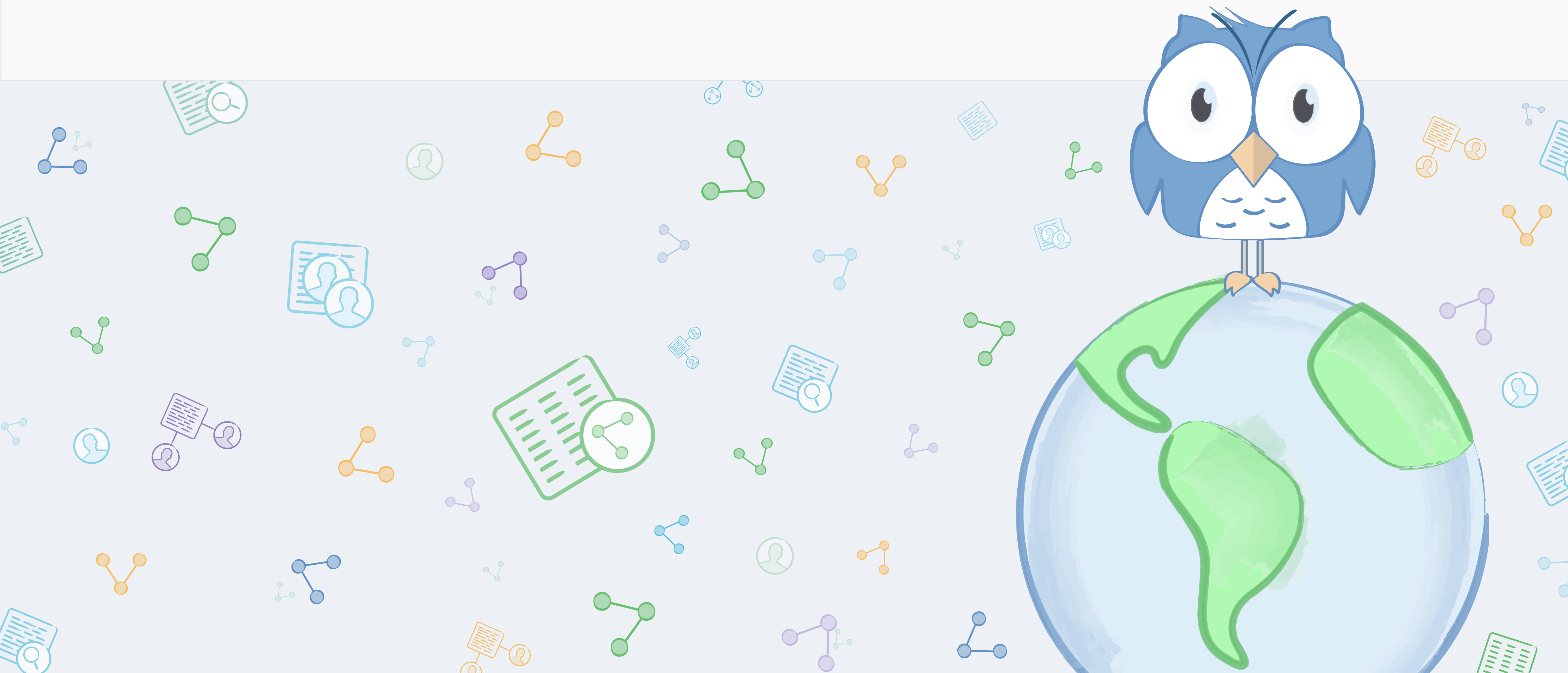


OVERVIEW





OUR PURPOSE

We're building the most **meaningful**, **collaborative**, and **abundant** data resource in the world by dismantling the barriers between data and people.

Our platform helps people who work with data solve problems faster by creating new ways to discover, prep, and collaborate.

data.world is the social network for data people.



