

Case Study 1 - International Conflicts Dashboard

Client: You have been approached by the United Nations in New York City to create an interactive dashboard for their internal department that studies world-wide conflicts. These are political scientists that are well equipped to analyze the data, but that do not have the visualization skills to create an effective dashboard.

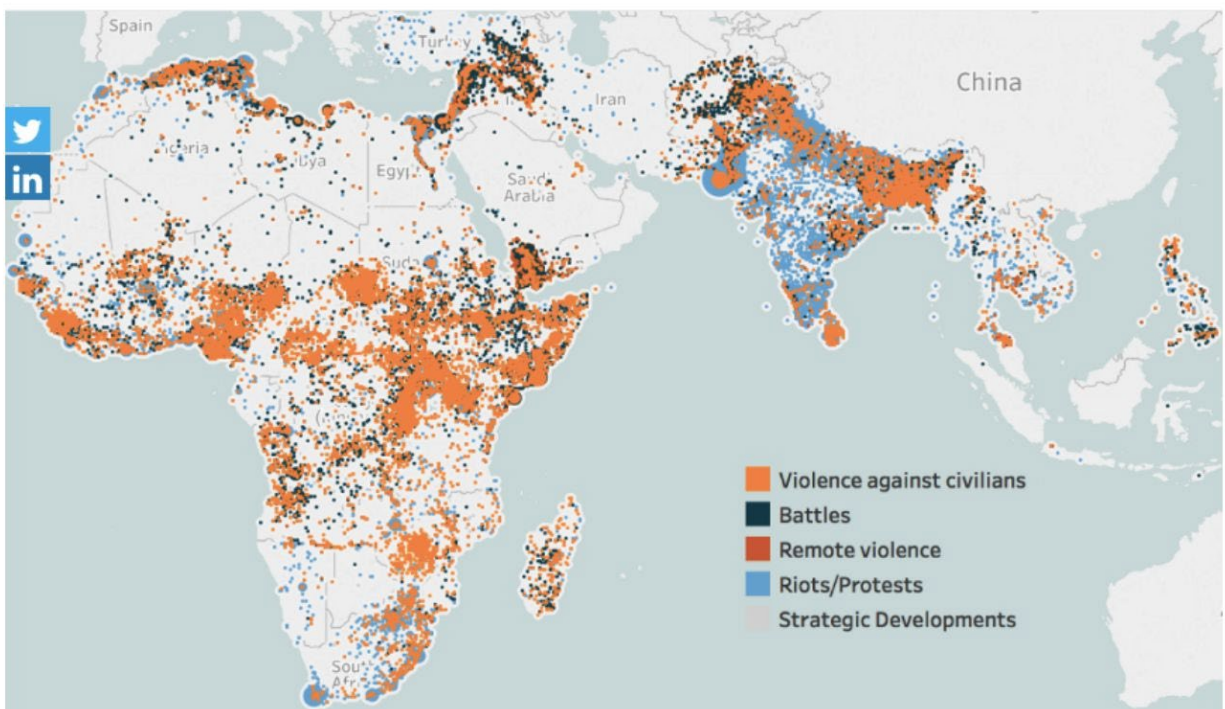
Goal: Create an interactive dashboard that gives an overview of world-wide conflicts and that allows expert users to drill down into the details of individual conflicts for exploratory data analysis. Your task is to highlight interesting facets of the data and provide effective filtering techniques to facilitate exploration.

The main criteria for your dashboard are:

- Convey as much of the information in the dataset as possible
- Allow expert users to interactively explore the data
- Give an overview and enable drill-down into details

Previous visualization: An intern has create this initial static visualization, but the UN is looking for a more sophisticated interactive dashboard.

Original Visualization



Data: [UN-conflicts.csv](#)

Data property description: [ACLED codebook](#)

Data fields:	Example
Iso: integer	466
Event_id_cnty: string	MLI2170
Event_id_no_cnty: integer	2170
Event_date: date	2018-07-28
Year: date	2018
Time_precision: integer	1
Event_type: string	Violence against civilians
Actor1: string	FLM: Macina Liberation Movement
Assoc_actor_1: string	[no data]
inter1: integer	3
Actor2: string	Civilians (Mali)
Assoc_actor_2: string	[no data]
Inter2: integer	7
Interaction: integer	37
region: string	Western Africa
country: string	Mali
admin1: string	Mopti
admin2: string	Mopti
admin3: string	Fatoma
location: string	Koundioli
Latitude: decimal	14.595
Longitude: decimal	-3.8739
Location_2: geopoint	POINT(-3.8739 14.595)
Geo_precision: integer	1
Source: string	Whatsapp
Source_scale: string	other
Notes: string	On July 28, presumed Katiba Macina militants abducted the village chief of Koundioli.
Fatalities: integer	0

Case Study 2 - Volcano Eruptions Website

Client: The National Geographic society hired you to create interactive visualizations on their website about volcano eruptions. The dataset contains information on every volcano eruption in the world since 1883.

