Exploratory Data Analysis (EDA) Paper

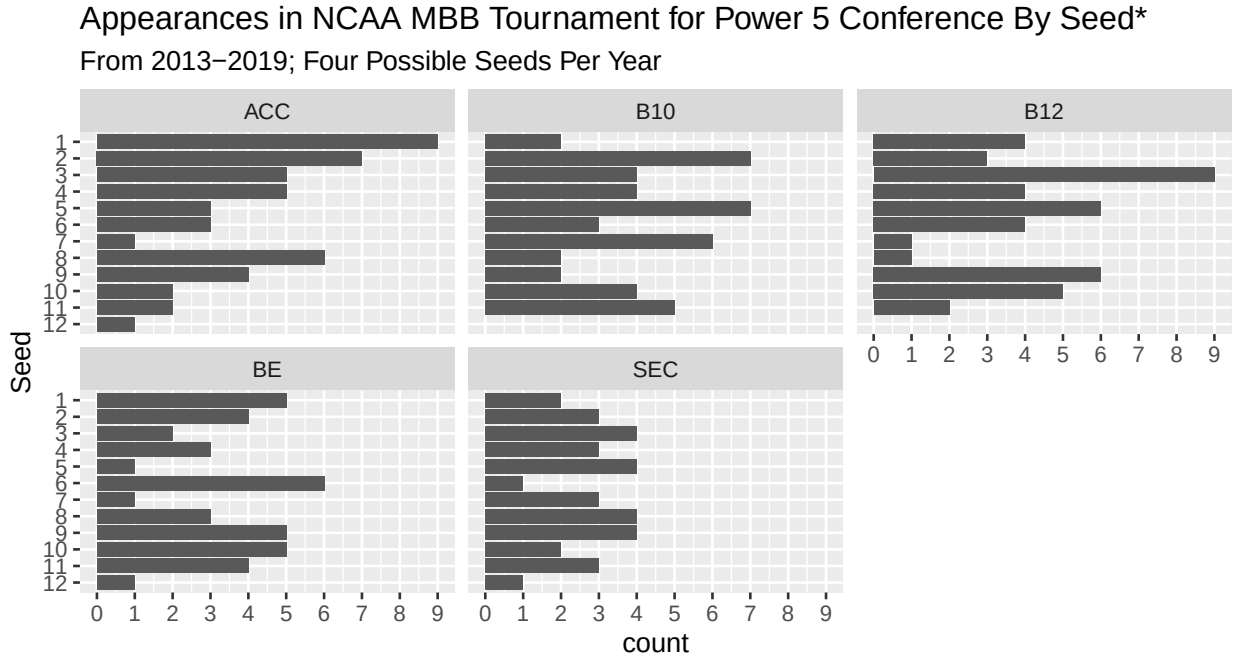
Carter Hall, Luis Medrano, Paul Charles, Correy Turral

June 14, 2021

Visualize Categorical-Categorical Relation

Research Question: Power 5 Conferences are recognized in MBB for their historic prowess; do they appear in the tournament frequently? If so, what seeds do they typically assume?
Graph:

```
powerFive <- c("ACC", "SEC", "B10", "B12", "BE")
pfc <- cbb %>% filter(CONF %in% powerFive & !is.na(SEED))
pfc$SEED <- as.factor(pfc$SEED)
ggplot(data = pfc) + geom_histogram(mapping = aes(y = reorder(SEED, desc(SEED))), stat = "count") +
  facet_wrap(CONF~.) +
  scale_x_continuous(breaks = c(0,1,2,3,4,5,6,7,8,9,10)) +
  labs(title = "Appearances in NCAA MBB Tournament for Power 5 Conference By Seed*",
       subtitle = "From 2013-2019; Four Possible Seeds Per Year",
       caption = "* - 1-seeds are generally representative of teams more favorited to win the tournamen
       y = "Seed")
```



