FinalProject

Ajay Karatam and Michael Rothschilds

5/13/2020

An Analysis of U.S Hospital Bed Capacity During the COVID-19 Pandemic

by Michael Rothschilds and Ajay Karatam

Introduction:

The COVID-19 pandemic has altered the lives of millions over the past few months. People have practiced Social Distancing at an unprecedented level with the intention of keeping people safe and “flattening the curve”. The purpose of “flattening the curve” is to prevent hospitals from being overrun with too many COVID-19 patients at once. In order to stop this from happening, public health and government officials need to project the number of cases, the rate of hospitalization, and the resulting number of hospital and ICU beds needed at any given time across all areas of the United States. The following data was collected from a team of researchers at the Harvard Global Data Institute. They made projections for different infection rates and used surveys conducted by The American Hospital Association to predict the true availability of hospital and ICU beds. According to an article from US news - New York, California, and Washington are currently the most affected by the virus. With our analysis spanning across the top 300 hospital markets in the US we can answer the probability of whether hospitals in these affected regions will eventually run out of beds.

Link to the GitHub repository

https://github.com/karatam1/GlobalHospitalAnalysis (https://github.com/karatam1/GlobalHospitalAnalysis)

Examine the original dataset here:

https://www.kaggle.com/mrmorj/hospital-bed-capacity-and-covid19 (https://www.kaggle.com/mrmorj/hospital-bed-capacity-and-covid19)

R Libraries Used

```
library(tidyverse)
library(lubridate)
library(dplyr)
library(broom)
library(leaflet)
```

Part 1: Dataframe Setup and Tidying

We begin our setup by downloading the dataset into our local directory and opening here as a .csv file. Some columns that required separation were split using regex like the ones seen below. This process also included re-arranging as well as omitting some columns for the sake of readability. Furthermore, we added two more columns that served as a mathematical computation of two other columns. The final tidy dataframe is setup as “df” which is what we will be using later on in the analysis.

```
csv_file <- "HRR_Scorecard.csv"
hcb <- read_csv(csv_file)
```

```
## Parsed with column specification:
## cols(
##   .default = col_character(),
##   `Total Hospital Beds` = col_number(),
##   `Total ICU Beds` = col_number(),
##   `Available Hospital Beds` = col_number(),
##   `Potentially Available Hospital Beds*` = col_number(),
##   `Available ICU Beds` = col_double(),
##   `Potentially Available ICU Beds*` = col_number(),
##   `Adult Population` = col_number(),
##   `Population 65+` = col_number(),
##   `Projected Infected Individuals` = col_number(),
##   `Projected Hospitalized Individuals` = col_number(),
##   `Projected Individuals Needing ICU Care` = col_number(),
##   `Hospital Beds Needed, Six Months` = col_number(),
##   `Hospital Beds Needed, Twelve Months` = col_number(),
##   `Hospital Beds Needed, Eighteen Months` = col_number(),
##   `ICU Beds Needed, Six Months` = col_number(),
##   `ICU Beds Needed, Twelve Months` = col_number(),
##   `ICU Beds Needed, Eighteen Months` = col_number()
## )
```

```
## See spec(...) for full column specifications.
```

```
#delete the first entry since it contains garbage values
hcb = hcb[-1,]
df <- hcb %>%
  #choose the relevant columns that we want to work with from hcb
  select(1:12)
df
```

```
## # A tibble: 305 x 12
##    HRR   `Total Hospital~ `Total ICU Beds` `Available Hosp~ `Potentially Av~
##    <chr>            <dbl>            <dbl>            <dbl>            <dbl>
##  1 Abil~              980              127              565              772
##  2 Akro~             1358              186              518              938
##  3 Alam~             2695              293              665             1680
##  4 Alba~              704               60              221              462
##  5 Alba~             4804              425             1579             3191
##  6 Albu~             2908              380             1102             2005
##  7 Alex~              917               43              402              660
##  8 Alle~             3267              334             1267             2267
##  9 Alto~              555               61              234              394
## 10 Amar~             1236              194              678              957
## # ... with 295 more rows, and 7 more variables: `Available ICU Beds` <dbl>,
## #   `Potentially Available ICU Beds*` <dbl>, `Adult Population` <dbl>,
## #   `Population 65+` <dbl>, `Projected Infected Individuals` <dbl>, `Projected
## #   Hospitalized Individuals` <dbl>, `Projected Individuals Needing ICU
## #   Care` <dbl>
```

```
#extract the state as a separate column from HRR
df$State <- str_extract(df$HRR, "([A-Z]{2})")

#extract the town as a separate column from HRR
df$Town <- sub(", [A-Z]{2}$", "", df$HRR)

#re-arrange the columns to make the data more presentable
df <- df[c(14,13,2,4,5,3,6,7,8,9,10,11,12)]

#calculate the percentage of occupied hospital beds
df$calc_hospital = ((df[c(4)] / df[c(3)])*100)
df$`Occupied Hospital Beds percentage` <- round(df$calc_hospital$`Available Hospital Beds`,digits=2)

#calculate the percent of occupied ICU beds
df$calc_ICU = ((df[c(7)] / df[c(6)])*100)
df$`Occupied ICU Beds percentage` <- round(df$calc_ICU$`Available ICU Beds`, digits=2)

#final columns re-arrangement
df <- df[c(1,2,3,4,5,15,6,7,8,17,9,10,11,12,13)]
head(df)
```

```
## # A tibble: 6 x 15
##   Town  State `Total Hospital~ `Available Hosp~ `Potentially Av~
##   <chr> <chr>            <dbl>            <dbl>            <dbl>
## 1 Abil~ TX                 980              565              772
## 2 Akron OH                1358              518              938
## 3 Alam~ CA                2695              665             1680
## 4 Alba~ GA                 704              221              462
## 5 Alba~ NY                4804             1579             3191
## 6 Albu~ NM                2908             1102             2005
## # ... with 10 more variables: `Occupied Hospital Beds percentage` <dbl>, `Total
## #   ICU Beds` <dbl>, `Available ICU Beds` <dbl>, `Potentially Available ICU
## #   Beds*` <dbl>, `Occupied ICU Beds percentage` <dbl>, `Adult
## #   Population` <dbl>, `Population 65+` <dbl>, `Projected Infected
## #   Individuals` <dbl>, `Projected Hospitalized Individuals` <dbl>, `Projected
## #   Individuals Needing ICU Care` <dbl>
```

==

Part 2: Data Analysis With a Focus on Hospital and ICU Bed Capacity

Given that our data deals with a statistical analysis of the hospital beds and ICU beds availability as well as population age distribtion across the top 300 US hospital markets; we decided to split out analysis into 3 parts to offer a detailed story.

