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 or 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.
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 counties 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 ‘coloremaps.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.

![Americas Happiness Ranks Visualization](image1)

**Figure 2: Americas Happiness Ranks Visualization**

![Europe, Middle East, and Asia Happiness Rank Visualization](image2)

**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.
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


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


[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.
CREATE TABLE happiness_report(
  Country VARCHAR(200),
  Region VARCHAR(50),
  Happiness_Rank INTEGER,
  Happiness_Score DOUBLE PRECISION,
  GDP_PerCapita DOUBLE PRECISION,
  Family DOUBLE PRECISION,
  LifeEXPECTancy DOUBLE PRECISION,
  Freedom DOUBLE PRECISION,
  Gov_Corruption DOUBLE PRECISION,
  Generosity DOUBLE PRECISION,
  Dystopia_Residual DOUBLE PRECISION,
  Year_of_Data INTEGER,
  PRIMARY KEY (Happiness_Rank, Year_of_Data)
);

CREATE TABLE religion_by_nation(
  Year_of_data INTEGER,
  Country VARCHAR(100),
  Code VARCHAR(10),
  christianity_protestant INTEGER,
  christianity_romancatholic INTEGER,
  christianity_easternorthodox INTEGER,
  christianity_anglican INTEGER,
  christianity_other INTEGER,
  judaism_orthodox INTEGER,
  judaism_conservative INTEGER,
  judaism_reform INTEGER,
  judaism_other INTEGER,
  islam_sunny INTEGER,
  islam_shia INTEGER,
  islam_ibadhi INTEGER,
  islam_alawite INTEGER,
  islam_ahmadiyya INTEGER,
  islam_other INTEGER,
  buddhism_mahayana INTEGER,
  buddhism_tearavada INTEGER,
  buddhism_other INTEGER,
  zoroastrianism_all INTEGER,
  hinduism_all INTEGER,
sikhism_all INTEGER,
shinto_all INTEGER,
baha’i_all INTEGER,
taoism_all INTEGER,
jainism_all INTEGER,
confucianism_all INTEGER,
syncretism_all INTEGER,
animism_all INTEGER,
noreligion_all INTEGER,
PRIMARY KEY (Year_of_Data, Country, Code)
);

/* Creating the world happiness report table */
/* Dropping country_location table and all dependencies if it exists */
DROP TABLE IF EXISTS country_location CASCADE;
/* Creating table with necessary attributes */
CREATE TABLE country_location(
   Country VARCHAR(200),
   Latitude DOUBLE PRECISION,
   Longitude DOUBLE PRECISION
);

/*-----------------------------*/
/* --- Creating better tables from the original dataset tables --- */
/*-----------------------------*/
DROP TABLE IF EXISTS happiness_report_by_factor CASCADE;
CREATE TABLE happiness_report_by_factor(
   country varchar(100),
   contribution_factor varchar(100),
   contribution DOUBLE PRECISION,
   year_of_data INTEGER
);

/* Creating the world religion_by_type table */
/* Dropping religion_by_type table and all dependencies if it exists */
DROP TABLE IF EXISTS religion_by_type CASCADE;
/* Creating table with necessary attributes */
CREATE TABLE religion_by_type(
   country varchar(100),
   religion_type varchar(200),
   population INTEGER,
   year_of_data INTEGER
);

/*-----------------------------*/
/* --- Populating Data into the flat unmodified tables using COPY --- */
/*-----------------------------*/
/* ---- Populating Happiness_Report ---- */
/* Copying in 2015 data */
\copy happiness_report FROM '2015.csv' DELIMITER ',' CSV HEADER;
/* Copying in 2016 data */
\copy happiness_report FROM '2016.csv' DELIMITER ',' CSV HEADER;
/* Copying in 2017 data */
\copy happiness_report FROM '2017.csv' DELIMITER ',' CSV HEADER;
/* ---- Populating religion_by_nation ---- */
\copy religion_by_nation FROM 'national.csv' DELIMITER ',' CSV HEADER;
/* ---- Populating country_location -----*/
\copy country_location FROM 'LocationCSVSScrubbed.csv' DELIMITER ',' CSV HEADER;

