

This report is prepared based on Walkerton Hospital Admissions data. It is part of the ECADS project led by Dr. Richard Davies (rfdavies@ottawaheart.ca). We investigate the extent to which the hospital admissions data reveals the effects of the Walkerton Crypto outbreak in mid/late May 2000.

Early Analysis of Walkerton Data

Andrew Moore	Professor	Computer Science, Carnegie Mellon	awm@cs.cmu.edu
Daniel Neill	Doctoral Graduate Student	Computer Science, Carnegie Mellon	neill@cs.cmu.edu
Robin Sabhnani	Senior Research Programmer and Masters Student	Computer Science, Carnegie Mellon	sabhnani+@cs.cmu.edu

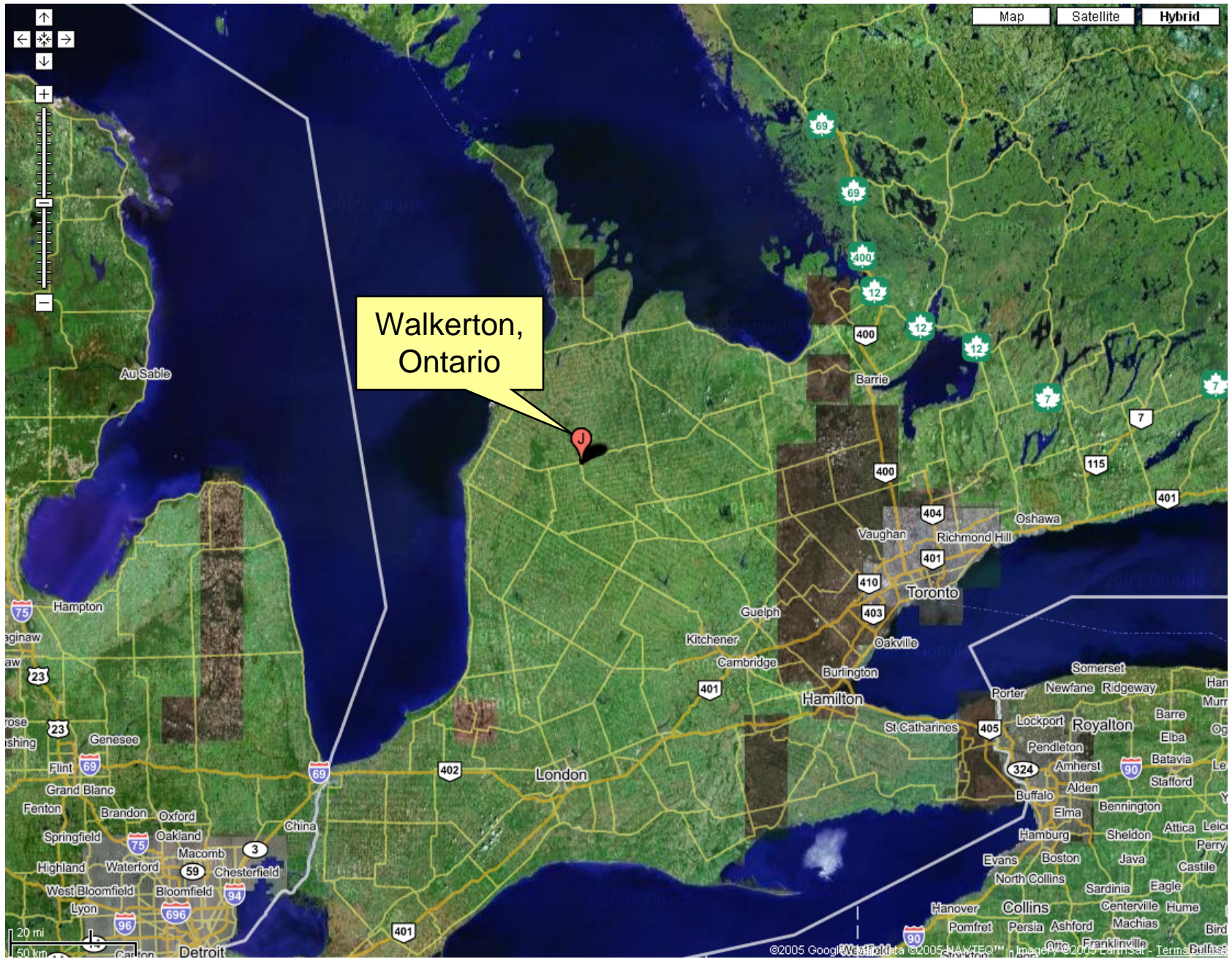
Version 11, June 18th, 2005

Note to other teachers and users of these slides. Andrew would be delighted if you found this source material useful in giving your own lectures. Feel free to use these slides verbatim, or to modify them to fit your own needs. PowerPoint originals are available. If you make use of a significant portion of these slides in your own lecture, please include this message, or the following link to the source repository of Andrew's tutorials: <http://www.cs.cmu.edu/~awm/tutorials> . Comments and corrections gratefully received.



Auton Lab: <http://www.autonlab.org>





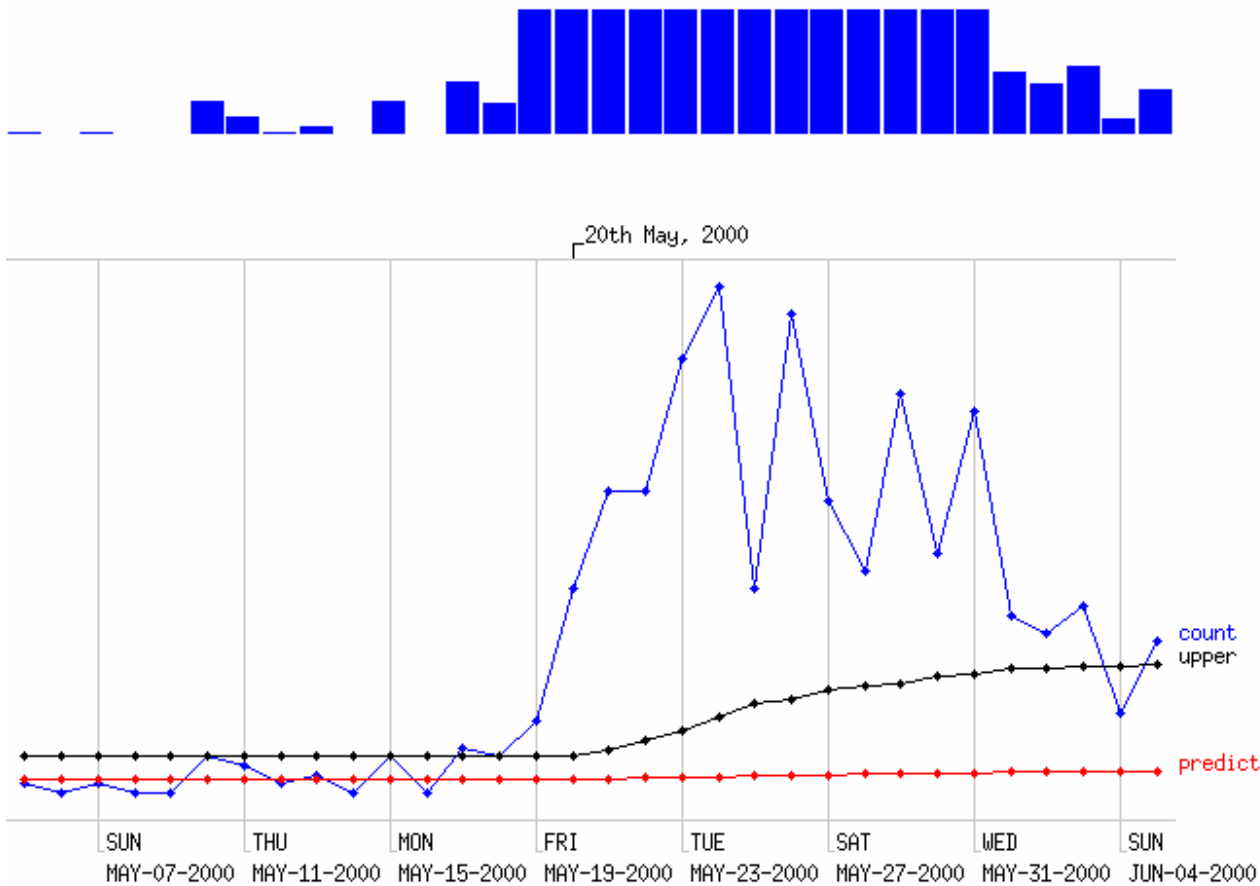


Walkerton,
Ontario

Hanover,
Ontario

Data	Tracking	Method
3 year data (excludes Hanover)	GI visits from city of Walkerton	Standard Control Chart

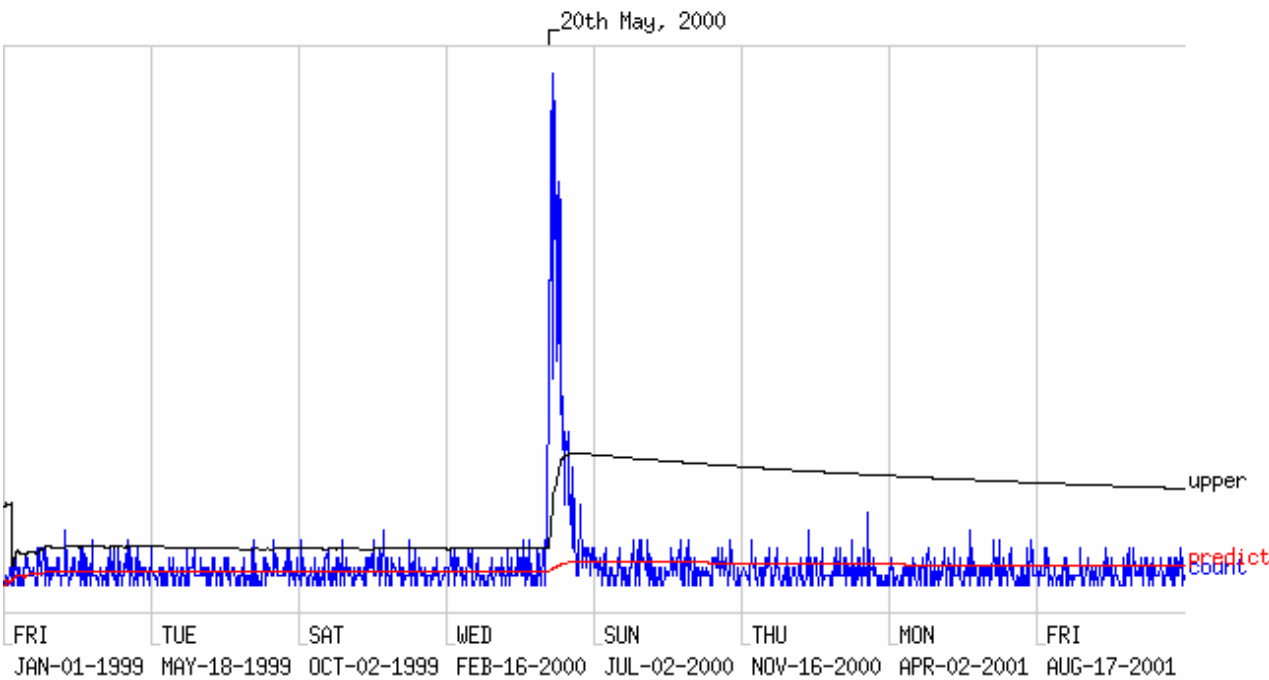
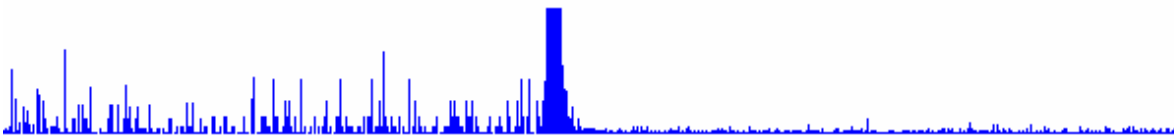
Bars show alarm levels: max = 10



- <---jump
- <---step
- jump---
- step---
- zoomout
- print
- quit

Data	Tracking	Method
3 year data (excludes Hanover)	GI visits from city of Walkerton	Standard Control Chart

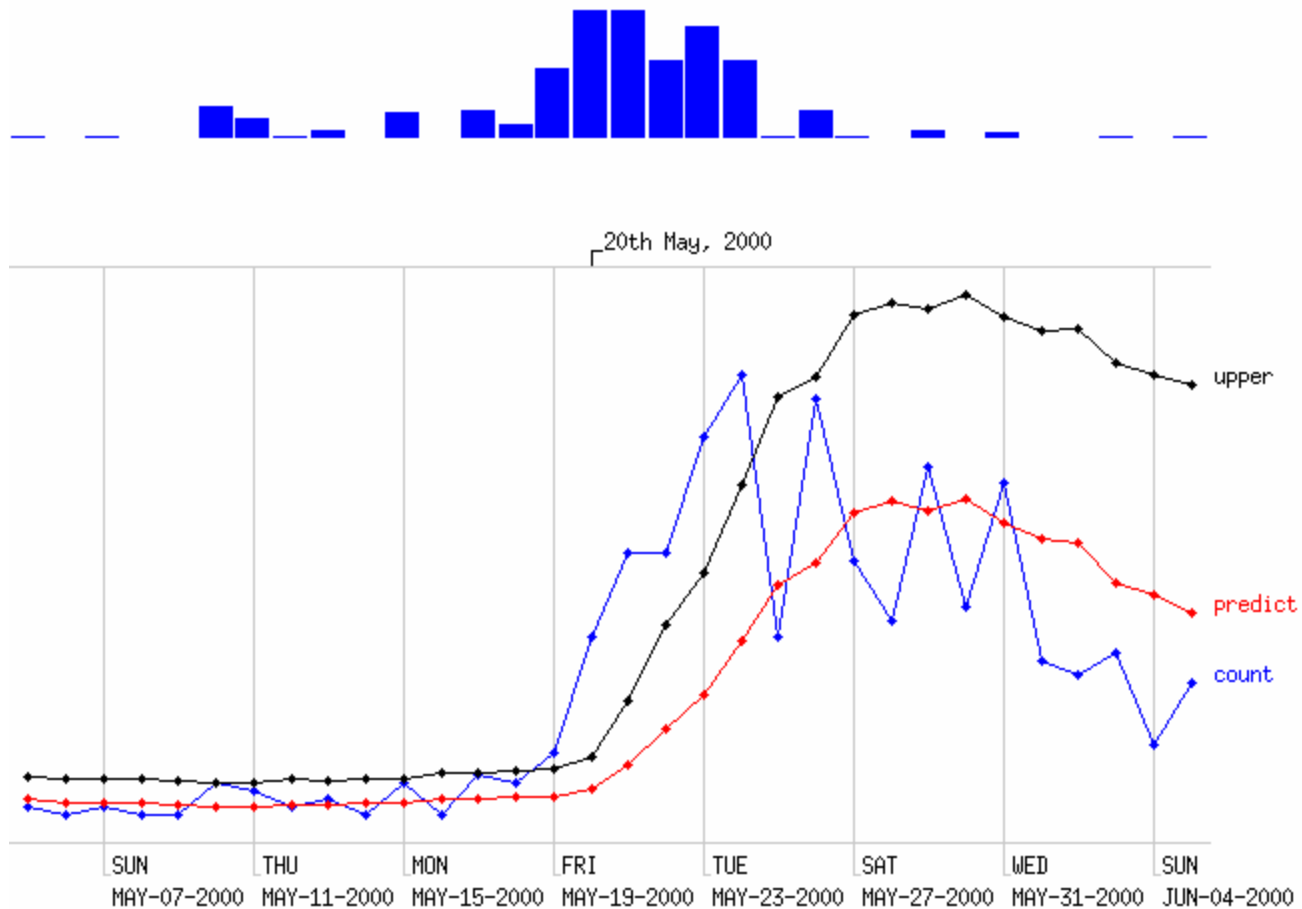
Bars show alarm levels; max = 10



- <---jump
- <---step
- jump---
- step---
- zoomout
- print
- quit

Data	Tracking	Method
3 year data (excludes Hanover)	GI visits from city of Walkerton	Moving Average (7 days)

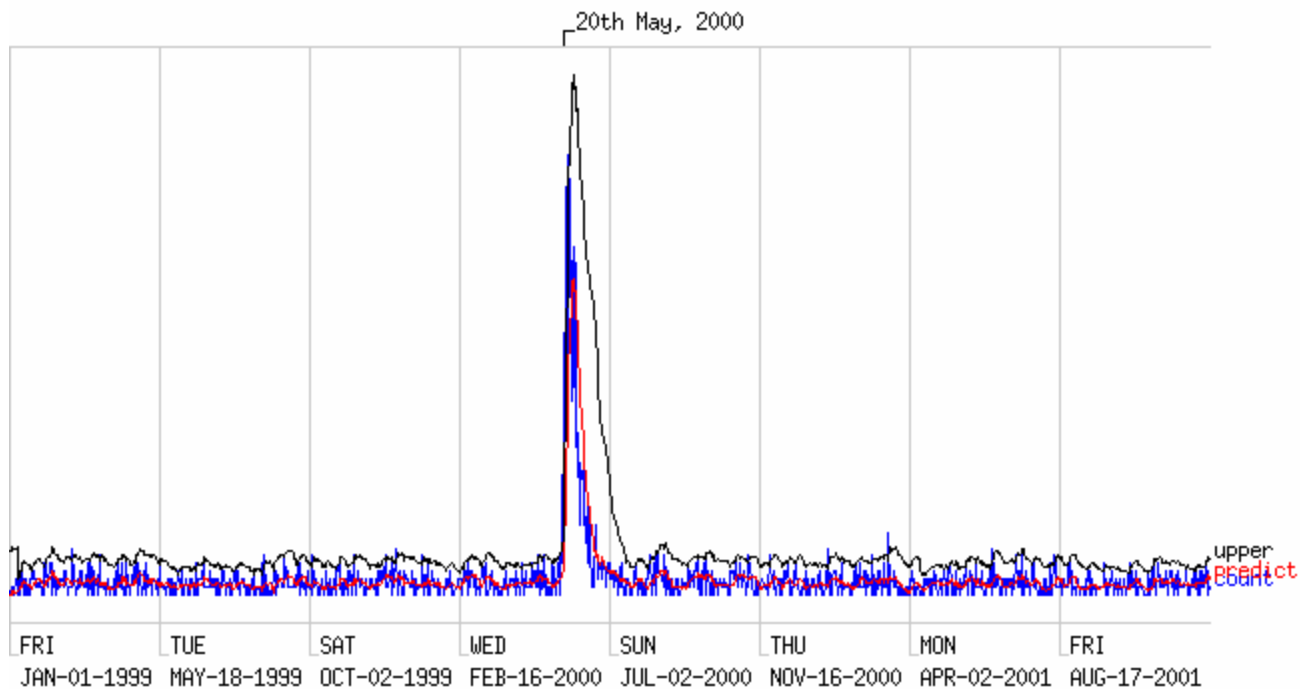
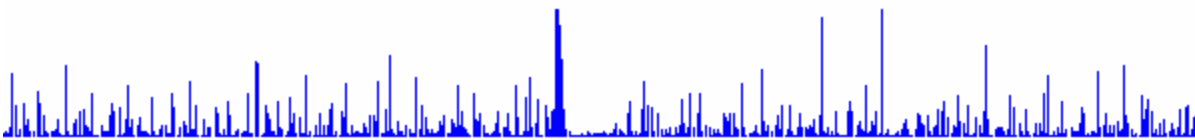
Bars show alarm levels; max = 10



