**Project Final Report**

**Housing Inspection Scores in relation to State Income**

**Group:** Internet Explorers

**Members:** Ian Braunfeld (ixb120030) & Paul Geevargheese Mathew (pxm142330)

**Project Type:** Custom Project

**1.** **Introduction:**

Our project takes a look at the possible correlation between State income and public housing inspection scores for housing units in those states. The point of this is to show whether or not a state’s overall income also affects its public housing quality of living and to show what kind of impact it has, if any.

**2. Target Audience:** For this project we identified 4 audiences that would benefit from this project.

1. The first being an audience looking for a certain quality of housing to move into based on inspection scores
2. The second being an audience looking for a location with a high Adjusted Gross Income to provide a higher paying job.
3. The third would be a combination of the last two, similar to a college student looking for a good income place to work with a good quality of housing as well.
4. It is also useful for research and showing how Adjusted Gross Income can correlate with housing inspection scores in each state.

This project is meant for people who tend to travel and don’t tend to settle down too long in a single place. It allows users to skip some of the research involved in moving to get an idea of the quality of living in certain states.

**3. Description of Data Sources:**

**3.1. Dataset 1356:**

Contains selected individual income tax return data items classified by state and county, each data point consists of a variety of attributes listed below.

* AGI: Adjusted Gross Income - an individual's total gross income (all income from whatever source) minus specific deductions such as dividends, taxable income, and taxable interest.
* County\_code: Number that represents a specific county (if “000” represents entire state)
* County\_name: Name of the county
* Dividends: amount generated in dividends
* Exmpt\_num: Number of exemptions
* Interest: Taxable interest
* Return\_num: Number of returns
* State\_abbrv: 2 letter abbreviation for state
* State\_code: numeric representation of state
* Wages\_salaries: Taxable income
* Year: The year the data was collected for

Note that some sources of income such as interest and dividends are listed, this is because Adjusted Gross Income does not include these sources of income like standard “gross income” would.

For this dataset we most notably used the AGI, Wage, Interest, State abbreviation, County and year in our queries to match the information to the map displayed.

**3.2. Dataset 1258:**

Contains the Department of Housing and Urban Development’s (HUD) physical inspections of Public Housing properties that are owned, subsidized, or insured by HUD. Each row of data contains a variety of attributes listed below.

* property\_id - Property identification
* property\_name - Name of the property
* address - Property address
* city - Property City
* cbsa\_name - Core Based Statistical Area name
* cbsa\_code - Core Based Statistical Area code #
* county\_name - County name
* county\_code - County code
* state\_name - State name
* state\_code - State code
* zip - zip code
* latitude - latitude #
* longitude - longitude #
* pha\_code - public housing administration code
* pha\_name - public housing administration
* inspection\_score - score given to the property
* inspection\_date - date score was given

For this dataset we most notably used Inspection score, County, State, and inspection date in our queries to match the information with the Google API Map as well as our graphs combining the data and displaying different years of data.

An important thing to note with this dataset is that inspections scores were taken between 2001-2009 and include numerous inspections for the same property across multiple years. It was difficult to be able to use the correct data to correlate with the other dataset since dataset 1356 was taken largely from 2007. Another thing to note because of this issue is that some years of data will have no data taken from some states, as seen in our graphs some years will display "null" information for their average inspection scores.

**4. Data Integration:**

We used a Jena Fuseki server to set up a local SPARQL endpoint. For each dataset we used we assigned a graph name and used it to query. We used XMLHTTP for connecting to the Fuseki server, this allowed us to send queries to the SPARQL endpoint. The dataset files were initially in CSV form and had to be converted into an array via the “jquery.csv” library for easier access to be used by the Google Visualization API.

For displaying the graphs we used the GeoChart, AreaChart, and BubbleChart components of Google Visualization API for displaying the data collected. Jquery packages were also used because Google API require it.

* For getting the Yearly state wise inspection score , we had to take the average of all the inspection score based on the county based on the state based on the year. For this Dataset 1258 was used. After getting the Inspection score based on state, the state id was used to find the county having highest inspection Score.

PREFIX dgp: <<http://data-gov.tw.rpi.edu/vocab/p/1258/>>

PREFIX xsd: <[http://www.w3.org/2001/XMLSchema#](http://www.w3.org/2001/XMLSchema)>

SELECT ?s (AVG(xsd:float(?score)) AS ?AverageInspectionScore) "

WHERE{ " +

graph <<http://data-gov.tw.rpi.edu/vocab/Dataset_1258>> " +

{

?o dgp:inspection\_score ?score .

