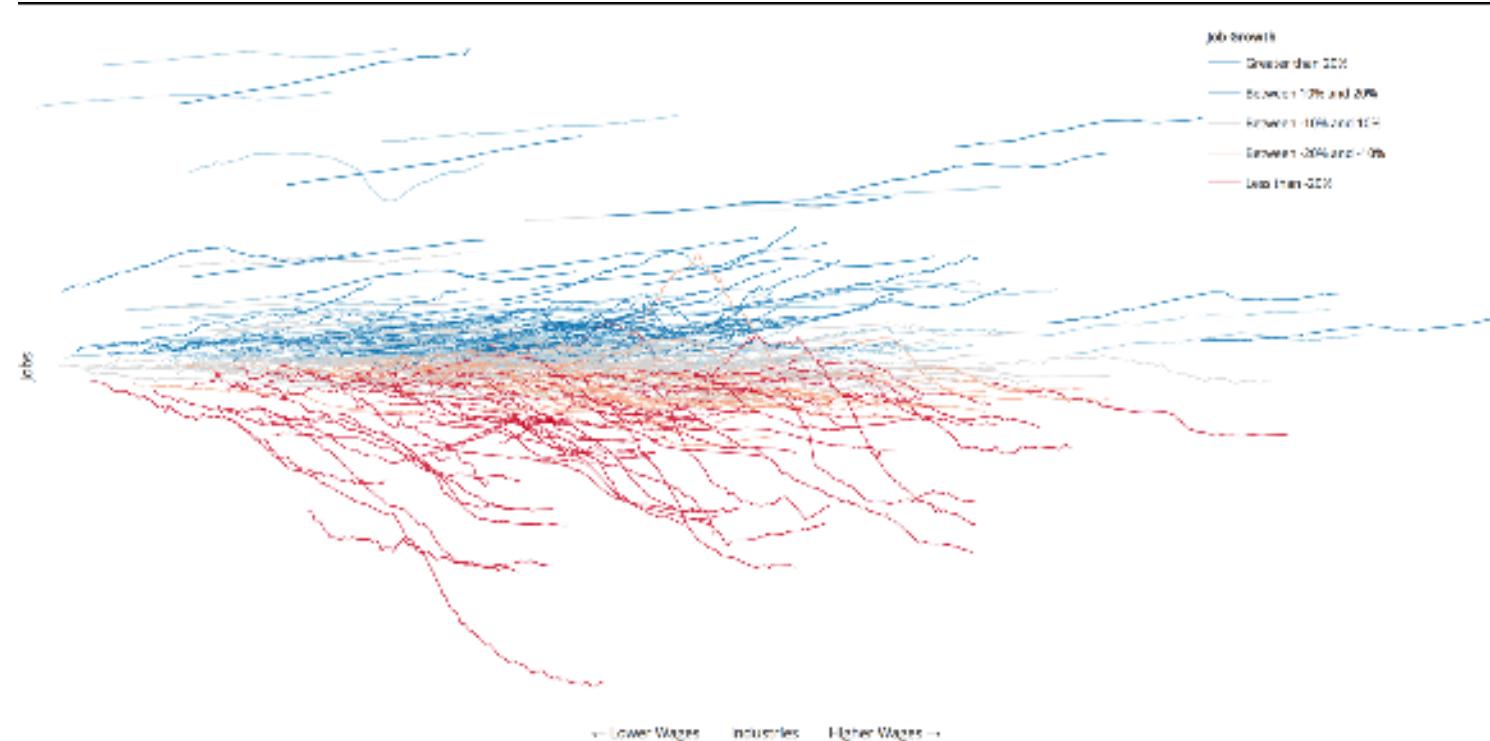


# Using Data Analytics to Understand (and Solve) Problems

David Yakobovitch  
[david@yakobovitch.com](mailto:david@yakobovitch.com)





# Projects That Matter

1887

1892

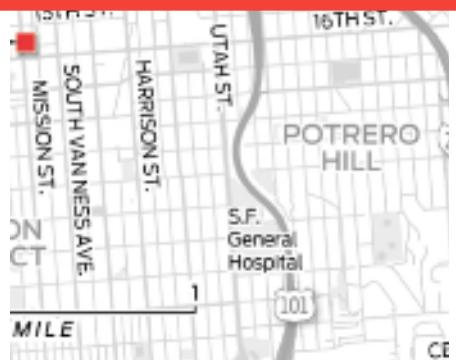
CHALLENGERS

LEADERS

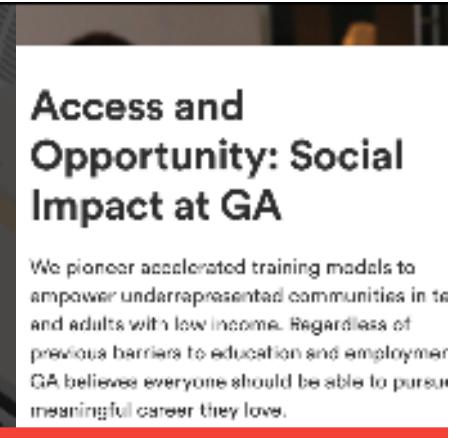
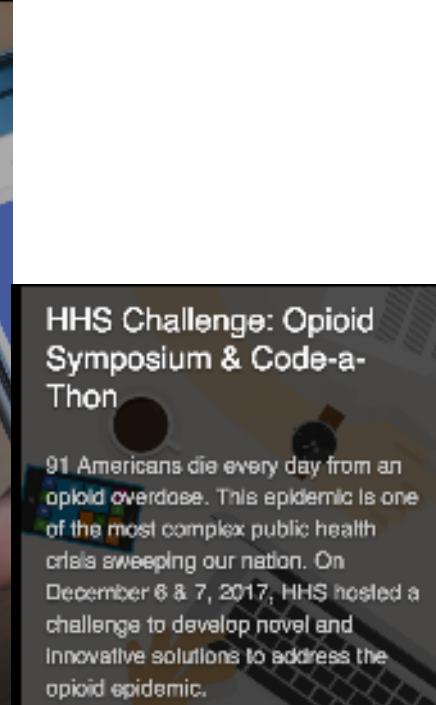
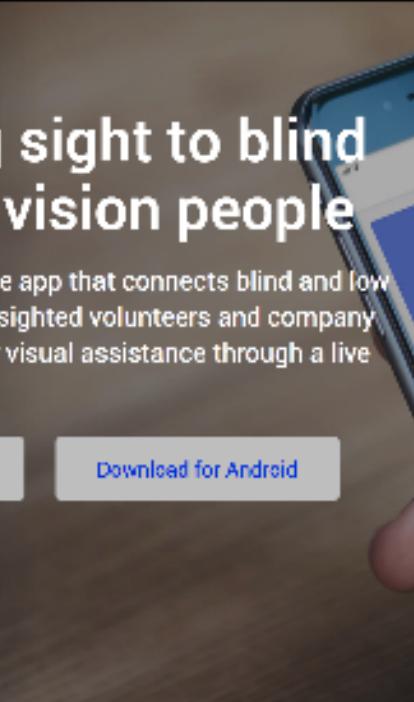


