# Analysis of college football teams' play schemes in 2005

## 1. Introduction

The two datasets chosen for this project are: 1) the game statistics for each team by each game, in the 2005-2006 Division I college football season. This dataset contains various variables about the team's general statistics for each game, as well as detailed offense, defense and special team data; 2) updated version of the team codes representing each team, in the 2013-2014 Division I season. This dataset contains 3 variables of the team's codes, their names, and their representative conference codes.

The datasets were acquired from Kaggle website. Link: https://www.kaggle.com/datasets/mhixon/college-football-statistics (https://www.kaggle.com/datasets/mhixon/college-football-statistics). Data in year 2005 were specifically extracted due to the interest of college football and Texas being the national champion that year!

We are going to focus on teams' offense and analyze their performances based on either rushing or passing. We will also see where the Texas team is located at among all the teams.

```
# load the data
library(readr)

game <- read_csv("C:\\Users\\Yukun\\Desktop\\game statistics_2005.csv")
```

```
## Rows: 1436 Columns: 68
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr  (1): Game Code
## dbl (67): Team Code, Rush Att, Rush Yard, Rush TD, Pass Att, Pass Comp, Pass...
##
## 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.
```

```
team <- read_csv("C:\\Users\\Yukun\\Desktop\\team code_2013.csv")
```

```
## Rows: 247 Columns: 3
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr (1): Name
## dbl (2): Team Code, Conference Code
##
## 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.
```

```
# call the tidyverse package
library(tidyverse)
```

```
## -- Attaching packages --------------------------------------- tidyverse 1.3.1 --
```

```
## v ggplot2 3.3.5     v dplyr   1.0.7
## v tibble  3.1.6     v stringr 1.4.0
## v tidyr   1.2.0     v forcats 0.5.1
## v purrr   0.3.4
```

