

# Draft Analysis

*Matthew Reyers*

*May 3, 2019*

The 2019 NFL Draft just finished in Nashville, Tennessee. From April 25th-27th, General managers and other staffers of the 32 NFL teams came together and made dreams come true for many incredible athletes. Hopefully, through it all, these managers also managed to make the dreams come true for their faithful fans by landing great young talent.

Evaluating the draft class shortly after it is over has been a long standing tradition. Many major sports channels go through the class, team by team, and assign grades to the performers. Looking back a few years on some of these previous reports, I found that some of them don't age as well as some might hope. Below is an excerpt from an A+ graded candidate taken in the first round.

Richardson has been my perfect pick for the Browns for some time now. He's a one-man-army, able to break tackles when offensive line protection breaks down, catch passes, put up significant yards, pass protect and—most importantly—score touchdowns.

This candidate was of course Trent Richardson, taken with the third pick of the 2012 NFL entry draft. He looked every bit the part of an NFL running back and was reasonably believed to be a possible top 6 pick. Now, unfortunately, this pick makes more fame as a bust than a home run.

It would be ideal to be able to predict this type of bust prior to picking in the draft, though it may be a fool's errand to try to predict NFL production from college and combine data. So much of a player's success can be associated with team, system, coaching, professionalism, and others that aren't as easily captured in the data accessible to me.

Instead, I wanted to dive deeper into how teams were picking players. Were teams picking players to fill immediate needs or were they taking a top talent that had fallen into their laps at their pick? This writeup will explore some of the different drafting behaviours expressed by teams with a hopefully interesting discussion.

In the spirit of reproducibility, I am doing this through RMarkdown and will be posting my data sets in the corresponding git repository. Hopefully this also allows others to explore a rather fun dataset without the pain of connecting and cleaning their data first. This will be my first major post so some stylistic choices may be weird. I am always happy to learn so feel free to let me know if there may be a better format for presenting my work.

## Preprocessing

The first thing to do is to load the libraries and the data. The webscraping code for connecting these data sources should be on the git repository. Thanks to Dr. Jack Davis of SFU for some baseline code to get me going on the scraping!

```
library(tidyverse)

## -- Attaching packages ----- tidyverse 1.2.1 --
## v ggplot2 3.0.0      v purrr   0.2.5
## v tibble  2.1.1      v dplyr   0.8.0.1
## v tidyr   0.8.1      v stringr 1.3.1
## v readr   1.1.1      vforcats 0.3.0

## -- Conflicts ----- tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```

library(rvest)

## Loading required package: xml2
##
## Attaching package: 'rvest'
## The following object is masked from 'package:purrr':
## 
##     pluck
## The following object is masked from 'package:readr':
## 
##     guess_encoding
library(XML)

##
## Attaching package: 'XML'
## The following object is masked from 'package:rvest':
## 
##     xml
library(ggimage)
library(RCurl)

## Loading required package: bitops
##
## Attaching package: 'RCurl'
## The following object is masked from 'package:tidyr':
## 
##     complete
# Dataset for draft picks
draft_res <- read_csv("draft_results_updated.csv") %>%
  mutate(Team = case_when(Team %in% "redskins" ~ "Redskins",
                         TRUE ~ Team)) # Small typo in scraped doc

## Parsed with column specification:
## cols(
##   X1 = col_integer(),
##   Round = col_integer(),
##   Pick = col_integer(),
##   Player = col_integer(),
##   Name = col_character(),
##   Team = col_character(),
##   Position = col_character(),
##   College = col_character(),
##   Year = col_integer()
## )

# Value for each draft pick
draft_valuation <- read_csv("draft_valuation.csv") %>% select(-X1)

## Parsed with column specification:
## cols(
##   X1 = col_integer(),

```

```

##   Pick = col_integer(),
##   Value = col_double()
## )

# Data for salaries and players, based on a player's salary cap hit
salary_and_stats <- read_csv("players_and_salaries.csv")

## Parsed with column specification:
## cols(
##   .default = col_integer(),
##   Player = col_character(),
##   Pay = col_character(),
##   Position = col_character(),
##   ShortName = col_character(),
##   abbr = col_character(),
##   playerID = col_character(),
##   sacks = col_double(),
##   X1_1 = col_character()
## )
## See spec(...) for full column specifications.

```

The draft valuation and draft results data sets are simply scraped. The salary and stats dataset is scraped salary data combined with nflscrapR player stats on a yearly basis. Details of this process are found in ‘salary\_stuff.R’. I do not currently make use of the salary component of the data though the intent was to also describe team need in terms of cap flexibility. This will take more time to address and I may get back to it in the future.

## Identifying Team Weaknesses

One of the biggest challenges of this work was to define what team need meant. Did a team need Guard help specifically or Offensive line help? I simplified this distinction by grouping players into their respective units. Given that the historical data is limited, this helps by making group sizes larger and limits the influence a single player can have on the analysis.

```

oline <- c("Guard", "Center", "Left Tackle", "Right Tackle", "Tackle")
dline <- c("Defensive Tackle", "Defensive End")
lbs <- c("Inside Linebacker", "Outside Linebacker", "Linebacker")
dBacks <- c("Safety", "Cornerback", "Free Safety", "Strong Safety")
rbs <- c("Running Back", "Fullback")
receivers <- c("Wide Receiver")
tightends <- c("Tight End")
qbs <- c("Quarterback")
spec_teams <- c("Kicker", "Punter", "Long Snapper")

salary_and_stats <- salary_and_stats %>%
  mutate(unit = case_when(Position %in% oline ~ "Offensive Line",
                         Position %in% dline ~ "Defensive Line",
                         Position %in% lbs ~ "Linebacker",
                         Position %in% dBacks ~ "Defensive Back",
                         Position %in% rbs ~ "Running Back",
                         Position %in% receivers ~ "Receiver",
                         Position %in% tightends ~ "Tight End",
                         Position %in% qbs ~ "Quarterback",
                         Position %in% spec_teams ~ "Special Teams",

```

```

    TRUE ~ "Position Missing"),
Pay = as.numeric(str_remove_all(substr(Pay, 2, 100), ","))

```

After grouping, I summarized the nflscrapR data by the units. I have already summarized at the player level as it originally comes in as boxscores. Summarizing at the unit level for a year is then just taking the column sums of all the counting stats for the players that make up the given unit for the given team.

```

stats_by_unit <- salary_and_stats %>% select(-ShortName, -Position, -Rank, -X1, -Player, -playerID, -X1,
  mutate(unit_stats = map(data, ~colSums(., na.rm = T))) %>% select(-data)
stat_names <- names(stats_by_unit$unit_stats[[1]])

stats_by_unit <- stats_by_unit %>% unnest() %>% mutate(stat = rep(stat_names, dim(stats_by_unit)[1]))
  spread(key = stat, value = unit_stats)

```

The next step is then to quantify the strength of each unit. This exercise is not too difficult for the skill position players of the NFL as they have many counting statistics that are reasonably well understood. The linemen and the defense tend to be harder to quantify. Offensive linemen especially get few counting statistics and often are not even assigned a playerID in the nflscrapR data for lack of showing up come game end. Further, I was unable to find advanced metrics like Wins Above Replacement for some of these positions. If this data is readily available, I would like to expand this work to include WAR. Finally, I had no certain way of measuring defensive contributions. How much is a tackle worth relative to an interception relative to a completion for short yardage? I imagine this could be pursued with an Expected Points approach but I am left wondering how to spread the points across the members that contributed. Did the Safety make a great play or did the Edge rusher put pressure on the Quarterback to make a bad throw? To compensate for these uncertainties, I had to make some assumptions. Be warned that the subjective part is coming and you may disagree with my statements.

