

CSCI 403 | Database Management

World Happiness & World Religions

Project 8 - Create

Ryan Fast
12-6-2017

1 The Datasets

In the data driven world we live in today, it is clear to see how there are countless sources available to us to find information on just about anything. Furthermore, with all the complexities, redundancies, and massive quantities of data available we can easily recognize just how important it is to manage and store key data in an intelligent and efficient way. For this project I strove to simulate just that by finding three very interesting datasets and manage to work through technical difficulties and obstacles to bring to life a useful database.

1.1 Data Description and Interest

After taking much time to cruise through the links provided on the course website and many other sites found in further searching, I settled on three somewhat related and very intriguing data sets from various sources.

The first dataset that caught my eye was a world happiness report. This dataset reports on and ranks 155 countries by their 'happiness score'. This score from zero to ten, ten being perfectly and completely happy and zero indicating a person with the worst possible life, is derived through a series of 'happiness indicators' or factors that contribute to the overall score. In my basic research into many of the attributes and validity of this dataset I found that this dataset is actually pretty widely recognized. The source of the information present in this dataset originates from the Gallup World Poll by Gallup Advanced Analytics [1]. Gallup's site itself describes the reasoning behind this poll was and still is to 'make decisions based on the will of the people' through determination of how people view their own state of being. This report first published in 2012 with annual publications from then till the most recent addition in 2017, continues to gain global attention by governments, organizations, and experts whom use happiness indicators to make critical decisions and analysis on the state of the world today. The most notable and interesting attributes present in this dataset are the contribution factors that tie into the country's rank and score out of ten. For example, the happiness report based posed questions to individuals to determine how seven key factors contributed to a person's overall score out of ten. The factors include, the contribution that the state of the economy had on a person's happiness indicated by gdp per capita, the contribution how content someone is with their family, how the feeling of life expectancy in their country contributed to their score, how much freedom they felt they had, how much trust in the government or feeling of corruption they sensed in the government, how much generosity they felt on a daily basis, and finally how much contribution their sense of their current state of being compared to the dystopia residual and how it contributed to their score. Dystopia residual is a fictional country that is set to have the worst possible living conditions imaginable. This is something that every country, regardless of the situation could compare positively to, to help scale the scores out of ten appropriately. The draw of this dataset is undoubtedly the meaning behind the happiness scores and the far-reaching implications of what a difference in score from one country to another or one year to another in the same country could mean about the state of the world and the people living in it. After understanding this

dataset in depth, I knew it would be extremely interesting and valuable to be able to make powerful queries against this data alone. Using psql and the course database framework provided, I knew I could ask some very engaging questions that would allow for even more meaning to be revealed by the data.

After researching and understanding the happiness report dataset described above, I sought to find another dataset to connect to this report and provide further interest and meaning behind any queries I could do against the database. Since the happiness report is a worldwide report, I wanted to get key aspect that was shared across the world that the happiness report did not take into account. To this end I found my next dataset, a world religions dataset. This dataset provides raw data on the populations of over 30 religions in approximately 200 countries around the globe. When investigating this dataset, I was pleasantly surprised how complete and in-depth it was. The with the large quantities of religion from very popular to small sects of major religions I believe this dataset effectively captures the crucial information required to do real analysis about religions across the world. This dataset is credited to be created by Zeev Maoz of the University of California-Davis, and Errol Henderson of Pennsylvania State University, and was published as part of something known as the Correlates of War Project. Intrigued why a project to seemingly produce data for correlations surrounding war would produce a world religions dataset I did a bit of further research. The Correlates of War Project site defines their goals as “seek[ing] to facilitate the collection, dissemination, and use of accurate and reliable quantitative data in international relations.” [2]. The site hosts several very interesting datasets that almost caused me to switch gears, however, nevertheless I felt that I could be sure of the validity of this world religions data set. Alone, this dataset offers a bit of variety, however simple, opportunities to make queries against. Paired with the happiness report dataset however, I believed there were some extremely powerful questions I could ask to reveal any correlations with happiness and religion.

Finally, with my two major and interesting data sets sourced I still had yet to feel that I could effectively visualize the key data and correlations I could possibly make. When searching and thinking of ways I would like to see this data presented, I stumbled upon software that could utilize CSV formatted inputs for values of data and display them on a map. The only thing I was missing was the latitude and longitude of the countries on both my happiness report and world religions datasets. Thus, for my third data set I found a very simple dataset that lists over 200 major countries and their respective latitudes and longitudes.

With all three of the datasets described above I believed my database was complete. Through the power of psql and relational databases I knew that I could make involved and powerful queries, answering interesting questions, and then take it a step further to visualize this data on a global map.

1.2 Source and Licensing Information

Before moving onto the real usage and results of bringing my datasets together, it is important to recognize just where this real-world information came from and make a comment on the possible licensing limitations or restrictions.

The happiness report dataset was found and taken from the site Kaggle. This site hosts a free service for users to create, share, modify, update, and manage everything dataset related. Kaggle takes in and reviews user submissions and grants licenses for user privacy and each user acknowledges that the dataset is hosted for public access after submission. This means that there appears to be no licensing issues or restrictions for users, such as myself, to use the datasets posted on there site. Additionally, with the happiness report the submission of this dataset has specific release information that confirms the above. This report was released under ‘CC0: Public Domain’. This means that ‘you can copy, modify, distribute and perform the work, even for commercial purposes, all without asking permission,’ which ensures all the data used from this report is free of any licensing issues [3].

Similarly, the world religions report was also found on Kaggle. As this dataset did not have additional release information, I researched the war collates project briefly described above. There site indicated that all software and data used is being published and distributed under a ‘GNU GPL License’. Researching this license, I found that ‘he GNU General Public License is intended to guarantee your freedom to share and change free software--to make sure the software is free for all its users.’ [4]. The rather dense licensing information seemingly indicates that, like the happiness report, there should be no issues with the usage of this dataset.

Finally, the geographical locations dataset was found on ‘developers.google.com’ in there public data, dataset publishing language section for creating an sharing data and code surrounding datasets. Similar to the happiness report this site linked to additional release information that indicated the information in the dataset was published under the ‘Creative Commons Attribution 3.0 License’. This license, much like the one described for the happiness report, allows for users to ‘copy and redistribute the material in any medium or format’ [5].

