

# Exploratory Data Analysis

*Kevin Thompson, Sean Kennedy, Sachin Chavan*

*September 29, 2019*

## 1 : Load required libraries

## 2 : Load Quarterback Crosssectional data

```
load("../analysis/data/.RData")
QBCrossSectional = getFootballData(URL) %>% filterUnwantedVariables() %>% filter(Position == 'QB')
summary(QBCrossSectional)
```

```
##      PlayerID      Name      Week      Position
## Min.   : 611   Length:453   Min.    : 1.000   Length:453
## 1st Qu.: 7242  Class :character 1st Qu.: 5.000   Class :character
## Median :13723  Mode  :character Median : 9.000   Mode  :character
## Mean   :11932                      Mean    : 9.049
## 3rd Qu.:16763                      3rd Qu.:13.000
## Max.   :19029                      Max.    :17.000
##      Opponent      TeamIsHome      GameDate      PassingCompletions
## Length:453        Mode :logical   Length:453      Min.    : 5.00
## Class :character   FALSE:226   Class :character 1st Qu.:18.00
## Mode  :character    TRUE :227   Mode  :character Median :21.00
##                                     Mean   :21.24
##                                     3rd Qu.:25.00
##                                     Max.   :44.00
##      Result      PassingAttempts PassingCompletionPercentage
## Length:453      Min.    :10.00   Min.    :38.70
## Class :character 1st Qu.:29.00   1st Qu.:57.10
## Mode  :character Median :33.00   Median :63.20
##                                     Mean   :33.62   Mean   :63.48
##                                     3rd Qu.:38.00   3rd Qu.:69.40
##                                     Max.    :66.00   Max.    :87.00
##      PassingYards PassingYardsPerAttempt PassingTouchdowns
## Min.    : 57.0   Min.    : 3.100   Min.    :0.000
## 1st Qu.:199.0   1st Qu.: 6.200   1st Qu.:1.000
## Median :241.0   Median : 7.200   Median :1.000
## Mean   :244.6   Mean   : 7.378   Mean   :1.587
## 3rd Qu.:291.0   3rd Qu.: 8.400   3rd Qu.:2.000
## Max.   :506.0   Max.   :14.100   Max.   :5.000
##      PassingInterceptions PassingRating RushingAttempts RushingYards
## Min.    :0.00      Min.    : 31.14   Min.    : 0.000   Min.    : -8.00
## 1st Qu.:0.00      1st Qu.: 77.92   1st Qu.: 1.000   1st Qu.: 0.00
## Median :0.00      Median : 92.94   Median : 3.000   Median : 8.00
## Mean   :0.66      Mean   : 93.88   Mean   : 3.185   Mean   :14.38
## 3rd Qu.:1.00      3rd Qu.:109.84   3rd Qu.: 5.000   3rd Qu.:23.00
## Max.    :4.00      Max.   :150.69   Max.   :14.000   Max.   :95.00
##      RushingYardsPerAttempt RushingTouchdowns FumblesLost FantasyPoints
## Min.    : -2.700      Min.    :0.0000   Min.    :0.0000   Min.    : 7.12
## 1st Qu.: 0.000      1st Qu.:0.0000   1st Qu.:0.0000   1st Qu.:11.86
```

```
## Median : 3.000      Median :0.0000      Median :0.0000      Median :15.86
## Mean   : 3.839      Mean    :0.1457      Mean    :0.1766      Mean    :16.89
## 3rd Qu.: 6.000      3rd Qu.:0.0000      3rd Qu.:0.0000      3rd Qu.:20.68
## Max.   :70.000      Max.    :2.0000      Max.    :3.0000      Max.    :37.64
##      Team
## Length:453
## Class :character
## Mode  :character
##
##
##
```

```
attach(QBCrossSectional)
```

```
## The following object is masked from package:ggplot2:
```

```
##
```

```
##      Position
```

Drop rushing yards per attempt

### 3 : Defensive stats

```
## 3.1 : Team Defensive Stats
```

```
team_defense = getFootballData("https://fantasydata.com/FantasyStatsNFL/FantasyStats_Read?sort=FantasyP
team_defense$StatSummary = c(NULL)
```

```
defensive_columns = c('Team', 'Week', 'TacklesForLoss', 'Sacks', 'QuarterbackHits', 'Interceptions', 'F
```

```
team_defense = team_defense %>% dplyr::select(defensive_columns) %>% rename('DefensiveFantasyPoints'='F
```

```
attach(team_defense)
```

```
## The following objects are masked from QBCrossSectional:
```

```
##
```

```
##      Team, Week
```

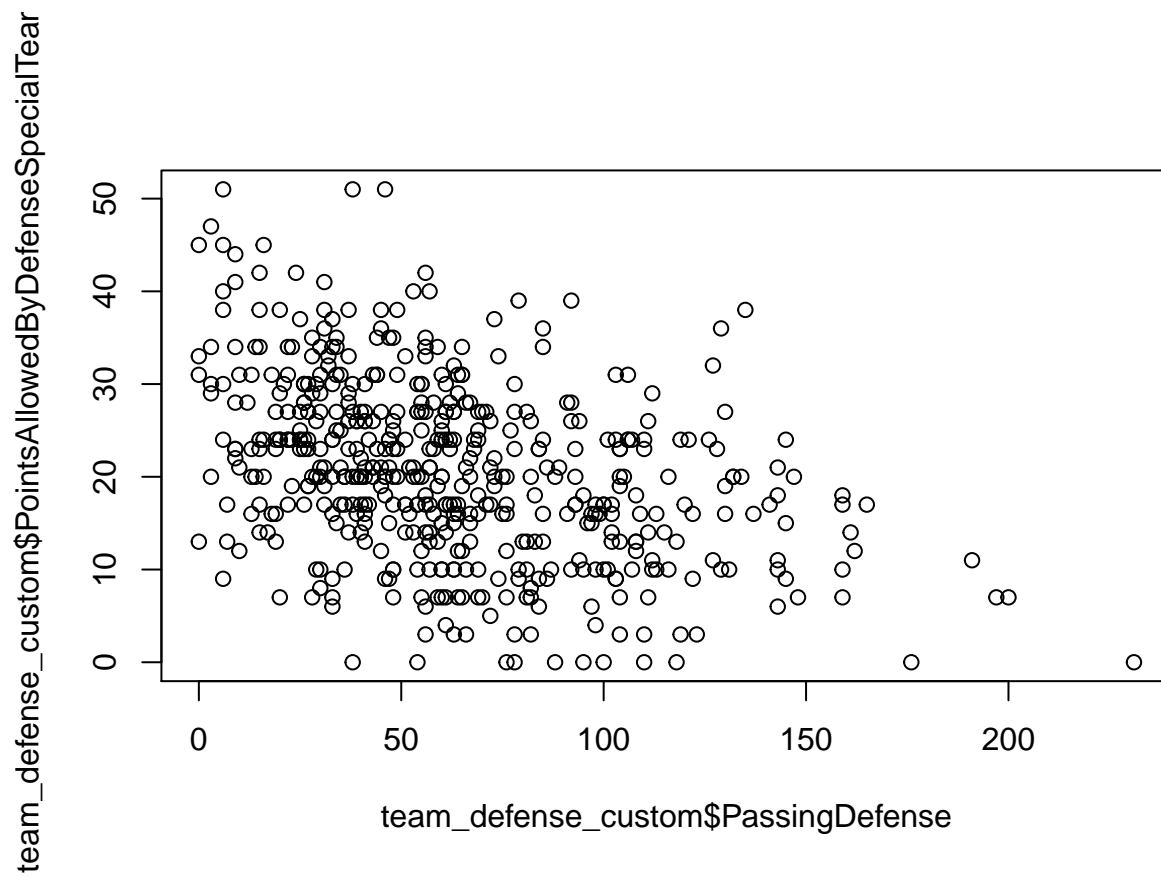
```
## 3.5 : Add defensive matchups
```

*#This only adds the current weeks matchup - we need next week's matchup as a target*

```
team_defense_custom = sqldf("SELECT Team
                             ,Week
                             ,(Sacks * 4
                             +QuarterbackHits * 3
                             +Interceptions * 7
                             +SackYards * 2) as PassingDefense
                             ,PointsAllowedByDefenseSpecialTeams
                             FROM team_defense")
```

*#scheme: weight interceptions, qbsacks, quarterbackhits, passesdefended*

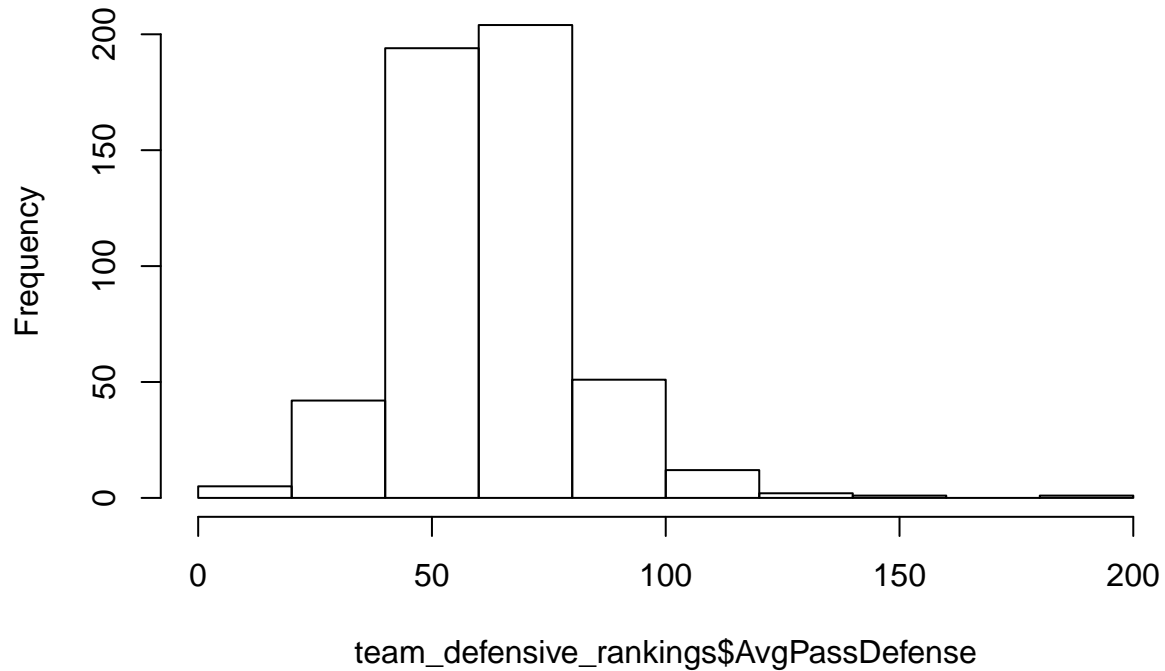
```
plot(team_defense_custom$PassingDefense, team_defense_custom$PointsAllowedByDefenseSpecialTeams)
```



```
team_defensive_rankings = team_defense_custom %>%
  #filter(Week > 1 & Week < 17)
  group_by(Team) %>%
  arrange(Week) %>%
  mutate(
    AvgPassDefense = cumsum(PassingDefense)/Week #Account for bye week....
  )
QBCrossSectional$Week = as.numeric(QBCrossSectional$Week)

hist(team_defensive_rankings$AvgPassDefense)
```

## Histogram of team\_defensive\_rankings\$AvgPassDefense



```
sqldf("SELECT * FROM team_defensive_rankings WHERE Team = 'LAC'")
```

##	Team	Week	PassingDefense	PointsAllowedByDefenseSpecialTeams
## 1	LAC	1	126	24
## 2	LAC	2	27	19
## 3	LAC	3	101	24
## 4	LAC	4	29	26
## 5	LAC	5	132	20
## 6	LAC	6	33	16
## 7	LAC	7	95	0
## 8	LAC	8	59	19
## 9	LAC	10	76	20
## 10	LAC	11	62	24
## 11	LAC	12	56	6
## 12	LAC	13	63	10
## 13	LAC	14	60	7
## 14	LAC	15	26	30
## 15	LAC	16	76	7
## 16	LAC	17	54	10
##	AvgPassDefense			
## 1			126.00000	
## 2			76.50000	
## 3			84.66667	
## 4			70.75000	
## 5			83.00000	
## 6			74.66667	
## 7			77.57143	
## 8			75.25000	
## 9			67.80000	
## 10			67.27273	

```

## 11      66.33333
## 12      66.07692
## 13      65.64286
## 14      63.00000
## 15      63.81250
## 16      63.23529

## 3.6 : Add some lag data for QB

### CumulativeVariables (these should definitley be combined into a weekly ranking)
QBCrossSectionalCumulativePassYards = QBCrossSectional %>% group_by(PlayerID)%>% filter(n() >= 8) %>% a
  mutate(CumulativeAveragePassingYards=cummean(PassingYards)
    , CumulativeAveragePassingTouchdowns=cummean(PassingTouchdowns)
    , CumulativeAveragePassingInterceptions=cummean(PassingInterceptions)
    , CumulativeAveragePassingRating=cummean(PassingRating)
    , CumulativeAverageCompletions = cummean(PassingCompletions) # not sure that completions matter
    , CumulativeAverageCompletionPercentage = cummean(PassingCompletionPercentage)
    , CumulativeMaxPassingTouchdowns = cummax(PassingTouchdowns)
    , CumulativeMaxPassingYards = cummax(PassingYards)
    , CumulativeMaxPassingAttempts = cummax(PassingAttempts)
    , CumulativeMaxPassingRating = cummax(PassingRating)
    , CumulativeMaxCompletions = cummax(PassingCompletions)
    , CumulativeMaxPassYardsPerAttempt = cummax(PassingYardsPerAttempt)
    , CumulativeMinPassingTouchdowns = cummin(PassingTouchdowns)
    , CumulativeMinPassingYards = cummin(PassingYards) #Let's get mins to capture downside risk
    , CumulativeMinPassingAttempts = cummin(PassingAttempts)
    , CumulativeMinPassingRating = cummin(PassingRating)
    , CumulativeMinCompletions = cummin(PassingCompletions)
    , CumulativeMinPassYardsPerAttempt = cummin(PassingYardsPerAttempt)
    , NextWeekFantasyPoints = lead(FantasyPoints) #Target Variable
    , NextOpponent = lead(Opponent)
    #, NextWeekDefensiveMatchup = lag(WeeklyRank) #Has to be last week's team ranking - not this we
  )

QBCrossSectionalDefensiveOverlayCumulativePassYards = QBCrossSectionalCumulativePassYards %>% left_join

#QBCrossSectionalDefensiveOverlayCumulativePassYards %>% filter(PlayerID == 6739) %>% write.csv('alex_s
attach(QBCrossSectionalDefensiveOverlayCumulativePassYards)

## The following objects are masked from team_defense:
##
##   PointsAllowedByDefenseSpecialTeams, Team, Week

## The following objects are masked from QBCrossSectional:
##
##   FantasyPoints, FumblesLost, GameDate, Name, Opponent,
##   PassingAttempts, PassingCompletionPercentage,
##   PassingCompletions, PassingInterceptions, PassingRating,
##   PassingTouchdowns, PassingYards, PassingYardsPerAttempt,
##   PlayerID, Position, Result, RushingAttempts,
##   RushingTouchdowns, RushingYards, RushingYardsPerAttempt, Team,
##   TeamIsHome, Week

## The following object is masked from package:ggplot2:

```

##

## Position

summary(QBCrossSectionalDefensiveOverlayCumulativePassYards)

```
##      PlayerID      Name      Week      Position
## Min.   : 611   Length:368   Min.   : 1.000   Length:368
## 1st Qu.: 6739   Class :character   1st Qu.: 4.000   Class :character
## Median :13320   Mode  :character   Median : 9.000   Mode  :character
## Mean   :11765
## 3rd Qu.:16763
## Max.   :18868
##
##      Opponent      TeamIsHome      GameDate      PassingCompletions
## Length:368      Mode :logical   Length:368      Min.   : 6.00
## Class :character FALSE:184      Class :character 1st Qu.:18.00
## Mode  :character TRUE :184      Mode  :character Median :21.00
##
##                                     Mean   :21.39
##                                     3rd Qu.:25.00
##                                     Max.   :44.00
##
##      Result      PassingAttempts PassingCompletionPercentage
## Length:368      Min.   :10.00   Min.   :39.30
## Class :character 1st Qu.:29.00   1st Qu.:57.40
## Mode  :character Median :33.00   Median :63.75
##                                     Mean   :33.56   Mean   :63.99
##                                     3rd Qu.:38.00   3rd Qu.:70.00
##                                     Max.   :66.00   Max.   :87.00
##
##      PassingYards PassingYardsPerAttempt PassingTouchdowns
## Min.   : 69.0   Min.   : 3.100   Min.   :0.000
## 1st Qu.:202.8   1st Qu.: 6.300   1st Qu.:1.000
## Median :242.0   Median : 7.300   Median :1.500
## Mean   :246.8   Mean   : 7.442   Mean   :1.617
## 3rd Qu.:292.2   3rd Qu.: 8.400   3rd Qu.:2.000
## Max.   :506.0   Max.   :14.100   Max.   :5.000
##
##      PassingInterceptions PassingRating      RushingAttempts      RushingYards
## Min.   :0.0000      Min.   : 31.14   Min.   : 0.000   Min.   : -8.00
## 1st Qu.:0.0000      1st Qu.: 79.29   1st Qu.: 1.000   1st Qu.: 0.00
## Median :0.0000      Median : 94.94   Median : 3.000   Median : 8.00
## Mean   :0.6114      Mean   : 95.45   Mean   : 3.258   Mean   :14.55
## 3rd Qu.:1.0000      3rd Qu.:110.83   3rd Qu.: 5.000   3rd Qu.:24.00
## Max.   :4.0000      Max.   :150.69   Max.   :14.000   Max.   :95.00
##
##      RushingYardsPerAttempt RushingTouchdowns FumblesLost      FantasyPoints
## Min.   : -2.700      Min.   :0.0000   Min.   :0.0000   Min.   : 7.32
## 1st Qu.: 0.000      1st Qu.:0.0000   1st Qu.:0.0000   1st Qu.:12.61
## Median : 3.000      Median :0.0000   Median :0.0000   Median :16.16
## Mean   : 3.818      Mean   :0.1495   Mean   :0.1793   Mean   :17.22
## 3rd Qu.: 6.050      3rd Qu.:0.0000   3rd Qu.:0.0000   3rd Qu.:20.70
## Max.   :70.000      Max.   :2.0000   Max.   :3.0000   Max.   :37.64
##
##      Team      CumulativeAveragePassingYards
## Length:368      Min.   :120.5
```

```

## Class :character    1st Qu.:221.9
## Mode  :character    Median :247.8
##                               Mean  :246.6
##                               3rd Qu.:271.5
##                               Max.   :369.0
##
## CumulativeAveragePassingTouchdowns CumulativeAveragePassingInterceptions
## Min.      :0.000                      Min.      :0.0000
## 1st Qu.:1.250                      1st Qu.:0.4125
## Median :1.667                      Median :0.5714
## Mean    :1.584                      Mean    :0.5933
## 3rd Qu.:2.000                      3rd Qu.:0.7500
## Max.    :4.000                      Max.    :2.0000
##
## CumulativeAveragePassingRating CumulativeAverageCompletions
## Min.      : 56.25                     Min.      :10.00
## 1st Qu.: 87.77                     1st Qu.:19.56
## Median : 97.01                     Median :21.85
## Mean    : 95.86                     Mean     :21.43
## 3rd Qu.:102.53                     3rd Qu.:23.13
## Max.    :148.57                     Max.     :29.00
##
## CumulativeAverageCompletionPercentage CumulativeMaxPassingTouchdowns
## Min.      :44.40                     Min.      :0.00
## 1st Qu.:61.60                      1st Qu.:2.00
## Median :64.13                      Median :3.00
## Mean    :64.42                      Mean    :2.91
## 3rd Qu.:67.37                      3rd Qu.:4.00
## Max.    :80.00                      Max.    :5.00
##
## CumulativeMaxPassingYards CumulativeMaxPassingAttempts
## Min.      :125.0                     Min.      :21.00
## 1st Qu.:288.0                      1st Qu.:39.00
## Median :332.0                      Median :44.00
## Mean    :331.3                     Mean     :43.31
## 3rd Qu.:368.0                      3rd Qu.:49.00
## Max.    :506.0                     Max.     :66.00
##
## CumulativeMaxPassingRating CumulativeMaxCompletions
## Min.      : 56.25                     Min.      :11.00
## 1st Qu.:110.80                     1st Qu.:25.00
## Median :125.96                     Median :28.00
## Mean    :123.69                     Mean     :27.42
## 3rd Qu.:141.79                     3rd Qu.:30.00
## Max.    :150.69                     Max.     :44.00
##
## CumulativeMaxPassYardsPerAttempt CumulativeMinPassingTouchdowns
## Min.      : 4.800                     Min.      :0.0000
## 1st Qu.: 8.500                      1st Qu.:0.0000
## Median :10.500                      Median :0.0000
## Mean    : 9.859                      Mean     :0.4647
## 3rd Qu.:10.900                      3rd Qu.:1.0000
## Max.    :14.100                      Max.     :4.0000
##

```

```
## CumulativeMinPassingYards CumulativeMinPassingAttempts
## Min. : 69.0 Min. :10.00
## 1st Qu.:128.0 1st Qu.:21.00
## Median :158.0 Median :24.00
## Mean :168.1 Mean :24.13
## 3rd Qu.:204.0 3rd Qu.:27.00
## Max. :369.0 Max. :41.00
##
## CumulativeMinPassingRating CumulativeMinCompletions
## Min. : 31.14 Min. : 6.00
## 1st Qu.: 59.66 1st Qu.:12.00
## Median : 68.58 Median :14.00
## Mean : 69.27 Mean :14.91
## 3rd Qu.: 77.92 3rd Qu.:17.00
## Max. :148.57 Max. :29.00
##
## CumulativeMinPassYardsPerAttempt NextWeekFantasyPoints NextOpponent
## Min. : 3.100 Min. : 7.32 Length:368
## 1st Qu.: 4.800 1st Qu.:12.79 Class :character
## Median : 5.800 Median :16.23 Mode :character
## Mean : 5.704 Mean :17.33
## 3rd Qu.: 6.200 3rd Qu.:20.91
## Max. :11.200 Max. :37.64
## NA's :28
##
## PassingDefense PointsAllowedByDefenseSpecialTeams AvgPassDefense
## Min. : 0.00 Min. : 0.00 Min. : 13.00
## 1st Qu.: 30.00 1st Qu.:16.00 1st Qu.: 46.22
## Median : 54.00 Median :21.00 Median : 60.17
## Mean : 56.76 Mean :21.48 Mean : 59.85
## 3rd Qu.: 76.00 3rd Qu.:27.00 3rd Qu.: 69.40
## Max. :200.00 Max. :51.00 Max. :200.00
## NA's :51 NA's :51 NA's :51
```

## 3.7 : Create indepenent structure for EDA from QBCrossSectional Data

```
eda_base <- QBCrossSectional %>% group_by(Week) %>%
  select(Week,
         FantasyPoints,
         PassingCompletions,
         PassingAttempts,
         PassingCompletionPercentage,
         PassingYards,
         PassingYardsPerAttempt,
         PassingTouchdowns,
         PassingInterceptions,
         PassingRating,
         RushingAttempts,
         RushingYards,
         RushingYardsPerAttempt,
         RushingTouchdowns,
         FumblesLost
        )

eda_base$Week <- as.factor(eda_base$Week)
```



```
## 3.8 : Create indepenent structure for derived features for EDA
      from QBCrossSectionalDefensiveOverlayCumulativePassYards Data
eda_derived <- QBCrossSectionalDefensiveOverlayCumulativePassYards %>% group_by(Week)%>%
  select(Week,
         FantasyPoints,
         CumulativeAveragePassingYards,
         CumulativeAveragePassingTouchdowns,
         CumulativeAveragePassingInterceptions,
         CumulativeAveragePassingRating,
         CumulativeAverageCompletions,
         CumulativeAverageCompletionPercentage,
         CumulativeMaxPassingTouchdowns,
         CumulativeMaxPassingYards,
         CumulativeMaxPassingAttempts,
         CumulativeMaxPassingRating,
         CumulativeMaxCompletions,
         CumulativeMaxPassYardsPerAttempt,
         CumulativeMinPassingTouchdowns,
         CumulativeMinPassingYards,
         CumulativeMinPassingAttempts,
         CumulativeMinPassingRating,
         CumulativeMinCompletions,
         CumulativeMinPassYardsPerAttempt)
eda_derived$Week <- as.factor(eda_derived$Week)
```

## 4 : Structures QBCrossSectional (Quarterbacks)

```
summary(QBCrossSectional)
```

```
##      PlayerID      Name      Week      Position
## Min.      : 611  Length:453  Min.      : 1.000  Length:453
## 1st Qu.: 7242  Class :character 1st Qu.: 5.000  Class :character
## Median :13723  Mode  :character  Median : 9.000  Mode  :character
## Mean    :11932                      Mean    : 9.049
## 3rd Qu.:16763                      3rd Qu.:13.000
## Max.    :19029                      Max.    :17.000
##      Opponent      TeamIsHome      GameDate      PassingCompletions
## Length:453      Mode :logical  Length:453      Min.      : 5.00
## Class :character FALSE:226      Class :character 1st Qu.:18.00
## Mode  :character TRUE :227      Mode  :character  Median :21.00
##                                     Mean    :21.24
##                                     3rd Qu.:25.00
##                                     Max.    :44.00
##      Result      PassingAttempts  PassingCompletionPercentage
## Length:453      Min.      :10.00  Min.      :38.70
## Class :character 1st Qu.:29.00  1st Qu.:57.10
## Mode  :character  Median :33.00  Median :63.20
##                                     Mean    :33.62  Mean    :63.48
##                                     3rd Qu.:38.00  3rd Qu.:69.40
##                                     Max.    :66.00  Max.    :87.00
##      PassingYards  PassingYardsPerAttempt  PassingTouchdowns
## Min.      : 57.0  Min.      : 3.100  Min.      :0.000
```

```
## 1st Qu.:199.0 1st Qu.: 6.200 1st Qu.:1.000
## Median :241.0 Median : 7.200 Median :1.000
## Mean :244.6 Mean : 7.378 Mean :1.587
## 3rd Qu.:291.0 3rd Qu.: 8.400 3rd Qu.:2.000
## Max. :506.0 Max. :14.100 Max. :5.000
## PassingInterceptions PassingRating RushingAttempts RushingYards
## Min. :0.00 Min. : 31.14 Min. : 0.000 Min. : -8.00
## 1st Qu.:0.00 1st Qu.: 77.92 1st Qu.: 1.000 1st Qu.: 0.00
## Median :0.00 Median : 92.94 Median : 3.000 Median : 8.00
## Mean :0.66 Mean : 93.88 Mean : 3.185 Mean :14.38
## 3rd Qu.:1.00 3rd Qu.:109.84 3rd Qu.: 5.000 3rd Qu.:23.00
## Max. :4.00 Max. :150.69 Max. :14.000 Max. :95.00
## RushingYardsPerAttempt RushingTouchdowns FumblesLost FantasyPoints
## Min. : -2.700 Min. :0.0000 Min. :0.0000 Min. : 7.12
## 1st Qu.: 0.000 1st Qu.:0.0000 1st Qu.:0.0000 1st Qu.:11.86
## Median : 3.000 Median :0.0000 Median :0.0000 Median :15.86
## Mean : 3.839 Mean :0.1457 Mean :0.1766 Mean :16.89
## 3rd Qu.: 6.000 3rd Qu.:0.0000 3rd Qu.:0.0000 3rd Qu.:20.68
## Max. :70.000 Max. :2.0000 Max. :3.0000 Max. :37.64
## Team
## Length:453
## Class :character
## Mode :character
##
##
##
```

## 5 : struture with Derived features

```
summary(eda_derived)
```

```
## Week FantasyPoints CumulativeAveragePassingYards
## 2 : 25 Min. : 7.32 Min. :120.5
## 3 : 24 1st Qu.:12.61 1st Qu.:221.9
## 4 : 24 Median :16.16 Median :247.8
## 13 : 24 Mean :17.22 Mean :246.6
## 14 : 24 3rd Qu.:20.70 3rd Qu.:271.5
## 7 : 23 Max. :37.64 Max. :369.0
## (Other):224
## CumulativeAveragePassingTouchdowns CumulativeAveragePassingInterceptions
## Min. :0.000 Min. :0.0000
## 1st Qu.:1.250 1st Qu.:0.4125
## Median :1.667 Median :0.5714
## Mean :1.584 Mean :0.5933
## 3rd Qu.:2.000 3rd Qu.:0.7500
## Max. :4.000 Max. :2.0000
##
## CumulativeAveragePassingRating CumulativeAverageCompletions
## Min. : 56.25 Min. :10.00
## 1st Qu.: 87.77 1st Qu.:19.56
## Median : 97.01 Median :21.85
## Mean : 95.86 Mean :21.43
## 3rd Qu.:102.53 3rd Qu.:23.13
```

```

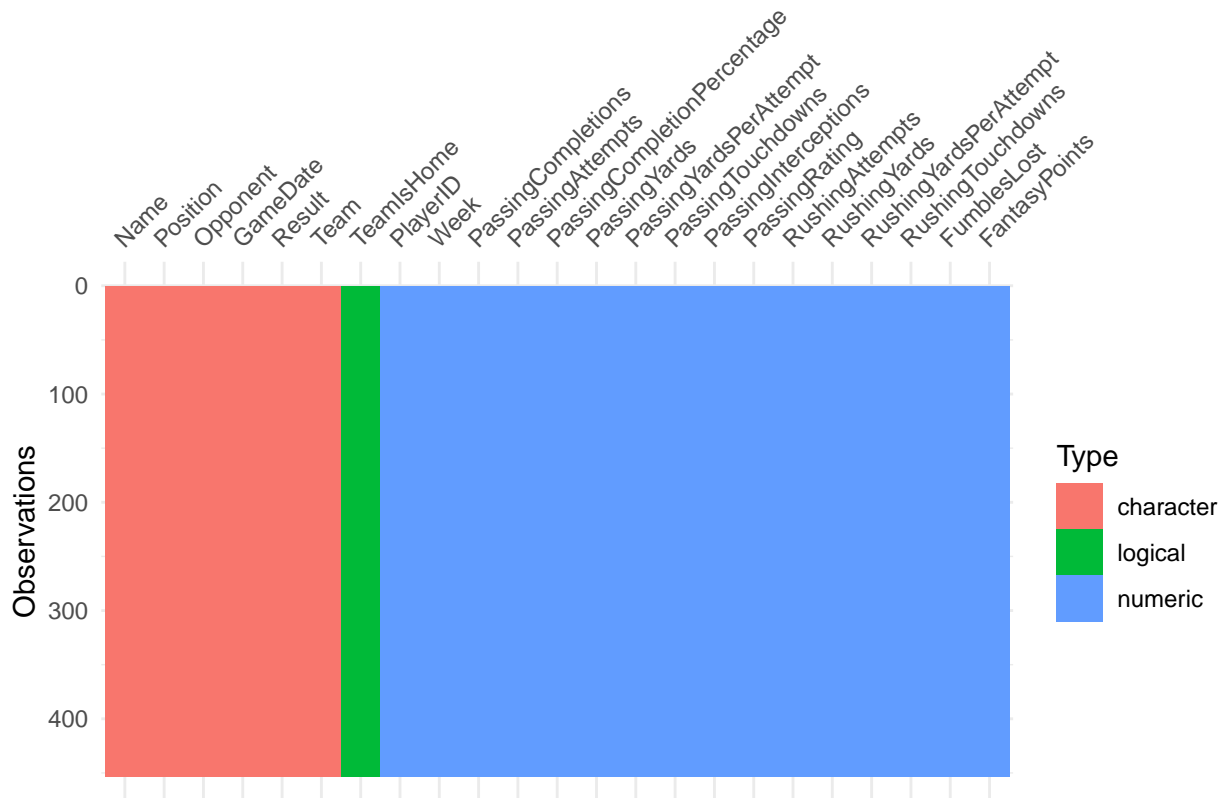
## Max. :148.57 Max. :29.00
##
## CumulativeAverageCompletionPercentage CumulativeMaxPassingTouchdowns
## Min. :44.40 Min. :0.00
## 1st Qu.:61.60 1st Qu.:2.00
## Median :64.13 Median :3.00
## Mean :64.42 Mean :2.91
## 3rd Qu.:67.37 3rd Qu.:4.00
## Max. :80.00 Max. :5.00
##
## CumulativeMaxPassingYards CumulativeMaxPassingAttempts
## Min. :125.0 Min. :21.00
## 1st Qu.:288.0 1st Qu.:39.00
## Median :332.0 Median :44.00
## Mean :331.3 Mean :43.31
## 3rd Qu.:368.0 3rd Qu.:49.00
## Max. :506.0 Max. :66.00
##
## CumulativeMaxPassingRating CumulativeMaxCompletions
## Min. : 56.25 Min. :11.00
## 1st Qu.:110.80 1st Qu.:25.00
## Median :125.96 Median :28.00
## Mean :123.69 Mean :27.42
## 3rd Qu.:141.79 3rd Qu.:30.00
## Max. :150.69 Max. :44.00
##
## CumulativeMaxPassYardsPerAttempt CumulativeMinPassingTouchdowns
## Min. : 4.800 Min. :0.0000
## 1st Qu.: 8.500 1st Qu.:0.0000
## Median :10.500 Median :0.0000
## Mean : 9.859 Mean :0.4647
## 3rd Qu.:10.900 3rd Qu.:1.0000
## Max. :14.100 Max. :4.0000
##
## CumulativeMinPassingYards CumulativeMinPassingAttempts
## Min. : 69.0 Min. :10.00
## 1st Qu.:128.0 1st Qu.:21.00
## Median :158.0 Median :24.00
## Mean :168.1 Mean :24.13
## 3rd Qu.:204.0 3rd Qu.:27.00
## Max. :369.0 Max. :41.00
##
## CumulativeMinPassingRating CumulativeMinCompletions
## Min. : 31.14 Min. : 6.00
## 1st Qu.: 59.66 1st Qu.:12.00
## Median : 68.58 Median :14.00
## Mean : 69.27 Mean :14.91
## 3rd Qu.: 77.92 3rd Qu.:17.00
## Max. :148.57 Max. :29.00
##
## CumulativeMinPassYardsPerAttempt
## Min. : 3.100
## 1st Qu.: 4.800
## Median : 5.800

```

```
## Mean    : 5.704
## 3rd Qu.: 6.200
## Max.    :11.200
##
```

## 6 : QBCrossSectional - Check Missing Data

```
vis_dat(QBCrossSectional)
```



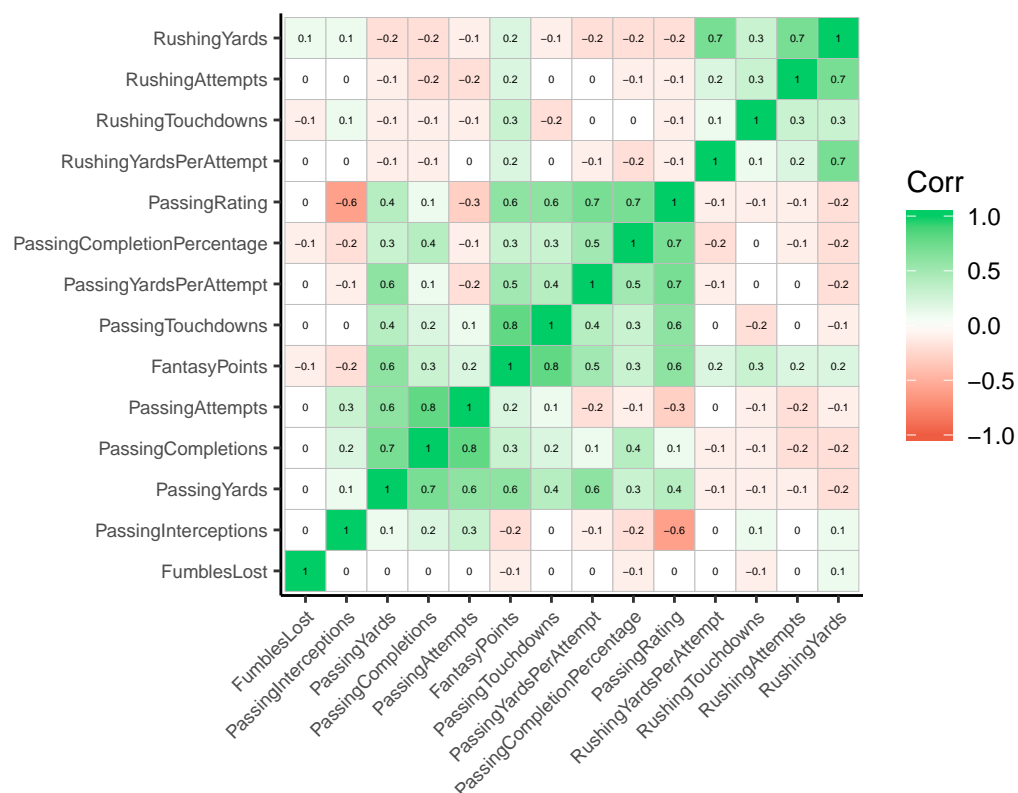
## 7 : Correlation

```
## 7.1 : Base Feature Set
```

```
corr <- round(cor(eda_base[-1]), 1)

ggcorrplot(corr, hc.order = TRUE,
  type = "full",
  lab = TRUE,
  lab_size = 1.5,
  method="square",
  colors = c("tomato2", "white", "springgreen3"),
  title = "Correlation of Base QB features",
  tl.cex = 7, pch=2, pch.col = 3, show.diag = T,
  ggtheme=theme_classic)
```

Correlogram of Base QB features

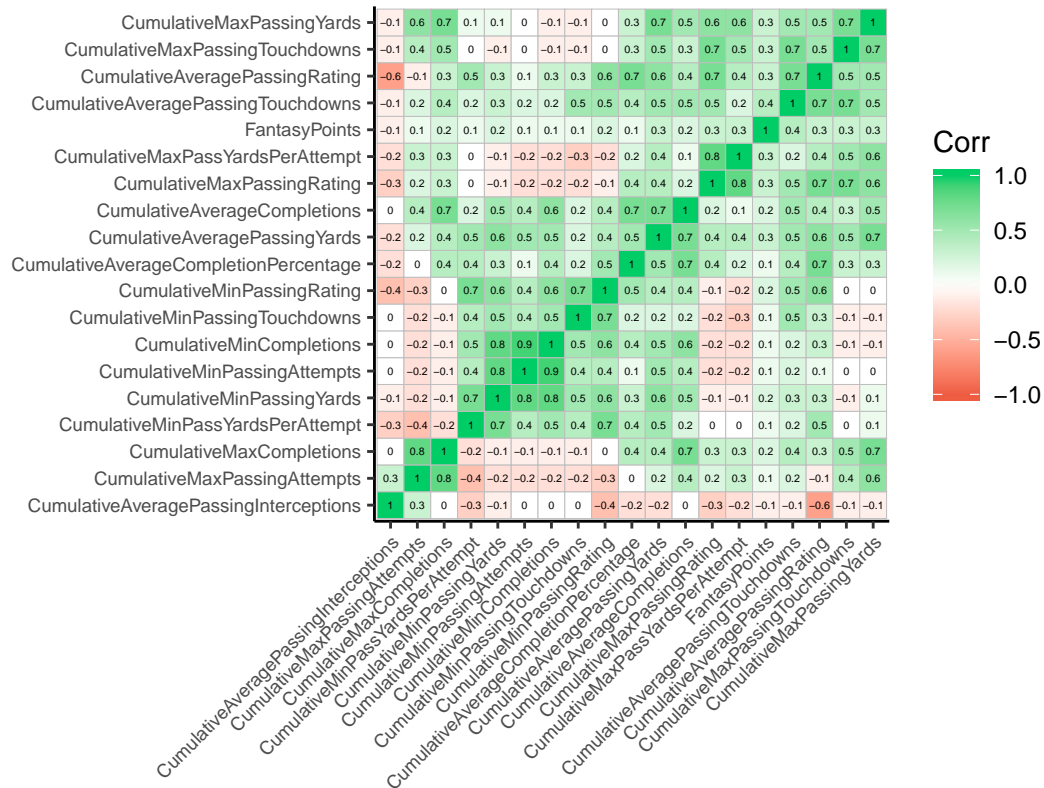


## 7.2 : Derived Feature Set

```
corr <- round(cor(eda_derived[-1]), 1)
```

```
ggcorrplot(corr, hc.order = TRUE,
  type = "full",
  lab = TRUE,
  lab_size = 1.5,
  method="square",
  colors = c("tomato2", "white", "springgreen3"),
  title = "Correlogram of Derived QB features",
  tl.cex = 7, pch=2, pch.col = 3, show.diag = T,
  ggtheme=theme_classic)
```

## Correlogram of Derived QB features

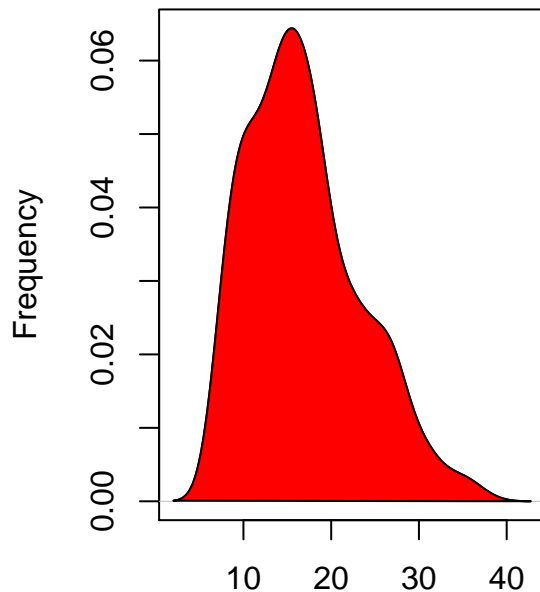


## 8 : Distributions

## 8.1 Density plot for FantasyPoints is approximately Normal

```
par(mfrow=c(1, 2)) # divide graph area in 2 columns
target <- QBCrossSectional$FantasyPoints
plot(density(target), main="CrossSectional Dataset: FantasyPoints", ylab="Frequency", sub=paste("Skewness:", round(skewness(target), 2)))
polygon(density(target), col="red")
```

## CrossSectional Dataset: FantasyPc

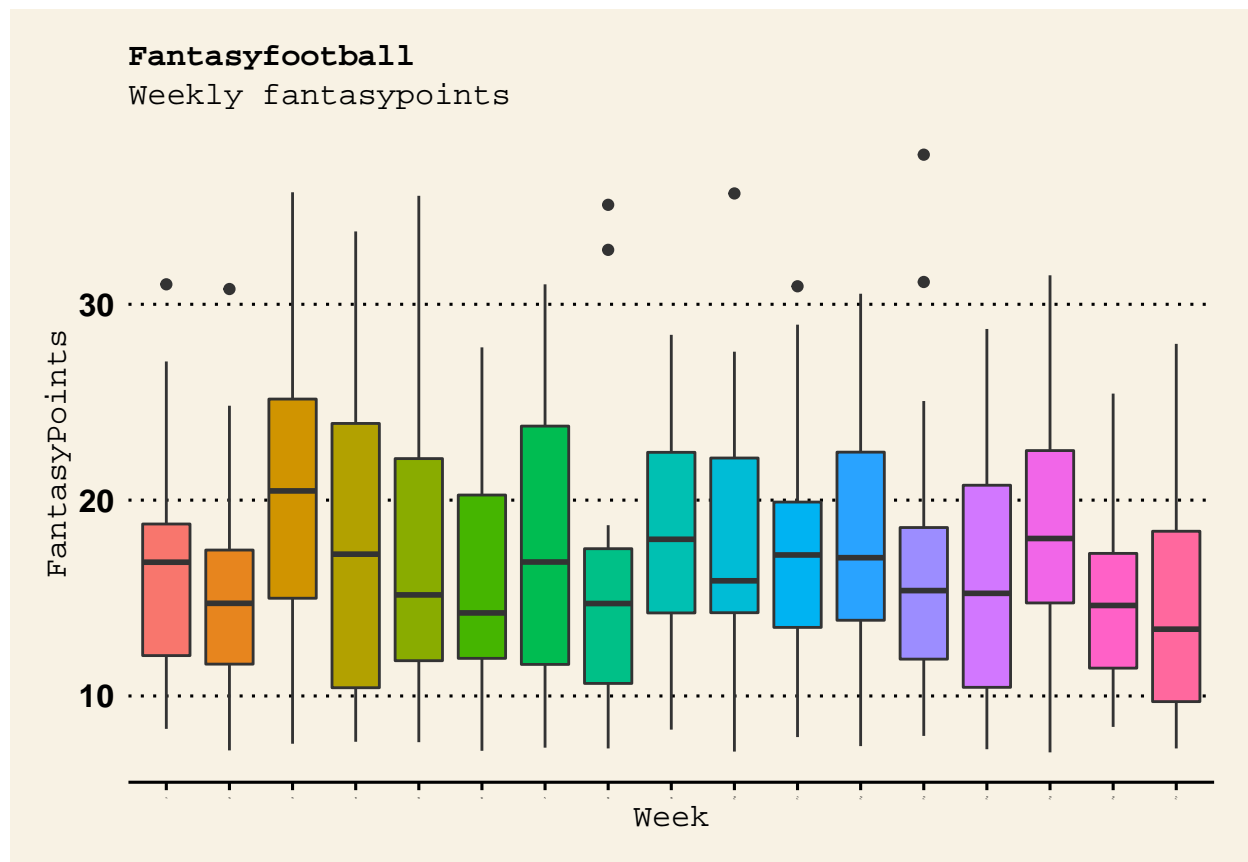


N = 453 Bandwidth = 1.7  
Skewness: 0.67

# 8.2 : Boxplots - Target and Individual Predictor Behavior for per Team

## 8.2.1 Base Feature set

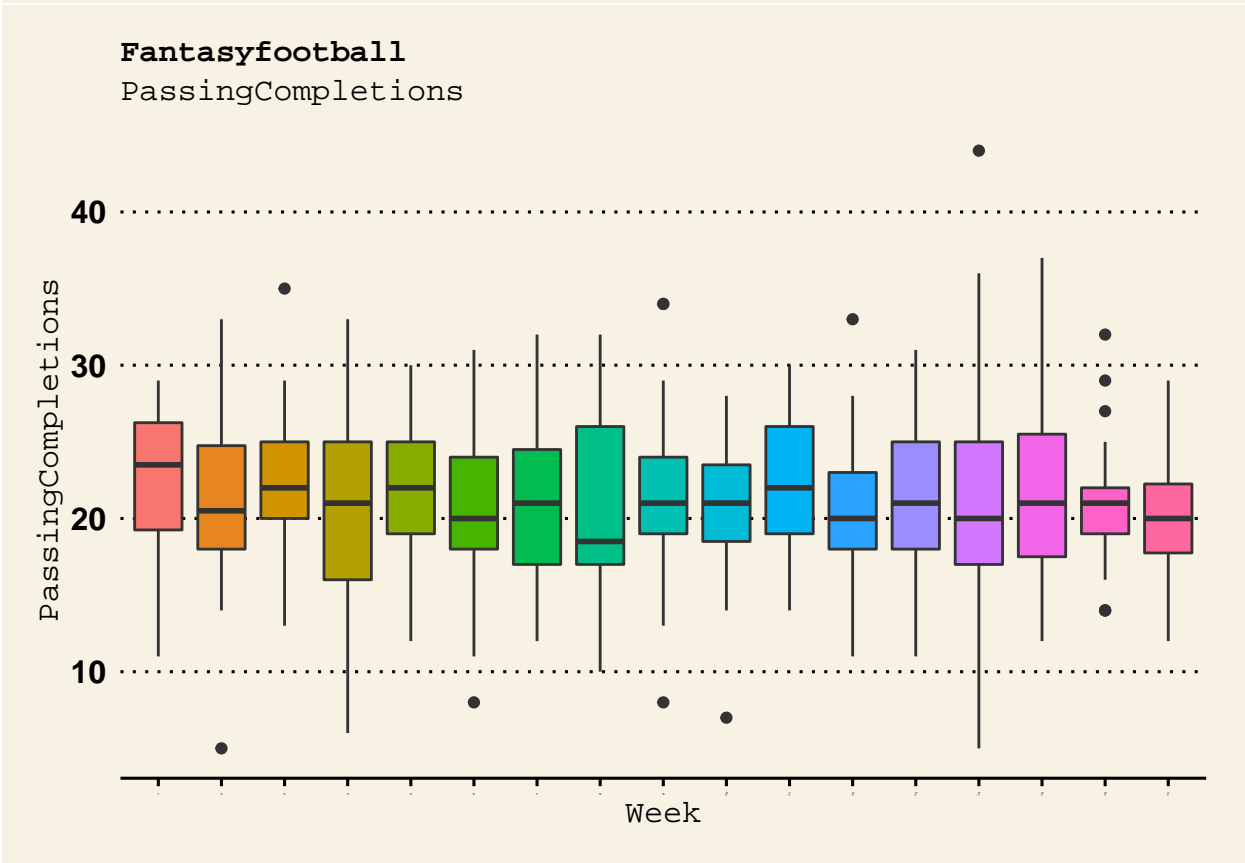
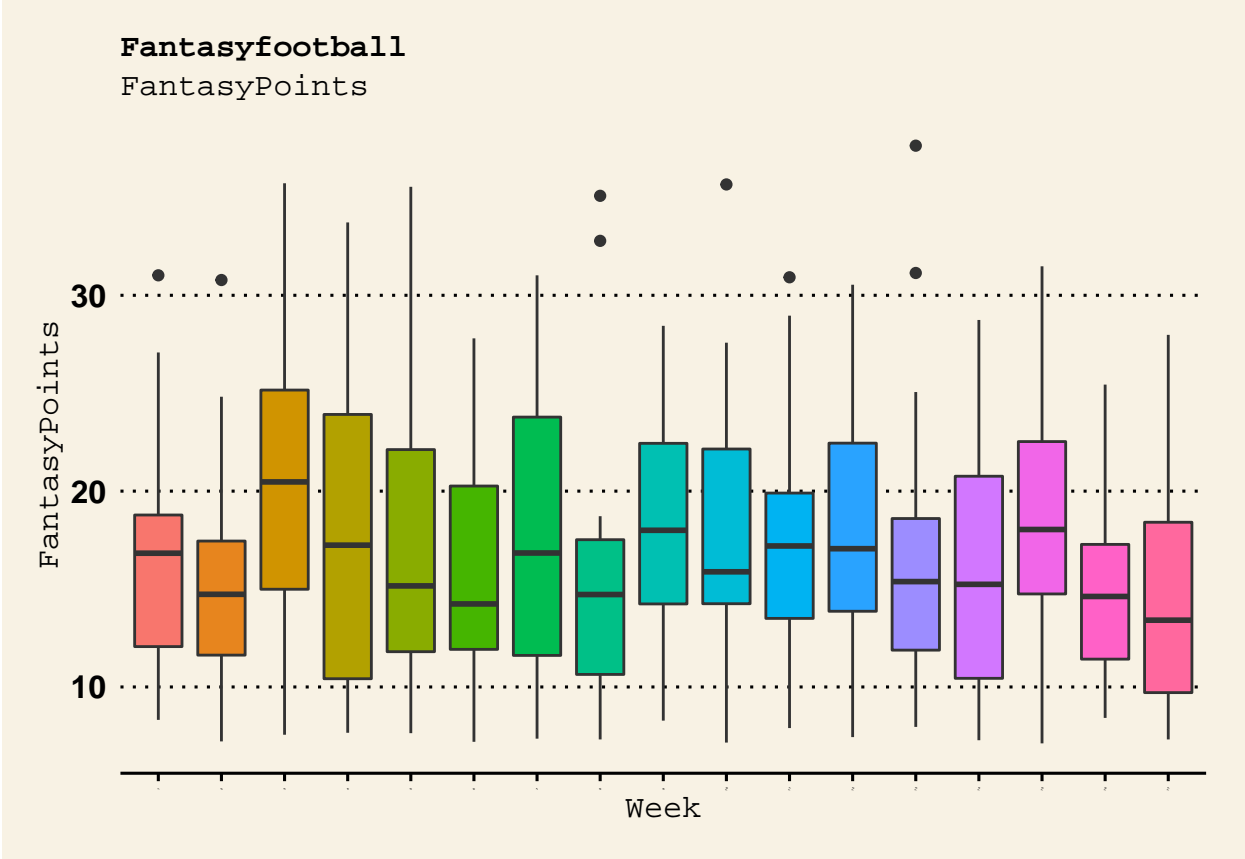
```
eda_base %>% ggplot(aes(y=FantasyPoints,x=Week,fill=Week,group=Week))+  
  geom_boxplot(show.legend = FALSE)+  
  xlab("Week")+ylab("FantasyPoints")+  
  labs(title="Fantasyfootball",  
        subtitle="Weekly fantasypoints",  
        caption="Source: Fantasyfootball")+  
  theme_wsj()+  
  theme(plot.title = element_text(size = rel(0.5)),  
        plot.subtitle = element_text(size = rel(0.5)),  
        axis.text.x = element_text(angle=65, vjust=0.6,size=1),  
        axis.title = element_text(size = rel(0.5)),  
        legend.position = "right",  
        legend.direction = "vertical",  
        legend.title = element_text(size = rel(0.5)))
```



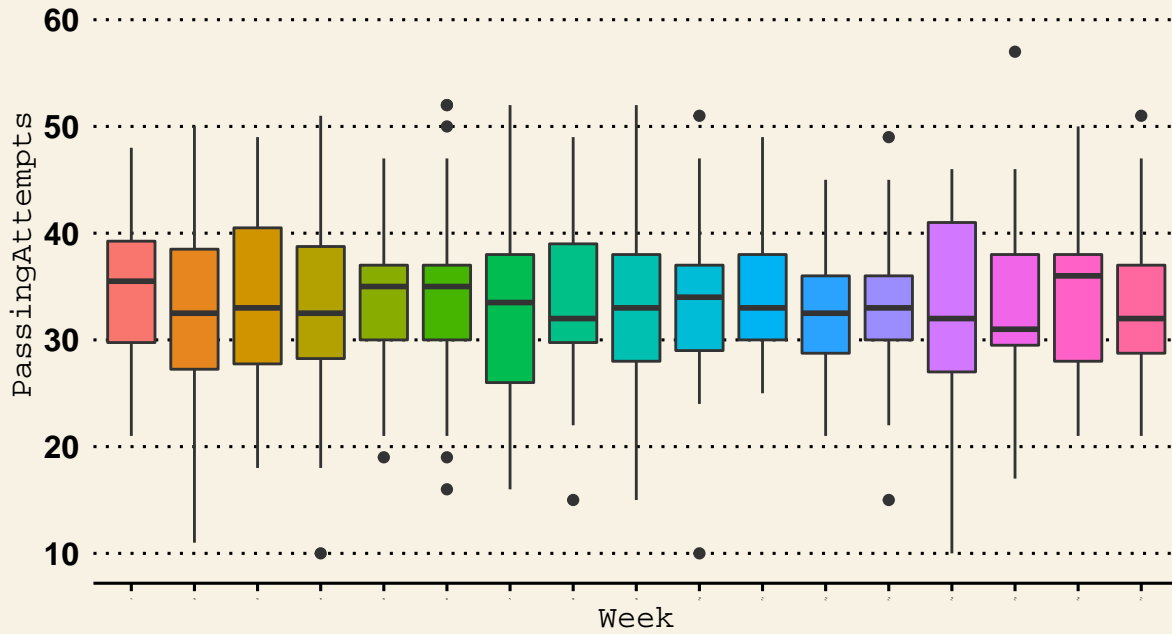
```
for (i in 2:15) {
  ggplotp <- eda_base %>%
    ggplot(aes_string(y=names(eda_base[i]),x="Week",fill="Week",group="Week"))+
    geom_boxplot(show.legend = FALSE)+
    xlab("Week")+ylab(names(eda_base[i]))+
    labs(title="Fantasyfootball",
         subtitle=names(eda_base[i]),
         aption="Source: Fantasyfootball")+
    theme_wsj()+
    theme(plot.title = element_text(size = rel(0.5)),
          plot.subtitle = element_text(size = rel(0.5)),
          axis.text.x = element_text(angle=65, vjust=0.6,size=1),
          axis.title = element_text(size = rel(0.5)),
          legend.position = "right",
          legend.direction = "vertical",
          legend.title = element_text(size = rel(0.5)))

  print(ggplotp)
}
```

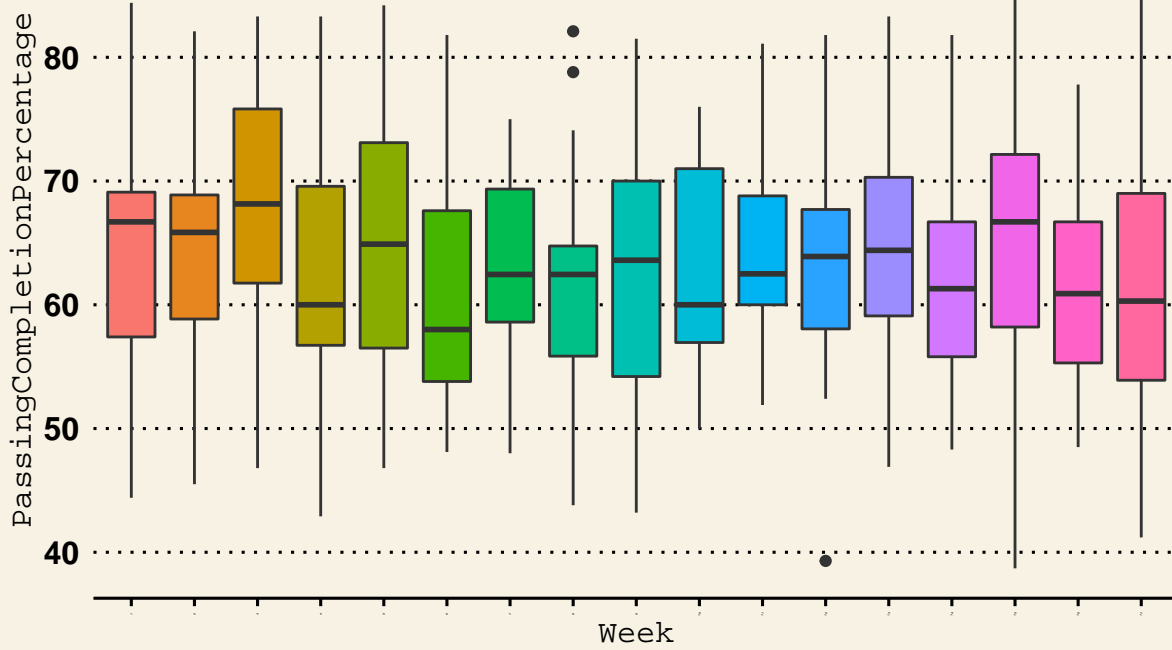


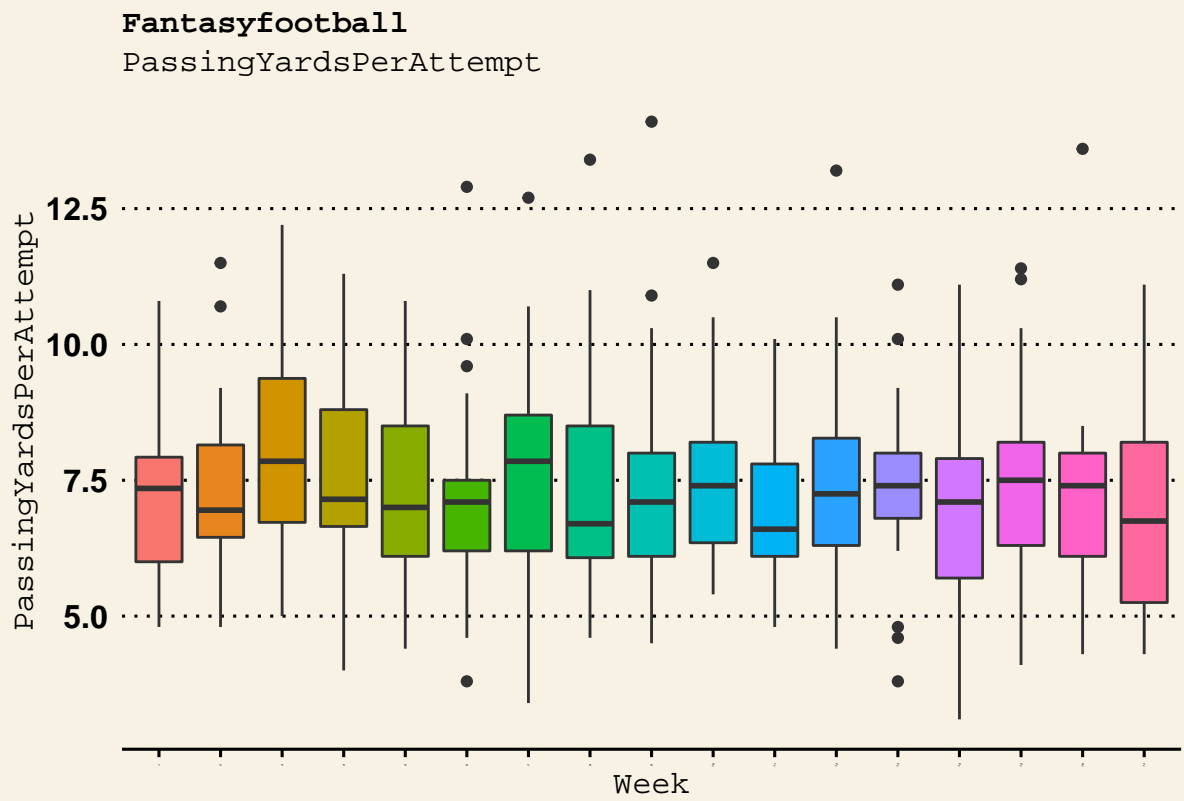
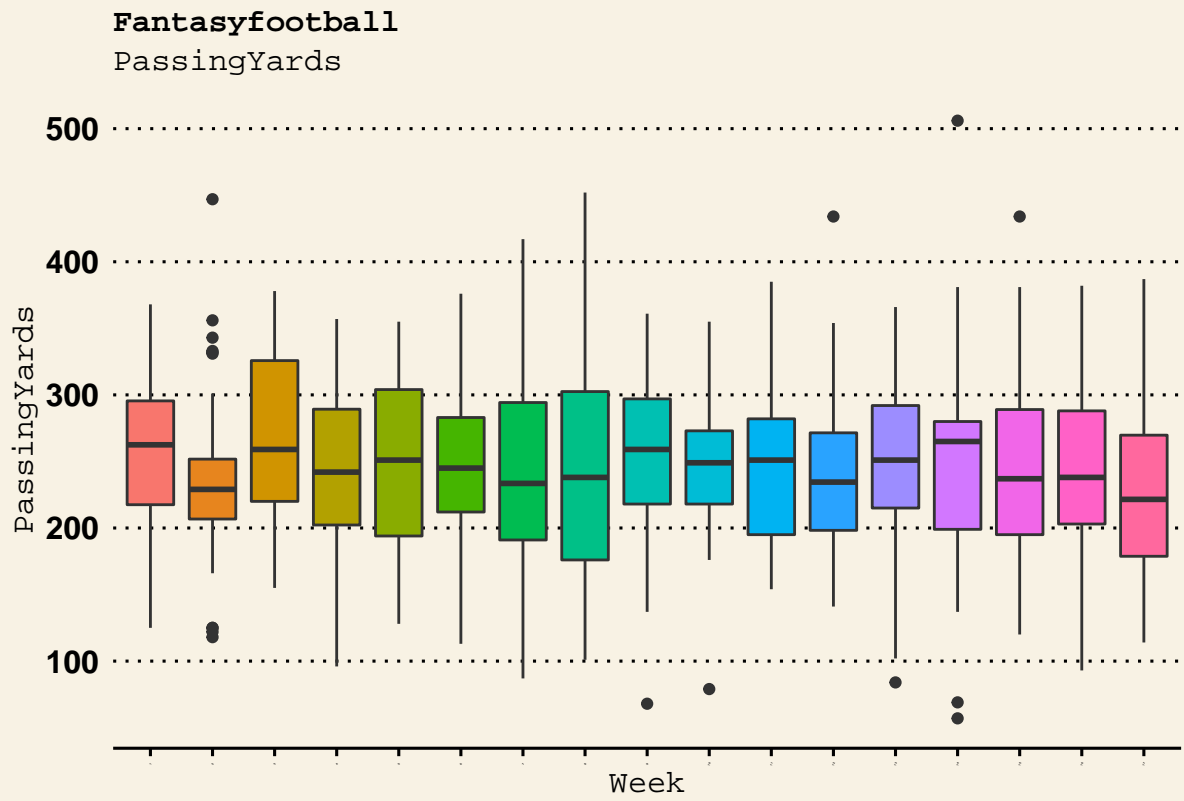


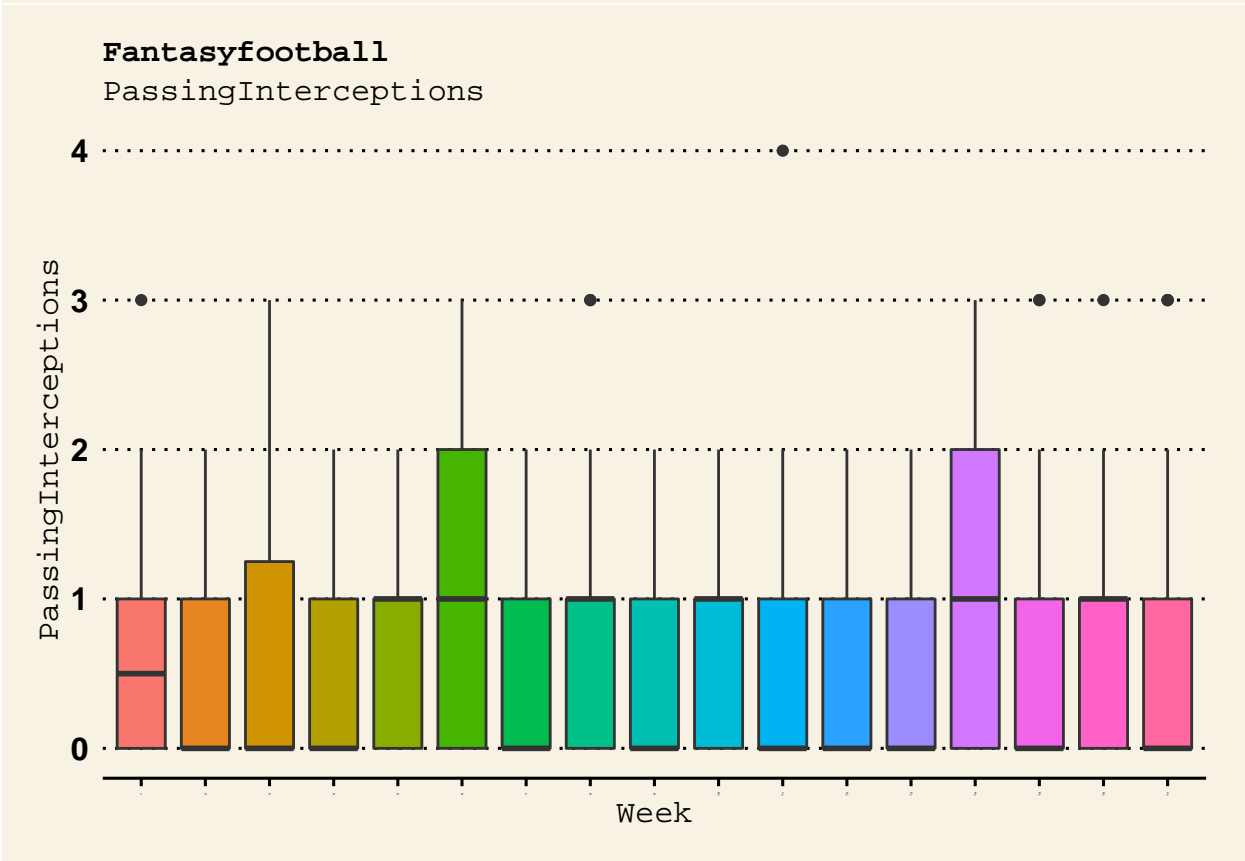
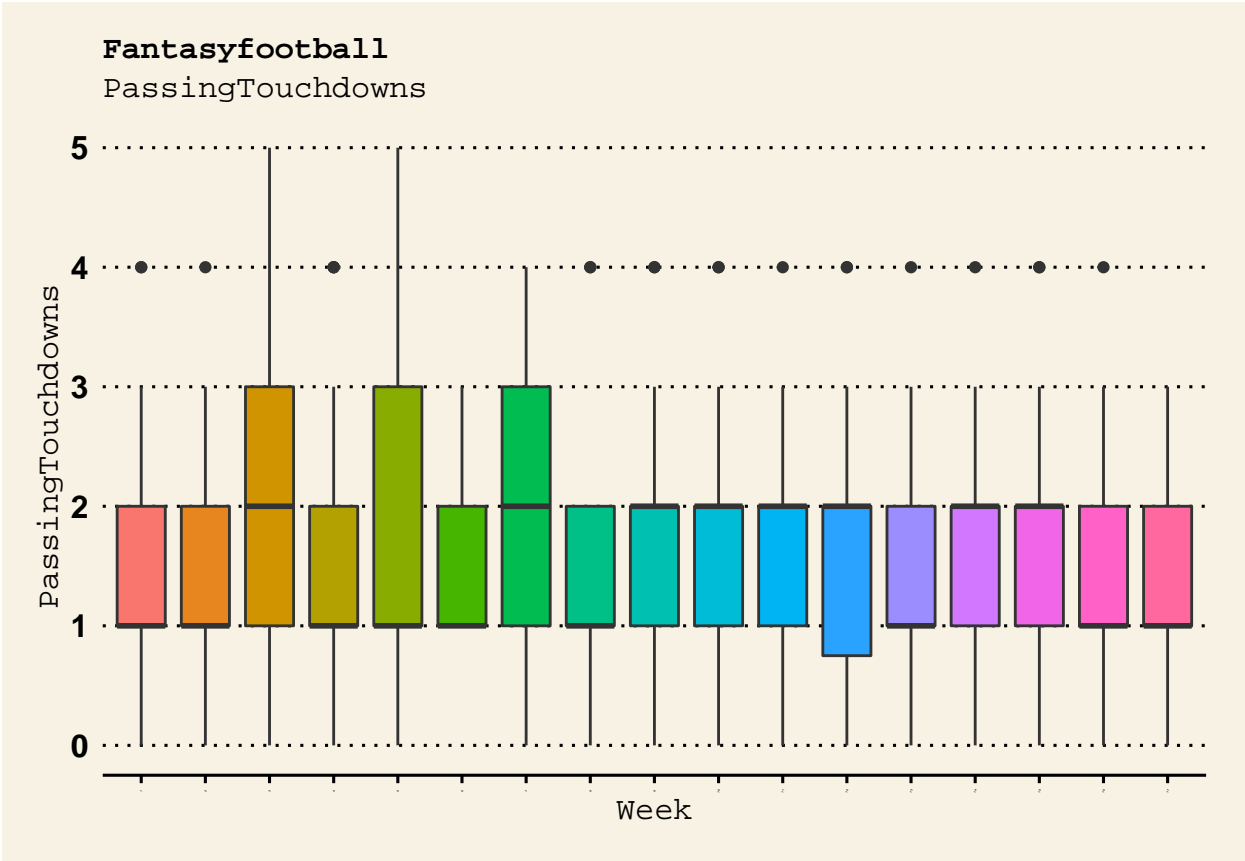
**Fantasyfootball**  
PassingAttempts