A convenient manner to connect all of these counting statistics together was to use fantasy points scored by unit, with some assumptions for the units that are not generally involved with scoring points. This is convenient for two reasons: I now have a way to weight each statistic that at least some people like and pro football reference has a dataset of fantasy points allowed by defenses to given skill position players of the opposing offence. This second point greatly simplifies the coding necessary as I can scrape the data rather than generate it from the original boxscores.

For the fantasy geeks out there, I used standard scoring rules. I imagine that 0.5 or 1 PPR approaches may change the results ever so slightly towards teams that use their running backs as safety valves more frequently but that this difference should not be extraordinary.

First, the easy players: the offensive skill position players and QB.

```

off_stats_fant <- stats_by_unit %>% filter(unit %in% c("Running Back", "Receiver", "Tight End", "Quarterback",
  mutate(fant_pt = 4 * pass.tds + 0.04*passyds - 1 * pass.ints + 2 * pass.twoptm +
    0.1 * recyds + 6 * rec.tds + 0.1*rushyds + 6 * rushtds - 1 * fumbloslost) %>%
group_by(Year, unit) %>%
arrange(desc(fant_pt)) %>%
mutate(rank = row_number(),
good_pos = case_when(rank < 16 ~ "Yes",
                      TRUE ~ "No")) %>%
ungroup()

```

Note that the subjective barrier I have made for a team need is the middle of the league. Football is a sport where a weakness in any unit can be used to expose a team. Seeing that the ultimate goal of a season is to win a Super Bowl, typically a team needs to be better than most in nearly all of the units I have described. Although coming in as rank 15 in a category is not stellar, I would suggest it is not a place of urgent weakness for a team. Relative to the rest of the team's roster, this may seem like a glaring fault but in comparison to the rest of the league the team may very well be in good shape.

The final piece of the offense to evaluate is the Offensive Linemen. Since these players have no counting stats in most instances, I had to gauge their performance based on the skill position players most readily impacted by them. This means that I define Offensive Line performance in terms of Running Back and Quarterback performance. Again, not a perfect measure. Some RBs and QBs make their Offensive Line look stellar while others have their play significantly enhanced by the Line in front of them. Perhaps something like Pro Football Focus grades/rankings would be useful in the future, though their rankings also have a layer of subjectivity to them.

```

o_linemen <- off_stats_fant %>% filter(unit %in% c("Running Back", "Quarterback")) %>%
  mutate(good_pos = case_when(rank < 16 ~ "Yes",
                               TRUE ~ "No")) %>%
  group_by(Year, abbr) %>%
  summarize(good_oline = (first(good_pos) == "Yes") + (last(good_pos) == "Yes")) %>%
  mutate(good_oline = case_when(good_oline == 2 ~ "Yes",
                               TRUE ~ "No"))

# Altogether now
offense_needs <- off_stats_fant %>% select(Year, abbr, unit, good_pos) %>% spread(unit, good_pos)

names(offense_needs) <- c("Year", "abbr", "good_qb", "good_rec", "good_rb", "good_te")

offense_needs <- offense_needs %>% replace_na(list(good_qb = "No", good_rec = "No", good_rb = "No", good_te = "No"))

## Joining, by = c("Year", "abbr")

```

Now onto defense. I have fantasy points against RB, WR, TE, and QB. As such, I first classify each team as being good or bad against the given position group as per the same specifications done for offense. This data is stored in def\_performance.

```
def_performance <- read_csv("def_performance.csv") %>% select(-X1)
```

```

## Parsed with column specification:
## cols(
##   X1 = col_integer(),
##   Tm = col_character(),
##   Year = col_integer(),
##   Rank_rb = col_integer(),
##   good_rb_def = col_character(),
##   Rank_wr = col_integer(),
##   good_wr_def = col_character(),
##   Rank_te = col_integer(),
##   good_te_def = col_character(),
##   Rank_qb = col_integer(),
##   good_qb_def = col_character(),
##   tot_rank = col_integer(),
##   good_def = col_character()
## )

```

I then have to translate these competencies into the player groups that I believe are most closely responsible for that ranking. Again, this is the subjective section.

For a team with a good defense against Running Backs, I claim that reflects primarily on good Defensive Line and Linebacker play. For a team with a good defense against Wide Receivers, I claim that reflects primarily on good Defensive Back play. For a team with a good defense against Tight Ends, I claim that reflects primarily on good Linebacker play. For a team with a good defense against Quarterbacks, I claim that reflects primarily on good Defensive Line, Defensive Back, and Linebacker play.

The following step makes these calculations for defense and then joins them to the offensive needs described earlier.

```
team_needs <- def_performance %>%
  mutate(good_line = case_when(good_rb_def == "Yes" & good_qb_def == "Yes" ~ "Yes",
                               TRUE ~ "No"),
         good_db = case_when(good_wr_def == "Yes" & good_qb_def == "Yes" ~ "Yes",
                               TRUE ~ "No"),
         good_lb=case_when(good_rb_def == "Yes" & good_te_def == "Yes" & good_qb_def == "Yes" ~ "Yes",
                           TRUE ~ "No")) %>%
  select(Tm, Year, good_def, good_line, good_db, good_lb) %>%
  mutate(Tm = case_when(Tm %in% "San Diego Chargers" ~ "LA Chargers",
                        Tm %in% "St. Louis Rams" ~ "LA Rams",
                        TRUE ~ Tm)) %>%
  arrange(Year, Tm) %>%
  filter(Year >= 2011) %>% # >=2011 due to spotrac salary data only going back that far, everything else
ungroup() %>%
  mutate(abbr = offense_needs$abbr) %>% # As long as both are arrange(Year, Tm) this will work
left_join(offense_needs) %>%
  mutate(logo = str_extract(Tm, "[A-Za-z0-9]+$")) %>%
  select(Year, Tm, abbr, logo, everything()) %>%
  mutate(abbr = case_when(abbr %in% "SEA" ~ "SF",
                         abbr %in% "SF" ~ "SEA",
                         TRUE ~ abbr)) %>% # Because alphabetical works on every abbr but these 2
  mutate(Year_start = Year,
        Year_end = Year + 1) %>% # Because the needs of a team are drafted in the Year_end draft, not just start
  select(-Year)

## Joining, by = c("Year", "abbr")
```

## Draft and Analysis

The last task to do prior to jumping into the analysis is to combine these team weaknesses and strengths with the draft picks made in the corresponding draft year. Of course the draft class relevant to addressing a team's weaknesses is the draft class following the end of the season in which those weaknesses arose. As such, draft picks are combined with the ending year of a given season. For example, the 2011-2012 season is paired with the 2012 draft class. Each pick is paired with their respective team as well as the value of draft capital spent to acquire them at that spot. This ignores trades that were made on draft day or leading up to it. Picking Trent Richardson 3rd overall is equivalent to trading up to pick Trent Richardson at 3rd overall.

```
need_and_draft <- draft_res %>% ungroup() %>% filter(Year >= 2012) %>% # Year >= 2012 bc end of season
left_join(team_needs, by = c("Team" = "logo", "Year" = "Year_end")) %>%
left_join(draft_valuation, by = c("Player"="Pick")) %>%
  mutate(addresses_need = case_when( (Position %in% "QB" & good_qb == "No") |
                                         (Position %in% c("DE", "DT") & good_line == "No") |
                                         (Position %in% c("RB") & good_rb == "No") |
                                         (Position %in% c("DB") & good_db == "No") |
                                         (Position %in% c("T", "C", "G") & good_oline == "No") |
                                         (Position %in% c("LB") & good_lb == "No") |
                                         (Position %in% c("WR") & good_rec == "No") |
                                         (Position %in% c("TE") & good_te == "No") ~ "Yes",
                                         TRUE ~ "No"),
                                         Value = case_when(!is.na(Value) ~ Value,
                                         TRUE ~ 1), # Low picks should only have value of 1
```

```

best_player_avail = case_when(addresses_need %in% "Yes" ~ "No",
                             TRUE ~ "Yes"),
value_for_need = case_when(addresses_need %in% "Yes" ~ Value,
                           TRUE ~ 0),
value_for_bpa = case_when(best_player_avail %in% "Yes" ~ Value,
                          TRUE ~ 0)) %>%
rename(Draft_year = Year) %>%
arrange(Draft_year, Team) %>%
group_by(Draft_year, Team) %>%
summarize(perc_need = sum(addresses_need == "Yes") / n() * 100,
          perc_bpa = 100 - perc_need,
          tot_value_for_need = sum(value_for_need),
          tot_value_for_bpa = sum(value_for_bpa))

