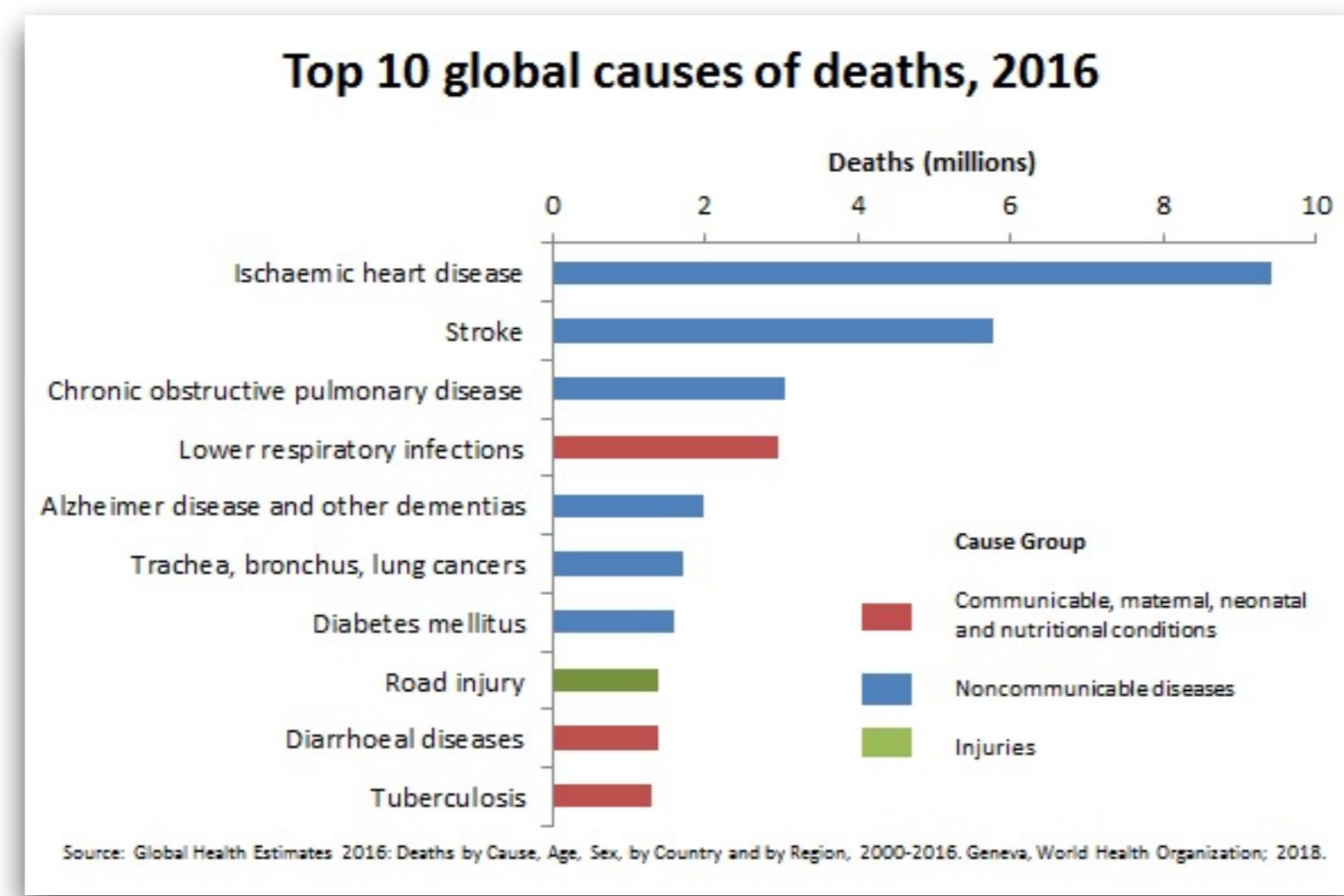


# Reproducible road safety research an exploration of the shifting spatial and temporal distribution of car-pedestrian crashes

Robin Lovelace  
Layik Hama  
Roger Beecham

call to action



call to action : road safety research

call to action

The screenshot shows the ROpenSci package page for `stats19`. The top navigation bar includes links for `stats19 0.2.1`, `Get started`, `Reference`, `Articles`, and `Changelog`. The main content area features a heading **stats19** and a brief description of the package's purpose: providing functions for downloading and formatting road crash data from the UK's official road traffic casualty database, `STATS19`. It highlights the package's goal of making road safety research more reproducible and accessible by automating the data handling process. Below this, there's a section on **Installation** with code snippets for installing the package via GitHub and CRAN. A detailed description of the `get_stats19()` function is provided, noting its parameters and the provision of data files categorized by year and type. To the right of the main content, there are sections for **Links** (CRAN download, GitHub issues), **License** (Full license, GPL-3), **Citation** (Citing `stats19`), **Developers** (Robin Lovelace, Malcolm Morgan, Layik Hama, Mark Padgham, All authors...), and **Dev status** (CRAN 0.2.1, build passing, codecov 90%, downloads 3198).

**stats19** 0.2.1

Get started Reference Articles Changelog

# stats19

**stats19** provides functions for downloading and formatting road crash data. Specifically, it enables access to the UK's official road traffic casualty database, [STATS19](#). (The name comes from the form used by the police to record car crashes and other incidents resulting in casualties on the roads.)

A full overview of STATS19 variables be found in a [document](#) provided by the UK's Department for Transport (DfT).

The raw data is provided as a series of `.csv` files that contain integers and which are stored in dozens of `.zip` files. Finding, reading-in and formatting the data for research can be a time consuming process subject to human error. **stats19** speeds up these vital but boring and error-prone stages of the research process with a single function: `get_stats19()`. By allowing public access to properly labelled road crash data, **stats19** aims to make road safety research more reproducible and accessible.

For transparency and modularity, each stage can be undertaken separately, as documented in the [stats19 vignette](#).

