

Elections Analysis 2015

By Mahnoor

Executive Summary

Overview: Analyze the election data for the 2015 Federal Elections. Take data given by Elections Canada and create a database to organize the data to make it easier to draw conclusions out of.

Steps:

1. Acquire raw data
2. Parse the raw data using Python and output into csv files
3. Create database and tables to import data into MySQL
4. Use SQL to look into data further and draw conclusions

Some questions that were answered included: What was the greatest margin that a candidate won by, how does the rejected vote count correlate with the margin of votes for each electoral districts and how many incumbent candidates re-ran in the election. For instance, it was seen that only 29 incumbent candidates re-ran for Liberals meaning most of the party was fairly new. For more analysis look at the later slides.

Raw Input Data

Elections Canada Link Format 2:

<http://www.elections.ca/content.aspx?section=res&dir=rep/off/42gedata&document=bypro&lang=e>

The other format of data was format 1 which only had 1 row for each polling station in each electoral district. This would have made it harder to analyze the results of each candidate as it had less detail. For example format 1 did not have the incumbent indicator or the electoral indicator which was needed to answer questions regarding incumbent candidates. Format 2 was chosen because it had more data that was needed for this analysis and in a more efficient format.

Sample Input Data pt 1 from Elections Canada

	A	B	C	D	E	F	G	H	I	J
1	Electoral District Number	Electoral District Name_English	Electoral District Name	Polling Station Number	Polling Station Name	Void Poll	No Poll	Merge With	Rejected Ballots	Electors
2	10001	Avalon	Avalon	1	Freshwater	N	N		0	133
3	10001	Avalon	Avalon	1	Freshwater	N	N		0	133
4	10001	Avalon	Avalon	1	Freshwater	N	N		0	133
5	10001	Avalon	Avalon	1	Freshwater	N	N		0	133
6	10001	Avalon	Avalon	1	Freshwater	N	N		0	133
7	10001	Avalon	Avalon	1	Freshwater	N	N		0	133
8	10001	Avalon	Avalon	2	Salmon Cove--Perry's	Y	N		0	0
9	10001	Avalon	Avalon	2	Salmon Cove--Perry's	Y	N		0	0
10	10001	Avalon	Avalon	2	Salmon Cove--Perry's	Y	N		0	0
11	10001	Avalon	Avalon	2	Salmon Cove--Perry's	Y	N		0	0
12	10001	Avalon	Avalon	2	Salmon Cove--Perry's	Y	N		0	0
13	10001	Avalon	Avalon	2	Salmon Cove--Perry's	Y	N		0	0
14	10001	Avalon	Avalon	3	Victoria	N	N		1	337
15	10001	Avalon	Avalon	3	Victoria	N	N		1	337
16	10001	Avalon	Avalon	3	Victoria	N	N		1	337
17	10001	Avalon	Avalon	3	Victoria	N	N		1	337
18	10001	Avalon	Avalon	3	Victoria	N	N		1	337
19	10001	Avalon	Avalon	3	Victoria	N	N		1	337
20	10001	Avalon	Avalon	4	Victoria	N	N		4	444
21	10001	Avalon	Avalon	4	Victoria	N	N		4	444
22	10001	Avalon	Avalon	4	Victoria	N	N		4	444
23	10001	Avalon	Avalon	4	Victoria	N	N		4	444
24	10001	Avalon	Avalon	4	Victoria	N	N		4	444
25	10001	Avalon	Avalon	4	Victoria	N	N		4	444
26	10001	Avalon	Avalon	5	Victoria	N	N		0	459
27	10001	Avalon	Avalon	5	Victoria	N	N		0	459
28	10001	Avalon	Avalon	5	Victoria	N	N		0	459
29	10001	Avalon	Avalon	5	Victoria	N	N		0	459
30	10001	Avalon	Avalon	5	Victoria	N	N		0	459
31	10001	Avalon	Avalon	5	Victoria	N	N		0	459
32	10001	Avalon	Avalon	6	Victoria	N	N		0	269
33	10001	Avalon	Avalon	6	Victoria	N	N		0	269
34	10001	Avalon	Avalon	6	Victoria	N	N		0	269
35	10001	Avalon	Avalon	6	Victoria	N	N		0	269
36	10001	Avalon	Avalon	6	Victoria	N	N		0	269
37	10001	Avalon	Avalon	6	Victoria	N	N		0	269
38	10001	Avalon	Avalon	7	Carbonear	N	N		2	343

