Chapter 4 Examples

Brooke Anderson

2/27/2020

```
library(tidyverse)
library(purrr)
```

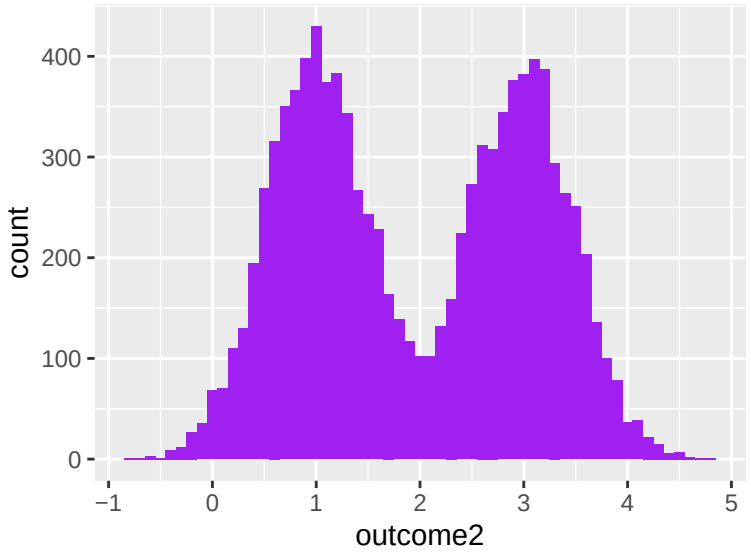
Mixture of two normals

```
coin_flips <- tibble(outcome1 = rbinom(n = 10000, size = 1, prob = 0.5)) %>%
  mutate(mean_norm = if_else(outcome1 == 0, 1, 3)) %>%
  mutate(outcome2 = map_dbl(mean_norm, ~ rnorm(n = 1, mean = .x, sd = 0.5)))

head(coin_flips)
```

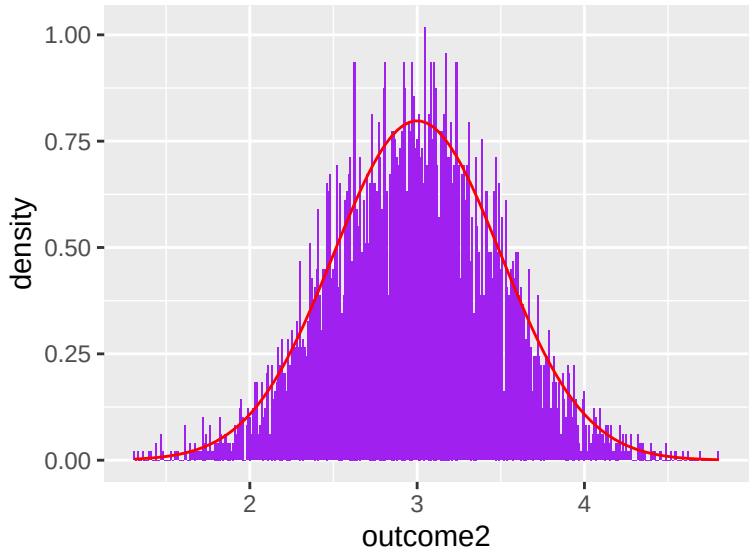
```
## # A tibble: 6 x 3
##   outcome1 mean_norm outcome2
##      <int>     <dbl>    <dbl>
## 1        0         1    1.10
## 2        0         1    1.40
## 3        1         3    3.95
## 4        0         1    1.99
## 5        0         1    0.160
## 6        0         1    0.807
```

```
ggplot(coin_flips, aes(x = outcome2)) +
  geom_histogram(fill = "purple", binwidth = 0.1)
```



Plot just for "heads" on initial coin flip:

```
ggplot(filter(coin_flips, outcome1 == 1), aes(x = outcome2)) +
  geom_histogram(aes(y = ..density..), fill = "purple", binwidth = 0.01) +
  stat_function(fun = dnorm, args = list(mean = 3, sd = 0.5), color = "red")
```



Two coins, with probs of heads of 0.125 and 0.25.