OPEN DATA IMPACTS SOCIETY TODAY

2,400,000 websites in 1998

At least 7.5x as many open datasets today as there were websites at Google's launch

> 18,000,000 open datasets in 2016

Government:

- 1.8M open government datasets from the US, Canada, UK and Australia
- Data.gov has served millions of people over 8 years

Academia and research:

- Harvard Dataverse (73,000 datasets)
- Montreal Neurological Institute (first scientific institute to open up all results and data)

Business:

- The Weather Company (uses open government data in forecasts)
- Healthgrades.com (helps people choose providers with data from Medicaid, Medicare, state governments, more)

Citizen empowerment:

- Panama Papers (searchable database of 320,000 offshore companies and trusts)
- OpenSecrets.org (database of campaign contributions, lobbying expenditures, and more)

Source: <http://www.internetlivestats.com/total-number-of-websites/> and data.world internal analysis

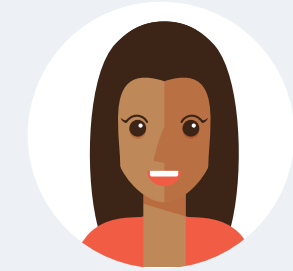
For more open data & data science facts, see our [By the Numbers](#) page

SOCIAL CHALLENGES



Will you share your data?

I am not sure that will be possible.

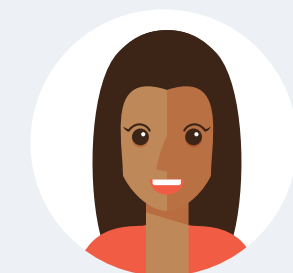


But your work was published in Science, which requires you to share you data.

I saved it on a USB drive! It is in a box at home. I just moved. There are so many boxes! I forgot to label the boxes.



— Later —



When I opened up the data, it was in hexadecimal. I cannot read hexadecimal.

You asked me for the data and I gave it to you. You will need the program that created the file, Cytosynth. The company went bankrupt in 2007.



- Mindset, reputation, and capital don't yet reward proper implementation or deep collaboration
- Researchers rewarded for patents and published articles, not datasets or collaboration
- Hoarding mentality is still prevalent
- Reproducible research movement not living up to expectations

From ["Data Sharing and Management Snafu in 3 Short Acts"](#)

Would you rather conduct a three-hour onboarding session, or send someone a link?



data.world/data4democracy/election-transparency

data.world

preview

Search

+

Add datasets

data4democracy/election-transparency

Updated Mar 10

64

Share

Download

Details

Contributors

Discussion

Launch query tool

Gathering the data from the #election-transparency channel on Data for Democracy slack.

Election Transparency

Slack: #election-transparency

Project Leads: @chris_dick, @rachelanddata, @scottcame

Maintainers (people with write/commit access):

- GitHub: @scottcame, @rachelanddata, @eric_bickel, @chris_dick
- data.world: @scottcame, @sharon, @chris_dick

GitHub: Head over to [Data4Democracy/election-transparency](#)

Project Description: This project analyzes elections in an effort to identify trends, outliers, and/or anomalies to enable insight and transparency into the democratic voting process.

Show more

Data dictionary

Aggregated metadata from this dataset's 16 tabular files and 1 other file

View data dictionary

17 files

Filter

Sort

AlaskaPrecinctBoroughMapping.csv

clean data

Request more info

Explore

Type	Column name	Empty	Distinct	Distribution
	County	0 (0%)	30	MOST COMMON 20 (361) NEXT MOST COMMON 90 (122)
	Precinct	0 (0%)	> 1,000	SAMPLE VALUE 01-110 40-330
	Year	0 (0%)	3	MOST COMMON 2010 (456) NEXT MOST COMMON 2013 (441)

Showing 1-10 of 1,335 rows, 3 columns See all

Switch to data preview

CountyCharacteristics.csv

County Characteristics File: Includes data from the Census Bureau, BEA, and BLS

Explore

Type	Column name	Empty	Distinct	Distribution
------	-------------	-------	----------	--------------

TAGS (5)

election transparency d4d

elections voting sxsw 2017

CONTRIBUTORS (19)

Data for Democracy

@data4democracy

jorge alarcon ochoa

@alarcj

Chris Dick

@chrisdick14

RECENT COMMENTS

@scottcame

last month

Thanks @evangold, we will look into it and report back here.