## Installation

Install and load the latest version with:

```
remotes::install_github("ropensci/stats19")
```

```
library(stats19)
#> Data provided under OGL v3.0. Cite the source and link to:
#> www.nationalarchives.gov.uk/doc/open-government-licence/version/3/
```

You can install the released version of `stats19` from [CRAN](#) with:

```
install.packages("stats19")
```

## get\_stats19()

`get_stats19()` requires `year` and `type` parameters, mirroring the provision of STATS19 data files, which are categorised by year (from 1979 onward) and type (with separate tables for crashes, casualties and vehicles, as outlined below). The following command, for example, gets crash data from 2017 (**note**: we follow the "crash not accident" campaign of [RoadPeace](#) in naming crashes, although the

### Links

Download from CRAN at [https://cloud.r-project.org/  
package=stats19](https://cloud.r-project.org/package=stats19)

Report a bug at [https://github.com/ropensci/stats19/  
issues](https://github.com/ropensci/stats19/issues)

### License

[Full license](#)  
[GPL-3](#)

### Citation

[Citing stats19](#)

### Developers

Robin Lovelace  
Author, maintainer

Malcolm Morgan  
Author

Layik Hama  
Author

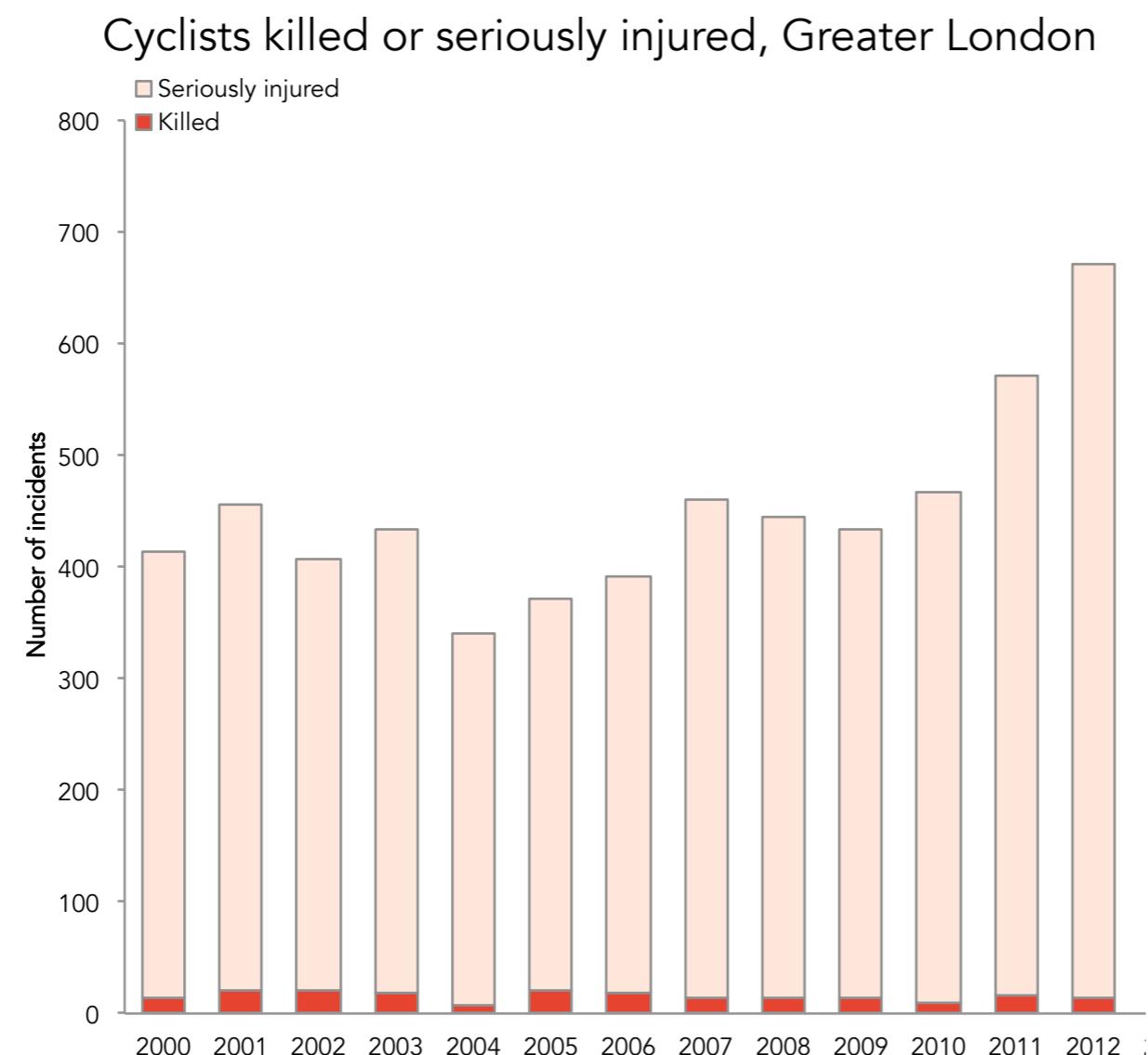
Mark Padgham  
Author

All authors...

### Dev status

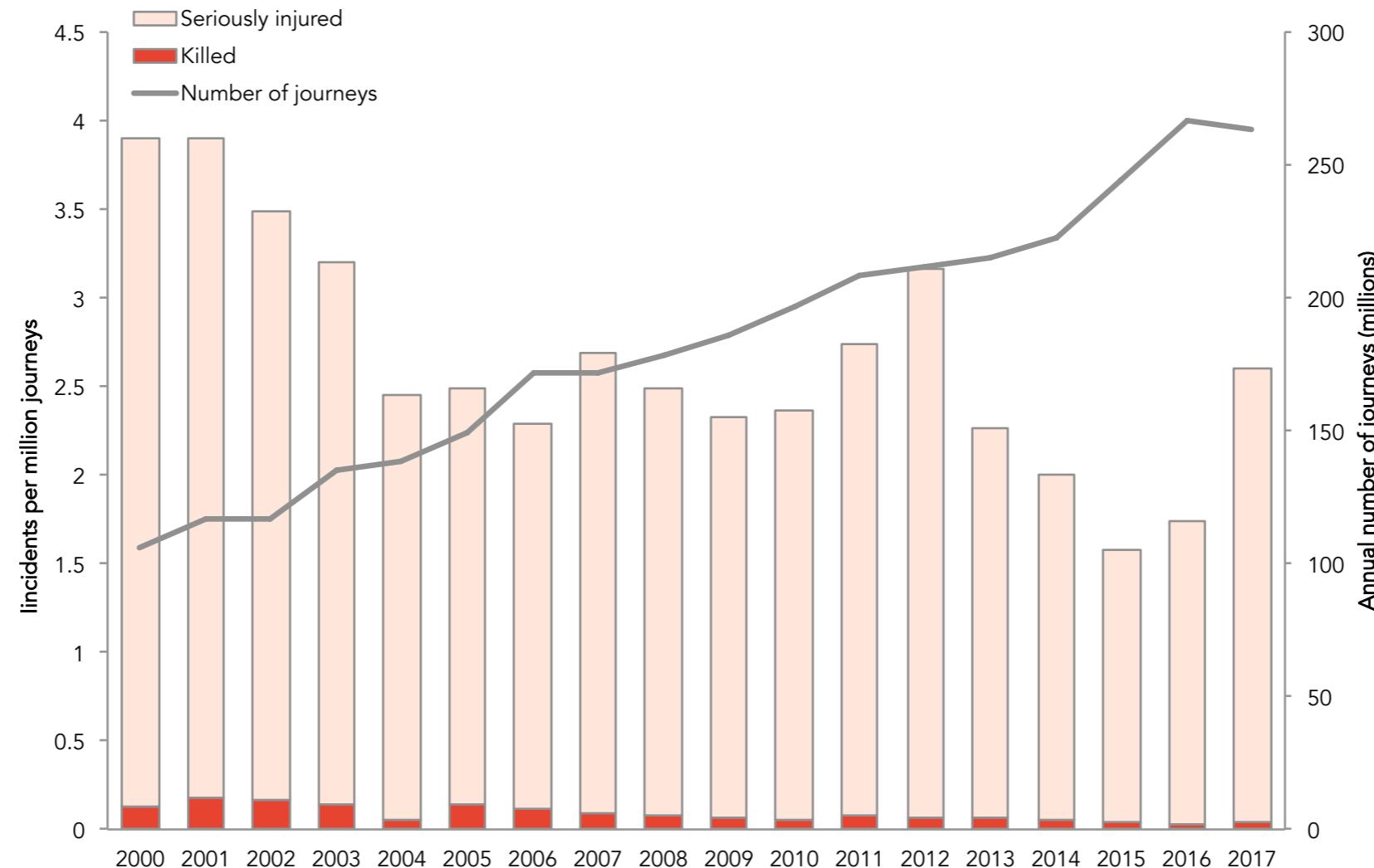
CRAN 0.2.1  
build passing  
 codecov 90%  
downloads 3198

