

Tableau challenges

Professor Roy A. Ruddle, School of Computing, University of Leeds.

January 2019

These challenges use off-street parking fines data from Leeds City Council (<https://datamillnorth.org/>) and you should use the “georeferenced car park fines.xlsx” data file that contains that data. The challenge dataset runs from April 2013 to March 2015, and has the following variables:

- PCN – Penalty Charge Notice
- Issue date – Date PCN was issued
- Area – The geographic area
- Car Park – The name of the car park
- Description of Offence – The parking offence
- Full Fine £ – Initial full charge (does not include any surcharges)
- Discount – the charge if payment made within 14 days.
- Paid in £ – Amount paid
- Post Code – Post code of the car park
- Latitude – Latitude of the car park
- Longitude – Longitude of the car park

Challenge 1 – Fiscal timeline

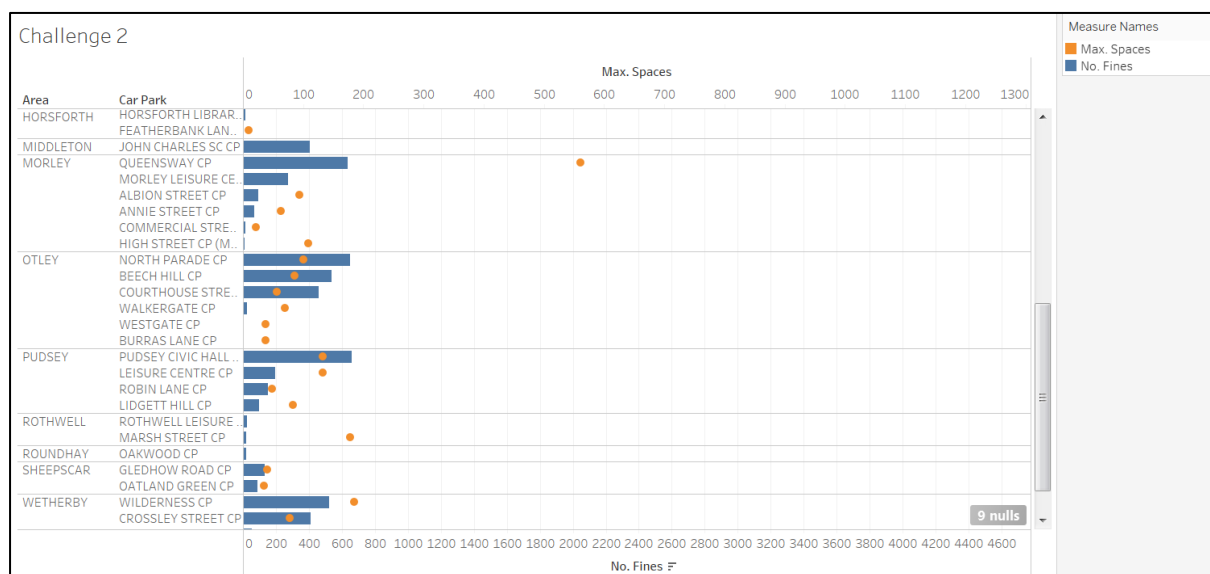
The Council have taken several initiatives to try to reduce parking offences, and wants to present a summary to the Executive Board. That summary needs to be in terms of the fiscal year (starts 1st April) rather than the calendar year.



Challenge 2 – Does car park size matter?

