UFC RESEARCH QUESTION 1

Due 5th December, 2024

Mu Niu, Nruta Choudhari

Research Question 1:

How does the reach of the fighter relate to the total number of strikes landed during a fight?

- Outcome Variable (include the name/description and type of variable): Reach (continuous)

- Independent Variable: Strikes landed (continuous variable)

- Interaction Term: Weight class, allowing analysis of the interaction between reach and weight class on striking effectiveness.

Data Cleaning:

```r
# Import the CSV into R
ufc <- read.csv("ufc-master.csv")

# View the data
head(ufc)
```

```
          RedFighter          BlueFighter RedOdds BlueOdds RedExpectedValue
1      Brandon Moreno          Amir Albazi    -180      150          55.5556
2    Erin Blanchfield      Rose Namajunas    -155      130          64.5161
3        Caio Machado     Brendson Ribeiro    -192      160          52.0833
4     Ariane da Silva  Jasmine Jasudavicius     190     -250         190.0000
5 Marc-Andre Barriault     Dustin Stoltzfus    -270      220          37.0370
6         Mike Malott         Trevin Giles    -375      295          26.6667
  BlueExpectedValue       Date                  Location Country Winner
1               150 2024-11-02 Edmonton, Alberta, Canada  Canada    Red
2               130 2024-11-02 Edmonton, Alberta, Canada  Canada    Red
3               160 2024-11-02 Edmonton, Alberta, Canada  Canada   Blue
4                40 2024-11-02 Edmonton, Alberta, Canada  Canada   Blue
```

```
5              220 2024-11-02 Edmonton, Alberta, Canada  Canada     Blue
6              295 2024-11-02 Edmonton, Alberta, Canada  Canada     Red
  TitleBout       WeightClass Gender NumberOfRounds BlueCurrentLoseStreak
1    False         Flyweight   MALE              5                     0
2    False Women's Flyweight FEMALE              3                     0
3    False Light Heavyweight   MALE              3                     2
4    False Women's Flyweight FEMALE              3                     0
5    False       Middleweight   MALE              3                     1
6    False       Welterweight   MALE              3                     2
  BlueCurrentWinStreak BlueDraws BlueAvgSigStrLanded BlueAvgSigStrPct
1                    5         0                2.72             0.34
2                    2         0                3.71             0.42
3                    0         0                3.16             0.42
4                    2         0                3.70             0.44
5                    0         0                3.47             0.44
6                    0         0                3.17             0.54
  BlueAvgSubAtt BlueAvgTDLanded BlueAvgTDPct BlueLongestWinStreak BlueLosses
1           0.5            1.39         0.32                    5          0
2           0.4            1.46         0.50                    3          5
3           0.4            1.27         0.50                    1          2
4           0.6            2.50         0.43                    2          2
5           1.1            2.31         0.41                    1          5
6           0.3            1.07         0.50                    3          6
  BlueTotalRoundsFought BlueTotalTitleBouts BlueWinsByDecisionMajority
1                    13                   0                          0
2                    52                   7                          0
3                     5                   0                          0
4                    24                   0                          0
5                    17                   0                          0
6                    32                   0                          0
  BlueWinsByDecisionSplit BlueWinsByDecisionUnanimous BlueWinsByKO
1                       1                           1            1
2                       2                           4            2
3                       0                           0            1
4                       0                           5            0
5                       0                           1            1
6                       2                           2            3
  BlueWinsBySubmission BlueWinsByTKODoctorStoppage BlueWins BlueStance
1                    2                           0        5   Orthodox
2                    3                           0       11   Orthodox
3                    0                           0        1   Orthodox
4                    1                           0        6   Orthodox
5                    1                           0        3   Orthodox
```

```
6                        0                        0       7    Orthodox
```

	BlueHeightCms	BlueReachCms	BlueWeightLbs	RedCurrentLoseStreak
1	165.10	172.72	125	2
2	165.10	165.10	125	1
3	190.50	205.74	205	2
4	170.18	172.72	125	1
5	182.88	190.50	185	2
6	182.88	187.96	170	1