- <---jump
- <---step
- jump---
- step---
- zoomout
- print
- quit

Data	Tracking	Method
3 year data (excludes Hanover)	GI visits from city of Walkerton	Moving Average (7 days)

Bars show alarm levels; max = 10



<---jump

<---step

jump--->

step--->

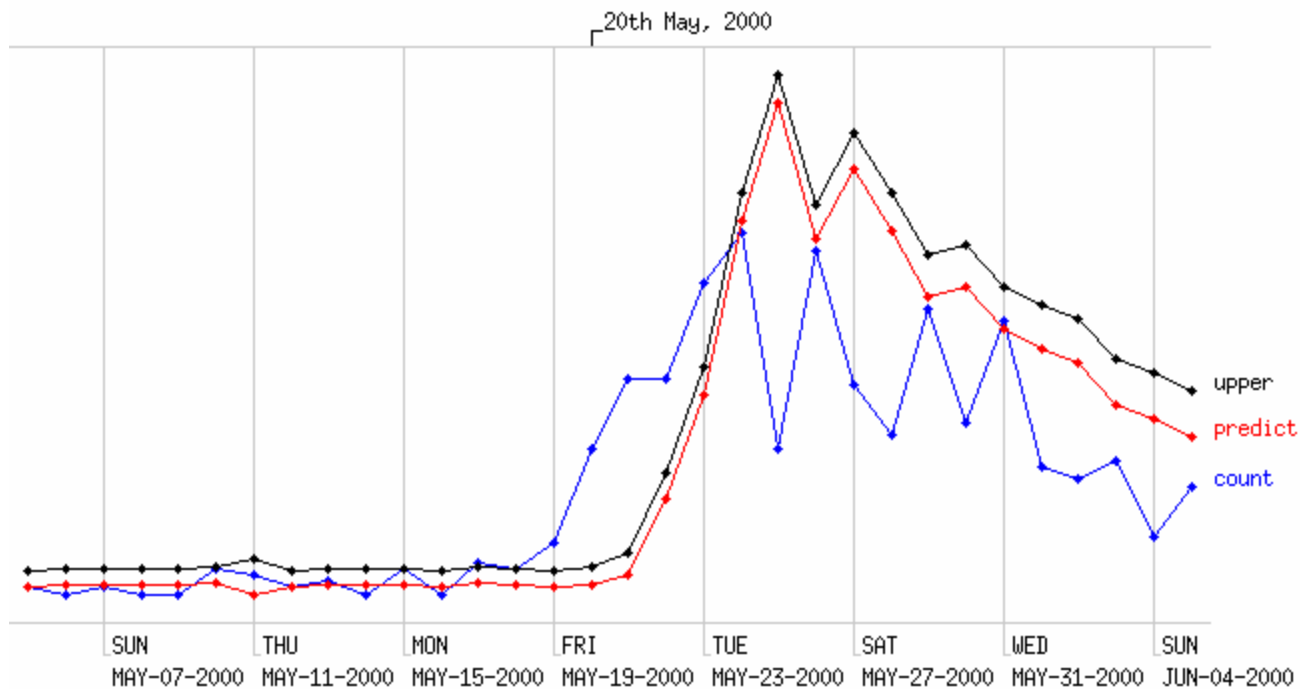
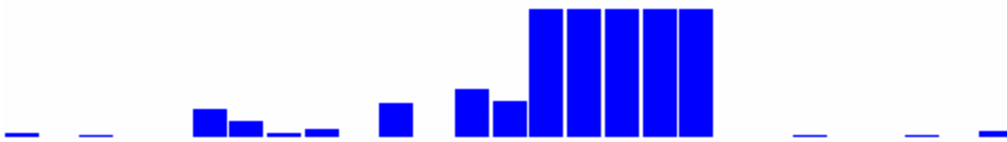
zoomout

print

quit

Data	Tracking	Method
3 year data (excludes Hanover)	GI visits from city of Walkerton	Regression w/ DOW + Season

Bars show alarm levels; max = 10



<---jump

<---step

jump--->

step--->

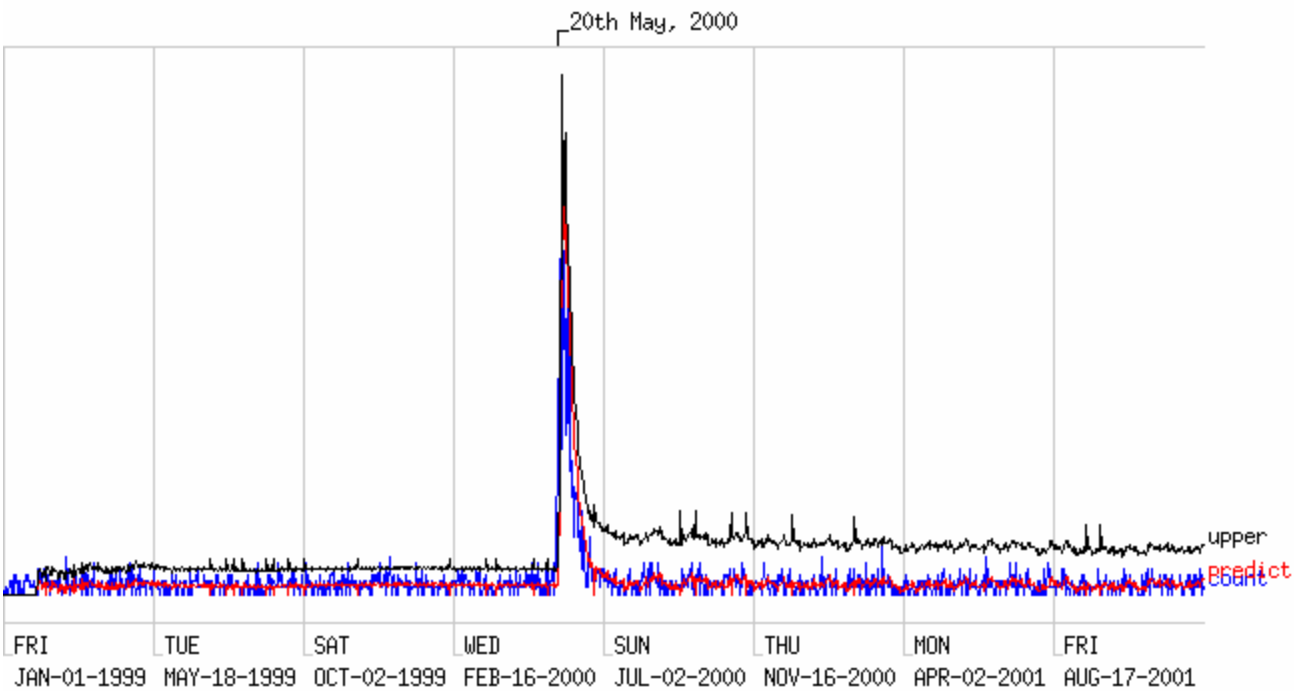
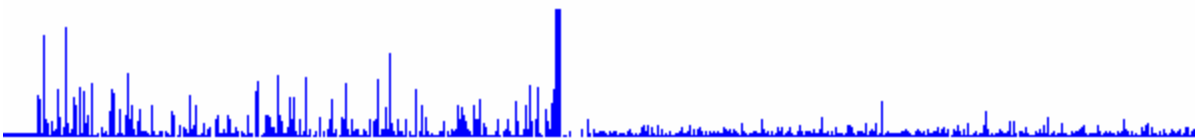
zoomout

print

quit

Data	Tracking	Method
3 year data (excludes Hanover)	GI visits from city of Walkerton	Regression w/ DOW + Season

Bars show alarm levels; max = 10



- <---jump
- <---step
- jump---
- step---
- zoomout
- print
- quit

Data	Tracking
3 year data (excludes Hanover)	GI visits from city of Walkerton

Performance of Univariate Algorithms: May 20th 2000

(False Positive Rate = Fraction of the days outside interval May 5-June 15 2000 in which there would have been an alarm, if the threshold was set such that the given date was detected)

```

cc      0.0% Standard Control Chart
ma1     0.0% Yesterday
ma3     0.0% Moving Average (3-day window)
ma7     0.1% Moving Average (7-day window)
ma14    0.1% Moving Average (14-day window)
ma28    0.1% Moving Average (28-day window)
regh    0.0% Regression (Hours of Daylight)
regm    0.0% Regression (HOD + Monday)
regtu   0.1% Regression (HOD + Monday + Tuesday)
regth   0.1% Regression (HOD + Monday + ... Thursday + Mean of Last Week)
regs    0.0% Regression (HOD + Monday + ... Saturday + Mean of Last Week)
regs84  0.1% Regression HOD + M..S + LastWeek learning from last 3 months
regs28  0.8% Regression HOD + M..S + LastWeek learning from last 4 weeks
regsh   0.0% Regression HOD + M..S + LastWeek avoiding times near holidays
cusum1  0.0% CUSUM With H = 1
cusum2  0.0% CUSUM With H = 2
cusum5  1.0% CUSUM With H = 5
cusum10 1.4% CUSUM With H = 10
cusum20 4.1% CUSUM With H = 20
sa1     0.3% Sickness/Availability with Window of 1 day
sa3     0.4% Sickness/Availability with Window of 3 days
sa7     0.5% Sickness/Availability with Window of 7 days
sa14    0.5% Sickness/Availability with Window of 14 days
sa28    0.5% Sickness/Availability with Window of 28 days

```

Data	Tracking
3 year data (excludes Hanover)	GI visits from city of Walkerton

Performance of Univariate Algorithms: May 19th 2000

(False Positive Rate = Fraction of the days outside interval May 5-June 15 2000 in which there would have been an alarm, if the threshold was set such that the given date was detected)

cc	0.0%	Standard Control Chart
ma1	2.1%	Yesterday
ma3	2.3%	Moving Average (3-day window)
ma7	0.8%	Moving Average (7-day window)
ma14	0.3%	Moving Average (14-day window)
ma28	0.5%	Moving Average (28-day window)
regh	0.0%	Regression (Hours of Daylight)
regm	0.0%	Regression (HOD + Monday)
regtu	0.1%	Regression (HOD + Monday + Tuesday)
regth	0.1%	Regression (HOD + Monday + ... Thursday + Mean of Last Week)
regs	0.0%	Regression (HOD + Monday + ... Saturday + Mean of Last Week)
regs84	0.1%	Regression HOD + M..S + LastWeek learning from last 3 months
regs28	0.8%	Regression HOD + M..S + LastWeek learning from last 4 weeks
regsh	0.0%	Regression HOD + M..S + LastWeek avoiding times near holidays
cusum1	0.2%	CUSUM With H = 1
cusum2	0.0%	CUSUM With H = 2
cusum5	4.4%	CUSUM With H = 5
cusum10	1.6%	CUSUM With H = 10
cusum20	1.4%	CUSUM With H = 20
sa1	0.3%	Sickness/Availability with Window of 1 day
sa3	0.4%	Sickness/Availability with Window of 3 days
sa7	0.5%	Sickness/Availability with Window of 7 days
sa14	0.5%	Sickness/Availability with Window of 14 days
sa28	0.5%	Sickness/Availability with Window of 28 days

Data	Tracking
3 year data (excludes Hanover)	GI visits from city of Walkerton

Performance of Univariate Algorithms: May 18th 2000

(False Positive Rate = Fraction of the days outside interval May 5-June 15 2000 in which there would have been an alarm, if the threshold was set such that the given date was detected)

cc	2.7%	Standard Control Chart
ma1	72.5%	Yesterday
ma3	27.7%	Moving Average (3-day window)
ma7	14.9%	Moving Average (7-day window)
ma14	11.2%	Moving Average (14-day window)
ma28	13.1%	Moving Average (28-day window)
regh	2.5%	Regression (Hours of Daylight)
regm	2.6%	Regression (HOD + Monday)
regtu	2.7%	Regression (HOD + Monday + Tuesday)
regth	2.5%	Regression (HOD + Monday + ... Thursday + Mean of Last Week)
regs	2.8%	Regression (HOD + Monday + ... Saturday + Mean of Last Week)
regs84	7.4%	Regression HOD + M..S + LastWeek learning from last 3 months
regs28	2.8%	Regression HOD + M..S + LastWeek learning from last 4 weeks
regsh	2.6%	Regression HOD + M..S + LastWeek avoiding times near holidays
cusum1	7.2%	CUSUM With H = 1
cusum2	11.2%	CUSUM With H = 2
cusum5	2.0%	CUSUM With H = 5
cusum10	2.4%	CUSUM With H = 10
cusum20	1.6%	CUSUM With H = 20
sa1	29.0%	Sickness/Availability with Window of 1 day
sa3	20.3%	Sickness/Availability with Window of 3 days
sa7	15.2%	Sickness/Availability with Window of 7 days
sa14	9.4%	Sickness/Availability with Window of 14 days
sa28	12.1%	Sickness/Availability with Window of 28 days

Data	Tracking
3 year data (excludes Hanover)	GI visits from city of Walkerton

Performance of Univariate Algorithms: May 17th 2000

(False Positive Rate = Fraction of the days outside interval May 5-June 15 2000 in which there would have been an alarm, if the threshold was set such that the given date was detected)

cc	1.0%	Standard Control Chart
ma1	0.7%	Yesterday
ma3	5.1%	Moving Average (3-day window)
ma7	6.7%	Moving Average (7-day window)
ma14	3.9%	Moving Average (14-day window)
ma28	4.6%	Moving Average (28-day window)
regh	0.8%	Regression (Hours of Daylight)
regm	0.7%	Regression (HOD + Monday)
regtu	0.8%	Regression (HOD + Monday + Tuesday)
regth	1.2%	Regression (HOD + Monday + ... Thursday + Mean of Last Week)
regs	1.4%	Regression (HOD + Monday + ... Saturday + Mean of Last Week)
regs84	6.5%	Regression HOD + M..S + LastWeek learning from last 3 months
regs28	14.7%	Regression HOD + M..S + LastWeek learning from last 4 weeks
regsh	1.2%	Regression HOD + M..S + LastWeek avoiding times near holidays
cusum1	2.3%	CUSUM With H = 1
cusum2	1.7%	CUSUM With H = 2
cusum5	4.4%	CUSUM With H = 5
cusum10	5.3%	CUSUM With H = 10
cusum20	3.2%	CUSUM With H = 20
sa1	5.4%	Sickness/Availability with Window of 1 day
sa3	16.3%	Sickness/Availability with Window of 3 days
sa7	21.1%	Sickness/Availability with Window of 7 days
sa14	15.8%	Sickness/Availability with Window of 14 days
sa28	17.7%	Sickness/Availability with Window of 28 days

Data	Tracking
3 year data (excludes Hanover)	GI visits from city of Walkerton

Performance of Univariate Algorithms: May 16th 2000

(False Positive Rate = Fraction of the days outside interval May 5-June 15 2000 in which there would have been an alarm, if the threshold was set such that the given date was detected)

