

How is internet and news consumption connected to attitudes towards climate change? (ALL Countries)

Final report

in accordance with the course of

EVERY NUMBER TELLS A STORY - INTRODUCTION TO STATISTICS WITH R

Submitted by

Samiul Hossain Sajal

Matriculation number: 7406871



Master of Science (M.Sc.)

Department of Physics

UNIVERSITY OF COLOGNE

Submission Date: 31 May, 2023

Contents

	Page
1 Introduction	2
2 Results and analyses	2
2.1 Univariate Analysis	2
2.1.1 Mean, median and mode	2
2.1.2 Statistical analysis on nominal variable (<i>cntry</i>)	3
2.1.3 Statistical analysis on ordinal variable (<i>clmchnng</i>)	4
2.1.4 Statistical analysis on metric variable (<i>netustm</i>)	4
2.2 Bivariate analysis	5
2.2.1 Correlation between two nominal variables (<i>gndr</i> with <i>climate_change</i>) . .	5
2.2.2 Correlation between internet and news consumption towards the climate change	6
2.2.3 Correlation between two freely selectable variables (<i>agea</i> with <i>netustm</i>) . .	6

1 Introduction

The data from the eighth round of the European Social Survey were subjected to some univariate and bivariate statistical analysis in this study. The primary goal is to determine whether the respondent's "Internet use in a typical day" and "News about politics and current affairs" are correlated with the factors related to climate change for all nations. For the nominal, ordinal, and metric variables, the mean, median, and mode values are also computed and include some graphical representations of them. Finally, a fascinating relationship between the respondents' internet usage and age is shown.

2 Results and analyses

2.1 Univariate Analysis

2.1.1 Mean, median and mode

In the selected topic of this report, *netustm* and *nwspol* are ratio variables, *clmchnng* is the ordinal variable and *cntry* is the nominal variable. So first of all, this report shows the mean, median, and mode values of these variables. The mean, mode, and median values of *netustm*, *nwspol*, *clmchnng* and *cntry* variables are shown in table: 2.1. In this table, the values are calculated in two aspects: (a) Filtered Data: In this column, filter out all the rows which contain NA values of *netustm* or *nwspol* or *clmchnng* variables. Since there are no NA values in *cntry* variables, it's not necessary to filter out the NA values of this variable. (b) Complete Data: In this column, the NA values are excluded only the variable in which mean, median, or mode is being calculated. The mean value of *clmchnng* is not calculated, because it's an ordinal variable. The mean and median of the *cntry* variable are also not calculated because it's a nominal variable. The mode value of *cntry* indicates

that the maximum participation of the survey is from "Germany (DE)". The maximum count of the *cntry* variable of filtered data is 2062 and without the filtered data is 2852.

Variable	Filtered Data			Complete Data		
	Mean	Median	Mode	Mean	Median	Mode
<i>netustm</i>	197.86	135.00	120.00	197.63	135.00	NA
<i>nwspol</i>	78.51	60.00	60.00	85.43	60.00	60.00
<i>clmchnng</i>		1.00	1.00		1.00	1.00
<i>cntry</i>			DE			DE

Table 2.1: Mean, median and mode values of *netustm*, *nwspol*, *clmchnng* and *cntry* variables

2.1.2 Statistical analysis on nominal variable (*cntry*)

In this section, the frequency of the *cntry* variable is analyzed from the filtered data (exclude the rows which contain NA values of either *netustm* or *nwspol* or *clmchnng* variables) and shown in the figure: 2.1. The figure shows that the maximum number of respondents is from Germany (DE) and the percentage is 7.01. This statement is also matched with the mean value of the *cntry* variable which is shown in the previous table: 2.1.

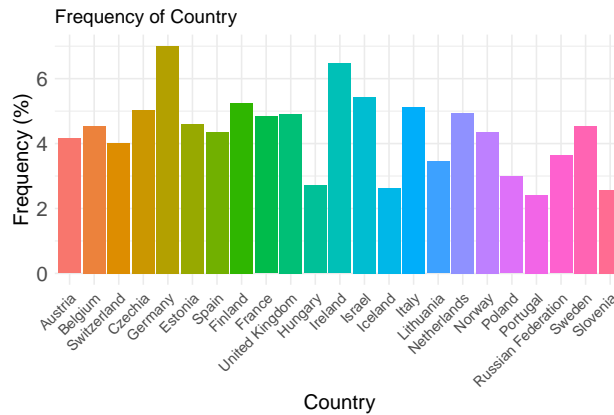


Figure 2.1: Proportion of country for the respondent of *netustm*, *nwspol* and *clmchnng* without their NA values

2.1.3 Statistical analysis on ordinal variable (*clmchnng*)

In this section, an ordinal variable *clmchnng* is analyzed with the NA values and illustrated through a barplot in the figure: 2.2. From the figure, it can be said that the maximum percentage (55.39%) of participants believe the climate is *definitelychanging* and the minimum percentage (2.20%) of the participants believe that the climate is *definitelynotchanging*.