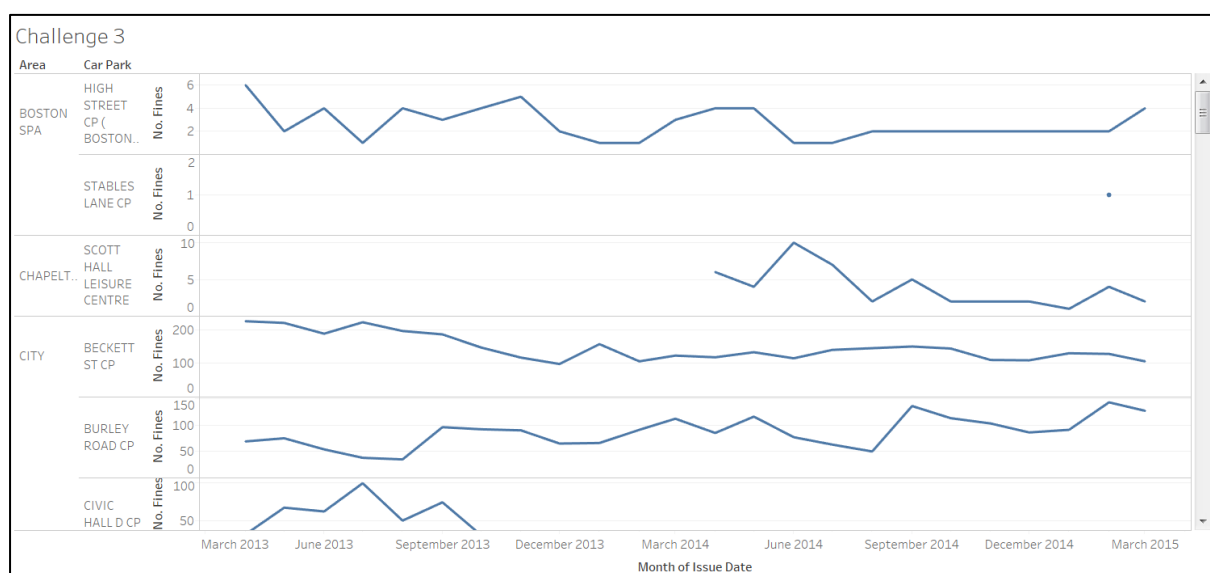
The Council want to gain a detailed understanding of parking violations, as a precursor to taking new initiatives. As a starting point, you have been asked to show the number of fines that are issued in the context of the size of each car park. The obvious choice is a scatter plot, but that is not very informative (try it!). However this dual axis chart is more effective.

By the way, someone is bound to notice the “9 nulls” message, so you had better be ready with an explanation!



Challenge 3 – Are there gaps in the data?

After noticing those nulls you decide to investigate how the number of fines varies longitudinally and whether there are any gaps in the data.

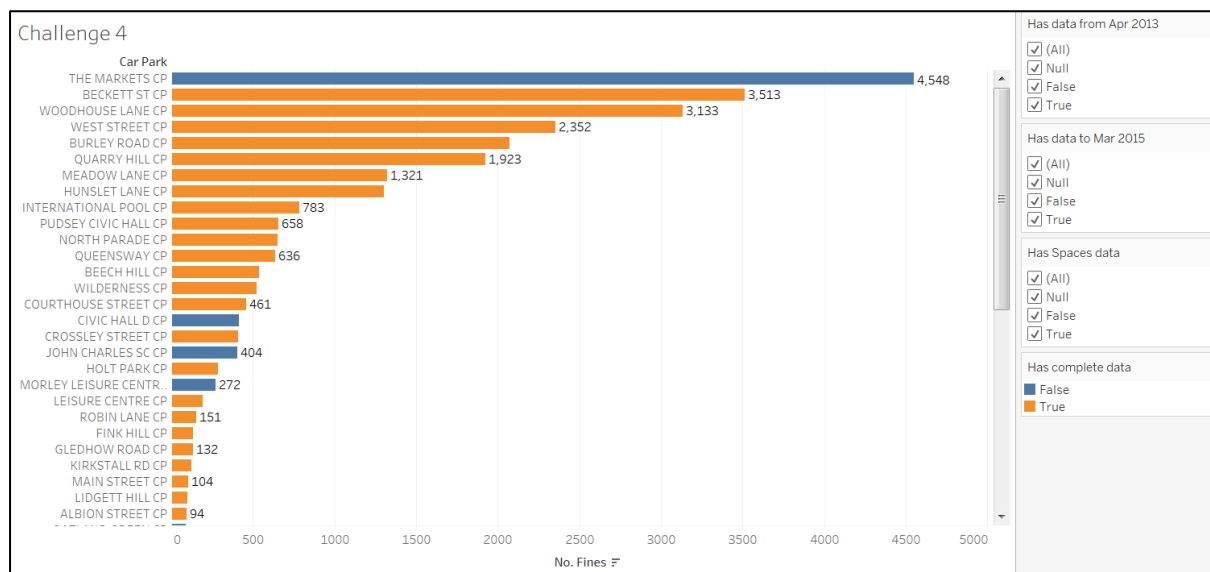


Challenge 4 – Data quality

There are clearly some issues with data quality, so you create three calculated fields to indicate whether each car park has: (a) a value for 'Spaces', (b) data from the beginning of the time period (April 2013), and (c) data to the end of the time period (March 2015).