```

For interesting plotting, I will combine team records for the season leading up to the draft and image recognition codes for use with ggimage.

```

master_record <- read_csv("master_record.csv") %>% select(-X1) %>% as_tibble() %>%
  mutate(wins = as.numeric(wins),
         losses = as.numeric(losses)) %>% right_join(need_and_draft, by = c("logo_ref" = "Team", "Year" = "Year"))
  rename(Draft_year = Year)

## Parsed with column specification:
## cols(
##   X1 = col_integer(),
##   wins = col_integer(),
##   losses = col_integer(),
##   logo_ref = col_character(),
##   Year = col_integer()
## )

team_abbr <- c("SF", "CHI", "CIN", "BUF", "DEN", "CLE", "TB", "ARI",
              "LAC", "KC", "IND", "DAL", "MIA", "PHI", "ATL", "NYG",
              "JAX", "NYJ", "DET", "GB", "CAR", "NE", "OAK", "LA",
              "BAL", "WAS", "NO", "SEA", "PIT", "HOU", "TEN", "MIN")

logo_translation <- data.frame(team = unique(master_record$logo_ref), team_code = team_abbr) %>%
  as_tibble() %>% mutate(team = as.character(team),
                         team_code = as.character(team_code))

master_record <- master_record %>% left_join(logo_translation, by = c("logo_ref" = "team"))

# Plotting tools
url.logo <- getURL("https://raw.githubusercontent.com/statsbylopez/BlogPosts/master/nfl_teamlogos.csv")
df.logos <- read.csv(text = url.logo) %>% mutate(team_code = as.character(team_code),
                                                 url = as.character(url))

master_record <- master_record %>% left_join(df.logos)

## Joining, by = "team_code"

```

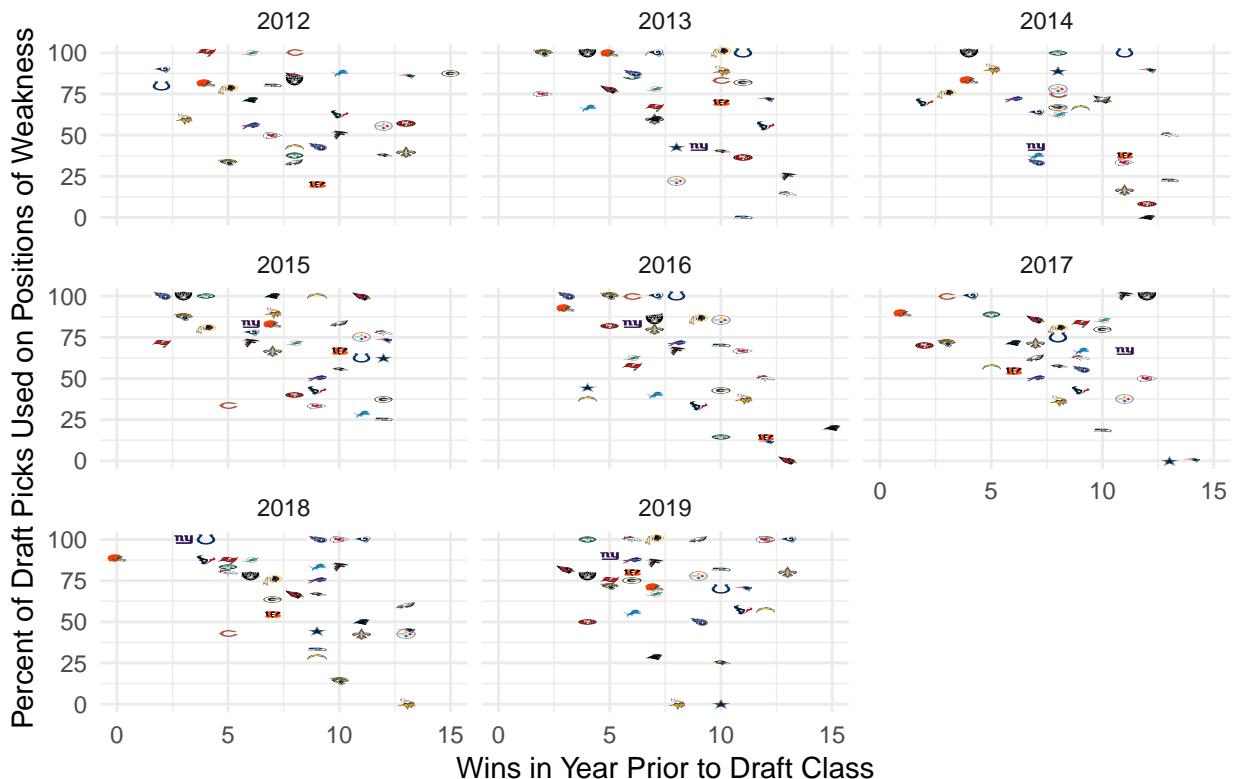
The first thing I want to investigate is if there is an association between the percentage of draft picks spent on positions of weakness and the team's season ending record.

```

master_record %>% ggplot(aes(x = wins, y = perc_need)) +
  geom_image(aes(image = url), size = 0.05) +
  xlab("Wins in Year Prior to Draft Class") +
  ylab("Percent of Draft Picks Used on Positions of Weakness") +
  theme_minimal() +
  ggtitle("Comparison of Drafting for Weakness and Previous Record") +
  facet_wrap(~ Draft_year)

```

## Comparison of Drafting for Weakness and Previous Record



There is a negative correlation -0.44 between the percentage of draft picks spent on positions of weakness and the team's season ending record. Part of the reason for this existence is the lack of positional weaknesses on good teams. Based on my labeling scheme described earlier, a team that has no immediate positional weaknesses does not draft for weakness.

