# Younger, Unmarried, Childless, and Less Happy Americans Use the Internet the Most*

An Analysis of the 2016 US General Social Survey

Sakura Ariga Aliyah Maxine Ramos Annie Yan

15 March 2023

**Abstract**

In this study, we investigated the 2016 United States General Social Survey to gain an understanding of the characteristics of the people that use the Internet the most, such as age, marital status, number of children, and level of happiness. We did this by comparing the average number of weekly Internet use across population groups. We found that young adults that have never been married or have children and who are not satisfied with their general happiness are those that use the Internet the most. In an era in which the Internet is becoming more and more used in the US, it is important to understand which population group uses the Internet the most in order to inform public policies surrounding Internet addiction and mental health.

## 1 Introduction

The Internet has become a widely used computer network that allows for connection, communication, and access to information. By the mid 1990s, the Internet began reaching millions of users from around the world, where using a web of information, known as the World Wide Web, was a popular way of browsing information and facilitating communication online ("A Short History of the Internet" 2020). Throughout the years, the different ways in which the Internet is used have expanded, and now users can also make purchases through electronic commerce platforms and share different forms of media on social media platforms.

This report aims to investigate the characteristics of American adults that use the Internet the most in comparison to those that use it the least. We aim to understand the reasons why these factors are correlated with a higher Internet use pattern and a lower Internet use pattern. Our estimand is the mean number of weekly hours of Internet usage of American adults. Using the survey data from the United States General Social Survey (GSS) recorded in 2016, we constructed different models on 6 factors and their relationship with total weekly hours of Internet use: age, race, sex, marital status, number of children, and happiness.

Through examining the relationships between the variables and the weekly number of hours of Internet use, we found that young adults tend to use the Internet more than older adults. This may be due to physical, psychological, and social barriers when using the Internet. We also found that those who have never been married use the Internet more than others with different marital statuses. This could be related to age and the difference in reasons for using the Internet. Additionally, as social media is a large part of Internet usage, we found that the using these platforms have a negative impact on emotional health, demonstrating a possible correlation between high Internet usage and a low level of happiness. These findings are important because using the Internet more regularly can improve the lives of individuals for its convenience and its accessibility. It is beneficial because it allows for easier and instantaneous communication, for learning, for increasing access to services, for encouraging connections and decreasing isolation, and for gathering information (Lane 2022). However, the consequences of increased Internet usage, such as Internet addiction and mental health issues, should also be kept in mind. Our research on which population groups use the Internet the most in

*Code and data are available at: https://github.com/AnnieYan0807/GSS-data-analysis.git.

the US can lay the foundation for subsequent research investigating the pros and cons of Internet use on different population groups, and the ways in which the pros and cons affect those who use the Internet more.

The remainder of this report is structured as follows: Section 2 discusses the dataset of interest, including the methodology of the survey and its strengths and weaknesses, Section 3 presents the results through data visualizations, Section 4 discusses our findings, its weaknesses, and next steps, and Section 5 consists of the supplementary survey as an extension of the survey data to improve and enhance the study.

# 2 Data

## 2.1 Survey

### 2.1.1 Key Features

The General Social Survey (GSS) is a national survey conducted in the United States of America that aims to better understand trends in the American adult population's social characteristics and attitudes to better inform research and public policy ("About the GSS," n.d.). Beginning in 1972, this interview-based survey is conducted biennially on five key themes: gender and marriage, current affairs, civil liberties, politics, and religion and spirituality ("Key Trends," n.d.). The 2016 GSS is unique in that it also includes questions about Internet-usage behaviours, which is why our analysis focuses on the GSS of this year. In 2016, there is data for 2,867 respondents for 961 variables.

### 2.1.2 Methodology

The GSS' target population is adults aged 18 years and older living in the United States. The sampling frame for the 2016 GSS is people living in the US aged 18+ who can speak English or Spanish and live in households ("Frequently Asked Questions," n.d.). Thus, those living outside households (e.g. in university dorms, nursing homes, institutions, military quarters) are not included in the GSS ("GSS 1972-2018 Cross-Section Codebook: Appendix a - Sampling Design and Weighting," n.d.). The frame was created using the NORC National Sampling Frame, which first uses US Census data to stratify US geographic regions by size and then chooses metropolitan areas and non-metropolitan counties within these regions, as the first stage of selection ("GSS 1972-2018 Cross-Section Codebook: Appendix a - Sampling Design and Weighting," n.d.). In the second stage of selection, these metropolitan areas and non-metropolitan counties were stratified by race and income and further divided into blocks ("GSS 1972-2018 Cross-Section Codebook: Appendix a - Sampling Design and Weighting," n.d.). Then, the sample is gathered by interviewers beginning from the northwest corner of a given block and moving from household to household until the equal-sex quota has been filled ("GSS 1972-2018 Cross-Section Codebook: Appendix a - Sampling Design and Weighting," n.d.).

