

Description

Available data

Data description

Data sources and methodology

Funding

License

References



Uncertainty in times of COVID-19: Raw survey data

Fabian Lange, Lars Vilhuber

2020-06-05

Data from a survey of consumer expectations

Description

Since April 24, 2020, Fabian Lange (<http://www.fabianlange.ca/>) and Lars Vilhuber (<https://lars.vilhuber.com>) have been conducting the survey "Uncertainty in COVID-19 times". The survey is a single-question survey focusing on people's anticipation about social distancing rules and firm closures during the 2020 COVID-19 health crisis.

We believe that this information is not otherwise available in a reliable and timely fashion. The information should be usable by policy-makers and researchers, to be included in models of future developments of society and the economy.

Citation

Please cite the data as

Lange, Fabian and Lars Vilhuber. 2020. "Uncertainty in times of COVID-19: Raw survey data [dataset]." Available at <https://labordynamicsinstitute.github.io/covid19-expectations-data> (<https://labordynamicsinstitute.github.io/covid19-expectations-data>) (accessed 2020-06-05).

Please cite this document as

Lange, Fabian and Lars Vilhuber. 2020. "Codebook for: Uncertainty in times of COVID-19: Raw survey data." Available at <https://labordynamicsinstitute.github.io/covid19-expectations-data> (<https://labordynamicsinstitute.github.io/covid19-expectations-data>) (accessed 2020-06-05).

This document is also available in PDF format at <https://labordynamicsinstitute.github.io/covid19-expectations-data/expectations-codebook.pdf> (<https://labordynamicsinstitute.github.io/covid19-expectations-data/expectations-codebook.pdf>).

We will be posting the data on Zenodo shortly. Data should be cite via DOI then.

Available data

Final files

List of files

Normalized files

Auxiliary files

Temporary files

Final files are uploaded after each wave is completed. Filenames in `final` (final/) tagged with geography, language, the question type, and date downloaded:

survey-[geography]-[language]-[question]-[date].xlsx

Data description

Topic	Answer
Geographic Coverage	United States of America, Canada
Time Periods	2020-04-24 - 2020-06-05
Date of Collection	2020-04-24 - 2020-06-05
Unit of Observation	Individual
Description of Variables	User_ID, Time_UTC, Survey_Completion, Publisher_Category, Gender, Age, Geography, Weight, Question_1_Answer, rt_Q1_ms

Reference period

The survey asks about point-in-time expectations. A new wave is launched every Friday. The list provides the dates of collection for each wave. Currently, data are available covering the period between 2020-04-24 and 2020-06-05.

Data Dictionary

Q1: Answer to primary question

This field captures the answer to the sole question of each survey, where answers differ across geographic scope (geotag), and languages. A consolidated (standardized) distribution is shown below, using the standardizer mapping (auxiliary/standardize_values.xlsx).

Standardized distribution	People, Canada, English	Business, Canada, French	People, Canada, French
Business, US, English	People, US, English		

The following tabulations are of unweighted data.



Question type

The actual question asked is encoded in the `tag` variable on normalized files, and differs by geographic target (`geotag`). On the original files, geographic target is not identifiable except through the file name, and the question text is on the "Overview" tab. On the normalized files, the variables `tag` and `geotag` allow to map back to the actual question:

Geographic target

Notes

Encoded in `geotag` on normalized files, and specifies the two-letter geocode (country or postal abbreviation) as targeted on the Google Survey platform. Note: `geotag = qc` also identifies the surveys that used the app.

```
## `summarise()` ungrouping output (override with `.groups` argument)
```

geotag	count	percent
canada	21845	36.16
ny	2003	3.32
qc	6492	10.75
us	30067	49.77

Age

```
## `summarise()` ungrouping output (override with `.groups` argument)
```

Age	count	percent
18-24	7612	12.60
25-34	9781	16.19
35-44	8589	14.22
45-54	7501	12.42
55-64	7496	12.41
65+	7227	11.96
Unknown	12201	20.20

Gender

```
## `summarise()` ungrouping output (override with `.groups` argument)
```

Gender	count	percent
Female	23048	38.15
Male	26000	43.04
Unknown	11359	18.80

Geography

Geography is as coded by Google Surveys. Precision may vary, having country, region, province, and sometimes city. Note that this may be different from the targeted geography.

Detailed geography	Country	Region	States/Province	City
--------------------	---------	--------	-----------------	------

The variable `Geography` corresponds to the geography as captured and recorded by Google. All other geography variables are derived from this variable, and are only available on the normalized files.

```
## `summarise()` ungrouping output (override with `.groups` argument)
```

Weight

See elsewhere in this document how weights are computed.

Response Time

Publisher Category

```
## `summarise()` ungrouping output (override with `.groups` argument)
```

Not tabulated

- `User_ID`
- `Time_UTC`
- `Survey_Completion`

Data structure

Data files are available for each completed cycle of the survey, in general once a week, and are stored under `final` (`final/`). Data from the preliminary study (assessing the questionnaire design) is stored under `preliminary` (`preliminary/`). We may make available data before the survey is completed for each cycle, under `temporary` (`temporary/`), however, once the final version from that cycle is available, these are deleted (this directory will be empty on Zenodo).

Data format

Native format is Office Open XML (XLSX, ECMA International (2016)). Normalized files are available in Stata and R formats.

Files are provided as downloaded from Google Surveys. Each file has 4 tabs.

Overview

Lists the questions asked by the client, in this case Lange and Vilhuber, as well as a survey ID.