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

```
# check the variables of datasets with glimpse()
glimpse(game)
```

```
## Rows: 1,436
## Columns: 68
## $ `Team Code`            <dbl> 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 8, 8, 8~
## $ `Game Code`            <chr> "0005055920050910", "0005041920050917", "05030~
## $ `Rush Att`             <dbl> 21, 39, 32, 24, 36, 33, 22, 42, 28, 44, 46, 32~
## $ `Rush Yard`            <dbl> 23, 102, 119, 90, 79, 160, 44, 153, 84, 228, 1~
## $ `Rush TD`              <dbl> 0, 0, 1, 0, 1, 2, 0, 2, 0, 2, 1, 3, 0, 2, 2, 2~
## $ `Pass Att`             <dbl> 46, 43, 39, 57, 33, 44, 45, 29, 44, 32, 18, 52~
## $ `Pass Comp`            <dbl> 26, 23, 20, 30, 12, 22, 21, 15, 22, 22, 7, 30,~
## $ `Pass Yard`            <dbl> 362, 319, 406, 270, 145, 217, 188, 205, 285, 2~
## $ `Pass TD`              <dbl> 3, 2, 5, 2, 0, 1, 0, 1, 2, 1, 2, 1, 4, 1, 2, 2~
## $ `Pass Int`             <dbl> 0, 3, 0, 3, 0, 2, 1, 0, 1, 1, 0, 1, 0, 1, 1, 0~
## $ `Pass Conv`            <dbl> 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0~
## $ `Kickoff Ret`          <dbl> 5, 2, 5, 5, 1, 2, 3, 2, 6, 1, 0, 3, 2, 3, 2, 0~
## $ `Kickoff Ret Yard`     <dbl> 144, 26, 95, 129, 15, 29, 26, 41, 132, 22, 0, ~
## $ `Kickoff Ret TD`       <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0~
## $ `Punt Ret`             <dbl> 2, 4, 5, 3, 3, 0, 1, 2, 1, 4, 2, 0, 0, 3, 4, 0~
## $ `Punt Ret Yard`        <dbl> -2, 16, 76, 6, 59, 0, 0, -11, 0, 56, 17, 0, 0,~
## $ `Punt Ret TD`          <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0~
## $ `Fum Ret`              <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0~
## $ `Fum Ret Yard`         <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 16, 0, 0, 0, 0, ~
## $ `Fum Ret TD`           <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0~
## $ `Int Ret`              <dbl> 0, 0, 1, 1, 1, 1, 0, 1, 0, 1, 1, 0, 0, 0, 1, 1~
## $ `Int Ret Yard`         <dbl> 0, 0, 29, 0, 0, 9, 0, 14, 0, 18, 32, 0, 0, 0, ~
## $ `Int Ret TD`           <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0~
## $ `Misc Ret`             <dbl> 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0~
## $ `Misc Ret Yard`        <dbl> 0, 0, 0, 0, 0, 13, 0, 0, 0, 0, 0, 7, 0, 0, 0, ~
## $ `Misc Ret TD`          <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0~
## $ `Field Goal Att`       <dbl> 1, 1, 2, 2, 4, 1, 0, 1, 1, 4, 0, 2, 1, 2, 2, 3~
## $ `Field Goal Made`      <dbl> 1, 1, 2, 1, 2, 1, 0, 1, 1, 2, 0, 1, 1, 1, 1, 3~
## $ `Off XP Kick Att`      <dbl> 3, 2, 4, 2, 1, 3, 0, 3, 2, 3, 5, 4, 4, 3, 4, 4~
## $ `Off XP Kick Made`     <dbl> 3, 2, 4, 2, 1, 2, 0, 3, 2, 3, 5, 4, 4, 3, 3, 4~
## $ `Off 2XP Att`          <dbl> 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0~
## $ `Off 2XP Made`         <dbl> 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0~
## $ `Def 2XP Att`          <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0~
## $ `Def 2XP Made`         <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0~
## $ Safety                 <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0~
## $ Points                 <dbl> 24, 17, 48, 17, 13, 23, 0, 24, 17, 27, 35, 31,~
## $ Punt                   <dbl> 7, 4, 7, 8, 8, 5, 8, 7, 8, 3, 5, 5, 8, 3, 5, 3~
```

