Automating Quality Assurance Assessments

Manual Prep. For TSAs

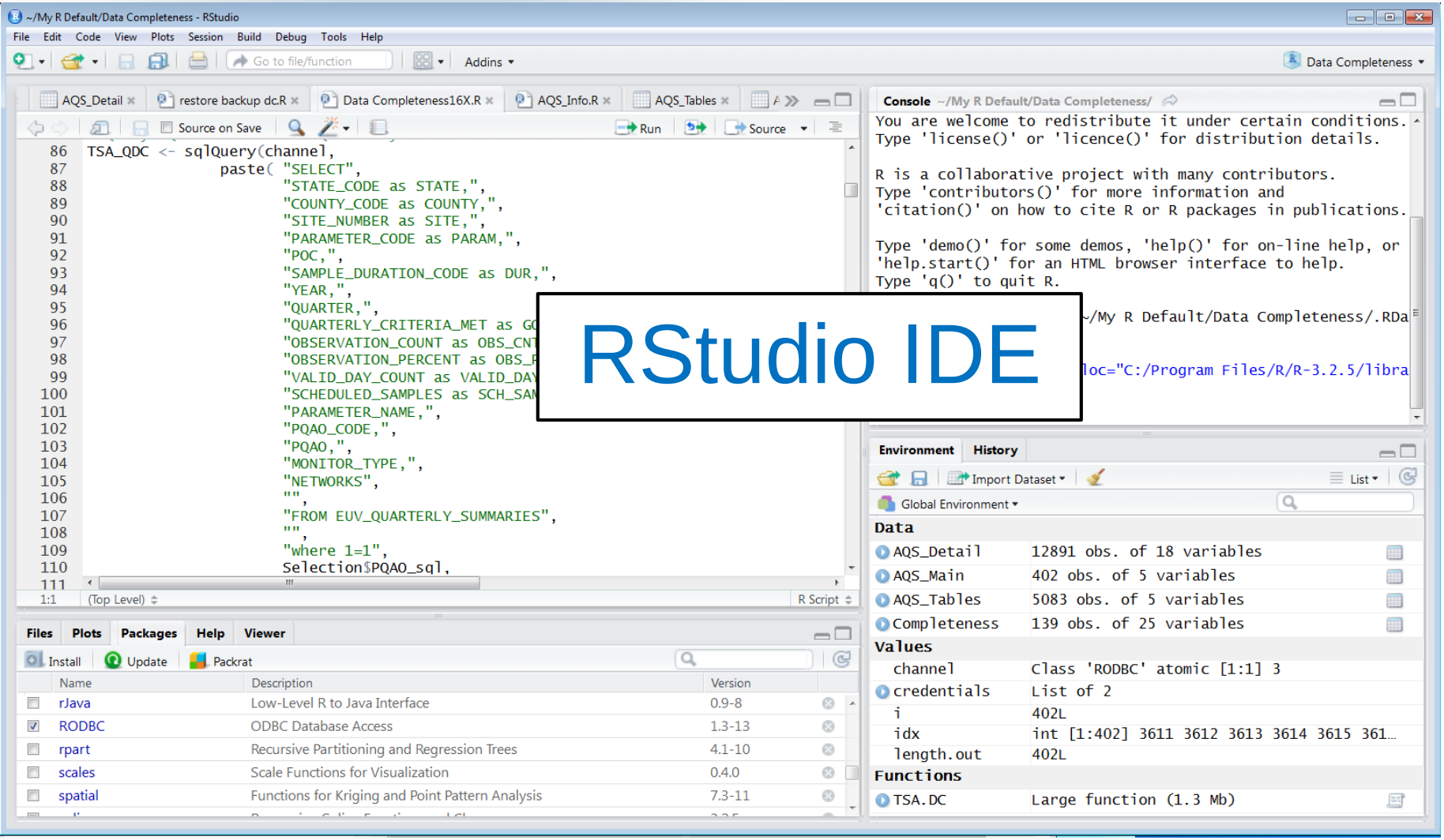
- AQS AMP Reports
  - Like FRM's ("Gold Standard")
  - Like FRM's (slow, labor intensive)
  - Most are PDF's intended for hardcopy
  - Many AMP Reports are needed for TSA assessments
    - 100's to 1000's of pages

