UCSF Stages Data - Prelim Analysis

Steph Reynolds (Stephanie.Reynolds@ucsf.edu)

Dec 27, 2021 13:28 PM

Background

Project

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

Description

This file imports demographic and staging data and returns a dataset indicating each patient's `max_stage`, `date_adm`, `date_disc`, `los`, `stage_adm`, `stage_disc`, `days_to_disc`. This file prepares the staging data for analysis.

Source Data

- Demographics and Daily Covid Stage Data (`dm_covid_stg_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 represents one day per patient in hospital...

Load required packages

```
library(here)
```

```
## here() starts at /Users/sreynolds2/Documents/GitHub/MS-Covid_Staging
```

```
library(tidyverse)
```

```
## -- Attaching packages --------------------------------------- tidyverse 1.3.1 --

## v ggplot2 3.3.5     v purrr   0.3.4
## v tibble  3.1.4     v dplyr   1.0.7
## v tidyr   1.1.3     v stringr 1.4.0
## v readr   2.0.1     v forcats 0.5.1
```

```
## -- Conflicts ---------------------------------------- tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
library(psych)
```

```
##
## Attaching package: 'psych'
```

```
## The following objects are masked from 'package:ggplot2':
##
##     %+%, alpha
```

```
library(tableone)
library(scales)
```

```
##
## Attaching package: 'scales'
```

```
## The following objects are masked from 'package:psych':
##
##     alpha, rescale
```

```
## The following object is masked from 'package:purrr':
##
##     discard
```

```
## The following object is masked from 'package:readr':
##
##     col_factor
```

```
library(DescTools) # Winsorized mean -- Winsorize(mean(df$var))
```

```
##
## Attaching package: 'DescTools'
```

```
## The following objects are masked from 'package:psych':
##
##     AUC, ICC, SD
```

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, stage_disc
## date (3): date_adm, date_disc, DOD
```

```
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```
## Rows: 1,117
## Columns: 17
## $ ID         <chr> "0055f0c4-990c-44ce-911e-4b1784666eeb", "00a80297-7016-4b06~
## $ age        <dbl> 53, 57, 51, 67, 66, 83, 55, 39, 52, 49, 39, 46, 64, 54, 72,~
## $ sex        <chr> "Male", "Male", "Male", "Male", "Female", "Female", "Female~
## $ zip        <chr> "95382", "94939", "93722", "94901", "94553-5927", "94110", ~
## $ race       <chr> "Other", "White or Caucasian", "White or Caucasian", "White~
## $ ethnicity  <chr> "Hispanic or Latino", "Not Hispanic or Latino", "Hispanic o~
## $ smoking    <chr> "Not Current Smoker", "Current Smoker", "Not Current Smoker~
## $ BMI        <dbl> 31.58, 42.03, 28.59, 27.21, 28.22, 22.37, 34.09, 35.82, 30.~
## $ LOS        <dbl> 9, 8, 9, 8, 2, 12, 12, 2, 4, 6, 5, 19, 5, 88, 3, 4, 7, 7, 9~
## $ stage      <dbl> 5, 5, 4, 4, 5, 4, 4, 4, 4, 4, 4, 4, 4, 9, 4, 4, 5, 5, 5, 5,~
## $ max_stage  <dbl> 6, 6, 5, 4, 5, 4, 6, 4, 5, 5, 5, 8, 5, 10, 5, 4, 10, 5, 5, ~
## $ date_adm   <date> 2020-10-19, 2021-01-04, 2020-12-04, 2020-08-10, 2020-12-07~
## $ date_disc  <date> 2020-10-28, 2021-01-12, 2020-12-13, 2020-08-18, 2020-12-09~
## $ stage_adm  <dbl> 5, 5, 4, 4, 5, 4, 4, 4, 4, 4, 4, 4, 4, 9, 4, 4, 5, 5, 5, 5,~
## $ stage_disc <dbl> 5, 5, 4, 4, 5, 4, 5, 4, 4, 4, 4, 5, 4, 10, 5, 4, 10, 4, 5, ~
## $ DOD        <date> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, 2021-0~
## $ death      <chr> "No", "No", "No", "No", "No", "No", "No", "No", "No", "No",~
```