call to action



call to action

### Bike journeys and KSI rate, Greater London



denominator neglect

stats19 package : why and how

reproducible road safety research

citizen science : tools to inform

**BETA** This is a new service – your [feedback](#) will help us to improve it

[Home](#) > Department for Transport > Road Safety Data

## Road Safety Data

**Published by:** Department for Transport  
**Last updated:** 08 November 2018  
**Topic:** Not added  
**Licence:** [Open Government Licence](#)

### Summary

These files provide detailed road safety data about the circumstances of personal injury road accidents in GB from 1979, the types of vehicles involved and the consequential casualties. The statistics relate only to personal injury accidents on public roads that are reported to the police, and subsequently recorded, using the STATS19 accident reporting form.

[View full summary](#)

### More from this publisher

[All datasets from Department for Transport](#)

### Related datasets

[Road Safety](#)

[Database of BAC levels in road fatalities](#)

[Scottish Road Accident Statistics](#)

[GM Road Casualty Accidents \(Full STATS19 data\)](#)

### Search

	accidents	
accident_index	..	crash time, location (~10m resolution), admin location, road context and weather and light conditions
	casualties	
accident_index	..	casualty sex, age, severity, passenger type, home area demos
vehicle_reference	..	
casualty_reference	..	
	vehicles	
accident_index	..	vehicles type, age, capacity, manoeuvre type, driver sex, age, home area demogs, purpose
vehicle_reference	..	

# stats19 package

The screenshot shows the top navigation bar of the stats19 package documentation. It includes a logo, the text "stats19 0.2.1", and links for "Get started", "Reference", "Articles", and "Changelog".

## stats19

**stats19** provides functions for downloading and formatting road crash data. Specifically, it enables access to the UK's official road traffic casualty database, **STATS19**. (The name comes from the form used by the police to record car crashes and other incidents resulting in casualties on the roads.)

A full overview of STATS19 variables be found in a [document](#) provided by the UK's Department for Transport (DfT).

The raw data is provided as a series of `.csv` files that contain integers and which are stored in dozens of `.zip` files. Finding, reading-in and formatting the data for research can be a time consuming process subject to human error. **stats19** speeds up these vital but boring and error-prone stages of the research process with a single function: `get_stats19()`. By allowing public access to properly labelled road crash data, **stats19** aims to make road safety research more reproducible and accessible.

For transparency and modularity, each stage can be undertaken separately, as documented in the [stats19 vignette](#).

## Installation

Install and load the latest version with:

```
remotes::install_github("ropensci/stats19")
```

```
library(stats19)
#> Data provided under OGL v3.0. Cite the source and link to:
#> www.nationalarchives.gov.uk/doc/open-government-licence/version/3/
```

You can install the released version of stats19 from [CRAN](#) with:

```
install.packages("stats19")
```

## get\_stats19()

`get_stats19()` requires `year` and `type` parameters, mirroring the provision of STATS19 data files, which are categorised by year (from 1979 onward) and type (with separate tables for crashes, casualties and vehicles, as outlined below). The following command, for example, gets crash data from 2017 (**note**: we follow the “crash not accident” campaign of [RoadPeace](#) in naming crashes, although the

## Links

Download from CRAN at  
[https://cloud.r-project.org/  
package=stats19](https://cloud.r-project.org/package=stats19)

Report a bug at  
[https://github.com/ropensci/stats19/  
issues](https://github.com/ropensci/stats19/issues)

## License

[Full license](#)  
[GPL-3](#)

## Citation

[Citing stats19](#)

