Project: Cricket Analysis

1.

The dataset chosen for project two reflects the most runs by cricket batsmen across all three international formats of all time. To obtain this dataset I used the tool kaggle. The data was already conveniently a csv so I simply used read.csv to read it in. Cleaning up the data was fairly simple as well because there were no NA variables. To make the data tidy I split the span the players played into start year and end year. Obtainment and processing gave me zero problems.

```
#read in most runs dataset
cricket = read.csv("most_runs_in_cricket.csv")
#include split of span
cricket = cricket %>% separate(Span, c("start_year", "end_year"), sep = "-")
head(cricket, n = 10)
```

```
##    X                     Player start_year end_year Mat Inns NO  Runs  HS
## 1  0        SR Tendulkar (INDIA)       1989     2013 664  782 74 34357 248
## 2  1 KC Sangakkara (Asia/ICC/SL)       2000     2015 594  666 67 28016 319
## 3  2        RT Ponting (AUS/ICC)       1995     2012 560  668 70 27483 257
## 4  3  DPMD Jayawardene (Asia/SL)       1997     2015 652  725 62 25957 374
## 5  4      JH Kallis (Afr/ICC/SA)       1995     2014 519  617 97 25534 224
## 6  5   R Dravid (Asia/ICC/INDIA)       1996     2012 509  605 72 24208 270
## 7  6             V Kohli (INDIA)       2008     2022 473  527 77 24130 254
## 8  7            BC Lara (ICC/WI)       1990     2007 430  521 38 22358 400
## 9  8     ST Jayasuriya (Asia/SL)       1989     2011 586  651 35 21032 340
## 10 9          S Chanderpaul (WI)       1994     2015 454  553 94 20988 203
##      Ave     BF    SR X100 X50 X0  X4s X6s
## 1  48.52 50817+ 67.58  100 164 34 4076 264
## 2  46.77  42086 66.56   63 153 28 3015 159
## 3  45.95  40130 68.48   71 146 39 2781 246
## 4  39.15  40100 64.73   54 136 47 2679 170
## 5  49.10  45346 56.30   62 149 33 2455 254
## 6  45.41  46564 51.98   48 146 21 2604  66
## 7  53.62  30483 79.15   71 125 33 2400 258
## 8  46.28  32839 68.08   53 111 33 2601 221
## 9  34.14  25910 81.17   42 103 53 2486 352
## 10 45.72  40150 52.27   41 125 21 2041 126
```

2.

Recently the T20 cricket international world cup ended and I found an interest into making my project based on cricket batsmen. I wanted to have a statistical answer to “Who is the greatest cricket batsmen of all time?”. To invesigate my question I will be creating new variables about average run rate, centuries, 50s, etc, to find a general conclusion on who the most effective batsman of all time is.

3.

