# Do Seat Belt Laws Predict a Reduction in Road Casualties?

## DATA 606 Final Project
## Spring 2022

**By Julian Adames-Ng**

# Abstract

To answer the proposed question, I gathered data using the "UKDriverDeaths" time-series from the R Documentation which includes monthly statistics on road deaths in the United Kingdom from January 1969 to December 1984. The mandatory seat belt law was introduced on January 1983. In order to manipulate the data to perform analyses, I created a data-frame from this data set. This allows me to parse the data as needed and analyze measures of center and variation as well as perform correlation analyses on the data prior to and after the passing of the seat belt law. Visualization of these analyses are provided.

# Loading the Data

## Time Series

- “UKDriverDeaths is a time series giving the monthly totals of car drivers in Great Britain killed or seriously injured Jan 1969 to Dec 1984. Compulsory wearing of seat belts was introduced on 31 Jan 1983.” - R Documentation
- Seatbelts is a multiple time series, with the columns:
    - DriversKilled - car drivers killed
    - drivers - same as UKDriverDeaths (total road deaths)
    - front - front-seat passengers killed or seriously injured
    - rear - rear-seat passengers killed or seriously injured
    - kms - distance driven
    - PetrolPrice - petrol price
    - VanKilled - number of van (‘light goods vehicle’) drivers
    - law - 0/1: was the law in effect that month?

```
UKDriverDeaths
```

```
##       Jan  Feb  Mar  Apr  May  Jun  Jul  Aug  Sep  Oct  Nov  Dec
## 1969 1687 1508 1507 1385 1632 1511 1559 1630 1579 1653 2152 2148
## 1970 1752 1765 1717 1558 1575 1520 1805 1800 1719 2008 2242 2478
## 1971 2030 1655 1693 1623 1805 1746 1795 1926 1619 1992 2233 2192
## 1972 2080 1768 1835 1569 1976 1853 1965 1689 1778 1976 2397 2654
## 1973 2097 1963 1677 1941 2003 1813 2012 1912 2084 2080 2118 2150
## 1974 1608 1503 1548 1382 1731 1798 1779 1887 2004 2077 2092 2051
## 1975 1577 1356 1652 1382 1519 1421 1442 1543 1656 1561 1905 2199
## 1976 1473 1655 1407 1395 1530 1309 1526 1327 1627 1748 1958 2274
## 1977 1648 1401 1411 1403 1394 1520 1528 1643 1515 1685 2000 2215
## 1978 1956 1462 1563 1459 1446 1622 1657 1638 1643 1683 2050 2262
## 1979 1813 1445 1762 1461 1556 1431 1427 1554 1645 1653 2016 2207
## 1980 1665 1361 1506 1360 1453 1522 1460 1552 1548 1827 1737 1941
## 1981 1474 1458 1542 1404 1522 1385 1641 1510 1681 1938 1868 1726
## 1982 1456 1445 1456 1365 1487 1558 1488 1684 1594 1850 1998 2079
## 1983 1494 1057 1218 1168 1236 1076 1174 1139 1427 1487 1483 1513
## 1984 1357 1165 1282 1110 1297 1185 1222 1284 1444 1575 1737 1763
```

Source - Harvey A.C. (1989). Forecasting, Structural Time Series Models and the Kalman Filter. Cambridge University Press, pp. 519-523.

Durbin, J. And Koopman, S. J. (2001). Time Series Analysis by State Space Methods. Oxford University Press. http://www.ssfpack.com/dkbook/

# Loading the Data

## Creating a Data-Frame

- "Seatbelts is more information on the same problem." - R Documentation
- This data-frame shows the previously mentioned variables involved in the study as column vectors with corresponding monthly dates for each data entry.

```
Seatbelts
```

```
##          DriversKilled drivers front rear   kms PetrolPrice VanKilled law
## Jan 1969           107    1687   867  269  9059  0.10297181        12   0
## Feb 1969            97    1508   825  265  7685  0.10236300         6   0
## Mar 1969           102    1507   806  319  9963  0.10206249        12   0
## Apr 1969            87    1385   814  407 10955  0.10087330         8   0
## May 1969           119    1632   991  454 11823  0.10101967        10   0
## Jun 1969           106    1511   945  427 12391  0.10058119        13   0
## Jul 1969           110    1559  1004  522 13460  0.10377398        11   0
## Aug 1969           106    1630  1091  536 14055  0.10407640         6   0
## Sep 1969           107    1579   958  405 12106  0.10377398        10   0
## Oct 1969           134    1653   850  437 11372  0.10302640        16   0
## Nov 1969           147    2152  1109  434  9834  0.10273011        13   0
## Dec 1969           180    2148  1113  437  9267  0.10199719        14   0
## Jan 1970           125    1752   925  316  9130  0.10127456        14   0
## Feb 1970           134    1765   903  311  8933  0.10070398         6   0
## Mar 1970           110    1717  1006  351 11000  0.10013961         8   0
## Apr 1970           102    1558   892  362 10733  0.09862110        11   0
## May 1970           103    1575   990  486 12912  0.09834929         7   0
## Jun 1970           111    1520   866  429 12926  0.09808018        13   0
## Jul 1970           120    1805  1095  551 13990  0.09727921        13   0
## Aug 1970           129    1800  1204  646 14926  0.09741062        11   0
## Sep 1970           122    1719  1029  456 12900  0.09742524        11   0
## Oct 1970           183    2008  1147  475 12034  0.09638063        14   0
## Nov 1970           169    2242  1171  456 10643  0.09573896        16   0
## Dec 1970           190    2478  1299  468 10742  0.09510631        14   0
## Jan 1971           134    2030   944  356 10266  0.09673597        17   0
## Feb 1971           108    1655   874  271 10281  0.09610922        16   0
## Mar 1971           104    1693   840  354 11527  0.09536725        15   0
```

# Loading the Data

## Creating a Data-Frame continued

```r
#Create a data frame from Seatbelts
sbelts <- data.frame(Seatbelts)

sbelts
```

```
##    DriversKilled drivers front rear   kms PetrolPrice VanKilled law
## 1            107    1687   867  269  9059  0.10297181        12   0
## 2             97    1508   825  265  7685  0.10236300         6   0
## 3            102    1507   806  319  9963  0.10206249        12   0
## 4             87    1385   814  407 10955  0.10087330         8   0
## 5            119    1632   991  454 11823  0.10101967        10   0
## 6            106    1511   945  427 12391  0.10058119        13   0
## 7            110    1559  1004  522 13460  0.10377398        11   0
## 8            106    1630  1091  536 14055  0.10407640         6   0
## 9            107    1579   958  405 12106  0.10377398        10   0
## 10           134    1653   850  437 11372  0.10302640        16   0
## 11           147    2152  1109  434  9834  0.10273011        13   0
## 12           180    2148  1113  437  9267  0.10199719        14   0
## 13           125    1752   925  316  9130  0.10127456        14   0
## 14           134    1765   903  311  8933  0.10070398         6   0
## 15           110    1717  1006  351 11000  0.10013961         8   0
## 16           102    1558   892  362 10733  0.09862110        11   0
## 17           103    1575   990  486 12912  0.09834929         7   0
## 18           111    1520   866  429 12926  0.09808018        13   0
## 19           120    1805  1095  551 13990  0.09727921        13   0
## 20           129    1800  1204  646 14926  0.09741062        11   0
## 21           122    1719  1029  456 12900  0.09742524        11   0
## 22           183    2008  1147  475 12034  0.09638063        14   0
## 23           169    2242  1171  456 10643  0.09573896        16   0
```

- I decided to use the previous data-frame to create a duplicate called “sbelts” that was easier for me to work with for parsing.
- The difference between “Seatbelts” and “sbelts” is that the row titles have been changed from a month/year date format in the former to indexed values in the latter.

