

Personal Reflection: Transition from a Data Analyst to a Data Reviewer

August 23, 2012

Duc Nguyen
Air Quality Meteorologist II
dnguyen@baaqmd.gov
415-749-4643



**BAY AREA AIR QUALITY
MANAGEMENT DISTRICT**

- Regulatory mandate
(attainment demonstration)
- Research purposes
- Monitor a situation /
Decision making support
- Provide trends
- Others
- Ultimately make the world a better
place!

WHY DO WE COLLECT DATA?



AQS data ARE NOT
“FINAL!”

**TRUST THE DATA NOT THE
SOURCE!**



Data Screening (as Data Analyst or Reviewer)



... is and should be a visual process!



What type of screenings do I look for?

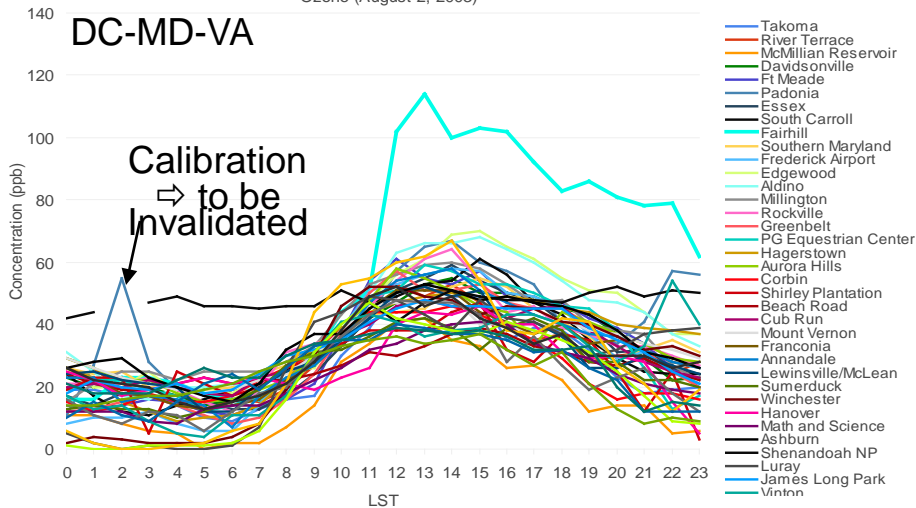
- Automatic: Min, max, avg, median, ROC, etc.
- **Temporal & spatial** trends, patterns, outliers.
 - Comparisons
 - Distributions
 - Correlations
- Data gaps
- NO sophisticated statistics.

Screenings are only done using
basic charts and tables!

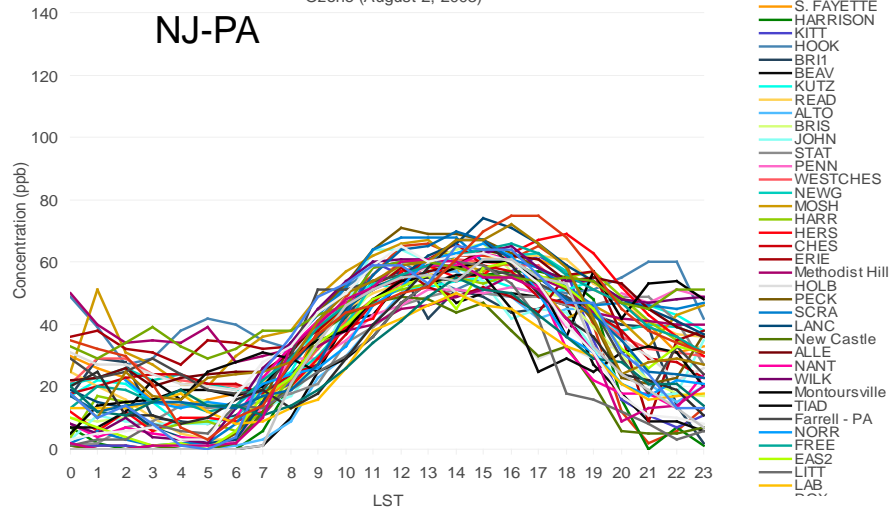


Examples of data screenings (1 of 4)

