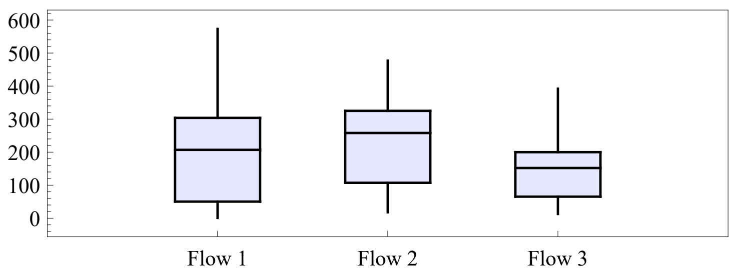
Name: Rufus Petrie

Final Problem 5: Traffic Flow Study

```
labelList = {"Flow 1", "Flow 2", "Flow 3"};
BoxPlot[flow1, flow2, flow3,
 Labels → labelList,
 AspectRatio → 1 / 3]
```



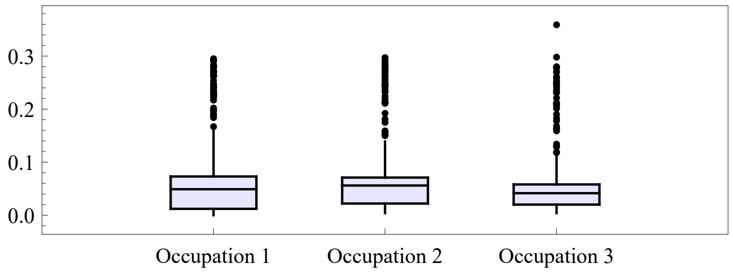

```
TableForm[Data,
 TableHeadings → {varNames, labelList}]
```

	Flow 1	Flow 2	Flow 3
Sample Size	576	576	576
Means	198.613	225.41	144.981
Medians	207.	258.	152.
SDs	147.434	122.848	83.7613

From these tables, it appears that the data for the flow of each lane is relatively well-behaved. From the means and medians, it appears that the middle lane typically has the highest amount of usage followed the the left lane and the right lane (in that order). Interestingly enough, it appears that the rightmost lane has the most variation in its usage, followed by the middle lane and the rightmost lane (in that order). This is probably due to the fact taht people generally change between the midlane and the fast lane, whereas they may switch to the rightmost lane left often for when they have to exit the I80.

Part (b) - Occupation

```
labelList = {"Occupation 1", "Occupation 2", "Occupation 3"};
BoxPlot[occ1, occ2, occ3,
 Labels → labelList,
 AspectRatio → 1 / 3]
```



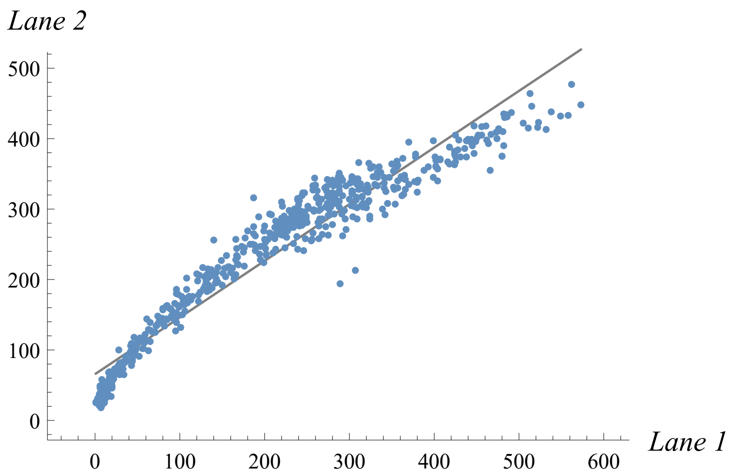

```
TableForm[Data,
 TableHeadings → {varNames, labelList}]
```

	Occupation 1	Occupation 2	Occupation 3
Sample Size	576	576	576
Means	198.613	225.41	144.981
Medians	207.	258.	152.
SDs	147.434	122.848	83.7613

From these tables, it appears that the occupation data is a bit more complicated than the flow data. In particular, we see that the flow of each lane has many upper outliers. This likely corresponds to the times during the week in which there was a great amount of congestion. From the summary statistics, we can see that like it was with usage, the middle lane has the most congestion, followed by the left lane and the right lane (in that order). However, unlike the usage numbers, there doesn't appear to be as much of a difference in the variability of lane congestion. This likely results from the fact that for high congestion periods, all three lanes clog up simultaneously.

