

# Chapter 1

## Introduction

### 1.1 Research Question

We are interested to know if socio-economic factors can be consolidated into a model that can forecast the actual 2016 election results.

We have two datasets, one from the presidential elections and one from the US Census Bureau (2014) and one from the MIT Election Data and Science Lab, 2018

We don't want to assess our model on the same dataset since otherwise we would have "optimism bias". It is important that our model is able to predict the target in unseen data, in other words, we want our model to be able to generalize. (pending to rewrite and find references). That said, we need to assess our model in different data, which we will call "honest assessment". In order to accomplish this we need to split data. WRITE ABOUT DATA SPLITTING into training and test set. We could split the data using a simple random sampling approach (SRS). SRS is a sampling technique where each member of the population is equally likely to be selected in the sample. By performing SRS we risk our samples to have a different distribution of our target outcome (in our case the binary response). In order to overcome this difficulty, we make use of stratified sampling. Stratified sampling is another sampling technique where the population can be divided into smaller groups or strata, where each of them represents defined characteristics of

---

the population (Gelman, Katz, & Tuerlinckx, 2002) (Thompson, 2012). Stratified sampling is preferred over SRS as the former will produce a more accurate representation of the entire population (Meng, 2013).

# References

- Gelman, A., Katz, J. N., & Tuerlinckx, F. (2002). The mathematics and statistics of voting power. *Statistical Science*, 17(4), 420–435. doi:10.1214/ss/1049993201
- Meng, X. (2013). *Scalable Simple Random Sampling and Stratified Sampling*.
- MIT Election Data and Science Lab. (2018). County presidential election returns 2000-2016. Version V2. doi:10.7910/DVN/VOQCHQ
- Thompson, S. K. (2012). *Sampling* (3rd). Wiley.