Lab 05

due February 18 at 11:59 PM

Group 2-8: Dav King, Eesha Yaqub, Reesey Du Pont, Vivian Zhang

```
library(tidyverse)
library(knitr)
library(viridis)
```

```
courage <- read_csv("courage.csv")
```

Exercise 1

There are 78 observations in the dataset, with each observation representing one game that the NC Courage played.

Exercise 2

```
seasonal_courage = courage %>%
  mutate(seasonal_category = case_when(
    game_number <= 9 ~ "early",
    game_number <= 17 ~ "middle",
    game_number <= 26 ~ "late"
  )) %>%
  mutate(win = if_else(result == "win", 1, 0)) %>%
  group_by(seasonal_category) %>%
  summarize(prop_win = mean(win))
seasonal_courage
```

```
## # A tibble: 3 x 2
##   seasonal_category prop_win
##   <chr>                <dbl>
## 1 early                0.593
## 2 late                 0.704
## 3 middle               0.75
```

This table shows the conditional probability that the Courage win a soccer game, given the point in the season that the game was played. The conditional probability that the team won a game given that it was early in the season is 0.593, given that it was late in the season is 0.704, and given that it was the middle of the season is 0.75.

Exercise 3

```
seasonal_courage %>%
   mutate(seasonal_category =
            factor(seasonal_category,
                   levels = c("early", "middle", "late"),
                   ordered = TRUE)) %>%
  arrange(seasonal_category)
```

```
## # A tibble: 3 x 2
##   seasonal_category prop_win
##   <ord>                <dbl>
## 1 early                0.593
## 2 middle               0.75
## 3 late                 0.704
```

Exercise 4

```
courage %>%
  mutate(Home_Game = case_when(home_team == "NC" ~ 1, T ~ 0)) %>%
  mutate(win = case_when( result == "win" ~ 1, T ~ 0 )) %>%
  mutate(pwin = mean(win)) %>%
  mutate(pHome_Game = mean(Home_Game)) %>%
  filter(result == "win") %>%
  mutate(pwin_athome = mean(Home_Game)) %>%
  mutate(pwin_Home_Game = ((pwin_athome)*pwin)/ pHome_Game) %>%
  summarize(pwin, pwin_Home_Game)
```

```
## # A tibble: 53 x 2
##     pwin pwin_Home_Game
##    <dbl>          <dbl>
##  1 0.679          0.738
##  2 0.679          0.738
##  3 0.679          0.738
##  4 0.679          0.738
##  5 0.679          0.738
##  6 0.679          0.738
##  7 0.679          0.738
##  8 0.679          0.738
##  9 0.679          0.738
## 10 0.679          0.738
## # ... with 43 more rows
```

P(win) = 0.679 Conditional Probability = 0.738 No because these two values are different. If the two events were independent, we would expect the probability of a win to be the exact same as the probability of a win given that it was a home game. Instead, the probability of winning at home is much higher. This suggests a strong home-field advantage for the Courage - they win much more frequently at home.

Exercise 5

```
courage <- courage %>%
  mutate(home_courage = if_else(home_team == "NC", "home", "away"))
courage %>%
count(result, home_courage) %>%
  pivot_wider(id_cols = c(result, home_courage),
              names_from = home_courage,
              values_from = n,
              values_fill = 0) %>%
  kable()
```

result	away	home
loss	9	5
tie	5	6
win	22	31

P(home) = (42/78) = 0.538

P(tie) = (11/78) = 0.141

P(home|tie) = (6/11) = 0.545

Finding P(tie|home) using Bayes' Theorem: P(tie|home) = ((0.545)(0.141))/(0.538) = 0.143