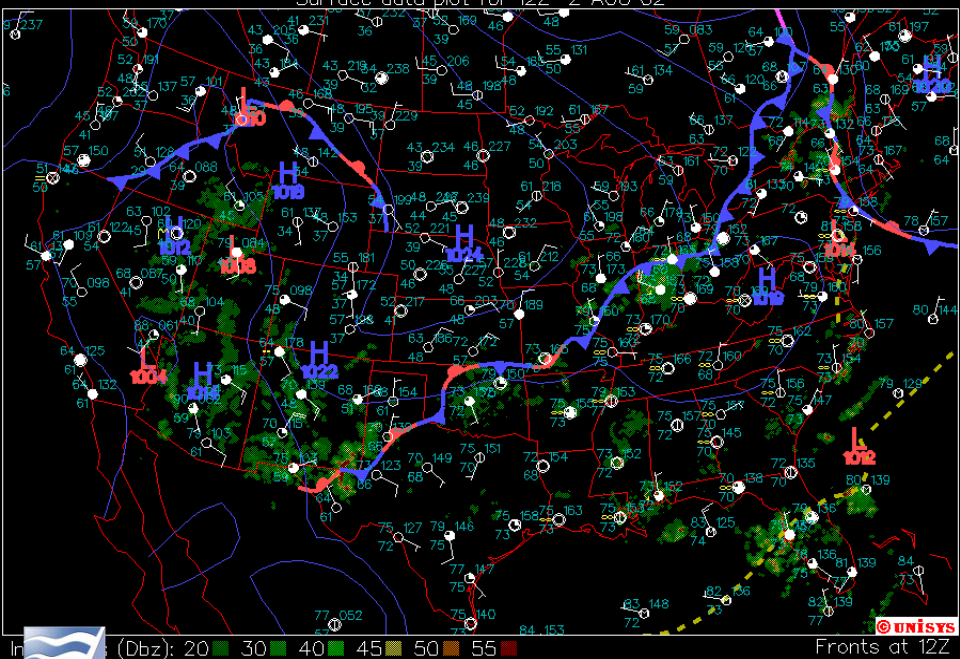
Ozone (August 2, 2003)



Ozone (August 2, 2003)



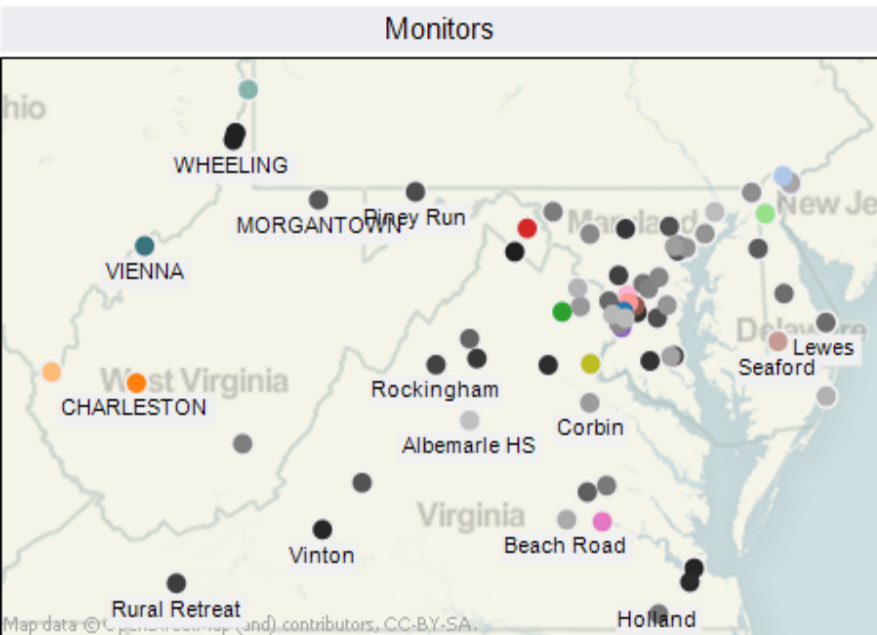
Surface data plot for 12Z 2 AUG 02



- Highly suspect ozone profile at Fairhill. Outside the norm for all sites in Mid-Atlantic.
 - Transport ozone?
 - Unique localized event?
- “Typical weather” couldn’t have done it alone.
- Based on my experience, no event like this ever occurred.
- Another value at Padonia (Calibration => invalid).

