

Assignment 2

Clement Cheah

2022-06-14

#Description of the data (Section 1)

The data that I would be using today talks about the most common unisex names used in America. The data contains 929 names given to male and female. Nearly 1 percent of Americans have a unisex name and nearly 3 million individuals have one of the 929 names in the dataset.

It is unsure on how the data was collected but it was partially derived from the SSA. The data set that was extracted was in a “.csv” file format. It is a flat file as it stores data in plain text format. It is also a fixed width file because each column is in integerial form except for column names. The dataset does not have a delimiter.

```
library(rmarkdown)

## Warning: package 'rmarkdown' was built under R version 4.1.3

library(dplyr)

## Warning: package 'dplyr' was built under R version 4.1.3

##
## Attaching package: 'dplyr'

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

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

library(ggplot2)

## Warning: package 'ggplot2' was built under R version 4.1.3
```

#Reading the data into R (Section 2)

To allow RStudio to read the data, I copied the raw github link and pasted the link into a code chuck. The data is read into RStudio as “unisexdata”

#Clean the data (Section 3)

1. First, I wanted to round the values in column total to a whole number.

```
round(unisexdata[, 'total'], digits = 0)
```

```
## [1] 176544 154861 136382 132929 121797 109870 94896 88964 80401 79211
## [11] 64152 53486 51288 44782 41921 41238 36361 32636 29529 27351
## [21] 24956 24790 24167 23160 22296 22185 21336 18844 18665 18367
## [31] 17890 16769 15093 14841 12546 12542 12515 11909 11841 11133
## [41] 10973 10195 9972 9838 9445 8999 8776 8387 8301 8282
## [51] 7923 7773 7729 7626 7585 7496 7149 7108 7064 6853
## [61] 6729 6559 6229 6140 6101 6095 6013 5847 5771 5748
## [71] 5686 5514 5450 5428 5232 4805 4797 4473 4429 4421
## [81] 4389 4249 4230 4123 4059 4018 4012 3999 3961 3913
## [91] 3844 3801 3782 3775 3739 3729 3704 3652 3635 3625
## [101] 3616 3615 3555 3533 3446 3407 3406 3336 3275 3270
## [111] 3201 3185 3156 3094 3064 3021 3021 3014 2965 2960
## [121] 2931 2900 2895 2894 2872 2856 2809 2798 2750 2685
## [131] 2549 2540 2535 2501 2493 2484 2481 2477 2473 2471
## [141] 2399 2396 2375 2343 2327 2323 2299 2299 2292 2268
## [151] 2266 2239 2204 2183 2156 2155 2140 2134 2106 2090
## [161] 2070 2061 2036 1997 1991 1978 1970 1959 1951 1946
## [171] 1928 1921 1911 1894 1866 1828 1795 1787 1777 1763
## [181] 1746 1740 1725 1711 1707 1676 1671 1666 1655 1653
## [191] 1648 1645 1643 1642 1612 1601 1595 1595 1585 1555
## [201] 1553 1547 1545 1538 1536 1535 1526 1518 1511 1465
## [211] 1464 1462 1459 1450 1448 1435 1420 1406 1392 1379
## [221] 1373 1363 1358 1354 1345 1343 1333 1324 1322 1319
## [231] 1313 1311 1308 1299 1294 1293 1288 1281 1279 1270
## [241] 1264 1262 1259 1242 1235 1227 1223 1219 1201 1199
## [251] 1195 1195 1192 1182 1177 1174 1173 1172 1151 1144
## [261] 1134 1131 1129 1111 1111 1109 1107 1099 1097 1094
## [271] 1090 1082 1081 1073 1070 1066 1065 1058 1056 1055
## [281] 1049 1043 1040 1040 1035 1033 1030 1021 1011 1006
## [291] 1001 1000 989 980 970 954 951 951 947 941
## [301] 936 926 914 911 910 902 893 888 884 882
## [311] 873 869 863 851 848 836 826 819 811 807
## [321] 804 803 803 800 798 797 796 795 794 790
## [331] 787 784 783 778 761 759 758 756 749 749
## [341] 745 741 736 736 735 731 730 729 726 719
## [351] 715 715 713 702 701 699 693 693 693 693
## [361] 689 683 683 674 673 673 666 666 665 664
## [371] 663 662 662 661 659 659 656 656 655 649
## [381] 646 646 645 643 639 634 633 628 625 624
## [391] 624 618 613 613 613 611 608 607 607 601
## [401] 598 595 592 591 590 590 589 587 583 583
## [411] 581 581 581 578 577 574 573 559 556 555
## [421] 554 542 541 540 537 537 536 535 535 534
## [431] 534 533 531 530 530 528 527 524 523 521
## [441] 520 519 515 514 513 511 509 506 504 503
## [451] 501 500 499 498 498 496 495 495 490 490
## [461] 489 487 485 484 483 480 480 480 479 477
## [471] 477 476 476 476 473 473 470 469 465 463
## [481] 462 460 460 460 457 451 449 448 445 444
## [491] 442 442 441 439 439 439 438 436 435 433
## [501] 430 429 428 427 425 421 419 418 418 415
```

```

## [511] 414 414 413 413 413 413 413 413 412 411 410
## [521] 409 407 403 403 400 394 394 394 393 391
## [531] 391 390 389 389 386 384 383 382 381 376
## [541] 376 376 373 371 369 368 367 367 366 366
## [551] 365 365 364 364 361 360 358 358 356 355
## [561] 353 352 349 349 346 345 345 344 343 342
## [571] 341 340 334 333 333 329 327 326 324 324
## [581] 323 323 321 320 319 318 316 316 315 315
## [591] 313 311 309 307 306 305 304 302 302 300
## [601] 300 299 299 298 296 295 294 293 293 292
## [611] 291 291 290 290 289 288 286 285 285 285
## [621] 284 284 280 279 279 278 277 275 274 274
## [631] 273 272 272 272 271 267 265 265 265 263
## [641] 262 262 260 260 259 258 257 257 256 255
## [651] 253 253 249 249 248 248 247 246 245 244
## [661] 243 242 242 240 240 240 238 238 238 237
## [671] 237 237 236 236 235 234 233 232 231 231
## [681] 231 229 229 227 227 227 226 226 225 224
## [691] 224 224 221 221 221 219 218 218 217 217
## [701] 215 215 214 214 213 213 213 210 210 210
## [711] 208 206 206 206 206 206 205 205 205 204
## [721] 204 203 201 201 199 198 197 197 196 195
## [731] 194 194 192 192 190 190 190 190 189 189
## [741] 187 187 186 185 183 181 180 179 179 179
## [751] 178 178 177 176 176 175 173 173 171 171
## [761] 169 169 168 168 168 168 166 166 164 164
## [771] 161 161 161 161 161 160 159 158 158 158
## [781] 157 157 157 156 156 156 155 155 155 155
## [791] 155 154 154 154 154 154 154 153 152 152
## [801] 151 151 150 149 149 148 147 147 147 146
## [811] 145 145 145 144 143 143 143 142 141 141
## [821] 141 141 141 141 140 139 139 137 135 135
## [831] 134 133 133 132 132 131 131 130 130 130
## [841] 129 129 129 128 128 128 128 128 127 127
## [851] 127 126 126 125 125 125 125 125 124 124
## [861] 124 123 123 123 122 122 122 121 121 121
## [871] 121 120 120 120 119 118 118 118 117 117
## [881] 117 117 117 116 114 114 113 113 112 112
## [891] 111 111 111 110 110 110 109 109 108 108
## [901] 107 107 107 106 106 106 106 104 104 104
## [911] 104 103 103 103 102 101 101 100 100 100

