SI618 Data Manipulation and Analysis

Project 1 Report

Analysis of Traffic Accidents in France & the U.K.

Name: Xinhao Liao
Date: 16 October 2019

Contents

1 Introduction

1.1 Motivation

With the mass production and popularization of automobiles, road accidents have become a big issue in the past decades.

The details of road accidents have long been recorded and stored by the police in many places. Nowadays with the help of data science, we can study the accumulated data, and try to answer question like:

- What is the trend of the frequency of fatal accidents over year?
- Which role in an accident is most likely to die?
- How can lighting conditions affect the severity of accidents?
- ...

We also want to compare data in different regions, and get to know how traffic safety varies in different regions.

1.2 Datasets Description

1.2.1 Accidents in France From 2005 to 2016

The dataset from [1] describes road accidents happening in France from 2005 to 2016. Three of the *csv* files provided, *caracteristics.csv*, *places.csv*, and *users.csv*, will be involved in this project. They respectively provide the basic information, the road conditions, and the victim information of the accidents.

1.2.2 Accidents in the UK From 2005 to 2015

The dataset from [2] describes road accidents happening in the UK from 2005 to 2015. This dataset provide information such as the road types, the casualties, the vehicles, and related details in separated *csv* files.

2 Data Manipulation Methods

2.1 Fatal accidents over years

In this section, we are interested in the number of fatal accidents in France and the UK over years, and try to find out a trend of number fatal accidents in the two countries, and of the overall fatal rate. We can do this by joining the datasets for causality (*france/users.csv*) and date (*france/caracteristics.csv*) of France and that of the UK (*Casualties0515.csv*), and then reduce them by the number *year* with the help of *pyspark*. After that, we can further combine the fatal counts for France and for the UK and calculate the overall fatal rate over years. The detailed procedure is coded in *fatal_number_year.py.*

2.2 Casualty analysis

With the help of *mrjob*, we can analyze the casualty in detail. Respectively mapping and reducing casualties with different severity (fatal, serious, or slight), we can count number of different severity in France (by applying *casualty_france.py* to *france/users.csv*) and in the UK (by applying *casualty_uk.py* to *uk/Casualties0515.csv*).

Similarly, we can obtain the role compositions of the fatal casualties in the two countries with the help of *mrjob*. That is, applying *fatal_rate_france.py* to *france/users.csv* and *fatal_role_uk.py* to *uk/Casualties0515.csv*.

2.3 Lighting conditions' effects on severity

In this part, we want to find out how lighting conditions affect the severity of accidents. For this analysis, we combine the data in the U.K. and the data in France. Also, we ignore the data with unknown or missing lighting conditions or in the dawn or twilight. This way, we can obtain the number of accidents of different severity (fatal, serious, or slight) with different lighting conditions. (The detailed procedures are coded in *lighting_effects.py*.) And then we can accordingly calculate the proportion of each severity condition in the datasets of three lighting conditions.

3 Analysis and Visualization

3.1 Fatal accidents over years

Following the procedures described in 2.1, we can obtain the number of fatal accidents in France and the UK, and the overall fatal rate over years from 2005 to 2015. The detailed procedure is coded in *fatal_number_year.py*.

We can visualize the results with *fatal_count_plot.py* and *fatal_rate_plot.py*. Since there's no data for accidents in 2016 in the UK, the results for accidents in 2016 in France are abandoned in the visualization to better compare the results. The results are shown in Fig. 1 and Fig. 2.