?o dgp:state\_name ?s.

?o dgp:inspection\_date ?date.

}FILTER((CONTAINS(STR(?date),'"+year+"')))"

} GROUP BY ?s

* For getting State wise AGI , we used Dataset 1356 and used the county code “000” to get the state wise AGI and we used state id to display the county having highest AGI.

PREFIX dgp1356: <<http://data-gov.tw.rpi.edu/vocab/p/1356/>>

              PREFIX skos:  <[http://www.w3.org/2004/02/skos/core#](http://www.w3.org/2004/02/skos/core)>

SELECT ?state (?state\_agi07 AS ?AGI)

WHERE {

    graph <<http://data-gov.tw.rpi.edu/vocab/Dataset_1356>> "+

    {

    ?s dgp1356:county\_code \"000\" . "+

    ?s dgp1356:year ?year ."+

    ?s dgp1356:county\_name ?state\_name1356 . "+

    ?s dgp1356:agi ?state\_agi07 . "+

    ?s dgp1356:state\_code  ?state\_fipscode . "+

    ?s dgp1356:state\_abbrv ?state. "+

FILTER(?year = '"+year+"') } "+

    }

* For getting performance on each state we used the dataset 1356 to display state wise AGI,Wage and Interest.Here we used county code 000 to get the statewise details .

 PREFIX dgp: <<http://data-gov.tw.rpi.edu/vocab/p/1258/>>

  PREFIX dgp1356: <<http://data-gov.tw.rpi.edu/vocab/p/1356/>>

  PREFIX xsd: <[http://www.w3.org/2001/XMLSchema#](http://www.w3.org/2001/XMLSchema)>

  SELECT ?Year (xsd:float(?state\_agi) AS ?AGI)   (xsd:float(?wage) AS ?WAGE) (xsd:float(?interest) AS ?Interest)

  WHERE{

  graph <<http://data-gov.tw.rpi.edu/vocab/Dataset_1356>>

    {

    ?s dgp1356:county\_code \"000\" .

    ?s dgp1356:year ?Year .

      ?s dgp1356:county\_name ?state\_name1356 .

      ?s dgp1356:agi ?state\_agi.

    ?s dgp1356:state\_code  ?state\_fipscode .

      ?s dgp1356:state\_abbrv ?state .

    ?s dgp1356:wages\_salaries ?wage.

      ?s dgp1356:interest  ?interest.

    ?s  dgp1356:dividends ?dividends.

    FILTER(?state = '"+state+"')

  }

  } ORDER BY DESC(?Year)

* For forming correlation between the statewise inspection score ,AGI,Wage ,interest and dividend ,we combined the datasets 1258 and 1356 based on the year( 2007 as we are depicting the latest correlation ) and state

PREFIX dgp: <<http://data-gov.tw.rpi.edu/vocab/p/1258/>>

PREFIX dgp1356: <<http://data-gov.tw.rpi.edu/vocab/p/1356/>>

PREFIX xsd: <[http://www.w3.org/2001/XMLSchema#](http://www.w3.org/2001/XMLSchema)>

SELECT ?st  (AVG(xsd:float(?score)) AS ?AverageInspectionScore) (?state\_agi AS ?GrossIncome) ?wage ?interest ?dividends

WHERE{

graph <<http://data-gov.tw.rpi.edu/vocab/Dataset_1258>>

{

?o dgp:inspection\_score ?score .

?o dgp:state\_name ?st.

?o dgp:inspection\_date ?date.

 }FILTER((CONTAINS(STR(?date),\"2007\")))

 graph <<http://data-gov.tw.rpi.edu/vocab/Dataset_1356>>

    {

   ?s dgp1356:county\_code \"000\" .

    ?s dgp1356:year \"2007\" .

    ?s dgp1356:county\_name ?state\_name1356 .

    ?s dgp1356:agi ?state\_agi.

    ?s dgp1356:state\_code  ?state\_fipscode .

    ?s dgp1356:state\_abbrv ?state.

    ?s dgp1356:state\_abbrv ?st.

    ?s dgp1356:wages\_salaries ?wage.

    ?s dgp1356:interest  ?interest.

