ESIP: VM Setup Help and Data Analysis

Presenter: Jill Minor
Indiana University Bloomington, USA

Help: Accessing Azure Windows VMs

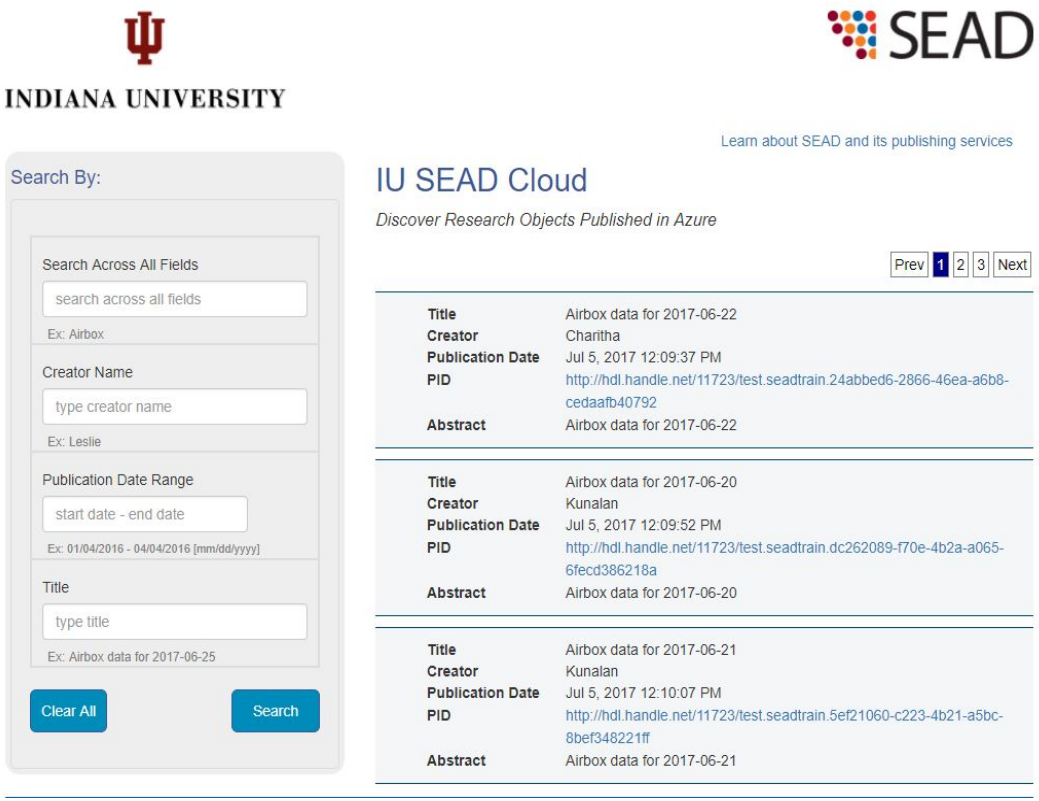
- Receive a username and password
- Open the link, download to your laptop, open the file and click Connect . (**Mac users will need to download Microsoft Remote Desktop in App store**)
- Note: **Windows OS Users**: In the credentials login you need to change the username by clicking on ‘more choices’-->’Use a different account’ and enter the provided username
- Click Yes on the Security Certificate Confirmation screen

Analysis: Retrieving PIDs and Metadata

Explore IU SEAD CLOUD AZURE SEARCH INTERFACE

http://d2i-dev.d2i.indiana.edu:8081/iusc-azure-search/search.html

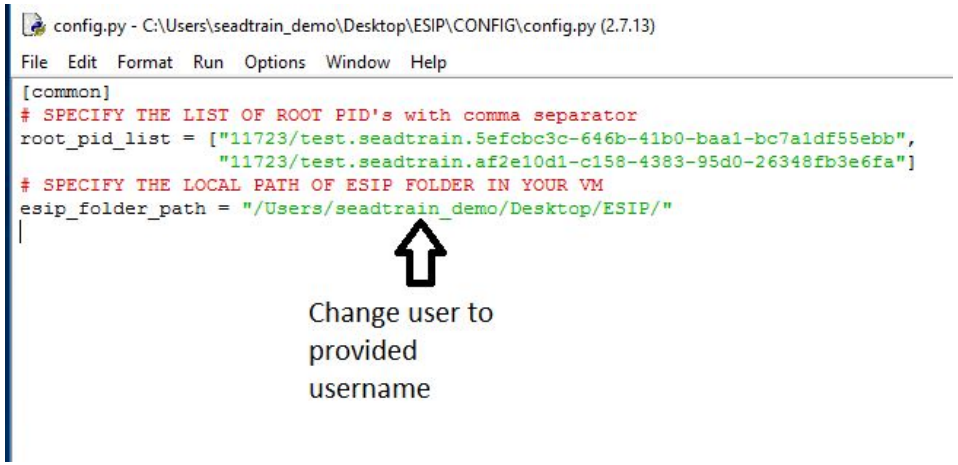
- Open up browser in vm
- Go to the link: **https://tinyurl.com/SEADTrain-search**
- Explore the Metadata.



Analysis: Updating PIDs

- Let's grab the most recent PID and add it to the CONFIG file on the VM.
  - In the Publication Date Range box enter in today's date for the start date and tomorrow's date for the end date.
  - Copy the PID
  - Open the Desktop/ESIP folder to paste it in the **root_pid_list** field of CONFIG/config.py
  - Update the path in **esip_folder_path** field, only changing the username is necessary.