# Parsing the Data

- I parsed the “sbelts” data-frame and created two subsets:
  - “DnoLaw” shows the data BEFORE the seat belt law was introduced. This is indicated in the “law” column where 0 means that the law was not yet passed.
  - “DLaw” shows the data AFTER the seat belt law was introduced. This is indicated in the “law” column where 1 means that the law had been passed.

```
DnoLaw <- sbelts %>% filter(law == 0)
DLaw <- sbelts %>% filter(law == 1)

DnoLaw
```

```
##      DriversKilled drivers front rear   kms PetrolPrice VanKilled law
## 1              107    1687   867  269  9059  0.10297181        12   0
## 2               97    1508   825  265  7685  0.10236300         6   0
## 3              102    1507   806  319  9963  0.10206249        12   0
## 4               87    1385   814  407 10955  0.10087330         8   0
## 5              119    1632   991  454 11823  0.10101967        10   0
## 6              106    1511   945  427 12391  0.10058119        13   0
## 7              110    1559  1004  522 13460  0.10377398        11   0
## 8              106    1630  1091  536 14055  0.10407640         6   0
## 9              107    1579   958  405 12106  0.10377398        10   0
## 10             134    1653   850  437 11372  0.10302640        16   0
```

```
DLaw
```

```
##      DriversKilled drivers front rear   kms PetrolPrice VanKilled law
## 1               95    1057   426  300 15511   0.1136570         3   1
## 2              100    1218   475  318 18308   0.1131444         2   1
## 3               89    1168   556  391 17793   0.1184955         6   1
## 4               82    1236   559  398 19205   0.1179694         3   1
## 5               89    1076   483  337 19162   0.1176866         7   1
## 6               60    1174   587  477 20997   0.1200592         6   1
## 7               84    1139   615  422 20705   0.1194377         8   1
## 8              113    1427   618  495 18759   0.1188813         8   1
## 9              126    1487   662  471 19240   0.1184624         4   1
## 10             122    1483   519  368 17504   0.1180166         3   1
## 11             118    1513   585  345 16591   0.1177066         5   1
## 12              92    1357   483  296 16224   0.1177761         5   1
## 13              86    1165   434  319 16670   0.1147970         3   1
```

```r
cat(" Overall Mean: \n", mean(sbelts$drivers),
    "\n Mean (Pre-Law): \n", mean(DnoLaw$drivers),
    "\n Mean (Post-Law): \n", mean(DLaw$drivers),

    "\n\n Mean - Driver Deaths (Pre-Law): \n", mean(DnoLaw$DriversKilled),
    "\n Mean - Driver Deaths (Post-Law): \n", mean(DLaw$DriversKilled),

    "\n\n Mean - Front Seat Passenger Deaths/Serious Injuries (Pre-Law): \n", mean(DnoLaw$front),
    "\n Mean - Front Seat Passenger Deaths/Serious Injuries (Post-Law): \n", mean(DLaw$front),

    "\n\n Mean - Rear Seat Passenger Deaths/Serious Injuries (Pre-Law): \n", mean(DnoLaw$rear),
    "\n Mean - Rear Seat Passenger Deaths/Serious Injuries (Post-Law): \n", mean(DLaw$rear))
```

```
##  Overall Mean:
##  1670.307
##  Mean (Pre-Law):
##  1717.751
##  Mean (Post-Law):
##  1321.696
##
##  Mean - Driver Deaths (Pre-Law):
##  125.8698
##  Mean - Driver Deaths (Post-Law):
##  100.2609
##
##  Mean - Front Seat Passenger Deaths/Serious Injuries (Pre-Law):
##  873.4556
##  Mean - Front Seat Passenger Deaths/Serious Injuries (Post-Law):
##  570.9565
##
##  Mean - Rear Seat Passenger Deaths/Serious Injuries (Pre-Law):
##  400.3195
##  Mean - Rear Seat Passenger Deaths/Serious Injuries (Post-Law):
##  407.7391
```

# Comparing Averages

- We see that there is a significant reduction in the average amount of deaths after the seat belt law was introduced.
- The mean for ALL deaths before the introduction of the law was 1717.751. After the law, the mean decreased to 1321.696.
- Parsing the data, we see that the mean for DRIVER deaths before the introduction of the law was 125.8698. After the law, the mean decreased to 100.2609.
- Parsing the data further, we can also compare the average FRONT SEAT passenger deaths/serious injuries. We see a significant reduction from 873.4556 to 570.9565.
- However, the average REAR SEAT passenger deaths/serious injuries increased from 400.3195 to 407.7391. When compared to the reduction in overall or front seat deaths, this increase is negligible.

# Comparing Variability

- We also see that the variability of road death data had reduced significantly post-law. This indicates more consistency in the data.

- The standard deviation for front seat passenger deaths decreased significantly and the standard deviation for rear seat passenger deaths decrease less so.

```
cat(" Overall Standard Deviation: \n", sd(sbelts$drivers),
    "\n Standard Deviation (Pre-Law): \n", sd(DnoLaw$drivers),
    "\n Standard Deviation (Post-Law): \n",sd(DLaw$drivers),

    "\n\n Standard Deviation - Driver Deaths (Pre-Law): \n", sd(DnoLaw$DriversKilled),
    "\n Standard Deviation - Driver Deaths (Post-Law): \n",sd(DLaw$DriversKilled),

    "\n\n Standard Deviation - Front (Pre-Law):\n", sd(DnoLaw$front),
    "\n Standard Deviation - Front (Post-Law):\n", sd(DLaw$front),

    "\n\n Standard Deviation - Rear (Pre-Law): \n", sd(DnoLaw$rear),
    "\n Standard Deviation - Rear (Post-Law): \n", sd(DLaw$rear))
```

```
##  Overall Standard Deviation:
##  289.611
##  Standard Deviation (Pre-Law):
##  266.892
##  Standard Deviation (Post-Law):
##  199.7233
##
##  Standard Deviation - Driver Deaths (Pre-Law):
##  24.26088
##  Standard Deviation - Driver Deaths (Post-Law):
##  22.2286
##
##  Standard Deviation - Front (Pre-Law):
##  151.5416
##  Standard Deviation - Front (Post-Law):
##  81.29099
##
##  Standard Deviation - Rear (Pre-Law):
##  84.88012
##  Standard Deviation - Rear (Post-Law):
##  69.91828
```

# Comparing Variability continued

- This reduction in variation is also evidenced by the range of road deaths in each scenario.  After the law, the maximum and minimum road deaths decreased by a good amount each.

```r
cat(" Maximum (Pre-Law):", max(range(DnoLaw$drivers)),
    "\n Maximum (Post-Law)", max(range(DLaw$drivers)),

    "\n\n Minimum (Pre-Law):", min(range(DnoLaw$drivers)),
    "\n Minimum (Post-Law)", min(range(DLaw$drivers)),

    "\n\n Range (Pre-Law):", max(range(DnoLaw$drivers))-min(range(DnoLaw$drivers)),
    "\n Range (Post-Law)", max(range(DLaw$drivers))-min(range(DLaw$drivers)))
```

```
##  Maximum (Pre-Law): 2654
##  Maximum (Post-Law) 1763
##
##  Minimum (Pre-Law): 1309
##  Minimum (Post-Law) 1057
##
##  Range (Pre-Law): 1345
##  Range (Post-Law) 706
```

