DSC 680 Project 2 R Code

Christine Hathaway

April 17, 2020

```
#Set the working directory
setwd("C:/Users/Christine/Documents/Bellevue/DSC 680/Project 2")
```

Import data from file

```
#Import file
df <- read.csv(file = "C:/Users/Christine/Documents/Bellevue/DSC 680/Project 2/parkinsons.csv", header =
```

Display first five records of file

```
head(df)
```

```
##             name MDVP.Fo.Hz. MDVP.Fhi.Hz. MDVP.Flo.Hz. MDVP.Jitter...
## 1 phon_R01_S01_1     119.992      157.302       74.997        0.00784
## 2 phon_R01_S01_2     122.400      148.650      113.819        0.00968
## 3 phon_R01_S01_3     116.682      131.111      111.555        0.01050
## 4 phon_R01_S01_4     116.676      137.871      111.366        0.00997
## 5 phon_R01_S01_5     116.014      141.781      110.655        0.01284
## 6 phon_R01_S01_6     120.552      131.162      113.787        0.00968
##   MDVP.Jitter.Abs. MDVP.RAP MDVP.PPQ Jitter.DDP MDVP.Shimmer MDVP.Shimmer.dB.
## 1          0.00007  0.00370  0.00554    0.01109      0.04374            0.426
## 2          0.00008  0.00465  0.00696    0.01394      0.06134            0.626
## 3          0.00009  0.00544  0.00781    0.01633      0.05233            0.482
## 4          0.00009  0.00502  0.00698    0.01505      0.05492            0.517
## 5          0.00011  0.00655  0.00908    0.01966      0.06425            0.584
## 6          0.00008  0.00463  0.00750    0.01388      0.04701            0.456
##   Shimmer.APQ3 Shimmer.APQ5 MDVP.APQ Shimmer.DDA     NHR    HNR status     RPDE
## 1      0.02182      0.03130  0.02971     0.06545 0.02211 21.033      1 0.414783
## 2      0.03134      0.04518  0.04368     0.09403 0.01929 19.085      1 0.458359
## 3      0.02757      0.03858  0.03590     0.08270 0.01309 20.651      1 0.429895
## 4      0.02924      0.04005  0.03772     0.08771 0.01353 20.644      1 0.434969
## 5      0.03490      0.04825  0.04465     0.10470 0.01767 19.649      1 0.417356
## 6      0.02328      0.03526  0.03243     0.06985 0.01222 21.378      1 0.415564
##        DFA   spread1  spread2       D2      PPE
## 1 0.815285 -4.813031 0.266482 2.301442 0.284654
## 2 0.819521 -4.075192 0.335590 2.486855 0.368674
## 3 0.825288 -4.443179 0.311173 2.342259 0.332634
## 4 0.819235 -4.117501 0.334147 2.405554 0.368975
## 5 0.823484 -3.747787 0.234513 2.332180 0.410335
## 6 0.825069 -4.242867 0.299111 2.187560 0.357775
```

Run str() function of file

```
str(df)
```

```
## 'data.frame':    195 obs. of  24 variables:
##  $ name            : Factor w/ 195 levels "phon_R01_S01_1",..: 1 2 3 4 5 6 7 8 9 10 ...
##  $ MDVP.Fo.Hz.     : num  120 122 117 117 116 ...
##  $ MDVP.Fhi.Hz.    : num  157 149 131 138 142 ...
##  $ MDVP.Flo.Hz.    : num  75 114 112 111 111 ...
##  $ MDVP.Jitter...  : num  0.00784 0.00968 0.0105 0.00997 0.01284 ...
##  $ MDVP.Jitter.Abs.: num  0.00007 0.00008 0.00009 0.00009 0.00011 0.00008 0.00003 0.00003 0.00006 0.
##  $ MDVP.RAP        : num  0.0037 0.00465 0.00544 0.00502 0.00655 0.00463 0.00155 0.00144 0.00293 0.0
##  $ MDVP.PPQ        : num  0.00554 0.00696 0.00781 0.00698 0.00908 0.0075 0.00202 0.00182 0.00332 0.0
##  $ Jitter.DDP      : num  0.0111 0.0139 0.0163 0.015 0.0197 ...
##  $ MDVP.Shimmer    : num  0.0437 0.0613 0.0523 0.0549 0.0643 ...
##  $ MDVP.Shimmer.dB.: num  0.426 0.626 0.482 0.517 0.584 0.456 0.14 0.134 0.191 0.255 ...
##  $ Shimmer.APQ3    : num  0.0218 0.0313 0.0276 0.0292 0.0349 ...
##  $ Shimmer.APQ5    : num  0.0313 0.0452 0.0386 0.0401 0.0483 ...
##  $ MDVP.APQ        : num  0.0297 0.0437 0.0359 0.0377 0.0447 ...
##  $ Shimmer.DDA     : num  0.0654 0.094 0.0827 0.0877 0.1047 ...
##  $ NHR             : num  0.0221 0.0193 0.0131 0.0135 0.0177 ...
##  $ HNR             : num  21 19.1 20.7 20.6 19.6 ...
##  $ status          : int  1 1 1 1 1 1 1 1 1 1 ...
##  $ RPDE            : num  0.415 0.458 0.43 0.435 0.417 ...
##  $ DFA             : num  0.815 0.82 0.825 0.819 0.823 ...
##  $ spread1         : num  -4.81 -4.08 -4.44 -4.12 -3.75 ...
##  $ spread2         : num  0.266 0.336 0.311 0.334 0.235 ...
##  $ D2              : num  2.3 2.49 2.34 2.41 2.33 ...
##  $ PPE             : num  0.285 0.369 0.333 0.369 0.41 ...
```

Count each variable

```
sapply(df, function(x) length(unique(x)))
```

```
##             name      MDVP.Fo.Hz.     MDVP.Fhi.Hz.     MDVP.Flo.Hz.
##              195              195              195              195
##   MDVP.Jitter... MDVP.Jitter.Abs.         MDVP.RAP         MDVP.PPQ
##              173               19              155              165
##       Jitter.DDP     MDVP.Shimmer MDVP.Shimmer.dB.     Shimmer.APQ3
##              180              188              149              184
##     Shimmer.APQ5         MDVP.APQ      Shimmer.DDA              NHR
##              189              189              189              185
##              HNR           status             RPDE              DFA
##              195                2              195              195
##          spread1          spread2               D2              PPE
##              195              194              195              195
```

Run desribe() function of file

```
describe(df)
```

```
## df
##
##  24  Variables      195  Observations
## ---------------------------------------------------------------------------
## name
```

```
##        n  missing distinct
##      195        0      195
##
## lowest : phon_R01_S01_1 phon_R01_S01_2 phon_R01_S01_3 phon_R01_S01_4 phon_R01_S01_5
## highest: phon_R01_S50_2 phon_R01_S50_3 phon_R01_S50_4 phon_R01_S50_5 phon_R01_S50_6
## ---------------------------------------------------------------------------
## MDVP.Fo.Hz.
##        n  missing distinct     Info     Mean      Gmd      .05      .10
##      195        0      195        1    154.2    46.69    101.9    110.7
##      .25      .50      .75      .90      .95
##    117.6    148.8    182.8    209.9    236.5
##
## lowest :  88.333  91.904  95.056  95.385  95.605
## highest: 243.439 244.990 245.510 252.455 260.105
## ---------------------------------------------------------------------------
## MDVP.Fhi.Hz.
##        n  missing distinct     Info     Mean      Gmd      .05      .10
##      195        0      195        1    197.1    84.42    115.8    125.3
##      .25      .50      .75      .90      .95
##    134.9    175.8    224.2    261.0    410.6
##
## lowest : 102.145 102.305 107.715 108.664 110.019
## highest: 565.740 581.289 586.567 588.518 592.030
## ---------------------------------------------------------------------------
## MDVP.Flo.Hz.
##        n  missing distinct     Info     Mean      Gmd      .05      .10
##      195        0      195        1    116.3    46.12    68.95    75.61
##      .25      .50      .75      .90      .95
##    84.29   104.31   140.02   187.88   220.19
##
## lowest :  65.476  65.750  65.782  65.809  66.004
## highest: 231.848 232.435 232.483 237.303 239.170
## ---------------------------------------------------------------------------
## MDVP.Jitter...
##        n  missing distinct     Info     Mean      Gmd      .05      .10
##      195        0      173        1  0.00622 0.004259 0.002211 0.002648
##      .25      .50      .75      .90      .95
## 0.003460 0.004940 0.007365 0.009882 0.015561
##
## lowest : 0.00168 0.00174 0.00178 0.00180 0.00183
## highest: 0.01936 0.02714 0.03011 0.03107 0.03316
## ---------------------------------------------------------------------------
## MDVP.Jitter.Abs.
##         n   missing  distinct      Info      Mean       Gmd       .05       .10
##       195         0        19     0.978 4.396e-05 3.302e-05     1e-05     1e-05
##       .25       .50       .75       .90       .95
##     2e-05     3e-05     6e-05     8e-05     1e-04
##
## lowest : 7.0e-06 9.0e-06 1.0e-05 2.0e-05 3.0e-05
## highest: 1.4e-04 1.5e-04 1.6e-04 2.2e-04 2.6e-04
##
## 7e-06 (1, 0.005), 9e-06 (5, 0.026), 1e-05 (20, 0.103), 2e-05 (28, 0.144), 3e-05
## (46, 0.236), 4e-05 (28, 0.144), 5e-05 (17, 0.087), 6e-05 (16, 0.082), 7e-05 (8,
## 0.041), 8e-05 (9, 0.046), 9e-05 (5, 0.026), 1e-04 (3, 0.015), 0.00011 (2,
```

```
## 0.010), 0.00012 (1, 0.005), 0.00014 (1, 0.005), 0.00015 (2, 0.010), 0.00016 (1,
## 0.005), 0.00022 (1, 0.005), 0.00026 (1, 0.005)
## ---------------------------------------------------------------------------
## MDVP.RAP
##        n  missing distinct     Info     Mean      Gmd      .05      .10
##      195        0      155        1 0.003306 0.002518 0.001118 0.001252
##      .25      .50      .75      .90      .95
## 0.001660 0.002500 0.003835 0.005400 0.008756
##
## lowest : 0.00068 0.00075 0.00076 0.00092 0.00093
## highest: 0.01159 0.01568 0.01800 0.01854 0.02144
## ---------------------------------------------------------------------------
## MDVP.PPQ
##        n  missing distinct     Info     Mean      Gmd      .05      .10
##      195        0      165        1 0.003446  0.00242 0.001315 0.001452
##      .25      .50      .75      .90      .95
## 0.001860 0.002690 0.003955 0.005712 0.009083
##
## lowest : 0.00092 0.00096 0.00100 0.00106 0.00107
## highest: 0.01154 0.01522 0.01628 0.01699 0.01958
## ---------------------------------------------------------------------------
## Jitter.DDP
##        n  missing distinct     Info     Mean      Gmd      .05      .10
##      195        0      180        1  0.00992 0.007553 0.003354 0.003758
##      .25      .50      .75      .90      .95
## 0.004985 0.007490 0.011505 0.016202 0.026271
##
## lowest : 0.00204 0.00225 0.00229 0.00276 0.00278
## highest: 0.03476 0.04705 0.05401 0.05563 0.06433
## ---------------------------------------------------------------------------
## MDVP.Shimmer
##        n  missing distinct     Info     Mean      Gmd      .05      .10
##      195        0      188        1  0.02971  0.01931  0.01121  0.01287
##      .25      .50      .75      .90      .95
##  0.01650  0.02297  0.03789  0.05593  0.06726
##
## lowest : 0.00954 0.00958 0.01015 0.01022 0.01024
## highest: 0.08143 0.08684 0.09178 0.09419 0.11908
## ---------------------------------------------------------------------------
## MDVP.Shimmer.dB.
##        n  missing distinct     Info     Mean      Gmd      .05      .10
##      195        0      149        1   0.2823   0.1931   0.1018   0.1198
##      .25      .50      .75      .90      .95
##   0.1485   0.2210   0.3500   0.5320   0.6527
##
## lowest : 0.085 0.089 0.090 0.093 0.094, highest: 0.833 0.891 0.930 1.018 1.302
## ---------------------------------------------------------------------------
## Shimmer.APQ3
##        n  missing distinct     Info     Mean      Gmd      .05      .10
##      195        0      184        1  0.01566  0.01049 0.005368 0.006358
##      .25      .50      .75      .90      .95
## 0.008245 0.012790 0.020265 0.030116 0.036227
##
## lowest : 0.00455 0.00468 0.00469 0.00476 0.00490
```

```
## highest: 0.04284 0.04421 0.05358 0.05551 0.05647
## ---------------------------------------------------------------------------
## Shimmer.APQ5
##        n  missing distinct     Info     Mean      Gmd      .05      .10
##      195        0      189        1  0.01788  0.01208 0.006383 0.007522
##      .25      .50      .75      .90      .95
## 0.009580 0.013470 0.022380 0.036972 0.042701
##
## lowest : 0.00570 0.00576 0.00582 0.00588 0.00606
## highest: 0.04962 0.05005 0.05426 0.05556 0.07940
## ---------------------------------------------------------------------------
## MDVP.APQ
##        n  missing distinct     Info     Mean      Gmd      .05      .10
##      195        0      189        1  0.02408  0.01617 0.009114 0.010654
##      .25      .50      .75      .90      .95
## 0.013080 0.018260 0.029400 0.044298 0.057718
##
## lowest : 0.00719 0.00726 0.00762 0.00802 0.00811
## highest: 0.06460 0.06824 0.08318 0.08808 0.13778
## ---------------------------------------------------------------------------
## Shimmer.DDA
##        n  missing distinct     Info     Mean      Gmd      .05      .10
##      195        0      189        1  0.04699  0.03148  0.01611  0.01906
##      .25      .50      .75      .90      .95
##  0.02474  0.03836  0.06080  0.09035  0.10868
##
## lowest : 0.01364 0.01403 0.01406 0.01407 0.01428
## highest: 0.12851 0.13262 0.16074 0.16654 0.16942
## ---------------------------------------------------------------------------
## NHR
##        n  missing distinct     Info     Mean      Gmd      .05      .10
##      195        0      185        1  0.02485  0.02982 0.002528 0.004066
##      .25      .50      .75      .90      .95
## 0.005925 0.011660 0.025640 0.052348 0.092044
##
## lowest : 0.00065 0.00072 0.00119 0.00135 0.00167
## highest: 0.16265 0.16744 0.21713 0.25930 0.31482
## ---------------------------------------------------------------------------
## HNR
##        n  missing distinct     Info     Mean      Gmd      .05      .10
##      195        0      195        1    21.89    4.894    13.48    16.02
##      .25      .50      .75      .90      .95
##    19.20    22.09    25.08    26.51    26.97
##
## lowest :  8.441  8.867  9.449 10.489 11.744, highest: 30.775 30.940 31.732 32.684 33.047
## ---------------------------------------------------------------------------
## status
##        n  missing distinct     Info      Sum     Mean      Gmd
##      195        0        2    0.557      147   0.7538    0.373
##
## ---------------------------------------------------------------------------
## RPDE
##        n  missing distinct     Info     Mean      Gmd      .05      .10
##      195        0      195        1   0.4985   0.1197   0.3309   0.3606
```

```
##      .25      .50      .75      .90      .95
##   0.4213   0.4960   0.5876   0.6375   0.6532
##
## lowest : 0.256570 0.263654 0.276850 0.296888 0.305062
## highest: 0.665318 0.671299 0.671378 0.677131 0.685151
## ---------------------------------------------------------------------------
## DFA
##        n  missing distinct     Info     Mean      Gmd      .05      .10
##      195        0      195        1   0.7181  0.06348   0.6323   0.6464
##      .25      .50      .75      .90      .95
##   0.6748   0.7223   0.7619   0.7898   0.8160
##
## lowest : 0.574282 0.582710 0.605417 0.623731 0.626710
## highest: 0.819521 0.821364 0.823484 0.825069 0.825288
## ---------------------------------------------------------------------------
## spread1
##        n  missing distinct     Info     Mean      Gmd      .05      .10
##      195        0      195        1   -5.684    1.231   -7.306   -7.052
##      .25      .50      .75      .90      .95
##   -6.450   -5.721   -5.046   -4.256   -3.734
##
## lowest : -7.964984 -7.777685 -7.695734 -7.682587 -7.517934
## highest: -3.269487 -2.931070 -2.929379 -2.839756 -2.434031
## ---------------------------------------------------------------------------
## spread2
##        n  missing distinct     Info     Mean      Gmd      .05      .10
##      195        0      194        1   0.2265  0.09419  0.08884  0.12135
##      .25      .50      .75      .90      .95
##  0.17435  0.21888  0.27923  0.33841  0.37314
##
## lowest : 0.006274 0.018689 0.056844 0.063412 0.066994
## highest: 0.396746 0.397749 0.414758 0.434326 0.450493
## ---------------------------------------------------------------------------
## D2
##        n  missing distinct     Info     Mean      Gmd      .05      .10
##      195        0      195        1    2.382   0.4299    1.849    1.925
##      .25      .50      .75      .90      .95
##    2.099    2.362    2.636    2.922    3.085
##
## lowest : 1.423287 1.512275 1.544609 1.743867 1.765957
## highest: 3.184027 3.274865 3.317586 3.413649 3.671155
## ---------------------------------------------------------------------------
## PPE
##        n  missing distinct     Info     Mean      Gmd      .05      .10
##      195        0      195        1   0.2066   0.1001  0.09159  0.10199
##      .25      .50      .75      .90      .95
##  0.13745  0.19405  0.25298  0.33408  0.36957
##
## lowest : 0.044539 0.056141 0.057610 0.068501 0.073581
## highest: 0.430788 0.444774 0.454721 0.457533 0.527367
## ---------------------------------------------------------------------------
```

