SQF Data Analysis

Subhashini
Suresh Krishna

Change points

Spike

Sharp increase followed by a sudden decrease in the value. E.g: ECG pulse

Step

Structures similar to stairs in a building or a half plateau.

Monotonic Change

Monotonic increase or decrease of data points

Methods Used for change point detection

Spike

- Calculate the first order derivative of entire data.
- Since the data is discrete the forward difference a[i+1]-a[i] /2h corresponds to the first order derivatives.
- Find the points where d\dx is zero.
- The points with zero values corresponds to sharp change in value.
- Plot the index of the change points and its corresponding value.

Methods Used for change point detection

Spike - Robustness

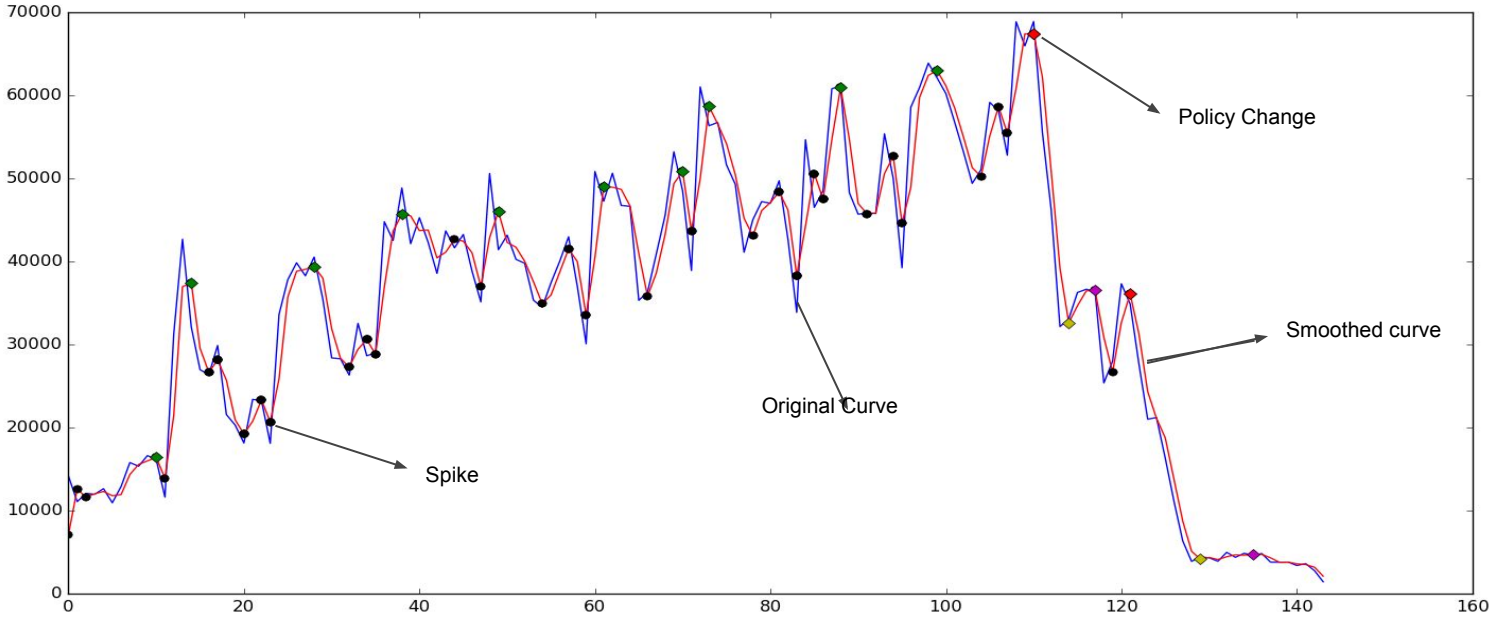
Highly accurate for sharp spikes (95 - 100%)

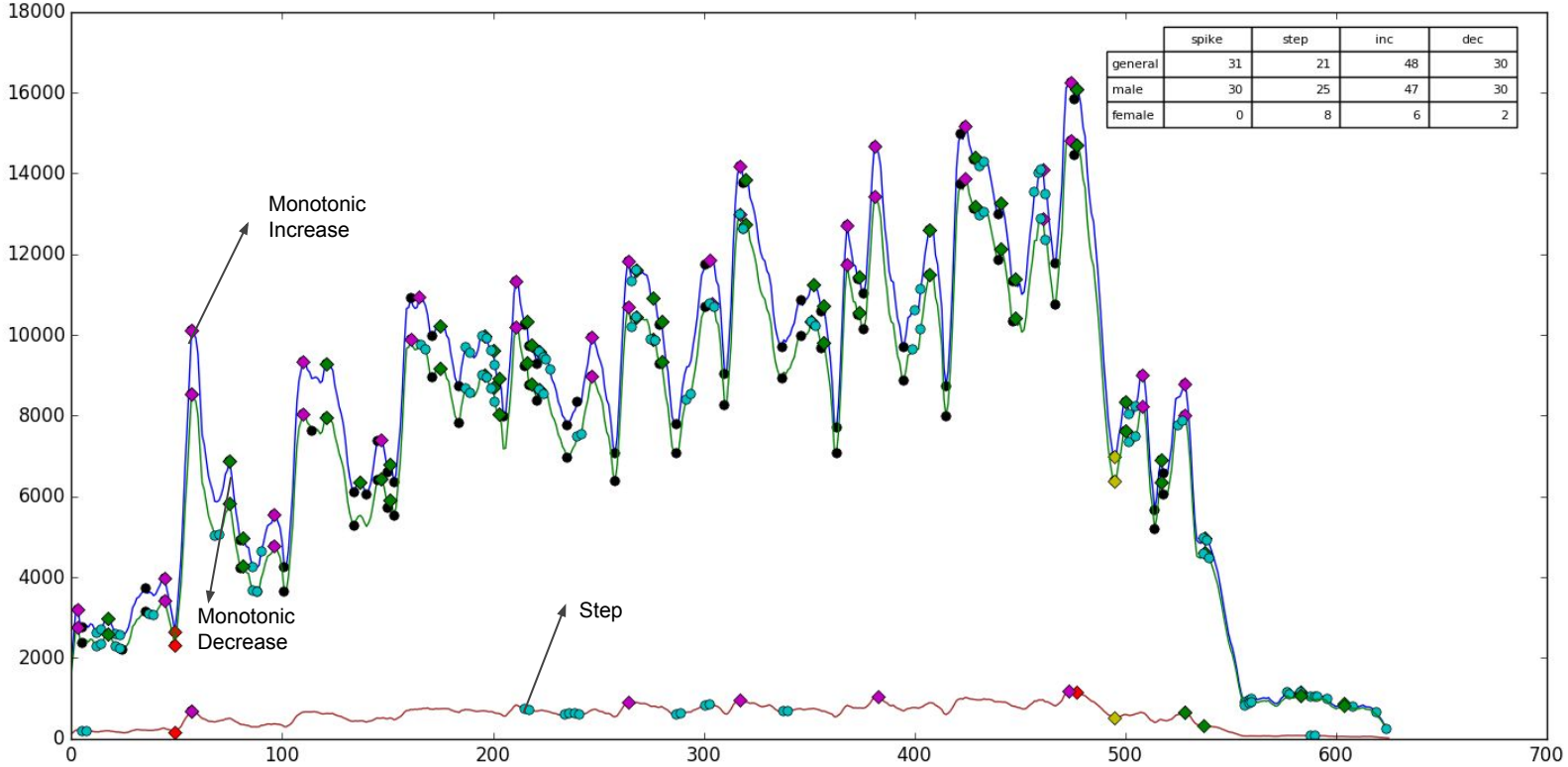
Moderately accurate for spikes with larger width (prediction depends on width threshold given by user)

Rejects very small spikes

Methods Used for change point detection

Spike - Robustness




Monotonic
Increase
Monotonic
Decrease
Step
spike step inc dec
general 31 21 48 30
male 30 25 47 30
female 0 8 6 2
0
2000
4000
6000
8000
10000
12000
14000
16000
18000
0
100
200
300
400
500
600
700

