RAC Foundation

Mobility • Safety • Economy • Environment

RCIP Police Area Collision Profiles

Area: Humberside

Dr. Craig Smith and Bruce Walton
Agilysis
March 2021

agilysis

The Royal Automobile Club Foundation for Motoring Ltd is a transport policy and research organisation which explores the economic, mobility, safety and environmental issues relating to roads and their users. The Foundation publishes independent and authoritative research with which it promotes informed debate and advocates policy in the interest of the responsible motorist.

RAC Foundation

89–91 Pall Mall

London

SW1Y 5HS

Tel no: 020 7747 3445

www.racfoundation.org

Registered Charity No. 1002705

RAC Foundation

Mobility • Safety • Economy • Environment

RCIP Police Area Collision Profiles

Area: Humberside

Dr. Craig Smith and Bruce Walton

Agilysis

March 2021

About the Authors

Dr. Craig Smith is a mathematician, with an established background in academic research. He studied for both his undergraduate master's degree and doctorate in mathematics at the University of Oxford. Since then, he has endeavoured to use his expertise to carry out robust and innovative analysis and research. As Agilysis' Data Scientist, Craig has extensive experience in handling a wide variety of data, and uses his background as a mathematician to explore the use of machine learning and artificial intelligence in advancing road safety research and unlocking the full potential of data.

Bruce Walton has been working with road safety data since 2002, coming from a background in analytical modelling, database design and IT training across several sectors. Since his appointment as project manager for the multi award-winning MAST Online project, Bruce has become recognised as expert in road casualty data, contributory factor analysis, resident risk, analytical architecture and enforcement data management. Bruce works with many road safety stakeholder organisations in the UK, and provides consultancy and training to international projects on road safety data architecture and reporting, such as the International Road Federation's 'World Road Statistics' programme. Bruce is also a member of the government's Standing Committee for Road Accident Statistics (SCRAS) in the UK.

Disclaimer

Contents

List of Figures

List of Tables

List of Abbreviations

ATS	automatic traffic signal
CF	contributory factor
DfT	Department for Transport
FSC	fatal and serious collision
IMD	Index of Multiple Deprivation
KSI	killed or seriously injured
LCV	light commercial vehicle
LSOA	Lower Layer Super Output Area
ONS	Office for National Statistics
P2W	powered two-wheeler
PSV	public service vehicle
RCIP	Road Collision Investigation Project

1 Introduction

This report forms part of the Road Collision Investigation Project (RCIP). The purpose of RCIP is to establish whether there is a business case for putting more resource into the investigation of road crashes – and, if there is, to establish how best to take this forward. The project, implemented by the RAC Foundation with government funding, began in the summer of 2018.[1]

RCIP's aims include developing an analytical framework and protocols and testing them in real-world environments. To address these aims, the RAC Foundation produced a research brief in February 2020. Agilysis successfully bid to undertake this research for three RCIP areas, with work commencing in April 2020. The project sought to apply deep learning models to road safety data to identify collision trends and types in a way which will provide value to the RCIP project.

This report is part of a series which delivers the results of this research. It contains a synthesis of the most significant findings of analysis carried out on data relating to one of the police force areas participating in RCIP. The intention is to test the validity and value of the methodology in a real-world environment.

An overview of RCIP and further explanation of how this report relates to the project is laid out in the accompanying methodology paper. That paper also contains a detailed description of the methodology used, and lessons learnt from the process.

1.1 Delivery

The research addresses these objectives by delivering four key outputs:

- comparator identification;
- trend analysis;
- collision type analysis; and
- synthesis.

The process used to create each of these components is also described in the methodology paper. This report sets out the results for Humberside. Appendix A includes a summary list of the input variables used, with an indication of how the model applied them when clustering collisions in Humberside.

The comparator identification process, and the process used to arrive at it, is described in the methodology paper. This process identified West Mercia Police as the force most comparable to Humberside.

The trend analysis has been supplied to RCIP investigators primarily by means of online dashboards. The output from this analysis is extensive; this report contains some synthesised key findings for Humberside. The collision type analysis output has been summarised in infographics which are also included in this report.

[1] www.racfoundation.org/collaborations/road-collision-investigation-project

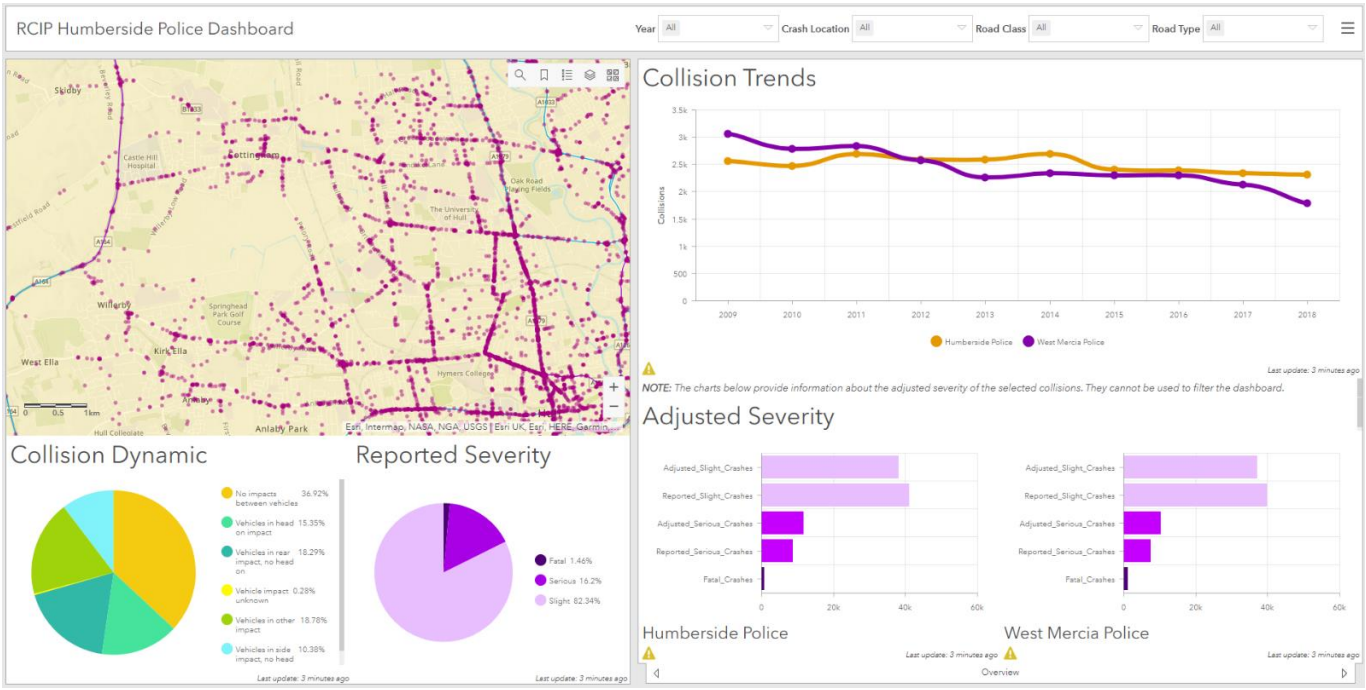
2 Area Profile of Humberside Police

2.1 Dashboards

Primary delivery of the trend analysis results was through a series of dashboards realised in ArcGIS Online. These dashboards are available to RCIP collision investigators via their logins to CrashMap Pro, an ArcGIS application developed by Agilysis. Figure 2.1 shows an example of the dashboard for Humberside, viewing the Collision Trends pane. Collision investigators with access to these credentials can view the dashboard containing results for Humberside Police, as well as the comparator area of West Mercia Police, at the following URL:

https://agilysis.maps.arcgis.com/apps/opsdashboard/index.html#/63c685c910174cf18da28e4545d8af2c

Figure 2.1: Example view of the Humberside dashboard



Source: Author's own

The dashboard provides complete interactive access to detailed analysis of all input variables. It allows investigators to filter collisions by any desired combination of variables and locate specific collisions which exhibit them on a map. The filter dropdown controls on the title bar apply overarching filters, which allow the subject and comparator areas to be examined either separately or together for any desired time period and/or road type. Individual collisions which fit the selected criteria can then be readily identified and examined in more detail if required. This allows investigators to view all reported collisions which exhibit a specific combination of characteristics.

The left pane of the dashboard maps all collisions, including those resulting in only slightly injured casualties. Each collision can be selected individually to show its ID (thereby facilitating further investigation of selected incidents in police records) and salient facts such as the number of casualties involved and types of vehicle conflict present. The map is accompanied by two pie charts showing the reported severity and collision dynamics of currently selected collisions. The map can be filtered using four drop down lists, covering year, location by area, class and type.

The right pane of the dashboard can be scrolled through several different views, most of which include multiple interactive controls used for filtering selected collisions (exceptions are noted in the list which follows, and also indicated on the dashboard itself). These controls are based on, but are more extensive in detail than, the input

variables used in the cluster analysis. For a listing of all input variables, along with how they were used in the Humberside cluster analysis compared to the national analysis described in the methodology paper, see Appendix A.

The panes included in the dashboard, and the controls available on them, are as follows:

- Overview pane (overview charts are for information only, and cannot be used for filtering the map)
    - trends over time of the selected collisions
    - comparative breakdown of recorded severity compared with adjusted severity according to Department for Transport (DfT) record level statistical adjustments which account for discrepancies with injury-based recording systems)
- Location analysis pane
    - traffic where DfT count point data is available, classified as busy (upper quartile), normal, or quiet (lower quartile)
    - road rurality following ONS classifications of rural, town or urban area
    - road class
    - road type
- Times and days
- Pedestrian collisions
    - pedestrian casualties by age and deprivation
    - pedestrian movement
- Actors (profile of involved persons)
    - young and Older drivers
    - vulnerable vehicles (cycles and horses)
    - deprived drivers (lower quartile of home community Index of Multiple Deprivation (IMD), as defined by the ONS)
    - working drivers
    - hit-and-run drivers
- Casualties by road user type and severity
- Attendant circumstances
    - weather and light conditions
    - junction types
- Vehicles involved
    - manoeuvres (overtaking, turning, lane changes and slow traffic)
    - run-offs[2]
    - vehicle type (including motorcycles by size)
- Contributory factor groups, as a Venn diagram (for information only, and cannot be used for filtering the map)
    - environmental factors (100, 700 and 900 series)
    - driver and Vehicle factors (200–600 series)
    - pedestrian (800 series)
- Contributory factor groups, as a bar chart (for information only, and cannot be used for filtering the map)
- Driver contributory factors, by vehicle type
- Driver contributory factors, by manoeuvre type

[2] Run-off-road collisions, referred to as 'run-offs' in this report, are collisions during which any involved vehicle leaves the carriageway , even if it later re-joins it.

3 Results of Collision Type Analysis

The process used to identify clusters of collisions which have characteristics in common is described in detail in the accompanying methodology paper. For Humberside, these groups were arranged subjectively into four overarching groups, then the clusters within each group were organised into families within which sibling clusters could be identified on the basis of the characteristics they shared.

