# COVID-19 Disease: mental health implications on international students in Ireland

Christopher Herrera Magana - x19127723[1]

[1]Cloud Competency Centre, National College of Ireland, Dublin, Ireland
[1]Total number of words ≈ 5700 August 28, 2020

***Abstract*—This study examines if there is an association between international English students' and workers' living conditions and the influence on mental well-being during this pandemic in Ireland. Three datasets were utilized to compare their cognitive behavior towards the number of sharing a house using a chi-square statistical analysis test to address the topic. The results infer that the chosen sample did not provide adequate data to assume that the incidence of mental health problems is related to the number of individuals who share a home. This research accentuates the importance of ethical concerns and the handling of data privacy following the GDPR.**

***Keywords*—MapReduce, Spark, Python, VMware, Chi-square, Hypothesis testing, Data Analytics.**

## I. Introduction

After roughly five months of the Covid-19 disease outbreak in Ireland, the current social distancing measures that have been implemented to reduce the spread of the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) have indeed flattened the curve of daily cases [1], [2]. As a consequence, the restrictions have increased mental health conditions such as depression, stress, and anxiety [3]–[5]. The side effects of relying on isolation techniques can be critical for patients that already have a physical or health condition [5], [6]. Similarly, due to the social distancing measures, the school's traditional teaching system has been changed to a virtual platform (online-classes) [7].

Being a student in the time of a pandemic emphasizes an already stressful phase of life full of relentless deadlines, hours, and continuous learning. A report declared that the transition from face-to-face to online classes puts pressure on teachers to adapt their classes(extra-work)—whereas for the student's it generates a challenge to cope with the deadlines and teaching method [8]–[10]. Moreover, online teaching-platform does not guarantee the same level of learning compared to traditional (face-to-face) courses [8]–[10]. One the one hand, it relies on how well the student can adapt to self-learning and take advantage and advance with other assignments [10], [11]. On the other hand, students might struggle with challenges out of their control, such as reduced learning/living environment, difficult topics and assignments, and the lack of interaction with peers, causing more stress and anxiety around failing assignments or deferral. [10], [12], [13].

As previously mentioned, the measures taken to reduce the pandemic (COVID-19 disease) have been successful [2]. Nevertheless, it has caused a domino effect that influences all sectors in Ireland [5], [14], [15]. In the education field, international English students have been affected by the closure of schools, resulting in arduous complaints to get a refund or lost it [13], [16]. Likewise, the hospitality sector, which employs many students, has temporary cessation of activities to avoid the increase of COVID-19 disease cases, putting more stress on international students to maintain themselves [15], [16].

The situation in Ireland regarding house-crisis has no changed, and it has represented an obstacle for the majority of citizens, but it affects international students [16]–[19]. It is a common practice for international students to live with a high number of people in houses meant for fewer. There are cases where international students decided to share a house with up 24 people due to difficulty finding a place [16], [19]. In other cases, the price plays a significant role in accepting those conditions [16], [19]. Sharing a room can become terrifying due to the lack of privacy, sleep, ability to focus, and currently, the potential threat of Covid-19 due to the inability to follow social distancing restrictions in tight living quarters. Consequently, all that could contribute to mental health conditions such as depression, stress, and anxiety [20].

Despite Ireland's government efforts to support (economically) people affected due to the pandemic, international students have suffered from the lack of home-sharing regulations [17], [20]. With the sudden increase of Covid-19 disease cases in Ireland, sharing a house/-room with many people could represent a threat [21]. Therefore, This research examines if there is a link between mental health and the number of other people international students live with during these times of a pandemic. Hopefully, results could provide insights into if there is a need for a different approach with social-distancing measures to enhance mental well-being in the housing-crisis in Ireland.

### *A. Statement of the problem*

This article's fundamental purpose is to statistically analyze if there is an association between international students' in Ireland's mental well-being with the number of people they share accommodation (Chi-square hypothesis test) during the Covid-19 outbreak.

### *B. Research Question*

Is there a (statistically) significant correlation linking the increase of mental well-being (depression on anxiety) and the number of people sharing a house concerning international students in Ireland during the pandemic?

### *C. Significance of Study*

This article provides a clear representation of how international students living in Ireland (housing conditions) are coping with the COVID-19 disease restrictions that might impact their mental health. By sharing this research, governmental and educational institutions could identify the negatives/positives effects of sharing the house depending on the number of inhabitants developing mental health conditions. Thus, the government can provide different measures based on the results obtained that could lead to manageable COVID-19 disease cases and better mental well-being for close-quarter housing.

### *D. Scope and Limitations of the Study*

The research was limited to international English students' and workers' opinions about the challenges they have experienced due to the COVID-19 disease restrictions in Ireland. Additionally, it was influenced by actions proposed by the government to assist the people affected financially (COVID-19 Pandemic Unemployment Payment). For this research, the cooperation of Fiachra' OLuain was required, who is the co-founder of the English Language Student's Union of Ireland (ELSU) movement, that aims to help the international English students to express their demands during this pandemic. Fiachra' OLuain provided two datasets (surveys) received by the 21st of June, 2020— this dataset follows ethical considerations and GPDR regulations.

Some limitations were observed in this study. First, the two datasets obtained from Fiachra were not sufficient to be considered big data (large volume of data– structured and unstructured) to produce a significant data analysis. Therefore, it was necessary to use another dataset to overcome this issue; [22] the dataset retrieved from the Nexoid project satisfied the big data characteristics but does not have related observations concerning international students' mental health or living accommodations. Second, the hardware capabilities used on a virtual machine (VMware) environment to carry out the data analysis were insufficient, leading to a lack of memory. However, it was possible to overcome this limitation by importing the dataset (Nexoid) using pandas function (read_csv) rather than importing from the MongoDB database.

This study is organized as follows: Section II. provides a general background of COVID-19 disease in Ireland, the social-distancing measures (mental health), the impact regarding international English students with house-crisis. Section III outlines the data analytics processes, MapReduce patterns, programming languages, and tools used for this research. Section IV analyzes the statistical hypothesis test results, and the obstacles proceeded with it—section V. outcomes, lessons learned, and the future work of the study. The appendix section illustrates the three dataset's insights in correlation with this study.

## II. Literature Review

To better appreciate the consequences of Covid-19 on international student's mental health, this study defines mental health [20]. As described, mental health is a psychological state in which a person understands his strengths, can endure life demands, performs productively, and contribute towards society [23]. Additionally, two mental health conditions critical in this study are depression and anxiety on students [13], [20]. Depression has a detrimental effect on the way a person feels, thoughts, and behaviors [24]. There are symptoms correlated to the depression condition like sadness, discourage moods, loose of appetite, or interest [24]. The latter condition is characterized by stress, concerned thoughts, and physical changes, such as elevated blood pressure [25].

In Ireland, many challenges arose after published the road-map (initially five phases) for lifting the COVID-19 restrictions slowly [26]. First, the closure of businesses caused anxiety in the population due to job losses or uncertainty in returning/getting a new job [26], [27]. Second, the actions proposed by the Health Service Executive (HSE) to wear personal protective equipment (gloves or masks) help to limit the spread of the virus to others in public spaces [28]. Third, the travel zone restriction from one's home (2km to 20km) and limit external interaction greatly influence depression and anxiety [5], [5], [26]. However, the measures taken contributed to control the spread of the virus despite causing stress, depression, and anxiety for the population.

One study presented that Irish citizens of 65+ years were the most affected by anxiety [4]. Others (one-in-four) suffered depression or anxiety during the first week of the lockdown [4], [29]. In contrast to another study in the UK regarding self-isolation restrictions, the most affected age-range between 24-34 years presented poor mental health (depression or anxiety) based on their measures (statistical test)— due to lower-income or physical problems (multi-morbidity) [30].

For students and international students, in particular, many situations caused mental health conditions [13]. First, online teaching (self-learning) caused anxiety and stress due to the traditional (face-to-face) teaching method's abrupt change [10], [11], [31]. Another situation was the broadband connection to access online-classes was not adequate, causing strain on the students [8]. The fear of failing assignments caused anxiety on students and also affecting student motivation [13], [31]. Despite the institution offer guidance on mental health conditions, some students prefer not to mention their mental health conditions because it may trigger a sense of inferiority [13], [31].

The actual situation in Ireland regarding the housing crisis is essential to improve. For instance, after the high cost of third-level education in Ireland for international students, other challenges are finding affordable accommodation that can be used for studying and living [12], [16]–[19]. As mentioned by [18], the price of on-campus accommodation for third level education keeps increasing. In contrast, some international English students did not get a refund due to the pandemic outbreak, and the situation gets worse to find accommodation [16], [19], [32]. Therefore, Ireland requires to change policy to fix the housing crisis to avoid further mental health damage on international students [32].

The household's current situation in Ireland can become a risk shortly due to the lack of measures on-place due to the sudden increase of COVID-19 disease cases in Ireland. As previously mentioned, many students share accommodation, the conditions where students' life are not adequate due to the limited room space [16], [19], [32]. Interestingly, [6] claimed that there is a relationship with the household conditions (limited space) to develop mental health conditions during this pandemic. Also, it described that not following social-distancing measures on household exposes a risk on COVID-19 cases [6].

To the best of our knowledge, recent studies have explained how Covid-19 restrictions have affected mental health with the use of isolation (self-quarantine) and social distancing measures in different fields. However, there are zero articles associated with the burden of international (English) students in Ireland finding affordable and adequate accommodation that could not impact on their mental health due to Covid-19 disease. Hence, this research intends to fill the literature gap in the actual situation regarding international English students associated with the house living conditions during this pandemic— encouraging an adequate number of people sharing a house/room to avoid the existing challenges (mental health and increase of COVID-19 cases).

## III. Research Methods & Specification

This study aims to examine if international students' living conditions are adequate or if the number of people they live with could affect their mental well-being. The research also relates to the spread of Covid-19 disease and an increase in cases where isolation measures cannot be observed in homes, representing another challenge to be controlled in the near future. Therefore, the study consists of a quantitative methodology approach, using different programming languages and data analysis tools that will be described in the section below.

