

Chapter 3: Finding Data

Answers to Exercises

Brian Fogarty

Answers to Exercises 1

(This is an example of finding and reading in an **open access** dataset from the UK Data Service (UKDS).)

There are several ways to find open access data on the UKDS' site. One simple way is to do a blank search where all the available datasets on UKDS are returned. Then on the left-hand menu select **Access: > Open**. This will return datasets that are open access.

For the example, I chose to download *Migrant Children in a Transforming Europe Survey Data, 2020-2021*.¹ On the study's site, we next click on **Access Data**. On the next page, we click on **Download from ReShare repository** which take us to the site that has the data and documentation.

The data file is in a SPSS format (.sav) and is named `MiCREATE_WP567_UK_MiCREATE_Survey_data.sav`. The documentation are in MS Word formats (.docx). I save all of the files to my working directory `"C:/QSSD/Exercises/Chapter 3 Exercises"`.

Now I'll read-in the data using the `read_sav()` from the **haven** package.

```
setwd("C:/QSSD/Exercises/Chapter 3 Exercises")

library(tidyverse)
library(haven)

data <- read_sav("MiCREATE_WP567_UK_MiCREATE_Survey_data.sav")
```

In the **Environment** window, I have an object called `data` that has 463 observations and 85 variables. Using `glimpse(data)`, the data looks to be correctly read-in, but I would want to go through all of the documentation before proceeding with data wrangling and analysis.

Answers to Exercises 2

(This is an example of finding and reading in a dataset from the Harvard Dataverse website.)

We can search or browse for relevant and interesting data on the Harvard Dataverse site. For this example, I went to *Comparative Political Studies*'s Dataverse and browsed the articles' replication materials. I chose to download the files for *Privileging one's own? Voting patterns and politicized spending in India*.²

Here, I'll read in the data file `Jensenius_Chhibber_2022_LS15.RData`, which, obviously, is saved as in `.RData` format. Given the format, we can simply use the `load()` function to read-in the data.

¹Arun, S., Szymczyk, A., Batool, F. (2022). Migrant Children in a Transforming Europe Survey Data, 2020-2021. [data collection]. UK Data Service. SN: 855768, DOI: 10.5255/UKDA-SN-855768

²Jensenius, Francesca; Pradeep Chhibber, 2022, "Replication Data for: Privileging one's own? Voting patterns and politicized spending in India", <https://doi.org/10.7910/DVN/RSVRFA>, Harvard Dataverse, V1.

```
load("Jensenius_Chhibber_2022_LS15.RData")
```

In the **Environment** window, I have an object called DTA that has 238,792 observations and 32 variables. Using `glimpse(DTA)`, the data looks to be correctly read-in, but I again would want to go through all of the documentation before proceeding with data wrangling and analysis.