Open Data Workshop

HELLO!

I am Roberto Martin

I am here to talk about open data

You can find me at @bobstring

What is open data?

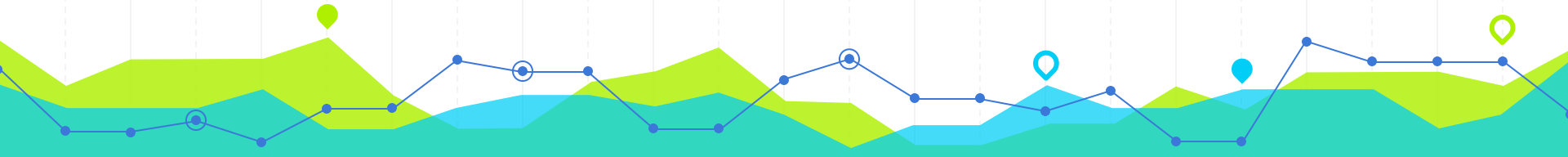Open data 101

1

OPEN DATA WORKFLOW

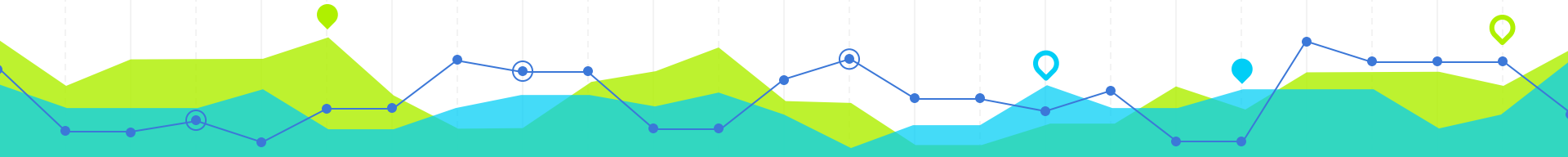**Obtaining** **Cleaning** **Displaying** **Storytelling**

Open data

Open data is the idea that certain data should be **freely available for everyone to use and republish as they wish**, without restrictions from copyright, patents, or other mechanisms of control [1]

Open data

Beneficial outcomes:

- ⊙ Increases **government transparency and accountability**
- ⊙ Enable and encourage the **development of applications** and services that address public and private demands [2]

FOIA and OPRA requests

◉ **F**reedom **o**f **I**nformation **A**ct (**FOIA**): is a **federal** freedom of information law that allows for the full or partial disclosure of previously unreleased information and documents controlled by the United States government

◉ **O**pen **P**ublic **R**ecords **A**ct (**OPRA**): is a **state statute** guaranteeing and providing process for freedom of information with respect to public records in the State of New Jersey [3]
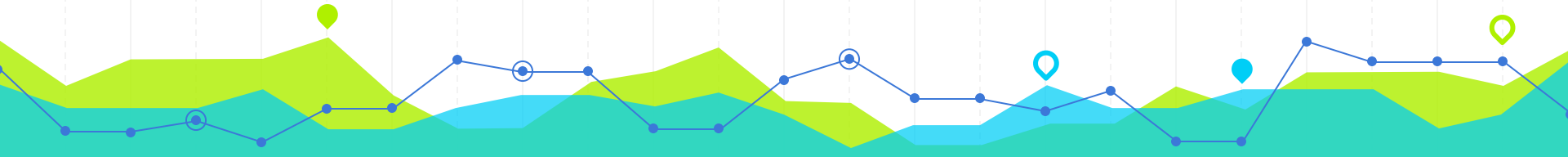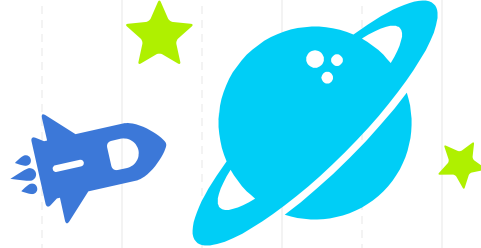
OPRA requests in Princeton

- Main page: http://www.princetonnj.gov/OPRA.html
- Princeton OPRA request form: http://www.princetonnj.gov/forms/PrincetonOPRARequest.pdf
- Where to file it: Office of the Clerk (400 Witherspoon Street Princeton, NJ 08540)
- 2 OPRA requests made (fire hydrants and OPRA log)

State and federal requests

- Most federal agencies: http://www.foia.gov/
- State of New Jersey: http://www.state.nj.us/opra/
- Mercer County: http://nj.gov/counties/mercer/home/opra.html

