Traffic data visualisation

Anni Niskanen

I found an interesting dataset which consists of vehicle counts on the road Stampfenbachstrasse in the center of Zurich. The counts are reported for each hour, both directions (either roughly north toward Schaffhauserplatz and roughly south toward Stampfenbachplatz), and for three vehicle types, classified by the length of the passing vehicle. The vehicle types are motorcycles (possibly including bicycles as well, but I'm not sure), passenger cars, and trucks. The traffic data has been collected since late 2007 and has continued to this day, however, due to construction work no data from most of 2018 and 2019 exists.

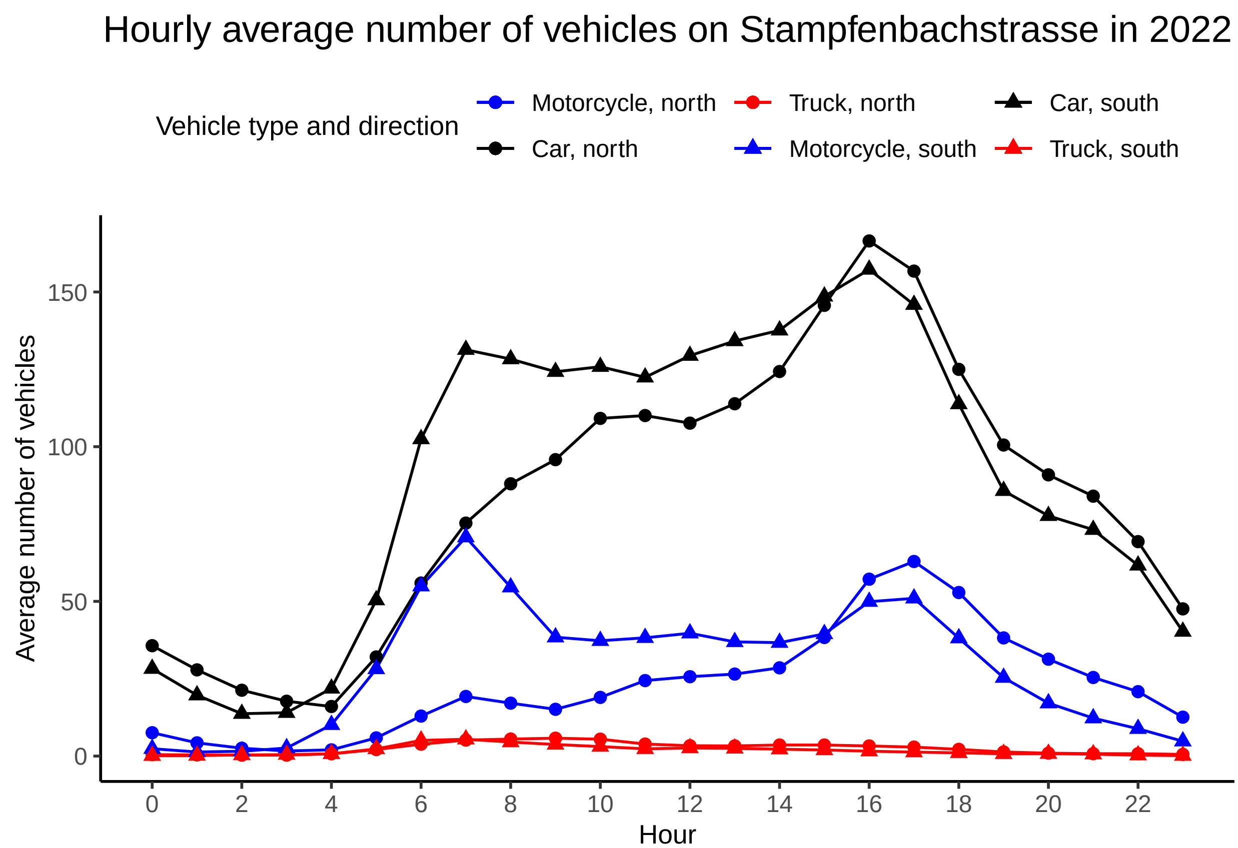
I set out to visualise two different phenomena in this data. Firstly, I wanted to plot what the daily trends of different vehicle types looked like. From such a visualisation, one could answer questions like "How much more prevalent are cars compared to other vehicles on this road?" and "Are there rush-hours on this road, and when?". I could visualise the trends by counting the average number of each vehicle type passing during each hour of the day. I chose to calculate the averages over the data from 2022, because it was the most recent year that had data from the entire year. Because the averages form time series, they would naturally be visualised as a line plot. In addition to vehicle types, I decided to plot curves separately for both directions, north and south. I chose to indicate vehicle type with colour and direction with shapes drawn on top of the curves (these choices will be discussed and justified later).

