Assignment 4: Model-Based RL and Exploration

Andrew ID: `hvz`
Collaborators: `shivamt`
NOTE: Please do **NOT** change the sizes of the answer blocks or plots.

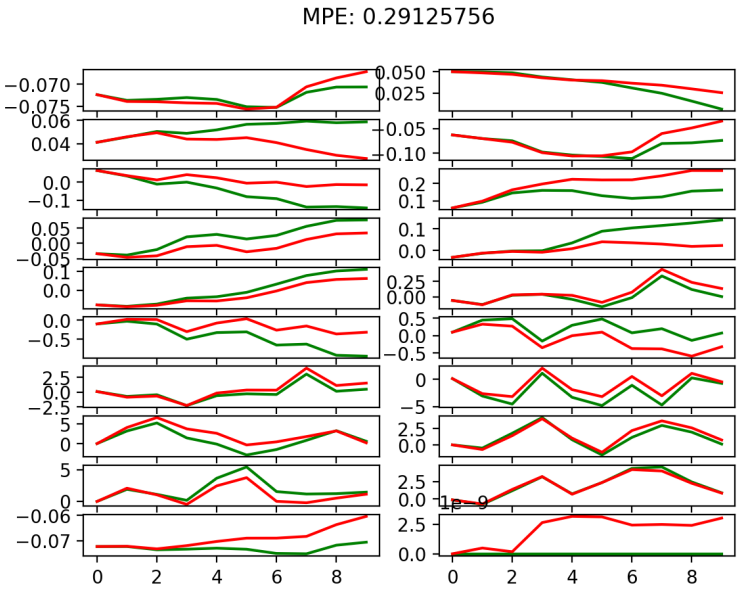
All the plots from tensorboard are with smoothing = 0.8

1 Problem 1: Dynamics Model Training – [10 points total]

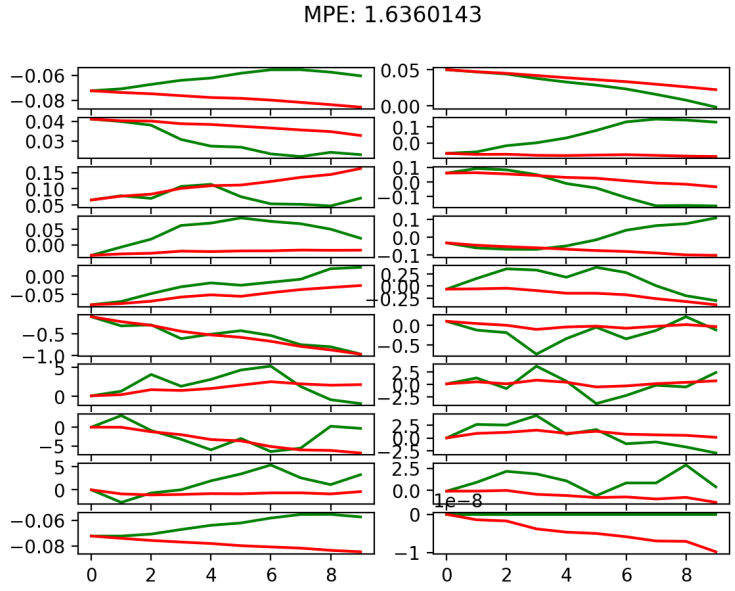
Theory questions

The superior performance of the largest model, characterized by two layers and 250 neurons each, which recorded a Mean Prediction Error (MPE) of 0.078, can primarily be explained by two key factors. Firstly, the enhanced capacity of a more extensive neural network enables it to capture more complex relationships within the data, thus improving its ability to develop an effective dynamics model. Secondly, extending the duration of the training process provides more opportunity for the model to refine its parameters, thereby converging towards a more precise representation of the actual dynamics involved. Combining a more complex network architecture and increased training iterations is crucial for achieving lower prediction errors.

