DA5020 Practicum 2: Report for Hospital Nursing Intervention Pilot Program

Group 1: Nishat Mohammad and Victoria Clendaniel

03/06/2024

Introduction:

ACME Integrated Delivery System, a network of healthcare organizations, prioritizes data-driven insights to improve on patient care and operational efficiency.

Critical care needs have warranted leadership plans to increase nursing staff in ICUs and SICUs, backed by some research on nurse-to-patient ratios.

To ensure cost-effective deployment, the focus is directed towards hospitals with ample ICU/SICU bed capacity that is assessed by licensed, census, and staffed beds.

Our task is to identify Top 10 Hospitals by ICU/SICU bed count volume and clarify the data set facts versus dimensions.

This report aims to:
i) Inform strategic decision-making
ii) Optimize resource allocation
iii) Enhance critical care services within ACME's healthcare network.

1. Identify the dimensions from each dimension table

We have been provided with two dimensional data sets:
1. Bed Type data
2. Business data

We will look at these on a high level at the dimensional tables and make few comments after the code chunk below.

```
library(RSQLite)
# Make a db for this practicum
conekshun <- dbConnect(SQLite(), dbname = "group1practicum2.sqlite")

# Get bed-type data from db
bed_type_dt <- read.csv("bed_type-1.csv", header = TRUE)
dbWriteTable(conekshun, "bed_type", bed_type_dt, overwrite = TRUE)
# Dimensions in bed_type
bed_type_tbl <- dbGetQuery(conekshun, "SELECT * FROM bed_type")
head(bed_type_tbl)
```

```
##   bed_id bed_code     bed_desc
## 1      1       BU         Burn
```

```
## 2      2       CC          CCU
## 3      3       DE    Detox ICU
## 4      4       IC          ICU
## 5      5       MS     Med/Surg
## 6      6       NE NeoNatal ICU
```

```
str(bed_type_tbl)
```

```
## 'data.frame':    20 obs. of  3 variables:
##  $ bed_id  : int  1 2 3 4 5 6 7 8 9 10 ...
##  $ bed_code: chr  "BU" "CC" "DE" "IC" ...
##  $ bed_desc: chr  "Burn" "CCU" "Detox ICU" "ICU" ...
```

```
bed_typ_dims <- dim(bed_type_tbl)
bed_typ_cols <- colnames(bed_type_tbl)
any(is.na(bed_type_tbl))
```

```
## [1] FALSE
```

```
# Get business data from db
biznes_dt <- read.csv("business-1.csv", header = TRUE)
dbWriteTable(conekshun, "business", biznes_dt, overwrite = TRUE)
# Dimensions in business
bizines_tbl <- dbGetQuery(conekshun, "SELECT * FROM business")
head(bizines_tbl)
```

```
##    ims_org_id                      business_name ttl_license_beds
## 1 INS00077200     140 Prescott Street Corporation              126
## 2 INS00000594                 366th Medical Group               10
## 3 INS00039181            7 Hills Pediatric Center               83
## 4 INS00011388 92 Brickroad Operating Company, LLC               49
## 5 INS00000593                  96th Medical Group               53
## 6 INS00036273    A and C Healthcare Services, Inc              130
##   ttl_census_beds ttl_staffed_beds bed_cluster_id
## 1             122              126              2
## 2               4               10              1
## 3              75               83              1
## 4              46               49              1
## 5              22               53              1
## 6             122              130              2
```

```
str(bizines_tbl)
```

```
## 'data.frame':    22202 obs. of  6 variables:
##  $ ims_org_id      : chr  "INS00077200" "INS00000594" "INS00039181" "INS00011388" ...
##  $ business_name   : chr  "140 Prescott Street Corporation" "366th Medical Group" "7 Hills Pediatric
##  $ ttl_license_beds: int  126 10 83 49 53 130 50 138 150 589 ...
##  $ ttl_census_beds : int  122 4 75 46 22 122 45 130 142 576 ...
##  $ ttl_staffed_beds: int  126 10 83 49 53 130 50 138 150 589 ...
##  $ bed_cluster_id  : int  2 1 1 1 1 2 1 2 2 6 ...
```

```
biz_dims <- dim(bizines_tbl)
biz_cols <- colnames(bizines_tbl)
any(is.na(bizines_tbl))
```

```
## [1] FALSE
```

In the chunk of code above, the three data files in csv format have been loaded on to the group 1 practicum 2 sql db and read into tables via RSqlite commands.

The general structures have been looked at with str() and all the data files have no missing values.

A few other details of the tables are specified below:

1. Bed Type Data: The dimensions are 20, 3, with variable names; bed_id, bed_code, bed_desc.
2. Business Data: The dimensions are 22202, 6, with the variable names; ims_org_id, business_name, ttl_license_beds, ttl_census_beds, ttl_staffed_beds, bed_cluster_id.

It is worth making a comment on the Primary Key for the tables:

Primary Key in a given table is one that is unique to each observation, it should not change over time and should not be a null value.