```
## $ `Punt Yard`             <dbl> 260, 126, 283, 309, 327, 185, 299, 282, 318, 8~
## $ Kickoff                 <dbl> 5, 4, 8, 4, 4, 5, 1, 5, 4, 6, 6, 6, 4, 5, 6, 8~
## $ `Kickoff Yard`          <dbl> 228, 260, 502, 253, 164, 325, 61, 269, 165, 38~
## $ `Kickoff Touchback`     <dbl> 0, 2, 1, 0, 1, 4, 0, 1, 0, 2, 3, 0, 0, 2, 3, 3~
## $ `Kickoff Out-Of-Bounds` <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1~
## $ `Kickoff Onside`        <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 2, 0, 0, 0~
## $ Fumble                  <dbl> 4, 0, 0, 2, 1, 5, 1, 1, 0, 3, 0, 1, 0, 2, 1, 1~
## $ `Fumble Lost`           <dbl> 2, 0, 0, 1, 1, 4, 0, 0, 0, 1, 0, 1, 0, 0, 1, 0~
## $ `Tackle Solo`           <dbl> 59, 39, 53, 45, 37, 30, 55, 43, 60, 32, 38, 42~
## $ `Tackle Assist`         <dbl> 18, 14, 26, 26, 30, 52, 16, 38, 27, 10, 24, 38~
## $ `Tackle For Loss`       <dbl> 2, 6, 6, 3, 10, 7, 8, 11, 11, 12, 10, 8, 7, 9,~
## $ `Tackle For Loss Yard`  <dbl> 4, 22, 22, 8, 57, 18, 24, 28, 37, 44, 30, 29, ~
## $ Sack                    <dbl> 0.0, 3.0, 5.0, 0.0, 5.0, 1.0, 1.0, 3.0, 5.5, 4~
## $ `Sack Yard`             <dbl> 0, 15, 12, 0, 45, 9, 3, 14, 24, 22, 5, 17, 0, ~
## $ `QB Hurry`              <dbl> 0, 0, 0, 0, 0, 3, 0, 1, 1, 0, 0, 0, 0, 3, 4, 3~
## $ `Fumble Forced`         <dbl> 1, 2, 1, 0, 1, 1, 0, 1, 0, 1, 2, 0, 1, 0, 0, 3~
## $ `Pass Broken Up`        <dbl> 2, 2, 3, 4, 4, 4, 3, 3, 1, 3, 1, 4, 3, 3, 3, 8~
## $ `Kick/Punt Blocked`     <dbl> 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 1, 1, 0, 0, 0, 0~
## $ `1st Down Rush`         <dbl> 4, 6, 6, 7, 4, 7, 3, 9, 4, 12, 7, 7, 3, 9, 8, ~
## $ `1st Down Pass`         <dbl> 12, 16, 14, 12, 5, 11, 10, 9, 15, 9, 3, 20, 18~
## $ `1st Down Penalty`      <dbl> 1, 4, 0, 3, 0, 3, 0, 2, 1, 2, 1, 0, 3, 1, 4, 0~
## $ `Time Of Possession`    <dbl> 1433, 2154, 1764, 1908, 1657, 1711, 1594, 1927~
## $ Penalty                 <dbl> 5, 8, 5, 5, 9, 10, 2, 8, 9, 10, 7, 8, 10, 7, 1~
## $ `Penalty Yard`          <dbl> 42, 49, 36, 50, 89, 108, 10, 84, 80, 85, 52, 7~
## $ `Third Down Att`        <dbl> 13, 14, 15, 20, 18, 17, 17, 16, 17, 17, 17, 14~
## $ `Third Down Conv`       <dbl> 2, 7, 6, 7, 3, 5, 5, 7, 6, 6, 5, 6, 8, 4, 6, 1~
## $ `Fourth Down Att`       <dbl> 3, 2, 0, 1, 3, 2, 3, 1, 2, 2, 6, 1, 1, 1, 2, 1~
## $ `Fourth Down Conv`      <dbl> 0, 0, 0, 1, 1, 1, 1, 1, 1, 2, 2, 1, 0, 0, 2, 1~
## $ `Red Zone Att`          <dbl> 2, 5, 4, 2, 5, 5, 0, 4, 3, 6, 5, 3, 2, 5, 6, 3~
## $ `Red Zone TD`           <dbl> 1, 2, 3, 0, 1, 3, 0, 3, 1, 2, 3, 3, 2, 3, 3, 1~
## $ `Red Zone Field Goal`   <dbl> 1, 0, 1, 1, 2, 1, 0, 1, 1, 2, 0, 0, 0, 1, 1, 2~
```

```
glimpse(team)
```

```
## Rows: 247
## Columns: 3
## $ `Team Code`       <dbl> 5, 8, 9, 28, 29, 30, 31, 37, 47, 51, 66, 67, 71, 77,~
## $ Name              <chr> "Akron", "Alabama", "UAB", "Arizona State", "Arizona~
## $ `Conference Code` <dbl> 875, 911, 24312, 905, 905, 818, 911, 911, 875, 25354~
```