My cleaned data includes a plethora of variables on all the players. My table includes the players overall all time run rank(int), player information(character: name and teams played), start_year(character), end_year(character), total mathces played(int), total runs scored(int), number of 100s scored(int), number of 50s scored(int), number of 4s hit(int), and number of 6s hit(int). There are no NA values in my cleaned dataset.

```
#Subset into my data
mycrick = cricket[,  c("X", "Player", "start_year", "end_year", "Mat", "Runs", "X100", "X50", "X4s", "X6
s")]
mycrick
```

```
##     X                       Player start_year end_year Mat  Runs X100 X50  X4s
## 1   0          SR Tendulkar (INDIA)       1989     2013 664 34357  100 164 4076
## 2   1   KC Sangakkara (Asia/ICC/SL)       2000     2015 594 28016   63 153 3015
## 3   2          RT Ponting (AUS/ICC)       1995     2012 560 27483   71 146 2781
## 4   3    DPMD Jayawardene (Asia/SL)       1997     2015 652 25957   54 136 2679
## 5   4        JH Kallis (Afr/ICC/SA)       1995     2014 519 25534   62 149 2455
## 6   5     R Dravid (Asia/ICC/INDIA)       1996     2012 509 24208   48 146 2604
## 7   6               V Kohli (INDIA)       2008     2022 473 24130   71 125 2400
## 8   7              BC Lara (ICC/WI)       1990     2007 430 22358   53 111 2601
## 9   8       ST Jayasuriya (Asia/SL)       1989     2011 586 21032   42 103 2486
## 10  9            S Chanderpaul (WI)       1994     2015 454 20988   41 125 2041
## 11 10 Inzamam-ul-Haq (Asia/ICC/PAK)       1991     2007 499 20580   35 129 2076
## 12 11       AB de Villiers (Afr/SA)       2004     2018 420 20014   47 109 2004
## 13 12             CH Gayle (ICC/WI)       1999     2021 483 19593   42 105 2332
## 14 13           HM Amla (SA/World)        2004     2019 349 18672   55  88 2138
## 15 14       SC Ganguly (Asia/INDIA)       1992     2008 424 18575   38 107 2022
## 16 15                SR Waugh (AUS)       1985     2004 493 18496   35  95 1705
## 17 16             LRPL Taylor (NZ)        2006     2022 450 18199   40  93 1766
## 18 17            Younis Khan (PAK)        2000     2017 408 17790   41  83 1691
## 19 18              AR Border (AUS)        1978     1994 429 17698   30 102 1661
## 20 19              TM Dilshan (SL)        1999     2016 497 17671   39  83 2011
## 21 20                JE Root (ENG)        2012     2022 314 17604   44  95 1756
## 22 21    Mohammad Yousuf (Asia/PAK)       1998     2010 381 17300   39  97 1747
## 23 22         MS Dhoni (Asia/INDIA)       2004     2019 538 17266   16 108 1486
## 24 23     V Sehwag (Asia/ICC/INDIA)       1999     2013 374 17253   38  72 2408
## 25 24         GC Smith (Afr/ICC/SA)       2002     2014 347 17236   37  90 2076
## 26 25              MJ Clarke (AUS)        2003     2015 394 17112   36  86 1672
## 27 26               ME Waugh (AUS)        1988     2002 372 16529   38  97 1495
## 28 27              DA Warner (AUS)        2009     2022 329 16466   43  84 1833
## 29 28            RG Sharma (INDIA)        2007     2022 420 16250   41  87 1528
## 30 29          Javed Miandad (PAK)        1975     1996 357 16213   31  93 1233
## 31 30               DL Haynes (WI)        1978     1994 354 16135   35  96 1586
## 32 31           KS Williamson (NZ)        2010     2022 324 15889   37  87 1605
## 33 32                AN Cook (ENG)        2006     2018 257 15737   38  76 1815
## 34 33             PA de Silva (SL)        1984     2003 401 15645   31  86 1449
## 35 34          M Azharuddin (INDIA)       1984     2000 433 15593   29  79 1342
## 36 35        AC Gilchrist (AUS/ICC)       1996     2008 396 15461   33  81 1866
## 37 36           SP Fleming (ICC/NZ)       1994     2008 396 15319   17  95 1760
## 38 37            IVA Richards (WI)        1974     1991 308 15261   35  90 1552
## 39 38          ML Hayden (AUS/ICC)        1993     2009 273 15066   40  69 1722
## 40 39   Tamim Iqbal (BAN/ICC/World)       2007     2022 378 14914   25  93 1728
## 41 40              BB McCullum (NZ)        2002     2016 432 14676   19  76 1552
## 42 41                 HH Gibbs (SA)        1996     2010 361 14661   35  66 1862
## 43 42               G Kirsten (SA)         1993     2004 286 14087   34  79 1581
## 44 43             MS Atapattu (SL)         1990     2007 360 14036   27  76 1419
## 45 44              AD Mathews (SL)         2008     2022 396 13936   16  83 1250
## 46 45              SPD Smith (AUS)         2010     2022 285 13887   40  67 1399
## 47 46        KP Pietersen (ENG/ICC)        2004     2014 277 13797   32  67 1531
## 48 47         Mushfiqur Rahim (BAN)        2005     2022 420 13509   17  73 1284
## 49 48              MJ Guptill (NZ)         2009     2022 367 13463   23  76 1385
## 50 49                DC Boon (AUS)         1984     1996 288 13386   26  69 1316
## 51 50                IR Bell (ENG)         2004     2015 287 13331   26  82 1467
## 52 51          SM Gavaskar (INDIA)         1971     1987 233 13214   35  72 1142
## 53 52         Shakib Al Hasan (BAN)        2006     2022 388 13205   14  91 1350
## 54 53               GA Gooch (ENG)         1975     1995 243 13190   28  69 1486
## 55 54             AJ Stewart (ENG)         1989     2003 303 13140   19  73 1590
## 56 55           Saleem Malik (PAK)         1982     1999 386 12938   20  76 1192
## 57 56            Saeed Anwar (PAK)         1989     2003 302 12876   31  68 1473
## 58 57         Mohammad Hafeez (PAK)        2003     2021 392 12780   21  64 1370
## 59 58             CG Greenidge (WI)        1974     1991 236 12692   30  65 1342
## 60 59              A Ranatunga (SL)        1982     2000 362 12561    8  87 1057
## 61 60              MEK Hussey (AUS)        2004     2013 302 12398   22  72 1126
## 62 61           RB Richardson (WI)         1983     1996 310 12197   21  71 1302
```

```
## 63 62                RR Sarwan (WI)       2000     2013 286 11944   20  71 1246
## 64 63         Shoaib Malik (ICC/PAK)      1999     2021 446 11867   12  61 1038
## 65 64                 NJ Astle (NZ)       1995     2007 308 11866   27  65 1339
## 66 65     Yuvraj Singh (Asia/INDIA)       2000     2017 402 11778   17  71 1245
## 67 66                A Flower (ZIM)       1992     2003 276 11580   16  82 1075
## 68 67                CL Hooper (WI)       1987     2003 329 11523   20  56 1042
## 69 68                DI Gower (ENG)       1978     1992 231 11401   25  51 1269
## 70 69              AJ Strauss (ENG)       2003     2012 231 11315   27  54 1330
## 71 70       F du Plessis (SA/World)       2011     2021 262 11198   23  66 1151
## 72 71  Shahid Afridi (Asia/ICC/PAK)       1996     2018 524 11196   11  51 1053
## 73 72                Q de Kock (SA)       2012     2022 261 11165   23  64 1287
## 74 73               MN Samuels (WI)       2000     2018 345 11134   17  64 1207
## 75 74           Misbah-ul-Haq (PAK)       2001     2017 276 11132   10  84  898
## 76 75            VVS Laxman (INDIA)       1996     2012 220 11119   23  66 1357
## 77 76               MA Taylor (AUS)       1989     1999 217 11039   20  68 1000
## 78 77              Babar Azam (PAK)       2015     2022 226 11017   26  74 1124
## 79 78               SR Watson (AUS)       2002     2016 307 10950   14  67 1168
## 80 79          EJG Morgan (ENG/IRE)       2006     2022 379 10859   16  64  917
## 81 80              S Dhawan (INDIA)       2010     2022 263 10746   24  54 1331
## 82 81       MV Boucher (Afr/ICC/SA)       1997     2012 467 10469    6  61 1034
## 83 82             JM Bairstow (ENG)       2011     2022 250 10453   23  46 1176
## 84 83         DB Vengsarkar (INDIA)       1976     1992 245 10376   18  58  737
## 85 84          ME Trescothick (ENG)       2000     2006 202 10326   26  52 1382
## 86 85             G Gambhir (INDIA)       2003     2016 242 10324   20  63 1188
## 87 86                MD Crowe (NZ)       1982     1995 220 10148   21  52 1037
## 88 87               GW Flower (ZIM)       1992     2010 288 10028   12  55  906
##    X6s
## 1  264
## 2  159
## 3  246
## 4  170
## 5  254
## 6   66
## 7  258
## 8  221
## 9  352
## 10 126
## 11 193
## 12 328
## 13 553
## 14  93
## 15 247
## 16  88
## 17 273
## 18 138
## 19  71
## 20 112
## 21  90
## 22 142
## 23 359
## 24 243
## 25  94
## 26 102
## 27  98
## 28 256
## 29 492
## 30  92
## 31  77
## 32 117
## 33  21
## 34 150
## 35  96
## 36 262
```

```
## 37  89
## 38 210
## 39 182
## 40 186
## 41 398
## 42 187
## 43  32
## 44  19
## 45 202
## 46 114
## 47 190
## 48 153
## 49 383
## 50  18
## 51  73
## 52  47
## 53 113
## 54  40
## 55  36
## 56  41
## 57 111
## 58 214
## 59 148
## 60 104
## 61 144
## 62  77
## 63  78
## 64 199
## 65 127
## 66 251
## 67  46
## 68 128
## 69  32
## 70  35
## 71 137
## 72 476
## 73 194
## 74 219
## 75 190
## 76   9
## 77  16
## 78 114
## 79 245
## 80 346
## 81 140
## 82 105
## 83 192
## 84  41
## 85  84
## 86  37
## 87  56
## 88  53
```

4.