```

2. Secondly, I would like to change columns ‘male_share’, ‘female_share’, and ‘gap’ to percentages.

```
round(unisexdata[, 'male_share'], digits = 2)
```

```

## [1] 0.58 0.51 0.48 0.42 0.34 0.56 0.43 0.48 0.35 0.37 0.33 0.65 0.62 0.37 0.64
## [16] 0.57 0.36 0.55 0.66 0.53 0.50 0.52 0.53 0.49 0.52 0.42 0.37 0.49 0.57 0.59
## [31] 0.34 0.41 0.57 0.61 0.61 0.56 0.46 0.45 0.34 0.52 0.48 0.42 0.54 0.64 0.41
## [46] 0.33 0.51 0.35 0.45 0.61 0.44 0.48 0.43 0.62 0.63 0.47 0.59 0.49 0.42 0.59
## [61] 0.35 0.49 0.66 0.59 0.38 0.57 0.36 0.34 0.35 0.59 0.56 0.45 0.56 0.55 0.45
## [76] 0.37 0.37 0.38 0.49 0.37 0.63 0.54 0.59 0.60 0.66 0.60 0.41 0.61 0.64 0.41
## [91] 0.66 0.57 0.43 0.39 0.40 0.64 0.62 0.35 0.39 0.55 0.41 0.65 0.56 0.55 0.60

