The world's anxiety of cybercrime in 2019*

Julie Nguyen

20 April 2023

Abstract

The emergence of the internet has altered the way we live, work, and communicate with each other. Despite being a tool to improve life in overall, technology and the internet also pose some potential risks that greatly affect the life of people. This paper aims at identifying the perceptions of risk of using the internet among people around the world by exploring the patterns and the relationship between gender, the number of children, education level and the pereception of risk towards online bullying. Findings show that females tend to be more worried about experiencing cybercrime than males, people having more children also witness the same pattern. Results also show that people having higher level of education are more anxious about fake news.

1 Introduction

The emergence of the internet has altered the way we live, work, and communicate with each other. It has become an integral part of our daily lives, offering numerous benefits such as instant access to information, online shopping, and communication with people from all over the world. However, along with these benefits, the Internet also poses various risks to its users.

The dangers of using the internet increase as it develops and expands. Some of the main dangers that people encounter include cybercrime, online scams, identity theft, cyberbullying, and pestering. Regardless of the person's age, gender, or location, these dangers can have significant financial and mental repercussions. Findings from Gray (2017) about emerging technology and its potential risk suggest that the Internet of Things (IoT) fell into the group of technology that had lots of risks, but not so much reward. On the one hand, they gave more access to the life of people. On the other hand, they could also "give cyber-criminals a window to our world".

Therefore, it is essential to recognize these dangers. Understanding and identifying risk is as important as keeping the world healthy. This paper focuses on identifying the perceptions of risk of using the internet among people around the world. The paper also analyzes the correlation between education, income, and living locations, and how these factors affect the perception of risk towards the internet. In exploring the patterns of risk perception, the paper hopes to provide insights into people in different demographics experience and perceive potential risks posed by digital technology, through which will improve digital well-being, raise awareness, and enhance the ways to protect internet users in the world.

This paper will firstly explore the patterns of risk posed by technology in the world and in different global regions from the data of "World Risk Poll" (2019) in 2019 using R Core Team (2022). Various groups (of gender, education, and income) are observed to see how they affect the respondent's perception of risk caused by the internet. Other packages such as tidyverse (Wickham et al. 2019), readr (Wickham, Hester, and Bryan 2022), haven (Wickham, Miller, and Smith 2023), janitor (Firke 2021), ggplot2 (Wickham 2016), tidymodels (Kuhn and Wickham 2020) etc. are also used to aggregate, visualize, and model the data.

The main findings show that females tend to experience more anxiety towards the prospect of cybercrime than males. People living in low-income countries tend to witness the same pattern. People who have less kids tend to be less worried about the potential of risk.

In the following paper, I will delve into the data from "World Risk Poll" (2019), and how it is obtained and arranged for proper analysis. This section will also guide us through the understanding of the initial survey and ethical concerns

*Code and data are available at: https://github.com/julieenguyennn/World-Risk-Analysis

regarding the data collection. Moving on to the Model section, we will discuss our model and its implications for how we interpret our results going forward. In the Results section, graphical data is shown as to provide a clear understanding of the results given. Lastly, the Discussion section will include what has been done, what was learnt as well as identify some weaknesses and propose the next steps for the paper.

2 Data

2.1 Dataset

To attempt an analysis on the correlation between the respondents' gender, education attainment, the number of children in household and the anxiety towards the prospect of being a victim of cybercrime, I manipulated the "World Risk Poll" (2019) data from the year 2019. These data are collected by the Llyod's Register Foundation. The dataset was gatherd in… and last updated on… The raw data includes 150,000 observations. Using R (R Core Team 2022), tidyverse (Wickham et al. 2019), readr (Wickham, Hester, and Bryan 2022), haven (Wickham, Miller, and Smith 2023), janitor (Firke 2021), ggplot2 (Wickham 2016), tidymodels (Kuhn and Wickham 2020), I cleaned and extracted the necessary data to complete an exploratory analysis and modelling.

2.2 Variables

09 variables were selected to conduct analysis for the paper including 04 variables for respondents' general information (gender, education attainment, global region, and the number of children in household) and 03 variables relating to the anxiety towards the risk of being online (cyber bullying, false information and fraudulent activities). I first viewed the presence of anxiety towards different gender, education, and global region to investigate the global trend of online risk and anxiety in the year 2019. Along with these variables, the remainings are used later on to explore whether there is a causal relationship between the respondent's identity and their being worried about risks posed by the internet.