The new variables created from my cleaned data include total years played(int), runs per year(int), and 100s per year(int). There are no missing values in this table. Total years played is derived by subtracting end year and start year by using the created function subtract. Runs per year is derived from the total run count divided by years played. 100s per year is derived from total centuries(X100) divided by years played.

```
#Subtract function to find span of years played
subtract = function(x,y){
  span1 = as.numeric(x)
  span2 = as.numeric(y)
  span = span1-span2
  return (span)
}

mycrick$years_played = subtract(mycrick$end_year, mycrick$start_year)

mycrick$runs_per_year = round(mycrick$Runs/mycrick$years_played, digits = 2)
mycrick$X100s_per_year = round(mycrick$X100/mycrick$years_played, digits = 2)
newvariables = mycrick[, c("Player", "years_played", "runs_per_year", "X100s_per_year")]
newvariables
```

```
##                          Player years_played runs_per_year X100s_per_year
## 1           SR Tendulkar (INDIA)           24       1431.54           4.17
## 2    KC Sangakkara (Asia/ICC/SL)           15       1867.73           4.20
## 3           RT Ponting (AUS/ICC)           17       1616.65           4.18
## 4     DPMD Jayawardene (Asia/SL)           18       1442.06           3.00
## 5         JH Kallis (Afr/ICC/SA)           19       1343.89           3.26
## 6      R Dravid (Asia/ICC/INDIA)           16       1513.00           3.00
## 7               V Kohli (INDIA)           14       1723.57           5.07
## 8              BC Lara (ICC/WI)           17       1315.18           3.12
## 9        ST Jayasuriya (Asia/SL)           22        956.00           1.91
## 10            S Chanderpaul (WI)           21        999.43           1.95
## 11 Inzamam-ul-Haq (Asia/ICC/PAK)           16       1286.25           2.19
## 12       AB de Villiers (Afr/SA)           14       1429.57           3.36
## 13             CH Gayle (ICC/WI)           22        890.59           1.91
## 14            HM Amla (SA/World)           15       1244.80           3.67
## 15       SC Ganguly (Asia/INDIA)           16       1160.94           2.38
## 16               SR Waugh (AUS)           19        973.47           1.84
## 17             LRPL Taylor (NZ)           16       1137.44           2.50
## 18            Younis Khan (PAK)           17       1046.47           2.41
## 19              AR Border (AUS)           16       1106.12           1.88
## 20              TM Dilshan (SL)           17       1039.47           2.29
## 21                JE Root (ENG)           10       1760.40           4.40
## 22    Mohammad Yousuf (Asia/PAK)           12       1441.67           3.25
## 23         MS Dhoni (Asia/INDIA)           15       1151.07           1.07
## 24     V Sehwag (Asia/ICC/INDIA)           14       1232.36           2.71
## 25          GC Smith (Afr/ICC/SA)           12       1436.33           3.08
## 26               MJ Clarke (AUS)           12       1426.00           3.00
## 27                ME Waugh (AUS)           14       1180.64           2.71
## 28              DA Warner (AUS)           13       1266.62           3.31
## 29            RG Sharma (INDIA)           15       1083.33           2.73
## 30          Javed Miandad (PAK)           21        772.05           1.48
## 31               DL Haynes (WI)           16       1008.44           2.19
## 32           KS Williamson (NZ)           12       1324.08           3.08
## 33                AN Cook (ENG)           12       1311.42           3.17
## 34             PA de Silva (SL)           19        823.42           1.63
## 35          M Azharuddin (INDIA)           16        974.56           1.81
## 36        AC Gilchrist (AUS/ICC)           12       1288.42           2.75
## 37           SP Fleming (ICC/NZ)           14       1094.21           1.21
## 38             IVA Richards (WI)           17        897.71           2.06
## 39           ML Hayden (AUS/ICC)           16        941.62           2.50
## 40   Tamim Iqbal (BAN/ICC/World)           15        994.27           1.67
## 41              BB McCullum (NZ)           14       1048.29           1.36
## 42                 HH Gibbs (SA)           14       1047.21           2.50
## 43                G Kirsten (SA)           11       1280.64           3.09
## 44              MS Atapattu (SL)           17        825.65           1.59
## 45               AD Mathews (SL)           14        995.43           1.14
## 46               SPD Smith (AUS)           12       1157.25           3.33
## 47        KP Pietersen (ENG/ICC)           10       1379.70           3.20
## 48         Mushfiqur Rahim (BAN)           17        794.65           1.00
## 49               MJ Guptill (NZ)           13       1035.62           1.77
## 50                 DC Boon (AUS)           12       1115.50           2.17
## 51                 IR Bell (ENG)           11       1211.91           2.36
## 52           SM Gavaskar (INDIA)           16        825.88           2.19
## 53          Shakib Al Hasan (BAN)           16        825.31           0.88
## 54                GA Gooch (ENG)           20        659.50           1.40
## 55              AJ Stewart (ENG)           14        938.57           1.36
## 56            Saleem Malik (PAK)           17        761.06           1.18
## 57              Saeed Anwar (PAK)           14        919.71           2.21
## 58         Mohammad Hafeez (PAK)           18        710.00           1.17
## 59             CG Greenidge (WI)           17        746.59           1.76
## 60              A Ranatunga (SL)           18        697.83           0.44
## 61              MEK Hussey (AUS)            9       1377.56           2.44
## 62            RB Richardson (WI)           13        938.23           1.62
```

```
## 63                RR Sarwan (WI)          13        918.77           1.54
## 64        Shoaib Malik (ICC/PAK)          22        539.41           0.55
## 65                 NJ Astle (NZ)          12        988.83           2.25
## 66     Yuvraj Singh (Asia/INDIA)          17        692.82           1.00
## 67                A Flower (ZIM)          11       1052.73           1.45
## 68                CL Hooper (WI)          16        720.19           1.25
## 69                DI Gower (ENG)          14        814.36           1.79
## 70              AJ Strauss (ENG)           9       1257.22           3.00
## 71       F du Plessis (SA/World)          10       1119.80           2.30
## 72  Shahid Afridi (Asia/ICC/PAK)          22        508.91           0.50
## 73                Q de Kock (SA)          10       1116.50           2.30
## 74               MN Samuels (WI)          18        618.56           0.94
## 75           Misbah-ul-Haq (PAK)          16        695.75           0.62
## 76            VVS Laxman (INDIA)          16        694.94           1.44
## 77               MA Taylor (AUS)          10       1103.90           2.00
## 78              Babar Azam (PAK)           7       1573.86           3.71
## 79               SR Watson (AUS)          14        782.14           1.00
## 80          EJG Morgan (ENG/IRE)          16        678.69           1.00
## 81              S Dhawan (INDIA)          12        895.50           2.00
## 82       MV Boucher (Afr/ICC/SA)          15        697.93           0.40
## 83             JM Bairstow (ENG)          11        950.27           2.09
## 84         DB Vengsarkar (INDIA)          16        648.50           1.12
## 85          ME Trescothick (ENG)           6       1721.00           4.33
## 86             G Gambhir (INDIA)          13        794.15           1.54
## 87                 MD Crowe (NZ)          13        780.62           1.62
## 88               GW Flower (ZIM)          18        557.11           0.67
```

5.