```

```

## [106] 0.38 0.53 0.47 0.50 0.49 0.47 0.35 0.66 0.60 0.54 0.53 0.40 0.45 0.57 0.59
## [121] 0.56 0.40 0.47 0.45 0.57 0.51 0.66 0.64 0.62 0.50 0.60 0.58 0.66 0.59 0.55
## [136] 0.52 0.38 0.63 0.37 0.49 0.37 0.65 0.34 0.46 0.52 0.44 0.53 0.47 0.51 0.51
## [151] 0.52 0.48 0.49 0.60 0.47 0.55 0.58 0.42 0.55 0.60 0.55 0.66 0.58 0.64 0.56
## [166] 0.38 0.46 0.46 0.46 0.52 0.45 0.35 0.57 0.48 0.38 0.59 0.65 0.46 0.58 0.62
## [181] 0.46 0.44 0.37 0.37 0.39 0.39 0.62 0.64 0.55 0.49 0.53 0.62 0.52 0.60 0.53
## [196] 0.42 0.54 0.56 0.49 0.36 0.56 0.57 0.47 0.41 0.60 0.41 0.47 0.65 0.62 0.56
## [211] 0.48 0.66 0.61 0.59 0.36 0.40 0.34 0.41 0.39 0.38 0.50 0.34 0.67 0.60 0.54
## [226] 0.62 0.48 0.39 0.50 0.66 0.61 0.56 0.57 0.37 0.65 0.59 0.39 0.59 0.42 0.53
## [241] 0.51 0.55 0.45 0.47 0.47 0.54 0.45 0.65 0.56 0.51 0.37 0.41 0.46 0.35 0.67
## [256] 0.35 0.47 0.66 0.64 0.37 0.47 0.67 0.52 0.56 0.42 0.42 0.59 0.41 0.66 0.59
## [271] 0.35 0.49 0.33 0.59 0.63 0.57 0.54 0.67 0.66 0.52 0.61 0.37 0.48 0.55 0.62
## [286] 0.38 0.47 0.38 0.63 0.39 0.45 0.36 0.39 0.52 0.64 0.45 0.49 0.54 0.54 0.45
## [301] 0.61 0.55 0.64 0.57 0.46 0.60 0.37 0.48 0.45 0.63 0.56 0.50 0.64 0.48 0.65
## [316] 0.58 0.53 0.61 0.59 0.60 0.38 0.63 0.52 0.37 0.43 0.47 0.60 0.57 0.35 0.58
## [331] 0.57 0.56 0.55 0.59 0.38 0.41 0.34 0.44 0.36 0.34 0.61 0.57 0.61 0.49 0.62
## [346] 0.50 0.65 0.50 0.56 0.59 0.42 0.40 0.43 0.64 0.53 0.49 0.53 0.47 0.59 0.51
## [361] 0.42 0.49 0.43 0.63 0.37 0.40 0.51 0.58 0.34 0.40 0.58 0.65 0.42 0.53 0.40
## [376] 0.43 0.55 0.61 0.43 0.66 0.45 0.47 0.63 0.41 0.34 0.55 0.64 0.47 0.52 0.40
## [391] 0.67 0.41 0.48 0.37 0.50 0.40 0.47 0.45 0.64 0.63 0.57 0.60 0.63 0.35 0.54
## [406] 0.43 0.61 0.55 0.59 0.38 0.46 0.59 0.65 0.57 0.48 0.59 0.62 0.66 0.60 0.59
## [421] 0.50 0.46 0.59 0.66 0.66 0.60 0.62 0.54 0.54 0.41 0.37 0.50 0.37 0.43 0.39
## [436] 0.51 0.55 0.65 0.43 0.65 0.49 0.64 0.36 0.65 0.40 0.35 0.46 0.35 0.61 0.62
## [451] 0.43 0.58 0.48 0.47 0.66 0.65 0.37 0.49 0.41 0.55 0.35 0.58 0.66 0.58 0.66
## [466] 0.59 0.52 0.65 0.57 0.37 0.36 0.63 0.64 0.64 0.47 0.44 0.64 0.64 0.57 0.56
## [481] 0.54 0.55 0.39 0.65 0.47 0.58 0.60 0.35 0.54 0.66 0.67 0.48 0.66 0.39 0.34
## [496] 0.51 0.44 0.34 0.50 0.58 0.57 0.47 0.53 0.59 0.40 0.55 0.36 0.37 0.35 0.61
## [511] 0.51 0.58 0.52 0.49 0.56 0.56 0.66 0.49 0.61 0.39 0.45 0.58 0.65 0.61 0.39
## [526] 0.35 0.66 0.49 0.57 0.66 0.64 0.54 0.60 0.62 0.42 0.55 0.40 0.38 0.34 0.57
## [541] 0.59 0.45 0.39 0.35 0.52 0.62 0.65 0.37 0.38 0.51 0.63 0.63 0.33 0.41 0.41
## [556] 0.53 0.37 0.36 0.57 0.63 0.56 0.43 0.46 0.50 0.48 0.33 0.41 0.59 0.43 0.57
## [571] 0.46 0.64 0.63 0.58 0.42 0.63 0.34 0.34 0.47 0.60 0.56 0.43 0.36 0.51 0.49
## [586] 0.35 0.45 0.40 0.59 0.35 0.61 0.34 0.66 0.59 0.41 0.52 0.65 0.59 0.63 0.54
## [601] 0.56 0.37 0.51 0.61 0.37 0.46 0.61 0.43 0.65 0.50 0.36 0.58 0.51 0.51 0.36
## [616] 0.50 0.58 0.53 0.36 0.55 0.53 0.40 0.50 0.62 0.51 0.56 0.62 0.35 0.45 0.52
## [631] 0.46 0.34 0.35 0.55 0.50 0.39 0.62 0.64 0.38 0.33 0.51 0.41 0.41 0.61 0.36
## [646] 0.65 0.39 0.48 0.52 0.38 0.66 0.58 0.44 0.34 0.38 0.50 0.48 0.37 0.53 0.43
## [661] 0.42 0.60 0.40 0.39 0.56 0.60 0.51 0.67 0.66 0.59 0.54 0.34 0.64 0.56 0.40
## [676] 0.39 0.62 0.60 0.59 0.48 0.56 0.52 0.58 0.59 0.63 0.53 0.58 0.41 0.40 0.48
## [691] 0.46 0.51 0.53 0.56 0.38 0.37 0.55 0.54 0.67 0.64 0.43 0.61 0.56 0.49 0.63
## [706] 0.45 0.38 0.55 0.34 0.36 0.49 0.63 0.54 0.48 0.50 0.34 0.46 0.40 0.35 0.34
## [721] 0.48 0.58 0.66 0.41 0.56 0.36 0.64 0.56 0.35 0.47 0.38 0.53 0.63 0.48 0.44
## [736] 0.50 0.62 0.56 0.66 0.63 0.51 0.38 0.63 0.58 0.42 0.57 0.66 0.57 0.40 0.42
## [751] 0.48 0.36 0.65 0.66 0.64 0.37 0.67 0.48 0.54 0.46 0.54 0.34 0.58 0.45 0.52
## [766] 0.54 0.38 0.38 0.59 0.43 0.54 0.40 0.44 0.46 0.39 0.58 0.43 0.52 0.53 0.35
## [781] 0.63 0.51 0.37 0.42 0.43 0.57 0.66 0.57 0.63 0.46 0.58 0.33 0.56 0.63 0.42
## [796] 0.57 0.43 0.51 0.46 0.47 0.51 0.41 0.57 0.41 0.48 0.41 0.38 0.62 0.59 0.41
## [811] 0.54 0.51 0.47 0.60 0.50 0.36 0.61 0.52 0.57 0.46 0.35 0.65 0.46 0.56 0.42
## [826] 0.56 0.37 0.66 0.36 0.47 0.65 0.66 0.36 0.66 0.65 0.60 0.47 0.60 0.45 0.43
## [841] 0.59 0.57 0.52 0.57 0.43 0.36 0.64 0.66 0.61 0.51 0.47 0.54 0.65 0.39 0.55
## [856] 0.35 0.49 0.35 0.65 0.55 0.36 0.36 0.35 0.61 0.45 0.52 0.40 0.60 0.65 0.45
## [871] 0.34 0.51 0.35 0.35 0.41 0.64 0.38 0.61 0.55 0.64 0.61 0.60 0.66 0.62 0.56
## [886] 0.47 0.38 0.62 0.63 0.47 0.60 0.48 0.64 0.55 0.64 0.51 0.63 0.56 0.37 0.37
## [901] 0.41 0.34 0.46 0.40 0.64 0.38 0.55 0.63 0.46 0.42 0.39 0.58 0.56 0.42 0.57