# 2. Tidying

The two datasets are already tidy since each variable has its own column, each observation has its own row, and each value has its own cell. We will check whether we need to deal with missing data below, and tidyr functions will be used later.

```
# check if there are any missing values in the datasets
sum(is.na(game))
```

```
## [1] 0
```

```
sum(is.na(team))
```

```
## [1] 0
```

No missing values are found for the two datasets.

# 3. Joining/Merging

In this part, we will apply different join functions in this part to the two datasets.

The difference in common IDs should represent: 1) teams that joined Division I during the 2005-2013 period; 2) teams that were in Division I in 2005, but did not play a single game during that year, thus no statistics included.

```
# apply left_join() to the 'team code' dataset and store it in a new dataset
game_new <- team %>% left_join(game, by = "Team Code")
game_new
```

```
## # A tibble: 1,517 x 70
##    `Team Code` Name  `Conference Code` `Game Code`      `Rush Att` `Rush Yard`
##          <dbl> <chr>             <dbl> <chr>                <dbl>       <dbl>
##  1           5 Akron               875 0005055920050910        21          23
##  2           5 Akron               875 0005041920050917        39         102
##  3           5 Akron               875 0503000520050924        32         119
##  4           5 Akron               875 0129000520051001        24          90
##  5           5 Akron               875 0005008620051008        36          79
##  6           5 Akron               875 0005041420051015        33         160
##  7           5 Akron               875 0725000520051022        22          44
##  8           5 Akron               875 0005007120051029        42         153
##  9           5 Akron               875 0005004720051105        28          84
## 10           5 Akron               875 0519000520051115        44         228
## # ... with 1,507 more rows, and 64 more variables: `Rush TD` <dbl>,
## #   `Pass Att` <dbl>, `Pass Comp` <dbl>, `Pass Yard` <dbl>, `Pass TD` <dbl>,
## #   `Pass Int` <dbl>, `Pass Conv` <dbl>, `Kickoff Ret` <dbl>,
## #   `Kickoff Ret Yard` <dbl>, `Kickoff Ret TD` <dbl>, `Punt Ret` <dbl>,
## #   `Punt Ret Yard` <dbl>, `Punt Ret TD` <dbl>, `Fum Ret` <dbl>,
## #   `Fum Ret Yard` <dbl>, `Fum Ret TD` <dbl>, `Int Ret` <dbl>,
## #   `Int Ret Yard` <dbl>, `Int Ret TD` <dbl>, `Misc Ret` <dbl>, ...
```

```
# find the number of total observations for each dataset
nrow(team)
```

```
## [1] 247
```

```
nrow(game)
```

```
## [1] 1436
```

```
nrow(game_new)
```

```
## [1] 1517
```

```
# find the number of unique IDs for each dataset
length(unique(team[["Team Code"]]))
```

```
## [1] 247
```

```
length(unique(game[["Team Code"]]))
```

```
## [1] 166
```

```
length(unique(game_new[["Team Code"]]))
```

```
## [1] 247
```

```
# find the number of IDs in common for 'team' and 'game' datasets
nrow(semi_join(team, game, by = "Team Code"))
```

```
## [1] 166
```

```
# find the number IDs that appear in 'team' dataset but not 'game' dataset
nrow(anti_join(team, game, by = "Team Code"))
```

```
## [1] 81
```

81 observations were added when joining the two datasets. This means there were 81 teams that either joined Division I during the 2005-2013 period or did not play a single game in 2005.