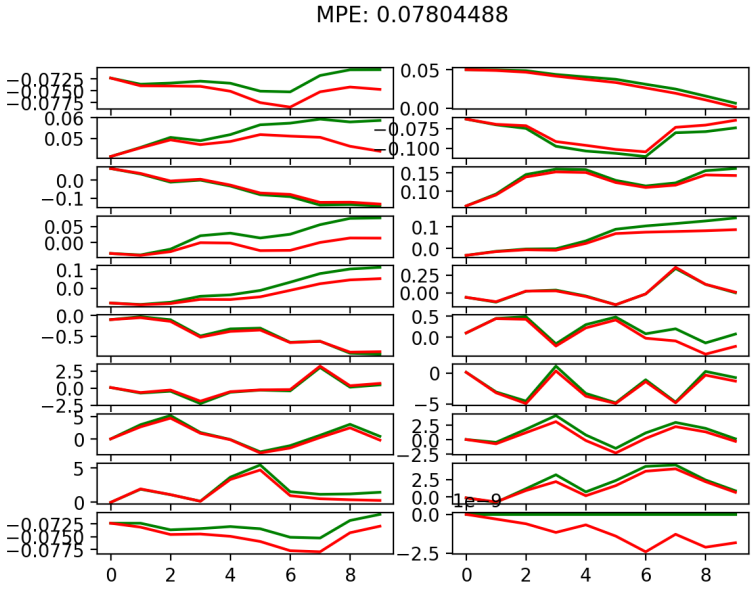
q1-cheetah-n500-arch1x32



q1-cheetah-n5-arch2x250

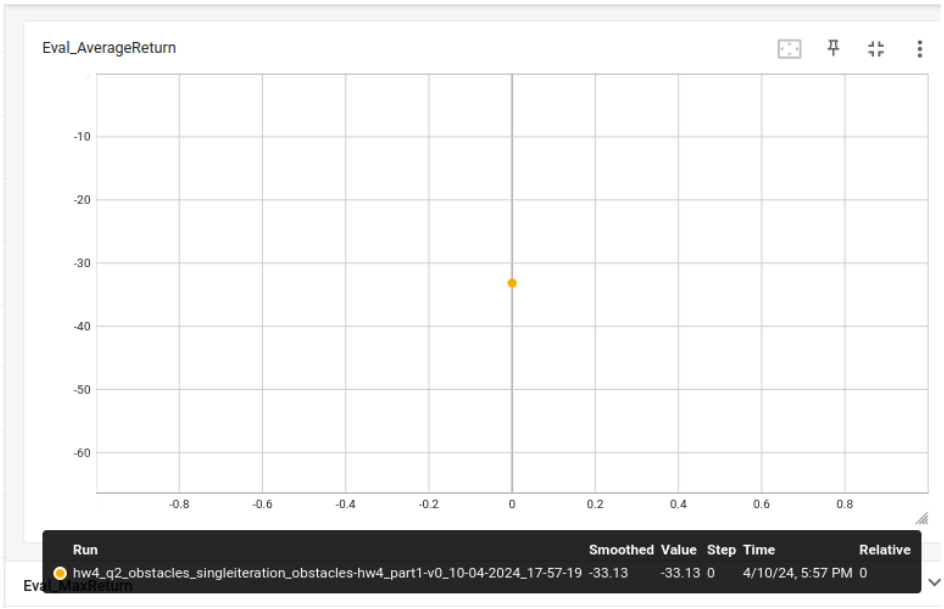


q1-cheetah-n500-arch2x250

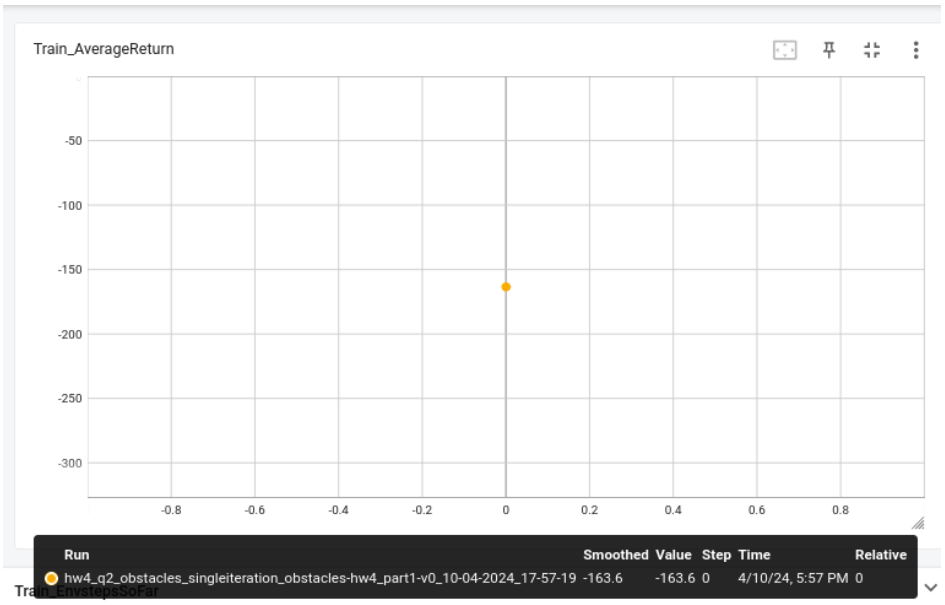


2 Problem 2: Action Selection