/*---------------------------------------------------------------------------------------
--------------------------------*/
/* ------------------------- Populating Data into the the better, more intelligent
 tables using queries-------------------*/
/*---------------------------------------------------------------------------------------
--------------------------------*/
/*---------------------------------------------------------------------------------------
-------------*/
/*-------------------------------Happiness_Report_By_Factor-------------------------------*/
/*---------------------------------------------------------------------------------------
-------------*/
/* ---- Populating and Updating Happiness_Report_by_factor for GDP_PerCapita ---- */
INSERT INTO happiness_report_by_factor (country, contribution, year_of_data)
SELECT hr.country, hr.GDP_PerCapita, hr.year_of_data
FROM happiness_report AS hr;

UPDATE happiness_report_by_factor
SET contribution_factor = 'GDP_PerCapita';

/* ---- Populating and Updating Happiness_Report_by_factor for family ---- */
INSERT INTO happiness_report_by_factor (country, contribution, year_of_data)
SELECT hr.country, hr.Family, hr.year_of_data
FROM happiness_report AS hr;

UPDATE happiness_report_by_factor
SET contribution_factor = 'Family'
WHERE contribution_factor IS NULL;

/* ---- Populating and Updating Happiness_Report_by_factor for Life_Expectancy ---- */
INSERT INTO happiness_report_by_factor (country, contribution, year_of_data)
SELECT hr.country, hr.Life_Expectancy, hr.year_of_data
FROM happiness_report AS hr;

UPDATE happiness_report_by_factor
SET contribution_factor = 'Life_Expectancy'
WHERE contribution_factor IS NULL;

/* ---- Populating and Updating Happiness_Report_by_factor for Freedom ---- */
INSERT INTO happiness_report_by_factor (country, contribution, year_of_data)
SELECT hr.country, hr.Freedom, hr.year_of_data
FROM happiness_report AS hr;

UPDATE happiness_report_by_factor
SET contribution_factor = 'Freedom'
WHERE contribution_factor IS NULL;

/* ---- Populating and Updating Happiness_Report_by_factor for Gov_Corruption ---- */
INSERT INTO happiness_report_by_factor (country, contribution, year_of_data)
SELECT hr.country, hr.Gov_Corruption, hr.year_of_data
FROM happiness_report AS hr;

UPDATE happiness_report_by_factor
SET contribution_factor = 'Gov_Corruption'
WHERE contribution_factor IS NULL;

/* ---- Populating and Updating Happiness_Report_by_factor for Generosity ---- */

INSERT INTO happiness_report_by_factor(country, contribution, year_of_data)
SELECT hr.country, hr.Generosity, hr.year_of_data
FROM happiness_report AS hr;

UPDATE happiness_report_by_factor
SET contribution_factor = 'Generosity'
WHERE contribution_factor IS NULL;

/* ---- Populating and Updating Happiness_Report_by_factor for Dystopia_Residual ---- */

INSERT INTO happiness_report_by_factor(country, contribution, year_of_data)
SELECT hr.country, hr.Dystopia_Residual, hr.year_of_data
FROM happiness_report AS hr;

UPDATE happiness_report_by_factor
SET contribution_factor = 'Dystopia_Residual'
WHERE contribution_factor IS NULL;

/*---------------------------------------------------------------------------------------
-------------
/*-----------------------------------Religion_by_Type------------------------------------
-------------
/*---------------------------------------------------------------------------------------
-------------
/*christianity_protestant*/

INSERT INTO religion_by_type(country, population, year_of_data)
SELECT rn.country, rn.christianity_protestant, rn.year_of_data
FROM religion_by_nation AS rn;

UPDATE religion_by_type
SET religion_type = 'christianity_protestant'
WHERE religion_type IS NULL;

/*christianity_romancatholic*/

INSERT INTO religion_by_type(country, population, year_of_data)
SELECT rn.country, rn.christianity_romancatholic, rn.year_of_data
FROM religion_by_nation AS rn;

UPDATE religion_by_type
SET religion_type = 'christianity_romancatholic'
WHERE religion_type IS NULL;

/*christianity_easternorthodox*/

INSERT INTO religion_by_type(country, population, year_of_data)
SELECT rn.country, rn.christianity_easternorthodox, rn.year_of_data
FROM religion_by_nation AS rn;

UPDATE religion_by_type
SET religion_type = 'christianity_easternorthodox'
WHERE religion_type IS NULL;

/*christianity_anglican*/

INSERT INTO religion_by_type(country, population, year_of_data)
SELECT rn.country, rn.christianity_anglican, rn.year_of_data
FROM religion_by_nation AS rn;

UPDATE religion_by_type
SET religion_type = 'christianity_anglican'
WHERE religion_type IS NULL;