# Rate of Fewer Deaths

```r
reduc_rate <- (mean(DnoLaw$drivers)-mean(DLaw$drivers))/mean(DnoLaw$drivers)
reduc_drivers <- (mean(DnoLaw$DriversKilled)-mean(DLaw$DriversKilled))/mean(DnoLaw$DriversKilled)
reduc_front <- (mean(DnoLaw$front)-mean(DLaw$front))/mean(DnoLaw$front)
reduc_rear <- (mean(DnoLaw$rear)-mean(DLaw$rear))/mean(DnoLaw$rear)

cat(" Average Total Death Reduction Rate: ", "\n", reduc_rate,
    "\n\n Average Reduction Rate in Driver Deaths: \n", reduc_drivers,
    "\n\n Average Reduction Rate in Front Seat Deaths: \n", reduc_front,
    "\n\n Average Reduction Rate in Rear Seat Deaths: \n", reduc_rear)
```

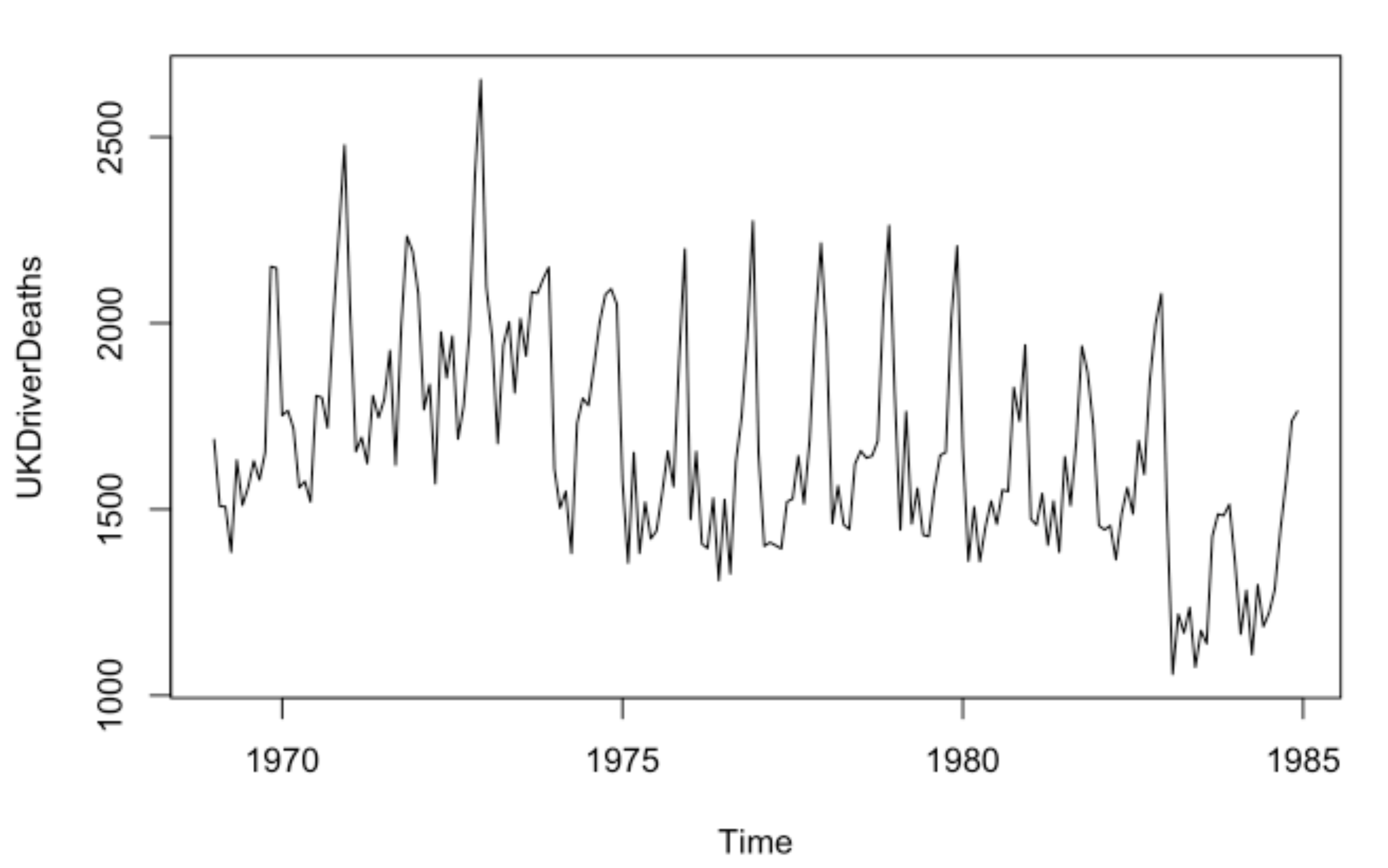
- We can compare the average reduction in deaths to the average number of deaths prior to the law's introduction using a ratio:

$$Reduction\ Rate = \frac{Mean\ Deaths\ (Before\ Law) - Mean\ Deaths\ (After\ Law)}{Mean\ Deaths\ (Before\ Law)}$$

```
##  Average Total Death Reduction Rate:
##  0.2305664
##
##  Average Reduction Rate in Driver Deaths:
##  0.2034559
##
##  Average Reduction Rate in Front Seat Deaths:
##  0.3463245
##
##  Average Reduction Rate in Rear Seat Deaths:
##  -0.0185342
```

