

Chapter 6

Problem 42

Part a)

Yes, the association is positive because the good riders are going to do well at both. However it will likely be a weak positive correlation because the riders who are good at sprinting and will do well in the short first stage even when they have fresh legs. The last time trial stage is will have a more distance oriented group take the top spots.

Part b)

Read the Data

```
## Loading required package: magrittr
```

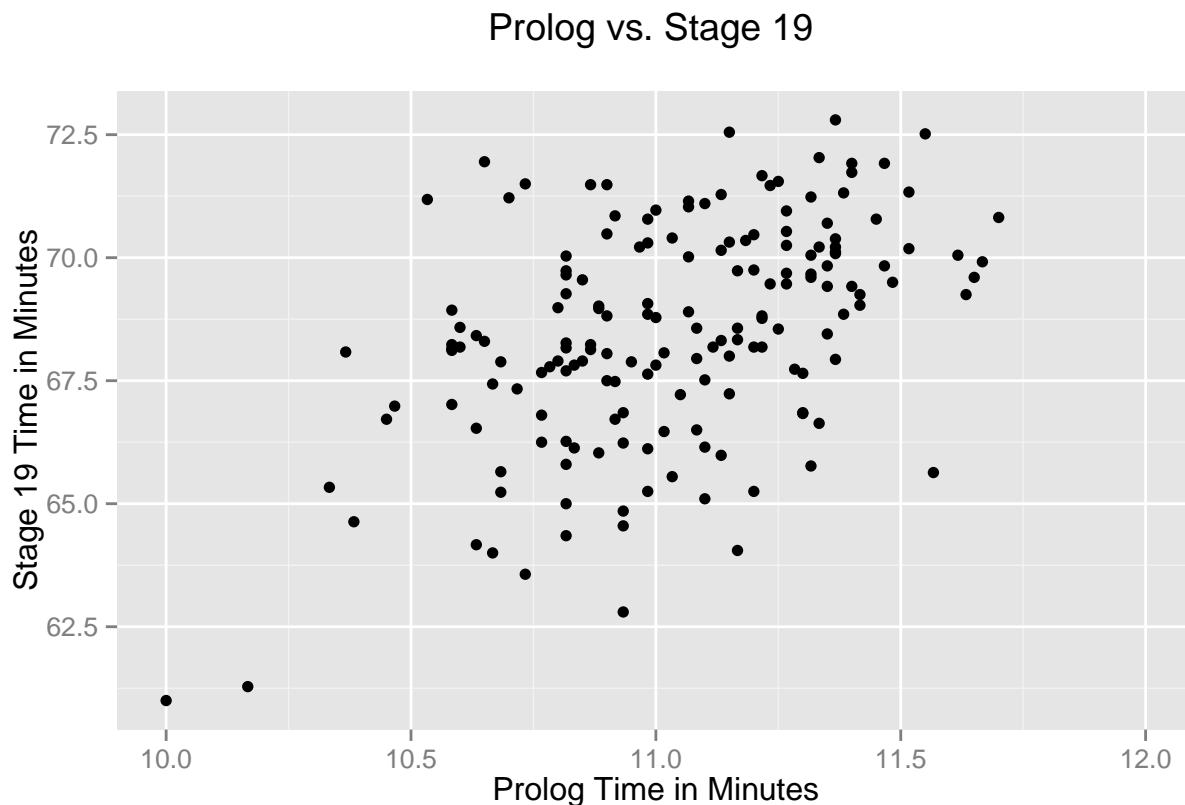
Graph the Scatter Plot

```
require(ggplot2)
```

```
## Loading required package: ggplot2
```

```
ggplot(data = tour, aes(x = PrologueTimeMinutes, y = Stage19TimeMinutes)) +  
  geom_point() +  
  xlim(10, 12) +  
  labs(x = "Prolog Time in Minutes", y = "Stage 19 Time in Minutes", title = "Prolog vs. Stage 19 \n")
```

```
## Warning: Removed 27 rows containing missing values (geom_point).
```



Based on the graph we can conclude weak positive.

Part c)

```
tour <- tour %>% na.omit # %>% is the pipe operator from the magrittr package
cor(tour$PrologueTimeMinutes, tour$Stage19TimeMinutes)
```

```
## [1] 0.481251
```

Part d)

As mentioned in part a, some riders are better suited to short fast time trials others are good at long hard efforts at the end of the tour. This accounts for some of the spread seen in the scatter plot.

