# CSSS 554: Assignment 2

## Department of Biostatistics @ University of Washington

### Alejandro Hernandez

### Winter Quarter 2025

1. After creating a DHS account, download and process the required data.

This assignment models the proportion of women with an unmet need for family planning across Admin-1-level regions of Armenia. Armenia is a landlocked country, located in West Asia. It is bordered on the north and east by Georgia and Azerbaijan. On the south and west, it shares borders with Iran and Turkey.

(a) Show the first six lines of the data object created using the getDHSindicator function.

```
##   cluster householdID             v022    v023    v024  weight strata value
## 1       1           1 tavush - rural yerevan yerevan 2213166  urban     1
## 2       1           1 tavush - rural yerevan yerevan 2213166  urban     1
## 3       1           6 tavush - rural yerevan yerevan 2213166  urban     0
## 4       1           6 tavush - rural yerevan yerevan 2213166  urban     0
## 5       1           7 tavush - rural yerevan yerevan 2213166  urban     1
## 6       1          11 tavush - rural yerevan yerevan 2213166  urban     0
```

(b) Show the first six lines of the data object created using the adminInfo() function for admin=1.

```
##   admin1.name population surveyWeight urban
## 1  Aragatsotn         NA           NA    NA
## 2      Ararat         NA           NA    NA
## 3     Armavir         NA           NA    NA
## 4      Erevan         NA           NA    NA
## 5 Gegharkunik         NA           NA    NA
## 6      Kotayk         NA           NA    NA
```
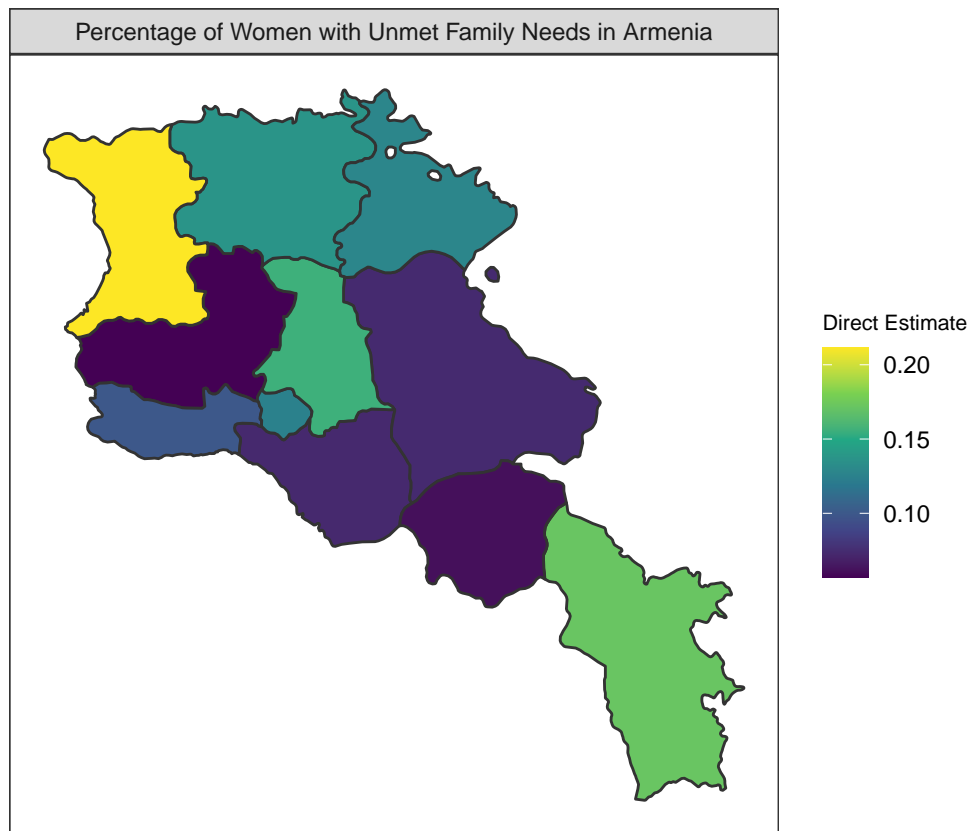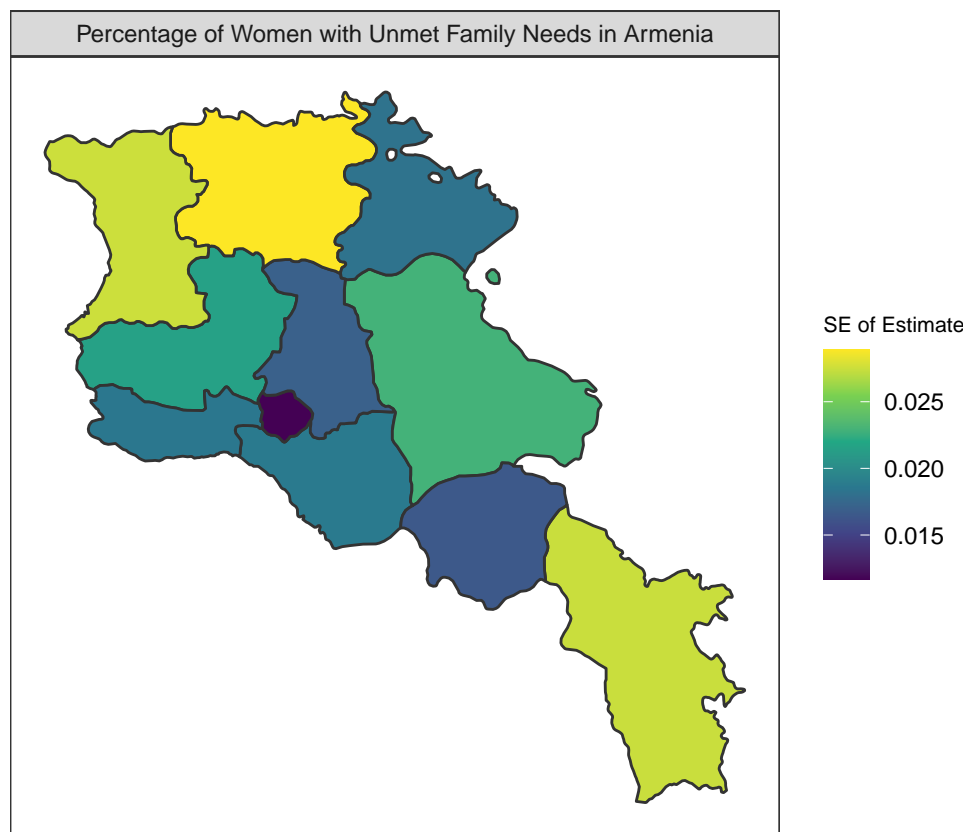
2. Obtain prevalence estimates at admin1, using the surveyPrev::directEST() function. Provide a table with admin names, prevalence estimates, variance of prevalence estimates. [Hint: These are the Hajek estimates.]

```
## Reading layer 'AMGE71FL' from data source
##   'C:\Users\alega\Desktop\RProjects\spatial-analysis\data\Armenia2016\AMGE71FL - GPS\AMGE71FL.shp'
##   using driver 'ESRI Shapefile'
## Simple feature collection with 313 features and 20 fields
## Geometry type: POINT
## Dimension:     XY
## Bounding box:  xmin: 44 ymin: 39 xmax: 46 ymax: 41
## Geodetic CRS:  WGS 84
```

```
##      Admin 1 Name Direct Estimate Variance of Estimate
## 1     Aragatsotn            0.057              0.00046
## 2         Ararat            0.073              0.00035
## 3        Armavir            0.100              0.00034
## 4         Erevan            0.125              0.00014
## 5     Gegharkunik           0.074              0.00052
## 6         Kotayk            0.156              0.00029
## 7           Lori            0.137              0.00084
## 8         Shirak            0.212              0.00076
## 9         Syunik            0.172              0.00075
## 10        Tavush            0.128              0.00033
## 11     Vayots Dzor          0.062              0.00027
```