	RedCurrentWinStreak	RedDraws	RedAvgSigStrLanded	RedAvgSigStrPct	RedAvgSubAtt
1	0	2	3.99	0.43	0.4
2	0	0	5.24	0.44	0.8
3	0	0	5.82	0.53	0.0
4	0	0	4.04	0.41	0.3
5	0	0	5.98	0.47	0.1
6	0	0	3.85	0.46	0.8

	RedAvgTDLanded	RedAvgTDPct	RedLongestWinStreak	RedLosses	RedTotalRoundsFought
1	1.59	0.46	3	5	53
2	1.86	0.31	6	1	19
3	0.00	0.00	1	2	9
4	0.52	0.55	3	6	29
5	0.21	0.25	2	7	31
6	2.27	0.56	4	1	8

	RedTotalTitleBouts	RedWinsByDecisionMajority	RedWinsByDecisionSplit
1	6	0	1
2	0	0	0
3	0	0	0
4	0	0	1
5	0	0	0
6	0	0	0

	RedWinsByDecisionUnanimous	RedWinsByKO	RedWinsBySubmission
1	2	3	3
2	3	0	3
3	1	0	0
4	3	0	2
5	2	2	1
6	0	1	3

	RedWinsByTKODoctorStoppage	RedWins	RedStance	RedHeightCms	RedReachCms
1	0	9	Orthodox	170.18	177.80
2	0	6	Orthodox	162.56	167.64
3	0	1	Southpaw	193.04	198.12
4	0	6	Orthodox	167.64	170.18
5	0	5	Orthodox	185.42	187.96
6	0	4	Orthodox	185.42	185.42

```
  RedWeightLbs RedAge BlueAge LoseStreakDif WinStreakDif LongestWinStreakDif
1          125     30      31            -2            5                   2
2          125     25      32            -1            2                  -3
3          250     30      28             0            0                   0
4          125     30      35            -1            2                  -1
5          185     34      32            -1            0                  -1
6          170     32      32             1            0                  -1
  WinDif LossDif TotalRoundDif TotalTitleBoutDif KODif SubDif HeightDif
1     -4      -5           -40                -6    -2     -1     -5.08
2      5       4            33                 7     2      0      2.54
3      0       0            -4                 0     1      0     -2.54
4      0      -4            -5                 0     0     -1      2.54
5     -2      -2           -14                 0    -1      0     -2.54
6      3       5            24                 0     2     -3     -2.54
  ReachDif AgeDif SigStrDif AvgSubAttDif AvgTDDif EmptyArena BMatchWCRank
1    -5.08      1     -1.27          0.1    -0.20         NA            3
2    -2.54      7     -1.53         -0.4    -0.40         NA            5
3     7.62     -2     -2.66          0.4     1.27         NA           NA
4     2.54      5     -0.34          0.3     1.98         NA           14
5     2.54     -2     -2.51          1.0     2.10         NA           NA
6     2.54      0     -0.68         -0.5    -1.20         NA           NA
  RMatchWCRank RWFlyweightRank RWFeatherweightRank RWStrawweightRank
1            2              NA                  NA                NA
2            3               3                  NA                NA
3           NA              NA                  NA                NA
4           13              13                  NA                NA
5           NA              NA                  NA                NA
6           NA              NA                  NA                NA
  RWBantamweightRank RHeavyweightRank RLightHeavyweightRank RMiddleweightRank
1                 NA               NA                    NA                NA
2                 NA               NA                    NA                NA
3                 NA               NA                    NA                NA
4                 NA               NA                    NA                NA
5                 NA               NA                    NA                NA
6                 NA               NA                    NA                NA
  RWelterweightRank RLightweightRank RFeatherweightRank RBantamweightRank
1                NA               NA                 NA                NA
2                NA               NA                 NA                NA
3                NA               NA                 NA                NA
4                NA               NA                 NA                NA
5                NA               NA                 NA                NA
6                NA               NA                 NA                NA
  RFlyweightRank RPFPRank BWFlyweightRank BWFeatherweightRank BWStrawweightRank
```