EvalAverageReturn

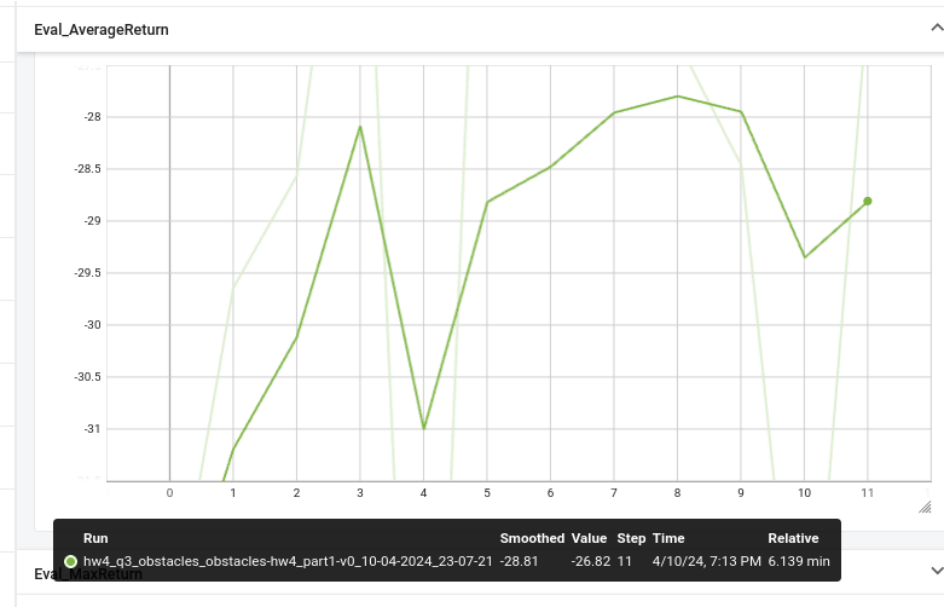


TrainAverageReturn

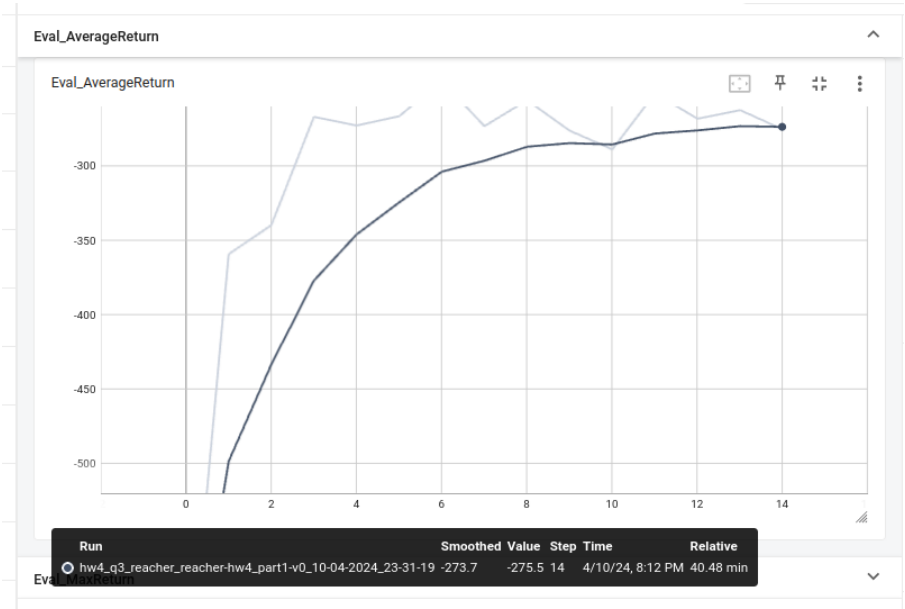


3 Problem 3: Iterative Model Training

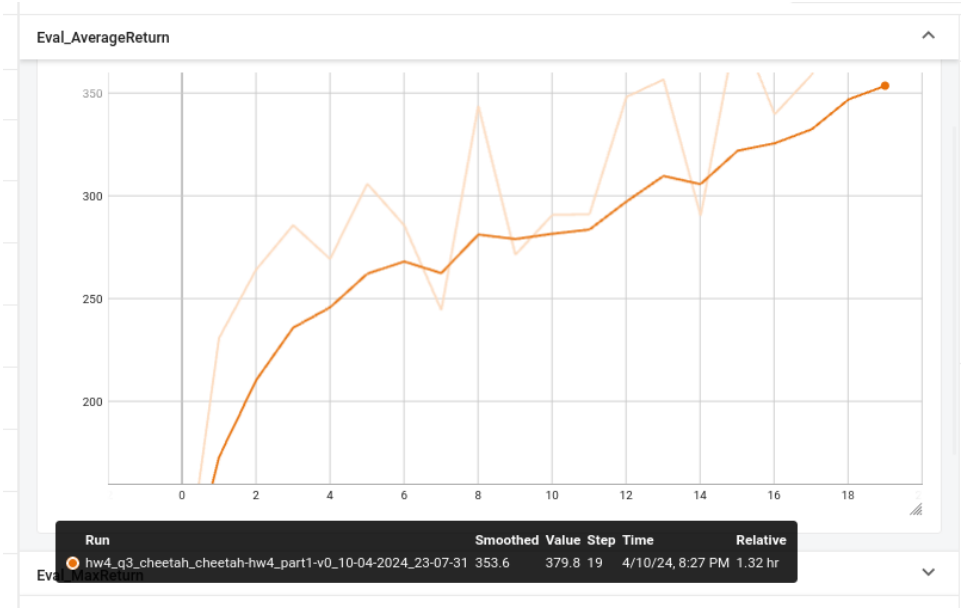
Obstacles



Reacher

