LAB 3 REPORT

**Student’s Name: Đinh Vũ Ngọc Linh**

**Student’s ID: ITDSIU21095**

1. **Exercise 1:**

First, we deal with Metro data.

* We remove all prefix X from all date columns.
* Then we make data tidy by pivot\_longer
* Make all date columns to be a column namely Date, and the columns Value with data of all date columns, so that the Metro data is tidy now

Now we deal with Zip data.

* We remove all prefix X from all date columns.
* Then we see that StateName column and State column have the same value so we can remove 1 column to make data tidy
* Make all date columns to be a column namely Date, and the columns Value with data of all date columns, so that the Zip data is tidy now!

Ảnh có chứa văn bản, ảnh chụp màn hình, Phông chữ

Mô tả được tạo tự động

1. **Exercise 2:**

Use intersect() to find and check if there are the common RegionID or StateName in the two datasets. Ảnh có chứa văn bản, ảnh chụp màn hình, Phông chữ, hàng

Mô tả được tạo tự động

The result of common between Zipdata’s RegionID, and Metrodata’s RegionID is

integer(0)

means the is no common between them.

The result of common between Zipdata’s StateName, and Metrodata’s StateName is

[1] "TX" "NJ" "NY" "CA" "IL" "TN" "GA" "WA" "VA" "NC" "OK" "HI" "AZ" "KS" "FL" "NM" "NV" "PA" "CO"

[20] "MO" "ID" "UT" "MD" "IN" "OH" "OR" "MA" "CT" "MI" "KY" "SC" "DE" "LA" "WI" "MN" "MT" "MS" "AL"

[39] "AR" "SD" "RI" "NE" "WV" "ME" "IA" "ND" "AK" "NH" "VT" "WY"

1. **Exercise 3:**

We merge the two datasets together, which matches pairs of observations whenever their keys or IDs are equal. We match on the variables RegionID and RegionID. We use merge().

We save the merged data set into a new object called old\_data\_metro\_zip



1. **Exercise 4:**

Now we merge the two datasets together, match all the variables that have the same name between two datasets. We use merge().

We save the new merge data set into new object called new\_data\_metro\_zip



1. **Exercise 5:**

We find the number of NA’s in new\_data\_metro\_zip

Ảnh có chứa văn bản, Phông chữ, ảnh chụp màn hình, Đồ họa

Mô tả được tạo tự động

The result is:

[1] "Number of NA's in data\_metro\_zip: 1592146"

1. **Exercise 6:**

We replace all NA in dataset with 0