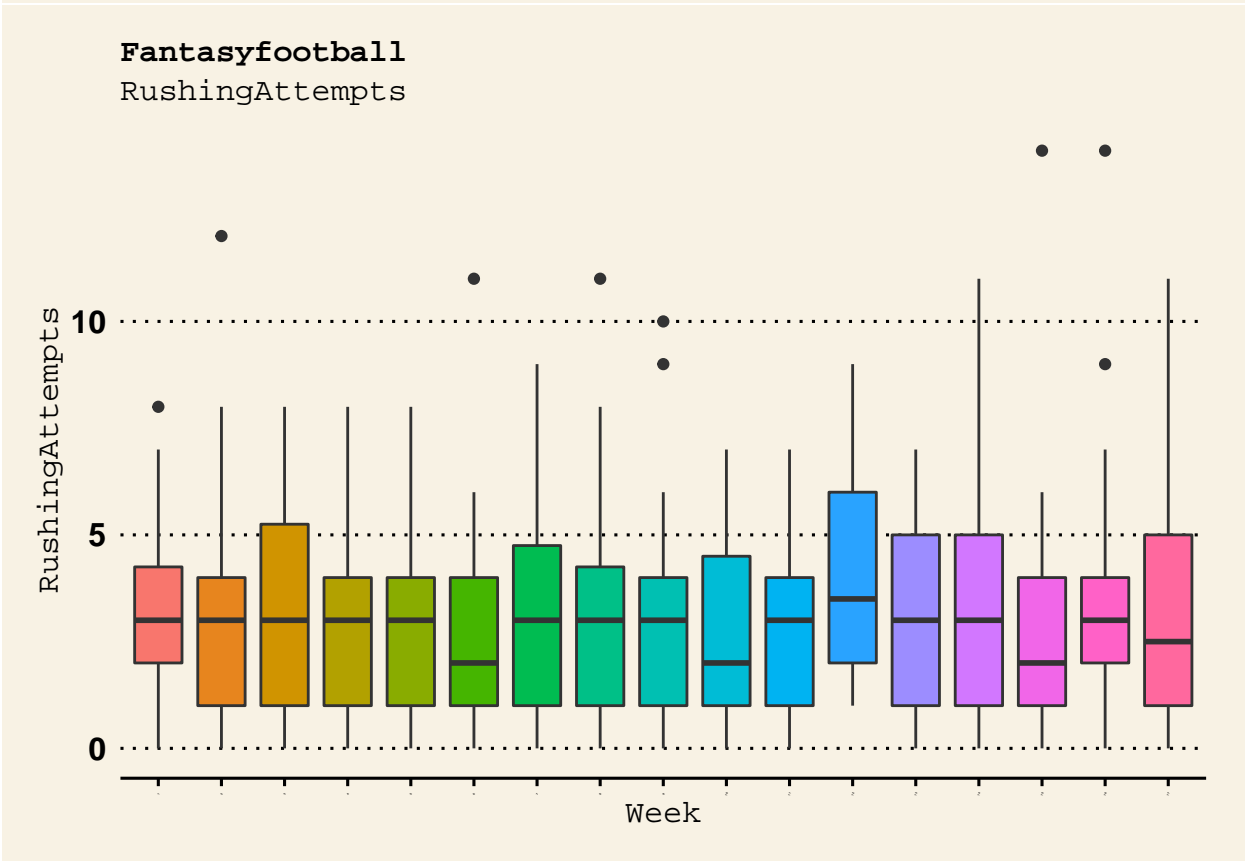
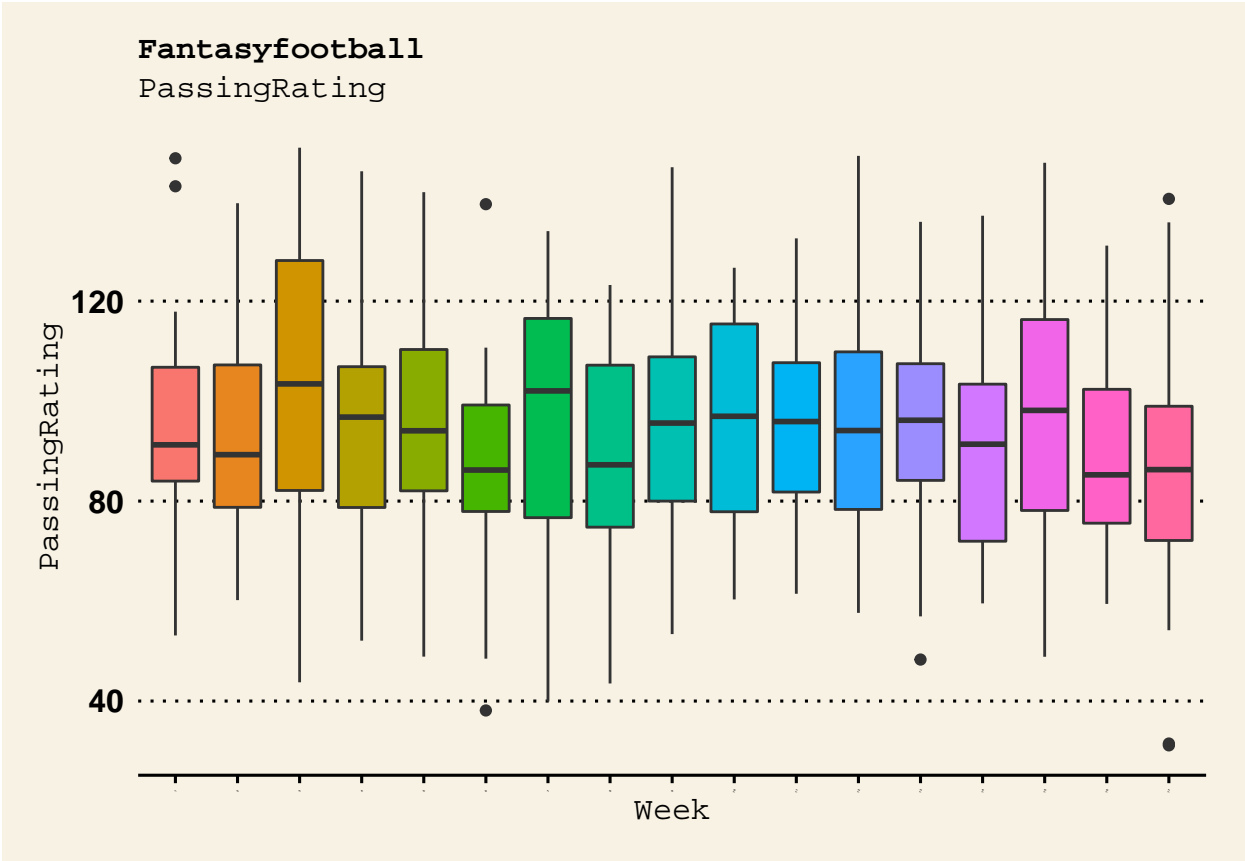


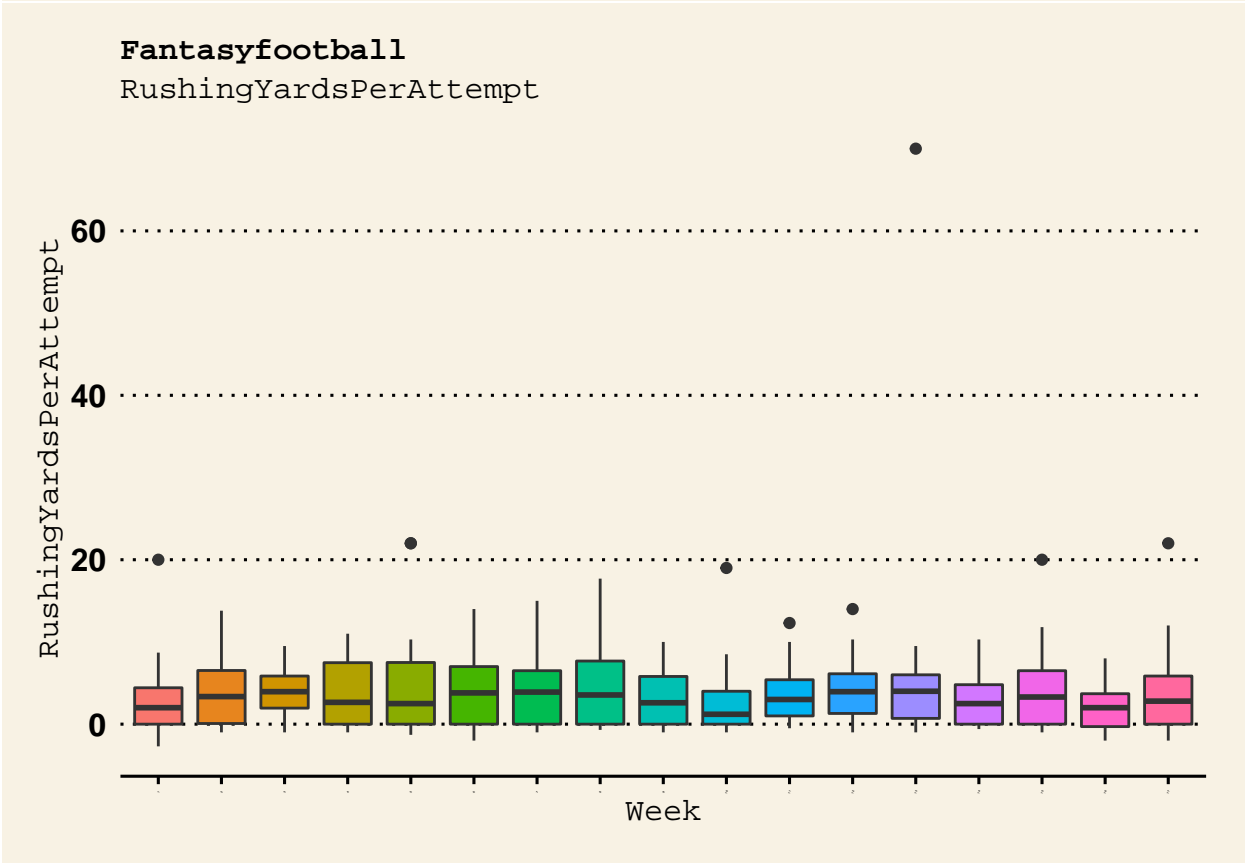
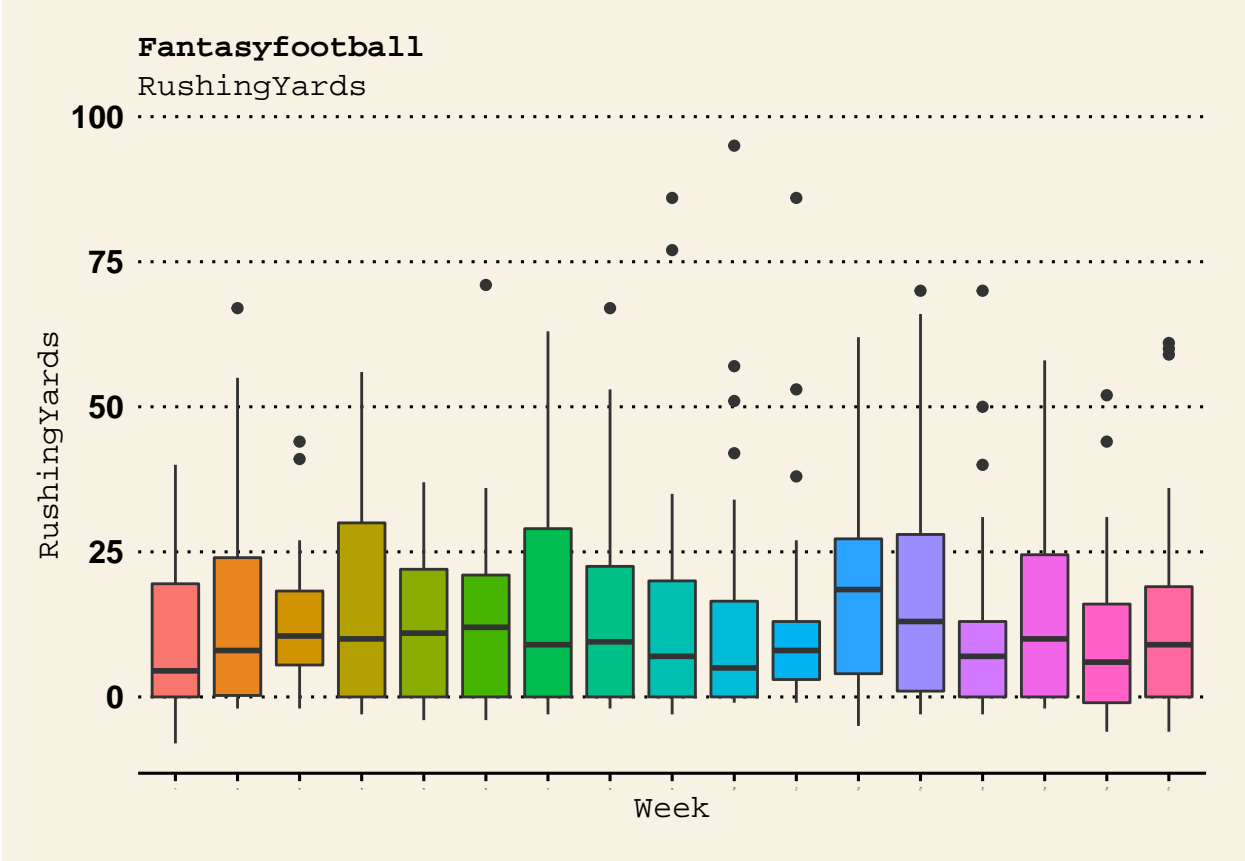
**Fantasyfootball**  
PassingCompletionPercentage

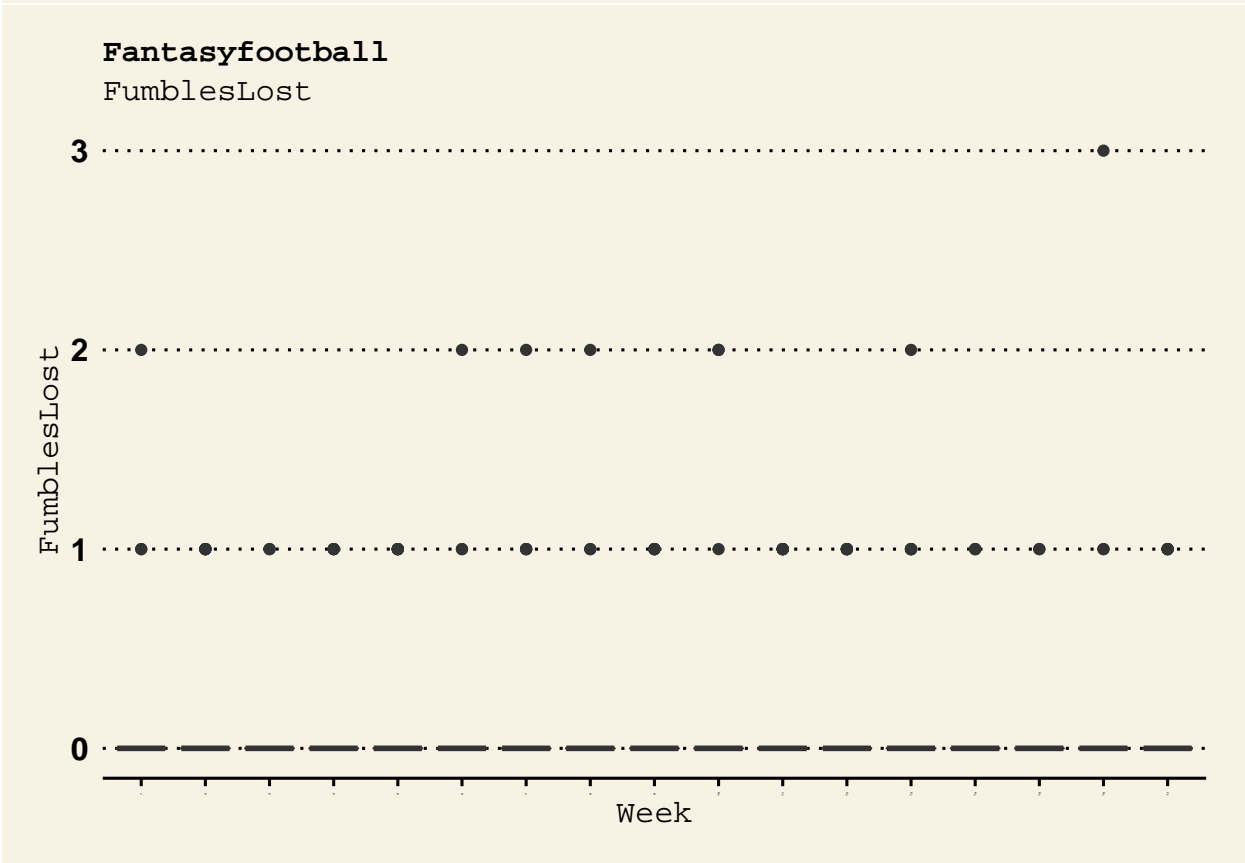
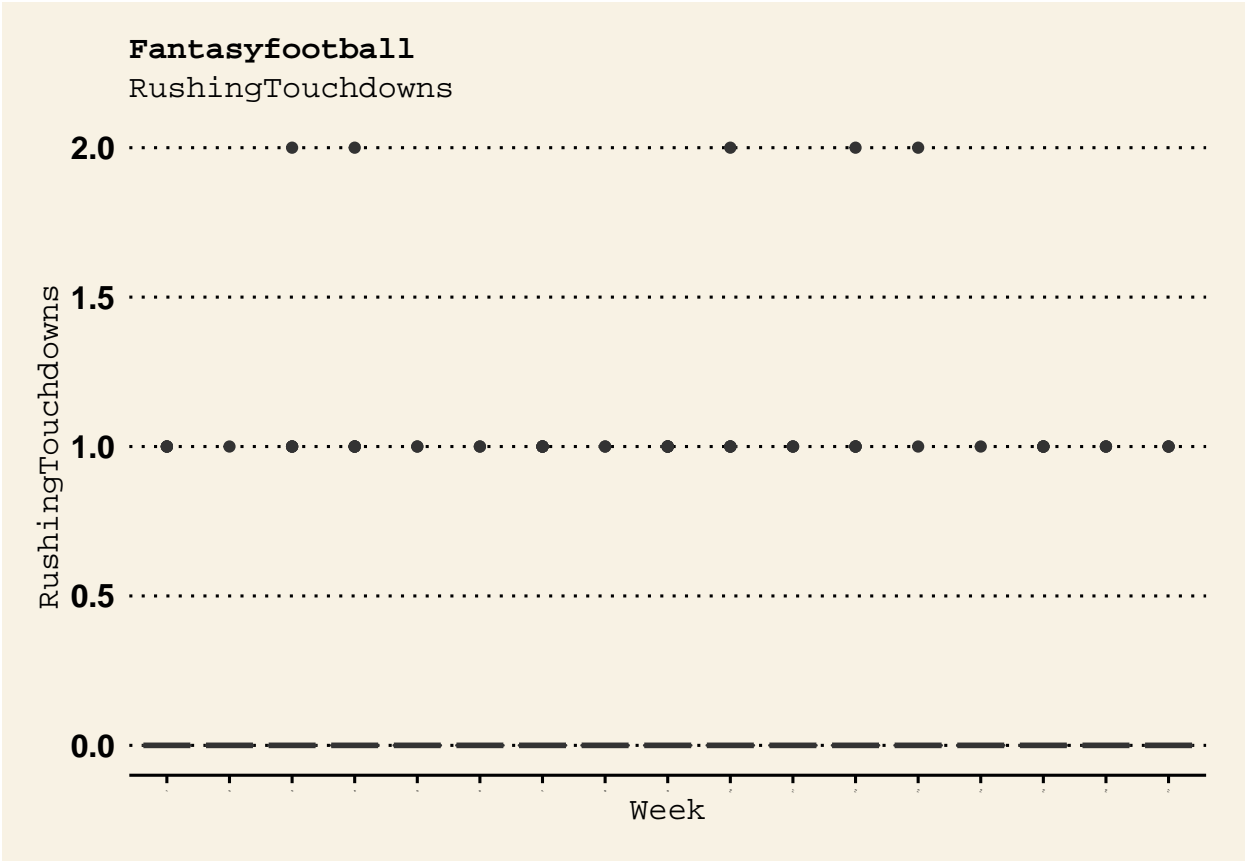






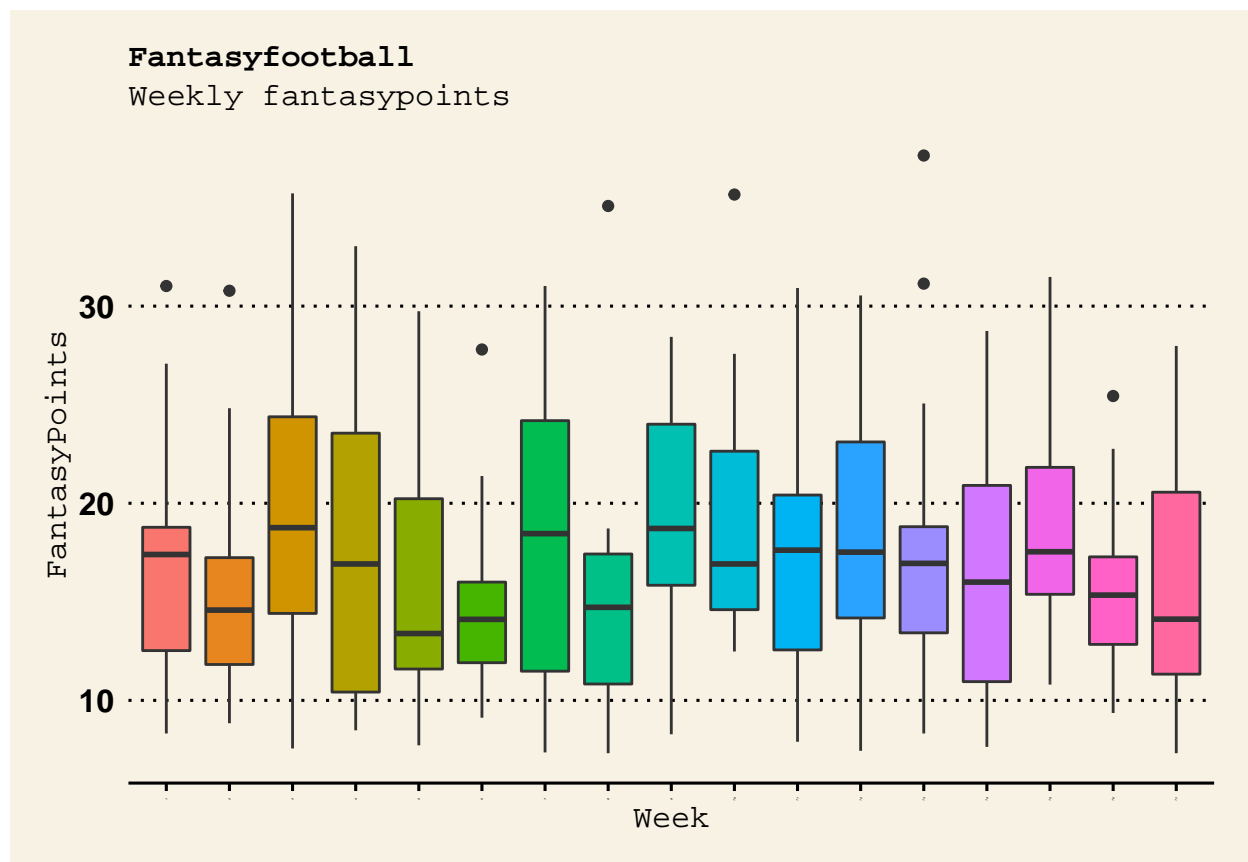






### ## 8.2.3 Boxplots Derived Feature set

```
eda_derived %>% ggplot(aes(y=FantasyPoints,x=Week,fill=Week,group=Week))+
  geom_boxplot(show.legend = FALSE)+
  xlab("Week")+ylab("FantasyPoints")+
  labs(title="Fantasyfootball",
        subtitle="Weekly fantasypoints",
        aption="Source: Fantasyfootball")+
  theme_wsj()+
  theme(plot.title = element_text(size = rel(0.5)),
        plot.subtitle = element_text(size = rel(0.5)),
        axis.text.x = element_text(angle=65, vjust=0.6,size=1),
        axis.title = element_text(size = rel(0.5)),
        legend.position = "right",
        legend.direction = "vertical",
        legend.title = element_text(size = rel(0.5)))
```



```
for (i in 2:20) {
  ggplotp <- eda_derived %>%
    ggplot(aes_string(y=names(eda_derived[i]),x="Week",fill="Week",group="Week"))+
    geom_boxplot(show.legend = FALSE)+
    xlab("Week")+ylab(names(eda_derived[i]))+
    labs(title="Fantasyfootball",
          subtitle=names(eda_derived[i]),
          aption="Source: Fantasyfootball")+
    theme_wsj()+
    theme(plot.title = element_text(size = rel(0.5)),
          plot.subtitle = element_text(size = rel(0.5)),
          axis.text.x = element_text(angle=65, vjust=0.6,size=1),
          axis.title = element_text(size = rel(0.5)),
          legend.position = "right",
          legend.direction = "vertical",
          legend.title = element_text(size = rel(0.5)))
}
```

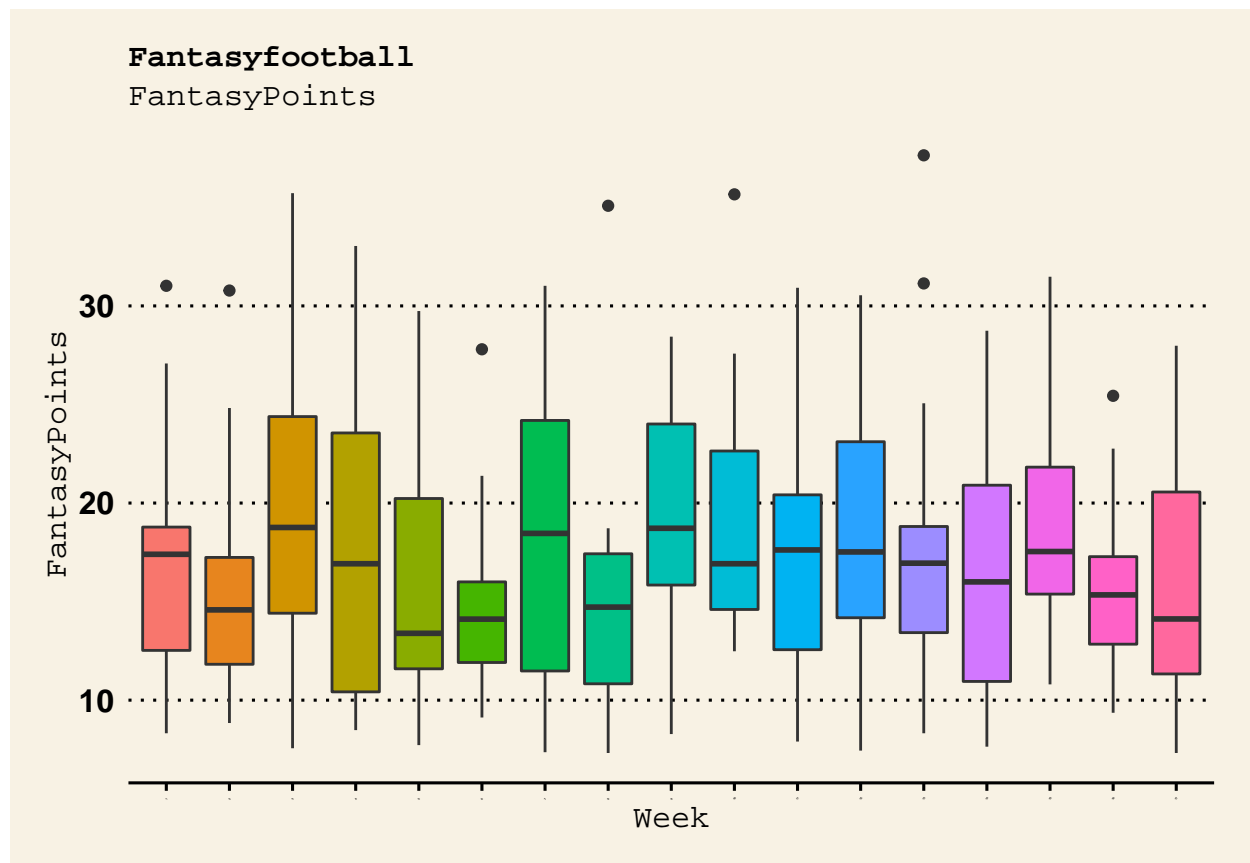


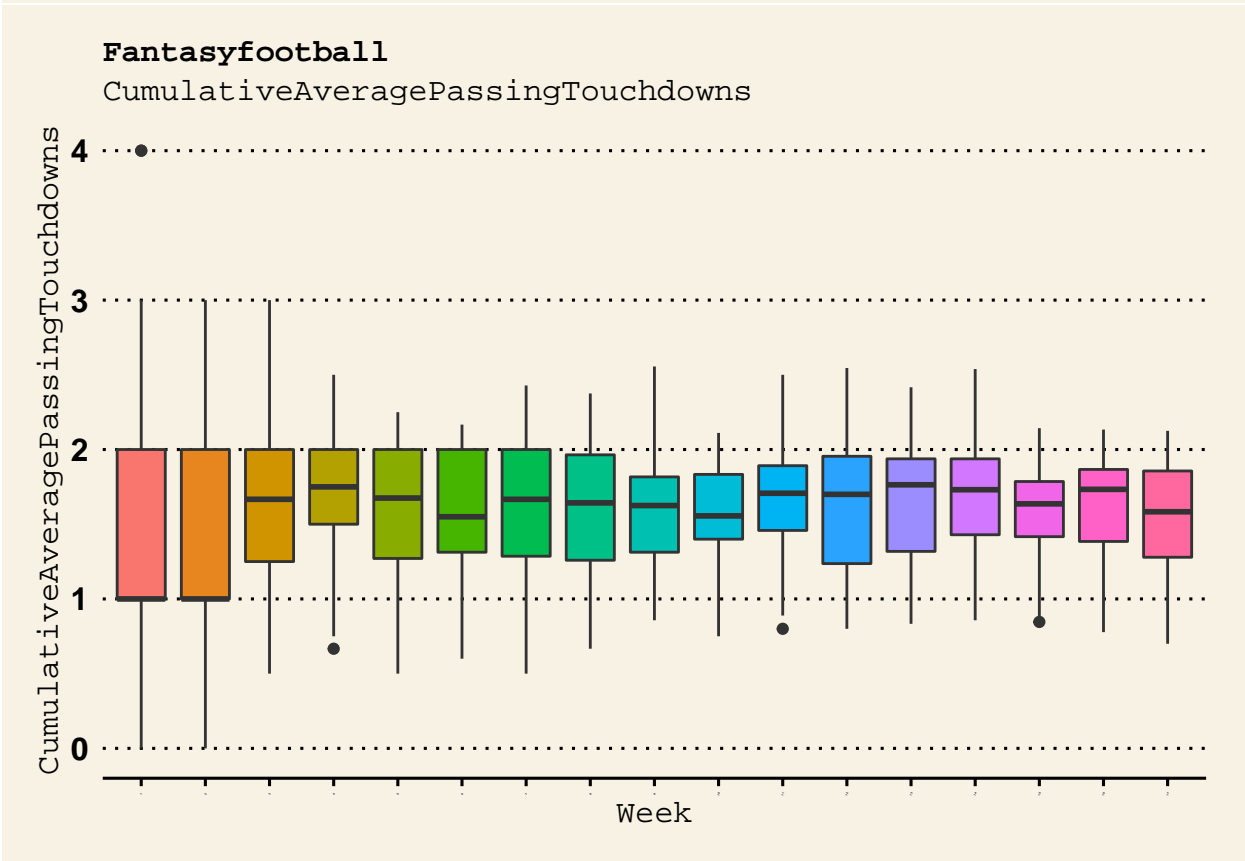
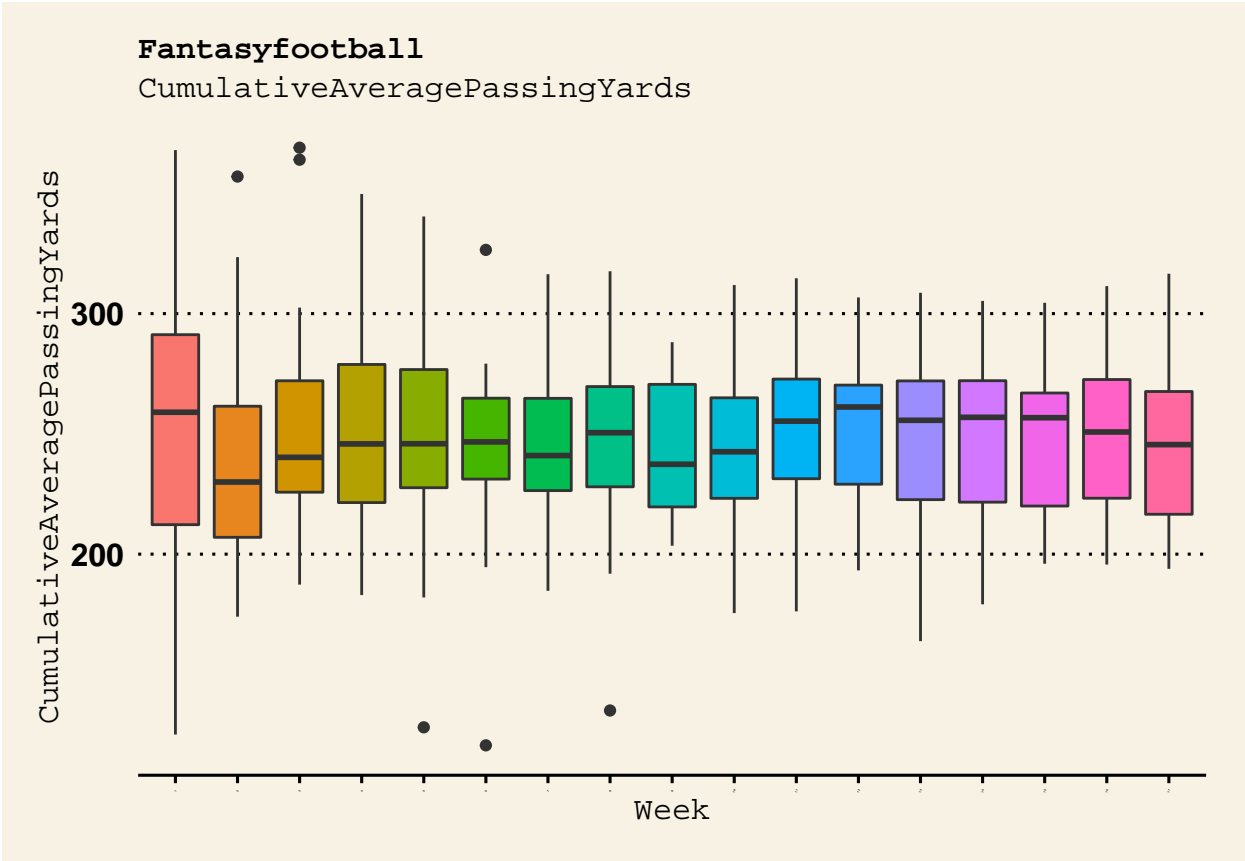
```

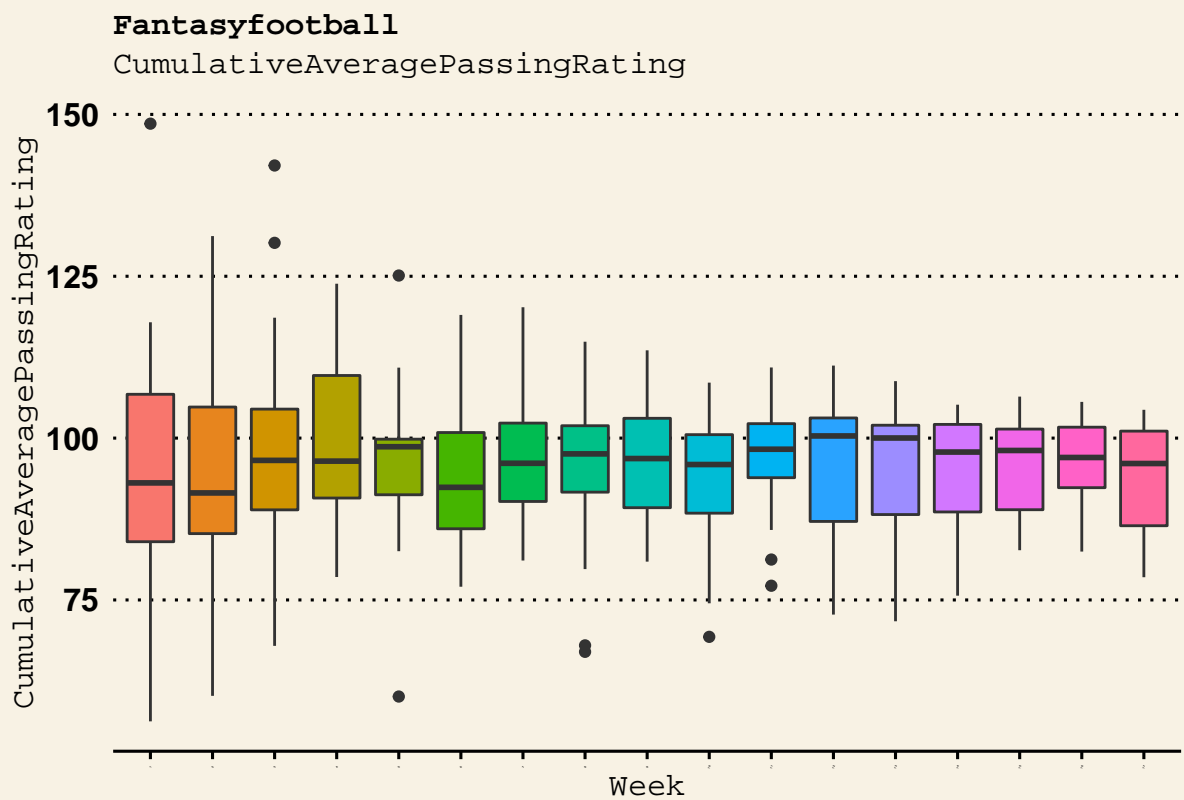
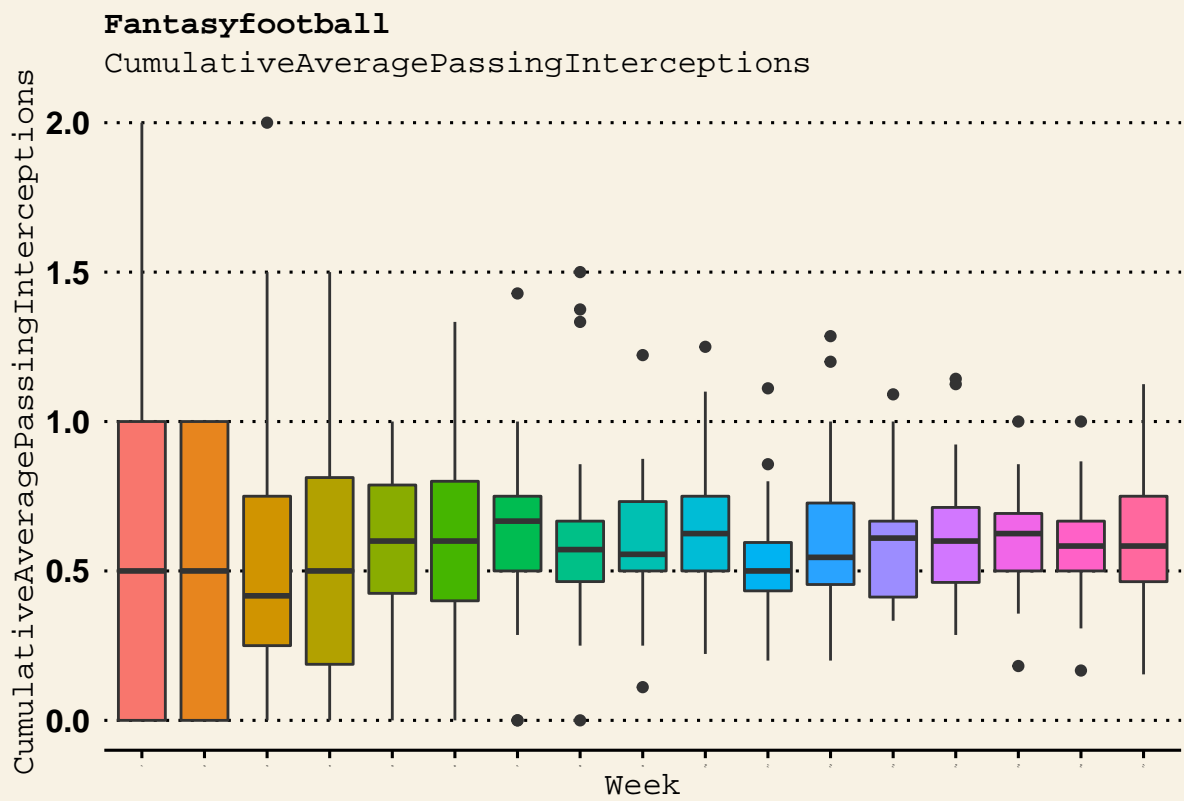
    theme(plot.title      = element_text(size = rel(0.5)),
          plot.subtitle   = element_text(size = rel(0.5)),
          axis.text.x     = element_text(angle=65, vjust=0.6,size=1),
          axis.title       = element_text(size = rel(0.5)),
          legend.position  = "right",
          legend.direction = "vertical",
          legend.title     = element_text(size = rel(0.5)))

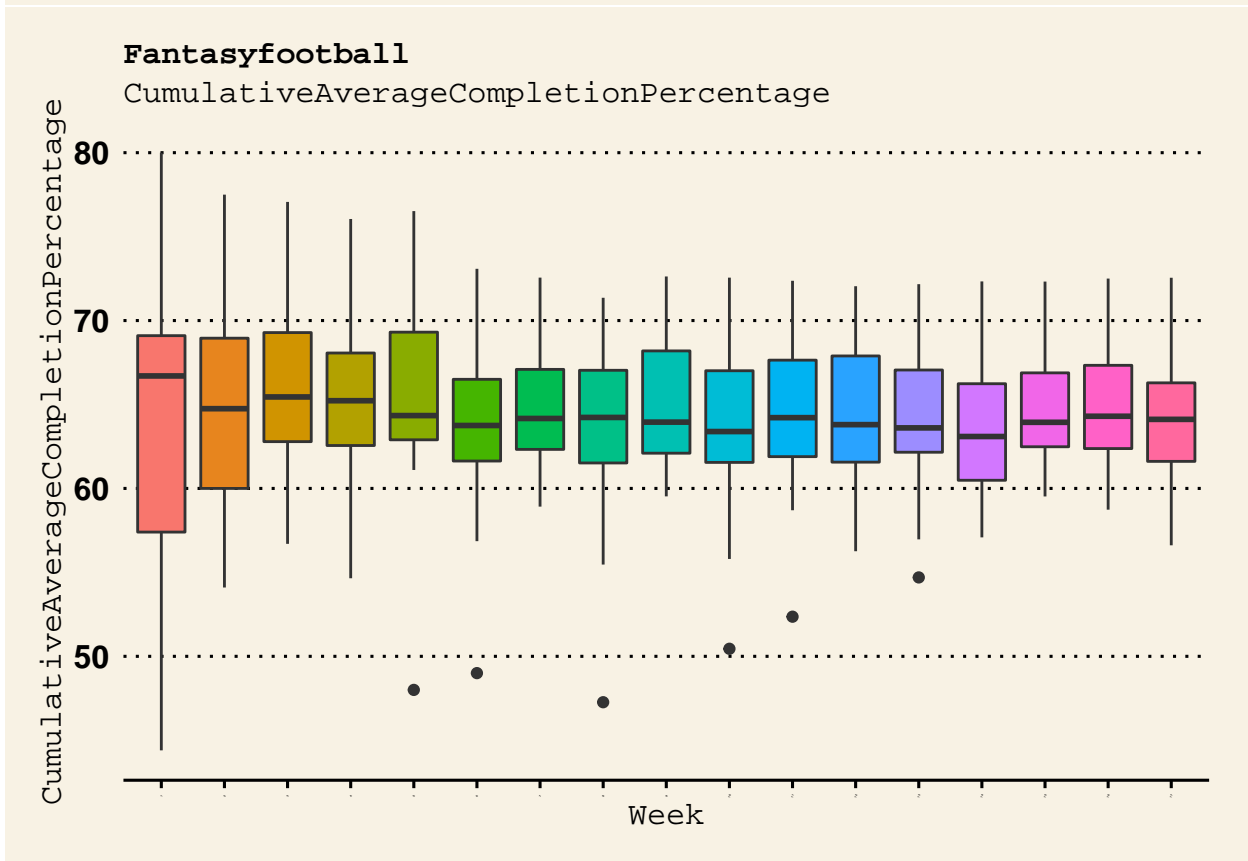
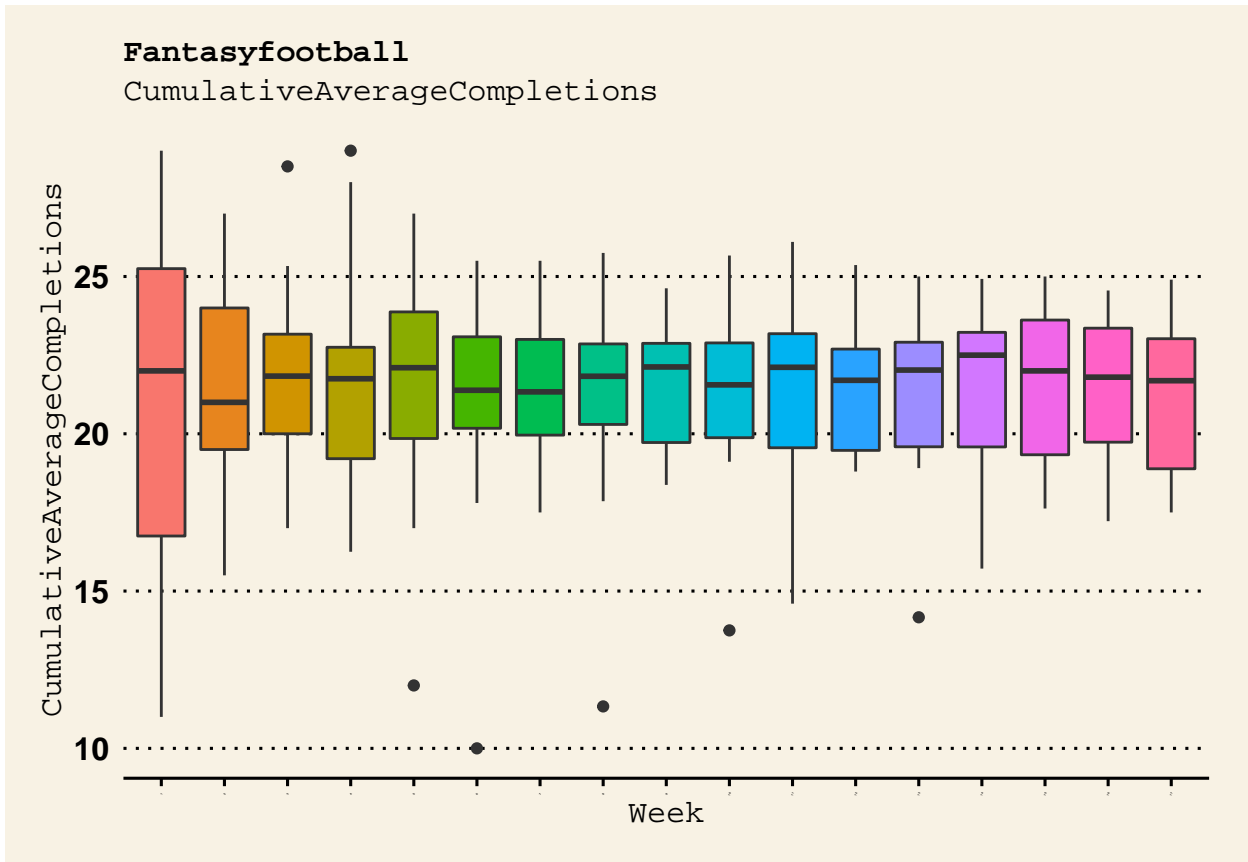
print(ggplotp)
}

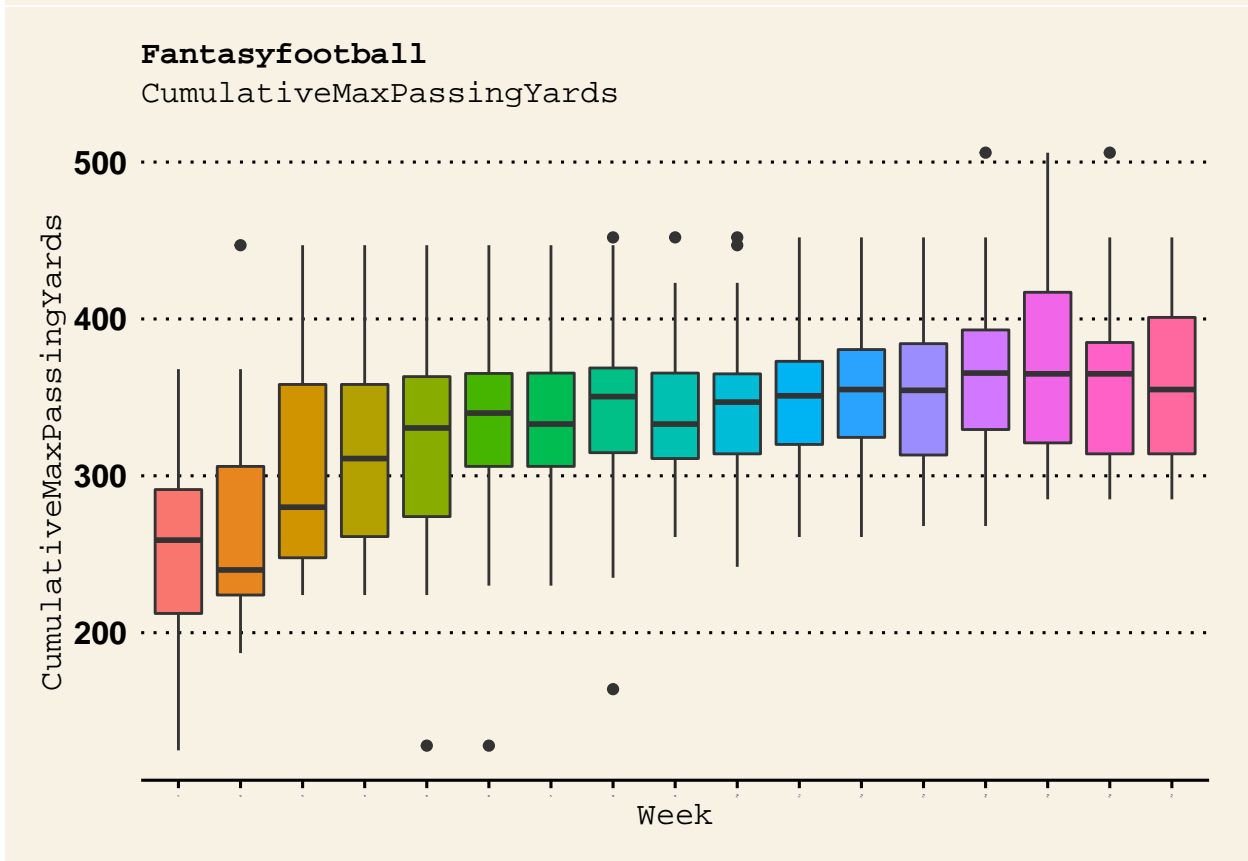
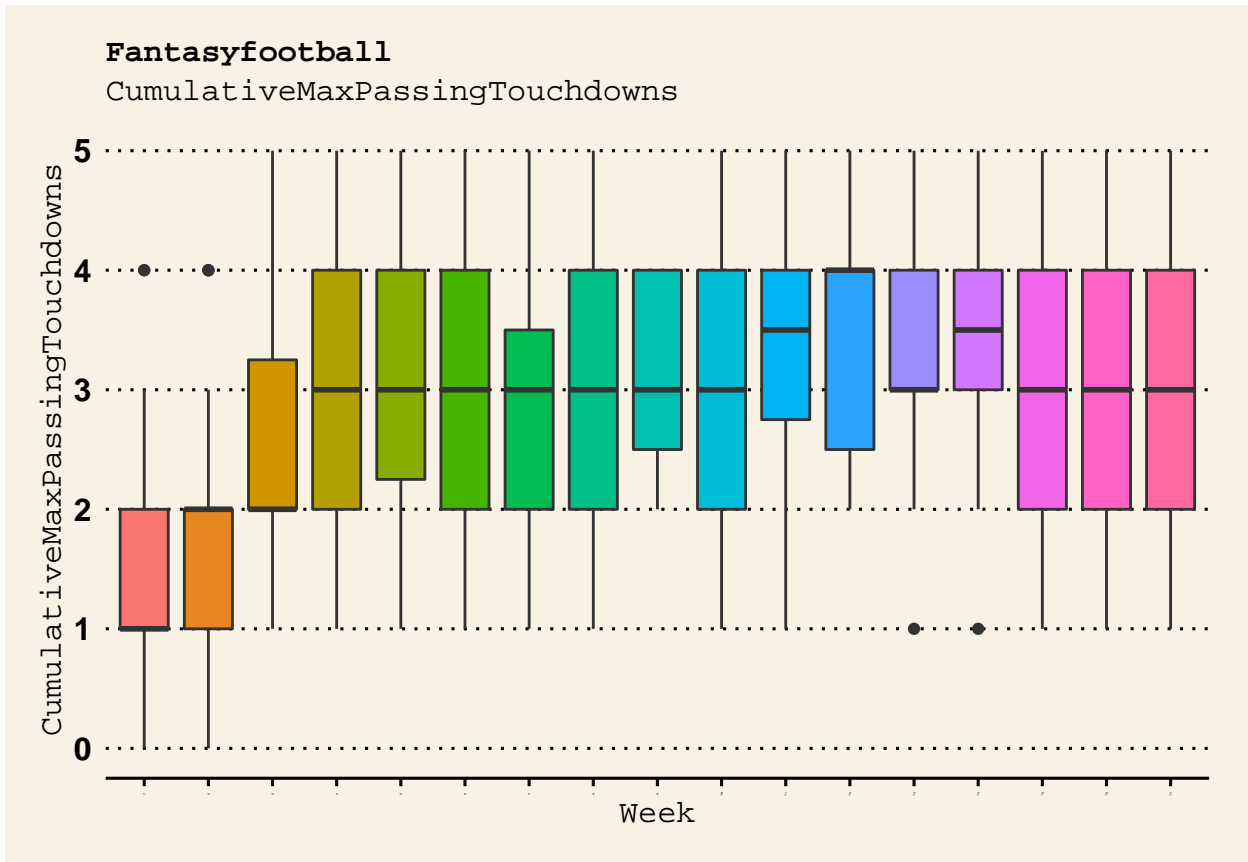
```





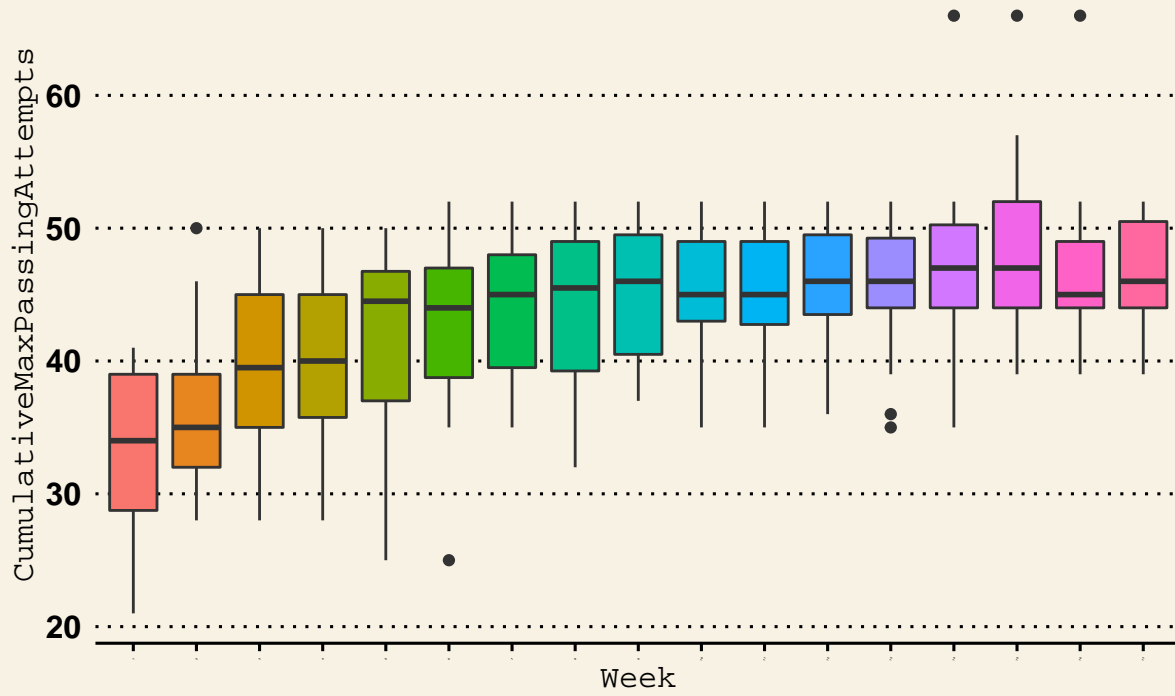






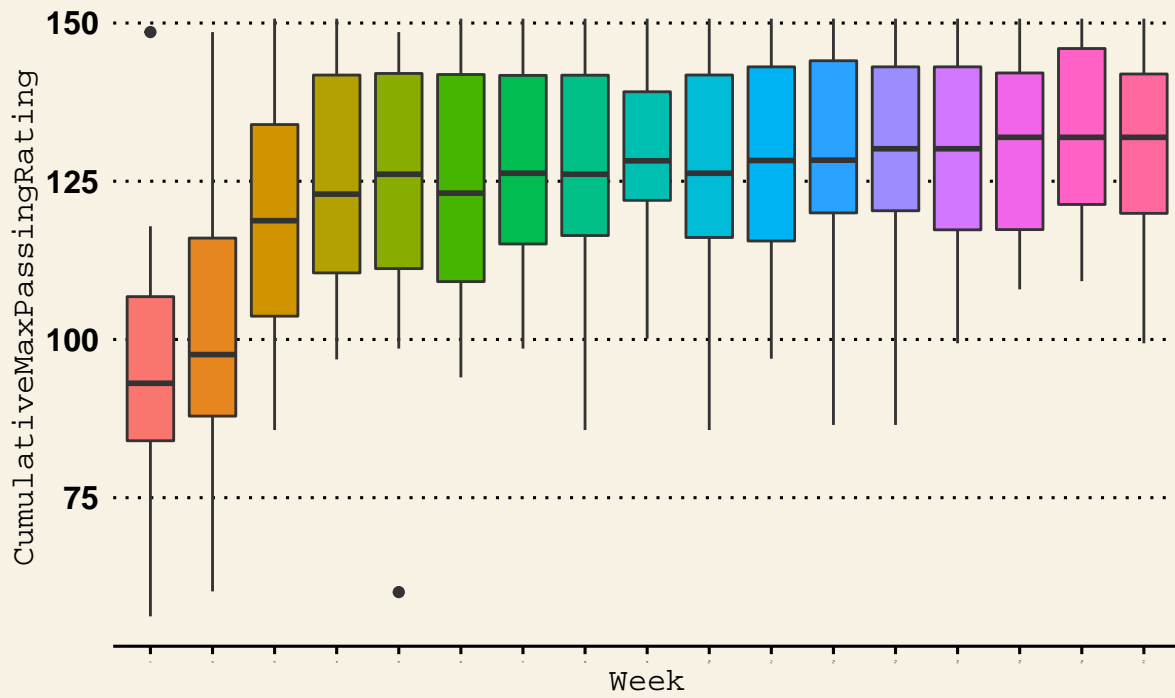
### Fantasyfootball

CumulativeMaxPassingAttempts

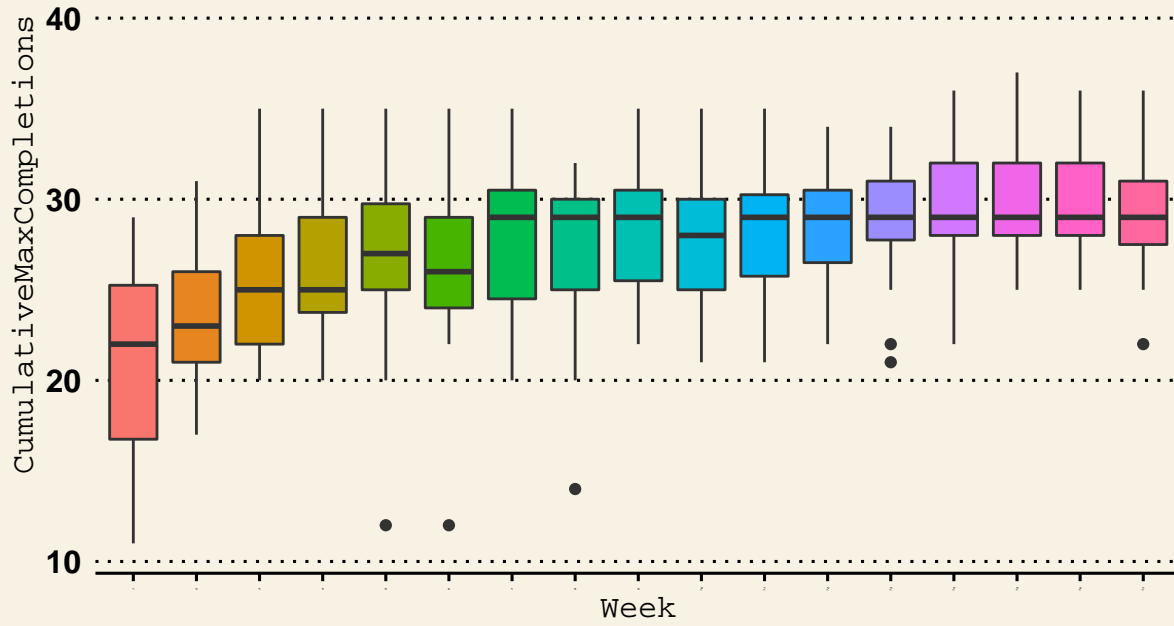


### Fantasyfootball

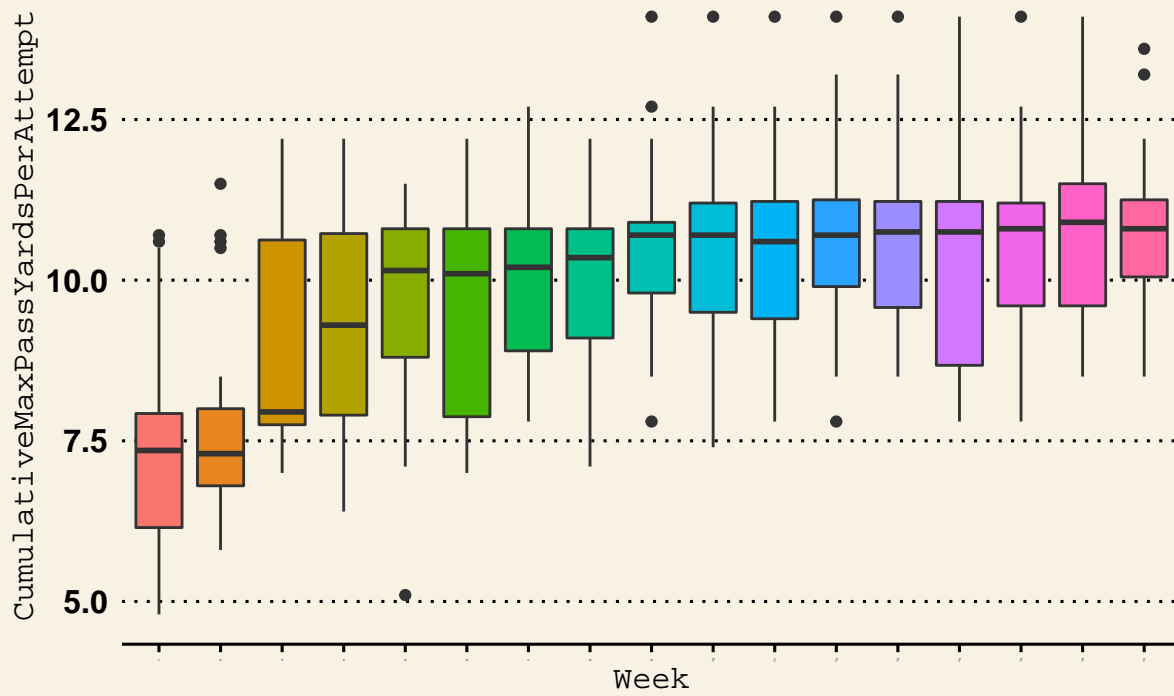
CumulativeMaxPassingRating

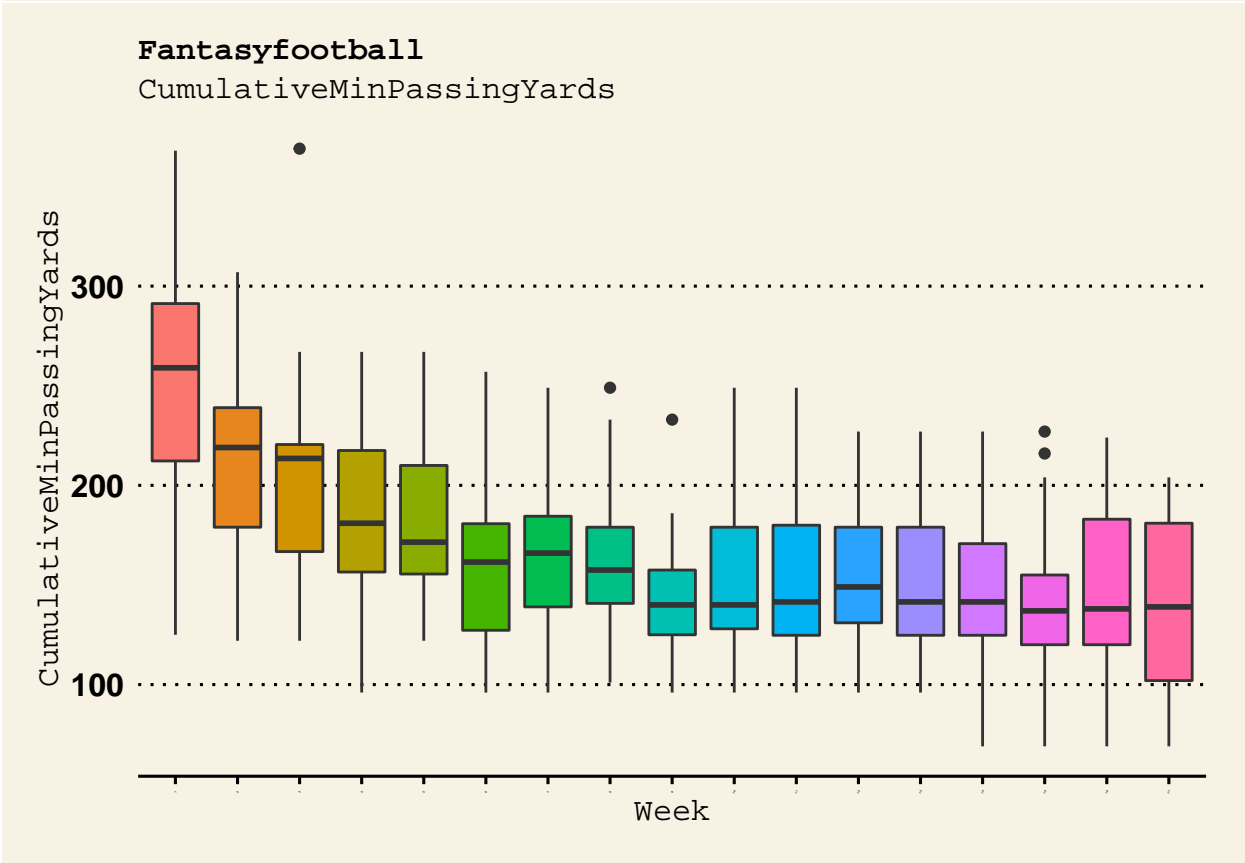
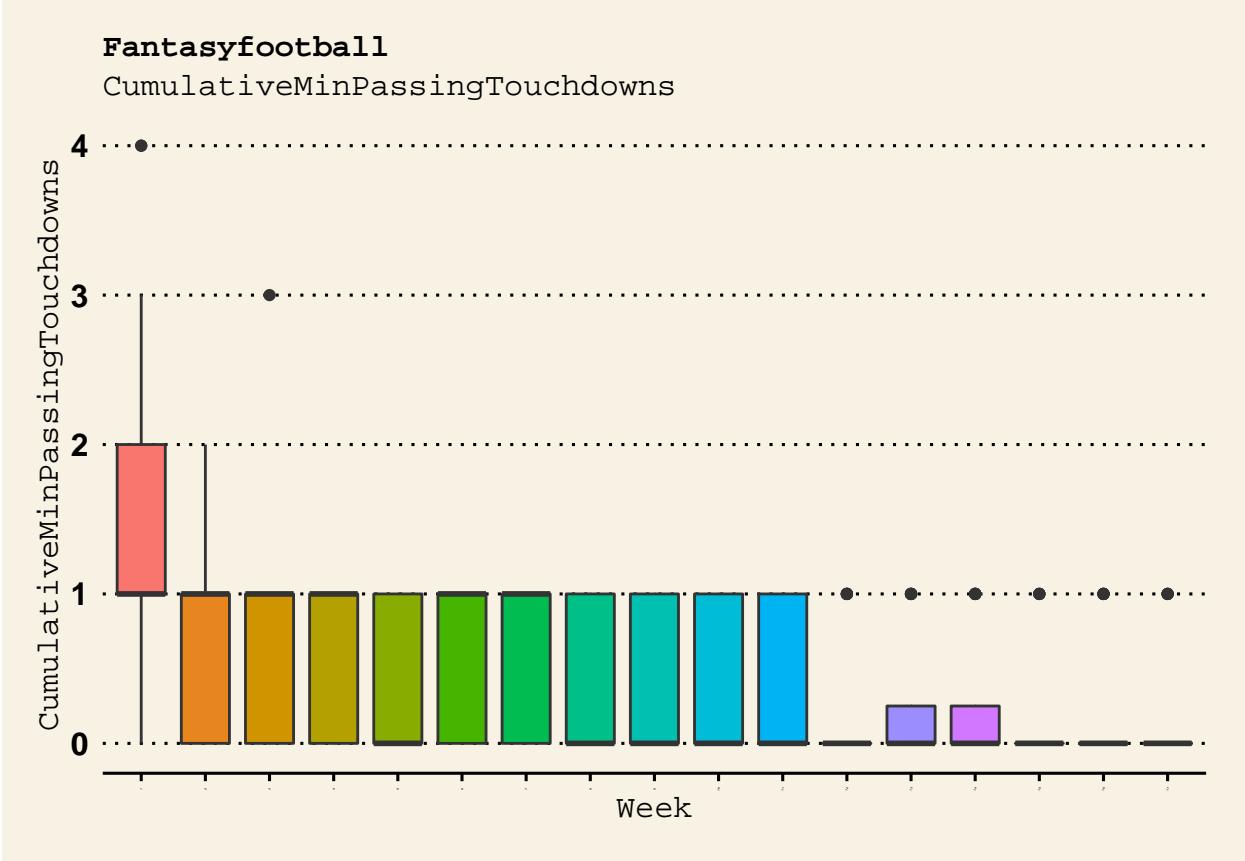