A few interesting trends can be seen from my visualisation. Firstly, traffic is at its highest during day and diminishes at night, as is expected. There are local maxima, i.e. rush-hours, at 7 a.m. and 16-17 p.m., and minima at 2-4 a.m. Another expected observation is that cars are the most common vehicle and there are fewer motorcycles and much fewer trucks. There is an interesting exception to this rule, though, as at around 5-7 a.m. there are momentarily almost the same number of motorcycles going south and cars going north.

More people go south than north in the morning rush-hour, and a bit more people drive north than south in the evening rush-hour. This could possibly be explained by the structure of the city of Zurich: although I'm no expert, a quick view of the map reveals that the city center is south the road, so people might commute to work this way, and drive back north when they are going home. It is also interesting that the difference between people driving north and south is much larger in the morning than in the evening. Perhaps a good number of people favor other roads and paths when they need to drive north, or there are quite a few people driving to the city center in the afternoon to spend the evening.

The last observation I can gather from the visualisation is that there seem to always be more trucks driving north than south. Unlike other traffic which has two maxima corresponding to the two rush-hours, trucks move mostly during the morning at 6-10 a.m., although there are a few moving in the evening as well. Truck traffic completely ceases at night.

I tried to follow Tufte's principles when making my visualisation, that is, I tried to use as little ink as possible while still displaying the data fully. The visualisation was made with R's ggplot package, which adds a grey background with white lines by default. The background and its lines were removed, as they only contribute to non-data ink (although they might help the viewer read the exact data point values from the axes). The shapes indicating vehicle direction were made sufficiently large for perception, but not too large in order to not waste data-ink. I believe the result is a visualisation with high data-ink ratio and data density, that is, time series fill the plot area quite nicely, leaving e.g. no large empty corners. Although it is not strictly a principle of Tufte, I also decided to make the visualisation wider than higher, as that makes it easier for a human viewer to read it.

