Adams_FinalProject

Grace Adams

2024-05-07

Final Project

Problem Decomposition

Hello! I decided to base my final project in a dataset I found on kaggle.com. The dataset contains statistics from all the NCAA Men's Division 1 College Basketball teams in the U.S. from 2013 to 2023. I chose the dataset while I was out of town traveling for the NCAA March Madness tournament, which I was very invested in. I thought it might be neat to look at the stats of some of the teams I was seeing, and of the PAC-12 conference as it is dissolving after the 2023-24 season.

Where I use things from each week:

- Week 2: Throughout entire project.
- Week 8: Mainly in the plotting section and Overarching Conclusions section.
- Week 9: Project is a GitHub repo.
- Week 5: Data Visualizations are the entirety of the plotting section.
- Week 3: I used the filter, mutate and select functions we learned in week 3 throughout the project to select the columns I wanted from the original dataframe, make new columns, and select rows that fit certain conditions.
- Week 11: I used the case_when() function that we learned in week 11 to make the W_PERC_CLASS column.

I've specified in each code chunk what I am doing and what week it came from, and I've also compiled a list in the README file for the project that tells which lines of code correspond to what week.

A handful of FYI's about the data:

- cbb.csv has seasons 2013-2019 and seasons 2021-2023 combined
- The 2020 season's data set is kept separate from the other seasons, because there was no postseason due to the Coronavirus. I don't use data from this season because it was so abnormal.
- The initial cbb.csv file had a lot of columns with statistics I didn't want to look at.
- Data Dictionary:
  - TEAM: The Division I college basketball school
  - CONF: The Athletic Conference in which the school participates in

    * A10 = Atlantic 10,
    * ACC = Atlantic Coast Conference,
    * AE = America East,
    * Amer = American,
    * ASun = ASUN,
    * B10 = Big Ten,
    * B12 = Big 12,
    * BE = Big East,
    * BSky = Big Sky,
    * BSth = Big South,
    * BW = Big West,
    * CAA = Colonial Athletic Association,
    * CUSA = Conference USA,
    * Horz = Horizon League,
    * Ivy = Ivy League,
    * MAAC = Metro Atlantic Athletic Conference,
    * MAC = Mid-American Conference,
    * MEAC = Mid-Eastern Athletic Conference,
    * MVC = Missouri Valley Conference,
    * MWC = Mountain West,
    * NEC = Northeast Conference,
    * OVC = Ohio Valley Conference,
    * P12 = Pac-12,
    * Pat = Patriot League,
    * SB = Sun Belt,
    * SC = Southern Conference,
    * SEC = South Eastern Conference,
    * Slnd = Southland Conference,
    * Sum = Summit League,
    * SWAC = Southwestern Athletic Conference,
    * WAC = Western Athletic Conference,
    * WCC = West Coast Conference)
- G: Number of games played
- W: Number of games won
- ADJOE: Adjusted Offensive Efficiency (An estimate of the offensive efficiency (points scored per 100 possessions) a team would have against the average Division I defense)
- ADJDE: Adjusted Defensive Efficiency (An estimate of the defensive efficiency (points allowed per 100 possessions) a team would have against the average Division I offense)
- BARTHAG: Power Rating (Chance of beating an average Division I team)
- EFG_O: Effective Field Goal Percentage Shot
- EFG_D: Effective Field Goal Percentage Allowed
- TOR: Turnover Percentage Allowed (Turnover Rate)
- TORD: Turnover Percentage Committed (Steal Rate)
- ORB: Offensive Rebound Rate
- DRB: Offensive Rebound Rate Allowed
- FTR : Free Throw Rate (How often the given team shoots Free Throws)
- FTRD: Free Throw Rate Allowed
- 2P_O: Two-Point Shooting Percentage
- 2P_D: Two-Point Shooting Percentage Allowed

- 3P_O: Three-Point Shooting Percentage
- 3P_D: Three-Point Shooting Percentage Allowed
- ADJ_T: Adjusted Tempo (An estimate of the tempo (possessions per 40 minutes) a team would have against the team that wants to play at an average Division I tempo)
- WAB: Wins Above Bubble (The bubble refers to the cut off between making the NCAA March Madness Tournament and not making it)
- POSTSEASON: Round where the given team was eliminated or where their season ended
  * R68 = First Four,
  * R64 = Round of 64,
  * R32 = Round of 32,
  * S16 = Sweet Sixteen,
  * E8 = Elite Eight,
  * F4 = Final Four,
  * 2ND = Runner-up,
  * Champion = Winner of the NCAA March Madness Tournament for that given year
- SEED: Seed in the NCAA March Madness Tournament
- YEAR: Season

In this project I load in packages, make sure my working directory is correct, load in the files from the cbb folder, explore the data, select the columns I need, turn a few columns into factors, make a few plots, make a few new columns in the dataframe, and make a few more final plots.

Loading packages

```
#install.packages("tidyverse")
library(tidyverse)
```

```
## -- Attaching core tidyverse packages ------------------------ tidyverse 2.0.0 --
## v dplyr     1.1.4     v readr     2.1.5
## v forcats   1.0.0     v stringr   1.5.1
## v ggplot2   3.5.0     v tibble    3.2.1
## v lubridate 1.9.3     v tidyr     1.3.1
## v purrr     1.0.2
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors
```

```
#install.packages("ggplot2")
library(ggplot2)

#install.packages("ggpubr")
library(ggpubr)

#install.packages("RColorBrewer)
library(RColorBrewer)
```

Reading in the data

```
#checking working directory
getwd()
```

```
## [1] "/Users/graceadams/Documents/Git/Adams_FinalProject"
```

```
# create object to store directory of the files you want to use
folder_path <- "./data_raw/cbb"
```

```
# create an object that is a list of all .csv files at the directory specified above
file_list <- list.files (folder_path, pattern = "cbb[0-9]{2}\\.csv")
```

```
# Loop through each file and read it into the environment
for (file in file_list) {
  # extract the two-digit number from the file name
  file_number <- gsub("cbb|\\.csv", "", file)

  # read the .csv file into a data frame with a name based on the file number
  assign(paste0("cbb", file_number), read.csv(file.path(folder_path, file)))
}

# load in cbb.csv file on its own
cbb <- read_csv("./data_raw/cbb/cbb.csv")
```

```
## Rows: 3523 Columns: 24
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr  (4): TEAM, CONF, POSTSEASON, SEED
## dbl (20): G, W, ADJOE, ADJDE, BARTHAG, EFG_O, EFG_D, TOR, TORD, ORB, DRB, FT...
##
## 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.
```