Note: The existing PIDs in Config file are samples. You can replace these with the new PIDs using Discovery UI.



Analysis: Running the Codes

1. Double click on pid_resolver.py to execute the program
    - Note the creation of two new folders within ESIP: PID-RESOLVER-FILES-OUTPUT and PID-RESOLVER-METADATA-OUTPUT
2. In the ANALYSIS subfolder within ESIP:
    - Double Click on csv-convert.py to create csv files from the text files.
    - Double Click on csv-merge.py to combine all of the separate device data files within one csv for each day
    - Double click on the final-csv-merge.py to create ‘final.csv’ for all devices and all days

Analysis: Data Fields

Note: We have added a PID field to the data that records the Child PID for that device on that day

Fields	Measurement & unit	Data type
PID	URI handle	string
device_id	12 character	string
date	year-month-day	date
time	hour:min:sec	time
device	name : LinkIt_Smart_7688_Duo	String:
s_t4	temperature : Celsius	float: %.2f
s_h4	relative humidity : %	float: %.2f
s_b2	barometer : [millibars]	float: %.6f
s_d2	dust sensor PM1: [ug/cm^3]	integer: 2 sig figs
s_d0	dust sensor PM2.5: [ug/cm^3]	integer: 2 sig figs
s_d1	dust sensor PM10 : [ug/cm^3]	integer: 2 sig figs
d_t5	device temperature	float: % .2f
d_h5	device humidity	float: %.2f
gps_lat	latitude	float : %.6f
gps_lon	longitude	float : %.6f
gps_fix	= 1	Integer
gps_num	# of satellites in gps fix = 15	Integer