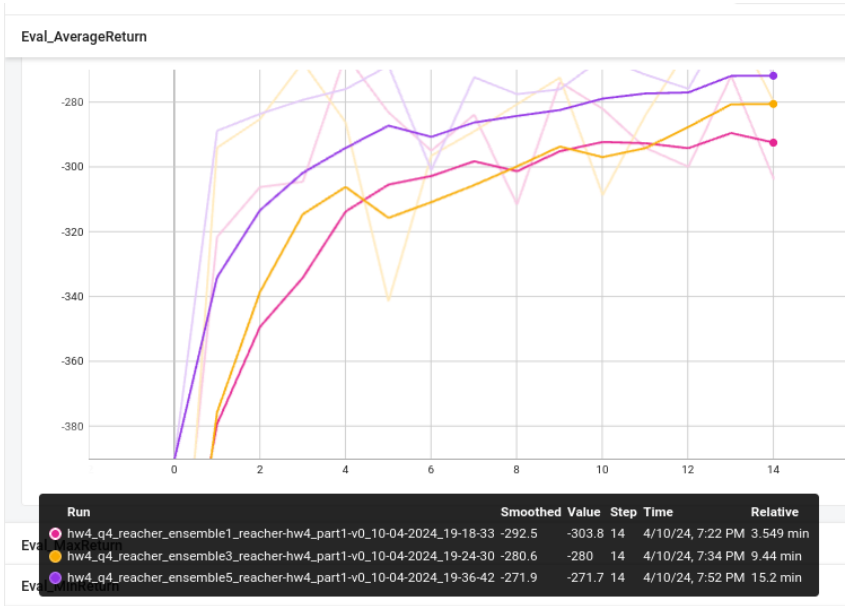


Cheetah



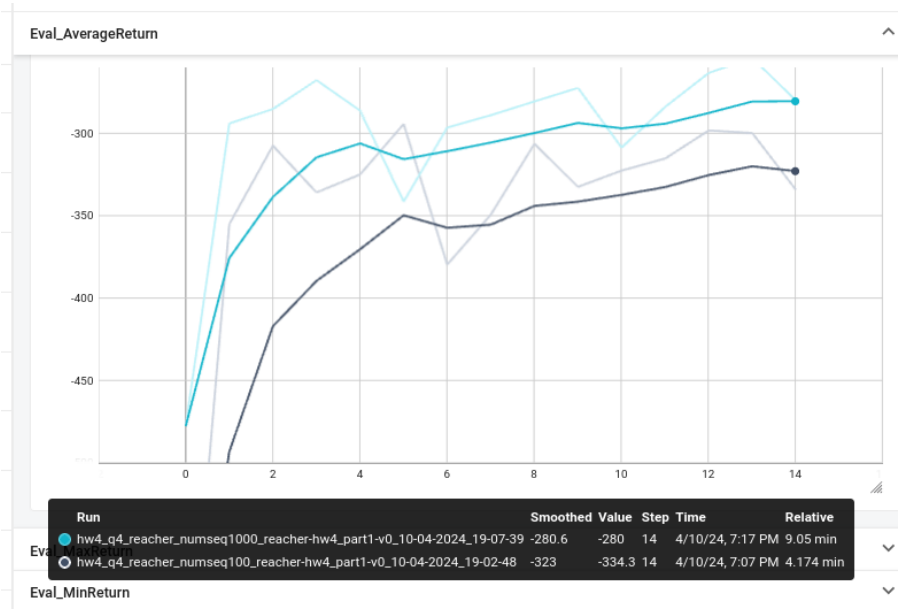
4 Problem 4: Hyper-parameter Comparison

effect of ensemble size



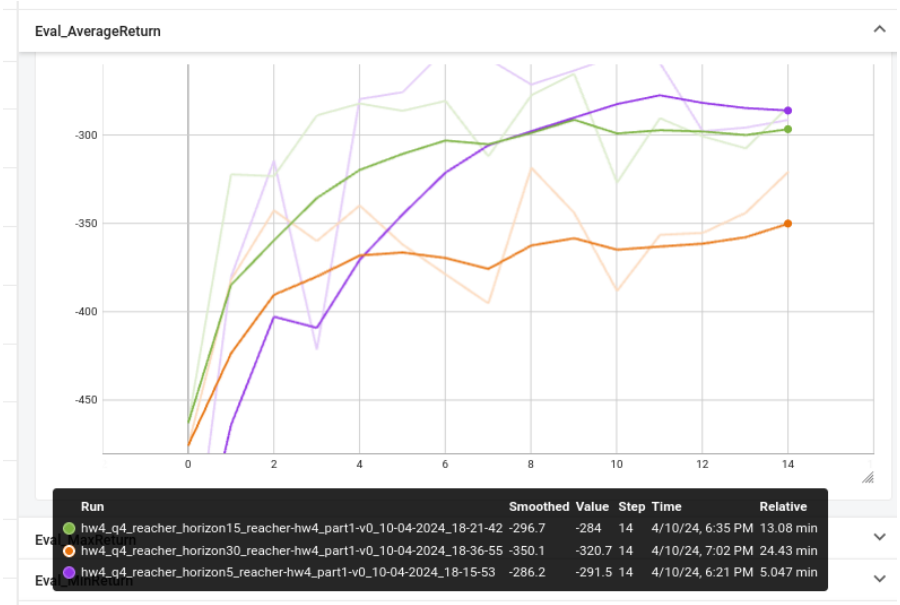
While the final reward was almost similar across different ensemble sizes, we can see that a larger ensemble size of 5 enabled the agent to achieve a higher reward sooner. The reward gradually increased to its maximum value with an ensemble size of 1.