/*christianity_other*/

INSERT INTO religion_by_type(country, population, year_of_data)
SELECT rn.country, rn.christianity_other, rn.year_of_data
FROM religion_by_nation AS rn;

UPDATE religion_by_type
SET religion_type = 'christianity_other'
WHERE religion_type IS NULL;

/*judaism_orthodox*/

INSERT INTO religion_by_type(country, population, year_of_data)
SELECT rn.country, rn.judaism_orthodox, rn.year_of_data
FROM religion_by_nation AS rn;

UPDATE religion_by_type
SET religion_type = 'judaism_orthodox'
WHERE religion_type IS NULL;

/*---here*/
SELECT rn.country, rn.judaism_orthodox, rn.year_of_data
FROM religion_by_nation AS rn;
UPDATE religion_by_type
SET religion_type = 'judaism_orthodox'
WHERE religion_type IS NULL;
/*judaism_conservative*/
INSERT INTO religion_by_type(country, population, year_of_data)
SELECT rn.country, rn.judaism_conservative, rn.year_of_data
FROM religion_by_nation AS rn;
UPDATE religion_by_type
SET religion_type = 'judaism_conservative'
WHERE religion_type IS NULL;
/*judaism_reform*/
INSERT INTO religion_by_type(country, population, year_of_data)
SELECT rn.country, rn.judaism_reform, rn.year_of_data
FROM religion_by_nation AS rn;
UPDATE religion_by_type
SET religion_type = 'judaism_reform'
WHERE religion_type IS NULL;
/*judaism_other*/
INSERT INTO religion_by_type(country, population, year_of_data)
SELECT rn.country, rn.judaism_other, rn.year_of_data
FROM religion_by_nation AS rn;
UPDATE religion_by_type
SET religion_type = 'judaism_other'
WHERE religion_type IS NULL;
/*islam_sunnii*/
INSERT INTO religion_by_type(country, population, year_of_data)
SELECT rn.country, rn.islam_sunnii, rn.year_of_data
FROM religion_by_nation AS rn;
UPDATE religion_by_type
SET religion_type = 'islam_sunnii'
WHERE religion_type IS NULL;
/*islam_shi’a*/
INSERT INTO religion_by_type(country, population, year_of_data)
SELECT rn.country, rn.islam_shi’a, rn.year_of_data
FROM religion_by_nation AS rn;
UPDATE religion_by_type
SET religion_type = 'islam_shi’a'
WHERE religion_type IS NULL;
/*islam_ibadhi*/
INSERT INTO religion_by_type(country, population, year_of_data)
SELECT rn.country, rn.islam_ibadhi, rn.year_of_data
FROM religion_by_nation AS rn;
UPDATE religion_by_type
SET religion_type = 'islam_ibadhi'
WHERE religion_type IS NULL;
/*islam_alawite*/
INSERT INTO religion_by_type(country, population, year_of_data)
SELECT rn.country, rn.islam_alawite, rn.year_of_data
FROM religion_by_nation AS rn;
UPDATE religion_by_type
SET religion_type = 'islam_alawite'
WHERE religion_type IS NULL;
/*islam_ahmadiyya*/
INSERT INTO religion_by_type(country, population, year_of_data)
SELECT rn.country, rn.islam_ahmadiyya, rn.year_of_data
FROM religion_by_nation AS rn;
UPDATE religion_by_type
SET religion_type = 'islam_ahmadiyya'
WHERE religion_type IS NULL;
/*islam_other*/
INSERT INTO religion_by_type(country, population, year_of_data)
SELECT rn.country, rn.islam_other, rn.year_of_data
FROM religion_by_nation AS rn;
UPDATE religion_by_type
SET religion_type = 'islam_other'
WHERE religion_type IS NULL;
/*buddhism_mahayana*/
INSERT INTO religion_by_type(country, population, year_of_data)
SELECT rn.country, rn.buddhism_mahayana, rn.year_of_data
FROM religion_by_nation AS rn;
UPDATE religion_by_type
SET religion_type = 'buddhism_mahayana'
WHERE religion_type IS NULL;
/*buddhism_mahayana*/
INSERT INTO religion_by_type(country, population, year_of_data)
SELECT rn.country, rn.buddhism_theravada, rn.year_of_data
FROM religion_by_nation AS rn;
UPDATE religion_by_type
SET religion_type = 'buddhism_theravada'
WHERE religion_type IS NULL;
/*buddhism_theravada*/
INSERT INTO religion_by_type(country, population, year_of_data)
SELECT rn.country, rn.buddhism_other, rn.year_of_data
FROM religion_by_nation AS rn;
UPDATE religion_by_type
SET religion_type = 'buddhism_other'
WHERE religion_type IS NULL;
/*buddhism_other*/
INSERT INTO religion_by_type(country, population, year_of_data)
SELECT rn.country, rn.zoroastrianism_all, rn.year_of_data
FROM religion_by_nation AS rn;
UPDATE religion_by_type
SET religion_type = 'zoroastrianism_all'
WHERE religion_type IS NULL;
/*zoroastrianism_all*/
INSERT INTO religion_by_type(country, population, year_of_data)
SELECT rn.country, rn.hinduism_all, rn.year_of_data
FROM religion_by_nation AS rn;
UPDATE religion_by_type
SET religion_type = 'hinduism_all'
WHERE religion_type IS NULL;
/*hinduism_all*/
INSERT INTO religion_by_type(country, population, year_of_data)
SELECT rn.country, rn.sikhism_all, rn.year_of_data
FROM religion_by_nation AS rn;
UPDATE religion_by_type
SET religion_type = 'sikhism_all'
WHERE religion_type IS NULL;
/*sikhism_all*/
INSERT INTO religion_by_type(country, population, year_of_data)