Problem 53

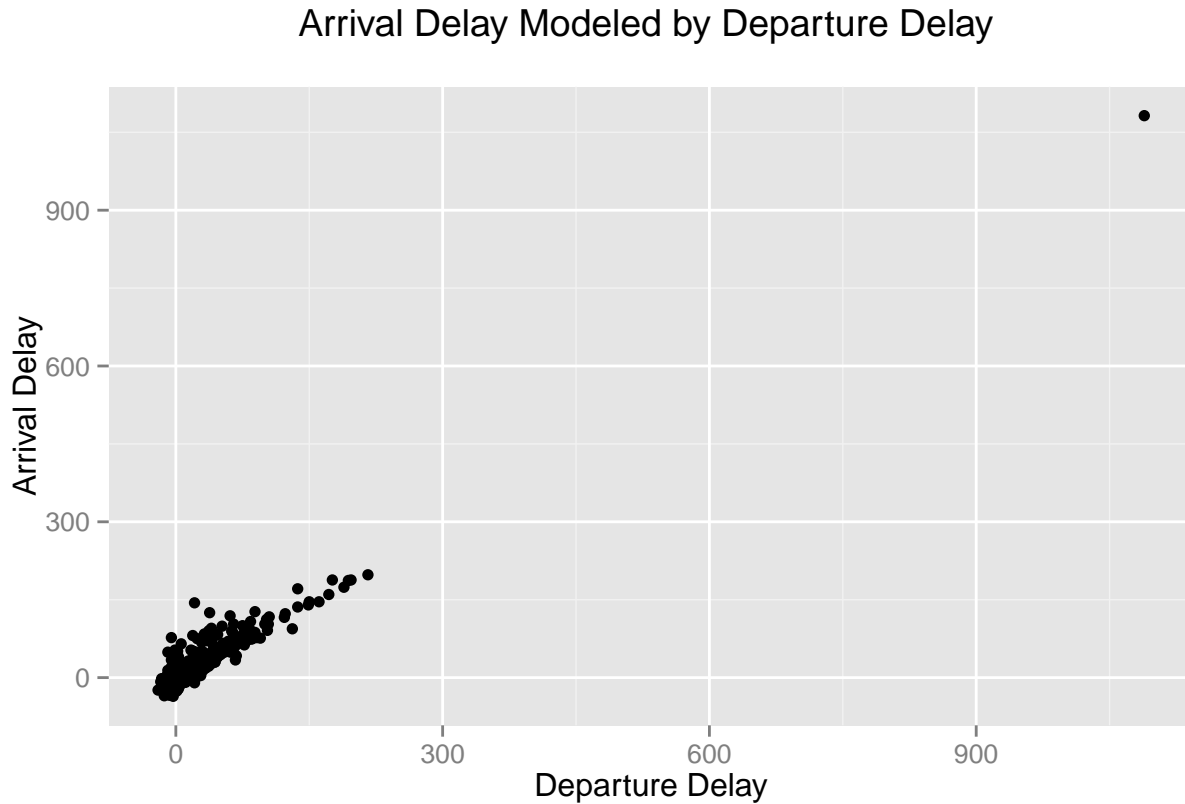
load the data

part a)

If the plane leave late you'd expect them to be late as well. I'd expect to see a positive correlation.

Part b)

```
ggplot(data = flight, aes(x = Departure_Delay., y = Arrival_Delay)) +  
  geom_point() +  
  labs(x = "Departure Delay", y = "Arrival Delay", title = "Arrival Delay Modeled by Departure Delay \n")
```



There is a strong positive relationship between the two.

Part C)

```
#----- With the outlier  
cor(flight$Arrival_Delay, flight$Departure_Delay.)
```

```
## [1] 0.9584526
```

Part D)

Let's take a look at the correlation with out the outlier.

```
#----- Without the outlier  
sub.flight <- flight[flight$Arrival_Delay <= 350, ] # subset that excludes the outlier  
cor(sub.flight$Arrival_Delay, sub.flight$Departure_Delay.)
```

```
## [1] 0.9069407
```

Removing the outlier reduces the strength of the correlation. Though not but very much.

Part E)

That shouldn't affect the correlations because everything would be scaled by the same factor. Let's check just in case.

```
cor(sub.flight$Arrival_Delay/60, sub.flight$Departure_Delay./60)
```

```
## [1] 0.9069407
```

It's the same as it was before. Scaling has no effect.