

ML Assignment 1

Sean Bradford

1. & 2.

```
nfl_data=read.csv('C:\\Users\\Sean\\OneDrive\\Desktop\\Grad School\\Machine Learning\\Module 2 - Intro
```

```
# Data Source: https://www.advancedsportsanalytics.com/nfl-raw-data
```

```
head(nfl_data)
```

```
##      game_id player_id pos      player team pass_cmp pass_att
## 1 201909050chi  RodgAa00 QB      Aaron Rodgers GNB      18      30
## 2 201909050chi  JoneAa00 RB      Aaron Jones GNB       0       0
## 3 201909050chi  ValdMa00 WR Marquez Valdes-Scantling GNB       0       0
## 4 201909050chi  AdamDa01 WR      Davante Adams GNB       0       0
## 5 201909050chi  GrahJi00 TE      Jimmy Graham GNB       0       0
## 6 201909050chi  DaviTr03 WR      Trevor Davis GNB       0       0
##      pass_yds pass_td pass_int pass_sacked pass_sacked_yds pass_long pass_rating
## 1         203         1         0          5          37         47         91.4
## 2          0         0         0          0           0          0          0.0
## 3          0         0         0          0           0          0          0.0
## 4          0         0         0          0           0          0          0.0
## 5          0         0         0          0           0          0          0.0
## 6          0         0         0          0           0          0          0.0
##      rush_att rush_yds rush_td rush_long targets rec rec_yds rec_td rec_long
## 1          3         8         0         10         0  0         0         0         0
## 2         13        39         0          9         1  1          0         0         0
## 3          1         0         0          0         6  4         52         0         47
## 4          0         0         0          0         8  4         36         0         11
## 5          0         0         0          0         5  3         30         1         16
## 6          0         0         0          0         1  1         28         0         28
##      fumbles_lost rush_scrambles designed_rush_att comb_pass_rush_play
## 1          0         1         2         38
## 2          0         0         13         13
## 3          0         0         1          1
## 4          0         0         0          0
## 5          0         0         0          0
## 6          0         0         0          0
##      comb_pass_play comb_rush_play Team_abbrev Opponent_abbrev two_point_conv
## 1          36         2      GNB      CHI          0
## 2           0        13      GNB      CHI          0
## 3           0         1      GNB      CHI          0
## 4           0         0      GNB      CHI          0
## 5           0         0      GNB      CHI          0
## 6           0         0      GNB      CHI          0
##      total_ret_td offensive_fumble_recovery_td pass_yds_bonus rush_yds_bonus
## 1          0          0          0          0
```

```

## 2          0          0          0          0
## 3          0          0          0          0
## 4          0          0          0          0
## 5          0          0          0          0
## 6          0          0          0          0
##   rec_yds_bonus Total_DKP Off_DKP Total_FDP Off_FDP Total_SDP Off_SDP
## 1          0      12.92  12.92      12.92  12.92      12.92  12.92
## 2          0       4.90   4.90       4.40   4.40       4.40   4.40
## 3          0       9.20   9.20       7.20   7.20       7.20   7.20
## 4          0       7.60   7.60       5.60   5.60       5.60   5.60
## 5          0      12.00  12.00      10.50  10.50      10.50  10.50
## 6          0       3.80   3.80       3.30   3.30       3.30   3.30
##   pass_target_yds pass_poor_throws pass_blitzed pass_hurried
## 1             300             5             8             6
## 2              0             0             0             0
## 3              0             0             0             0
## 4              0             0             0             0
## 5              0             0             0             0
## 6              0             0             0             0
##   rush_yds_before_contact rush_yac rush_broken_tackles rec_air_yds rec_yac
## 1                   6         2                   0         0.0         0
## 2                  21        18                   1        -1.0         1
## 3                   0         0                   0        81.0         2
## 4                   0         0                   0        63.2        19
## 5                   0         0                   0        58.0        17
## 6                   0         0                   0        18.0        10
##   rec_drops offense off_pct vis_team home_team vis_score home_score   OT
## 1          0      61      100     GNB      CHI         10         3 FALSE
## 2          0      37      61     GNB      CHI         10         3 FALSE
## 3          0      41      67     GNB      CHI         10         3 FALSE
## 4          0      59      97     GNB      CHI         10         3 FALSE
## 5          0      33      54     GNB      CHI         10         3 FALSE
## 6          0      17      28     GNB      CHI         10         3 FALSE
##   Roof Surface Temperature Humidity Wind_Speed Vegas_Line Vegas_Favorite
## 1 outdoors  grass         65        69         10        -3.5          CHI
## 2 outdoors  grass         65        69         10        -3.5          CHI
## 3 outdoors  grass         65        69         10        -3.5          CHI
## 4 outdoors  grass         65        69         10        -3.5          CHI
## 5 outdoors  grass         65        69         10        -3.5          CHI
## 6 outdoors  grass         65        69         10        -3.5          CHI
##   Over_Under game_date
## 1          47  9/5/2019
## 2          47  9/5/2019
## 3          47  9/5/2019
## 4          47  9/5/2019
## 5          47  9/5/2019
## 6          47  9/5/2019

```

The dataset contains statistics for passing, rushing, and receiving per game for NFL players since 2019.