Sample Input Data pt 2 from Elections Canada

K	L	M	N	O	P	Q	R
Candidate's Family	Candidate's Middle	Candidate's First	Political Affiliation Name	Political Affilia	Incumbent	Elected	Candidate Poll
Andrews		Scott	No Affiliation	Aucune appart	Y	N	39
Baldwin		Jeannie	NDP-New Democratic Part	NPD-Nouveau P	N	N	9
Barnett		Lorraine E.	Conservative	Conservateur	N	N	4
Byrne-Puumala		Krista	Green Party	Parti Vert	N	N	0
McCreath		Jennifer	Forces et Démocratie - Ali	Forces et Dém	N	N	0
McDonald		Ken	Liberal	Libéral	N	Y	55
Andrews		Scott	No Affiliation	Aucune appart	Y	N	0
Baldwin		Jeannie	NDP-New Democratic Part	NPD-Nouveau P	N	N	0
Barnett		Lorraine E.	Conservative	Conservateur	N	N	0
Byrne-Puumala		Krista	Green Party	Parti Vert	N	N	0
McCreath		Jennifer	Forces et Démocratie - Ali	Forces et Dém	N	N	0
McDonald		Ken	Liberal	Libéral	N	Y	0
Andrews		Scott	No Affiliation	Aucune appart	Y	N	71
Baldwin		Jeannie	NDP-New Democratic Part	NPD-Nouveau P	N	N	16
Barnett		Lorraine E.	Conservative	Conservateur	N	N	17
Byrne-Puumala		Krista	Green Party	Parti Vert	N	N	1
McCreath		Jennifer	Forces et Démocratie - Ali	Forces et Dém	N	N	1
McDonald		Ken	Liberal	Libéral	N	Y	86
Andrews		Scott	No Affiliation	Aucune appart	Y	N	59
Baldwin		Jeannie	NDP-New Democratic Part	NPD-Nouveau P	N	N	29
Barnett		Lorraine E.	Conservative	Conservateur	N	N	38
Byrne-Puumala		Krista	Green Party	Parti Vert	N	N	1
McCreath		Jennifer	Forces et Démocratie - Ali	Forces et Dém	N	N	0
McDonald		Ken	Liberal	Libéral	N	Y	87
Andrews		Scott	No Affiliation	Aucune appart	Y	N	61
Baldwin		Jeannie	NDP-New Democratic Part	NPD-Nouveau P	N	N	22
Barnett		Lorraine E.	Conservative	Conservateur	N	N	14
Byrne-Puumala		Krista	Green Party	Parti Vert	N	N	0
McCreath		Jennifer	Forces et Démocratie - Ali	Forces et Dém	N	N	1
McDonald		Ken	Liberal	Libéral	N	Y	85
Andrews		Scott	No Affiliation	Aucune appart	Y	N	59
Baldwin		Jeannie	NDP-New Democratic Part	NPD-Nouveau P	N	N	9
Barnett		Lorraine E.	Conservative	Conservateur	N	N	13
Byrne-Puumala		Krista	Green Party	Parti Vert	N	N	0
McCreath		Jennifer	Forces et Démocratie - Ali	Forces et Dém	N	N	0
McDonald		Ken	Liberal	Libéral	N	Y	45
Andrews		Scott	No Affiliation	Aucune appart	Y	N	45

