

Michael Root

Data Mining

Homework 2

Part 1.)

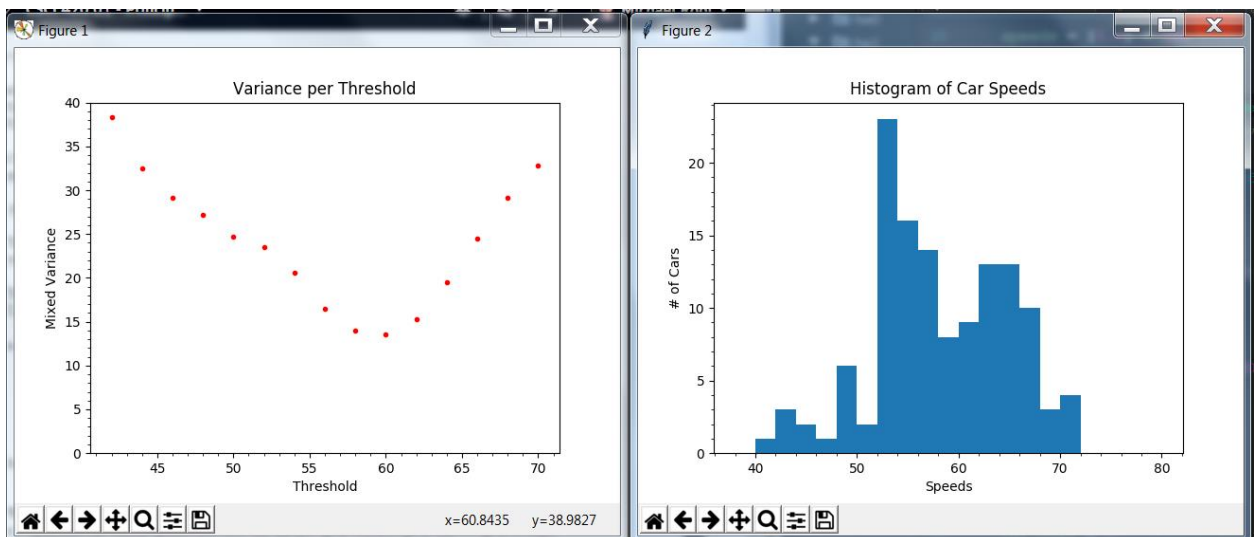
A.) Ethics

I don't have any ethical dilemmas with monitoring traffic on a road. People set out with full intention of being in public, so it's free to the public to monitor what happens in free space. People are already recorded in their travels whether they know it or not. Be it toll roads, police seeing them pass, or them being seen on a gas stations security camera.

B.)

Considering we already have that type of software available and being used in intersections, I would have no qualms with writing software similar to it. Like I said before, people are aware they are in public domain when they are driving on roads, so there should be aware that they may become data points while doing so.

C.)→ E.)



F.)

The best speed to separate would be a threshold greater than or equal to 60.

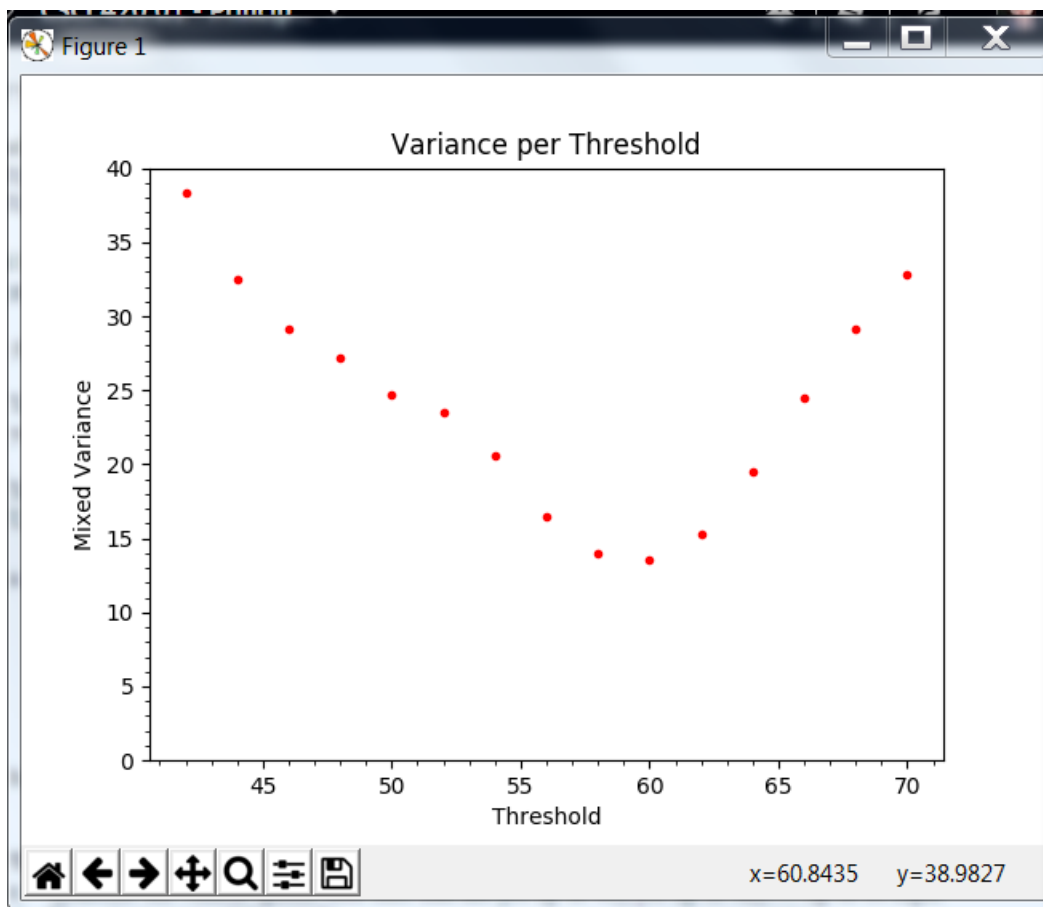
G.)

The minimum mixed variance at that threshold is 13.5962318349.

H.)

In my program, the first minimum mixed variant will remain if another is discovered, as I tell my program only to get a new threshold if the new minimum mixed variant is less than the current one. Not if it is less than or equal to the current one.

Part 2.)



Part 3.)

A.) Unknown Data Mode = 7, 8, 16

Unknown Data Median = 14.5

Unknown Data Average = 15.8

B.) Unknown Data 2 Mode = 7, 8

Unknown Data 2 Median = 14.0

Unknown Data 2 Average = 15.8

The mode changes because there is now one less value that can be considered the mode. The Median goes down by .5 because one of the higher values is taken away from it. But the average stays the same because we took out a value from the middle. By pulling a value from the center of mass of the data we didn't upset anything, so it stays the same.

C.) Threshold: 20

Minimum Mixed Variance: Greater than or equal to 40.9128