Driving in Bogotá, How did we do it in 2017?

Juan Carlos Méndez
jc.mendez (at) uniandes.edu.co

Url: http://jcmendez.gkudos.com/hw4/
Video:

The Customer

According to World Health Organization, road traffic crashes are predicted to become the seventh leading cause of death by 2030. Moreover road traffic crashes cost most countries 3% of their gross domestic product.

People is starting to get worried. I would like to have a tool to get an idea about trends and characteristics of traffic crashes in Bogotá using temporal historical data and ...

I need it for tomorrow

Description

- Create a dashboard that allows to explore some facts about traffic accidents in Bogotá using temporal data.
- Data:
  - Accidentes de Tránsito en la ciudad de Bogotá - 2017
  - Url https://bit.ly/2NOn9SY
  - Source: Secretaría Distrital de Movilidad (datos.gov.co)

Dataset

- Dataset Name: Traffic accidents in the city of Bogotá - 2017
- Dataset Type: Table, Temporal, Static
- Attributes:
  - FechaOcurrencia: Date, Quantitative, ordered
  - HoraOcurrencia: Time, Quantitative, ordered, cyclic
  - GravedadNombre: Categorical
  - Localidad: Categorical
  - Totalmuertos: Quantitative, ordered, sequential

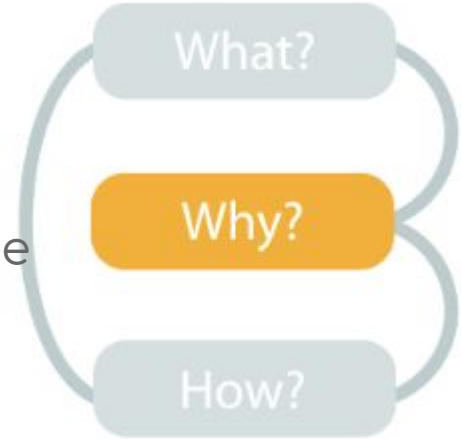
Derived Attributes

- Derivation: *FechaOcurrencia* and *HoraOcurrencia* needed to be processed to be used as a timestamp
- Attributes:
  - Event_time: date, quantitative, ordered, sequential
  - Event_date: date, quantitative,ordered, sequential
  - Event_hour: Quantitative, ordered, cyclic
  - Event_day: Quantitative, ordered, cyclic
  - Event_day_name: Categorical, Ordinal
  - event_month: Quantitative, ordered, cyclic
  - event_month_name: Categorical, Ordinal

Derived Attributes

- Derivation: Moving average of total number of events per day. Used to smooth the time series.
- Attributes:
  - Trend: quantitative, ordered, sequential
  - Total events: quantitative, ordered, sequential

Tasks



Main Task:

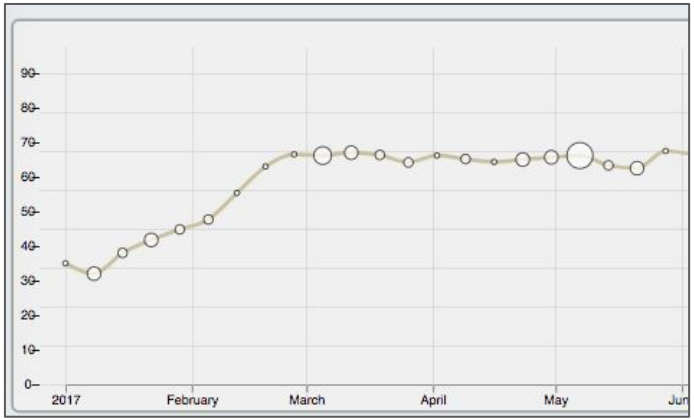
- **Discover Trends** in traffic accidents related to time

Secondary Tasks:

- **Explore Features** of traffic accidents
- **Identify Features** of traffic accidents according to relevant attributes of data
- **Identify Features** of traffic accidents
- **Summarize Trends** of traffic accidents related to time and relevant attributes
- **Derive Features** to use data as a time series.