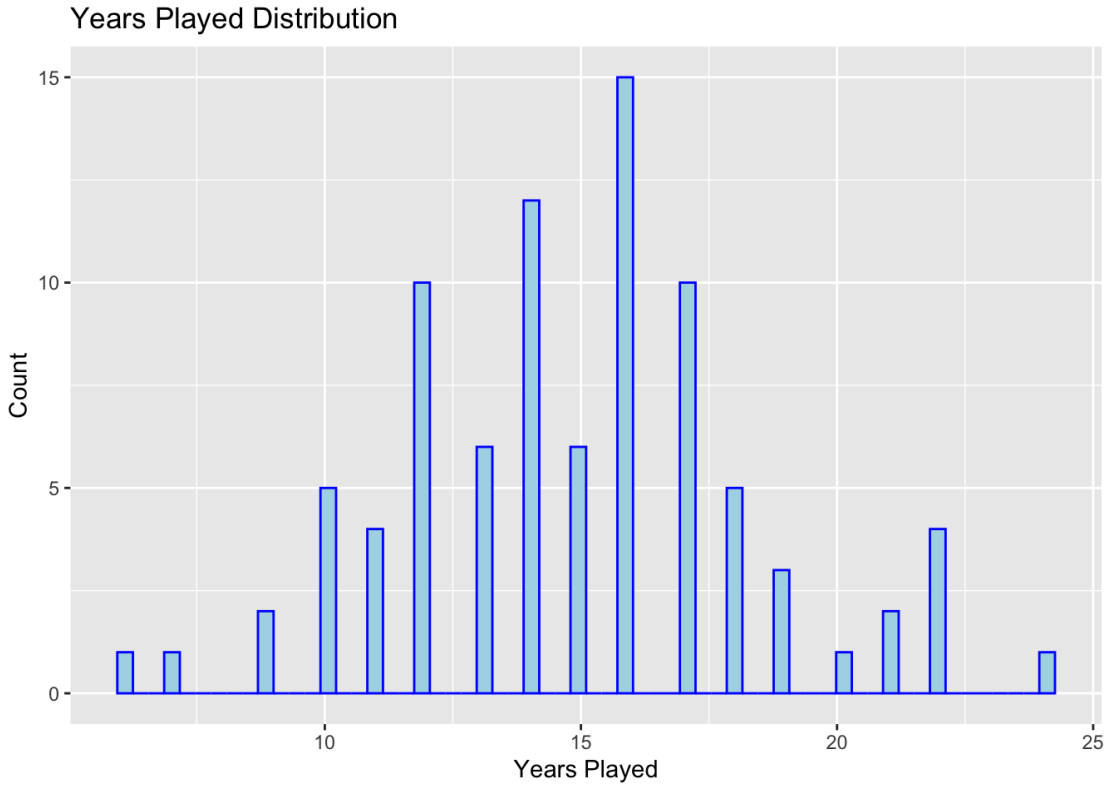
I created all the functions I used in my code to create univariate and bivariate plots. My first function was a histogram function, my second was a density plot, and my third was a dotplot. For my bivariate plots I created a scatterplot function and a line graph function.

```
#Histogram function
makehist = function(var, color.want = "blue", color.fill = "lightblue", binw = 30, t, xa, ya){
  yp = ggplot()+geom_histogram(aes(x = var), color = color.want, fill = color.fill, bins = binw) + labs(tit
le = t, x = xa, y = ya)


  print(paste("Median of Distribution: ", median(var)))
  print(paste("Mean of Distribution ", mean(var)))
  return (yp)

}

#Density Plot function
makedense = function(var, color.want = "blue",color.fill = "lightblue", t, xa, ya){
  yp = ggplot()+ geom_density(aes(x = var), color = color.want, fill = color.fill)+labs(title = t, x = xa,
y = ya)
  return(yp)
}

#Dotplot Function
makedot = function(var, color.want = "blue", color.fill = "lightblue",binw, t, xa, ya){
  yp = ggplot()+geom_dotplot(aes(x = var), color = color.want, fill = color.fill, binwidth = binw)+labs(tit
le = t, x = xa, y = ya)
  return(yp)
}

#ScatterPlot function
makescat = function(var1, var2, color.want = "black", t, xa, ya){
  yp = ggplot()+geom_point(aes(x = var1, y = var2), color = color.want)+labs(title = t, x = xa, y= ya)
  print(paste("Correlation is:", cor(var1, var2)))
  return(yp)
}

#Line graph Function
makeline = function(var1,var2, t, xa, ya){
  yp = ggplot()+geom_line(aes(x = var1, y = var2))+labs(title = t, x = xa, y = ya)
  return(yp)
}
```

For my first graph I made a histogram of total years played for each batsmen in the dataset. The histogram was symmetrical with a bell shape and a median of 15 years and a mean of 14.99 years. The histogram is symmetrical without very much skew which is why the mean and median are very close to each other.

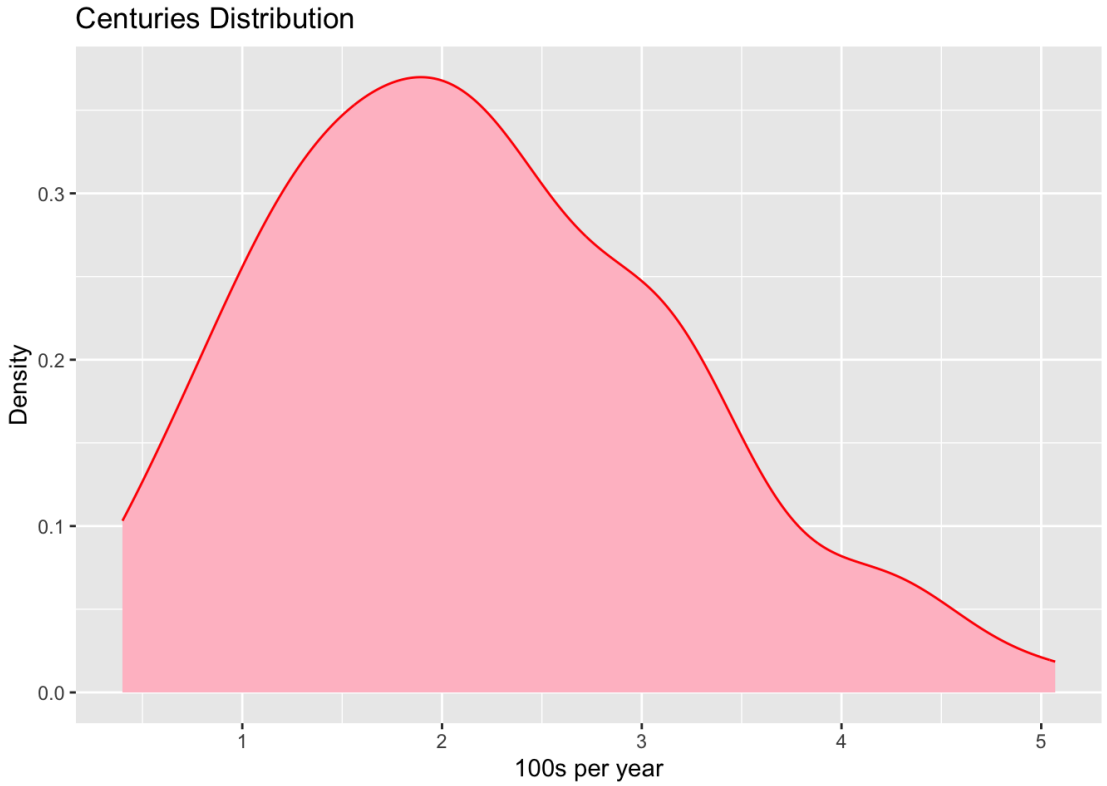
```
makehist(mycrick$years_played, "blue", "lightblue", 60, "Years Played Distribution", "Years Played", "Coun
t")
```

```
## [1] "Median of Distribution:  15"
## [1] "Mean of Distribution  14.9090909090909"
```



