Cardiotocography (CTG) is used during pregnancy to monitor fetal heart rate and uterine contractions. It is monitor fetal well-being and allows early detection of fetal distress.

CTG interpretation helps in determining if the pregnancy is high or low risk. An abnormal CTG may indicate the need for further investigations and potential intervention.

In this project, I will create a model to classify the outcome of Cardiotocogram test to ensure the well being of the fetus.

In [1]:
```
import pandas as pd
import numpy as np
import seaborn as sns
import matplotlib.pyplot as plt
from sklearn.model_selection import train_test_split
```

In [2]:
```
fetal_health = pd.read_csv('fetal_health.csv')
```

In [3]:
```
fetal_health.head()
```

Out[3]:

	baseline value	accelerations	fetal_movement	uterine_contractions	light_decelerations	severe_decelerations	prolongued_decelerations	abnorma
0	120	0.000	0.0	0.000	0.000	0.0	0.0	
1	132	0.006	0.0	0.006	0.003	0.0	0.0	
2	133	0.003	0.0	0.008	0.003	0.0	0.0	
3	134	0.003	0.0	0.008	0.003	0.0	0.0	
4	132	0.007	0.0	0.008	0.000	0.0	0.0	

5 rows × 22 columns

In [4]:
```
fetal_health.tail()
```

Out[4]:

	baseline value	accelerations	fetal_movement	uterine_contractions	light_decelerations	severe_decelerations	prolongued_decelerations	abnor
2121	140	0.000	0.000	0.007	0.0	0.0	0.0	
2122	140	0.001	0.000	0.007	0.0	0.0	0.0	
2123	140	0.001	0.000	0.007	0.0	0.0	0.0	
2124	140	0.001	0.000	0.006	0.0	0.0	0.0	
2125	142	0.002	0.002	0.008	0.0	0.0	0.0	

5 rows × 22 columns

In [5]:
```
fetal_health.isnull().sum()
```

Out[5]:
```
baseline value                                                0
accelerations                                                 0
fetal_movement                                                0
uterine_contractions                                          0
light_decelerations                                           0
severe_decelerations                                          0
prolongued_decelerations                                      0
abnormal_short_term_variability                               0
mean_value_of_short_term_variability                          0
percentage_of_time_with_abnormal_long_term_variability        0
mean_value_of_long_term_variability                           0
histogram_width                                               0
histogram_min                                                 0
histogram_max                                                 0
histogram_number_of_peaks                                     0
histogram_number_of_zeroes                                    0
histogram_mode                                                0
histogram_mean                                                0
histogram_median                                              0
histogram_variance                                            0
histogram_tendency                                            0
fetal_health                                                  0
dtype: int64
```

In [6]:
```
fetal_health.describe()
```

Out[6]:

	baseline value	accelerations	fetal_movement	uterine_contractions	light_decelerations	severe_decelerations	prolongued_decelerations	a
count	2126.000000	2126.000000	2126.000000	2126.000000	2126.000000	2126.000000	2126.000000	
mean	133.303857	0.003178	0.009481	0.004366	0.001889	0.000003	0.000159	
std	9.840844	0.003866	0.046666	0.002946	0.002960	0.000057	0.000590	
min	106.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	
25%	126.000000	0.000000	0.000000	0.002000	0.000000	0.000000	0.000000	
50%	133.000000	0.002000	0.000000	0.004000	0.000000	0.000000	0.000000	
75%	140.000000	0.006000	0.003000	0.007000	0.003000	0.000000	0.000000	
max	160.000000	0.019000	0.481000	0.015000	0.015000	0.001000	0.005000	

8 rows × 22 columns

In [7]:
```
fetal_health.duplicated().sum()
```

Out[7]:
```
13
```

In [8]:
```
fetal_health = fetal_health.drop_duplicates()
```

In [9]:
```
fetal_health.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Index: 2113 entries, 0 to 2125
Data columns (total 22 columns):
 #   Column                                                Non-Null Count  Dtype
---  ------                                                --------------  -----
 0   baseline value                                        2113 non-null   int64
 1   accelerations                                         2113 non-null   float64
 2   fetal_movement                                        2113 non-null   float64
 3   uterine_contractions                                  2113 non-null   float64
 4   light_decelerations                                   2113 non-null   float64
 5   severe_decelerations                                  2113 non-null   float64
 6   prolongued_decelerations                              2113 non-null   float64
 7   abnormal_short_term_variability                       2113 non-null   int64
 8   mean_value_of_short_term_variability                  2113 non-null   float64
 9   percentage_of_time_with_abnormal_long_term_variability  2113 non-null   int64
 10  mean_value_of_long_term_variability                   2113 non-null   float64
 11  histogram_width                                       2113 non-null   int64
 12  histogram_min                                         2113 non-null   int64
 13  histogram_max                                         2113 non-null   int64
 14  histogram_number_of_peaks                             2113 non-null   int64
 15  histogram_number_of_zeroes                            2113 non-null   int64
 16  histogram_mode                                        2113 non-null   int64
 17  histogram_mean                                        2113 non-null   int64
 18  histogram_median                                      2113 non-null   int64
 19  histogram_variance                                    2113 non-null   int64
 20  histogram_tendency                                    2113 non-null   int64
 21  fetal_health                                          2113 non-null   int64
dtypes: float64(8), int64(14)
memory usage: 379.7 KB
```

In [10]:
```
fetal_health.shape
```

Out[10]:
```
(2113, 22)
```

In [11]:
```
fetal_health.columns
```

Out[11]:
```
Index(['baseline value', 'accelerations', 'fetal_movement',
       'uterine_contractions', 'light_decelerations', 'severe_decelerations',
       'prolongued_decelerations', 'abnormal_short_term_variability',
       'mean_value_of_short_term_variability',
       'percentage_of_time_with_abnormal_long_term_variability',
       'mean_value_of_long_term_variability', 'histogram_width',
       'histogram_min', 'histogram_max', 'histogram_number_of_peaks',
       'histogram_number_of_zeroes', 'histogram_mode', 'histogram_mean',
       'histogram_median', 'histogram_variance', 'histogram_tendency',
       'fetal_health'],
      dtype='object')
```

In [12]:
```
df = fetal_health[['baseline value', 'accelerations', 'fetal_movement',
       'uterine_contractions', 'light_decelerations', 'severe_decelerations',
       'prolongued_decelerations', 'abnormal_short_term_variability',
       'mean_value_of_short_term_variability',
       'percentage_of_time_with_abnormal_long_term_variability',
       'mean_value_of_long_term_variability', 'histogram_width',
       'histogram_min', 'histogram_max', 'histogram_number_of_peaks',
       'histogram_number_of_zeroes', 'histogram_mode', 'histogram_mean',
       'histogram_median', 'histogram_variance', 'histogram_tendency',
       'fetal_health']]
```

In [13]:
```
for i in df.columns:
    print(i)
    print(df[i].unique())
```

```
baseline value
[120 132 133 134 122 151 150 131 130 129 128 124 115 114 116 158 156 148
 149 146 144 142 136 141 138 140 154 145 139 125 123 159 143 119 121 127
 126 118 135 137 147 157 117 152 112 106 110 160]
accelerations
[0.    0.006 0.003 0.007 0.001 0.005 0.009 0.002 0.008 0.004 0.01  0.015
 0.013 0.014 0.011 0.017 0.012 0.016 0.019 0.018]
fetal_movement
[0.    0.072 0.222 0.408 0.38  0.441 0.383 0.451 0.469 0.34  0.425 0.334
 0.135 0.099 0.108 0.112 0.089 0.103 0.085 0.109 0.079 0.065 0.055 0.058
 0.047 0.038 0.012 0.018 0.02  0.005 0.003 0.006 0.001 0.004 0.009 0.01
 0.002 0.008 0.007 0.028 0.026 0.107 0.013 0.016 0.029 0.05  0.053 0.011
 0.015 0.022 0.021 0.017 0.019 0.025 0.014 0.024 0.023 0.035 0.054 0.03
 0.048 0.088 0.043 0.052 0.091 0.033 0.092 0.084 0.115 0.041 0.04  0.027
 0.031 0.063 0.06  0.071 0.306 0.298 0.139 0.189 0.157 0.235 0.36  0.455
 0.443 0.47  0.477 0.446 0.481 0.369 0.335 0.43  0.346 0.323 0.375 0.353
 0.045 0.032 0.051 0.036 0.037 0.049]
uterine_contractions
[0.    0.006 0.008 0.01  0.013 0.002 0.003 0.001 0.004 0.005 0.007 0.009
 0.012 0.011 0.015 0.014]
light_decelerations
[0.    0.003 0.009 0.008 0.001 0.002 0.005 0.004 0.012 0.01  0.014 0.006
 0.007 0.015 0.011 0.013]
severe_decelerations
[0.    0.001]
prolongued_decelerations
[0.    0.002 0.003 0.001 0.004 0.005]
abnormal_short_term_variability
[73 17 16 26 29 83 84 86 64 28 21 19 24 18 23 30 34 80 87 22 27 40 20 43
 61 70 57 58 39 41 33 25 44 46 45 52 53 56 65 36 35 38 62 51 67 68 72 63
 60 47 50 69 54 55 66 48 31 32 37 49 59 14 12 13 74 75 77 71 76 79 78 81
 42 82 15]
mean_value_of_short_term_variability
[0.5 2.1 2.4 5.9 6.3 0.3 1.9 2.  1.4 1.5 2.3 1.7 2.5 0.4 0.2 4.4 6.  4.5
 6.9 2.9 3.4 3.2 3.7 3.6 2.2 1.6 1.8 4.7 4.9 5.  7.  1.3 4.1 5.4 1.2 0.8
 1.1 0.9 1.  0.7 2.8 3.9 5.2 4.8 4.3 0.6 5.3 2.6 3.3 3.8 2.7 3.1 5.7 4.
 3.5 3.  4.2]
percentage_of_time_with_abnormal_long_term_variability
[43  0  6  5  9  8 79 72 14 71  1 40 69 54 53 38 29 18 21 37 20  2  3 27
 67 68 75 74  7 30 49 39 31 32 19 16 11 34 58 13 10 15 12 25 26  4 46 57
 33 23 22 51 52 59 78 84 62 24 45 35 41 61 56 44 81 42 77 88 91 47 55 82
 86 60 28 17 70 48 50 36 85 64 66 90 73 65 63]
mean_value_of_long_term_variability
[ 2.4 10.4 13.4 23.  19.9  0.  15.6 13.6 10.6 27.6 29.5 12.9  5.4  7.9
  8.7 10.9 13.9  8.8  7.8  8.5  6.7  4.   6.8  2.9  4.8  3.4 10.5  5.
 12.5  6.3 15.1 21.7 12.4 24.2 18.8 16.2 19.6 12.2 13.5 13.3 11.4 15.2
 15.7 22.1 21.1 16.6 14.9 12.8 22.2  3.3  6.2  5.1 13.2  5.5  6.4  5.2
 11.6 11.2  5.7 11.9 10.3 13.  11.5 17.7 22.3 21.3 25.9 18.4 15.8  1.2
 23.4  7.1  6.5  4.9  5.3  4.5  4.6  9.5 14.6 17.3 11.  17.2 13.1 12.1
 19.3  4.4 10.1 19.4  5.8  3.9  2.1  7.4  3.1  3.8  4.7 10.8  8.   8.6
  4.3  7.2  9.1  8.1 26.3 12.7 35.7 16.9 18.   9.   9.6  8.3 11.3  8.9
  3.7  3.2  8.4  8.2 41.8 50.7 25.8 15.9  5.6 10.7 14.  10.  14.1 14.3
 15.3 23.3 12.3  9.8  7.7  7.  14.8  7.6  7.5  2.5  6.1 17.9 12.6 15.4
 13.8  3.5 18.7 11.7  0.9  1.3 15.5 14.2  9.3  9.2  9.7 19.7 14.5  3.6
  6.9  2.8  7.3 16.4 16.7  4.2  9.4 11.8  9.9  6.6  3.  28.4 21.4 18.5
 19.  16.1 10.2 11.1 14.4 20.5  4.1  1.6 29.6 29.3 23.1 34.7 27.3 16.
 20.   5.9 18.9 17.1  2.7 19.8 24.7 21.5 25.6 26.1  1.4 36.9 29.  28.
```

```
 17.   6.  16.5 23.8 40.8 25.2  1.9 21.9  2.   1.7 12.  18.1 20.8 17.6
 15.  13.7 24.1 17.5 22.5 27.  17.8 20.4 29.1 28.9  2.3  0.5 25.3 22.7
 24.9 19.1 25.1 25.4 16.3  2.2 14.7 17.4 23.6  0.4  0.7 19.2  1.8  0.6
  1.1  2.6 33.5 27.4  0.3  0.2  1.  21.   0.1  1.5  0.8]
histogram_width
[ 64 130 117 150  68  66  87 107 125  99 112 128 141 145  16  12  24  10
 143 149 144 126 134 138 129  98  93 142 148 120  55  31  18  20 100  44
  28  26 111 103  60  37  30 104 108  92 119 140 109 136 113  15  17  25
  21  35  83  96  84  29  82 105  62 122 118  54 146  53  56  94  85  77
  91  11  23  40  42  22 147  90 121 106  52  48  46  39  33  34  81  65
  57 124  72  89 139 127  19  27  69  63  75  38  80  36  71  74  14   8
 110  76  79  58 114   7  78   9   5  97 132  13  95  41  47 123  51 135
 102 116  86 115 133  88  45  43 137  67  59  32 131 101  70  49  73   3
   6  61  50 162 176 161 153 163 158 180]
histogram_min
[ 62  68  53  50  56  88  71  67  59  65  54  57 114 118 122  55  51  52
  77 142 130 136  58 148 143  78 134  74  81 119 132 138  60  63  73 135
 139 128 121 101  72  97 117  80 111  95  92  75  61 153 156 159 133 149
 147 146 155  93 115 113 116  83  86  91  69 129 140 145 103 109 112  90
  82 125  64 120  89 141 107 106 110  66 131 144 126 137 108  99 151 154
 123  70 102  98 124  79  84 127  76 158 100 105  96  94 152 150  87  85
 104]
histogram_max
[126 198 170 200 130 186 154 158 174 178 177 182 199 146 128 191 193 194
 181 195 188 153 148 197 161 184 166 163 171 156 185 180 179 169 168 172
 176 196 187 167 152 165 150 142 189 144 138 173 183 141 147 149 143 192
 135 151 133 131 136 145 175 140 137 139 129 160 159 155 164 190 157 162
 134 132 127 213 210 125 123 122 238 205 204 211 228 230]
histogram_number_of_peaks
[ 2  6  5 11  9  0  1  7  3 13 10  8  4 12 14 15 16 18]
histogram_number_of_zeroes
[ 0  1  3  2  4 10  5  8  7]
histogram_mode
[120 141 137  76  71 122 150 135 143 134 133 129  75 126 128 124 123 121
 125 119 117 127 116 170 151 154 149 147 140 142 153 162 156 144 146 152
 148 145 167 165 161 131 132 136 139 114 163 159 138 158 157 186 187 180
 176 115 160  90 108  97 130 155 107 112 109 110 169  99  86 106 113 111
  98  95  91  88 164  89 179 105  67  60 100  93  77 103 104  69]
histogram_mean
[137 136 135 134 107 122 148 125 127 128 124 129 104  99 126 123 119 113
 120 112 116 115 117 118 168 171 142 131 152 155 153 154 145 140 143 156
 158 157 147 150 164 163 151 149 130 132 114 105 162 165 169 172 167 161
 146 121 133 160 141 144 139 138 159 166 178 180 182 175 173  98  90  87
 106 110 109 108 111  91 103 101  83  84  78  85  82 100 170 102  97  92
  93  95  81  79  76  89  88  75  94  96  86  80  73]
histogram_median
[121 140 138 137 107 106 123 151 141 135 133 132 129 120 102 125 127 117
 126 115 130 118 119 124 170 172 152 136 154 150 156 155 147 144 159 160
 153 149 158 166 165 148 145 157 128 134 113 164 174 168 162 163 139 122
 161 131 143 146 142 169 180 183 186 178 171 176 177 116  91 167  78  77
 112 114 111 110  86  79 108 105 109  95  94  97 101  98  93  99 100  92
  87  90  82 103 104]
histogram_variance
[ 73  12  13  11 170 215   3   1   9  10   7  76  43  70  45  36  27 138
  34 148   0  89  56  66  35  25  24  21  19  14  23  28  75  72 108   8
   4   2   5  18  15   6  16  55  20  86  60 117  39  30 137  65  41 136
  38 157 106 177  52  29  54  44  37  42  17  22  26  33  53  62  74  69
  79  83 100  31  49  51  95  61  47  68  40 104  48 103 129  57  85  46
  32  82  91  90  92  64  63  59 127 114  78  84  94  97  80  71 101 126
 134 144  50  87 115  96 116  77 119 110  98 113 195 182 128  88 109 121
 269 250 254 243 241 190 147]
histogram_tendency
[ 1  0 -1]
fetal_health
[2 1 3]
```

In [14]:
```
for i in df.columns:
    print(i)
    print(df[i].value_counts())
```

```
baseline value
baseline value
133    136
130    111
122    106
138    102
125     91
128     85
120     78
142     77
132     76
144     76
136     72
140     69
134     67
135     64
146     61
127     60
137     59
```

```
129     57
131     56
143     56
123     53
148     51
121     44
139     39
141     39
126     38
145     36
115     28
150     26
110     21
149     18
119     17
152     17
112     16
147     14
151     14
159     12
114     11
158     10
124     10
118      9
154      8
106      7
116      5
157      4
156      4
117      2
160      1
Name: count, dtype: int64
accelerations
accelerations
0.000    886
0.003    159
0.002    159
0.001    143
0.004    117
0.006    112
0.005    109
0.008    103
0.007     90
0.009     60
0.010     50
0.011     36
0.012     24
0.013     22
0.014     20
0.015      9
0.016      7
0.017      4
0.018      2
0.019      1
Name: count, dtype: int64
fetal_movement
fetal_movement
0.000    1302
0.001     164
0.002     112
0.003      87
0.004      48
         ...
0.079       1
0.109       1
0.103       1
0.031       1
0.099       1
Name: count, Length: 102, dtype: int64
uterine_contractions
uterine_contractions
0.000    323
0.005    290
0.004    242
0.006    231
0.007    216
0.003    211
0.008    160
0.002    159
0.001    118
0.009     82
0.010     49
0.011     16
0.012     11
0.013      2
0.014      2
0.015      1
Name: count, dtype: int64
light_decelerations
```

```
light_decelerations
0.000    1218
0.001     163
0.003     118
0.002     115
0.004     114
0.005     107
0.006      74
0.008      55
0.007      54
0.009      37
0.010      15
0.011      13
0.012      12
0.013       8
0.014       7
0.015       3
Name: count, dtype: int64
severe_decelerations
severe_decelerations
0.000    2106
0.001       7
Name: count, dtype: int64
prolongued_decelerations
prolongued_decelerations
0.000    1935
0.002      72
0.001      70
0.003      24
0.004       9
0.005       3
Name: count, dtype: int64
abnormal_short_term_variability
abnormal_short_term_variability
60    62
58    61
65    59
63    58
64    58
      ..
14     4
86     4
12     2
82     2
87     1
Name: count, Length: 75, dtype: int64
mean_value_of_short_term_variability
mean_value_of_short_term_variability
0.8    122
1.3    121
0.5    120
0.4    118
0.7    117
0.6    113
0.9    112
1.2    106
1.5    100
1.0     99
1.1     97
1.4     95
0.3     84
1.7     78
1.6     76
1.9     59
1.8     51
2.2     47
0.2     46
2.1     44
2.0     38
2.3     27
2.4     27
2.5     26
2.7     25
2.8     22
2.6     21
3.0     16
2.9     13
3.2     13
3.4     12
3.1     10
3.3      7
3.8      6
3.6      4
3.5      3
4.9      3
3.7      3
4.1      3
4.2      3
5.0      2
```

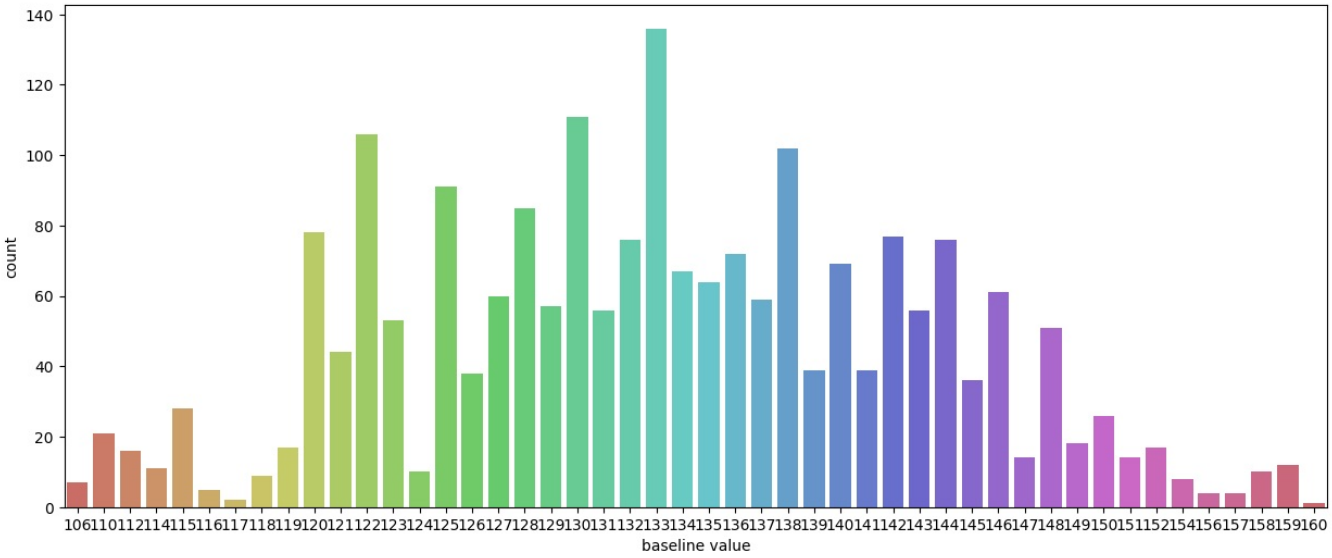
```
3.9      2
6.3      2
4.0      2
4.4      2
4.5      2
5.4      2
4.3      2
4.8      2
7.0      1
5.3      1
4.7      1
6.0      1
5.7      1
5.2      1
5.9      1
6.9      1
Name: count, dtype: int64
percentage_of_time_with_abnormal_long_term_variability
percentage_of_time_with_abnormal_long_term_variability
0     1235
1       52
2       44
5       43
4       39
      ...
85       1
79       1
86       1
88       1
63       1
Name: count, Length: 87, dtype: int64
mean_value_of_long_term_variability
mean_value_of_long_term_variability
0.0     137
7.1      29
6.7      29
6.5      25
5.2      25
       ...
24.1      1
17.5      1
22.5      1
27.0      1
0.8       1
Name: count, Length: 249, dtype: int64
histogram_width
histogram_width
39     39
102    35
27     30
31     29
98     28
       ..
6       1
162     1
131     1
146     1
180     1
Name: count, Length: 154, dtype: int64
histogram_min
histogram_min
50     76
52     50
120    48
71     47
60     45
       ..
155     2
149     2
158     1
159     1
156     1
Name: count, Length: 109, dtype: int64
histogram_max
histogram_max
157    71
171    65
158    62
156    58
159    58
       ..
123     2
122     2
134     2
213     1
205     1
Name: count, Length: 86, dtype: int64
histogram_number_of_peaks
histogram_number_of_peaks
```