```
## # A tibble: 20 x 17
##    ID       age sex    zip   race  ethnicity smoking   BMI   LOS stage max_stage
##    <chr>  <dbl> <chr>  <chr> <chr> <chr>     <chr>   <dbl> <dbl> <dbl>     <dbl>
##  1 0055f~    53 Male   95382 Other Hispanic~ Not Cu~  31.6     9     5         6
##  2 00a80~    57 Male   94939 Whit~ Not Hisp~ Curren~  42.0     8     5         6
##  3 01abe~    51 Male   93722 Whit~ Hispanic~ Not Cu~  28.6     9     4         5
##  4 01b4c~    67 Male   94901 Whit~ Hispanic~ Not Cu~  27.2     8     4         4
##  5 01ef2~    66 Female 9455~ Whit~ Not Hisp~ Not Cu~  28.2     2     5         5
##  6 020b8~    83 Female 94110 Other Hispanic~ Not Cu~  22.4    12     4         4
##  7 0219f~    55 Female 94606 Asian Not Hisp~ Not Cu~  34.1    12     4         6
##  8 027bf~    39 Male   94110 Other Not Hisp~ Not Cu~  35.8     2     4         4
##  9 02ca4~    52 Female 96002 Whit~ Not Hisp~ Not Cu~  30.1     4     4         5
## 10 02cc5~    49 Female 95205 Other Hispanic~ Not Cu~  30.8     6     4         5
## 11 02e1c~    39 Female 94901 Whit~ Hispanic~ Not Cu~  22.4     5     4         5
## 12 02e48~    46 Female 9413~ Other Hispanic~ Not Cu~  57.0    19     4         8
## 13 03207~    64 Male   94134 Asian Not Hisp~ Not Cu~  29.5     5     4         5
## 14 0360f~    54 Female 95111 Other Hispanic~ Smokin~  44.6    88     9        10
## 15 04486~    72 Male   94112 Asian Not Hisp~ Not Cu~  23.5     3     4         5
## 16 04943~    81 Male   94116 Asian Not Hisp~ Not Cu~  19.6     4     4         4
## 17 049a4~    91 Male   94116 Asian Not Hisp~ Not Cu~  20.6     7     5        10
## 18 04b2f~    66 Male   94112 Other Hispanic~ Not Cu~  26.3     7     5         5
## 19 05775~    62 Female 9412~ Whit~ Not Hisp~ Not Cu~  23.4     9     5         5
## 20 05776~    66 Male   95762 Whit~ Not Hisp~ Not Cu~  26.0   110     5        10
## # ... with 6 more variables: date_adm <date>, date_disc <date>,
## #   stage_adm <dbl>, stage_disc <dbl>, DOD <date>, death <chr>
```

```
##            vars    n   mean     sd median trimmed    mad   min    max   range
## ID*           1 1117 559.00 322.59  559.0  559.00 413.65  1.00 1117.0 1116.00
## age           2 1117  58.63  19.63   60.0   58.69  22.24 18.00  104.0   86.00
```

```
## sex*          3 1117   1.52   0.50    2.0    1.52   0.00  1.00    2.0    1.00
## zip*          4 1117 252.00 145.97  232.0  246.67 170.50  1.00  542.0  541.00
## race*         5 1117   3.93   1.91    4.0    4.03   2.97  1.00    6.0    5.00
## ethnicity*    6 1117   1.71   0.51    2.0    1.73   0.00  1.00    3.0    2.00
## smoking*      7 1117   2.03   0.36    2.0    2.00   0.00  1.00    3.0    2.00
## BMI           8 1076  28.46   7.86   27.2   27.69   6.83 11.55   79.2   67.65
## LOS           9 1117  12.97  18.38    7.0    9.13   5.93  1.00  231.0  230.00
## stage        10 1117   4.99   1.37    5.0    4.67   1.48  4.00    9.0    5.00
## max_stage    11 1117   5.90   1.96    5.0    5.64   1.48  4.00   10.0    6.00
## date_adm     12 1117    NaN     NA     NA     NaN     NA   Inf   -Inf    -Inf
## date_disc    13 1117    NaN     NA     NA     NaN     NA   Inf   -Inf    -Inf
## stage_adm    14 1117   4.99   1.37    5.0    4.67   1.48  4.00    9.0    5.00
## stage_disc   15 1117   4.85   1.69    4.0    4.37   0.00  4.00   10.0    6.00
## DOD          16   94    NaN     NA     NA     NaN     NA   Inf   -Inf    -Inf
## death*       17 1117   1.08   0.28    1.0    1.00   0.00  1.00    2.0    1.00
##              skew kurtosis   se  Q0.25 Q0.75
## ID*          0.00    -1.20 9.65 280.00 838.0
## age         -0.06    -0.75 0.59  44.00  72.0
## sex*        -0.06    -2.00 0.01   1.00   2.0
## zip*         0.29    -1.02 4.37 134.00 369.0
## race*       -0.36    -1.29 0.06   2.00   6.0
## ethnicity* -0.28    -0.64 0.02   1.00   2.0
## smoking*     0.36     4.48 0.01   2.00   2.0
## BMI          1.44     4.17 0.24  23.04  32.3
## LOS          4.63    32.09 0.55   4.00  14.0
## stage        1.78     2.37 0.04   4.00   5.0
## max_stage    1.01    -0.44 0.06   5.00   7.0
## date_adm       NA       NA   NA     NA    NA
## date_disc      NA       NA   NA     NA    NA
## stage_adm    1.78     2.37 0.04   4.00   5.0
## stage_disc   2.39     4.48 0.05   4.00   5.0
## DOD            NA       NA   NA     NA    NA
## death*       2.99     6.96 0.01   1.00   1.0
```

Create table one for categorical and continuous variables

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

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

