Analysis on whether the red zone is an arbitrary field marker

are penalties drive killers?

Intro

Historically, the red zone is at the 20 yard marker, but are teams dramatically more likely to score there? Do defenses tighten up in the red zone? How does playcalling differ?

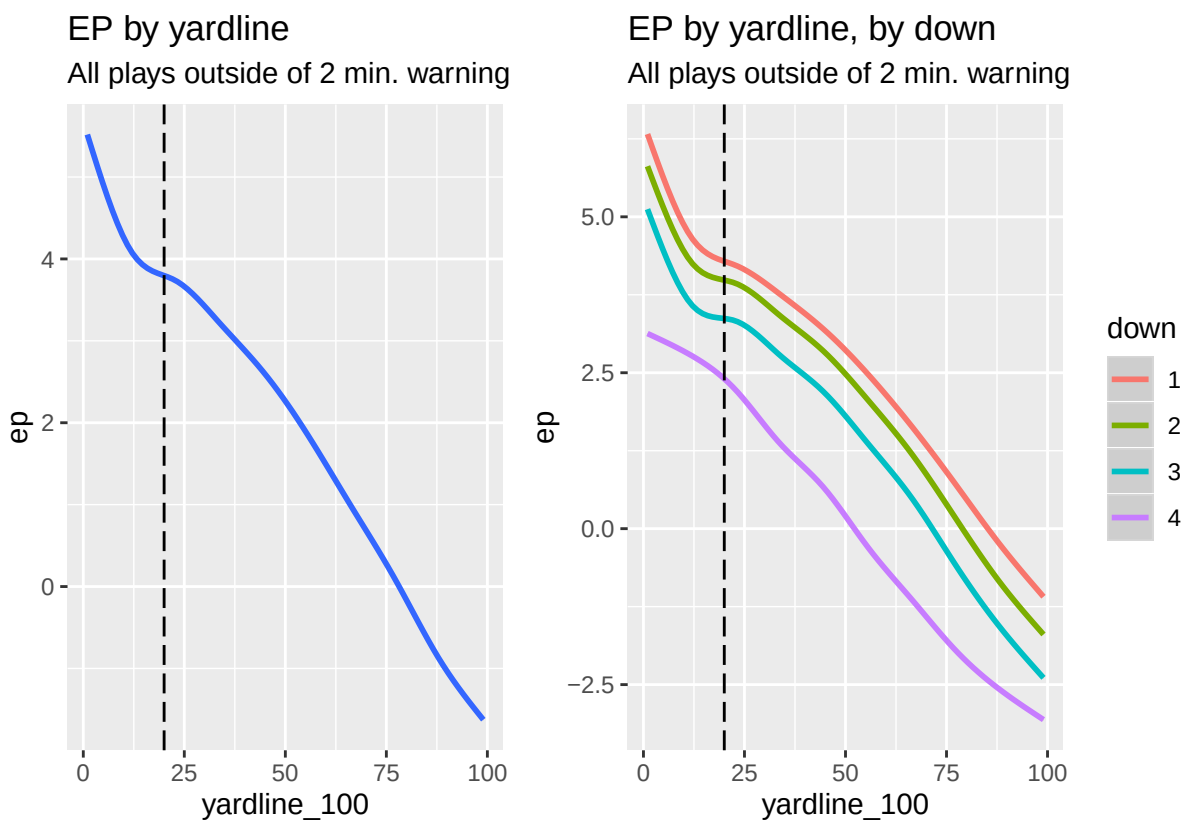
Note: Data sourced from Lee Sharpe's nflfastR repository: https://github.com/leesharpe/nfldata/blob/master/UPDATING-NFLFASTR.md

```
# The following transformations select the relevant columns for analysis as well
filtered_pbp <-
  pbp %>%
    filter(!is.na(ep)) %>%
    filter(yardline_100 < 100) %>%
    filter(!is.na(down)) %>%
    mutate(down = as.factor(down)) %>%
    select(season, game_id, play_id, down, time, half_seconds_remaining, play_type, pass,
           rush, ep, success, epa, series_success, yardline_100)

pbp_standard <-
    filtered_pbp %>%
        filter(half_seconds_remaining > 120)

pbp_under2 <-
    filtered_pbp %>%
        filter(half_seconds_remaining <= 120)
```