```

```

## [916] 0.35 0.56 0.54 0.46

round(unisexdata[, 'female_share'], digits =2)

## [1] 0.42 0.49 0.52 0.58 0.66 0.44 0.57 0.52 0.65 0.63 0.67 0.35 0.38 0.63 0.36
## [16] 0.43 0.64 0.45 0.34 0.47 0.50 0.48 0.47 0.51 0.48 0.58 0.63 0.51 0.43 0.41
## [31] 0.66 0.59 0.43 0.39 0.39 0.44 0.54 0.55 0.66 0.48 0.52 0.58 0.46 0.36 0.59
## [46] 0.67 0.49 0.65 0.55 0.39 0.56 0.52 0.57 0.38 0.37 0.53 0.41 0.51 0.58 0.41
## [61] 0.65 0.51 0.34 0.41 0.62 0.43 0.64 0.66 0.65 0.41 0.44 0.55 0.44 0.45 0.55
## [76] 0.63 0.63 0.62 0.51 0.63 0.37 0.46 0.41 0.40 0.34 0.40 0.59 0.39 0.36 0.59
## [91] 0.34 0.43 0.57 0.61 0.60 0.36 0.38 0.65 0.61 0.45 0.59 0.35 0.44 0.45 0.40
## [106] 0.62 0.47 0.53 0.50 0.51 0.53 0.65 0.34 0.40 0.46 0.47 0.60 0.55 0.43 0.41
## [121] 0.44 0.60 0.53 0.55 0.43 0.49 0.34 0.36 0.38 0.50 0.40 0.42 0.34 0.41 0.45
## [136] 0.48 0.62 0.37 0.63 0.51 0.63 0.35 0.66 0.54 0.48 0.56 0.47 0.53 0.49 0.49
## [151] 0.48 0.52 0.51 0.40 0.53 0.45 0.42 0.58 0.45 0.40 0.45 0.34 0.42 0.36 0.44
## [166] 0.62 0.54 0.54 0.54 0.48 0.55 0.65 0.43 0.52 0.62 0.41 0.35 0.54 0.42 0.38
## [181] 0.54 0.56 0.63 0.63 0.61 0.61 0.38 0.36 0.45 0.51 0.47 0.38 0.48 0.40 0.47
## [196] 0.58 0.46 0.44 0.51 0.64 0.44 0.43 0.53 0.59 0.40 0.59 0.53 0.35 0.38 0.44
## [211] 0.52 0.34 0.39 0.41 0.64 0.60 0.66 0.59 0.61 0.62 0.50 0.66 0.33 0.40 0.46
## [226] 0.38 0.52 0.61 0.50 0.34 0.39 0.44 0.43 0.63 0.35 0.41 0.61 0.41 0.58 0.47
## [241] 0.49 0.45 0.55 0.53 0.53 0.46 0.55 0.35 0.44 0.49 0.63 0.59 0.54 0.65 0.33
## [256] 0.65 0.53 0.34 0.36 0.63 0.53 0.33 0.48 0.44 0.58 0.58 0.41 0.59 0.34 0.41
## [271] 0.65 0.51 0.67 0.41 0.37 0.43 0.46 0.33 0.34 0.48 0.39 0.63 0.52 0.45 0.38
## [286] 0.62 0.53 0.62 0.37 0.61 0.55 0.64 0.61 0.48 0.36 0.55 0.51 0.46 0.46 0.55
## [301] 0.39 0.45 0.36 0.43 0.54 0.40 0.63 0.52 0.55 0.37 0.44 0.50 0.36 0.52 0.35
## [316] 0.42 0.47 0.39 0.41 0.40 0.62 0.37 0.48 0.63 0.57 0.53 0.40 0.43 0.65 0.42
## [331] 0.43 0.44 0.45 0.41 0.62 0.59 0.66 0.56 0.64 0.66 0.39 0.43 0.39 0.51 0.38
## [346] 0.50 0.35 0.50 0.44 0.41 0.58 0.60 0.57 0.36 0.47 0.51 0.47 0.53 0.41 0.49
## [361] 0.58 0.51 0.57 0.37 0.63 0.60 0.49 0.42 0.66 0.60 0.42 0.35 0.58 0.47 0.60
## [376] 0.57 0.45 0.39 0.57 0.34 0.55 0.53 0.37 0.59 0.66 0.45 0.36 0.53 0.48 0.60
## [391] 0.33 0.59 0.52 0.63 0.50 0.60 0.53 0.55 0.36 0.37 0.43 0.40 0.37 0.65 0.46
## [406] 0.57 0.39 0.45 0.41 0.62 0.54 0.41 0.35 0.43 0.52 0.41 0.38 0.34 0.40 0.41
## [421] 0.50 0.54 0.41 0.34 0.34 0.40 0.38 0.46 0.46 0.59 0.63 0.50 0.63 0.57 0.61
## [436] 0.49 0.45 0.35 0.57 0.35 0.51 0.36 0.64 0.35 0.60 0.65 0.54 0.65 0.39 0.38
## [451] 0.57 0.42 0.52 0.53 0.34 0.35 0.63 0.51 0.59 0.45 0.65 0.42 0.34 0.42 0.34
## [466] 0.41 0.48 0.35 0.43 0.63 0.64 0.37 0.36 0.36 0.53 0.56 0.36 0.36 0.43 0.44
## [481] 0.46 0.45 0.61 0.35 0.53 0.42 0.40 0.65 0.46 0.34 0.33 0.52 0.34 0.61 0.66
## [496] 0.49 0.56 0.66 0.50 0.42 0.43 0.53 0.47 0.41 0.60 0.45 0.64 0.63 0.65 0.39
## [511] 0.49 0.42 0.48 0.51 0.44 0.44 0.34 0.51 0.39 0.61 0.55 0.42 0.35 0.39 0.61
## [526] 0.65 0.34 0.51 0.43 0.34 0.36 0.46 0.40 0.38 0.58 0.45 0.60 0.62 0.66 0.43
## [541] 0.41 0.55 0.61 0.65 0.48 0.38 0.35 0.63 0.62 0.49 0.37 0.37 0.67 0.59 0.59
## [556] 0.47 0.63 0.64 0.43 0.37 0.44 0.57 0.54 0.50 0.52 0.67 0.59 0.41 0.57 0.43
## [571] 0.54 0.36 0.37 0.42 0.58 0.37 0.66 0.66 0.53 0.40 0.44 0.57 0.64 0.49 0.51
## [586] 0.65 0.55 0.60 0.41 0.65 0.39 0.66 0.34 0.41 0.59 0.48 0.35 0.41 0.37 0.46
## [601] 0.44 0.63 0.49 0.39 0.63 0.54 0.39 0.57 0.35 0.50 0.64 0.42 0.49 0.49 0.64
## [616] 0.50 0.42 0.47 0.64 0.45 0.47 0.60 0.50 0.38 0.49 0.44 0.38 0.65 0.55 0.48
## [631] 0.54 0.66 0.65 0.45 0.50 0.61 0.38 0.36 0.62 0.67 0.49 0.59 0.59 0.39 0.64
## [646] 0.35 0.61 0.52 0.48 0.62 0.34 0.42 0.56 0.66 0.62 0.50 0.52 0.63 0.47 0.57
## [661] 0.58 0.40 0.60 0.61 0.44 0.40 0.49 0.33 0.34 0.41 0.46 0.66 0.36 0.44 0.60
## [676] 0.61 0.38 0.40 0.41 0.52 0.44 0.48 0.42 0.41 0.37 0.47 0.42 0.59 0.60 0.52
## [691] 0.54 0.49 0.47 0.44 0.62 0.63 0.45 0.46 0.33 0.36 0.57 0.39 0.44 0.51 0.37
## [706] 0.55 0.62 0.45 0.66 0.64 0.51 0.37 0.46 0.52 0.50 0.66 0.54 0.60 0.65 0.66
## [721] 0.52 0.42 0.34 0.59 0.44 0.64 0.36 0.44 0.65 0.53 0.62 0.47 0.37 0.52 0.56
## [736] 0.50 0.38 0.44 0.34 0.37 0.49 0.62 0.37 0.42 0.58 0.43 0.34 0.43 0.60 0.58