Semi-Automated Prep. For TSAs

- AMP-504 -- Extract QA Data
  - Used for uploading QA Data to AQS
  - Can be imported with Spreadsheets for Sorting and Filtering results to identify monitor*dates of failed checks
  - Manually reconcile back to AMP-350

Automated QA Assessments

- Using the “R” scripting language
- R is “*Procedural*” code
  - Re-enforces Consistency in Assessments
  - Assessment Process is Transparent
- R is “Free”
- Leveraging work products made by NADG (AQS Team) for SQL access to AQS tables
  - End User Views
  - Required to be inside EPA Firewall


~/My R Default/Data Completeness - RStudio
File Edit Code View Plots Session Build Debug Tools Help
Addins
Data Completeness
AQS_Detail
restore backup dc.R
Data Completeness16X.R
AQS_Info.R
AQS_Tables
Source on Save
Run
Source
TSA_QDC <- sqlQuery(channel,
paste("SELECT",
"STATE_CODE as STATE,",
"COUNTY_CODE as COUNTY,",
"SITE_NUMBER as SITE,",
"PARAMETER_CODE as PARAM,",
"POC,",
"SAMPLE_DURATION_CODE as DUR,",
"YEAR,",
"QUARTER,",
"QUARTERLY_CRITERIA_MET as GO
"OBSERVATION_COUNT as OBS_CNT
"OBSERVATION_PERCENT as OBS_P
"VALID_DAY_COUNT as VALID_DAY
"SCHEDULED_SAMPLES as SCH_SAM
"PARAMETER_NAME,",
"PQAO_CODE,",
"PQAO,",
"MONITOR_TYPE,",
"NETWORKS",
"",
"FROM EUV_QUARTERLY_SUMMARIES",
"",
"where 1=1",
Selection$PQAO_sql,
1:1
(Top Level)
R Script
Console ~/My R Default/Data Completeness/
You are welcome to redistribute it under certain conditions.
Type 'license()' or 'licence()' for distribution details.
R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.
Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.
~/My R Default/Data Completeness/.RDa
loc="C:/Program Files/R/R-3.2.5/libra
RStudio IDE
Environment History
Import Dataset
List
Global Environment
Data
AQS_Detail 12891 obs. of 18 variables
AQS_Main 402 obs. of 5 variables
AQS_Tables 5083 obs. of 5 variables
Completeness 139 obs. of 25 variables
Values
channel Class 'RODBC' atomic [1:1] 3
credentials List of 2
i 402L
idx int [1:402] 3611 3612 3613 3614 3615 361...
length.out 402L
Functions
TSA.DC Large function (1.3 Mb)
Files Plots Packages Help Viewer
Install Update Packrat
Name Description Version
rJava Low-Level R to Java Interface 0.9-8
RODBC ODBC Database Access 1.3-13
rpart Recursive Partitioning and Regression Trees 4.1-10
scales Scale Functions for Visualization 0.4.0
spatial Functions for Kriging and Point Pattern Analysis 7.3-11

R – Example Code

```
# Douglas Jager
# US EPA Region 4 SESD

library(RODBC)
library(reshape2)
library(dplyr)
library(XLConnect)
credentials <- list(userid=NULL, passwd=NULL)
source("C:/Users/DJAGER/Documents/My R Default/AQS_Credentials.R", chd

setwd("C:/Users/DJAGER/Documents/My R Default/Data Completeness")



Selection   <- list(Title_Row = "Data Completeness", SubTitle_Row = "Pr
                   st yr = 2013, end.yr = 2015,
                   PQAO = "1025",
                   State = "", County = "",
                   Region = "04")


# Defining / Refining Selection$ List for SQL Criteria
# if Region is "1" to "10" then Selection$State_list gets defined as be
#    if State is also input, these if-else if statements will be over
if (Selection$Region == "01") {
  Selection$State_list <- "IN ('09', '23', '25', '33', '44', '50')"
} else if (Selection$Region == "02") {
  Selection$State_list <- "IN ('34', '36', '72', '78')"
```