I think what may lead to really interesting results would be to look at how team outcomes change from year to year considering how the team drafted. I can only do this to a limited scope as I do not have free agency and retirement captured. These factors may very well change teams more extremely than drafting does. Anyways, to the best I can do right now, here is a comparison between the percentage of draft picks used on areas of weakness against the win differential experienced by the end of the following season.

```

win_diff_master <- master_record %>% group_by(team_code) %>%
  arrange(Draft_year) %>%
  mutate(win_next_year = lead(wins, default = 8),
        win_diff = win_next_year - wins) %>%
  filter(Draft_year < 2019)

win_diff_master %>%
  ggplot(aes(x = win_diff, y = perc_need)) +

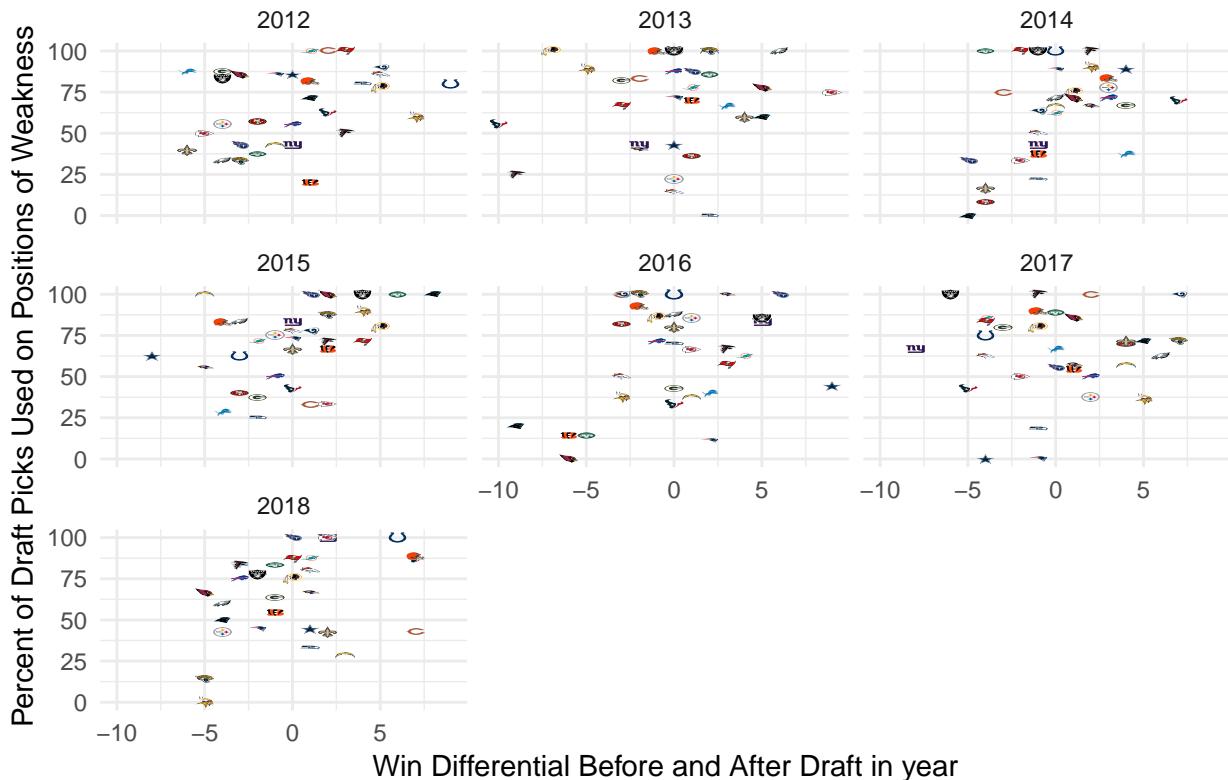
```

```

geom_image(aes(image = url), size = 0.05) +
xlab("Win Differential Before and After Draft in year") +
ylab("Percent of Draft Picks Used on Positions of Weakness") +
theme_minimal() +
ggtitle("Comparison of Drafting for Weakness and Change in Record") +
facet_wrap(~ Draft_year)

```

## Comparison of Drafting for Weakness and Change in Record



This time a general positive trend is noticeable, helping us to potentially identify teams that focus more on drafting for weaknesses improving in terms of wins year over year. The correlation is 0.27.

So far this review has only investigated the percentage of picks that went into addressing positions of weakness. Concluding here would be disingenuous to the nature of the draft as there are different expectations for first round draft picks and seventh round draft picks, along with everyone in between. Capturing this difference is achieved through a value function, in which each pick is assigned a value. The first pick is worth 3000, the second pick 2600, the third pick 2200, so on and so forth as the value continues to decay. A more useful exercise may therefore be to capture how much value, or draft capital, a team actually places in addressing positions of weakness. Using two firsts on OL help and 5 4th+ rounders on positions of depth would reflect as a best player available strategy earlier in this analysis. In this case, however, the overwhelming quantity of draft capital has been spent on positions of weakness, that being the OL. This seems to be a more reasonable way of classifying how teams address weakness. The following plot will demonstrate how teams currently perform in this manner, with respect to league based unit weakness.

```

master_record %>% group_by(team_code) %>%
  arrange(Draft_year) %>%
  mutate(win_next_year = lead(wins, default = 8),
         win_diff = win_next_year - wins) %>%
  filter(Draft_year < 2019) %>% # this uses the default as no results accumulated yet, no need for this

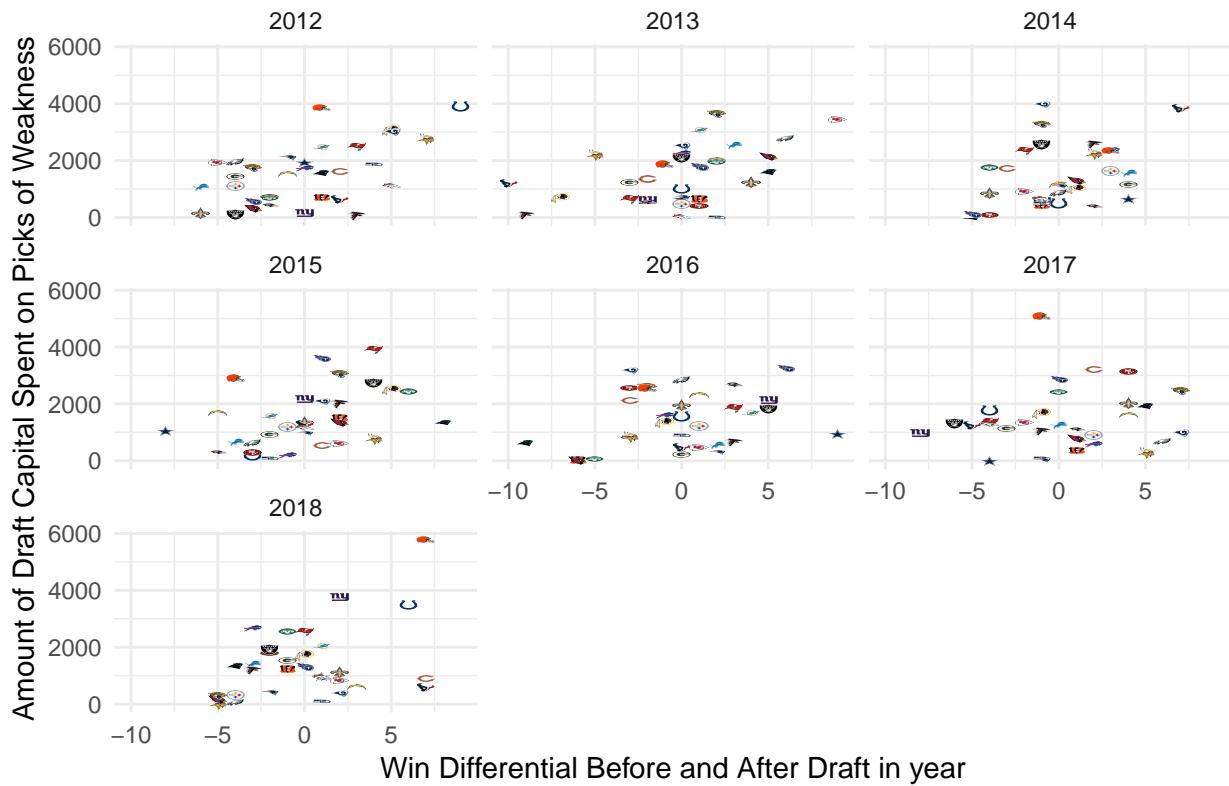
```

```

ggplot(aes(x = win_diff, y = tot_value_for_need)) +
  geom_image(aes(image = url), size = 0.05) +
  xlab("Win Differential Before and After Draft in year") +
  ylab("Amount of Draft Capital Spent on Picks of Weakness") +
  theme_minimal() +
  ggtitle("Comparison of Drafting for Weakness and Change in Record") +
  facet_wrap(~ Draft_year)