Examples of data screenings (2 of 4)

Outlier Ozone Data in DC, DE, VA & WV (2000-2010)

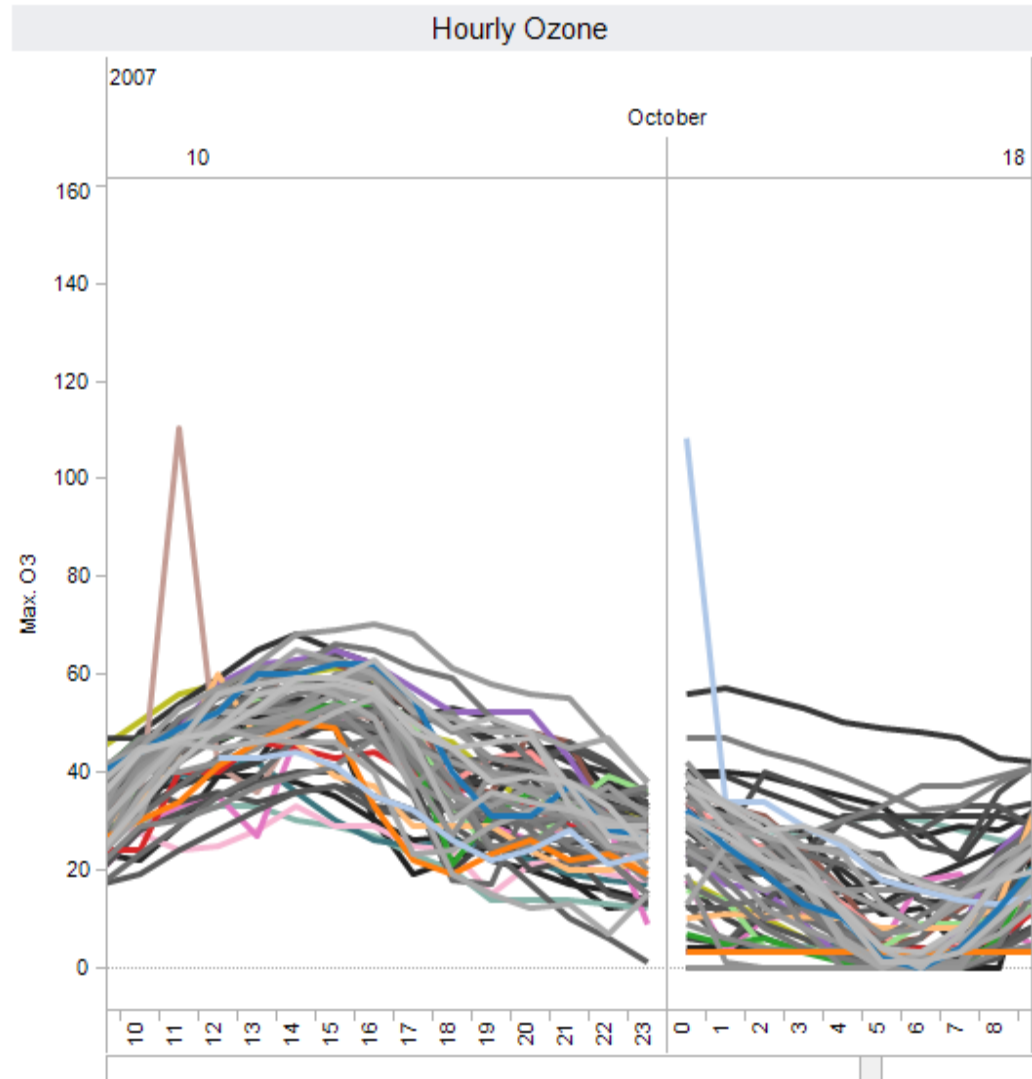


Site

<input type="checkbox"/> 516500008	<input type="checkbox"/> Ashburn	<input type="checkbox"/> BELLFNT2	<input type="checkbox"/> Cub Run/Cha
<input type="checkbox"/> Albemarle HS	<input type="checkbox"/> Assateague Islan..	<input type="checkbox"/> Calvert Co / Bars..	<input type="checkbox"/> Davidsonville
<input type="checkbox"/> Aldino	<input checked="" type="checkbox"/> Aurora Hills	<input checked="" type="checkbox"/> CHARLESTON	<input type="checkbox"/> Edgewood
