

Unconventional Sources of Data

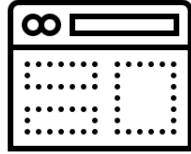


Getting Data

Unstructured Data Types



Text files and
documents



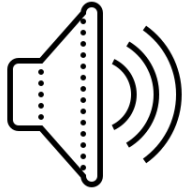
Websites and
applications



Sensor
data



Image
files



Audio
files



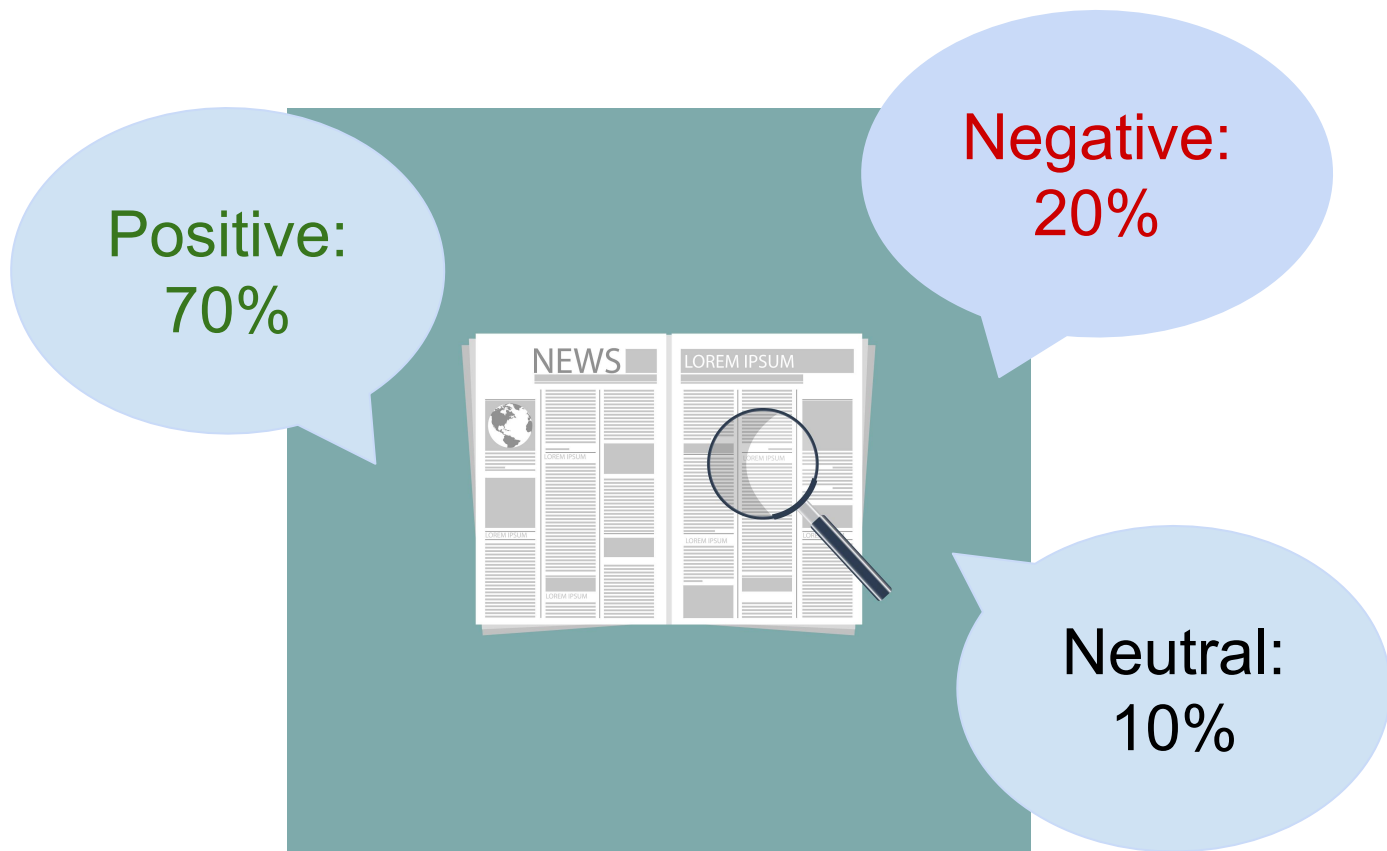
Video
files



Email
data



Social media
data





David Robinson

Chief Data Scientist at
DataCamp, works in R and
Python.

