

Zoning out: an analysis of zoning and property values in Washington, D.C.

Brian Bontempo, Derrick Lee, and Aaron R. Williams

Sampling Frame

Download the data

```
library(tidyverse)
library(knitr)

# file path to csv with addresses
aru_file_path <-
  "https://opendata.arcgis.com/datasets/c3c0ae91dca54c5d9ce56962fa0dd645_68.csv"

ap_file_path <-
  "https://opendata.arcgis.com/datasets/aa514416aaf74fdc94748f1e56e7cc8a_0.csv"

# create a directory for downloading the data
if (!dir.exists("data/")) {
  dir.create("data")
}

# if the data doesn't already exist, download the data
if (!file.exists("data/aru.csv")) {
  download.file(aru_file_path, "data/aru.csv")
}

if (!file.exists("data/ap.csv")) {
  download.file(ap_file_path, "data/ap.csv")
}
```

Address Residential Units

The first dataset is Address Residential Units

The dataset does not contain a variable for quadrant, so we extract quadrant from the full address.

```
aru <- read_csv("data/aru.csv") %>%
  rename_all(tolower) %>%
  select(unit_id, address_id, fulladdress, status, unitnum, unittype)

# extract quadrant
aru <- aru %>%
  mutate(quadrant = str_sub(fulladdress, start = -2, end = -1))
```

Address Residential Units contains residential units with status set to “RETIRED”. We drop these cases as well.

```
count(aru, status) %>%  
  kable()
```

| status | n |
|----------|--------|
| ACTIVE | 244046 |
| ASSIGNED | 47 |
| RETIRE | 7087 |

```
aru <- aru %>%  
  filter(status != "RETIRE")
```

Address Points

```
# load the data and convert the variable names to lower case  
ap <- read_csv("data/ap.csv", guess_max = 10000) %>%  
  rename_all(tolower) %>%  
  select(address_id, status, type_, entrancetype, quadrant, fulladdress,  
         objectid_1, assessment_nbhd, cfsa_name, census_tract, vote_prcnct,  
         ward, zipcode, anc, census_block, census_blockgroup, latitude,  
         longitude, active_res_unit_count, res_type, active_res_occupancy_count)
```

Address Points contains residential units, non-residential units, and mixed-use units. Residential units and mixed-use units contain residences that belong to our sampling frame. We drop non-residential units.

```
count(ap, res_type) %>%  
  kable()
```

| res_type | n |
|-----------------|--------|
| MIXED USE | 473 |
| NON RESIDENTIAL | 15807 |
| RESIDENTIAL | 131370 |

```
ap <- ap %>%  
  filter(res_type != "NON RESIDENTIAL")
```

Address points contains residential units with status set to “RETIRED”. We drop these cases as well.

```
count(ap, status) %>%
  kable()
```

| status | n |
|-----------|--------|
| ACTIVE | 128490 |
| ASSIGNED | 668 |
| RETIRE | 2675 |
| TEMPORARY | 10 |

```
ap <- ap %>%
  filter(status != "RETIRE")
```

After the above filtering, there are 98 observations from Address Points and 3,706 observations in Address Residential Units that have missing addresses. We investigated joining the two datasets on `address_id` to fill in the address but all records missing an address in one dataset were missing an address in the other dataset.

We dropped the missing values which represented about 1.5 percent of observations in Address Residential Units and 0.07 percent of observations in Address Points.

```
ap <- ap %>%
  filter(!is.na(fulladdress))

aru <- aru %>%
  filter(!is.na(fulladdress))
```

Merge variables

Address Points has interesting variables not present in Address Residential Units. So we merge the Address Points dataset with the Address Residential Units dataset. The join works for all but 572 cases, most of which are in a new building at the Wharf.

```
aru_expanded <- aru %>%
  select(-status) %>%
  left_join(ap, by = c("fulladdress", "address_id")) %>%
  select(quadrant = quadrant.x, everything(), -quadrant.y)

anti_join(aru, ap, by = c("fulladdress", "address_id"))
```

```
## # A tibble: 572 x 7
##   unit_id address_id fulladdress      status unitnum unittype quadrant
##   <dbl>    <dbl> <chr>          <chr> <chr>    <chr>    <chr>
## 1  223379    276680 600 WATER STREET SW ACTIVE 6-12    RENTAL    SW
## 2  223380    276680 600 WATER STREET SW ACTIVE 6-13    RENTAL    SW
## 3  223381    276680 600 WATER STREET SW ACTIVE 6-14    RENTAL    SW
## 4  223384    276680 600 WATER STREET SW ACTIVE 1-1     RENTAL    SW
```

```
## 5 223389      276680 600 WATER STREET SW ACTIVE 1-6      RENTAL  SW
## 6 223392      276680 600 WATER STREET SW ACTIVE 1-9      RENTAL  SW
## 7 223494      276680 600 WATER STREET SW ACTIVE 8-16     RENTAL  SW
## 8 223497      276680 600 WATER STREET SW ACTIVE 9-3      RENTAL  SW
## 9 223503      276680 600 WATER STREET SW ACTIVE 9-9      RENTAL  SW
## 10 223508     276680 600 WATER STREET SW ACTIVE 9-14     RENTAL  SW
## # ... with 562 more rows
```

```
rm(aru)
```

Combination

Next, we need to drop addresses in the Address Points dataset that exist in the Address Residential Units dataset so we don't over count addresses in multi-dwelling units.

```
ap <- ap %>%
  filter(!address_id %in% unique(aru_expanded$address_id))
```

Finally, we can combine the two datasets to create a sampling frame that contains approximately every residential address in Washington D.C.

```
sampling_frame <- bind_rows(ap, aru_expanded)

rm(ap, aru_expanded)

#summarize_all(addresses, list(~sum(is.na(.))))

write_csv(sampling_frame, "sampling_frame.csv")
```

Pilot survey

```
set.seed(20190714)

pilot_sample <- sampling_frame %>%
  group_by(quadrant) %>%
  sample_n(25)

write_csv(pilot_sample, "data/pilot_sample.csv")

rm(pilot_sample)

# load the completed pilot survey and clean the values
pilot_sample <- read_csv("data/pilot_sample_completed.csv") %>%
  mutate(land_value = ifelse(!is.na(rf_land_value),
                             rf_land_value,
                             land_value),
```

```

    improvement_value = ifelse(!is.na(rf_improvement_value),
                                rf_improvement_value,
                                improvement_value)) %>%
mutate(property_value = land_value + improvement_value) %>%
mutate(property_value = ifelse(unittype == "RENTAL" &
                                active_res_occupancy_count > 4 &
                                property_value > 2000000,
                                property_value / active_res_occupancy_count,
                                property_value
                                ))

```

```

pilot_sample %>%
  summarize(mean = mean(property_value, na.rm = TRUE),
            s_squared_h = var(property_value, na.rm = TRUE),
            missing_prop = mean(is.na(property_value))) %>%
  kable(caption = "Pilot survey summary statistics")

```

Table 4: Pilot survey summary statistics

| mean | s_squared_h | missing_prop |
|----------|--------------|--------------|
| 535087.4 | 297224769021 | 0.17 |

```

pilot_sample %>%
  group_by(quadrant) %>%
  summarize(mean = mean(property_value, na.rm = TRUE),
            s_squared_h = var(property_value, na.rm = TRUE),
            missing_prop = mean(is.na(property_value))) %>%
  kable(caption = "Pilot survey summary statistics by quadrant")

```

Table 5: Pilot survey summary statistics by quadrant

| quadrant | mean | s_squared_h | missing_prop |
|----------|----------|--------------|--------------|
| NE | 408489.5 | 55231295979 | 0.08 |
| NW | 928130.1 | 728182282168 | 0.12 |
| SE | 496448.4 | 136823871969 | 0.28 |
| SW | 283103.1 | 25025018879 | 0.20 |

Picking stratum sizes

Condition 1: Sample mean

We begin with a derivation of Exact Optimal Sample Allocation for \bar{y} .