The power of open data

Ben Wellington: How we found the worst
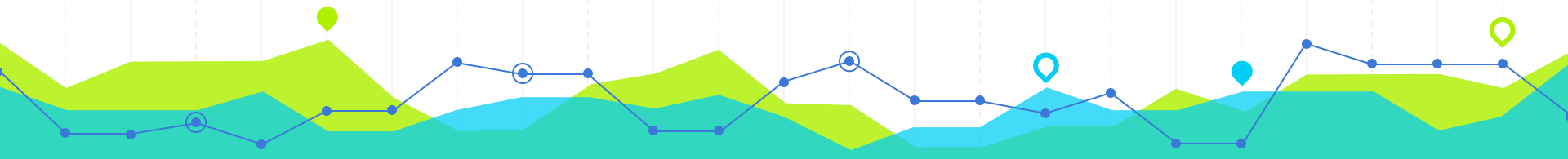place to park in New York City, using big data
https://youtu.be/lz_kIDxbzGA

Data portals

Some agencies, cities and organisations provide a data portal and/or an online form to do OPRA/FOIA requests:

- Jersey City data portal: http://www.cityofjerseycity.com/data/
- NYC data portal: https://nycopendata.socrata.com/
- Code for NYC (Beta NYC) data portal: http://data.beta.nyc/
- Code for Jersey City data portal: https://data.openjerseycity.org/

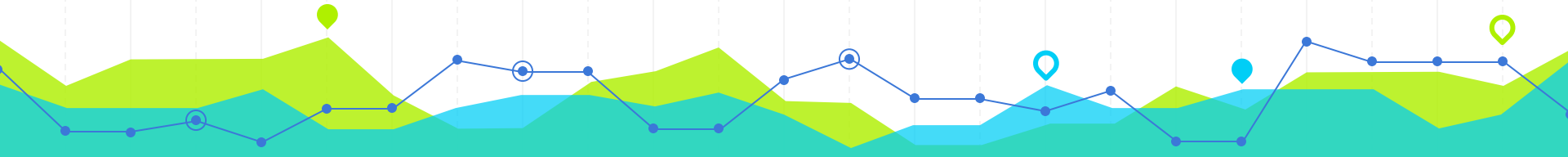# Cleaning data

2

OPEN DATA WORKFLOW

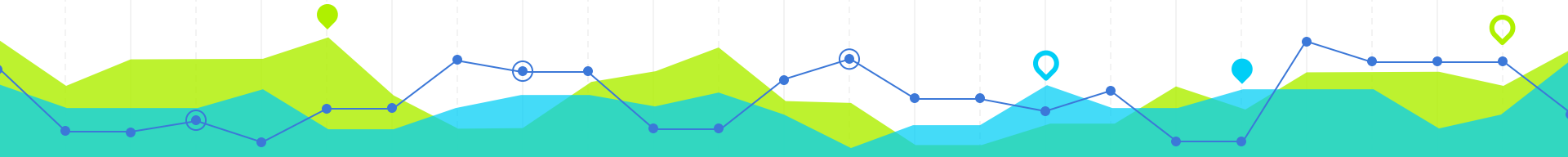**Obtaining** → **Cleaning** → **Displaying** → **Storytelling**

What is cleaning the data?

The process of transforming the raw data obtained from an agency or organization into a more understandable format
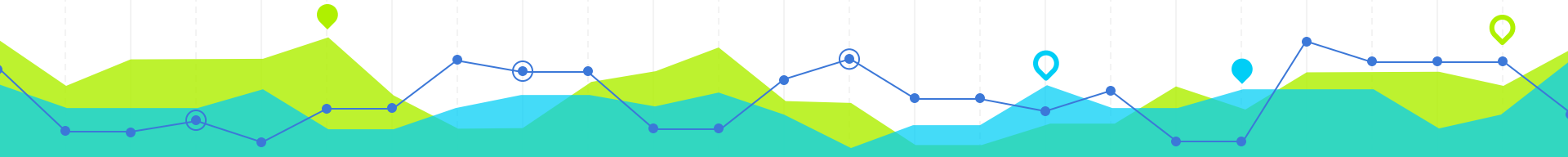
Examples:

- ◉ Changing plain addresses into longitude and latitude
- ◉ Convert PDF into spreadsheet format

Tools for cleaning data

Basic tools:
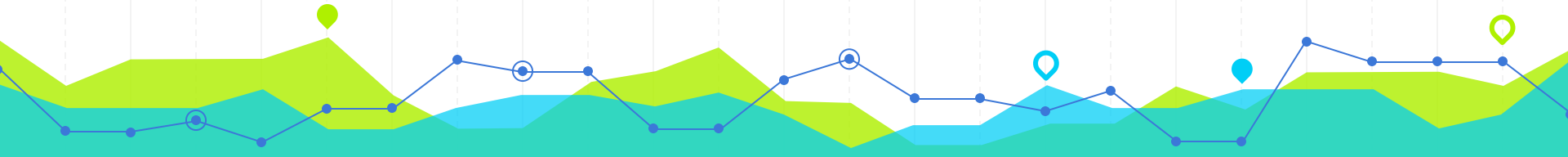
- ◉ Notepad
- ◉ Excel/Spreadsheet program

