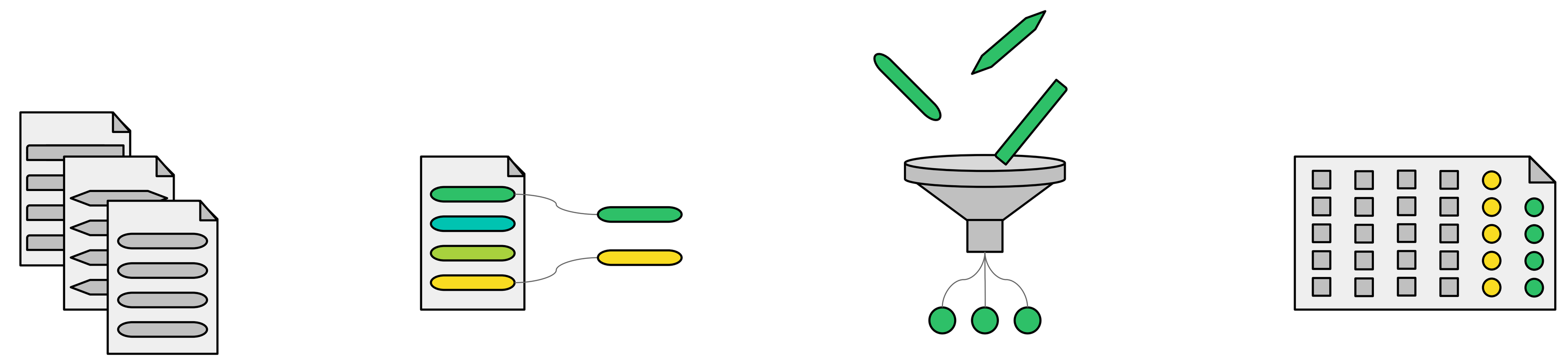
Demographic Predictors of Climate Opinion

Energy and Environment Transitions (ENVENT) Lab at UCSB
Undergrads: **Emma Franzblau**, **Annie Adams**, **Andrew Bissell**, **Johnny Yu**
Sponsor: **Prof. Matto Mildenberger**
Mentor: **Adam Waterbury**

Introduction / Background / Data

Goal: Clean and organize covariate data from 140+ surveys, merge demographic information into existing data processing routines. Then develop predictive models to understand how demographic attributes are differentially predictive of climate opinion in different countries, continents, and time periods.

Data Pre-Processing Pipeline:



Data	Extract	Standardize	Merge
Given 149 surveys that included climate opinion and demographic questions from the past 23 years	Identify and pull relevant demographic variables, along with their unique responses, from each of the surveys	Determine standardization techniques and standardize ~4,000 unique demographic responses	Add standardized demographics to given dataset of climate opinion variables with 3 million+ responses (*megapoll*)

Data

	source	year	action_policy_shouldstop_afrobarometer	concern_climatenegative_afrobarometer	concern_countryimpact_afrobarometer	happening_aware_afrobarometer	human_maincause_afrobarometer	respondent_gender
1	afrobarometer7_2016	2016	-2	-1	-2	0	-1	1
2	afrobarometer7_2016	2016	2	2	-2	1	2	0
3	afrobarometer7_2016	2016	NA	NA	-2	1	2	1
4	afrobarometer7_2016	2016	NA	0	-2	1	2	0
5	afrobarometer7_2016	2016	3	2	-2	1	2	1
6	afrobarometer7_2016	2016	-2	-1	-2	0	-1	0