**Fantasyfootball**  
CumulativeMaxCompletions

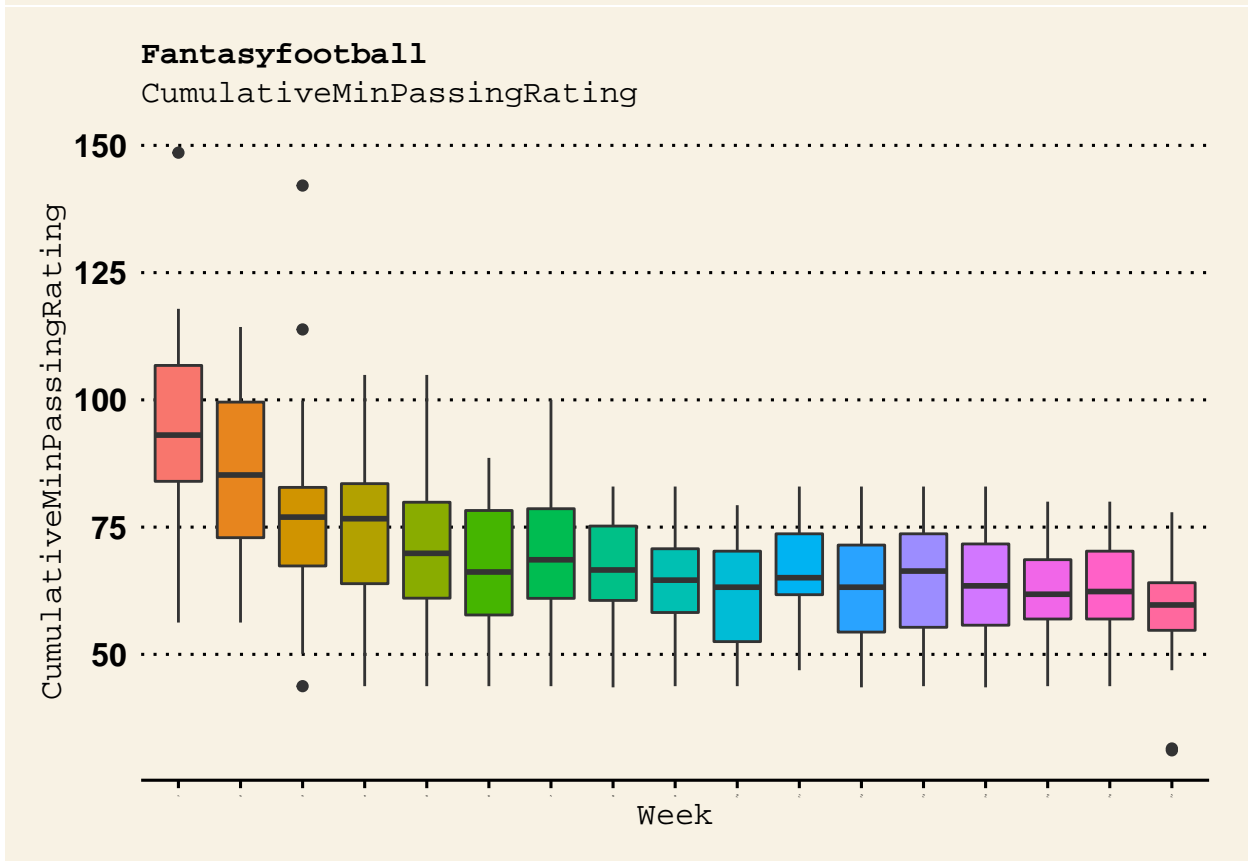
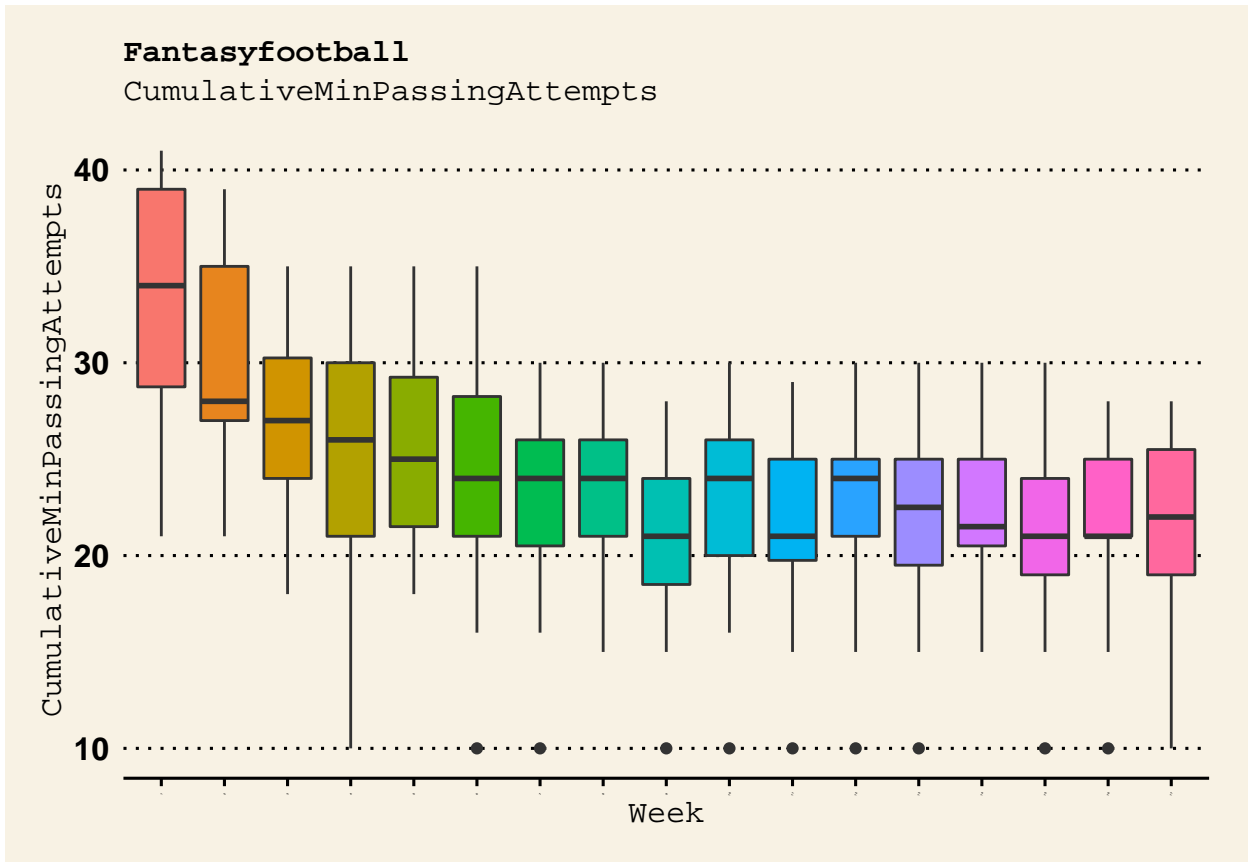


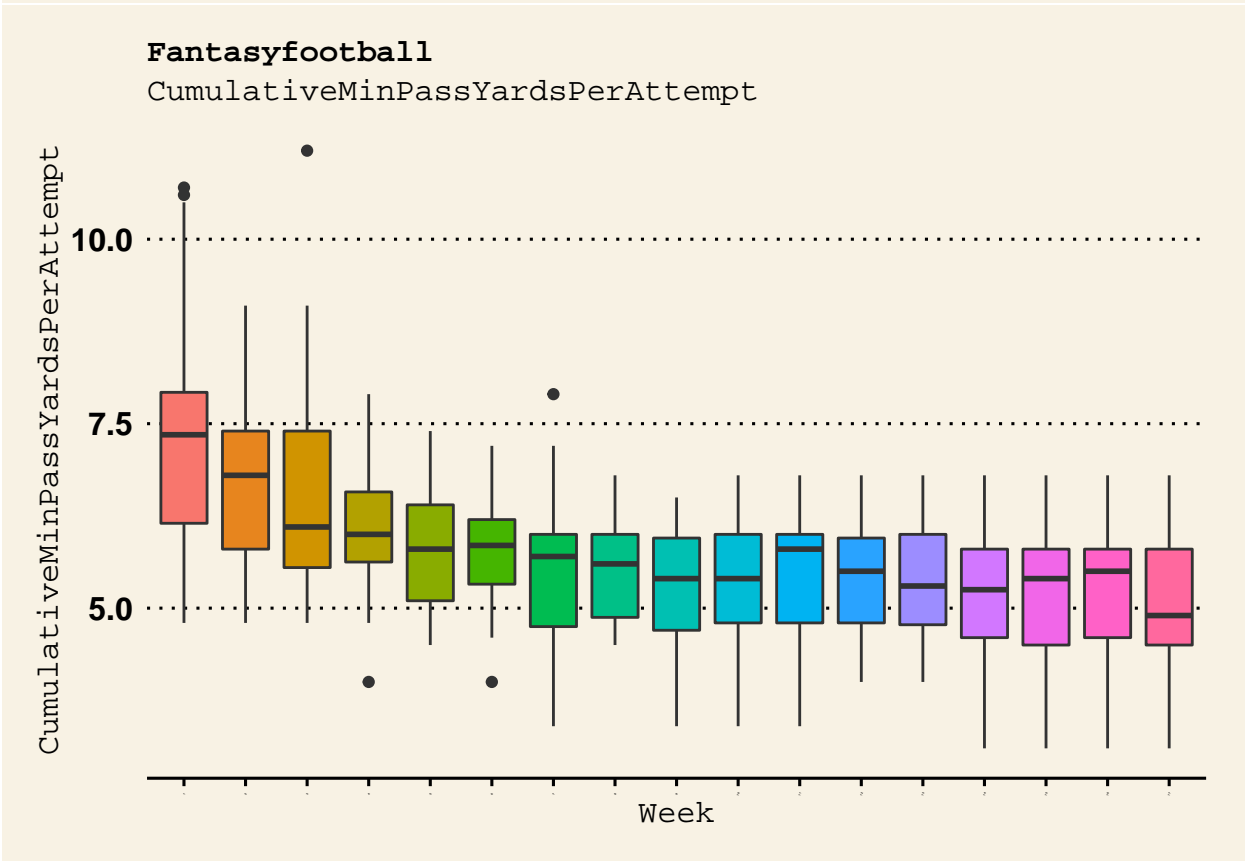
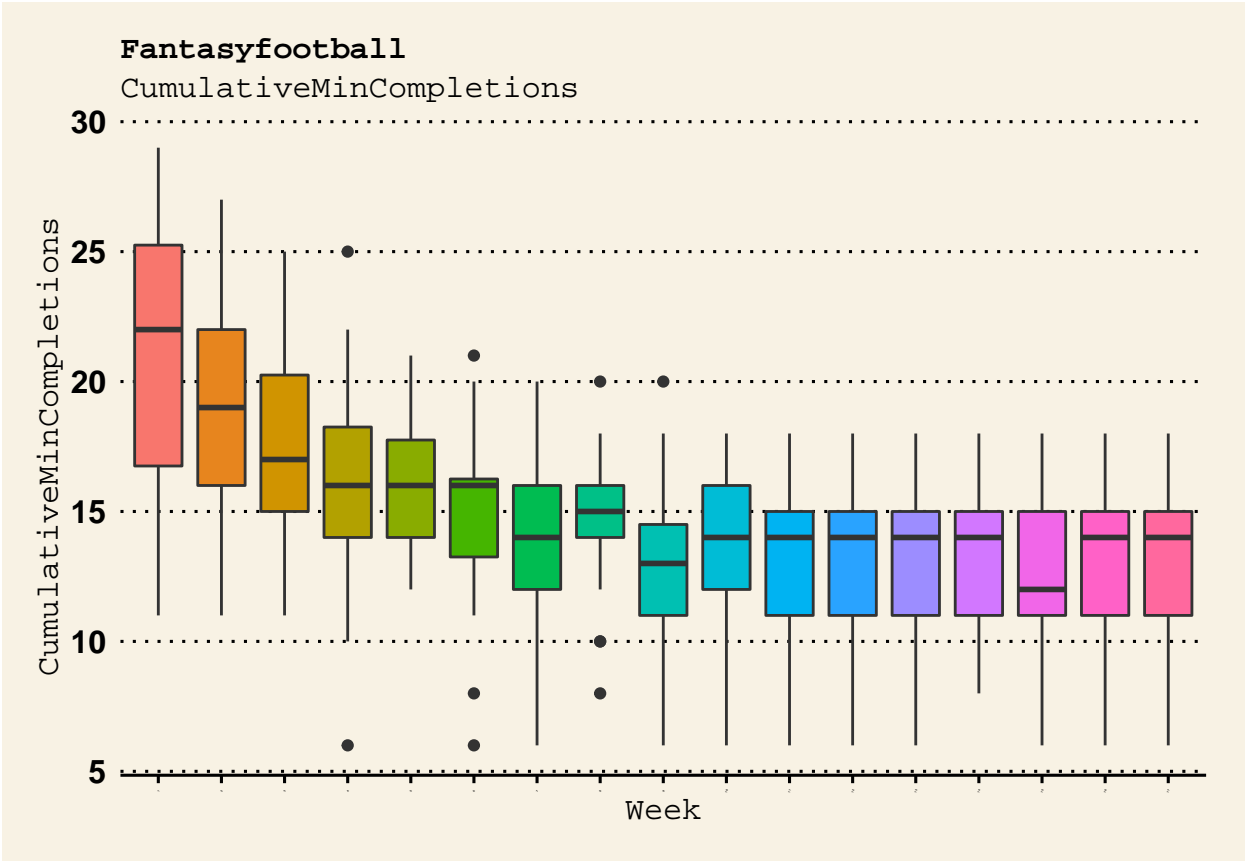
**Fantasyfootball**  
CumulativeMaxPassYardsPerAttempt





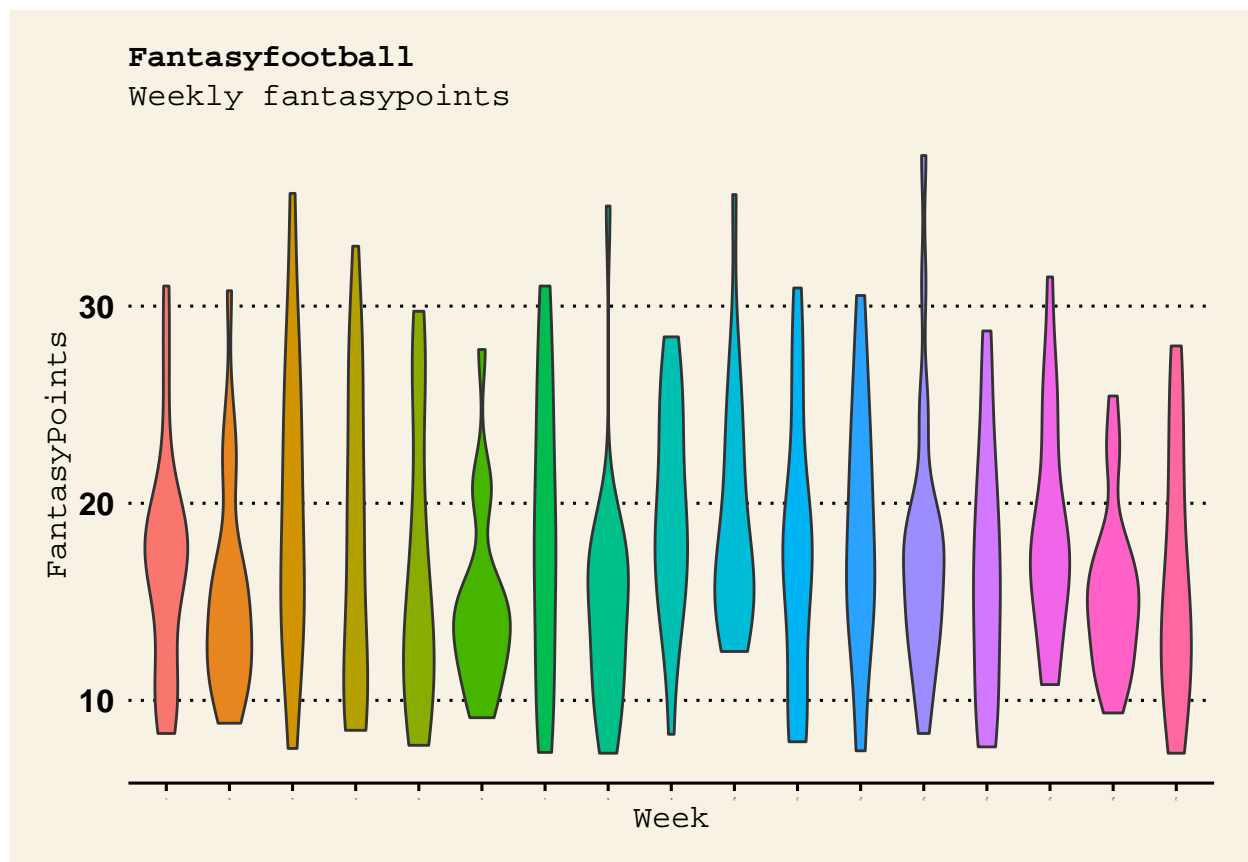






### ## 8.3 Violin plots Derived Feature set

```
eda_derived %>% ggplot(aes(y=FantasyPoints,x=Week,fill=Week,group=Week))+
  geom_violin(show.legend = FALSE)+
  xlab("Week")+ylab("FantasyPoints")+
  labs(title="Fantasyfootball",
        subtitle="Weekly fantasypoints",
        aption="Source: Fantasyfootball")+
  theme_ws() +
  theme(plot.title = element_text(size = rel(0.5)),
        plot.subtitle = element_text(size = rel(0.5)),
        axis.text.x = element_text(angle=65, vjust=0.6,size=1),
        axis.title = element_text(size = rel(0.5)),
        legend.position = "right",
        legend.direction = "vertical",
        legend.title = element_text(size = rel(0.5)))
```



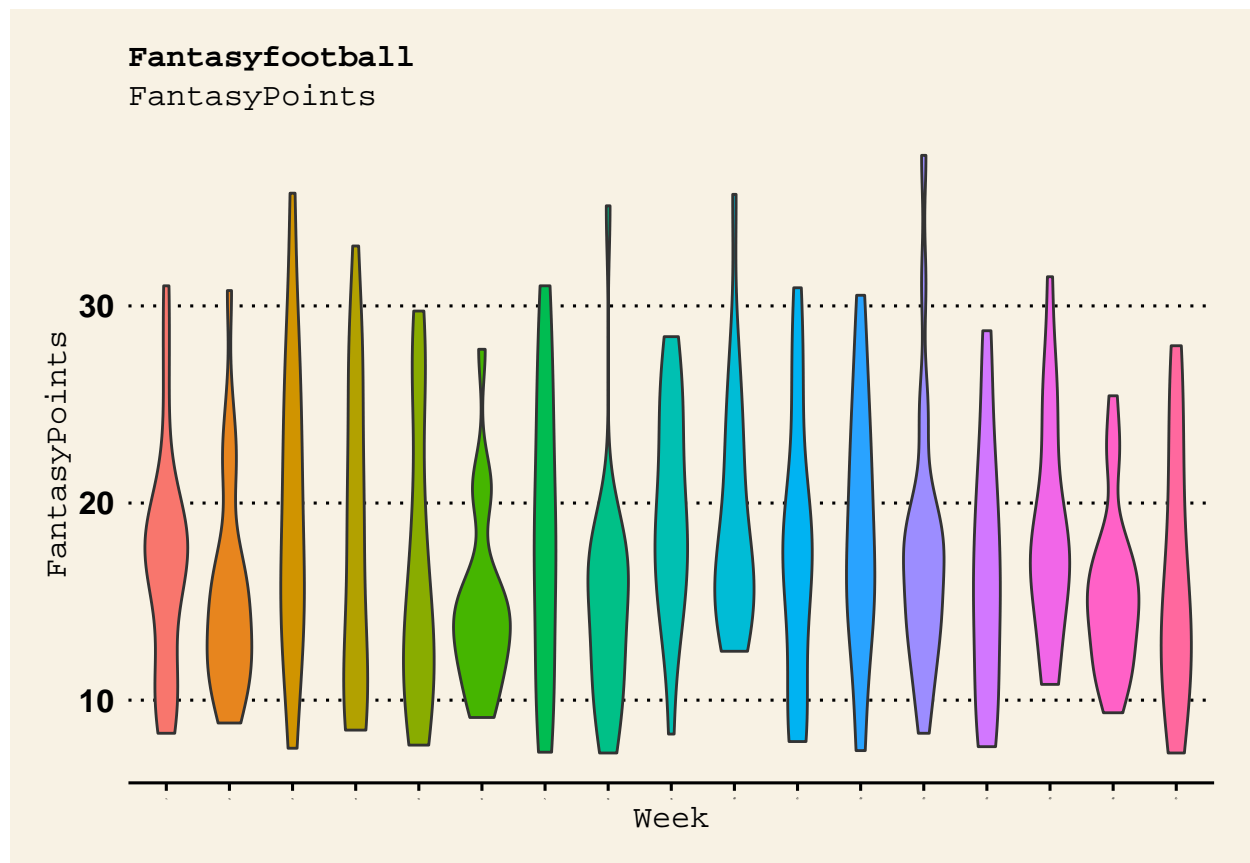
```
for (i in 2:20) {
  ggplotp <- eda_derived %>%
    ggplot(aes_string(y=names(eda_derived[i]),x="Week",fill="Week",group="Week"))+
    geom_violin(show.legend = FALSE)+
    xlab("Week")+ylab(names(eda_derived[i]))+
    labs(title="Fantasyfootball",
          subtitle=names(eda_derived[i]),
          aption="Source: Fantasyfootball")+
    theme_ws() +
```

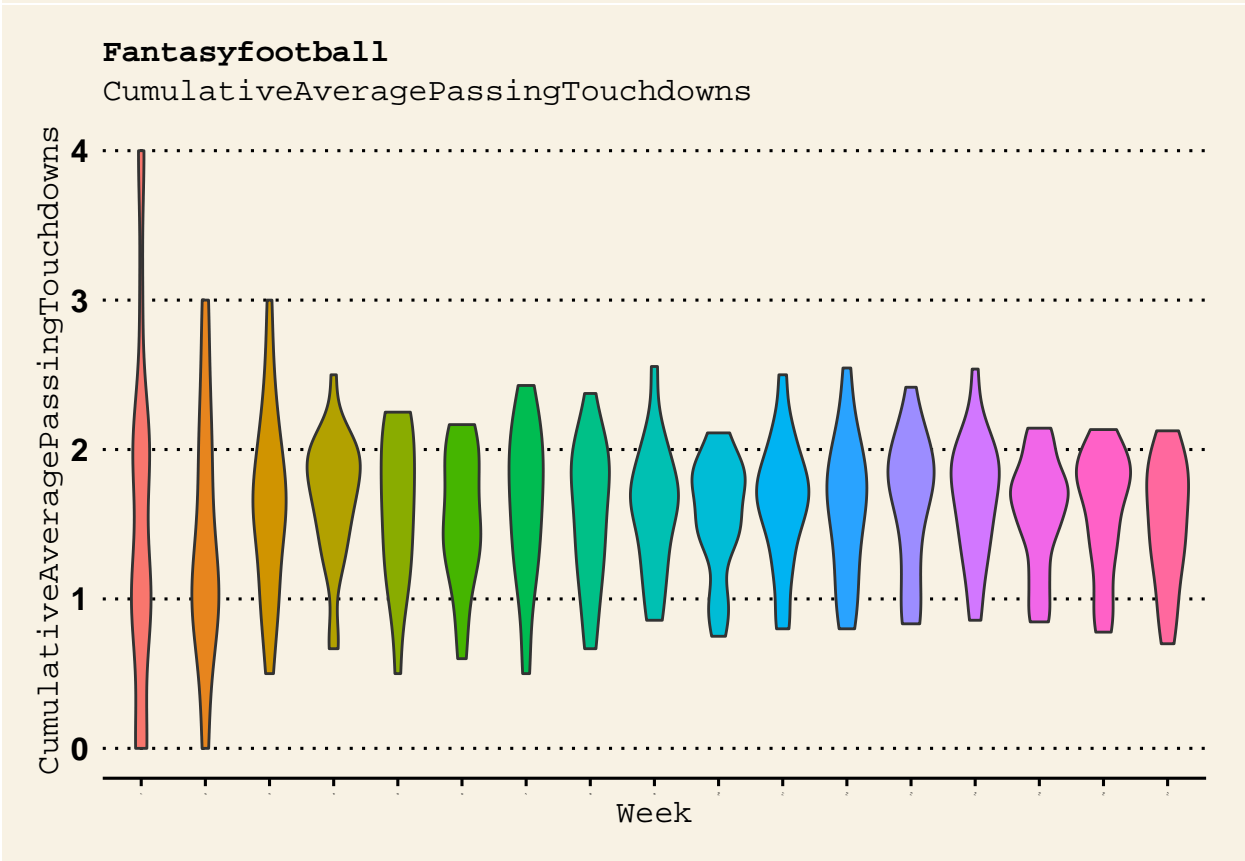
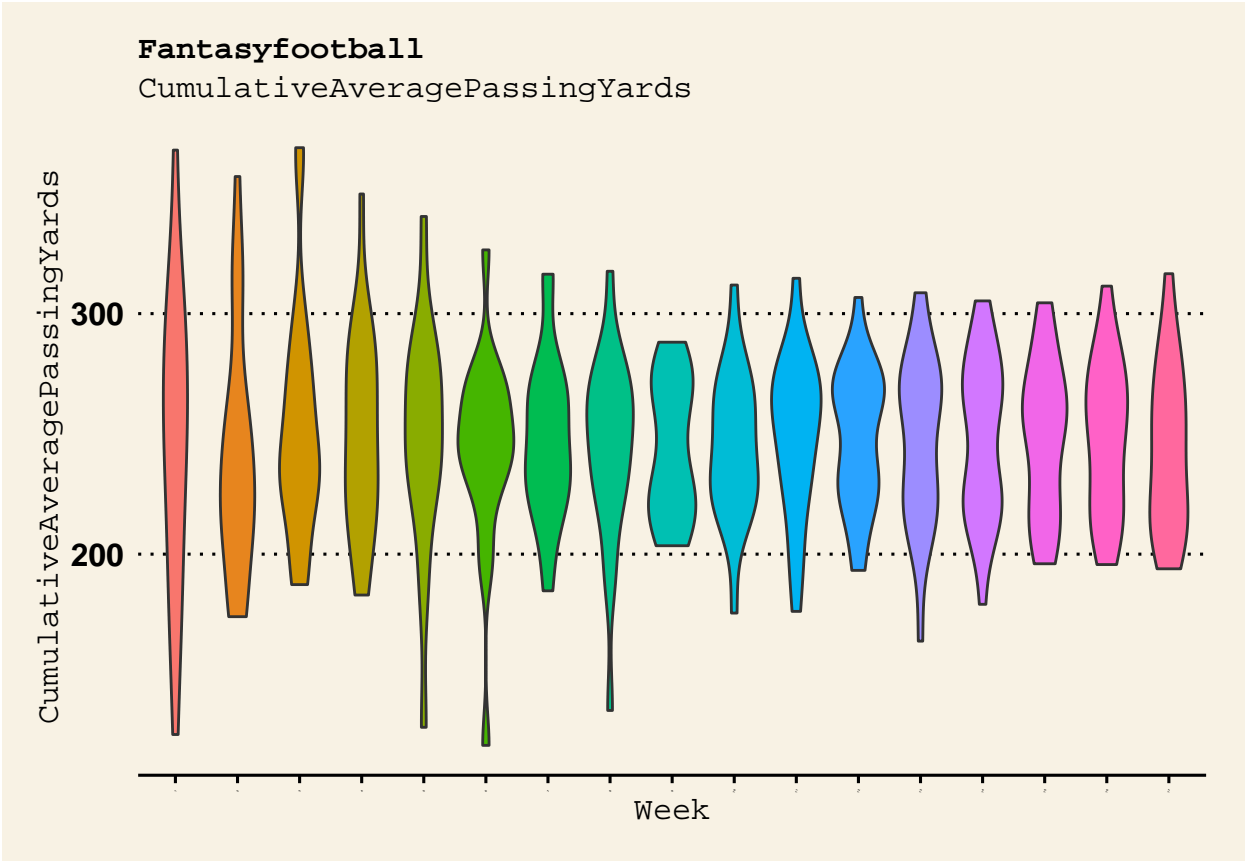
```

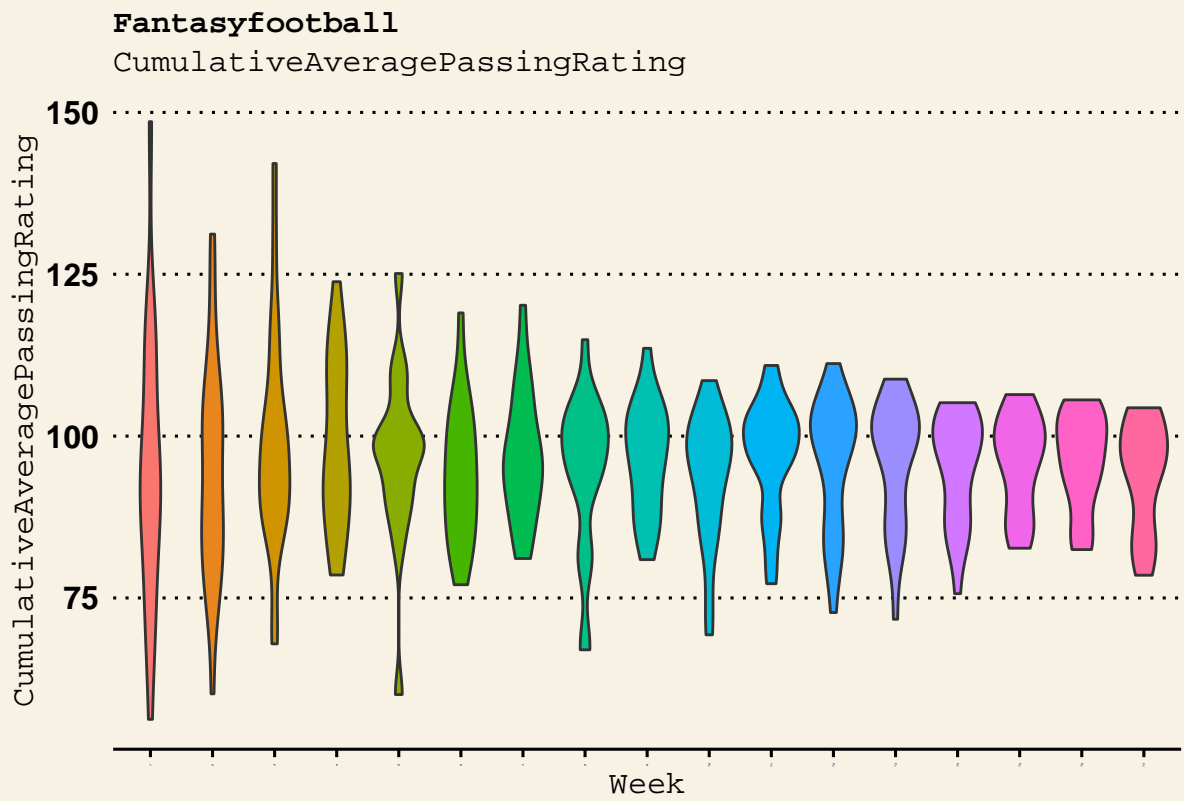
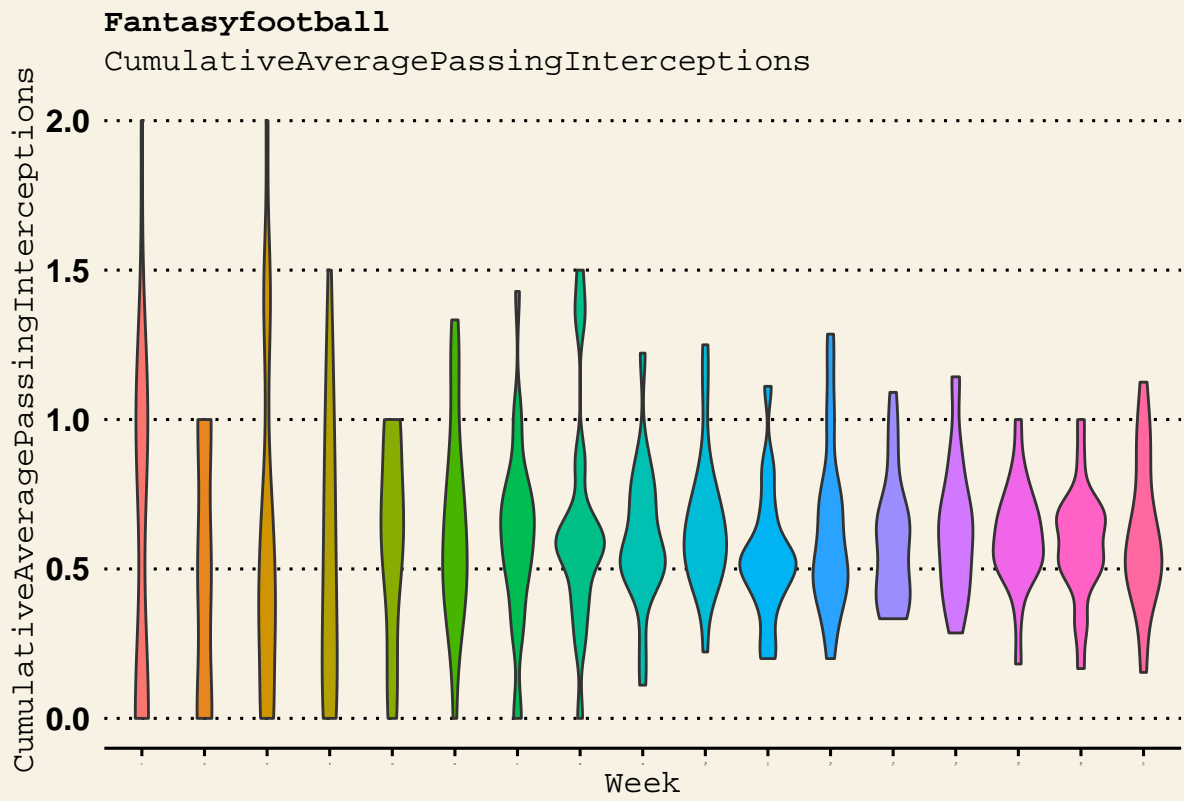
    theme(plot.title      = element_text(size = rel(0.5)),
          plot.subtitle   = element_text(size = rel(0.5)),
          axis.text.x     = element_text(angle=65, vjust=0.6,size=1),
          axis.title       = element_text(size = rel(0.5)),
          legend.position  = "right",
          legend.direction = "vertical",
          legend.title     = element_text(size = rel(0.5)))

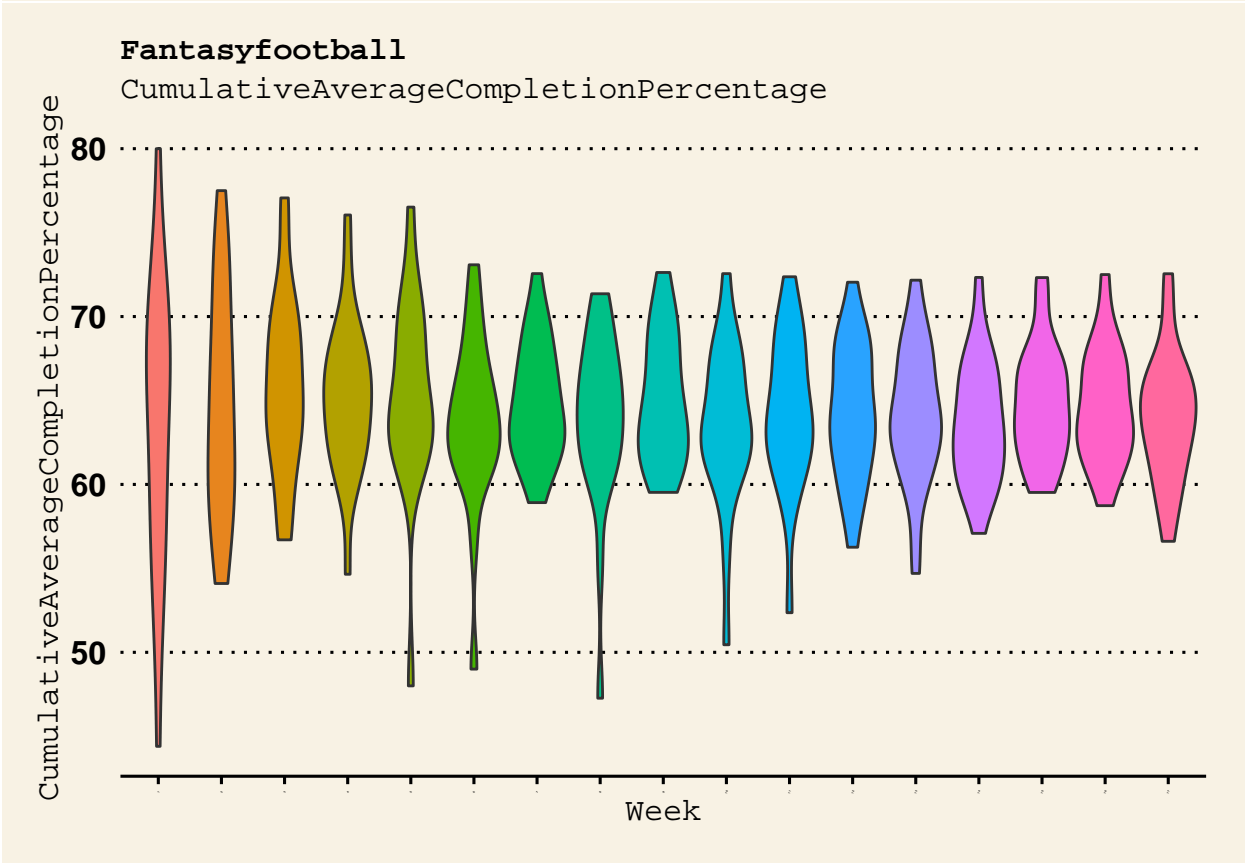
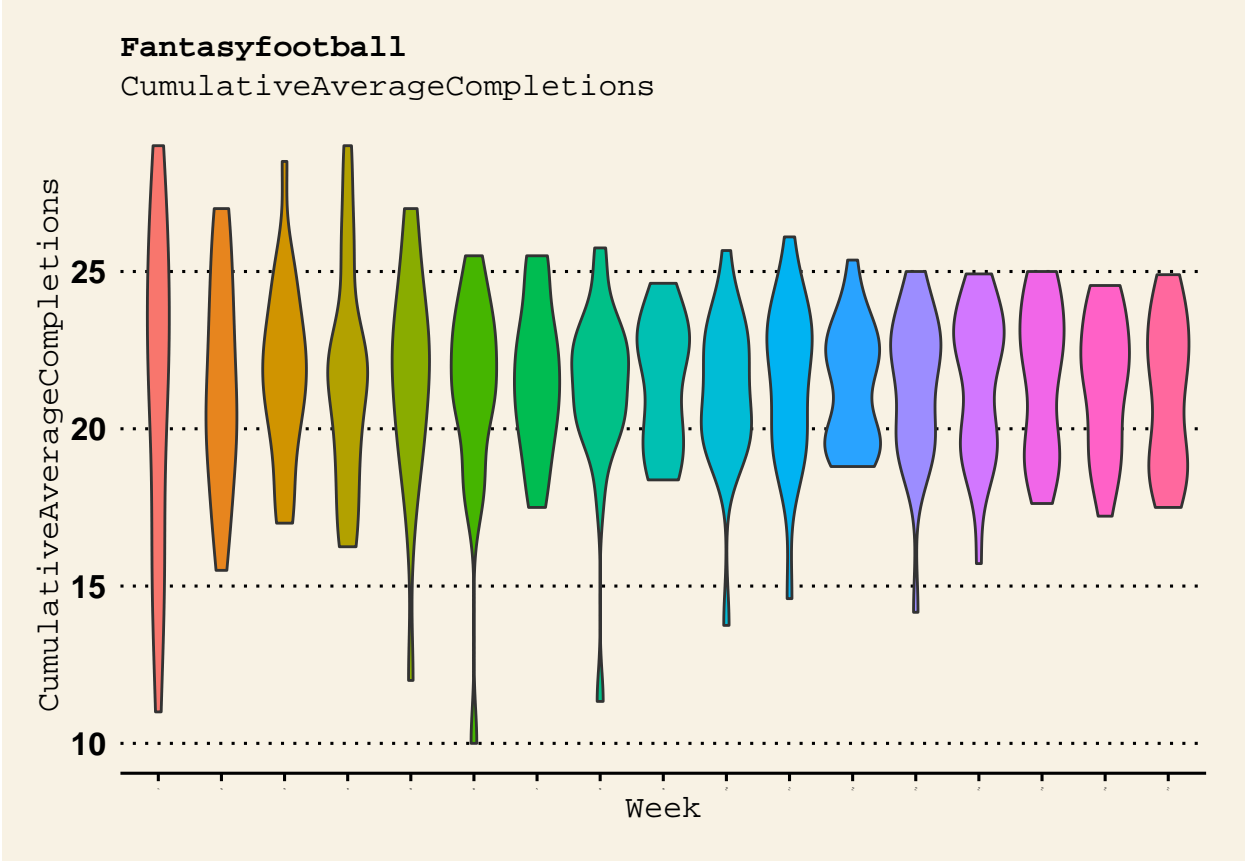
    print(ggplotp)
  }

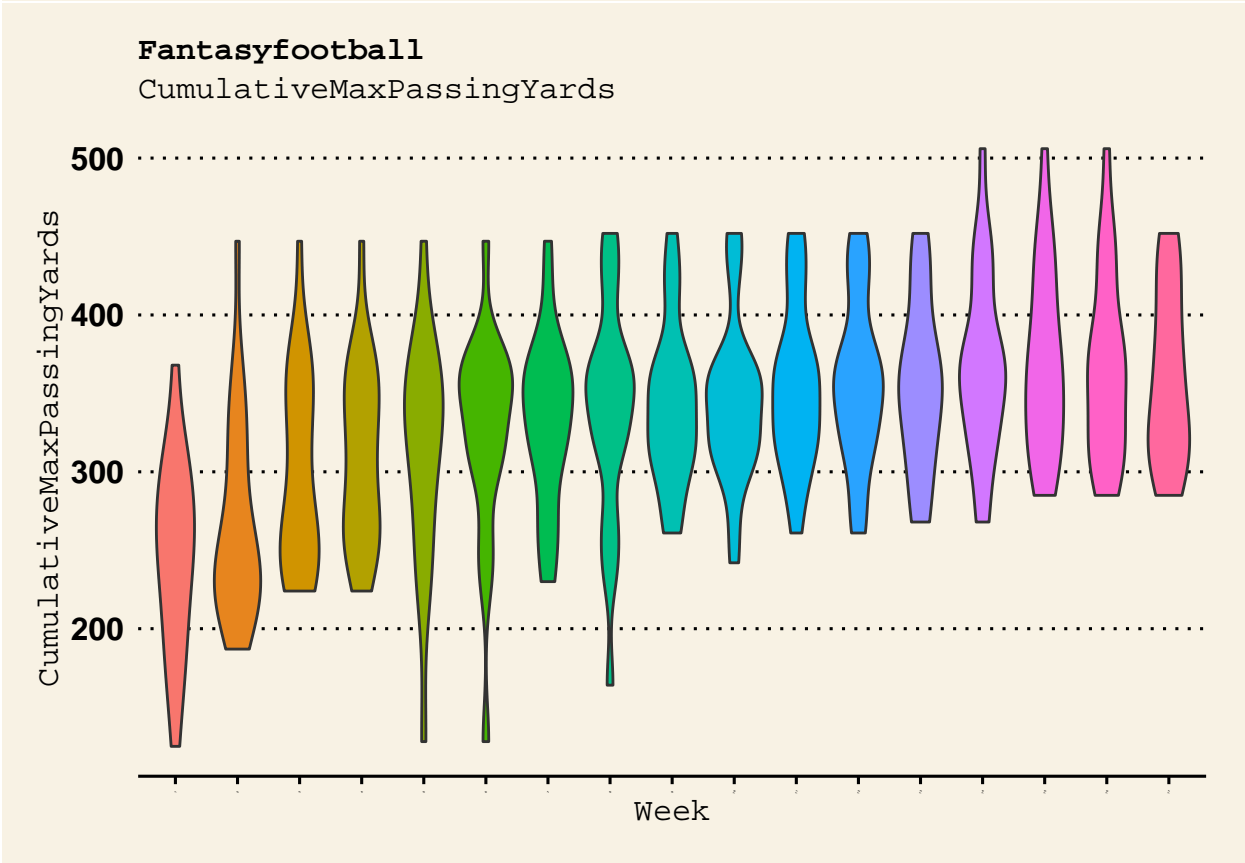
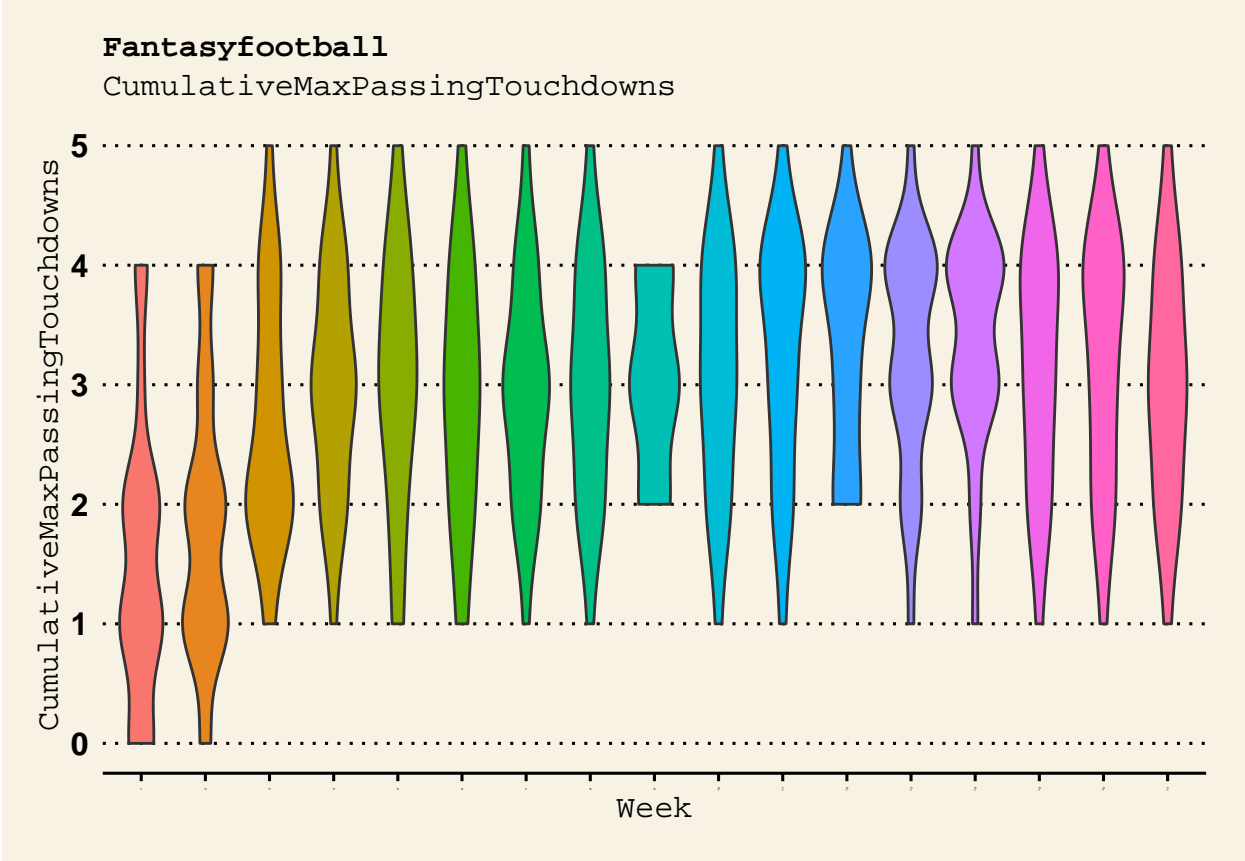
```