```

```

## [751] 0.52 0.64 0.35 0.34 0.36 0.63 0.33 0.52 0.46 0.54 0.46 0.66 0.42 0.55 0.48
## [766] 0.46 0.62 0.62 0.41 0.57 0.46 0.60 0.56 0.54 0.61 0.42 0.57 0.48 0.47 0.65
## [781] 0.37 0.49 0.63 0.58 0.57 0.43 0.34 0.43 0.37 0.54 0.42 0.67 0.44 0.37 0.58
## [796] 0.43 0.57 0.49 0.54 0.53 0.49 0.59 0.43 0.59 0.52 0.59 0.62 0.38 0.41 0.59
## [811] 0.46 0.49 0.53 0.40 0.50 0.64 0.39 0.48 0.43 0.54 0.65 0.35 0.54 0.44 0.58
## [826] 0.44 0.63 0.34 0.64 0.53 0.35 0.34 0.64 0.34 0.35 0.40 0.53 0.40 0.55 0.57
## [841] 0.41 0.43 0.48 0.43 0.57 0.64 0.36 0.34 0.39 0.49 0.53 0.46 0.35 0.61 0.45
## [856] 0.65 0.51 0.65 0.35 0.45 0.64 0.64 0.65 0.39 0.55 0.48 0.60 0.40 0.35 0.55
## [871] 0.66 0.49 0.65 0.65 0.59 0.36 0.62 0.39 0.45 0.36 0.39 0.40 0.34 0.38 0.44
## [886] 0.53 0.62 0.38 0.37 0.53 0.40 0.52 0.36 0.45 0.36 0.49 0.37 0.44 0.63 0.63
## [901] 0.59 0.66 0.54 0.60 0.36 0.62 0.45 0.37 0.54 0.58 0.61 0.42 0.44 0.58 0.43
## [916] 0.65 0.44 0.46 0.54

round(unisexdata[, 'gap'], digits = 2)

## [1] 0.17 0.02 0.04 0.16 0.33 0.12 0.13 0.03 0.30 0.26 0.33 0.29 0.25 0.26 0.27
## [16] 0.14 0.28 0.11 0.33 0.06 0.00 0.05 0.05 0.02 0.03 0.15 0.27 0.03 0.15 0.19
## [31] 0.32 0.17 0.14 0.22 0.22 0.12 0.07 0.09 0.32 0.03 0.04 0.15 0.09 0.28 0.17
## [46] 0.33 0.02 0.30 0.11 0.21 0.12 0.05 0.14 0.25 0.27 0.05 0.17 0.02 0.16 0.19
## [61] 0.30 0.02 0.32 0.19 0.25 0.14 0.27 0.32 0.31 0.18 0.12 0.10 0.11 0.09 0.10
## [76] 0.26 0.27 0.23 0.02 0.26 0.25 0.09 0.18 0.20 0.32 0.19 0.18 0.22 0.28 0.17
## [91] 0.32 0.14 0.14 0.22 0.20 0.28 0.24 0.29 0.22 0.10 0.18 0.30 0.11 0.11 0.21
## [106] 0.24 0.05 0.06 0.00 0.01 0.05 0.30 0.32 0.21 0.08 0.05 0.20 0.10 0.14 0.17
## [121] 0.11 0.20 0.07 0.09 0.14 0.03 0.32 0.29 0.24 0.00 0.20 0.15 0.32 0.18 0.10
## [136] 0.04 0.24 0.27 0.26 0.03 0.26 0.29 0.32 0.07 0.05 0.13 0.06 0.06 0.01 0.02
## [151] 0.05 0.04 0.01 0.21 0.06 0.10 0.15 0.16 0.09 0.19 0.10 0.32 0.15 0.28 0.13
## [166] 0.24 0.08 0.08 0.08 0.03 0.09 0.29 0.14 0.03 0.25 0.18 0.30 0.09 0.15 0.25
## [181] 0.08 0.13 0.26 0.27 0.21 0.21 0.24 0.27 0.10 0.01 0.05 0.25 0.04 0.20 0.05
## [196] 0.16 0.09 0.13 0.02 0.29 0.11 0.14 0.07 0.18 0.19 0.17 0.06 0.30 0.24 0.12
## [211] 0.03 0.33 0.22 0.17 0.27 0.19 0.31 0.18 0.22 0.24 0.00 0.32 0.33 0.19 0.09
## [226] 0.25 0.04 0.22 0.01 0.32 0.23 0.11 0.14 0.26 0.30 0.19 0.23 0.19 0.16 0.05
## [241] 0.01 0.10 0.09 0.07 0.05 0.08 0.10 0.31 0.11 0.03 0.26 0.18 0.08 0.29 0.33
## [256] 0.30 0.06 0.32 0.29 0.26 0.07 0.33 0.05 0.12 0.16 0.16 0.19 0.19 0.33 0.18
## [271] 0.30 0.03 0.33 0.18 0.26 0.15 0.09 0.33 0.32 0.04 0.22 0.27 0.03 0.10 0.24
## [286] 0.23 0.05 0.25 0.26 0.22 0.11 0.27 0.22 0.04 0.27 0.09 0.01 0.07 0.08 0.09
## [301] 0.23 0.09 0.28 0.15 0.07 0.20 0.26 0.04 0.09 0.26 0.11 0.00 0.28 0.04 0.29
## [316] 0.17 0.06 0.21 0.17 0.20 0.24 0.27 0.05 0.27 0.14 0.06 0.21 0.14 0.30 0.17
## [331] 0.15 0.13 0.09 0.17 0.24 0.18 0.32 0.12 0.29 0.32 0.22 0.14 0.22 0.03 0.23
## [346] 0.01 0.30 0.00 0.13 0.17 0.15 0.21 0.14 0.28 0.06 0.02 0.06 0.06 0.18 0.02
## [361] 0.16 0.02 0.13 0.26 0.27 0.20 0.01 0.16 0.32 0.20 0.16 0.31 0.15 0.07 0.21
## [376] 0.13 0.10 0.22 0.15 0.31 0.09 0.07 0.26 0.17 0.31 0.09 0.28 0.06 0.04 0.21
## [391] 0.33 0.19 0.05 0.26 0.00 0.20 0.06 0.10 0.28 0.26 0.14 0.21 0.27 0.31 0.07
## [406] 0.15 0.22 0.10 0.19 0.24 0.09 0.18 0.30 0.14 0.04 0.19 0.24 0.32 0.21 0.18
## [421] 0.00 0.09 0.18 0.32 0.32 0.20 0.24 0.08 0.08 0.18 0.25 0.00 0.25 0.14 0.22
## [436] 0.02 0.10 0.31 0.14 0.30 0.02 0.28 0.28 0.31 0.20 0.30 0.07 0.30 0.22 0.24
## [451] 0.13 0.17 0.03 0.05 0.32 0.29 0.25 0.01 0.19 0.11 0.29 0.16 0.32 0.17 0.33
## [466] 0.18 0.05 0.30 0.15 0.27 0.27 0.27 0.28 0.28 0.07 0.13 0.28 0.28 0.13 0.12
## [481] 0.08 0.11 0.22 0.30 0.06 0.17 0.21 0.30 0.08 0.33 0.33 0.03 0.32 0.22 0.32
## [496] 0.03 0.11 0.32 0.01 0.15 0.15 0.06 0.05 0.18 0.19 0.09 0.28 0.26 0.31 0.21
## [511] 0.03 0.15 0.04 0.03 0.13 0.12 0.32 0.03 0.23 0.23 0.10 0.15 0.30 0.22 0.23
## [526] 0.31 0.31 0.02 0.14 0.32 0.28 0.09 0.21 0.24 0.16 0.10 0.21 0.25 0.32 0.15
## [541] 0.18 0.10 0.21 0.30 0.04 0.25 0.31 0.27 0.24 0.03 0.27 0.25 0.33 0.17 0.18
## [556] 0.06 0.26 0.29 0.14 0.26 0.13 0.13 0.07 0.00 0.04 0.33 0.18 0.18 0.13 0.14
## [571] 0.09 0.28 0.26 0.17 0.15 0.26 0.33 0.32 0.06 0.20 0.13 0.15 0.28 0.02 0.03

