



2019 Advanced Institute on Health Investigation
and Air Sensing for Asian Pollution (AI on Hi-ASAP)
September 2 – 6, 2019
Academia Sinica, Taipei, Taiwan

Demonstration of Data Analysis

Shih-Chun Candice LUNG 龍世俊

- Research Center for Environmental Changes, Academia Sinica, Taiwan
- Center for Sustainability Science, Academia Sinica, Taiwan

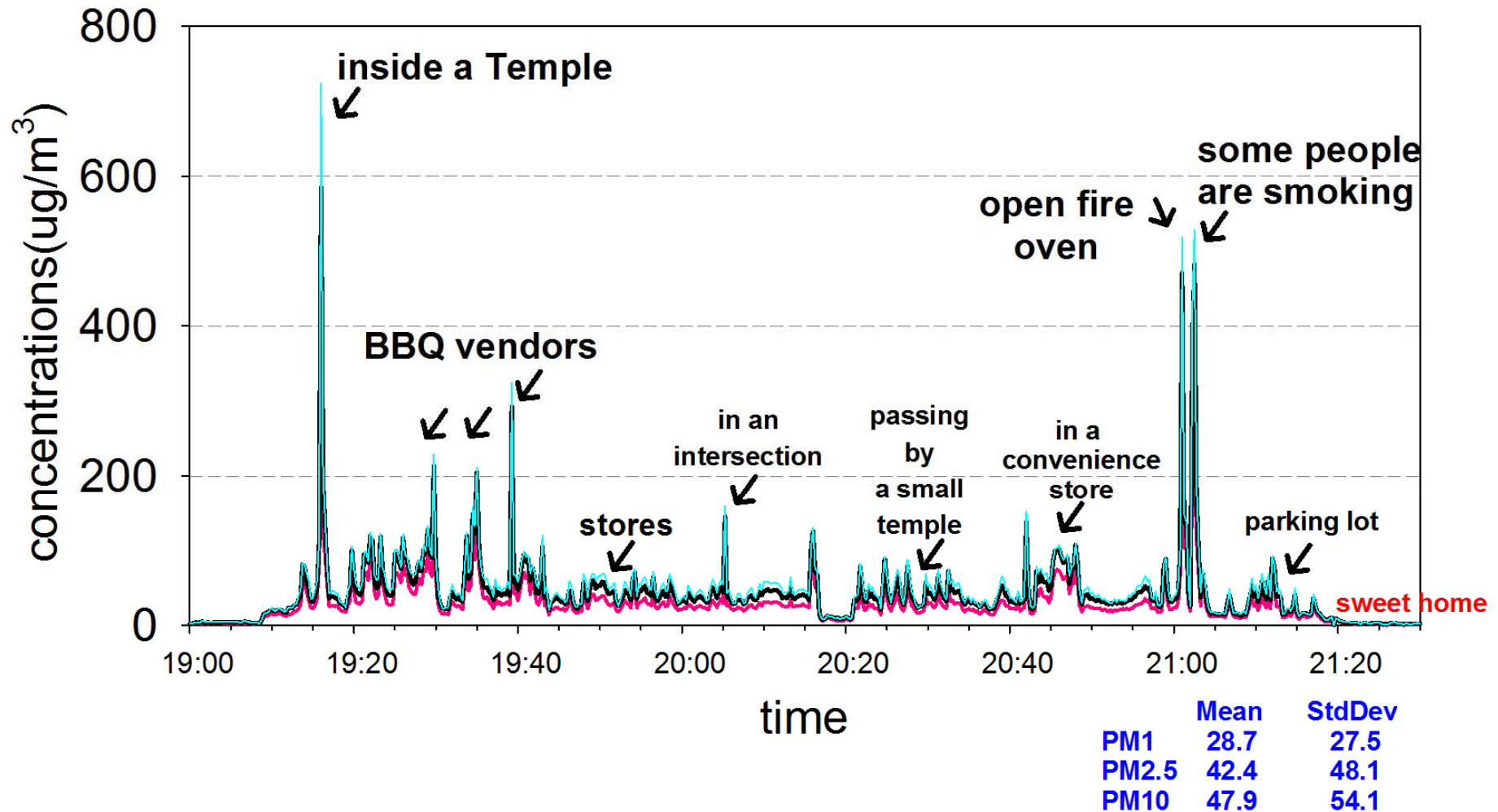
Example 1: Plotting the data in time series



