

# VoterRepresentationAnalysis

October 17, 2018

```
In [1]: import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import statsmodels.regression.linear_model as sm
```

## 1 How Toronto's New Wards Change Voter Representation

This analysis will look at the census profile data for each ward

**Notes:** \* Think about a grouping wards into big, medium, and small \* Measure change in inverse of size of ward (i.e. how much "representation" does each person have) \* Histogram of percent of immigrants in each region (look at "dilution" / wards with high proportion of "protected group" that don't exist anymore) \* IQSS consulting \* Mention that this is sample from stats. Also, mention that well designed sample is as good as entire population \* subtraction is more interpretable than division

**Factors to Analyze:** \* Overall increase in ward size \* Income \* Language \* First-gen immigrants \* Race \* Education \* Employment \* Rent vs. Home owners

## 2 Overall Increase in Size of Ward

```
In [2]: full_before = pd.read_csv('./Data/FullPop/full_47.csv', index_col=0).transpose()
full_after = pd.read_csv('./Data/FullPop/full_25.csv', index_col=0).transpose()
```

```
In [3]: full_before.head()
```

```
Out[3]:
```

	total
Toronto	2731570
Ward 1	60735
Ward 2	57210
Ward 3	64065
Ward 4	66060

```
In [4]: full_after.head()
```

```
Out[4]:
```

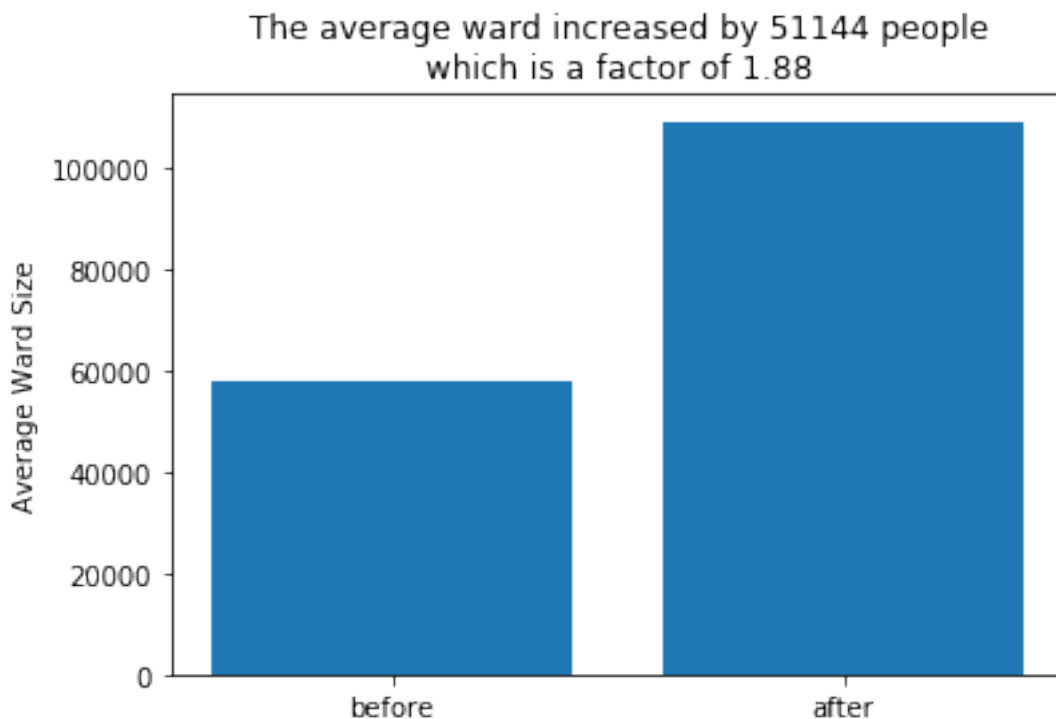
	total
Toronto	2731570
Ward 1	118040
Ward 2	118020

```
Ward 3    129080
Ward 4    108805
```

```
In [5]: toronto_pop = full_after.loc['Toronto'][0]
        print('The population of Toronto is ', toronto_pop)
```

The population of Toronto is 2731570

```
In [6]: avg_ward_before = np.mean(full_before[1:].total)
        avg_ward_after = np.mean(full_after[1:].total)
        label = ['before', 'after']
        plt.bar(label, [avg_ward_before, avg_ward_after])
        plt.ylabel('Average Ward Size')
        plt.title('The average ward increased by {} people\nwhich is a factor of {}'.format(int(avg_ward_after - avg_ward_before), round(avg_ward_after/avg_ward_before)))
        plt.show()
```