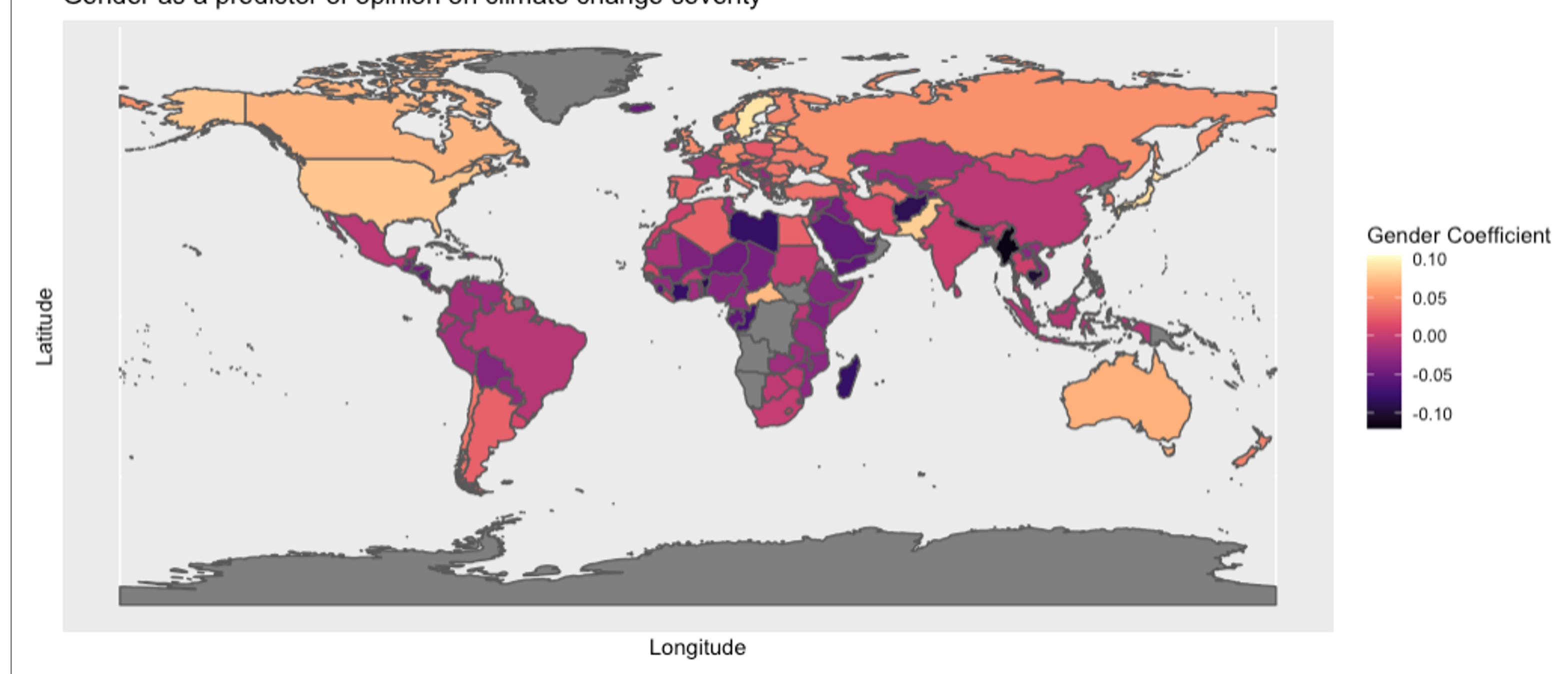
The features of our dataset are different questions pertaining to climate change and respondent demographics which we merged into the final dataset, such as **gender, age, level of education, race, income,** etc.

Our dataset encounters numerous missing values in surveys that have country responses as variables rather than observations. To fix this, we created modification variables that shift the country responses to observations in order to obtain a dataset that we are able to model on.

Since we are working with self reported survey data there is also likely to be a good amount of bias in our data. Due to differences in question wording across surveys, further generalization (*such as the conversion from ordinal to binary format*) of a respondent's answers to these questions leaves more room for potential information loss and errors.