The anxiety towards different types of cybercrime were measure by asking the respondents "When using the Internet or social media, do you worry about any of the following things happening to you?". The answered is scaled as "Yes", "No". Gender of the respondents was recorded as 1 for "Male", and 2 for "Female". For the variable regarding the number of children in household, it is categorized into from 0 to 5 as the number of children accordingly, except for 5 was coded as "5 and above".

Lastly, the education attainment of the respondent is divided into 3 levels "Up to 8 years of basic education", "9-15 years of education", and "Completed four years of education beyond high school and/or received a 4-year college degree". Other types of responses that fell into "Do not Know" and "Refused" in all questions are omitted from analysis.

2.3 Plot

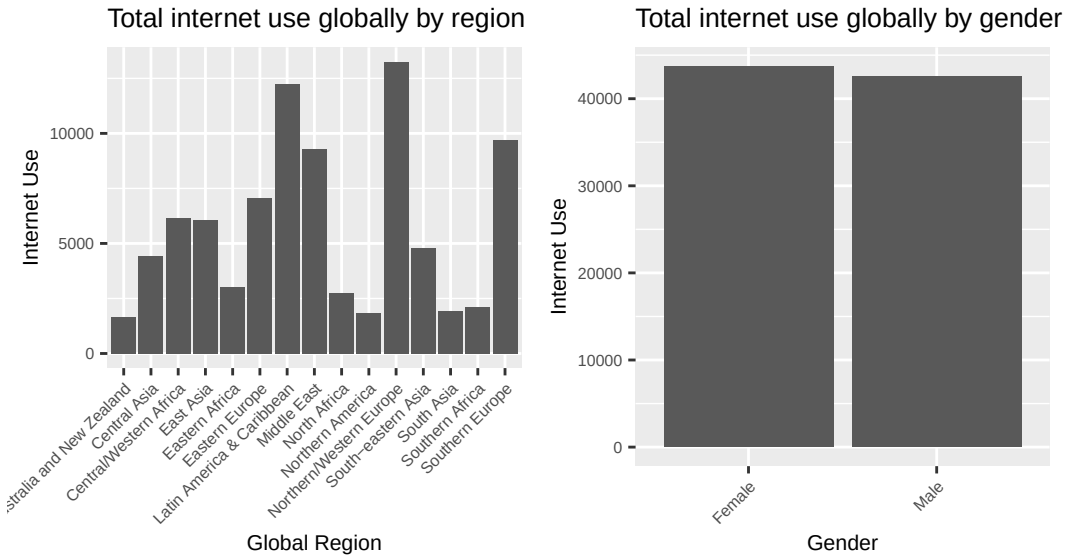

Figure 1: Anxiety towards cyber risks by region

Figure 1 depicts the demographics of total internet use globally by region and gender. A total number of 86,248 observations is recorded as having used the internet in the last 30 days from the time of survey. Among 86,248 observations, it is recorded the the majority comes from Europe, with 15.47% in Northern/Western Europe, 11.23% in Southern Europe, and 8.21% in Eastern Europe. On the other hand, the region of Australia and New Zealand only witnesses a total of 1666 respondents, which takes up 1.94% of the total observations. Regarding the gender, the survey witnesses a quite equal proportion of female and male respondents, 50.75% and 49.25% respectively.