CONS

R Scripts can be intimidating at first

PROS

Easier to Learn than Python or VBA

R Script called from Function

```r
credentials <- list(userid=NULL, passwd=NULL)
setwd("C:/Users/DJAGER/Documents/My R Default/Data Completeness")

source("C:/Users/DJAGER/Documents/My R Default/AQS_Credentials.R", chdir = TRUE)

source("TSA.DC.Function -R0.3.R")


# Example input for Function

# TSA.DC(startYR = 2013,              # required user input (numeric)
#        endYR = 2015,                # required user input (numeric)
#        PQAO="",                     # optional user input (quoted text),
#        St_Code="01",                # optional user input (quoted text),
#        Cnty_Code="089",             # optional user input (quoted text),
#        Reg_Code="04",               # optional user input (quoted text),
#        Title_A="Main Title",        # Default Title_A
#        Title_B="Subtitle",          # Default Title_B
#        userid=credentials$userid,   # Default AQS userid from Credentials
#        passwd=credentials$passwd)   # Default AQS passwd from Credentials

# function returns completeness as both R data.frame as well as exported ex
#                                  with file name: "Data.Completeness" appended with query date and extension .xlsx

#                                        example: "Data.Completeness.2016-03-09.xlsx"


Completeness <- TSA.DC(startYR = 2013,
                       endYR = 2015,
                       PQAO="",
                       St_Code="",
                       Cnty_Code="",
                       Reg_Code="04")
```

Only have to edit once

Using a Function to Shield the User from "Intimidating" script body.

Inputs to Function "TSA.DC()"

R dataframe to Excel (raw)

Automated Data Completeness Assessment

R Program Connecting Directly to AQS Oracle Tables, then Exporting to Excel

Query Date: 2016-03-29

							2013					2014					2015
AQS Site	Param	Name	POC	Dur	Monitor Type	Network	Qtr 1	Qtr 2	Qtr 3	Qtr 4	Annual	Qtr 1	Qtr 2	Qtr 3	Qtr 4	Annual	Qtr 1
13-247-0001	42601	NO	1	1	SLAMS	UNOFFICIAL PAMS	97	97	97	98	97	98	98	97	97	97	98
13-247-0001	42602	NO2	1	1	SLAMS	UNOFFICIAL PAMS	97	97	97	98	97	98	98	97	97	97	98
13-247-0001	42603	NOX	1	1	SLAMS	UNOFFICIAL PAMS	97	97	97	98	97	98	98	97	97	97	98
13-247-0001	44201	O3	1	W	SLAMS	UNOFFICIAL PAMS	34	100	100	33	100	35	99	99	33	99	35
13-261-1001	44201	O3	2	W	SLAMS	--	100	100	100	100	100	100	99	100	100	100	100
13-295-0002	88101	LC25	1	7	SLAMS	--	80	65	77	90	78	97	100	97	100	98	100
13-295-0002	88502	PM2.5-AQI	3	X	OTHER	--	97	99	100	100	99	93	91	99	100	96	97
13-303-0001	81102	--	1	--	--	--	--	--	--	--	--	--	--	--	--	--	--
13-303-0001	88101	LC25	1	7	SLAMS	--	90	81	97	94	90	80	93	97	100	93	97
13-319-0001	88101	LC25	1	7	SLAMS	--	87	81	90	87	86	80	87	97	94	89	97
21-013-0002	44201	O3	1	W	SPM	--	100	100	100	100	100	100	97	95	100	96	100
21-013-0002	88101	LC25	1	7	SPM	--	93	100	100	100	98	100	93	100	100	98	93
21-013-1002	44201	O3	1	W	NON-EPA FEDEI	--	--	69	94	--	59	0	100	100	--	75	--
21-015-0003	44201	O3	1	W	SLAMS	--	100	100	99	100	100	100	100	100	100	100	100
21-019-0002	81102	PM10	1	7	SLAMS	--	98	99	98	100	98	87	100	100	94	95	93
21-019-0002	81102	PM10	2	7	SLAMS	--	100	100	100	100	100	100	100	100	88	97	100
21-019-0016	14129	Pb-LC	1	7	--	--	100	100	53	--	62	--	--	--	--	--	--
21-019-0017	42401	SO2	1	1	SLAMS	--	99	100	99	99	99	99	99	99	99	99	99
21-019-0017	42401	SO2	1	Y	SLAMS	--	98	99	99	98	98	98	98	99	99	98	99
21-019-0017	42406	SO2M	1	1	SLAMS	--	99	100	99	99	99	99	99	99	99	99	99
21-019-0017	42601	NO	1	1	SLAMS	--	94	94	91	93	93	94	95	94	91	93	95
21-019-0017	42602	NO2	1	1	SLAMS	--	94	94	91	93	93	94	95	94	91	93	95
21-019-0017	44201	O3	1	W	SLAMS	--	100	100	100	100	100	100	100	100	100	100	100
21-019-0017	88101	LC25	1	7	SLAMS	--	100	100	97	90	97	90	100	100	100	98	100
21-019-0017	88502	PM2.5-AQI	3	X	SPM	--	100	96	100	100	99	99	99	96	99	98	97
21-029-0006	44201	O3	1	W	SLAMS	--	100	100	97	100	99	100	100	84	71	90	100