1/8 prob. of picking coin 1:

```
coin_flips2 <- tibble(outcome1 = rbinom(n = 100, size = 1, prob = 1 / 8)) %>%
  mutate(prob_2 = if_else(outcome1 == 0, 0.125, 0.25)) %>%
  mutate(outcome2 = map_dbl(prob_2, ~ sum(rbinom(n = 2, size = 1, prob = .x))))

head(coin_flips2)
```

```
## # A tibble: 6 x 3
##   outcome1 prob_2 outcome2
##      <int>  <dbl>    <dbl>
## 1        1  0.25         1
## 2        0  0.125        0
## 3        0  0.125        0
## 4        0  0.125        0
## 5        0  0.125        0
## 6        0  0.125        0
```

```
coin_flips2 %>%
  group_by(outcome2) %>%
  count()
```

```
## # A tibble: 3 x 2
## # Groups:   outcome2 [3]
##   outcome2     n
##      <dbl> <int>
## 1        0    82
## 2        1    16
## 3        2     2
```

1/4 prob. of picking coin 1:

```
coin_flips2 <- tibble(outcome1 = rbinom(n = 100, size = 1, prob = 1 / 4)) %>%
  mutate(prob_2 = if_else(outcome1 == 0, 0.125, 0.25)) %>%
  mutate(outcome2 = map_dbl(prob_2, ~ sum(rbinom(n = 2, size = 1, prob = .x))))

head(coin_flips2)
```

```
## # A tibble: 6 x 3
##   outcome1 prob_2 outcome2
##      <int>  <dbl>    <dbl>
## 1        0  0.125        0
## 2        0  0.125        0
## 3        1  0.25         1
## 4        0  0.125        1
## 5        0  0.125        0
## 6        0  0.125        2
```

```
coin_flips2 %>%
  group_by(outcome2) %>%
  count()
```

```
## # A tibble: 3 x 2
## # Groups:   outcome2 [3]
##   outcome2     n
##      <dbl> <int>
## 1        0    77
## 2        1    20
## 3        2     3
```

Mixture of two normals with mean parameters unknown, standard deviation of 1:

```
two_norms <- tibble(u = sample(2, 100, replace = TRUE)) %>%
  mutate(mu = if_else(u == 1, -0.5, 1.5)) %>%
  mutate(y = map_dbl(mu, ~ rnorm(n = 1, mean = .x, sd = 1)))

head(two_norms)
```

```
## # A tibble: 6 x 3
##       u    mu      y
##   <int> <dbl>  <dbl>
## 1     2   1.5  1.11
## 2     1  -0.5 -0.237
## 3     1  -0.5  0.419
## 4     1  -0.5 -1.64
## 5     2   1.5  0.417
## 6     2   1.5  3.56
```

```
two_norms %>%
  group_by(u) %>%
  summarize(n(),
            mean(y))
```

```
## # A tibble: 2 x 3
##       u `n()` `mean(y)`
##   <int> <int>     <dbl>
## 1     1    54    -0.550
## 2     2    46     1.43
```

```r
library("mixtools")
gm <- normalmixEM(two_norms$y, k = 2, lambda = c(0.5, 0.5),