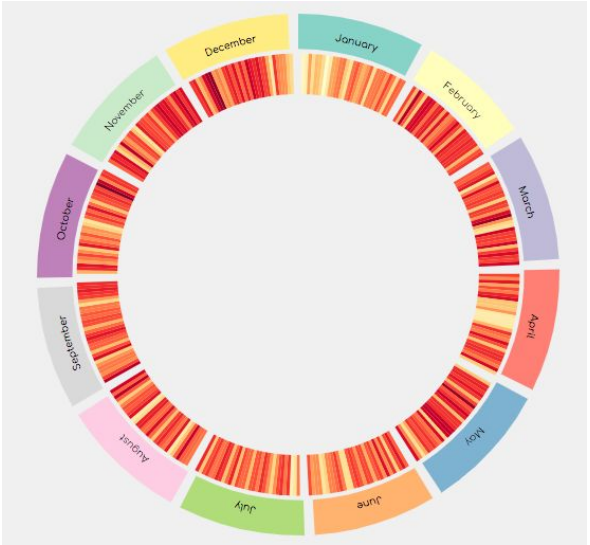
Abstractions

- Idioms: Line Chart
  - Mark: line
    - Attributes:
      - Event_date, total events
    - Channels:
      - Position on a common scale, X position, Y Position,
    - Encode -> arrange -> express
    - Reduce -> filter, aggregate
  - Mark: Area
    - Attributes:
      - Totalmuertos
    - Channels:
      - Position on a common scale, X position, Y Position, Size Area
    - Encode -> arrange -> express
    - Reduce -> filter

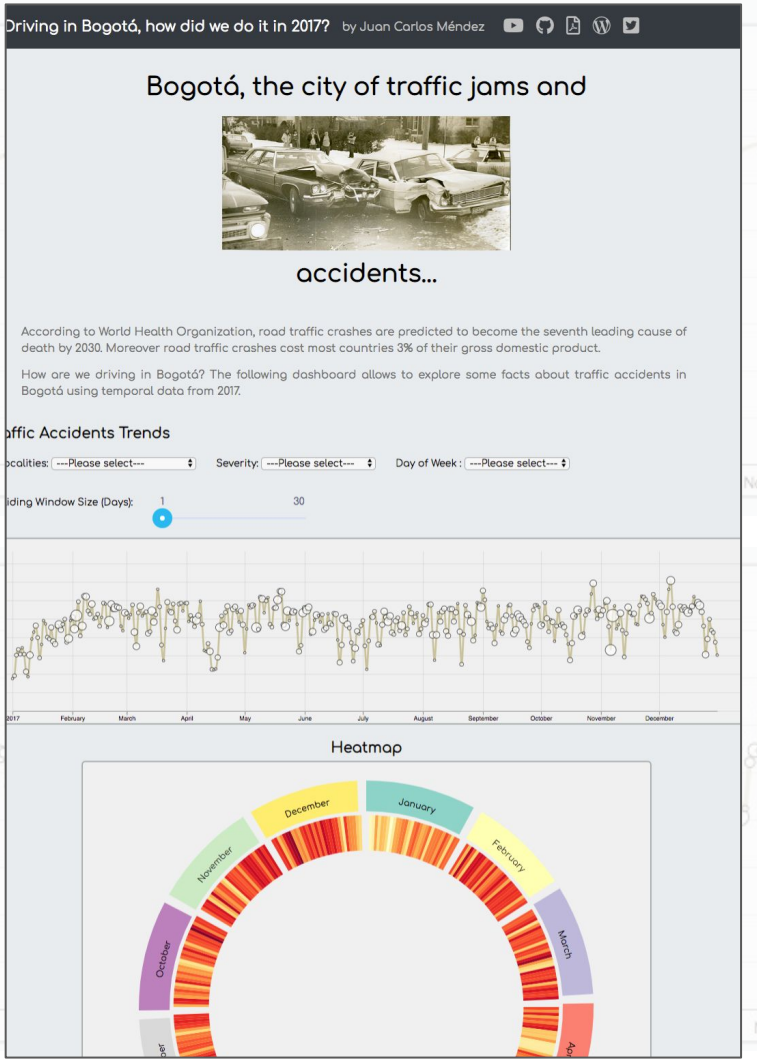


Abstractions

- Idioms: Radial Timeline / Heatmap
  - Mark: Area
    - Attributes:
      - Date, total of events
    - Channels:
      - Position on a common scale, radial position, spatial region, color saturation
    - Encode -> arrange, express
    - Reduce -> filter
  - Mark: Line
    - Attributes:
      - Month
    - Channels:
      - Radial position, spatial region, color hue
    - Encode -> arrange, separate, order, align

