Trend Flagging to Aid Resource Allocation Decisions

Fred S. Roberts

Director

Command, Control, and Interoperability Center for Advanced Data Analysis (*CCICADA*)

Source: wikimedia commons

Command, Control, and Interoperability Center for Advanced Data Analysis
A Department of Homeland Security Center of Excellence

Flagging Trends

- Decision makers need to be able to flag trends so as to respond to changes.
- Response might be rethink policies or reallocate resources.
- US Coast Guard a case in point.
  - Looks at changing shipping volumes or vessel behavior patterns
  - Reallocates resources to missions such as search & rescue, environmental protection, narcotics interdiction

Credit: commons.wikipedia.org

CCICADA
Command, Control, and Interoperability
Center for Advanced Data Analysis
A Department of Homeland Security Center of Excellence

Flagging Trends

- Increased shipping in the Arctic due to changing ice conditions will allow more eco-tourism and cruise ships – requiring preparation for more search & rescue.
- Increased petrochemical shipping in the Houston Ship Channel increases risk of oil spill.
- Increased frequency of vessels anchoring offshore for unusually long periods of time might suggest smuggling of narcotics or weapons.

Credit: commons.wikipedia.org

Stacks of cocaine
Credit: En.wikipedia.org

Flagging Trends

- We introduce a simple-to-use tool called TrendFlagger that allows a decision maker to get evidence that a trend is appearing – without statistical sophistication.
- We illustrate TrendFlagger with Houston Ship Channel data.
- Illustrations will use source of data that USCG uses to study shipping trends and vessel behavior: The Automatic Identification System (AIS).
  - AIS automatically transmits data identifying a ship, its location, course, destination port, estimated ETA, etc.
  - For over 1,000,000 ships worldwide.
  - Updates as frequently as every 2 seconds while ship is in motion, every 3 minutes while in anchor.
  - Original use: safety (collision avoidance).
  - Now, many national security uses.

Command, Control, and Interoperability
Center for Advanced Data Analysis
A Department of Homeland Security Center of Excellence

The TrendFlagger

- TrendFlagger developed by Paul Kantor.
- Two choices by the decision maker:
  - How long to wait to decide whether there is a trend.
  - How strong an increase or decrease merits attention.
- First choice involves a "window size"
- Second involves two thresholds, one for uptrend and one for downtrend.
- Use a "moving window."
- If latest data point confirms rate above upper target level, color data point **RED**. If below lower target level, color it **YELLOW**. Otherwise, WHITE.

The TrendFlagger

- More precisely: look at change per unit, i.e., slope of line going through the last *n* data points if *n* is window size.
- Thresholds give upward and downward slopes that trigger a **RED** or **YELLOW**.
- This is a decision support tool.
- Gives decision maker a quick, easily visualized tool to:
  - Spot trends
  - Suggest more sophisticated analysis
  - Spur review of policies or resource allocations
- Depends heavily on human intervention and experimentation with choice of the three parameters.

CCICADA
Command, Control, and Interoperability
Center for Advanced Data Analysis
A Department of Homeland Security Center of Excellence

The TrendFlagger

- AIS data covers zones.
- Zone 15 includes Houston Ship Channel.
- Compare total number ships arriving in Zone 15 per month vs. number arriving in Houston.
- First 4 entries pink because window size is 5.
- Last 5 entries in Total Column: slope of regression line = -116.4.
- 116.4 > 42.81, so yellow.

		Total		Houston/Total
	UpTrend	42.81		0.005
	DownTrend	42.81		0.005
Year.Month	AveragingWindow	5		5
13.01	1	1801		0.198778456
13.02	2	1644		0.196
13.03	3	1874		0.191
13.04	4	1818		0.187
13.05	5	1983		0.184
13.06	6	1861		0.202
13.07	7	1931		0.190
13.08	8	1960		0.201
13.09	9	1893		0.192
13.10	10	2070		0.189
13.11	11	1947		0.216
13.12	12	2026		0.194
14.01	1	1915		0.207
14.02	2	1816		0.172
14.03	3	2026		0.187
14.04	4	2075		0.193
14.05	5	1969		0.198
14.06	6	1849		0.206
14.07	7	1918		0.209
14.08	8	1866		0.214
14.09	9	1714		0.235
14.10	10	1820		0.221
14.11	11	422		0.358
14.12	12	1930		0.216
		alpha		0.3

The TrendFlagger

- If December 2014 entry in Total column were 2300, slope would be -42.4. Thus, white.
- If entry were 2700, slope 37.6. White.
- If entry were 2750, slope 47.6. Red.