```

Comparison of Drafting for Weakness and Change in Record

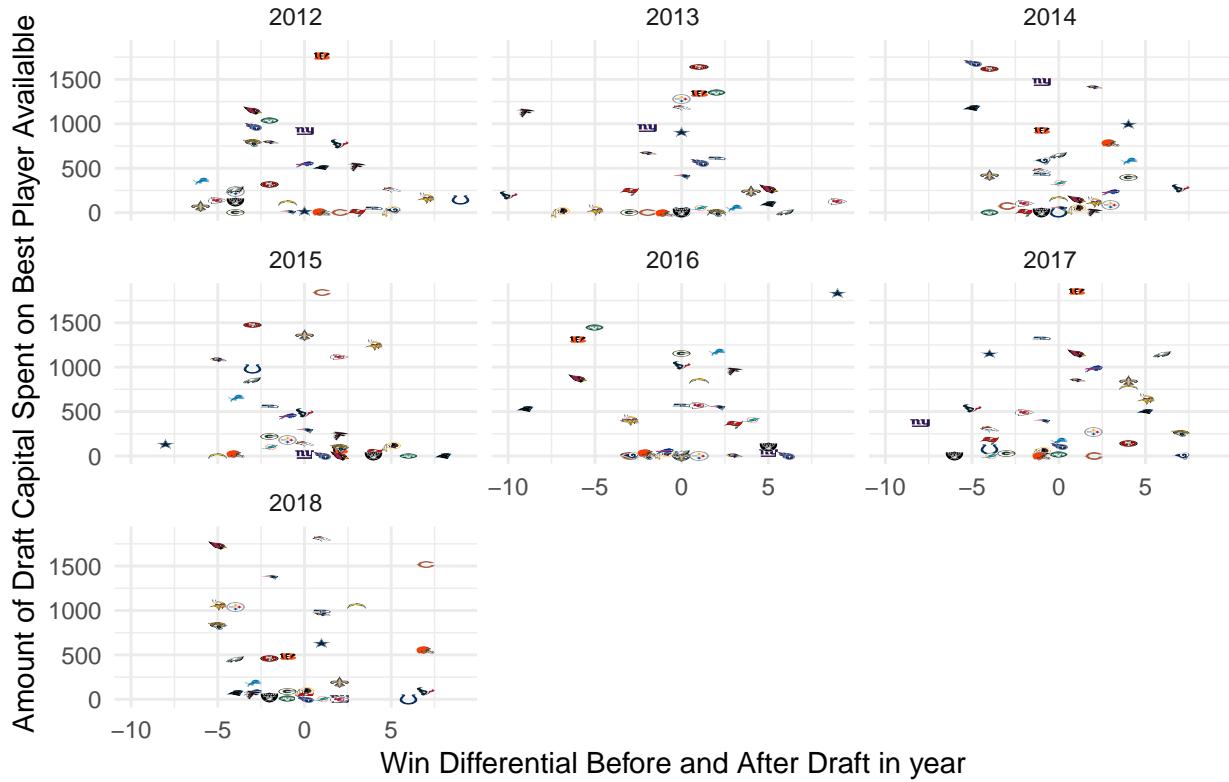


```

master_record %>% group_by(team_code) %>%
  arrange(Draft_year) %>%
  mutate(win_next_year = lead(wins, default = 8),
         win_diff = win_next_year - wins) %>%
  filter(Draft_year < 2019) %>% # this uses the default as no results accumulated yet, no need for this
  ggplot(aes(x = win_diff, y = tot_value_for_bpa)) +
  geom_image(aes(image = url), size = 0.05) +
  xlab("Win Differential Before and After Draft in year") +
  ylab("Amount of Draft Capital Spent on Best Player Available") +
  theme_minimal() +
  ggtitle("Comparison of Drafting BPA and Change in Record") +
  facet_wrap(~ Draft_year)

```

## Comparison of Drafting BPA and Change in Record



```
corr_val <- master_record %>% group_by(team_code) %>%
  arrange(Draft_year) %>%
  mutate(win_next_year = lead(wins, default = 8),
        win_diff = win_next_year - wins) %>%
  filter(Draft_year < 2019)
corr_val_need <- cor(corr_val$tot_value_for_need, corr_val$win_diff)
corr_val_bpa <- cor(corr_val$tot_value_for_bpa, corr_val$win_diff)
lm(win_diff ~ tot_value_for_need, data = corr_val) %>% summary()
```

```
##
## Call:
## lm(formula = win_diff ~ tot_value_for_need, data = corr_val)
##
## Residuals:
##      Min       1Q   Median       3Q      Max 
## -9.655 -2.260  0.099  2.157  9.649 
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)    
## (Intercept) -1.8534800  0.3822061 -4.849 2.33e-06 ***
## tot_value_for_need 0.0012682  0.0002129   5.957 9.96e-09 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.354 on 222 degrees of freedom
## Multiple R-squared:  0.1378, Adjusted R-squared:  0.1339
```

```
## F-statistic: 35.49 on 1 and 222 DF, p-value: 9.959e-09
```

The correlation in the plot for positions of weakness is 0.3712501. Though not perfect, there is something to be said about the existence of an association between an improved record and spending draft capital on positions of weakness. For example, the same comparison using the total value for BPA has a correlation of -0.070345. Furthermore, only the position of weakness generates statistically significant results. This suggests that there is an association between drafting for weakness and the number of wins on the following season, while there is no evidence that the same exists for best player available strategies. Lets take a quick dive into some of the largest changes on record to explore where this analysis fits and where it falls flat.

```
best_5 <- master_record %>% group_by(team_code) %>%
  arrange(Draft_year) %>%
  mutate(win_next_year = lead(wins, default = 8),
         win_diff = win_next_year - wins) %>%
  filter(Draft_year < 2019) %>%
  ungroup() %>%
  arrange(desc(win_diff)) %>%
  slice(1:5)
best_5

## # A tibble: 5 x 13
##   wins losses logo_ref Draft_year perc_need perc_bpa tot_value_for_n~
##   <dbl>  <dbl> <chr>        <dbl>      <dbl>      <dbl>
## 1     2      2  Colts       2012       80        20      3906.
## 2     2      2  Chiefs      2013       75        25      3451
## 3     4      12 Cowboys    2016      44.4      55.6      950.
## 4     7      8 Panthers   2015      100        0      1348.
## 5     3      13 Vikings   2012       60        40      2748.
## # ... with 6 more variables: tot_value_for_bpa <dbl>, team_code <chr>,
## #   team <fct>, url <chr>, win_next_year <dbl>, win_diff <dbl>
```

The 5 biggest turn arounds for teams in this data set were the 2012 Colts, 2013 Chiefs, and 2016 Cowboys (9 more wins), the 2015 Panthers (8 more wins), and the 2012 Vikings (7 more wins). As a reminder, the year referenced is the year that the team had the better season. For example, the 2015 Panthers went 15-1 while the 2014 Panthers went 7-8-1 for a difference of 8 wins.