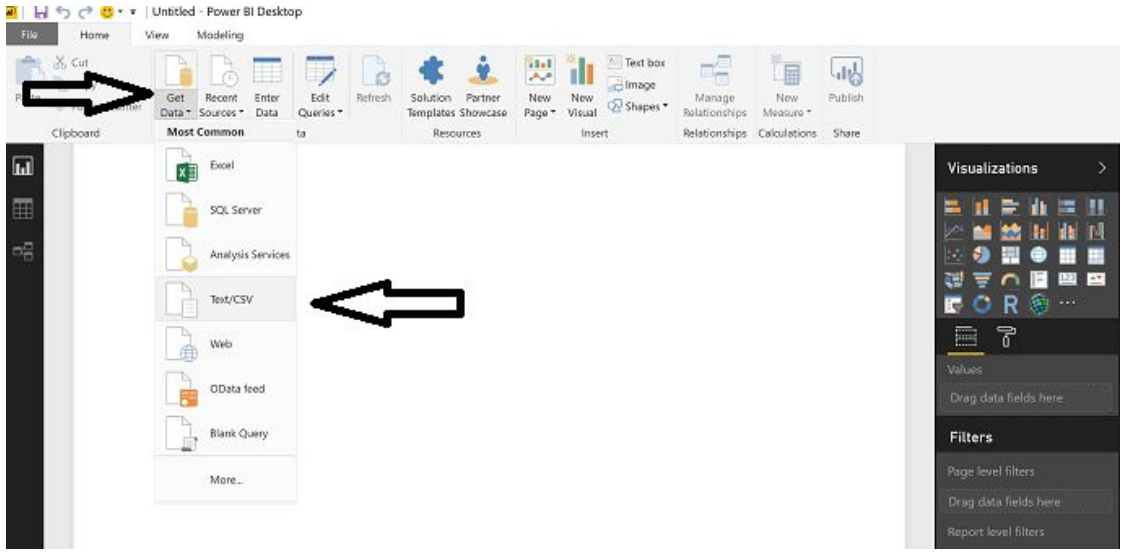
Analysis: Open Power BI Desktop

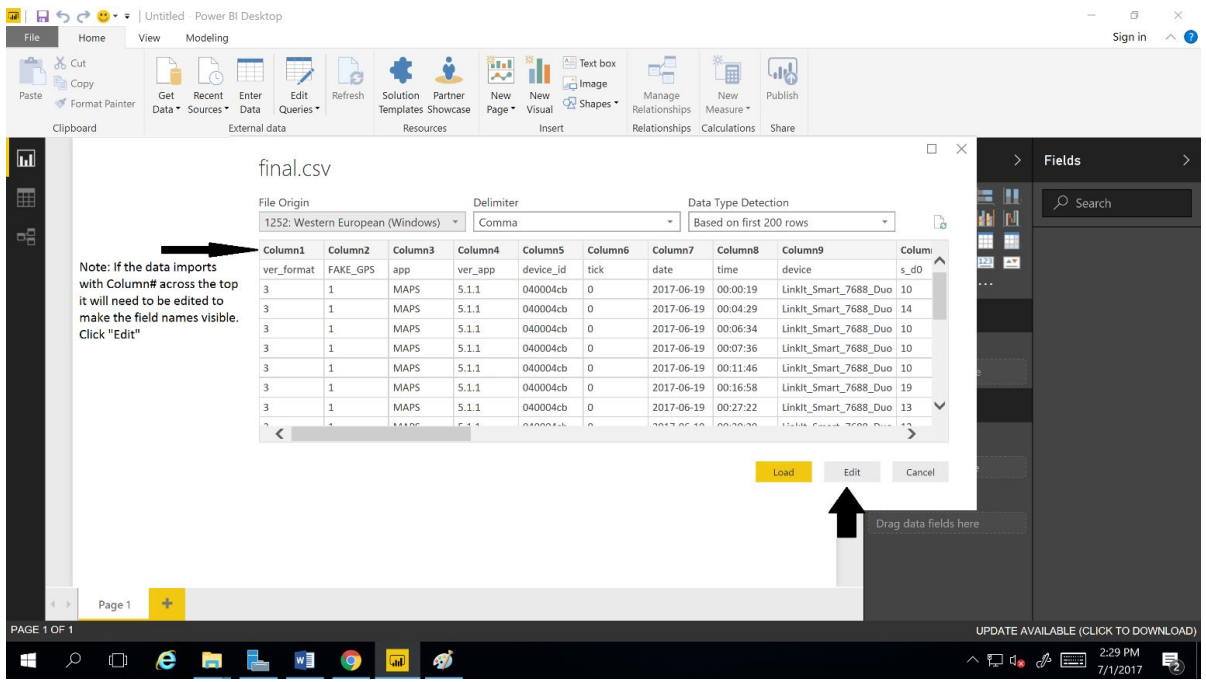
Importing Data

- Click “Get Data”
- Select txt/csv -> connect
- Select: Desktop -> ESIP -> ANALYSIS -> final.csv -> open

Check that “Column#” is NOT across the top of the table. See image below.

Otherwise Click “Load”

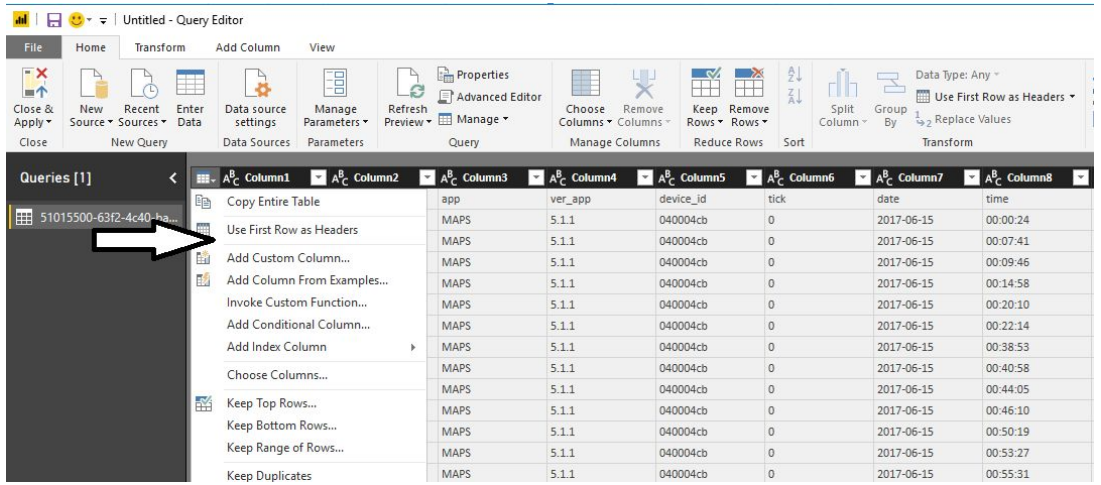


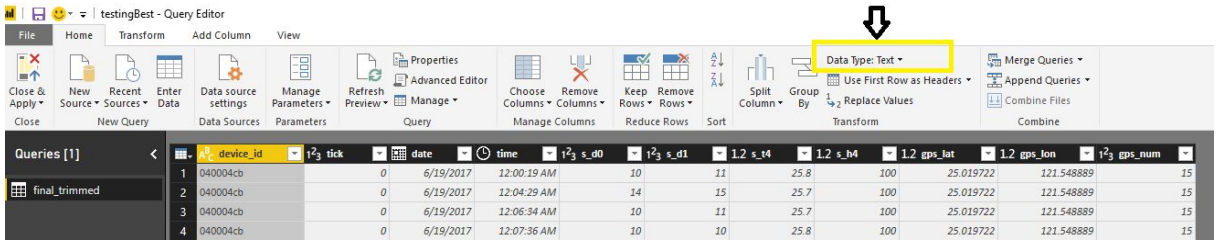


DATA TO INSIGHT CENTER
PERVASIVE TECHNOLOGY INSTITUTE

Analysis: If Editing is needed:

1. Instead of clicking “Load” click “Edit”. If you already clicked “Load” you can still click “ Edit Queries” in the ribbon.
2. In the editing window click on the pull down menu next to the small table icon in the upper left corner. Select “Use First Row as Headers”.
3. When the table imports the Column# as headers all fields are recognized as text data types. To manipulate numerical fields the data types for all necessary columns need to be changed to decimal or whole number.