# 4. Wrangling

In this part, we will explore the data with six core dplyr functions

```
# take a look at the Texas team statistics with filter()
game_new %>%
  filter(Name == "Texas")
```

```
## # A tibble: 13 x 70
##    `Team Code` Name  `Conference Code` `Game Code`      `Rush Att` `Rush Yard`
##          <dbl> <chr>             <dbl> <chr>                <dbl>       <dbl>
##  1         703 Texas             25354 0671070320050903        52         418
##  2         703 Texas             25354 0703051820050910        38         112
##  3         703 Texas             25354 0574070320050917        47         361
##  4         703 Texas             25354 0703043420051001        50         349
##  5         703 Texas             25354 0522070320051008        40         203
##  6         703 Texas             25354 0157070320051015        47         145
##  7         703 Texas             25354 0700070320051022        40         205
##  8         703 Texas             25354 0703052120051029        49         367
##  9         703 Texas             25354 0703005120051105        54         347
## 10         703 Texas             25354 0328070320051112        53         336
## 11         703 Texas             25354 0703069720051125        42         174
## 12         703 Texas             25354 0703015720051203        57         268
## 13         703 Texas             25354 0703065720060104        36         289
## # ... with 64 more variables: `Rush TD` <dbl>, `Pass Att` <dbl>,
## #   `Pass Comp` <dbl>, `Pass Yard` <dbl>, `Pass TD` <dbl>, `Pass Int` <dbl>,
## #   `Pass Conv` <dbl>, `Kickoff Ret` <dbl>, `Kickoff Ret Yard` <dbl>,
## #   `Kickoff Ret TD` <dbl>, `Punt Ret` <dbl>, `Punt Ret Yard` <dbl>,
## #   `Punt Ret TD` <dbl>, `Fum Ret` <dbl>, `Fum Ret Yard` <dbl>,
## #   `Fum Ret TD` <dbl>, `Int Ret` <dbl>, `Int Ret Yard` <dbl>,
## #   `Int Ret TD` <dbl>, `Misc Ret` <dbl>, `Misc Ret Yard` <dbl>, ...
```

```
# select certain columns on offense for further analysis , with select()
game_offense <- game_new %>%
  select(Name, `Rush Att`, `Rush Yard`, `Pass Att`, `Pass Comp`, `Pass Yard`)

game_offense
```

```
## # A tibble: 1,517 x 6
##    Name  `Rush Att` `Rush Yard` `Pass Att` `Pass Comp` `Pass Yard`
##    <chr>      <dbl>       <dbl>      <dbl>       <dbl>       <dbl>
##  1 Akron         21          23         46          26         362
##  2 Akron         39         102         43          23         319
##  3 Akron         32         119         39          20         406
##  4 Akron         24          90         57          30         270
##  5 Akron         36          79         33          12         145
##  6 Akron         33         160         44          22         217
##  7 Akron         22          44         45          21         188
##  8 Akron         42         153         29          15         205
##  9 Akron         28          84         44          22         285
## 10 Akron         44         228         32          22         270
## # ... with 1,507 more rows
```

```
# create summary statistics for each team, with group_by() and summarize(), and drop rows with NA values
game_offense <- game_offense %>%
  group_by(Name) %>%
  summarize(`Total Rush Att` = sum(`Rush Att`),
            `Total Rush Yard` = sum(`Rush Yard`),
            `Total Pass Att` = sum(`Pass Att`),
            `Total Pass Comp` = sum(`Pass Comp`),
            `Total Pass Yard` = sum(`Pass Yard`)) %>%
  drop_na()

game_offense
```

```
## # A tibble: 166 x 6
##    Name      `Total Rush Att` `Total Rush Ya~` `Total Pass Att` `Total Pass Co~`
##    <chr>                <dbl>            <dbl>            <dbl>            <dbl>
##  1 Air Force              588             2712              205              121
##  2 Akron                  422             1395              541              284
##  3 Alabama                463             1710              351              209
##  4 Appalach~               69              309               57               32
##  5 Arizona                382             1342              368              211
##  6 Arizona ~              447             1748              493              312
##  7 Arkansas               481             2386              280              150
##  8 Arkansas~              481             2318              299              162
##  9 Army                   472             1543              312              174
## 10 Auburn                 481             2329              339              195
## # ... with 156 more rows, and 1 more variable: `Total Pass Yard` <dbl>
```

```
# calculate each team's yard/rush and yard/pass, and create a categorical variable 'Play Dominance', all with mutate()
# 'Play Dominance' will be defined by comparing total rush attempts and total pass attempts
game_offense <- game_offense %>%
  mutate(`Yard per Rush Att` = `Total Rush Yard`/`Total Rush Att`,
         `Yard per Pass Att` = `Total Pass Yard`/`Total Pass Comp`,
         `Play Dominance` = case_when(`Total Rush Att` > `Total Pass Att` ~ "run dominant",
                                      `Total Rush Att` < `Total Pass Att` ~ "pass dominant",
                                      `Total Rush Att` == `Total Pass Att` ~ "equal"))

game_offense
```

```
## # A tibble: 166 x 9
##    Name      `Total Rush Att` `Total Rush Ya~` `Total Pass Att` `Total Pass Co~`
##    <chr>                <dbl>            <dbl>            <dbl>            <dbl>
##  1 Air Force              588             2712              205              121
##  2 Akron                  422             1395              541              284
##  3 Alabama                463             1710              351              209
##  4 Appalach~               69              309               57               32
##  5 Arizona                382             1342              368              211
##  6 Arizona ~              447             1748              493              312
##  7 Arkansas               481             2386              280              150
##  8 Arkansas~              481             2318              299              162
##  9 Army                   472             1543              312              174
## 10 Auburn                 481             2329              339              195
## # ... with 156 more rows, and 4 more variables: `Total Pass Yard` <dbl>,
## #   `Yard per Rush Att` <dbl>, `Yard per Pass Att` <dbl>,
## #   `Play Dominance` <chr>
```

```
# call the tidyverse package again
library(tidyverse)

# arrange the dataset based on yard/rush
game_offense %>%
  arrange(desc(`Yard per Rush Att`))
```

```
## # A tibble: 166 x 9
##    Name      `Total Rush Att` `Total Rush Ya~` `Total Pass Att` `Total Pass Co~`
##    <chr>                <dbl>            <dbl>            <dbl>            <dbl>
##  1 USC                    525             3380              481              312
##  2 Nicholls~               65              408               11                4
##  3 Texas                  605             3574              336              218
##  4 Californ~              483             2823              321              167
##  5 Texas A&M              452             2584              321              166
##  6 Navy                   672             3832              147               69
##  7 Memphis                597             3215              239              135
##  8 Minnesota              610             3277              323              185
##  9 Toledo                 490             2602              378              241
## 10 Washingt~              440             2332              381              216
## # ... with 156 more rows, and 4 more variables: `Total Pass Yard` <dbl>,
## #   `Yard per Rush Att` <dbl>, `Yard per Pass Att` <dbl>,
## #   `Play Dominance` <chr>
```

```
# arrange the dataset based on yard/pass
game_offense %>%
  arrange(desc(`Yard per Pass Att`))
```

```
## # A tibble: 166 x 9
##    Name      `Total Rush Att` `Total Rush Ya~` `Total Pass Att` `Total Pass Co~`
##    <chr>                <dbl>            <dbl>            <dbl>            <dbl>
##  1 Illinois~               34               89               25               10
##  2 Navy                   672             3832              147               69
##  3 Eastern ~               31               48               37               10
##  4 Eastern ~               32               88               40               23
##  5 Southern~               35              117               30               15
##  6 Richmond                24               89               31               16
##  7 Air Force              588             2712              205              121
##  8 Wisconsin              572             2186              337              200
##  9 Georgia                455             2108              363              201
## 10 Texas St~               30              115               36               26
## # ... with 156 more rows, and 4 more variables: `Total Pass Yard` <dbl>,
## #   `Yard per Rush Att` <dbl>, `Yard per Pass Att` <dbl>,
## #   `Play Dominance` <chr>
```

```
# count and caculate the frequency for different 'Play Dominance' types
# pivot function is used here for tidier table
game_pivot <- game_offense %>%
  group_by(`Play Dominance`) %>%
  summarize(count = n()) %>%
  pivot_wider(names_from = `Play Dominance`,
              values_from = count)
```

```
# install and call the kableExtra package
install.packages("kableExtra", repos="http://cran.us.r-project.org")
```

```
## Installing package into 'C:/Users/Yukun/Documents/R/win-library/4.1'
## (as 'lib' is unspecified)
```

```
## package 'kableExtra' successfully unpacked and MD5 sums checked
##
## The downloaded binary packages are in
##  C:\Users\Yukun\AppData\Local\Temp\Rtmpa6i6aD\downloaded_packages
```

```
library(kableExtra)
```

```
## Warning: package 'kableExtra' was built under R version 4.1.3
```

```
##
## Attaching package: 'kableExtra'
```

```
## The following object is masked from 'package:dplyr':
##
##     group_rows
```

```
# style the table
game_pivot %>%
  kable(caption = "count of play_dominance types") %>%
  kable_classic(full_width = F, html_font = "Cambria")
```

