Nikita Shvachko
Statistics 133
Date: 02/05/08

Homework 1

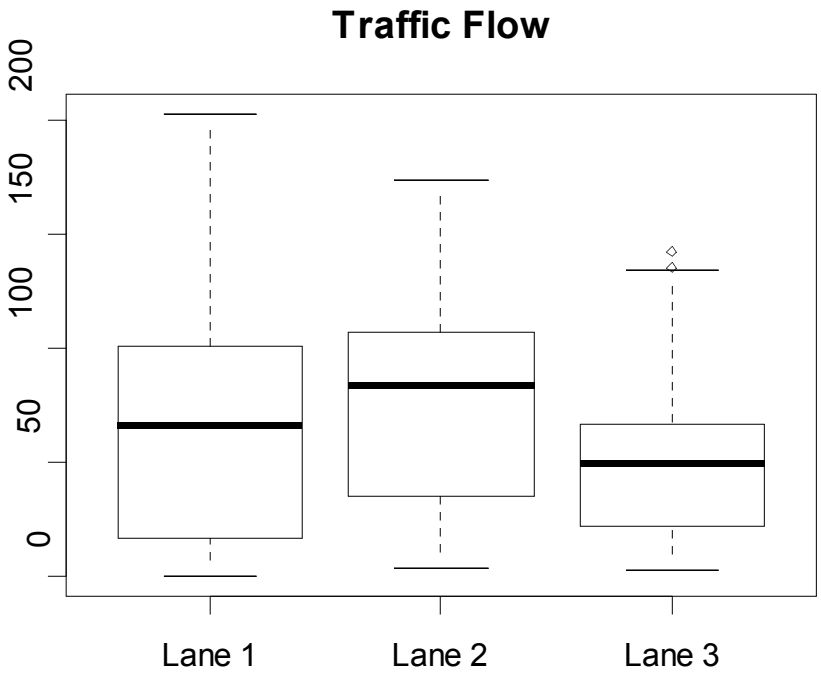
1) I used the "read.csv" commands because it has the right defaults set for reading in comma separated values. Here is the full command I issued:

```
> tflow =
read.csv("http://www.stat.berkeley.edu/users/nolan/stat133/data/f
low-occ-table.txt")
```

2) Explore the data visually to answer the following questions:

1. To plot a boxplot I issued the following command:

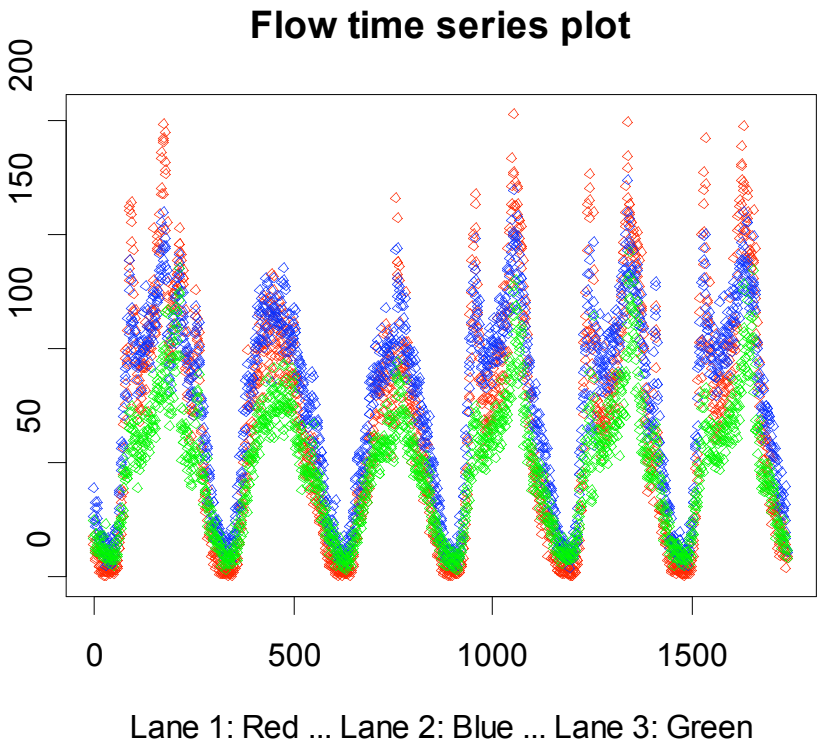
```
> boxplot(tflow$Flow1, tflow$Flow2, tflow$Flow3, main = "Traffic
Flow", ylab="Number of cars", names=c("Lane 1","Lane 2","Lane
3"))
```




The statement "The flow in lane 2 is typically about 50% higher than lane 3" is correct because the median of Lane 2 is about 80 and the median of Lane 3 is about 50 so it is reasonable to conclude that Lane 2 typically has 50% higher flow than lane 3.