## Developers

Robin Lovelace  
Author, maintainer

Malcolm Morgan  
Author

Layik Hama  
Author

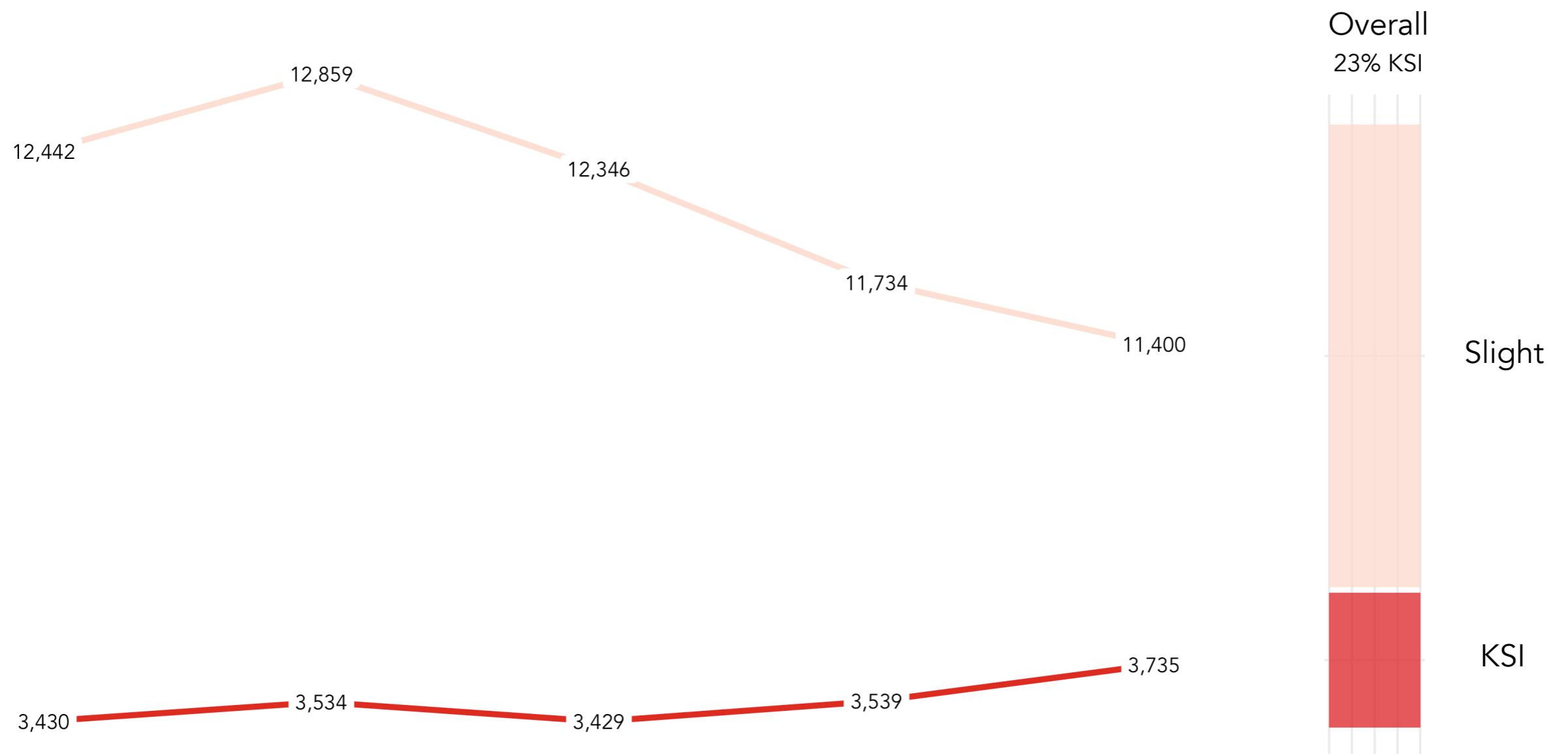
Mark Padgham  
Author

[All authors...](#)

## Dev status

CRAN 0.2.1  
 build passing  
 [codecov](#) 90%  
 downloads 3198

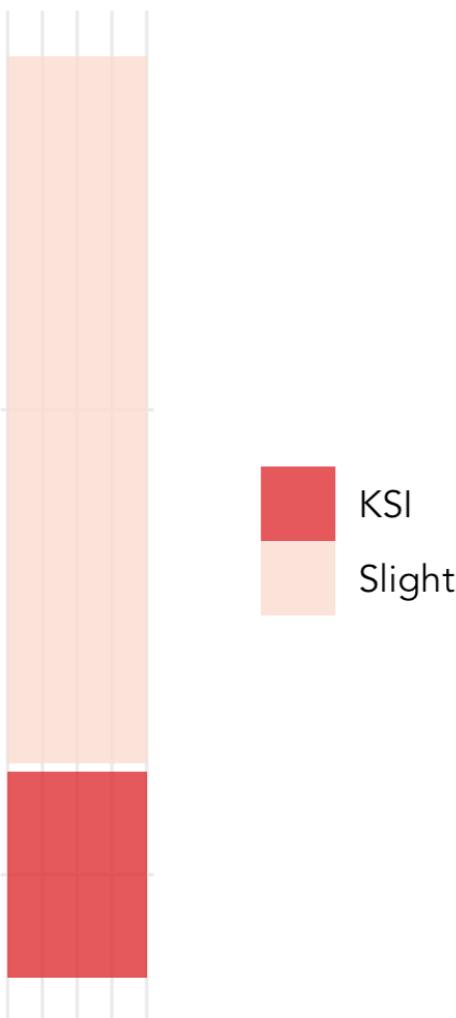
```
install.packages("stats19")  
  
get_stats19(  
  year = <yyyy>,  
  type = <"Accidents">  
    <"Casualties">  
    <"Vehicles">  
)
```