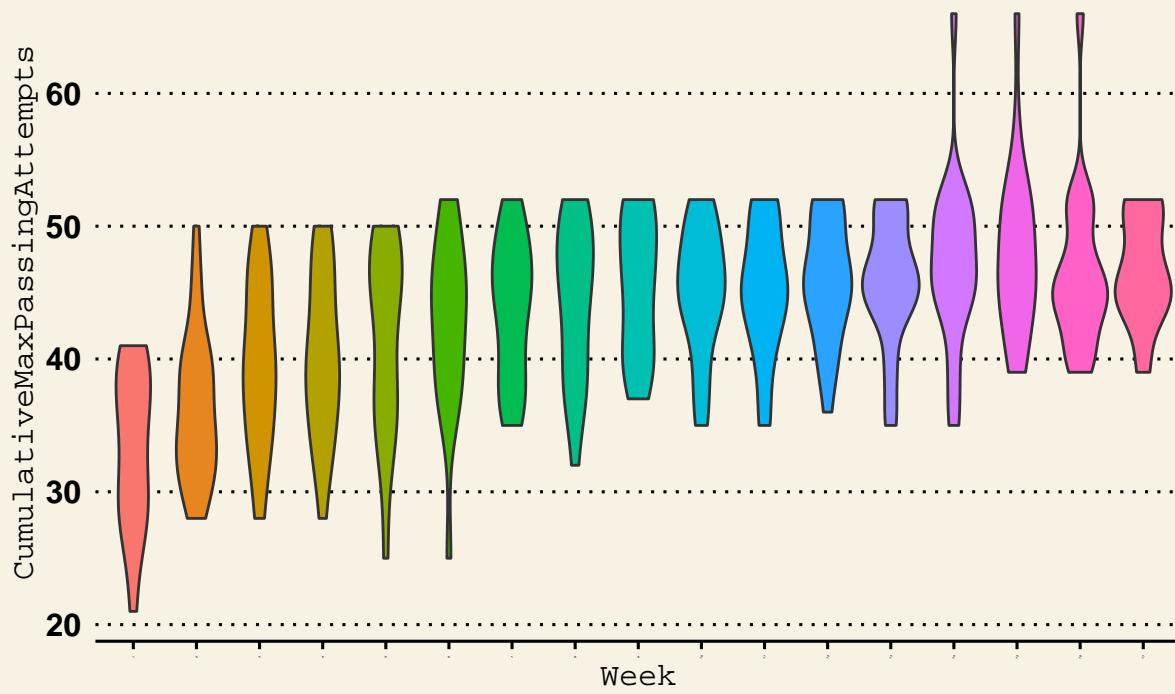






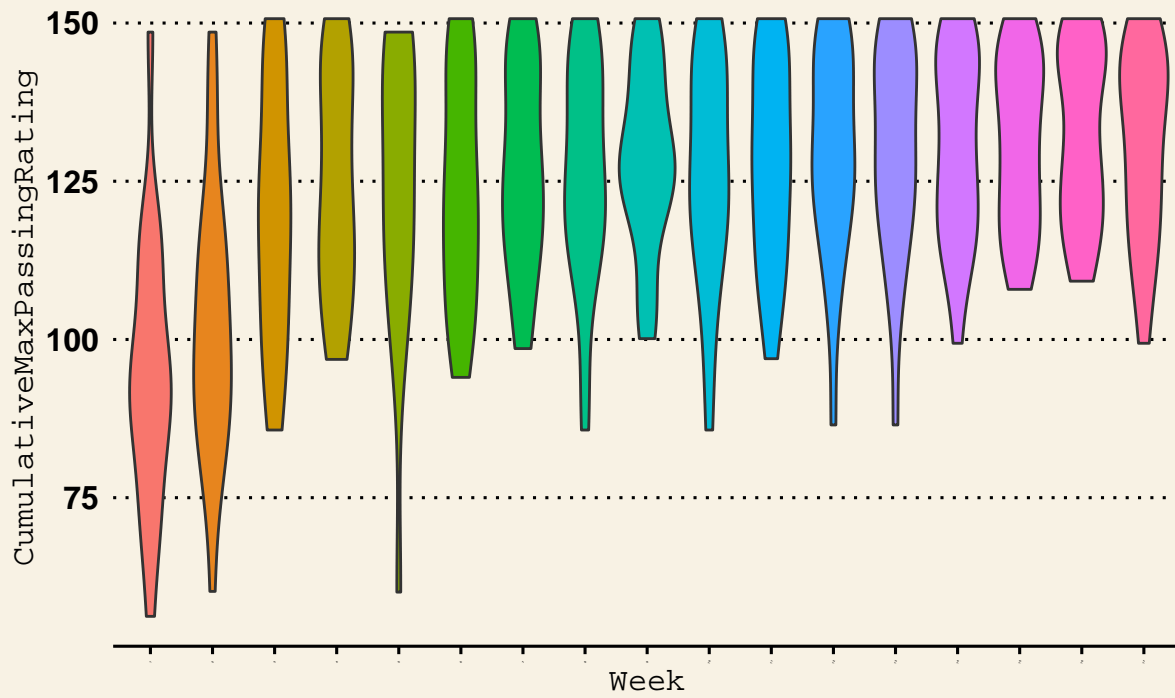
## Fantasyfootball

CumulativeMaxPassingAttempts

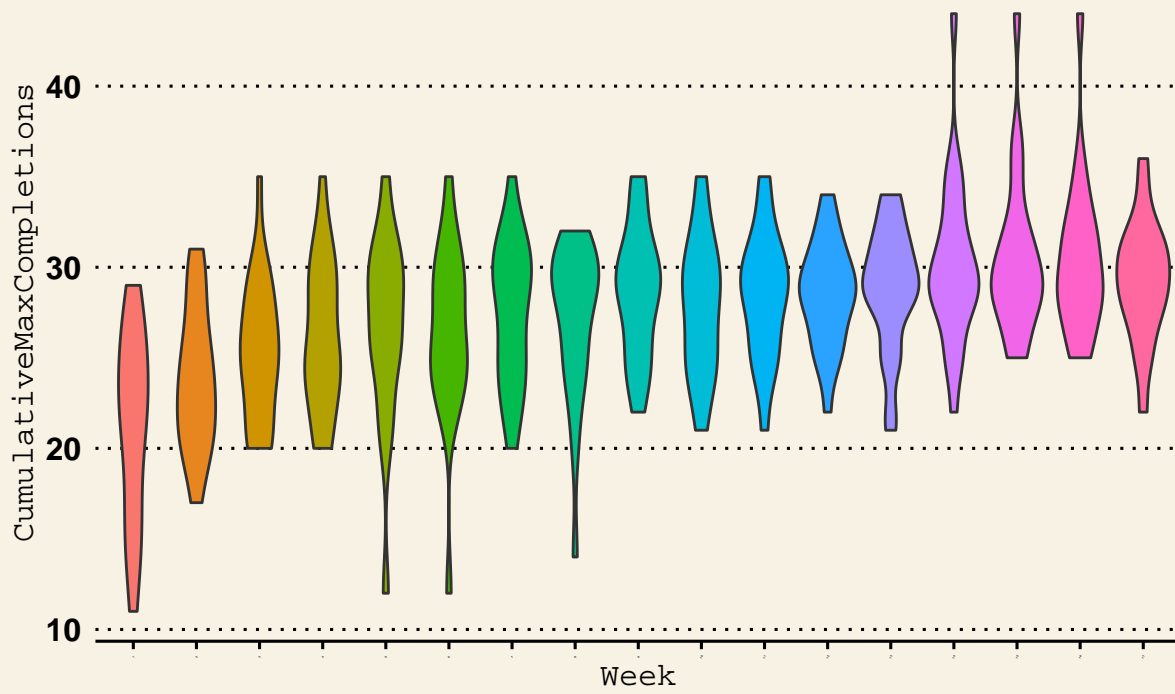


## Fantasyfootball

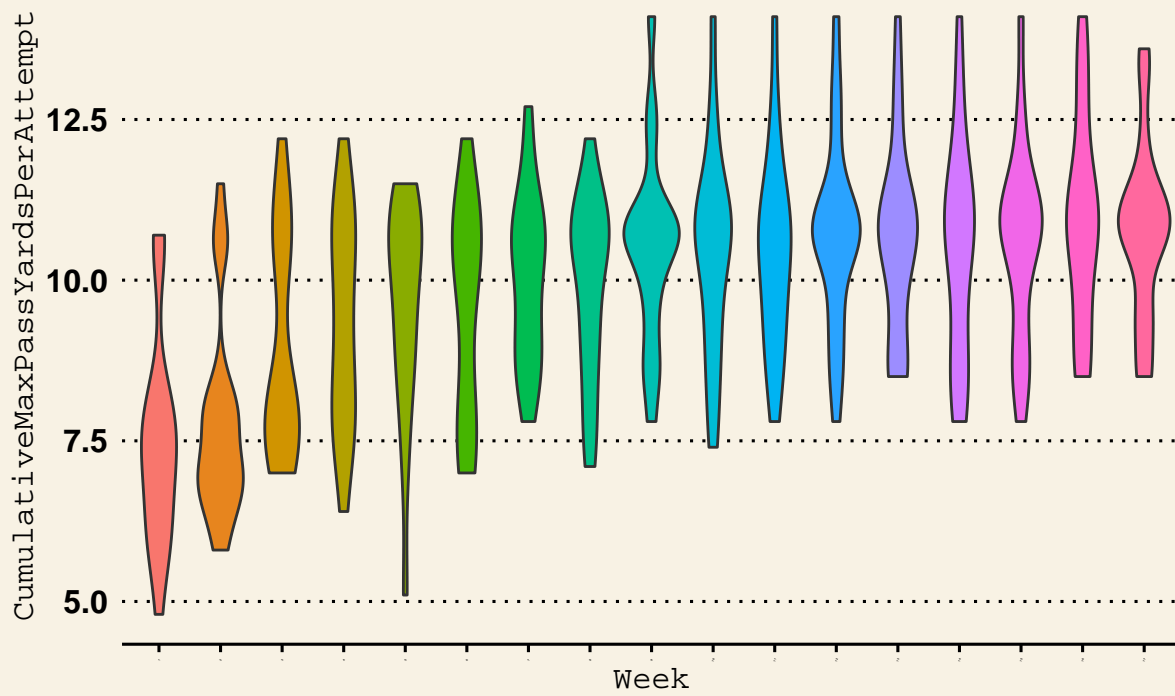
CumulativeMaxPassingRating



**Fantasyfootball**  
CumulativeMaxCompletions

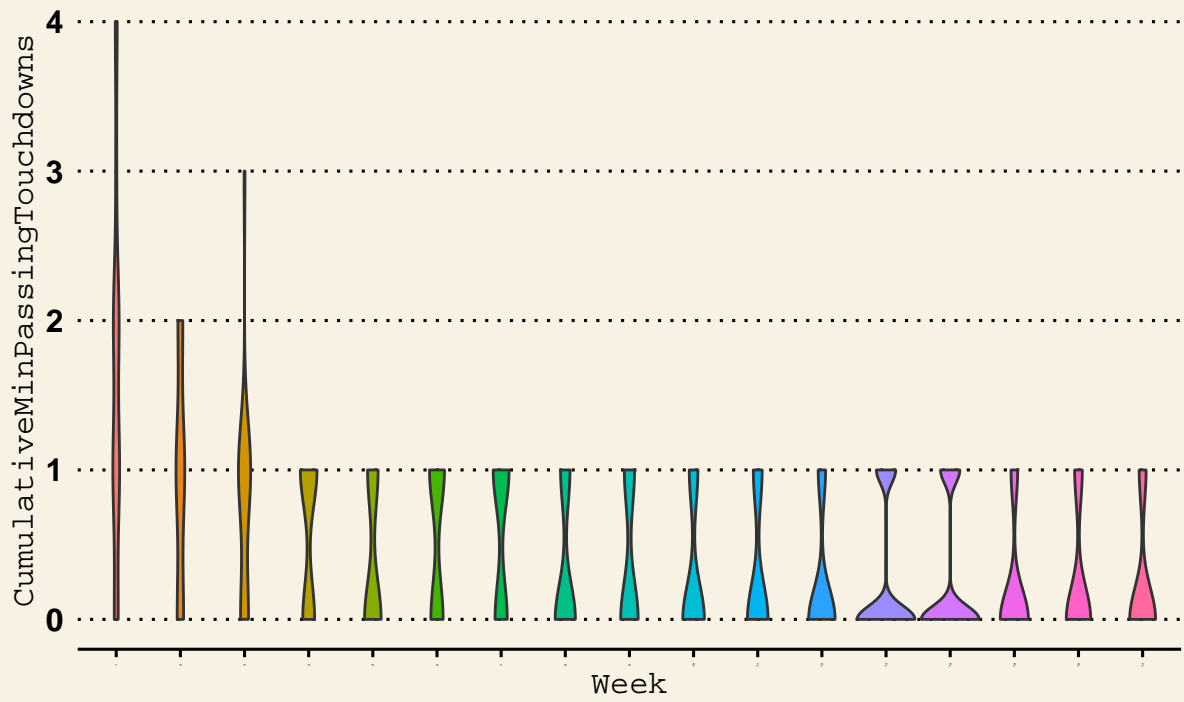


**Fantasyfootball**  
CumulativeMaxPassYardsPerAttempt



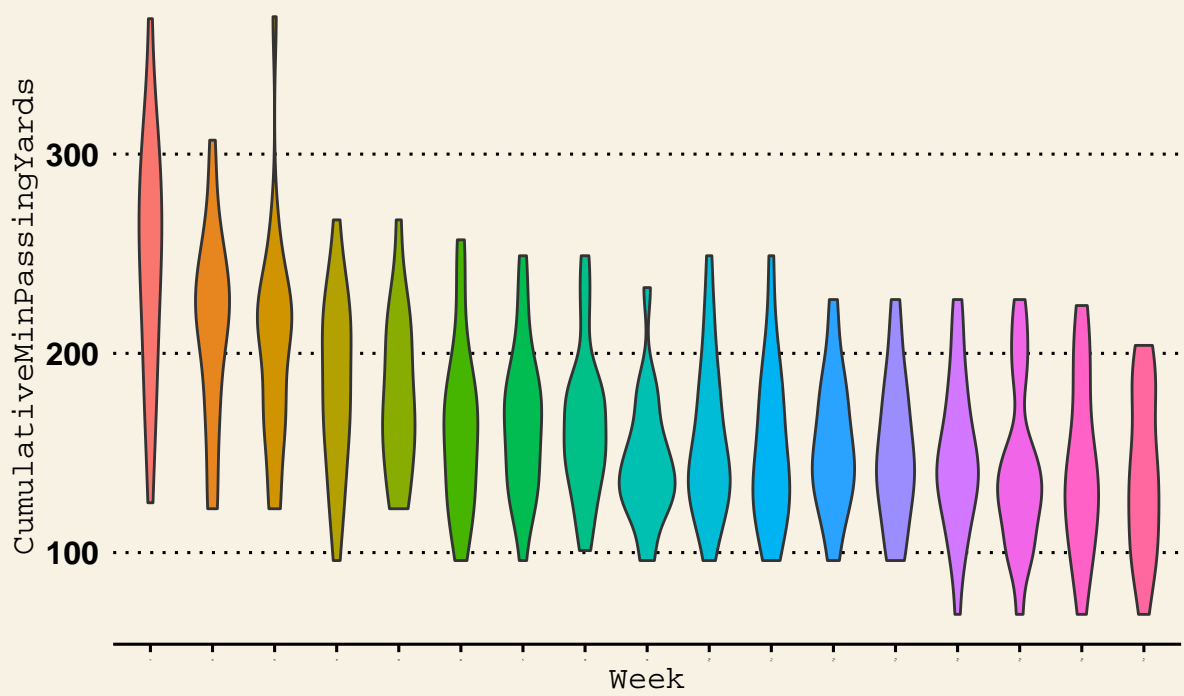
## Fantasyfootball

CumulativeMinPassingTouchdowns



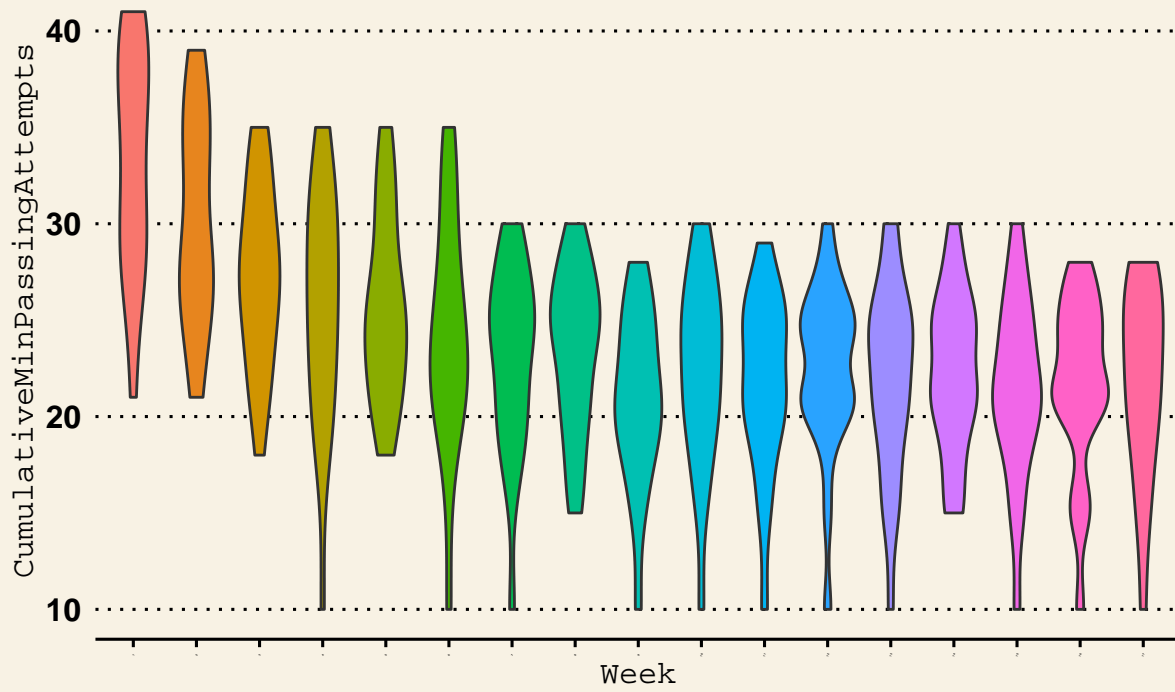
## Fantasyfootball

CumulativeMinPassingYards



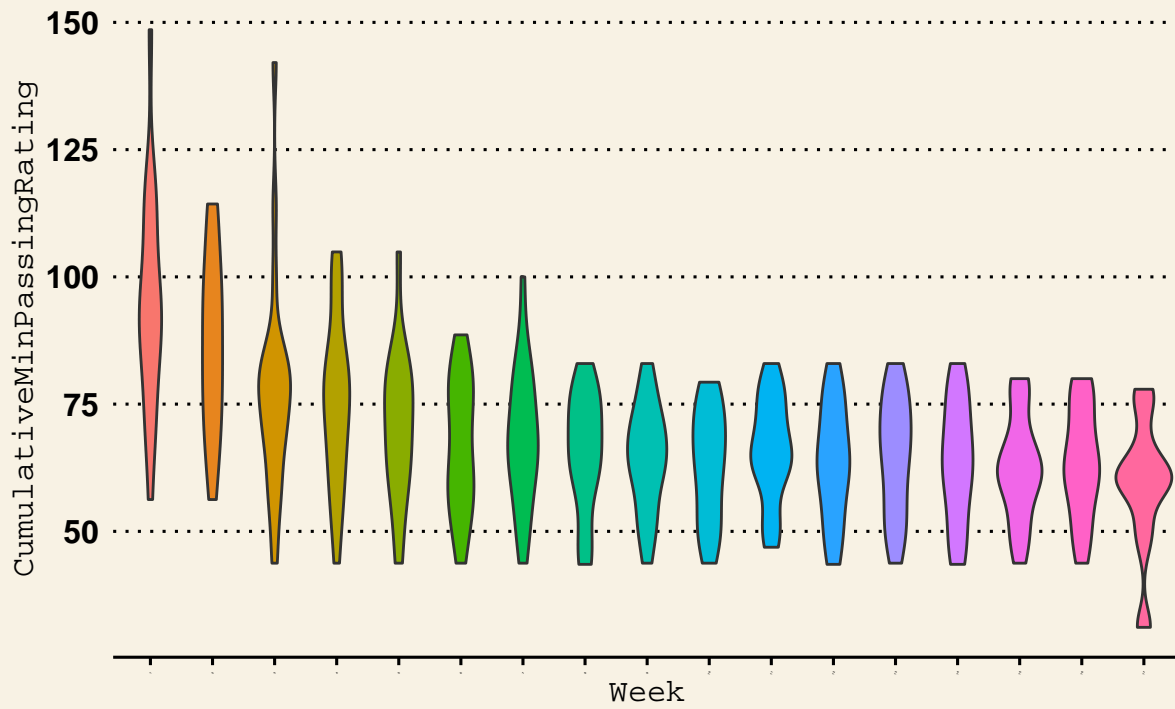
## Fantasyfootball

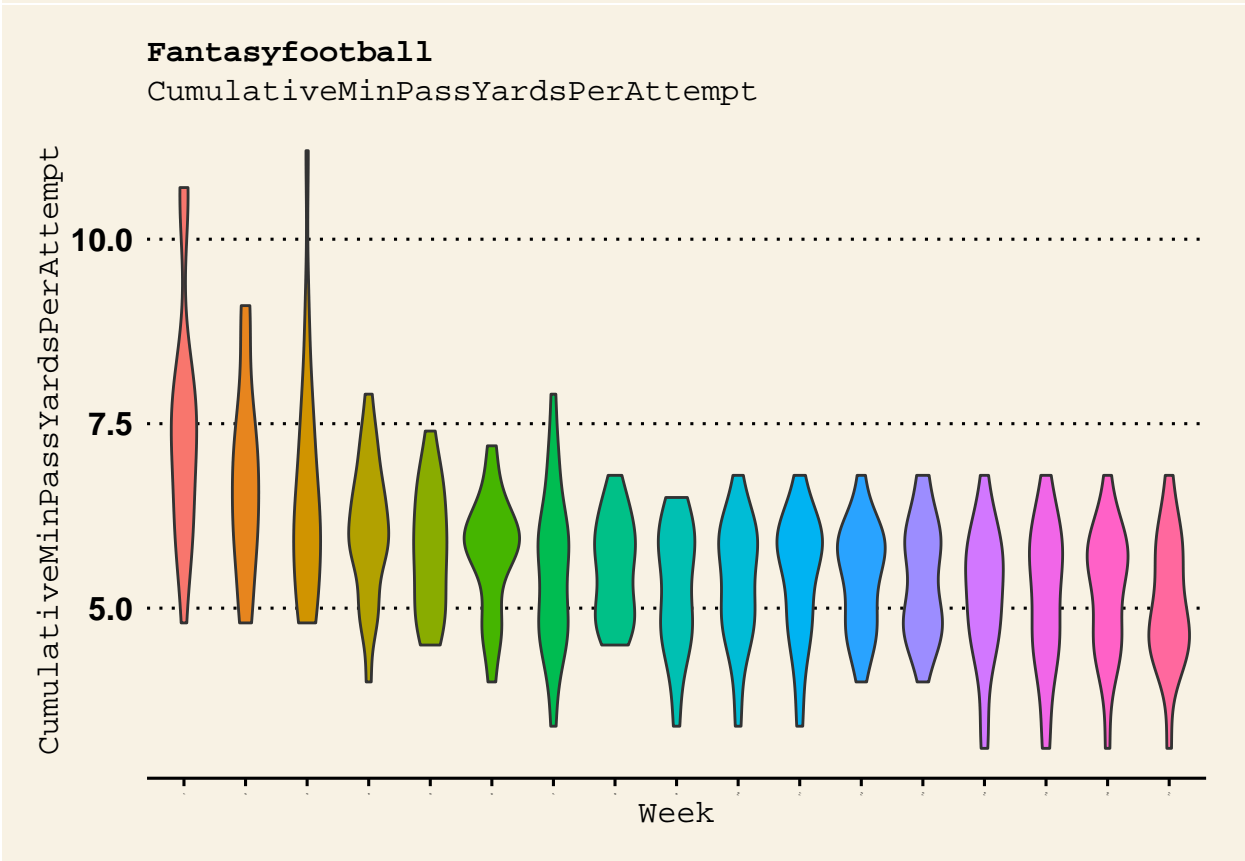
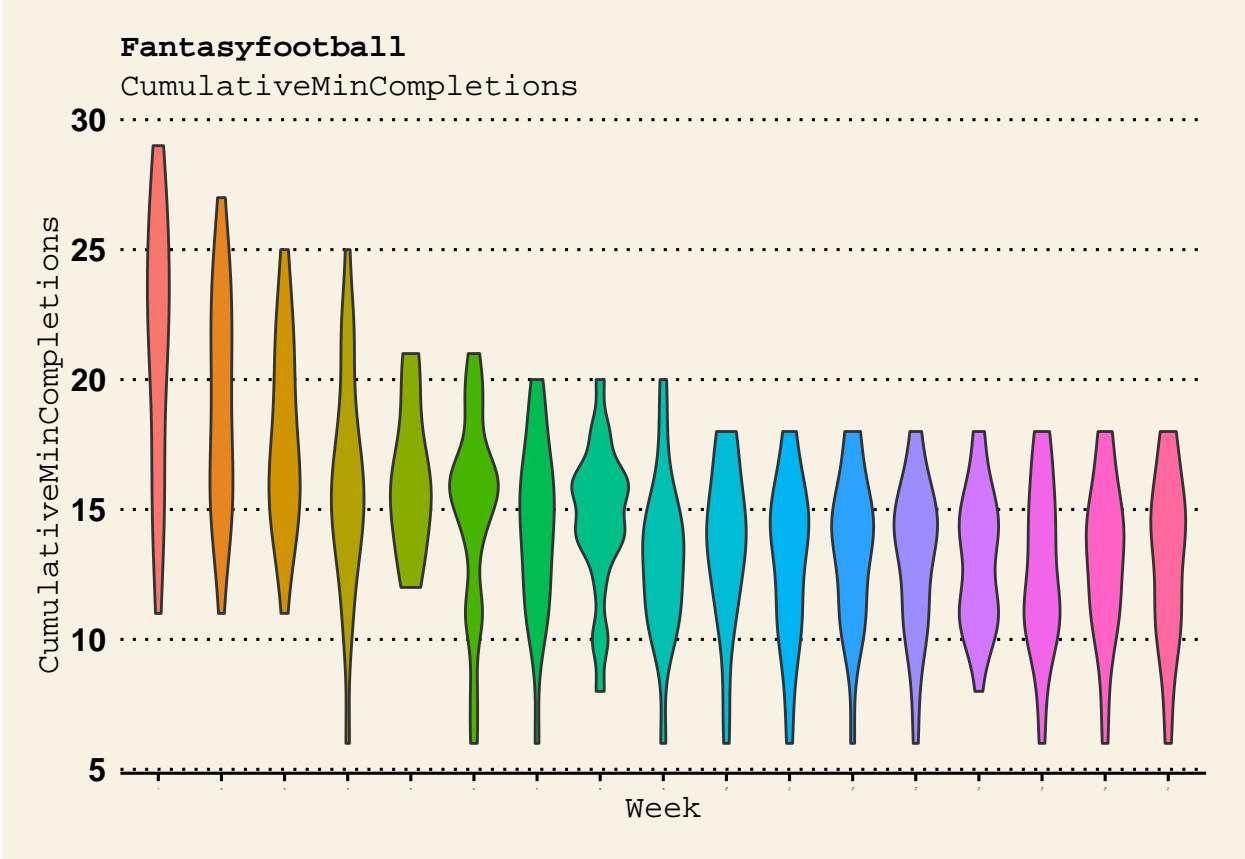
CumulativeMinPassingAttempts



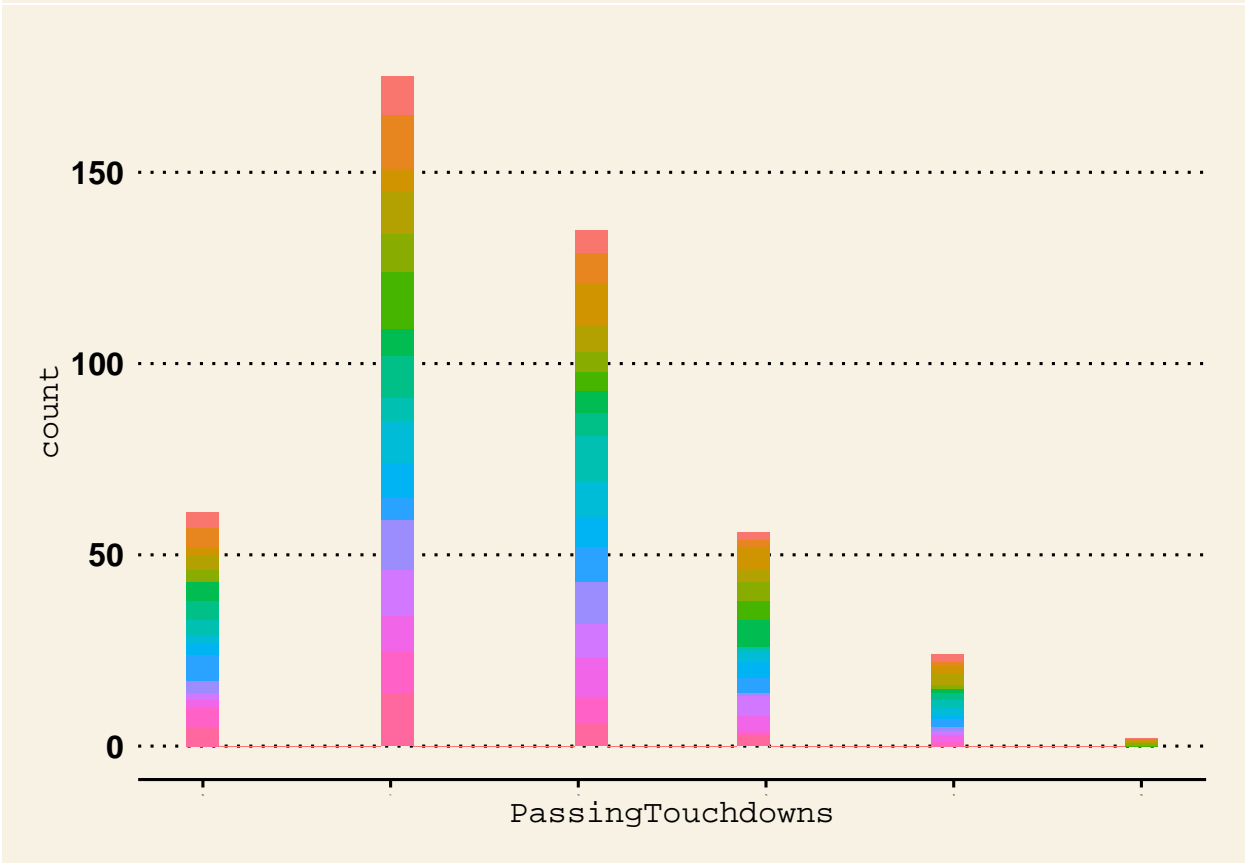
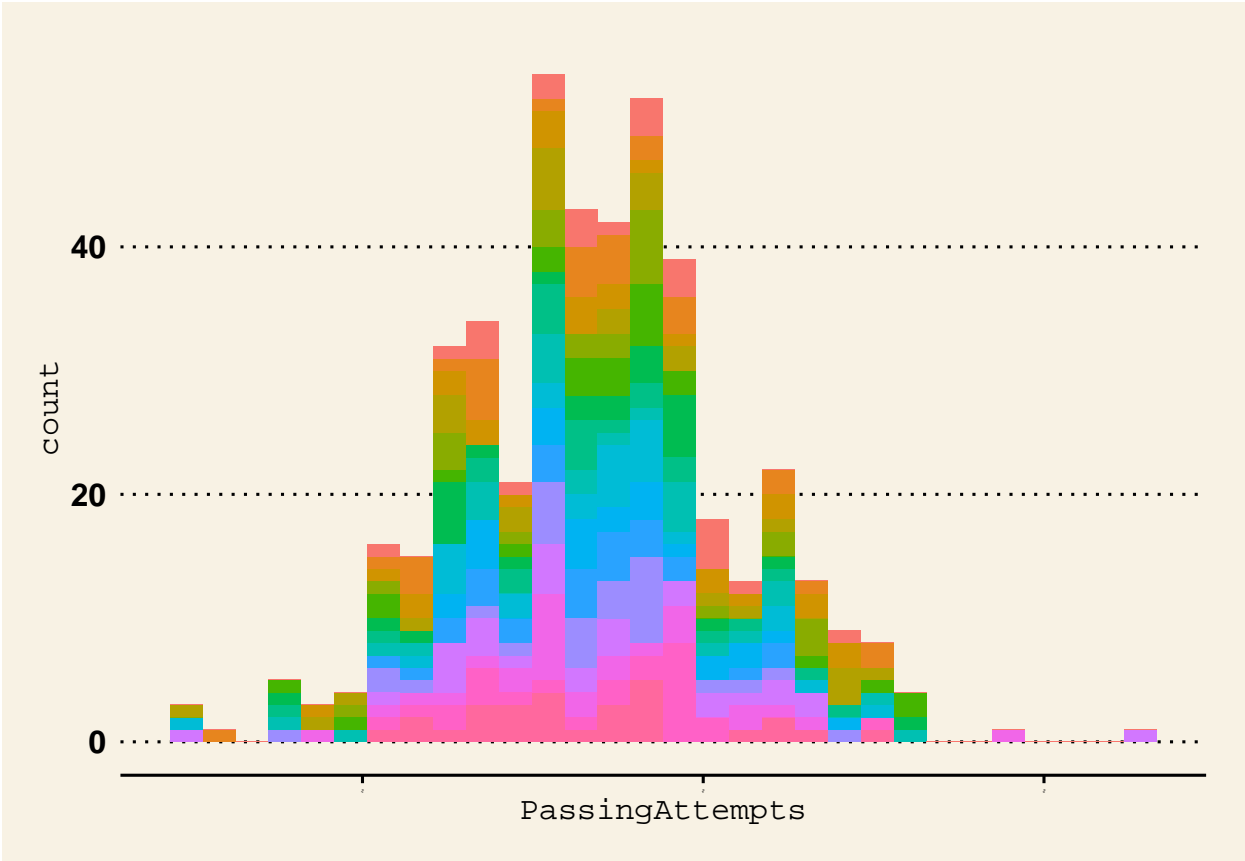
## Fantasyfootball

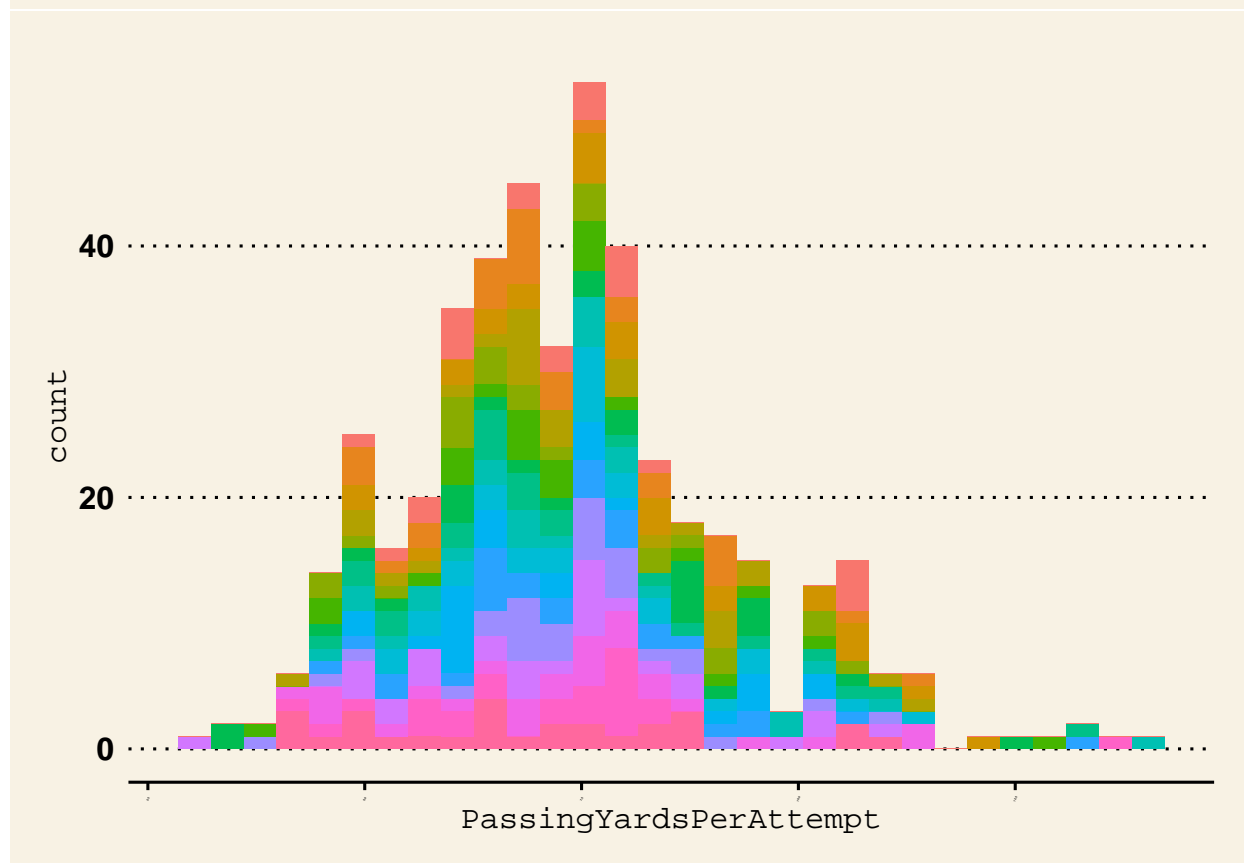
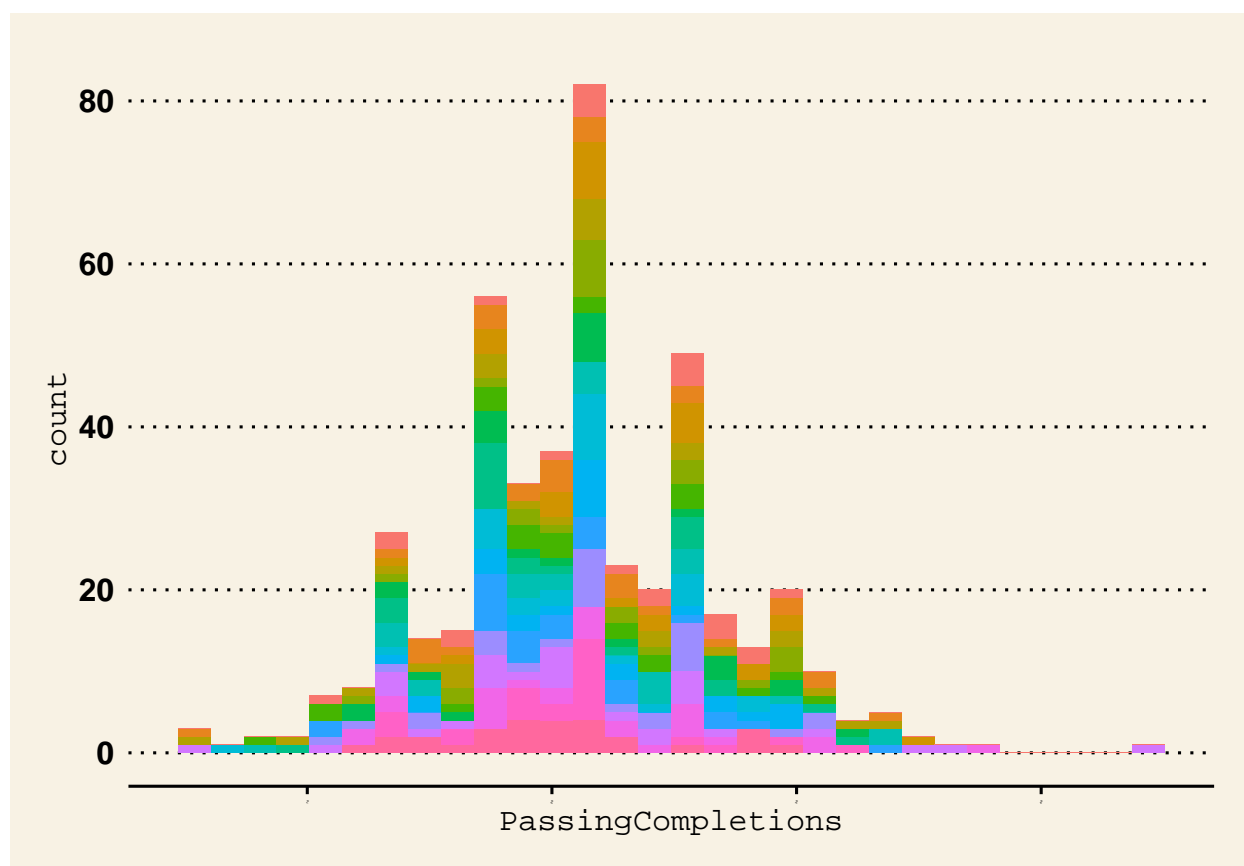
CumulativeMinPassingRating









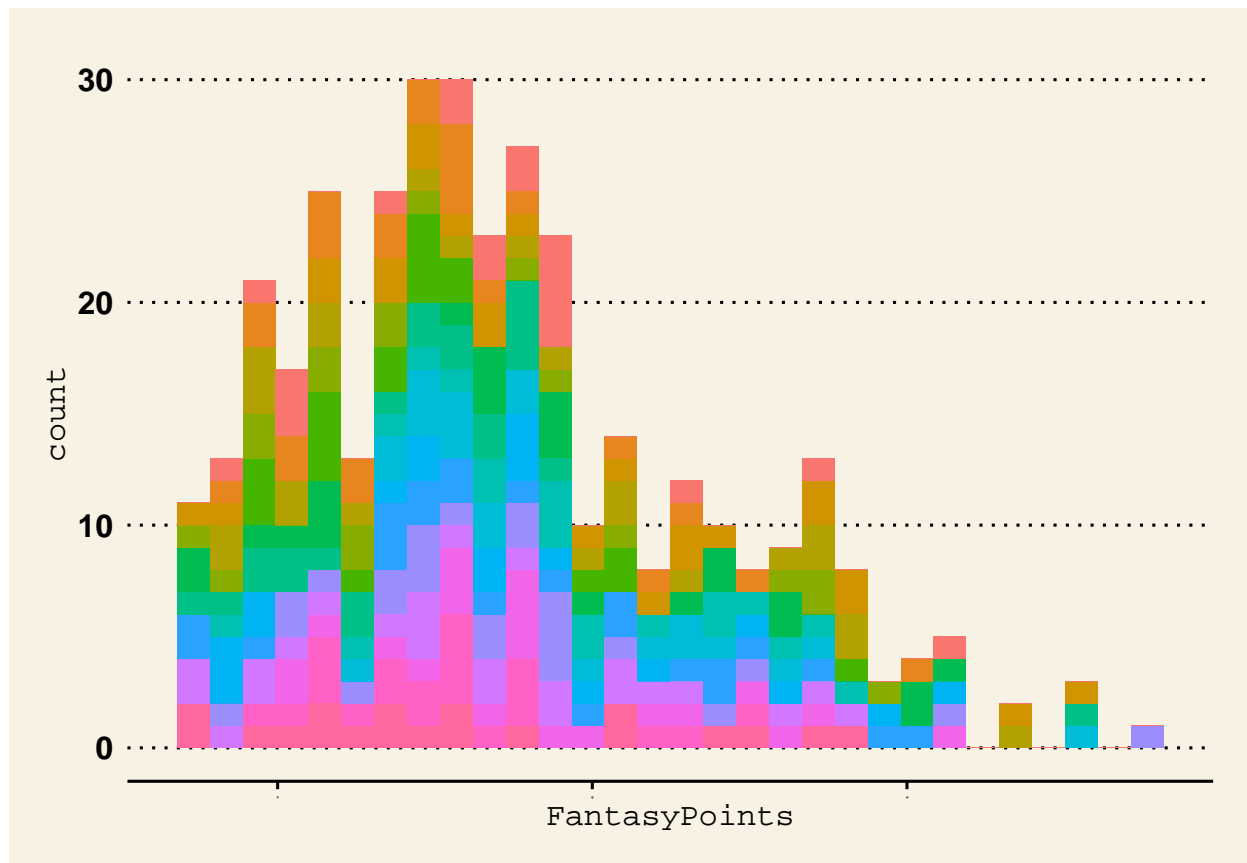


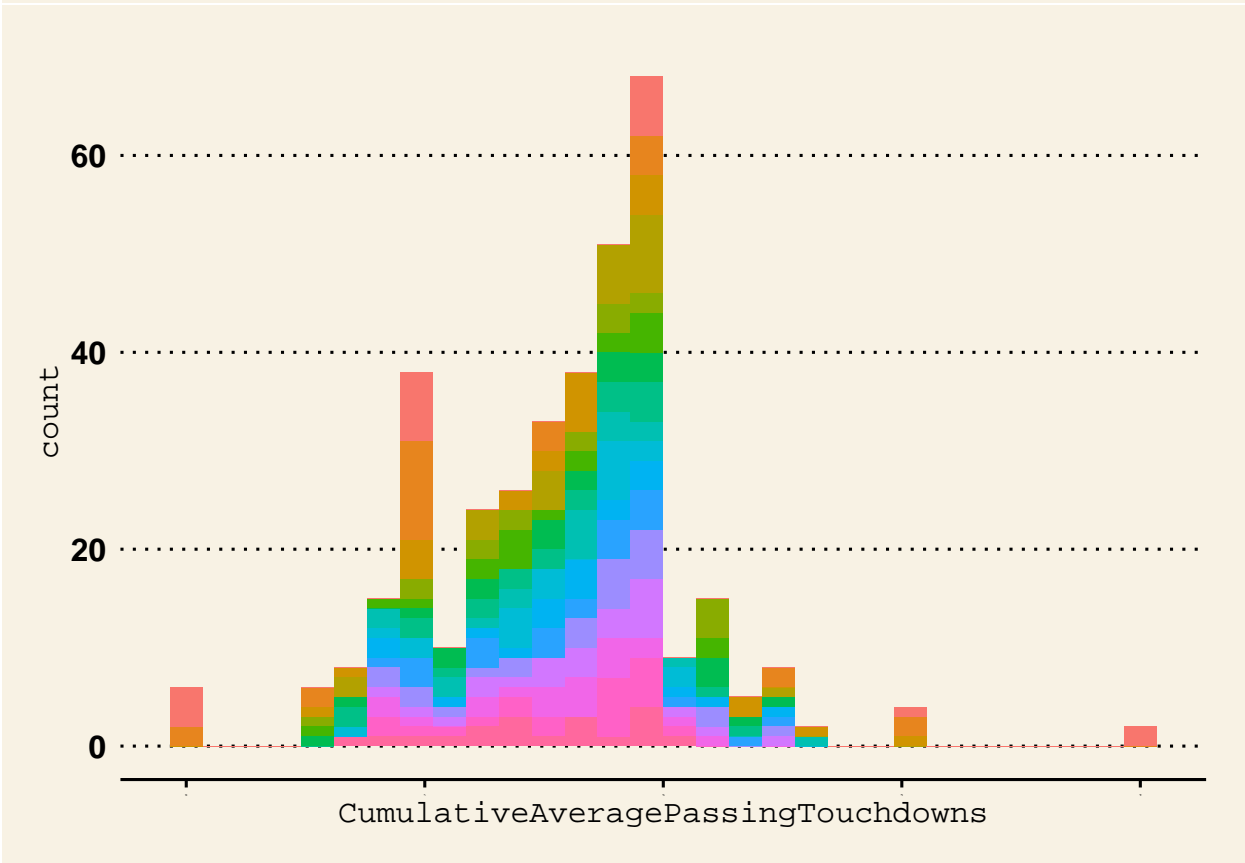
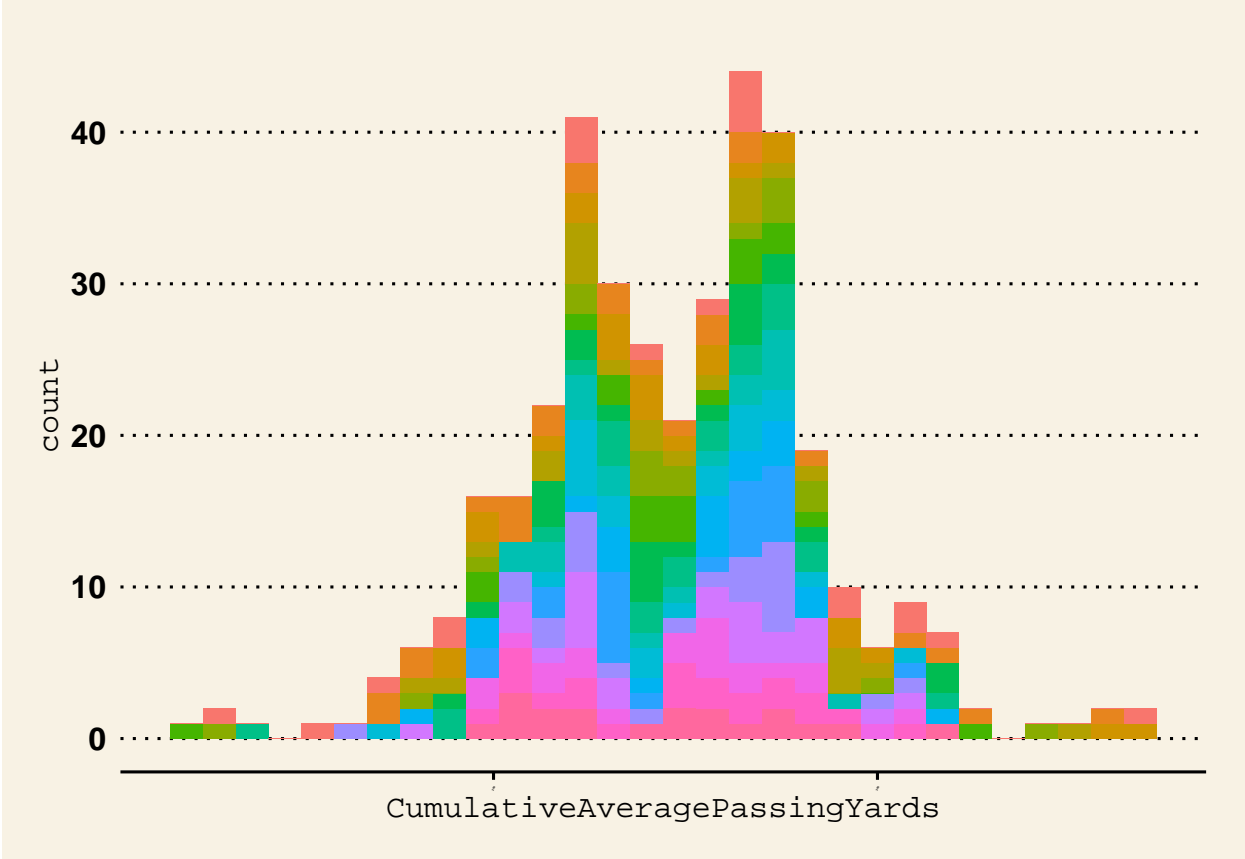


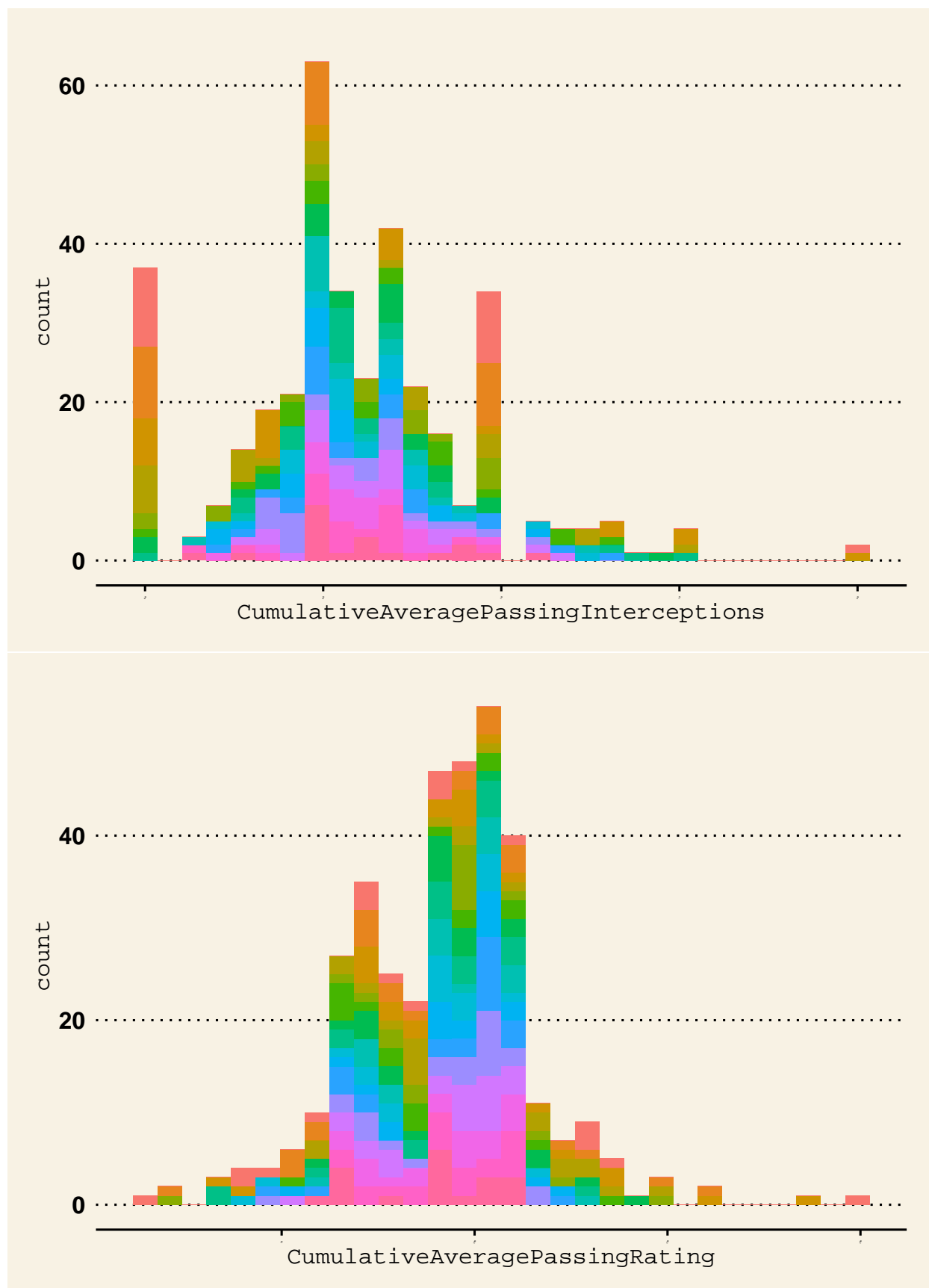
```
### 8.4.2 : Derived feature Set
```

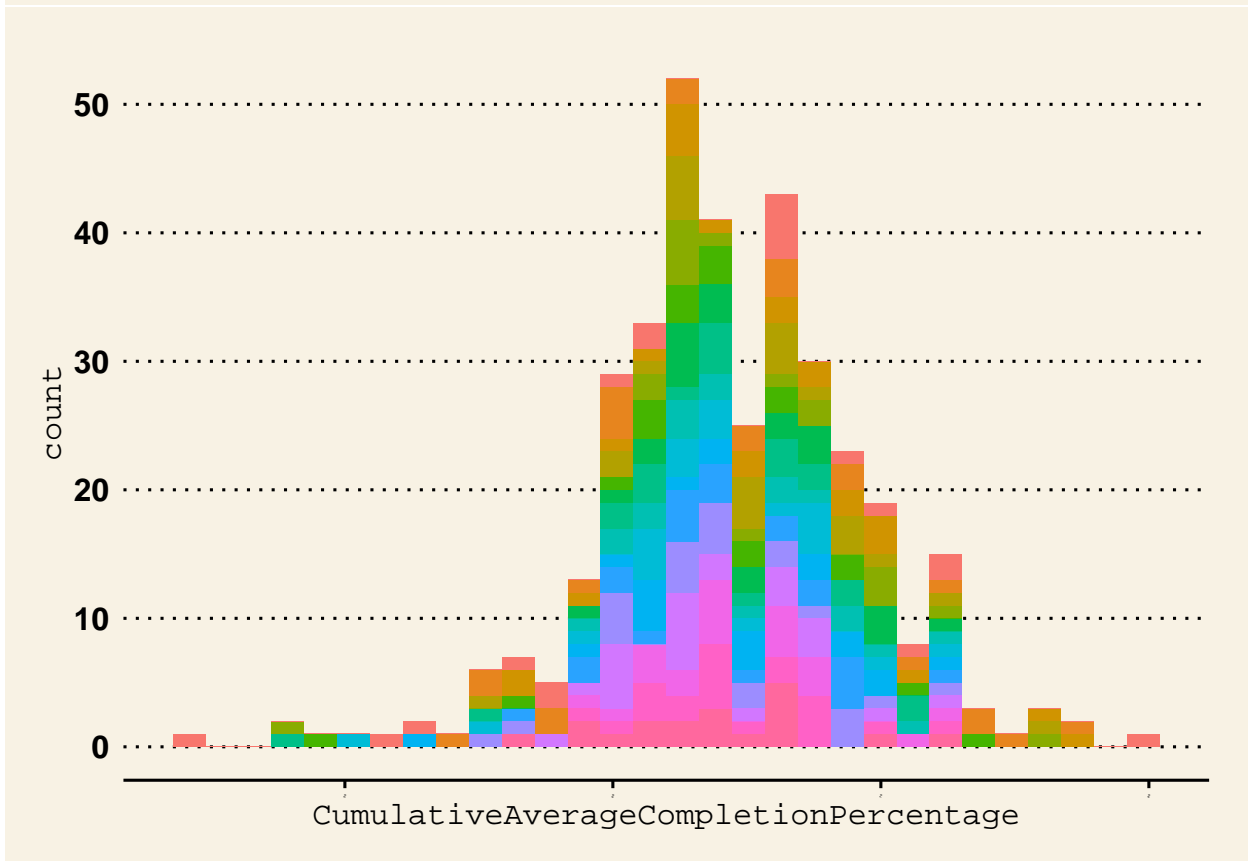
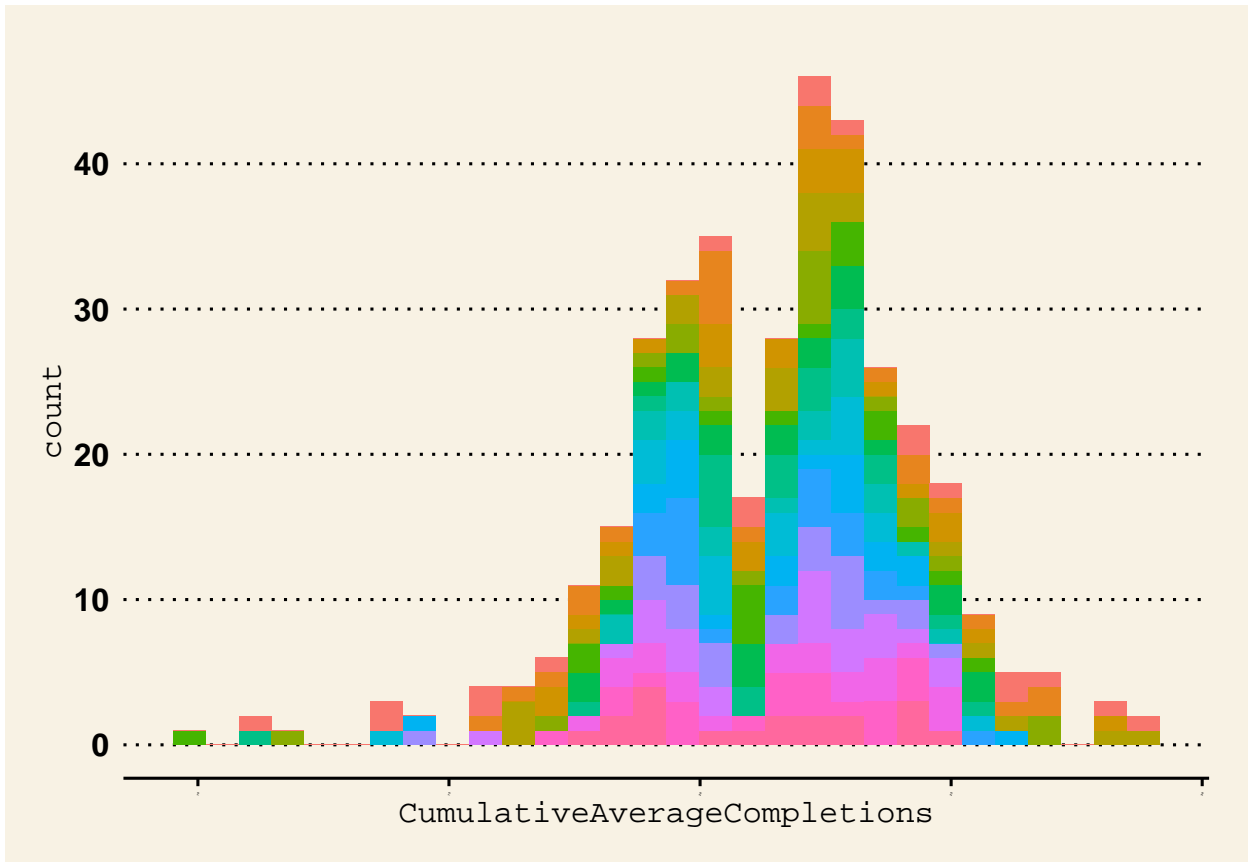
```
features_to_keep = names(eda_derived[-1])
```

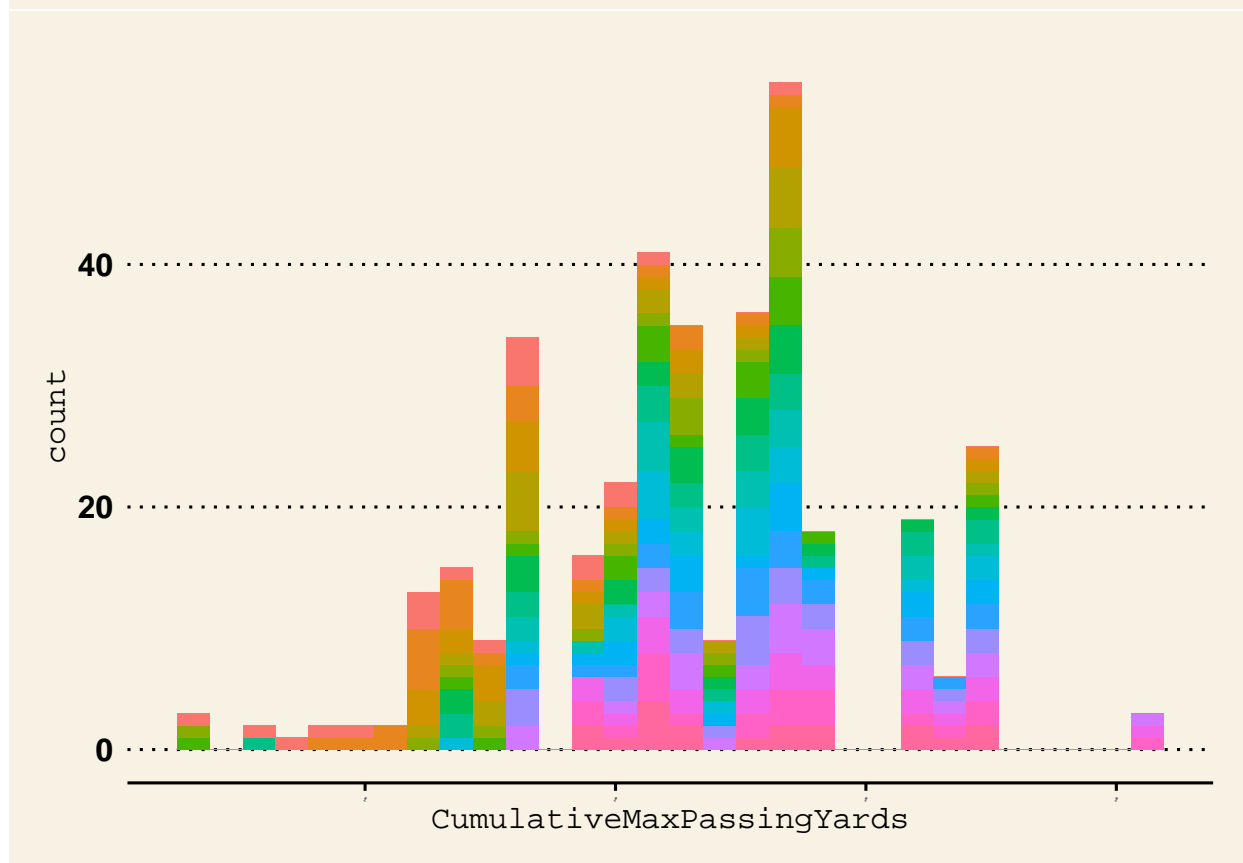
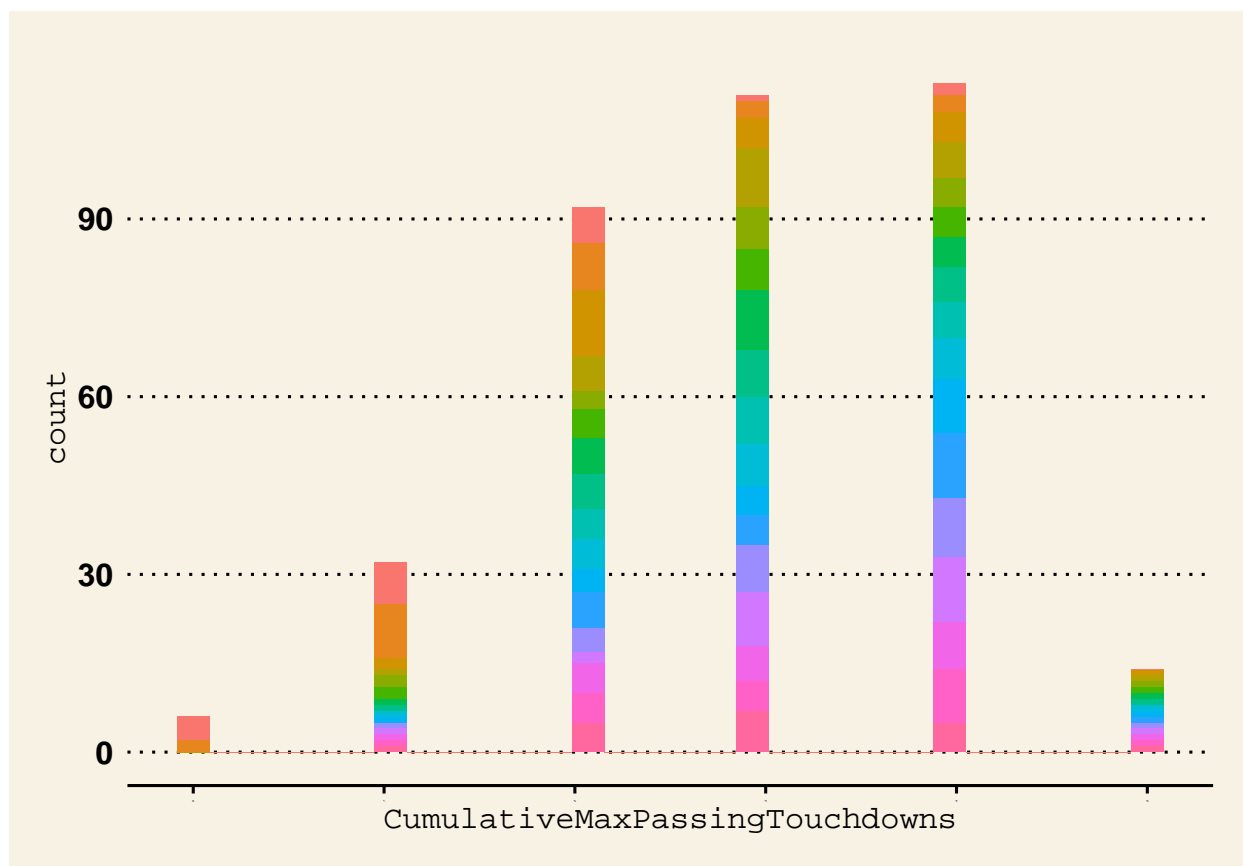
```
for(f in features_to_keep){  
  hist = eda_derived %>% ggplot(aes_string(x=f,fill="Week"))+  
    geom_histogram(bins=30,show.legend = FALSE)+  
    theme_wsj()+  
    theme(plot.title = element_text(size = rel(0.5)),  
          plot.subtitle = element_text(size = rel(0.5)),  
          axis.text.x = element_text(angle=65, vjust=0.6,size=1),  
          axis.title = element_text(size = rel(0.5)),  
          legend.position = "right",  
          legend.direction = "vertical",  
          legend.title = element_text(size = rel(0.5)))  
  print(hist)  
}
```

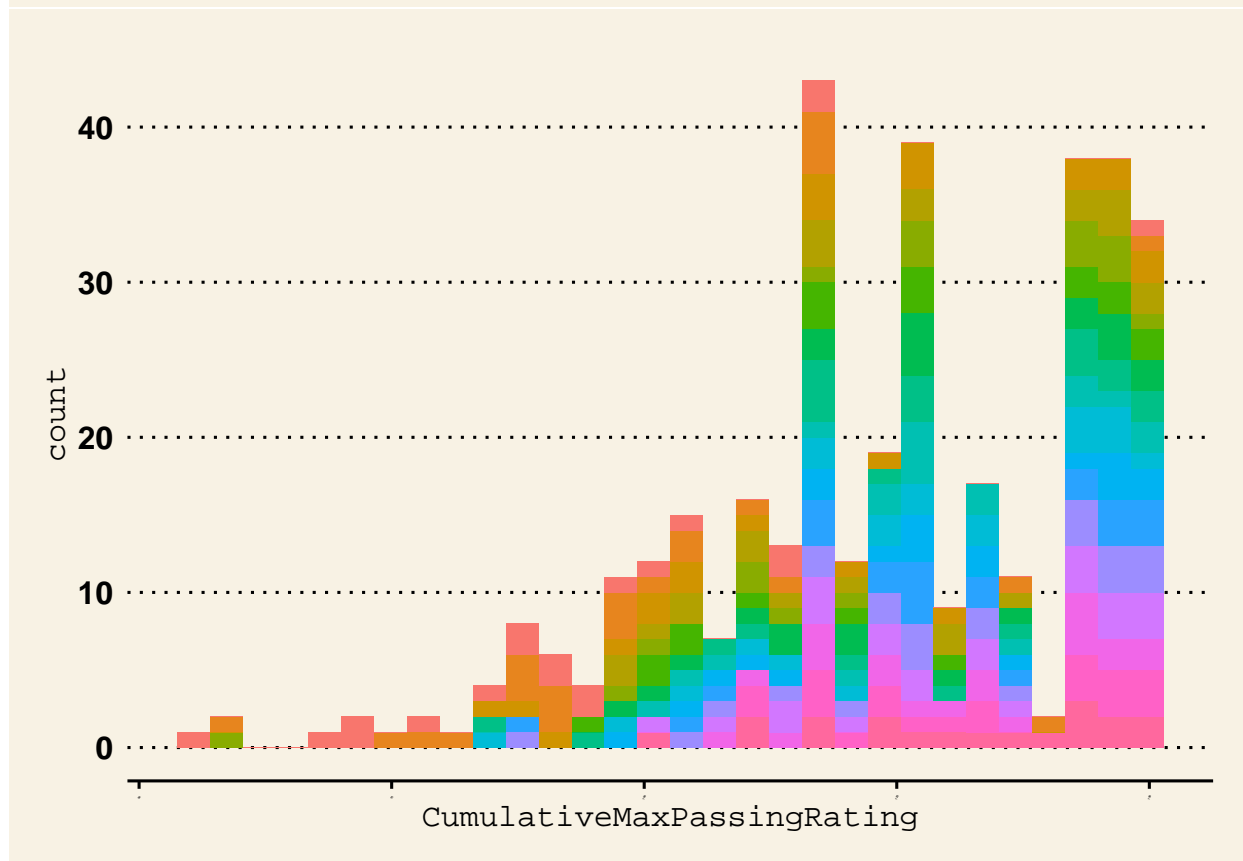
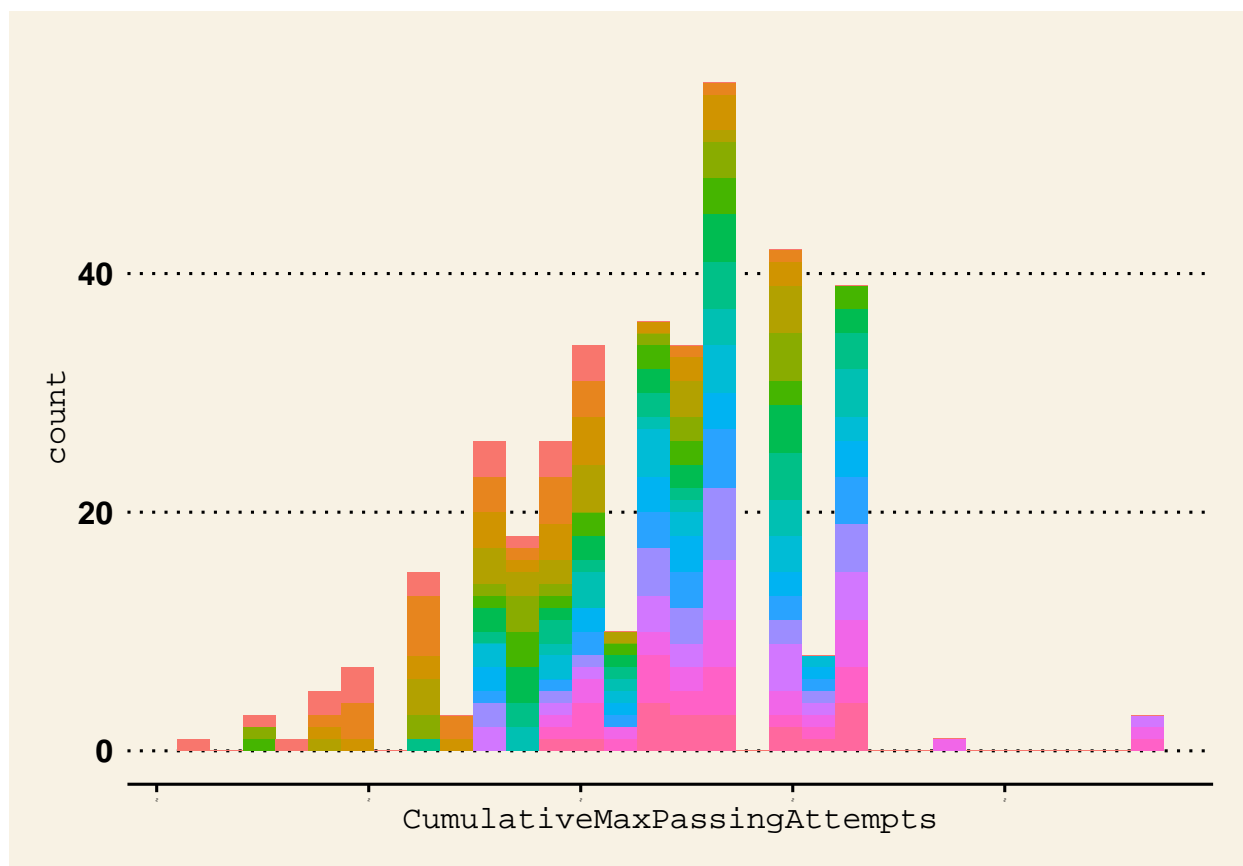


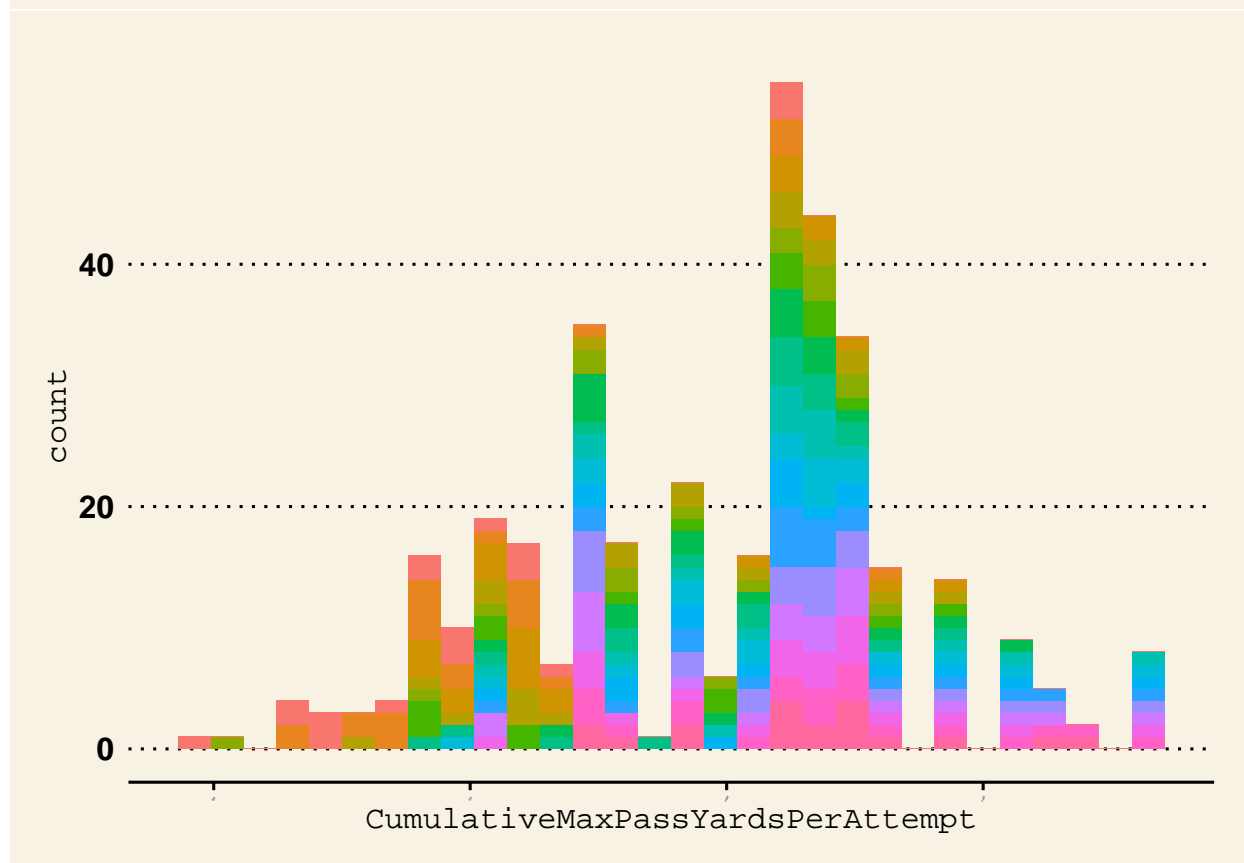
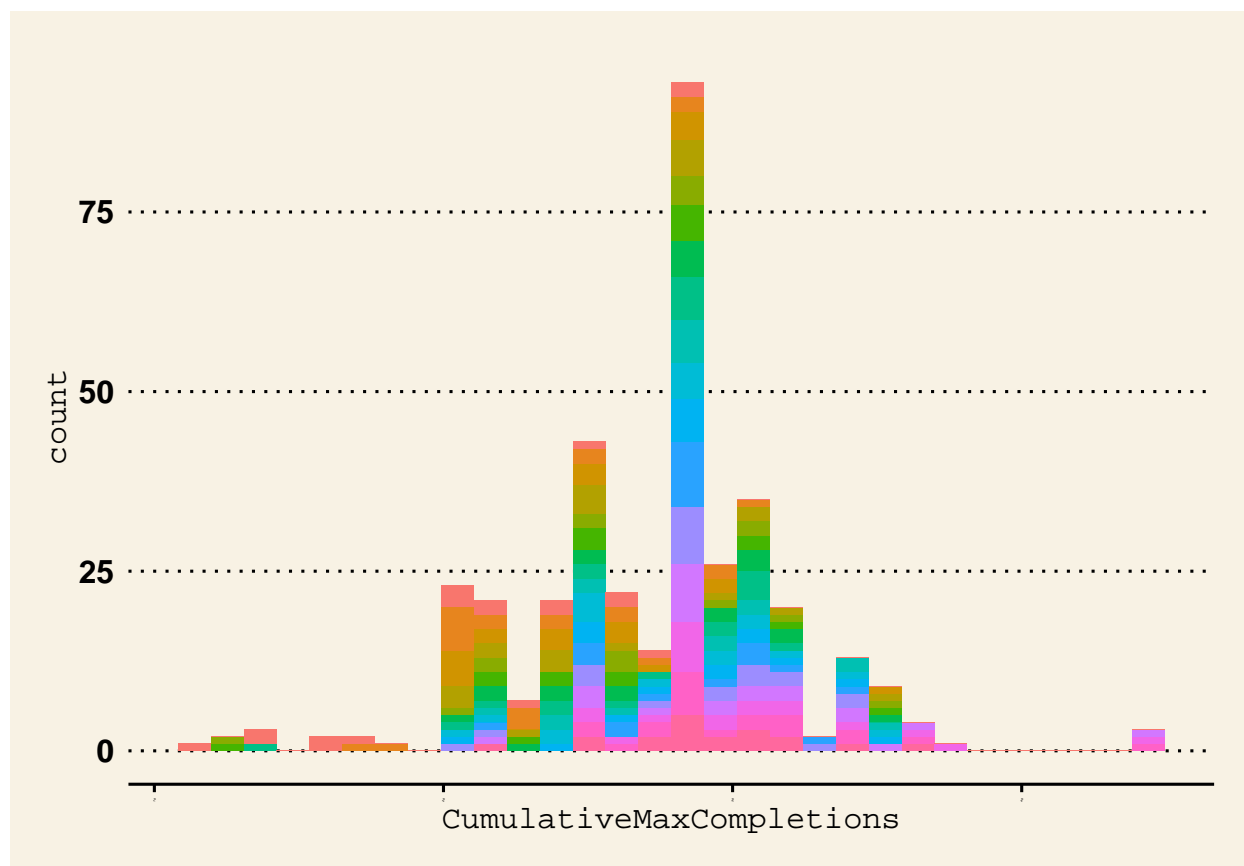


