Session 2: Practice Problems

Andrew McCormack

2018-11-13

Problem 1

1. Load the civil war dataset.
2. Select the `pop`, `mtn`, and `polity2` variables. Create a data frame with just those three variables along with the variables `country` and `year`. Do this using the `select()` function from `dplyr`.

Problem 2

1. There is an issue with the mountainous terrain variable, `mtn`. Instead of `NA` values, missing data was coded as `-99`. This won't do. Ask R for a summary of the mountain variable to verify this.
2. Create a new variable called `mtn2` that is the same as `mtn` except with the `-99` values replaced with `NA`s.
3. Bonus: there are two ways to solve problem 2—one using `dplyr` and the other using the basic functions from last week. Try using `dplyr` to solve this problem (hint: `mutate()` and `ifelse()` might come in handy here).

Problem 3

1. Using the original data, create a "dummy" variable for observations where the value is 1 if the country is located in Sub-Saharan Africa and 0 otherwise.
2. Select a numeric variable and calculate its average separately for countries in Sub-Saharan Africa and countries elsewhere. Try doing this using the pipe operator in `dplyr`. What is the difference in the two means?

Problem 4

1. Filter the original dataset to include only the years 1989 and 1999.
2. Find the means for `polity2` by region and year (hint: you can use two variables in `group_by()`). Did any regions experience democratic backsliding between 1989 and 1999? If so, which ones?

*Bonus: complete steps 1 and 2 simultaneously using the pipe operator.