STA304-final-report

Brian Diep

26/11/2020

Abstract

General statement about topic. Important because TBD. We find TBD. Our findings imply that French Canadians still deal with many socioeconomic challenges as compared to English Canadians

Introduction

Many second language (L2) language learners. Although traditionally, we e. Although a key part of the electoral tradition is placed on the significance of each individual's vote, we can still find that almost all voters fall into broad categories that constitute voting blocs. These blocs can be united from many disparate categories such as age, race, and state of residence and many voters will identify with multiple blocs simultaneously. Although this is not an easy task, an analysis of these blocs can allow pollsters to predict the outcome of the federal election. For our purposes we use publicly available data from the Duolingo and census data from the 2018 American Community Survey (ACS) to use a multilevel regression with poststratification.

In order to create a Bayesian multilevel logistic model, variables associated with voter preference were extracted from the Nationscape study. In particular, age, state, income, race, gender, Hispanic identity and employment were decided upon as predictors since exploratory analysis of the data showed a significant relationship (p-value less than 0.05 via logistic regression models) between these variables and voting intentions. Poststratification was performed using the ACS data to incorporate the population distributions of the U.S. population. Our findings suggest that there is merit to an approach to second language acquisition that uses half life regression.

will continue to further explore the data sets used in our modelling, summarizing features of the data and visualizing variables of interest to our model. Next the features of our model and post-stratification process are described prior to simulating the election on the Electoral College. Finally, the prediction of Joe Biden's win will be investigated further and a discussion follows into strengths and weaknesses of the model, why this information is useful and further avenues of study.

```
unique(data$ui_language)
```

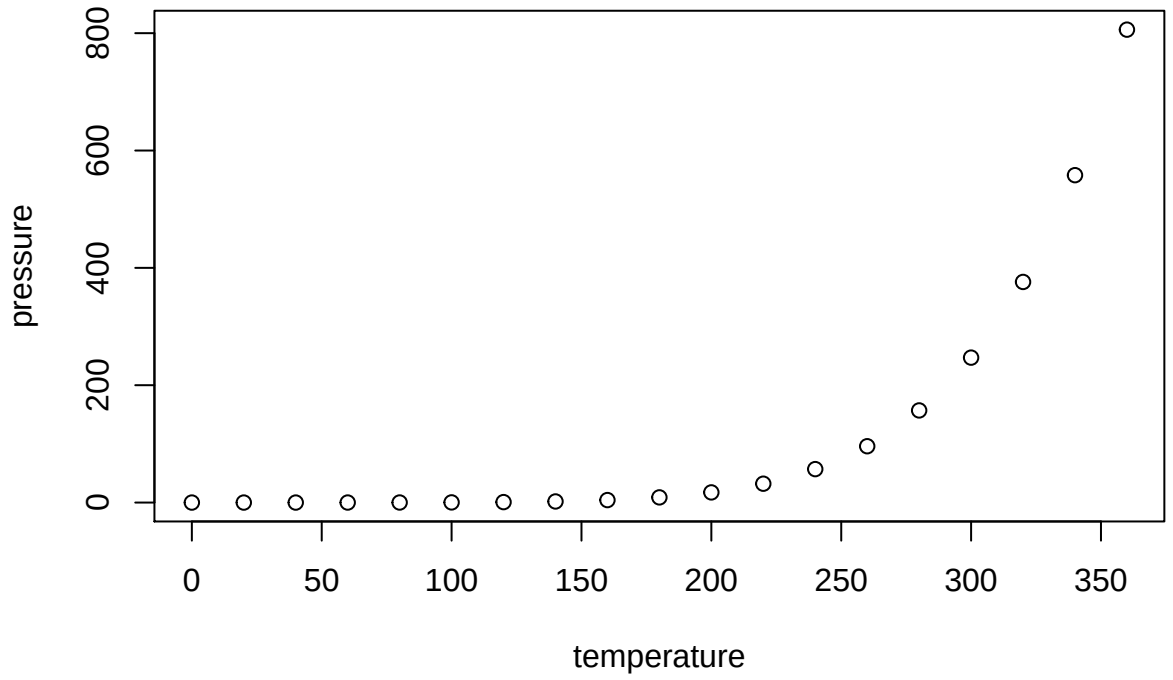
```
## [1] "en" "pt" "es" "it"
```

```
unique(data$learning_language)
```

```
## [1] "de" "es" "pt" "en" "fr" "it"
```

Including Plots

You can also embed plots, for example:



Note that the `echo = FALSE` parameter was added to the code chunk to prevent printing of the R code that generated the plot.