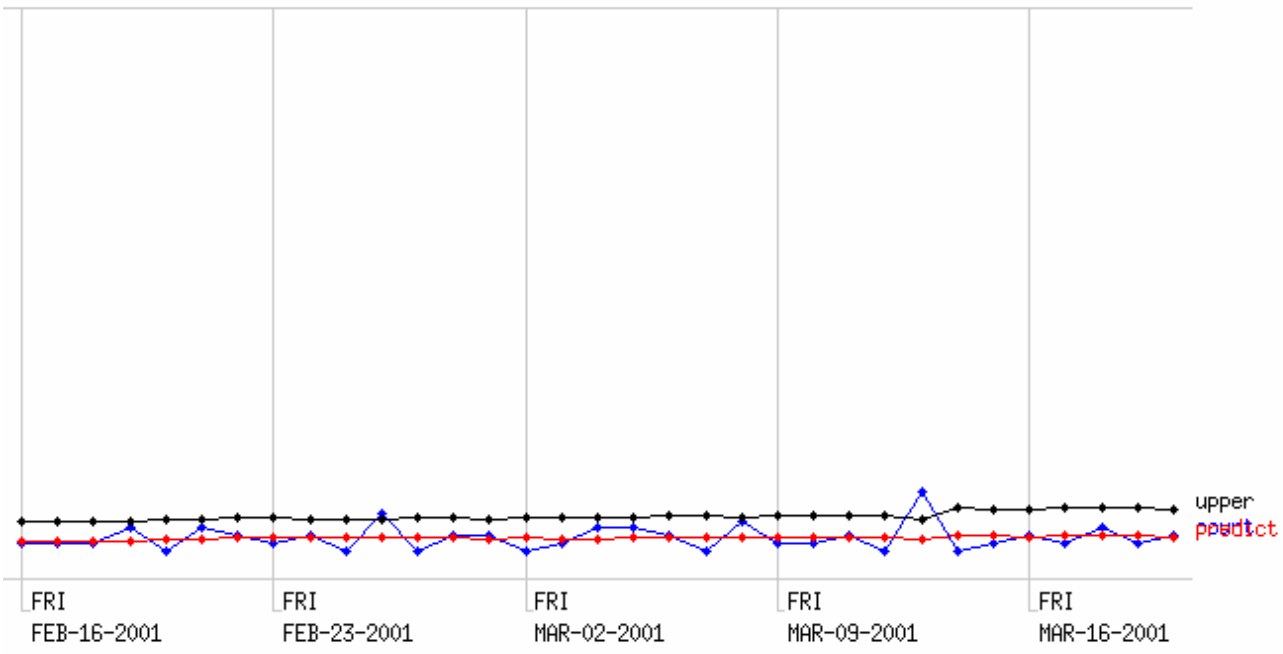
```

cc      97.9% Standard Control Chart
ma1     98.3% Yesterday
ma3     87.1% Moving Average (3-day window)
ma7     91.2% Moving Average (7-day window)
ma14    83.0% Moving Average (14-day window)
ma28    88.2% Moving Average (28-day window)
regh    94.3% Regression (Hours of Daylight)
regm    93.2% Regression (HOD + Monday)
regtu   90.9% Regression (HOD + Monday + Tuesday)
regth   89.7% Regression (HOD + Monday + ... Thursday + Mean of Last Week)
regs    91.1% Regression (HOD + Monday + ... Saturday + Mean of Last Week)
regs84  85.4% Regression HOD + M..S + LastWeek learning from last 3 months
regs28  53.1% Regression HOD + M..S + LastWeek learning from last 4 weeks
regsh   91.9% Regression HOD + M..S + LastWeek avoiding times near holidays
cusum1  100.0% CUSUM With H = 1
cusum2  29.2% CUSUM With H = 2
cusum5  23.4% CUSUM With H = 5
cusum10 15.3% CUSUM With H = 10
cusum20  8.5% CUSUM With H = 20
sa1     97.3% Sickness/Availability with Window of 1 day
sa3     92.9% Sickness/Availability with Window of 3 days
sa7     94.6% Sickness/Availability with Window of 7 days
sa14    87.7% Sickness/Availability with Window of 14 days
sa28    95.6% Sickness/Availability with Window of 28 days

```

The biggest Walkerton GI blip outside the outbreak period

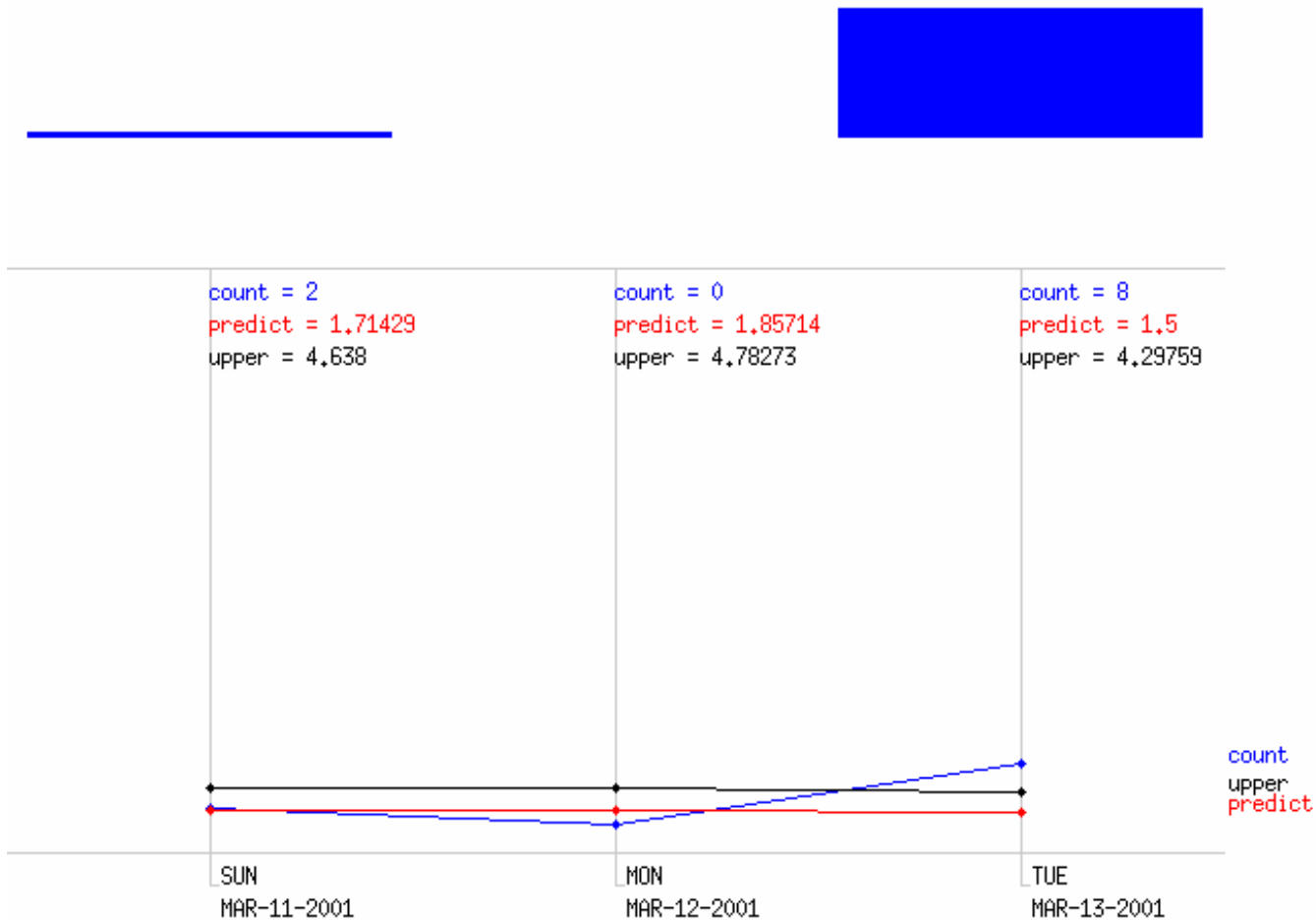
Bars show alarm levels; max = 10



- <---jump
- <---step
- jump--->
- step--->
- zoomout
- print
- quit

The biggest Walkerton GI blip outside the outbreak period

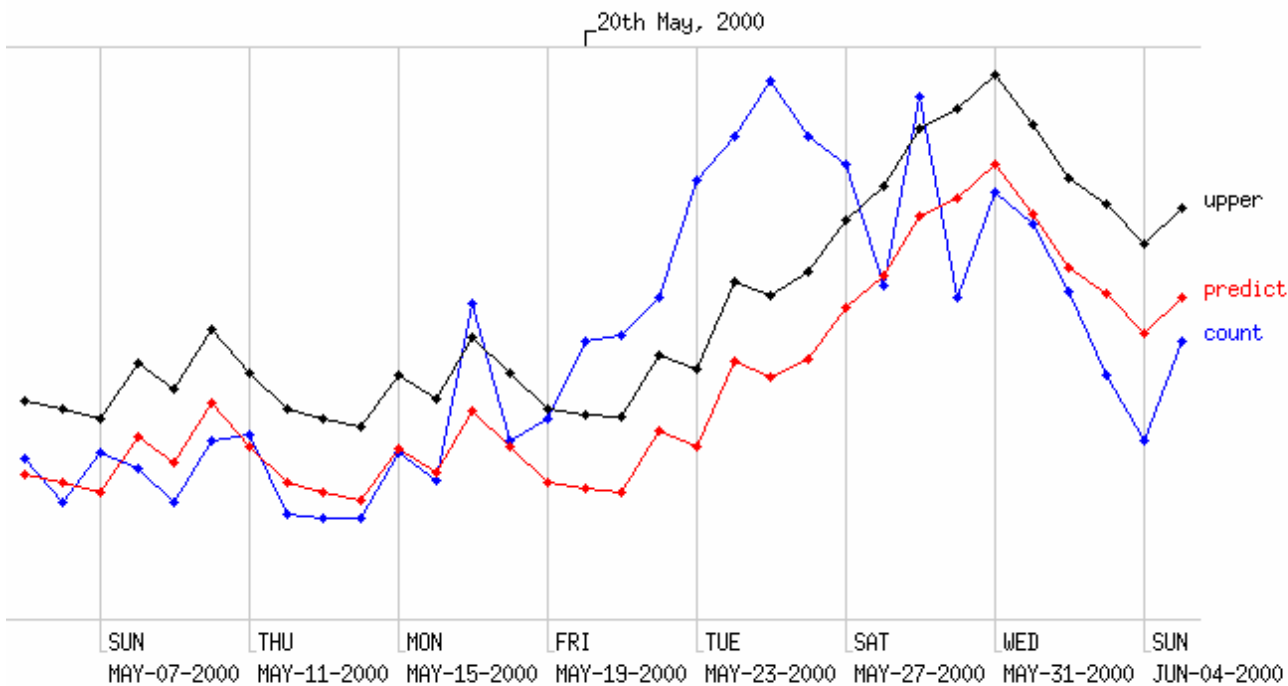
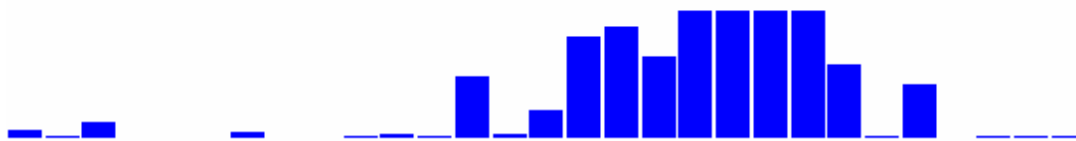
Bars show alarm levels: max = 10



- <---jump
- <---step
- jump---
- step---
- zoomout
- print
- quit

Data	Tracking	Method
3 year data (excludes Hanover)	Walkerton visits (all patients from Walkerton)	Regression w/ DOW + Season

Bars show alarm levels; max = 10



<---jump

<---step

jump--->

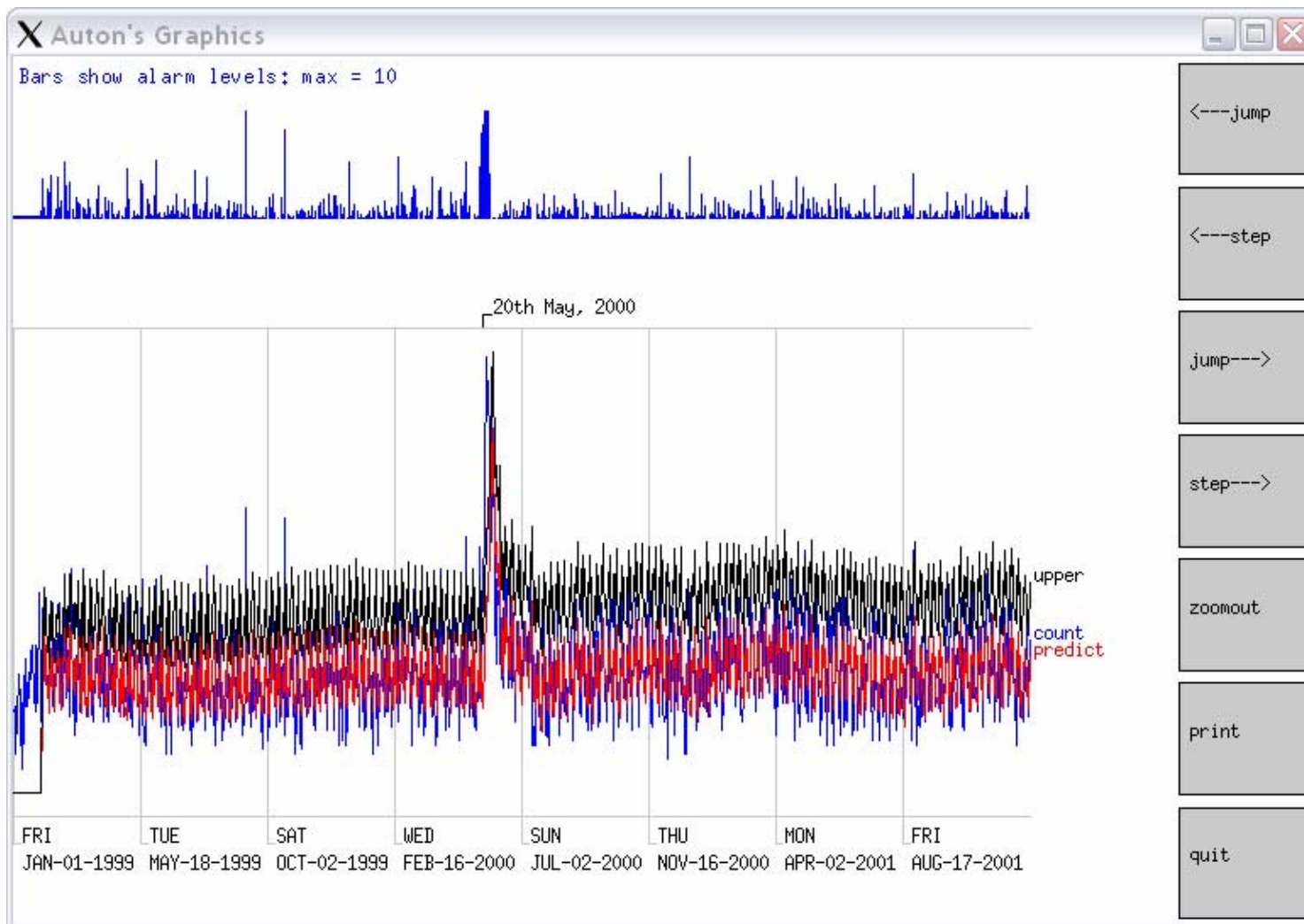
step--->

zoomout

print

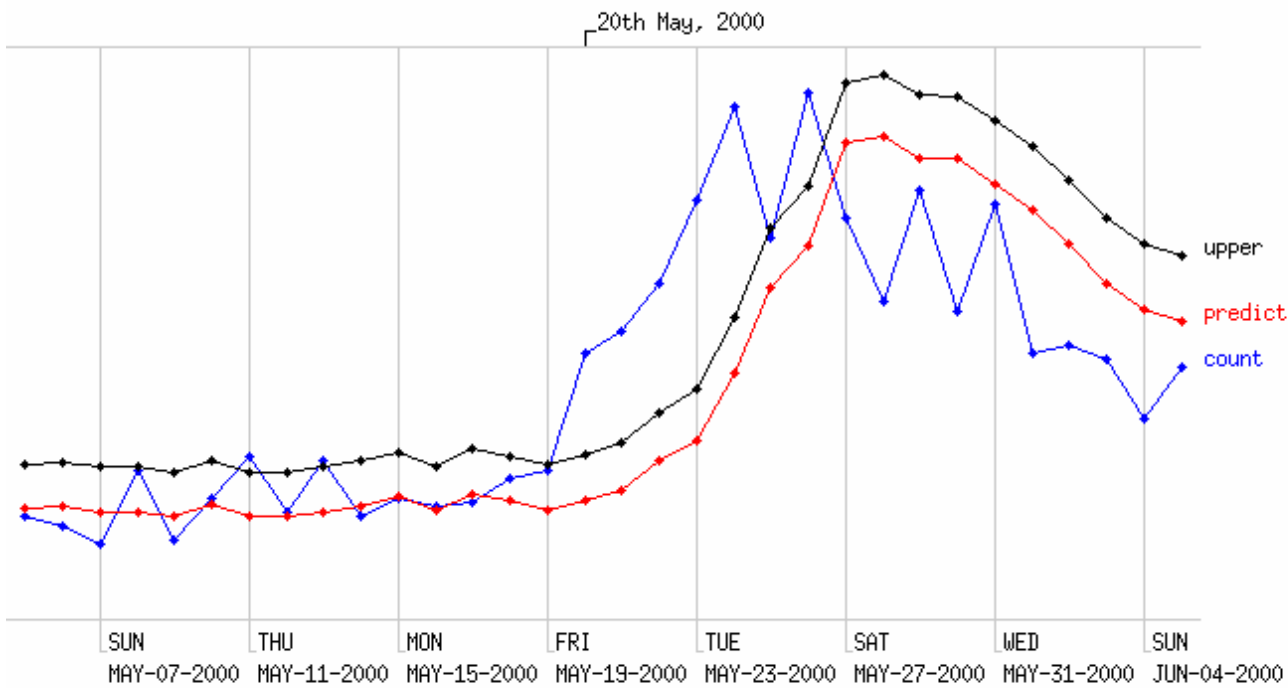
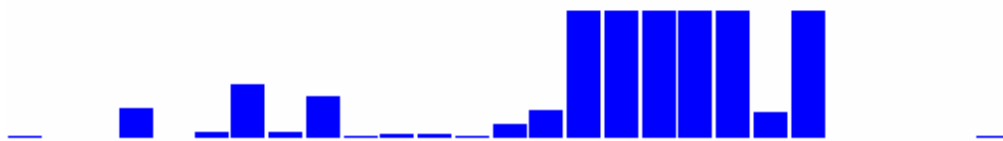
quit

Data	Tracking	Method
3 year data (excludes Hanover)	Walkerton visits (all patients from Walkerton)	Regression w/ DOW + Season



Data	Tracking	Method
3 year data (excludes Hanover)	GI visits from everywhere	Regression w/ DOW + Season