Data Exploration

```
# two different ways of checking out the data
summary(cbb)
```

```
##      TEAM               CONF                 G               W
##  Length:3523        Length:3523        Min.   : 5.00   Min.   : 0.00
##  Class :character   Class :character   1st Qu.:29.00   1st Qu.:11.00
##  Mode  :character   Mode  :character   Median :31.00   Median :16.00
##                                        Mean   :30.49   Mean   :15.99
##                                        3rd Qu.:33.00   3rd Qu.:21.00
##                                        Max.   :40.00   Max.   :38.00
##      ADJOE           ADJDE          BARTHAG           EFG_O
##  Min.   : 76.6   Min.   : 84.0   Min.   :0.0050   Min.   :39.20
##  1st Qu.: 98.2   1st Qu.: 98.4   1st Qu.:0.2813   1st Qu.:47.90
```

```
##  Median :102.8   Median :103.2   Median :0.4756   Median :49.80
##  Mean   :103.2   Mean   :103.2   Mean   :0.4941   Mean   :49.89
##  3rd Qu.:107.9   3rd Qu.:107.8   3rd Qu.:0.7143   3rd Qu.:51.90
##  Max.   :129.1   Max.   :124.0   Max.   :0.9842   Max.   :61.00
##      EFG_D            TOR             TORD            ORB             DRB
##  Min.   :39.60   Min.   :11.9   Min.   :10.20   Min.   :14.40   Min.   :18.40
##  1st Qu.:48.10   1st Qu.:17.3   1st Qu.:17.10   1st Qu.:26.50   1st Qu.:27.30
##  Median :50.10   Median :18.6   Median :18.50   Median :29.40   Median :29.50
##  Mean   :50.09   Mean   :18.7   Mean   :18.63   Mean   :29.31   Mean   :29.52
##  3rd Qu.:52.00   3rd Qu.:20.0   3rd Qu.:20.10   3rd Qu.:32.10   3rd Qu.:31.70
##  Max.   :60.10   Max.   :27.1   Max.   :28.50   Max.   :43.60   Max.   :40.40
##       FTR             FTRD           2P_O            2P_D            3P_O
##  Min.   :19.60   Min.   :16.5   Min.   :37.70   Min.   :37.70   Min.   :24.90
##  1st Qu.:30.60   1st Qu.:30.2   1st Qu.:46.80   1st Qu.:47.10   1st Qu.:32.30
##  Median :34.30   Median :34.1   Median :49.00   Median :49.30   Median :34.10
##  Mean   :34.53   Mean   :34.8   Mean   :49.11   Mean   :49.29   Mean   :34.19
##  3rd Qu.:38.10   3rd Qu.:38.8   3rd Qu.:51.30   3rd Qu.:51.50   3rd Qu.:36.00
##  Max.   :58.60   Max.   :60.7   Max.   :64.00   Max.   :61.20   Max.   :44.10
##       3P_D            ADJ_T            WAB          POSTSEASON
##  Min.   :26.10   Min.   :57.20   Min.   :-25.20   Length:3523
##  1st Qu.:32.70   1st Qu.:65.70   1st Qu.:-12.60   Class :character
##  Median :34.30   Median :67.70   Median : -7.90   Mode  :character
##  Mean   :34.37   Mean   :67.74   Mean   : -7.58
##  3rd Qu.:36.00   3rd Qu.:69.70   3rd Qu.: -3.00
##  Max.   :43.10   Max.   :83.40   Max.   : 13.10
##      SEED                YEAR
##  Length:3523         Min.   :2013
##  Class :character    1st Qu.:2015
##  Mode  :character    Median :2018
##                      Mean   :2018
##                      3rd Qu.:2021
##                      Max.   :2023
```

```
glimpse(cbb)
```

```
## Rows: 3,523
## Columns: 24
## $ TEAM       <chr> "North Carolina", "Wisconsin", "Michigan", "Texas Tech", "G~
## $ CONF       <chr> "ACC", "B10", "B10", "B12", "WCC", "SEC", "B10", "ACC", "AC~
## $ G          <dbl> 40, 40, 40, 38, 39, 40, 38, 39, 38, 39, 40, 40, 40, 40, 36,~
## $ W          <dbl> 33, 36, 33, 31, 37, 29, 30, 35, 35, 33, 35, 36, 32, 35, 27,~
## $ ADJOE      <dbl> 123.3, 129.1, 114.4, 115.2, 117.8, 117.2, 121.5, 125.2, 123~
## $ ADJDE      <dbl> 94.9, 93.6, 90.4, 85.2, 86.3, 96.2, 93.7, 90.6, 89.9, 91.5,~
## $ BARTHAG    <dbl> 0.9531, 0.9758, 0.9375, 0.9696, 0.9728, 0.9062, 0.9522, 0.9~
## $ EFG_O      <dbl> 52.6, 54.8, 53.9, 53.5, 56.6, 49.9, 54.6, 56.6, 55.2, 51.7,~
## $ EFG_D      <dbl> 48.1, 47.7, 47.7, 43.0, 41.1, 46.0, 48.0, 46.5, 44.7, 48.1,~
## $ TOR        <dbl> 15.4, 12.4, 14.0, 17.7, 16.2, 18.1, 14.6, 16.3, 14.7, 16.2,~
## $ TORD       <dbl> 18.2, 15.8, 19.5, 22.8, 17.1, 16.1, 18.7, 18.6, 17.5, 18.6,~
## $ ORB        <dbl> 40.7, 32.1, 25.5, 27.4, 30.0, 42.0, 32.5, 35.8, 30.4, 41.3,~
## $ DRB        <dbl> 30.0, 23.7, 24.9, 28.7, 26.2, 29.7, 29.4, 30.2, 25.4, 25.0,~
## $ FTR        <dbl> 32.3, 36.2, 30.7, 32.9, 39.0, 51.8, 28.4, 39.8, 29.1, 34.3,~
## $ FTRD       <dbl> 30.4, 22.4, 30.0, 36.6, 26.9, 36.8, 22.7, 23.9, 26.3, 31.6,~
## $ `2P_O`     <dbl> 53.9, 54.8, 54.7, 52.8, 56.3, 50.0, 53.4, 55.9, 52.5, 51.0,~
## $ `2P_D`     <dbl> 44.6, 44.7, 46.8, 41.9, 40.0, 44.9, 47.6, 46.3, 45.7, 46.3,~
```

```
## $ `3P_O`      <dbl> 32.7, 36.5, 35.2, 36.5, 38.2, 33.2, 37.9, 38.7, 39.5, 35.5,~
## $ `3P_D`      <dbl> 36.2, 37.5, 33.2, 29.7, 29.0, 32.2, 32.6, 31.4, 28.9, 33.9,~
## $ ADJ_T       <dbl> 71.7, 59.3, 65.9, 67.5, 71.5, 65.9, 64.8, 66.4, 60.7, 72.8,~
## $ WAB         <dbl> 8.6, 11.3, 6.9, 7.0, 7.7, 3.9, 6.2, 10.7, 11.1, 8.4, 8.9, 1~
## $ POSTSEASON <chr> "2ND", "2ND", "2ND", "2ND", "2ND", "2ND", "2ND", "Champions~
## $ SEED        <chr> "1", "1", "3", "3", "1", "8", "4", "1", "1", "1", "2", "1",~
## $ YEAR        <dbl> 2016, 2015, 2018, 2019, 2017, 2014, 2013, 2015, 2019, 2017,~
```

What am I working with here?