Goal: Create prototypes of engaging interactive visualizations that are aimed at the interested general public. National Geographic wants to convey the facts and numbers, however, the visualizations need to be engaging enough to capture the interest of their typical audience.

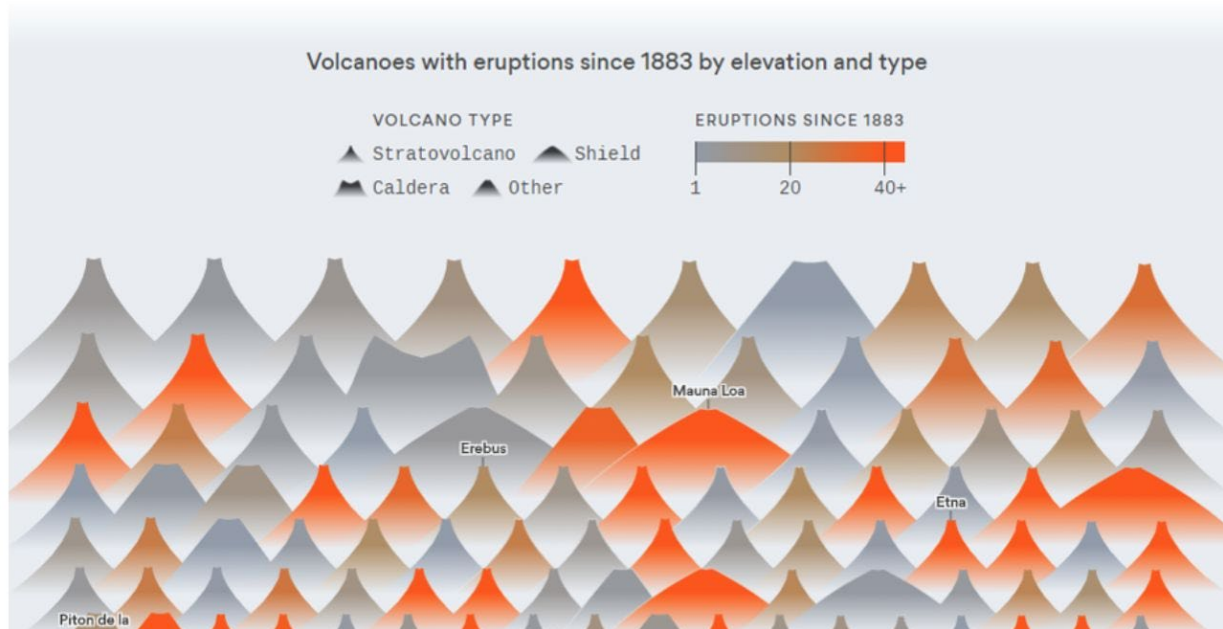
The main criteria for your visualizations on their website are:

- Accurately convey the facts about volcano eruptions in the data
- Well designed and engaging, ideally revealing some surprising facts
- Appropriate messaging with interesting titles, captions, and annotations

Previous visualization: A design consultant created an initial prototype shown below. While visually pleasing and engaging, it does not support interactive filtering and may be a bit too abstract for an interested reader.

Original Visualization

Here's every volcano that has erupted since Krakatoa



Data: [volcano_eruptions.xlsx](#)

Data fields:

Volcano_number: integer
Volcano_name: string
country: string
primary_volcano_type: string
activity_evidence: string
last_known_eruption: string
region: string
subregion: string
Latitude: decimal
Longitude: decimal
Location: geopoint
Elevation_m: integer
Dominant_rock_type: string
Tectonic_setting: string

Examples:

211020
Vesuvius
Italy
Stratovolcano
Eruption observed
1944 CE
Mediterranean and Western Asia
Italy
40.821
14.426
POINT(14.426 40.821)
1281
Phono-tephrite / Tephri-phonolite
Subduction zone / Continental crust (>25 km)

Case Study 3 - Health Data in NYC

Client: The Health Department of the city of New York wants to publish interactive visualizations on their website on rat sightings as part of their annual health report to the public. The commissioner on rodents and marsupials has pushed for this for years and created a large dataset. She created an initial visualization, however, she wants something more informative than just a scatter plot.

