

World Happiness Report

Jerry Chang Dr. Allison Theobold, Data 301 Winter 2025
California Polytechnic State University, San Luis Obispo

Introduction / Background / Goals

Primary Question: Do certain features contribute more greatly to higher happiness scores or rankings?

Secondary Question: Does the surveyed rating of a predictor have effects on other predictors?
Within this data, we could also derive predicted happiness scores of hypothetical countries, depending on a collection of features.

This data [1] was collected by the Gallup World Poll from over 150 countries, using representative surveys of about 1,000 respondents per country. Surveys are done via face-to-face interviews or telephone calls, depending on the country's infrastructure [2]

Methods

World Happiness Report:

- Dataset pulled from Kaggle.
- The dataset contains 5 years of records, spanning 2015-2019 in a CSV for each respective year.
- Each observation in the dataset is a Country.

Analysis:

- A mix of numerical values and categorical values.
- Response Variables:** Happiness Rank describes the ranking of Happiness of a country that year, and Happiness Score describes the calculated happiness index for that country in that year.
- Predictors:** 'Country', 'Economy (GDP per Capita)', 'Social support', 'Health (Life Expectancy)', 'Freedom', 'Trust (Government Corruption)', 'Generosity', 'year'.
- Data Cleaning:** The dataset was cleaned by renaming inconsistent column names, removing unnecessary columns, and adding a 'Year' column to track records over time.
- The five datasets were then merged using an OUTER join, and missing values were filled with 0 to ensure consistency for analysis and modeling.
- Model Tuning:** Scale with Mean and without Standard Deviation, use Euclidean Distance, use K= 9 Neighbors in KNN, and weight by Distance.

Model

We have chosen to train a **KNN-Regression model** for this data. Based on the combination of features, we predict a numerical score on a continuous scale. Using GridSearchCV, we were able to tune hyperparameters to find an ideal KNN model for the dataset for an **RMSE of 0.524**.

Using the tuned model, we may predict happiness for hypothetical weightings and regions to better understand how increasing the focus on a feature would perform for overall happiness. For example, when decreasing the weighting for Generosity and shifting that sum to increase the weighting for Social Support, we found that predicted Happiness increased for the hypothetical case, keeping all other predictors stable.

