Chapter 6

Jonathan Nelson

March 6, 2015

Problem 42

Part a)

Yes, the association is possitive because the good riders are going to do well at both. How ever it will likely be a weak possitive corrilation 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

```
require(magrittr)
```

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

```
tour <- read.table("C:\\Users\\Jonathan\\Google Drive\\Stats Camp\\Stine&Foster\\Data by Chapter\\Chapt
```

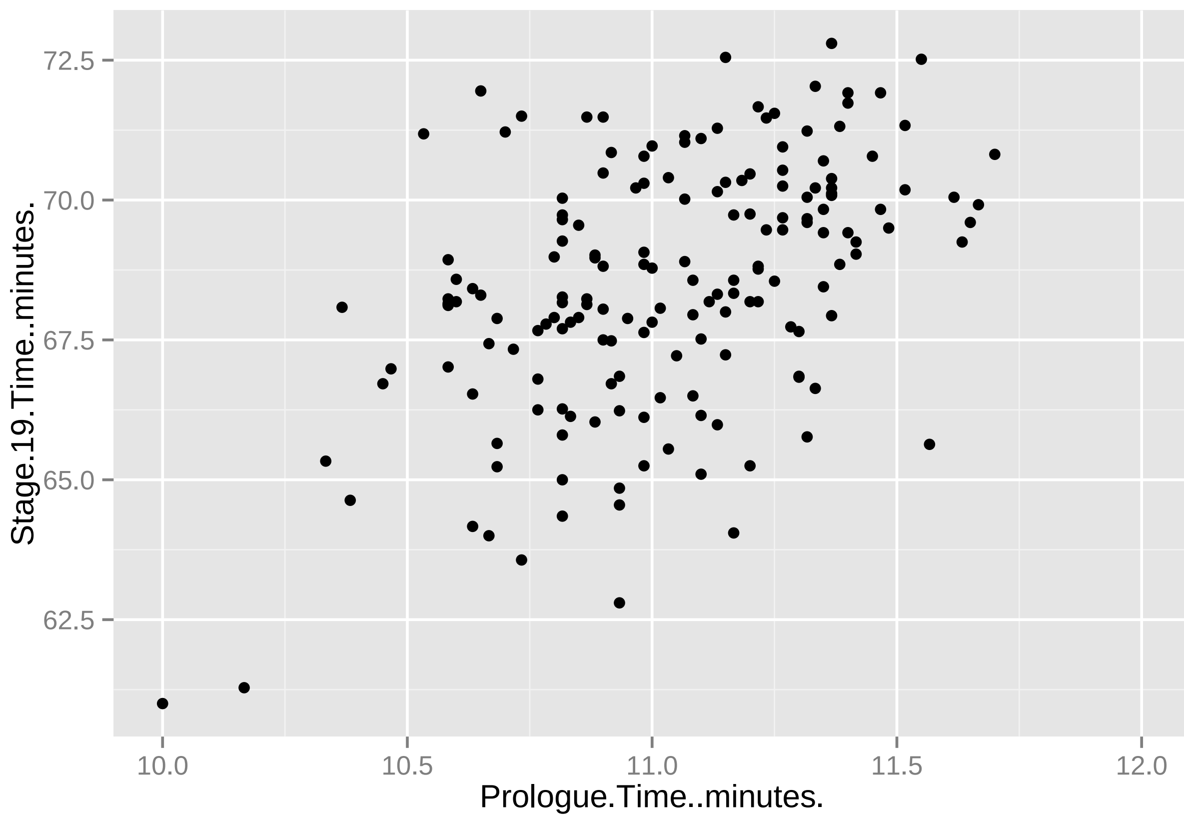
Graph the Scatter Plot

```
require(ggplot2)
```

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

```
ggplot(data = tour, aes(x = Prologue.Time..minutes., y = Stage.19.Time..minutes.)) +
  geom_point() +
  xlim(10, 12)
```

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



Based on the graph we can conclue weak possitive.