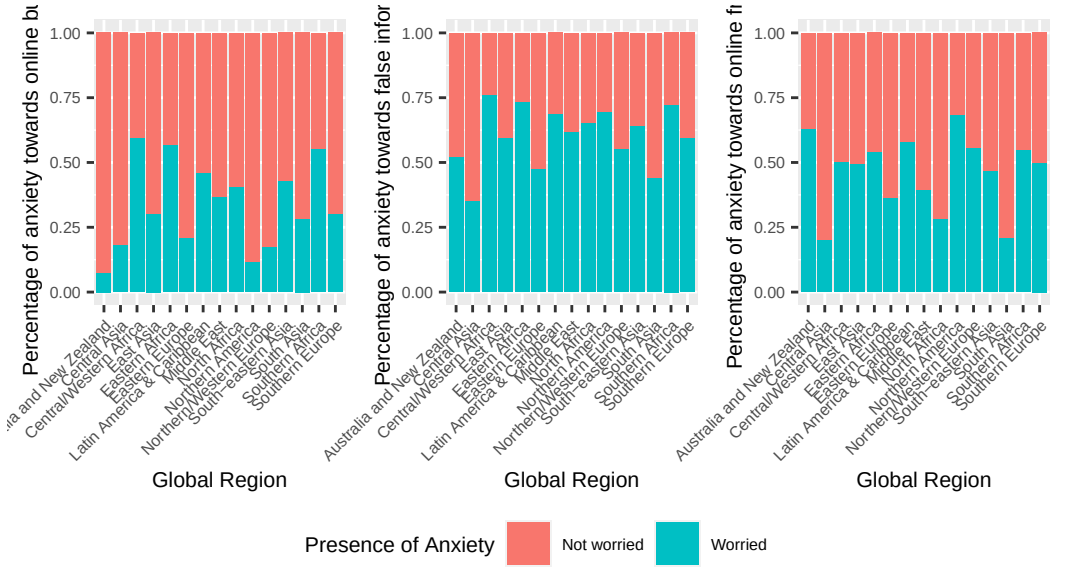

Figure 2: Anxiety towards cyber risks by region

Figure 2 illustrates the proportion of respondents feeling anxious towards the three types of cybercrime. In general, people were more worried about the prospect of being a victim of false information. A majority of respondents in Africa reported to have been worried about being online bullied, with a percentage of xxx% in Central/Western Africa, xxx% in Eastern Africa, xxx% in North Africa, and xxx% in Southern Africa. People from Australia and New Zealand were deeply concerned with confronting false information and fraudulent activities online. On the other hand, the

region of Central Asia and South Asia recorded similar attitudes of being "not worried" than "worried" for any of the three types of cybercrime.

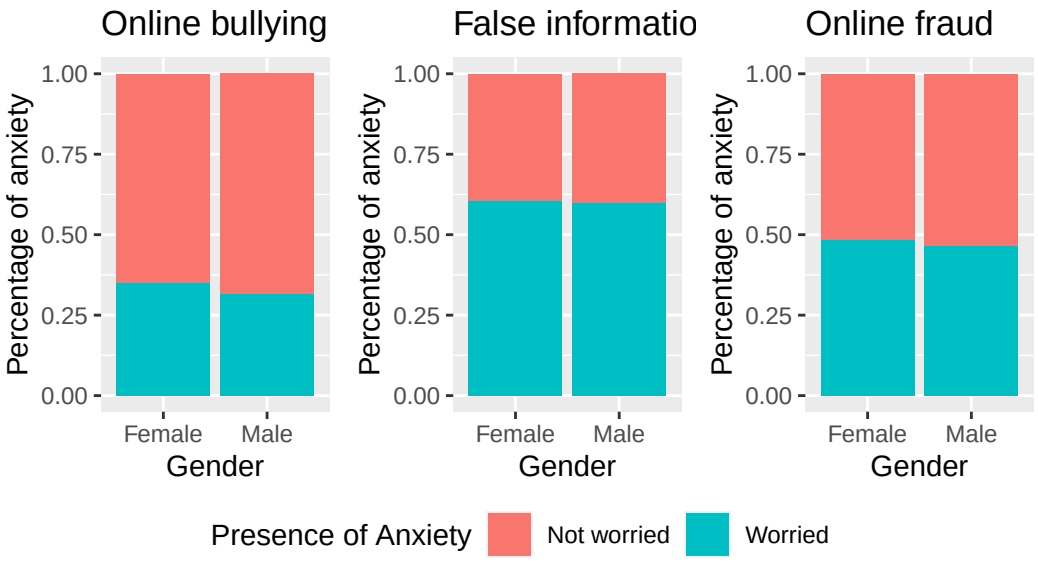

Figure 3: Anxiety towards cyber risks by gender

Figure 3 shows the percentage of people feeling worried about three types of cybercrime by gender. There is an overall trend, although not too distinct, that female respondents feel more anxious than male towards the online risks. It is also obvious that both male and female respondents felt more anxious towards the prospect of facing false information when using the internet and social media.