```

```

## [586] 0.30 0.09 0.19 0.18 0.29 0.22 0.32 0.33 0.17 0.19 0.03 0.30 0.19 0.25 0.07
## [601] 0.12 0.27 0.02 0.23 0.26 0.07 0.22 0.15 0.29 0.00 0.28 0.17 0.01 0.02 0.27
## [616] 0.01 0.16 0.05 0.28 0.11 0.06 0.20 0.00 0.23 0.03 0.13 0.23 0.30 0.11 0.03
## [631] 0.09 0.32 0.29 0.10 0.00 0.22 0.25 0.28 0.24 0.33 0.01 0.18 0.19 0.23 0.28
## [646] 0.29 0.22 0.05 0.04 0.23 0.32 0.17 0.12 0.32 0.25 0.01 0.04 0.25 0.06 0.14
## [661] 0.15 0.20 0.21 0.22 0.12 0.20 0.02 0.33 0.32 0.17 0.08 0.31 0.28 0.13 0.21
## [676] 0.21 0.24 0.20 0.18 0.04 0.12 0.04 0.16 0.17 0.27 0.07 0.17 0.18 0.20 0.03
## [691] 0.08 0.02 0.07 0.11 0.24 0.27 0.11 0.08 0.33 0.28 0.14 0.23 0.12 0.02 0.25
## [706] 0.11 0.23 0.09 0.31 0.28 0.03 0.26 0.07 0.05 0.00 0.32 0.08 0.20 0.30 0.33
## [721] 0.05 0.17 0.32 0.17 0.12 0.27 0.28 0.12 0.31 0.07 0.24 0.06 0.25 0.05 0.12
## [736] 0.01 0.24 0.12 0.33 0.25 0.03 0.23 0.27 0.15 0.15 0.13 0.33 0.15 0.20 0.16
## [751] 0.04 0.27 0.30 0.33 0.28 0.25 0.33 0.05 0.08 0.08 0.09 0.33 0.15 0.10 0.04
## [766] 0.09 0.25 0.24 0.18 0.14 0.09 0.20 0.12 0.09 0.23 0.16 0.14 0.03 0.06 0.31
## [781] 0.26 0.01 0.27 0.17 0.13 0.13 0.32 0.14 0.26 0.08 0.17 0.33 0.12 0.25 0.16
## [796] 0.13 0.13 0.01 0.09 0.06 0.01 0.17 0.14 0.18 0.03 0.18 0.25 0.25 0.18 0.17
## [811] 0.09 0.02 0.07 0.19 0.00 0.27 0.23 0.05 0.13 0.07 0.30 0.30 0.09 0.13 0.16
## [826] 0.13 0.26 0.32 0.28 0.05 0.29 0.33 0.28 0.32 0.31 0.21 0.05 0.19 0.10 0.14
## [841] 0.18 0.14 0.04 0.13 0.15 0.27 0.27 0.32 0.23 0.02 0.06 0.09 0.29 0.22 0.10
## [856] 0.30 0.02 0.30 0.31 0.10 0.28 0.27 0.31 0.23 0.11 0.04 0.21 0.20 0.30 0.10
## [871] 0.33 0.03 0.31 0.31 0.18 0.28 0.24 0.22 0.09 0.27 0.22 0.20 0.31 0.23 0.11
## [886] 0.05 0.25 0.25 0.26 0.05 0.20 0.04 0.27 0.10 0.28 0.03 0.25 0.13 0.27 0.27
## [901] 0.19 0.32 0.08 0.20 0.29 0.24 0.09 0.26 0.08 0.16 0.21 0.15 0.13 0.16 0.15
## [916] 0.29 0.12 0.09 0.07