[Candice Lung, 2017]

July 4, 2017, Taipei night market
AS-LUNG sampling rate: every 15 seconds

— PM1 — PM2.5 — PM10



Example 2: Description statistics in different grouping PM₁₀ levels from PEMs(ug/m³)

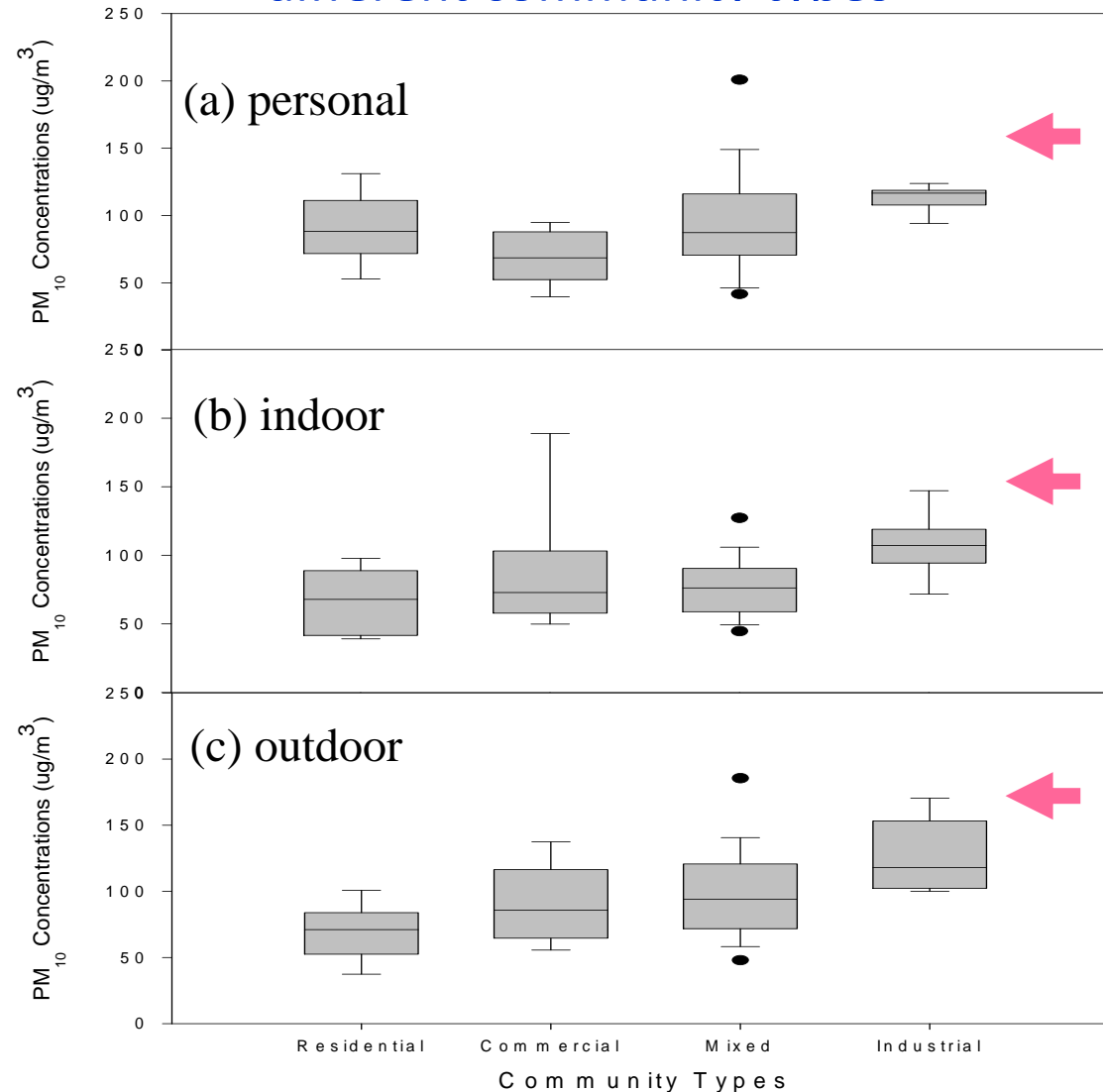
		NM	NR	CC	CM	SM	SI	overall
personal	N	8	8	8	9	8	6	47
	mean	48.5	91.1	68.7	129.7	85.8	98.2	87.4
	SD	26.2	28.9	21.6	51.8	33.0	37.0	42.2
indoor	N	8	8	6	9	8	6	45
	mean	53.6	67.0	88.9	86.3	83.6	107.9	79.8
	SD	18.7	25.1	63.5	24.5	24.0	26.9	34.4
outdoor	N	7	8	6	7	7	6	41
	mean	57.8	69.4	90.8	103.0	132.3	127.3	95.5
	SD	24.7	23.4	34.8	21.8	51.6	29.4	41.3