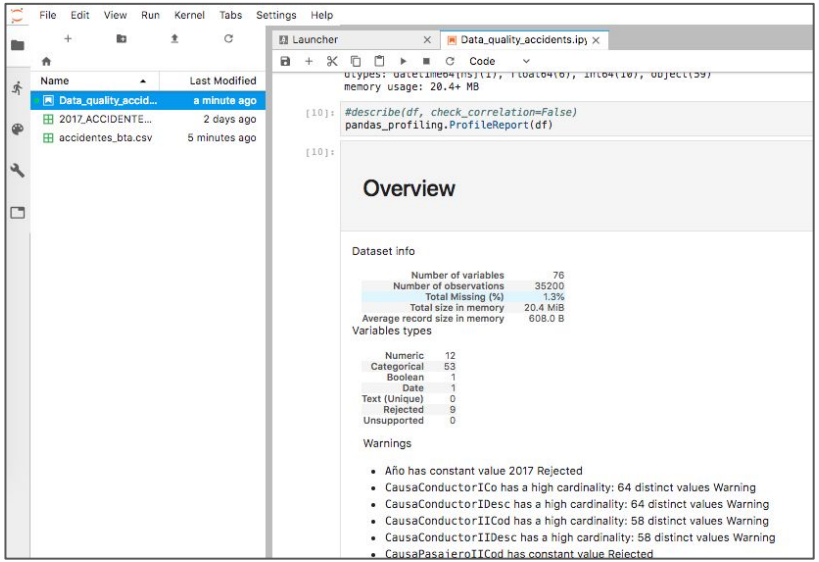


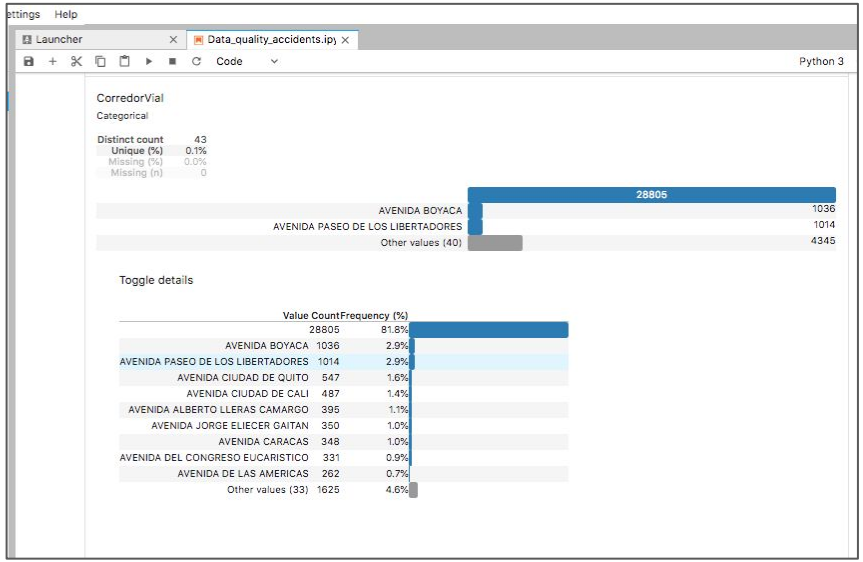
Demo


Driving in Bogotá, how did we do it in 2017? by Juan Carlos Méndez
Bogotá, the city of traffic jams and
accidents...
According to World Health Organization, road traffic crashes are predicted to become the seventh leading cause of death by 2030. Moreover road traffic crashes cost most countries 3% of their gross domestic product.
How are we driving in Bogotá? The following dashboard allows to explore some facts about traffic accidents in Bogotá using temporal data from 2017.
Heatmap
December
January
February
March
November
October

Tech

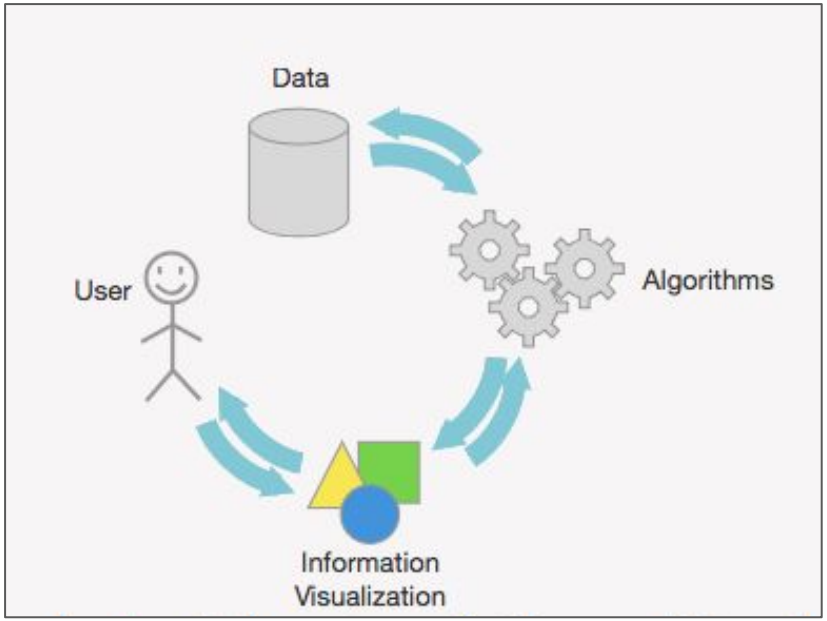
- Derivation / Data Quality Checking:
  - Jupyter lab