1. Bed Type Data: The Primary key is bed_id.
2. Business Data: The Primary Key is ims_org_id.

2. Identify the Facts variables from the single Fact Table

We want to look more closely at the Bed Fact Table to get to know the dimensional and factual variables therein through the code below.

```
#  Get bed fact data
bed_fact_dt <- read.csv("bed_fact-1.csv", header = TRUE)
dbWriteTable(conekshun, "bed_fact", bed_fact_dt, overwrite = TRUE)
# Dimensions in bed fact
bed_fact_tbl <- dbGetQuery(conekshun, "SELECT * FROM bed_fact")
head(bed_fact_tbl)
```

```
##    ims_org_id bed_id license_beds census_beds staffed_beds
## 1 INS00000519      2           10           7           10
## 2 INS00000519      5          566         394          566
## 3 INS00000519     15           25          17           25
## 4 INS00000519      4           38          26           38
## 5 INS00000519      6           32          22           32
## 6 INS00000519     18          671         466          671
```

```
str(bed_fact_tbl)
```

```
## 'data.frame':    51458 obs. of  5 variables:
##  $ ims_org_id  : chr  "INS00000519" "INS00000519" "INS00000519" "INS00000519" ...
##  $ bed_id      : int  2 5 15 4 6 18 18 5 5 18 ...
##  $ license_beds: int  10 566 25 38 32 671 25 25 25 25 ...
##  $ census_beds : int  7 394 17 26 22 466 14 14 16 16 ...
##  $ staffed_beds: int  10 566 25 38 32 671 25 25 25 25 ...
```

```r
bed_fact_dims <- dim(bed_fact_tbl)
bed_fact_cols <- colnames(bed_fact_tbl)
any(is.na(bed_fact_tbl))
```

```
## [1] FALSE
```

```r
factual_vars <- bed_fact_cols[3:5]
dimensional_vars <- bed_fact_cols[1:2]
```

Bed Facts Data: 1. The dimensions are 51458, 5, with the variable names; ims_org_id, bed_id, license_beds, census_beds, staffed_beds.

2. The Primary Key is ims_org_id.
3. The Variables are:

i) License Beds- The number of beds approved by state regulatory authorities at the institution

ii) Census Beds- The number of beds currently occupied by patients

iii) Staffed Beds- The number of beds staffed by a hospital nurse

The Variables that can be called the fact license_beds, census_beds, staffed_beds.

The Variables that are more of dimensions are the ims_org_id, bed_id.

This table carries more aggregate data than the business table which is could be normalized on the less factual data therein. It would be easier to compute the data by merging these two tables on the fist column which carries the ims_org_id variables.

3. Analysis and Interpretation for Leadership

3.1. Analysis for Leadership

We will extract from the data provided, the ICU beds with bed_id of 4 and SICU bed with bed_id of 15. these will be discussed based on the categories they fall under including the License beds, Census beds and the Staffed beds. Summary reports will be provided for each category mentioned.

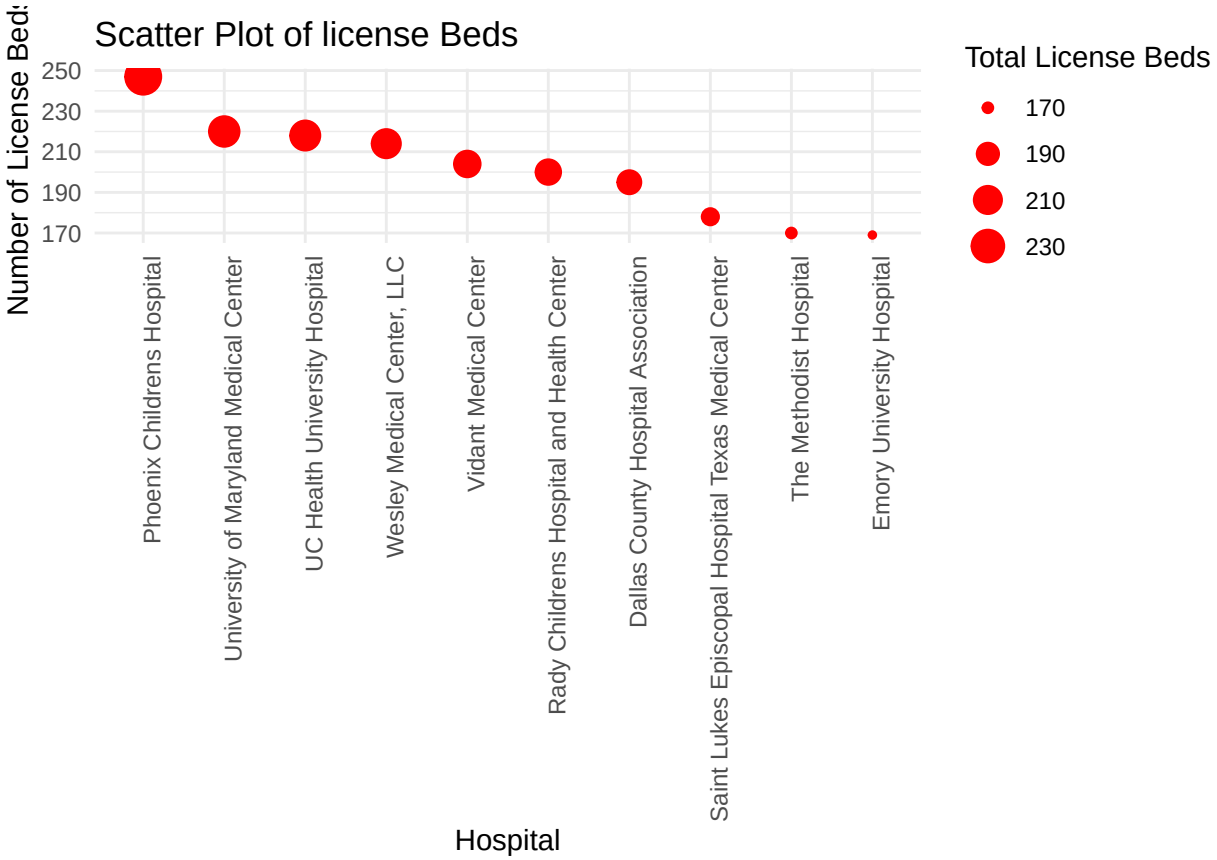
3.1.1. License beds: For the licensed beds category, we will list the top 10 hospitals in descending order through a table with the business name and the total number of licensed beds.

```r
# sql query to get the data
top10license_beds_dt <- dbGetQuery(conekshun,
"SELECT b.business_name, SUM(f.license_beds) AS total_license_beds
FROM bed_fact AS f
INNER JOIN business AS b ON f.ims_org_id = b.ims_org_id
INNER JOIN bed_type AS t ON f.bed_id = t.bed_id
WHERE t.bed_id IN (4, 15)
GROUP BY b.business_name
ORDER BY total_license_beds DESC
```

```
LIMIT 10
")

# View data in table
knitr::kable(top10license_beds_dt)
```