Overall, all research and information found surrounding the datasets used for this project ensures that there should be no licensing and usage and distribution of any of the data present.

1.3 Database Significance

To bring back the focus to the primary goal of this project, it is crucial to discuss the significant aspects of the tables and relations that result from the chosen datasets. The following provides a description for the major tables created to hold the bulk data in the datasets.

One of the primary tables holding the some of the most valuable raw data copied into my schema is the ‘Happiness_Report’ table. This table holds the core information and attributes for the happiness report by the Gallup World Poll. The most significant attributes present in this table are the seven key factors and their contributions to each countries’ overall rank and score out of

ten. Each factor is of integer type and to add up all the factors' contributions you would achieve the total score out of ten that a country's population on average sees their current state of being. Additionally, it is important to note that the datasets pulled from the world happiness report provided data for 2015, 2016, and 2017. Thus, notably, an attribute for the year the data was published was added in this table to provide opportunity to see changes in the data across a timeline (e.g. changes in the countries ranks or scores over time).

The next primary table for this database and schema is the 'Religion_by_nation' table. This table holds the raw information copied from the world religions dataset pulled of the war correlates project. This table notably contains the year of the population data pulled for that specific country and then the over 30 religions to list their populations for each country and year. No additional attributes were necessary to be added to organize this table into my desired schema.

Finally, the last primary table added to the database was the 'Country_Location' table. This table holds the simple dataset pulled together for the over 200 countries geographical locations. This table simply holds attributes of country, latitude, and longitude.

When the initial creation and population of the primary tables was successful I began trying to come up with involved and interesting query mashups to get results from the database. I quickly realized that the primary tables could not effectively accomplish many of the questions I had. Specifically, I thought that asking questions such as 'what factor contributed the most to the happiest country and what is the predominate religion of the that country?' were very interesting questions. While this is easy to figure out just looking at the CSV datasets, I found it was actually impossible to have psql query and report back the attribute name of with the greatest population. For this error and many others, I will describe in the later sections, I created two additional tables to help psql provide by intelligent responses to posed questions.

First, I added the 'religion_by_type' table. This table's most significant attribute is the religion type where I could now store a string for every instance of a population and country. To populate this table I had to do some interesting updates and alters after querying the 'religion_by_nation' table. With this table I can now report the religion name with the largest population of a single country. Similarly, the next and final secondary table I added was the 'Happiness_by_factor' table. Like the 'religion_by_type' table, this helped me report intelligent results by allowing me to store and query on a string of each contribution factor.

The figure below shows a basic visual representation of the table described above that make up the entirety of my database for this project.

Happiness_Report	Religion_by_nation	Country_Location
County Region Happiness_rank Happiness_score Gdp_percapita Family Life_expectancy Freedom Gov_Corruption Generosity Dystopia_Residual Year_of_Data	Year_of_Data Country Code Christianity_Protestant Christianity_RomanCatholic ... Judaism_orthodox Judaism_conservative ... Islam_sunni Islam_shi'a ...	Country Latitude Longitude
Religion by type	Happiness report by factor	
Country Religion_Type Population Year_of_Data	Country Contributon_factor Contribution Year_of_data	

Figure 1: Database Table Visualization

2 Data Usage and Visualization

The most powerful and clever uses of my datasets came in the form of query mash ups and data visualization on a world map. I believe the combination of the two provides insight into just how useful a database like this could be and how much potential this set could have.

2.1 Query Mashups

When selecting the happiness report and world religion datasets I knew I could ask some very interesting and involved questions. The focal point of what I wanted to accomplish was to see if there were any correlations to the scores and ranks of countries around the globe and the predominate religions present in the area. When I first selected the happiness report I was immediately curious how factors such as family, generosity, and government corruption could affect how many of us perceive our happiness. Alone, I thought these correlations were very interesting in themselves. I believe there were numerous worthy questions to be asked such as the following:

- How did the happiest country change from 2015 to 2016 to 2017?
- What country in 2017 was freedom the largest contributor out of all other countries and where does that country rank?

- What are the largest contributions to the happiness score for all the countries ranked in 2015?
 - How did these change from 2016 and 2017?

I believe all of these queries are very intriguing and my database most certainly allows for these questions to be answered. However, I believe the real value of my database comes from the addition of the world religions dataset. With this I felt I was basically investigating another possible contribution factor that the world happiness report did not take into account. Religion is undoubtedly present in everyday life culture and across the globe. Especially, in countries such as the middle east, the very religion that is predominate in the area would drastically affect your rights and quality of life. For example, women's rights in the countries with a predominate Islamic belief is often a topic of conversation for organizations hoping to improve quality of life across the globe. Thus, with the addition of the world religions dataset my database now offered a brand-new level of complexity to produce even more valuable and involved answers to questions. The following list provides some interesting questions I was able to answer with the final database schema:

- Where does the United States rank on the happiness scale in 2017 and what is the predominant religion using the most recent data available?
- What are the happiness score and rank of country with the most non-religious populations in 2017?
- What are the predominate religions in the countries that have a happiness score of 4 or lower? Order by score.
- How much does trust in the government do countries contribute to happiness in the top 5 countries with the most population the sunni sect of Islam have?
 - How about protestant sect of Christianity? (using the most recent data)
- Of the countries that generosity contributed 0.4 or more to the overall score what is the predominant religion in those countries? Order by generosity contribution
- Of the range dates available in the regional data set for religion what how much has the protestant religion of the United States change in population from earliest date to most recent?

The queries created to answer these questions can be seen in the appendix of this write-up.

2.2 World Map Data visualization

To take this a step further, I wanted my database to be able to do more than just offer query results. As previously touched on, this is where the addition of the country locations dataset came into scope. Using a free online software and outputs of query results of my database tables I was able to map and visualize the core data available in my database. The software used was Colore Maps at 'coloremeps.com'. This site allows the user to focus on geographical locations around the world and then supply markers and values to visualize information. Through a join on the happiness report and country locations table and similarly the world religions and country locations table I can now visualize data in the database in a dynamic way.

Using the happiness report data and countries latitude and longitude I was able to visualize where the countries that I had information on in my database were located relative to one another and further show the rank of each country.



