

lab 8: Analysis of seatbelts data

Background

This lab worksheet is based on the "Road Casualties in Great Britain 1969-84" inbuilt R data set, which is called `Seatbelts`.

- This shows data on road accident injuries and deaths month by month, before and after a law was introduced on 31st January 1983 making seatbelt wearing for drivers and front-seat passengers compulsory.

Seatbelts were not compulsory for rear-seat passengers until much later. The data set starts in January 1969 and observations run until December 1984.

At the time the law was introduced there was a huge public debate about whether forcing people to wear seatbelts restricted people's freedom, and whether seatbelts were at all effective – many people argued that if you are wearing a seatbelt, you are more likely to be trapped in the car after an accident, and possibly burned to death.

People wrote to the newspapers describing incidents when (they claimed) if they had been wearing a seatbelt, they would certainly have died.

These fears turned out to be groundless. Years later, it became so widely accepted that seatbelts are a good idea that a typical scene in action movies was that the when the hero or heroine was driving in a car with the villain, they would intelligently fasten their seatbelt and deliberately crash the car, ensuring that they survived and the stupid villain was injured because he had not fastened his seatbelt....

Load the data:

```
#install.packages("datasets")
library(datasets)
data() # this shows all the datasets available.
summary(Seatbelts)

## DriversKilled      drivers      front      rear
## Min.   : 60.0   Min.   :1057   Min.   : 426.0   Min.   : 224.0
## 1st Qu.:104.8   1st Qu.:1462   1st Qu.: 715.5   1st Qu.:344.8
## Median :118.5   Median :1631   Median : 828.5   Median :401.5
## Mean   :122.8   Mean   :1670   Mean   : 837.2   Mean   :401.2
## 3rd Qu.:138.0   3rd Qu.:1851   3rd Qu.: 950.8   3rd Qu.:456.2
## Max.   :198.0   Max.   :2654   Max.   :1299.0   Max.   :646.0
## kms      PetrolPrice   VanKilled      law
## Min.   : 7685   Min.   : 0.08118   Min.   : 2.000   Min.   : 0.0000
## 1st Qu.:12085   1st Qu.: 0.09258   1st Qu.: 6.000   1st Qu.: 0.0000
## Median :14087   Median : 0.10448   Median : 8.000   Median : 0.0000
## Mean   :14994   Mean   : 0.10362   Mean   : 9.057   Mean   : 0.1198
## 3rd Qu.:17203   3rd Qu.: 0.11486   3rd Qu.:12.000   3rd Qu.: 0.0000
## Max.   :21626   Max.   : 0.13303   Max.   :17.000   Max.   : 1.0000

names(Seatbelts)

## NULL

str(Seatbelts)

## Time-Series [1:192, 1:8] from 1969 to 1985: 107 97 102 87 119 106 110 106 107 134 ...
## - attr(*, "dimnames")=List of 2
## ..$ : chr [1:8] "DriversKilled" "drivers" "front" "rear" ...

head(Seatbelts, 10)
```

	DriversKilled	drivers	front	rear	kms	PetrolPrice	VanKilled	law
## [1,]	107	1687	867	269	9059	0.1029718	12	0
## [2,]	97	1508	825	265	7685	0.1023630	6	0
## [3,]	102	1507	806	319	9963	0.1020625	12	0
## [4,]	87	1385	814	407	10955	0.1008733	8	0
## [5,]	119	1632	991	454	11823	0.1010197	10	0
## [6,]	106	1511	945	427	12391	0.1005812	13	0
## [7,]	110	1559	1004	522	13460	0.1037740	11	0
## [8,]	106	1630	1091	536	14055	0.1040764	6	0
## [9,]	107	1570	950	405	12106	0.1037740	10	0
## [10,]	134	1653	850	437	11372	0.1030264	16	0

Oops – this does not seem to be a data frame, but we can convert it into one.

```
seatbelts <- as.data.frame(Seatbelts)
str(seatbelts)

## 'data.frame':   192 obs. of  8 variables:
## $ DriversKilled: num  107 97 102 87 119 106 110 106 107 134 ...
## $ drivers      : num  1687 1508 1507 1385 1632 ...
## $ front        : num  867 825 806 814 995 ...
## $ rear         : num  269 265 319 407 454 427 522 536 405 437 ...
## $ kms          : num  9059 7685 9963 10955 11823 ...
## $ PetrolPrice  : num  0.103 0.102 0.102 0.101 0.101 ...
## $ VanKilled    : num  12 6 12 8 10 13 11 6 10 16 ...
## $ law          : num  0 0 0 0 0 0 0 0 0 0 ...

names(seatbelts)

## [1] "DriversKilled" "drivers"      "front"      "rear"
## [5] "kms"          "PetrolPrice" "VanKilled"  "law"

head(seatbelts, 10)
```