3 Model

In order to investigate and validate the relationship between a person's identity and their anxiety towards online crime, I utilized multiple binary logistic regression using R (R Core Team 2022) and tidymodels (Kuhn and Wickham 2020). This model is appropriate for an explanatory analysis of binary outcome variable, in this case the presence and absence of anxiety. Using the logistic regression also allows me to identify risk factors for anxiety towards cybercrime that may be more prevalent in certain populations, as well as the estimation of the probability of getting worried towards those crimes. I will be using four standard logistic regression models, each with separate dependent variable and a predictor as follows:

$$log(\frac{\pi_i}{\pi_i - 1}) = \beta_0 + \beta_1 * x_1$$

In this model, π is the probability of respondents recorded as worrying about the cybercrime and x_1 is the gender of the respondents, for $x_1 = 1$ is male and $x_1 = 0$ is female.

A similar and expanded model is also used to demonstrate the likelihood of a person's education attainment and the number of children in household towards the anxiety of cybercrime.

$$log(\frac{\pi_i}{\pi_i - 1}) = \beta_0 + \beta_1 * x_1 + \beta_2 * x_2 + \beta_3 * x_3 + \beta_4 * x_4 + \beta_5 * x_5$$

Similarly, π is the probability of respondents recorded as worrying about the cybercrime, while x_i is the independent variables of education level or the number of children in household.

4 Results

4.1 Anxiety towards online bullying

Examining the anxiety towards online bullying, I choose 02 independent variables that are most suitable for the explanation of the likelihood of getting anxious, which are the respondent's gender and the number of children in household. It is generally believed that gender would affect the rise of anxiety when browsing online, as well as the number of children that a person is raising. In the current era, parents are getting more worried as their children go online easier and more frequently.

Table 1 shows that the odds ratio at the intercept is 0.54, which represents that female respondents have an odd ratio of 0.54 of getting anxious over cyberbullying. The p-value for both male and female respondents are lower than 0.001, which indicate that these variables have statistical significance.

Moreover, results from Table 1 shows summary of model in the number of children in a household. It is apparent that odds ratio get higher when the number of children in household increase. This also means that people or more likely to get worried when they have more kids in the house.

Table 1: Summary model of anxiety towards online bullying

	Gender			Children in Household		
Predictors	Odds Ratios	std. Error	Statistic	Odds Ratios	std. Error	Statistic
(Intercept)	0.54 ***	0.01	-61.32	0.39 ***	0.00	-88.73
gender [Male]	0.85 ***	0.01	-10.90			
children in household [1]				1.41 ***	0.03	17.72
children in household [2]				1.53 ***	0.03	20.71
children in household [3]				1.88 ***	0.05	22.35
children in household [4]				2.06 ***	0.08	17.54
children in household [5]				2.64 ***	0.11	24.37
Observations	86248			86248		
R^2 Tjur	0.001			0.017		
* p<0.05 ** p<0.01 *** p<0.001						

To predict the characteristics of people getting anxious over the prospect of online bullying, I plotted a model for both gender variable and number of children variable. Figure 4 displays the predicted probability of anxiety among genders and the number of children that one is raising. The graph on the left shows that . On the other hand, the graph on the right side indicates the more children there is in a household, the more likely that the parent/guardian is worried about cyberbullying.