## Gartner Magic Quadrants

# Break throughs



## Navigation Centers

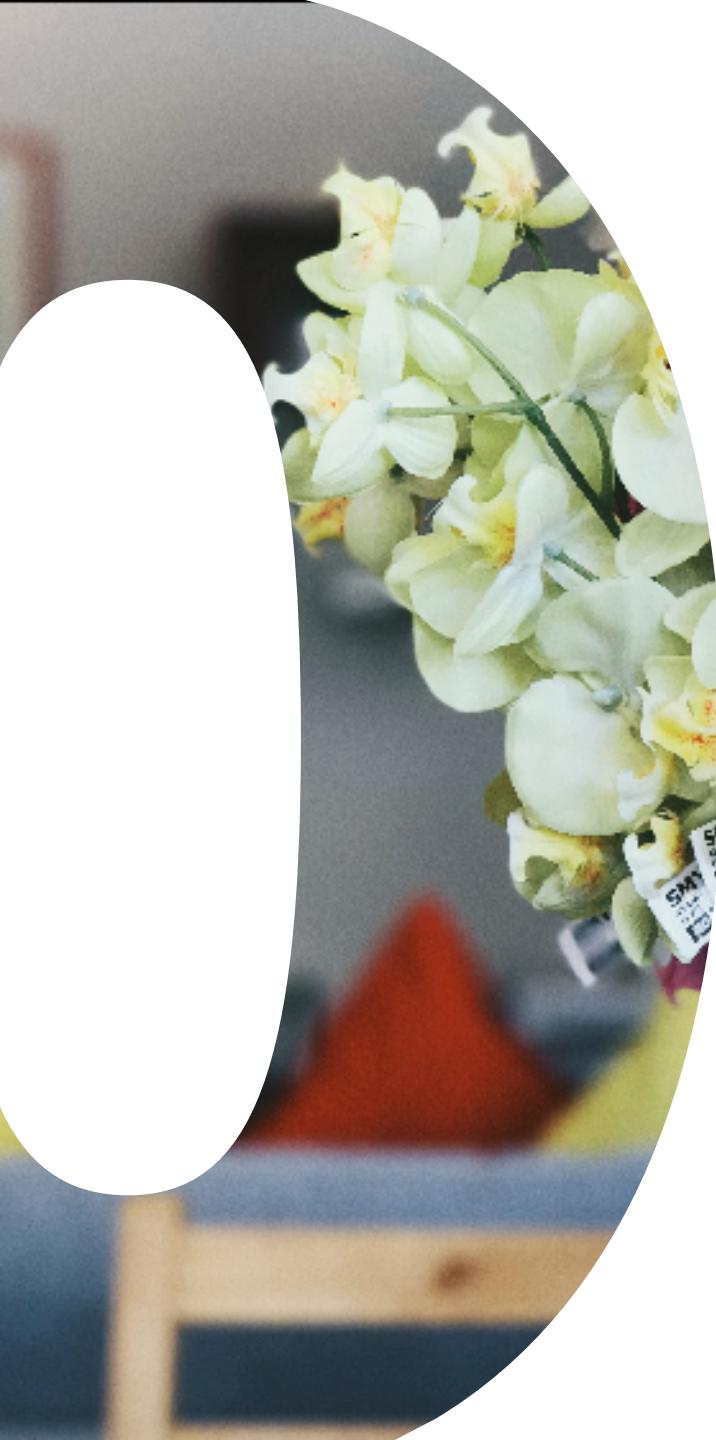


Across all  
industries



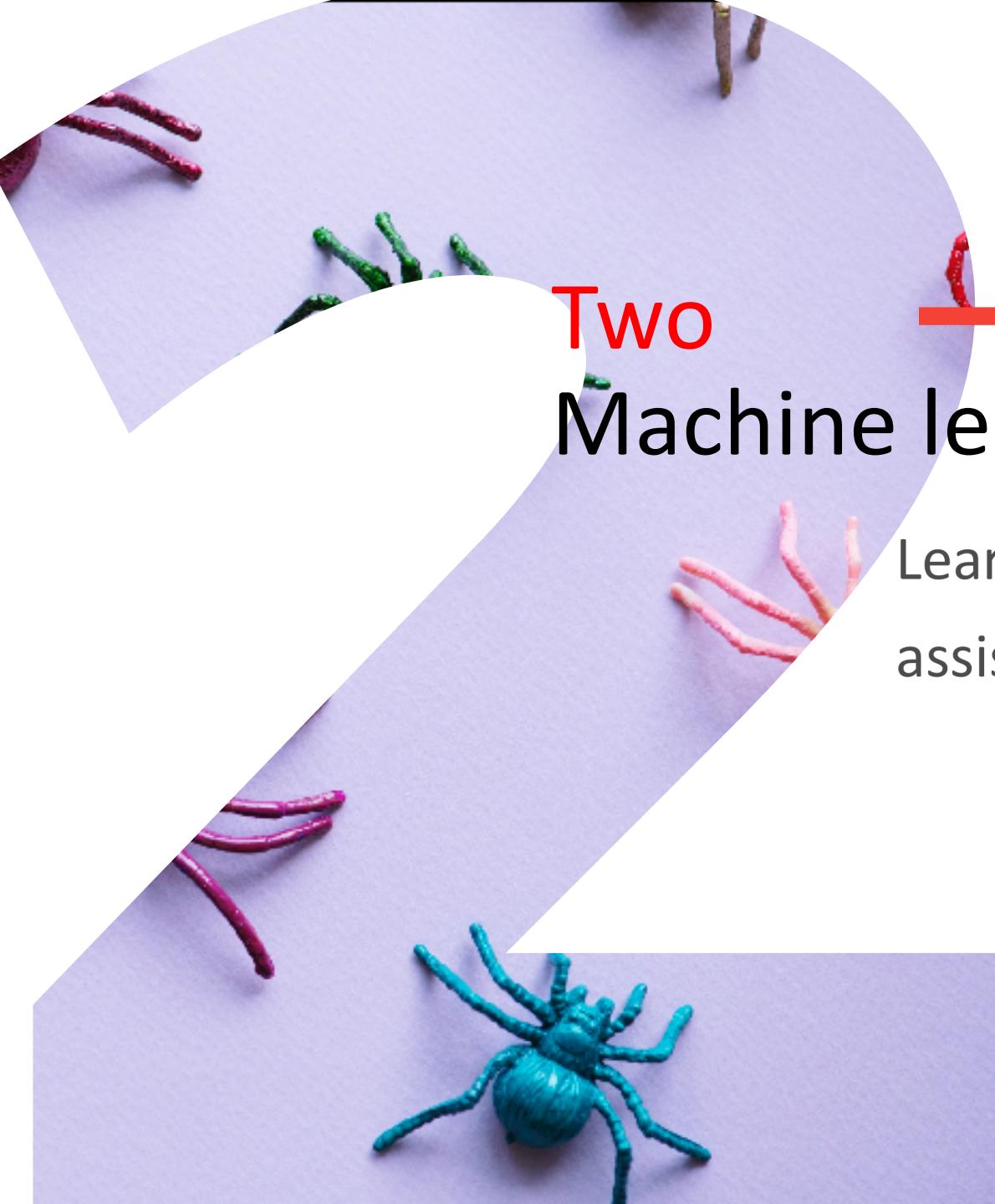
# Data First

Yesterday → Today



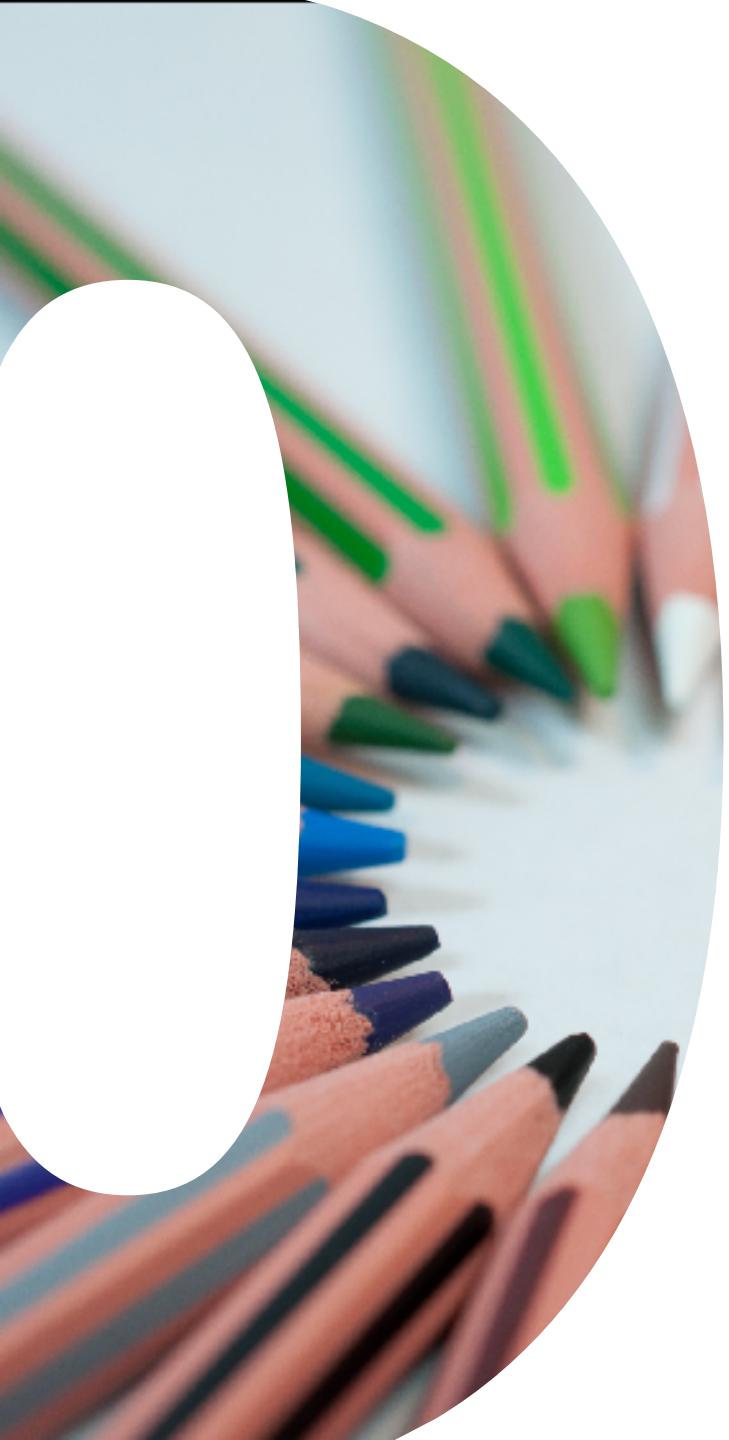
# One Artificial intelligence

Self-learning at or  
above human level



# Two Machine learning

Learning with human  
assistance



# Three Data science



Scientific method  
applied to data

# Start here

## DATA SCIENCE WORKFLOW



# Individual Exercise

---

What challenges are your organization experiencing?



# IRT Scoring (1 to 10)

How **impactful** will this challenge be to your organization and community?

How much will this improve your organizational **readiness**?

How **time** sensitive is this goal for your organization to achieve in the next 6 months?

# Rank your challenges

---

Add all 3 numbers.

Which ranked the highest?

How close was second?

Third?

# SMART Goals: Independent

---

Specific:	What will you accomplish?
Measurable:	By how much will you increase?
Actionable:	With what method will you grow?
Realistic:	How plausible is this growth hack to achieve?