<input type="checkbox"/> Alexandria	<input type="checkbox"/> BCSP	<input type="checkbox"/> Clifton Park	<input type="checkbox"/> Essex
<input type="checkbox"/> Annandale	<input type="checkbox"/> Beach Road	<input type="checkbox"/> Corbin	<input type="checkbox"/> Fairhill

Site

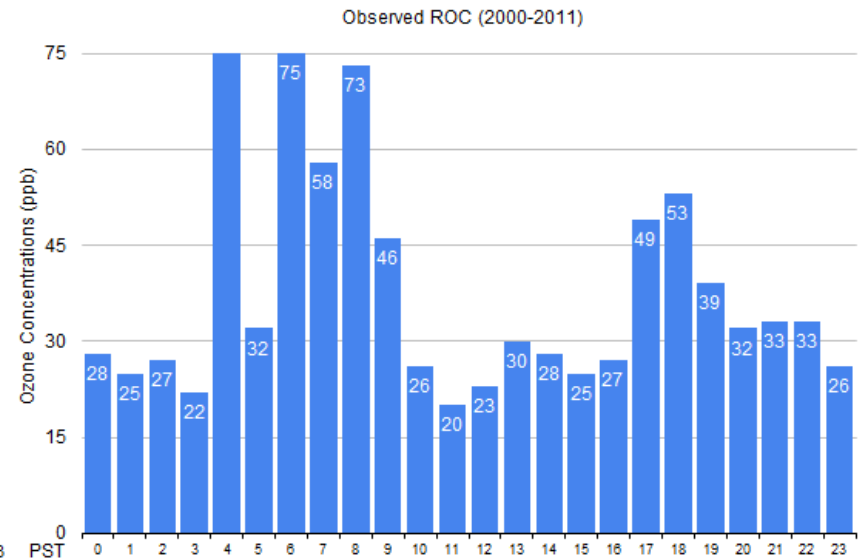
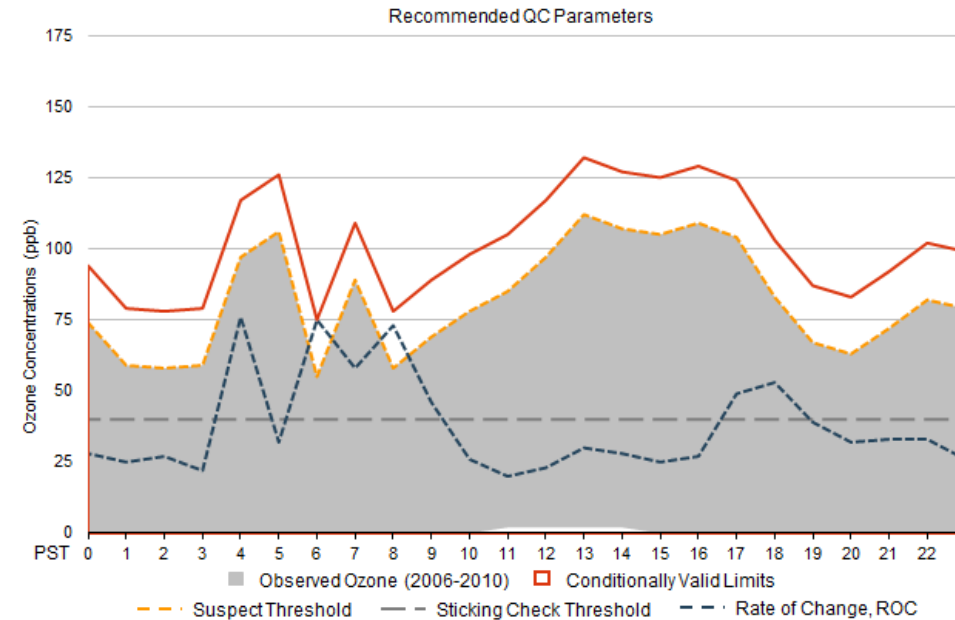
<input checked="" type="checkbox"/> 516500008	<input type="checkbox"/> Outlier
<input checked="" type="checkbox"/> Albemarle HS	<input checked="" type="checkbox"/> No
<input checked="" type="checkbox"/> Aldino	<input checked="" type="checkbox"/> Yes
<input checked="" type="checkbox"/> Alexandria	