3.1 How to read the cluster diagrams

The clusters in each group are shown by the following diagrams. In each diagram:

- each coloured area shows a family of collisions within the group that have been grouped together based on similar characteristics;
- each of the inner boxes within that family represents sibling or 'Grandsibling' clusters that divide up these shared characteristics down to another level of separation; and
- all collision totals are additive, so percentages are based on the overall total for the entire group (and may not add up to 100% due to rounding).
- The key for the associated meanings represented by each infographic within the diagrams can be found in Appendix B.

3.2 Collisions involving cyclist casualties

3.1.1 Cluster list

Table 3.1 summarises all clusters in this group textually.

Table 3.1: Textual summary of clusters, collisions involving cyclist casualties

Family	Sibling	GrandSibling	Cluster ID	Count	Cluster FSCs as % of group
Cyclist casualty, cyclist contributory factor (CF)			H24	469	10.4%
Cyclist casualty, cyclist CF	Deprived driver, driver observation CF		H23	48	1.1%
Cyclist casualty, cyclist CF	Deprived driver, driver observation CF	Other impact	H15	312	6.9%
Cyclist casualty, cyclist CF	Deprived driver, driver observation CF	Other impact, uncontrolled junction	H17	491	10.9%
Cyclist casualty, other impact			H1	935	20.8%
Cyclist casualty, other impact	Deprived driver, driver observation CF		H18	342	7.6%
Cyclist casualty, other impact	Deprived driver, driver observation CF	Uncontrolled junction, right turn, driver turning CF	H16	420	9.4%
Cyclist casualty, other impact	Driver observation CF, driver turning CF, uncontrolled junction		H21	1,462	32.6%
Cyclist casualty, working driver, LCV/PSV, LCV/PSV CF, weekend			H30	3	0.1%

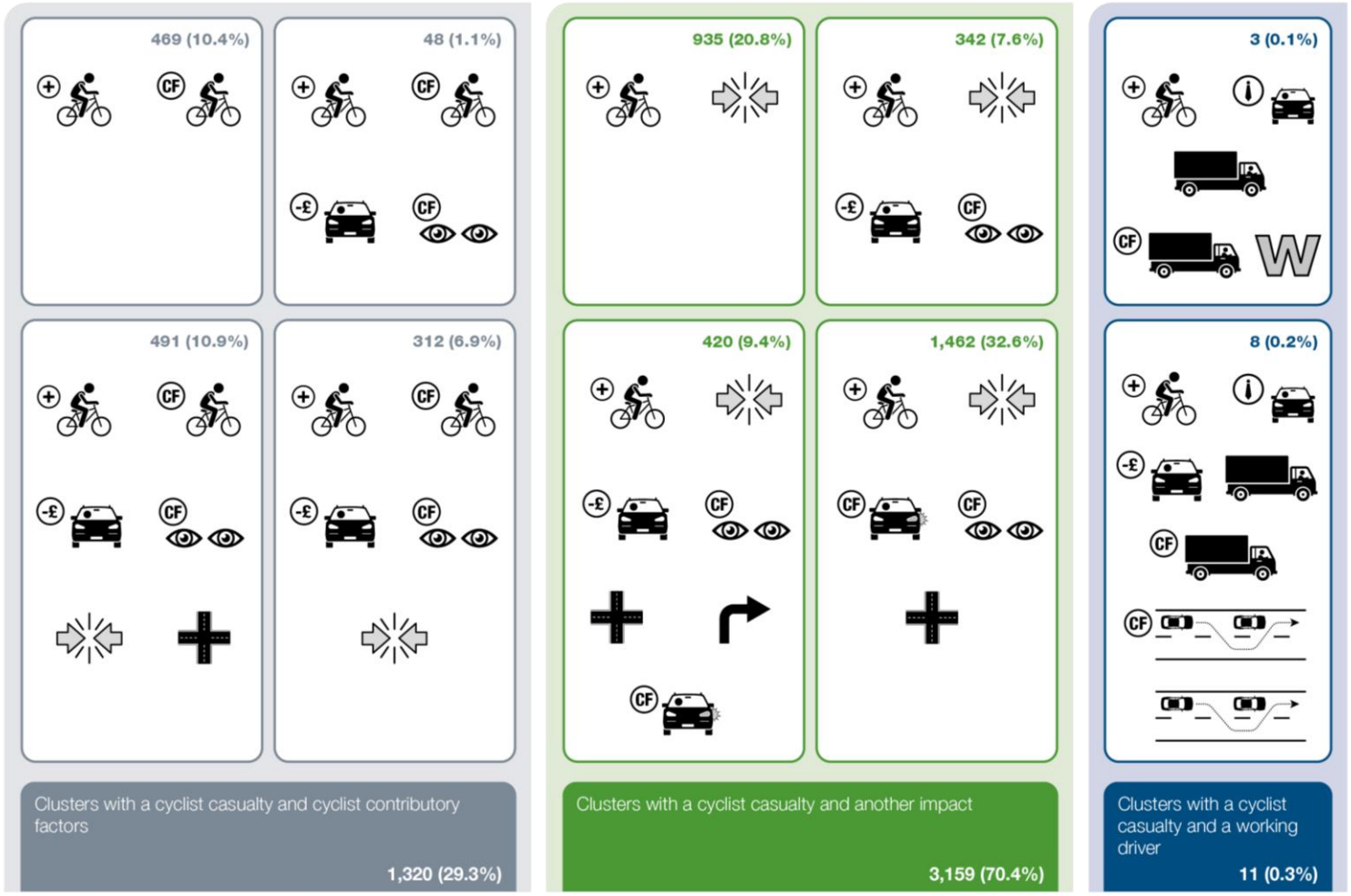
Cyclist casualty, working driver, LCV/PSV, LCV/PSV CF, weekend	Deprived driver, working driver, vehicle overtaking, driver overtaking CF		H27	8	0.2%

Source: Author's own

3.1.2 Cluster infographic

Figure 3.1 summarises all clusters in this group diagrammatically.

Figure 3.1: Cluster family diagram, collisions involving cyclist casualties

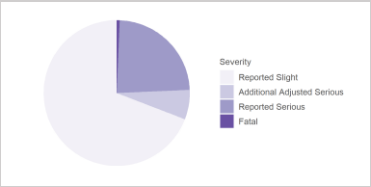
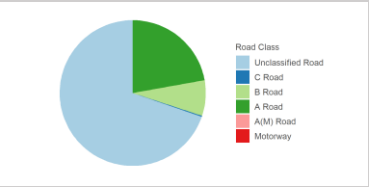
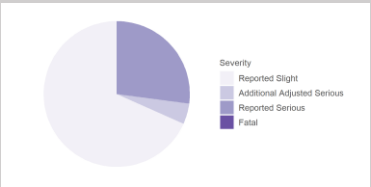
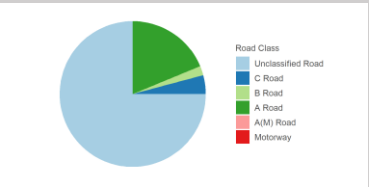
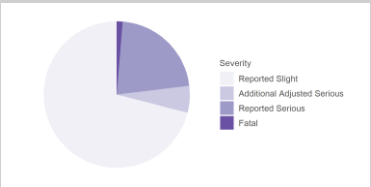
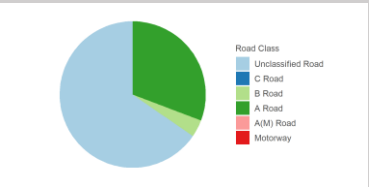
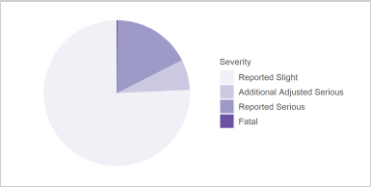
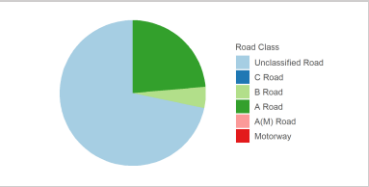
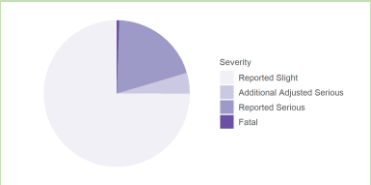
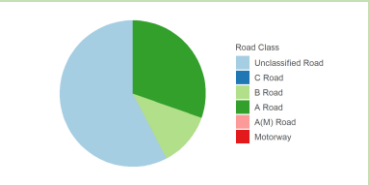


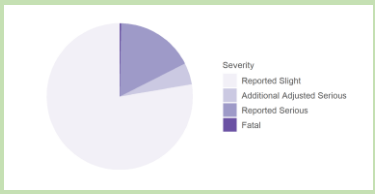
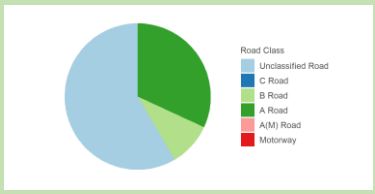
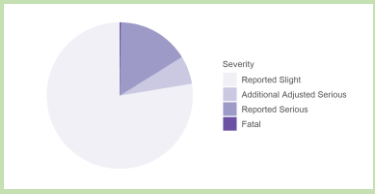
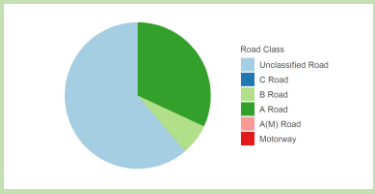
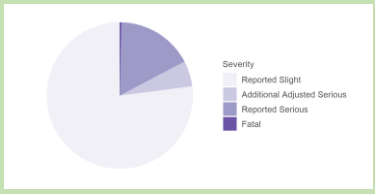
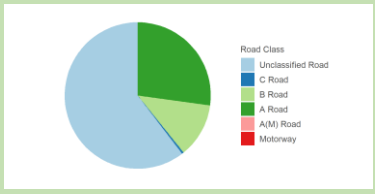
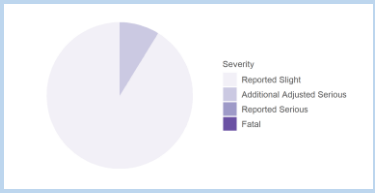
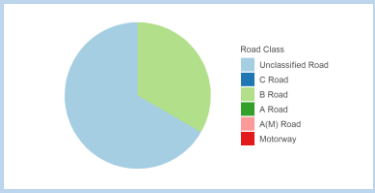
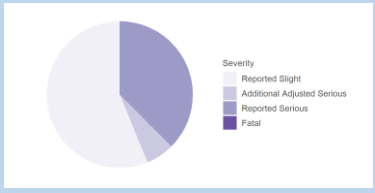
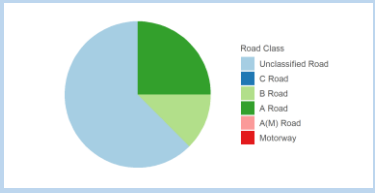
Source: Author's own

3.2.3 Cluster details

Table 3.2 summarises the most salient clusters in this group by severity and road class.