2. In order to plot the three flows, I used the following commands:

```
> plot(tflow$Flow1, col = "red", main = "Flow time series plot",
xlab="Lane 1: Red ... Lane 2: Blue ... Lane 3: Green", ylab =
"Number of Cars")

> points(tflow$Flow2, col = "blue")

> points(tflow$Flow3, col = "green")
```

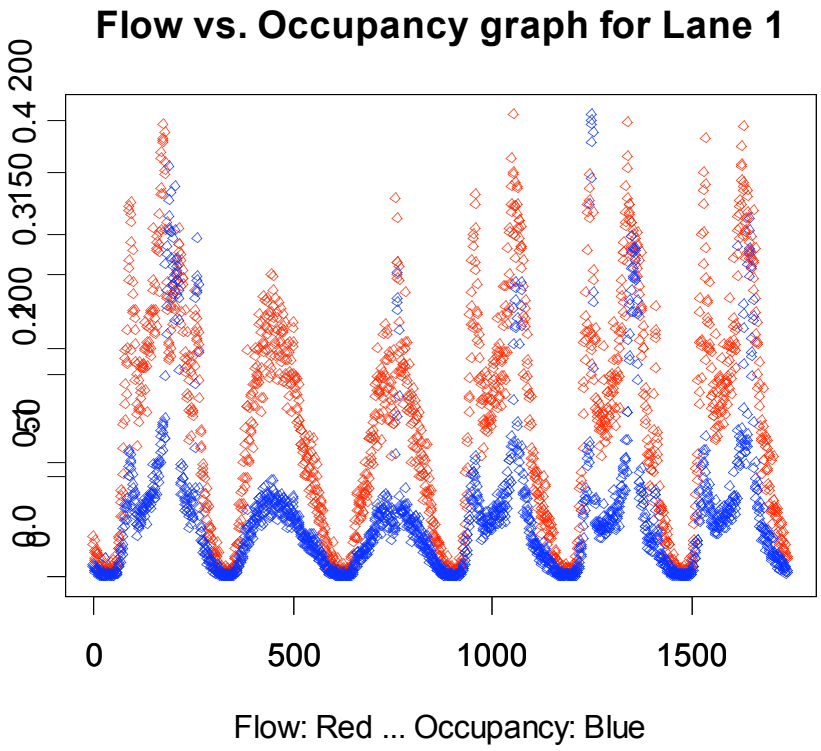



Lane 1 (in red) is the lane that typically serves the most traffic because it generally has the most cars at peak traffic hours.

The statement "When one lane is congested, the others are too" is accurate because the time series plot clearly shows the flow for the three lanes overlapping each other, when graphed with respect to time.

3. To plot Flow vs Occupancy in Lane 1, I issued the following commands:

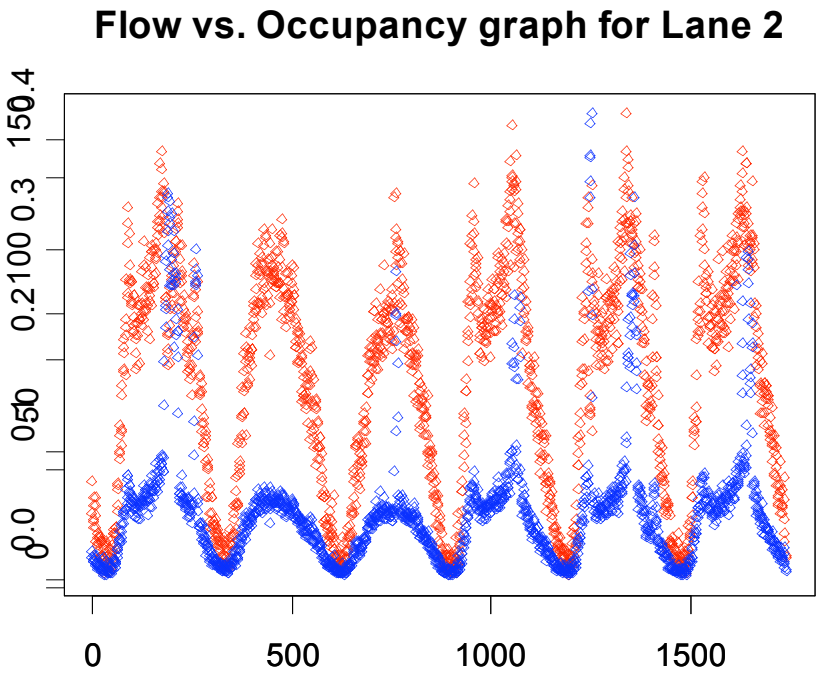
```
> plot(tflow$Flow1, col = "red", main = "Flow vs. Occupancy graph
for Lane 1", xlab = "Flow: Red ... Occupancy: Blue", ylab = "Flow
(0-200) + Occupancy (0-0.5)")

> par(new = T)

> plot(tflow$Occ1, col = "blue", xlab = "", ylab = "")
```



