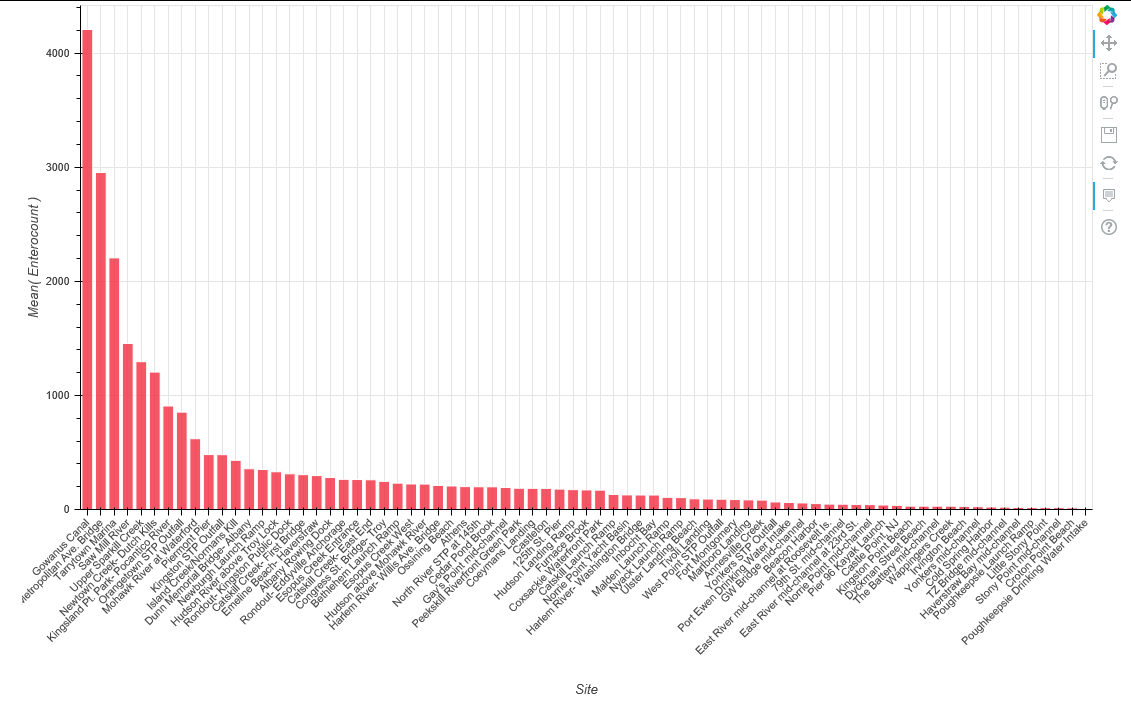
Angus Huang | Data 608 | Homework 4

03/26/2017

In order to present the results in a clean and concise manner, the graphs from jupyter notebook are attached below. Since the graph from "bokeh" module are rendered as html and does not display within the 'Jupyter Notebook', they are presented below.

1) Create lists & graphs of the best and worst places to swim in the dataset.

As shown in below graph, the worst place to swim is near "Gowanus Canal" and the best place to swim is "Poughkeepsie Drinking Water Intake", although this might be a legal swimming place.

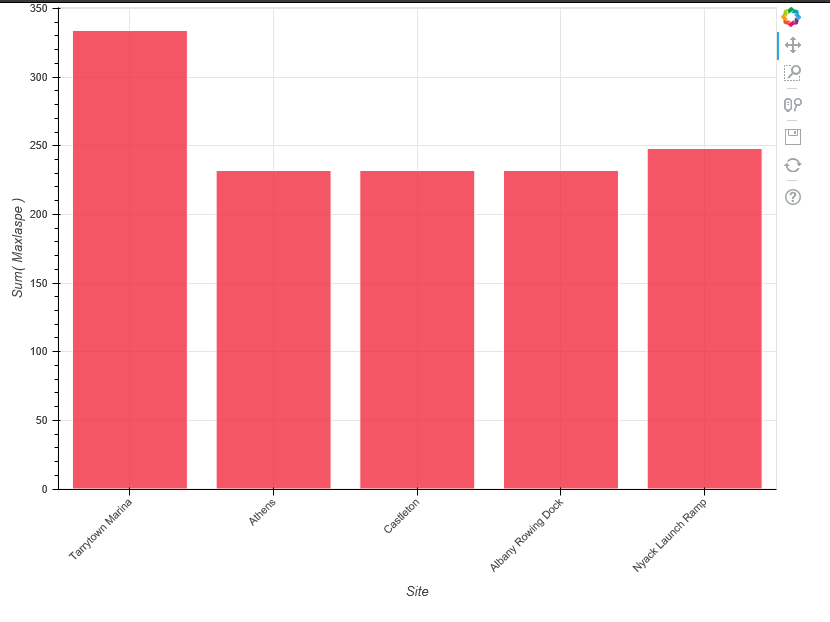


2) The testing of water quality can be sporadic. Which sites have been tested most regularly?

Which ones have long gaps between tests? Pick out 5-10 sites and visually compare how

regularly their water quality is tested.

The sites that were most regular tested includes: "Piermont Pier", Upper Sparkill Creek" and "125th St. Pier" while "Tarrytown Marina" and "Nyack Launch Ramp" were among the longest lapse days and less frequently tested site.



3)

3. Is there a relationship between the amount of rain and water quality? Show this

relationship graphically. If you can, estimate the effect of rain on quality at different sites and

create a visualization to compare them.

The linear regression graph below shows high rainfall is correlated with higher entero count meanings low water quality. Although, the correlation is not high, it does seems to imply the rain water carrier fecal matters to the river.

