13s
9s	7s
21s	14s
16s	13s
16s	13s
14s	7s
11s	6s
19s	8s
21s	15s
17s	9s
18s	16s
26s	14s
18s	10s
21s	12s
9s	6s
19s	13s
20s	14s
38s	19s
13s	11s
20s	13s
8s	7s
19s	13s
42s	14s


 tfidf lda_150 lsi_150 glove_pre glove_tf d2v_new

Filter triplets

Settings

Triplets

 tfidf lda_150 lsi_150 glove_pre glove_tf d2v_new

Filter triplets

Settings

Mean RT Median RT

		A	B
17s	15s	● 24	2
18s	13s	● 26	2
16s	9s	● 17	8
15s	11s	● 26	0
15s	14s	● 20	5
14s	9s	● 24	2
21s	12s	● 26	1
18s	8s	● 26	1
29s	16s	● 22	4
11s	10s	● 25	0
11s	6s	● 26	0
23s	11s	● 24	2
13s	8s	● 24	1
13s	10s	● 26	1
18s	10s	● 28	2
12s	10s	● 28	0
11s	6s	● 25	0
28s	13s	● 22	3
16s	12s	● 18	9
16s	14s	● 24	1
9s	8s	● 23	3
12s	8s	● 26	0
18s	12s	● 19	6
17s	11s	● 24	1
42s	12s	● 24	1
29s	11s	● 26	0
27s	15s	● 18	7
8s	6s	● 27	0
13s	12s	● 18	8
12s	11s	● 25	1
15s	8s	● 25	1
14s	7s	● 26	0
7s	6s	● 26	0
12s	9s	● 26	0
17s	14s	● 23	2
23s	18s	● 22	4
29s	26s	● 17	12
15s	12s	● 28	0
17s	13s	● 24	1
9s	7s	● 26	0
21s	14s	● 24	3
16s	13s	● 23	3
16s	13s	● 26	1
14s	7s	● 25	1
11s	6s	● 25	0
19s	8s	● 25	1
21s	15s	● 26	0
17s	9s	● 26	1
18s	16s	● 21	6
26s	14s	● 23	3
18s	10s	● 26	0
21s	12s	● 23	2
9s	6s	● 20	7
19s	13s	● 14	12
20s	14s	● 25	1
38s	19s	● 26	1
13s	11s	● 26	0
20s	13s	● 23	3
8s	7s	● 27	0
19s	13s	● 25	0
42s	14s	● 24	3

Rising House Prices, incomparable to inflated Salary 1990-2015

Sheet 1
Percentage of 20-24 year old population enrolled in tertiary education, 1990-2015.
SUM(% of 20-24 Population in Tertiary Education.) | % (as a decimal)
Year

Year % of 20-24 Population in Tertiary Education. Men Median Marriage Age
 Women Median Marriage Age Price % Change \$ Change Sum of annual income Year1
 Year2 Year3

Sothebystest2.1.1.1

Incline Village1
AVG(Average Days on Market) Area YEAR(Date)

Date Area Type City Zip code State Median Sales Price Average List Price
 Average Sold Price Average Days on Market Highest Sold Price Lowest Sold Price
 Properties Sold % Sold Price to Average List Price Total Dollar Volume Sold