Step 2.1: Hospital and ICU Dataframe Setup

Most of the attributes for the Hospital and ICU are very similar ex: ‘Hospital Beds Availble’, ‘ICU Beds Available’ or ‘Total Hospital Beds’, ‘Total ICU beds’. It made sense to pull the relevant attributes from the tidy dataframe (df) and used it in our analysis for the Hospital and ICU. In the process of setting up this new dataframe, we decided to combine the regions for every state as one entry, we achieved this by adding up all the attributes ex: ‘Total Hospital Beds’, ‘Total ICU Beds’. That way we end up with a more interpretable dataframe of 51 entries and 10 columns. Combining the attributes helps us generalize the analysis to each state rather than every region and this was what we were aiming to do with this project to begin with. The final two columns that I added were the rate of hospital beds availablity and rate of icu beds availability, both of these attributes help us in understanding how each state’s hospital and icu ward compare with one another.

```
beds_df <- df %>%
  select(1:10)%>%
  group_by(State)%>%
  summarize(`No. of Regions` = n_distinct(Town),
            `Total Hospital Beds` = sum(`Total Hospital Beds`),
            `Hospital Beds Available` = sum(`Available Hospital Beds`),
            `Potential Hospital Beds Available` = sum(`Potentially Available Hospital Beds*`),
            `Total ICU Beds` = sum(`Total ICU Beds`),
            `ICU Beds Available` = sum(`Available ICU Beds`),
            `Potential ICU Beds Available` = sum(`Potentially Available ICU Beds*`)
            )%>%
  mutate(`Rate of Hospital beds availability` = (`Hospital Beds Available`/`Total Hospital Beds`)*100)%>%
  mutate(`Rate of ICU beds availability` = (`ICU Beds Available`/`Total ICU Beds`)*100)%>%
  select(1,2,3,4,5,9,6,7,8,10)%>%
  arrange(State)
beds_df
```

```
## # A tibble: 51 x 10
##    State `No. of Regions` `Total Hospital~ `Hospital Beds ~ `Potential Hosp~
##    <chr>            <int>            <dbl>            <dbl>            <dbl>
##  1 AK                   1             1583              533             1058
##  2 AL                   6            14793             5282            10037
##  3 AR                   5             8560             4063             6311
##  4 AZ                   4            12590             4763             8676
##  5 CA                  24            68074            22585            45328
##  6 CO                   7            10335             4417             7376
##  7 CT                   3             7034             1731             4382
##  8 DC                   1             5055             1595             3325
##  9 DE                   1             1845              601             1223
## 10 FL                  17            53744            18464            36105
## # ... with 41 more rows, and 5 more variables: `Rate of Hospital beds
## #   availability` <dbl>, `Total ICU Beds` <dbl>, `ICU Beds Available` <dbl>,
## #   `Potential ICU Beds Available` <dbl>, `Rate of ICU beds availability` <dbl>
```

```
beds_df$`Rate of Hospital beds availability` = round(beds_df$`Rate of Hospital beds availability`, digits=2)
beds_df$`Rate of ICU beds availability` = round(beds_df$`Rate of ICU beds availability`, digits=2)
```

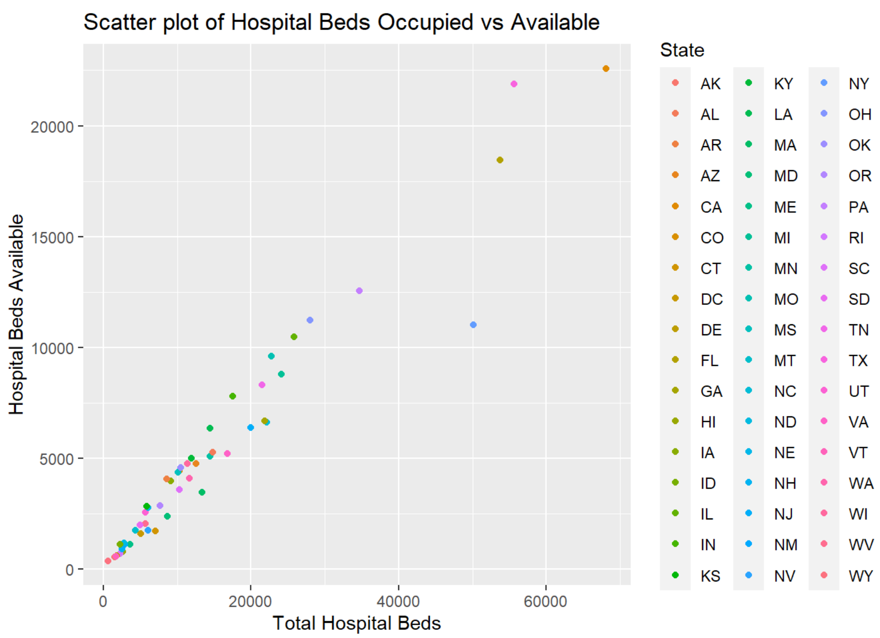
Step 2.2: Hospital Analysis - Scatter plot of Hospital Beds Occupied vs Available