Timely:	By when will you accomplish this goal?

Reduce veteran homelessness.

Vs.

By November 2, 2018, shorten the length of homelessness by 40% in 3 counties through text message alerts with over 5,000 veterans.

# SMART Goals: Demo Workshop

---

Specific:	What will you accomplish?
Measurable:	By how much will you increase?
Actionable:	With what method will you grow?
Realistic:	How plausible is this growth hack to achieve?
Timely:	By when will you accomplish this goal?

Reduce veteran homelessness.

Vs.

By November 2, 2018, shorten the length of homelessness by 40% in 3 counties through text message alerts with over 5,000 veterans.

# SMART Goals: Group Share

---

Specific:	What will you accomplish?
Measurable:	By how much will you increase?
Actionable:	With what method will you grow?
Realistic:	How plausible is this growth hack to achieve?
Timely:	By when will you accomplish this goal?

Reduce veteran homelessness.

Vs.

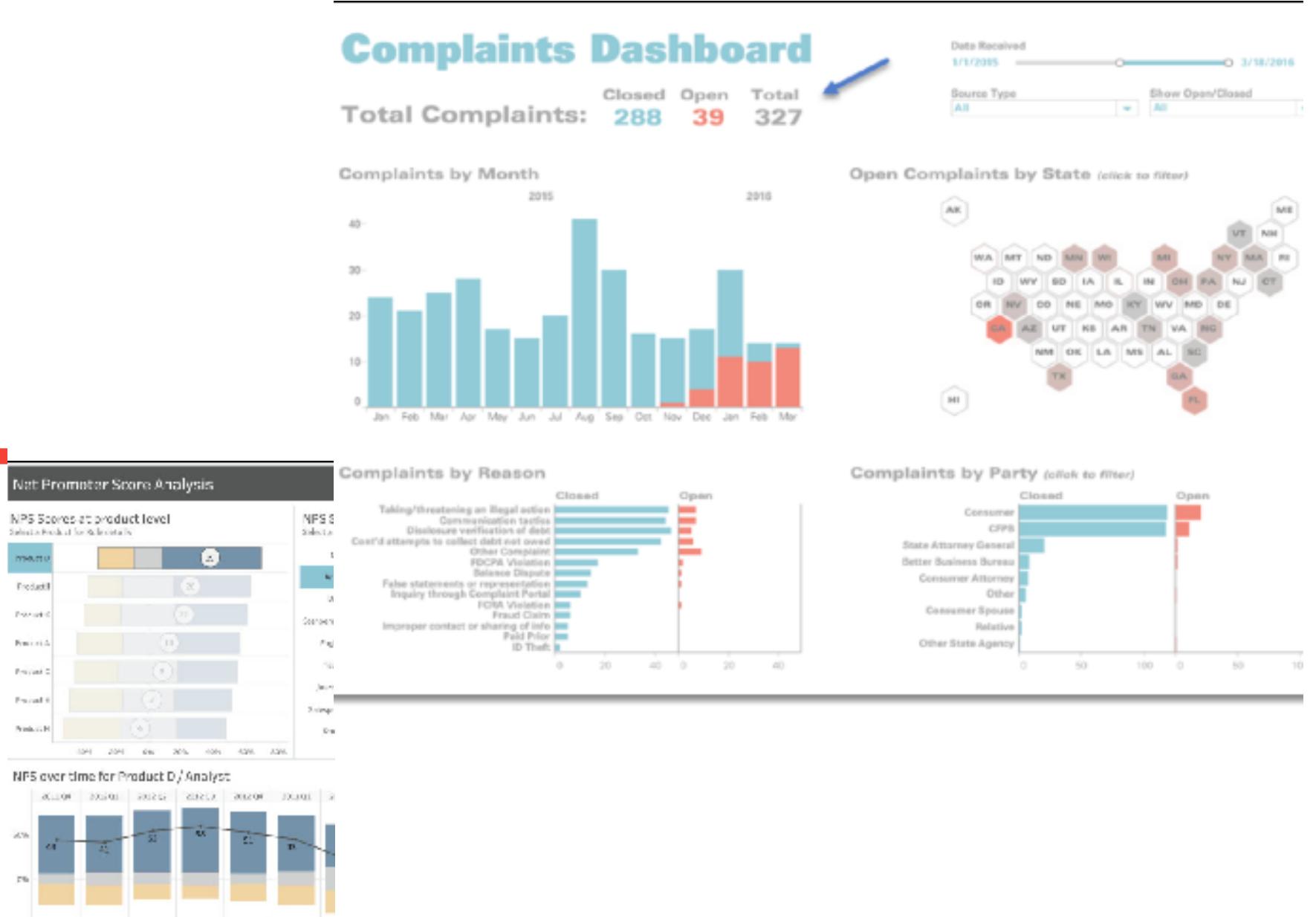
By November 2, 2018, shorten the length of homelessness by 40% in 3 counties through text message alerts with over 5,000 veterans.

