

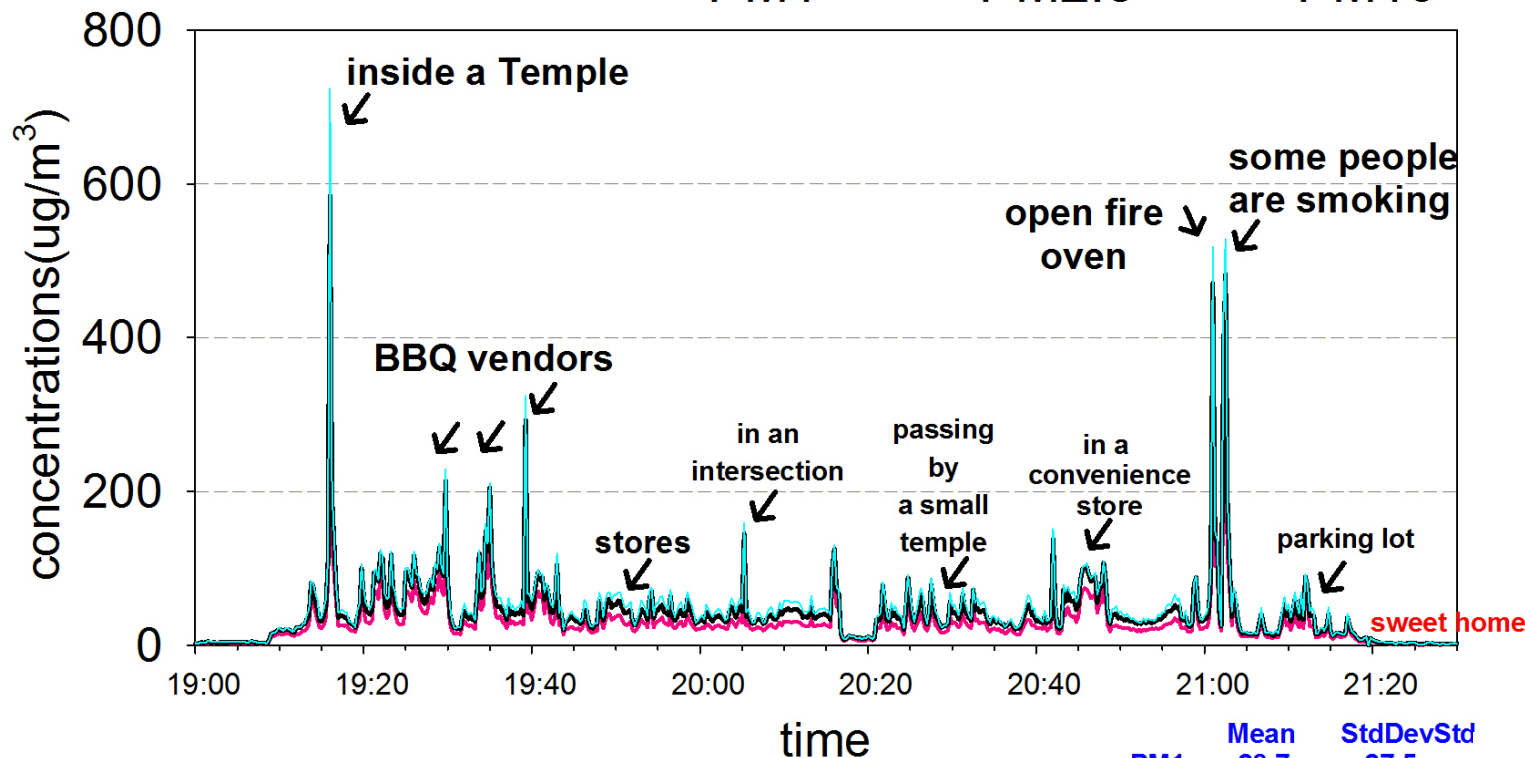
QA/QC of the survey data

- Quality Assurance/Quality Control
- Introduction of Core Questions in Questionnaire
- Some tricks to prepare dataset after survey

Joanne HU
SC Candice LUNG

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

— PM1 — PM2.5 — PM10



■ Unadjusted PM concentrations

	Mean	StdDevStd
PM1	28.7	27.5
PM2.5	42.4	48.1
PM10	47.9	54.1

Survey Data

中央研究院
ACADEMIA SINICA

2020

Impact of Air Pollution on Individual and Community

Questionnaire

All responses given will be used for academic research only.