- Email
- Twitter
- Github
- Stack Overflow

Subscribe

Recommended Blogs

- DataCamp
- R Bloggers

Text analysis of Trump's tweets confirms he writes only the (angrier) Android half

I don't normally post about politics (I'm not particularly savvy about polling, which is where data science has had the largest impact on politics). But this weekend I saw a hypothesis about Donald Trump's twitter account that simply begged to be investigated with data:



JSON: key-value pairs

`{"Name": "Isabela"}`

key



value



These are all
nested within
attributes

```
"attributes": {  
  "Take-out": true, A key-value pair  
  "Wi-Fi": "free",  
  "Drive-Thru": true,  
  "Good For": {  
    "dessert": false,  
    "latenight": false,  
    "lunch": false,  
    "dinner": false,  
    "breakfast": false,  
    "brunch": false  
  },  
}
```

These are all
nested within
"Good For"



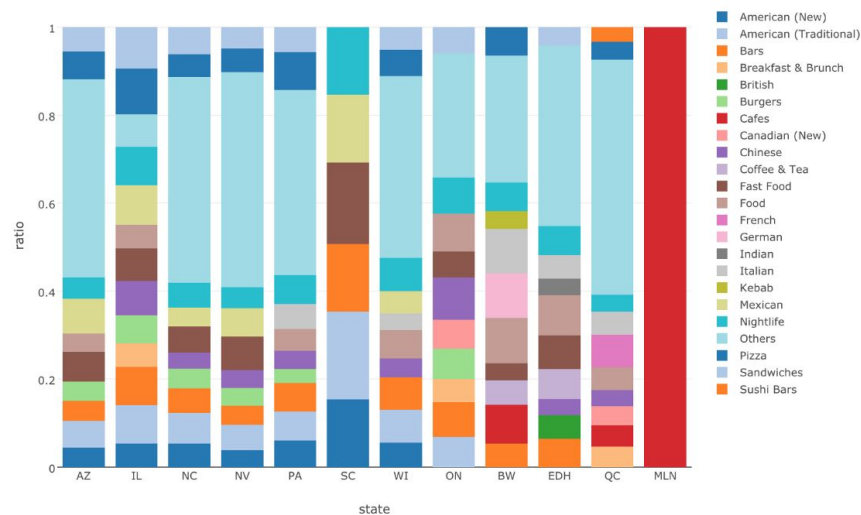
Kan Nishida

Follow

CEO / Founder at Exploratory(<https://exploratory.io/>). Having fun analyzing interesting data and learning something new everyday.

Mar 29, 2016 · 10 min read

Working with JSON data in very simple way



A node

\$node

An opening tag

<tag>

An element

<tag2> more content </tag2>

<tag3> more content </tag3>

</tag>

A closing tag



José Roberto Ayala Solares

Follow

Research Scientist working on Deep Medicine at @UniofOxford / Inženiero je profesio, matematikisto je penso, esploristo je datumoj kaj vojaĝanto je koro...

Aug 2, 2017 · 5 min read

Web scraping tutorial in R

A couple of days ago, [Kevin Markham](#) from [Data School](#), published a nice [tutorial](#) about web scraping using 16 lines of Python code.

Web scraping the President's lies in 16 lines of Python

Note: This tutorial is available as a Jupyter notebook, and the dataset of lies is available as a CSV file, both of...