		Total		Houston/Total
	UpTrend	42.81		0.005
	DownTrend	42.81		0.005
Year.Month	**AveragingWindow**	5		5
13.01	1	1801		0.198778456
13.02	2	1644		0.196
13.03	3	1874		0.191
13.04	4	1818		0.187
13.05	5	1983		0.184
13.06	6	1861		0.202
13.07	7	1931		0.190
13.08	8	1960		0.201
13.09	9	1893		0.192
13.10	10	2070		0.189
13.11	11	1947		0.216
13.12	12	2026		0.194
14.01	1	1915		0.207
14.02	2	1816		0.172
14.03	3	2026		0.187
14.04	4	2075		0.193
14.05	5	1969		0.198
14.06	6	1849		0.206
14.07	7	1918		0.209
14.08	8	1866		0.214
14.09	9	1714		0.235
14.10	10	1820		0.221
14.11	11	422		0.358
14.12	12	1930		0.216
		alpha		0.3

Moving Averages

- TrendFlagger similar to moving averages used to identify trends in stock prices, commodity prices, etc.
- Moving average chooses a window and reports average value (price) over the window.
- Average is “moving” – each time period the oldest value is dropped and newest one added.
- Used to identify uptrends and downtrends.
- No right timeframe to use with moving averages. Literature suggests:
  - Figure out what is best for you.
  - Experiment with different time periods.
- Moving average should not be used alone.
- Use complementary tools.

The TrendFlagger

- This figure shows the total data, but with different parameters in each column.
- In first column, window size has been changed from 5 to 3.
- Slope of regression line through last three points is 55.
- Now December 2014 entry here is red, not yellow.
- Second Total column changes thresholds. Now December 2014 entry is white.

			Total	Total
		UpTrend	42.81	55.26
		DownTrend	42.81	55.26
Year.Month		AveragingWindow	3	3
13.01		1	1801	1801
13.02		2	1644	1644
13.03		3	1874	1874
13.04		4	1818	1818
13.05		5	1983	1983
13.06		6	1861	1861
13.07		7	1931	1931
13.08		8	1960	1960
13.09		9	1893	1893
13.10		10	2070	2070
13.11		11	1947	1947
13.12		12	2026	2026
14.01		1	1915	1915
14.02		2	1816	1816
14.03		3	2026	2026
14.04		4	2075	2075
14.05		5	1969	1969
14.06		6	1849	1849
14.07		7	1918	1918
14.08		8	1866	1866
14.09		9	1714	1714
14.10		10	1820	1820
14.11		11	422	422
14.12		12	1930	1930
	alpha			0.3

The TrendFlagger

- So how does one choose the parameters?
- There is no "right way" to do this.
- Experiment with different values.
- Use knowledge of the subject matter.
- E.g.: if we feel we can handle a 10% increase in tanker traffic, we might use a 10% uptrend threshold.
- We might feel that if there are other important ways to spend our money, we may reduce oil spill if we see a 15% decrease.
- One heuristic: relate thresholds to standard deviation.
  - If not all data is in, use s.d. s of first n data points. Then choose thresholds using $s/\sqrt{n}$.
  - Sometimes multiply this by a parameter α.
    - See earlier figures.
  - If all data is in, use s.d. of all data points.

What Can we Learn from TrendFlagger?

- Comparing trends in two different data sets can be interesting.
- During 2013 after first 4 months, trend in Total is either Red or White.
- Reversal of trend in Feb. 2014, and especially in second half of 2014.

		Total		Houston/Total
	UpTrend	42.81		0.005
	DownTrend	42.81		0.005
Year.Month	AveragingWindow	5		5
13.01	1	1801		0.198778456
13.02	2	1644		0.196
13.03	3	1874		0.191
13.04	4	1818		0.187
13.05	5	1983		0.184
13.06	6	1861		0.202
13.07	7	1931		0.190
13.08	8	1960		0.201
13.09	9	1893		0.192
13.10	10	2070		0.189
13.11	11	1947		0.216
13.12	12	2026		0.194
14.01	1	1915		0.207
14.02	2	1816		0.172
14.03	3	2026		0.187
14.04	4	2075		0.193
14.05	5	1969		0.198
14.06	6	1849		0.206
14.07	7	1918		0.209
14.08	8	1866		0.214
14.09	9	1714		0.235
14.10	10	1820		0.221
14.11	11	422		0.358
14.12	12	1930		0.216
		alpha		0.3

What Can we Learn from TrendFlagger?