### 3 Function to calculate average increase (in group vs. out group)

```
In [7]: def scale_populations(df_before, df_after, full_before = full_before, full_after = full_after):
        #scale population
        scale_factors_before = full_before[1:].total/df_before[1:].total
```

```

df_before_scaled = df_before[1:].mul(scale_factors_before, axis = 0)

scale_factors_after = full_after[1:].total/df_after[1:].total
df_after_scaled = df_after[1:].mul(scale_factors_after, axis = 0)
return df_before_scaled, df_after_scaled

In [8]: def get_increases(total_before, protected_before, majority_before, total_after, protected_after, majority_after):
    # get protected group ward size increase
    avg_ward_size_protected_before = np.average(total_before, weights = protected_before)
    avg_ward_size_protected_after = np.average(total_after, weights = protected_after)
    avg_increase_protected = avg_ward_size_protected_after - avg_ward_size_protected_before

    # get majority group ward size increase
    avg_ward_size_majority_before = np.average(total_before, weights = majority_before)
    avg_ward_size_majority_after = np.average(total_after, weights = majority_after)
    avg_increase_majority = avg_ward_size_majority_after - avg_ward_size_majority_before

    return int(avg_increase_protected), int(avg_increase_majority)

```

## 4 Income

The low income threshold is provided by the census data and needs to be defined.

```

In [9]: income_before = pd.read_csv('./Data/Income/Income_47.csv', index_col=0).transpose()
        income_after = pd.read_csv('./Data/Income/Income_25.csv', index_col=0).transpose()

```

```

In [10]: income_before.head()

```

```

Out[10]:
           total  low_income  prop_low_income
Toronto  2691665.0    543365.0             20.2
Ward 1     60040.0     13875.0             23.1
Ward 2     56300.0     11345.0             20.2
Ward 3     63165.0      7575.0             12.0
Ward 4     65335.0      7850.0             12.0

```

```

In [11]: income_after.head()

```

```

Out[11]:
           total  low_income  prop_low_income
Toronto  2691665.0    543365.0             20.2
Ward 1     116955.0     26300.0             22.5
Ward 2     116055.0     13525.0             11.7
Ward 3     127525.0     19640.0             15.4
Ward 4     106445.0     18535.0             17.4

```

```

In [12]: income_before, income_after = scale_populations(income_before, income_after)

```

```

In [13]: income_before.head()

```

```
Out [13]:
```

	total	low_income	prop_low_income
Ward 1	60735.0	14035.611676	23.367397
Ward 2	57210.0	11528.373890	20.526501
Ward 3	64065.0	7682.931608	12.170981
Ward 4	66060.0	7937.108747	12.133160
Ward 5	49520.0	7285.027545	14.852969

```
In [14]: # calculated high income
income_before['high_income'] = income_before.total - income_before.low_income
income_after['high_income'] = income_after.total - income_after.low_income
```

```
In [56]: # Initialize params
total_before = income_before.total
protected_before = income_before.low_income
majority_before = income_before.high_income

total_after = income_after.total
protected_after = income_after.low_income
majority_after = income_after.high_income

protected_increase, majority_increase = get_increases(total_before, protected_before,
```

```
In [57]: print("If you are low-income, your ward size increased by {} people on average".format(
print("If you are high-income, your ward size increased by {} people on average".format(
print("This is a disparity of {} people on average".format(protected_increase-majority_incre
```

If you are low-income, your ward size increased by 50281 people on average  
If you are high-income, your ward size increased by 49989 people on average  
This is a disparity of 292 people on average

```
In [61]: # lost majorities
prop_protected_before = protected_before / total_before
print("wards with majority low-income before: {}/47".format(len([p for p in prop_protected_before if p > 0.5])))

prop_protected_after = protected_after / total_after
print("wards with majority low-income before: {}/25".format(len([p for p in prop_protected_after if p > 0.5])))
```

wards with majority low-income before: 0/47  
wards with majority low-income before: 0/25