Parsing Raw Input Data into Output CSV Files

Using Python

```
import os
import csv
import datetime
```

```
politicalPartyMap = {}
candidateMap = {}
pollingStationMap = {}
```

```
def addPollingStation(ed_code, ps_num, pollingStationName, rejectedBallots,
electors, voidPollIndicator, noPollHeldIndicator):
```

```
    ed_ps_key = ed_code, ps_num
```

```
    if ed_ps_key in pollingStationMap:
        return pollingStationMap[ed_code, ps_num]
```

```
    else:
```

```
        pollingStationMap[ed_code, ps_num] = ed_ps_key
```

```
        with open("C:/pyspace/OUTPollingStationOUT.csv", 'a', newline="") as
```

```
politicalPartyOut:
```

```
            politicalPartyFile = csv.writer(politicalPartyOut, delimiter=";", quotechar='')
```

```
            writetofile = [[ed_code, ps_num, pollingStationName, rejectedBallots,
```

```
electors, voidPollIndicator, noPollHeldIndicator]]
```

```
            politicalPartyFile.writerows(writetofile)
```

```
        return ed_ps_key
```

```
# end def addPollingStation
```

```
for root, dirs, files in os.walk(directory):
```

```
    for file in files:
```

```
        if file.endswith(".csv"):
```

```
            # print("This is a csv file", file)
```

```
            f = open(directory+"/"+file, 'r')
```

```
            g= open(directory+"/"+file)
```

```
            startTime = datetime.datetime.now()
```

```
            print("Processing ", file, "...")
```

```
            rowsProcessed = 0
```

```
with open(directory+"/"+file) as csvfile:
    inputFile = csv.reader(csvfile, delimiter=',')
    next(inputFile)
```

```
for row in inputFile:
```

```
    rowsProcessed += 1
```

```
    ed_code = row[0]
```

```
    ps_num = row[3]
```

```
    pollingStationName = row[4]
```

```
addPollingStation(ed_code, ps_num, pollingStationName, rejectedBallots,
electors, voidPollIndicator, noPollHeldIndicator)
```

```
if candid > prevCandid:
```

```
    prevCandid = candid
```

```
endTime = datetime.datetime.now()
```

```
print("Processed ", rowsProcessed, "rows, time taken: ", endTime -
startTime)
```

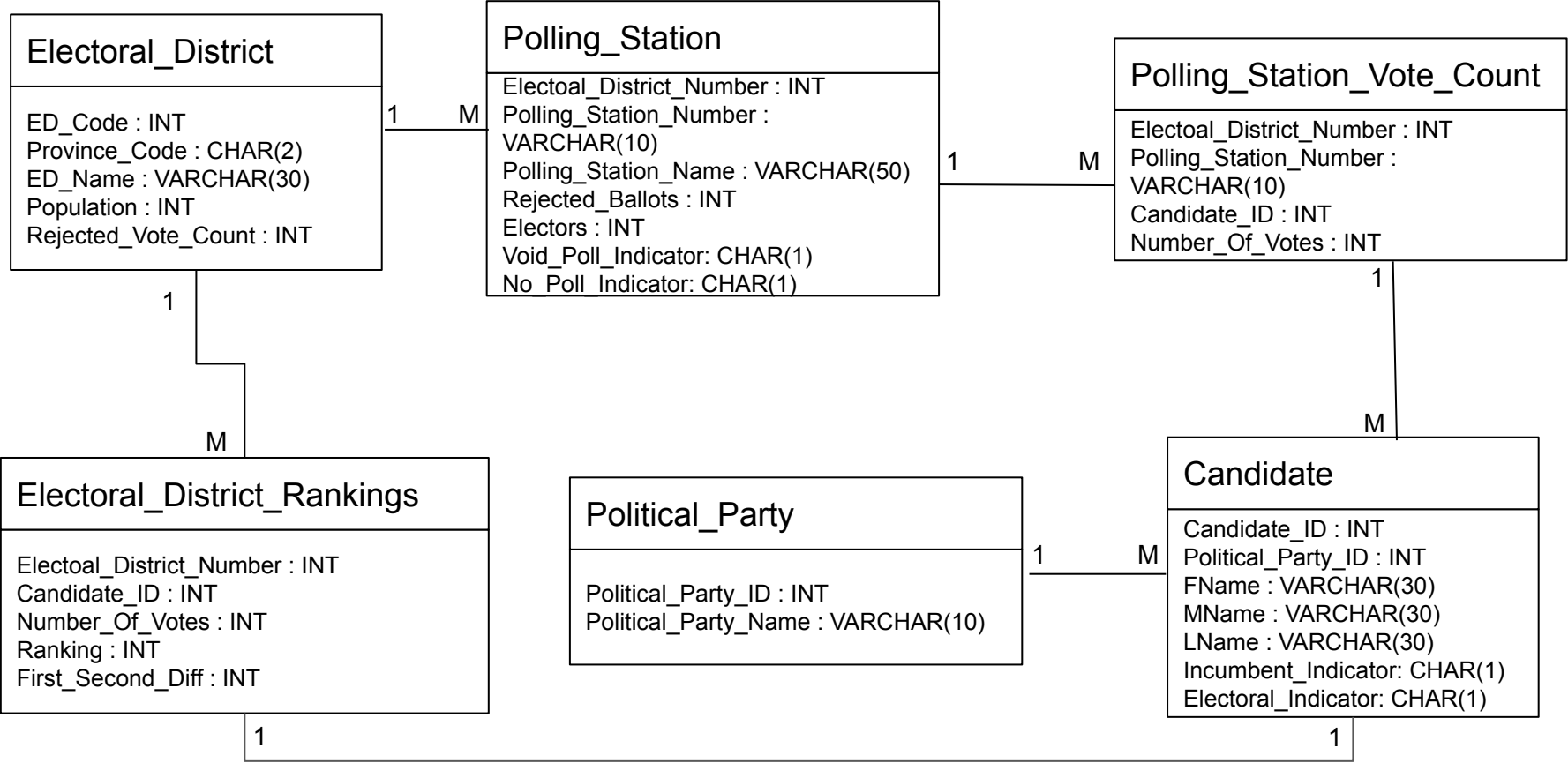
```
# end for row
```

```
#end for file
```

```
# end for root, dirs, files
```

