Data science project

Alireza Sadeghi

Introduction

-Objective: Reproduce the graph and table using Rstudio. These figures show “Labor force participation rate%female” and the table shows “the annual household labor income”
-Tools: These figures have been replicated in Rstudio by packages called ggplot2(**Wickham 2016**) , dplyr, (**Wickham et al. 2014**), modelsummary, -(**Arel-Bundock 2022**), kableExtra,(**Zhu 2017**).
-Result: To understand the female’s labor force participation and also the annual household labor income

Introduction

-We used ggplot2 ,dplyr for replicating the graph and we used modelsummary, dplyr to replicate the table.
-This paper (**Choi, Kim, and Lim 2025**)asks this question that do people with stronger strategic thinking abilities earn more or perform better economically? And does this effect go beyond individuals to benefit households?
-This has been study on 3000 people in South Korea and Singapore.

Challenges

_ Old version of Rstudio
_ Pushing to github
_ Making a reference list

Original graph

Replicated graph

Replication code of graph

```r
if (!require(pacman)) install.packages("pacman")
# Load required packages
library(ggplot2)
library(dplyr)

# Simulated example data (replace with actual dataset if available)
set.seed(123)  # For reproducibility
data <- data.frame(
  year = rep(2000:2020, 7),
  flfp = runif(21 * 7, min = 52, max = 81),
  country = rep(c("Australia", "Germany", "Japan", "Singapore",
                  "South Korea", "Sweden", "U.S."), each = 21)
)

# Define custom colors for countries
country_colors <- c(
  "Australia"   = "navy",
  "Germany"     = "maroon",
  "Japan"       = "forestgreen",
  "Singapore"   = "orange",
  "South Korea" = "lightgreen",
  "Sweden"      = "red",
  "U.S."        = "purple"  # 'lavender' is hard to see; replaced with darker purple
)

# Define custom shapes for countries
country_shapes <- c(
```

Replicated to a Table

Female Labor Force Participation

Country	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009
Australia	60.3	74.9	63.9	77.6	79.3	53.3	67.3	77.9	68.0	65.2
Germany	72.1	70.6	80.8	71.0	72.5	67.8	69.2	60.4	56.3	79.9
Japan	64.0	62.7	56.4	56.0	58.8	65.5	59.7	76.9	53.3	64.8
Singapore	60.0	75.6	65.0	75.5	75.6	75.0	64.8	73.9	70.2	72.6
South Korea	55.0	64.6	80.6	77.9	77.7	57.1	55.8	70.9	62.0	71.0
Sweden	77.8	78.5	69.7	63.9	56.3	79.1	60.7	53.8	79.5	72.9
U.S.	56.5	54.6	56.1	72.0	70.0	77.9	71.5	73.4	67.1	71.1

Replicated Code

```r
library(dplyr)
library(tidyr)
library(knitr)
library(kableExtra)

# Sample data
set.seed(123)
data <- data.frame(
  year = rep(2000:2020, 7),
  flfp = runif(21 * 7, min = 52, max = 81),
  country = rep(c("Australia", "Germany", "Japan", "Singapore",
                  "South Korea", "Sweden", "U.S."), each = 21)
)

# Pivot to wide format
flfp_wide <- data %>%
  pivot_wider(names_from = year, values_from = flfp)

# Create enhanced table
flfp_wide %>%
  kable(
    format = "html",
    digits = 1,
    caption = "📊 Female Labor Force Participation Rate (%) by Country and Year",
    col.names = c("Country", as.character(2000:2020))
  ) %>%
  kable_styling(
```

References

·el-Bundock, Vincent. 2022. “**Modelsummary**: Data and Model Summaries in *R*.” *Journal of Statistical Software* 103 (1). **https://doi.org/10.18637/jss.v103.i01**.

ıoi, Syngjoo, Seonghoon Kim, and Wooyoung Lim. 2025. “Strategic Thinking Skills: A Key to Collective Economic Success.” *American Economic Journal: Microeconomics* 17 (2): 214–40. **https://doi.org/10.1257/mic.20220259**.

'ickham, Hadley. 2016. *Ggplot2: Elegant Graphics for Data Analysis*. Springer. **https://ggplot2.tidyverse.org**.

'ickham, Hadley, Romain François, Lionel Henry, Kirill Müller, and Davis Vaughan. 2014. “Dplyr: A Grammar of Data Manipulation.” The R Foundation. **https://doi.org/10.32614/cran.package.dplyr**.

ıu, Hao. 2017. “kableExtra: Construct Complex Table with ’Kable’ and Pipe Syntax.” The R Foundation. **https://doi.org/10.32614/cran.package.kableextra**.

Speaker notes