Exploring Open Government Data using Neo4j

Robin Bramley
Contents

• Open Government Data
• Deriving the graph model
• ETL Cypher-style
• What does the graph reveal?
Open Government Data
Find data published by government departments and agencies, public bodies and local authorities. You can use this data to learn more about how government works, carry out research or build applications and services.
Road Safety Data

Published by Department for Transport. Licensed under OGL Open Government Licence.

These files provide detailed road safety data about the circumstances of personal injury road accidents in GB from 1979, the types (including Make and Model) of vehicles involved and the consequential casualties. The statistics relate only to personal injury accidents on public roads that are reported to the police, and subsequently recorded, using the STATS19 accident reporting form.

All the data variables are coded rather than containing textual strings. The lookup tables are available in the "Additional resources" section towards the bottom of the table.

Please note that the 2015 data were revised on the 29th September 2016.


Also includes: Results of breath-test screening data from recently introduced digital breath testing devices, as provided by Police Authorities in England and Wales

Results of blood alcohol levels (milligrams / 100 millilitres of blood) provided by matching coroners' data (provided by Coroners in England and Wales and by Procurators...
Openness Scoring

**Reason:** Content of file appeared to be format "CSV" which receives openness score: 3.

- ★★★★★ Linked data - data URIs and linked to other data (e.g. RDF)
- ★★★★☆ Linkable data - served at URIs (e.g. RDF)
- ★★★☆☆ Structured data in open format (e.g. CSV)
- ★★★☆☆☆ Structured data but proprietry format (e.g. Excel)
- ★☆☆☆☆ Unstructured data (e.g. PDF)

**Score updated:** 26/03/2017

**BRONZE:** data is openly licensed, available with no restrictions, accessible and legally reusable.
Deriving the graph model

Some simple rules
1. Know your data!

 Subject to local directions, boxes with a gray background need not be completed if already recorded