Decomposition of $V(\bar{y}_h)$:

By Wright (12.4), $V(\bar{y}_{str}) = \sum_{h=1}^H (\frac{N_h}{N})^2 V(\bar{y}_h) = \sum_{h=1}^H (\frac{N_h}{N})^2 \frac{N_h - n_h}{N_h} \frac{S_h^2}{n_h}$

$$\begin{aligned}
V(\bar{y}_h) &= \left(\frac{N_h}{N}\right)^2 \frac{N_h - n_h}{N_h} \frac{S_h^2}{n_h} \\
V(\bar{y}_h) &= \left(\frac{N_h^2}{N^2}\right) \left(1 - \frac{n_h}{N_h}\right) \frac{S_h^2}{n_h} \\
V(\bar{y}_h) &= \left(\frac{N_h^2 S_h^2}{N^2}\right) \left(\frac{1}{n_h}\right) - \frac{N_h^2 n_h S_h^2}{N^2 N_h n_h} \\
V(\bar{y}_h) &= \left(\frac{N_h^2 S_h^2}{N^2}\right) \left(\frac{1}{n_h}\right) - \frac{N_h S_h^2}{N^2} \\
V(\bar{y}_h) &= \left(\frac{N_h^2 S_h^2}{N^2}\right) \left(1 - \frac{1}{1 \cdot 2} - \frac{1}{2 \cdot 3} - \dots - \frac{1}{n_h(n_h - 1)}\right) - \frac{N_h S_h^2}{N^2} \\
V(\bar{y}_h) &= \frac{N_h(N_h - 1)S_h^2}{N^2} - \frac{N_h^2 S_h^2}{N^2 \cdot 1 \cdot 2} - \frac{N_h^2 S_h^2}{N^2 \cdot 2 \cdot 3} - \dots - \frac{N_h^2 S_h^2}{N^2 n_h(n_h - 1)}
\end{aligned}$$

Decomposition of $V(\bar{y}_{str})$

$$\begin{aligned}
V(\bar{y}_{str}) &= \sum_{h=1}^H \frac{N_h(N_h - 1)S_h^2}{N^2} \\
&- \frac{N_1^2 S_1^2}{N^2 \cdot 1 \cdot 2} - \frac{N_1^2 S_1^2}{N^2 \cdot 2 \cdot 3} - \dots - \frac{N_1^2 S_1^2}{N^2 n_1(n_1 - 1)} \\
&\dots \\
&- \frac{N_h^2 S_h^2}{N^2 \cdot 1 \cdot 2} - \frac{N_h^2 S_h^2}{N^2 \cdot 2 \cdot 3} - \dots - \frac{N_h^2 S_h^2}{N^2 n_h(n_h - 1)} \\
&\dots \\
&- \frac{N_H^2 S_H^2}{N^2 \cdot 1 \cdot 2} - \frac{N_H^2 S_H^2}{N^2 \cdot 2 \cdot 3} - \dots - \frac{N_H^2 S_H^2}{N^2 n_H(n_H - 1)}
\end{aligned}$$

For a desired bound V_0 on the sampling variance $V(\bar{y}_{str})$, we may find an optimal allocation using the following algorithm:

- 1) Assign, for each stratum, 1 unit to be selected for the sample.
- 2) Fill in the following table and number these values starting from 1, in decreasing order.

| | | | |
|---|---|---|-----|
| $\frac{N_1^2 S_1^2}{N^2 \cdot 1 \cdot 2}$ | $\frac{N_1^2 S_1^2}{n^2 \cdot 2 \cdot 3}$ | $\frac{N_1^2 S_1^2}{N^2 \cdot 3 \cdot 4}$ | ... |
| $\frac{N_2^2 S_2^2}{N^2 \cdot 1 \cdot 2}$ | $\frac{N_2^2 S_2^2}{N^2 \cdot 2 \cdot 3}$ | $\frac{N_2^2 S_2^2}{N^2 \cdot 3 \cdot 4}$ | ... |
| . | . | . | ... |
| . | . | . | ... |
| . | . | . | ... |
| $\frac{N_H^2 S_H^2}{N^2 \cdot 1 \cdot 2}$ | $\frac{N_H^2 S_H^2}{N^2 \cdot 2 \cdot 3}$ | $\frac{N_H^2 S_H^2}{N^2 \cdot 3 \cdot 4}$ | ... |

- 3) Since the initial allocation is $(n_{11}, n_{21}, \dots, n_{H1}) = (1, 1, \dots, 1)$, compute $V(\bar{y}_{str} | n_{11} = 1, n_{21} = 1, \dots, n_{H1} = 1) = \sum_{h=1}^H \frac{N_h(N_h - 1)S_h^2}{N^2}$
- 4) Pick value (1) from the table and increase the associated stratum's sample

size by 1, so that the updated allocation is $(n_{12}, n_{22}, \dots, n_{H2})$, where exactly one of the n_{h2} 's is equal to 2 and the rest are equal to 1. Then, compute $V(\bar{y}_{str}|n_{12}, \dots, n_{H2}) = V(\bar{y}_{str}|n_{11}, \dots, n_{H1}) - \frac{1}{N^2}$ where "(1)" represents the largest value from the table. If $V(\bar{y}_{str}|n_{12}, \dots, n_{H2}) \leq V_0$, then stop with $n_1 = n_{12}, \dots, n_H = n_{H2}$. Otherwise, go to step 5.

- 5) Pick value (2) from the table and increase the associated stratum's sample size by 1, so that the updated allocation is (n_{13}, \dots, n_{H3}) . Then compute $V(\bar{y}_{str}|n_{13}, \dots, n_{H3}) = V(\bar{y}_{str}|n_{12}, \dots, n_{H2}) - \frac{(2)}{N^2}$, where "(2)" represents the second value from the table. If $V(\bar{y}_{str}|n_{13}, \dots, n_{H3}) \leq V_0$, then stop with $n_1 = n_{13}, \dots, n_H = n_{H3}$. Otherwise, continue until step j , where $V(\bar{y}_{str}|n_{1j}, \dots, n_{Hj}) \leq V_0$. The final allocation is n_{1j}, \dots, n_{Hj} and $n = n_{1j} + \dots + n_{Hj}$.

```
# find Nh and s2 for each strata
# (1) and (2)
s_squared_h <- pilot_sample %>%
  group_by(stratum = quadrant) %>%
  summarize(s_squared_h = var(property_value, na.rm = TRUE),
            missing_prop = mean(is.na(property_value)))

Nh <- sampling_frame %>%
  count(stratum = quadrant) %>%
  rename(Nh = n)

strata <- left_join(s_squared_h, Nh, by = "stratum") %>%
  # adjust N because of missingness
  mutate(Nh = Nh * (1 - missing_prop)) %>%
  mutate(N = sum(Nh))

rm(s_squared_h, Nh)

kable(strata)
```

| stratum | s_squared_h | missing_prop | Nh | N |
|---------|--------------|--------------|-----------|----------|
| NE | 55231295979 | 0.08 | 68953.08 | 297153.2 |
| NW | 728182282168 | 0.12 | 166931.60 | 297153.2 |
| SE | 136823871969 | 0.28 | 49423.68 | 297153.2 |
| SW | 25025018879 | 0.20 | 11844.80 | 297153.2 |

$$\text{Step 3: } \hat{V}(\bar{y}|1, 1, 1, 1) = \sum_{h=1}^H \left(\frac{N_h}{N}\right)^2 \frac{N_h - n_h}{N_H} \frac{s_h^2}{n_h} = \sum_{h=1}^H \left(\frac{N_h}{N}\right)^2 \frac{N_h - 1}{N_H} \frac{s_h^2}{1}$$

