* Why do the sponsors want the project in the first place? What do they lack, and what do they need?
* What are they doing to solve the problem now, and why isn’t that good enough?
* What resources will you need: what kind of data and how much staff?
* Will you have domain experts to collaborate with, and what are the computational resources?
* How do the project sponsors plan to deploy your results? What are the constraints that have to be met for successful deployment?
* What data is available to me?
* Will it help me solve the problem?
* Is it enough?
* Is the data quality good enough?

You can think of the null model as being “the obvious guess” that your model must do better than. In situations where there’s a working model or solution already in place that you’re trying to improve, the null model is the existing solution. Since this is the simplest possible model, its error rate is called the base error rate. The limit on prediction accuracy due to unexplainable variance is known as the Bayes rate. You can think of the Bayes rate as describing the best accuracy you can achieve given your data.

**Introduction**