```

```

unisexdata$male_share <- unisexdata$male_share*100
unisexdata$female_share<-unisexdata$female_share*100
unisexdata$gap<-unisexdata$gap*100

```

3. Next, I would like to rename all columns of the dataset except for total to better reflect the underlying data.

```

colnames(unisexdata)[colnames(unisexdata)=='X'] <- 'No.'
colnames(unisexdata)[colnames(unisexdata)=='male_share'] <- 'Male %'
colnames(unisexdata)[colnames(unisexdata)=='female_share'] <- 'Female %'
colnames(unisexdata)[colnames(unisexdata)=='gap'] <- 'Difference'

```

#Characteristics of the data (Section 4)

This dataframe has r nrow(unisexdata)rows and r ncol(unisexdata)columns. The names of the columns and a brief description of each are in the table below: I wanted to include a table using Markdown directly or kable from the knitr package with 2 columns.

```
library(knitr)
```

```
## Warning: package 'knitr' was built under R version 4.1.3
```

```
kable(summary(unisexdata), caption="The descriptive summary statistics of each column in the dataset")
```

Table 1: The descriptive summary statistics of each column in the dataset

No.	name	total	Male %	Female %	Difference
Min. : 1.0	Length:919	Min. : 100.2	Min. :33.36	Min. :33.37	Min. : 0.00717
1st Qu.:230.5	Class :character	1st Qu.: 224.7	1st Qu.:41.86	1st Qu.:40.81	1st Qu.: 8.95653
Median :460.0	Mode :character	Median : 489.6	Median :51.42	Median :48.58	Median :17.56914
Mean :460.0	NA	Mean : 3142.0	Mean :50.72	Mean :49.28	Mean :17.38627
3rd Qu.:689.5	NA	3rd Qu.: 1316.1	3rd Qu.:59.19	Qu.:58.14	Qu.:25.95527
Max. :919.0	NA	Max. :176544.3	Max. :66.63	Max. :66.64	Max. :33.28785

```
kable(head(unisexdata,n=10), caption="The first 10 rows of each column in the dataset")
```

Table 2: The first 10 rows of each column in the dataset

No.	name	total	Male %	Female %	Difference
1	Casey	176544.33	58.42866	41.57134	16.857313
2	Riley	154860.67	50.76391	49.23609	1.527814
3	Jessie	136381.83	47.78343	52.21657	4.433146
4	Jackie	132928.79	42.11326	57.88674	15.773480
5	Avery	121797.42	33.52131	66.47869	32.957385
6	Jaime	109870.19	56.17929	43.82071	12.358580
7	Peyton	94896.40	43.37194	56.62806	13.256125
8	Kerry	88963.93	48.39488	51.60512	3.210231
9	Jody	80400.52	35.20680	64.79320	29.586394
10	Kendall	79210.87	37.23667	62.76333	25.526652

Here is the table a column name in the datafram with a very brief description of what each column measures.

```
text_tbl <- data.frame(
  Names = c("name", "total", "Male %", "Female %", "Difference"),
  Description = c(
    "First names from SSA",
    "Total number of Americans using the name",
    "Percentage of male counterparts with the name",
    "Percentage of female counterparts with the name",
    "Difference between Male % and Female %"
  )
)
text_tbl
```

