

SPARQL queries on statistics.gov.scot

Three Scenarios to fetch its data

Standard Dimension: Time (years)

Two Topics: Environment (Scenario 1) & Housing (Scenario 2 & 3)

Scenario	First Row	Second Row	Third Row	Fourth Row	[...]
1	Year	Percent of renewable energy of the total energy consumption	Total emissions (Metric Tons of CO2eq)	Transport emissions (Metric Tons of CO2eq)	
2	Year	Count_of_Births	Mean Price of a house in GBP		
3	Year	Count_of_Births	Number of Dwellings in Scotland (Country)	Number of Dwellings in Aberdeenshire (Council Area)	Number of Dwellings in [...] (Council Area)

Scenario 1

Scenario	First Row	Second Row	Third Row	Fourth Row	[...]
1	Year	Percent of renewable energy of the total energy consumption	Total emissions (Metric Tons of CO2eq)	Transport emissions (Metric Tons of CO2eq)	

We are trying to figure out if renewable energy led to lower carbon dioxide emissions.

With the accession of the emissions from transport we will try also to conclude if transport became “greener” than it was thanks to renewable energy (of course is not the only one variable which affects transport emissions).

Comparing with the best countries in this field (The Netherlands) we will try to understand if Scotland has become better in this field and if there are any prospects in “green transport”.

Scenario 1 - the query (1)

```
SELECT ?Year ?Percent_of_renewable ?Metric_Tons_of_CO2eq ?Transport_Emissions_Metric_Tons_of_CO2eq
WHERE {
#We fetch the first dataset of renewable energy consumption#
?a qb:dataSet <http://statistics.gov.scot/data/renewable-electricity>;
  <http://purl.org/linked-data/sdmx/2009/dimension#refArea> <http://statistics.gov.scot/id/statistical-geography/S92000003>;
#We assing the triplets to our desired variable#
  <http://statistics.gov.scot/def/measure-properties/ratio> ?Percent_of_renewable;
#We assign the available years in a dummy variable#
  <http://purl.org/linked-data/sdmx/2009/dimension#refPeriod> ?dummy1.
#And we get from this variable the actual string#
  ?dummy1 rdfs:label ?Year.

#we follow the same procedure for the next dataset but,#
  ?b qb:dataSet <http://statistics.gov.scot/data/greenhouse-gas>;
  <http://purl.org/linked-data/sdmx/2009/dimension#refArea> <http://statistics.gov.scot/id/statistical-geography/S92000003>;
  <http://statistics.gov.scot/def/measure-properties/count> ?Metric_Tons_of_CO2eq;
#We link the second dataset to the reference_periods of the first one (!)#
  <http://purl.org/linked-data/sdmx/2009/dimension#refPeriod> ?dummy1.
```

[...]

Scenario 1 - the query (2)

[...]

#in the same way for the third dataset (emissions of transport)#

?c qb:dataSet <http://statistics.gov.scot/data/greenhouse-gas-emissions-by-source-sector>;

<http://purl.org/linked-data/sdmx/2009/dimension#refArea> <http://statistics.gov.scot/id/statistical-geography/S92000003>;

<http://statistics.gov.scot/def/dimension/greenhouseGasSourceSector> <http://statistics.gov.scot/def/concept/greenhouse-gas-source-sector-transport-excluding-international-aviation-and-shipping>;

<http://statistics.gov.scot/def/dimension/pollutant> <http://statistics.gov.scot/def/concept/pollutant/co2>;

<http://statistics.gov.scot/def/measure-properties/count> ?Transport_Emissions_Metric_Tons_of_CO2eq;

#We link the third dataset to the reference periods of the first one#

<http://purl.org/linked-data/sdmx/2009/dimension#refPeriod> ?dummy1.

}

#We order the results by year#

ORDER BY ASC(?Year)

LIMIT 100

Scenario 1 - the results

Year	Percent_of_renewable	Metric_Tons_of_CO2eq	Transport_Emissions_Metric_Tons_of_CO2eq
2000	12.2	74.712	10.864
2001	10.4	73.879	10.846
2002	12.3	69.489	11.174
2003	9	69.506	11.213
2004	14.1	67.196	11.315
2005	15.5	65.856	11.468
2006	16.8	68.989	11.647
2007	19.7	64.163	11.831
2008	22.1	61.867	11.395
2009	27.2	57.247	10.94
<div><div>LOAD MORE ROWS</div><div>Ready</div></div>			

Scenario 2

Scenario	First Row	Second Row	Third Row
2	Year	Count_of_Births	Mean Price of a house in GBP

In this dataset we will try to predict if the total counts of births in Scotland can affect the mean price of housing. Of course this approach might be ingenuous as the mean price is affected from many other variables (and also immigrant tendencies).