Topline

This tab contains a weighted summary of the responses to the questions (similar to the above summary).

Complete responses

This tab contains the actual microdata for any complete responses. Note that for a single-question survey, this is identical to the "All responses". A complete response might have a weight of zero.

All responses

All responses, whether complete or not, are recorded on this tab. In the case of a single-question survey, this is identical to the "Complete responses" tab.

Data sources and methodology

Target population

- All Canadians aged 18 and older from the ten provinces and three territories are eligible to participate.
- All US residents aged 18 and older are eligible to participate.

Instrument design

Each individual is asked one of two questions: how long they expect "social distancing rules" or "business closures" to remain in effect:

- How much longer do you expect social distancing rules (restrictions on gatherings, stay-at-home rules) to stay in place in your province/state?
- How much longer do you expect the closure of non-essential businesses to stay in place in your province/state?

Five response choices are offered:

- "less than 1 month",
- "1-2 months",
- "2-3 months",
- "3-6 months",
- "more than 6 months".

An additional answer allows respondents to affirm that "such measures are not implemented in their province/state". See questionnaires for visual representation of the questions.

Questionnaires



Data collection

Data is collected via Google Surveys. For English-language surveys, data is collected via a web form. For French-language surveys, the Android Google Survey app is used, as web-collection in French is not possible via Google Surveys. See Sostek and Slatkin (2018) and Google (2020) for more details.

The survey questionnaire was approved by McGill University Research Ethics Board under REB File # 20-04-070. Exemption was issued by Cornell University Institutional Review Board under Protocol ID# 2004009539.

Sampling

Google Surveys is an online non-probability survey. It uses stratified sampling for collection, based (in the US) on the target internet population from the 2017 Current Population Survey (CPS) Computer and Internet Use Supplement (Sostek and Slatkin 2018; Google 2020).

Data are collected directly from survey respondents.

For each country, we plan to collect 2500 responses per question, per week. For Canada, a French-language variant is fielded. In order to determine the split, we use Statistics Canada statistics on "Language spoken most often at home" by other language(s) spoken regularly at home and age" (Statistics Canada 2017),¹ combining responses for "French" and "French and non-official language" (i.e., no English mentioned).

For 2016, 20.4% spoke French and no English as the language spoken most often at home. We thus target 510 responses via the French-language questionnaire, and 1990 in English.

Imputation

All demographics are imputed by Google Surveys, if collected via web. Demographics for respondents via the app are collected through the app.

Weighting

Weights are provided by Google Surveys, based on the imputed demographics. For the US, the US Census Bureau's Current Population Survey (CPS) Computer and Internet Use Supplement (<https://www.census.gov/programs-surveys/cps/technical-documentation/complete.2017.html>) is used (currently the 2017 version). For Canada, Google (2020) points to a "combination of government data and internal Google data sources." Google uses post-stratification weighting to align the weighted demographics with the target population.

Quality evaluation

A preliminary survey was conducted to allow for choice of either a two-question variant, or a one-question variant that included both social distancing and business closures ("How much longer do you expect social distancing rules (restrictions on gatherings, closure of non-essential businesses, stay-at-home rules) to stay in place in your province?"). See "Uncertainty in times of COVID-19: Choosing whether to ask 1 or 2 questions (evaluation.md)" for more information.

Privacy and disclosure control

Privacy and disclosure control are described in Google (2020). For most respondents, no direct or indirect identifiers are collected, and are imputed based on other information available to Google, but not the sponsors of the survey.

Response rates

The specific response rates are not known. Google (2020) reports response rates in general for this type of data collection.

Funding

We acknowledge generous funding by Lange's Canada Research Chair in Labour and Personnel Economics (<https://www.chairs-chaires.gc.ca/chairholders-titulaires/profile-eng.aspx?profileid=3277>), and by the Cornell Atkinson Center for Sustainability (<https://www.atkinson.cornell.edu/>) under its "Rapid Response Fund" program.

License

These data are licensed under a Creative Commons Attribution-NonCommercial 4.0 International (<https://creativecommons.org/licenses/by-nc/4.0/>) license. See citation for attribution.

References

ECMA International. 2016. "Standard ECMA-376: Office Open XML File Formats." <https://www.ecma-international.org/publications/standards/Ecma-376.htm> (<https://www.ecma-international.org/publications/standards/Ecma-376.htm>).

Google. 2020. "Methodology Google Surveys." <https://support.google.com/surveys/answer/6189786> (<https://support.google.com/surveys/answer/6189786>).

Sostek, Katrina, and Brett Slatkin. 2018. "How Google Surveys Works." Whitepaper. Google. https://services.google.com/fh/files/misc/white_paper_how_google_surveys_works.pdf (https://services.google.com/fh/files/misc/white_paper_how_google_surveys_works.pdf).

Statistics Canada. 2017. "Language Highlight Tables, 2016 Census." Catalogue 98-402-X2016005. Statistics Canada. <https://www12.statcan.gc.ca/census-recensement/2016/dp-pd/hlt-fst/lang/Table.cfm?Lang=E&T=31&Geo=00> (<https://www12.statcan.gc.ca/census-recensement/2016/dp-pd/hlt-fst/lang/Table.cfm?Lang=E&T=31&Geo=00>).

1. Table can be downloaded from here (https://www12.statcan.gc.ca/census-recensement/2016/dp-pd/hlt-fst/lang/Tables/Download/_file.cfm?Lang=E&T=31&Geo=00&SP=1&view=1&age=1&rl=1&OFT=csv).↩