Final Project Report

Raleigh Goodwin

```
## here() starts at /Users/raleighgoodwin/Dropbox (University of Oregon)/UO Dropbox/DROPBOX Desktop/654-
```

```
## Warning: package 'tidyverse' was built under R version 4.1.2
```

```
## Warning: package 'tibble' was built under R version 4.1.2
```

```
## Warning: package 'tidyr' was built under R version 4.1.2
```

```
## Warning: package 'readr' was built under R version 4.1.2
```

```
## Warning: package 'purrr' was built under R version 4.1.2
```

```
## Warning: package 'dplyr' was built under R version 4.1.2
```

```
## Warning: package 'stringr' was built under R version 4.1.2
```

```
## Warning: package 'forcats' was built under R version 4.1.2
```

```
## Warning: package 'lubridate' was built under R version 4.1.2
```

```
## -- Attaching core tidyverse packages ----------------------- tidyverse 2.0.0 --
## v dplyr     1.1.2      v readr     2.1.4
## v forcats   1.0.0      v stringr   1.5.0
## v ggplot2   3.4.4      v tibble    3.2.1
## v lubridate 1.9.2      v tidyr     1.3.0
## v purrr     1.0.1
```

```
## -- Conflicts --------------------------------------------- tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors
##
## Attaching package: 'psych'
##
##
## The following objects are masked from 'package:ggplot2':
##
##     %+%, alpha
```

Introduction

For this project, I am working with a data set that was collected in my lab last year. This study examined the potential relationship between participants' evaluation of politicized scientific information and trait-level science curiosity. Months prior to data collection for this particular study, the participants completed a baseline assessment that included common measures related to cognitive processing. These included scores on Raven's Progressive Matrices (RPM; a nonverbal general cognitive ability task; `rav_scored`), Need for Cognitive (NFC; `nfc_scored`), Actively Open-minded Thinking (AOT; `aot_scored`), and verbal logic (VL; `vl_scored`). (All of these scores are simply the number of items correctly answered.) They also provided demographic information (i.e., age, gender, race, education) and reported their political ideology (`conservatism_f`, where higher values indicate greater levels of conservatism).

Later, the same participants (N=538) were invited back to participate in this study. This time, they completed a task in which they saw a series of graphs depicting longitudinal data about a variety of different environmental metrics. They were then asked to make projections about relative future values (e.g., whether future levels will be higher or lower than current levels) of those metrics, given the information given to them by the graphs. These graphs differed in two crucial ways. The first variable was the political or apolitical nature of the metrics each graph displayed. The political graphs displayed metrics related to climate change (e.g., atmospheric CO2 levels); the apolitical graphs showed environmental metrics unrelated to any political platform (e.g., population levels of harmless, lake-dwelling bacteria). Second, the graphs differed in whether the slope between the final two data points was congruent or incongruent with the global trend of the data. In instances where the local and global trends are incongruent, people can be susceptible to endpoint bias, a phenomenon in which people overweight recent, local trends in longitudinal data as opposed to overarching, robust global trends. In this study (or in the aspect of the study I'm focusing on for this project), we tested whether conservative participants responded with more endpoint bias on politicized items when local trends were incongruent with global trends but congruent with conservative ideology. Participants provided estimates for future values of each metric 10 years from now and 20 years from now. Performance on these items is represented in the data by the variables ending in `_ch`, which indicate the extent to which a participant responded in line with the local trend over the global trend (1 = heavily weighted global trend, 7 = heavily weighted local trend).

After this task, participants completed a science curiosity inventory (the score of which is represented by `sc_scored` in the data, where higher values indicate greater science curiosity). This score was calculated using a formula developed by the creator of the scale. Lastly, they answered three open response questions: 1) one in which they reported what they were thinking and feeling during the graph task, 2) one in which they reported what they think the graph portion of the experiment may have been studying, and 3) one in which they reported what they thought the science curiosity inventory may have been studying (it was embedded in a larger "general interest" inventory). I generated sentence embeddings for each question and they are included here, labeled `o1`, `s1`, and `s2`, respectively. From my initial visual inspection, the responses to these items seemed interesting because there seemed to be a distinct split between participants on whether they picked up on the political nature of the study. Some provided responses that were very much at face value (e.g., "the graph portion was studying how we understand graphs"; "the interest inventory measured what kind of topics I'm interested in"). Others correctly guessed that we were interested in their evaluation of climate change information based on political ideology, though neither climate change nor politics was ever explicitly mentioned in the study (political ideology was only reported in the baseline).

