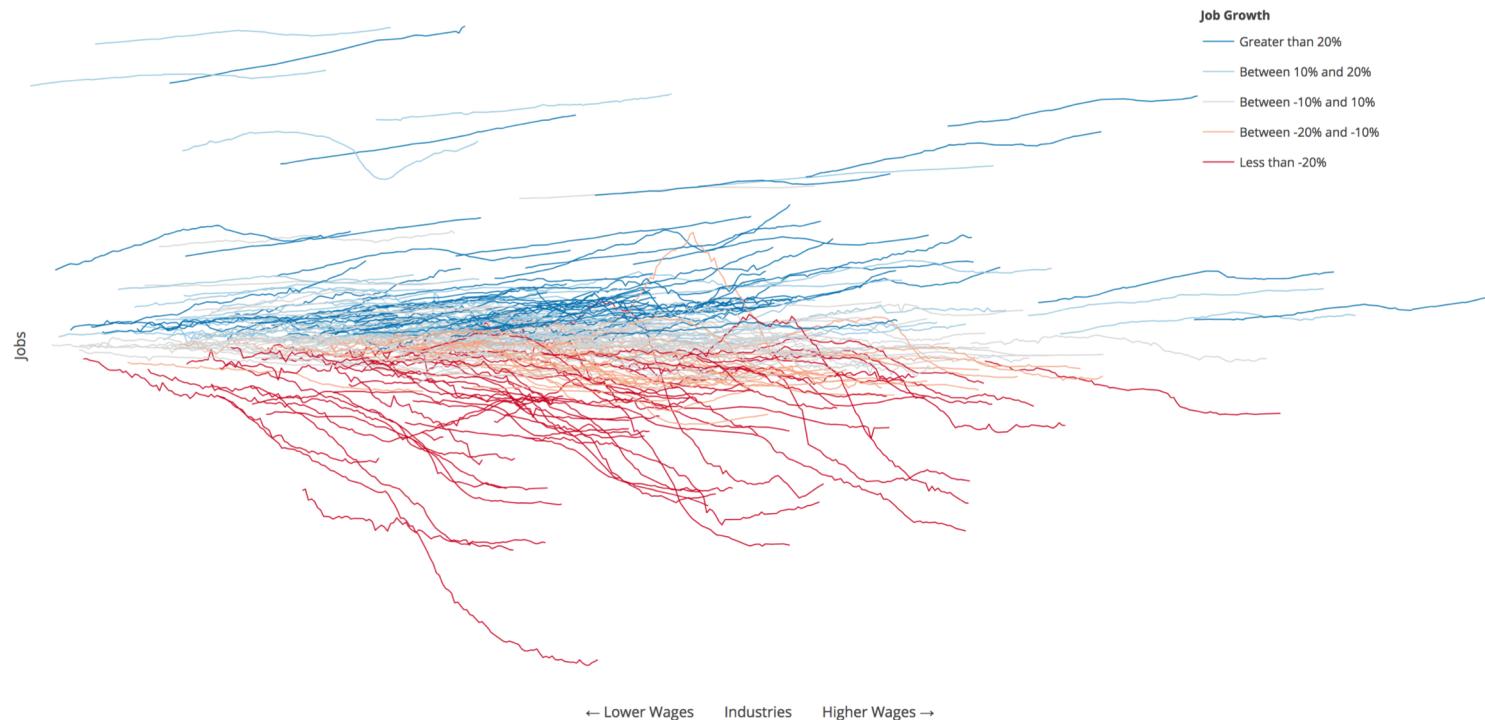


Using Data Analytics to Understand (and Solve) Problems

David Yakobovitch
david@yakobovitch.com



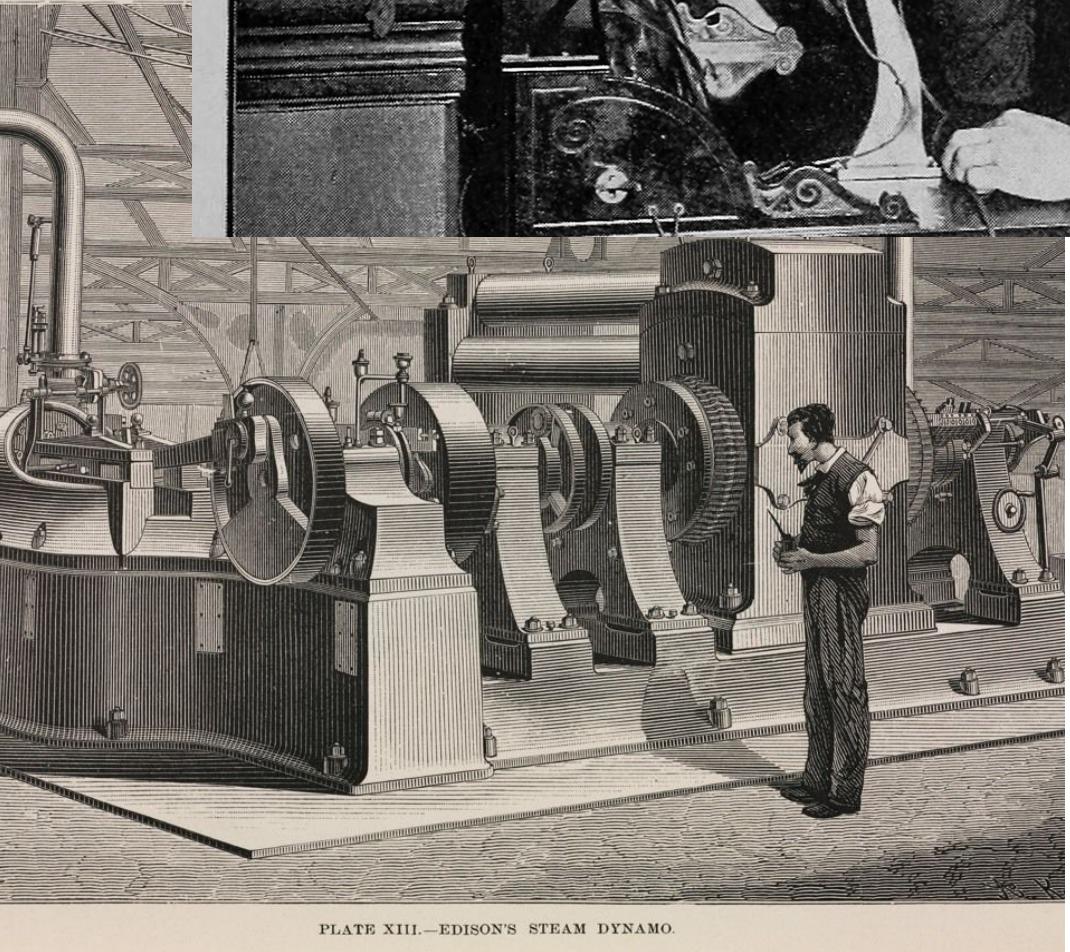


PLATE XIII.—EDISON'S STEAM DYNAMO.



