

Week 4 Practice

Elena Leib & Willa Voorhies

This week, you will be working with data from the 2020 World Happiness Report. They use a measure called “ladder score” to measure happiness in countries. They also collect a bunch of other measures for each country.

- 1) Load in world-happiness_2020.csv, which has the information about each country. Look at the data frame, in particular the column names (use colnames())
- 2) Load in population.csv, which holds general information about each country (e.g., population). Look at the data frame.

Do the next steps in one pipe:

- 3) After looking at the world happiness data frame, we see a bunch of columns that we won't need, and they all start with “Explained.” Remove all these columns.

One of these helper functions might make the task easier!

	Helper functions for select - ?select
<code>select(iris, contains("."))</code>	Select columns whose name contains a character string.
<code>select(iris, ends_with("Length"))</code>	Select columns whose name ends with a character string.
<code>select(iris, everything())</code>	Select every column.
<code>select(iris, matches(".t."))</code>	Select columns whose name matches a regular expression.

- 4) Add the population information for all the countries in the happiness dataframe.
- 5) Add a column that categorizes countries by size. Large countries have over 100 million citizens, small countries have fewer than 1 million citizens, and medium countries have populations in between.
- 6) Create a new variable that categorizes countries as below average or above average for ladder score.
- 7) Now create another new variable that categorizes countries as below average or above average compared to the other countries in their region.

<End pipe>

Summarizing data

- 8) Pick two variables and summarize them in a new data frame. Get the mean, median, and sd.
- 9) In another new data frame, get the mean, median, sd for these variables by region.
- 10) In a third new data frame, get the mean, median, sd for these variables by region and population category.