count of play_dominance types

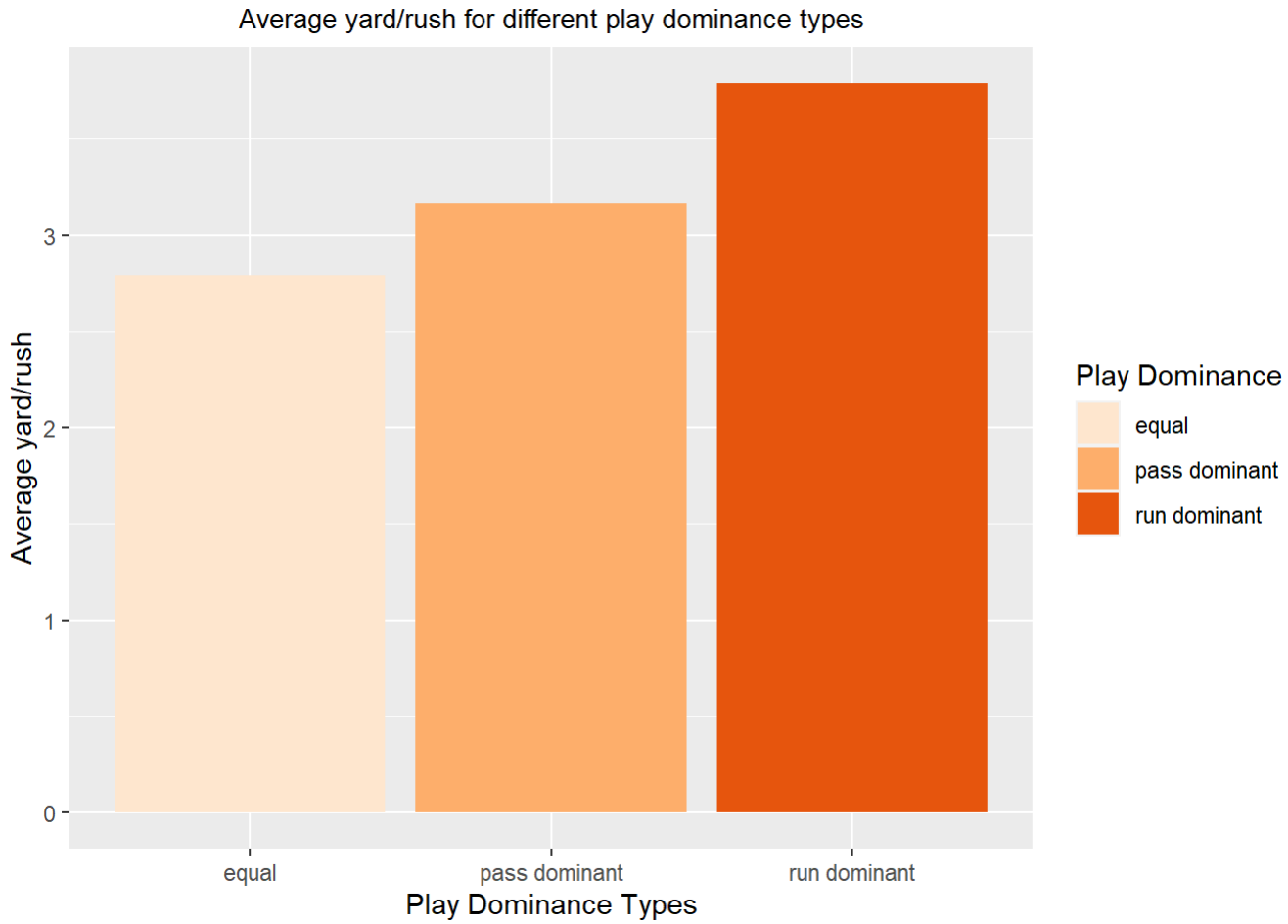
| equal | pass dominant | run dominant |
|---|---|---|
| 2 | 45 | 119 |

The results show that most of the teams in 2005 were predominantly using run plays instead of pass plays. Texas, as the champion that year, was pretty effective on both rushing and passing, ranking 3rd on yard/rush and 16th on yard/pass among all Division I teams.

