Exploring Qualtrics Surveys with R

Ambreen Chaudhri

1/13/2023

## R Markdown Installation

This is an R Markdown document. Markdown is a simple formatting syntax for authoring HTML, PDF, and MS Word documents. For more details on using R Markdown see <http://rmarkdown.rstudio.com>.

To install R Markdown, you will need to run the following install packages lines and load library lines:

# install packages  
#install.packages('rmarkdown')  
  
# load packages  
library(rmarkdown)  
  
# load additional packages  
library(qualtRics)  
library(dplyr)  
library(stringr)  
library(vtable)  
#library(stargazer)

## Including Code Chunks and Results

You can embed an R code chunk in the space below. Note that this line will show you both the R command you used and the results.

# Clear Workspace  
rm(list=ls())   
  
# set working directory  
setwd("~/MKTG\_531")  
  
# RUN THIS STEP FIRST TIME ONLY, TO GENERATE ".Renviron" FILE  
#qualtrics\_api\_credentials(api\_key = "INSERT YOUR API KEY HERE",  
# base\_url = "ca1.qualtrics.com",  
# install = TRUE,  
# overwrite = TRUE)  
  
readRenviron("~/.Renviron")  
outpath = "~/MKTG\_531"  
  
# Show all surveys and select one project by name  
surveys <- all\_surveys()  
surveys

## # A tibble: 86 × 6  
## id name ownerId lastModified creationDate isActive  
## <chr> <chr> <chr> <chr> <chr> <lgl>   
## 1 SV\_0H63qPlpEPK7WPs Post-Survey: D… UR\_4OW… 2022-08-10T… 2022-07-21T… TRUE   
## 2 SV\_2gHUqeygiXUF74F Faculty Subscr… UR\_0fk… 2018-10-03T… 2017-07-11T… TRUE   
## 3 SV\_37AJQNTlXAdBUaN Academic Props… UR\_cIS… 2021-05-11T… 2020-06-15T… TRUE   
## 4 SV\_51MlEKTaJKqujrg Academic Props… UR\_cIS… 2022-05-18T… 2021-11-30T… TRUE   
## 5 SV\_6A4GI1rCxDTISF0 Querying Large… UR\_4OW… 2022-11-29T… 2022-11-29T… TRUE   
## 6 SV\_6iepepDgfuyzn5b Emotion and Ul… UR\_3Uf… 2018-03-27T… 2018-03-05T… FALSE   
## 7 SV\_7O3AFgjw9hgFPz8 Demographics S… UR\_1Sb… 2023-01-06T… 2022-07-06T… TRUE   
## 8 SV\_9RGpeLaha178FAV Trust Game X R… UR\_3Xk… 2018-03-28T… 2018-03-13T… TRUE   
## 9 SV\_b3MA3u5En0VMmRE Academic Props… UR\_cIS… 2022-05-18T… 2022-05-18T… TRUE   
## 10 SV\_b3NY495VdBuAwdv Reaction to Mo… UR\_3Xk… 2018-04-11T… 2018-03-06T… TRUE   
## # … with 76 more rows

myproject <- filter(surveys, name=="Demographics Survey - eLab - 2022 pool refresh")  
myproject

## # A tibble: 1 × 6  
## id name ownerId lastModified creationDate isActive  
## <chr> <chr> <chr> <chr> <chr> <lgl>   
## 1 SV\_7O3AFgjw9hgFPz8 Demographics Su… UR\_1Sb… 2023-01-06T… 2022-07-06T… TRUE

myproject$id[1]

## [1] "SV\_7O3AFgjw9hgFPz8"

#Fetch the survey for that project as .RDS file  
mysurvey <- fetch\_survey(surveyID = myproject$id[1],  
 save\_dir = outpath,  
 force\_request = TRUE,  
 verbose = TRUE)

## | | | 0% | |==== | 6% | |============= | 18% | |================== | 25% | |=========================== | 39% | |=============================== | 45% | |================================== | 49% | |========================================= | 58% | |============================================= | 64% | |================================================== | 72% | |======================================================= | 79% | |============================================================= | 87% | |=============================================================== | 89% | |======================================================================| 100%

##   
## ── Column specification ────────────────────────────────────────────────────────  
## cols(  
## .default = col\_character(),  
## StartDate = col\_datetime(format = ""),  
## EndDate = col\_datetime(format = ""),  
## Progress = col\_double(),  
## `Duration (in seconds)` = col\_double(),  
## Finished = col\_logical(),  
## RecordedDate = col\_datetime(format = ""),  
## RecipientLastName = col\_logical(),  
## RecipientFirstName = col\_logical(),  
## RecipientEmail = col\_logical(),  
## ExternalReference = col\_logical(),  
## LocationLatitude = col\_double(),  
## LocationLongitude = col\_double(),  
## `Q2 Birth year` = col\_double(),  
## `Q9 Zip code` = col\_double(),  
## MTurkCode = col\_double()  
## )  
## ℹ Use `spec()` for the full column specifications.