(Wright 12.5)

```
# Let the initial allocation be (n_11, n_21, n_31, n_41) = (1, 1, 1, 1)
# (3) and (6)
starting_variance <- strata %>%
  mutate(strata_variance = Nh * (Nh - 1) * s_squared_h / N^2)
```

```
kable(starting_variance)
```

| stratum | s_squared_h | missing_prop | Nh | N | strata_variance |
|---------|--------------|--------------|-----------|----------|-----------------|
| NE | 55231295979 | 0.08 | 68953.08 | 297153.2 | 2973894581 |
| NW | 728182282168 | 0.12 | 166931.60 | 297153.2 | 229802057101 |
| SE | 136823871969 | 0.28 | 49423.68 | 297153.2 | 3784970868 |
| SW | 25025018879 | 0.20 | 11844.80 | 297153.2 | 39758730 |

```
starting_variance <- starting_variance %>%
  summarize(V = sum(strata_variance)) %>%
  pull()

starting_variance
```

```
## [1] 236600681280
```

Step 3:

$$\text{Priority value} = \frac{N_1^2 \cdot s_1^2}{N_1^2 \cdot n_h(n_h - 1)}$$

```
# create a table of priority values
# (4) and (5)
n_strata <-
  tibble(n = c(1:500, 1:500, 1:500, 1:500),
    stratum = c(rep("NE", 500), rep("NW", 500), rep("SE", 500), rep("SW", 500))) %>%
  left_join(strata, by = "stratum")

# step 2
priority_values <- n_strata %>%
  group_by(stratum) %>%
  mutate(priority_value = (Nh ^ 2 * s_squared_h) / (n * lag(n) * N ^ 2)) %>%
  ungroup() %>%
  arrange(desc(priority_value))

kable(head(select(priority_values, -missing_prop), n = 10))
```

| n | stratum | s_squared_h | Nh | N | priority_value |
|----|---------|--------------|----------|----------|----------------|
| 2 | NW | 728182282168 | 166931.6 | 297153.2 | 114901716867 |
| 3 | NW | 728182282168 | 166931.6 | 297153.2 | 38300572289 |
| 4 | NW | 728182282168 | 166931.6 | 297153.2 | 19150286144 |
| 5 | NW | 728182282168 | 166931.6 | 297153.2 | 11490171687 |
| 6 | NW | 728182282168 | 166931.6 | 297153.2 | 7660114458 |
| 7 | NW | 728182282168 | 166931.6 | 297153.2 | 5471510327 |
| 8 | NW | 728182282168 | 166931.6 | 297153.2 | 4103632745 |
| 9 | NW | 728182282168 | 166931.6 | 297153.2 | 3191714357 |
| 10 | NW | 728182282168 | 166931.6 | 297153.2 | 2553371486 |
| 11 | NW | 728182282168 | 166931.6 | 297153.2 | 2089122125 |

Step 4:

```
# (7)
priority_values <- priority_values %>%
  mutate(agg_priority_value = cumsum(priority_value)) %>%
  mutate(marginal_variance = starting_variance - agg_priority_value) %>%
  mutate(marginal_sd = sqrt(marginal_variance))

kable(head(select(priority_values, -missing_prop, -N), n = 100), digits = 0)
```

| n | stratum | s_squared_h | Nh | priority_value | agg_priority_value | marginal_variance | marginal_sd |
|----|---------|--------------|--------|----------------|--------------------|-------------------|-------------|
| 2 | NW | 728182282168 | 166932 | 114901716867 | 114901716867 | 121698964413 | 348854 |
| 3 | NW | 728182282168 | 166932 | 38300572289 | 153202289155 | 83398392124 | 288788 |
| 4 | NW | 728182282168 | 166932 | 19150286144 | 172352575300 | 64248105980 | 253472 |
| 5 | NW | 728182282168 | 166932 | 11490171687 | 183842746987 | 52757934293 | 229691 |
| 6 | NW | 728182282168 | 166932 | 7660114458 | 191502861444 | 45097819835 | 212362 |
| 7 | NW | 728182282168 | 166932 | 5471510327 | 196974371771 | 39626309508 | 199064 |
| 8 | NW | 728182282168 | 166932 | 4103632745 | 201078004517 | 35522676763 | 188475 |
| 9 | NW | 728182282168 | 166932 | 3191714357 | 204269718874 | 32330962406 | 179808 |
| 10 | NW | 728182282168 | 166932 | 2553371486 | 206823090360 | 29777590920 | 172562 |
| 11 | NW | 728182282168 | 166932 | 2089122125 | 208912212485 | 27688468795 | 166399 |
| 2 | SE | 136823871969 | 49424 | 1892523726 | 210804736211 | 25795945069 | 160611 |
| 12 | NW | 728182282168 | 166932 | 1740935104 | 212545671315 | 24055009965 | 155097 |
| 2 | NE | 55231295979 | 68953 | 1486968855 | 214032640170 | 22568041110 | 150227 |
| 13 | NW | 728182282168 | 166932 | 1473098934 | 215505739104 | 21094942176 | 145241 |
| 14 | NW | 728182282168 | 166932 | 1262656229 | 216768395333 | 19832285946 | 140827 |
| 15 | NW | 728182282168 | 166932 | 1094302065 | 217862697399 | 18737983881 | 136887 |
| 16 | NW | 728182282168 | 166932 | 957514307 | 218820211706 | 17780469574 | 133343 |
| 17 | NW | 728182282168 | 166932 | 844865565 | 219665077271 | 16935604008 | 130137 |
| 18 | NW | 728182282168 | 166932 | 750991614 | 220416068885 | 16184612395 | 127219 |
| 19 | NW | 728182282168 | 166932 | 671939865 | 221088008749 | 15512672530 | 124550 |
| 3 | SE | 136823871969 | 49424 | 630841242 | 221718849991 | 14881831288 | 121991 |
| 20 | NW | 728182282168 | 166932 | 604745878 | 222323595870 | 14277085410 | 119487 |
| 21 | NW | 728182282168 | 166932 | 547151033 | 222870746902 | 13729934377 | 117175 |
| 22 | NW | 728182282168 | 166932 | 497410030 | 223368156932 | 13232524348 | 115033 |
| 3 | NE | 55231295979 | 68953 | 495656285 | 223863813217 | 12736868063 | 112858 |
| 23 | NW | 728182282168 | 166932 | 454156984 | 224317970201 | 12282711079 | 110827 |
| 24 | NW | 728182282168 | 166932 | 416310568 | 224734280769 | 11866400511 | 108933 |
| 25 | NW | 728182282168 | 166932 | 383005723 | 225117286492 | 11483394788 | 107161 |
| 26 | NW | 728182282168 | 166932 | 353543744 | 225470830236 | 11129851043 | 105498 |
| 27 | NW | 728182282168 | 166932 | 327355319 | 225798185555 | 10802495725 | 103935 |
| 4 | SE | 136823871969 | 49424 | 315420621 | 226113606176 | 10487075104 | 102406 |
| 28 | NW | 728182282168 | 166932 | 303972796 | 226417578972 | 10183102308 | 100911 |
| 29 | NW | 728182282168 | 166932 | 283009155 | 226700588127 | 9900093153 | 99499 |
| 30 | NW | 728182282168 | 166932 | 264141878 | 226964730005 | 9635951275 | 98163 |
| 4 | NE | 55231295979 | 68953 | 247828143 | 227212558147 | 9388123133 | 96892 |
| 31 | NW | 728182282168 | 166932 | 247100466 | 227459658613 | 9141022666 | 95609 |
| 32 | NW | 728182282168 | 166932 | 231656687 | 227691315301 | 8909365979 | 94389 |
| 33 | NW | 728182282168 | 166932 | 217616888 | 227908932189 | 8691749091 | 93230 |
| 34 | NW | 728182282168 | 166932 | 204815895 | 228113748083 | 8486933196 | 92125 |
| 35 | NW | 728182282168 | 166932 | 193112129 | 228306860212 | 8293821067 | 91070 |
| 5 | SE | 136823871969 | 49424 | 189252373 | 228496112585 | 8104568695 | 90025 |