	DriversKilled	drivers	front	rear	kms	PetrolPrice	VanKilled	law
## 1	107	1687	867	269	9059	0.1029718	12	0
## 2	97	1508	825	265	7685	0.1023630	6	0
## 3	102	1507	806	319	9963	0.1020625	12	0
## 4	87	1385	814	407	10955	0.1008733	8	0
## 5	119	1632	991	454	11823	0.1010197	10	0
## 6	106	1511	945	427	12391	0.1005812	13	0
## 7	110	1559	1004	522	13460	0.1037740	11	0
## 8	106	1630	1091	536	14055	0.1040764	6	0
## 9	107	1570	950	405	12106	0.1037740	10	0
## 10	134	1653	850	437	11372	0.1030264	16	0

Good. Now we have a data frame.

The natural question to ask is whether there was a reduction in the number of serious injuries and deaths of drivers and front seat passengers when the seat belt law was introduced.

There are 8 variables in the `dataset`. The variables are:

1. `DriversKilled`, `VanKilled` - the number of drivers killed each month.
2. `drivers`, `front`, `rear` – the numbers of drivers, front-seat passengers, and rear-seat passengers killed or severely injured each month.
3. `PetrolPrice` - petrol prices.
4. `kms` - a measure of how much driving there was in the country that month (one might assume that the number of accidents varies with the amount of driving) - something to investigate.
5. `law` - 0 for each month before the law was introduced, 1 for each month afterwards.

Now what relationships do you expect?

Plot a line graph of drivers vs time. (Use a vector of row-numbers for time - in

- this lab we will not spend time on representing dates, but you can if you want to.)

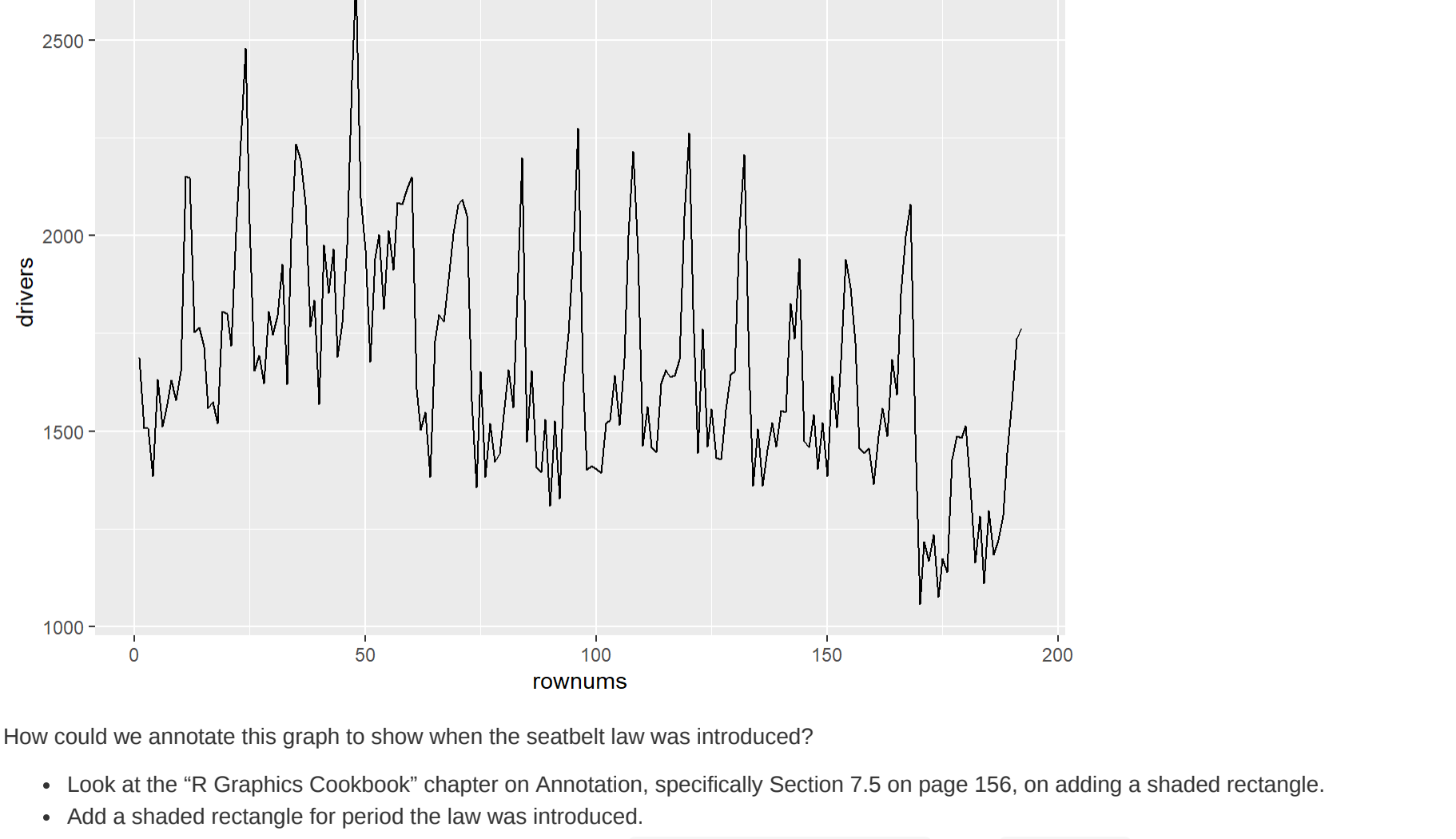
```
rownums <- 1:nrow(seatbelts)

Your basic plot (which you should vary and elaborate in many ways) is:
```

```
#install.packages("ggplot2")
library(ggplot2)

## Warning: package 'ggplot2' was built under R version 4.0.5

p <- ggplot( seatbelts, aes(x=rownums, y=drivers) ) + geom_line()
p
```