First, the 2012 Colts. The draft class for the 2012 Colts contained players like Andrew Luck, T.Y. Hilton, and Coby Fleener. Based on the classification of positions of weakness, the Colts had holes at nearly every position with the exception of Wide Receiver. To address these weaknesses, the Colts spent their first 3 picks at shoring up the other positions before grabbing T.Y. Hilton as the Best Player Available with their second pick of the 3rd round. The win differential realized by the Colts in 2012 can largely be tied to them hitting with the highest value pick of the draft on Andrew Luck. Stepping in to replace Peyton Manning is no easy task but with Peyton having been sidelined the previous season with injury, it is not surprising that the Colts were able to rebound so strongly with a more than competent quarterback. This may point to a downside of the analysis, as the injury to a figurehead of the offense artificially creates positions of weakness, especially when that position is the quarterback. Further, some quick googling suggests that the Colts had a decent free agency period. Talents were signed, but it would be hard to argue any of these talents having a greater effect on the team moving forward than Andrew Luck. Regardless, the 2012 Colts represent a remarkable turnaround and are a highlight of how drafting for positions of weakness can push a team forward.

The 2013 Chiefs did similar to the 2012 Colts, spending heavily on positions of weakness by throwing the number one overall pick at Eric Fisher. Though Fisher is not necessarily the game changer that Luck is, his addition addressed a weakness in the Chiefs Offensive Line. The bigger change of course points to the main weakness of this approach, that of which being a lack of Free Agency knowledge. The Chiefs brought in both QB Alex Smith and Head Coach Andy Reid in 2013. These two additions likely have been more impactful than Eric Fisher in the scheme of the team and helped facilitate the turn around.

Each of the 3 remaining teams have a story composed of some of the aspects above. The 2016 Cowboys went BPA with Ezekiel Elliott and struck gold with Dak Prescott in the later rounds to shore up an aging Tony Romo. The 2015 Panthers put their limited number of picks entirely into positions of weakness and were greatly aided by an MVP season from Cam Newton. The 2012 Vikings cleaned up some weaknesses in their Offensive Line and Defensive Back roles but also welcomed an MVP season from Adrian Peterson.

The common theme with these changing teams is that a lot changes, so much so that it is hard to distinguish what causes each result. By taking Matt Kalil, did the Vikings do enough on the Offensive Line to push Adrian Peterson to an MVP season? Do the Cowboys dominate to the same extent with BPA Ezekiel Elliott or an available player at a position of weakness like Jack Conklin or DeForest Buckner? These are things we will likely never have the answers to, all the more unfortunate for this work.

Naturally, lets then look at the 5 biggest draft capital investors for positions of weakness. The top 5 are as follows.

```
best_5_val <- master_record %>% group_by(team_code) %>%
  arrange(Draft_year) %>%
  mutate(win_next_year = lead(wins, default = 8),
        win_diff = win_next_year - wins) %>%
  filter(Draft_year < 2019) %>%
  ungroup() %>%
  arrange(desc(tot_value_for_need)) %>%
  slice(1:5)
best_5_val

## # A tibble: 5 x 13
##   wins losses logo_ref Draft_year perc_need perc_bpa tot_value_for_n~
##   <dbl>  <dbl> <chr>       <dbl>      <dbl>      <dbl>      <dbl>
## 1     0     16 Browns     2018      88.9     11.1     5788
## 2     1     15 Browns     2017      90        10       5097.
## 3     7      9 Rams      2014      63.6     36.4     3974.
## 4     2     14 Colts     2012      80        20       3906.
## 5     2     14 Buccane~    2015      71.4     28.6     3901
## # ... with 6 more variables: tot_value_for_bpa <dbl>, team_code <chr>,
## #   team <fct>, url <chr>, win_next_year <dbl>, win_diff <dbl>
```

With great draft capital comes great responsibility and so we see teams that either spent a number 1 overall pick on a position of weakness or had multiple first round picks and spent both on positions of weakness. Of the teams above, two teams did worse the following year (2014 Rams and 2017 Browns, 1 fewer win each) while the rest improved (Bucs +4, Browns 2018 +7, Colts +9). Lets dive in a bit to these teams and their drafts.

We will begin with the elephant in the room, Cleveland. Possessing back to back drafts of enormous draft capital (2 first round picks in each), the Browns had gaping holes at almost every roster position. The first of these drafts, 2017, lead to the worst season on record with 0 wins and 16 losses. The second draft, 2018, lead to an enormous win differential of positive 7. All of their first round picks, save for Jabrill Peppers, are still on the team and play important roles. All of them also happened to address an area of weakness on the Browns.

We have already covered the 2012 Colts in depth and so we will skip them for ease of reading. The 2014 Rams had a 6-10 season following a draft class that included Lamarcus Joyner and Aaron Donald. Taking them into orbit would then be the eventual hiring of Sean McVay and drafting of Jared Goff the following year.

The 2015 Buccaneers drafted Jameis Winston and Kwon Alexander, two important upgrades to their lacking 2014 team. Although only moving up to 6 wins, the 2015 Buccaneers improved in part due to their new additions. No head coaching or major free agency changes happened here, possibly suggesting that the 2015 draft class predominantly helped push this team forward.

## Position of Greatest Weakness

The final question I hope to answer in this report, or at least open up to discussion, is the priority with which to address positions of need. Football has been suggested to be a weak link sport in which a team is defined by its weakest member rather than its greatest contributor. To that extent, we would expect any and every position of weakness to be valuable to fill. Draft pundits and members of the football community, however, recognize the value a franchise QB can have on a team.

```
need_and_draft_by_pos <- draft_res %>% ungroup() %>% filter(Year >= 2012) %>% # Year >= 2012 bc end of
  left_join(team_needs, by = c("Team" = "logo", "Year" = "Year_end")) %>%
  left_join(draft_valuation, by = c("Player" = "Pick")) %>%
  mutate(addresses_need = case_when( (Position %in% "QB" & good_qb == "No") |
    (Position %in% c("DE", "DT") & good_line == "No") |
    (Position %in% c("RB") & good_rb == "No") |
    (Position %in% c("DB") & good_db == "No") |
    (Position %in% c("T", "C", "G") & good_oline == "No") |
    (Position %in% c("LB") & good_lb == "No") |
    (Position %in% c("WR") & good_rec == "No") |
    (Position %in% c("TE") & good_te == "No") ~ "Yes",
    TRUE ~ "No"),
    Value = case_when(!is.na(Value) ~ Value,
      TRUE ~ 1), # Low picks should only have value of 1
  best_player_avail = case_when(addresses_need %in% "Yes" ~ "No",
    TRUE ~ "Yes"),
  value_for_need = case_when(addresses_need %in% "Yes" ~ Value,
    TRUE ~ 0),
  value_for_bpa = case_when(best_player_avail %in% "Yes" ~ Value,
    TRUE ~ 0)) %>%
  rename(Draft_year = Year) %>%
  arrange(Draft_year, Team) %>%
  group_by(Draft_year, Team, Position) %>%
  summarize(tot_value_for_need = sum(value_for_need),
    tot_value_for_bpa = sum(value_for_bpa)) %>%
  select(-tot_value_for_bpa)

master_record_2 <- read_csv("master_record.csv") %>% select(-X1) %>% as_tibble() %>%
  mutate(wins = as.numeric(wins),
    losses = as.numeric(losses)) %>% right_join(need_and_draft_by_pos, by = c("logo_ref" = "Team",
  rename( Draft_year = Year)

## Warning: Missing column names filled in: 'X1' [1]
## Parsed with column specification:
## cols(
##   X1 = col_integer(),
##   wins = col_integer(),
##   losses = col_integer(),
##   logo_ref = col_character(),
##   Year = col_integer()
## )
team_abbr <- c("SF", "CHI", "CIN", "BUF", "DEN", "CLE", "TB", "ARI",
  "LAC", "KC", "IND", "DAL", "MIA", "PHI", "ATL", "NYG",
  "JAX", "NYJ", "DET", "GB", "CAR", "NE", "OAK", "LA",
  "BAL", "WAS", "NO", "SEA", "PIT", "HOU", "TEN", "MIN")
```

```

logo_translation <- data.frame(team = unique(master_record_2$logo_ref), team_code = team_abbr) %>%
  as_tibble() %>% mutate(team = as.character(team),
                           team_code = as.character(team_code))

master_record_2 <- master_record_2 %>% left_join(logo_translation, by = c("logo_ref" = "team")) %>%
  spread(key = Position, value = tot_value_for_need) %>%
  replace(is.na(.), 0) %>%
  group_by(Draft_year, team_code) %>%
  slice(1) %>%
  select(-K, -P)

master_record_2 <- master_record_2 %>% left_join(df.logos) %>% group_by(team_code) %>%
  arrange(Draft_year) %>%
  mutate(win_next_year = lead(wins, default = 8),
        win_diff = win_next_year - wins) %>%
  filter(Draft_year < 2019) %>%
  ungroup()

## Joining, by = "team_code"
win_diff_lm <- lm(win_diff ~ C + DB + DE + DT + G + LB + QB + RB + T + TE + WR, data = master_record_2)

win_sum <- win_diff_lm %>% summary()
sort(win_sum$coefficients[,1], decreasing = T)

##          C           G           TE          T           LB
## 2.329969e-03 2.325051e-03 2.280558e-03 1.912001e-03 1.491467e-03
##          RB           DE           DB           WR           QB
## 1.487355e-03 1.255340e-03 1.052013e-03 9.202920e-04 9.074018e-04
##          DT  (Intercept)
## 7.089255e-05 -1.869578e+00

library(car)

## Loading required package: carData
##
## Attaching package: 'car'
##
## The following object is masked from 'package:dplyr':
## 
##     recode
##
## The following object is masked from 'package:purrr':
## 
##     some
Anova(win_diff_lm, type = "II")

## Anova Table (Type II tests)
##
## Response: win_diff
##             Sum Sq Df F value    Pr(>F)
## C            21.84  1  1.9198 0.1673324
## DB           45.64  1  4.0125 0.0464382 *