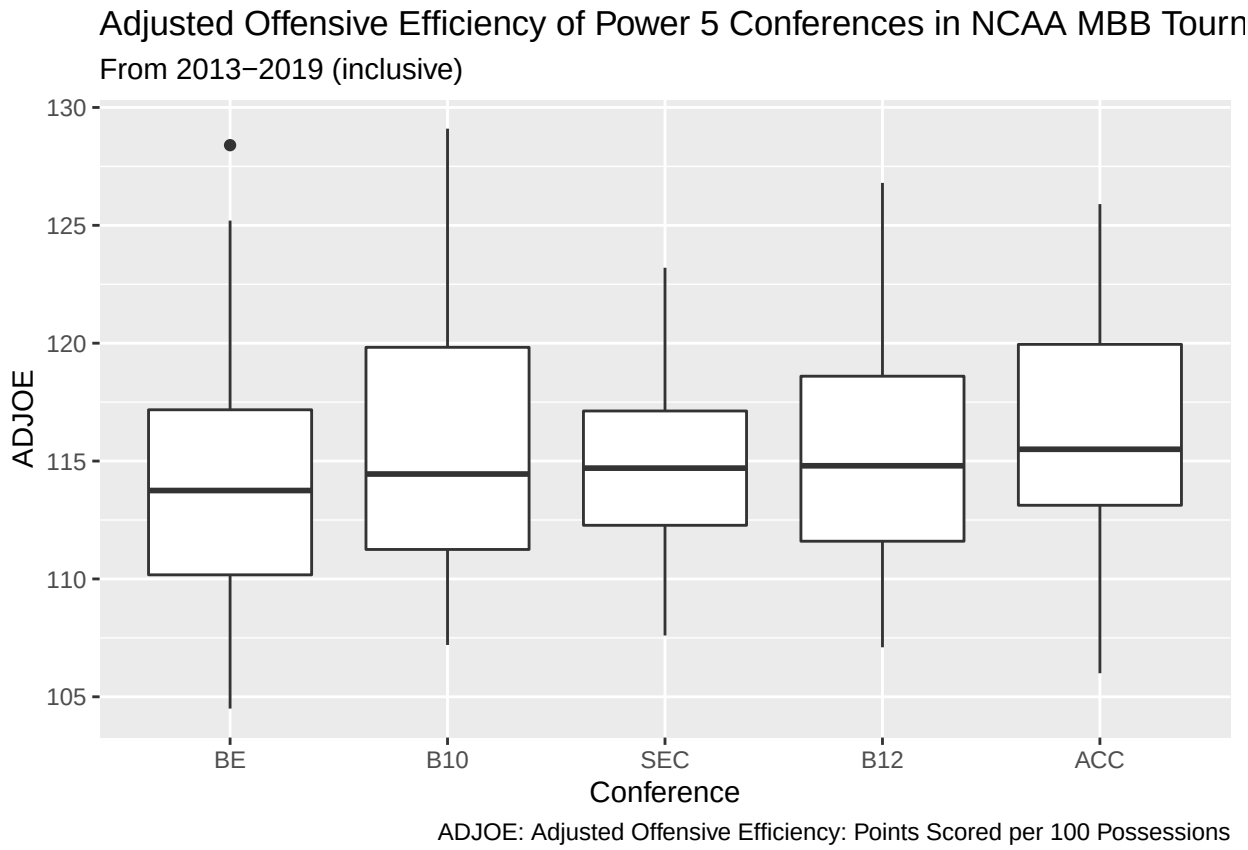
Description (1-sentence): The Power-5 Conferences have appeared frequently throughout this time period, assuming many of the higher seeds the tournament offered.

Visualize Categorical-Continuous Relation

Research Question: From 2013-2019, which of the Power 5 Conference(s) were the most efficient on offense in the NCAA Tournament?

Graph:

```
ggplot(data = pfc, mapping = aes(x = reorder(CONF, ADJOE, FUN = median), y = ADJOE)) +
  geom_boxplot() +
  labs(title = "Adjusted Offensive Efficiency of Power 5 Conferences in NCAA MBB Tournament",
       subtitle = "From 2013-2019 (inclusive)",
       x = "Conference",
       caption = "ADJOE: Adjusted Offensive Efficiency: Points Scored per 100 Possessions")
```



