I. The Problem

- **What is the problem you want to solve?**
- I want to build a predictive model of one-year survival after diagnosis with NSCLC (non-small cell lung cancer) using both clinical and genomic data.

II. Stakeholders and Significance

- **Who is your client and why do they care about this problem? In other words, what will your client do or decide based on your analysis that they wouldn't have done otherwise?**
- Prognostic accuracy for life expectancy is highly valued by patients, their families, and healthcare professionals. Accurate prognostication helps to determine course of treatment and helps end-of-life decision making. Hospitals, healthcare professionals, and, most importantly, patients benefit significantly from more accurate life expectancy forecasts.

III. The Data

- **What data are you using? How will you acquire the data?**
- The simulated dataset was provided by the US Department of Veteran Affairs.

IV. My Approach

- **Briefly outline how you'll solve this problem. Your approach may change later, but this is a good first step to get you thinking about a method and solution.**
- I will split the dataset into train, validation, and test sets. Cox proportional hazards regression with elastic net regularization will be used for my survival model. For model evaluation, I will use both concordance index and average partial log-likelihood.

V. Deliverables

- **What are your deliverables? Typically, this includes code, a paper, or a slide deck.**
- My deliverables will include a final report, a presentation slide deck, and all the Jupyter notebooks that I create.