V506 R Introductory Lab Exercise – Fall 2024 - Solution

Luis Navarro

August 2024

In this document I lay out the solutions for the in-lab exercise. I strongly recommend for you to run *line-by-line* the code below and understand what is doing. After running each line see what appears in the console and any potential changes in your environment.

I will provide solutions using base R functions as well as more simple-elegant solutions using the packages described in the lab.

1. Set and confirm your working directory, preferably to your cloud storage folder that you will use for V506.

Using base R we can use the `getwd` and `setwd` commands. Hint: If your code shows error, check whether you wrote the backslashes correctly.

- getwd (get working directory): shows the current folder in which R is looking/dropping files.
- setwd (changes the working directory)

```
# get wd shows my current directory
getwd()
```

```
## [1] "/Users/luisenriquenavarro/Library/CloudStorage/OneDrive-IndianaUniversity/V506/Fall24/Lab1"
```

Use setwd to change it to the folder where I store all the V506 materials.

```
setwd('/Users/luisenriquenavarro/Library/CloudStorage/OneDrive-IndianaUniversity/V506/Fall24/Lab1')
```

Verify it changed (or verify is the folder you are using):

```
getwd()
```

```
## [1] "/Users/luisenriquenavarro/Library/CloudStorage/OneDrive-IndianaUniversity/V506/Fall24/Lab1"
```

You can also use the `here` library to set your master data folder and just set the folder path using the here command.

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

```
library(here)
```

```
## here() starts at /Users/luisenriquenavarro/Library/CloudStorage/OneDrive-IndianaUniversity/V506/Fall2
```

```r
# Create here object. R understands this is an address in your computer.
v506_path <- here('/Users/luisenriquenavarro/Library/CloudStorage/OneDrive-IndianaUniversity/V506')
```

The benefit of using here is that you can explore the folders easily with the syntax of the `here` function.

```r
v506_lab1 <- here(v506_path, 'Fall24','Lab1')
print(v506_lab1)
```

```
## [1] "/Users/luisenriquenavarro/Library/CloudStorage/OneDrive-IndianaUniversity/V506/Fall24/Lab1"
```

2. Install and load the packages, "rmarkdown", "dplyr" , and "tidyverse" to R.

Let's install also the packages we discussed in class.

```r
# install packages
install.packages("rmarkdown")
install.packages("tidyverse")
# install rio: to import/export data.
install.packages("rio")
# install pacman
install.packages("pacman")
```

You can load the libraries using the library function.

```r
library(rmarkdown)
library(tidyverse)
```

```
## -- Attaching core tidyverse packages ------------------------ tidyverse 2.0.0 --
## v dplyr     1.1.4     v readr     2.1.5
## v forcats   1.0.0     v stringr   1.5.1
## v ggplot2   3.5.1     v tibble    3.2.1
## v lubridate 1.9.3     v tidyr     1.3.1
## v purrr     1.0.2
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors
```

```r
library(here)
library(rio)
```

```
## Warning: package 'rio' was built under R version 4.4.1
```

Alternatively, you can install `pacman` and use p_load to load all libraries together

```r
# load pacman
library(pacman)

# use p_load to load all libraries
p_load(rmarkdown, tidyverse, here, rio)
```

Verify on the packages tab. The loaded packages should have a checkmark.

3.Import the Banking.csv file (download from the R Labs>Introductory Lab folder in Canvas and save to your working directory) to a data frame called atm. This file contains the variables indicated below. Confirm that the variables listed below are the same as what you imported using the names() or colnames() function.

We can load the data using base R read_csv function and creating the file path manually.

Let's name this data frame: `atm`

```
# Base R solution
banking_filepath <- '/Users/luisenriquenavarro/Library/CloudStorage/OneDrive-IndianaUniversity/V506/Fal

atm <- read.csv(file = banking_filepath)
```

We can also do it with `rio` and `here`. Notice I will overwrite the previous objects (i.e. use the same names, and assign new information to them.)

```
banking_filepath <- here(v506_lab1, 'banking.csv')

atm <- rio::import(file = banking_filepath)
```

4. Confirm the structure of the data frame and, using the head() and tail() functions, print the first 8 and last 10 observations in the Console Window.