1	2	NA	NA	NA	NA
2	NA	7	5	NA	NA
3	NA	NA	NA	NA	NA
4	NA	NA	14	NA	NA
5	NA	NA	NA	NA	NA
6	NA	NA	NA	NA	NA

	BWBantamweightRank	BHeavyweightRank	BLightHeavyweightRank	BMiddleweightRank
1	NA	NA	NA	NA
2	NA	NA	NA	NA
3	NA	NA	NA	NA
4	NA	NA	NA	NA
5	NA	NA	NA	NA
6	NA	NA	NA	NA

	BWelterweightRank	BLightweightRank	BFeatherweightRank	BBantamweightRank
1	NA	NA	NA	NA
2	NA	NA	NA	NA
3	NA	NA	NA	NA
4	NA	NA	NA	NA
5	NA	NA	NA	NA
6	NA	NA	NA	NA

	BFlyweightRank	BPFPRank	BetterRank	Finish	FinishDetails	FinishRound
1	3	NA	Red	U-DEC		5
2	NA	6	Red	U-DEC		5
3	NA	NA	neither	S-DEC		3
4	NA	NA	Red	SUB	D'Arce Choke	3
5	NA	NA	neither	KO/TKO	Punch	1
6	NA	NA	neither	U-DEC		3

	FinishRoundTime	TotalFightTimeSecs	RedDecOdds	BlueDecOdds	RSubOdds	BSubOdds
1	5:00	1500	130	215	800	1100
2	5:00	1500	140	180	450	1600
3	5:00	900	300	500	800	600
4	2:28	748	300	-110	2000	450
5	4:28	268	150	350	1800	800
6	5:00	900	275	550	215	2000

	RKOOdds	BKOOdds
1	400	1600
2	1600	650
3	200	350
4	800	750
5	225	800
6	240	650

```
# checking for nulls
colSums(is.na(ufc))  # Returns a count of missing values for each column)  # Total number of
```

RedFighter	BlueFighter
0	0
RedOdds	BlueOdds
219	219
RedExpectedValue	BlueExpectedValue
219	219
Date	Location
0	0
Country	Winner
0	0
TitleBout	WeightClass
0	0
Gender	NumberOfRounds
0	0
BlueCurrentLoseStreak	BlueCurrentWinStreak
0	0
BlueDraws	BlueAvgSigStrLanded
0	930
BlueAvgSigStrPct	BlueAvgSubAtt
765	832
BlueAvgTDLanded	BlueAvgTDPct
833	842
BlueLongestWinStreak	BlueLosses
0	0
BlueTotalRoundsFought	BlueTotalTitleBouts
0	0
BlueWinsByDecisionMajority	BlueWinsByDecisionSplit
0	0
BlueWinsByDecisionUnanimous	BlueWinsByKO
0	0
BlueWinsBySubmission	BlueWinsByTKODoctorStoppage
0	0
BlueWins	BlueStance
0	0
BlueHeightCms	BlueReachCms
0	0
BlueWeightLbs	RedCurrentLoseStreak
0	0
RedCurrentWinStreak	RedDraws

```
                           0                                 0
           RedAvgSigStrLanded                   RedAvgSigStrPct
                         455                               357
                 RedAvgSubAtt                    RedAvgTDLanded
                         357                               357
                  RedAvgTDPct              RedLongestWinStreak
                         367                                 0
                    RedLosses             RedTotalRoundsFought
                           0                                 0
           RedTotalTitleBouts        RedWinsByDecisionMajority
                           0                                 0
        RedWinsByDecisionSplit    RedWinsByDecisionUnanimous
                           0                                 0
                  RedWinsByKO              RedWinsBySubmission
                           0                                 0
   RedWinsByTKODoctorStoppage                          RedWins
                           0                                 0
                    RedStance                     RedHeightCms
                           0                                 0
                  RedReachCms                     RedWeightLbs
                           0                                 0
                      RedAge                          BlueAge
                           0                                 0
               LoseStreakDif                    WinStreakDif
                           0                                 0
          LongestWinStreakDif                          WinDif
                           0                                 0
                     LossDif                   TotalRoundDif
                           0                                 0
            TotalTitleBoutDif                           KODif
                           0                                 0
                      SubDif                       HeightDif
                           0                                 0
                    ReachDif                          AgeDif
                           0                                 0
                    SigStrDif                   AvgSubAttDif
                           0                                 0
                    AvgTDDif                      EmptyArena
                           0                              1436
                 BMatchWCRank                     RMatchWCRank
                        5289                              4716
                RWFlyweightRank            RWFeatherweightRank
                        6382                              6469
```