Figure 2: Americas Happiness Ranks Visualization

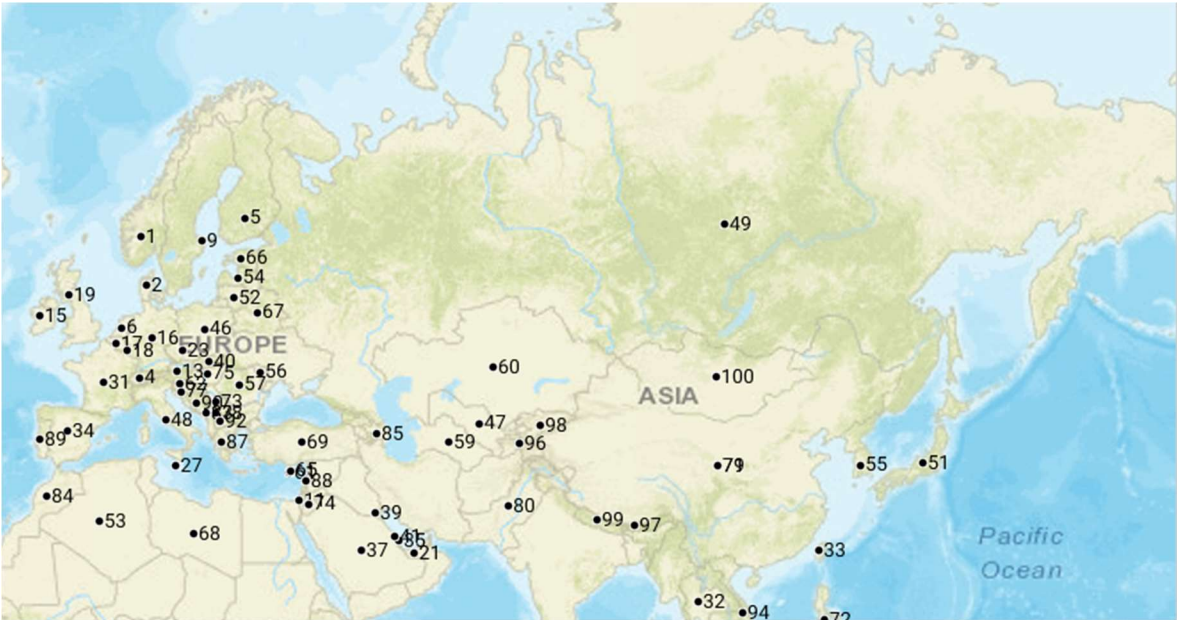


Figure 3: Europe, Middle East, and Asia Happiness Rank Visualization

Furthermore, with the usage of the world religions and country locations datasets I could produce a bubble map based on populations for each religion and label the countries by name.



Figure 4: Global Visualization of Christianity Population

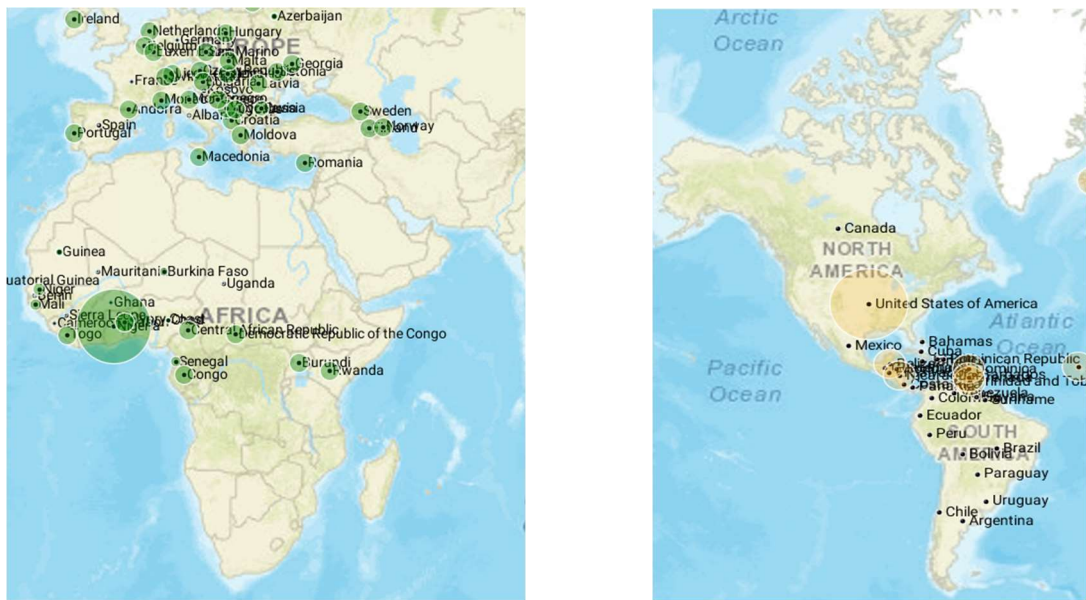


Figure 5 -
 LEFT: Islamic Population Visual in Africa and Europe
 RIGHT: Judaism Population Visual in the Americas

These visualizations allow for even more valuable correlations and information about the happiness of areas around the world. For example, we can see that geographical location may have a contribution partial to happiness. Such that maybe the climate or that the proximity of the happier countries seems to concentrate possible indicating that happiness can be influenced by the state of surrounding areas.

