

Colombia's Open Data Explorer

Juan Carlos Méndez

jc.mendez (at) uniandes.edu.co

Url: <http://jcmendez.gkudos.com/hw5/>

Video: https://youtu.be/8T0eAf4_CHw

The Customer

*Looks like there are lots of data published in
datos.gov.co.*

*I would like to have a quick view of the kind of
information it is published in there.*



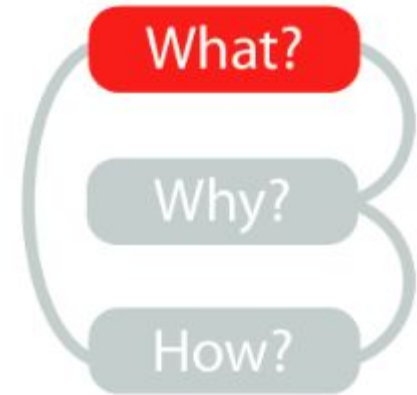
MATT GROENING

Description

- Create a dashboard that allows to explore some facts about datos.gov.co open data catalog
- Data:
 - Datos Abiertos Gobierno Digital Colombia
 - Socrata Catalog : <https://www.datos.gov.co/api>
 - Datos.gov.co

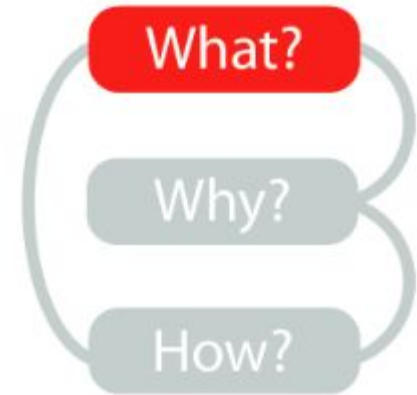
Dataset

- Dataset Name: datos.gov.co
- Dataset Type: Table, Temporal, Static
- Attributes:
 - Publication Date : Date, Quantitative, ordered
 - Dataset Name: Categorical
 - Category: Categorical
 - Organization: Categorical
 - State: Categorical
 - Municipality: Categorical
 - Sector: Categorical
 - Order: Categorical
 - permalink : categorical
 - Id: categorical



Derived Attributes

- year: date, quantitative, ordered, sequential
- Unique Id's: It is necessary to generate unique ids for tree and network visualizations
 - Category_id: categorical
 - Organization_id: categorical
 - Sector_id: categorical
 - municipality_id: categorical
 - State_id: categorical
 - Order_id: categorical



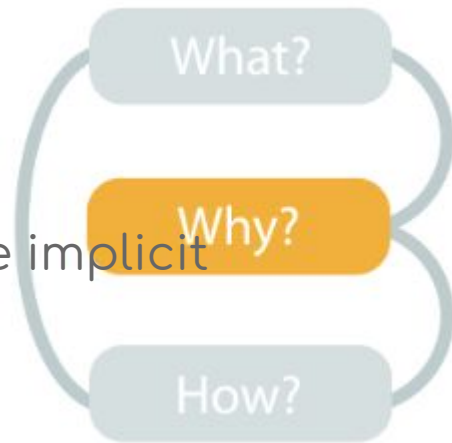
Tasks

Main Task:

- Discover Distribution about data according to the implicit hierarchy.

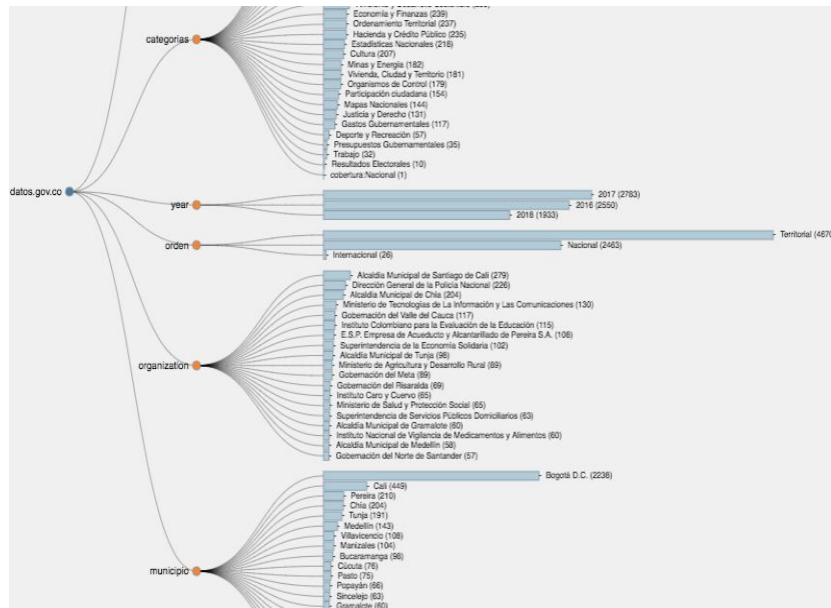
Secondary Tasks:

- Identify Paths / dependencies between dataset categories and organizations
- Explore open datasets as **Features** belonging to organizations
- Locate clusters of organizations according to **similarities** in categories
- Derive **Features** to use as inputs for network / tree visualizations



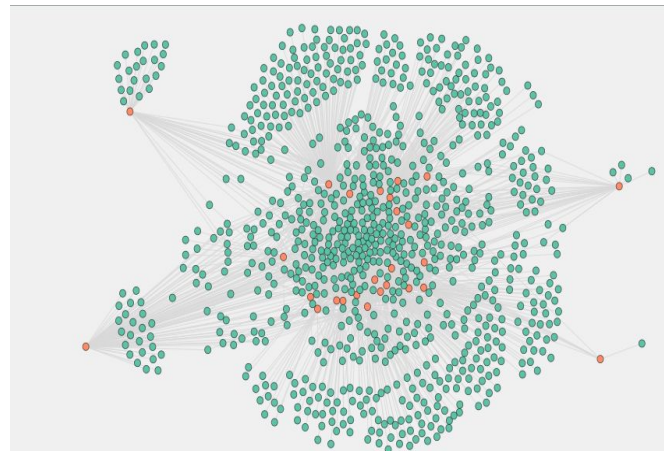
Abstractions

- Idioms: Tree Histogram
 - Mark: line
 - Attributes:
 - Category, state, municipality, year, order, organization
 - Channels:
 - Position on a common scale, X position, Y Position,
 - Encode -> arrange -> separate, order align
 - Facet: juxtapose
 - Mark: point
 - Attributes:
 - Hierarchy type
 - Channels:
 - Position on a common scale, X position, Y Position
 - Color Hue
 - Encode -> arrange -> separate, order align



Abstractions

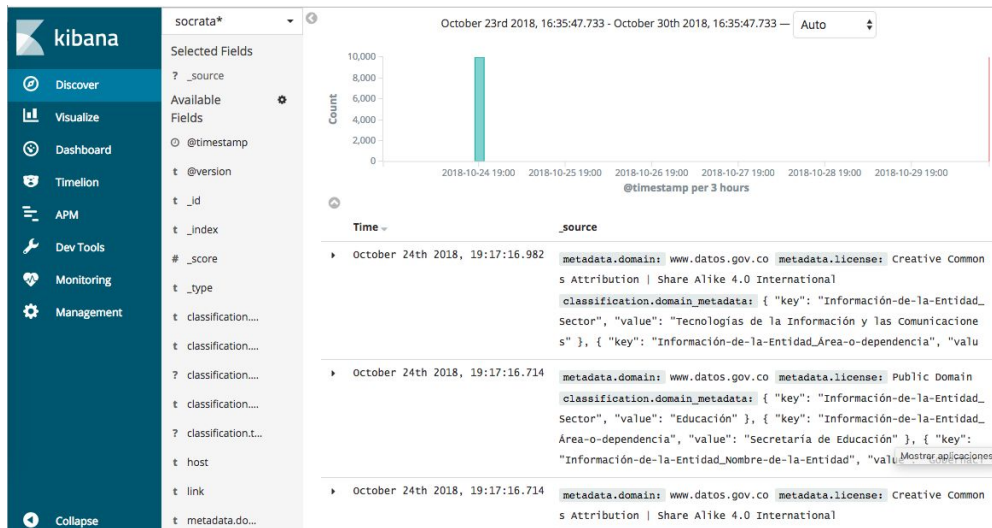
- Idioms: Force Directed Placement
 - Mark: point
 - Attributes:
 - Category, organization
 - Channels:
 - Position X, position y, color hue
 - Encode -> arrange, separate
 - Reduce -> filter
 - Mark: Line
 - Attributes:
 - Link between organization and category
 - Channels:
 - Length, angle
 - Encode -> arrange, separate