```

```

## DE      76.52   1  6.7274 0.0101561 *
## DT      0.09    1  0.0076 0.9304322
## G       87.90   1  7.7283 0.0059252 **
## LB      52.24   1  4.5928 0.0332456 *
## QB      61.01   1  5.3643 0.0215087 *
## RB      44.89   1  3.9470 0.0482434 *
## T       156.33  1 13.7446 0.0002674 ***
## TE      28.26   1  2.4849 0.1164374
## WR      22.23   1  1.9546 0.1635562
## Residuals 2411.22 212
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

The above linear model is with respect to the effect of draft capital spent on a given position of weakness and the change in record the following year. A positive coefficient means that spending draft capital on that position is associated with an improved record the following year (a positive win differential). A draft of only best players available would have a value of 0 for each of the predictors above as none of the capital was used on positions of weakness.

Interpreting the model above, the positions with the greatest effect on win differential are Centres, Guards, Tight Ends, Tackles and Linebackers. This may be because players in these positions are ready to start in the NFL earlier than players at other positions such as QB or because they are drafted a few positions later in the first round where they cost less draft capital but are still impactful. Regardless, it seems to be the case that investments in the Offensive Line and Blocking (Tight Ends) pay the largest immediate dividends.

Looking a little closer at the results through a Type 2 Sums of Square Anova comparison, we see that the effects for Tight Ends and Centres could reasonably be chalked up to random chance as they are insignificant in our test. Rather significance exists for Tackles, Guards, Defensive Ends, Quarterbacks, Linebackers, Defensive Backs, and Running Backs. Some of these positions, such as Running Back, tend to be drafted later overall and their depressed market value may be the reason we see their effect as statistically significant.

## Conclusion

Many draft analysts and fans alike see the draft as a chance to improve teams in the NFL with the best available talent. But drafting for the best available player at a given draft position is not always the best approach as it neglects the current areas of weakness on a team.

Since the ultimate goal of the NFL is to win the super bowl and doing so requires a team to win more games, we investigated the relationship between win differential year over year and different approaches to viewing draft strategy. We found that drafting for positional weaknesses was positively correlated at a statistically significant level (5%) with win differential while drafting for best player available was not statistically significant, suggesting it may not be associated with win differential. We investigated further by looking at the amount of draft capital teams were pouring into their positions of weakness and again found that teams who used more draft capital tended to have better win differentials than those that didn't, significant at the 5% level again. No such trend was found for best player available approaches. Finally, we took a brief look into the priority of weaknesses to address. The preliminary results seem to suggest that Offensive Line improvements yield the largest and most immediate benefit to a team with respect to win differential.

This work is still yet incomplete. There is some discussion below based on Twitter feedback from some initial findings, discussing alternative ways to just weakness / need. Future work could benefit from using WAR or non-fantasy dependent stats to define positional strengths and weaknesses within a roster. Finally, an extension to positional need based on salary issues could be of interest. Teams are still businesses at the end of the day and money is limited.

Thank you for reading this far. Please feel free to reach out and continue the discussion on Twitter! I look forward to seeing what others think of these ideas!

## In Response to Feedback on Good teams going BPA

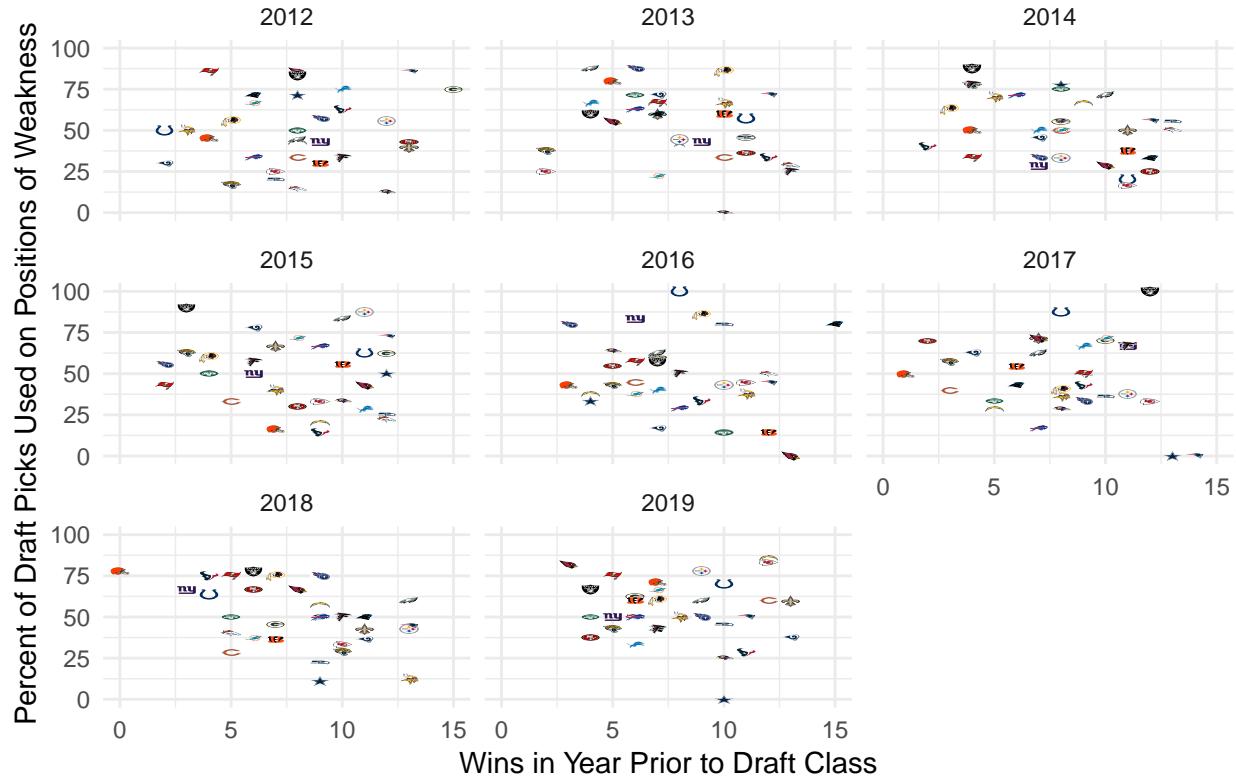
As prompted in an early discussion on Twitter, it was suggested that good teams are good because they take the best player available rather than drafting for need. To investigate this idea we could consider a team, regardless of record, to have immediate weaknesses in any category that ranks above their average rank. For a simplified example, if we assume Team A has an average rank of 10.3 then any positional group with a rank > 10.3 would be a positional weakness. This should give us some grounds to explore this idea.