To interprete this graph, we know that Hospital Beds Available <= Total Beds Available as a universal truth, therefore, a point which is higher on the y-axis is an indication that there are more beds available. Additionally, the closer the x axis is to 0 and a higher y axis point means that the hospital is operating extremely efficiently. From this plot, there seems be more concentration of points around x <= 20,000 and y <= 5,000; this means that roughly 25% hospital beds are available for most of these states. Ofcourse there are a few states which have more total beds but a higher ratio of occupied beds (>25%).

```
  hosp_df <- beds_df %>%
    select(1:6)%>%
    group_by(State, `No. of Regions`)%>%
    arrange(`Rate of Hospital beds availability`)
  hosp_df
```

```
## # A tibble: 51 x 6
## # Groups:   State, No. of Regions [51]
##    State `No. of Regions` `Total Hospital~ `Hospital Beds ~ `Potential Hosp~
##    <chr>             <int>            <dbl>            <dbl>            <dbl>
##  1 NY                   10            50102            11003            30554
##  2 CT                    3             7034             1731             4382
##  3 MA                    3            13352             3473             8412
##  4 MD                    3             8710             2368             5539
##  5 NV                    2             6051             1748             3899
##  6 NC                    9            22158             6610            14384
##  7 HI                    1             2623              795             1709
##  8 GA                    7            21861             6681            14270
##  9 ME                    2             3618             1110             2364
## 10 RI                    1             2249              690             1470
## # ... with 41 more rows, and 1 more variable: `Rate of Hospital beds
## #   availability` <dbl>
```

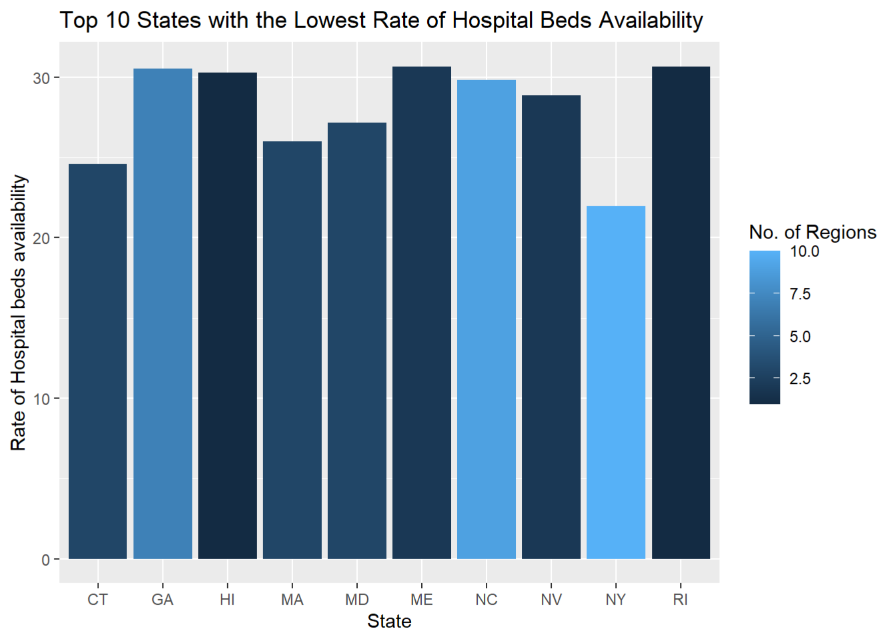
```
ggplot(hosp_df, mapping=aes(x=`Total Hospital Beds`, y=`Hospital Beds Available`))+
  geom_point(mapping=aes(color = State))+
  ggtitle("Scatter plot of Hospital Beds Occupied vs Available")
```



Step 2.3: Hospital Analysis - Bargraph of the Top 10 States with the Lowest Rate of Hospital Beds Availability

The states included in this graph gives the reader an understanding of the population demographics for these regions, it is easy to guess that there could be a significant older adult population. Also the number of regions for each of these states is another indicator of the intensity of beds occupancy. From the graph, New York has the most hospitals as well as the least availability, Rhode Island also has fewer hospitals and a tad lower availability rate. An interesting feature about the states in this list is the fact that most of them belong in the East coast.

```
hosp_df[1:10,] %>%
  ggplot(mapping=aes(x=State, y=`Rate of Hospital beds availability`, fill= `No. of Regions`))+
  geom_col(mapping=aes())+
  ggtitle("Top 10 States with the Lowest Rate of Hospital Beds Availability")
```



Step 2.4: Hospital Analysis - Linear Regression Model of Hospital: Potential Available Beds vs Available Beds

This regression plot features a unique attribute which is “Potential Available Beds”, this attribute was part of our original dataset and it is a numerical value that represents the scenario if non-covid patients took up 50% less beds. By plotting this against the current available beds, the regression analysis will help us understand the correlation. Off the bat, it seems like the concentration lies around lower x and y values. The straight linear regression curve is a strong indication that many hospitals can promise 50% more hospital beds.

```
ggplot(hosp_df, mapping=aes(x=`Potential Hospital Beds Available`, y=`Hospital Beds Available`))+
  geom_point(mapping=aes(color = State))+
  geom_smooth(method=lm)+
  ggtitle("Linear Regression Model of Hospital: Potential Available Beds vs Available Beds")
```