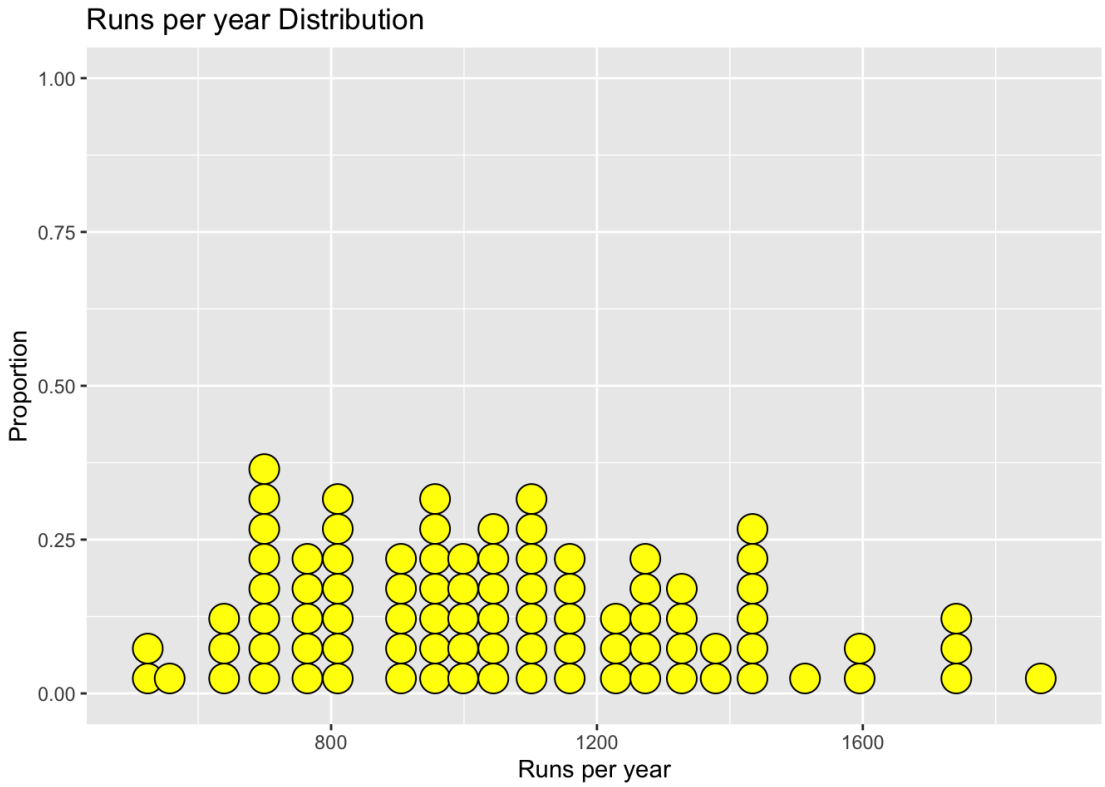
My second univariate graph is a density plot of 100s hit per year. This is a smoothed distribution along a numeric axis and shows that most of the players in the chart have around 2 centuries per year(peak is highest concentration of numeric variable.)

```
makedense(mycrick$X100s_per_year, "Red","pink", "Centuries Distribution", "100s per year", "Density")
```



The dot plot for runs per year shows each observation of runs per year for each batsmen. The highest frequency of runs per year is at 700 runs per year. The dotplot also shows a slight bell shape with a slight right skew.

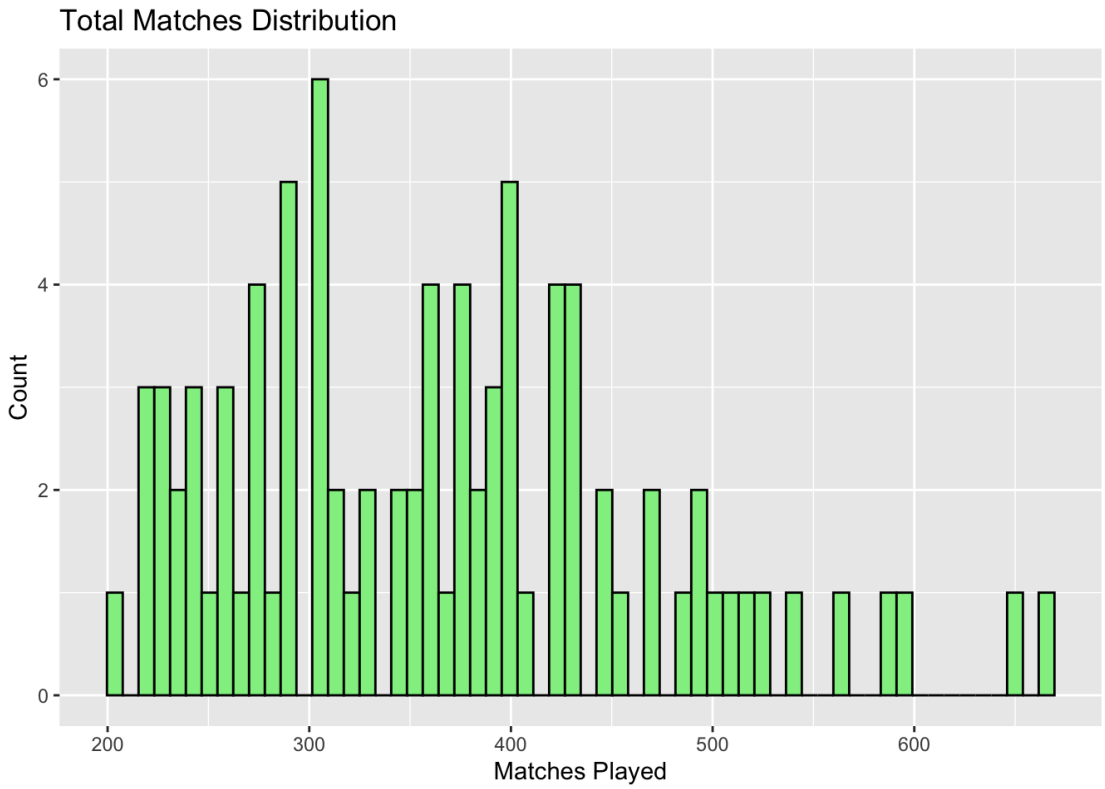
```
makedot(mycrick$runs_per_year, "black", "yellow", 45, "Runs per year Distribution", "Runs per year", "Propo
rtion")
```



This is a histogram of matches played for each batsmen in the dataset. The histogram is skewed to the right with a bimodal peaks. The median of the distribution is 360.5 matches played, while the mean is 365.43 matches played. Since the histogram is skewed the median is a better indicator of the actual midpoint of the data.