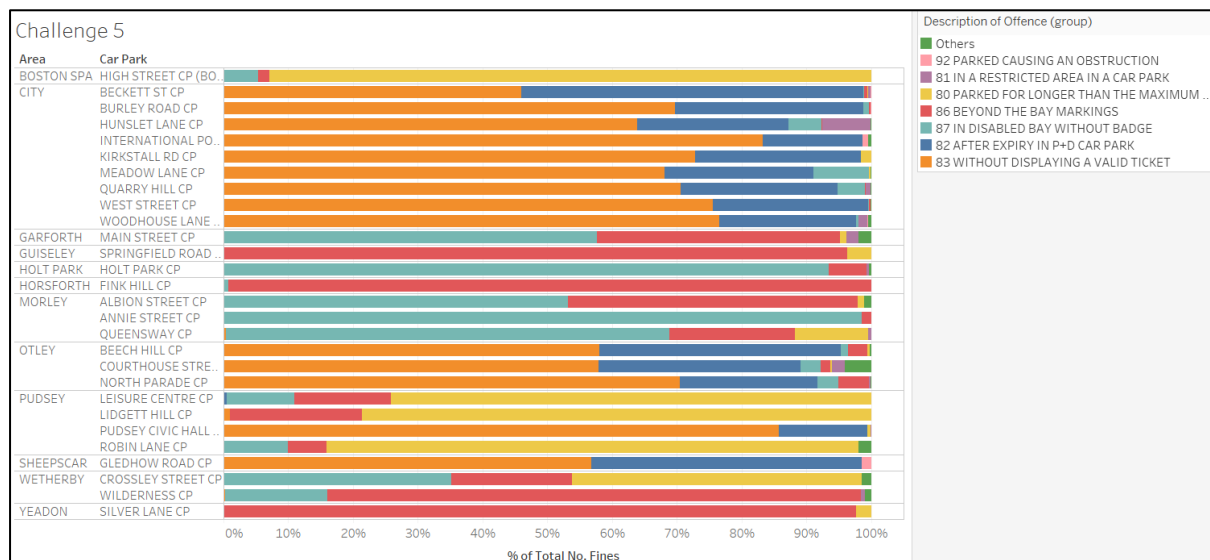
Hint: IFNULL() is useful for (a), and for (b) and (c) you can use level of detail to calculate the first/last issue date for a car park and then compare that date with the beginning/end of the time period.

That allows you to explain the quality issues to your colleagues. Clearly those issues will need to be investigated but, in the following analysis, you will only include car parks that have complete data and, therefore, satisfies all three quality criteria.



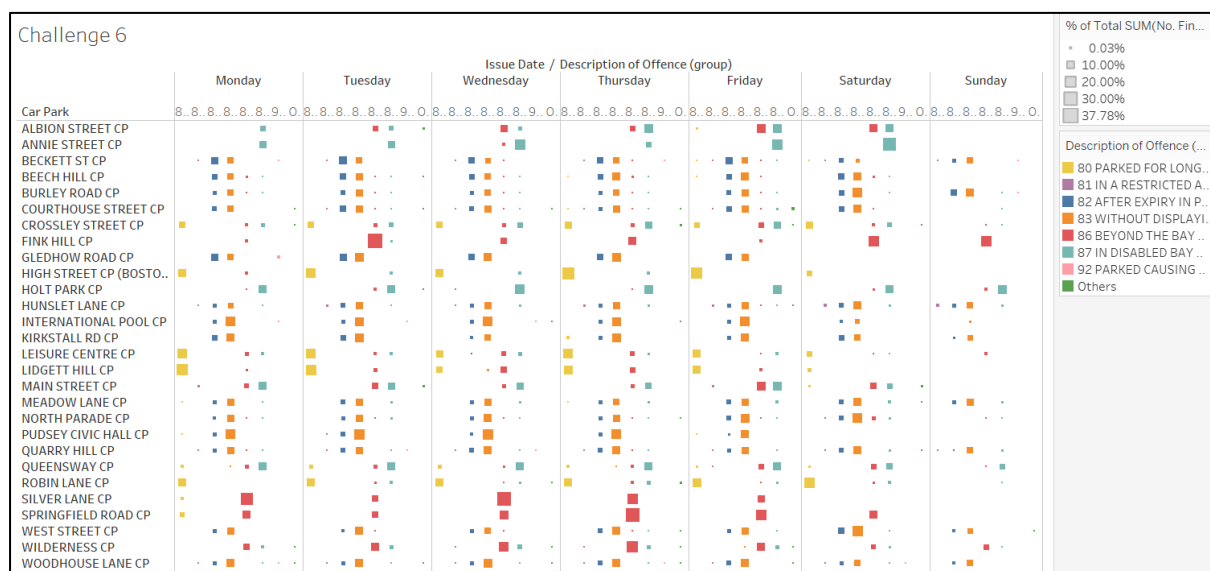
Challenge 5 – What are the most common offences in each car park?

