Cycling safety in Scotland- Cycle collision hotspots

David Corner, Matt Pearce and George Macklon, Sustrans - Research and Monitoring Unit

1 Introduction

Using data from the STATS19 dataset, and accounting for the volume of cycling in an area, this paper identifies 20 of the most dangerous locations for cyclists across Scotland, based on the volume and severity of collisions over the past decade.

The paper seeks to help identify locations where an investment in cycle safety is most needed.
It must be stressed that this is only one approach to ranking dangerous locations, and the principle that underlies any ‘hot spot’ identification is to understand, in detail, the local characteristics of cycle safety.

2 Background

The STATS19 dataset holds a record of all reported road traffic collisions where a casualty has occurred. There are a number of weaknesses with this dataset (particularly underreporting of collisions and near misses), but it is nevertheless the most comprehensive dataset of its kind in Scotland.
This paper shows how this dataset can be used to identify locations where there is a high rate of cycle collisions, and in so doing, identify locations that would benefit from investment in improving cycle safety.

3 Methodology

There are three main steps taken to identify these ‘hotspots’:

1. Locations where there are ‘clusters’ of collisions are identified
2. These locations are scored using a number of variables
3. The locations are ranked by the sum of these scores

3.1 Identifying collision clusters

There are a number of spatial clustering methodologies available but the one used here is Density Based Clustering of Applications with Noise (DBSCAN). This algorithm groups together closely packed sets of points in some space – that is, points with many nearby neighbours. It is a commonly used method and used in mainstream scientific literature.

DBSCAN requires two parameters: ε (eps (the distance parameter)) and the minimum number of points required to form a dense region (minPts)[i]. For the purposes of this exercise, a distance of 20m was used for ε, and minPts was designated as 3. These values were assumed to represent the size of a typical junction and an appropriate size of cluster.

This algorithm was applied to the Easting and Northing co-ordinates of all cycle collisions in Scotland included in the STATS19 dataset (2005-2014).

This resulted in the identification of 253 collision clusters across Scotland, ranging from a single cluster of 22 collisions, to 134 clusters of the minimum required three collisions.

3.2 Scoring location variables

3.2.1 Variables used

When the list of collision clusters were ranked simply by the number of collisions in the cluster, it became apparent that this approach was too simplistic to accurately rank the different levels of risk at each location. The clusters with the greatest number of collisions were all located in Edinburgh, the area in Scotland with the second highest population size[ii] and with the joint highest proportion of residents who cycle as their main mode of transport[iii] (see Table 1). This paper concentrates on identifying locations that have a disproportionately large number of collisions relative to the cycling activity at that location. These are the true cycle safety 'hot spots'.

Table 1: Top 10 ranked locations (based on number of cycle collisions)

Location	Area	Number of cycle collisions 2005-2014
A7/Chambers street	Edinburgh	22
B901/London street roundabout	Edinburgh	15
Dalmeny street/Easter road	Edinburgh	15
Home street/A700	Edinburgh	14
Mayfield road/W Savile terrace	Edinburgh	13
Leith walk/London road roundabout	Edinburgh	11
Picardy Place roundabout	Edinburgh	9
Fountainbridge/Earl Grey street	Edinburgh	9
Dumbrae roundabout	Edinburgh	8
Corstorphine road/Murrayfield road	Edinburgh	8

Previous research by Sustrans has indicated that there are two variables that have the largest impact on the levels of cycling at any one location:

- The population 'Gravity' of a location – that is, the size of the population divided by the square of the distance from the location
- The proportion of commuters in the immediate vicinity who reported that they cycle to work in the 2011 Census

These variables have been calculated for each of the cluster locations. In the case of population gravity, this was calculated at 1km intervals up to 10km from the cluster location. In the case of the Census data, the value used was the maximum proportion reported of the Lower Super Output Areas (LSOAs – a geographical unit used by the Census) within 1km of the cluster location.

In addition, the number of collisions in each cluster was recorded and weighted according to how recently each collision had occurred. Collisions occurring in 2014 were weighted at 100% and collisions in 2005 weighted at 10%. Intervening years were weighted according to a curved line drawn between these values. This was done to account for any improvements in cycle safety in the past decade with lower values ascribed to collisions up to 2011, and greater prominence given to collisions that occurred after 2011, due to investment in cycling with an increased focus on safety following the publication of the Cycling Action Plan for Scotland (CAPS, 2013).

Finally, the number of serious and fatal collisions in each cluster was also recorded. This was to make sure that locations with a smaller volume of more severe collisions were not unfairly excluded from the ranking. These variables were also weighted according to the date of the collisions, using the same method outlined above.

3.2.2 Scoring method

To allow for weighting to be assigned to the different variables (so the clusters can be ranked according to different priorities[iv]), each variable was scored using the same system.

The percentile rank of each cluster was calculated for each variable[v], and the cluster received a score out of 10 according to which decile it fell into, with the minimum score being 1, the maximum score 10. Deciles were selected to give the greatest segmentation of the data without losing usability.

