Frederick Peña Sims, Alexandra Salo and Sofia Villamil, 20.03.2024

Survey Research Methodology Assignment

In this assignment we seek to explain Cross-Country Differences in support for transgender individuals obtaining official documents (cq19 in the questionnaire) to match their inner gender and then to continue and create a model that is able to predict the support levels of other countries. First we load the data and proceed to clean it. Then we add extra country level variables, we chose to add;

- Legal variables - we included a variable that showed if a country has general legal protections for transgender individuals. We were able to find a variable that shows this, which is a binary variable created by the Harvard Dataverse, TRIP (Williamson, 2023).
- Economic variables - usually, economic variables are quite important in general. As such, we wanted to see if they may be influential for this analysis as well. For this purpose we chose two benchmark economic variables, GDP and unemployment (World Bank, 2024).

Narrowing down the variables

Since the dataset is very large, we first tried reducing variables according to its significance and correlation. We quickly realized that this wasn't enough and we were still left with hundreds of variables, so opted for another approach to perform our feature selection. This approach consisted of performing a literature review to find evidence of variables related to the support of transgender rights. After the literature analisis, variables for transgender rights support were gathered and we sought out its approximate counterpart in the dataset.

Variables that exist in the dataset and are as well prominent in the literature are the following:

Harrison and Michelson (2019) show that men are more likely to oppose transgender rights citing perceived threats to their notion of masculinity as an effect channel. Further, according to Norton and Herek (2013), a binary conception of gender, political ideology, religion, and contact with a transgender people are all significant variables and are associated with the target variable (see also Earle et al., 2021; Campbell et al., 2019; Flores, 2015). Flores et al. (2016) note the importance of age in the support for transgender rights while Flores (2015) shows that information about transgender people increases support for transgender rights. Lastly, Earle et al. (2021) who show in their 77 country multilevel analysis that more expanded legal rights also increases support for transgender rights. As such, we added information on the legal situation concerning transgender rights to our data (see Williamson, 2023).

In addition to these variables, we developed our own hypotheses for variables we assumed to be associated with and possibly be predictors of the dependent variable. These include solidarity among minority groups, education, the place of residence and economic situation of individuals and the country.

Explaining cross country differences

In order to explain these differences we start with the descriptive analysis, we conduct associative tests to check for statistically significant associations between each of our inferred predictors and the target variable. The descriptive analysis oscillates between individual and country level variables. Whenever reasonable, we inquire for country level differences for statistically significant using individual-level variables. As such, we create three additional country-level variables out of our individual-level variables and check for the strength of the association. To make our work more palpable, we use various visualizations. This extensive descriptive analysis forms the basis for our following modeling process.

To see the results and interpretations of the full descriptive analysis, we give the full results in the rmd file. Here we will highlight the most important variables in which we were able to see country level differences. We find that opposing the introduction of a third gender option (binary conception of gender) is highly associated with the lack of support. We demonstrate the convergence of both variables on the country level by visualizing the share of ‘No’s’ concerning our target variable and the binary conception of gender. Furthermore, on the individual level, we inquire about the role of religion showing that non-believers and protestants are more prone to support transgender rights while

specifically orthodox Christians oppose them. While not fully illuminating country differences seen in the support for transgender rights, high-rejection countries are also found to partially converge according to the share of orthodox Christians among the respondents. Following, we present results in line with literature on the effect of information on transgender people showing that greater awareness about the discrimination the community faces is associated with greater support for the expansion of transgender people. We extrapolate the share of people who indicate discrimination against transgender people is very or fairly widespread by country and show that this share is strongly associated with the share of people who support the expansion of transgender people by country. Similarly, supporting the notion that lesbian, gay and bi-sexual people have equal rights is associated with greater support of transgender rights. It is important to note that, different to the previous variables, the association is two-ways where a rejection that LGB people should have equal rights is associated with a rejection of transgender rights. Again, we extrapolate the share by country, this time for those agreeing with the notion that LGB people should have equal rights correlating it with the share of people supporting transgender rights by country. The results suggest a strong correlation. Moving on to legal protections for the transgender community, we find a statistically significant association between general legal provisions to protect the transgender community and support to expand transgender rights, with the former being positively associated with the latter. We decided to not include the ideology of the government as a variable as we thought this much better represents the legal and political culture of a given country vis-a-vis the transgender community. Lastly, we show that there is a statistically significant difference between those supporting and opposing the expansion of transgender rights on the country level with regards their unemployment rates and GDP.