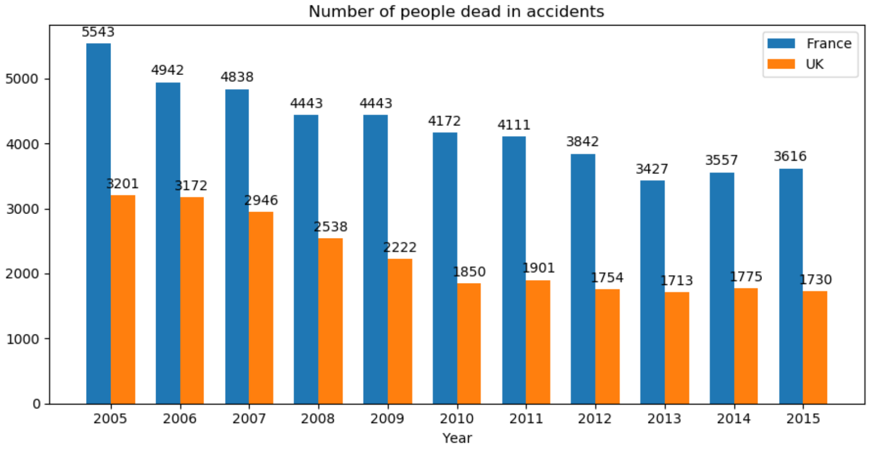

Figure 1: Number of dead people in accidents from 2005 to 2015.

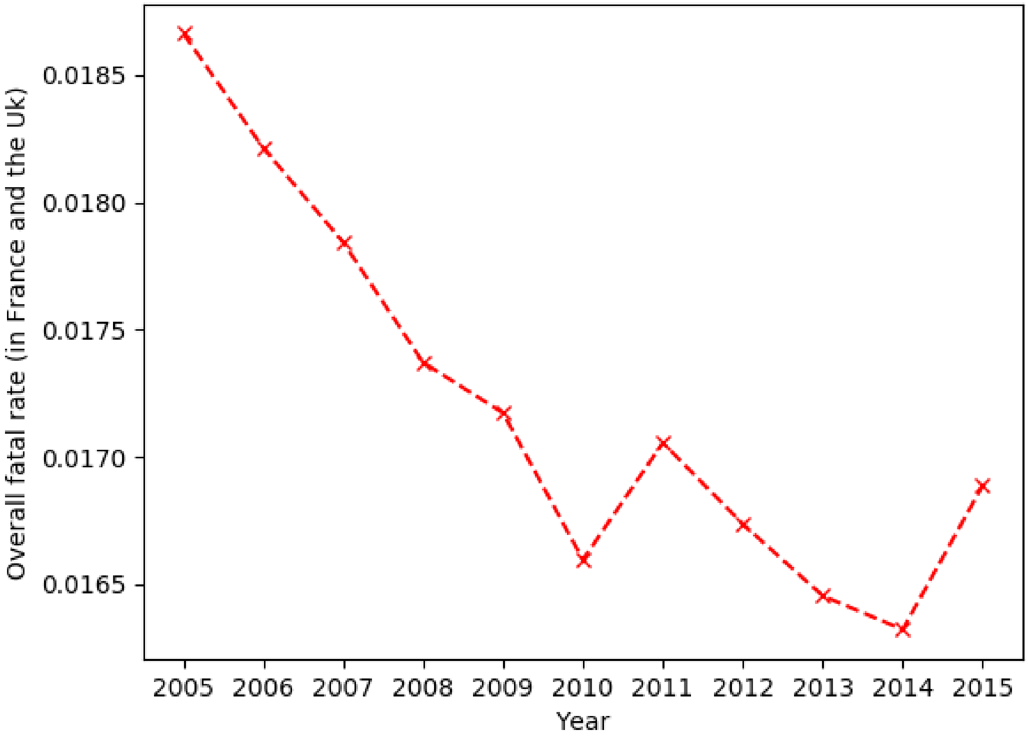

Figure 2: Overall fatal rates for accidents from 2005 to 2015 in France and the UK.

According to Fig. 1, we can see that there has been a declining trend in the number of people dead in accidents in both countries in the decade from 2005 to 2015, though the numbers have slightly increased since 2014. And the overall fatal rate in the two countries, as shown in Fig. 2, generally declined from 2005 to 2015, which is consistent with the result in Fig. 1. Generally speaking, traffic in both countries is getting safer and safer.