Part (c) - Scatter Plots

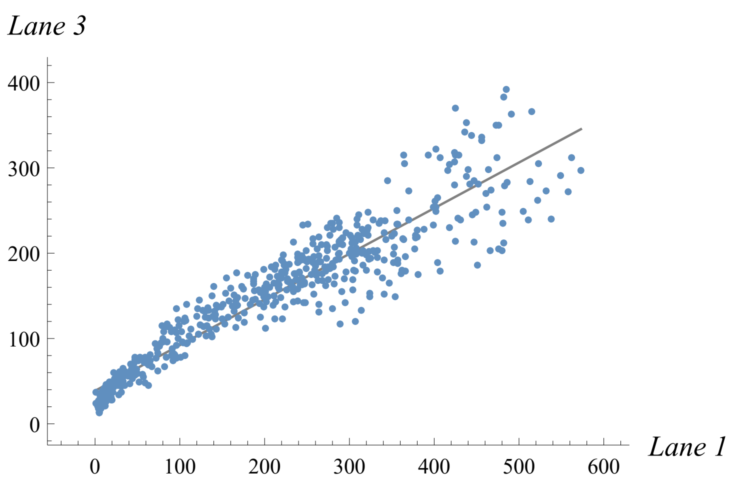
```
Pairs = Transpose[{flow1, flow2}];
ScatterPlot[Pairs, Correlation → True,
 AxesLabel → {"Lane 1", "Lane 2"}]
Clear[x]; Remove[f];
f[x_] = Fit[Pairs, x, x]
```




```
Correlation:     0.965
```

```
1.01751 x
```

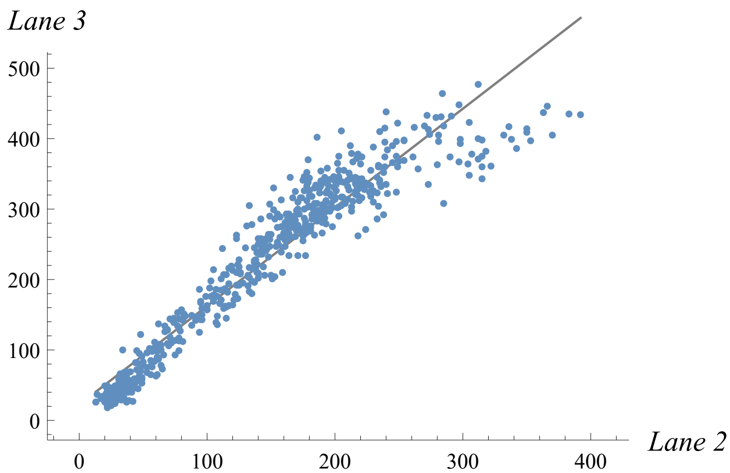
```
Pairs = Transpose[{flow1, flow3}];
ScatterPlot[Pairs, Correlation → True,
 AxesLabel → {"Lane 1", "Lane 3"}]
Clear[x]; Remove[f];
f[x_] = Fit[Pairs, x, x]
```




```
Correlation:     0.942
```

```
0.660838 x
```

```
Pairs = Transpose[{flow3, flow2}];
ScatterPlot[Pairs, Correlation → True,
 AxesLabel → {"Lane 2", "Lane 3"}]
Clear[x]; Remove[f];
f[x_] = Fit[Pairs, x, x]
```



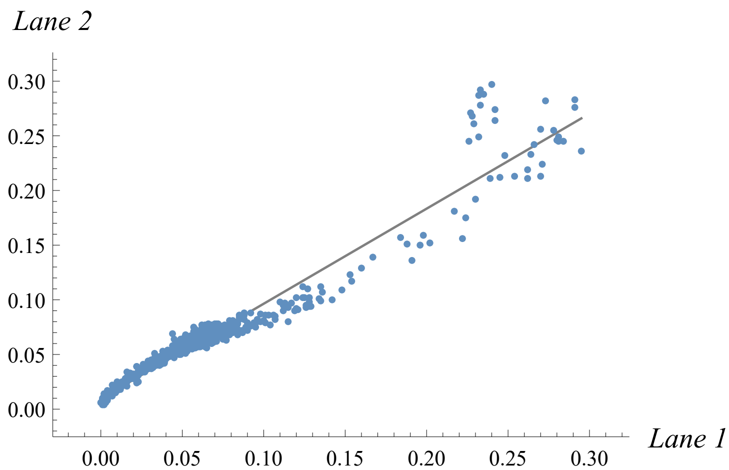

```
Correlation:     0.9547

1.51614 x
```