                  mu = c(-0.01, 0.01), sigma = c(1, 1))
```

```
## number of iterations= 589
```

```r
gm
```

```
## $x
##   [1]  1.114969083 -0.237377508  0.419131861 -1.641665882  0.416734692
##   [6]  3.560189754  2.014466693 -1.348484064 -0.668900868  0.323761878
##  [11] -2.062943256  1.231862211  3.034630294  0.726562156 -0.839246421
##  [16]  0.064212944 -0.685752754 -0.552683873 -0.334355293  2.758706543
##  [21]  0.919049832 -1.597107626  0.001112722  0.882887949 -1.682874570
##  [26] -0.619745967  2.112393229  0.978058822  0.789745159 -1.898032045
##  [31]  1.730520808 -1.370068472  1.401421436 -2.024884308 -1.043908946
##  [36]  1.970138899 -0.802383138  0.433892107  1.054301340 -0.935552147
##  [41] -0.259858448 -0.270546092 -1.255056860 -0.223189703  1.388418861
##  [46] -0.426655313  0.753170090  1.641574614  2.047021700 -0.498870165
##  [51]  2.335764524  0.125063594  0.877747828  0.965965344  1.038444549
##  [56]  2.717682300 -1.141051417  1.613353772 -0.091317013  0.156568713
##  [61]  1.145782112 -0.337768176  0.091076174 -1.653433242 -0.488890208
##  [66] -0.564034956  0.136434406 -0.092509686 -1.742838189  0.082836761
##  [71] -1.101097588 -0.081888839  0.689207841 -0.182144863 -0.051943114
##  [76]  0.045643644 -0.939997528  1.095975524  0.290931116  1.868954132
##  [81]  1.002827416  1.854304387  3.043613723  0.272559706 -1.700284097
##  [86] -1.909196179  1.730795624  1.331238239  1.585618247  1.976140973
##  [91]  1.466888947 -0.462335821  1.444047782  0.701762442  2.062787826
##  [96]  2.887550005  0.395172935  0.711743138 -0.479819391  1.053991990
##
## $lambda
## [1] 0.2977096 0.7022904
##
## $mu
## [1] -0.8656287  0.8833632
##
## $sigma
## [1] 0.7461086 1.1170704
##
## $loglik
## [1] -166.5423
##
## $posterior
##              comp.1    comp.2
##   [1,] 1.877062e-02 0.9812294
##   [2,] 4.241263e-01 0.5758737
##   [3,] 1.357749e-01 0.8642251
##   [4,] 8.262326e-01 0.1737674
##   [5,] 1.365301e-01 0.8634699
##   [6,] 2.562626e-07 0.9999997
##   [7,] 6.155400e-04 0.9993845
##   [8,] 7.911465e-01 0.2088535
##   [9,] 6.168210e-01 0.3831790
##  [10,] 1.680131e-01 0.8319869
##  [11,] 8.501837e-01 0.1498163
```

```