## 5 Visible Minority

```
In [62]: race_before = pd.read_csv('./Data/Race/race_before.csv', index_col=0).transpose()
race_after = pd.read_csv('./Data/Race/race_after.csv', index_col=0).transpose()
```

```
In [63]: race_before.head()
```

```

Out [63]:
      total  visible_minority  south_asian  chinese  black  filipino  \
Toronto  2691665          1385850        338960   299465  239855   152715
Ward 1    60045           49875         20615     700   13935    2190
Ward 2    56295           35405         12210     805   11990    2050
Ward 3    63170           18215          4020    2450   3320    1975
Ward 4    65330           18755          4315    1620   5135    1375

      latin_american  arab  southeast_asian  west_asian  korean  japanese  \
Toronto          77165  36030          41650        60320   41640    13410
Ward 1           2410   3120          1065         1715    140      40
Ward 2           3045   525           925         1015    325     100
Ward 3           1630   675           540          840   1380    275
Ward 4           2495   385           305          590   1055    285

      other_visible_minoirty  multiple_visible_minorities  \
Toronto          36975          47670
Ward 1           2715          1235
Ward 2           1300          1115
Ward 3            380           735
Ward 4            555           650

      not_visible_minority
Toronto          1305815
Ward 1           10165
Ward 2           20895
Ward 3           44955
Ward 4           46575

```

```
In [64]: race_after.head()
```

```

Out [64]:
      total  visible_minority  south_asian  chinese  black  filipino  \
Toronto  2691665          1385855        338965   299465  239850   152715
Ward 1    116955           88495         33825    1635   27365    4460
Ward 2    116055           31445          6910    3530   6825    2835
Ward 3    127525           35355          7050    4445   6325    4255
Ward 4    106445           27865          5010    3710   5625    2805

      latin_american  arab  southeast_asian  west_asian  korean  japanese  \
Toronto          77165  36030          41645        60325   41640    13415
Ward 1           5605   3705          2015         2790    480     145
Ward 2           3725   885           715         1315   2165     500
Ward 3           3060  1155          1315          915   2425     655
Ward 4           2320   700          1595          585   1015     730

      other_visible_minoirty  multiple_visible_minorities  \
Toronto          36975          47670
Ward 1           4115          2355
Ward 2            805          1235

```

Ward 3	2185	1575
Ward 4	2255	1515

	not_visible_minority
Toronto	1305815
Ward 1	28460
Ward 2	84615
Ward 3	92165
Ward 4	78580

```
In [65]: race_before, race_after = scale_populations(race_before, race_after)
```

```
In [66]: race_before.head()
```

```
Out[66]:
```

	total	visible_minority	south_asian	chinese	black \
Ward 1	60735.0	50448.132651	20851.894829	708.043967	14095.132401
Ward 2	57210.0	35980.460965	12408.457234	818.084199	12184.881428
Ward 3	64065.0	18473.072265	4076.955833	2484.711889	3367.038151
Ward 4	66060.0	18964.569111	4363.215980	1638.101944	5192.378693
Ward 5	49520.0	14398.286064	3324.235870	1606.545603	2404.766374

	filipino	latin_american	arab	southeast_asian \
Ward 1	2215.166125	2437.694229	3155.853110	1077.238321
Ward 2	2083.320011	3094.492406	533.533173	940.034639
Ward 3	2002.982033	1653.094032	684.563479	547.650784
Ward 4	1390.364304	2522.879229	389.302005	308.408082
Ward 5	1404.464395	1111.446644	464.786778	535.515201

	west_asian	korean	japanese	other_visible_minority \
Ward 1	1734.707719	141.608793	40.459655	2746.199101
Ward 2	1031.497469	330.282441	101.625366	1321.129763
Ward 3	851.901219	1399.552003	278.896232	385.383885
Ward 4	596.592683	1066.788612	288.184601	561.201592
Ward 5	399.110386	1530.765150	267.757600	823.480922