Run stat.desc() function of file

```
stat.desc(df)
```

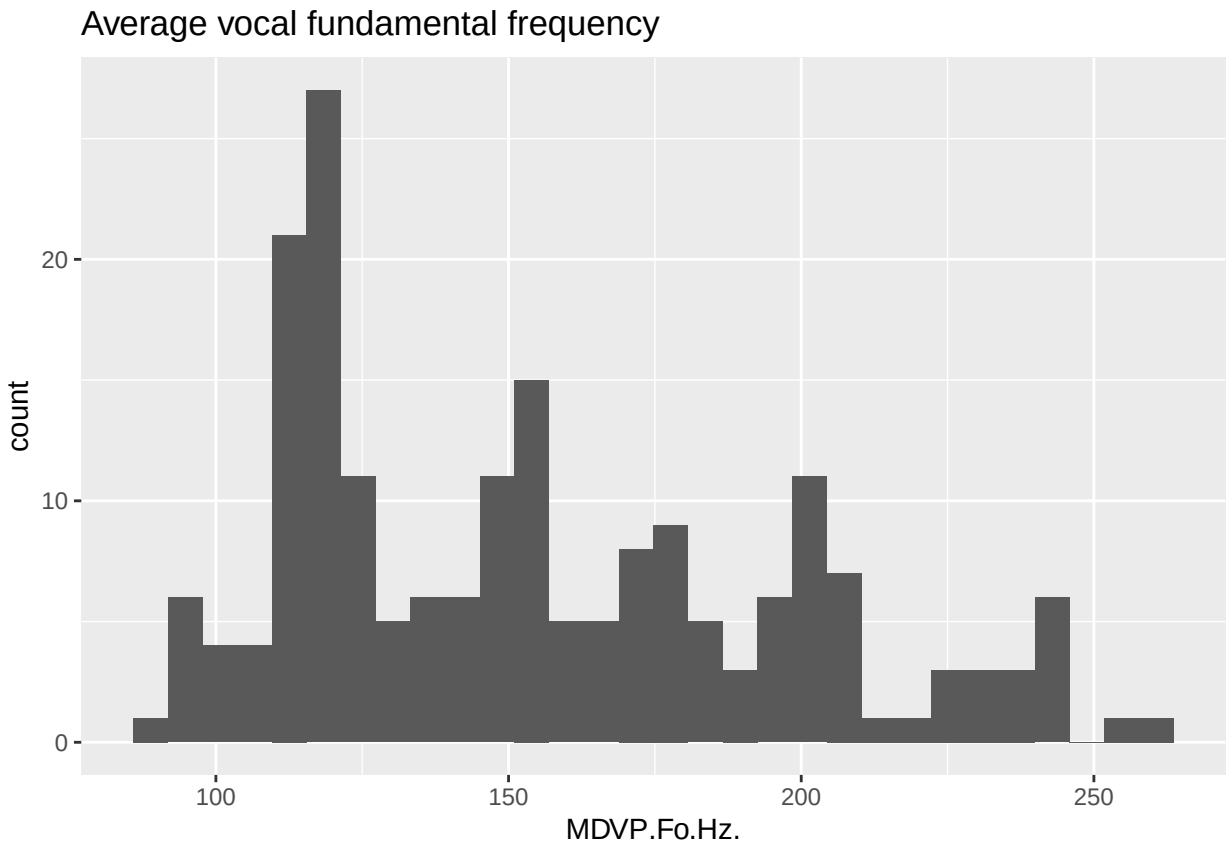
```
##          name  MDVP.Fo.Hz. MDVP.Fhi.Hz. MDVP.Flo.Hz. MDVP.Jitter...
## nbr.val    NA 1.950000e+02 1.950000e+02 1.950000e+02   1.950000e+02
## nbr.null   NA 0.000000e+00 0.000000e+00 0.000000e+00   0.000000e+00
## nbr.na     NA 0.000000e+00 0.000000e+00 0.000000e+00   0.000000e+00
## min        NA 8.833300e+01 1.021450e+02 6.547600e+01   1.680000e-03
## max        NA 2.601050e+02 5.920300e+02 2.391700e+02   3.316000e-02
## range      NA 1.717720e+02 4.898850e+02 1.736940e+02   3.148000e-02
## sum        NA 3.007458e+04 3.843546e+04 2.268330e+04   1.212990e+00
## median     NA 1.487900e+02 1.758290e+02 1.043150e+02   4.940000e-03
## mean       NA 1.542286e+02 1.971049e+02 1.163246e+02   6.220462e-03
## SE.mean    NA 2.964004e+00 6.551846e+00 3.116633e+00   3.471821e-04
## CI.mean    NA 5.845809e+00 1.292199e+01 6.146834e+00   6.847359e-04
## var        NA 1.713137e+03 8.370703e+03 1.894113e+03   2.350440e-05
## std.dev    NA 4.139006e+01 9.149155e+01 4.352141e+01   4.848134e-03
## coef.var   NA 2.683682e-01 4.641769e-01 3.741376e-01   7.793849e-01
##          MDVP.Jitter.Abs.     MDVP.RAP     MDVP.PPQ   Jitter.DDP MDVP.Shimmer
## nbr.val      1.950000e+02 1.950000e+02 1.950000e+02 1.950000e+02 1.950000e+02
## nbr.null     0.000000e+00 0.000000e+00 0.000000e+00 0.000000e+00 0.000000e+00
## nbr.na       0.000000e+00 0.000000e+00 0.000000e+00 0.000000e+00 0.000000e+00
## min          7.000000e-06 6.800000e-04 9.200000e-04 2.040000e-03 9.540000e-03
## max          2.600000e-04 2.144000e-02 1.958000e-02 6.433000e-02 1.190800e-01
## range        2.530000e-04 2.076000e-02 1.866000e-02 6.229000e-02 1.095400e-01
## sum          8.572000e-03 6.447500e-01 6.720400e-01 1.934390e+00 5.793280e+00
## median       3.000000e-05 2.500000e-03 2.690000e-03 7.490000e-03 2.297000e-02
## mean         4.395897e-05 3.306410e-03 3.446359e-03 9.919949e-03 2.970913e-02
## SE.mean      2.493649e-06 2.125267e-04 1.975744e-04 6.375817e-04 1.350373e-03
## CI.mean      4.918142e-06 4.191596e-04 3.896696e-04 1.257482e-03 2.663297e-03
## var          1.212565e-09 8.807685e-06 7.611952e-06 7.926954e-05 3.555839e-04
## std.dev      3.482191e-05 2.967774e-03 2.758977e-03 8.903344e-03 1.885693e-02
## coef.var     7.921456e-01 8.975820e-01 8.005483e-01 8.975192e-01 6.347185e-01
##          MDVP.Shimmer.dB. Shimmer.APQ3 Shimmer.APQ5     MDVP.APQ  Shimmer.DDA
## nbr.val      195.00000000 1.950000e+02 1.950000e+02 1.950000e+02 1.950000e+02
## nbr.null       0.00000000 0.000000e+00 0.000000e+00 0.000000e+00 0.000000e+00
## nbr.na         0.00000000 0.000000e+00 0.000000e+00 0.000000e+00 0.000000e+00
## min            0.08500000 4.550000e-03 5.700000e-03 7.190000e-03 1.364000e-02
## max            1.30200000 5.647000e-02 7.940000e-02 1.377800e-01 1.694200e-01
## range          1.21700000 5.192000e-02 7.370000e-02 1.305900e-01 1.557800e-01
## sum           55.03900000 3.054510e+00 3.486260e+00 4.695890e+00 9.163560e+00
## median         0.22100000 1.279000e-02 1.347000e-02 1.826000e-02 3.836000e-02
## mean           0.28225128 1.566415e-02 1.787826e-02 2.408149e-02 4.699262e-02
## SE.mean        0.01395545 7.270830e-04 8.610354e-04 1.213581e-03 2.181223e-03
## CI.mean        0.02752389 1.434002e-03 1.698192e-03 2.393506e-03 4.301955e-03
## var            0.03797716 1.030867e-04 1.445695e-04 2.871919e-04 9.277580e-04
## std.dev        0.19487729 1.015316e-02 1.202371e-02 1.694674e-02 3.045912e-02
## coef.var       0.69043899 6.481781e-01 6.725323e-01 7.037247e-01 6.481682e-01
##                     NHR          HNR       status         RPDE          DFA
## nbr.val  1.950000e+02  195.0000000 195.00000000 1.950000e+02 1.950000e+02
## nbr.null 0.000000e+00    0.0000000  48.00000000 0.000000e+00 0.000000e+00
## nbr.na   0.000000e+00    0.0000000   0.00000000 0.000000e+00 0.000000e+00
## min      6.500000e-04    8.4410000   0.00000000 2.565700e-01 5.742820e-01
## max      3.148200e-01   33.0470000   1.00000000 6.851510e-01 8.252880e-01
```

```
## range    3.141700e-01   24.6060000   1.00000000 4.285810e-01 2.510060e-01
## sum      4.845180e+00 4267.7650000 147.00000000 9.721443e+01 1.400293e+02
## median   1.166000e-02   22.0850000   1.00000000 4.959540e-01 7.222540e-01
## mean     2.484708e-02   21.8859744   0.75384615 4.985355e-01 7.180990e-01
## SE.mean  2.894425e-03    0.3169356   0.03092743 7.443421e-03 3.962681e-03
## CI.mean  5.708581e-03    0.6250817   0.06099716 1.468042e-02 7.815467e-03
## var      1.633651e-03   19.5873894   0.18651864 1.080388e-02 3.062054e-03
## std.dev  4.041845e-02    4.4257643   0.43187803 1.039417e-01 5.533583e-02
## coef.var 1.626688e+00    0.2022192   0.57289943 2.084941e-01 7.705877e-02
##                 spread1      spread2           D2          PPE
## nbr.val    195.0000000 1.950000e+02 195.00000000 1.950000e+02
## nbr.null     0.0000000 0.000000e+00   0.00000000 0.000000e+00
## nbr.na       0.0000000 0.000000e+00   0.00000000 0.000000e+00
## min         -7.9649840 6.274000e-03   1.42328700 4.453900e-02
## max         -2.4340310 4.504930e-01   3.67115500 5.273670e-01
## range        5.5309530 4.442190e-01   2.24786800 4.828280e-01
## sum      -1108.4573650 4.416952e+01 464.45608700 4.027757e+01
## median      -5.7208680 2.188850e-01   2.36153200 1.940520e-01
## mean        -5.6843967 2.265103e-01   2.38182609 2.065516e-01
## SE.mean      0.0780714 5.972811e-03   0.02741281 6.453579e-03
## CI.mean      0.1539777 1.177998e-02   0.05406539 1.272818e-02
## var          1.1885530 6.956521e-03   0.14653511 8.121492e-03
## std.dev      1.0902078 8.340576e-02   0.38279905 9.011932e-02
## coef.var    -0.1917895 3.682205e-01   0.16071662 4.363041e-01
```

