

Experimenting with Mapping the Digital Divide in Canada

Assessing digital divide with insufficient publicly available data

Zein Hajj-Ali, Electrical and Computer Engineering MSc.
Leslie Zijun Rao, Communications and Media Studies, MA.
Supervisor: Michael Genkin, DATA5000 Data Science Seminar

1. Introduction

- **Digital Divide** is a complex social issue, referring to the unequal access to the Internet, so as well as to knowledge & opportunities. Lacking broadband infrastructures, quality network services, and proper digital devices can all result in this inaccessibility.
- The state has devised a series of program to close this gap. However, it is often still an unmonitored problem, as the details of digital divide and many of its related factors are not officially published, meanwhile inaccessibility issues are reported by Canadians across the country.
- **Research questions:**
 - How to map the digital divide in Canada?
 - What are the barriers preventing us from obtaining an accurate assessment?

2. Methodology

(A) Availability of Internet Access
The operational definition & the subject to assess

(B) Dataset Exploration
Identify potentially useful datasets related to access Internet availability

(C) Processing Data
Data cleaning, compilation, correlations to assess who & where tend to be more vulnerable to digital divide

(D) Discussion
Reflect on the mapping process, discuss the limitations of this project.

3. Use of Datasets & Variables

National Broadband Data (NBD)

- Geographic location (25Km² Hexagonal areas)
- Wireless/wired network
- Highest internet speed availability
- Population of area

Note: **Pseudo Household Demographic Distribution (PHH)** was used by CTRC to merge the NBD with the demographic information

2016 Census - Ontario

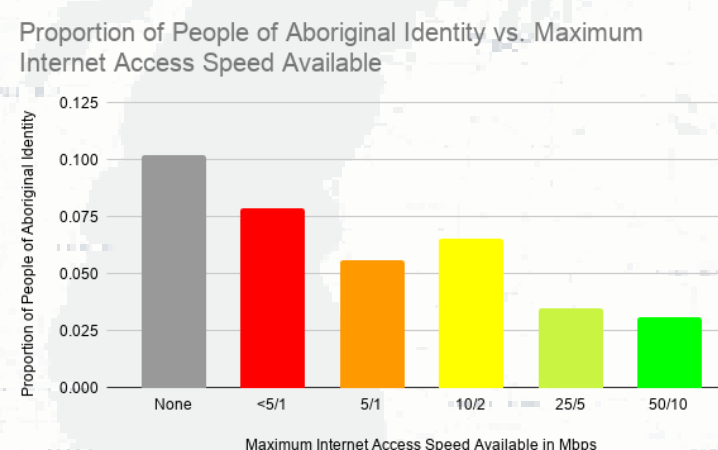
- Dissemination area (geographic)
- Aboriginal identity
- Average age
- Citizenship
- Median income (household and individual)
- Population
- Number of private dwellings (regularly occupied and otherwise)
- Visible Minority

4. Processing Data

- Dataset size: **1.3 million records** of Ontario extracted from *the NBD*, as does *the PHH*. Over **27.3 million records** for Ontario extracted from *the 2016 Census*
- **Changing values** to retain information & increase readability
- **Dimensionality reduced** to make processing easier & less computationally intensive
- Resulting dataset **merged with 2016 Census data** along **the Dissemination Area** to allow access to demographic data
- Each records demographic data proportioned according to the population in that area.
- **Pearson, Kendall & Spearman correlation tests** run between every demographic category and speeds (Wireless, Wired, Either)
- **Averages of the demographic data** were also calculated for each level of access

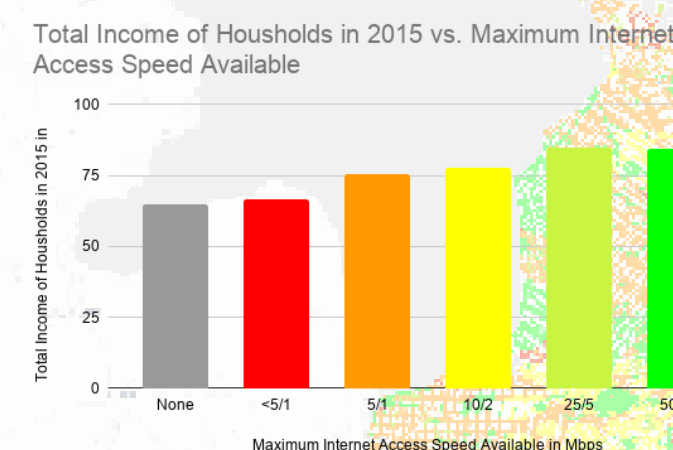
5. Data Outcomes: Lacking Detailed Results

- The correlation tests **don't show a particularly strong linear or monotonic relationship** between the census variables and the level of available internet access.
- From the variables that have some correlation, we can see that their **proportional averages have a slight trend** when compared to the level of access.



- The areas with a higher level of internet access have a lower average number of people that identify as aboriginal.

- Furthermore, there is a very slight downward trend when looking at age with relation to the level of internet access, while Canadian citizenship has a slight upward trend.



- The areas with a higher level of internet access have a higher average household income

6. Reflection: Why The Lack?

- **General absence of usable/related datasets**
 - E.g. **Retail Fixed Internet Sector and Broadband Availability Data** (CRTC & ISED, 2018): contains information concerning a.) broadband service availability b.) Residential Internet service subscribers
 - E.g. **2018 Canadian Internet Use Survey Microdata** (by ISED): contains variables concerning a.) access to digital devices b.) Internet access location
 - However, geographic unit of both datasets are by provinces/territories, therefore are not applicable to analyze digital divide more detailed.
- **Different ID formats used in different datasets**
 - The PHH must be used to connect the IDs of the NBD to the Geocode used in the Census data.
 - However, Hexagonal areas used to map NBD & PHH doesn't correctly correspond to the shape of Census division areas. Therefore, the result rendered from matching the two datasets might miss or misrepresent the detailed information.
- **Misleading variable labelling in the Census data**
 - Some variables are labeled with the same name, despite they are the subcategories under different main categories. Therefore, there is no way of distinguishing between different census variables of the same record.

7. Conclusion

The public need wider scope of open data that speaks to the the digital divide in Canada.

The data with provinces/territories geographic unit cannot provide new information to the public, besides the already-well-circulated knowledge of the rural/urban digital disparity.

This inaccessible knowledge about digital divide not only imposes a barrier towards the public understanding of the issue, but also hampers the generation of community-based solutions.

References

Stoiciu, A. (no date). The Role of e-Governance in Bridging the Digital Divide. *United Nations*.
Statistics Canada. (2020). National Broadband Data. Statistics Canada. *The Government of Canada*.
Statistics Canada. (2016). 2016 Census. *Statistics Canada*.
Innovation, Science and Economic Development Canada. (2017). Pseudo-Household Demographic Distribution. *The Government of Canada*.

Acknowledgement

Thanks for professor Michael Genkin for his kind support and feedback.