##  [12,] 1.264825e-02 0.9873518
##  [13,] 4.721063e-06 0.9999953
##  [14,] 6.170223e-02 0.9382978
##  [15,] 6.756260e-01 0.3243740
##  [16,] 2.764090e-01 0.7235910
##  [17,] 6.231298e-01 0.3768702
##  [18,] 5.704575e-01 0.4295425
##  [19,] 4.715334e-01 0.5284666
##  [20,] 1.952434e-05 0.9999805
##  [21,] 3.506307e-02 0.9649369
##  [22,] 8.220243e-01 0.1779757
##  [23,] 3.062937e-01 0.6937063
##  [24,] 3.914067e-02 0.9608593
##  [25,] 8.298060e-01 0.1701940
##  [26,] 5.978121e-01 0.4021879
##  [27,] 4.034538e-04 0.9995965
##  [28,] 2.919471e-02 0.9708053
##  [29,] 5.154463e-02 0.9484554
##  [30,] 8.439390e-01 0.1560610
##  [31,] 1.983651e-03 0.9980163
##  [32,] 7.943749e-01 0.2056251
##  [33,] 6.941073e-03 0.9930589
##  [34,] 8.490691e-01 0.1509309
##  [35,] 7.320666e-01 0.2679334
##  [36,] 7.428943e-04 0.9992571
##  [37,] 6.638348e-01 0.3361652
##  [38,] 1.311875e-01 0.8688125
##  [39,] 2.289321e-02 0.9771068
##  [40,] 7.040459e-01 0.2959541
##  [41,] 4.352295e-01 0.5647705
##  [42,] 4.404881e-01 0.5595119
##  [43,] 7.758154e-01 0.2241846
##  [44,] 4.170946e-01 0.5829054
##  [45,] 7.276057e-03 0.9927239
##  [46,] 5.149783e-01 0.4850217
##  [47,] 5.724068e-02 0.9427593
##  [48,] 2.813978e-03 0.9971860
##  [49,] 5.354629e-04 0.9994645
##  [50,] 5.473812e-01 0.4526188
##  [51,] 1.484939e-04 0.9998515
##  [52,] 2.486597e-01 0.7513403
##  [53,] 3.975197e-02 0.9602480
##  [54,] 3.032243e-02 0.9696776
##  [55,] 2.409492e-02 0.9759051
##  [56,] 2.395701e-05 0.9999760
##  [57,] 7.538704e-01 0.2461296
##  [58,] 3.138775e-03 0.9968612
##  [59,] 3.514907e-01 0.6485093
##  [60,] 2.347861e-01 0.7652139
##  [61,] 1.694161e-02 0.9830584
##  [62,] 4.731738e-01 0.5268262
##  [63,] 2.640140e-01 0.7359860
##  [64,] 8.272836e-01 0.1727164
##  [65,] 5.429975e-01 0.4570025
```

```
##  [66,] 5.751997e-01 0.4248003
##  [67,] 2.436109e-01 0.7563891
##  [68,] 3.520813e-01 0.6479187
##  [69,] 8.344846e-01 0.1655154
##  [70,] 2.677926e-01 0.7322074
##  [71,] 7.452653e-01 0.2547347
##  [72,] 3.468271e-01 0.6531729
##  [73,] 6.843968e-02 0.9315603
##  [74,] 3.966806e-01 0.6033194
##  [75,] 3.320814e-01 0.6679186
##  [76,] 2.850986e-01 0.7149014
##  [77,] 7.052762e-01 0.2947238
##  [78,] 1.998392e-02 0.9800161
##  [79,] 1.801336e-01 0.8198664
##  [80,] 1.132735e-03 0.9988673
##  [81,] 2.699871e-02 0.9730013
##  [82,] 1.203040e-03 0.9987970
##  [83,] 4.502085e-06 0.9999955
##  [84,] 1.871373e-01 0.8128627
##  [85,] 8.312268e-01 0.1687732
##  [86,] 8.444806e-01 0.1555194