Projects That Matter

1887

1892

CHALLENGERS

LEADERS

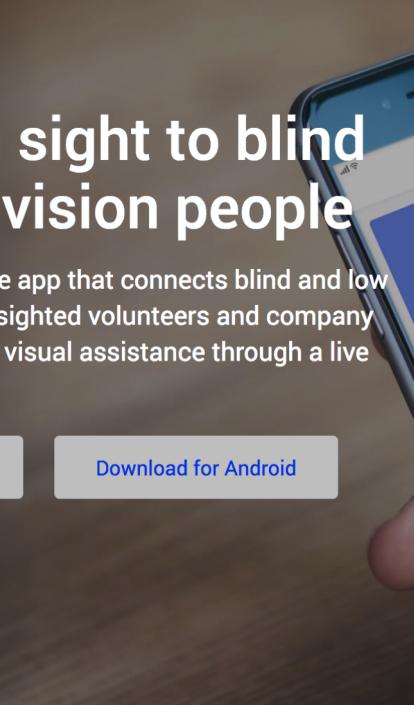


Gartner Magic Quadrants

Break throughs



Navigation Centers



Access and Opportunity: Social Impact at GA

We pioneer accelerated training models to empower underrepresented communities in tech and adults with low income. Regardless of previous barriers to education and employment, GA believes everyone should be able to pursue a meaningful career they love.

Across all
industries





One
Artificial intelligence

Self-learning at or
above human level



Two Machine learning

Learning with human
assistance

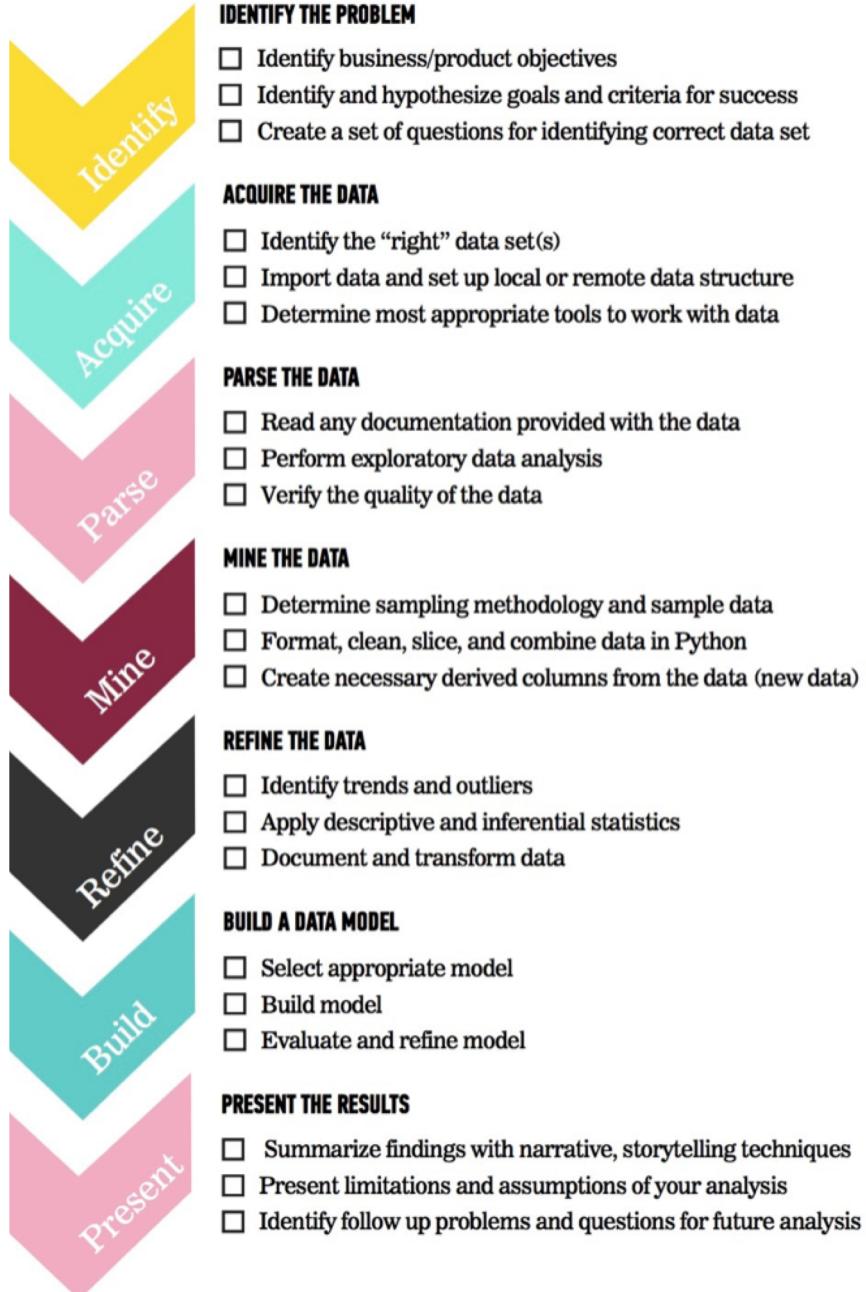


Three Data science

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.