Create graphs of fundamental frequencies

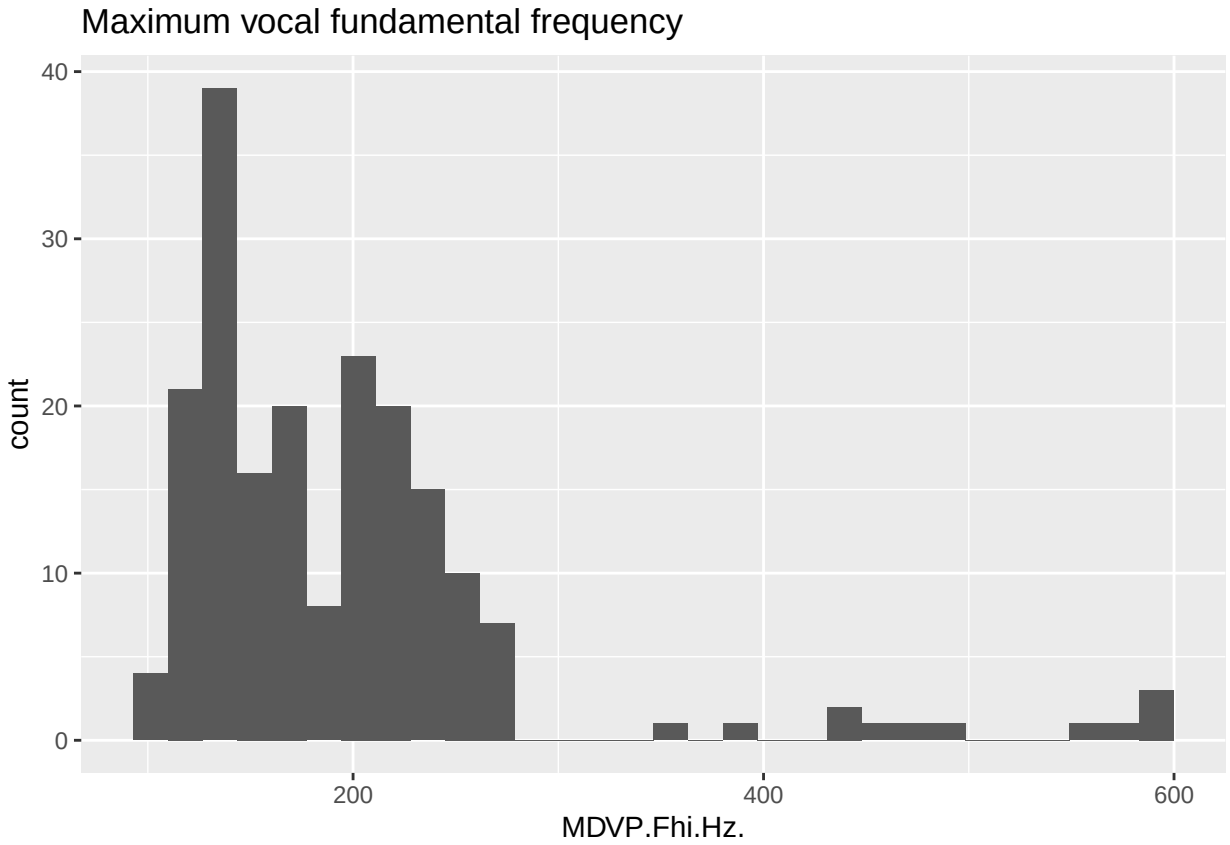
```
ggplot(df, aes(x=MDVP.Fo.Hz.)) + geom_histogram() + ggtitle("Average vocal fundamental frequency")
```

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



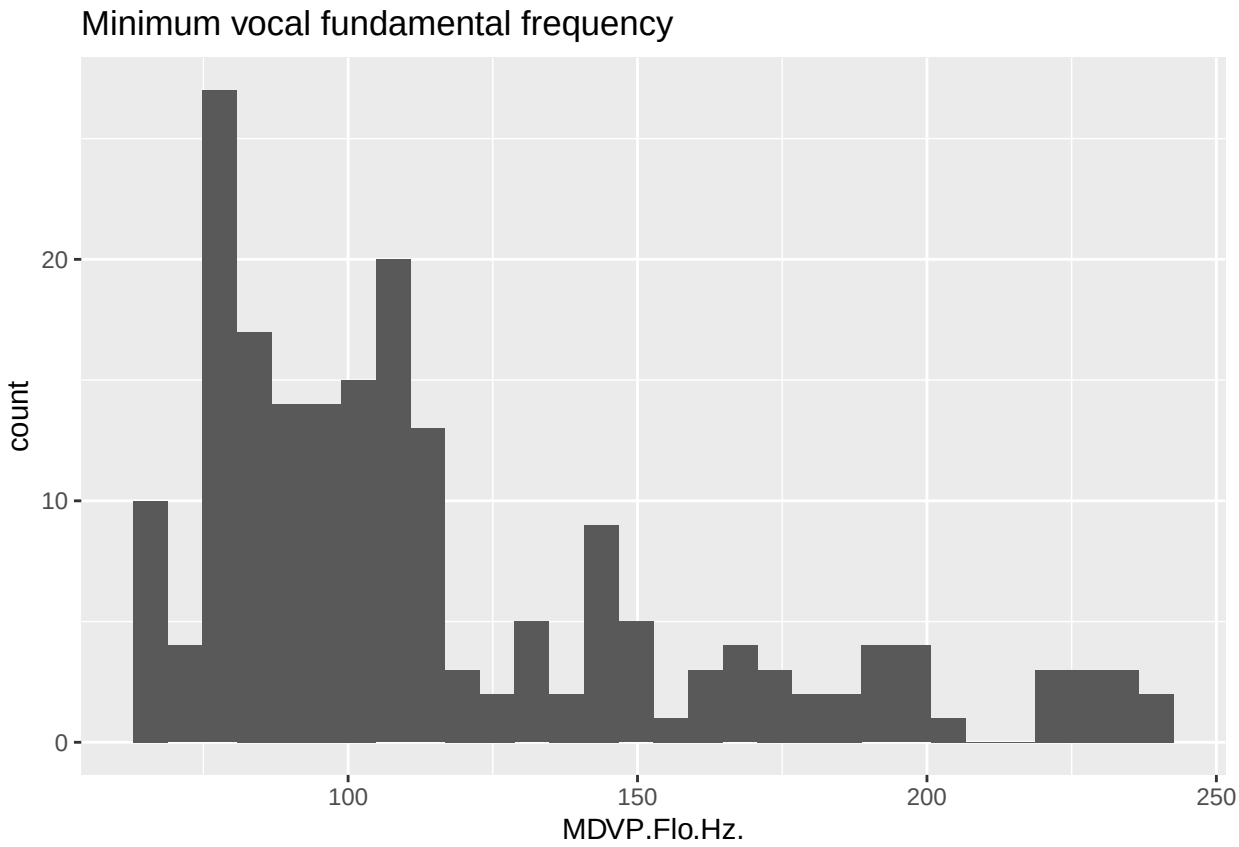

```
ggplot(df, aes(x=MDVP.Fhi.Hz.)) + geom_histogram() + ggtitle("Maximum vocal fundamental frequency")
```

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




```
ggplot(df, aes(x=MDVP.Flo.Hz.)) + geom_histogram() + ggtitle("Minimum vocal fundamental frequency")
```

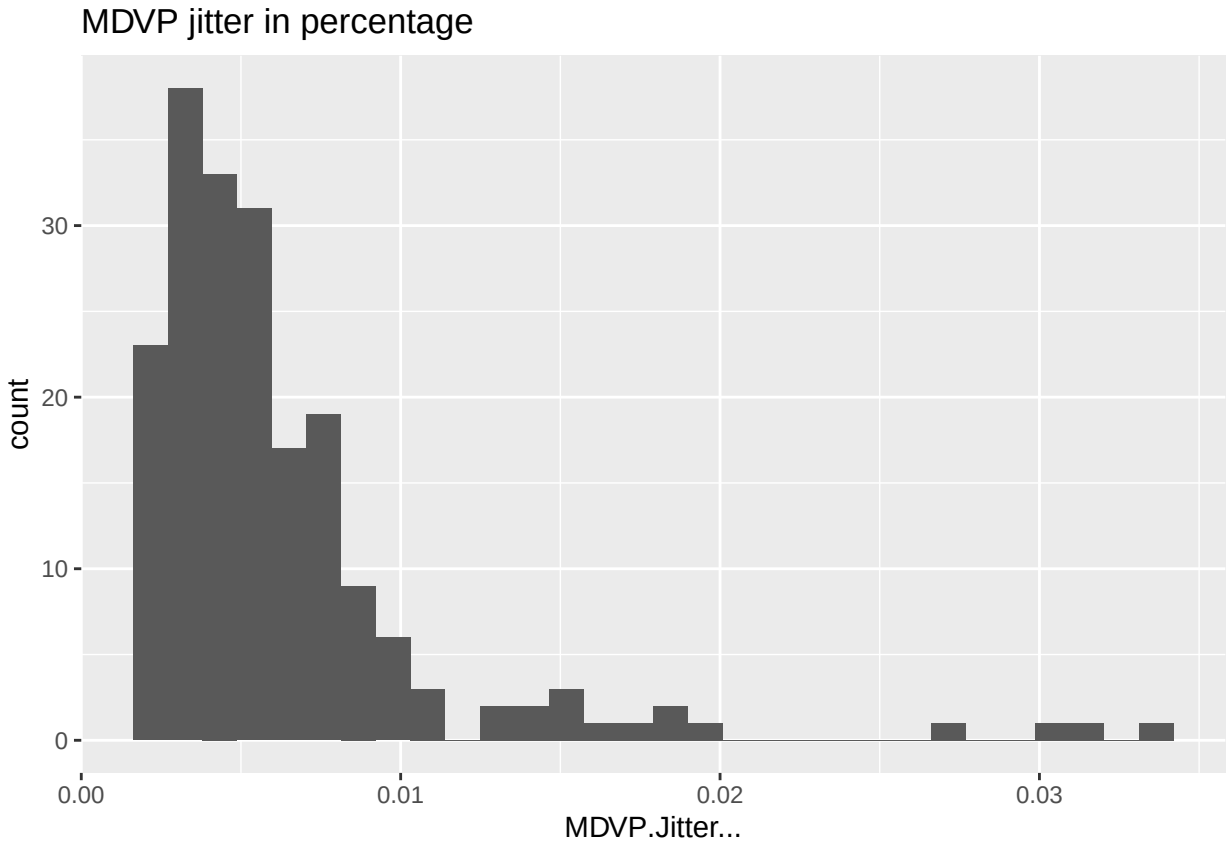
```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```



Create graphs of measures of variation in fundamental frequencies

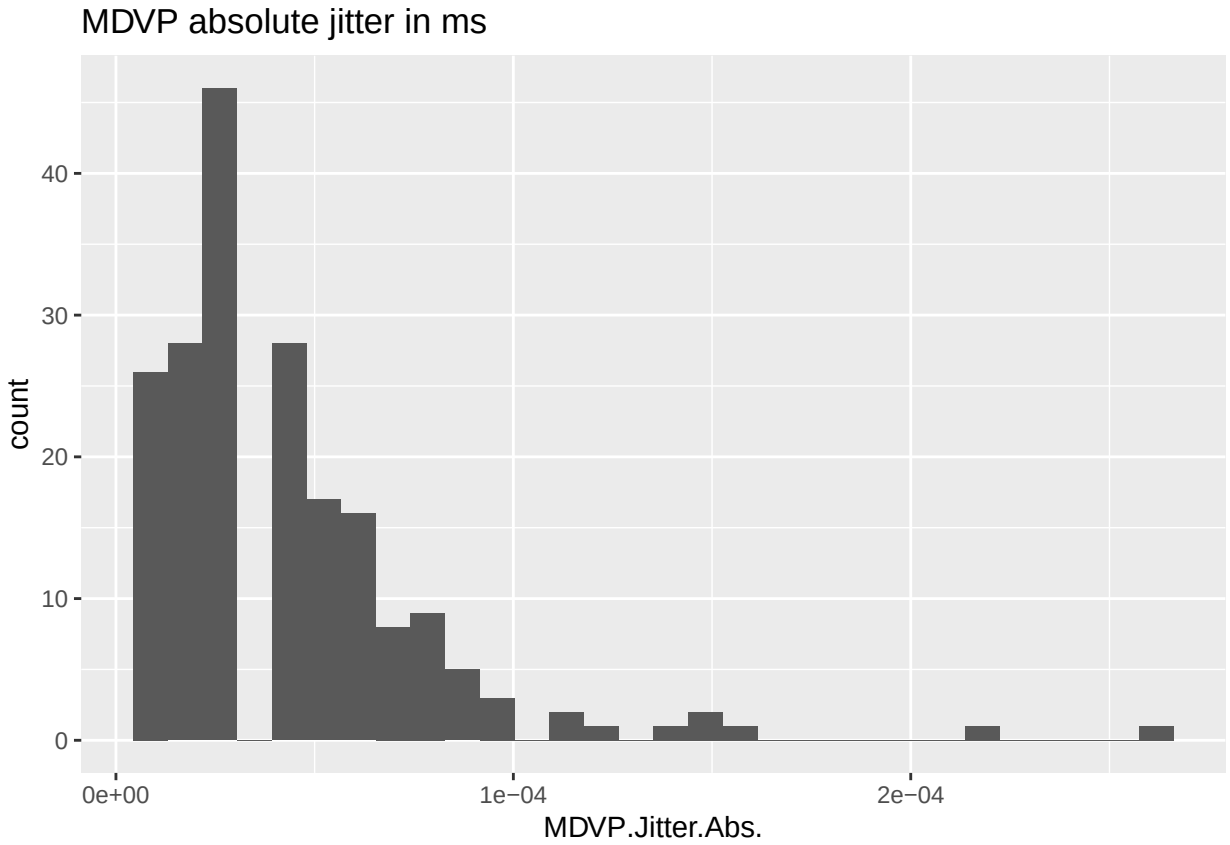
```
ggplot(df, aes(x=MDVP.Jitter...)) + geom_histogram() + ggtitle("MDVP jitter in percentage")
```

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



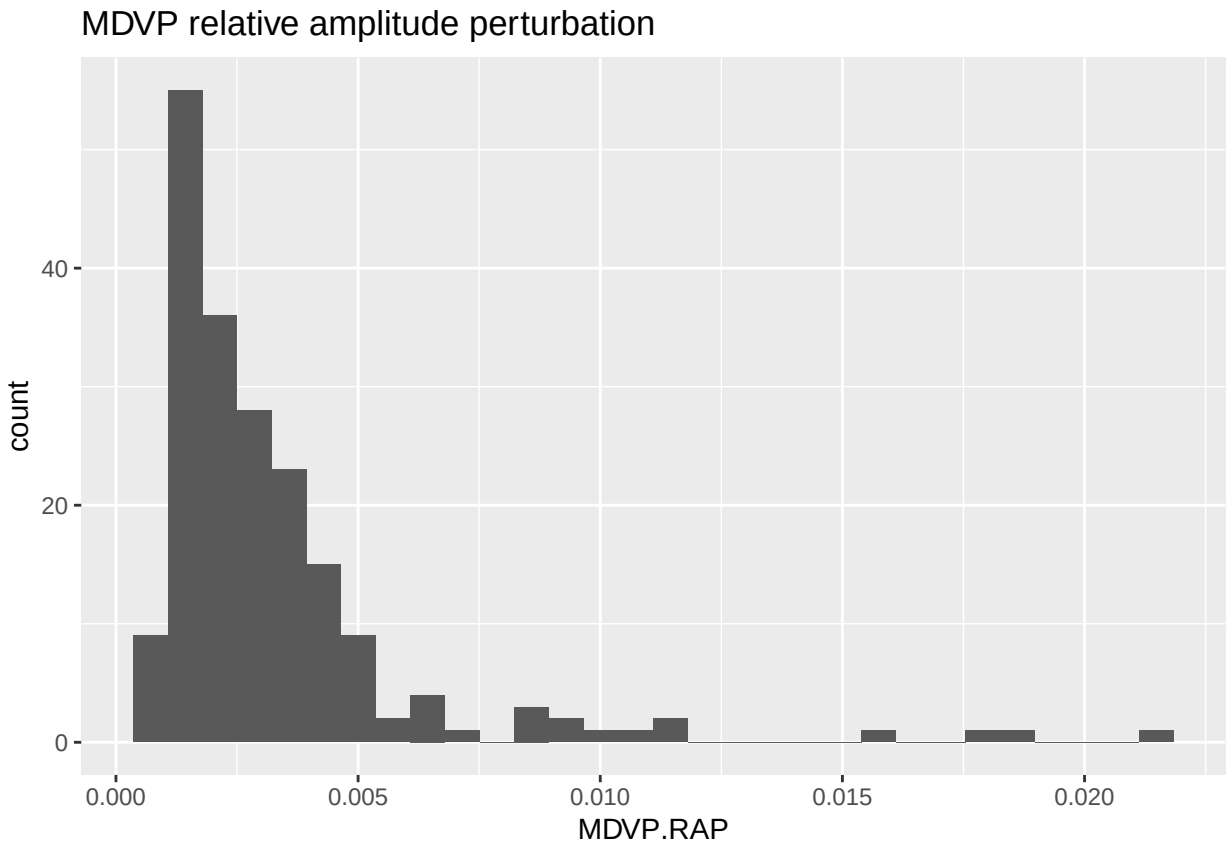