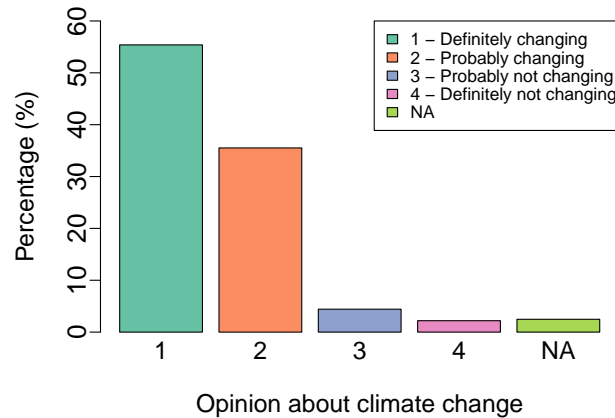


Figure 2.2: Percentage of different opinions about climate change including the NA values.

2.1.4 Statistical analysis on metric variable (*netustm*)

This section represents a histogram of a ratio variable (*netustm*) in the figure: 2.3. This distribution indicates that most of the respondents use the internet from 0 to 700 minutes.

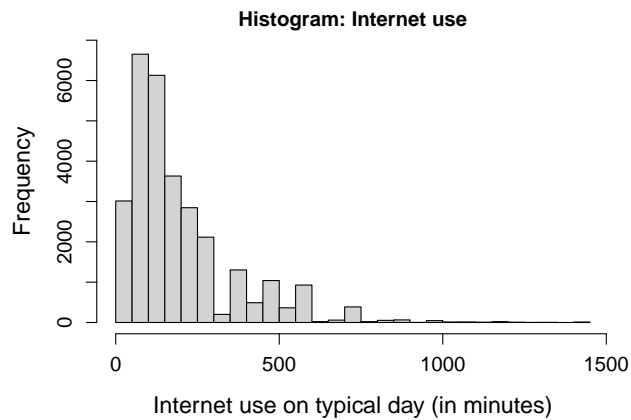


Figure 2.3: This histogram represents the distribution of *netustm* variable with break = 40.

2.2 Bivariate analysis

2.2.1 Correlation between two nominal variables (*gndr* with *climate_change*)

To find the correlation between two nominal variables, a new nominal variable is introduced as *climate_change*. In this variable, the *clmchnng* variable is divided into two categories: (a) **Changing** - The respondents of "Definitely changing" and "Probably changing" of *clmchnng* variable are grouped together in this category (b) **Not Changing** - the respondents of "Definitely not changing" and "Probably not changing" of *clmchnng* variable are grouped together in this category. Now a crosstable is being plotted to analyze the correlation between two nominal variables *climate_change* and *gndr* which is given in the R script. The graphical representation is also illustrated in the figure: 2.4.

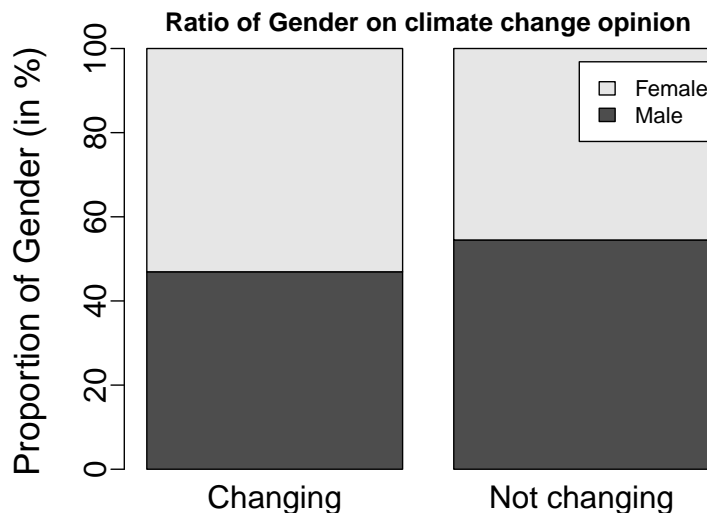


Figure 2.4: The ratio of different opinions about climate change from the respondents of different countries

From the crosstable, it is found that $\chi^2 = 62.545$, Cramer's $V = 0.038$, and $p = 0.000$. Cramer's V , a measure of connection between categorical variables, is equal to 0.038. It is developed from the chi-square statistic and takes the sample size into consideration. Cramer's V has a range of 0 to 1, with 0 denoting no association and 1 denoting the ideal association. The figure of 0.038 in this instance points to an extremely weak correlation between the variables.