##

===

Part 3: ICU Analysis

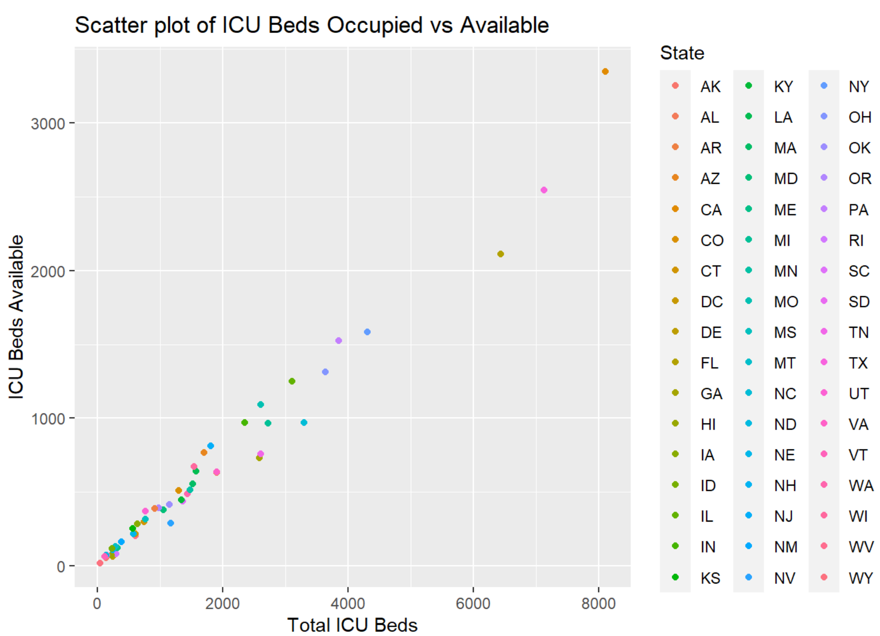
Step 3.1: Scatter plot of ICU Beds Occupied vs Available

ICU beds occupancy is an interesting factor to observe simply because of its dependent nature, in other words, ICU beds occupancy is dependant on demographic factors like number of old people or number of hospitals in a given state. From the graph below, similar to the results from the Hospital beds version of this graph, has a lower concentration at lower x and y values. From color inspection is seems like most states are maintaining similar hospital beds and icu beds availability rate. The states within the concetration seem to show that roughy 50% of the icu beds are available. States with more hospital beds show lower availability rate (~30%).

```
icu_df <- beds_df %>%
  select(1,2,7,8,9,10)%>%
  group_by(State, `No. of Regions`)%>%
  arrange(`Rate of ICU beds availability`)
icu_df
```

```
## # A tibble: 51 x 6
## # Groups:   State, No. of Regions [51]
##    State `No. of Regions` `Total ICU Beds` `ICU Beds Avail~ `Potential ICU ~
##    <chr>            <int>            <dbl>            <dbl>            <dbl>
##  1 NV                   2             1167              290              729
##  2 DE                   1              237               62              149
##  3 RI                   1              293               82              188
##  4 GA                   7             2582              730             1656
##  5 TN                   7             2601              760             1679
##  6 NC                   9             3294              970             2133
##  7 SC                   5             1358              437              897
##  8 FL                  17             6433             2113             4269
##  9 AL                   6             1903              633             1268
## 10 HI                   1              219               73              146
## # ... with 41 more rows, and 1 more variable: `Rate of ICU beds
## #   availability` <dbl>
```

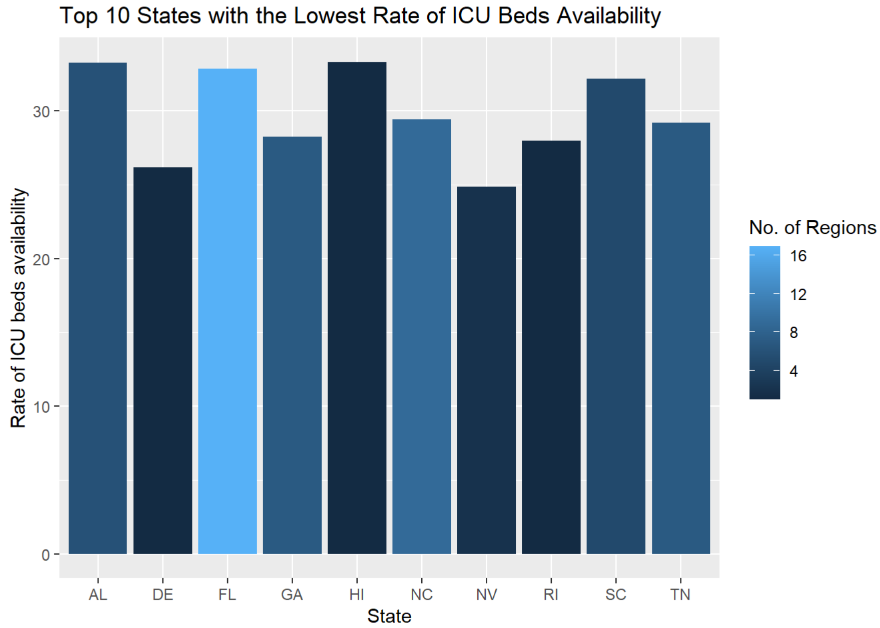
```
ggplot(icu_df, mapping=aes(x=`Total ICU Beds`, y=`ICU Beds Available`))+
  geom_point(mapping=aes(color = State))+
  ggtitle("Scatter plot of ICU Beds Occupied vs Available")
```



Step 3.2: Bargraph of the Top 10 States with the Lowest Rate of ICU Beds Availability

This plot just like the hospital plot version, compares the top 10 lowest ICU beds availability rates with respect to region. From observations, Nevada has the lowest availability rate while also having the least number of regions, Florida on the other hand has roughly 8% higher availability rate and significantly more regions. It is interesting to find that Georgia, Rhode Island, Nevada, Hawaii, North Carolina; all of which were featured in both hospital and ice bargraphs; this is an indication of hospital inefficiency in these regions as well higher demand.

