Maastricht University

Building & Mining Knowledge Graphs Project Report

Building knowledge graph for exploring dependencies between suicide, climate change and economics

Maastricht University

Faculty of Science and Engineering

Department of Data Science and Knowledge Engineering

Author

Martyna Mikos, i6155316

1 Introduction

Knowledge graphs (KG) play an important role in answering questions based on information from various sources like texts and databases. Reason for that being the structured knowledge format that allows for efficient querying (Moschitti et al., 2017). Knowledge graphs are without a doubt a valuable tool for uncovering interesting facts about the gathered data.

Literature review shows that there is a relation between the economic factors, climate change and suicide. In this project all of those factor will be thought investigated in order to confirm those relations.

In order to perform the analysis a Knowledge Graph is created by interlinking 3 different datasets, which contain information about number of suicides, temperature and economic data per country over the period of 37 years. Moreover, in order to thoroughly analyse the outputted data some statistical testing will be performed to proof (or disproof) the dependencies between the variables.

2 Definition of Knowledge Graph

In the recent years Knowledge Graphs became more popular and commonly used in research, technology and business. This increase in popularity is caused by the introduction of the Google Knowledge Graph. Since then, the term *Knowledge Graphs* started being widely used and interpreted in many different ways. Therefore, it is worth looking for more generic definition that is not reffering just to the google product.

So what is a Knowledge graph to me?

Knowledge graph starts with a triple $< subject, predicate, object >$, indicating a nodes and edges in the graph making the information linked, structured and easily accessible. It lets the user of the graph understand the data and makes it possible to answer questions about it. For example: *Where does the author of this project study and lives?*. Below there's an example of such graph:

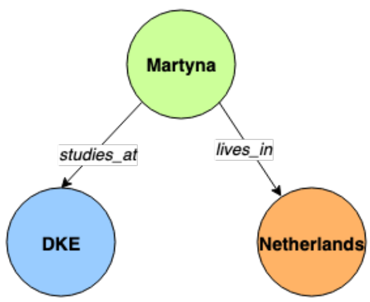

Figure 1: Example of a knowledge graph

Above graph is represented in an RDF in the form of triples defining the relation using URIs or literals:

subject	predicate	object
<http://mylink.com/Martyna>	<http://mylink.com/lives_in>	<http://mylink.com/Netherlands>
<http://mylink.com/Martyna>	<http://mylink.com/studies_at>	<http://mylink.com/DKE>

Moreover, in order to simplify the use of URIs we need to specify the prefix declaration to be able to determine identifiers, literals and datatypes (in case the literals are not a string) also referring to a specific ontology that is a dictionary containing information about properties, types and relationships between the entities. Thanks to this URIs will contain information about the predicates/classes in a formalised, explicit way which makes the definition of things unified in a shared vocabulary.

PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX ont: <http://mylink.com/>
ont:Martyna ont:studies_at ont:DKE .
ont:Martyna rdf:type ont:Student .
ont:DKE rdf:type ont:UMDepartment .
ont:Netherlands ont:lives_in ont:Netherlands .
ont:Netherlands rdf:type ont:Country

To sum up Knowledge Graphs is an organised flexible structure of data with meaning of this data being encoded in the graph as an ontology making it easy to understand what kind of knowledge this data provides. The unified encoding of data in a shared vocabulary makes it easy to communicate between the user and provider of the data.

3 Innovation

There hasn't been much work done on the topic in terms of using a knowledge graph as a tool. Although the topic has been analysed and discussed many times in the papers from economics and social perspective, the idea of using knowledge graphs to analyse the problem hasn't been done before.

4 Problem statement and research questions

Having a tool for continuous analysis of mental health statistics can be beneficial for improving healthcare, support medical decision making and help governments create awareness and develop adequate social and economic care. Therefore it is of high importance to establish protective factors, which are the aspects that cause a decrease in the chance of a person committing suicide.

Having a database that contains all the useful information about the possible risk factors can be a valuable weapon for decreasing the suicide rate. Unfortunately such data is usually scattered in many different data sources or hardly accessible.

Moreover, merging the data sets from different sources can be a challenging task due to following characteristics:

- Heterogenity - problem of achieving integration and interoperability between various knowledge resources.
- Sparsity (missing data) - some data is simply unavaliable for some countries in some years. For example high income countries do not report to Worldbank data like poverty, aid and external debt. Data can be also missing due to political conflicts and due to the fact that some indicators are derived form surveys being performed only every few years.
- Incorrectness and incosistency - resources can contain noisy data and bias.

Furthermore the properties of gathered data are not the only problem encountered when building a knowledge graph. The issue is also the lack of a shared vocabulary focused on economic and social indicators that would be topic-specific, consistent and formalised. Consequently this project will also drag attention to the need of creating such ontology in the future.

In this project the focus will be put on integrating and employing information form three different sources of data and creating a Knowledge Graph to explore relationships and dependencies among different resources like economic indicators, temperature and suicide statistics.

