Fantasy vs. Reality

Jake Berberian

Data Import and Clean

```
scores <- pbp %>%
  group_by(game_id) %>%
  summarize(n = n(),
            home_team = home_team[n],
            home_score = total_home_score[n],
            away_team = away_team[n],
            away_score = total_away_score[n]) %>%
  select(-n)

scores %>%
  head(5) %>%
  kbl(booktabs = TRUE) %>%
  kable_styling(latex_options = "striped")
```

game_id	home_team	home_score	away_team	away_score
2010_01_ARI_STL	LA	13	ARI	17
2010_01_ATL_PIT	PIT	15	ATL	9
2010_01_BAL_NYJ	NYJ	9	BAL	10
2010_01_CAR_NYG	NYG	31	CAR	18
2010_01_CIN_NE	NE	38	CIN	24

Columns that are important:

- QB/RB/WR/TE
  - play_type
  - yards_gained
  - td_team
  - pass_touchdown
  - rush_touchdown
  - return_touchdown
- Kickers
  - field_goal_result
  - kick_distance
- D/ST
  - interception

- fumble_forced
- safety
- fumble_lost
- sack

Skill Positions

Here's a tricky thing: *return_touchdown* counts all types of returns (interceptions, fumbles, punts, etc.). So we'll need to deal with that on our own because a fumble return TD is only worth 6 points to the D/ST, but a kickoff return TD is worth 12 (6 to the individual player, 6 to the D/ST).

```
tds_tos <- pbp %>%
  group_by(game_id, posteam) %>%
  summarize(pass_td = sum(pass_touchdown),
            rush_td = sum(rush_touchdown),
            fumble = sum(fumble_lost),
            int = sum(interception)) %>%
  drop_na() %>%
  mutate(ft_pts = 6*(pass_td + rush_td) - 2*fumble - 1*int) %>%
  rename(score_team = posteam) %>%
  select(game_id, score_team, ft_pts)

yardage <- pbp %>%
  group_by(game_id, posteam, play_type) %>%
  summarize(yrds = sum(yards_gained)) %>%
  drop_na() %>%
  filter(play_type %in% c("pass", "run", "qb_kneel")) %>%
  mutate(ft_pts = case_when(
    play_type == "pass" ~ yrds * 0.14,
    play_type == "run" ~ yrds * 0.1,
    play_type == "qb_kneel"~ yrds * 0.1)) %>%
  rename(score_team = posteam) %>%
  ungroup() %>%
  select(game_id, yrds, ft_pts, score_team) %>%
  group_by(game_id, score_team) %>%
  summarize(ft_pts = sum(ft_pts))


returns <- pbp %>%
  filter(return_touchdown == 1) %>%
  group_by(game_id, play_type, posteam, defteam) %>%
  summarize(kick_ret = sum(play_type == "kickoff" | play_type == "punt"),
            def_ret = sum(play_type == "pass" | play_type == "run")) %>%
  ungroup() %>%
  mutate(score_team = case_when(
    kick_ret != 0 ~ posteam,
    def_ret != 0 ~ defteam)) %>%
  ungroup() %>%
  mutate(ft_pts = 6 * def_ret + 12 * kick_ret) %>%
  group_by(game_id, score_team) %>%
  summarize(ft_pts = sum(ft_pts))
```

All looks good: checked a few randomly selected games using ProFootballReference boxscores.

Kickers

```
fgs <- pbp %>%
  filter(play_type == "field_goal" | play_type == "extra_point") %>%
  select(game_id, posteam, play_type, field_goal_result, kick_distance) %>%
  group_by(game_id, posteam) %>%
  summarize(fg19 = sum(field_goal_result == "made" & kick_distance < 20),
            fg29 = sum(field_goal_result == "made" & kick_distance > 20 & kick_distance < 30),
            fg39 = sum(field_goal_result == "made" & kick_distance > 30 & kick_distance < 40),
            fg49 = sum(field_goal_result == "made" & kick_distance > 40 & kick_distance < 50),
            fg50 = sum(field_goal_result == "made" & kick_distance > 49)) %>%
  mutate(ft_pts = 3 * fg19 + 3 * fg29 + 3 * fg39 + 4 * fg49 + 5 * fg50) %>%
  ungroup() %>%
  select(game_id, posteam, ft_pts) %>%
  rename(score_team = posteam)


pats <- pbp %>%
  filter(play_type == "extra_point") %>%
  select(game_id, posteam, extra_point_result) %>%
  group_by(game_id, posteam) %>%
  summarize(pats = sum(extra_point_result == "good")) %>%
  mutate(ft_pts = 1 * pats) %>%
  ungroup() %>%
  select(-pats) %>%
  rename(score_team = posteam)

kickers <- full_join(fgs, pats, by = c("game_id", "score_team")) %>%
  mutate(ft_pts.x = replace_na(ft_pts.x, 0),
         ft_pts.y = replace_na(ft_pts.y, 0),
         ft_pts = ft_pts.x + ft_pts.y) %>%
  select(-c(ft_pts.x, ft_pts.y))
```