### *A. Design method*

The research proposal consists of three stages. The first step involves all the configuration of software, tools, libraries, and the virtual environment required to complete this project. Second, it will continue by processing and analyzing the datasets with the use of

MapReduce patterns to facilitate observations and patterns that will be required for further interpretation. After processing the data, it will be stored using a SQL database (MySQL). In the third stage, the data storage on MySQL will be used in statistical analysis (hypothesis test). It will provide visual insights into different interpretations: school payment, government support, homesharing, mental health conditions, English level, and international student home countries. All these aspects are critical, which will provide a better comprehension of the circumstances that international students face and the consequences.

*1) Phase one:* Before proceeding with this stage's description, it is recommended to define more in-depth virtual machines (VM) and the technology behind them. Per definition, ""virtualization is a technology that enables us to create or simulate (virtual) computing environment [33]"". Similarly, a virtual machine (VM) refers to an image that works as a real computer. In other words, a VM is a virtual world, which acts as a virtual computing device with its CPU, memory, network, and storage, built on a real hardware system (off-or on-site) [34].

During this process, it was necessary to install and configure a virtual machine using VMware. The operating system for this VM is a distro of Linux—Ubuntu 18.04 image. It was downloaded from the official website and configured assigning 5 GB of memory RAM, 50GB of hard disk, four processors, and host-only network settings. After the VM was ready to use, the installation of Apache Spark and Jupyter notebook proceeded according to a web-page guide manual installation [35], [36]. All the set-up configuration is the foundation for installing the database (MongoDB and MYSQL). In the case of MongoDB, the installation also included the Robo 3T (Robomongo) tool for the database's graphical user interface(GUI) [37]. Subsequently, the completion of MySQL workbench [38] and all the settings needed to enable the research Python application and the connectivity with both databases through Python DB API.

*2) Phase two:* MongoDB (non-relational database), will contain the three raw datasets to carry out the analysis. First, using Jupyter notebook with Python programming language permits reading the raw datasets to conduct manipulation and examination of the data. While cleaning and organizing the raw datasets, it is essential to distinguish significant variables, either categorical, ordinal, or numerical. As previously mentioned, the variables represent meaningful information about international students' lives in Dublin, Ireland. It is imperative to validate what type of variables are required to extract those pieces of information. Therefore, MapReduce (framework) could contribute to reducing the burden on processing data (big data) in a parallelizable manner [39].

*3) Phase three:* In the third stage, after the process of cleaning and organizing the data into different tables using diverse MapReduce patterns, each table will contain a specific segment of three datasets transmitted to the MySQL database. From MySQL, the exporting process begins, and R programming language and Tableau software facilitate the visual representation of the data and their insights that this study aims to produce. Moreover, before carrying out the statistical analysis, some requirements (conditions/assumptions) are recommended to present a suitable hypothesis test analysis.

### *B. MapReduce Design Patterns*

MapReduce design patterns serve as a template to solve generic issues with the benefits of MapReduce data manipulation [39]. For this research, it will be required to use five MapReduce patterns: MapReduce (Key: Value), Summarization, Join (Merge), Filtering, and Data organization.

*1) Map Pattern:* This step intends to group data into a key/value pair representation [39]. In the three datasets, the variable type is categorical (nominal), where there is no rank or order between the countries. Consequently, it will show a relationship country (key) with a value pair.

*2) Reducer Pattern:* The reducer iterates through all the key/value pairs associated with simplifying the groups [39]. In other words, it will count the number of values with the same key. For instance, the three datasets' outcome is the total number of students/workers associated with a country.

```
countries = db_student.groupby(['country']).groups.
    keys() # Key = Countries

list_country = list(countries)

my_dict = {} # Dictionary total number of students
    per country

for x in range(0,len(list_country)): # iterate
    through the total number of countries

    counter = 0

    for i in range(0,len(db_student.iloc[:,0])): #
        iterate through all the rows per Country

        value = list_country[x]

        if i < 1077: # 1077 total number of row
            values
```

```
            if value == (db_student['country'].iloc[
                i]): #

                counter = counter + 1 # count the
                    number of students per country

        elif i == 1077: #1077

            my_dict.__setitem__(value,counter) #
                updating the dictionary Key:Value =
                Country: Total_number_of_students

            #my_dict

        else:
            pass
```

*3) Filtering Pattern:* The filtering excludes unwanted parts of the data, but it does not change the data; it just takes a subset of the data to understand smaller and essential pieces of the data to analyze [39]. For this study, several sections (Gender, Age, accommodation, and college payment ) will be filtered to proceed with deeper insights into the data— data cleansing.

*4) Summarization Pattern:* The summarization is used to express the data's different operations such as statistical, calculation, indexing, or counting in an analytical way [39]. For instance, it can show the total amount a student pays for the English course in Ireland or the total number of students per country.

```
table_student_homesharing = db_student.groupby('
    country')['homeshare'].value_counts().unstack().
    fillna(0)
```

*5) Data Organization Pattern:* This MapReduce pattern is about ordering the data depending on the different elements (variables) used for either visual representation or further data analysis (statistics) [39]. In this case, there will be several tables that contain significant value to advance with a hypothesis test.

### C. Merge Pattern

```
table_student_cost_edu = pd.merge(db_country, db_
    student_schoolCost,how='outer', on='country',
    indicator=True)
```

### D. Source of Data

Two of three datasets were obtained from Fiachra 'O Luain (ELSU), were created using two questionnaires. The first questionnaire focuses on international English students and workers in Ireland and their challenges during this Covid-19 disease— it consisted of 31 questions. Seventeen questions have three to five options classified as a categorical variable (nominal variable type) [40]. Eleven questions are open responses that are modified (cleaned and reduced) and classified as a nominal variable type except for the cost of education (discrete variable type) [40]. Finally, the remaining three are questions are dates that are interpreted as the continuous variable type [40]. The second dataset is a follow-up that intents to check the progress of the situation towards international English workers in Ireland for the second survey— made up of 33 questions. Fourteen questions (three options) are the variable type of nominal, seventeen binary types (yes/no), ten are open response as a nominal variable, one as a continuous variable (date), and the last one (multiple options) nominal variable type [40].

The third dataset was retrieved from the Nexoid research project by the 20th of July. Nexoid is a tech firm, with a data analysis team dedicated to developing a dataset focus on COVID-19 [41]. Despite their expertise in science and information, the company is not a pharmacy practitioner. Nexoid's data focuses on the risk involved with Covid-19 around different parts of the world. Therefore, their app, Survive COVID-19 Calculator, provides insights into the different data patterns to provide the best COVID-19 predictions [41]. The calculator app provides a web survey made up of 31 questions of multiple options. It is vital to mention that the Survival COVID-19 Calculator was created using two different equations; chance of contamination and the death rate after being infected [41]. Seven hundred seventy-five thousand rows and 60 columns form the master_dataset. This dataset has several categorical variable types that are represented into the categories of nominal and binary types. For this research, specific values such as age, gender, house_count and their physical health condition will be selected to emphasize Ireland's actual situation regarding Covid-19 disease.

### E. Sample Selection

Before explaining the process of selecting the values (variables) to analyze, there are critical aspects to consider. Firstly, the surveys managed by Fiachra (ELSU) might contain missing values, errors in the data due to the lack of English comprehension by students, and limited area coverage, mostly in Dublin. Besides, the mental health conditions responses were not well defined by the participants as a consequence; this study took into consideration (depression or anxiety) to observe if there is a change in mental well-being. Secondly, Nexoid datasets are more restrictive in response options, just having one open-ended question. However, their questionnaire is based on opinions and assumptions of

the responders, potentially resulting in errors or biased interpretations that can impact this study's outcome.

The sample selection represents a subset of the population of the datasets used. However, due to the datasets' conditions and characteristics, it was a challenge in the manipulation of the data to select an appropriate sample for the statistical analysis. Moreover, just one dataset was classified as big data, whereas the other two are a smaller sample of respondents. Despite all these obstacles, two different actions were used to solve this dilemma. The first measure was to combined two datasets; one is the workers' dataset (MapReduce output) with particular variables (mental_health_household and Home_share) the other is randomly created to have similar variables related to the first one. The second measure was to use a random Python function to select a specific number of variables for the hypothesis test( random sampling selection). These considerations outline the research population made-up of one dataset with 1,776 cases (Mental_health_household and Home_share) for the statistical analysis. A compilation of the three datasetC (specific observations) shows the data's insights regarding international English students', workers and people affected due to COVID-19 disease (restrictions or infection) in Ireland. Thus the sample population randomly selected for the statistical analysis was about 1,111 cases (mental well-being).

### *F. Data Analysis*

The following step involves the use of MapReduce patterns and Python programming libraries to analyze and extract parts of the data. Before proceeding to extract variables is recommended to check and modify the type of the variable. As previously mentioned, the variable's type is related to the field of statistics. However, those variables are different in the field of programming. Some variables can be represented as an object, string, int, float, chart, and Boolean. The following illustration shows the variable type before and after cleaning the data. PICTURE

The second part of the data analysis conveys a visual representation of international English students in Ireland. There are up to seven factors that are critical in the life of those students. Therefore, the following illustrations show the student's burden and how this will impact mental health conditions while sharing a home with a different number of people. Similarly, what measures might be necessary to reduce those challenges and the spread of the Covid-19 disease.

## IV. HYPOTHESIS TESTING

Before proceeding with the hypothesis test, it was vital to select which statistics test is adequate for the hypothesis research; and what requirements/conditions need to be satisfied. The first step is to know the type of variables the study is using. In this project, the two variables are categorical type (nominal). The reason for this is to determine if there is a significant relationship between mental health (depression) with the number of people live with regarding international English students. This verifies if the data follows a normal distribution (parametric test) or distribution-free (non-parametric test). These are defined by the following, "A parametric test is based on assumptions about the distribution of population from which the sample was taken. Non-parametric statistics are not based on assumptions; that is, the data can be collected from a sample that does not follow a specific distribution [42]."