Part c)

```
tour <- tour %>% na.omit # %>% is the pipe operator from the magrittr package

cor(tour$Prologue.Time..minutes., tour$Stage.19.Time..minutes.)
```

```
## [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 acounts for some of the spread seen in the scatter plot.

Problem 53

load the data

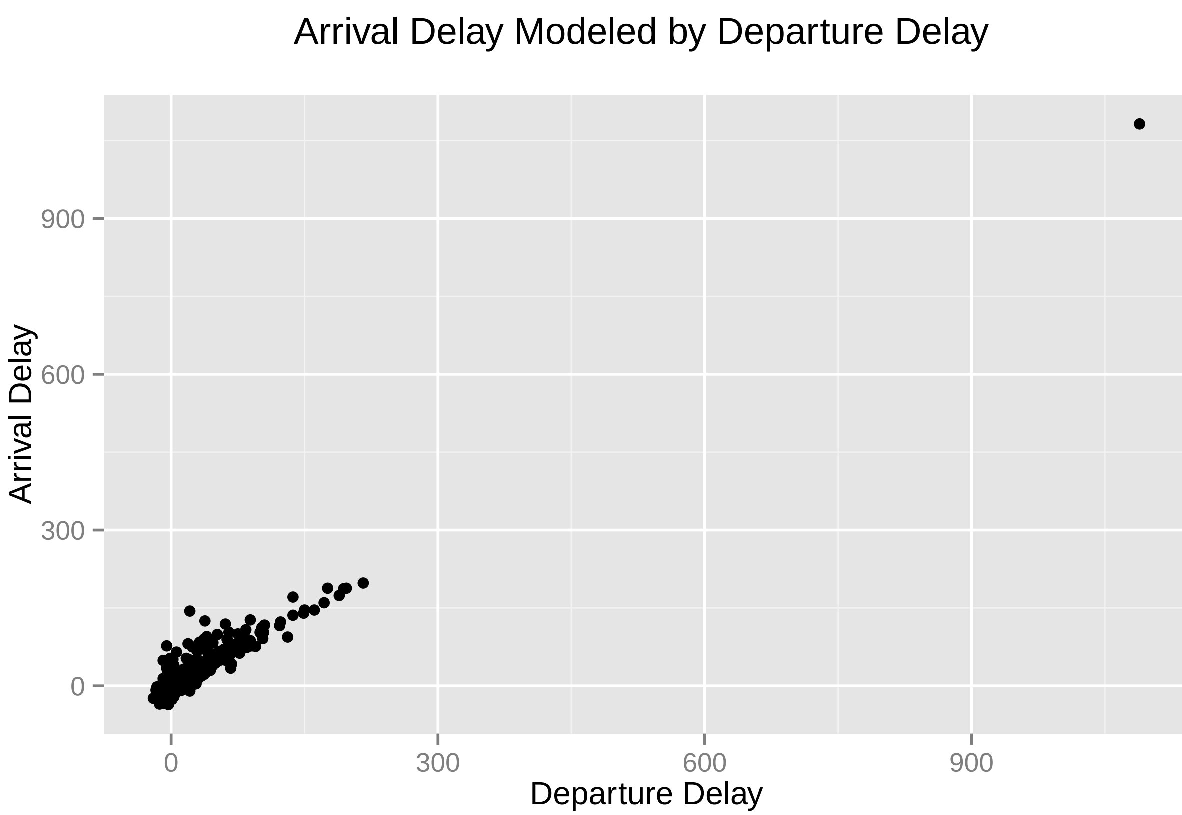
```
load("C:\\Users\\Jonathan\\Google Drive\\Stats Camp\\Stine&Foster\\Data by Chapter\\Chapter 6\\06_fligh
```

part a)

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

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 possitive reltionship 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 oulier reduces the stregth 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.