## Warning: 38 parsing failures.  
## row col expected actual  
## 14 -- value in level set < 1 year  
## 16 -- value in level set < 1 year  
## 42 -- value in level set < 1 year  
## 49 -- value in level set < 1 year  
## 71 -- value in level set < 1 year  
## ... ... .................. ........  
## See problems(...) for more details.

nrow(mysurvey)

## [1] 6705

# save the approximate day/time of when you downloaded the survey  
Sys.time()

## [1] "2023-01-12 14:26:38 CST"

download\_time <- format(Sys.time(), "%a %b %d %X %Y")  
download\_message <- paste("I downloaded the", myproject$name, "survey data on:", download\_time, sep=" ")  
download\_message

## [1] "I downloaded the Demographics Survey - eLab - 2022 pool refresh survey data on: Thu Jan 12 02:26:38 PM 2023"

# save download time to log file  
logfile\_path <- paste(outpath, "/logfile.text", sep="")  
writeLines(download\_message, logfile\_path)  
  
  
# Read RDS file into dataframe "mysurvey", and save copy as CSV  
rdspath <- paste(outpath, "/", myproject$id[1], ".rds", sep="")  
mysurvey <- readRDS(file = rdspath)  
csvpath <- paste(outpath, "/", myproject$id[1], ".csv", sep="")  
ret <- write.csv(x=mysurvey, file=csvpath)  
  
  
# Fix problem of spaces in column names!  
names(mysurvey) <- str\_replace\_all(names(mysurvey),c(" "=".", ","=""))  
colnames(mysurvey)

## [1] "StartDate" "EndDate" "Status"   
## [4] "IPAddress" "Progress" "Duration.(in.seconds)"  
## [7] "Finished" "RecordedDate" "ResponseId"   
## [10] "RecipientLastName" "RecipientFirstName" "RecipientEmail"   
## [13] "ExternalReference" "LocationLatitude" "LocationLongitude"   
## [16] "DistributionChannel" "UserLanguage" "Consent.text.6"   
## [19] "Q1.Gender" "Q2.Birth.year" "Q3.Hispanic"   
## [22] "Q4.Race\_1" "Q4.Race\_2" "Q4.Race\_3"   
## [25] "Q4.Race\_4" "Q4.Race\_5" "Q4.Race\_6"   
## [28] "Q5.marital" "Q6.Eng.first" "Q7.Eng.skill"   
## [31] "Q8.Home.lang" "Q8.Home.lang\_5\_TEXT" "Q9.Zip.code"   
## [34] "Q10.Citizenship" "Q11.Years.in.US" "Q12.School"   
## [37] "Q13.Edu.level" "Q14.Parent.edu" "Q15.Employment"   
## [40] "Q16.Household.inc" "Q17.Size.household" "Q18.Under.18.house"   
## [43] "Q19.Pol.party" "Q20.Pol.orientation" "Q21.Religious"   
## [46] "Q22.Religious.group" "Q23.Pet" "Q24.Surroundings"   
## [49] "MID" "MTurkCode"

### FILTERING  
# Eliminate test responses  
mysample <- filter(mysurvey, Status=="IP Address")  
nrow(mysample)

## [1] 6468

# Eliminate outliers for response time  
quant\_duration <- quantile(mysample$`Duration.(in.seconds)`, probs=seq(0,1,.05), na.rm=FALSE, names=FALSE)  
quant\_duration

## [1] 2.00 96.00 112.00 122.00 131.00 141.00 150.00  
## [8] 158.00 170.00 180.00 191.00 204.00 220.00 237.00  
## [15] 257.00 286.00 325.00 384.00 480.00 690.65 377912.00

mysample <- filter(mysample, `Duration.(in.seconds)` > quant\_duration[2])  
mysample <- filter(mysample, `Duration.(in.seconds)` < quant\_duration[20])  
nrow(mysample)

## [1] 5816

# Proportion of respondents who are female  
females <- filter(mysample, Q1.Gender == "Female")  
pct\_female <- nrow(females) / nrow(mysample)  
pct\_female

## [1] 0.5428129

# Create a table  
partyXgender <- table(mysample$Q19.Pol.party, mysample$Q1.Gender,   
 exclude = c(NA,"I do not identify with either of these.",   
 "I do not identify with any of these"))  
partyXgender

##   
## Male Female  
## Democratic 1194 1571  
## Republican 742 811  
## An independent or third party 535 577

## Embedding the results of a R code within text

R markdown also allows you to embed the results of some R code within your text. For instance:

The proportion of females respondents was 54.3 percent in our sample.

# Creating a PDF or HTML document.

Finally, when you click the **Knit** button a document will be generated that includes both content as well as the output of any embedded R code chunks within the document