This python code is defining a function which will take care of making seeing if the Electoral District code and Polling Station number is already seen in the file, if it is not, they are added to the dictionary. It opens an output csv file and puts that specific row with the new Electoral District code and Polling Station number into the file. It later opens the folder with input files and then each specific file. It takes values from the file and puts it into variables and then the functions are called to add new values into the output files.

Data Model



Electoral District Output File

	A	B	C	D
1	ED CODE	PROVINCE_CODE	ED_NAMEEE	POPULATION
2	10001	NL	Avalon	81540
3	10002	NL	Bonavista--Burin--Trinity	76704
4	10003	NL	Coast of Bays--Central--Notre Dame	78092
5	10004	NL	Labrador	26728
6	10005	NL	Long Range Mountains	87592
7	10006	NL	St. John's East	81936
8	10007	NL	St. John's South--Mount Pearl	81944
9	11001	PE	Cardigan	36005
10	11002	PE	Charlottetown	34562
11	11003	PE	Egmont	34598
12	11004	PE	Malpeque	35039
13	12001	NS	Cape Breton--Canso	75247
14	12002	NS	Central Nova	74597
15	12003	NS	Cumberland--Colchester	82321
16	12004	NS	Dartmouth--Cole Harbour	91212
17	12005	NS	Halifax	92643
18	12006	NS	Halifax West	87275
19	12007	NS	Kings--Hants	83306
20	12008	NS	Sackville--Preston--Chezzetcook	85583
21	12009	NS	South Shore--St. Margarets	92561
22	12010	NS	Sydney--Victoria	73328
23	12011	NS	West Nova	83654
24	13001	NB	Acadie--Bathurst	79340
25	13002	NB	Beauséjour	80416