# 5. Creating Visualizations

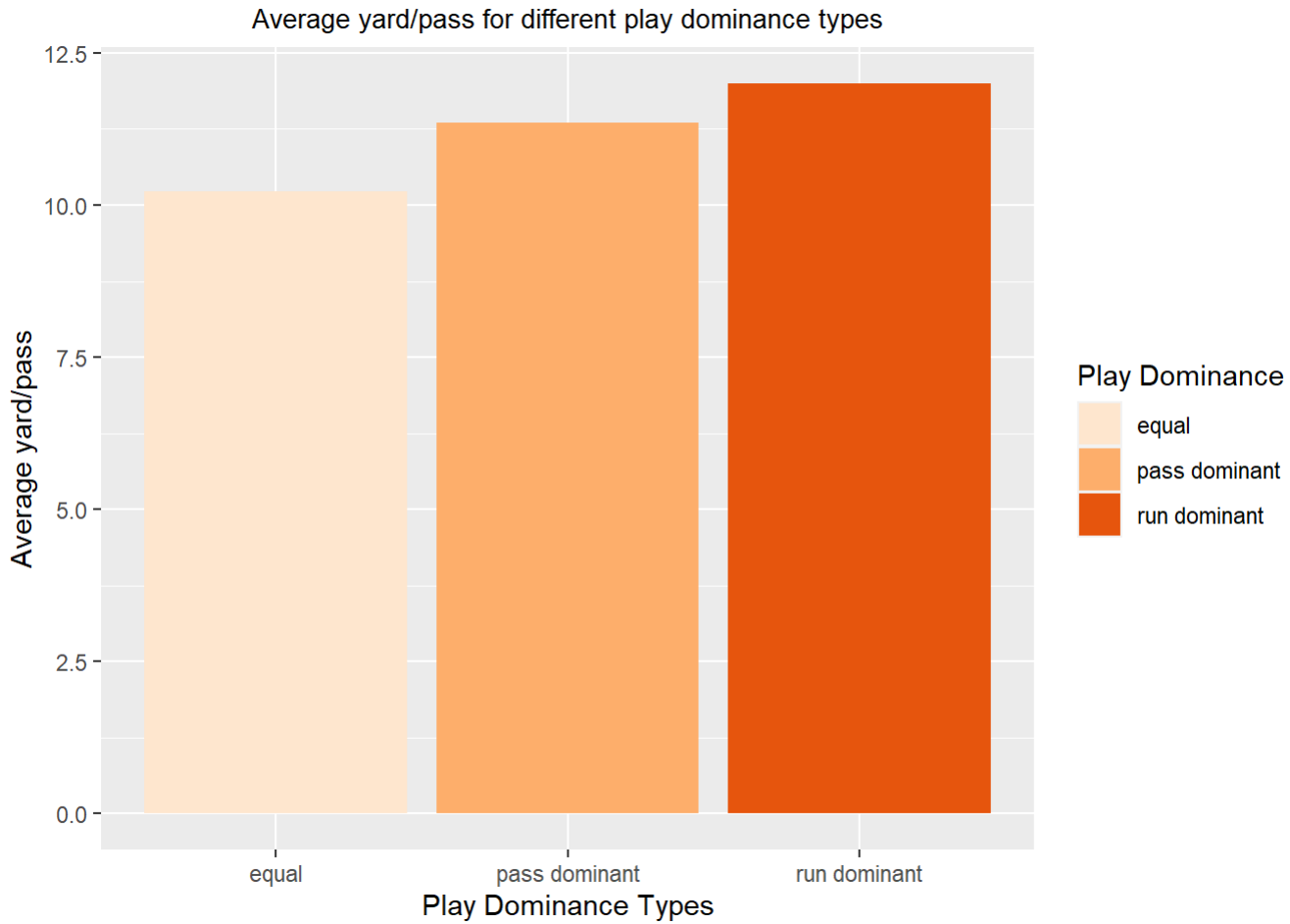
In this part, we will create 3 ggplots analyzing the relationship among yard/rush, yard/pass and different play dominance types.

```
# create a bar graph of average yard/rush for different play dominance types
# label the axes, add a title, and change the theme and scales
ggplot(game_offense, aes(x = `Play Dominance`, y = `Yard per Rush Att`, fill = `Play Dominance`)) +
  geom_bar(stat = "summary", fun = "mean") +
  labs(title = "Average yard/rush for different play dominance types",
       x = "Play Dominance Types",
       y = "Average yard/rush") +
  theme(plot.title = element_text(size = 10, hjust = 0.5)) +
  scale_fill_brewer(palette = "Oranges")
```