Tech



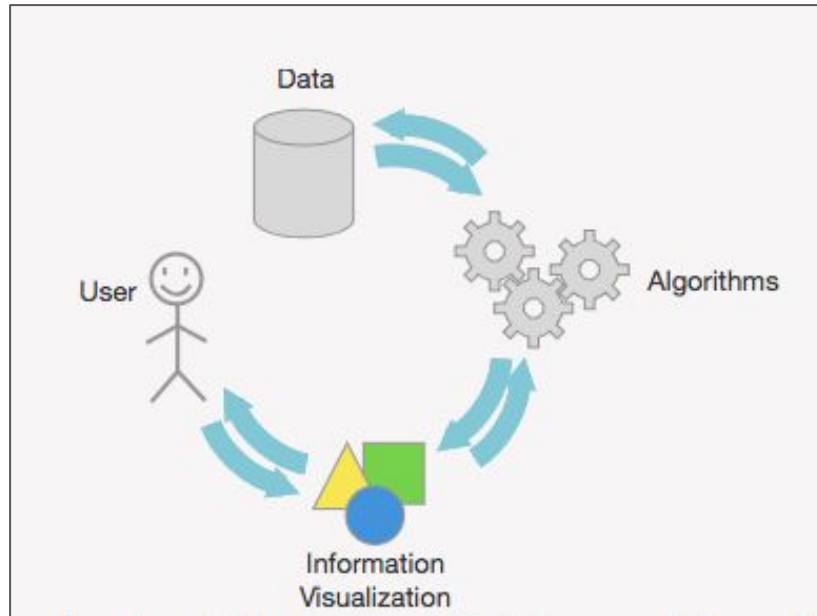
- Derivation / ETL / Data Quality Checking:
 - Postgresql (Carto)
 - Python
 - Logstash / Elasticsearch / Kibana



Tech



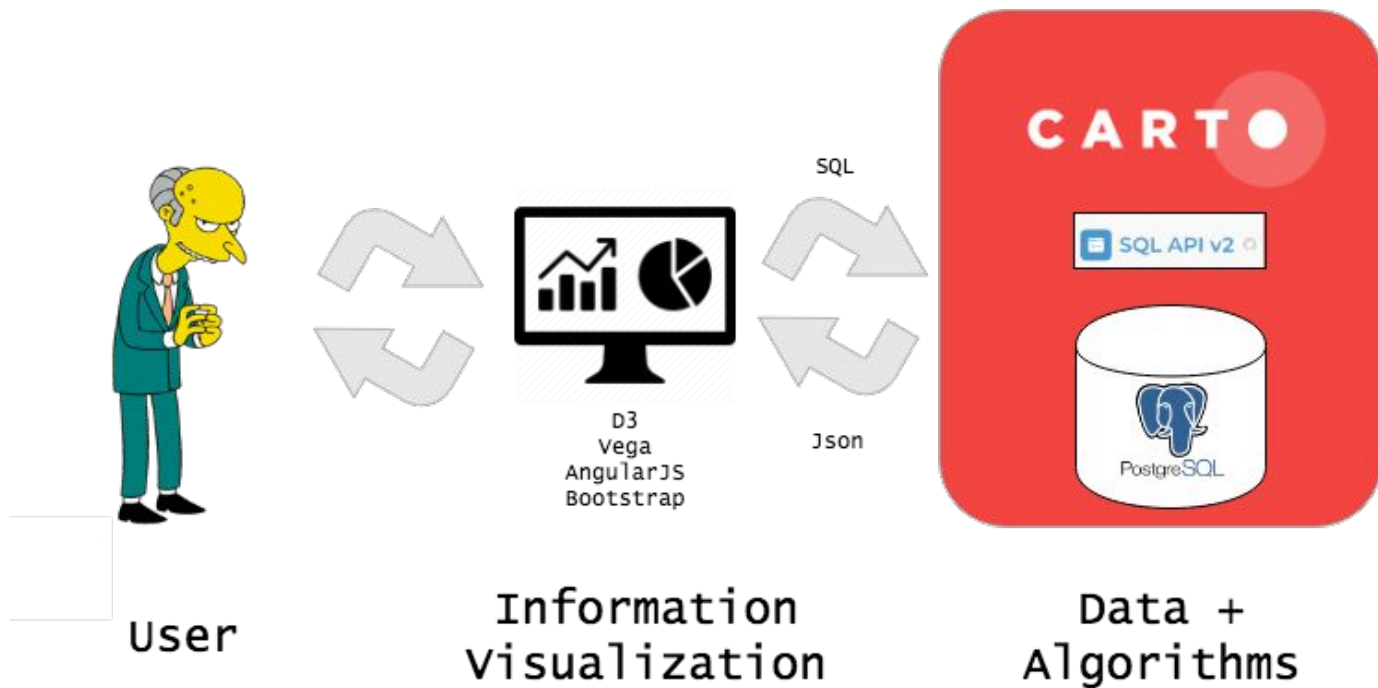
- Visual Analytics Reference Architecture