Any personal information provided will be kept strictly confidential.

- Survey methods
 - Questionnaire by face-to-face interview
 - TAD recorded by subjects and confirmed by interviewer within 24 hours
- The importance of survey data
 - Find out the important factors that influence subject's $PM_{2.5}$ exposure concentrations
 - Truthfully record what happened
 - Find out some factors or truth unveiled in subject's exposure patterns

Time-activity Diary (TAD)

Time-activity Diary (TAD)

Date: [continental time] 20 / / (yyyy/mm/dd)

Subject No. _____

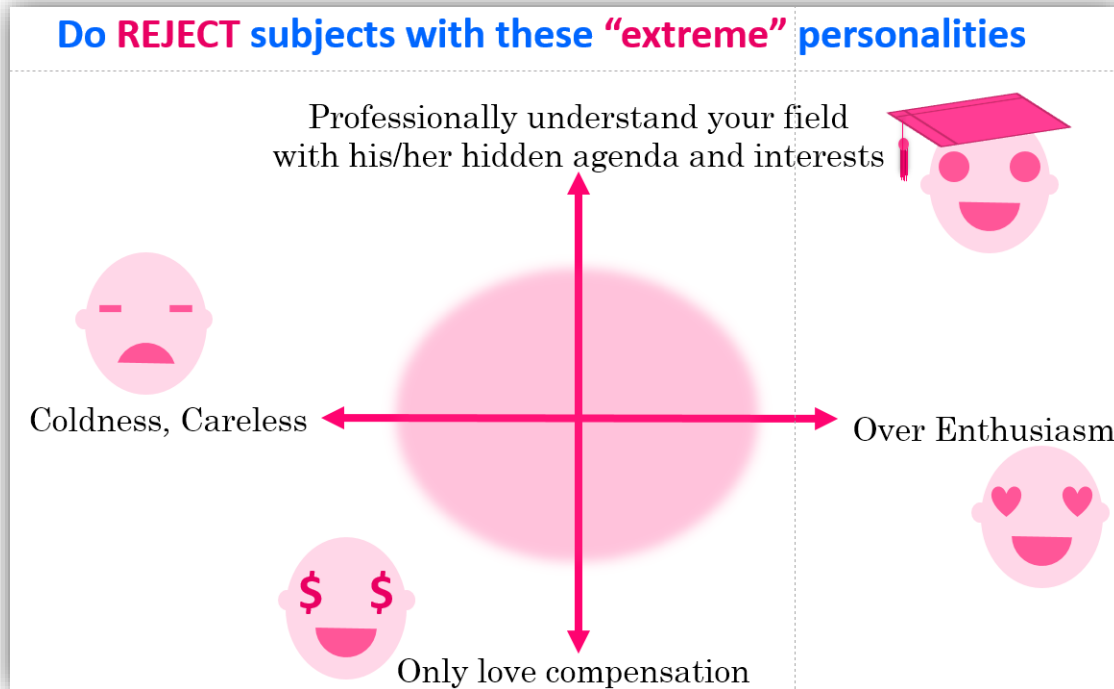
Interviewer No. _____

Weather code: Sunny, Cloudy, Rainy, Windy

Items Time (hh:mm)	A. Location		B. Ventilation status	C. Activity type		D. Air quality and weather		
	18	16		Activity 1	Activity 2	Quality 1	Quality 2	Weather
1900 - 1910	18	16	9	8	6	3	65.sticky tofu	S, W
Set your watch first and write down your environment information below ! Make sure that each blank has been filled with only one right code.								