| n | stratum | s_squared_h | Nh | priority_value | agg_priority_value | marginal_variance | marginal_sd |
|----|---------|--------------|--------|----------------|--------------------|-------------------|-------------|
| 36 | NW | 728182282168 | 166932 | 182383678 | 228678496263 | 7922185017 | 89007 |
| 37 | NW | 728182282168 | 166932 | 172525100 | 228851021363 | 7749659917 | 88032 |
| 38 | NW | 728182282168 | 166932 | 163444832 | 229014466195 | 7586215085 | 87099 |
| 39 | NW | 728182282168 | 166932 | 155063046 | 229169529241 | 7431152039 | 86204 |
| 5 | NE | 55231295979 | 68953 | 148696886 | 229318226126 | 7282455153 | 85337 |
| 40 | NW | 728182282168 | 166932 | 147309893 | 229465536020 | 7135145260 | 84470 |
| 41 | NW | 728182282168 | 166932 | 140124045 | 229605660065 | 6995021215 | 83636 |
| 42 | NW | 728182282168 | 166932 | 133451471 | 229739111536 | 6861569744 | 82835 |
| 43 | NW | 728182282168 | 166932 | 127244426 | 229866355962 | 6734325317 | 82063 |
| 6 | SE | 136823871969 | 49424 | 126168248 | 229992524211 | 6608157069 | 81291 |
| 44 | NW | 728182282168 | 166932 | 121460589 | 230113984799 | 6486696480 | 80540 |
| 45 | NW | 728182282168 | 166932 | 116062340 | 230230047139 | 6370634140 | 79816 |
| 46 | NW | 728182282168 | 166932 | 111016152 | 230341063291 | 6259617989 | 79118 |
| 47 | NW | 728182282168 | 166932 | 106292060 | 230447355351 | 6153325929 | 78443 |
| 48 | NW | 728182282168 | 166932 | 101863224 | 230549218575 | 6051462704 | 77791 |
| 6 | NE | 55231295979 | 68953 | 99131257 | 230648349832 | 5952331447 | 77151 |
| 49 | NW | 728182282168 | 166932 | 97705542 | 230746055374 | 5854625906 | 76516 |
| 50 | NW | 728182282168 | 166932 | 93797320 | 230839852694 | 5760828586 | 75900 |
| 7 | SE | 136823871969 | 49424 | 90120177 | 230929972871 | 5670708409 | 75304 |
| 51 | NW | 728182282168 | 166932 | 90118994 | 231020091865 | 5580589415 | 74703 |
| 52 | NW | 728182282168 | 166932 | 86652878 | 231106744743 | 5493936536 | 74121 |
| 53 | NW | 728182282168 | 166932 | 83382959 | 231190127702 | 5410553578 | 73556 |
| 54 | NW | 728182282168 | 166932 | 80294701 | 231270422403 | 5330258877 | 73009 |
| 55 | NW | 728182282168 | 166932 | 77374894 | 231347797296 | 5252883984 | 72477 |
| 56 | NW | 728182282168 | 166932 | 74611504 | 231422408801 | 5178272479 | 71960 |
| 57 | NW | 728182282168 | 166932 | 71993557 | 231494402357 | 5106278922 | 71458 |
| 7 | NE | 55231295979 | 68953 | 70808041 | 231565210398 | 5035470881 | 70961 |
| 58 | NW | 728182282168 | 166932 | 69511020 | 231634721419 | 4965959861 | 70470 |
| 8 | SE | 136823871969 | 49424 | 67590133 | 231702311552 | 4898369728 | 69988 |
| 59 | NW | 728182282168 | 166932 | 67154715 | 231769466266 | 4831215013 | 69507 |
| 60 | NW | 728182282168 | 166932 | 64916224 | 231834382491 | 4766298789 | 69038 |
| 61 | NW | 728182282168 | 166932 | 62787823 | 231897170314 | 4703510966 | 68582 |
| 62 | NW | 728182282168 | 166932 | 60762410 | 231957932724 | 4642748556 | 68138 |
| 63 | NW | 728182282168 | 166932 | 58833444 | 232016766168 | 4583915111 | 67705 |
| 64 | NW | 728182282168 | 166932 | 56994899 | 232073761067 | 4526920212 | 67282 |
| 65 | NW | 728182282168 | 166932 | 55241210 | 232129002277 | 4471679002 | 66871 |
| 66 | NW | 728182282168 | 166932 | 53567234 | 232182569511 | 4418111768 | 66469 |
| 8 | NE | 55231295979 | 68953 | 53106031 | 232235675542 | 4365005738 | 66068 |
| 9 | SE | 136823871969 | 49424 | 52570103 | 232288245645 | 4312435634 | 65669 |
| 67 | NW | 728182282168 | 166932 | 51968212 | 232340213858 | 4260467422 | 65272 |
| 68 | NW | 728182282168 | 166932 | 50439735 | 232390653593 | 4210027687 | 64885 |
| 69 | NW | 728182282168 | 166932 | 48977714 | 232439631307 | 4161049973 | 64506 |
| 70 | NW | 728182282168 | 166932 | 47578351 | 232487209657 | 4113471622 | 64136 |
| 71 | NW | 728182282168 | 166932 | 46238115 | 232533447773 | 4067233507 | 63775 |
| 72 | NW | 728182282168 | 166932 | 44953723 | 232578401496 | 4022279783 | 63421 |
| 73 | NW | 728182282168 | 166932 | 43722114 | 232622123611 | 3978557669 | 63076 |
| 74 | NW | 728182282168 | 166932 | 42540436 | 232664664046 | 3936017233 | 62738 |
| 10 | SE | 136823871969 | 49424 | 42056083 | 232706720129 | 3893961150 | 62402 |
| 75 | NW | 728182282168 | 166932 | 41406024 | 232748126153 | 3852555126 | 62069 |
| 9 | NE | 55231295979 | 68953 | 41304690 | 232789430844 | 3811250436 | 61735 |
| 76 | NW | 728182282168 | 166932 | 40316392 | 232829747236 | 3770934044 | 61408 |
| 77 | NW | 728182282168 | 166932 | 39269213 | 232869016448 | 3731664831 | 61087 |