2.2.2 Correlation between internet and news consumption towards the climate change

The main task of this report is to find the correlation between *netustm* and *nwspol* towards the variable regarding climate change variables (*clmchnng*, *enefvap*, *rdcenr*, *wrcmch*). To analyze the correlation between the variables corresponding to the selected topic of this report, a correlation matrix is created from which a certain part is included in this report.

Variable	nwspol	netustm	clmchnng	agea	gndr
nwspol	1.00	0.02	0.00	0.13	-0.03
netustm	0.02	1.00	-0.03	-0.25	-0.01
clmchnng	0.00	-0.03	1.00	0.04	-0.04
agea	0.13	-0.25	0.04	1.00	0.04
gndr	-0.03	-0.01	-0.04	0.04	1.00

Table 2.2: Correlation Matrix

From the correlation table: 2.2, it can be said that there is no significant relation between the *netustm* & *nwspol* with any of the variables related to climate change.

2.2.3 Correlation between two freely selectable variables (*agea* with *netustm*)

From the table: 2.2, a significant negative correlation has been seen between *agea* and *netustm*. So this section will analyze this correlation a little deeper. At first, a scatter plot is represented in the figure: 2.5 to see the relationship between the internet use time of the respondents with their age. From this figure, it can be said that long-time internet users are younger and the time of internet use is reduced with their age. So it indicates a negative correlation between age and internet use time in a typical day.

To validate the previous findings about the negative correlation between *agea* and *netustm*, the respondents of *netustm* variable are categorized into 7 categories: (a) less than 10min (b) 11min to 30min (c) 31min to 1hour (d) 1h to 2h (e) 2h to 3h (f) 3h to 4h and (g) more than 4h. Now a bar plot is illustrated in the figure: 2.6 which shows the percentage of respondents of different

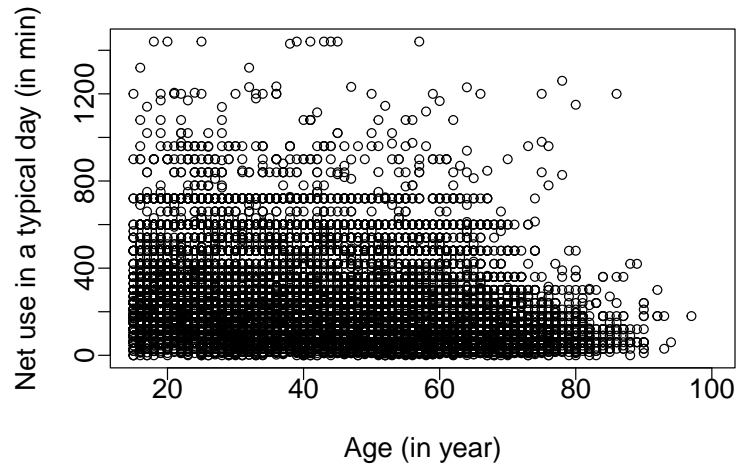


Figure 2.5: Net use in a typical day as a function of Age of the respondents

categories corresponding to their age. The bar chart shows that the percentage of internet users of all categories more than 3h is being decreased with age. On the other hand, the percentage of internet users of all categories less than 2h is increasing. So this bar plot also validates the negative correlation between *netustm* and *agea*.

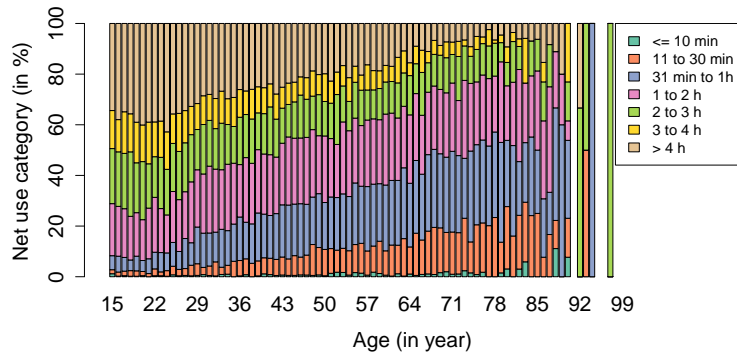


Figure 2.6: The percentage of different categories of net use in a typical day as a function of the age of the respondents