Table 3.2: Summary of clusters by severity and road class, collisions involving cyclist casualties

	Collisions involving cyclist casualties	
H24	**Cyclist casualty, cyclist contributory factor (CF)**	
All collisions		**Percentage adjusted fatal and serious collisions (FSCs)**
469		10.4%
		
H23	**Cyclist casualty, cyclist CF, deprived driver, driver observation CF**	
All collisions		**Percentage adjusted FSCs**
48		1.1%
		
H15	**Cyclist casualty, cyclist CF, deprived driver, driver observation CF, other impact**	
All collisions		**Percentage adjusted FSCs**
312		6.9%
		
H17	**Cyclist casualty, cyclist CF, deprived driver, driver observation CF, other impact, uncontrolled junction**	
All collisions		**Percentage adjusted FSCs**
491		10.9%
		
H1	**Cyclist casualty, other impact**	
All collisions		**Percentage adjusted FSCs**
935		20.8%
		
H18	**Cyclist casualty, deprived driver, driver observation CF, other impact**	
All collisions		**Percentage adjusted FSCs**
342		7.6%

	
H16	**Cyclist casualty, deprived driver, driver observation CF, other impact, uncontrolled junction, right turn, driver turning CF**
All collisions	**Percentage adjusted FSCs**
420	9.4%
	
H21	**Cyclist casualty, driver observation CF, driver turning CF, other impact, uncontrolled junction**
All collisions	**Percentage adjusted FSCs**
1,462	32.6%
	
H30	**Cyclist casualty, working driver, LCV/PSV, LCV/PSV CF, weekend**
All collisions	**Percentage adjusted FSCs**
3	0.1%
	
H27	**Cyclist casualty, deprived driver, working driver, LCV/PSV, LCV/PSV CF, vehicle overtaking, driver overtaking CF**
All collisions	**Percentage adjusted FSCs**
8	0.2%
	

Source: Author's own

3.3 Single-vehicle collisions

3.3.1 Cluster list

Table 3.3 summarises all clusters in this group textually.

Table 3.3: Textual summary of clusters, single-vehicle collisions

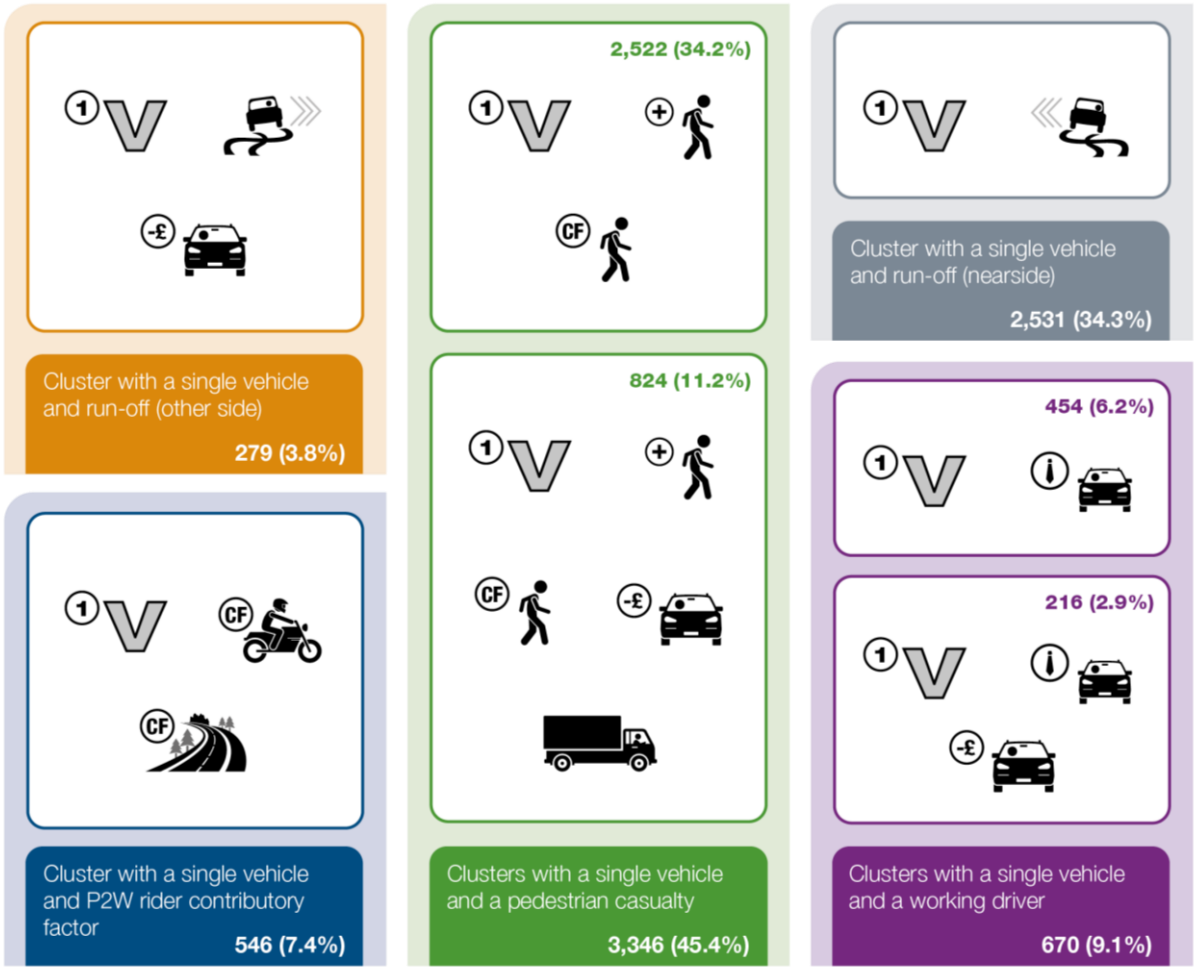
Family	Sibling	GrandSibling	Cluster ID	Count	Cluster FSCs as % of group
Single vehicle, runoff (other), deprived driver			H25	279	3.8%
Single vehicle, P2W (powered two-wheeler) rider contributory factor (CF), environment CF			H9	546	7.4%
Single vehicle, pedestrian casualty, pedestrian CF			H12	2,522	34.2%
Single vehicle, pedestrian casualty, pedestrian CF	Deprived driver, LCV/PSV		H14	824	11.2%
Single vehicle, runoff (nearside)			H4	2,531	34.3%
Single vehicle, working driver			H20	454	6.2%
Single vehicle, working driver	Deprived driver		H13	216	2.9%

Source: Author's own

3.3.2Cluster infographic

Figure 3.2 summarises all clusters in this group diagrammatically.

Figure 3.2: Cluster family diagram, single-vehicle collisions

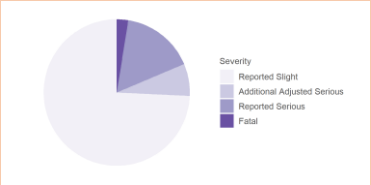
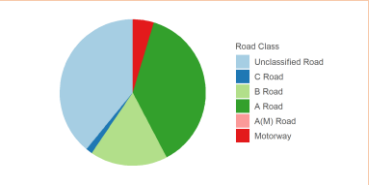
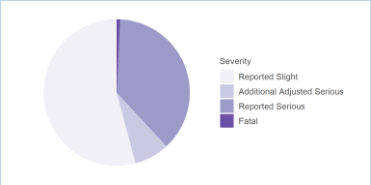
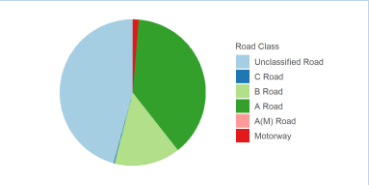
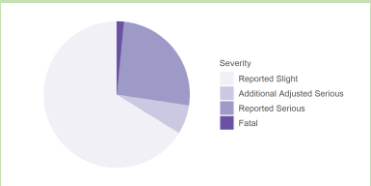
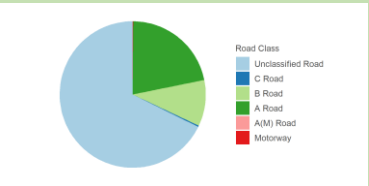
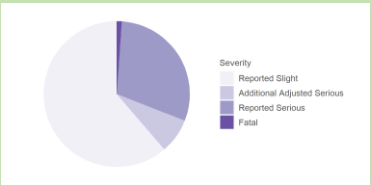
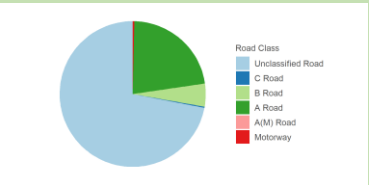
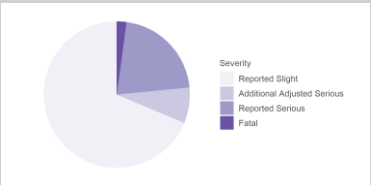
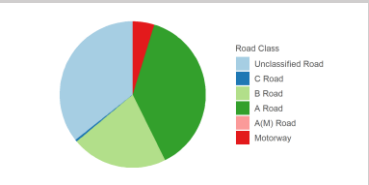


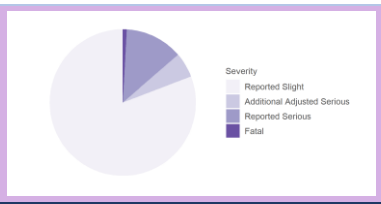
Source: Author's own

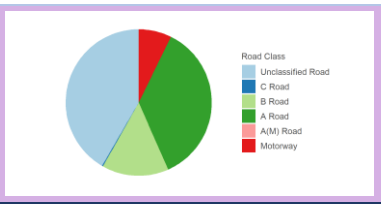
3.3.3 Cluster details

Table 3.4 summarises the most salient clusters in this group by severity and road class.

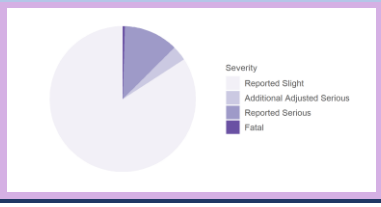
Table 3.4: Summary of clusters by severity and road class, single-vehicle collisions

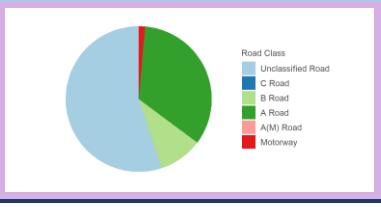
	Single-vehicle collisions	
H25	**Single vehicle, runoff (other), deprived driver**	
All collisions		**Percentage adjusted FSCs**
279		3.8%
		
H9	**Single vehicle, P2W (powered two-wheeler) rider contributory factor (CF), environment CF**	
All collisions		**Percentage adjusted FSCs**
546		7.4%
		
H12	**Single vehicle, pedestrian casualty, pedestrian CF**	
All collisions		**Percentage adjusted FSCs**
2,522		34.2%
		
H14	**Single vehicle, pedestrian casualty, pedestrian CF, deprived driver, LCV/PSV**	
All collisions		**Percentage adjusted FSCs**
824		11.2%
		
H4	**Single vehicle, runoff (nearside)**	
All collisions		**Percentage adjusted fatal and serious collisions (FSCs)**
2,531		34.3%
		
H20	**Single vehicle, working driver**	
All collisions		**Percentage adjusted FSCs**
454		6.2%





H13	Single vehicle, working driver, deprived driver
All collisions	**Percentage adjusted FSCs**
216	2.9%





Source: Author's own

3.4 Collisions involving working drivers

3.4.1 Cluster list

Table 3.5 summarises all clusters in this group textually.