3 Technical Challenge

Reflecting back on this project, there was quite a bit of technical challenge throughout each aspect. The first challenge presented in this project was how messy the real-world data sets were from retrieving them from their original sources. In attempts to keep the unnecessary data to a minimum and inch towards have my database in a more normal form I had to do a lot of extra processing in excel and within the database itself to properly populate attributes and tables. For example, the happiness report and world religions datasets had a lot of redundant data. In early attempts to copy over the bulk data into the database and apply key restraints to the incoming data I had to overcome many instances of duplicate information in the datasets. I was able to circumvent this problem by using two methods. First, I used a bit of pre-processing in excel to identify as many duplicates I could. For the smaller datasets this was a sufficient method and was ultimately successfully. However, for the larger datasets I used temporary tables in the database and used helpful sql queries to populate the real tables I wanted to use. Operators such as `SELECT DISTINCT`, proved very useful for cases such as this. Similarly, I noticed in the raw CSV files of the datasets that there were many instances of inferred attributes that were unnecessary and could be calculated in a sql query rather than take up space in the database. For example, the world religions database had about twice as many columns in the CSV than what I ingested into the database, due to the fact that there were totals and sums of populations for each religion. I simply deleted these attributes in the CSV and to keep my data in a more normal state. One of the biggest challenges presented itself when I was ready to query on my primary tables. As described above, I realized that many of the queries I wanted to do were impossible under the current state of my datasets. After hitting this obstacle, I was almost ready and looking to switch datasets completely. Luckily, using some database management and psql skills I was able to think through this issue by creating better, more intelligent tables and populate them with the useful information from the primary tables. Another challenge arose when trying to take my data to the next level and visualize it with the online software I found. I did not know of the best way to take large query results and kick it back out to a usable format to manipulate further. This problem was overcome by a bit of extra research and help from canvas. The solution was as easy as copying results back out of sql into a CSV, essentially reversing what was done to bulk load data into the database. This perfectly fit the software I intended to use as the markers you are able to place on a world map required a comma delineated format of latitude, longitude, label, and optional marker size value. Finally, the last technical challenge I faced was the query mashing I wanted to do with my datasets. I believe I had to get pretty fancy and really test my sql skills with the documentation provided on postgresql to overcome this challenge. I used brand new aggregate functions such as `GREATEST` that I never used before and I had to have a lot of 'fun' with subqueries to achieve the results I really wanted.

4 Works Cited Links

[1] <http://analytics.gallup.com/213704/world-poll.aspx>

[2] <http://www.correlatesofwar.org/data-sets>

[3] <https://creativecommons.org/publicdomain/zero/1.0/>

[4] <http://www.gnu.org/licenses/gpl-2.0.html>

[5] <https://creativecommons.org/licenses/by/3.0/>

5 Appendix and Notes

Please see the following attachments.

- Project8_dbCreation.sql
 - This is the script used to create and populate the primary and secondary tables that make up my database.
- Project8_queryMashup.sql
 - This is the script used that will report results to numerous interesting sample questions I posed in this write-up.

Additionally, with in this submissions zip file you can see the raw data of all the datasets used in this project.

*NOTE: the all data and tables were created under my schema. Username: rfast.

```

1  /*
2  *  project8_dbCreation.sql
3  *
4  *  author: Ryan Fast
5  *
6  */
7
8  /*-----*/
9  /* -----CREATING THE DATABASE -----*/
10 /*-----*/
11
12 /*-----*/
13 /* --- Creating flat dataset tables --- */
14 /*-----*/
15
16 /* Creating the world happiness report table */
17 /* Dropping happiness_report table and all dependencies if it exists */
18 DROP TABLE IF EXISTS happiness_report CASCADE;
19
20 /* Creating table with necessary attributes */
21
22 CREATE TABLE happiness_report(
23     Country VARCHAR(200),
24     Region VARCHAR(50),
25     Happiness_Rank INTEGER,
26     Happiness_Score DOUBLE PRECISION,
27     GDP_PerCapita DOUBLE PRECISION,
28     Family DOUBLE PRECISION,
29     Life_Expectancy DOUBLE PRECISION,
30     Freedom DOUBLE PRECISION,
31     Gov_Corruption DOUBLE PRECISION,
32     Generosity DOUBLE PRECISION,
33     Dystopia_Residual DOUBLE PRECISION,
34     Year_of_Data INTEGER,
35     PRIMARY KEY(Happiness_Rank, Year_of_data)
36
37 );
38
39
40 /* Creating the world religions by nation report table */
41 /* Dropping religion_by_region table and all dependencies if it exists */
42 DROP TABLE IF EXISTS religion_by_nation CASCADE;
43
44 /* Creating table with necessary attributes */
45
46 CREATE TABLE religion_by_nation(
47     Year_of_data INTEGER,
48     Country VARCHAR(100),
49     Code VARCHAR(10),
50     christianity_protestant INTEGER,
51     christianity_romancatholic INTEGER,
52     christianity_easternorthodox INTEGER,
53     christianity_anglican INTEGER,
54     christianity_other INTEGER,
55     judaism_orthodox INTEGER,
56     judaism_conservative INTEGER,
57     judaism_reform INTEGER,
58     judaism_other INTEGER,
59     islam_sunni INTEGER,
60     islam_shi'a INTEGER,
61     islam_ibadhi INTEGER,
62     islam_alawite INTEGER,
63     islam_ahmadiyya INTEGER,
64     islam_other INTEGER,
65     buddhism_mahayana INTEGER,
66     buddhism_theravada INTEGER,
67     buddhism_other INTEGER,
68     zoroastrianism_all INTEGER,
69     hinduism_all INTEGER,

```

```

70  sikhism_all  INTEGER,
71  shinto_all  INTEGER,
72  baha'i_all  INTEGER,
73  taoism_all  INTEGER,
74  jainism_all  INTEGER,
75  confucianism_all  INTEGER,
76  syncretism_all  INTEGER,
77  animism_all  INTEGER,
78  noreligion_all  INTEGER,
79  PRIMARY KEY (Year_of_Data, Country, Code)
80  );
81
82  /* Creating the world happiness report table */
83  /* Droping country_location table and all dependencies if it exists */
84  DROP TABLE IF EXISTS country_location CASCADE;
85
86  /* Creating table with nessesary attributes */
87
88  CREATE TABLE country_location(
89      Country VARCHAR(200),
90      Latitude DOUBLE PRECISION,
91      Longitude DOUBLE PRECISION
92
93      );
94
95  /*-----*/
96  /* --- Creating better tables from the original dataset tables --- */
97  /*-----*/
98
99  DROP TABLE IF EXISTS happness_report_by_factor CASCADE;
100
101  CREATE TABLE happness_report_by_factor (
102      country varchar(100),
103      contribution_factor varchar(100),
104      contribution DOUBLE PRECISION,
105      year_of_data INTEGER
106
107  );
108
109
110  /* Creating the world religion_by_type table */
111  /* Droping religion_by_type table and all dependencies if it exists */
112  DROP TABLE IF EXISTS religion_by_type CASCADE;
113  /* Creating table with nessesary attributes */
114  CREATE TABLE religion_by_type (
115      country varchar(100),
116      religion_type varchar(200),
117      population INTEGER,
118      year_of_data INTEGER
119
120  );
121
122  /*-----*/
123  /* ----- Populating Data into the flat unmodified tables using COPY ----- */
124  /*-----*/
125  /* ---- Populating Happiness_Report ---- */
126  /* Copying in 2015 data */
127  \copy happiness_report FROM '2015.csv' DELIMITER ',' CSV HEADER;
128
129  /* Copying in 2016 data */
130  \copy happiness_report FROM '2016.csv' DELIMITER ',' CSV HEADER;
131
132  /* Copying in 2017 data */
133  \copy happiness_report FROM '2017.csv' DELIMITER ',' CSV HEADER;
134
135  /* ---- Populating religion_by_nation ---- */