Bars show alarm levels; max = 10



<---jump

<---step

jump---

step---

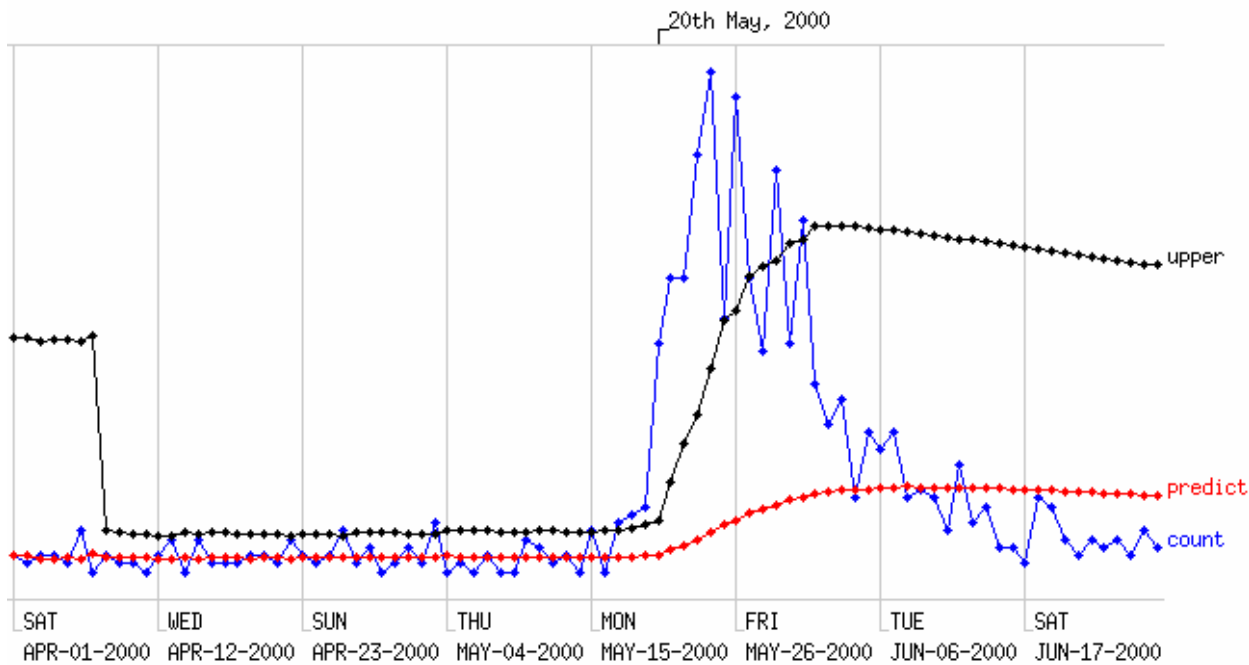
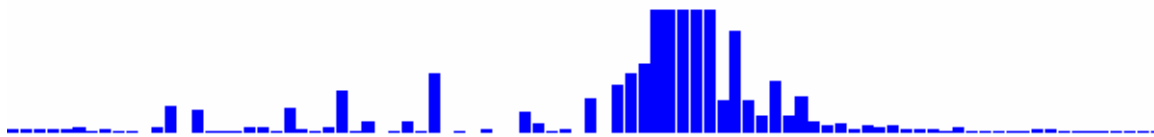
zoomout

print

quit

Data	Tracking	Method
3 month data (includes Hanover)	GI visits from city of Walkerton	Standard Control Chart

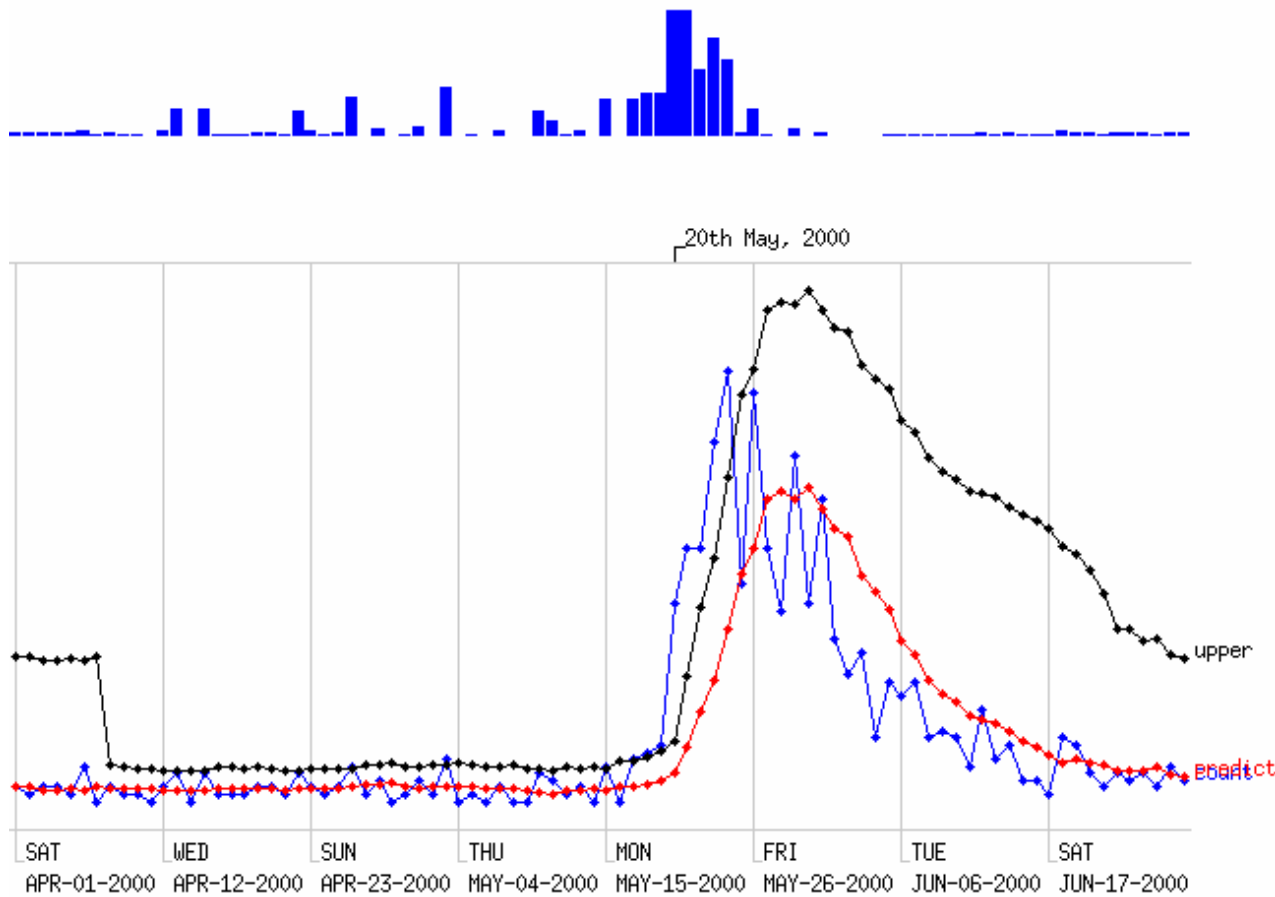
Bars show alarm levels: max = 10



- <---jump
- <---step
- jump---
- step---
- zoomout
- print
- quit

Data	Tracking	Method
3 month data (includes Hanover)	GI visits from city of Walkerton	Moving Average (7 days)

Bars show alarm levels; max = 10



- <---jump
- <---step
- jump---
- step---
- zoomout
- print
- quit

Data	Tracking
3 month data (includes Hanover)	GI visits from city of Walkerton

Performance of Univariate Algorithms: May 20th 2000

(False Positive Rate = Fraction of the days outside interval May 5-June 15 2000 in which there would have been an alarm, if the threshold was set such that the given date was detected)

```

cc      0.0% Standard Control Chart
ma1     0.0% Yesterday
ma3     0.0% Moving Average (3-day window)
ma7     0.0% Moving Average (7-day window)
ma14    0.0% Moving Average (14-day window)
ma28    0.0% Moving Average (28-day window)
regh    0.0% Regression (Hours of Daylight)
regm    0.0% Regression (HOD + Monday)
regtu   0.0% Regression (HOD + Monday + Tuesday)
regth   0.0% Regression (HOD + Monday + ... Thursday + Mean of Last Week)
regs    0.0% Regression (HOD + Monday + ... Saturday + Mean of Last Week)
regs84  0.0% Regression HOD + M..S + LastWeek learning from last 3 months
regs28  0.0% Regression HOD + M..S + LastWeek learning from last 4 weeks
regsh   0.0% Regression HOD + M..S + LastWeek avoiding times near holidays
cusum1  100.0% CUSUM With H = 1
cusum2  100.0% CUSUM With H = 2
cusum5  100.0% CUSUM With H = 5
cusum10 100.0% CUSUM With H = 10
cusum20 100.0% CUSUM With H = 20
sa1     1.8% Sickness/Availability with Window of 1 day
sa3     1.8% Sickness/Availability with Window of 3 days
sa7     1.8% Sickness/Availability with Window of 7 days
sa14    3.5% Sickness/Availability with Window of 14 days
sa28    3.5% Sickness/Availability with Window of 28 days

```

Data	Tracking
3 month data (includes Hanover)	GI visits from city of Walkerton

Performance of Univariate Algorithms: May 19th 2000

(False Positive Rate = Fraction of the days outside interval May 5-June 15 2000 in which there would have been an alarm, if the threshold was set such that the given date was detected)

```

cc      0.0% Standard Control Chart
ma1    31.6% Yesterday
ma3    5.3% Moving Average (3-day window)
ma7    1.8% Moving Average (7-day window)
ma14   0.0% Moving Average (14-day window)
ma28   1.8% Moving Average (28-day window)
regh   0.0% Regression (Hours of Daylight)
regm   0.0% Regression (HOD + Monday)
regtu  0.0% Regression (HOD + Monday + Tuesday)
regth  0.0% Regression (HOD + Monday + ... Thursday + Mean of Last Week)
regs   0.0% Regression (HOD + Monday + ... Saturday + Mean of Last Week)
regs84 0.0% Regression HOD + M..S + LastWeek learning from last 3 months
regs28 0.0% Regression HOD + M..S + LastWeek learning from last 4 weeks
regsh  0.0% Regression HOD + M..S + LastWeek avoiding times near holidays
cusum1 100.0% CUSUM With H = 1
cusum2 100.0% CUSUM With H = 2
cusum5 100.0% CUSUM With H = 5
cusum10 100.0% CUSUM With H = 10
cusum20 100.0% CUSUM With H = 20
sa1    1.8% Sickness/Availability with Window of 1 day
sa3    1.8% Sickness/Availability with Window of 3 days
sa7    1.8% Sickness/Availability with Window of 7 days
sa14   3.5% Sickness/Availability with Window of 14 days
sa28   3.5% Sickness/Availability with Window of 28 days

```


Data	Tracking
3 month data (includes Hanover)	GI visits from city of Walkerton

Performance of Univariate Algorithms: May 18th 2000

(False Positive Rate = Fraction of the days outside interval May 5-June 15 2000 in which there would have been an alarm, if the threshold was set such that the given date was detected)

```

cc      0.0% Standard Control Chart
ma1    31.6% Yesterday
ma3     7.0% Moving Average (3-day window)
ma7     1.8% Moving Average (7-day window)
ma14    0.0% Moving Average (14-day window)
ma28    1.8% Moving Average (28-day window)
regh    1.8% Regression (Hours of Daylight)
regm    1.8% Regression (HOD + Monday)
regtu   1.8% Regression (HOD + Monday + Tuesday)
regth   0.0% Regression (HOD + Monday + ... Thursday + Mean of Last Week)
regs    0.0% Regression (HOD + Monday + ... Saturday + Mean of Last Week)
regs84  0.0% Regression HOD + M..S + LastWeek learning from last 3 months
regs28  0.0% Regression HOD + M..S + LastWeek learning from last 4 weeks
regsh   0.0% Regression HOD + M..S + LastWeek avoiding times near holidays
cusum1 100.0% CUSUM With H = 1
cusum2 100.0% CUSUM With H = 2
cusum5 100.0% CUSUM With H = 5
cusum10 100.0% CUSUM With H = 10
cusum20 100.0% CUSUM With H = 20
sa1     7.0% Sickness/Availability with Window of 1 day
sa3     5.3% Sickness/Availability with Window of 3 days
sa7     3.5% Sickness/Availability with Window of 7 days
sa14    3.5% Sickness/Availability with Window of 14 days
sa28    5.3% Sickness/Availability with Window of 28 days

```

Data	Tracking
3 month data (includes Hanover)	GI visits from city of Walkerton

Performance of Univariate Algorithms: May 17th 2000

(False Positive Rate = Fraction of the days outside interval May 5-June 15 2000 in which there would have been an alarm, if the threshold was set such that the given date was detected)

```

cc      1.8% Standard Control Chart
ma1     0.0% Yesterday
ma3     3.5% Moving Average (3-day window)
ma7     3.5% Moving Average (7-day window)
ma14    1.8% Moving Average (14-day window)
ma28    1.8% Moving Average (28-day window)
regh    1.8% Regression (Hours of Daylight)
regm    1.8% Regression (HOD + Monday)
regtu   1.8% Regression (HOD + Monday + Tuesday)
regth   1.8% Regression (HOD + Monday + ... Thursday + Mean of Last Week)
regs    1.8% Regression (HOD + Monday + ... Saturday + Mean of Last Week)
regs84  1.8% Regression HOD + M..S + LastWeek learning from last 3 months
regs28  1.8% Regression HOD + M..S + LastWeek learning from last 4 weeks
regsh   1.8% Regression HOD + M..S + LastWeek avoiding times near holidays
cusum1  100.0% CUSUM With H = 1
cusum2  100.0% CUSUM With H = 2
cusum5  100.0% CUSUM With H = 5
cusum10 100.0% CUSUM With H = 10
cusum20 100.0% CUSUM With H = 20
sa1     3.5% Sickness/Availability with Window of 1 day
sa3     3.5% Sickness/Availability with Window of 3 days
sa7     5.3% Sickness/Availability with Window of 7 days
sa14    7.0% Sickness/Availability with Window of 14 days
sa28    7.0% Sickness/Availability with Window of 28 days

```

Data	Tracking
3 month data (includes Hanover)	GI visits from city of Walkerton

Performance of Univariate Algorithms: May 16th 2000

(False Positive Rate = Fraction of the days outside interval May 5-June 15 2000 in which there would have been an alarm, if the threshold was set such that the given date was detected)

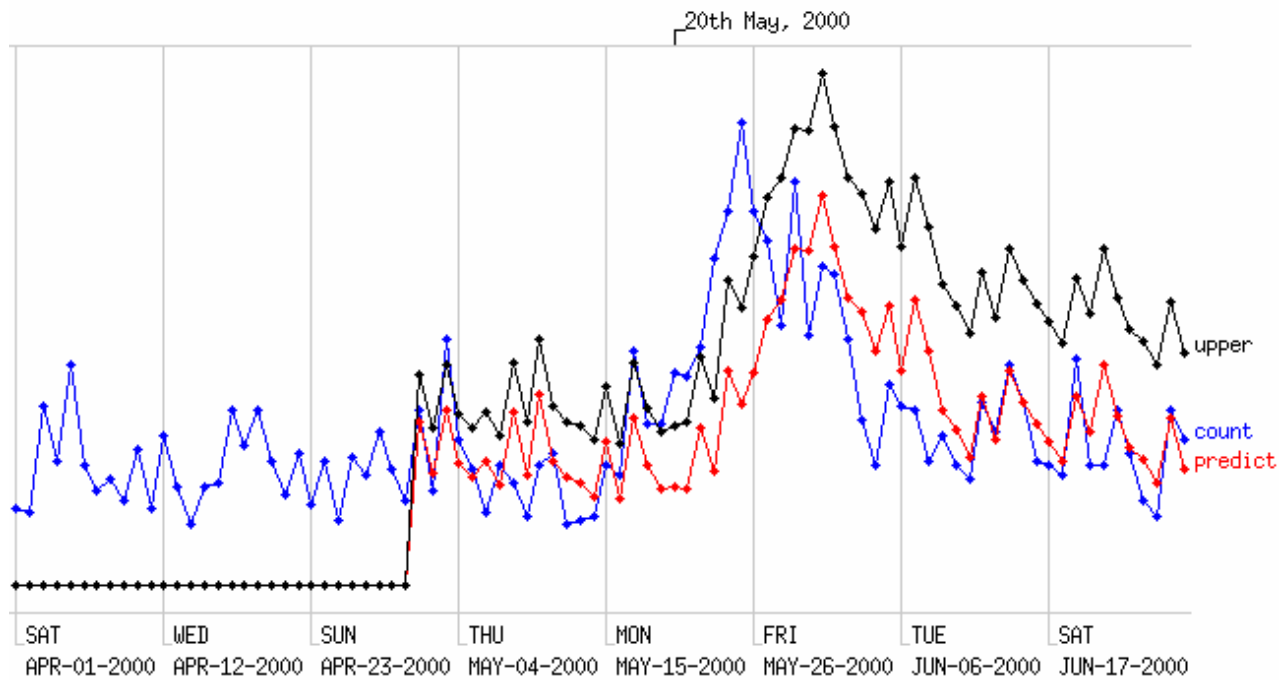
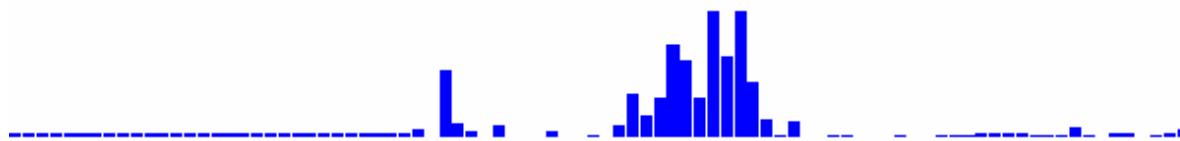
```

cc 93.0% Standard Control Chart
ma1 98.2% Yesterday
ma3 94.7% Moving Average (3-day window)
ma7 94.7% Moving Average (7-day window)
ma14 93.0% Moving Average (14-day window)
ma28 93.0% Moving Average (28-day window)
regh 77.2% Regression (Hours of Daylight)
regm 78.9% Regression (HOD + Monday)
regtu 57.9% Regression (HOD + Monday + Tuesday)
regth 86.0% Regression (HOD + Monday + ... Thursday + Mean of Last Week)
regs 86.0% Regression (HOD + Monday + ... Saturday + Mean of Last Week)
regs84 86.0% Regression HOD + M..S + LastWeek learning from last 3 months
regs28 89.5% Regression HOD + M..S + LastWeek learning from last 4 weeks
regsh 86.0% Regression HOD + M..S + LastWeek avoiding times near holidays
cusum1 100.0% CUSUM With H = 1
cusum2 100.0% CUSUM With H = 2
cusum5 100.0% CUSUM With H = 5
cusum10 100.0% CUSUM With H = 10
cusum20 100.0% CUSUM With H = 20
sa1 98.2% Sickness/Availability with Window of 1 day
sa3 98.2% Sickness/Availability with Window of 3 days
sa7 98.2% Sickness/Availability with Window of 7 days
sa14 100.0% Sickness/Availability with Window of 14 days
sa28 80.7% Sickness/Availability with Window of 28 days