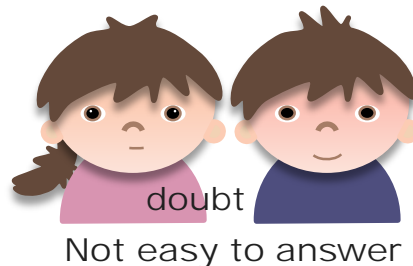



The mostly important **key step** to make sure good quality survey data is **when you recruit subjects**



When interviewing

- QA/QC tricks for trained interviewers
 - Better to ask questionnaire with **face-to-face** interview
 - Always care subject's **facial expressions** when interviewing
 - Not longer than **one hour** when interviewing



 中央研究院
ACADEMIA SINICA

2020

Air-Quality-Survey-for
Personal-and-Community-Environments

Questionnaire

All responses given will be used for academic research only.
Any personal information provided will be kept strictly confidential.

◎ Subject No.: _____

◎ Interviewer name or No.: _____



Questionnaire with core questions



Questionnaire

- Questionnaire for Hi-ASAP collaborators
 - These are **core questions**
 - Including three sections: A, B and C
 - We arrange order of these questions based on subject's attention and psychological reasons
- There are 65 questions so far in this version

The image shows a questionnaire form from Academia Sinica. At the top left is the logo and name '中央研究院 ACADEMIA SINICA'. The title '2020 Air Quality Survey for Personal and Community Environments' is centered in blue. Below it, 'Questionnaire' is written in black. A disclaimer states: 'All responses given will be used for academic research only. Any personal information provided will be kept strictly confidential.' At the bottom, there are two fields: '④ Subject No.: _____' and '④ Interviewer name or No.: _____'.

Section A: Household and community environment

A → Household and Community Environment

A1. → How many air-conditioners are in your house? _____

A2. → How many air-cleaners are in your house? _____

A3. → How many total floor area in your house? (For multi-floor residence, please add up the total)
□ around: _____ m²

A4. → Is your living room carpeted?

□(01) Yes, entirely → □(02) Yes, with a small piece → □(03) No

A5. → What is the main construction material of your residence?

□(01) Reinforced concrete → □(02) Brick and concrete

□(03) Steel and reinforced concrete → □(04) Light-weight steel wall and structure →

□(05) Wood → □(06) Other (Please specify): _____

A6. → What type of window in your living room?

□(01) Non-airtight (wood) → □(02) Non-airtight (metal)

□(03) Airtight → □(04) No window

□(05) Other (Please specify): _____

A7. → Usually ventilation status in your **living room** in the past year.

(The hot, mild or cold days depend on the feeling of subject)

the windows (door)	Fully open	Half open	Slightly open or all closed (Without air-conditioning or heating on)	Slightly open or all closed (With air-conditioning or heating on)
A7a. On hot days	□(01)	□(02)	□(03)	□(04)
A7b. On mild days	□(01)	□(02)	□(03)	□(04)
A7c. On cold days	□(01)	□(02)	□(03)	□(04)

A8. → Ventilation status in your **bedroom** in the past year.

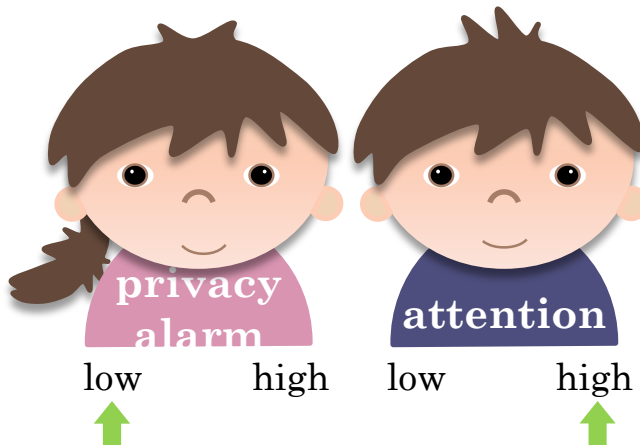
(The hot, mild or cold days depend on the feeling of subject)

the windows (door)	Fully open	Half open	Slightly open or all closed (Without air-conditioning or heating on)	Slightly open or all closed (With air-conditioning or heating on)
A8a. On hot days	□(01)	□(02)	□(03)	□(04)
A8b. On mild days	□(01)	□(02)	□(03)	□(04)
A8c. On cold days	□(01)	□(02)	□(03)	□(04)

- Section A collects household and community characteristics which may influence air pollution concentrations

- How many air-conditioners and air-cleaners
- Ventilation status in living room
- Community PM_{2.5} emission sources

- Subject's facial expressions



Section B: Personal daily routine

B → Personal daily routine

B1. → Last month, how long on average do you use the following means of transport per day?