business_name	total_license_beds
Phoenix Childrens Hospital	247
University of Maryland Medical Center	220
UC Health University Hospital	218
Wesley Medical Center, LLC	214
Vidant Medical Center	204
Rady Childrens Hospital and Health Center	200
Dallas County Hospital Association	195
Saint Lukes Episcopal Hospital Texas Medical Center	178
The Methodist Hospital	170
Emory University Hospital	169

```
# Visualize with a scatterplot
library(ggplot2)

# Create a scatter plot
ggplot(top10license_beds_dt, aes(x = reorder(business_name, -total_license_beds), y = total_license_beds
  geom_point(color = "red") +
  labs(x = "Hospital", y = "Number of License Beds", size = "Total License Beds") +
  ggtitle("Scatter Plot of license Beds") +
  theme_minimal() +
  theme(axis.text.x = element_text(angle = 90, hjust = 1))
```



The Table above shows the top 10 Hospitals with the total number of beds. We arrived at this table by Joining the three tables based on the primary key in each table thus getting the facts and dimensions together through SQL queries that filtered the bed_id, 4 and 15 for ICU and SICU respectively, thereafter. For the License bed Category Phoenix Childrens Hospital had the highest total of 247 license beds, followed by University of Maryland Medical Center with 220 license beds.This data has been visualized with an abbreviated word cloud and a scatter plot for clarity.

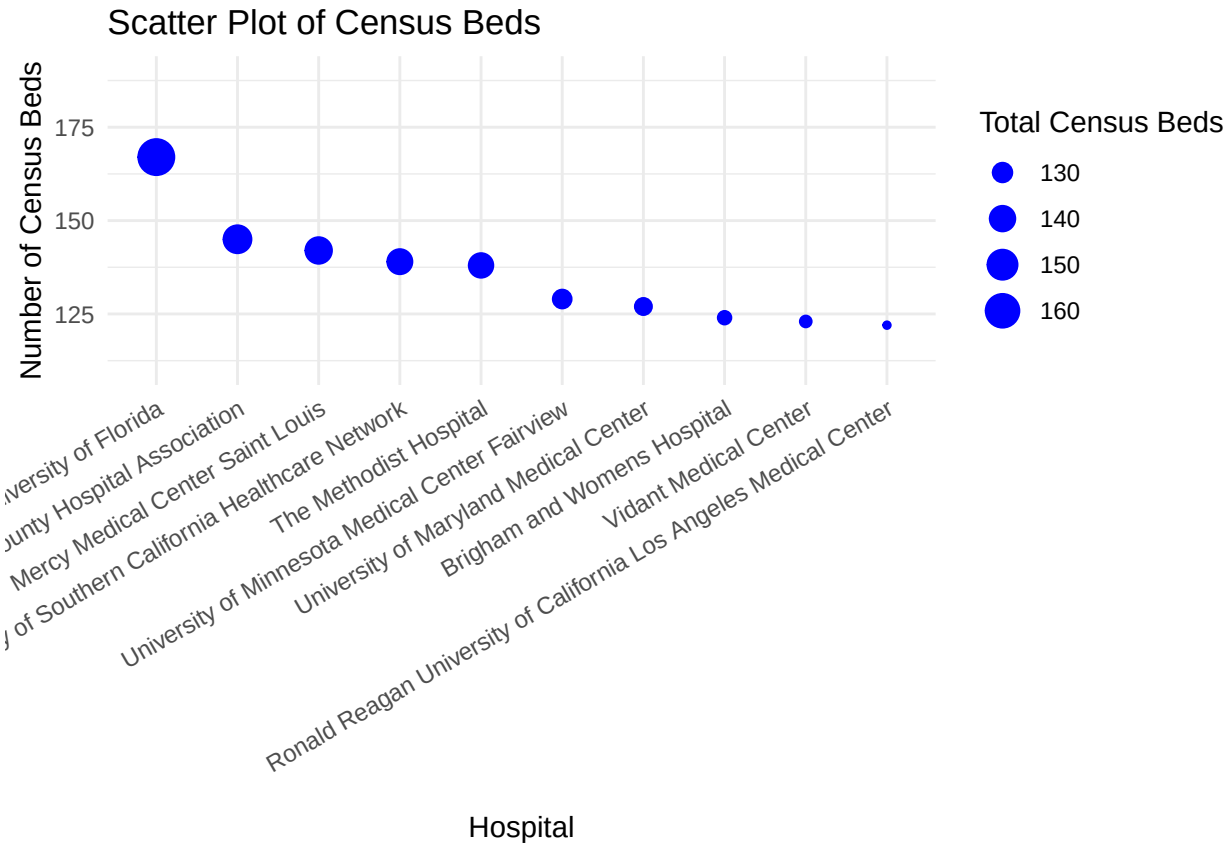
Let us move forward to carry out similar SQL Queries for the Census beds and the Staffed beds categories.

