Module 2 Report

Exploratory Data Analysis (EDA) of Two Data Sets

ALY 6000

Professor: Maria Ayala

Analyst: Syed Faizan

Date: 1/21/2024

CONTENTS

Module 2 Report

Introduction

Module 2 Project of ALY 6000 was a two-part project in which I practiced working with different datasets. In Part One, I analyzed happiness and freedom indices from a range of countries. In Part Two, Major League Baseball provided the background for the data analysis. I wrangled datasets through a range of operations, including column and row selection; data filtering, sorting, and augmenting. I have finally summarized the data utilizing built-in descriptive statistics. After completing the R script, I included the analysis and my final conclusion in a report.

Problems and their Solutions

1. Read the data set **2015.csv** and store it in a variable called **data_2015**. You can test that you loaded it correctly with the code utilizing the head function below.

```
Country        Region Happiness.Rank Happiness.Score Standard.Error Economy..GDP.per.Capita.  Family
1 Switzerland Western Europe            1         7.587       0.03411                1.39651 1.34951
2     Iceland Western Europe            2         7.561       0.04884                1.30232 1.40223
3     Denmark Western Europe            3         7.527       0.03328                1.32548 1.36058
4      Norway Western Europe            4         7.522       0.03880                1.45900 1.33095
5      Canada  North America            5         7.427       0.03553                1.32629 1.32261
6     Finland Western Europe            6         7.406       0.03140                1.29025 1.31826
  Health..Life.Expectancy. Freedom Trust..Government.Corruption. Generosity Dystopia.Residual
1                  0.94143 0.66557                       0.41978    0.29678          2.51738
2                  0.94784 0.62877                       0.14145    0.43630          2.70201
3                  0.87464 0.64938                       0.48357    0.34139          2.49204
4                  0.88521 0.66973                       0.36503    0.34699          2.46531
5                  0.90563 0.63297                       0.32957    0.45811          2.45176
6                  0.88911 0.64169                       0.41372    0.23351          2.61955


>
```

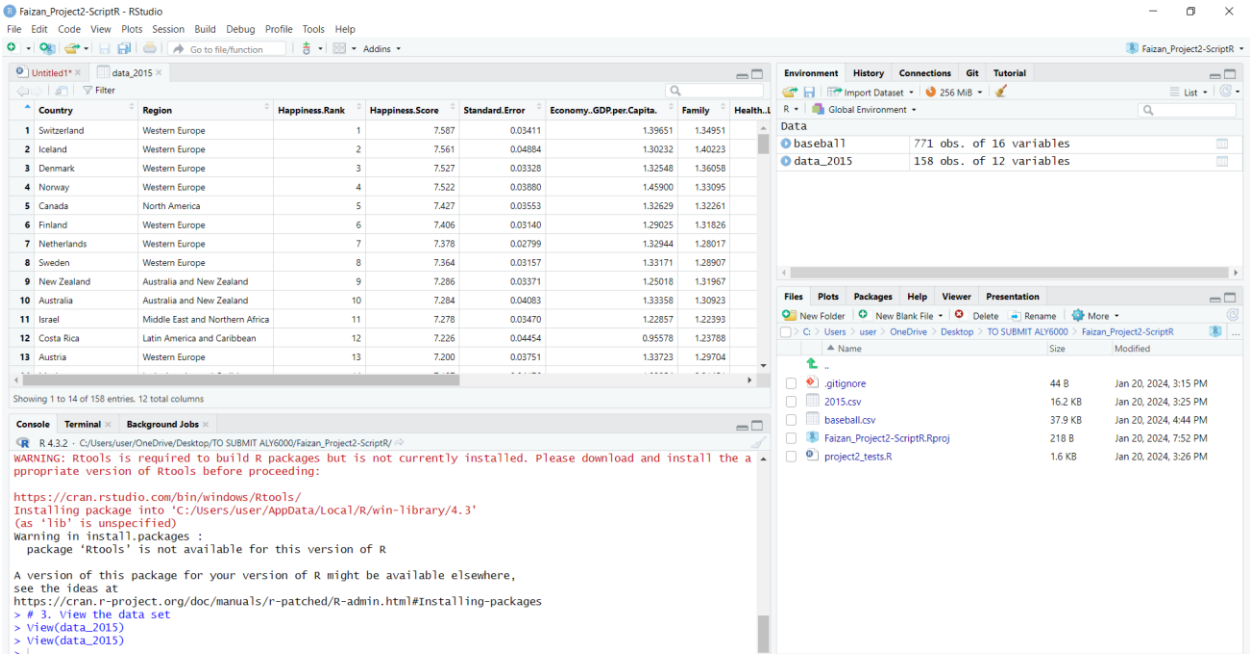
2. Use the function **names** to produce the column names for your data set.

```
# 2. Get column names
> names(data_2015)
 [1] "Country"                   "Region"                    "Happiness.Rank"
 [4] "Happiness.Score"           "Standard.Error"            "Economy..GDP.per.Capita."
 [7] "Family"                    "Health..Life.Expectancy."  "Freedom"
[10] "Trust..Government.Corruption." "Generosity"               "Dystopia.Residual"


>
```

3. Use the **view** function to view the data set in a separate tab.

The dataset opened in another tab as seen below-



4. Use the **glimpse** function to view your data set in another configuration.

```
glimpse(data_2015)
Rows: 158
Columns: 12
$ Country                   <chr> "Switzerland", "Iceland", "Denmark", "Norway", "Canada", "Finland", "…
$ Region                    <chr> "Western Europe", "Western Europe", "Western Europe", "Western Europe…
$ Happiness.Rank            <int> 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20…
$ Happiness.Score           <dbl> 7.587, 7.561, 7.527, 7.522, 7.427, 7.406, 7.378, 7.364, 7.286, 7.284,…
$ Standard.Error            <dbl> 0.03411, 0.04884, 0.03328, 0.03880, 0.03553, 0.03140, 0.02799, 0.0315…
$ Economy..GDP.per.Capita.  <dbl> 1.39651, 1.30232, 1.32548, 1.45900, 1.32629, 1.29025, 1.32944, 1.33
17…
$ Family                    <dbl> 1.34951, 1.40223, 1.36058, 1.33095, 1.32261, 1.31826, 1.28017, 1.2890…
$ Health..Life.Expectancy.  <dbl> 0.94143, 0.94784, 0.87464, 0.88521, 0.90563, 0.88911, 0.89284, 0.9108
…
$ Freedom                   <dbl> 0.66557, 0.62877, 0.64938, 0.66973, 0.63297, 0.64169, 0.61576, 0.6598…
$ Trust..Government.Corruption. <dbl> 0.41978, 0.14145, 0.48357, 0.36503, 0.32957, 0.41372, 0.31814, 0.4
384…
$ Generosity                <dbl> 0.29678, 0.43630, 0.34139, 0.34699, 0.45811, 0.23351, 0.47610, 0.3626…
$ Dystopia.Residual         <dbl> 2.51738, 2.70201, 2.49204, 2.46531, 2.45176, 2.61955, 2.46570, 2.3711…
```

5. Install and load the **janitor** package. Janitor has a function called **clean_names** that can be given a data frame to make the names more R friendly. Be sure to store the resulting converted data frame in a variable.

The following output was recorded-

```
> data_2015 <- clean_names(data_2015)
> data_2015
```

	country	region	happiness_rank	happiness_score	standard_error
1	Switzerland	Western Europe	1	7.587	0.03411
2	Iceland	Western Europe	2	7.561	0.04884
3	Denmark	Western Europe	3	7.527	0.03328
4	Norway	Western Europe	4	7.522	0.03880
5	Canada	North America	5	7.427	0.03553
6	Finland	Western Europe	6	7.406	0.03140
7	Netherlands	Western Europe	7	7.378	0.02799

8	Sweden	Western Europe	8	7.364	0.03157
9	New Zealand	Australia and New Zealand	9	7.286	0.03371
10	Australia	Australia and New Zealand	10	7.284	0.04083
11	Israel	Middle East and Northern Africa	11	7.278	0.03470
12	Costa Rica	Latin America and Caribbean	12	7.226	0.04454
13	Austria	Western Europe	13	7.200	0.03751
14	Mexico	Latin America and Caribbean	14	7.187	0.04176
15	United States	North America	15	7.119	0.03839
16	Brazil	Latin America and Caribbean	16	6.983	0.04076
17	Luxembourg	Western Europe	17	6.946	0.03499
18	Ireland	Western Europe	18	6.940	0.03676
19	Belgium	Western Europe	19	6.937	0.03595
20	United Arab Emirates	Middle East and Northern Africa	20	6.901	0.03729
21	United Kingdom	Western Europe	21	6.867	0.01866
22	Oman	Middle East and Northern Africa	22	6.853	0.05335
23	Venezuela	Latin America and Caribbean	23	6.810	0.06476
24	Singapore	Southeastern Asia	24	6.798	0.03780
25	Panama	Latin America and Caribbean	25	6.786	0.04910
26	Germany	Western Europe	26	6.750	0.01848
27	Chile	Latin America and Caribbean	27	6.670	0.05800
28	Qatar	Middle East and Northern Africa	28	6.611	0.06257
29	France	Western Europe	29	6.575	0.03512
30	Argentina	Latin America and Caribbean	30	6.574	0.04612
31	Czech Republic	Central and Eastern Europe	31	6.505	0.04168
32	Uruguay	Latin America and Caribbean	32	6.485	0.04539
33	Colombia	Latin America and Caribbean	33	6.477	0.05051
34	Thailand	Southeastern Asia	34	6.455	0.03557
35	Saudi Arabia	Middle East and Northern Africa	35	6.411	0.04633
36	Spain	Western Europe	36	6.329	0.03468
37	Malta	Western Europe	37	6.302	0.04206
38	Taiwan	Eastern Asia	38	6.298	0.03868
39	Kuwait	Middle East and Northern Africa	39	6.295	0.04456
40	Suriname	Latin America and Caribbean	40	6.269	0.09811
41	Trinidad and Tobago	Latin America and Caribbean	41	6.168	0.10895
42	El Salvador	Latin America and Caribbean	42	6.130	0.05618
43	Guatemala	Latin America and Caribbean	43	6.123	0.05224
44	Uzbekistan	Central and Eastern Europe	44	6.003	0.04361
45	Slovakia	Central and Eastern Europe	45	5.995	0.04267
46	Japan	Eastern Asia	46	5.987	0.03581
47	South Korea	Eastern Asia	47	5.984	0.04098
48	Ecuador	Latin America and Caribbean	48	5.975	0.04528
49	Bahrain	Middle East and Northern Africa	49	5.960	0.05412
50	Italy	Western Europe	50	5.948	0.03914

51	Bolivia	Latin America and Caribbean	51	5.890	0.05642
52	Moldova	Central and Eastern Europe	52	5.889	0.03799
53	Paraguay	Latin America and Caribbean	53	5.878	0.04563
54	Kazakhstan	Central and Eastern Europe	54	5.855	0.04114
55	Slovenia	Central and Eastern Europe	55	5.848	0.04251
56	Lithuania	Central and Eastern Europe	56	5.833	0.03843
57	Nicaragua	Latin America and Caribbean	57	5.828	0.05371
58	Peru	Latin America and Caribbean	58	5.824	0.04615
59	Belarus	Central and Eastern Europe	59	5.813	0.03938
60	Poland	Central and Eastern Europe	60	5.791	0.04263
61	Malaysia	Southeastern Asia	61	5.770	0.04330
62	Croatia	Central and Eastern Europe	62	5.759	0.04394
63	Libya	Middle East and Northern Africa	63	5.754	0.07832
64	Russia	Central and Eastern Europe	64	5.716	0.03135
65	Jamaica	Latin America and Caribbean	65	5.709	0.13693
66	North Cyprus	Western Europe	66	5.695	0.05635
67	Cyprus	Western Europe	67	5.689	0.05580
68	Algeria	Middle East and Northern Africa	68	5.605	0.05099
69	Kosovo	Central and Eastern Europe	69	5.589	0.05018
70	Turkmenistan	Central and Eastern Europe	70	5.548	0.04175
71	Mauritius	Sub-Saharan Africa	71	5.477	0.07197
72	Hong Kong	Eastern Asia	72	5.474	0.05051
73	Estonia	Central and Eastern Europe	73	5.429	0.04013
74	Indonesia	Southeastern Asia	74	5.399	0.02596
75	Vietnam	Southeastern Asia	75	5.360	0.03107
76	Turkey	Middle East and Northern Africa	76	5.332	0.03864
77	Kyrgyzstan	Central and Eastern Europe	77	5.286	0.03823
78	Nigeria	Sub-Saharan Africa	78	5.268	0.04192
79	Bhutan	Southern Asia	79	5.253	0.03225
80	Azerbaijan	Central and Eastern Europe	80	5.212	0.03363
81	Pakistan	Southern Asia	81	5.194	0.03726
82	Jordan	Middle East and Northern Africa	82	5.192	0.04524
83	Montenegro	Central and Eastern Europe	82	5.192	0.05235

	economy_gdp_per_capita	family	health_life_expectancy	freedom	trust_government_corruption	generosity
1	1.39651	1.34951	0.94143	0.66557	0.41978	0.29678
2	1.30232	1.40223	0.94784	0.62877	0.14145	0.43630
3	1.32548	1.36058	0.87464	0.64938	0.48357	0.34139
4	1.45900	1.33095	0.88521	0.66973	0.36503	0.34699
5	1.32629	1.32261	0.90563	0.63297	0.32957	0.45811
6	1.29025	1.31826	0.88911	0.64169	0.41372	0.23351
7	1.32944	1.28017	0.89284	0.61576	0.31814	0.47610
8	1.33171	1.28907	0.91087	0.65980	0.43844	0.36262