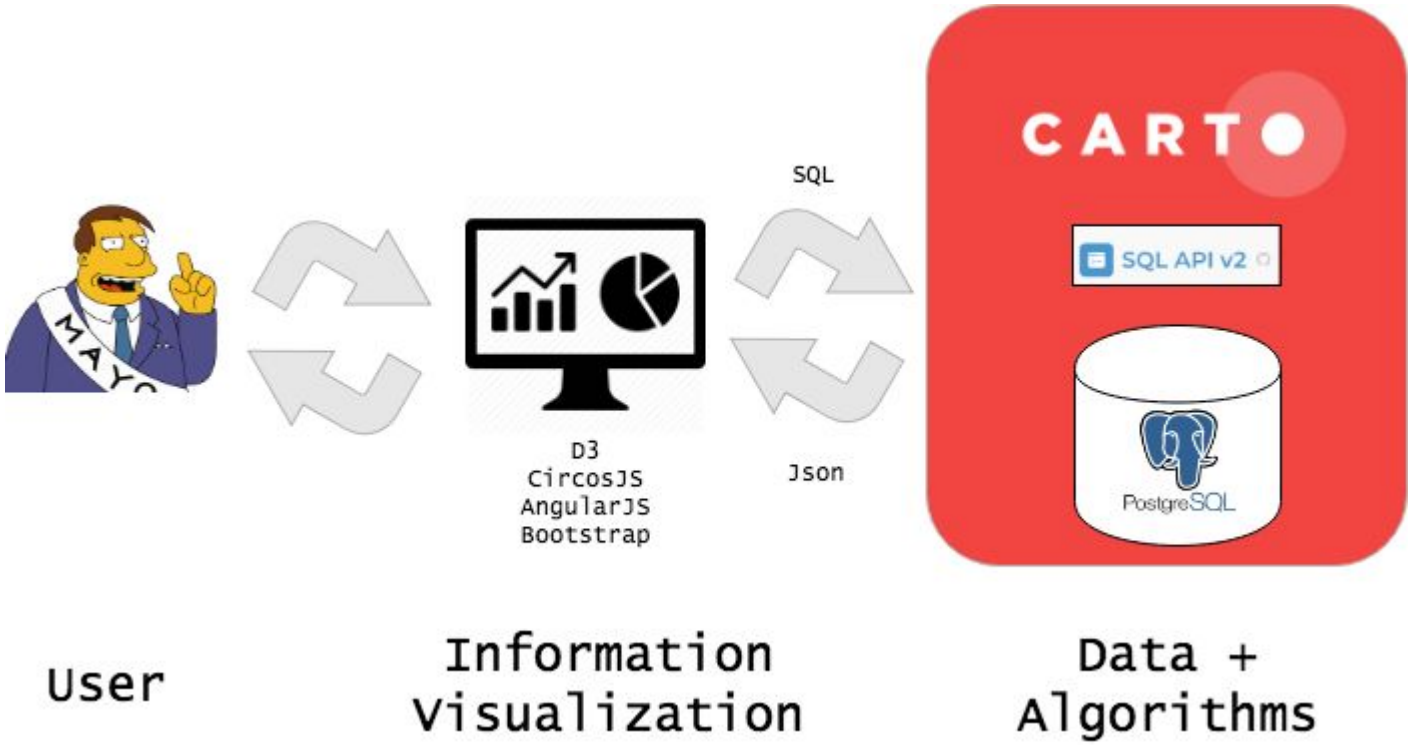
Tech

- Visual Analytics Reference Architecture



Tech

- Visual Analytics Architecture: Actual Implementation



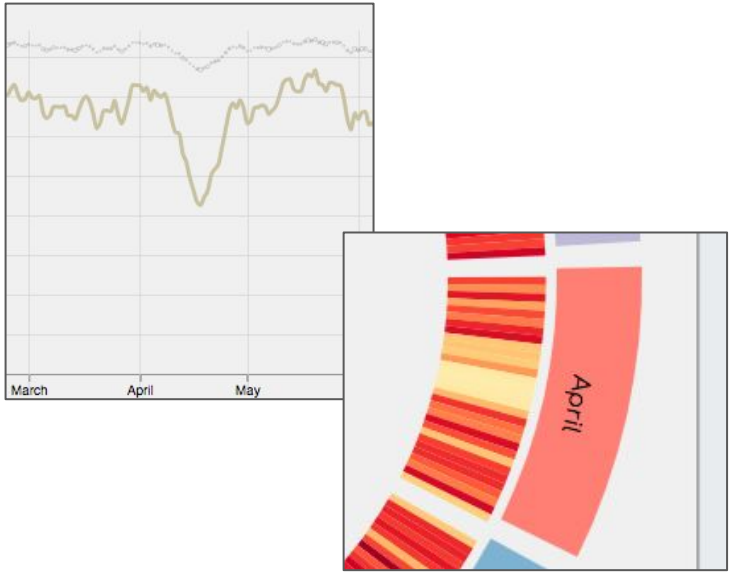
Tech

- Visual Analytics Architecture: Actual Implementation
  - Dynamic SQL

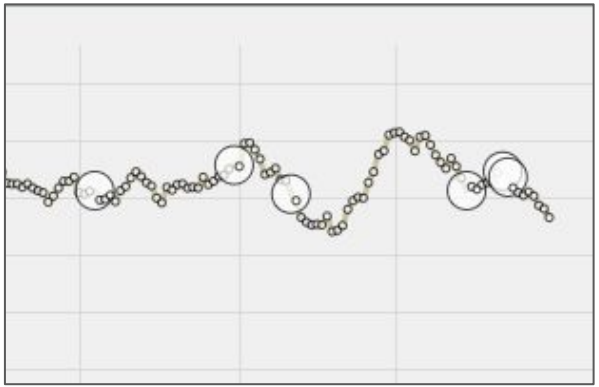
```
let query = ` with events as (
  SELECT to_char(event_time, 'YYYY-MM-DD')  as event_day, cartodb_id as id, totalmuertos
  FROM kudosg.accidentes_bta where 1 = 1 ${localityFilter}  ${severityFilter} ${dayFilter}
    )
  , events_agg as (
   select event_day, count(id) as total , sum(totalmuertos) as totalmuertos
   from events
  group by event_day
  order by event_day
  )
  select  event_day, total, totalmuertos, avg(total) over
    (order by event_day rows between ${$scope.slider.value} preceding and 0 following )::float as trend
  from events_agg
  order by 1 asc `