3.1.2. Census beds: List of Top 10 Hospitals ordered by total icu or sicu census beds. Include just two variables, hospital_name (business_name) and the total census beds from above as one summary fact. Include only 10 rows again.

```
# sql query to get the data
top10census_bed_dt <- dbGetQuery(conekshun,
"SELECT b.business_name, SUM(f.census_beds) AS total_census_beds
FROM bed_fact AS f
INNER JOIN business AS b ON f.ims_org_id = b.ims_org_id
INNER JOIN bed_type AS t ON f.bed_id = t.bed_id
WHERE t.bed_id IN (4, 15)
GROUP BY b.business_name
ORDER BY total_census_beds DESC
LIMIT 10")

# View data in table
knitr::kable(top10census_bed_dt)
```

business_name	total_census_beds
Shands Hospital at the University of Florida	167
Dallas County Hospital Association	145
Mercy Medical Center Saint Louis	142
Los Angeles County University of Southern California Healthcare Network	139
The Methodist Hospital	138
University of Minnesota Medical Center Fairview	129
University of Maryland Medical Center	127
Brigham and Womens Hospital	124
Vidant Medical Center	123
Ronald Reagan University of California Los Angeles Medical Center	122

```
#Create Scatter plot
ggplot(top10census_bed_dt, aes(x = reorder(business_name, -total_census_beds), y = total_census_beds, si
  geom_point(color = "blue") +
  labs(x = "Hospital", y = "Number of Census Beds", size = "Total Census Beds") +
  ggtitle("Scatter Plot of Census Beds") +
  theme_minimal() +
  theme(axis.text.x = element_text(angle = 30, hjust = 1)) +
  ylim(110, 190)
```



The table above shows the top 10 hospitals with the total Census Beds in descending order.
This chart was obtained by running similar SQL queries as the license bed categories. The Shands Hospital at the University of Florida topped this category with 167 Census Beds and Dallas County Hospital Association

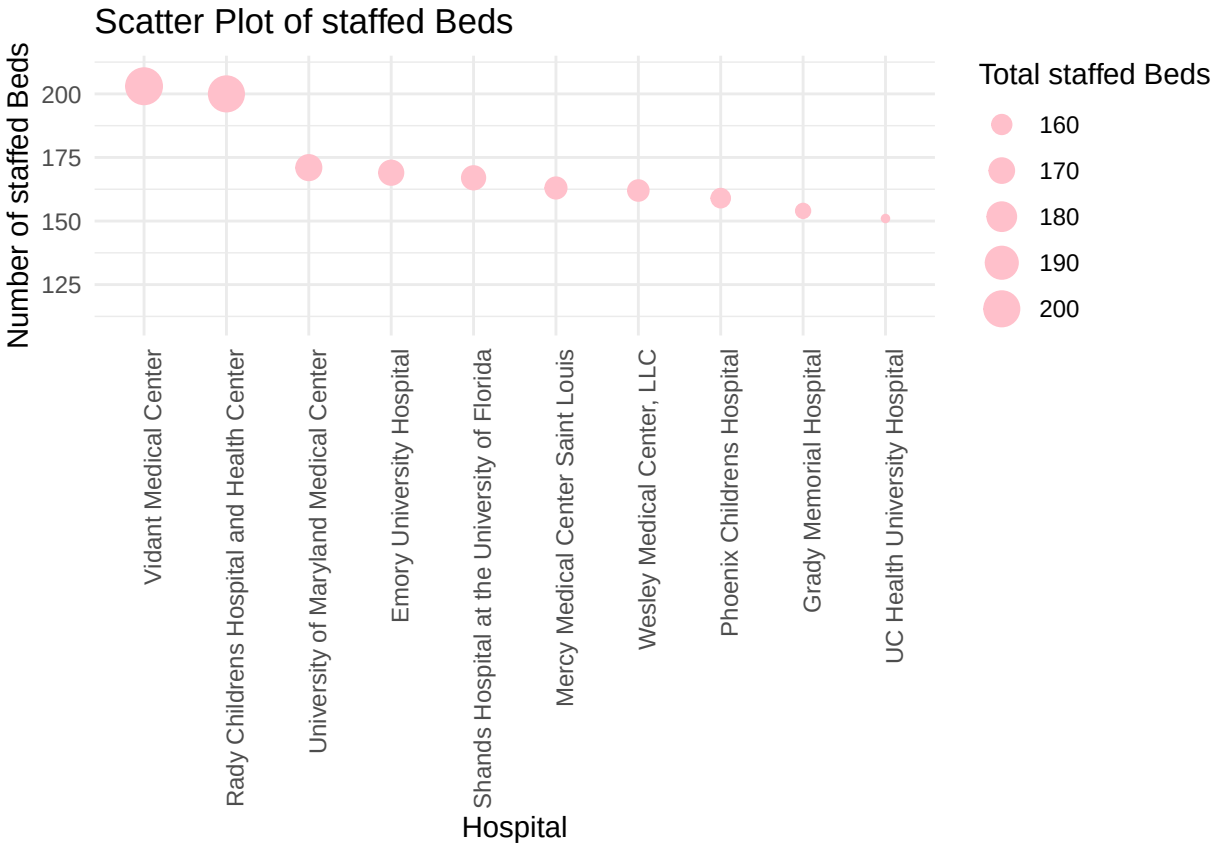
took second place with 145. Please look at the visualization with word cloud and scatter plot for more insights.