```

Data	Tracking	Method
3 month data (includes Hanover)	Walkerton visits (all patients from Walkerton)	Regression w/ DOW + Season

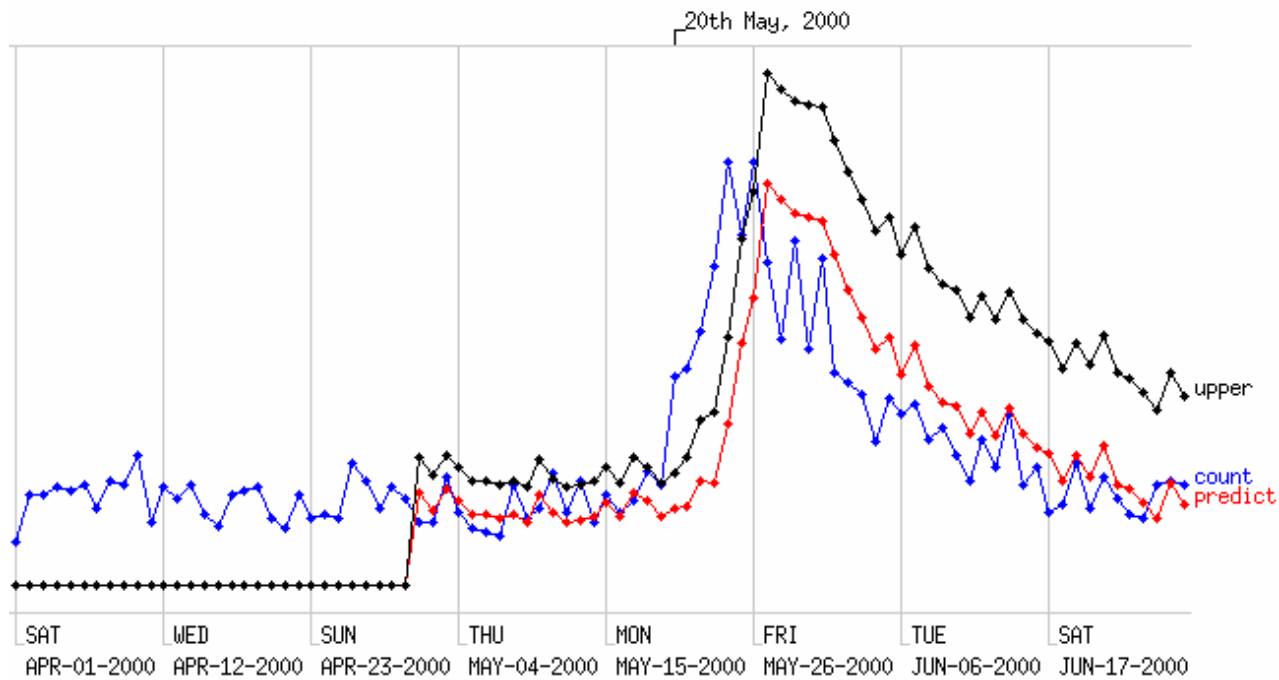
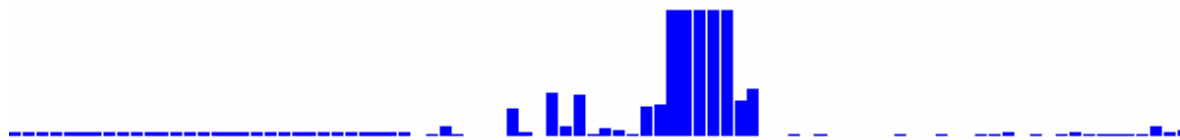
Bars show alarm levels; max = 10



- <---jump
- <---step
- jump---
- step---
- zoomout
- print
- quit

Data	Tracking	Method
3 month data (includes Hanover)	GI visits from everywhere	Regression w/ DOW + Season

Bars show alarm levels; max = 10



<---jump

<---step

jump--->

step--->

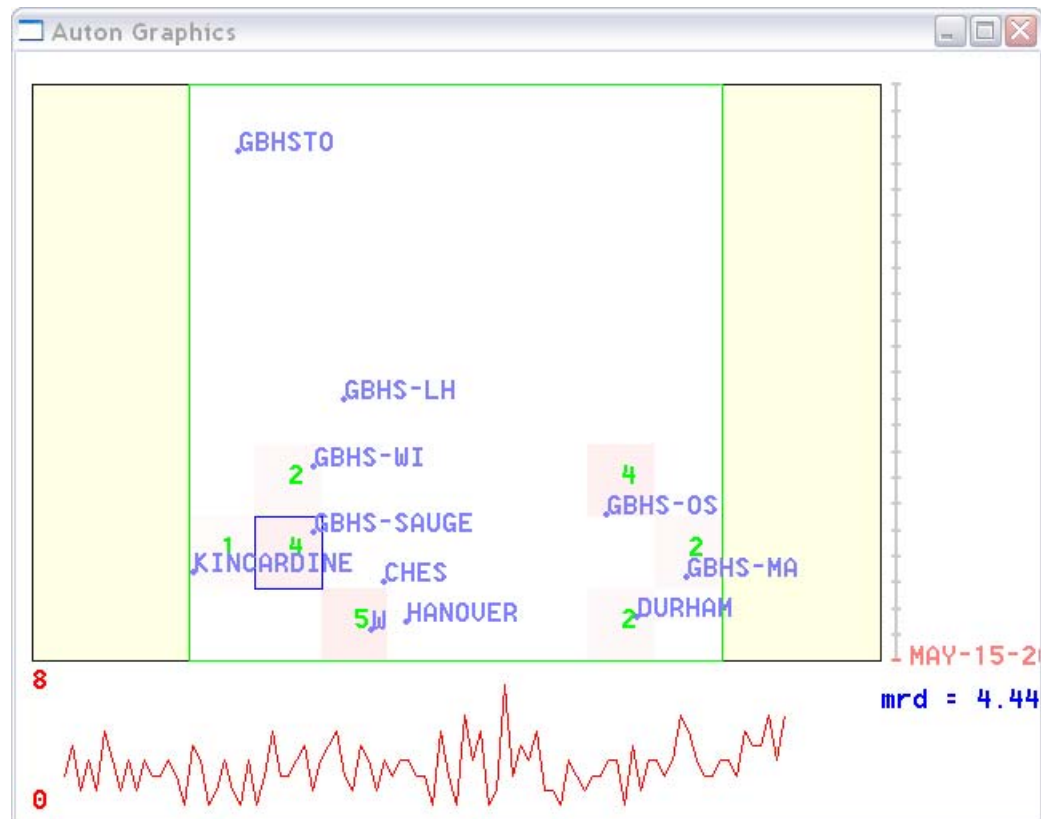
zoomout

print

quit

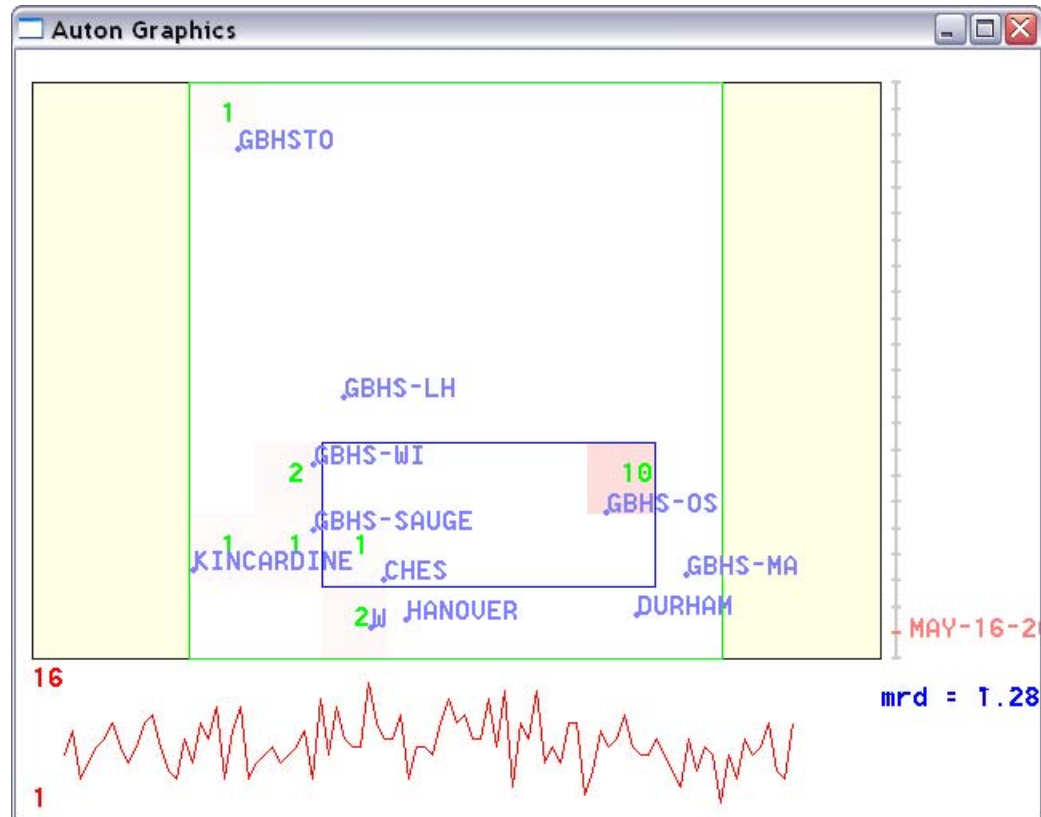
Spatial Scan: Hospital

Data	Tracking	Method
3 year data (excludes Hanover)	GI visits from everywhere	Fast Spatial Scan on Hospital Long/Latitude



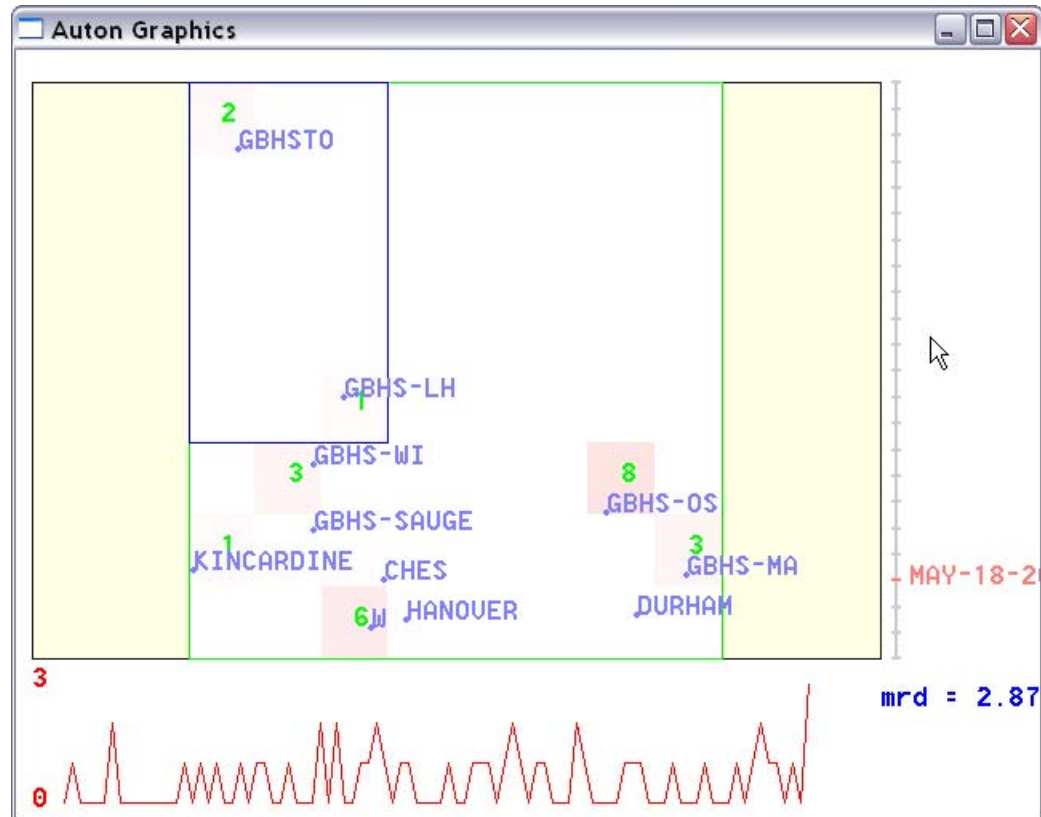
Spatial Scan: Hospital

Data	Tracking	Method
3 year data (excludes Hanover)	GI visits from everywhere	Fast Spatial Scan on Hospital Long/Latitude



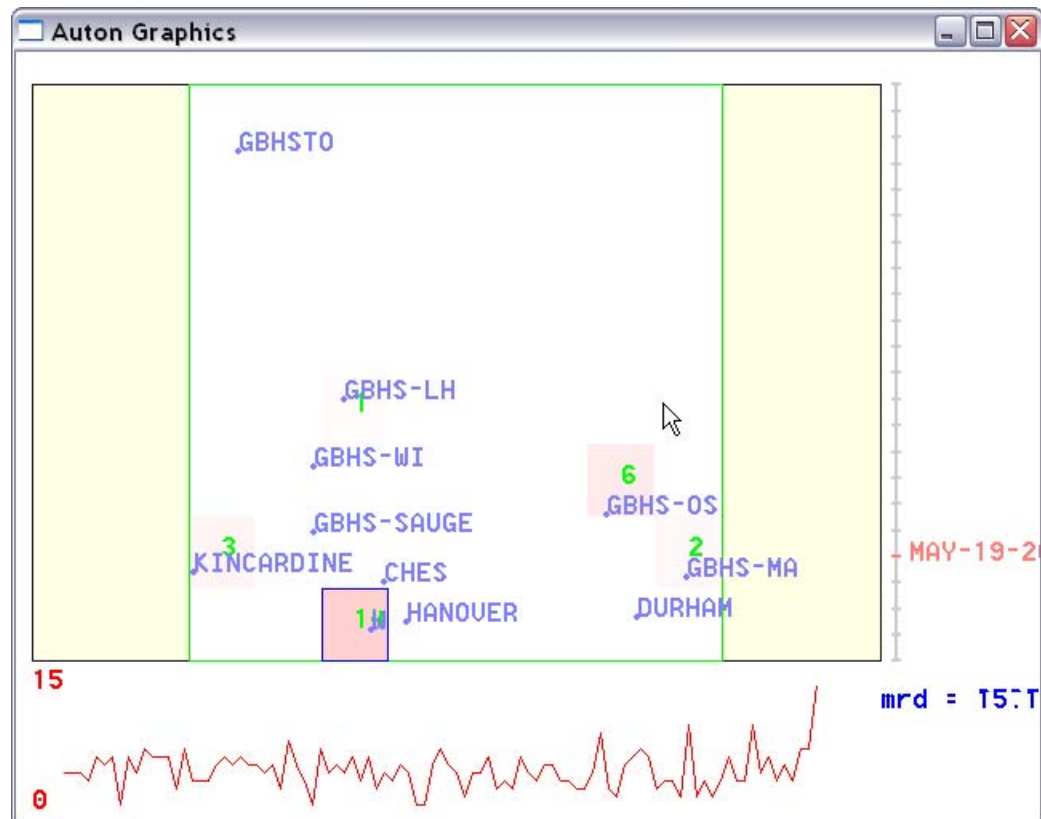
Spatial Scan: Hospital

Data	Tracking	Method
3 year data (excludes Hanover)	GI visits from everywhere	Fast Spatial Scan on Hospital Long/Latitude



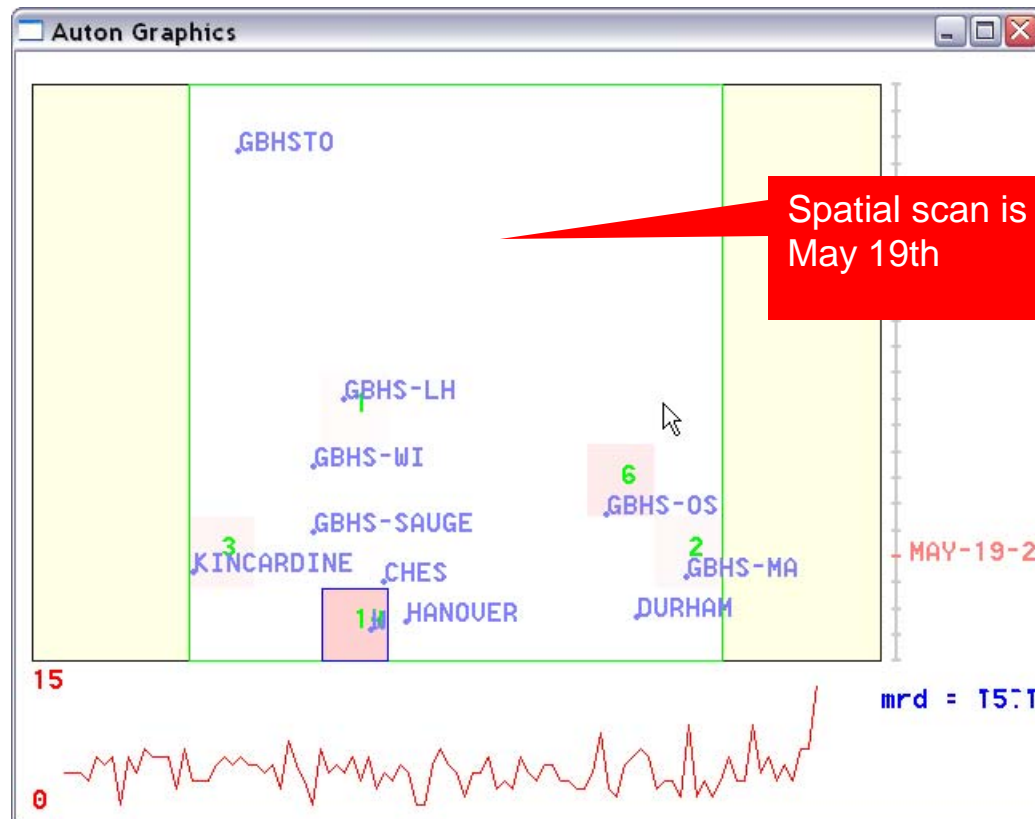
Spatial Scan: Hospital

Data	Tracking	Method
3 year data (excludes Hanover)	GI visits from everywhere	Fast Spatial Scan on Hospital Long/Latitude



