

Veterans Affairs - State-wise distribution of Veteran related information for FY07 and FY08

Team Name: Logic, Proof, and Trust

Team Members:

1. Shreevatsa Hosakere Vittal Rao (sxh163630)
2. Avinash Chandrasekharan (axc166930)
3. Piyush Supe (pgs160230)
4. Aayush Bhandari (axb167830)

Type of Project: Custom Project

1. INTRODUCTION

The Geographic Distribution of Expenditures (GDX) by the Department of Veteran Affairs is an annual report that shows estimated expenditures for major geographical areas (state, county, and congressional district). The major issues considered are: Compensation and Pension, Readjustment (Education) and Vocational Rehabilitation, Insurance, Construction and Veteran Population.

The Veterans Health Administration 2008 Hospital Report Card by the Department of Veteran Affairs is an annual report that shows estimated medical expenditures for major geographical areas (state, county, and congressional district). The major issues considered are: Wait times, Medical Center Staffing, Age wise distribution of Veteran Population. Using these datasets, various information regarding the performance of these departments all across United States can be obtained and can be used to improve the conditions to better serve the veterans.

2. TARGET AUDIENCE

Datasets 1152 and 1154 were helpful in obtaining average expenditure and average population of the veterans in 2 years for every state. This helps the Veteran affairs department in predicting the expenditure and population for the upcoming years. Also, this helps us in identifying an increase/decrease in expenditure and population values of the two years.

Datasets 1154 and 1205 were helpful in getting number of medical staff available per veteran in every state. Datasets 1154 and 1204 were helpful in getting number of veterans that are served within 30 days of them reporting. Veterans' health administrative department can analyze and understand the quality of care received by veterans and identify the necessity for medical professionals in each state.

Dataset 1538 helps us in knowing about the number of people between the specified age ranges. This information can be used by the veteran's affairs department to identify in a particular state, the distribution of veterans across age groups so that proper policies can be formed to aid the veterans.

3. DESCRIPTION OF DATA SOURCES

Datasets under consideration are:

- Dataset 1152 – <http://data-gov.tw.rpi.edu/raw/1152/data-1152.rdf> (No. of triples: 44,007)
Dataset 1154 – <http://data-gov.tw.rpi.edu/raw/1154/data-1154.rdf> (No. of triples: 40,654)
Dataset 1204 - <http://data-gov.tw.rpi.edu/raw/1204/data-1204.rdf> (No. of triples: 1118)
Dataset 1205 - <http://data-gov.tw.rpi.edu/raw/1205/data-1205.rdf> (No. of triple: 1257)
Dataset 1538 – <http://data-gov.tw.rpi.edu/raw/1538/data-1538.rdf> (No. of triples: 53,488)

Dataset 1152:

```

<rdf:Description rdf:about="#entry690">
  <medical_care>2384.26</medical_care>
<education_vocational_rehabilitation_employment>59.36</education_vocational_rehabilitation_employment>
  <general_operating_expenses>0</general_operating_expenses>
  <veteran_population>1687.04</veteran_population>
  <compensation_pensions>2008.84</compensation_pensions>
  <rdf:type rdf:resource="http://data-gov.tw.rpi.edu/2009/data-gov-twc.rdf#DataEntry"/>
  <fips>17193</fips>
  <unique_patients>542</unique_patients>
  <insurance_indemnities>59.61</insurance_indemnities>
  <county>WHITE</county>
  <total_expenditure>4512.07</total_expenditure>
  <loan_guaranty>0</loan_guaranty>
  <construction>0</construction>
  <state>ILLINOIS</state>
</rdf:Description>

```

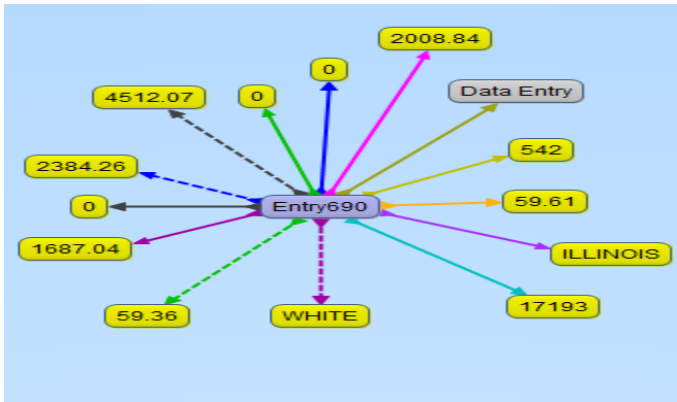


Figure 1: Triples in DATASET 1152

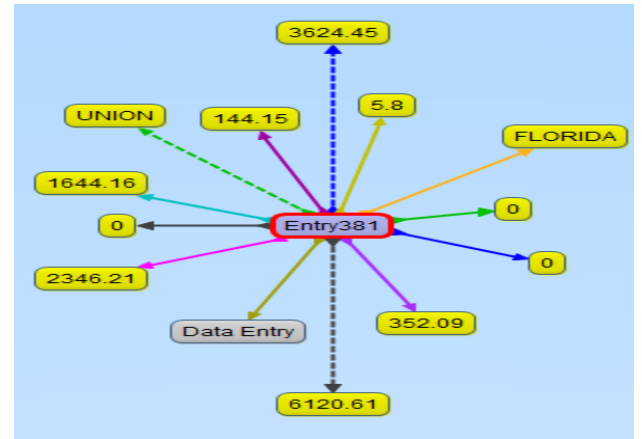


Figure 2: Triples in DATASET 1154

Dataset 1154:

```

<rdf:Description rdf:about="#entry381">
  <insurance_indemnities>5.8</insurance_indemnities>
  <loan_guaranty>0</loan_guaranty>
  <veteran_population>1644.16</veteran_population>
<education_vocational_rehabilitation_employment>144.15</education_vocational_rehabilitation_employment>
  <general_operating_expenses>0</general_operating_expenses>
  <state>FLORIDA</state>
  <unique_patients>352.09</unique_patients>
  <medical_care>3624.45</medical_care>
  <compensation_pensions>2346.21</compensation_pensions>
  <total_expenditure>6120.61</total_expenditure>
  <rdf:type rdf:resource="http://data-gov.tw.rpi.edu/2009/data-gov-twc.rdf#DataEntry"/>
  <county>UNION</county>
  <construction>0</construction>
</rdf:Description>