3. Using the mapPlot() function, provide two maps, one of the prevalence estimates and the other of the standard errors (which are the square roots of the variances you put in a table in the previous part). And to reiterate, these are at admin1.



Percentage of Women with Unmet Family Needs in Armenia

Percentage of Women with Unmet Family Needs in Armenia

End of report. Code appendix begins on the next page.

## Code Appendix

```r
# Clear environment
rm(list=ls())

# Setup options
knitr::opts_chunk$set(echo=FALSE, warning=FALSE, message=FALSE)
options(knitr.kable.NA = '-', digits = 2)
labs = knitr::all_labels()
labs = labs[!labs %in% c("setup", "allcode")]

# Load relevant packages
library(devtools)
devtools::install_github("richardli/surveyPrev")
library(surveyPrev)
library(dplyr)
library(kableExtra)
library(labelled)
library(naniar)
library(sjlabelled)
library(haven)
library(dplyr)
library(geodata)
library(sf)
library(INLA)
library(SUMMER)
library(knitr)


#### QUESTION 1 ####

# Data settings
indicator <- "unmet_family"
year <- 2010
country <- "Armenia"

## Load data locally
file_path <- "../data/Armenia2016/AMBR72DT/AMBR72FL.DTA" # data file
dhsData <- as.data.frame(read_dta(file_path))
indicator <- "unmet_family"
data <- surveyPrev::getDHSindicator(dhsData, indicator = indicator)
# View head
head(data)

# Get admin 1 info
poly.adm1 <- geodata::gadm(country = country, level = 1, path = tempdir())
poly.adm1 <- sf::st_as_sf(poly.adm1)
admin.info <- surveyPrev::adminInfo(poly.adm = poly.adm1, admin = 1, by.adm = "NAME_1")
head(admin.info$data)

#### QUESTION 2 ####

# Get geographic data locally
file_path_geo <- "../data/Armenia2016/AMGE71FL - GPS/AMGE71FL.shp" # shapefile
```

```r
geo <- st_read(file_path_geo)

## Get prevalence estimates at admin 1 level
# Get cluster information at Admin 1
source("../R/cluster.info.revised.R")
cluster.info <- cluster.info.revised(geo = geo, poly.adm1 = poly.adm1,
                                     by.adm1 = "NAME_1")

# Estimate prevalence
res_ad1 <- directEST(data = data, cluster.info = cluster.info, admin = 1)
plot_ad1 <- res_ad1$res.admin1 %>%
  select("Admin 1 Name"=admin1.name, "Direct Estimate"=direct.est,
         "Variance of Estimate"=direct.var)
plot_ad1

#### QUESTION 3 ####

# Arrange all estimates for mapPlot
plot_ad1$"SE of Estimate" <- sqrt(plot_ad1$"Variance of Estimate")
plot_ad1$model <- "Percentage of Women with Unmet Family Needs in Armenia"

# Map direct estimates and uncertainty
g1 <- SUMMER::mapPlot(data = plot_ad1, geo = poly.adm1,
             by.data = "Admin 1 Name", by.geo = "NAME_1", is.long = TRUE,
             variable = "model",
             value = "Direct Estimate",
             legend.label = "Direct Estimate")
g2 <- mapPlot(data = plot_ad1, geo = poly.adm1,
             by.data = "Admin 1 Name",  by.geo = "NAME_1", is.long = TRUE,
             variable = "model", value = "SE of Estimate",
             legend.label = "SE of Estimate")
g1
g2
```

**End of document.**