The task is to convert gathered datasets to RDF and integrate them into a Knowledge graph in order to answer the following questions:

1. Does the temperature influence the suicide rate?
2. What other economic/demographic factors can influence that rate?
3. Are there any correlations and dependencies found in the data?
4. Based on the results can we draw any conclusions on how to create awareness and identify risks associated with the suicide rate studied?

5 Significance

It is important to identify and asses suicide risks and ways of protection in order to succeed in suicide prevention. Protective factors and risk identification will be a huge asset in providing changes and public awareness campaigns in order to decrease suicide rates. Those factors are usually identified on various levels like: individual, family, and community (Center and Rodgers, 2011).

In this project the focus will be put on identifying factors out of an individual/family scope. It has been proven that economic factors and climate has also a significant effect on suicide rates. Economist have shown that suicidal behavior is not related only to social isolation and mental illness (or mental illness in family history) but also factors like welfare, inflation and unemployment (Marcotte, 2003). Identification which economic factors influence suicide the most can help the government introduce the protective factors accordingly.

6 Related work

There is a lot of existing work on the topic. Current study shows a strong link between cold climate and suicide. Interestingly this trend has been noticed in Europe from south to north-east. It is also shown that sudden changes of temperature twice a year can cause depression and also longer periods of light can lead to an impulsive suicide act. Link between suicide and temperature is proved to be stronger than the influence of the economic factors. Nonetheless both are proven to be existent. Moreover, the economic factors that seem to influence the suicide rate the most are unemployment and GDP per capita (Fountoulakis et al., 2016).

The methods used in this project are quite commonly used in research papers regarding building knowledge graphs and combining multiple datasets. Usage is spread though various fields like developing KG as a tool for integrating multiple sources of medical data and analysis of health care quality (Huang et al., 2017). Further research shows that KG are also used for analysing enterprise information from many different websites and databases in order to achieve completeness and valuable investment analysis and data visualization (Ruan et al., 2016).

7 Methodology

In order to answer the research questions multiple datasets from different sources were pre-processed, converted to RDF and integrated.

All datasets include data for the time period between 1979 and 2016. They cover statistics by country and year.

The source of the first dataset is WHO Mortality Database, 1979-2016 that contains the suicide data.

Second dataset was retrieved from The World Bank, 1979-2016 and contains the economic indicators listed below:

- Educational attainment, at least completed upper secondary
- Unemployment
- Poverty headcount ratio
- Net migration
- Population
- GNI per capita

The third dataset contains average temperature data per country. Data was retrieved from Climate Change Knowledge Portal, 1979-2016.

Firstly, the datasets were pre-processed. Missing data from the suicide dataset was removed, since its the core of this research. The temperature dataset, contained information about the temperature per month and since all the other attributes are gathered yearly, the average temperature per year was calculated for each country. Furthermore, name of the country in the suicide dataset wasn't matching the ones present in remaining two (e.g. "Czechia" and "Czech Republic"). There were also no ISO code assailable to link them. The solution was to check which names differ and change them in the suicide dataset and assign the ISO codes to each country for the easier linking by using a function in excel - $VLOOKUP$.

Secondly, a mapping file *mapping.ttl* was defined to map the data sources to triples and generate the RDF.

Definition of the entities In order to generate the RDF file triples for each entity were defined. In order to do that the logical source indicating the file path, file format and the class of that entity was indicated.

Another challenging task was to find and select a shared vocabulary in order to define the entities in the Knowledge Graph.

Unfortunately, there is no shared vocabulary concerning strictly economic indicators, therefore a National Cancer Institute vocabulary: NCI Metathesaurus (NCIm) biomedical terminology database was used for almost all entity classes. It contains all the definitions and accurately matches the indicators used in this project.

Other ontologies that were also used are the DBpedia and GeoNames, which were used for the definition of the remaining entities i.e. *db:Year*, *db:Continent*, *db:Country*, *db:GrossDomesticProduct PerCapita*, *db:Population*, *gn:countryCode* due to the fact that NCIm ontology uses codes to refer to predicates, a different ontology was used when available, especially because country code (ISO3) and year was used as a predicate in every triple.

$< CountryTriple >$

Countries were extracted from the WHO dataset, the instances are defined by DBpedia ontology class linking each one to its *GeoNames* web page through the generated URIs.

`http://geonames.org/countries/{codecountry}/{country}.html`

$< YearTriple >$

As there is no URI that links country to the specific year, thus one was created in a following way:

`http://dbpedia.org/page/{year}/{codecountry}`

The *WHO*, *Climate Change* and *World Bank* dataset were much more challenging becasue the specific URIs don't exist. In that case the avaliable APIs were used. Therefore three different APIs were used: GHO API, Climate Data API and World Bank API. All of those follow the same template {countrycode/ISO3} and {year}, having for each indicator a specific code in oder to access the data. All of those are listed below.