9	1.25018 1.31967	0.90837 0.63938	0.42922	0.47501
10	1.33358 1.30923	0.93156 0.65124	0.35637	0.43562
11	1.22857 1.22393	0.91387 0.41319	0.07785	0.33172
12	0.95578 1.23788	0.86027 0.63376	0.10583	0.25497
13	1.33723 1.29704	0.89042 0.62433	0.18676	0.33088
14	1.02054 0.91451	0.81444 0.48181	0.21312	0.14074
15	1.39451 1.24711	0.86179 0.54604	0.15890	0.40105
16	0.98124 1.23287	0.69702 0.49049	0.17521	0.14574
17	1.56391 1.21963	0.91894 0.61583	0.37798	0.28034
18	1.33596 1.36948	0.89533 0.61777	0.28703	0.45901
19	1.30782 1.28566	0.89667 0.58450	0.22540	0.22250
20	1.42727 1.12575	0.80925 0.64157	0.38583	0.26428
21	1.26637 1.28548	0.90943 0.59625	0.32067	0.51912
22	1.36011 1.08182	0.76276 0.63274	0.32524	0.21542
23	1.04424 1.25596	0.72052 0.42908	0.11069	0.05841
24	1.52186 1.02000	1.02525 0.54252	0.49210	0.31105
25	1.06353 1.19850	0.79661 0.54210	0.09270	0.24434
26	1.32792 1.29937	0.89186 0.61477	0.21843	0.28214
27	1.10715 1.12447	0.85857 0.44132	0.12869	0.33363
28	1.69042 1.07860	0.79733 0.64040	0.52208	0.32573
29	1.27778 1.26038	0.94579 0.55011	0.20646	0.12332
30	1.05351 1.24823	0.78723 0.44974	0.08484	0.11451
31	1.17898 1.20643	0.84483 0.46364	0.02652	0.10686
32	1.06166 1.20890	0.81160 0.60362	0.24558	0.23240
33	0.91861 1.24018	0.69077 0.53466	0.05120	0.18401
34	0.96690 1.26504	0.73850 0.55664	0.03187	0.57630
35	1.39541 1.08393	0.72025 0.31048	0.32524	0.13706
36	1.23011 1.31379	0.95562 0.45951	0.06398	0.18227
37	1.20740 1.30203	0.88721 0.60365	0.13586	0.51752
38	1.29098 1.07617	0.87530 0.39740	0.08129	0.25376
39	1.55422 1.16594	0.72492 0.55499	0.25609	0.16228
40	0.99534 0.97200	0.60820 0.59657	0.13633	0.16991
41	1.21183 1.18354	0.61483 0.55884	0.01140	0.31844
42	0.76454 1.02507	0.67737 0.40350	0.11776	0.10692
43	0.74553 1.04356	0.64425 0.57733	0.09472	0.27489
44	0.63244 1.34043	0.59772 0.65821	0.30826	0.22837
45	1.16891 1.26999	0.78902 0.31751	0.03431	0.16893
46	1.27074 1.25712	0.99111 0.49615	0.18060	0.10705
47	1.24461 0.95774	0.96538 0.33208	0.07857	0.18557
48	0.86402 0.99903	0.79075 0.48574	0.18090	0.11541
49	1.32376 1.21624	0.74716 0.45492	0.30600	0.17362
50	1.25114 1.19777	0.95446 0.26236	0.02901	0.22823
51	0.68133 0.97841	0.53920 0.57414	0.08800	0.20536

52	0.59448	1.01528	0.61826	0.32818	0.01615	0.20951
53	0.75985	1.30477	0.66098	0.53899	0.08242	0.34240
54	1.12254	1.12241	0.64368	0.51649	0.08454	0.11827
55	1.18498	1.27385	0.87337	0.60855	0.03787	0.25328
56	1.14723	1.25745	0.73128	0.21342	0.01031	0.02641
57	0.59325	1.14184	0.74314	0.55475	0.19317	0.27815
58	0.90019	0.97459	0.73017	0.41496	0.05989	0.14982
59	1.03192	1.23289	0.73608	0.37938	0.19090	0.11046
60	1.12555	1.27948	0.77903	0.53122	0.04212	0.16759
61	1.12486	1.07023	0.72394	0.53024	0.10501	0.33075
62	1.08254	0.79624	0.78805	0.25883	0.02430	0.05444
63	1.13145	1.11862	0.70380	0.41668	0.11023	0.18295
64	1.13764	1.23617	0.66926	0.36679	0.03005	0.00199
65	0.81038	1.15102	0.68741	0.50442	0.02299	0.21230
66	1.20806	1.07008	0.92356	0.49027	0.14280	0.26169
67	1.20813	0.89318	0.92356	0.40672	0.06146	0.30638
68	0.93929	1.07772	0.61766	0.28579	0.17383	0.07822
69	0.80148	0.81198	0.63132	0.24749	0.04741	0.28310
70	0.95847	1.22668	0.53886	0.47610	0.30844	0.16979
71	1.00761	0.98521	0.70950	0.56066	0.07521	0.37744
72	1.38604	1.05818	1.01328	0.59608	0.37124	0.39478
73	1.15174	1.22791	0.77361	0.44888	0.15184	0.08680
74	0.82827	1.08708	0.63793	0.46611	0.00000	0.51535
75	0.63216	0.91226	0.74676	0.59444	0.10441	0.16860
76	1.06098	0.94632	0.73172	0.22815	0.15746	0.12253
77	0.47428	1.15115	0.65088	0.43477	0.04232	0.30030
78	0.65435	0.90432	0.16007	0.34334	0.04030	0.27233
79	0.77042	1.10395	0.57407	0.53206	0.15445	0.47998
80	1.02389	0.93793	0.64045	0.37030	0.16065	0.07799
81	0.59543	0.41411	0.51466	0.12102	0.10464	0.33671
82	0.90198	1.05392	0.69639	0.40661	0.14293	0.11053
83	0.97438	0.90557	0.72521	0.18260	0.14296	0.16140

	dystopia_residual
1	2.51738
2	2.70201
3	2.49204
4	2.46531
5	2.45176
6	2.61955
7	2.46570
8	2.37119
9	2.26425
10	2.26646

11	3.08854
12	3.17728
13	2.53320
14	3.60214
15	2.51011
16	3.26001
17	1.96961
18	1.97570
19	2.41484
20	2.24743
21	1.96994
22	2.47489
23	3.19131
24	1.88501
25	2.84848
26	2.11569
27	2.67585
28	1.55674
29	2.21126
30	2.83600
31	2.67782
32	2.32142
33	2.85737
34	2.31945
35	2.43872
36	2.12367
37	1.64880
38	2.32323
39	1.87634
40	2.79094
41	2.26882
42	3.03500
43	2.74255
44	2.23741
45	2.24639
46	1.68435
47	2.21978
48	2.53942
49	1.73797
50	2.02518
51	2.82334
52	3.10712
53	2.18896

54	2.24729
55	1.61583
56	2.44649
57	2.32407
58	2.59450
59	2.13090
60	1.86565
61	1.88541
62	2.75414
63	2.09066
64	2.27394
65	2.32038
66	1.59888
67	1.88931
68	2.43209
69	2.76579
70	1.86984
71	1.76145
72	0.65429
73	1.58782
74	1.86399
75	2.20173
76	2.08528
77	2.23270
78	2.89319
79	1.63794
80	2.00073
81	3.10709
82	1.87996
83	2.10017

[reached 'max' / getOption("max.print") -- omitted 75 rows]

6. Select from the data set the **country**, **region**, **happiness_score**, and **freedom columns**. St ore this new table as **happy_df**.

The following new table was the output-

```
> happy_df <- select(data_2015, country, region, happiness_score, freedom)
> happy_df
                country                       region happiness_score freedom
1           Switzerland               Western Europe           7.587 0.66557
2               Iceland               Western Europe           7.561 0.62877
3               Denmark               Western Europe           7.527 0.64938
4                Norway               Western Europe           7.522 0.66973
5                Canada               North America            7.427 0.63297
6               Finland               Western Europe           7.406 0.64169
7           Netherlands               Western Europe           7.378 0.61576
8                Sweden               Western Europe           7.364 0.65980
9           New Zealand      Australia and New Zealand           7.286 0.63938
10            Australia      Australia and New Zealand           7.284 0.65124
11               Israel Middle East and Northern Africa           7.278 0.41319
12           Costa Rica     Latin America and Caribbean           7.226 0.63376
13              Austria               Western Europe           7.200 0.62433
14               Mexico     Latin America and Caribbean           7.187 0.48181
15        United States               North America            7.119 0.54604
16               Brazil     Latin America and Caribbean           6.983 0.49049
17           Luxembourg               Western Europe           6.946 0.61583
18              Ireland               Western Europe           6.940 0.61777
19              Belgium               Western Europe           6.937 0.58450
20 United Arab Emirates Middle East and Northern Africa           6.901 0.64157
21       United Kingdom               Western Europe           6.867 0.59625
22                 Oman Middle East and Northern Africa           6.853 0.63274
23            Venezuela     Latin America and Caribbean           6.810 0.42908
24            Singapore               Southeastern Asia        6.798 0.54252
25               Panama     Latin America and Caribbean           6.786 0.54210
26              Germany               Western Europe           6.750 0.61477
27                Chile     Latin America and Caribbean           6.670 0.44132
28                Qatar Middle East and Northern Africa           6.611 0.64040
29               France               Western Europe           6.575 0.55011
30            Argentina     Latin America and Caribbean           6.574 0.44974
31       Czech Republic      Central and Eastern Europe           6.505 0.46364
32              Uruguay     Latin America and Caribbean           6.485 0.60362
33             Colombia     Latin America and Caribbean           6.477 0.53466
34             Thailand               Southeastern Asia        6.455 0.55664
35         Saudi Arabia Middle East and Northern Africa           6.411 0.31048
36                Spain               Western Europe           6.329 0.45951
37                Malta               Western Europe           6.302 0.60365
38               Taiwan                Eastern Asia            6.298 0.39740
39               Kuwait Middle East and Northern Africa           6.295 0.55499
40             Suriname     Latin America and Caribbean           6.269 0.59657
41  Trinidad and Tobago     Latin America and Caribbean           6.168 0.55884
42          El Salvador     Latin America and Caribbean           6.130 0.40350
43            Guatemala     Latin America and Caribbean           6.123 0.57733
```

44	Uzbekistan	Central and Eastern Europe	6.003	0.65821
45	Slovakia	Central and Eastern Europe	5.995	0.31751
46	Japan	Eastern Asia	5.987	0.49615
47	South Korea	Eastern Asia	5.984	0.33208
48	Ecuador	Latin America and Caribbean	5.975	0.48574
49	Bahrain	Middle East and Northern Africa	5.960	0.45492
50	Italy	Western Europe	5.948	0.26236
51	Bolivia	Latin America and Caribbean	5.890	0.57414
52	Moldova	Central and Eastern Europe	5.889	0.32818
53	Paraguay	Latin America and Caribbean	5.878	0.53899
54	Kazakhstan	Central and Eastern Europe	5.855	0.51649
55	Slovenia	Central and Eastern Europe	5.848	0.60855
56	Lithuania	Central and Eastern Europe	5.833	0.21342
57	Nicaragua	Latin America and Caribbean	5.828	0.55475
58	Peru	Latin America and Caribbean	5.824	0.41496
59	Belarus	Central and Eastern Europe	5.813	0.37938
60	Poland	Central and Eastern Europe	5.791	0.53122
61	Malaysia	Southeastern Asia	5.770	0.53024
62	Croatia	Central and Eastern Europe	5.759	0.25883
63	Libya	Middle East and Northern Africa	5.754	0.41668
64	Russia	Central and Eastern Europe	5.716	0.36679
65	Jamaica	Latin America and Caribbean	5.709	0.50442
66	North Cyprus	Western Europe	5.695	0.49027
67	Cyprus	Western Europe	5.689	0.40672
68	Algeria	Middle East and Northern Africa	5.605	0.28579
69	Kosovo	Central and Eastern Europe	5.589	0.24749
70	Turkmenistan	Central and Eastern Europe	5.548	0.47610
71	Mauritius	Sub-Saharan Africa	5.477	0.56066
72	Hong Kong	Eastern Asia	5.474	0.59608
73	Estonia	Central and Eastern Europe	5.429	0.44888
74	Indonesia	Southeastern Asia	5.399	0.46611
75	Vietnam	Southeastern Asia	5.360	0.59444
76	Turkey	Middle East and Northern Africa	5.332	0.22815
77	Kyrgyzstan	Central and Eastern Europe	5.286	0.43477
78	Nigeria	Sub-Saharan Africa	5.268	0.34334
79	Bhutan	Southern Asia	5.253	0.53206
80	Azerbaijan	Central and Eastern Europe	5.212	0.37030
81	Pakistan	Southern Asia	5.194	0.12102
82	Jordan	Middle East and Northern Africa	5.192	0.40661
83	Montenegro	Central and Eastern Europe	5.192	0.18260
84	China	Eastern Asia	5.140	0.51697
85	Zambia	Sub-Saharan Africa	5.129	0.48827
86	Romania	Central and Eastern Europe	5.124	0.35068
87	Serbia	Central and Eastern Europe	5.123	0.20107
88	Portugal	Western Europe	5.102	0.51469
89	Latvia	Central and Eastern Europe	5.098	0.29671