Methods Used for change point detection

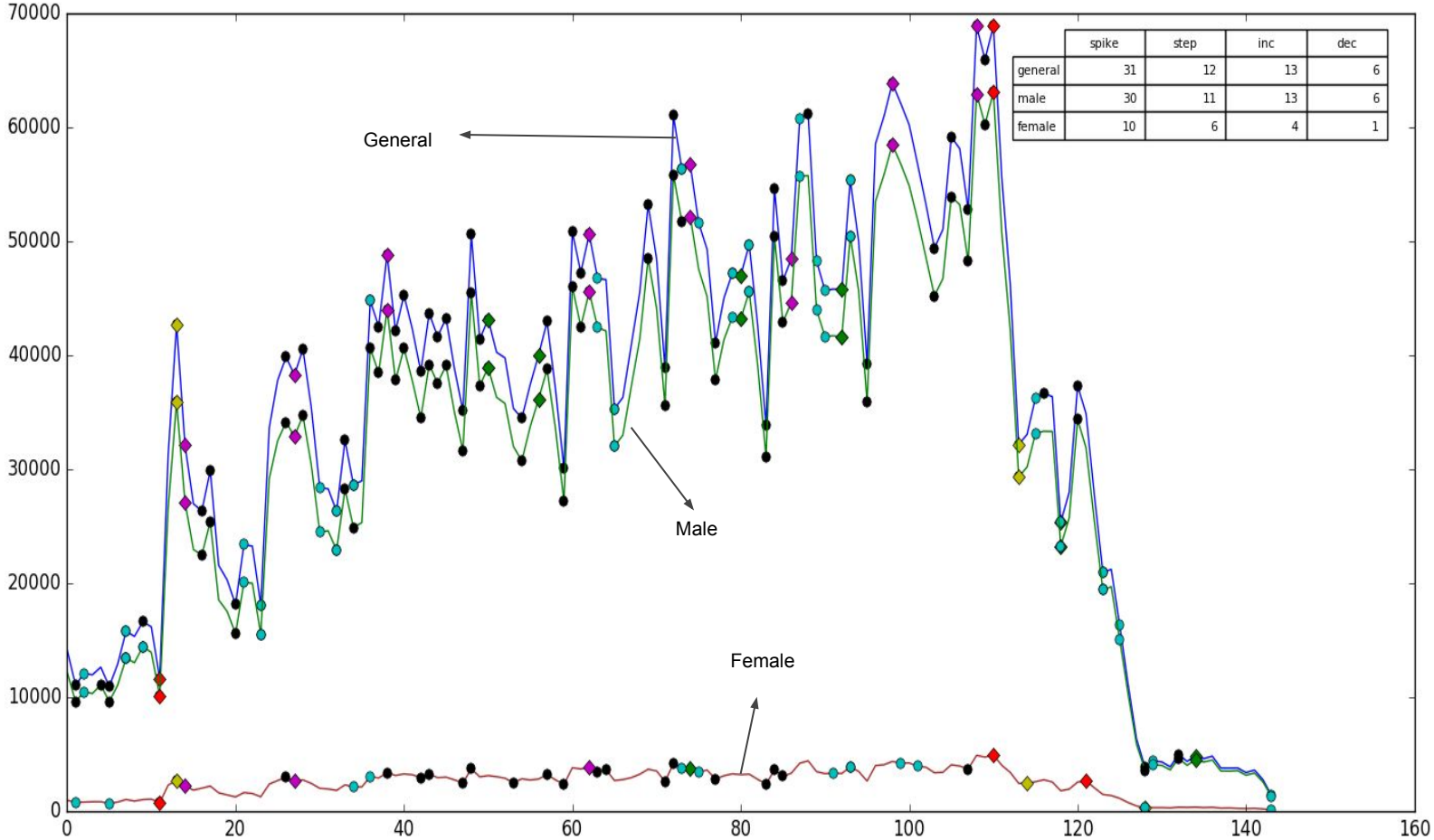
Step

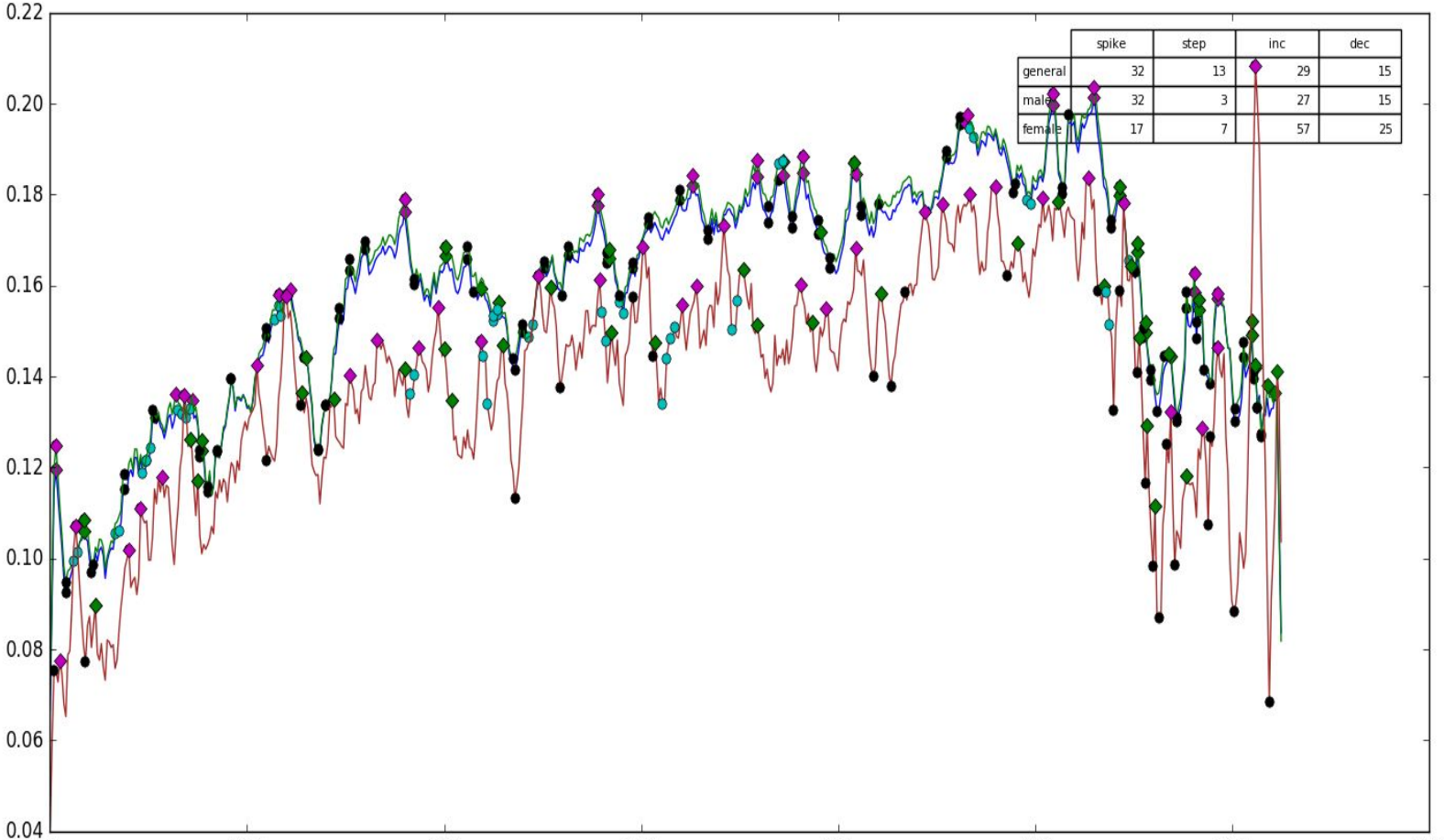
- Smooth the data with a reasonable threshold.
- Find the difference between every point
- Find the change in slope with very less threshold .
- Append all the points within the threshold.
- If the difference is greater break the loop and mark the beginning and end points.
- Check the angle between start and end points reject acute angles to avoid spikes.
- Plot the index of the change points and its corresponding value.

Methods Used for change point detection

Step - Robustness

- Finds flat regions with greater accuracy (90 - 95%)
- Finds step with slope with moderate accuracy (65 - 75 %)
- Rarely finds steps with small spikes in between.
- **False positive** on slopes with very small gradient. Extremely small spikes or spikes with very low slopes .


70000
60000
50000
40000
30000
20000
10000
0
0
20
40
60
80
100
120
140
160
General
Male
Female
spike step inc dec
general 31 12 13 6
male 30 11 13 6
female 10 6 4 1


0.22
0.20
0.18
0.16
0.14
0.12
0.10
0.08
0.06
0.04
0
100
200
300
400
500
600
700
spike step inc dec
general 32 13 29 15
male 32 3 27 15
female 17 7 57 25

Methods Used for change point detection

Monotonic Increase / Decrease

- Smooth the data with high threshold.
- Find the local maxima and minima
- Maximum point if there are points lower by some X threshold - increasing trend
- Minimum point if there are points greater than some X threshold - decreasing trend
- Plot the index of the change points and its corresponding value.

Methods Used for change point detection

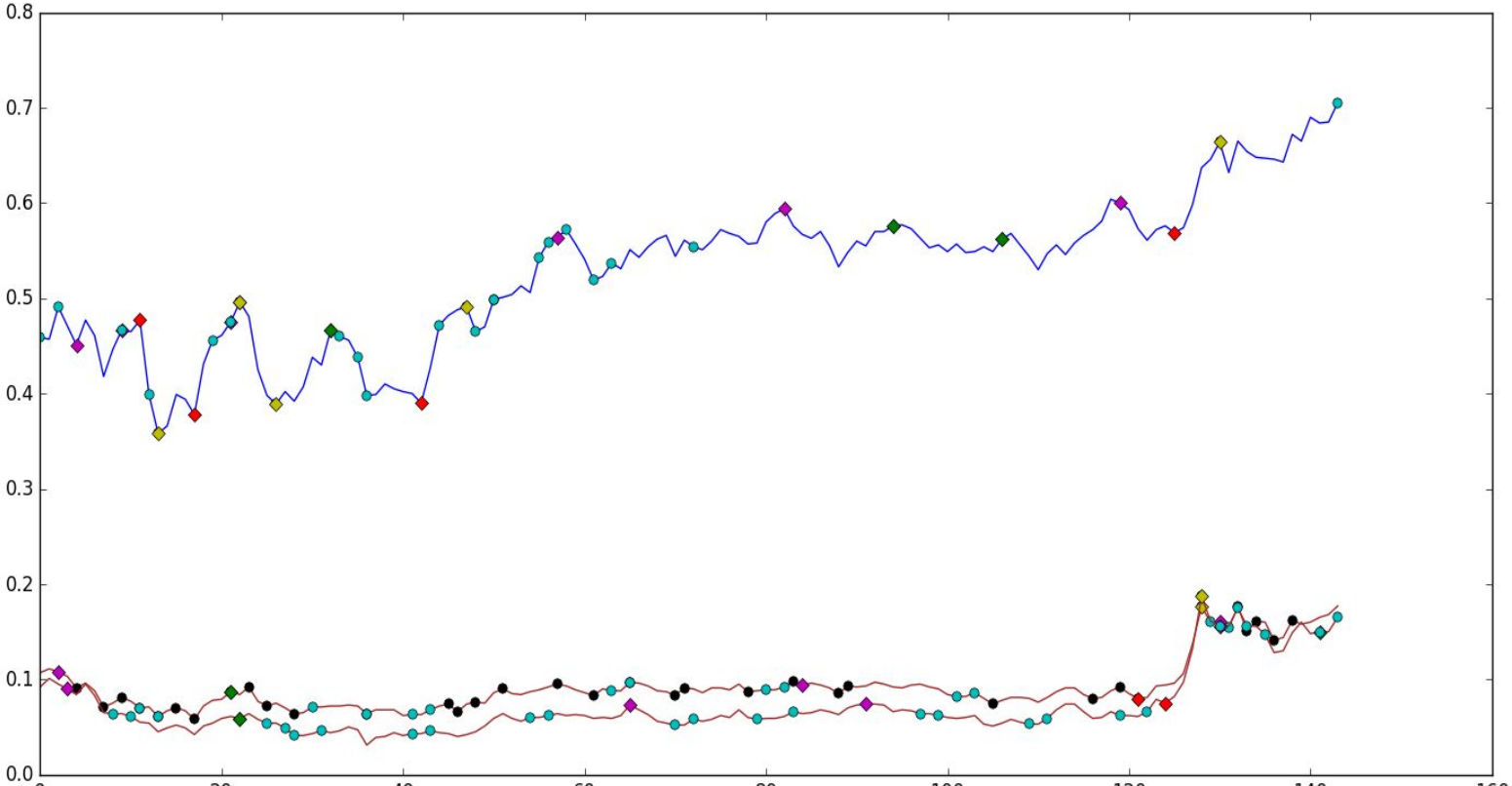
Monotonic Trend - Robustness

- Highly accurate if the increase / decrease is high (85-95%) .
- Highly accurate if there is a single sharp change in trend (80 - 95%)
- Moderately accurate if there is a sharp change in trend continuously (65 - 75%)

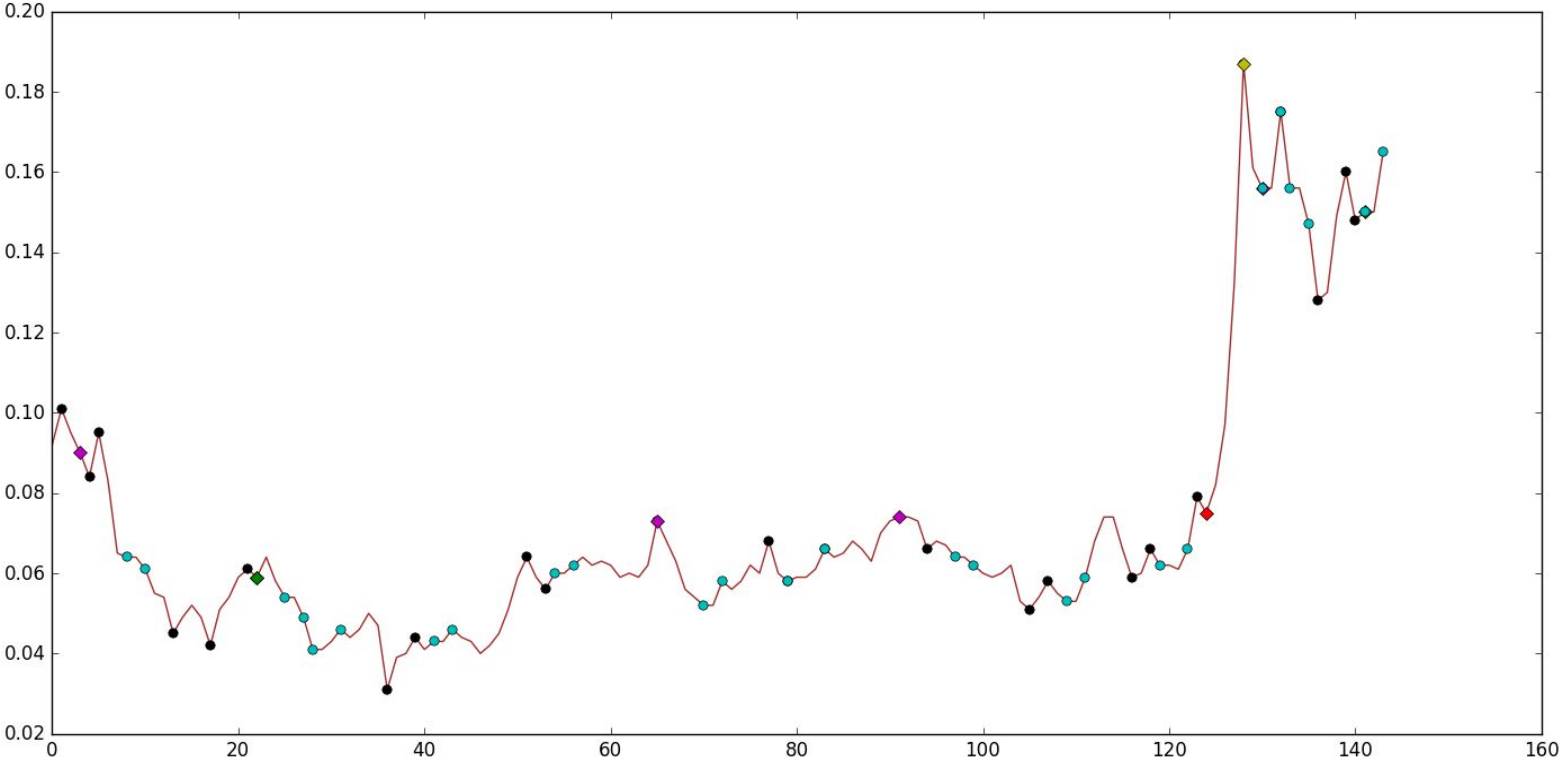
Anomalies

- Fraction of people frisked is very high compared to arrested and searched
  - policy for search and arrest is very different from frisk.
- There is a huge jump in fraction of people arrested or searched after 2012 but frisk fraction is unaffected.
  - implies policy change for search and arrest around 2012
- The fraction of frisk rf_vaact plot has a huge jump in between 2004 - 2005
  - policy for frisk with reason “ ACTIONS OF ENGAGING IN A VIOLENT CRIME” has changed in between 2004 - 2005.