```
ggplot(df, aes(x=MDVP.Jitter.Abs.)) + geom_histogram() + ggtitle("MDVP absolute jitter in ms")
```

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



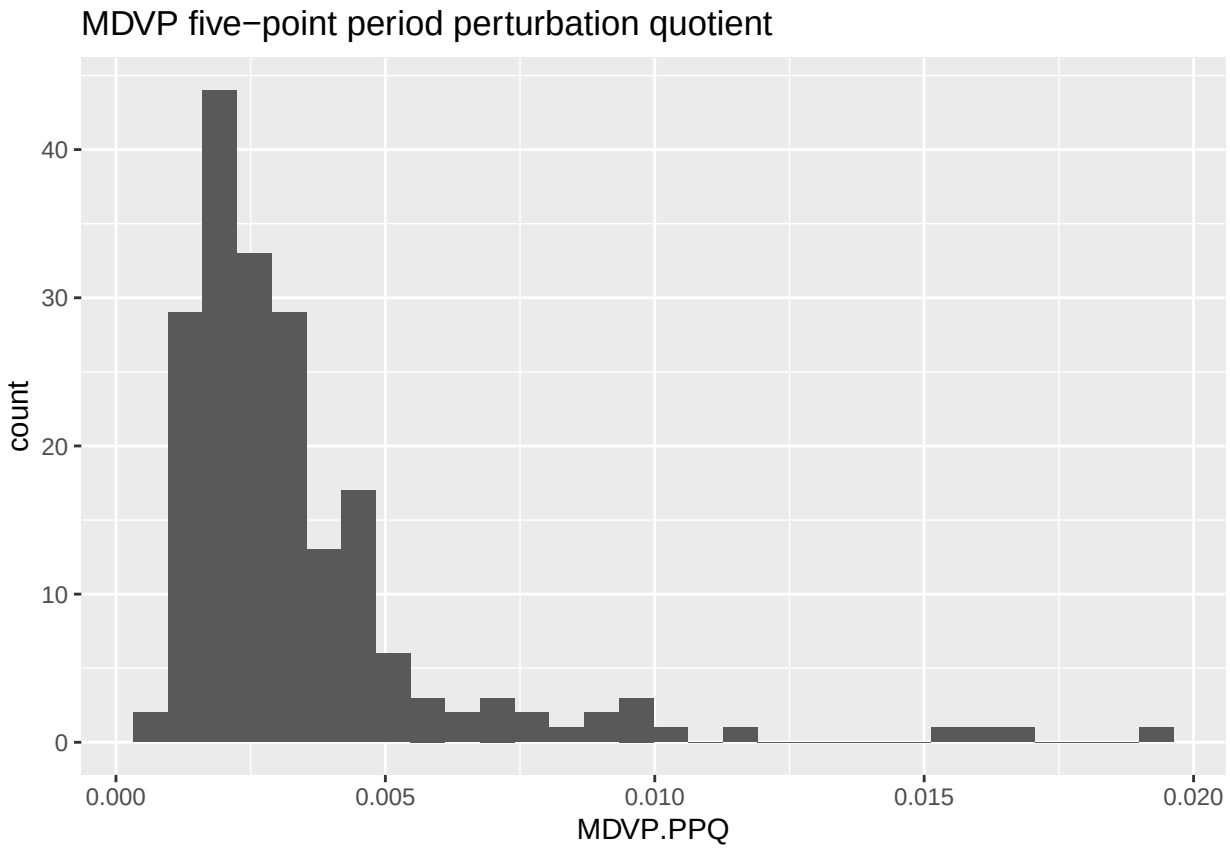

```
ggplot(df, aes(x=MDVP.RAP)) + geom_histogram() + ggtitle("MDVP relative amplitude perturbation")
```

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



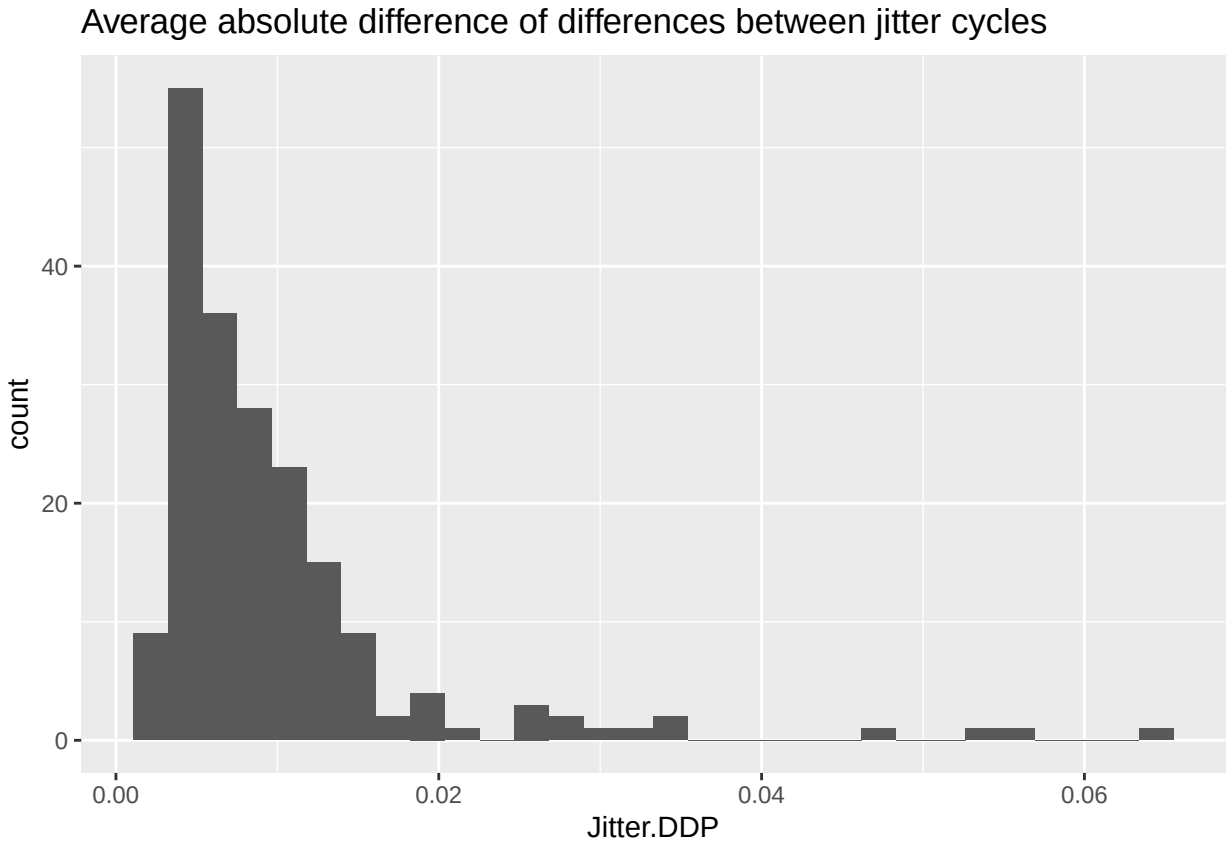

```
ggplot(df, aes(x=MDVP.PPQ )) + geom_histogram() + ggtitle("MDVP five-point period perturbation quotient
```

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




```
ggplot(df, aes(x=Jitter.DDP )) + geom_histogram() + ggtitle("Average absolute difference of differences
```

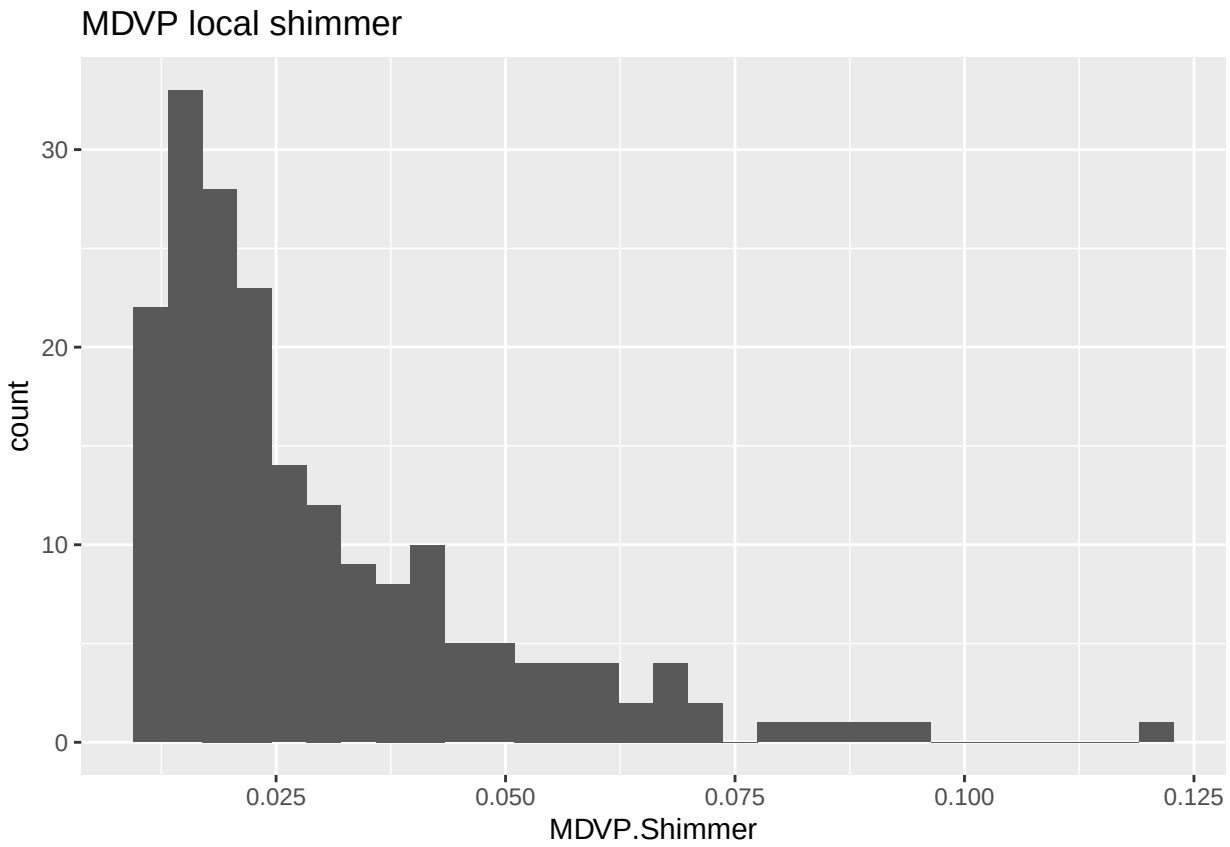
```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```



Create graphs of measures of variation in amplitude

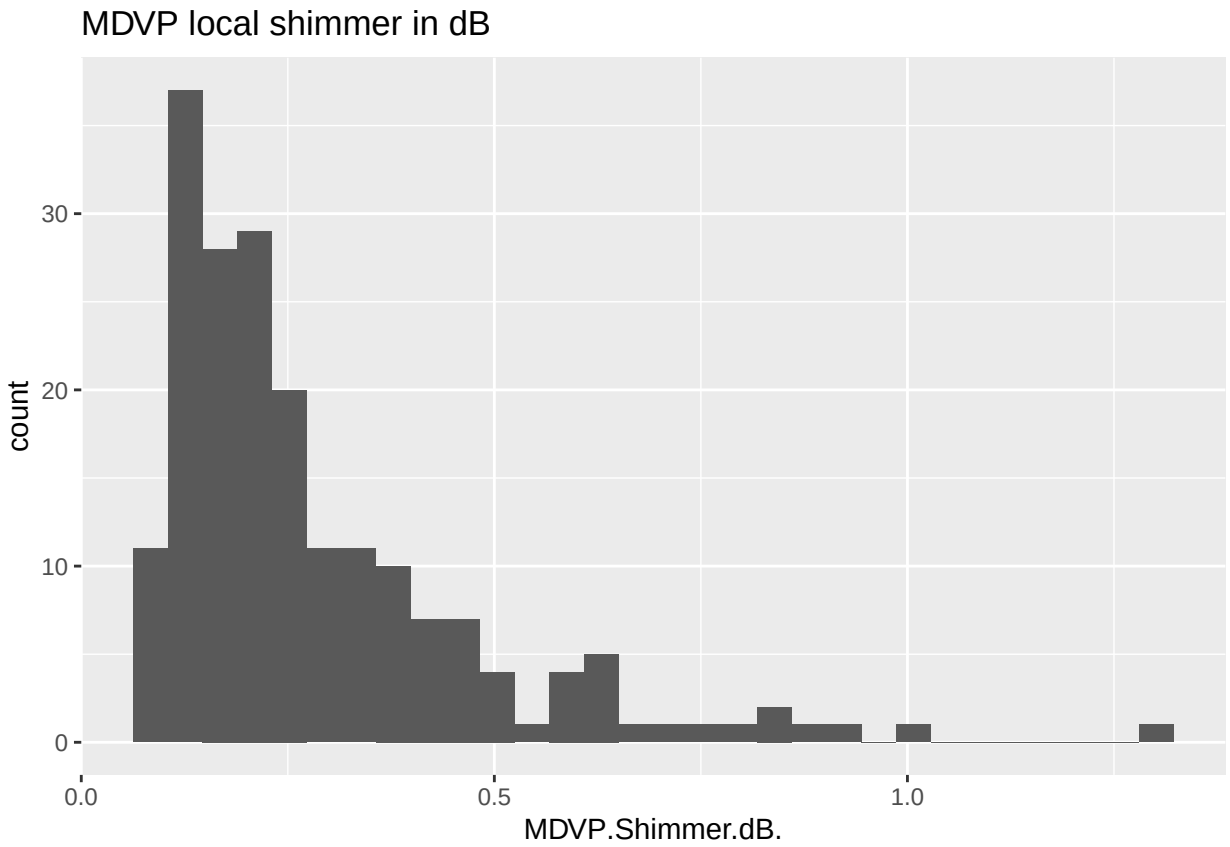
```
ggplot(df, aes(x=MDVP.Shimmer)) + geom_histogram() + ggtitle("MDVP local shimmer")
```