Taking a data-driven approach, you group the six least common offences into 'Others', and calculate the percentage of fines for each offence in each car park. This reveals dramatic differences between some of the car parks, which will help the Council to target those new initiatives at specific locations (e.g., respect for the disabled in Holt Park, Garforth and Morley).



Challenge 6 – When are offences committed?

To gain more detailed insights you decide to visualize how types of offence vary with the days of the week. You could do this in several ways (experiment for yourself!), but a heat map is compact and allows you to see all of the data with the glance of an eye, rather than having to scroll.

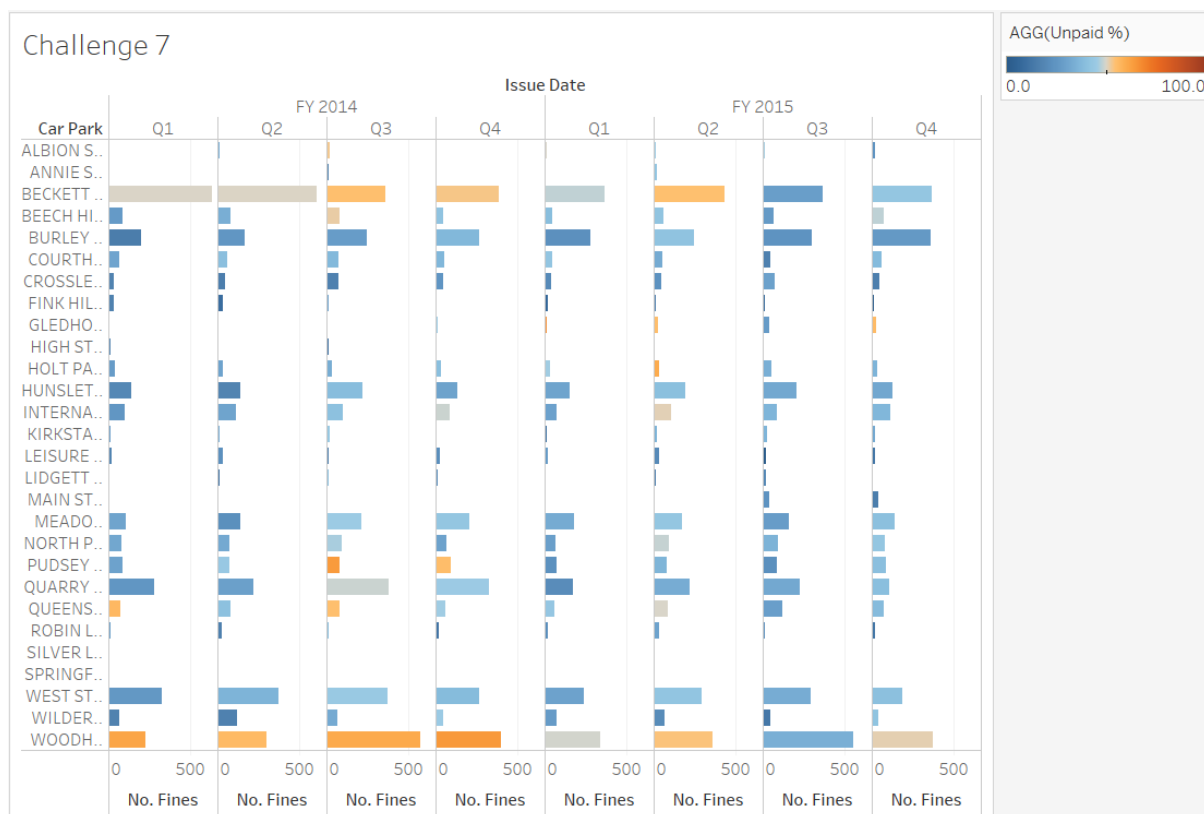


Challenge 7 – Are people paying their fines?

A crude way of differentiating between people who have vs. have not paid their fine is to check whether “Paid in £” is zero. Having categorised whether each fine is *paid*, *unpaid* or the data is *missing*, you can count the number of fines in each category, and then calculate the % of unpaid fines.

You decide to visualize the % unpaid fines in the context of the number of unpaid fines. This allows you to explain that the Council’s initiatives are reducing the % unpaid, but there is still scope for improvement.

It is interesting that Q3 in 2015 stands out as the quarter with the lowest % unpaid. Why might that be?



Challenge 8 – Where are people not paying?

You need a simple visualization that will draw the Executive Board's attention to the large number of unpaid fines in some city car parks, compared with the low number in other car parks. A map will be much more appropriate than bar charts and heat maps.

