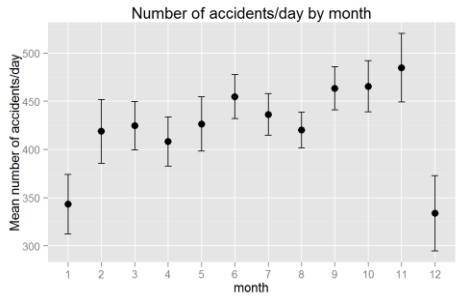
Analysis of the Accident Dataset

We have the dataset of the road accidents in Great Britain in 2010. There were 32 variables collected. We will analyze if the mean number of accidents/day varied across the months of the year.

From the result of the hypothesis test, we got F-statistic = 11.29 on 11 and 353 DF, p-value: < 2.2e-16. Therefore, we had enough evidence that the mean number of accidents/day varied across the months. The Figure 1 shows the mean number of accidents/day by month along with a 95% confidence interval. We found that the highest accident rate occurred in Sept., Oct. and Nov. and the lowest occurred in Dec. and Jan.

Figure 2 presents a histogram of the number of accidents as it varies by minute of the hour. Its shape is odd. The distribution has high frequencies at every multiples of 5 and extremely high frequencies at 0 and 30 minute. It is because the eyewitness trend to tell the police a time of multiples of 5 or 30 (It is easy to remember) instead of a very specific time. The police should not analyze the accidents rate based on the every specific minutes. It is untrusted.



Fingure1: *Estimated mean number of accidents by month along with 95% CI*

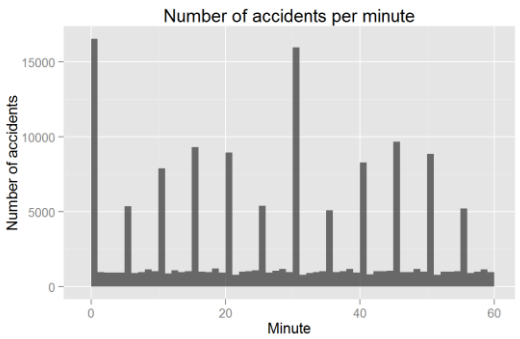


Figure2: *a histogram of the number of accidents as it varies by minute of the hour.*