3.1.3. Staffed beds: Here we will examine the Staffed beds category to get the top 10 Hospitals by running SQL queries in similitude with the other categories.

```
# sql query to get the data
top10stafd_bed_dt <- dbGetQuery(conekshun,
"SELECT b.business_name, SUM(f.staffed_beds) AS total_staffed_beds
FROM bed_fact AS f
INNER JOIN business AS b ON f.ims_org_id = b.ims_org_id
INNER JOIN bed_type AS t ON f.bed_id = t.bed_id
WHERE t.bed_id IN (4, 15)
GROUP BY b.business_name
ORDER BY total_staffed_beds DESC
LIMIT 10")

# View data in table
knitr::kable(top10stafd_bed_dt)
```

business_name	total_staffed_beds
Vidant Medical Center	203
Rady Childrens Hospital and Health Center	200
University of Maryland Medical Center	171
Emory University Hospital	169
Shands Hospital at the University of Florida	167
Mercy Medical Center Saint Louis	163
Wesley Medical Center, LLC	162
Phoenix Childrens Hospital	159
Grady Memorial Hospital	154
UC Health University Hospital	151

```
# Visualize with a scatterplot
ggplot(top10stafd_bed_dt, aes(x = reorder(business_name, -total_staffed_beds), y = total_staffed_beds, s
  geom_point(color = "pink") +
  labs(x = "Hospital", y = "Number of staffed Beds", size = "Total staffed Beds") +
  ggtitle("Scatter Plot of staffed Beds") +
  theme_minimal() +
  theme(axis.text.x = element_text(angle = 90, hjust = 1)) +
  ylim(110, 210)
```



The Table above shows the top 10 hospitals with the highest number of ICU and SICU staffed beds in descending order.
We arrived at this list of hospitals and corresponding total beds and found Vidant Medical Center topped the chart with 203 Staffed beds while Rady Childrens Hospital and Health Center came in second place with 200 Staffed beds. A better understanding of this table can be derived form the word cloud and scatter plot of the data.

3.2. Interpretation of Findings:

From the top-performers charts generated above, we want to develop deeper insight into the business data to bring to leadership's attention.

Summary- Top 2 Candidates for each bed type:
1. Licensed beds-
Phoenix Childrens Hospital University of Maryland Medical Center
2. Census Beds-
Shands Hospital at the University of Florida
Dallas County Hospital Association

3. Staffed beds-
   Vidant Medical Center
   Rady Childrens Hospital and Health Center

Based on the results from step 3a, it is clear that there are two hospitals which appear on all three "top-performer" lists for the highest number of each bed type (licensed beds, census beds, and staffed beds). The intersection of the lists that could be considered for the pilot program are the University of Maryland

Medical Center, and Vidant Medical Center. All other hospitals listed were either on only one or two of the "top-performer" lists.