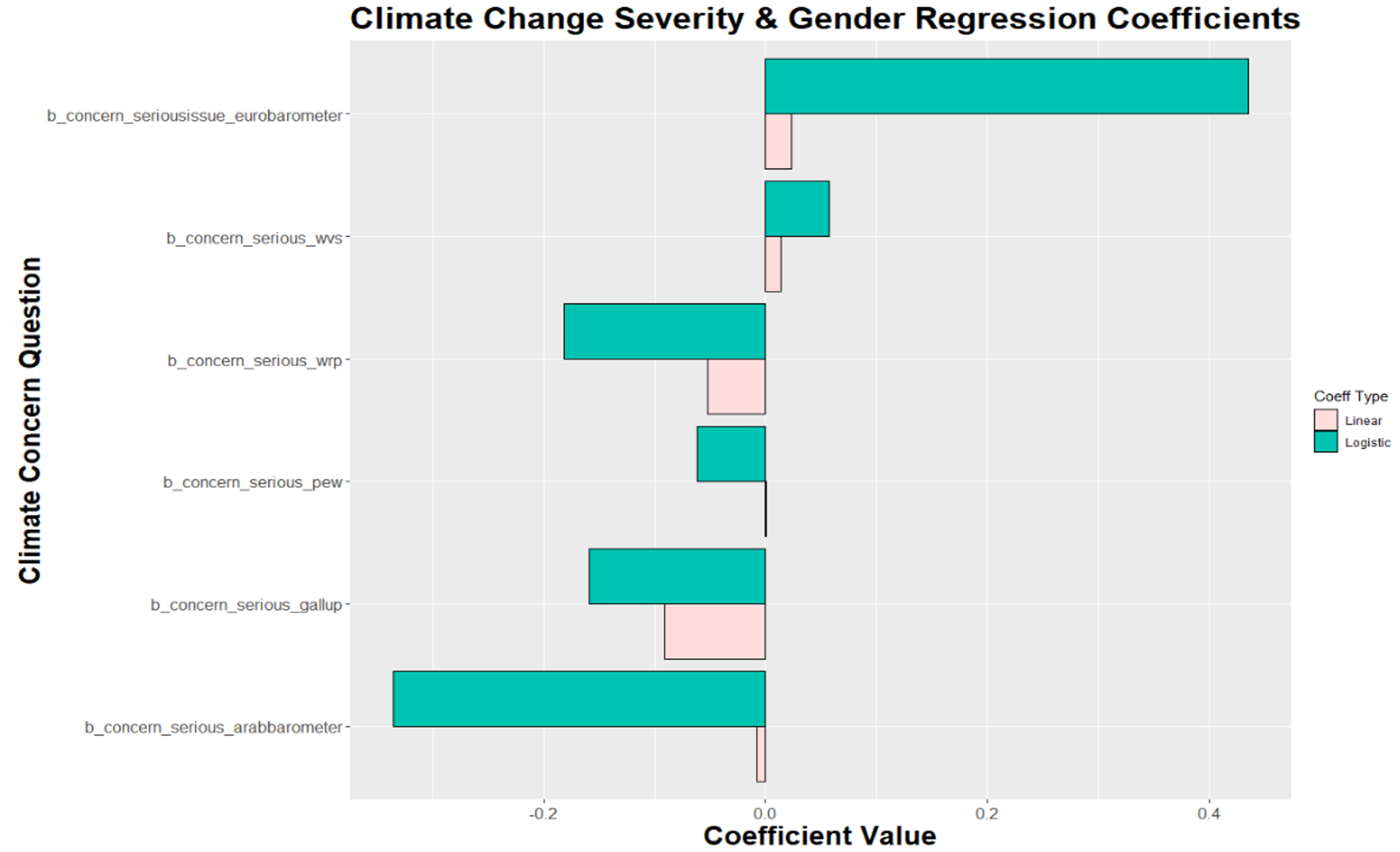
Models and Analysis

Gender as a predictor of opinion on climate change severity



Our results from our linear regression model above display the **countries in which gendered individuals are more likely to see climate change as a threat**. A coefficient score of .10 signifies that a person who identifies as female will, on average, be 10% more likely to see climate change as a severe threat in comparison to males.