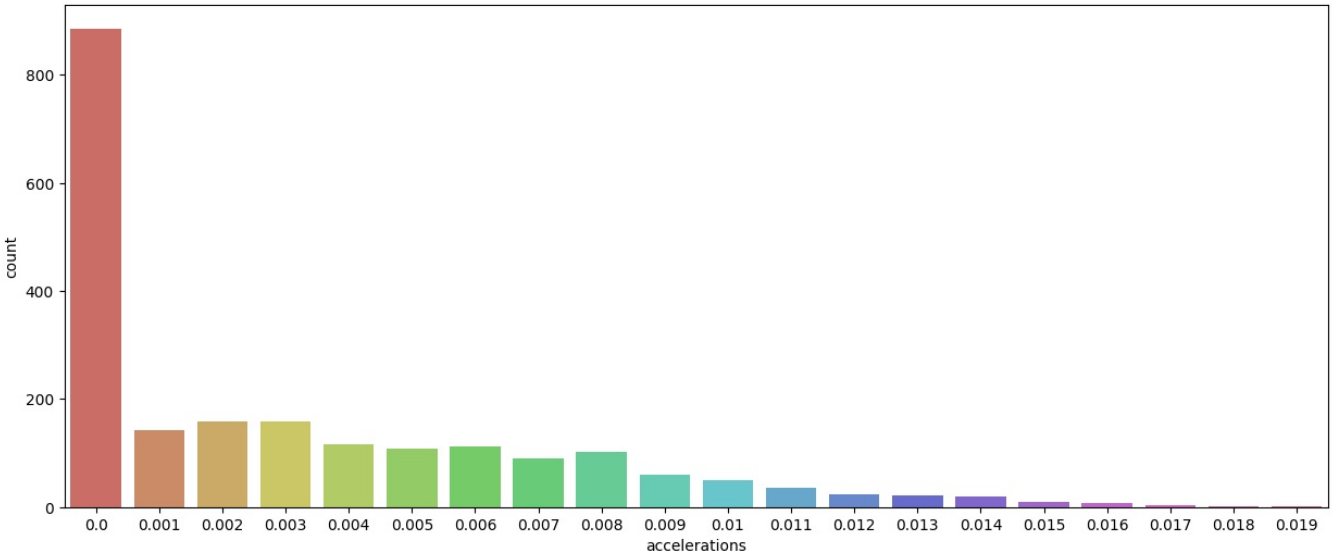
```
1     351
2     329
3     267
4     257
5     210
6     158
7     143
0     107
8     106
9      67
10     49
11     28
12     22
13     10
14      5
16      2
15      1
18      1
Name: count, dtype: int64
histogram_number_of_zeroes
histogram_number_of_zeroes
0     1611
1      366
2      108
3       21
4        2
5        2
10       1
8        1
7        1
Name: count, dtype: int64
histogram_mode
histogram_mode
133    140
136     89
150     88
142     86
148     79
      ...
164      1
97       1
95       1
98       1
69       1
Name: count, Length: 88, dtype: int64
histogram_mean
histogram_mean
143    65
144    63
135    63
141    61
140    60
       ..
178     1
169     1
165     1
168     1
73      1
Name: count, Length: 103, dtype: int64
histogram_median
histogram_median
137    68
146    68
142    68
145    67
147    64
       ..
178     1
186     1
183     1
180     1
104     1
Name: count, Length: 95, dtype: int64
histogram_variance
histogram_variance
1      246
0      186
2      163
3      158
4      106
      ...
127      1
126      1
134      1
144      1
147      1
Name: count, Length: 133, dtype: int64
histogram_tendency
histogram_tendency
```

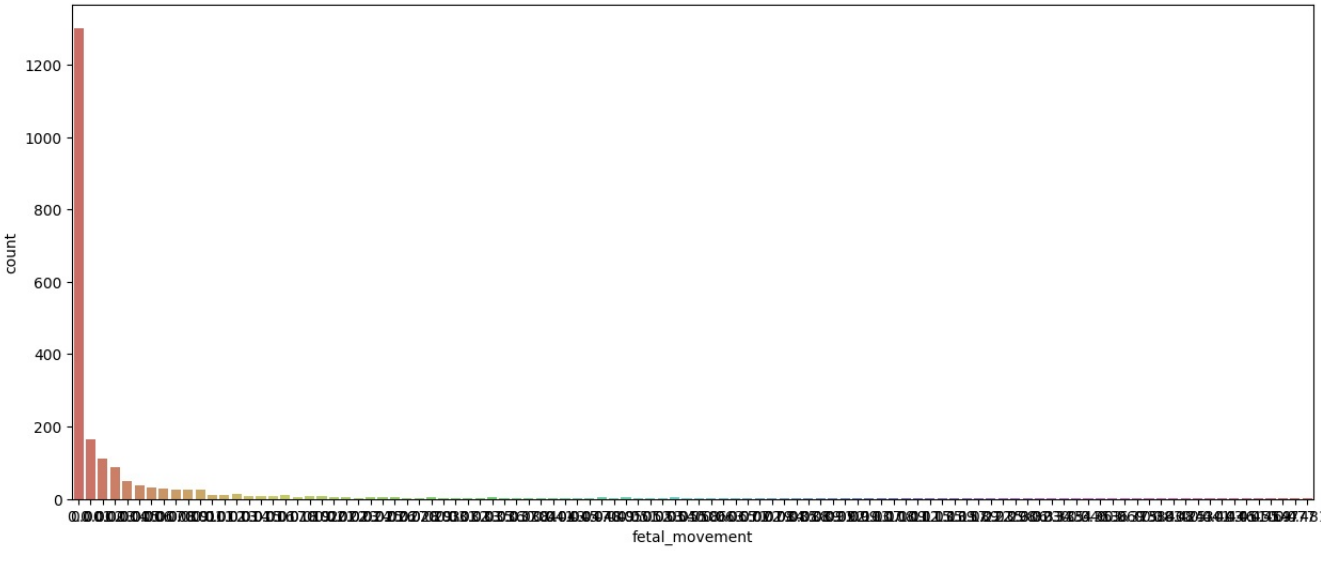
```
 0    1110
 1     838
-1     165
Name: count, dtype: int64
fetal_health
fetal_health
1    1646
2     292
3     175
Name: count, dtype: int64
```

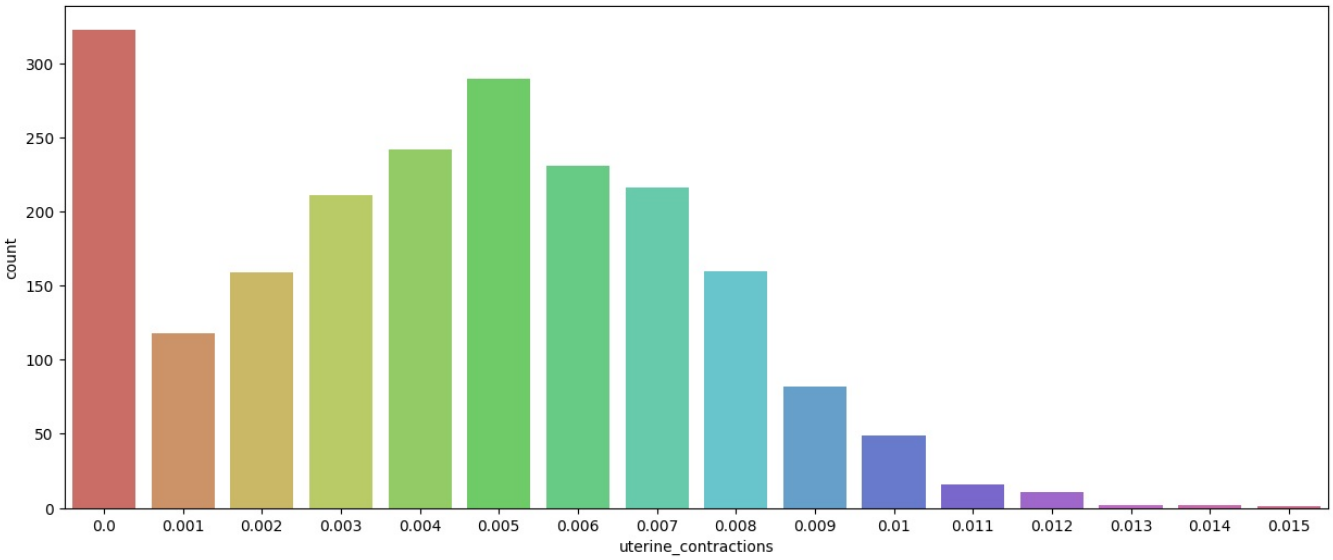
In [15]:
```
import warnings
warnings.filterwarnings('ignore')
```

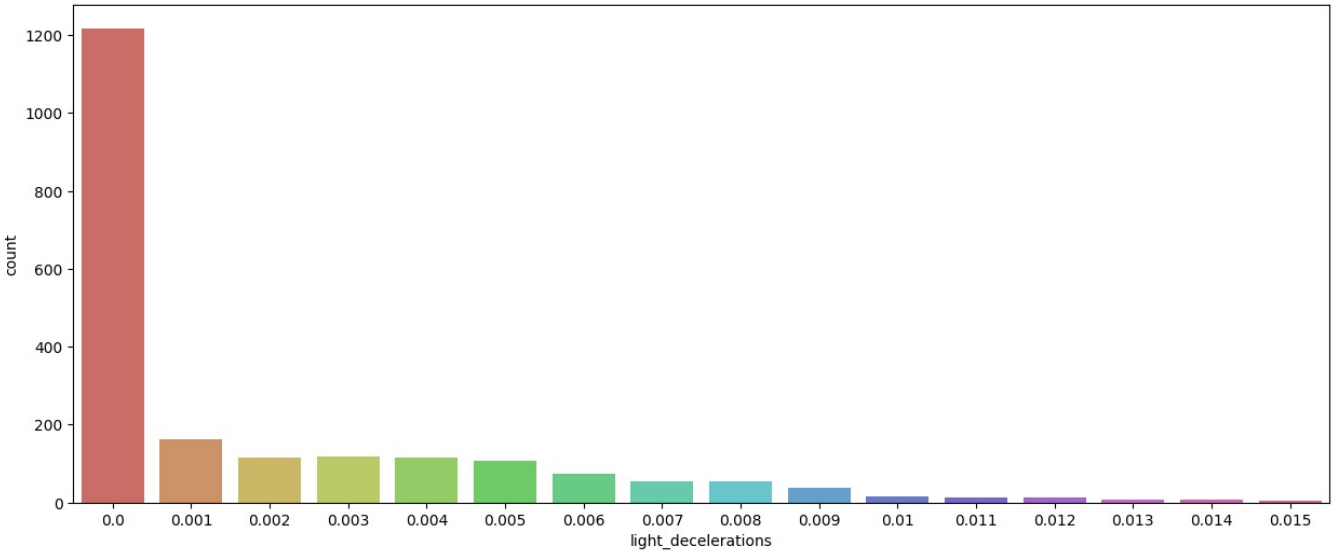
In [16]:
```
for i in df.columns:
    plt.figure(figsize=(15,6))
    sns.countplot(data=df,  x=df[i], palette = 'hls')
    plt.show()
```

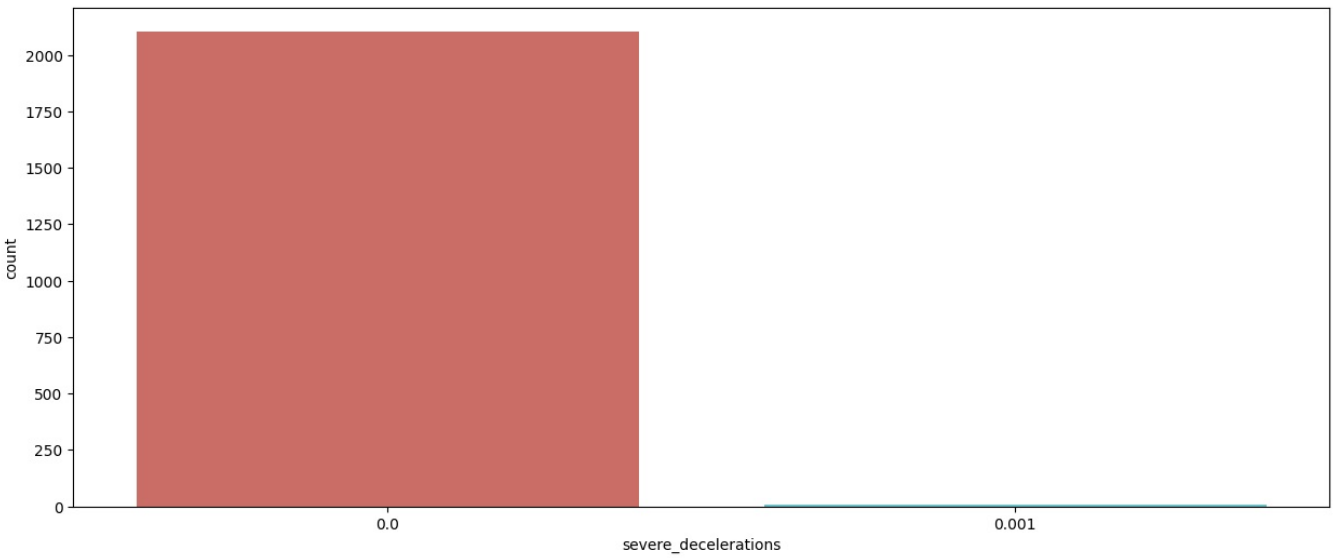


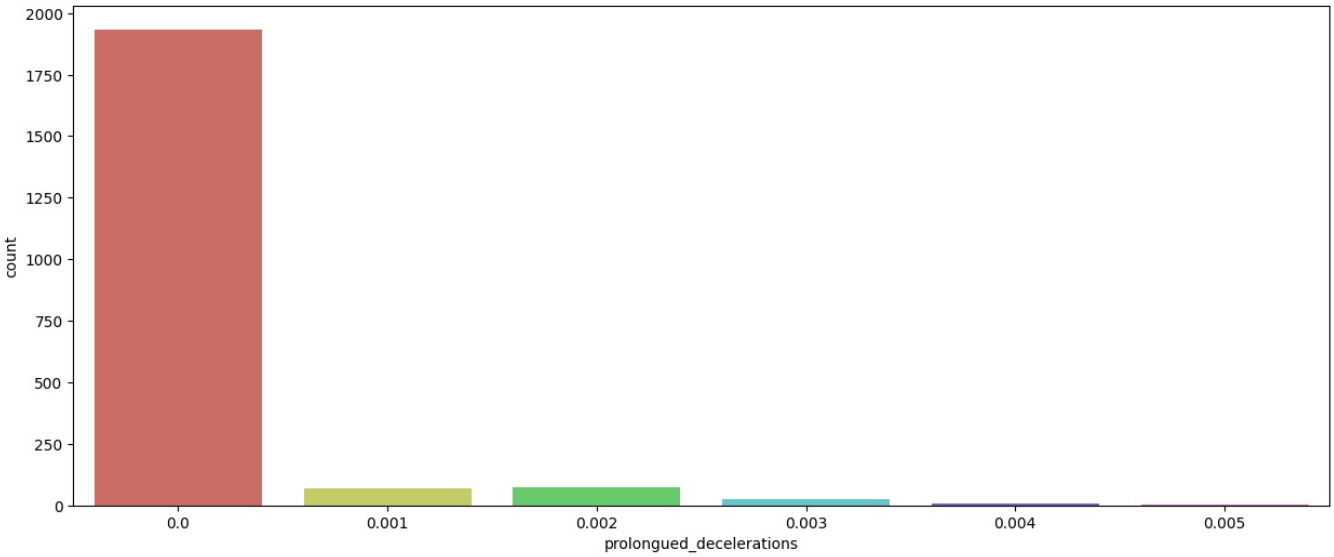


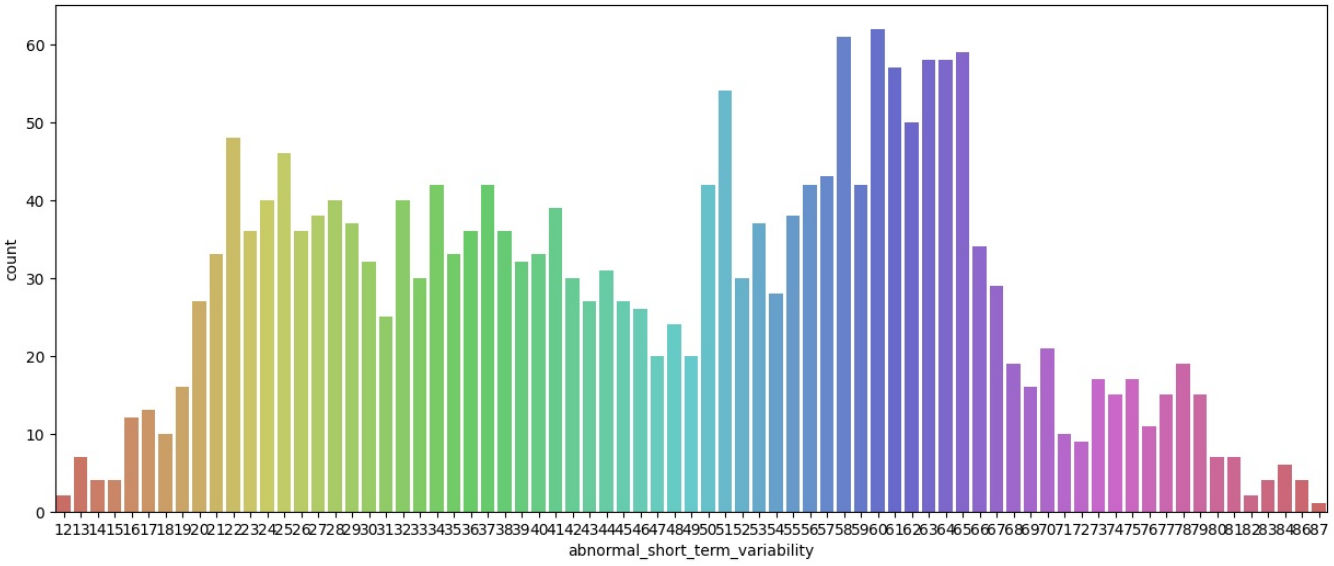


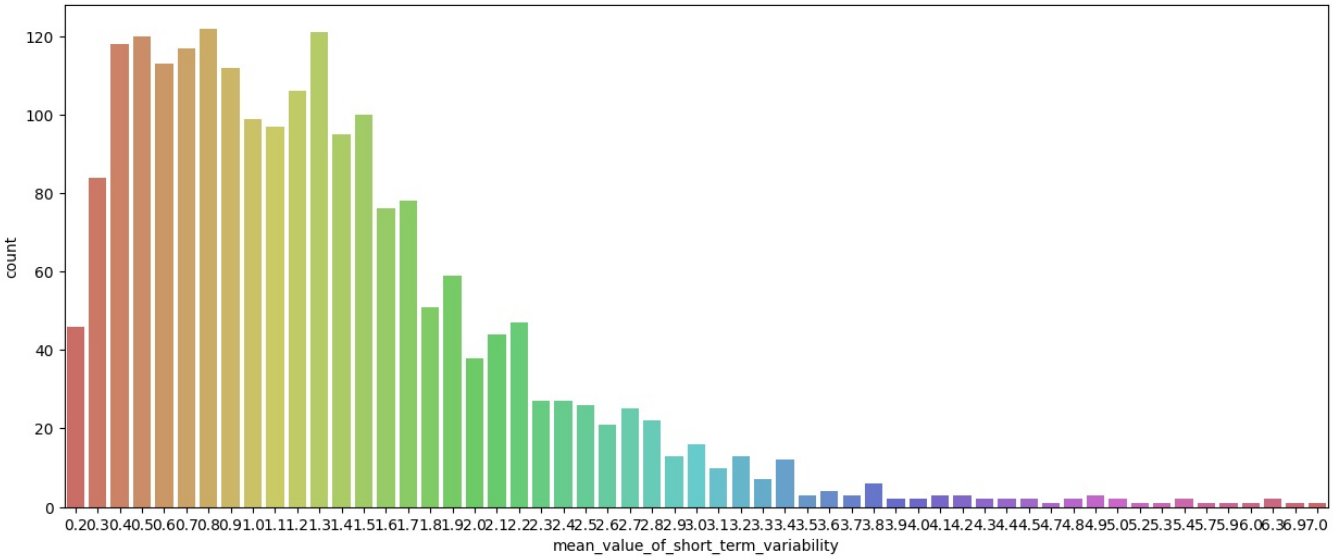

count
0
50
100
150
200
250
300
0.0
0.001
0.002
0.003
0.004
0.005
0.006
0.007
0.008
0.009
0.01
0.011
0.012
0.013
0.014
0.015
uterine_contractions

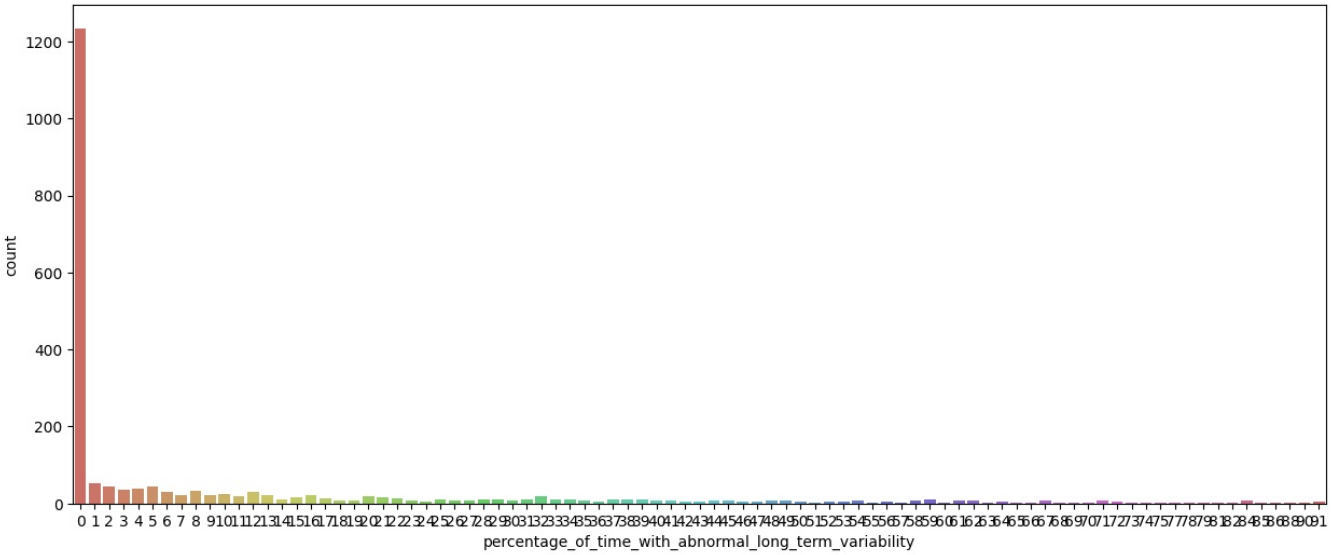

count
0
200
400
600
800
1000
1200
0.0
0.001
0.002
0.003
0.004
0.005
0.006
0.007
0.008
0.009
0.01
0.011
0.012
0.013
0.014
0.015
light_decelerations