90	Philippines	Southeastern Asia	5.073	0.62545
91	Somaliland region	Sub-Saharan Africa	5.057	0.46582
92	Morocco	Middle East and Northern Africa	5.013	0.41691
93	Macedonia	Central and Eastern Europe	5.007	0.33457
94	Mozambique	Sub-Saharan Africa	4.971	0.34037
95	Albania	Central and Eastern Europe	4.959	0.35733
96	Bosnia and Herzegovina	Central and Eastern Europe	4.949	0.09245
97	Lesotho	Sub-Saharan Africa	4.898	0.31767
98	Dominican Republic	Latin America and Caribbean	4.885	0.57672
99	Laos	Southeastern Asia	4.876	0.59591
100	Mongolia	Eastern Asia	4.874	0.43626
101	Swaziland	Sub-Saharan Africa	4.867	0.30658
102	Greece	Western Europe	4.857	0.07699
103	Lebanon	Middle East and Northern Africa	4.839	0.33916
104	Hungary	Central and Eastern Europe	4.800	0.32112
105	Honduras	Latin America and Caribbean	4.788	0.40148
106	Tajikistan	Central and Eastern Europe	4.786	0.47216
107	Tunisia	Middle East and Northern Africa	4.739	0.26268
108	Palestinian Territories	Middle East and Northern Africa	4.715	0.24499
109	Bangladesh	Southern Asia	4.694	0.40820
110	Iran	Middle East and Northern Africa	4.686	0.30033
111	Ukraine	Central and Eastern Europe	4.681	0.25123
112	Iraq	Middle East and Northern Africa	4.677	0.00000
113	South Africa	Sub-Saharan Africa	4.642	0.33207
114	Ghana	Sub-Saharan Africa	4.633	0.42342
115	Zimbabwe	Sub-Saharan Africa	4.610	0.25861
116	Liberia	Sub-Saharan Africa	4.571	0.28531
117	India	Southern Asia	4.565	0.39786
118	Sudan	Sub-Saharan Africa	4.550	0.10081
119	Haiti	Latin America and Caribbean	4.518	0.24425
120	Congo (Kinshasa)	Sub-Saharan Africa	4.517	0.22605
121	Nepal	Southern Asia	4.514	0.38282
122	Ethiopia	Sub-Saharan Africa	4.512	0.43450
123	Sierra Leone	Sub-Saharan Africa	4.507	0.40840
124	Mauritania	Sub-Saharan Africa	4.436	0.24232
125	Kenya	Sub-Saharan Africa	4.419	0.42215
126	Djibouti	Sub-Saharan Africa	4.369	0.46074
127	Armenia	Central and Eastern Europe	4.350	0.19847
128	Botswana	Sub-Saharan Africa	4.332	0.49495
129	Myanmar	Southeastern Asia	4.307	0.44017
130	Georgia	Central and Eastern Europe	4.297	0.40577
131	Malawi	Sub-Saharan Africa	4.292	0.43054
132	Sri Lanka	Southern Asia	4.271	0.53726
133	Cameroon	Sub-Saharan Africa	4.252	0.49309
134	Bulgaria	Central and Eastern Europe	4.218	0.30587
135	Egypt	Middle East and Northern Africa	4.194	0.17288

136	Yemen	Middle East and Northern Africa	4.077	0.35571
137	Angola	Sub-Saharan Africa	4.033	0.10384
138	Mali	Sub-Saharan Africa	3.995	0.38857
139	Congo (Brazzaville)	Sub-Saharan Africa	3.989	0.41466
140	Comoros	Sub-Saharan Africa	3.956	0.22917
141	Uganda	Sub-Saharan Africa	3.931	0.45727
142	Senegal	Sub-Saharan Africa	3.904	0.36772
143	Gabon	Sub-Saharan Africa	3.896	0.31914
144	Niger	Sub-Saharan Africa	3.845	0.47692
145	Cambodia	Southeastern Asia	3.819	0.66246
146	Tanzania	Sub-Saharan Africa	3.781	0.32878
147	Madagascar	Sub-Saharan Africa	3.681	0.19184
148	Central African Republic	Sub-Saharan Africa	3.678	0.48879
149	Chad	Sub-Saharan Africa	3.667	0.23501
150	Guinea	Sub-Saharan Africa	3.656	0.37725
151	Ivory Coast	Sub-Saharan Africa	3.655	0.46866
152	Burkina Faso	Sub-Saharan Africa	3.587	0.39493
153	Afghanistan	Southern Asia	3.575	0.23414
154	Rwanda	Sub-Saharan Africa	3.465	0.59201
155	Benin	Sub-Saharan Africa	3.340	0.48450
156	Syria	Middle East and Northern Africa	3.006	0.15684
157	Burundi	Sub-Saharan Africa	2.905	0.11850
158	Togo	Sub-Saharan Africa	2.839	0.36453

>

7. Slice the first 10 rows from **happy_df** and store it as **top_ten_df**.

```
country                region happiness_score freedom
1  Switzerland          Western Europe          7.587 0.66557
2      Iceland          Western Europe          7.561 0.62877
3      Denmark          Western Europe          7.527 0.64938
4       Norway          Western Europe          7.522 0.66973
5       Canada          North America           7.427 0.63297
6      Finland          Western Europe          7.406 0.64169
7  Netherlands          Western Europe          7.378 0.61576
8       Sweden          Western Europe          7.364 0.65980
9  New Zealand Australia and New Zealand          7.286 0.63938
10   Australia Australia and New Zealand          7.284 0.65124
```

8.From **happy_df** filter the table for freedom values under 0.20. Store this new table as **no_freedom_df.**

```
country                    region happiness_score freedom
1             Pakistan                Southern Asia          5.194 0.12102
2           Montenegro      Central and Eastern Europe          5.192 0.18260
3  Bosnia and Herzegovina      Central and Eastern Europe          4.949 0.09245
4               Greece                Western Europe          4.857 0.07699
5                 Iraq Middle East and Northern Africa          4.677 0.00000
6                Sudan           Sub-Saharan Africa          4.550 0.10081
7              Armenia      Central and Eastern Europe          4.350 0.19847
8                Egypt Middle East and Northern Africa          4.194 0.17288
9               Angola           Sub-Saharan Africa          4.033 0.10384
10          Madagascar           Sub-Saharan Africa          3.681 0.19184
11            Syria Middle East and Northern Africa          3.006 0.15684
12             Burundi           Sub-Saharan Africa          2.905 0.11850
```

```
>
```

9. Arrange the values in **happy_df** in descending order by their freedom values. Store this new table as **best_freedom_df**.

```
> best_freedom_df <- arrange(happy_df, desc(freedom))
```

```
> best_freedom_df
                country                      region happiness_score freedom
1                Norway              Western Europe           7.522 0.66973
2           Switzerland              Western Europe           7.587 0.66557
3              Cambodia           Southeastern Asia           3.819 0.66246
4                Sweden              Western Europe           7.364 0.65980
5            Uzbekistan  Central and Eastern Europe           6.003 0.65821
6             Australia  Australia and New Zealand           7.284 0.65124
7               Denmark              Western Europe           7.527 0.64938
8               Finland              Western Europe           7.406 0.64169
9  United Arab Emirates Middle East and Northern Africa           6.901 0.64157
10               Qatar Middle East and Northern Africa           6.611 0.64040
11         New Zealand  Australia and New Zealand           7.286 0.63938
12          Costa Rica Latin America and Caribbean           7.226 0.63376
13              Canada              North America           7.427 0.63297
14                Oman Middle East and Northern Africa           6.853 0.63274
15             Iceland              Western Europe           7.561 0.62877
16         Philippines           Southeastern Asia           5.073 0.62545
17             Austria              Western Europe           7.200 0.62433
18             Ireland              Western Europe           6.940 0.61777
19          Luxembourg              Western Europe           6.946 0.61583
20         Netherlands              Western Europe           7.378 0.61576
21             Germany              Western Europe           6.750 0.61477
22            Slovenia  Central and Eastern Europe           5.848 0.60855
23               Malta              Western Europe           6.302 0.60365
24             Uruguay Latin America and Caribbean           6.485 0.60362
25            Suriname Latin America and Caribbean           6.269 0.59657
26      United Kingdom              Western Europe           6.867 0.59625
27           Hong Kong                Eastern Asia           5.474 0.59608
28                Laos           Southeastern Asia           4.876 0.59591
29             Vietnam           Southeastern Asia           5.360 0.59444
30              Rwanda         Sub-Saharan Africa           3.465 0.59201
31             Belgium              Western Europe           6.937 0.58450
32           Guatemala Latin America and Caribbean           6.123 0.57733
33  Dominican Republic Latin America and Caribbean           4.885 0.57672
34             Bolivia Latin America and Caribbean           5.890 0.57414
35           Mauritius         Sub-Saharan Africa           5.477 0.56066
36 Trinidad and Tobago Latin America and Caribbean           6.168 0.55884
37            Thailand           Southeastern Asia           6.455 0.55664
38              Kuwait Middle East and Northern Africa           6.295 0.55499
39           Nicaragua Latin America and Caribbean           5.828 0.55475
40              France              Western Europe           6.575 0.55011
41       United States              North America           7.119 0.54604
42           Singapore           Southeastern Asia           6.798 0.54252
43              Panama Latin America and Caribbean           6.786 0.54210
44            Paraguay Latin America and Caribbean           5.878 0.53899
```

45	Sri Lanka	Southern Asia	4.271	0.53726
46	Colombia	Latin America and Caribbean	6.477	0.53466
47	Bhutan	Southern Asia	5.253	0.53206
48	Poland	Central and Eastern Europe	5.791	0.53122
49	Malaysia	Southeastern Asia	5.770	0.53024
50	China	Eastern Asia	5.140	0.51697
51	Kazakhstan	Central and Eastern Europe	5.855	0.51649
52	Portugal	Western Europe	5.102	0.51469
53	Jamaica	Latin America and Caribbean	5.709	0.50442
54	Japan	Eastern Asia	5.987	0.49615
55	Botswana	Sub-Saharan Africa	4.332	0.49495
56	Cameroon	Sub-Saharan Africa	4.252	0.49309
57	Brazil	Latin America and Caribbean	6.983	0.49049
58	North Cyprus	Western Europe	5.695	0.49027
59	Central African Republic	Sub-Saharan Africa	3.678	0.48879
60	Zambia	Sub-Saharan Africa	5.129	0.48827
61	Ecuador	Latin America and Caribbean	5.975	0.48574
62	Benin	Sub-Saharan Africa	3.340	0.48450
63	Mexico	Latin America and Caribbean	7.187	0.48181
64	Niger	Sub-Saharan Africa	3.845	0.47692
65	Turkmenistan	Central and Eastern Europe	5.548	0.47610
66	Tajikistan	Central and Eastern Europe	4.786	0.47216
67	Ivory Coast	Sub-Saharan Africa	3.655	0.46866
68	Indonesia	Southeastern Asia	5.399	0.46611
69	Somaliland region	Sub-Saharan Africa	5.057	0.46582
70	Czech Republic	Central and Eastern Europe	6.505	0.46364
71	Djibouti	Sub-Saharan Africa	4.369	0.46074
72	Spain	Western Europe	6.329	0.45951
73	Uganda	Sub-Saharan Africa	3.931	0.45727
74	Bahrain	Middle East and Northern Africa	5.960	0.45492
75	Argentina	Latin America and Caribbean	6.574	0.44974
76	Estonia	Central and Eastern Europe	5.429	0.44888
77	Chile	Latin America and Caribbean	6.670	0.44132
78	Myanmar	Southeastern Asia	4.307	0.44017
79	Mongolia	Eastern Asia	4.874	0.43626
80	Kyrgyzstan	Central and Eastern Europe	5.286	0.43477
81	Ethiopia	Sub-Saharan Africa	4.512	0.43450
82	Malawi	Sub-Saharan Africa	4.292	0.43054
83	Venezuela	Latin America and Caribbean	6.810	0.42908
84	Ghana	Sub-Saharan Africa	4.633	0.42342
85	Kenya	Sub-Saharan Africa	4.419	0.42215
86	Morocco	Middle East and Northern Africa	5.013	0.41691
87	Libya	Middle East and Northern Africa	5.754	0.41668
88	Peru	Latin America and Caribbean	5.824	0.41496
89	Congo (Brazzaville)	Sub-Saharan Africa	3.989	0.41466
90	Israel	Middle East and Northern Africa	7.278	0.41319