Climate Change Severity & Gender Regression Coefficients



The bar chart displayed above conveys the similarities between linear regression and logistic regression results. While the interpretations of the coefficient magnitudes themselves slightly differ, **both models predict the correct positive/negative relationship of gender and its effect on climate opinion**.

The difference in interpretation of the two magnitude types is due to the fact that linear regression coefficients represent the direct magnitude change in the response value when the binary predictor value increases from 0 to 1, whereas logistic regression coefficients represent the odds ratio between the predictor values - the ratio is equal to e^{β} which, in the context of our project, translates to "Females are e^{β} times the odds of Males for viewing climate change as a threat."

Discussion and Additional Notes

Over the course of the capstone, we had several discussions with the ENVENT team in regards to the types of models we could try to incorporate.

When presenting results in the political science space, there is a need to focus on the immediate **readability and interpretability of results**. In our case, logistic regression may possess just enough extra complexity compared to linear regression to not be as immediately verifiable for a reviewer, particularly one who isn't in the environmental politics space.

Additionally, in areas of dense common support, the best linear approximation of the conditional expectation function is sufficiently similar to that of the logit - **in the end, we'd achieve the same learning of the world from both approaches.** And so because both SHOULD yield similar results, our workflow consists of running a linear regression model and then utilizing logit for robustness. We then made sure to take a look taking a look at **model summary statistics** in order to further validate our findings.

```
Call:
lm(formula = b_concern_seriousissue_eurobarometer ~ respondent_gender,
    data = concern_serious)

Residuals:
     Min       1Q   Median       3Q      Max
-0.95432  0.04568  0.04568  0.06887  0.06887

Coefficients:
                  Estimate Std. Error t value Pr(>|t|)
(Intercept)       0.931126   0.002734 340.571  < 2e-16 ***
respondent_gender 0.023195   0.003800   6.104 1.06e-09 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.2313 on 14840 degrees of freedom
  (3078259 observations deleted due to missingness)
Multiple R-squared:  0.002505,  Adjusted R-squared:  0.002437
F-statistic: 37.26 on 1 and 14840 DF,  p-value: 1.058e-09
```

```
Call:
glm(formula = as.factor(b_concern_seriousissue_eurobarometer) ~
    respondent_gender, family = binomial(), data = concern_serious)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-2.4844   0.3058   0.3058   0.3778   0.3778

Coefficients:
                  Estimate Std. Error z value Pr(>|z|)
(Intercept)        2.60412    0.04667  55.794  < 2e-16 ***
respondent_gender  0.43524    0.07186   6.057 1.39e-09 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 6478.7  on 14841  degrees of freedom
Residual deviance: 6441.4  on 14840  degrees of freedom
  (3078259 observations deleted due to missingness)
AIC: 6445.4

Number of Fisher Scoring iterations: 5
```

Conclusion & Next Steps

At the beginning of W22, our team had set out to complete the megapoll modifications and begin conducting a preliminary exploratory analysis for seven demographic variables (*see Data section*). The **data-preprocessing task** ended up becoming our **largest obstacle** over the course of this capstone and subsequently we ran out of time to analyze every variable. With that said however, **we were ultimately still able to set the foundation for further exploration of the data in the next iteration of the project as a result of completing the necessary proof of concepts** for modifying the demographic variable values and merging them into the megapoll.

Here are some possible next steps and directions the ENVENT lab can now take in the future with the dataset as a result of our work:

- **Continue merging demographic data** into the megapoll for analysis.
- Continue generating exploratory plots to pinpoint potential areas of interest for further investigation.
- Begin **region-specific** or **time-series** explorations.
- Develop more complex supervised and unsupervised ML models such as **SVMs** or **PCA** which could make predictions of an individual's climate score based on their demographic info.
- **Further improve standardizations** for concern scores and demographics.

Data Science Capstone

UC SANTA BARBARA