| n | stratum | s_squared_h | Nh | priority_value | agg_priority_value | marginal_variance | marginal_sd |
|----|---------|--------------|--------|----------------|--------------------|-------------------|-------------|
| 78 | NW | 728182282168 | 166932 | 38262310 | 232907278758 | 3693402521 | 60773 |
| 79 | NW | 728182282168 | 166932 | 37293644 | 232944572402 | 3656108877 | 60466 |
| 80 | NW | 728182282168 | 166932 | 36361303 | 232980933705 | 3619747574 | 60164 |
| 81 | NW | 728182282168 | 166932 | 35463493 | 233016397198 | 3584284082 | 59869 |
| 82 | NW | 728182282168 | 166932 | 34598530 | 233050995728 | 3549685552 | 59579 |
| 11 | SE | 136823871969 | 49424 | 34409522 | 233085405250 | 3515276030 | 59290 |
| 83 | NW | 728182282168 | 166932 | 33764830 | 233119170080 | 3481511200 | 59004 |

```
rm(n_strata)
```

Condition 2: Sample means within strata

We are interested in comparing \bar{y}_h from the four different quadrants.

$$n = \frac{N\sigma^2}{(N-1)\frac{e^2}{z_{\frac{\alpha}{2}}^2} + \sigma^2}$$

We can use s^2 from our pilot survey as an unbiased estimate for σ^2 .

$$n = \frac{Ns^2}{(N-1)\frac{e^2}{z_{\frac{\alpha}{2}}^2} + s^2}$$

We want \$50,000 precision at a 90% confidence level for the mean of property value in each strata.

```
condition2 <- strata %>%
  mutate(n = (N * s_squared_h) / ((N - 1) * (50000 ^ 2 / qnorm(0.95)) + s_squared_h))

condition2 %>%
  kable()
```

| stratum | s_squared_h | missing_prop | Nh | N | n |
|---------|--------------|--------------|-----------|----------|-----------|
| NE | 55231295979 | 0.08 | 68953.08 | 297153.2 | 36.33464 |
| NW | 728182282168 | 0.12 | 166931.60 | 297153.2 | 478.33170 |
| SE | 136823871969 | 0.28 | 49423.68 | 297153.2 | 89.99514 |
| SW | 25025018879 | 0.20 | 11844.80 | 297153.2 | 16.46414 |

Condition 3: Sample proportion

We begin with a derivation of Exact Optimal Sample Allocation for \hat{p} .

Decomposition of $V(\hat{p}_{str})$

By Wright (12.14), $V(\hat{p}_{str}) = \sum_{h=1}^H (\frac{N_h}{N})^2 V(p_h) = \sum_{h=1}^H (\frac{N_h}{N})^2 \frac{N_h - n_h}{N_h - 1} \frac{p(1-p)}{n_h}$

$$V(\hat{p}_h) = (\frac{N_h}{N})^2 \frac{N_h - n_h}{N_h - 1} \frac{p(1-p)}{n_h}$$

$$V(\hat{p}_h) = \frac{N_h^2}{N^2} \frac{N_h}{N_h-1} \frac{p(1-p)}{n_h} - \frac{N_h^2}{N^2} \frac{n_h}{N_h-1} \frac{p(1-p)}{n_h}$$

$$V(\hat{p}_h) = \frac{N_h^2}{N^2} \frac{N_h}{N_h-1} \frac{p(1-p)}{n_h} - \frac{N_h^2}{N^2} \frac{n_h}{N_h-1} \frac{p(1-p)}{n_h}$$

$$V(\hat{p}_h) = \frac{N_h^3 p(1-p)}{N^2(N_h-1)} \frac{1}{n_h} - \frac{N_h^2 p(1-p)}{N^2(N_h-1)}$$

$$V(\hat{p}_h) = \frac{N_h^3 p(1-p)}{N^2(N_h-1)} \left(1 - \frac{1}{1 \cdot 2} - \frac{1}{2 \cdot 3} - \dots - \frac{1}{n_h(n_h-1)}\right) - \frac{N_h^2 p(1-p)}{N^2(N_h-1)}$$

$$V(\hat{p}_h) = \frac{N_h^3 p(1-p)}{N^2(N_h-1)} - \frac{N_h^3 p(1-p)}{N^2(N_h-1) \cdot 1 \cdot 2} - \frac{N_h^3 p(1-p)}{N^2(N_h-1) \cdot 2 \cdot 3} - \dots - \frac{N_h^3 p(1-p)}{N^2(N_h-1) \cdot n_h(n_h-1)} - \frac{N_h^2 p(1-p)}{N^2(N_h-1)}$$

$$V(\hat{p}_h) = \frac{(N_h^3 - N_h^2) p(1-p)}{N^2(N_h-1)} - \frac{N_h^3 p(1-p)}{N^2(N_h-1) \cdot 1 \cdot 2} - \frac{N_h^3 p(1-p)}{N^2(N_h-1) \cdot 2 \cdot 3} - \dots - \frac{N_h^3 p(1-p)}{N^2(N_h-1) \cdot n_h(n_h-1)}$$

$$V(\hat{p}_h) = \frac{N_h^2(N_h-1) p(1-p)}{N^2(N_h-1)} - \frac{N_h^3 p(1-p)}{N^2(N_h-1) \cdot 1 \cdot 2} - \frac{N_h^3 p(1-p)}{N^2(N_h-1) \cdot 2 \cdot 3} - \dots - \frac{N_h^3 p(1-p)}{N^2(N_h-1) \cdot n_h(n_h-1)}$$

Decomposition of $V(\hat{p}_{str})$

$$\begin{aligned} V(\hat{p}_{str}) &= \sum_{h=1}^H \frac{N_h^2(N_h-1) p(1-p)}{N^2(N_h-1)} \\ &- \frac{N_1^3 p(1-p)}{N^2(N_1-1) \cdot 1 \cdot 2} - \frac{N_1^3 p(1-p)}{N^2(N_1-1) \cdot 2 \cdot 3} - \dots - \frac{N_1^3 p(1-p)}{N^2(N_1-1) n_1(n_1-1)} \\ &\dots \\ &- \frac{N_h^3 p(1-p)}{N^2(N_h-1) \cdot 1 \cdot 2} - \frac{N_h^3 p(1-p)}{N^2(N_h-1) \cdot 2 \cdot 3} - \dots - \frac{N_h^3 p(1-p)}{N^2(N_h-1) n_h(n_h-1)} \\ &\dots \\ &- \frac{N_H^3 p(1-p)}{N^2(N_H-1) \cdot 1 \cdot 2} - \frac{N_H^3 p(1-p)}{N^2(N_H-1) \cdot 2 \cdot 3} - \dots - \frac{N_H^3 p(1-p)}{N^2(N_H-1) n_H(n_H-1)} \end{aligned}$$

For a desired bound on V_0 on the sampling variance $V(\hat{p}_{str})$, we may find an optimal allocation using the following algorithm:

- 1) Assign, for each stratum, 1 unit to be selected for the sample.
- 2) Fill in the following table and number these values starting from 1, in decreasing order.
We assume $p_h = 0.5$ because that is where the variance reaches its global maximum.

| | | | |
|--|--|--|---------|
| $\frac{\frac{1}{4} N_1^3}{N^2(N_1-1) \cdot 1 \cdot 2}$ | $\frac{\frac{1}{4} N_1^3}{N^2(N_1-1) \cdot 2 \cdot 3}$ | $\frac{\frac{1}{4} N_1^3}{N^2(N_1-1) \cdot 3 \cdot 4}$ | \dots |
| $\frac{\frac{1}{4} N_2^3}{N^2(N_2-1) \cdot 1 \cdot 2}$ | $\frac{\frac{1}{4} N_2^3}{N^2(N_2-1) \cdot 2 \cdot 3}$ | $\frac{\frac{1}{4} N_2^3}{N^2(N_2-1) \cdot 3 \cdot 4}$ | \dots |
| \cdot | \cdot | \cdot | \dots |
| \cdot | \cdot | \cdot | \dots |
| \cdot | \cdot | \cdot | \dots |
| $\frac{\frac{1}{4} N_H^3}{N^2(N_H-1) \cdot 1 \cdot 2}$ | $\frac{\frac{1}{4} N_H^3}{N^2(N_H-1) \cdot 2 \cdot 3}$ | $\frac{\frac{1}{4} N_H^3}{N^2(N_H-1) \cdot 3 \cdot 4}$ | \dots |

- 3) Since the initial allocation is $(n_{11}, n_{21}, \dots, n_{H1}) = (1, 1, \dots, 1)$, compute $V(\hat{p}_{str} | n_{11} = 1, n_{21} = 1, \dots, n_{H1} = 1) = \frac{1}{N^2} \sum_{h=1}^H ((N_h^2 - N_h) S_h^2)$

- 4) Pick value (1) from the table and increase the associated stratum's sample size by 1, so that the updated allocation is $(n_{12}, n_{22}, \dots, n_{H2})$, where exactly one of the n_{h2} 's is equal to 2 and the rest are equal to 1. Then, compute $V(\hat{p}_{str}|n_{12}, \dots, n_{H2}) = V(\hat{p}_{str}|n_{11}, \dots, n_{H1}) - \frac{1}{N^2}$ where "(1)" represents the largest value from the table. If $V(\hat{p}_{str}|N_{12}, \dots, n_{H2}) \leq V_0$, then stop with $n_1 = n_{12}, \dots, N_H = N_{H2}$. Otherwise, go to step 5.
- 5) Pick value (2) from the table and increase the associated stratum's sample size by 1, so that the updated allocation is (n_{13}, \dots, n_{H3}) . Then compute $V(\hat{p}_{str}|n_{13}, \dots, n_{H3}) = V(\hat{p}_{str}|n_{12}, \dots, n_{H2}) - \frac{(2)}{N^2}$, where "(2)" represents the second value from the table. If $V(\hat{p}_{str}|n_{13}, \dots, N_H = n_{H3}) \leq V_0$, then stop with $n_1 = n_{13}, \dots, N_H = n_{H3}$. Otherwise, continue until step j , where $V(\hat{p}_{str}|n_{1j}, \dots, n_{Hj}) \leq V_0$. The final allocation is n_{1j}, \dots, n_{Hj} and $n = n_{1j} + \dots + n_{Hj}$.

```
#
strata <- sampling_frame %>%
  count(stratum = quadrant) %>%
  rename(Nh = n) %>%
  mutate(N = sum(Nh),
         s_squared_h = 0.5 * (1 - 0.5))

kable(strata)
```

| stratum | Nh | N | s_squared_h |
|---------|--------|--------|-------------|
| NE | 74949 | 348094 | 0.25 |
| NW | 189695 | 348094 | 0.25 |
| SE | 68644 | 348094 | 0.25 |
| SW | 14806 | 348094 | 0.25 |

```
# Let the initial allocation be (n_11, n_21, n_31, n_41) = (1, 1, 1, 1)
# (3) and (6)
starting_variance <- strata %>%
  mutate(strata_variance = (Nh / N) ^ 2 * ((Nh - 1) / Nh) * (s_squared_h / 1))

kable(starting_variance)
```

| stratum | Nh | N | s_squared_h | strata_variance |
|---------|--------|--------|-------------|-----------------|
| NE | 74949 | 348094 | 0.25 | 0.0115897 |
| NW | 189695 | 348094 | 0.25 | 0.0742432 |
| SE | 68644 | 348094 | 0.25 | 0.0097218 |
| SW | 14806 | 348094 | 0.25 | 0.0004523 |

```
starting_variance <- starting_variance %>%
  summarize(V = sum(strata_variance)) %>%
  pull()

starting_variance
```

```
## [1] 0.09600692
```

```
# create a table of priority values
# (4) and (5)

n_strata <- tibble(n = c(1:500, 1:500, 1:500, 1:500),
                  stratum = c(rep("NE", 500),
                              rep("NW", 500),
                              rep("SE", 500),
                              rep("SW", 500))) %>%
  left_join(strata, by = "stratum")

# step 2
priority_values <- n_strata %>%
  group_by(stratum) %>%
  mutate(priority_value = (0.25 * Nh ^ 3) / (N ^ 2 * (Nh - 1) * n * lag(n))) %>%
  ungroup() %>%
  arrange(desc(priority_value))

kable(head(priority_values, n = 10))
```

| n | stratum | Nh | N | s_squared_h | priority_value |
|---|---------|--------|--------|-------------|----------------|
| 2 | NW | 189695 | 348094 | 0.25 | 0.0371220 |
| 3 | NW | 189695 | 348094 | 0.25 | 0.0123740 |
| 4 | NW | 189695 | 348094 | 0.25 | 0.0061870 |
| 2 | NE | 74949 | 348094 | 0.25 | 0.0057950 |
| 2 | SE | 68644 | 348094 | 0.25 | 0.0048610 |
| 5 | NW | 189695 | 348094 | 0.25 | 0.0037122 |
| 6 | NW | 189695 | 348094 | 0.25 | 0.0024748 |
| 3 | NE | 74949 | 348094 | 0.25 | 0.0019317 |
| 7 | NW | 189695 | 348094 | 0.25 | 0.0017677 |
| 3 | SE | 68644 | 348094 | 0.25 | 0.0016203 |

```
# (7)
priority_values <- priority_values %>%
  mutate(agg_priority_value = cumsum(priority_value)) %>%
  mutate(marginal_variance = starting_variance - agg_priority_value) %>%
  mutate(marginal_sd = sqrt(marginal_variance))

kable(head(select(priority_values, -N), n = 100), align = "l")
```

| n | stratum | Nh | s_squared_h | priority_value | agg_priority_value | marginal_variance | marginal_sd |
|---|---------|--------|-------------|----------------|--------------------|-------------------|-------------|
| 2 | NW | 189695 | 0.25 | 0.0371220 | 0.0371220 | 0.0588849 | 0.2426622 |
| 3 | NW | 189695 | 0.25 | 0.0123740 | 0.0494960 | 0.0465110 | 0.2156640 |
| 4 | NW | 189695 | 0.25 | 0.0061870 | 0.0556830 | 0.0403240 | 0.2008083 |
| 2 | NE | 74949 | 0.25 | 0.0057950 | 0.0614780 | 0.0345289 | 0.1858197 |
| 2 | SE | 68644 | 0.25 | 0.0048610 | 0.0663390 | 0.0296679 | 0.1722438 |
| 5 | NW | 189695 | 0.25 | 0.0037122 | 0.0700512 | 0.0259557 | 0.1611078 |
| 6 | NW | 189695 | 0.25 | 0.0024748 | 0.0725260 | 0.0234809 | 0.1532349 |