The ambient PM₁₀ from Taiwan EPA stations were from 37-98ug/m³.

Current 24-hr PM10 standard is 125ug/m³.

Example 3: Personal/indoor/outdoor comparisons (ANOVA)

Figure 1: (a) personal, (b) indoor, and (c) outdoor PM_{10} levels ($\mu g/m^3$) among different community types



middle lines represent median values; dots represent the 5% and 95% percentiles

Example 4: Comparisons with two-sample t test or non-parametric tests

Table 6 Personal PM₁₀ levels (µg/m³) classified by different (a) locations (b) traffic-related activities and (c) pollution sources during the sampling day

(a) Locations		N	mean	SD
% time spent outdoors^{1*}	<16%	24	72.7	48.9
	>16%	23	101.5	27.9
kitchen^{2*}	no	22	73.1	36.4
	yes	25	100.1	43.5
(b) Traffic-related activity		N	mean	SD
time spent on transportation^{1†}	<4 hour	24	75.9	28.0
	>4 hour	23	99.5	51.0
rode a motorcycle*	no	30	77.4	37.7
	yes	17	105.1	44.9

***: $p < 0.05$, †: $0.05 < p < 0.1$, Wilcoxon rank sum test was used**

1: these variables were classified by the median values

2: two subjects reported cooking but did not select kitchen in the “location” since they did not have a separated kitchen

Example 5: Comparisons between exposed/not-exposed subjects

Table 6 (continued)

(c) Pollution source		N	mean	SD
cooking or in the kitchen ^{2†}	no	20	74.3	34.6
	yes	27	97.2	45.2
burning incense [†]	no	37	83.0	41.0
	yes	10	103.8	44.6
Passing by factories*	no	34	76.8	34.0
	yes	13	115.2	49.8
ETS exposure duration ¹	<0.5 hr	23	79.1	30.6
(Environmental Tobacco Smoke)	>0.5 hr	24	95.4	50.2