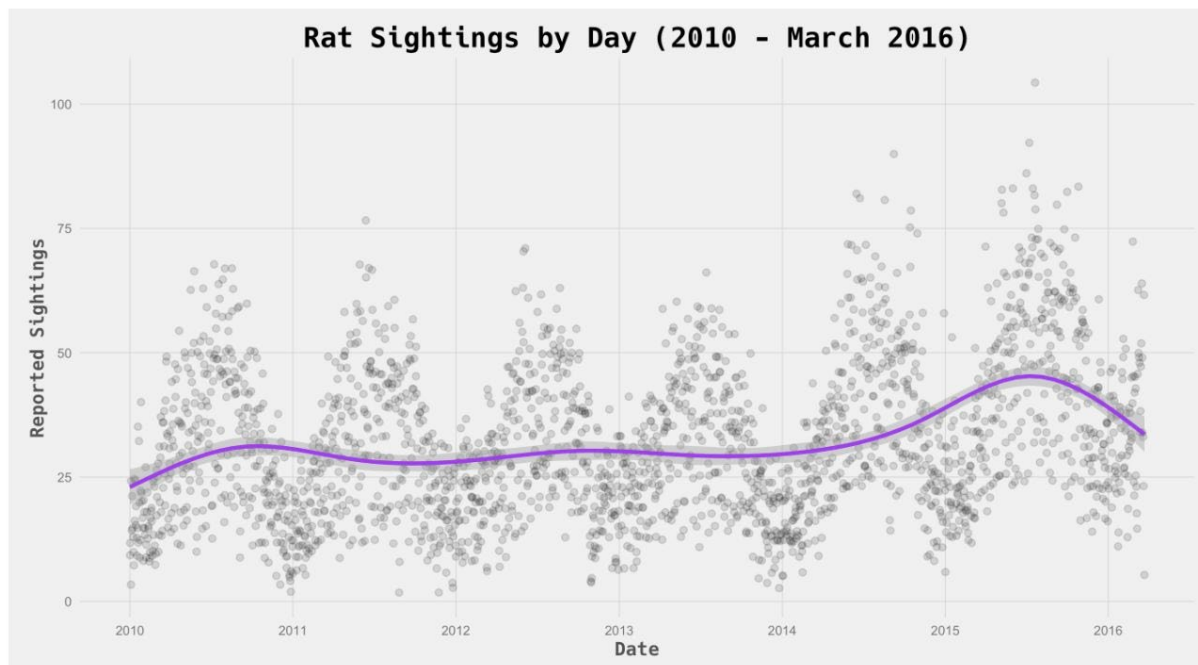
Goal: Create several interactive visualizations that show where rat sightings are most common. Users should be able to explore the dataset, but you can choose which aspects you want to highlight in your messaging.

The main criteria for your visualizations on their website are:

- Accurately convey the facts about rat sightings in the data
- Highlight aspects of the data that might be interesting or surprising
- Informative titles, captions, and annotations

Previous visualization: Created by the commissioner herself, this initial visualization shows the average number of rat sightings over time. However, it is a bit unprecise (e.g., average number of sightings per week/day/month?) and should only serve as an initial starting point.

Original Visualization



Data: [rat_sightings.csv](#)

Data fields:

Address_type: string
agency_name: string
agency: string
borough: string
city: string
Closed_date: datetime
community_board: string 10
complaint_type: string
Created_date: datetime
cross_street_1: string
cross_street_2: string
descriptor: string
Due_date: datetime
facility_type: string
incident_address: string
Incident_zip: integer
intersection_street_1: string
intersection_street_2: string
landmark: string
Latitude: decimal
location_type: string
Longitude: decimal
Location: geopoint
park_borough: string
park_facility_name: string
Resolution_action_updated_date: datetime
status: string
street_name: string
Unique_key: integer

Examples:

ADDRESS
Department of Health and Mental Hygiene
DOHMH
MANHATTAN
NEW YORK
2013-08-30T00:00:00
MANHATTAN
Rodent
2013-05-09T00:00:00
WEST 113 STREET
WEST 114 STREET
Rat Sighting
2013-05-10T16:36:00
N/A
2098 8 AVENUE
10026
[no data]
[no data]
[no data]
40.80279815
3+ Family Apt. Building
-73.95654015
POINT(-73.95654015 40.80279815)
MANHATTAN
Unspecified
[no data]
Pending
8 AVENUE
26264065

Case Study 4 - Flight delay in the U.S.