Car-pedestrian casualties  
2013-2017

Overall

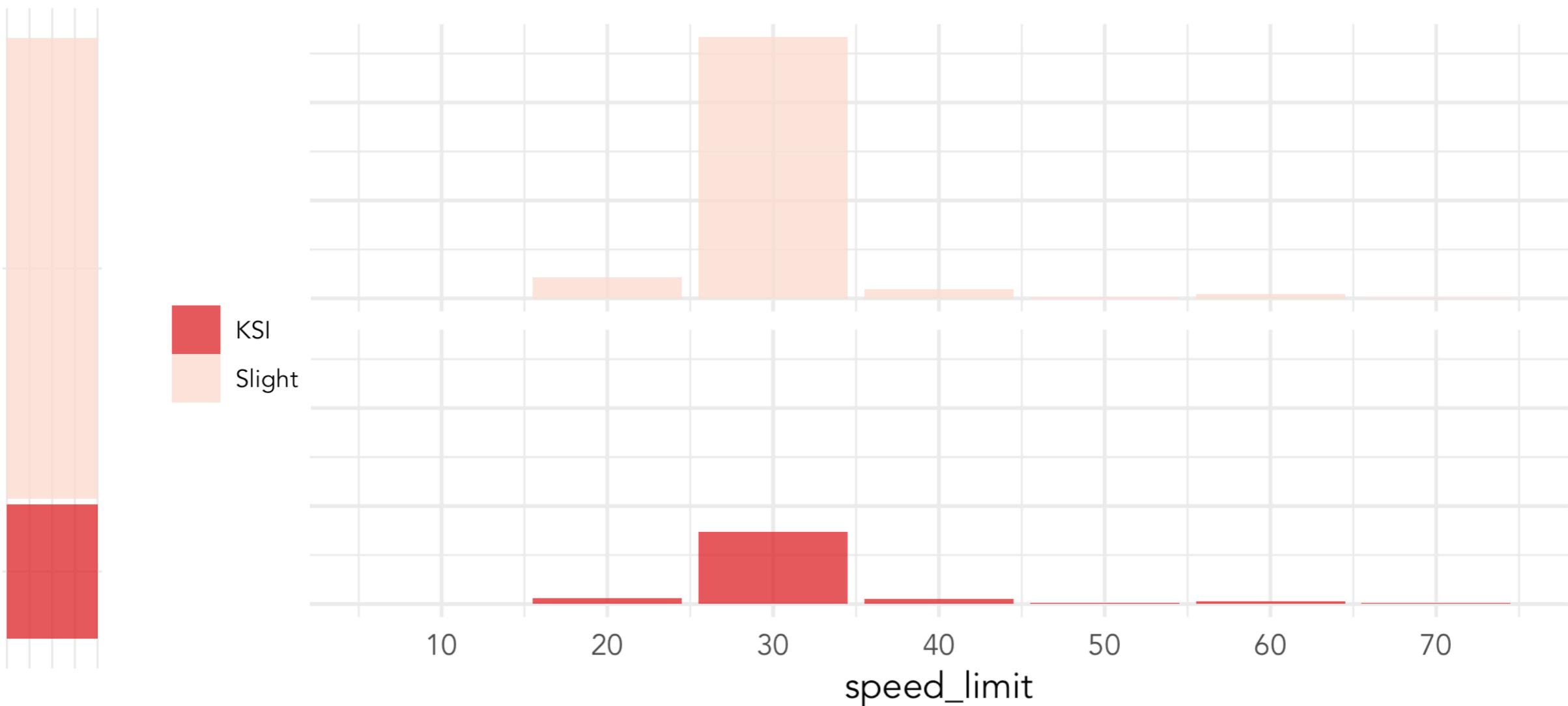
23% KSI



Car-pedestrian casualties  
by mph road

Overall

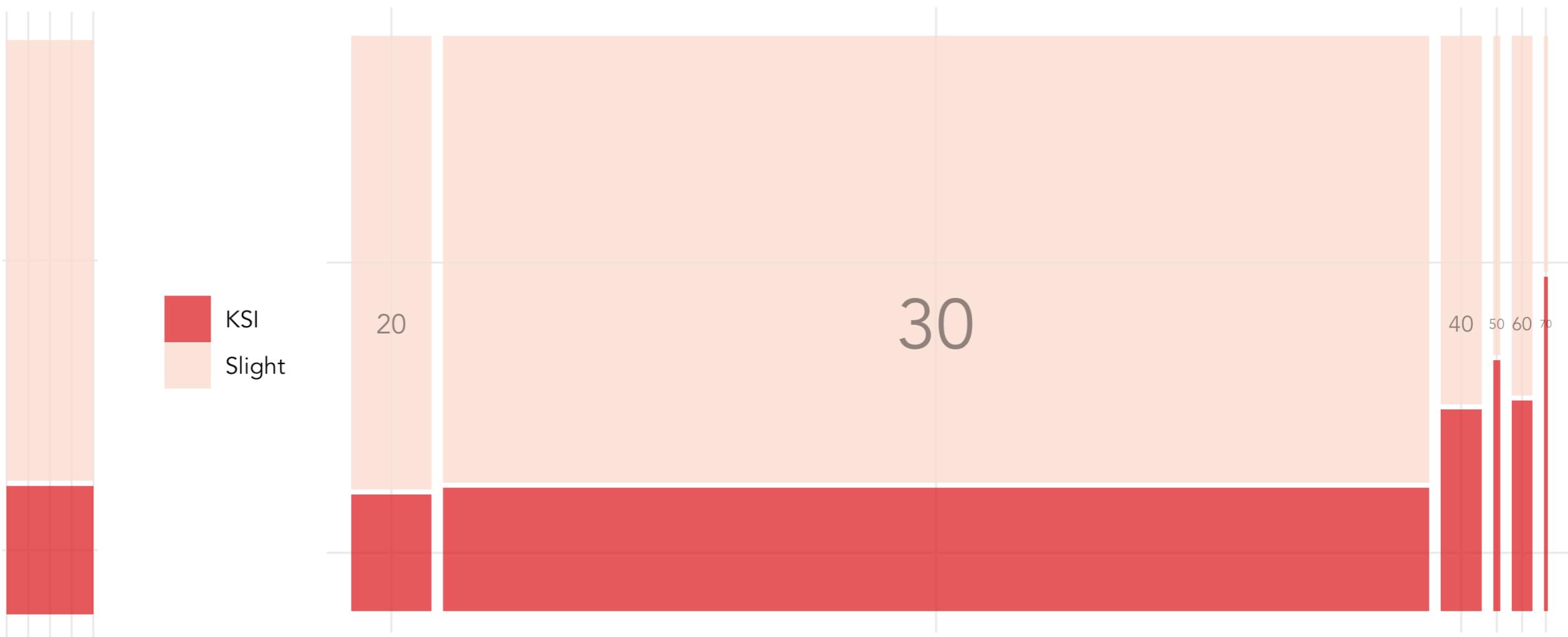
23% KSI



Car-pedestrian casualties  
by mph road

Overall

23% KSI



Car-pedestrian casualties  
by mph road

Overall

23% KSI

Rural

20

30

40

50

60

70

KSI  
Slight

Urban

20

30

40

50

60

70

Car-pedestrian casualties  
by mph road

Overall

23% KSI

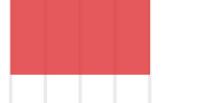
30

60

Daylight

Darkness -  
lights litDarkness  
- no  
lighting

Daylight

Darkness -  
lights litDarkness  
- no  
lighting

Car-pedestrian casualties

30mph | 60mph by lighting condition

Overall

23% KSI

30

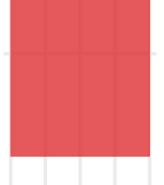
60

Dry

Wet or  
damp

Dry

Wet or  
damp



Car-pedestrian casualties

30mph | 60mph by surface condition



$$\text{risk-ratio}_{\text{force}} = \frac{\text{ksi}_{\text{force}}}{\text{ksi}_{\text{nat}}}$$