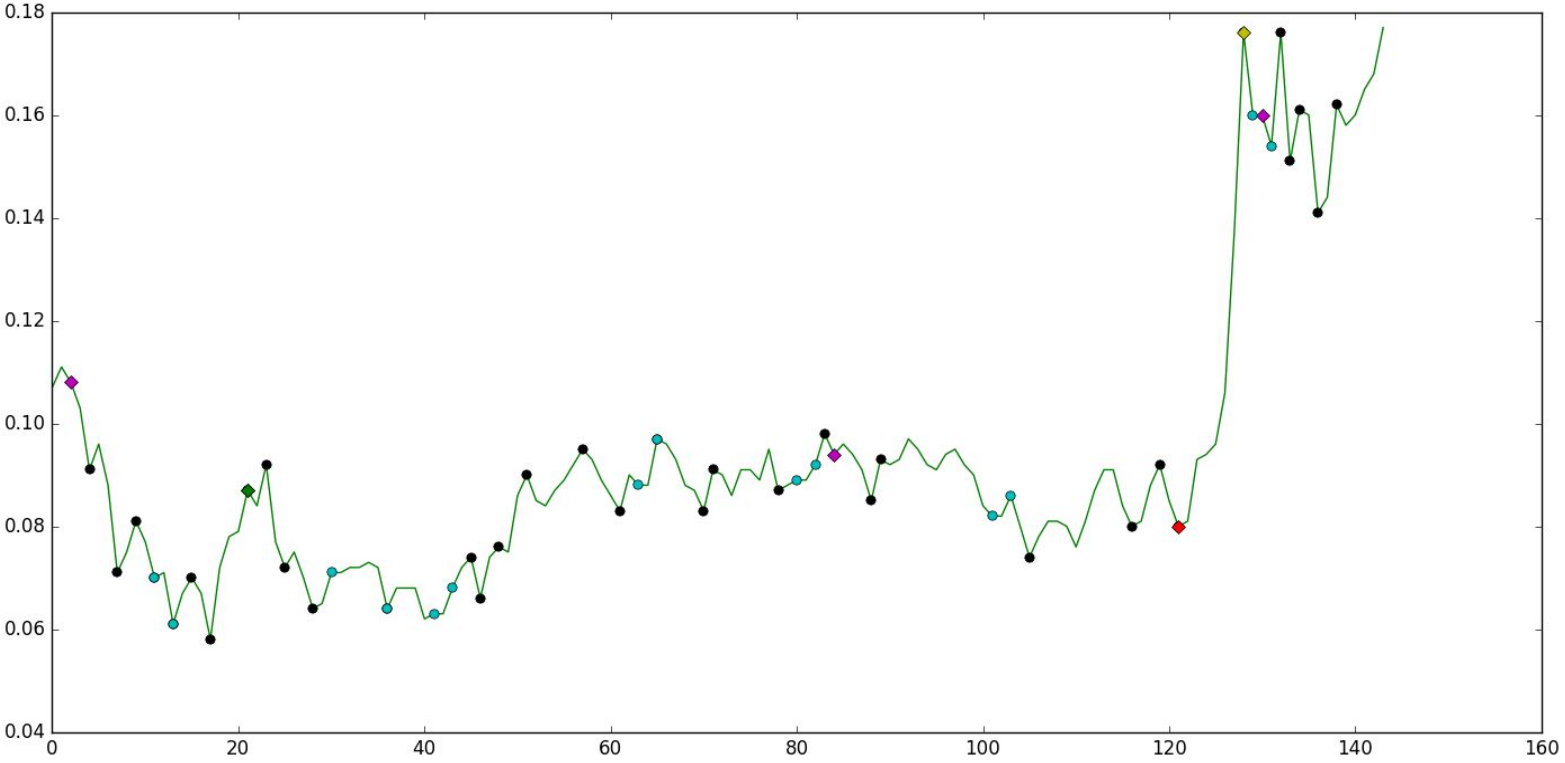
Frisk Vs Arrest Vs Search



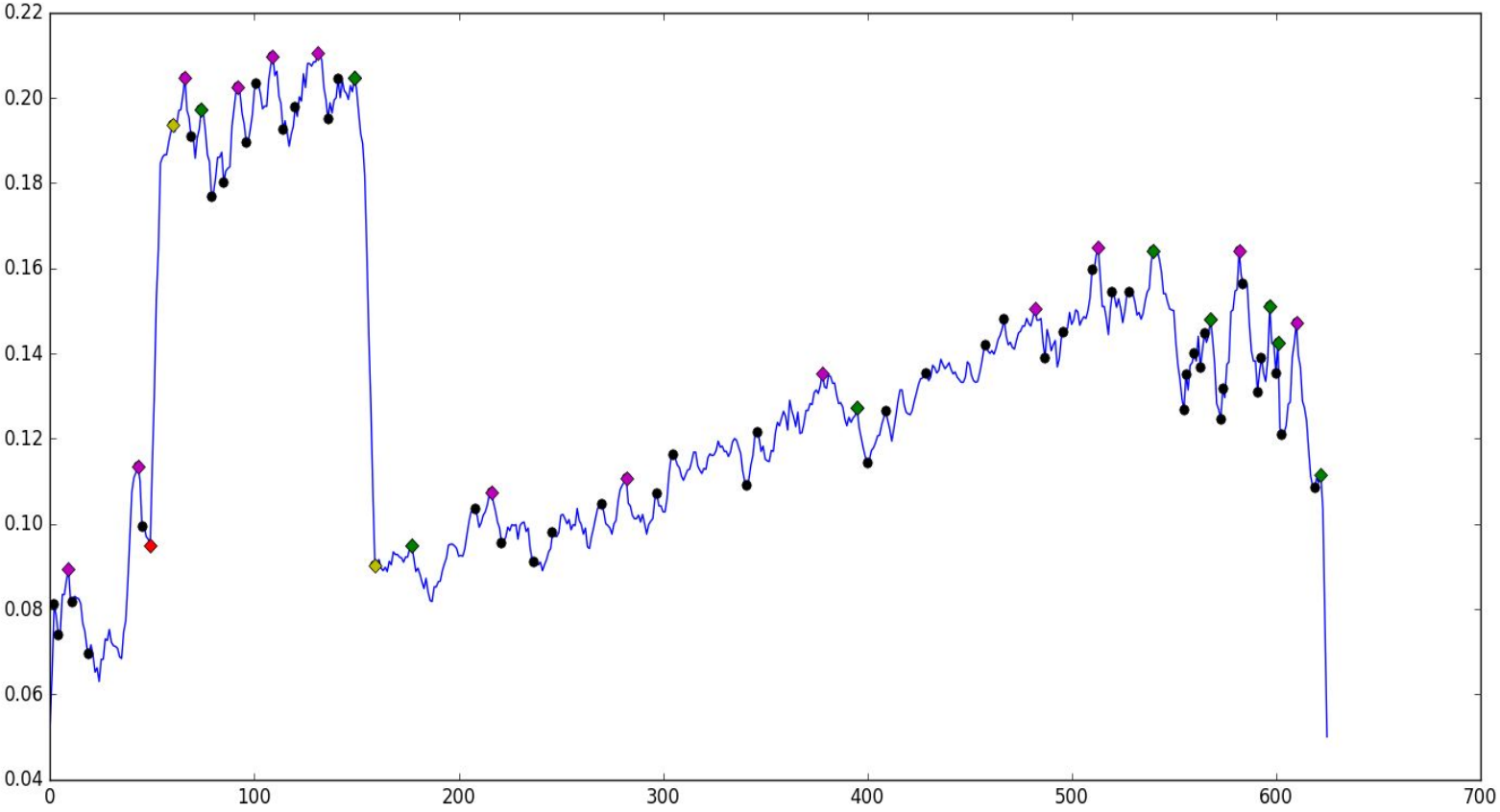
Monthly Arrest Fraction



Monthly Search Fraction



Rf_vcact fraction



Phase 2.a

KI - Divergence

KL - Divergence Calculation

Wikipedia quotes

> “a measure of the difference between two probability distributions *P* and *Q*.”

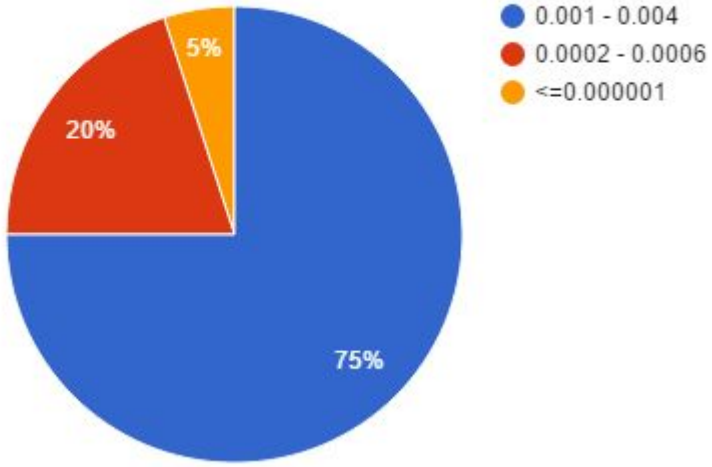
$$D_{\mathrm{KL}}(P\|Q) = \sum_{i} P(i) \log \frac{P(i)}{Q(i)}.$$

Methodology for KL - Divergence

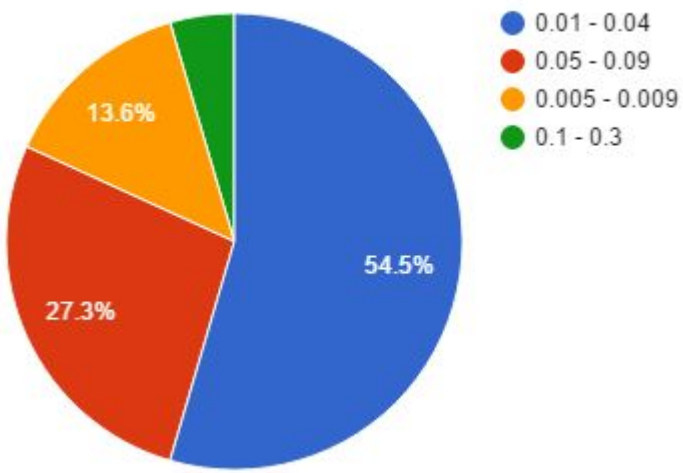
- In this dataset, pk is probability of the main data (eg stop)
- Qk is the probability for which divergence has to be computed (eg male stop, female stop, black stop etc.,..
- Compute KL-divergence using the formula given in last slide
- Plot the graphs with divergence value
- More divergence value signifies the group qk is treated differently when compared to pk
- Less divergence value signifies pk and qk treated relatively same

