

交通事故統計の再集計と社会貢献

水越明哉

交通事故統計グラフ	223
グラフ作成のわけ	223
Excel ファイルが使いにくい点	224
改善希望点、あるいは愚痴	224
整然データ	225
社会の一員としての貢献	225
2023 年の交通事故統計	226
参考	226

交通事故統計グラフ

警察庁のウェブページで公開されている交通事故統計を再集計してグラフにすることを私は趣味にしている。毎月 1 回、前月分の統計の Excel ファイルが公開されるので、過去のファイルとまとめて月毎の時系列データとして csv ファイルを生成して、GitHub に載せて公開している。Excel ファイルにはいろいろな切り口のデータが入っているが、その中から都道府県別と、年代別の情報を収集して 2 つの csv ファイルにしている。この趣味を始めたのは 2020 年頃であるが、そのときに公開されていた過去ファイルと合わせて、2014 年 10 月からの月毎のデータになっている。

グラフ作成のわけ

なんでそんなことをしているのか、と思う人もいるであろう。ひとつは自分なりのオープンソースソフトウェアプロダクトを作りたいと思ったからである。ruby のスクリプトで実装しているが、自分で自由に設計できるソフトウェアはなかなか楽しいものである。ただ、最近は毎月のデータ更新以外はできていなくて雑な設計が残っていたりする。

また、交通事故の死亡者数の減少傾向がどのように推移してくのか、気になっていたという点もある。2004 年（平成 16 年）にピークだった交通事故の死亡者数は 2022 年（令和 4 年）まで毎年減少の傾向がある。今後、この傾向がどのようになっていくのか、年間の統計を待たずに月毎の統計を見て毎月確認したい、という思いがある。ちなみに、単純な死亡者数ではなく、人口 10 万人あたり死者数は 1970 年（昭和 45 年）がピークである。1970 年（昭和 45 年）の 16.33 に対して 2004 年

(平成 16 年) には 5.83、2022 年 (令和 4 年) には 2.08 とかなり減っている。

もうひとつ、理由がある。これが一番大きい理由である。せっかく警察庁から公開されているデータなのに、簡単には使いこなせない形なのである。使いやすい形にして公開して社会に貢献できたらと思ったのである。念の為に書いておくと、警察庁や日本の政府には、こうした情報を公開していただいととても感謝している。公開されていなければ見ることも再集計もできない。であるがやはり、もっと使いやすい形になっていてほしい。待っていても実現しそうにないので、代わりにやってみよう、という感じである。

Excel ファイルが使いにくい点

なにが使いにくいのか。まずは Excel ファイルであることである。ただファイルを開いて目で見ただけであれば Excel ファイルでも良いのですが、ruby などのプログラミング言語で機械的に処理しようとする Excel ファイルは扱いづらい。何らかのライブラリ、私の場合は roo というライブラリを使っているが、そういうものを使って読むことになる。csv ファイルのようなテキスト形式であればずっと楽に取り扱えるのである。

(じつは、Excel ファイルではなく csv ファイルで公開されていることに最近気が付いた。2020 年に ruby スクリプトを作っていたときには Excel ファイルだけだったのであるが、2021 年度からは csv ファイルも公開するようになっていたようである。毎月見ていたページなのに、2024 年のこのコラムを書くまでになるまで 3 年間気が付かなかったのである。csv ファイルを見たときはかなりショックであった。今までの Excel ファイルを加工していた苦労は無駄だったのか、そしてこのコラムのネタも無くなってしまうじゃないかと。でも、改めて考えると、2015 年から 2020 年の過去の Excel ファイルも使ってデータを出したいので、csv だけでなく Excel ファイルも扱わないとならないのである。そしてあらためて csv ファイルを見ると、Excel ファイルの表をそのまま csv 出力したようなファイルになっている。後で書く整然データの形ではなく雑然データだということである。そういう意味では、結局はあまり変わらないのかもしれない。)

また別の理由として、月毎の Excel ファイルになっていて、何ヶ月もまたいで見るには複数の Excel ファイルを開かなければならないということである。例えば 2023 年の月毎の都道府県別交通事故死亡者数を調べたければ、2023 年 1 月から 2023 年 12 月の 12 個のファイルを開かなければならないのである。できれば、最新の 1 ファイルに過去のデータも全部入っていてほしい。そうすればその 1 ファイルを読むだけで過去からの推移も過去との比較もできるからである。

