

## UCSF Stages Data - Prelim Analysis

Steph Reynolds (Stephanie.Reynolds@ucsf.edu)

Dec 29, 2021 10:24 AM

# Background

## Project

MINDSCAPE: Modeling of infectious network dynamics for surveillance, control and prevention enhancement

## Description

This file summarizes the data, creates descriptive tables for the continuous and categorical variables, and generates plots to show the distribution of length of stay (LOS) by admission stage.

## Source Data

- ▶ Demographics and Daily Covid Stage Data  
(dm\_stg\_unique\_pt\_11.08.2021.csv)
  - ▶ This file contains data on patient demographics and COVID stage (based on WHO Clinical Progression Scale, which aims to capture patient clinical trajectory and resource usage over the course of the clinical illness – in this case, COVID-19).
  - ▶ Each row contains a different patient as shown by their unique encounter ID.

Load required packages

## Import and preview data

```
## Rows: 1117 Columns: 17

## -- Column specification -----
## Delimiter: ","
## chr  (7): ID, sex, zip, race, ethnicity, smoking, death
## dbl  (7): age, BMI, LOS, stage, max_stage, stage_adm, st
## date (3): date_adm, date_disc, DOD

##
## i Use 'spec()' to retrieve the full column specification
## i Specify the column types or set 'show_col_types = FALS
```

ID	age	sex	zip
0055f0c4-990c-44ce-911e-4b1784666eeb	53	Male	95382
00a80297-7016-4b06-8dfc-21a1490baafa	57	Male	94939
01abe61d-6df2-4f18-9a85-8b8d12f94f08	51	Male	93722
01b4c414-b32c-46dc-8c2a-29f399147f70	67	Male	94901

## Create table one for categorical and continuous variables

```
# Define categorical and continuous variables
#cat_vars <- c("sex", "race", "ethnicity", "smoking", "death")
cat_vars <- c("sex", "race", "ethnicity", "smoking", "death")
cont_vars <- c("age", "BMI", "LOS")

t1 <- CreateCatTable(data = df, cat_vars)
#print(t1, varLabels = T, showAllLevels = T, digits = 1)
kableone(t1)
```

	Overall
n	1117
sex = Male (%)	576 (51.6)
race (%)	
Asian	221 (19.8)
Black or African American	110 ( 9.8)
Native Hawaiian or Other Pacific Islander	18 ( 1.6)
Other	341 (30.5)
Unknown	33 ( 3.0)

## Stage transition matrices

*# Create categorical table of stage\_adm vs. stage\_disc*

```
CreateCatTable(strata = 'stage_adm', vars = 'stage_disc', c
```

	4	5	6	7	8
n	517	397	79	11	50
stage_disc (%)					
4	435 (84.1)	200 (50.4)	29 (36.7)	4 (36.4)	20 (40.0)
5	55 (10.6)	162 (40.8)	32 (40.5)	0 ( 0.0)	14 (28.0)
6	3 ( 0.6)	5 ( 1.3)	2 ( 2.5)	0 ( 0.0)	0 ( 0.0)
7	3 ( 0.6)	1 ( 0.3)	2 ( 2.5)	6 (54.5)	5 (10.0)
8	0 ( 0.0)	0 ( 0.0)	0 ( 0.0)	1 ( 9.1)	1 ( 2.0)
9	0 ( 0.0)	0 ( 0.0)	1 ( 1.3)	0 ( 0.0)	0 ( 0.0)
10	21 ( 4.1)	29 ( 7.3)	13 (16.5)	0 ( 0.0)	10 (20.0)

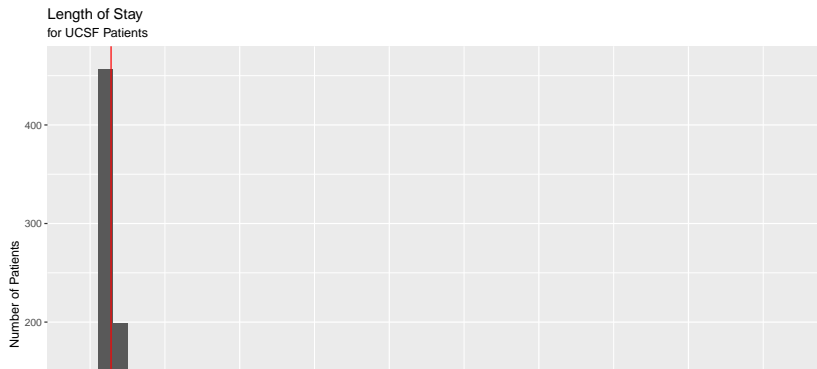
*# Create categorical table of stage\_adm vs. max\_stage*

```
CreateCatTable(strata = 'stage_adm', vars = 'max_stage', da
```