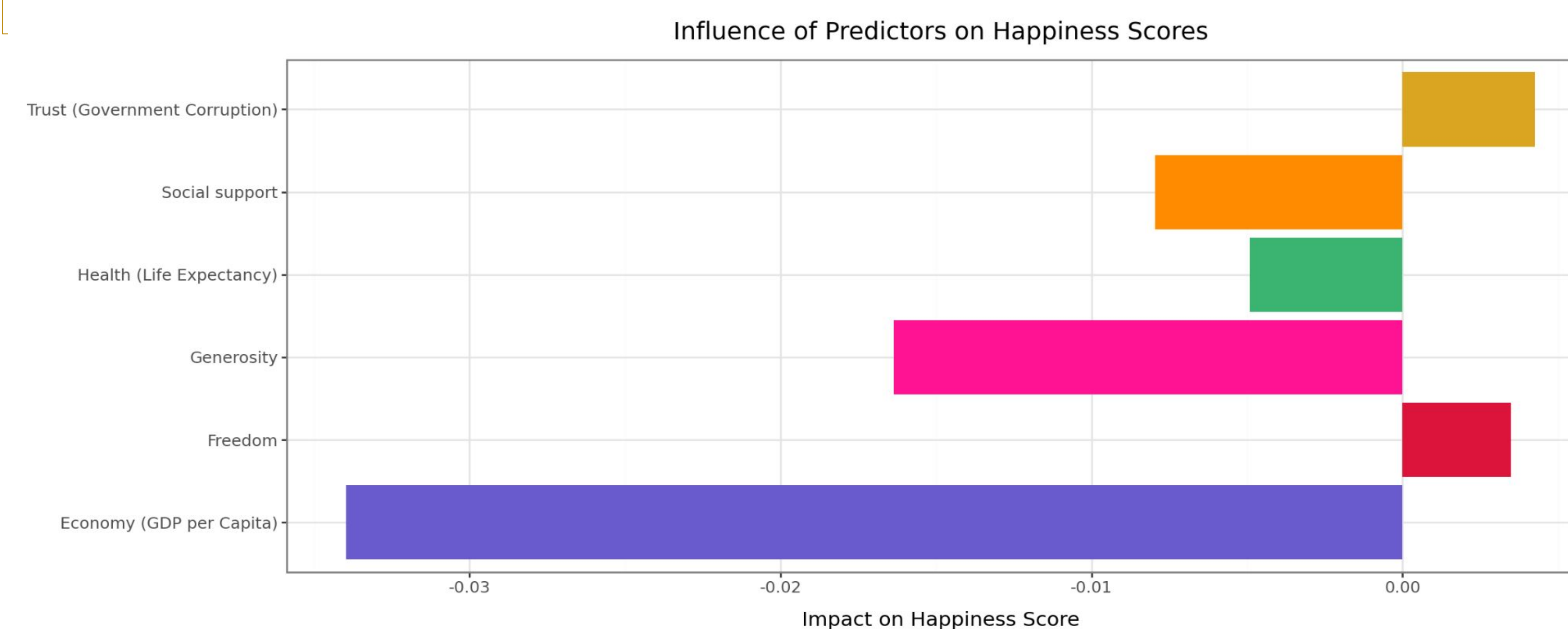


Figure 1. Looking at the changes in Happiness Scores after adjusting each predictor by a 0.1 increase, we may see that people's perception of the Economy is the most influential predictor for Happiness Scores of a country. For a 0.1 increase in weighting for Freedom, there is an associated 0.0339 decrease in Happiness Score, while holding all other variables constant.

Results & Implications

- From our KNN regression model with K = 9, Euclidean Distance, and Distance weighting, we have found that emphasis on the **Economy is the most influential predictor on a country's overall happiness**.
 - Other influential predictors: Generosity, Social Support.
 - Economy, Generosity, and Social Support have a negative relationship with Happiness Score, which may potentially be explained by negative sentiment that results in higher weightings for those features and poor effects on a people's Happiness.
- From calculating correlations **between Predictors**, we find **some strong correlations**
 - A strong, positive relationship between the importance of Health and the Economy when deciding on Happiness in a country.
 - This reflects an underlying effect predictors have on each other in rating Happiness.
- This model could potentially be used to **identify social programs or legislation** that governmental bodies could prioritize to lift morale and address the grievances of their constituents, or to **place a country amongst rankings on happiness**.
- A limitation is that the change in what citizens of a country hold to be important could shift dramatically between years, at a rate faster than what annually collected data reflects and much faster than the action of legislation.

Ethics & Limitations

Happiness and contentment are essential for our health as individuals and as a collective. Without happiness of some level, our health would suffer and would reflect in our communities. That being said, different **cultures perceive happiness differently** between self and society, community, or even within. This dataset does not capture the fine grained reasons for happiness in different countries, and thus we may only interpret happiness in terms of the available predictors.

Historical events greatly affect the societal stratification of a people, the wealth amongst the people, and perception on a global scale. When we view predictors such as GDP and Government Corruption, we must understand that people make do within their setting to find their happiness.

Institutions in a country affect the educational, racial, political, monetary, etc. opportunities respondents have, and in turn, would affect goals necessary for happiness. This skews the data on an individual level and if the respondents are not varied, this would be reflected for the perceived happiness of the overall country and in the weighting in an observation.

Furthermore, these ratings are done based on **feeling rather than hard, repeatable measurements** and are thus subject to human bias. The types of respondents are those technologically reachable, and this could skew the scores and importance since it is a voluntary survey. With all these factors in mind, our findings should be **taken as reference or supplemental information**, rather than an absolute guide for interested parties.

