Summary of Recodes

Simon J. Kiss, Matthew J. Arp and Scott Lu

09/03/2021

## Warning: Unknown or uninitialised column: `prov`.

## Introduction

This document tracks and summarizes basic data wrangling and recodes conducted at LISPOP for the 2021 Canadian Public Health Workforce and General Population Survey. The purpose of this document is to facilitate multi-institutional collaboration. LISPOP staff will take on the task of basic variable recodes and export data files for analysis by project researchers so that they can dedicate their attention to substantive analyses.

## Scientific Literacy Questions

We have recoded variables Q14\_1, Q14\_2, Q14\_3 and Q14\_4 into variables know1, know2, ‘know3’ and know4 such that respondents were coded 1 if they were correct and 0 if they were incorrect.

If Q14\_1 = 1 or 2 (Definitely or Probably false), it is correct “Ordinary tomatoes do not contain genes, while genetically modified tomatoes do.”

If Q14\_2 = 1 or 2 (Definitely or Probably false), it is correct “Antibiotics kill viruses as well as bacteria.”

If Q14\_3 = 3 or 4 (Definitely or Probably true), it is correct “The cloning of living things produces genetically identical copies.”

If Q14\_4 = 3 or 4 (Definitely or Probably true), it is correct “All plants and animals have DNA.”

We have also provided a variable mean\_know that has the proportion of correct responses (so it runs from 0 to 1) each respondent got.

## Cognitive Reflection test Questions

We have recoded variables Q18\_1, Q19\_1, Q20\_1 and Q21\_1 into variables crt1, crt2 etc.such that respondents were coded 1 if they got the correct response and 0 if they were incorrect.

CRT Question Q18\_1 coded 1 when answer (X) is a variation of “Second Place” “If you’re running a race and you pass the person in second place, what place are you in?”

CRT Question Q19\_1 coded 1 when answer (X) is a variation of “Eight” “A farmer had 15 sheep and all but 8 died. How many are left?”

CRT Question Q20\_1 coded 1 when answer (X) is a variation of “Emily” “Emily’s father has three daughters. The first two are named April and May. What is the third daughter’s name?”

CRT Question Q21\_1 coded 1 when answer (X) is a variation of “None” “How many cubic feet of dirt are there in a hole that is 3’ deep x 3’ wide x 3’ long?”

You can see the some fo the results of this here:

## # A tibble: 2,207 × 8  
## Q18\_1 Q19\_1 Q20\_1 Q21\_1 crt1 crt2 crt3 crt4  
## <chr> <chr> <chr> <chr> <dbl> <dbl> <dbl> <dbl>  
## 1 second 8 june 27 1 1 0 0  
## 2 First 7 June 3’ 0 0 0 0  
## 3 Second 7 Emily Idk 1 0 1 0  
## 4 1 8 Emily 16 0 1 1 0  
## 5 first 8 Emily 27 0 1 1 0  
## 6 first 7 June 3' 0 0 0 0  
## 7 first 7 June 27 0 0 0 0  
## 8 First 8 Emily 9 0 1 1 0  
## 9 second 15 Emily 3 1 0 1 0  
## 10 1st 8 emily 27 0 1 1 0  
## # … with 2,197 more rows

We have also provided a variable mean\_crt that is the proportion of correct responses, again scaled 0 to1.

## Most Important Problem in Public Health

We have provided more meaningful variable names (e.g. Obesity, Vaccine Hesitancy , etc. ) and *appended* these to the dataset. The original variables are Q1\_1 through Q1\_9. In addition, because respondents were given the option to add other open ended responses, we have gone through these and coded these into some meaningful categories. The categories are currently these:

1 = COVID-19

2 = Public health messaging/gov’t handling of COVID

3 = Healthcare (access to care, short supply)

4 = Long term care

5 = Chronic disease (cancer, heart disease)