$< SuicideTriple >$

`http://apps.who.int/gho/athena/api/GHO/SA_0000001400_ARCHIVED?filter=COUNTRY:{codecountry};YEAR:{year}`

$< TemperatureTriple >$

`http://climatedataapi.worldbank.org/climateweb/rest/v1/country/annualavg/tas/{Year}/{Year}/{ISO3}`

$< UnemploymentTriple >$

`http://api.worldbank.org/v2/country/{countrycode}/indicator/SL.UEM.TOTL.NE.ZS?format=json&date={time}`

$< EducationTriple >$

`http://api.worldbank.org/v2/country/{countrycode}/indicator/SE.SEC.CUAT.UP.ZS?format=json&date={time}`

$< PovertyRatioTriples >$

`http://api.worldbank.org/v2/country/{countrycode}/indicator/SI.POV.NAHC?format=json&date={time}`

$< NetMigrationTriples >$

`http://api.worldbank.org/v2/country/{countrycode}/indicator/SM.POP.NETM?format=json&date={time}`

$< GDPTriples >$

`http://api.worldbank.org/v2/country/{countrycode}/indicator/NY.GDP.PCAP.PP.CD?format=json&date={time}`

$< PopulationTriples >$

`http://api.worldbank.org/v2/country/{countrycode}/indicator/SP.POP.TOTL?format=json&date={time}`

To sum up 10 types of triples were created taking into account how the graph should look like making each of the subject class a node. Therefore, each indicator e.g. suicide, unemployment etc. has its own triple defined.

Adding attributes to the entities Selection of the attributes (predicates) for each triple was performed in the next step. Each entity has come common information like year and country code. The reason is that these two attributes are the common link between the datasets thus are a predicate in each entity.

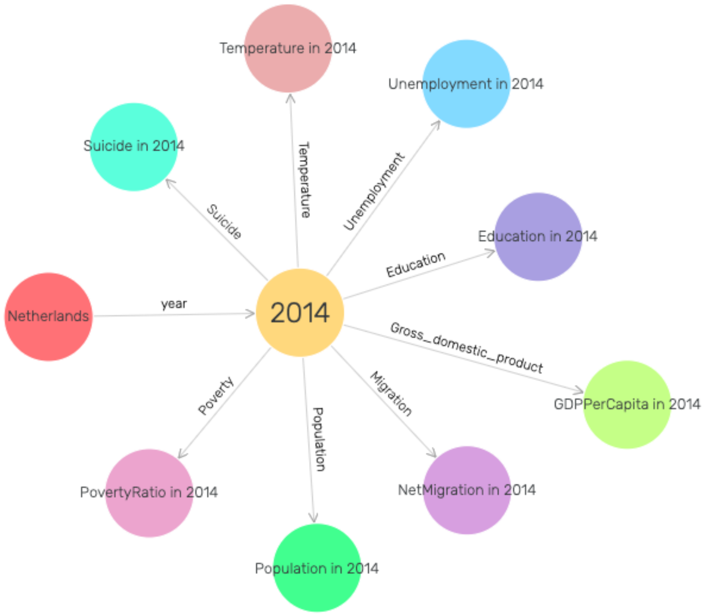

Figure 2: Example of the entities assigned to Netherlands in 2014

Apart from that, a label was created for each subject e.g. "Unemployment in {year}" to indicate, what data is associated with each node and enable a visual search in the graph representation, which can be seen in Figure 2.

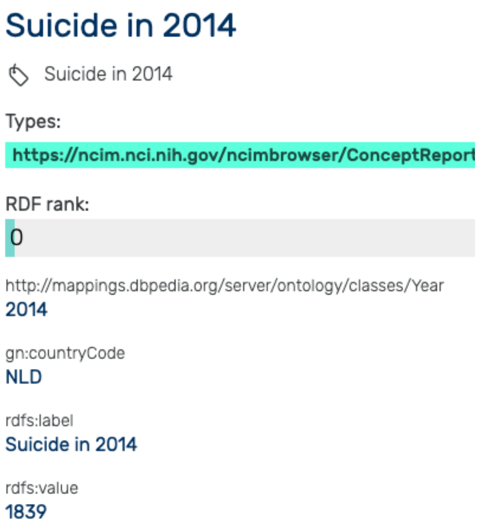

Figure 3: Example of the country with its GDP values.

Moreover, in most cases except the label, the value of an attribute was added indicating the datatype (e.g. float). Example of the information representation is presented in Figure 3.

Datasets linking The next step in generating the Knowledge Graph is linking all the datasets and triples in order to create a relation between the entities and visualise them. This will make querying simple and more powerful.

So as to generate this relationship among the datasets the $INSERT$ operation was performed in order to to add triples to the Graph Store based on bindings between triples as showed in Query 1. The relation between the year and the country code was used to join them together. Next step was to join year with all the indicators. This was also accomplished by using $INSERT$ as shown in Query 2 that links year to the suicide. The remaining triples were joined in the same way that is why the queries are not shown in this report.

The output is the Knowledge Graph with total of 357 427 statements presented in Figure 4. Due to these operation queries can now be performed and research questions answered. Tools that will be used for the graph exploration are SPARQL queries in GraphDB. Moreover STATA will be used to perform statistical tests and analysis and also fitting a regression line to see if the dependencies actually exists.