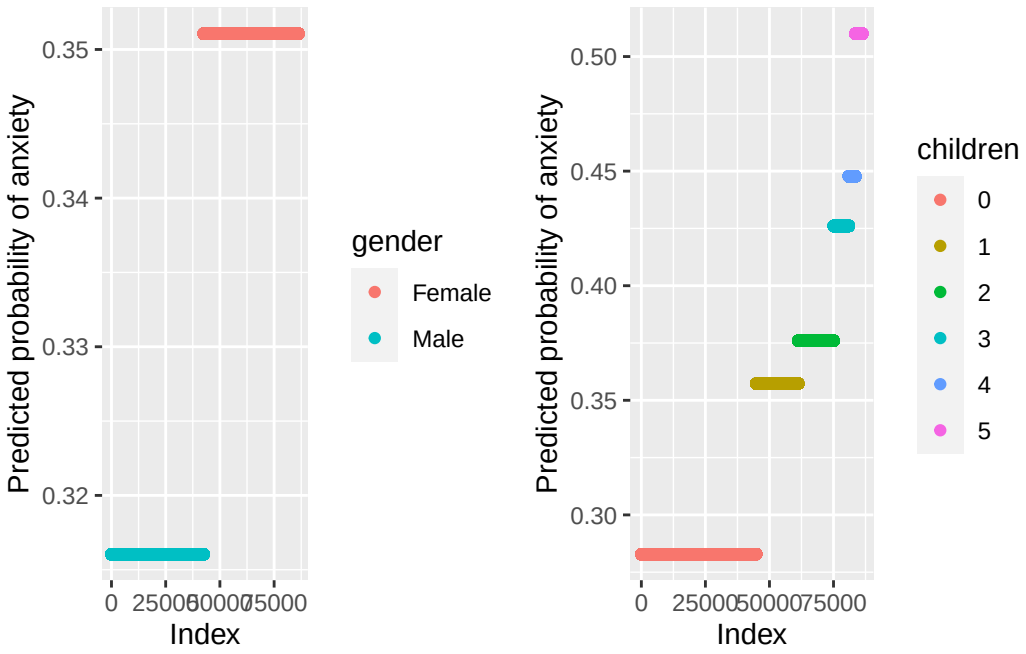

Figure 4: Predicted probability of getting anxious over online bullying

4.2 Anxiety towards false information

The respondent's education attainment is chosen as an independent variable for examining the correlation and the likelihood of experiencing anxiety. The summary of the model Table 2 concludes that the group of people pursuing education from 9-15 years have the highest odds ratio of 1.54, while that of 8-year basic education and higher education are 0.90 and 0.96 respectively.

Table 2: Summary model of anxiety towards false information

	Education Level		
Predictors	Odds Ratios	std. Error	Statistic
(Intercept)	1.54 ***	0.01	48.26
education [Completed 4 years beyond high-school]	0.96 *	0.02	-2.28
education [Up to 8 years]	0.90 ***	0.02	-5.12
Observations	86248		
R^2 Tjur	0.000		
* p<0.05 ** p<0.01 *** p<0.001			

Figure 5 illustrates the predicted probability of anxiety confronting false information between different education attainment. The model predicts that people falling into the group of 9-15 years of education are prone to feel worried about getting false information or being a victim of it, with a probability of 0.61%, while that of the basic education group is 0.58%.

4.3 Anxiety towards online fraudulent activities

The respondent's education attainment is once again chosen as an independent variable for examining the correlation and the likelihood of experiencing anxiety towards online fraudulent activities. It is believed that the higher the education level one possesses, the lower chance that one gets anxiety over frauds.

The ordinal logistic regression model is used to analyze this relationship. Table 3 shows that odds ratio at the intercept is 0.89, which means people pursuing education from 9-15 years, the odds ratio of getting worried is 0.89. In addition,