91	Sierra Leone	Sub-Saharan Africa	4.507	0.40840
92	Bangladesh	Southern Asia	4.694	0.40820
93	Cyprus	Western Europe	5.689	0.40672
94	Jordan	Middle East and Northern Africa	5.192	0.40661
95	Georgia	Central and Eastern Europe	4.297	0.40577
96	El Salvador	Latin America and Caribbean	6.130	0.40350
97	Honduras	Latin America and Caribbean	4.788	0.40148
98	India	Southern Asia	4.565	0.39786
99	Taiwan	Eastern Asia	6.298	0.39740
100	Burkina Faso	Sub-Saharan Africa	3.587	0.39493
101	Mali	Sub-Saharan Africa	3.995	0.38857
102	Nepal	Southern Asia	4.514	0.38282
103	Belarus	Central and Eastern Europe	5.813	0.37938
104	Guinea	Sub-Saharan Africa	3.656	0.37725
105	Azerbaijan	Central and Eastern Europe	5.212	0.37030
106	Senegal	Sub-Saharan Africa	3.904	0.36772
107	Russia	Central and Eastern Europe	5.716	0.36679
108	Togo	Sub-Saharan Africa	2.839	0.36453
109	Albania	Central and Eastern Europe	4.959	0.35733
110	Yemen	Middle East and Northern Africa	4.077	0.35571
111	Romania	Central and Eastern Europe	5.124	0.35068
112	Nigeria	Sub-Saharan Africa	5.268	0.34334
113	Mozambique	Sub-Saharan Africa	4.971	0.34037
114	Lebanon	Middle East and Northern Africa	4.839	0.33916
115	Macedonia	Central and Eastern Europe	5.007	0.33457
116	South Korea	Eastern Asia	5.984	0.33208
117	South Africa	Sub-Saharan Africa	4.642	0.33207
118	Tanzania	Sub-Saharan Africa	3.781	0.32878
119	Moldova	Central and Eastern Europe	5.889	0.32818
120	Hungary	Central and Eastern Europe	4.800	0.32112
121	Gabon	Sub-Saharan Africa	3.896	0.31914
122	Lesotho	Sub-Saharan Africa	4.898	0.31767
123	Slovakia	Central and Eastern Europe	5.995	0.31751
124	Saudi Arabia	Middle East and Northern Africa	6.411	0.31048
125	Swaziland	Sub-Saharan Africa	4.867	0.30658
126	Bulgaria	Central and Eastern Europe	4.218	0.30587
127	Iran	Middle East and Northern Africa	4.686	0.30033
128	Latvia	Central and Eastern Europe	5.098	0.29671
129	Algeria	Middle East and Northern Africa	5.605	0.28579
130	Liberia	Sub-Saharan Africa	4.571	0.28531
131	Tunisia	Middle East and Northern Africa	4.739	0.26268
132	Italy	Western Europe	5.948	0.26236
133	Croatia	Central and Eastern Europe	5.759	0.25883
134	Zimbabwe	Sub-Saharan Africa	4.610	0.25861
135	Ukraine	Central and Eastern Europe	4.681	0.25123
136	Kosovo	Central and Eastern Europe	5.589	0.24749

```
137  Palestinian Territories Middle East and Northern Africa          4.715 0.24499
138                   Haiti     Latin America and Caribbean           4.518 0.24425
139              Mauritania             Sub-Saharan Africa           4.436 0.24232
140                    Chad             Sub-Saharan Africa           3.667 0.23501
141             Afghanistan                  Southern Asia           3.575 0.23414
142                 Comoros             Sub-Saharan Africa           3.956 0.22917
143                  Turkey Middle East and Northern Africa           5.332 0.22815
144       Congo (Kinshasa)             Sub-Saharan Africa           4.517 0.22605
145              Lithuania      Central and Eastern Europe           5.833 0.21342
146                 Serbia      Central and Eastern Europe           5.123 0.20107
147                Armenia      Central and Eastern Europe           4.350 0.19847
148             Madagascar             Sub-Saharan Africa           3.681 0.19184
149             Montenegro      Central and Eastern Europe           5.192 0.18260
150                  Egypt Middle East and Northern Africa           4.194 0.17288
151                  Syria Middle East and Northern Africa           3.006 0.15684
152               Pakistan                  Southern Asia           5.194 0.12102
153                Burundi             Sub-Saharan Africa           2.905 0.11850
154                 Angola             Sub-Saharan Africa           4.033 0.10384
155                  Sudan             Sub-Saharan Africa           4.550 0.10081
156  Bosnia and Herzegovina      Central and Eastern Europe           4.949 0.09245
157                 Greece                  Western Europe           4.857 0.07699
158                   Iraq Middle East and Northern Africa           4.677 0.00000

>
```

10. Create a new column **data_2015** called **gff_stat.** For each row, the **gff_stat** is the sum of the family, freedom, and generosity values. Store the resulting table back into the **data_2015** variable.

```
View(data_2015)

> data_2015$gff_stat <- data_2015$family + data_2015$freedom + data_2015$generosity
> data_2015
       country                region happiness_rank happiness_score standard_error
1  Switzerland        Western Europe              1           7.587        0.03411
2      Iceland        Western Europe              2           7.561        0.04884
3      Denmark        Western Europe              3           7.527        0.03328
4       Norway        Western Europe              4           7.522        0.03880
5       Canada         North America              5           7.427        0.03553
6      Finland        Western Europe              6           7.406        0.03140
7  Netherlands        Western Europe              7           7.378        0.02799
8       Sweden        Western Europe              8           7.364        0.03157
9  New Zealand Australia and New Zealand          9           7.286        0.03371
```

10	Australia	Australia and New Zealand	10	7.284	0.04083
11	Israel	Middle East and Northern Africa	11	7.278	0.03470
12	Costa Rica	Latin America and Caribbean	12	7.226	0.04454
13	Austria	Western Europe	13	7.200	0.03751
14	Mexico	Latin America and Caribbean	14	7.187	0.04176
15	United States	North America	15	7.119	0.03839
16	Brazil	Latin America and Caribbean	16	6.983	0.04076
17	Luxembourg	Western Europe	17	6.946	0.03499
18	Ireland	Western Europe	18	6.940	0.03676
19	Belgium	Western Europe	19	6.937	0.03595
20	United Arab Emirates	Middle East and Northern Africa	20	6.901	0.03729
21	United Kingdom	Western Europe	21	6.867	0.01866
22	Oman	Middle East and Northern Africa	22	6.853	0.05335
23	Venezuela	Latin America and Caribbean	23	6.810	0.06476
24	Singapore	Southeastern Asia	24	6.798	0.03780
25	Panama	Latin America and Caribbean	25	6.786	0.04910
26	Germany	Western Europe	26	6.750	0.01848
27	Chile	Latin America and Caribbean	27	6.670	0.05800
28	Qatar	Middle East and Northern Africa	28	6.611	0.06257
29	France	Western Europe	29	6.575	0.03512
30	Argentina	Latin America and Caribbean	30	6.574	0.04612
31	Czech Republic	Central and Eastern Europe	31	6.505	0.04168
32	Uruguay	Latin America and Caribbean	32	6.485	0.04539
33	Colombia	Latin America and Caribbean	33	6.477	0.05051
34	Thailand	Southeastern Asia	34	6.455	0.03557
35	Saudi Arabia	Middle East and Northern Africa	35	6.411	0.04633
36	Spain	Western Europe	36	6.329	0.03468
37	Malta	Western Europe	37	6.302	0.04206
38	Taiwan	Eastern Asia	38	6.298	0.03868
39	Kuwait	Middle East and Northern Africa	39	6.295	0.04456
40	Suriname	Latin America and Caribbean	40	6.269	0.09811
41	Trinidad and Tobago	Latin America and Caribbean	41	6.168	0.10895
42	El Salvador	Latin America and Caribbean	42	6.130	0.05618
43	Guatemala	Latin America and Caribbean	43	6.123	0.05224
44	Uzbekistan	Central and Eastern Europe	44	6.003	0.04361
45	Slovakia	Central and Eastern Europe	45	5.995	0.04267
46	Japan	Eastern Asia	46	5.987	0.03581
47	South Korea	Eastern Asia	47	5.984	0.04098
48	Ecuador	Latin America and Caribbean	48	5.975	0.04528
49	Bahrain	Middle East and Northern Africa	49	5.960	0.05412
50	Italy	Western Europe	50	5.948	0.03914
51	Bolivia	Latin America and Caribbean	51	5.890	0.05642
52	Moldova	Central and Eastern Europe	52	5.889	0.03799
53	Paraguay	Latin America and Caribbean	53	5.878	0.04563
54	Kazakhstan	Central and Eastern Europe	54	5.855	0.04114
55	Slovenia	Central and Eastern Europe	55	5.848	0.04251

56	Lithuania	Central and Eastern Europe	56	5.833	0.03843
57	Nicaragua	Latin America and Caribbean	57	5.828	0.05371
58	Peru	Latin America and Caribbean	58	5.824	0.04615
59	Belarus	Central and Eastern Europe	59	5.813	0.03938
60	Poland	Central and Eastern Europe	60	5.791	0.04263
61	Malaysia	Southeastern Asia	61	5.770	0.04330
62	Croatia	Central and Eastern Europe	62	5.759	0.04394
63	Libya	Middle East and Northern Africa	63	5.754	0.07832
64	Russia	Central and Eastern Europe	64	5.716	0.03135
65	Jamaica	Latin America and Caribbean	65	5.709	0.13693
66	North Cyprus	Western Europe	66	5.695	0.05635
67	Cyprus	Western Europe	67	5.689	0.05580
68	Algeria	Middle East and Northern Africa	68	5.605	0.05099
69	Kosovo	Central and Eastern Europe	69	5.589	0.05018
70	Turkmenistan	Central and Eastern Europe	70	5.548	0.04175
71	Mauritius	Sub-Saharan Africa	71	5.477	0.07197
72	Hong Kong	Eastern Asia	72	5.474	0.05051
73	Estonia	Central and Eastern Europe	73	5.429	0.04013
74	Indonesia	Southeastern Asia	74	5.399	0.02596
75	Vietnam	Southeastern Asia	75	5.360	0.03107
76	Turkey	Middle East and Northern Africa	76	5.332	0.03864

	economy_gdp_per_capita	family	health_life_expectancy	freedom	trust_government_corruption	generosity
1	1.39651	1.34951	0.94143	0.66557	0.41978	0.29678
2	1.30232	1.40223	0.94784	0.62877	0.14145	0.43630
3	1.32548	1.36058	0.87464	0.64938	0.48357	0.34139
4	1.45900	1.33095	0.88521	0.66973	0.36503	0.34699
5	1.32629	1.32261	0.90563	0.63297	0.32957	0.45811
6	1.29025	1.31826	0.88911	0.64169	0.41372	0.23351
7	1.32944	1.28017	0.89284	0.61576	0.31814	0.47610
8	1.33171	1.28907	0.91087	0.65980	0.43844	0.36262
9	1.25018	1.31967	0.90837	0.63938	0.42922	0.47501
10	1.33358	1.30923	0.93156	0.65124	0.35637	0.43562
11	1.22857	1.22393	0.91387	0.41319	0.07785	0.33172
12	0.95578	1.23788	0.86027	0.63376	0.10583	0.25497
13	1.33723	1.29704	0.89042	0.62433	0.18676	0.33088
14	1.02054	0.91451	0.81444	0.48181	0.21312	0.14074
15	1.39451	1.24711	0.86179	0.54604	0.15890	0.40105
16	0.98124	1.23287	0.69702	0.49049	0.17521	0.14574
17	1.56391	1.21963	0.91894	0.61583	0.37798	0.28034
18	1.33596	1.36948	0.89533	0.61777	0.28703	0.45901
19	1.30782	1.28566	0.89667	0.58450	0.22540	0.22250
20	1.42727	1.12575	0.80925	0.64157	0.38583	0.26428
21	1.26637	1.28548	0.90943	0.59625	0.32067	0.51912
22	1.36011	1.08182	0.76276	0.63274	0.32524	0.21542
23	1.04424	1.25596	0.72052	0.42908	0.11069	0.05841