Polling Station Output File

	A	B	C	D	E	F	G
1	10001	1	Freshwater	0	133	N	N
2	10001	2	Salmon Cove--Perry's Cove	0	0	Y	N
3	10001	3	Victoria	1	337	N	N
4	10001	4	Victoria	4	444	N	N
5	10001	5	Victoria	0	459	N	N
6	10001	6	Victoria	0	269	N	N
7	10001	7	Carbonear	2	343	N	N
8	10001	8	Carbonear	0	426	N	N
9	10001	9	Carbonear	0	381	N	N
10	10001	10	Carbonear	3	365	N	N
11	10001	11	Carbonear	1	306	N	N
12	10001	12	Carbonear	0	421	N	N
13	10001	13	Harbour Grace--Bristol's Hope	0	481	N	N
14	10001	14	Harbour Grace	1	417	N	N
15	10001	15	Harbour Grace	0	210	N	N
16	10001	16	Carbonear	0	339	N	N
17	10001	17	Carbonear	0	366	N	N
18	10001	18	Carbonear	0	299	N	N
19	10001	19	Carbonear	0	320	N	N
20	10001	20	Carbonear	1	373	N	N
21	10001	21	Bryant's Cove	0	280	N	N
22	10001	22	Upper Island Cove	0	489	N	N
23	10001	23	Upper Island Cove	3	289	N	N
24	10001	24	Upper Island Cove	0	234	N	N
25	10001	25	Upper Island Cove	0	351	N	N

Polling Station Vote Count Output File

	A	B	C	D
1	10001	1	1	39
2	10001	1	2	9
3	10001	1	3	4
4	10001	1	4	0
5	10001	1	5	0
6	10001	1	6	55
7	10001	2	1	0
8	10001	2	2	0
9	10001	2	3	0
10	10001	2	4	0
11	10001	2	5	0
12	10001	2	6	0
13	10001	3	1	71
14	10001	3	2	16
15	10001	3	3	17
16	10001	3	4	1
17	10001	3	5	1
18	10001	3	6	86
19	10001	4	1	59
20	10001	4	2	29
21	10001	4	3	38
22	10001	4	4	1
23	10001	4	5	0
24	10001	4	6	87
25	10001	5	1	61

Candidate Output File

	A	B	C	D	E	F	G
1	1	1	Scott	Andrews	Y	N	
2	2	2	Jeannie	Baldwin	N	N	
3	3	3	Lorraine E.	Barnett	N	N	
4	4	4	Krista	Byrne-Puumala	N	N	
5	5	5	Jennifer	McCreath	N	N	
6	6	6	Ken	McDonald	N	Y	
7	7	2	Jenn	Brown	N	N	
8	8	4	Tyler John	Colbourne	N	N	
9	9	6	Judy M.	Foote	Y	Y	
10	10	3	Mike	Windsor	N	N	
11	11	2	Claudette	Menchenton	N	N	
12	12	3	Kevin George	O'Brien	N	N	
13	13	4	Elizabeth	Perry	N	N	
14	14	6	Scott	Simms	Y	Y	
15	15	6	Yvonne	Jones	Y	Y	
16	16	3	Peter	Penashue	N	N	
17	17	2	Edward	Rudkowski	N	N	
18	18	2	Devon	Babstock	N	N	
19	19	4	Terry	Cormier	N	N	
20	20	6	Gudie	Hutchings	N	Y	
21	21	3	Wayne	Ruth	N	N	
22	22	7	Sean	Burton	N	N	
23	23	2	Jack	Harris	Y	N	
24	24	4	David Anthony	Peters	N	N	
25	25	3	Deanne	Stapleton	N	N	

Political Party Output File

	A	B
1	1	No Affiliation
2	2	NDP-New Democratic Party
3	3	Conservative
4	4	Green Party
5	5	Forces et Démocratie - Allier les forces de nos régions
6	6	Liberal
7	7	Communist
8	8	Christian Heritage Party
9	9	Independent
10	10	Marxist-Leninist
11	11	Rhinoceros
12	12	Libertarian
13	13	Bloc Québécois
14	14	ATN
15	15	United Party
16	16	Animal Alliance/Environment Voters
17	17	CAP
18	18	PC Party
19	19	Radical Marijuana
20	20	Pirate
21	21	PACT
22	22	Seniors Party
23	23	The Bridge
24	24	Canada Party
25	25	Democratic Advancement

Creating Election Database

Using MySQL

Sample SQL Code

create table Electoral_District (ED_Code INT, Province_Code char(2), ED_Name varchar(30), Population INT);

load data infile "/home/user/Documents/ElectoralDistrictInfo.csv"

-> **into** table Electoral_District

-> Columns terminated by ","

-> Enclosed by ""

-> lines terminated by "\n"