The non-parametric test has three conditions regarding the data for it uses; However, at least one condition would be sufficient to proceed with is used [43].

- Condition 1: The variables should be categorical type, either nominal or ordinal type [43].
- Condition 2: "The sample sizes of the study groups are unequal [43]."
- Condition 3: The data variables interval or ratio type contradicts one of the parametric test assumptions [43].
  - A: "The distribution of the data was seriously skewed or kurtotic [43]."
  - B: "The data violate the assumptions of equal variance or homoscedasticity [43]."
  - C: "the continuous data were collapsed into a small number of categories, and thus the data are no longer interval or ratio [43]"

In this study, the following tests were introduced on the data to confirm and validate the conditions needed to use a non-parametric test. First, a descriptive statistics (Fig. 1) test summarizes the data that provides a basic understanding of the patterns regarding their measures and distribution (skewed or kurtotic) [44]. The second test is a visual representation of the data as a histogram plot. The third test is the Shapiro-Wilk test which aims to check if the variables or groups come from a normal distribution. The results obtained from this test could provide enough evidence to select the non-parametric test. To have a better perception about the aforementioned, in the appendix section (A) the images show that the conditions to use the non-parametric test were satisfy

. Therefore, this research used the chi-square test of the independence hypothesis test ($\chi^2$ for Independence).

### A. Assumptions of the Chi-square Test for Independence

It was required to satisfies the following six assumptions on $\chi^2$ to present an accurate outcome.

*1) Assumption#1:* The data should be interpreted in a frequency table.

As it can be observed on (Fig.1) illustrates the dataset's frequency table used for the hypothesis test. Therefore, the evidence provided certifies this assumption.

*2) Assumption#2:* The variable's type (categorical) are mutually exclusive [45].

As previously mentioned in the literature reviewed, some studies claimed that household living conditions could contribute to developing mental health conditions [6]. However, there is no significant evidence that both events can happen at the same time [6]. There might be cases where people previously had mental health conditions and got worse sharing with other people [3], [46]. Similarly, there might be situations that help people from isolation in a friendly house environment or physical activities, reducing mental health impacts [3], [46]. Hence, this study satisfies the assumption regarding variables (Mental Health and Sharing Room people) that are mutually exclusive.

*3) Assumption#3:* "Each subject may contribute data to one and only one cell in the $\chi^2$ [45]."

The datasets obtained from Fiachra Ó Luain (ELSU) and download from Nexoid (web-application COVID-19 survival calculator) , which both used surveys, might have deleted duplicated user's responses. Thus, this might provide enough evidence to suffice this assumption.

*4) Assumption#4:* "The study groups must be independent [45]."

The datasets students and workers satisfy this assumption where there is no relationship between the observations in each group. In the student's dataset, the observation for mental well-being was generated randomly. Whereas, in the worker's dataset, it includes those two variables (observations).

*5) Assumption#5:* "There are 2 variables, and both are measured as categories, usually at the nominal level [45]."

As previously mentioned, the two variables are the categorical type. The depression variable is a binary type, and the number of people sharing a room is a nominal type. Accordingly, this study fills the assumption.

*6) Assumption#6:* The predicted frequency count for every cell in the contingency table (sample date) should be at least five [47].

The dataset (random sampling) used for the statistical analysis provides enough count frequency in the contingency table to meet this assumption's requirements. The Fig.1 certifies this assumption.

```
> freq(sample_selection)
Frequencies
sample_selection$mental_health
Type: Character

             Freq   % Valid   % Valid Cum.   % Total   % Total Cum.
---------- ------ --------- -------------- --------- --------------
        No    456     41.04          41.04     41.04          41.04
       Yes    655     58.96         100.00     58.96         100.00
     <NA>      0                                0.00         100.00
     Total   1111    100.00         100.00    100.00         100.00

sample_selection$home_sharing
Type: Numeric

             Freq   % Valid   % Valid Cum.   % Total   % Total Cum.
---------- ------ --------- -------------- --------- --------------
         1    110      9.90           9.90      9.90           9.90
         2    132     11.88          21.78     11.88          21.78
         3    145     13.05          34.83     13.05          34.83
         4    157     14.13          48.96     14.13          48.96
         5    116     10.44          59.41     10.44          59.41
         6    135     12.15          71.56     12.15          71.56
         7    114     10.26          81.82     10.26          81.82
         8     86      7.74          89.56      7.74          89.56
         9     54      4.86          94.42      4.86          94.42
        10     62      5.58         100.00      5.58         100.00
     <NA>      0                                0.00         100.00
     Total   1111    100.00         100.00    100.00         100.00
```

Fig. 1. Distribution of frequencies

Overall, the six assumptions before mentioned are significant to obtained adequate results. Hence, this study considered that the chi-square test could be affected by the evidence provided to suffice the assumptions.

### B. Stating Hypothesis

The hypothesis test for this study are as follows:

The null hypothesis ($H_0$): There is no (statistically) connection between international English student's men-

tal health conditions (depression or anxiety) with the number of people sharing a house.

$$H_0 : \rho_{MentalHealth} = \rho_{SharingHouse}$$

The alternative hypothesis ($H_a$): There is a (statistically) connection between international English student's mental health conditions (depression or anxiety) with the number of people sharing a house.

$$H_a : \rho_{MentalHealth} \neq \rho_{SharingHouse}$$

### C. The significance level of alpha

As mentioned by "Fisher saw the P-value as an index measuring the strength of evidence against the null hypothesis. He advocated ($\alpha$) $<$ 0.05 (5%significance) as a standard level for concluding that there is evidence against a hypothesis [48]." Hence, this study's selected significance value is 0.05 showing that it could implicate the probability of rejecting the null hypothesis when it is true (type 1 error) [49].

### D. Chi-square statistics analysis

Chi-square ($\chi^2$ "key-square") was used to investigate if there is any relationship between the categorical variables' distributions. In this study, the statistics analysis's categorical variables were mental_health (symptoms) and home_sharing. The first step was to compute a contingency table that detailed the frequency distribution of the observations (categorical variables) as it can be seen in the following image.

| | sample_selection$home_sharing | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| **sample_selection$mental_health** | **1** | **2** | **3** | **4** | **5** | **6** | **7** | **8** | **9** | **10** |
| **No** | 37 | 51 | 53 | 54 | 51 | 66 | 51 | 35 | 26 | 32 |
| **Yes** | 73 | 81 | 92 | 103 | 65 | 69 | 63 | 51 | 28 | 30 |

Fig. 2. Sample dataset contingency table

Moreover, the importance of the contingency table is to check if there is a difference between the observed frequencies' patterns with the expected frequencies— as the formula below shows the comparison [50].

$$\tilde{\chi_c}^2 = \sum \frac{(O_i - E_i)^2}{E_i}$$

Where:

- c is degrees of freedom
- O is Observed value(s)
- E is expected value(s)
- i is the "$ith$" position in the contingency table

Before to calculate the chi-square statistics it is required to obtained the expected frequency. As the following formula illustrates the procedure to calculate it [51].

$$E(r, c) = \frac{n(r) \times c(r)}{n}$$

Where:

- r is row in cell
- c is column in cell
- n is total values in the data

With the use of chisq.test function provided the expected frequencies values to obtain the chi-square statistic. This function also shows the degrees of freedom and the p-value that were used to obtain the critical value located on an X2 distribution table. The image below illustrates the previously mention.

```
> chisq.test(freq_table_sample_dataset)

        Pearson's Chi-squared test

data:  freq_table_sample_dataset
X-squared = 15.366, df = 9, p-value = 0.08136
```

Fig. 3. Chi-square statistics results

Before proceeding to make a decision based on this information, whether to accept or reject the null hypothesis, it is crucial to state the rule decision based on the chi-square value and critical value. The decision rule is as follows:

Reject $H_0$ if $\chi^2 \geq$ Critical value

Fail to reject $H_0$ if $\chi^2 \leq$ Critical value

The Fig.4 represents the chi-square distribution with 9 degrees of freedom, a p-value of 0.08, and a critical value of 16.92.

With a significance level of 0.05, this study fails to reject the null hypothesis due to the difference between the chi-square statistics' critical value and also the difference between the p-value and alpha. In other words, these results indicate that the sample selected did not provide enough evidence to conclude that there is a relationship between developing mental health conditions associated with the number of people sharing a house. At the same

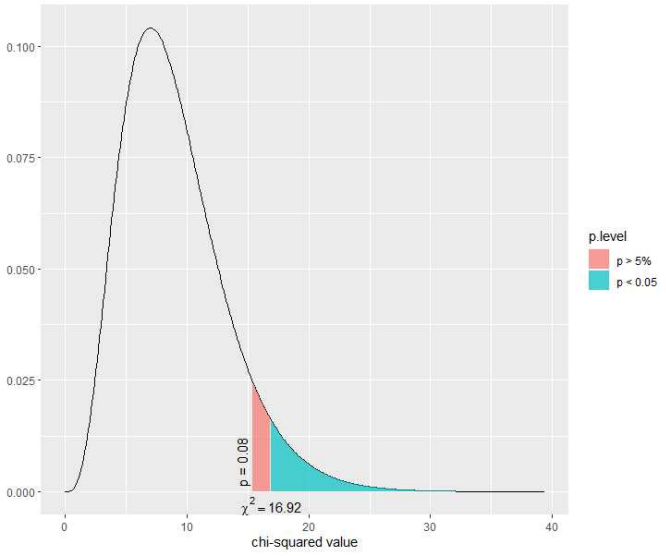


Fig. 4. Chi-square critical value

time, though, the lack of evidence does not prove that the relationship does not exist.

The following image illustrates the contribution per cell to the total chi-square statistics score. In the case of number of people sharing a house for one and six people were the one's that most contributed with roughly 40% for not cause an affectation on their mental well-being. Whereas, in the case of mental health affectation contribution were about 21%. Thus the results obtained account for most of the difference between expected and observed values.