	multiple_visible_minorities	not_visible_minority
Ward 1	1249.191856	10281.809893
Ward 2	1133.122835	21234.620304
Ward 3	745.413567	45591.927735
Ward 4	657.263126	47095.430889
Ward 5	535.515201	35121.713936

```
In [67]: race_before.columns
```

```
Out[67]: Index(['total', 'visible_minority', 'south_asian', 'chinese', 'black',
                'filipino', 'latin_american', 'arab', 'southeast_asian', 'west_asian',
                'korean', 'japanese', 'other_visible_minority',
                'multiple_visible_minorities', 'not_visible_minority'],
                dtype='object')
```

```

In [68]: # Initialize params
protected = ['visible_minority',
             'south_asian',
             'chinese',
             'black',
             'filipino',
             'latin_american',
             'arab',
             'southeast_asian',
             'west_asian',
             'korean',
             'japanese',
             'other_visible_minoirty',
             'multiple_visible_minorities']

data = []
for col in protected:
    total_before = race_before.total
    protected_before = race_before[col]
    majority_before = race_before.not_visible_minority

    total_after = race_after.total
    protected_after = race_after[col]
    majority_after = race_after.not_visible_minority

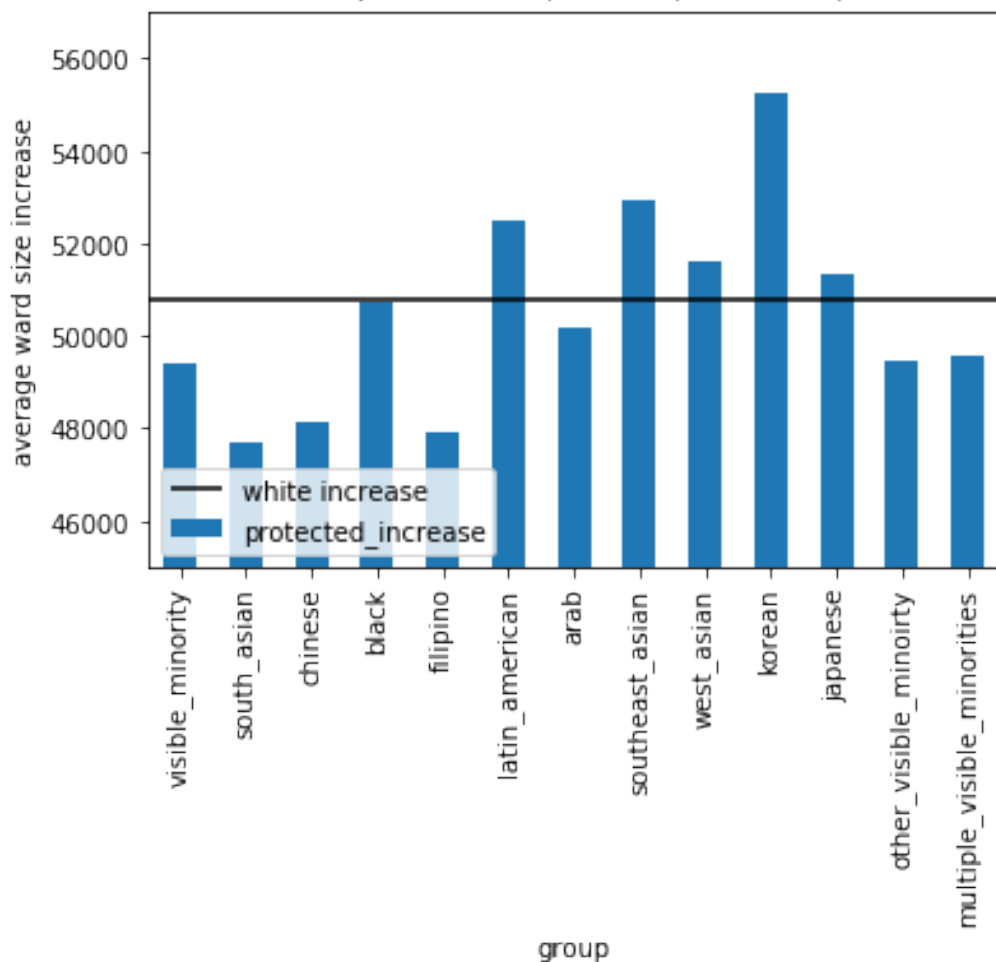
    protected_increase, majority_increase = get_increases(total_before,
                                                         protected_before,
                                                         majority_before,
                                                         total_after,
                                                         protected_after,
                                                         majority_after)

    data.append([col, protected_increase, majority_increase])

In [73]: race_results =pd.DataFrame(data, columns = ['group', 'protected_increase', 'majority_in
race_results['disparity'] = race_results.protected_increase - race_results.majority_in
race_results[['protected_increase','group']].plot(x = 'group', kind = 'bar')
plt.ylabel('average ward size increase')
plt.hlines(race_results.majority_increase[0], xmin = -1, xmax = 15, label = "white in
plt.legend(loc = 3)
plt.ylim(45000,57000)
plt.title("Some groups (S. asians, chinese, filipino) have more representation \n whi
plt.show()

