

## Marking of homework questions - example

```
library(dplyr)
library(ggplot2)
library(lubridate)
Ambulance <- read.csv("Ambulance.csv")
Ambulance <- Ambulance %>% mutate( Call=ymd_hms(Call) )
Hospitals <- read.csv("Hospital.csv")
PatientsAmbulance <- inner_join( Ambulance, Hospitals )
PatientsAmbulance <- PatientsAmbulance %>%
  mutate( Arrival=ymd_hms(Arrival), Wait=as.numeric(Arrival-Call) )
PatientsAmbulance <- PatientsAmbulance %>%
  mutate( Category2 = case_when( Category2==1 ~ "Category 1",
                                 Category2==2 ~ "Category 2",
                                 Category2==3 ~ "Category 3",
                                 Category2==4 ~ "Category 4" ) )
```

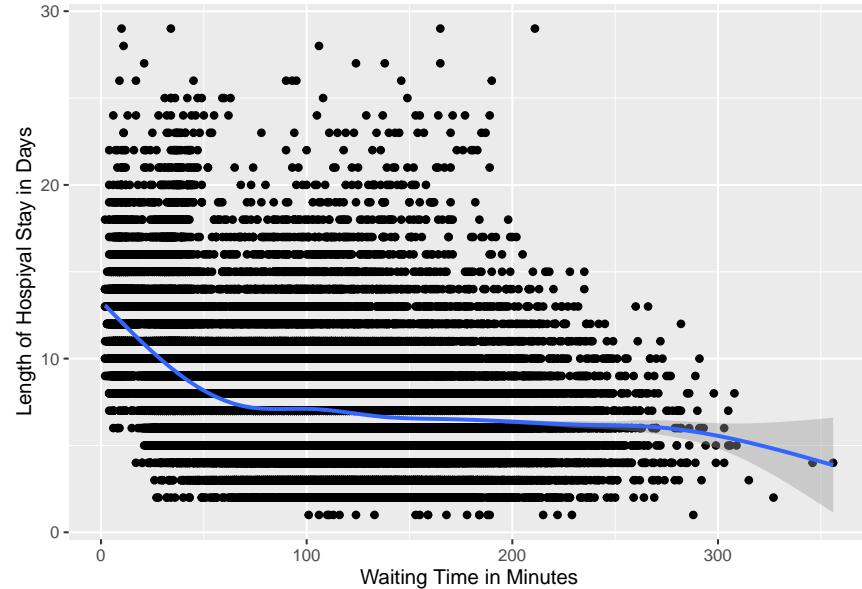
### Answer 1

We create a plot of waiting time versus length of stay:

```
ggplot( PatientsAmbulance, aes(x=Wait,y=Length) ) + geom_point() + geom_smooth() +
  labs( x="Waiting Time in Minutes", y="Length of Hospital Stay in Days")
```

```
## `geom_smooth()` using method = 'gam' and formula = 'y ~ s(x, bs = "cs")'
```



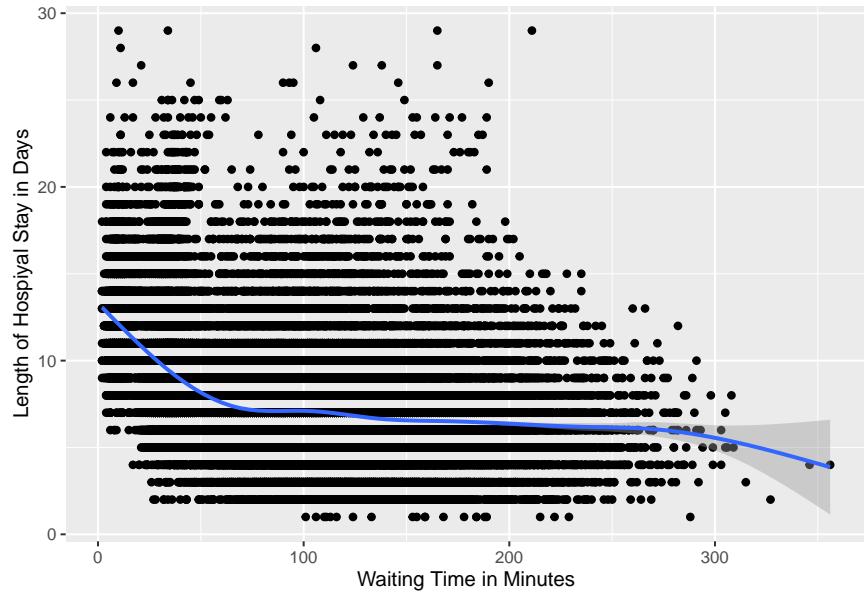
We find that in fact the opposite is observed: The length of the hospital stay decrease as the time taken to reach the patient increases. This would imply that the more severe a patient's case, the faster the ambulance is likely to get to them - that is, the ambulance service is prioritizing the care of the patients who need it more urgently.