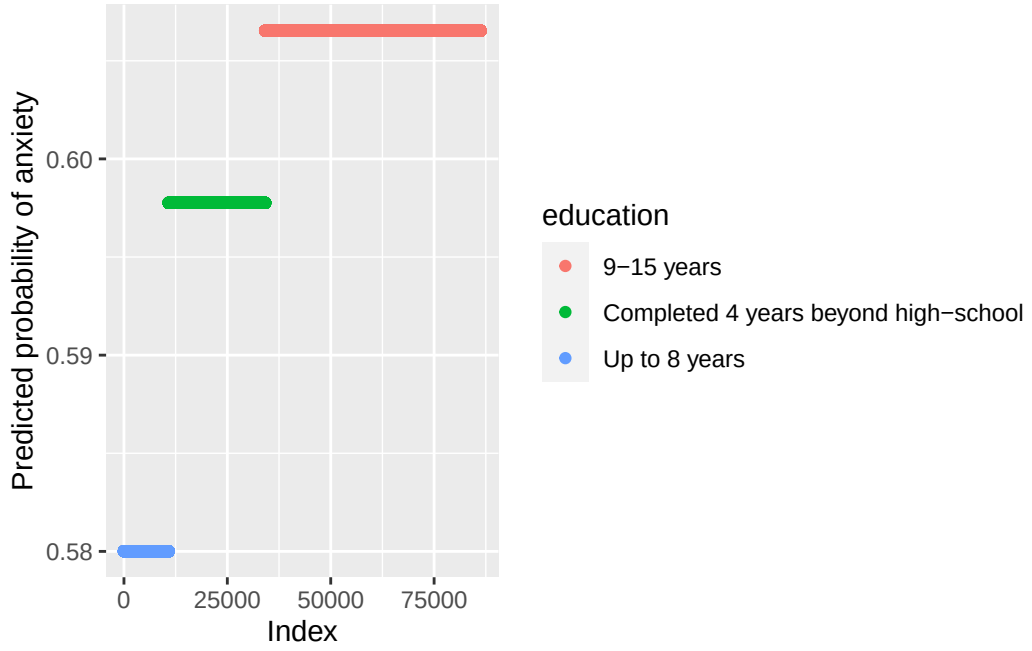

Figure 5: Predicted probability of anxiety towards false information among education groups

Figure 6 predicts that people pursuing or obtained higher education have a slight probability of worrying about the prospect of being a victim of online fraud.

Table 3: Summary model of anxiety towards online fraudulent activities

	Education Level		
Predictors	Odds Ratios	std. Error	Statistic
(Intercept)	0.89 ***	0.01	-13.59
education [Completed 4 years beyond high-school]	1.18 ***	0.02	10.73
education [Up to 8 years]	0.76 ***	0.02	-12.95
Observations	86248		
R^2 Tjur	0.004		
* p<0.05 ** p<0.01 *** p<0.001			

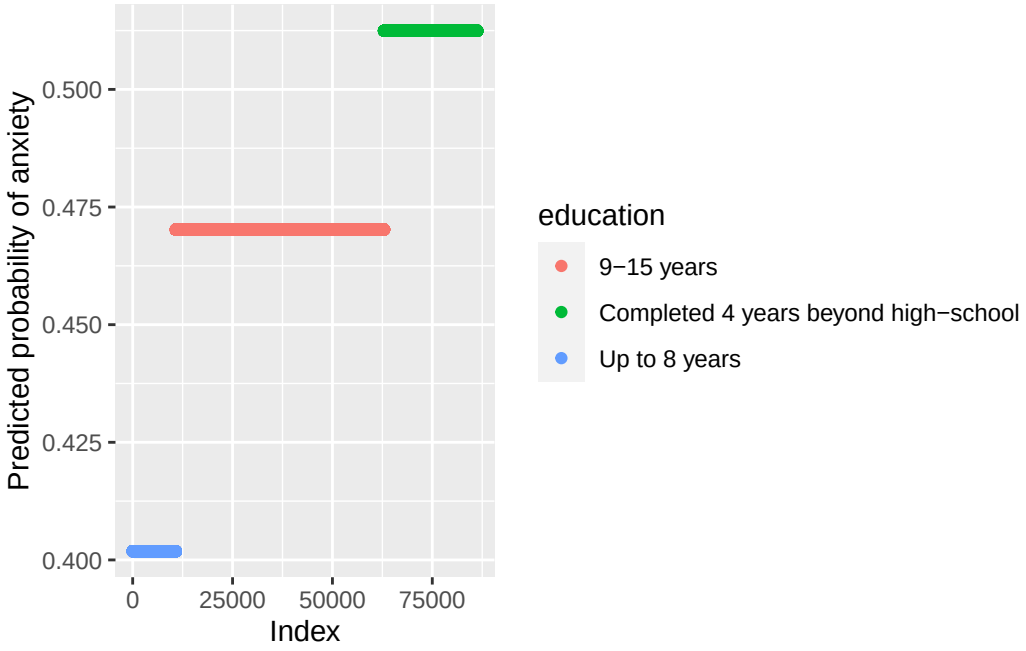

Figure 6: Predicted probability of anxiety towards online frauds among education groups

5 Discussion

5.1 Fake news rank the top worry for internet users worldwide

The data and findings of this paper sparks the existence of anxiety towards false information. Internet users across the world regard receiving false information as a greater worry than other risks such as online bullying and internet fraud. 75% per cent of internet users across all parts of the world, socio-economic groups and all ages, regard false information as a major concern. The concern is more prevalent in areas of substantial economic disparity and where ethnic, religious, or political division exists, resulting in a deterioration of social cohesion and trust. A similar pattern is also witnessed in terms of gender, people are extremely worried about fake news all over the world.