SELECT rn.country, rn.shinto_all, rn.year_of_data
FROM religion_by_nation AS rn;
UPDATE religion_by_type
SET religion_type = 'shinto_all'
WHERE religion_type IS NULL;
/*shinto_all*/
INSERT INTO religion_by_type(country, population, year_of_data)
SELECT rn.country, rn.baha’i_all, rn.year_of_data
FROM religion_by_nation AS rn;
UPDATE religion_by_type
SET religion_type = 'baha’i_all'
WHERE religion_type IS NULL;
/*baha’i_all*/
INSERT INTO religion_by_type(country, population, year_of_data)
SELECT rn.country, rn.taoism_all, rn.year_of_data
FROM religion_by_nation AS rn;
UPDATE religion_by_type
SET religion_type = 'taoism_all'
WHERE religion_type IS NULL;
/*taoism_all*/
INSERT INTO religion_by_type(country, population, year_of_data)
SELECT rn.country, rn.jainism_all, rn.year_of_data
FROM religion_by_nation AS rn;
UPDATE religion_by_type
SET religion_type = 'jainism_all'
WHERE religion_type IS NULL;
/*jainism_all*/
INSERT INTO religion_by_type(country, population, year_of_data)
SELECT rn.country, rn.baha’i_all, rn.year_of_data
FROM religion_by_nation AS rn;
UPDATE religion_by_type
SET religion_type = 'baha’i_all'
WHERE religion_type IS NULL;
/*baha’i_all*/
INSERT INTO religion_by_type(country, population, year_of_data)
SELECT rn.country, rn.taoism_all, rn.year_of_data
FROM religion_by_nation AS rn;
UPDATE religion_by_type
SET religion_type = 'taoism_all'
WHERE religion_type IS NULL;
/*taoism_all*/
INSERT INTO religion_by_type(country, population, year_of_data)
SELECT rn.country, rn.jainism_all, rn.year_of_data
FROM religion_by_nation AS rn;
UPDATE religion_by_type
SET religion_type = 'jainism_all'
WHERE religion_type IS NULL;
/*jainism_all*/
INSERT INTO religion_by_type(country, population, year_of_data)
SELECT rn.country, rn.baha’i_all, rn.year_of_data
FROM religion_by_nation AS rn;
UPDATE religion_by_type
SET religion_type = 'baha’i_all'
WHERE religion_type IS NULL;
/*baha’i_all*/
INSERT INTO religion_by_type(country, population, year_of_data)
SELECT rn.country, rn.taoism_all, rn.year_of_data
FROM religion_by_nation AS rn;
UPDATE religion_by_type
SET religion_type = 'taoism_all'
WHERE religion_type IS NULL;
/*taoism_all*/
INSERT INTO religion_by_type(country, population, year_of_data)
SELECT rn.country, rn.jainism_all, rn.year_of_data
FROM religion_by_nation AS rn;
UPDATE religion_by_type
SET religion_type = 'jainism_all'
WHERE religion_type IS NULL;
/*jainism_all*/
/*confucianism_all*/
INSERT INTO religion_by_type(country, population, year_of_data)
SELECT rn.country, rn.confucianism_all, rn.year_of_data
FROM religion_by_nation AS rn;
UPDATE religion_by_type
SET religion_type = 'confucianism_all'
WHERE religion_type IS NULL;
/*syncretism_all*/
INSERT INTO religion_by_type(country, population, year_of_data)
SELECT rn.country, rn.syncretism_all, rn.year_of_data
FROM religion_by_nation AS rn;
UPDATE religion_by_type
SET religion_type = 'syncretism_all'
WHERE religion_type IS NULL;
/*animism_all*/
INSERT INTO religion_by_type(country, population, year_of_data)
SELECT rn.country, rn.animism_all, rn.year_of_data
FROM religion_by_nation AS rn;
UPDATE religion_by_type
SET religion_type = 'animism_all'
WHERE religion_type IS NULL;
/*noreligion_all*/
INSERT INTO religion_by_type(country, population, year_of_data)
SELECT rn.country, rn.noreligion_all, rn.year_of_data
FROM religion_by_nation AS rn;
UPDATE religion_by_type
SET religion_type = 'noreligion_all'
WHERE religion_type IS NULL;
/* Issues/bugs list that I ran into and had to resolve */
/*
- Datasets had alot of redundant data. SOLVED: Post processing in the CSV to eliminate redundancies
- Datasets had many 'inferred' attributes that could be calculated by the more significant attributes. SOLVED: By taking out the attributes when populating database to keep the tables in more normal form
- Datasets had many errors and duplicated information making it difficult to put table contrainsts such as a primary key. SOLVED: Pre-processing the data 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 in identifying duplicates that I missed.
- Relating the datasets for query mash-ups. Becuase the two datasets that I pulled together were so different it was easy to identify interesting queries
  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:
  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 schema and data itself to better suit the questions I wanted to ask about my two datasets.
- 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 for intelligent/useful results.
- 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 based. e.g. COPY tablename TO file.csv WITH (FORMAT csv);
*/
/*
  * project8_queryMashUp.sql
  * author: Ryan Fast
  */