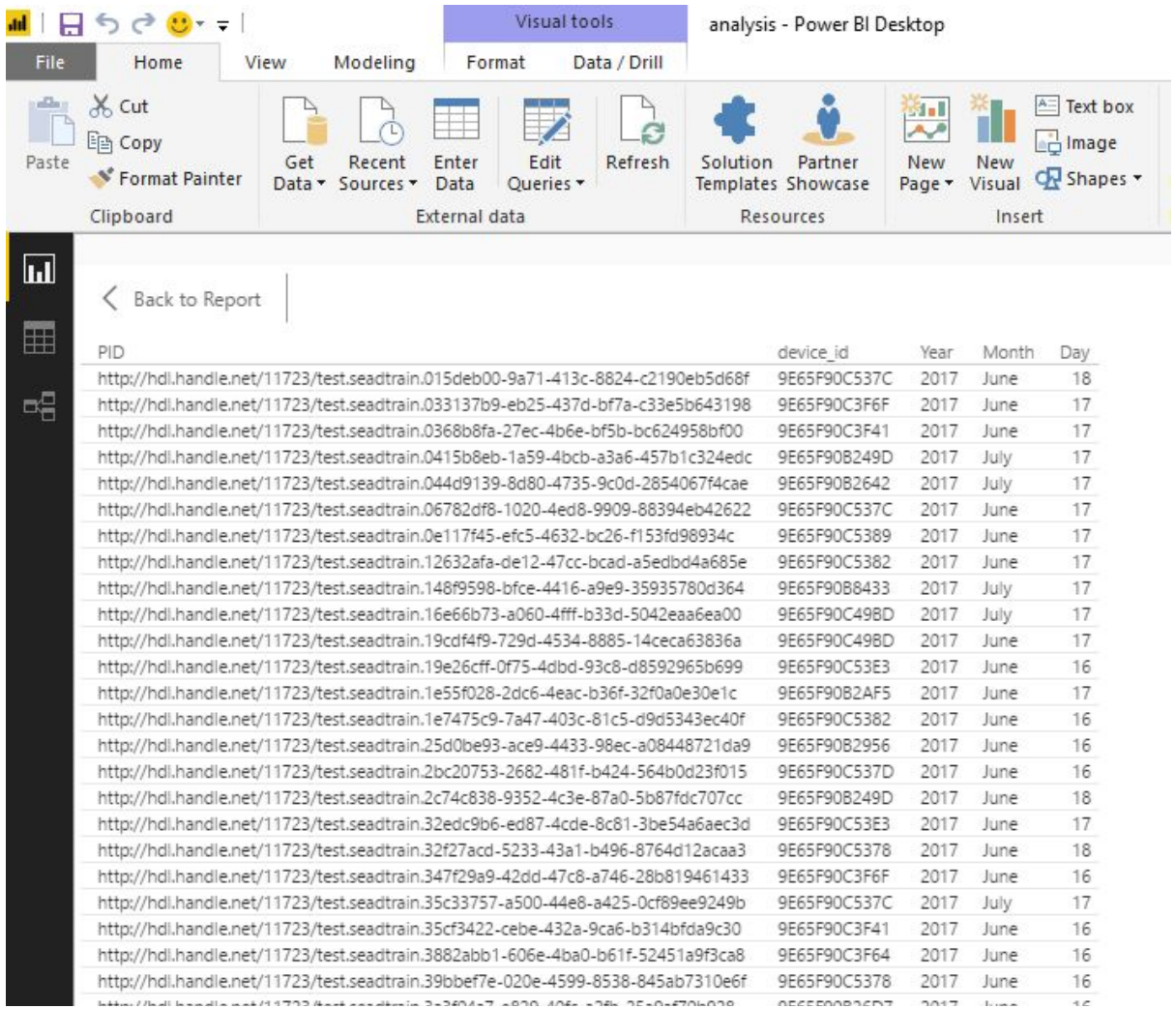




Change the s_d0, s_h4, and s_t4 to decimal. Change date to date, and time to time.