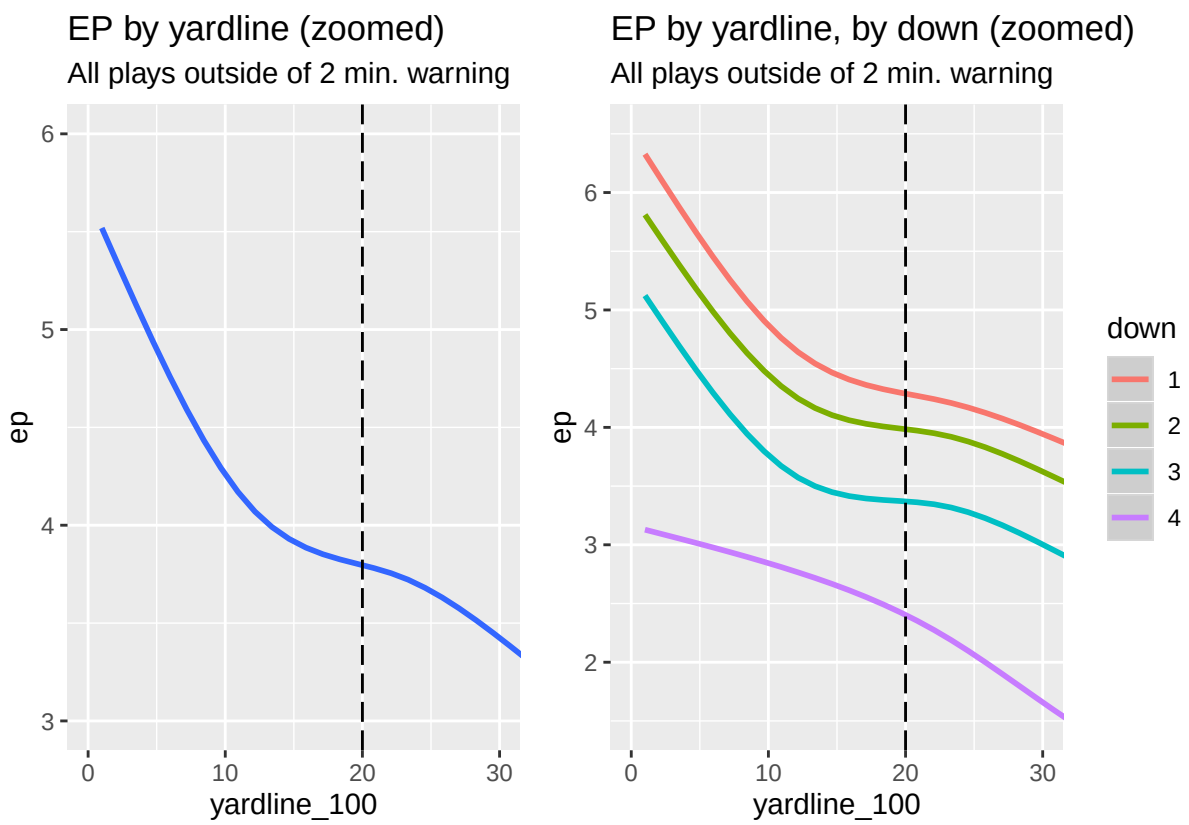
Question 1: Are teams more likely to score within the red zone compared to other places on the field?

The following charts display expected points (https://www.advancedfootballanalytics.com/index.php/home/stats/stats-explained/glossary) from @nflscrapR by field position. The first charts show the full data, while the second show a zoomed in version between the 0 and 30 yard line.

```
## `geom_smooth()` using method = 'gam' and formula 'y ~ s(x, bs = "cs")'
## `geom_smooth()` using method = 'gam' and formula 'y ~ s(x, bs = "cs")'
```




```
## `geom_smooth()` using method = 'gam' and formula 'y ~ s(x, bs = "cs")'
## `geom_smooth()` using method = 'gam' and formula 'y ~ s(x, bs = "cs")'
```