```
##
##                  level                                Overall
##   n                                                   1117
##   sex (%)        Female                                541 (48.4)
##                  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)
##                  White or Caucasian                         394 (35.3)
##   ethnicity (%)  Hispanic or Latino                         357 (32.0)
##                  Not Hispanic or Latino                     729 (65.3)
##                  Unknown                                     31 ( 2.8)
##   smoking (%)    Current Smoker                              58 ( 5.2)
##                  Not Current Smoker                         969 (86.8)
##                  Smoking Status Unknown                      90 ( 8.1)
##   death (%)      No                                        1023 (91.6)
##                  Yes                                         94 ( 8.4)
##   max_stage (%)  4                                         251 (22.5)
##                  5                                         476 (42.6)
##                  6                                         107 ( 9.6)
##                  7                                           6 ( 0.5)
##                  8                                          68 ( 6.1)
##                  9                                         115 (10.3)
##                  10                                         94 ( 8.4)
##   stage_adm (%)  4                                         517 (46.3)
##                  5                                         397 (35.5)
##                  6                                          79 ( 7.1)
##                  7                                          11 ( 1.0)
##                  8                                          50 ( 4.5)
##                  9                                          63 ( 5.6)
##   stage_disc (%) 4                                         705 (63.1)
##                  5                                         282 (25.2)
##                  6                                          11 ( 1.0)
##                  7                                          20 ( 1.8)
##                  8                                           3 ( 0.3)
##                  9                                           2 ( 0.2)
##                  10                                         94 ( 8.4)
```

```
# tableone::summary.CreateCatTable
summary(t1)
```

```
## strata: Overall
##         var    n miss p.miss                                      level freq
##         sex 1117    0    0.0                                     Female  541
##                                                                    Male  576
##
##        race 1117    0    0.0                                      Asian  221
##                                                 Black or African American  110
##                                     Native Hawaiian or Other Pacific Islander   18
##                                                                   Other  341
##                                                                 Unknown   33
##                                                        White or Caucasian  394
##
##   ethnicity 1117    0    0.0                            Hispanic or Latino  357
##                                                      Not Hispanic or Latino  729
##                                                                 Unknown   31
##
##     smoking 1117    0    0.0                                Current Smoker   58
##                                                        Not Current Smoker  969
```

```
##                                                Smoking Status Unknown   90
##
##       death 1117    0    0.0                                     No 1023
##                                                                  Yes   94
##
##   max_stage 1117    0    0.0                                      4  251
##                                                                   5  476
##                                                                   6  107
##                                                                   7    6
##                                                                   8   68
##                                                                   9  115
##                                                                  10   94
##
##   stage_adm 1117    0    0.0                                      4  517
##                                                                   5  397
##                                                                   6   79
##                                                                   7   11
##                                                                   8   50
##                                                                   9   63
##
##  stage_disc 1117    0    0.0                                      4  705
##                                                                   5  282
##                                                                   6   11
##                                                                   7   20
##                                                                   8    3
##                                                                   9    2
##                                                                  10   94
##
##  percent cum.percent
##     48.4        48.4
##     51.6       100.0
##
##     19.8        19.8
##      9.8        29.6
##      1.6        31.2
##     30.5        61.8
##      3.0        64.7
##     35.3       100.0
##
##     32.0        32.0
##     65.3        97.2
##      2.8       100.0
##
##      5.2         5.2
##     86.8        91.9
##      8.1       100.0
##
##     91.6        91.6
##      8.4       100.0
##
##     22.5        22.5
##     42.6        65.1
##      9.6        74.7
##      0.5        75.2
```

```
##      6.1        81.3
##     10.3        91.6
##      8.4       100.0
##
##     46.3        46.3
##     35.5        81.8
##      7.1        88.9
##      1.0        89.9
##      4.5        94.4
##      5.6       100.0
##
##     63.1        63.1
##     25.2        88.4
##      1.0        89.3
##      1.8        91.1
##      0.3        91.4
##      0.2        91.6
##      8.4       100.0
##
```

```r
# tableone::print.CreateContTable
t2 <- tableone::CreateContTable(data = df, cont_vars)
print(t2, nonnormal = "LOS", digits = 1)
```

```
##
##                       Overall
##   n                   1117
##   age (mean (SD))     58.6 (19.6)
##   BMI (mean (SD))     28.5 (7.9)
##   LOS (median [IQR])  7.0 [4.0, 14.0]
```

```r
# tableone::summary.CreateContTable
summary(t2)
```

```
## strata: Overall
##        n miss p.miss mean   sd median p25 p75 min max   skew  kurt
## age 1117    0    0.0   59 19.6     60  44  72  18 104 -0.064 -0.74
## BMI 1117   41    3.7   28  7.9     27  23  32  12  79  1.441  4.21
## LOS 1117    0    0.0   13 18.4      7   4  14   1 231  4.638 32.31
```

```r
# Stage transition matrices
# Create categorical table of stage_adm vs. stage_disc
CreateCatTable(strata = 'stage_adm', vars = 'stage_disc', data = df)
```