24	1.52186	1.02000	1.02525	0.54252	0.49210	0.31105
25	1.06353	1.19850	0.79661	0.54210	0.09270	0.24434
26	1.32792	1.29937	0.89186	0.61477	0.21843	0.28214
27	1.10715	1.12447	0.85857	0.44132	0.12869	0.33363
28	1.69042	1.07860	0.79733	0.64040	0.52208	0.32573
29	1.27778	1.26038	0.94579	0.55011	0.20646	0.12332
30	1.05351	1.24823	0.78723	0.44974	0.08484	0.11451
31	1.17898	1.20643	0.84483	0.46364	0.02652	0.10686
32	1.06166	1.20890	0.81160	0.60362	0.24558	0.23240
33	0.91861	1.24018	0.69077	0.53466	0.05120	0.18401
34	0.96690	1.26504	0.73850	0.55664	0.03187	0.57630
35	1.39541	1.08393	0.72025	0.31048	0.32524	0.13706
36	1.23011	1.31379	0.95562	0.45951	0.06398	0.18227
37	1.20740	1.30203	0.88721	0.60365	0.13586	0.51752
38	1.29098	1.07617	0.87530	0.39740	0.08129	0.25376
39	1.55422	1.16594	0.72492	0.55499	0.25609	0.16228
40	0.99534	0.97200	0.60820	0.59657	0.13633	0.16991
41	1.21183	1.18354	0.61483	0.55884	0.01140	0.31844
42	0.76454	1.02507	0.67737	0.40350	0.11776	0.10692
43	0.74553	1.04356	0.64425	0.57733	0.09472	0.27489
44	0.63244	1.34043	0.59772	0.65821	0.30826	0.22837
45	1.16891	1.26999	0.78902	0.31751	0.03431	0.16893
46	1.27074	1.25712	0.99111	0.49615	0.18060	0.10705
47	1.24461	0.95774	0.96538	0.33208	0.07857	0.18557
48	0.86402	0.99903	0.79075	0.48574	0.18090	0.11541
49	1.32376	1.21624	0.74716	0.45492	0.30600	0.17362
50	1.25114	1.19777	0.95446	0.26236	0.02901	0.22823
51	0.68133	0.97841	0.53920	0.57414	0.08800	0.20536
52	0.59448	1.01528	0.61826	0.32818	0.01615	0.20951
53	0.75985	1.30477	0.66098	0.53899	0.08242	0.34240
54	1.12254	1.12241	0.64368	0.51649	0.08454	0.11827
55	1.18498	1.27385	0.87337	0.60855	0.03787	0.25328
56	1.14723	1.25745	0.73128	0.21342	0.01031	0.02641
57	0.59325	1.14184	0.74314	0.55475	0.19317	0.27815
58	0.90019	0.97459	0.73017	0.41496	0.05989	0.14982
59	1.03192	1.23289	0.73608	0.37938	0.19090	0.11046
60	1.12555	1.27948	0.77903	0.53122	0.04212	0.16759
61	1.12486	1.07023	0.72394	0.53024	0.10501	0.33075
62	1.08254	0.79624	0.78805	0.25883	0.02430	0.05444
63	1.13145	1.11862	0.70380	0.41668	0.11023	0.18295
64	1.13764	1.23617	0.66926	0.36679	0.03005	0.00199
65	0.81038	1.15102	0.68741	0.50442	0.02299	0.21230
66	1.20806	1.07008	0.92356	0.49027	0.14280	0.26169
67	1.20813	0.89318	0.92356	0.40672	0.06146	0.30638
68	0.93929	1.07772	0.61766	0.28579	0.17383	0.07822
69	0.80148	0.81198	0.63132	0.24749	0.04741	0.28310

70	0.95847	1.22668	0.53886	0.47610	0.30844	0.16979
71	1.00761	0.98521	0.70950	0.56066	0.07521	0.37744
72	1.38604	1.05818	1.01328	0.59608	0.37124	0.39478
73	1.15174	1.22791	0.77361	0.44888	0.15184	0.08680
74	0.82827	1.08708	0.63793	0.46611	0.00000	0.51535
75	0.63216	0.91226	0.74676	0.59444	0.10441	0.16860
76	1.06098	0.94632	0.73172	0.22815	0.15746	0.12253

	dystopia_residual	gff_stat
1	2.51738	2.31186
2	2.70201	2.46730
3	2.49204	2.35135
4	2.46531	2.34767
5	2.45176	2.41369
6	2.61955	2.19346
7	2.46570	2.37203
8	2.37119	2.31149
9	2.26425	2.43406
10	2.26646	2.39609
11	3.08854	1.96884
12	3.17728	2.12661
13	2.53320	2.25225
14	3.60214	1.53706
15	2.51011	2.19420
16	3.26001	1.86910
17	1.96961	2.11580
18	1.97570	2.44626
19	2.41484	2.09266
20	2.24743	2.03160
21	1.96994	2.40085
22	2.47489	1.92998
23	3.19131	1.74345
24	1.88501	1.87357
25	2.84848	1.98494
26	2.11569	2.19628
27	2.67585	1.89942
28	1.55674	2.04473
29	2.21126	1.93381
30	2.83600	1.81248
31	2.67782	1.77693
32	2.32142	2.04492
33	2.85737	1.95885
34	2.31945	2.39798
35	2.43872	1.53147
36	2.12367	1.95557
37	1.64880	2.42320
38	2.32323	1.72733

```
39          1.87634  1.88321
40          2.79094  1.73848
41          2.26882  2.06082
42          3.03500  1.53549
43          2.74255  1.89578
44          2.23741  2.22701
45          2.24639  1.75643
46          1.68435  1.86032
47          2.21978  1.47539
48          2.53942  1.60018
49          1.73797  1.84478
50          2.02518  1.68836
51          2.82334  1.75791
52          3.10712  1.55297
53          2.18896  2.18616
54          2.24729  1.75717
55          1.61583  2.13568
56          2.44649  1.49728
57          2.32407  1.97474
58          2.59450  1.53937
59          2.13090  1.72273
60          1.86565  1.97829
61          1.88541  1.93122
62          2.75414  1.10951
63          2.09066  1.71825
64          2.27394  1.60495
65          2.32038  1.86774
66          1.59888  1.82204
67          1.88931  1.60628
68          2.43209  1.44173
69          2.76579  1.34257
70          1.86984  1.87257
71          1.76145  1.92331
72          0.65429  2.04904
73          1.58782  1.76359
74          1.86399  2.06854
75          2.20173  1.67530
76          2.08528  1.29700
 [ reached 'max' / getOption("max.print") -- omitted 82 rows ]
```

11. Group the **happy_df** data set by region. Run a summary that provides the number of countries in each region in a column called **country_count**, the **mean** happiness for each region in a column called **mean_happiness**, and the

mean freedom of each region in a column called **mean_freedom.** Store your resulting table in a variable called **regional_stats_df.**

```
regional_stats_df
# A tibble: 10 × 4
   region                        country_count mean_happiness mean_freedom
   <chr>                                 <int>          <dbl>        <dbl>
 1 Australia and New Zealand                 2           7.28        0.645
 2 Central and Eastern Europe               29           5.33        0.358
 3 Eastern Asia                              6           5.63        0.462
 4 Latin America and Caribbean              22           6.14        0.502
 5 Middle East and Northern Africa          20           5.41        0.362
 6 North America                             2           7.27        0.590
 7 Southeastern Asia                         9           5.32        0.557
 8 Southern Asia                             7           4.58        0.373
 9 Sub-Saharan Africa                       40           4.20        0.366
10 Western Europe                           21           6.69        0.550


>
```

Assignment Part 2

12. Download the **baseball.csv** data set that represents batting statistics from the 1986 Major League Baseball season. Read this data set in a **variable** called **baseball.**

```
baseball <- read.csv("baseball.csv")
baseball
```
was run to create and read the dataset.

13. Spend time with the data using various exploration functions to get a general feel for what you are working with. For more information on this data set and its various columns, see Baseball Reference's 1986 Major League Standard Batting.

I used various exploration functions like str(), summary(), head(), tail(), dim() to understand and carry out a preliminary exploration of the database

```
str(baseball)
summary(baseball)
head(baseball)
tail(baseball)
```

```
dim(baseball)
```

14. Remove (**filter)** from **baseball** any player with 0 at bats (AB). Store the result in **baseball.**

```
baseball <- baseball[baseball$AB > 0, ]
> baseball
        Last  First Age   G  PA  AB   R   H X2B X3B HR RBI SB CS  BB  SO
1       Acker    Jim  27  21  28  28   1   3   1   0  0   0  0  0   0  21
2      Adduci    Jim  26   3  13  11   2   1   1   0  0   0  0  0   1   2
3      Aguayo   Luis  27  62 146 133  17  28   6   1  4  13  1  1   8  26
4    Aguilera   Rick  24  32  57  51   4   8   0   0  2   6  0  0   3  12
6     Aldrete   Mike  25  84 256 216  27  54  18   3  2  25  1  3  33  34
7   Alexander  Doyle  35  18  45  38   2   8   1   0  0   5  0  0   0   8
8    Allanson   Andy  24 101 324 293  30  66   7   3  1  29 10  1  14  36
9       Almon   Bill  33 102 230 196  29  43   7   2  7  27 11  4  30  38
10    Amelung     Ed  27   8  11  11   0   1   0   0  0   0  0  0   0   4
11   Andersen  Larry  33  48   7   6   0   0   0   0  0   0  0  0   0   3
12   Anderson   Dave  25  92 241 216  31  53   9   0  1  15  5  1  22  39
13   Anderson   Rick  29  15  12  11   1   1   0   0  0   0  0  0   0   4
15      Armas   Tony  32 121 453 425  40 112  21   4 11  58  0  3  24  77
16    Asadoor  Randy  23  15  60  55   9  20   5   0  0   7  1  2   3  13
17      Ashby   Alan  34 120 361 315  24  81  15   0  7  38  1  0  39  56
18 Assenmacher   Paul  25  61   8   6   0   0   0   0  0   0  0  0   2   3
20    Backman  Wally  26 124 440 387  67 124  18   2  1  27 13  7  36  32
22     Bailey   Mark  24  57 182 153   9  27   5   0  4  15  1  1  28  45
23     Baines Harold  27 145 618 570  72 169  29   2 21  88  2  1  38  89
25      Baker  Dusty  37  83 271 242  25  58   8   0  4  19  0  1  27  37
26      Baker   Doug  25  13  30  24   1   3   1   0  0   0  0  0   2   7
27    Balboni  Steve  29 138 562 512  54 117  25   1 29  88  0  0  43 146
28     Baller    Jay  25  36   6   5   0   0   0   0  0   0  0  0   0   1
29      Bando  Chris  30  92 290 254  28  68   9   0  2  26  0  1  22  49
30   Barfield  Jesse  26 158 671 589 107 170  35   2 40 108  8  8  69 146
31     Bargar   Greg  27  22   2   2   0   0   0   0  0   0  0  0   0   2
32    Barrett  Marty  28 158 713 625  94 179  39   4  4  60 15  7  65  31
33       Bass  Kevin  27 157 640 591  83 184  33   5 20  79 22 13  38  72
34      Bathe   Bill  25  39 112 103   9  19   3   0  5  11  0  0   2  20
35     Baylor    Don  37 160 687 585  93 139  23   1 31  94  3  5  62 111
36      Beane  Billy  24  80 194 183  20  39   6   0  3  15  2  3  11  54
38   Bedrosian  Steve  28  68   6   5   0   1   0   0  0   0  0  0   1   1
39       Bell  Buddy  34 155 655 568  89 158  29   3 20  75  2  8  73  49
40       Bell George  26 159 690 641 101 198  38   6 31 108  7  8  41  62
41       Bell    Jay  20   5  16  14   3   5   2   0  1   4  0  0   2   3
42       Bell  Terry  23   8   5   3   0   0   0   0  0   0  0  0   2   1
43    Belliard Rafael  24 117 350 309  33  72   5   2  0  31 12  2  26  54
```

```
44    Benedict  Bruce  30  64 183 160  11  36  10   1  0  13  1  0  15  10
45    Beniquez   Juan  36 113 395 343  48 103  15   0  6  36  2  3  40  49
46   Berenguer   Juan  31  46  10   7   0   1   0   0  0   0  0  0   0   1
47     Berenyi  Bruce  31  14  12  11   0   0   0   0  0   0  0  0   0   3
48     Bergman   Dave  33  65 151 130  14  30   6   1  1   9  0  0  21  16
49   Bernazard   Tony  29 146 636 562  88 169  28   4 17  73 17  8  53  77
50       Berra   Dale  29  42 121 108  10  25   7   0  2  13  0  0   9  14
51  Biancalana  Buddy  26 100 209 190  24  46   4   4  2   8  5  1  15  50
52    Bielecki   Mike  26  31  54  48   3   3   0   0  0   1  0  0   2  26
53  Bilardello   Dann  27  79 212 191  12  37   5   0  4  17  1  0  14  32
54    Bittiger   Jeff  24   3   4   3   1   1   0   0  1   1  0  0   0   1
55        Blue   Vida  36  28  53  43   3   4   1   0  1   3  0  0   6  20
56      Bochte  Bruce  35 125 473 407  57 104  13   1  6  43  3  2  65  68
57       Bochy  Bruce  31  63 142 127  16  32   9   0  8  22  1  0  14  23
58      Bockus  Randy  25   6   1   1   0   0   0   0  0   0  0  0   0   1
59      Boever    Joe  25  11   2   2   0   1   0   0  0   0  0  0   0   0
60       Boggs   Wade  28 149 693 580 107 207  47   2  8  71  0  4 105  44
61       Bonds  Barry  21 113 484 413  72  92  26   3 16  48 36  7  65 102
62     Bonilla   Juan  30 102 316 284  33  69  10   1  1  18  0  0  25  21
63     Bonilla  Bobby  23 138 496 426  55 109  16   4  3  43  8  5  62  88
64     Bonnell  Barry  32  17  53  51   4  10   2   0  0   4  0  1   1  13
66       Boone    Bob  38 144 503 442  48  98  12   2  7  49  1  0  43  30
67      Bosley   Thad  29  87 139 120  15  33   4   1  1   9  3  0  18  24
68      Boston  Daryl  23  56 224 199  29  53  11   3  5  22  9  5  21  33
69     Bradley   Phil  27 143 615 526  88 163  27   4 12  50 21 12  77 134
 [ reached 'max' / getOption("max.print") -- omitted 664 rows ]
> dim(baseball)
[1] 726  16
```

