

# PracTools Project

Sunghee Lee, Raphael Nishimura

Assigned 2/18/2025; Due 4/1/2025

## Introduction

In this project, you will design a sample of census tracts, block groups (BGs), and persons from Prince George's County in the state of Maryland in the U.S. for a survey whose goals are to estimate the proportions of persons in different age groups with different levels of civic awareness. For example, persons will be asked the name of their district representative in the U.S. House of Representatives, the name of their local delegate to the Maryland House of Delegates, whether the constitution of Maryland contains a provision protecting the right to keep and bear arms, and other questions related to knowledge of Maryland laws.

Based on considering analytic subgroups, the desired precision of estimates, and the available budget, it has been determined that these sample sizes of respondents are to be selected:

Age Group	n (respondents)	Expected Response Rate
18-44 yrs	200	0.50
45-64 yrs	200	0.60
65+ yrs	200	0.65
Total	600	
Sample tracts	30	
Sample blockgroups/tract	1	

The sample design will use census tracts as PSUs, block groups (BGs) as SSUs, and persons as elements. The goals of the sample design are to select a sample that will yield the responding sizes above while 1) achieving a self-weighting sample in each of the age groups above and 2) obtaining an equal workload in each sample PSU.

You should pay particular attention to tracts where there are not 2 BGs and to geographic areas that have small population counts and decide how they should be handled in the frame.

Use systematic sampling with probabilities proportional to size to select the PSUs and SSUs. Sort the file in order by tract before selecting the sample. This method of selection is available in the R **sampling** package. When using R to select the sample, include the statement

```
set.seed(-1234)
```

at the beginning of your code.

# Deliverables

The deliverables for the project will be

1. A sampling report (details of the report below) [filename: REPORT\_GROUPNAME.pdf];
2. Text files giving the units used for the area frame and relevant census counts and measures of size [filename: FRAME#\_GROUPNAME.pdf]; and
3. Text file for the selected sample along with relevant census counts, measures of size, selection probabilities, and weights [filename: SAMPLE\_GROUPNAME.pdf]

## Sampling Report

### Contents

This will be a technical report written to a client whose staff includes managers and technical personnel. Managers will be more interested in understanding the broad outline of the steps used in weighting. Technical personnel will be interested in understanding the details of sample selection and weight computation, including appropriate formulae. You should consider how to structure your report to serve these audiences. Below is a list of topic areas that should be included in your report. The order of the sections in your report does not have to be the same as that given below. You should construct your report in a way that presents topics in an order that seems logical.

### Topic Areas for the Sampling Report

1. Title Page (project title, date of submission, and name of project contact person)
2. Introduction (overview of the document)
3. Sample Design
  - Goals of the sample design
  - Target population
  - Area sampling frame: units, data available, source of the data
  - Assigning measures of size to units
  - Descriptive statistics on the frame you created
  - Discuss whether a design that uses the same sampling rate for all persons could be used to achieve the sample size targets for each age group. Describe any advantages or disadvantages that this design might have. Support your conclusions with numerical calculations.
4. Sample Selection
  - Method of selection
  - Selected units and characteristics of each
  - Selection probabilities of units at each stage of selection
  - Description of how persons should be selected from area listings
5. Anticipated Precision
  - Precision of estimated proportions of persons who correctly answer questions related to the government of Maryland, e.g., are able to name their representative in the House of Delegates
6. General description of weighting plan

7. General explanation of how an analyst should estimate variances from the sample
  - Discuss how to define strata, PSUs, and any other characteristics that should be considered
8. Maps
  - Prince George's County
  - Selected tracts and block groups
9. Appendix
  - Codebook of frame and sample files, i.e. a list of the variables on the files and a description of each variable
  - Listing of the sample PSUs and sample SSUs with their selection probabilities and census data. On each sample SSU, list the sampling rate you will use to select persons in each domain.

## **Format**

- 10 to 15 pages long
- Line spacing: Single
- 1 inch margins
- Pages should be numbered
- Use informative section titles

## Data files and other information

### Background information about geographies in Census 2020

- Census geography hierarchy:  
<https://www.census.gov/programs-surveys/geography/guidance/hierarchy.html>
- Census geography definition:  
<https://www.census.gov/programs-surveys/geography/about/glossary.html>
- Collection of census geography reference maps:  
<https://www.census.gov/programs-surveys/geography/geographies/reference-maps.html>

