

Task 1 Documentation

Use of World Religion Dataset: We opted to use the world religion dataset split by nation rather than the global version. We imputed the data in between each set of 5 years by finding the difference in values for each set of 5 years and incrementing the values by 20% of the difference each year. We chose to only focus on the percentage of population columns, as it was independent of population size, unlike the totals.

Year Cutoff: We chose to start observing our data at 1900, as there is a lot of missing data before 1900 and the data that exists is likely less accurate.

Naming Conventions: Regarding countries that carried different names in different datasets, we chose to establish a common name for such countries and made the appropriate replacements. Additionally, countries pre- and post-split were not always renamed accordingly so we made the appropriate changes in the appropriate years. We made these changes using Python. The country names and years are as follows:

Myanmar (Burma) , Vietnam (1976 onward), North Vietnam (1954 to 1976), South Vietnam (1954 to 1976), Yemen (after 1990), North Yemen (until 1990), South Yemen (until 1990), Congo (Brazzaville), Congo (Kinshasa), Germany (after 1990), West Germany (1949 to 1990), East Germany (1949 to 1990), Czechoslovakia (after 1992), Czech Republic (until 1992), Slovakia (until 1992), United States of America, Ivory Coast, Cape Verde, Turkey, Palestine (West Bank), Palestine (Gaza), Timor-Leste, Gambia

Merging: Different datasets had different ways of identifying countries. The v-dem and bti datasets had both the country name and abbreviations, while the fiw dataset only contained the country name and the wrp dataset only contained the country abbreviation and COW code. The first three datasets were therefore merged on country name, while the last, which was identified later, was merged in on country COW code.

Column Removal: We removed the following columns from the v-dem dataset because...

V-Dem Factors:

v2x_civlib	CivLib
v2x_clphy	Violence
v2x_clpol	PolLib
v2x_freexp_altinf	FreeExpress
v2xcl_rol	LawEqual
v2xeg_eqprotec	EqProtect
v2x_corr	Corrupt_vdem
v2x_execorr	Ecorrupt
v2x_pubcorr	Pubcorrupt
v2x_rule	RuleLaw
v2xcl_acjst	Justice
v2xcl_prpty	Property
v2juhcind	CourtIndep

Task 2 Documentation

To run the below visualizations, be sure to download the data files

Country_and_Territory_Ratings_and_Statuses_FIW_1973-2024.xlsx.zip,

V-Dem-CY-Full+Others-v14.csv.zip, free-fair-elections-bti.csv.zip, mutated_NWR.csv.zip, and imputed_cn.csv.zip.

Visualization 1: The code for this visualization is found in Statistics Capstone.ipynb. Run the code until the code chunk titled # map visualization new/updated, which will display the first visualization. Adjust the sliders to change the year, and hover over the country to display its name and value. Click between the variables to change the display.

Visualization 2: This code is under the v2simulation.R file. Load the imputed_cn dataset and run the shiny app. Use the Country and Year Dropdown menus to select a specific country and year, the sliders should automatically update to reflect the values on the dataset from the chosen country and year. Adjust the slider values and the main panel will update in real time to show the score originally stored in the dataset, the model's prediction after any slider adjustments, and a text label ("Free", "Partly Free", or "Not Free") based on the adjusted score. The freedom house scores are as follows, Free: Score between 1.0–2.5, Partly Free: Score between 3.0–5.0, and Not Free: Score between 5.5–7.0.

Visualization 3: The code for this visualization is found in the bubble.R file. Select variables for the three axes: x, y, bubble size. Use the slider to select the year of interest, and click the play button to begin the animation from the selected year until the most recent. One can hover over bubbles for further information about the given datapoint. As well, there is a menu on the right which allows for the selection of countries to be excluded in the visualization.

Visualization 4: The code for this visualization is found in Statistics Capstone.ipynb. Run the code until the code chunk titled # display graph over time for country using dropdown (new, multi-selection for countries)), which will display a line graph with options to multi-select countries. The subsequent code chunk will display a line graph with options to multi-select variables. Select as many as desired, but keep in mind the view may be less interpretable with many selections.