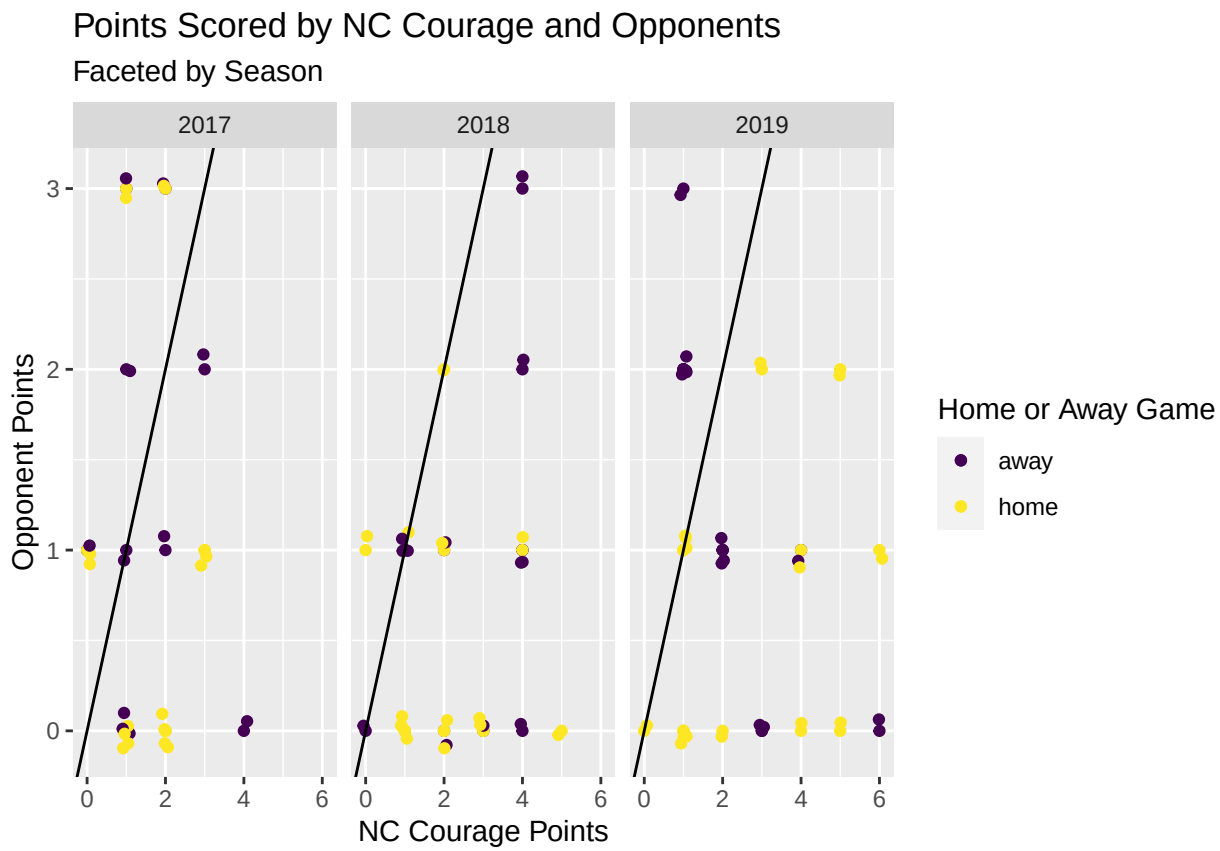
Checking Bayes' Theorem using the contingency table: P(tie|home) = (6/42) = 0.143

The probability of a tie is not independent of the Courage playing at home or away because when comparing the conditional probability of a tie occurring when the courage plays at home (or P(tie|home)) is not equal to the marginal probability of a tie occurring (or P(tie)).

Exercise 6

```
courage <- courage %>%
  mutate(total_pts = home_pts + away_pts) %>%
  mutate(courage_pts = if_else(home_courage == "home", home_pts, away_pts)) %>%
  mutate(opponent_pts = if_else(home_courage == "away", home_pts, away_pts))
```

```
courage %>%
  ggplot(mapping = aes(x = courage_pts,
                       y = opponent_pts, color = home_courage)) +
  geom_point() +
  scale_color_viridis(discrete = TRUE, option = "D") +
  labs(title = "Points Scored by NC Courage and Opponents",
       subtitle = "Faceted by Season", x = "NC Courage Points",
       y = "Opponent Points", color = "Home or Away Game") +
  geom_jitter(width = 0.1, height = 0.1) +
  geom_abline(slope = 1, intercept = 0) +
  facet_wrap(~ season)
```



The graphs show that the majority of the highest-scoring games for NC Courage were at home, while most of the games where the opponent outscored NC Courage (or where the opponent just scored high in general, even if they still scored lower than NC Courage) were away games. Additionally, NC Courage seems to have won the majority of their games, since in the graphs for each season, more points lie to the right of the y = x line rather than to the left (indicating an NC Courage win).