##  [87,] 1.981484e-03 0.9980185
##  [88,] 8.932351e-03 0.9910676
##  [89,] 3.491675e-03 0.9965083
##  [90,] 7.243008e-04 0.9992757
##  [91,] 5.459627e-03 0.9945404
##  [92,] 5.311825e-01 0.4688175
##  [93,] 5.939725e-03 0.9940603
##  [94,] 6.611203e-02 0.9338880
##  [95,] 5.003231e-04 0.9994997
##  [96,] 1.015730e-05 0.9999898
##  [97,] 1.434517e-01 0.8565483
##  [98,] 6.430756e-02 0.9356924
##  [99,] 5.389860e-01 0.4610140
## [100,] 2.291614e-02 0.9770839
##
## $all.loglik
##   [1] -182.5484 -167.8822 -167.8822 -167.8821 -167.8821 -167.8820 -167.8820
##   [8] -167.8820 -167.8819 -167.8819 -167.8818 -167.8818 -167.8817 -167.8817
##  [15] -167.8816 -167.8815 -167.8814 -167.8814 -167.8813 -167.8812 -167.8811
##  [22] -167.8810 -167.8808 -167.8807 -167.8806 -167.8804 -167.8802 -167.8801
##  [29] -167.8799 -167.8797 -167.8794 -167.8792 -167.8789 -167.8787 -167.8784
##  [36] -167.8780 -167.8777 -167.8773 -167.8769 -167.8764 -167.8759 -167.8754
##  [43] -167.8749 -167.8743 -167.8736 -167.8729 -167.8721 -167.8713 -167.8704
##  [50] -167.8695 -167.8685 -167.8674 -167.8662 -167.8649 -167.8635 -167.8620
##  [57] -167.8604 -167.8587 -167.8568 -167.8548 -167.8527 -167.8504 -167.8479
##  [64] -167.8452 -167.8424 -167.8393 -167.8360 -167.8324 -167.8286 -167.8245
##  [71] -167.8201 -167.8154 -167.8103 -167.8048 -167.7989 -167.7926 -167.7858
##  [78] -167.7785 -167.7706 -167.7621 -167.7529 -167.7429 -167.7321 -167.7204
##  [85] -167.7077 -167.6939 -167.6788 -167.6623 -167.6443 -167.6244 -167.6025
##  [92] -167.5783 -167.5515 -167.5218 -167.4888 -167.4522 -167.4114 -167.3662
##  [99] -167.3162 -167.2614 -167.2018 -167.1380 -167.0710 -167.0025 -166.9346
## [106] -166.8698 -166.8106 -166.7591 -166.7164 -166.6827 -166.6573 -166.6389
## [113] -166.6260 -166.6171 -166.6111 -166.6070 -166.6042 -166.6021 -166.6005
```

```
## [120] -166.5992 -166.5982 -166.5972 -166.5963 -166.5954 -166.5946 -166.5938
## [127] -166.5930 -166.5922 -166.5914 -166.5906 -166.5899 -166.5891 -166.5884
## [134] -166.5877 -166.5870 -166.5862 -166.5855 -166.5849 -166.5842 -166.5835
## [141] -166.5828 -166.5822 -166.5815 -166.5808 -166.5802 -166.5796 -166.5790
## [148] -166.5783 -166.5777 -166.5771 -166.5765 -166.5759 -166.5754 -166.5748
## [155] -166.5742 -166.5737 -166.5731 -166.5726 -166.5720 -166.5715 -166.5710
## [162] -166.5705 -166.5700 -166.5695 -166.5690 -166.5685 -166.5680 -166.5675
## [169] -166.5671 -166.5666 -166.5662 -166.5657 -166.5653 -166.5648 -166.5644
## [176] -166.5640 -166.5636 -166.5632 -166.5628 -166.5624 -166.5620 -166.5616
## [183] -166.5612 -166.5608 -166.5605 -166.5601 -166.5598 -166.5594 -166.5591
## [190] -166.5587 -166.5584 -166.5581 -166.5577 -166.5574 -166.5571 -166.5568
## [197] -166.5565 -166.5562 -166.5559 -166.5556 -166.5554 -166.5551 -166.5548
## [204] -166.5545 -166.5543 -166.5540 -166.5538 -166.5535 -166.5533 -166.5530
## [211] -166.5528 -166.5526 -166.5524 -166.5521 -166.5519 -166.5517 -166.5515
## [218] -166.5513 -166.5511 -166.5509 -166.5507 -166.5505 -166.5503 -166.5501
## [225] -166.5500 -166.5498 -166.5496 -166.5495 -166.5493 -166.5491 -166.5490
## [232] -166.5488 -166.5487 -166.5485 -166.5484 -166.5482 -166.5481 -166.5480
## [239] -166.5478 -166.5477 -166.5476 -166.5474 -166.5473 -166.5472 -166.5471
## [246] -166.5470 -166.5469 -166.5468 -166.5466 -166.5465 -166.5464 -166.5463
## [253] -166.5462 -166.5461 -166.5461 -166.5460 -166.5459 -166.5458 -166.5457
## [260] -166.5456 -166.5455 -166.5455 -166.5454 -166.5453 -166.5452 -166.5452
## [267] -166.5451 -166.5450 -166.5449 -166.5449 -166.5448 -166.5448 -166.5447
## [274] -166.5446 -166.5446 -166.5445 -166.5445 -166.5444 -166.5444 -166.5443
## [281] -166.5443 -166.5442 -166.5442 -166.5441 -166.5441 -166.5440 -166.5440
## [288] -166.5439 -166.5439 -166.5439 -166.5438 -166.5438 -166.5437 -166.5437
## [295] -166.5437 -166.5436 -166.5436 -166.5436 -166.5435 -166.5435 -166.5435
## [302] -166.5434 -166.5434 -166.5434 -166.5434 -166.5433 -166.5433 -166.5433
## [309] -166.5432 -166.5432 -166.5432 -166.5432 -166.5432 -166.5431 -166.5431
## [316] -166.5431 -166.5431 -166.5431 -166.5430 -166.5430 -166.5430 -166.5430
## [323] -166.5430 -166.5429 -166.5429 -166.5429 -166.5429 -166.5429 -166.5429
## [330] -166.5428 -166.5428 -166.5428 -166.5428 -166.5428 -166.5428 -166.5428
## [337] -166.5428 -166.5427 -166.5427 -166.5427 -166.5427 -166.5427 -166.5427
## [344] -166.5427 -166.5427 -166.5427 -166.5427 -166.5426 -166.5426 -166.5426
## [351] -166.5426 -166.5426 -166.5426 -166.5426 -166.5426 -166.5426 -166.5426
## [358] -166.5426 -166.5426 -166.5426 -166.5425 -166.5425 -166.5425 -166.5425
## [365] -166.5425 -166.5425 -166.5425 -166.5425 -166.5425 -166.5425 -166.5425
## [372] -166.5425 -166.5425 -166.5425 -166.5425 -166.5425 -166.5425 -166.5425
## [379] -166.5425 -166.5425 -166.5425 -166.5425 -166.5424 -166.5424 -166.5424
## [386] -166.5424 -166.5424 -166.5424 -166.5424 -166.5424 -166.5424 -166.5424
## [393] -166.5424 -166.5424 -166.5424 -166.5424 -166.5424 -166.5424 -166.5424
## [400] -166.5424 -166.5424 -166.5424 -166.5424 -166.5424 -166.5424 -166.5424
## [407] -166.5424 -166.5424 -166.5424 -166.5424 -166.5424 -166.5424 -166.5424
## [414] -166.5424 -166.5424 -166.5424 -166.5424 -166.5424 -166.5424 -166.5424
## [421] -166.5424 -166.5424 -166.5424 -166.5424 -166.5424 -166.5424 -166.5424
## [428] -166.5424 -166.5424 -166.5424 -166.5424 -166.5424 -166.5424 -166.5424
## [435] -166.5424 -166.5424 -166.5424 -166.5424 -166.5424 -166.5424 -166.5424
## [442] -166.5423 -166.5423 -166.5423 -166.5423 -166.5423 -166.5423 -166.5423
## [449] -166.5423 -166.5423 -166.5423 -166.5423 -166.5423 -166.5423 -166.5423
## [456] -166.5423 -166.5423 -166.5423 -166.5423 -166.5423 -166.5423 -166.5423
## [463] -166.5423 -166.5423 -166.5423 -166.5423 -166.5423 -166.5423 -166.5423
## [470] -166.5423 -166.5423 -166.5423 -166.5423 -166.5423 -166.5423 -166.5423
## [477] -166.5423 -166.5423 -166.5423 -166.5423 -166.5423 -166.5423 -166.5423
## [484] -166.5423 -166.5423 -166.5423 -166.5423 -166.5423 -166.5423 -166.5423
## [491] -166.5423 -166.5423 -166.5423 -166.5423 -166.5423 -166.5423 -166.5423
```

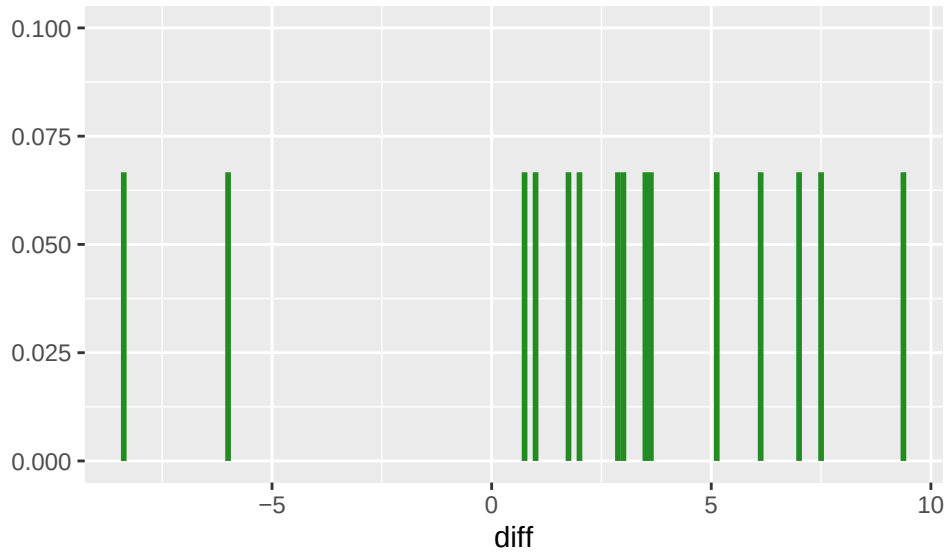
```
## [498] -166.5423 -166.5423 -166.5423 -166.5423 -166.5423 -166.5423 -166.5423
## [505] -166.5423 -166.5423 -166.5423 -166.5423 -166.5423 -166.5423 -166.5423
## [512] -166.5423 -166.5423 -166.5423 -166.5423 -166.5423 -166.5423 -166.5423
## [519] -166.5423 -166.5423 -166.5423 -166.5423 -166.5423 -166.5423 -166.5423
## [526] -166.5423 -166.5423 -166.5423 -166.5423 -166.5423 -166.5423 -166.5423
## [533] -166.5423 -166.5423 -166.5423 -166.5423 -166.5423 -166.5423 -166.5423
## [540] -166.5423 -166.5423 -166.5423 -166.5423 -166.5423 -166.5423 -166.5423
## [547] -166.5423 -166.5423 -166.5423 -166.5423 -166.5423 -166.5423 -166.5423
## [554] -166.5423 -166.5423 -166.5423 -166.5423 -166.5423 -166.5423 -166.5423
## [561] -166.5423 -166.5423 -166.5423 -166.5423 -166.5423 -166.5423 -166.5423
## [568] -166.5423 -166.5423 -166.5423 -166.5423 -166.5423 -166.5423 -166.5423
## [575] -166.5423 -166.5423 -166.5423 -166.5423 -166.5423 -166.5423 -166.5423
## [582] -166.5423 -166.5423 -166.5423 -166.5423 -166.5423 -166.5423 -166.5423
## [589] -166.5423 -166.5423
##
## $restarts
## [1] 0
##
## $ft
## [1] "normalmixEM"
##
## attr(,"class")
## [1] "mixEM"
```