@evangold

last month

I was looking into stats on voter turn out, by comparing election results ...
Continue reading

View comments

QUERIES (18)

<>

AlaskaPrecinctBoroughMapping...

SQL

<>

CountyCharacteristics (CountyC...

SQL

<>

HouseElectionResults1980to201...

SQL

<>

PartyRegistration (PartyRegistra...

SQL

<>

PresidentialElectionResults1992...

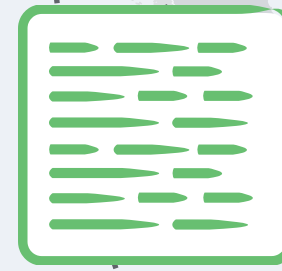
SQL

WHAT WOULD IT TAKE TO SOLVE ZIKA FASTER?



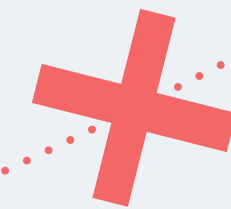
Dr. Almeida
Recife, Brazil

Trying to ID a Zika-free *Aedes aegypti* replacement. Finds that another *Aedes* mosquito is Zika-resistant, but doesn't thrive in Recife's climate. Shares data via email with colleagues.



Hidalgo Flores
Mexico City, Mexico

Analyzing data from field agents after a week of prep. Finds evidence that Zika-resistant *Aedes vexans* mosquito thrives in warm coastal climates and overtakes *Aedes aegypti*. Shares data via FTP with colleagues. Posts data to Mexico's open data portal.



Flores has found the mosquito that Dr. Almeida needs, but Dr. Almeida won't know it until she attends Flores' conference session months later.

What kept this life-saving knowledge from reaching Dr. Almeida sooner?

THE NETWORK EFFECT

What if people around the world could look for other insights in the same data, post their findings for all to see, and discuss the data in context?

Networks leverage the wisdom of the crowd.



Dr. Almeida
Recife, Brazil

What if the coastal climate data in Dr. Almeida's file was automatically matched to the coastal climate data in Flores'?

Networks detect patterns instantly and new datasets increase the value of existing datasets.



Hidalgo Flores
Mexico City, Mexico

What if Flores' data prep took an hour not a week?

Networks normalize data quickly and simplify transformation.

What if Flores knew there was a place where his research would be seen and used, where his contribution would be recognized?

Networks extend reach and enhance discovery.

THIS IS DATA'S NETWORK EFFECT, AND IT CAN HELP US SOLVE EVERY PROBLEM WE FACE, FASTER.



INTRODUCING



WITH DATA.WORLD...

- **FINDING** the data you're looking for is **EASY**
- **DISCOVERING** data you didn't know existed is **EXCITING**
- **PREPARING** data is **PAINLESS** and **QUICK**
- **COLLABORATING** and **SHARING** is **SEAMLESS**, **SECURE**, and **SOCIAL**
- **COMBINING** open data with closed data is **SAFE** and **PRODUCTIVE**

Until today, doing any one of these things well would have required a custom data warehouse, millions of dollars, and thousands of development hours.

data.world is affordable enough to put on a personal credit card and starts helping new users solve problems faster within minutes of signing up.

WE MAKE OPEN DATA MEANINGFUL

data.world links datasets together using semantic technology. This gives concepts within data an independent existence, allowing people and machines to work with data without needing to learn everything about it first.

Linked data...

- Makes data discovery easier
- Makes data interoperable
- Helps people and machines unlock data's value faster



DATA CAN BECOME WELL DOCUMENTED OVER TIME

Reliable
social proof

Methodology
Rich Summaries

Data Dictionary

Valid
Conversations
and collaboration

The screenshot displays the Data.world interface for the dataset 'data4democracy/election-transparency'. The page includes a header with the dataset name, update date (Mar 10), and an 'OPEN' button. Below the header, there are tabs for 'Details', 'Contributors', and 'Discussion'. The main content area features a description of the dataset, project leads, maintainers, and a project description. A 'Data dictionary' section is visible, showing aggregated metadata from 16 tabular files and 1 other file. The 'Data dictionary' table lists columns: County, Precinct, and Year, with their respective data types, empty values, distinct values, and distributions. A 'Data dictionary' button is present. The right sidebar shows contributors, recent comments, and queries. Annotations highlight specific features: a blue circle around the '64' likes and 'Share' button indicates 'Reliable social proof'; a blue circle around the 'Data dictionary' button indicates 'Methodology Rich Summaries'; a blue circle around the 'Data dictionary' section indicates 'Data Dictionary'; and a blue circle around the 'RECENT COMMENTS' section indicates 'Valid Conversations and collaboration'.