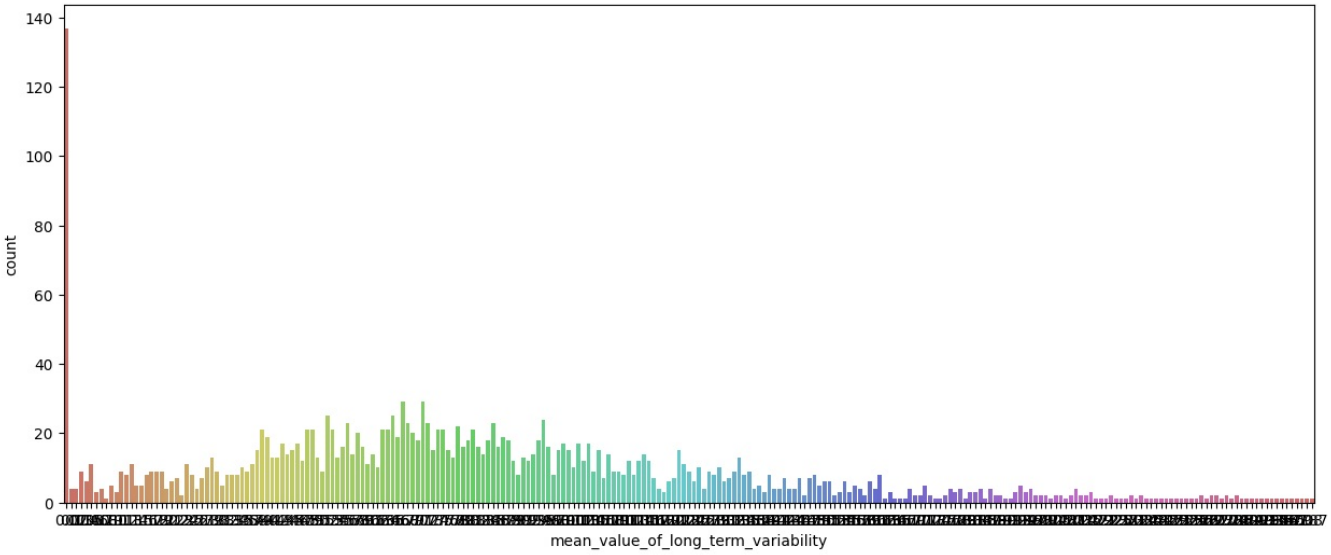

count
0
250
500
750
1000
1250
1500
1750
2000
0.0
0.001
severe_decelerations

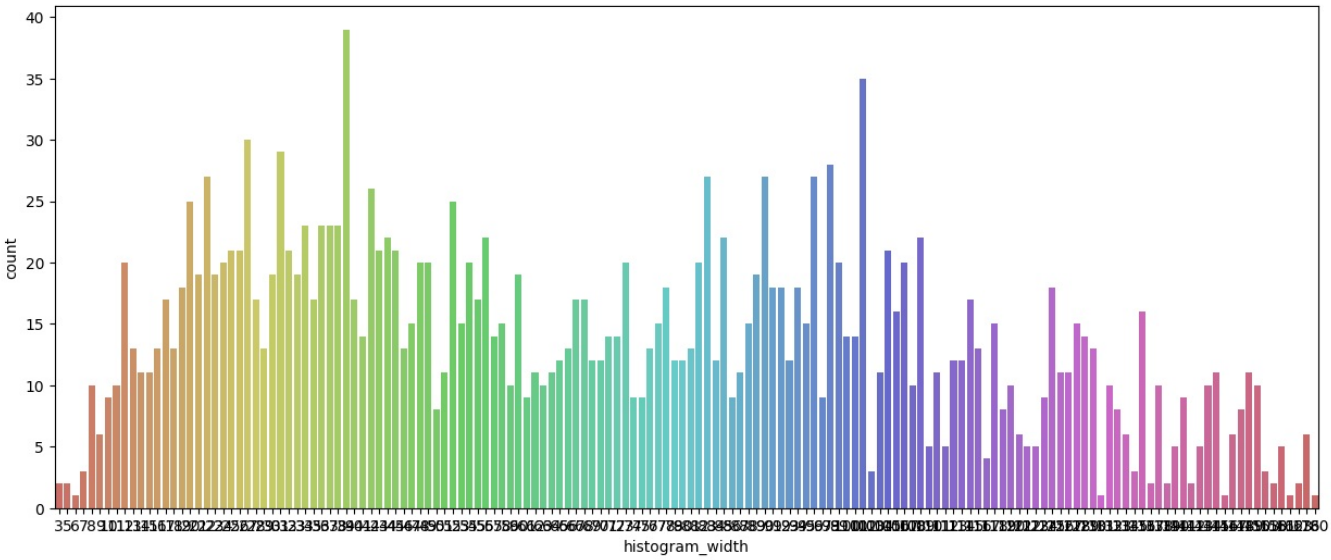

2000
1750
1500
1250
1000
750
500
250
0
count
0.0
0.001
0.002
0.003
0.004
0.005
prolongued_decelerations