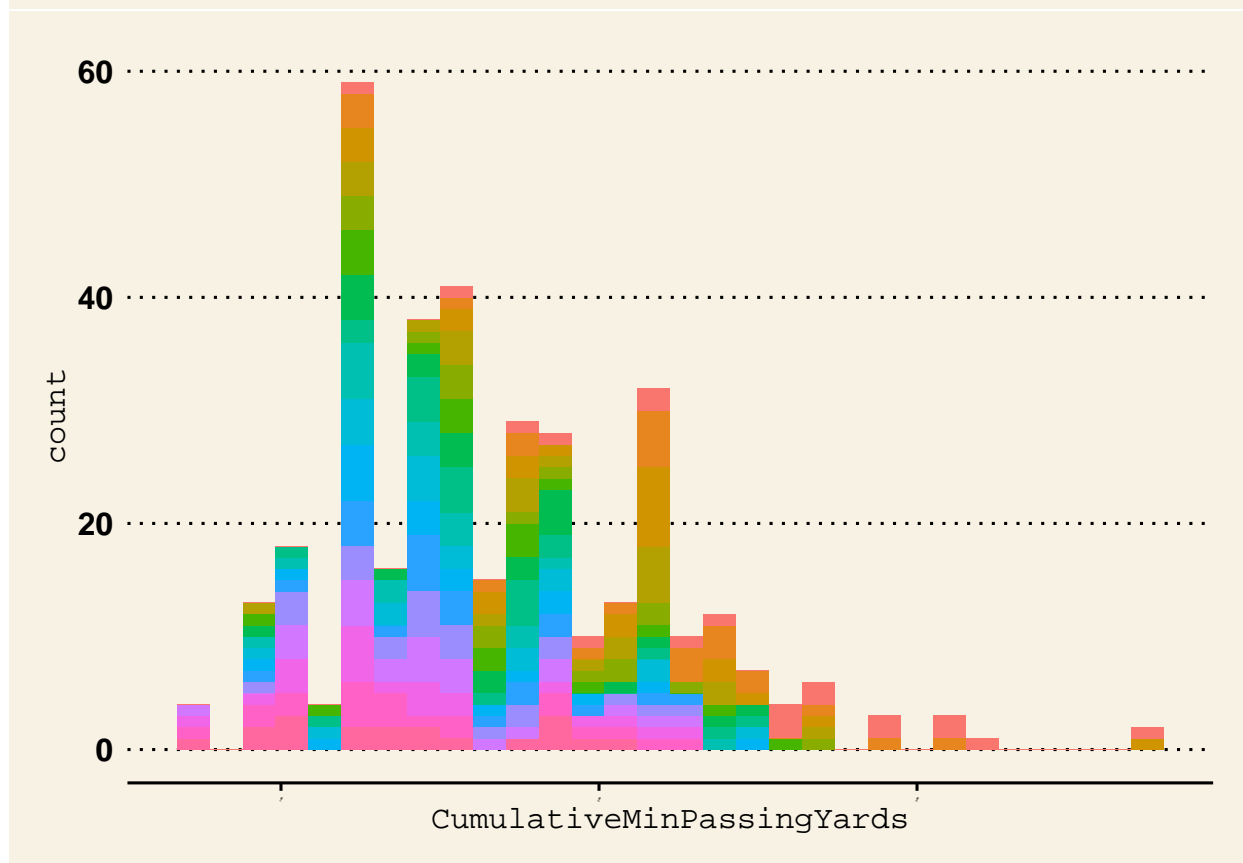
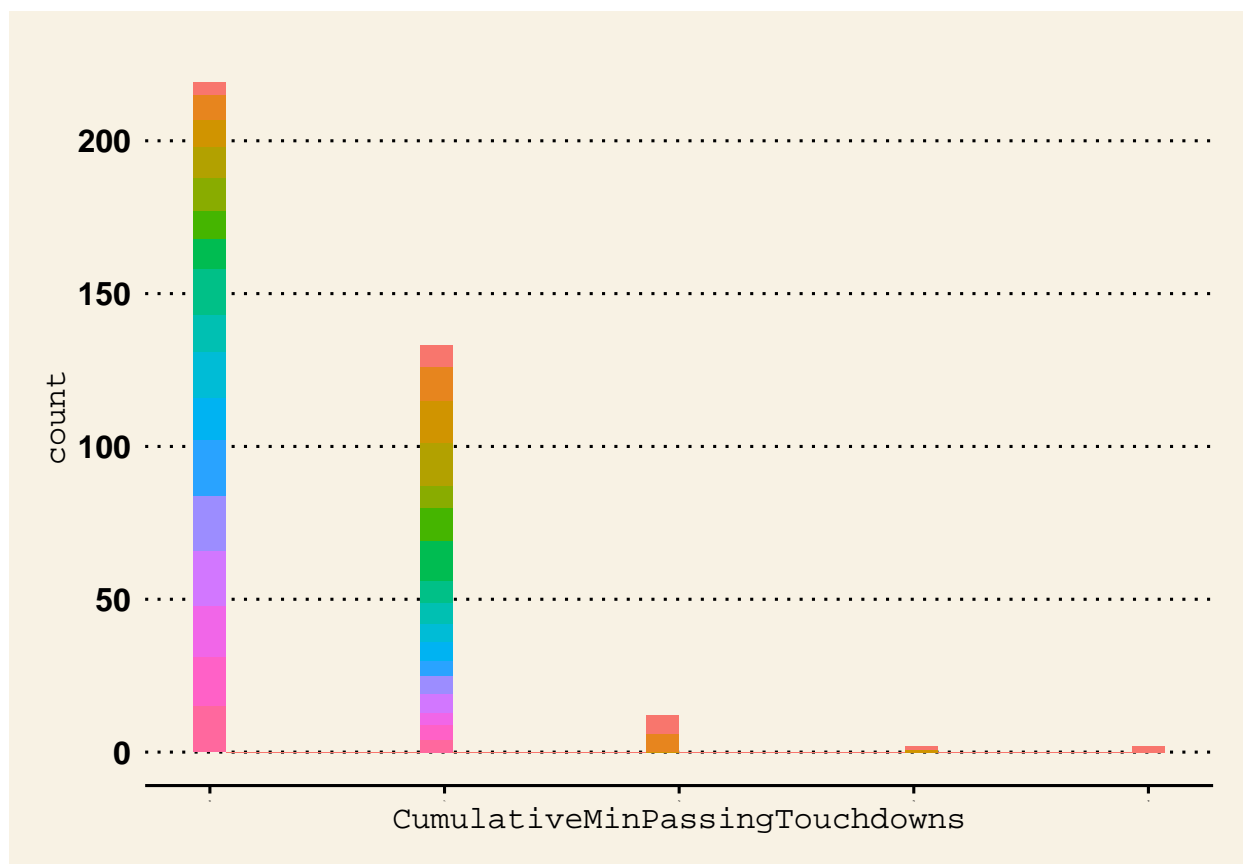




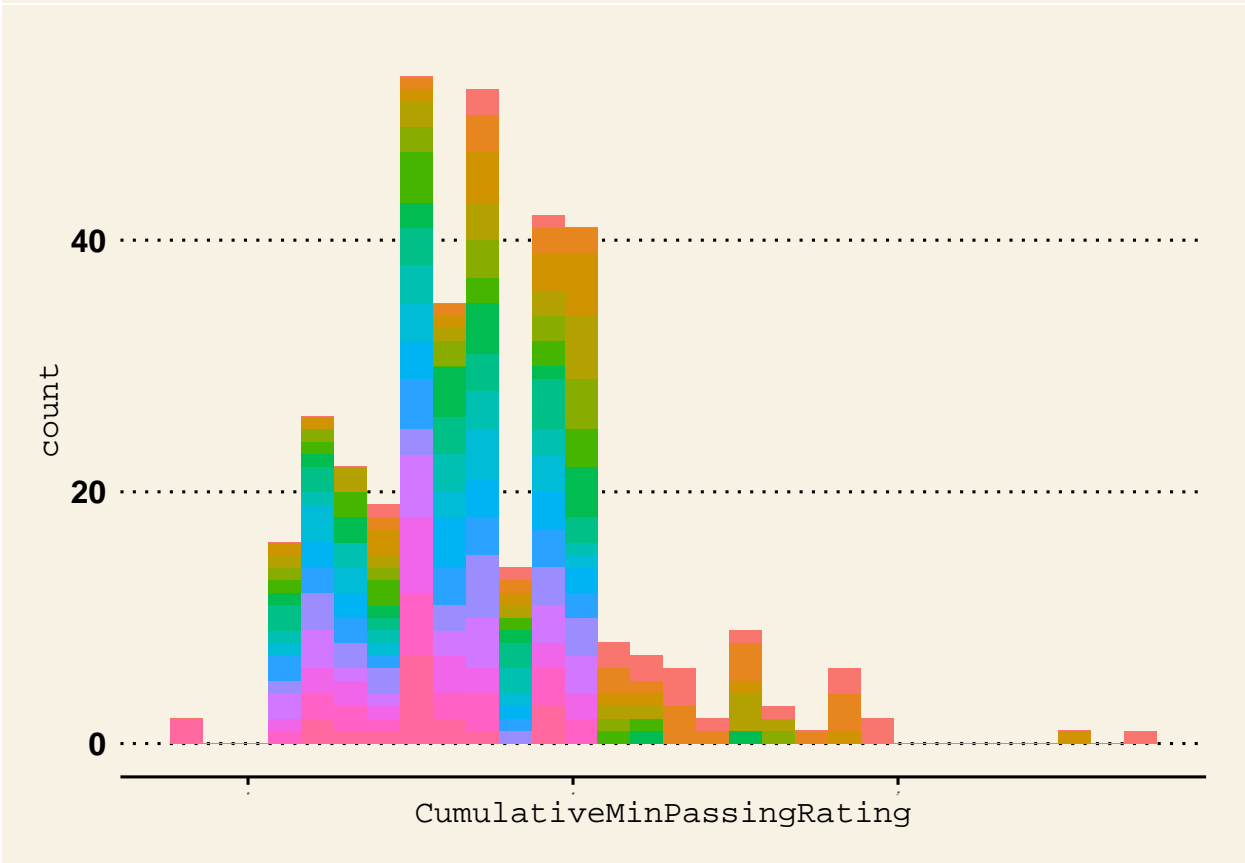
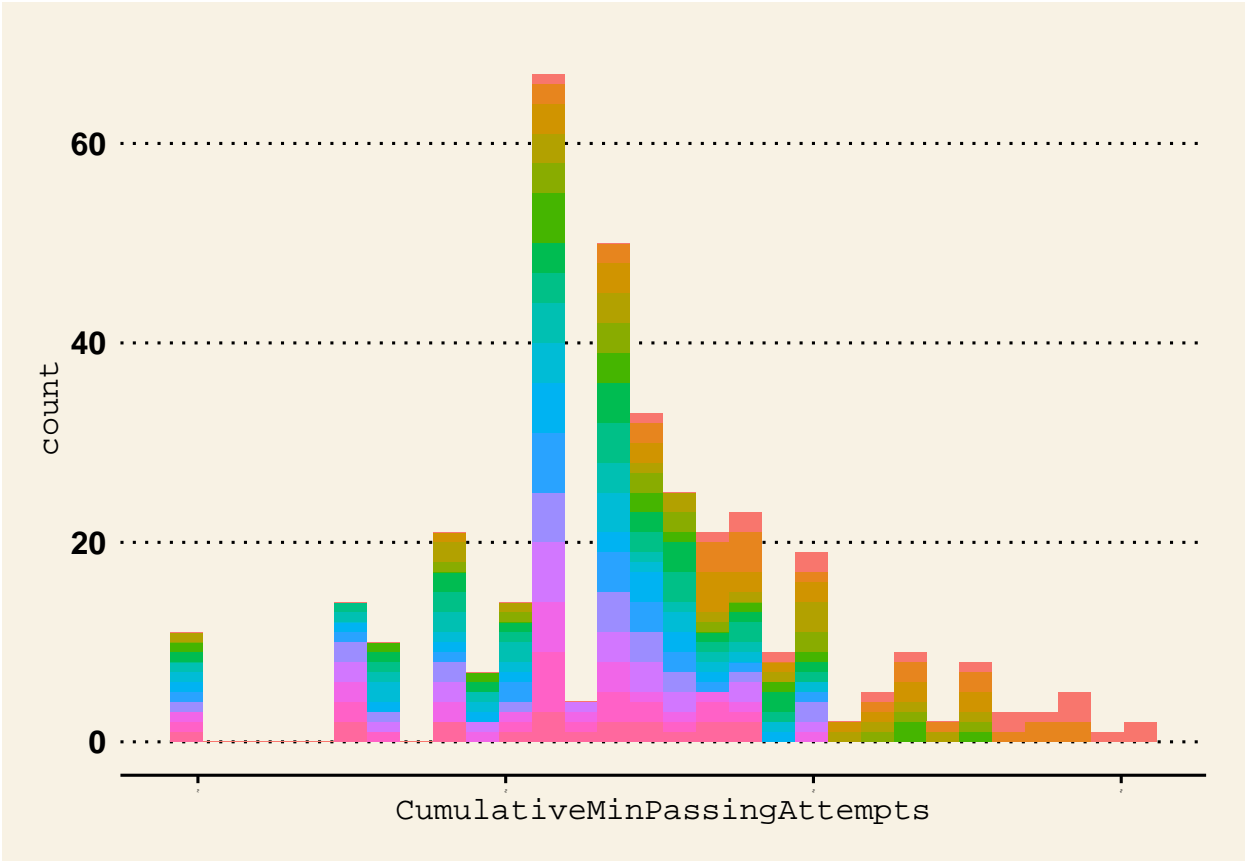


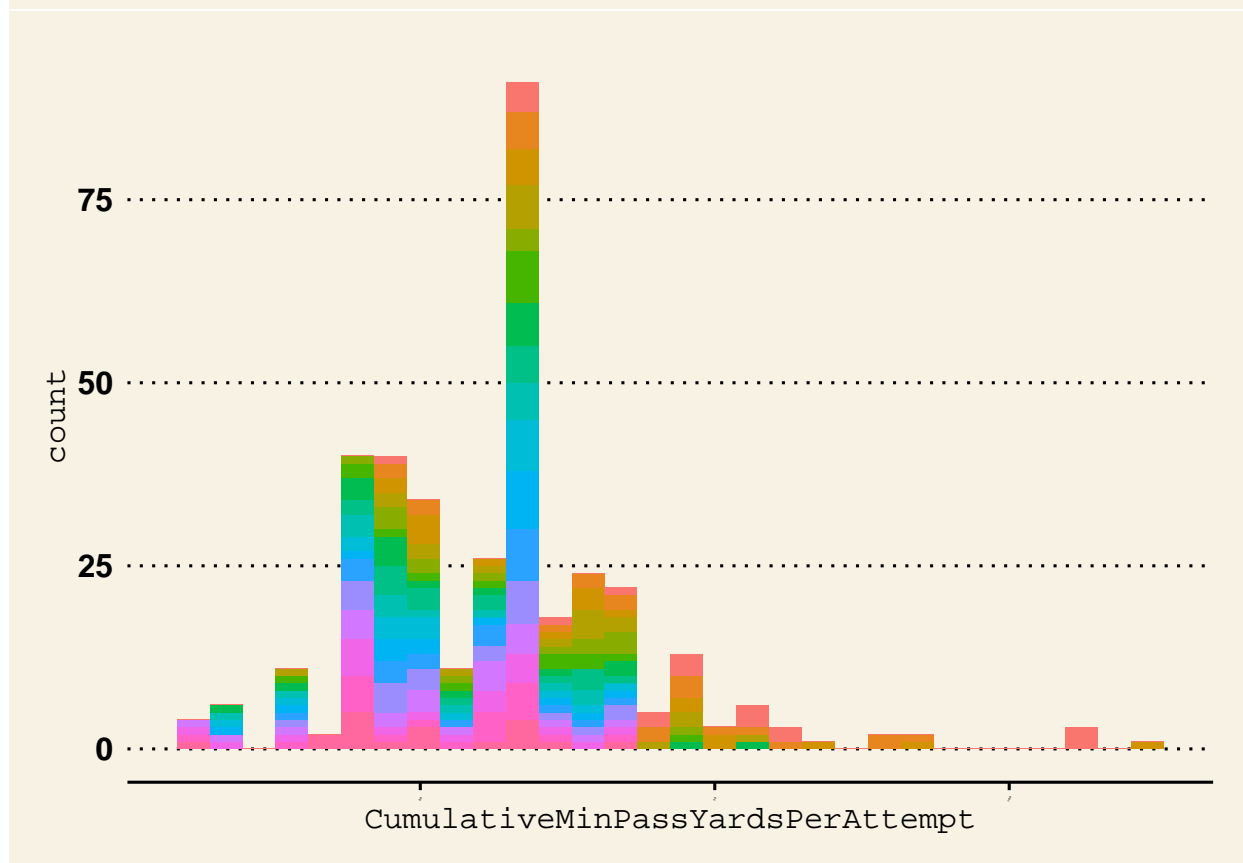
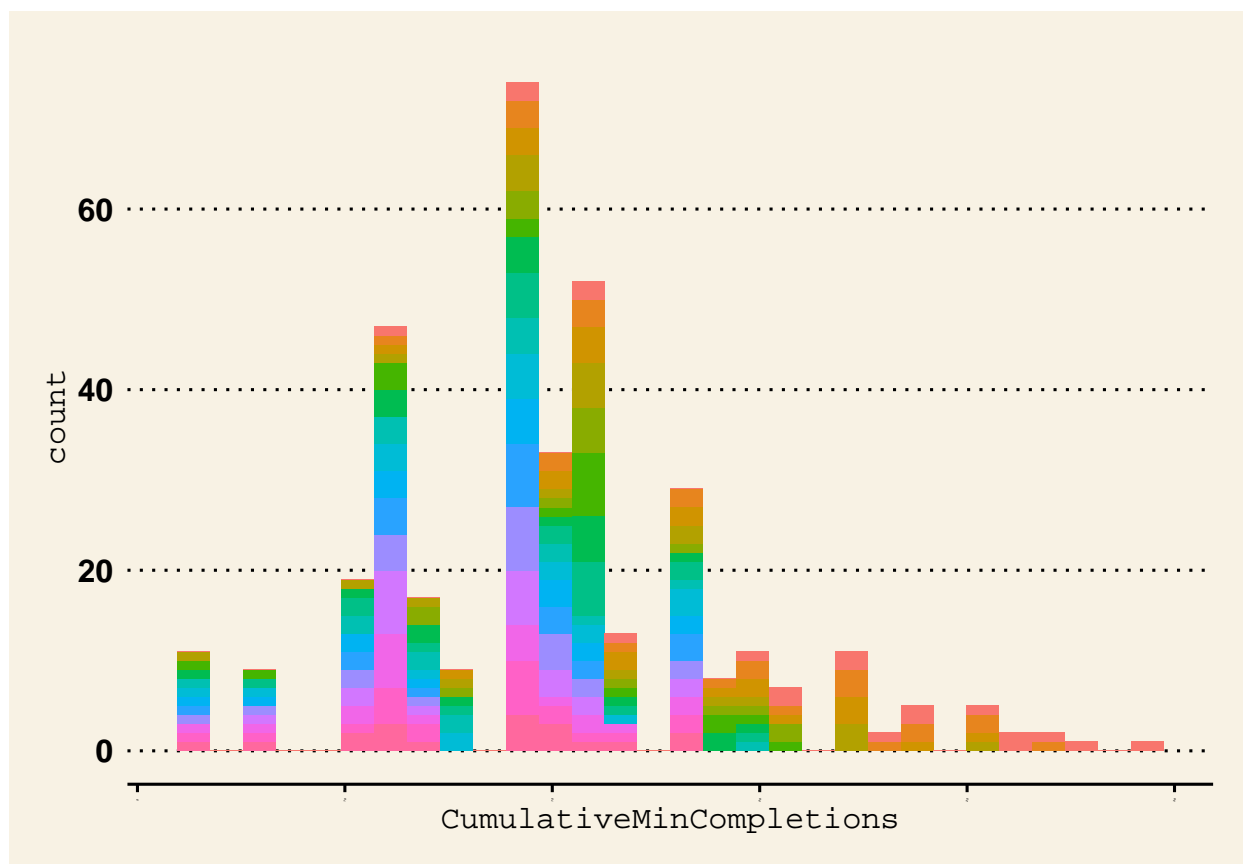






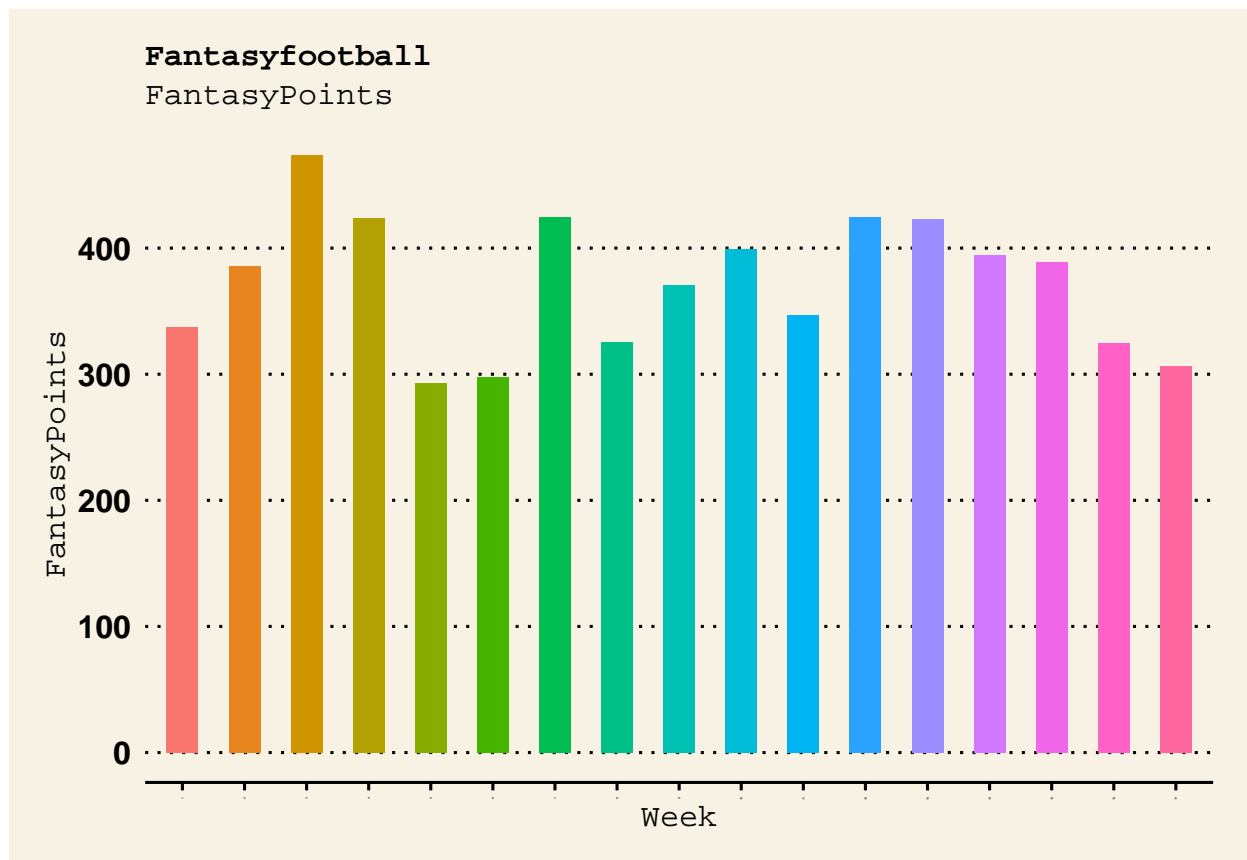


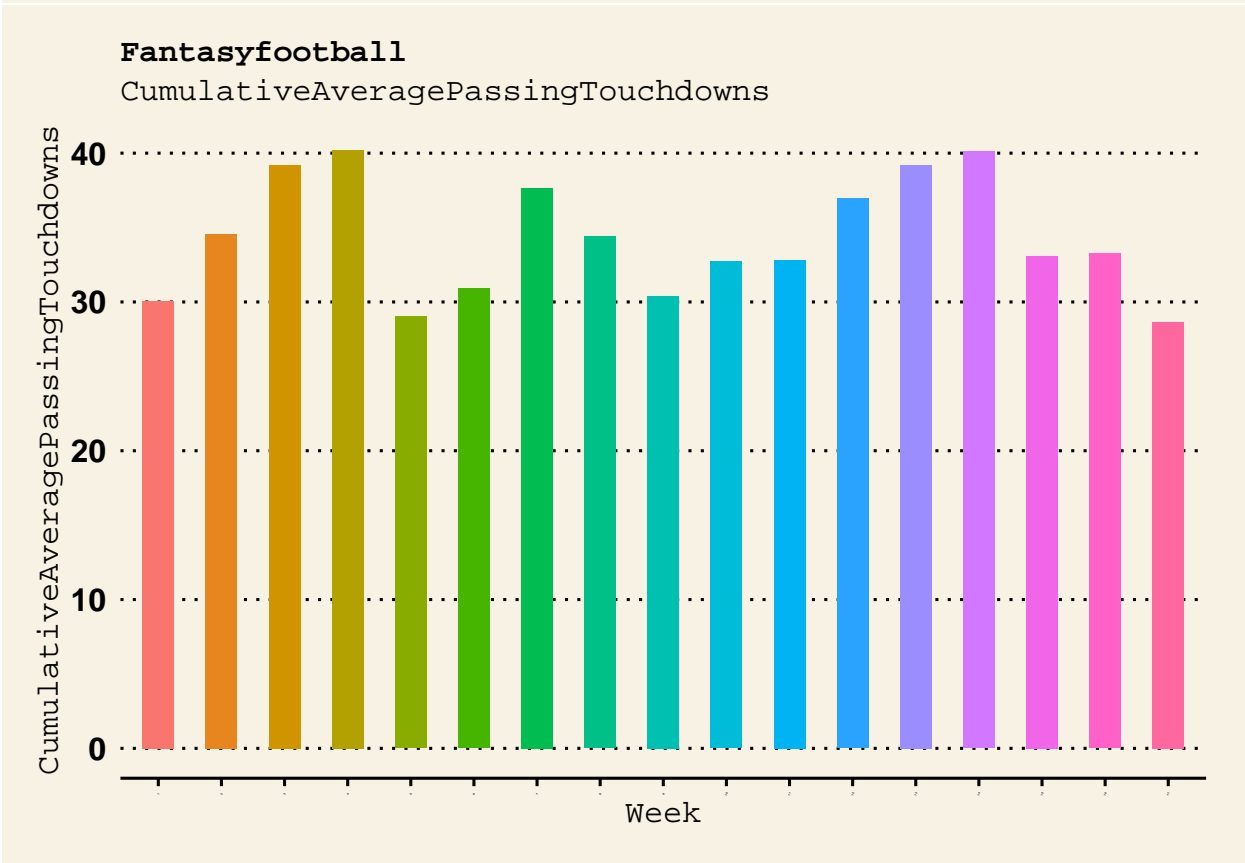
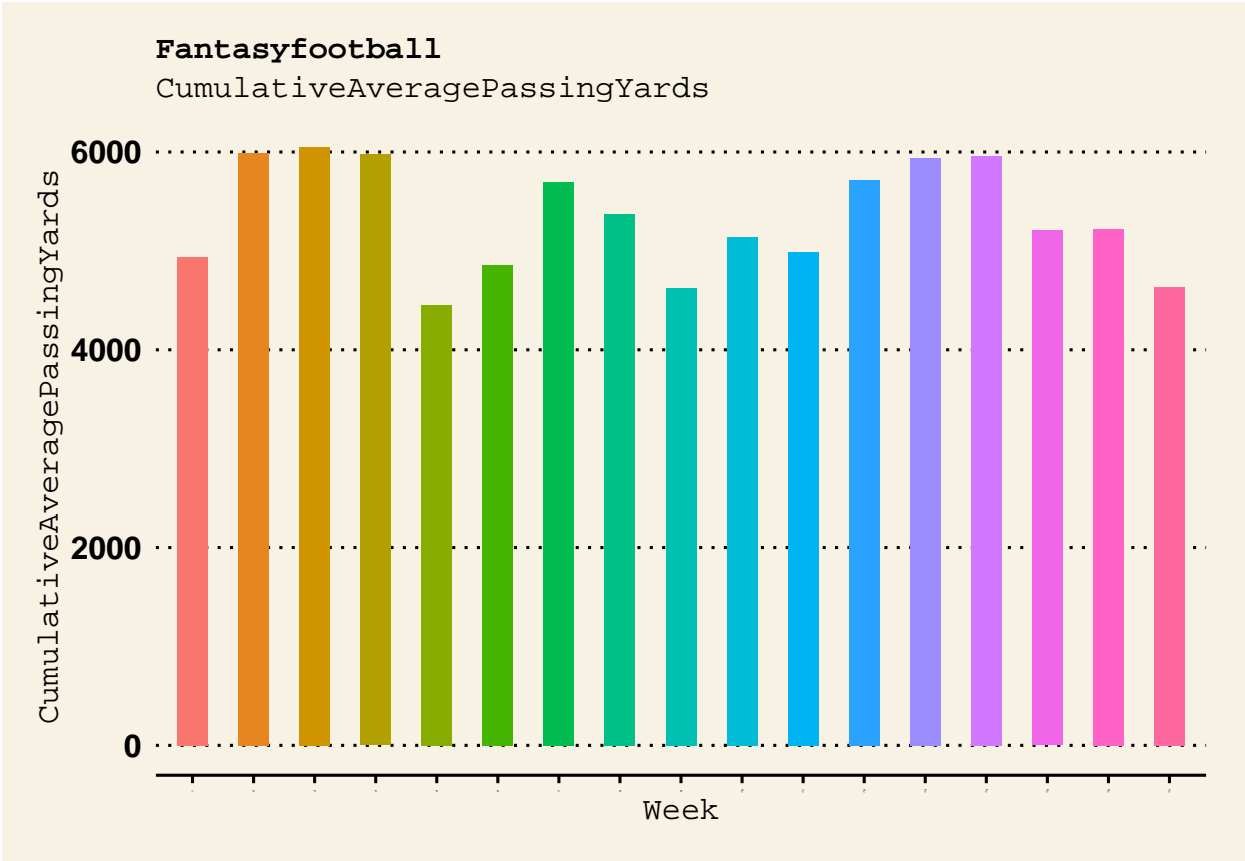


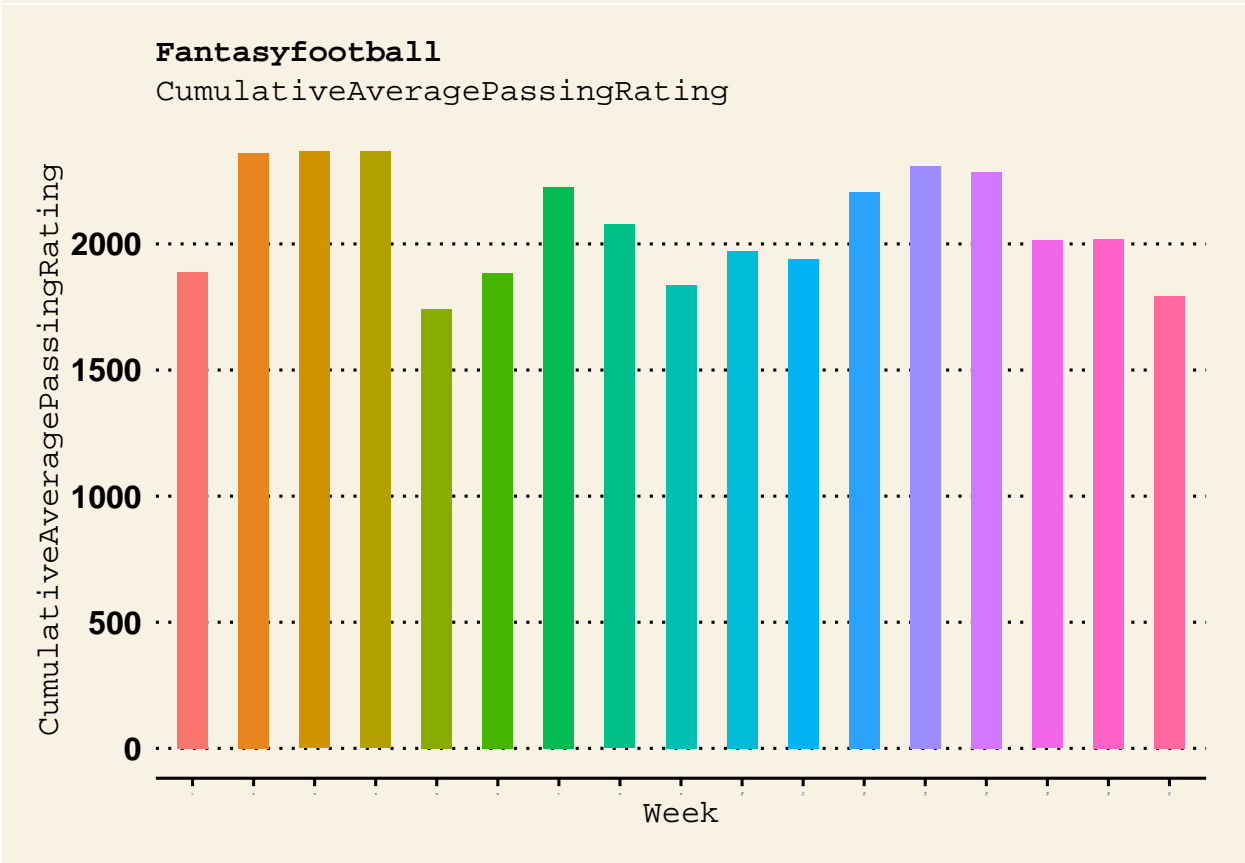
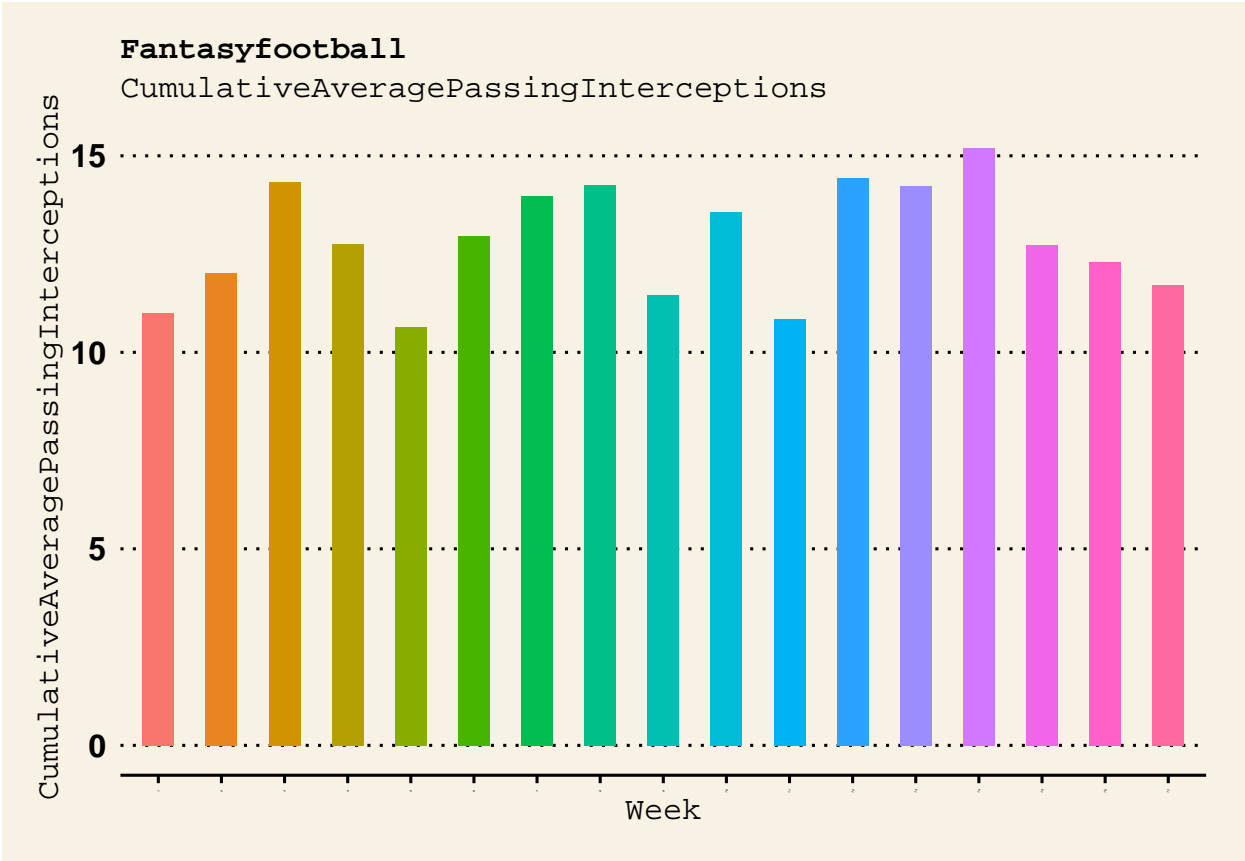


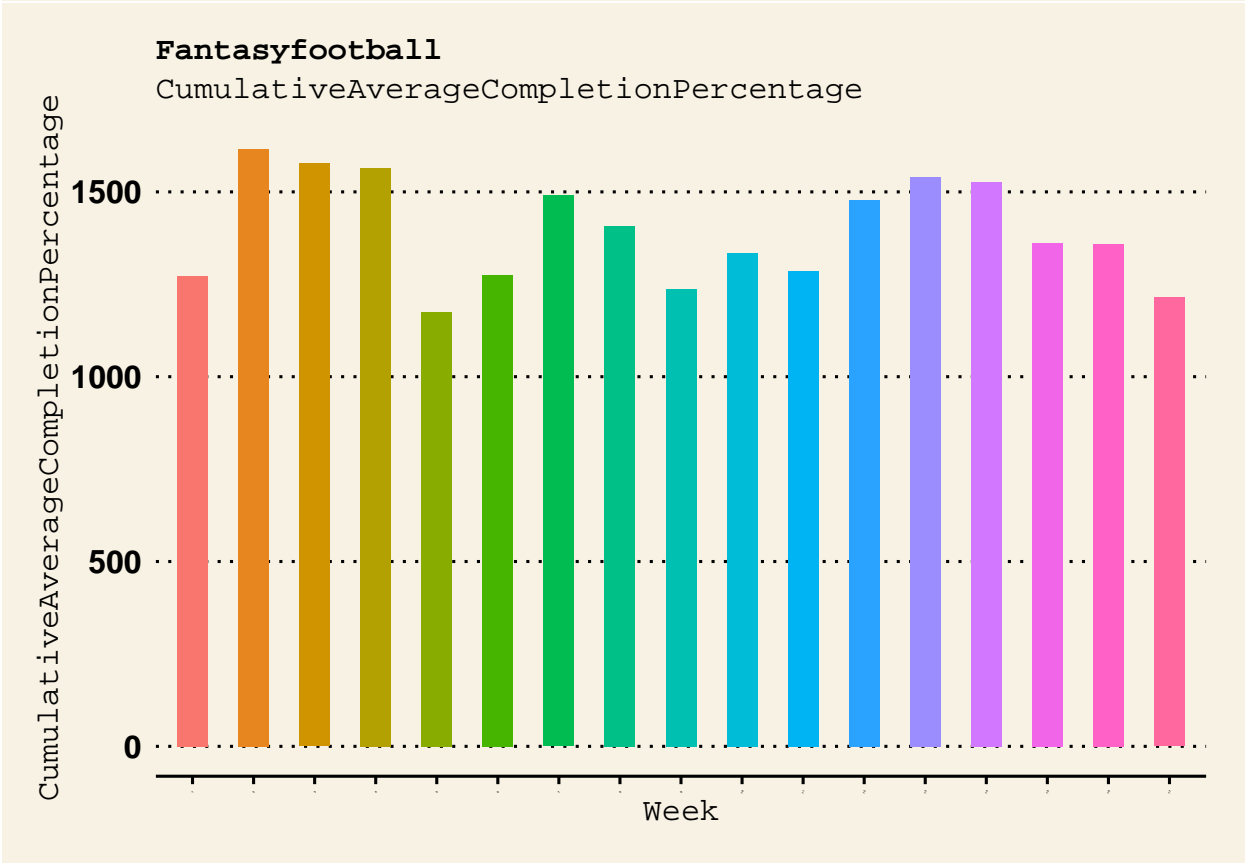
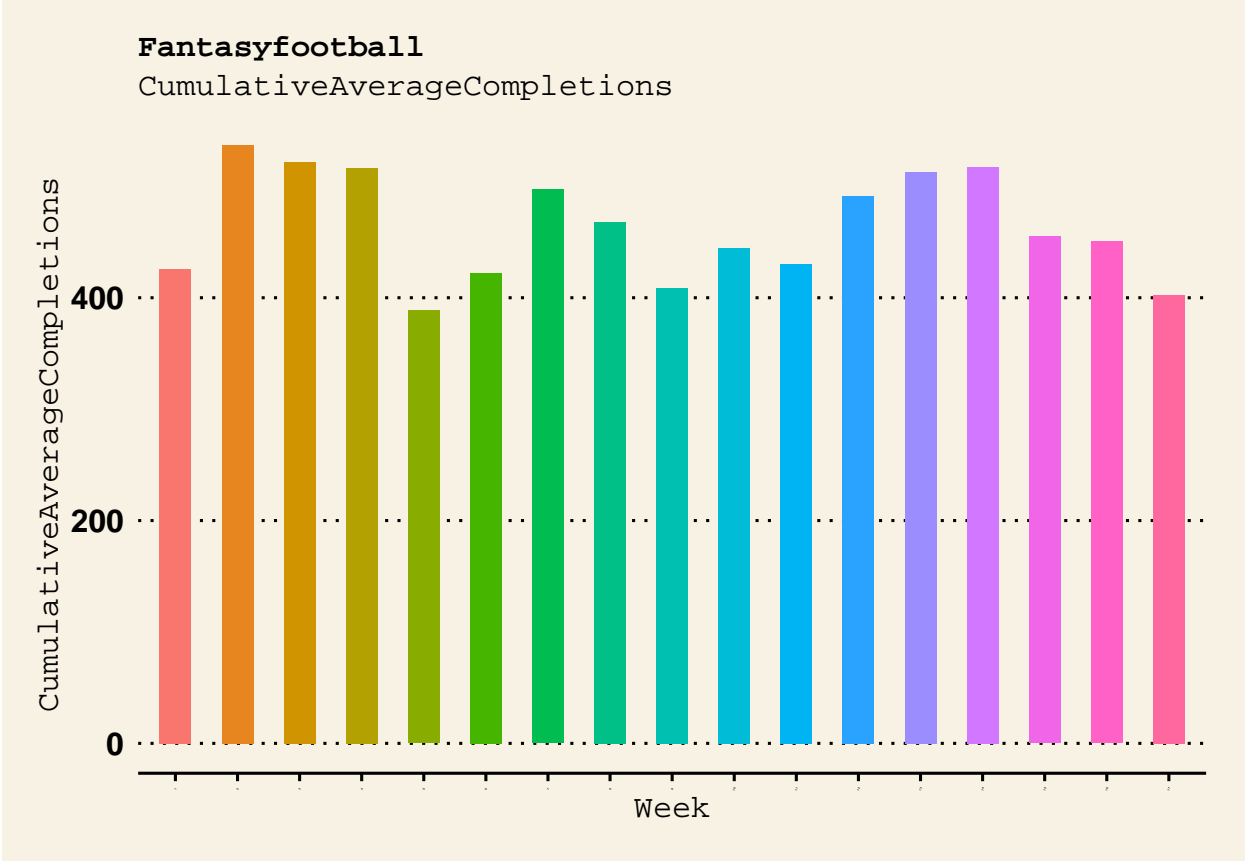
## 8.5 Bar plots Derived Feature set

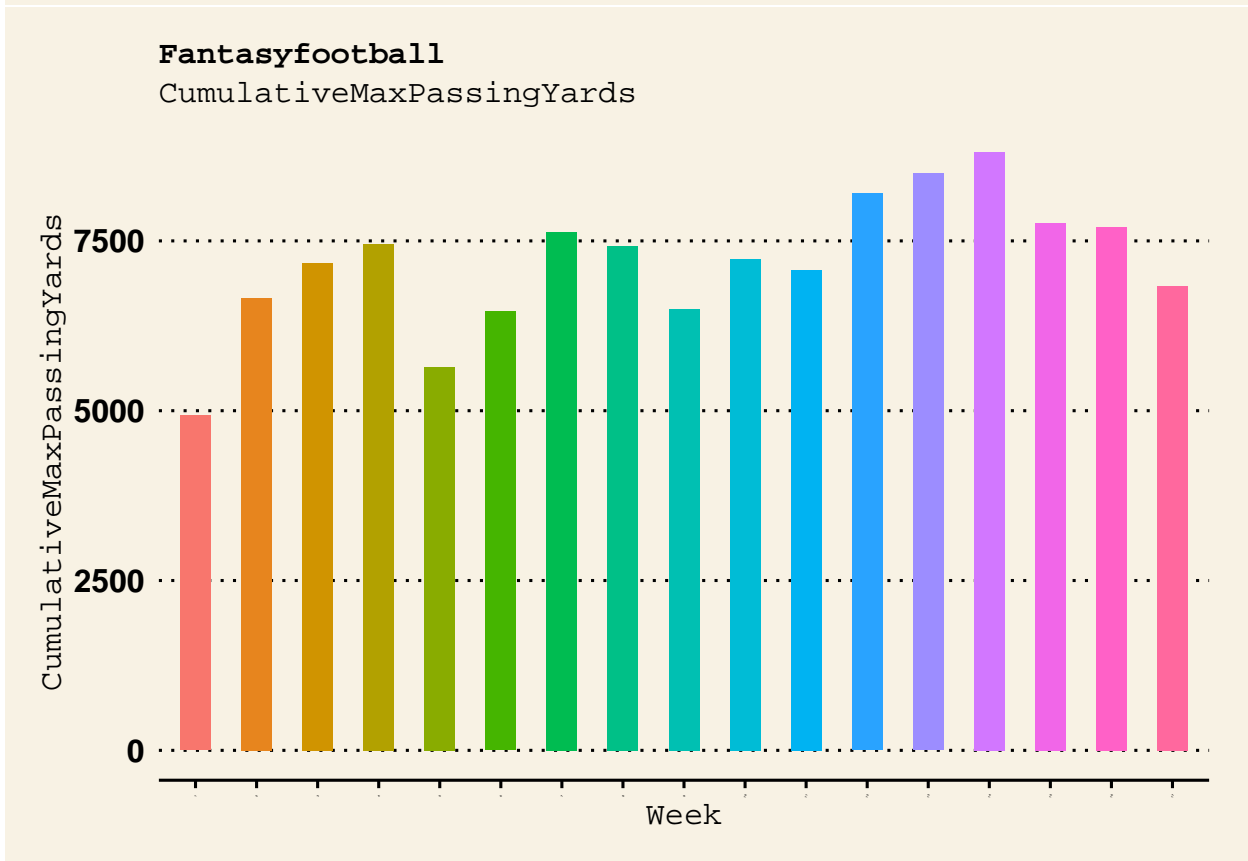
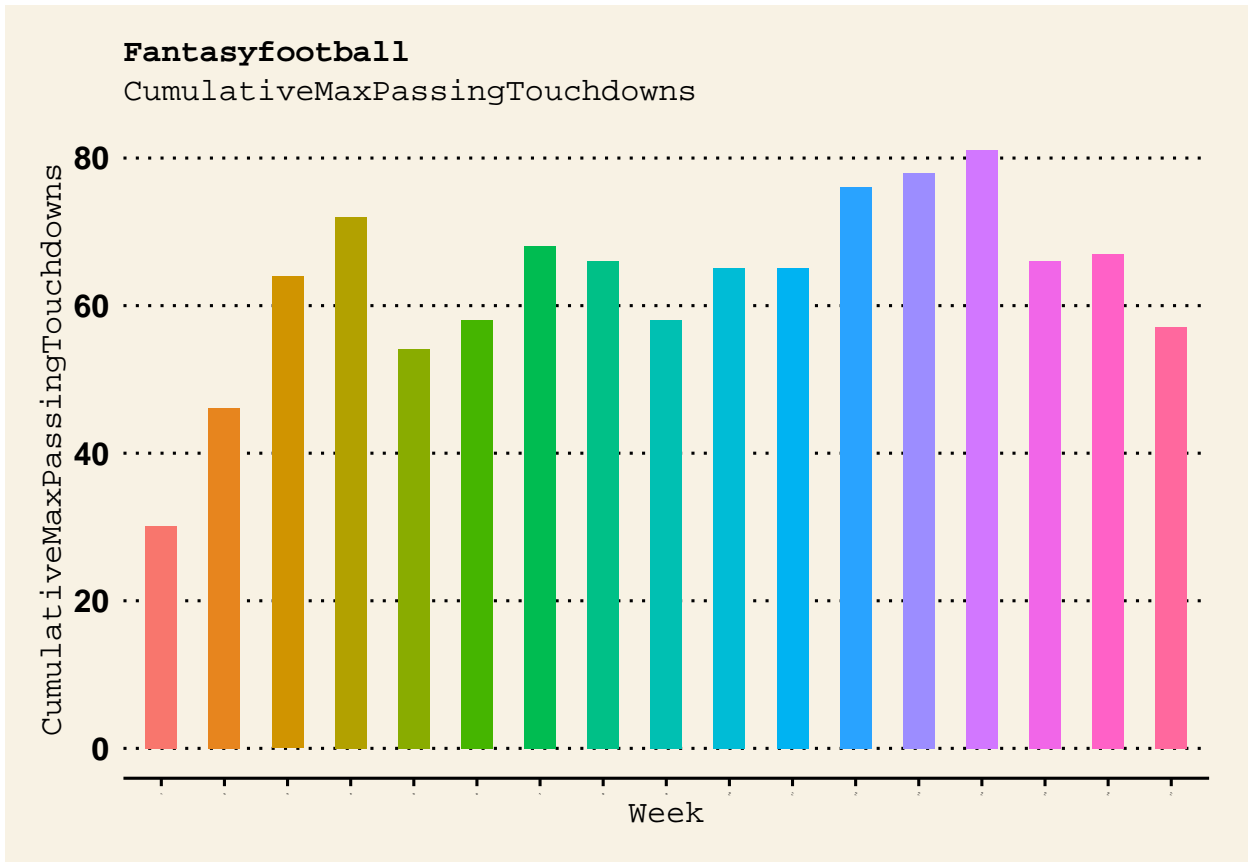
```
for(p in 2:20){
  bar_plot <- eda_derived %>%
    ggplot(aes_string(x="Week",
                      y=names(eda_derived[p]),
                      fill="Week",group="Week"))+
    ggtitle(names(eda_derived[p]))+
    geom_bar(stat="identity", width=.5, show.legend = FALSE)+
    xlab("Week")+
    ylab(names(eda_derived[p]))+
    labs(title="Fantasyfootball",
         subtitle=names(eda_derived[p]),
         aption="Source: Fantasyfootball")+
    theme_wsj()+
    theme(plot.title = element_text(size = rel(0.5)),
          plot.subtitle = element_text(size = rel(0.5)),
          axis.text.x = element_text(angle=65, vjust=0.6,size=1),
          axis.title = element_text(size = rel(0.5)),
          legend.position = "right",
          legend.direction = "vertical",
          legend.title = element_text(size = rel(0.5)))
  print(bar_plot)
}
```

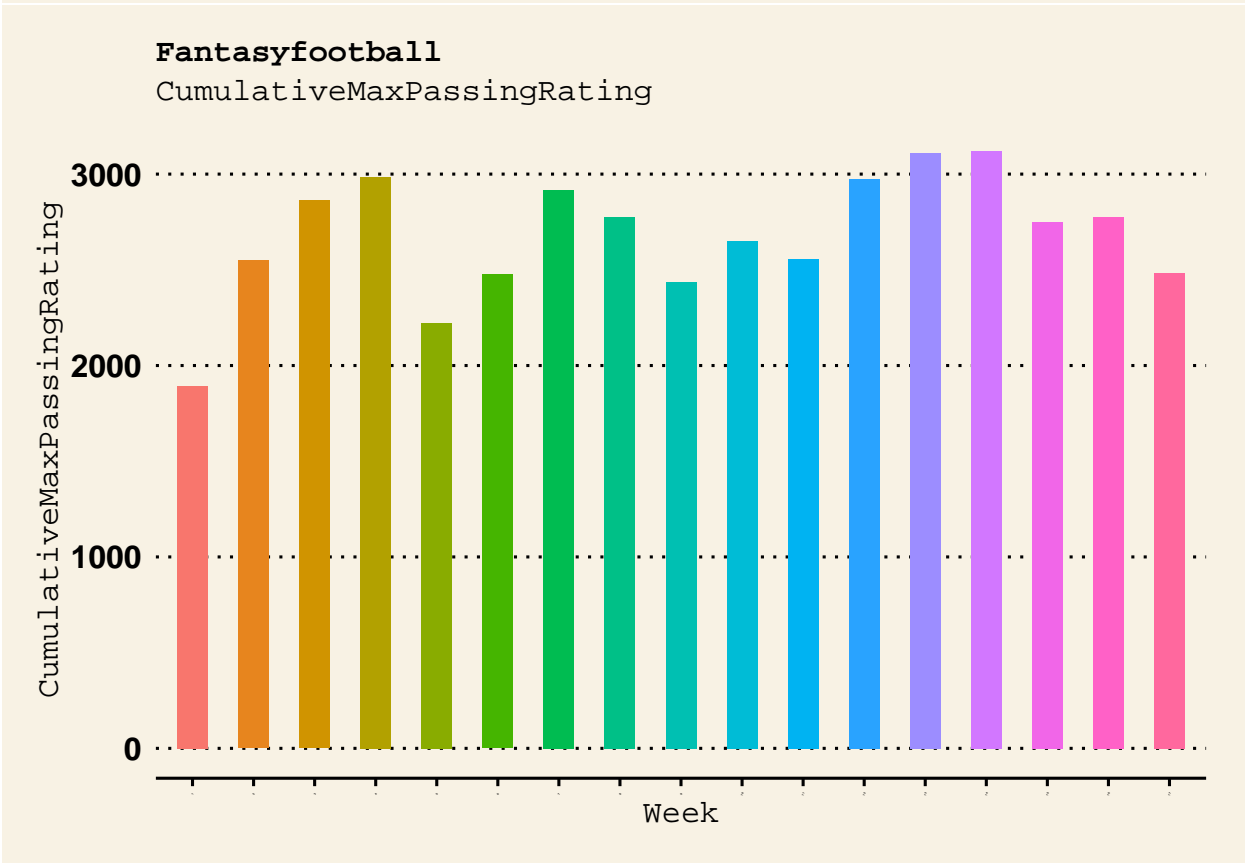
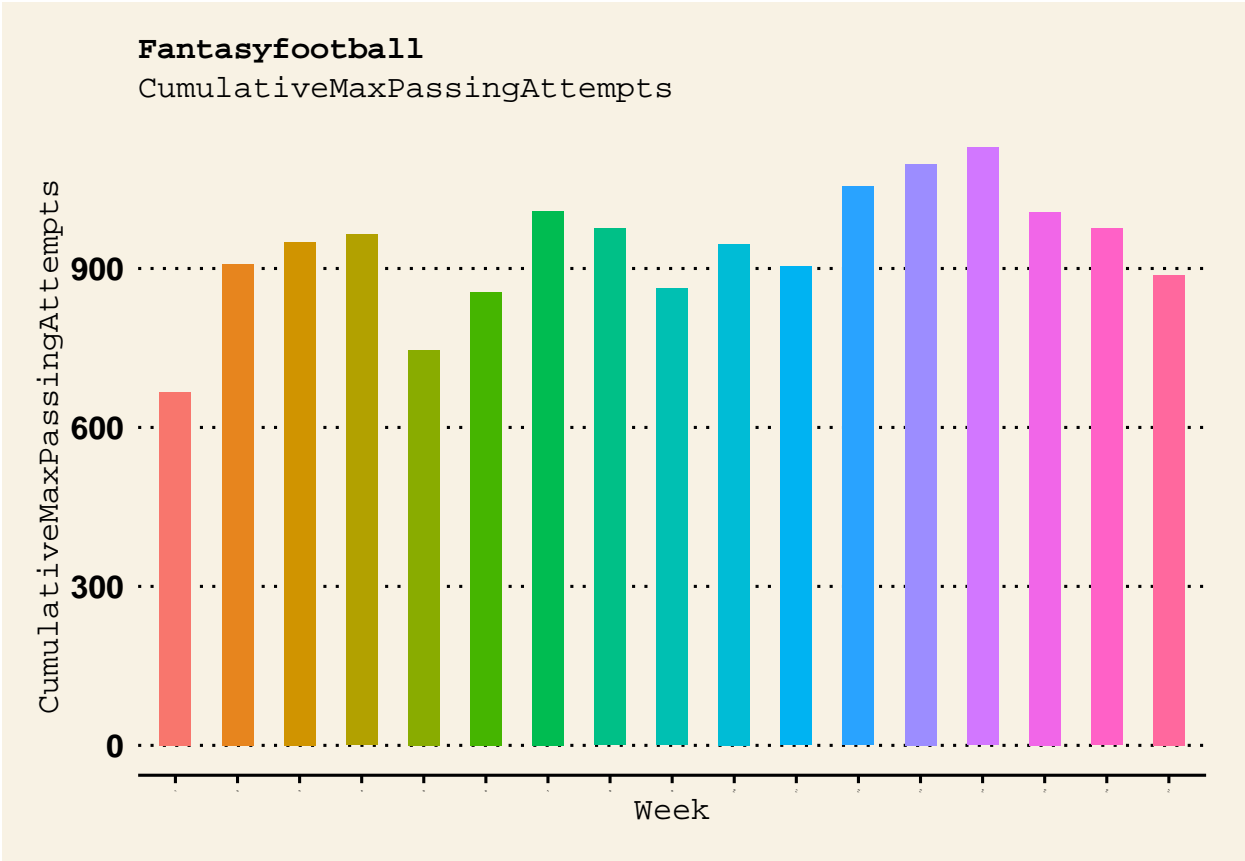




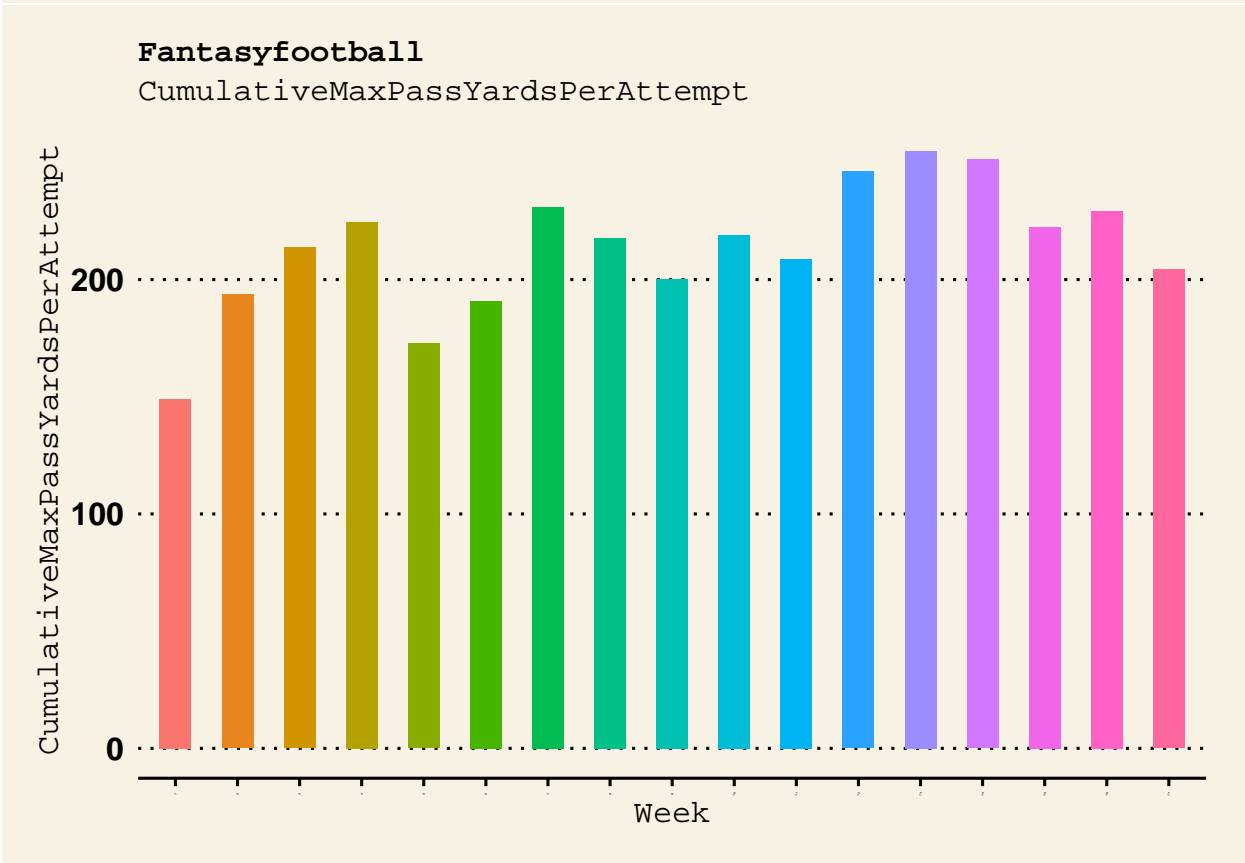
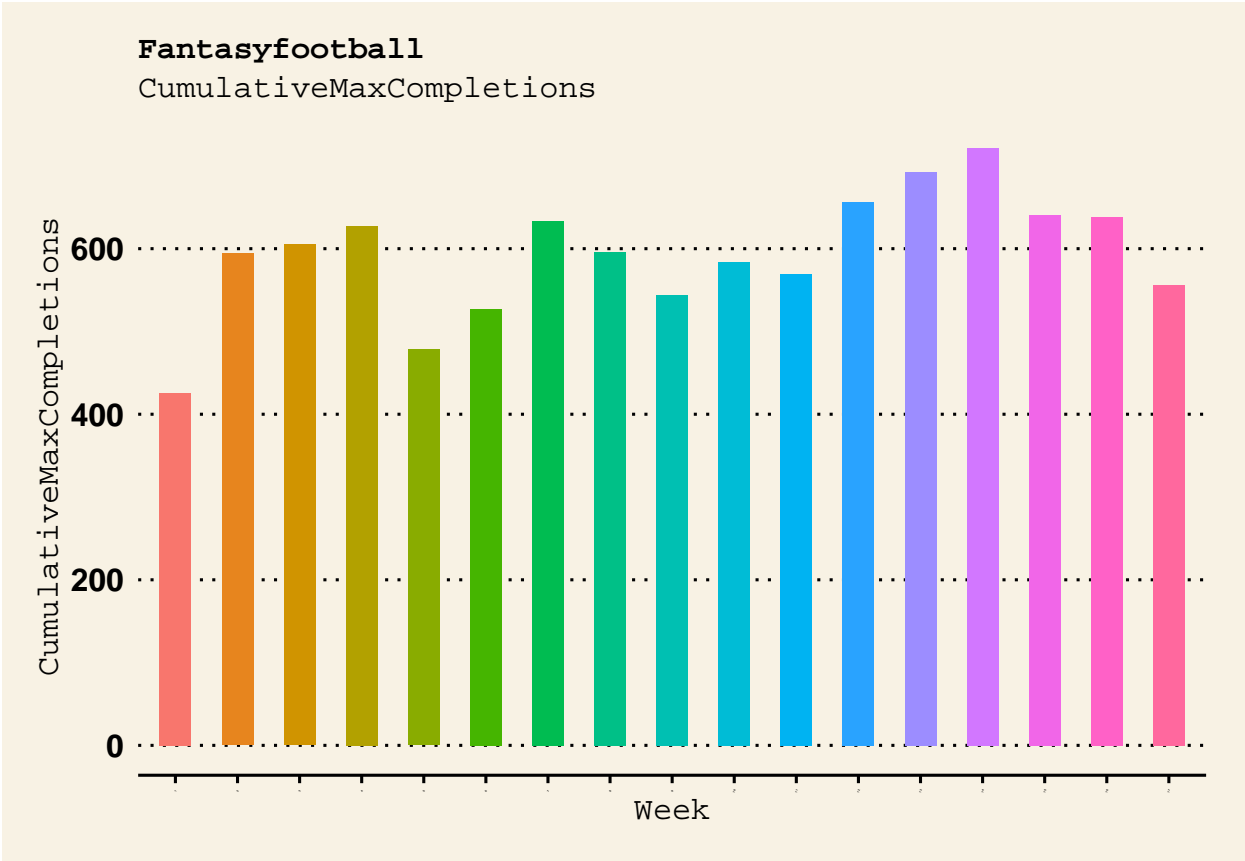


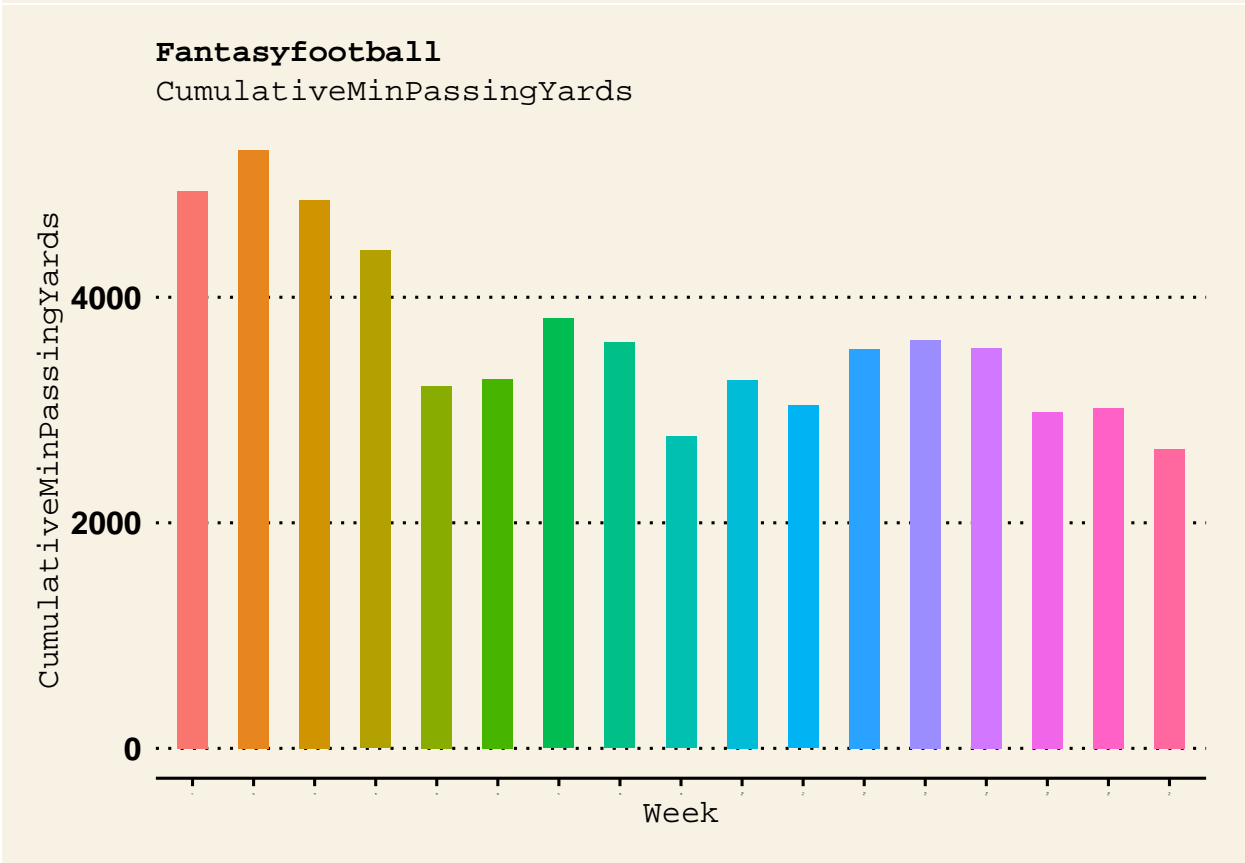
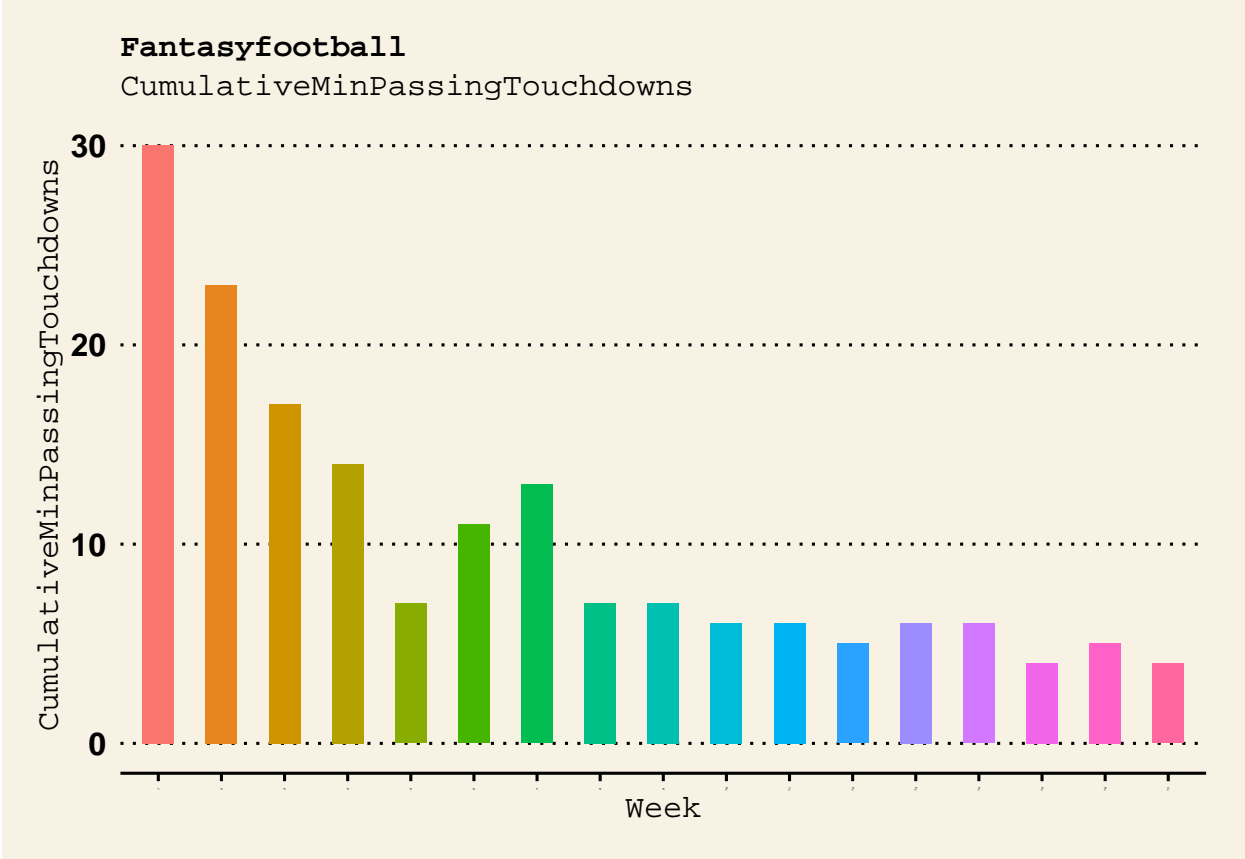