However we will try figure out if there is a significant relation between these two variables. Is worth to mention that count of births might not affect the demand in the real estate, but maybe is an indicator of a future demand (in 30 years for example).

Scenario 2 - the query

```
SELECT ?Year ?Count_of_births ?Mean_Price_of_a_house_in_GBP
WHERE {
  ?a qb:dataSet <http://statistics.gov.scot/data/births>;
    <http://statistics.gov.scot/def/dimension/gender> <http://statistics.gov.scot/def/concept/gender/all>;
    <http://statistics.gov.scot/def/dimension/timePeriod> <http://statistics.gov.scot/def/concept/time-period/calendar-year>;
    <http://purl.org/linked-data/sdmx/2009/dimension#refArea> <http://statistics.gov.scot/id/statistical-geography/S92000003>;
    <http://statistics.gov.scot/def/measure-properties/count> ?Count_of_births.
  ?a <http://purl.org/linked-data/sdmx/2009/dimension#refPeriod> ?dummy1.
    ?dummy1 rdfs:label ?Year.

  ?b qb:dataSet <http://statistics.gov.scot/data/house-sales-prices>;
    <http://purl.org/linked-data/sdmx/2009/dimension#refArea> <http://statistics.gov.scot/id/statistical-geography/S92000003>;
    <http://statistics.gov.scot/def/measure-properties/mean> ?Mean_Price_of_a_house_in_GBP;

    #We link the second dataset to the ref_periods of the first one(!)#
    <http://purl.org/linked-data/sdmx/2009/dimension#refPeriod> ?dummy1.

}
ORDER BY ASC(?Year)
LIMIT 100
```


Scenario 2 - the results

Scenario	First Row	Second Row	Third Row
2	Year	Count_of_Births	Mean Price of a house in GBP
	2006	55690	144112
	2007	57781	160736
	2008	60041	165214
	2009	59046	159075
	2010	58791	163396
	2011	58590	163264
	2012	58027	158491

A quick view in the results shows that both variables decreased in 2009 (financial crisis). However, the price of housing became unstable afterwards.

Scenario 3

Scenario	First Row	Second Row	Third Row	Fourth Row	[...]
3	Year	Count_of_Births	Number of Dwellings in Scotland (Country)	Number of Dwellings in Aberdeenshire (Council Area)	Number of Dwellings in [...] (Council Area)

In this dataset we will try to figure out if the total counts of births in Scotland are related with the number of dwellings in Scotland, but also in the 3 most populated council areas.

Scenario 3 - the query

```
SELECT ?Year ?Count_of_births ?Dwellings_Scotland ?Dwellings_Aberdeenshire ?Dwellings_City_of_Edinburgh ?Dwellings_Glasgow_City
WHERE {
```

```
?a qb:dataSet <http://statistics.gov.scot/data/births>;
  <http://statistics.gov.scot/def/dimension/gender> <http://statistics.gov.scot/def/concept/gender/all>;
  <http://statistics.gov.scot/def/dimension/timePeriod> <http://statistics.gov.scot/def/concept/time-period/calendar-year>;
  <http://purl.org/linked-data/sdmx/2009/dimension#refArea> <http://statistics.gov.scot/id/statistical-geography/S92000003>;
  <http://statistics.gov.scot/def/measure-properties/count> ?Count_of_births.
?a <http://purl.org/linked-data/sdmx/2009/dimension#refPeriod> ?dummy1.
  ?dummy1 rdfs:label ?Year.
```

```
?b qb:dataSet <http://statistics.gov.scot/data/dwellings-council-tax>;
  <http://purl.org/linked-data/cube#measureType> <http://statistics.gov.scot/def/measure-properties/count>;
  <http://statistics.gov.scot/def/dimension/councilTaxBand> <http://statistics.gov.scot/def/concept/council-tax-band/total-dwellings>;
  <http://purl.org/linked-data/sdmx/2009/dimension#refPeriod> ?dummy1;
  <http://purl.org/linked-data/sdmx/2009/dimension#refArea> <http://statistics.gov.scot/id/statistical-geography/S92000003>;
  <http://statistics.gov.scot/def/measure-properties/count> ?Dwellings_Scotland.
```