Car-pedestrian casualties  
30mph roads by Police Force

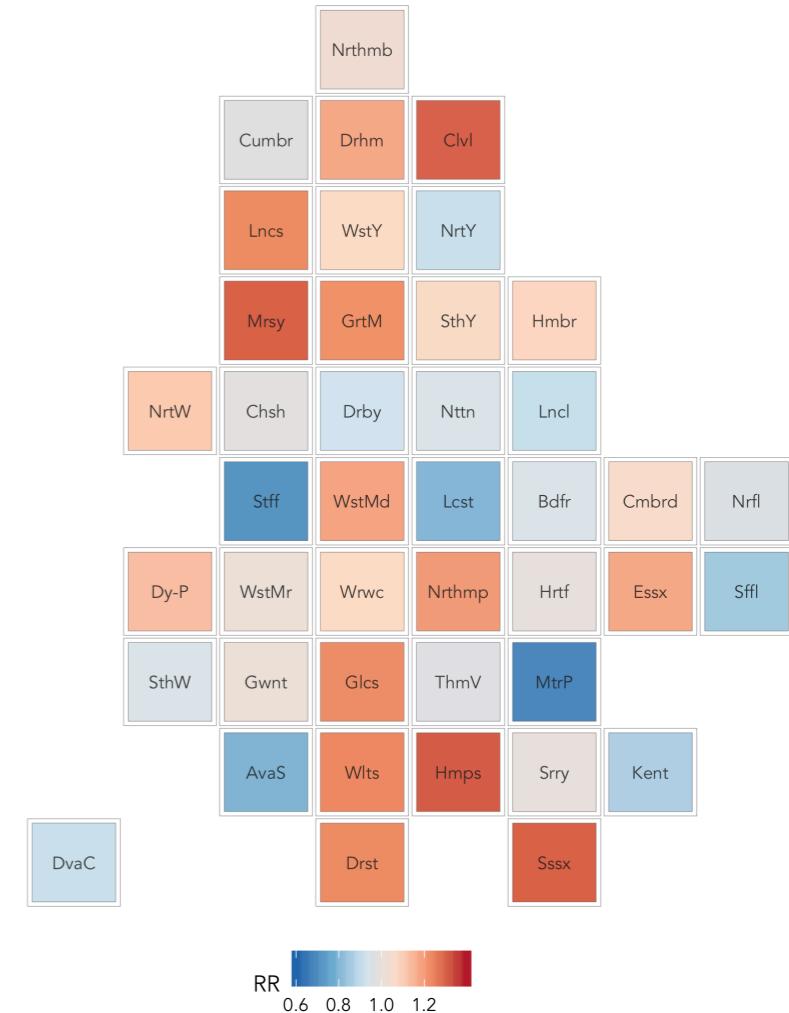


$$\text{risk-ratio}_{\text{force}} = \frac{\text{ksi}_{\text{force}}}{\text{ksi}_{\text{nat}}}$$

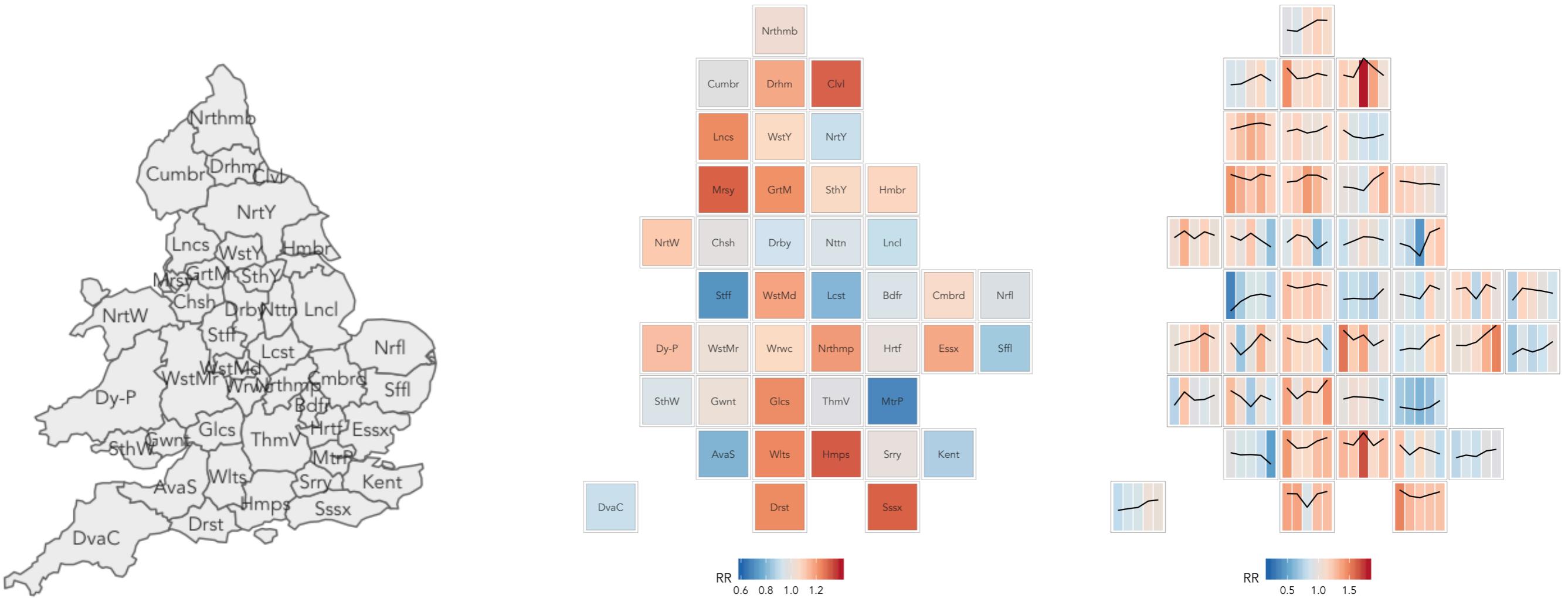
Car-pedestrian casualties  
30mph roads by Police Force