Zea mays example

```
library("HistData")
head(ZeaMays)
```

```
##   pair pot  cross   self   diff
## 1    1   1 23.500 17.375  6.125
## 2    2   1 12.000 20.375 -8.375
## 3    3   1 21.000 20.000  1.000
## 4    4   2 22.000 20.000  2.000
## 5    5   2 19.125 18.375  0.750
## 6    6   2 21.500 18.625  2.875
```

```
ggplot(ZeaMays, aes(x = diff, ymax = 1 / 15, ymin = 0)) +
  geom_linerange(size = 1, col = "forestgreen") +
  ylim(0, 0.1)
```

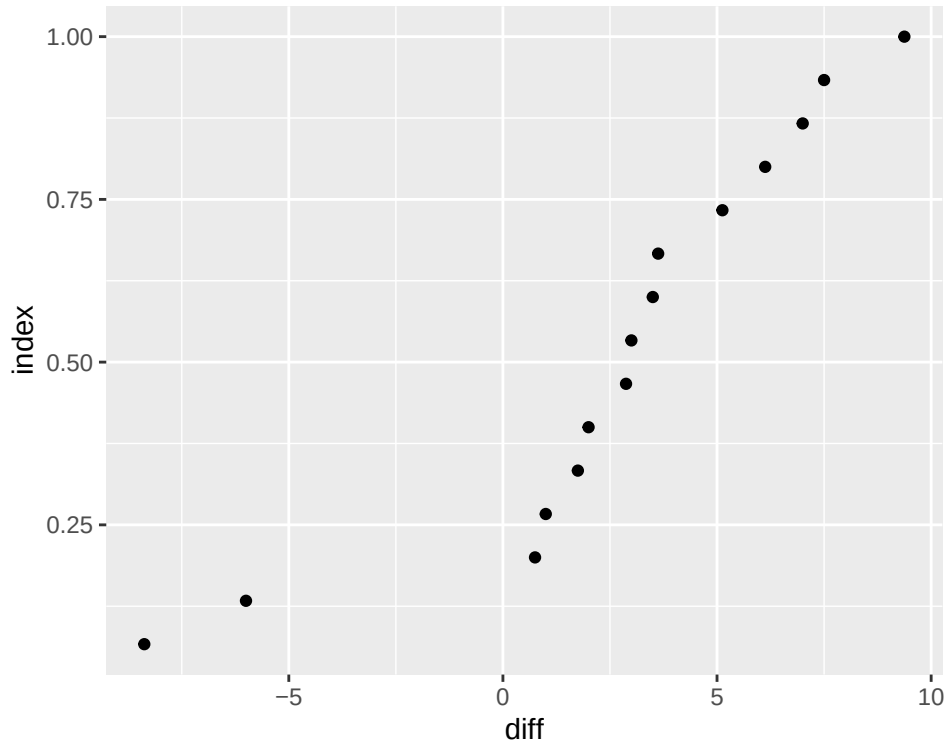



```
ordered_zm <- ZeaMays %>%
  dplyr::select(diff) %>%
  arrange(diff) %>%
  mutate(index = 1:n() / n())

head(ordered_zm)
```

```
##     diff      index
## 1 -8.375 0.06666667
## 2 -6.000 0.13333333
## 3  0.750 0.20000000
## 4  1.000 0.26666667
## 5  1.750 0.33333333
## 6  2.000 0.40000000
```

```
ggplot(ordered_zm, aes(x = diff, y = index)) +
  geom_point()
```



Bootstrap calc of median:

```
bs_zm <- tibble(index = 1:1000) %>%
  mutate(bs_sample = map(index, ~ sample(ZeaMays$diff, replace = TRUE, size = 15))) %>%
  mutate(sample_median = map_dbl(bs_sample, median))
bs_zm
```

```
## # A tibble: 1,000 x 3
##    index bs_sample  sample_median
##    <int> <list>             <dbl>
##  1     1 <dbl [15]>          1
##  2     2 <dbl [15]>          3.62
##  3     3 <dbl [15]>          3
##  4     4 <dbl [15]>          3
##  5     5 <dbl [15]>          3.62
##  6     6 <dbl [15]>          3.62
##  7     7 <dbl [15]>          2.88
##  8     8 <dbl [15]>          3
##  9     9 <dbl [15]>          3.5
## 10    10 <dbl [15]>          2
## # ... with 990 more rows
```

```
bs_zm$bs_sample[[1]]
```

```
##  [1]  1.000  0.750  5.125  2.875  3.625  1.000  1.000  0.750 -6.000  0.750
## [11] -6.000  7.000  9.375  3.500  3.000
```

```
bs_zm$bs_sample[[2]]
```

```
##  [1] 2.000 0.750 6.125 6.125 6.125 9.375 9.375 6.125 3.500 3.000 3.625 0.750
## [13] 2.000 9.375 1.750
```

```
ggplot(bs_zm, aes(x = sample_median)) +
  geom_histogram()
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```