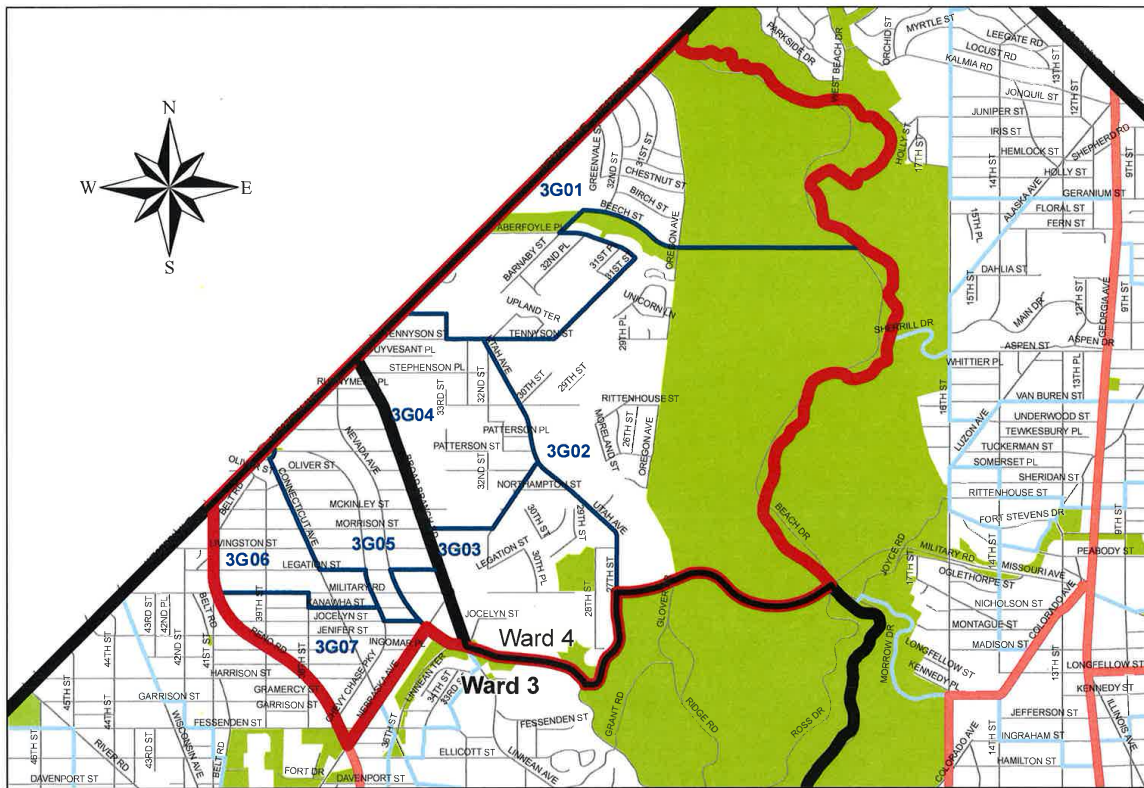# Collapsing ANC Election Data to Contest Level

Where do we see ward discrepancies?

- ward_check comes from the full SMD designation ("2B01" e.g.)
- ward is from the initial data

```
##   ward anc smd ward_check
## 1    2   D  04          6
## 2    4   G  01          3
## 3    4   G  02          3
## 4    4   G  03          3
## 5    4   G  04          3
```


ANC 3G 2013 Boundaries

- we see ANC 3G SMDs 1-4 are in ward 4, which agrees with maps
- also seeing 6D04... which is accurate but ignorable – looks like it includes a section of hain's point that's in ward 2.

```
##   contest_name votes
## 1         3G01  4125
## 2         3G02  3888
## 3         3G03  3429
```

```
## 4        3G04  4118
## 5        6D04     3
```

this is seeing the vacuous part of 6D04, cool

```
data %<>% filter(!(contest_name=="6D04" & ward==2))
```

Recode all of ANC 3G to Ward 3

- We currently don't work with any ward-level data
- And this is the naming convention used by shapefules in impute_turnout.R

```
data %<>% mutate(ward=ifelse(contest_name %in% c("3G01", "3G02", "3G03", "3G04"), 3, ward))
```

Peek at Candidate-Level Data

```
data.cand %>% ungroup %>% skim_without_charts
```

Table 1: Data summary

Name	Piped data
Number of rows	2797
Number of columns	9
Column type frequency:	
character	5
numeric	4
Group variables	None

Variable type: character

skim_variable	n_missing	complete_rate	min	max	empty	n_unique	whitespace
contest_name	0	1	4	4	0	296	0
candidate	0	1	7	30	0	1252	0
year	0	1	4	4	0	4	0
anc	0	1	1	1	0	7	0
smd	0	1	2	2	0	12	0

Variable type: numeric

skim_variable	n_missing	complete_rate	mean	sd	p0	p25	p50	p75	p100
votes	0	1.00	276.48	294.73	0	24	174	473	2161
ward	0	1.00	4.49	2.26	1	3	5	6	8
over_votes	681	0.76	0.29	0.58	0	0	0	0	3
under_votes	681	0.76	165.13	120.67	0	96	139	200	1008

Peek at contest-level data

```
data.cont %>% ungroup %>% skim_without_charts
```

Table 4: Data summary

Name	Piped data
Number of rows	1184
Number of columns	13
Column type frequency:	
character	5
numeric	8
Group variables	None

Variable type: character

skim_variable	n_missing	complete_rate	min	max	empty	n_unique	whitespace
contest_name	0	1	4	4	0	296	0
year	0	1	4	4	0	4	0
anc	0	1	1	1	0	7	0
smd	0	1	2	2	0	12	0
winner	0	1	7	30	0	864	0

Variable type: numeric

skim_variable	n_missing	complete_rate	mean	sd	p0	p25	p50	p75	p100
smd_anc_votes	0	1.00	653.15	298.03	2	458.75	665.5	855.0	2216
explicit_candidates	0	1.00	1.36	0.65	0	1.00	1.0	2.0	4
over_votes	296	0.75	0.26	0.55	0	0.00	0.0	0.0	3
under_votes	296	0.75	172.04	126.99	0	99.00	145.0	207.0	1008
ward	0	1.00	4.43	2.28	1	2.00	4.0	6.0	8
winner_votes	0	1.00	534.03	265.06	2	345.00	521.0	711.5	2161
write_in_votes	0	1.00	32.47	41.87	0	14.00	22.0	35.0	482
smd_ballots	296	0.75	810.70	334.29	4	588.25	808.5	1028.0	2668