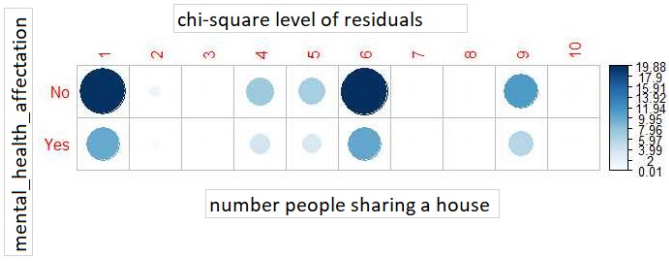


Fig. 5. Chi-square contribution per cell

Similarly as the Figure.9(B the expected frequency shows a slightly similar behaviour but not the same as the observed frequency. With this evidence it suggests that it is required to gather more information and also to validate the procedure of cleaning data to avoid error or wrong interpretations on the results.

## V. Conclusions and Future work

This research provided different outlooks regarding international English students, workers, and people affected due to the Covid-19 disease in Ireland. With the sudden increase of Covid-19 cases in Ireland, it is imperative to analyze other segments of the population that can induce the virus's spread due to the limited living conditions. Similarly, the isolation and social-distancing measures taken to control and reduce the pandemic has affected the people's mental well-being. Therefore, this study could contribute to further examining if there is any relationship between the living conditions among international English students and workers that can cause a change in their mental health.

The results obtained showed all the factors where international English students and workers need to deal with strenuous situations due to the housing crisis in Ireland. Additionally, it is required to implement different approaches to regulate the number of people who share a house. This factor is generated due to the student visa restrictions, limiting to 20-hours work per week that cannot afford to pay more than they earn for a room with adequate living conditions [12], [16]. Despite the results obtained in the chi-square statistical analysis, it provided the foundation for future work. There was not sufficient data to reject the null hypothesis. Consequently, this could imply that if measures are not taken on time, it will cause another predicament in which might not be able to control due to the housing regulations and mortgage.

In the data analysis field, notwithstanding the challenges of analyzing and manipulating the data, it was required to use different programming languages (R and Python) to solve this issue. One of the most significant challenges was finding adequate datasets (considered big data) to proceed with a reasonable and reproducible statistical analysis. There were many aspects in the survey (open-question) that might affect the interpretation and the statistics results. Another obstacle was the hardware requirements to examine the datasets through Apache Spark. One advantage of using Apache Spark (Pyspark) was the MapReduce-like processing function, which reduces the burden of handling a massive amount of data (unstructured/structured) and repetitive operations. The MapReduce patters used (MapReduce Key: Value, Filtering, Summarization, and Merging) contributed to clean and organize the datasets in an efficient manner in order to carry out a statistical analysis. As previously mentioned, due to the difference in variables type on the three datasets, it was necessary to aggregate similar-

like (random sampling) values to visualize this study and compact (summarize) to represent meaningful insights.

This study's future work examines international English students, and workers continue with the use of Machine Learning to detect behavior. This justification is based on studies that suggest that people are afraid to express their mental well-being due to the fear of being treated differently [10]. Therefore, the use of artificial intelligence to detect facial expression or voice recognition patterns would help college counselors to take a diverse approach for the students to open.

### *A. Ethical considerations*

This research aims to provide insights into the international English student's/workers' mental well-being and the living conditions in Ireland during this pandemic (SARS-CoV-2). It does not reflect personal inclinations on the subject matter, neither to discriminate nor to criticize the measures taken in Ireland. The purpose is to share knowledge and minimize the gap regarding the mental consequences in the field of international English students in Ireland during COVID-19. Similarly, this study consolidates the use of open-source software and tools to encourage replication and future research direction. Furthermore, the information retrieved and the values presented are cited mainly using academic peer literature and reliable web sources. Finally, the contributions, interpretation, and analysis are considered unbiased; relying on personal knowledge and skills used in the completion of this research.

## REFERENCES


[1] C. Pope, "What science got wrong on covid-19: Masks, warning signs and transmission without symptoms." https://www.irishtimes.com/life-and-style/health-family/what-science-got-wrong-on-covid-19-masks-warning-signs-and-transmission-without-symptoms-1.4333309, 08 2020. (Accessed on 08/25/2020).

[2] C. Brennan, "Coronavirus ireland good news: Dr tony holohan says covid-19 curve has been flattened and no peak expected - irish mirror online." https://www.irishmirror.ie/news/irish-news/health-news/coronavirus-ireland-good-news-dr-21887090, 04 2020. (Accessed on 08/21/2020).

[3] L. Jacob, M. A. Tully, Y. Barnett, G. F. Lopez-Sanchez, L. Butler, F. Schuch, R. López-Bueno, D. McDermott, J. Firth, I. Grabovac, A. Yakkundi, N. Armstrong, T. Young, and L. Smith, "The relationship between physical activity and mental health in a sample of the uk public: A cross-sectional study during the implementation of covid-19 social distancing measures," *Mental Health and Physical Activity*, vol. 19, p. 100345, 2020. JCR Impact Factor: (1.797).

[4] P. Hyland, M. Shevlin, O. McBride, J. Murphy, T. Karatzias, R. P. Bentall, A. Martinez, and F. Vallières, "Anxiety and depression in the republic of ireland during the covid-19 pandemic," *Acta Psychiatrica Scandinavica*, vol. n/a, no. n/a. JCR Impact Factor: (4.694).

[5] A. Nolan, "The economic shock from covid-19 will also affect our health." https://www.irishtimes.com/opinion/the-economic-shock-from-covid-19-will-also-affect-our-health-1.4241070, 04 2020. (Accessed on 08/23/2020).

[6] J. Mikolai, K. Keenan, and H. Kulu, "Intersecting household level health and socio-economic vulnerabilities and the covid-19 crisis: An analysis from the uk," *SSM - Population Health*, p. 100628, 2020. JCR Impact Factor: (2.05).

[7] E. Carroll and S. McCoy, "School leader wellbeing and workload during covid-19 lockdown — education matters." https://educationmatters.ie/school-leader-wellbeing-workload/, 07 2020. (Accessed on 08/23/2020).

[8] U. Nations, "Education in ireland during covid-19 — united nations western europe." https://unric.org/en/education-in-irel and-during-covid-19/, 04 2020. (Accessed on 08/21/2020).

[9] I. Times, "Challenge of online learning all about creative collaboration." https://www.irishtimes.com/sponsored/innovation-partner-profiles/challenge-of-online-learning-all-about-creative-collaboration-1.4320894, 08 2020. (Accessed on 08/20/2020).

[10] N. Kemp and R. Grieve, "Face-to-face or face-to-screen? undergraduates' opinions and test performance in classroom vs. online learning," *Frontiers in psychology*, vol. 5, pp. 1278–1278, Nov 2014. JCR Impact Factor: (2.129).

[11] J. Cavanaugh and S. Jacquemin, "A large sample comparison of grade based student learning outcomes in online vs. face-to-face courses," *Online Learning*, vol. 19, no. 2, 2015.

[12] J. Casey, "Radical change in store for irish third-level over next five years due to covid-19 pandemic." https://www.irishexaminer.com/news/arid-30999421.html, 05 2020. (Accessed on 08/23/2020).

[13] A. Price, "Ahead - mental health at third level: What are the issues facing students and what can we do about them?." https://www.ahead.ie/journal/Mental-health-at-third-level-What-are-the-issues-facing-students-and-what-can-we-do-about-them. (Accessed on 08/22/2020).

[14] L. Molloy, "'we must now prepare for the mental health damage that covid-19 will do'." https://www.thejournal.ie/readme/coronavirus-mental-health-5-5091400-May2020/, 05 2020. (Accessed on 08/19/2020).

[15] PWC, "Covid-19 and the irish hospitality sector: Impact and options." https://www.pwc.ie/publications/2020/hospitality-sector-covid19-impact-and-options.pdf, 03 2020. (Accessed on 08/22/2020).

[16] S. Malekmian, "Union formed for english language students in ireland — hotpress." https://www.hotpress.com/culture/union-formed-english-language-students-ireland-22816058, 05 2020. (Accessed on 08/22/2020).

[17] R. Hearne, "Why fixing ireland's housing crisis requires a change of policy — maynooth university." https://www.maynoothuniversity.ie/research/spotlight-research/why-fixing-irelands-housing-crisis-requires-change-policy. (Accessed on 08/23/2020).

[18] G. Gataveckaite, "Revealed: How much your college charges for student accommodation as rents are hiked - independent.ie." https://www.independent.ie/irish-news/education/revealed-how-much-your-college-charges-for-student-accommodation-as-rents-are-hiked-38381820.html, 08 2019. (Accessed on 08/26/2020).

[19] G. MacNamee, "Student housing crisis: 'we have students who are sleeping in cars, on their friends' floors'." https://www.thejournal.ie/ucd-student-housing-5024454-Feb2020/, 02 2020. (Accessed on 08/22/2020).

[20] SPR, "Consequences of student mental health issues — suicide prevention resource center." https://www.sprc.org/colleges-universities/consequences/. (Accessed on 08/21/2020).

[21] F. Bowers, "Warning covid-19 could rise to 'unsustainable' level." https://www.rte.ie/news/2020/0819/1159989-covid19-ireland/, 08 2020. (Accessed on 08/24/2020).

[22] Nexoid, "Nexoid covid-19 dataset." https://www.covid19survivalcalculator.com/en/download. (Accessed on 08/15/2020).

[23] WHO, "Mental health: strengthening our response." https://www.who.int/news-room/fact-sheets/detail/mental-health-strengthening-our-response, 03 2018. (Accessed on 08/17/2020).

[24] APA, "What is depression?." https://www.psychiatry.org/patients-families/depression/what-is-depression. (Accessed on 08/23/2020).

[25] APA, "Anxiety." https://www.apa.org/topics/anxiety#:~:text=Anxiety$%$20is$%$20an$%$20emotion$%$20characterized,certain$%$20situations$%$20out$%$20of$%$20worry. (Accessed on 08/23/2020).

[26] D. of the Taoiseach, "gov.ie - government publishes roadmap to ease covid-19 restrictions and reopen ireland's society and economy." https://www.gov.ie/en/press-release/e5e599-government-publishes-roadmap-to-ease-covid-19-restrictions-and-reope/, 05 2020. (Accessed on 08/23/2020).