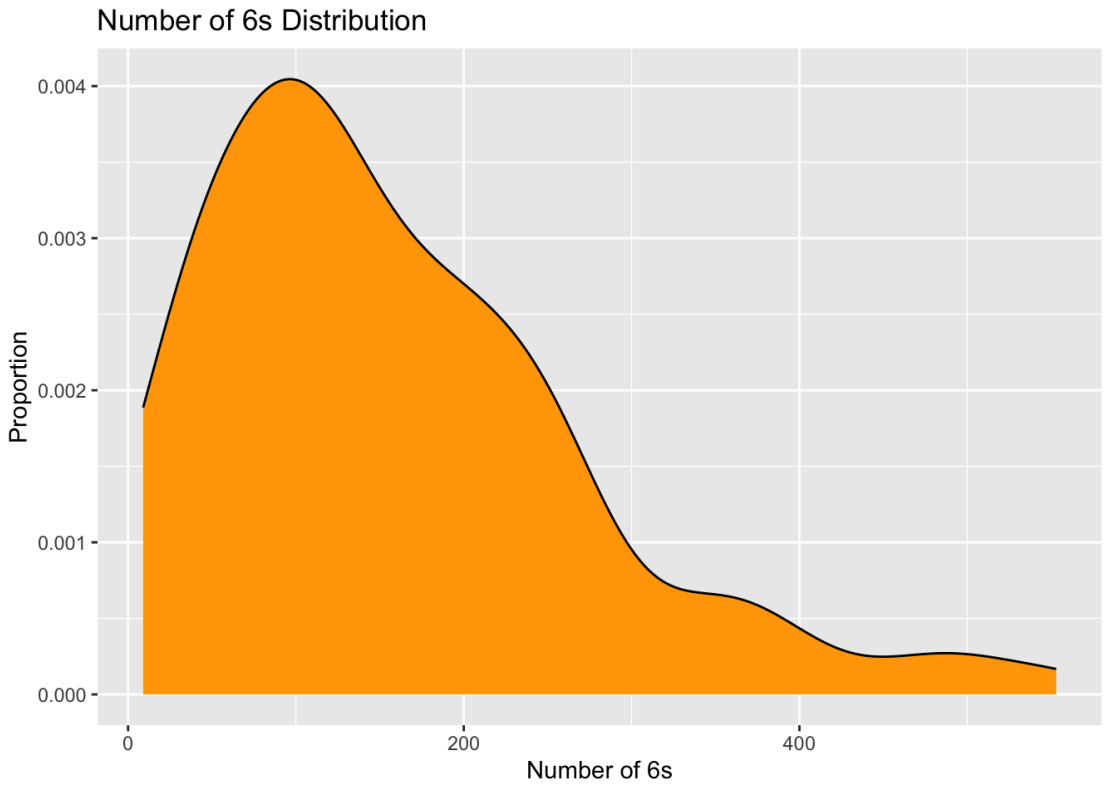
```
makehist(mycrick$Mat, "black", "lightgreen", 60, "Total Matches Distribution", "Matches Played", "Count")
```

```
## [1] "Median of Distribution:  360.5"
## [1] "Mean of Distribution  365.431818181818"
```



The last plot is another density plot on the number of 6s hit by each batsmen. In the dataset about 90 6s were hit at the highest density(peak).

```
makedense(mycrick$X6s, "black", "orange", "Number of 6s Distribution", "Number of 6s", "Proportion")
```



6.

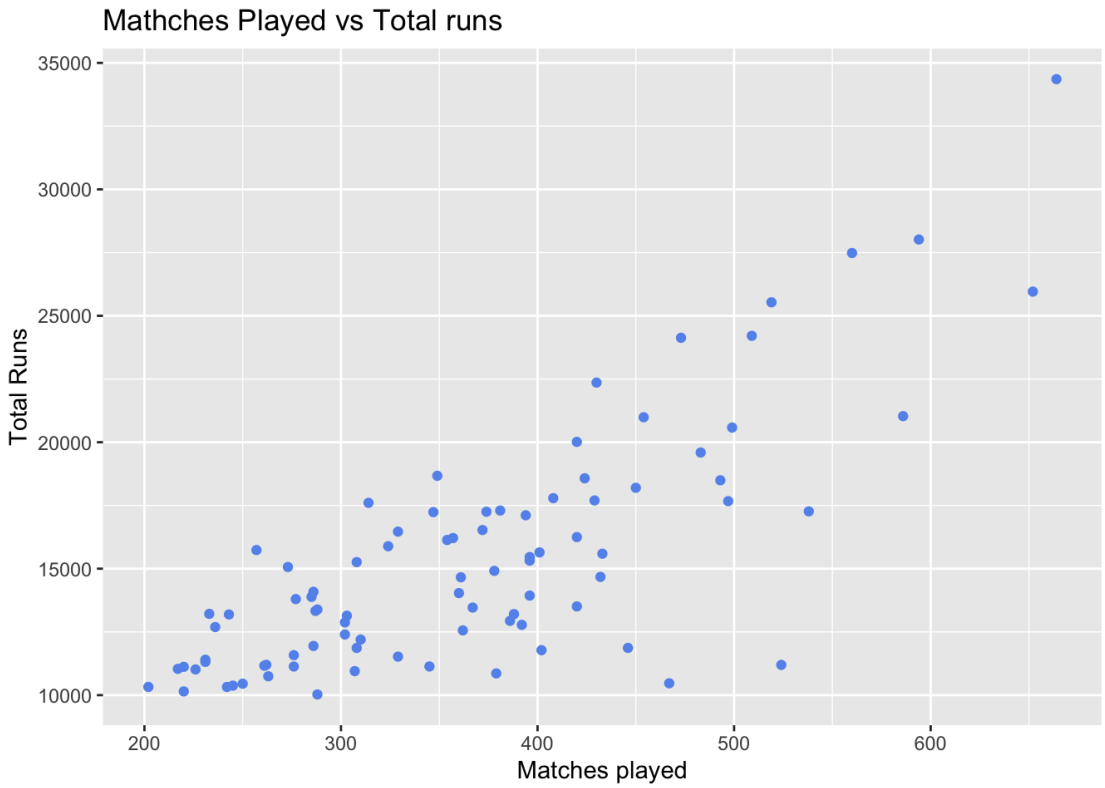
I created three new interesting variables using for loops. In order to discover the total proportion of runs that were boundaries(4s or 6s) I created three new variables. The code is below.

```
for (i in 1:88){
  mycrick$propboundaries[i] = (cricket$X6s[i]+cricket$X4s[i])/cricket$Runs[i]
}
for (i in 1:88){
  mycrick$propX4[i] = cricket$X4s[i]/cricket$Runs[i]
}

for (i in 1:88){
  mycrick$propX6[i] = cricket$X6s[i]/cricket$Runs[i]
}
```

The first scatterplot shows total matches played vs total runs scored. Visually there is a positive association between both variables. The correlation of .78 indicates that there is a fairly strong correlation between total matches played and total runs scored. Initutively that makes sense(more games played=more runs scored),

```
makescat(mycrick$Mat, mycrick$Runs, "cornflowerblue",
"Mathches Played vs Total runs", "Matches played", "Total Runs")
```