Analysis: Power BI Service exports your data to your blog or website

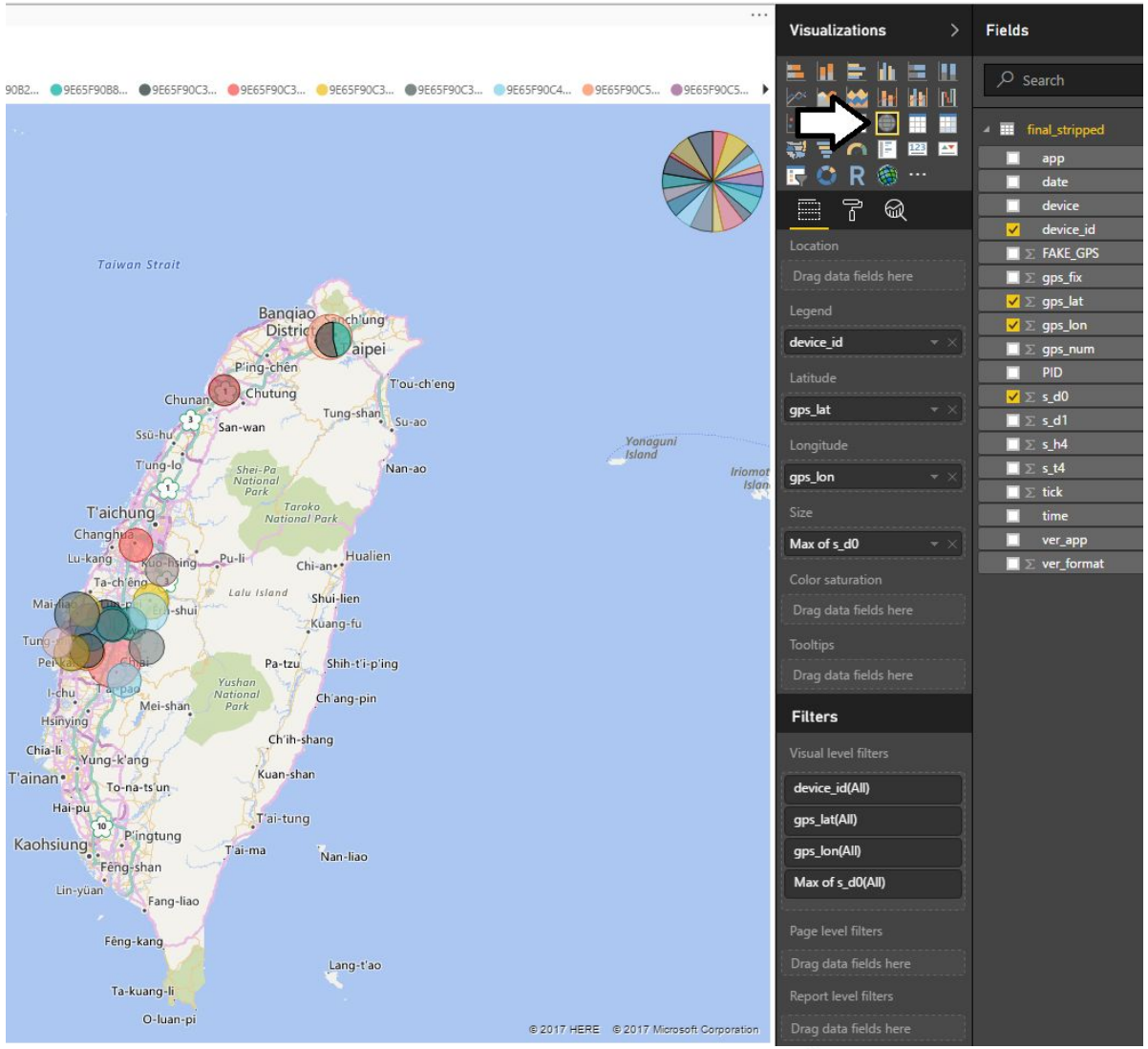
You can make your data accessible by importing a PID table that continuously updates as new data is added.



PID	device_id	Year	Month	Day
http://hdl.handle.net/11723/test.seadtrain.015deb00-9a71-413c-8824-c2190eb5d68f	9E65F90C537C	2017	June	18
http://hdl.handle.net/11723/test.seadtrain.033137b9-eb25-437d-bf7a-c33e5b643198	9E65F90C3F6F	2017	June	17
http://hdl.handle.net/11723/test.seadtrain.0368b8fa-27ec-4b6e-bf5b-bc624958bf00	9E65F90C3F41	2017	June	17
http://hdl.handle.net/11723/test.seadtrain.0415b8eb-1a59-4bcb-a3a6-457b1c324edc	9E65F90B249D	2017	July	17
http://hdl.handle.net/11723/test.seadtrain.044d9139-8d80-4735-9c0d-2854067f4cae	9E65F90B2642	2017	July	17
http://hdl.handle.net/11723/test.seadtrain.06782df8-1020-4ed8-9909-88394eb42622	9E65F90C537C	2017	June	17
http://hdl.handle.net/11723/test.seadtrain.0e117f45-efc5-4632-bc26-f153fd98934c	9E65F90C5389	2017	June	17
http://hdl.handle.net/11723/test.seadtrain.12632afa-de12-47cc-bcad-a5edbd4a685e	9E65F90C5382	2017	June	17
http://hdl.handle.net/11723/test.seadtrain.148f9598-bfce-4416-a9e9-35935780d364	9E65F90B8433	2017	July	17
http://hdl.handle.net/11723/test.seadtrain.16e66b73-a060-4fff-b33d-5042eaa6ea00	9E65F90C49BD	2017	July	17
http://hdl.handle.net/11723/test.seadtrain.19cdf4f9-729d-4534-8885-14ceca63836a	9E65F90C49BD	2017	June	17
http://hdl.handle.net/11723/test.seadtrain.19e26cff-0f75-4dbd-93c8-d8592965b699	9E65F90C53E3	2017	June	16
http://hdl.handle.net/11723/test.seadtrain.1e55f028-2dc6-4eac-b36f-32f0a0e30e1c	9E65F90B2AF5	2017	June	17
http://hdl.handle.net/11723/test.seadtrain.1e7475c9-7a47-403c-81c5-d9d5343ec40f	9E65F90C5382	2017	June	16
http://hdl.handle.net/11723/test.seadtrain.25d0be93-ace9-4433-98ec-a08448721da9	9E65F90B2956	2017	June	16
http://hdl.handle.net/11723/test.seadtrain.2bc20753-2682-481f-b424-564b0d23f015	9E65F90C537D	2017	June	16
http://hdl.handle.net/11723/test.seadtrain.2c74c838-9352-4c3e-87a0-5b87fdc707cc	9E65F90B249D	2017	June	18
http://hdl.handle.net/11723/test.seadtrain.32edc9b6-ed87-4cde-8c81-3be54a6aec3d	9E65F90C53E3	2017	June	17
http://hdl.handle.net/11723/test.seadtrain.32f27acd-5233-43a1-b496-8764d12acaa3	9E65F90C5378	2017	June	18
http://hdl.handle.net/11723/test.seadtrain.347f29a9-42dd-47c8-a746-28b819461433	9E65F90C3F6F	2017	June	16
http://hdl.handle.net/11723/test.seadtrain.35c33757-a500-44e8-a425-0cf89ee9249b	9E65F90C537C	2017	July	17
http://hdl.handle.net/11723/test.seadtrain.35cf3422-cebe-432a-9ca6-b314bfda9c30	9E65F90C3F41	2017	June	16
http://hdl.handle.net/11723/test.seadtrain.3882abb1-606e-4ba0-b61f-52451a9f3ca8	9E65F90C3F64	2017	June	16
http://hdl.handle.net/11723/test.seadtrain.39bbef7e-020e-4599-8538-845ab7310e6f	9E65F90C5378	2017	June	16

Analysis: Where are AirBox sensors located?