#we fetch each geographical area seperately#

```
?c qb:dataSet <http://statistics.gov.scot/data/dwellings-council-tax>;
  <http://purl.org/linked-data/cube#measureType> <http://statistics.gov.scot/def/measure-properties/count>;
  <http://statistics.gov.scot/def/dimension/councilTaxBand> <http://statistics.gov.scot/def/concept/council-tax-band/total-dwellings>;
  <http://purl.org/linked-data/sdmx/2009/dimension#refPeriod> ?dummy1;
  <http://purl.org/linked-data/sdmx/2009/dimension#refArea> <http://statistics.gov.scot/id/statistical-geography/S12000034>;
  <http://statistics.gov.scot/def/measure-properties/count> ?Dwellings_Aberdeenshire.
```


Scenario 3 - the query

```
?d qb:dataSet <http://statistics.gov.scot/data/dwellings-council-tax>;
  <http://purl.org/linked-data/cube#measureType> <http://statistics.gov.scot/def/measure-properties/count>;
  <http://statistics.gov.scot/def/dimension/councilTaxBand> <http://statistics.gov.scot/def/concept/council-tax-band/total-dwellings>;
  <http://purl.org/linked-data/sdmx/2009/dimension#refPeriod> ?dummy1;
  <http://purl.org/linked-data/sdmx/2009/dimension#refArea> <http://statistics.gov.scot/id/statistical-geography/S12000036>;
  <http://statistics.gov.scot/def/measure-properties/count> ?Dwellings_City_of_Edinburgh.

?e qb:dataSet <http://statistics.gov.scot/data/dwellings-council-tax>;
  <http://purl.org/linked-data/cube#measureType> <http://statistics.gov.scot/def/measure-properties/count>;
  <http://statistics.gov.scot/def/dimension/councilTaxBand> <http://statistics.gov.scot/def/concept/council-tax-band/total-dwellings>;
  <http://purl.org/linked-data/sdmx/2009/dimension#refPeriod> ?dummy1;
  <http://purl.org/linked-data/sdmx/2009/dimension#refArea> <http://statistics.gov.scot/id/statistical-geography/S12000046>;
  <http://statistics.gov.scot/def/measure-properties/count> ?Dwellings_Glasgow_City.

}
ORDER BY ASC(?Year)
LIMIT 100
```

But can we find a way to avoid repetitive lines for each geographical area?

Scenario 3 - the use of OPTIONAL (1)

We will try to create the same query with the use of OPTIONAL

```
SELECT ?Year ?Count_of_births ?Dwellings_Scotland ?Dwellings_Aberdeenshire ?Dwellings_City_of_Edinburgh ?Dwellings_Glasgow_City
WHERE {
```

```
?a qb:dataSet <http://statistics.gov.scot/data/births>;
  <http://statistics.gov.scot/def/dimension/gender> <http://statistics.gov.scot/def/concept/gender/all>;
  <http://statistics.gov.scot/def/dimension/timePeriod> <http://statistics.gov.scot/def/concept/time-period/calendar-year>;
  <http://purl.org/linked-data/sdmx/2009/dimension#refArea> <http://statistics.gov.scot/id/statistical-geography/S92000003>;
  <http://statistics.gov.scot/def/measure-properties/count> ?Count_of_births.
?a <http://purl.org/linked-data/sdmx/2009/dimension#refPeriod> ?dummy1.
  ?dummy1 rdfs:label ?Year.
```

```
?b qb:dataSet <http://statistics.gov.scot/data/dwellings-council-tax>;
  <http://purl.org/linked-data/cube#measureType> <http://statistics.gov.scot/def/measure-properties/count>;
  <http://statistics.gov.scot/def/dimension/councilTaxBand> <http://statistics.gov.scot/def/concept/council-tax-band/total-dwellings>;
  <http://purl.org/linked-data/sdmx/2009/dimension#refPeriod> ?dummy1.
```

[...]

