

## Exercise

# Standardizing Data

## Summary

1. Module: Understanding Data<sup>1</sup>
2. Lesson: Organizing Data
3. Sub-topic: Standardizing a data table
4. Objective: Identify and correct common data entry mistakes
5. Time Allotment: 30 minutes

## Steps

### Procedure

Each of the following data tables breaks several of the basic rules of data standardization. Please identify the mistakes and decide how to correct them.

## Further Practice

1. City crime data

Offender ID	Date/Time	Location	Offense
729	07-10-2014/5:32 PM	Corner of Preah Sihanouk Blvd and Charles de Gaulle Blvd	Auto theft
8923	03/12/17/03:20	Wat Botum Park	Assault
3287	03-14-2010/12 PM	--	Car theft

---

<sup>1</sup> This lesson was adapted from the World Bank's Introduction to Data Literacy training manual by Eva Constantaras, and adapted by Yan Naung Oak, Open Development Cambodia and Open Development Initiative, and is licensed under a [Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License](#). For full terms of use, see [here](#).

	02-31-2019/0212	11°34'34.8"N 104°55'17.0"E	Unknown
--	-----------------	-------------------------------	---------

## 2. Healthcare resources

Country	Cambodia	Thailand	Myanmar	Vietnam
Doctors per 1,000	1.658	4.70	3.59	88
Hospital bed density	N/A	21 per 10,000	6/10,00	--
Births attended by skilled birth	2009-2014: 89.0	2013-2016: 99.1	2015-2016: 60.2	2011-2014: 93.8

3. Go to page 34 of this [report](#). Look at Table 3.6 Prevalence of a disadvantage in Health. What data standardization is needed?

**Table 3.6** | Prevalence of a disadvantage in Health

		Population with a disadvantage (%)		
Domain	Indicator	Union	Urban	Rural
Health	<i>A disadvantage in any indicator</i>	12.1	8.2	13.8
	Disability	5.1	4.4	5.4
	Child and youth mortality	7.4	4.0	8.8

## Assessment

Data standardization is necessary for any type of quality data analysis. Try to find a dataset with a complex format and identify the steps necessary for data standardization. We will be going over how to make these changes in the data cleaning lab with tools that allow transformations on an entire dataset.