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



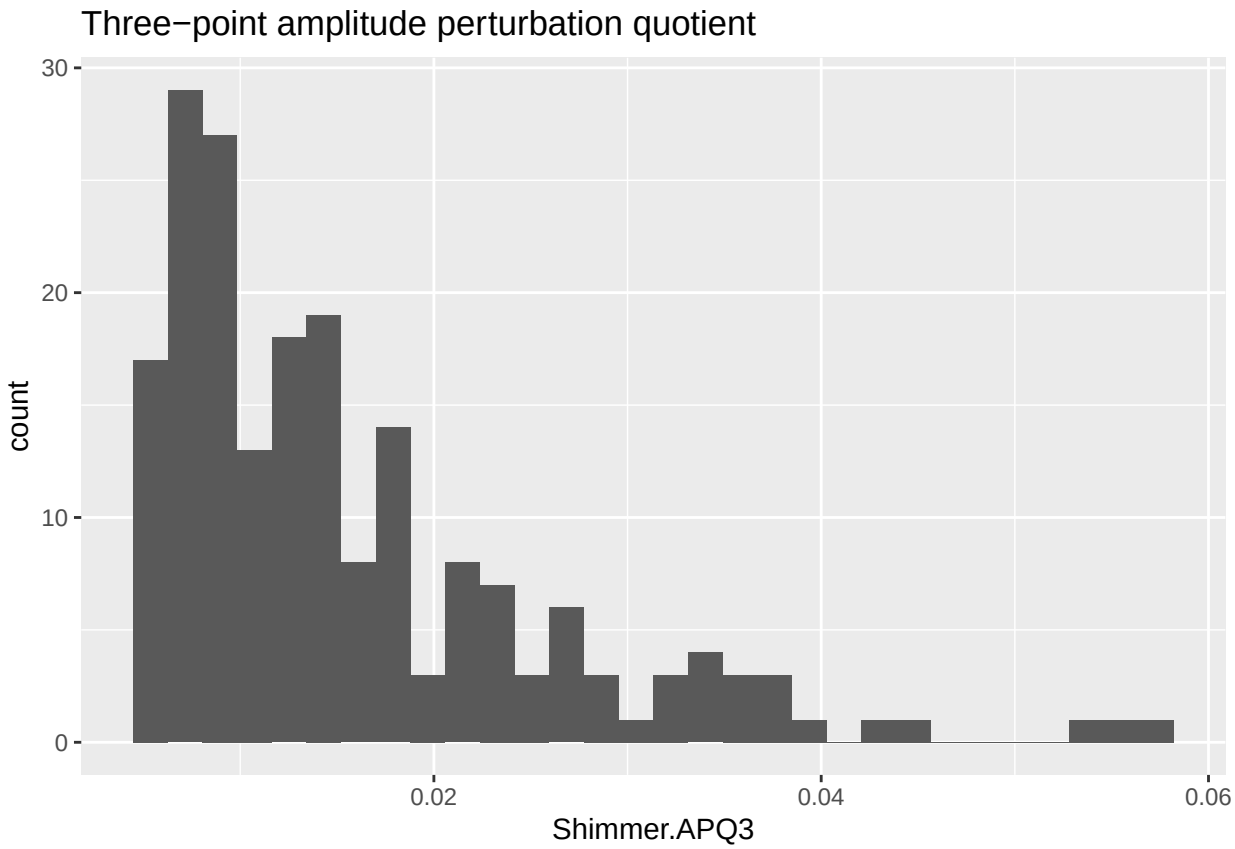

```
ggplot(df, aes(x=MDVP.Shimmer.dB.)) + geom_histogram() + ggtitle("MDVP local shimmer in dB")
```

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



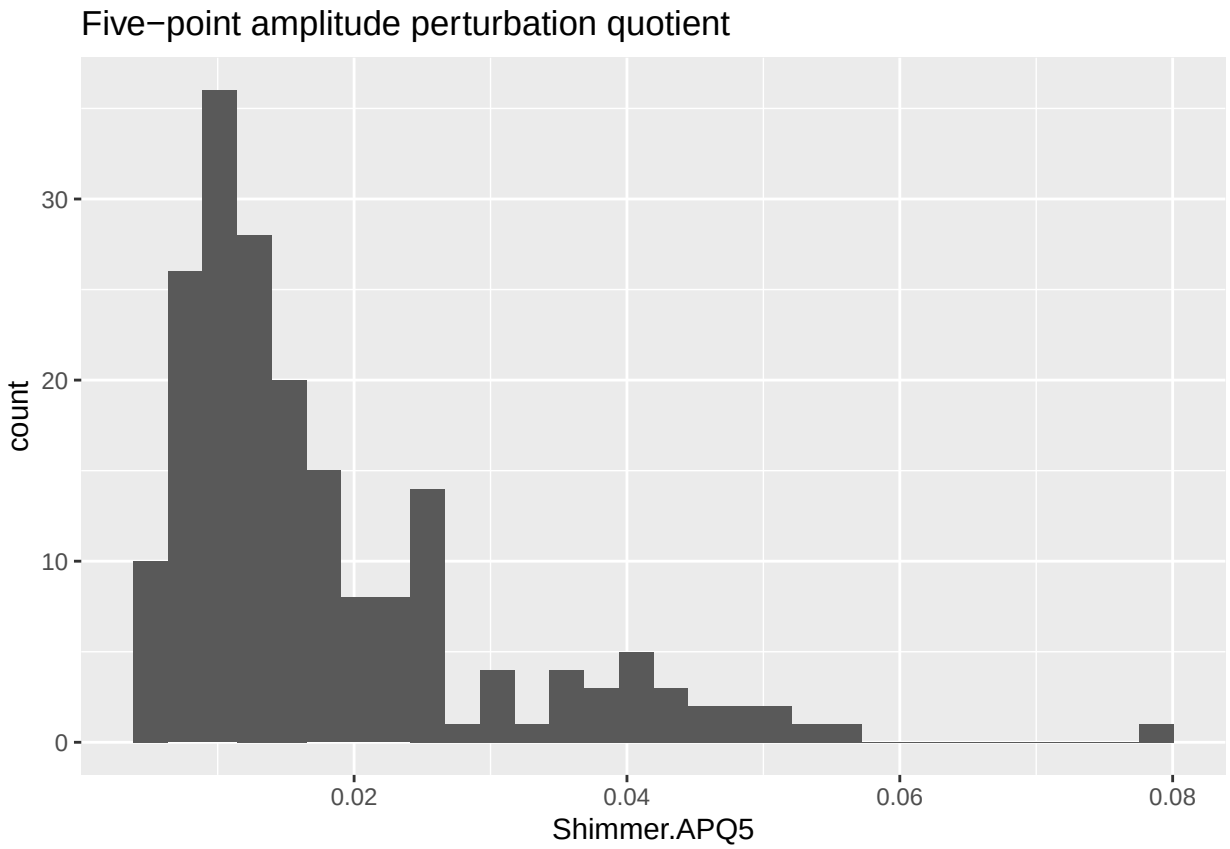

```
ggplot(df, aes(x=Shimmer.APQ3)) + geom_histogram() + ggtitle("Three-point amplitude perturbation quotie
```

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



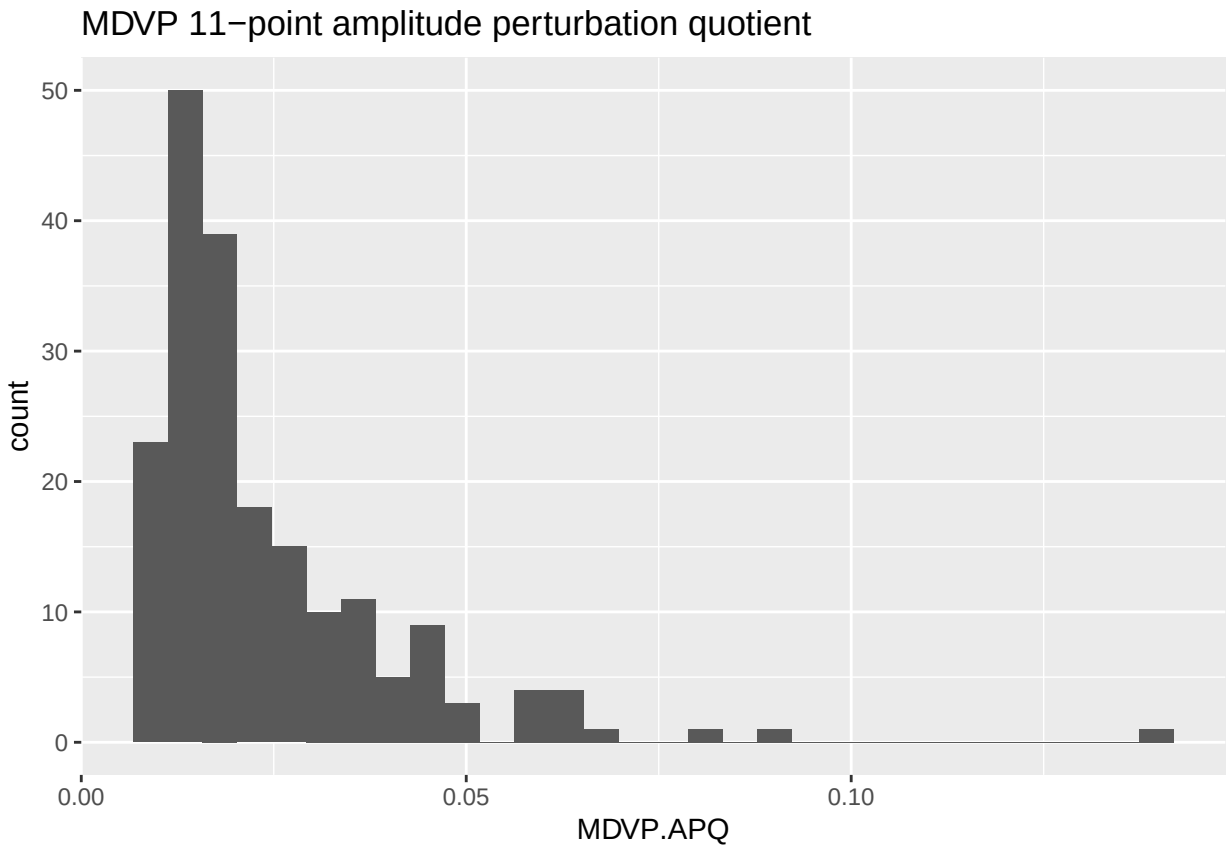

```
ggplot(df, aes(x=Shimmer.APQ5)) + geom_histogram() + ggtitle("Five-point amplitude perturbation quotien
```

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



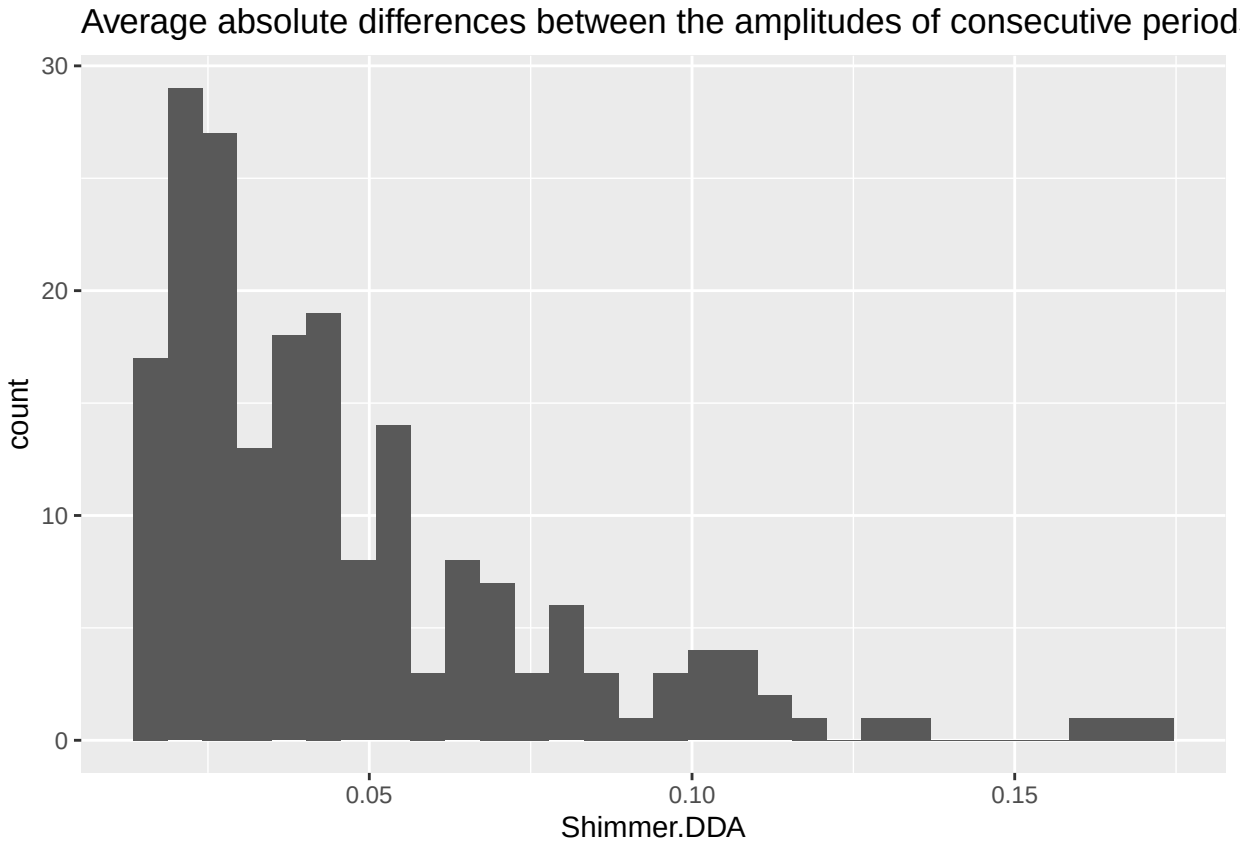

```
ggplot(df, aes(x=MDVP.APQ)) + geom_histogram() + ggtitle("MDVP 11-point amplitude perturbation quotient
```