Client: You have been granted a research fellowship from the U.S. Department of Transportation to develop an interactive dashboard for analyzing flight delays. The government decision makers will use your system to explore the collected data. They are familiar with the terminologies in the data and skilled at statistics.

Goal: Your client is aiming to use the tool to review their existing policies and propose new ones if necessary. They also intend to use the tool to help them decide the annual budget. Your task is to create an interactive dashboard that allows your client to easily confirm their hypothesis and identify outliers.

The main criteria for your interactive exploration tool are:

- Convey as much of the information in the dataset as possible
- Allow the user to easily filter the data with various of conditions
- Allow the user to easily identify outliers

Previous Visualization: The U.S. Department of Transportation has an existing online tool. However, they want a better visual tool to support the filtering and outlier identification process.

Original Visualization

Airline On-Time Statistics and Delay Causes

[Delay Cause Definition](#) [Understanding Delay Data](#) [Database Tables](#) [Flight Delays at a Glance](#)

Select a carrier: [Show all reporting carriers](#) Select an airport: [Show all airports \(by state\)](#) Period from: Period to:

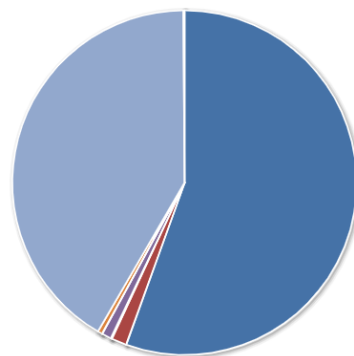
On-Time Arrival Performance National (April, 2020)

[Most Recent Month](#) [Year to Date](#)

[View Tabular Version](#) [Download Raw Data](#)

More Topics:

- [Flight Delays by Cause](#)
- [Weather's Share of Delayed Flights](#)
- [Weather's Share of National Aviation System \(NAS\) Delays](#)
- [National Aviation System \(NAS\) Delay by Cause](#)



- On Time: 55.51%
- Air Carrier Delay: 1.44%
- Weather Delay: 0.09%
- National Aviation System Delay: 0.91%
- Security Delay: 0.01%
- Aircraft Arriving Late : 0.43%
- Cancelled: 41.51%
- Diverted: 0.1%

Image from Bureau of Transportation Statistics Website

Data: [airline_delay_causes.csv](#)

Data fields:

Date: date
Carrier Code: string
Carrier: string
Airport Code: string
Airport: string
Total Operations: integer
Arrivals Delayed: integer
Carrier Delays: decimal
Weather Delays: decimal
NAS Delays: decimal
Security Delays: decimal
Late Aircraft Delays: decimal
Arrival Cancelled: integer
Arrival Diverted: integer
Arrival Delay Minutes: integer
Carrier Delay Minutes: integer
Weather Delay Minutes: integer
NAS Delay Minutes: integer
Security Delay Minutes: integer
Late Aircraft Delay Minutes: integer

Examples:

2/1/2008
AS
Alaska Airlines Inc.
BOS
Boston, MA: Logan International
58
6
2.5
0
3.5
0
0
0
1
616
368
0
248
0
0

A special note:

Carrier Delays, Weather Delays, NAS Delays, Security Delays, and Late Aircraft Delays are recorded as decimals. This is because a flight delay can have multiple types, and the decimal values are reflecting partial reasons. For example, if a flight was 30 minutes late, 15 minutes due to weather and 15 minutes due to security it will contribute .5 to *Arrivals Delayed* and .5 to *Security Delays*.

Case Study 5 - Pharmacy supply chain

Client: You recently joined a team working on optimizing the supply chain of a pharma company. The team has access to the order history of a certain period (including orders from competitors). Your job is to create an interactive dashboard for the team to explore the data, form hypotheses, and to propose possible improvements.

Goal: Create an interactive dashboard that gives an overview of historical demand in the national regions, as well as in state and zip code level resolutions. The user should be able to filter the data in various ways. The team also intends to use the dashboard to identify cases where there are obvious shipping delays, outliers in profit and regions at disadvantage comparing to competitors.

The main criteria for your dashboard are:

- Allow users to interactively explore the data in different geographic resolutions
- Highlight shipping delays, outliers in profit and regions at disadvantage
- Highlight other aspects that might be interesting or unexpected
- Allow users to interactively filter the data according to time

Previous visualization: The team does not have a previous system but found an interesting visualization online in a brainstorming session. However, it only shows the zip code level information, and it is not interactive for filtering and exploration. Temporal information is also essential for this project which is not available in the online figure.