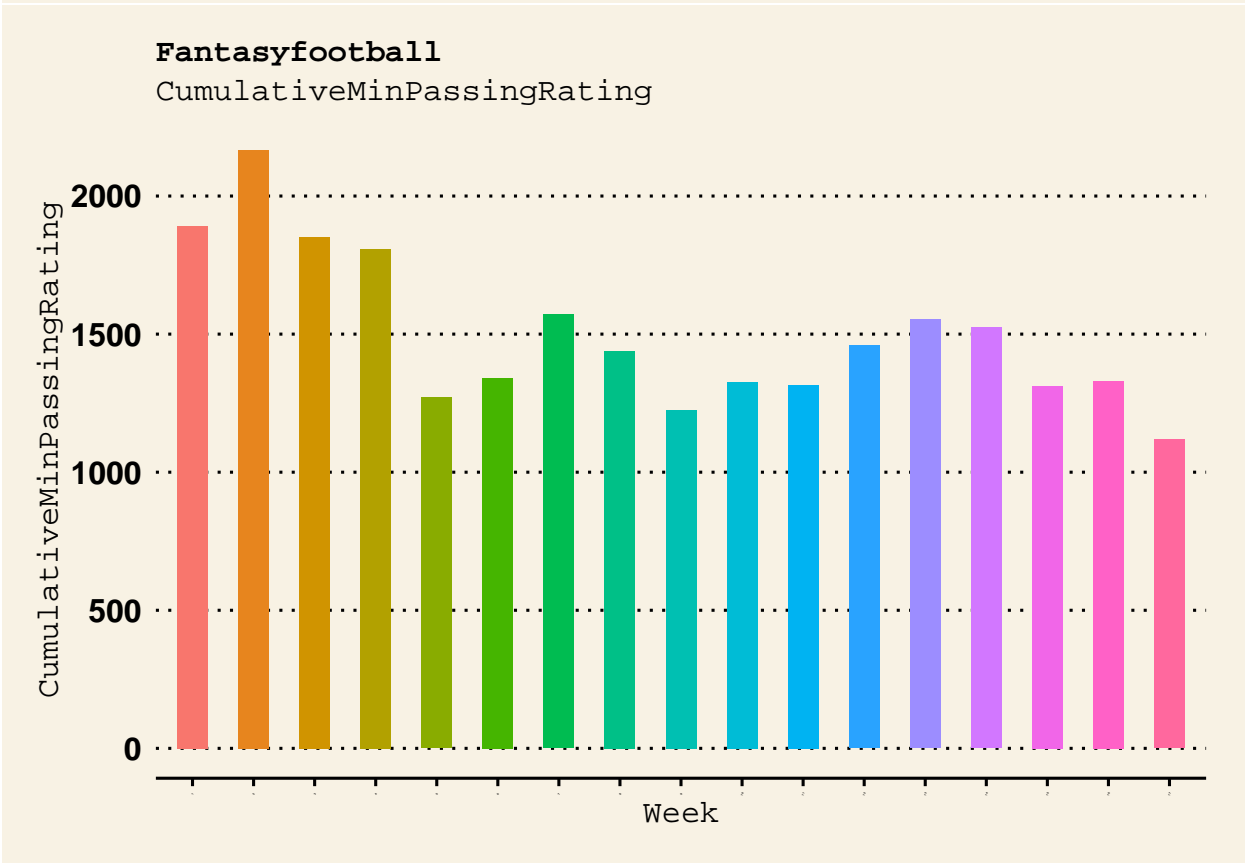
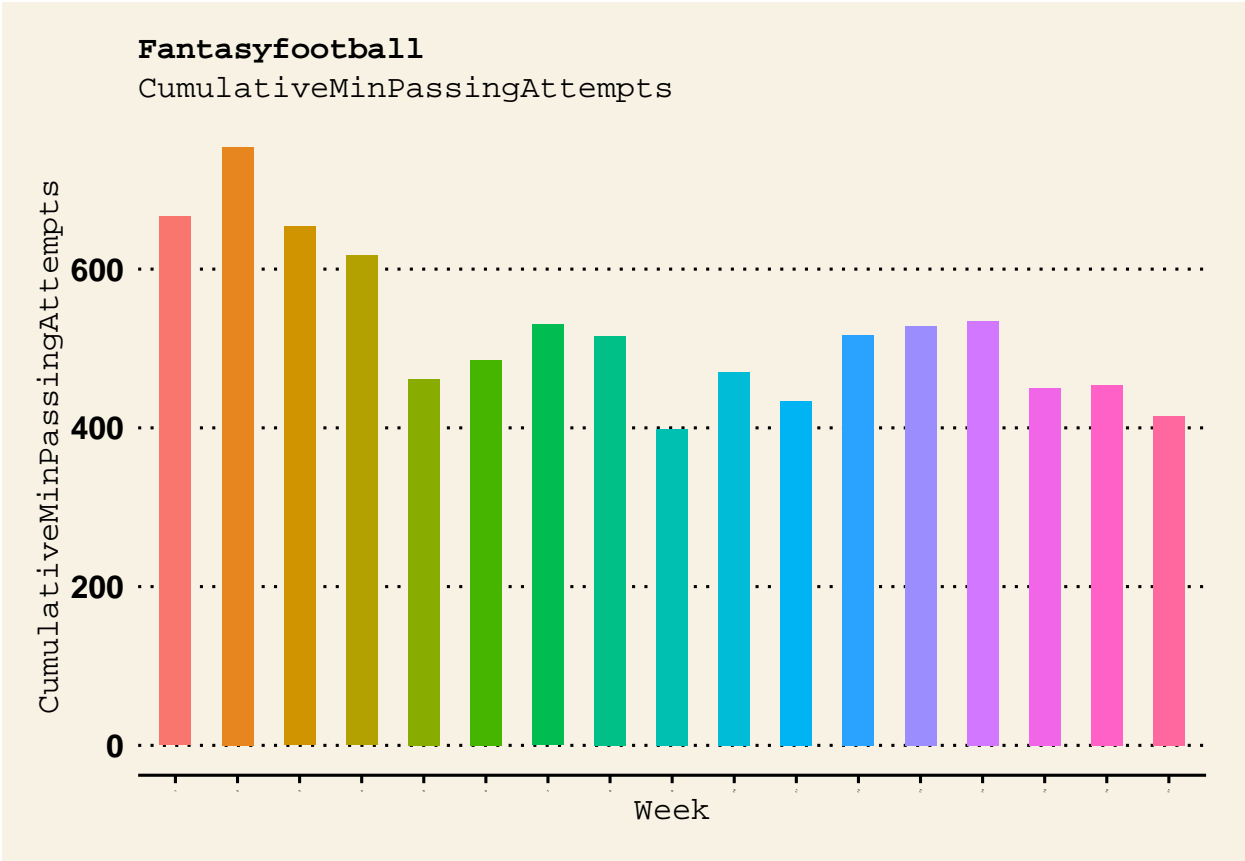


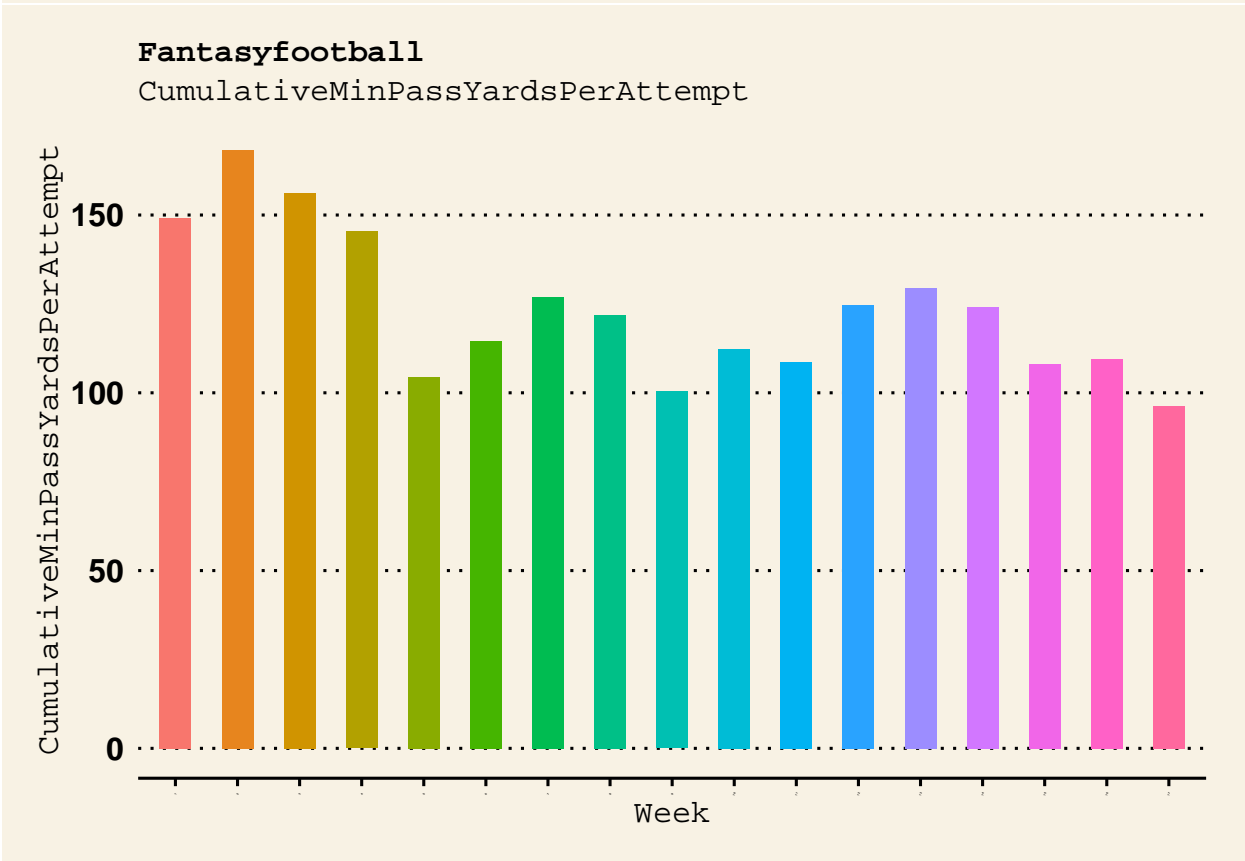
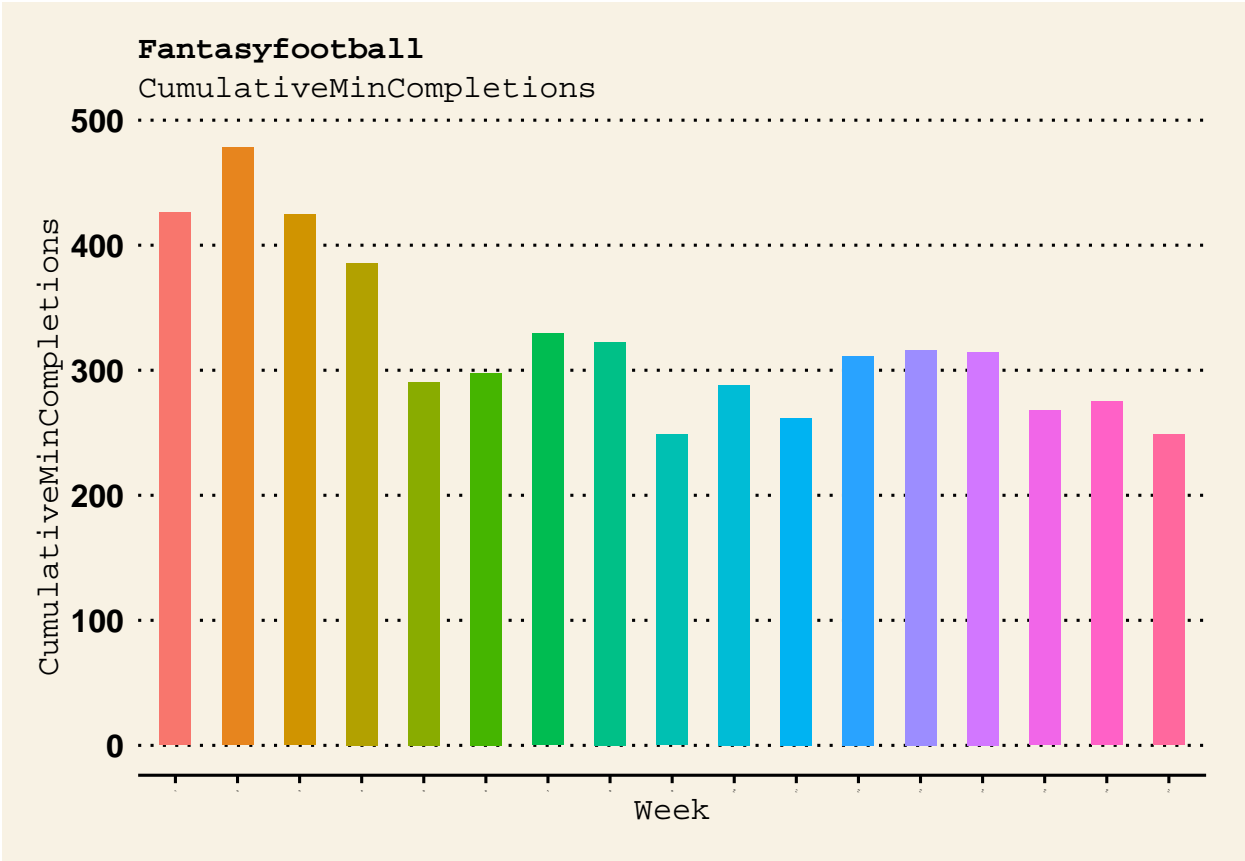










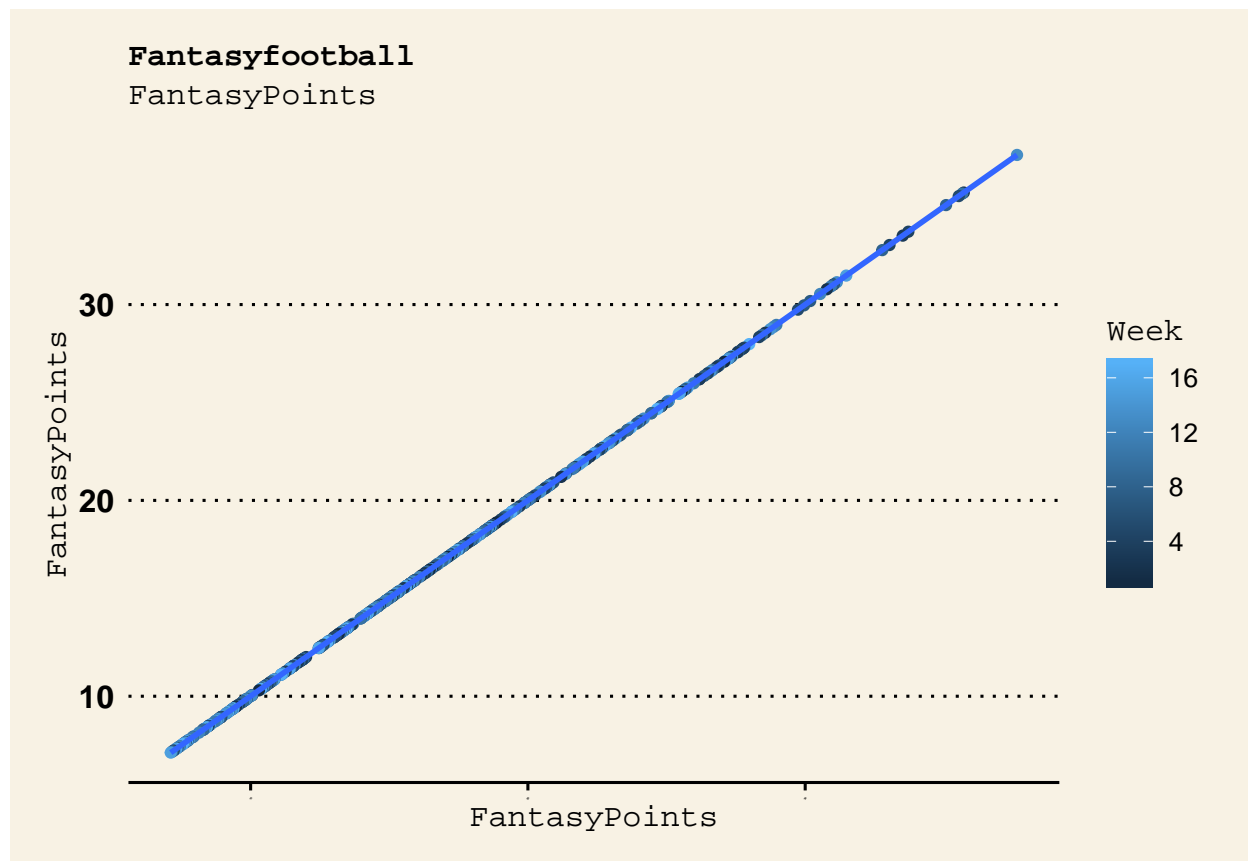


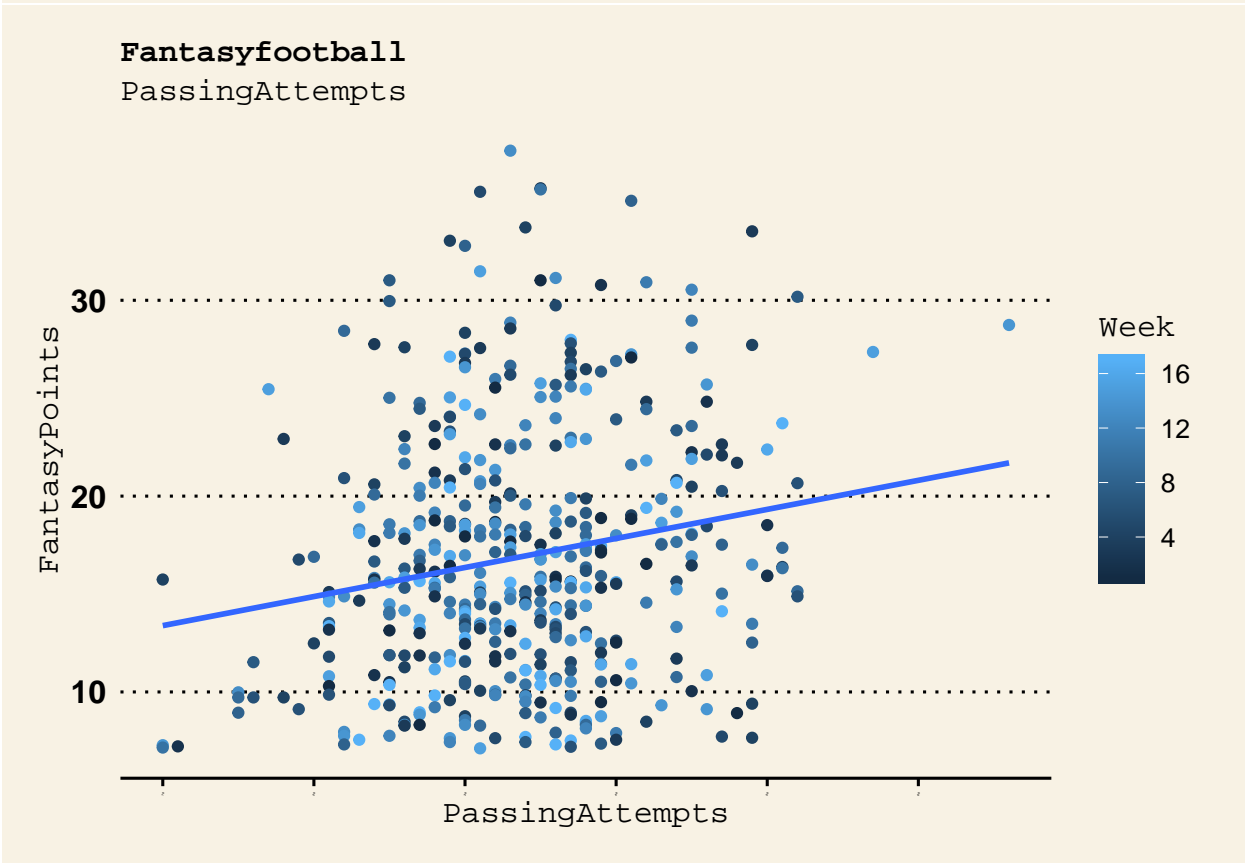
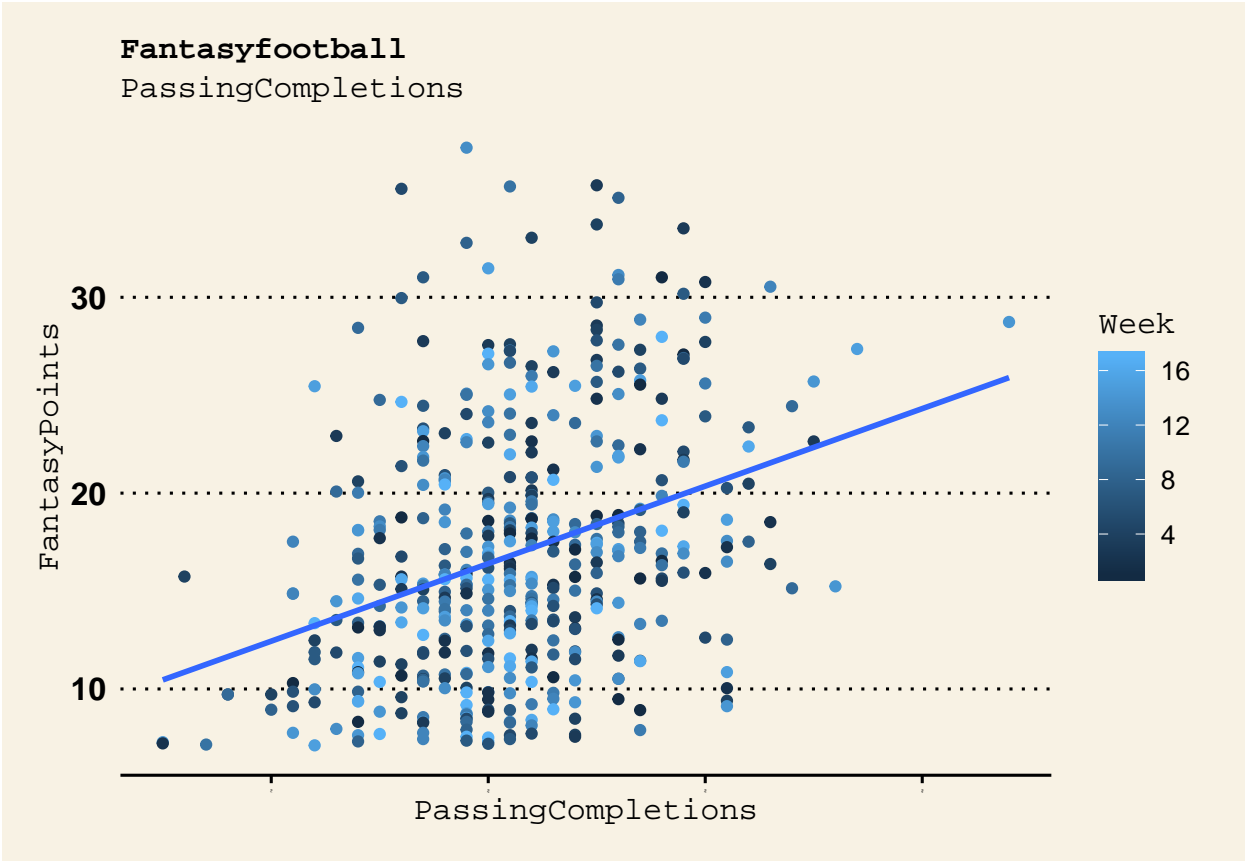
## 9 : Scatterplots

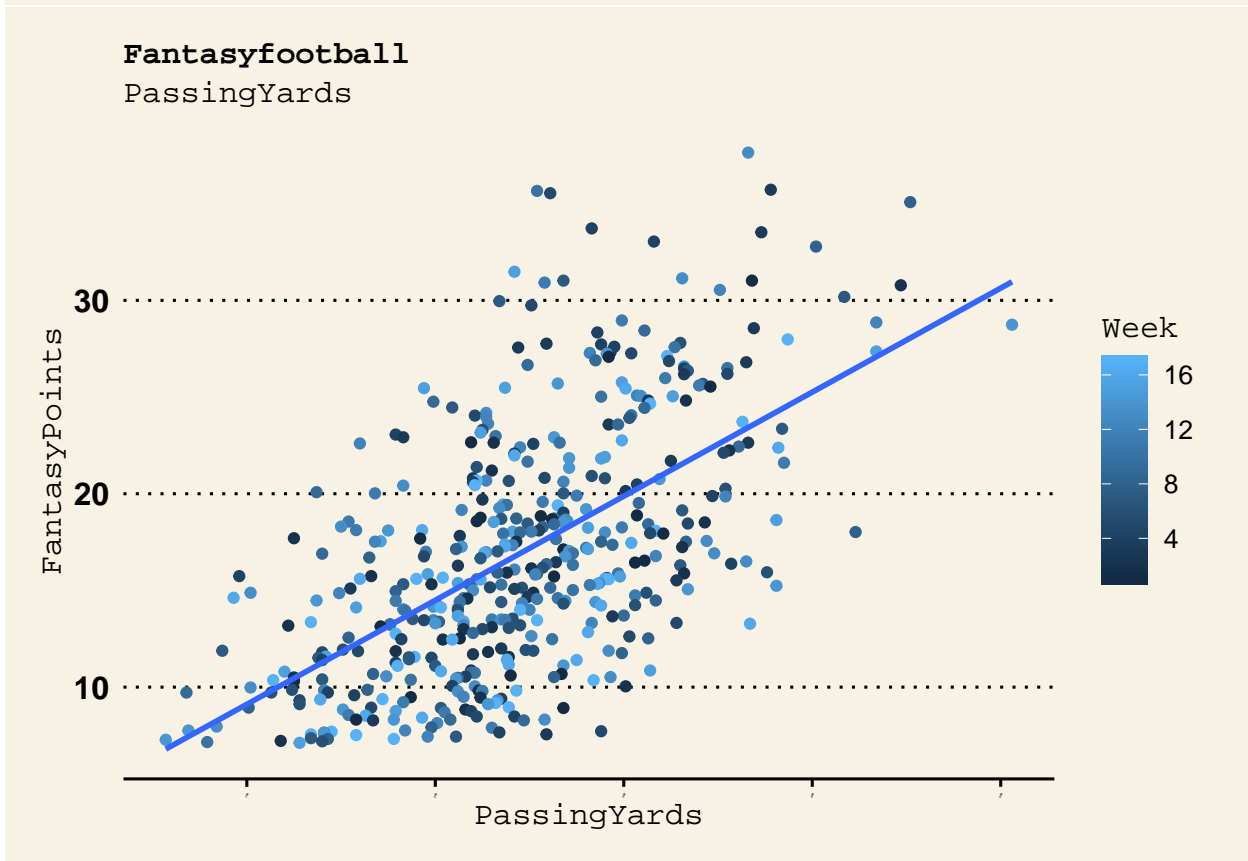
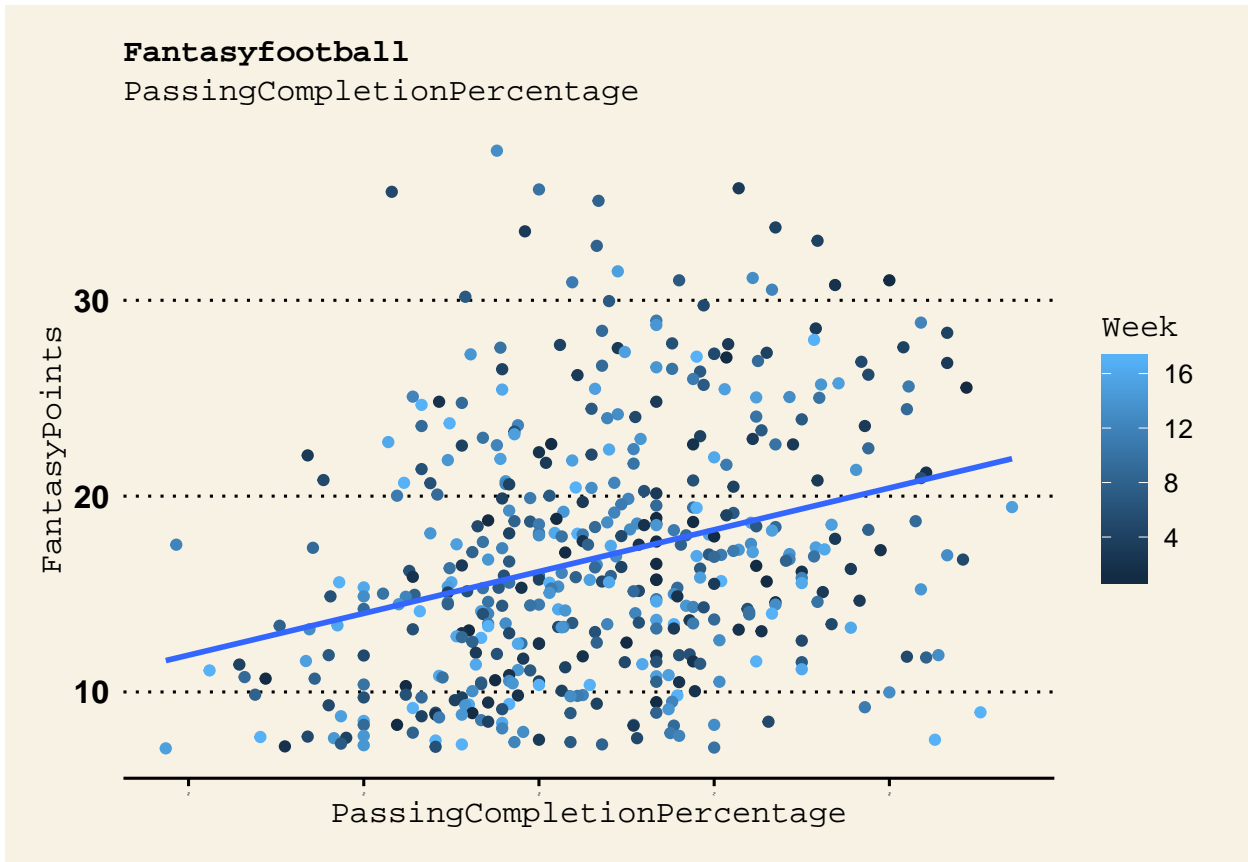
We are trying to capture what the NEXT value for fantasy points is likely to be - these are all going to be the current week since the fantasy score is a linear combination of the predictors for any given week. We need to shift the data before running these charts.

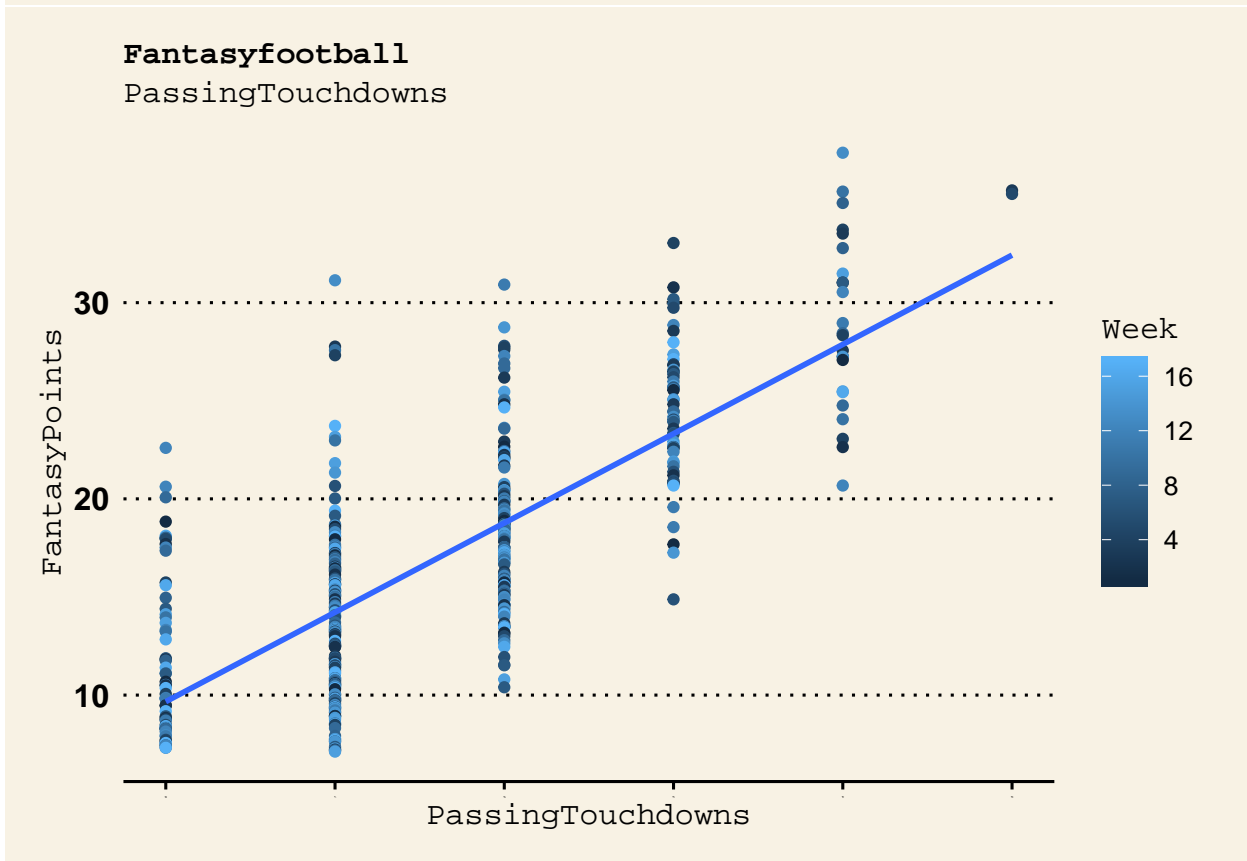
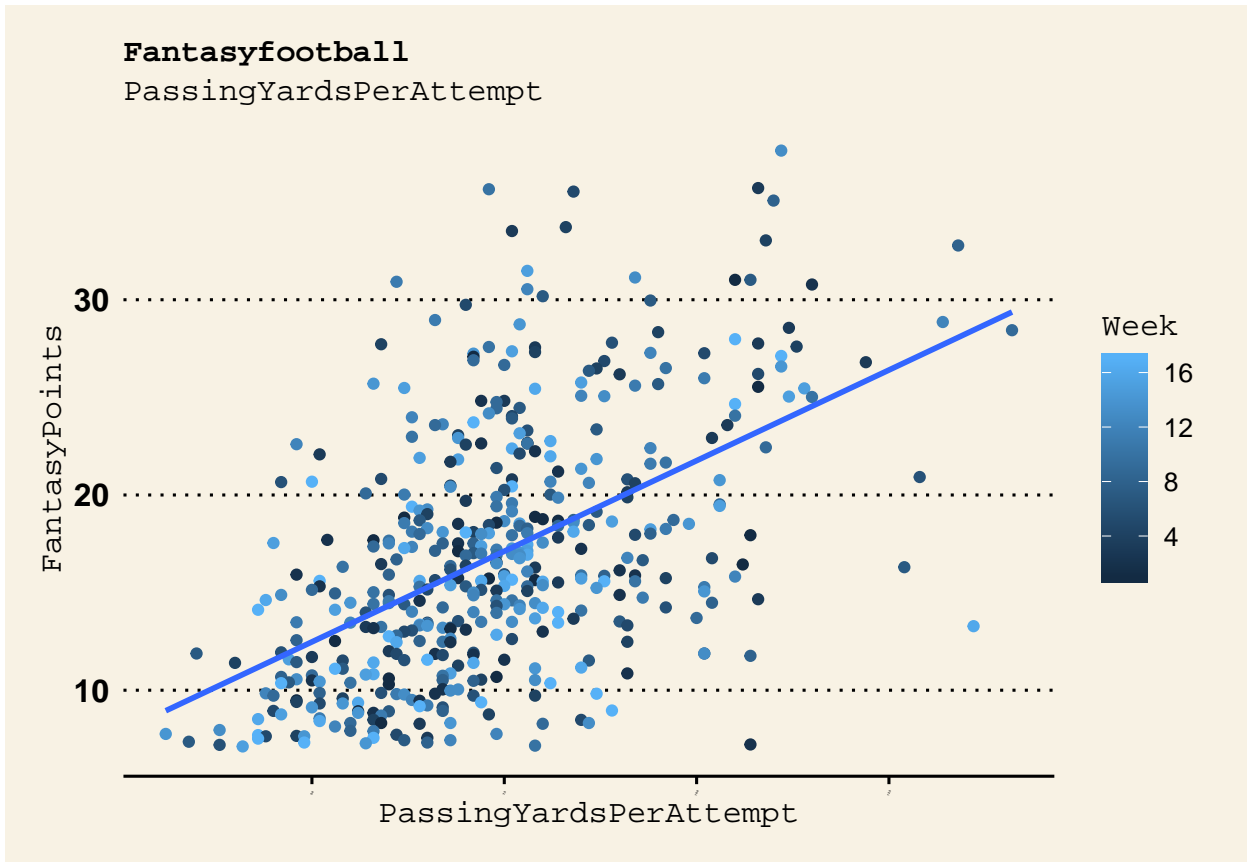
## 9.1 : Scatterplots for base features

```
for (i in 2:15) {  
  base_scatter <- eda_base %>%  
    ggplot(aes_string(y="FantasyPoints",x=names(eda_base[i]),color="as.numeric(Week)")) +  
    geom_point()+geom_smooth(method="lm",se=F)+  
    labs(title="Fantasyfootball",  
         subtitle=names(eda_base[i]),  
         aption="Source: Fantasyfootball")+  
    labs(color="Week")+  
    theme_wsj()+  
    theme(plot.title = element_text(size = rel(0.5)),  
          plot.subtitle = element_text(size = rel(0.5)),  
          axis.text.x = element_text(angle=65, vjust=0.6,size=1),  
          axis.title = element_text(size = rel(0.5)),  
          legend.position = "right",  
          legend.direction = "vertical",  
          legend.title = element_text(size = rel(0.5)))  
  print(base_scatter)  
}
```

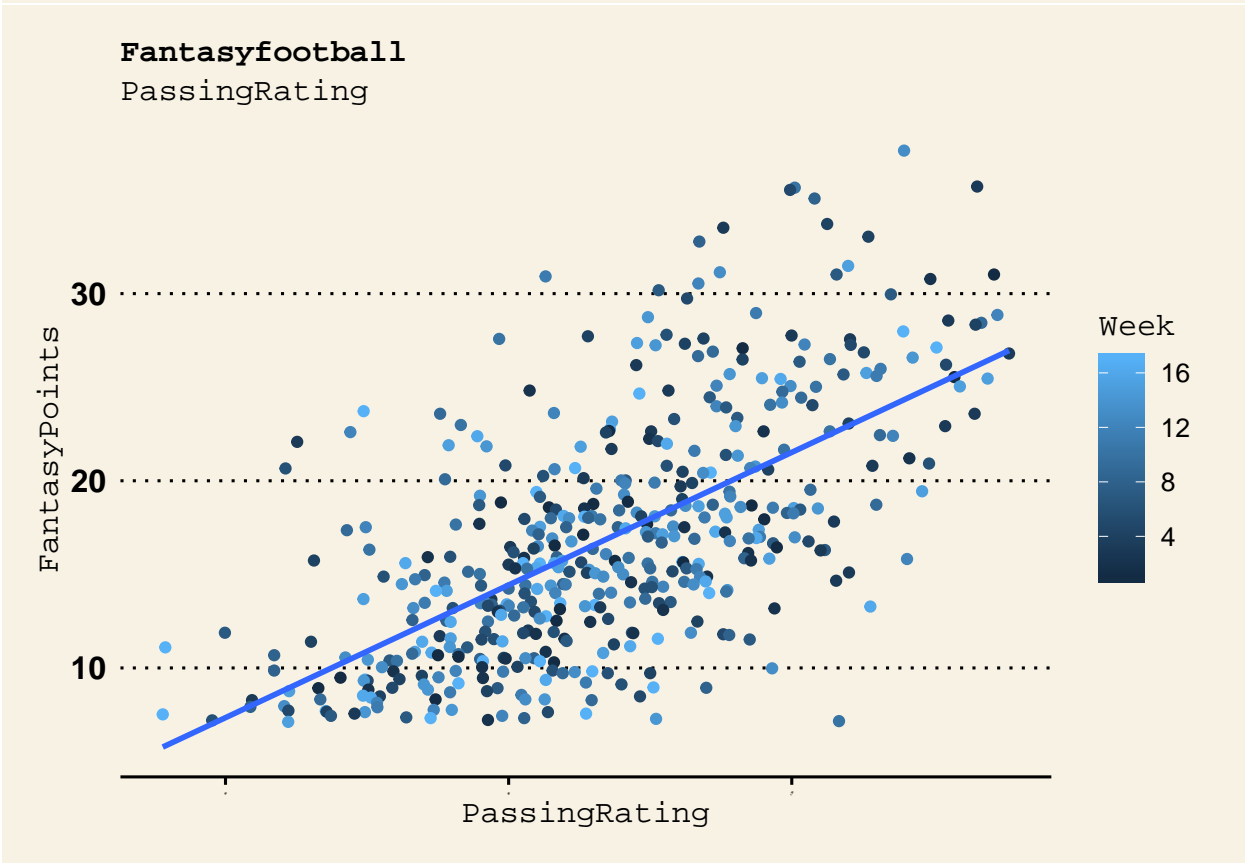
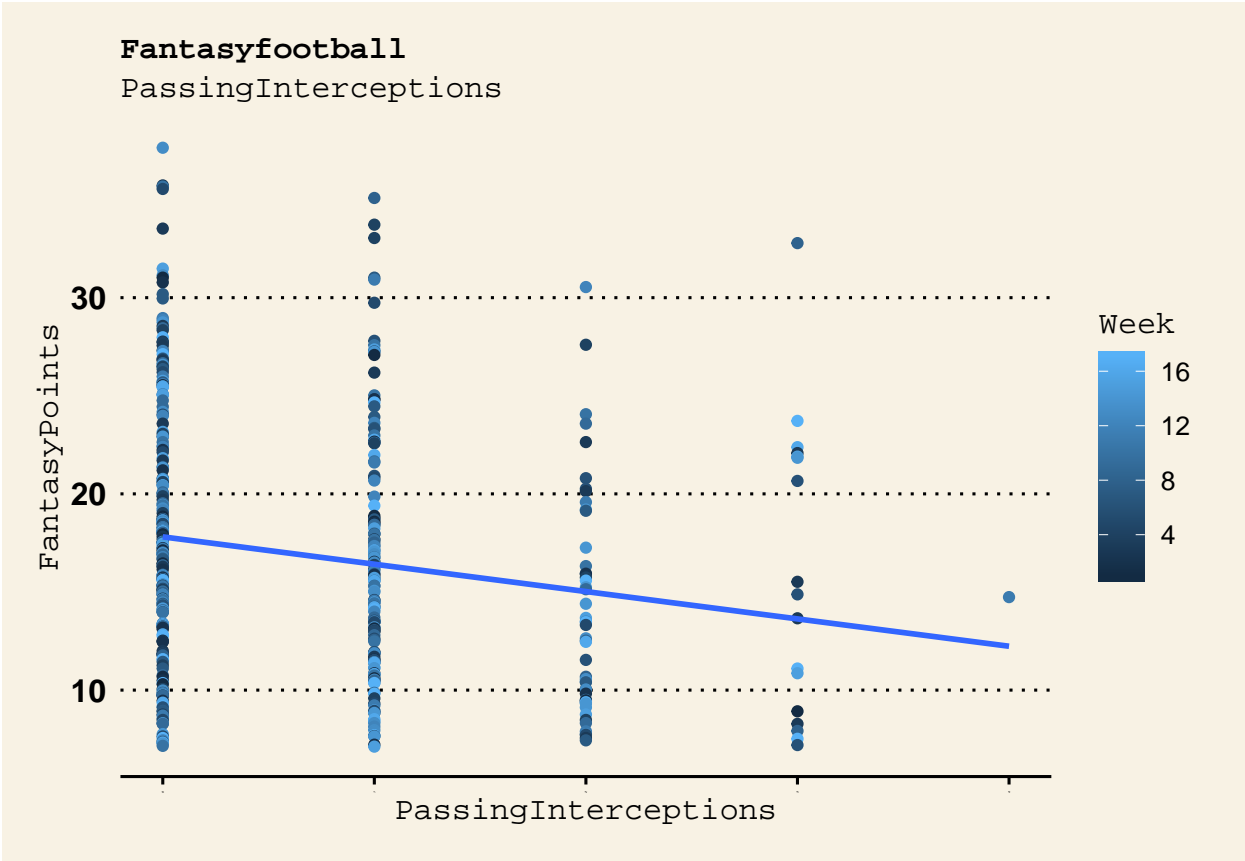


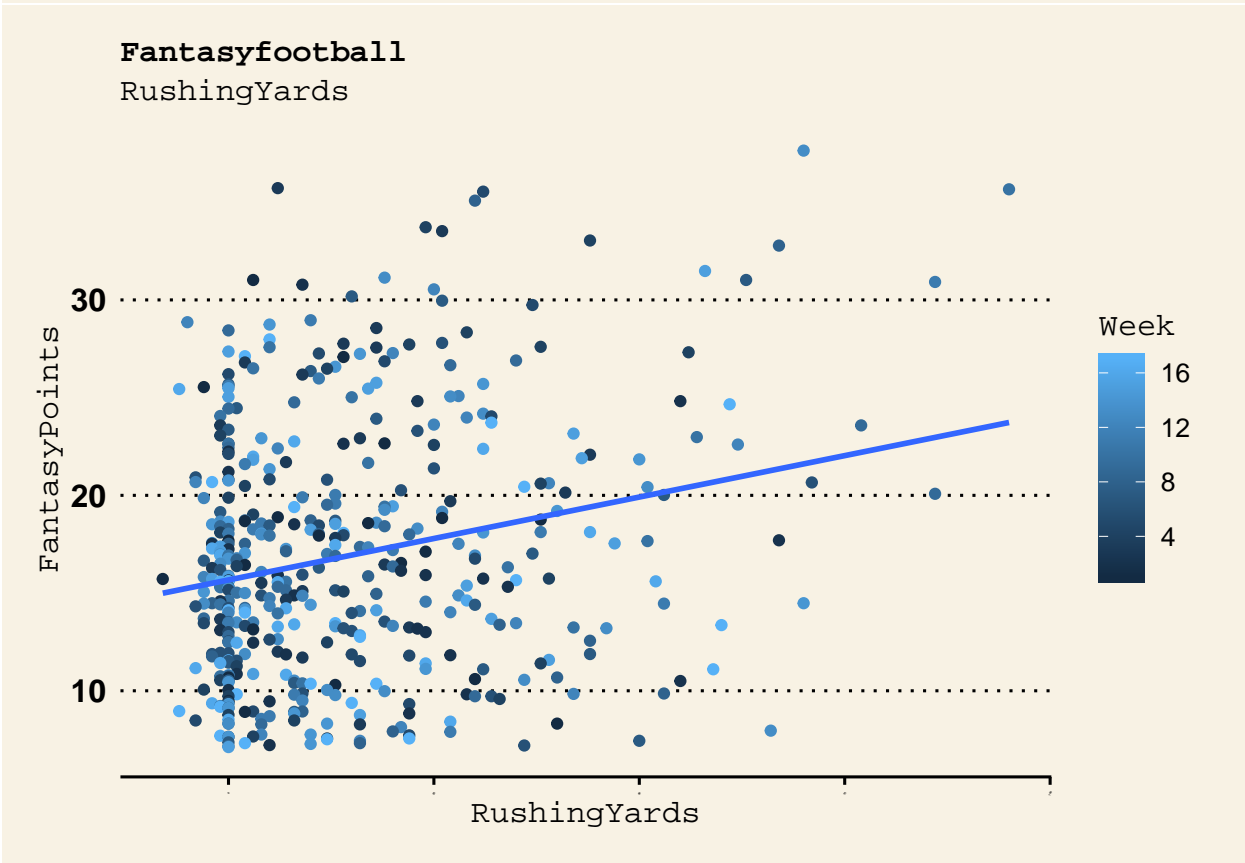
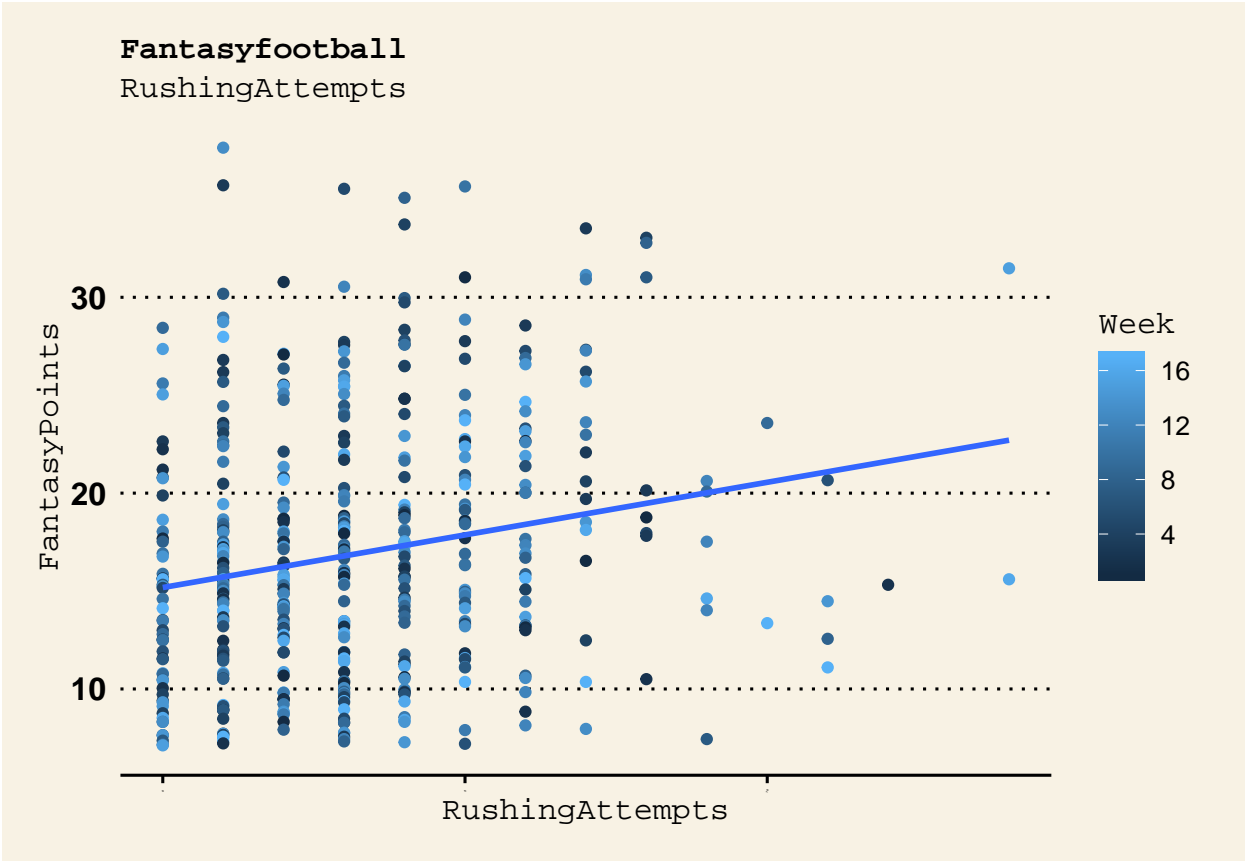


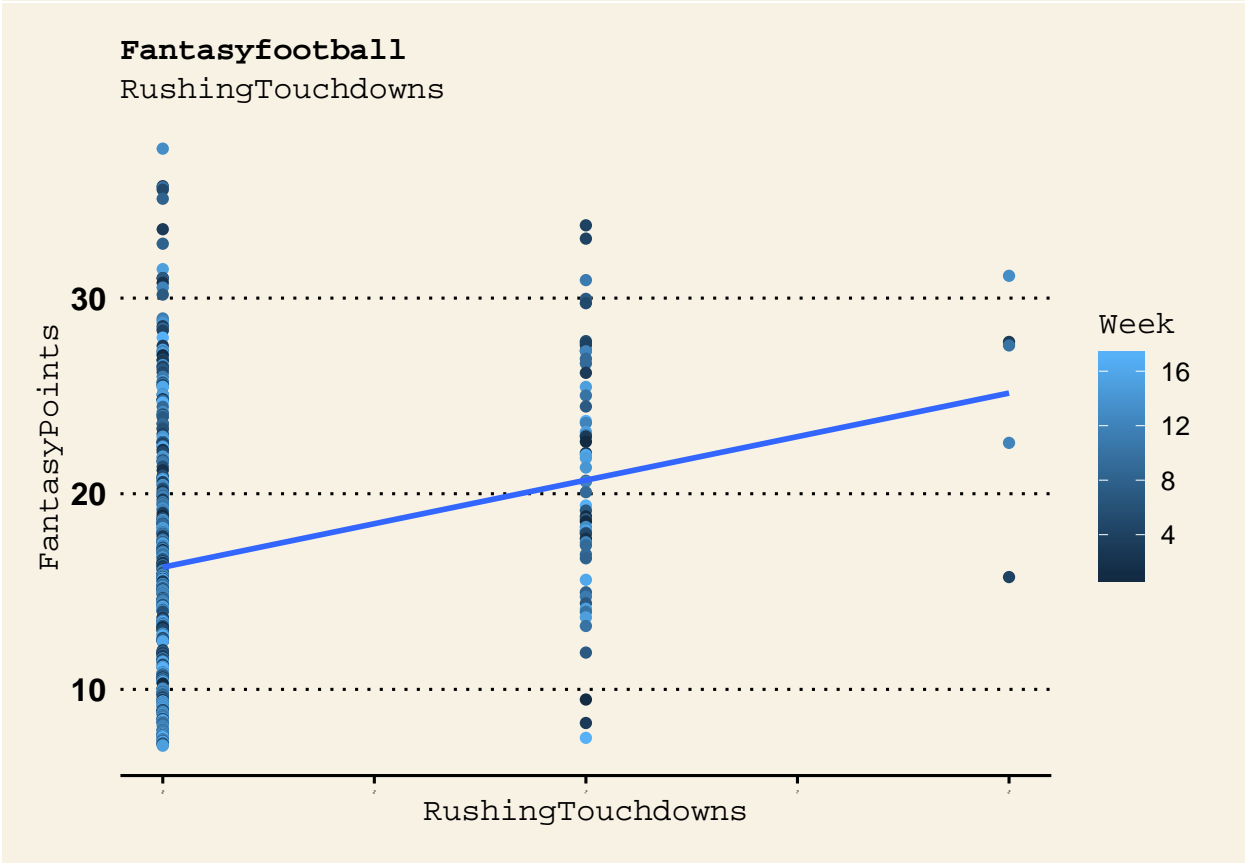
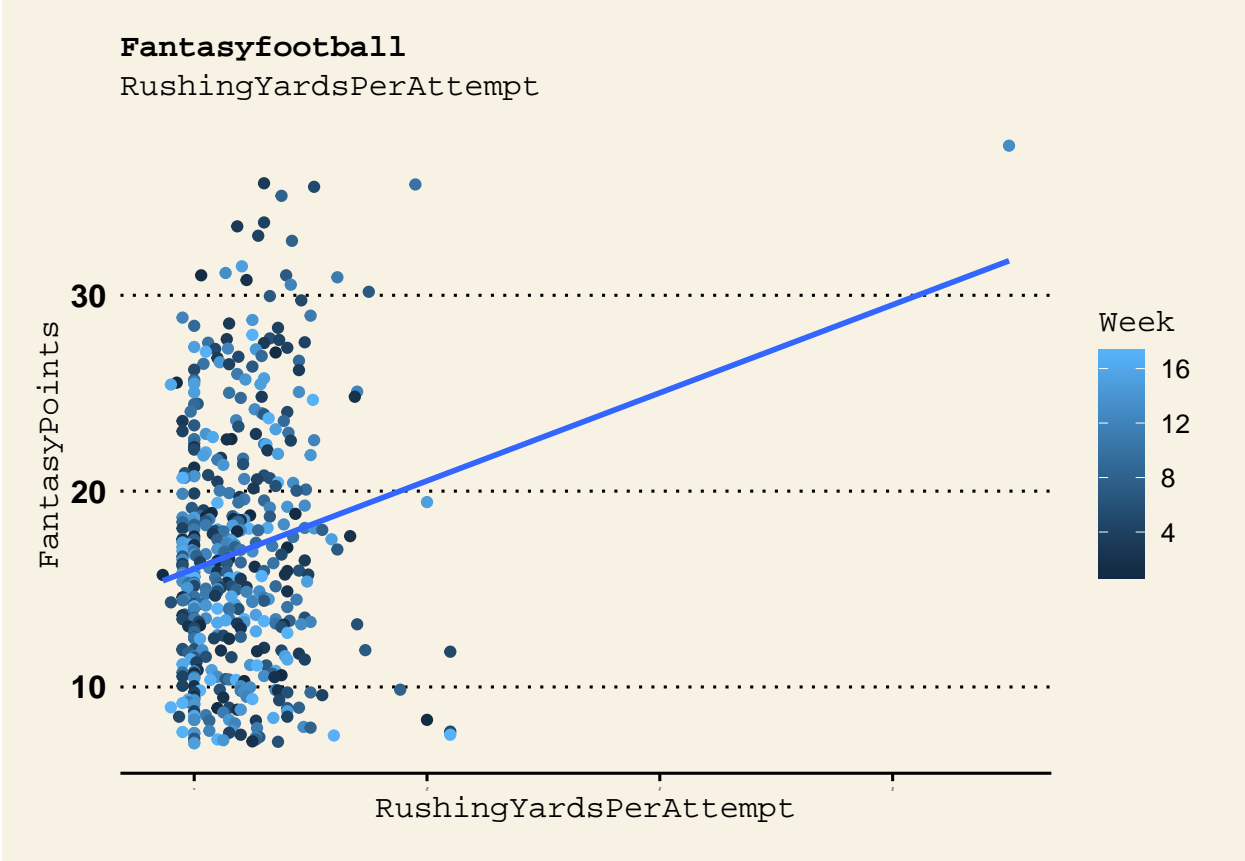


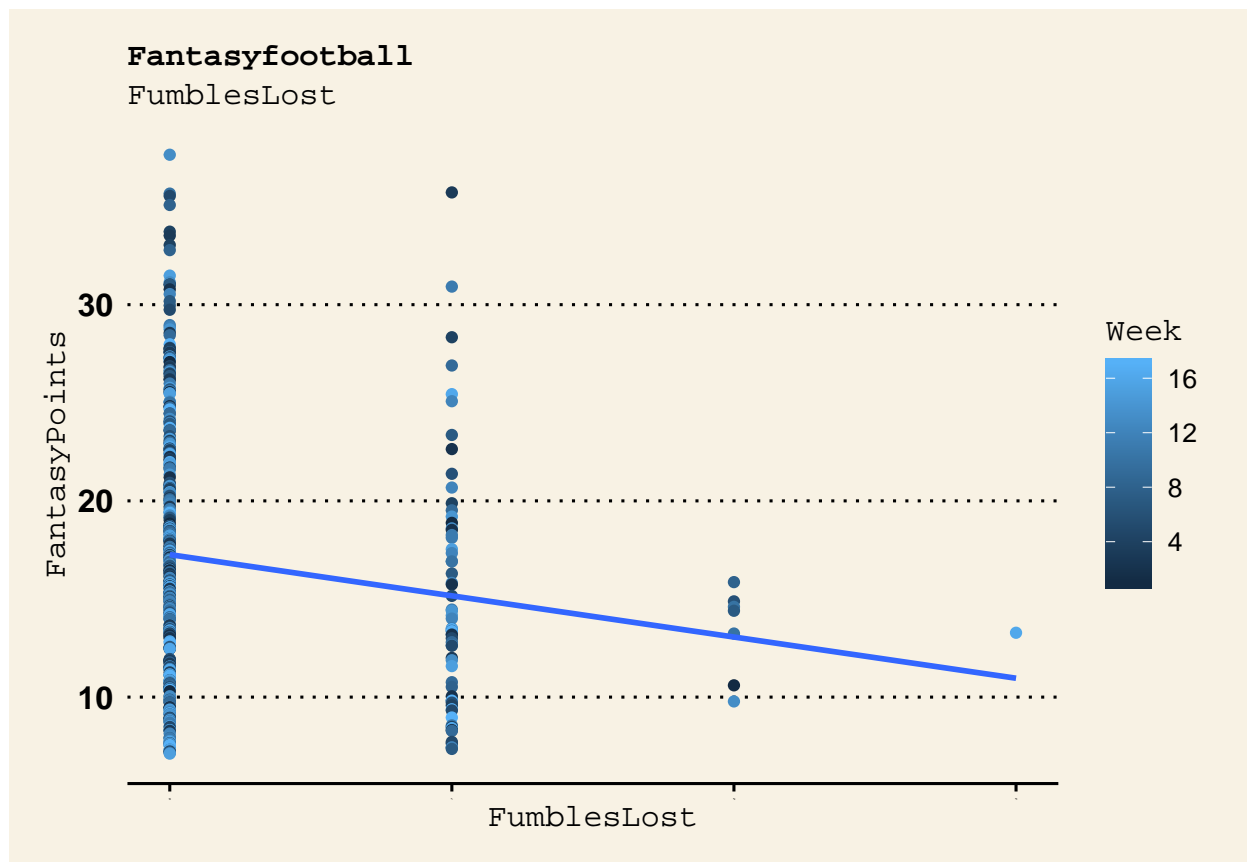






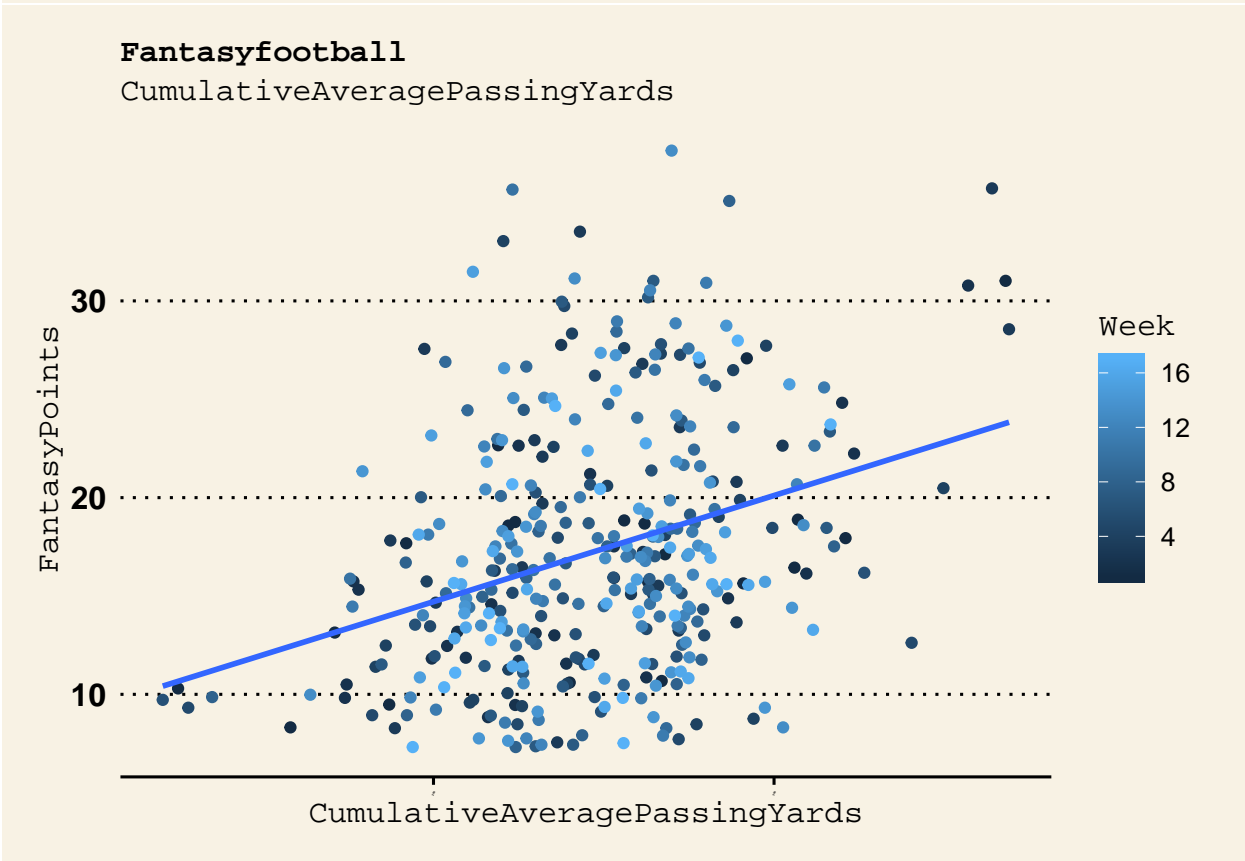
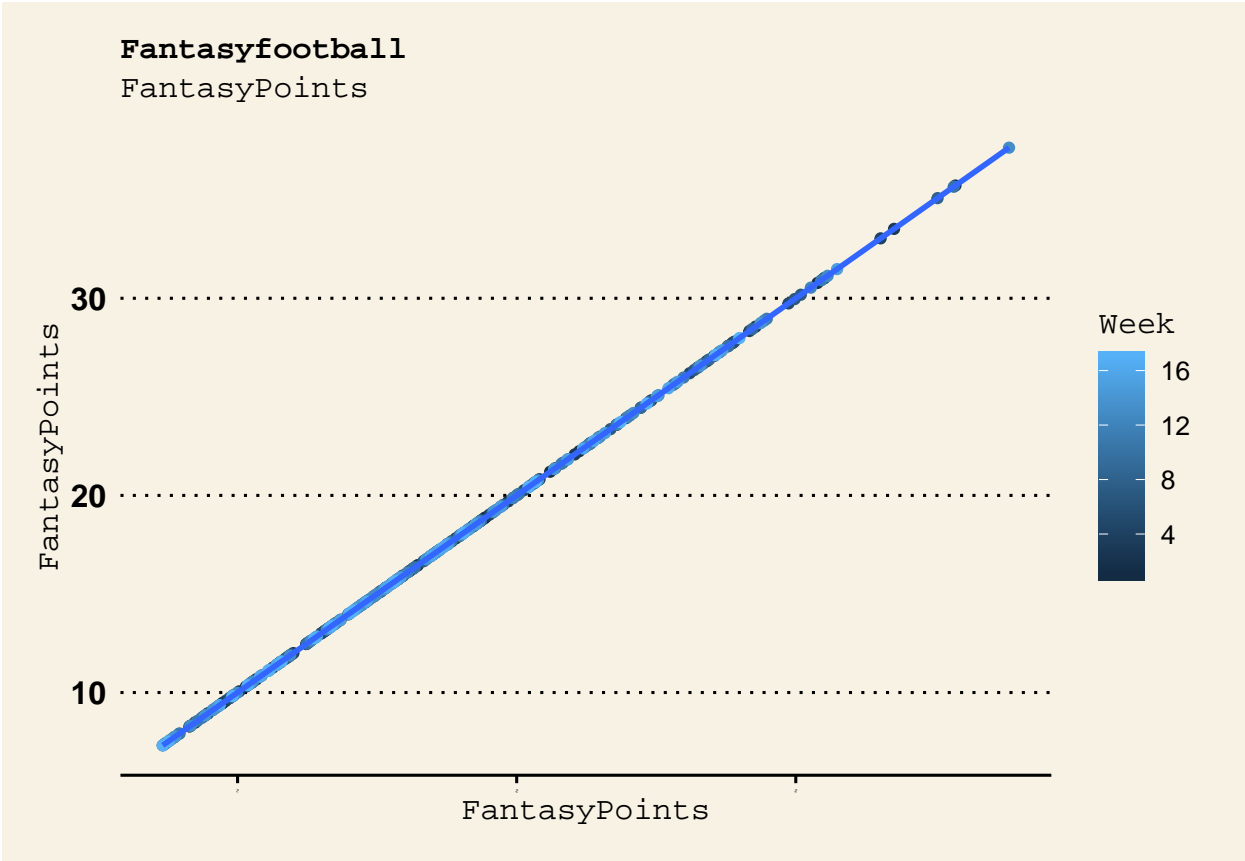