Example 6: Evaluating factors with multiple regressions

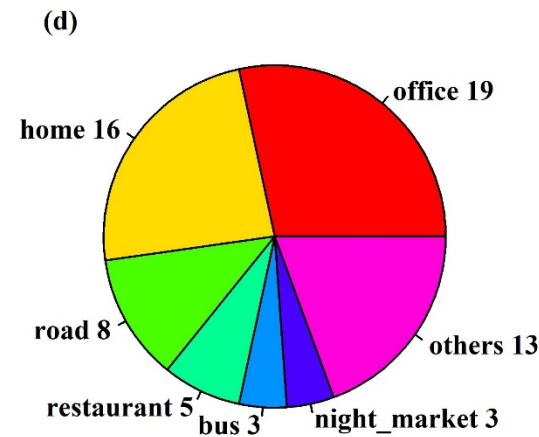
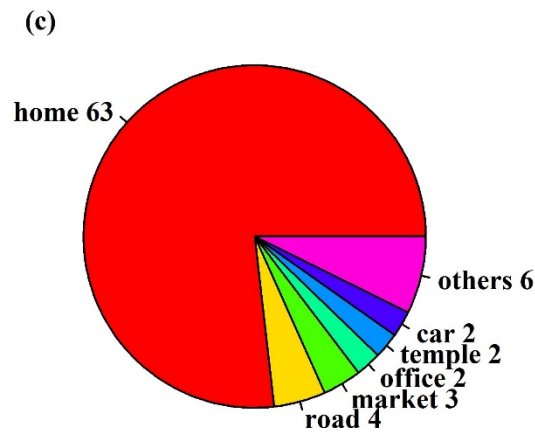
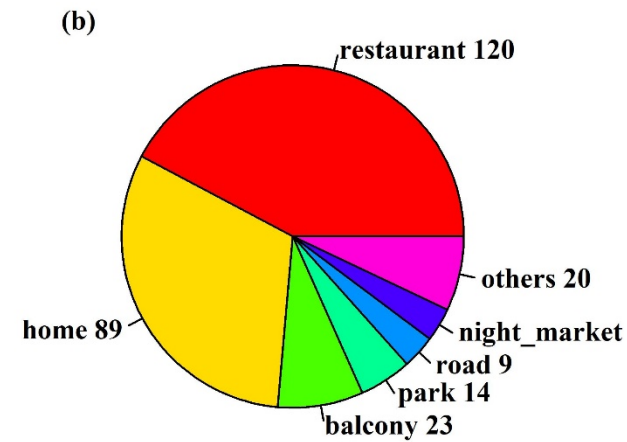
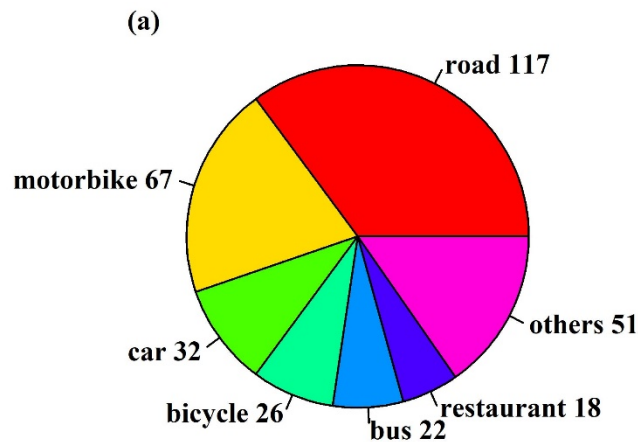
Table 8 Best-fit stepwise regression ($R^2 = 0.54$) for personal PM_{10} exposures ($\mu\text{g}/\text{m}^3$)

Variable	Parameter Estimate	Standard Error	<i>p</i> value
Intercept	50.7	8.49	<.0001
time spent on transportation (hr)	3.15	1.05	0.0048
passing by factories	24.2	10.4	0.0255
in the kitchen	23.3	9.02	0.0142
environmental tobacco exposure (hr)	1.89	0.89	0.0412

Note: the variables are ranked in the order of entering the regression model, passing by factories and being in the kitchen were dichotomous variables while the time spent on transportation and ETS exposure duration were continuous ones