改善希望点、あるいは愚痴

もうちょっと細かい改善希望点も書いておく。ただの愚痴なのかもしれない。まず

は、複数のフォーマットが使われている点である。統計する項目が変わってフォーマットが変わるのは、やむを得ないと思う。例えば「原付」の区分が2024年から「一般原付」と「特定小型原付」に分かれた。「特定小型原付」は電動キックボード等のことであろう。こういうのは時代に合わせて変わっていくのは必要なことである。一方、改善してほしいのは、1月から11月のフォーマットと、12月のフォーマットが違うことである。1月から11月は、「交通事故統計月報」という月次データで、12月は「交通死亡事故の発生状況及び道路交通法違反取締り状況等について」という年次データなのである。1年間のまとめとしての年次データはあって良いのだが、それはそれとして、毎月同じフォーマットで12月の月次データを作って欲しいところである。

また、Excelのシートをまたいでいるデータも直してほしい。例えば「表2-5 年齢層別・状態別死者数の推移」と「表2-6 65歳以上高齢者の年齢層別・状態別死者数の推移」などである。65歳以上をまとめた表と、65歳以上が詳細に書かれた表が別にあるのである。全年代の詳細な情報を得るためには2つのシートを読んでマージする必要がある。シートをまたがずに、2つをまとめたシートにしてほしいところである。

整然データ

結局、どのようなデータであれば嬉しいのか、ということも書いておく。それは整然データです。交通事故統計のExcelファイルは、例えば都道府県別であれば、行が都道府県、列が「発生件数」「死者数」「負傷者数」であり、中に人数が入っている表になっている。人数が縦横に2次元に入っている。それを、各行に都道府県と「発生件数」または「死者数」または「負傷者数」と、そして人数を入れた表にしたものが整然データである。人数が1次元に入っている。整然データではない形式を雑然データというらしい。雑然データにも、そのまま人間が目で見える場合はわかりやすいというメリットがある。機械的に処理する場合は整然データのほうが使いやすい。Excelファイルをそのまま見たい人もいるだろうから、今のままのものも残したうえで、整然データに毎月追加される形で公開していただけると嬉しい。私の公開しているcsvファイルは、警察庁の公開しているExcelファイルを整然データに変換したものと言えるのかもしれない。ただ、いまあらためて生成したcsvファイルを見ると、完全な整然データとは言えないかもしれない。いずれ改善して整然データにしたいところである。

社会の一員としての貢献

上にも書いたように、交通事故統計を公開している警察庁や日本の政府には感謝している。情報公開の責任を果たしていると思う。ただ、そのデータを利用しようとする人が、みんなが同じ苦勞をして手元で整然データを作るのは、広い目で見ると無駄が多い気がしてしまうのである。従って、整然データの形に再集計して公開す

る人がいると、社会への貢献と言えるのかなと思う。もしかして、そのうち、警察庁がそれをみて、整然データで公開を始めてくれるかもしれない。その時を夢見て、毎月の更新をしていこうと思っている。

2023 年の交通事故統計

ちなみに、2024 年 3 月に公開された 2023 年（令和 5 年）の交通事故統計を見ると、2004 年（平成 16 年）以来下がり続けていた交通事故死亡者数が、2023 年に増加に転じてしまった。今まで色々なことをして死亡者数を下げてきたが、ついに下げ止まってしまったということなのだろう。シートベルト着用率の向上や、飲酒運転の厳罰化などがこれまでの減少傾向を作ってきたと思う。一度下げ止まってしまったのですが、これから期待するのは AI による自動運転であろう。もっともっと交通事故が減ってくれることに期待する。

参考

統計表 | 警察庁 Web サイト

<https://www.npa.go.jp/publications/statistics/koutsuu/toukeihyo.html>

整然データ

https://ja.wikipedia.org/wiki/Tidy_data

私が再集計して公開している交通事故統計

<https://github.com/Akiyah/monthly-traffic-accidents-in-japan>