# Visualizing the Data

```
plot.ts(UKDriverDeaths)
```

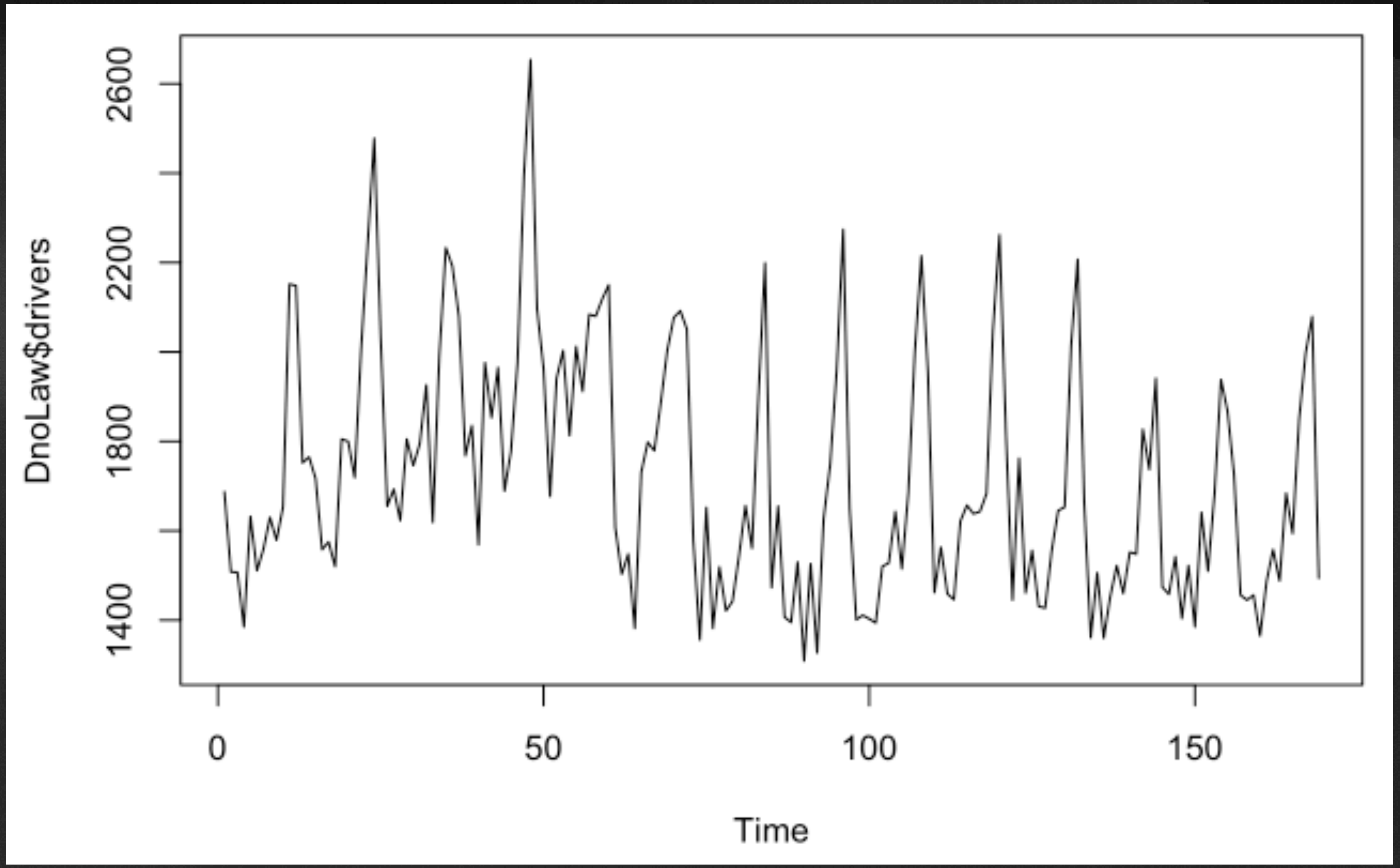



- In the first graph to the left, we see the visualized time-series data on total road casualties.
- It shows the entire timeline of data with Time (in years) on the horizontal axis.

# Visualizing the Data

```
plot.ts(DnoLaw$drivers)
```

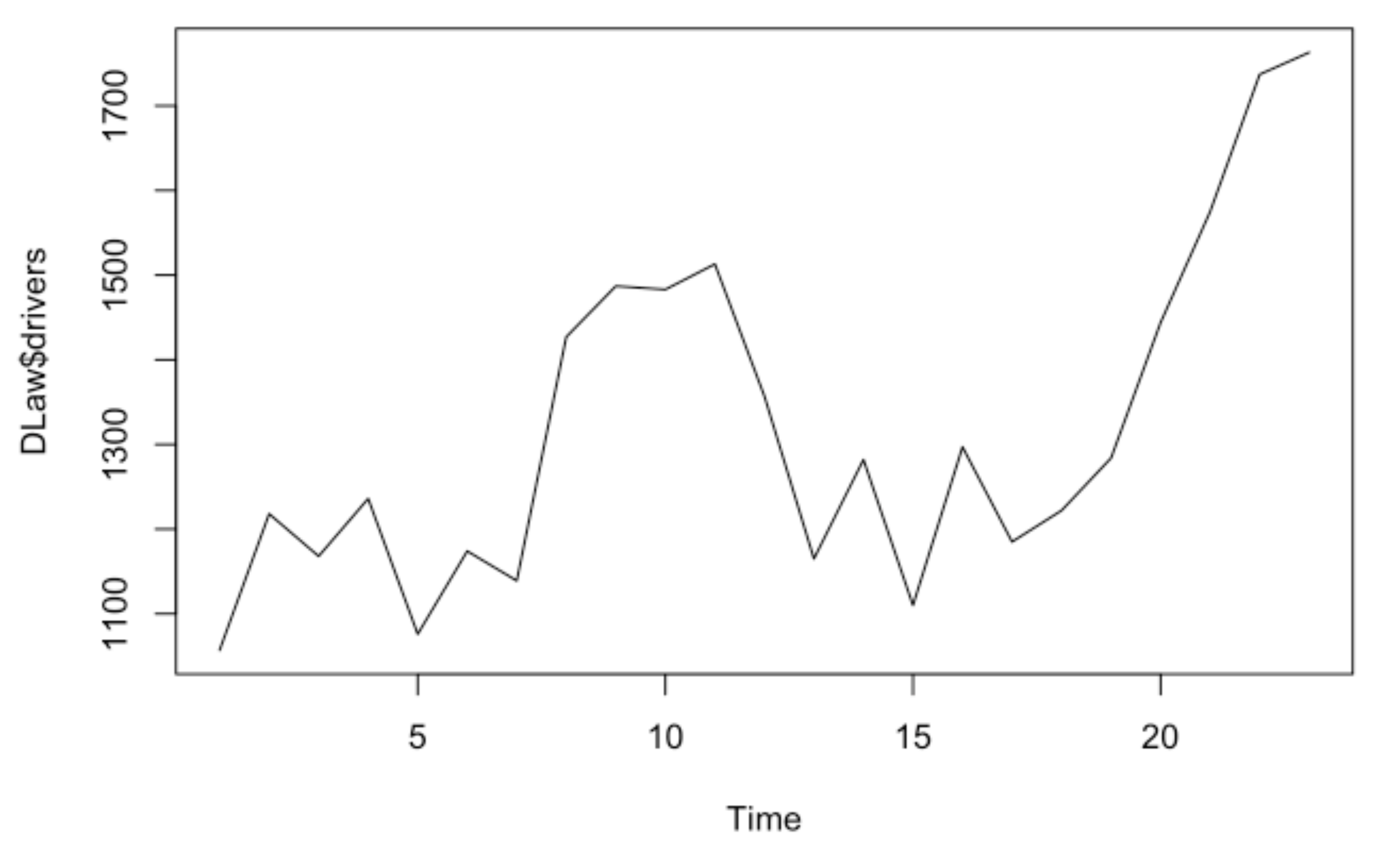
- The second graph shows the timeline of data BEFORE the law was enacted (the first 169 months), with Time (in months SINCE THE START OF THE STUDY) on the horizontal axis.



# Visualizing the Data

```
plot.ts(DLaw$drivers)
```

- The third graph shows the timeline of data AFTER the law was enacted (the last 23 months), with Time (in months SINCE ENACTMENT OF THE LAW) on the horizontal axis.



