

Class3_Exercise

PA 434

1/25/2020

Exercise

We are going to open up last week's dataset and use dplyr to extract some information

Upload the ImmigrationData and RegionData datasets from Blackboard.

Remember to report at the top of your code any packages that you need to complete the exercise.

Merge

Let's do one more "join" of two datasets to make sure that it's clear in your head.

Merge the two dataset so that **all observations contained in ImmigrationData are kept**. How many rows to you expect the dataset to have? Remember to add the code lines needed to make sure the merge was done correctly (e.g., check dataset size)

Analysis

Perform the following actions and answer the following questions using dplyr. Use **pipes** when writing your code.

Filter, arrange, and select

1. Filter all observations from countries in Africa collected in 1990. How many are there? (you can just look at the number of rows)
2. Which countries have a number of female migrants between 1 and 2 million excluded?
3. Filter countries located in Africa or Oceania. How many are there? (you can just look at the number of rows)
4. Which country has the highest number of migrants among "developed" countries in 2010? In asnwering these questions, keep only relevant columns in the output
5. Rerganize the dataset so that all columns containing information about the immigrant population come right after the year column. All columns should be kept in the dataset.
6. Filter observations from the following countries: Italy, Germany, Spain, France, Portugal, and Greece in 2015. Use **%in%** to write your filter function.