TRACKING YOUR DATA

---

IN THE CLOUD

# CREATE DASHBOARDS





davidyakovitch Rename open\_data\_sets to open\_data\_sets.md

91ab771 3 days ago

1 contributor

88 lines (75 sloc) | 4.57 KB

Raw

Blame

History



The following are drop-down guides for open data resources, to assist you during your data science journey for creating projects and accessing open source data recommendations.

▼ Open Data Sets

1. [Data is Plural](#)
2. [India Open Data Gov](#)
3. [Canada Open Data Gov](#)
4. [US Open Data Gov](#)
5. [Quandl Financial Data](#)
6. [UCI Machine Learning Datasets](#)
7. [Open India](#)
8. [FiveThirtyEight](#)

# EXPLORE DATA SOURCES

---

9. [Data Catalogs](#)

10. [world](#)
11. [Wikipedia ML Listings](#)
12. [Cool Datasets on Twitter](#)
13. [Public Data Science Datasets](#)
14. [OpenML](#)
15. [Github: Awesome Public Datasets](#)
16. [Kaggle Datasets](#)
17. [data.ny.gov](#)

# EXPLORE DATA SOURCES

---

Visit:  
[bit.ly/opendatasets](https://bit.ly/opendatasets)

[Home](#)[Environments](#)[Learning](#)[Community](#)

Applications on

base (root)

Channels

Refresh



jupyterlab

0.32.0

An extensible environment for interactive and reproducible computing, based on the Jupyter Notebook and Architecture.

[Launch](#)

notebook

5.4.1

Web-based, interactive computing notebook environment. Edit and run human-readable docs while describing the data analysis.

[Launch](#)

vscode

1.24.1

Streamlined code editor with support for development operations like debugging, task running and version control.

[Launch](#)

glueviz

0.13.3

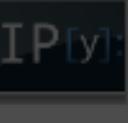
Multidimensional data visualization across files. Explore relationships within and among related datasets.



orange3

3.13.0

Component based data mining framework. Data visualization and data analysis for novice and expert. Interactive workflows with a large toolbox.



qtconsole

4.3.1

PyQt GUI that supports inline figures, proper multiline editing with syntax highlighting, graphical calltips, and more.

[Documentation](#)[Developer Blog](#)[Feedback](#)



PM Dashboard x

Secure | https://tychobra.shinyapps.io/wc-pm-dashboard/

Andy

Apps actuarial stats blogs edu tychobra G Web Dev caret book learnr DT tricks React Table plumber Using databases wit... Postgres Guide react-sounds next.js go intro Firebase Web Codelab Other Bookmarks

PM Dashboard

Overview Individual Claims Tour this Tab

16,634,301 Average Predicted Payments

502,320 Standard Deviation of Prediction

16,493,430 Actual Claim Payments

Metric Payments Status

Claim Payments Simulation Between Age 1 and Age 2

Actual Payments = 16,493,430

20% Confidence Level = 16,221,697

87% Confidence Level = 17,276,913

Number of Observations

Simulated Payments

Confidence Interval

0% 20% 87% 100%

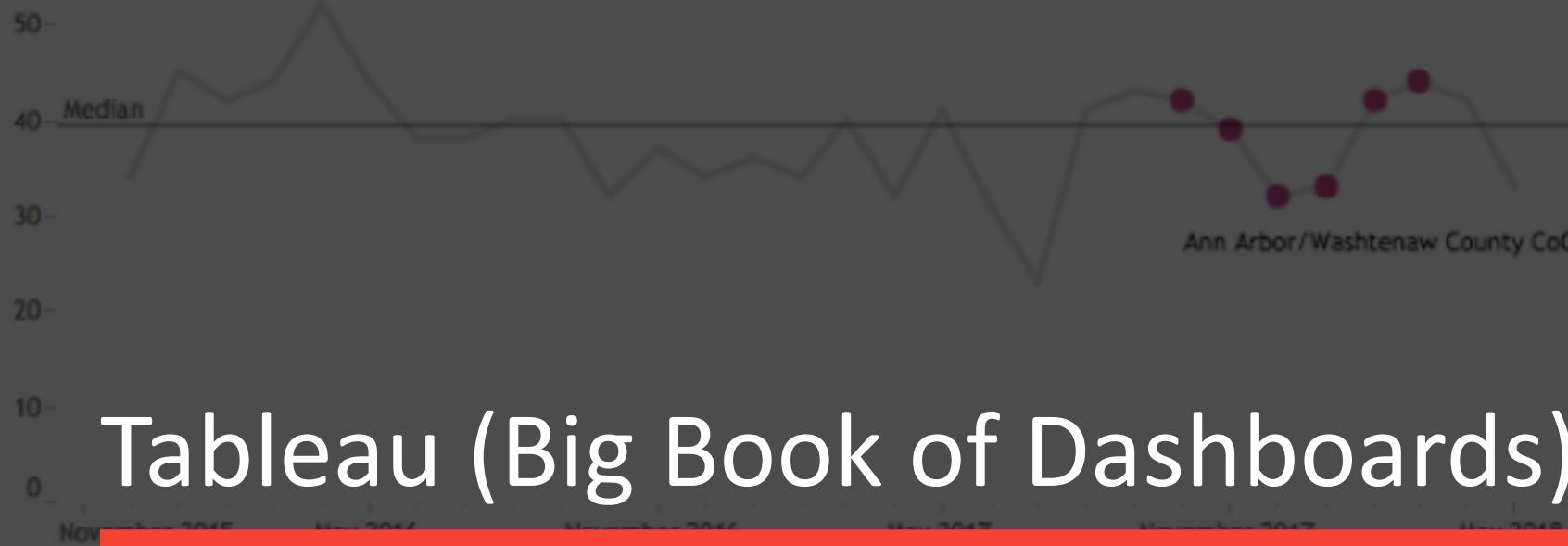
R Shiny Apps

bit.ly/rshinyapp

https://tychobra.shinyapps.io/wc-pm-dashboard/#shiny-tab-claims

## Actively Homeless

Monthly trend with signal indicators for Veteran subpopulation(s)



### INSTRUCTIONS

Select your community from the drop-down menu below and use filters to adjust the data. You'll be able to see additional information by hovering over the charts.