	(1) Working/School day	(02) Holidays
a. → Walking	_____ min	_____ min
b. → Manpower bicycle or tricycle	_____ min	_____ min
c. → Scooter or motorcycle	_____ min	_____ min
d. → Car or truck	_____ min	_____ min
e. → Engine-powered bus	_____ min	_____ min
f. → Train or light rail (tram)	_____ min	_____ min
g. → subway	_____ min	_____ min

B2. → Excluding time spent on transport, how long on average do you spend outdoors each working/school day? ·
_____ hr. _____ min

B3. → Excluding time spent on transport, how long on average do you spend outdoors each non-working day/holiday? · _____ hr. _____ min

B4. → In the past month, how often do you or your family fry, stir-fry or deep-fry with the kitchen hood on?
☐ (01) Always → ☐ (02) Often → ☐ (03) Sometimes → ☐ (04) Seldom → ☐ (05) Never

B5. → Do you belong to a religion and, if yes, which religion do you belong to?
☐ (01) No religion ☐ (05) Other Christian ☐ (09) Hindu
☐ (02) Catholic ☐ (06) Jewish ☐ (10) Other Asian religions
☐ (03) Protestant ☐ (07) Islamic ☐ (11) Other religions
☐ (04) Orthodox ☐ (08) Buddhist

B6. → How often do you worship at your religion place (such as temple, church, mosque...)?
☐ (01) 3 times or more every day ☐ (05) 1-2 times a week
☐ (02) 2 times every day ☐ (06) 1-3 times a month
☐ (03) Once a day ☐ (07) Less than once a month
☐ (04) 3-6 times a week ☐ (08) Never (Go to B9)

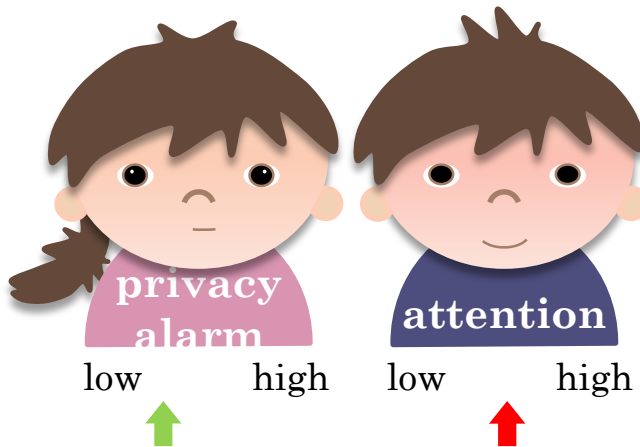
B7. → How long on average each time (at this religion place)? · _____ hr. _____ min

B8. → If joss sticks are used in your religion, how many joss sticks on average do you use each time?
_____ joss sticks · ☐ (00) Not used

- Section B collects some activities and durations which are important factors for air pollution exposure

- tricks: interviewers should **read out this question word by word** and **fully understand the definition** in order to respond to subjects' suspicious and confusion.

- Subject's facial expressions



Section C: Personal health and social status

C→ Personal health and social status

C1.→ Sex : ☐ (01) Male · ☐ (02) Female ·

C2.→ In which year were you born? (in AD year) _____

C3.→ What is your height? _____ cm

What is your weight? _____ kg

C4.→ How many people (including yourself) live in your residence? _____

C5.→ What is your ethnicity? → {Each country lists its own top four ethnicity groups for choices}

☐ (01)

☐ (02)

☐ (03)

☐ (04)

☐ (05) Other (Please specify) _____

C6.→ How many years of education have you completed? _____ years

C7.→ What is your present marital status?

☐ (01) Never married

☐ (02) Cohabiting

☐ (03) Married

☐ (04) Divorced

☐ (05) Separated

☐ (06) Widowed

☐ (07) Other (Please specify) _____

C8.→ Overall speaking, how do you consider your physical health?