3.

```
summary(nfl_data[,c("player", "team", "pass_yds", "pass_td")])
```

```
##      player      team      pass_yds      pass_td
## Length:19973   Length:19973   Min.   : -2.00   Min.   :0.0000
## Class :character Class :character 1st Qu.:  0.00   1st Qu.:0.0000
## Mode  :character Mode  :character Median :  0.00   Median :0.0000
##                                     Mean  : 20.57   Mean  :0.1317
##                                     3rd Qu.:  0.00   3rd Qu.:0.0000
##                                     Max.   :525.00   Max.   :5.0000
```

Qualitative Variables: 'player' & 'team' Quantitative Variables: 'pass_yds' & 'pass_td'

4.

```
library(dplyr)
```

```
##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':
##
##      filter, lag

## The following objects are masked from 'package:base':
##
##      intersect, setdiff, setequal, union
```

```
nfl_data=nfl_data%>%mutate(pass_cmp=(pass_cmp/pass_att))
head(nfl_data)
```

```
##      game_id player_id pos      player team pass_cmp pass_att
## 1 201909050chi  RodgAa00 QB      Aaron Rodgers GNB      0.6      30
## 2 201909050chi  JoneAa00 RB      Aaron Jones GNB      NaN      0
## 3 201909050chi  ValdMa00 WR Marquez Valdes-Scantling GNB      NaN      0
## 4 201909050chi  AdamDa01 WR      Davante Adams GNB      NaN      0
## 5 201909050chi  GrahJi00 TE      Jimmy Graham GNB      NaN      0
## 6 201909050chi  DaviTr03 WR      Trevor Davis GNB      NaN      0
##      pass_yds pass_td pass_int pass_sacked pass_sacked_yds pass_long pass_rating
## 1      203      1      0      5      37      47      91.4
## 2      0      0      0      0      0      0      0.0
## 3      0      0      0      0      0      0      0.0
## 4      0      0      0      0      0      0      0.0
## 5      0      0      0      0      0      0      0.0
## 6      0      0      0      0      0      0      0.0
##      rush_att rush_yds rush_td rush_long targets rec rec_yds rec_td rec_long
## 1      3      8      0      10      0  0      0      0      0
## 2     13     39      0      9      1  1      0      0      0
## 3      1      0      0      0      6  4     52      0     47
## 4      0      0      0      0      8  4     36      0     11
## 5      0      0      0      0      5  3     30      1     16
```

## 6	0	0	0	0	1	1	28	0	28
##	fumbles_lost	rush_scrambles	designed_rush_att	comb_pass_rush_play					
## 1	0		1	2			38		
## 2	0		0	13			13		
## 3	0		0	1			1		
## 4	0		0	0			0		
## 5	0		0	0			0		
## 6	0		0	0			0		
##	comb_pass_play	comb_rush_play	Team_abbrev	Opponent_abbrev	two_point_conv				
## 1	36	2	GNB	CHI	0				
## 2	0	13	GNB	CHI	0				
## 3	0	1	GNB	CHI	0				
## 4	0	0	GNB	CHI	0				
## 5	0	0	GNB	CHI	0				
## 6	0	0	GNB	CHI	0				
##	total_ret_td	offensive_fumble_recovery_td	pass_yds_bonus	rush_yds_bonus					
## 1	0		0	0			0		
## 2	0		0	0			0		
## 3	0		0	0			0		
## 4	0		0	0			0		
## 5	0		0	0			0		
## 6	0		0	0			0		
##	rec_yds_bonus	Total_DKP	Off_DKP	Total_FDP	Off_FDP	Total_SDP	Off_SDP		
## 1	0	12.92	12.92	12.92	12.92	12.92	12.92		
## 2	0	4.90	4.90	4.40	4.40	4.40	4.40		
## 3	0	9.20	9.20	7.20	7.20	7.20	7.20		
## 4	0	7.60	7.60	5.60	5.60	5.60	5.60		
## 5	0	12.00	12.00	10.50	10.50	10.50	10.50		
## 6	0	3.80	3.80	3.30	3.30	3.30	3.30		
##	pass_target_yds	pass_poor_throws	pass_blitzed	pass_hurried					
## 1	300		5	8		6			
## 2	0		0	0		0			
## 3	0		0	0		0			
## 4	0		0	0		0			
## 5	0		0	0		0			
## 6	0		0	0		0			
##	rush_yds_before_contact	rush_yac	rush_broken_tackles	rec_air_yds	rec_yac				
## 1		6	2	0	0.0	0			
## 2		21	18	1	-1.0	1			
## 3		0	0	0	81.0	2			
## 4		0	0	0	63.2	19			
## 5		0	0	0	58.0	17			
## 6		0	0	0	18.0	10			
##	rec_drops	offense	off_pct	vis_team	home_team	vis_score	home_score	OT	
## 1	0	61	100	GNB	CHI	10	3	FALSE	
## 2	0	37	61	GNB	CHI	10	3	FALSE	
## 3	0	41	67	GNB	CHI	10	3	FALSE	
## 4	0	59	97	GNB	CHI	10	3	FALSE	
## 5	0	33	54	GNB	CHI	10	3	FALSE	
## 6	0	17	28	GNB	CHI	10	3	FALSE	
##	Roof	Surface	Temperature	Humidity	Wind_Speed	Vegas_Line	Vegas_Favorite		
## 1	outdoors	grass	65	69	10	-3.5	CHI		
## 2	outdoors	grass	65	69	10	-3.5	CHI		
## 3	outdoors	grass	65	69	10	-3.5	CHI		

```
## 4 outdoors grass 65 69 10 -3.5 CHI
## 5 outdoors grass 65 69 10 -3.5 CHI
## 6 outdoors grass 65 69 10 -3.5 CHI
## Over_Under game_date
## 1 47 9/5/2019
## 2 47 9/5/2019
## 3 47 9/5/2019
## 4 47 9/5/2019
## 5 47 9/5/2019
## 6 47 9/5/2019
```