Examples of data screenings (3 of 4)

AIRNow DMC's Real-Time Quality-Control (QC) Checks for 2012 ULATIS DRIVE, VACAVILLE (060953)

Ozone



NOTE: Interpolated data during calibration hours were also included in graphs.

	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23
Invalid Minimum	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Invalid Maximum	94	79	78	79	117	126	75	109	78	89	98	105	117	132	127	125	129	124	103	87	83	92	102	99
Suspect Threshold	74	59	58	59	97	106	55	89	58	69	78	85	97	112	107	105	109	104	83	67	63	72	82	79
ROC	28	25	27	22	76	32	75	58	73	46	26	20	23	30	28	25	27	49	53	39	32	33	33	26
Sticking Check Threshold	40	40	40	40	40	40	40	40	40	40	40	40	40	40	40	40	40	40	40	40	40	40	40	40

	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23
Minimum (2007-2011)	0	0	0	0	0	0	0	0	0	0	0	0	2	2	2	2	0	0	0	0	0	0	0	0
Maximum (2007-2011)	74	59	58	59	97	106	55	89	58	69	78	85	97	112	107	105	109	104	83	67	63	72	82	79
ROC (2000-2011)	28	25	27	22	76	32	75	58	73	46	26	20	23	30	28	25	27	49	53	39	32	33	33	26

Criteria for updating QC parameters

1. Invalid minimum: change to 0 ppb.
2. Invalid maximum: 20 ppb above 2005-2011 Maximum.
3. Suspect threshold: 2005-2011 hourly maximum.
4. ROC: 2000-2011 maximum hourly ROC.
5. Sticking check threshold: unchanged at 40 ppb.

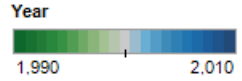
Screenings and visual analysis are happening at the same time!



Examples of data screenings (4 of 4)

Seasonal Pattern - 230090102

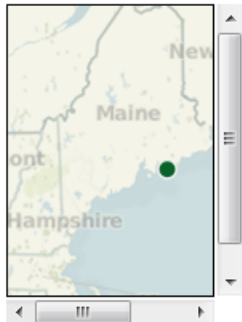
Diurnal Pattern - 230090102



AQSCode

- 090099005
- 090110008
- 090110124
- 090131001
- 230010014
- 230031100
- 230050014
- 230050027
- 230050029
- 230052003
- 230090001
- 230090101
- 230090102
- 230090103
- 230090301
- 230090401
- 230092003
- 230092004
- 230112001