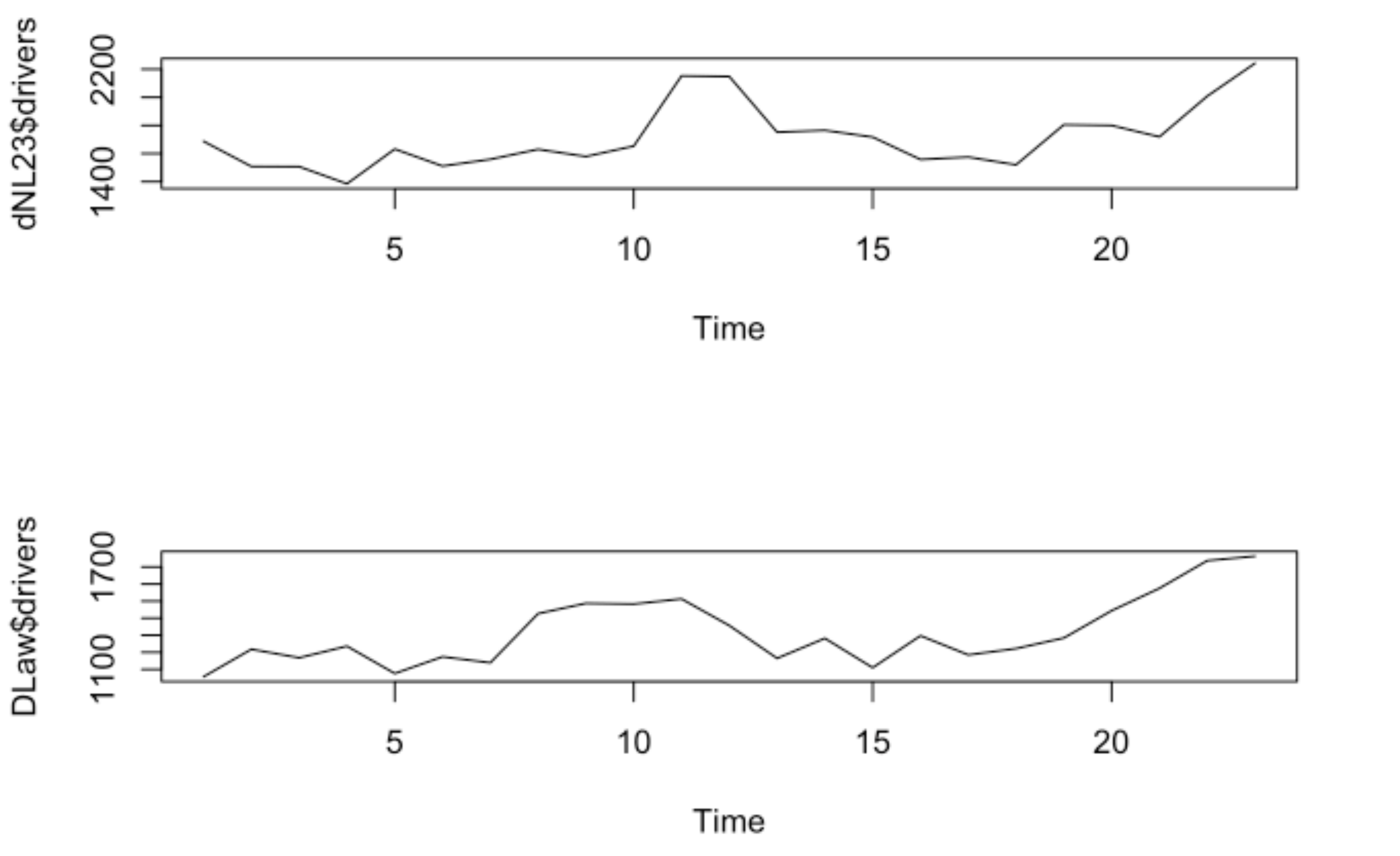
# Visualizing the Data

- Since we're limited to 23 months worth of data after the law was enacted, for the following graphs I chose to zoom in on four particular 22-26 month time periods before the enactment of the law. I did this in hopes of having a more comparable amount of data points, being sure to include the lowest dips in road deaths for each period. I paired each 22-26 month pre-law period with the 23 month post-law for comparison.

- We can visually see that the scaling for the number of deaths, regardless of time frame, decreases significantly after the passing of the seat belt law.

**First 23 Months (Pre-Law) vs Final 23 Months (Post-Law)**

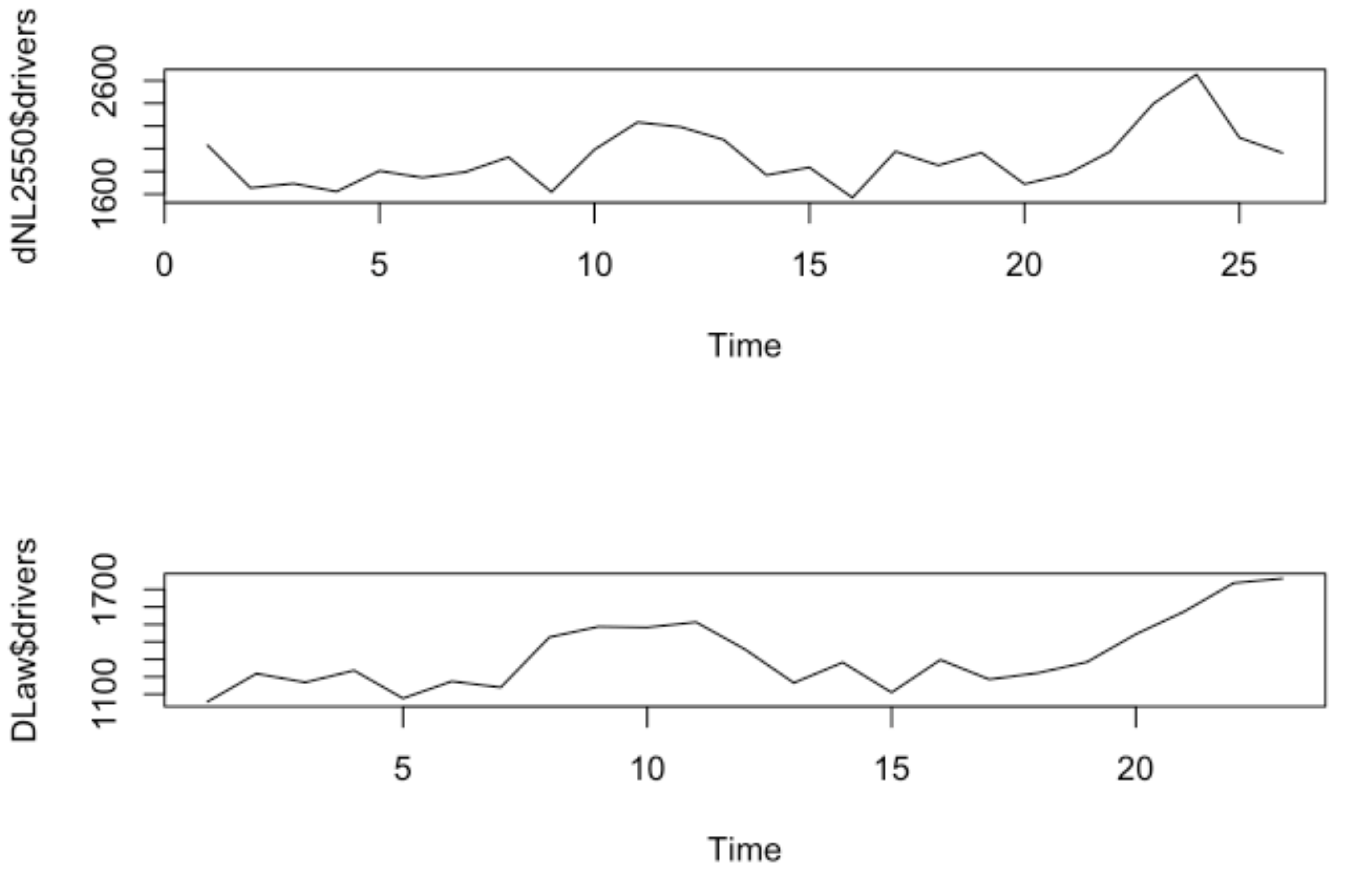
```
#first 23 months
dNL23 <- DnoLaw %>% slice(1:23)
par(mfrow=c(2,1))
plot.ts(dNL23$drivers)
plot.ts(DLaw$drivers)
```



