Warming Up with R

Day 2

DIMA Units of Regional Bureaux: EHAGL, SA and WCA

15 February 2022

Today will be about:

- Understanding basics
- Recap of packages
- Data import and export
- Organising your project

1. Understanding basics

How R works?

- R stores objects in memory (RAM) of your computer temporarily.
  - Objects appear in the Environment pane.
  - Objects could be anything.
  - Focus on 'data objects': vector, list, data frame, etc.

Basic code writing in R

Recap of packages

"Packages are the fundamental units of reproducible R code."

- Install by writing a code

```
install.packages("tidyverse")
```

- Install by clicking in RStudio
- Let R know the package to use

```
library(tidyverse)
```

In RStudio, documents of packages are accessible in Help tab.

```
?`tidyverse`
```

Data import and export

Setting working directory

- Absolute and relative path: Recommended to use relative path.
- `here()` package

```
install.packages("here")
library(here)
here()
```

Data import and export

Import

- Use the dropdown menu (File > Import Dataset)
- Navigate in Files tab
- **Write a script with packages and functions**

The packages we need are: `readxl` to read excel, `writexl` to export data frame to excel, `haven` to read data in various formats such as .csv, Stata (.dta) and SPSS (.sav), and `pdftools` to read pdf.

Data import and export

Import

- csv: Refugee data of ASR 2020 from Population Statistics Reference
- Stata (.dta): Survey data, Understanding the socioeconomic conditions of refugees in Kalobeyei, Kenya 2018, from RIDL
- excel: Resettlement data from Resettlement Statistics Report
- pdf: South Africa government data of PoCs

Export

Data import and export

Import directly from packages

- UNHCR open data from Refugee Data Finder

```
install.packages("remotes") # install a required package
library(remotes)
remotes::install_github('unhcr/unhcrdatapackage') # install unhcrdatapackage
library(unhcrdatapackage)

popdata <- unhcrdatapackage::end_year_population_totals_long
head(popdata)
```

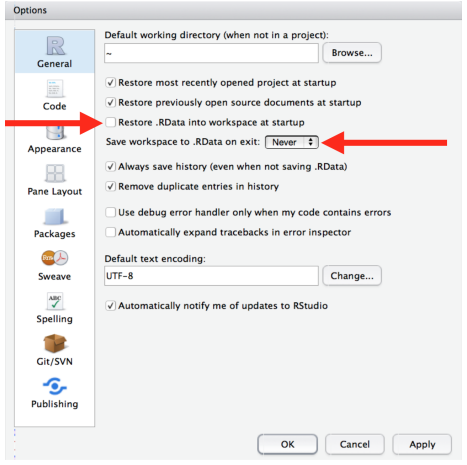
- World Bank open data

```
install.packages("wbstats")
library(wbstats)

wb_ind <- c("NY.GDP.PCAP.CD", "NY.GDP.PCAP.KD.ZG", "FP.CPI.TOTL.ZG", "SI.POV.DDAY")
wb_df <- wb_data(indicator = wb_ind, country = "COD",
                 start_date = 1960, end_date = 2021)
```

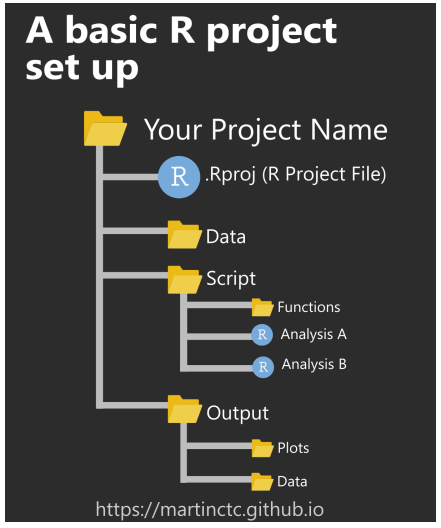
Organising your project

- Not recommended to store your workspace as .RData
  - Change it in Tools > Global Options in RStudio



Organising your project

- Your project will be usually consisted of:
  - Data
  - Script
  - Results
  - Auxiliary files



- Recommend to start with Github for version control

Further reading

R for Beginners

R for Data Science

Project-oriented workflow

Tomorrow (Day 3) will be about:

- Data manipulation: Selecting column, filtering, sorting, grouping, summarise, rename, create new columns, and so on.