☐ (01) Very good → ☐ (02) Good → ☐ (03) Fair → ☐ (04) Poor → ☐ (05) Very poor

C9.→ Which of the following have you been diagnosed with? {Multiple choices allowed} ·

☐ (01) High blood pressure and other cardiovascular diseases

☐ (02) Asthma

☐ (03) Rhinitis or nasal allergies

☐ (04) Bronchitis, pneumonia and other lung diseases

☐ (05) Skin disease

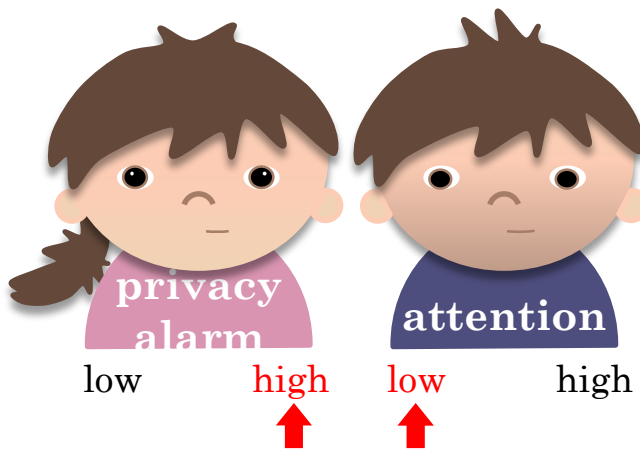
☐ (06) Renal disease

☐ (07) Diabetes

☐ (08) Sleep disorder

☐ (09) None of the above

- Section C collects certain **private information** and some questions that subject may reject or give you a random answer
- Avoid** subject's facial expressions look like these
 - tired, over-cautious, inpatient, nervous, cranky
 - Solutions:** “How about we continue this section tomorrow?”



After interviewing questionnaire

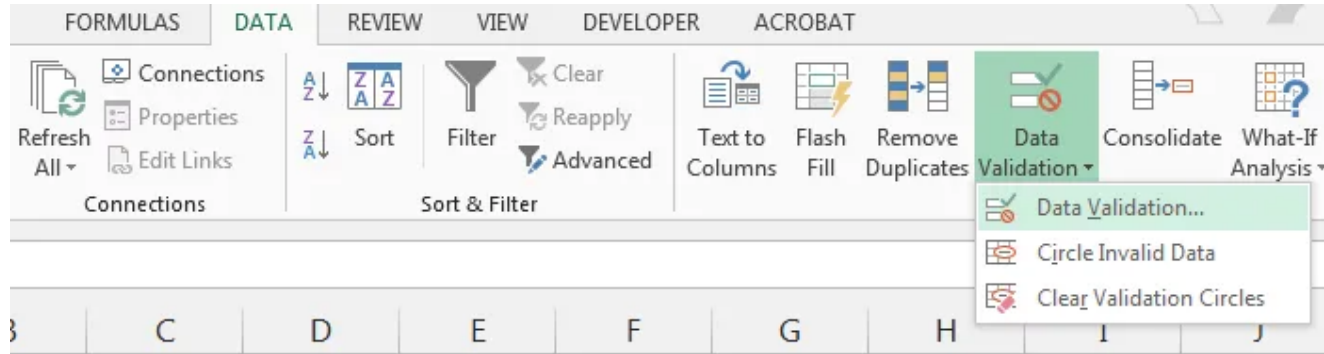
- **Within two days**: Check and ask subject if needed
 - Missing or unreasonable answers
- **Around 20 days later**: Choose some questions to ask subject again for check his consistency and reliability
 - Choose the questions with the answers that will stay the same for the past 20 days
 - eg. Birth year, Height, air-conditioner number, etc.