Thus, the sample for the 2016 GSS was an area-probability sample gathered by the clustered sampling approach, as households are selected and then one random member from the household is selected to be included in the survey ("GSS 1972-2018 Cross-Section Codebook: Appendix a - Sampling Design and Weighting," n.d.). A key tradeoff of this approach is that, while this method is much easier than solely random sampling, it can lead to bias in that there is more bias if the clusters are similar to each other.

Non-response is handled by inputting it in the GSS dataset as one of the following options depending on the method of non-response: "NA", "no answer", "do not know", or "skipped".

### 2.1.3 Strengths & Weaknesses

A strength of this 2016 GSS survey was that it was delivered both in English and in Spanish, allowing for a greater reach than had the survey just been conducted in English, like in the years previous to 2006 ("GSS 1972-2018 Cross-Section Codebook: Appendix a - Sampling Design and Weighting," n.d.). Of the people who had been excluded from this survey due to the English-only delivery, Spanish-speakers made up the majority, so this dual language delivery addressed this key exclusion ("GSS 1972-2018 Cross-Section Codebook: Appendix a - Sampling Design and Weighting," n.d.).

A weakness of this sample is the lack of a gender variable, as the sample only provides a variable on sex. A gender variable would be more useful for data analysis, as it better reflects the respondent's lived experience. The lack of a gender variable that better accommodates for people's lived experience might lead to those who feel uncomfortable with the binary categories of sex to refrain from participating in the survey, leading to selection bias.

Another weakness of the 2016 GSS survey is that the interviewer physically fills out the responses for the respondent. This can potentially lead to measurement error. When the interviewer fills out the survey on behalf of the respondent, there may be a difference in the observed responses of the interviewer and the actual responses of the interviewee. This creates difficulty in verifying responses ("For Survey Participants," n.d.).

A third weakness of the survey is the role of selection bias that exists due to the inability to reach the entire population. The survey may also be susceptible to volunteering bias because the data is reliant on participants willingly providing their answers, which could lead to biased results.

## 2.2 Questionnaire

A strength of the questionnaire was the clear wording of the question regarding Internet usage per week in hours. The survey question was formatted as follows: "Not counting e-mail, about how many minutes or hours per week do you use the Web? (Include time you spend visiting regular web sites and time spent using interactive Internet services like chat rooms, Usenet groups, discussion forums, bulletin boards, and the like.)" ("GSS Data Explorer: NORC at the University of Chicago" (n.d.)). The providing of examples was likely helpful in stimulating the respondents' memory of what exactly Internet usage can look like, as if one is flat out asked whether they use the Internet, it can be hard to conceptualize what that looks like.

A key weakness of this questionnaire is the race variable, as it only allows for three options: "Black", "White", or "Other", which does not take into account the largest ethnic minority group, Hispanics and Latino Americans, as well as other ethnic groups, such as Asians and Indigenous Americans in the United States. In comparison, while not without fault itself, the US Census identifies 5 minimum categories of race (White, Black or African American, American Indian or Alaska Native, Asian, Native Hawaiian or Other Pacific Islander) and also allows for the reporting of more than one race (Bureau 2022). The questionnaire should at least have followed these US Census minimum race categories.

Another weakness of this questionnaire is that questions are allowed non-responses. This results in truncated data, in which the dataset has been incompletely collected. Having no response for a question can only be observed as null, despite there being a possible actual value (Alexander 2023). There are a variety of reasons as to why a respondent may not want to answer a question, but can only be observed as a non-response by researchers to avoid biased assumptions and inaccurate results.

## 2.3 Overview of Data

### 2.3.1 Data Source

In this report, the 2867 observations from the 2016 United States General Social Survey are used. The survey data was obtained from the University of Chicago National Opinion Research Centre (NORC) ("GSS Data Explorer: NORC at the University of Chicago," n.d.).

This report was created using the R statistical programming language (R Core Team 2020). The here package was utilized to retrieve files from another folder within the same R project (Müller 2020). For the results and analysis of this report, all figures were created using the tidyverse package (Wickham et al. 2019). Additionally, the tables were created using the packages knitr (Xie 2023) and kableExtra (Zhu 2021), and the graphs were created using the packages lessR (Gerbing 2021), tidyr (Wickham, Vaughan, and Girlich 2023), and ggplot2 (Wickham 2016).