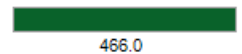
Map



Altitude

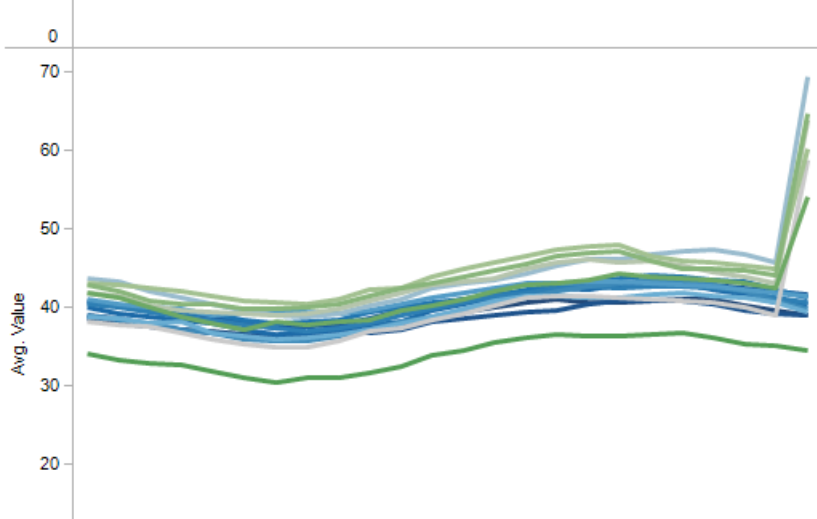
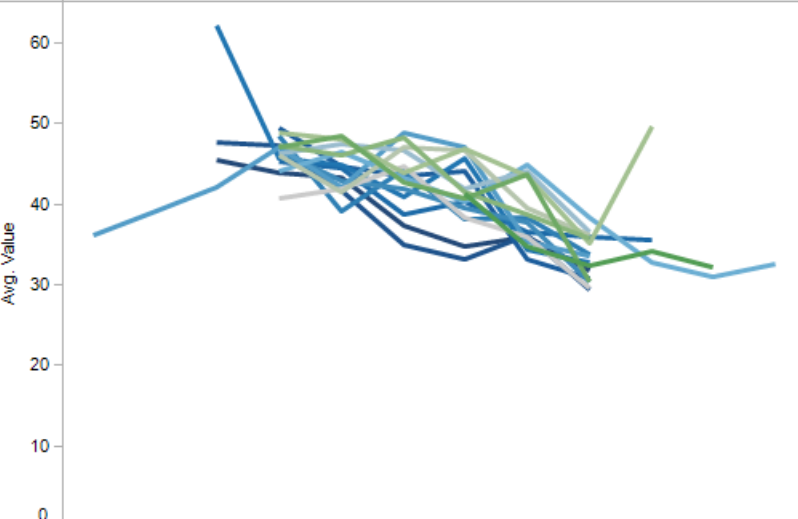
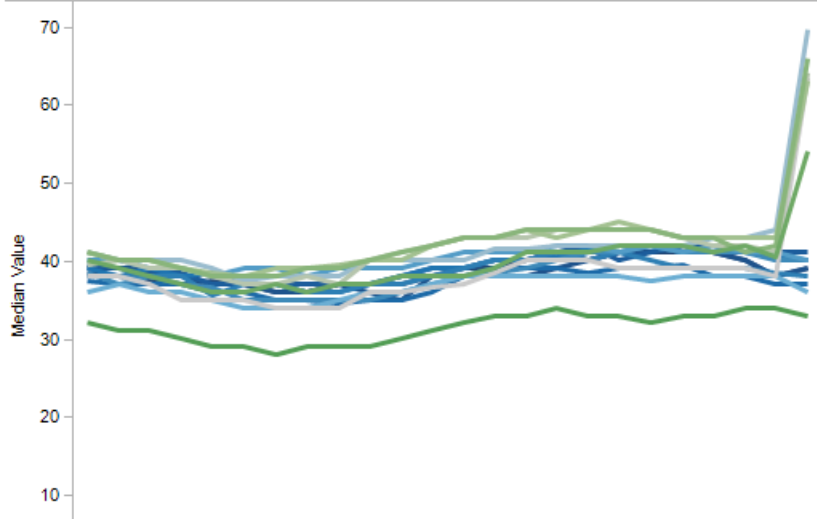
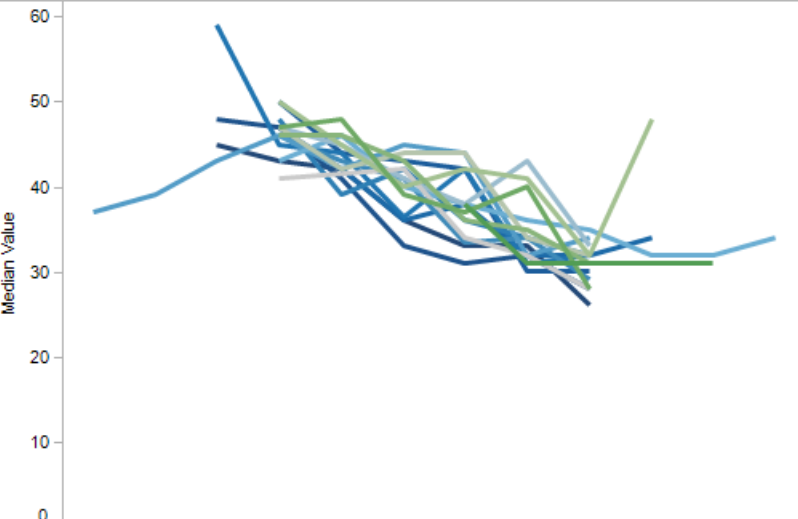