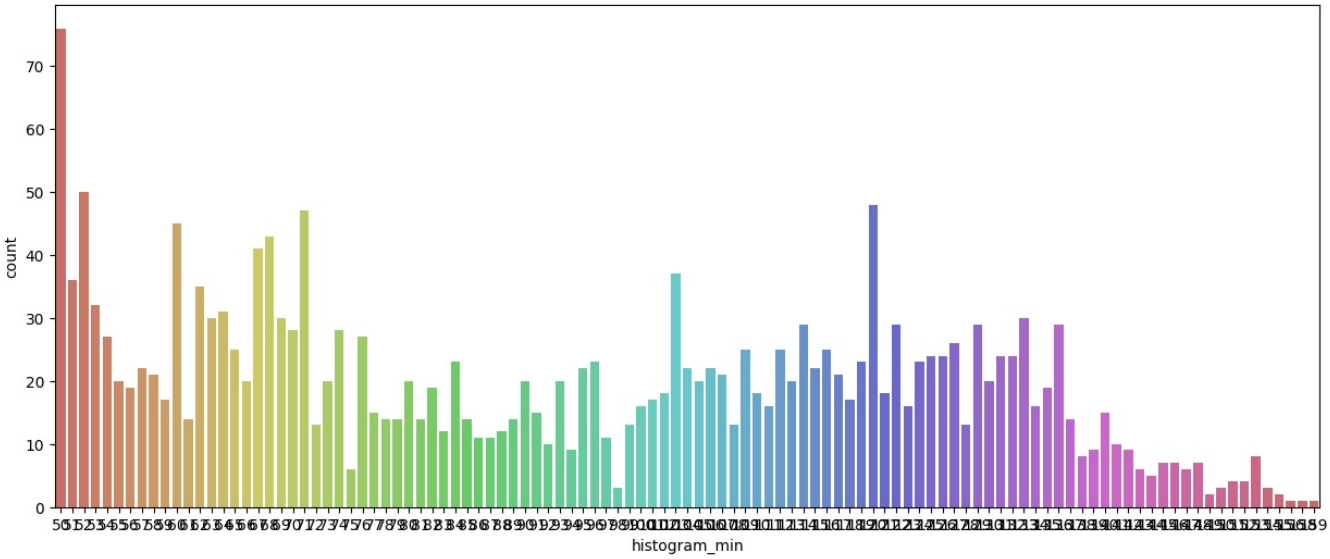

60
50
40
30
20
10
0
count
abnormal_short_term_variability

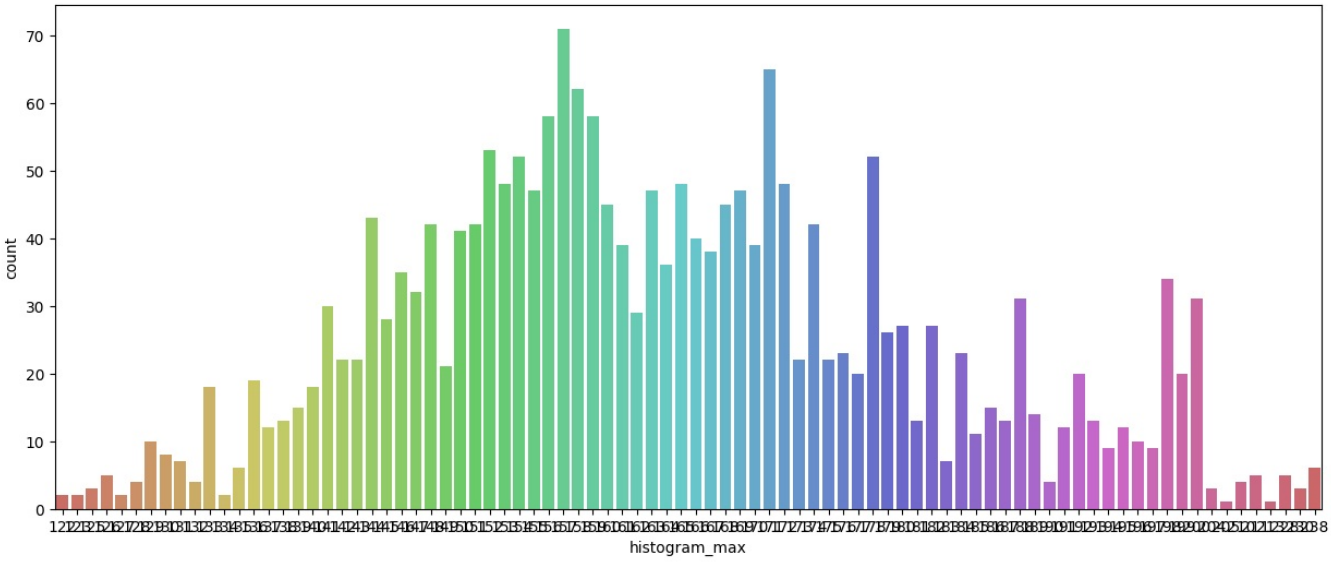

count
mean_value_of_short_term_variability

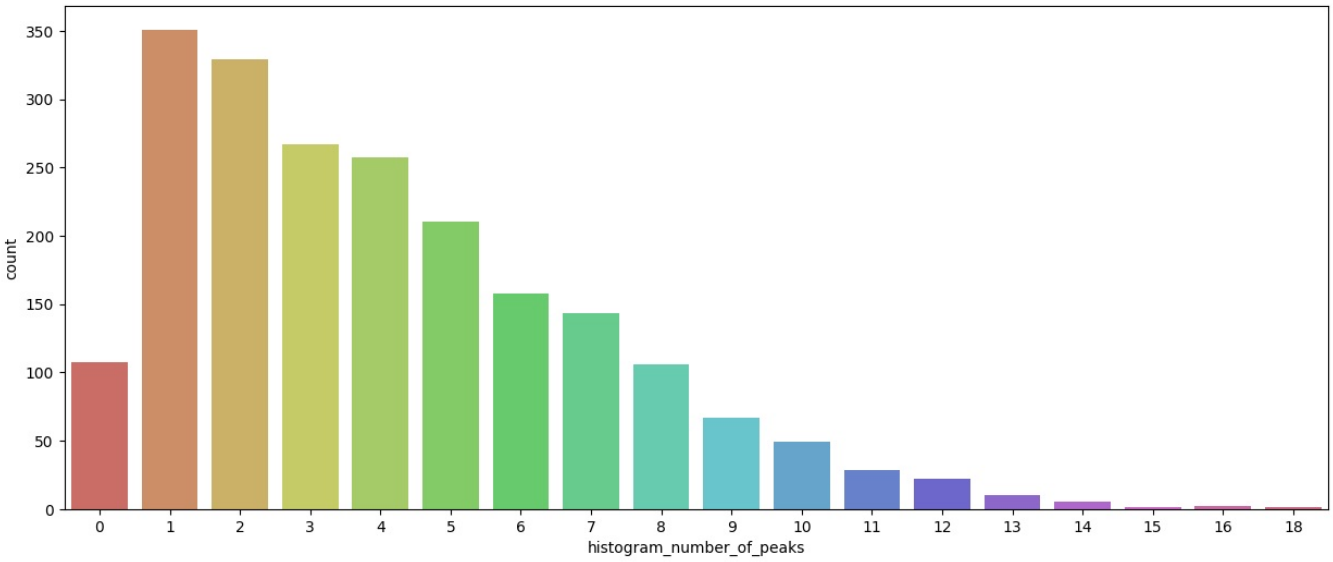

count
percentage_of_time_with_abnormal_long_term_variability

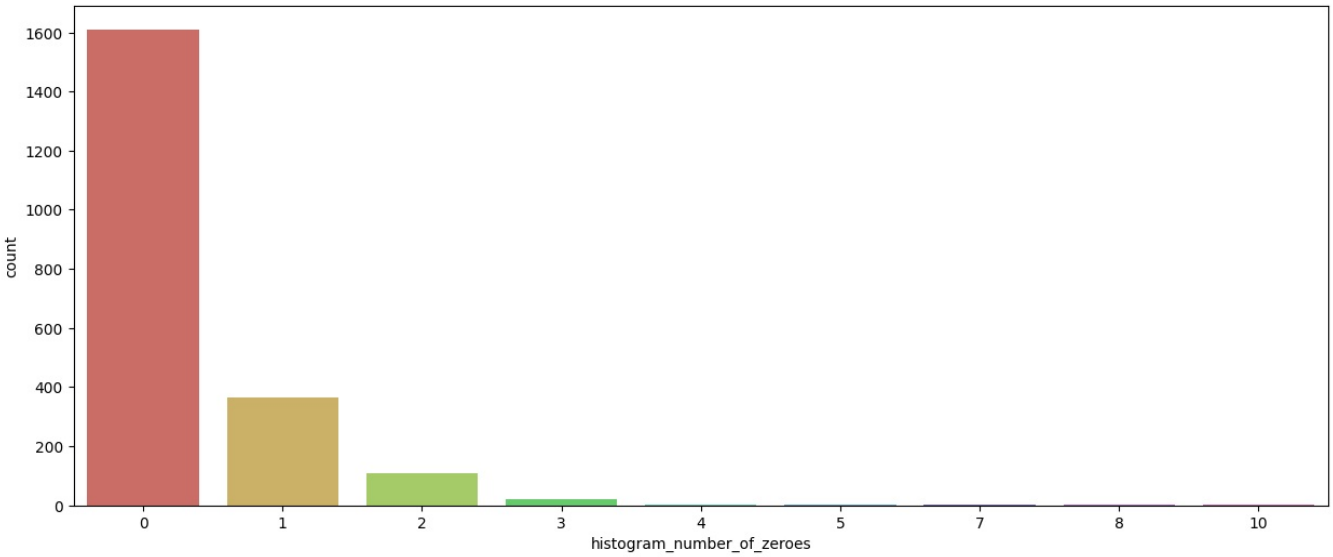

count
mean_value_of_long_term_variability

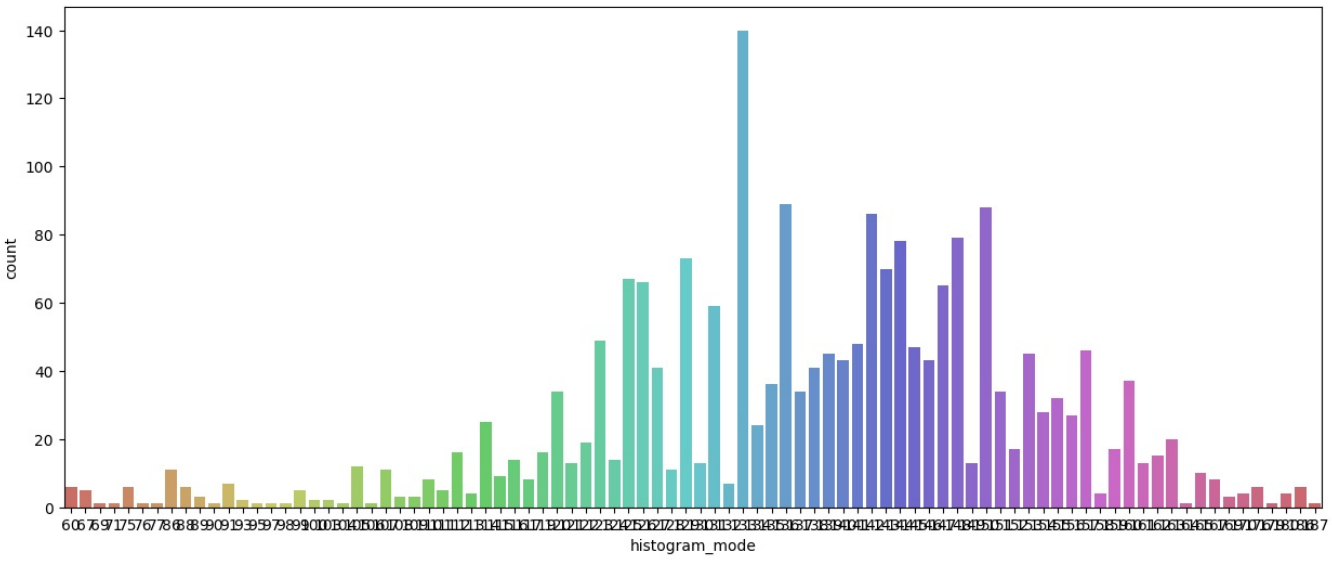

count
histogram_width

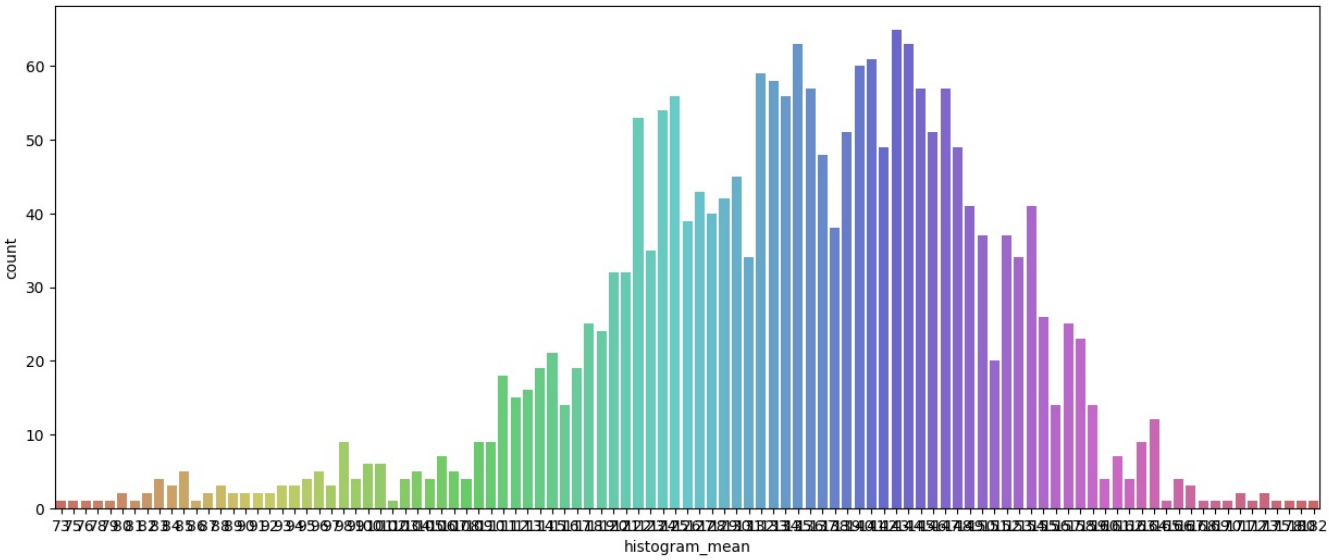

count
histogram_min

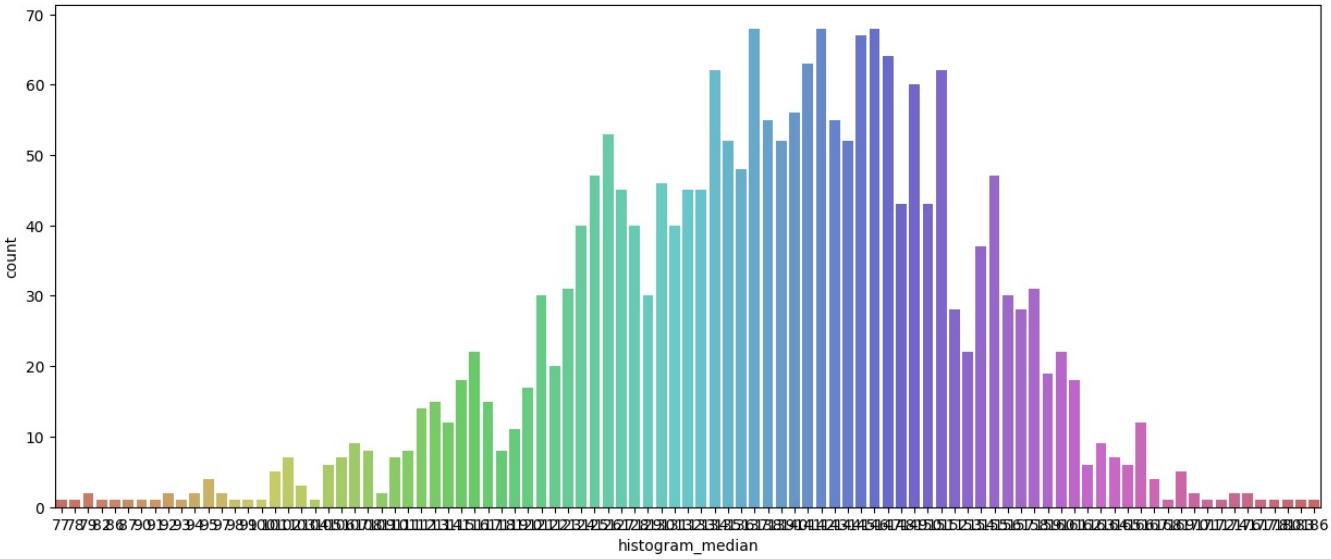

count
histogram_max

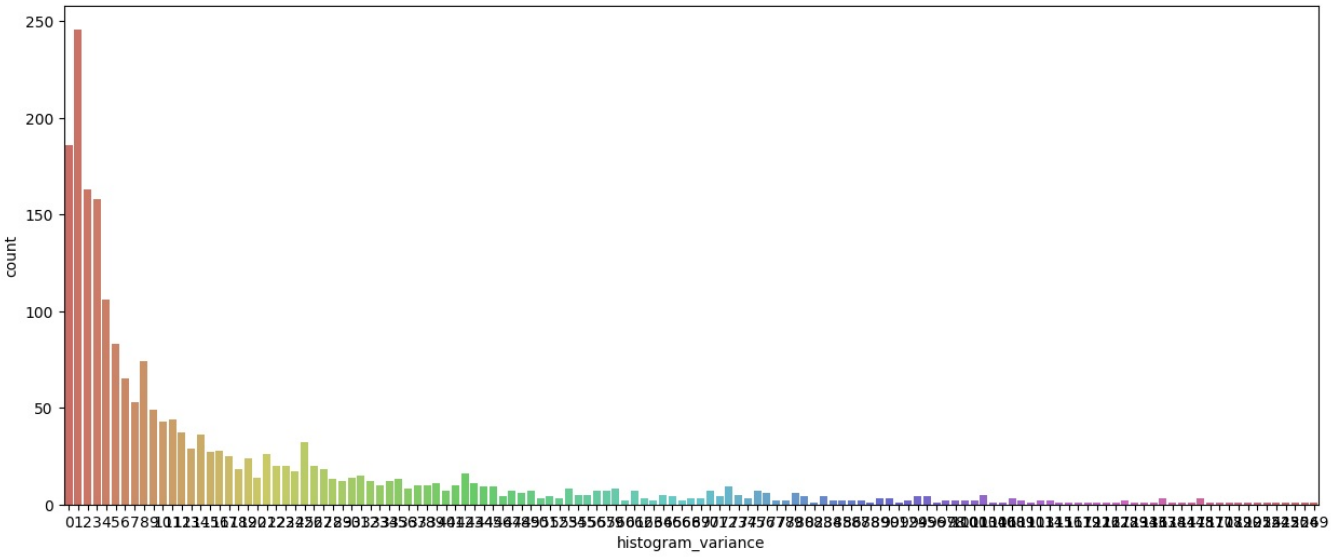

count
histogram_number_of_peaks

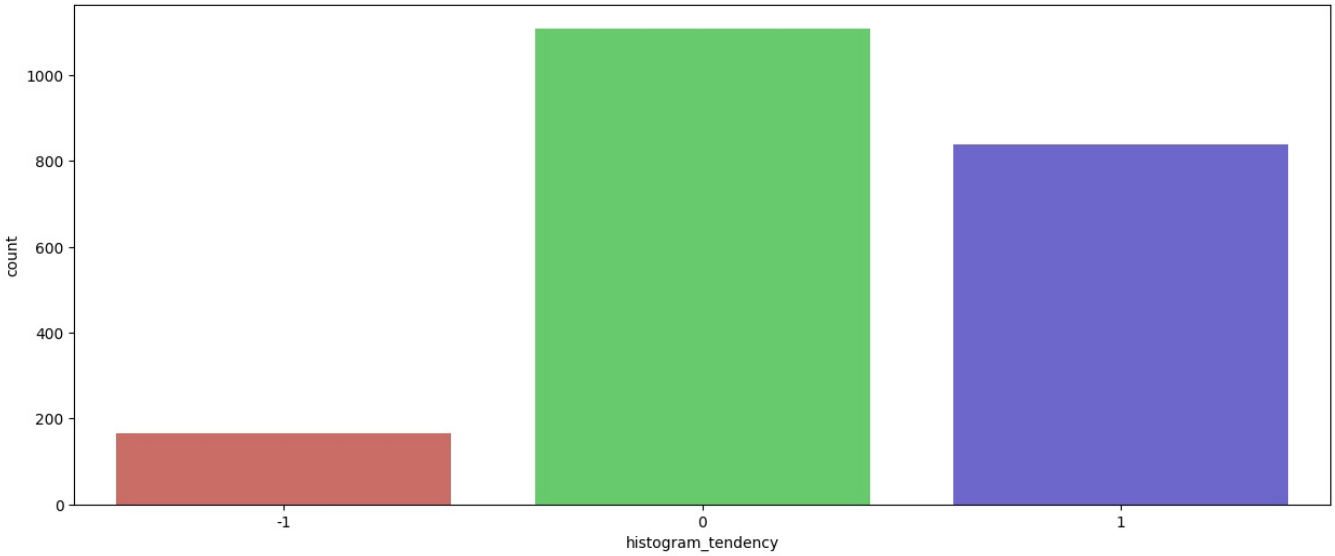

count
histogram_number_of_zeroes

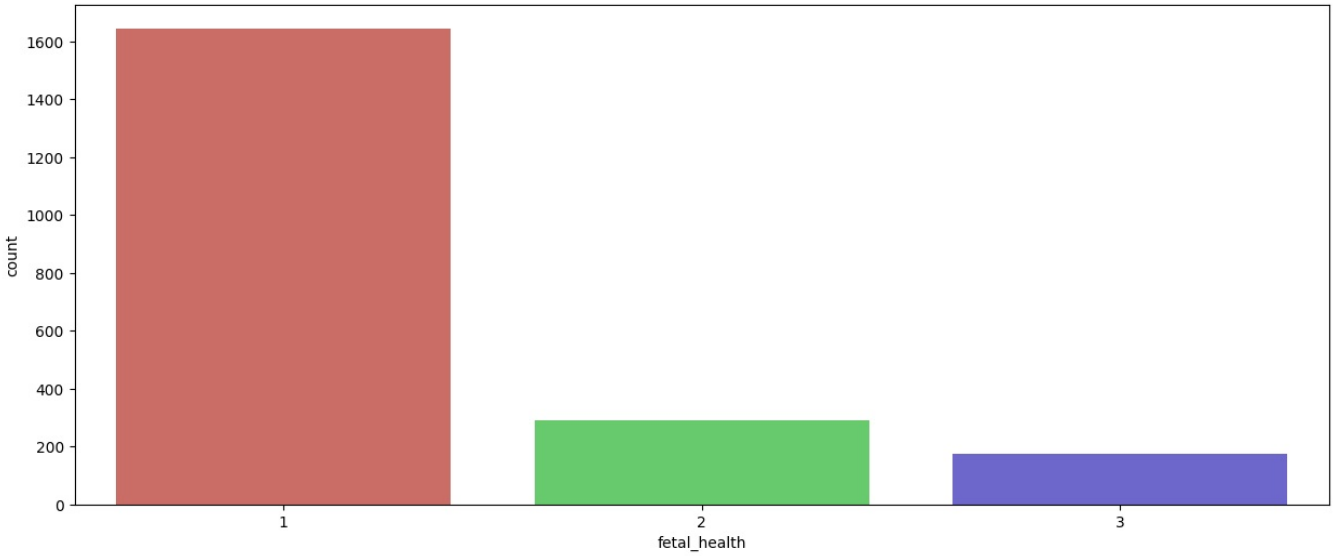

count
histogram_mode


count
histogram_mean

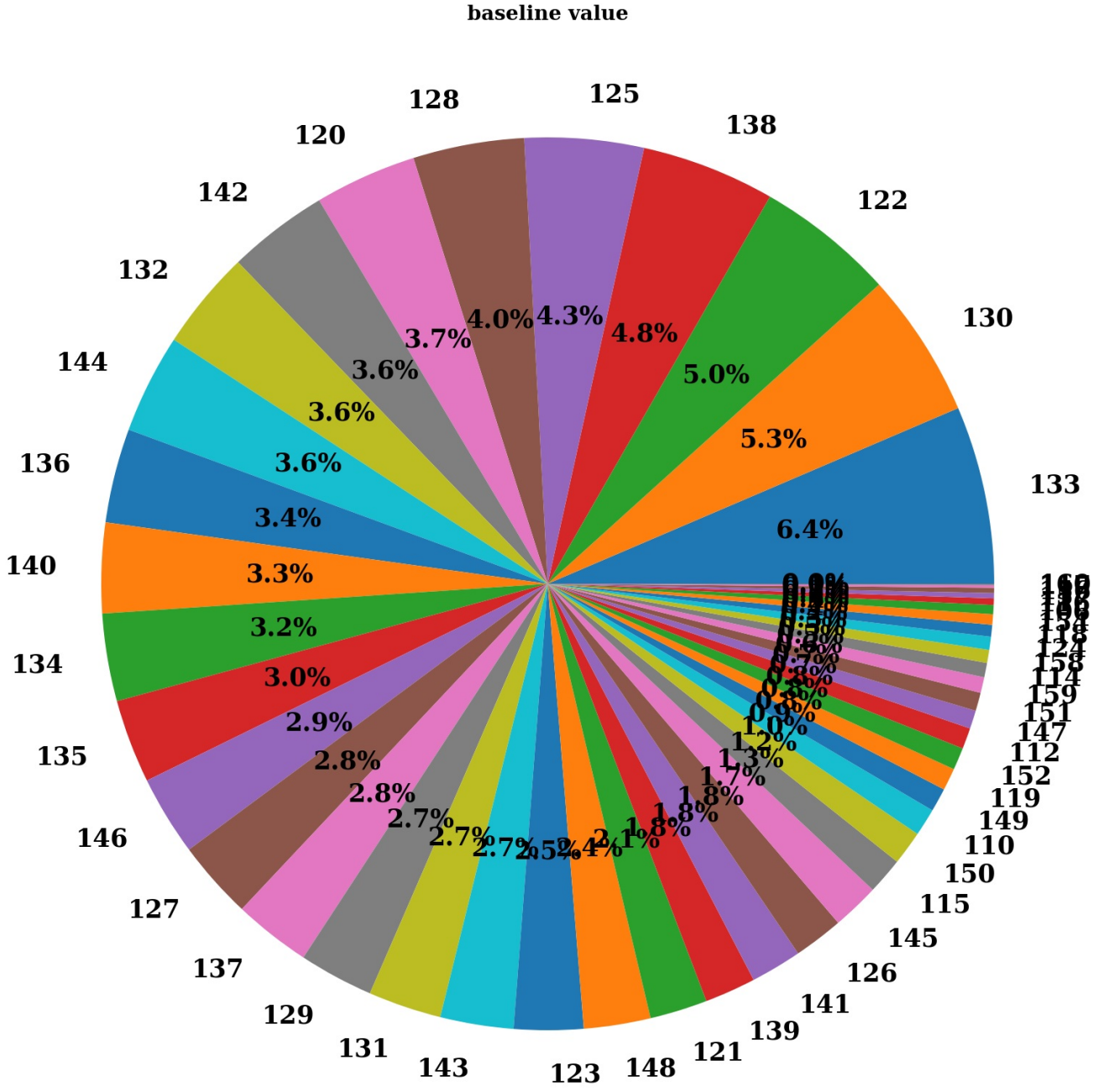

count
histogram_median

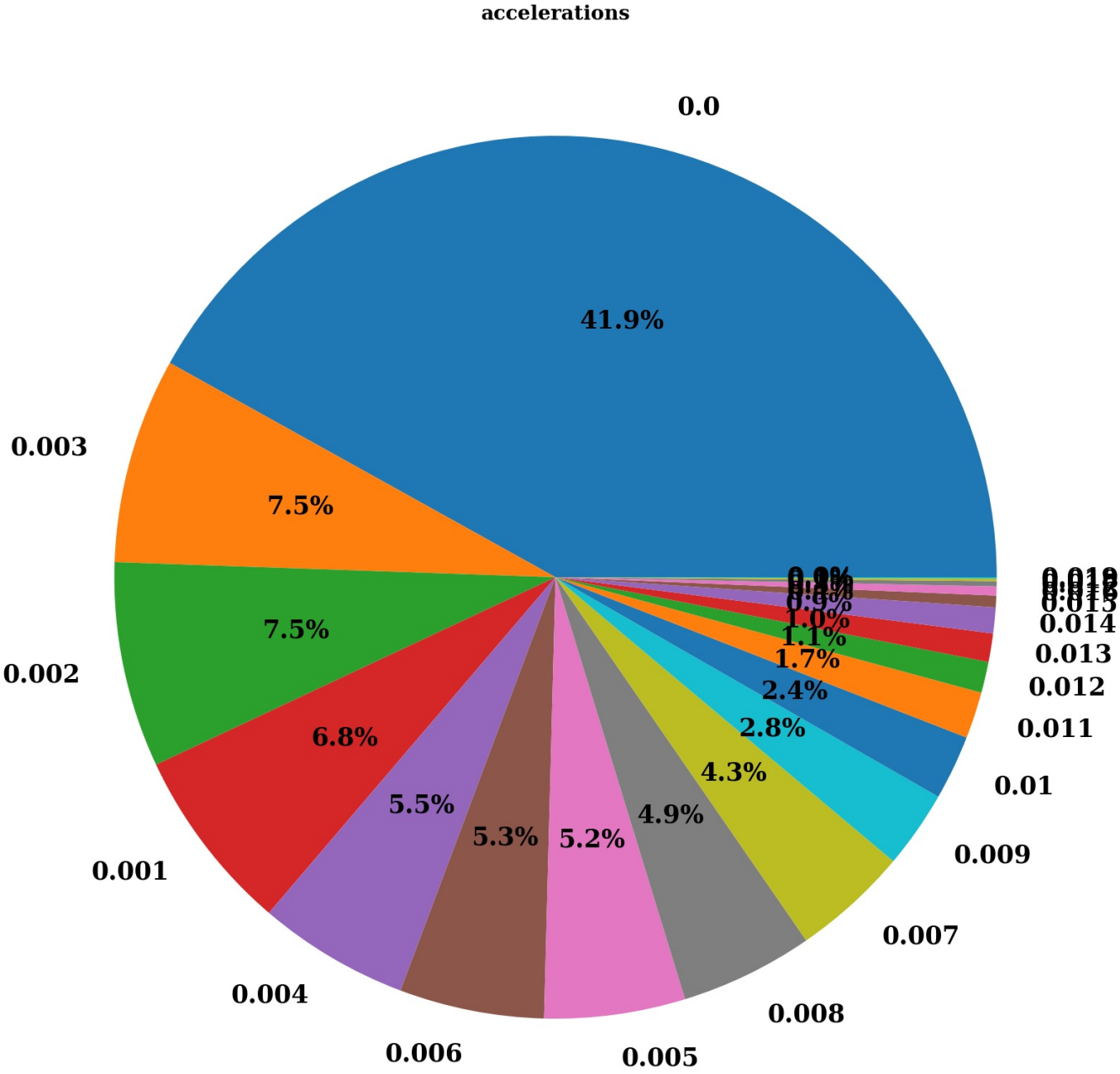

count
histogram_variance

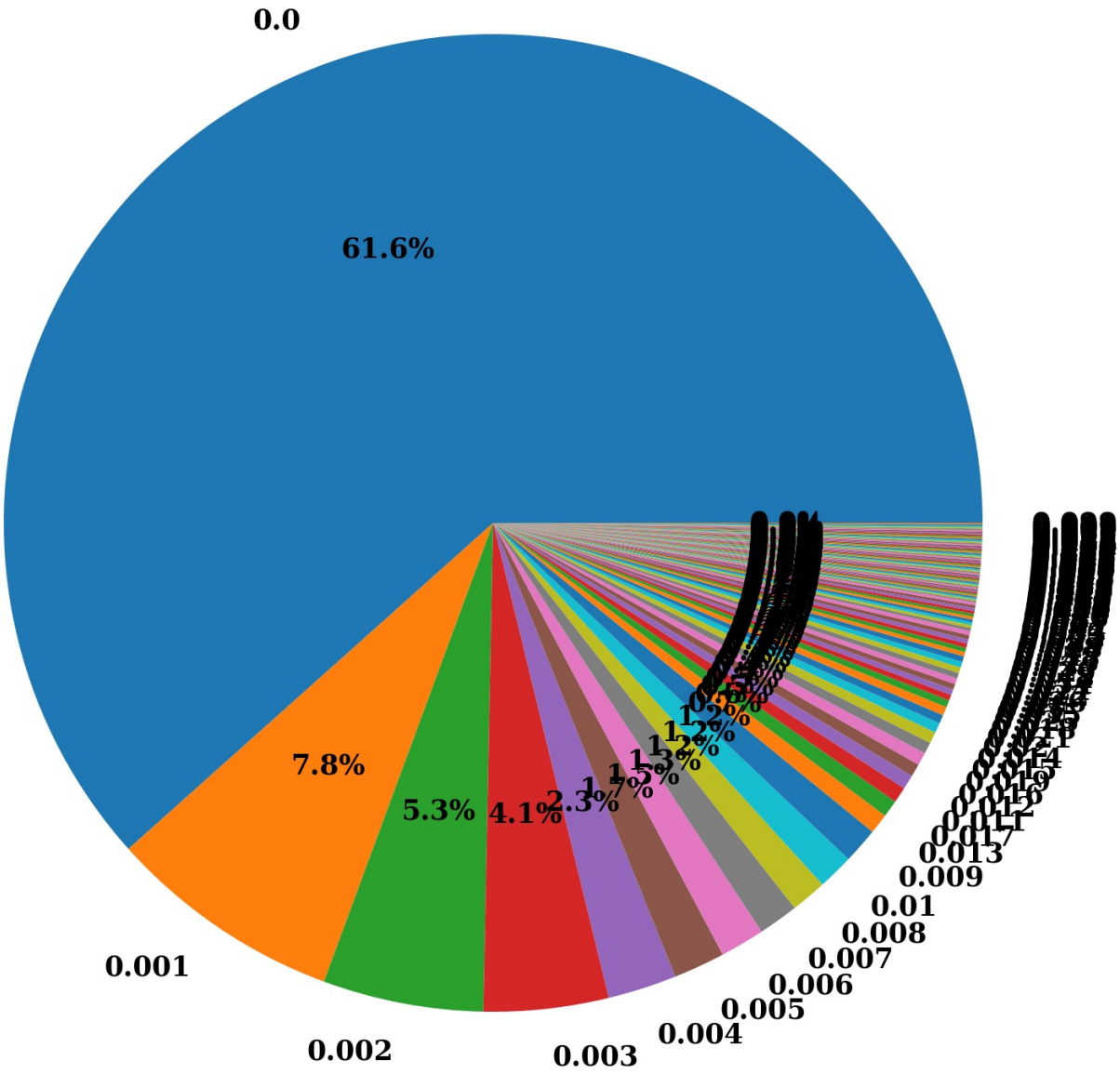


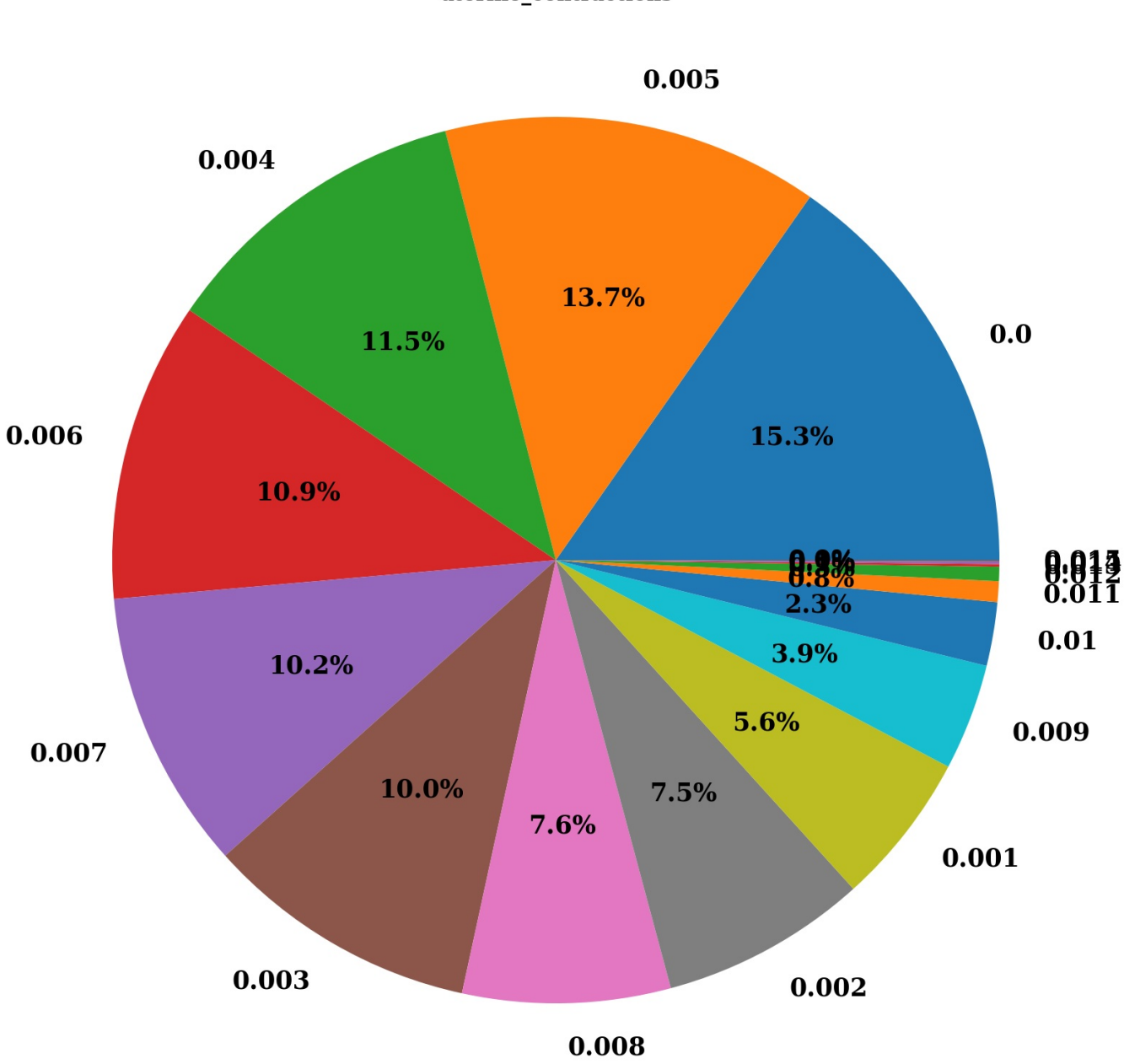


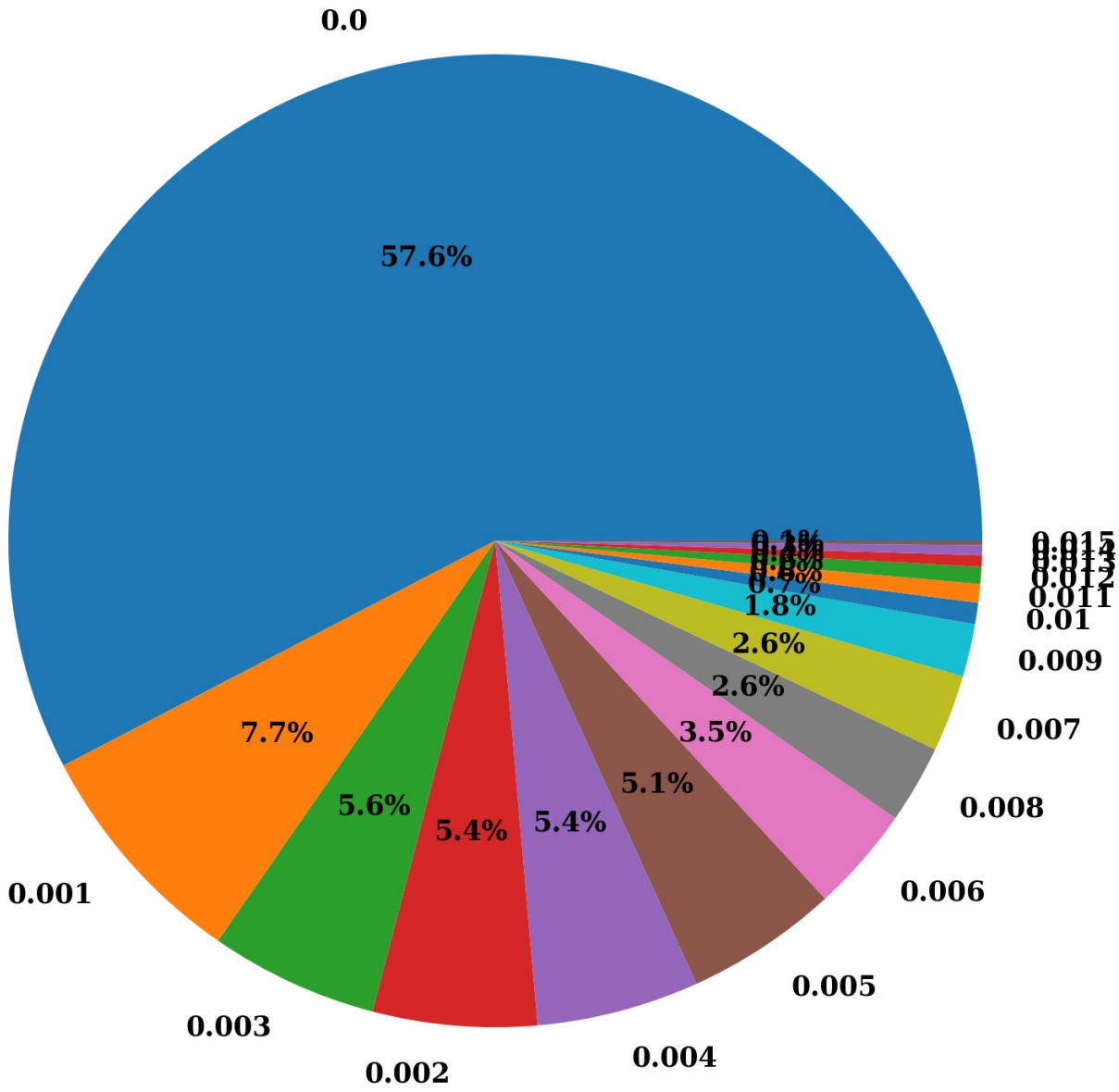
In [17]:
```
for i in df.columns:
    plt.figure(figsize=(30,20))
    plt.pie(df[i].value_counts(), labels=df[i].value_counts().index, autopct='%1.1f%%', textprops={ 'fontsize':
                                        'color': 'black',
                                        'weight': 'bold',
                                        'family': 'serif' })
    hfont = {'fontname':'serif', 'weight': 'bold'}
    plt.title(i, size=20, **hfont)
    plt.show()
```