- 4 character variables (TEAM, CONF, POSTSEASON, SEED)
- 20 numeric variables (G, W, ADJOE, ADJDE, BARTHAG, EFG_O, EFG_D, TOR, TORD, ORB, DRB, FTR, FTRD, X2P_O, X2P_D, X3P_O, X3P_D, ADJ_T, WAB, YEAR)
  - Very happy that there does not appear to be any NAs in the numeric stuff or in the character stuff!

Let's just jump right into it!

I'm going to select the columns I anticipate using. (Week 3)

```
# using the select function from Week 3
cbb <- cbb %>%
  select(TEAM, CONF, G, W, ORB, DRB, POSTSEASON, SEED, YEAR)
```

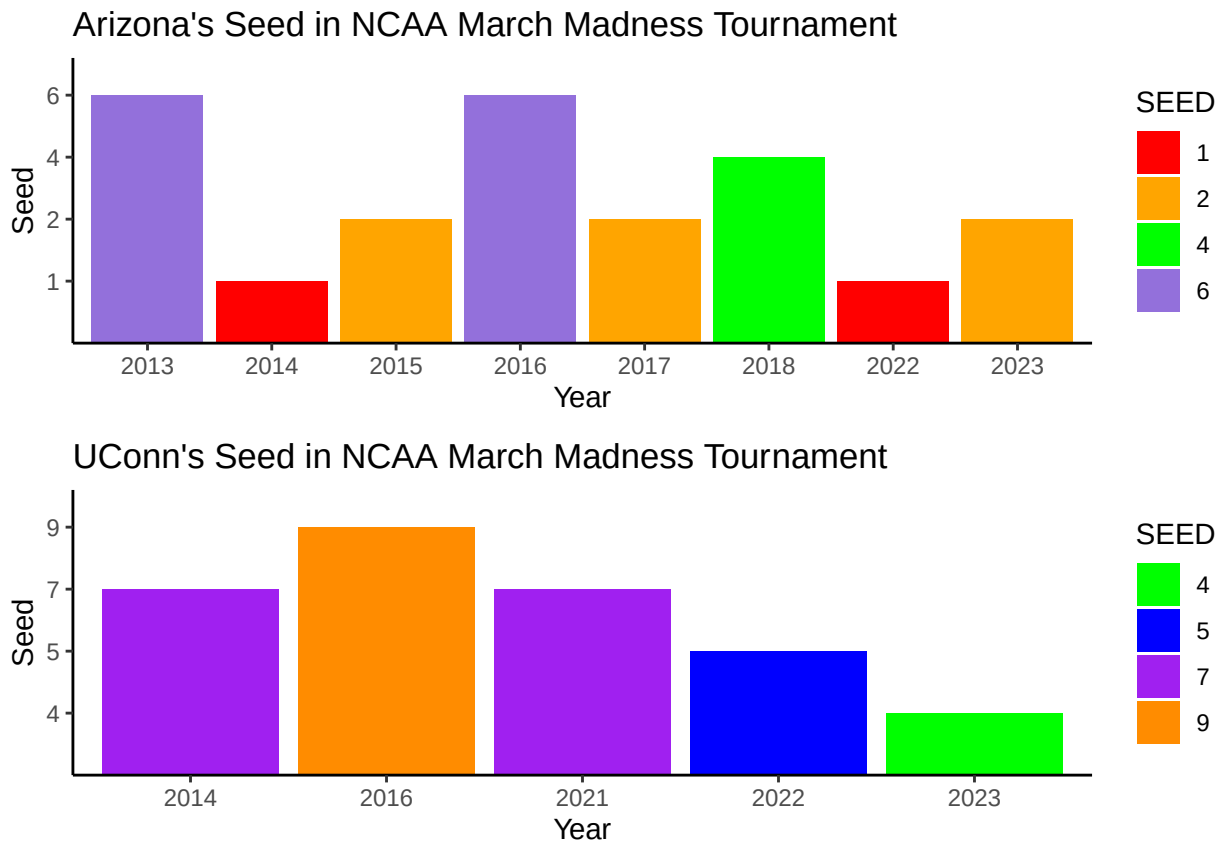
My life will be easier if I turn TEAM, SEED, YEAR, POSTSEASON, and CONF into factors and give SEED and POSTSEASON their own specific color palettes.

```
#making TEAM, SEED, YEAR, POSTSEASON, and CONF factors and giving SEED and POSTSEASON their own specifi
cbb$TEAM <- as.factor(cbb$TEAM)

cbb$SEED <- as.factor(cbb$SEED)
seed <- c("1" = "red", "2" = "orange", "3" = "yellow", "4"= "green", "5"= "blue", "6" = "mediumpurple",

cbb$YEAR <- as.factor(cbb$YEAR)


cbb$POSTSEASON <- factor(cbb$POSTSEASON, levels = c("R68", "R64", "R32", "S16", "E8", "F4", "2ND", "Cha
cbb$POSTSEASON <- as.factor(cbb$POSTSEASON)
postseason = c('R68'='darkorchid3','R64'='firebrick3','R32'='aquamarine3','S16'='slateblue3','E8' = 'gr



cbb$CONF <- as.factor(cbb$CONF)
```

Let's see the difference in seed for Arizona and UConn over the years...

```
# using the tidyverse pipe and the filter function from Week 3 to get rows where TEAM is Arizona, and w
arizona_seed <- cbb %>%
  filter(TEAM == "Arizona", !is.na(SEED)) %>%
  ggplot(aes(x=YEAR, y=SEED, fill=SEED))+
  scale_fill_manual(values=seed)+
```