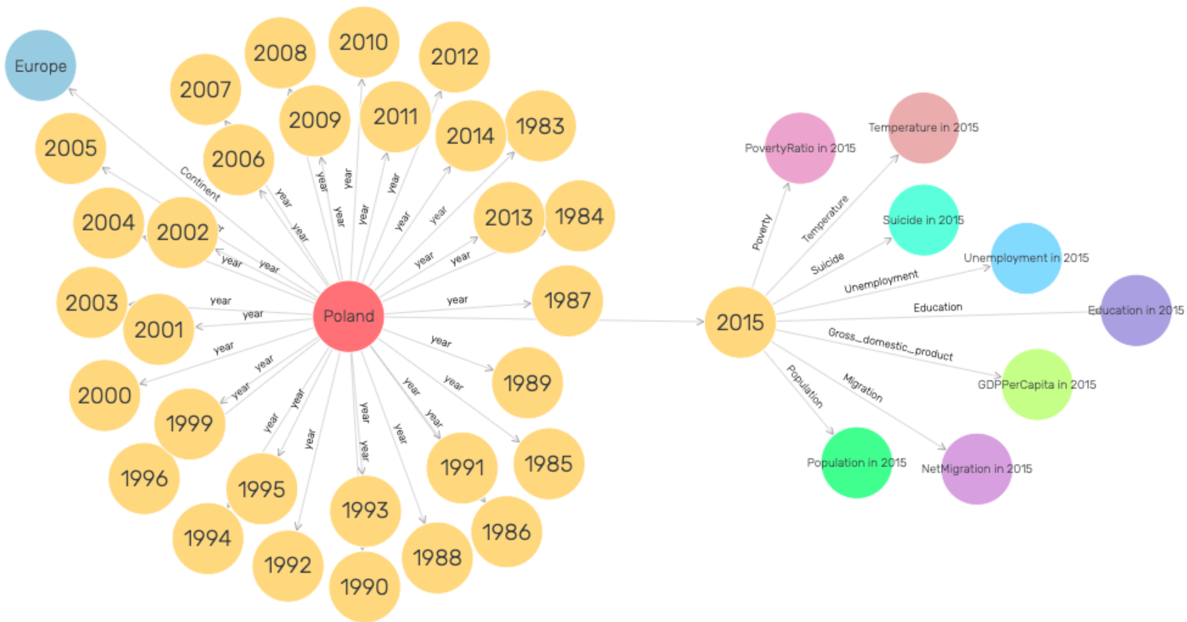

Figure 4: Knowledge Graph created

8 Results & findings

Firstly a very simple query was done to find out which countries have the highest suicide rate. The results are visualised in Figure 5. Interestingly most of the countries are European countries. Through the Query 4 top 20 countries were displayed and confirmed that 13 of them were in fact European. That could also be due to the missing data that is not available for south of Asia and Africa as it can be seen in Figure 6.

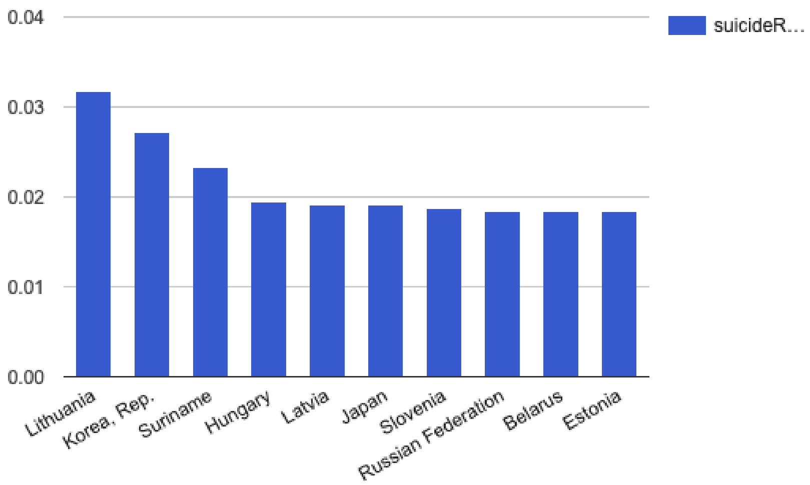

Figure 5: Top 10 countries with highest suicide rate per population (Query 4)