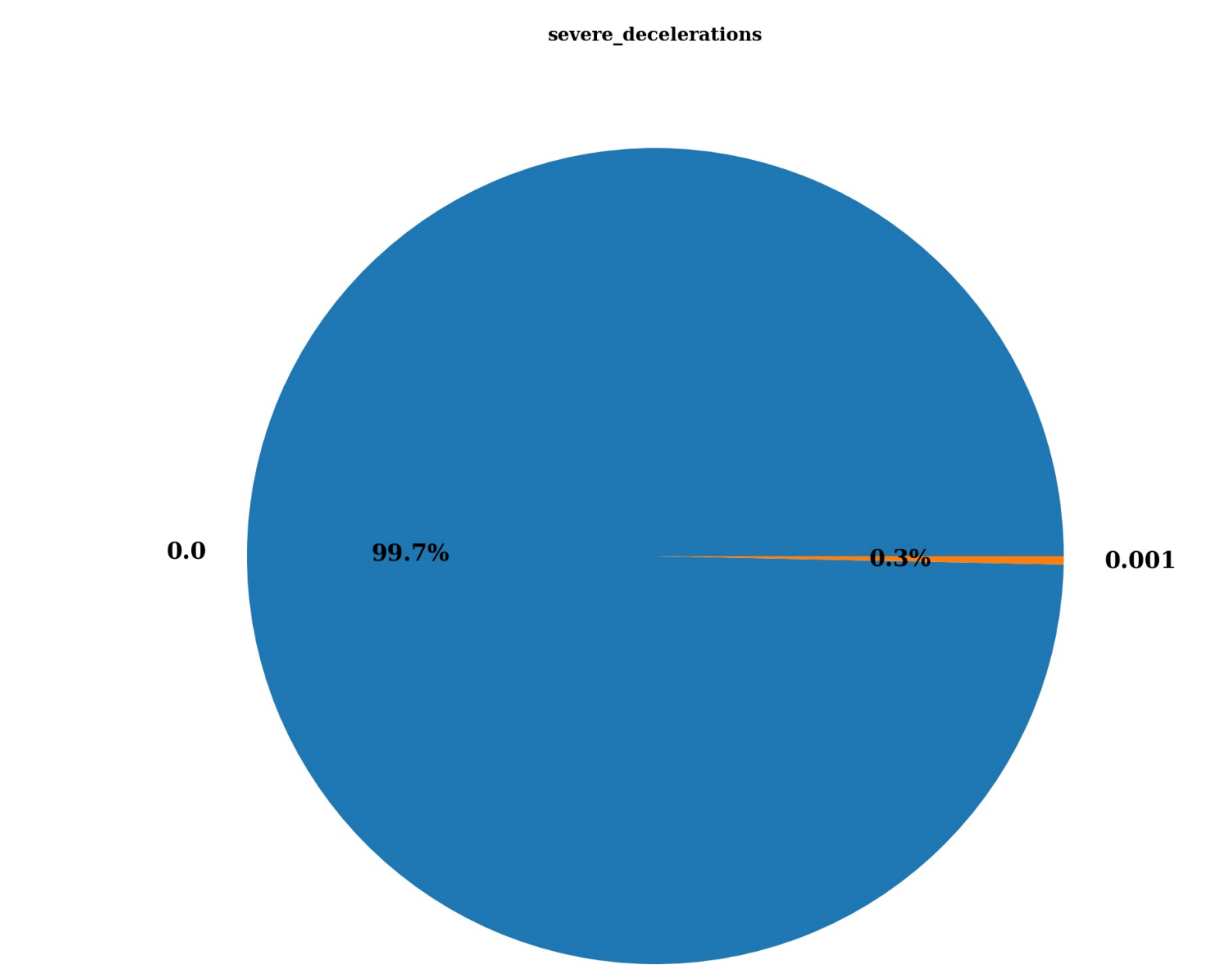

baseline value
128
125
120
138
142
122
132
130
144
136
133
140
134
135
146
127
137
129
131
143
123
148
121
139
141
126
145
115
150
110
149
119
152
112
147
151
159
114
158
124
118
6.4%
5.3%
5.0%
4.8%
4.3%
4.0%
3.7%
3.6%
3.6%
3.6%
3.4%
3.3%
3.2%
3.0%
2.9%
2.8%
2.8%
2.7%
2.7%
2.7%
2.4%
2.1%
1.8%
1.8%
1.8%
1.7%
1.3%
1.2%
1.0%

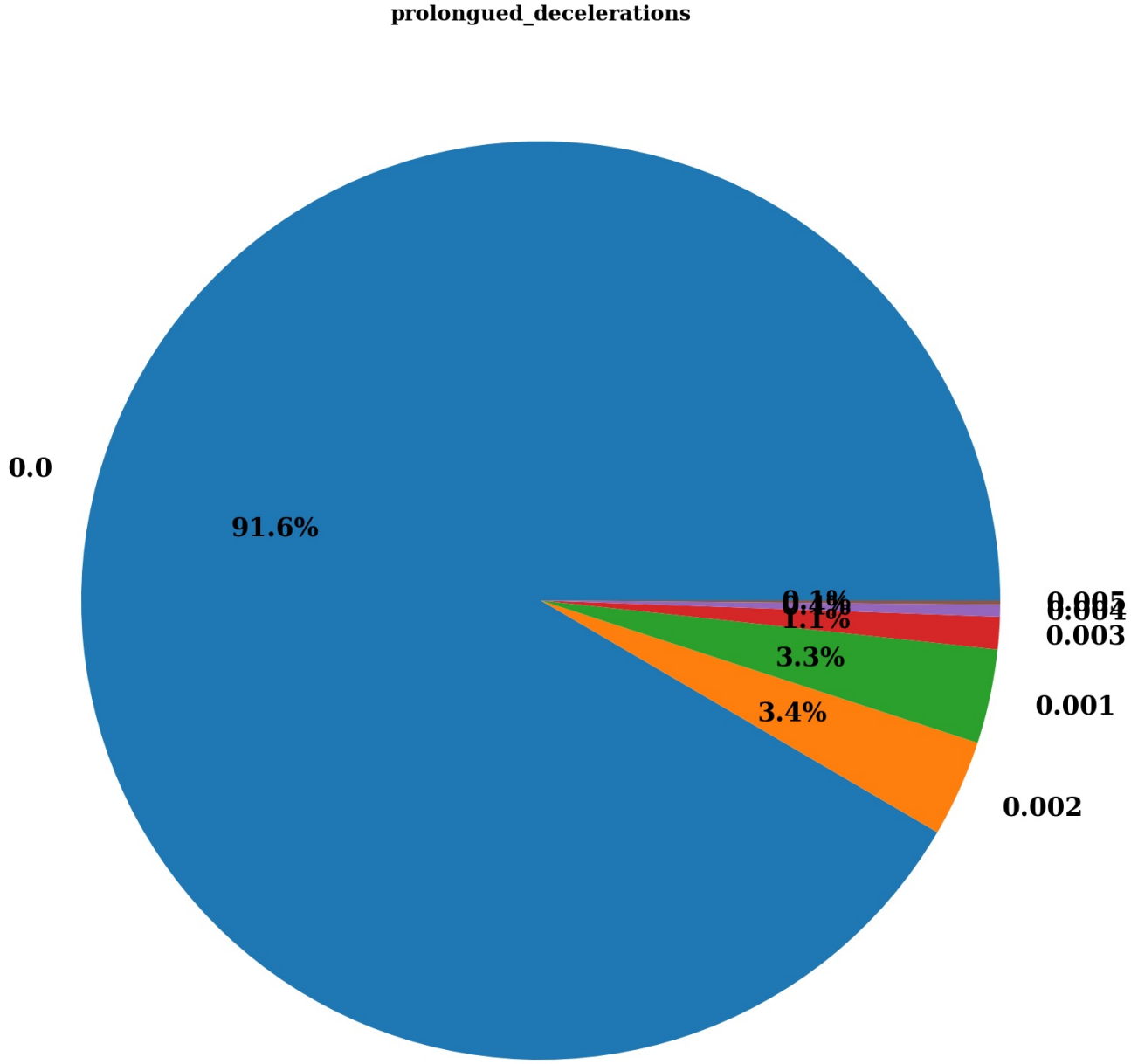

accelerations
0.0
41.9%
0.003
7.5%
7.5%
0.002
6.8%
0.001
5.5%
5.3%
5.2%
4.9%
4.3%
2.8%
2.4%
1.7%
1.1%
1.0%
0.9%
0.004
0.006
0.005
0.008
0.007
0.009
0.01
0.011
0.012
0.013
0.014
0.015

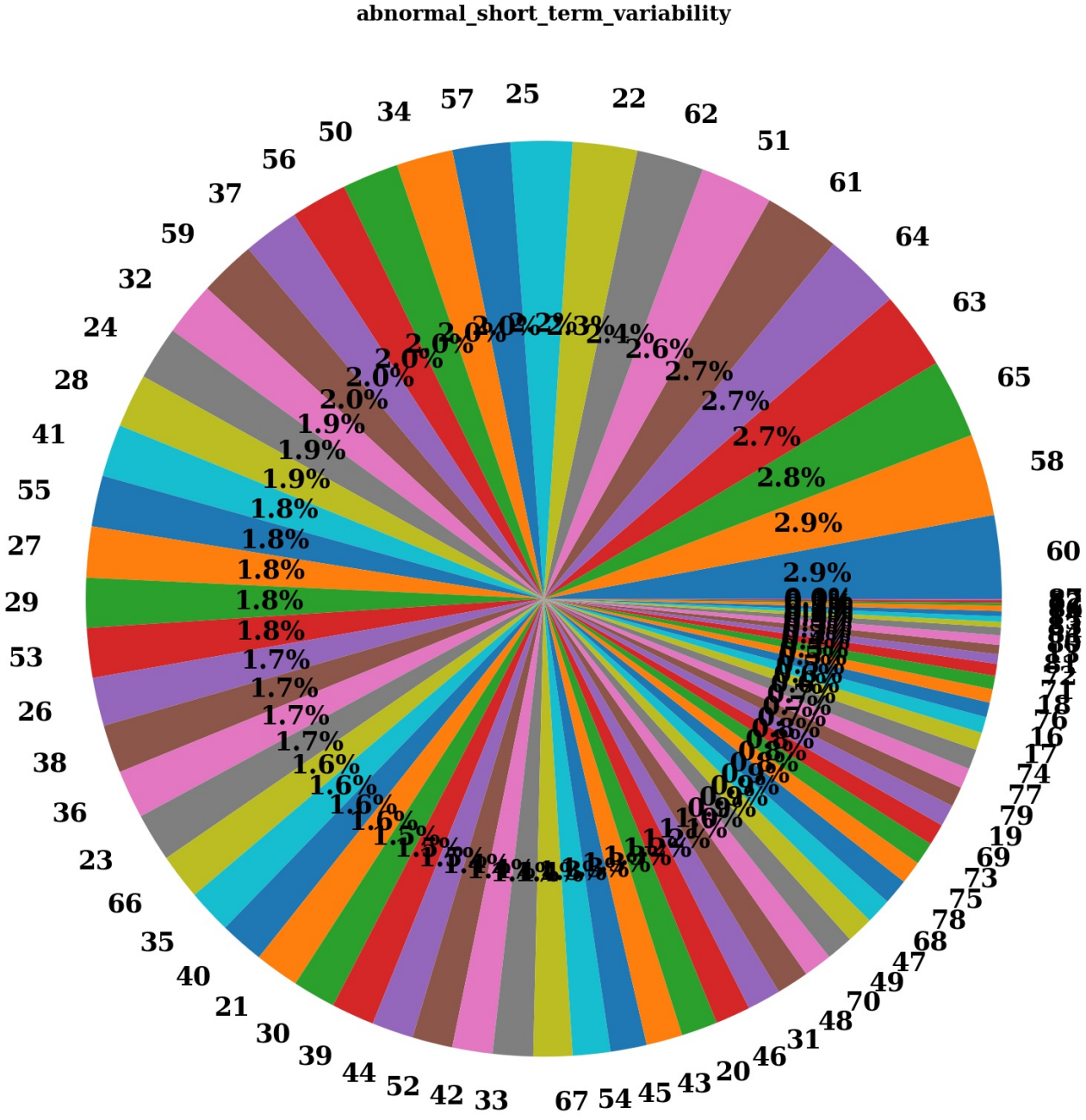

fetal_movement
0.0
61.6%
0.001
7.8%
0.002
5.3%
0.003
4.1%
0.004
2.3%
0.005
0.006
0.007
0.008
0.01
0.009
0.013

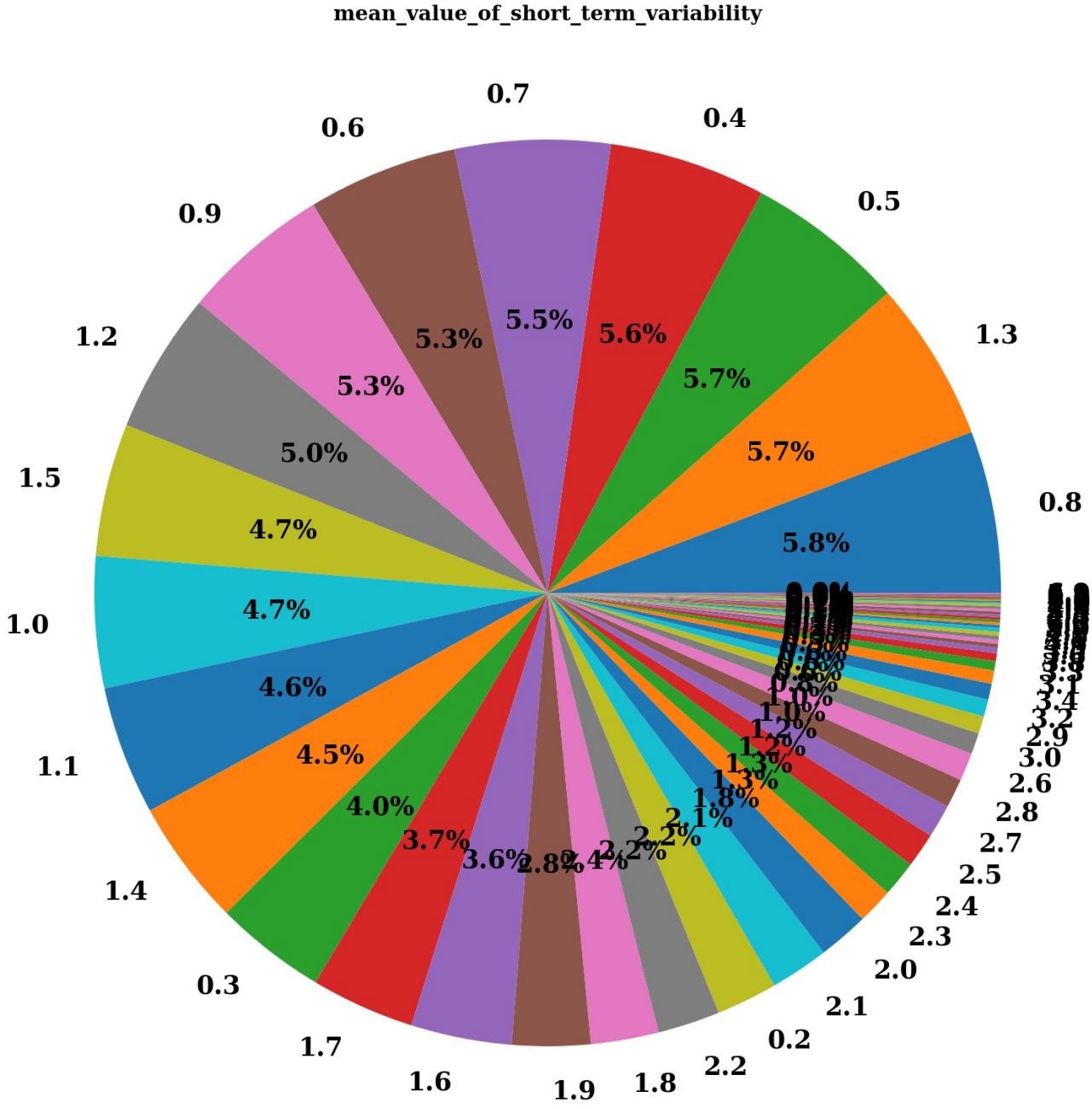

uterine_contractions
0.005
0.004
0.0
0.006
0.007
0.009
0.001
0.003
0.002
0.008
0.01
0.011
13.7%
11.5%
15.3%
10.9%
10.2%
10.0%
7.6%
7.5%
5.6%
3.9%
2.3%

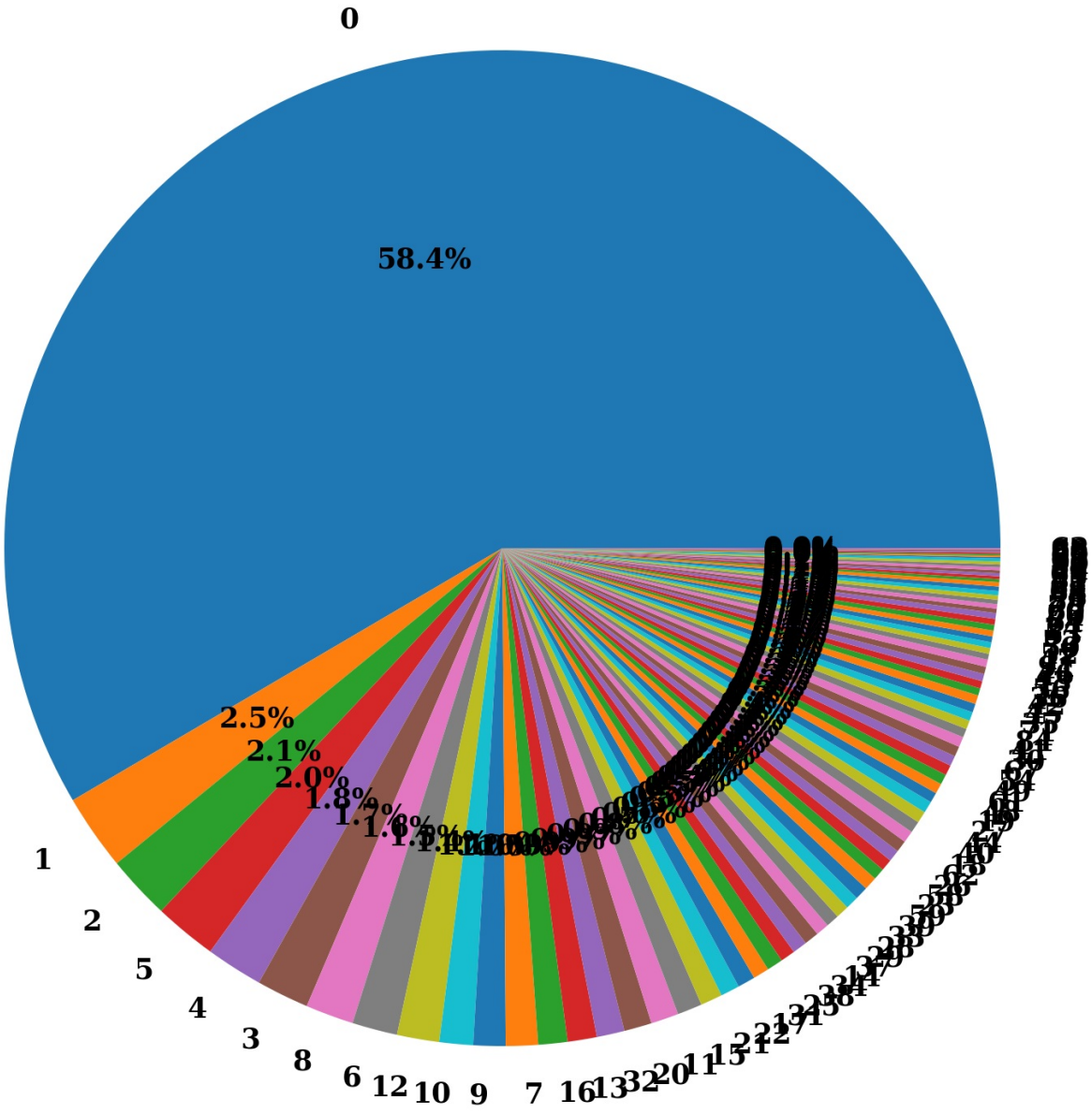

light_decelerations
0.0
57.6%
0.001
7.7%
0.003
5.6%
0.002
5.4%
0.004
5.4%
0.005
5.1%
0.006
3.5%
0.008
2.6%
0.007
2.6%
0.009
1.8%
0.01
0.011
0.012
0.013
0.014
0.015


severe_decelerations
0.0
99.7%
0.3%
0.001

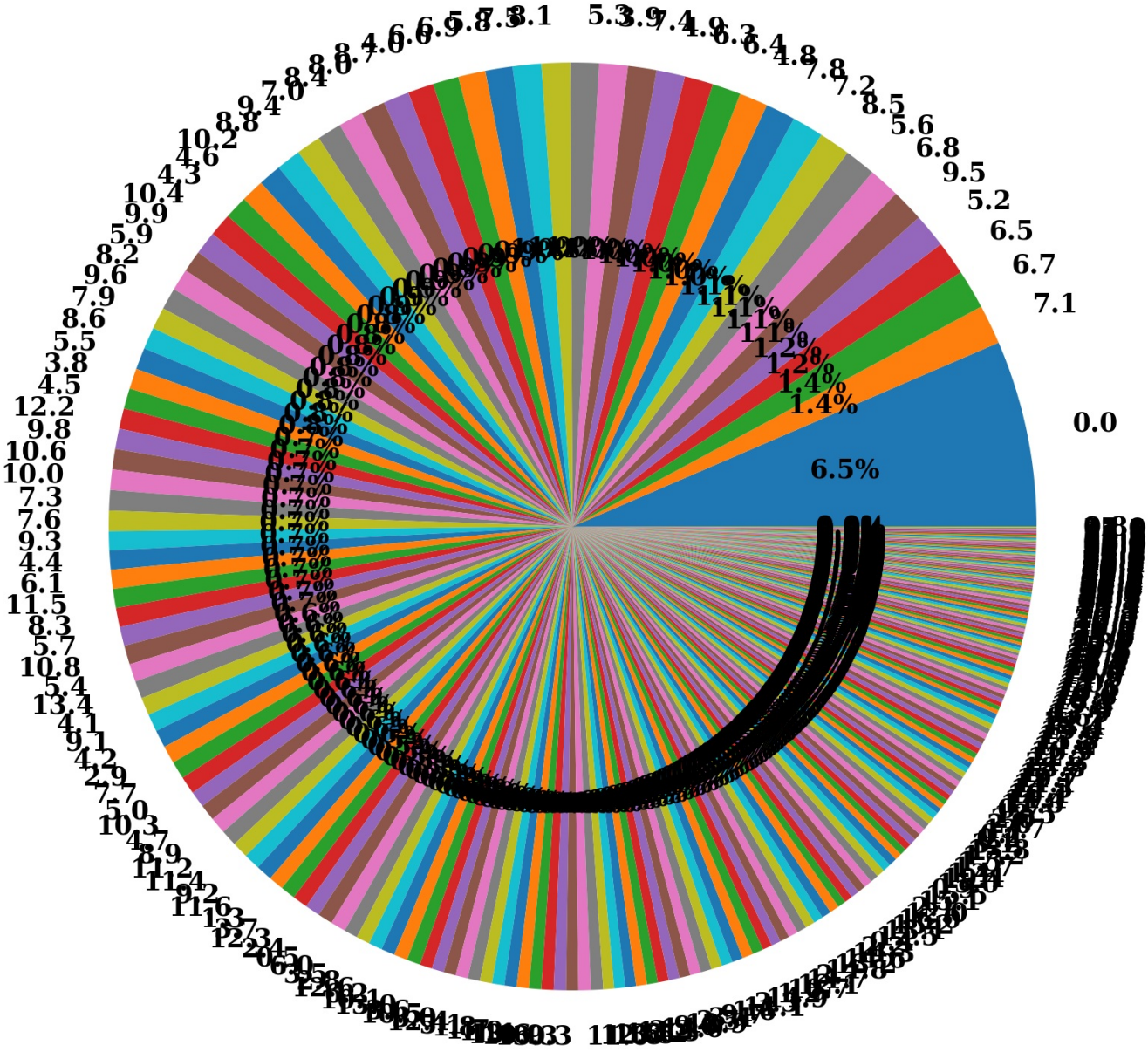

prolongued_decelerations
0.0
91.6%
0.1%
0.4%
1.1%
3.3%
3.4%
0.005
0.004
0.003
0.001
0.002