```

Some groups (S. asians, chinese, filipino) have more representation while others (L. american, Korean, SE. Asian) have less



```
In [125]: threshold = .5
for col in protected:
    print("-----")
    total_before = race_before.total
    protected_before = race_before[col]
    prop_protected_before = protected_before/total_before
    print("wards with majority {} before: {}/47={}"
          .format(col,
                  len([p for p in prop_protected_before if p > threshold]),
                  len([p for p in prop_protected_before if p > threshold])/47
                  ))

total_after = race_after.total
protected_after = race_after[col]
```



```

prop_protected_after = protected_after/total_after
print("wards with majority {} after: {}/47={}"
      .format(col,
              len([p for p in prop_protected_after if p > threshold]),
              len([p for p in prop_protected_after if p > threshold])/25
            )
    )

```

```

-----
wards with majority visible_minority before: 24/47=0.5106382978723404
wards with majority visible_minority after: 12/47=0.48
-----

```

```

wards with majority south_asian before: 0/47=0.0
wards with majority south_asian after: 0/47=0.0
-----

```

```

wards with majority chinese before: 1/47=0.02127659574468085
wards with majority chinese after: 0/47=0.0
-----

```

```

wards with majority black before: 0/47=0.0
wards with majority black after: 0/47=0.0
-----

```

```

wards with majority filipino before: 0/47=0.0
wards with majority filipino after: 0/47=0.0
-----

```

```

wards with majority latin_american before: 0/47=0.0
wards with majority latin_american after: 0/47=0.0
-----

```

```

wards with majority arab before: 0/47=0.0
wards with majority arab after: 0/47=0.0
-----

```

```

wards with majority southeast_asian before: 0/47=0.0
wards with majority southeast_asian after: 0/47=0.0
-----

```

```

wards with majority west_asian before: 0/47=0.0
wards with majority west_asian after: 0/47=0.0
-----

```

```

wards with majority korean before: 0/47=0.0
wards with majority korean after: 0/47=0.0
-----

```

```

wards with majority japanese before: 0/47=0.0
wards with majority japanese after: 0/47=0.0
-----

```

```

wards with majority other_visible_minoirty before: 0/47=0.0
wards with majority other_visible_minoirty after: 0/47=0.0
-----

```

```

wards with majority multiple_visible_minorities before: 0/47=0.0
wards with majority multiple_visible_minorities after: 0/47=0.0

```

## 6 Language

```
In [81]: lang_before = pd.read_csv('./Data/Language/language_before.csv', index_col=0).transpose()
        lang_after = pd.read_csv('./Data/Language/language_after.csv', index_col=0).transpose()
```

```
In [82]: lang_before.head()
```

```
Out[82]:
```

	total	english	non_english
Toronto	2691665	2557220	134445
Ward 1	60040	56105	3940
Ward 2	56300	54115	2180
Ward 3	63165	61410	1755
Ward 4	65330	63855	1475

```
In [83]: lang_after.head()
```

```
Out[83]:
```

	total	english	non_english
Toronto	2691665	2557220	134445
Ward 1	116960	110765	6185
Ward 2	116055	113025	3030
Ward 3	127520	124255	3265
Ward 4	106445	103735	2710

```
In [84]: lang_before, lang_after = scale_populations(lang_before, lang_after)
```

```
In [85]: lang_before.head()
```

```
Out[85]:
```

	total	english	non_english
Ward 1	60735.0	56754.449950	3985.607928
Ward 2	57210.0	54989.682948	2215.236234
Ward 3	64065.0	62284.994063	1780.005937
Ward 4	66060.0	64568.518292	1491.481708
Ward 5	49520.0	48145.847786	1384.256274

```
In [86]: # Initialize params
```

```
total_before = lang_before.total
protected_before = lang_before.non_english
majority_before = lang_before.english
```

```
total_after = lang_after.total
protected_after = lang_after.non_english
majority_after = lang_after.english
```

```
protected_increase, majority_increase = get_increases(total_before, protected_before,
```

```
In [87]: print("If you are english speaking, your ward size increased by {} people on average"
          print("If you are not english speaking, your ward size increased by {} people on average"
          print("This is a disparity of {} people on average".format(protected_increase-majority_increa
```