5.2 Education attainment and anxiety towards cybercrime

The analysis spotted out some interesting correlations between the education level of the respondents and their anxiety towards cybercrime. It is widely believed that the higher the education level one gets, the less worried they are, as people attained more knowledge, have higher awareness and are able to identify the threats. That is also the reason why governments and organizations have conducted campaigns for people pursuing lower level of education or the illiterates, in order to make up for their lack of knowledge and awareness of how to protect themselves from online threats. Nevertheless, the findings point out that higher-educated people are prone to feel more worried about the online environment and the crime. While people pursuing higher education are less worried about facing false information, they are more likely to feel anxious about the prospect of being a fraud victim. On the other hand, people that obtained basic 8-year education do not see either false information or online fraudulent activities as threats or risks to their life.

The findings has illustrated the current situation of cyber risks and predicted the global anxiety over cybercrime in 2019. This highlights the importance of educating individuals from all backgrounds on the risks of cybercrime and how to stay safe online. On the other hand, this also lays the ground for the World Risk Poll 2021 analysis, in order to see how this pattern has changed throughout the COVID-19 period - a time when everybody stayed at home and communicated mainly through digital devices.

5.3 Weaknesses & Ethical Concerns

The weaknesses and ethical concerns of this paper and analysis fall into both the original dataset and the treatment of data and modelling.

Firstly, weakness emerged from the original dataset can be accounted for how survey questions were designed, which could have introduced to the appearance of non-sampling errors. The type of non-sampling errors in the survey mostly falls into non-response error, where respondents left the answer blank or provide incomplete answers. In addition, the binary options for measuring anxiety in the survey could introduce bias and inaccuracies in the results. This is because people may have different levels of interpretations of what constitutes "worried" or "not worried".

The reliance on self-selection by respondents may have introduced biases into the sample, as those who choose to participate may have different characteristics or opinions than those who do not. Additionally, the reliance on web-based responses may have excluded individuals who do not have access to or are not comfortable using the internet, further potentially reducing the representativeness of the sample.

Regarding the data treatment and modelling, there are also a few weaknesses that can be found. Firstly is the omission of NA values. Any responses that are recorded as NA or "Don't Know" were omitted during the data cleaning process, which lower the total number of respondents from 154,195 to 86,248. This significantly affects the overall analysis of the trend as well as the modelling for prediction.

References

Firke, Sam. 2021. *Janitor: Simple Tools for Examining and Cleaning Dirty Data*. https://CRAN.R-project.org/package=janitor.

Gray, Alex. 2017. "What New Technologies Carry the Biggest Risks?" World Economic Forum. https://www.weforum.org/agenda/2017/01/what-emerging-technologies-have-the-biggest-negative-consequences/.

Kuhn, Max, and Hadley Wickham. 2020. *Tidymodels: A Collection of Packages for Modeling and Machine Learning Using Tidyverse Principles*. https://www.tidymodels.org.

R Core Team. 2022. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.

Wickham, Hadley. 2016. *Ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York. https://ggplot2.tidyverse.org.

Wickham, Hadley, Mara Averick, Jennifer Bryan, Winston Chang, Lucy D'Agostino McGowan, Romain François, Garrett Grolemund, et al. 2019. "Welcome to the tidyverse." *Journal of Open Source Software* 4 (43): 1686. https://doi.org/10.21105/joss.01686.

Wickham, Hadley, Jim Hester, and Jennifer Bryan. 2022. *Readr: Read Rectangular Text Data*. https://CRAN.R-project.org/package=readr.

Wickham, Hadley, Evan Miller, and Danny Smith. 2023. *Haven: Import and Export 'SPSS', 'Stata' and 'SAS' Files*. https://CRAN.R-project.org/package=haven.

"World Risk Poll." 2019. The Lloyd's Register Foundation. https://wrp.lrfoundation.org.uk/.