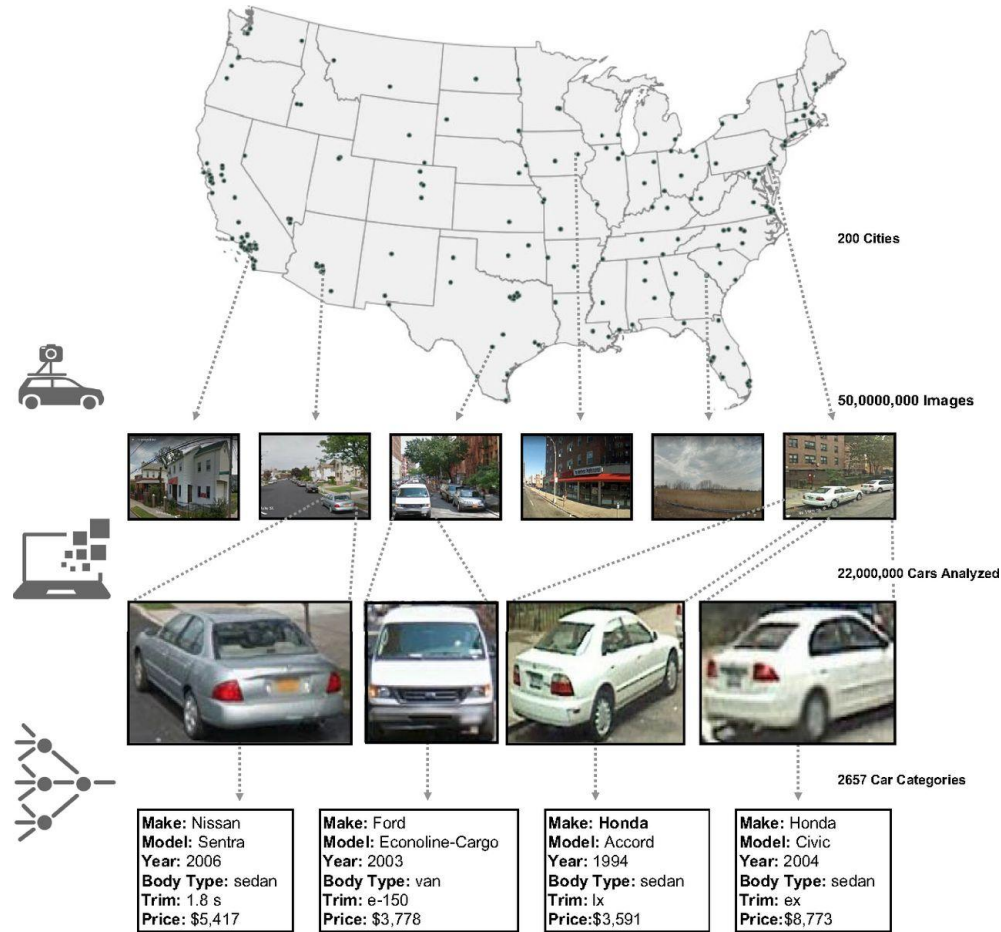
www.dataschool.io

The tutorial is simple and really well-made. I strongly encourage you to have a look at it. In fact, such a tutorial motivated me to replicate the results but this





Source: Google



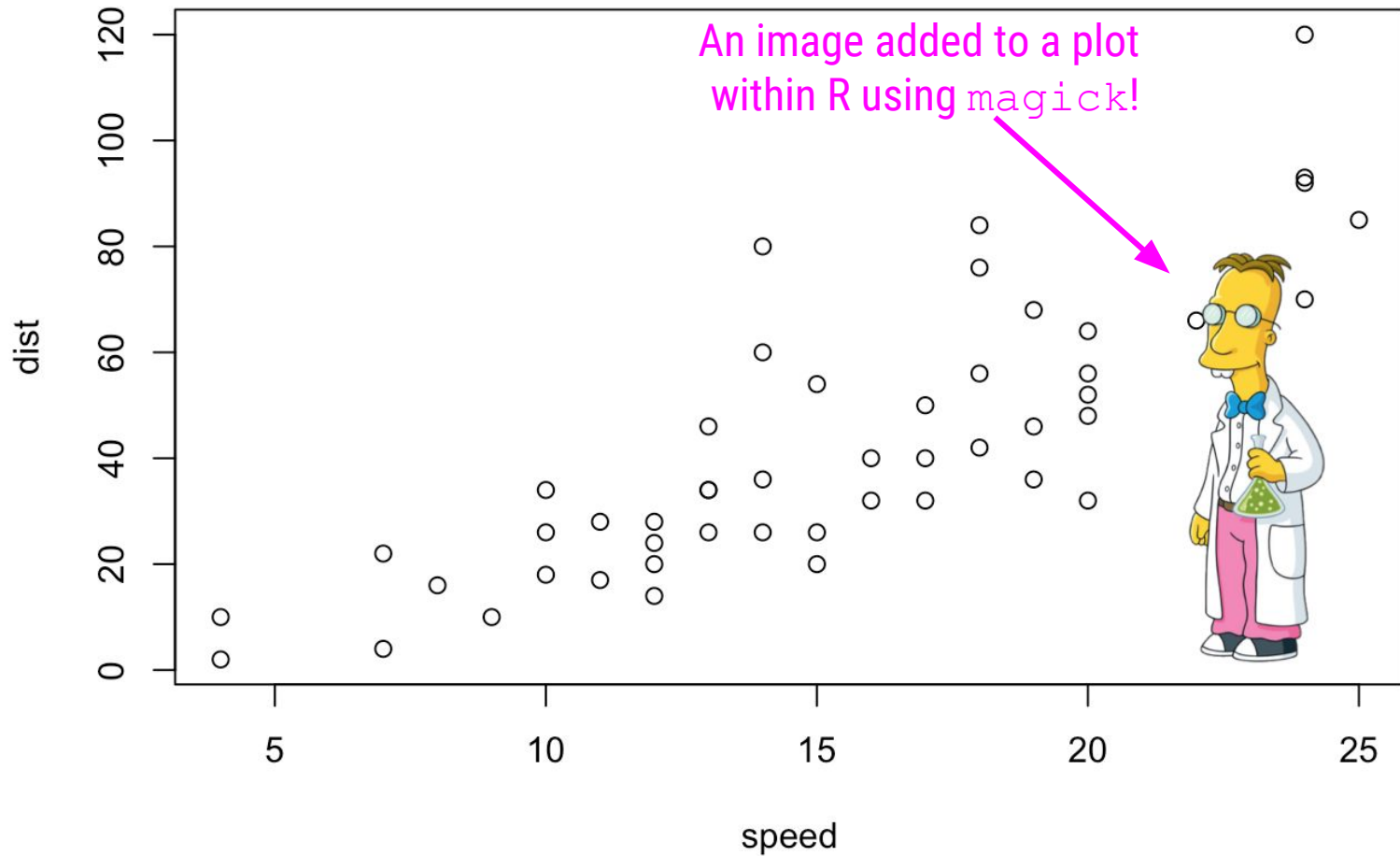
Source: Demography with deep learning and street view, Gebru et al.