If you are english speaking, your ward size increased by 47397 people on average  
 If you are not english speaking, your ward size increased by 50187 people on average  
 This is a disparity of -2790 people on average

```
In [90]: # lost majorities
prop_protected_before = protected_before / total_before
print("Wards with majority not english speaking before: {}/47".format(len([p for p in

prop_protected_after = protected_after / total_after
print("Wards with majority not english speaking before: {}/25".format(len([p for p in
```

Wards with majority not english speaking before: 0/47  
 Wards with majority not english speaking before: 0/25

## 7 First generation immigrants

```
In [105]: immigrant_before = pd.read_csv('./Data/Immigration/immigration_before.csv', index_col=0)
immigrant_after = pd.read_csv('./Data/Immigration/immigration_after.csv', index_col=0)
```

```
In [106]: immigrant_before.head()
```

```
Out[106]:
```

	total	first_gen	second_gen_plus
Toronto	2691665	1377465	1314205
Ward 1	60040	39450	20585
Ward 2	56295	30490	25810
Ward 3	63165	28290	34870
Ward 4	65335	26895	38440

```
In [107]: immigrant_after.head()
```

```
Out[107]:
```

	total	first_gen	second_gen_plus
Toronto	2691665	1377465	1314205
Ward 1	116960	71535	45425
Ward 2	116055	50030	66025
Ward 3	127525	51980	75545
Ward 4	106445	37785	68650

```
In [111]: immigrant_before, immigrant_after = scale_populations(immigrant_before, immigrant_after)
```

```
In [112]: immigrant_before.head()
```

```
Out[112]:
```

	total	first_gen	second_gen_plus
Ward 1	60735.0	39906.658061	20823.284061
Ward 2	57210.0	30985.574207	26229.507061
Ward 3	64065.0	28693.087153	35366.841605
Ward 4	66060.0	27193.444555	38866.555445
Ward 5	49520.0	21839.926546	27685.125485

```
In [113]: # Initialize params
total_before = immigrant_before.total
protected_before = immigrant_before.first_gen
majority_before = immigrant_before.second_gen_plus

total_after = immigrant_after.total
protected_after = immigrant_after.first_gen
majority_after = immigrant_after.second_gen_plus

protected_increase, majority_increase = get_increases(total_before, protected_before, majority_before, total_after, protected_after, majority_after)

In [116]: print("If you are first gen canadian, your ward size increased by {} people on average".format(protected_increase))
print("If you are not first gen canadian, your ward size increased by {} people on average".format(majority_increase))
print("This is a disparity of {} people on average".format(protected_increase-majority_increase))
```

If you are first gen canadian, your ward size increased by 49799 people on average  
If you are not first gen canadian, your ward size increased by 50308 people on average  
This is a disparity of -509 people on average

```
In [122]: # lost majorities
prop_protected_before = protected_before / total_before
print("Wards with majority first gen canadian before: {}/47={}"
      .format(len([p for p in prop_protected_before if p > threshold]),
              len([p for p in prop_protected_before if p > threshold])/47))

prop_protected_after = protected_after / total_after
print("Wards with majority first gen canadian after: {}/25={}"
      .format(len([p for p in prop_protected_after if p > threshold]),
              len([p for p in prop_protected_after if p > threshold])/25))
```

Wards with majority first gen canadian before: 27/47=0.574468085106383  
Wards with majority first gen canadian after: 12/25=0.48

I think this is pretty significant

## 8 Education

```
In [127]: education_before = pd.read_csv('./Data/Education/education_before.csv', index_col=0)
education_after = pd.read_csv('./Data/Education/education_after.csv', index_col=0).transpose()

In [128]: education_before.head()
```

```
Out[128]:
```

	total	post_secondary	no_post_secondary
Toronto	2294785	1356355	938435
Ward 1	48650	20630	28020
Ward 2	46555	21875	24680
Ward 3	53825	32465	21365
Ward 4	55085	33430	21660

```
In [129]: education_after.head()
```

```
Out[129]:
```

	total	post_secondary	no_post_secondary
Toronto	2294790	1356355	938430
Ward 1	95295	42350	52940
Ward 2	98790	58480	40305
Ward 3	110720	71165	39555
Ward 4	90620	62430	28195

```
In [130]: education_before, education_after = scale_populations(education_before, education_after)
```

```
In [131]: education_before.head()
```

```
Out[131]:
```

