Meeting 7/17/20 Notes

1. Ideas for improving the dashboard
   1. General (Page 1)
      1. Add a Texas map with county-level cases per 100k people in 1995, 2000, 2005, 2010 and 2015
      2. Use R’s lowess() function to add general trend lines to the temporal trends plots
         1. f is the smoothing span – takes in some kind of window. For example, a window might be the 3 years before and after 2010.
            1. The smaller the interval, the more sensitive to jumps in the data
         2. Add a legend that explains what the line is and what the points mean. Make the points bigger.
      3. Fix the x-axis and y-axis labels
      4. Change the legend to be categorical instead of a continuous gradient
   2. COVID-19 (Page 4)
      1. Two plots next to each other of
         1. COVID-19 mortality rates / COVID-19 case rates
         2. 2015 lung cancer rates
2. Models
   1. <https://becarioprecario.bitbucket.io/inla-gitbook/ch-spatial.html#sec:irrlattice>
      1. Section talks about besag, bym, and proper version of besag’s model which will be very applicable to this dataset
   2. Besag assumes there’s a strong spatial dependency
   3. The MLE of a poisson distribution is the SIR
   4. Dr. Bauer has data for Leroux model
   5. Would be best to focus on models that can be implemented by INLA
      1. CAR Bayes package takes a long time, would also need to write the code
   6. SIR is useful for assessing potential hotspots
   7. Bernardinelli model **assumes a linear time trend** – should try random walk or spline model as they are more flexible because the trend isn’t quite linear always
3. Assessing model fit
   1. **Moran’s I** is the correlation between 2 continuous variables
      1. Way to quantify a correlation with a single number
      2. Problem is that it’s only one number which isn’t always very informative
         1. Will say there are clusters in the data but not where those clusters are
      3. There are plenty of R packages that can implement this
   2. Use **names(res1)** where res1 is output from INLA to see whether they output WAIC or some other model adequacy assessments
4. Covariates to consider
   1. Remember that this is county-level data not individual-level data. Because of this, it’s hard to make associations on specific causations such as occupation/ pesticide-use
      1. Most air quality data is at specific locations as opposed to representing the entire county
      2. Need county-level, socioeconomic metrics for this analysis
   2. Insurance data, % below poverty line, maybe % land that’s farmland?
   3. **Rural Urban Continuum Code**: it’s play that they don’t have the data for every year, but don’t use averages of the data for missing years. Just use the most recent previous year
   4. **Poverty data**: Use the averages for the 2005 and 1995 data of 2000 and 2010, etc.