We can also notice that the number of dead people in accidents in France is generally twice as many as that in the UK. Considering that the two countries have similar populations (about 67.06 million in France in September 2019 and about 66.44 million in the UK in June 2018 according to [3]), traffic in France seems to be much more dangerous than that in the UK.

3.2 Casualty analysis

Following what is described in 2.2, we can obtain number of different severity casualty and then the role compositions of the dead in the two countries. The result can be shown in pie charts (generated with *casualty_piechart.py* and *fatal_roles_piechart.py*) in Fig. 3 and Fig. 4.

The result in Fig. 3 clearly shows that the severity of accidents in France is much greater than that in the UK. The serious rate in France is 21.0% which is twice as much as that in the UK. And the fatal rate in France is 2.7%, also far greater than 1.0% in the UK. This is consistent with what we have seen in 3.1, where the number of people dead in France is much greater than that in the UK. This again implies that traffic in France is much more dangerous than that in the UK.

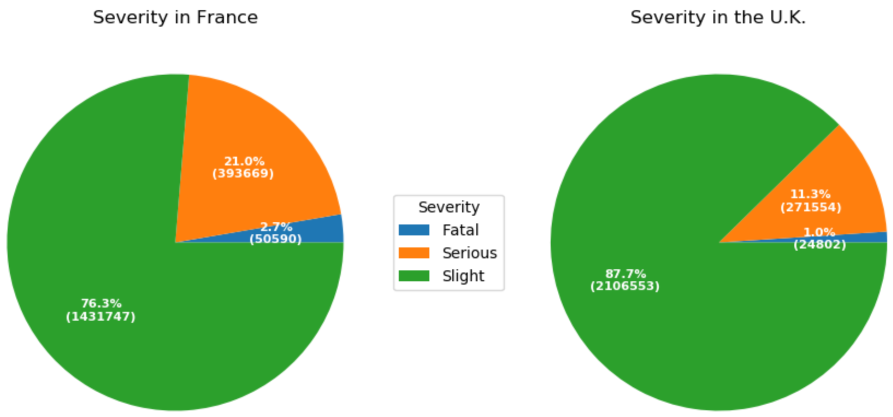

Figure 3: Severity composition of the casualty.

For the result in Fig. 4, as we can see, the role compositions are similar in the two countries, with drivers as the most number of the dead in accidents, and passengers making up of 17.3% of the dead. In France, 69.4% of the dead are drivers, which is greater than 60.2% in the UK. Another distinction is that 22.6% of the dead in the UK are pedestrians, which is far far greater than 13.3% in France. A driver is always the most dangerous in an accident as expected.

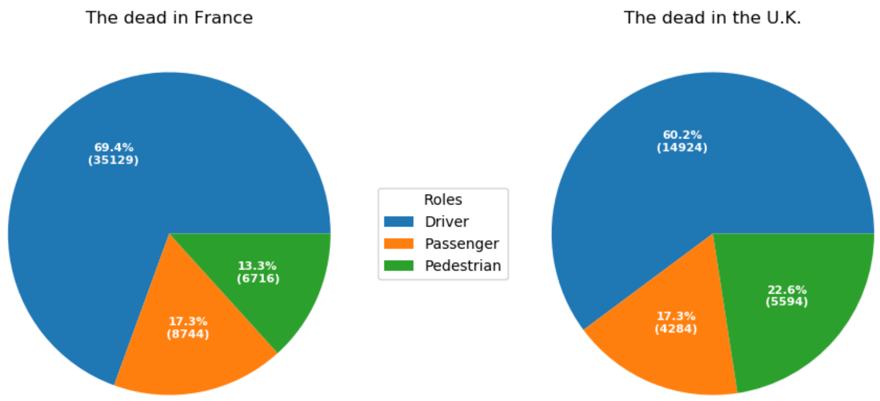

Figure 4: Roles composition of the dead.