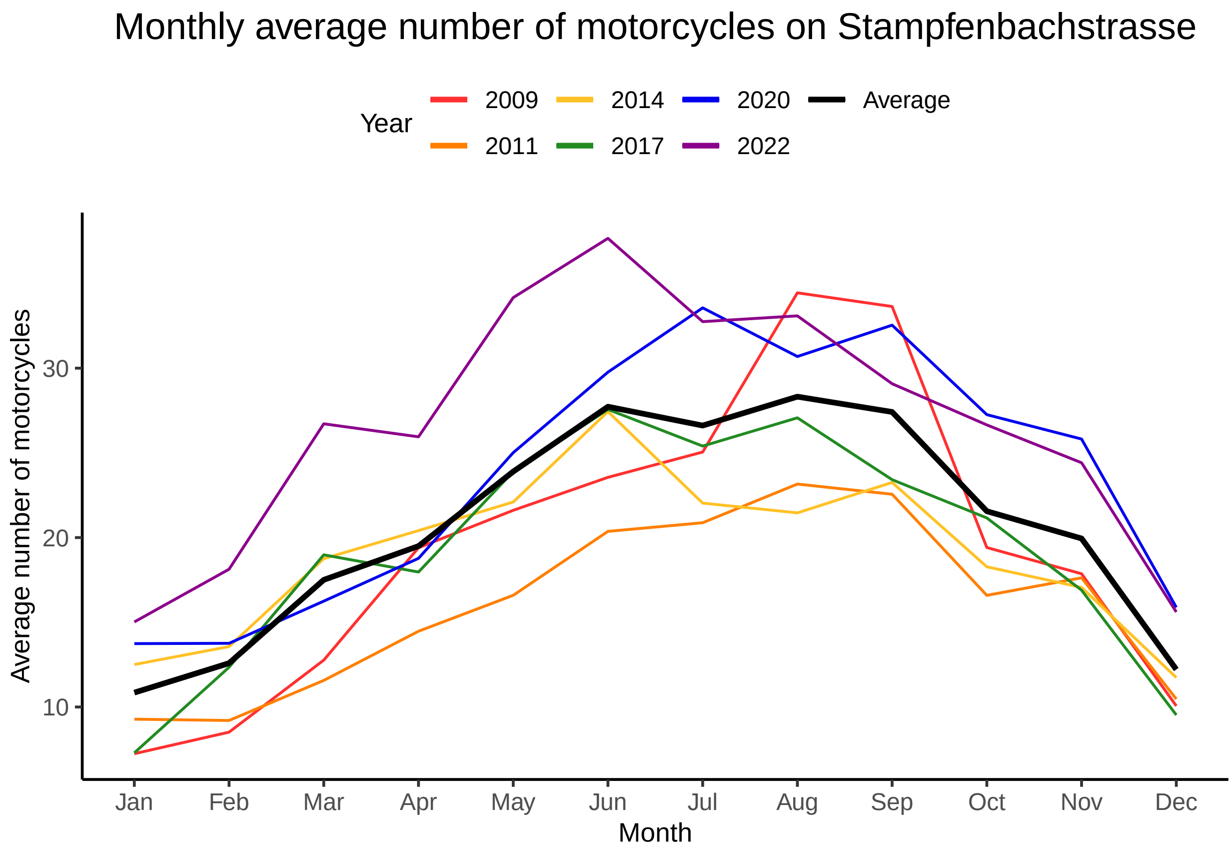


Another thing I considered when making the visualisation were Gestalt laws. The Gestalt law of similarity states that similar objects are perceived to be grouped together. I utilised this law by grouping vehicle types by colour and directions by shape, although the law is much stronger in the first case, that is, long lines of the same colour are perceived to be similar much stronger than two lines with the same shapes drawn on them. Nonetheless, due to the Gestalt law of similarity, the same vehicle types (regardless of direction) are considered a group by the viewer, and the times series for all vehicles driving to the same direction (regardless of vehicle type) are perceived as a group, too. The same vehicle types are viewed even stronger as a group because of another Gestalt law, the law of proximity. It states that things that are near to each other appear to belong in the same group. Because all the sets of two time series of the same vehicle type are quite close to each other most of the time, while maintaining a distance from all other time series, the three time series sets are considered separate groups, and the three vehicle types are further distinguished.

Colour and shape were chosen to represent vehicle type and direction, respectively, because they are intuitive. This is because they are pre-attentive features, that is, they are processed by the visual cortex quickly, subconsciously, and in parallel. Therefore colour and shape are perceived independently of each other and can intuitively express two separate attributes of a data point (here vehicle type and direction).

Not much thought was put into the colours of the visualisation. Black was chosen to represent passenger cars because it is perhaps a good "default colour" (although it is actually not a colour at all), similarly to how one might consider the car itself to be a default vehicle or mode of transportation. Blue and red were chosen because they are simple primary colours. Therefore, although the colour choices of the visualisation could likely be better, it can be pointed out that this colour combination is good in the sense that it is not problematic for colour-blind people.

After making the first visualisation, I wanted to visualise if anything interesting happened in the data on a longer timescale. Therefore I set out to visualise the monthly trend of motorcycles on the road, which could provide answers to questions like "How does the number of motorcycles change across seasons, especially between summer and winter?". This visualisation, similarly to the first one, would display time series data and should therefore also be a line plot. Unlike in the first visualisation, however, in this one the data from multiple years could be easily shown separately. Therefore I chose 6 years from the full data's time interval (2007-2023) so that there was recorded data from the entire year for each chosen year. Additionally, the chosen years should be somewhat evenly part, which in the case of this data turned out to be 2-3 years. I calculated and recorded the monthly averages of passing motorcycles for the chosen years. In addition, I calculated the average of the years' monthly averages, and plotted it with a curve a little wider than the other curves for clarity.



As one might suspect, in every year there are more motorcycles in summer compared to other seasons, especially winter. The maxima are quite widely distributed from June to September, and the minima are in December to February. Another interesting observation is that the total number of motorcycles seems to be increasing along the years, as the curves of the three latest years are mostly above the curves of earlier years. It is difficult, even impossible, to say if this is due to a supposed increase in motorcycle popularity, or if perhaps the same people who already own a motorcycle have used it more and more along the years. Finally, yearly variation, for example in weather, likely affects the trends. For example, perhaps spring came later than usual in 2011, and therefore fewer motorcycles than usual were in traffic in February to April 2011. Additionally, perhaps the summer of 2009 had exceptionally good weather for motorcycling, or a motorcycling event was in town then.

This visualisation was made similarly to the first one with Tufte's principles and human vision in mind: data-ink ratio was maximised and the visualisation has larger width than height. As these aspects were already discussed with the first visualisation, I will not discuss them more here and will instead concentrate on a (somewhat) new topic: colours. The average time series was first decided to the plotted in black in order to separate it from the other time series. After that, however, it was not trivial which colours should be utilised to separate the different years' curves from each other. An ordinal colourmap would not have worked, as then some curves would unavoidably have had very little contrast with the background and would have been much more difficult to see than others. However, the six time series to be plotted can be ordered by year, so I thought perhaps a nominal colourmap would be too arbitrary. I settled on a (slightly altered) discrete spectral colourmap, which should be semi-ordinal for anyone familiar with how e.g. a rainbow looks. It emphasises the increasing trend of motorcycles across the years. This colourmap can be problematic for colour-blind people, though. I tried to alleviate the problem by setting different luminance values for red and green, and blue and yellow, respectively, so that they might be separated by luminance instead of colour by a colour-blind person. However, as I had to consider the contrast to the background and the similarity of colours to each other, too, I didn't succeed very well, especially with red and green. Aside from the red-green colour-blindness issue, though, the colours of this colourmap are easily distinguishable from each other and the background. Therefore my choice of colours for the visualisation, although improvable, is adequate.