```

Dataset 1204:

```

<rdf:Description rdf:about="#entry123">
  <specialty_care_seen_in_30_days>98</specialty_care_seen_in_30_days>
  <primary_care_seen_in_30_days>91</primary_care_seen_in_30_days>

```

```

<state>CALIFORNIA</state>
<city>Palo Alto</city>
<facility_name>VIRGINIA Palo Alto HCSPaloAlto MenloPk Livermore</facility_name>
<visn_name>Sierra Pacific Network</visn_name>
<visn>21</visn>
<rdf:type rdf:resource="http://data-gov.tw.rpi.edu/2009/data-gov-twc.rdf#DataEntry"/>
</rdf:Description>

```

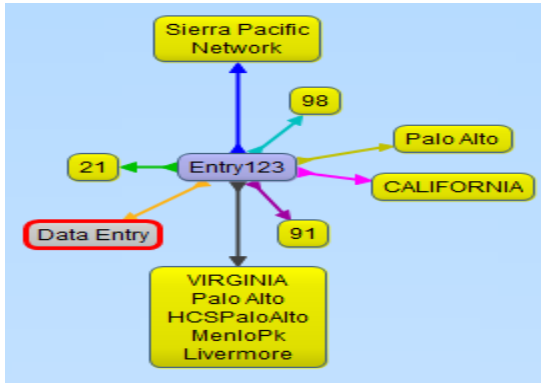


Figure 3: Triples in DATASET 1204

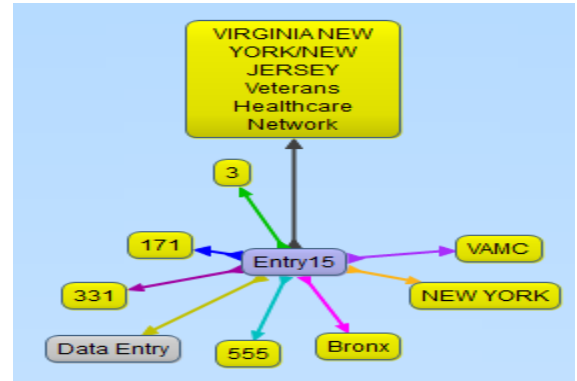


Figure 4: Triples in DATASET 1205

Dataset 1205:

```

<rdf:Description rdf:about="#entry15">
  <staffing_other_health_professionals>555</staffing_other_health_professionals>
  <staffing_nursing>331</staffing_nursing>
  <staffing_physicians>171</staffing_physicians>
  <state>NEW YORK</state>
  <city>Bronx</city>
  <facility_name>VAMC</facility_name>
  <visn_name>VIRGINIA NEW YORK/NEW JERSEY Veterans Healthcare Network</visn_name>
  <visn>3</visn>
  <rdf:type rdf:resource="http://data-gov.tw.rpi.edu/2009/data-gov-twc.rdf#DataEntry"/>
</rdf:Description>

```

Dataset 1538:

```

<rdf:Description rdf:about="#entry2816">
  <total_pension>44</total_pension>
  <age_75_c_p>136</age_75_c_p>
  <rdf:type rdf:resource="http://data-gov.tw.rpi.edu/2009/data-gov-twc.rdf#DataEntry"/>
  <age_35_44_c_p>56</age_35_44_c_p>
  <age_65_74_c_p>78</age_65_74_c_p>
  <compensation_30_50>152</compensation_30_50>
  <state>VERMONT</state>
  <county>WASHINGTON</county>
  <missing_age_c_p>0</missing_age_c_p>
  <total_compensation>582</total_compensation>
  <compensation_60_90>118</compensation_60_90>
  <compensation_100>60</compensation_100>
  <total_c_p>626</total_c_p>
  <age_55_64_c_p>211</age_55_64_c_p>
  <age_45_54_c_p>104</age_45_54_c_p>
  <compensation_30>251</compensation_30>

```

```
<age_35_c_p>41</age_35_c_p>
</rdf:Description>
```

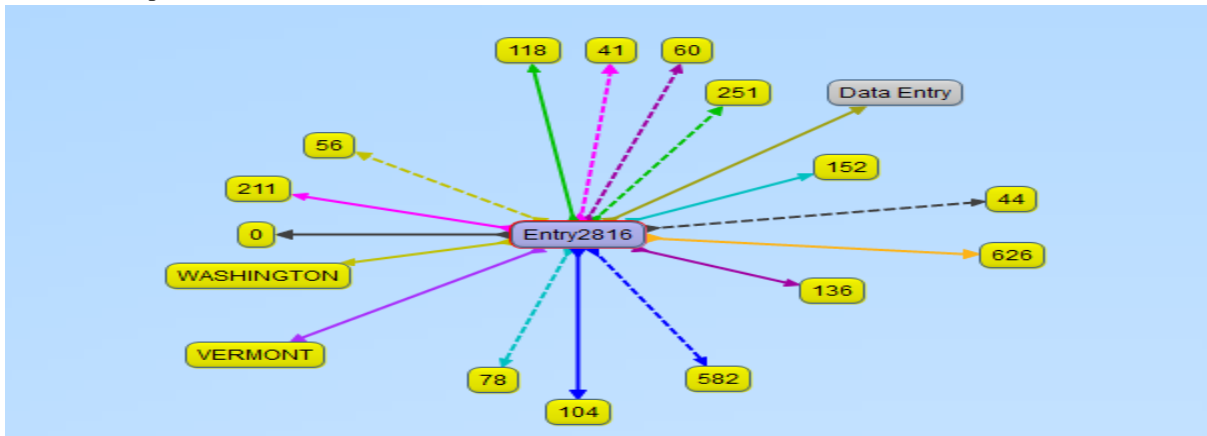


Figure 5: Triples in DATASET 1538

4. DATA INTEGRATION

Different datasets have useful attributes which would help us in our analysis and prediction. These datasets can be integrated to get more useful information. To perform this analysis, the above datasets from LOGD site are uploaded to Fuseki server. Apache JENA Fuseki: SPARQL server. Provides REST-style SPARQL HTTP Update, SPARQL Query, and SPARQL Update using the SPARQL protocol over HTTP.