4. Further Analysis and Recommendations

4.1. Drill down investigation

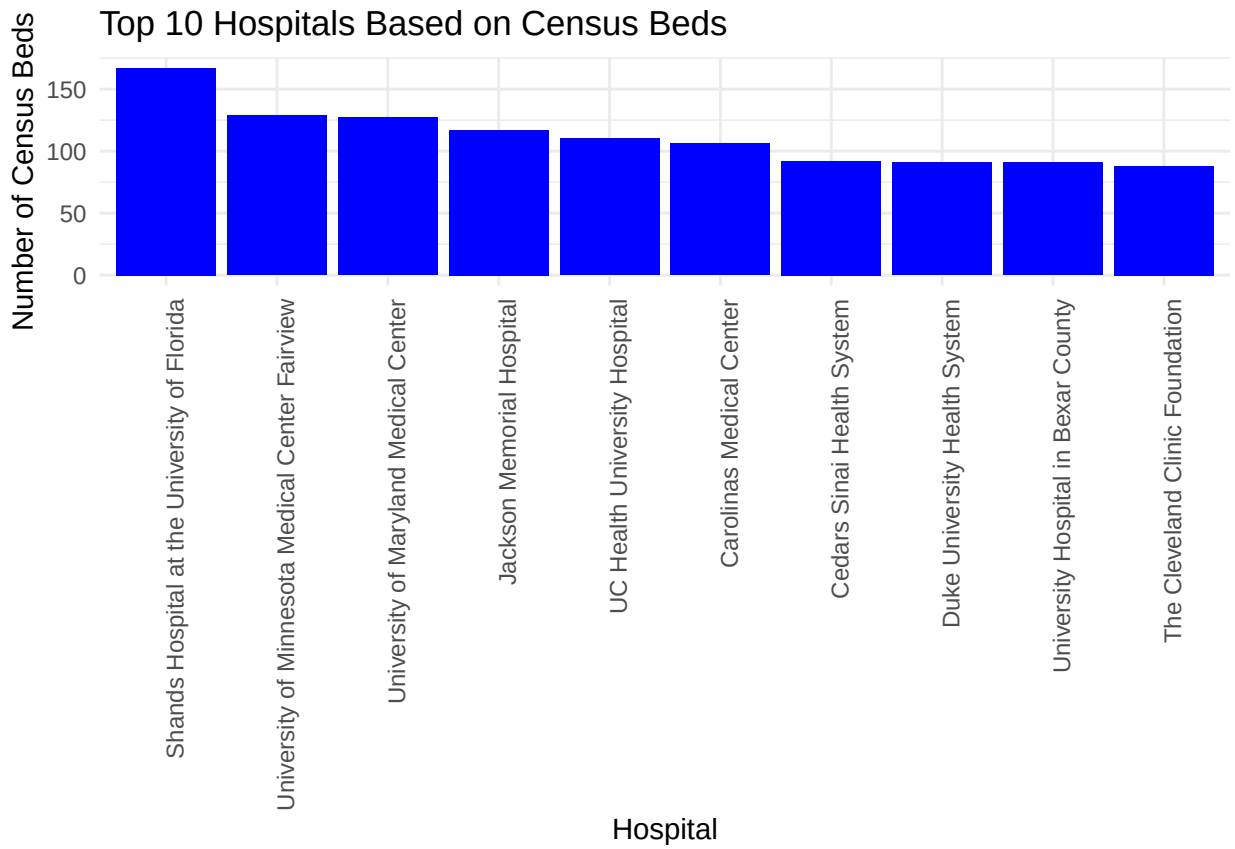
Based on the analyses in the previous sections, we would like to submit a final investigation to further drill down the top choice for the pilot program.The following analyses will investigate and determine the best recommendation to leadership.

First we will start by investigating the data for census beds which contain both SICU and ICU bed types.

```
# SQL Query to get the Census beds data

censusbeds_minreq <- dbGetQuery(conekshun,
"SELECT b.business_name, SUM(f.census_beds) AS total_census_beds
FROM bed_fact AS f
INNER JOIN business AS b ON f.ims_org_id = b.ims_org_id
INNER JOIN bed_type AS t ON f.bed_id = t.bed_id
WHERE t.bed_id IN (4, 15)
GROUP BY b.business_name
HAVING COUNT(DISTINCT t.bed_id) = 2
ORDER BY total_census_beds DESC
LIMIT 10")

# Visualize the data
knitr::kable(censusbeds_minreq)
```

business_name	total_census_beds
Shands Hospital at the University of Florida	167
University of Minnesota Medical Center Fairview	129
University of Maryland Medical Center	127
Jackson Memorial Hospital	117
UC Health University Hospital	110
Carolinas Medical Center	106
Cedars Sinai Health System	92
University Hospital in Bexar County	91
Duke University Health System	91
The Cleveland Clinic Foundation	88

```
library(ggplot2)

ggplot(censusbeds_minreq, aes(x = reorder(business_name, -total_census_beds), y = total_census_beds)) +
  geom_bar(stat = "identity", fill = "blue") +
  labs(x = "Hospital", y = "Number of Census Beds") +
  ggtitle("Top 10 Hospitals Based on Census Beds") +
  theme_minimal() +
  theme(axis.text.x = element_text(angle = 90, hjust = 1))
```