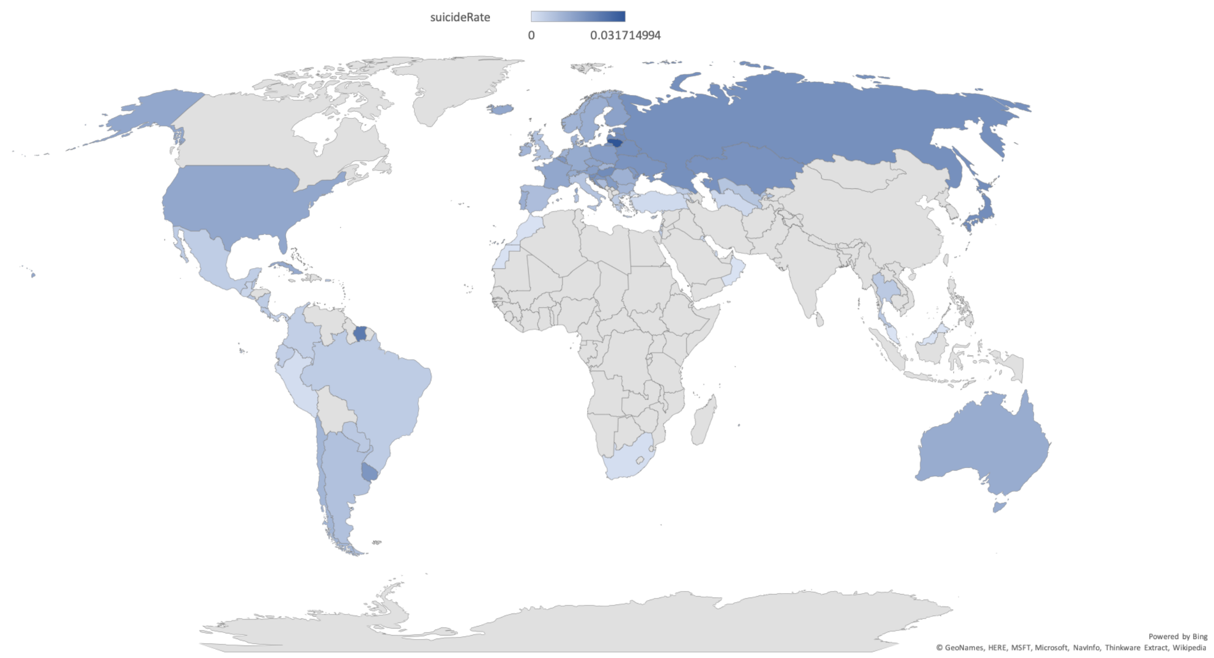

Figure 6: Map of the countries and the suicide rate per population (Query 4)

Next analysis run involved connecting to an endpoint using $SERVICE$ in order to take information about the area and see if population per km^2 is somehow related to the suicide rate. Query 3 was used and generated the table shown in Figure 7. Query is constructed in an unusual way because when connecting to the enpoint all the sources were displayed i.e. DBpedia, GeoNames and more. Data from DBpedia wanted to be used but just filtering did not help because to our surprise DBpedia has sometimes 2 different values for the same country. To solve this problem the maximum area value was chosen. Unfortunately result didn't show any visible relation between the population density and suicide rate.

country_name	suicide	pop_per_km2	suicide_rate
Singapore	355	7606.347118965374	0.006
Malta	32	1375.1835738734178	0.007
Mauritius	120	618.1049042823529	0.01
Korea, Rep.	13,834	506.40317131910984	0.027
Netherlands	1,839	405.9650970223624	0.011
Israel	382	395.55608641078476	0.005
Puerto Rico	220	388.2770158453427	0.006
Belgium	1,899	367.1707130923197	0.017
Japan	24,357	336.733948242346	0.019
El Salvador	386	298.52050470022743	0.006
St. Lucia	14	285.92757843053096	0.008
United Kingdom	4,788	266.45107096835574	0.007
Germany	10,217	226.73503101769475	0.013

Figure 7: Suicides per populated area (Query 3)

Next query checks the difference in data between year 2007 and 2015. The reason for choosing these dates is to capture the effect of the Financial crisis in 2007. The top queried data with respect to suicide difference is presented in Figure 8. Temperature overlaps with GDP and unemployment overlaps with suicide but it can be clearly visible that these changes are correlated with each other especially for countries with the highest suicide rate. We can see that for Russia, Japan and Ukraine there is a positive difference, therefore the amount of suicides dropped significantly for these countries as well as unemployment. On the other hand the GPD greatly increased in the same time. Consequently we can see a trend here that these variables might be in deed correlated. For all the other countries trend is similar: increase in GDP and drop in suicides and unemployment. In all cases except Russia there's also an increase of average temperatures associated with lower amount of suicides. Unfortunately other indicators included in the graph were not used due to the fact that there was too much missing data and it could not be interpreted properly.