NOTE: Numbers reflect self-reported community data (submitted using the form below).

Got questions? Email us at  
[bfzdatasupport@community.solutions](mailto:bfzdatasupport@community.solutions)

### DASHBOARD FILTERS

#### Select Community

Ann Arbor/Washtenaw County CoC

#### Select Subpopulation

Veteran

(All)

Chronic

Veteran

Cancel

Apply

#### Select Type of Signal

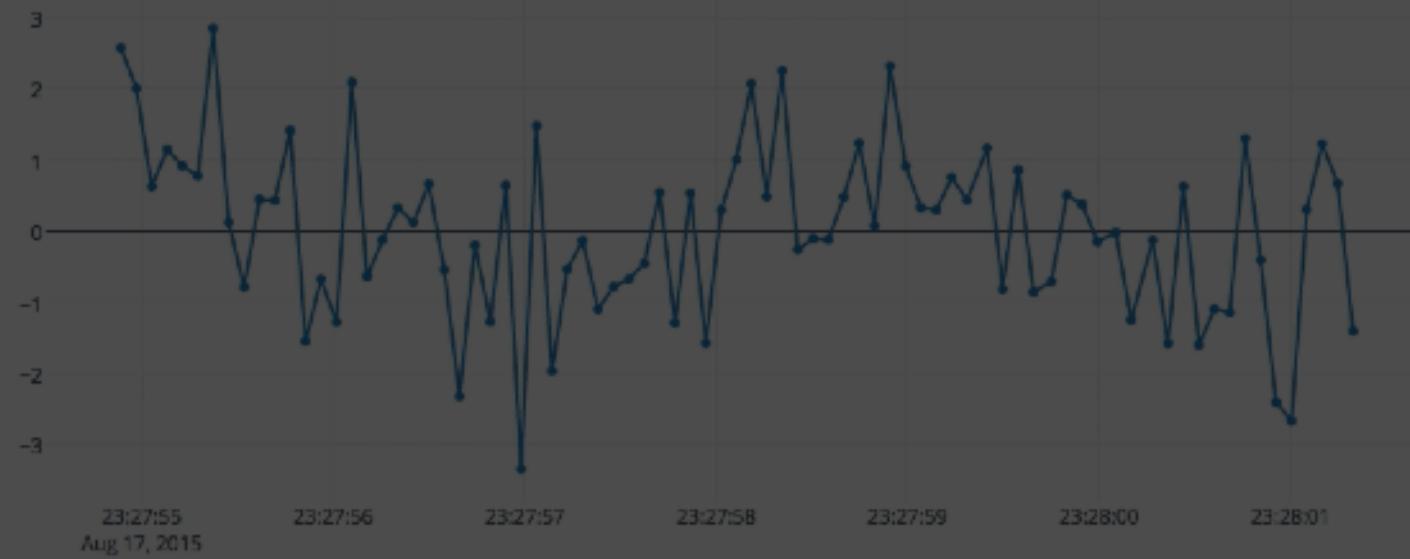
Trend



## Actively Homeless Population

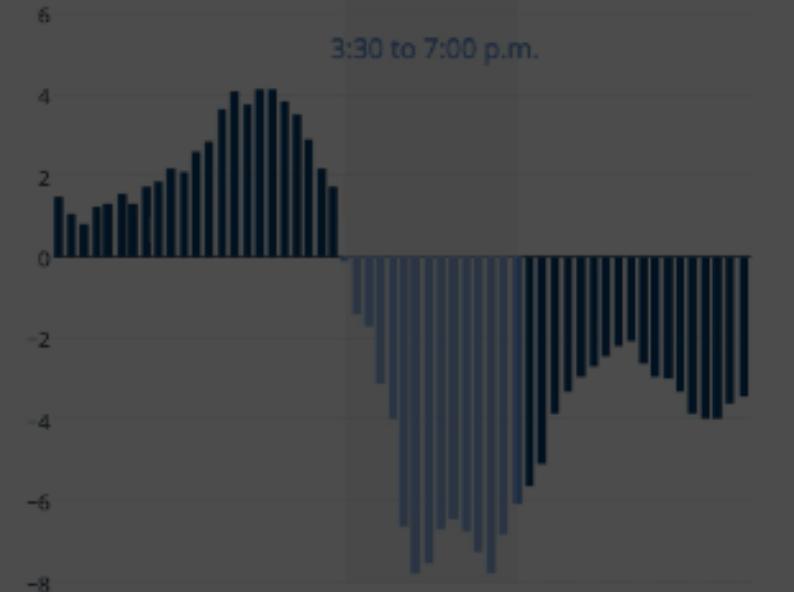
Monthly count for Veteran subpopulation(s)

[bit.ly/bigbookdash](http://bit.ly/bigbookdash)



23:27:55  
Aug 17, 2015

[Edit chart »](#) - Source:



3:30 to 7:00 p.m.

10 15 20 [Edit chart »](#)

[Make a bar chart in Python](#)

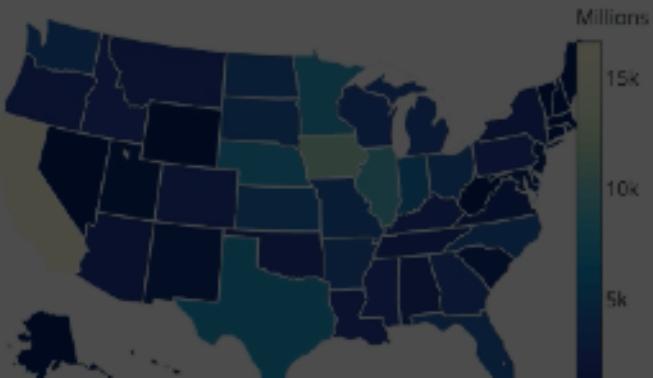
[Stream live data to charts with Python, MATLAB, or Node.js.](#)