15. Add a new column batting average called **BA.** Batting average is computed by the number of hits (H) divided by the number of at bats (AB). Store the result in **baseball.**

```
> baseball$BA <- baseball$H / baseball$AB
> baseball
         Last  First Age   G  PA  AB   R   H X2B X3B HR RBI SB CS  BB  SO        BA
1       Acker    Jim  27  21  28  28   1   3   1   0  0   0  0  0   0  21 0.10714286
2      Adduci    Jim  26   3  13  11   2   1   1   0  0   0  0  0   1   2 0.09090909
3      Aguayo   Luis  27  62 146 133  17  28   6   1  4  13  1  1   8  26 0.21052632
4    Aguilera   Rick  24  32  57  51   4   8   0   0  2   6  0  0   3  12 0.15686275
6     Aldrete   Mike  25  84 256 216  27  54  18   3  2  25  1  3  33  34 0.25000000
7   Alexander  Doyle  35  18  45  38   2   8   1   0  0   5  0  0   0   8 0.21052632
8    Allanson   Andy  24 101 324 293  30  66   7   3  1  29 10  1  14  36 0.22525597
9       Almon   Bill  33 102 230 196  29  43   7   2  7  27 11  4  30  38 0.21938776
10    Amelung     Ed  27   8  11  11   0   1   0   0  0   0  0  0   0   4 0.09090909
```

11	Andersen	Larry	33	48	7	6	0	0	0	0	0	0	0	0	0	3	0.00000000
12	Anderson	Dave	25	92	241	216	31	53	9	0	1	15	5	1	22	39	0.24537037
13	Anderson	Rick	29	15	12	11	1	1	0	0	0	0	0	0	0	4	0.09090909
15	Armas	Tony	32	121	453	425	40	112	21	4	11	58	0	3	24	77	0.26352941
16	Asadoor	Randy	23	15	60	55	9	20	5	0	0	7	1	2	3	13	0.36363636
17	Ashby	Alan	34	120	361	315	24	81	15	0	7	38	1	0	39	56	0.25714286
18	Assenmacher	Paul	25	61	8	6	0	0	0	0	0	0	0	0	2	3	0.00000000
20	Backman	Wally	26	124	440	387	67	124	18	2	1	27	13	7	36	32	0.32041344
22	Bailey	Mark	24	57	182	153	9	27	5	0	4	15	1	1	28	45	0.17647059
23	Baines	Harold	27	145	618	570	72	169	29	2	21	88	2	1	38	89	0.29649123
25	Baker	Dusty	37	83	271	242	25	58	8	0	4	19	0	1	27	37	0.23966942
26	Baker	Doug	25	13	30	24	1	3	1	0	0	0	0	0	2	7	0.12500000
27	Balboni	Steve	29	138	562	512	54	117	25	1	29	88	0	0	43	146	0.22851562
28	Baller	Jay	25	36	6	5	0	0	0	0	0	0	0	0	0	1	0.00000000
29	Bando	Chris	30	92	290	254	28	68	9	0	2	26	0	1	22	49	0.26771654
30	Barfield	Jesse	26	158	671	589	107	170	35	2	40	108	8	8	69	146	0.28862479
31	Bargar	Greg	27	22	2	2	0	0	0	0	0	0	0	0	0	2	0.00000000
32	Barrett	Marty	28	158	713	625	94	179	39	4	4	60	15	7	65	31	0.28640000
33	Bass	Kevin	27	157	640	591	83	184	33	5	20	79	22	13	38	72	0.31133672
34	Bathe	Bill	25	39	112	103	9	19	3	0	5	11	0	0	2	20	0.18446602
35	Baylor	Don	37	160	687	585	93	139	23	1	31	94	3	5	62	111	0.23760684
36	Beane	Billy	24	80	194	183	20	39	6	0	3	15	2	3	11	54	0.21311475
38	Bedrosian	Steve	28	68	6	5	0	1	0	0	0	0	0	0	1	1	0.20000000
39	Bell	Buddy	34	155	655	568	89	158	29	3	20	75	2	8	73	49	0.27816901
40	Bell	George	26	159	690	641	101	198	38	6	31	108	7	8	41	62	0.30889236
41	Bell	Jay	20	5	16	14	3	5	2	0	1	4	0	0	2	3	0.35714286
42	Bell	Terry	23	8	5	3	0	0	0	0	0	0	0	0	2	1	0.00000000
43	Belliard	Rafael	24	117	350	309	33	72	5	2	0	31	12	2	26	54	0.23300971
44	Benedict	Bruce	30	64	183	160	11	36	10	1	0	13	1	0	15	10	0.22500000
45	Beniquez	Juan	36	113	395	343	48	103	15	0	6	36	2	3	40	49	0.30029155
46	Berenguer	Juan	31	46	10	7	0	1	0	0	0	0	0	0	0	1	0.14285714
47	Berenyi	Bruce	31	14	12	11	0	0	0	0	0	0	0	0	0	3	0.00000000
48	Bergman	Dave	33	65	151	130	14	30	6	1	1	9	0	0	21	16	0.23076923
49	Bernazard	Tony	29	146	636	562	88	169	28	4	17	73	17	8	53	77	0.30071174
50	Berra	Dale	29	42	121	108	10	25	7	0	2	13	0	0	9	14	0.23148148
51	Biancalana	Buddy	26	100	209	190	24	46	4	4	2	8	5	1	15	50	0.24210526
52	Bielecki	Mike	26	31	54	48	3	3	0	0	0	1	0	0	2	26	0.06250000
53	Bilardello	Dann	27	79	212	191	12	37	5	0	4	17	1	0	14	32	0.19371728
54	Bittiger	Jeff	24	3	4	3	1	1	0	0	1	1	0	0	0	1	0.33333333
55	Blue	Vida	36	28	53	43	3	4	1	0	1	3	0	0	6	20	0.09302326
56	Bochte	Bruce	35	125	473	407	57	104	13	1	6	43	3	2	65	68	0.25552826
57	Bochy	Bruce	31	63	142	127	16	32	9	0	8	22	1	0	14	23	0.25196850
58	Bockus	Randy	25	6	1	1	0	0	0	0	0	0	0	0	0	1	0.00000000
59	Boever	Joe	25	11	2	2	0	1	0	0	0	0	0	0	0	0	0.50000000
60	Boggs	Wade	28	149	693	580	107	207	47	2	8	71	0	4	105	44	0.35689655
61	Bonds	Barry	21	113	484	413	72	92	26	3	16	48	36	7	65	102	0.22276029

```
62     Bonilla   Juan  30 102 316 284  33  69  10   1  1  18  0  0  25  21 0.24295775
63     Bonilla  Bobby  23 138 496 426  55 109  16   4  3  43  8  5  62  88 0.25586854
64     Bonnell  Barry  32  17  53  51   4  10   2   0  0   4  0  1   1  13 0.19607843
 [ reached 'max' / getOption("max.print") -- omitted 668 rows ]
```

16. On-base percentage (OBP) is arguably a better statistic than batting average. Create a column called **OBP** that computes this stat as (H + BB) / (AB + BB). Store the result in **baseball.**

```
> baseball$OBP <- (baseball$H + baseball$BB) / (baseball$AB + baseball$BB)
> baseball
         Last  First Age   G  PA  AB   R   H X2B X3B HR RBI SB CS  BB  SO        BA       OBP
1       Acker    Jim  27  21  28  28   1   3   1   0  0   0  0  0   0  21 0.10714286 0.10714286
2      Adduci    Jim  26   3  13  11   2   1   1   0  0   0  0  0   1   2 0.09090909 0.16666667
3      Aguayo   Luis  27  62 146 133  17  28   6   1  4  13  1  1   8  26 0.21052632 0.25531915
4    Aguilera   Rick  24  32  57  51   4   8   0   0  2   6  0  0   3  12 0.15686275 0.20370370
6     Aldrete   Mike  25  84 256 216  27  54  18   3  2  25  1  3  33  34 0.25000000 0.34939759
7   Alexander  Doyle  35  18  45  38   2   8   1   0  0   5  0  0   0   8 0.21052632 0.21052632
8    Allanson   Andy  24 101 324 293  30  66   7   3  1  29 10  1  14  36 0.22525597 0.26058632
9       Almon   Bill  33 102 230 196  29  43   7   2  7  27 11  4  30  38 0.21938776 0.32300885
10    Amelung     Ed  27   8  11  11   0   1   0   0  0   0  0  0   0   4 0.09090909 0.09090909
11   Andersen  Larry  33  48   7   6   0   0   0   0  0   0  0  0   0   3 0.00000000 0.00000000
12   Anderson   Dave  25  92 241 216  31  53   9   0  1  15  5  1  22  39 0.24537037 0.31512605
13   Anderson   Rick  29  15  12  11   1   1   0   0  0   0  0  0   0   4 0.09090909 0.09090909
15      Armas   Tony  32 121 453 425  40 112  21   4 11  58  0  3  24  77 0.26352941 0.30289532
16    Asadoor  Randy  23  15  60  55   9  20   5   0  0   7  1  2   3  13 0.36363636 0.39655172
17      Ashby   Alan  34 120 361 315  24  81  15   0  7  38  1  0  39  56 0.25714286 0.33898305
18 Assenmacher   Paul  25  61   8   6   0   0   0   0  0   0  0  0   2   3 0.00000000 0.25000000
20    Backman  Wally  26 124 440 387  67 124  18   2  1  27 13  7  36  32 0.32041344 0.3782505
9
22     Bailey   Mark  24  57 182 153   9  27   5   0  4  15  1  1  28  45 0.17647059 0.30386740
23     Baines Harold  27 145 618 570  72 169  29   2 21  88  2  1  38  89 0.29649123 0.34046053
25      Baker  Dusty  37  83 271 242  25  58   8   0  4  19  0  1  27  37 0.23966942 0.31598513
26      Baker   Doug  25  13  30  24   1   3   1   0  0   0  0  0   2   7 0.12500000 0.19230769
27    Balboni  Steve  29 138 562 512  54 117  25   1 29  88  0  0  43 146 0.22851562 0.28828829
28     Baller    Jay  25  36   6   5   0   0   0   0  0   0  0  0   0   1 0.00000000 0.00000000
29      Bando  Chris  30  92 290 254  28  68   9   0  2  26  0  1  22  49 0.26771654 0.32608696
30   Barfield  Jesse  26 158 671 589 107 170  35   2 40 108  8  8  69 146 0.28862479 0.36322188
31     Bargar   Greg  27  22   2   2   0   0   0   0  0   0  0  0   0   2 0.00000000 0.00000000
32    Barrett  Marty  28 158 713 625  94 179  39   4  4  60 15  7  65  31 0.28640000 0.35362319
33       Bass  Kevin  27 157 640 591  83 184  33   5 20  79 22 13  38  72 0.31133672 0.35294118
34      Bathe   Bill  25  39 112 103   9  19   3   0  5  11  0  0   2  20 0.18446602 0.20000000
35     Baylor    Don  37 160 687 585  93 139  23   1 31  94  3  5  62 111 0.23760684 0.31066461
36      Beane  Billy  24  80 194 183  20  39   6   0  3  15  2  3  11  54 0.21311475 0.25773196
```

```
38   Bedrosian  Steve  28  68   6   5   0   1   0   0  0   0  0  0   1   1 0.20000000 0.33333333
39        Bell  Buddy  34 155 655 568  89 158  29   3 20  75  2  8  73  49 0.27816901 0.36037441
40        Bell George  26 159 690 641 101 198  38   6 31 108  7  8  41  62 0.30889236 0.35043988
41        Bell    Jay  20   5  16  14   3   5   2   0  1   4  0  0   2   3 0.35714286 0.43750000
42        Bell  Terry  23   8   5   3   0   0   0   0  0   0  0  0   2   1 0.00000000 0.40000000
43    Belliard Rafael  24 117 350 309  33  72   5   2  0  31 12  2  26  54 0.23300971 0.29253731
44    Benedict  Bruce  30  64 183 160  11  36  10   1  0  13  1  0  15  10 0.22500000 0.29142857
45    Beniquez   Juan  36 113 395 343  48 103  15   0  6  36  2  3  40  49 0.30029155 0.37336815
46   Berenguer   Juan  31  46  10   7   0   1   0   0  0   0  0  0   0   1 0.14285714 0.14285714
47     Berenyi  Bruce  31  14  12  11   0   0   0   0  0   0  0  0   0   3 0.00000000 0.00000000
48     Bergman   Dave  33  65 151 130  14  30   6   1  1   9  0  0  21  16 0.23076923 0.33774834
49   Bernazard   Tony  29 146 636 562  88 169  28   4 17  73 17  8  53  77 0.30071174 0.3609756
1
50       Berra   Dale  29  42 121 108  10  25   7   0  2  13  0  0   9  14 0.23148148 0.29059829
51  Biancalana  Buddy  26 100 209 190  24  46   4   4  2   8  5  1  15  50 0.24210526 0.29756098
52    Bielecki   Mike  26  31  54  48   3   3   0   0  0   1  0  0   2  26 0.06250000 0.10000000
53  Bilardello   Dann  27  79 212 191  12  37   5   0  4  17  1  0  14  32 0.19371728 0.24878049
54    Bittiger   Jeff  24   3   4   3   1   1   0   0  1   1  0  0   0   1 0.33333333 0.33333333
55        Blue   Vida  36  28  53  43   3   4   1   0  1   3  0  0   6  20 0.09302326 0.20408163
56      Bochte  Bruce  35 125 473 407  57 104  13   1  6  43  3  2  65  68 0.25552826 0.35805085
57       Bochy  Bruce  31  63 142 127  16  32   9   0  8  22  1  0  14  23 0.25196850 0.32624113
58      Bockus  Randy  25   6   1   1   0   0   0   0  0   0  0  0   0   1 0.00000000 0.00000000
59      Boever    Joe  25  11   2   2   0   1   0   0  0   0  0  0   0   0 0.50000000 0.50000000
60       Boggs   Wade  28 149 693 580 107 207  47   2  8  71  0  4 105  44 0.35689655 0.45547445
61       Bonds  Barry  21 113 484 413  72  92  26   3 16  48 36  7  65 102 0.22276029 0.32845188
 [ reached 'max' / getOption("max.print") -- omitted 671 rows ]
```

