

# Tables vs LongForm Dataframes

*Jimmy Oh*

Department of Statistics  
University of Auckland

## Abstract

A comparison and brief commentary of two different types of ‘tables’, one that is intended for a human (Table) and one that takes machine readability into consideration (LongForm Dataframe).

## 1 Introduction

A ‘table’ is a very broad term that encompasses a diverse range of presentation formats. This article compares two such ‘tables’, looking in particular at the impact of structure to readability by humans and computers.

The first, which we will call a *Table*, is an example of a common ‘table’ intended for human consumption. The second, which we will call a *LongForm Dataframe*<sup>1</sup>, is a very simple type of ‘table’ with minimal structure, but one that is highly suited to machine processing.

We hope this article will convince the reader that Tables are suited to present data, but are unsuited for providing data for re-use. Conversely, LongForm Dataframes are suited for providing data for re-use (including using it to create a Table of the data), but are unsuited for direct human consumption.

---

<sup>1</sup>Mainly to differentiate it from other ‘tables’. A LongForm Dataframe is still a form of ‘table’, but one that is structured long form.

## 2 Table

Presentation of data can be done in a variety of ways, and one effective way is to present it in a tabular format, using layout to communicate hierarchical relationships.

A human reader, with minimal training<sup>2</sup>, can easily decipher and interpret these relationships. The layout helps the reader easily group the data by category, and also to easily find a specific value.

			<b>Full-time</b>
1	<b><i>New Zealand</i></b>		
2	<b>Auckland</b>		
3	Accounting	Male	250
4		Female	200
5	Economics	Male	40
6		Female	30
7	Statistics	Male	120
8		Female	130
9	<b>Wellington</b>		
10	Economics	Male	140
11		Female	120
12	Statistics	Male	150
13		Female	155
14	<b><i>Australia</i></b>		
15	<b>Sydney</b>		
16	Accounting	Male	720
17		Female	490
18	Economics	Male	480
19		Female	430

Table 1: An example of a Table as defined above. The data is entirely fictional and purely for demonstration purposes. While it is easy for a human reader to understand that the value 250 corresponds to Male Accountants in Auckland, New Zealand, this is not easy for a computer to process.

Unfortunately, such tables are very difficult to read in and analyse with a computer, as the hierarchical relationships are often solely based on visual layout - something that is easy for a human to decipher but not so simple for a computer.

---

<sup>2</sup>Often covered in mandatory education, and hence can be assumed in the vast majority of readers.

### 3 LongForm Dataframe

One very simple way to store data is to store it in *long form*. Because this is such a simple and basic format, there is no single explicit name for this format, but we will call it the *LongForm Dataframe*. The key feature of the LongForm is that each row contains all the information for that individual or observation. This eliminates all need to decipher relationships to understand the data, making it ideal for processing with a computer.

	Country	City	Profession	Gender	Full-time
1	New Zealand	Auckland	Accounting	Male	250
2	New Zealand	Auckland	Accounting	Female	200
3	New Zealand	Auckland	Economics	Male	40
4	New Zealand	Auckland	Economics	Female	30
5	New Zealand	Auckland	Statistics	Male	120
6	New Zealand	Auckland	Statistics	Female	130
7	New Zealand	Wellington	Economics	Male	140
8	New Zealand	Wellington	Economics	Female	120
...					
13	Australia	Sydney	Economics	Male	480
14	Australia	Sydney	Economics	Female	430

Table 2: An example of a LongForm Dataframe. The data is entirely fictional and purely for demonstration purposes.

The same reason that makes LongForm so easy to manipulate with a computer (lack of structure) makes it difficult to use directly, making it a poor choice for a human reader. However, because it is easy to manipulate with the computer, this also makes it easy to represent this data in several different ways that may be more appropriate for direct human consumption. With a small amount of training, and with the aid of a computer package, it is easy to convert a LongForm Dataframe into any number formats for presentation (including Tables). The converse however is not true. It is often extremely difficult to take something that is already made for direct human consumption, and convert that to another form, whether that is for computer consumption, or even a different way of presenting the data for human consumption.