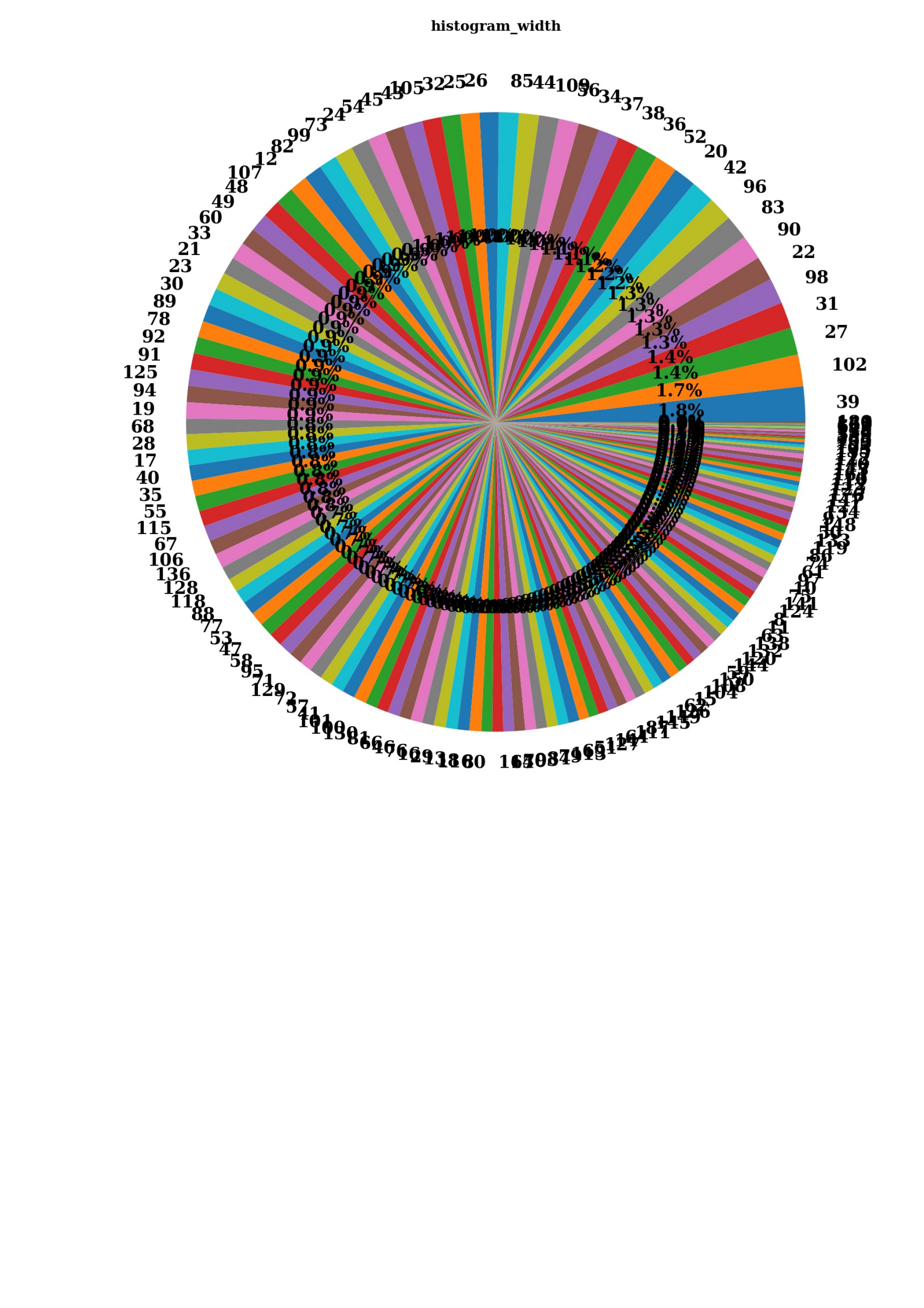

abnormal_short_term_variability
25
22
62
51
61
64
63
65
58
60
18
76
16
17
74
77
79
19
69
73
75
78
68
47
49
70
48
31
46
20
43
45
54
67
33
42
52
44
39
30
21
40
35
66
23
36
38
26
53
29
27
55
41
28
24
32
59
37
56
50
34
57
2.9%
2.8%
2.7%
2.7%
2.7%
2.6%
2.4%
2.0%
1.9%
1.9%
1.9%
1.8%
1.8%
1.8%
1.8%
1.8%
1.7%
1.7%
1.7%
1.7%
1.6%
1.6%
1.6%

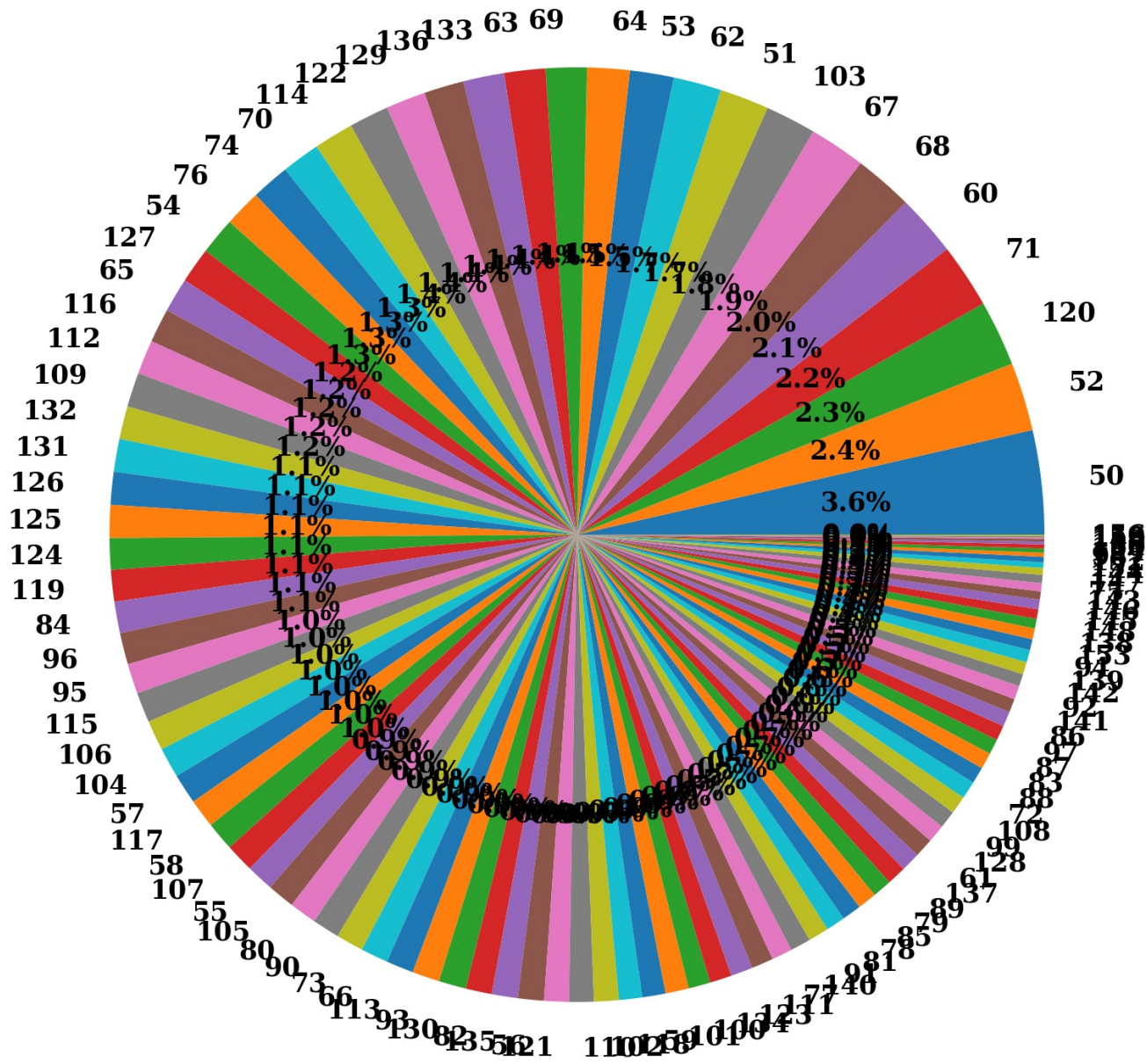

mean_value_of_short_term_variability
0.7
0.4
0.6
0.5
0.9
1.3
1.2
5.5%
5.6%
5.3%
5.7%
5.3%
5.7%
1.5
5.0%
0.8
4.7%
5.8%
1.0
4.7%
4.6%
1.1
4.5%
4.0%
1.4
3.7%
3.6%
2.8%
0.3
1.7
1.6
1.9
1.8
2.2
0.2
2.1
2.0
2.3
2.4
2.5
2.7
2.8
2.6
3.0
2.9

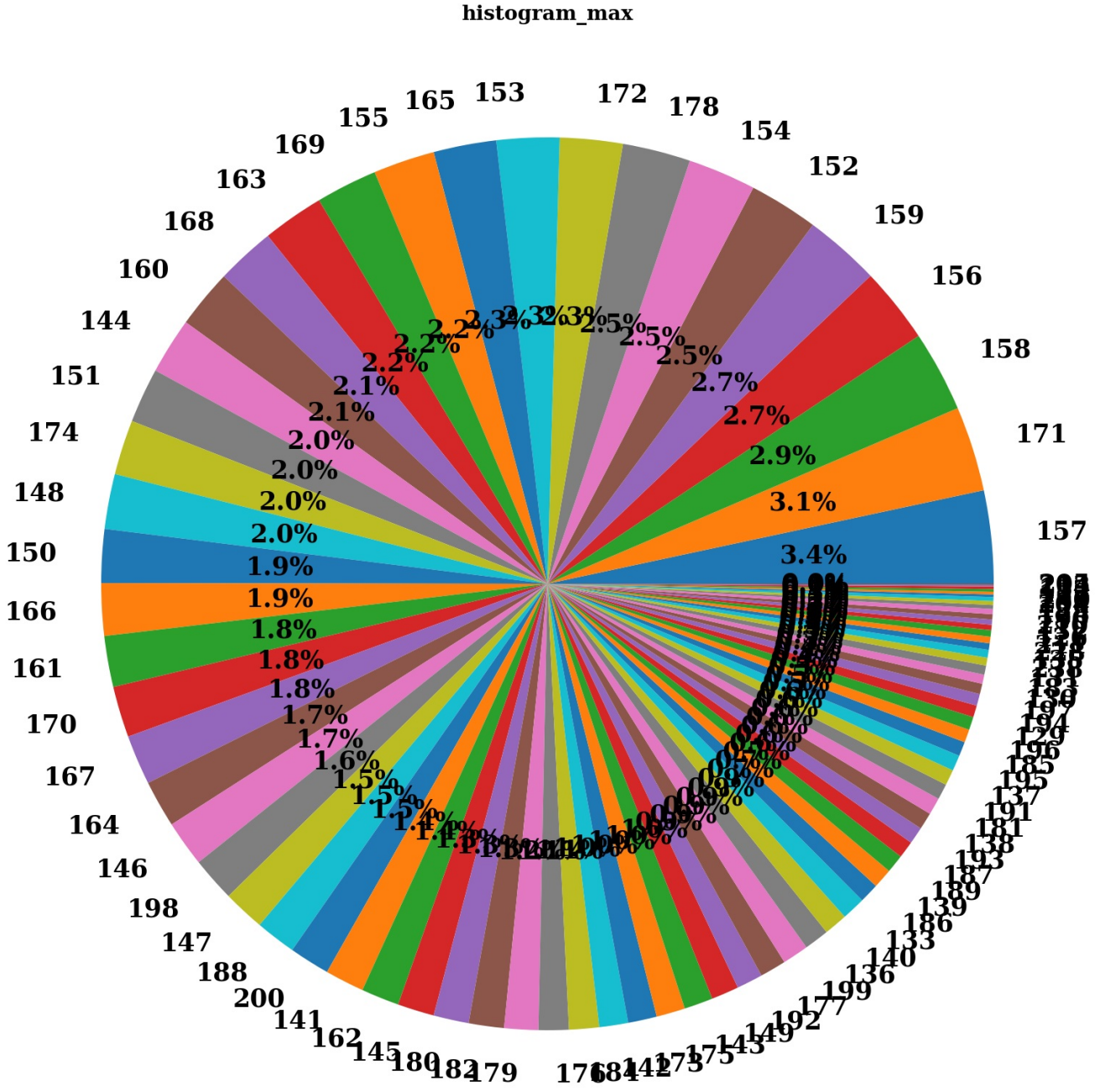

percentage_of_time_with_abnormal_long_term_variability
0
58.4%
2.5%
2.1%
2.0%
1.8%
1
2
5
4
3
8
6
12
10
9
7
16
13
32
20
11
15

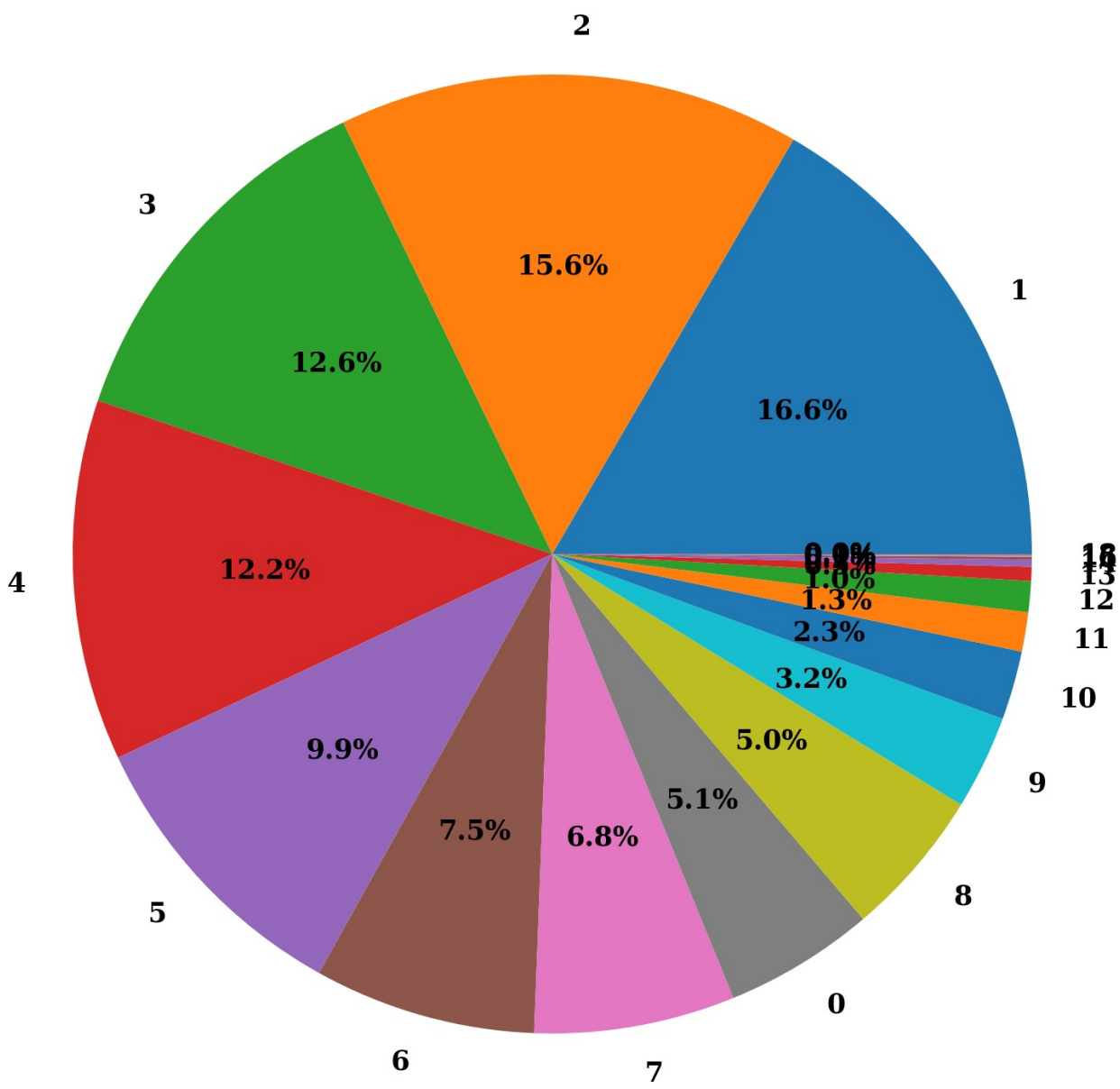
mean_value_of_long_term_variability



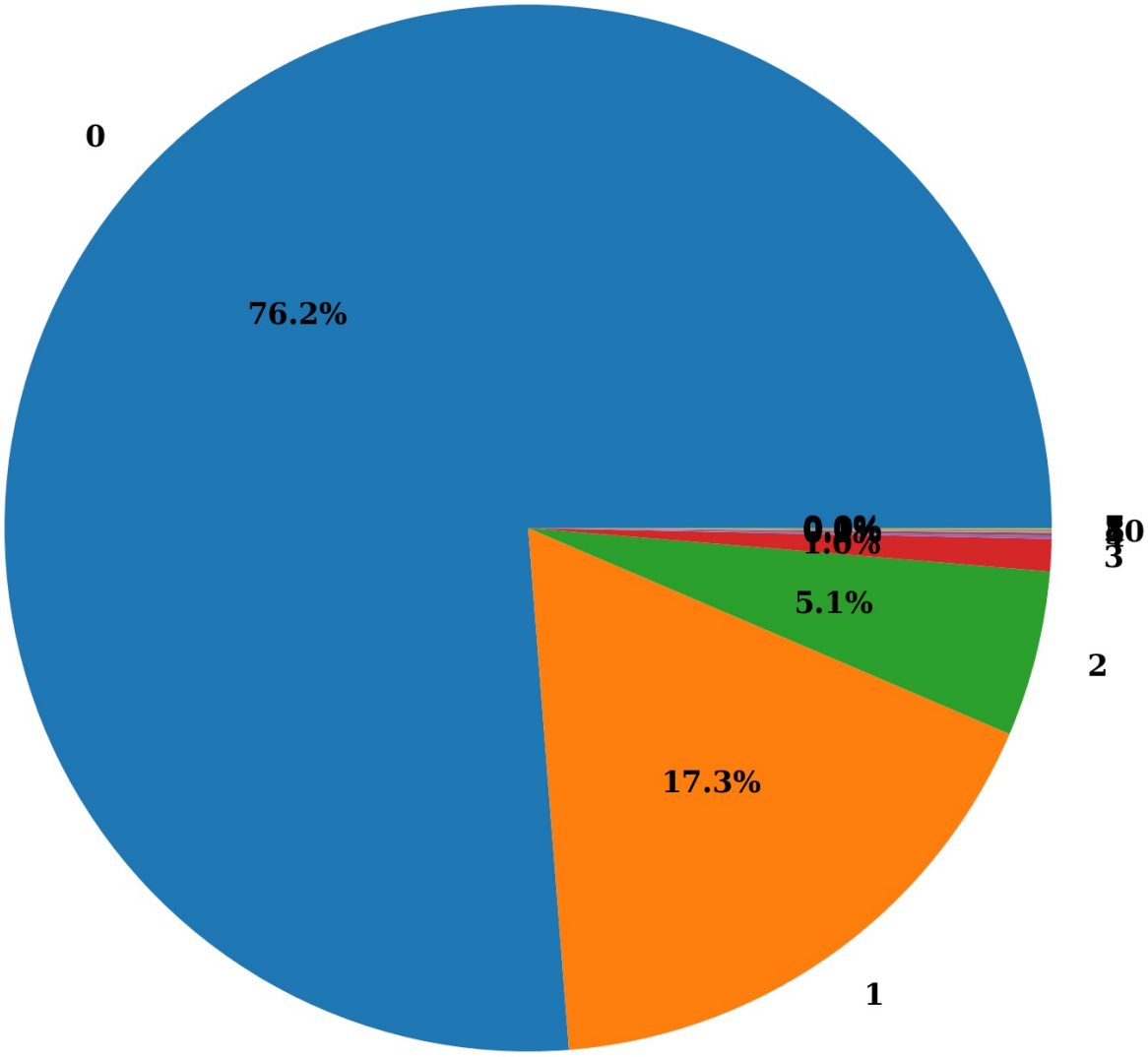

histogram_width

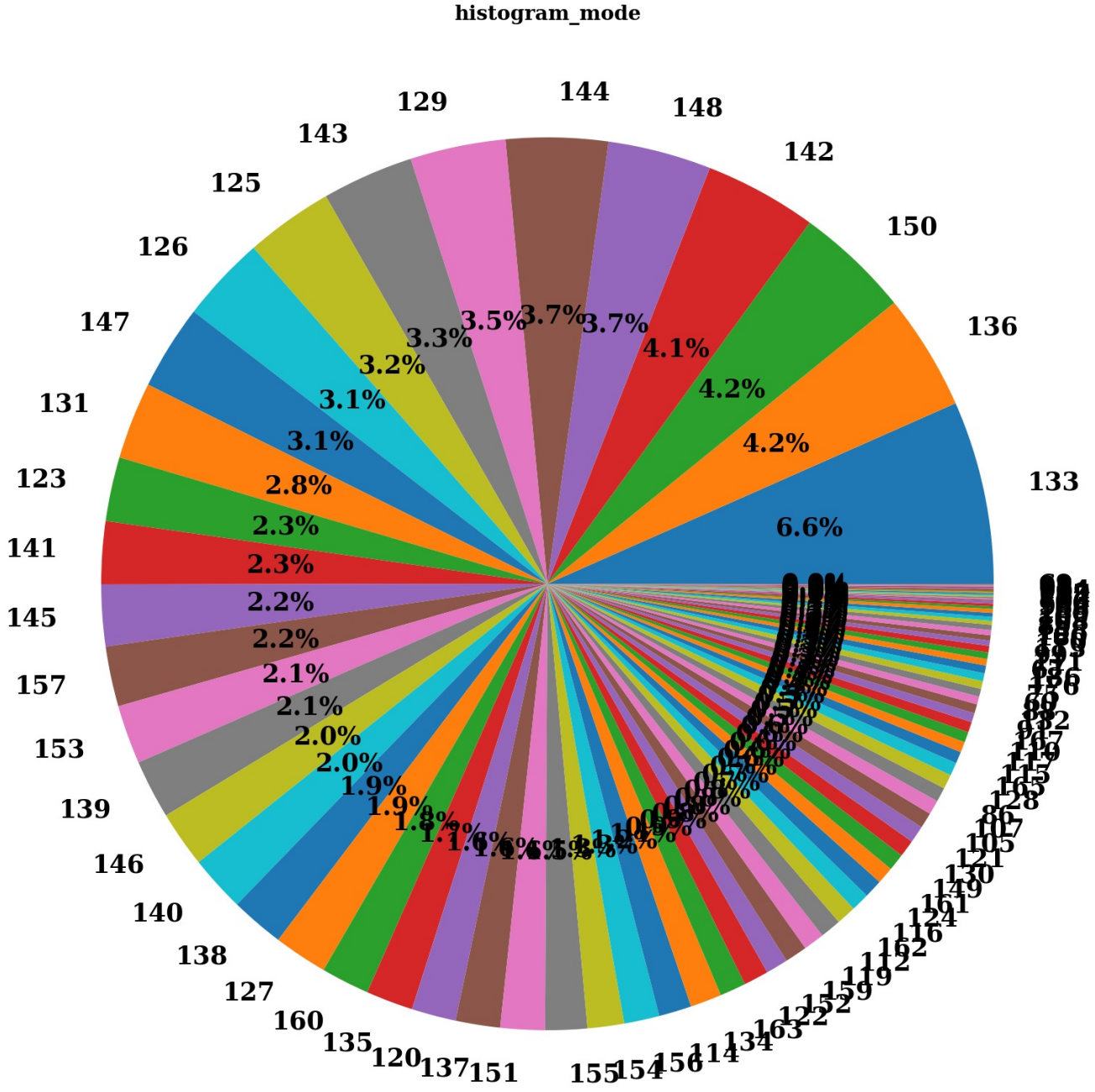

histogram_min
64
53
62
51
103
67
68
60
71
120
52
50
69
63
133
136
129
122
114
70
74
76
54
127
65
116
112
109
132
131
126
125
124
119
84
96
95
115
106
104
57
117
58
107
55
105
80
90
73
66
113
93
130
82
135
56
121
110
102
59
101
100
104
123
111
77
146
91
81
78
85
79
89
137
61
128
99
108
72
88
83
87
86
141
142
139
94
143
97
141
3.6%
2.4%
2.3%
2.2%
2.1%
2.0%
1.9%
1.8%