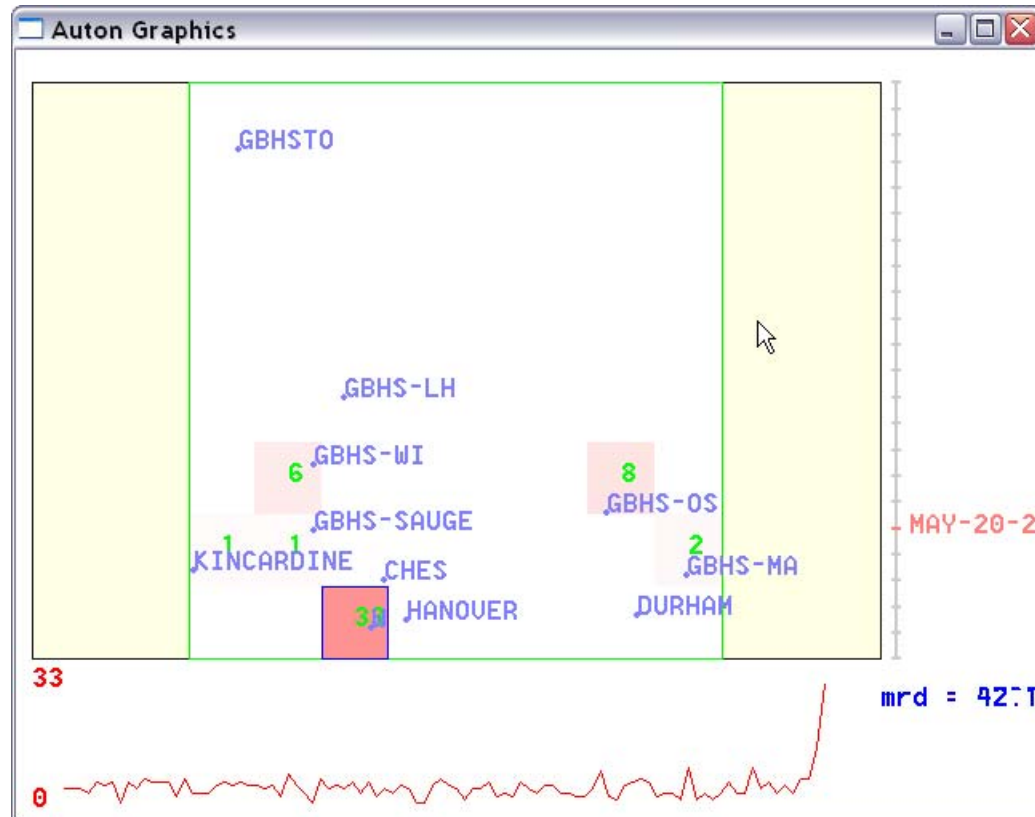
Spatial Scan: Hospital

Data	Tracking	Method
3 year data (excludes Hanover)	GI visits from everywhere	Fast Spatial Scan on Hospital Long/Latitude



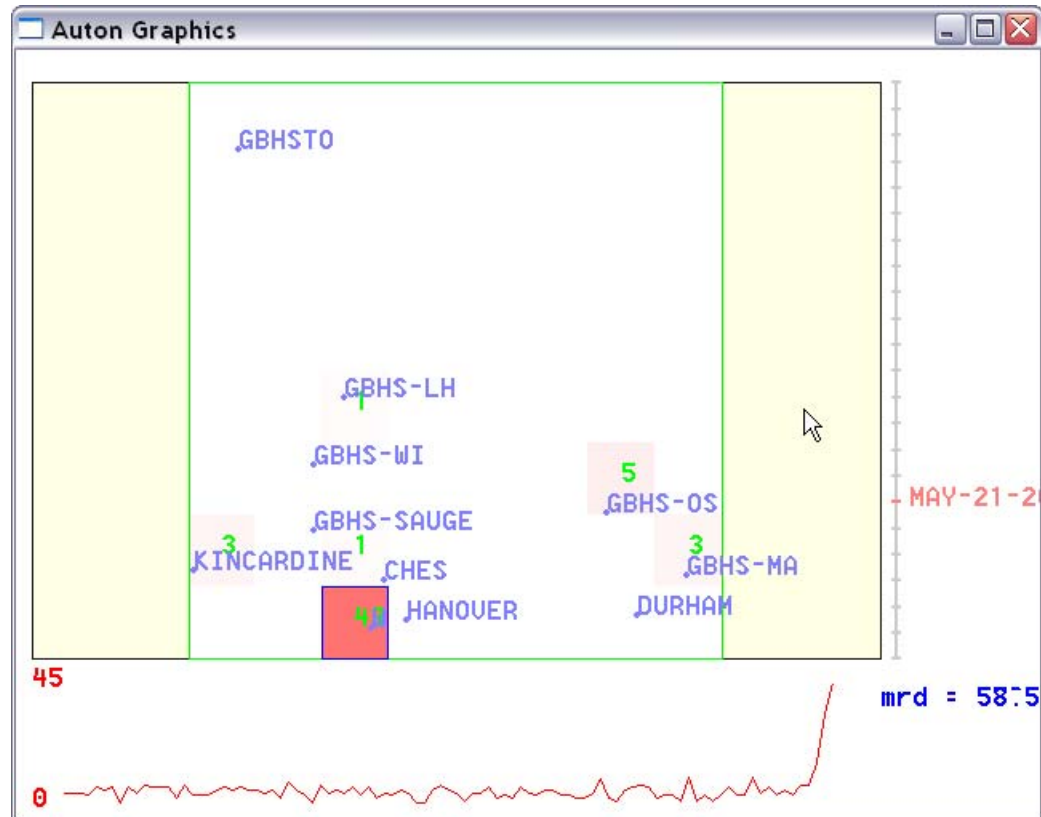
Spatial Scan: Hospital

Data	Tracking	Method
3 year data (excludes Hanover)	GI visits from everywhere	Fast Spatial Scan on Hospital Long/Latitude



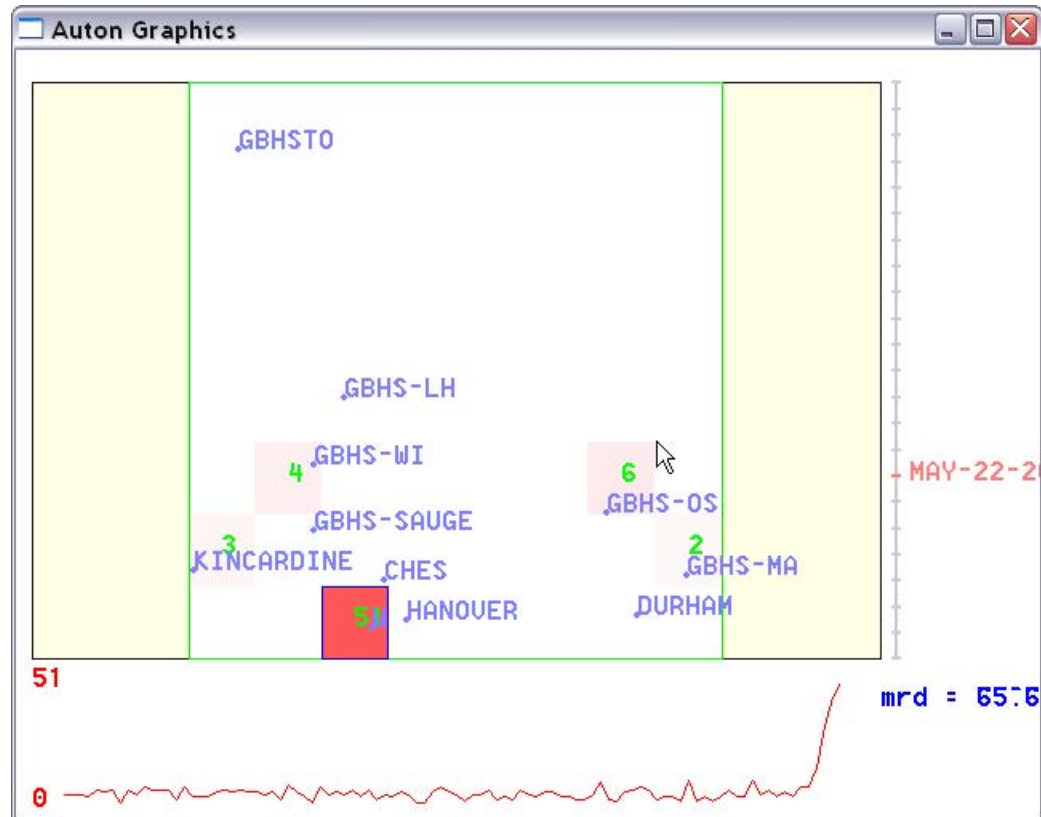
Spatial Scan: Hospital

Data	Tracking	Method
3 year data (excludes Hanover)	GI visits from everywhere	Fast Spatial Scan on Hospital Long/Latitude



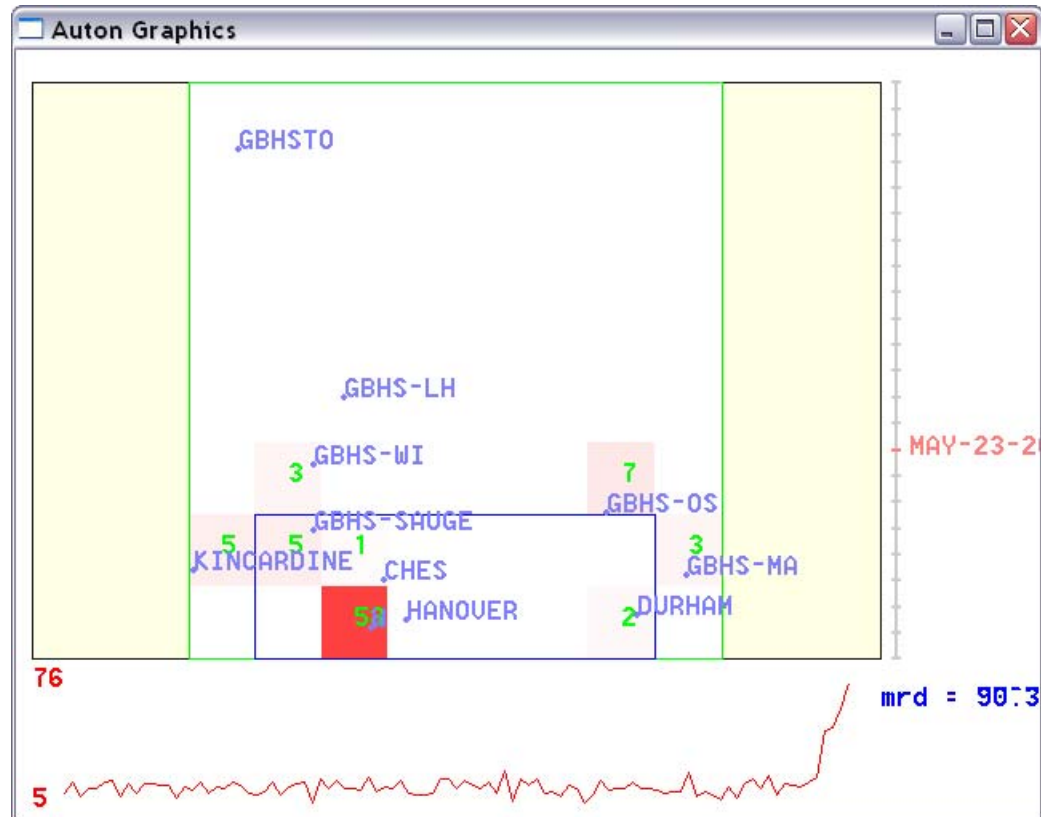
Spatial Scan: Hospital

Data	Tracking	Method
3 year data (excludes Hanover)	GI visits from everywhere	Fast Spatial Scan on Hospital Long/Latitude



Spatial Scan: Hospital

Data	Tracking	Method
3 year data (excludes Hanover)	GI visits from everywhere	Fast Spatial Scan on Hospital Long/Latitude



Spatial Scan: Hospital

Data	Tracking	Method
3 year data (excludes Hanover)	GI visits from everywhere	Fast Spatial Scan on Hospital Long/Latitude

Total number of non-attack days: 990				
Method/Date	05/18/00	05/19/00	05/20/00	05/21/00
all_mean_1	210	7	0	0
all_mean_3	416	10	0	0
adj_EWLR_1	230	4	0	0
adj_EWLR_3	489	27	0	0
adj_EWMA_1	203	1	0	0
adj_EWMA_3	412	2	0	0
strat_mean_1	302	6	0	0
strat_mean_3	526	15	0	0

Spatial Scan: Hospital

Data	Tracking	Method
3 year data (excludes Hanover)	GI visits from everywhere	Fast Spatial Scan on Hospital Long/Latitude

Total number of non-attack days: 990				
Method/Date	05/18/00	05/19/00	05/20/00	05/21/00
all_mean_1	210	7	0	0
all_mean_3	416	10	0	0
adj_EWLR_1	230	4	0	0
adj_EWLR_3	489	27	0	0
adj_EWMA_1	203	1	0	0
adj_EWMA_3	412	2	0	0
strat_mean_1	302	6	0	0
strat_mean_3	526	15	0	0

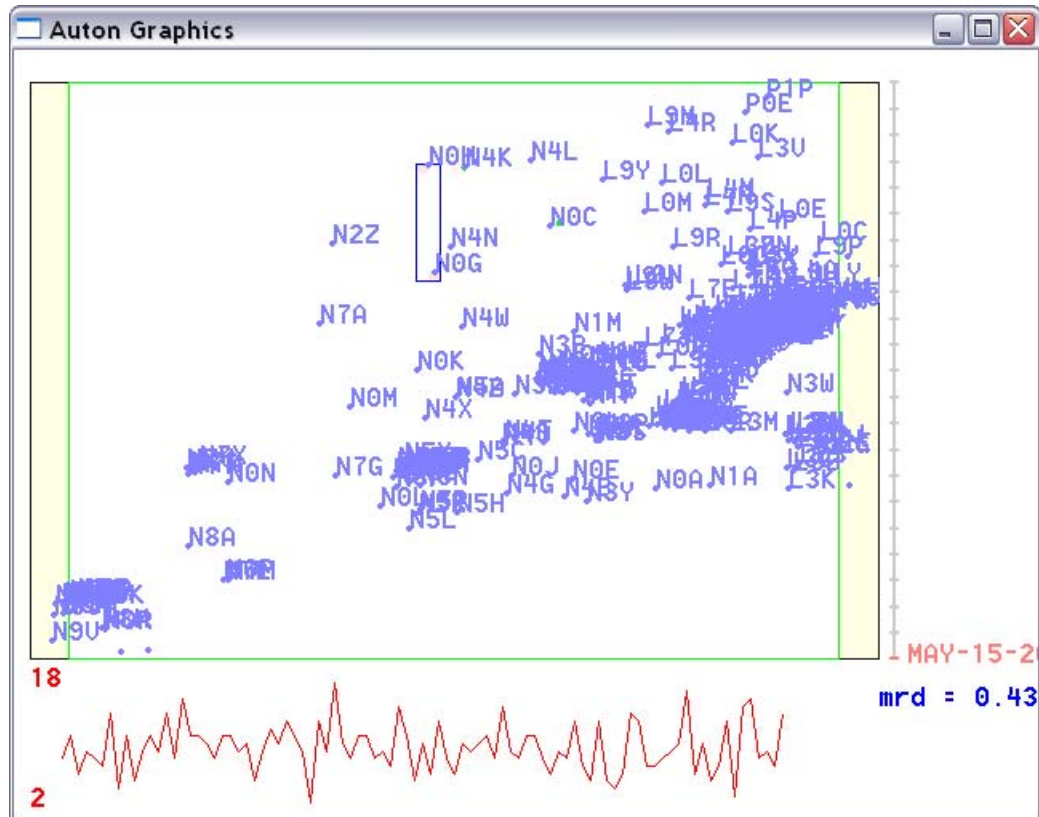
we tested many variants of spatial scan, listed (with cryptic names) in the first column

Column 2 shows that if you set your alarm threshold high enough that there'd be an alarm on May 18th, all methods would produce hundreds of additional alarms on other occasions throughout the three years.