	4	5	6	7	8
n	517	397	79	11	50

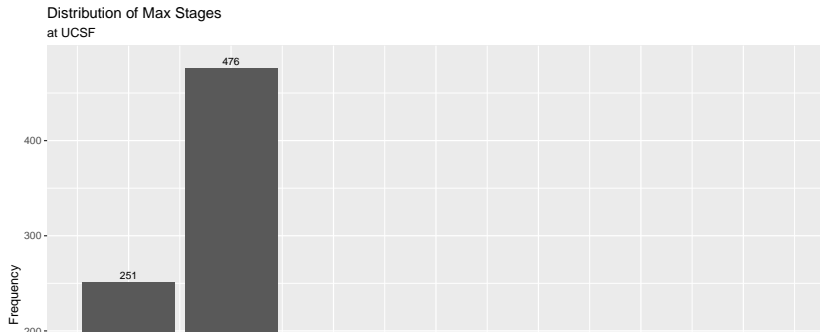
# Create histogram of LOS

```
ggplot(df, aes(x = LOS)) +  
  geom_histogram(binwidth = 5) +  
  labs(title = 'Length of Stay',  
        subtitle = 'for UCSF Patients',  
        x = 'Number of Days',  
        y = 'Number of Patients') +  
  geom_vline(xintercept = median(df$LOS), color = 'red')
```



## Create barplot showing distribution of max stages

```
ggplot(df, aes(x = max_stage)) +  
  geom_bar() +  
  labs(title = 'Distribution of Max Stages',  
        subtitle = 'at UCSF',  
        x = 'WHO Clinical Progression Stage',  
        y = 'Frequency') +  
  scale_x_continuous(breaks = 4:10) +  
  geom_text(stat = 'count', aes(label = after_stat(count)))
```

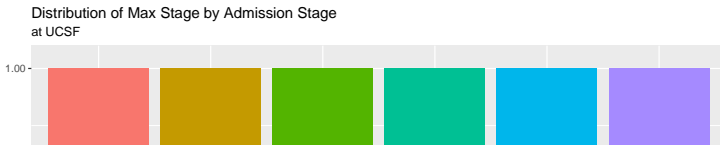




Create proportional stacked bar chart to show proportion of max\_stage within stage\_adm

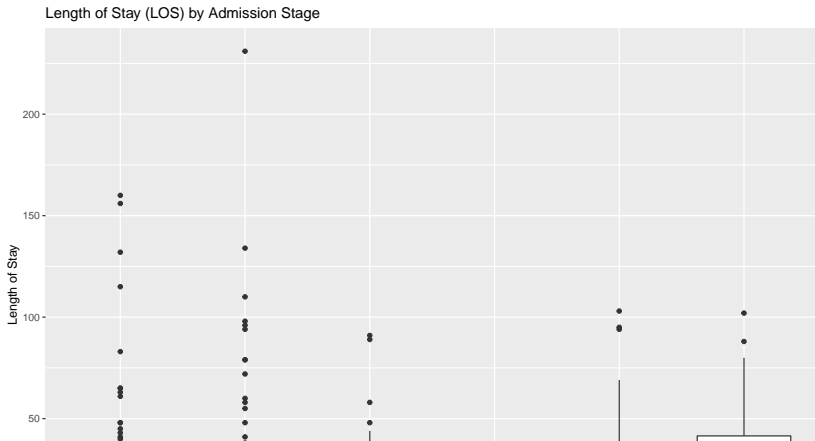
```
df %>%  
  group_by(stage_adm, max_stage) %>%  
  summarize(count = n()) %>%  
  ggplot(aes(x = as.factor(stage_adm), y = count, fill = as.factor(max_stage))) +  
  geom_col(position = 'fill') +  
  labs(title = 'Distribution of Max Stage by Admission Stage',  
       subtitle = 'at UCSF',  
       x = 'Admission Stage',  
       y = 'Proportion',  
       fill = 'Max Stage') #+
```

## 'summarise()' has grouped output by 'stage\_adm'. You can



## Create boxplot of LOS grouped by admission stage

```
ggplot(df, aes(x = as.factor(stage_adm), y = LOS)) +  
  geom_boxplot() +  
  labs(title = 'Length of Stay (LOS) by Admission Stage',  
        x = 'Admission Stage',  
        y = 'Length of Stay')
```



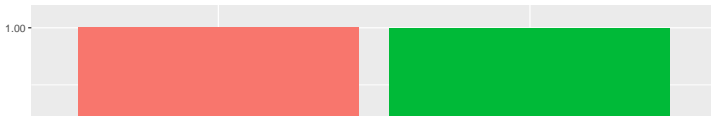
Re-run the proportional stacked bar chart and boxplot from above using stage

categories (moderate, severe, dead)

```
# Proportional stacked bar chart showing which proportion of  
df %>%  
  group_by(stgcat_adm, stgcat_max) %>%  
  summarize(count = n()) %>%  
  ggplot(aes(x = stgcat_adm, y = count, fill = stgcat_max))  
  geom_col(position = 'fill') +  
  labs(title = 'Distribution of Max Stage by Admission Stage',  
       subtitle = 'at UCSF',  
       x = 'Admission Stage',  
       y = 'Proportion',  
       fill = 'Max Stage')
```

## 'summarise()' has grouped output by 'stgcat\_adm'. You can

Distribution of Max Stage by Admission Stage  
at UCSF



End of Document