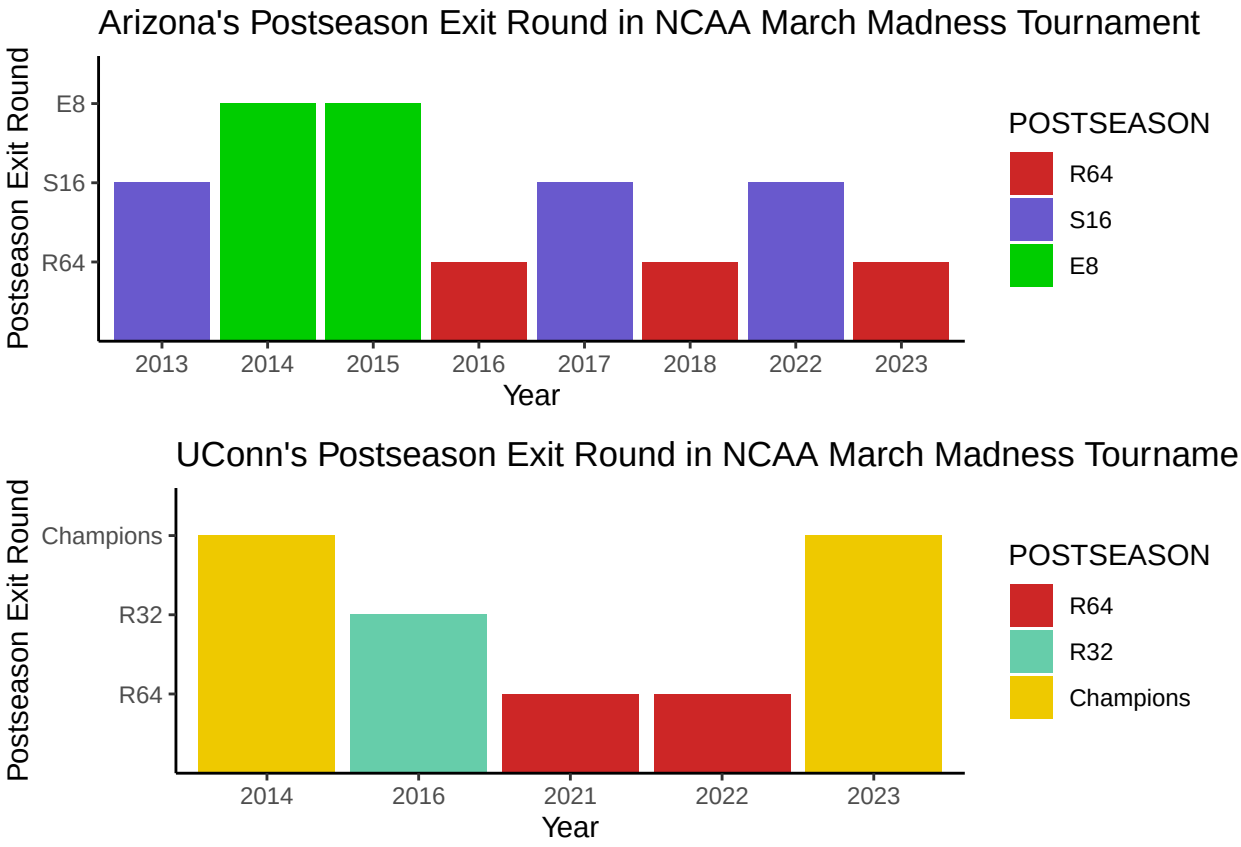
```
  geom_col()+
  labs(x="Year", y="Seed", title = "Arizona's Seed in NCAA March Madness Tournament")+
  theme_classic()

# using the tidyverse pipe and the filter function from Week 3 to get rows where TEAM is Connecticut, a
uconn_seed <- cbb %>%
  filter(TEAM == "Connecticut", !is.na(SEED)) %>%
  ggplot(aes(x=YEAR, y=SEED, fill=SEED))+
  scale_fill_manual(values=seed)+
  geom_col()+
  labs(x="Year", y="Seed", title = "UConn's Seed in NCAA March Madness Tournament")+
  theme_classic()

# using ggarrange to paste both previous plots into one and saving that to an object called seed_plot.
seed_plot <- ggarrange(arizona_seed, uconn_seed, ncol = 1, nrow = 2)

#calling seed_plot to see the plot
seed_plot
```



Let's see the difference in postseason exit round for Arizona and UConn...

```
# using the tidyverse pipe and the filter function from Week 3 to get rows where TEAM is Arizona, and w
arizona_postseason <- cbb %>%
  filter(TEAM == "Arizona", !is.na(POSTSEASON), !POSTSEASON == "N/A") %>%
  ggplot(aes(x=YEAR, y=POSTSEASON, fill=POSTSEASON))+
  scale_fill_manual(values = postseason)+
```

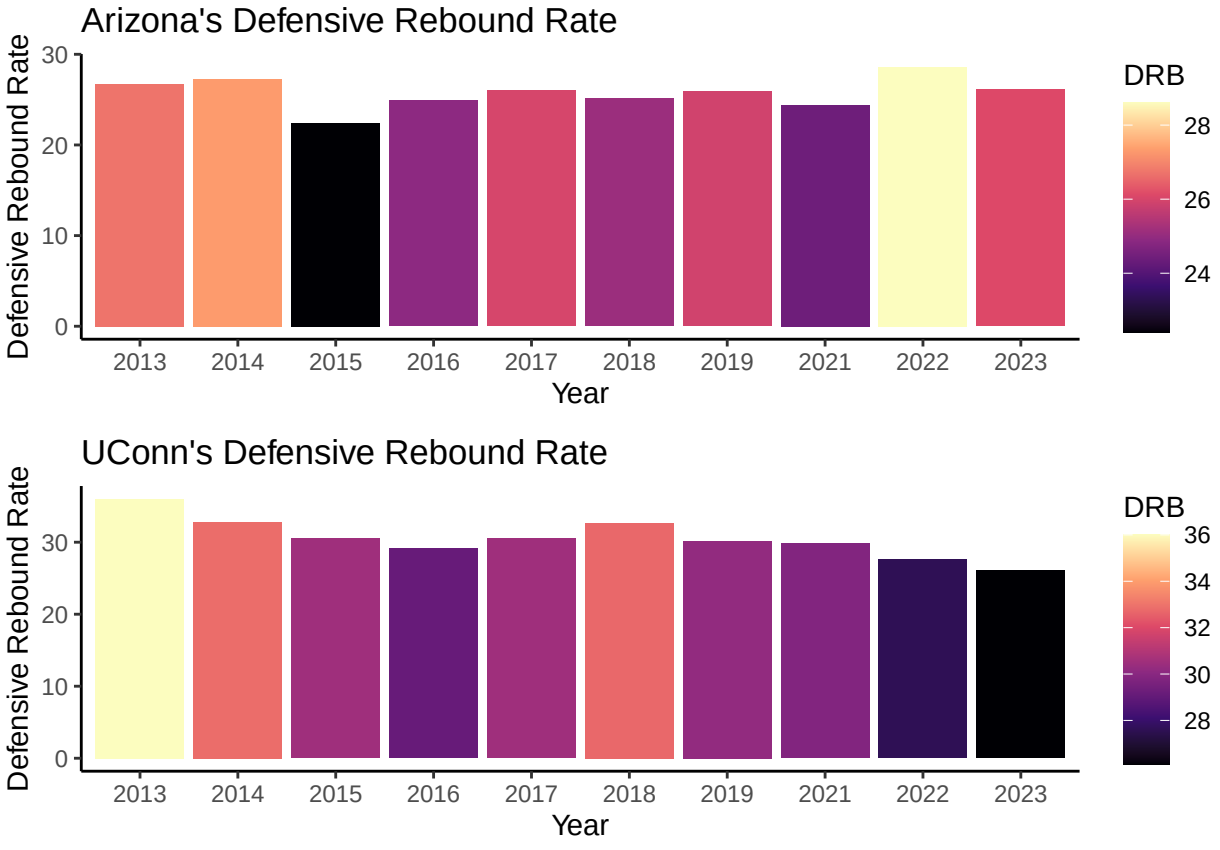
```
  geom_col()+
  labs(x="Year", y="Postseason Exit Round", title = "Arizona's Postseason Exit Round in NCAA March Madne
  theme_classic()

# using the tidyverse pipe and the filter function from Week 3 to get rows where TEAM is Connecticut, a
uconn_postseason <- cbb %>%
  filter(TEAM == "Connecticut", !is.na(POSTSEASON), !POSTSEASON == "N/A") %>%
  ggplot(aes(x=YEAR, y=POSTSEASON, fill = POSTSEASON))+
  scale_fill_manual(values = postseason)+
  geom_col()+
  labs(x="Year", y="Postseason Exit Round", title = "UConn's Postseason Exit Round in NCAA March Madnes
  theme_classic()

# using ggarrange to paste both previous plots into one and saving that to an object called postseason_
postseason_plot <- ggarrange(arizona_postseason, uconn_postseason, ncol=1, nrow=2)

# calling postseason_plot to see the plot
postseason_plot
```