# Visualizing the Data

## Months 25-50 (Pre-Law) vs Final 23 Months (Post-Law)

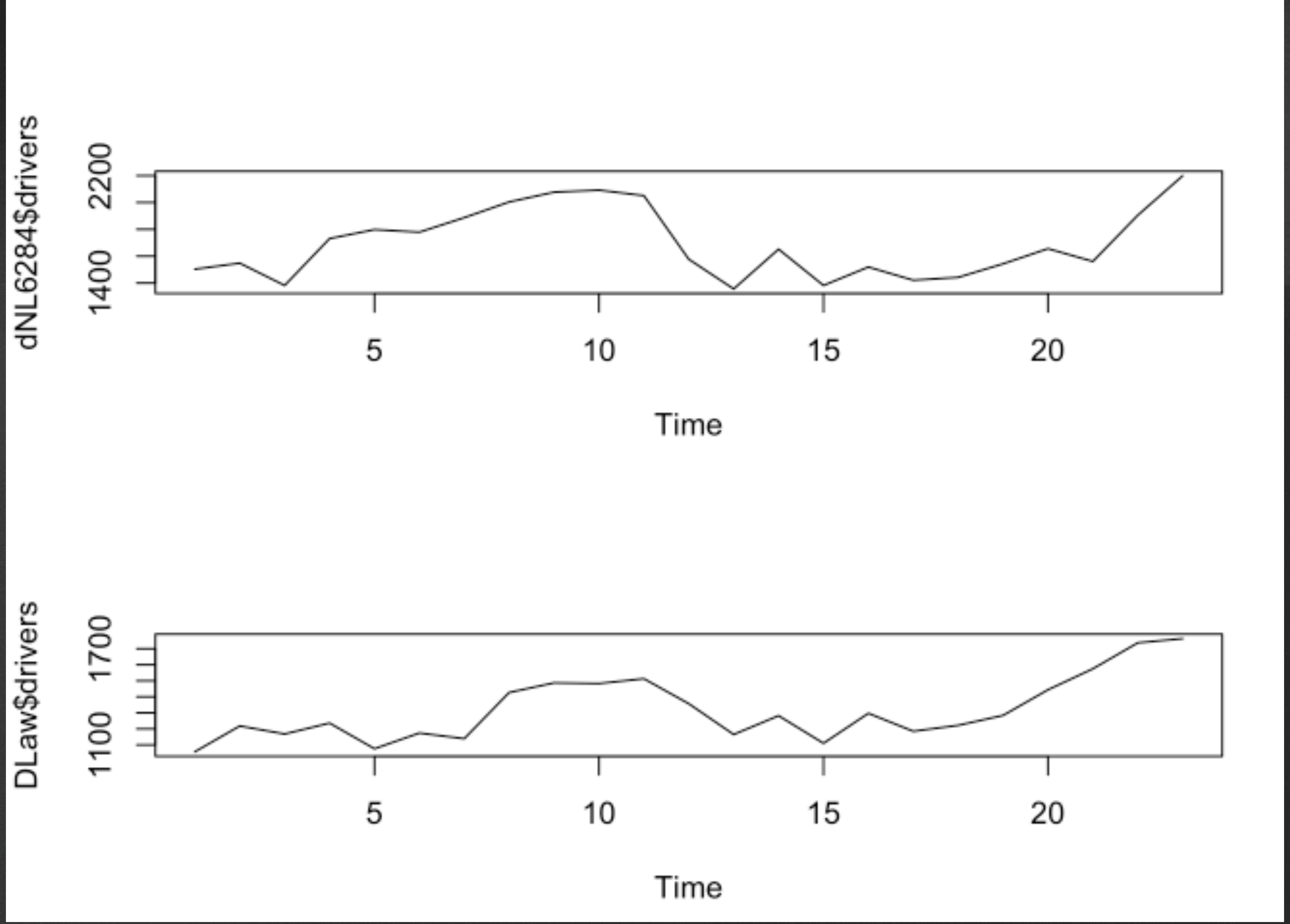
```r
#months 25-50
dNL2550 <- DnoLaw %>% slice(25:50)
par(mfrow=c(2,1))
plot.ts(dNL2550$drivers)
plot.ts(DLaw$drivers)
```



# Visualizing the Data

## Months 62-84 (Pre-Law) vs Final 23 Months (Post-Law)

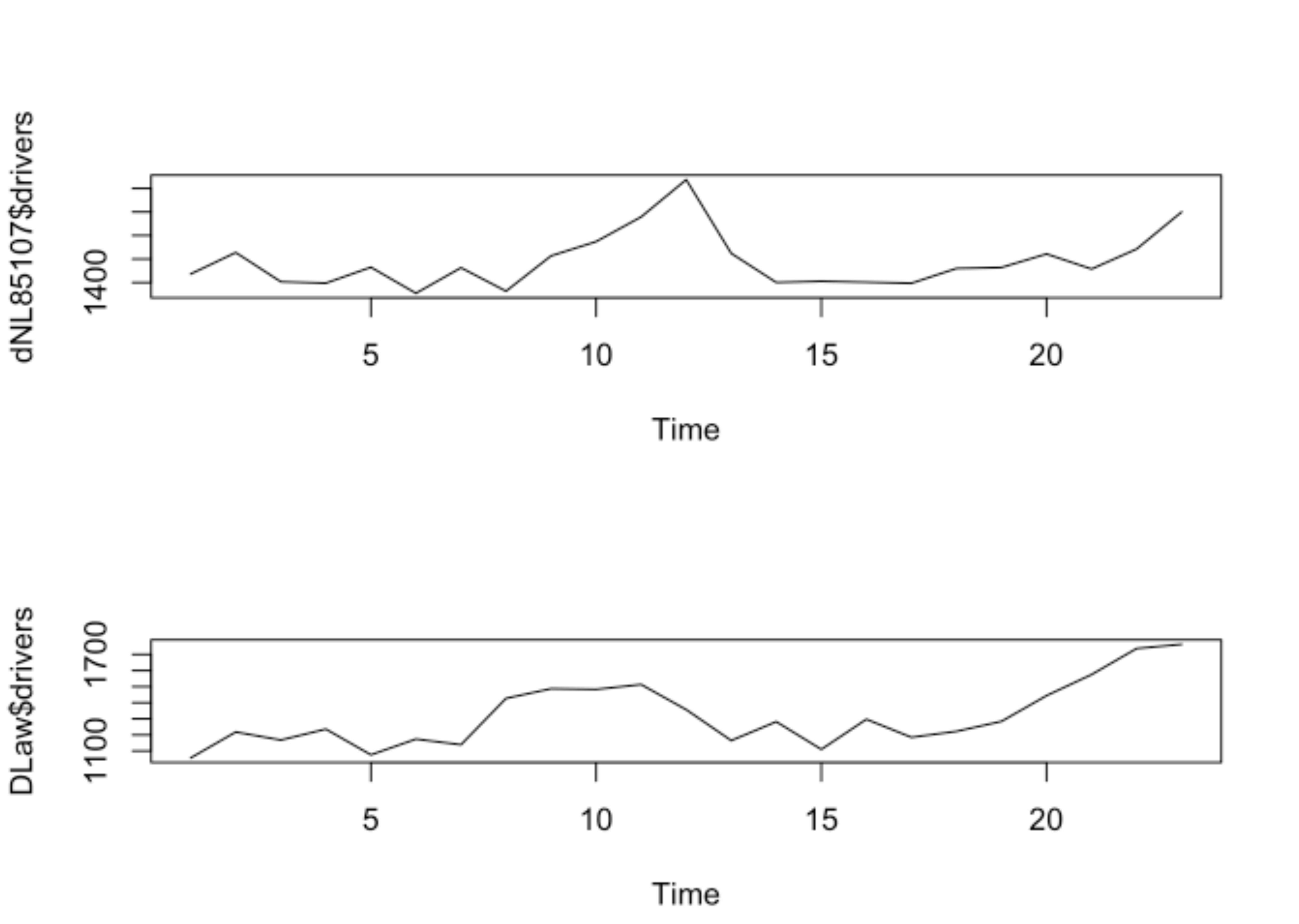
```r
#months 62-84
dNL6284 <- DnoLaw %>% slice(62:84)
par(mfrow=c(2,1))
plot.ts(dNL6284$drivers)
plot.ts(DLaw$drivers)
```



# Visualizing the Data

## Months 85-107 (Pre-Law) vs Final 23 Months (Post-Law)

```r
#months 85-107
dNL85107 <- DnoLaw %>% slice(85:107)
par(mfrow=c(2,1))
plot.ts(dNL85107$drivers)
plot.ts(DLaw$drivers)
```