4.1 COMBINING DATA:

- Datasets 1152 and 1154: When integrated we get the overall information on the total expenditure and total population for two years (2007, 2008) for every state and county.
- Datasets 1154 and 1205: When integrated we get the overall information on veteran population, medical staff available for every state and county during the year 2008.
- Datasets 1154 and 1204: When integrated we get the overall information on veteran population, medical assistance available and number of veterans taken care of within the first 30 days for every state and county during the year 2008.

Dataset 1152,1154:

Attribute	Description
total_expenditure	total expenditure of the veterans in every state
veteran_population	total population of the veterans in every state
state	state's name

Dataset 1205

Attribute	Description
staffing_nursing	total number of nursing staff
staffing_physicians	total number of physicians
state	state's name

Dataset 1204

Attribute	Description
specialty_care_seen_in_30_days	total number of veterans in special care attended in 30 days

primary_care_seen_in_30_days	total number of veterans in primary care attended in 30 days
state	state's name

Dataset 1538:

Attribute	Description
age_35_44_c_p	Number of people receiving pension between ages of 35 and 44
age_45_54_c_p	Number of people receiving pension between ages of 45 and 54
age_65_74_c_p	Number of people receiving pension between ages of 65 and 74
state	state's name

5 DATA PRODUCT RESULTS:

SPARQL Queries:

The data is being integrated using the “**state**” property in every dataset. The SPARQL queries shown below are executed in FUSEKI server’s sparql endpoint and the corresponding results are retrieved:

Query 1:

```
SELECT ?state ?veteran_population_1 ?veteran_population_2
WHERE {
  {
    SELECT ?state (sum(xsd:decimal(?vp1152)) AS ?veteran_population_1)
      WHERE {
        ?s <http://data-gov.tw.rpi.edu/vocab/p/1152/veteran_population> ?vp1152.
        ?s <http://data-gov.tw.rpi.edu/vocab/p/1152/state> ?state.
      } group by ?state order by ?state
    }
  {
    SELECT ?state1 (sum(xsd:decimal(?vp1154)) AS ?veteran_population_2)
      WHERE {
        ?s <http://data-gov.tw.rpi.edu/vocab/p/1154/veteran_population> ?vp1154.
        ?s <http://data-gov.tw.rpi.edu/vocab/p/1154/state> ?state1.
      } group by ?state1 order by ?state1
    }
  FILTER (?state = ?state1)
}
```

Result:

This output shows the veteran expenditure and population in each state for two years.

“exp_year1, exp_year2” – expenditures in two years

“population_year1, population_year2” – population in two years.

```
{
  "head": {
    "vars": [ "state" , "veteran_population_1" , "veteran_population_2" ]
  },
  "results": {
    "bindings": [
      {
        "state": { "type": "literal" , "value": "ALABAMA" } ,
        "veteran_population_1": { "type": "literal" , "datatype": "http://www.w3.org/2001/XMLSchema#decimal" , "value": "416947.01" } ,
        "veteran_population_2": { "type": "literal" , "datatype": "http://www.w3.org/2001/XMLSchema#decimal" , "value": "413578.90" }
      }
    ]
  }
}
```

Figure 6: JSON output showing population of veterans for each state during the given years

Query 2:

PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>

```

PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
SELECT ?state ?exp_year1 ?exp_year2
WHERE {
  {
    SELECT ?state (sum(xsd:decimal(?exp_yr1)) AS ?exp_year1)
      WHERE {
        ?s <http://data-gov.tw.rpi.edu/vocab/p/1152/total_expenditure> ?exp_yr1.
        ?s <http://data-gov.tw.rpi.edu/vocab/p/1152/state> ?state.
      } group by ?state order by ?state
    }
  {
    SELECT ?state1 (sum(xsd:decimal(?exp_yr2)) AS ?exp_year2)
      WHERE {
        ?s <http://data-gov.tw.rpi.edu/vocab/p/1154/total_expenditure> ?exp_yr2.
        ?s <http://data-gov.tw.rpi.edu/vocab/p/1154/state> ?state1.
      } group by ?state1 order by ?state1
    }
  }
  FILTER (?state = ?state1)
}

```

Result:

This query output shows the number of people of different ages receiving compensation and pension in every state.

```

{
  "head": {
    "vars": [ "state" , "exp_year1" , "exp_year2" ]
  } ,
  "results": {
    "bindings": [
      {
        "state": { "type": "literal" , "value": "ALABAMA" } ,
        "exp_year1": { "type": "literal" , "datatype": "http://www.w3.org/2001/XMLSchema#decimal" , "value": "1534602.94" } ,
        "exp_year2": { "type": "literal" , "datatype": "http://www.w3.org/2001/XMLSchema#decimal" , "value": "1714683.44" }
      } ,

```

Figure 7: JSON output showing expenditure for each state during the given years

Query 3:

```

PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
SELECT (SUM(xsd:integer(?age35to44)) AS ?betweenageof35and44) (SUM(xsd:integer(?age45to54)) AS
  ?agebetween45and54) (SUM(xsd:integer(?age55to64)) AS ?agebetween55and64)
(SUM(xsd:integer(?age65to74)) AS ?agebetween65and74) ?state
WHERE {
  graph ?ff{
    ?s <http://data-gov.tw.rpi.edu/vocab/p/1538/state> ?state.
    ?s <http://data-gov.tw.rpi.edu/vocab/p/1538/age_35_44_c_p> ?age35to44.
    ?s <http://data-gov.tw.rpi.edu/vocab/p/1538/age_45_54_c_p> ?age45to54.
    ?s <http://data-gov.tw.rpi.edu/vocab/p/1538/age_65_74_c_p> ?age65to74.
    ?s <http://data-gov.tw.rpi.edu/vocab/p/1538/age_55_64_c_p> ?age55to64.
  }
  } group by ?state order by ?state

```

Result:

This query output shows the number veterans in each age category in every state.

```
{
  "head": {
    "vars": [ "betweenageof35and44" , "agebetween45and54" , "agebetween55and64" , "agebetween65and74" , "state" ]
  },
  "results": {
    "bindings": [
      {
        "betweenageof35and44": { "type": "literal" , "datatype": "http://www.w3.org/2001/XMLSchema#integer" , "value": "8049" } ,
        "agebetween45and54": { "type": "literal" , "datatype": "http://www.w3.org/2001/XMLSchema#integer" , "value": "14345" } ,
        "agebetween55and64": { "type": "literal" , "datatype": "http://www.w3.org/2001/XMLSchema#integer" , "value": "23685" } ,
        "agebetween65and74": { "type": "literal" , "datatype": "http://www.w3.org/2001/XMLSchema#integer" , "value": "10026" } ,
        "state": { "type": "literal" , "value": "ALABAMA" }
      }
    ]
  }
}
```

Figure 8: JSON output showing number of people across different age gaps getting medical benefits

Query 4:

PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>

PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>

PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>

prefix fiftyfour: <http://data-gov.tw.rpi.edu/vocab/p/1154/>

prefix onetwofive: <http://data-gov.tw.rpi.edu/vocab/p/1205/>

SELECT distinct ?state (((xsd:decimal(?veteran_population1))/(xsd:decimal(?staffing_nursing1)+
xsd:decimal(?staffing_physicians1)))) AS ?div)

WHERE {

{

SELECT ?state (SUM(xsd:decimal(?veteran_population)) as ?veteran_population1)

WHERE {

graph ?1154 {

?x fiftyfour:state ?state .

?x fiftyfour:veteran_population ?veteran_population .

}

}

group by ?state

}

{

SELECT ?state1 (SUM(xsd:decimal(?staffing_nursing)) as ?staffing_nursing1)

(SUM(xsd:decimal(?staffing_physicians)) as ?staffing_physicians1)

WHERE {

graph ?1205new {

?y onetwofive:state ?state1 .

?y onetwofive:staffing_nursing ?staffing_nursing .

?y onetwofive:staffing_physicians ?staffing_physicians.

}

}

group by ?state1

}

FILTER (?state = ?state1)

}

ORDER BY ?state

Result:

This query output shows the number of patients and their total expenditure for medical care in each state. Also specifies the medical facility being provided in each state.

```
[
  "head": {
    "vars": [ "state" , "div" ]
  },
  "results": {
    "bindings": [
      {
        "state": { "type": "literal" , "value": "ALABAMA" } ,
        "div": { "type": "literal" , "datatype": "http://www.w3.org/2001/XMLSchema#decimal" , "value": "338.444271685761047463175122" }
      }
    ]
  }
]
```

Figure 9:JSON output showing number of veterans per hospital staff (doctors and nurses) in every state

Query 5:

PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>

SELECT ?state ?veteran_population_2 (xsd:decimal(?specialty_care_seen)+xsd:decimal(?primary_care_seen) AS ?Total_cases)

WHERE {

{

SELECT ?state (sum(xsd:decimal(?vp1154)) AS ?veteran_population_2)

WHERE {

?s <http://data-gov.tw.rpi.edu/vocab/p/1154/veteran_population> ?vp1154.

?s <http://data-gov.tw.rpi.edu/vocab/p/1154/state> ?state.

} group by ?state order by ?state

}

{

SELECT ?state1 (sum(xsd:decimal(?scs30)) AS ?specialty_care_seen) (sum(xsd:decimal(?pcs30)) AS

?primary_care_seen)

WHERE {

?s1 <http://data-gov.tw.rpi.edu/vocab/p/1204/specialty_care_seen_in_30_days> ?scs30.

?s1 <http://data-gov.tw.rpi.edu/vocab/p/1204/primary_care_seen_in_30_days> ?pcs30.

?s1 <http://data-gov.tw.rpi.edu/vocab/p/1204/state> ?state1.

} group by ?state1 order by ?state1

}

FILTER (?state = ?state1)

}

}

Result:

This query output shows the number of patients and their total expenditure for medical care in each state. Also specifies the medical facility being provided in each state.

```
{
  "head": {
    "vars": [ "state" , "veteran_population_2" , "specialty_care_seen" , "primary_care_seen" ]
  },
  "results": {
    "bindings": [
      {
        "state": { "type": "literal" , "value": "ALABAMA" } ,
        "veteran_population_2": { "type": "literal" , "datatype": "http://www.w3.org/2001/XMLSchema#decimal" , "value": "413578.90" } ,
        "specialty_care_seen": { "type": "literal" , "datatype": "http://www.w3.org/2001/XMLSchema#decimal" , "value": "294" } ,
        "primary_care_seen": { "type": "literal" , "datatype": "http://www.w3.org/2001/XMLSchema#decimal" , "value": "298" }
      }
    ]
  }
}
```

Figure 10:JSON output showing average wait time for veterans in every state

6. BUILDING VISUALIZATION

Based on the results obtained above sparql queries, a web page has been designed using html, jsp, and bootstrap. The json output is used as input to query url in building the visualization charts.

Query 1:

Veteran poulation difference for FY2007 and FY2008

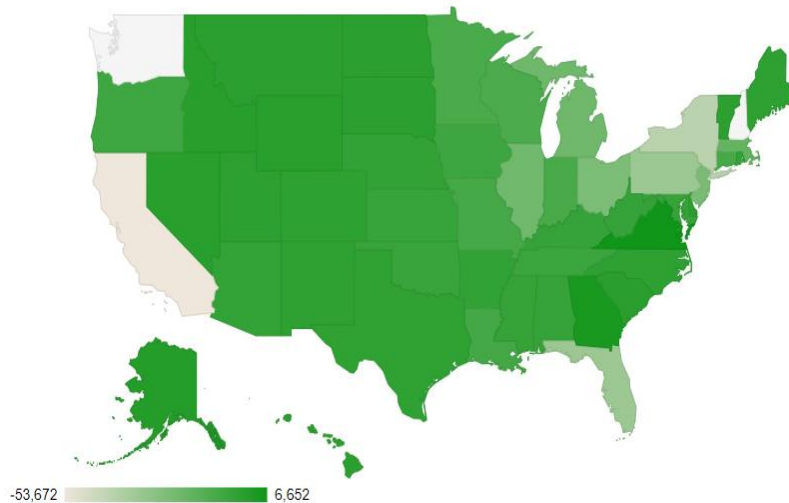


Figure 11: Difference in veteran population from 2007 to 2008

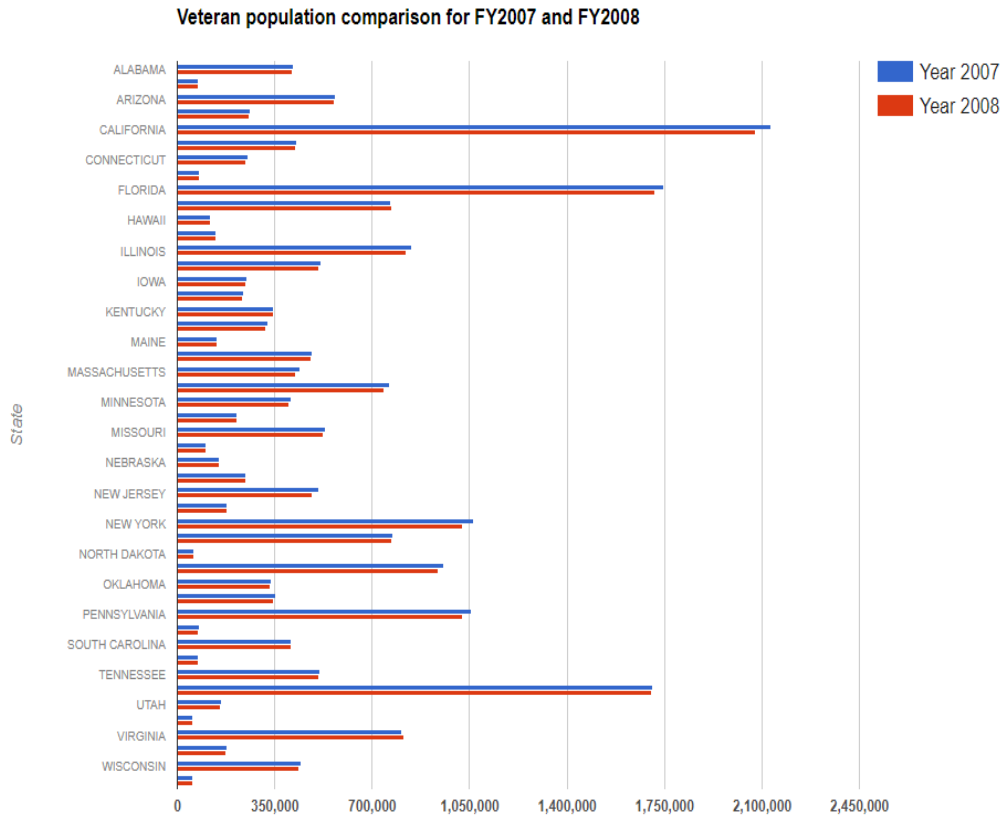


Figure 12: Same results in bar chart

Query 2:

Veteran expenditure difference for FY2007 and FY2008

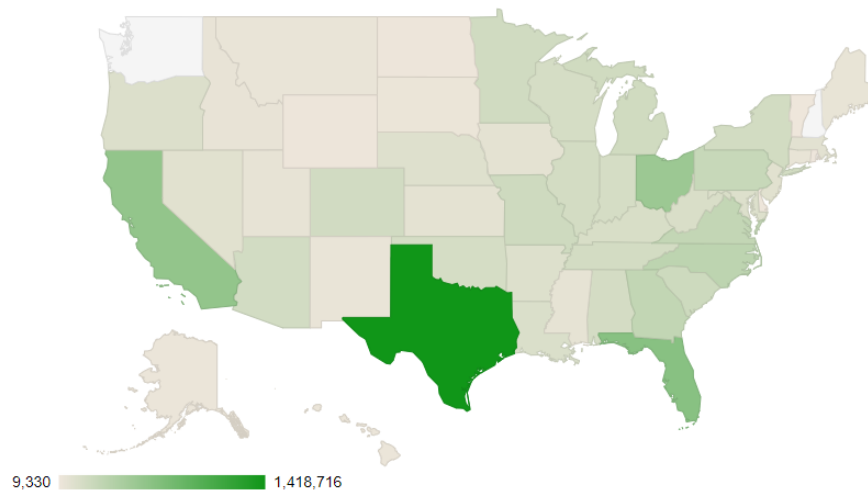


Figure 13: Difference in expenditure from 2007 to 2008

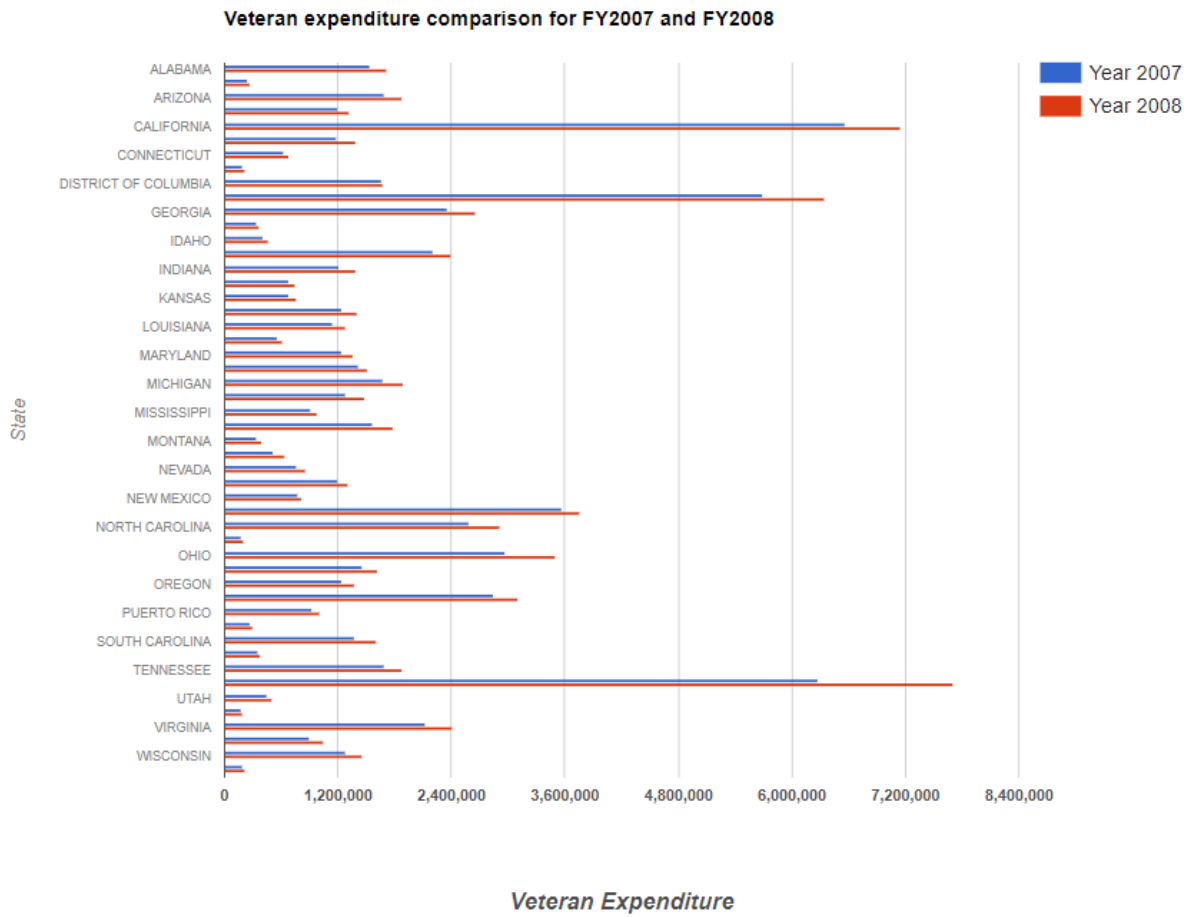
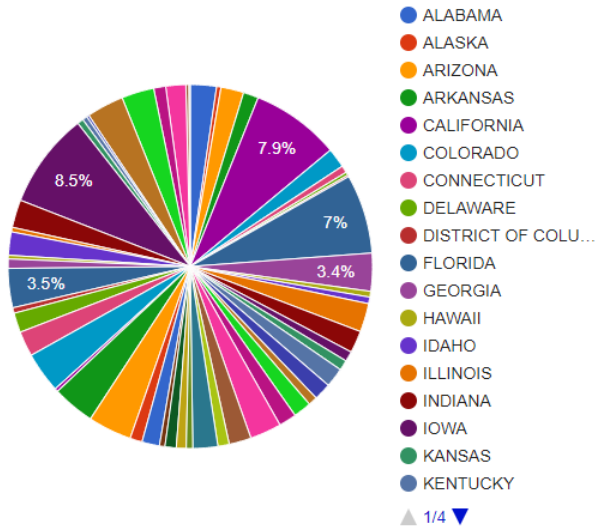


Figure 14: Same results in bar chart

Query 3: State wise population showing the veteran age distribution each state.

Between age 55 and 64



Between age 65 and 74

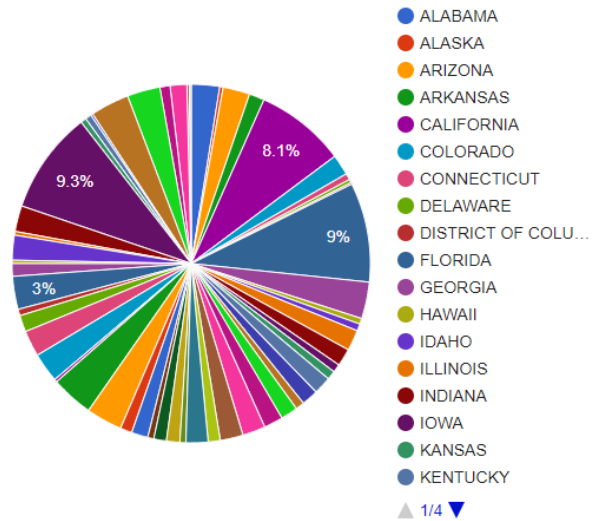
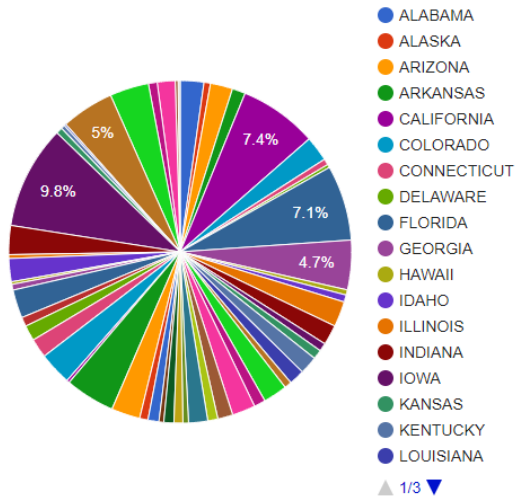


