1. Q
   1. In this case a more flexible method would be better than an inflexible one. This is because of there being few predictors, decreasing the likelihood of a flexible method following the errors too closely and producing an overly complex model.
   2. In this case, the inflexible method would be optimal, as there are too many predictors to have the flexible method analyze and still result in a coherent model, while the inflexible method would be much simpler and clearly highlight the relationship between each predictor and the response.
   3. An inflexible method would not perform very well, as most would estimate the shape/function of *f* to be linear or close to linear. However, if *f* is highly non-linear, the inflexible method will never accurately reflect the true *f*, while the flexible method would note the differing shapes and curvatures in *f* better.
   4. If the variance of the error terms is extremely high, then flexible methods would follow the errors far too closely, resulting in a model with many qualities not in the true *f*. This in turn would increase the test MSE, which we wish to minimize, and therefore an inflexible method would be better suited here.
2. Q
   1. This situation would be considered regression, as the response is a quantitative value (CEO Salary). Inference is the goal here, as we wish to find which factors impact CEO salary the most and not to predict a CEO’s salary given their predictor values. (*n*=500 top firms, *p*=3)
   2. This scenario is classification since it involves qualitative responses (success or failure). Prediction is the goal here, as we wish to know whether or not a new product would be considered a success or a failure based on the predictors. (*n*=20, *p=*13)
   3. % change is quantitative and so this is a regression problem. We wish to predict the % change of the USD/Euro exchange rate in relation to the weekly changes in the world stock markets, making this a prediction situation. (*n*=52, *p*=3)
3. Q
   1. Squared Bias: squared bias steadily decreases as we go from inflexible to flexible approaches because of more linear/less flexible methods making assumptions about the shape of the true *f*, thus increasing the inherent bias.
   2. Variance: Increases as flexibility increases, due to the nature of flexible methods meaning they follow the test data more closely and changing even a single point will affect flexible methods much more than it would to inflexible ones.
   3. Training Error: Decreases as flexibility increases, since flexible methods follow the training data very closely and will have minimal to no error when analyzing the training data.
   4. Test Error: Decreases as flexibility increases until some point in the middle, and then starts to increase again. The phenomenon at play here is known as *overfitting*, as extremely flexible methods follow the patterns of the training data too closely, picking up patterns that do not exist, and so make more errors on test data than if they were less flexible. However, if a method is too inflexible, it will not pick up certain trends in the true *f* as a result of the assumed model or form of said *f*, consequently ending in poorer estimates.
   5. Irreducible Error: By definition, all estimates of *f*, no matter how accurate, will contain some inherent error that cannot be altered by changing the estimation f(x), as it is only an estimation. Therefore, irreducible error is not affected by the method in which we use to estimate *f*, nor its flexibility, meaning that the graph of irreducible error is simply a straight, horizontal line.
4. Q (Skipped)
5. Flexible approaches for regression are useful when the true *f* is very non-linear, as any inflexible method would then do a bad job of estimating *f,* especially compared to flexible methods. Few predictors are also where flexible methods thrive as they are less likely to include non-existent trends in the model. Inflexible methods are generally more useful for regression when there are many predictors, as inflexible methods tend to still remain simple despite a large number of predictor-response relationships being recorded. If *f* is not overly non-linear, inflexible methods tend to still have reasonable estimates for the true *f*.
6. Parametric approaches make assumptions about the shape of *f* to be an equation of some functional form or shape. The parameters (usually the coefficients) must then be estimated, but this is a greatly simplified task as we know what form our estimate will take, leaving only the numerical values to be computed using a training set. Conversely, non-parametric procedures make no such assumptions, allowing a wider range of possible fits for the estimate of *f*. On occasion, using parametric approaches will always result in a poor estimate for *f*, if the assumed shape is far from the actual shape of *f*. However, very large amounts of data/observations are needed to accurately estimate *f*, non-parametrically, much more than using a parametric method.
7. Let the test point be point , and be the distance between points 1 and 2, respectively.
   1. With , our prediction for point 0 would be green. Only the single nearest neighbor is taken into consideration, or the point with the least Euclidean distance from our test point. In this case, this would be point 5, which has a response of green.
   2. With , our prediction for point 0 would be red. Similar to in *b*, the 3 nearest neighbors are the three points with the least Euclidean distance from our test point. In order from least to greatest distance, we have points 5 (green), 3 (red), and 2 (red). Therefore, as of the nearest neighbors are red, the predicted response for point 0 is red.
   3. If the Bayes decision boundary is highly non-linear in actuality, then we would expect the best value for to be small. The larger is, the more neighbors we would consider for every given test point, creating more clearly defined areas with less curvature in their boundaries as increases. It is less flexible, and thus does not make an accurate prediction for the true boundary line of the data, as in this case the boundary line is very non-linear. Therefore, larger values would be worse than smaller values, so the best value would be relatively small.