In addition to these country level variables, we provide further insights into demographic variables contributing to country-level differences according to the literature. This includes gender, political ideology, age, age when completing formal education, contact with transgender people, being sexual minority, the personal economic situation and the place of residency. Their full interpretations can be found in the rmd file.

Moving on, these statistically significant variables were chosen, in order to further explain country level differences as we undertake statistical modeling. First, however, we imputed the variable ‘age when stopped formal education’ as we considered that this variable of the variables containing missing data had some of the highest rates of missing data (however ultimately not very high) and since it was a numerical variable, we wanted to test to see if it would affect the variables significance in the regression model. However we came to the conclusion that it did not affect the significance of the variable, as it was insignificant in the regression model with and without imputation. Otherwise the variables we had chosen had quite low levels of missing data so we didn’t deem imputation necessary for any other variables.

For our modeling approach, we draw on the variables chosen through our previous descriptive analysis. Since we have multilevel data (on the individual and country-level) we adjust our statistical modeling approach accordingly. As such, we draw on hierarchical (multilevel) logistic regression modeling. We considered adding weights to our model, however since the data is balanced and each country is very similarly represented in the data we decided against using weights. We develop this model by, firstly, performing an individual level analysis to get an idea of the individual level variables and to assess significance.As part of it, we undertake feature engineering as we suspect non-linear relationships for some variables. We see that some of the variables are insignificant in the analysis and thus choose to remove these variables: education variable, type of area lived in and more. After the individual level analysis we move to include the country level variables in the analysis, since many variables produce very similar results in both levels of the analysis, most will be shared in the deeper level analysis. We conduct our analysis with all variables employing a generalized linear mixed-effects model (GLMM), where the country is deployed as the random effects variable, noting clear overfitting and possible scaling issues, while ensuring there is no multicollinearity. Step by step, we remove the non-significant variables in the model comparing each model by BIC and AIC. Finally, we plot the residuals of our best models finding a large cluster around the zero point, which is not unexpected for binary data. Regarding the results and their interpretation,

we find that *gender* and *age* play significant roles in both levels of analysis, confirming literature cited. For the case of *gender*, females appear to show higher odds of support compared to males. This could be related to the potential influence of the traditional gender roles and perceptions we have on our society. Moreover, the *age* variables show that there is a slight increase in support for our dependent variable for each additional year of age, up until a certain age, after which it starts decreasing (non linear relationship). This might suggest that life experiences or generational shifts could gradually influence views towards more inclusive attitudes, however, age influence is minimal in comparison to the other variables represented in our regression.

The categorical variable *political Ideology* significantly impacts attitudes of support for our target variable in both levels of analysis. The findings suggest that an alignment towards a more conservative view, results in being less supportive to allow changes in civil documents for transgender individuals. It showed a 22.8% decrease of support for our target variables if the individual identified himself as conservative. Additionally, *religious affiliation* further complicates the picture, as identifying as Catholic decreases the support for the target variable by about 21,2%. In contrast, individuals identifying with the Jewish faith are slightly more supportive, but, we need to take into consideration that in our data set the proportion of individuals identified as Jewish is really small in comparison to the Catholic faith individuals. Thus if we see demographic differences between countries in relation to religion and political ideology, you could argue that these drive country level differences as well.

Concentrating on the economic aspect, individuals who established that they had not had *difficulties in paying the bills*, are slightly more supportive of our dependent variables. Indicating that economic factors may intersect with views on rights and equality on both levels of the analysis.