From these results, it appears that each of the pair comparisons has a high correlation (above 0.94). Furthermore, most of the points on the scatter plots hug the correlation line pretty closely, so making statements of the form like "The flow in lane 2 is typically 50% higher than the flow in lane 3" is probably a decent summary of the relationships. As we can see from the regression output coefficient from the lane 3/2 comparison, the flow in lane 2 is actually around 51% higher than the flow for lane on average.

Part (d) - Simultaneous Congestion

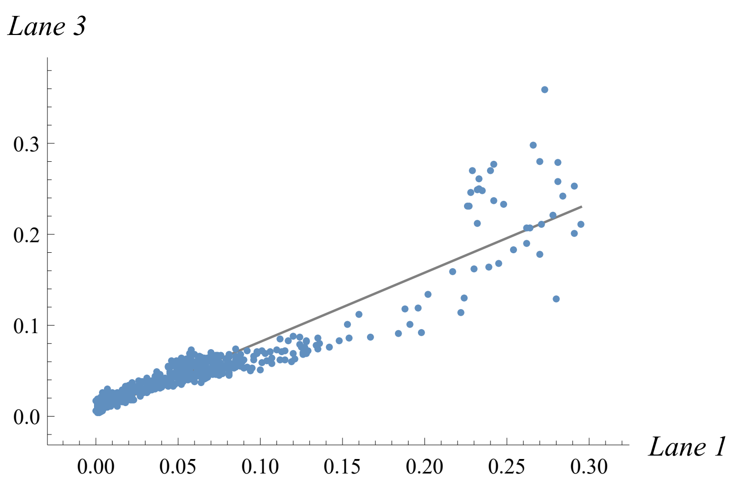
```
Pairs = Transpose[{occ1, occ2}];
ScatterPlot[Pairs, Correlation → True,
 AxesLabel → {"Lane 1", "Lane 2"}]
Clear[x]; Remove[f];
f[x_] = Fit[Pairs, x, x]
```




```
Correlation:     0.9739
```

```
0.944933 x
```

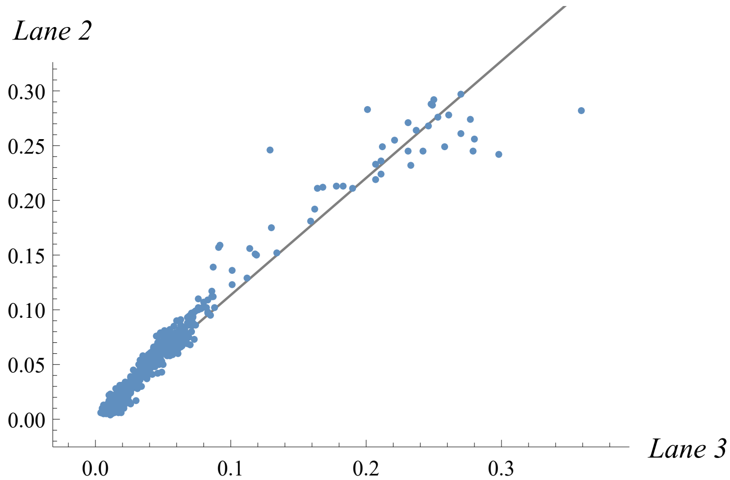
```
Pairs = Transpose[{occ1, occ3}];
ScatterPlot[Pairs, Correlation → True,
 AxesLabel → {"Lane 1", "Lane 3"}]
Clear[x]; Remove[f];
f[x_] = Fit[Pairs, x, x]
```




```
Correlation:     0.9351
```

```
0.806747 x
```

```
Pairs = Transpose[{occ3, occ2}];
ScatterPlot[Pairs, Correlation → True,
 AxesLabel → {"Lane 3", "Lane 2"}]
Clear[x]; Remove[f];
f[x_] = Fit[Pairs, x, x]
```



