

Impact of Internet Infrastructure on Canadian Household Productivity and Recreation*

Brief Overview of Canadian Internet Use Trends

Haocong Wu

26 April 2022

Abstract

Internet access is critical to many users. This paper examines the quality of internet access that users have and how they are impacted by subpar service. With data collected by Statistics Canada in the Canadian Internet Use Survey, and modelled with linear regression, stability positively impacts user household internet usage, whilst speed has diminishing returns past a certain point.

Keywords: internet use, internet infrastructure, statistics canada, canadian residents, linear regression

1 Introduction

Internet access is becoming, if not already considered, a human right. It is a standard point of communication used by the majority of the Canadian population. The platform that the internet provides can be used to host a variety of activities, ranging from those of productive nature, to others of recreational and leisurely type. The growing importance of internet access necessitates a revision of the bare minimums of providing internet service.

This paper examines the causal relations between individual internet usage amounts to factors such as connection speed, nature of use, and necessity of use.

The following sections of the paper is organized as stated: the Data section discusses the source of the data and its surveying methodology, the Modelling section attempts to formalize a model incorporating various factors to explain the response variable of user household internet use duration, the Results section discusses the impact of the selected factors, and the Discussion section comments on the results and its validity, potential bias, and weaknesses of the paper.

2 Data

2.1 Data Source

The Canadian Internet Use Survey (CIUS) replaced the Household Internet Use Survey (HIUS) in 2005 (Statistics Canada 2019). The 2018 of the CIUS aims to record and track the quickly increasing Canadian internet utilization. Aspects of internet technologies that Canadians interact with such as online government services, social networks, smartphone use, and online commercial platforms are presented to respondents for their sentiments.

*Code and data are available at: <https://github.com/hcgw0318/2018-Canadian-Internet-Use-Analysis>

Survey data was provided by the portal, maintained by the Ontario Council of University Libraries.

The particular set of data used in this report is a subset of the topics covered by the 2018 CIUS.

2.2 Survey Methodology

Based on the 2018 CIUS documentation (Ontario Data Documentation 2020), the survey methodology can be summarized by the following.

The CIUS attempts to gauge the population habits regarding various facets of internet use. The CIUS' stated target population includes all individuals at and above the age of 15 in Canada, excluding those in the Canadian Territories and excluding institutionalized individuals.

The sampling frame of the GSS relies on the availability of telephone access. Households without telephones, as well as households with telephone services not covered by the current frame, were excluded from the sampling frame.

The sampling process consisted of location based random sampling without replacement. A target minimum sample size for each province was determined, and this minimum sample size was allocated to the strata within the province.

Each respondent was contacted to complete either an electronic questionnaire, or participate in a computer assisted telephone interview. If the respondent refused or was absent, further contact attempts were made through telephone contact until reasonable confirmation of non-response, and the respondent is removed from the pool.

The 2018 CIUS had a target sample size of 15,000. The actual number of respondents was 13,810, the total households contacted was 33,248, and the overall response rate was 41.6%.

2.3 Data Processing Tools and Methodology

Data processing and analysis in this project is done in the R statistical programming language (R Core Team 2020). The following R packages were imported and used for their corresponding purpose:

- **tidyverse** (Wickham et al. 2019) : Used for general logical commands, file importing, and data manipulation
- **knitr** (Xie 2021) : Used to generate and knit the Markdown document to PDF
- **ggplot2** (Wickham 2016) : Used to generate graphs and apply styling
- **kableExtra** (Zhu 2020) : Used to assist in table formatting
- **bookdown** (Xie 2020) : Used to improve the Markdown document formatting
- **here** (Müller 2020) : Used to simplify file access in a project environment
- **forcats** (Wickham 2022) : Used to sort data tables by frequency

2.4 Variables of Interest

Are there any that are very similar that you nonetheless don't use? Did you construct any variables by combining various ones? What do the data look like? Plot the actual data that you're using (or as close as you can get to it). Discuss these plots and the other features of these data.

3 Simulation

```
set.seed(318)
population <- 13810
sim_dataset <- tibble(age_group = sample(x = c(1:6),
                                         size = population,
                                         replace = TRUE),
                     gender = sample(x = c(1:2),
                                     size = population,
                                     replace = TRUE))

sim_dataset <- sim_dataset %>%
  mutate(age_group=case_when(age_group==1~"15 to 24 years",
                             age_group==2~"25 to 34 years",
                             age_group==3~"35 to 44 years",
                             age_group==4~"45 to 54 years",
                             age_group==5~"55 to 64 years",
                             age_group==6~"65 years and over",),
         gender=case_when(gender==1~'Male',
                          gender==2~'Female'))
```

4 Modelling

Linear regression modelling

5 Results

6 Discussion

7 Appendix

7.1 Datasheet

References

- Müller, Kirill. 2020. *Here: A Simpler Way to Find Your Files*. <https://CRAN.R-project.org/package=here>.
- Ontario Data Documentation, Scholars Portal, Extraction Service and Infrastructure (ODESI). 2020. *Canadian Internet Use Survey, 2018 Study Documentation*.
- R Core Team. 2020. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. <https://www.R-project.org/>.
- Statistics Canada. 2019. *Canadian Internet Use Survey (CIUS)*. <https://www23.statcan.gc.ca/imdb/p2SV.pl?Function=getSurvey&Id=1196799>.
- Wickham, Hadley. 2016. *Ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York. <https://ggplot2.tidyverse.org>.
- . 2022. *Forcats: Tools for Working with Categorical Variables (Factors)*. <https://forcats.tidyverse.org>.
- Wickham, Hadley, Mara Averick, Jennifer Bryan, Winston Chang, Lucy D'Agostino McGowan, Romain François, Garrett Grolemond, et al. 2019. "Welcome to the tidyverse." *Journal of Open Source Software* 4 (43): 1686. <https://doi.org/10.21105/joss.01686>.

- Xie, Yihui. 2020. *Bookdown: Authoring Books and Technical Documents with r Markdown*. <https://github.com/rstudio/bookdown>.
- . 2021. *Knitr: A General-Purpose Package for Dynamic Report Generation in r*. <https://yihui.org/knitr/>.
- Zhu, Hao. 2020. *kableExtra: Construct Complex Table with 'Kable' and Pipe Syntax*. <https://CRAN.R-project.org/package=kableExtra>.