### Prince George's County, MD population data

- PrinceGeorgeMD.xlsx available on Canvas includes both tract- and blockgroup-level aggregate data on 1) housing counts, 2) person counts and 3) person counts by Sex x Age from the 2020 Decennial Census Decennial Census: Demographic and Housing Characteristics (DHC) Summary File. (Note that there are many types of summary files from 2020 Decennial Census; See <https://api.census.gov/data.html> for detail)
- PrinceGeorgeMD.xlsx was prepared by using `tidycensus` package to download the data for Prince George's County, MD as follows:
  - Note: In order to download data using `tidycensus`, a census Application Programming Interface (API) key is required. You can obtain this from [https://api.census.gov/data/key\\_signup.html](https://api.census.gov/data/key_signup.html).

```
library(tidycensus)
library(tidyverse)

# Once you have the api key, you can set with:
census_api_key("YOUR KEY", install = TRUE)
# No need to set repeatedly.

# You can look up what variables are available
# in 2020 Decennial DHC
var2020dhc <- load_variables(2020, "dhc", cache = TRUE)
View(var2020dhc)
# From this, you can determine what variable you want.
# In our case, 1) housing counts (H1_001N),
#                2) person counts (many options) and
#                3) person counts by Sex x Age ("P_12" series).

# For tract level housing counts:
hh_tract <- get_decennial(geography = "tract",
                          variable = c(TotHH = "H1_001N"),
                          year = 2020,
```

```

        sumfile="dhc",
        state="MD", county="Prince",
        output="wide")
# For tract level person counts -- Sex x Age:
person_tract<- get_decennial(geography = "tract",
                             table="P12",
                             year = 2020,
                             sumfile="dhc",
                             state="MD", county="Prince",
                             output = "wide")%>%
rename(TotPerson      =P12_001N,
       TotMale        =P12_002N,
       MaleLT5Yrs     =P12_003N,
       Male5to9Yrs    =P12_004N,
       Male10to14Yrs  =P12_005N,
       Male15to17Yrs  =P12_006N,
       Male18to19Yrs  =P12_007N,
       Male20Yrs      =P12_008N,
       Male21Yrs      =P12_009N,
       Male22to24Yrs  =P12_010N,
       Male25to29Yrs  =P12_011N,
       Male30to34Yrs  =P12_012N,
       Male35to39Yrs  =P12_013N,
       Male40to44Yrs  =P12_014N,
       Male45to49Yrs  =P12_015N,
       Male50to54Yrs  =P12_016N,
       Male55to59Yrs  =P12_017N,
       Male60to61Yrs  =P12_018N,
       Male62to64Yrs  =P12_019N,
       Male65to66Yrs  =P12_020N,
       Male67to69Yrs  =P12_021N,
       Male70to74Yrs  =P12_022N,
       Male75to79Yrs  =P12_023N,
       Male80to84Yrs  =P12_024N,
       MaleGE85Yrs    =P12_025N,
       TotFemale      =P12_026N,
       FemaleLT5Yrs   =P12_027N,
       Female5to9Yrs  =P12_028N,
       Female10to14Yrs=P12_029N,
       Female15to17Yrs=P12_030N,
       Female18to19Yrs=P12_031N,
       Female20Yrs    =P12_032N,
       Female21Yrs    =P12_033N,
       Female22to24Yrs=P12_034N,