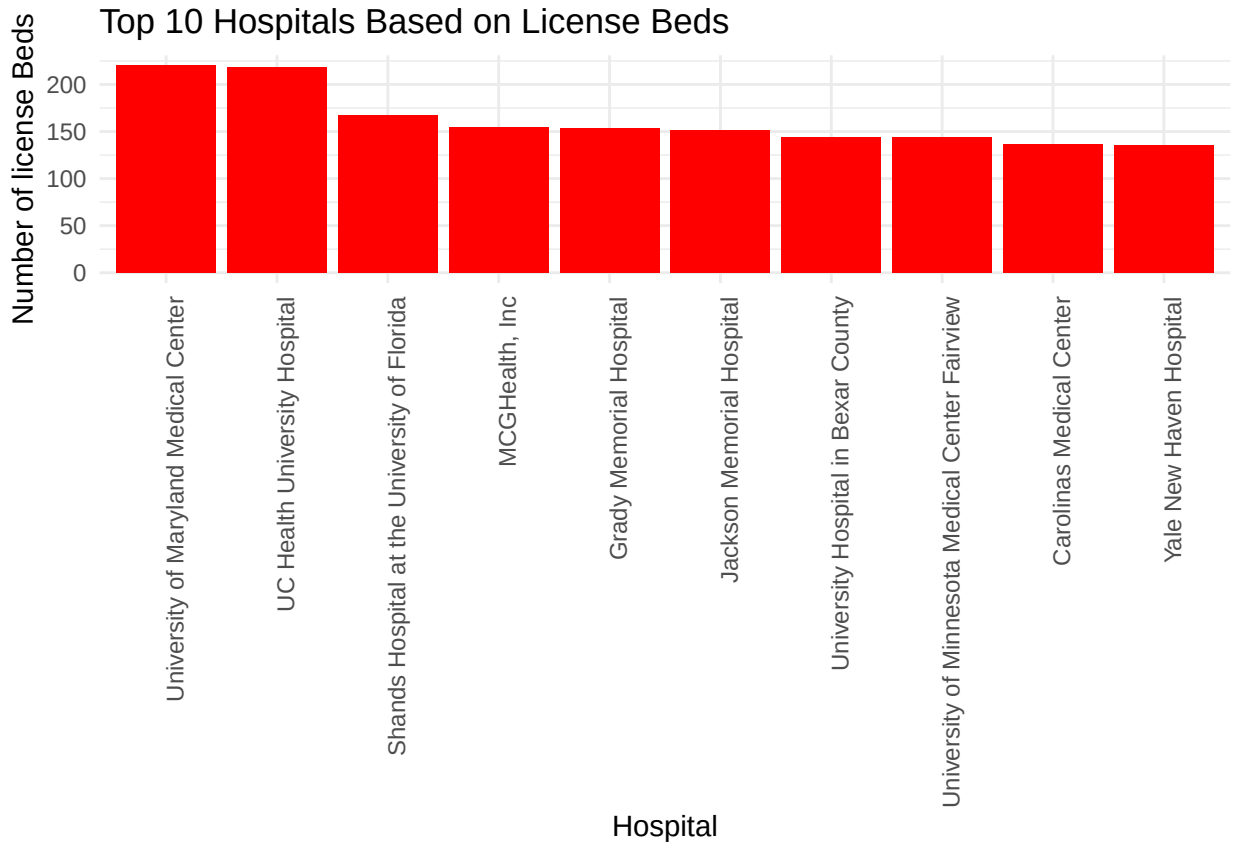
Next we will conduct the same investigation, this type observing the licensed bed types in the data.

```
# sql query to get the data
licensebeds_minreq <- dbGetQuery(conekshun,
"SELECT b.business_name, SUM(f.license_beds) AS total_license_beds
FROM bed_fact AS f
INNER JOIN business AS b ON f.ims_org_id = b.ims_org_id
INNER JOIN bed_type AS t ON f.bed_id = t.bed_id
WHERE t.bed_id IN (4, 15)
GROUP BY b.business_name
HAVING COUNT(DISTINCT t.bed_id) = 2
ORDER BY total_license_beds DESC
LIMIT 10
")

# Visualize the data
knitr::kable(licensebeds_minreq)
```

business_name	total_license_beds
University of Maryland Medical Center	220
UC Health University Hospital	218
Shands Hospital at the University of Florida	167
MCGHealth, Inc	155
Grady Memorial Hospital	154
Jackson Memorial Hospital	151

business_name	total_license_beds
University of Minnesota Medical Center Fairview	144
University Hospital in Bexar County	144
Carolinas Medical Center	137
Yale New Haven Hospital	136

```
ggplot(licensebeds_minreq, aes(x = reorder(business_name, -total_license_beds), y = total_license_beds)
  geom_bar(stat = "identity", fill = "red") +
  labs(x = "Hospital", y = "Number of license Beds") +
  ggtitle("Top 10 Hospitals Based on License Beds") +
  theme_minimal() +
  theme(axis.text.x = element_text(angle = 90, hjust = 1))
```



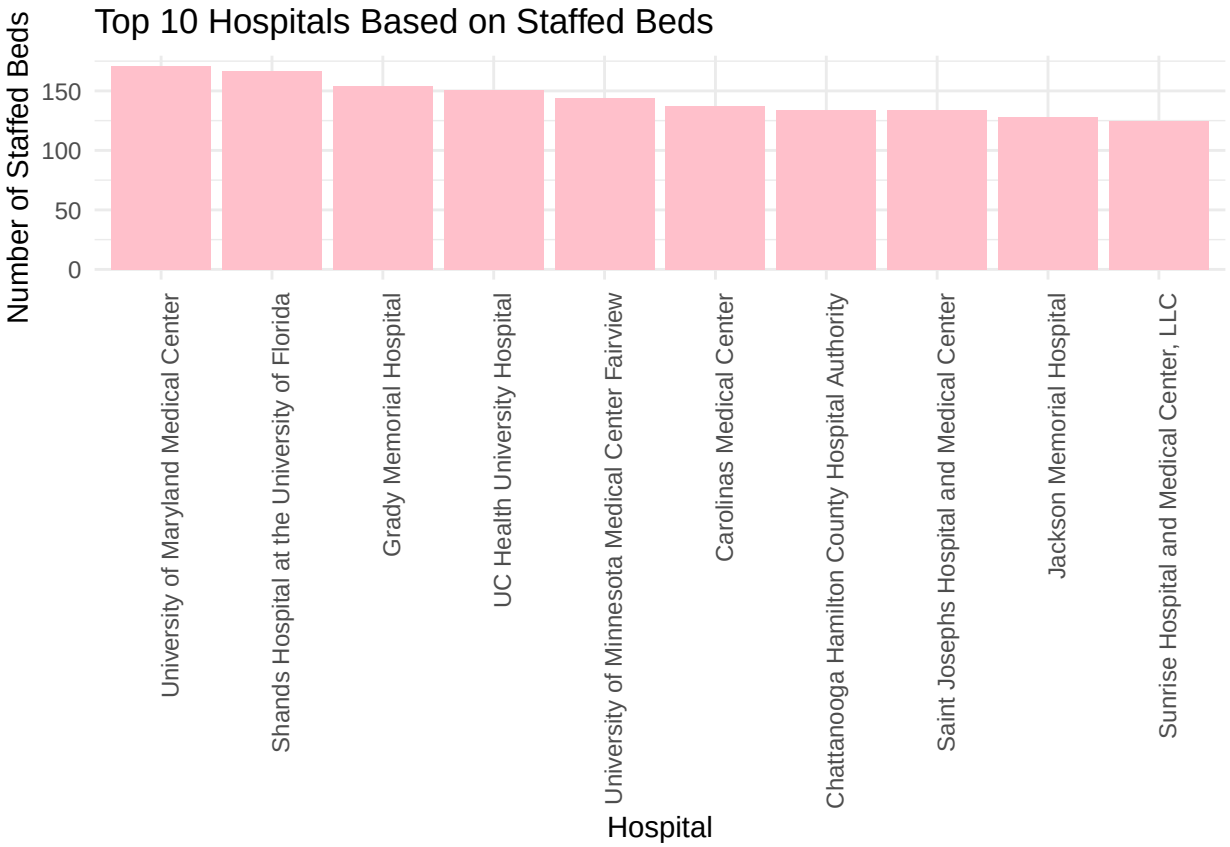
The next analysis follows the same as the previous two, except investigating staffed beds which have both SICU and ICU bed types.

```
# sql query to get the data
staffedbed_minreq <- dbGetQuery(conekshun,
"SELECT b.business_name, SUM(f.staffed_beds) AS total_staffed_beds
FROM bed_fact AS f
INNER JOIN business AS b ON f.ims_org_id = b.ims_org_id
INNER JOIN bed_type AS t ON f.bed_id = t.bed_id
WHERE t.bed_id IN (4, 15)
GROUP BY b.business_name
```

```
HAVING COUNT(DISTINCT t.bed_id) = 2
ORDER BY total_staffed_beds DESC
LIMIT 10")

# Visualize the data
knitr::kable(staffedbed_minreq)
```