Figure 15: Veterans in the age between 35 and 44

Figure 16: Veterans in the age between 45 and 54

Between age 35 and 44



Between age 45 and 54

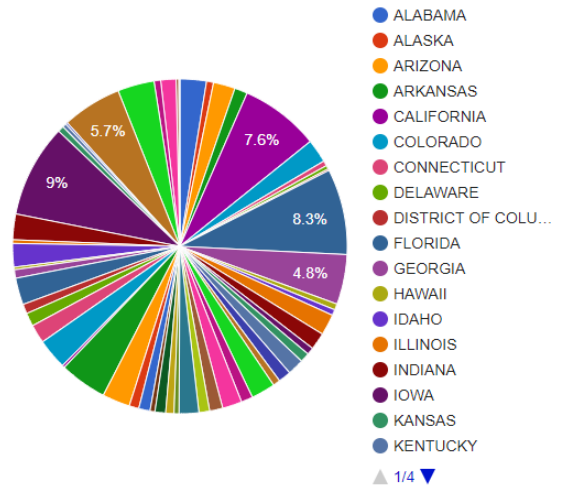
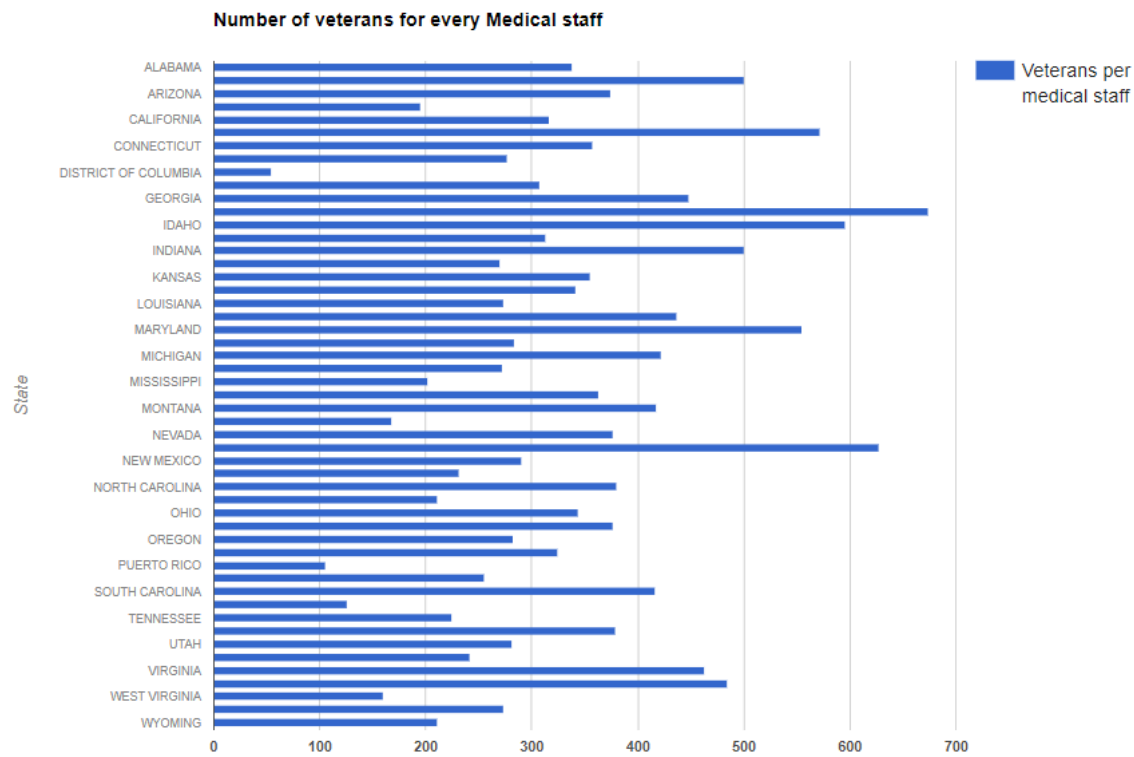


Figure 17: Veterans in the age between 55 and 64

Figure 18: Veterans in the age between 65 and 74

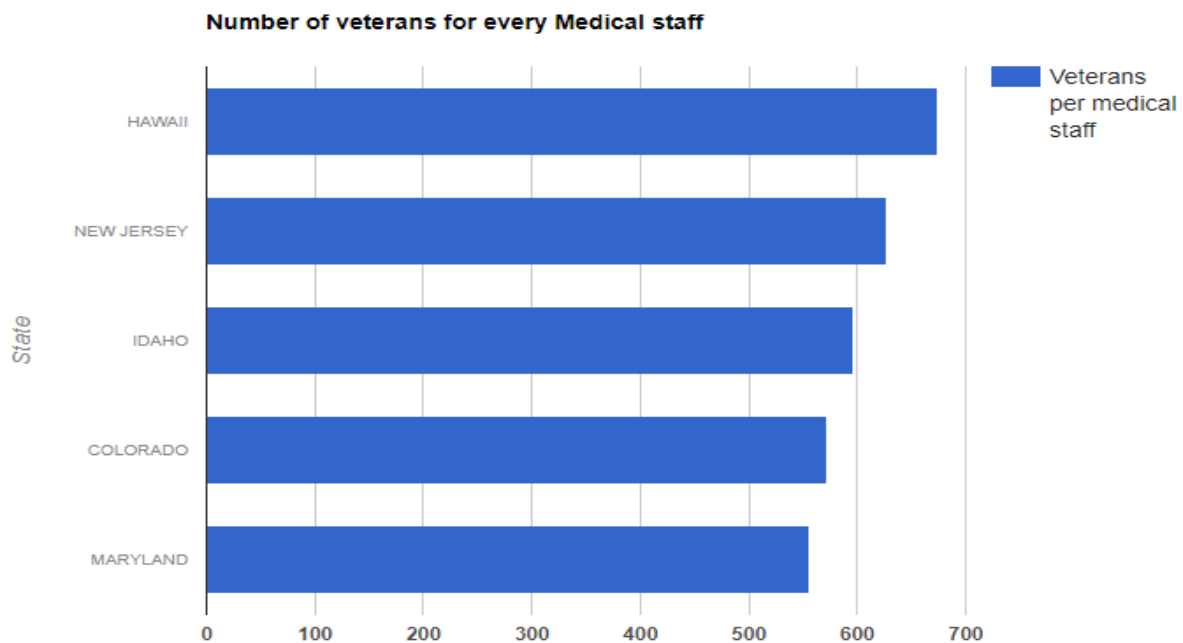
Query 4:



Number of Veterans

Figure 19: Number of veterans per medical staff in each state

Bottom 5 states with maximum number of veterans per medical staff,



Number of Veterans

Figure 20: Bottom 5 states with max number of veterans per medical staff

Next is, graph showing states which has attended most number of veterans in 30 day duration

Total patients attended in 30 days compared to veteran population(in thousands)