```
## Parsed with column specification:
## cols(
##   .default = col_character(),
##   X1 = col_integer(),
##   dline_rank = col_integer(),
##   db_rank = col_integer(),
##   lb_rank = col_integer(),
##   qb_rank = col_integer(),
##   wr_rank = col_integer(),
##   rb_rank = col_integer(),
##   te_rank = col_integer(),
##   oline_rank = col_integer(),
##   Year_start = col_integer(),
##   Year_end = col_integer(),
##   avg_rank = col_double()
## )
## See spec(...) for full column specifications.

## Joining, by = "Pick"

## Parsed with column specification:
## cols(
##   X1 = col_integer(),
##   wins = col_integer(),
##   losses = col_integer(),
##   logo_ref = col_character(),
##   Year = col_integer()
## )
## Joining, by = "team_code"
```

## Comparison of Drafting for Weakness and Previous Record

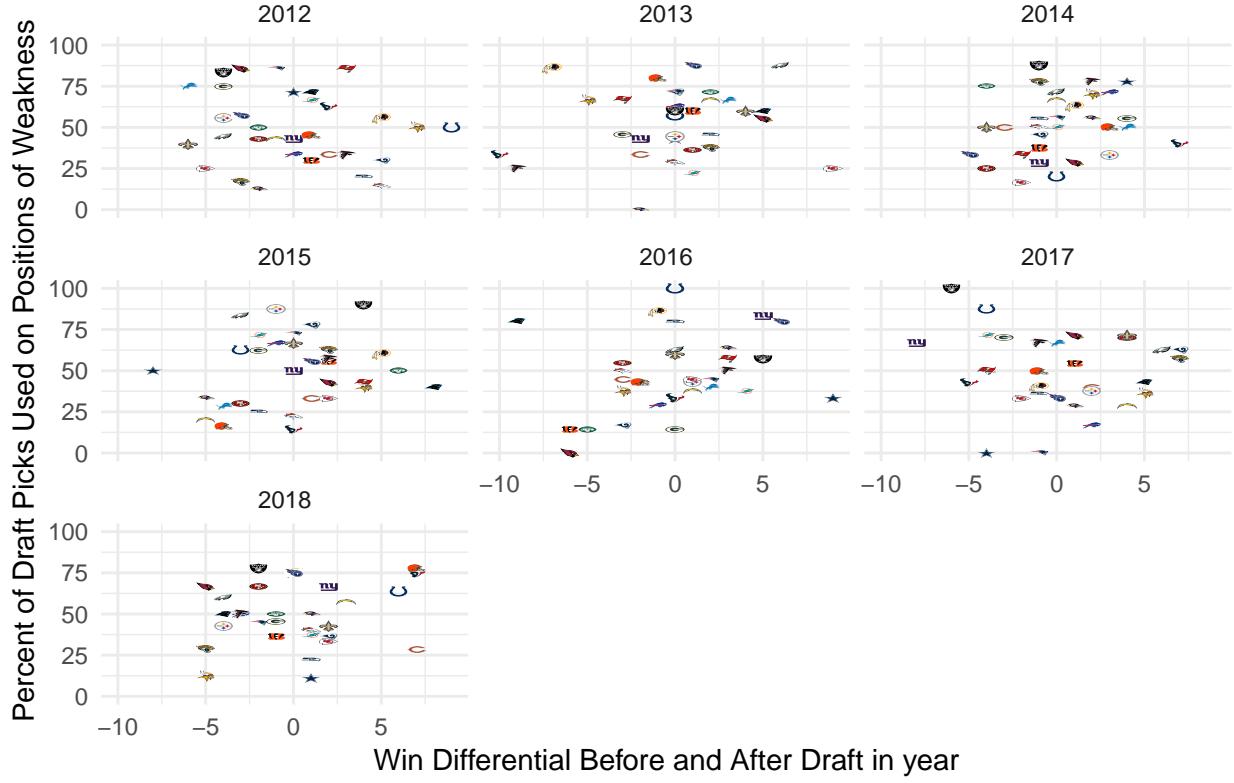


Describing positional weakness in this new way, we see that the negative correlation is much less obvious, being only -0.19. I think the biggest takeaway from this plot is that regardless of the number of wins a team has, teams still tend to put a number of picks into the positions they deem relatively weaker than elsewhere on the team rather than just going for the best player available outright. Revolutionary, I know.

```
win_diff_master_avg <- master_record_avg %>% group_by(team_code) %>%
  arrange(Draft_year) %>%
  mutate(win_next_year = lead(wins, default = 8),
         win_diff = win_next_year - wins) %>%
  filter(Draft_year < 2019)

win_diff_master_avg %>% # this uses the default as no results accumulated yet, no need for this in plot
ggplot(aes(x = win_diff, y = perc_need)) +
  geom_image(aes(image = url), size = 0.05) +
  xlab("Win Differential Before and After Draft in year") +
  ylab("Percent of Draft Picks Used on Positions of Weakness") +
  theme_minimal() +
  ggtitle("Comparison of Drafting for Weakness and Change in Record") +
  facet_wrap(~ Draft_year)
```

## Comparison of Drafting for Weakness and Change in Record



Considering instead the relative weaknesses of a team in comparison to their other position groups, the trend previously seen is much more subdued. There is only a correlation of 0.09. This result is not statistically significant at the 5% level, unlike all previous results. Therefore there does not appear to be an association between the percent of draft picks used on positions of weakness and win differential when only looking within the team for weakness. Perhaps this hints at good teams drafting for best player available or evaluating talent differently.

% Review this section ^ ^

Again, considering the alternative way of capturing team weaknesses:

```
master_record_avg %>% group_by(team_code) %>%
  arrange(Draft_year) %>%
  mutate(win_next_year = lead(wins, default = 8),
         win_diff = win_next_year - wins) %>%
  filter(Draft_year < 2019) %>% # this uses the default as no results accumulated yet, no need for this
  ggplot(aes(x = win_diff, y = tot_value_for_need)) +
  geom_image(aes(image = url), size = 0.05) +
  xlab("Win Differential Before and After Draft in year") +
  ylab("Amount of Draft Capital Spent on Picks of Weakness") +
  theme_minimal() +
  ggtitle("Comparison of Drafting for Weakness and Change in Record") +
  facet_wrap(~ Draft_year)
```

## Comparison of Drafting for Weakness and Change in Record