Automated Data Completeness Assessment

R Program Connecting Directly to AQS Oracle Tables, then Exporting to Excel

Query Date: 2016-03-29							2013					2014					
AQS Site	Param	Name	POC	Dur	Monitor Type	Network	Qtr 1	Qtr 2	Qtr 3	Qtr 4	Annual	Qtr 1	Qtr 2	Qtr 3	Qtr 4	Annual	Qtr 1
37-189-0003	88101	LC25	1	7	SLAMS	--	97	100	93	100	98	100	97	87	94	94	90
37-189-0003	88101	LC25	3	X	SPM	--	--	--	--	--	--	--	--	--	--	--	61
37-191-0005	88101	LC25	1	7	SLAMS	--	100	100	100	100	100	100	100	93	100	98	100
37-191-0005	88101	LC25	3	X	SPM	--	100	91	98	100	97	89	100	83	100	93	--
37-191-0005	88501	--	3	--	--	--	--	--	--	--	--	--	--	--	--	--	--
37-191-0005	88502	PM2.5-AQI	3	X	SLAMS	--	--	--	--	--	--	--	--	--	66	20	100
37-199-0004	44201	O3	1	W	SPM	--	--	96	99	100	98	--	82	99	100	93	--
45-001-0001	44201	O3	1	W	SLAMS	--	--	93	99	97	95	--	77	90	94	85	--
45-003-0003	44201	O3	2	W	SLAMS	--	--	97	99	100	98	--	99	99	84	95	--
45-007-0004	88501	--	3	--	--	--	--	--	--	--	--	--	--	--	--	--	--
45-007-0005	44201	O3	1	W	SLAMS	--	--	99	99	100	99	--	99	99	100	99	--
45-015-0002	44201	O3	1	W	SLAMS	--	--	99	99	90	98	--	97	100	90	97	--
45-019-0003	14129	--	2	--	--	--	--	--	--	--	--	--	--	--	--	--	--
45-019-0003	42401	SO2	1	1	SLAMS	--	88	99	99	99	96	99	99	99	99	99	99
45-019-0003	42401	SO2	1	Y	SLAMS	--	87	99	99	99	96	99	99	98	99	99	99
45-019-0003	42401	SO2M	3	H	SLAMS	--	--	--	--	--	--	--	--	97	96	49	97
45-019-0003	42602	NO2	2	1	SPM	--	88	99	99	99	96	88	99	99	99	96	99
45-019-0003	81102	PM10	3	X	SLAMS	--	99	99	90	100	97	100	100	100	100	100	97
45-019-0006	88101	--	1	--	--	--	--	--	--	--	--	--	--	--	--	--	--
45-019-0008	88101	--	1	--	--	--	--	--	--	--	--	--	--	--	--	--	--
45-019-0009	88101	--	1	--	--	--	--	--	--	--	--	--	--	--	--	--	--
45-019-0046	42101	--	1	--	--	--	--	--	--	--	--	--	--	--	--	--	--
45-019-0046	42401	SO2	2	1	SPM	--	99	99	99	97	98	98	99	38	98	83	98
45-019-0046	42401	SO2	2	Y	SPM	--	98	99	99	96	98	97	99	37	97	82	97
45-019-0046	42401	SO2M	5	H	SPM	--	--	--	--	--	--	--	--	37	96	33	96
45-019-0046	42602	NO2	1	1	SPM	--	66	65	99	78	77	91	77	72	95	84	62
45-019-0046	44201	O3	1	W	SLAMS	--	--	99	82	97	91	--	84	99	97	92	--