RWStrawweightRank	RWBantamweightRank
6334	6324
RHeavyweightRank	RLightHeavyweightRank
6295	6296
RMiddleweightRank	RWelterweightRank
6296	6290
RLightweightRank	RFeatherweightRank
6295	6303
RBantamweightRank	RFlyweightRank
6299	6292
RPFPRank	BWFlyweightRank
6228	6406
BWFeatherweightRank	BWStrawweightRank
6477	6380
BWBantamweightRank	BHeavyweightRank
6371	6332
BLightHeavyweightRank	BMiddleweightRank
6360	6341
BWelterweightRank	BLightweightRank
6360	6359
BFeatherweightRank	BBantamweightRank
6355	6360
BFlyweightRank	BPFPRank
6348	6411
BetterRank	Finish
0	0
FinishDetails	FinishRound
0	622
FinishRoundTime	TotalFightTimeSecs
0	622
RedDecOdds	BlueDecOdds
1077	1107
RSubOdds	BSubOdds
1326	1350
RKOOdds	BKOOdds
1324	1351

```
# check missing
sum(is.na(ufc$BlueReachCms))
```

```
[1] 0
```

```
sum(is.na(ufc$RedReachCms))
```

```
[1] 0
```

```
sum(is.na(ufc$BlueAvgSigStrLanded))
```

```
[1] 930
```

```
sum(is.na(ufc$RedAvgSigStrLanded))
```

```
[1] 455
```

```
sum(is.na(ufc$WeightClass))
```

```
[1] 0
```

```
filtered_ufc_blue = ufc[!is.na(ufc$BlueAvgSigStrLanded),
                        c("BlueReachCms","BlueAvgSigStrLanded","WeightClass")]
colnames(filtered_ufc_blue) <- c("ReachCms", "AvgSigStrLanded", "WeightClass")
filtered_ufc_red = ufc[!is.na(ufc$RedAvgSigStrLanded),
                       c("RedReachCms","RedAvgSigStrLanded","WeightClass")]
colnames(filtered_ufc_red) <- c("ReachCms", "AvgSigStrLanded", "WeightClass")

# appending the two data sets
filtered_ufc = rbind(filtered_ufc_blue, filtered_ufc_red)

# exclude outlier(one observation with 0 cm reach)
filtered_ufc = filtered_ufc[filtered_ufc$ReachCms > 0,]
head(filtered_ufc)
```

```
  ReachCms AvgSigStrLanded        WeightClass
1   172.72            2.72          Flyweight
2   165.10            3.71 Women's Flyweight
3   205.74            3.16 Light Heavyweight
4   172.72            3.70 Women's Flyweight
5   190.50            3.47        Middleweight
6   187.96            3.17        Welterweight
```

Model:

Multiple Linear Regression

Model formulation:

$StrikesLanded = \beta_0 + \beta_1 Reach + \beta 2 WeightClass + \beta_3(Reach * WeightClass) + \epsilon$

```
filtered_ufc$WeightClass = factor(filtered_ufc$WeightClass)

model = lm(AvgSigStrLanded ~ ReachCms * WeightClass, data = filtered_ufc)

summary(model)
```

```
Call:
lm(formula = AvgSigStrLanded ~ ReachCms * WeightClass, data = filtered_ufc)

Residuals:
    Min      1Q  Median      3Q     Max
-34.935 -16.105  -6.059  12.260 130.914

Coefficients:
                                        Estimate Std. Error t value
(Intercept)                              71.85306   16.49692   4.356
ReachCms                                 -0.29303    0.09437  -3.105
WeightClassCatch Weight                 -19.84947   39.70888  -0.500
WeightClassFeatherweight                 49.92197   23.88948   2.090
WeightClassFlyweight                    166.59933   28.61306   5.822
WeightClassHeavyweight                  -45.12146   24.61544  -1.833
WeightClassLight Heavyweight            -48.72561   25.66726  -1.898
WeightClassLightweight                   18.48720   22.06127   0.838
WeightClassMiddleweight                 -20.55274   23.67575  -0.868
WeightClassWelterweight                 -11.30735   22.02135  -0.513
WeightClassWomen's Bantamweight          19.00361   37.66184   0.505
WeightClassWomen's Featherweight        -46.34315   99.19127  -0.467
WeightClassWomen's Flyweight            -92.08390   32.17156  -2.862
WeightClassWomen's Strawweight         -128.57207   29.24545  -4.396
ReachCms:WeightClassCatch Weight          0.06821    0.22096   0.309
ReachCms:WeightClassFeatherweight        -0.26447    0.13472  -1.963
ReachCms:WeightClassFlyweight            -0.99003    0.16650  -5.946
ReachCms:WeightClassHeavyweight           0.24808    0.13223   1.876
ReachCms:WeightClassLight Heavyweight     0.27927    0.13842   2.018
```

```
ReachCms:WeightClassLightweight                -0.07826    0.12403   -0.631
ReachCms:WeightClassMiddleweight                0.12431    0.12972    0.958
ReachCms:WeightClassWelterweight                0.09101    0.12239    0.744
ReachCms:WeightClassWomen's Bantamweight       -0.11287    0.21972   -0.514
ReachCms:WeightClassWomen's Featherweight       0.22516    0.56802    0.396
ReachCms:WeightClassWomen's Flyweight           0.48017    0.18905    2.540
ReachCms:WeightClassWomen's Strawweight         0.78436    0.17620    4.452
                                              Pr(>|t|)
(Intercept)                                    1.34e-05 ***
ReachCms                                        0.00191 **
WeightClassCatch Weight                         0.61717
WeightClassFeatherweight                        0.03667 *
WeightClassFlyweight                            5.95e-09 ***
WeightClassHeavyweight                          0.06682 .
WeightClassLight Heavyweight                     0.05767 .
WeightClassLightweight                          0.40205
WeightClassMiddleweight                         0.38536
WeightClassWelterweight                         0.60763
WeightClassWomen's Bantamweight                 0.61386
WeightClassWomen's Featherweight                0.64036
WeightClassWomen's Flyweight                    0.00421 **
WeightClassWomen's Strawweight                  1.11e-05 ***
ReachCms:WeightClassCatch Weight                0.75756
ReachCms:WeightClassFeatherweight               0.04965 *
ReachCms:WeightClassFlyweight                   2.83e-09 ***
ReachCms:WeightClassHeavyweight                 0.06067 .
ReachCms:WeightClassLight Heavyweight            0.04366 *
ReachCms:WeightClassLightweight                 0.52807
ReachCms:WeightClassMiddleweight                0.33790
ReachCms:WeightClassWelterweight                0.45714
ReachCms:WeightClassWomen's Bantamweight        0.60747
ReachCms:WeightClassWomen's Featherweight       0.69183
ReachCms:WeightClassWomen's Flyweight           0.01110 *
ReachCms:WeightClassWomen's Strawweight         8.61e-06 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 19.82 on 11544 degrees of freedom
Multiple R-squared:  0.0325,    Adjusted R-squared:  0.0304
F-statistic: 15.51 on 25 and 11544 DF,  p-value: < 2.2e-16
```

```
par(mfrow = c(2, 2), mar = c(2, 2, 2, 2))
plot(model)
```