GEA1000 QUANTITATIVE REASONING WITH DATA
TUTORIAL 2

Please work on the problems before coming to class. In class, you will engage in group work.

A Case Study on Recycling Rates in Singapore

You are a student of a new course NEA1000 *Environmental Matters of Singapore* and you have been assigned to do a group project on recycling of Paper/Cardboard waste in Singapore with your classmates, Chuan and Tammy. The file `recyclingstats.pdf` was compiled by Chuan using data found on the National Environment Agency website. For the tutorial discussion, we assume that the data refers to domestic recycling only.

1. You have been tasked to understand recycling rates of Paper/Cardboard in Singapore from 2016 to 2021. While researching, you come across the following article from The Straits Times.

> SINGAPORE – The domestic recycling rate in 2022 fell to 12 per cent – the lowest in more than a decade – because less paper, cardboard, textile and leather were exported for recycling, the National Environment Agency (NEA) said on Wednesday.
>
> Between 2012 and 2018, Singapore's domestic recycling rate stood at between 19 per cent and 22 per cent. The rate declined from 2018 as China banned the import of 24 recyclables, including paper and plastics.
>
> The domestic recycling rate is the percentage of waste recycled from households, shophouses, education institutions, petrol stations, hawker centres and places of worship.

   a. You would like to see if the data found by Chuan reflects a decline in domestic recycling rates from 2018, as suggested by the second paragraph of the article. Referring only to the file `recyclingstats.pdf`, fill up the following 2x2 contingency table, and determine which period, before 2018 or 2018 onwards, is positively associated with recycling of Paper/Cardboard.

	Before 2018	**2018 onwards**
Recycled ('000 tonnes)		
Disposed ('000 tonnes)		
Total ('000 tonnes)		

b. Your groupmate, Tammy, looks at your table in (a) and points out that the amount (in tonnes) of Paper/Cardboard recycled from 2018 onwards was significantly greater than that before 2018. She claims that, as such, the period from 2018 onwards is positively associated with recycling of Paper/Cardboard. Is Tammy's claim correct? How can you reconcile her claim with your answer in (a)?

c. From the file `recyclingstats.pdf`, Tammy computed the rate of Paper/Cardboard waste generated before 2018 to be 2328.1/(2328.1+4343) =34.9%. Upon looking at her answer and yours in (a), Chuan immediately claimed that more than 34.9% of all the recycled Paper/Cardboard waste was done before 2018. Is Chuan's claim correct? Describe his thought process in leading to the said conclusion using the rules you learnt in Chapter 2.

d. During a group meeting, Tammy shares that she has found data related to Singapore's domestic recycling rate from 2012 to 2018, presented in the table below. Chuan, who has read the same article from The Straits Times, suspects Tammy's data to be false. Based on the above excerpt of the news article, explain Chuan's suspicion.

% of waste recycled in…	Households	Shophouses	Educational Institutions	Petrol Kiosks	Hawker Centres	Places of Worship
2012	3.9	9.1	10.2	8.9	5.6	14.3
2013	4.1	9.3	12.3	9.1	5.6	14.2
2014	3.8	9.7	15.6	8.7	7.5	15.6
2015	4.5	10.3	16.1	10.2	8.2	16.1
2016	5.3	10.1	16.9	11.2	9.3	15.2
2017	5.7	10.1	16.8	11.5	9.5	14.9
2018	6.1	10.2	17.1	12.3	10.1	15.6

Tammy decided to conduct a study to understand the prevalence of Paper/Cardboard recycling amongst NUS students. She stood at the main entrance of Town Plaza in front of the spiral staircase over a period of two weeks during her free time. Knowing that she would not be able to interview everyone who was going to walk past her, she decided beforehand to only survey one student out of every seven who walks past her, starting from the first person she sees when she arrives at the entrance. She collated the information she gathered in the data set recycling_survey.csv. Below is a brief description of the variables:

Variable	**Description**
`year`	The student's year of study.