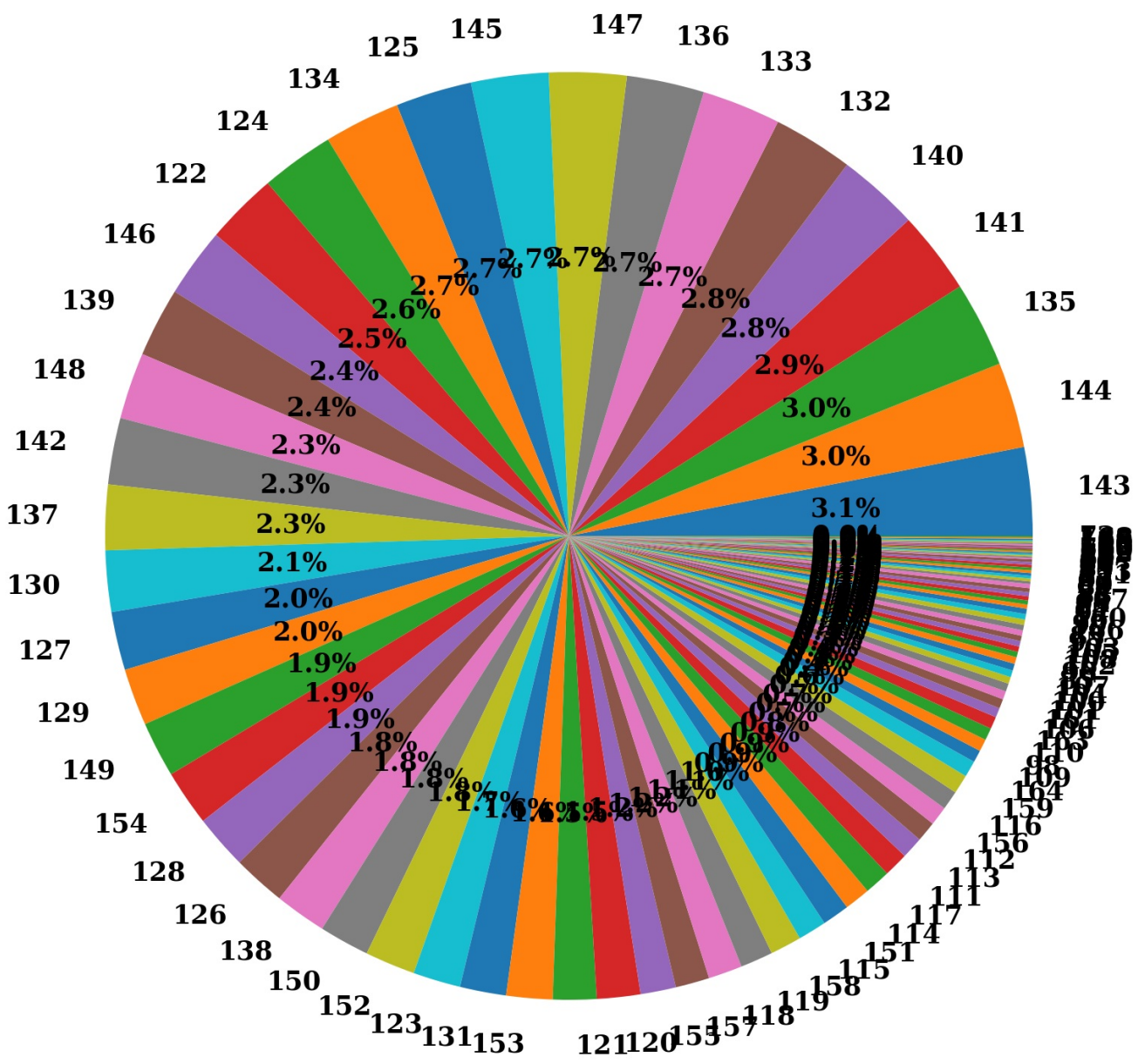

histogram_max
153
172
178
154
152
159
156
158
171
157
165
155
169
163
168
160
144
151
174
148
150
166
161
170
167
164
146
198
147
188
200
141
162
145
180
182
179
176
184
142
173
175
143
149
192
177
199
136
140
133
186
139
189
187
193
138
181
191
137
195
185
196
129
194
197
3.4%
3.1%
2.9%
2.7%
2.7%
2.5%
2.5%
2.5%
2.3%
2.2%
2.2%
2.2%
2.1%
2.1%
2.0%
2.0%
2.0%
2.0%
1.9%
1.9%
1.8%
1.8%
1.8%
1.7%
1.7%
1.6%
1.5%
1.5%
1.5%
1.4%


histogram_number_of_peaks
2
15.6%
3
12.6%
1
16.6%
4
12.2%
1.3%
2.3%
11
3.2%
10
5.0%
9
9.9%
5.1%
7.5%
6.8%
8
5
0
6
7

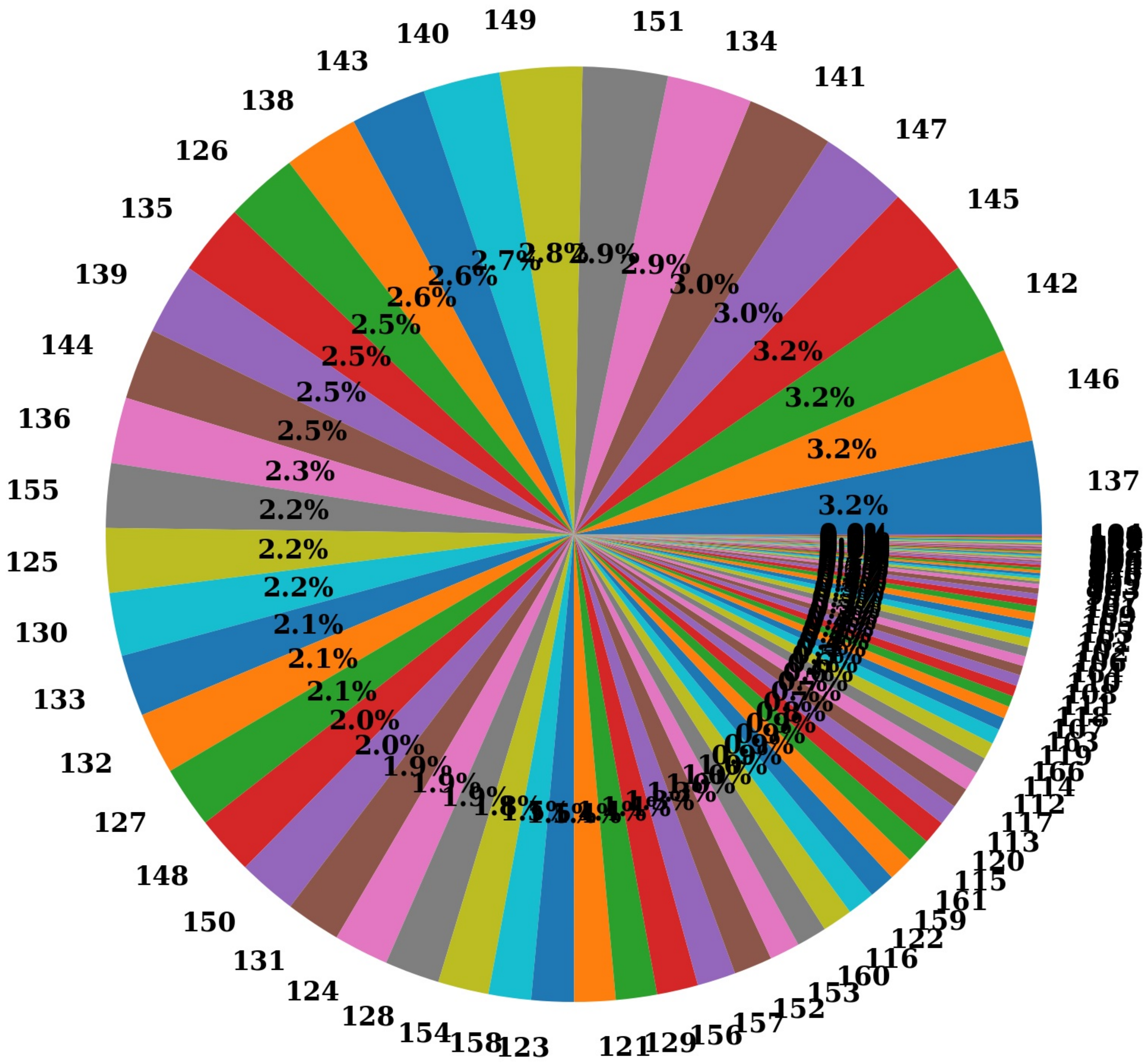
histogram_number_of_zeroes



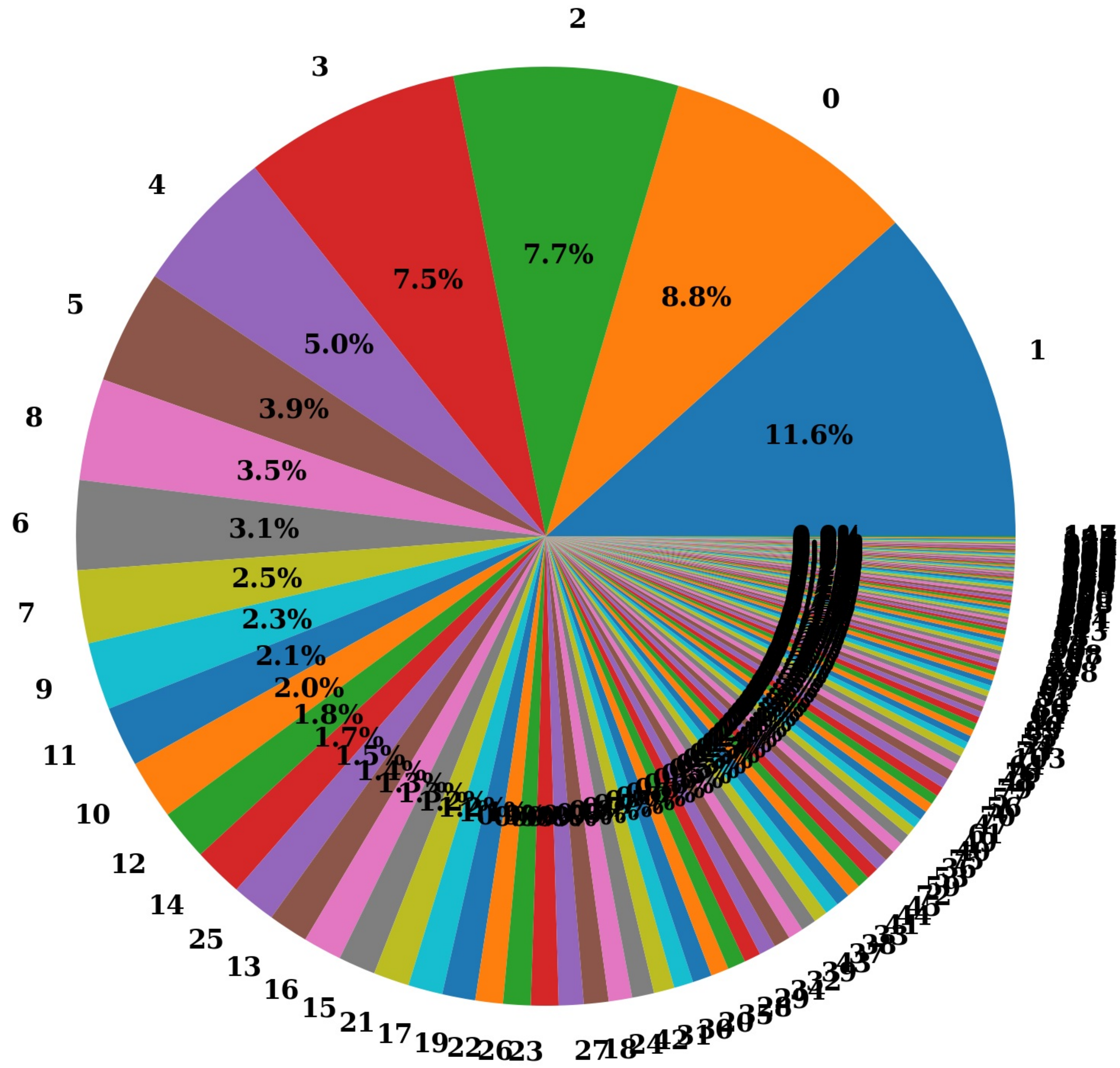

histogram_mode
144
148
142
150
136
133
129
143
125
126
147
131
123
141
145
157
153
139
146
140
138
127
160
135
120
137
151
155
154
156
114
134
163
122
152
159
119
112
162
116
124
151
149
130
121
105
107
86
128
165
115
117
69
92
3.5%
3.7%
3.7%
4.1%
4.2%
4.2%
6.6%
3.3%
3.2%
3.1%
3.1%
2.8%
2.3%
2.3%
2.2%
2.2%
2.1%
2.1%
2.0%
2.0%
1.9%
1.9%

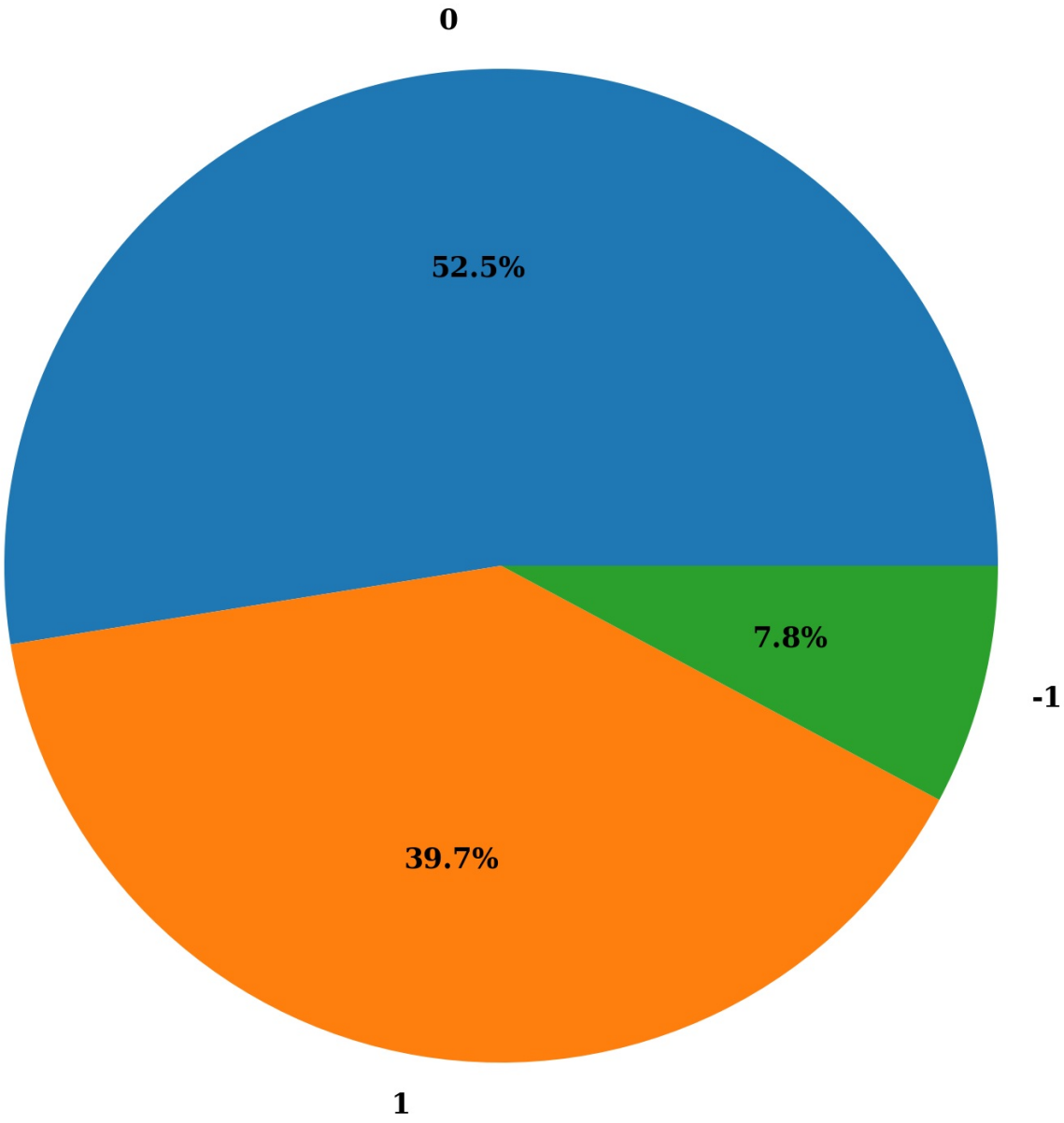

histogram_mean
147
136
133
132
140
141
135
144
143
145
125
134
124
122
146
139
148
142
137
130
127
129
149
154
128
126
138
150
152
123
131
153
121
120
155
157
118
119
158
115
151
114
117
111
113
112
156
116
159
164
109
110
3.1%
3.0%
3.0%
2.9%
2.8%
2.8%
2.7%
2.7%
2.7%
2.7%
2.7%
2.6%
2.5%
2.4%
2.4%
2.3%
2.3%
2.3%
2.1%
2.0%
2.0%
1.9%
1.9%
1.9%
1.8%
1.8%
1.8%
1.8%