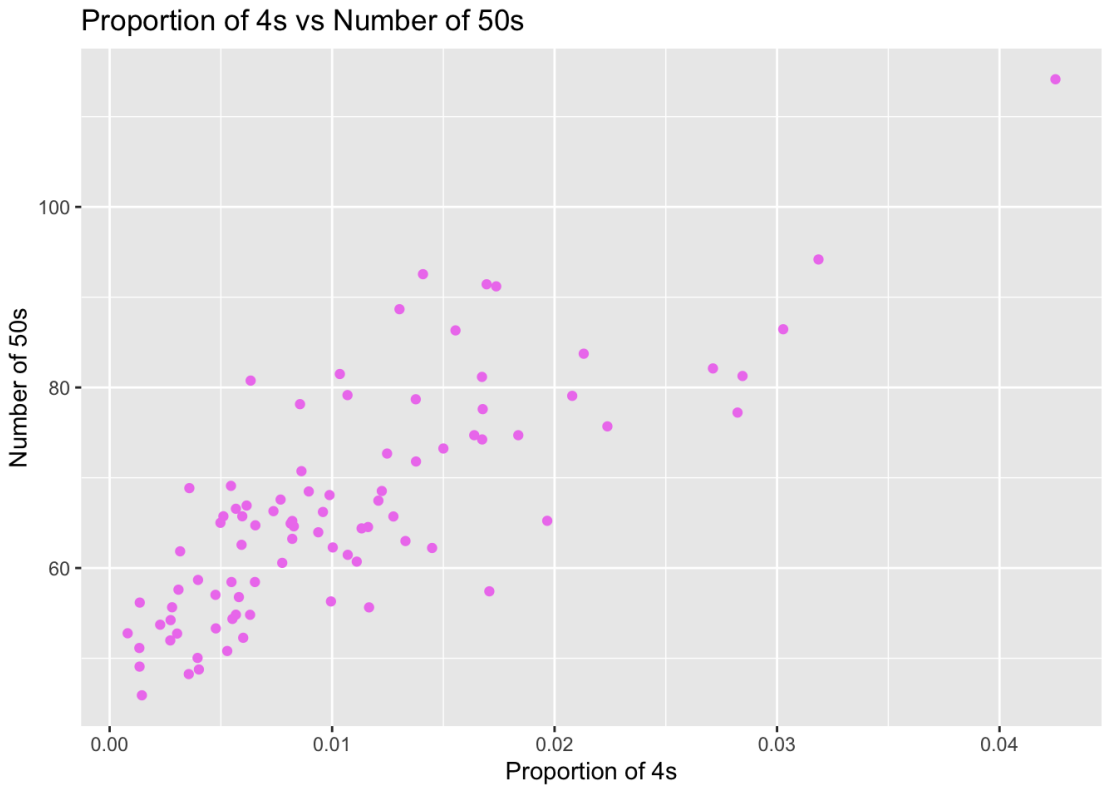
```
## [1] "Correlation is: 0.780139246123781"
```



This scatterplot shows the relationship between the proportion of total runs that were 6s and the batsmens strike rate. In cricket the more runs you score per ball the higher the strike rate. The scatterplot shows a positive association and the correlation of .79 indicates that there is a fairly strong relationship between proportion of 6s and strike rate.

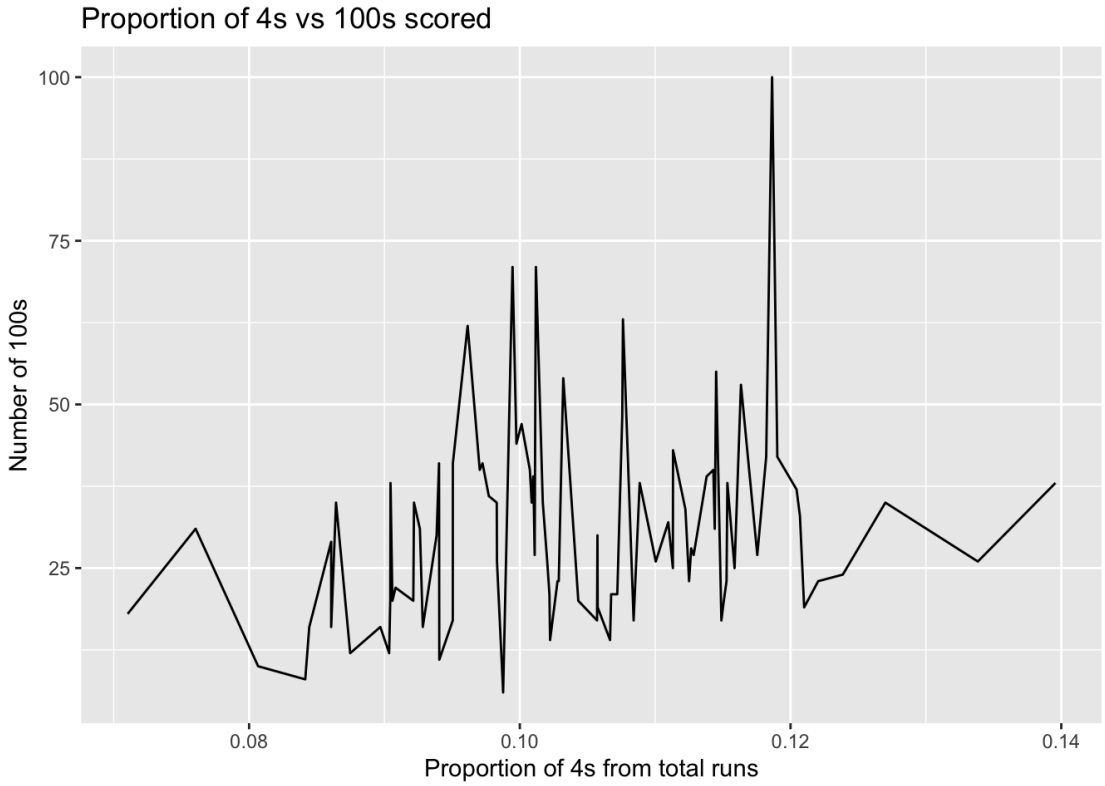
```
makescat(mycrick$propX6, cricket$SR, "violet", "Proportion of 4s vs Number of 50s", "Proportion of 4s", "Nu
mber of 50s")
```

```
## [1] "Correlation is: 0.789626548461695"
```



