Lab 3

Name here

Q1. (6 points)

We previously calculated the marginal probability that dealer "busts" in black jack. Now consider the conditional probability of a bust, give a specific card showing. In particular, what is the probability that dealer busts given they have

- a Jack showing
- a 6 showing
- a 2 showing

To simplify the calculation (albeit not being entirely correct), you can treat all Aces as being worth 11 points. Note that the dealer must continue to take cards until the total is 17 or more, at which point the dealer must stand.

```
library(tidyverse)
```

```
## -- Attaching packages --------------------------------------- tidyverse 1.3.2 --
## v ggplot2 3.3.6      v purrr   0.3.5
## v tibble  3.1.8      v dplyr   1.0.10
## v tidyr   1.2.1      v stringr 1.4.1
## v readr   2.1.2      v forcats 0.5.2
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
blackjack <- function(faceup, num_hands = 10000){
  bust <- rep(FALSE, num_hands)
  values <- tibble(cards = rep(c(2:10,'Jack',"Queen","King","Ace"), each = 4),
                   points = rep(c(2:10,10,10,10,11), each = 4))
  remaining_deck <- values %>% filter(cards != faceup) %>%
    bind_rows(values %>% filter(cards == faceup) %>% slice(1:3))

  points <- values %>%
                   filter(cards == faceup) %>%
                   slice(1) %>%
                   select(points) %>% pull()
  for (i in 1:num_hands){
    shuffle <- c(points,sample(remaining_deck$points))
    num_cards <- 2
    total <- sum(shuffle[1:num_cards])
    while (total < 17){
      num_cards <- num_cards + 1
      total <- sum(shuffle[1:num_cards])
    }
    if (total > 21) {
      bust[i] = TRUE
```

```
    }
  }
  mean(bust)
}
jack <- blackjack("Jack")
six <- blackjack("6")
two <- blackjack('2')
```

- With a jack, the dealer will bust with probability equal to 0.23
- With a six, the dealer will bust with probability equal to 0.46
- With a two, the dealer will bust with probability equal to 0.42

Q2. (12 points)

Use a dataset containing homes in the Seattle, WA area http://www.math.montana.edu/ahoegh/teaching/stat408/datasets/SeattleHousing.csv for this question.

```
library(tidyverse)
seattle <- read_csv('http://www.math.montana.edu/ahoegh/teaching/stat408/datasets/SeattleHousing.csv') %
  mutate(bath_category = case_when(
      bathrooms <= 2 ~ '0-2 bathrooms',
      bathrooms > 2 & bathrooms <=4 ~ '2-4 bathrooms',
      bathrooms > 4 ~ 'more than 4 bathrooms'),
    bed_category = case_when(
      bedrooms == 0 ~ '0 beds',
      bedrooms > 0 & bedrooms <= 2 ~ '1-2 beds',
      bedrooms > 2 & bedrooms <= 4 ~ '3-4 beds',
      bedrooms > 4 ~ 'more than 4 beds'),
    floor_category = case_when(
      floors < 2 ~ 'less than 2 floors',
      floors >= 2 & floors <= 3 ~ '2-3 floors',
      floors > 3 ~ 'more than 3 floors'
    )
   )
```

```
## Rows: 869 Columns: 14
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## dbl (14): price, bedrooms, bathrooms, sqft_living, sqft_lot, floors, waterfr...
##
## 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.
```

a. (2 points) Create a the two-by-two table containing the joint probabilities of bathrooms (grouped as: 0 - 2, more than 2 - 4, more than 4) and floors: (less than 2, 2 - 3, more than 3).

```
round(table(seattle$bath_category, seattle$floor_category) / nrow(seattle), 3)
```

```
##
##                          2-3 floors less than 2 floors more than 3 floors
##   0-2 bathrooms               0.084              0.464              0.001
##   2-4 bathrooms               0.291              0.132              0.000
##   more than 4 bathrooms       0.028              0.000              0.000
```

b.(2 points) Compute the joint probability of 0 - 2 bathrooms and less than 2 floors

The probability is 0.464

c. (2 points) Compute marginal probability of having 0 - 2 bathrooms

The probability is 0.549

d. (2 points) Compute conditional probability of having less than 2 floors given having 0 - 2 bathrooms

The probability is 0.845

e. (2 points) Compute conditional probability of having 0 - 2 bathrooms given there are less than 2 floors

The probability is 0.778

f. (2 points) Are bathrooms and number of floors independent? Why or why not.

No the probability of the number of bathrooms changes based on the number of floors.

Q3. (6 points)

Bayes Rule is commonly applied with two binary variables such as test outcomes and disease outcomes. However, it can also apply with three categories.

Suppose we are interested in understanding the probability that an injured person was skiing, snowboarding, or telemark skiing. (Formally, P(skiing | injured), P(snowboarding|injured), and P(telemark|injured)).

Suppose

- Prob(Skier) = .7
- Prob(Snowboard) = .25
- Prob(Telemark) = .05

and

- Prob(injured|skier) = .02
- Prob(injured|snowboard) = .01
- Prob(injured|telemark) = .10

Calculate

- P(skiing | injured)

```
p_ski <- .7
p_snowboard <- .25
p_telemark <- .05
p_inj_ski <- .02
p_inj_sb <- .01
p_inj_t <- .10

p_ski_inj <- p_ski * p_inj_ski / (p_ski * p_inj_ski + p_snowboard * p_inj_sb + p_telemark * p_inj_t)
```

This probability would be 0.65

- P(snowboard | injured)

```
p_sb_inj <- p_snowboard * p_inj_sb / (p_ski * p_inj_ski + p_snowboard * p_inj_sb + p_telemark * p_inj_t
```

This probability would be 0.12

- P(telemark | injured)

```
p_tele_inj <- p_telemark * p_inj_t / (p_ski * p_inj_ski + p_snowboard * p_inj_sb + p_telemark * p_inj_t)
```

This probability would be 0.23