data.world

data4democracy/election-transparency **OPEN**

Updated Mar 10

Details Contributors Discussion

Gathering the data from the #election-transparency channel on Data for Democracy slack.

Election Transparency

Slack: #election-transparency

Project Leads: @chris_dick, @rachelanddata, @scottcaine

Maintainers (people with write/commit access):

- GitHub: @scottcaine, @rachelanddata, @eric_bickel, @chris_dick
- data.world: @scottcaine, @sharon, @chris_dick

GitHub: Head over to Data4Democracy/election-transparency

Project Description: This project analyzes elections in an effort to identify trends, outliers, and/or anomalies to enable insight and transparency into the democratic voting process.

Show more

Data dictionary

Aggregated metadata from this dataset's 16 tabular files and 1 other file

View data dictionary

Filter Sort

AlaskaPrecinctBoroughMapping.csv

clean data Request more info

Type	Column name	Empty	Distinct	Distribution
County	County	0 (0%)	30	MOST COMMON 20 (361) NEXT MOST COMMON 90 (122)
Precinct	Precinct	0 (0%)	> 1,000	SAMPLE VALUE 01-110 40-330
Year	Year	0 (0%)	3	MOST COMMON 2010 (456) NEXT MOST COMMON 2013 (441)

Showing 1-10 of 1,335 rows, 3 columns See all

Switch to data preview

CountyCharacteristics.csv

County Characteristics File: Includes data from the Census Bureau, BEA, and BLS

Explore

RECENT COMMENTS

@scottcaine • last n
Thanks @evangold,
it and report back h

@evangold • last mo
I was looking into
out, by compar

Continue

QUERIES (18)

<> AlaskaPrecinctBoroughMapping.csv SQL

<> CountyCharacteristics.csv SQL

<> HouseElectionResults1980to201... SQL

<> PartyRegistration (PartyRegistra... SQL

<> PresidentialElectionResults1992... SQL

Joins from clinicaltrials.gov, cancer.gov & census.gov to examine cancer trials, mortality, incidence and demographics

Intelligence of dogs by breed

The Dollar-And-Cents Case Against Hollywood’s Exclusion of Women

178,000 WWII air missions (digitized paper mission reports by location, mission and more than 60 other attributes)

The most comprehensive Product Hunt API dataset ever released

4 million Facebook posts from mainstream media sources

Complaints received by Bureau of Consumer Financial Protection about financial products and services from 2012-2016


Hate crime laws and data from the ADL and FBI, 1997-2015

Curated collection of 353 data visualization tools, resources and technologies

Using smartphones to collect movement patterns in Parkinson’s patients

TOP DATASETS SHOW A SPECTRUM OF INTERESTS

Hate Crime Laws and Statistics




@dkelley

CONTRIBUTOR

• 3 months ago

It looks like the 2015 anti-Islamic crimes are second only to the

Like




@mlieberman

CONTRIBUTOR

• 3 months ago

What do you think is the connection between a 67% increase in

Like



@dkelley


CONTRIBUTOR

• 3 months ago

Looks like every year race based crimes are the most numerous

numerous within that.