```
##                  Stratified by stage_adm
##                   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)
```

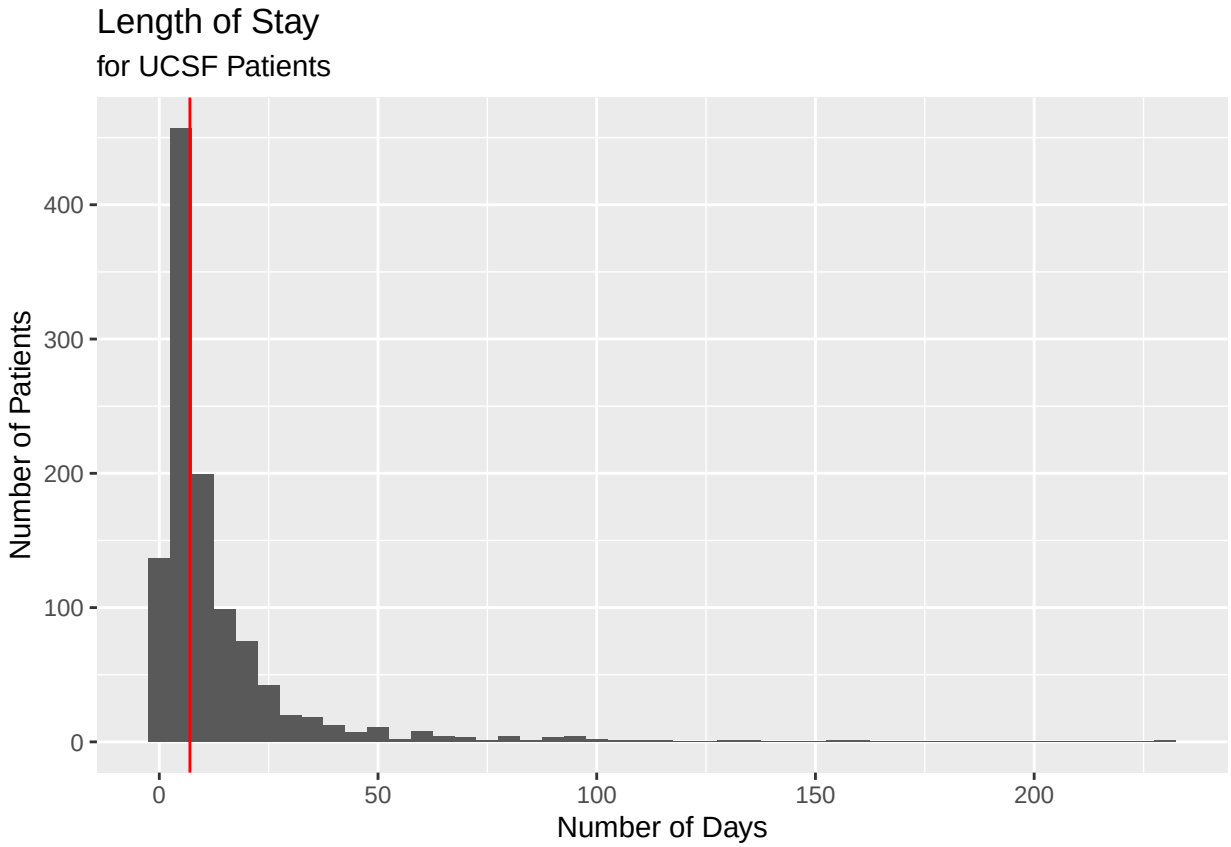
```
##      8             0 ( 0.0)    0 ( 0.0)   0 ( 0.0)  1 ( 9.1)   1 ( 2.0)
##      9             0 ( 0.0)    0 ( 0.0)   1 ( 1.3)  0 ( 0.0)   0 ( 0.0)
##      10           21 ( 4.1)   29 ( 7.3)  13 (16.5)  0 ( 0.0)  10 (20.0)
##                  Stratified by stage_adm
##                   9          p      test
##   n                63
##   stage_disc (%)              <0.001
##      4            17 (27.0)
##      5            19 (30.2)
##      6             1 ( 1.6)
##      7             3 ( 4.8)
##      8             1 ( 1.6)
##      9             1 ( 1.6)
##      10           21 (33.3)
```

```
# Create categorical table of stage_adm vs. max_stage
CreateCatTable(strata = 'stage_adm', vars = 'max_stage', data = df)
```