17. Determine the 10 players who struck out the most this season. Store these results as **strikeout_artist.**

```
Last  First Age   G  PA  AB   R   H X2B X3B HR RBI SB CS BB  SO       BA       OBP
315 Incaviglia   Pete  22 153 606 540  82 135  21   2 30  88  3  2 55 185 0.2500000 0.3193277
158       Deer    Rob  25 134 546 466  75 108  17   3 33  86  5  2 72 179 0.2317597 0.3345725
103    Canseco   Jose  21 157 682 600  85 144  29   1 33 117 15  7 65 175 0.2400000 0.3142857
544    Presley    Jim  24 155 660 616  83 163  33   4 27 107  0  4 32 172 0.2646104 0.3009259
676  Tartabull  Danny  23 137 578 511  76 138  25   6 25  96  4  8 61 157 0.2700587 0.3479021
27     Balboni  Steve  29 138 562 512  54 117  25   1 29  88  0  0 43 146 0.2285156 0.2882883
30    Barfield  Jesse  26 158 671 589 107 170  35   2 40 108  8  8 69 146 0.2886248 0.3632219
607    Samuel   Juan  25 145 633 591  90 157  36  12 16  78 42 14 26 142 0.2656514 0.2965964
```

```
475     Murphy   Dale  30 160 692 614  89 163  29   7 29  83  7  7 75 141 0.2654723 0.3454282
668 Strawberry Darryl  24 136 562 475  76 123  27   5 27  93 28 12 72 141 0.2589474
0.3564899
```

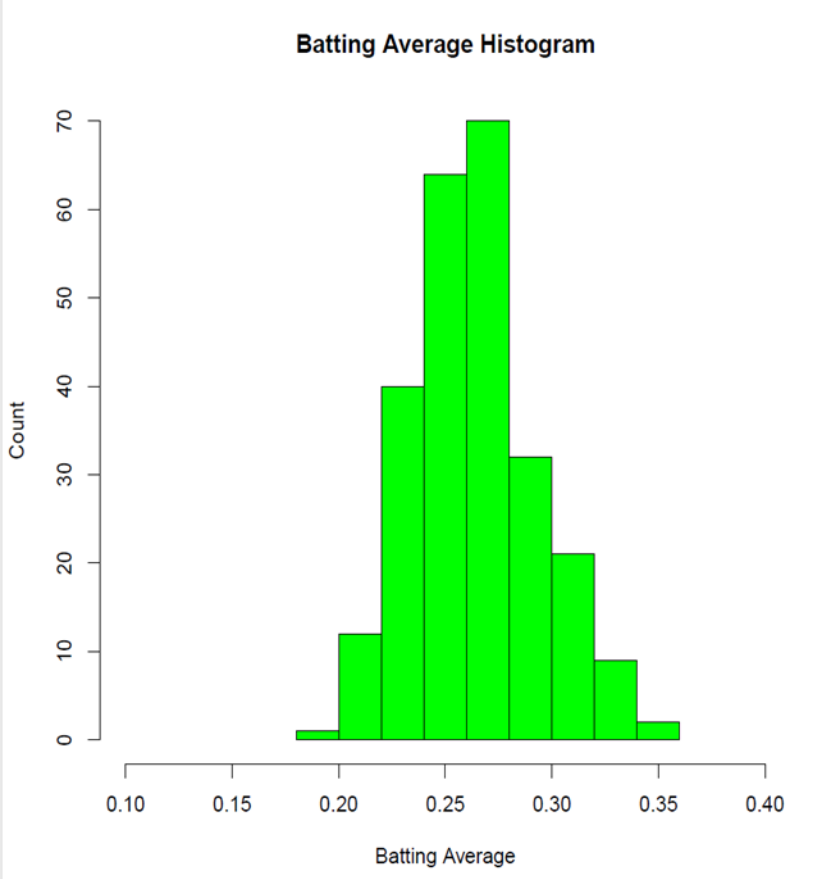
18. To be eligible for end-of-season awards, a player must have either at least 300 at bats or appear in at least 100 games. Keep only the players who are eligible to be considered and store them in a variable called **eligible_df.**

```
eligible_df
        Last   First Age   G  PA  AB   R   H X2B X3B HR RBI  SB CS  BB  SO        BA       OBP
8    Allanson    Andy  24 101 324 293  30  66   7   3  1  29  10  1  14  36 0.2252560 0.2605863
9       Almon    Bill  33 102 230 196  29  43   7   2  7  27  11  4  30  38 0.2193878 0.3230088
15      Armas    Tony  32 121 453 425  40 112  21   4 11  58   0  3  24  77 0.2635294 0.3028953
17      Ashby    Alan  34 120 361 315  24  81  15   0  7  38   1  0  39  56 0.2571429 0.3389831
20    Backman   Wally  26 124 440 387  67 124  18   2  1  27  13  7  36  32 0.3204134 0.3782506
23     Baines  Harold  27 145 618 570  72 169  29   2 21  88   2  1  38  89 0.2964912 0.3404605
27    Balboni   Steve  29 138 562 512  54 117  25   1 29  88   0  0  43 146 0.2285156 0.2882883
30    Barfield   Jesse  26 158 671 589 107 170  35   2 40 108   8  8  69 146 0.2886248 0.3632219
32    Barrett   Marty  28 158 713 625  94 179  39   4  4  60  15  7  65  31 0.2864000 0.3536232
33       Bass   Kevin  27 157 640 591  83 184  33   5 20  79  22 13  38  72 0.3113367 0.3529412
35     Baylor     Don  37 160 687 585  93 139  23   1 31  94   3  5  62 111 0.2376068 0.3106646
39       Bell   Buddy  34 155 655 568  89 158  29   3 20  75   2  8  73  49 0.2781690 0.3603744
40       Bell  George  26 159 690 641 101 198  38   6 31 108   7  8  41  62 0.3088924 0.3504399
43    Belliard  Rafael  24 117 350 309  33  72   5   2  0  31  12  2  26  54 0.2330097 0.2925373
45    Beniquez    Juan  36 113 395 343  48 103  15   0  6  36   2  3  40  49 0.3002915 0.3733681
49   Bernazard    Tony  29 146 636 562  88 169  28   4 17  73  17  8  53  77 0.3007117 0.3609756
51  Biancalana   Buddy  26 100 209 190  24  46   4   4  2   8   5  1  15  50 0.2421053 0.2975610
56      Bochte   Bruce  35 125 473 407  57 104  13   1  6  43   3  2  65  68 0.2555283 0.3580508
60       Boggs    Wade  28 149 693 580 107 207  47   2  8  71   0  4 105  44 0.3568966 0.4554745
61       Bonds   Barry  21 113 484 413  72  92  26   3 16  48  36  7  65 102 0.2227603 0.3284519
62     Bonilla    Juan  30 102 316 284  33  69  10   1  1  18   0  0  25  21 0.2429577 0.3042071
63     Bonilla   Bobby  23 138 496 426  55 109  16   4  3  43   8  5  62  88 0.2558685 0.3504098
66       Boone     Bob  38 144 503 442  48  98  12   2  7  49   1  0  43  30 0.2217195 0.2907216
69     Bradley    Phil  27 143 615 526  88 163  27   4 12  50  21 12  77 134 0.3098859 0.3980100
73      Bream     Sid  25 154 591 522  73 140  37   5 16  77  13  7  60  73 0.2681992 0.3436426
74      Brenly     Bob  32 149 560 472  60 116  26   0 16  62  10  6  74  97 0.2457627 0.3479853
75       Brett  George  33 124 529 441  70 128  28   4 16  73   1  2  80  45 0.2902494 0.3992322
77       Brock    Greg  29 115 367 325  33  76  13   0 16  52   2  5  37  60 0.2338462 0.3121547
79      Brooks   Hubie  29  80 338 306  50 104  18   5 14  58   4  2  25  60 0.3398693 0.3897281
81       Brown   Chris  24 116 463 416  57 132  16   3  7  49  13  9  33  43 0.3173077 0.3674833
86   Brunansky     Tom  25 157 655 593  69 152  28   1 23  75  12  4  53  98 0.2563238 0.3173375
88     Buckner    Bill  36 153 681 629  73 168  39   2 18 102   6  4  40  25 0.2670906 0.3109118
89    Buechele   Steve  24 153 513 461  54 112  19   2 18  54   5  8  35  98 0.2429501 0.2963710
```

```
94        Bush   Randy  27 130 402 357  50  96  19   7  7  45   5  3  39  63 0.2689076 0.3409091
96      Butler   Brett  29 161 683 587  92 163  17  14  4  51  32 15  70  65 0.2776831 0.3546423
97      Cabell    Enos  36 107 298 277  27  71  11   0  2  29  10  4  14  26 0.2563177 0.2920962
102  Cangelosi    John  23 137 525 438  65 103  16   3  2  32  50 17  71  61 0.2351598 0.3418468
103    Canseco    Jose  21 157 682 600  85 144  29   1 33 117  15  7  65 175 0.2400000 0.3142857
106     Carter    Gary  32 132 573 490  81 125  14   2 24 105   1  0  62  63 0.2551020 0.3387681
107     Carter     Joe  26 162 709 663 108 200  36   9 29 121  29  7  32  95 0.3016591 0.3338129
118      Clark    Will  22 111 458 408  66 117  27   2 11  41   4  7  34  76 0.2867647 0.3416290
121    Coleman   Vince  24 154 670 600  94 139  13   8  0  29 107 14  60  98 0.2316667 0.3015152
122      Coles Darnell  24 142 587 521  67 142  30   2 20  86   6  2  45  84 0.2725528 0.3303887
123    Collins    Dave  33 124 476 419  44 113  18   2  1  27  27 12  44  49 0.2696897 0.3390929
124 Concepcion    Dave  38  90 346 311  42  81  13   2  3  30  13  2  26  43 0.2604502 0.3175074
126     Cooper   Cecil  36 134 589 542  46 140  24   1 12  75   1  2  41  87 0.2583026 0.3104631
135       Cruz    Jose  38 141 536 479  48 133  22   4 10  72   3  4  55  86 0.2776618 0.3520599
142      Davis    Eric  24 132 487 415  97 115  15   3 27  71  80 11  68 100 0.2771084 0.3788820
143      Davis   Glenn  25 158 654 574  91 152  32   3 31 101   3  1  64  72 0.2648084 0.3385580
144      Davis    Jody  29 148 581 528  61 132  27   2 21  74   0  1  41 110 0.2500000 0.3040422
146      Davis   Chili  26 153 618 526  71 146  28   3 13  70  16 13  84  96 0.2775665 0.3770492
147      Davis   Alvin  25 135 562 479  66 130  18   1 18  72   0  3  76  68 0.2713987 0.3711712
149      Davis    Mike  27 142 533 489  77 131  28   3 19  55  27  4  34  91 0.2678937 0.3154876
152     Dawson   Andre  31 130 546 496  65 141  32   2 20  78  18 12  37  79 0.2842742 0.3339587
156   DeCinces    Doug  35 140 572 512  69 131  20   3 26  96   2  2  52  74 0.2558594 0.3244681
 [ reached 'max' / getOption("max.print") -- omitted 196 rows ]
```

```
>
```

19. For eligible players, create a histogram of batting average.



20. Important statistics for baseball players include the on-base percentage (OBP), the number of home runs (HR), the number of runs batted-in (RBI) among others. Analyze the eligible players and select a player that in your opinion is deserving of the Most Valuable Player (MVP) award. This choice must be supported by your data. In your report, you should present your data analysis supported by relevant data points and statistics that supports your recommendation. Produce a concise, written executive summary that focuses on the baseball data analysis. In addition to the title page and citations, it contains an introduction, presentation of written key findings, and a conclusion that contains your recommendations as supported by the data. Your executive summary should adhere to basic APA guidelines.