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




```
ggplot(df, aes(x=Shimmer.DDA)) + geom_histogram() + ggtitle("Average absolute differences between the a
```

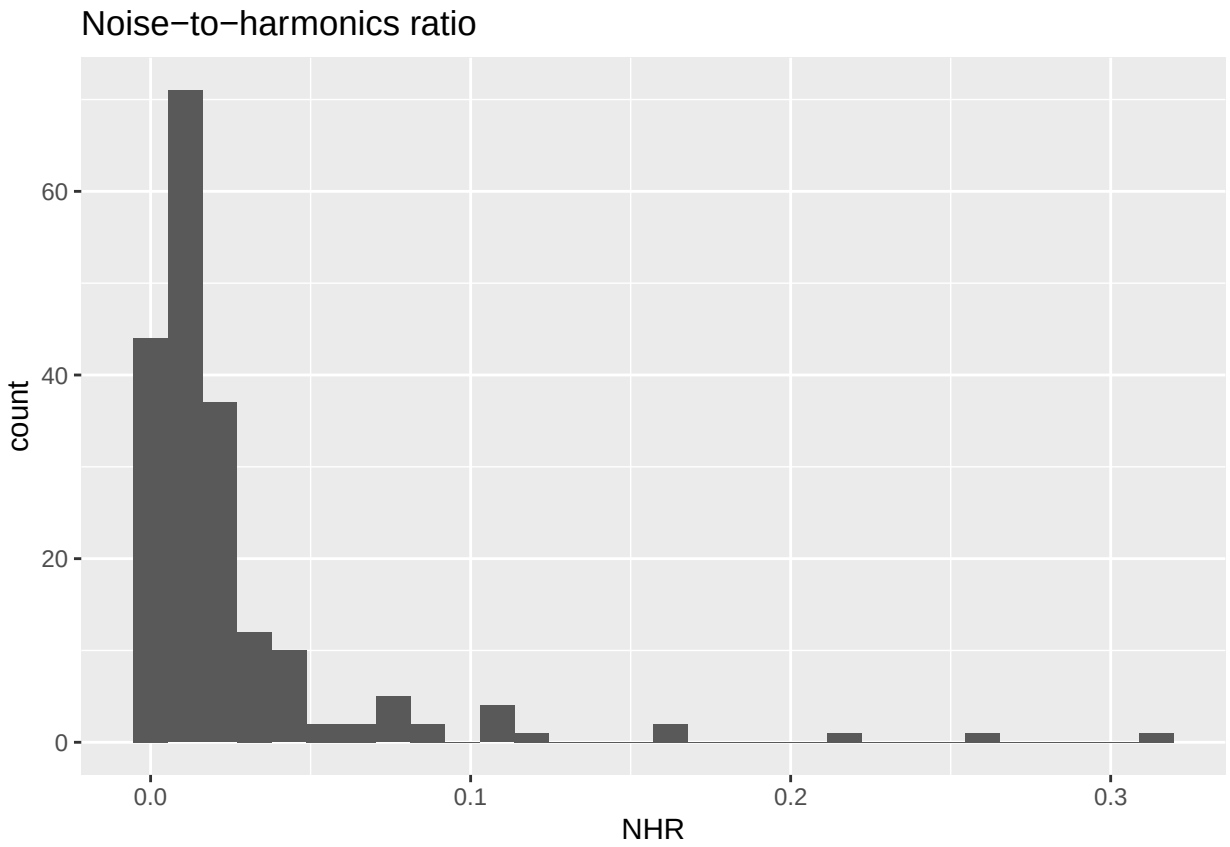
```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```



Create graphs of measures of ratio of noise to tonal components in the voice

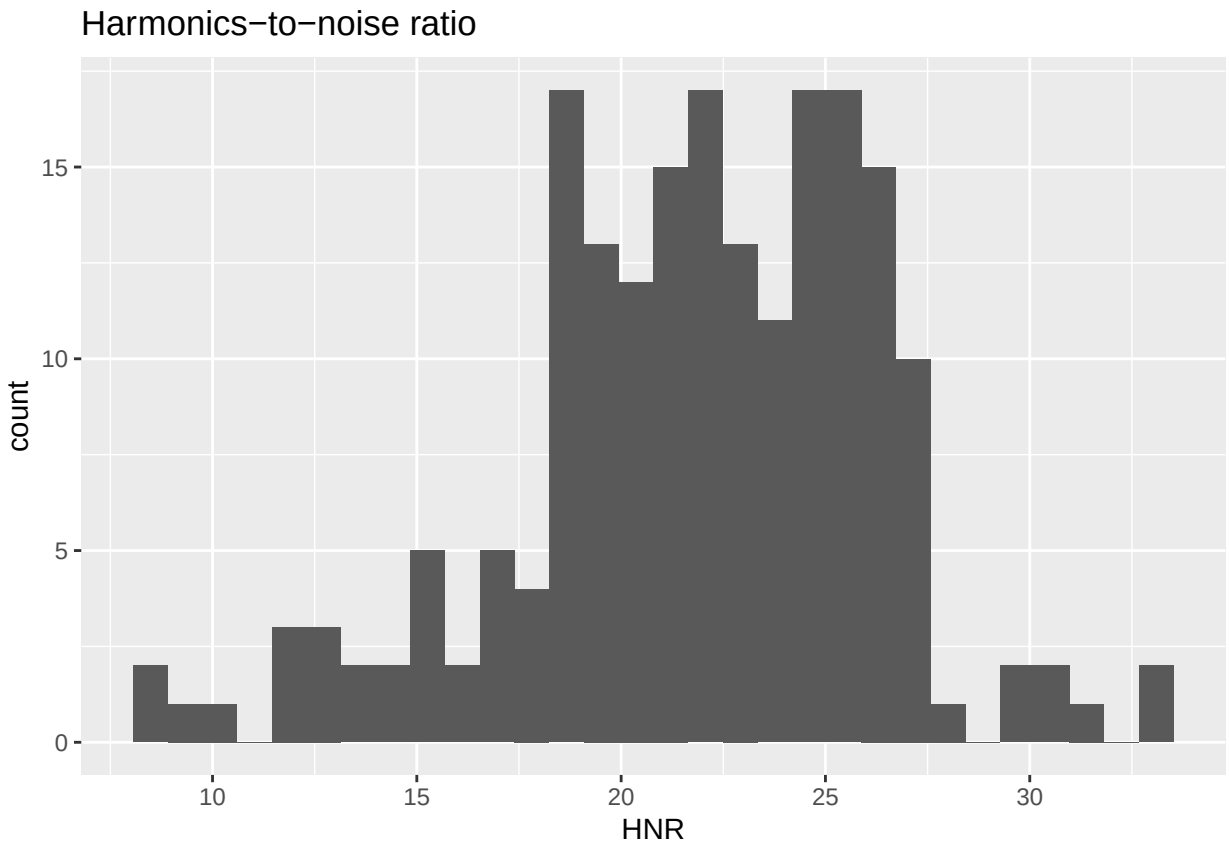
```
ggplot(df, aes(x=NHR)) + geom_histogram() + ggtitle("Noise-to-harmonics ratio")
```

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




```
ggplot(df, aes(x=HNR)) + geom_histogram() + ggtitle("Harmonics-to-noise ratio")
```

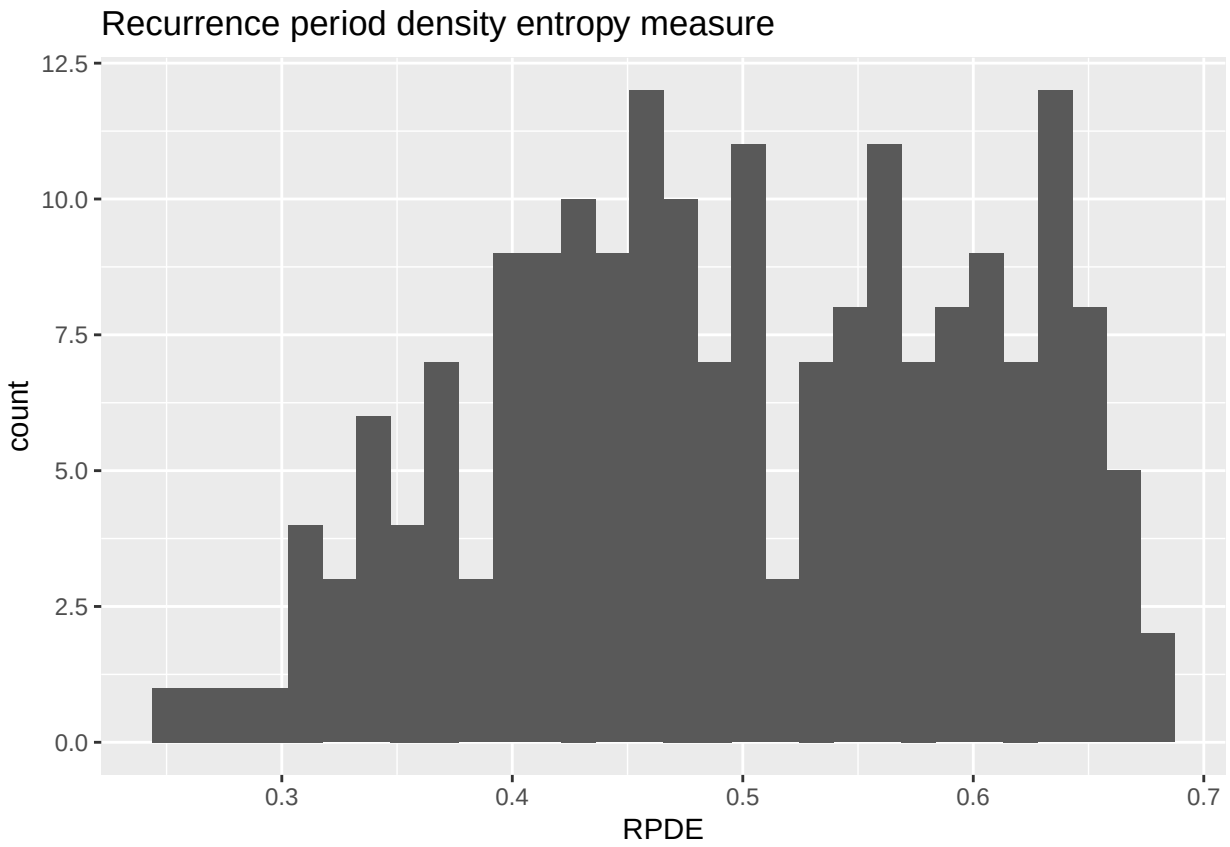
```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```



Create graphs of nonlinear dynamical complexity measures

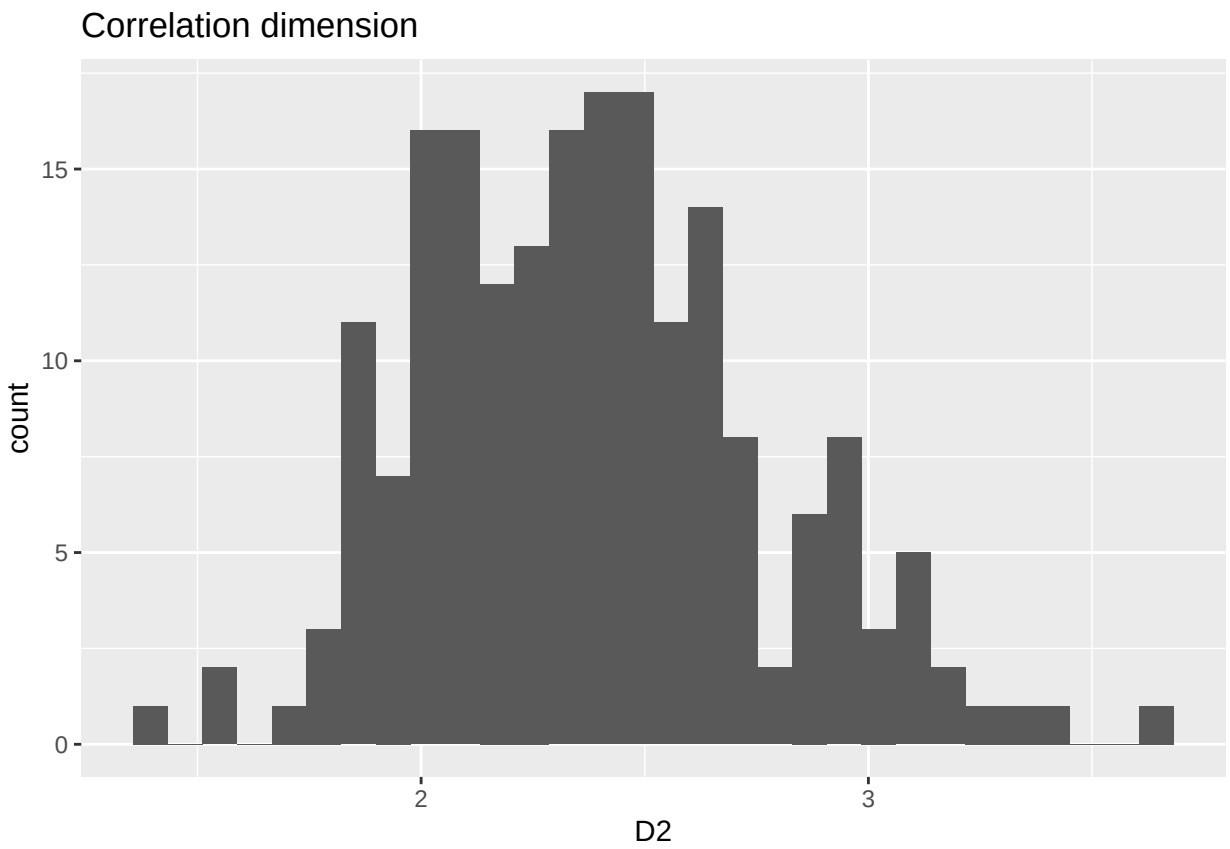
```
ggplot(df, aes(x=RPDE)) + geom_histogram() + ggtitle("Recurrence period density entropy measure")
```

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




```
ggplot(df, aes(x=D2 )) + geom_histogram() + ggtitle("Correlation dimension")
```