```
##                 Stratified by stage_adm
##                  4           5           6          7         8
##   n              517         397         79         11        50
##   max_stage (%)
##      4           251 (48.5)    0 ( 0.0)   0 ( 0.0)  0 ( 0.0)   0 ( 0.0)
##      5           203 (39.3)  273 (68.8)   0 ( 0.0)  0 ( 0.0)   0 ( 0.0)
##      6            10 ( 1.9)   55 (13.9)  42 (53.2)  0 ( 0.0)   0 ( 0.0)
##      7             1 ( 0.2)    1 ( 0.3)   0 ( 0.0)  4 (36.4)   0 ( 0.0)
##      8            16 ( 3.1)   18 ( 4.5)   5 ( 6.3)  4 (36.4)  25 (50.0)
##      9            15 ( 2.9)   21 ( 5.3)  19 (24.1)  3 (27.3)  15 (30.0)
##      10           21 ( 4.1)   29 ( 7.3)  13 (16.5)  0 ( 0.0)  10 (20.0)
##                  Stratified by stage_adm
##                   9          p      test
##   n                63
##   max_stage (%)               <0.001
##      4             0 ( 0.0)
##      5             0 ( 0.0)
##      6             0 ( 0.0)
##      7             0 ( 0.0)
##      8             0 ( 0.0)
##      9            42 (66.7)
##      10           21 (33.3)
```

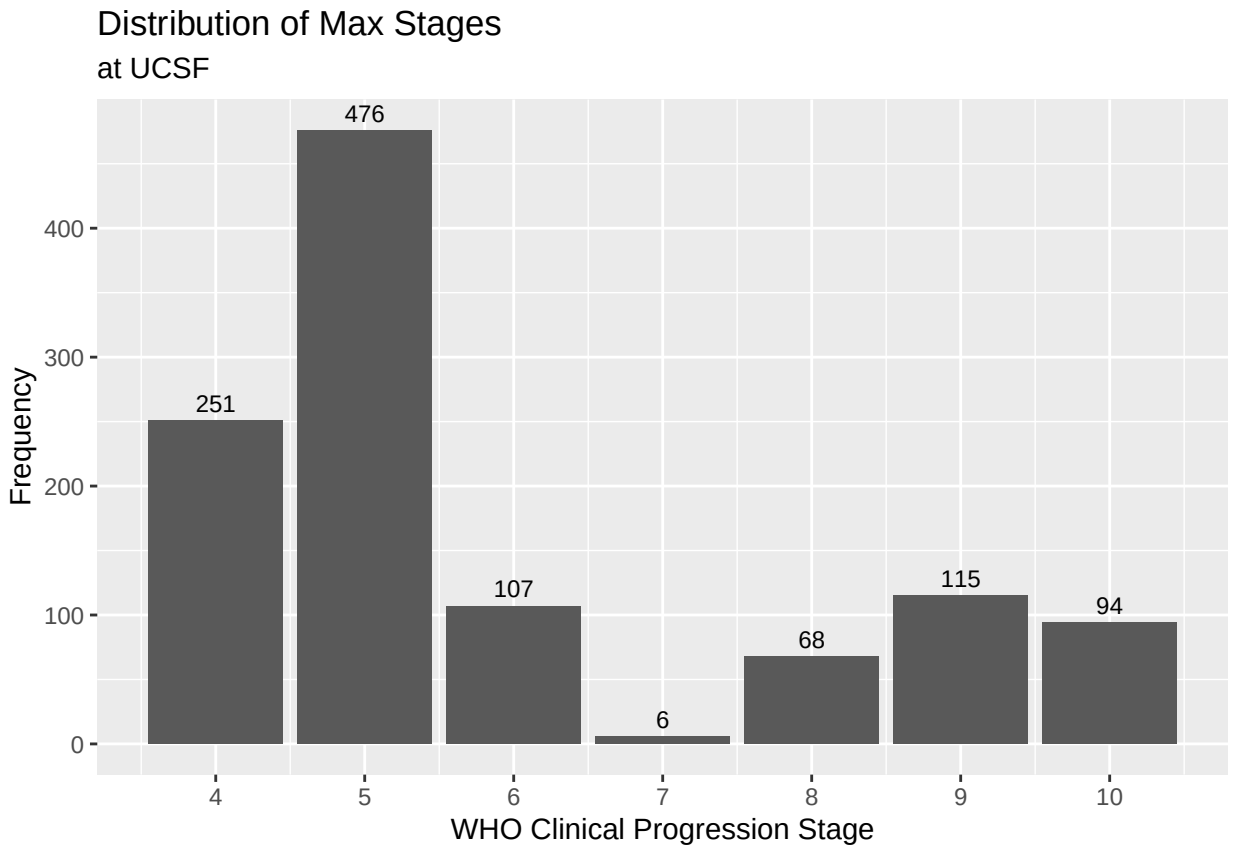
Create histogram of LOS

```
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 dstribution 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)), size = 3, vjust = -0.5)
```



Run correlation stats

```
# max_stage and LOS
cor.test(x = df$max_stage, y = df$LOS, method = 'pearson')
```

```
##
##  Pearson's product-moment correlation
##
## data:  df$max_stage and df$LOS
## t = 18.781, df = 1115, p-value < 2.2e-16
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  0.4443494 0.5335412
## sample estimates:
##       cor
## 0.4902276
```

```
# max_stage and stage_disc
cor.test(x = df$max_stage, y = df$stage_disc, method = 'pearson')
```

```
##
##  Pearson's product-moment correlation
##
## data:  df$max_stage and df$stage_disc
## t = 33.663, df = 1115, p-value < 2.2e-16
## alternative hypothesis: true correlation is not equal to 0
```