Let's see the difference in defensive rebound rate for Arizona and UConn...

```
# using the tidyverse pipe and the filter function from Week 3 to get rows where TEAM is Arizona, and w
arizona_drb <- cbb %>%
  filter(TEAM == "Arizona", !is.na(DRB), !DRB == "N/A") %>%
  ggplot(aes(x=YEAR, y=DRB, fill=DRB))+
  geom_col()+
```

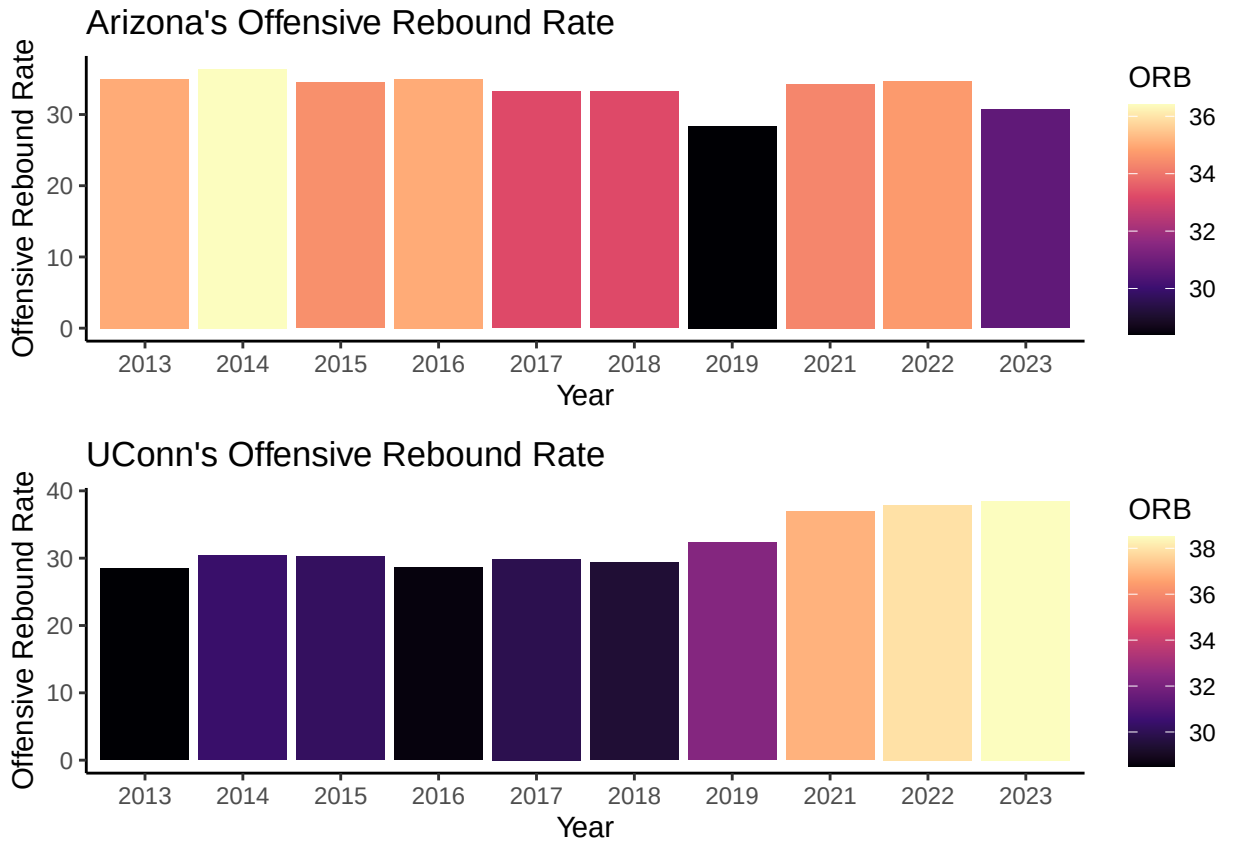
```
  scale_fill_viridis_c(option = "magma")+
  labs(x="Year", y="Defensive Rebound Rate", title = "Arizona's Defensive Rebound Rate")+
  theme_classic()

# using the tidyverse pipe and the filter function from Week 3 to get rows where TEAM is Connecticut, a
uconn_drb <- cbb %>%
  filter(TEAM == "Connecticut", !is.na(DRB), !DRB == "N/A") %>%
  ggplot(aes(x=YEAR, y=DRB, fill=DRB))+
  geom_col()+
  scale_fill_viridis_c(option = "magma")+
  labs(x="Year", y="Defensive Rebound Rate", title = "UConn's Defensive Rebound Rate")+
  theme_classic()

# using ggarrange to paste both previous plots into one and saving that to an object called drb_plot.
drb_plot <- ggarrange(arizona_drb, uconn_drb, ncol=1, nrow=2)

# calling drb_plot to see the plot
drb_plot
```



Let's see the difference in offensive rebound rate for Arizona and UConn...

```
# using the tidyverse pipe and the filter function from Week 3 to get rows where TEAM is Arizona, and w
arizona_orb <- cbb %>%
  filter(TEAM == "Arizona", !is.na(ORB), !ORB == "N/A") %>%
  ggplot(aes(x=YEAR, y=ORB, fill=ORB))+
  geom_col()+
```