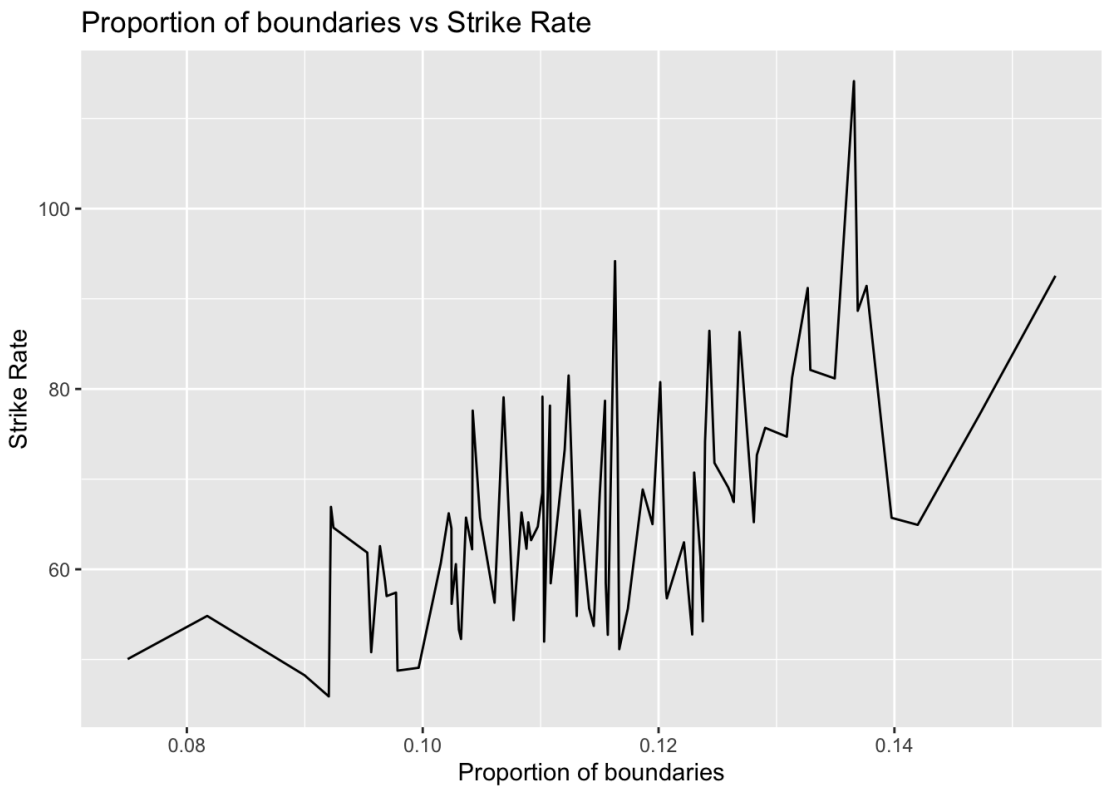
This line chart shows the realtionship between proportion of 4s from total runs and number of 100s scored. The line chart shows a zig zag pattern however there is an interesting peak at around .119 proportion of 4s. The batsmen with the most 100s have a proportion of .11 4s in their total run count. You can't conclude that that is a sweet spot, however it is an interesting peak.

```
makeline(mycrick$propX4, mycrick$X100, "Proportion of 4s vs 100s scored", "Proportion of 4s from total run
s", "Number of 100s")
```



Similarly this second line chart shows the proportion of total boundaries(4s and 6s) by total runs and strike rate. This chart also shows a zig zag inconclusive pattern but there is a very high peak at .138 proportion of 4s. Nothing is conclusive however the peak is a notable point.

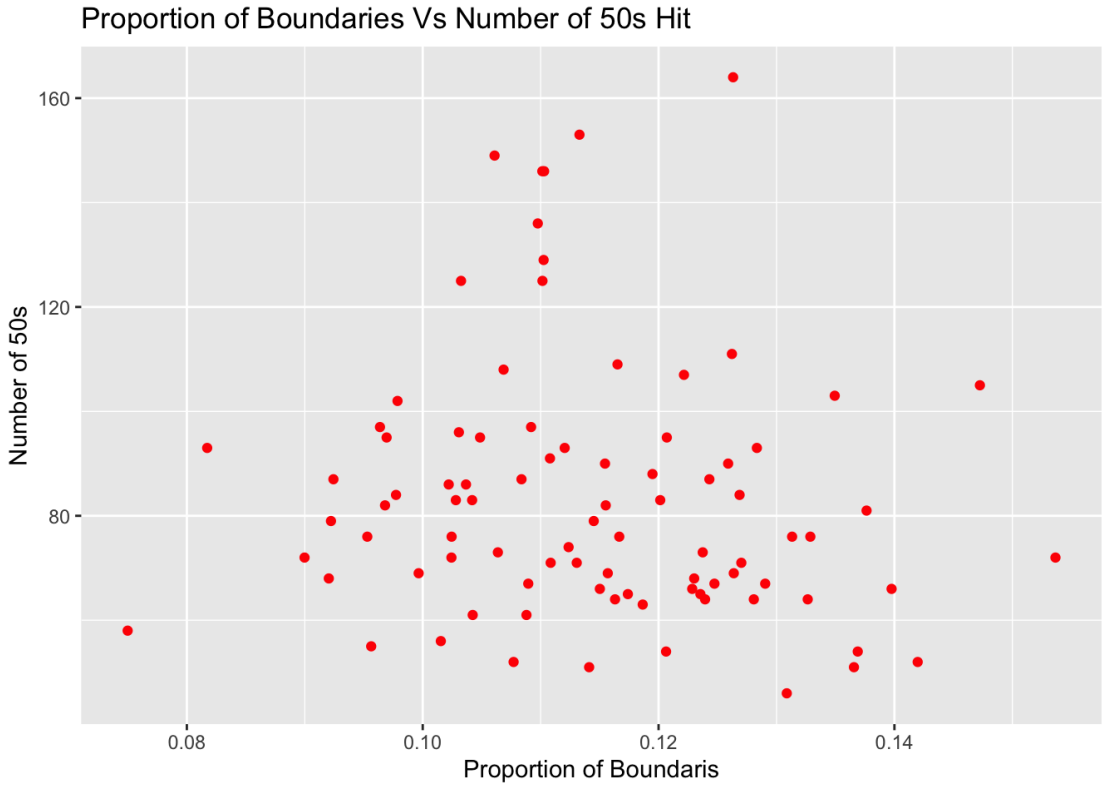
```
makeline(mycrick$propboundaries, cricket$SR, "Proportion of boundaries vs Strike Rate", "Proportion of boun
daries", "Strike Rate")
```



The last graph is the relationship between proportion of boundaries and number of 50s hit. The plot shows a scattered pattern with no visible associaiton. The correlation of -.11 tells us theres no correlation between proportion of boundaries hit and total number of 50s hit.

```
makescat(mycrick$propboundaries, mycrick$X50, "Red", "Proportion of Boundaries Vs Number of 50s Hit", "Prop
ortion of Boundaris", "Number of 50s")
```

```
## [1] "Correlation is: -0.111903977633802"
```



7.

All requirements are clearly indicated throughout the project

8.

My initial dataset had no NA values so I did not have to handle NA values throughout my experience

9.

1. I had a problem creating my for loop variables because when initializing them, the new variable would not be created. I fixed the issue when I figured out that the for loop was not iterating through the whole list just the first value.
2. I had a problem initializing my histogram function, however it was a simple fix because I kept forgetting to put an xaxis aesthetic input.

#sources: https://www.kaggle.com/datasets/dheerajmukati/most-runs-in-cricket?select=most_runs_in_cricket.csv (https://www.kaggle.com/datasets/dheerajmukati/most-runs-in-cricket?select=most_runs_in_cricket.csv)
https://www.google.com/search?q=what+does+a+density+plot+show&oq=what+does+a+density+&aqs=chrome.0.0i512l2j69i57j0i512l7.3287j1j7&sourceid=chrome&ie=UTF-8 (https://www.google.com/search?q=what+does+a+density+plot+show&oq=what+does+a+density+&aqs=chrome.0.0i512l2j69i57j0i512l7.3287j1j7&sourceid=chrome&ie=UTF-8)

12. Github Link: https://github.com/vhalaharivi26/Project2 (https://github.com/vhalaharivi26/Project2)