Scenario 3 - the use of OPTIONAL (2)

[...]

#And we make seperate OPTIONAL requests for every territory#

```
OPTIONAL{?b <http://purl.org/linked-data/sdmx/2009/dimension#refArea> <http://statistics.gov.scot/id/statistical-geography/S92000003  
<http://statistics.gov.scot/def/measure-properties/count> ?Dwellings_Scotland. }
```

```
OPTIONAL{?b <http://purl.org/linked-data/sdmx/2009/dimension#refArea> <http://statistics.gov.scot/id/statistical-geography/S1200003  
<http://statistics.gov.scot/def/measure-properties/count> ?Dwellings_Aberdeenshire.}
```

```
OPTIONAL{?b <http://purl.org/linked-data/sdmx/2009/dimension#refArea> <http://statistics.gov.scot/id/statistical-geography/S1200003  
<http://statistics.gov.scot/def/measure-properties/count> ?Dwellings_City_of_Edinburgh.}
```

```
OPTIONAL{?b <http://purl.org/linked-data/sdmx/2009/dimension#refArea> <http://statistics.gov.scot/id/statistical-geography/S1200004  
<http://statistics.gov.scot/def/measure-properties/count> ?Dwellings_Glasgow_City.}
```

```
}
```

```
ORDER BY ASC(?Year)
```

```
LIMIT 100
```

Though, if we try to make this query in the endpoint we don't get correct results :(

Scenario 3 - the results
without the use of OPTIONAL

Year	Count_of_births	Dwellings_Scotland	Dwellings_Aberdeenshire	Dwellings_City_of_Edinburgh	Dwellings_Glasgow_City
2003	52432	2348611	99672	219151	287632
2004	53957	2382158	101851	222702	293719
2005	54386	2402785	103270	224700	294650
2006	55690	2424049	104573	226934	296644
2007	57781	2447256	105922	228535	299251
2008	60041	2465998	107142	230903	299388
2009	59046	2479954	108457	232303	298972
2010	58791	2493838	109955	234001	299459

Scenario 3 - Idea for an extended query

In the same approach we could also fetch the data for the births in each territory (council area).

By this way we will be able to make comparisons between the council areas and the tendencies of the country.

Can we aggregate counts of quartiles?

1998-Q2

Data

API

URI

This is a linked data resource: it has a permanent unique uri at which both humans and machines can find it on the Internet, and which can be used as an identifier in queries on our [SPARQL endpoint](#).

`http://reference.data.gov.uk/id/quarter/1998-Q2`

ALL METADATA

In Graph

`http://statistics.gov.scot/graph/vocab/external-ontology/reference-time`

<div>Has beginning</div> <div><code>http://www.w3.org/2006/time#hasBeginning</code></div>	<div>1998-04-01 ▶</div> <div><code>http://reference.data.gov.uk/id/gregorian-instant/1998-04-01T00:00:00</code></div>
<div>Has end</div> <div><code>http://www.w3.org/2006/time#hasEnd</code></div>	<div>1998-07-01 ▶</div> <div><code>http://reference.data.gov.uk/id/gregorian-instant/1998-07-01T00:00:00</code></div>
<div>Label <code>rdfs:label</code></div> <div><code>http://www.w3.org/2000/01/rdf-schema#label</code></div>	<div>1998-Q2 <code>xsd:string</code></div>
<div>Type <code>rdf:type</code></div> <div><code>http://www.w3.org/1999/02/22-rdf-syntax-ns#type</code></div>	<div><code>http://reference.data.gov.uk/def/intervals/Interval</code></div> <div>Time interval ▶ <code>http://www.w3.org/2006/time#Interval</code></div>

1998-Q3

Data

API

URI

This is a linked data resource: it has a permanent unique uri at which both humans and machines can find it on the Internet, and which can be used as an identifier in queries on our [SPARQL endpoint](#).