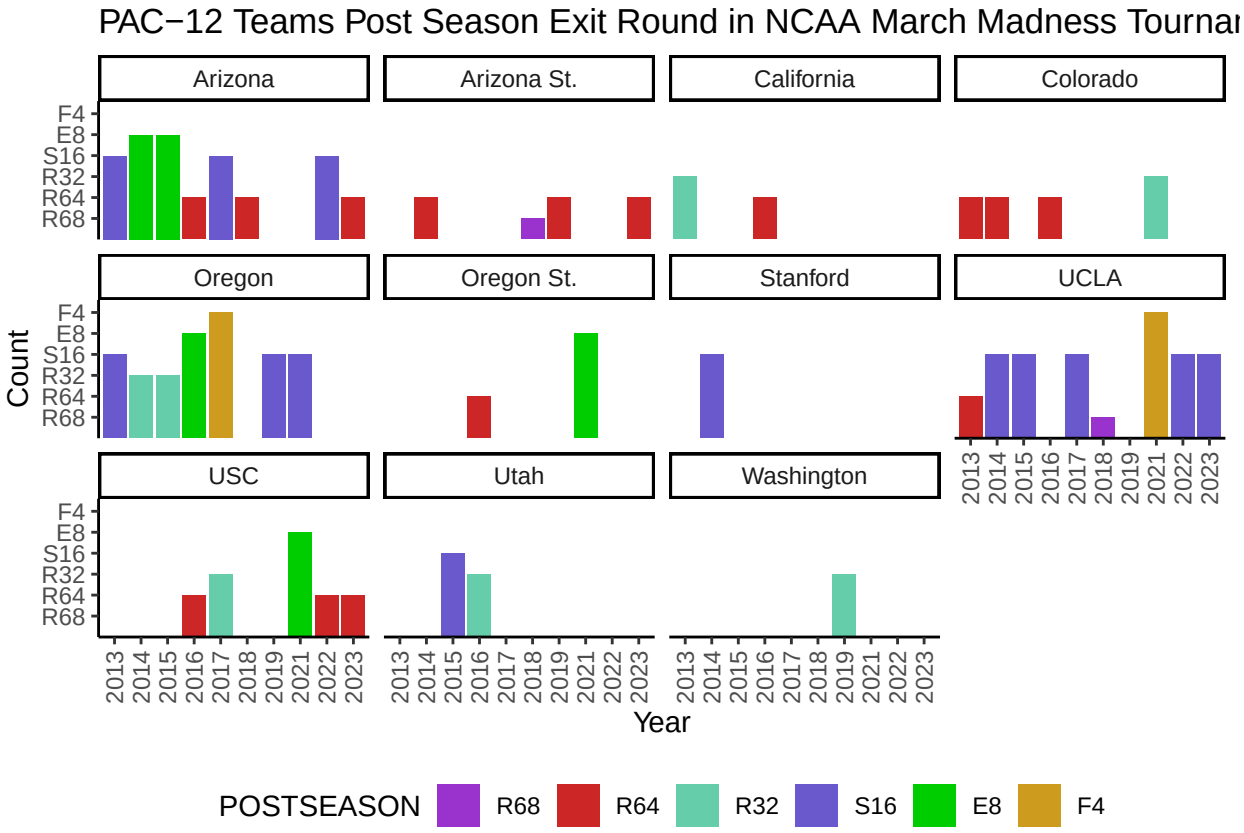
```
  scale_fill_viridis_c(option = "magma")+
  labs(x="Year", y="Offensive Rebound Rate", title = "Arizona's Offensive Rebound Rate")+
  theme_classic()

# using the tidyverse pipe and the filter function from Week 3 to get rows where TEAM is Connecticut, a
uconn_orb <- cbb %>%
  filter(TEAM == "Connecticut", !is.na(ORB), !ORB == "N/A") %>%
  ggplot(aes(x=YEAR, y=ORB, fill=ORB))+
  geom_col()+
  scale_fill_viridis_c(option = "magma")+
  labs(x="Year", y="Offensive Rebound Rate", title = "UConn's Offensive Rebound Rate")+
  theme_classic()

# using ggarrange to paste both previous plots into one and saving that to an object called orb_plot.
orb_plot <- ggarrange(arizona_orb, uconn_orb, ncol=1, nrow=2)

# calling orb_plot to see the plot
orb_plot
```



Let's check out when each school in the Pac-12 exited the NCAA March Madness Tournament if they made it in...

```
# using the tidyverse pipe and the filter function from Week 3 to get rows where CONF is P12, and where
pac12_postseason <- cbb %>%
  filter(CONF == "P12", !is.na(POSTSEASON), !POSTSEASON == "N/A") %>%
  ggplot(aes(x=YEAR, y=POSTSEASON, fill = POSTSEASON))+
```

```
  geom_col()+
  scale_fill_manual(values = postseason)+
  theme_classic()+
  labs(x="Year", y= "Count", title = "PAC-12 Teams Post Season Exit Round in NCAA March Madness Tournam
  theme(legend.position = "bottom", axis.text.x = element_text(angle = 90, vjust = 0.5, hjust=1))+
  guides(fill = guide_legend(nrow = 1))+
  facet_wrap(~TEAM)

# calling pac12_postseason to see the plot
pac12_postseason
```



I think it would be neat to see the percentage of games each team won in each year... Let's make a new column called W_PERC to show that.

```
# using the tidyverse pipe and the mutate function to divide the number of wins each team had in each y
cbb <- cbb %>%
  mutate(W_PERC = cbb$W/cbb$G)
```