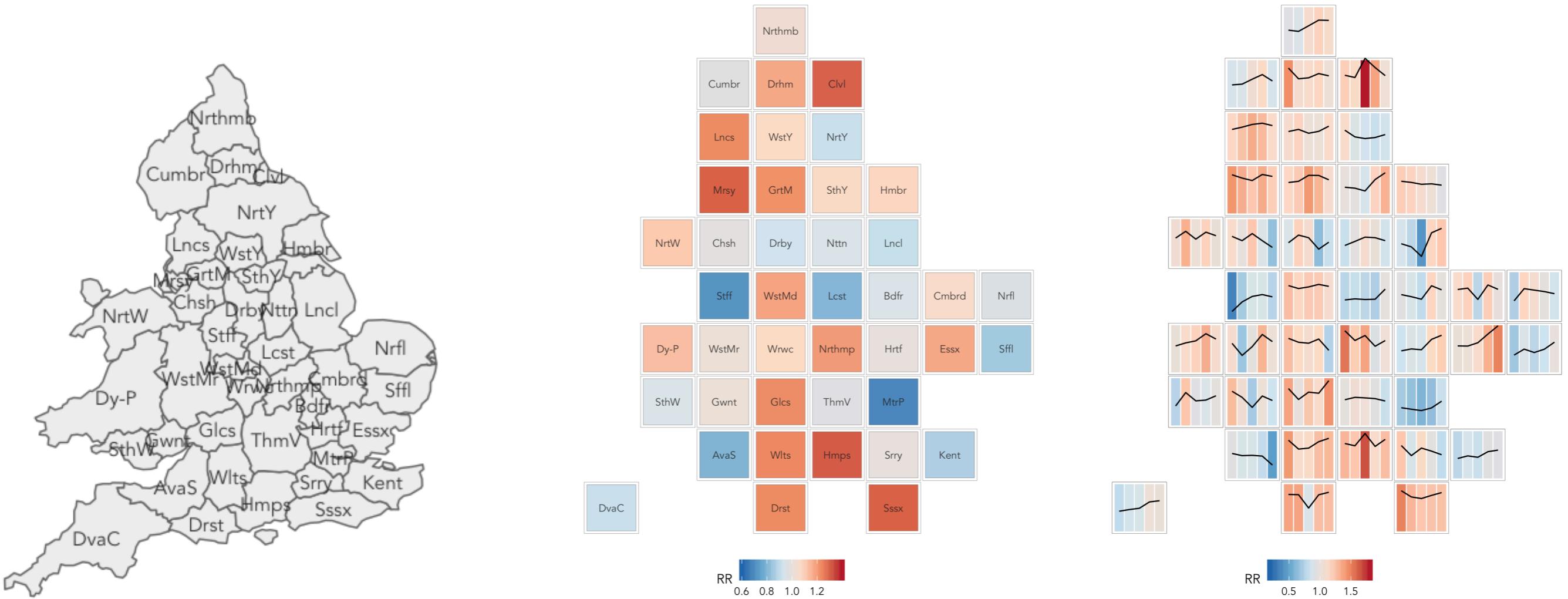
$$\text{risk-ratio}_{\text{force}} = \frac{\text{ksi}_{\text{force}}}{\text{ksi}_{\text{nat}}}$$



Car-pedestrian casualties  
30mph roads by Police Force



Car-pedestrian casualties  
30mph roads by Police Force, year-on-year



Car-pedestrian casualties  
30mph roads by Police Force, year-on-year

# call to action : exposure

denominator

vehicle miles travelled

traffic volume  
on network

temporal traffic volume  
on network

datasets

National Travel Survey

Census  
Travel demand survey  
TfL count data

ATC data

DfT

TfL

DfT



Cycling injury risk in London: A case-control study exploring the impact of cycle volumes, motor vehicle volumes, and road characteristics including speed limits

Rachel Aldred<sup>a,\*</sup>, Anna Goodman<sup>b</sup>, John Gulliver<sup>c</sup>, James Woodcock<sup>d</sup>

<sup>a</sup> University of Westminster, United Kingdom  
<sup>b</sup> London School of Hygiene and Tropical Medicine, United Kingdom  
<sup>c</sup> Imperial College, London, United Kingdom  
<sup>d</sup> Centre for Diet and Activity Research, University of Cambridge, United Kingdom

**ARTICLE INFO**  
**Keywords:** Cycling, Injury, Motor traffic, Risk, Safety in numbers  
  
**ABSTRACT**  
Cycling injury risk is an important topic, but few studies explore cycling risk in relation to exposure. This is largely because of a lack of exposure data, in other words, much cycling is done at different locations. This paper aims to address this by using a case-control study of cycling injuries in London in 2010–2011, using modelled cyclist flow data alongside datasets covering some characteristics of the London road network. A multilevel binary logistic regression model is used to investigate factors associated with injury risk, comparing injury risk across different locations, and examining the relationship between injury risk and road characteristics and numbers: for each increase of a natural logarithmic unit (2.71828) in cycling flows, an 18% decrease in injury odds was found. Conversely, increased motor traffic volume is associated with higher odds of cycling injury, with one logarithmic unit increase in traffic per hour associated with 21% higher injury odds. Residential areas with 20 mph speed limits were associated with 21% lower injury odds. Residential areas with junctions with reduced injury odds, and junctions with substantially higher injury odds. Bus lanes do not affect injury odds once other factors are controlled for. These data suggest that speed limits of 20 mph may reduce cycling injury risk, as may motor traffic reduction. Further, building cycle routes that generate new cycle trips should generate safety in numbers benefits.

Aldred et al. 2018

# call to action : cumulative knowledge building

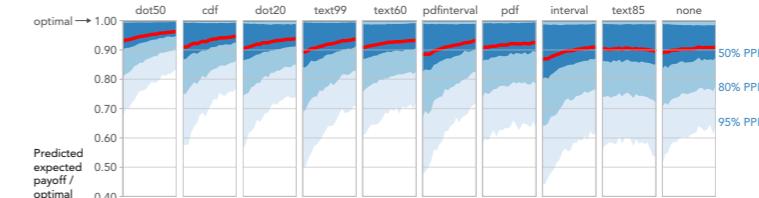
## Researcher-Centered Design of Statistics: Why Bayesian Statistics Better Fit the Culture and Incentives of HCI

**Matthew Kay**  
 Computer Science &  
 Engineering | dub,  
 University of Washington  
[mjskay@uw.edu](mailto:mjskay@uw.edu)

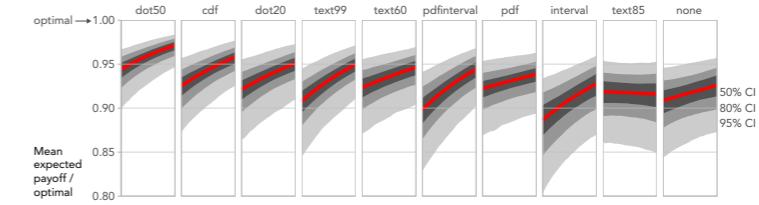
**Gregory L. Nelson**  
 Computer Science &  
 Engineering | dub,  
 University of Washington  
[glnelson@uw.edu](mailto:glnelson@uw.edu)

**Eric B. Hekler**  
 School of Nutrition &  
 Health Promotion  
 Arizona State University  
[ehekler@asu.edu](mailto:ehekler@asu.edu)

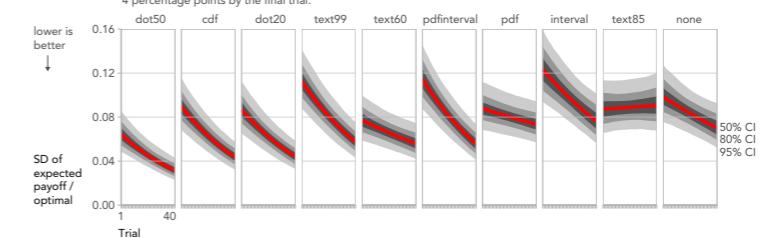
1. Posterior predictive intervals and predicted mean for performance in each condition. These intervals are what we would predict 50%, 80%, or 95% of new observations of performance to fall into. Performance improves with additional trials, especially for dot50, cdf, and dot20. Meanwhile, performance in text depends on the risk threshold, with text85 performing similarly to no uncertainty.