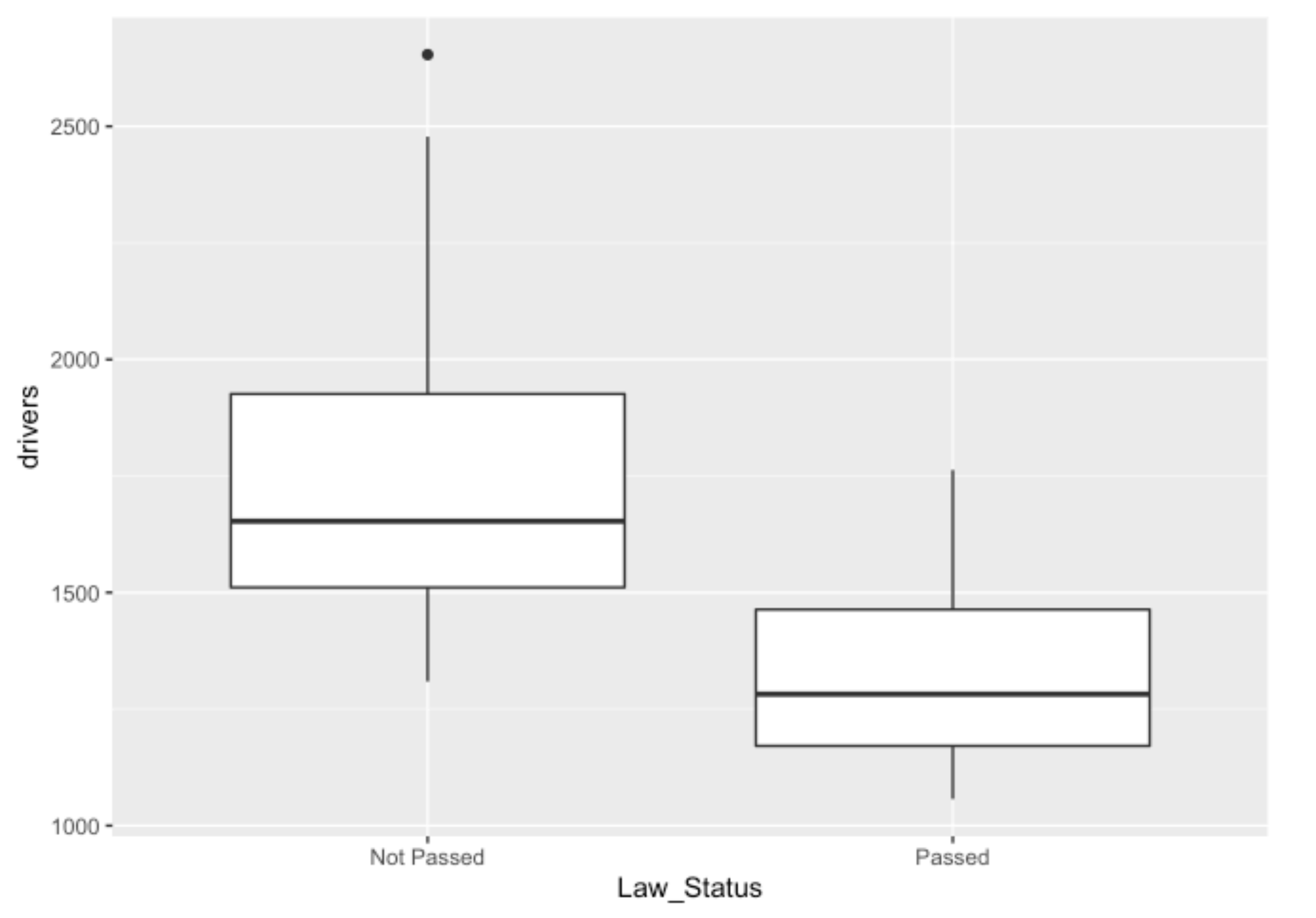


# Visualizing the Data

## Boxplots of Summary Data

```
sbelts <- sbelts %>%
  mutate(Law_Status = ifelse(sbelts$law > 0, "Passed", "Not Passed"))

p <- ggplot(sbelts, aes(x = Law_Status, y = drivers))+
geom_boxplot()
p
```

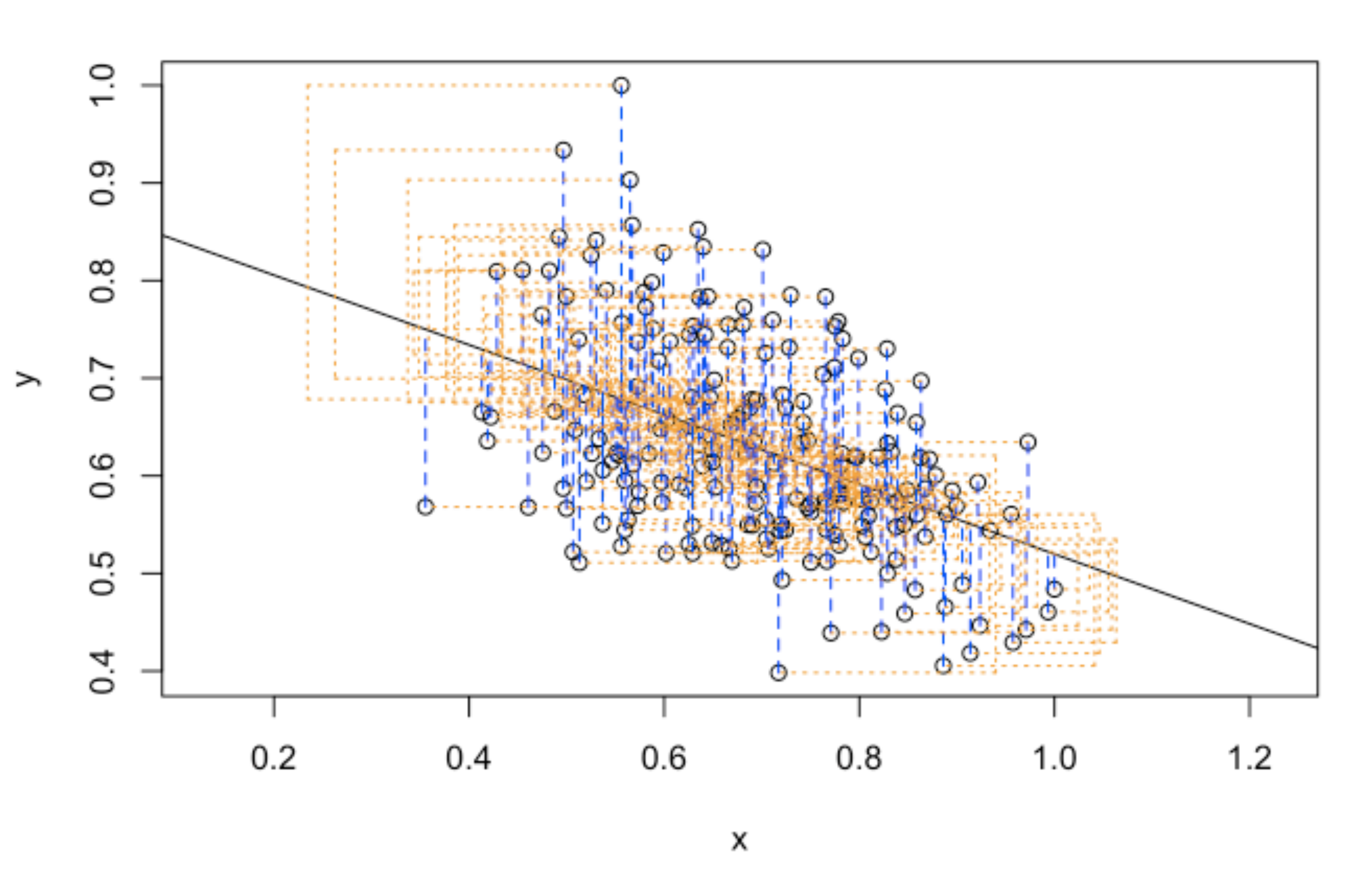



- The boxplot to the left is a comparison of the summary data pre-law and post law.
- Within the image, the plot on the left represents the data corresponding to pre-law statistics.
- The plot on the right represents data corresponding to post-law statistics.
- We see that the data before the law was passed is centered around much higher values than after it was passed. The third quartile indicator for the post-law data is smaller than the first quartile indicator for the pre-law data. This signifies to me a dramatic reduction in casualties.

```
sbelts <- sbelts %>%
  mutate(deaths_norm = sbelts$drivers/max(sbelts$drivers))
sbelts <- sbelts %>%
  mutate(kms_norm = sbelts$kms/max(sbelts$kms))
sbelts <- sbelts %>%
  mutate(p_norm = sbelts$PetrolPrice/max(sbelts$PetrolPrice))

DATA606::plot_ss(x = sbelts$kms_norm, y = sbelts$deaths_norm, showSquares =TRUE)
```