Value Distribution

Male kl

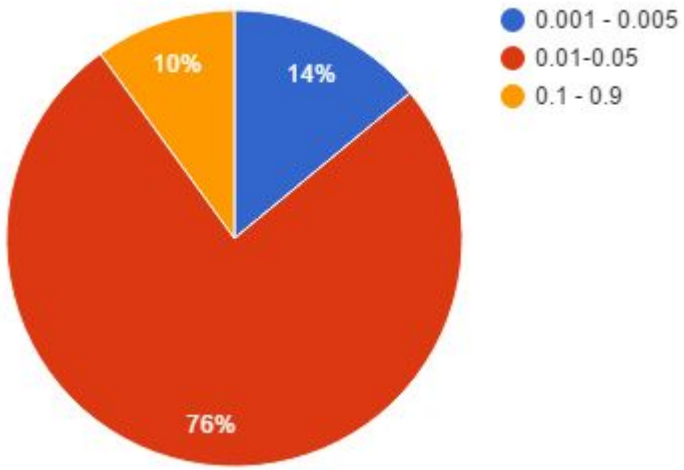


Female kl

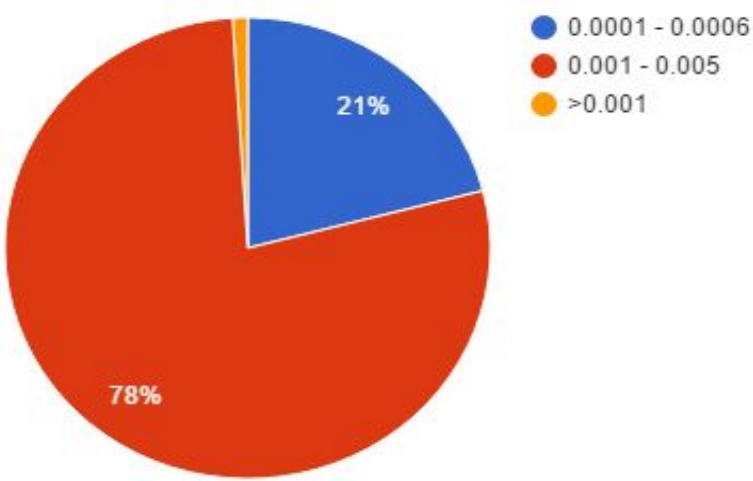


Value Distribution

White kl

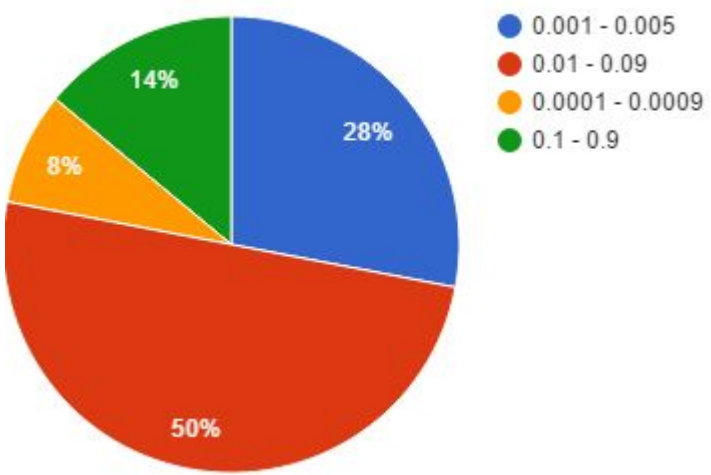


Black kl

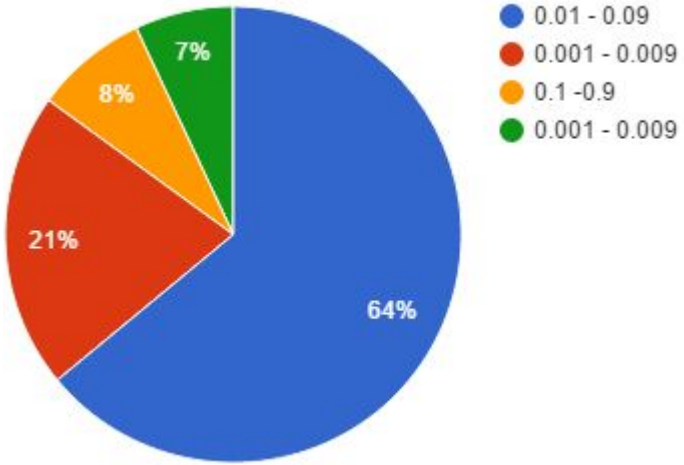


Value Distribution

White Hispanic kl



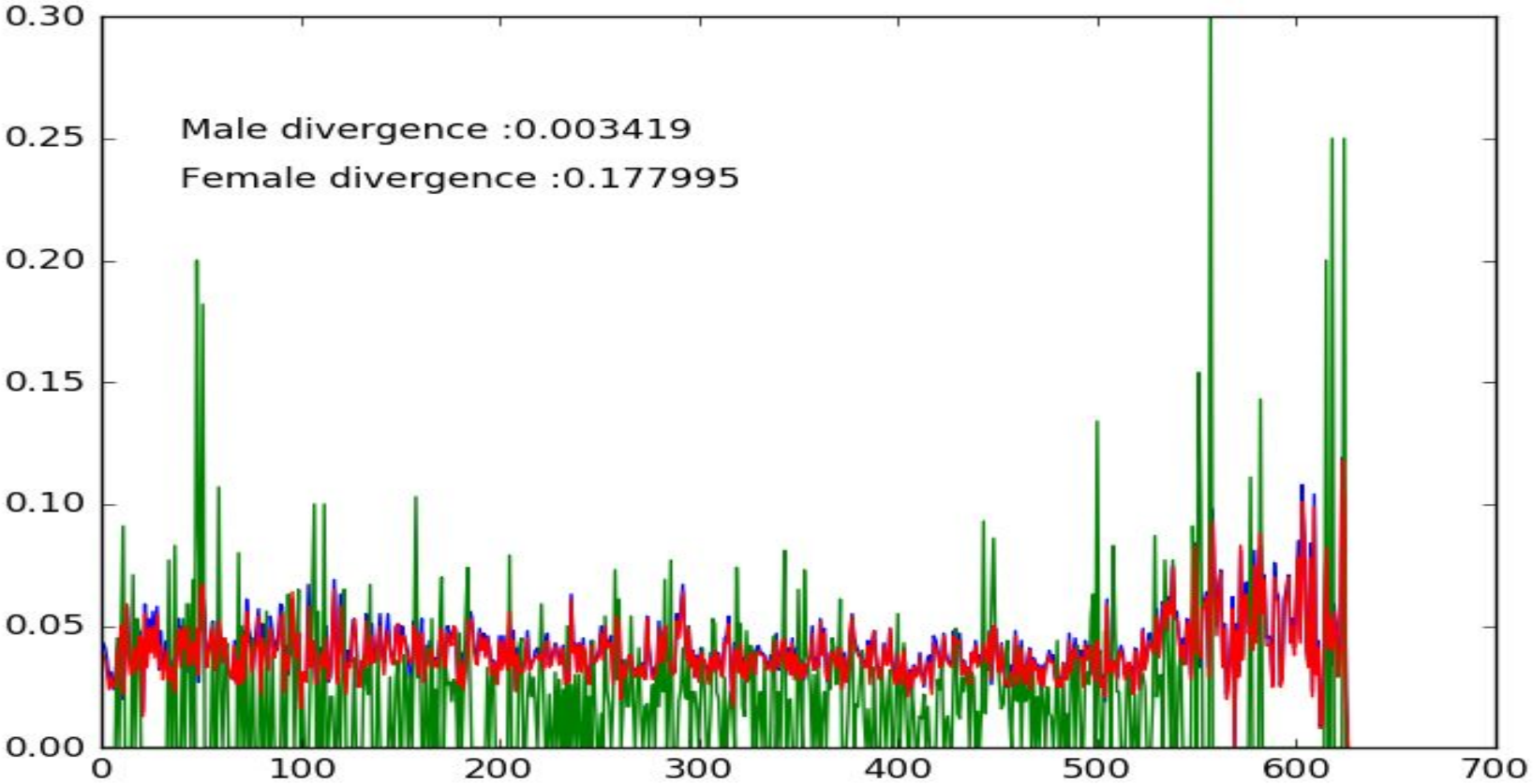
Black Hispanic kl



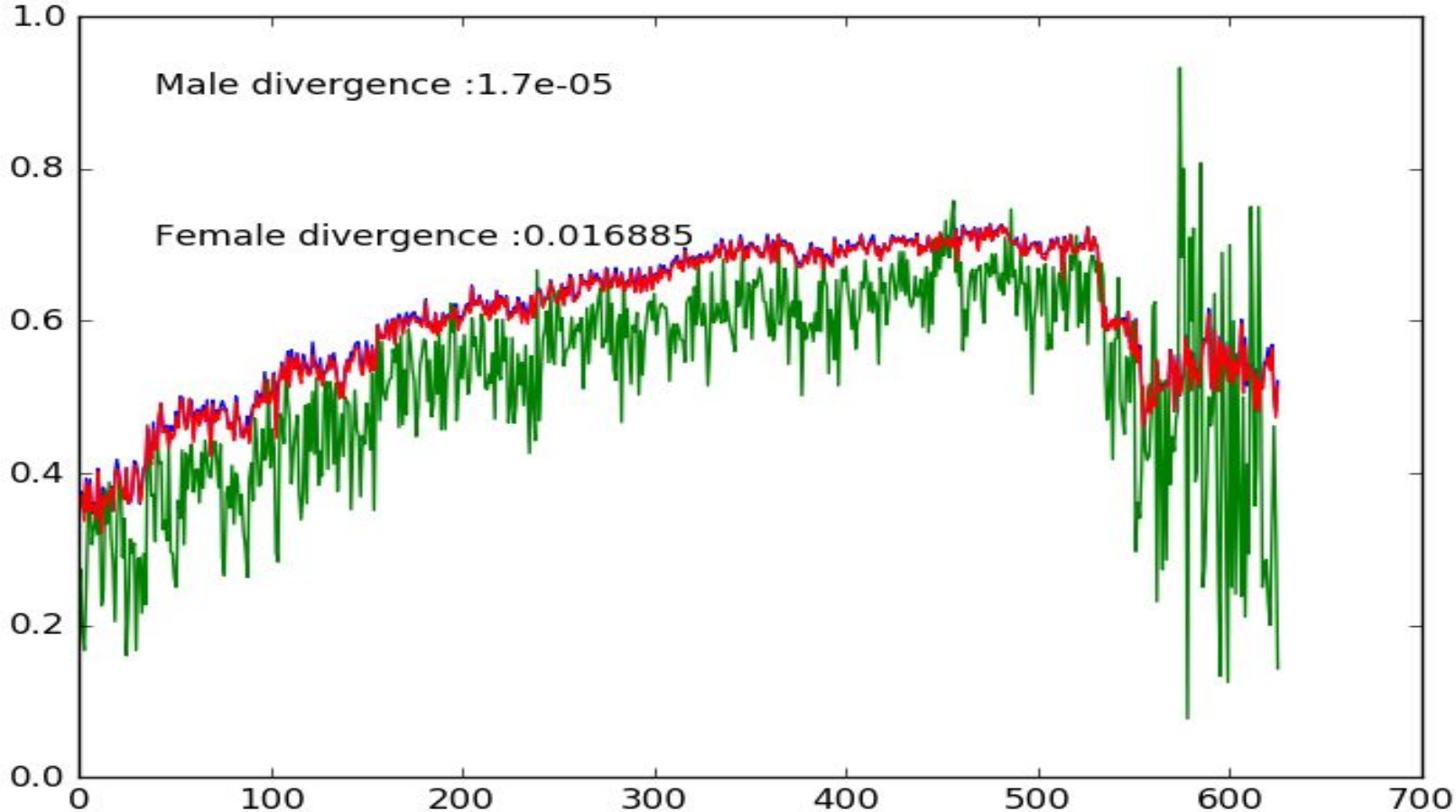
KL - Divergence findings

- Divergence of female is very high than male in most cases.
- Divergence of race - black is mostly very less
- Divergence of whites are generally high
- The top three of maximum divergence in Gender are
    - {"weeksearchsb_admisfemalefl": 0.177195, "weeksearchsb_outlnfemalefl": 0.14856, "weekfriskrf_knowlfemalekl": 0.182248}
- The top three of maximum divergence in Race are
    - {('divergencevaluesearchWsb_admis': 0.404236, 'divergencevaluesearchQsb_admis', 0.348035), ('divergencevaluefriskedWrf_verbl': 0.399998)}

