Meeting 7/13/20 Notes

1. Selecting models for the analysis
   1. Focus of the analysis should be at the county-level
   2. Purpose of modeling is to get rid of uncertainty and unnecessary noise (smooth)
      1. Dr. Bauer will send paper on alternate smoothing methods
   3. Will be separate distributions for each histologic type that’s being investigated
   4. log(uit) which is log of the true number of cases for county i in year j will be equal to the sum of the following components:
      1. alpha (overall value regardless of where/ when)
      2. **Spatial component**: (u­i + vi)
         1. This pattern is going to be the same across years
         2. Marginal
      3. **Time component**: (β\*tt)
         1. All counties; every year, the linear increase for this term is the same across all th counties
         2. Marginal
      4. **Space-time interaction component**: (𝛿i \* tt)
2. Use of mixture/ bivariate model?
   1. South Carolina study used model with a mixture parameter to compare models that took into account different levels of spatial/ spatio-temporal analysis
      1. Wanted to determine whether data regarding lung and bronchus cancer could be used to better understand the spatial patterns of other less common lung cancers.
   2. In this sort of bivariate model, we would need to make the assumption that they have similar patterns (i.e. when one type of cancer goes up, the other will as well). This makes it difficult to model the specific differences between the two types
3. Other data that can be included in the analysis
   1. Important to consider the *county-level* data, **not** the data specifically regarding cancer patients. If looking to see if a covariate (i.e. radon levels) is associated with higher lung cancer incidences, need to find radon-levels for the **entire county** rather than for cancer patients specifically as this is a county-level, **areal data** analysis
   2. tidycensus is a useful R package for some of these datasets
4. COVID-19 Mortality calculation
   1. For the most recent date, divide the total cumulative deaths by total cumulative cases
   2. Best we can do as it’s difficult to adjust for lag because there is so much variation in how COVID-19 data is reported and the lag for specific counties
5. Paper/ conference/ final end-goal
   1. Focus on CPRIT presentation as it’s what’s coming soonest
   2. Healthier TX Summit is a local conference that may be of interest
   3. <https://healthiertexassummit.com/>