##	Names	Description
## 1	name	First names from SSA
## 2	total	Total number of Americans using the name
## 3	Male %	Percentage of male counterparts with the name
## 4	Female %	Percentage of female counterparts with the name
## 5	Difference	Difference between Male % and Female %

```
#Summary statistics (Section 5)
```

I wanted to check if there are any missing values. I used function colSums to see the number of total missing values in each column. The output indicates that there is no missing values. Hence, the summary function can be used without removing any columns.

```
colSums(is.na(unisexdata))
```

```
##      No.     name    total   Male % Female % Difference
##      0       0       0       0       0       0       0
```

I wanted to pick three columns to use summary function

```
data_pick3 <- select(unisexdata,'total','Male %','Female %')
```

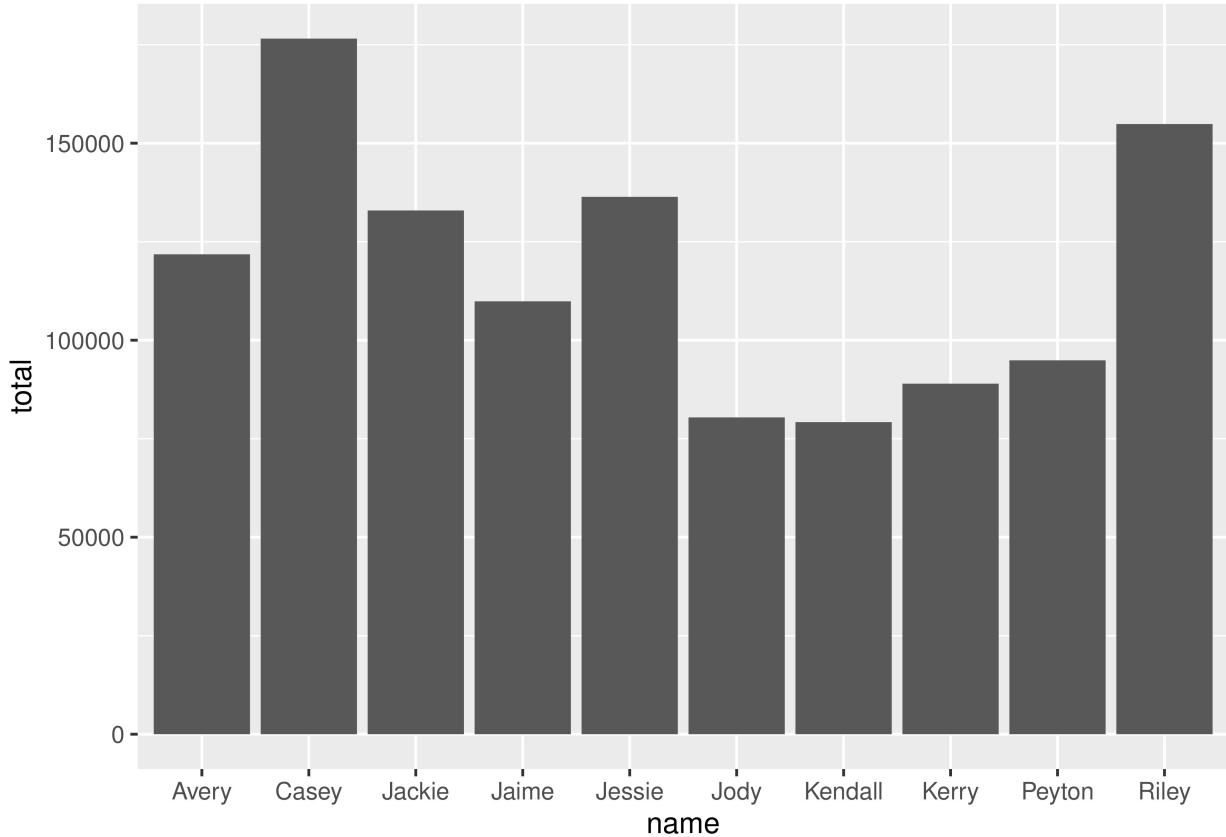
```
Summarytable<-summary(unisexdata)
print(Summarytable)
```

```
##      No.     name    total   Male %
##  Min. : 1.0 Length:919   Min. : 100.2 Min. :33.36
##  1st Qu.:230.5 Class :character 1st Qu.: 224.7 1st Qu.:41.86
##  Median :460.0 Mode  :character Median : 489.6 Median :51.42
##  Mean   :460.0                   Mean   : 3142.0 Mean  :50.72
##  3rd Qu.:689.5                   3rd Qu.: 1316.1 3rd Qu.:59.19
##  Max.   :919.0                   Max.  :176544.3 Max. :66.63
##      Female %   Difference
##  Min.   :33.37 Min.   : 0.00717
##  1st Qu.:40.81 1st Qu.: 8.95653
##  Median :48.58 Median :17.56914
##  Mean   :49.28 Mean   :17.38627
##  3rd Qu.:58.14 3rd Qu.:25.95527
##  Max.   :66.64 Max.   :33.28785
```

Top 10 unisex names that are used in America.

```
# I would like to showcase the Top 10 names that are the most commonly used in America
commonnames <- unisexdata %>%
  filter(row_number(desc(total))<=10)

library(ggplot2)
ggplot(data = commonnames) +
  geom_bar(mapping = aes(x = `name`, y = `total`), stat = "identity")
```



Bottom 10 unisex names that are used in America.

```
#I would like to showcase the Bottom 10 names that are not commonly used in America
uncommonnames <- unisexdata %>%
  filter(row_number(total) <= 10)

library(ggplot2)
ggplot(data = uncommonnames) +
  geom_bar(mapping = aes(x = `name`, y = `total`), stat = "identity")
```