Altitude



SDateTime

SDateTime



January February March April May June July August September October November December

0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23

Data Screening is a visual process!

A dashboard
concept can
be applied to
data screening
and reviewing
process.

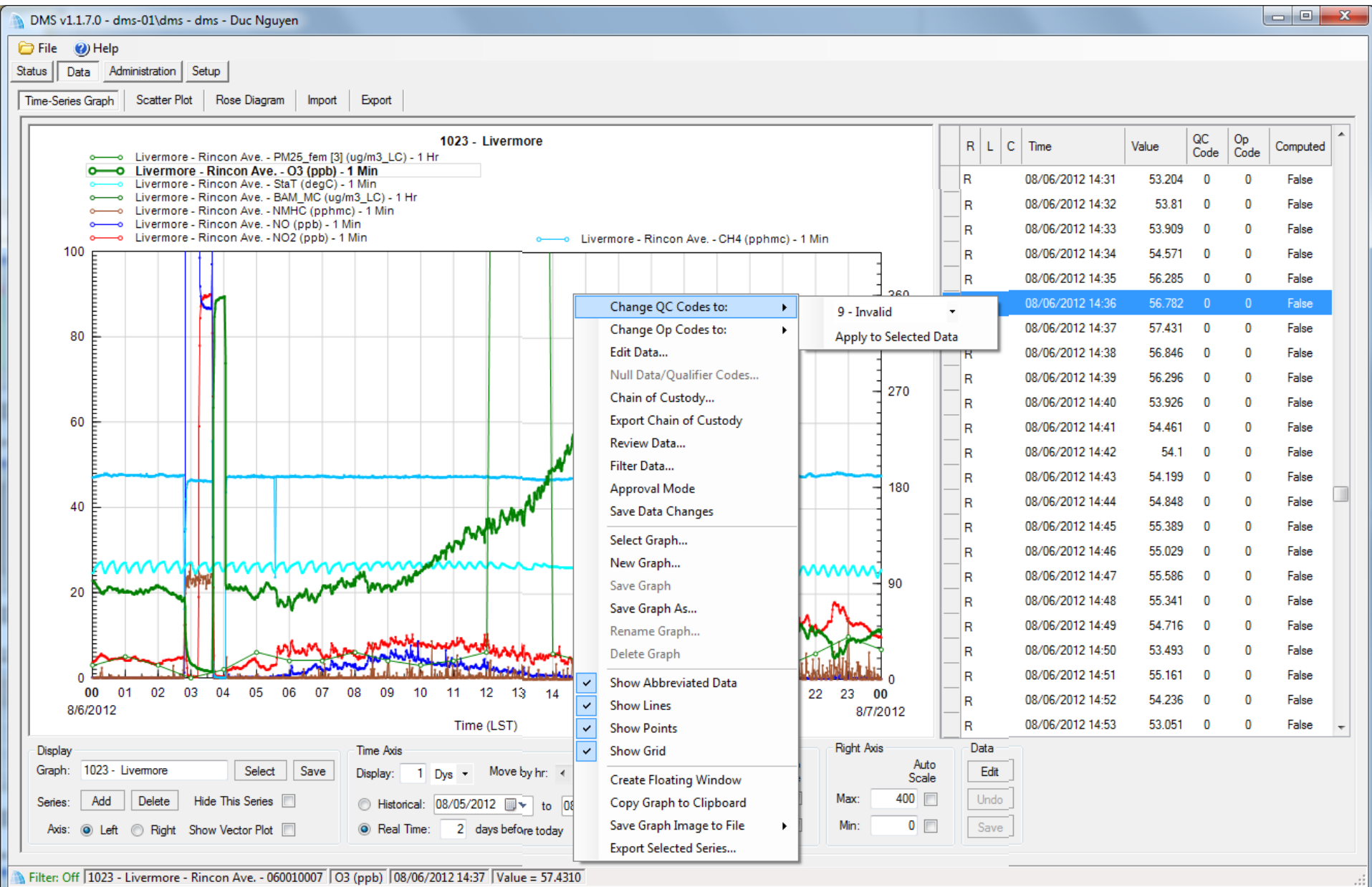


- NO Air Quality Data Management software is truly designed for data screening!
- Comparing to several other commercial software, DMS (not commercial) is the “best” although still limited by its design.

WARNING !!!



DMS



Sample Dashboard Display

WeatherSpark Beta | [Forecasts](#) | [Dashboard](#) | [Averages](#) | [More](#) | [Save](#)

<http://weatherspark.com> | San Francisco, CA, U... | Pleasant Hill, Calif... x

70°F

Mostly cloudy
At 1:56pm, now 2:09pm

San Francisco, CA, USA

Forecast | Daily | 1 quarter | 1 year | Averages

History

0.050" | 0.000" | 0.017" Me

80°F | 71°F 75th %-ile | 69°F | 61°F | 55°F

60°F | 67°F | 63°F | 57°F | 55°F | 53°F

50°F | 56°F | 53°F

99°F Max | 1pm Wednesday Aug 1, 2012 | 53°F Min

60°F 3°F Max | 58°F | 56°F | 54°F | 52°F | 50°F | 48°F

51°F | 51°F | 51°F | 50°F | 51°F | 51°F | 51°F | 53°F Mean

49°F | 49°F | 49°F

43°F Min

100% | 80% | 60% | 40% | 20% | 0%