```
## 95 percent confidence interval:
##  0.6796087 0.7378902
## sample estimates:
##       cor
## 0.7099629
```

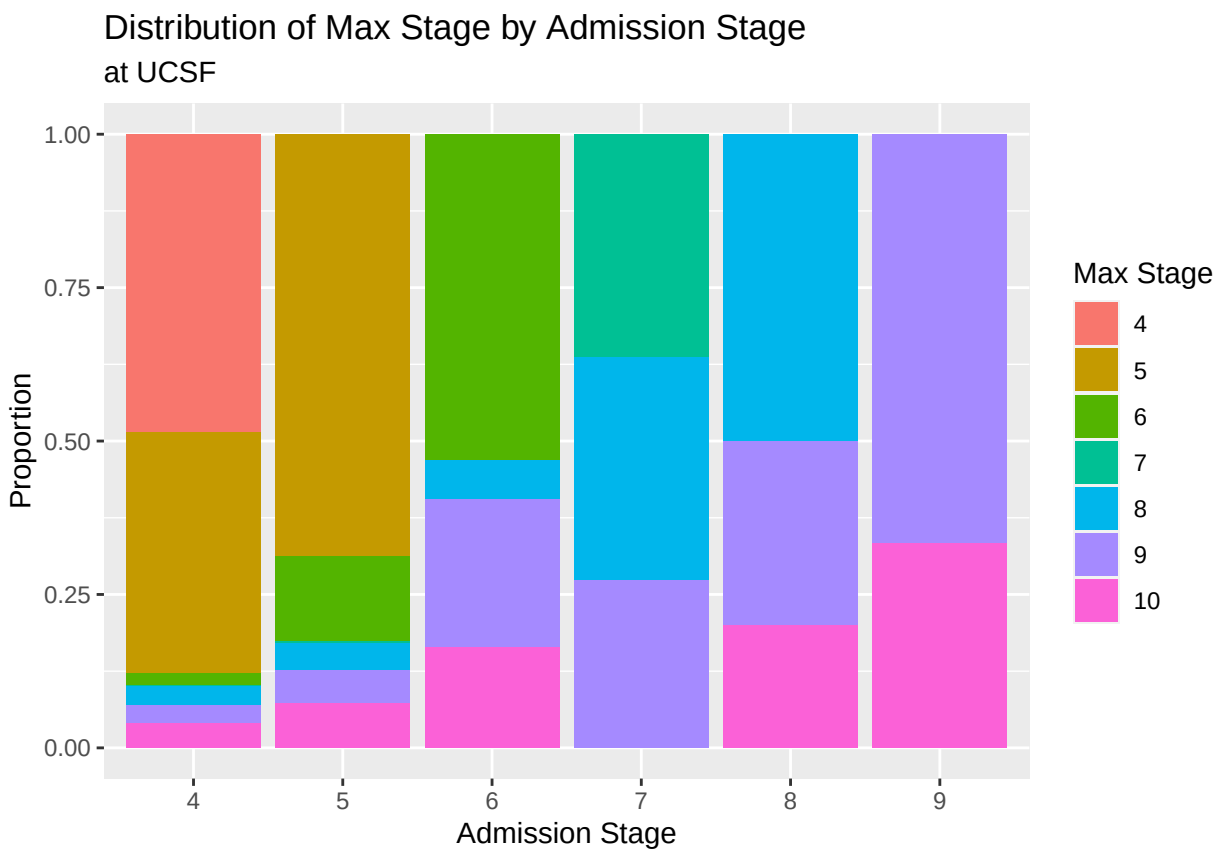
```
# max_stage and stage_adm
cor.test(x = df$max_stage, y = df$stage_adm, method = 'pearson')
```

```
##
##  Pearson's product-moment correlation
##
## data:  df$max_stage and df$stage_adm
## t = 28.528, df = 1115, p-value < 2.2e-16
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  0.6143151 0.6822269
## sample estimates:
##       cor
## 0.6495647
```

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') #+
```

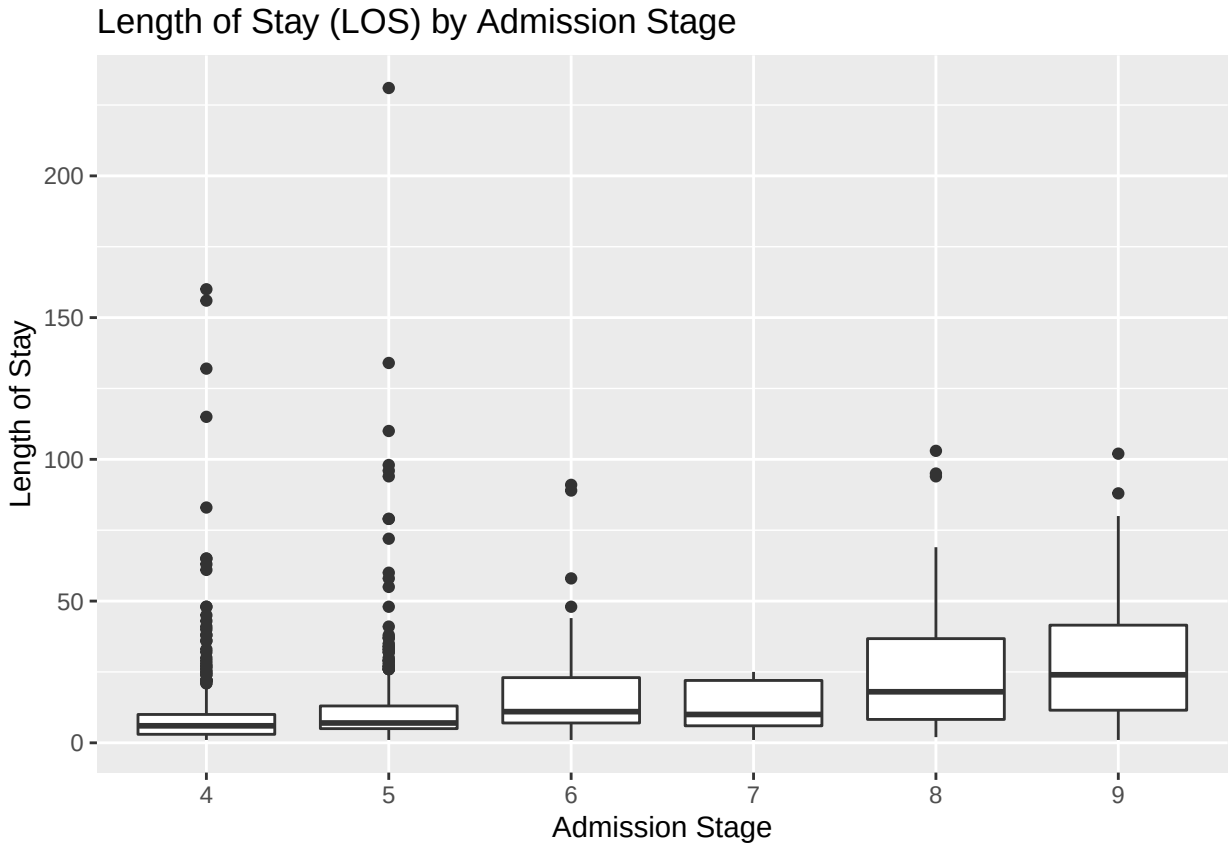
```
## `summarise()` has grouped output by 'stage_adm'. You can override using the `.groups` argument.
```




```
  #geom_text(aes(label = stat(y), group = stage_adm), stat = 'summary', fun = sum, vjust = -1)
```