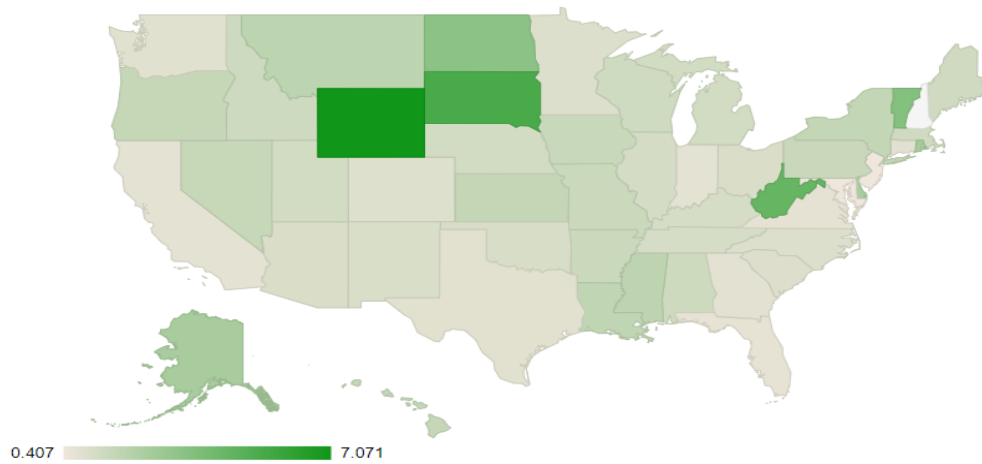


Figure 21: Number of patients attended in 30 days

Veteran population comparison for FY2007 and FY2008

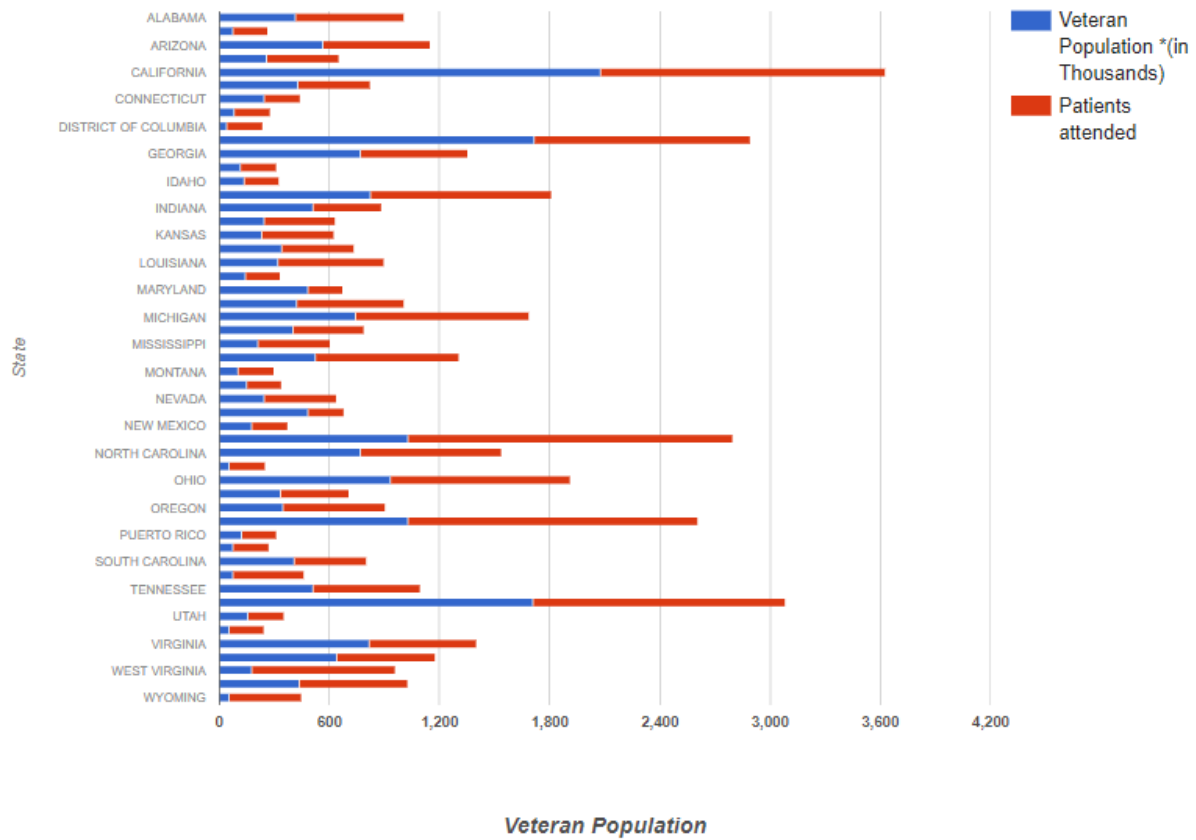


Figure 22: Number of patients attended in 30 days in comparison to veteran population

7. CUSTOM PROJECT JUSTIFICATION

We justify this as a custom project by looking at the following:

- a) Framework used - SPARQL endpoint of Apache Jena FUSEKI server - uses local host to access the server in web page whereas a logd project uses the LOGD SPARQL endpoint.
- b) In our project, datasets are taken from the logd site and loaded to the server to use the triples in them, where as in logd project the datasets are already present on the logd site.
- c) Localhost JSON output query was used in our visualisation whereas in logd project the query is present on the logd site.

8. SUMMARY

First step in the process is to identify the datasets that would help us in this project (Source of datasets: logd site). After some research we have concluded that datasets numbered 1152, 1154, 1538, 1204 and 1205 would help us in this project.

Once the datasets are chosen, pre-processing was done to change the case of States in different files to one common format, the underlying RDF and ontology information of the datasets is understood. We have also used Gruff to get a visual understanding of the semantics of the datasets. Then we have analyzed if there is a chance to integrate two datasets based on certain attributes to get more useful information.

Fuseki, an Apache Jena server is used to load the RDF datasets and SPARQL queries are written to query the interested RDF data and get desired results. SPARQL supports various output formats such as text, JSON. Also, Google visualization API's are used to improve the visualization of the results.

A webpage has been created using JavaScript (Server side scripting language), Bootstrap (front end design framework) and Google visualization packages for visualization of the results.