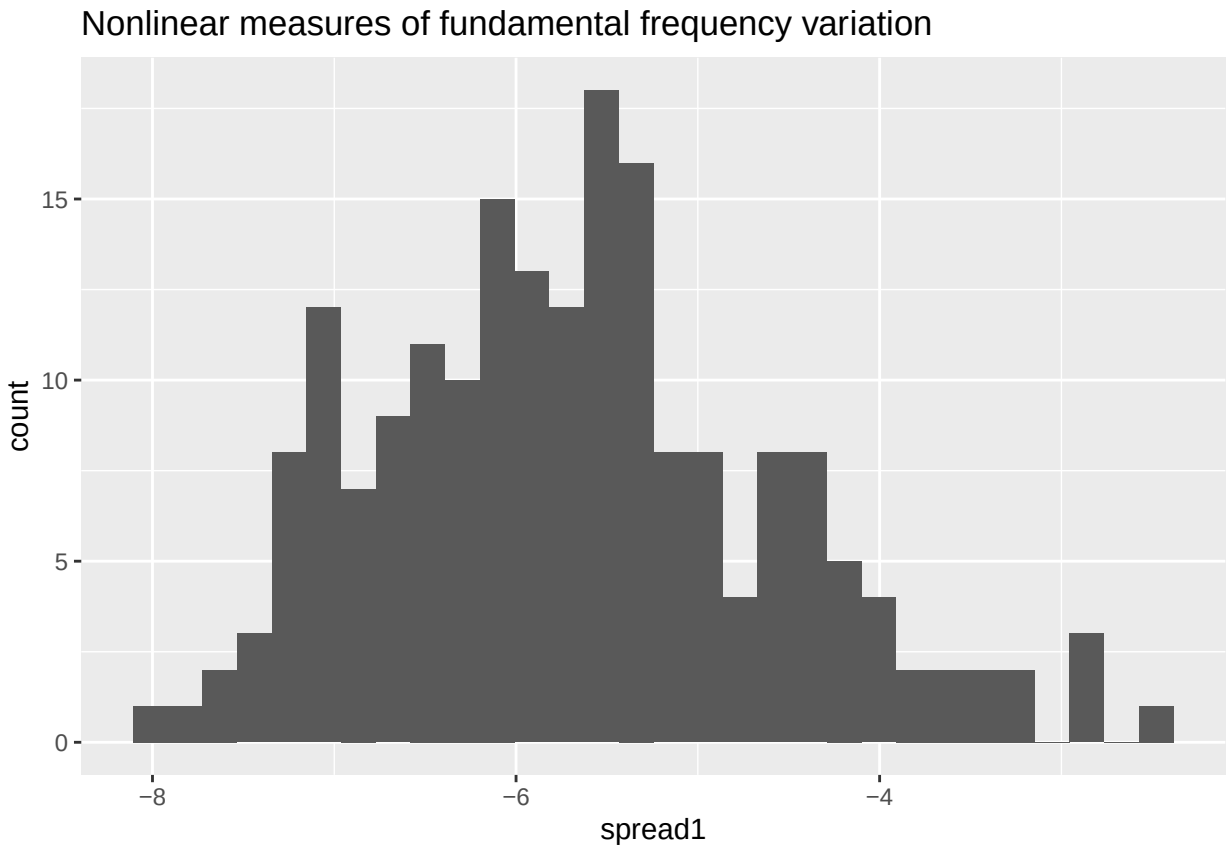
```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```



Create graphs of nonlinear dynamical complexity measures

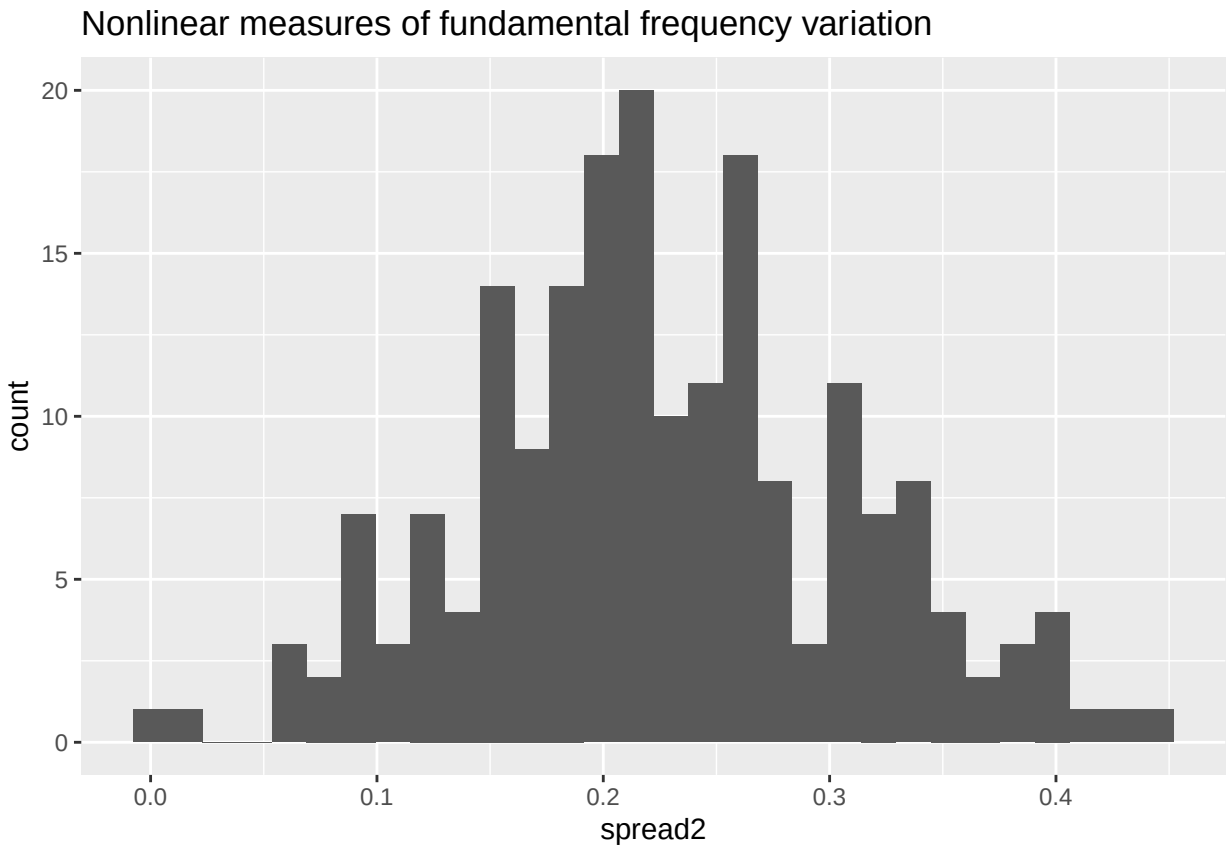
```
ggplot(df, aes(x=spread1)) + geom_histogram() + ggtitle("Nonlinear measures of fundamental frequency va
```

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



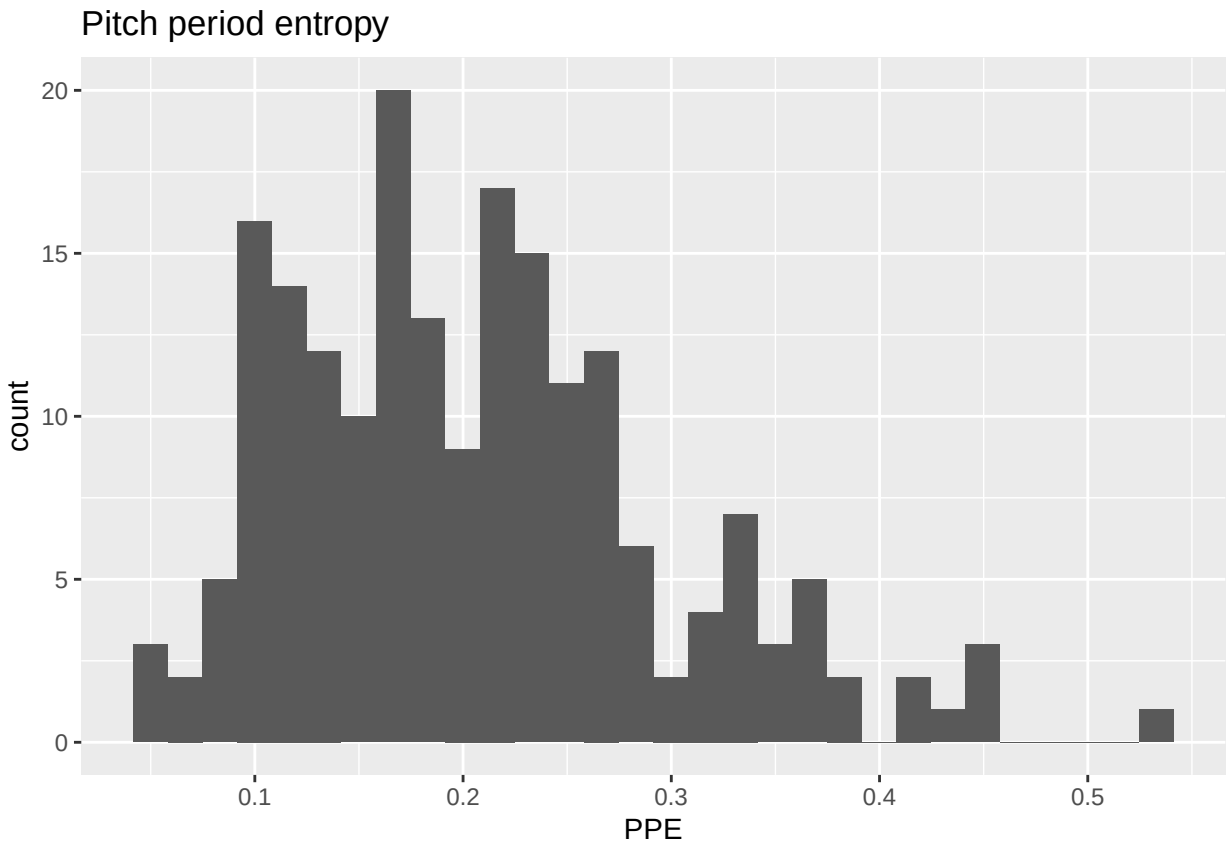

```
ggplot(df, aes(x=spread2)) + geom_histogram() + ggtitle("Nonlinear measures of fundamental frequency va
```

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




```
ggplot(df, aes(x=PPE)) + geom_histogram() + ggtitle("Pitch period entropy")
```

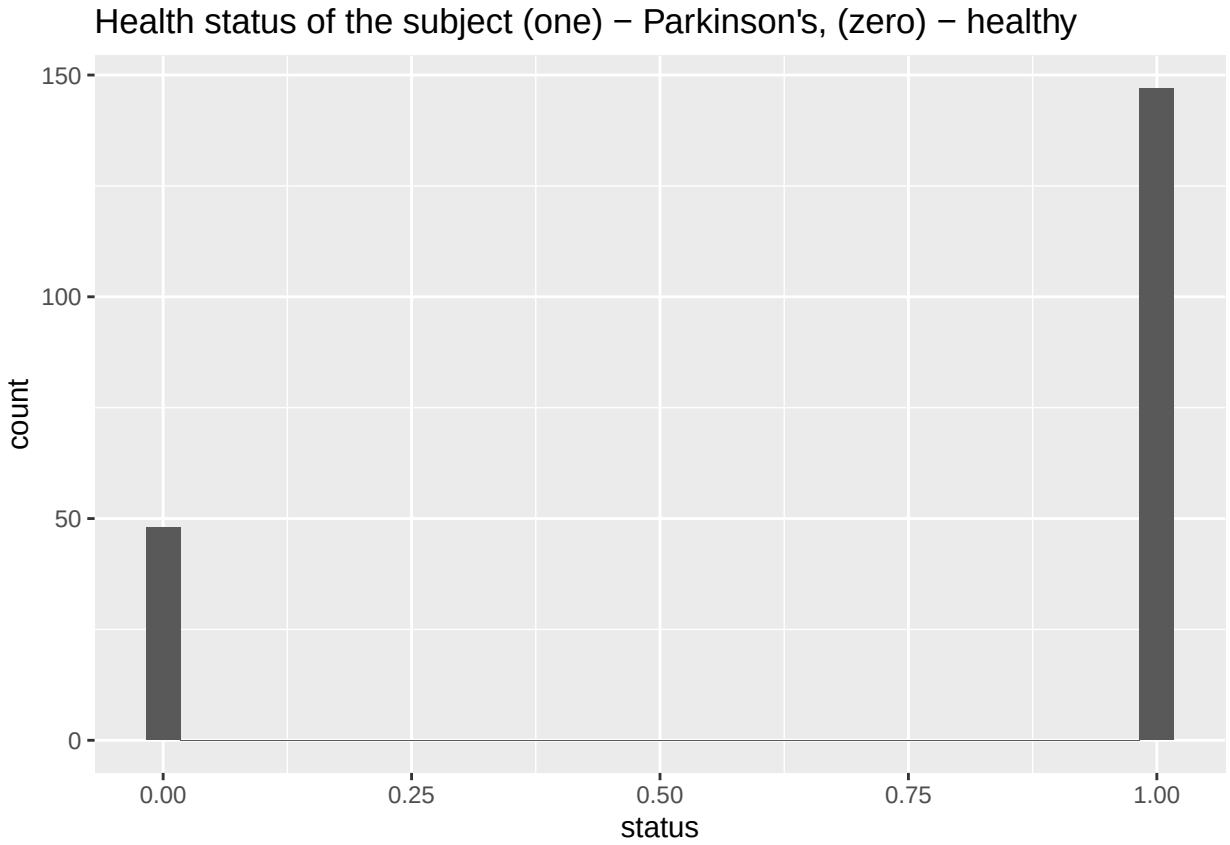
```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```



