

Caroline Siekman & Eli Weise
Data Science
Professor Robin Belton
March 6, 2025

Final Data Science Project Proposal

Data Set 1:

<https://www.icpsr.umich.edu/web/ICPSR/studies/36368>

Source of the data:

- Inter-University Consortium for Political and Social Research
- Lead Investigators: Barry Rabe (University of Michigan), Christopher Borick (Muhlenberg College), Sarah B. Mills (University of Michigan)

When and how it was originally collected: The National Surveys on Energy and Environment (NSEE) is a biannual national opinion survey. This data set contains 18 survey batches collected from Fall 2008 to Spring 2017.

The cases: Each case represents an individual's responses to the survey questions. There are 15,136 total cases.

General Description of Relevant Variables: There are 321 columns containing survey responses. There are a myriad of applicable variables, many of them containing descriptors of the person that could make for interesting comparisons on pressing climate issues.

Personal Characteristics: DEMOG_AGE (age), WORLDVIEWS_TRUMP (Agreement with Donald Trump on important issues facing the country today), DEMOG_STATE_EE (State), DEMOG_INCOME (family income), DEMOG_GENDER (gender), DEMOG_EDU (highest level of education), DEMOG_POLVIEWS (conservative to liberal), DEMOG_NEWS (Primary source for national & world news), DEMOG_RACE (race), DEMOG_RELIG (religious affiliation) and more

As well as survey questions showing people's views on the climate crisis such as:

GW_PREVENT (Can humans prevent further climate change), CCVSGW_CONCERN (How concerned an individual is about climate change), WILLING2PAY_TEMPRISE_OPEN (How much would you be willing to pay/month to prevent temperatures from rising), INTL_SUPPORT_V1 (Support for international greenhouse gas treaty), BELIEVER_PROBLEM (Is global warming a serious problem), INTL_ECONOMICIMPACT (Overall, the U.S. economy would benefit from leaving the international climate treaty designed to reduce greenhouse gas emissions), FRACKING_SUPPORT (Support for fracking in the United States), CDR_SUPPORT (Support for use of carbon dioxide removal), GW_TECHSOLVEPROB (Technological innovations are going to solve the climate problem), and more

Research Question

This data set contains incredibly relevant information to our current society. The climate crisis is one of the largest issues that we are facing, and understanding the general population's stance and views on climate change is important for policy making decisions. Additionally, understanding where anti-global warming stances come from, whether it be the media, political parties, education, state, can provide a large amount of interesting data. Two areas that we think could be interested in looking into would be how partisan polarization may have shifted climate policy in the period from 2008-2017. As well as perceived economic risk (belief climate policy will hurt the economy), reduces support of climate legislation, even among climate believers (with this question maybe also looking into what conditions respondents are more likely to pay for renewable energy policies).

We anticipate that in this data set, there will have been a partisan shift in climate support, with a larger number of republicans shifting to be anti-climate in the more recent years (Trump era) and Democrats swinging hard the other way in support of the climate agenda. We also predict that perceived economic risk will reduce the support of climate legislation in both political parties, with very few respondents being willing to pay for renewable energy policies.

Dataset 2:

<https://github.com/rfordatascience/tidytuesday/tree/main/data/2023/2023-12-19>

Identify the source of the data:

IMDb database

When and how it was originally collected:

This data was collected from IMDb on December 19, 2023

The cases:

Each case represents an episode of a TV show that contains “holiday”, “Christmas”, “Hanukkah”, or “Kwanzaa” in the title. There are 2289 total episodes.

General description of relevant variables:

year: the year the episode was released

christmas: boolean representing if the title contains “Christmas” or variations

hanukkah: boolean representing if the title contains “Hanukkah” or variations

kwanzaa: boolean representing if the title contains “Kwanzaa”

holiday: boolean representing if the title contains “holiday”

Research Question:

We chose this dataset because it may show general sentiment towards winter holidays over time. There is one main question that we will be asking: “Do non-Christmas centered holiday specials become more common as time gets closer to the present?” We believe that the data will show that they did become more common over time. A secondary question that we can ask is if the instances of “Christmas” in titles decreased while titles with “holiday” increased. We also believe that to be the case as certain parts of society become more open to making people of all backgrounds feel represented.