© Copyright 2017 Ixxus Ltd. Commercial in Confidence

Source: STATS19 Form
2. Know your data format

<table>
<thead>
<tr>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Acc_Index</td>
<td>accru</td>
<td>Vehicle_Reference</td>
<td>Vehicle_Type</td>
<td>Towing_and_Articulation</td>
</tr>
<tr>
<td>2</td>
<td>201101BS70001</td>
<td>2011</td>
<td>1</td>
<td>9</td>
<td>0</td>
</tr>
<tr>
<td>3</td>
<td>201101BS70002</td>
<td>2011</td>
<td>2</td>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td>4</td>
<td>201101BS70003</td>
<td>2011</td>
<td>11</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>5</td>
<td>201501BS70004</td>
<td>2015</td>
<td>1,9,0,9,0,2,0,0,0,0,1,1,6,1,6,1668,2,10,-1,1,Vauxhall</td>
<td>1,9,0,9,0,2,0,0,0,0,1,1,6,1,6,1668,2,10,-1,1,Vauxhall</td>
<td>1,9,0,9,0,2,0,0,0,0,1,1,6,1,6,1668,2,10,-1,1,Vauxhall</td>
</tr>
<tr>
<td>6</td>
<td>201501BS70009</td>
<td>2015</td>
<td>1,3,0,18,0,8,0,0,0,1,1,6,1,7,124,1,3,-1,1,Yamaha</td>
<td>1,3,0,18,0,8,0,0,0,1,1,6,1,7,124,1,3,-1,1,Yamaha</td>
<td>1,3,0,18,0,8,0,0,0,1,1,6,1,7,124,1,3,-1,1,Yamaha</td>
</tr>
<tr>
<td>7</td>
<td>201501BS70010</td>
<td>2015</td>
<td>1,9,0,9,0,2,0,0,0,0,1,1,6,1,7,124,1,3,-1,1,Renault</td>
<td>1,9,0,9,0,2,0,0,0,0,1,1,6,1,7,124,1,3,-1,1,Renault</td>
<td>1,9,0,9,0,2,0,0,0,0,1,1,6,1,7,124,1,3,-1,1,Renault</td>
</tr>
<tr>
<td>8</td>
<td>201501BS70011</td>
<td>2015</td>
<td>2,9,0,18,0,8,0,0,0,0,1,1,6,1,7,1995,2,3,-1,1,Mini</td>
<td>2,9,0,18,0,8,0,0,0,0,1,1,6,1,7,1995,2,3,-1,1,Mini</td>
<td>2,9,0,18,0,8,0,0,0,0,1,1,6,1,7,1995,2,3,-1,1,Mini</td>
</tr>
<tr>
<td>9</td>
<td>201501BS70012</td>
<td>2015</td>
<td>1,9,0,13,0,8,0,0,0,0,1,1,6,2,7,124,2,8,-1,1,Vauxhall</td>
<td>1,9,0,13,0,8,0,0,0,0,1,1,6,2,7,124,2,8,-1,1,Vauxhall</td>
<td>1,9,0,13,0,8,0,0,0,0,1,1,6,2,7,124,2,8,-1,1,Vauxhall</td>
</tr>
<tr>
<td>10</td>
<td>201501BS70012</td>
<td>2015</td>
<td>2,5,0,7,0,8,0,0,0,0,1,1,6,2,7,124,2,8,-1,1,Aprilia</td>
<td>2,5,0,7,0,8,0,0,0,0,1,1,6,2,7,124,2,8,-1,1,Aprilia</td>
<td>2,5,0,7,0,8,0,0,0,0,1,1,6,2,7,124,2,8,-1,1,Aprilia</td>
</tr>
<tr>
<td>11</td>
<td>201501BS70013</td>
<td>2015</td>
<td>1,8,0,18,0,8,0,0,0,0,1,1,6,1,7,1598,2,5,-1,1,Skoda</td>
<td>1,8,0,18,0,8,0,0,0,0,1,1,6,1,7,1598,2,5,-1,1,Skoda</td>
<td>1,8,0,18,0,8,0,0,0,0,1,1,6,1,7,1598,2,5,-1,1,Skoda</td>
</tr>
<tr>
<td>12</td>
<td>201501BS70013</td>
<td>2015</td>
<td>2,8,0,9,0,8,0,0,0,0,3,1,1,6,1,7,1978,1,6,-1,1,Toyota</td>
<td>2,8,0,9,0,8,0,0,0,0,3,1,1,6,1,7,1978,1,6,-1,1,Toyota</td>
<td>2,8,0,9,0,8,0,0,0,0,3,1,1,6,1,7,1978,1,6,-1,1,Toyota</td>
</tr>
<tr>
<td>13</td>
<td>201501BS70014</td>
<td>2015</td>
<td>1,9,0,8,0,8,2,10,0,0,4,1,1,6,1,7,1461,2,11,-1,1,Renault</td>
<td>1,9,0,8,0,8,2,10,0,0,4,1,1,6,1,7,1461,2,11,-1,1,Renault</td>
<td>1,9,0,8,0,8,2,10,0,0,4,1,1,6,1,7,1461,2,11,-1,1,Renault</td>
</tr>
<tr>
<td>14</td>
<td>201501BS70016</td>
<td>2015</td>
<td>1,4,0,18,0,8,0,0,0,0,1,1,6,2,6,124,1,8,-1,1,Piaggio</td>
<td>1,4,0,18,0,8,0,0,0,0,1,1,6,2,6,124,1,8,-1,1,Piaggio</td>
<td>1,4,0,18,0,8,0,0,0,0,1,1,6,2,6,124,1,8,-1,1,Piaggio</td>
</tr>
<tr>
<td>15</td>
<td>201501BS70018</td>
<td>2015</td>
<td>1,11,0,4,0,1,0,0,0,0,4,1,1,7,4500,2,3,-1,1,Alexander Dennis</td>
<td>1,11,0,4,0,1,0,0,0,0,4,1,1,7,4500,2,3,-1,1,Alexander Dennis</td>
<td>1,11,0,4,0,1,0,0,0,0,4,1,1,7,4500,2,3,-1,1,Alexander Dennis</td>
</tr>
<tr>
<td>16</td>
<td>201501BS70019</td>
<td>2015</td>
<td>1,11,0,4,0,1,0,0,0,0,4,1,1,7,4500,2,3,-1,1,Alexander Dennis</td>
<td>1,11,0,4,0,1,0,0,0,0,4,1,1,7,4500,2,3,-1,1,Alexander Dennis</td>
<td>1,11,0,4,0,1,0,0,0,0,4,1,1,7,4500,2,3,-1,1,Alexander Dennis</td>
</tr>
<tr>
<td>17</td>
<td>201501BS70022</td>
<td>2015</td>
<td>1,9,0,18,0,8,0,0,0,0,1,1,6,1,7,12435,1,14,-1,1,Volvo</td>
<td>1,9,0,18,0,8,0,0,0,0,1,1,6,1,7,12435,1,14,-1,1,Volvo</td>
<td>1,9,0,18,0,8,0,0,0,0,1,1,6,1,7,12435,1,14,-1,1,Volvo</td>
</tr>
<tr>
<td>18</td>
<td>201501BS70023</td>
<td>2015</td>
<td>2,8,0,6,0,8,0,0,0,0,4,1,1,7,680,1,7,-1,1,Honda</td>
<td>2,8,0,6,0,8,0,0,0,0,4,1,1,7,680,1,7,-1,1,Honda</td>
<td>2,8,0,6,0,8,0,0,0,0,4,1,1,7,680,1,7,-1,1,Honda</td>
</tr>
<tr>
<td>19</td>
<td>201501BS70024</td>
<td>2015</td>
<td>1,5,0,18,0,8,0,0,0,0,4,1,1,7,820,2,10,-1,1,Volvo</td>
<td>1,5,0,18,0,8,0,0,0,0,4,1,1,7,820,2,10,-1,1,Volvo</td>
<td>1,5,0,18,0,8,0,0,0,0,4,1,1,7,820,2,10,-1,1,Volvo</td>
</tr>
<tr>
<td>20</td>
<td>201501BS70029</td>
<td>2015</td>
<td>2,9,0,13,0,8,0,0,0,0,4,1,1,7,1390,1,7,-1,1,Volkswagen</td>
<td>2,9,0,13,0,8,0,0,0,0,4,1,1,7,1390,1,7,-1,1,Volkswagen</td>
<td>2,9,0,13,0,8,0,0,0,0,4,1,1,7,1390,1,7,-1,1,Volkswagen</td>
</tr>
</tbody>
</table>
3. Know your *lookup* data!