business_name	total_staffed_beds
University of Maryland Medical Center	171
Shands Hospital at the University of Florida	167
Grady Memorial Hospital	154
UC Health University Hospital	151
University of Minnesota Medical Center Fairview	144
Carolinas Medical Center	137
Saint Josephs Hospital and Medical Center	134
Chattanooga Hamilton County Hospital Authority	134
Jackson Memorial Hospital	128
Sunrise Hospital and Medical Center, LLC	125

```
ggplot(staffedbed_minreq, aes(x = reorder(business_name, -total_staffed_beds), y = total_staffed_beds))
  geom_bar(stat = "identity", fill = "pink") +
  labs(x = "Hospital", y = "Number of Staffed Beds") +
  ggtitle("Top 10 Hospitals Based on Staffed Beds") +
  theme_minimal() +
  theme(axis.text.x = element_text(angle = 90, hjust = 1))
```



Above you will see the outputs for total census beds, total licensed beds, and the total staffed beds in which there are at least one bed present in the ICU and SICU. Based on the data, it appears as though there have been some movements of hospital positions when adding this new constraint to the query.

4.2. Final recommendation

In addition to the tables in the previous section, we have decided to take the analysis one step further and consider another variable- bed availability. While top performing hospitals with high staffed bed numbers or high census numbers can be attractive as a pilot choice, we must ensure that there is still adequate room to grow or expand to fully assess the program. Below you will see an assessment of measured bed availability in relation to census numbers (current beds in use).

```
top_selection <- dbGetQuery(conekshun,
"SELECT b.business_name,
  SUM(f.census_beds) AS total_census_beds,
  SUM(f.license_beds) - SUM(f.staffed_beds) AS Bed_Availability
FROM bed_fact AS f
INNER JOIN business AS b ON f.ims_org_id = b.ims_org_id
INNER JOIN bed_type AS t ON f.bed_id = t.bed_id
WHERE t.bed_id IN (4, 15)
GROUP BY b.business_name
HAVING COUNT(DISTINCT t.bed_id) = 2
ORDER BY Bed_Availability DESC
LIMIT 10")

knitr::kable(top_selection)
```

business_name	total_census_beds	Bed_Availability
UC Health University Hospital	110	67
Yale New Haven Hospital	64	61
MCGHealth, Inc	88	59
Hennepin Health Care System, Inc	47	56
University of Maryland Medical Center	127	49
University of Texas Medical Branch at Galveston	49	40
Allegheny General Hospital	51	40
Georgetown University Hospital	37	36
Via Christi Health	31	35
Northwest Texas Healthcare System, Inc	45	34

```
library(wordcloud)

# Visualize with word cloud
abbreviated_names4 <- sapply(strsplit(top_selection$business_name, " "), function(words) paste0(substr(
wordcloud(words = abbreviated_names4,
          freq = top_selection$Bed_Availability,
          min.freq = 20,
          random.order = FALSE,
          colors = "green",
          family="Times")
```

Our final recommendation for leadership is to consider UC Health University Hospital because it has high usage (total census beds), yet still has the room to grow. We determined room to grow (bed_availability)

by subtracting licensed beds from staffed beds. UC Health University Hospital is capable of adding 67 beds to its roster, while still currently being a very active hospital when looking at the total census beds. With this information, that is our final recommendation to hospital leadership.