- During 2013, in months when Total is showing upward trend, proportion going to Houston shows no trend.
- In latter half of 2014, Total shows downtrend, while proportion to Houston uptrend.
- TrendFlagger calls attention to these "opposites;" that calls for explanation.

		Total		Houston/Total
	UpTrend	42.81		0.005
	DownTrend	42.81		0.005
Year.Month	AveragingWindow	5		5
13.01	1	1801		0.198778456
13.02	2	1644		0.196
13.03	3	1874		0.191
13.04	4	1818		0.187
13.05	5	1983		0.184
13.06	6	1861		0.202
13.07	7	1931		0.190
13.08	8	1960		0.201
13.09	9	1893		0.192
13.10	10	2070		0.189
13.11	11	1947		0.216
13.12	12	2026		0.194
14.01	1	1915		0.207
14.02	2	1816		0.172
14.03	3	2026		0.187
14.04	4	2075		0.193
14.05	5	1969		0.198
14.06	6	1849		0.206
14.07	7	1918		0.209
14.08	8	1866		0.214
14.09	9	1714		0.235
14.10	10	1820		0.221
14.11	11	422		0.358
14.12	12	1930		0.216
		alpha		0.3

What Can we Learn from TrendFlagger?

- AIS code xy for vessel
- x = vessel type, y = cargo type
- x = 8 for tanker
- 1 = hazard-ous cargo

		Total	Total_80s	Total_81	Hous	Hous_80s	Hous_81
	UpTrend	25.01	8.47	1.61	12.49	6.44	0.99
	DownTrend	25.01	8.47	1.61	12.49	6.44	0.99
Year.Month	AveragingWindow	3	3	3	3	3	3
13.01	1	1801	528	46	358	163	12
13.02	2	1644	488	44	323	149	13
13.03	3	1874	538	53	358	150	15
13.04	4	1818	541	52	340	156	13
13.05	5	1983	567	57	365	176	17
13.06	6	1861	527	46	375	155	19
13.07	7	1931	561	49	366	162	17
13.08	8	1960	566	61	393	170	24
13.09	9	1893	553	51	364	163	16
13.10	10	2070	590	50	392	180	15
13.11	11	1947	586	45	421	217	19
13.12	12	2026	626	57	393	195	15
14.01	1	1915	566	57	396	201	20
14.02	2	1816	512	39	312	138	8
14.03	3	2026	575	50	378	175	16
14.04	4	2075	595	56	401	181	22
14.05	5	1969	568	57	390	194	20
14.06	6	1849	528	39	381	172	17
14.07	7	1918	561	47	401	174	13
14.08	8	1866	564	39	399	190	16
14.09	9	1714	483	39	402	173	14
14.10	10	1820	488	40	403	189	19
14.11	11	422	154	18	151	74	12
14.12	12	1930	580	50	417	198	28
alpha	0.4						

What Can we Learn from TrendFlagger?

- Total hazardous tanker traffic can have opposite trend from Total traffic
- Total tanker traffic never has opposite trend from Total traffic
- No opposite trends in Total and hazardous heading to Houston

		Total	Total_80s	Total_81	Hous	Hous_80s	Hous_81
	UpTrend	25.01	8.47	1.61	12.49	6.44	0.99
	DownTrend	25.01	8.47	1.61	12.49	6.44	0.99
Year.Month	AveragingWindow	3	3	3	3	3	3
13.01	1	1801	528	46	358	163	12
13.02	2	1644	488	44	323	149	13
13.03	3	1874	538	53	358	150	15
13.04	4	1818	541	52	340	156	13
13.05	5	1983	567	57	365	176	17
13.06	6	1861	527	46	375	155	19
13.07	7	1931	561	49	366	162	17
13.08	8	1960	566	61	393	170	24
13.09	9	1893	553	51	364	163	16
13.10	10	2070	590	50	392	180	15
13.11	11	1947	586	45	421	217	19
13.12	12	2026	626	57	393	195	15
14.01	1	1915	566	57	396	201	20
14.02	2	1816	512	39	312	138	8
14.03	3	2026	575	50	378	175	16
14.04	4	2075	595	56	401	181	22
14.05	5	1969	568	57	390	194	20
14.06	6	1849	528	39	381	172	17
14.07	7	1918	561	47	401	174	13
14.08	8	1866	564	39	399	190	16
14.09	9	1714	483	39	402	173	14
14.10	10	1820	488	40	403	189	19
14.11	11	422	154	18	151	74	12
14.12	12	1930	580	50	417	198	28
alpha	0.4						

Learning from Anomalies

- Look for situations where expect similar trends.
- Consider Total arrivals in Texas City in 2013 vs. 2014.
- Nov. 2013 upward, Nov. 2014 downward.
- Similar anomaly in same month if consider all tankers coming to Texas City.
- This led us to observe that Total and tanker traffic to Texas City is MUCH lower in Nov. 2014 than Nov. 2013. Why?