/*---------------------------------------------------------*/
/* ----------Interesting Query Mash Ups! ------------------*/
/*---------------------------------------------------------*/
/* The following queries provide very interesting results with the two datasets
ingested into
my schema. Basic Descriptions/Questions to answer are provided for the expected
results of the query */
/*---------------------------------------------------------*/
/*Where does the United States rank on the happiness scale in 2017 and what is the
predominant religion using the most
recent data available? */

SELECT h.country, happiness_rank, religion_type
FROM happiness_report AS h, religion_by_type AS rt
WHERE h.country = 'United States' AND rt.country = 'United States of America' AND
h.year_of_data = 2017
AND rt.population = (SELECT GREATEST(
  christianity_protestant ,
  christianity_roman_catholic ,
  christianity_easternorthodox ,
  christianity_english ,
  christianity_orthodox ,
  jainism_orthodox ,
  judaism_orthodox ,
  judaism_conservative ,
  judaism_reform ,
  judaism_other ,
  islam_sunn ,
  islam_shi'a ,
  islam_ibadhi ,
  islam_alawite ,
  islam_ahmadiyya ,
  islam_other ,
  buddhism_mahayana ,
  buddhism_theravada ,
  buddhism_other ,
  zoroastrianism_all ,
  hinduism_all ,
  sikhism_all ,
  shinto_all ,
  baha'i_all ,
  taoism_all ,
  confucianism_all ,
  syncretism_all ,
  animism_all ,
  noreligion_all )
FROM religion_by_nation
WHERE code = 'USA'
AND year_of_data = (SELECT MAX(year_of_data) FROM religion_by_nation));

/*What country in 2017 was freedom the largest contributor out of all other countries
and where does
that country rank? */

SELECT country, happiness_rank
FROM happiness_report
WHERE year_of_data = 2017 AND freedom = (
SELECT MAX(freedom) FROM happiness_report
WHERE year_of_data = 2017;

/* What are the happiness score and rank of country with the most non-religious populations in 2017? */

SELECT h.country, happiness_rank, happiness_score
FROM happiness_report AS h, religion_by_nation AS rn
WHERE h.country = rn.country AND h.year_of_data = 2017
AND rn.noreligion_all = (SELECT MAX(noreligion_all) FROM religion_by_nation);

/* What are the predominate religions in the countries that have a happiness score of 4 or lower? Order by score. */