Defense and Special Teams

I don't think blocked punts count for additional points, so we won't include them. Need to add two point conversions.

```
turnovers <- pbp %>%
  group_by(game_id, defteam) %>%
  summarize(int = sum(interception),
            sack = sum(sack),
            sfty = sum(safety),
            fmbl = sum(fumble_lost)) %>%
  drop_na() %>%
  mutate(ft_pts = sack + 2*(int+sfty+fmbl)) %>%
  rename(score_team = defteam) %>%
  select(game_id, score_team, ft_pts)

points_allowed <- scores %>%
  pivot_longer(cols = c(away_score, home_score), names_to = "team", values_to = "score") %>%
  mutate(def_team = case_when(
```

```
    team == "home_score" ~ away_team,
    team == "away_score" ~ home_team)) %>%
  rename(pts_allowed = score) %>%
  select(game_id, def_team, pts_allowed) %>%
  mutate(ft_pts = case_when(
    pts_allowed == 0 ~ 10,
    pts_allowed < 7 & pts_allowed > 0 ~ 7,
    pts_allowed < 14 & pts_allowed > 6 ~ 4,
    pts_allowed < 21 & pts_allowed > 13 ~ 1,
    pts_allowed < 28 & pts_allowed > 20 ~ 0,
    pts_allowed < 35 & pts_allowed > 27 ~ -1,
    pts_allowed > 34 ~ -4,
  )) %>%
  rename(score_team = def_team) %>%
  select(-pts_allowed)


blks <- pbp %>%
  filter(field_goal_result == "blocked") %>%
  group_by(game_id, defteam, field_goal_result) %>%
  summarize(blk = sum(field_goal_result == "blocked")) %>%
  mutate(ft_pts = 2 * blk) %>%
  ungroup() %>%
  rename(score_team = defteam) %>%
  select(game_id, score_team, ft_pts)
```

Fantasy Points: Total

```
fantasy <- yardage %>%
  bind_rows(tds_tos) %>%
  bind_rows(returns) %>%
  bind_rows(kickers) %>%
  bind_rows(turnovers) %>%
  bind_rows(points_allowed) %>%
  bind_rows(blks) %>%
  group_by(game_id, score_team) %>%
  summarize(ft_pts = sum(ft_pts))

fantasy %>%
  head(10) %>%
  kbl(booktabs = TRUE) %>%
  kable_styling(latex_options = "striped")
```

game_id	score_team	ft_pts
2010_01_ARI_STL	ARI	71.44
2010_01_ARI_STL	LA	64.10
2010_01_ATL_PIT	ATL	53.98
2010_01_ATL_PIT	PIT	67.84
2010_01_BAL_NYJ	BAL	49.52
2010_01_BAL_NYJ	NYJ	37.00
2010_01_CAR_NYG	CAR	51.62
2010_01_CAR_NYG	NYG	88.92
2010_01_CIN_NE	CIN	75.44
2010_01_CIN_NE	NE	96.92