Table 3.5: Textual summary of clusters, collisions involving working drivers

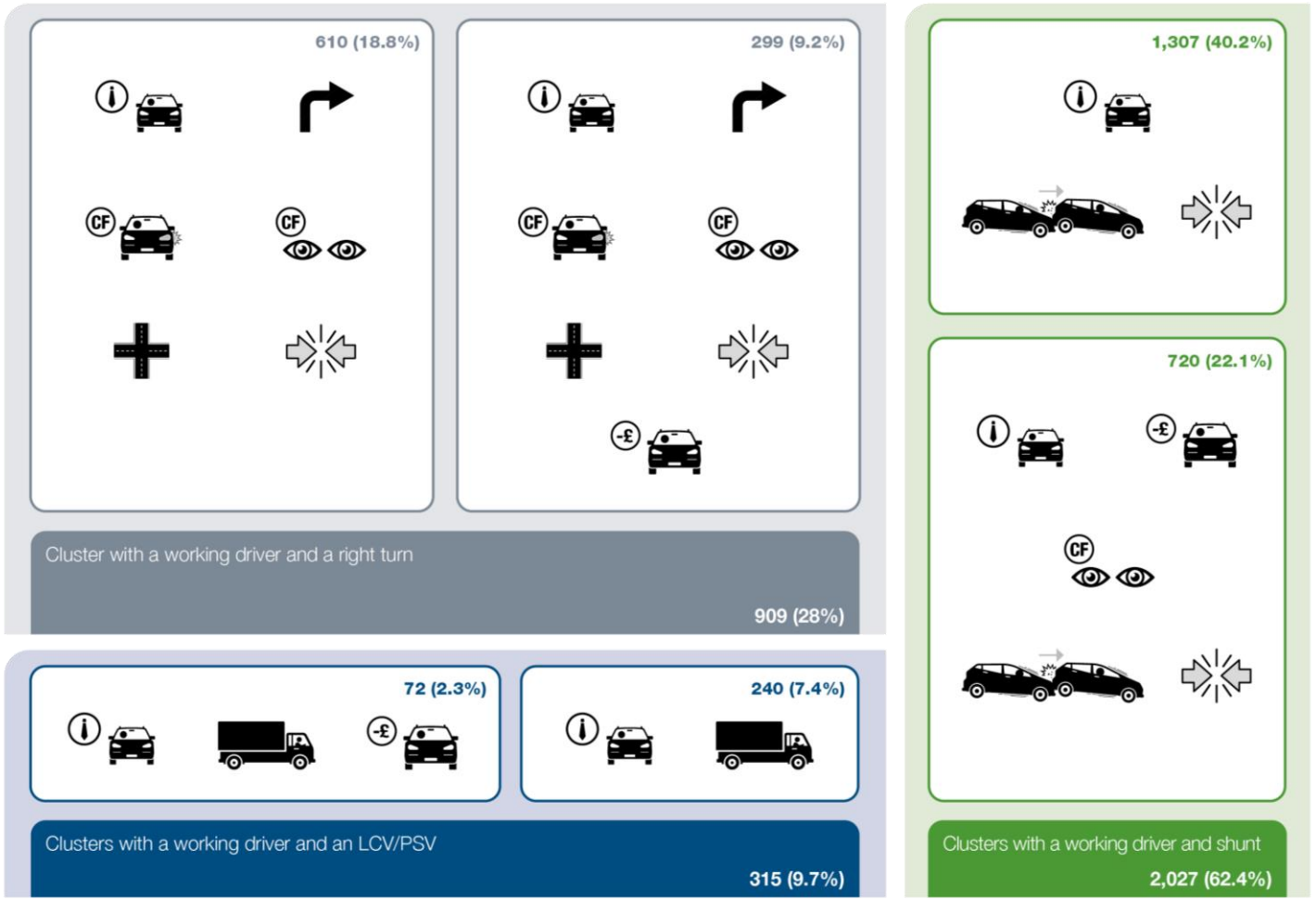
Family	Sibling	GrandSibling	Cluster ID	Count	Cluster FSCs as % of group
Working driver, right turn, driver turning contributory factor (CF), driver observation CF, uncontrolled junction, other impact			H22	610	18.8%
Working driver, right turn, driver turning CF, driver observation CF, uncontrolled junction, other impact	**Deprived driver**		H7	299	9.2%
Working driver, LCV/PSV			H26	240	7.4%
Working driver, LCV/PSV	Deprived driver		H28	75	2.3%
Working driver, shunt, other impact			H2	1,307	40.2%
Working driver, shunt, other impact	Deprived driver, driver observation CF		H11	720	22.1%

Source: Author's own

3.4.2 Cluster infographic

Figure 3.3 summarises all clusters in this group diagrammatically.

Figure 3.3: Cluster family diagram, collisions involving working drivers

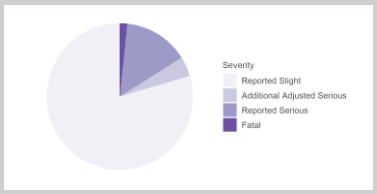
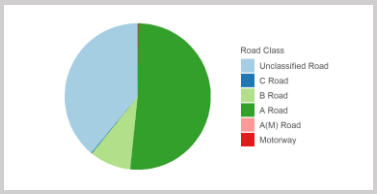
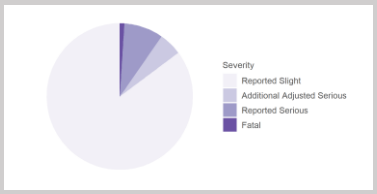
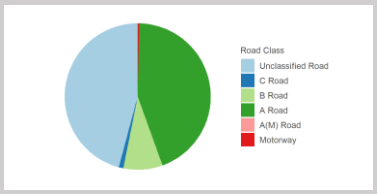
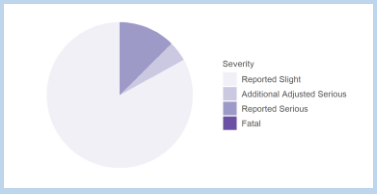
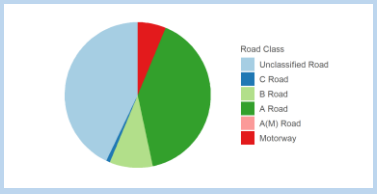
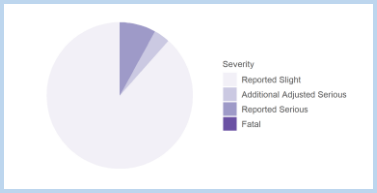
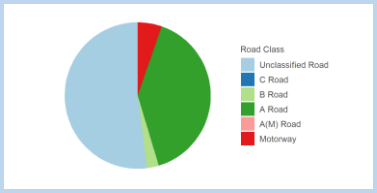
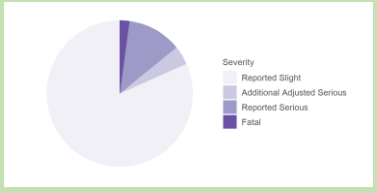
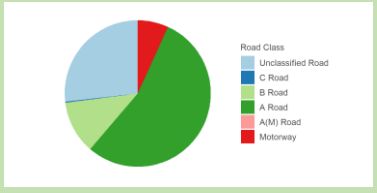


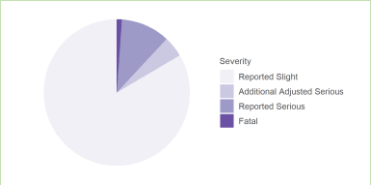
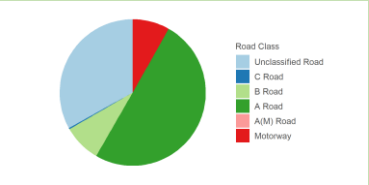
Source: Author's own

3.4.3 Cluster details

Table 3.6 summarises the most salient clusters in this group by severity and road class.

Table 3.6: Summary of clusters by severity and road class, collisions involving working drivers

	Collisions involving working drivers	
H22	**Working driver, right turn, driver turning contributory factor (CF), driver observation CF, uncontrolled junction, other impact**	
All collisions		**Percentage adjusted fatal and serious collisions (FSCs)**
610		18.8%
		
H7	**Working driver, deprived driver, right turn, driver turning CF, driver observation CF, uncontrolled junction, other impact**	
All collisions		**Percentage adjusted FSCs**
299		9.2%
		
H26	**Working driver, LCV/PSV**	
All collisions		**Percentage adjusted FSCs**
240		7.4%
		
H28	**Working driver, deprived driver, LCV/PSV**	
All collisions		**Percentage adjusted FSCs**
75		2.3%
		
H2	**Working driver, shunt, other impact**	
All collisions		**Percentage adjusted FSCs**
1,307		40.2%
		
H11	**Working driver, deprived driver, driver observation CF, shunt, other impact**	
All collisions		**Percentage adjusted FSCs**

720	22.1%
	

Source: Author's own

3.5 Other collisions

3.5.1 Cluster list

Table 3.7 summarises all clusters in this group textually.

Table 3.7: Textual summary of clusters, all other collisions

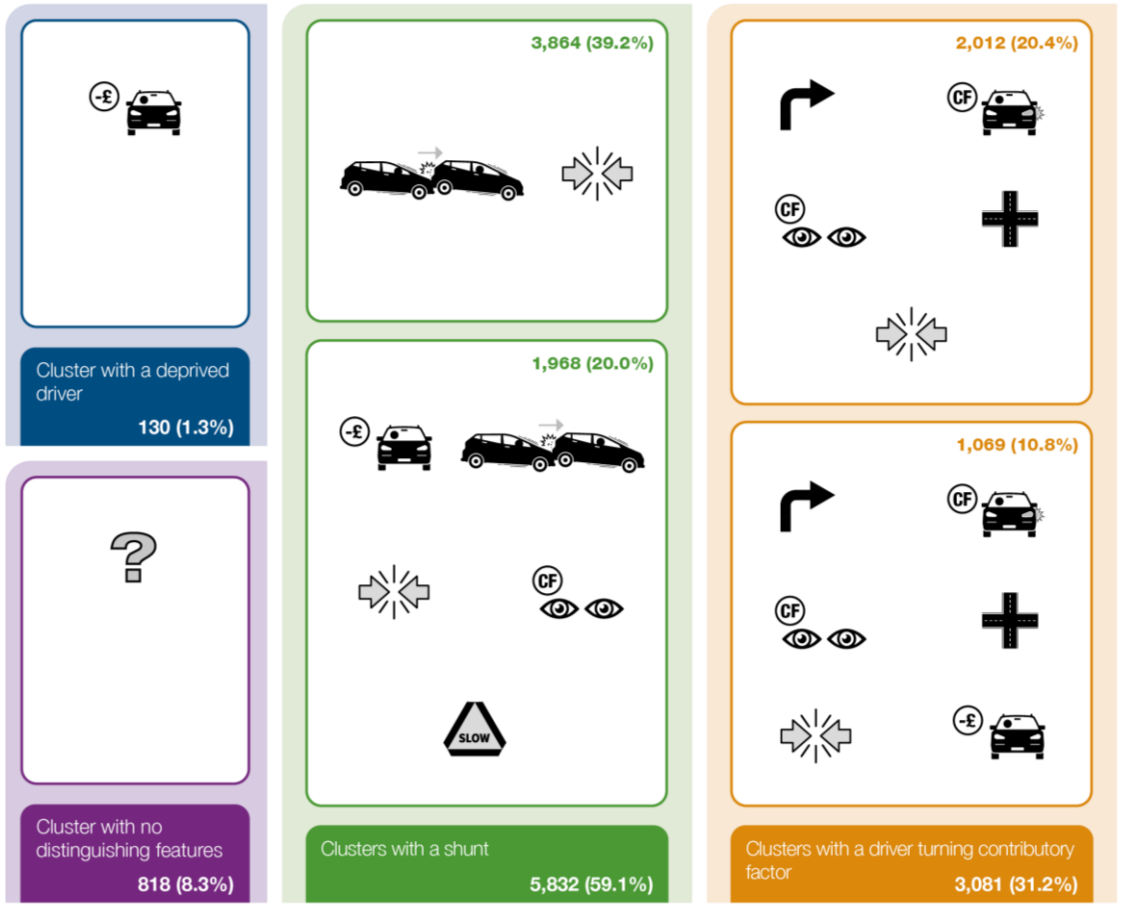
Family	Sibling	GrandSibling	Cluster ID	Count	Cluster FSCs as % of group
Deprived driver, not clustered further			H19	130	1.3%
No distinguishing features			H8	818	8.3%
Shunt, other impact			H3	3,864	39.2%
Shunt, other impact	Deprived driver, slow-vehicle manoeuvre, driver observation contributory factor (CF)		H10	1,968	20.0%
Right turn, driver turning CF, driver observation CF, uncontrolled junction, other impact			H6	2,012	20.4%
Right turn, driver turning CF, driver observation CF, uncontrolled junction, other impact	Deprived driver		H5	1,069	10.8%