OpenRefine

- OpenRefine (formerly Google Refine) is a powerful tool for working with messy data: cleaning it; transforming it from one format into another; extending it with web services; and linking it to databases. It is **open source software** and is licensed under the BSD license.
- http://openrefine.org/
- Good tutorial: http://is.gd/refine

Uses of OpenRefine

- Working with big amount of data. With 80000 records splitting a column in Excel takes 5 min, with OpenRefine takes seconds
- Easy syncing with other services (transforming addresses to lat and long with open street maps)
- Finding and fixing common mistakes: Us, US, United States, us

Displaying the data

3

OPEN DATA WORKFLOW

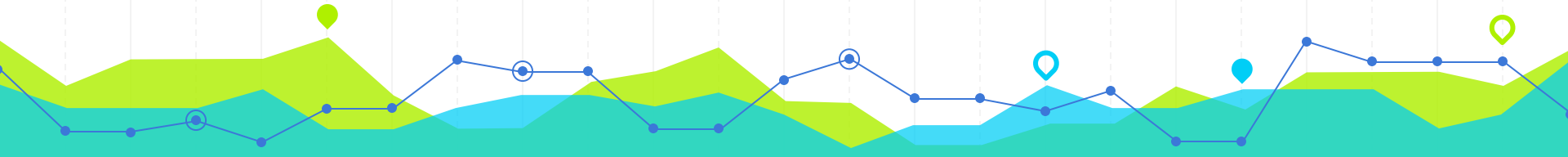**Obtaining** **Cleaning** **Displaying** **Storytelling**

What is displaying the data?

The process of presenting the data in a graphic format. It could be showing the data in a map or using a chart to show valuable information about the data.

We are going to take a look at a couple of tools to map our data:

- ◉ Google fusion tables
- ◉ CartoDB

Google Fusion Tables

A Google product to plot data in a map.

- ◉ Go to drive.google.com and create a "Google Fusion Table"
- ◉ Example: https://goo.gl/fglTLs

Characteristics:

- ◉ Simple and intuitive
- ◉ Can be connected with Google Forms so users can add data dynamically.

CartoDB

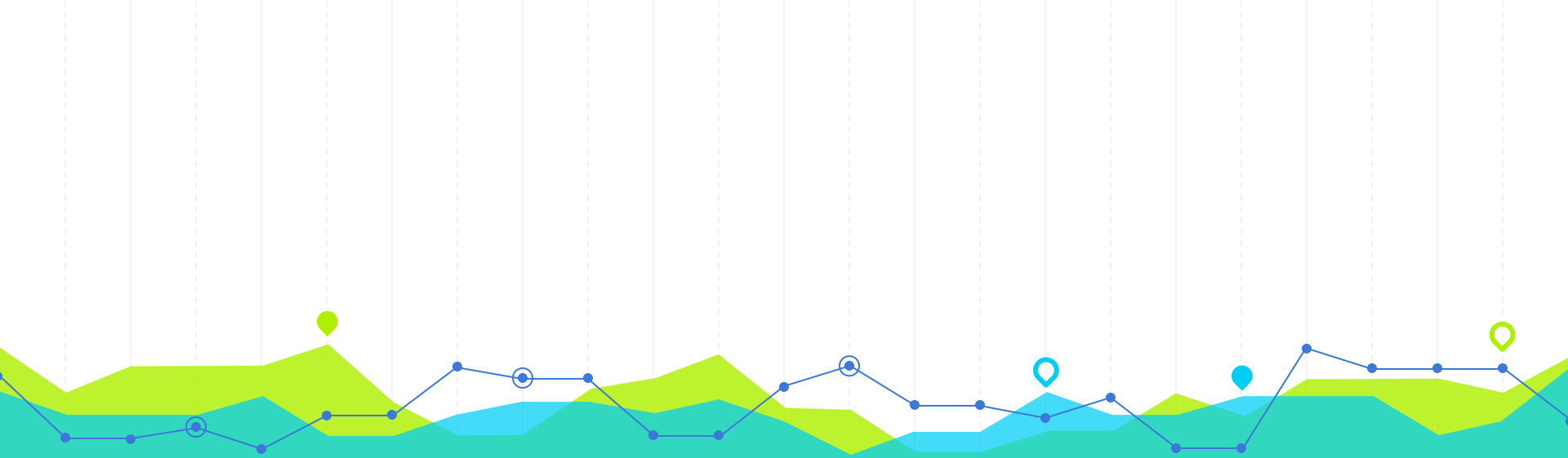A popular tool to map and analyze location data.