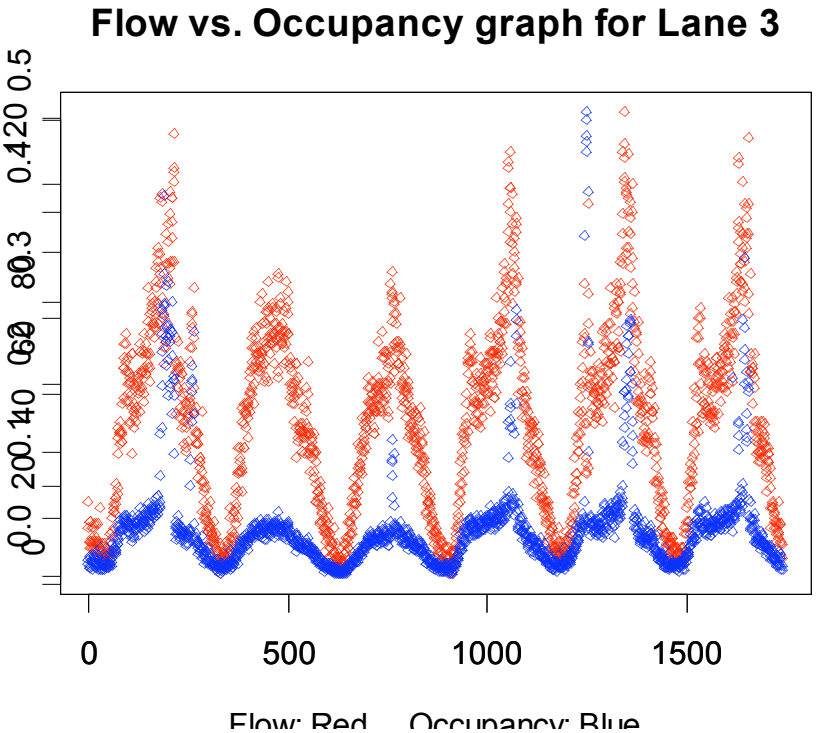

The Flow and Occupancy have a very high correlation, as the Flow increases, the occupancy also increases. This of course represents traffic: when there are more cars, the speeds go down and the time that a car spends in a certain space is increased. On the flip side when there are fewer cars, the speeds increase and the time spent (occupancy) goes down.

To plot Flow vs Occupancy in Lane 2 and 3, I issued the following commands:

```
> plot(tflow$Flow2, col = "red", main = "Flow vs. Occupancy graph for
Lane 2", xlab = "Flow: Red ... Occupancy: Blue", ylab = "Flow (0-200) +
Occupancy (0-0.5)")

> par(new = T)

> plot(tflow$Occ2, col = "blue", xlab = "", ylab = "")
```




```
> plot(tflow$Flow3, col = "red", main = "Flow vs. Occupancy graph for
Lane 3", xlab = "Flow: Red ... Occupancy: Blue", ylab = "Flow (0-200) +
Occupancy (0-0.5)")

> par(new = T)

> plot(tflow$Occ3, col = "blue", xlab = "", ylab = "")
```



After looking at Flow vs. Occupancy plots for lanes 2 and 3, we can clearly see that they have the same relationship found in the Flow vs. Occupancy plot for lane 1. This means that the correlation between the number of cars (Flow) and the time spent occupying a stretch of road (Occupancy) applies to all lanes. This relationship makes perfect sense when thinking about traffic congestion in the real world.