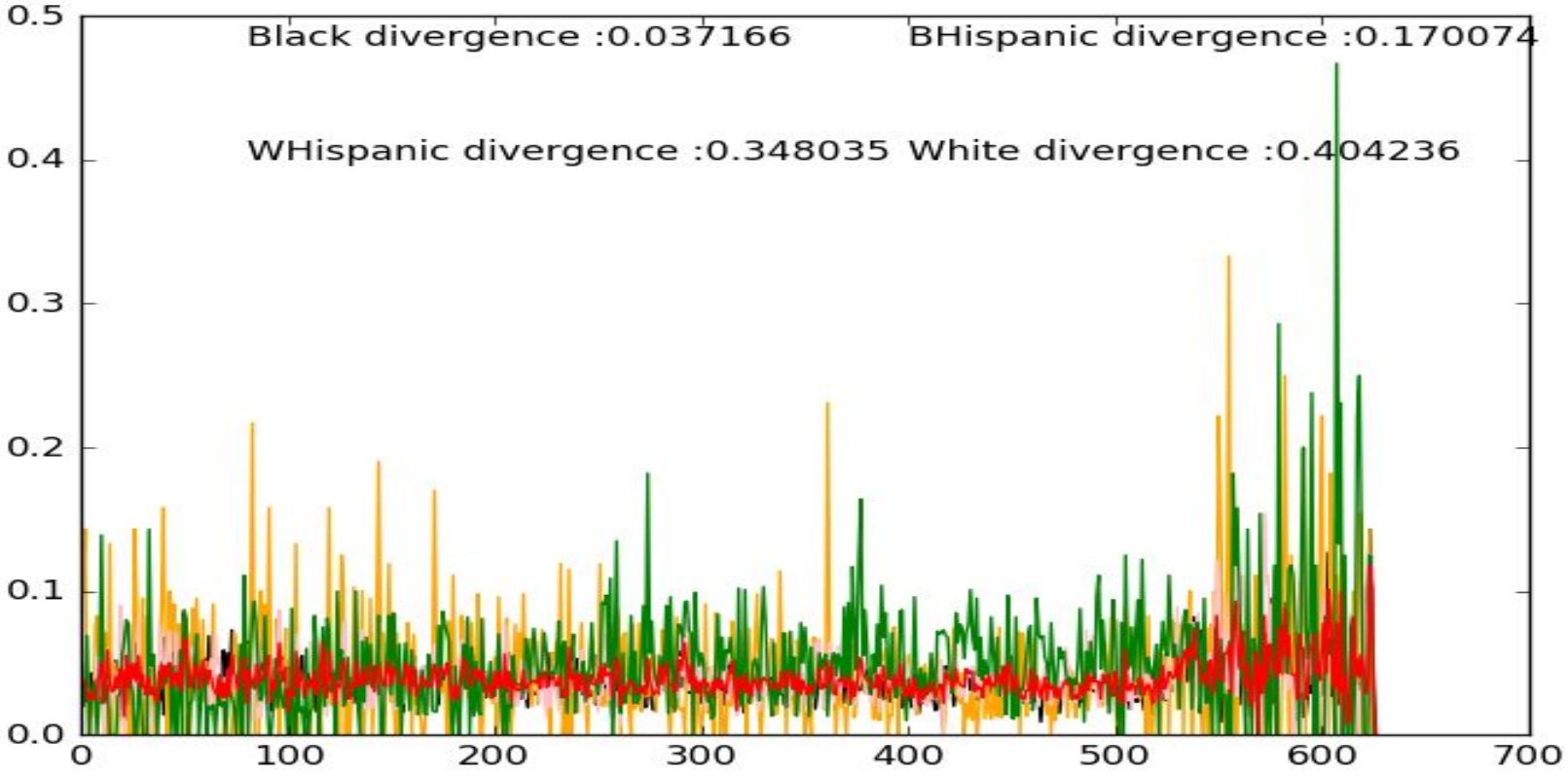
Week search sb_admis



Week frisk rf_furt - High Divergence between male and female



Week frisk divergence



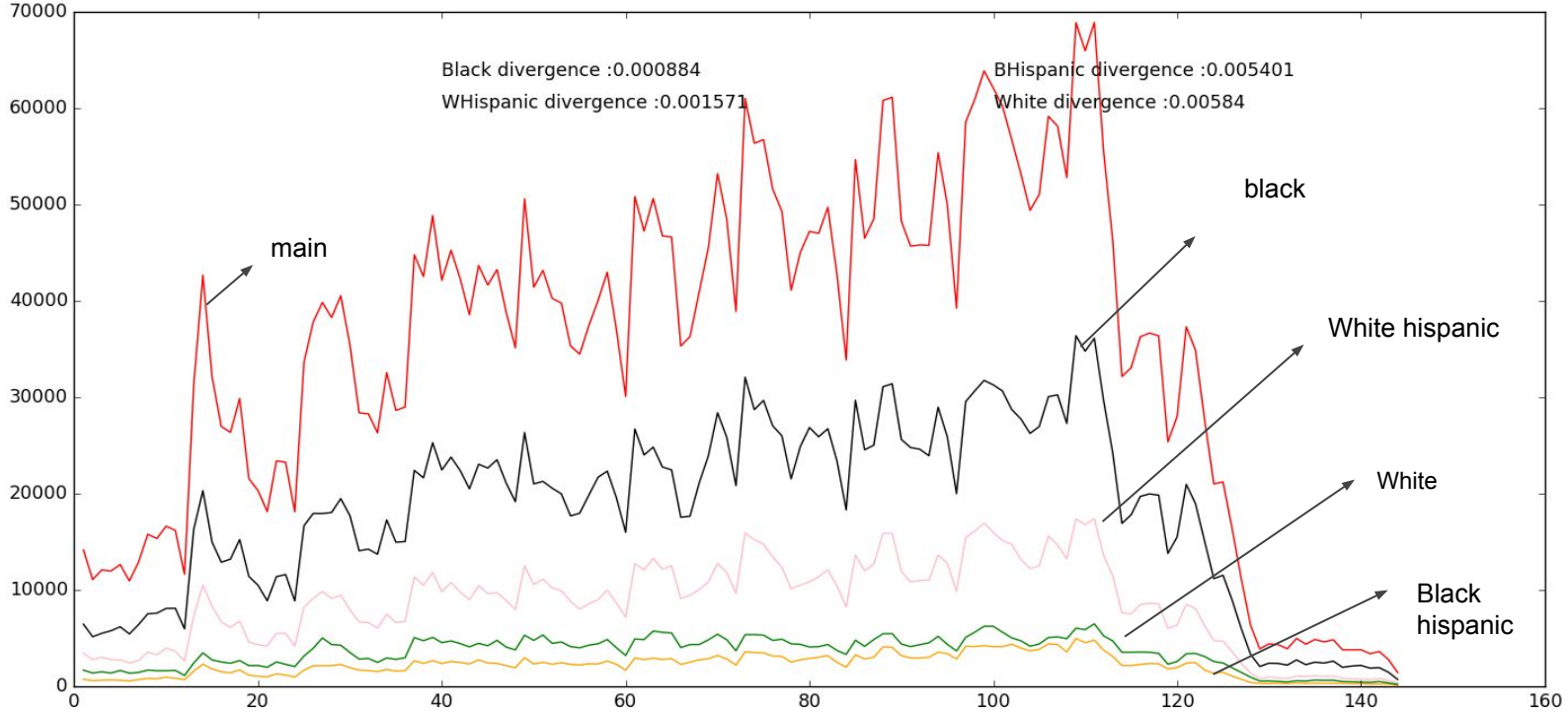
Phase 2.b

Policy change for Sub - Groups

Findings

- There is a major policy change in 2012
- The number of people stopped, frisked, arrested had hugely reduced.
- The proportion of black people searched, frisked or arrested is higher than white in almost all cases

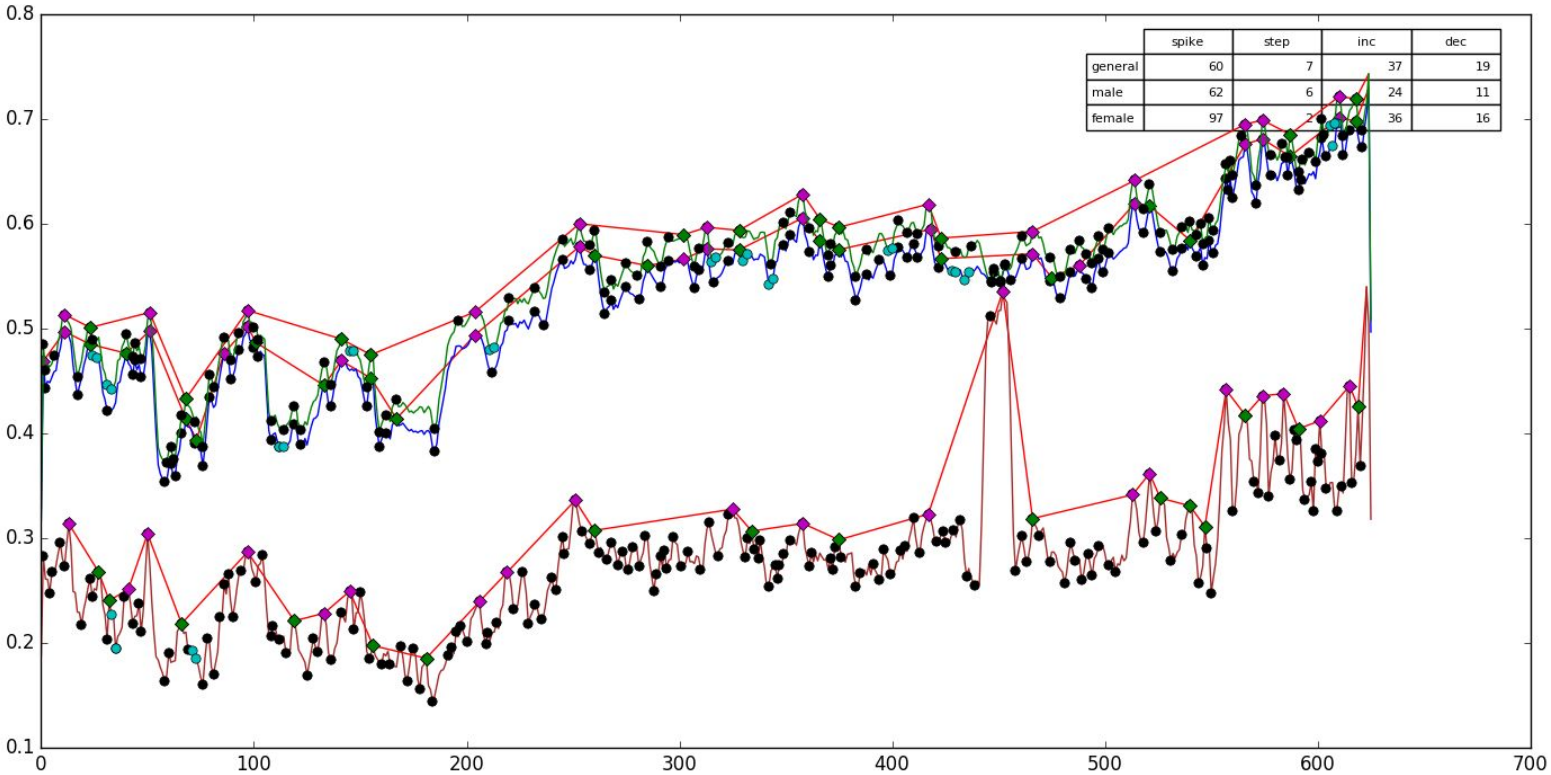
Month stop race



Findings

- For weekly frisk fraction
  - Year 2011
    - There is a sudden huge increase for female but there is a decrease in male frisk
    - Policy change to stop more female