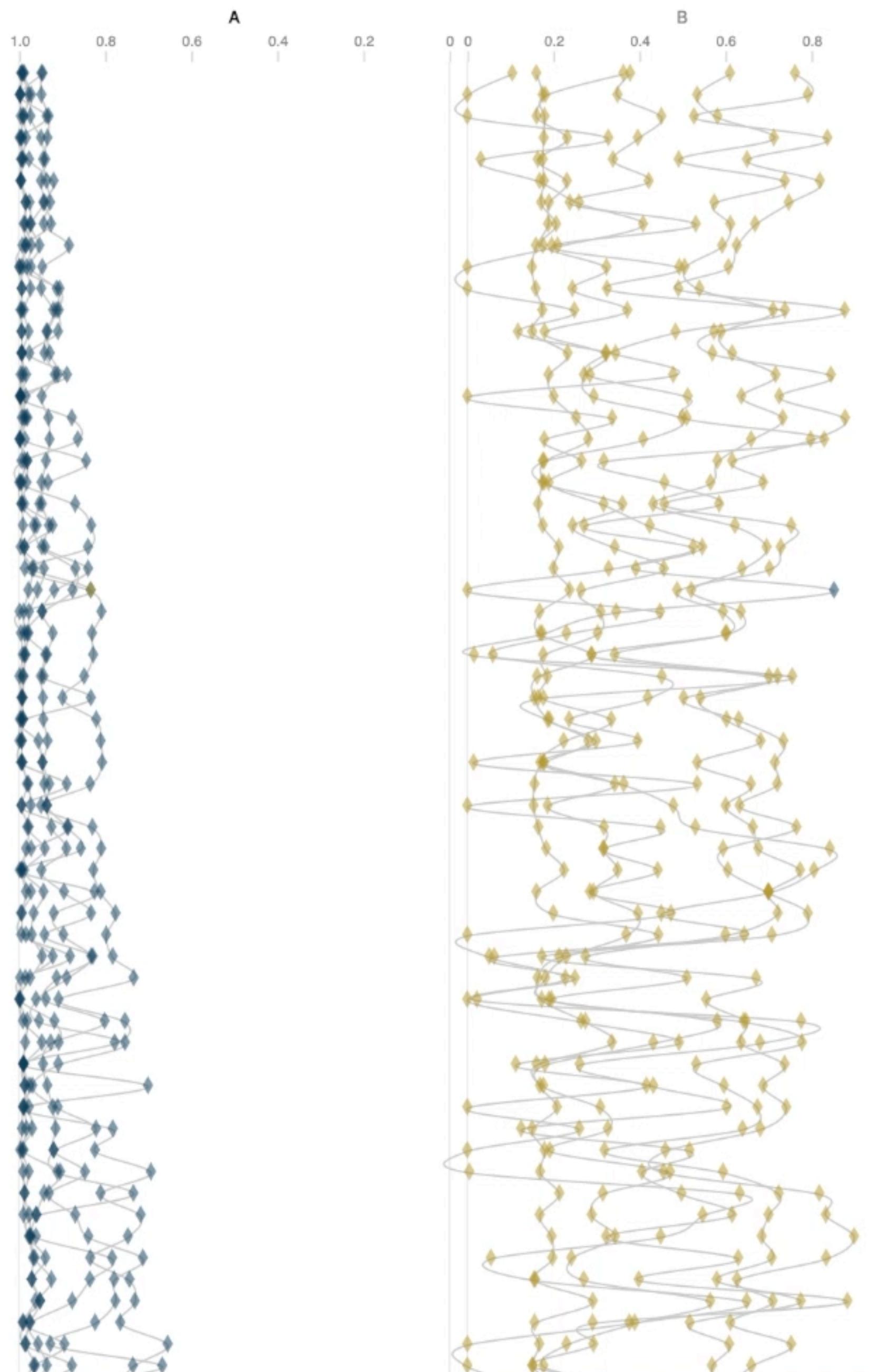
1990-2015 Australian Social and Housing Trends.

Sheet 3
Average Residential House Price in the Greater Sydney Region, 1990-2015.
SUM(Price) | Average House Price (Sydney) AUD\$ Year (Sheet12) | Year
SUM(Women Median Marriage Age) SUM(% of 20-24 Population in Tertiary Education.)

Year % of 20-24 Population in Tertiary Education. Men Median Marriage Age Women Median Marriage Age
 Price % Change \$ Change Sum of annual income Year1 Year2 Year3