2. Quantile credible intervals (Bayesian analog to confidence intervals) and posterior median of the mean performance in each condition. These intervals show the uncertainty in the location of the red line in chart #1, above.



3. Quantile credible intervals and posterior median of the standard deviation of performance in each condition. Not only does mean performance improve, but variance in performance also improves over time: people get better on average and more consistent; dot50, the best-performing condition, achieves an SD in performance likely less than 4 percentage points by the final trial.



## Uncertainty Displays Using Quantile Dotplots or CDFs Improve Transit Decision-Making

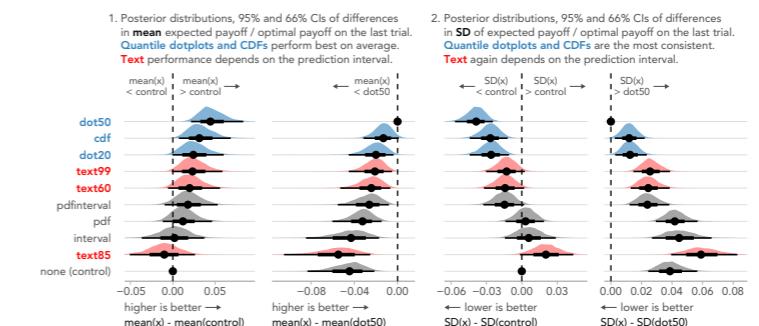
Michael Fernandes<sup>1</sup>, Logan Walls<sup>1</sup>, Sean Munson<sup>1</sup>, Jessica Hullman<sup>1</sup>, and Matthew Kay<sup>2</sup>

<sup>1</sup>University of Washington  
 Seattle, WA, USA  
[mfern, logan.w.gm, smunson, jhullman@uw.edu](mailto:mfern, logan.w.gm, smunson, jhullman@uw.edu)

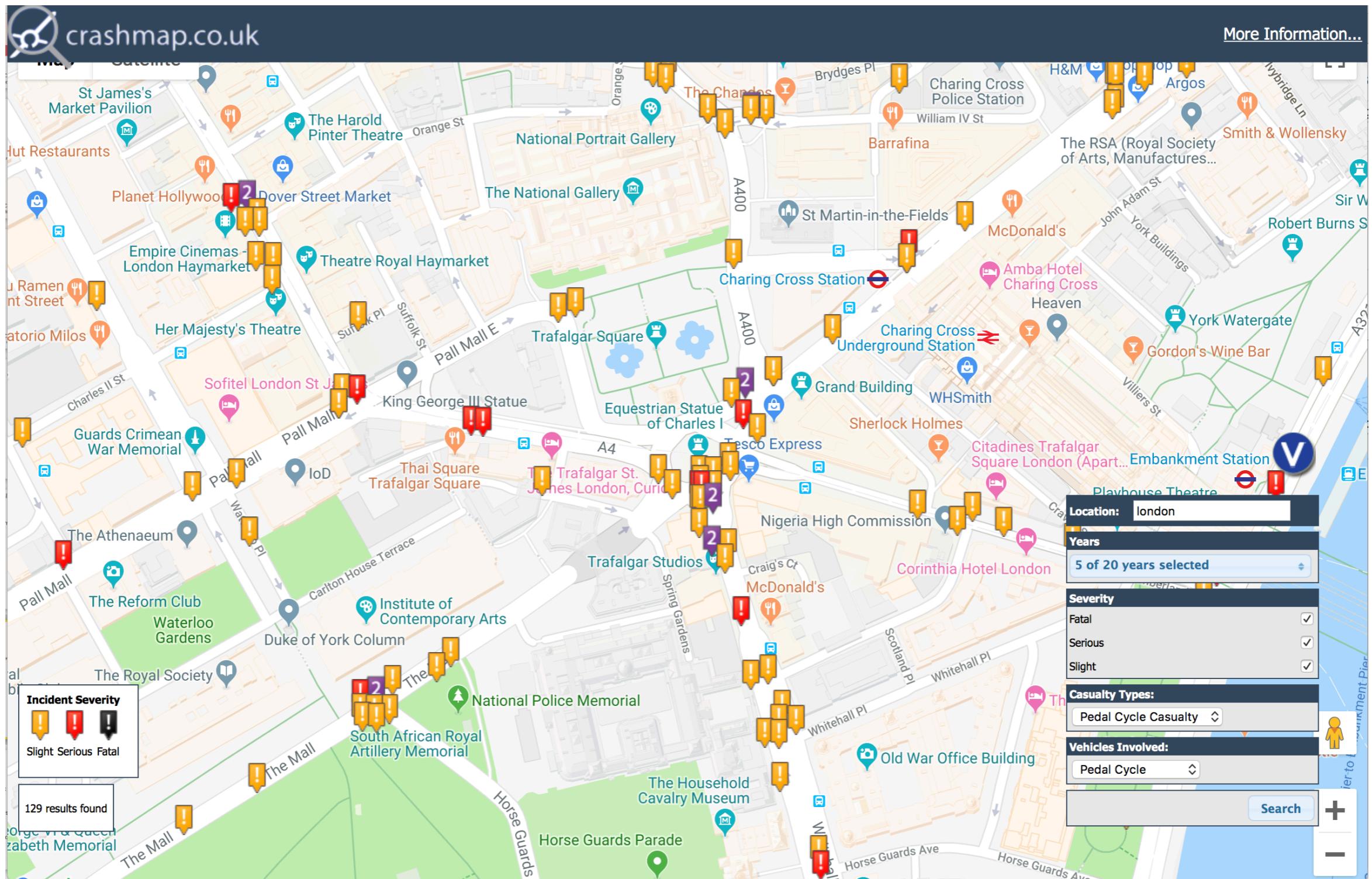
<sup>2</sup>University of Michigan  
 Ann Arbor, MI, USA  
[mjskay@umich.edu](mailto:mjskay@umich.edu)

## Beyond Weber's Law: A Second Look at Ranking Visualizations of Correlation

Matthew Kay and Jeffrey Heer



# call to action : citizen science



**paper**

<https://github.com/Robinlovelace/stats19-gisruk>

**stats19**

<https://itsleeds.github.io/stats19/>