Like



@albert


CONTRIBUTOR

• 3 months ago

randomly answered a question on open data se about this, if its

<http://opendata.stackexchange.com/questions/10088/hate-cr>

Like



@gswider


CONTRIBUTOR

• 3 months ago

thanks, @albert! I just posted the link to this workspace on the

dataset as a resource as well

Like




@ian

CONTRIBUTOR

• 3 months ago

@albert was helpful and much appreciated

World War II THOR Data



@acu192

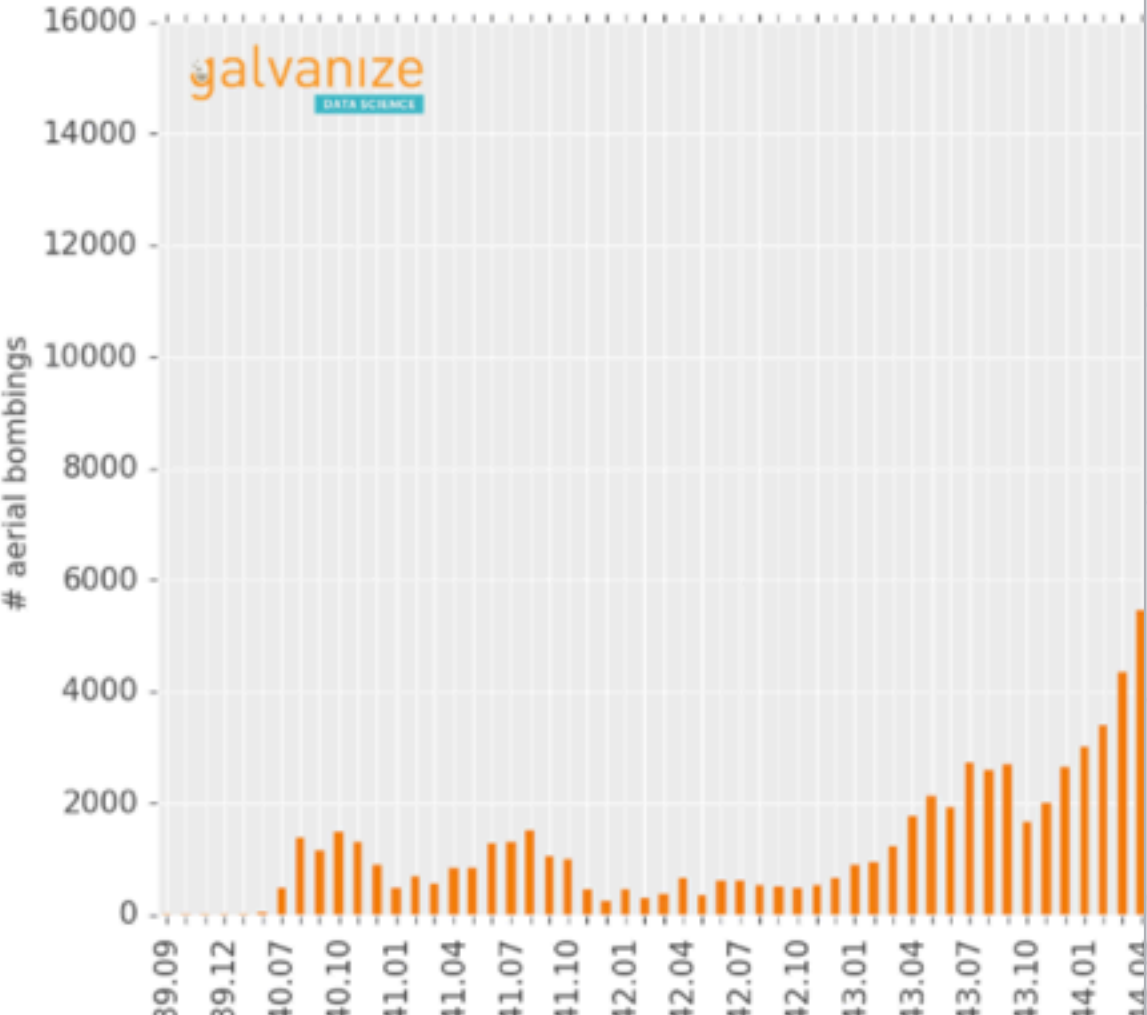
CONTRIBUTOR

• 2 months ago • Edited


Not too surprising, but things heated up (um... literally) toward the end of the war. For

Tuesday, 1944-06-06. See the chart below:

Timeline of WWII Aerial Bombing




galvanize DATA SCIENCE



@datamil

@datamil


Owner since Nov 22



Ryan Henning

@acu192


Contributor since Dec 13



Jonathan Ortiz

@databeats


Contributor since Nov 29



Ian Greenleigh

@ian


Contributor since Nov 24



Kuang Keng

@kuangkeng


Viewer since Nov 30



marylazzari

@marylazzari

Viewer since Dec 7




Noah Rippner

@noahrippner

Contributor since Nov 30

RECENT COMMENTS




@wojtekpajak

• last month

Could you please add something like 'by allied planes' in the title of ...

Continue reading




@tle

• last month

@ian, btw this is the sparql query that let you discover those IDs [http...](http://...)

Continue reading




@ian

• last month

Have you seen this story? [German city of 54,000 is evacuated after the](#)

...

Continue reading




@ian

• 2 months ago

@hhaveliw Good stuff! It's a nice visual summary of the data. Thanks for ...

Continue reading




@hhaveliw

• 2 months ago

Hey guys I couldn't post an image in this comment. But here you go, a

Intelligence of Dogs




@mara

CONTRIBUTOR

• 2 months ago

@albert Awesome! Yeah, I think this is indeed the place for it (hopefuly

Like




@albert

CONTRIBUTOR

• 2 months ago

@mara cool. done.

Like



@sya


CONTRIBUTOR

• 2 months ago

@albert @mara

Using Tabula I converted a couple of the tables on Heterozygosity in dataset.

Like




@albert

CONTRIBUTOR

• 2 months ago

cool! sorry but i have to get up on my stool here and say that tabula is users. appreciate the effort though.

Like




@dash

• 2 months ago

Hi @albert. Tabula is available on GitHub and released under the MIT another tool. [https://github.com/tabulapdf/tabula/blob/master/](https://github.com/tabulapdf/tabula/blob/master/README.md)

Like



@sya

CONTRIBUTOR

• 2 months ago

It's an open-source [Knight Prototype Fund](#) project, @albert.