Column 3 shows that detecting May 19th involves far fewer false alarms, especially for the EWMA models

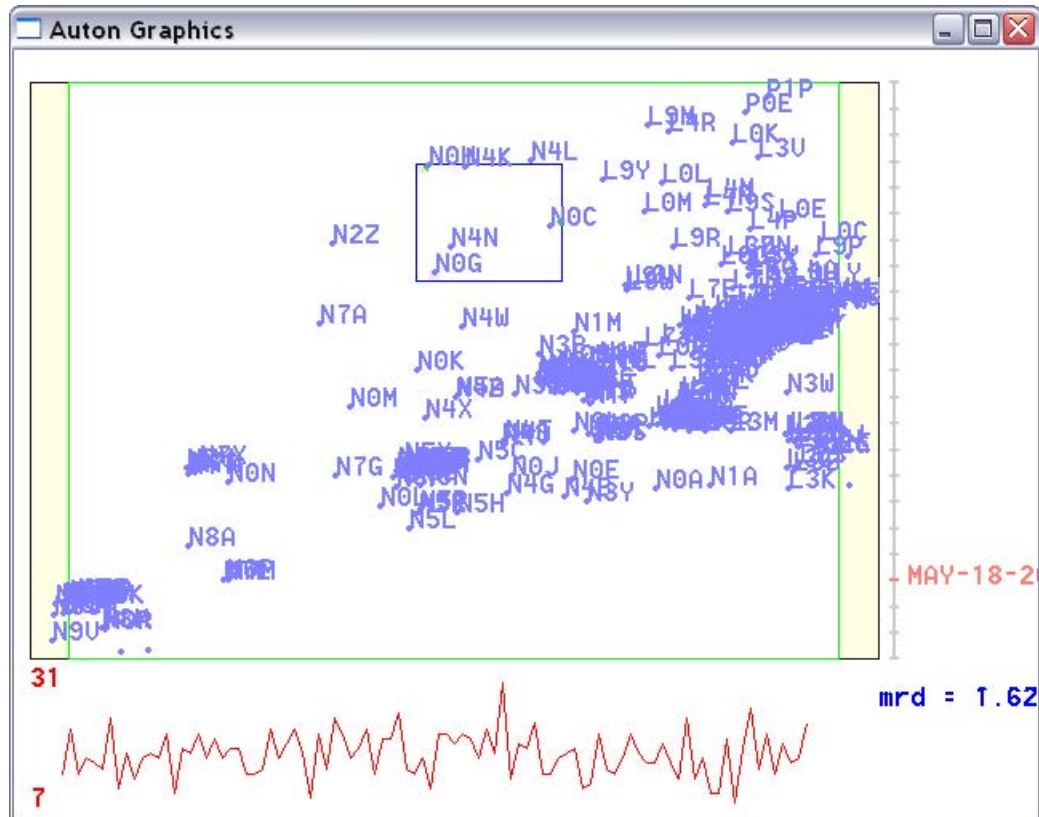
Spatial Scan: Home

Data	Tracking	Method
3 year data (excludes Hanover)	GI visits from everywhere	Fast Spatial Scan on Patient Home 3- character Postcode



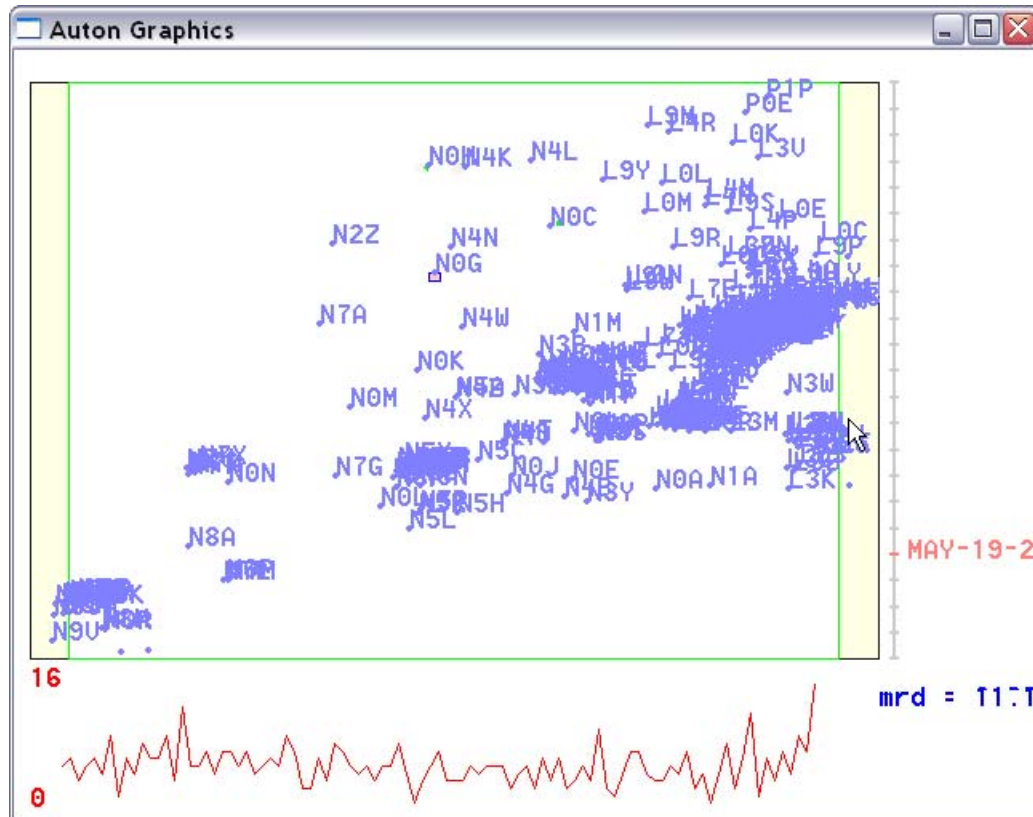
Spatial Scan: Home

Data	Tracking	Method
3 year data (excludes Hanover)	GI visits from everywhere	Fast Spatial Scan on Patient Home 3- character Postcode



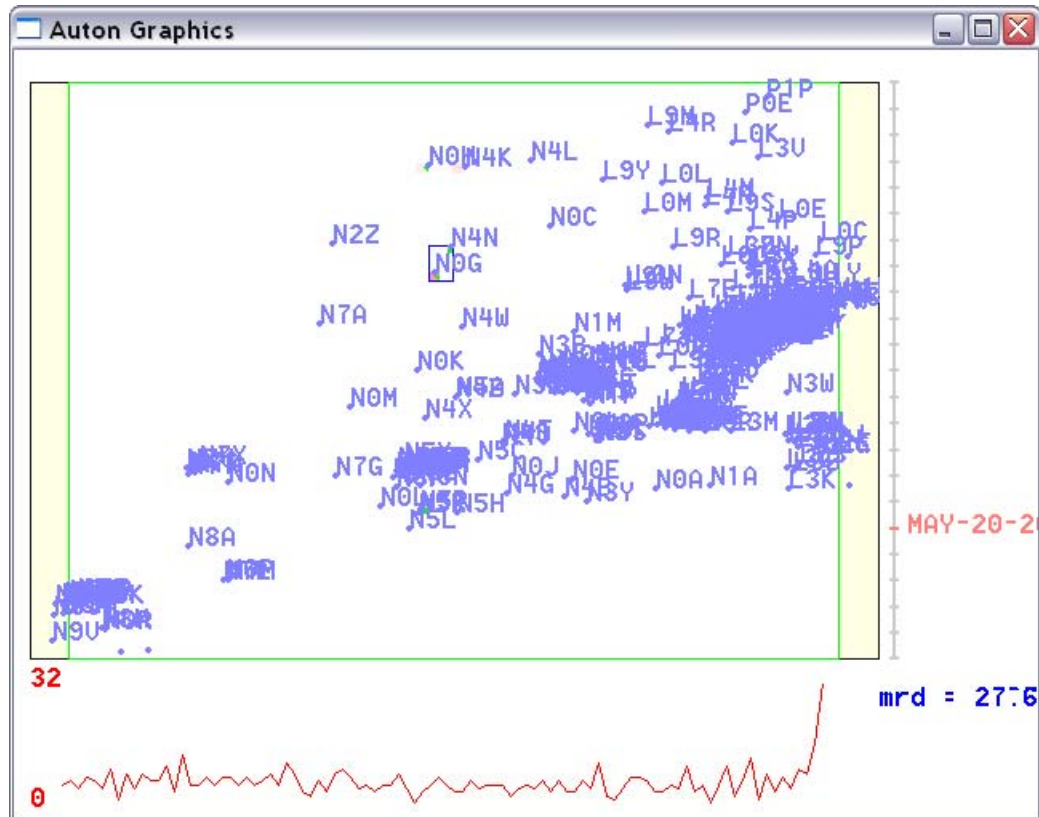
Spatial Scan: Home

Data	Tracking	Method
3 year data (excludes Hanover)	GI visits from everywhere	Fast Spatial Scan on Patient Home 3- character Postcode



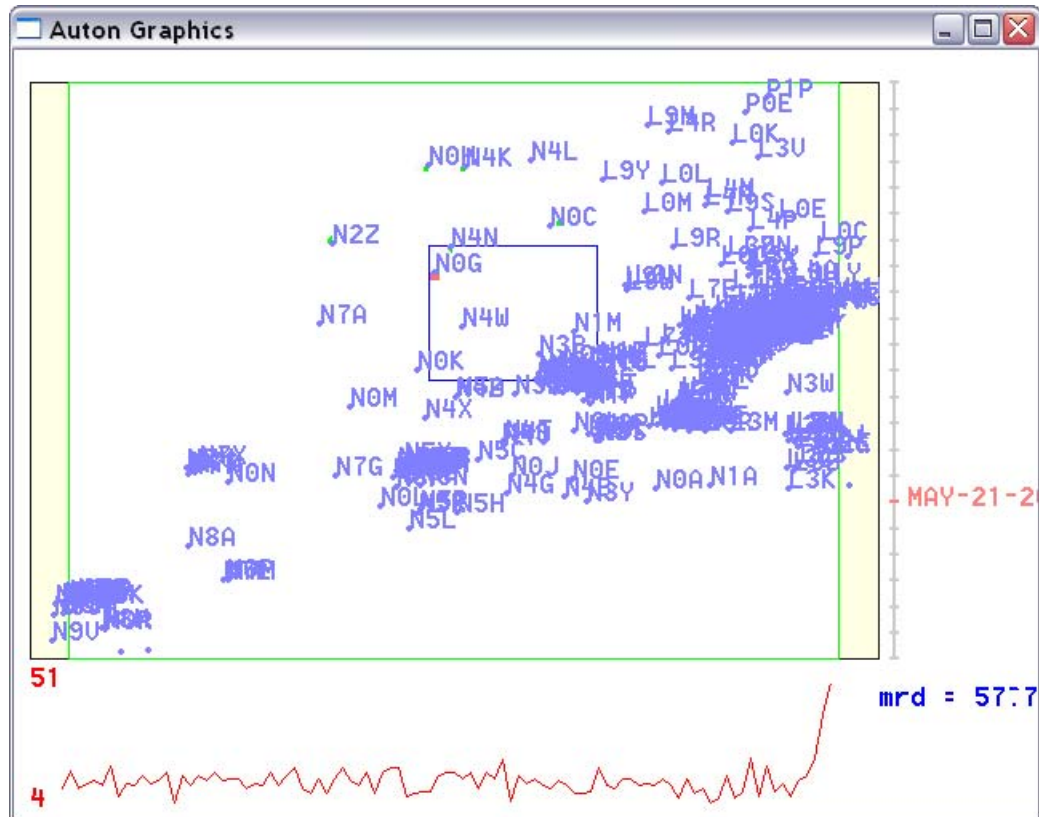
Spatial Scan: Home

Data	Tracking	Method
3 year data (excludes Hanover)	GI visits from everywhere	Fast Spatial Scan on Patient Home 3- character Postcode



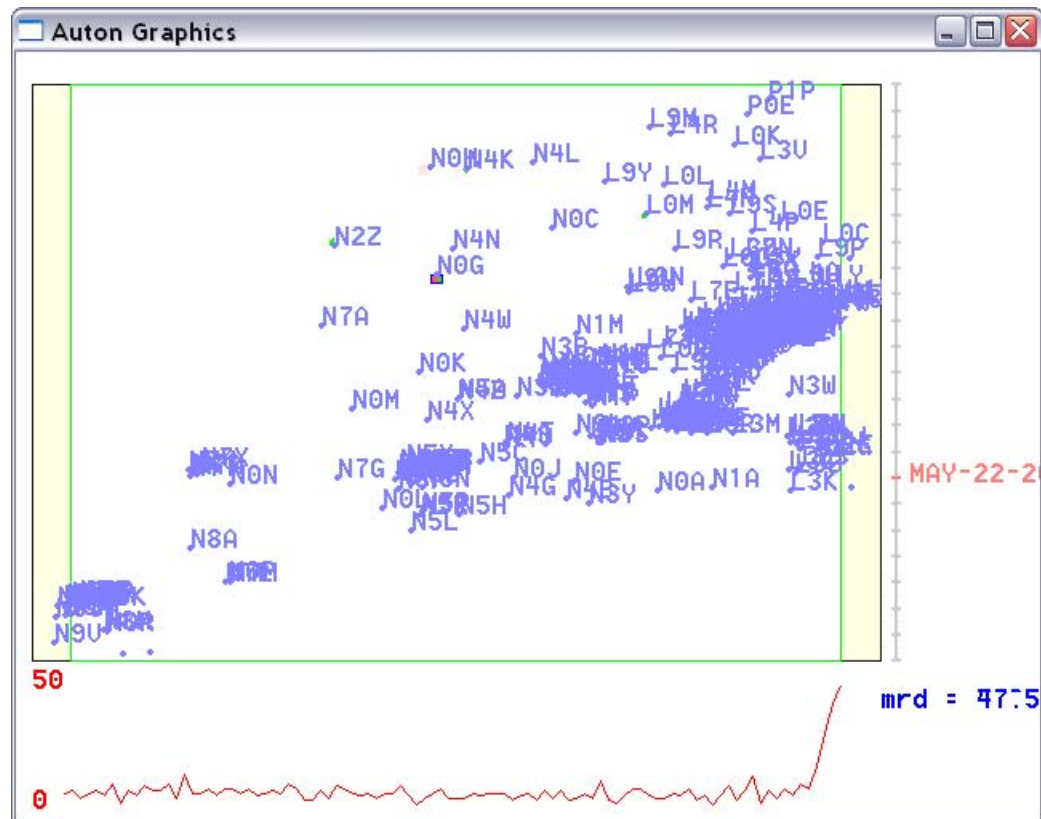
Spatial Scan: Home

Data	Tracking	Method
3 year data (excludes Hanover)	GI visits from everywhere	Fast Spatial Scan on Patient Home 3- character Postcode



Spatial Scan: Home

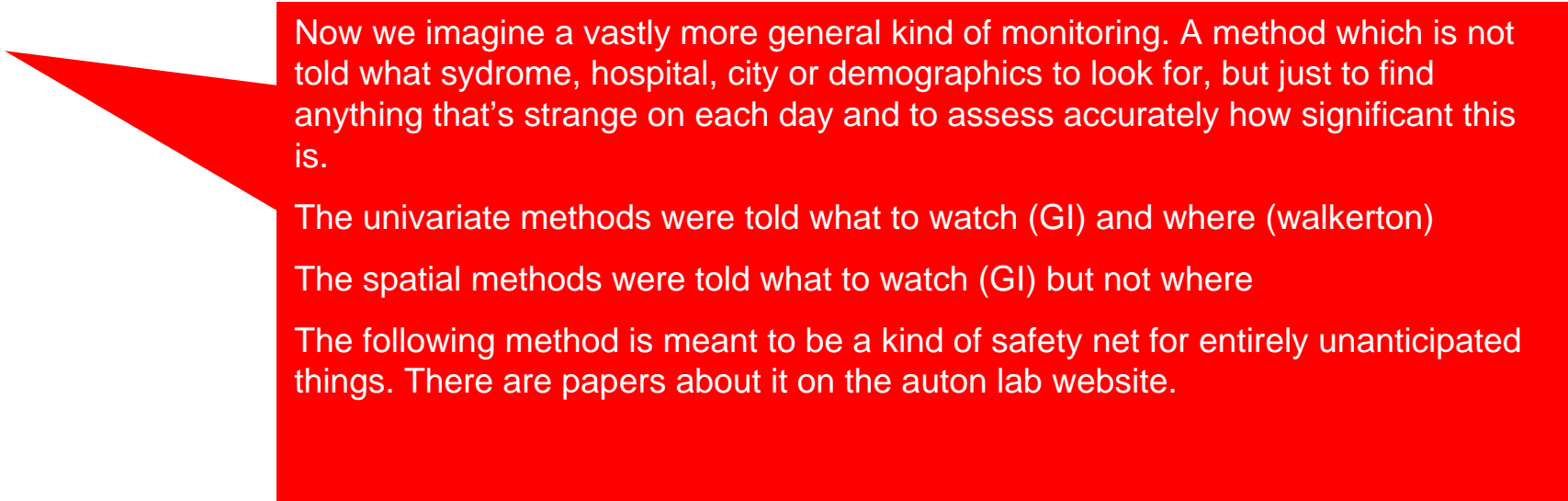
Data	Tracking	Method
3 year data (excludes Hanover)	GI visits from everywhere	Fast Spatial Scan on Patient Home 3- character Postcode



Spatial Scan: Home

Data	Tracking	Method
3 year data (excludes Hanover)	GI visits from everywhere	Fast Spatial Scan on Patient Home 3- character Postcode

Total number of non-attack days: 990				
Method/Date	05/18/00	05/19/00	05/20/00	05/21/00
all_mean_1	396	15	0	0
all_mean_3	583	25	0	0
adj_EWLR_1	351	6	1	0
adj_EWLR_3	696	30	6	0
adj_EWMA_1	390	2	0	0
adj_EWMA_3	685	8	0	0
strat_mean_1	438	3	0	0
strat_mean_3	738	11	0	0



Now we imagine a vastly more general kind of monitoring. A method which is not told what syndrome, hospital, city or demographics to look for, but just to find anything that's strange on each day and to assess accurately how significant this is.

The univariate methods were told what to watch (GI) and where (walkerton)

The spatial methods were told what to watch (GI) but not where

The following method is meant to be a kind of safety net for entirely unanticipated things. There are papers about it on the auton lab website.