We need to get ready for a join with the scores dataframe.

```
new_scores <- scores %>%
  pivot_longer(cols = c(away_score, home_score), names_to = "team", values_to = "score") %>%
  mutate(score_team = case_when(
    team == "home_score" ~ home_team,
    team == "away_score" ~ away_team)) %>%
  select(game_id, score_team, score)

new_scores %>%
  head(10) %>%
  kbl(booktabs = TRUE) %>%
  kable_styling(latex_options = "striped")
```

game_id	score_team	score
2010_01_ARI_STL	ARI	17
2010_01_ARI_STL	LA	13
2010_01_ATL_PIT	ATL	9
2010_01_ATL_PIT	PIT	15
2010_01_BAL_NYJ	BAL	10
2010_01_BAL_NYJ	NYJ	9
2010_01_CAR_NYG	CAR	18
2010_01_CAR_NYG	NYG	31
2010_01_CIN_NE	CIN	24
2010_01_CIN_NE	NE	38

```
final <- right_join(fantasy, new_scores, by = c("game_id", "score_team"))

final %>%
  head(10) %>%
  kbl(booktabs = TRUE) %>%
  kable_styling(latex_options = "striped")
```

game_id	score_team	ft_pts	score
2010_01_ARI_STL	ARI	71.44	17
2010_01_ARI_STL	LA	64.10	13
2010_01_ATL_PIT	ATL	53.98	9
2010_01_ATL_PIT	PIT	67.84	15
2010_01_BAL_NYJ	BAL	49.52	10
2010_01_BAL_NYJ	NYJ	37.00	9
2010_01_CAR_NYG	CAR	51.62	18
2010_01_CAR_NYG	NYG	88.92	31
2010_01_CIN_NE	CIN	75.44	24
2010_01_CIN_NE	NE	96.92	38

Results time!

```
final_reveal <- final %>%
  group_by(game_id) %>%
  summarize(W_game = score_team[which.max(score)],
            W_ft = score_team[which.max(ft_pts)]) %>%
  mutate(diff = W_game == W_ft) %>%
  count(diff)

final_reveal %>%
  kbl(booktabs = TRUE) %>%
  kable_styling(latex_options = "striped")
```

diff	n
FALSE	407
TRUE	2263

```
final_reveal %>%
  summarize(diff, prop = round(n / sum(n), 4)) %>%
  kbl(booktabs = TRUE) %>%
  kable_styling(latex_options = "striped")
```

diff	prop
FALSE	0.1524
TRUE	0.8476

Some Brief Analysis

Let's look at the largest difference in fantasy points.

```
analysis <- final %>%
  group_by(game_id) %>%
  summarize(W_score = max(score),
            L_score = min(score),
            W_ft = max(ft_pts),
            L_ft = min(ft_pts)) %>%
  mutate(diff_score = W_score - L_score,
         diff_ft = W_ft - L_ft)

analysis %>%
  arrange(desc(diff_score)) %>%
  head(10) %>%
  kbl(booktab = TRUE) %>%
  kable_styling(latex_options = "striped")
```

game_id	W_score	L_score	W_ft	L_ft	diff_score	diff_ft
2012_14_ARI_SEA	58	0	141.66	14.84	58	126.82
2011_07_IND_NO	62	7	142.54	29.08	55	113.46
2014_13_OAK_STL	52	0	119.84	21.72	52	98.12
2019_01_BAL_MIA	59	10	144.42	31.16	49	113.26
2012_16_TEN_GB	55	7	130.72	24.04	48	106.68
2010_07_OAK_DEN	59	14	128.00	37.60	45	90.40
2011_02_KC_DET	48	3	118.98	27.34	45	91.64
2011_05_TB_SF	48	3	112.00	31.64	45	80.36
2018_01_BUF_BAL	47	3	102.98	19.10	44	83.88
2013_16_CHI_PHI	54	11	126.40	41.82	43	84.58

```
analysis %>%
  arrange(desc(diff_ft)) %>%
  head(10) %>%
  kbl(booktab = TRUE) %>%
  kable_styling(latex_options = "striped")
```

game_id	W_score	L_score	W_ft	L_ft	diff_score	diff_ft
2012_14_ARI_SEA	58	0	141.66	14.84	58	126.82
2011_07_IND_NO	62	7	142.54	29.08	55	113.46
2019_01_BAL_MIA	59	10	144.42	31.16	49	113.26
2012_16_TEN_GB	55	7	130.72	24.04	48	106.68
2012_05_BUF_SF	45	3	129.50	23.00	42	106.50
2014_09_SD_MIA	37	0	115.46	13.92	37	101.54
2013_03_NYG_CAR	38	0	112.52	11.60	38	100.92
2018_11_PHI_NO	48	7	124.52	25.12	41	99.40
2014_13_OAK_STL	52	0	119.84	21.72	52	98.12
2013_10_DAL_NO	49	17	132.82	35.46	32	97.36