100% ▾

\$

%

.0

.00

123

▼

Arial



10



B

I



A



A

B

C

D

E

F

G

H

I

J

K

L

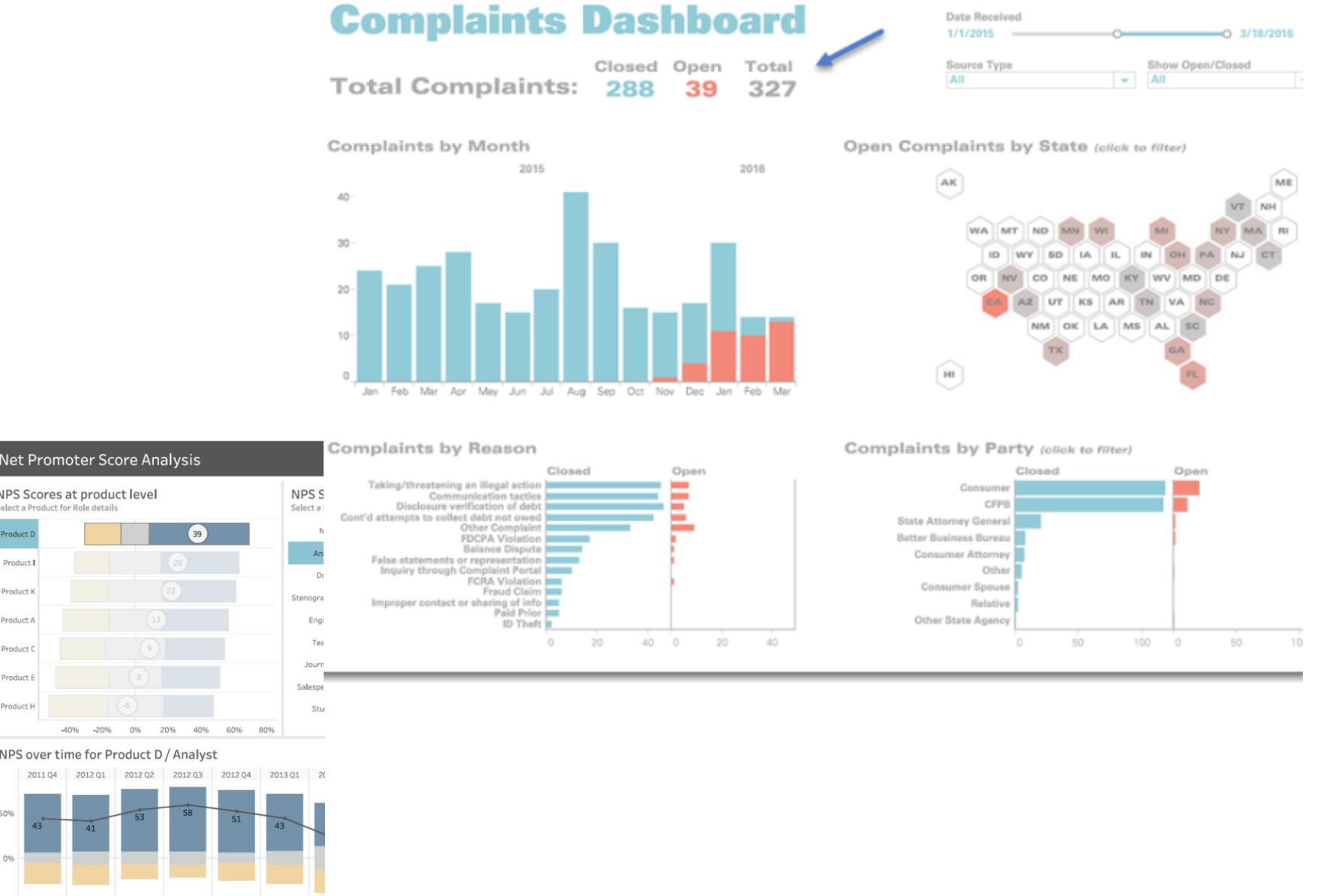
M

N

TRACKING YOUR DATA

IN THE CLOUD

CREATE DASHBOARDS



**davyakobovitch** 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. [OpenInde](#)
8. [FiveThirtyEight](#)

EXPLORE DATA SOURCES

9. [DataCamp](#)
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

Home

Environments

Learning

Community

Applications on

base (root)

Channels

Refresh

lab
jupyterlab
 0.32.0

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

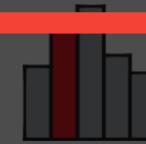
jupyter
notebook
 5.4.1

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

vscode
 1.24.1

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

EXPLORE TOOLS FOR DATA



glueviz

0.13.3

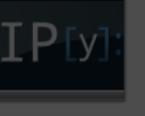
Multidimensional data visualization across files. Explore relationships within and among related datasets.



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

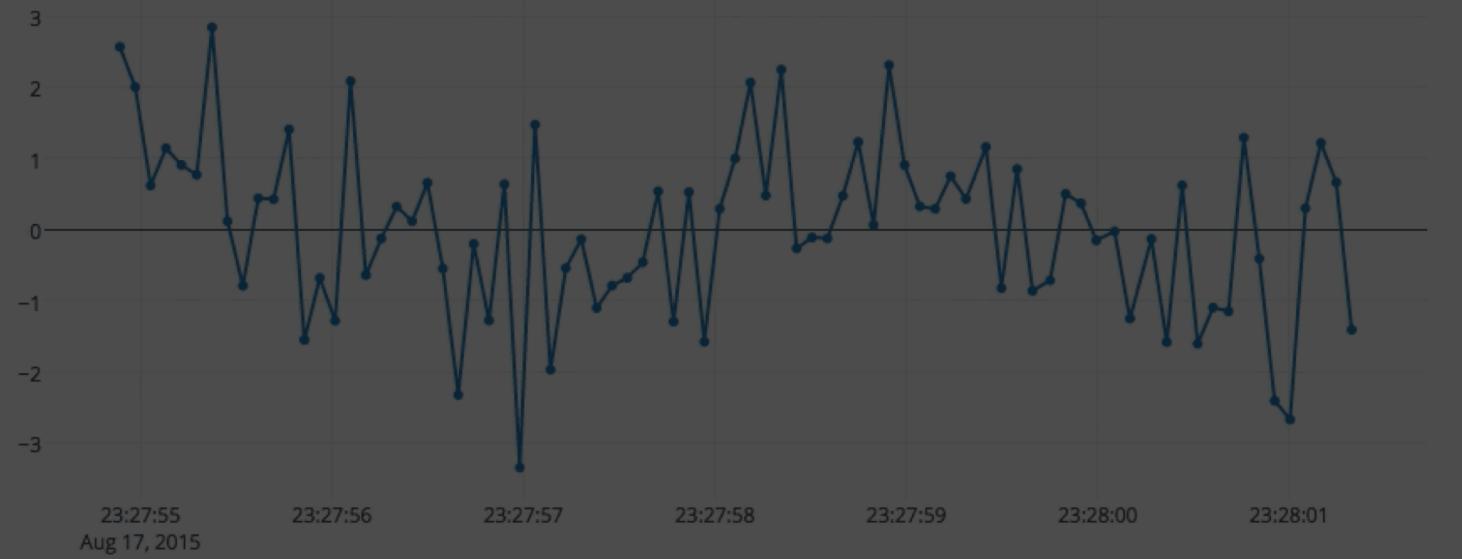
4.3.1

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

Documentation

Developer Blog

Feedback



[Edit chart »](#) - Source:

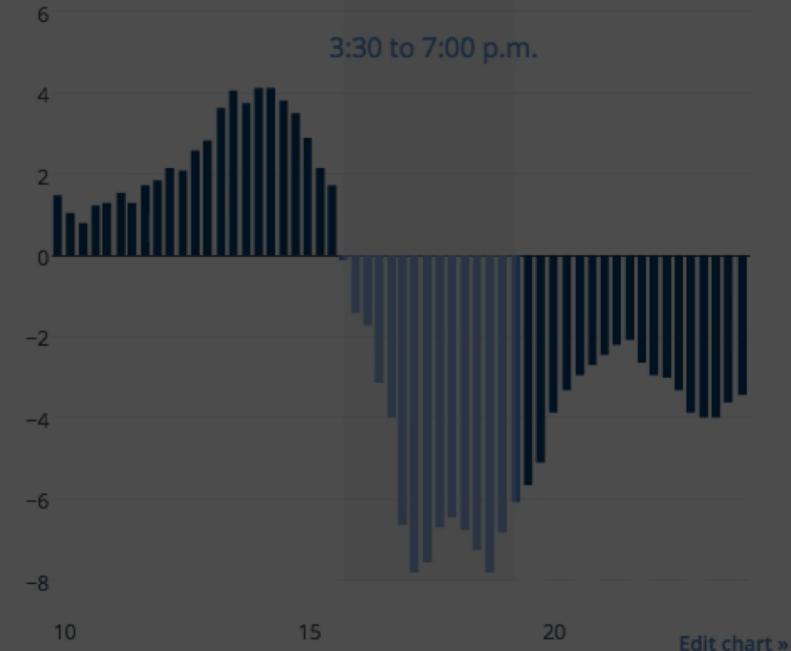
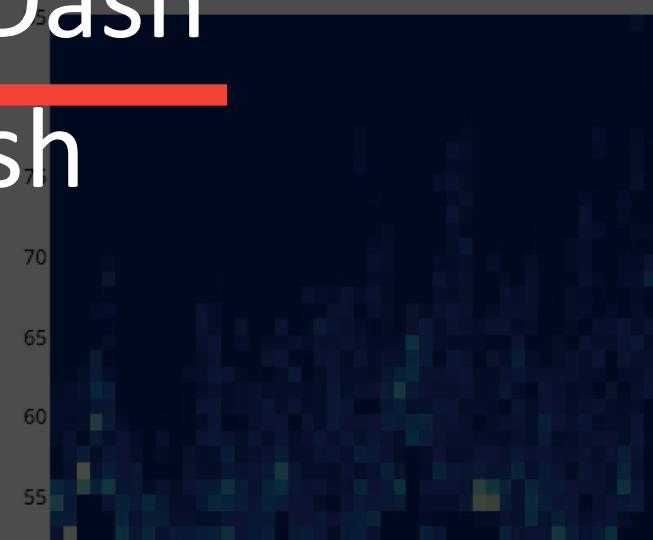
Stream live data to charts with [Python](#), [MATLAB](#), or [Node.js](#).