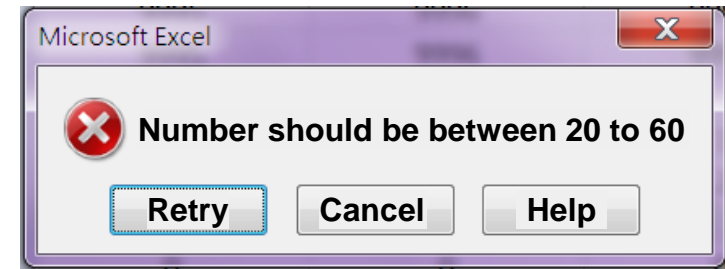
Some tricks to process or classify your data

Some tricks: (1) to Key-in your data

- Use “Data validation” tool provided by Microsoft Excel software



- eg. Set up range of subjects' age between 20 to 60. There will be a warning when key-in the wrong value.
- Double key-in efforts may reduce typo errors



Some tricks: (2) to screen your data by some simple formula

- Does any subject's age over your **recruit criteria**?
- Does any subject's have the **unreasonable** height, weights, responses?
- Does any subject's questionnaire contain too many “ I don't know” answers?
 - Subjects have rights to answer “ I don't know” or “I don't want to answer”. But please consider to **rule out** the questionnaire of the subject **with too many "No" answers**.



Some tricks: (3) Process and classify data

- Process original data → Pretest → Refine and revise dataset format
- Make sure that dataset is clean and neat before you analyze
 - Missing data
 - Extreme high/low or unreasonable values
 - Classify data
 - Dummy variables (0 or 1) transforming
 - Recalculating data
 - eg. Body Mass Index (BMI) = $\text{weight}(\text{kg}) / \text{height}(\text{meter})^2$
 - Consistent data format
 - eg. “0” means “none of this kind of air pollution source”
“1” means “yes, do have this kind of air pollution source”

Questionnaire (shortly called QA)

Subject's ID

Education levels

Marriage status

Considering to classify into 3 groups (1-3, 4-7, 8-10)

QA_Example.csv

	A	B	C	D	E	F	G	H	DY	DZ	EA	EB	EC	ED	EE
1	S_no	Loc	Season	A_Gender	A_Birth	A_FBirthPlace	A_Educ	A_Marriage	C_Height	C_Weight	C_SelfPerceived_Health	C_SocietyClass	C_Employment	C_Personal_Income	C_Family_Income
2	LA001	1	1	2	48	1	4	3	158	75	4	1	3	4	1
3	LA002	1	1	2	66	1	4	3	155	48	4	5	1	4	1
4	LA003	1	1	1	66	1	6	3	163	78	3	7	1	10	1
5	LA004	1	1	2	41	3	5	1	168	80	3	6	10	4	
6	LA005	1	1	1	37	3	6	4	166	72	2	4	10	98	9
7	LA006	1	1	1	53	1	3	3	173	75	4	5	1	9	
8	LA007	1	1	2	63	1	5	3	158	52	4	6	1	3	
9	LA008	1	1	2	46	1	4	3	158	63	2	8	14	1	
10	LA009	1	1	1	42	1	7	3	180	90	3	5	1	13	1
11	LA010	1	1	2	58	1	6	3	168	55	2	6	1	10	2
12	LA011	1	1	1	52	1	4	3	170	75	2	5	1	7	1
13	LA012	1	1	2	39	3	4	3	148	52	4	6	10	3	
14	LA013	1	1	2	52	1	4	1	145	43.6	26.12861083	4	6		

GK

C_BMI

30.0432623

19.97918835

29.35752192

28.3446712

26.12861083

25.05930703

20.82999519

Recalculated into BMI

19.48696145

25.95155709

23.73995617

20.7372176

Missing data codes:
 95: unexpected situation
 96: interviewers' error
 97: subject doesn't know
 98: reject to answer
 99: skip

TAD (shortly called QB)