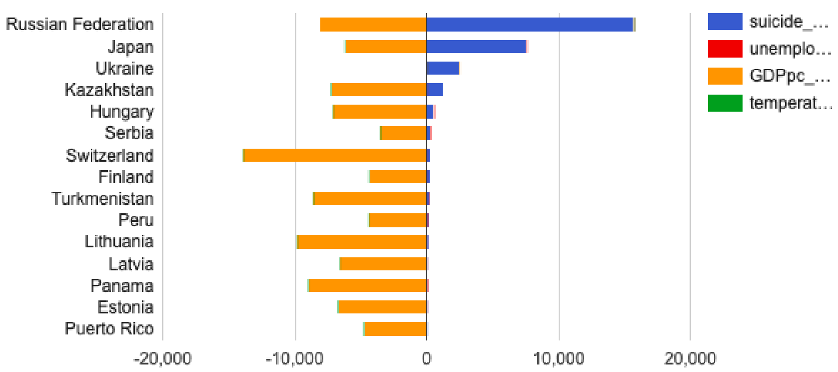

Figure 8: Difference in suicide, unemployment and GDP between 2007 and 2015 (Query 5)

Last step of the project is to perform a statistical analysis on the joined datasets. It was done by running a Query 6 for the whole dataset. For the purpose of detecting dependencies, Spearman correlation measure was used. All the calculations were done in STATA. Figure 9b displays the correlation matrix of the economic indicators just for Europe. Star by the correlation measure indicates that this correlation is significant on the confidence level $\alpha = 5\%$. Looking at this matrix we can see that suicide is negatively correlated with temperature and GDP per capita and positively with unemployment. Of course there's also a positive relation with population, which is expected. This means that if temperature decreases the number of suicides increases. Also if GDP increases the number of suicides drops. It's the opposite for unemployment. When it increases, then so does suicide. Second Figure 9a shows correlation for the whole world in 2014, which shows that only correlation with temperature is the significant one. This shows that such analysis of factors influencing suicide must be continent-specific.

	suicide	temper~e	unempl~t	gdppc	popula~n
suicide	1.0000				
temperature	-0.2721*	1.0000			
unemployment	0.1546	-0.5554*	1.0000		
gdppc	0.1674	0.3099*	-0.3879*	1.0000	
population	0.4681*	-0.4100*	0.1014	-0.3315*	1.0000

(a) Correlation matrix in 2014

	suicide	temper~e	unempl~t	gdppc	popula~n
suicide	1.0000				
temperature	-0.0889*	1.0000			
unemployment	0.1089*	0.2663*	1.0000		
gdppc	-0.1907*	-0.0774*	-0.4595*	1.0000	
population	0.6467*	0.1525*	0.0749*	0.0003	1.0000

(b) Correlation matrix in Europe

Figure 9: Comparison of correlation statistics

Lastly the regression line was fitted for the significant variables. The variables education, poverty and migration had too much missing data therefore were removed. The statistical test also showed that GDP is irrelevant, therefore was removed from this model. Results are presented in Figure 10. To state whether the variable is relevant we used p-value measure ($P > |t|$). All the values are smaller than 5%, which means that the relationship with the dependent variable (suicide) is significant. Moreover we can see that R^2 statistics is equal to 59%. That means that independent variables used for modeling explain the dependent variable in 59%. Lastly, the p-value associated with F statistics ($Prob > F$) is also smaller than $\alpha = 5\%$. We can assume that independent variables predict the dependent one reliably.

Source	SS	df	MS			
				Number of obs	=	2,350
				F(3, 2346)	=	1121.11
Model	7.2522e+10	3	2.4174e+10	Prob > F	=	0.0000
Residual	5.0586e+10	2,346	21562502.8	R-squared	=	0.5891
				Adj R-squared	=	0.5886
Total	1.2311e+11	2,349	52408441.1	Root MSE	=	4643.5

suicide	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
temperature	-116.1639	11.00155	-10.56	0.000	-137.7377	-94.59009
unemployment	-110.6799	17.73782	-6.24	0.000	-145.4633	-75.89644
population	.0001147	2.14e-06	53.68	0.000	.0001105	.0001189
_cons	2907.194	276.0432	10.53	0.000	2365.88	3448.508

Figure 10: Regression line fitting for relevant variables

Based on the above result we can derive a regression equation that is:

$$suicide_i = 2907.2 - 116.2(temperature)_i - 110.7(unemployment)_i + 0.0001(population)_i$$

Interestingly, by fitting the regression line only for Europe, it was discovered that the model explains the dependent variable in almost 80% and GDP is the relevant variable for the model, whereas unemployment is not. The results are shown in Figure 11. Which confirms again that such analysis is most useful when we focus on one continent at a time.

Source	SS	df	MS
Model	4.2940e+10	4	1.0735e+10
Residual	1.1497e+10	753	15268363.5
Total	5.4437e+10	757	71911091.9

Number of obs	=	758
F(4, 753)	=	703.08
Prob > F	=	0.0000
R-squared	=	0.7888
Adj R-squared	=	0.7877
Root MSE	=	3907.5

suicide	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
temperature	-475.5134	31.0792	-15.30	0.000	-536.5256	-414.5012
unemployment	-30.79763	30.58377	-1.01	0.314	-90.83722	29.24195
gdppc	-.0874854	.0095157	-9.19	0.000	-.1061659	-.0688048
population	.0002058	4.79e-06	42.96	0.000	.0001964	.0002152
_cons	5947.304	558.4178	10.65	0.000	4851.063	7043.545

Figure 11: Regression line fitting for relevant variables

9 Conclusions

It this research we asked four research questions, which the thorough statistical analysis and querying helped us answer. Based on the results we can say that temperature in deed influences the suicide rate exhibiting relevant correlation with the variable suicide. Unfortunately due to missing data and sparsity we were unable to determine whether education, poverty and migration has any effect on the suicide rate. However, it was confirmed that there is a strong relation with unemployment and GDP. Results indicate that GDP is an important factor in Europe but not necessarily if we analyse the whole wold data at once. Looking at those results we can identify the risks and show that governmental social help can help prevent (at least to some extent) suicides by providing good social aid for unemployed and poor. Moreover in countries with lower temperatures - if that's in deed a relevant factor in a specific country help can be provided by creating awareness and making counseling and help more accesible.