	total	post_secondary	no_post_secondary
Ward 1	60735.0	25754.636177	34980.363823
Ward 2	57210.0	26881.511116	30328.488884
Ward 3	64065.0	38641.341849	25429.609382
Ward 4	66060.0	40090.511028	25975.485159
Ward 5	49520.0	32661.128497	16864.741584

```
In [132]: # Initialize params
```

```
total_before = education_before.total
```

```
protected_before = education_before.no_post_secondary
```

```
majority_before = education_before.post_secondary
```

```
total_after = education_after.total
```

```
protected_after = education_after.no_post_secondary
```

```
majority_after = education_after.post_secondary
```

```
protected_increase, majority_increase = get_increases(total_before, protected_before, total_after, protected_after, majority_after)
```

```
In [133]: print("If you don't have a post-secondary degree, your ward size increased by {} people on average".format(protected_increase))
print("If you have a post-secondary degree, your ward size increased by {} people on average".format(majority_increase))
print("This is a disparity of {} people on average".format(protected_increase-majority_increase))
```

If you don't have a post-secondary degree, your ward size increased by 49261 people on average

If you have a post-secondary degree, your ward size increased by 50598 people on average

This is a disparity of -1337 people on average

```
In [135]: # lost majorities
```

```
prop_protected_before = protected_before / total_before
```

```
print("Wards with majority no post-secondary degree before: {}/47={}"
```

```
.format(len([p for p in prop_protected_before if p > threshold]),
```

```
len([p for p in prop_protected_before if p > threshold])/47))
```

```
prop_protected_after = protected_after / total_after
```

```
print("Wards with majority no post-secondary degree after: {}/25={}"
```

```
.format(len([p for p in prop_protected_after if p > threshold]),
```

```
len([p for p in prop_protected_after if p > threshold])/25))
```

Wards with majority no post-secondary degree before: 13/47=0.2765957446808511  
Wards with majority no post-secondary degree after: 5/25=0.2

## 9 Renter vs. Home Owner

```
In [139]: housing_before = pd.read_csv('./Data/Housing/home_before.csv', index_col=0).transpose()  
housing_after = pd.read_csv('./Data/Housing/home_after.csv', index_col=0).transpose()
```

```
In [140]: housing_before.head()
```

```
Out[140]:
```

	total	own	rent
Toronto	1112905	587080	525825
Ward 1	18305	9925	8385
Ward 2	19465	11705	7765
Ward 3	24530	17615	6915
Ward 4	25675	15700	9975

```
In [141]: housing_after.head()
```

```
Out[141]:
```

	total	own	rent
Toronto	1112905	587080	525825
Ward 1	37895	20945	16945
Ward 2	45045	30175	14870
Ward 3	59730	33860	25870
Ward 4	50315	21375	28945

```
In [142]: housing_before, housing_after = scale_populations(housing_before, housing_after)
```

```
In [143]: housing_before.head()
```

```
Out[143]:
```

	total	own	rent
Ward 1	60735.0	32930.613220	27820.976509
Ward 2	57210.0	34402.417159	22822.278448
Ward 3	64065.0	46005.094782	18059.905218
Ward 4	66060.0	40395.014606	25664.985394
Ward 5	49520.0	28338.746890	21192.454196

```
In [144]: # Initialize params  
total_before = housing_before.total  
protected_before = housing_before.rent  
majority_before = housing_before.own  
  
total_after = housing_after.total  
protected_after = housing_after.rent  
majority_after = housing_after.own  
  
protected_increase, majority_increase = get_increases(total_before, protected_before,
```

```
In [145]: print("If you rent, your ward size increased by {} people on average".format(protected_increase-renter_ward_size))
          print("If you own your home, your ward size increased by {} people on average".format(protected_increase-homeowner_ward_size))
          print("This is a disparity of {} people on average".format(protected_increase-majority_ward_size))
```

If you rent, your ward size increased by 50615 people on average

If you own your home, your ward size increased by 49567 people on average

This is a disparity of 1048 people on average

```
In [146]: # lost majorities
          prop_protected_before = protected_before / total_before
          print("Wards with majority renters before: {}/47={}"
                .format(len([p for p in prop_protected_before if p > threshold]),
                        len([p for p in prop_protected_before if p > threshold])/47))

          prop_protected_after = protected_after / total_after
          print("Wards with majority renters after: {}/25={}"
                .format(len([p for p in prop_protected_after if p > threshold]),
                        len([p for p in prop_protected_after if p > threshold])/25))
```

Wards with majority renters before: 17/47=0.3617021276595745

Wards with majority renters after: 9/25=0.36