Tech



- Visual Analytics Architecture: Actual Implementation



Tech



- Visual Analytics Architecture: Actual Implementation
 - Dynamic SQL (details on demand)

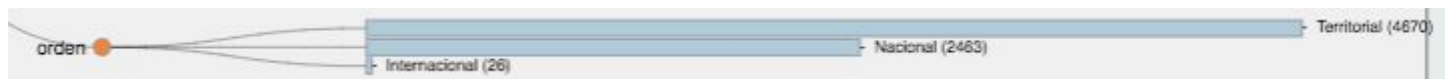
```
$scope.loadDatasets = function (d) {  
  $scope.selectedNode = d  
  
  var filterColumn = ""  
  if (d.id.substring(0, 1) == 'o') {  
    filterColumn = "organization_code"  
  } else {  
    filterColumn = "category_code"  
  }  
  
  let query = ` select resource_id as id, permalink as link, INITCAP(resource_name) as label, resource_createdat as creationDate  
    from datos_gov where ${filterColumn} = '${d.id}'  
    order by resource_name asc `
```



```
  // $log.log(query);  
  
  $scope.progressbar.start();  
  d3.json($scope.remoteServiceUrl + query).then($scope.parseDatasets);  
}
```

Insights

- Datos.gov.co has some datasets related to international topics (not only national or regional)

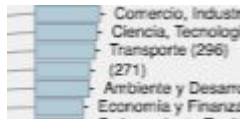


- Medellín is not one of the municipalities on the top 3 of open data publishing in colombia

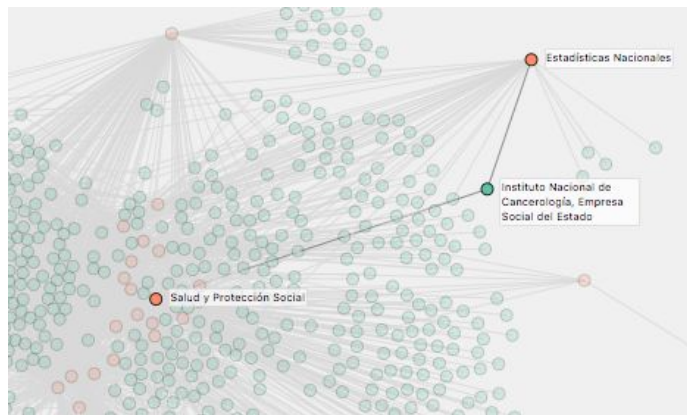
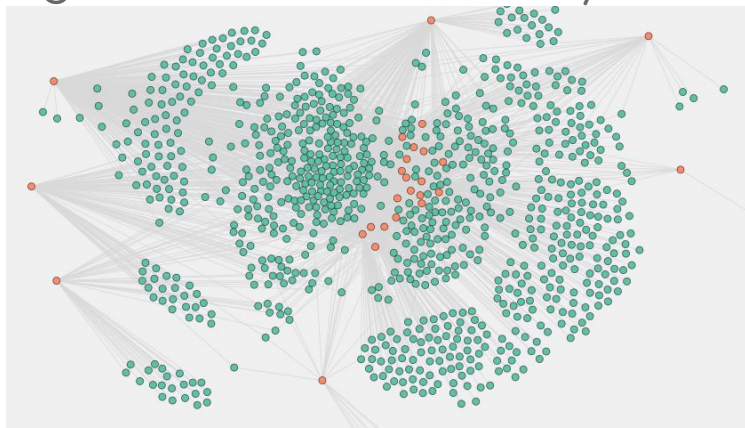


Insights

- There are data quality issues (nulls in categories, organizations, etc)



- Thank to force layout it is “easy” to recognize clusters of organizations related by their published datasets



Difficulties

- Using “the force” has problems with nulls (it was needed to remove null data)

Achievements

- Learn about *“the force”*.
- Learn about *Vega*
- Implementation of *Visual Analytics Reference Architecture*
- Implementation of *“visualization mantra”*



Thanks