6 = Mental health

7 = Access to housing and food

8 = Drug abuse

9 = Inequality

10 = Economy

11 = Corrupt gov’t

12 = Climate change/environmental

13 = Abortion

14 = Reliance on meat

15 = Domestic abuse

16 = Misinformation

17 = Internet addiction

18 = Don’t know

## Trust

Q33-36 were scaled from 0-1 using the following code:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| pos | variable | label | col\_type | levels | value\_labels |
| 103 | Q33 | Q33 How much of the time do you think you can trust the government in Ottawa to do what is right: | dbl+lbl | NA | [1] Never |
| 103 | Q33 | Q33 How much of the time do you think you can trust the government in Ottawa to do what is right: | dbl+lbl | NA | [2] Some of the time |
| 103 | Q33 | Q33 How much of the time do you think you can trust the government in Ottawa to do what is right: | dbl+lbl | NA | [3] About half of the time |
| 103 | Q33 | Q33 How much of the time do you think you can trust the government in Ottawa to do what is right: | dbl+lbl | NA | [4] Most of the time |
| 103 | Q33 | Q33 How much of the time do you think you can trust the government in Ottawa to do what is right: | dbl+lbl | NA | [5] Always |
| 104 | Q34 | Q34 Do you think that people in government waste: | dbl+lbl | NA | [1] A lot of the money we pay in taxes |
| 104 | Q34 | Q34 Do you think that people in government waste: | dbl+lbl | NA | [2] Waste some of it |
| 104 | Q34 | Q34 Do you think that people in government waste: | dbl+lbl | NA | [3] Or don’t waste very much of it |
| 105 | Q35 | Q35 Would you say the government is: | dbl+lbl | NA | [1] Pretty much run by a few big interests looking out for themselves |
| 105 | Q35 | Q35 Would you say the government is: | dbl+lbl | NA | [2] Run for the benefit of all the people |
| 106 | Q36 | Q36 Generally speaking, would you say that most people can be trusted, or that you need to be very careful when dealing with people? | dbl+lbl | NA | [1] Most people can be trusted |
| 106 | Q36 | Q36 Generally speaking, would you say that most people can be trusted, or that you need to be very careful when dealing with people? | dbl+lbl | NA | [2] You need to be very careful when dealing with people |

full %>%  
 mutate(  
 trust\_ottawa=case\_when(  
 #Q33 is scaled from 0-1  
 Q33 == 1 ~ 0,  
 Q33 == 2 ~ .25,  
 Q33 == 3 ~ .5,  
 Q33 == 4 ~ .75,  
 Q33 == 5 ~ 1,  
 ),  
 trust\_waste=case\_when(  
 #Q34 is scaled from 0-1  
 Q34 == 1 ~ 0,  
 Q34 == 2 ~ .5,  
 Q34 == 3 ~ 1,  
 ),  
 trust\_interests=case\_when(  
 #Q35 is scaled from 0-1  
 Q35 == 1 ~ 0,  
 Q35 == 2 ~ 1,  
 ),  
 trust\_people=case\_when(  
 #Q36 is scaled from 0-1  
 Q36 == 1 ~ 0,  
 Q36 == 2 ~ 1,  
 )  
 )->full

#### Covid case counts

## COVID Case counts

Dr. Tim Gravelle (Survey Monkey and now affiliated with Laurier) helped me merge respondents` FSA with their health regions. With respondents health regions, I was able to draw on [this](https://github.com/ccodwg/Covid19CanadaData) data-set to get several variables of COVID19 severity that you will see in the In the recoded data. They include:

|  |
| --- |
| x |
| Comm\_Name |
| HR |
| health\_region |
| HR\_UID |
| HR\_NAME |
| FSA\_population |
| area.km2 |
| pop.km2 |
| province |
| province\_fsa\_bad |
| date\_report |
| cases |
| cumulative\_cases |
| deaths |
| cumulative\_deaths |
| numtotal\_last14 |
| numdeaths\_last14 |
| avgtotal\_last14 |
| avgdeaths\_last14 |
| numtotal\_last7 |
| numdeaths\_last7 |
| avgtotal\_last7 |
| avgdeaths\_last7 |
| province\_full |
| province\_short |
| health\_region\_esri |
| health\_region\_population |
| case\_trend |

They are pretty self-explanatory, although the variables Comm\_Name is just the generic name for the community the health region was in; HR\_UID is a Statistics Canada identifier for health regions, HR\_Name and health\_region are both versions of the health region name from the different datasets involved in merging this.But I will point out **case\_trend** . It is just the result of the avgtotal\_last7 divided by the avgtotal\_last14. So, I’m trying to capture the trend of COVID severity. So, if the average number of cases in the respondent’s health region was the same in the last seven days as in the previous 14 days, then they would get a 1, so their health region would be stable. But, if the seven day average was double the average 14 day total, then they would get a two, cases would be rising.

## Province FSA mismatch

As part of the process of merging our data with the health region file in order to access the covid19 case counts, it was discovered that 8 cases feature reported provinces that do not match the reported FSA. These have been flagged in the variable province\_fsa\_bad. For the moment, the assigned health region and rural variable assume that the reported FSA is correct and the reported province is a mistake.

# Rural

Matt can you add in a section here describing how we added in the rural variable?