LEGEND

- 75% to 79% Data Completeness
- 0% to 74% Data Completeness
- Missing Required Non-NAAQS Parameter
- NAAQS Exclusion

R -- PQAO Network Summary

Regulatory Network

Demonstration Project: PQAO Network Summary

State	PQAO	Lead (TSP) LC 14129	Carbon monoxide 42101	Sulfur dioxide 42401	Nitrogen dioxide 42602	Ozone 44201	PM10 81102	PM10-Cont 81102-C	PM2.5 88101
47	0170	0	0	0	0	2	0	0	3
47	0581	4	0	0	0	2	2	0	5
47	0673	1	3	1	1	3	3	0	4
47	0682	0	1	1	2	2	3	0	3
47	0745	0	1	2	1	3	0	0	0
47	1025	2	0	5	0	9	0	1	14
47	1026	0	1	1	0	0	0	0	0
47	1344	0	0	0	0	2	0	0	0

Query Date:
07-05-2016

Industrial Monitor

PQAO Site Summary

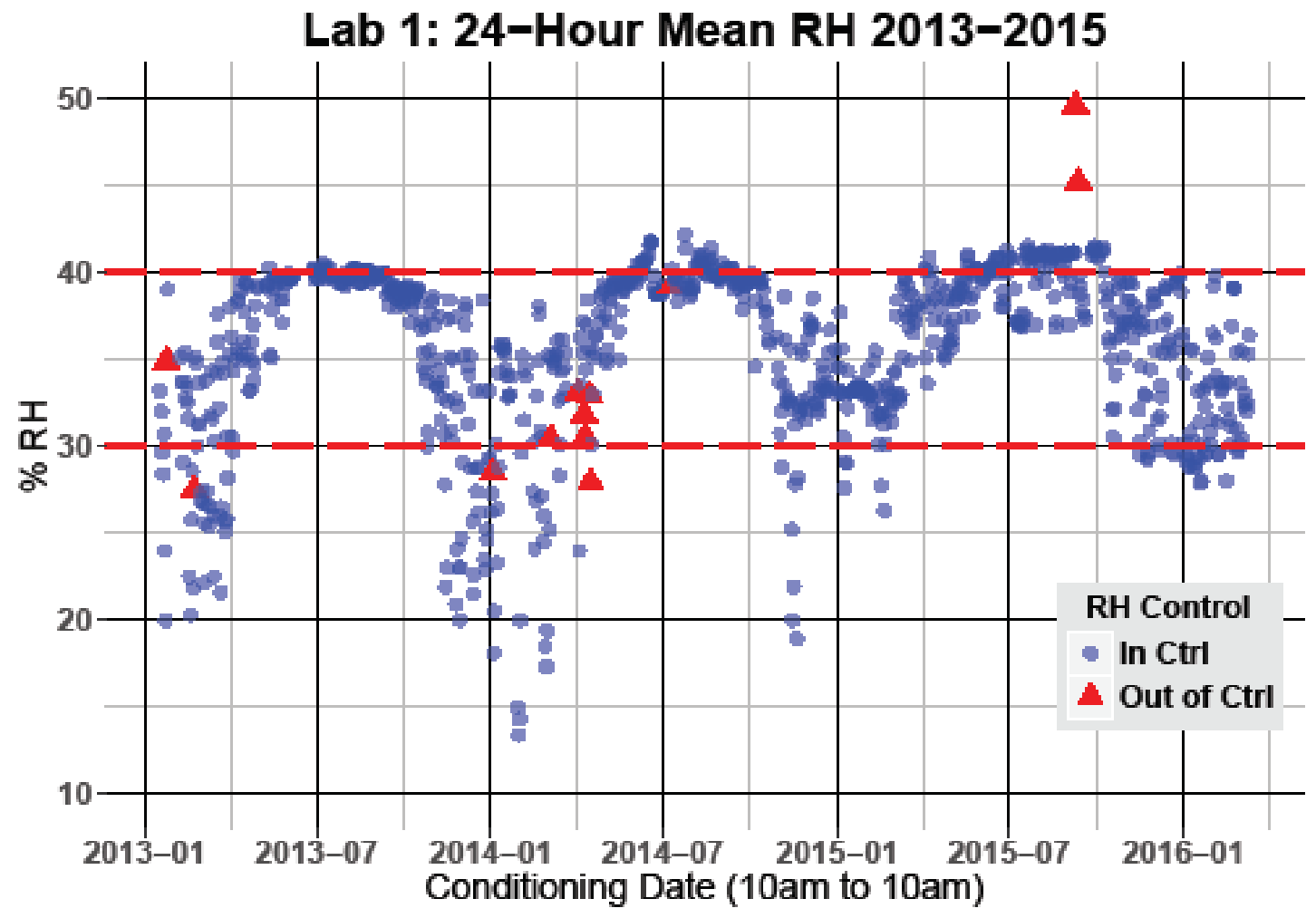
State	PQAO	AQS ID	Lead (TSP) LC 14129	Carbon monoxide 42101	Sulfur dioxide 42401	Ozone 44201	PM10-Cont 81102-C	PM2.5 88101
47	1025	47-001-0101			X	X		
47	1025	47-009-0011						X
47	1025	47-009-0102				X		
47	1025	47-011-0102			X			
47	1025	47-045-0004						X
47	1025	47-089-0002				X		
47	1025	47-099-0002						X
47	1025	47-105-0108				X		X
47	1025	47-107-0101			X			
47	1025	47-107-1002						X
47	1025	47-113-0006						X
47	1025	47-119-2007						X
47	1025	47-125-1009						X
47	1025	47-141-0005						X
47	1025	47-145-0004						X
47	1025	47-163-0007		X	X			
47	1025	47-163-1007						X
47	1025	47-163-2002				X		
47	1025	47-163-2003				X		
47	1025	47-163-3004	X					
47	1025	47-165-0007				X		X
47	1025	47-173-0107					X	
47	1025	47-187-0106				X		
47	1025	47-189-0103				X		