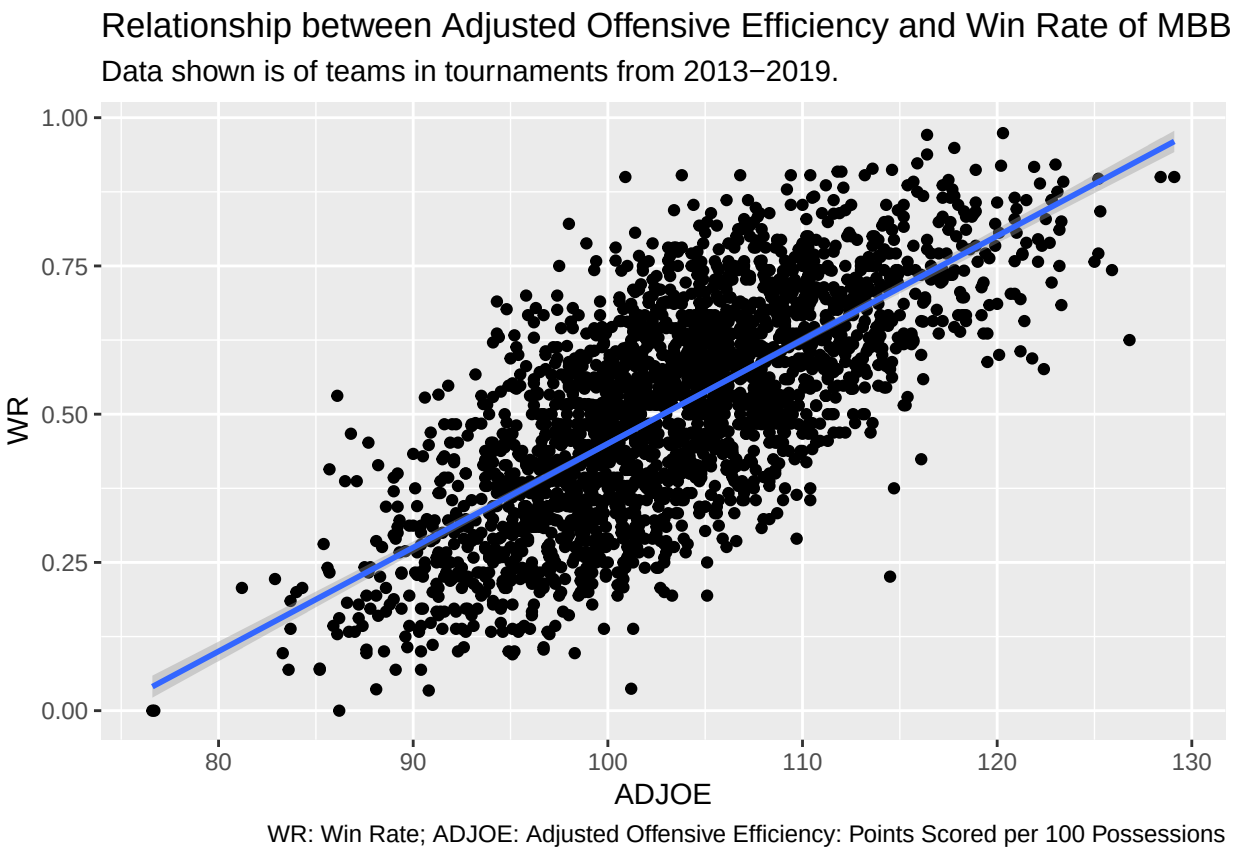
Description (1-sentence): Based on the result of the boxplot above, we can determine that the ACC was the most efficient conference offensively from 2013 to 2019.

Visualize Continuous-Continuous Relation

Research Question: From 2013-2019, is there any relation between win rate and offensive efficiency from all teams in the NCAA Tournament?

Graph:

```
cbb <- cbb %>% mutate(WR = round(W / G, 3))
pfc <- cbb %>% filter(CONF %in% powerFive & !is.na(SEED))
pfc$SEED <- as.factor(pfc$SEED)
ggplot(data = cbb, mapping = aes(x = ADJOE, y = WR)) + geom_point() +
  geom_smooth(method = lm) +
  labs(title = "Relationship between Adjusted Offensive Efficiency and Win Rate of MBB Teams in NCAA Tou
       subtitle = "Data shown is of teams in tournaments from 2013-2019.",
       caption = "WR: Win Rate; ADJOE: Adjusted Offensive Efficiency: Points Scored per 100 Possessions
```



Description (1-sentence): There seems to be a positive strong correlation between offensive efficiency and win rate, with the points of highest win rate showing up at a high offensive efficiency and those with close to zero or zero win rate at a low offensive efficiency.