Source: Author's own

3.5.2 Cluster infographic

Figure 3.4 summarises all clusters in this group diagrammatically.

Figure 3.4: Cluster family diagram, all other collisions

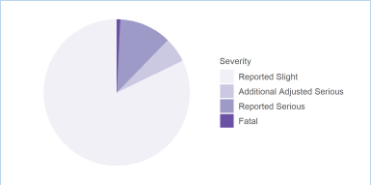
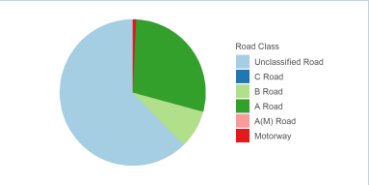
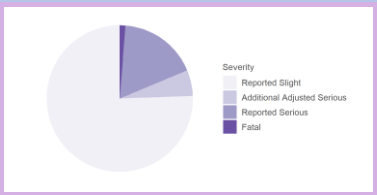
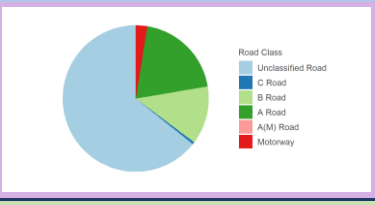
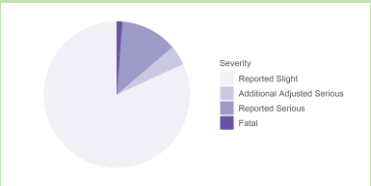
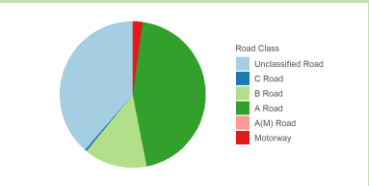
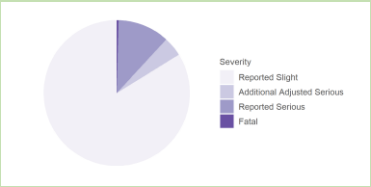
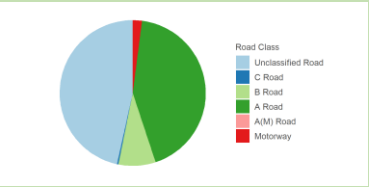
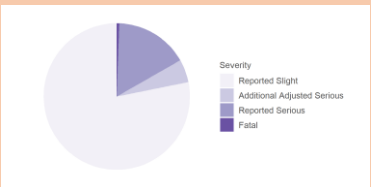
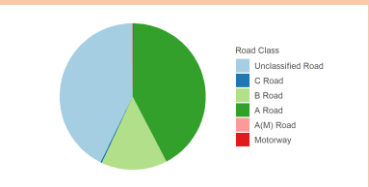


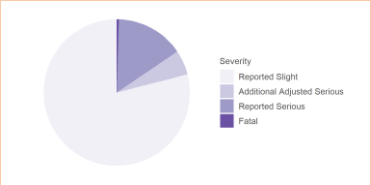
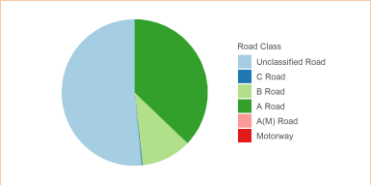
Source: Author's own

3.5.3 Cluster details

Table 3.8 summarises the most salient clusters in this group by severity and road class.

Table 3.8: Summary of clusters by severity and road class, all other collisions

	Other collisions
H19	**Deprived driver**
All collisions	**Percentage adjusted fatal and serious collisions (FSCs)**
130	1.3%
	
H8	**No distinguishing features**
All collisions	**Percentage adjusted FSCs**
818	8.3%
	
H3	**Shunt, other impact**
All collisions	**Percentage adjusted FSCs**
3,864	39.2%
	
H10	**Deprived driver, slow-vehicle manoeuvre, driver observation contributory factor (CF), shunt, other impact**
All collisions	**Percentage adjusted FSCs**
1,968	20.0%
	
H6	**Right turn, driver turning CF, driver observation CF, uncontrolled junction, other impact**
All collisions	**Percentage adjusted FSCs**
2,012	20.4%
	
H5	**Deprived driver, right turn, driver turning CF, driver observation CF, uncontrolled junction, other impact**
All collisions	**Percentage adjusted FSCs**

1,069	10.8%
	

Source: Author's own

4 Synthesis of Key Findings From Collision Type Analysis

Since the full analytic output is available to investigators interactively, this report will concentrate on identifying and summarising key findings. Informed by the main clusters of collisions which have been identified by the collision type analysis, a detailed analysis of all input variables for Humberside was undertaken. The objective was to identify key commonalities and key differences between collision patterns in Humberside and its identified comparator area of West Mercia, while also considering the national context.

During this process, in view of the importance placed on collisions resulting in fatalities or very serious injury in the work of RCIP, this analysis considered only collisions involving killed and/or seriously injured casualties. However, all collisions, including those resulting in only slightly injured casualties, are reported in the dashboard. This allows investigators to view all reported collisions which exhibit a specific combination of characteristics.

In addition, the analysis applied DfT severity adjustment data[3] at individual casualty level where appropriate, to ensure that recent changes in data collection procedures did not distort the results or the validity of comparisons. Consequently, in police force areas which have not yet applied injury-based reporting practices, some casualties originally reported as slight have been split probabilistically between the serious and slight injury categories.

4.1 General observations

In general, the clusters produced by the model for Humberside were both less numerous and had fewer related sibling clusters than both the comparator area of West Mercia and also the other RCIP areas. This is probably in part due to the smaller sample of collisions in the area in comparison to others, although it may also reflect a greater degree of homogeneity in the characteristics of the collisions.

One notable feature of the Humberside clusters was the frequency with which deprived drivers were identified as an associated demographic. To some extent this is likely to be an artefact of the nature of the area, as an unusually high proportion of collision-involved drivers in Humberside generally are in the most deprived quartile of the population. The fact that deprived drivers were not present across the board in all clusters is an indication that there may be meaningful correlation with particular behaviours and collision types, but more detailed further analysis than is possible here would be required to confirm this. However, the frequent identification of working drivers is more likely to represent a genuine local trend.

The pattern of input variables relating to assigned contributory factors was particularly interesting in Humberside. Because reporting practices differ between police force areas, it is hard to be certain that comparative analyses between areas are not being improperly equated. However, the pattern of cyclist contributory factors appears to represent a genuine local trend, and those allocated to pedestrians (mentioned in section 4.3) also constitute a notable feature of the Humberside clusters. The allocation of contributory factors to drivers does, on the other hand, appear to be more consistent with how these were treated by regional and national models.

4.2 Collisions involving cyclist casualties

The most striking outcome from the cluster analysis of collisions resulting in cyclist casualties in Humberside was a clear division into two cluster families: those where contributory factor(s) were assigned to a cyclist by attending officers, and those where no factor was assigned to cyclists. This clear dichotomy is not present in the national cluster model, where allocations of contributory factors to cyclists did not feature. However, a similar trend is clearly present in cluster siblings in West Mercia, albeit not in families.

[3] Severity adjustments are explained in https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/743845/severity-reporting-methodology-report.odt and available for download from https://data.gov.uk/dataset/cb7ae6f0-4be6-4935-9277-47e5ce24a11f/road-safety-data

In both force areas, clusters where turning and/or observation factors were applied to drivers, but none were applied to cyclists, contained a much higher proportion of collisions than clusters where the cyclist contributed but the driver did not. Nationally a similar proportion of clusters included allocation of observation factors to drivers, although turning factors and cyclist factors did not feature. It appears that networks like those in both Humberside and West Mercia networks tend to experience a modest proportion of collisions where the cyclist has contributed, and a much larger share where the cyclist was not at fault but motorists have erred. In both areas, such errors often involved poor right turns at uncontrolled junctions.

A unique feature of these cases in Humberside, where drivers had contributory factors assigned to them, consisted in clusters where a driver from a deprived community was involved. Over-representation of deprived drivers is not in itself surprising, since it reflects the demography of the area: in general, deprived drivers are involved in collisions in Humberside at nearly double the frequency that they are in West Mercia. However, the strong association made by the model between deprivation and observation errors in collisions with cyclists may indicate the recurrence of specific behaviours. The Humberside model did not identify clusters, found both in West Mercia and nationally, which featured variables such as night, slow traffic, and side impacts. This may mean that the association with deprived drivers was stronger than other features such as these.

4.3 Single-vehicle collisions

As is the case both nationally and in West Mercia, clusters of single-vehicle collisions in Humberside frequently feature run-offs. However, there are some notable differences in other features of these clusters in Humberside. The most striking is a large cluster family involving pedestrian casualties.

Pedestrians did not feature as a distinct cluster group in Humberside because they were entirely concentrated in this one sibling family, which was not the case in West Mercia or nationally. This family featured widespread assignment of contributory factors to pedestrian behaviour; this correlation was more ubiquitous in Humberside than is the case either nationally or in West Mercia. A notable sibling cluster within this family highlighted involvement of large vehicles (heavy goods vehicles or buses) and/or deprived drivers. This association does not occur either nationally or in West Mercia. This analysis suggests that pedestrian collisions have a particular common character in Humberside which is less pronounced elsewhere: namely that pedestrians, who have often behaved unwisely and/or are from deprived backgrounds, conflict with vehicles which are often heavy.