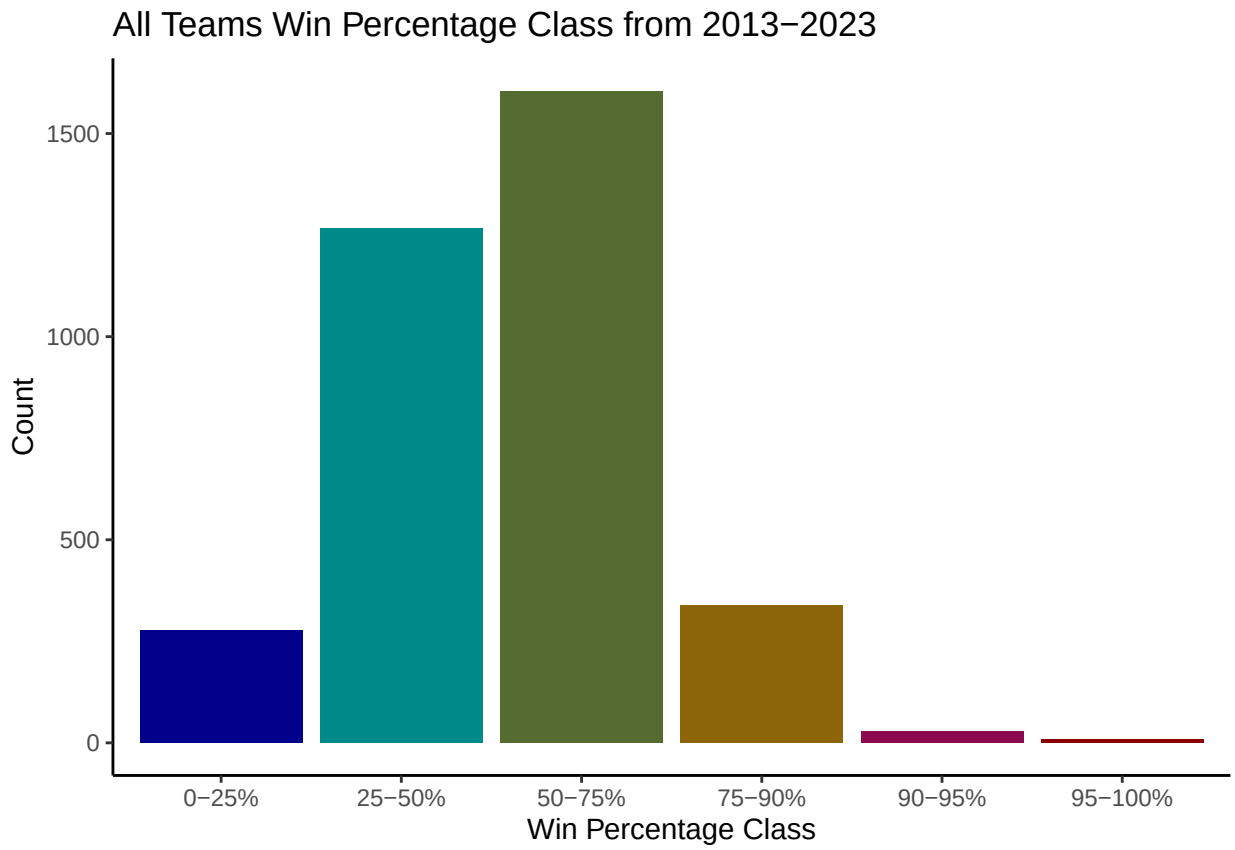
Alright now that that worked, let's make another new column called W_PERC_CLASS telling me if the win percentage is between 0 and 25% then 25 and 50%, then 50 and 75%, then 75 and 90%, then 90 and 95% and finally 95-100%.

```
# using the tidyverse pipe and the mutate and case_when functions to make 6 classes of win percentage b
cbb <- cbb %>%
  mutate(W_PERC_CLASS = case_when(W_PERC >= .95 ~ "95-100%",
```

```
                                    W_PERC >= .90 & W_PERC < .95 ~ "90-95%",
                                    W_PERC >= .75 & W_PERC < .90 ~ "75-90%",
                                    W_PERC >= .50 & W_PERC < .75 ~ "50-75%",
                                    W_PERC >= .25 & W_PERC < .50 ~ "25-50%",
                                    W_PERC >= .0 & W_PERC < .25 ~ "0-25%",
                                    TRUE ~ NA))
```

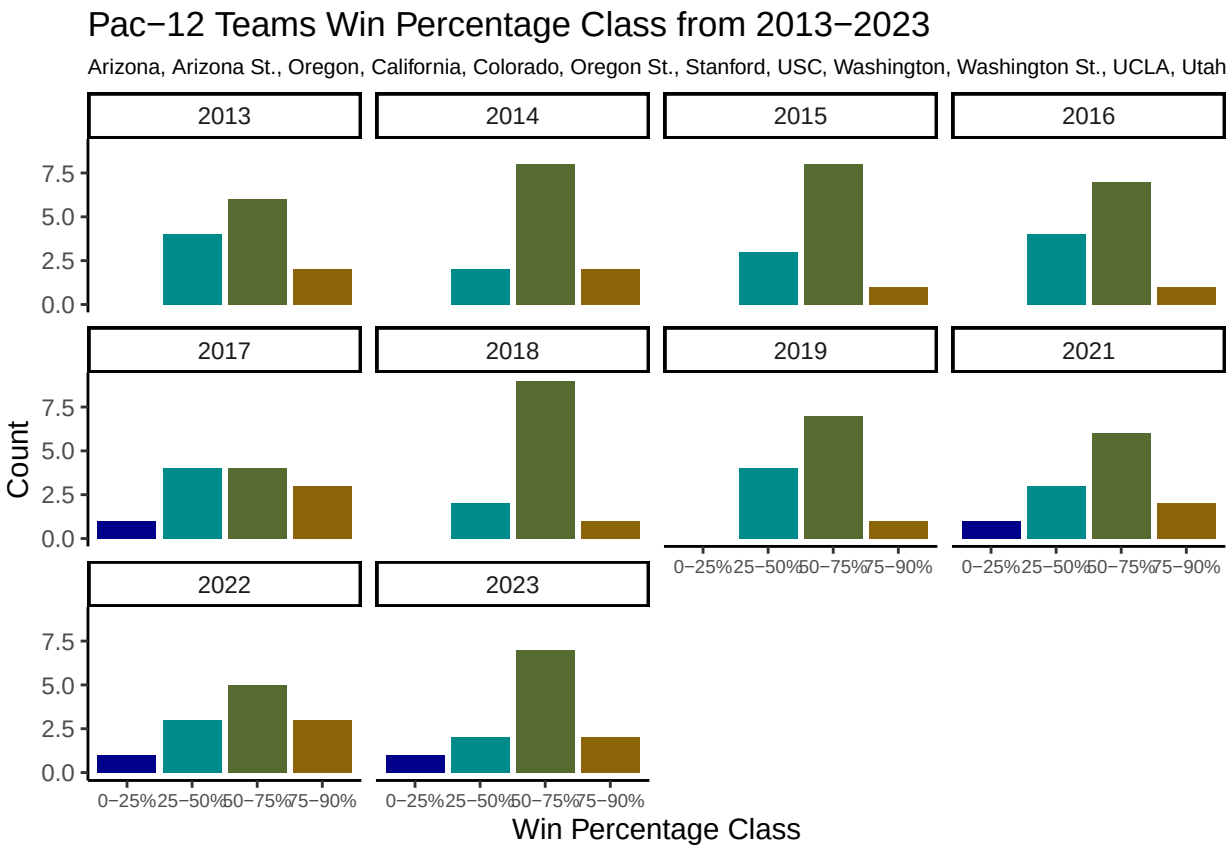
Let's make a plot showing the win percentage categories...

```
# making W_PERC_CLASS a factor and giving it its own specific color palette.
cbb$W_PERC_CLASS <- as.factor(cbb$W_PERC_CLASS)
w_perc_class = c('0-25%'='darkblue','25-50%'='darkcyan','50-75%'='darkolivegreen','75-90%'='darkgoldenr

# plotting W_PERC_CLASS for all teams and then saving it to an object called all_teams_w_perc_class.
all_teams_w_perc_class <- cbb %>%
  ggplot(aes(x=W_PERC_CLASS, fill = W_PERC_CLASS))+
  geom_bar()+
  scale_fill_manual(values = w_perc_class)+
  theme_classic()+
  labs(x="Win Percentage Class", y="Count", title = "All Teams Win Percentage Class from 2013-2023")+
  theme(legend.position = "none")

# calling all_teams_w_perc_class to see the plot
all_teams_w_perc_class
```



Let's make a plot showing the Pac-12 schools win percentage class...