Python Plotly Dash

bit.ly/plotlydash

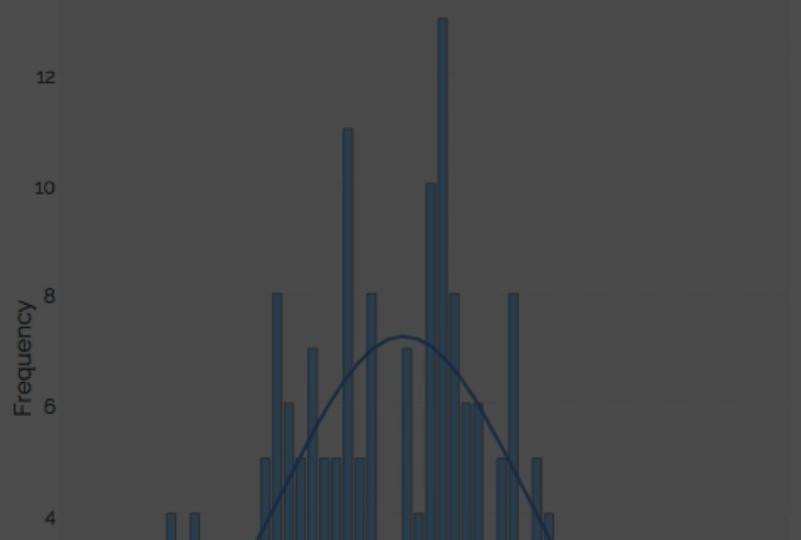


Cron Job Example



[Make a bar chart in Python](#)

Histogram Fit Example



PM Dashboard x [Secure | https://tychobra.shinyapps.io/wc-pm-dashboard/](https://tychobra.shinyapps.io/wc-pm-dashboard/)

Andy

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)

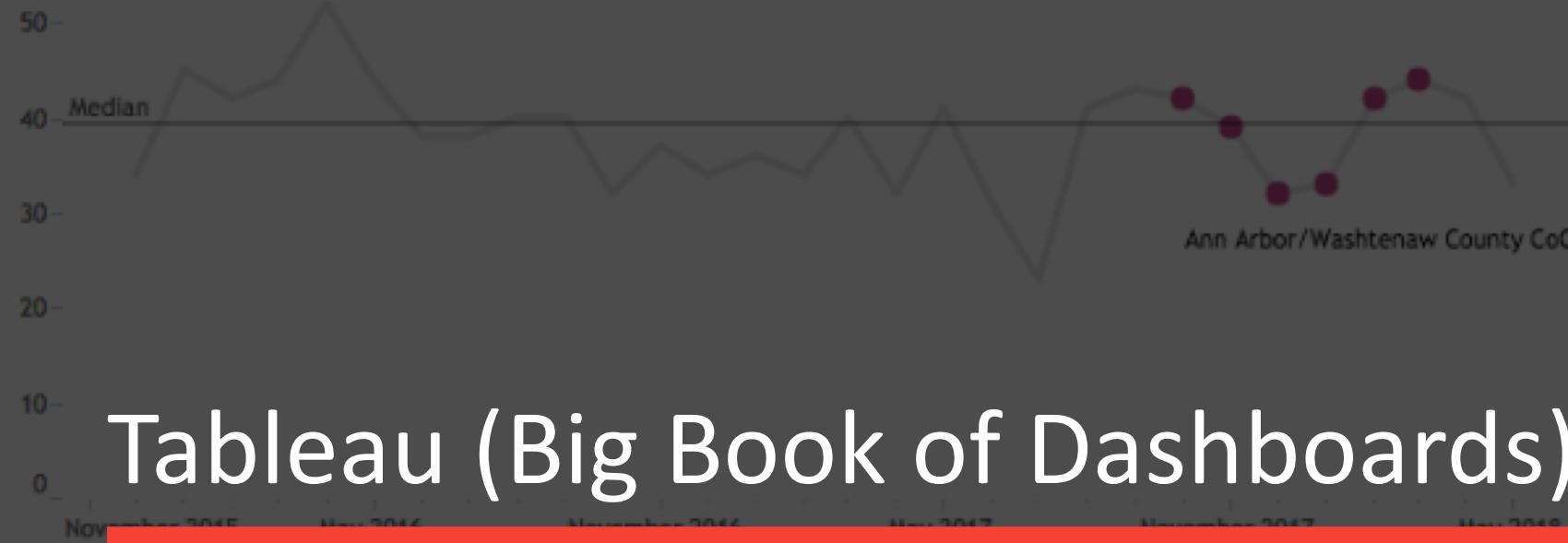
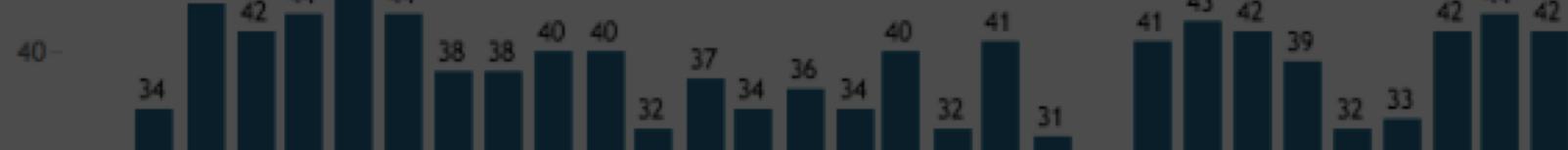


Tableau (Big Book of Dashboards)

bit.ly/bigbookdash

Actively Homeless Population

Monthly count 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

DASHBOARD FILTERS

Select Community

Ann Arbor/Washtenaw County CoC

Select Subpopulation

Veteran

(All)