		Texas City 2013	Texas City 2014
	UpTrend	3.5	3.5
	DownTrend	3.5	3.5
Month	AveragingWindow	4	4
1		32	27
2		22	21
3		26	26
4		25	21
5		19	18
6		21	21
7		26	26
8		25	25
9		31	31
10		31	21
11		38	11
12		33	23

		Texas City 80s 2013	Texas City 80s 2014
	UpTrend	2	2
	DownTrend	2	2
Month	AveragingWindow	3	3
1		25	17
2		10	11
3		20	17
4		20	11
5		11	13
6		13	13
7		17	22
8		15	16
9		19	23
10		20	16
11		29	9
12		25	15

What Can we Learn from TrendFlagger?

- This led us to look at the Total and the Houston entries for Nov. 2014.
- See they too are MUCH lower than other entries.

		Total	Total_80s	Total_81		Hous	Hous_80s	Hous_81
	UpTrend	25.01	8.47	1.61		12.49	6.44	0.99
	DownTrend	25.01	8.47	1.61		12.49	6.44	0.99
Year.Month	AveragingWindow	3	3	3		3	3	3
13.01	1	1801	528	46		358	163	12
13.02	2	1644	488	44		323	149	13
13.03	3	1874	538	53		358	150	15
13.04	4	1818	541	52		340	156	13
13.05	5	1983	567	57		365	176	17
13.06	6	1861	527	46		375	155	19
13.07	7	1931	561	49		366	162	17
13.08	8	1960	566	61		393	170	24
13.09	9	1893	553	51		364	163	16
13.10	10	2070	590	50		392	180	15
13.11	11	1947	586	45		421	217	19
13.12	12	2026	626	57		393	195	15
14.01	1	1915	566	57		396	201	20
14.02	2	1816	512	39		312	138	8
14.03	3	2026	575	50		378	175	16
14.04	4	2075	595	56		401	181	22
14.05	5	1969	568	57		390	194	20
14.06	6	1849	528	39		381	172	17
14.07	7	1918	561	47		401	174	13
14.08	8	1866	564	39		399	190	16
14.09	9	1714	483	39		402	173	14
14.10	10	1820	488	40		403	189	19
14.11	11	422	154	18		151	74	12
14.12	12	1930	580	50		417	198	28
alpha	0.4							

What Can we Learn from TrendFlagger?

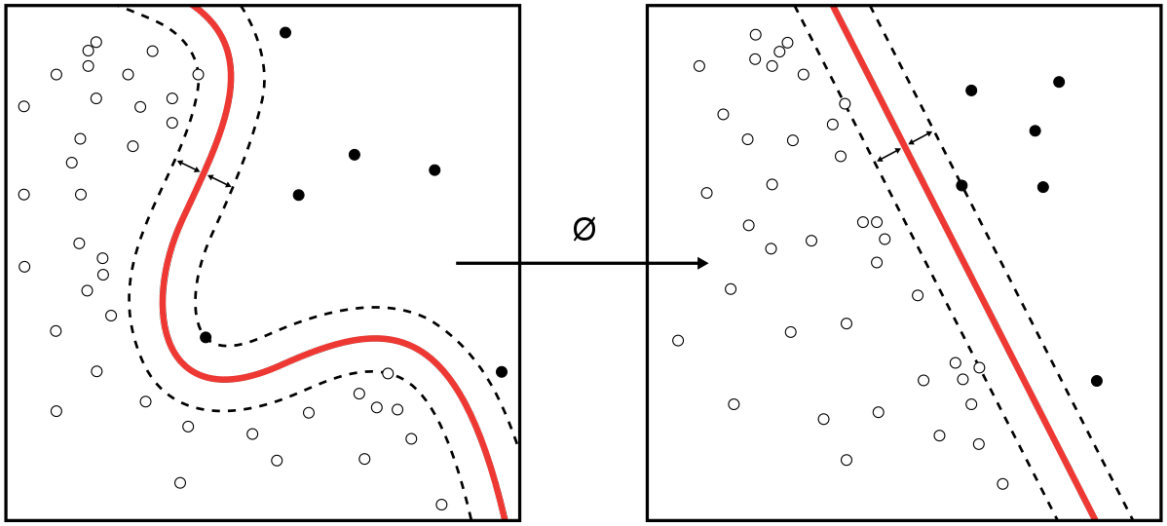
- This led us to discover that magnitude of AIS data for Nov. 2014 was lower for various zones.
- We found that there was an AIS tracking system update in Nov. 2014.
- That probably makes all Nov. 2014 data suspect.
- TrendFlagger led us to discover this problem.