PQAO Site Summary
(/w method codes)

State	PQAO	AQS ID	Lead (TSP) LC 14129	Carbon monoxide 42101	Sulfur dioxide 42401	Ozone 44201	PM10-Cont 81102-C	PM2.5 88101
47	1025	47-001-0101			100	087		
47	1025	47-009-0011						118
47	1025	47-009-0102				053		
47	1025	47-011-0102			600			
47	1025	47-045-0004						118
47	1025	47-089-0002				087		
47	1025	47-099-0002						118
47	1025	47-105-0108				087		118
47	1025	47-107-0101			600			
47	1025	47-107-1002						118
47	1025	47-113-0006						118
47	1025	47-119-2007						118
47	1025	47-125-1009						118
47	1025	47-141-0005						118
47	1025	47-145-0004						118
47	1025	47-163-0007		054	060			
47	1025	47-163-1007						118
47	1025	47-163-2002				087		
47	1025	47-163-2003				087		
47	1025	47-163-3004	192					
47	1025	47-165-0007				047		118
47	1025	47-173-0107					079	
47	1025	47-187-0106				047		
47	1025	47-189-0103				047		

LEGEND		
Param	Method	Equipment
14129	192	Pb-TSP/ICP SPECTRA (ICP-MS)
42101	054	THERMO ELECTRON 48, 48C, 48i
42401	060	THERMO ELECTRON 43A, 43B, 43C
42401	100	API MODEL 100 A SO2 ANALYZER
42401	600	Teledyne API 100 EU
44201	047	THERMO ELECTRON 49
44201	053	MONITOR LABS 8810
44201	087	MODEL 400 OZONE ANALYZER
81102	079	RUPRCHT&PATSHNCK TEOM SER 1400
88101	118	R & P CO PLUS MODEL 2025PM SEQ