QB_Original_Example.csv

Original data after key-in and QA/QC

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	
1	S_no	Year	Month	Day	Date	WD	Hour	Minute_3	Place1	Place2	Ventilatio	Ventilatio	Activity1	Activity2	AP1	AP2	AP3	Sensor	
2	LA001	2019	3	4	1	1	10	1	1	0	3	0	10	0	0	0	0	0	0
3	LA001	2019	3	4	1	1	10	2	1	0	3	0	10	0	0	0	0	0	1
4	LA001	2019	3	4	1	1	11	1	1	0	3	0	4	0	0	0	0	0	1
5	LA001	2019	3	4	1	1	11	2	1	0	3	0	8	0	0	0	0	0	1
6	LA001	2019	3	4	1	1	12	1	1	0	3	0	10	0	0	0	0	0	1
7	LA001	2019	3	4	1	1	12	2	1	0	3	0	10	0	0	0	0	0	1
8	LA001	2019	3	4	1	1	13	1	1	0	3	0	10	0	0	0	0	0	1
9	LA001	2019	3	4	1	1	13	2	1	0	3	0	1	0	0	0	0	0	1
10	LA001	2019	3	4	1	1	14	1	3	0	3	0	3	0	0	0	0	0	1
11	LA001	2019	3	4	1	1	14	2	3	0	3	0	3	0	0	0	0	0	1

QB_Processed_Example.csv

Classified and also transformed into dummy variables

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U
1	S_no	Year	Month	Day	Date	WD	Hour	Minute_3	Loc_Home	Loc_In	Loc_Out	Loc_Trans	Loc_Trans	Loc_Trans	Loc_In_All	Loc_Out_#	Vent_Clos	Vent_Ope	Vent_AC	Act_Sleep	Act_Co
2	LA001	2019	3	4	1	1	10	1	1	1	0	0	0	0	1	0	0	1	0	0	0
3	LA001	2019	3	4	1	1	10	2	1	1	0	0	0	0	1	0	0	1	0	0	0
4	LA001	2019	3	4	1	1	11	1	1	1	0	0	0	0	1	0	0	1	0	0	0
5	LA001	2019	3	4	1	1	11	2	1	1	0	0	0	0	1	0	0	1	0	0	0
6	LA001	2019	3	4	1	1	12	1	1	1	0	0	0	0	1	0	0	1	0	0	0
7	LA001	2019	3	4	1	1	12	2	1	1	0	0	0	0	1	0	0	1	0	0	0
8	LA001										0	0	0	0	1	0	0	1	0	0	0
9	LA001										0	0	0	0	1	0	0	1	0	1	0
10	LA001										0	0	0	0	1	0	0	1	0	0	0
11	LA001										0	0	0	0	1	0	0	1	0	0	0
12	LA001	2019	3	4	1	1	15	1	0	1	0	0	0	0	1	0	0	1	0	0	0
13	LA001	2019	3	4	1	1	15	2	0	1	0	0	0	0	1	0	0	1	0	0	0
14	LA001	2019	3	4	1	1	16	1	0	1	0	0	0	0	1	0	0	1	0	0	0

Time information for connecting to AS-LUNG and other environment data

Take home messages

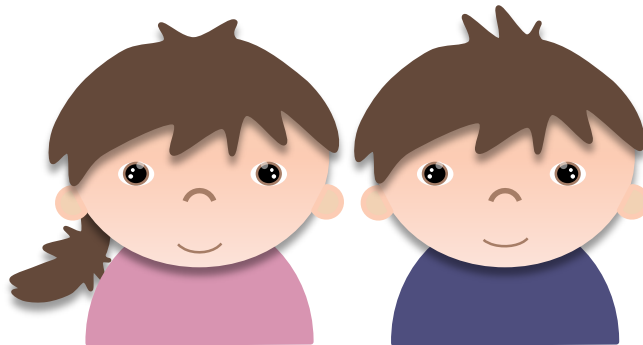
- For international comparison, core questions should be the same for all subjects across different areas. Please translate the core questions based on faithfulness, expressiveness, and elegance in your own language.
- You may design more questions based on knowledge of local PM_{2.5} sources or customs. But do keep the questionnaire not too long. Otherwise, your subjects will lose patience and give your random answers
- When designing questions, always think how I analyze these questions with environment data to fit our research purposes; be sure to use plain languages for your subjects
- Think about the possibility to collaborate with scientists in other fields using the same questions or combining other datasets
- Make sure the consistency of time-resolution of questionnaire and environment data



QA/QC of the survey data

- Quality Assurance/Quality Control
- Introduction of Core Questions in Questionnaire
- Some tricks to prepare dataset after survey

Thank you for your attention!
Any question and comment are welcome



Joanne HU
SC Candice LUNG