SELECT h.country, h.happiness_score, rt.religion_type
FROM happiness_report AS h, religion_by_type AS rt
WHERE h.happiness_score <= 4
AND h.country = rt.country
AND h.year_of_data = 2017
AND rt.population =
(SELECT GREATEST(
christianity_protestant, 
christianity_romancatholic, 
christianity_easternorthodox, 
christianity_anglican, 
christianity_other, 
judaism_orthodox, 
judaism_conservative, 
judaism_reform, 
judaism_other, 
islam_sunny, 
islam_shi'a, 
islam_ibadhi, 
islam_alawite, 
islam_ahmadiyya, 
islam_other, 
buddhism_mahayana, 
buddhism_theravada, 
buddhism_other, 
zoroastrianism_all, 
hinduism_all, 
sikhism_all, 
shinto_all, 
baha'i_all, 
taoism_all, 
jainism_all, 
confucianism_all, 
syncretism_all, 
animism_all, 
noreligion_all
)
FROM religion_by_nation AS rn
WHERE rt.country = rn.country
AND year_of_data = (SELECT MAX(year_of_data) FROM religion_by_nation))
ORDER BY h.happiness_score DESC;

/* What are the largest contributions to the happiness score for all the countries ranked in 2015? */

SELECT country, contribution_factor AS largest_contribution_factor
FROM happiness_report_by_factor AS h1
WHERE year_of_data = 2015 AND contribution =
(SELECT GREATEST(
GDP_PerCapita, 
Family, 
Life_Expectancy, 
Freedom,
Gov_Corruption,
Generosity,
Dystopia_Residual
) FROM happiness_report AS h2
WHERE h2.country = h1.country AND year_of_data = 2015);

/* How much does trust in the government do countries contribute to happiness in the
the top 5 counties with
the most population the sunni sect of Islam have? How about protestant sect of
Christianity?
(using the most recent data)*/

/*Sunni*/
SELECT h.country AS Sunni_country, h.Gov_Corruption
FROM happiness_report AS h, religion_by_nation AS rn
WHERE h.country = rn.country AND h.year_of_data = (SELECT MAX(year_of_data) FROM happiness_report)
AND rn.year_of_data = (SELECT MAX(year_of_data) FROM religion_by_nation)
ORDER BY rn.islam_sunni DESC LIMIT 5;

/*Protestant*/
SELECT h.country AS protestant_country, h.Gov_Corruption
FROM happiness_report AS h, religion_by_nation AS rn
WHERE h.country = rn.country AND h.year_of_data = (SELECT MAX(year_of_data) FROM happiness_report)
AND rn.year_of_data = (SELECT MAX(year_of_data) FROM religion_by_nation)
ORDER BY rn.christianity_protestant DESC LIMIT 5;

/* How did the happiest country change from 2015 to 2016 to 2017? */

SELECT h.country AS Happiest_Country_2015, h1.country AS Happiest_Country_2016, h2.country AS Happiest_Country_2017
FROM happiness_report AS h, happiness_report AS h1, happiness_report AS h2
WHERE h.year_of_data = 2015 AND h.happiness_rank = 1
AND h1.year_of_data = 2016 AND h1.happiness_rank = 1
AND h2.year_of_data = 2017 AND h2.happiness_rank = 1;

/* 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*/

SELECT h.country, rt.religion_type
FROM happiness_report AS h, religion_by_type AS rt
WHERE h.country = rt.country AND h.generosity >= 0.4
AND h.year_of_data = (SELECT MAX(year_of_data) FROM happiness_report)
AND rt.population = (SELECT GREATEST(christianity_protestant, christianity_romancatholic, christianity_easternorthodox, christianity_anglican, christianity_other, judaism_orthodox, judaism_conservative, judaism_reform, judaism_other, islam_sunni, islam_shi’a, islam_ibadhi, islam_alawite, islam_ahmadiyya, islam_other, buddhism_mahayana, buddhism_theravada, buddhism_other, ...
/* 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? */

SELECT rn1.christianity_protestant - rn2.christianity_protestant AS Population_Change
FROM religion_by_nation AS rn1, religion_by_nation AS rn2
WHERE rn1.code = 'USA' AND rn2.code = 'USA'
AND rn1.year_of_data = (SELECT MAX(year_of_data) FROM religion_by_nation)
AND rn2.year_of_data = (SELECT MIN(year_of_data) FROM religion_by_nation);