3.3 Lighting conditions' effects on severity

Following what's described in 2.3, we can obtain the proportion of each severity condition in the datasets of three lighting conditions, and obtain a visualization (as coded in *lighting_effects_plot.py*) as shown in Fig. 5.

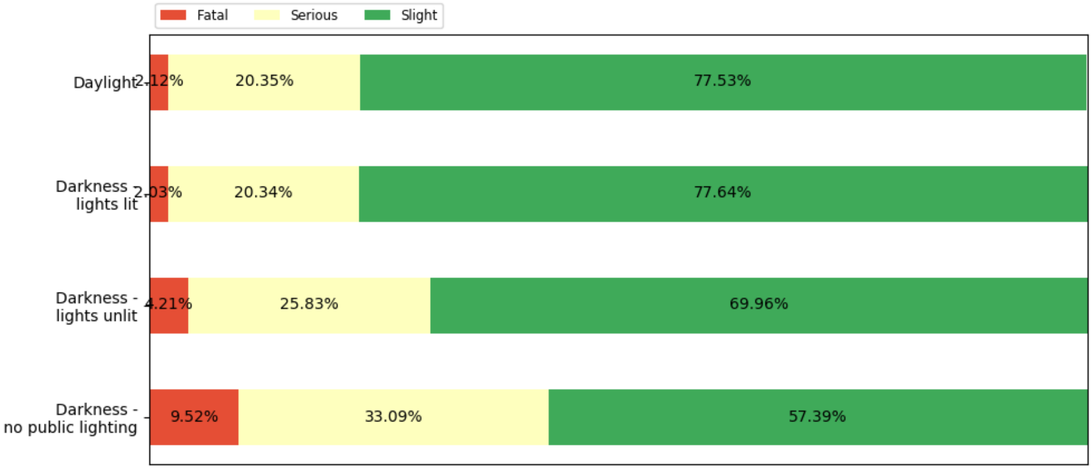

Figure 5: Severity composition of accidents in different lighting conditions.

As we can see, darkness with no public lighting corresponds to the most proportion of fatal or serious accidents (9.52% for fatal and 33.09% for serious accidents). And the fatal and serious proportion is generally growing for darker conditions as expected. It's absolutely more dangerous when driving in the darkness.

We can also notice that the daylight and the darkness with lights lit conditions correspond to almost the same severity compositions. It seems to imply that driving with public lights lit is almost as safe as driving in the day.

4 Challenges and Conclusion

The greatest challenge I met in this project is the inconsistency between the two datasets. Though both recording relevant information of accidents, what recorded in the two datasets are somehow different, and to be worse, even for something common, different criteria might be applied for measuring (mostly due to different customs in the two countries). The inconsistency makes many of my initial proposed ideas about data comparing or combining unrealistic.

So I have to dig deeper into the description of the datasets and try to find some information commonly recorded and equivalently measured. Then it came to me the challenge of understanding the different customs in the two countries. For instance, I have no idea what is road class A, B, C in the U.K., or what does VL, VU, and PL means in terms of vehicles in France, or how I can map the given road numbers to locations in the two countries, etc.

Though I had to abandon some analysis or visualization due to the inconsistency and the complex transformation between the datasets, after analyzing the data we can still reach some interesting conclusions like:

- The fatal rate of accidents is generally declining from 2005 to 2015 in both France and the UK.
- Fatal rate of accidents in France is far more greater than that in the U.K.
- Most of the dead in accidents are drivers.
- The lighting condition of darkness with no public light corresponds to the greatest fatal rate.
- Driving with public lights lit seems to be almost as safe as driving in the day
- ...

References

[1] "Accidents in France from 2005 to 2016", *Kaggle*. `https://www.kaggle.com/ahmedlahlou/accidents-in-france-from-2005-to-2016`

[2] "UK Car Accidents 2005-2015", *Kaggle*. `https://www.kaggle.com/silicon99/dft-accident-data`

[3] "List of countries and dependencies by population", *Wikipedia*. `https://en.wikipedia.org/wiki/List_of_countries_and_dependencies_by_population`