How could we annotate this graph to show when the seatbelt law was introduced?

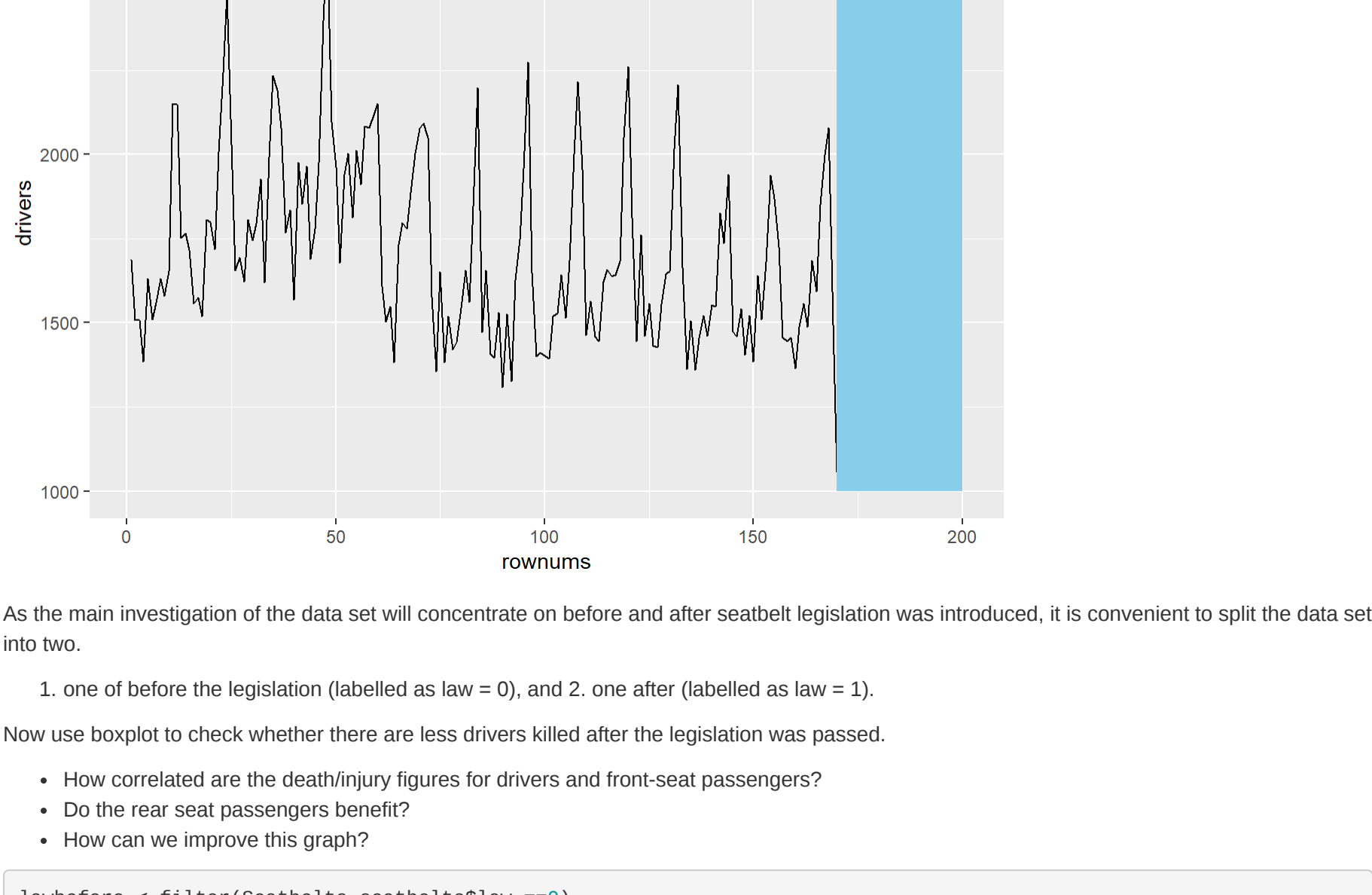
- Look at the "R Graphics Cookbook" chapter on Annotation, specifically Section 7.5 on page 156, on adding a shaded rectangle.
- Add a shaded rectangle for period the law was introduced.
- (You may find the row number when the law starts using `which.max(seatbelts$law)`. Type `?which.max` to find out what `which.max` does.)

```
which.max(seatbelts$law)

## [1] 170

#?which.max

p <- ggplot( seatbelts, aes(x=rownums, y=drivers) ) + geom_line() + annotate("rect", xmin =170, xmax =200, ymin =1000, ymax =2500, fill = "skyblue") + ggtitle("Plot of drivers vs time, and a shaded rectangle for period the law was introduced")
p
```