One single-vehicle collision cluster in Humberside involved powered two-wheeler riders. A similar cluster is present in West Mercia, although not nationally, suggesting a network characteristic may be responsible. In both areas contributory factors related to the road environment were prominent features in these collisions.

The cluster family of single-vehicle collisions involving working drivers present in Humberside is also present in the national clusters where it forms a larger proportion of single-vehicle collisions. A corresponding cluster was not present in West Mercia.

4.4 Collisions involving working drivers

The cluster group of collisions in Humberside which involve working drivers shows little similarity with the comparable group in West Mercia, nor with a comparable family in the national clusters. Most notably the largest cluster group, distinguished by working drivers involved in front-to-rear ‘shunt’ impacts, is unique to Humberside; it may reflect some characteristics of parts of the local network predominantly used by working drivers. One cluster within that group notably adds observation contributory factors to the profile.

However, two other smaller but more specific groups, involving right turns at an uncontrolled junction where attending officers often assigned both observation and turning manoeuvre contributory factors to involved drivers, are similar to clusters present in the output both nationally and in West Mercia. The circumstances which precipitate these events do not appear to represent an unusual local trend.

The smaller collision cluster representing working drivers of large vehicles in Humberside is comparable to one identified in West Mercia, but in the latter case head-on impacts and overtaking manoeuvres were often present. This feature was not present in Humberside, reinforcing the impression from the large 'shunt' clusters that Humberside working driver collisions less frequently occur in free-flowing traffic than they do elsewhere.

4.5 Other collision clusters

Two other cluster families of collisions in Humberside of notable size were identified by the model but have not been segregated into groups, as they were not further subdivided into multiple siblings. Both have considerable similarities to clusters which are present in both national and the West Mercia outputs. They can be interpreted as generalised versions of cluster families referred to in the working driver group, meaning that these characteristic collision types occur in association with both working and non-working drivers. For both these collision profiles, the West Mercia model was able to identify considerably more detailed cluster groups.

One cluster family involves right turns at uncontrolled junctions with both driver observation and turning contributory factors assigned by attending officers. It is notable that the family contains a specific cluster of deprived drivers. The other is a family of shunt collision clusters, with one specific cluster relating to slow-moving vehicles, involvement of a deprived driver and the assignment of observation contributory factors in a single cluster. It is hard to ascertain whether the recurrence of deprived drivers relates solely to the demography of Humberside in general, or whether it to some extent relates to a prevalence of deprived communities in parts of the network where such collisions more commonly occur.

Appendix A: Input Variables

For an explanation of how these input variables were applied during machine learning, see section 2.2 Input data in the accompanying methodology paper.

Table A.1: Collision input variables

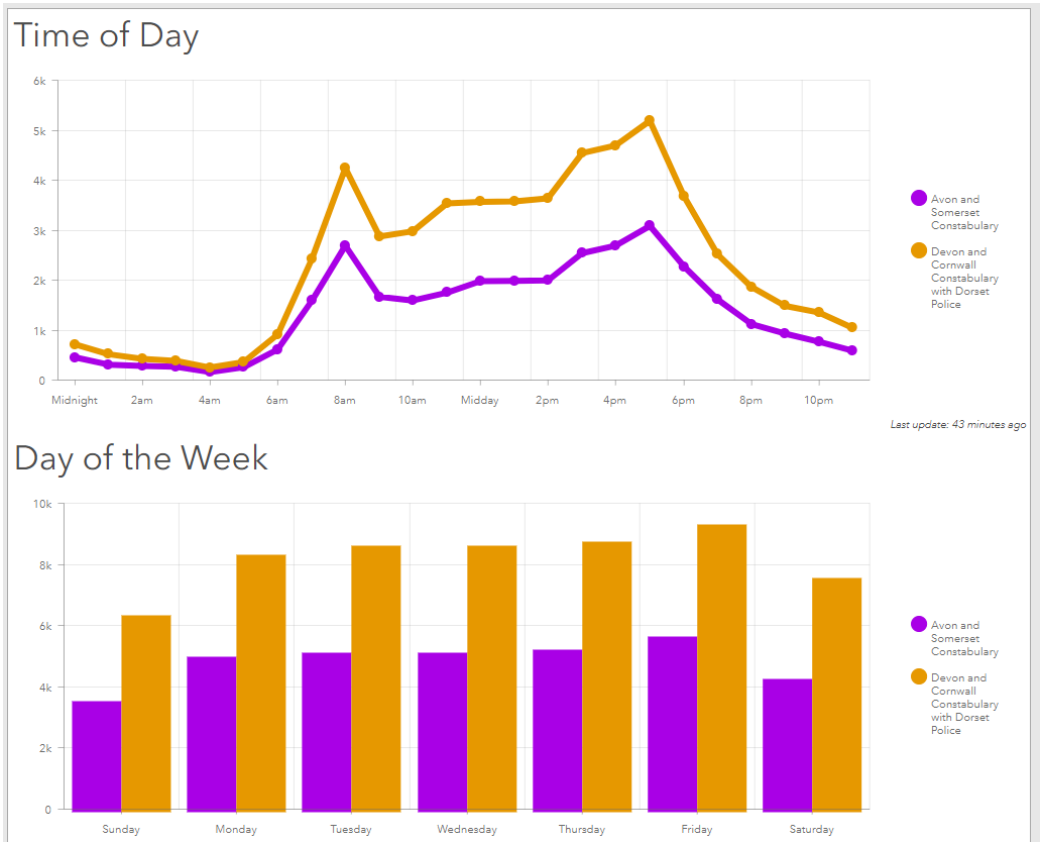
Group	Title	Type	Definition	Model usage
101a	Severity_Fatal	Boolean	True: at least one casualty was killed	Used subtly
101b	Severity_Serious_Adjusted	Continuous	Probability that at least one casualty would have been classified as serious if injury-based reporting had been in place	Used subtly
102a	Junction_Controlled	Boolean	True: junction with ATS (automatic traffic signal) or authorised person	Used subtly
102b	Junction_Uncontrolled_Roundabout	Boolean	True: junction with roundabout or mini-roundabout	Used subtly
102c	Junction_Uncontrolled_Other	Boolean	True: junction with Give Way or Stop (not at roundabout)	Used extensively
103	Weather_Adverse	Boolean	True: any inclement weather conditions (rain, snow, fog, other with or without high winds)	Used subtly
104a	Date_PH	Boolean	True: was a weekday public holiday (Christmas, Easter or bank holiday)	Ignored as irrelevant
104b	Date_Weekend	Boolean	True: was a Saturday or Sunday	Used subtly in Humberside, but ignored as irrelevant nationally
105a	Time_Rush_AM_7to9	Boolean	True: was at or after 7 a.m. and before 9 a.m.	Used subtly in Humberside, but used moderately nationally
105b	Time_Night_7to7	Boolean	True: was at or after 7 p.m. and before 7 a.m. the following day	Used extensively in Humberside, but used moderately nationally
106a	Night_Streetlights	Boolean	True: was dark, and streetlights were present and lit	Used extensively
106b	Night_NoStreetlights	Boolean	True: was dark, and no lit streetlights were present	Used subtly in Humberside, but used moderately nationally
107	Vehicles_Single	Boolean	True: only one vehicle was involved	Used extensively
108	Population_Density_Raw	Continuous	Population per square km of Lower Layer Super Output Area (LSOA) / data zone in mid-2018	Ignored as irrelevant
109	Dynamics_HeadOn	Boolean	True: at least one vehicle had a front impact; *and* at least one other vehicle travelling in the opposite direction also had an impact	Used moderately

110	Dynamics_Shunt	Boolean	True: at least one vehicle had a rear impact; *and* at least one other vehicle travelling in the same direction also had an impact	Used extensively
111	Dynamics_SideImpact	Boolean	True: at least one vehicle had a side impact; *and* at least one other vehicle travelling in an adjacent direction also had an impact	Used moderately
112	Dynamics_OtherImpact	Boolean	True: at least two vehicles had impacts	Used extensively
113	Vehicles_Count	Continuous	Number of vehicles involved	Used subtly
114	Casualties_Count	Continuous	Number of casualties resulting (of all severities)	Used subtly

Source: Author's own

Figure A.1 shows some collision variables applied to trend analysis in an area dashboard.

Figure A.1: Collision variables applied to trend analysis in an area dashboard



Source: Author's own

Table A.2: Vehicle input variables

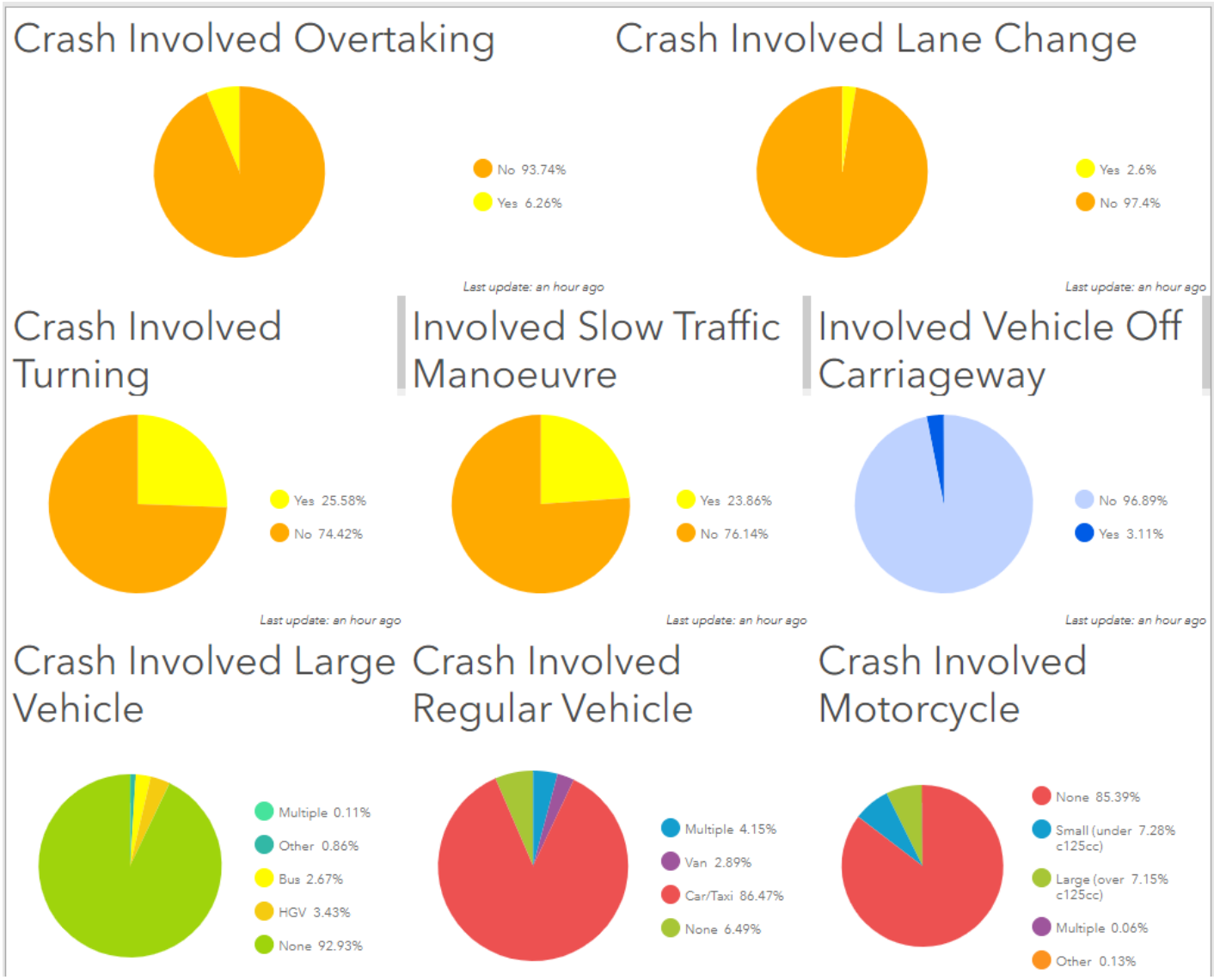
Group	Title	Type	Definition	Model usage
201a	Runoff_Nearside	Boolean	True: vehicle left carriageway to the nearside (whether rebounded or not)	Used extensively
201b	Runoff_Other	Boolean	True: vehicle left carriageway in any other fashion	Used moderately
202	Vehicle_HitRun	Boolean	True: vehicle was hit-and-run (excluding non-stop vehicles not hit)	Used subtly
203	Vehicle_NotInMainCway	Boolean	True: any vehicle on a footway; any vehicle on, entering or leaving a hard shoulder; a vehicle other than a bus in a bus lane or busway; or any vehicle other than a tram on a tram track	Ignored as irrelevant
204a	Vehicle_Overtaking	Boolean	True: vehicle was overtaking (offside or nearside)	Used subtly
204b	Vehicle_LeftTurn	Boolean	True: vehicle was turning left, or waiting to do so	Used moderately
204c	Vehicle_RightTurn	Boolean	True: vehicle was turning right, or waiting to do so	Used extensively