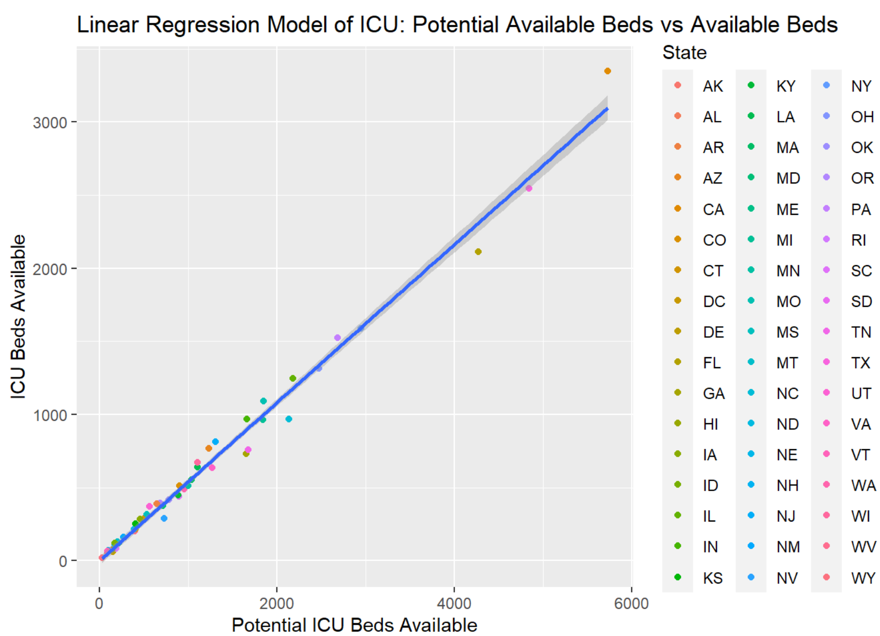
```
icu_df[1:10,] %>%
  ggplot(mapping=aes(x=State, y=`Rate of ICU beds availability`, fill= `No. of Regions`))+
  geom_col(mapping=aes())+
  ggtitle("Top 10 States with the Lowest Rate of ICU Beds Availability")
```



Step 3.3: Linear Regression Model of ICU: Potential Available Beds vs Available Beds

As discussed earlier in the hospital plot, the potential available beds attribute introduces an efficient method to increase hospital beds occupancy for covid patients. Interestingly enough, the linear regression is similar to the one we observed earlier, this means that most hospitals can promise 50% more ICU beds for most states in the concentraion. However, unlike hospital beds, ICU beds can be optimized to offer 50% more even for hospitals with 4000 or more beds. This is a nod to the 1:2 ratio nature of the relation.

```
ggplot(icu_df, mapping=aes(x=`Potential ICU Beds Available`, y=`ICU Beds Available`))+
  geom_point(mapping=aes(color = State))+
  geom_smooth(method=lm)+
  ggtitle("Linear Regression Model of ICU: Potential Available Beds vs Available Beds")
```



##

==

Part 4: Data Analysis with a Focus on State Population and Projected Infection Rates

The purpose of this section is to take the entire dataset and use it to create a smaller dataset that focuses on the population.

Step 4.1:

Obtain this smaller dataset that includes Town, State, Adult Population, Population 65+, Projected Infected Individual, Projected Hospitalized Individuals, and Projected Individuals Needing ICU Care.

```
pop_df <- df %>% select(Town, State, `Adult Population`, `Population 65+`,
`Projected Infected Individuals`, `Projected Hospitalized Individuals`,
`Projected Individuals Needing ICU Care`)
```

Step 4.2:

Turn the region data into statewide data by grouping by state and using summarize to add the totals for each state. The statewide data allows the opportunity to compare the risks that states are facing based on total population. The dataset contains all of the large regional hospitals, so we felt that the transition to statewide data would be seamless.

```
states_pop_df <- pop_df %>%
  group_by(State) %>%
  summarize(Adult_Population = sum(`Adult Population`),
  `Population 65+` = sum(`Population 65+`),
  Projected_Infected_Individuals = sum(`Projected Infected Individuals`),
  Projected_Hospitalized_Individuals = sum(`Projected Hospitalized Individuals`),
  Projected_ICU_Care = sum(`Projected Individuals Needing ICU Care`))
```

Step 4.3:

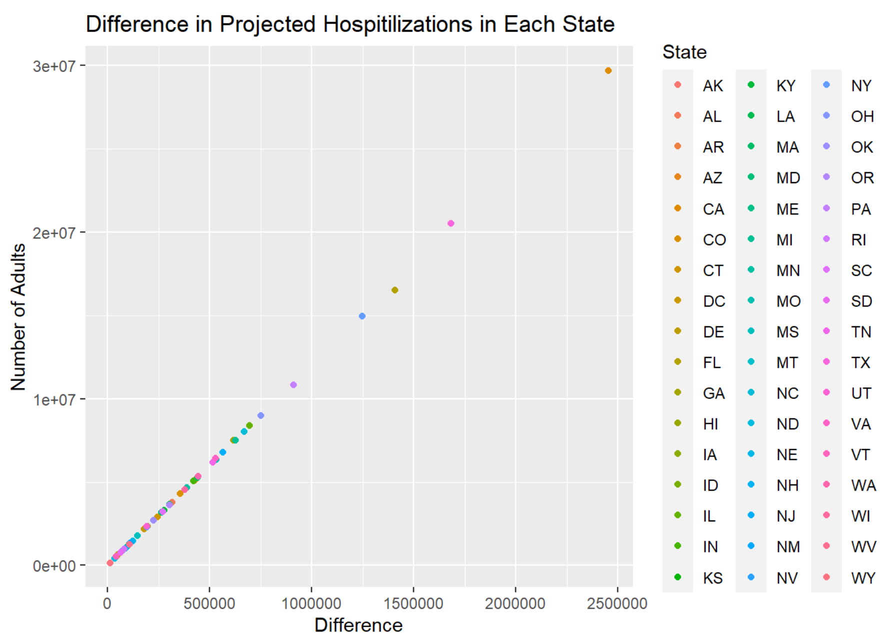
Once statewide data exists, we next wanted to show how much the risk would increase if 60 percent of the adult population contracted the virus in each state. In order to provide a snapshot, we tripled the number of Projected Infected Individuals, Projected Hospitalized Individuals, and Projected Individuals Needing ICU Care. This represents a 200 percent increase in each category over the original twenty percent. This is a plausible estimate according to various projection models.

```
states_pop_df <- states_pop_df %>%
  mutate(DoomsDay_Projected_Infected_Individuals = Projected_Infected_Individuals * 3,
  Doomsday_Projected_Hospitalized_Individuals = Projected_Hospitalized_Individuals *3,
  Doomsday_Projected_ICU_Care = Projected_ICU_Care * 3)
```

Step 4.4:

After doing this, we created a scatterplot showing the difference in the number of hospitalizations with the exact same adult population to emphasize how much worse this plausible scenario could make the situation. This scatterplot highlighted the risks that higher population states face if they aren't properly equipped.

