* **Data Sources**
  + *World Happiness Report*
    - The happiness scores and rankings from this report use data from the Gallup World Poll. While the Happiness Report was published between 2012 and 2017, we will be looking at the data from 2016.
    - <https://www.kaggle.com/unsdsn/world-happiness?select=2016.csv>
  + *World Bank World Development Indicators (WDI) – Health Systems*
    - The WDI data is compiled by the World Bank from a variety of international resources. It covers topics like Agriculture and Food Security, Public Sector Management, Education, and 15 others. For the purposes of this project we will be looking at the Health Systems data for 2016.
    - <https://www.kaggle.com/danevans/world-bank-wdi-212-health-systems/data>
* **Data Cleanup & Analysis**
  + By selecting these two datasets we hope to look at how health expenditures per country relate to happiness scores and ranking.
  + To that end, we’ll be joining our two datasets using the countries available, making sure to keep only those present in the World Happiness Report.
  + We’ve decided to drop territories, which are included, in the World Bank dataset. Although they’re included in the values, there is no actual data for them.
  + We’ve dropped Somalia, due to missing values in the Happiness Report, as well as no values in the World Bank dataset.
  + We used postgresql, a relational database, to hold our data and Pandas to carry out our work. This will have one database with two tables, one for each dataset.
* **ETL** 
  + **Extract**
    - Extracting our data sets were coming from Kaggle as csvs. Importing both csvs into a jupyter notebook allowed us to view the data in whole and start the cleaning process.
  + **Transform**
    - Data cleaning was required for both data sets. In the World Happiness data seet, all data was pretty straight forward. We wanted to extract the columns: Country, Happiness Rank, Happiness Score, and Economy (GDP per Capita) for all countries. In order for the column titles to be compatible with PgAdmin, we needed to rename the column names. Moving on the searching for missing values, we noticed that just for one country, Somalia, there was no integer in the Economy column. With that, we decided to drop Somalia as the clean up for this data set.
    - In the World Bank data set, there was much more cleaning needed. We started by extracting and renaming the columns: Country Region, Province State, World Bank name, Health exp pct, Health exp public pct, health exp out of pocket pct. Next we needed to find the missing values in the data set. There were missing values in the Country column that were dropped. The Province column was dropped. Lastly, four countries: Libya, Liechtenstein, Somalia, Syria were dropped from the data set due to not having values in multiple columns, current health, domestic government health expenditure, out of pocket expenditure. This concluded out clean up
  + **Load**
    - To load our data, we decided to use PgAdmin. It was our decision to load the tables into PgAdmin then merge them there. We decided to load into PgAdmin because we were combining our data by the country column. This would allow us to sort the data by country alphabetical. Another option to sort the data is by Happiness rank. In conclusion, once we merged the two data sets it not only gave us plenty ideas on what we could add later or if more time was given but also build a concrete analysis on both data sets together.