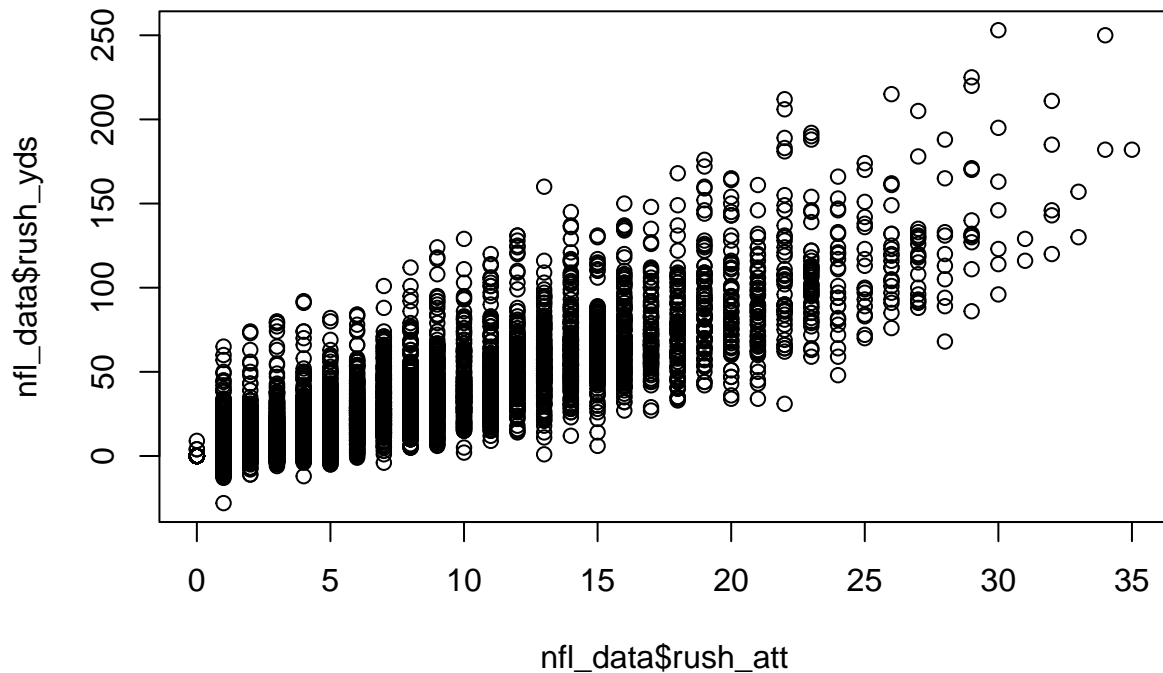
```
summary(nfl_data$pass_cmp)
```

```
## Min. 1st Qu. Median Mean 3rd Qu. Max. NA's
## 0.000 0.571 0.652 0.636 0.724 1.000 17963
```

Changed variable pass_cmp to return the ratio of passes completed instead of the total number of passes completed.

5.

```
plot(nfl_data$rush_att,nfl_data$rush_yds)
```



Created scatterplot to show relationship between rushing attempts and rushing yards.