In the case of this project, I chose an outcome variable that was not included in our preregistered analyses: climate change risk perception (`cc_risk_sc`, where higher values indicate greater climate change risk perception). While not featured in our formal hypotheses, this variable is (obviously, by definition) a more proximal measure of climate change beliefs than political ideology. Studying climate change risk perception has obvious broader implications from a social and environmental context; however, I also chose to focus on this variable out of curiosity. My preregistered analyses were focused on predicting endpoint bias, using information like political ideology, and the cognitive variables included in this data frame were also not part of my preregistered analyses. I thought this project could be a beneficial opportunity to see the data through a different lens, especially since I would like to examine this data in a more exploratory fashion for

hypothesis generation.

This data does not contain any missing values that would necessitate imputation; the variables I'm examining had validation requirements on the Qualtrics survey, so participants were not able to move on in the study without answering each question.

	mean	sd	median	min	max	skew	kurtosis	se
climate_risk	8.478	2.863	9	1	11	-1.149	0.332	0.123
conservatism	3.705	1.807	4	1	7	0.149	-1.208	0.078
science_curiosity	24.621	5.774	25	9	40	-0.161	-0.299	0.249
need_for_cognition	21.364	6.681	23	6	30	-0.634	-0.623	0.288
actively_openminded_thinking	33.056	5.272	34	9	42	-0.745	0.996	0.227
verbal_logic	2.918	0.918	3	0	4	-0.458	-0.603	0.040
ravens_matrices	5.625	1.563	6	1	10	-0.089	0.427	0.067
arctic_ice_bias	2.582	1.311	2	1	7	1.124	0.909	0.057
global_temperature_bias	2.355	1.160	2	1	7	1.302	2.031	0.050
co2_levels_bias	1.916	1.017	2	1	7	1.621	3.660	0.044
icesheets_mass_bias	1.924	1.196	2	1	7	1.829	3.728	0.052

	median	min	max	skew	kurtosis
climate_risk	0.182	-2.612	0.881	-1.149	0.332
conservatism	0.163	-1.497	1.823	0.149	-1.208
science_curiosity	0.066	-2.706	2.664	-0.161	-0.299
need_for_cognition	0.245	-2.300	1.292	-0.634	-0.623
actively_openminded_thinking	0.179	-4.563	1.697	-0.745	0.996
verbal_logic	0.089	-3.179	1.179	-0.458	-0.603
ravens_matrices	0.240	-2.958	2.799	-0.089	0.427
arctic_ice_bias	-0.444	-1.207	3.370	1.124	0.909
global_temperature_bias	-0.306	-1.168	4.003	1.302	2.031
co2_levels_bias	0.082	-0.901	4.999	1.621	3.660
icesheets_mass_bias	0.064	-0.773	4.245	1.829	3.728

Methods

I tested four modeling approaches:

- Linear regression (unpenalized)
- Linear regression with ridge penalty
- Linear regression with elastic net penalty
- Decision tree algorithm

For every approach, I employed 10-fold cross-validation to examine model performance. The latter three approaches all required hyperparameter tuning. For all hyperparameters, I essentially triangulated the optimal value through trial and error of testing different ranges and levels of granularity. For ridge regression, `alpha` was fixed to 0; for elastic net, both `alpha` and `lamba` were tuned. I chose ridge regression over lasso because I believe (?) that the ridge penalty can work well when you suspect that a large portion of features in the data may be unrelated to the outcome variable, and I suspected that the sentence embeddings would add a significant amount of noise to the data.

For the decision tree algorithm, in addition to the `complexity parameter`, I attempted to write some for loops to tune `minsplit` and `maxdepth` as well. While I **believe** (?) these loops ended up working correctly,

I haven't been able to get R to actually successfully run through the loops without terminating the session. This is something I'd like to continue working on, as well as applying Random Forests as well.

For each model, I planned to judge performance based on comparing $RMSE$, MAE, and R^2.

Results

In short, the results of all models I tested were underwhelming.

Provide results of model evaluation.

- Compare and contrast results from different approaches and discuss model performance
- Discuss final model selection and evidence for selection

Discussion

I do not believe the models I present in this project are anywhere close to optimal, and I'm curious to keep working on this further. This is not necessarily surprising as the model I tested through these approaches was not based on specific theoretical rationale. On the other hand, based on my previous/preregistered analyses of this data, I was anticipating

I'm curious as to whether the sheer volume of completely uninterpretable sentence embedding features could have had an effect here, though I suppose the entire point of these approaches is to reduce noise and prevent overfitting with unrelated predictors. Overall, I think there's a lot more that I could do in terms of data cleaning and feature selection, although it's also possible that even in "optimal" form this model, regardless of algoritm, could simply be a bad fit for this data. Though not the main goal of this assignment, I think that attempting to work through those `for` loops (even if I still haven't gotten them to fully cooperate) was actually quite a helpful exercise in functional programming in addition to manual tuning of hyperparameters.

- What variables were most important predictors? Was this expected or surprising?
- Were different models close in performance, or were there significant gaps?