Control Charting with R

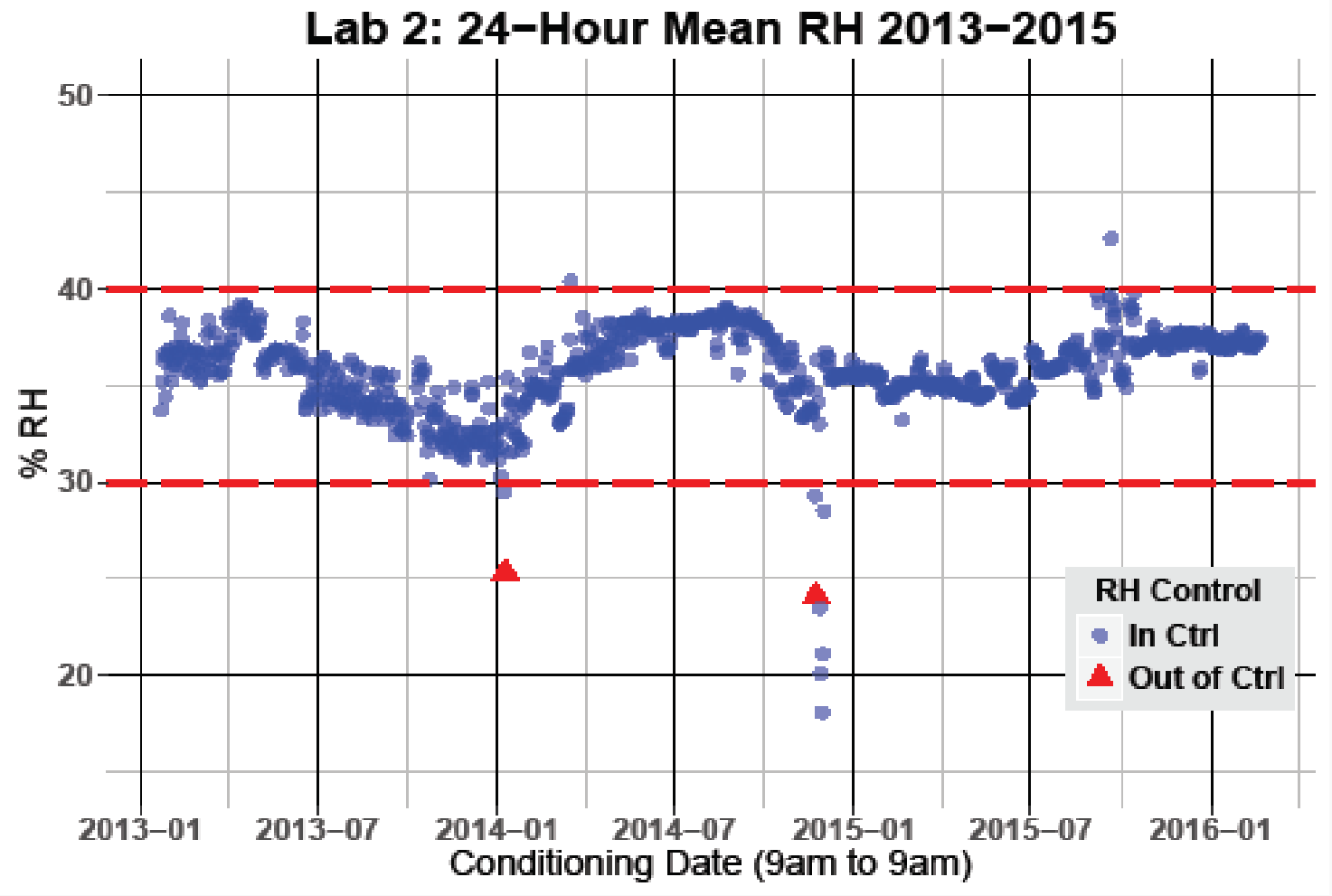


Daily RH Means (Lab 1)

Computed from:

- 150 Tab Delimited files
- 238,425 minute readings

Control Charting with R



Daily RH Means (Lab 2)

Computed from:

- 528 Tab Delimited files
- 7,467,869 minute readings

R -- QA Assessment Tools in Development

- Flowrate Verifications and Audits
  - Reconciling results with routine measurements
- 1-Pt Precision and Annual Performance Eval. Audits
  - Reconciling results with routine measurements
- Inspection of routine measurements for Test Atmospheres
  - Time of day
  - Extreme Values vs Null Data Coding
  - Rate of Change
- AQS Metadata for Appendix E
- Export to ArcGIS or Google Earth KML for Site Evaluations

Why Automate QA Assessments?

Decreasing resources and increasing demands require that we improve the efficiency and effectiveness of data quality assessments.

Automated Data Analysis Tools:

- Drive consistency in data evaluations
- Enhance the speed in performing QA assessments
- Liberate limited staffing resources for other high value activities in TSAs

& enable R4 to do Quarterly QA Assessments like R1

Doug Jager

jager.doug@epa.gov

706.355.8618