As the main investigation of the data set will concentrate on before and after seatbelt legislation was introduced, it is convenient to split the data set into two.

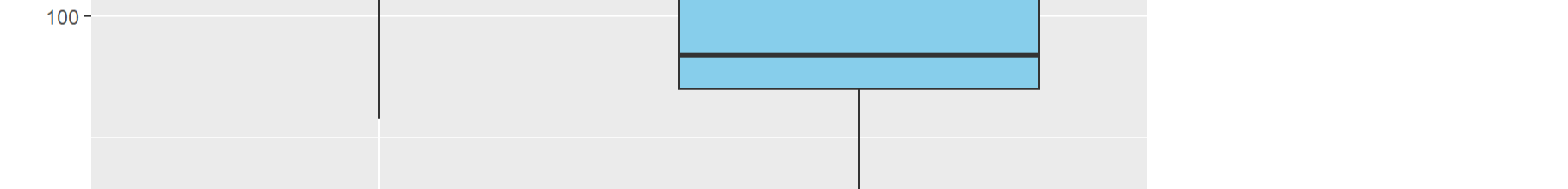
1. one of before the legislation (labelled as law = 0), and 2. one after (labelled as law = 1).

Now use `boxplot` to check whether there are less drivers killed after the legislation was passed.

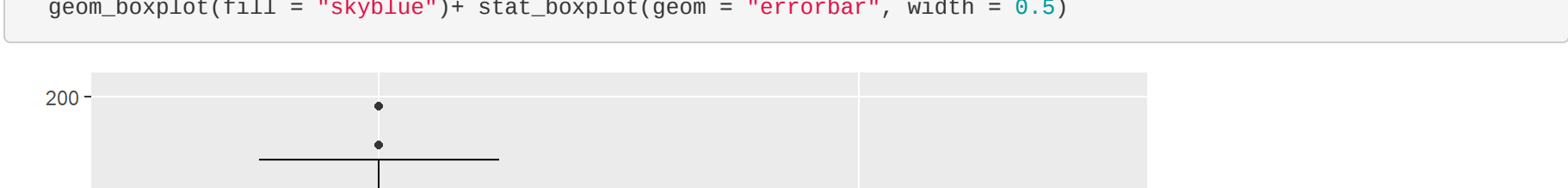
- How correlated are the death/injury figures for drivers and front-seat passengers?
- Do the rear seat passengers benefit?
- How can we improve this graph?

```
lawBefore <- filter(seatbelts, seatbelts$law ==0)
lawAfter <- filter(seatbelts, seatbelts$law==1)
```

```
ggplot(seatbelts, aes(x=factor(law), y =DriversKilled)) +
  geom_boxplot(fill = "skyblue")
```



```
ggplot(seatbelts, aes(x=factor(law), y =DriversKilled)) +
  geom_boxplot(fill = "skyblue")+ stat_boxplot(geom = "errorbar", width = 0.5)
```



```
ggplot(seatbelts, aes(x=factor(law), y =DriversKilled)) +
  geom_boxplot(fill = "skyblue")+ stat_boxplot(geom = "errorbar", width = 0.5)+
  theme_grey()+ylab ("Monthly Driver Mortality or badly injured")+
  xlab("Before(0) and after(1) seatbelt law introduced")+
  ggtitle("Boxplot of before and after seatbelt legislation was introduced")
```