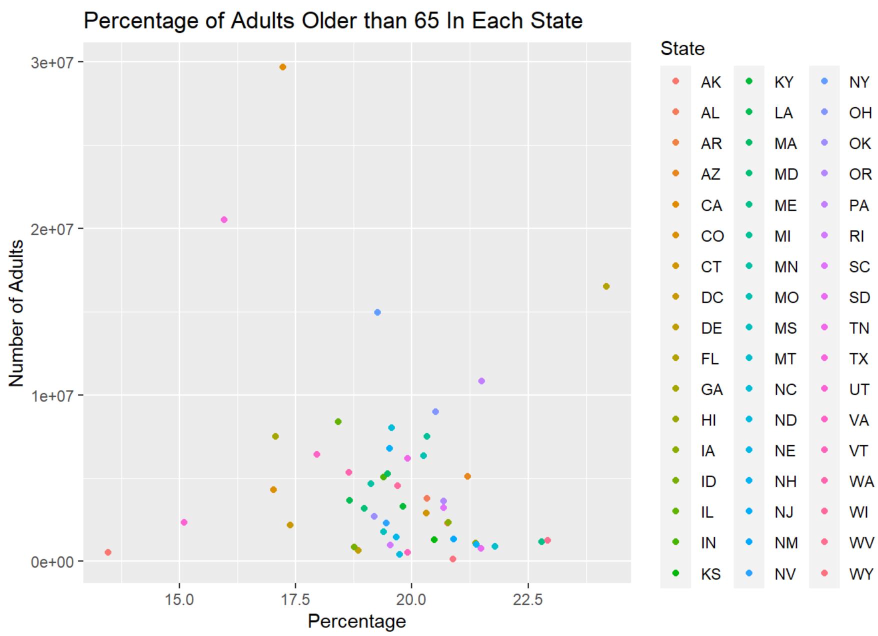
```
states_pop_df %>%
  mutate(difference_in_hospitilzations =
  Doomsday_Projected_Hospitalized_Individuals - Projected_Hospitalized_Individuals) %>%
  ggplot(mapping=aes(x = difference_in_hospitilzations, y= Adult_Population, color= State))+
  geom_point()+
  labs(title = "Difference in Projected Hospitilizations in Each State",
  x="Difference", y="Number of Adults")
```



Step 4.5:

When focusing solely on population, one of the most important things to consider is the proportion of people who are older than 65. According to the CDC, this demographic is at highest risk for this virus as they are much more likely to be hospitalized when they contract the virus. In order to depict this across the fifty states, we created a scatterplot with the adult population mapped to the y-axis and the percentage of adults older than 65 on the x axis for each state.

```
states_pop_df %>%
  mutate(Percentage_Adults_Older_65 = (`Population 65+` / Adult_Population) * 100) %>%
  ggplot(mapping=aes(x=Percentage_Adults_Older_65, y= Adult_Population, color = State))+
  geom_point()+
  labs(title = "Percentage of Adults Older than 65 In Each State", x="Percentage",
  y="Number of Adults")
```



Step 4.6: Furthermore, we used the proportion of adult population and the projected number of hospitalizations to calculate a risk level by state just according to the population. The purpose of this is to depict how a more frequent older population puts a state at risk. When you combine that with questionable hospital supplies, a state can be in major trouble. A study done by researchgate details the struggles hospitals go through to compensate for a lack of supplies. It decreases the probability that a patient will receive the best possible care when in critical condition.

```
states_pop_df <- states_pop_df %>%
  mutate(Risk_Level = (Projected_Hospitalized_Individuals + `Population 65+`) / Adult_Population) %>%
  arrange(desc(Risk_Level))

select(states_pop_df, State, Risk_Level)
```

```
## # A tibble: 51 x 2
##    State Risk_Level
##    <chr>      <dbl>
##  1 FL         0.284
##  2 WV         0.272
##  3 ME         0.270
##  4 MT         0.260
##  5 PA         0.257
##  6 SD         0.257
##  7 NH         0.256
##  8 HI         0.256
##  9 AZ         0.254
## 10 NM         0.251
## # ... with 41 more rows
```

===

Part 5: Merging the Two Data Frames To Calculate a Preparedness Score

Step 5.1:

We merged the beds data frame from parts 2 and 3 with the population data frame from part 4.

```
merge_df <- merge(beds_df, states_pop_df, by="State")
```

Step 5.2:

In order to quantify which states were the most prepared and which states were the least prepared, we used the merged dataset to calculate a preparedness score. In order to calculate the score, we derived a formula that uses the percentage of population 65 and older, the rate of hospital bed availability, and the rate of ICU bed availability. We felt that these factors differentiated the states the most and concluded that the rate of available hospital and ICU beds was more than twice as important as the percentage of elderly population in a state. This formula produced a score on a scale of 0 to 100 and is represented by the continuous attribute "preparedness_score". The states who were the most prepared had scores in the 65-90 range, while the states who were the least prepared had scores in the 43-55 range.

```
merge_df <- merge_df %>%
  mutate(Preparedness_Score = ((`Population 65+`/Adult_Population)*.3 +
                         ((`Rate of Hospital beds availability`/100) +
                         (`Rate of ICU beds availability`/100)) *.70)*100)
```

Step 5.3:

Once we calculated our preparedness_score, we created a separate data frame to prepare for Logistic Regression. This data frame includes all of the pertinent variables in our regression model, including the building blocks for the calculation of each preparedness_score.

```
fit_df <- select(merge_df, State, `Rate of Hospital beds availability`,
`Rate of ICU beds availability`, Adult_Population, `Population 65+`, Preparedness_Score)

fit_df <- fit_df %>%
  mutate(Percentage_65_older = (`Population 65+` / Adult_Population) * 30,
  Hospital_Bed_Availability = ((`Rate of Hospital beds availability`/100) +
  (`Rate of ICU beds availability`/100))*70)
```

Step 5.4:

We perform logistic regression using the fit_df created in step 5.3. The purpose of doing this is to quantify the relationship between preparedness_scores and the two independent variables. The preparedness score is the dependent variable whilethe Percentage_65_older and the Hospital_Bed_Availability are the independent variables. Using the broom library, the tidy function with our model produces a p value of 0.25, which tells us that our relationship is relatively strong, but not strong enough to be statistically significant.

```
lm_df <- glm(Preparedness_Score ~ Percentage_65_older + Hospital_Bed_Availability, data=fit_df)
broom::tidy(lm_df)
```

```
## # A tibble: 3 x 5
##   term                      estimate std.error statistic p.value
##   <chr>                        <dbl>     <dbl>     <dbl>   <dbl>
## 1 (Intercept)              -2.39e-14  2.05e-14  -1.16e 0   0.250
## 2 Percentage_65_older       1.00e+ 0  3.08e-15   3.25e14   0
## 3 Hospital_Bed_Availability  1.00e+ 0  1.91e-16   5.22e15   0
```

```
lm_df%>%
  tidy()%>%
  knitr::kable(digits=4)
```

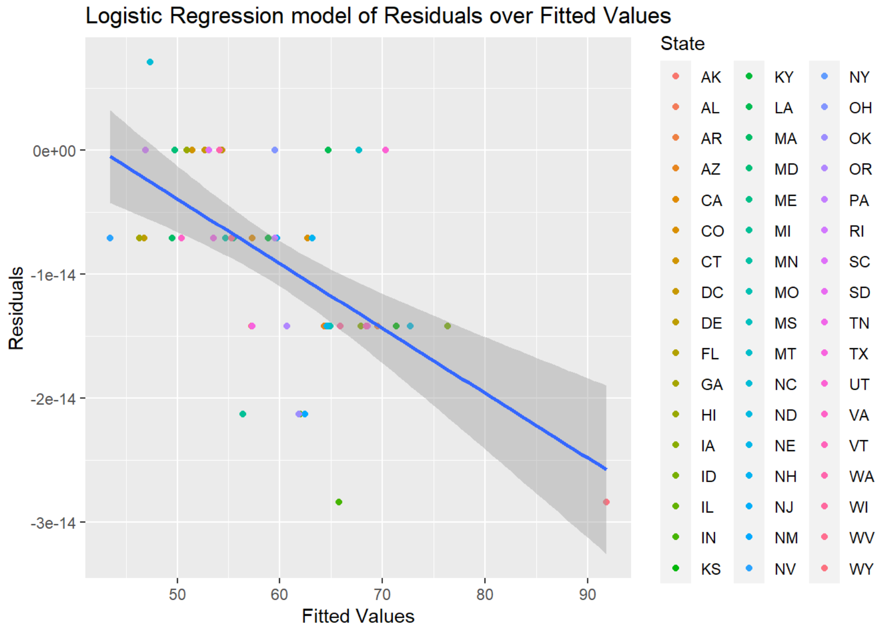
term	estimate	std.error	statistic	p.value
(Intercept)	0	0	-1.164000e+00	0.2502
Percentage_65_older	1	0	3.251182e+14	0.0000
Hospital_Bed_Availability	1	0	5.222330e+15	0.0000

Step 5.5:

We used the results of the logistic regression to graph the residuals versus the fitted values. Our residuals appear to skew a little below zero, but they are within a reasonable range of zero.

```
ggplot(lm_df, mapping=aes(x=lm_df$fitted.values, y=lm_df$residuals))+
    geom_point(mapping=aes(color = lm_df$data$State))+
    geom_smooth(method=lm)+
    labs(title="Logistic Regression model of Residuals over Fitted Values",
        y = "Residuals",
         x = "Fitted Values",
        color = "State")
```

```
## `geom_smooth()` using formula 'y ~ x'
```




```
fit_df$fitted_values <- lm_df$fitted.values
```

===

Part 6: Visually Depicting the Tiers of Preparedness_Score

In order to assess each state's preparedness score with respect to one another, we decided to utilize the leaflet library. We started off by downloading a csv file from kaggle with the co-ordinates for each state and then merged it with our data to create a new data frame (map_df). Using the preparedness score for each state, we color coded them in order to show visually which states could be at risk. As the preparedness score ranges roughly from 40-95, with 40 being considered least prepared, the color sequence follows that range as well with Red indicating a state is poorly equipped where as blue is an indication of a strongly equipped state (the sequence follows Red->Orange->Yellow->Green->Blue). An interesting feature of the map is the popup option, clicking a circle for any state will reveal details about it's preparedness score. By examining a map containing impacted states directly from CDC.gov, our map analysis is in sync in terms of impacted states. Our map shows that most east-coast states were at higher risk and this reflects very accurately with the map from the CDC website, additionally, states in the mid-west of America shown to have a higher preparedness score which once again happens to be a mirror reflection of the map from the CDC.