[27] C. Ruane, "Coronavirus: managing the return to work in ireland." https://www.pinsentmasons.com/out-law/guides/coronavirus-return-to-work-ireland, 06 2020. (Accessed on 08/23/2020).

[28] L. Kingston, "All you need to know about wearing face masks in public." https://www.rte.ie/brainstorm/2020/0421/1132931-face-masks-guide-coronavirus/, 06 2020. (Accessed on 08/26/2020).

[29] P. Hyland, "Covid-19 mental health survey by maynooth university and trinity college finds high rates of anxiety — maynooth university." https://www.maynoothuniversity.ie/research/spotlight-research/covid-19-mental-health-survey-maynooth-university-and-trinity-college-finds-high-rates-anxiety. (Accessed on 08/16/2020).

[30] L. Smith, L. Jacob, A. Yakkundi, D. McDermott, N. C. Armstrong, Y. Barnett, G. F. López-Sánchez, S. Martin, L. Butler, and M. A. Tully, "Correlates of symptoms of anxiety and depression and mental wellbeing associated with covid-19: a cross-sectional study of uk-based respondents," *Psychiatry Research*, vol. 291, p. 113138, 2020. JCR Impact Factor: (2.208).

[31] P. Sahu, "Closure of universities due to coronavirus disease 2019 (covid-19): Impact on education and mental health of students and academic staff," *Cureus*, vol. 12, pp. e7541–e7541, Apr 2020.

[32] R. Hearne, "Rory hearne: 'students' futures at risk as housing crisis takes toll on education and mental health' - independent.ie." https://www.independent.ie/opinion/comment/rory-hearne-students-futures-at-risk-as-housing-crisis-takes-toll-on-education-and-mental-health-38347777.html, 07 2019. (Accessed on 08/21/2020).

[33] M. Azure, "What is virtualization - definition — microsoft azure." https://azure.microsoft.com/en-us/overview/what-is-virtualization/. (Accessed on 08/19/2020).

[34] R. Hat, "What is a virtual machine (vm)?." https://www.redhat.com/en/topics/virtualization/what-is-a-virtual-machine. (Accessed on 08/15/2020).

[35] J. Mutai, "Install apache spark on ubuntu 20.04/18.04 & debian 10/9 — computingforgeeks." https://computingforgeeks.com/how-to-install-apache-spark-on-ubuntu-debian/, 05 2020. (Accessed on 08/15/2020).

[36] D. Malcom, "How to install jupyter notebook on ubuntu 18.04 vps or dedicated server — hostadvice." https://hostadvice.com/how-to/how-to-install-jupyter-notebook-on-ubuntu-18-04-vps-or-dedicated-server/, 11 2018. (Accessed on 08/16/2020).

[37] P. K. Das, "Install robomongo (robo 3t) on ubuntu 18.04 — techrunnr." https://www.techrunnr.com/install-robomongo-robo-3t-on-ubuntu-18-04/, 07 2019. (Accessed on 08/15/2020).

[38] linuxize, "How to install and use mysql workbench on ubuntu 18.04 — linuxize." https://linuxize.com/post/how-to-install-and-use-mysql-workbench-on-ubuntu-18-04/, 06 2019. (Accessed on 08/15/2020).

[39] D. Miner and A. Shook, *MapReduce design patterns*. Sebastopol, CA: O'Reilly, 2013. https://learning.oreilly.com/library/view/mapreduce-design-patterns/9781449341954/. (Accessed on 07/20/2020).

[40] R. Bevans, "Types of variables — definitions and easy examples." https://www.scribbr.com/methodology/types-of-variables/, 08 2020. (Accessed on 08/15/2020).

[41] Nexoid, "Covid-19 survival." https://www.covid19survivalcalculator.com/en/home. (Accessed on 08/15/2020).

[42] IBM, "Ibm knowledge center." https://www.ibm.com/support/knowledgecenter/en/SSFMBX/com.ibm.swg.im.dashdb.analytics.doc/doc/r_statistics.html. (Accessed on 08/14/2020).

[43] M. L. McHugh, "The chi-square test of independence," *Biochemia medica*, vol. 23, no. 2, pp. 143–149, 2013. JCR Impact Factor: (2.202).

[44] statistics laerd, "Understanding descriptive and inferential statistics — laerd statistics." https://statistics.laerd.com/statistical-guides/descriptive-inferential-statistics.php. (Accessed on 08/13/2020).

[45] datanovia, "Homogeneity of variance test in r: The definitive guide - datanovia." https://www.datanovia.com/en/lessons/homogeneity-of-variance-test-in-r/. (Accessed on 08/15/2020).

[46] D. L. Dawson and N. Golijani-Moghaddam, "Covid-19: Psychological flexibility, coping, mental health, and wellbeing in the uk during the pandemic," *Journal of Contextual Behavioral Science*, vol. 17, pp. 126 – 134, 2020. JCR Impact Factor: (1.890).

[47] stattrek, "Chi-square test of independence." https://stattrek.com/chi-square-test/independence.aspx. (Accessed on 08/15/2020).

[48] J. A. C. Sterne and G. D. Smith, "Sifting the evidence-what's wrong with significance tests?," *Physical Therapy*, vol. 81, pp. 1464–1469, Aug. 2001. JCR Impact Factor: (3.043).

[49] J. Frost, "Significance level - statistics by jim." https://statisticsbyjim.com/glossary/significance-level/. (Accessed on 08/26/2020).

[50] sampath kumar gajawada, "Chi-square test for feature selection in machine learning — by sampath kumar gajawada — towards data science." https://towardsdatascience.com/chi-square-test-for-feature-selection-in-machine-learning-206b1f0b8223, 10 2019. (Accessed on 08/27/2020).

[51] A. Hayes, "Chi-square statistic: How to calculate it / distribution - statistics how to." https://www.statisticshowto.com/probability-and-statistics/chi-square/, 03 2020. (Accessed on 08/27/2020).

## APPENDIX

### A. Conditions for the non-parametric test

As we can see from the three figures (Fig.6-Fig.8) the dataset does not show high skewness and the kurtosis is

Descriptive Statistics
sample_selection$home_sharing
N: 1111

| | home_sharing |
|---|---|
| Mean | 4.88 |
| Std.Dev | 2.59 |
| Min | 1.00 |
| Q1 | 3.00 |
| Median | 5.00 |
| Q3 | 7.00 |
| Max | 10.00 |
| MAD | 2.97 |
| IQR | 4.00 |
| CV | 0.53 |
| Skewness | 0.28 |
| SE.Skewness | 0.07 |
| Kurtosis | -0.90 |
| N.Valid | 1111.00 |
| Pct.Valid | 100.00 |

Fig. 6. Descriptive statistics of home_Sharing variable

Shapiro-Wilk normality test

data: sample_selection$home_sharing
W = 0.94862, p-value < 2.2e-16

Fig. 7. Shapiro-wilk test for normality

on the acceptable range value -3 to 3. For the Shapiro test the p-value is so small that it suggest that the data does not follow a normal distribution. Similarly, with the histogram the behaviour does not represent a bell curve. Therefore with this evidence the decision taken for the chi-square hypothesis test is supported by this evidence.

### *B. Chi-square statistics expected values*

### *C. Insights of the datasets post-MapReduce output*

The first variable to analyze is the cost of education and what are the difference between other countries and also how big is that difference.

The Figure.10 used a box-plot chart to analyze the distribution of cost of education among the different countries. The out-liners dot the distance comparing to the whiskers. Brazil and Mexico are willing to pay more 10,000 euros for their English education. Whereas, Argentina and Spain invested around 2,300 euros for their courses. However, this information does not provide more details about the duration of the course or if it includes accommodation.

The Figure.11 Analyze what is the normal distribution of sharing a house with other people. This information is relevant after consider the price student's pay for their education. The average among number of people living in a house is about four to eight people. Likewise, Brazil does is the most affected in living conditions in cases where sharing with 24 people.

These two observations generate an idea about how bad international English students living conditions are. In the Figure.12 illustrates the level of English among students and also if the students presented were experienced mental health conditions (anxiety or depression). This mosaic plot presented the difference between English language which might represent an issue for the students communication or comprehension. The students that complain or express their mental well-being affected were in the category of intermediate to upper/advanced level. Whereas, elementary people did not present any mental well-being affection. This premise can be compare with the image (Fig.13). The options to describe their feelings towards mental health were a bit extended. Countries like Brazil and Mexico were the ones that participated in all the options to express their mental well-being. Whereas other countries like South Korea or Russia were not experience any mental health condition.

The last two images (Fig.14 and Fig.15) show the difference between international students living conditions and national residents. In the Fig.14 there are few cases where people share a house with up to six people. Nevertheless, this information is significant due to the difference on jobs and salary as the Fig.15 illustrates. AS Rory Hearne [17] mentioned the house crisis in Ireland has affected everyone, professional sometimes cannot buy a house due to the mortgage conditions and the high cost of the houses [17]. As a result, international students which are working part-time can afford to spend half of their salary to be able to maintained themselves [18], [32].