Credit: En.wikipedia.org

Command, Control, and Interoperability
Center for Advanced Data Analysis
A Department of Homeland Security Center of Excellence

Looking for Explanations

- When TrendFlagger highlights an upward or downward trend, we want to look further.
- One approach: statistical tests, e.g., multiple time series exhibiting trends.
- Trend flagging will require development of new statistical methods to provide early warning of changed shipping patterns.
- E.g., variants of sequential change detection methods.
- Alternatively, we can look for "explanations" for changes in data.



Credit: En.wikipedia.org

Looking for Explanations

- See uptrend in Zone 15 Total in April 2014. Why?
- We found that in late March 2014, there was an oil spill in the region.
- Likely that some vessels were delayed, leading to uptrend in April that wasn't seasonal.
- See downtrend in Total in June 2014. Why?
- We found that in May 2014, there was major flooding in region.
- Likely impacted downtrend in vessel traffic into June.
- TrendFlagger led to need to understand what happened.

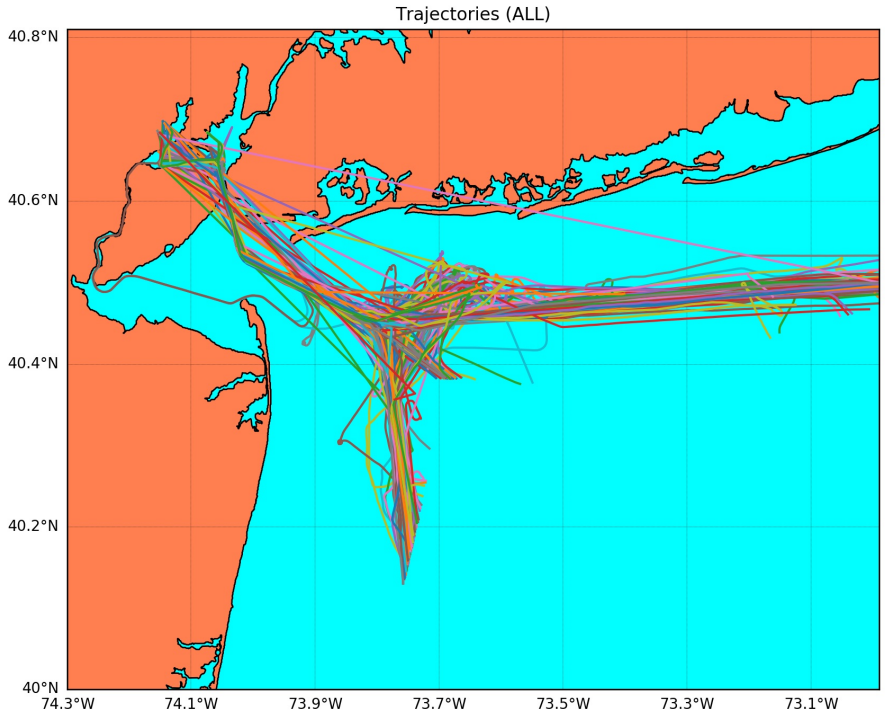
Year.Month		Total
	UpTrend	108
	DownTrend	108
	WindowSize	3
13.01		1801
13.02		1644
13.03		1874
13.04		1818
13.05		1983
13.06		1861
13.07		1931
13.08		1960
13.09		1893
13.10		2070
13.11		1947
13.12		2026
14.01		1915
14.02		1816
14.03		2026
14.04		2075
14.05		1969
14.06		1849
14.07		1918
14.08		1866
14.09		1714
14.10		1820

Aside: i-Group and i-Detect

- In another project, we are developing two statistical tools that use AIS data that can help with anomalies.
  - **i-Group** (individualized group learning to group similar vessels)
  - **i-Detect** (individualized detection to identify vessels deviating from the normal in their group)
- These methods give early warning of abnormalities for vessels.
- i-Group focuses on each individual vessel and forms one individualized group for each vessel, by locating vessels that share similar characteristics.
- **i-Detect**: detect outliers based on a vessel's own individualized baseline distribution