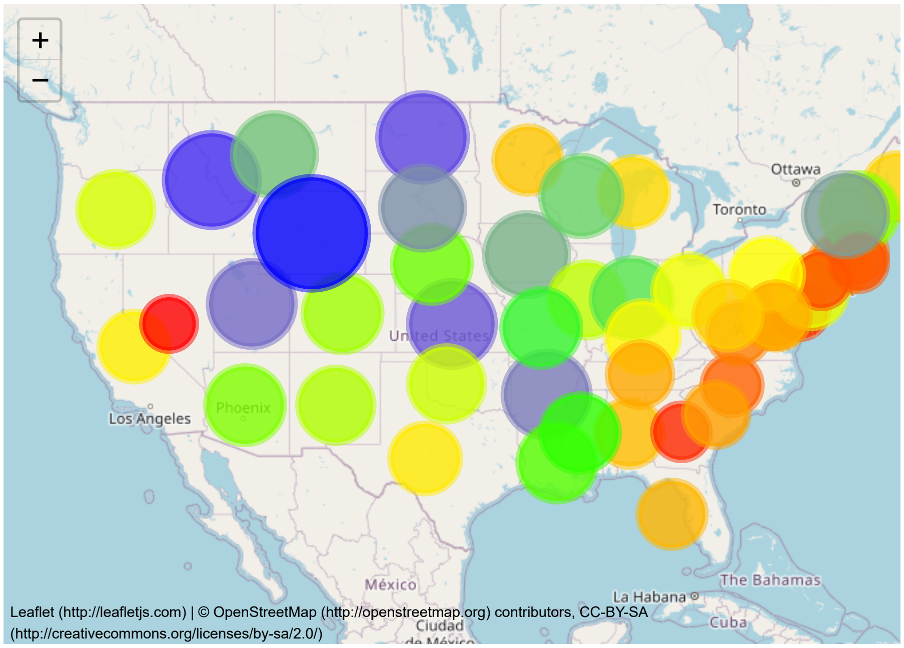
```
#https://www.kaggle.com/washimahmed/usa-latlong-for-state-abbreviations
csv_file <- "statelatlong.csv"
sll <- read_csv(csv_file)
```

```
## Parsed with column specification:
## cols(
##   State = col_character(),
##   Latitude = col_double(),
##   Longitude = col_double(),
##   City = col_character()
## )
```

```
map_df <- merge(sll, fit_df, by = "State")

map_df$Preparedness_Score = round(map_df$Preparedness_Score, digits = 2)
map_df$Preparedness_Score = as.character(map_df$Preparedness_Score)
map_df$Preparedness_Score = paste(", Preparedness Score = ", map_df$Preparedness_Score)
map_df$Preparedness_Score = paste(map_df$City , map_df$Preparedness_Score)
map_df$Preparedness_Score = paste(map_df$Preparedness_Score, ", Adult Population = ")
map_df$Adult_Population = as.character(map_df$Adult_Population)
map_df$Preparedness_Score = paste(map_df$Preparedness_Score, map_df$Adult_Population)
map_df$Preparedness_Score = paste("State = ",map_df$Preparedness_Score)

pal <- colorFactor(c("red", "orange", "yellow", "green", "blue"), domain = c(map_df$fitted_values))

prep_map <- leaflet(map_df) %>%
  addTiles() %>%
  #Longitude and Latitude of Kansas was used as the default map co-ordinates.
  setView(lat=38.49, lng=-98.32, zoom=4) %>%
  addCircleMarkers(radius = ~map_df$fitted_values/2.2, popup = map_df$Preparedness_Score,
                   color = ~pal(map_df$fitted_values), fillOpacity = 0.8)
```

```
## Assuming "Longitude" and "Latitude" are longitude and latitude, respectively
```

```
prep_map
```



==

Part7: Conclusion

Covid-19 rocked the world over the past few months and many people are suffering from the virus as we speak, this is the worst nightmare for many hospitals around the US. By inspecting the beds occupancy rate for the top 300 US hospital markets, we were able to determine that states with low adult population, specifically population with ages over 65; those regions had a low occupancy rate and therefore are better prepared to handle a sudden surge in cases. The safest states that we determined happened to be located around mid-west US: Wyoming, Utah, Kansas, Idaho. With that being said, the dichotomy between states which are well-prepared and not is drastic. States along the east-coast -> south: Rhode Island, New York, Delaware, Washington D.C, Georgia, Florida; all these states are least prepared against the virus and this is due to a combination of high rates of hospital occupancy and high population of old-age people. Our findings are accurate based on the current scenario when comparing it with the map from CDC.gov which pointed out the states that are most affected by the virus, however, these scenarios are subject to change if proper pre-cautions are not taken in place to “lower the curve”.

Part 8: References

Main dataset link = https://www.kaggle.com/mrmorj/hospital-bed-capacity-and-covid19 (https://www.kaggle.com/mrmorj/hospital-bed-capacity-and-covid19)

Dataset link for longitudes and latitudes = https://www.kaggle.com/washimahmed/usa-latlong-for-state-abbreviations (https://www.kaggle.com/washimahmed/usa-latlong-for-state-abbreviations)

“Cases in the U.S”. (2020, May 7). Retrieved from https://www.cdc.gov/coronavirus/2019-ncov/cases-updates/cases-in-us.html (https://www.cdc.gov/coronavirus/2019-ncov/cases-updates/cases-in-us.html)

“U.S. Hospital Beds Were Already Maxed Out Before Coronavirus Pandemic.” (2020, March 26) Retrieved from https://www.usnews.com/news/health-news/articles/2020-03-26/us-hospital-beds-were-already-maxed-out-before-coronavirus-pandemic (https://www.usnews.com/news/health-news/articles/2020-03-26/us-hospital-beds-were-already-maxed-out-before-coronavirus-pandemic)

“COVID-19 Guidance for Older Adults.” Centers for Disease Control and Prevention, Centers for Disease Control and Prevention, 18 May 2020, www.cdc.gov/aging/covid19-guidance.html.

Stowell, Andrew, et al. “Hospital out-Lying through Lack of Beds and Its Impact on Care and Patient Outcome.” Scandinavian Journal of Trauma, Resuscitation and Emergency Medicine, vol. 21, no. 1, 2013, doi:10.1186/1757-7241-21-17 (doi:10.1186/1757-7241-21-17).