    ?s  dgp1356:dividends ?dividends

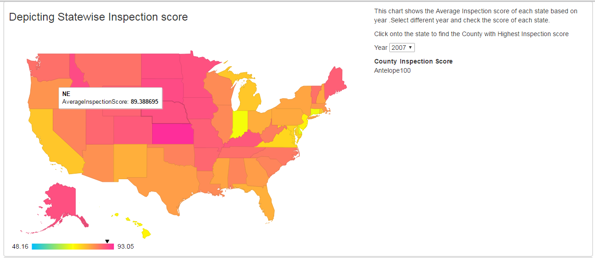
 }

} GROUP BY ?st ?state\_agi ?wage ?interest ?dividends

**5. Data Product Results:**

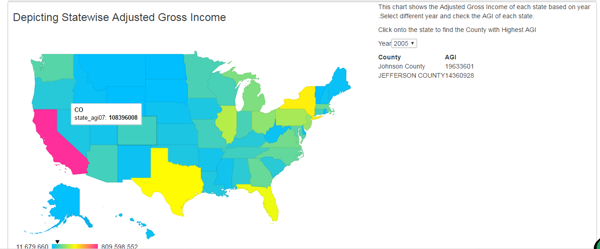
The result product we created with the datasets are 4 different graphs. The first two graphs display maps of the United States each showcasing one of the datasets. The third graph goes more in depth on the Gross Income dataset for a selected state, and the fourth graph combines the two datasets to show the correlation between Adjusted Gross Income and average inspection score for the year 2007. More details are explained below:

**5.1. Graph Depicting State-wise Inspection Score:**



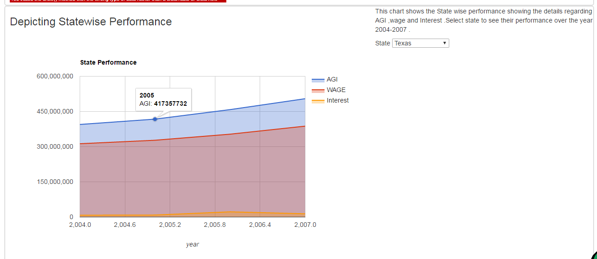
* This graph depicts a heat map of the United States for a selected year where the darker pink/orange colors depict a high average state housing inspection score and the lighter blue/yellow colors depicts a low average state housing inspection score.
* If you move your cursor over a state on the map you will be able to see the exact average inspection score for that state.
* Upon clicking on a state the highest scoring county will be displayed on the right side of the map along with the score.
* You can change the year of inspections scores via the dropdown box on the right side of the map.

**5.2. Graph Depicting State-wise Adjusted Gross Income:**



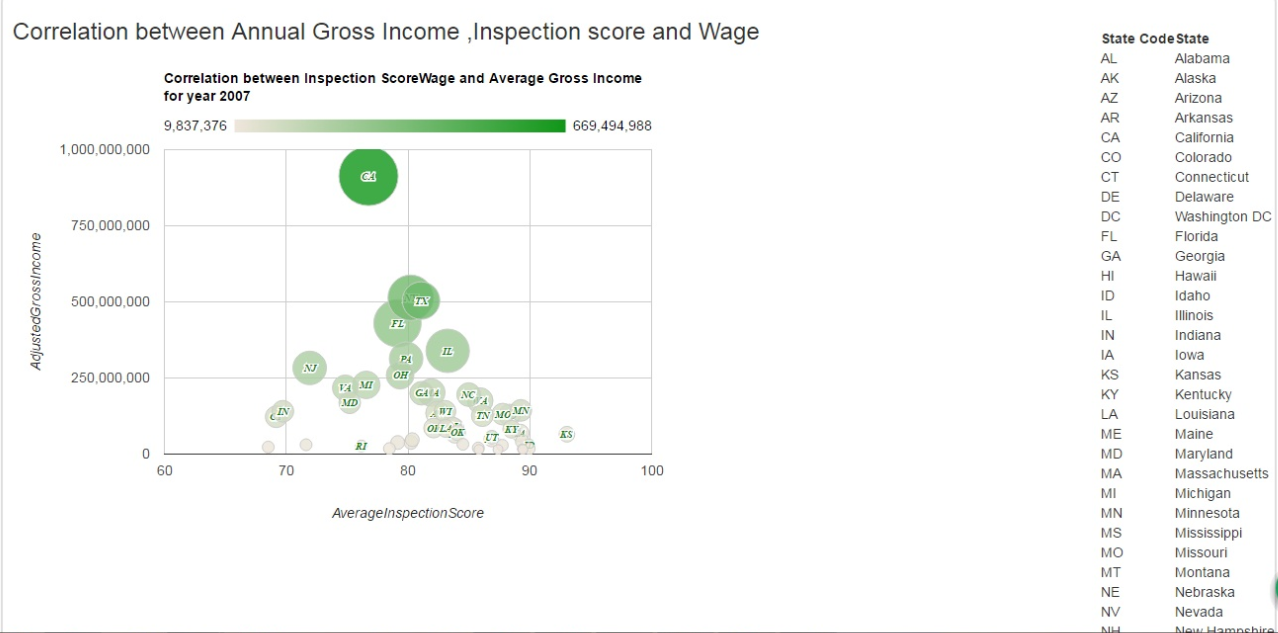
* This graph depicts another heat map of the United States for a selected year where the darker orange colors depict a high Adjusted Gross Income(relative to the highest and lowest) and the lighter blue/yellow colors depicts a low Adjusted Gross Income.
* If you move your cursor over a state on the map you will be able to see the exact Adjusted Gross Income for that state.
* Upon clicking on a state the highest AGI county will be displayed on the right side of the map along with the AGI value.
* You can change the year of Adjusted Gross Income via the dropdown box on the right side of the map.