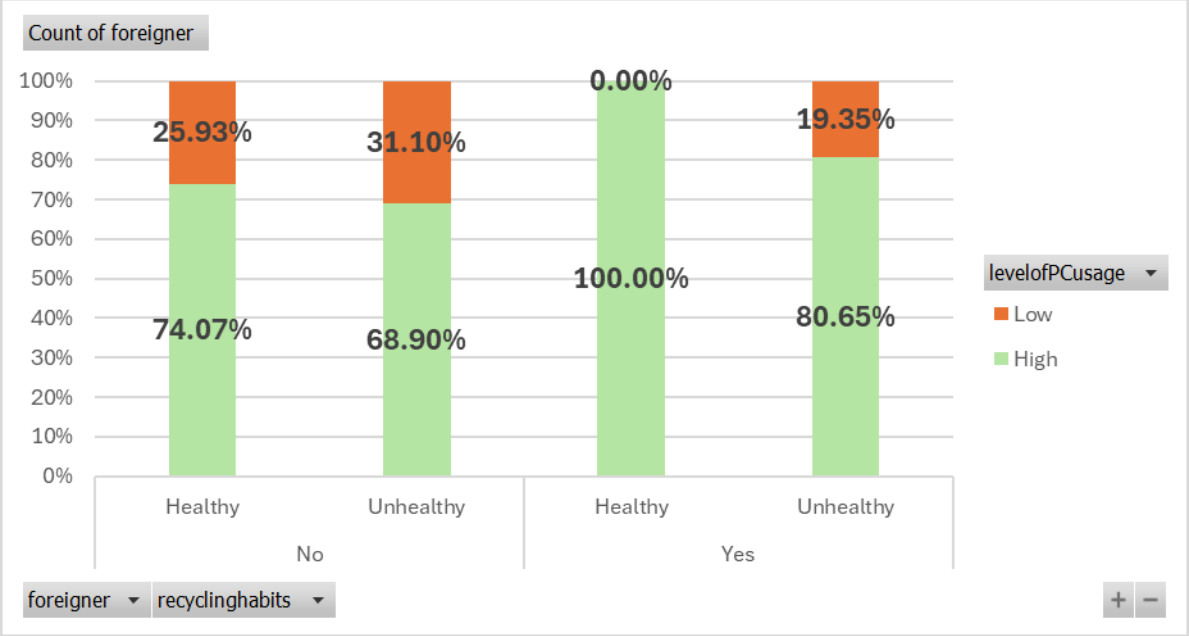
`use_monthly`	The average monthly usage of Paper/Cardboard of a student. • Often – At least once a day for the whole month • Sometimes – Once in two to six days inclusive • Seldom – At most once a week, but at least once in two weeks • Rarely – Less frequent than once in two weeks
`recycle_monthly`	The average number of times a student brings his/her recycled materials to a designated recycling bin in a month.
`stay_campus`	Whether a student stays on campus.
`foreigner`	A student is a "foreigner" if he/she is not a Singapore Citizen or a Singapore Permanent Resident.
`faculty`	The faculty which a student belongs to in his/her course of study. Students from NUS College are represented as "NUSC".

e. Chuan hypothesises that students who have a lower average monthly usage of Paper/Cardboard would bring their recycled materials to a designated recycling bin more frequently, as these behaviours both suggest that a student is more environmentally conscious.

   He suggests considering an average monthly usage of Paper/Cardboard of a student to be **"low"** if the student's response was **"Rarely" or "Seldom"** (otherwise, it is considered "high"); and considering a student to have **"healthy"** recycling habits if the student brings the recycled materials to a designated recycling bin **more than twice on average each month** (otherwise, it is considered **"unhealthy"**). What is the association between having healthy recycling habits and a high average monthly usage of Paper/Cardboard within Tammy's data set?

f. Chuan is not convinced that his hypothesis is entirely flawed. In fact, he suspects that the year of a student may have influenced the overall association between recycling habits and the usage of Paper/Cardboard. In his own exploration of the data, Chuan claims to have observed Simpson's Paradox when the data was sliced by a student's year. Verify if Chuan's claim is true.

g. Upon seeing what Chuan did, Tammy decided to explore the data set as well. She sliced the data set using whether a student is a foreigner as a third variable and analysed the association between recycling habits and the use of Paper/Cardboard. She then produced the following stacked bar plots:



Tammy claims that from her plots, Simpson's Paradox is not observed, and thus "whether a student is a foreigner" is **not** a confounder in the association between recycling habits and the usage of Paper/Cardboard. Based only on Tammy's plots above, is Tammy's claim true?

h. "We can have a good understanding of recycling of Paper/Cardboard amongst NUS students from the study because 2000 students responded to the survey, and systematic random sampling was used," Tammy asserts. Would you agree with her assertion?