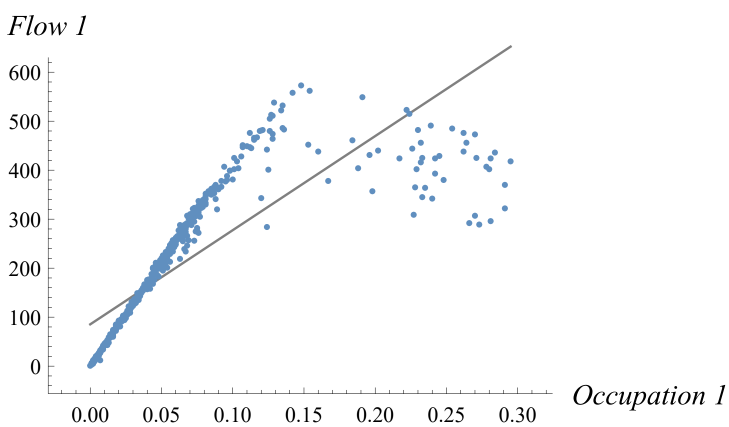

```
Correlation:     0.9737

1.1337 x
```

As we can see from the scatterplot outputs, there appear to be strong positive correlations between the congestion in the lanes, with each correlation being above 0.9. Furthermore, we can see from the regression outputs that the congestion in each lane rises steadily with respect to the congestion in the other lanes. The lowest coefficient is 0.8, which describes the linear relationship between lane 1 and 3. These two lanes are the farthest apart, so it makes sense that the congestion in those two lanes would have the lowest linear relationship.

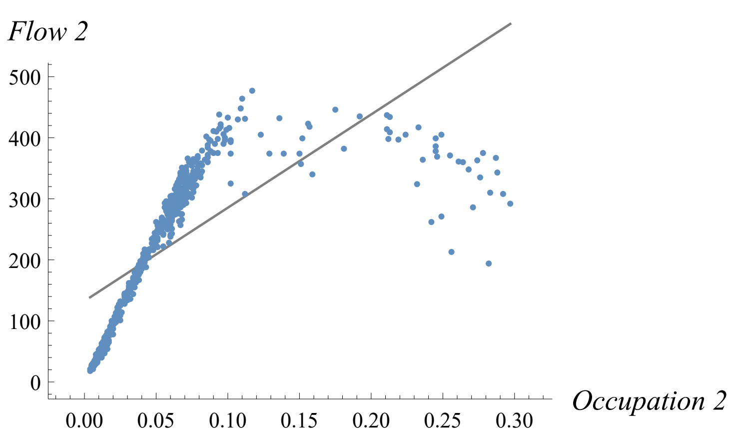
Part (e) - Pair Comparison

```
ScatterPlot[pairs1, Correlation → True,
 AxesLabel → {"Occupation 1", "Flow 1"}]
```



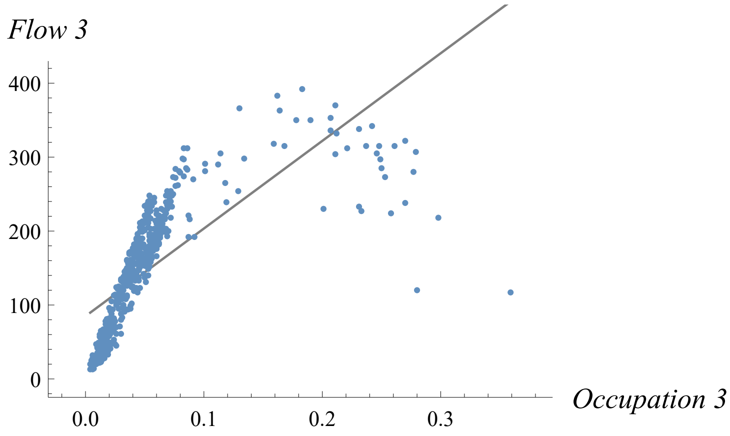

```
Correlation:     0.8179
```

```
ScatterPlot[pairs2, Correlation → True,
 AxesLabel → {"Occupation 2", "Flow 2"}]
```




```
Correlation:     0.6969
```

```
ScatterPlot[pairs3, Correlation → True,
 AxesLabel → {"Occupation 3", "Flow 3"}]
```




```
Correlation:     0.7216
```

Based on the scatter plots, this hypothesis seems to be false. In each graph, flow initially increases as occupation increases. However, at some critical level of occupation flow actually starts to decrease. The flow/occupation relationship is not the same in all lanes. In particular, it takes the longest for the flow of lane 1 to start decreasing, and lane 3 is the quickest one to start slowing down as occupancy decreases.