Aside: i-Group and i-Detect

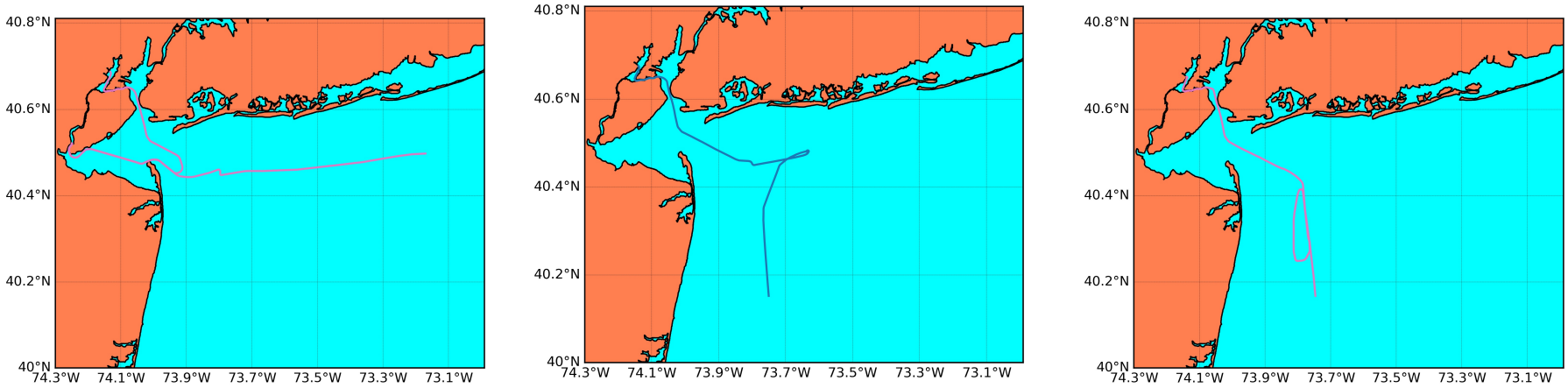
- We focused on 534 vessels/ voyages (tankers, cargo vessels) arriving in Port of Newark between July and November 2014
- Investigated behaviors starting from crossing the 12 nautical mile US territorial sea (TS) boundary to arrival
- The trajectory, a functional feature, is used as the standard feature for i-Group
- Outliers in duration (time spent from TS boundary to port) are detected (based on the two standard deviation rule)



Command, Control, and Interoperability Center for Advanced Data Analysis
A Department of Homeland Security Center of Excellence

Aside: i-Group and i-Detect

- Looked for outliers that have abnormal time duration (from territorial sea boundary to Port) compared to vessels in their clique (vessels with similar trajectory)
- Outliers were detected by the 2 standard deviation rule: Vessels in a clique with time duration at least two s.d. from the clique mean
- 95 detected outliers: (a) 50 vessels had a prior dock before the Port of Newark (left); (b) 18 vessels were anchored somewhere outside the port for an extremely long time (middle); (c) the other 27 vessels were traveling too fast/slow compared with their cliques (right)



Future Directions

- Allow for monthly variations in traffic patterns. Automatically compare data in same month in earlier year.
- TrendFlagger weighs all data points equally. Can we have more recent data have greater influence on conclusions?
  - Compare Exponential Moving Average
- Choice of parameters in TrendFlagger should be primarily dependent upon experience and goals of the user. But could there be both theory and heuristics for parameter choice? (E.g., $s/\sqrt{n}$ or $\alpha s/\sqrt{n}$)
- Trend flagging will require development of new statistical methods including variants of sequential change point detection methods and streaming algorithms suited to particular spatio-temporal characteristics relevant to the application area being studied.

Thanks to my Collaborators and Co-Authors

Dennis Egan

Paul Kantor

Joonhee Lee

Christie Nelson

Helen Roberts

Command, Control, and Interoperability
Center for Advanced Data Analysis
A Department of Homeland Security Center of Excellence

Trend Flagging to Aid Resource Allocation Decisions

For More Information:

Fred Roberts
froberts@dimacs.rutgers.edu

CCICADA Center
www.ccicada.org

Command, Control, and Interoperability
Center for Advanced Data Analysis
A Department of Homeland Security Center of Excellence