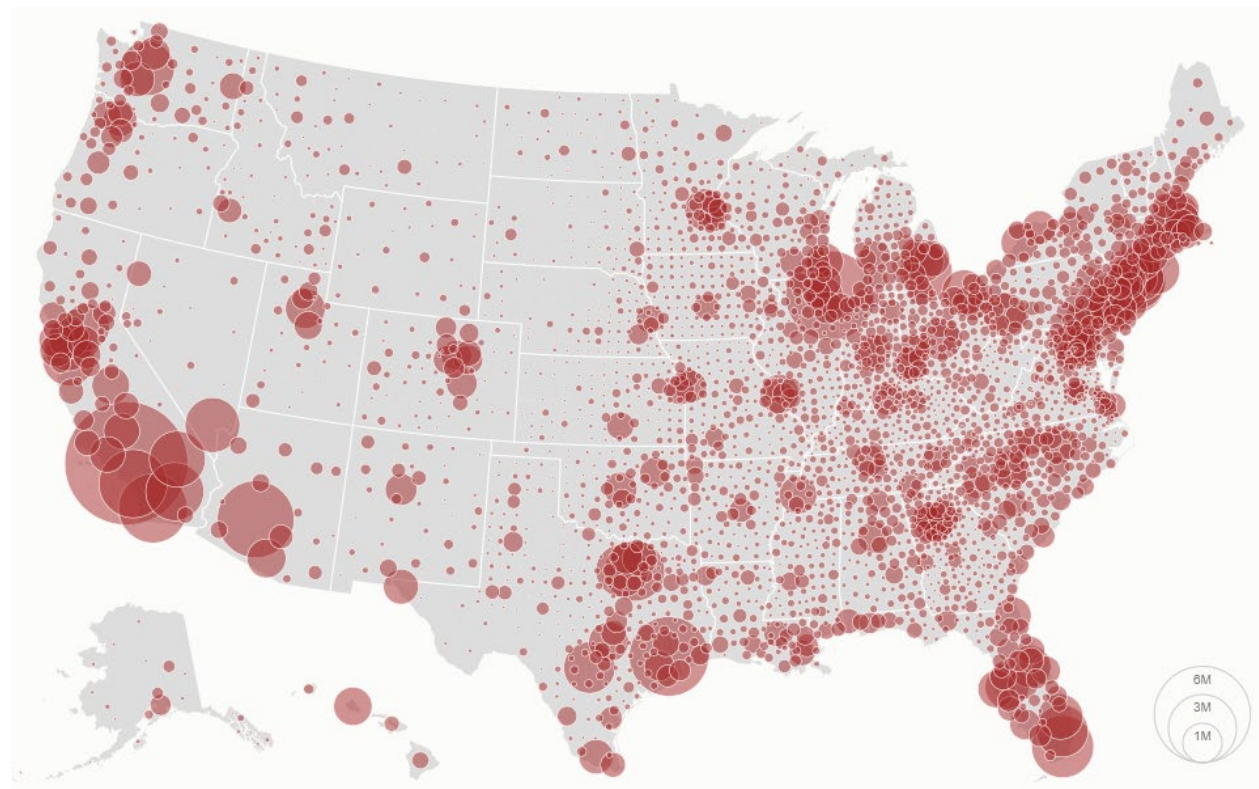


Image obtained online, created by Mike Bostock

Data: [drugs_prescriptions_sales.xlsx](#)

There are two tables in this xlsx file.

Rx table

Data fields:

Date: date
Dispensed Pharmacy: string
Drug Dosage Form: string
Drug Name: string
Drug Route: string
Drug MDR Unit Type: string
Drug Strength: string
Drug Supplier: string
Order ID: integer
Order Subitem: string
TableauCeuticalSales: decimal
Discount: decimal
Gross Profit Margin: decimal
Net Profit Margin: decimal
Rx Gross Sales: decimal
Rx Net Sales: decimal
Rx Net Profit: decimal
Rx Quantity: integer
Unit Price: decimal
Number of Records: integer

Examples:

6/3/2009
Bryan Spruell
TABLET
RISEDRONATE SODIUM
ORAL
TAB
35MG
TableauCeutical
154986
154986-003
78120
0%
42%
42%
78120
78120
32810.4
108500
0.72
1

Distribution table

Data fields:

Order Date: date
Dispensed Pharmacy: string
Dispensed Region: string
Dispensed State: string
Dispensed Zip Code: integer
Order Priority: string
Ship Date: date
Order ID: integer
Number of Records: integer

Examples:

9/20/2009
Mark Cousins
East
Massachusetts
01001
Medium
9/23/2009
155387
1