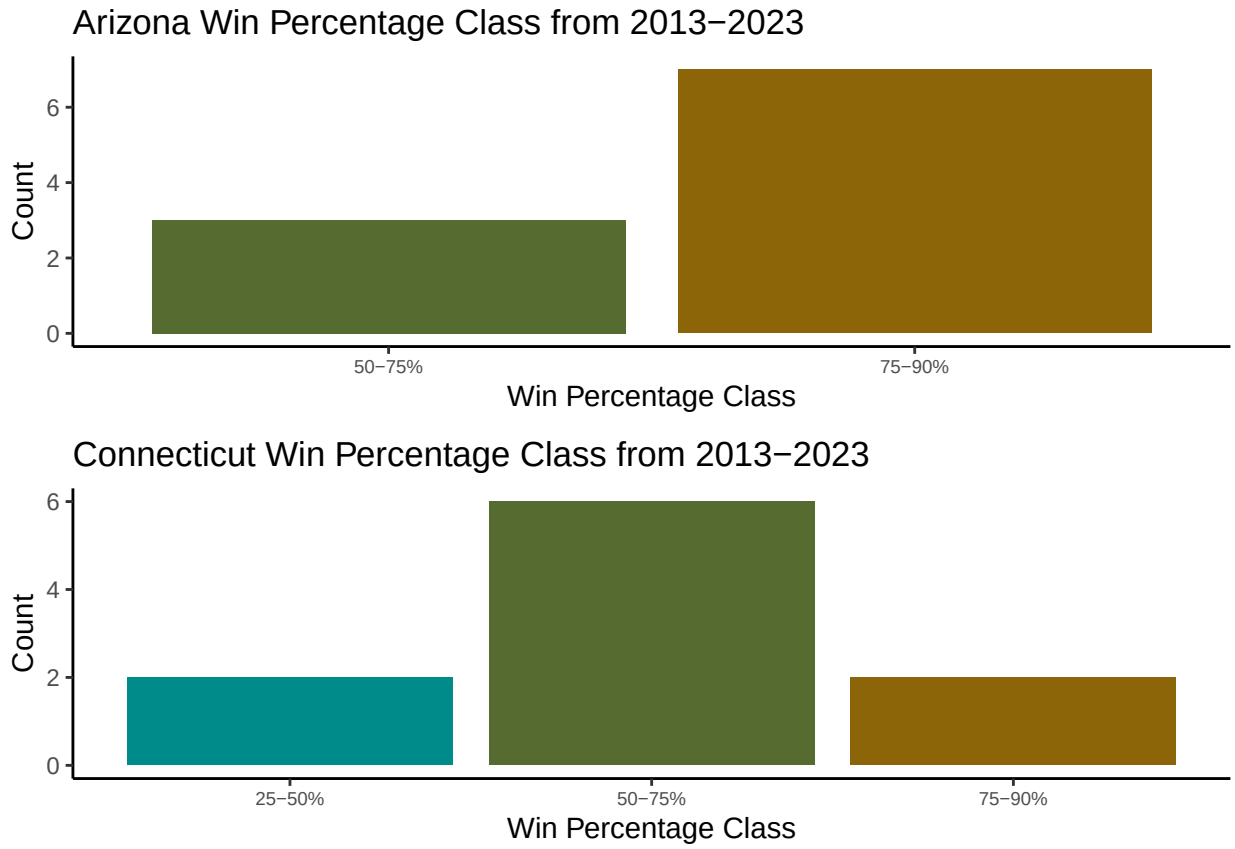
```r
# using the tidyverse pipe and the filter function from Week 3 to get rows where CONF is P12. Then plot
p12_w_perc_class <- cbb %>%
  filter(CONF == "P12") %>%
  ggplot(aes(x=W_PERC_CLASS, fill = W_PERC_CLASS))+
  geom_bar()+
  scale_fill_manual(values = w_perc_class)+
  theme_classic()+
  labs(x="Win Percentage Class", y="Count", title = "Pac-12 Teams Win Percentage Class from 2013-2023",
  theme(legend.position = "none", plot.subtitle=element_text(size=8), axis.text.x = element_text(size=7
  facet_wrap(~YEAR)

# calling p12_w_perc_class to see the plot.
p12_w_perc_class
```



Let's see the difference between win percentage class for Arizona and UConn...

```r
# using the tidyverse pipe and the filter function from Week 3 to get rows where TEAM is Arizona. Then
arizona_w_perc_class <-
  cbb %>%
  filter(TEAM == "Arizona") %>%
  ggplot(aes(x=W_PERC_CLASS, fill = W_PERC_CLASS))+
  geom_bar()+
  scale_fill_manual(values = w_perc_class)+
  theme_classic()+
  labs(x="Win Percentage Class", y="Count", title = "Arizona Win Percentage Class from 2013-2023")+
  theme(legend.position = "none", plot.subtitle=element_text(size=8), axis.text.x = element_text(size=7
```

```r
# using the tidyverse pipe and the filter function from Week 3 to get rows where TEAM is Connecticut Th
uconn_w_perc_class <-
  cbb %>%
  filter(TEAM == "Connecticut") %>%
  ggplot(aes(x=W_PERC_CLASS, fill = W_PERC_CLASS))+
  geom_bar()+
  scale_fill_manual(values = w_perc_class)+
  theme_classic()+
  labs(x="Win Percentage Class", y="Count", title = "Connecticut Win Percentage Class from 2013-2023")+
  theme(legend.position = "none", plot.subtitle=element_text(size=8), axis.text.x = element_text(size=7

# using ggarrange to paste both previous plots into one and saving that to an object called uconn_az_wi
uconn_az_win_perc_class <- ggarrange(arizona_w_perc_class, uconn_w_perc_class, nrow=2, ncol =1)

# calling uconn_az_win_perc_class to see the plot
uconn_az_win_perc_class
```



Overarching Conclusions from Plots

The plots I made comparing the statistics of Arizona and UConn were interesting because it showed me that there isn't a lot of carry over of some statistic performance year to year. College basketball has a lot of turnover because each player only gets 5 years to play up to 4 seasons, meaning that each year the team can be drastically different and perform differently in each category. These conclusions constitute accomplishing a larger task as we learned how to do in Week 8.