`http://reference.data.gov.uk/id/quarter/1998-Q3`

ALL METADATA

In Graph

`http://statistics.gov.scot/graph/vocab/external-ontology/reference-time`

<div>Has beginning</div> <div><code>http://www.w3.org/2006/time#hasBeginning</code></div>	<div>1998-07-01 ▶</div> <div><code>http://reference.data.gov.uk/id/gregorian-instant/1998-07-01T00:00:00</code></div>
<div>Has end</div> <div><code>http://www.w3.org/2006/time#hasEnd</code></div>	<div>1998-10-01 ▶</div> <div><code>http://reference.data.gov.uk/id/gregorian-instant/1998-10-01T00:00:00</code></div>
<div>Label <code>rdfs:label</code></div> <div><code>http://www.w3.org/2000/01/rdf-schema#label</code></div>	<div>1998-Q3 <code>xsd:string</code></div>
<div>Type <code>rdf:type</code></div> <div><code>http://www.w3.org/1999/02/22-rdf-syntax-ns#type</code></div>	<div><code>http://reference.data.gov.uk/def/intervals/Interval</code></div> <div>Time interval ▶ <code>http://www.w3.org/2006/time#Interval</code></div>

In order to aggregate the counts of quartiles per year, we need to find the appropriate class or predicate which refers to a common object (e.g. in our case maybe year :1998)

Can we aggregate counts of quartiles?

1998-Q2

Data

API

URI

This is a linked data resource: it has a permanent unique uri at which both humans and machines can find it on the Internet, and which can be used an identifier in queries on our [SPARQL endpoint](#).

<http://reference.data.gov.uk/id/quarter/1998-Q2>

ALL METADATA

In Graph

<http://statistics.gov.scot/graph/vocab/external-ontology/reference-time>

<div>Has beginning</div> <div>http://www.w3.org/2006/time#hasBeginning</div>	<div>1998-04-01</div> <div>http://reference.data.gov.uk/id/gregorian-instant/1998-04-01T00:00:00</div>
<div>Has end</div> <div>http://www.w3.org/2006/time#hasEnd</div>	<div>1998-07-01</div> <div>http://reference.data.gov.uk/id/gregorian-instant/1998-07-01T00:00:00</div>
<div>Label</div> <div><div>rdfs:label</div><div>http://www.w3.org/2000/01/rdf-schema#label</div></div>	<div>1998-Q2</div> <div>xsd:string</div>
<div>Type</div> <div><div>rdf:type</div><div>http://www.w3.org/1999/02/22-rdf-syntax-ns#type</div></div>	<div>http://reference.data.gov.uk/def/intervals/Interval</div>
	<div>Time interval</div> <div>http://www.w3.org/2006/time#Interval</div>

1998-Q3

Data

API

URI

This is a linked data resource: it has a permanent unique uri at which both humans and machines can find it on the Internet, and which can be used an identifier in queries on our [SPARQL endpoint](#).

<http://reference.data.gov.uk/id/quarter/1998-Q3>

ALL METADATA

In Graph

<http://statistics.gov.scot/graph/vocab/external-ontology/reference-time>

<div>Has beginning</div> <div>http://www.w3.org/2006/time#hasBeginning</div>	<div>1998-07-01</div> <div>http://reference.data.gov.uk/id/gregorian-instant/1998-07-01T00:00:00</div>
<div>Has end</div> <div>http://www.w3.org/2006/time#hasEnd</div>	<div>1998-10-01</div> <div>http://reference.data.gov.uk/id/gregorian-instant/1998-10-01T00:00:00</div>
<div>Label</div> <div><div>rdfs:label</div><div>http://www.w3.org/2000/01/rdf-schema#label</div></div>	<div>1998-Q3</div> <div>xsd:string</div>
<div>Type</div> <div><div>rdf:type</div><div>http://www.w3.org/1999/02/22-rdf-syntax-ns#type</div></div>	<div>http://reference.data.gov.uk/def/intervals/Interval</div>
	<div>Time interval</div> <div>http://www.w3.org/2006/time#Interval</div>

However, we conclude that there is no a common object between these two subjects “1998-Q2” & “1998-Q3” (except that they belong to the reference-time in the ontology) so maybe there is no a direct way to group and aggregate counts of quartiles.

There is no a distinct object or a class that connect or refers to these individual quartiles.

Maybe a use of a function, with aim to keep and bound the first four digits of the label (1998) could help to create a module in order to group and aggregate different quartiles, but in any case this will fetch data only for a specific year.

For the course of Information Systems Development

Kokovidis Symeon

March 2017