Chronic

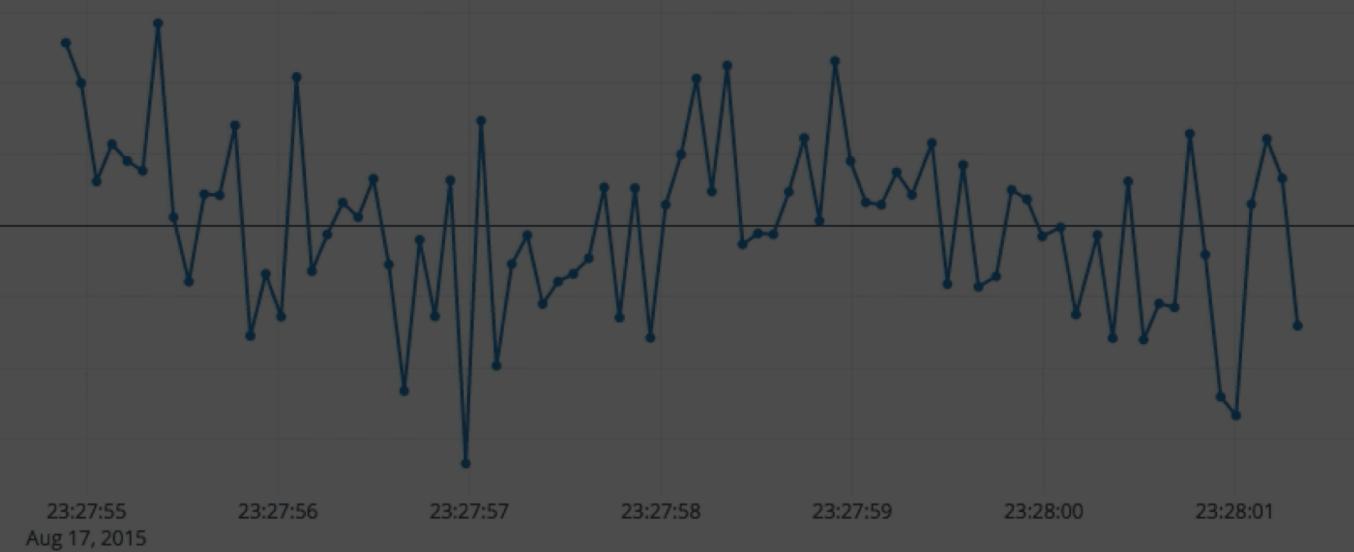
Veteran

Cancel

Apply

Select Type of Signal

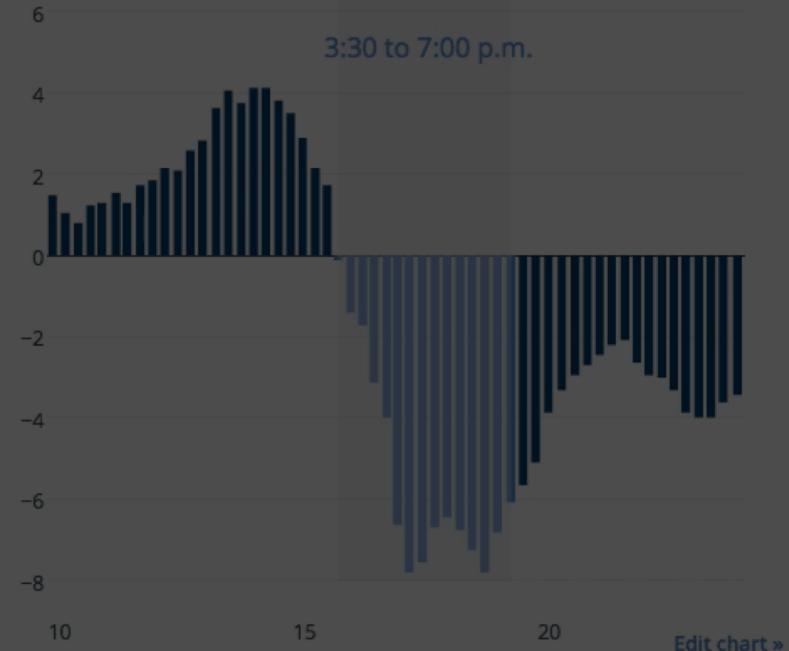
Trend



[Edit chart »](#) - Source:

23:27:55
Aug 17, 2015

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



3:30 to 7:00 p.m.

[Edit chart »](#)

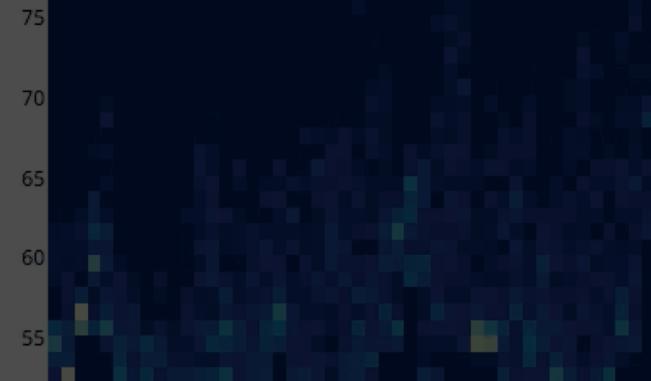
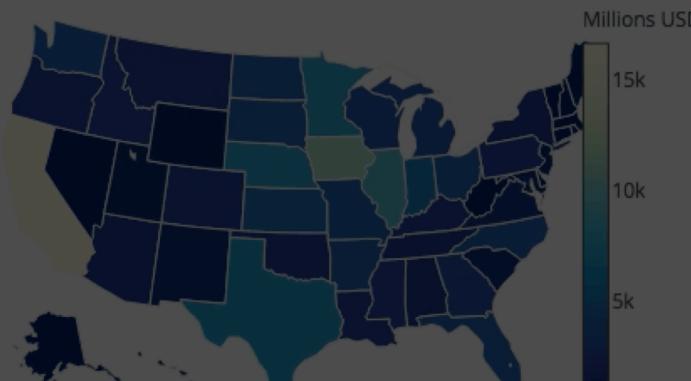
Make a bar chart in Python

[Histogram Fit Example](#)

Map Example

Cron Job Example

EXPLORE WHAT ARE GOOD GRAPHS?



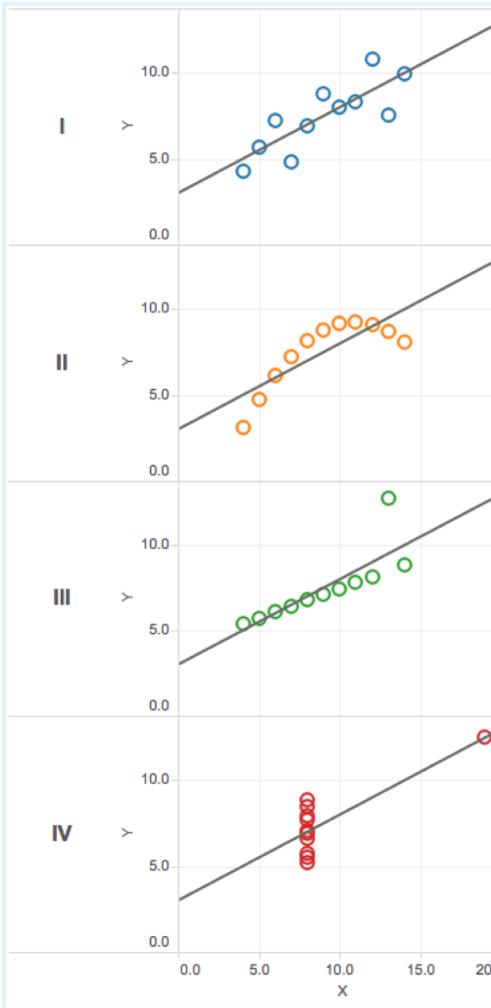
Power of Visualization

Anscombe's Quartet: The power of visualization

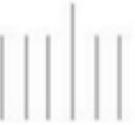
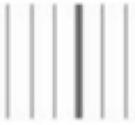
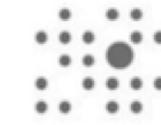
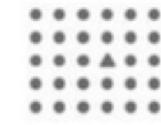
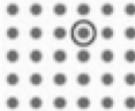
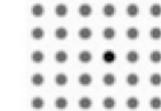
These four data sets have identical summary statistics, yet the plots show vastly different stories

