**Reducing Traffic Mortality in the USA**

How can we find a good strategy for reducing traffic-related deaths?

#### Project Description

While the rate of fatal road accidents has been decreasing steadily since the 80s, the past ten years have seen a stagnation in this reduction. Coupled with the increase in number of miles driven in the nation, the total number of traffic related-fatalities has now reached a ten year high and is rapidly increasing.

By looking at the demographics of traﬃc accident victims for each US state, we find that there is a lot of variation between states. Now we want to understand if there are patterns in this variation in order to derive suggestions for a policy action plan. In particular, instead of implementing a costly nation-wide plan we want to focus on groups of states with similar profiles. How can we find such groups in a statistically sound way and communicate the result effectively?

This project lets you apply skills from:

* [Introduction to Shell for Data Science](https://www.datacamp.com/courses/introduction-to-shell-for-data-science), including how to navigate the file system and view files
* [pandas Foundations](https://www.datacamp.com/courses/pandas-foundations), including reading, exploring, filtering, and grouping data
* [Manipulating DataFrames with pandas](https://www.datacamp.com/courses/manipulating-dataframes-with-pandas), including how to reshape data into the long format and how to perform multiple aggregations
* [Merging DataFrames with pandas](https://www.datacamp.com/courses/merging-dataframes-with-pandas), including how two merge two DataFrames
* [Unsupervised Learning in Python](https://www.datacamp.com/courses/unsupervised-learning-in-python), including KMeans clustering, dimensionally reduction through PCA, and visualizations using matplotlib
* [Supervised Learning with scikit-learn](https://www.datacamp.com/courses/supervised-learning-with-scikit-learn), including multivariate regression
* [Intermediate Python for Data Science](https://www.datacamp.com/courses/intermediate-python-for-data-science), including visualizations using matplotlib
* [Data Visualization With Seaborn](https://www.datacamp.com/courses/data-visualization-with-seaborn), including statistical visualizations using seaborn

We recommend that you review the appropriate sections of those courses before starting this project.

#### Project Tasks

* 1 The raw data files and their format
* 2 Read in and get an overview of the data
* 3 Create a textual and a graphical summary of the data
* 4 Quantify the association of features and accidents
* 5 Fit a multivariate linear regression
* 6 Perform PCA on standardized data
* 7 Visualize the first two principal components
* 8 Find clusters of similar states in the data
* 9 KMeans to visualize clusters in the PCA scatter plot
* 10 Visualize the feature differences between the clusters
* 11 Compute the number of accidents within each cluster
* 12 Make a decision when there is no clear right choice