Bios 6301: Assignment 1

Yiqing Pan

Due Thursday, 05 September, 1:00 PM

50 points total.

Add your name as `author` to the file's metadata section.

Submit a single quarto file (named `homework1.qmd`) by email to huiding.chen@vanderbilt.edu. Place your R code in between the appropriate chunks for each question. Check your output by using the `Render` button in RStudio.

Create a Data Set

A data set in R is called a data.frame. This particular data set is made of three categorical variables, or factors: `gender`, `smoker`, and `exercise`. In addition `exercise` is an ordered factor. `age` and `los` (length of stay) are continuous variables.

```
##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':
##
##     filter, lag

## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```r
gender <- c('M','M','F','M','F','F','M','F','M')
age <- c(34, 64, 38, 63, 40, 73, 27, 51, 47)
smoker <- c('no','yes','no','no','yes','no','no','no','yes')
exercise <- factor(c('moderate','frequent','some','some','moderate','none','none',
                     'moderate','moderate'),
                   levels=c('none','some','moderate','frequent'), ordered=TRUE
)
los <- c(4,8,1,10,6,3,9,4,8)
x <- data.frame(gender, age, smoker, exercise, los)
x
```

```
##   gender age smoker exercise los
## 1      M  34     no moderate   4
## 2      M  64    yes frequent   8
## 3      F  38     no     some   1
## 4      M  63     no     some  10
```

```
## 5      F  40    yes moderate    6
## 6      F  73     no     none    3
## 7      M  27     no     none    9
## 8      F  51     no moderate    4
## 9      M  47    yes moderate    8
```