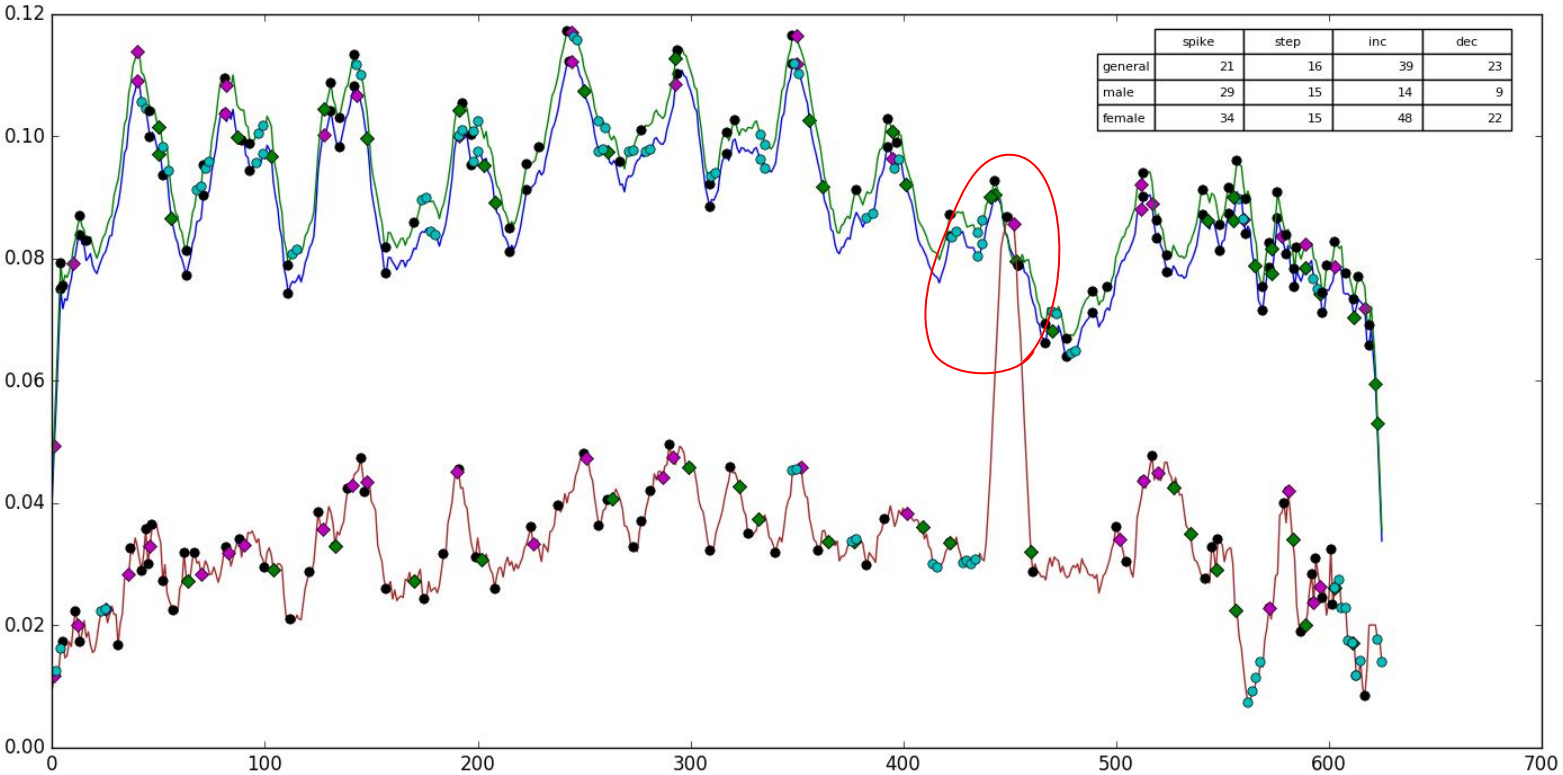
Weekly Frisk fraction Gender



Findings

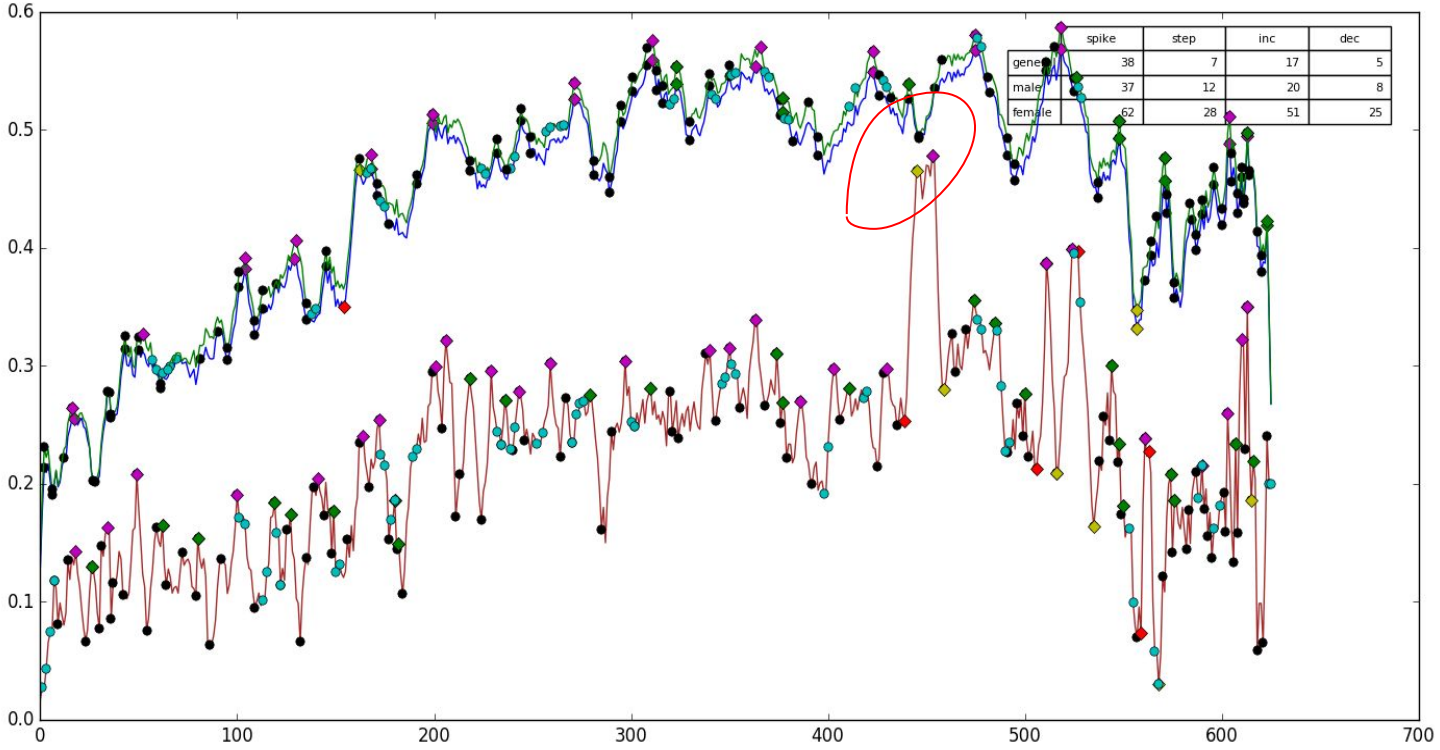
- For weekly stop reason “SUSPICIOUS BULGE” and “Search Hard Object”
  - Year 2011
    - There is a sudden huge increase for female but there is a decrease in male frisk
    - Policy change to stop more female

Weekly Stop cs_Bulg



	spike	step	inc	dec
general	21	16	39	23
male	29	15	14	9
female	34	15	48	22

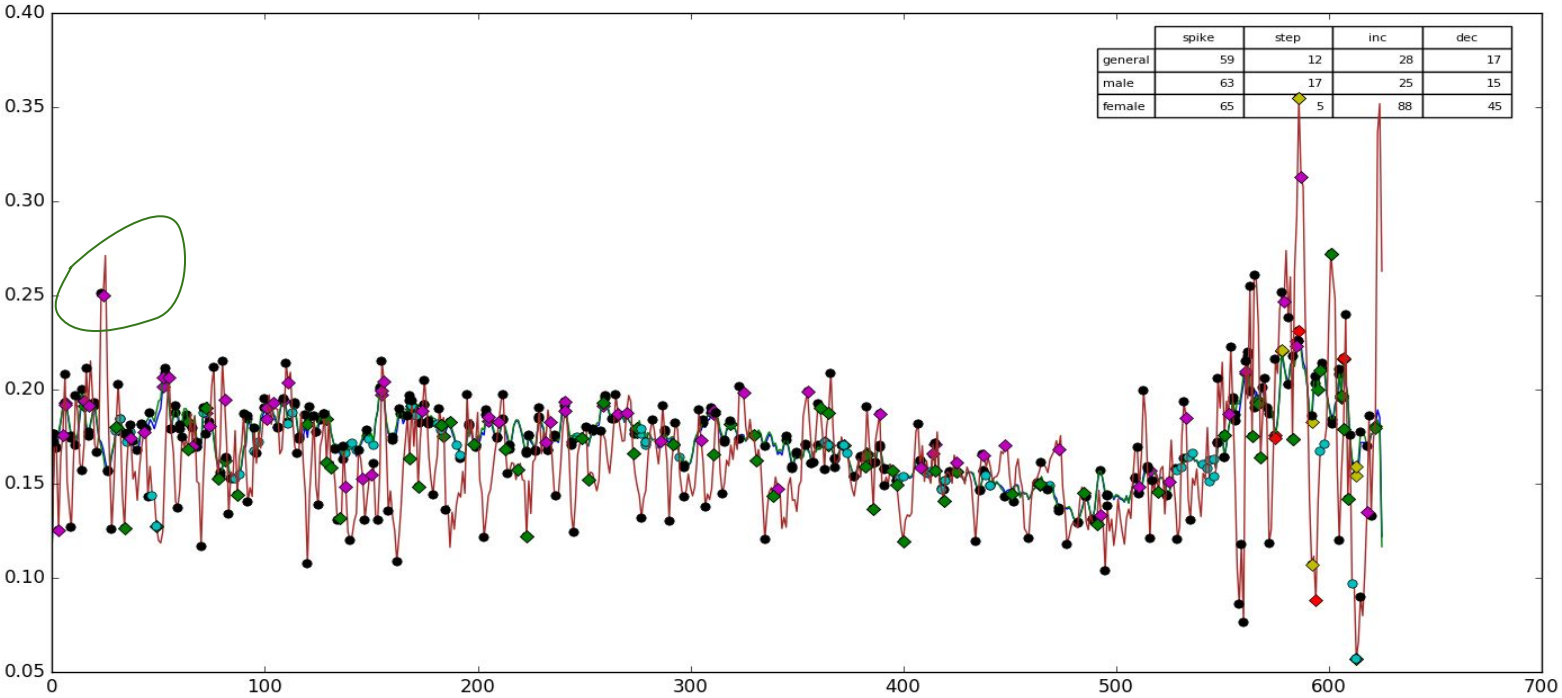
week_search_hdobj_gender



Findings

- For the frisk reason rfcmp “REFUSE TO COMPLY WITH THE OFFICER”
  - Throught the graph, we could see the monotonic trend complies with the main curve , but for the females the variance is very random
  - In the year 2003, there is a sharp increase in female frisked ratio.

Week_frisk_rfcmp_gender

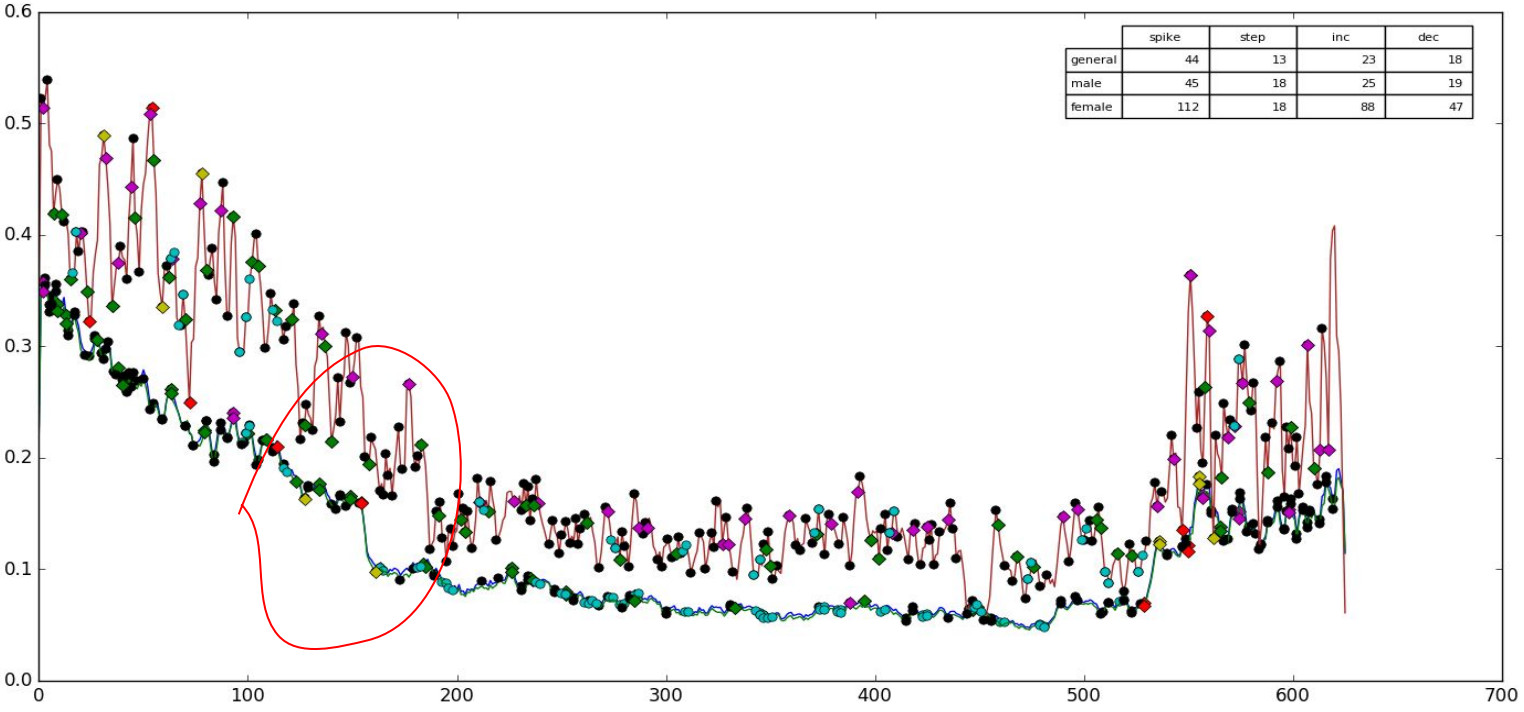


	spike	step	inc	dec
general	59	12	28	17
male	63	17	25	15
female	65	5	88	45

Findings

- For the frisk reason “OTHER SUSPICIONS OF WEAPONS”
    - In year 2005-2006, there is a monotonic increase in number of female frisked whereas a monotonic decrease in number of males frisked