I	II	III	IV
(4, 4.3)	(4, 3.1)	(4, 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, 7.3)	(7, 6.4)	(8, 6.6)
(6, 7.2)	(14, 8.1)	(8, 6.8)	(8, 6.9)
(13, 7.6)	(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, 8.8)	(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, 10.8)	(11, 9.3)	(13, 12.7)	(19, 12.5)

Summary Statistics						
Plot	sum X	sum Y	avg X	avg Y	stdev X	stdev 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

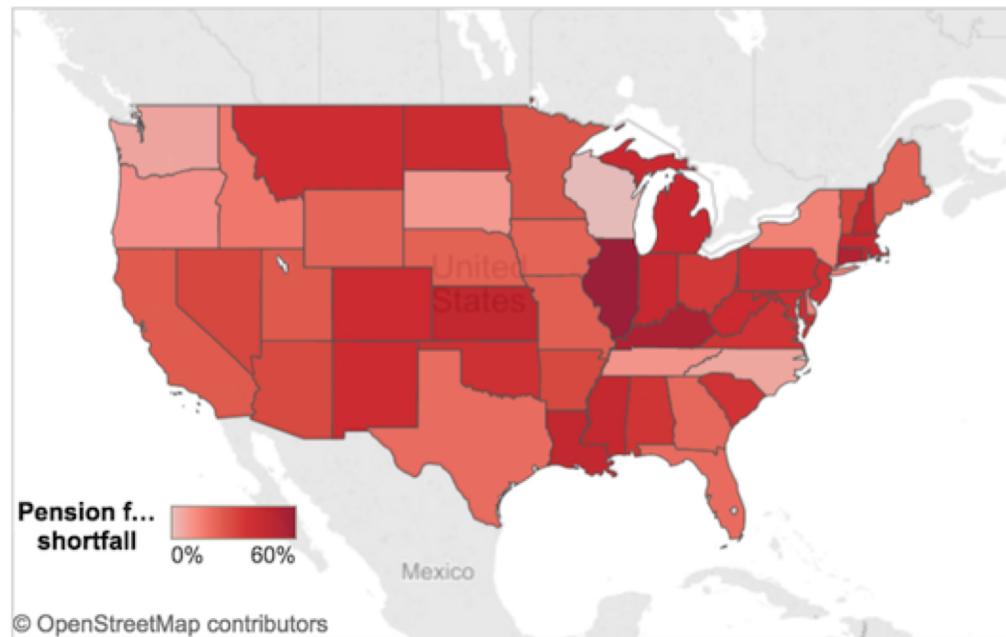
Pensions in Peril

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

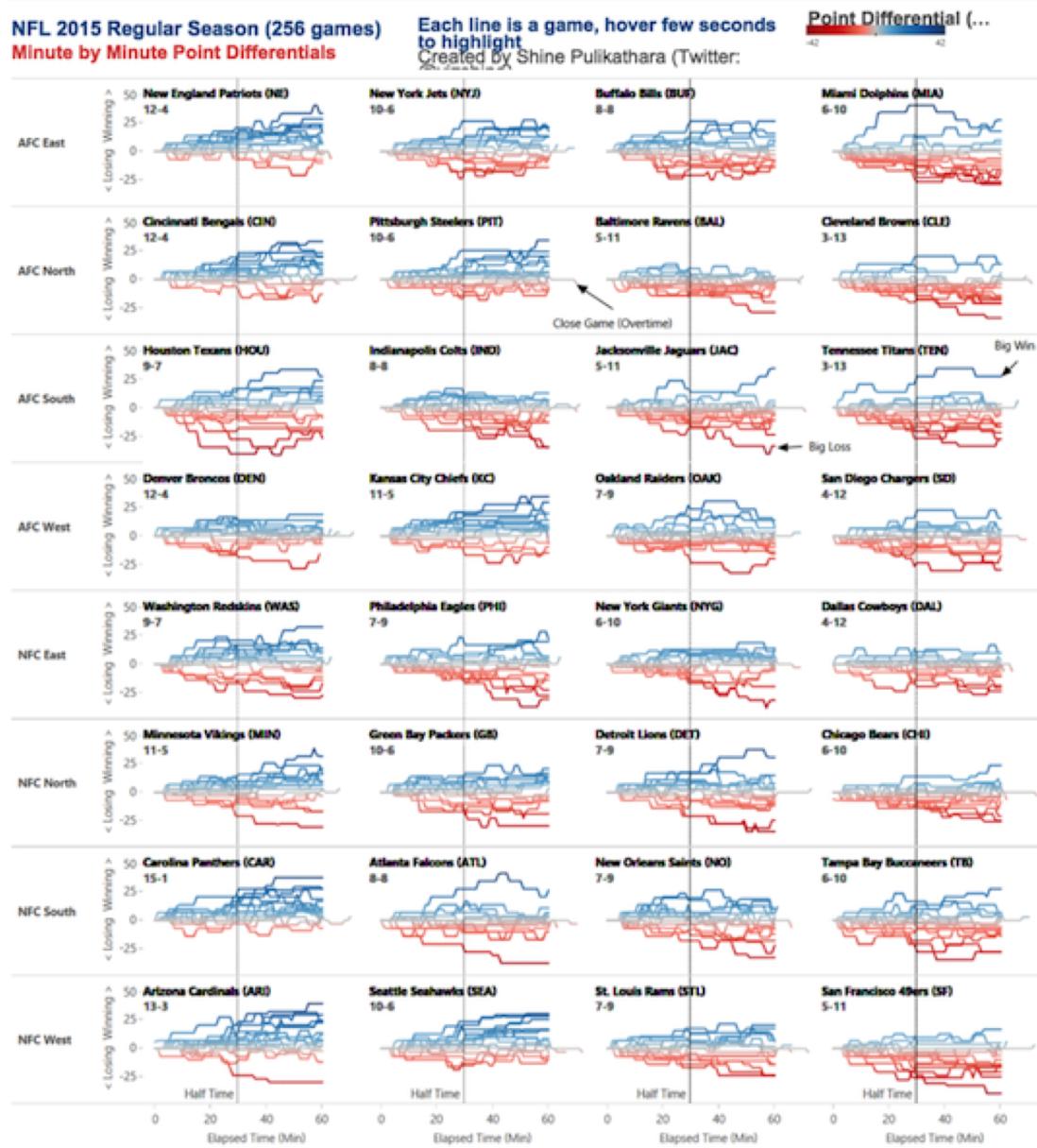


(Dropdown for AK, HI)