| n | stratum | Nh | s_squared_h | priority_value | agg_priority_value | marginal_variance | marginal_sd |
|----|---------|--------|-------------|----------------|--------------------|-------------------|-------------|
| 3 | NE | 74949 | 0.25 | 0.0019317 | 0.0744577 | 0.0215493 | 0.1467966 |
| 7 | NW | 189695 | 0.25 | 0.0017677 | 0.0762254 | 0.0197815 | 0.1406469 |
| 3 | SE | 68644 | 0.25 | 0.0016203 | 0.0778457 | 0.0181612 | 0.1347635 |
| 8 | NW | 189695 | 0.25 | 0.0013258 | 0.0791715 | 0.0168354 | 0.1297513 |
| 9 | NW | 189695 | 0.25 | 0.0010312 | 0.0802027 | 0.0158042 | 0.1257149 |
| 4 | NE | 74949 | 0.25 | 0.0009658 | 0.0811685 | 0.0148384 | 0.1218130 |
| 10 | NW | 189695 | 0.25 | 0.0008249 | 0.0819934 | 0.0140135 | 0.1183785 |
| 4 | SE | 68644 | 0.25 | 0.0008102 | 0.0828036 | 0.0132033 | 0.1149056 |
| 11 | NW | 189695 | 0.25 | 0.0006749 | 0.0834786 | 0.0125284 | 0.1119302 |
| 5 | NE | 74949 | 0.25 | 0.0005795 | 0.0840581 | 0.0119489 | 0.1093108 |
| 12 | NW | 189695 | 0.25 | 0.0005625 | 0.0846205 | 0.0113864 | 0.1067071 |
| 5 | SE | 68644 | 0.25 | 0.0004861 | 0.0851066 | 0.0109003 | 0.1044045 |
| 13 | NW | 189695 | 0.25 | 0.0004759 | 0.0855825 | 0.0104244 | 0.1020999 |
| 14 | NW | 189695 | 0.25 | 0.0004079 | 0.0859905 | 0.0100164 | 0.1000822 |
| 6 | NE | 74949 | 0.25 | 0.0003863 | 0.0863768 | 0.0096301 | 0.0981331 |
| 15 | NW | 189695 | 0.25 | 0.0003535 | 0.0867303 | 0.0092766 | 0.0963149 |
| 6 | SE | 68644 | 0.25 | 0.0003241 | 0.0870544 | 0.0089525 | 0.0946177 |
| 16 | NW | 189695 | 0.25 | 0.0003093 | 0.0873638 | 0.0086432 | 0.0929685 |
| 7 | NE | 74949 | 0.25 | 0.0002760 | 0.0876397 | 0.0083672 | 0.0914724 |
| 17 | NW | 189695 | 0.25 | 0.0002730 | 0.0879127 | 0.0080942 | 0.0899680 |
| 18 | NW | 189695 | 0.25 | 0.0002426 | 0.0881553 | 0.0078516 | 0.0886093 |
| 7 | SE | 68644 | 0.25 | 0.0002315 | 0.0883868 | 0.0076201 | 0.0872934 |
| 2 | SW | 14806 | 0.25 | 0.0002262 | 0.0886129 | 0.0073940 | 0.0859882 |
| 19 | NW | 189695 | 0.25 | 0.0002171 | 0.0888300 | 0.0071769 | 0.0847165 |
| 8 | NE | 74949 | 0.25 | 0.0002070 | 0.0890370 | 0.0069699 | 0.0834861 |
| 20 | NW | 189695 | 0.25 | 0.0001954 | 0.0892324 | 0.0067745 | 0.0823076 |
| 21 | NW | 189695 | 0.25 | 0.0001768 | 0.0894091 | 0.0065978 | 0.0812267 |
| 8 | SE | 68644 | 0.25 | 0.0001736 | 0.0895828 | 0.0064242 | 0.0801509 |
| 9 | NE | 74949 | 0.25 | 0.0001610 | 0.0897437 | 0.0062632 | 0.0791403 |
| 22 | NW | 189695 | 0.25 | 0.0001607 | 0.0899044 | 0.0061025 | 0.0781184 |
| 23 | NW | 189695 | 0.25 | 0.0001467 | 0.0900512 | 0.0059558 | 0.0771736 |
| 9 | SE | 68644 | 0.25 | 0.0001350 | 0.0901862 | 0.0058207 | 0.0762937 |
| 24 | NW | 189695 | 0.25 | 0.0001345 | 0.0903207 | 0.0056862 | 0.0754071 |
| 10 | NE | 74949 | 0.25 | 0.0001288 | 0.0904495 | 0.0055575 | 0.0745483 |
| 25 | NW | 189695 | 0.25 | 0.0001237 | 0.0905732 | 0.0054337 | 0.0737137 |
| 26 | NW | 189695 | 0.25 | 0.0001142 | 0.0906874 | 0.0053195 | 0.0729349 |
| 10 | SE | 68644 | 0.25 | 0.0001080 | 0.0907954 | 0.0052115 | 0.0721905 |
| 27 | NW | 189695 | 0.25 | 0.0001058 | 0.0909012 | 0.0051057 | 0.0714543 |
| 11 | NE | 74949 | 0.25 | 0.0001054 | 0.0910066 | 0.0050003 | 0.0707131 |
| 28 | NW | 189695 | 0.25 | 0.0000982 | 0.0911048 | 0.0049021 | 0.0700153 |
| 29 | NW | 189695 | 0.25 | 0.0000914 | 0.0911962 | 0.0048107 | 0.0693593 |
| 11 | SE | 68644 | 0.25 | 0.0000884 | 0.0912846 | 0.0047223 | 0.0687192 |
| 12 | NE | 74949 | 0.25 | 0.0000878 | 0.0913724 | 0.0046345 | 0.0680773 |
| 30 | NW | 189695 | 0.25 | 0.0000853 | 0.0914577 | 0.0045492 | 0.0674476 |
| 31 | NW | 189695 | 0.25 | 0.0000798 | 0.0915376 | 0.0044694 | 0.0668532 |
| 3 | SW | 14806 | 0.25 | 0.0000754 | 0.0916130 | 0.0043940 | 0.0662870 |
| 32 | NW | 189695 | 0.25 | 0.0000748 | 0.0916878 | 0.0043191 | 0.0657200 |
| 13 | NE | 74949 | 0.25 | 0.0000743 | 0.0917621 | 0.0042448 | 0.0651523 |
| 12 | SE | 68644 | 0.25 | 0.0000737 | 0.0918357 | 0.0041712 | 0.0645846 |
| 33 | NW | 189695 | 0.25 | 0.0000703 | 0.0919060 | 0.0041009 | 0.0640380 |
| 34 | NW | 189695 | 0.25 | 0.0000662 | 0.0919722 | 0.0040347 | 0.0635193 |
| 14 | NE | 74949 | 0.25 | 0.0000637 | 0.0920359 | 0.0039710 | 0.0630160 |