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')
```



Transform admission stage, discharge stage, and max stage to stage categories where # 4-5 = moderate, 6-9 = severe, and 10 = dead

```
# Collapse levels of `stage_disc` to moderate, severe, and dead --> `stgcat_disc`
df$stgcat_disc <- fct_collapse(as.character(df$stage_disc),
                           "Moderate" = c("4", "5"),
                           "Severe" = c("6", "7", "8", "9"),
                           "Dead" = "10")
df$stgcat_disc <- factor(df$stgcat_disc, levels=c("Moderate", "Severe", "Dead"))
levels(df$stgcat_disc)
```

```
## [1] "Moderate" "Severe"   "Dead"
```

```
# Collapse levels of `stage_adm` to moderate, severe, and dead --> `stgcat_adm`
df$stgcat_adm <- fct_collapse(as.character(df$stage_adm),
                           "Moderate" = c("4", "5"),
                           "Severe" = c("6", "7", "8", "9"),
                           "Dead" = "10")
```

```
## Warning: Unknown levels in `f`: 10
```

```
df$stgcat_adm <- factor(df$stgcat_adm, levels=c("Moderate", "Severe", "Dead"))
levels(df$stgcat_adm)
```

```
## [1] "Moderate" "Severe"   "Dead"
```

```
# Collapse levels of `max_stage` to moderate, severe, and dead --> `stgcat_max`
df$stgcat_max <- fct_collapse(as.character(df$max_stage),
                              "Moderate" = c("4", "5"),
                              "Severe" = c("6", "7", "8", "9"),
                              "Dead" = "10")
df$stgcat_max <- factor(df$stgcat_max, levels=c("Moderate", "Severe", "Dead"))
levels(df$stgcat_max)
```

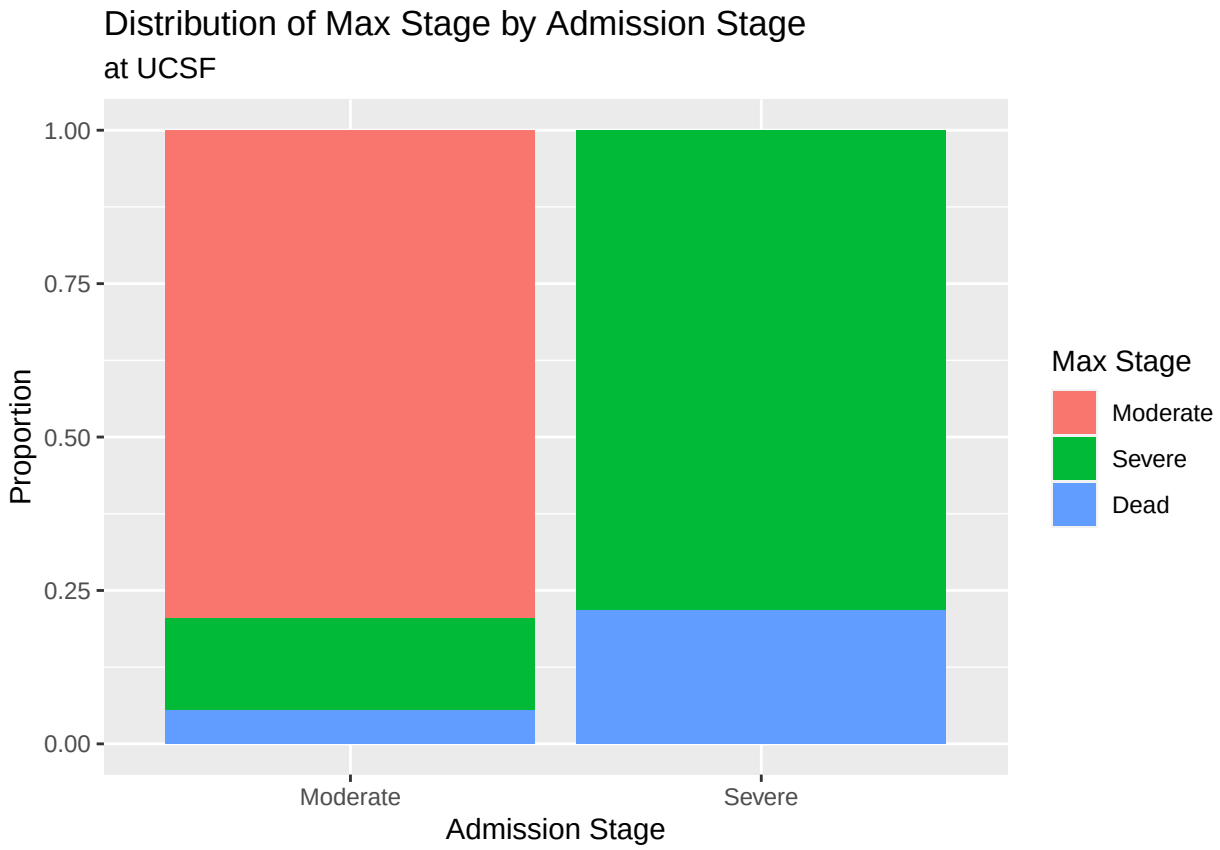
```
## [1] "Moderate" "Severe"   "Dead"
```

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

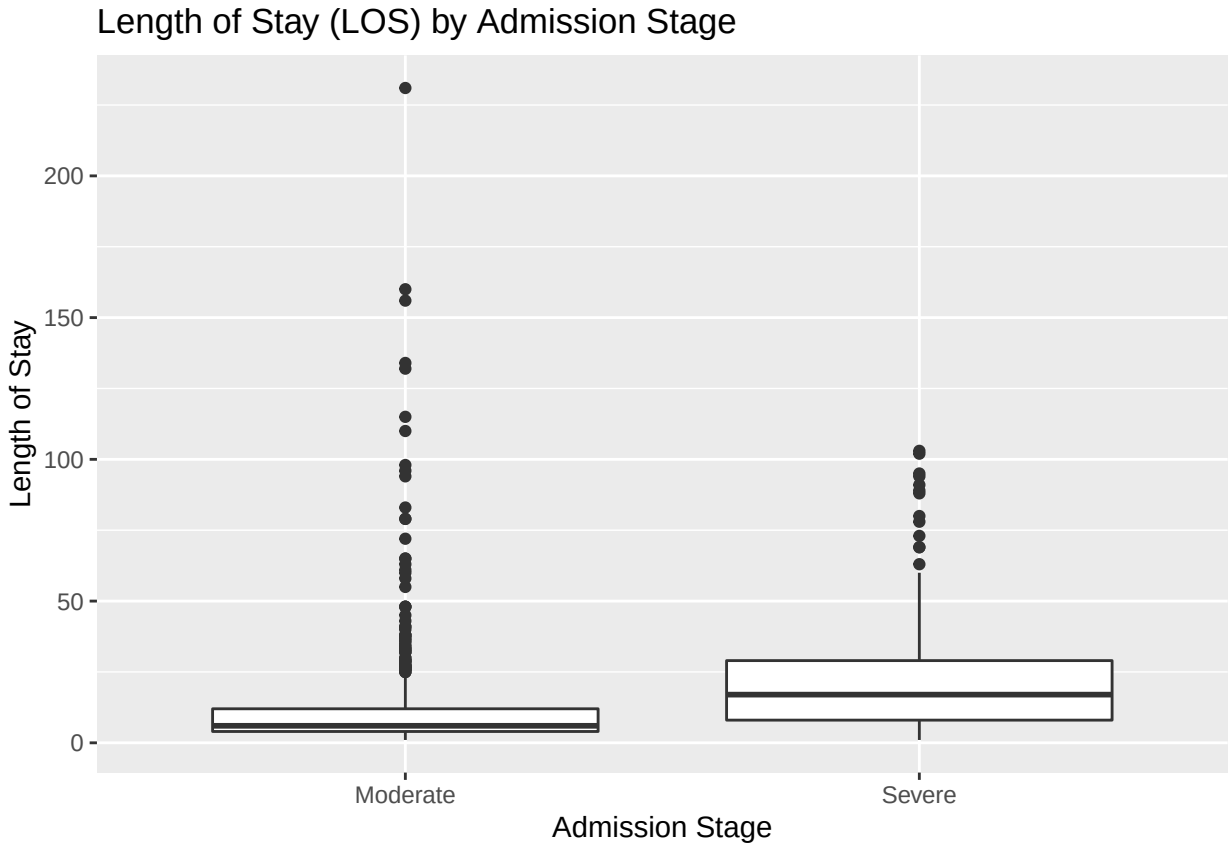
categories (moderate, severe, dead)

```
# Proportional stacked bar chart showing which proportion of patients admitted at moderate and severe s
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 override using the `.groups` argument.
```




```
# Boxplot showing median LOS by admission stage category (moderate vs. severe)
ggplot(df, aes(x = stgcat_adm, y = LOS)) +
  geom_boxplot() +
  labs(title = 'Length of Stay (LOS) by Admission Stage',
       x = 'Admission Stage',
       y = 'Length of Stay')
```



End of Document