Map Example

Cron Job Example

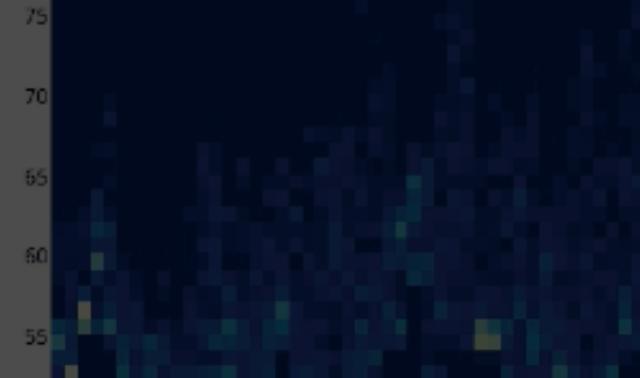
Histogram Fit Example

# EXPLORE WHAT ARE GOOD GRAPHS?



Millions USD

15k  
10k  
5k



12

40  
35  
30  
25  
20  
15  
10  
5

Frequency

12  
10  
8  
6  
4  
2  
0

12

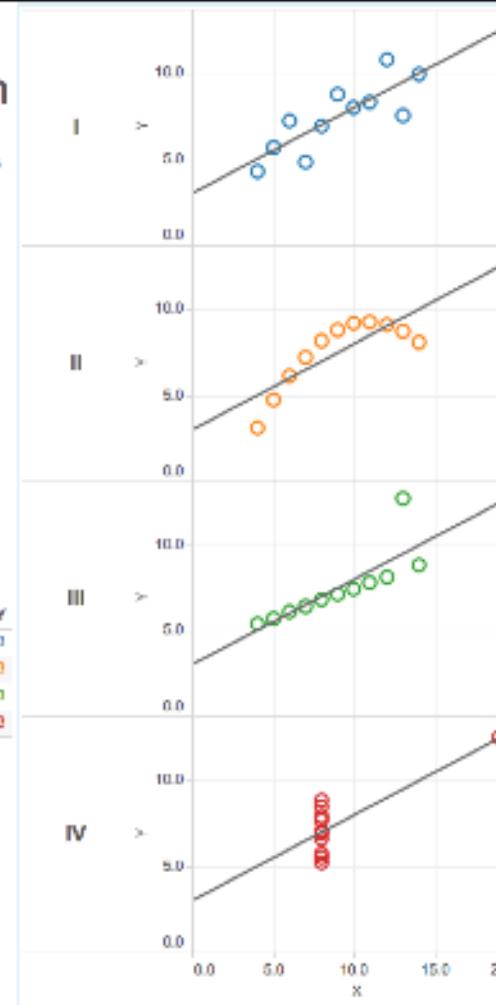
# Power of Visualization

## Anscombe's Quartet: The power of visualization

These four data sets have identical summary statistics, yet the plots show vastly different stories

I	II	III	IV
(6, 4.3)	(4, 3.1)	[6, 5.4]	[8, 5.3]
(7, 4.8)	(5, 4.7)	[5, 5.7]	[8, 5.6]
(5, 5.7)	(6, 6.1)	[6, 6.1]	[8, 5.8]
(8, 7.0)	(7, 7.2)	[7, 6.4]	[8, 5.6]
(6, 7.2)	(14, 8.1)	[6, 6.8]	[8, 5.9]
(13, 7.0)	(8, 8.1)	[9, 7.1]	[8, 7.0]
(10, 8.0)	(13, 8.7)	[10, 7.5]	[8, 7.7]
(11, 8.3)	(9, 8.6)	[11, 7.8]	[8, 7.9]
(9, 8.8)	(12, 9.1)	[12, 8.2]	[8, 8.5]
(14, 10)	(10, 9.1)	[14, 8.8]	[8, 8.8]
(12, 10.0)	(11, 9.3)	[13, 12.7]	[19, 12.5]

Summary Statistics						
Plot	sum X	sum Y	avg X	avg Y	sdev X	sdev Y
I	99.0	82.5	9.00	7.50	3.32	2.03
II	99.0	82.5	9.00	7.50	3.32	2.03
III	99.0	82.5	9.00	7.50	3.32	2.03
IV	99.0	82.5	9.00	7.50	3.32	2.03



# Attributes of Good Visualizations

---

Length	Width	Orientation	Size	Shape	Curvature
--------	-------	-------------	------	-------	-----------



Enclosure	2-D Position	Spatial Grouping	Color (Hue)	Color (Intensity)
-----------	--------------	------------------	-------------	-------------------



# Attributes of Good Visualizations

---

Length	Width	Orientation	Size	Shape	Curvature
--------	-------	-------------	------	-------	-----------



Enclosure	2-D Position	Spatial Grouping	Color (Hue)	Color (Intensity)
-----------	--------------	------------------	-------------	-------------------



# Example 1

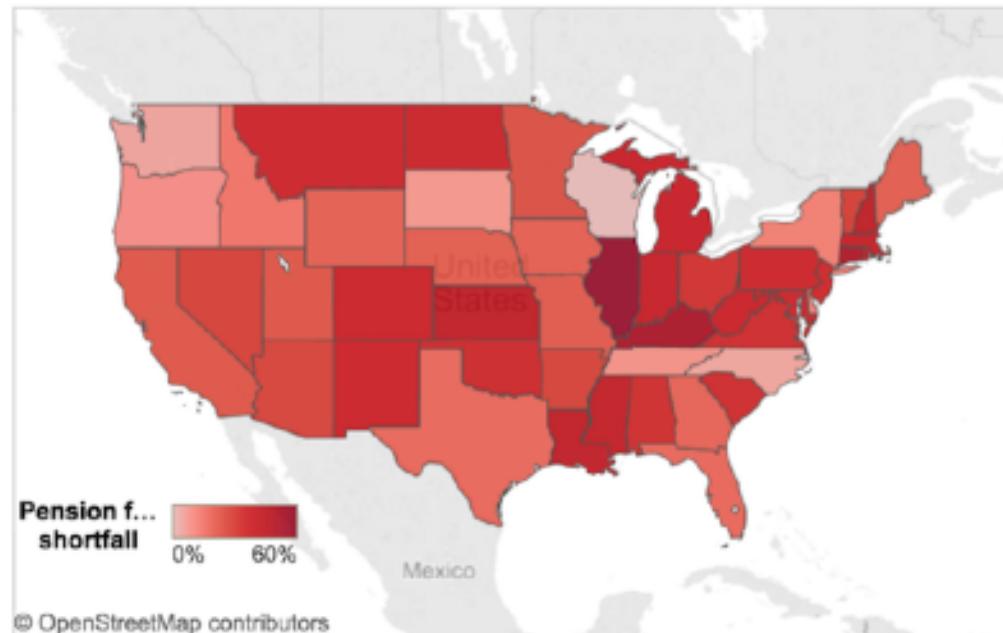
## Pensions in Peril

Despite recent stock market gains, states continue to shortchange their pension plans, leaving many of them badly underfunded. (SOURCE: Pew Charitable Trusts)