effect of the number of candidate action sequences



We can see that the larger number of candidate sequences (1000) resulted in a better final reward.

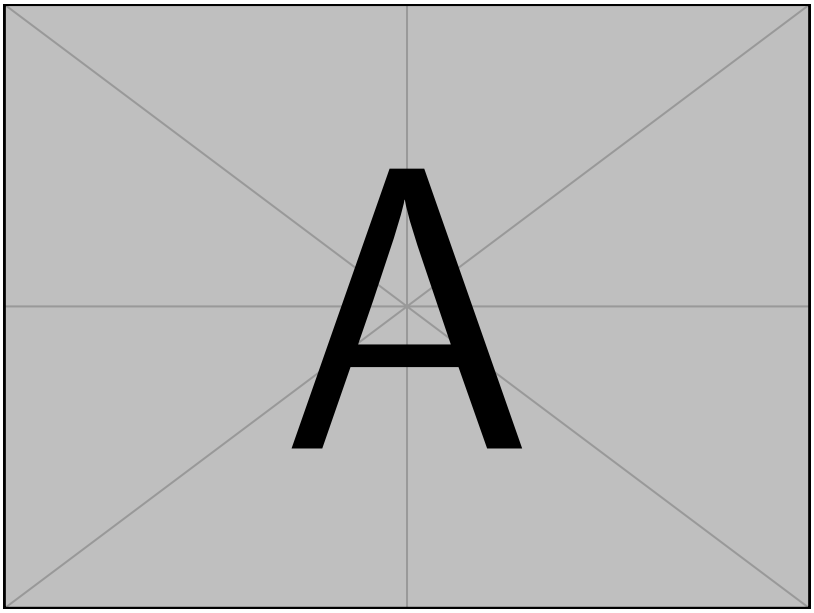
efffect of planning horizon



The agent achieved its best reward when the horizon was set to 5. Conversely, the agent received a lower reward when the horizon was 30, indicating that a lower horizon size resulted in better performance.

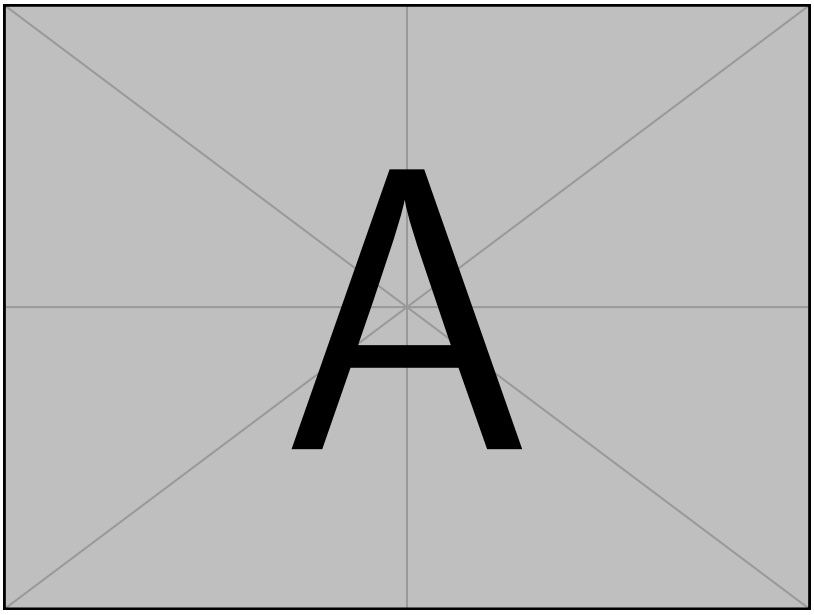
5 Problem 5: Hyper-parameter Comparison (Bonus)