## Road Casualties vs Distance Traveled Scatterplot




```
cor(sbelts$kms, sbelts$drivers)
```

```
## [1] -0.4447631
```

# Comparison of Other Variables Using Scatterplots

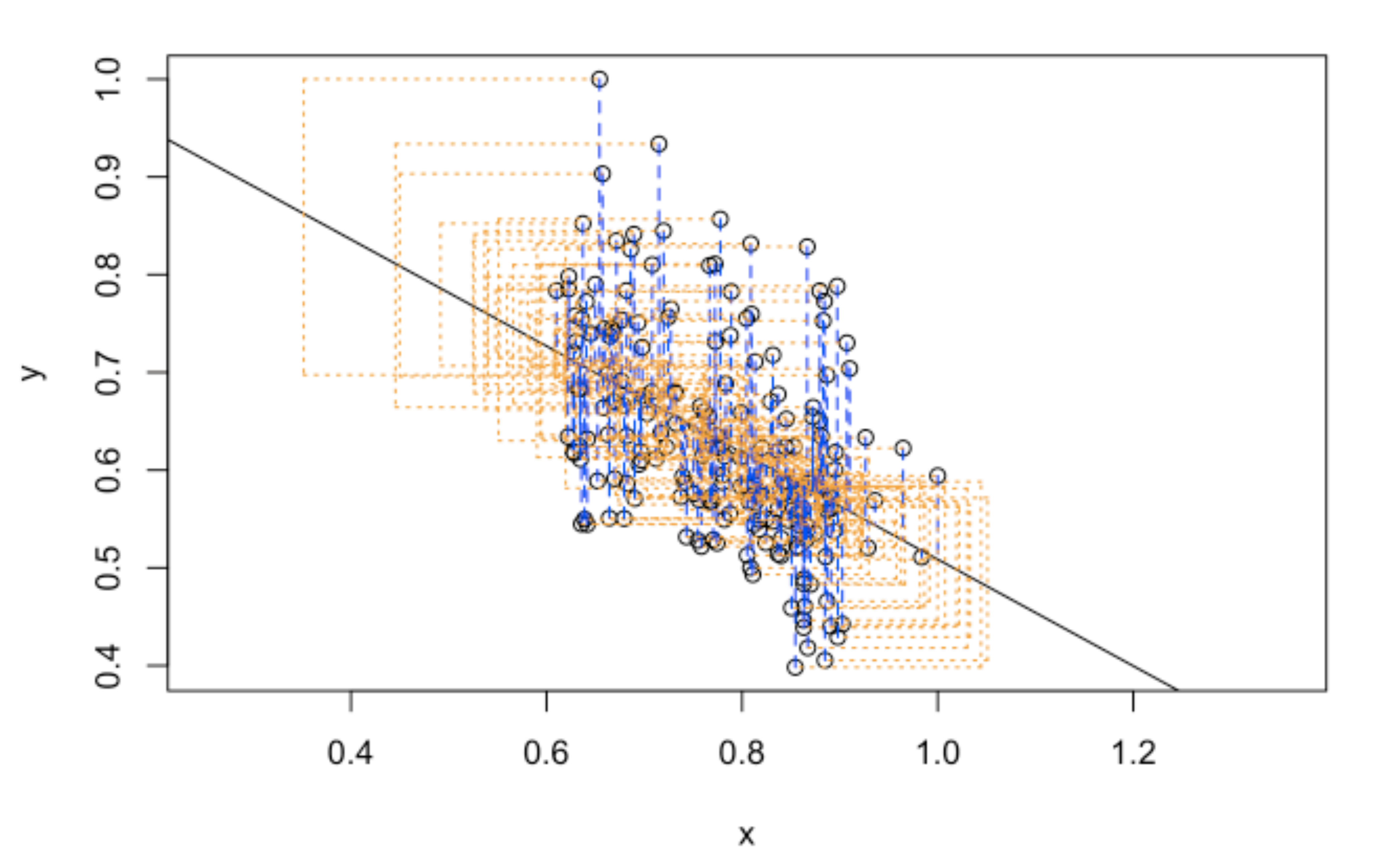
- In addition to casualty count, the data set also included information on distance traveled and price of petroleum.
- There was another variable called "VanKilled" which represents the number of van ('light goods vehicle') drivers killed, but I left that out of my analysis as it would not give any additional information that the other variables already give.
- I used scatterplots to visualize the correlation between these variables and road deaths. I also normalized the data to get a better visual. There seems to be a linear correlation between both distance traveled vs road deaths and petroleum price vs road deaths, using the cor() function gives us the correlation coefficient for the Least Squares Regression line of each.
- We see that the values for r are -0.04447631 and -0.4576675, respectively. Although there seems to be a linear relationship in each case, the values for r indicate a weak correlation. This still does not rule out their involvement in the reduction of road casualties.
- The graph to the left shows a negative linear relationship between casualties and distance traveled. In other words, more distance traveled (ie. driving experience) correlates to lower death.

# Comparison of Other Variables

```
DATA606::plot_ss(x = sbelts$p_norm, y = sbelts$deaths_norm, showSquares =TRUE)
```

- The graph to the right shows a negative linear relationship between casualties and petrol price. In other words, higher petrol prices (ie. driving experience) correlates to lower death.

## Road Casualties vs Petrol Price Scatterplot




```
cor(sbelts$PetrolPrice, sbelts$drivers)
```

```
## [1] -0.4576675
```

# Conclusion

The analysis involved in this study determined that, with regard to measures of central tendency, there was a significant reduction in the average overall number of road casualties after the passing of the seat belt law. This is consistent with the mean data for front seat passenger and driver deaths. Although the mean rear seat passenger deaths increased, it was very slight and is negligible when compared to the overall data. The mean analysis was verified with the visualization of the summary data using box plots. In my analysis of the spread of data, the standard deviation and range decreased after the passing of the law, regardless of the type of casualty. This indicates a narrowing of the data meaning more consistent values that are easier to rely on. I also used narrowed down time frames for the comparison of time series graphs to have much more comparable visualizations. In the visualizations for all comparisons, the scaling for the number of deaths decreased after the law was passed. Finally, there were two previously unused variables that were included in the study. I considered these as possible influencers in road death casualties and used scatterplots, least squares regression, and the correlation coefficient function in R Studio to determine the strength of the linear correlation. Given the values of r, I determined that there is a linear correlation between both of these variables and road casualties. Although the strengths of these correlations were weak, it may still represent a degree of influence on survival outcomes, but based on my findings, there is sufficient evidence to conclude that seat belt laws *predict* a reduction in road casualties.