84% | 86% | 63% | 63% | 83% | 86% | 86% | 86% | 67% | 55% Mean

Humidity

86%

Wind

6a 9a 12p 3p 6p 9p 12a 3a 6a 9a 12p 3p 6p 9p

Tue Jul 31, 2012 | Wed Aug 1, 2012 | Thu Aug 2, 2012

1950s | 1960s | 1970s | 1980s | 1990s | 2000s

Intuitive Data Selection

Map Overlay

- Radar (US only)
- Hide

Weather Stations

- Temperature
- Dew Point
- Wind
- Pressure
- Humidity
- Hide

Beautiful Graphs and Data Displays

Intuitive Data Selection

Icons

- Sun & Moon
 - Sun & Moon Options
- Clouds
- Precipitation Probability
- Precipitation Amount
- Precipitation Rate
- ColorTemp
- Temperature
 - Temperature Options
- Dew Point
 - Dew Point Options
- Humidity
- Pressure
- Wind Direction
- Wind Speed
 - Wind Speed Options
- Wind Icons

Show All

Reset to Defaults

Spatial Display

Beautiful Wind Vector

Seamless Navigation

POWERED BY Google

Map data ©2012 Google - [Terms of Use](#)

Software features that will improve data screening ...

- Geospatial views
- Multi-chart displays
- Default and user defined formatting (colors, chart design, labels, scales and other visual components)
- Interactivity (navigation, zooming, panning, highlighting, rescaling, structured aggregation, data on demand, time synchronization, etc.)
- “Real” dashboard support with linked data and display interface.