204d	Vehicle_SlowManeouvre	Boolean	True: vehicle was stopping, stationary or moving off	Used moderately in Humberside, but used extensively nationally
204e	Vehicle_LaneChange	Boolean	True: vehicle was changing lane (to left or right)	Used subtly
205a	Vehicle_Moped	Boolean	True: vehicle was a motorcycle with engine size 50cc or under	Used subtly in Humberside, but ignored as irrelevant nationally
205b	Vehicle_MC_MidSize	Boolean	True: vehicle was a motorcycle with engine size over 50cc up to 500cc (includes vehicles which were electric or of unknown engine size)	Used subtly
205c	Vehicle_MC_Large	Boolean	True: vehicle was a motorcycle with engine size over 500cc	Used subtly
205d	Vehicle_Large_GV_PSV	Boolean	True: vehicle was a bus, coach or tram; or a goods vehicle over 3.5 tonnes mgw or of unknown weight	Used moderately in Humberside, but used extensively nationally
206a	Driver_Young_Under25	Boolean	True: driver/rider of motor vehicle was aged 16–24 inclusive	Used extensively in Humberside, but used moderately nationally
206b	Driver_Old_70Plus	Boolean	True: driver/rider of motor vehicle was aged over 69	Used extensively
207	Driver_Deprived_BottomQuintile	Boolean	True: driver's home postcode was in a LSOA classified by the ONS in the most deprived quintile of the Index of Multiple Deprivation	Used moderately
208	Driver_Working	Boolean	True: driver was recorded as working; and/or was driving a large vehicle; and/or was on a commuting journey in a taxi or light goods vehicle	Used extensively

Source: Author's own

Figure A.2 shows some of these vehicle variables applied to trend analysis in an area dashboard.

Figure A.2: Vehicle variables applied to trend analysis in an area dashboard



Source: Author's own

Table A.3: Casualty input variables

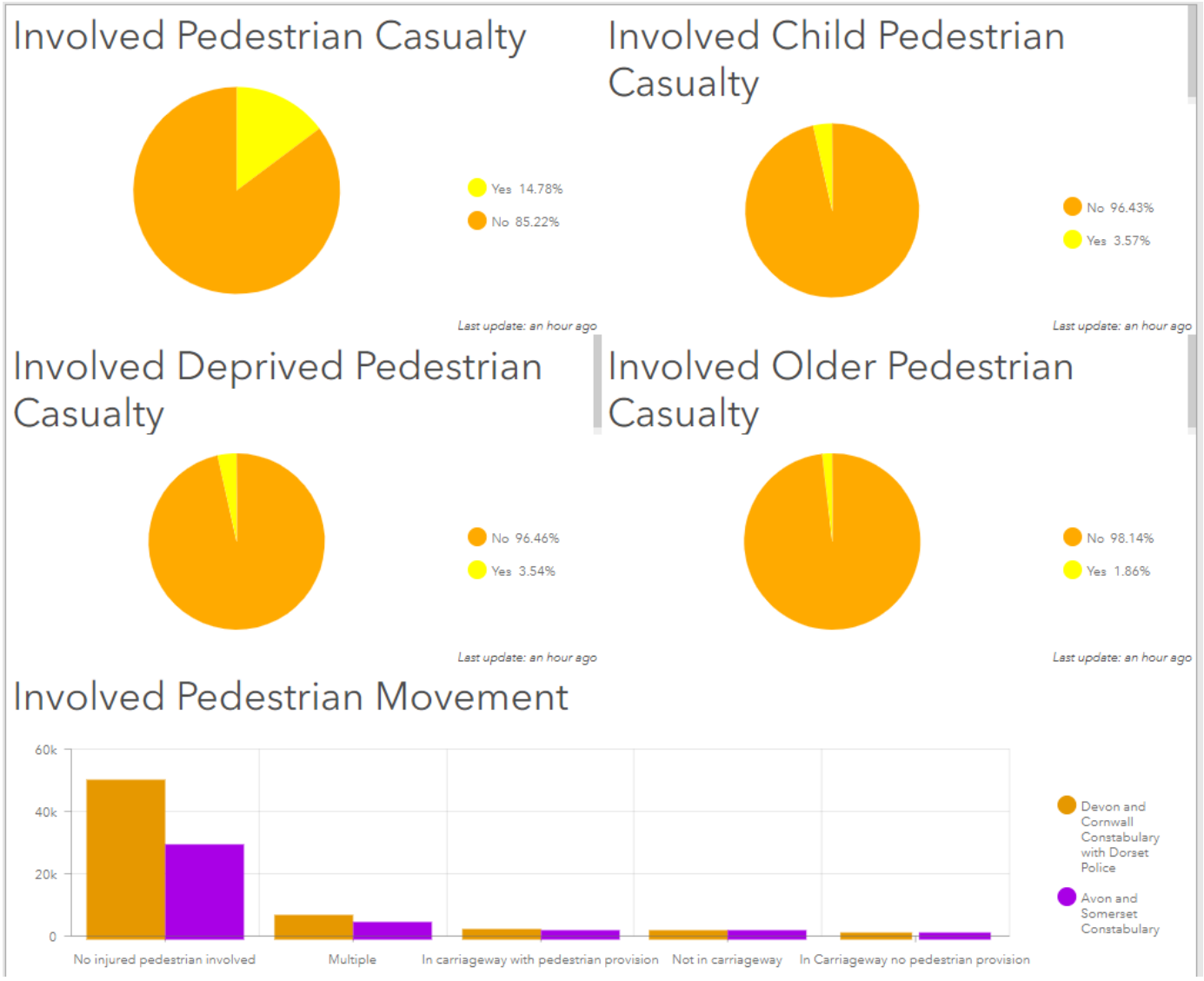
Group	Title	Type	Definition	Model usage
301a	Casualty_PCUser	Boolean	True: casualty was rider or pillion passenger on a cycle	Used extensively
301b	Casualty_HorseRider	Boolean	True: casualty was rider or pillion passenger on a horse	Ignored as irrelevant
301c	Casualty_MobilityScooterUser	Boolean	True: casualty was rider or pillion passenger on a mobility scooter	Ignored as irrelevant
302	Casualty_Pedestrian	Boolean	True: casualty was a pedestrian	Used extensively
303a	Casualty_ChildPedestrian_Under16	Boolean	True: casualty was a pedestrian aged under 16	Used moderately in Humberside, but used extensively nationally

303b	Casualty_OldPedestrian_70Plus	Boolean	True: casualty was a pedestrian aged over 69	Used moderately
304a	Casualty_Pedestrian_CrossingOrRefuge	Boolean	True: casualty was a pedestrian on a crossing, refuge or central island	Used moderately
304b	Casualty_Pedestrian_Footway	Boolean	True: casualty was a pedestrian on a footway	Used moderately in Humberside, but used subtly nationally
305	Casualty_Pedestrian_InCway_Masked	Boolean	True: casualty was a pedestrian anywhere in the carriageway who was masked by a stationary or parked vehicle	Used moderately in Humberside, but used subtly nationally

Source: Author's own

Figure A.3 shows some casualty variables applied to trend analysis in an area dashboard.