Create graphs of Parkinson's status and signal fractal scaling exponent

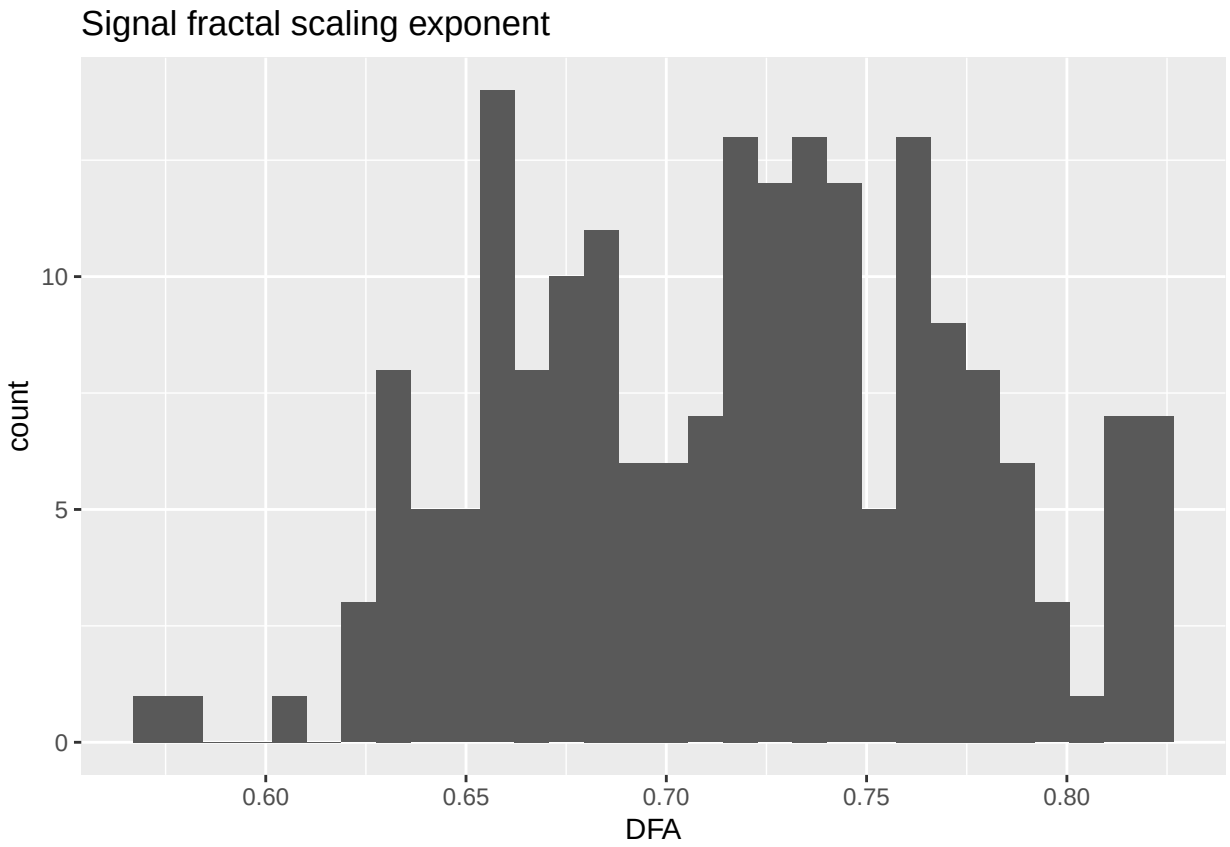
```
ggplot(df, aes(x=status)) + geom_histogram() + ggtitle("Health status of the subject (one) - Parkinson's
```

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




```
ggplot(df, aes(x=DFA)) + geom_histogram() + ggtitle("Signal fractal scaling exponent")
```

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



Create Training and Testing Datasets

```
#create a list of random number ranging from 1 to number of rows from actual data and 70% of the data i

data = sort(sample(nrow(df), nrow(df)*.7))

#creating training data set by selecting the output row values
train<-df[data,]

#creating test data set by not selecting the output row values
test<-df[-data,]
```

Create regression model

```
model <- glm(status ~ NHR, data = train, family = "binomial")
```

Output model results

```
summary(model)
```

```
##
## Call:
## glm(formula = status ~ NHR, family = "binomial", data = train)
##
## Deviance Residuals:
```

```
##      Min        1Q    Median        3Q       Max
## -2.81417   0.00007   0.45811   0.82621   1.11225
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept) -0.05248    0.35145  -0.149  0.88130
## NHR         89.91209   27.90332   3.222  0.00127 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 148.40  on 135  degrees of freedom
## Residual deviance: 125.84  on 134  degrees of freedom
## AIC: 129.84
##
## Number of Fisher Scoring iterations: 7
```

Make predictions

```
#Make predictions
predict <- model %>% predict(test, type = "terms")

head(predict)
```

```
##           NHR
## 1  -0.4646472
## 2  -0.7181992
## 3  -1.2756542
## 4  -1.2360929
## 9  -1.4905441
## 10 -1.5337019
```

```
summary(predict)
```

```
##       NHR
##  Min.   :-2.3312
##  1st Qu.:-1.9104
##  Median :-1.5337
##  Mean   :-0.7223
##  3rd Qu.:-0.7249
##  Max.   : 7.1815
```

Create regression model using all variables

```
model2 <- glm(status ~ MDVP.Fo.Hz.+MDVP.Flo.Hz.+MDVP.Jitter...+MDVP.Jitter.Abs.+MDVP.RAP + MDVP.PPQ + J
```

Output model results

```
summary(model2)
```

```
##
## Call:
## glm(formula = status ~ MDVP.Fo.Hz. + MDVP.Flo.Hz. + MDVP.Jitter... +
##     MDVP.Jitter.Abs. + MDVP.RAP + MDVP.PPQ + Jitter.DDP + MDVP.Shimmer +
##     MDVP.Shimmer.dB. + Shimmer.APQ3 + Shimmer.APQ5 + MDVP.APQ +
##     Shimmer.DDA + NHR + HNR + RPDE + DFA + spread1 + spread2 +
##     D2 + PPE, family = "binomial", data = train)
##
## Deviance Residuals:
##      Min        1Q    Median        3Q       Max
## -2.25111   0.00011   0.09361   0.34962   1.90785
##
## Coefficients:
##                     Estimate Std. Error z value Pr(>|z|)
## (Intercept)      -6.562e+00  2.301e+01  -0.285    0.775
## MDVP.Fo.Hz.      -2.114e-02  2.969e-02  -0.712    0.477
## MDVP.Flo.Hz.      6.436e-04  1.548e-02   0.042    0.967
## MDVP.Jitter...   -1.429e+02  1.766e+03  -0.081    0.936
## MDVP.Jitter.Abs. -7.947e+04  1.207e+05  -0.659    0.510
## MDVP.RAP          3.575e+04  1.638e+05   0.218    0.827
## MDVP.PPQ         -2.508e+03  2.465e+03  -1.017    0.309
## Jitter.DDP       -1.084e+04  5.456e+04  -0.199    0.843
## MDVP.Shimmer      4.752e+02  1.059e+03   0.449    0.654
## MDVP.Shimmer.dB.  6.967e+00  2.658e+01   0.262    0.793
## Shimmer.APQ3      1.463e+05  1.525e+05   0.960    0.337
## Shimmer.APQ5     -1.708e+02  4.981e+02  -0.343    0.732
## MDVP.APQ          1.839e+02  4.357e+02   0.422    0.673
## Shimmer.DDA      -4.907e+04  5.088e+04  -0.964    0.335
## NHR              -2.804e+01  6.060e+01  -0.463    0.644
## HNR               4.214e-03  2.626e-01   0.016    0.987
## RPDE             -1.899e+00  5.896e+00  -0.322    0.747
## DFA               1.295e+01  1.151e+01   1.125    0.261
## spread1           1.174e+00  2.394e+00   0.490    0.624
## spread2           6.837e+00  7.716e+00   0.886    0.376
## D2                2.412e+00  1.899e+00   1.270    0.204
## PPE               1.856e+01  3.559e+01   0.521    0.602
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 148.402  on 135  degrees of freedom
## Residual deviance:  56.906  on 114  degrees of freedom
## AIC: 100.91
##
## Number of Fisher Scoring iterations: 9
```

Make predictions

```
#Make predictions
predict <- model %>% predict(test, type = "terms")

head(predict)
```

```
##           NHR
```

```
## 1  -0.4646472
## 2  -0.7181992
## 3  -1.2756542
## 4  -1.2360929
## 9  -1.4905441
## 10 -1.5337019
```

```
summary(predict)
```

```
##       NHR
##  Min.   :-2.3312
##  1st Qu.:-1.9104
##  Median :-1.5337
##  Mean   :-0.7223
##  3rd Qu.:-0.7249
##  Max.   : 7.1815
```

Create regression model using vocal fundamental frequency variables

```
model3 <- glm(status ~ MDVP.Fo.Hz.+MDVP.Flo.Hz.+MDVP.Jitter...+MDVP.Jitter.Abs.+MDVP.RAP+ MDVP.PPQ + Ji
```

Output model results

```
summary(model3)
```

```
##
## Call:
## glm(formula = status ~ MDVP.Fo.Hz. + MDVP.Flo.Hz. + MDVP.Jitter... +
##     MDVP.Jitter.Abs. + MDVP.RAP + MDVP.PPQ + Jitter.DDP + MDVP.Shimmer +
##     MDVP.Shimmer.dB. + Shimmer.APQ3, family = "binomial", data = train)
##
## Deviance Residuals:
##      Min        1Q    Median        3Q       Max
## -2.49448   0.00005   0.12638   0.47939   1.33014
##
## Coefficients:
##                    Estimate Std. Error z value Pr(>|z|)
## (Intercept)      -1.395e-01  2.739e+00  -0.051  0.95939
## MDVP.Fo.Hz.      -2.369e-02  1.612e-02  -1.470  0.14160
## MDVP.Flo.Hz.      1.421e-03  8.984e-03   0.158  0.87433
## MDVP.Jitter...    1.211e+03  1.009e+03   1.200  0.23003
## MDVP.Jitter.Abs. -7.938e+03  7.302e+04  -0.109  0.91344
## MDVP.RAP         -4.061e+04  1.190e+05  -0.341  0.73297
## MDVP.PPQ         -2.222e+03  1.336e+03  -1.664  0.09617 .
## Jitter.DDP        1.383e+04  3.971e+04   0.348  0.72765
## MDVP.Shimmer      1.165e+03  4.271e+02   2.727  0.00639 **
## MDVP.Shimmer.dB. -3.473e+01  2.068e+01  -1.679  0.09315 .
## Shimmer.APQ3     -1.295e+03  5.506e+02  -2.352  0.01866 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
```

```
##
##     Null deviance: 148.402  on 135  degrees of freedom
## Residual deviance:  80.188  on 125  degrees of freedom
## AIC: 102.19
##
## Number of Fisher Scoring iterations: 8
```

Create regression model using vocal fundamental frequency variables

```
model4 <- glm(status ~ Shimmer.DDA + HNR + NHR + DFA+RPDE+ spread1 + spread2 + D2 + PPE, data = train, f
```

Output model results

```
summary(model4)
```

```
##
## Call:
## glm(formula = status ~ Shimmer.DDA + HNR + NHR + DFA + RPDE +
##     spread1 + spread2 + D2 + PPE, family = "binomial", data = train)
##
## Deviance Residuals:
##      Min        1Q    Median        3Q       Max
## -2.02207   0.01448   0.13554   0.43582   1.90946
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)  -1.93800   18.33488  -0.106   0.9158
## Shimmer.DDA  20.11160   22.69021   0.886   0.3754
## HNR           0.03457    0.18632   0.186   0.8528
## NHR         -10.61286   28.23185  -0.376   0.7070
## DFA          15.95033    8.35878   1.908   0.0564 .
## RPDE         -0.31963    4.86046  -0.066   0.9476
## spread1       2.40726    1.88809   1.275   0.2023
## spread2       6.50613    6.03280   1.078   0.2808
## D2            2.41694    1.44105   1.677   0.0935 .
## PPE          -7.59345   26.35854  -0.288   0.7733
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 148.40  on 135  degrees of freedom
## Residual deviance:  69.53  on 126  degrees of freedom
## AIC: 89.53
##
## Number of Fisher Scoring iterations: 8
```

Create regression model using vocal fundamental frequency variables

```
model5 <- glm(status ~  Shimmer.APQ5 +MDVP.APQ, data = train, family = "binomial")
```

```
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
```

Output model results

```
summary(model5)
```

```
##
## Call:
## glm(formula = status ~ Shimmer.APQ5 + MDVP.APQ, family = "binomial",
##     data = train)
##
## Deviance Residuals:
##      Min        1Q    Median        3Q       Max
## -2.71065   0.00004   0.09935   0.52402   1.53873
##
## Coefficients:
##               Estimate Std. Error z value Pr(>|z|)
## (Intercept)    -3.4160     0.8716  -3.919 8.88e-05 ***
## Shimmer.APQ5 -554.0847   151.7804  -3.651 0.000262 ***
## MDVP.APQ      709.7304   162.4452   4.369 1.25e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 148.402  on 135  degrees of freedom
## Residual deviance:  88.754  on 133  degrees of freedom
## AIC: 94.754
##
## Number of Fisher Scoring iterations: 8
```

Based on XGBoost restuls in Python

```
modeltest <- glm(status ~   MDVP.Fo.Hz.+MDVP.Fhi.Hz.+Shimmer.APQ5+ spread1+ PPE, data = train, family =
```

Output model results

```
summary(modeltest)
```

```
##
## Call:
## glm(formula = status ~ MDVP.Fo.Hz. + MDVP.Fhi.Hz. + Shimmer.APQ5 +
##     spread1 + PPE, family = "binomial", data = train)
##
## Deviance Residuals:
##      Min        1Q    Median        3Q       Max
## -2.56593   0.01421   0.19500   0.45886   1.67202
##
## Coefficients:
##                Estimate Std. Error z value Pr(>|z|)
## (Intercept)   20.076998  16.037568   1.252    0.211
## MDVP.Fo.Hz.   -0.011187   0.010169  -1.100    0.271
## MDVP.Fhi.Hz.   0.001901   0.005875   0.324    0.746
## Shimmer.APQ5  63.396783  52.495443   1.208    0.227
```

```
## spread1        2.669416   1.883337   1.417    0.156
## PPE          -11.000771  26.101101  -0.421    0.673
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 148.402  on 135  degrees of freedom
## Residual deviance:  79.098  on 130  degrees of freedom
## AIC: 91.098
##
## Number of Fisher Scoring iterations: 7
```

Make predictions

```
#Make predictions
predict <- modeltest %>% predict(test, type = "terms")

head(predict)
```

```
##    MDVP.Fo.Hz. MDVP.Fhi.Hz. Shimmer.APQ5   spread1        PPE
## 1    0.4116710  -0.07510303   0.81065373 2.2975321 -0.8602506
## 2    0.3847317  -0.09155431   1.69060107 4.2671316 -1.7845353
## 3    0.4487013  -0.12490370   1.27218231 3.2848211 -1.3880675
## 4    0.4487685  -0.11204995   1.36537558 4.1541913 -1.7878465
## 9    0.6831001  -0.12308402  -0.36408866 0.4672548 -0.2762967
## 10   0.6906404  -0.14583477  -0.08007107 1.7667240 -0.7140283
```

```
summary(predict)
```

```
##   MDVP.Fo.Hz.       MDVP.Fhi.Hz.        Shimmer.APQ5        spread1
##  Min.   :-0.9628   Min.   :-0.159765   Min.   :-0.8085   Min.   :-5.61636
##  1st Qu.:-0.2124   1st Qu.:-0.121842   1st Qu.:-0.6351   1st Qu.:-2.11257
##  Median : 0.2222   Median :-0.062869   Median :-0.3673   Median :-0.27294
##  Mean   : 0.0947   Mean   : 0.001917   Mean   :-0.1330   Mean   :-0.09421
##  3rd Qu.: 0.4360   3rd Qu.: 0.038342   3rd Qu.: 0.1570   3rd Qu.: 1.68902
##  Max.   : 0.7259   Max.   : 0.751507   Max.   : 2.2662   Max.   : 6.34267
##       PPE
##  Min.   :-2.762053
##  1st Qu.:-0.488243
##  Median : 0.109126
##  Mean   :-0.003518
##  3rd Qu.: 0.810288
##  Max.   : 1.637408
```