

Notes on Tidy Data

GW Dean

2023-01-26

This document is written to help me understand the contents in Hadley Wickham's 2014 paper "Tidy Data" from the *Journal of Statistical Software*. It is intended to be an informal summary/commentary that will help me get at the main essence of the work. In addition, it will serve as a way to practice using RStudio and Github as a preparation for my upcoming presentation at UMass Dartmouth.

High Level Overview of the Document

The first step in understanding a paper such as this is to gain a high-level overview of the structure of the document. The version I have can be found at the following link:

<https://vita.had.co.nz/papers/tidy-data.pdf>

Looks like a final pre-print, as there is no date listed. The website for the *Journal of Statistical Software* can be found here:

<https://vita.had.co.nz/papers/tidy-data.pdf>

R Markdown

This is an R Markdown document. Markdown is a simple formatting syntax for authoring HTML, PDF, and MS Word documents. For more details on using R Markdown see <http://rmarkdown.rstudio.com>.

When you click the **Knit** button a document will be generated that includes both content as well as the output of any embedded R code chunks within the document. You can embed an R code chunk like this:

```
summary(cars)
```

```
##      speed      dist
##  Min.   : 4.0    Min.   : 2.00
##  1st Qu.:12.0    1st Qu.: 26.00
##  Median :15.0    Median : 36.00
##  Mean   :15.4    Mean   : 42.98
##  3rd Qu.:19.0    3rd Qu.: 56.00
##  Max.   :25.0    Max.   :120.00
```

Including Plots

You can also embed plots, for example:



Note that the `echo = FALSE` parameter was added to the code chunk to prevent printing of the R code that generated the plot.