The head and tail functions create new data frames with the top or bottom observations in the data frame.

```
atm_top8 <- head(atm, n = 8)
print(atm_top8)
```

```
##   Balance ATM Services Debit Interest City
## 1    1756  13        4     0        1    2
## 2     748   9        2     1        0    1
## 3    1501  10        1     0        0    1
## 4    1831  10        4     0        1    3
## 5    1622  14        6     0        1    4
## 6    1886  17        3     0        1    1
## 7     740   6        3     0        0    3
## 8    1593  10        8     1        0    1
```

```
atm_last10 <- tail(atm, n = 10)
print(atm_last10)
```

```
##    Balance ATM Services Debit Interest City
## 51    1338  14        4     1        0    2
## 52    2076  12        5     1        0    2
## 53    1708  13        3     1        0    1
## 54    2138  18        5     0        1    4
## 55    2375  12        4     0        0    2
## 56    1455   9        5     1        1    3
## 57    1487   8        4     1        0    4
## 58    1125   6        4     1        0    2
## 59    1989  12        3     0        1    2
## 60    2156  14        5     1        0    2
```

5. *Using the nrow() function, determine the number of rows in the data frame.*

Function `nrow` shows the number of rows in the data frame.

```
nrow(atm)
```

```
## [1] 60
```

6. *Generate and print basic descriptive statistics for the atm data frame.*

The `summary` shows the descriptive statistics of all the columns in a data frame.

```
summary(atm)
```

```
##     Balance           ATM           Services          Debit
##  Min.   :  32   Min.   : 2.0   Min.   :0.000   Min.   :0.0000
##  1st Qu.:1124   1st Qu.: 7.0   1st Qu.:3.000   1st Qu.:0.0000
##  Median :1604   Median :10.0   Median :4.000   Median :0.0000
##  Mean   :1500   Mean   :10.3   Mean   :4.417   Mean   :0.4333
##  3rd Qu.:1924   3rd Qu.:13.0   3rd Qu.:6.000   3rd Qu.:1.0000
##  Max.   :2557   Max.   :20.0   Max.   :9.000   Max.   :1.0000
##     Interest           City
##  Min.   :0.0000   Min.   :1.0
##  1st Qu.:0.0000   1st Qu.:1.0
##  Median :0.0000   Median :2.0
##  Mean   :0.2667   Mean   :2.4
##  3rd Qu.:1.0000   3rd Qu.:3.0
##  Max.   :1.0000   Max.   :4.0
```

7. *Create and print a new variable called balatm, which represents the balance per ATM transaction (i.e., balance divided by the number of ATM transactions per month). Make this into a data.frame with 1 column and 60 rows.*

We need to create a need variable. We can use the dollar sign syntax to do it.

```
# first create vectors of the variables, so you understand how the operation is taking place
atm_balance <- atm$Balance
atm_transactions <- atm$ATM

# compute the balance per transaction
atm_balance_transcation <- atm_balance/atm_transactions

# we can store this as an independent data frame. In this case, I named this new variable balance_atm i
atm_balance_transcation_df <- data.frame(balance_atm = atm_balance_transcation)
```

I can also add this variable to my existing data frame.

```
atm$balance_atm <- atm$Balance/atm$ATM
```

8. *Write the balatm vector to your working directory as "balatm.R", and confirm its presence on your storage media.*

We can export using both base R and libraries here and rio.

Task: export dataframe `atm_balance_transcation_df` to your folder.

Base R solution:

```
# check the help file to verify the syntax of the function
# you need to specify object x (the data frame you want to export) and the file: where is going to be s
# In my case, I created a folder called clean data to store the output from this exercise.
export_path <- '/Users/luisenriquenavarro/Library/CloudStorage/OneDrive-IndianaUniversity/V506/Fall24/La
write.csv(atm_balance_transcation_df,
          file = export_path)
```

Using `rio` and `here`

```
export_path <- here(v506_lab1, 'clean_data','atm_balance_transcation_df.csv')
rio::export(atm_balance_transcation_df,
            file = export_path)
```

Extra: try exporting this to an excel and an Rds file.

Without rio: you will need install the library writexl and use the `write_xlsx` function, which has similar syntax to `write.csv`. For the Rds file, you can use `saveRDS`.

With rio: just change the file extension at the end of the file.

```
# Excel
rio::export(atm_balance_transcation_df,
            file = here(v506_lab1, 'clean_data','atm_balance_transcation_df.xlsx'))
# Rds
rio::export(atm_balance_transcation_df,
            file = here(v506_lab1, 'clean_data','atm_balance_transcation_df.Rds'))
```

9. If you have time, transfer your code to an R Markdown program and generate an HTML Report that includes Steps 1-9 above, along with some comments that document various parts of the program.

If all your code is correct, then you should be able to knit this document into an HTML Report.

Extra: try knitting into a pdf file. For that, just change the YAML preamble at the top of this document.

```
---
title: "V506 R Introductory Lab Exercise - Fall 2024 - Solution"
author: Luis Navarro
date: August 2024
output: pdf_document
---
```