**5.3. Graph Depicting State-wise Performance:**



* This graph depicts 3 specific values related to the state income over the course of time data was collected for that state.
* The X axis represents the year, the Y axis represents the amount of money gained
* The 3 values depicted are Adjusted Gross Income, Wage (taxable income, and Interest (taxable interest
* You can change the state for which the graph is displaying information for by using the drop down menu on the right side of the graph.
* By moving your cursor over the lines on the graph you can see the exact data points for certain years.
* This graph allows users to see a basic trend the state has been following over the course of a few years

**5.4. Graph Depicting Correlation between Adjusted Gross income, Inspection Score, and Wage:**



* The final graph we created depicts the correlation between the two datasets, the X axis is the average inspection score while the Y axis is the Adjusted Gross Income.
* Each data point in the graph represents a different state in the United States with an additional depiction of information via the color and size of these data points. The darker green and larger a data point is, the higher the Wage for that state is.
* Upon scrolling over a data point you will see the State abbreviation followed by that states Average Inspection Score, AGI, Wage, and Interest values.
* Along the right side of the graph a legend is shown in case a user is unfamiliar with State Abbreviations.

**6. Summary:**

Our Project took information from two different datasets about state income and housing inspection scores and threw them together to show users 4 different graphs of information. The use of this information can range from seeing where housing needs improvements in the U.S. to deciding where the best place to move is. We were able to show users what states have high AGI’s while also having high inspection scores, or even let them choose what on the spectrum they are looking for.

The conclusions we came across with our graphs was that the gross income doesn’t necessarily dictate the housing inspection scores. While there is a slight correlation showing lower income states having high inspection scores there also exist low income states with low inspection scores. Overall we were able to fulfill the goal we set out to make and the use of the data gathered and shown is up to the user and what they are looking to get out of it.

**6.1. Changes Made:**

No changes made to dataset or expected results.

**6.2. Problems:**

As stated previously the only major problem we came across in this project was the inspection score dataset had multiple data points taken for a specific property/county over numerous years which made it hard to select the exact housing score we wanted for our graphs while not letting previous years of the same property/county interfere with our data findings.

Another flaw with this dataset was there was not a fixed number of inspections per county in a given year and this could cause the state average to change year to year depending on what counties got inspected and how many times.

One other slight performance problem we had was the load time for the 4th graph was significantly higher than the rest and sometimes would not appear initially if scrolled to first on our web page.

**6.3. Lessons Learned:**

We thought initially the AGI would significantly affect the housing inspection scores but by creating these graphs we realized that a state like California with by far the highest AGI had a pretty ordinary average inspection score. At the same time there were low AGI states like Kansas and Connecticut which were on completely opposite sides of the inspection score spectrum while having about the same income.

We wanted to be able to show county maps of selected states and the scores/incomes of counties similar to the first two maps but unfortunately Google Visualization API doesn’t accommodate county maps. The only alternative would have been using the longitude/latitude values provided in dataset 1258 but those would only show housing locations not entire counties.

Learned how to segregate datasets based on graph name which let us use common attribute names without clashing with other datasets.

**6.4. Things that could have been done differently:**

While the project came out the way we wanted it we did think of additional information we could have shown if we had time. Information such as adding data from the “multihousing inspection scores” dataset or even factoring in county population to see if any additional relationships exist or could be seen between the data.