Triplets

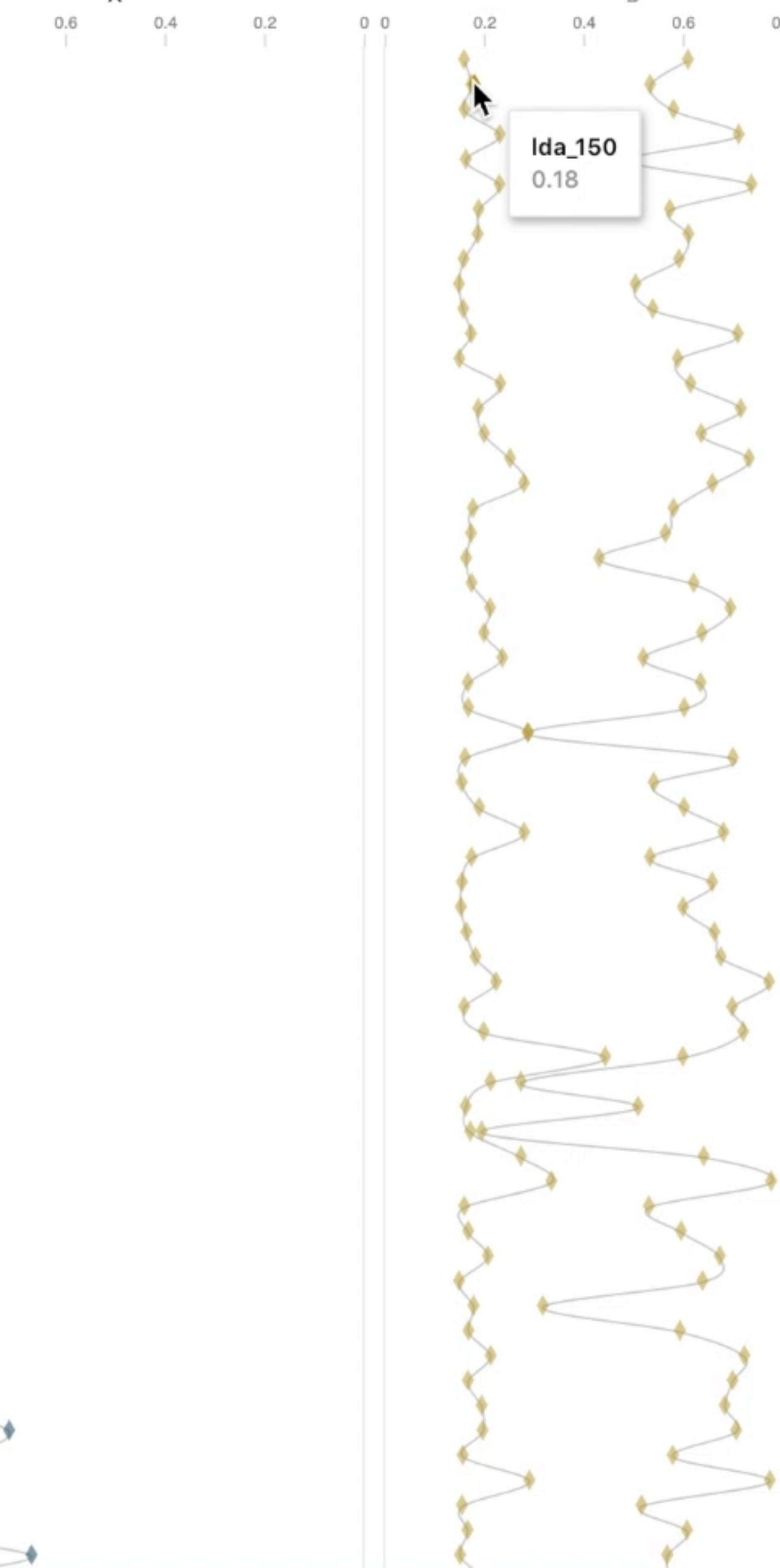
Mean RT	Median RT	A	B
17s	15s	● 24	2
16s	11s	● 25	0
18s	13s	● 26	2
18s	16s	● 22	3
29s	16s	● 22	4
15s	8s	● 25	1
24s	12s	● 27	0
9s	6s	● 26	0
26s	15s	● 23	2
16s	11s	● 23	2
13s	10s	● 23	3
19s	15s	● 16	10
17s	13s	● 24	1
14s	10s	● 26	1
17s	14s	● 23	2
14s	10s	● 25	1
15s	12s	● 28	0
17s	11s	● 24	1
14s	10s	● 22	4
17s	12s	● 25	0
19s	8s	● 26	0
21s	16s	● 23	2
17s	16s	● 22	5
23s	16s	● 24	3
16s	11s	● 23	2
21s	12s	● 26	1
13s	8s	● 24	1
12s	8s	● 26	0
34s	14s	● 15	12
17s	15s	● 24	2
11s	6s	● 26	0
19s	13s	● 25	0
18s	16s	● 27	1
13s	10s	● 26	1
8s	7s	● 27	0
8s	6s	● 27	0
12s	12s	● 25	1
20s	12s	● 24	1
11s	9s	● 26	1
25s	18s	● 23	4
13s	11s	● 24	2
13s	11s	● 26	0
16s	15s	● 18	10
11s	6s	● 25	0
20s	17s	● 25	1
18s	12s	● 19	6
14s	7s	● 25	1
10s	9s	● 26	0
11s	10s	● 25	0
14s	7s	● 25	1
38s	19s	● 26	1
7s	6s	● 26	0
56s	13s	● 27	2
23s	18s	● 22	4
18s	8s	● 26	1
20s	13s	● 23	3
42s	12s	● 24	1
18s	13s	● 23	2
26s	14s	● 23	3
9s	8s	● 23	3
42s	14s	● 24	3



Triplets

Mean RT Median RT

Mean RT	Median RT
17s	15s
16s	11s
18s	13s
18s	16s
29s	16s
15s	8s
24s	12s
9s	6s
26s	15s
16s	11s
13s	10s
19s	15s
17s	13s
14s	10s
17s	14s
14s	10s
15s	12s
17s	11s
14s	10s
17s	12s
19s	8s
21s	16s
17s	16s
23s	16s
16s	11s
21s	12s
13s	8s
12s	8s
34s	14s
17s	15s
11s	6s
19s	13s
18s	16s
13s	10s
8s	7s
8s	6s
12s	12s
20s	12s
11s	9s
25s	18s
13s	11s
13s	11s
16s	15s
11s	6s
20s	17s
18s	12s
14s	7s
10s	9s
11s	10s
14s	7s
38s	19s
7s	6s
56s	13s
23s	18s
18s	8s
20s	13s
42s	12s
18s	13s
26s	14s
9s	8s
42s	14s

A**B**

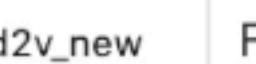
tfidf



lda_150



lsi_150



glove_pre



glove_tf



d2v_new

Filter triplets

Settings