```

```

136 \copy religion_by_nation FROM 'national.csv' DELIMITER ',' CSV HEADER;
137
138 /* ---- Populating country_location ----*/
139 \copy country_location FROM 'LocationCSVSScrubbed.csv' DELIMITER ',' CSV HEADER;
140
141
142 /*-----*/
143 /* ----- Populating Data into the the better, more inteligent
tables using queries-----*/
144 /*-----*/
145 /*-----*/
146 /*-----Happiness_Report_By_Factor-----*/
147 /*-----*/
148 /* ---- Populating and Updating Happiness_Report_by_factor for GDP_PerCapita ---- */
149
150 INSERT INTO happiness_report_by_factor(country, contribution, year_of_data)
151 SELECT hr.country, hr.GDP_PerCapita, hr.year_of_data
152 FROM happiness_report AS hr;
153
154 UPDATE happiness_report_by_factor
155 SET contribution_factor = 'GDP_PerCapita';
156
157
158 /* ---- Populating and Updating Happiness_Report_by_factor for family ---- */
159
160 INSERT INTO happiness_report_by_factor(country, contribution, year_of_data)
161 SELECT hr.country, hr.Family, hr.year_of_data
162 FROM happiness_report AS hr;
163
164 UPDATE happiness_report_by_factor
165 SET contribution_factor = 'Family'
166 WHERE contribution_factor IS NULL;
167
168
169 /* ---- Populating and Updating Happiness_Report_by_factor for Life_Expectancy ---- */
170
171 INSERT INTO happiness_report_by_factor(country, contribution, year_of_data)
172 SELECT hr.country, hr.Life_Expectancy, hr.year_of_data
173 FROM happiness_report AS hr;
174
175 UPDATE happiness_report_by_factor
176 SET contribution_factor = 'Life_Expectancy'
177 WHERE contribution_factor IS NULL;
178
179
180 /* ---- Populating and Updating Happiness_Report_by_factor for Freedom ---- */
181
182 INSERT INTO happiness_report_by_factor(country, contribution, year_of_data)
183 SELECT hr.country, hr.Freedom, hr.year_of_data
184 FROM happiness_report AS hr;
185
186 UPDATE happiness_report_by_factor
187 SET contribution_factor = 'Freedom'
188 WHERE contribution_factor IS NULL;
189
190
191 /* ---- Populating and Updating Happiness_Report_by_factor for Gov_Corruption ---- */
192
193 INSERT INTO happiness_report_by_factor(country, contribution, year_of_data)
194 SELECT hr.country, hr.Gov_Corruption, hr.year_of_data
195 FROM happiness_report AS hr;
196
197 UPDATE happiness_report_by_factor
198 SET contribution_factor = 'Gov_Corruption'

```



```

199 WHERE contribution_factor IS NULL;
200
201
202 /* ---- Populating and Updating Happiness_Report_by_factor for Generosity ---- */
203
204 INSERT INTO happiness_report_by_factor(country, contribution, year_of_data)
205 SELECT hr.country, hr.Generosity, hr.year_of_data
206 FROM happiness_report AS hr;
207
208 UPDATE happiness_report_by_factor
209 SET contribution_factor = 'Generosity'
210 WHERE contribution_factor IS NULL;
211
212
213 /* ---- Populating and Updating Happiness_Report_by_factor for Dystopia_Residual ---- */
214
215 INSERT INTO happiness_report_by_factor(country, contribution, year_of_data)
216 SELECT hr.country, hr.Dystopia_Residual, hr.year_of_data
217 FROM happiness_report AS hr;
218
219 UPDATE happiness_report_by_factor
220 SET contribution_factor = 'Dystopia_Residual'
221 WHERE contribution_factor IS NULL;
222
223
224 /*-----*/
225 /*-----Religion_by_Type-----*/
226 /*-----*/
227 /*christianity_protestant*/
228 INSERT INTO religion_by_type(country, population, year_of_data)
229 SELECT rn.country, rn.christianity_protestant, rn.year_of_data
230 FROM religion_by_nation AS rn;
231 UPDATE religion_by_type
232 SET religion_type = 'christianity_protestant'
233 WHERE religion_type IS NULL;
234 /*christianity_romancatholic*/
235 INSERT INTO religion_by_type(country, population, year_of_data)
236 SELECT rn.country, rn.christianity_romancatholic, rn.year_of_data
237 FROM religion_by_nation AS rn;
238 UPDATE religion_by_type
239 SET religion_type = 'christianity_romancatholic'
240 WHERE religion_type IS NULL;
241 /*christianity_easternorthodox*/
242 INSERT INTO religion_by_type(country, population, year_of_data)
243 SELECT rn.country, rn.christianity_easternorthodox, rn.year_of_data
244 FROM religion_by_nation AS rn;
245 UPDATE religion_by_type
246 SET religion_type = 'christianity_easternorthodox'
247 WHERE religion_type IS NULL;
248 --here
249 /*christianity_anglican*/
250 INSERT INTO religion_by_type(country, population, year_of_data)
251 SELECT rn.country, rn.christianity_anglican, rn.year_of_data
252 FROM religion_by_nation AS rn;
253 UPDATE religion_by_type
254 SET religion_type = 'christianity_anglican'
255 WHERE religion_type IS NULL;
256 /*christianity_other*/
257 INSERT INTO religion_by_type(country, population, year_of_data)
258 SELECT rn.country, rn.christianity_other, rn.year_of_data
259 FROM religion_by_nation AS rn;
260 UPDATE religion_by_type
261 SET religion_type = 'christianity_other'
262 WHERE religion_type IS NULL;
263 /*judaism_orthodox*/
264 INSERT INTO religion_by_type(country, population, year_of_data)