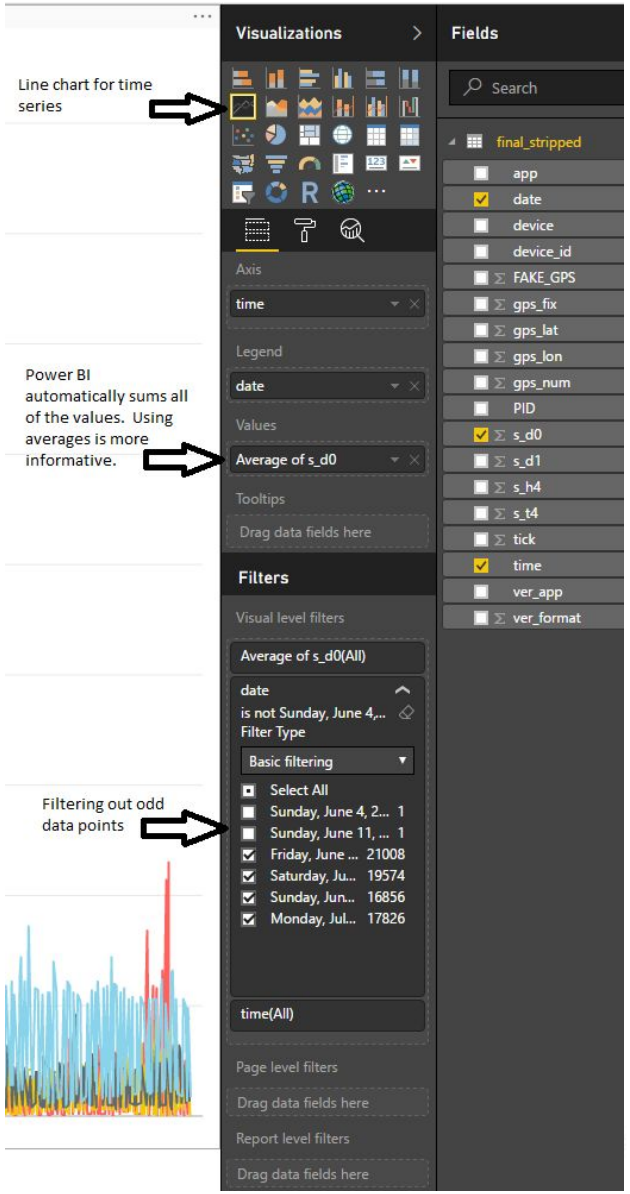
1. Click on the map icon in the visualization area.(sorry, not ArcGIS)
2. Drag device_id under Legend
3. gps_lat under Latitude
4. gps_lon under Longitude
5. We can play with the bubbles, basing them on various stat functions of PM2.5 (s_d0)
6. Thoughts?



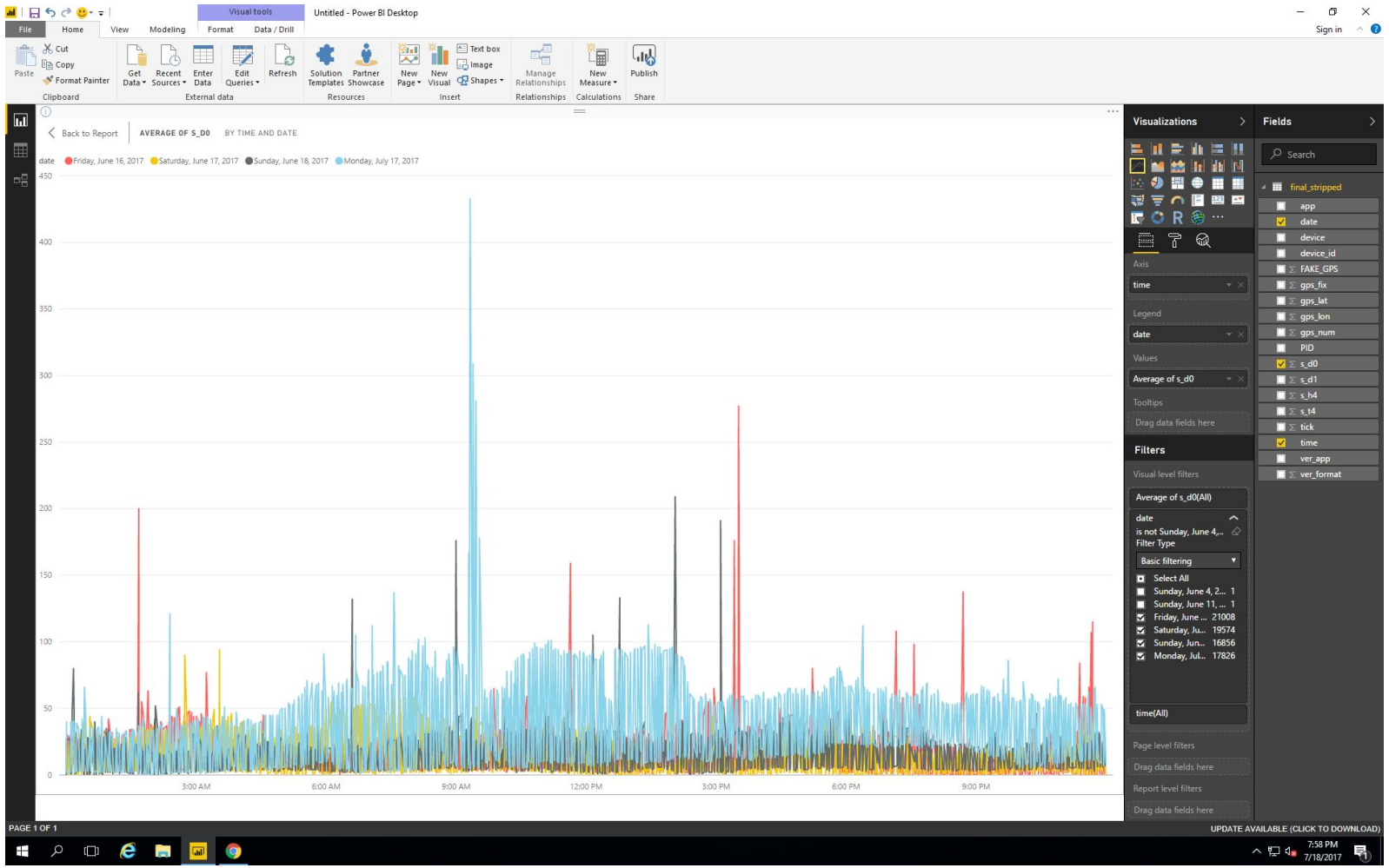
DATA TO INSIGHT CENTER
PERVASIVE TECHNOLOGY INSTITUTE