USER STORIES

Enhancing data education and knowledge creation

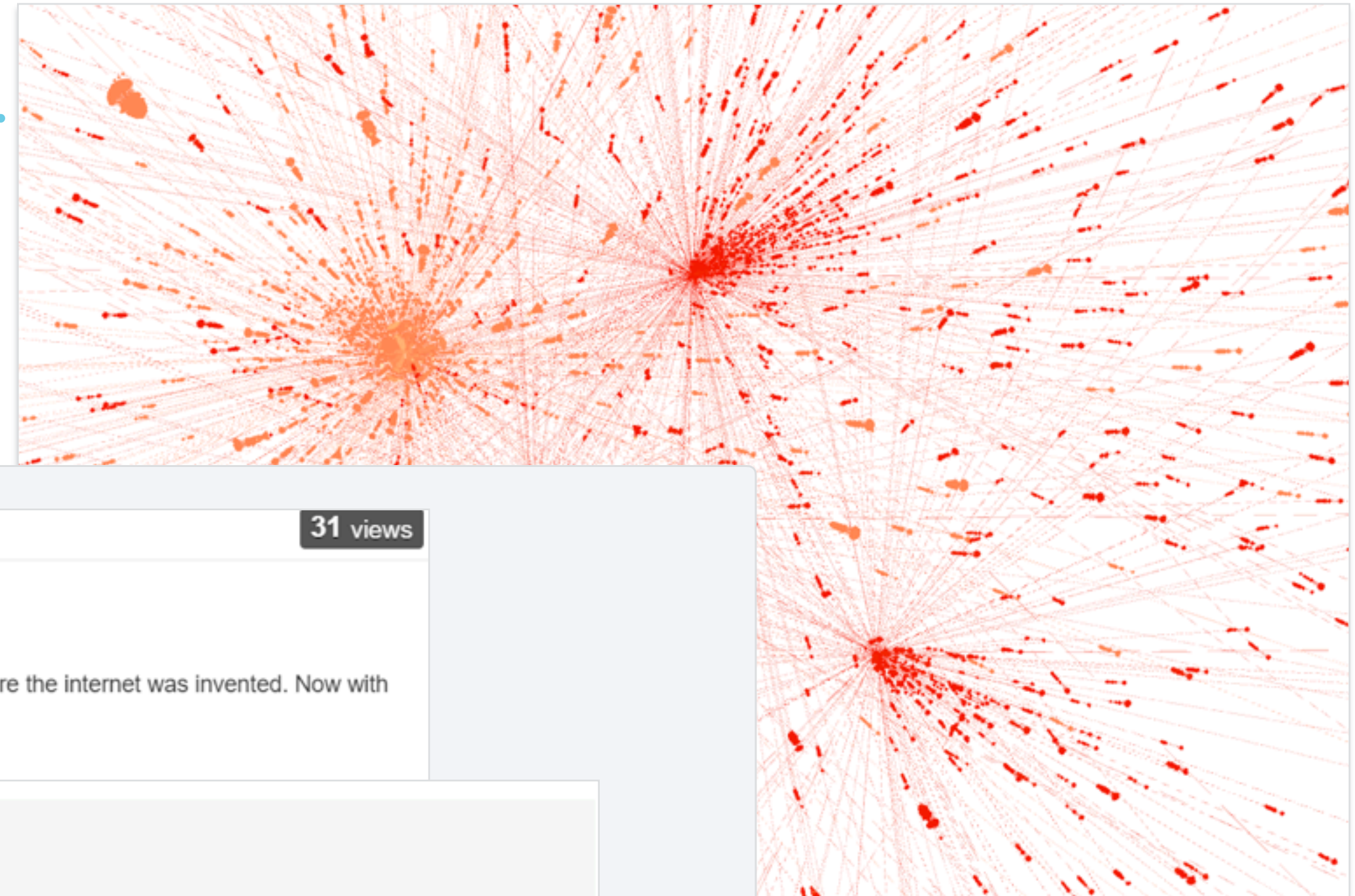


Julie Schellberg uses data from data.world in her predictive analytics tutorial.

I have created a DXP called [lending_club_loans \(2\)](#) to accompany this post using data from [Data.World](#) on [Lending Club](#) loans. I have dabbled a bit with Lending Club myself, so I found this data set particularly interesting.



Vadim Markovtsev publishes metadata on 452,000,000 GitHub commits to data.world, and uses the dataset to prove the famous “Six Degrees of Separation” theory.



University of Texas CS professor Dr. Philip Cannata uses data.world extensively in his course Data Visualization.

note ★

31 views

Here's a link to the paper I was talking about last class

- [The Irresistible Move towards Interoperable Database Systems](#) - this is a paper I wrote and delivered at a conference in 1991 before the internet was invented. Now with data.world, most all of it is starting to come true.
- I put the link under the Week 1 class notes, however, because I wanted it to be next to the data.world link.

#pin

class_notes

Resolved

Unresolved

Wendy He 16 hours ago
How cool!
"If you can call anyone in the world from anywhere in the world, shouldn't you be able to locate and manipulate any relevant data in the world from anywhere in the world?"

Philip Cannata 7 hours ago From this article:
"In short, data.world makes data discoverable, interoperable, and social. And that could mean an explosion of data-driven insights is at hand."

That's exactly what I was talking about and working on at MCC (Austin's first high tech software company) in the early 90s.

Reply to this followup discussion

A NEW KIND OF COMPANY



As a public benefit corporation, data.world is legally structured to pursue its public benefit mission and **shareholder value with equal vigor through growth, IPO, and beyond**. This rare, powerful corporate structure allows data.world to make a profound impact by democratizing access to the world's data and creating a sustainable archive. data.world's specific public benefit purpose can be found at data.world/benefit-statement.

data.world is also a Certified B Corporation, a prestigious designation awarded to over 2,000 companies globally that meet the highest standards of overall social and environmental performance, transparency and accountability and aspire to use the power of business to solve social and environmental problems.

Top-5 VC-backed Certified B Corporations in tech (by funding)

THANK YOU

Carolee Mitchell
carolee@data.world



data.world