```



```

265 SELECT rn.country, rn.judaism_orthodox, rn.year_of_data
266 FROM religion_by_nation AS rn;
267 UPDATE religion_by_type
268 SET religion_type = 'judaism_orthodox'
269 WHERE religion_type IS NULL;
270 /*judaism_conservative*/
271 INSERT INTO religion_by_type(country, population, year_of_data)
272 SELECT rn.country, rn.judaism_conservative, rn.year_of_data
273 FROM religion_by_nation AS rn;
274 UPDATE religion_by_type
275 SET religion_type = 'judaism_conservative'
276 WHERE religion_type IS NULL;
277 /*judaism_reform*/
278 INSERT INTO religion_by_type(country, population, year_of_data)
279 SELECT rn.country, rn.judaism_reform, rn.year_of_data
280 FROM religion_by_nation AS rn;
281 UPDATE religion_by_type
282 SET religion_type = 'judaism_reform'
283 WHERE religion_type IS NULL;
284 /*judaism_other*/
285 INSERT INTO religion_by_type(country, population, year_of_data)
286 SELECT rn.country, rn.judaism_other, rn.year_of_data
287 FROM religion_by_nation AS rn;
288 UPDATE religion_by_type
289 SET religion_type = 'judaism_other'
290 WHERE religion_type IS NULL;
291 /*islam_sunni*/
292 INSERT INTO religion_by_type(country, population, year_of_data)
293 SELECT rn.country, rn.islam_sunni, rn.year_of_data
294 FROM religion_by_nation AS rn;
295 UPDATE religion_by_type
296 SET religion_type = 'islam_sunni'
297 WHERE religion_type IS NULL;
298 /*islam_shi'a*/
299 INSERT INTO religion_by_type(country, population, year_of_data)
300 SELECT rn.country, rn.islam_shi'a, rn.year_of_data
301 FROM religion_by_nation AS rn;
302 UPDATE religion_by_type
303 SET religion_type = 'islam_shi'a'
304 WHERE religion_type IS NULL;
305 /*islam_ibadhi*/
306 INSERT INTO religion_by_type(country, population, year_of_data)
307 SELECT rn.country, rn.islam_ibadhi, rn.year_of_data
308 FROM religion_by_nation AS rn;
309 UPDATE religion_by_type
310 SET religion_type = 'islam_ibadhi'
311 WHERE religion_type IS NULL;
312 /*islam_alawite*/
313 INSERT INTO religion_by_type(country, population, year_of_data)
314 SELECT rn.country, rn.islam_alawite, rn.year_of_data
315 FROM religion_by_nation AS rn;
316 UPDATE religion_by_type
317 SET religion_type = 'islam_alawite'
318 WHERE religion_type IS NULL;
319 /*islam_ahmadiyya*/
320 INSERT INTO religion_by_type(country, population, year_of_data)
321 SELECT rn.country, rn.islam_ahmadiyya, rn.year_of_data
322 FROM religion_by_nation AS rn;
323 UPDATE religion_by_type
324 SET religion_type = 'islam_ahmadiyya'
325 WHERE religion_type IS NULL;
326 /*islam_other*/
327 INSERT INTO religion_by_type(country, population, year_of_data)
328 SELECT rn.country, rn.islam_other, rn.year_of_data
329 FROM religion_by_nation AS rn;
330 UPDATE religion_by_type
331 SET religion_type = 'islam_other'
332 WHERE religion_type IS NULL;
333 /*buddhism_mahayana*/

```

```

334 INSERT INTO religion_by_type(country, population, year_of_data)
335 SELECT rn.country, rn.buddhism_mahayana, rn.year_of_data
336 FROM religion_by_nation AS rn;
337 UPDATE religion_by_type
338 SET religion_type = 'buddhism_mahayana'
339 WHERE religion_type IS NULL;
340 /*buddhism_theravada*/
341 INSERT INTO religion_by_type(country, population, year_of_data)
342 SELECT rn.country, rn.buddhism_theravada, rn.year_of_data
343 FROM religion_by_nation AS rn;
344 UPDATE religion_by_type
345 SET religion_type = 'buddhism_theravada'
346 WHERE religion_type IS NULL;
347 /*buddhism_other*/
348 INSERT INTO religion_by_type(country, population, year_of_data)
349 SELECT rn.country, rn.buddhism_other, rn.year_of_data
350 FROM religion_by_nation AS rn;
351 UPDATE religion_by_type
352 SET religion_type = 'buddhism_other'
353 WHERE religion_type IS NULL;
354 /*zoroastrianism_all*/
355 INSERT INTO religion_by_type(country, population, year_of_data)
356 SELECT rn.country, rn.zoroastrianism_all, rn.year_of_data
357 FROM religion_by_nation AS rn;
358 UPDATE religion_by_type
359 SET religion_type = 'zoroastrianism_all'
360 WHERE religion_type IS NULL;
361 /*hinduism_all*/
362 INSERT INTO religion_by_type(country, population, year_of_data)
363 SELECT rn.country, rn.hinduism_all, rn.year_of_data
364 FROM religion_by_nation AS rn;
365 UPDATE religion_by_type
366 SET religion_type = 'hinduism_all'
367 WHERE religion_type IS NULL;
368 /*sikhism_all*/
369 INSERT INTO religion_by_type(country, population, year_of_data)
370 SELECT rn.country, rn.sikhism_all, rn.year_of_data
371 FROM religion_by_nation AS rn;
372 UPDATE religion_by_type
373 SET religion_type = 'sikhism_all'
374 WHERE religion_type IS NULL;
375 /*shinto_all*/
376 INSERT INTO religion_by_type(country, population, year_of_data)
377 SELECT rn.country, rn.shinto_all, rn.year_of_data
378 FROM religion_by_nation AS rn;
379 UPDATE religion_by_type
380 SET religion_type = 'shinto_all'
381 WHERE religion_type IS NULL;
382 /*baha'i_all*/
383 INSERT INTO religion_by_type(country, population, year_of_data)
384 SELECT rn.country, rn.baha'i_all, rn.year_of_data
385 FROM religion_by_nation AS rn;
386 UPDATE religion_by_type
387 SET religion_type = 'baha'i_all'
388 WHERE religion_type IS NULL;
389 /*taoism_all*/
390 INSERT INTO religion_by_type(country, population, year_of_data)
391 SELECT rn.country, rn.taoism_all, rn.year_of_data
392 FROM religion_by_nation AS rn;
393 UPDATE religion_by_type
394 SET religion_type = 'taoism_all'
395 WHERE religion_type IS NULL;
396 /*jainism_all*/
397 INSERT INTO religion_by_type(country, population, year_of_data)
398 SELECT rn.country, rn.jainism_all, rn.year_of_data
399 FROM religion_by_nation AS rn;
400 UPDATE religion_by_type
401 SET religion_type = 'jainism_all'
402 WHERE religion_type IS NULL;