Exploratory Data Analysis and Most Valuable Player (MVP) recommendation based on the baseball.csv dataset

Analyst: Syed Faizan

Instructor: Prof. Maria Ayala

Submission Date: 1/21/2024

Introduction and Key Findings

143. Introduction:

The baseball dataset is a data frame with contains information about various players and their performance statistics. The dataset includes key offensive metrics such as Home Runs, RBIs, and more.

2. Data Overview:

- The dataset has a total of 771 rows, each representing a player's performance.
- There are 16 columns providing information on player names, age, at-bats, hits, runs, and various other statistics.

Data Manipulation:

Two new variables namely "**batting avg**" and "**on-base percentage**" were created in order to facilitate analysis of the dataset "baseball".

A **missing data pattern** showed the limitations of these two variables due to their missing values ("NA") as a result of 0 at-bats among 45 players.

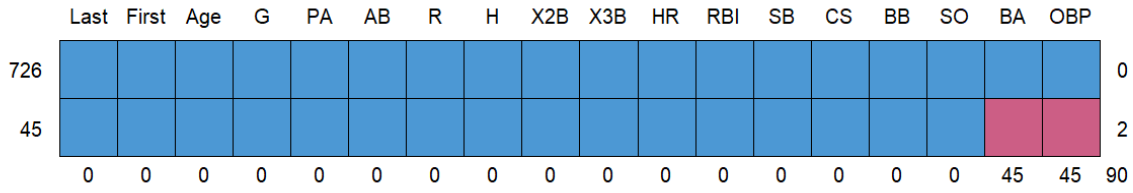


Fig 1. **Missing Data Pattern of "Baseball".**

I therefore ventured to create a new dataset namely "**baseball1**" with all the players with 0 at bats removed, in order to facilitate a deeper analysis of performance without the inconvenience of having NA values in the Dataframe.

I then worked towards finding out the players with the best performance through deploying summarizing and descriptive tools to analyze the Dataframe.

3. Descriptive Statistics:

3.1 Summarizing Statistics:

Below are some summarized statistics for key offensive metrics:

	Metric	Minimum	FirstQuartile	Median	Mean	ThirdQuartile	Maximum
1	Batting Average	0	0.1604	0.2347	0.2088	0.2689	1
2	Home Runs	0	0.0000	1.0000	5.2520	7.7500	40
3	RBIs	0	1.0000	11.0000	23.9600	41.7500	121
4	OBP	0	0.2033	0.3000	0.2672	0.3396	1

Table 1. Summarized key statistics

Using these key data, the ten best players were determined.

	Last	First	Age	G	PA	AB	R	H	X2B	X3B	HR	RBI	SB	CS	BB	SO	BA	OBP	CompositeScore
618	Schmidt	Mike	36	160	657	552	97	160	29	1	37	119	1	2	89	84	0.2898551	0.3884555	19.53248
30	Barfield	Jesse	26	158	671	589	107	170	35	2	40	108	8	8	69	146	0.2886248	0.3632219	19.02442
103	Canseco	Jose	21	157	682	600	85	144	29	1	33	117	15	7	65	175	0.2400000	0.3142857	18.49029
107	Carter	Joe	26	162	709	663	108	200	36	9	29	121	29	7	32	95	0.3016591	0.3338129	18.12081
518	Parker	Dave	35	162	700	637	89	174	31	3	31	116	1	6	56	126	0.2731554	0.3318903	18.00883
218	Gaetti	Gary	27	157	661	596	91	171	34	1	34	108	14	15	52	108	0.2869128	0.3441358	17.81801
430	Mattingly	Don	25	162	742	677	117	238	53	2	31	113	0	0	53	35	0.3515510	0.3986301	17.76021
40	Bell	George	26	159	690	641	101	198	38	6	31	108	7	8	41	62	0.3088924	0.3504399	17.22869
353	Kingman	Dave	37	144	604	561	70	118	19	0	35	94	3	3	33	126	0.2103387	0.2542088	16.56040
143	Davis	Glenn	25	158	654	574	91	152	32	3	31	101	3	1	64	72	0.2648084	0.3385580	16.50749

Table 2. Top Ten Players.

Of the ten best players using the selected via key statistics and the composite score **Mike Schmidt** was determined to be the most suitable candidate for MVP.

Key Insights:

- Players in the dataset have a diverse range of batting averages, with some exhibiting high averages.
- There is a positive correlation between Batting Average and Home Runs, indicating that players with higher averages tend to hit more home runs.
- The distribution of RBIs shows variation among players, with some players consistently contributing more RBIs.

3.2 Visualization of Statistics –

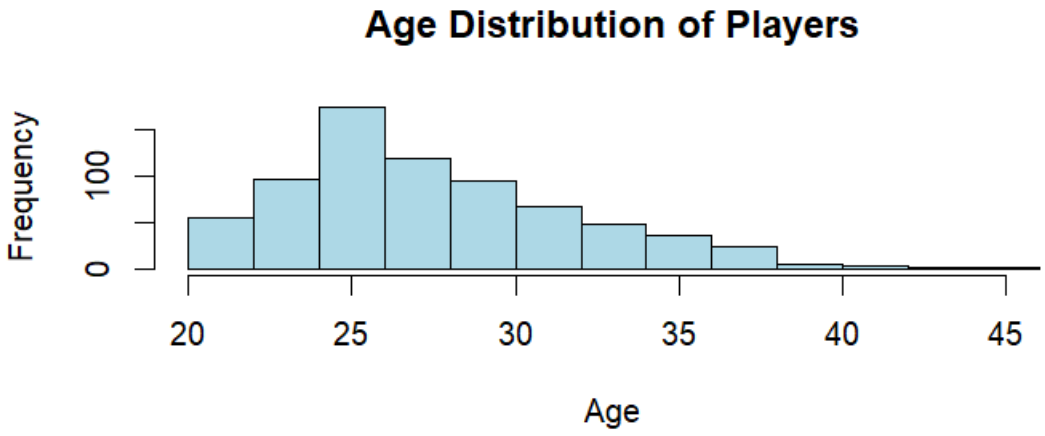


Fig 2. Age Distribution of the players.

The age distribution of players ranges from 20 to 43.

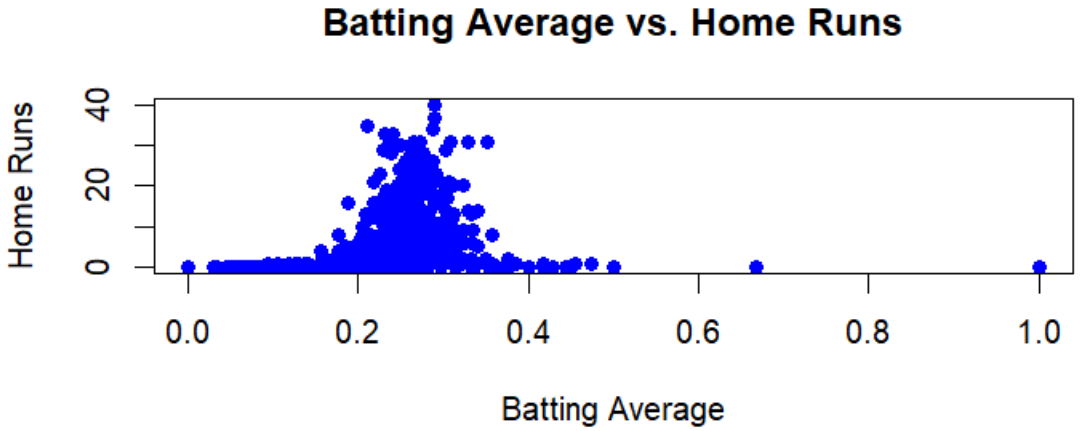


Fig 3. BA vs HR.

The scatter plot shows the relationship between Batting Average and Home Runs. Players with higher batting averages tend to have a higher number of home runs.

The boxplot displays the distribution of RBIs among players, showing the median, quartiles, and potential outliers.

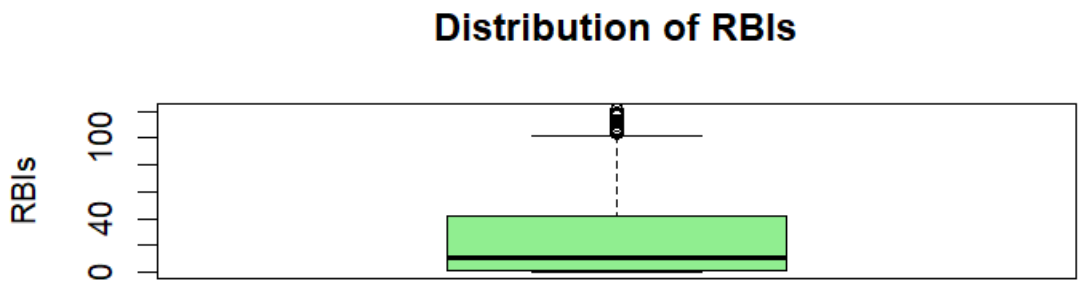


Fig 4. Distribution of RBI.

4. Conclusion

This exploratory data analysis provides an initial understanding of the baseball dataset, highlighting key statistics and relationships among offensive metrics. Further analysis and modeling could provide deeper insights into player performance and contributions to the team. Based on a basic exploratory analysis and key offensive statistics the ten best players were determined to be

618 – Mike Schmidt

30 – Jesse Barfield

103 – Jose Canseco

107 – Joe Carter

518 – Dave Parker

218 – Gary Gaetti

430 – Don Mattingly

40 – George Bell

353 – Dave Kingman

143 – Glenn Davis

Among whom Mike Schmidt was, in conclusion, found to be the most eligible for the title of MVP for the following reasons.

1. **Batting Performance:**
    - Schmidt's batting statistics, including his batting average, number of hits, home runs, and RBIs are comparatively high.
2. **Run Production:**
    - Schmidt's role in run production. The number of runs he scored and the runs his RBIs contributed to. This reflects his impact on the team's offensive success.
3. **Comparisons with Peers:**
    - Comparing Schmidt's statistics with those of other players in the dataset, he stands out in key metrics, especially in comparison to players in similar positions, and this fact strengthens his MVP case.

5. Recommendations

A few recommendations as to what data might be needed to further illuminate our grasp of the data related to baseball players are included below –

1. **Team Wins and Standings:**
    - Assessing how Schmidt's performance correlates with his team's success is very important and this data will help us understand Schmidt's performances better. If the team performed well and secured a high position in the standings, Schmidt's contributions likely played a crucial role.
2. **Situational Performance:**
    - Evaluating Schmidt's performance in critical situations, such as clutch hits or key defensive plays. This can demonstrate his ability to shine when the team needs it the most.
3. **Leadership and Intangibles:**
    - Missing in the dataset is information regarding Schmidt's leadership qualities, influence in the clubhouse, and other intangibles that may contribute to team morale and success.
4. **Fan and Expert Opinion:**
    - External factors, such as fan and expert opinions or awards are widely recognized to add credibility to MVP candidacy. Data on fan's opinion of him and expert evaluations might help us a great deal in making a more informed decision.

Bibliography

Arends, Danny. Live video- Data Analysis using R . YouTube. https://www.youtube.com/live/FbFMFcV2yBM?si=2Gv24rK3D7YRt7pA

Ayala, Professor Maria. (2024, January). Module 2 Assignment — Project 2- Exploratory Data Analysis of two Data sets. Canvas. Retrieved 21st January, 2024 from https://northeastern.instructure.com/courses/165041/assignments/2099213

Bluman, A. (2018). *Elementary statistics: A step by step approach* (10th ed.). McGraw Hill.ISBN 13: 978-1-259-755330

DataCamp. (n.d.). DataCamp: Learn R, Python & Data Science. https://www.datacamp.com/

Kabacoff, R. I. (2022). *R in action: Data analysis and graphics with R and tidyverse* (3rd ed.). Manning Publications. ISBN 978-1-617-29605-5

No Author. (n.d.). freecodecamp - Datascience [Playlist]. YouTube. https://youtube.com/playlist?list=PLWKjhJtqVAblQe2CCWqV4Zy3LY01Z8aF1&si=KCcE3aqRGVACzqFF