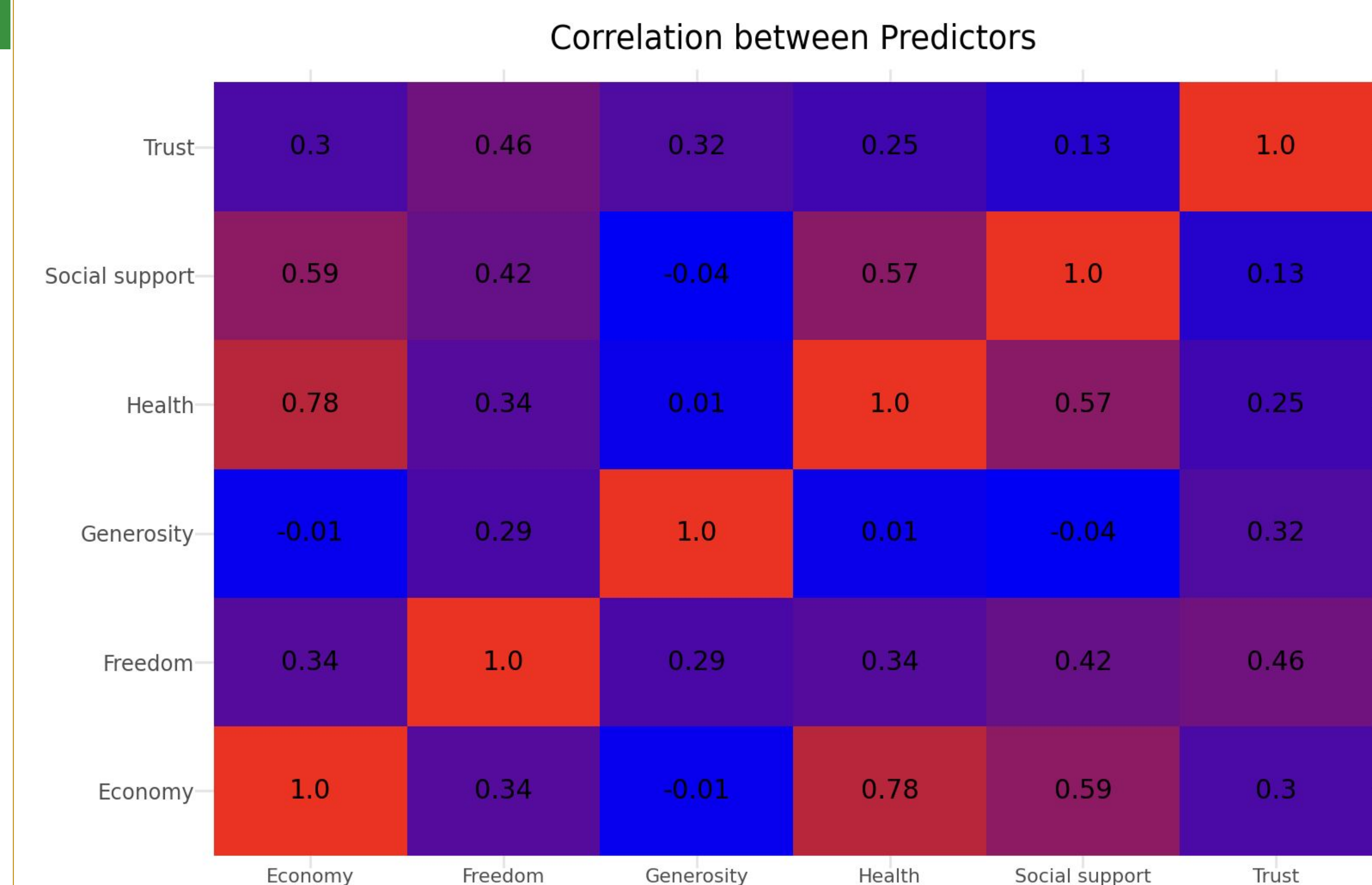


Figure 2. To answer our secondary question, we calculate the correlation between all the different predictors across the years. There is a strong, positive relationship between the importance of Health and the Economy when rating Happiness in a country. There is a moderately strong, positive relationship between the importance of Social Support and both Economy and Health.

References

- [1] World Happiness Report (2019) [www.kaggle.com](https://www.kaggle.com/datasets/unsdsn/world-happiness). Available at: <https://www.kaggle.com/datasets/unsdsn/world-happiness>.
[2] Inc, G. (2014) How Does the Gallup World Poll Work?, Gallup.com. Available at: <https://www.gallup.com/178667/gallup-world-poll-work.aspx>.

Acknowledgements

We would like to extend a special thanks to Dr. Theobold for her instruction and guidance throughout the quarter. In addition, we would also like to express gratitude for Cal Poly SLO for the resources provided for learning in DATA 301.