1. Oral presentation (5 mins). Suggest these slides (~ 45 sec. each after title slide):

**a). Title (with your name)**

Shiwei Julia Huang

**b). Problem area – why it is of interest (in general or to you), what you might want to predict? This could be a hypothesis.**

**Interest:**

As a Chinese citizen living in the US, I am very concerned with the ongoing coronavirus outbreak in China. Wuhan coronavirus is currently spreading in China and around the world, which leads to many social issues such as the supply of medical surgical masks, treatments, vaccines, and the potential slowdown in the Chinese economy and its impact on the global economy.

I want to use the existing dataset of coronavirus in China to predict three crucial areas:

* number of patients who might be infected
* number of patients who might be severely ill

The prediction is generated by the existing dataset since the outbreak. We currently have the numbers on the following areas in the existing dataset since January 27th:

* Number of patients who are infected by Coronavirus
* Number of patients who might be infected
* Number of patients who have been cured
* Number of patients who are severely ill
* Number of death

**c). The data – where it might come from, why it may be applicable, and any preliminary assessment you’ve made?**

The data will come from baidu.com. Baidu is a Chinese multinational technology company specialized in artificial intelligence and is the largest internet company in China. Baidu has created an epidemic map which shows the real-time locations of confirmed and suspected coronavirus cases daily.

My second resource is the Johns Hopkins CSSE map. The online dashboard from Johns Hopkins uses CDC and WHO data to track the outbreak in real time.

**d). How you plan to conduct your analysis: distribution, pattern/ relationship and model construction. What techniques do you think you will use?**

I am planning to use the Susceptible-Infected-Recovered (SIR) model. An SIR model is an epidemiological model that computes the theoretical number of people infected with a contagious illness in a closed population over time. The SIR model was used during the Ebola outbreak in west Africa.

Reference:

<http://mathworld.wolfram.com/SIRModel.html>

**e). How do you plan to apply the model? What are the possible uncertainties?**

I am planning to use the data from Baidu.com to the SIR model.

The possible uncertainties come from couple sources. First, China Center for Disease Control and Prevention changes their data collection method from time to time, which causes uncertainties in our data sources. However, I have no immediate resolution to the data source since Baidu is the only public published data from the government we could use as for now.

Another uncertainty that I have includes any action from the Chinese government side. Chinese government has already taken severe control in Hubei, China. The lockdown of the city of Wuhan changes the way the disease could spread around the city. In addition, Chinese government had pledged to build a hospital within 10 days for the deceased people. If we increase the number of domestic and international medical support, we could certainly alleviate the current outbreak situation. In addition, the Chinese government might take other actions to lock down its nearby cities. If they did, the amount of death toll will certainly be lowered.

**f). What do you want to predict and what decisions (prescriptions) may be possible? What would a good outcome be?**

I want to predict the toll of disease and death in the future 2-3 months. Due to Coronavirus, a number of travel bans have been generated for Chinese citizens. For example, many universities have sanctioned travel bans to and from China. Inside China, the city of Wuhan is still being locked down. It would be crucial to understand from the data model on a potential time expectancy of the lock down and timeframe of the travel ban.