Nonetheless it was confirmed that those dependencies exist and using a Knowledge Graph is a useful, flexible structure that keeps the vocabulary of the data and makes accessing it very easy. Especially when we have to integrate multiple datasets from different sources.

For the future work some other factors can be investigated and if relevant put into the graph. If kept up to date, cleared of missing data and inconsistencies the KG can be a valuable analysis tool in a field of economics.

References

Center, Suicide Prevention Resource and P. Rodgers (2011). "Understanding risk and protective factors for suicide: A primer for preventing suicide". In: *Newton, MA: Education Development Center, Inc.*

Fountoulakis, Konstantinos N et al. (2016). "Relationship of suicide rates with climate and economic variables in Europe during 2000–2012". In: *Annals of general psychiatry* 15.1, p. 19.

Huang, Zhisheng et al. (2017). "Constructing knowledge graphs of depression". In: *International Conference on Health Information Science*. Springer, pp. 149–161.

Marcotte, Dave E (2003). "The economics of suicide, revisited". In: *Southern Economic Journal*, pp. 628–643.

Moschitti, Alessandro et al. (2017). "Question Answering and Knowledge Graphs". In: *Exploiting Linked Data and Knowledge Graphs in Large Organisations*. Springer International Publishing, pp. 181–212. ISBN: 978-3-319-45654-6.

Ruan, Tong et al. (2016). "Building and exploring an enterprise knowledge graph for investment analysis". In: *International Semantic Web Conference*. Springer, pp. 418–436.

WHO Mortality Database (1979-2016). *Suicide, crude death rate per 100 000, both sexes.* data retrieved from WHO Mortality Database, http://apps.who.int/healthinfo/statistics/mortality/whodpms/.

Climate Change Knowledge Portal (1979-2016). *Average temperature by Country.* data retrieved from Climate Change Knowledge Portal, https://climateknowledgeportal.worldbank.org/download-data.

The World Bank (1979-2016). *Economic indicators.* data retrieved from World Development Indicators, https://data.worldbank.org/indicator.


```
PREFIX db: <http://mappings.dbpedia.org/server/ontology/classes/>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX gn: <http://www.geonames.org/ontology#>

INSERT
{
 ?country db:year ?year
}
WHERE
{
    ?country a db:Country .
    ?country gn:countryCode ?country_code .
    ?country rdfs:label ?country_name .

    ?year a db:Year .
    ?year gn:countryCode ?year_code .
          ?year db:Country ?year_country .
    ?year rdfs:label ?year_label .

    FILTER(?country_code=?year_code)
}
```

Listing 1: Linking country to the year

```
PREFIX db: <http://mappings.dbpedia.org/server/ontology/classes/>
PREFIX dbo: <http://dbpedia.org/page/>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX ncit: <https://ncim.nci.nih.gov/ncimbrowser/ConceptReport.jsp?dictionary=NCI%20Metathesauru
PREFIX gn: <http://www.geonames.org/ontology#>

INSERT
{
    ?year dbo:Suicide ?suicide
}

WHERE
{
    ?suicide a ncit:C0038661 .
    ?suicide rdfs:value ?suicide_val .
    ?suicide gn:countryCode ?suicide_country .
    ?suicide db:Year ?suicide_year .

    ?year a db:Year .
    ?year gn:countryCode ?year_country .
    ?year rdfs:label ?year_val .

    FILTER(?suicide_year=?year_val)
    FILTER(?suicide_country=?year_country)
}
```

Listing 2: Linking year to suicide

```
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX db: <http://mappings.dbpedia.org/server/ontology/classes/>
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX dbpont: <http://dbpedia.org/ontology/>
PREFIX dbo: <http://dbpedia.org/page/>
PREFIX onto: <http://data.ontotext.com/resource/leak/>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
PREFIX gn: <http://www.geonames.org/ontology#>

SELECT DISTINCT
?country_name ?suicide
(xsd:float(?population)*1000000/xsd:double(?maxArea) AS ?pop_per_km2)
(xsd:float(?suicide*100)/xsd:float(?population) AS ?suicide_rate)

WHERE
{
    {
            SELECT DISTINCT ?country_code ?country_name ?suicide ?population
            WHERE
            {
                ?country a db:Country .
                ?country gn:countryCode ?country_code .
                ?country rdfs:label ?country_name .
                ?country db:year ?year.
                ?year rdfs:label 2014 .
                ?year dbo:Suicide ?suic.
                ?suic rdfs:value  ?suicide .
                ?year dbo:Population ?pop.
                ?pop rdfs:value ?population .
            }
    }

        SERVICE <http://factforge.net/repositories/ff-news>
        {
            ?country_service a dbpont:Country .
            ?country_service onto:countryCode ?iso.
            {
                SELECT ?iso (MAX(?area) AS ?maxArea)
                WHERE
                {
                    ?country_service onto:countryCode ?iso .
                    ?country_service dbpont:areaTotal ?area.
                }
                GROUP BY ?iso
            }
        }

    FILTER(?iso=?country_code)
    FILTER(?suicide>10)
}
ORDER BY DESC (?pop_per_km2)
```