Across the clusters it was apparent that there were no fatal collisions recorded within any of the clusters and that only a couple of dozen clusters contained collisions in which there had been a serious injury to a cyclist. This meant that fatal collisions were irrelevant for the purpose of cluster comparison and so the ‘fatal collision’ variable was not included in the calculation of clusters’ final scores. The presence of only a small number of clusters which included collisions involving a serious cyclist casualty led to disproportionate differences, between clusters, in serious casualty collision scores relative to the number of serious casualty collisions. In order to reduce the impact of this on the final cluster scores, this variable received a lower weighting then the cluster size, population size and proportion of residents that cycle to work variables[vi].

Ascending scoring was used for the weighted cluster sizes with larger clusters receiving larger scores. However, as this paper concentrates on identifying collision clusters that are disproportionately large relative to the level of cycling in the area, population and cycle to work proportions were scored using a descending method.

The scores for each variable were summed for each cluster, and the clusters ranked accordingly, with high scoring clusters at the top of the ranking.

4 Results

Table 2 shows the top 20 high scoring locations identified by this method, along with the total number of cycle collisions that have occurred at each location between 2005 and 2014.

Table 2: Top 20 ranked locations (using scoring method)

Location	Area	Number of cycle collisions 2005-2014
A761/Arkleston road	Paisley	3
A726/Parkway roundabout	nr Erskine	4
A199/B1361/A6094 roundabout	Wallyford	3
B959/Robertson street	Dundee	4
Barrhead road/Peat road/Braidcraft road/Brockburn road roundabout	Glasgow	3
Glasgow road/Viewlands road	Perth	3
A8/Cathedral street	Glasgow	3
High street	Tranent	3
Main street/B6482/Newbattle road mini roundabouts	Newtongrange	3
Mearns road/A727	Clarkston	3
Dumbrae roundabout	Edinburgh	8
A7/Craigmillar Castle road	Edinburgh	7
A70/Juniper avenue	Edinburgh	3
Burdiehouse road/Straiton road/Lang long roundabout	Edinburgh	3
A823/Laburnum road roundabout	Dunfermline	3
Spiersbridge roundabout	Thornliebank	3
A8/Springhill parkway	Glasgow	3
B9006/Tower road	Inverness	5
University road west/Hermitage road	Stirling	7
Starlaw road/Boghall roundabout	Bathgate	5

The top 20 locations are not widely spread across the country and are typically within urban areas, in particular there was a high representation in and around Glasgow. Although Glasgow has a high population (which would contribute to a lower total score), the proportion of the population that cycle to work is relatively low (which would contribute to a higher score).

As Table 3 shows, the majority of the top 20 cluster locations are at either T or staggered junctions (7), or at roundabouts (8). There is a noticeable divergence between the types of location represented by the top 20 clusters and the types represented across all collision clusters. In particular, 40% of the top 20 clusters were located at roundabouts, whereas 18.5% were at roundabouts across the total selection of clusters identified.

Table 3: Cluster location types

Location type	Total clusters	Clusters in top 20
T or staggered junction	100	7
Crossroads	68	1
Roundabout	47	8
Not at junction or within 20 metres	16	1
Mini-roundabout	9	1
More than 4 arms (not roundabout)	8	1
Other junction	5	1

5 Conclusion

This paper has shown that it is possible to identify cycle safety ‘hotspots’ using a methodology that attempts to account for different levels of cycling across the country. It also highlights that the main risk to cyclists occurs when they directly interact with other vehicles – at junctions and roundabouts – rather than when simply sharing the carriageway.

It does not claim to be the definitive list of dangerous locations, nor that the variables included are the only ones that could be used, and as ever when using the STATS19 dataset it is important to recognise the limitations of the data in representing cycle safety, with a large volume of low severity incidents going unreported, and no attempt to measure the ‘near misses’ that contribute so much to feelings of insecurity.[vii]

The fact that there were no fatal cycle collisions contained within the collision clusters that were identified also highlights the difference in the nature of collision severities in as much as fatal collisions do not occur at the same locations as serious and slight collisions. In addition, due to their infrequency, it is not possible to define fatal collision ‘hotspots’ by combining the collision frequency in a limited location size with other relevant variables, in the way that serious and slight collision ‘hotspots’ can be. A different approach would be required to prioritise the dangerousness of locations based on fatal cycle collisions.

6 References

Scottish Government. (2013). *Cycling Action Plan for Scotland (CAPS)*. http://www.transport.gov.scot/system/files/uploaded_content/documents/tsc_basic_pages/Environment/CAPS_2013_-_final_draft_-_19_June_2013_0.pdf (accessed June 2016)

[i] It is important to note that this algorithm starts from an arbitrary point – different clusters may be identified if the process starts from a different point
[ii] Mid-year population estimates (2014)
[iii] Scottish household survey (2014)
[iv] In the ranking reported in this paper, a flat weighting across the variables was applied
[v] The percentile rank of a value is the percentage of values in its frequency distribution that are equal to or lower than it
[vi] The weights applied to each variable before summing for a final score were: serious collisions (0.1), collision cluster size (0.3), population size (0.3) and proportion who cycle to work (0.3).