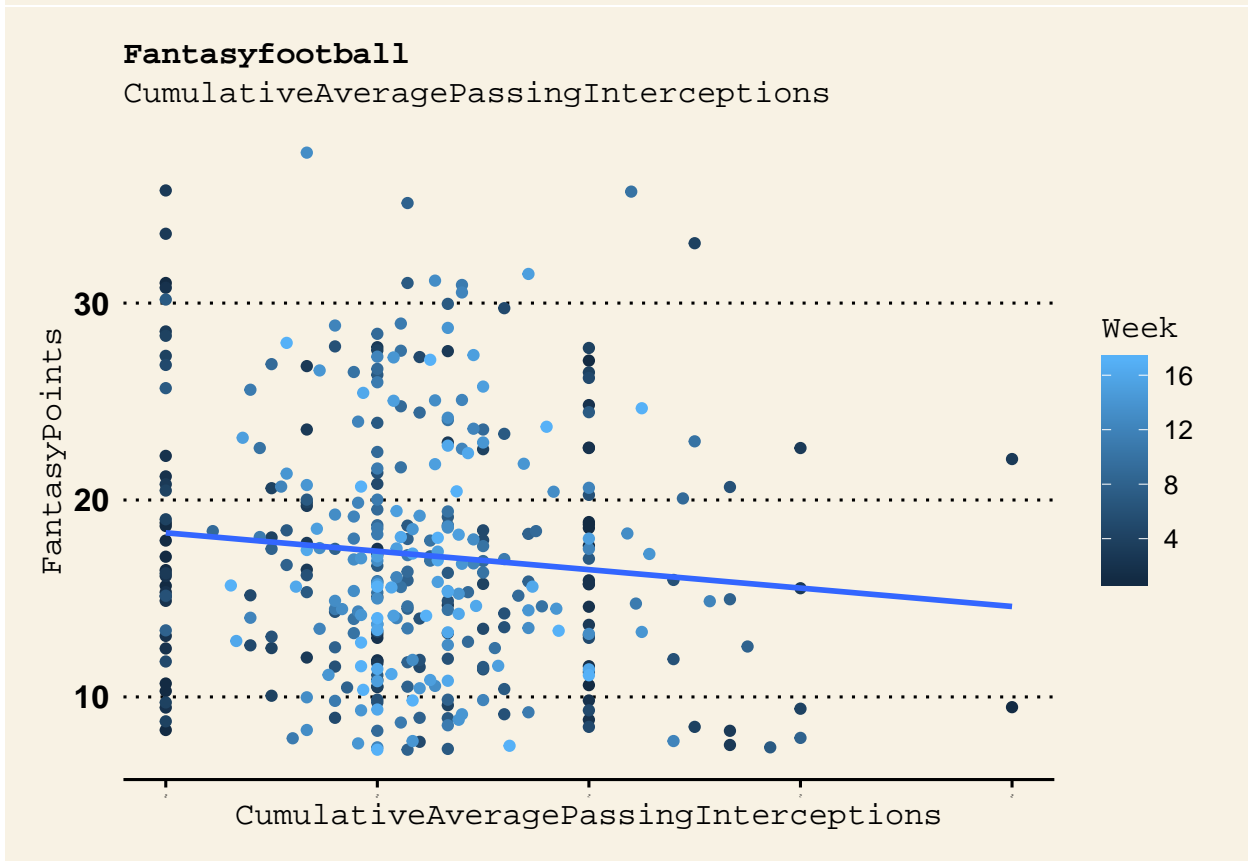
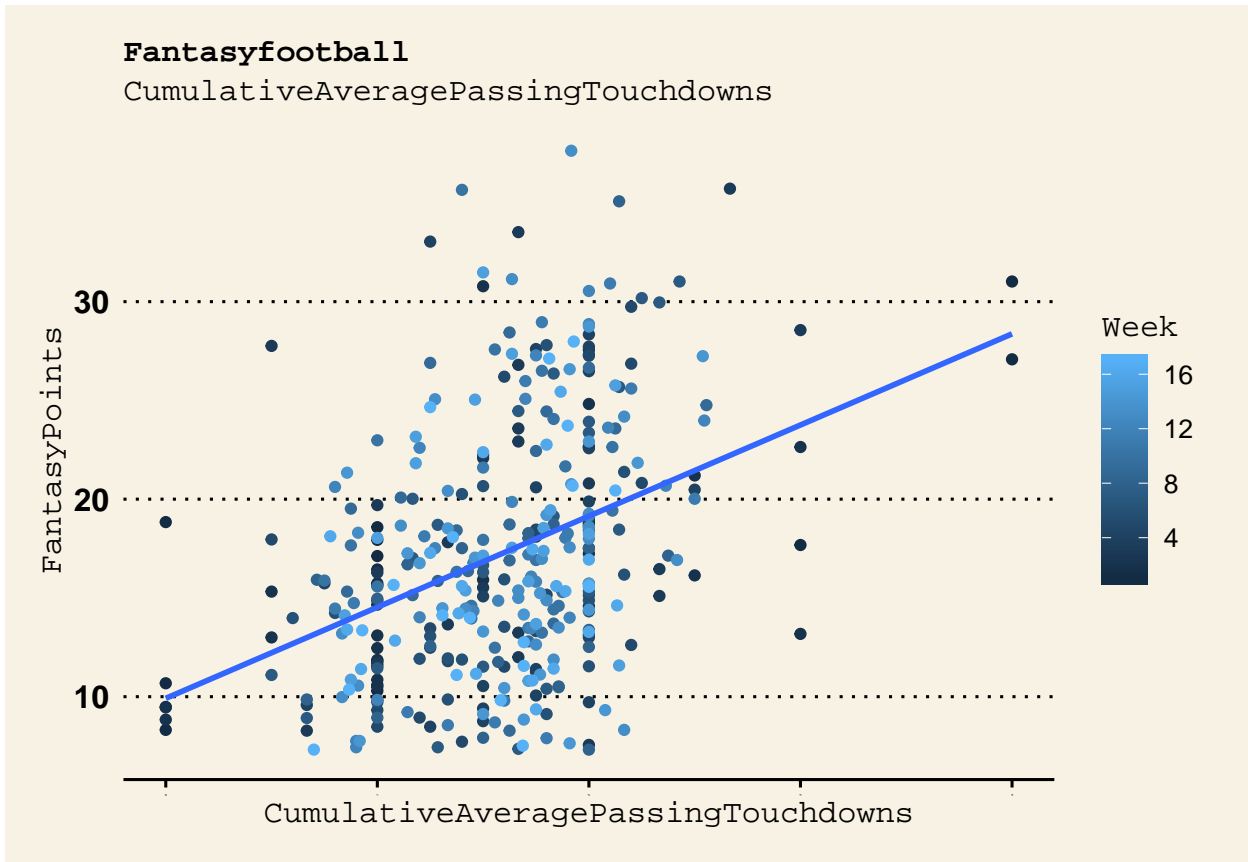


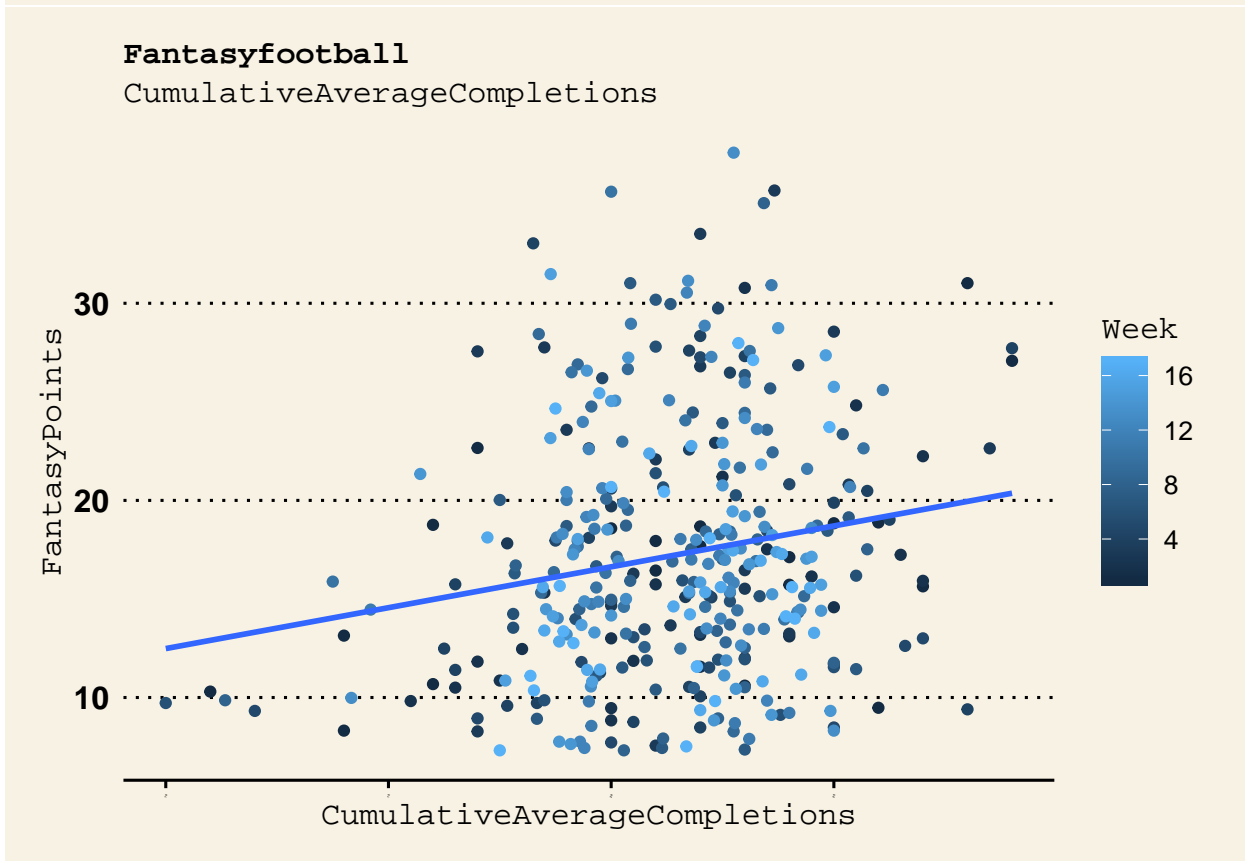
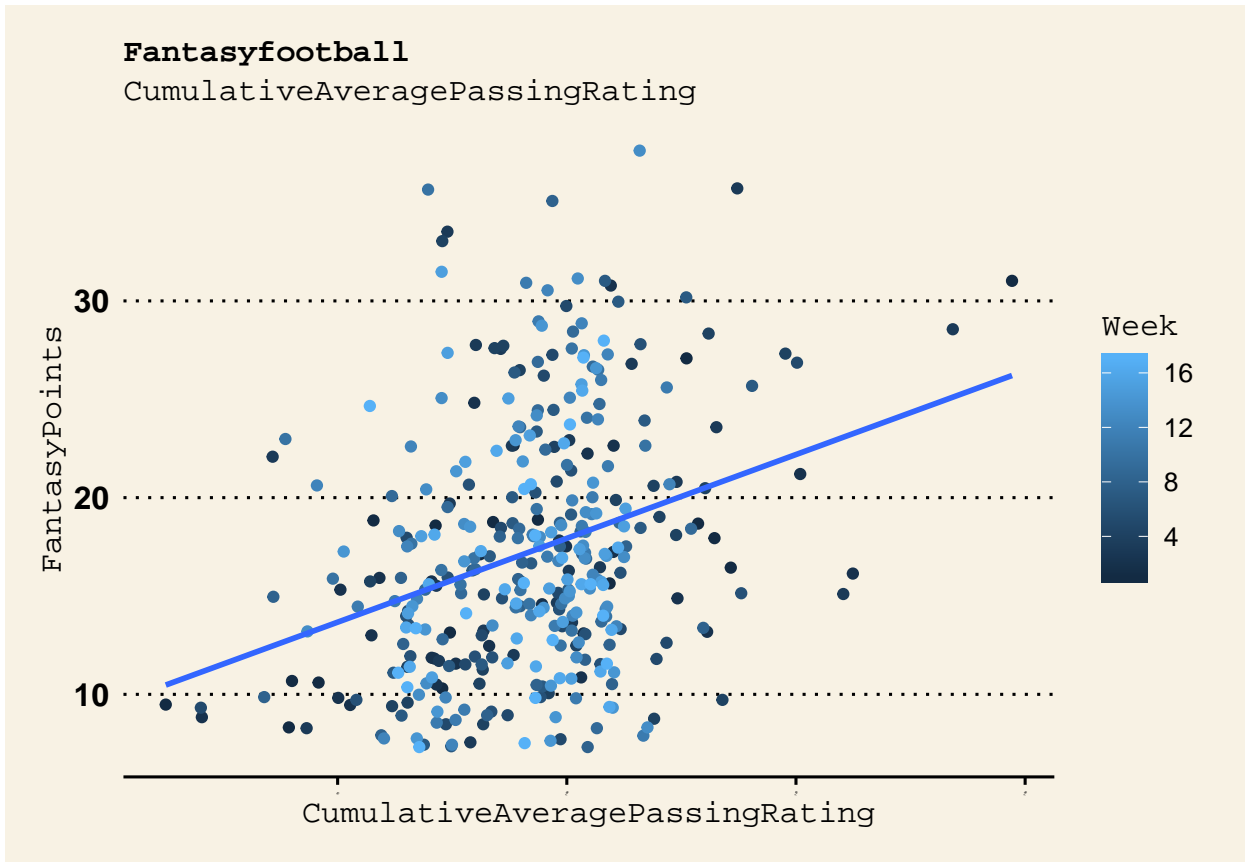


## 9.2 : Scatterplots for Derived features

```
for (i in 2:20) {
  derived_scatter <- eda_derived %>%
    ggplot(aes_string(y="FantasyPoints",x=names(eda_derived[i]),color="as.numeric(Week)"))+
    geom_point()+
    geom_smooth(method="lm",se=F)+
    xlab(names(eda_derived[i]))+
    ylab("FantasyPoints")+
    labs(title="Fantasyfootball",
         subtitle=names(eda_derived[i]),
         caption="Source: Fantasyfootball")+
    labs(color="Week")+
    theme_ws()+
    theme(plot.title = element_text(size = rel(0.5)),
          plot.subtitle = element_text(size = rel(0.5)),
          axis.text.x = element_text(angle=65, vjust=0.6,size=1),
          axis.title = element_text(size = rel(0.5)),
          legend.position = "right",
          legend.direction = "vertical",
          legend.title = element_text(size = rel(0.5))
    )
  print(derived_scatter)
}
```

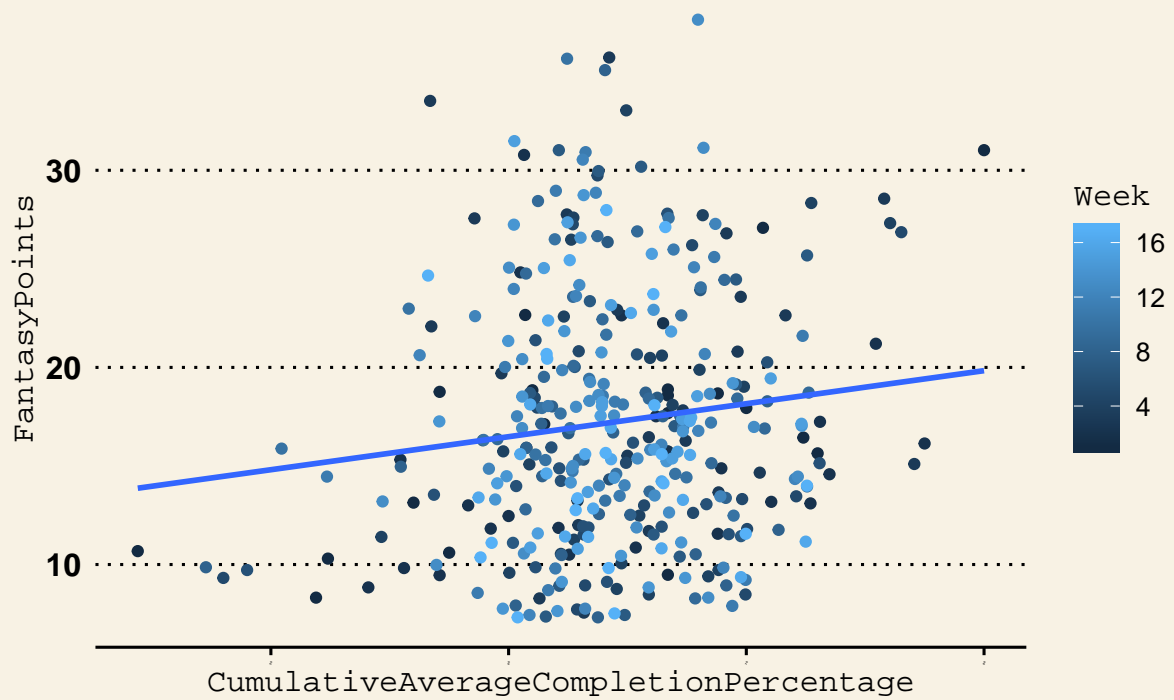






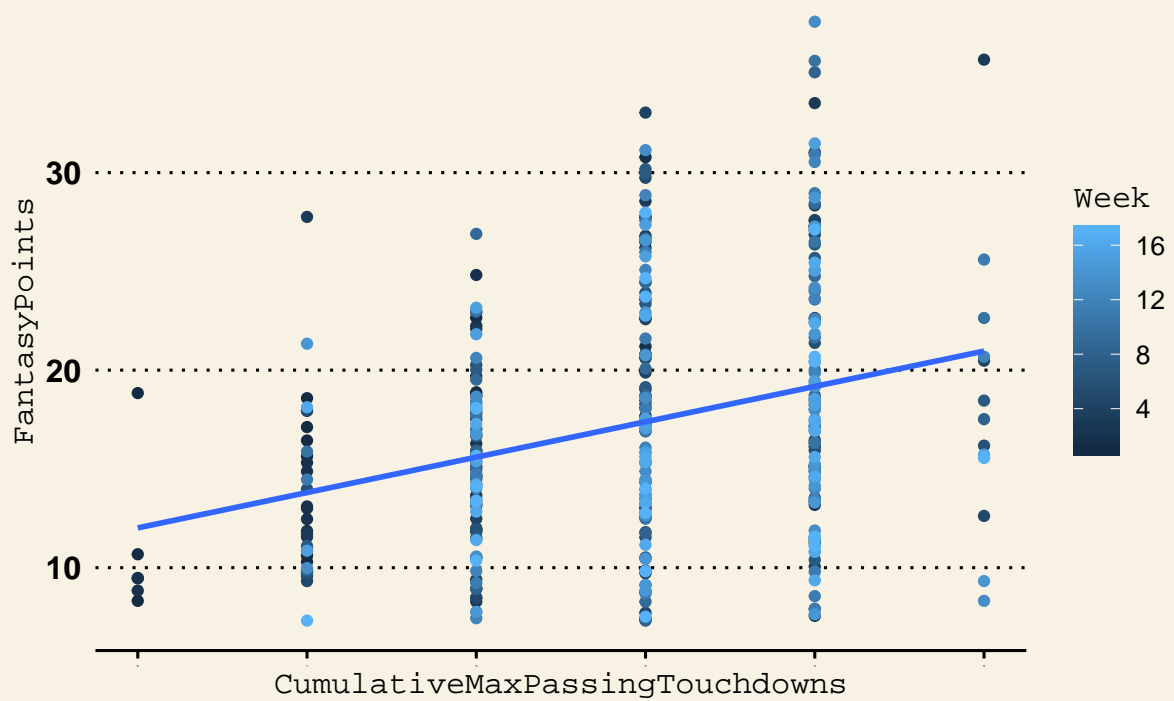
### Fantasyfootball

CumulativeAverageCompletionPercentage

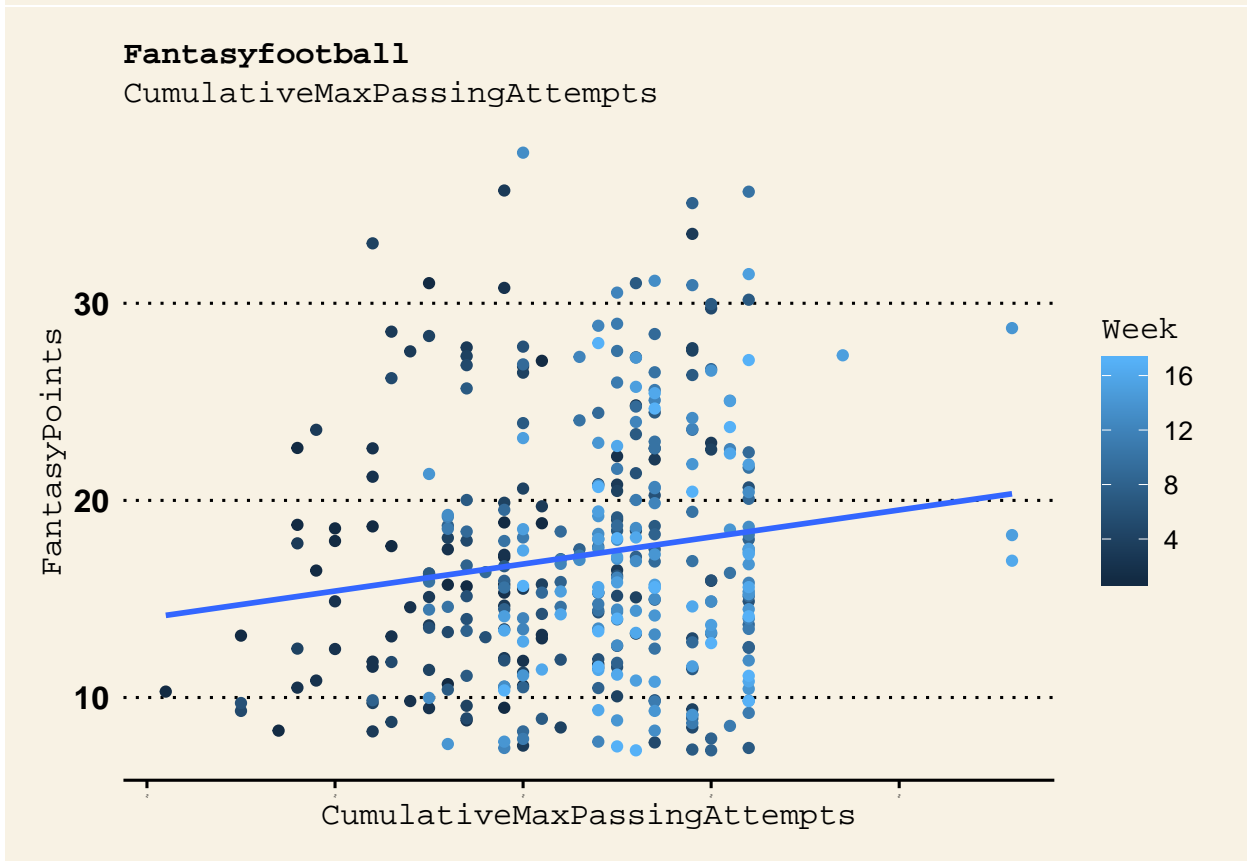
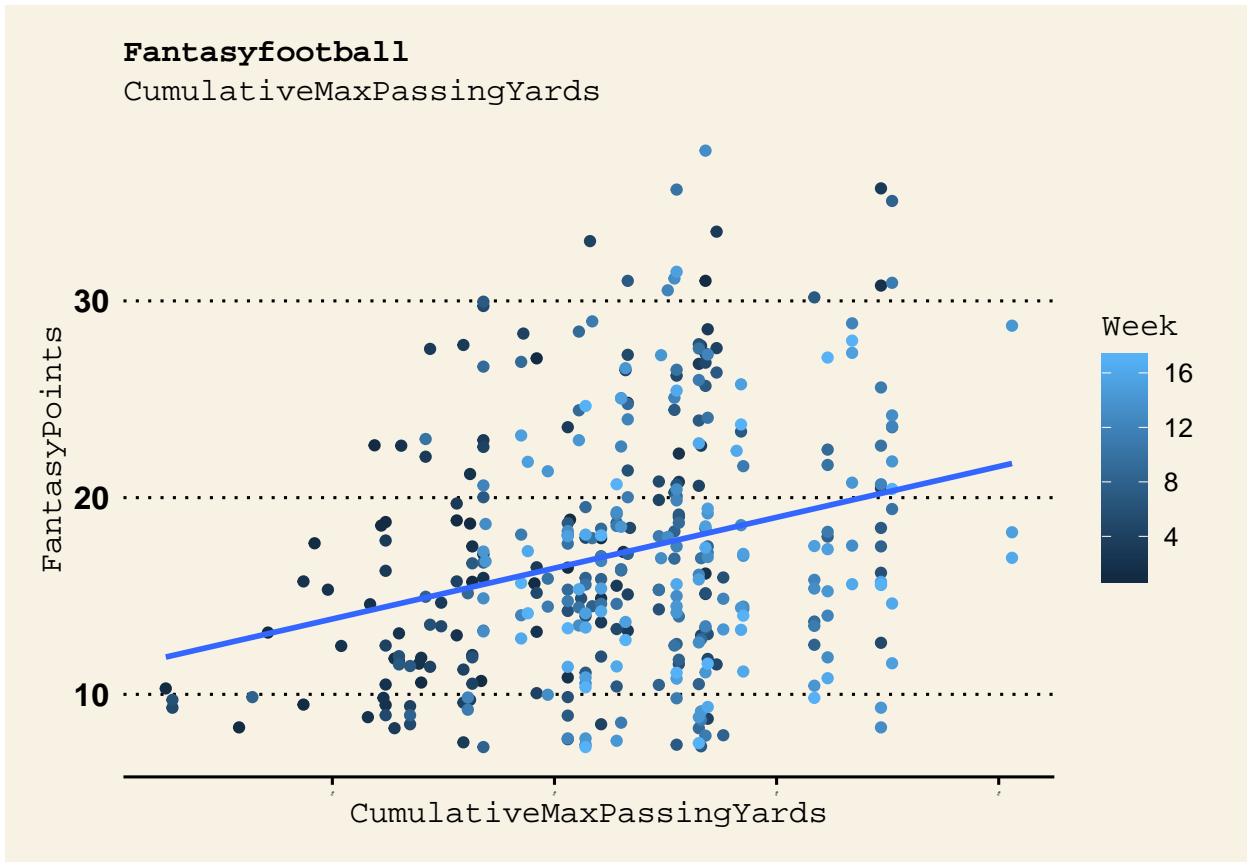


### Fantasyfootball

CumulativeMaxPassingTouchdowns

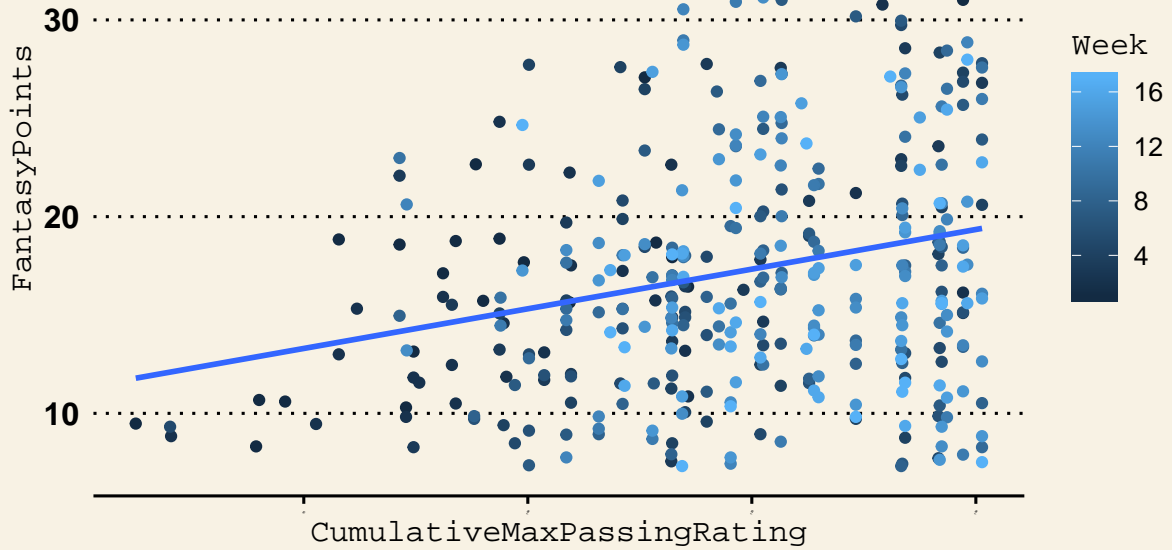






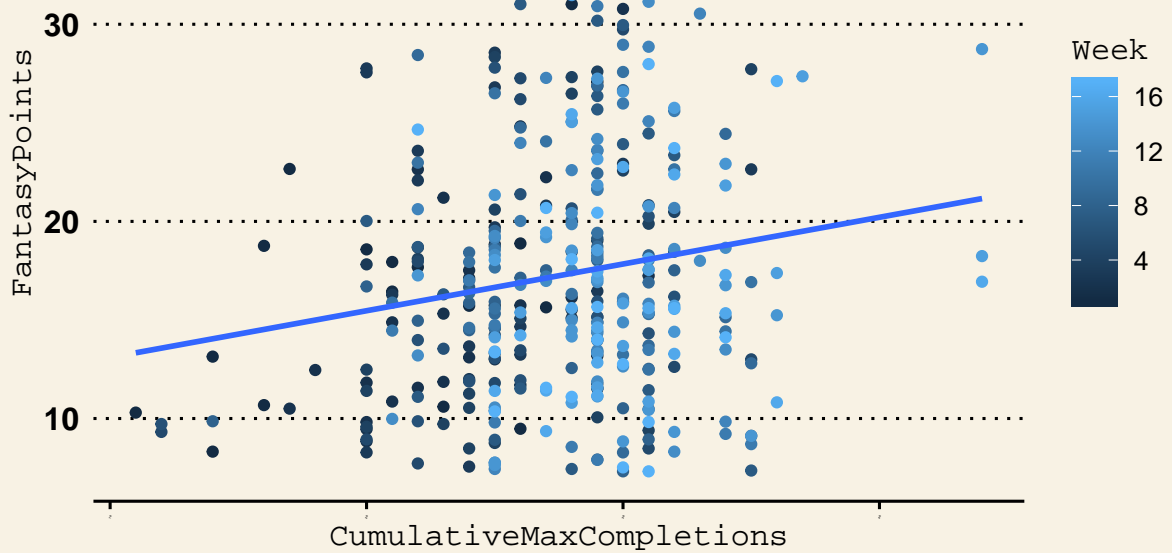
### Fantasyfootball

CumulativeMaxPassingRating



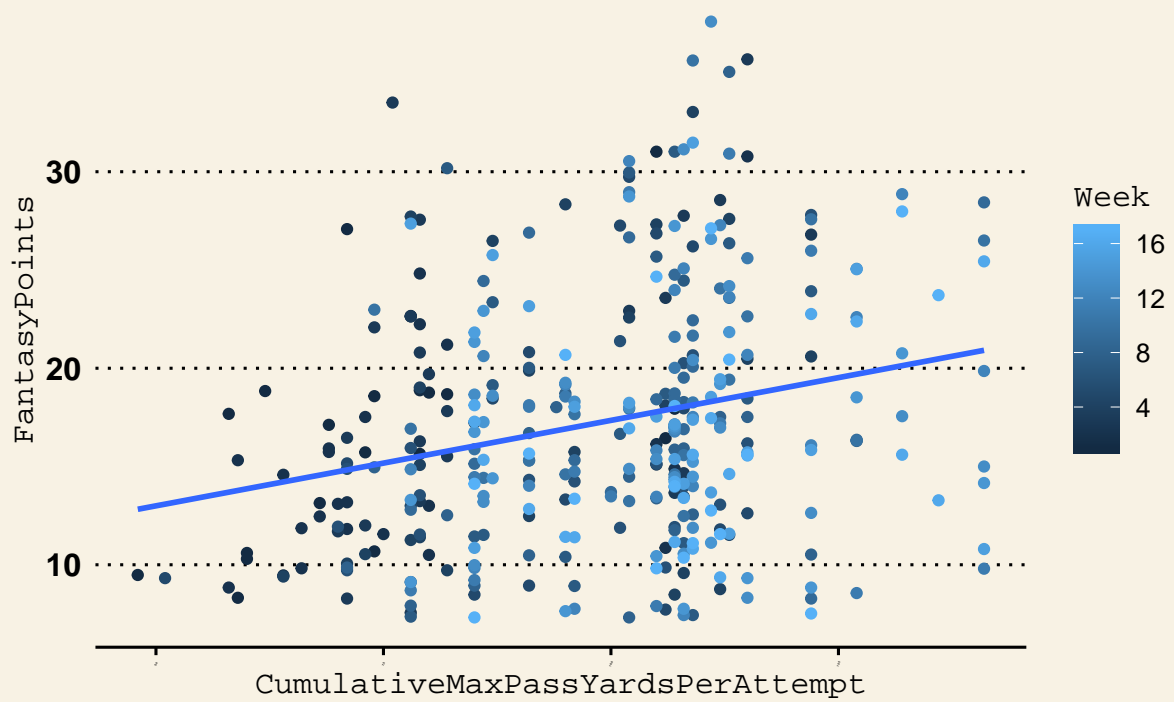
### Fantasyfootball

CumulativeMaxCompletions



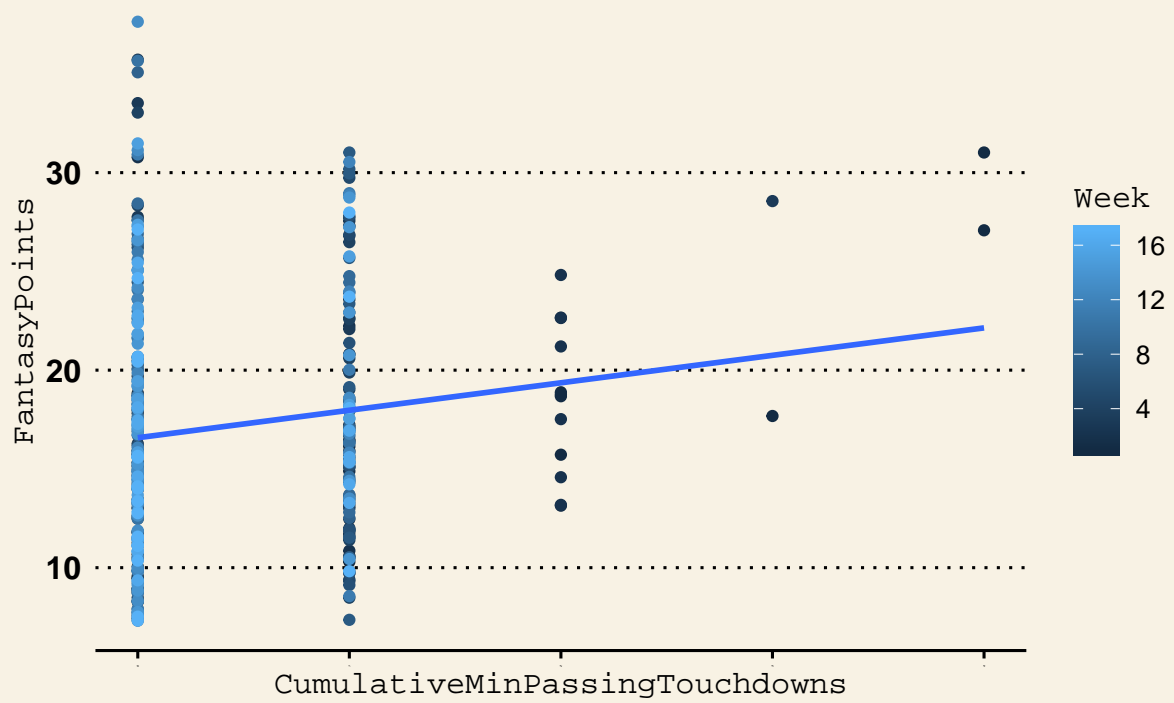
### Fantasyfootball

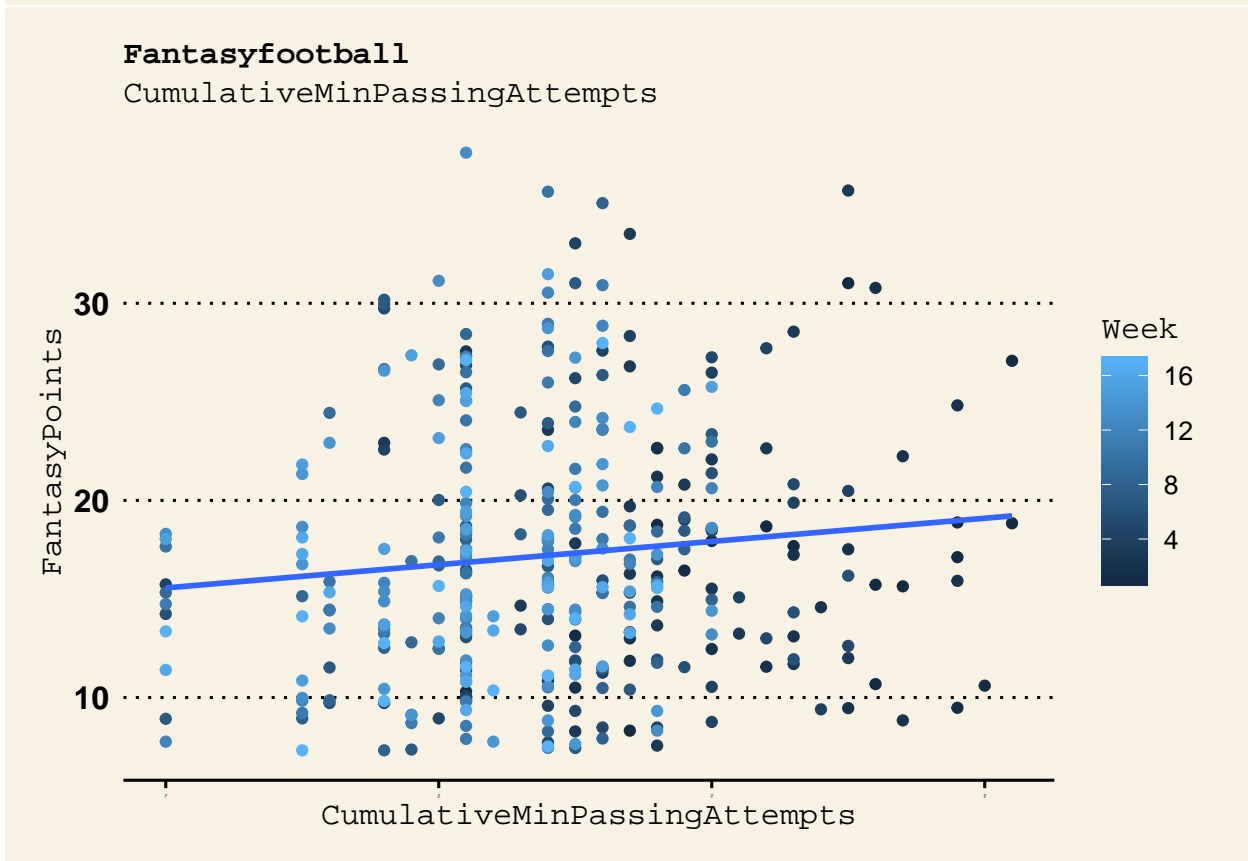
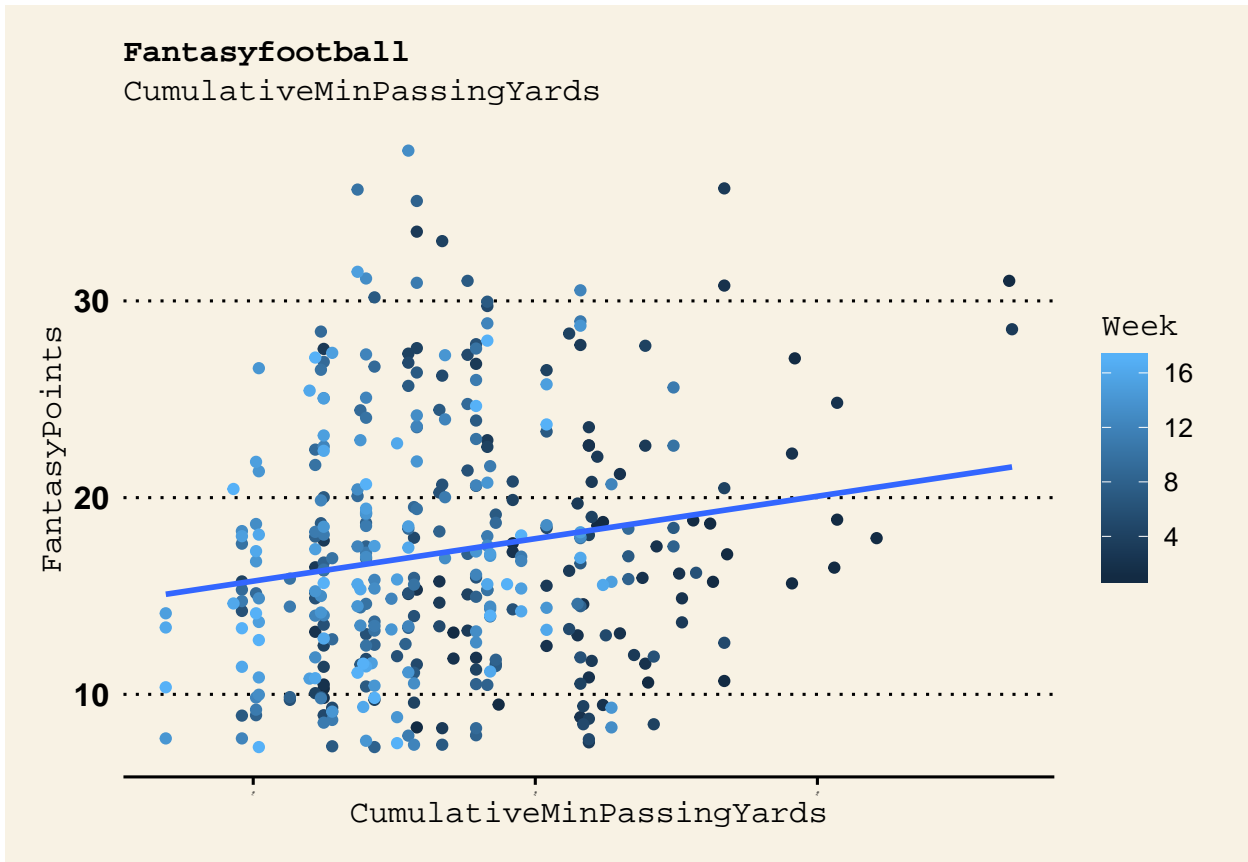
CumulativeMaxPassYardsPerAttempt

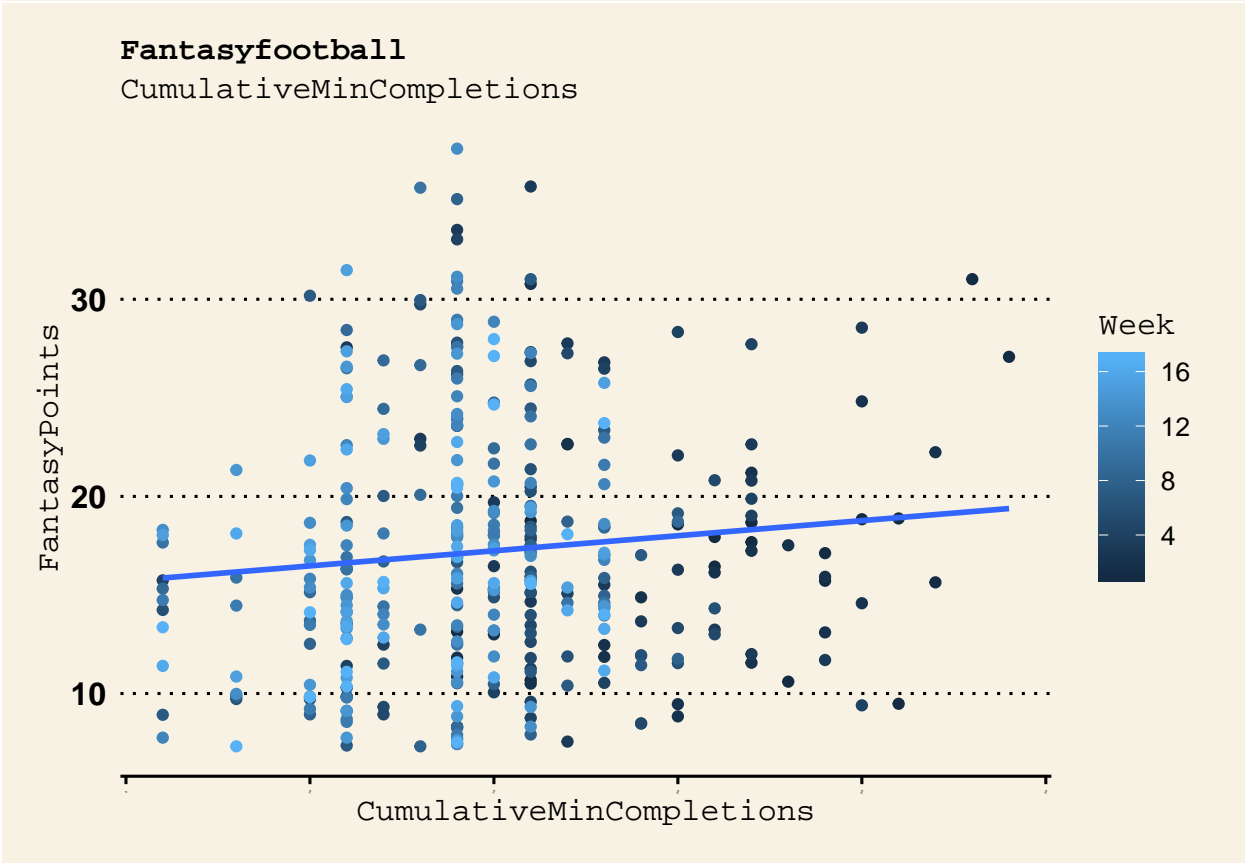
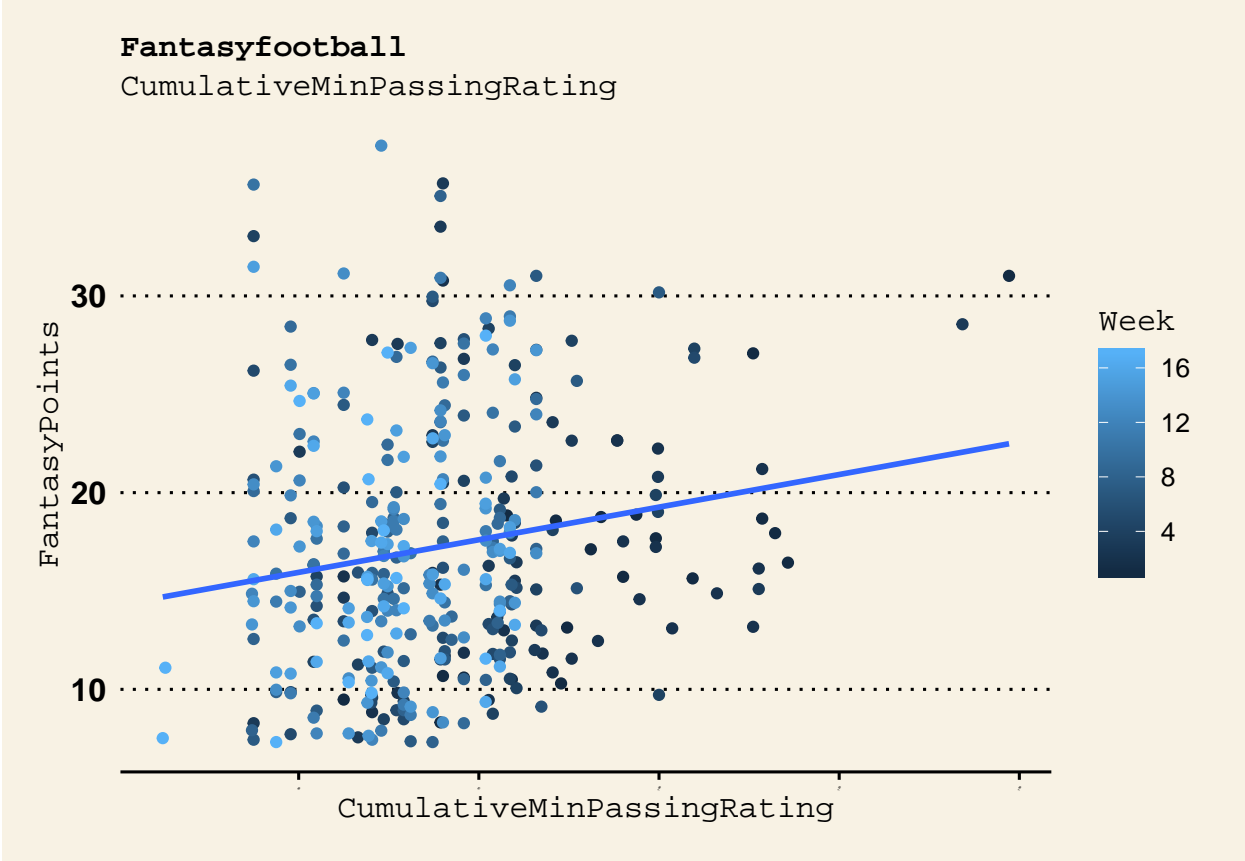


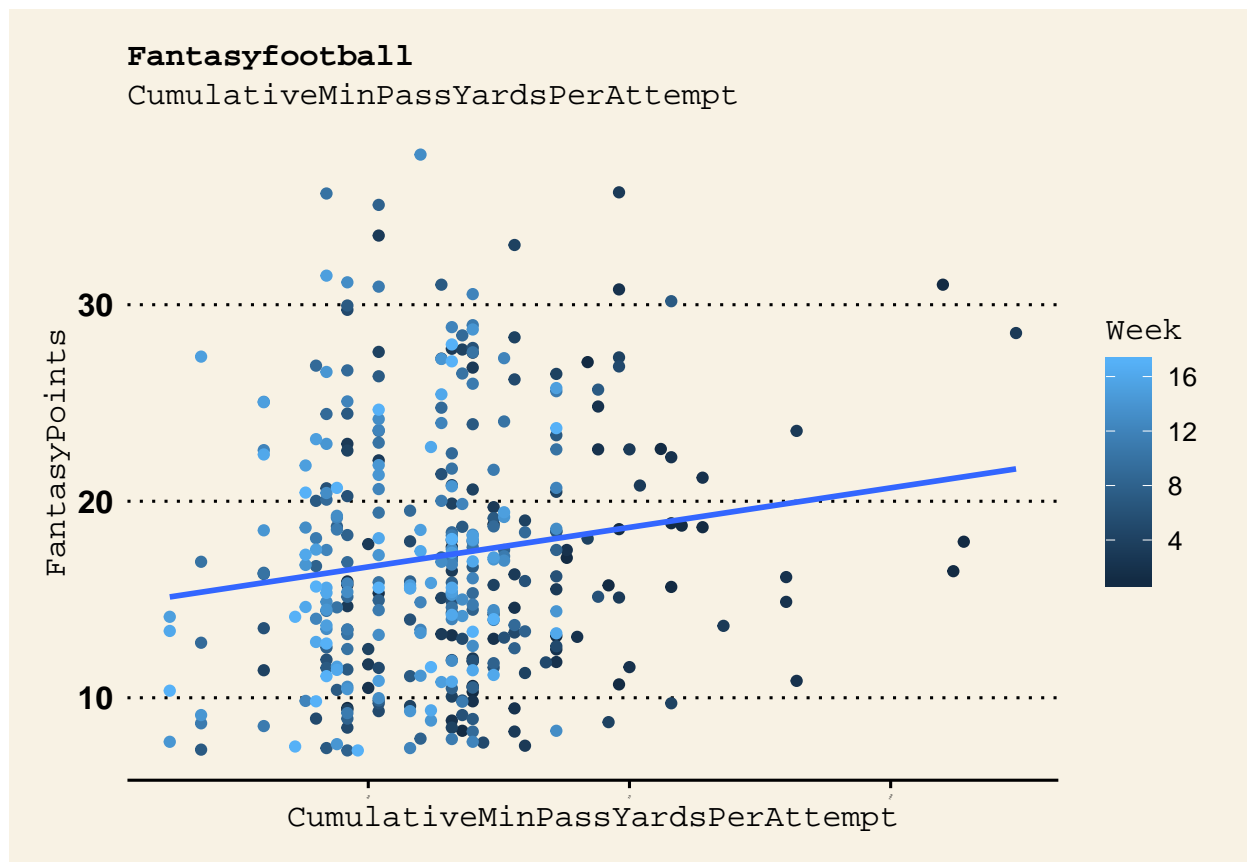
### Fantasyfootball

CumulativeMinPassingTouchdowns









## 10 : Line plots

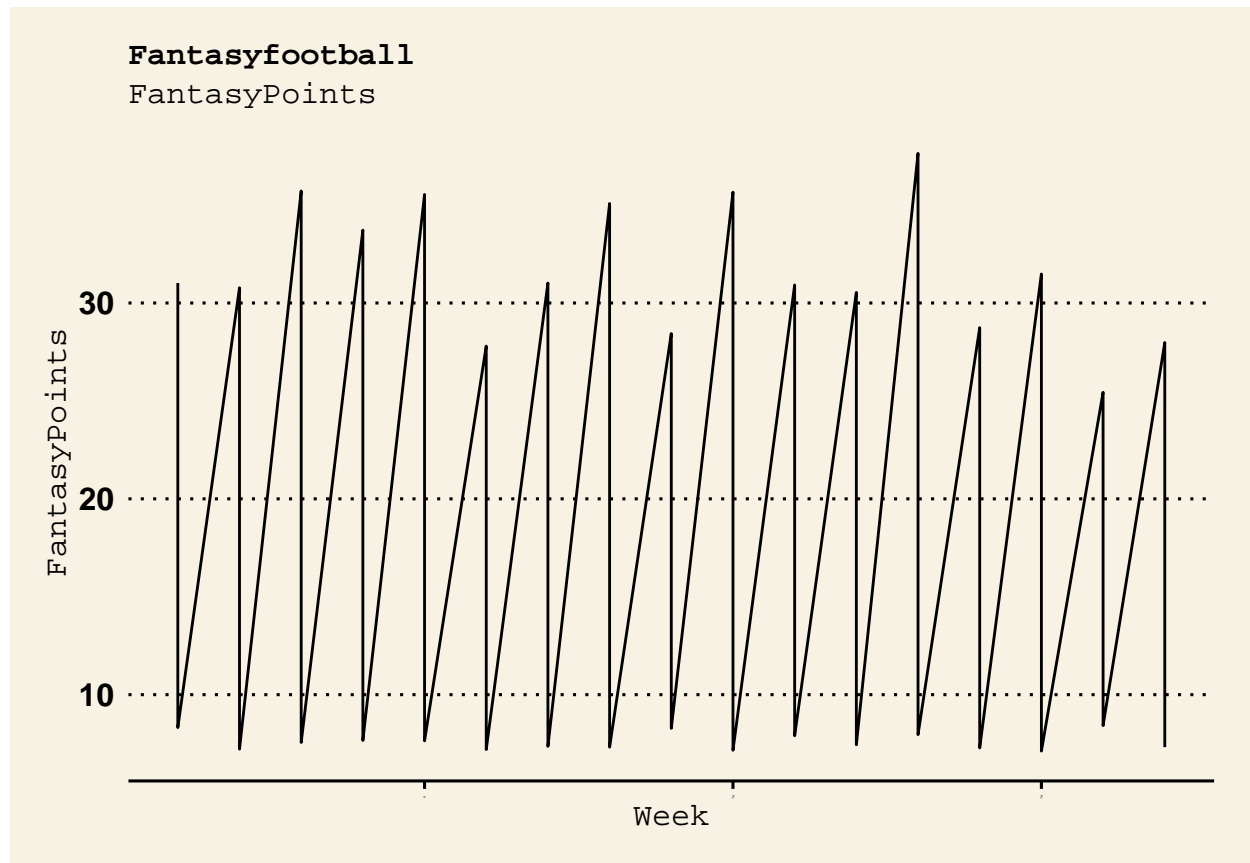
## 10.1 Base Features

```
line_ds <- eda_base %>% group_by(Week) %>% arrange(Week)

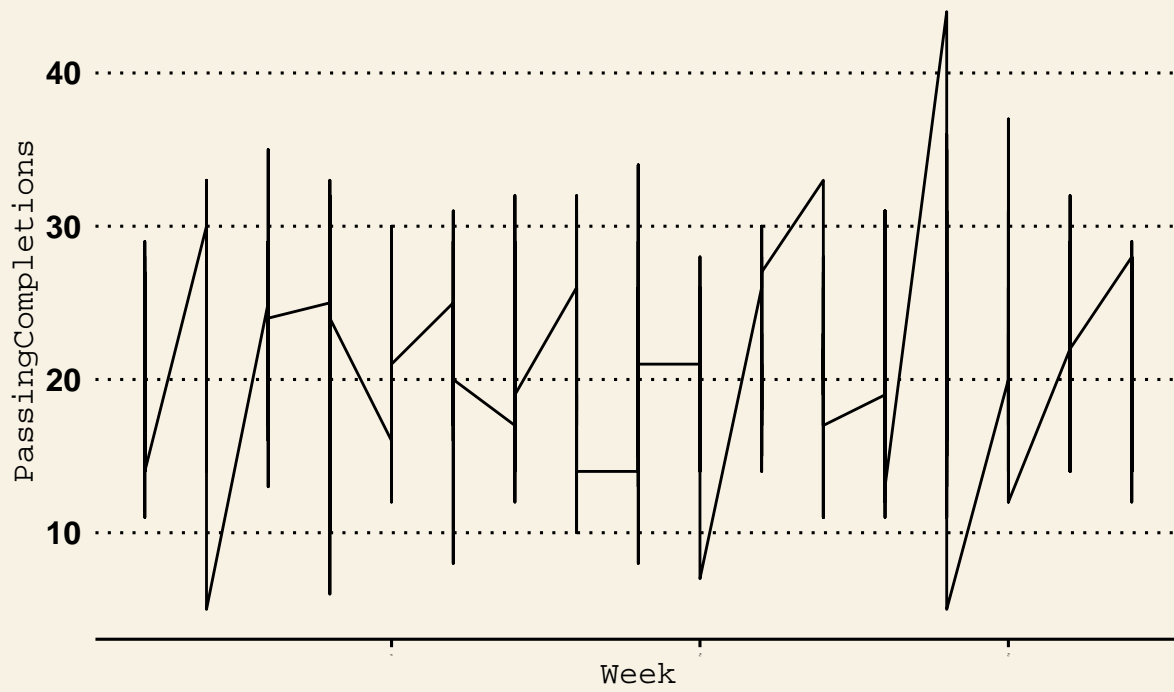
for(p in 2:15){

  line_plot <- line_ds %>%
    ggplot(aes_string(x="as.numeric(Week)",y=names(line_ds[p])))+
    ggtitle(names(line_ds[p]))+
    geom_line(show.legend = FALSE)+
    xlab("Week")+
    ylab(names(line_ds[p]))+
    labs(title="Fantasyfootball",
         subtitle=names(line_ds[p]),
         caption="Source: Fantasyfootball")+
    theme_ws() +
    theme(plot.title = element_text(size = rel(0.5)),
          plot.subtitle = element_text(size = rel(0.5)),
          axis.text.x = element_text(angle=65, vjust=0.6,size=1),
          axis.title = element_text(size = rel(0.5)),
          legend.position = "right",
          legend.direction = "vertical",
          legend.title = element_text(size = rel(0.5)))
```

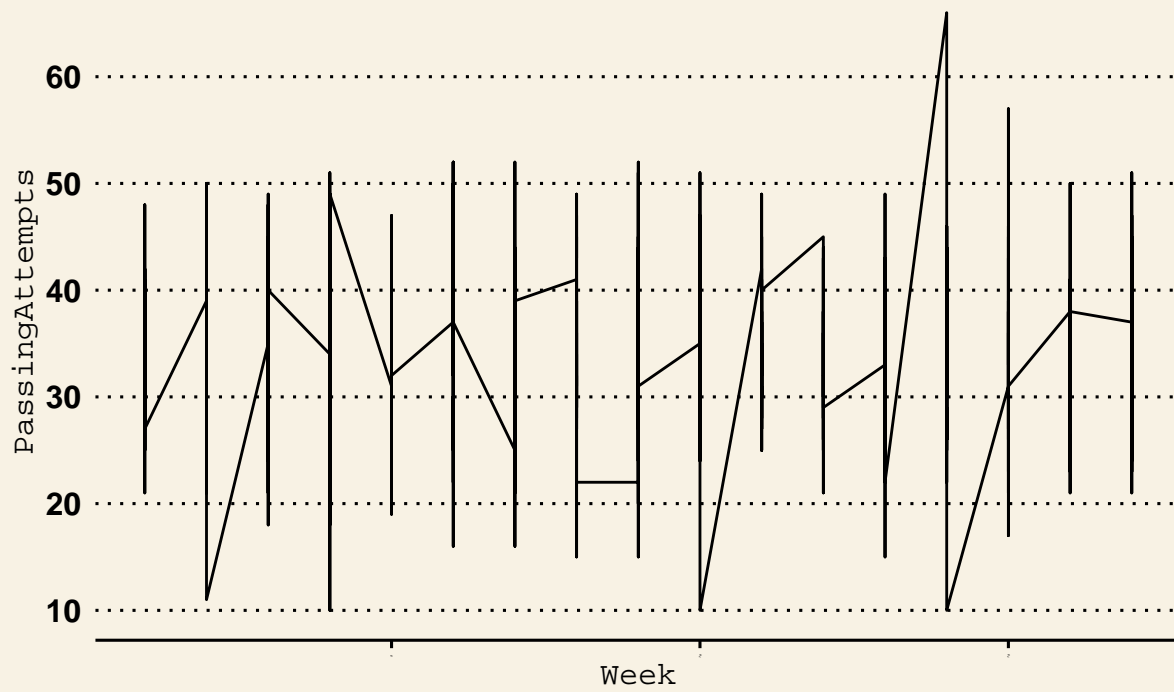
```
print(line_plot)
}
```



**Fantasyfootball**  
PassingCompletions



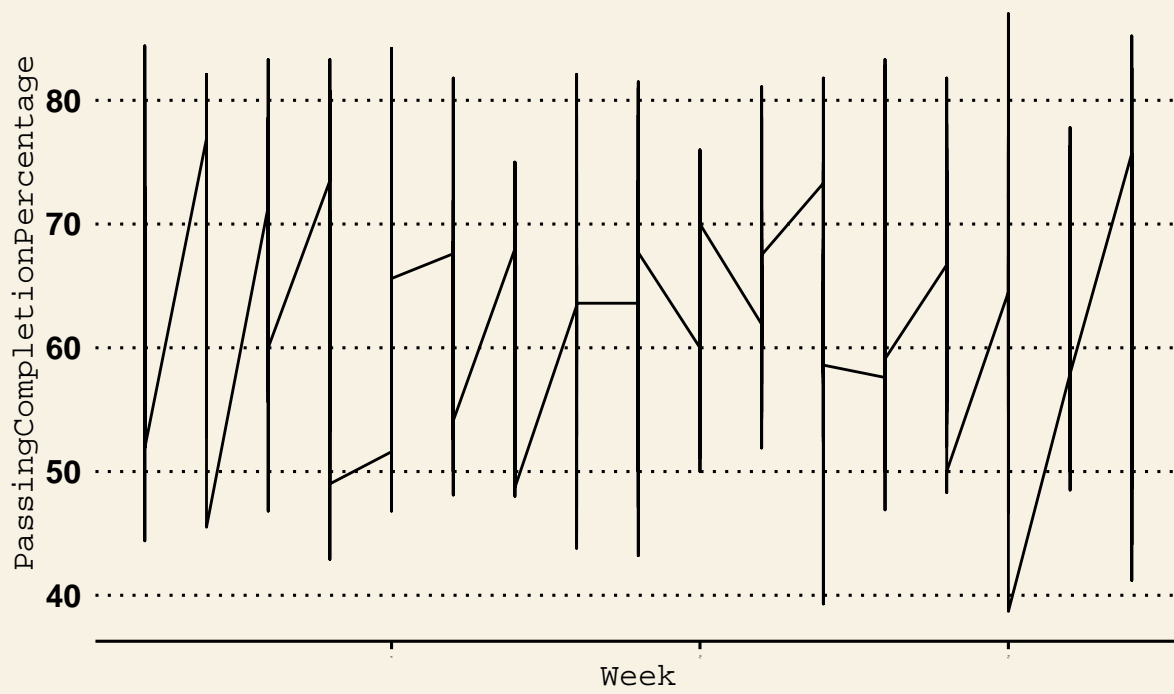
**Fantasyfootball**  
PassingAttempts





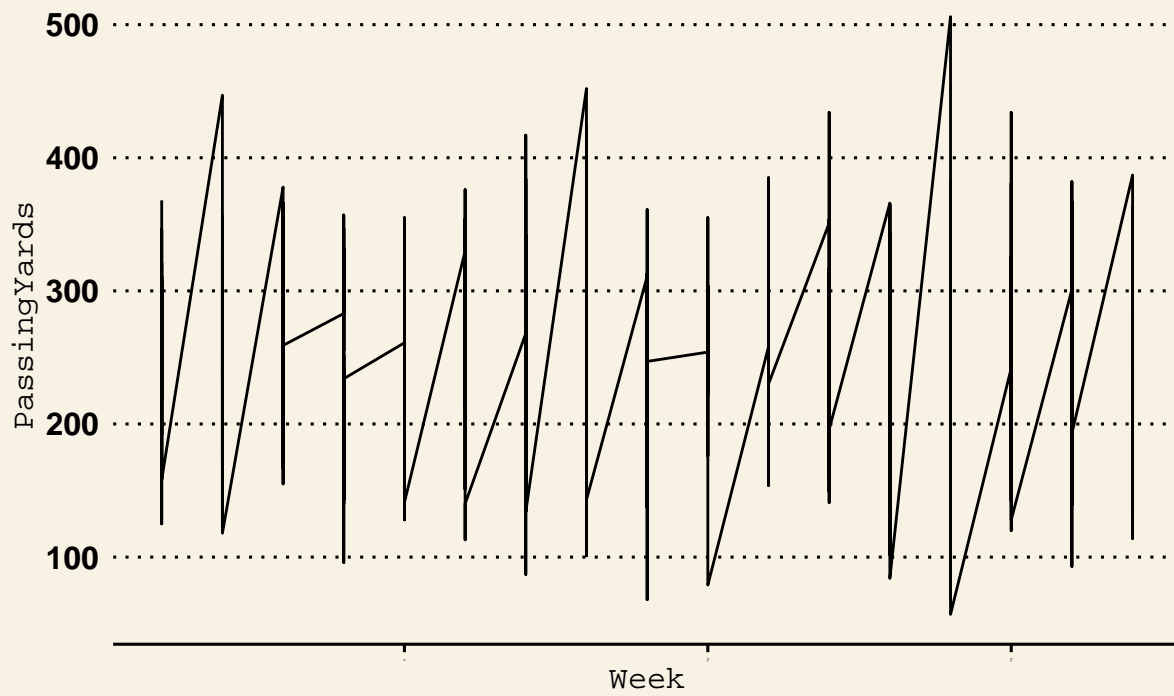
## Fantasyfootball

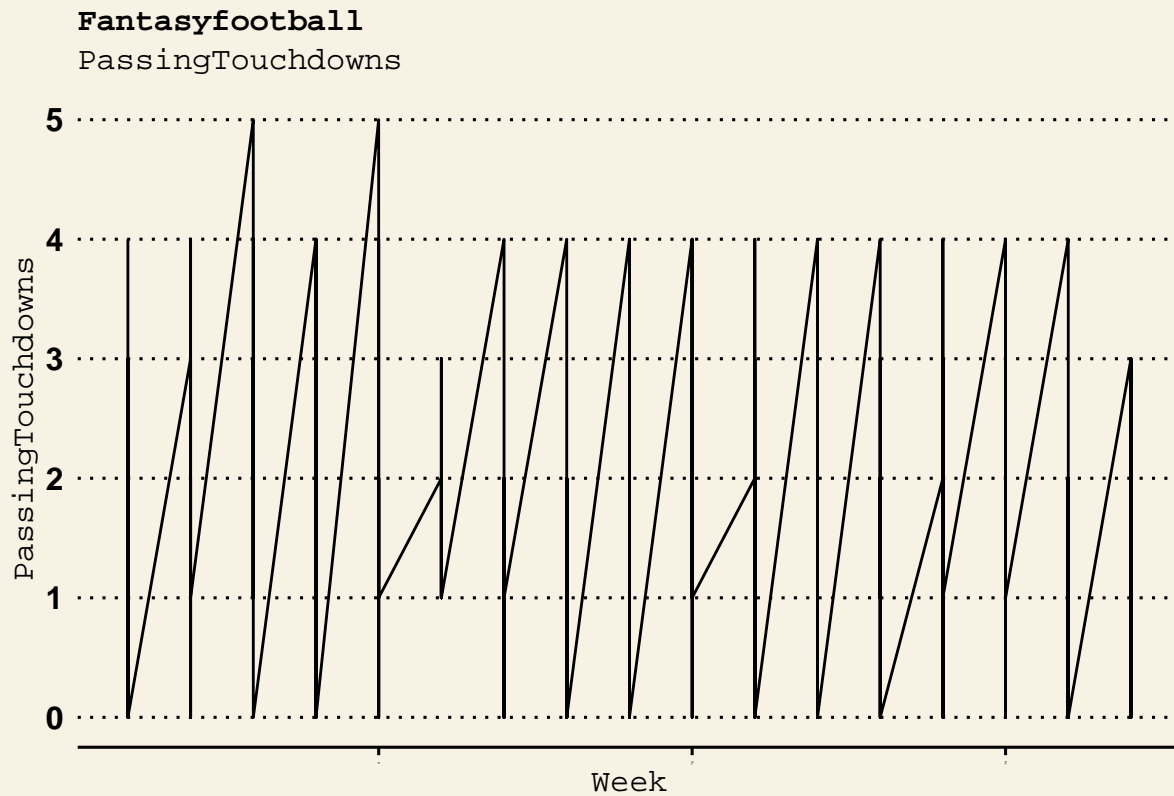
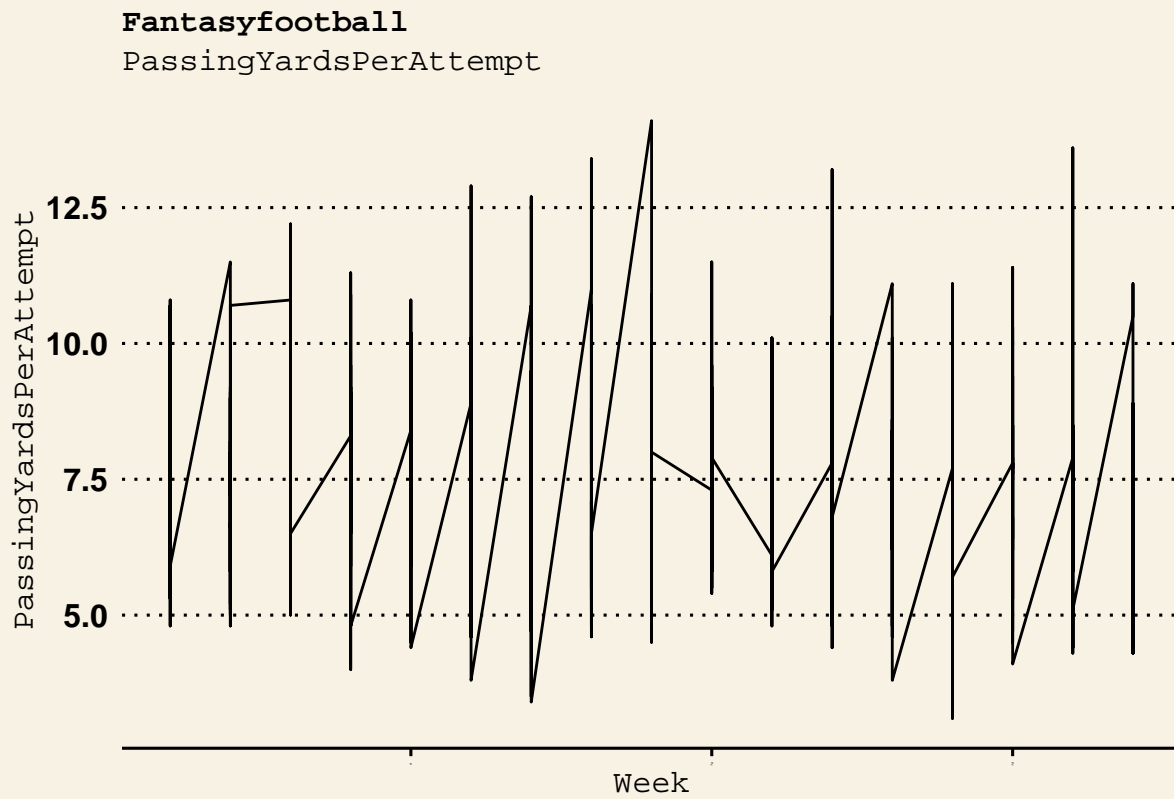
PassingCompletionPercentage