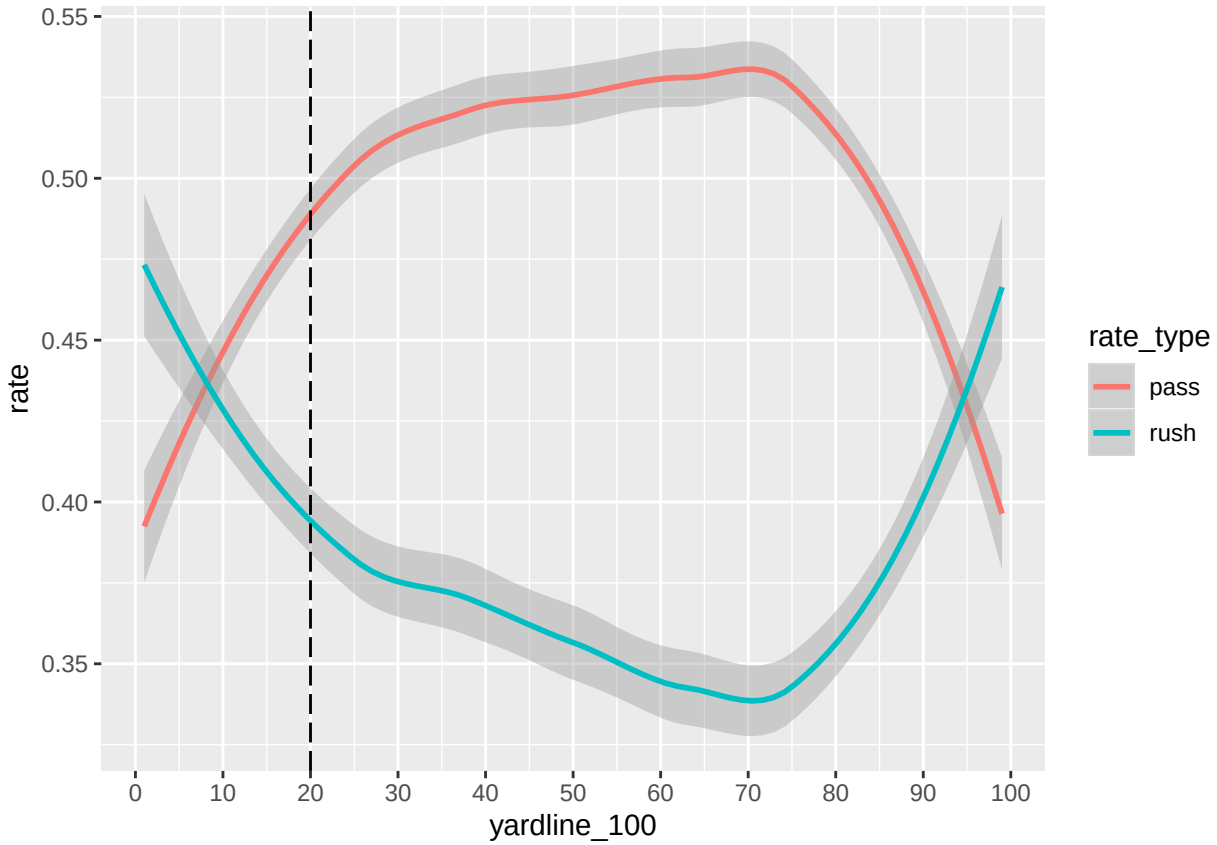


Conclusion: Based on the graph, between the 100 yard line and the 25 yard line, expected points increases fairly linearly. Between 15 and 25, it flattens a bit and rises again past the ~15 yard line.