Plot

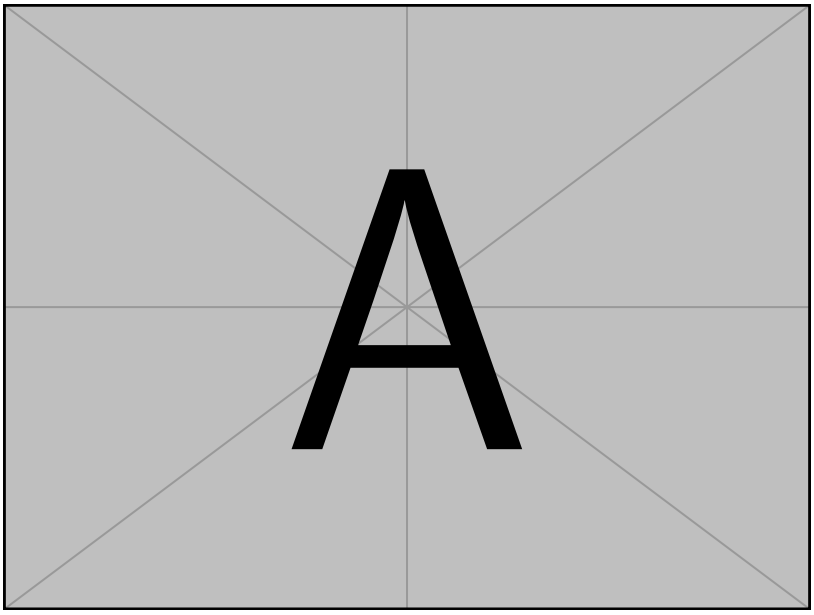


6 Problem 6: Exploration (Bonus)

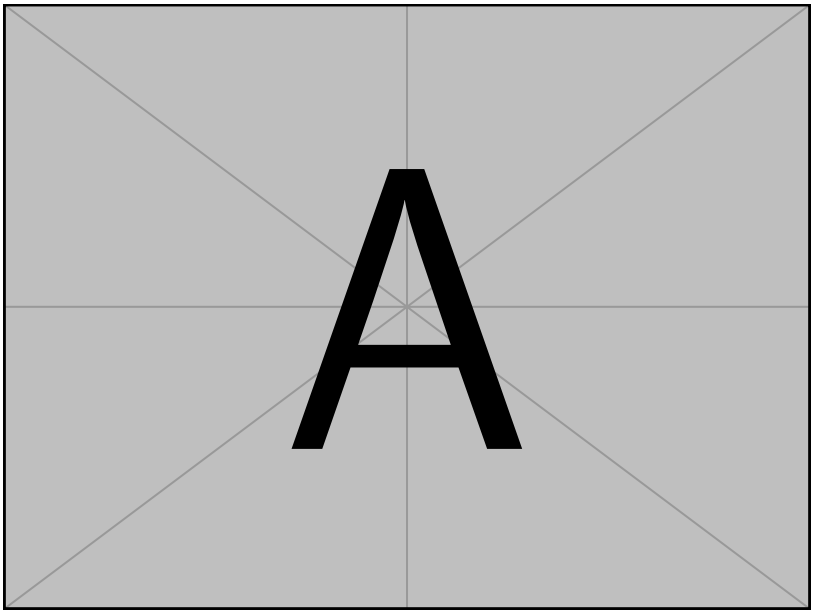
Plot



Plot



Plot



Plot