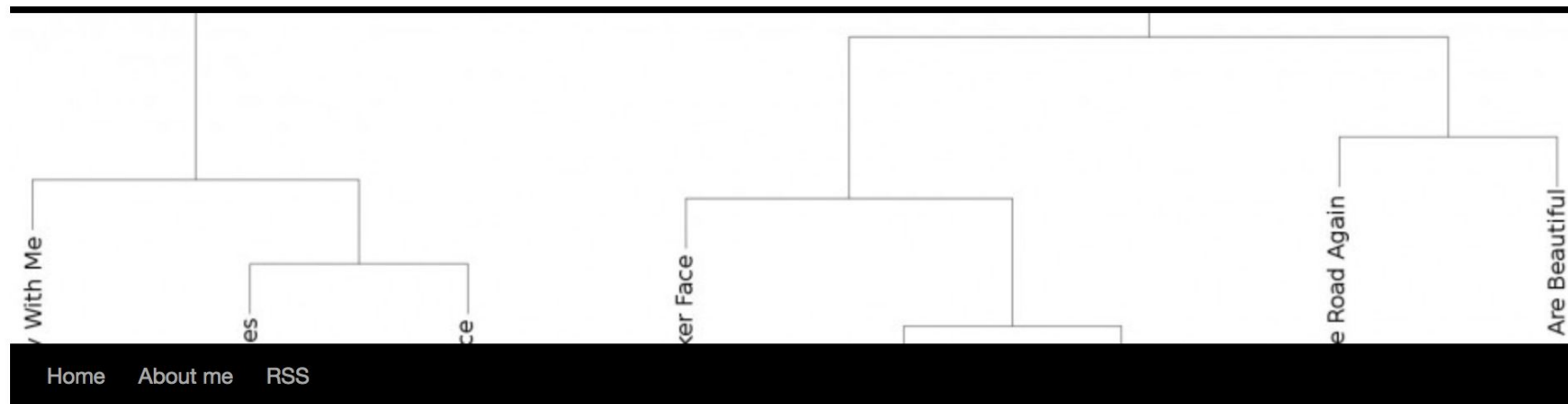
The cars in this image of a Brooklyn neighborhood can reveal a lot about the residents there. // Google Street View

Google Street View Can Reveal How Your Neighborhood Votes

LINDA POON DEC 6, 2017







← tRanscribing music from audio files

Network visualization – part 1: Cytoscape →

A simple way to cluster music

Posted on [January 25, 2013](#) by [Vessy](#)

In my last blog, I discussed the [tuneR](#) functions that provide an option to transcribe musical notes from audio frequencies. In this blog, I'll write about functions for comparison of audio spectrum distributions, available in the [seewave](#) library.

The idea I wanted to test here is simple: can I use similarities between audio spectrum distributions to find songs that are similar to each other? More specifically, can I use

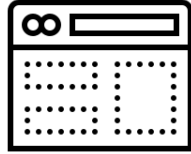
Recent Posts

- [Network visualization – part 6: D3 and R \(networkD3\)](#)
- [A few thoughts on the existing code parallelization](#)
- [Network visualization – part 5: Cytoscape \(an update\) – RCy3](#)
- [Networks, mazes, and R \(Rmaze\)](#)
- [One function to run them all... Or just eval](#)

Unstructured Data Types



Text files and
documents



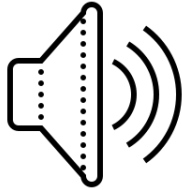
Websites and
applications



Sensor
data



Image
files



Audio
files



Video
files



Email
data



Social media
data