Generating Qualtrics Dictionaries

Working with qualtrics files can be tricky because surveys can utilizes question families that contain multiple levels of responses. As a result, variables belong to variable groups. The raw qualtrics names are not very helpful, so dictionary crosswalk files have been create to facilitate data preparation workflow.

Example Variable Family:

QUALTRICS VARIABLE NAME	NEW NAME	LABEL
Staff#1_1_1	Staff_Fulltime_2021	Number of full time staff in 2021
Staff#1_2_1	Staff_Parttime_2021	Number of part time staff in 2021
Staff#1_3_1	Staff_Boardmmbr_2021	Number of Board Members in 2021
Staff#2_1_1	Staff_Fulltime_2022	Number of full time staff in 2022
Staff#2_2_1	Staff_Parttime_2022	Number of part time staff in 2022
Staff#2_3_1	Staff_Boardmmbr_2022	Number of Board Members in 2022

Some utility scripts have been written to extract variable dictionary elements from qualtrics survey file exports and convert them into a basic crosswalk file.

```
# source( "../data-dictionaries/R/00-data-processing-utils.R" )

URL <- "https://raw.githubusercontent.com/UrbanInstitute/nccs-nptrends/main/data-dictionaries/
source( URL )

###########
###########   DATA DICTIONARY
###########

# USE LEGACY = TRUE
# FOR THE YEAR 2 DATA DICTIONARY,
# FALSE FOR SUBSEQUENT YEARS

d <-
  read_survey(
    "data-raw/wave-02-qualtrics-download-29mar23.csv",
    legacy = T )
```

```r
head( as.data.frame(d) )

# exports data dictionary

dd <- extract_colmap( d )

# add group variables
# and factor labels

dd <-
  dd %>%
  mutate( group_name = append_groups(qname) ) %>%
  group_by( group_name ) %>%
  mutate( group_n = n(),
          group_name = ifelse( group_n > 1, group_name, "" ),
          is_group = ifelse( group_n > 1, "1", "0" ),
          group_levels = ifelse( group_n > 1, get_categories(description), "" ) ) %>
  ungroup() %>%
  mutate( type = ifelse( group_n > 1, "factor", "" ) ) %>%
  select( qname, is_group, group_name, group_n, group_levels,
          description, main, sub )


write.csv( dd, "../data-dictionaries/dd-nptrends-wave-02.csv" )
```

After generating the skeleton of the crosswalk file from the script, it would be completed by a research assistant by providing information for fields marked **user** in the table:

DD VARIABLE	DESCRIPTION	SOURCE
q	question number (order)	qualtrix
vname_raw	variable name (from qualtrix export)	qualtrix
vname	variable name (final)	user
vlabel	variable label	user
type	data type (numeric, character, factor, logical, date)	user
group	group name	r script
group_lev1	factor levels	user
group_lev2	second factor level for double-grouped variables (e.g. finances_1_1)	user

DD VARIABLE	DESCRIPTION	SOURCE
group_lev_draft	parsed categories (clean up and use as group_lev labels)	r script
add_noise	add noise to this variable to anonymize?	user
description	survey question full	qualtrix
main	survey question sub	qualtrix
sub	survey question response categories (kindof)	qualtrix

These steps need to be completed each year. Most of the completed dictionaries can be reused if questions are not changing, but note that changes to the order of the questions in the survey can change the qualtrics naming conventions, and any new questions added would need documentation.

These dictionary crosswalk files are utilized in subsequent steps.