```

```

403  /*confucianism_all*/
404  INSERT INTO religion_by_type(country, population, year_of_data)
405  SELECT rn.country, rn.confucianism_all, rn.year_of_data
406  FROM religion_by_nation AS rn;
407  UPDATE religion_by_type
408  SET religion_type = 'confucianism_all'
409  WHERE religion_type IS NULL;
410  /*syncretism_all*/
411  INSERT INTO religion_by_type(country, population, year_of_data)
412  SELECT rn.country, rn.syncretism_all, rn.year_of_data
413  FROM religion_by_nation AS rn;
414  UPDATE religion_by_type
415  SET religion_type = 'syncretism_all'
416  WHERE religion_type IS NULL;
417  /*animism_all*/
418  INSERT INTO religion_by_type(country, population, year_of_data)
419  SELECT rn.country, rn.animism_all, rn.year_of_data
420  FROM religion_by_nation AS rn;
421  UPDATE religion_by_type
422  SET religion_type = 'animism_all'
423  WHERE religion_type IS NULL;
424  /*noreligion_all*/
425  INSERT INTO religion_by_type(country, population, year_of_data)
426  SELECT rn.country, rn.noreligion_all, rn.year_of_data
427  FROM religion_by_nation AS rn;
428  UPDATE religion_by_type
429  SET religion_type = 'noreligion_all'
430  WHERE religion_type IS NULL;
431
432
433  /* Issues/bugs list that I ran into and had to resolve */
434  /*
435
436  - Datasets had alot of redundant data. SOLVED: Post processing in the CSV to eliminate
  redundancies
437
438  - Datasets had many 'inferred' attributes that could be calculated by the more
  significant attributes. SOLVED: By taking out the attributes when populating
439  database to keep the tables in more normal form
440
441  - Datasets had many errors and duplicated information making it difficult to put table
  constraints such as a primary key. SOLVED: Pre-processing the data
442  in excel allowed me to identify errors. Additionally, when injesting the data into
  the my schema the output error messages from psql helped my greatly
443  in identifying duplicates that I missed.
444
445  - Relating the datasets for query mash-ups. Becuase the two datasets that I pulled
  together were so diffierent it was easy to identify interesting queries
446  I wanted to make against my database, however, it was difficult to find ways in many
  of the joins to effectively accomplish several of the queries. SOLVED:
447  I was able to solve this by much trail and error, where I would work hard to change
  my queries up to fetch the data I wanted, before finally modifying my
448  schema and data itself to better suit the questions I wanted to ask about my two
  datasets.
449
450  - Can't query and report back an attribute name. SOLVED: need to add two additional
  table to hold a string for the religion/factor name to be able to query
451  for intelligent/useful results.
452
453  - Difficulty in finding easy way to get output for data visulization; SOLVED: from
  canvas we can used the copy function to get bulk data back out of the data
454  based. e.g. COPY tablename TO file.csv WITH (FORMAT csv);
455
456
457  */
458
459
460
461

```

```

1  /*
2  *  project8_queryMashUp.sql
3  *
4  *  author: Ryan Fast
5  *
6  */
7
8
9  /*-----*/
10 /* -----Interesting Query Mash Ups! -----*/
11 /*-----*/
12 /* The following queries provide very interesting results with the two datasets
13    ingested into
14    my schema. Basic Descriptions/Questions to answer are provided for the expected
15    results of the query */
16 /*-----*/
17
18 /*Where does the United States rank on the happiness scale in 2017 and what is the
19    predominant religion using the most
20    recent data available? */
21
22 SELECT h.country, happiness_rank, religion_type
23 FROM happiness_report AS h, religion_by_type AS rt
24 WHERE h.country = 'United States' AND rt.country = 'United States of America' AND
25 h.year_of_data = 2017
26 AND rt.population =
27 (SELECT GREATEST(
28     christianity_protestant ,
29     christianity_romancatholic ,
30     christianity_easternorthodox ,
31     christianity_anglican ,
32     christianity_other ,
33     judaism_orthodox ,
34     judaism_conservative ,
35     judaism_reform ,
36     judaism_other ,
37     islam_sunni ,
38     islam_shi'a ,
39     islam_ibadhi ,
40     islam_alawite ,
41     islam_ahmadiyya ,
42     islam_other ,
43     buddhism_mahayana ,
44     buddhism_theravada ,
45     buddhism_other ,
46     zoroastrianism_all ,
47     hinduism_all ,
48     sikhism_all ,
49     shinto_all ,
50     baha'i_all ,
51     taoism_all ,
52     jainism_all ,
53     confucianism_all ,
54     syncretism_all ,
55     animism_all ,
56     noreligion_all
57 )
58 FROM religion_by_nation
59 WHERE code = 'USA'
60 AND year_of_data = (SELECT MAX(year_of_data) FROM religion_by_nation));
61
62 /* What country in 2017 was freedom the largest contributor out of all other contries
63    and where does
64    that country rank? */
65
66 SELECT country, happiness_rank
67 FROM happiness_report
68 WHERE year_of_data = 2017 AND freedom = (