frisk_rf_othsw_gender

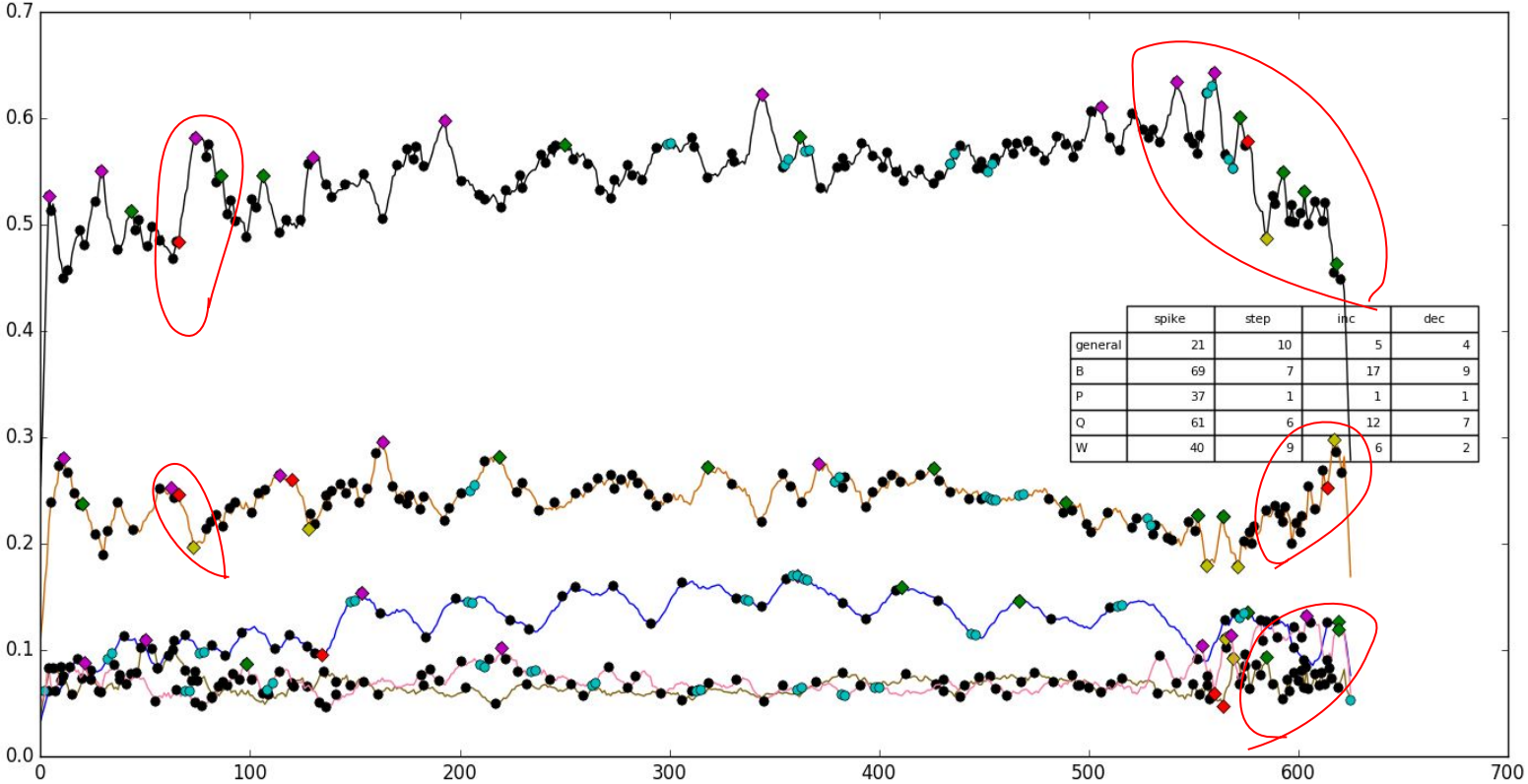


	spike	step	inc	dec
general	44	13	23	18
male	45	18	25	19
female	112	18	88	47

Findings

- For the frisk reason “INAPPROPRIATE ATTIRE FOR SEASON”
  - year 2004
    - huge increasing spike for black race whereas for other race there is a considerable decrease in frisk count
    - there was a policy change for blacks.
  - years 2013 - 2014
    - sharp fall in the number of blacks frisked and a gradual increase for other races
    - some policy change to reduce blacks frisked

Week frisk rf_attir

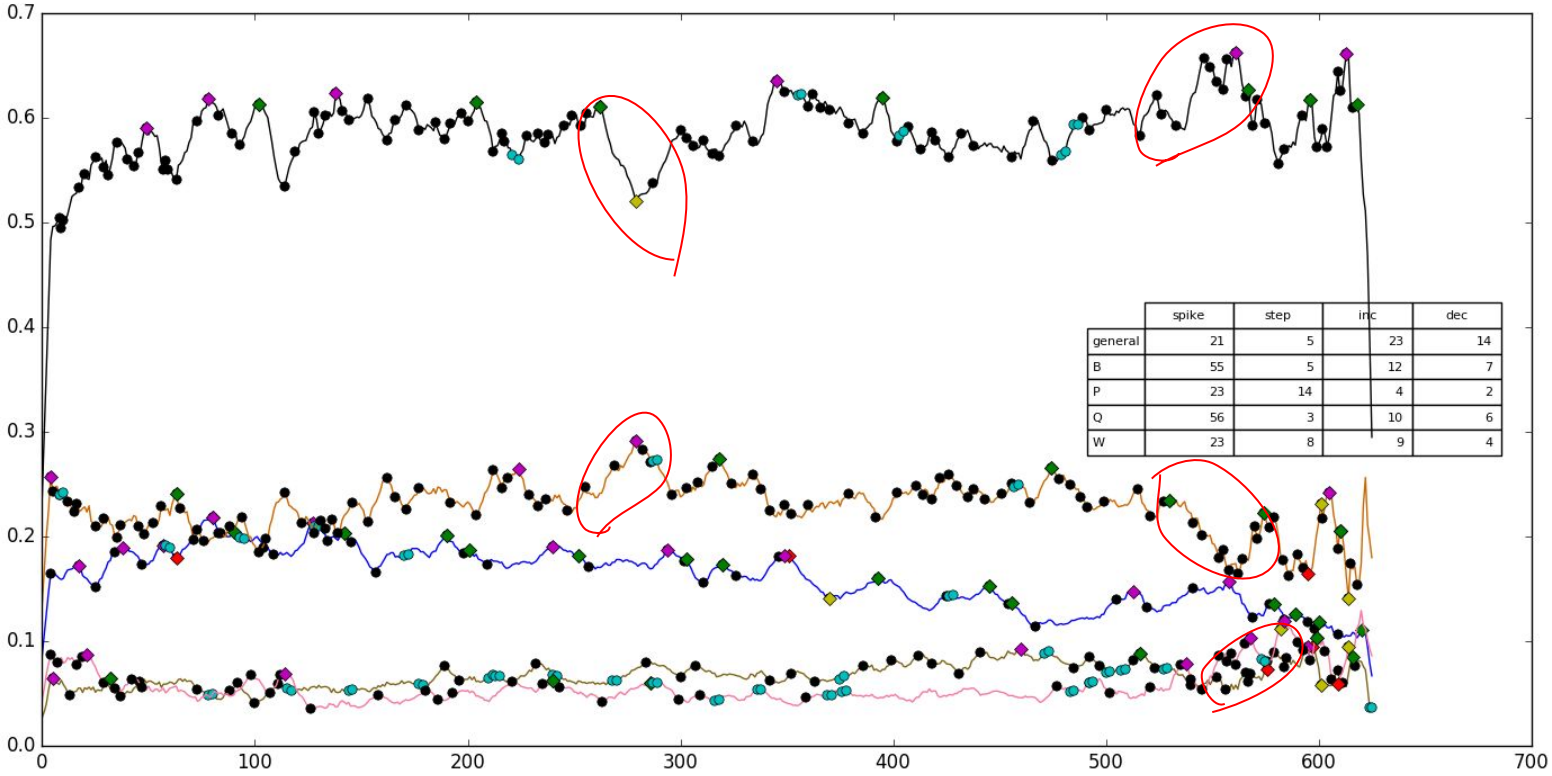


	spike	step	inc	dec
general	21	10	5	4
B	69	7	17	9
P	37	1	1	1
Q	61	6	12	7
W	40	9	6	2

Findings

- For the frisk reason “Frisk for reason bulge”
  - year 2008
    - huge decreasing spike for black race whereas for white hispanic there is a monotonic increase in frisk count
    - there was a policy change to frisk more white hispanics.
  - years 2012 - 2013
    - Monotonic decrease in the number of white hispanics frisked and a gradual increase for other races
    - some policy change to support white hispanics.

Week frisk rf_bulg

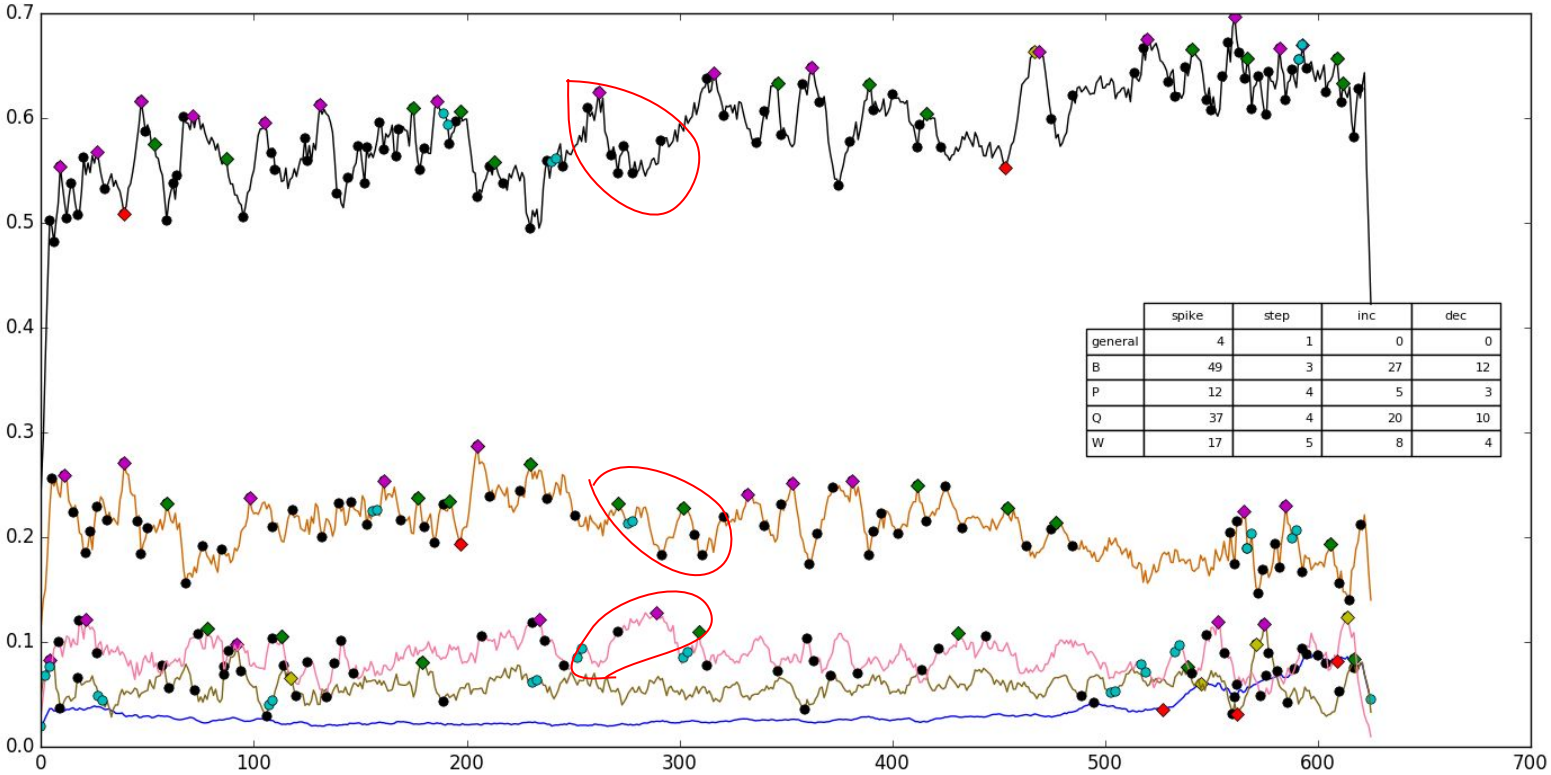


	spike	step	inc	dec
general	21	5	23	14
B	55	5	12	7
P	23	14	4	2
Q	56	3	10	6
W	23	8	9	4

Findings

- For the frisk reason “INAPPROPRIATE ATTIRE FOR SEASON”
  - year 2007 - 2008
    - Monotonic increase in white whereas monotonic decrease for other races.
    - there was a policy change to frisk more white people.