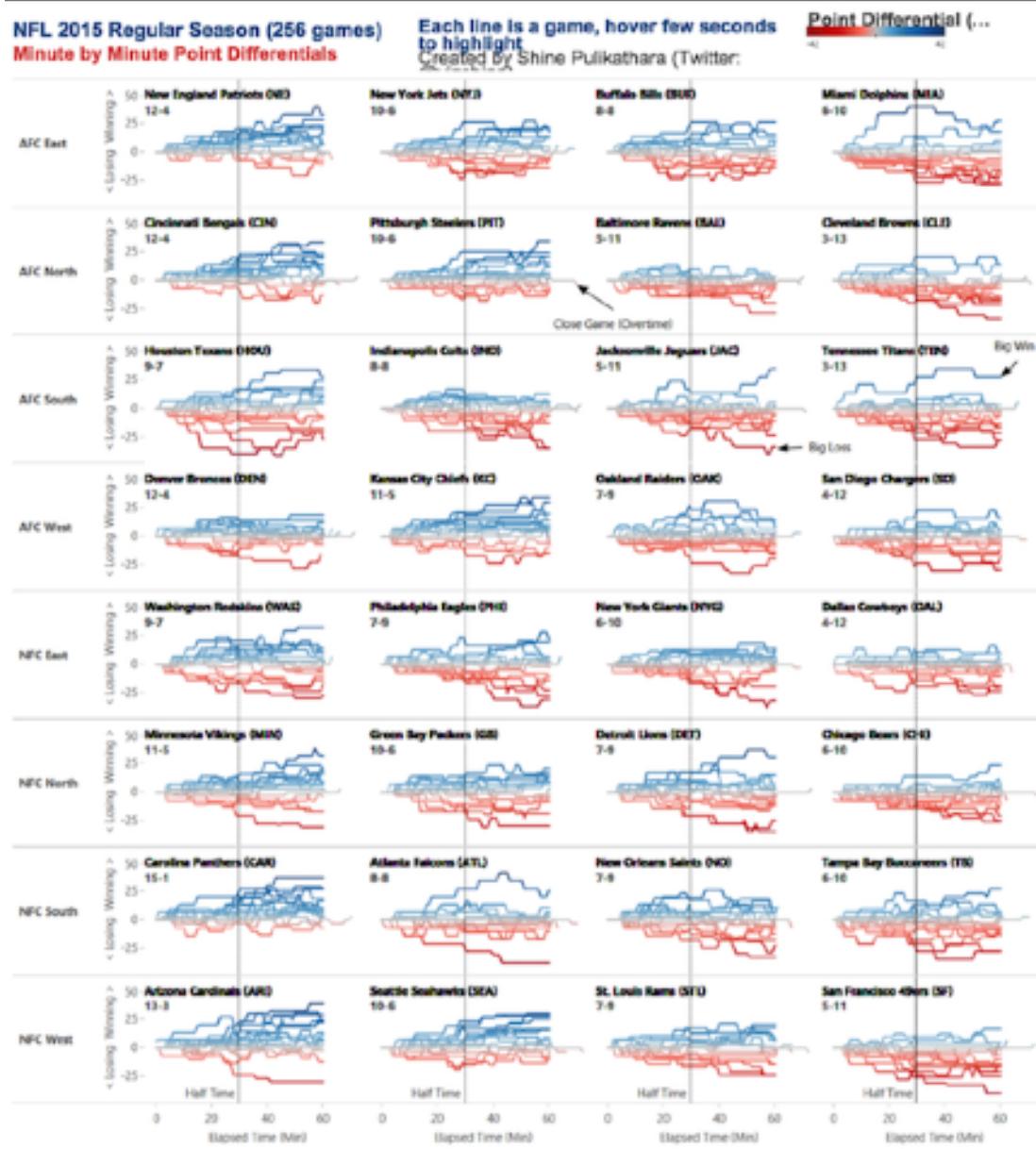
CNBC

(Dropdown for AK, HI)