```

```

65     SELECT MAX(freedom) FROM happiness_report
66     WHERE year_of_data = 2017);
67
68
69  /* What are the happiness score and rank of country with the most non-religious
70  populations in 2017? */
71
72     SELECT h.country, happiness_rank, happiness_score
73     FROM happiness_report AS h, religion_by_nation AS rn
74     WHERE h.country = rn.country AND h.year_of_data = 2017
75     AND rn.noreligion_all = (SELECT MAX(noreligion_all) FROM religion_by_nation);
76
77  /* What are the predominate religions in the countries that have a happiness score of 4
78  or lower? Order by score. */
79
80     SELECT h.country, h.happiness_score, rt.religion_type
81     FROM happiness_report AS h, religion_by_type AS rt
82     WHERE h.happiness_score <= 4
83     AND h.country = rt.country
84     AND h.year_of_data = 2017
85     AND rt.population =
86     (SELECT GREATEST(
87     christianity_protestant ,
88     christianity_romancatholic ,
89     christianity_easternorthodox ,
90     christianity_anglican ,
91     christianity_other ,
92     judaism_orthodox ,
93     judaism_conservative ,
94     judaism_reform ,
95     judaism_other ,
96     islam_sunni ,
97     islam_shi'a ,
98     islam_ibadhi ,
99     islam_alawite ,
100    islam_ahmadiyya ,
101    islam_other ,
102    buddhism_mahayana ,
103    buddhism_theravada ,
104    buddhism_other ,
105    zoroastrianism_all ,
106    hinduism_all ,
107    sikhism_all ,
108    shinto_all ,
109    baha'i_all ,
110    taoism_all ,
111    jainism_all ,
112    confucianism_all ,
113    syncretism_all ,
114    animism_all ,
115    noreligion_all
116    )
117    FROM religion_by_nation AS rn
118    WHERE rt.country = rn.country
119    AND year_of_data = (SELECT MAX(year_of_data) FROM religion_by_nation))
120    ORDER BY h.happiness_score DESC;
121
122  /* What are the largest contributions to the happiness score for all the countries
123  ranked in 2015? */
124
125     SELECT country, contribution_factor AS largest_contribution_factor
126     FROM happiness_report_by_factor AS h1
127     WHERE year_of_data = 2015 AND contribution =
128     (SELECT GREATEST(
129     GDP_PerCapita,
130     Family,
131     Life_Expectancy,
132     Freedom,

```

```

131     Gov_Corruption,
132     Generosity,
133     Dystopia_Residual
134 ) FROM happiness_report AS h2
135 WHERE h2.country = h1.country AND year_of_data = 2015);
136
137
138 /* How much does trust in the government do countries contribute to happiness in the
139 top 5 countries with
140 the most population the sunni sect of Islam have? How about protestant sect of
141 Christianity?
142 (using the most recent data)*/
143
144 /*Sunni*/
145 SELECT h.country AS Sunni_country, h.Gov_Corruption
146 FROM happiness_report AS h, religion_by_nation AS rn
147 WHERE h.country = rn.country
148 AND h.year_of_data = (SELECT MAX(year_of_data) FROM happiness_report)
149 AND rn.year_of_data = (SELECT MAX(year_of_data) FROM religion_by_nation)
150 ORDER BY rn.islam_sunni DESC
151 LIMIT 5;
152
153 /*Protestant*/
154 SELECT h.country AS protestant_country, h.Gov_Corruption
155 FROM happiness_report AS h, religion_by_nation AS rn
156 WHERE h.country = rn.country
157 AND h.year_of_data = (SELECT MAX(year_of_data) FROM happiness_report)
158 AND rn.year_of_data = (SELECT MAX(year_of_data) FROM religion_by_nation)
159 ORDER BY rn.christianity_protestant DESC
160 LIMIT 5;
161
162 /* How did the happiest country change from 2015 to 2016 to 2017? */
163
164 SELECT h.country AS Happiest_Country_2015, h1.country AS Happiest_Country_2016,
165 h2.country AS Happiest_Country_2017
166 FROM happiness_report AS h, happiness_report AS h1, happiness_report AS h2
167 WHERE h.year_of_data = 2015 AND h.happiness_rank = 1
168 AND h1.year_of_data = 2016 AND h1.happiness_rank = 1
169 AND h2.year_of_data = 2017 AND h2.happiness_rank = 1;
170
171 /* Of the countries that generosity contributed 0.4 or more to the overall score what
172 is the predominant religion
173 in those countries? Order by generosity contribution*/
174
175 SELECT h.country, rt.religion_type
176 FROM happiness_report AS h, religion_by_type AS rt
177 WHERE h.country = rt.country AND h.generosity >= 0.4
178 AND h.year_of_data = (SELECT MAX(year_of_data) FROM happiness_report)
179 AND rt.population =
180 (SELECT GREATEST(
181     christianity_protestant ,
182     christianity_romancatholic ,
183     christianity_easternorthodox ,
184     christianity_anglican ,
185     christianity_other ,
186     judaism_orthodox ,
187     judaism_conservative ,
188     judaism_reform ,
189     judaism_other ,
190     islam_sunni ,
191     islam_shi'a ,
192     islam_ibadhi ,
193     islam_alawite ,
194     islam_ahmadiyya ,
195     islam_other ,
196     buddhism_mahayana ,
197     buddhism_theravada ,
198     buddhism_other ,

```

```

196         zoroastrianism_all ,
197         hinduism_all ,
198         sikhism_all ,
199         shinto_all ,
200         baha'i_all ,
201         taoism_all ,
202         jainism_all ,
203         confucianism_all ,
204         syncretism_all ,
205         animism_all ,
206         noreligion_all
207     )
208     FROM religion_by_nation AS rn
209     WHERE rt.country = rn.country
210           AND year_of_data = (SELECT MAX(year_of_data) FROM religion_by_nation))
211     ORDER BY h.generosity DESC;
212
213
214     /* Of the range dates available in the regional data set for religion what how much has
the protestant religion of the United States
215     change in population from earliest date to most recent? */
216
217     SELECT rn1.christianity_protestant - rn2.christianity_protestant AS Population_Change
218     FROM religion_by_nation AS rn1, religion_by_nation AS rn2
219     WHERE rn1.code = 'USA' AND rn2.code = 'USA'
220           AND rn1.year_of_data = (SELECT MAX(year_of_data) FROM religion_by_nation)
221           AND rn2.year_of_data = (SELECT MIN(year_of_data) FROM religion_by_nation);
222
223
224

```