Question 2: Does playcalling change when teams enter the redzone?

```
# The following transformations find the pass rate and run rate for each spot on the field
pass_rates <- pbp_standard %>%
  group_by(yardline_100) %>%
  summarise(rate = mean(pass, na.rm = TRUE), rate_type = "pass")
rush_rates <- pbp_standard %>%
  group_by(yardline_100) %>%
  summarise(rate = mean(rush, na.rm = TRUE), rate_type = "rush")
rates <- rbind(pass_rates, rush_rates)
rates$rate_type = as.factor(rates$rate_type)
```

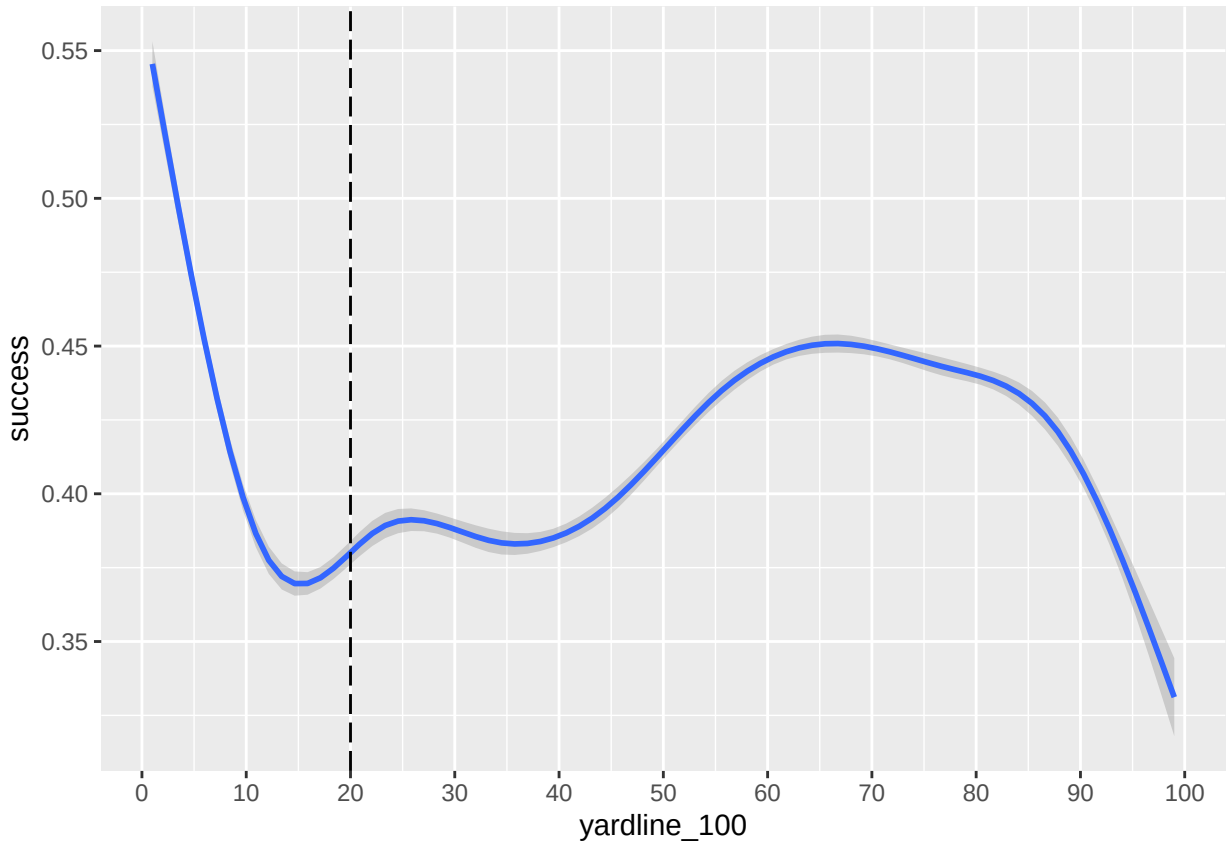
```
## `geom_smooth()` using method = 'loess' and formula 'y ~ x'
```



Conclusion: Rushing is favored closer to the goal line while passing is favored on other parts of the field. No significant changes occur near the 20 yard line.

Question 3: Do defenses clamp down in the red zone?

```
## `geom_smooth()` using method = 'gam' and formula 'y ~ s(x, bs = "cs")'
```



Success is defined by plays that have a positive EPA.

Conclusion: The graph shows that in the middle of the field, success rate gradually drops until around the 15 yardline, at which point is skyrockets up. This may indicate that once offenses reach the 15 yard line, they become significantly more likely to score.