## Answer 2

We create a plot of waiting time versus length of stay:

```
ggplot( PatientsAmbulance, aes(x=Wait,y=Length) ) + geom_point() + geom_smooth() +  
  labs( x="Waiting Time in Minutes", y="Length of Hospital Stay in Days")
```

```
## `geom_smooth()` using method = 'gam' and formula = 'y ~ s(x, bs = "cs")'
```

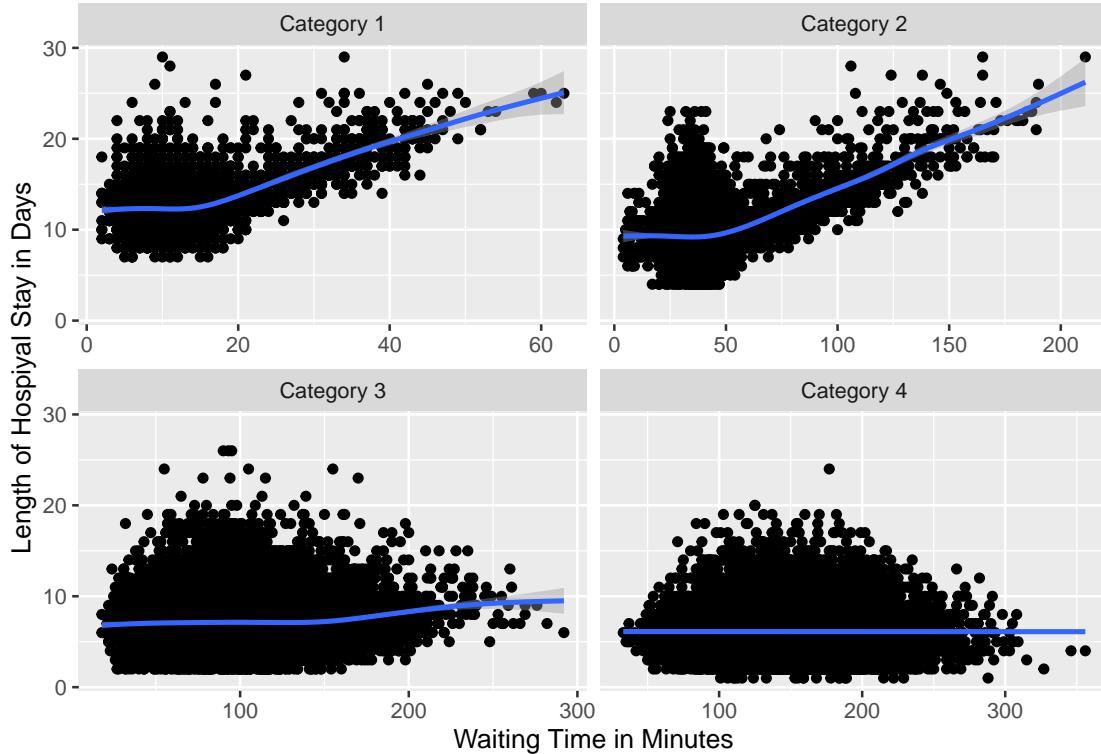


There is a fairly weak correlation between the length of stay in hospital and the time taken for the ambulance to arrive. From 0-100 minutes the length of stay decreases as the time for arrival increases and from 100-300 minutes the length of stay in hospital remains fairly constant.

### Answer 3

Since the effect may vary across the different ambulance categories, I create a plot of waiting time against length of stay for each ambulance response category:

```
ggplot( PatientsAmbulance, aes(x=Wait,y=Length) ) +  
  facet_wrap(~Category2, scales = "free_x") + geom_point() + geom_smooth() +  
  labs( x="Waiting Time in Minutes", y="Length of Hospital Stay in Days")  
  
## `geom_smooth()` using method = 'gam' and formula = 'y ~ s(x, bs = "cs")'
```



The plots show that for the most severe cases (Categories 1 and 2), we see a strong increase in the length of hospital stay for patients with a long waiting. For Category 1, the average length of stay starts to increase for patients that wait more than 15 minutes. For patients in Category 2, the average length of stay tends to increase if patients wait for more than 50 minutes. Finally, there is a small increase in the average for patients in Category 3 waiting more than 150 minutes, and no effect is visible for patients in Category 4. In summary, the data supports the statement that the length of stay tends to increase with the time the patient is waiting.