Furthermore, respondents without *contact* with transgender individuals show a decreasing support for our dependent variable by 45.5% in both levels. We find that the individuals who responded with an ‘automatic refusal’ tend to be the least supportive individuals towards our dependent variable. The findings underscore the crucial importance of representation and visibility for the transgender community, as well as the development of personal relationships that foster understanding and empathy. These will most likely increase the support for allowing transgender individuals to change their civil documents to match their inner gender identity.

Further, we find that not being part of a *sexual minority* is associated with an extremely high non-supportive attitude regarding the legal provision that transgender individuals should be able to change their civil documents.

Indicators on the general support in a country for supporting the possibility of a third gender indication (*binary gender conception*) and country-level *awareness* of discrimination against transgender individuals are strongly associated with higher supportive attitudes towards allowing transgender individuals to change their civil documents.

Unemployment is significantly and positively associated with the dependent variable, suggesting that higher unemployment rates within a country might be associated with greater support for transgender rights, which could indicate a variety of social dynamics but may also be a coincidental association.

In conclusion, the main reasons for cross-country differences in support levels for transgender individuals obtaining official documents according to our mixed effects model appear to be related to demographic factors (gender and age), political and religious affiliations, social determinants, awareness regarding LGBTI issues, and economic conditions within a country.

Predicting support for transgender rights

Finally, since we were able to model the variables and find the variables of significance, we employ multiple prediction models to predict the support of a transgender person including their true identity in their official documents. In order to predict as accurately as we can, we test out various prediction models testing their performance while considering the replicability of the analysis (by customizing and validating the parameters in order to run models in a realistic time frame). We employ logistic regression, k-nearest neighbor, decision trees, random forest, gradient boosting and neural networks.

For each model we evaluate the success of the model using accuracy, specificity and sensitivity after adopting the optimal threshold provided by the ROC curve.

We find the Gradient Boosting model to be the best predictive model outperforming the other tested models. The model not only demonstrates one of the highest accuracy levels but more importantly achieves the most balanced rates of specificity and sensitivity, which, considering we want to predict support for transgender rights, is the prime objective. We further inquire about the importance of variables feeding into the Gradient Boosting model, using a variable importance plot. And creating a partial dependence plot for the variable single out as the most significant one.

Bibliography

Campbell, M., Hinton, J. D. X. & Anderson, J. R. (2019) *A systematic review of the relationship between religion and attitudes toward transgender and gender-variant people*. International Journal of Transgenderism, 20:1, 21-38, DOI: 10.1080/15532739.2018.1545149

Earle, M. et al. (2021). *A multilevel analysis of LGBT (Lesbian, Gay, Bisexual, Transgender) rights support across 77 countries: The role of contact and country laws* . British Journal of Social Psychology. doi:10.1111/bjso.12436

Flores, A. R. (2015) *Attitudes toward transgender rights: perceived knowledge and secondary interpersonal contact, Politics, Groups, and Identities*, 3:3, 398-416, DOI: 10.1080/21565503.2015.1050414

Flores, A. R., Brown, T. N. T., & Park, A. S. (2016). *Public Support for Transgender Rights: A Twenty-three Country Survey*. The Williams Institute at UCLA School of Law. http://www.jstor.org/stable/resrep34965

Harrison, B.F., Michelson, M.R. (2019) *Gender, Masculinity Threat, and Support for Transgender Rights: An Experimental Study.* Sex Roles **80**, 63–75. https://doi.org/10.1007/s11199-018-0916-6

Norton, A.T., Herek, G.M. (2013) Heterosexuals' Attitudes Toward Transgender People: Findings from a National Probability Sample of U.S. Adults. *Sex Roles* **68**, 738–753.

Williamson, Myles, 2023, "Replication Data for: A Global Analysis of Transgender Rights: Introducing the Trans Rights Indicator Project (TRIP)", https://doi.org/10.7910/DVN/FXXLTS, Harvard Dataverse, V1, UNF:6:E4O2rb6DFd44VSZjm9SQNQ== [fileUNF]

World Bank. (n.d.). DataBank: World Development Indicators [Economy & Growth]. Retrieved March 13, 2024, from https://databank.worldbank.org/source/world-development-indicators)