Final Project - Predicting March Madness

Anmol Sapru and Rohit Gunda

```r
library(tidyverse)
library(tidymodels)
library(Stat2Data)

cbb <- read_csv("data/cbb.csv")

#Remove non-postseason teams and R68 losers
cbb <- cbb[!is.na(cbb$POSTSEASON),]
cbb <- filter(cbb, !grepl("R68", POSTSEASON))

filter(cbb, POSTSEASON == "F4") |>
  arrange(YEAR)
```

```
# A tibble: 14 x 24
   TEAM       CONF      G     W ADJOE ADJDE BARTHAG EFG_O EFG_D  TOR  TORD   ORB
   <chr>      <chr> <dbl> <dbl> <dbl> <dbl>   <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>
 1 Syracuse   BE       40    30   113   87    0.953  49    42.6  18.9  23.6  38.8
 2 Wichita ~  MVC      39    30   111.  91    0.904  50    45.3  19.4  20    38
 3 Florida    SEC      39    36   116.  88.4  0.958  52.2  45.4  17.5  21.3  35.3
 4 Wisconsin  B10      38    30   123.  95.9  0.944  53.3  47.2  12.7  15.3  28.1
 5 Michigan~  B10      39    27   116.  92.6  0.933  53.2  44.9  17.5  16.5  33.5
 6 Kentucky   SEC      39    38   120.  84    0.984  51.5  39.6  16.3  21.3  39.5
 7 Syracuse   ACC      37    23   112.  93.6  0.886  50    47.3  18.1  20.4  33.5
 8 Oklahoma   B12      37    29   118.  94.1  0.933  54.3  47.2  18.3  17.7  31.2
 9 Oregon     P12      38    33   118.  93.8  0.932  55.5  46.4  17.1  19.3  32.4
10 South Ca~  SEC      36    26   109.  89    0.908  47.3  45.5  18.4  24.3  34
11 Kansas     B12      39    31   122.  96.4  0.938  57.5  49.1  16.6  18.2  29.4
12 Loyola C~  MVC      37    32   110.  95.4  0.844  57.7  47.4  19.1  19.4  22.1
13 Michigan~  B10      39    32   120.  91    0.960  55.2  43.9  18.5  14.9  33.9
14 Auburn     SEC      39    30   121.  95.2  0.942  54.4  51.7  16.8  24.6  31.6
# ... with 12 more variables: DRB <dbl>, FTR <dbl>, FTRD <dbl>, `2P_O` <dbl>,
```

```
#    `2P_D` <dbl>, `3P_O` <dbl>, `3P_D` <dbl>, ADJ_T <dbl>, WAB <dbl>,
#    POSTSEASON <chr>, SEED <dbl>, YEAR <dbl>
```

```r
#separated
cbb <- mutate(cbb, round_64 = if_else(POSTSEASON == "R64", FALSE, TRUE))

cbb <- mutate(cbb, round_32 =
        case_when(POSTSEASON == "R32" ~ FALSE,
        POSTSEASON %in% c("S16", "E8", "F4", "2ND", "Champions") ~ TRUE,
        TRUE ~ NA))

cbb <- mutate(cbb, sweet_sixteen =
        case_when(POSTSEASON == "S16" ~ FALSE,
        POSTSEASON %in% c("E8", "F4", "2ND", "Champions") ~ TRUE,
        TRUE ~ NA))

cbb <- mutate(cbb, elite_eight =
        case_when(POSTSEASON == "E8" ~ FALSE,
        POSTSEASON %in% c("F4", "2ND", "Champions") ~ TRUE,
        TRUE ~ NA))

cbb <- mutate(cbb, final_four =
        case_when(POSTSEASON == "F4" ~ FALSE,
        POSTSEASON %in% c("2ND", "Champions") ~ TRUE,
        TRUE ~ NA))

cbb <- mutate(cbb, champ_game =
        case_when(POSTSEASON == "2ND" ~ FALSE,
        POSTSEASON %in% c("Champions") ~ TRUE,
        TRUE ~ NA))
```