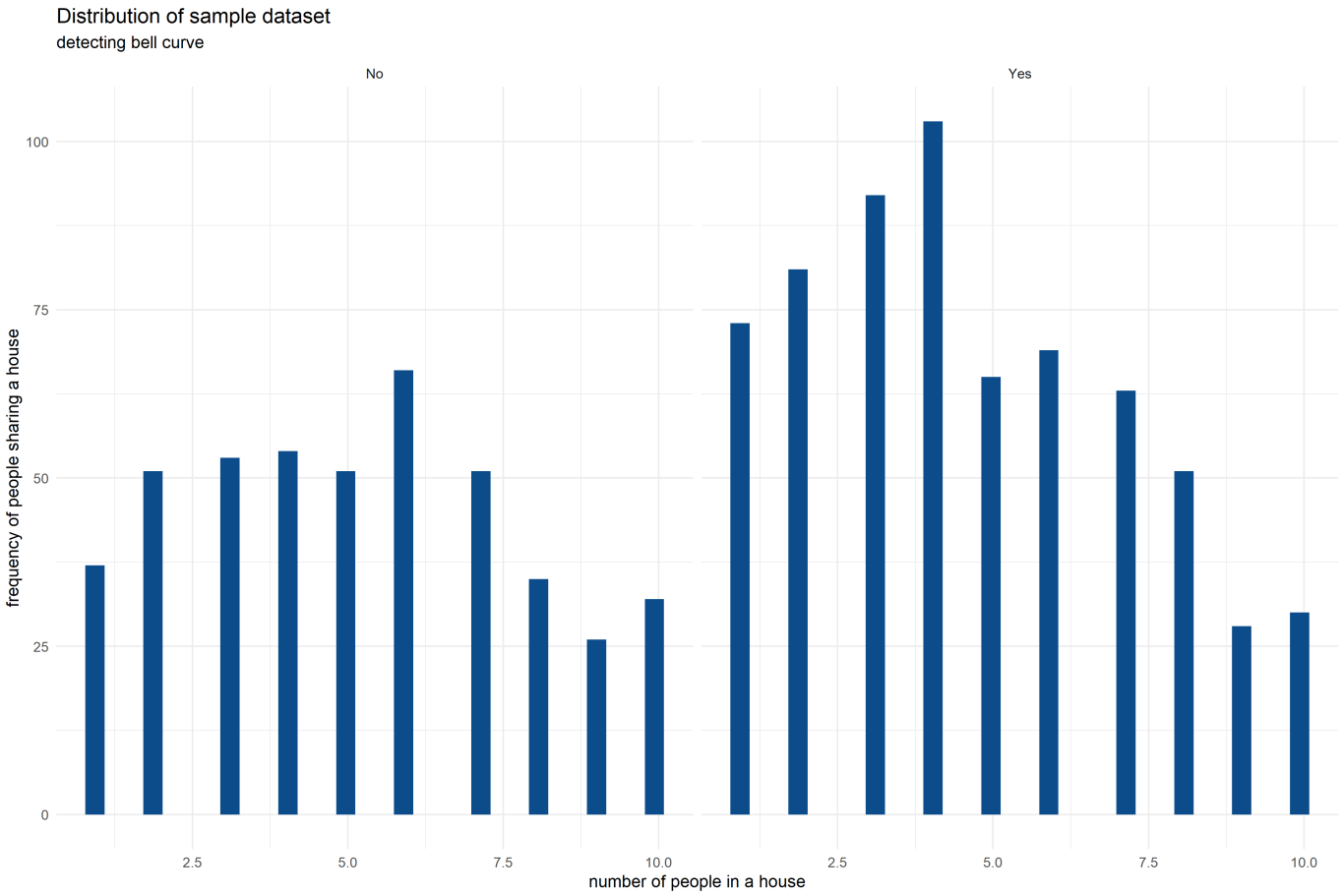


Fig. 8. Histogram of the sample dataset distribution.