Listing 3: Using SERVICE to get information about the area

```
PREFIX db: <http://mappings.dbpedia.org/server/ontology/classes/>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
PREFIX gn: <http://www.geonames.org/ontology#>
PREFIX dbo: <http://dbpedia.org/page/>

SELECT DISTINCT
?country ?suicideRate
WHERE
{
    ?count a db:Country .
    ?count rdfs:label ?country .
    ?count db:year ?year.

    ?year rdfs:label 2014 .
    ?year dbo:Suicide ?suic.
    ?suic rdfs:value  ?suicide .

    ?year dbo:Population ?pop.
    ?pop rdfs:value ?population .
    BIND((xsd:float(?suicide)*100/(xsd:float(?population))) AS ?suicideRate)
}
ORDER BY desc (?suicideRate)
LIMIT 20
```

Listing 4: Suicide rate in 2014

```
PREFIX db: <http://mappings.dbpedia.org/server/ontology/classes/>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
PREFIX gn: <http://www.geonames.org/ontology#>
PREFIX dbo: <http://dbpedia.org/page/>

SELECT DISTINCT
?country
((?suicide2007-?suicide2015) AS ?suicide_2007_2015_difference)
((?unemployment2007-?unemployment2015) AS ?unemployment_2007_2015_difference)
((?GDPpc2007-?GDPpc2015) AS ?GDPpc_2007_2015_difference)
((?temperature2007-?temperature2015) AS ?temperature_2007_2015_difference)
{
    ?count a db:Country .
    ?count rdfs:label ?country .
    ?count db:year ?year7.
    ?count db:year ?year15.

    ?year7 rdfs:label 2007 .
    ?year15 rdfs:label 2015 .
    ?year7 dbo:Suicide ?suic7.
    ?year15 dbo:Suicide ?suic15.
        ?suic7 rdfs:value ?suicide2007 .
    ?suic15 rdfs:value ?suicide2015 .

    ?year7 dbo:Unemployment ?unem7 .
    ?year15 dbo:Unemployment ?unem15 .
    ?unem7 rdfs:value ?unemployment2007.
    ?unem15 rdfs:value ?unemployment2015.

    ?year7 dbo:Gross_domestic_product ?GDP7 .
    ?year15 dbo:Gross_domestic_product ?GDP15 .
    ?GDP7 rdfs:value ?GDPpc2007 .
    ?GDP15 rdfs:value ?GDPpc2015 .

    ?year7 dbo:Temperature ?temp7 .
    ?year15 dbo:Temperature ?temp15 .
    ?temp7 rdfs:value ?temperature2007 .
    ?temp15 rdfs:value ?temperature2015 .
}
ORDER BY DESC (?suicide_2007_2015_difference)
LIMIT 10
```

Listing 5: Difference between 2007 and 2015

```
PREFIX db: <http://mappings.dbpedia.org/server/ontology/classes/>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
PREFIX ncit: <https://ncim.nci.nih.gov/ncimbrowser/ConceptReport.jsp?dictionary=NCI%20Metathesauru
PREFIX gn: <http://www.geonames.org/ontology#>
PREFIX dbo: <http://dbpedia.org/page/>

SELECT DISTINCT
?continent ?country
?year_int ?suicide
?temperature ?unemployment
?education ?poverty
?GDPpc ?migration ?population

WHERE
{
    ?count a db:Country .
    ?count rdfs:label ?country .
    ?count ncit:C0454690 ?continent .

    ?count db:year ?year.
    ?year rdfs:label ?year_int.

    ?year dbo:Suicide ?suic.
    ?suic rdfs:value  ?suicide .

    ?year dbo:Temperature ?temp .
    ?temp rdfs:value ?temperature .

    ?year dbo:Unemployment ?unem .
    ?unem rdfs:value ?unemployment.

    ?year dbo:Education ?edu.
    ?edu rdfs:value ?education .

    ?year dbo:Poverty ?pov.
    ?pov rdfs:value ?poverty .

    ?year dbo:Gross_domestic_product ?GDP .
    ?GDP rdfs:value ?GDPpc .

    ?year dbo:Migration ?mig.
    ?mig rdfs:value ?migration .

    ?year dbo:Population ?pop.
    ?pop rdfs:value ?population .
}
```

Listing 6: Query of the whole database