Figure A.3: Casualty variables applied to trend analysis in an area dashboard



Source: Author's own

Table A.4: Contributory factor input variables

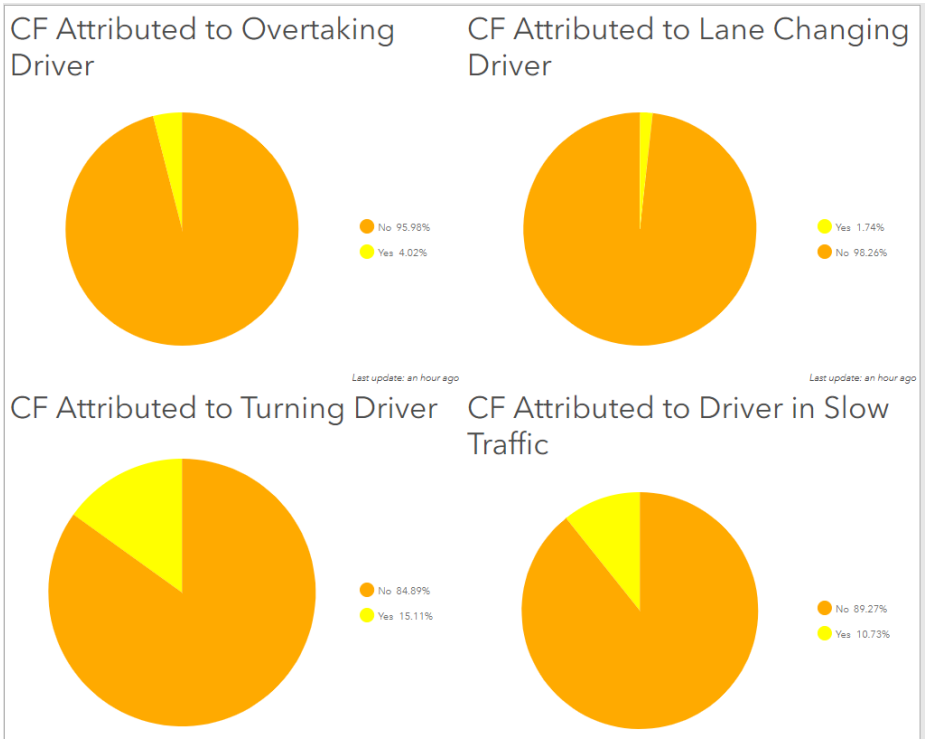
Group	Title	Type	Definition	Model usage
401a	Pedestrian_Casualty_Contributed	Boolean	True: any injured pedestrian or vehicle passenger had a pedestrian contributory factor (CF) assigned to them	Used extensively
401b	Pedestrian_Uninjured_Contributed	Boolean	True: any uninjured pedestrian had a pedestrian CF assigned to them	Ignored as irrelevant
402a	Driver_Contributed_Overtaking	Boolean	True: any overtaking driver or rider had any driver/rider CF assigned to them	Used subtly
402b	Driver_Contributed_Turning	Boolean	True: any turning driver or rider had any driver/rider CF assigned to them	Used extensively
402c	Driver_Contributed_LaneChange	Boolean	True: any lane-changing driver or rider had any driver/rider CF assigned to them	Used subtly
403a	Cyclist_Contributed	Boolean	True: any cyclist had any CF assigned to them	Used moderately in Humberside, but used extensively nationally
403b	P2W_Rider_Contributed	Boolean	True: any motorcyclist had any CF assigned to them	Used subtly
403c	Large_GV_PSV_Driver_Contributed	Boolean	True: any large vehicle driver had any CF assigned to them	Used moderately in Humberside, but used extensively nationally
404	Environmental_Factor_Contributed	Boolean	True: any participant had an environmental, vision-affected or other specific CF assigned to them	Used moderately in Humberside, but used extensively nationally
405	Vehicle_Factor_Contributed	Boolean	True: any driver or rider had a vehicle defect CF assigned to them	Ignored as irrelevant
406	Driver_Crime_Contributed	Boolean	True: any driver or rider had a crime-related CF assigned to them	Ignored as irrelevant
407	Driver_Intoxicated_Contributed	Boolean	True: any driver or rider had an intoxication CF assigned to them	Ignored as irrelevant
408	Driver_SpeedChoice_Contributed	Boolean	True: any driver or rider had a speed choice CF assigned to them	Used subtly
409	Driver_MobilePhone_Contributed	Boolean	True: any driver or rider had mobile phone CF assigned to them	Ignored as irrelevant
410	Driver_CloseFollowing_Contributed	Boolean	True: any driver or rider had close following CF assigned to them	Used subtly
411	Driver_Disobeyed_Contributed	Boolean	True: any driver or rider had any 'disobeyed sign or marking' CF assigned to them	Used subtly
412	Driver_Observation_Contributed	Boolean	True: any driver or rider had any observation CF assigned to them	Used moderately in Humberside, but used extensively nationally

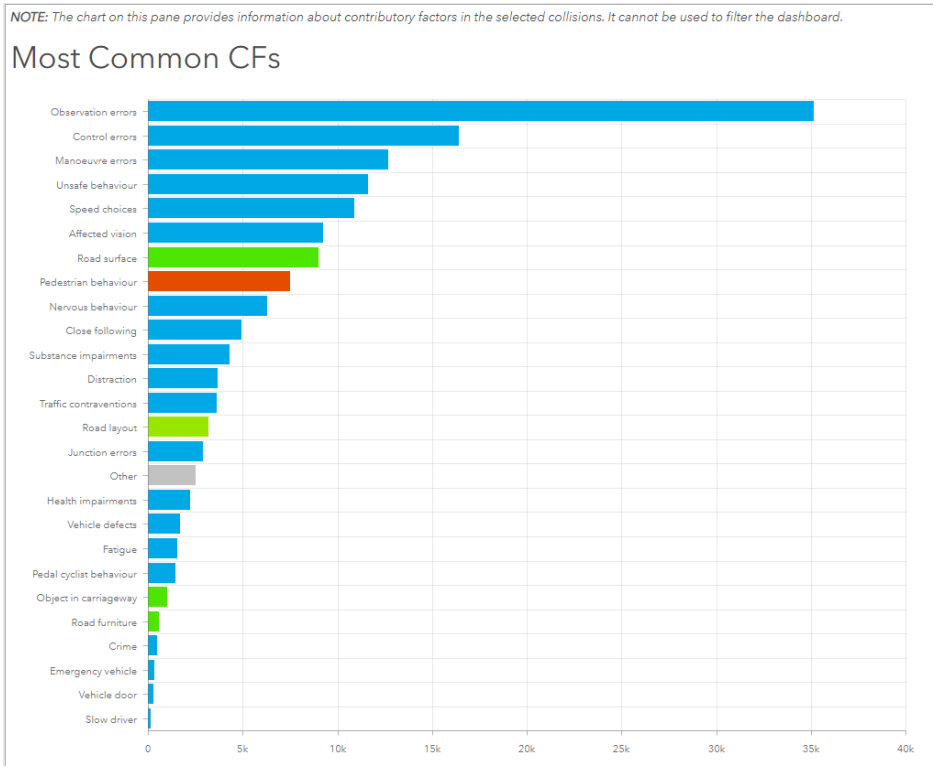
413	Driver_Fatigue_Contributed	Boolean	True: any driver or rider had fatigue CF assigned to them	Ignored as irrelevant
414	Driver_Distracted_Contributed	Boolean	True: any driver or rider had any distraction CF assigned to them	Ignored as irrelevant
415	Driver_Careless_Contributed	Boolean	True: any driver or rider had aggressive and/or careless CF assigned to them	Used subtly in Humberside, but ignored as irrelevant nationally

Source: Author's own

Figure A.4 shows some contributory factor (CF) variables applied to trend analysis in an area dashboard.

Figure A.4: Contributory factor (CF) variables applied to trend analysis in an area dashboard



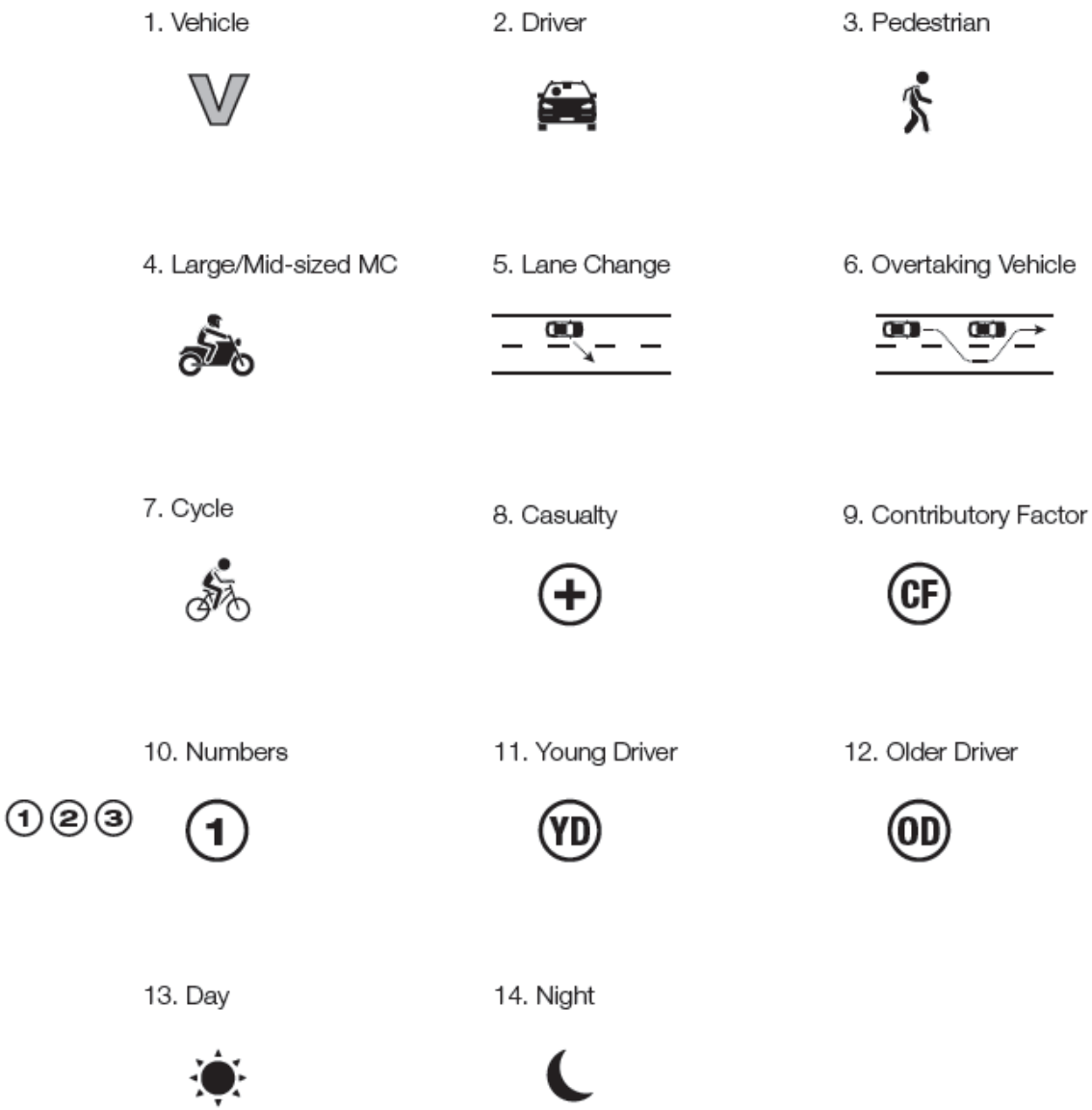


Source: Author's own

Appendix B: Infographics Key

Figure A.5 shows the icon definitions

Figure A.5: Icon directory (source: author's own)



15. Other Impact

16. Driver Observation CF

17. Shunt

18. Uncontrolled Junction

19. Right Turn

20. Driver Turning CF

21. Environment CF

22. Runoff (nearside)

23. Young Driver

24. Older Driver

25. Night (streetlights)

26. Night (no streetlights)

27. Runoff (other)

28. Pedestrian footway

29. Slow Vehicle Manouever

30. Working Driver

31. Head On

32. Roundabout

33. Cyclist Casualty

34. Pedestrian Casualty

35. Deprived Driver

36. Hit and Run

37. LGV/PSV

38. Side Impact

39. Vehicle not in Carriageway

40. Pedestrian CF

41. Single Vehicle

42. Child Pedestrian Casualty

43. LGV/PSV CF

44. Driver Overtaking CF

45. P2W rider CF

46. Large MC

47. Mid-sized MC

48. Moped

49. Left Turn

50. Controlled Junction

51. Pedestrian on Crossing

52. Adverse Weather

53. Weekend

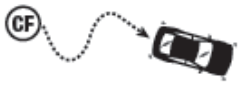
54. Close following CF

55. No distinguishing Features

56. Speed Choice CF

57. Lane Change CF

58. Careless driver CF



59. More likely hit and run

60. Cyclist CF

61. Working Young or Old

The Royal Automobile Club Foundation for Motoring Ltd is a transport policy and research organisation which explores the economic, mobility, safety and environmental issues relating to roads and their users. The Foundation publishes independent and authoritative research with which it promotes informed debate and advocates policy in the interest of the responsible motorist.

RAC Foundation

89–91 Pall Mall

London

SW1Y 5HS

Tel no: 020 7747 3445

www.racfoundation.org

Registered Charity No. 1002705