## Fantasyfootball

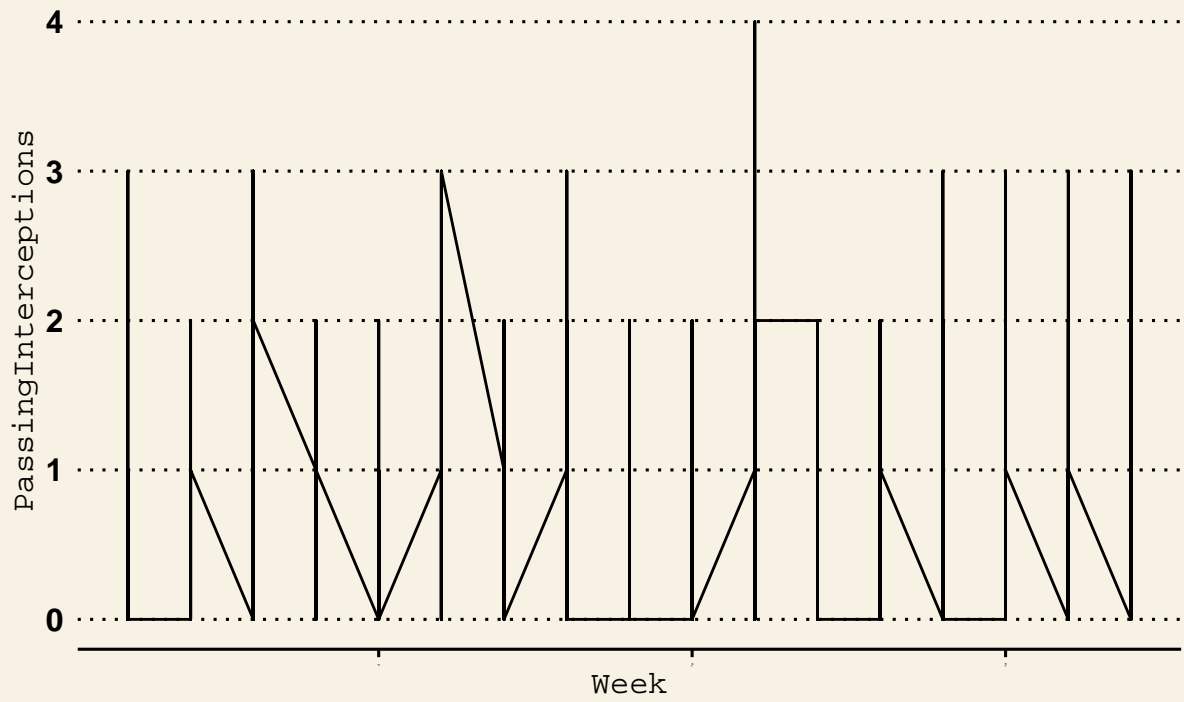
PassingYards





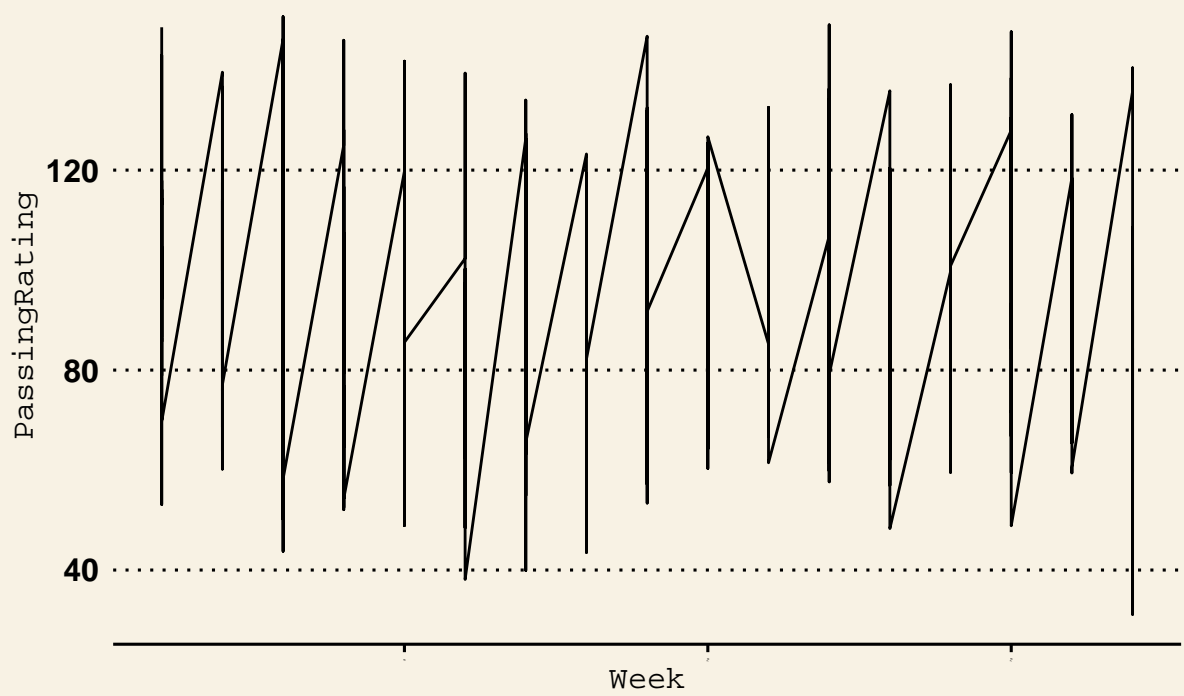
### Fantasyfootball

PassingInterceptions

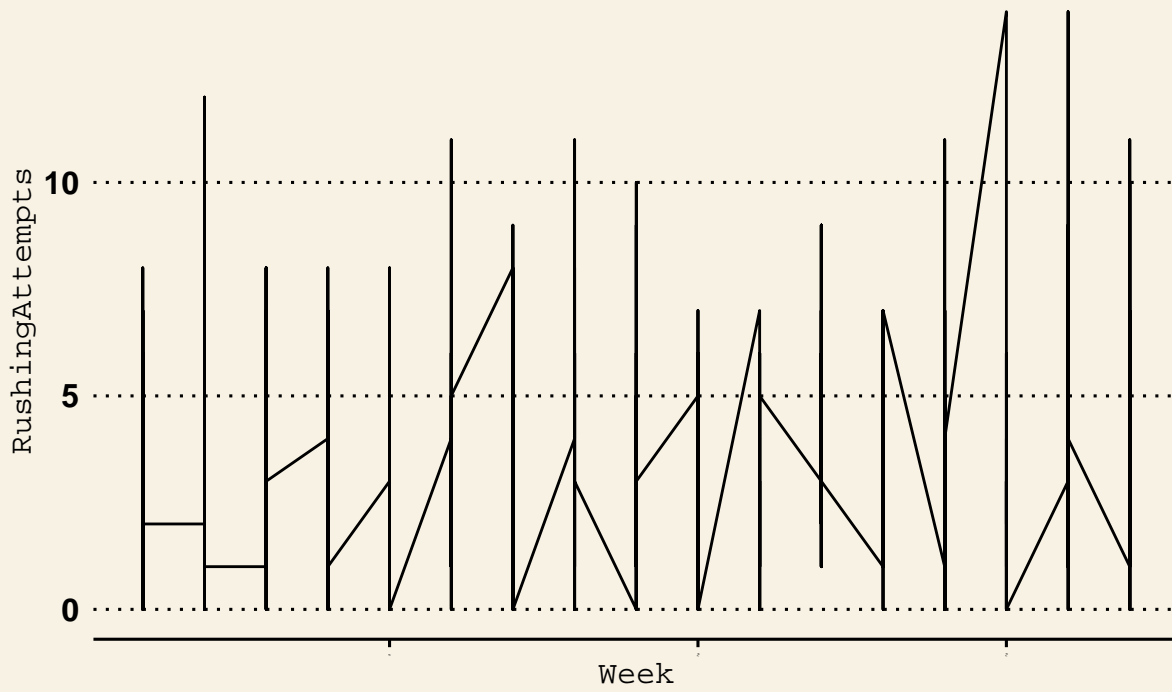


### Fantasyfootball

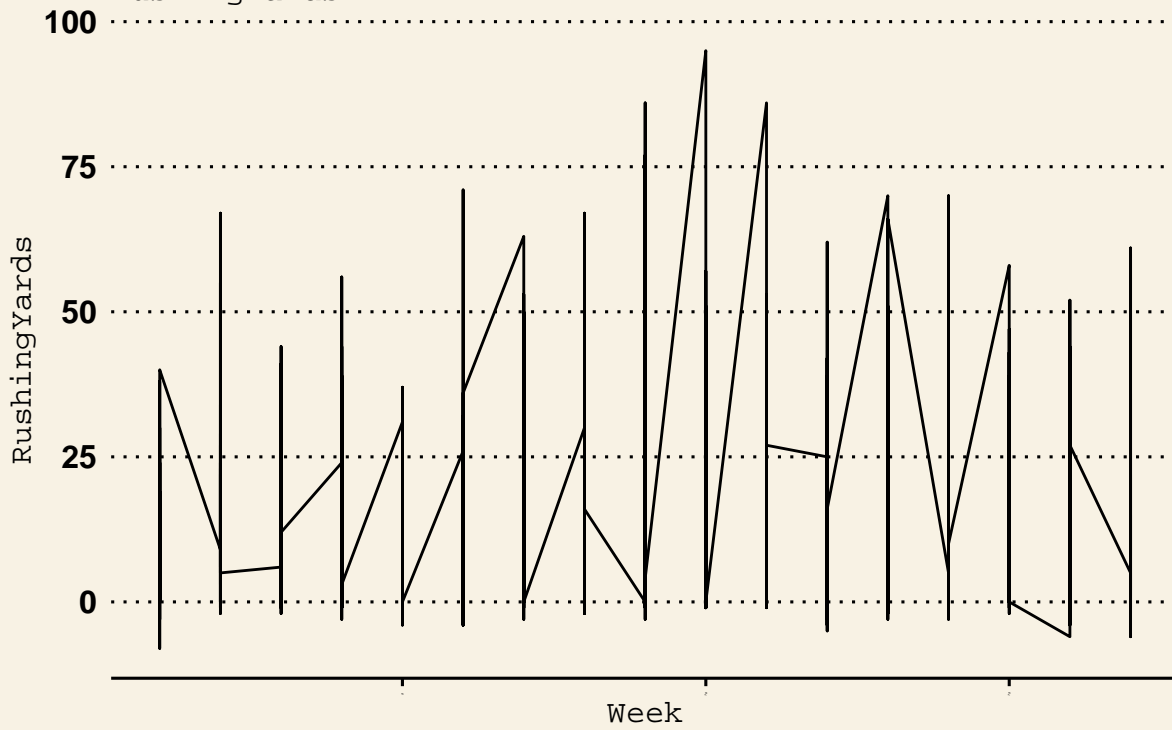
PassingRating



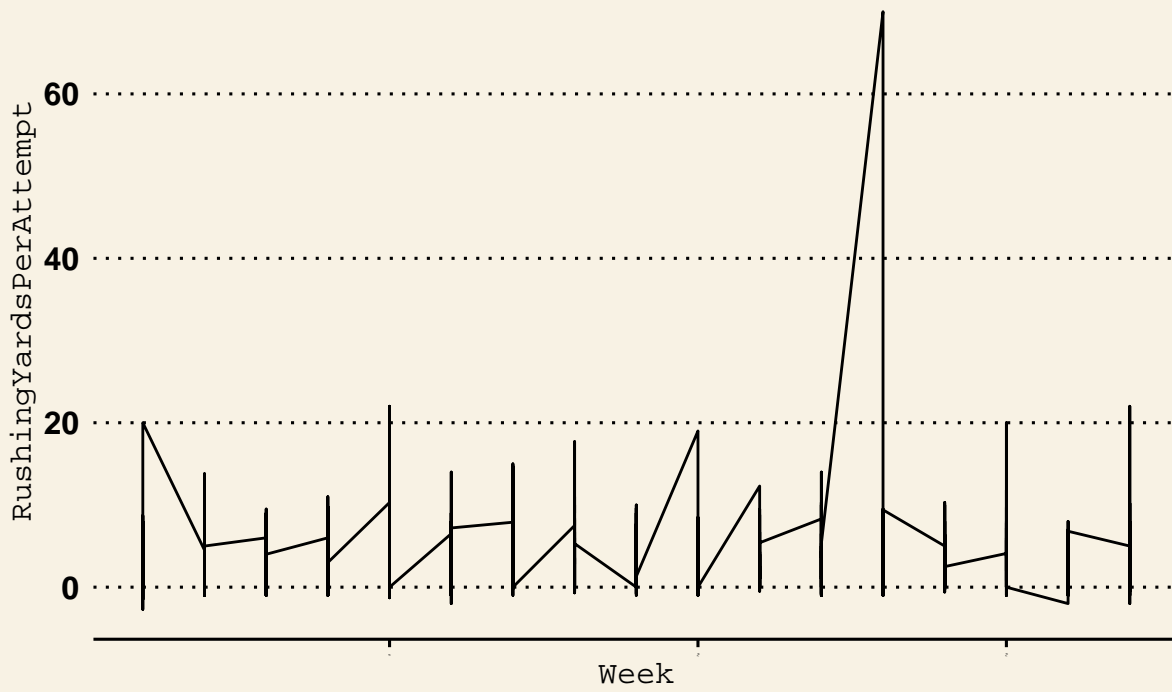
**Fantasyfootball**  
RushingAttempts



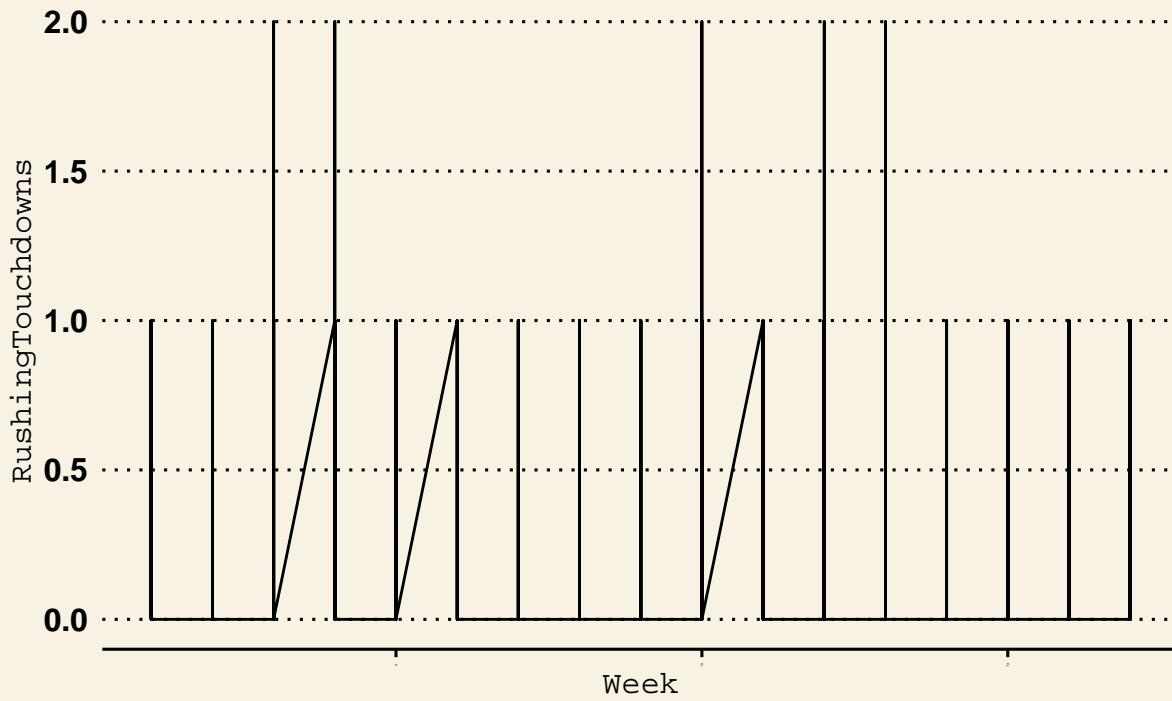
**Fantasyfootball**  
RushingYards

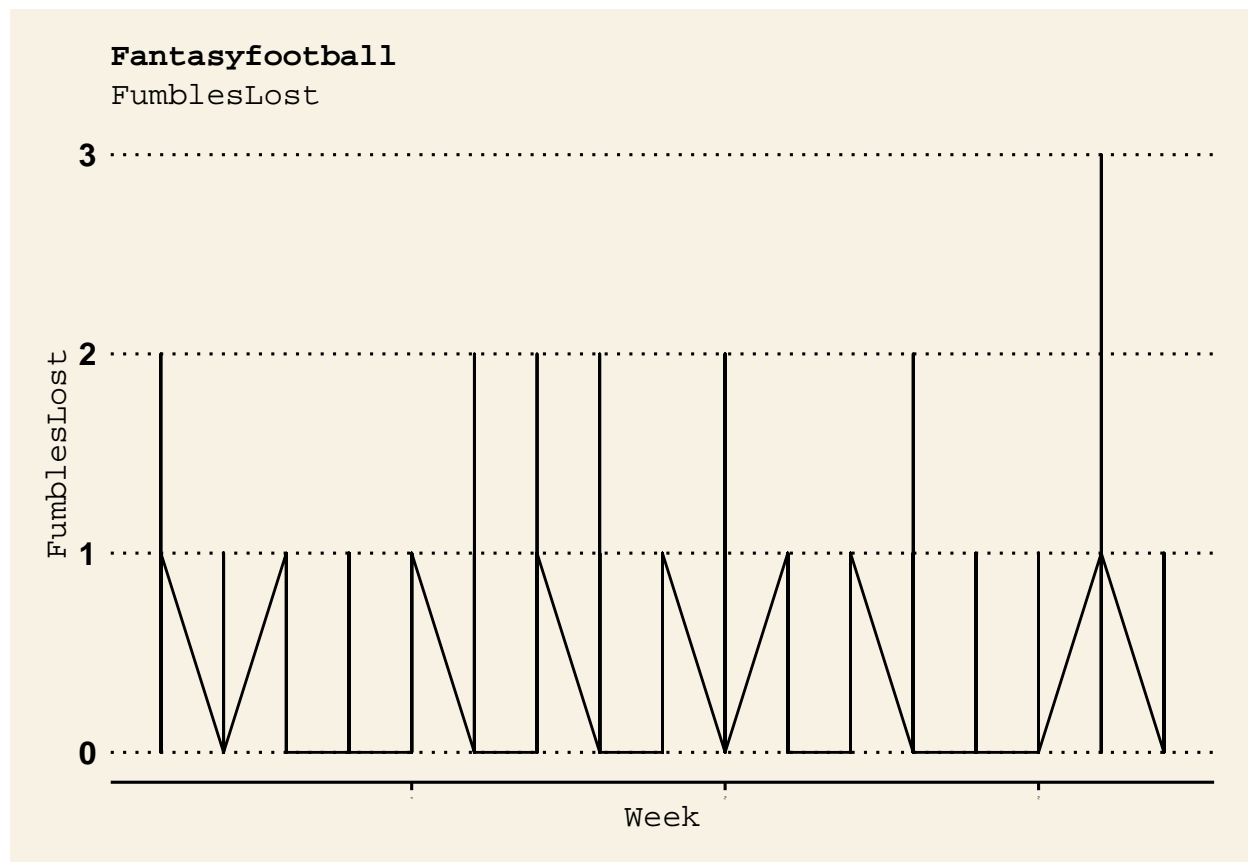


**Fantasyfootball**  
RushingYardsPerAttempt



**Fantasyfootball**  
RushingTouchdowns





## 10.2 Derived Features

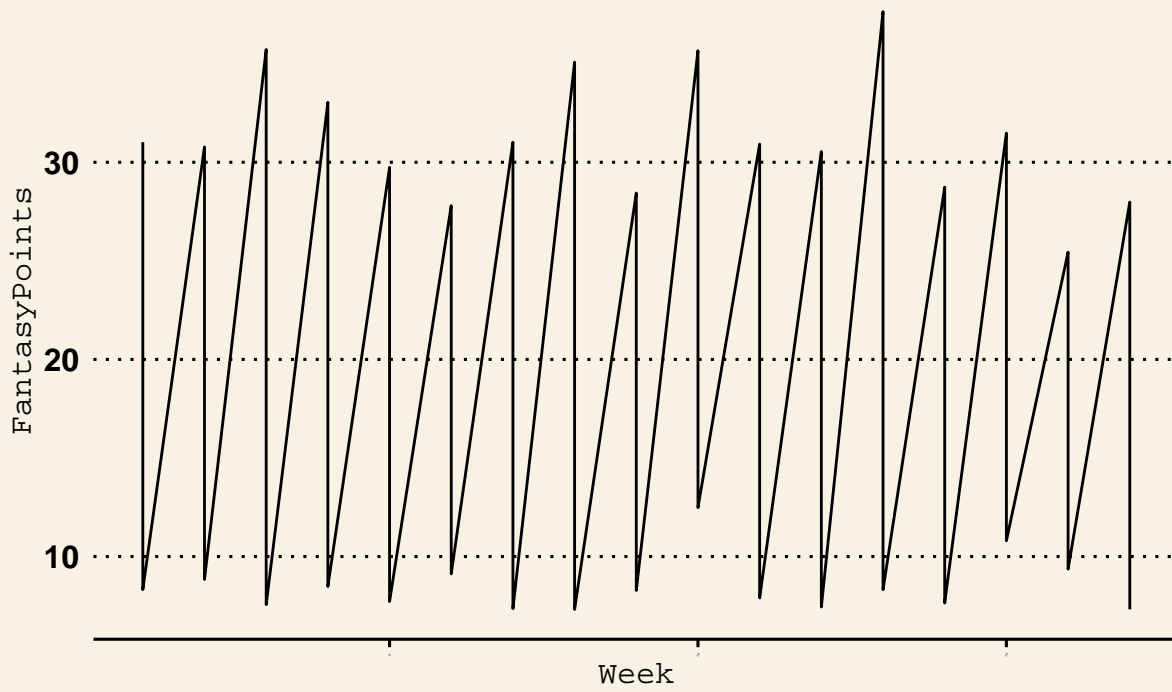
```
line_ds <- eda_derived %>% group_by(Week) %>% arrange(Week)

for(p in 2:20){

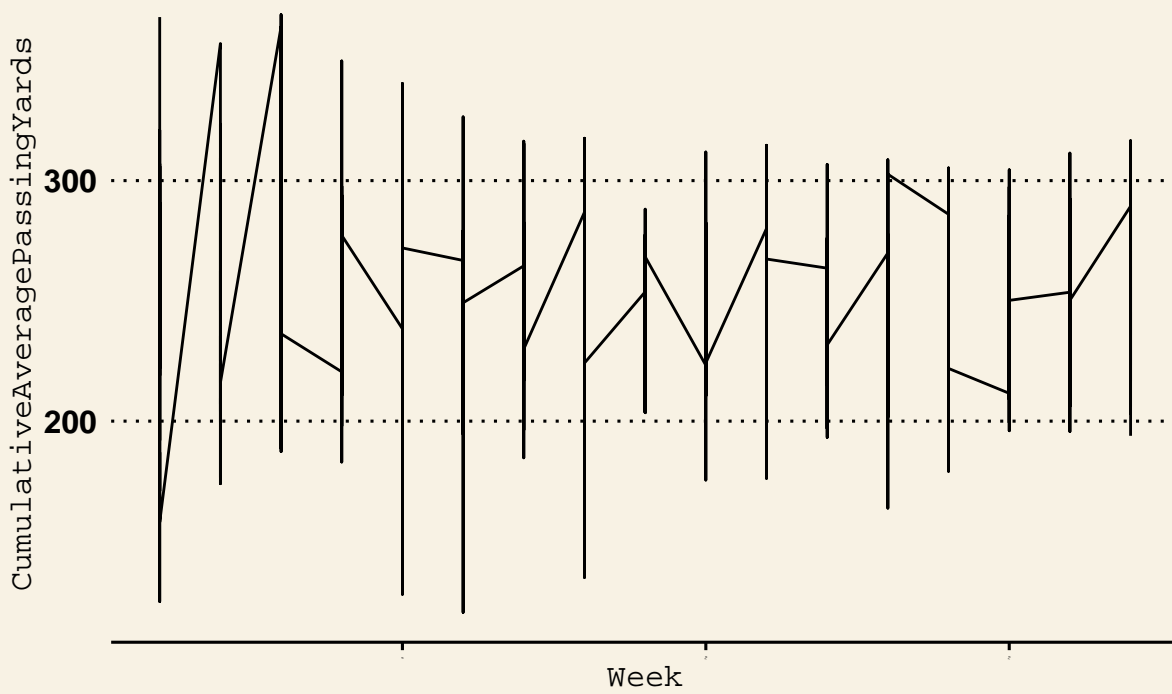
  line_plot <- line_ds %>%
    ggplot(aes_string(x="as.numeric(Week)",y=names(line_ds[p])))+
    ggtitle(names(line_ds[p]))+
    geom_line(show.legend = FALSE)+
    xlab("Week")+
    ylab(names(line_ds[p]))+
    labs(title="Fantasyfootball",
         subtitle=names(line_ds[p]),
         caption="Source: Fantasyfootball")+
    theme_ws()+
    theme(plot.title = element_text(size = rel(0.5)),
          plot.subtitle = element_text(size = rel(0.5)),
          axis.text.x = element_text(angle=65, vjust=0.6,size=1),
          axis.title = element_text(size = rel(0.5)),
          legend.position = "right",
          legend.direction = "vertical",
          legend.title = element_text(size = rel(0.5)))

  print(line_plot)
}
```

**Fantasyfootball**  
FantasyPoints

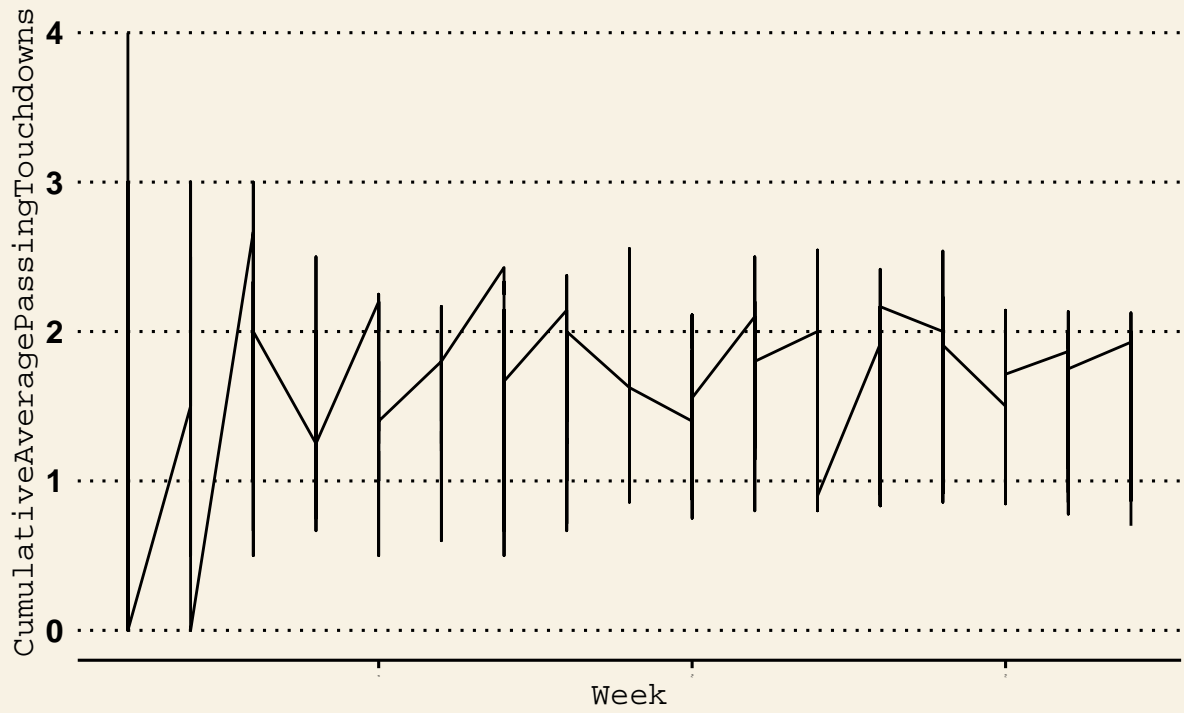


**Fantasyfootball**  
CumulativeAveragePassingYards



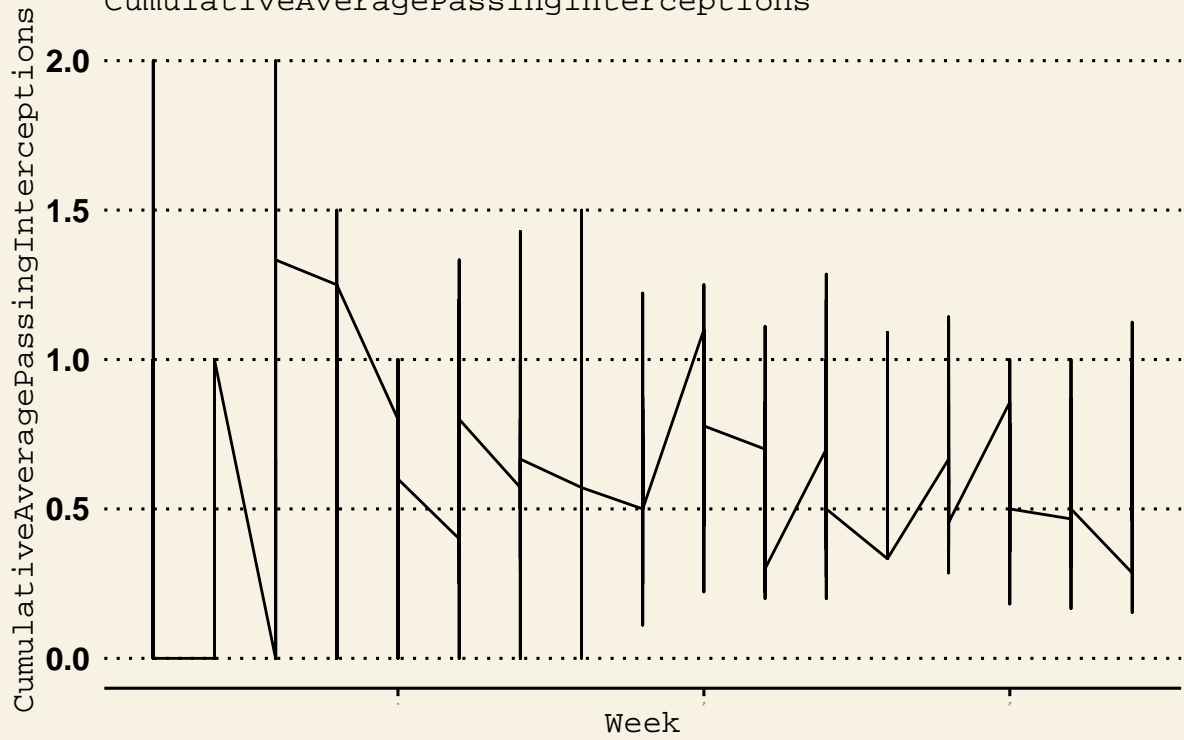
### Fantasyfootball

CumulativeAveragePassingTouchdowns

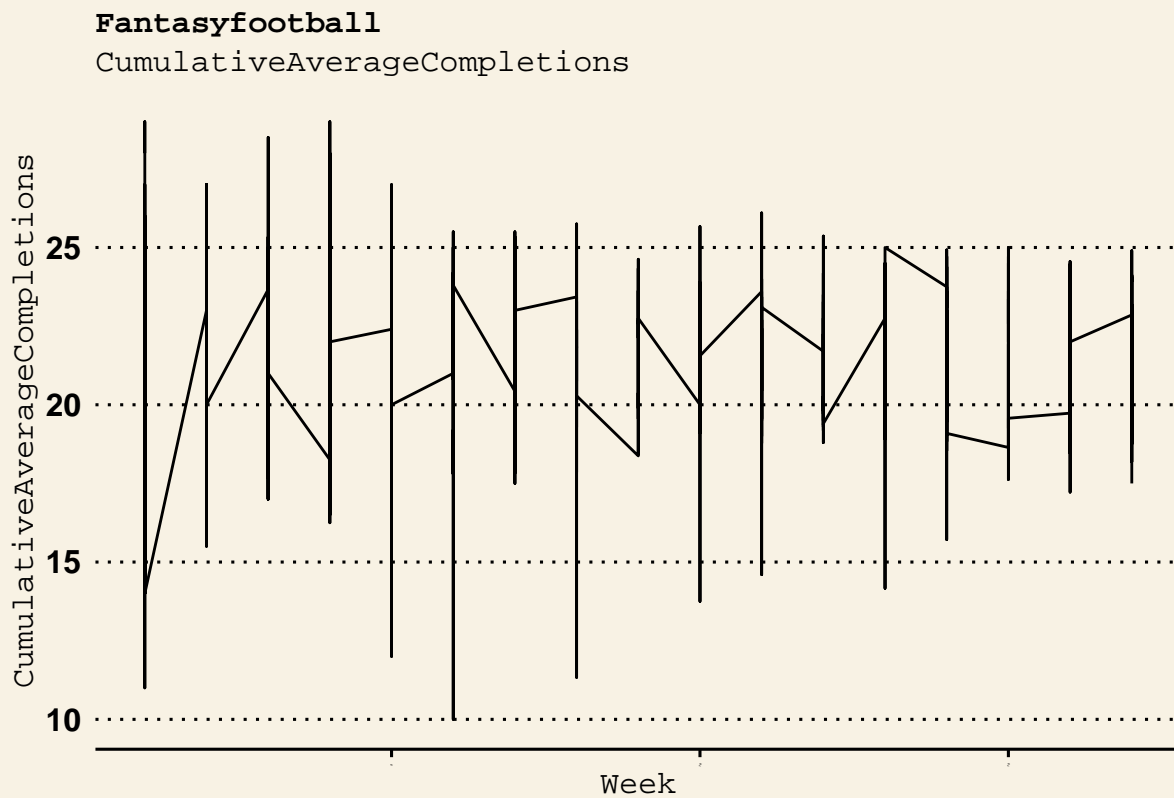
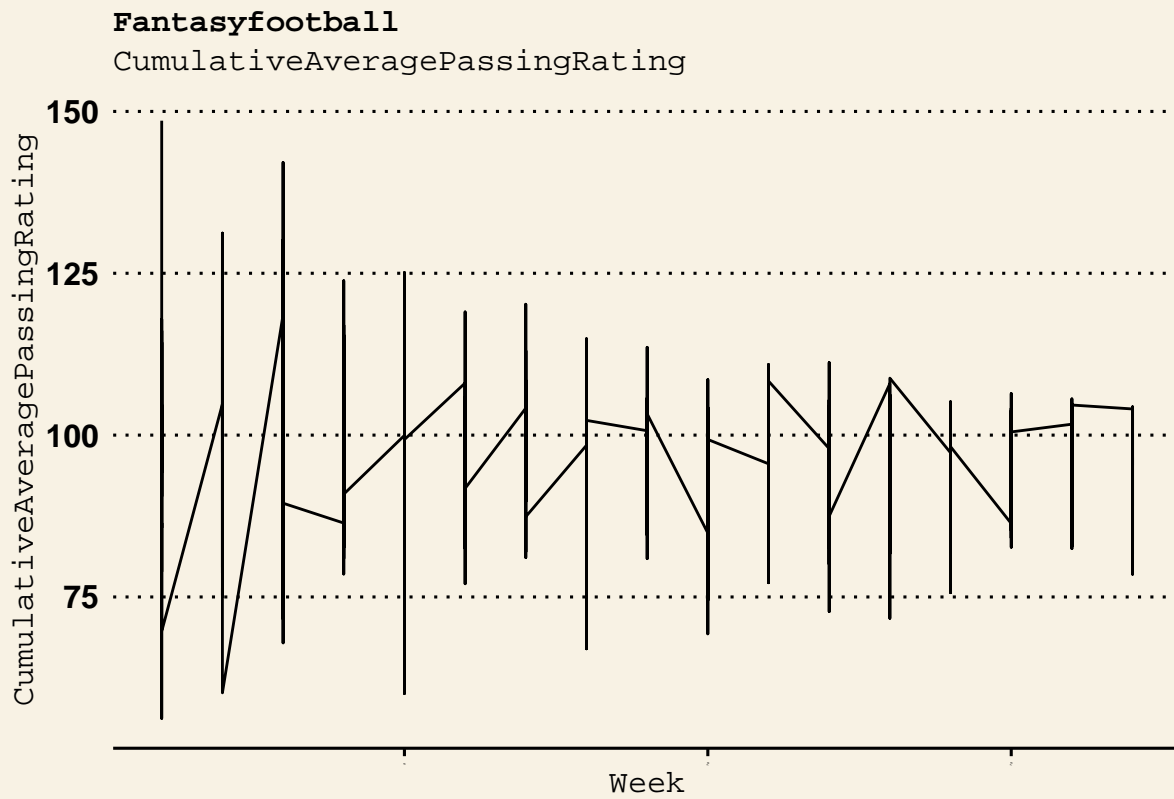


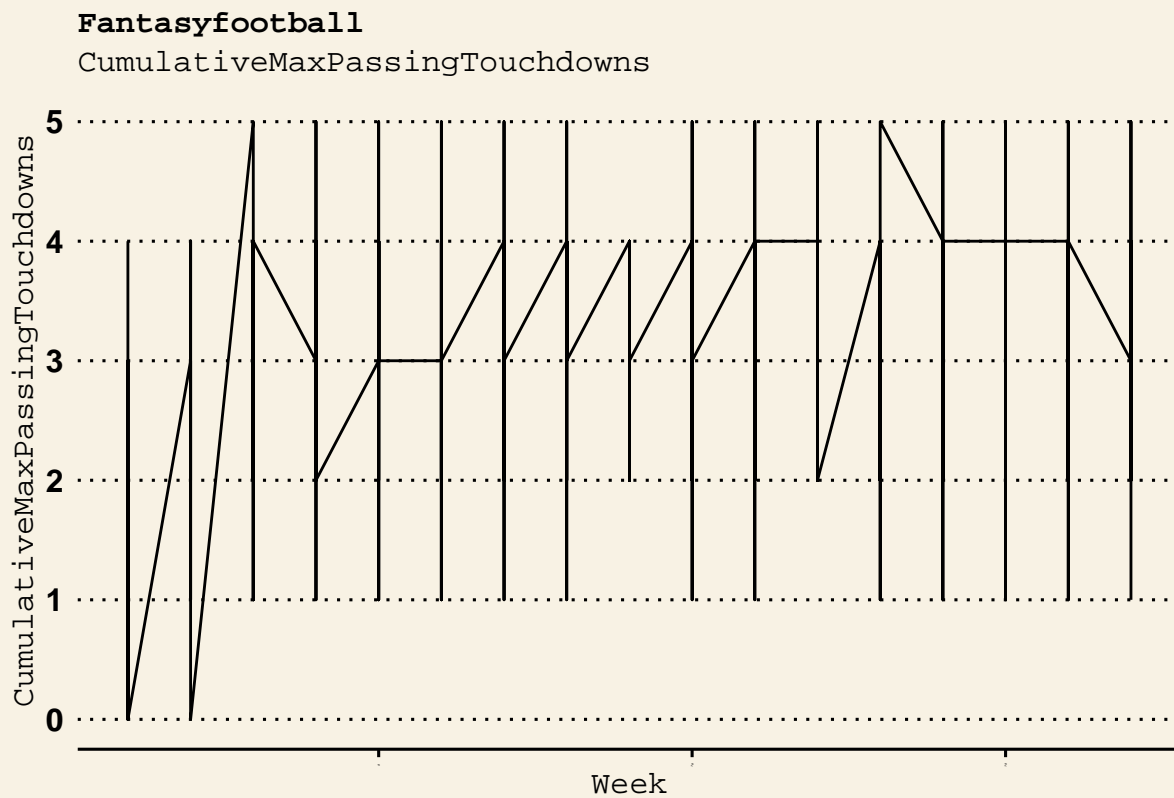
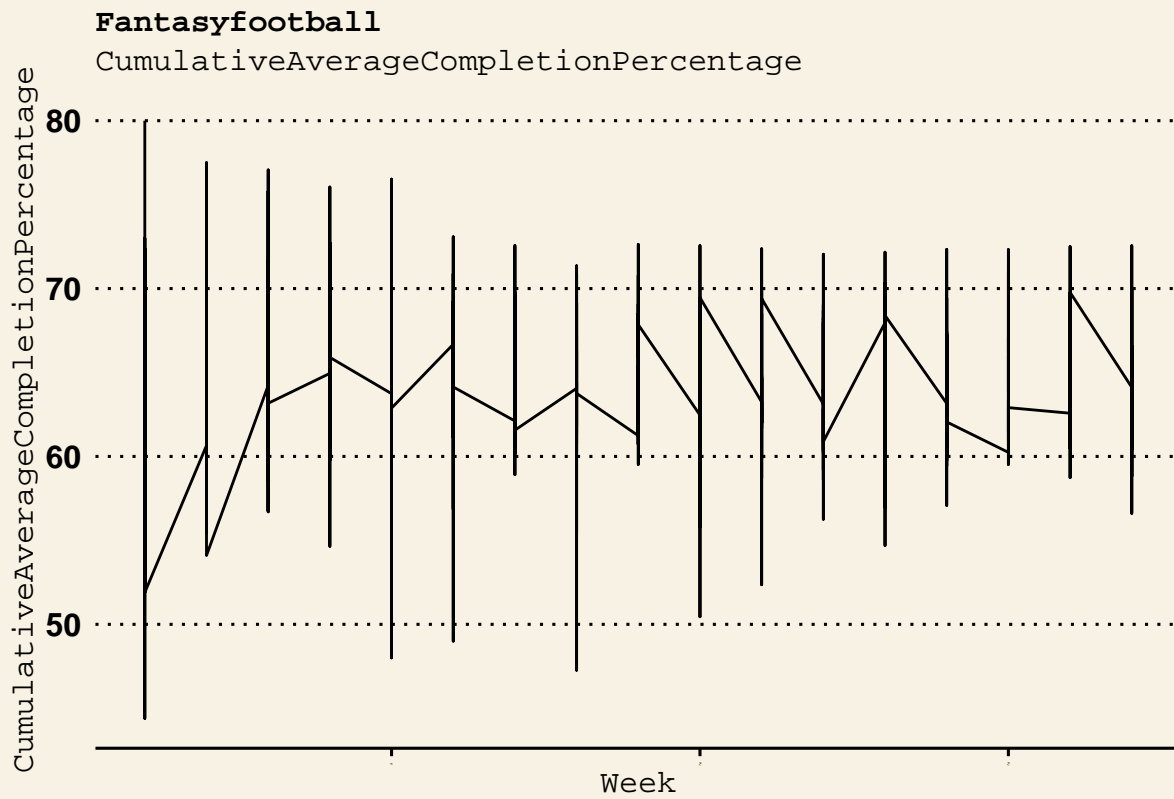
### Fantasyfootball

CumulativeAveragePassingInterceptions



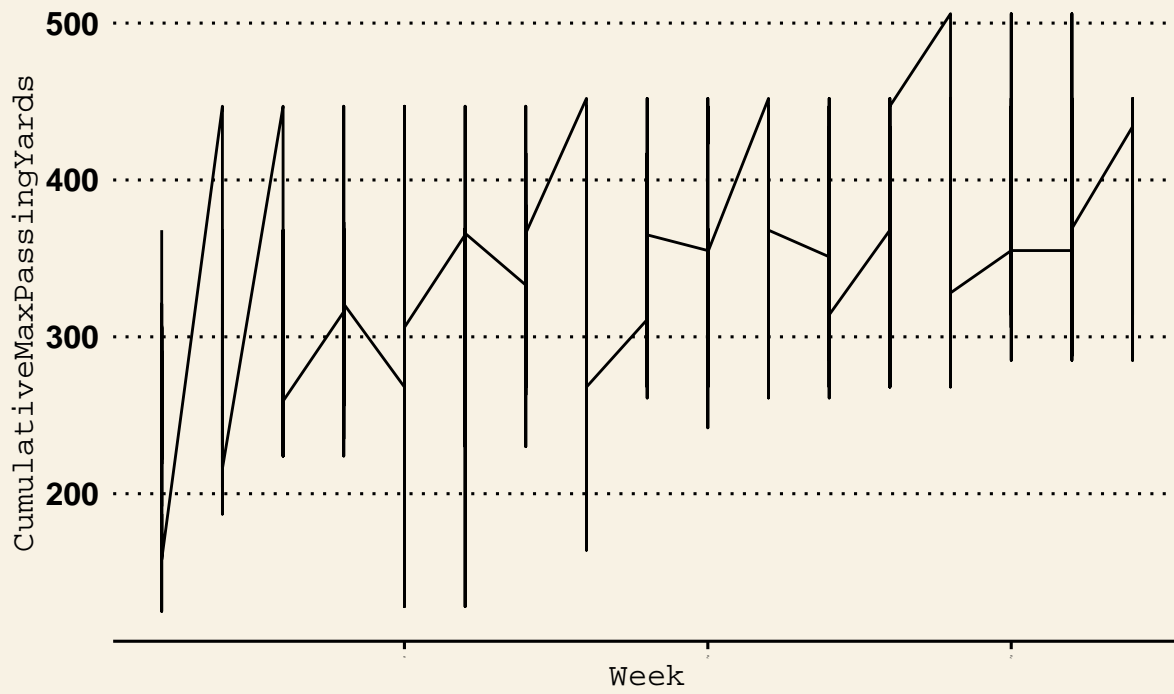






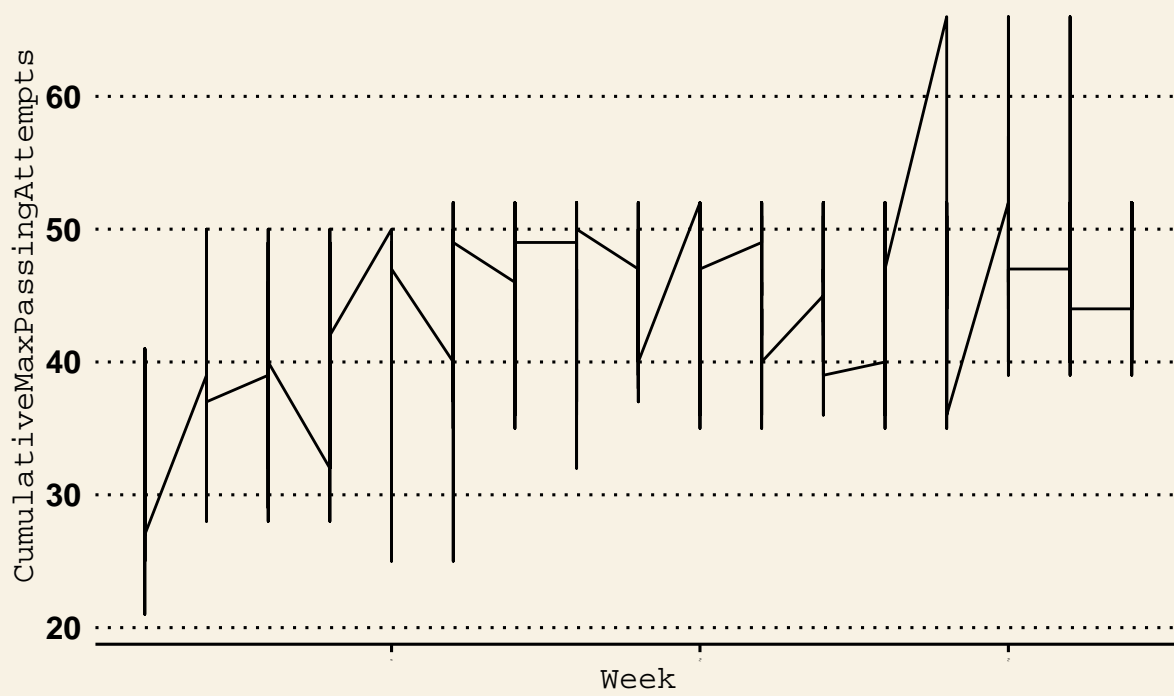
### Fantasyfootball

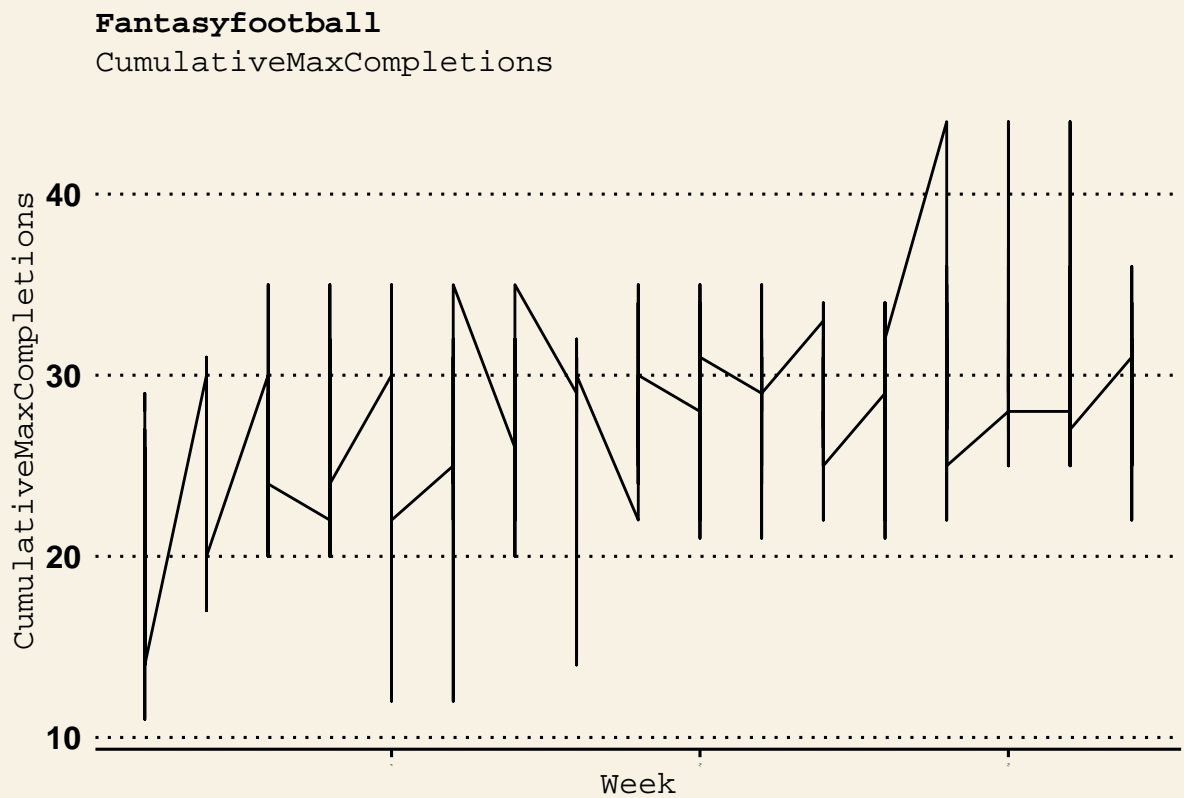
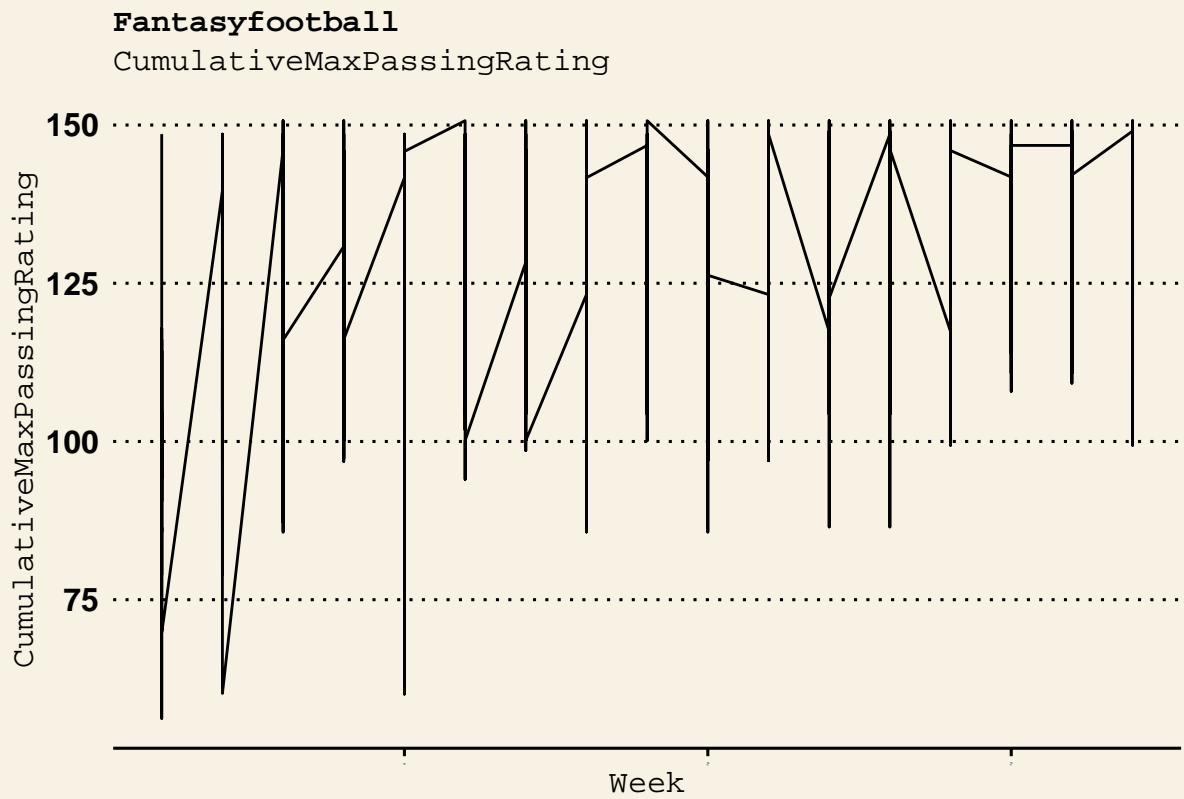
CumulativeMaxPassingYards

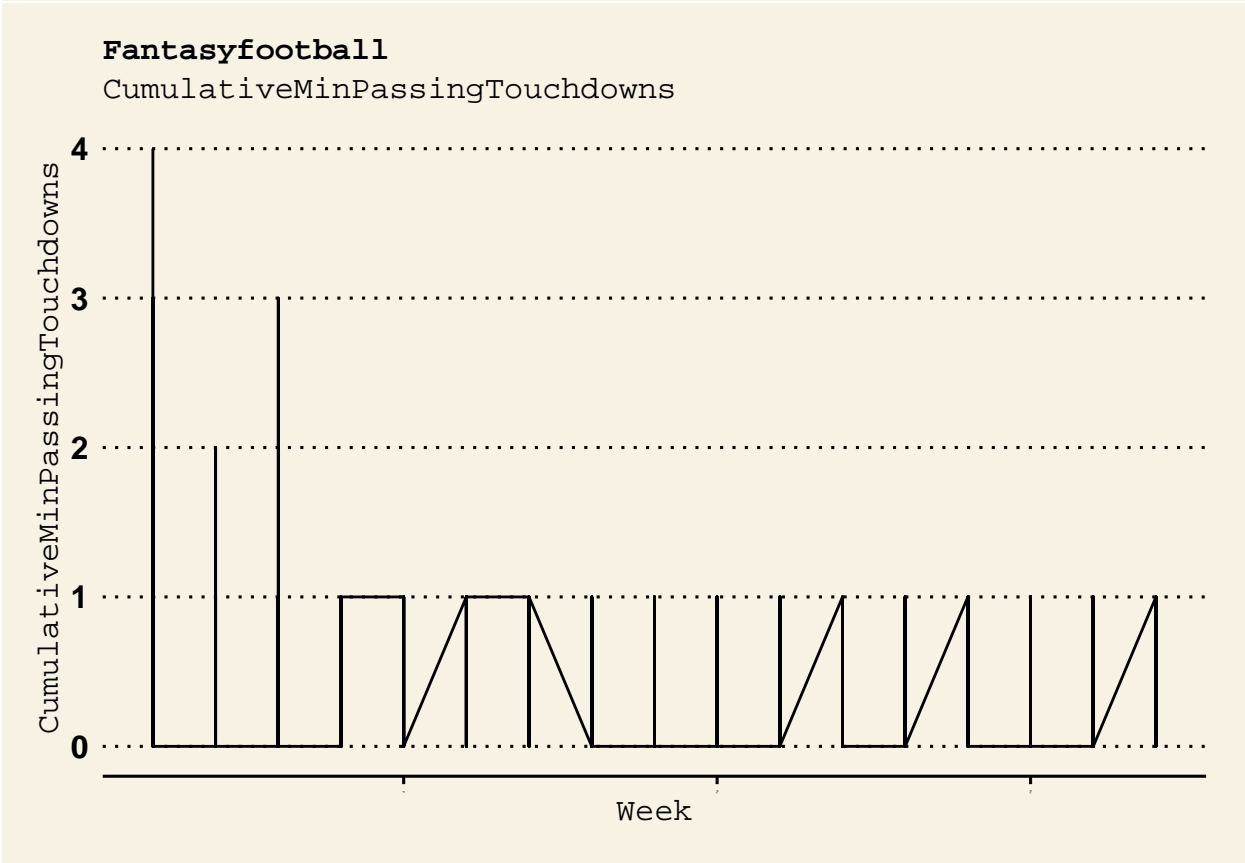
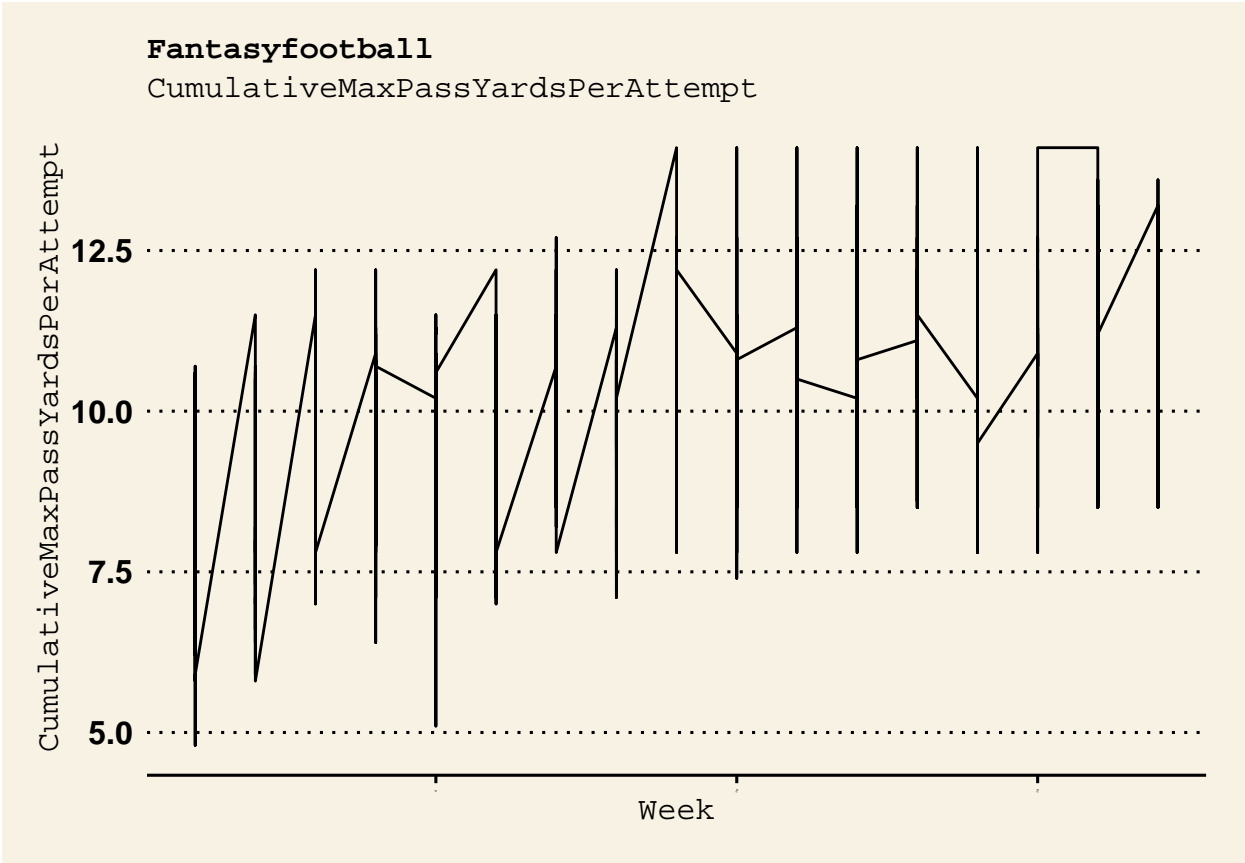


### Fantasyfootball

CumulativeMaxPassingAttempts

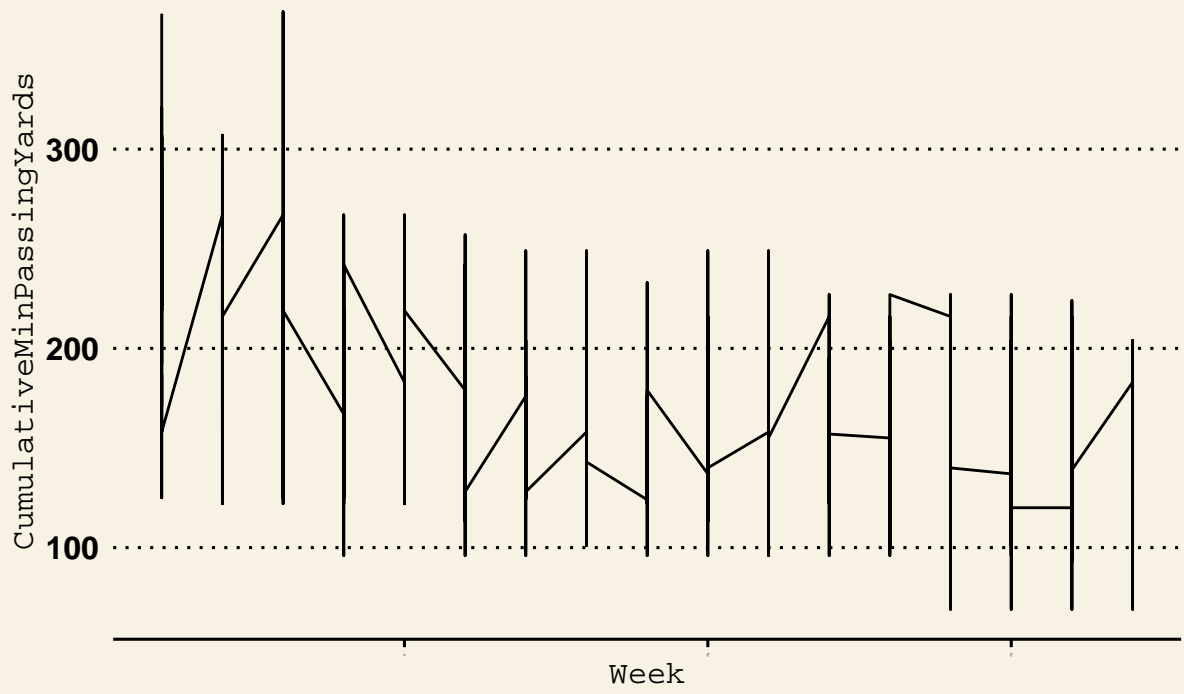






# Fantasyfootball

CumulativeMinPassingYards



# Fantasyfootball

CumulativeMinPassingAttempts