//$log.log(query);
$scope.trends = d3.json($scope.remoteServiceUrl + query).then($scope.parseTrends);
}
```

Insights

- There is a significant drop in traffic accidents during holy week. It should be correlated with more accidents in tourist places outside Bogotá.

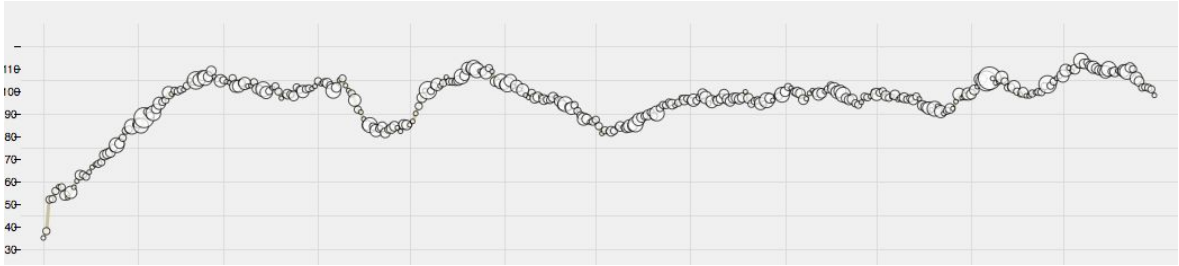


- There are data quality issues (events with injuries *AND* deaths)



Insights

- In general, traffic accidents are increasing in all localities, but the timeframe of 1 year is not enough to give more conclusions.



- Time series looks different between localities. It probably is related to land use. It would be useful to include spatial data for subsequent analysis.

Difficulties

- circosJS has few documentation
- Create Responsive charts with D3 can be a little bit tricky

Achievements

- Learn about D3's transitions.
- Implementation of Visual Analytics Reference Architecture

Thanks