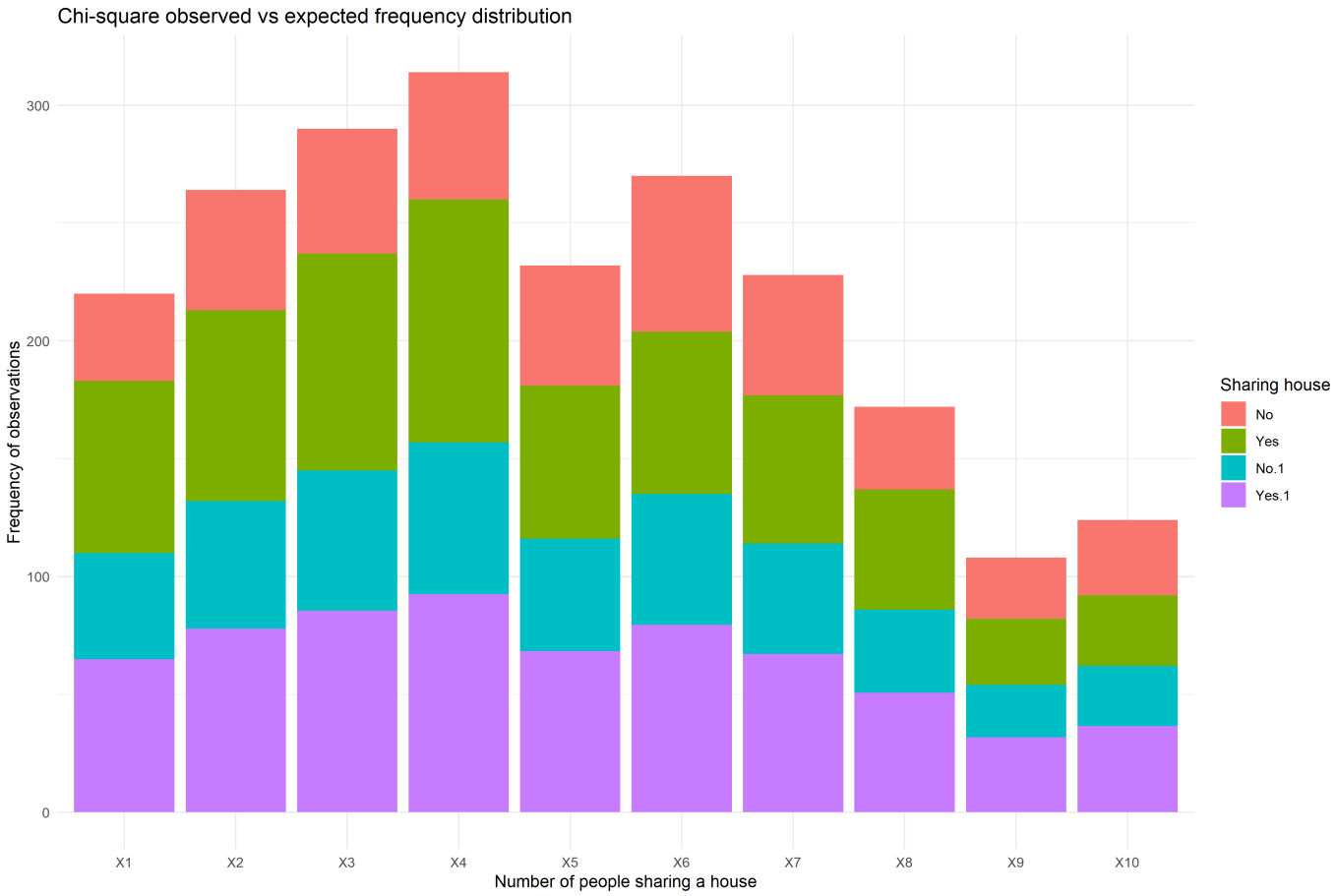


Fig. 9. Comparison between observed vs expected values chi-square

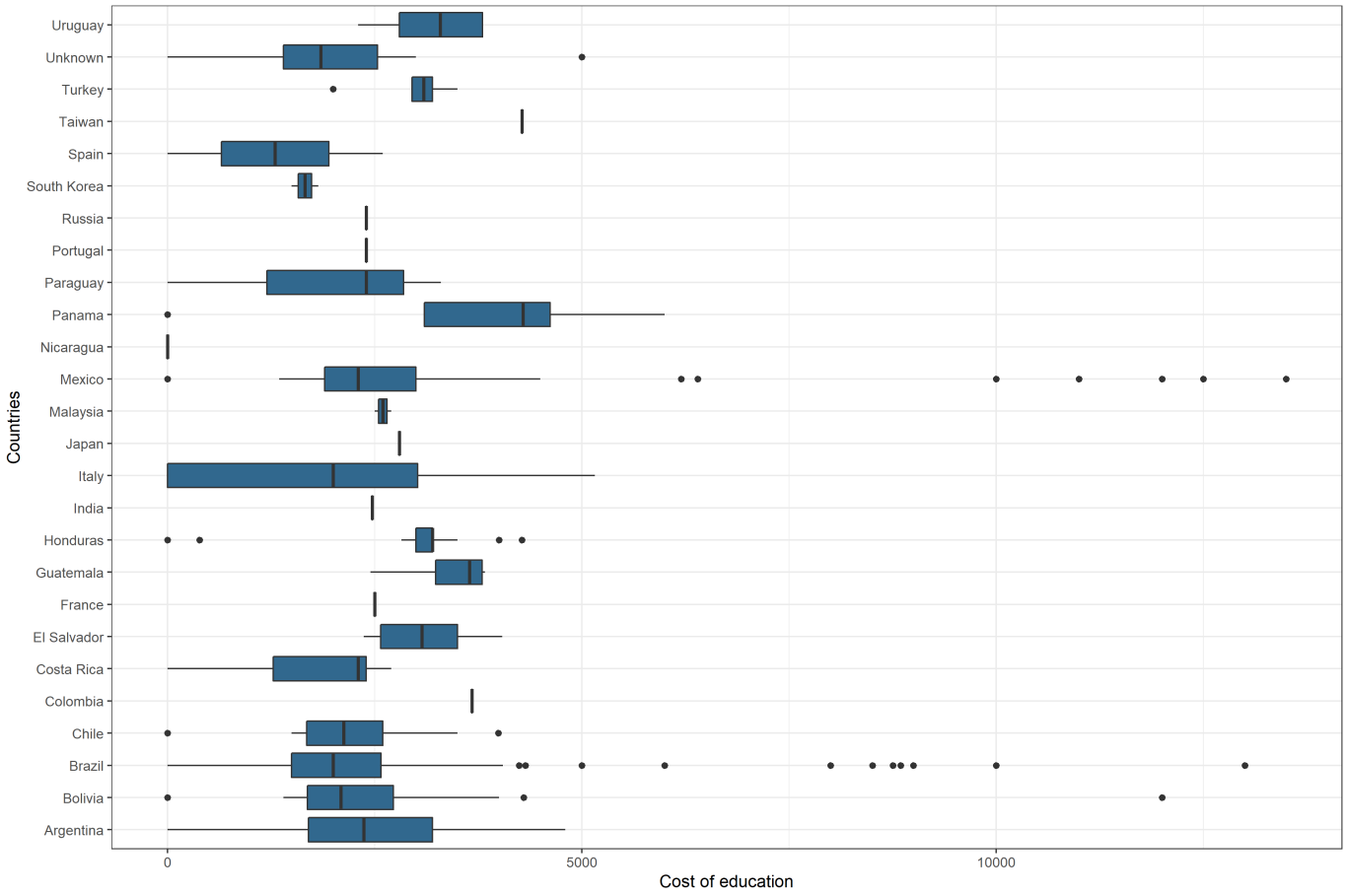


Fig. 10. Comparison between the cost of education per country

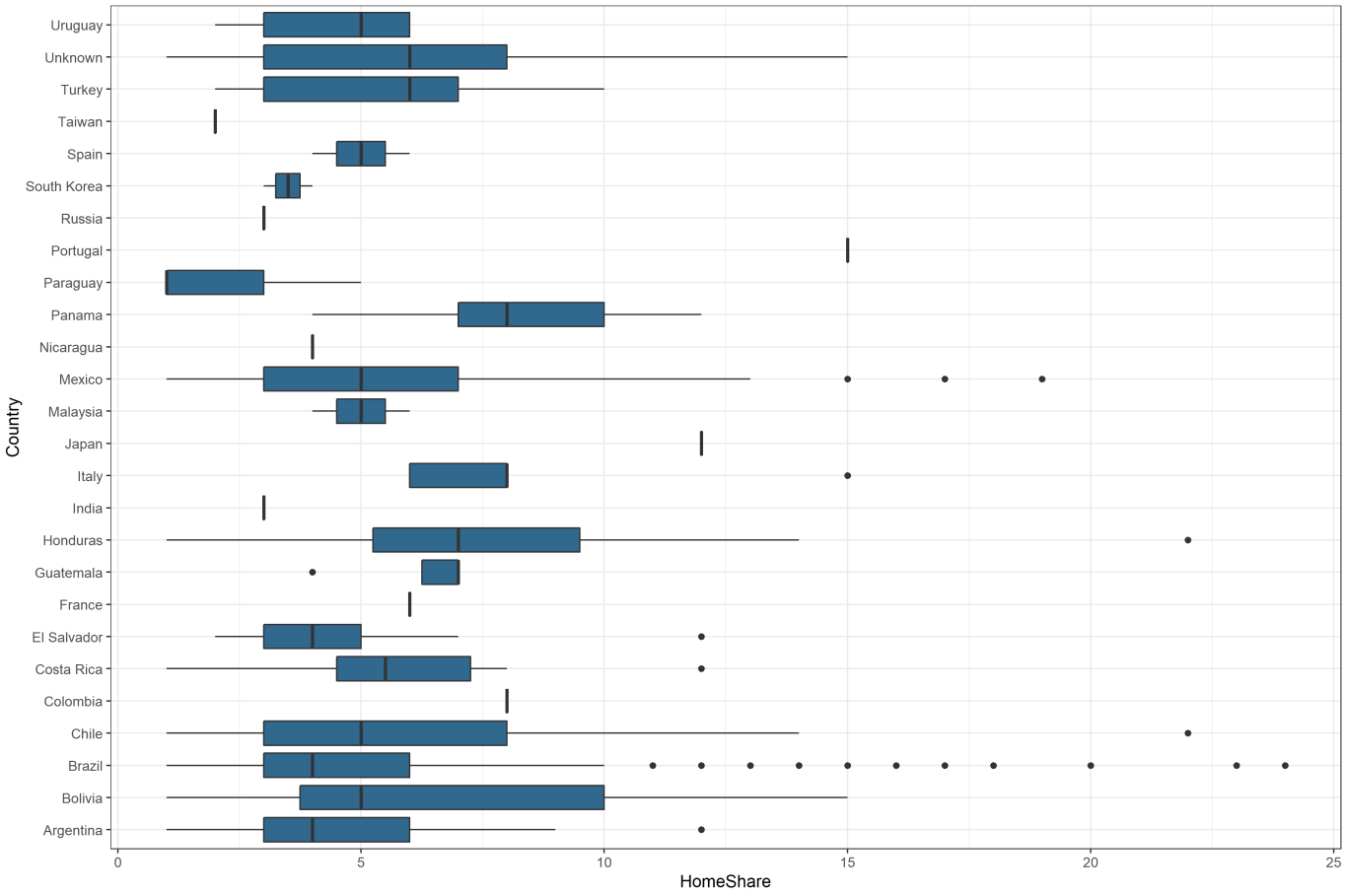


Fig. 11. Comparison between the number of people sharing a house per country