- ◎ http://cartodb.com
- ◎ Example: http://cdb.io/1DnNM5E

Characteristics:

- ◎ Many options and maps
- ◎ Limited geotagging
- ◎ SQL queries on the data
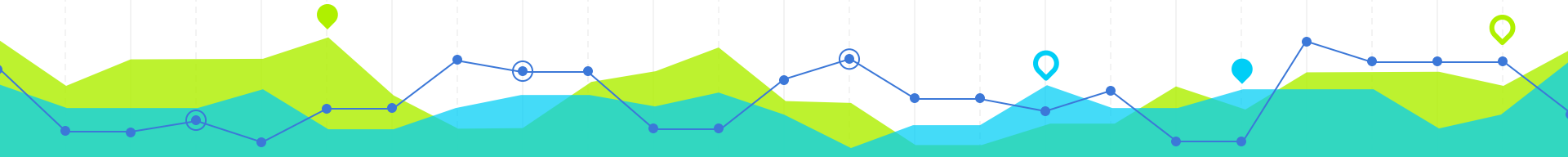
Storytelling

4

OPEN DATA WORKFLOW

Obtaining

Cleaning

Showing

Storytelling

What is storytelling?

Once you have the data relaying what you want in the way you want, it is time to let it tell its story. Are there any interesting findings? Any conclusions from the analyses?

Examples

The New York Times article about New York Health department restaurant ratings.

- ⊙ Article: http://nyti.ms/WRuqQN
- ⊙ CartoDB map: https://mgrimshaw.cartodb.com/viz/f2083de2-4424-11e4-afd2-0e73339ffa50/public_map

Examples

WNYC article about putting a face and a short story behind fatal accidents in NYC. WNYC used public data from NYPD to create the stories. This demonstrates that data does not always have to be a map
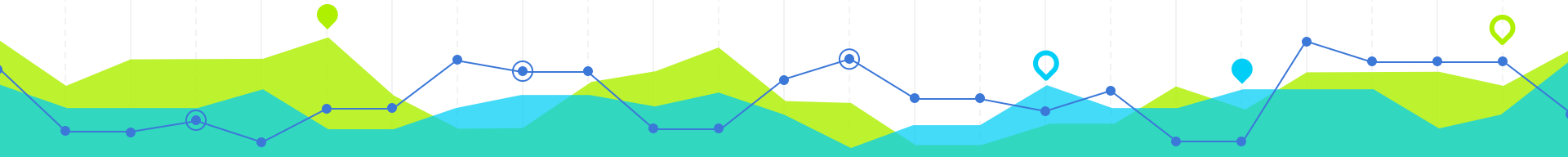
- ◉ Article: http://www.wnyc.org/story/killed-cars-who-dying-traffic-crashes-and-why/

Workshop

Let's practice!

5

Jersey City Fire Hydrants Tickets

⊙ Check: http://robertmm.com/open-data-workshop-CfP/ for useful information related to the workshop
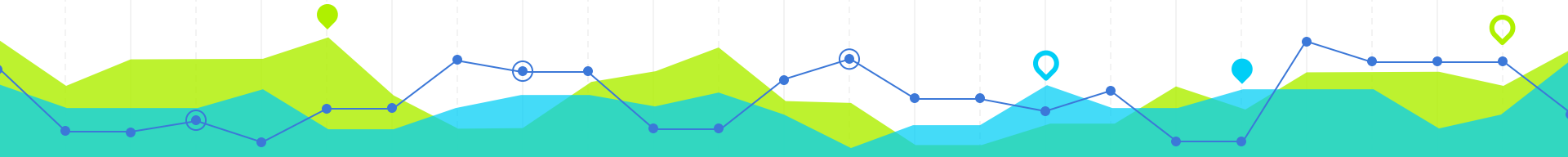
THANKS!

Any questions?

You can find me at

@bobstring / contact@robertmm.com

Works cited

1. Auer, S. R.; Bizer, C.; Kobilarov, G.; Lehmann, J.; Cyganiak, R.; Ives, Z. (2007). "DBpedia: A Nucleus for a Web of Open Data". The Semantic Web. Lecture Notes in Computer Science 4825. p. 722
2. Brito, Jerry. "Hack, Mash, & Peer: Crowdsourcing Government Transparency". Colum. Sci. & Tech. L. Rev. 119 (2008).
3. Freedom of Information Act (United States) (https://en.wikipedia.org/wiki/Freedom_of_Information_Act_(United_States))
4. Open Public Records Act (https://en.wikipedia.org/wiki/Open_Public_Records_Act)

CREDITS

Special thanks to all the people who made and released the template, SlidesCarnival.