You can compile the different vector sum binaries as follows:

```
nvcc vector_sum_gpu.cu -o vecto_sum_gpu
nvcc vector_sum_gpu_v2.cu -o vecto_sum_gpu
nvcc vector_sum_gpu_v3.cu -o vecto_sum_gpu
```

To run:

```
./vector_sum_gpu <array length>
```

These three different vector sum files add two vectors together with multiple levels of correctness and usage of threads and blocks. vector_sum_gpu can only add two arrays of length up to 512. vector_sum_gpu_v2 has the same limitation, but it uses a 2 dimensional allocation of threads (it sets the block dimensions). vector_sum_gpu_v3 will work for arrays of any length by using multiple kernel calls.