

Session 2: Practice Problems

Andrew McCormack

2018-11-13

```
setwd("/Users/andrewmccormack/Documents/DSC/")
civilwar <- read.csv("fearon03.csv")
```

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`.

```
cw <- civilwar %>%
  select(country, year, pop, mtn, polity2)
```

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.

```
summary(cw$mtn)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## -99.00    1.70    9.70   16.94   27.20   94.30
```

2. Create a new variable called `mtn2` that is the same as `mtn` except with the `-99` values replaced with `NA`s. Don't alter the original mountain variable.

```
cw$mtn2 <- cw$mtn
cw$mtn2[cw$mtn == -99] <- NA
```

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).

```
cw <- cw %>%
  mutate(mtn2 = ifelse(mtn == -99, NA, mtn))
```

```
## Warning: package 'bindrcpp' was built under R version 3.4.4
```

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.

```
unique(civilwar$region)
```

```
## [1] western democracies and japan
## [2] latin america and the caribbean
## [3] e. europe and the former soviet union
## [4] n. africa and the middle east
```

```
## [5] sub-saharan africa
## [6] asia
## 6 Levels: asia ... western democracies and japan
```

```
cw2 <- civilwar %>%
  mutate(sub_saharan_africa = ifelse(region == "sub-saharan africa", 1,0))
```

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?

```
# Base method
cw2$gdppc[cw2$sub_saharan_africa == 1] %>%
  mean(na.rm = TRUE)
```

```
## [1] 1.103149
```

```
cw2$gdppc[cw2$sub_saharan_africa == 0] %>%
  mean(na.rm = TRUE)
```

```
## [1] 4.530286
```

```
# dplyr method
cw2 %>%
  group_by(sub_saharan_africa) %>%
  summarise(gdppc_mean = mean(gdppc, na.rm = TRUE))
```

```
## # A tibble: 2 x 2
##   sub_saharan_africa gdppc_mean
##           <dbl>         <dbl>
## 1             0         4.53
## 2             1         1.10
```

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?

```
civilwar %>%
  filter(year %in% c(1989, 1999)) %>%
  group_by(region, year) %>%
  summarise(democracy_mean = mean(polity2, na.rm = T))
```

```
## # A tibble: 12 x 3
## # Groups:   region [?]
##   region                year democracy_mean
##   <fct>              <int>         <dbl>
## 1 asia                1989         -1.17
## 2 asia                1999          1.65
## 3 e. europe and the former soviet union 1989          -3
## 4 e. europe and the former soviet union 1999          3.15
## 5 latin america and the caribbean      1989          4.96
## 6 latin america and the caribbean      1999          6.83
## 7 n. africa and the middle east          1989         -4.35
## 8 n. africa and the middle east          1999         -3.33
## 9 sub-saharan africa                    1989         -5.76
```

## 10 sub-saharan africa	1999	0.190
## 11 western democracies and japan	1989	9.95
## 12 western democracies and japan	1999	9.95

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