| n | stratum | Nh | s_squared_h | priority_value | agg_priority_value | marginal_variance | marginal_sd |
|----|---------|--------|-------------|----------------|--------------------|-------------------|-------------|
| 35 | NW | 189695 | 0.25 | 0.0000624 | 0.0920983 | 0.0039086 | 0.0625190 |
| 13 | SE | 68644 | 0.25 | 0.0000623 | 0.0921606 | 0.0038463 | 0.0620186 |
| 36 | NW | 189695 | 0.25 | 0.0000589 | 0.0922195 | 0.0037874 | 0.0615417 |
| 37 | NW | 189695 | 0.25 | 0.0000557 | 0.0922753 | 0.0037316 | 0.0610872 |
| 15 | NE | 74949 | 0.25 | 0.0000552 | 0.0923305 | 0.0036765 | 0.0606337 |
| 14 | SE | 68644 | 0.25 | 0.0000534 | 0.0923839 | 0.0036230 | 0.0601916 |
| 38 | NW | 189695 | 0.25 | 0.0000528 | 0.0924367 | 0.0035702 | 0.0597514 |
| 39 | NW | 189695 | 0.25 | 0.0000501 | 0.0924868 | 0.0035201 | 0.0593307 |
| 16 | NE | 74949 | 0.25 | 0.0000483 | 0.0925351 | 0.0034718 | 0.0589223 |
| 40 | NW | 189695 | 0.25 | 0.0000476 | 0.0925827 | 0.0034242 | 0.0585171 |
| 15 | SE | 68644 | 0.25 | 0.0000463 | 0.0926290 | 0.0033780 | 0.0581201 |
| 41 | NW | 189695 | 0.25 | 0.0000453 | 0.0926742 | 0.0033327 | 0.0577294 |
| 42 | NW | 189695 | 0.25 | 0.0000431 | 0.0927173 | 0.0032896 | 0.0573547 |
| 17 | NE | 74949 | 0.25 | 0.0000426 | 0.0927600 | 0.0032470 | 0.0569821 |
| 43 | NW | 189695 | 0.25 | 0.0000411 | 0.0928011 | 0.0032058 | 0.0566202 |
| 16 | SE | 68644 | 0.25 | 0.0000405 | 0.0928416 | 0.0031653 | 0.0562613 |
| 44 | NW | 189695 | 0.25 | 0.0000392 | 0.0928808 | 0.0031261 | 0.0559115 |
| 18 | NE | 74949 | 0.25 | 0.0000379 | 0.0929187 | 0.0030882 | 0.0555718 |
| 4 | SW | 14806 | 0.25 | 0.0000377 | 0.0929564 | 0.0030505 | 0.0552316 |
| 45 | NW | 189695 | 0.25 | 0.0000375 | 0.0929939 | 0.0030130 | 0.0548911 |
| 46 | NW | 189695 | 0.25 | 0.0000359 | 0.0930298 | 0.0029772 | 0.0545634 |
| 17 | SE | 68644 | 0.25 | 0.0000357 | 0.0930655 | 0.0029414 | 0.0542349 |
| 47 | NW | 189695 | 0.25 | 0.0000343 | 0.0930998 | 0.0029071 | 0.0539173 |
| 19 | NE | 74949 | 0.25 | 0.0000339 | 0.0931337 | 0.0028732 | 0.0536022 |
| 48 | NW | 189695 | 0.25 | 0.0000329 | 0.0931666 | 0.0028403 | 0.0532943 |
| 18 | SE | 68644 | 0.25 | 0.0000318 | 0.0931984 | 0.0028085 | 0.0529954 |
| 49 | NW | 189695 | 0.25 | 0.0000316 | 0.0932300 | 0.0027769 | 0.0526967 |
| 20 | NE | 74949 | 0.25 | 0.0000305 | 0.0932605 | 0.0027464 | 0.0524065 |
| 50 | NW | 189695 | 0.25 | 0.0000303 | 0.0932908 | 0.0027161 | 0.0521166 |
| 51 | NW | 189695 | 0.25 | 0.0000291 | 0.0933199 | 0.0026870 | 0.0518365 |
| 19 | SE | 68644 | 0.25 | 0.0000284 | 0.0933483 | 0.0026586 | 0.0515616 |
| 52 | NW | 189695 | 0.25 | 0.0000280 | 0.0933763 | 0.0026306 | 0.0512894 |
| 21 | NE | 74949 | 0.25 | 0.0000276 | 0.0934039 | 0.0026030 | 0.0510197 |
| 53 | NW | 189695 | 0.25 | 0.0000269 | 0.0934308 | 0.0025761 | 0.0507550 |
| 54 | NW | 189695 | 0.25 | 0.0000259 | 0.0934568 | 0.0025501 | 0.0504988 |
| 20 | SE | 68644 | 0.25 | 0.0000256 | 0.0934824 | 0.0025245 | 0.0502448 |
| 22 | NE | 74949 | 0.25 | 0.0000251 | 0.0935075 | 0.0024995 | 0.0499946 |
| 55 | NW | 189695 | 0.25 | 0.0000250 | 0.0935325 | 0.0024745 | 0.0497439 |
| 56 | NW | 189695 | 0.25 | 0.0000241 | 0.0935566 | 0.0024504 | 0.0495010 |
| 57 | NW | 189695 | 0.25 | 0.0000233 | 0.0935798 | 0.0024271 | 0.0492655 |
| 21 | SE | 68644 | 0.25 | 0.0000231 | 0.0936030 | 0.0024039 | 0.0490300 |

```
rm(pilot_sample, n_strata)
```

Condition 4: Sample proportion within strata

We are interested in comparing \hat{p}_h from the four different quadrants.

$$n = \frac{Np(1-p)}{(N-1)\frac{e^2}{z^2} + p(1-p)}$$

We can assume that $p = 0.5$.

$$n = \frac{\frac{1}{4}N}{(N-1)\frac{e^2}{z_{\frac{\alpha}{2}}^2} + \frac{1}{4}}$$

We want 0.1 precision at a 90% confidence level for the mean of proportion with multi-family zoning in each strata.

```
condition4 <- strata %>%
  mutate(n = (N * 0.25) / ((N - 1) * (0.1 ^ 2 / qnorm(0.95)) + 0.25))

condition4 %>%
  kable()
```

| stratum | Nh | N | s_squared_h | n |
|---------|--------|--------|-------------|---------|
| NE | 74949 | 348094 | 0.25 | 41.1166 |
| NW | 189695 | 348094 | 0.25 | 41.1166 |
| SE | 68644 | 348094 | 0.25 | 41.1166 |
| SW | 14806 | 348094 | 0.25 | 41.1166 |

Combining the above conditions

We want to sample at a rate that meets the four different requirements from above

1. $V_0 > V(\bar{y}_{str})$ for the sample mean
2. \$50,000 precision at a 90% confidence level for \bar{y}_h in each strata
3. $V_0 > V(\hat{p}_h)$ for the sample proportion
4. 0.1 precision at a 90% confidence level for \hat{p} in each strata

```
tibble(`2.` = condition2$n,
       `4.` = condition4$n) %>%
  kable(caption = "Recommended strata sizes across the four conditions")
```

Table 18: Recommended strata sizes across the four conditions

| | 2. | 4. |
|--|-----------|---------|
| | 36.33464 | 41.1166 |
| | 478.33170 | 41.1166 |
| | 89.99514 | 41.1166 |
| | 16.46414 | 41.1166 |