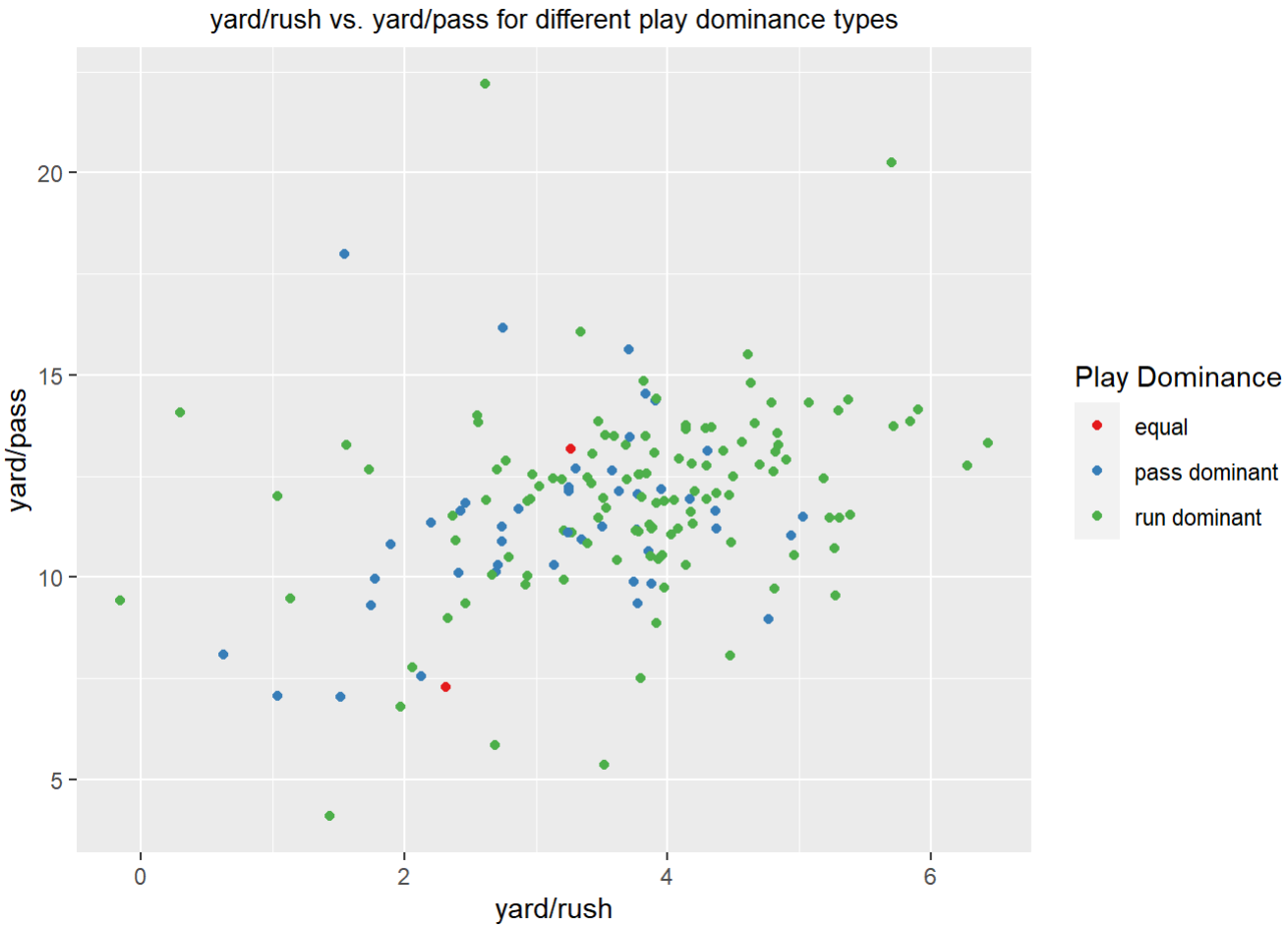

The bar graph shows that in 2005, run-dominant-teams have the highest average yard/rush in total, following pass dominant and equal.

```
# create a bar graph of average yard/pass for different play dominance types
# label the axes, add a title, and change the theme and scales
ggplot(game_offense, aes(x = `Play Dominance`, y = `Yard per Pass Att`, fill = `Play Dominance`)) +
  geom_bar(stat = "summary", fun = "mean") +
  labs(title = "Average yard/pass for different play dominance types",
       x = "Play Dominance Types",
       y = "Average yard/pass") +
  theme(plot.title = element_text(size = 10, hjust = 0.5)) +
  scale_fill_brewer(palette = "Oranges")
```




Opposing to the last graph, this bar graph shows that in 2005, pass-dominant-teams do not have the highest average yard/pass in total. Instead they are second and pass-dominant-teams still are the highest, with equal the third.

```
# create a scatter plot of yard/rush vs. yard/pass for different play dominance types
# label the axes, add a title, and change the theme and scales
ggplot(game_offense, aes(x = `Yard per Rush Att`, y = `Yard per Pass Att`)) +
  geom_point(aes(color = `Play Dominance`)) +
  labs(title = "yard/rush vs. yard/pass for different play dominance types",
       x = "yard/rush",
       y = "yard/pass") +
  theme(plot.title = element_text(size = 10, hjust = 0.5)) +
  scale_colour_brewer(palette = "Set1")
```




The scatter plot shows that when comparing yard/rush and yard/pass, there is no significant difference among teams with different play dominance types. There is also a general trend for all teams: Teams with lower yard/rush tend to have lower yard/pass, and teams with higher yard/rush also tend to have higher yard/pass.