Analysis: Preliminary Visualization: PM2.5 over Time

- Particulate Matter at 2.5 microns is the unit of interest.
- In the Visualization panel click on the Line Graph
- Select the fields of interest: time, s_d0, date
- They may pop up in the appropriate categories or may need to be dragged so that time is the Axis variable, date is the Legend, and s_d0 is in the Values.



Analysis: PM 2.5 Over Time



Clicking on individual days within the legend allows you to view trends separately.

Patterns?

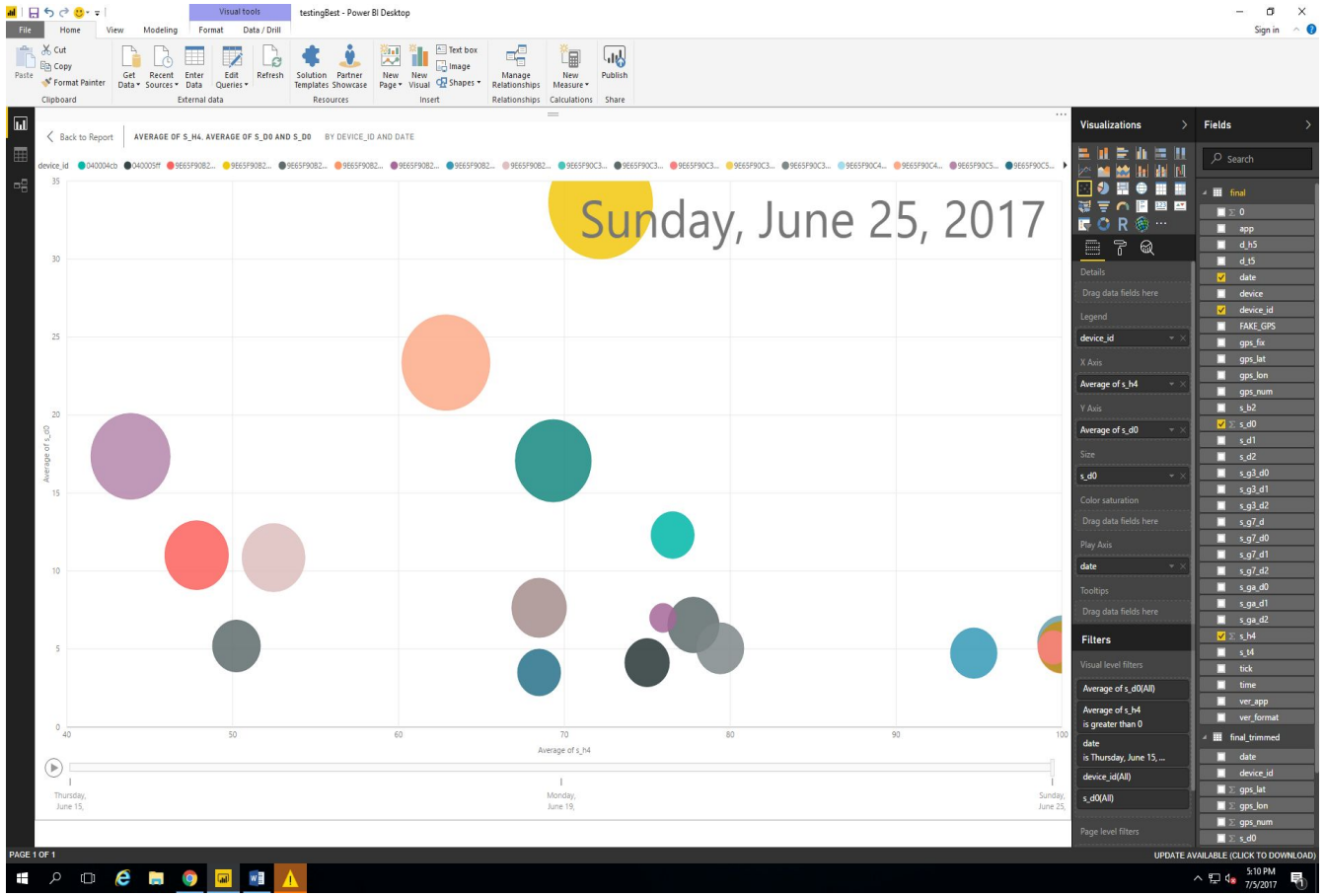
Analysis: PM2.5 vs Humidity

Click outside of the graphs to create a new chart in the dashboard.

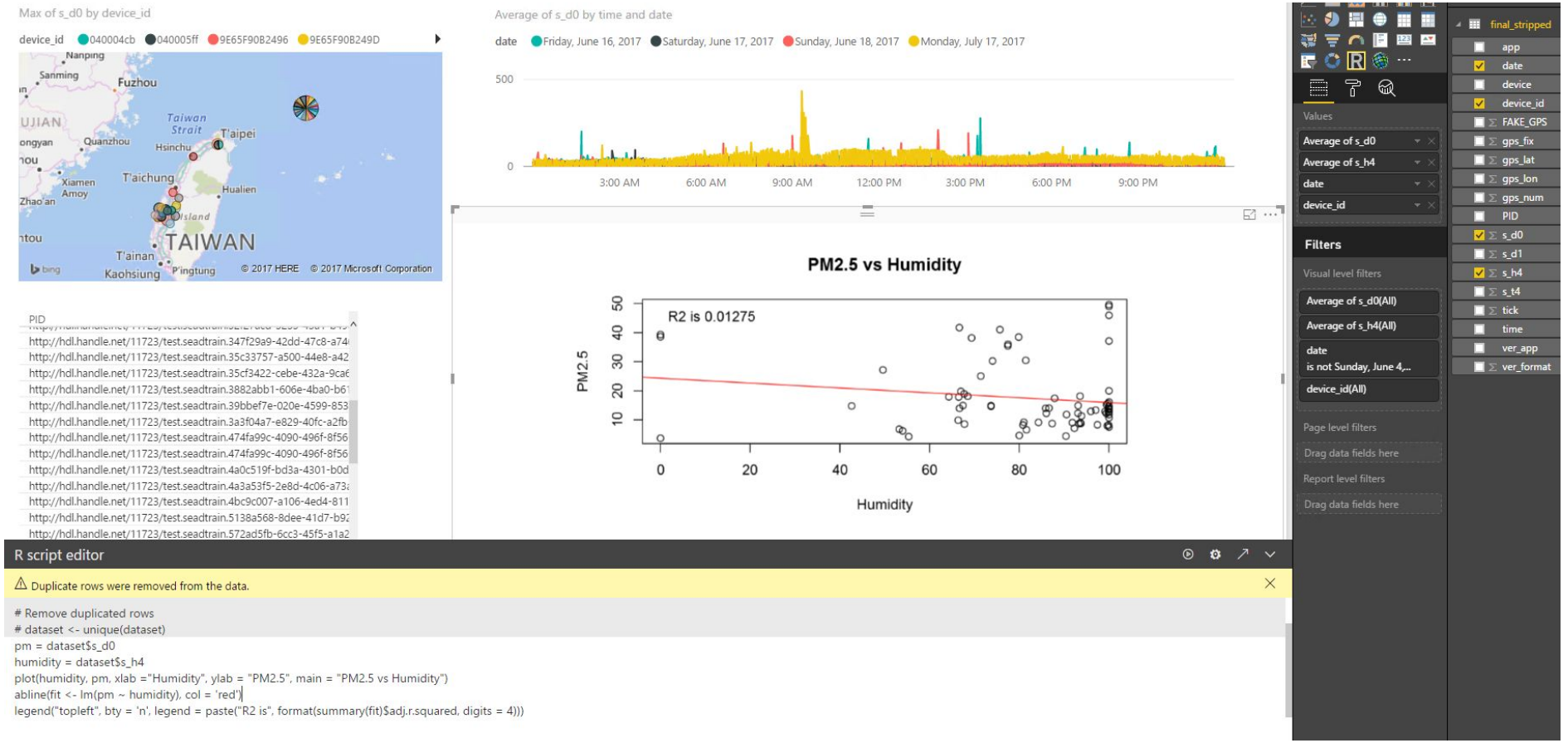
Click on the “Scatter Chart” icon under Visualizations.

Under “Fields” click in this order: device_id for the Legend, average s_h4 for the X axis, average s_d0 for the PM2.5. Drag the date field to the Play Axis. Average s_d0 for the Size.

Note: interactive capability



Analysis: Statistical Inference with R



R is pre-loaded on the VM, select R in visualization pane.

Select desired data fields, they will be loaded as 'dataset' into R

Use the R script editor to create desired plots and analysis

DATA TO INSIGHT CENTER
PERVASIVE TECHNOLOGY INSTITUTE

Thank You!

Please go to: **https://tinyurl.com/SEADTrain**

and fill out the course evaluation