**Project #1 STAT 5385**

**Project data description and links**

**Impact and Hospitalization data descriptor:** The following dataset provides **facility-level data** for hospital utilization aggregated on a weekly basis (Friday to Thursday). These are derived from reports with facility-level granularity across two main sources: (1) HHS TeleTracking, and (2) reporting provided directly to HHS Protect by state/territorial health departments on behalf of their healthcare facilities.  
  
The hospital population includes all hospitals registered with Centers for Medicare & Medicaid Services (CMS) as of June 1, 2020. It includes non-CMS hospitals that have reported since July 15, 2020. It does not include psychiatric, rehabilitation, Indian Health Service (IHS) facilities, U.S. Department of Veterans Affairs (VA) facilities, Defense Health Agency (DHA) facilities, and religious non-medical facilities.  
  
For a given entry, the term “collection\_week” signifies the start of the period that is aggregated. For example, a “collection\_week” of 2020-11-20 means the average/sum/coverage of the elements captured from that given facility starting and including Friday, November 20, 2020, and ending and including reports for Thursday, November 26, 2020.  
  
Reported elements include an append of either “\_coverage”, “\_sum”, or “\_avg”.

* A “\_coverage” append denotes how many times the facility reported that element during that collection week.
* A “\_sum” append denotes the sum of the reports provided for that facility for that element during that collection week.
* A “\_avg” append is the average of the reports provided for that facility for that element during that collection week.

The file will be updated weekly. No statistical analysis is applied to impute non-response. For averages, calculations are based on the number of values collected for a given hospital in that collection week. Suppression is applied to the file for sums and averages less than four (4). In these cases, the field will be replaced with “-999,999”.

**Link:** <https://healthdata.gov/Hospital/COVID-19-Reported-Patient-Impact-and-Hospital-Capa/anag-cw7u>

**Vulnerability data descriptor:** This crosswalk weighs communities down to the census tract with COVID-19 data, considering disease morbidity and mortality, as well as social and economic contextual data. To do this, this spreadsheet crosswalks four key open datasets: Hardest Hit Area, Low Income Area, Tribal Community, and Rural Community, and uses the FCC's scoring methodology to weigh them to provide an evaluation of the most vulnerable communities in the country.

**Link**: <https://healthdata.gov/Health/COVID-19-Community-Vulnerability-Crosswalk-Crosswa/x2y5-9muu>

**Research Question: How is hospital capacity associated with community vulnerabilities?**

Measures of hospital bed capacity: inpatient beds 7 day avg, inpatient beds covid 7 day avg, total icu beds day avg

Measures of community vulnerability: metro/micro, poverty percentage, HHA, Tribal community, rural community

For each hospital bed capacity variable (listed above), choose the best fitting community vulnerability measure and report out best fitting model in detail.

Follow the R Markdown template for the final report.