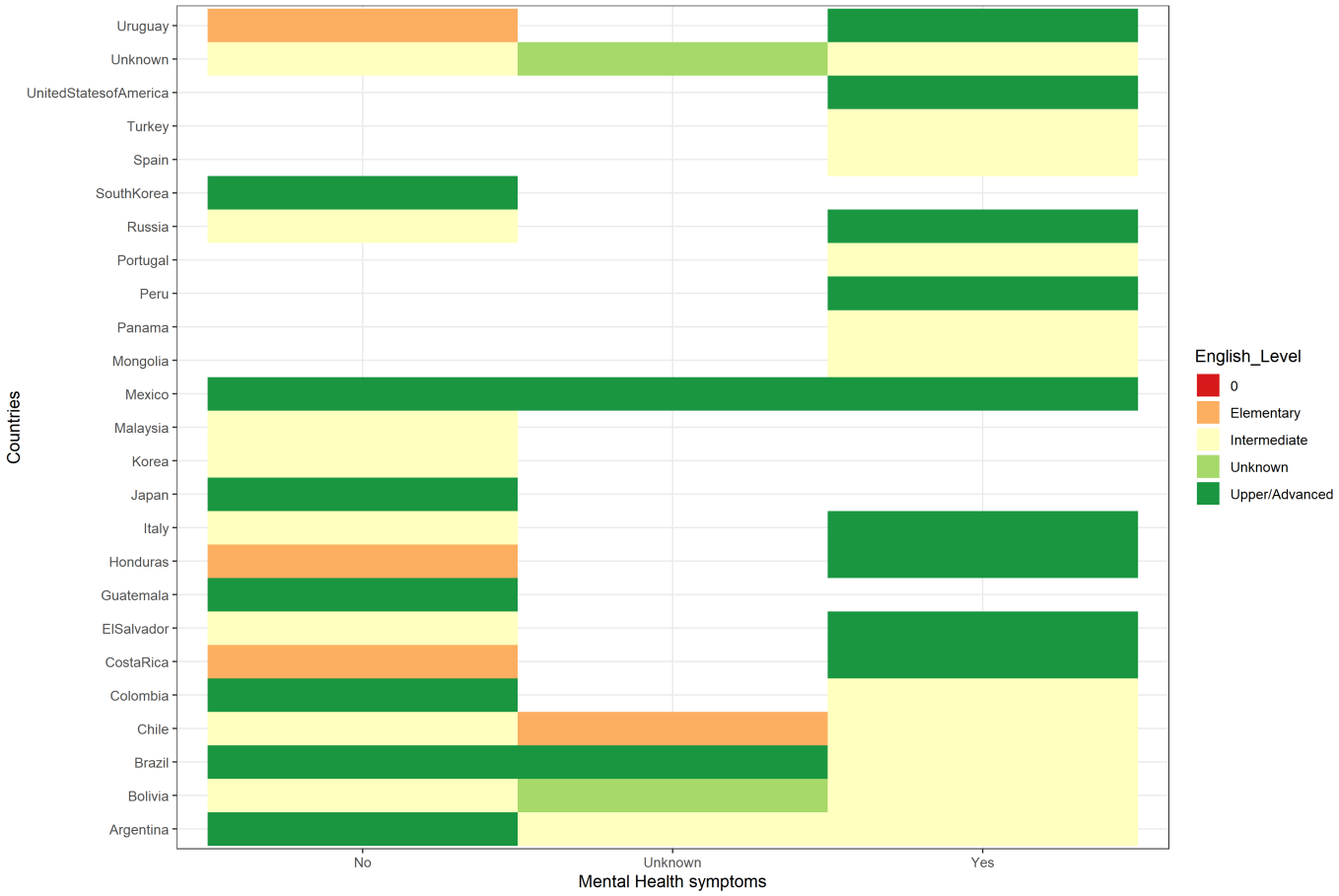


Fig. 12. Comparison Mental health associated with English level

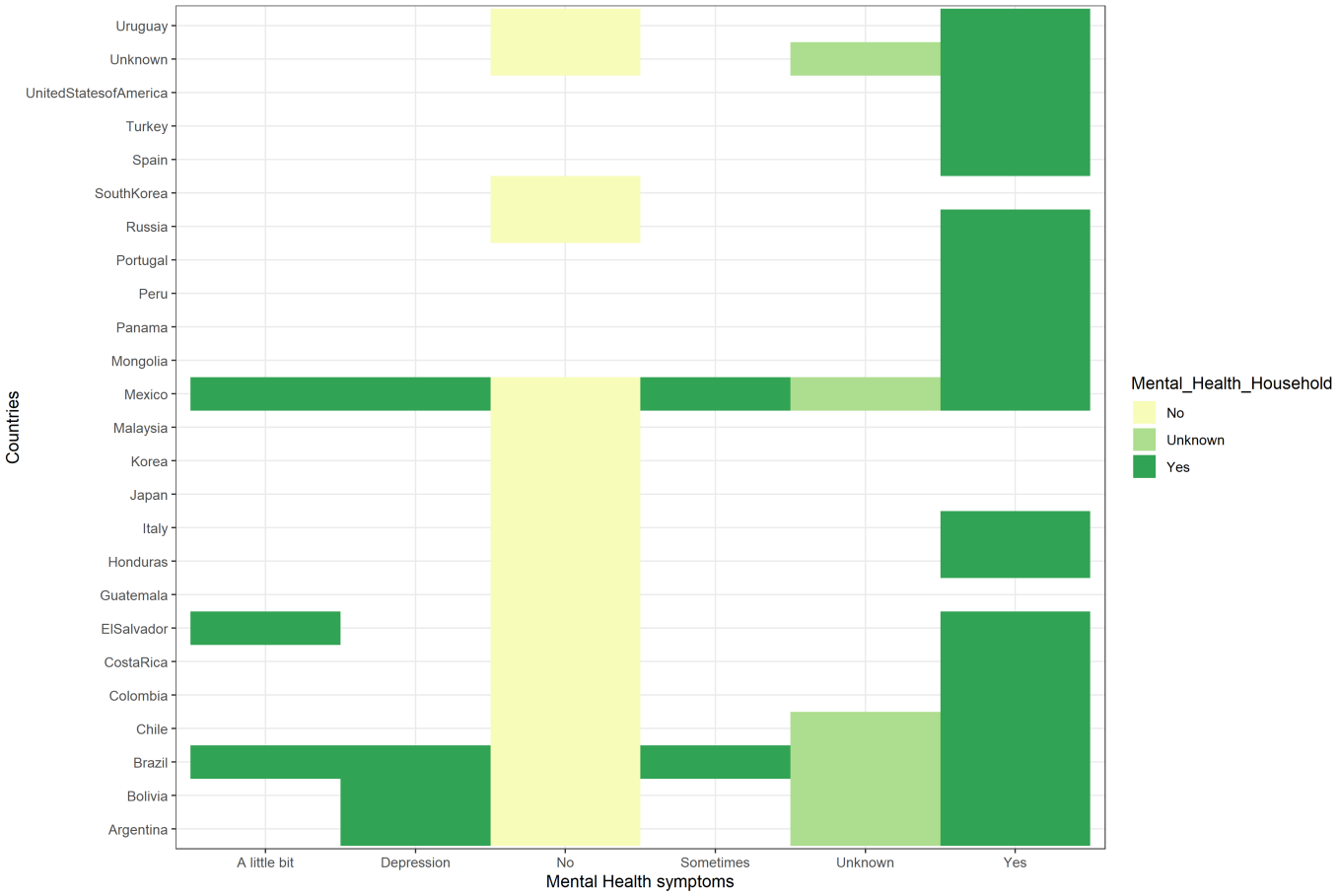


Fig. 13. Comparison Mental health conditions associated with countries

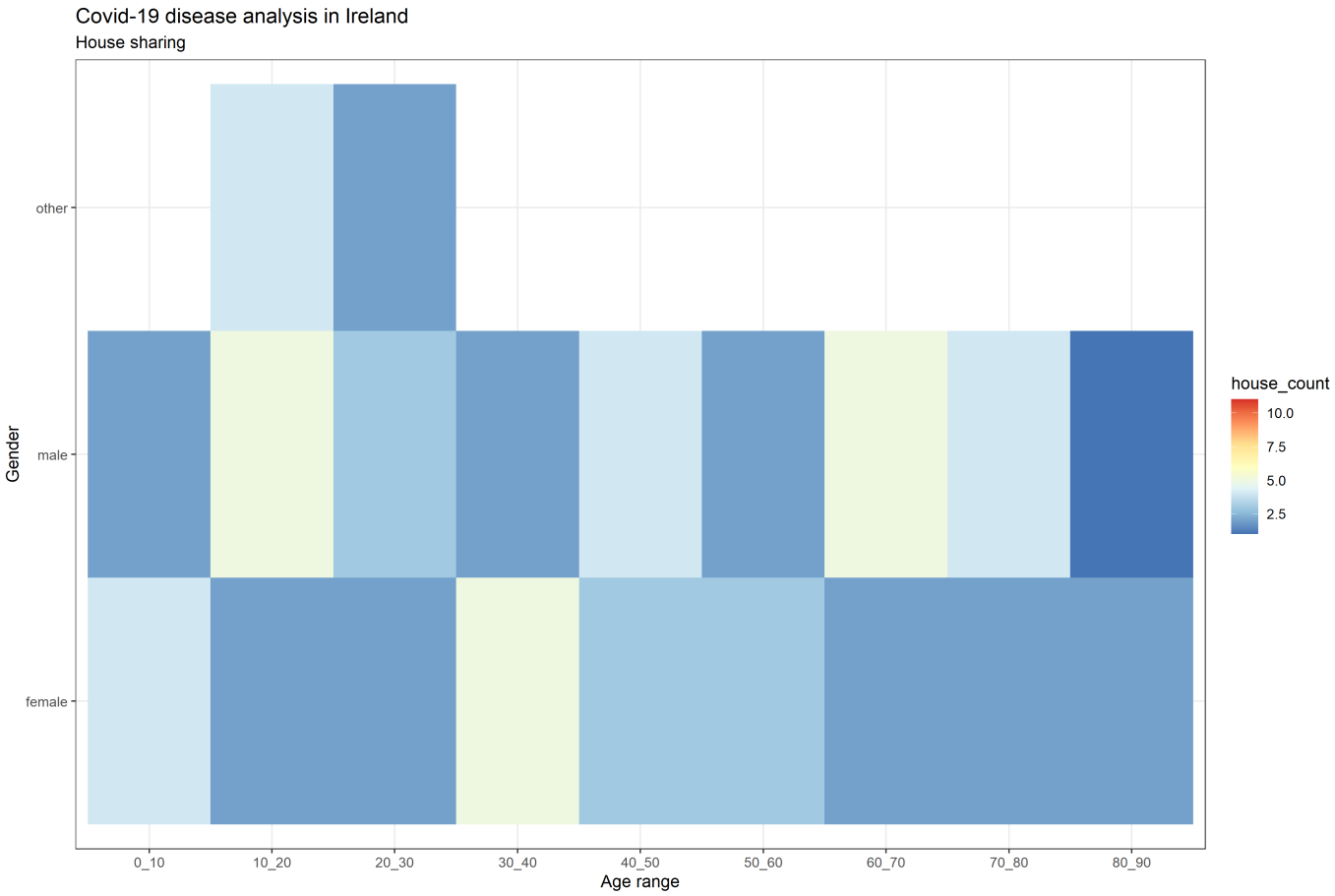


Fig. 14. nationals sharing a house

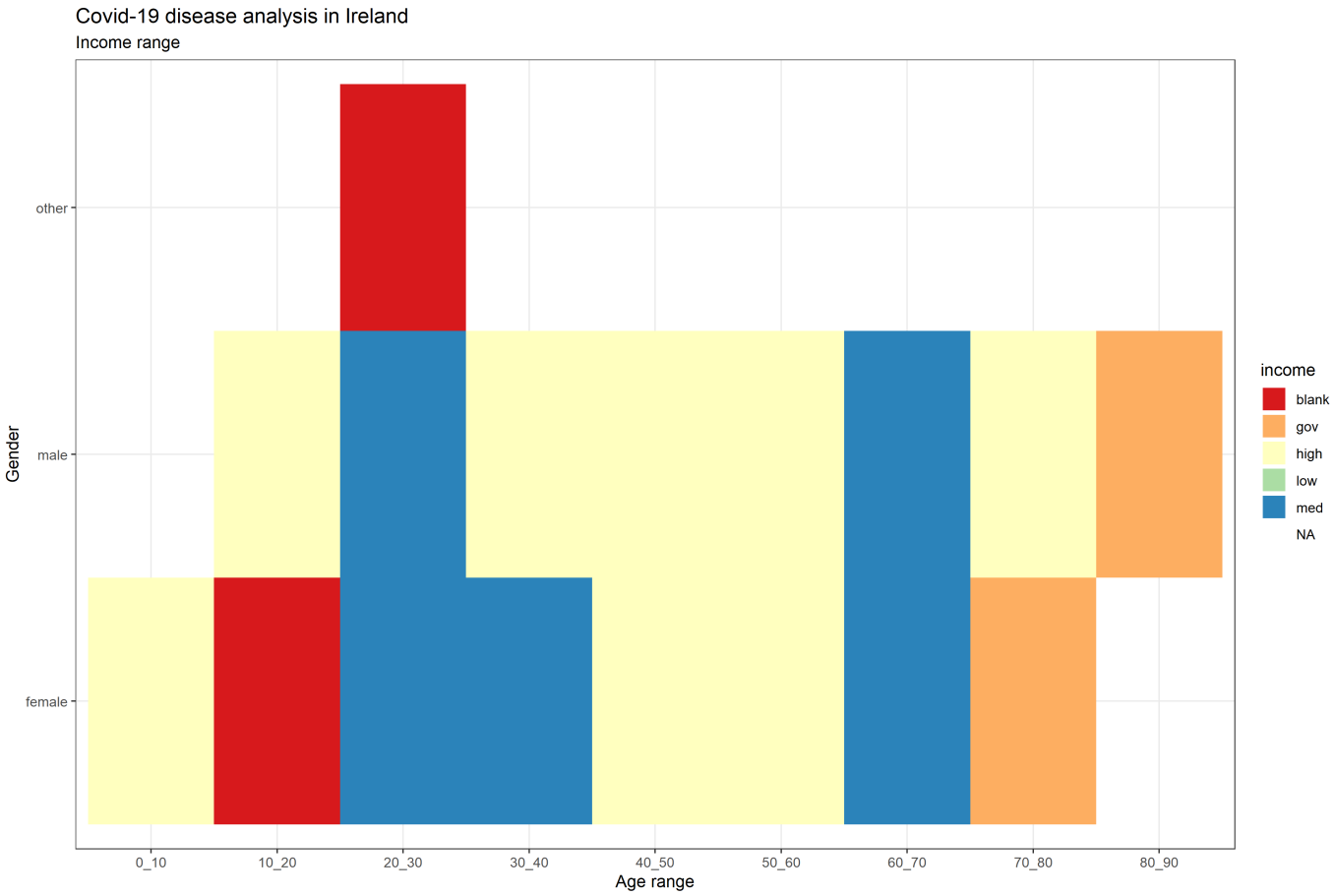


Fig. 15. Nationals income