**ADDITIONAL LINKS (5)**

- **Road Safety - Blood Alcohol Content Data Guide**
  - **XLS**
  - [Details](#) [Download XLS](#) (17.9 kB)

- **Road Safety - Digital Breath Test Data Guide**
  - **XLS**
  - [Details](#) [Download XLS](#) (17.9 kB)

- **Published statistics and supporting documents**
  - **HTML**
  - [Details](#) [Link](#)

- **Brief guide to road accidents and safety data**
  - **DOC**
  - [Details](#) [Download doc](#) (50.2 kB)

- **Lookup up tables for variables**
  - **XLS**
  - [Details](#) [Download XLS](#) (820.2 kB)
4. Decide what’s important...

Data you don’t have...

Data you have
Simple rules recap

1. Know your data!
2. Know your data *format*!
3. Know your *lookup* data!
4. Decide what’s important...

Then start doodling 😊
Implementing ETL in Cypher
Steps:

1. Convert to Cypher
2. Load your lookup data
3. Profile your queries
4. Optimise
5. Run
Demo start state
Optimise: 141.6k rows loaded in 1 hour


Cypher version: CYpher 3.0, planner: COST, runtime: INTERPRETED.

© Copyright 2017 Ixxus Ltd. Commercial in Confidence
USING PERIODIC COMMIT 1000
LOAD CSV FROM "file:///MakeModel2015v2.csv" AS csvLine
MATCH (g:Gender), (a:AgeBand), (vt:VehicleType), (p:Propulsion)
WHERE g.code = csvLine[15]
AND a.code = csvLine[16]
AND vt.code = csvLine[3]
AND p.code = csvLine[18]
MERGE (mf:Manufacturer { name : RTRIM(csvLine[22]) } )
MERGE (i:Incident { ref : csvLine[0] } )
CREATE (v:Vehicle { incidentRef : csvLine[0], index : csvLine[2],
model : RTRIM(csvLine[23]), age : csvLine[19], capacity : 
csvLine[17] } )
CREATE (v)-[:MADE_BY]->(mf)
CREATE (v)-[:OF_TYPE]->(vt)
CREATE (v)-[:PROPULSION]->(p)
CREATE (v)<-[:INVOLVED]-(i)
CREATE (v)-[:DRIVER_AGE]->(a)
CREATE (v)-[:DRIVER_GENDER]->(g);

Optimised query: ~50 seconds
Neo4j provides:

• A labelled property graph
• Built-in ETL capabilities (e.g. LOAD CSV)
• An intuitive query language (Cypher)
• Database browser web application
  • Complete with data visualisation