Week frisk rf_Knowl

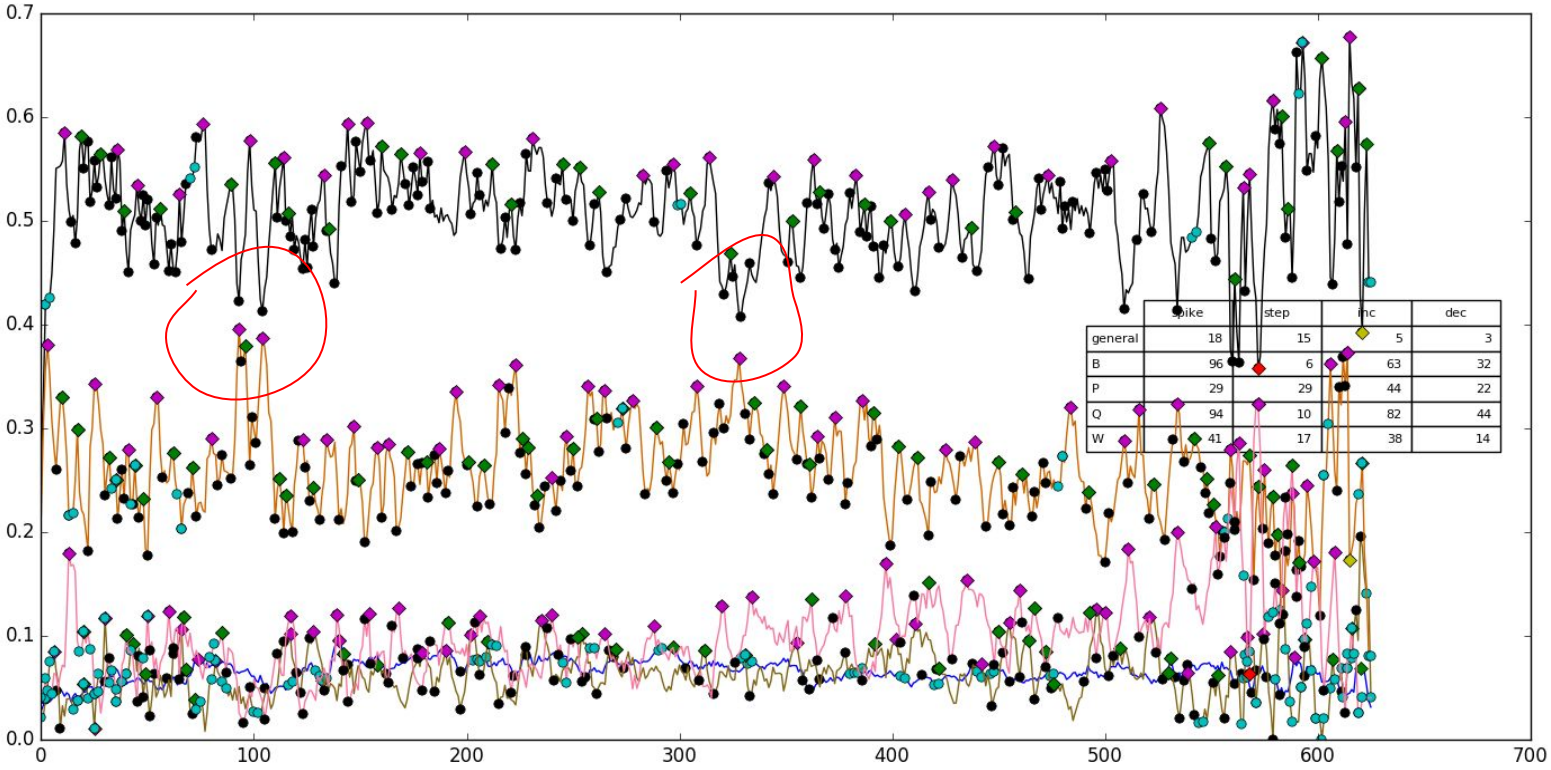


	spike	step	inc	dec
general	4	1	0	0
B	49	3	27	12
P	12	4	5	3
Q	37	4	20	10
W	17	5	8	4

Findings

- For the search reason “OUTLINE OF WEAPON”
  - year 2004 - 2005
    - Monotonic increase in white hispanic whereas monotonic decrease for other races.
    - there was a policy change to search more white people.
  - year 2009- 2010
    - Monotonic decrease in black whereas monotonic increase for other races.
    - there was a policy change to search less black people.

week_sb_outln

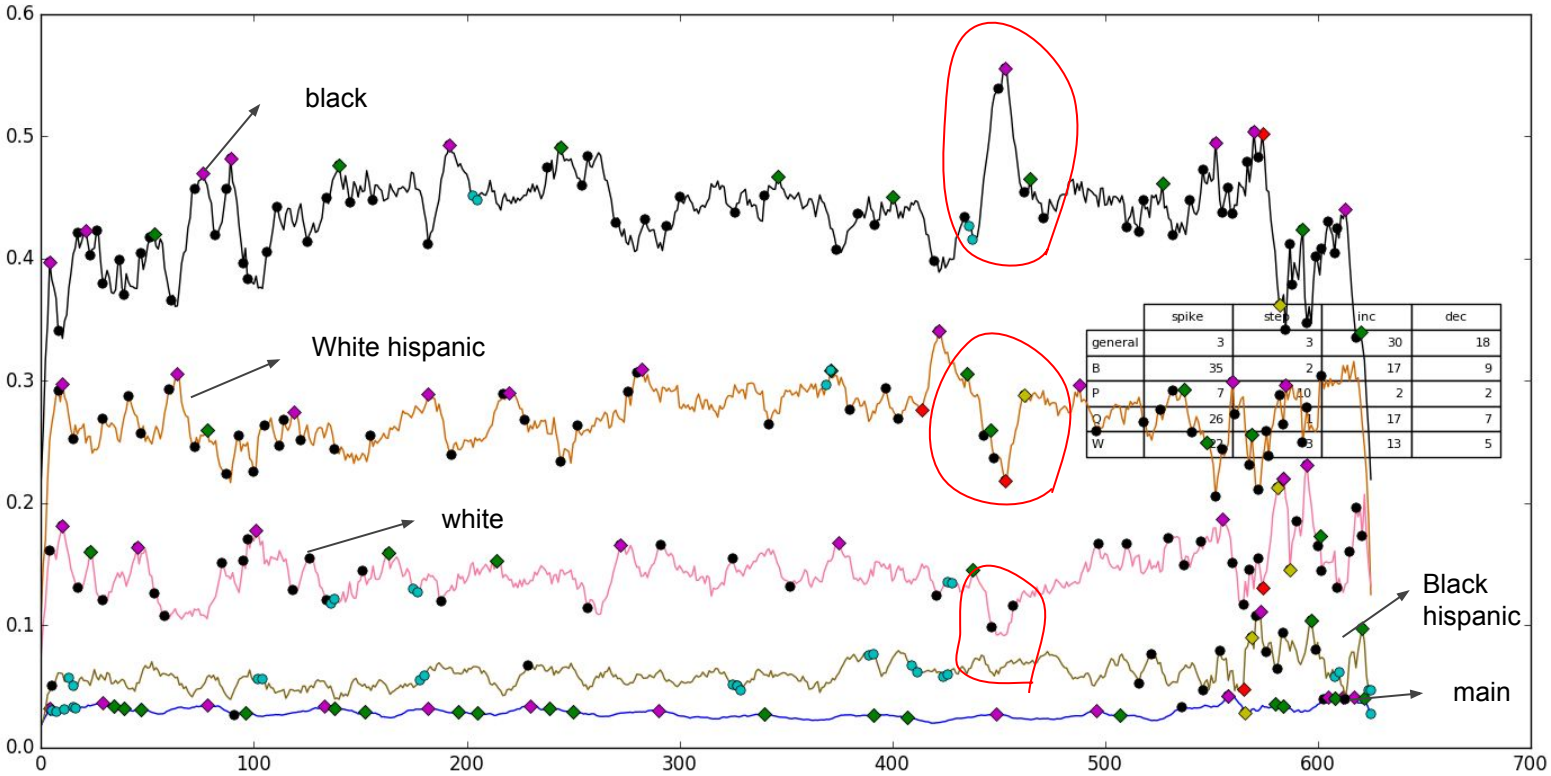


	spike	step	inc	dec
general	18	15	5	3
B	96	6	63	32
P	29	29	44	22
Q	94	10	82	44
W	41	17	38	14

Findings

- For the stop reason “CARRYING SUSPICIOUS OBJECT”
  - year 2011 - 2012
    - Sharp spike in black whereas monotonic/sharp decrease for other races.
    - there was a policy change to stop more black people.

Week_stop - cs_objcs



	spike	step	inc	dec
general	3	3	30	18
B	35	2	17	9
P	7	10	2	2
Q	26	1	17	7
W	2	3	13	5

Findings

- For the stop reason “SUSPICIOUS BULGE”
  - year 2007
    - Gradual decrease in black whereas gradual increase for other races.
    - there was a policy change to stop less black people .
  - year 2014
    - Monotonic decrease in black stop but monotonic increase for others
    - there was a policy change to stop less black people .

Week_stop - cs_bulg

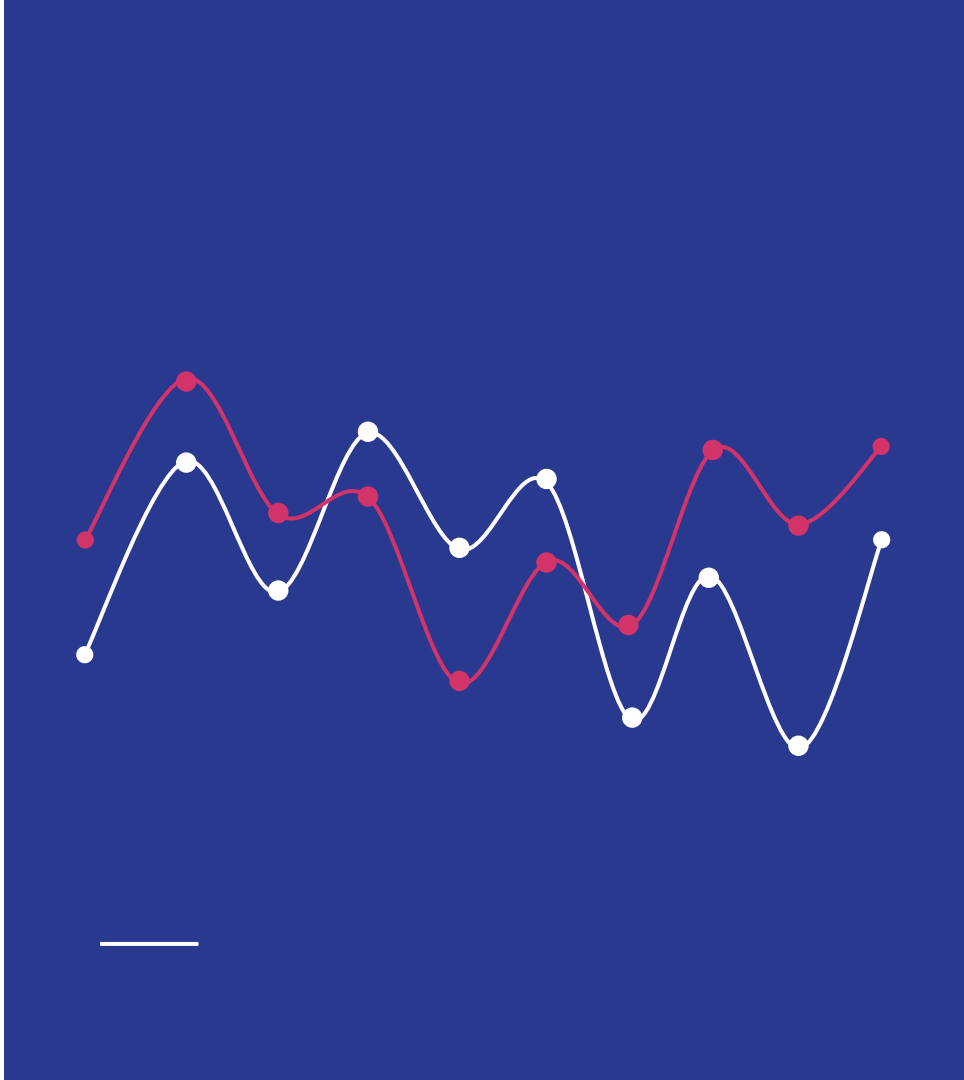
	spike	step	inc	dec
general	15	20	1	1
B	59	2	13	7
P	24	10	2	0
Q	51	4	8	3
W	22	14	0	0

Findings

To conclude

- year 2007
  - there was a policy change to stop less black people .
- year 2007
  - Policy change to stop more female.
- Black hispanics are very rarely affected by policy change.

Thank You



Anomalies - Extra for races

- Rf_cmp - White and Black hispanic has higher divergence
- Sb_outln - White hispanic has higher divergence
- Cs_casng - Black hispanic has higher divergence
- Cs_drgtr - Black and white has relatively same divergence