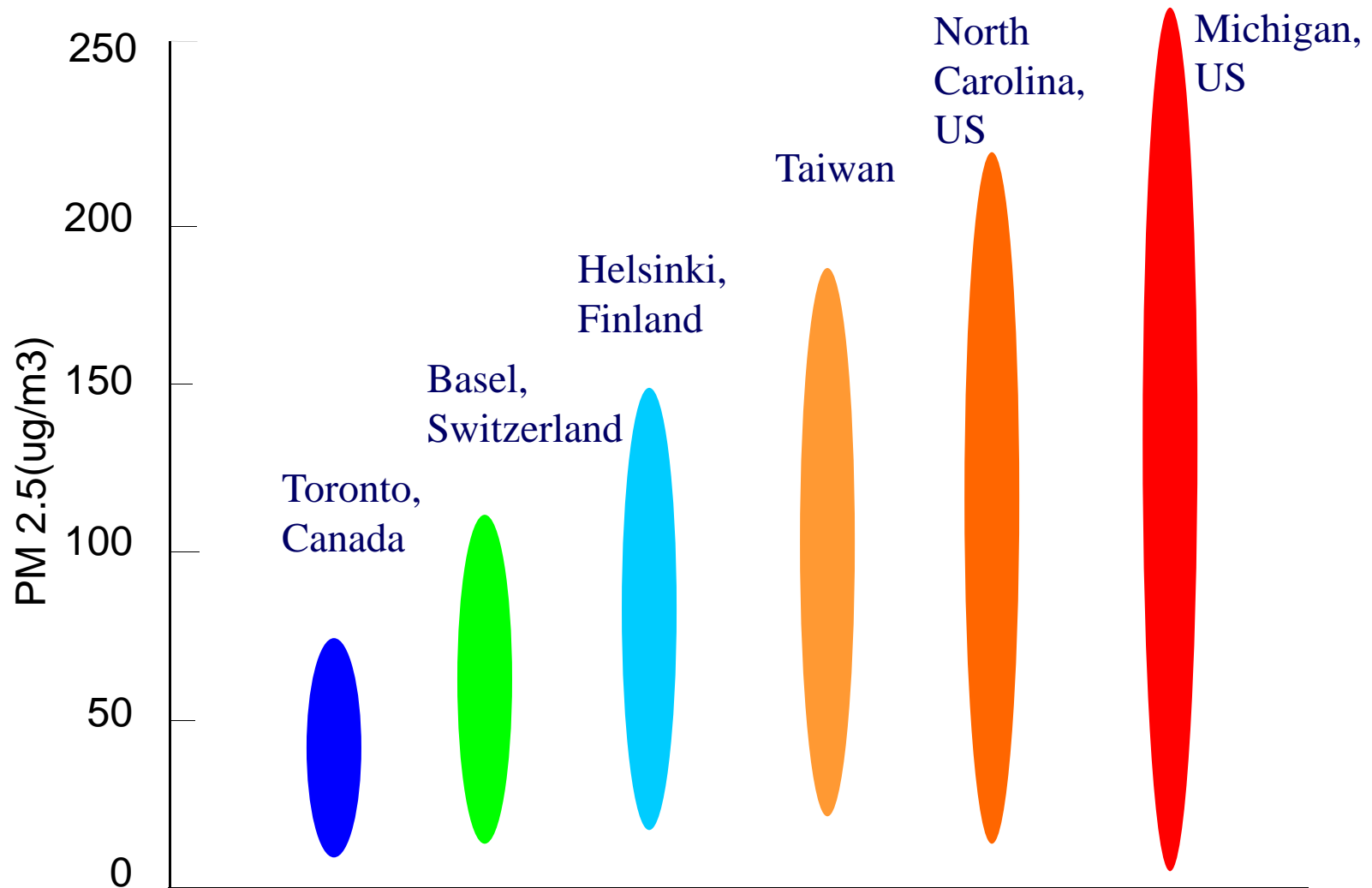
Example 7: Graphic presentations

The counted numbers of the exposure micro-environments based on the responses at 30-min intervals of time-activity diaries of 33 subjects indicating (a) vehicle emission (n=333), (b) cooking (n=284), (c) incense-burning (n=82), and (d) Environmental Tobacco Smoke (n=67)



Example 8: International comparison

PM_{2.5} exposure levels among different studies



Important Notes for Data Analysis

- Think about **how to analyze your data** when you plan your study, consider the time and spatial resolution of your data
- **Plan data analysis before** not after **data collection**
- **Collect information** that can be used in data analysis or explain your results
 - Ex. weather condition, exposure sources/behaviors, etc.
- **Data cleaning** is the first step of data analysis
- Plot your data **graphically** before conducting any statistical analysis

Demonstration of Data Analysis

Any questions?

Shih-Chun Candice LUNG 龍世俊

Research Center for Environmental Changes, Academia Sinica

Center for Sustainability Science , Academia Sinica

Sc.D, Harvard University, School of Public Health