WSARE

Data	Tracking	Method
3 year data (excludes Hanover)	All data from everywhere	What's Strange About Recent Events

The most surprising thing about MAR-01-2000 is:

Normally 1.5% of records (20/1335) have sendng = CHES and syndrome = other
But recently 10.0% of records (39/390) have sendng = CHES and syndrome = other
Pvalue = 0.00099975
Which means that in a world where nothing changes we'd
expect to have a result this significant about once
every 1000 times we ran the program

The most surprising thing about MAR-23-2000 is:

Normally 2.3% of records (28/1241) have sendng = CHES and syndrome = other
But recently 9.6% of records (45/470) have sendng = CHES and syndrome = other
Pvalue = 0.0005
Which means that in a world where nothing changes we'd
expect to have a result this significant about once
every 1999 times we ran the program

The most surprising thing about MAR-31-2000 is:

Normally 22.1% of records (270/1223) have sendng = GBHS-OS
But recently 39.5% of records (96/243) have sendng = GBHS-OS
Pvalue = 0.0005
Which means that in a world where nothing changes we'd
expect to have a result this significant about once
every 1999 times we ran the program

The most surprising thing about MAY-20-2000 is:

Normally 0.2% of records (2/1042) have city = WALKERTON and syndrome = gastrointestinal
But recently 5.8% of records (23/396) have city = WALKERTON and syndrome = gastrointestinal
Pvalue = 0.00099975
Which means that in a world where nothing changes we'd
expect to have a result this significant about once
every 1000 times we ran the program

The most surprising thing about MAY-21-2000 is:

Normally 0.8% of records (7/906) have sendng = W and syndrome = gastrointestinal
But recently 10.2% of records (43/422) have sendng = W and syndrome = gastrointestinal
Pvalue = 0.0005
Which means that in a world where nothing changes we'd
expect to have a result this significant about once
every 1999 times we ran the program

Results 1-5 of the 14
days in 2000 in which
WSARE issued an
alert with a Pvalue
exceeding 1 in 1000

*9 of the 15 (shown in green)
were during or just after the
outbreak period*

WSARE

Data	Tracking	Method
3 year data (excludes Hanover)	All data from everywhere	What's Strange About Recent Events

The most surprising thing about MAY-22-2000 is:

Normally 0.9% of records (12/1360) have sendng = W and syndrome = gastrointestinal
But recently 11.6% of records (51/438) have sendng = W and syndrome = gastrointestinal
Pvalue = 0.0005

Which means that in a world where nothing changes we'd expect to have a result this significant about once every 1999 times we ran the program

The most surprising thing about MAY-23-2000 is:

Normally 0.4% of records (5/1262) have sendng = W and syndrome = gastrointestinal
But recently 9.8% of records (58/594) have sendng = W and syndrome = gastrointestinal
Pvalue = 0.0005

Which means that in a world where nothing changes we'd expect to have a result this significant about once every 1999 times we ran the program

The most surprising thing about MAY-24-2000 is:

Normally 1.4% of records (18/1331) have sendng = W and syndrome = gastrointestinal
But recently 13.7% of records (76/554) have sendng = W and syndrome = gastrointestinal
Pvalue = 0.0005

Which means that in a world where nothing changes we'd expect to have a result this significant about once every 1999 times we ran the program

The most surprising thing about MAY-26-2000 is:

Normally 1.5% of records (18/1216) have sendng = W and syndrome = gastrointestinal
But recently 15.6% of records (77/494) have sendng = W and syndrome = gastrointestinal
Pvalue = 0.0005

Which means that in a world where nothing changes we'd expect to have a result this significant about once every 1999 times we ran the program

The most surprising thing about JUN-14-2000 is:

Normally 5.5% of records (84/1533) have sendng = W and syndrome = other
But recently 14.5% of records (82/567) have sendng = W and syndrome = other
Pvalue = 0.0005

Which means that in a world where nothing changes we'd expect to have a result this significant about once every 1999 times we ran the program

Results 6-10 of the 14 days in 2000 in which WSARE issued an alert with a Pvalue exceeding 1 in 1000

9 of the 15 (shown in green) were during or just after the outbreak period

WSARE

Data	Tracking	Method
3 year data (excludes Hanover)	All data from everywhere	What's Strange About Recent Events

The most surprising thing about JUN-15-2000 is:

Normally 1.7% of records (28/1624) have sendng = CHES and syndrome = other
But recently 9.0% of records (45/501) have sendng = CHES and syndrome = other
Pvalue = 0.0005

Which means that in a world where nothing changes we'd expect to have a result this significant about once every 1999 times we ran the program

The most surprising thing about JUN-16-2000 is:

Normally 0.2% of records (3/1446) have sendng = GBHSTO and syndrome = other
But recently 4.5% of records (20/441) have sendng = GBHSTO and syndrome = other
Pvalue = 0.0005

Which means that in a world where nothing changes we'd expect to have a result this significant about once every 1999 times we ran the program

The most surprising thing about JUL-13-2000 is:

Normally 3.1% of records (45/1436) have sendng = CHES
But recently 11.2% of records (54/481) have sendng = CHES
Pvalue = 0.00099975

Which means that in a world where nothing changes we'd expect to have a result this significant about once every 1000 times we ran the program

The most surprising thing about AUG-14-2000 is:

Normally 1.9% of records (25/1336) have sendng = DURHAM
But recently 9.3% of records (44/474) have sendng = DURHAM
Pvalue = 0.00099975

Which means that in a world where nothing changes we'd expect to have a result this significant about once every 1000 times we ran the program

Results 11-14 of the 14 days in 2000 in which WSARE issued an alert with a Pvalue exceeding 1 in 1000
9 of the 15 (shown in green) were during or just after the outbreak period

WSARE

Data	Tracking	Method
3 year data (excludes Hanover)	All data from everywhere	What's Strange About Recent Events

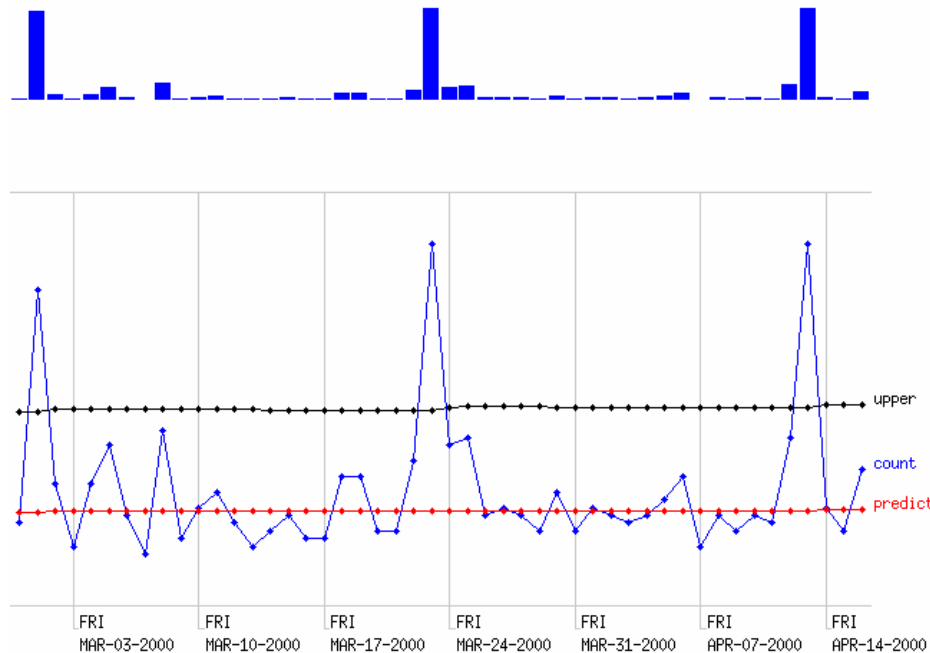
The most surprising thing about MAR-01-2000 is:

Normally 1.5% of records (20/1335) have sendng = CHES and syndrome = other
 But recently 10.0% of records (39/390) have sendng = CHES and syndrome = other
 Pvalue = 0.00099975
 Which means that in a world where nothing changes we'd expect to have a result this significant about once every 1000 times we ran the program

The most surprising thing about MAR-23-2000 is:

Normally 2.3% of records (28/1241) have sendng = CHES and syndrome = other
 But recently 9.6% of records (45/470) have sendng = CHES and syndrome = other
 Pvalue = 0.0005
 Which means that in a world where nothing changes we'd expect to have a result this significant about once every 1999 times we ran the program

Looking at the signal detected in alerts 1 and 2



- <---jump
- <---step
- jump---
- step---
- zoomout
- print
- quit

did something happen roughly every three weeks that sent a bunch of folks to CHES with "other" injury?

WSARE

The most surprising thing about MAY-15-2000 is:

Normally 3.4% of records (46/1345) have sendng = DURHAM
But recently 7.6% of records (34/449) have sendng = DURHAM
Pvalue = 0.0838333
Which means that in a world where nothing changes we'd expect to have a result this significant about once every 11 times we ran the program

The most surprising thing about MAY-16-2000 is:

Normally 18.4% of records (233/1263) have sendng = KINCARDINE
But recently 24.9% of records (105/422) have sendng = KINCARDINE
Pvalue = 0.4005
Which means that in a world where nothing changes we'd expect to have a result this significant about once every 2 times we ran the program

The most surprising thing about MAY-17-2000 is:

Normally 12.0% of records (156/1305) have sendng = KINCARDINE and syndrome = other
But recently 16.9% of records (80/473) have sendng = KINCARDINE and syndrome = other
Pvalue = 0.444944
Which means that in a world where nothing changes we'd expect to have a result this significant about once every 2 times we ran the program

The most surprising thing about MAY-18-2000 is:


Normally 23.8% of records (352/1480) have sendng = KINCARDINE and age = D
But recently 32.2% of records (156/485) have sendng = KINCARDINE and age = D
Pvalue = 0.2005
Which means that in a world where nothing changes we'd expect to have a result this significant about once every 4 times we ran the program

The most surprising thing about MAY-19-2000 is:

Normally 0.0% of records (0/1265) have city = MISSISSAUGA
But recently 1.0% of records (4/407) have city = MISSISSAUGA
Pvalue = 1.0005
Which is thoroughly insignificant

Data	Tracking	Method
3 year data (excludes Hanover)	All data from everywhere	What's Strange About Recent Events

The period leading up to May 20th...
Nothing interesting or significant



Now the same analysis with the 3-month-including-hanover data...

WSARE

Data	Tracking	Method
3 month data (includes Hanover)	All data from everywhere	What's Strange About Recent Events

The most surprising thing about MAY-20-2000 is:

Normally 0.2% of records (2/1123) have city = WALKERTON and syndrome = gastrointestinal
But recently 6.5% of records (28/434) have city = WALKERTON and syndrome = gastrointestinal
Pvalue = 0.0005

Which means that in a world where nothing changes we'd expect to have a result this significant about once every 1999 times we ran the program

The most surprising thing about MAY-21-2000 is:

Normally 0.7% of records (7/1031) have sendng = W and syndrome = gastrointestinal
But recently 9.3% of records (43/463) have sendng = W and syndrome = gastrointestinal
Pvalue = 0.0005

Which means that in a world where nothing changes we'd expect to have a result this significant about once every 1999 times we ran the program

The most surprising thing about MAY-22-2000 is:

Normally 0.8% of records (12/1485) have sendng = W and syndrome = gastrointestinal
But recently 10.8% of records (51/474) have sendng = W and syndrome = gastrointestinal
Pvalue = 0.0005

Which means that in a world where nothing changes we'd expect to have a result this significant about once every 1999 times we ran the program

The most surprising thing about MAY-23-2000 is:

Normally 0.1% of records (1/1390) have city = WALKERTON and syndrome = gastrointestinal
But recently 7.8% of records (51/656) have city = WALKERTON and syndrome = gastrointestinal
Pvalue = 0.0005

Which means that in a world where nothing changes we'd expect to have a result this significant about once every 1999 times we ran the program

The most surprising thing about MAY-24-2000 is:

Normally 1.2% of records (18/1460) have sendng = W and syndrome = gastrointestinal
But recently 12.6% of records (76/603) have sendng = W and syndrome = gastrointestinal
Pvalue = 0.0005

Which means that in a world where nothing changes we'd expect to have a result this significant about once every 1999 times we ran the program

Results 1-5 of the 9 days in the 3 month date in which WSARE issued an alert with a Pvalue exceeding 1 in 1000

All 9 were during or just after the outbreak period

WSARE

Data	Tracking	Method
3 month data (includes Hanover)	All data from everywhere	What's Strange About Recent Events

The most surprising thing about MAY-26-2000 is:

Normally 4.9% of records (66/1347) have syndrome = gastrointestinal
But recently 22.0% of records (122/554) have syndrome = gastrointestinal
Pvalue = 0.0005

Which means that in a world where nothing changes we'd expect to have a result this significant about once every 1999 times we ran the program

The most surprising thing about JUN-14-2000 is:

Normally 5.0% of records (84/1666) have sendng = W and syndrome = other
But recently 13.4% of records (80/597) have sendng = W and syndrome = other
Pvalue = 0.00099975

Which means that in a world where nothing changes we'd expect to have a result this significant about once every 1000 times we ran the program

The most surprising thing about JUN-15-2000 is:

Normally 1.5% of records (28/1818) have sendng = CHES and syndrome = other
But recently 8.3% of records (45/544) have sendng = CHES and syndrome = other
Pvalue = 0.0005

Which means that in a world where nothing changes we'd expect to have a result this significant about once every 1999 times we ran the program

The most surprising thing about JUN-16-2000 is:

Normally 0.2% of records (3/1581) have sendng = GBHSTO and syndrome = other
But recently 4.4% of records (21/478) have sendng = GBHSTO and syndrome = other
Pvalue = 0.0005

Which means that in a world where nothing changes we'd expect to have a result this significant about once every 1999 times we ran the program

Results 6-9 of the 9 days in the 3 month date in which WSARE issued an alert with a Pvalue exceeding 1 in 1000

All 9 were during or just after the outbreak period

WSARE

Data	Tracking	Method
3 month data (includes Hanover)	All data from everywhere	What's Strange About Recent Events

The most surprising thing about MAY-15-2000 is:

Normally 3.1% of records (46/1477) have sendng = DURHAM
 But recently 6.9% of records (34/490) have sendng = DURHAM
 Pvalue = 0.1505
 Which means that in a world where nothing changes we'd expect to have a result this significant about once every 6 times we ran the program

The most surprising thing about MAY-16-2000 is:

Normally 0.0% of records (0/1392) have city = LONDON
 But recently 0.8% of records (4/475) have city = LONDON
 Pvalue = 0.5005
 Which means that in a world where nothing changes we'd expect to have a result this significant about once every 1 times we ran the program

The most surprising thing about MAY-17-2000 is:

Normally 10.8% of records (156/1442) have sendng = KINCARDINE and syndrome = other
 But recently 15.4% of records (80/520) have sendng = KINCARDINE and syndrome = other
 Pvalue = 0.6005
 Which is thoroughly insignificant

The most surprising thing about MAY-18-2000 is:

Normally 8.4% of records (135/1615) have sendng = HANOVER
 But recently 12.6% of records (70/555) have sendng = HANOVER
 Pvalue = 0.308192
 Which means that in a world where nothing changes we'd expect to have a result this significant about once every 3 times we ran the program

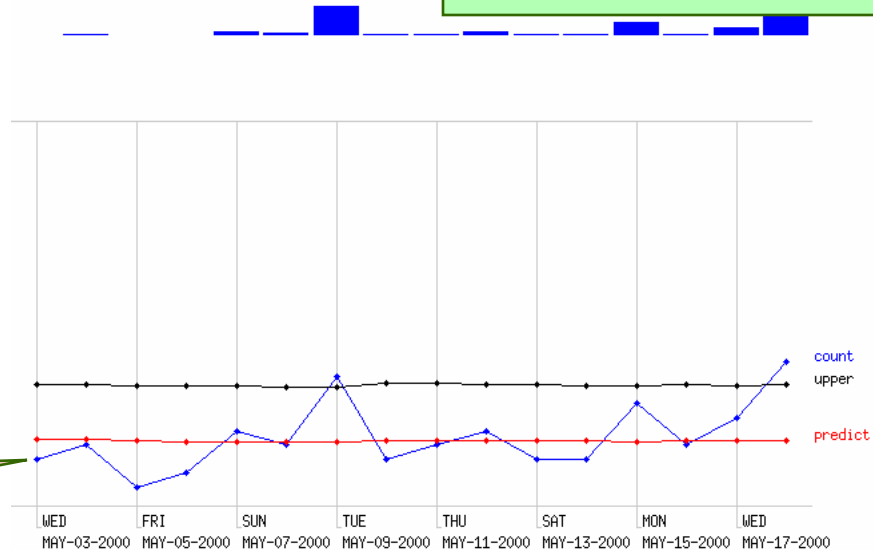
The most surprising thing about MAY-19-2000 is:

Normally 0.0% of records (0/1381) have city = MISSISSAUGA
 But recently 0.9% of records (4/460) have city = MISSISSAUGA
 Pvalue = 1.0005
 Which is thoroughly insignificant

Hanover Gastro cases

The period leading up to May 20th. May 18th is possibly of interest, though notice that it is not considered significant

Bars show alarm levels: max = 10



- <---step
- jump---
- step---
- zoomout
- print
- quit

What next?

- Want to try some other univariate methods
- Get better spatial coding of home locations?
- Should do a search for other syndromes/locations with increase early in outbreak
- Search on other spatial regions centered on Walkerton
- Run WSARE on data in which all the syndrome=other records are removed
- Analyse the chief complaint strings to see if there was a pattern in those strings in the days leading up to May 19th that is more specific than the GI syndrome coded by CoCo
- Get hold of data about which three-letter postcodes are in which water supply regions and allow home-water-region as another feature in the data
- Methods which look at multi-sized time windows (almost everything here looked one day at a time)
- See how sensitivity increases with multivariate methods that fuse ED with other data (absenteeism, over-the-counter sales, prescriptions, physician appointments...)