-> **ignore** 1 rows;

```
mysql> show tables;
```

Tables_in_elections
Candidate
Electoral_District
Electoral_District_Rankings
Political_Party
Polling_Station
Polling_Station_Vote_Count

Adding to the Databases

Creating new columns and tables

Rejected Ballots Added to Electoral District Table

The sum of all the rejected ballots at each polling station in every electoral district was put into the Electoral District table to show the rejected count at the Electoral District level.

```
alter table Electoral_District Add Rejected_Count INT
```

```
update Electoral_District
```

```
set Rejected_Count =
```

```
(select sum(Polling_Station.Rejected_Ballots)
```

```
from Polling_Station
```

```
where Electoral_District.ED_Code = Polling_Station.Electoral_District_Number);
```

```
mysql> select * from Electoral_District;
```

ED_Code	Province_Code	ED_Name	Population	Rejected_Count
10001	NL	Avalon	81540	162
10002	NL	Bonavista--Burin--Trinity	76704	173
10003	NL	Coast of Bays--Central--Notre	78092	145
10004	NL	Labrador	26728	53

New Table: Electoral_District_Rankings

A new table called Electoral_District_Rankings was created to rank candidates in each Electoral District

create table Electoral_District_Rankings (Electoral_District_Number INT, Candidate_ID INT, Number_Of_Votes INT, Ranking INT, First_Second_Diff INT);

load data local infile "/home/zahid/ED_Rank_Input_2.csv" **into** table Electoral_District_Rankings **fields** terminated by "," **enclosed** by "" **lines** terminated by "\n" **ignore** 1 rows;

select Electoral_District_Number, Candidate_ID, Number_Of_Votes

from Electoral_District_Rankings

order by Electoral_District_Number, Number_Of_Votes **desc**;

	A	B	C	D	E
1	Electoral_District_Number	Candidate_ID	Number_Of_Votes	Rankings	1st and 2nd DIFF
2	10001	6	23528	1	16027
3	10001	1	7501	2	0
4	10001	2	6075	3	0
5	10001	3	4670	4	0
6	10001	4	228	5	0
7	10001	5	84	6	0
8	10002	9	28704	1	25170

To get Rankings column:
=IF(C3<C2, D2+1)

To get 1st and 2nd DIFF column:
=IF(D2 = 1, C2 - C3)

Analysis Questions

Brainstorming Questions

- What percentage of each province voted for independent parties?
- Ratio of total votes to electors in electoral district
- In each province, what was the percentage difference between the winning party win than the runner up?
- Who won by the biggest margin in the whole country?
- Number of seats won by each party
- Number of seats won by each party in each province
- Number of candidates in each party
- How many rejected votes were there in each Electoral District?
- How many incumbents won from each party

Seats Won by Political Parties

Political_Party_Name	Seats
Liberal	184
Conservative	99
NDP	44
Bloc Qu'Appelle	10
Green Party	1

Query 1

Total Candidates in Each Party

Political_Party_Name	Total_Candidates
Conservative	338
Liberal	338
Green Party	336
NDP	336
Bloc Qu'Appelle	78
Independent	73
Libertarian	72
Marxist-Leninist	70
Christian Heritage Party	30
Rhinoceros	27

- The Conservative and Liberal party both have candidates from each and every electoral district in the country, this shows the size of the parties telling us that they are tied as the biggest parties
- The Green Party and NDP both have 336 candidates which is 2 less than the total amount of electoral districts. A follow up question could be seeing **which electoral districts do not have candidates from these parties.**
- Looking back at slide 19, the table shows the amount of seats won by each party. Even though the Conservative and Liberal parties have the same number of candidates, there was a difference of 85 seats between them. Also with Green Party and NDP. Green Party lost against NDP by 43 seats.
- In general, looking at all the top four parties, they all have a similar amount of candidates, yet Liberal won 184 seats vs Green party who won only 1 seat
- The next four parties from Bloc Q - Marxist have between 70-78 candidates. This is around 260 seats less than the top four parties. This may seem like a huge difference in the amount of seats, yet looking at the table in slide 19, Bloc Q still won 9 more seats than Green Party.
- This shows that number of candidates in the Green Party does not have a positive correlation with the amount of seats won by the party
- The last two parties are significantly smaller than the total amount of electoral districts and looking at the table in slide 19, none of the rest of the parties won any seats. One could say it is because the size of the party is too small, but looking at the results for Green Party, one could also argue that that is not the case and that the number of candidates is not a determining factor