```

```

    Female25to29Yrs=P12_035N,
    Female30to34Yrs=P12_036N,
    Female35to39Yrs=P12_037N,
    Female40to44Yrs=P12_038N,
    Female45to49Yrs=P12_039N,
    Female50to54Yrs=P12_040N,
    Female55to59Yrs=P12_041N,
    Female60to61Yrs=P12_042N,
    Female62to64Yrs=P12_043N,
    Female65to66Yrs=P12_044N,
    Female67to69Yrs=P12_045N,
    Female70to74Yrs=P12_046N,
    Female75to79Yrs=P12_047N,
    Female80to84Yrs=P12_048N,
    FemaleGE85Yrs =P12_049N)

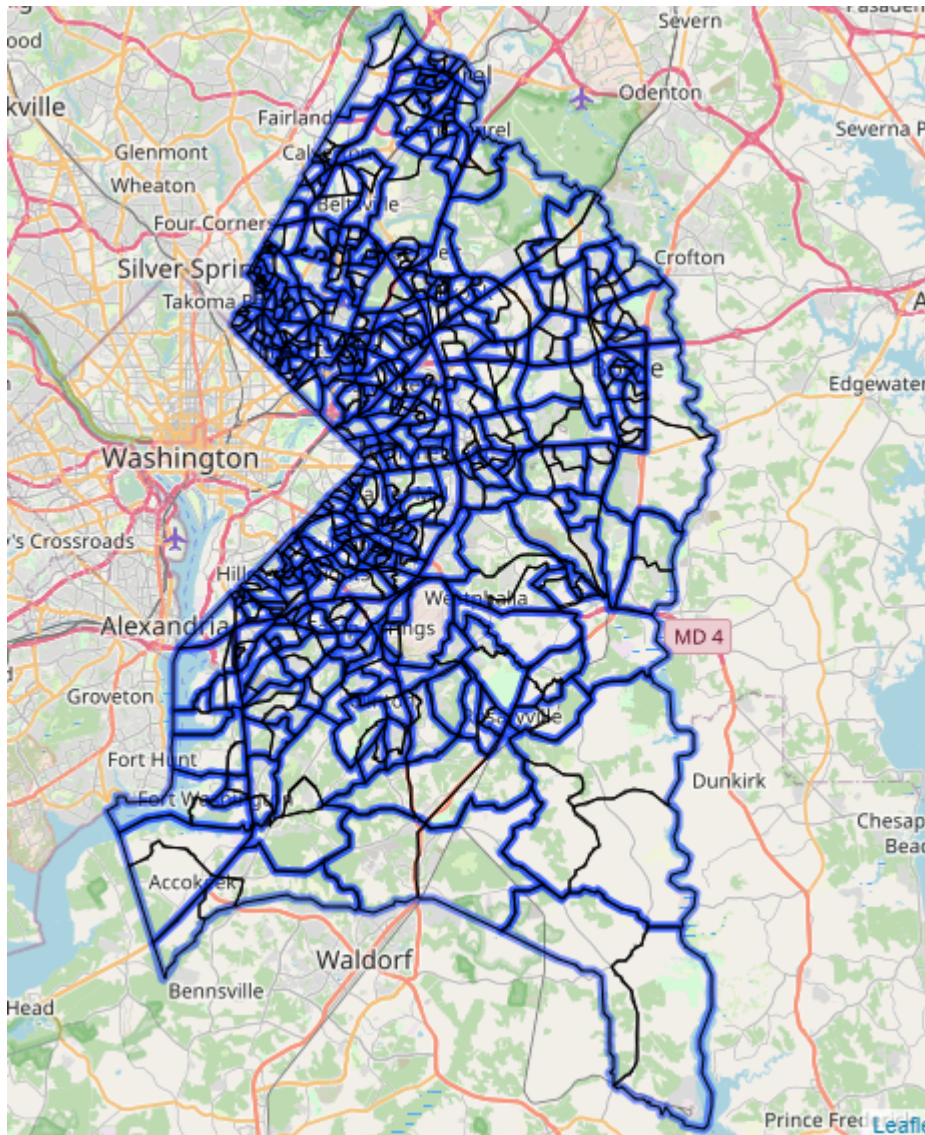
dim(hh_tract)
dim(person_tract)

tract<-left_join(hh_tract, person_tract, by=c("GEOID", "NAME"))%>%
  rename(Tract=GEOID)

# Similar code was used to prepare the blockgroup-level data.

```

## Maps of the Prince George's County, MD



- Blue: Census tracts; Black: Census blockgroups
- The map is interactive, generated using the code below. It requires various packages. The main ones here are `tigris` and `leaflet`.

```
library(glue)
library(tigris)
library(sf)
options(tigris_class = "sf")
options(tigris_use_cache = T)
library(leaflet)

# To get PG tract map shapefile from tigris
PG_tract <- tracts(state="MD", county="Prince", year=2020)
```



```

PG_tract[1:2,]

st_crs(PG_tract) #Check the GIS projection type

PG_tract_trans <- st_transform(PG_tract,
                               "+proj=longlat +datum=WGS84")
st_crs(PG_tract_trans)

# To get block group map shapefile
PG_bg <- block_groups(state="MD", county="Prince", year=2020)

PG_bg_trans <- st_transform(PG_bg, "+proj=longlat +datum=WGS84")

# To create PG blockgroup map layered with tracts
leaflet() %>%
  addTiles() %>%
  addPolygons(data=PG_tract_trans,
              popup = ~NAME,
              fillOpacity = 0,
              fillColor= "snow") %>%
  addPolygons(data=PG_bg_trans,
              popup = ~GEOID,
              color = "black",
              weight = 1,
              opacity= 1,
              smoothFactor = 1,
              fillOpacity = 0,
              fillColor= "snow")

```

- Simpler version

```
require(ggplot2)
ggplot() +
  geom_sf(data = PG_tract_trans, fill = NA,
          color = "black", size=1)+
  geom_sf(data = PG_bg_trans, fill=NA,
          linetype="dotted") +
  ggtitle("Prince George's County Tracts (Solid)
          \nand Block Groups (Dotted)")
```