Contiguous US ▾



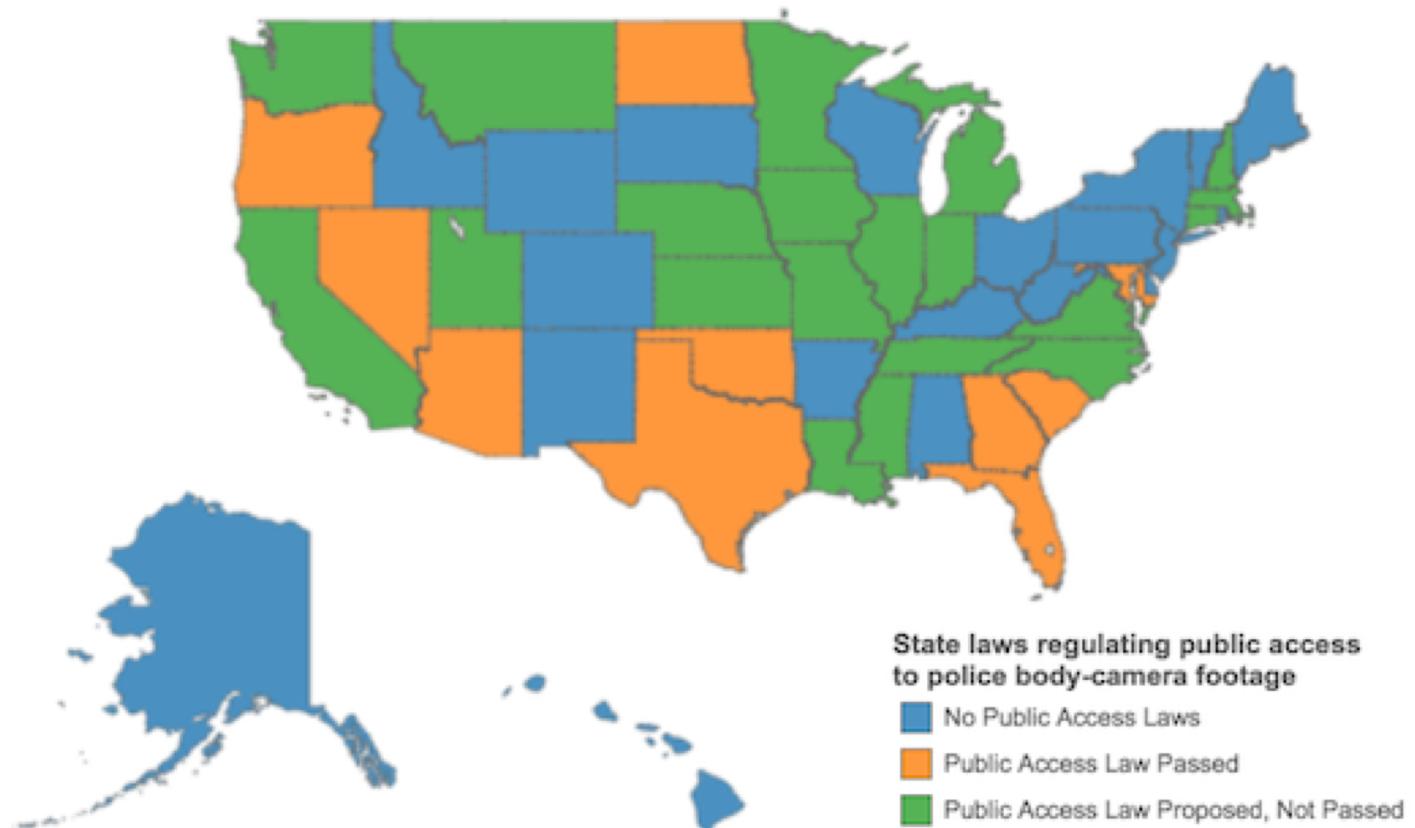
Example 2



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

Are Film Sequels Profitable?

Box Office Stats For Major Film Franchises

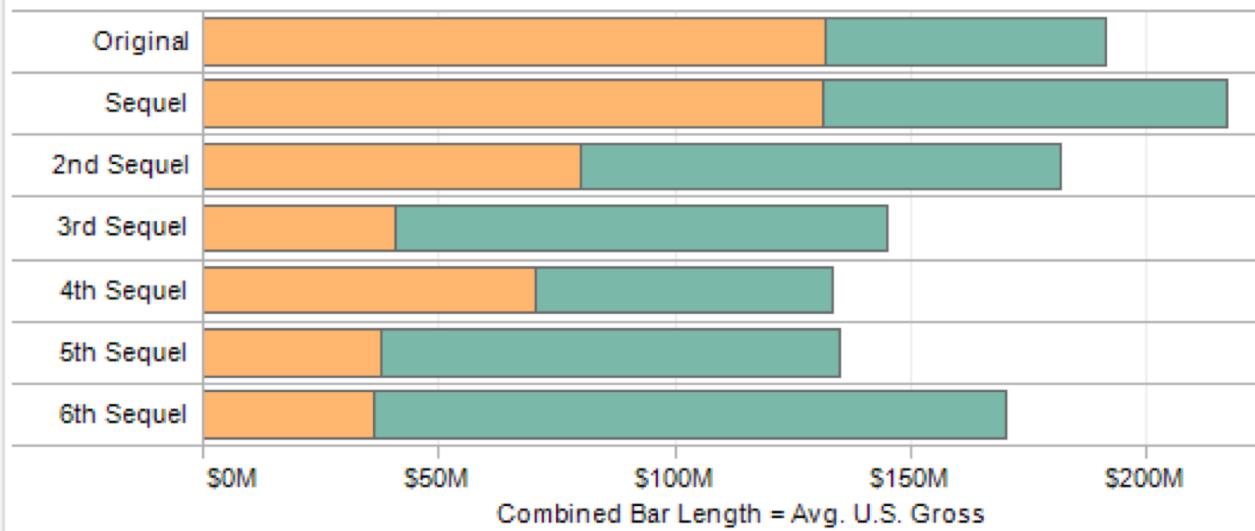
Select Movie Franchise:

(All) ▾

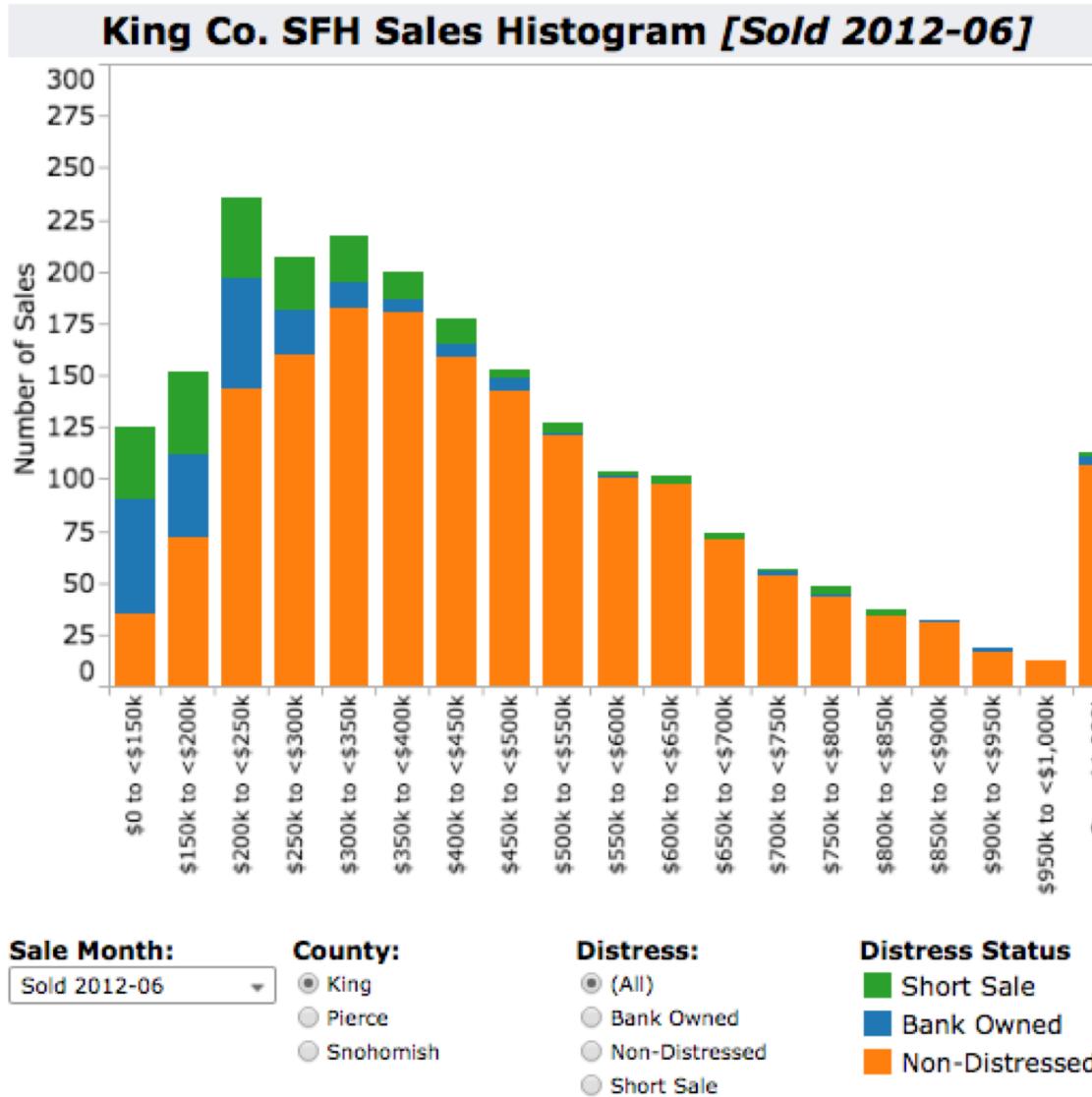
Click to Highlight Average:

Estimated Budget

Profit



Example 5





100% ▾

\$

%

.0

.00

123

▼

Arial

▼

10

▼

B

I

S

A



A

B

C

D

E

F

G

H

I

J

K

L

M

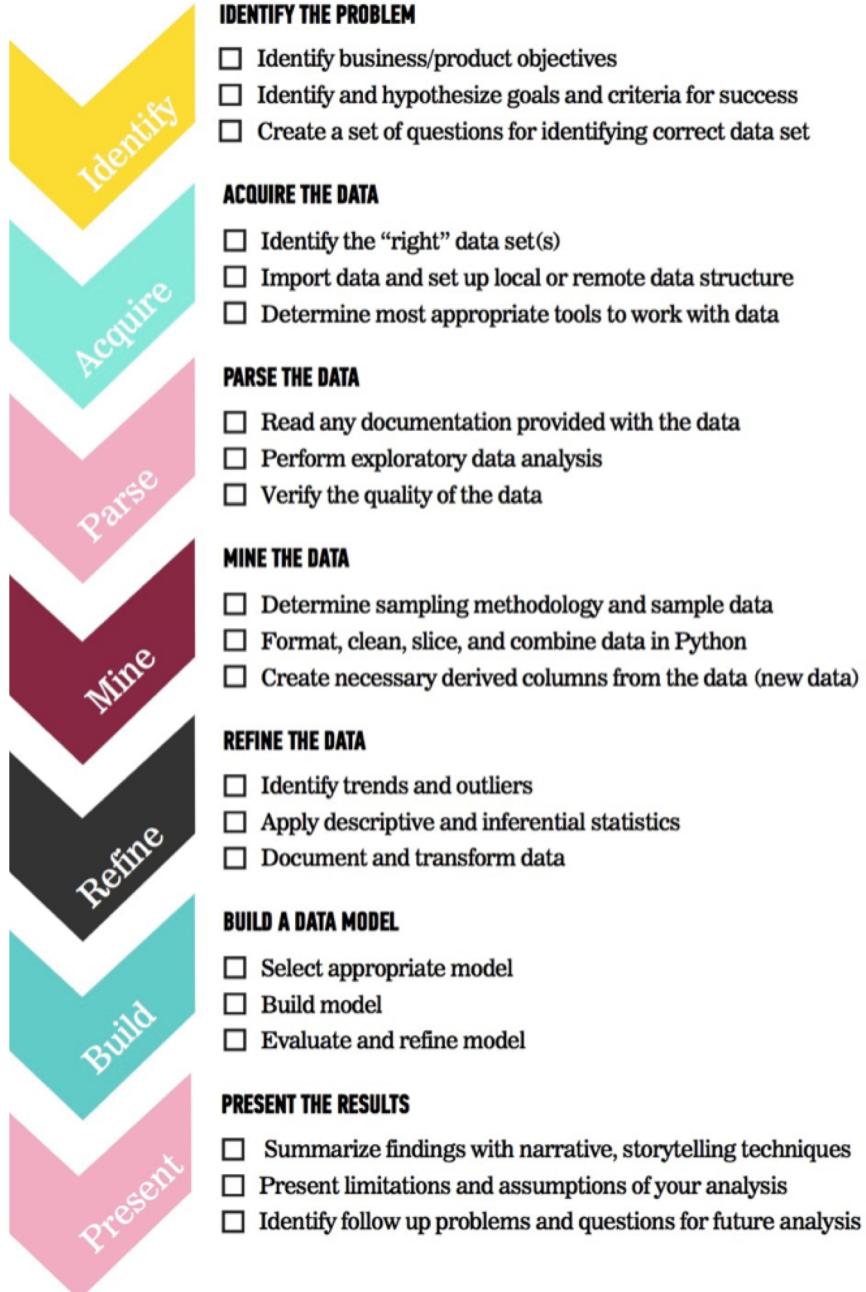
N

MAKE YOUR WORKFLOW



Continue here

DATA SCIENCE WORKFLOW



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



SHARE



100% ▾

\$

%

.0

.00

123

▼

Arial

▼

10

▼

B

I

S

A



A

B

C

D

E

F

G

H

I

J

K

L

M

N

SHARE YOUR WORKFLOW



Using Data Analytics to Understand (and Solve) Problems

David Yakobovitch
david@yakobovitch.com