Total Incumbent Candidates who Won

Political_Party_Name	Total_Incumbents_Won
Conservative	63
Liberal	29
NDP	28
Bloc QuÃ©bÃ©cois	1
Green Party	1

Query 3

- Looking at this data, we can see that for Green Party, the one seat that the party won was from an incumbent candidate, meaning that area is a strong supporter of the Green Party. A follow up question could be which area is this electoral district in which Green Party won from an incumbent candidate.
- In the table on slide 20, the Liberal party had 338 candidates and according to this table, Liberals only have 29 incumbent candidates that won in the election. This could mean that the Liberal party was fairly new and brought a lot of new candidates. This was also proven in slide 22 where it shows that the total amount of incumbent candidates for Liberal was also 29 meaning there only were 29 incumbent candidates. This tells us two things, firstly that all the incumbent ran from Liberals won and secondly that the Liberal party had $(338-29 =)$ 309 new candidates. A follow up question could be to look at data from previous elections and seeing if that is a trend or strategy that the Liberal party uses.
- On the other hand, the NDP had only 28 incumbent candidates that won. The total candidates were 336 and the total incumbent candidates was 82 as seen in the table in slide 22. This means that $82/336$ were incumbent and $28/82$ incumbents won. This means that the party is also fairly new at 254 new candidates, not as much as Liberal but they still have a lot of new candidates

Total Incumbent Candidates from each Party

Political_Party_Name	Total_Incumbents
Conservative	127
NDP-New Democratic Party	82
Liberal	29
Green Party	4
Forces et D�mocratie - A	2
Bloc Qu�b�cois	1
No Affiliation	1
Independent	1

- The Conservative party has 127/338 candidates that are incumbent which is a lot more than Liberal and NDP. Out of the 127, 63 incumbent candidates won. The Conservative Party had the most largest amount of incumbent candidates and largest amount of incumbent candidates who won. A little more than a third of their total candidates were incumbent
- Bloc Q only had 1 incumbent candidate and that one candidate also won as shown in slide 21.
- The rest of the incumbent candidates from the other parties all ran, but none of them won

Ratio of Total Votes Casted to Electors

Political_Party_Name	Total_Incumbents
Conservative	127
NDP-New Democratic Party	82
Liberal	29
Green Party	4
Forces et D�mocratie - A	2
Bloc Qu�b�cois	1
No Affiliation	1
Independent	1

•

Query 4

Appendix

Number	Query Text
1	<pre>select p.Political_Party_Name, count(r.Ranking) as Seats from Political_Party p, Electoral_District_Rankings r, Candidate c where p.Political_Party_ID = c.Political_Party_ID and r.Candidate_ID = c.Candidate_ID and r.Ranking = 1 group by p.Political_Party_Name order by count(r.Ranking) desc;</pre>
2	<pre>select p.Political_Party_Name, count(c.Candidate_ID) as Total_Candidates from Political_Party p, Candidate c where p.Political_Party_ID = c.Political_Party_ID group by p.Political_Party_Name order by count(c.Candidate_ID) desc;</pre>
3	<pre>select p.Political_Party_Name, count(c.Incumbent_Indicator) as Total_Incumbents_Won from Political_Party p, Candidate c where p.Political_Party_ID = c.Political_Party_ID and c.Incumbent_Indicator = "Y" and c.Electoral_Indicator = "Y" group by p.Political_Party_Name order by count(c.Incumbent_Indicator) desc;</pre>
4	<pre>select p.Political_Party_Name, count(c.Incumbent_Indicator) as Total_Incumbents from Political_Party p, Candidate c where p.Political_Party_ID = c.Political_Party_ID and c.Incumbent_Indicator = "Y" group by p.Political_Party_Name order by count(c.Incumbent_Indicator) desc;</pre>

Number	Query Text
5	
6	
7	
8	