Contiguous US ▾



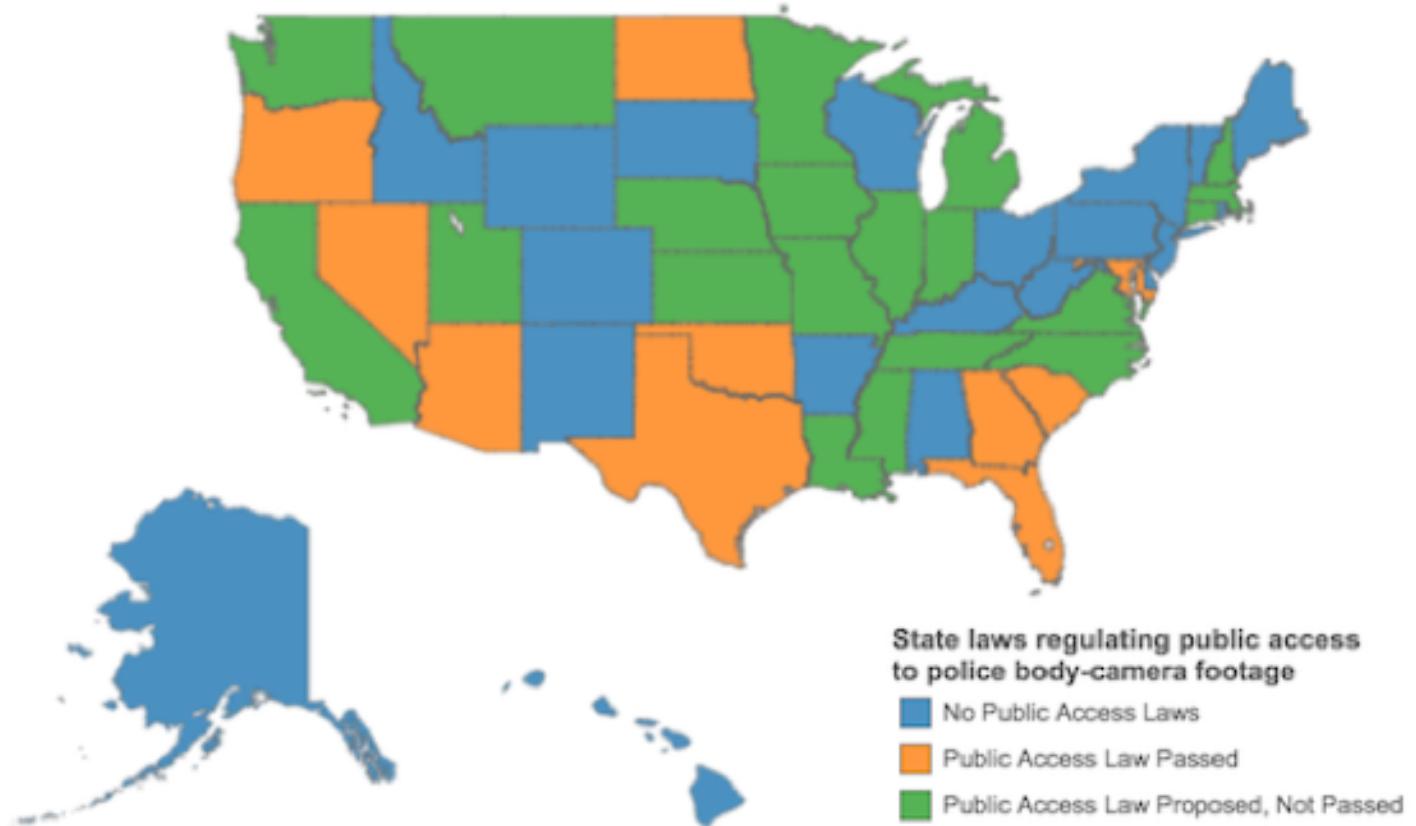
# Example 2



# Example 3

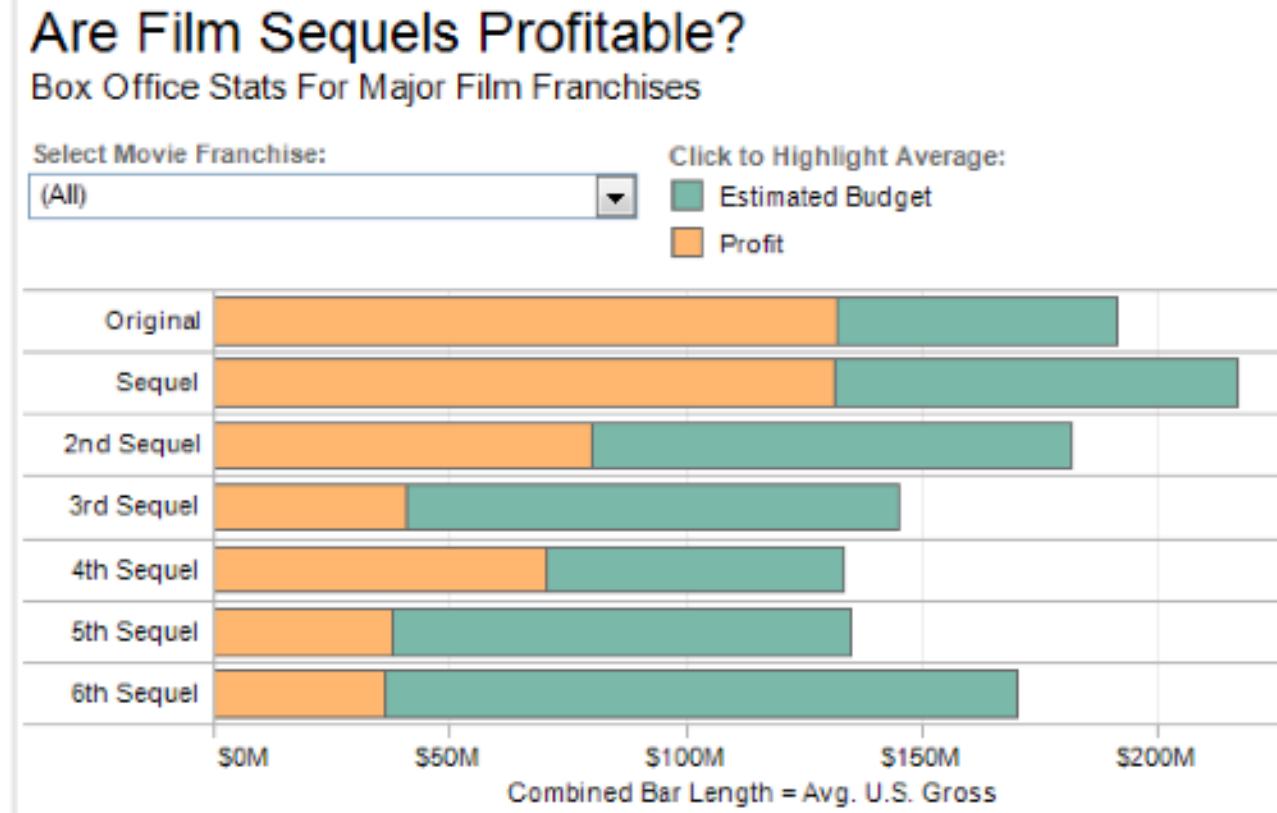
## Body Camera Laws

Ten states have passed laws that control the public's access to footage from police body cameras. Hover over each state for more information.



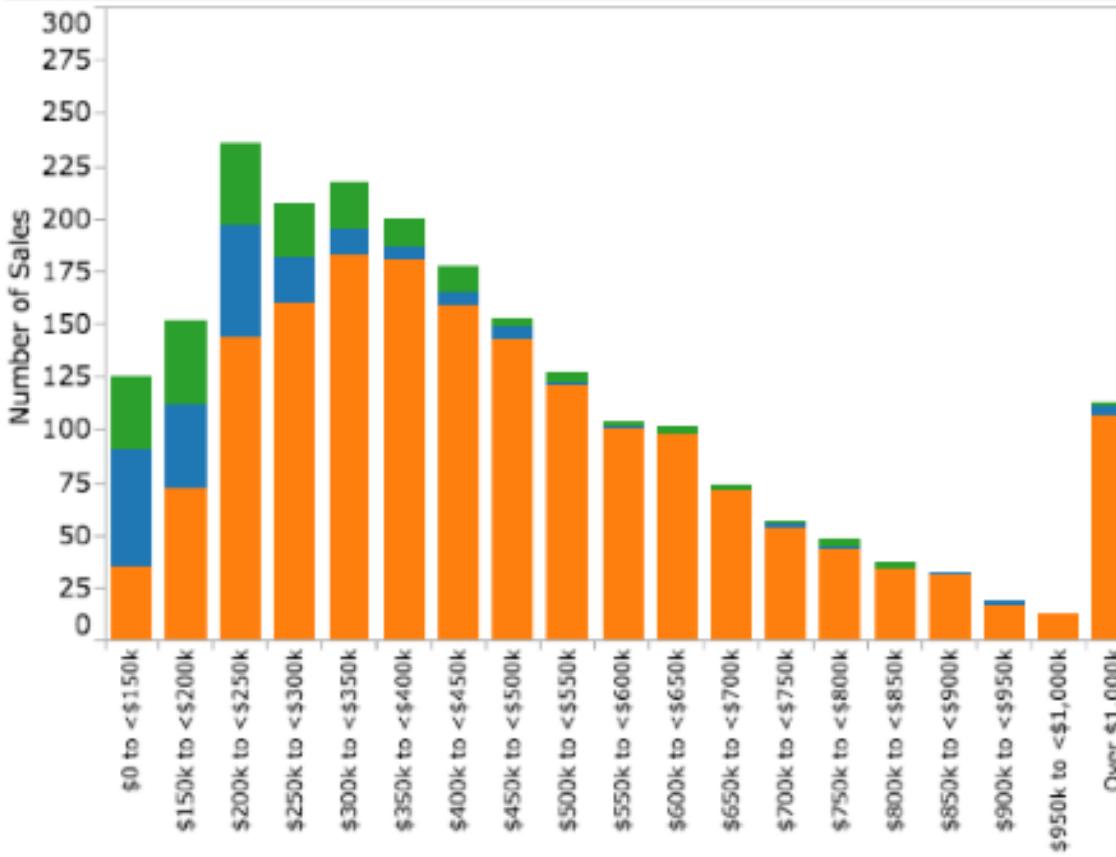
Source: Reporters Committee for Freedom of the Press

# Example 4



# Example 5

King Co. SFH Sales Histogram [Sold 2012-06]



Sale Month:  
Sold 2012-06

County:  
 King  
 Pierce  
 Snohomish

Distress:  
 (All)  
 Bank Owned  
 Non-Distressed  
 Short Sale

Distress Status  
 Short Sale  
 Bank Owned  
 Non-Distressed

100%  \$ % . . 123 

Arial

10

B

I

S

A

H

E

M

L

D

V

C

+

I

Y

-

Σ

A

B

C

D

E

F

G

H

I

J

K

L

M

# MAKE YOUR WORKFLOW

---

# Continue here



# The 5 steps

01

02

## Identify the problem

What is the challenge you would like to solve? What is the hypothesis and critical goals for success?

## Acquire the data

Identify the right data sets and tools to work with. Read documentation and review the data.

# The 5 steps

01

02

03

04

05

## Identify the problem

What is the challenge you would like to solve? What is the hypothesis and critical goals for success?

## Acquire the data

Identify the right data sets and tools to work with. Read documentation and review the data.

## Refine the data

Clean the data and add calculations to better explain and understand your data.

## Build data models

Whether visualizations, or data science models, explore the insights and trends that you can reveal from your data.

## Communicate your results

Create a dashboard, a report, or a presentation to share the outcomes with both your internal and external stakeholders



100% \$ % .0 .00 123

Arial

10

B

I

S

A

A

B

C

D

E

F

G

H

I

J

K

L

M

# SHARE YOUR WORKFLOW

---

# Using Data Analytics to Understand (and Solve) Problems

David Yakobovitch  
[david@yakobovitch.com](mailto:david@yakobovitch.com)