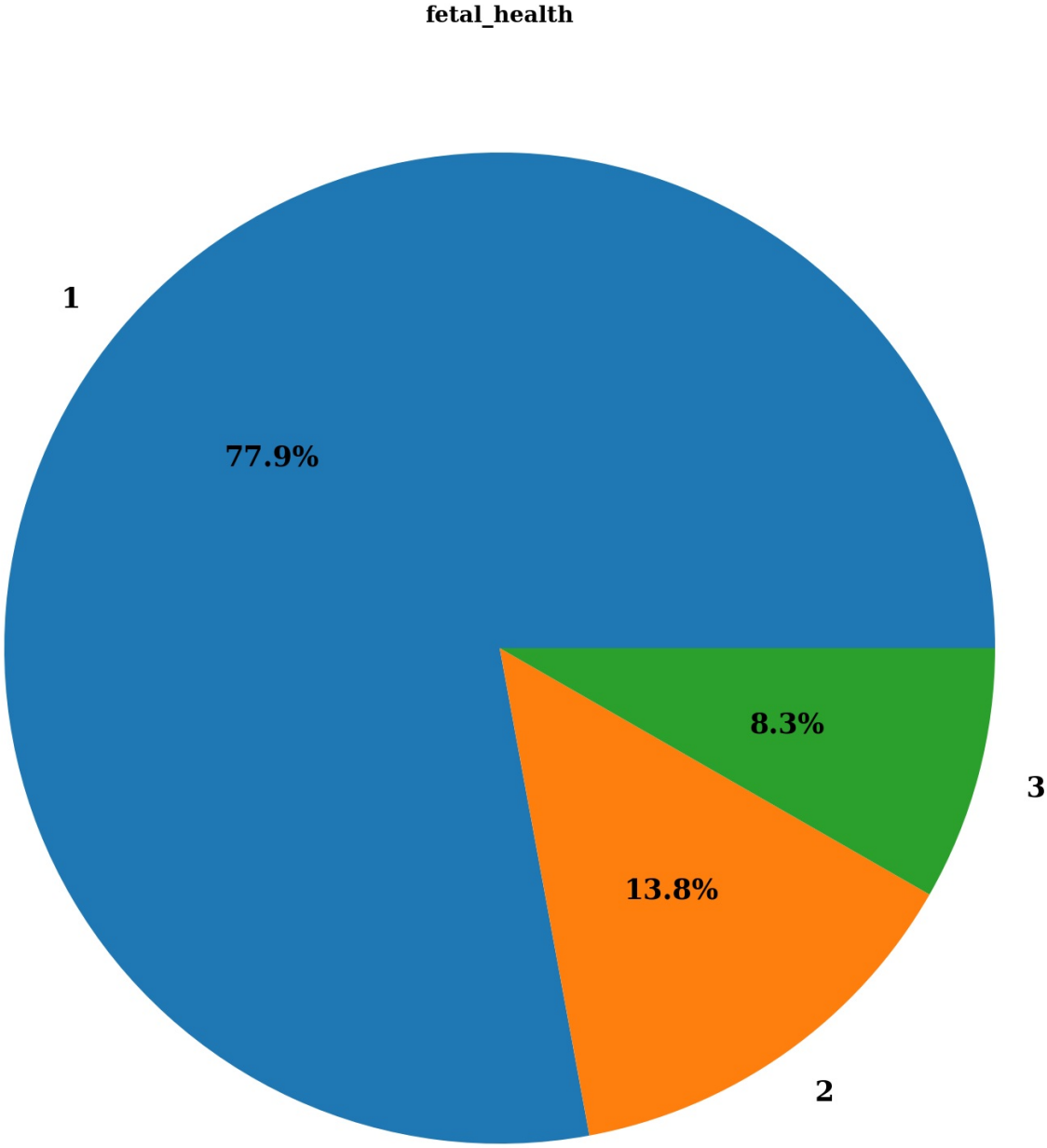
histogram_median



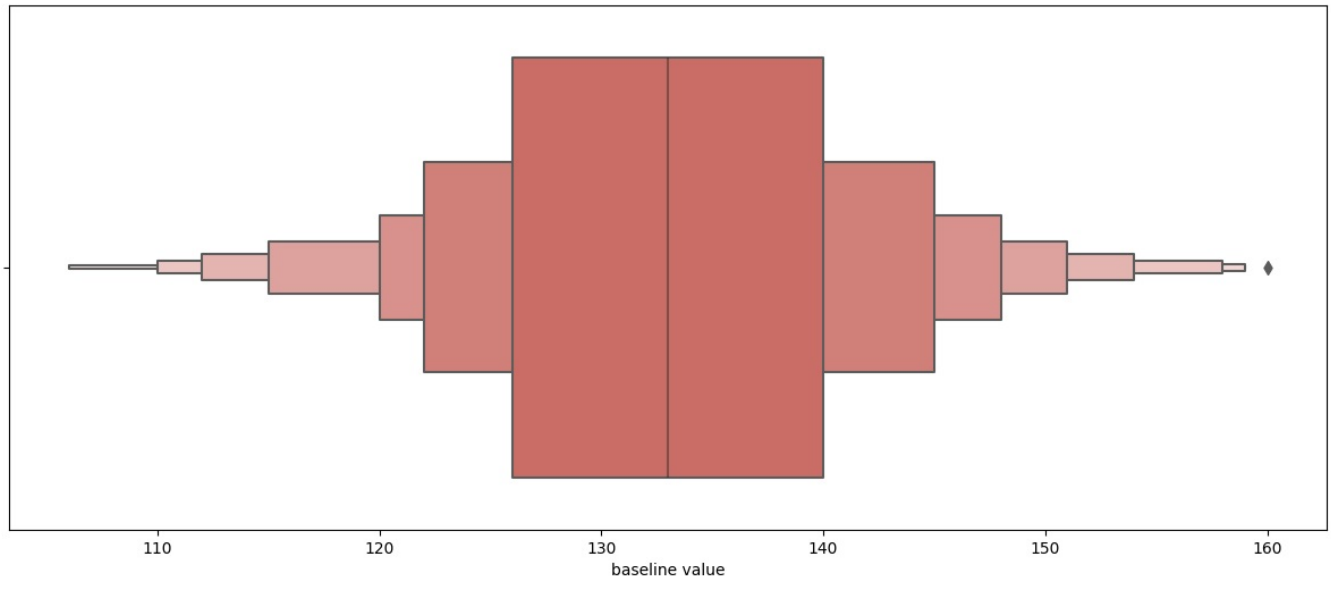
histogram_variance

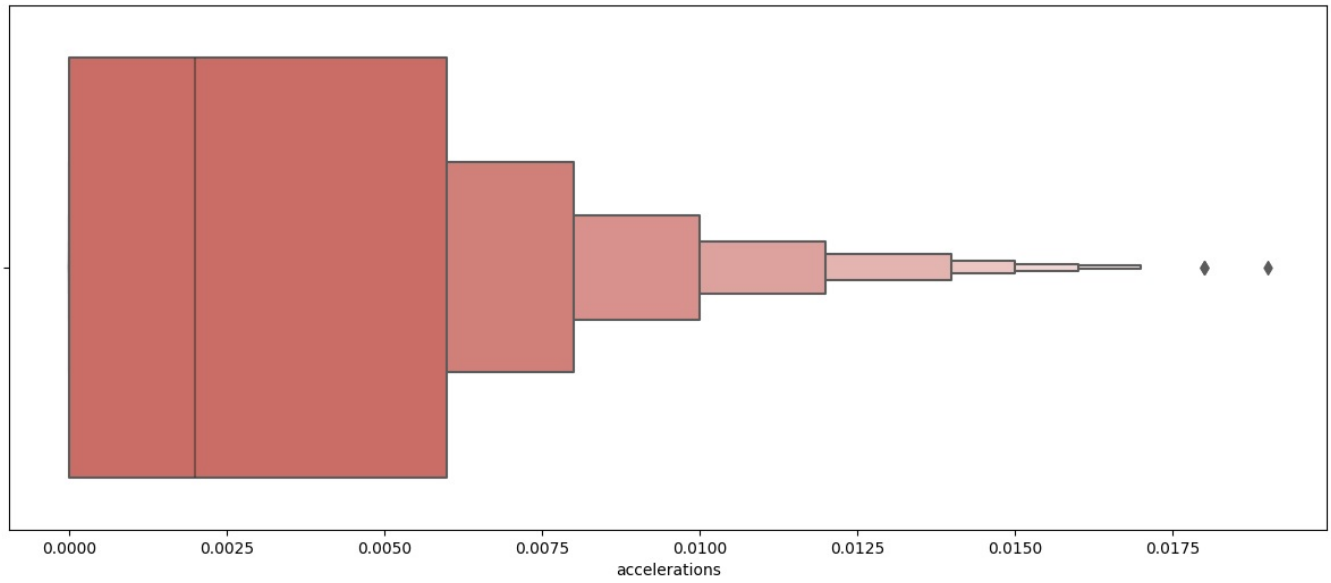


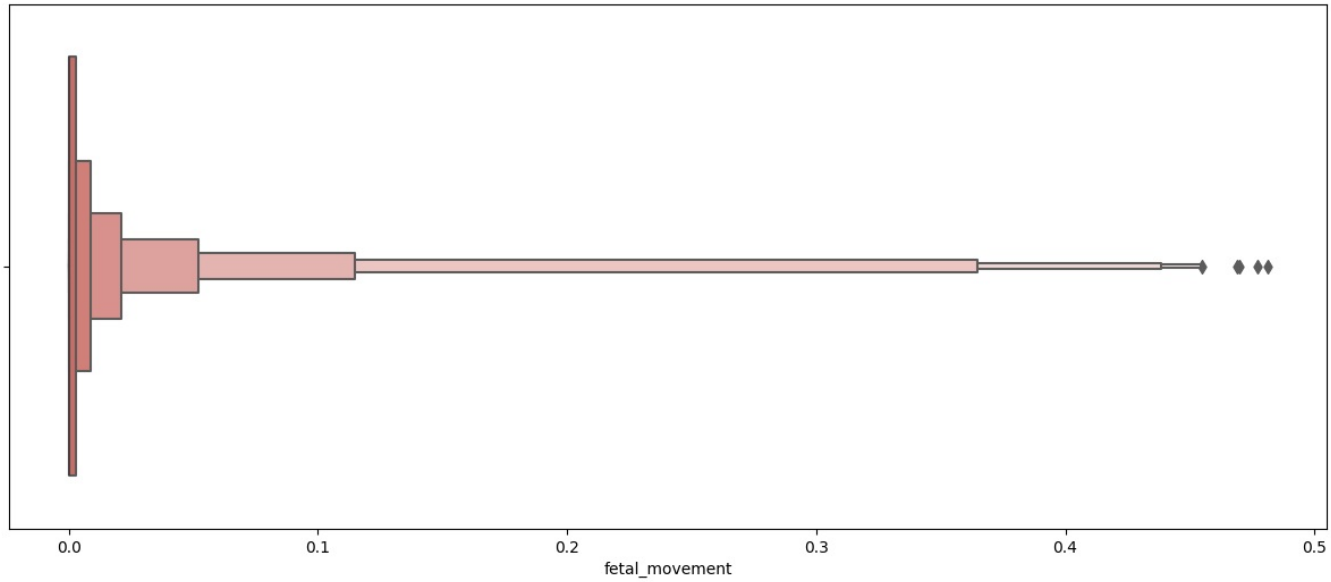

histogram_tendency
0
52.5%
7.8%
-1
39.7%
1

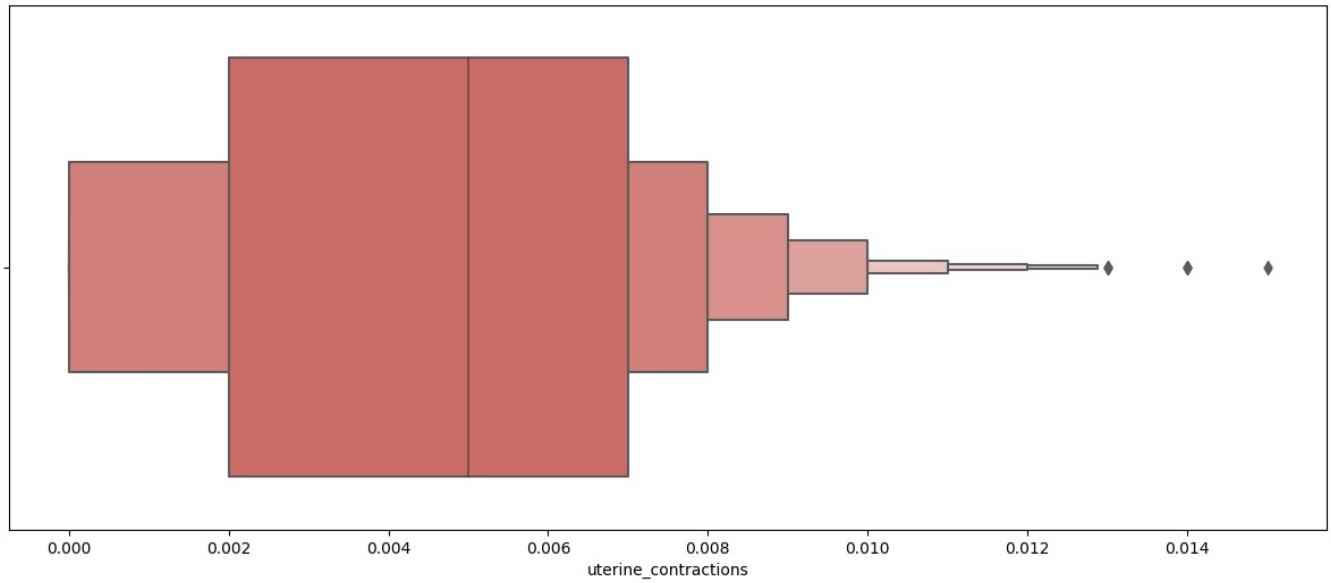


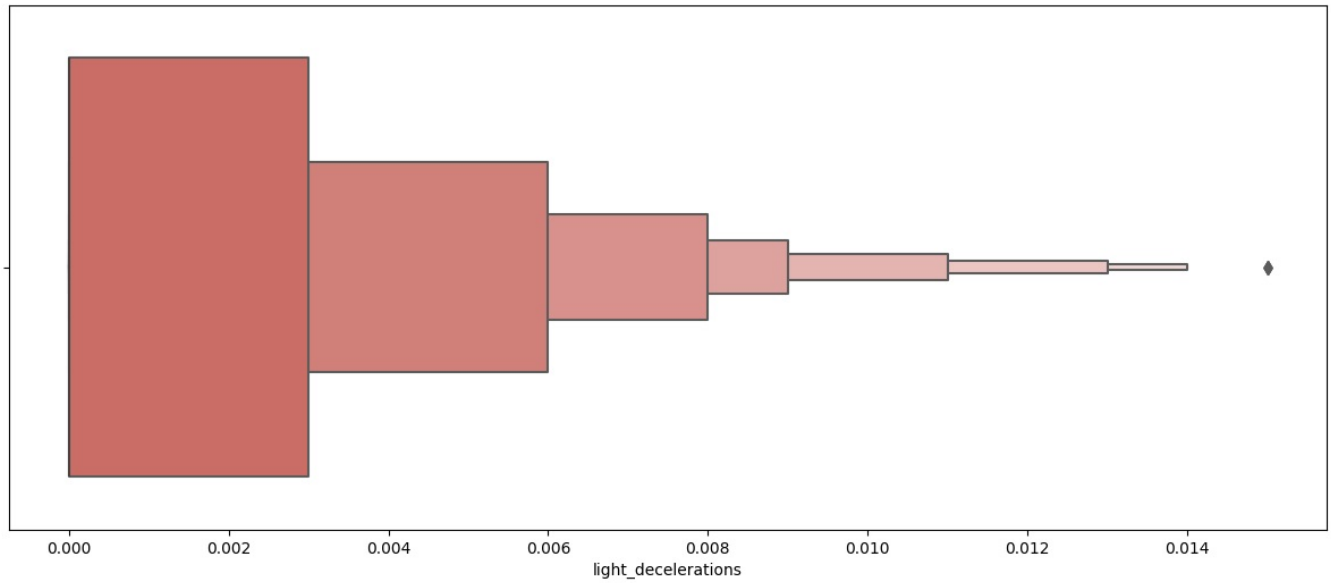
In [18]:
```
for i in df:
    plt.figure(figsize=(15,6))
    sns.boxenplot(x =df[i], data = df, palette = 'hls')
    plt.show()
```

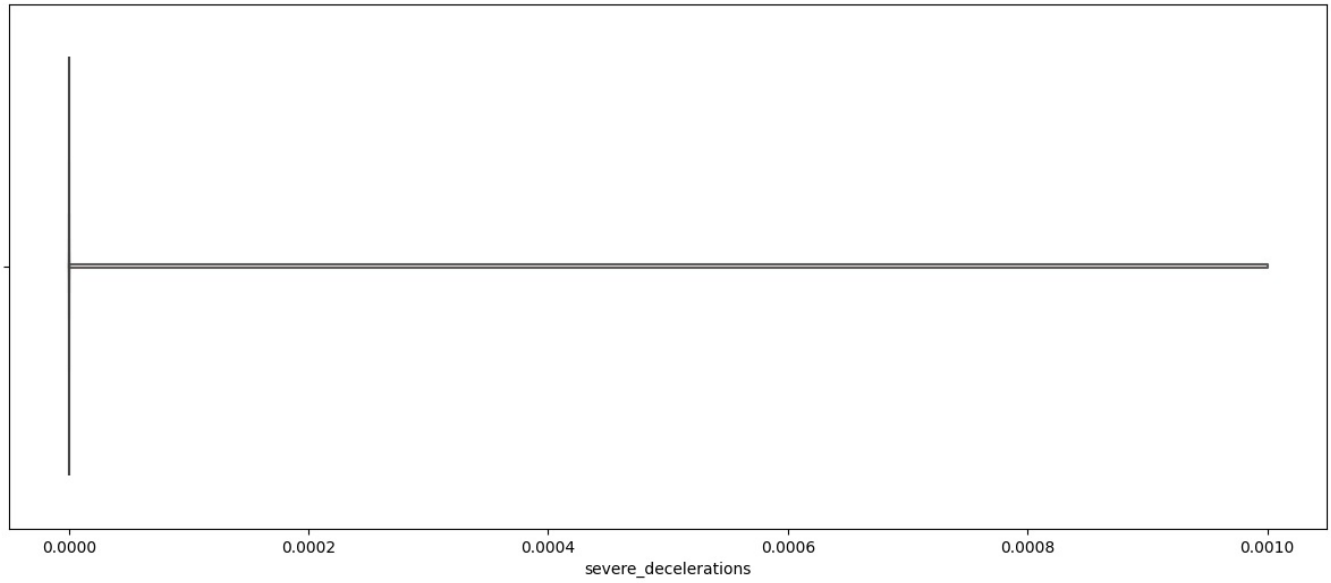


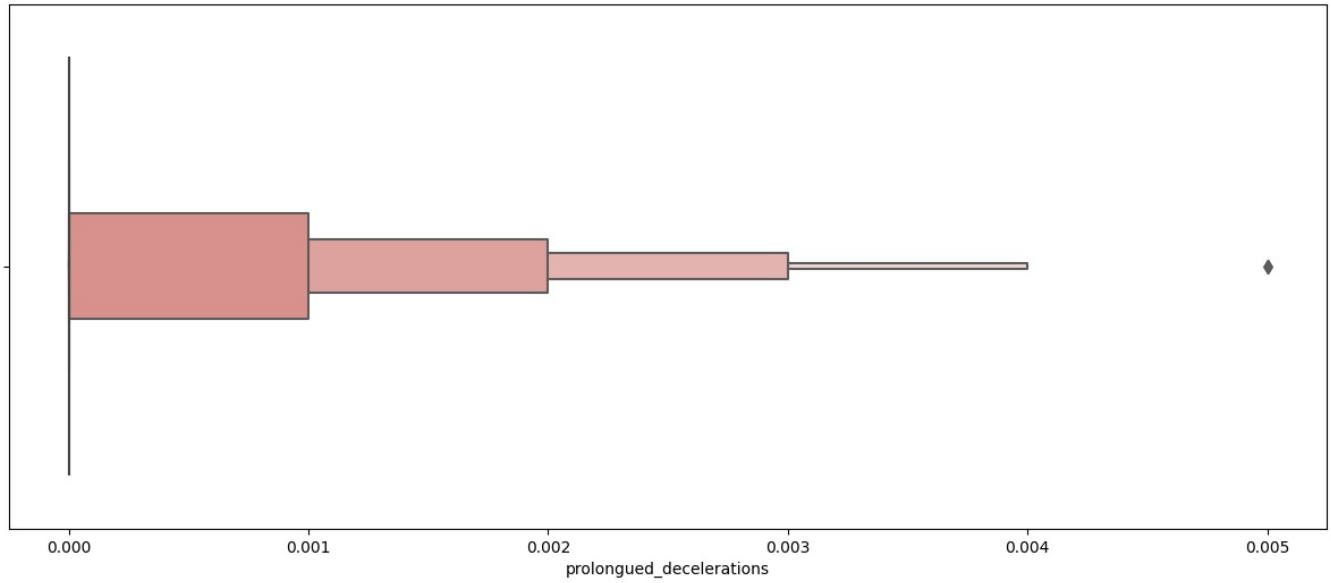

0.0000
0.0025
0.0050
0.0075
0.0100
0.0125
0.0150
0.0175
accelerations

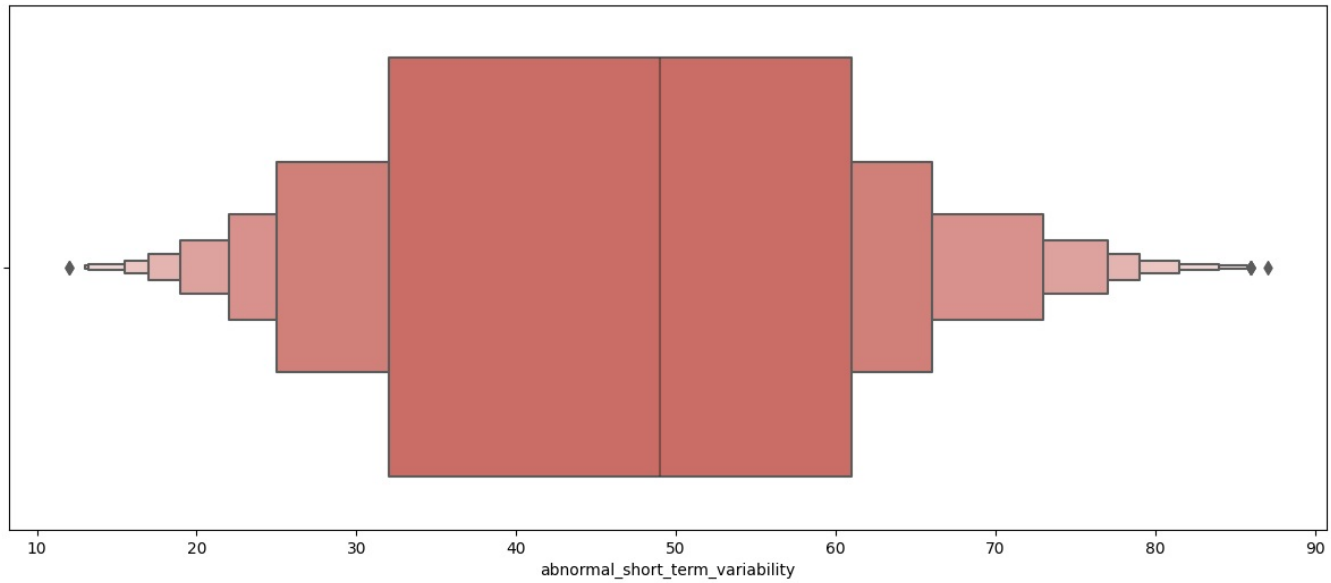

0.0
0.1
0.2
0.3
0.4
0.5
fetal_movement

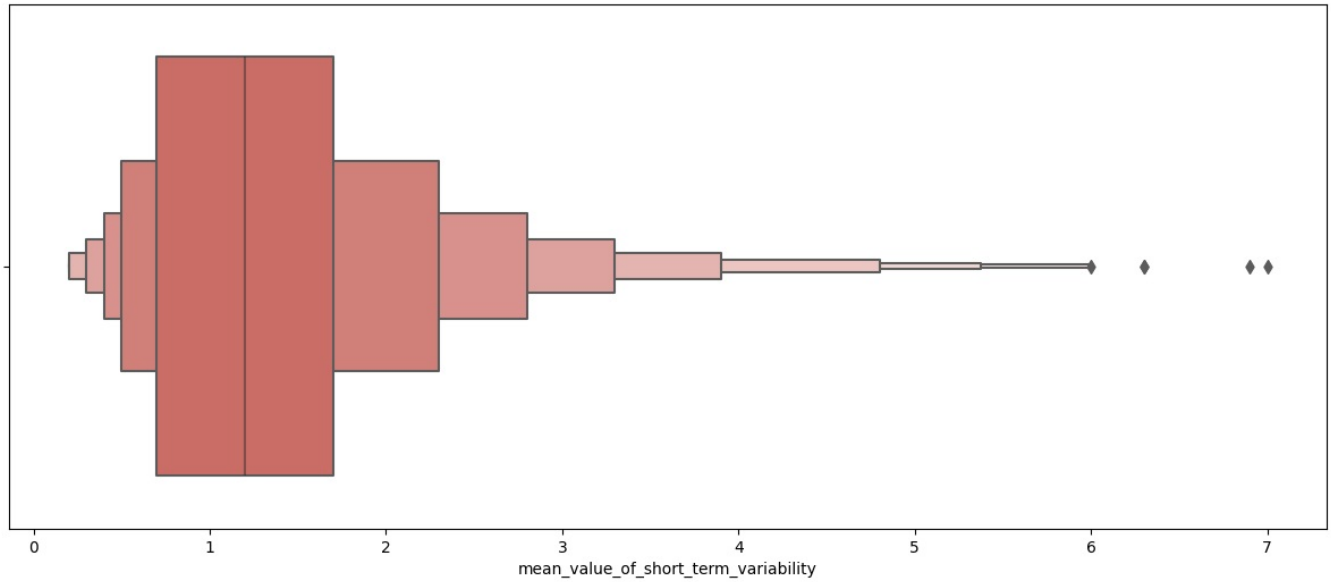

0.000
0.002
0.004
0.006
0.008
0.010
0.012
0.014
uterine_contractions

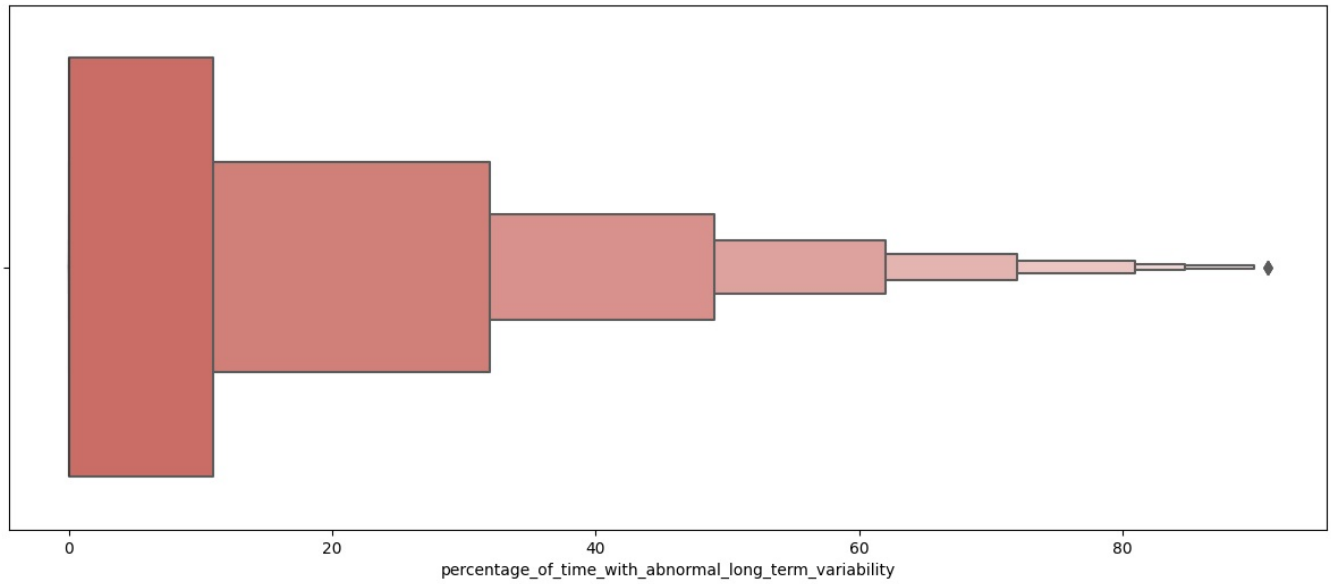

0.000
0.002
0.004
0.006
0.008
0.010
0.012
0.014
light_decelerations

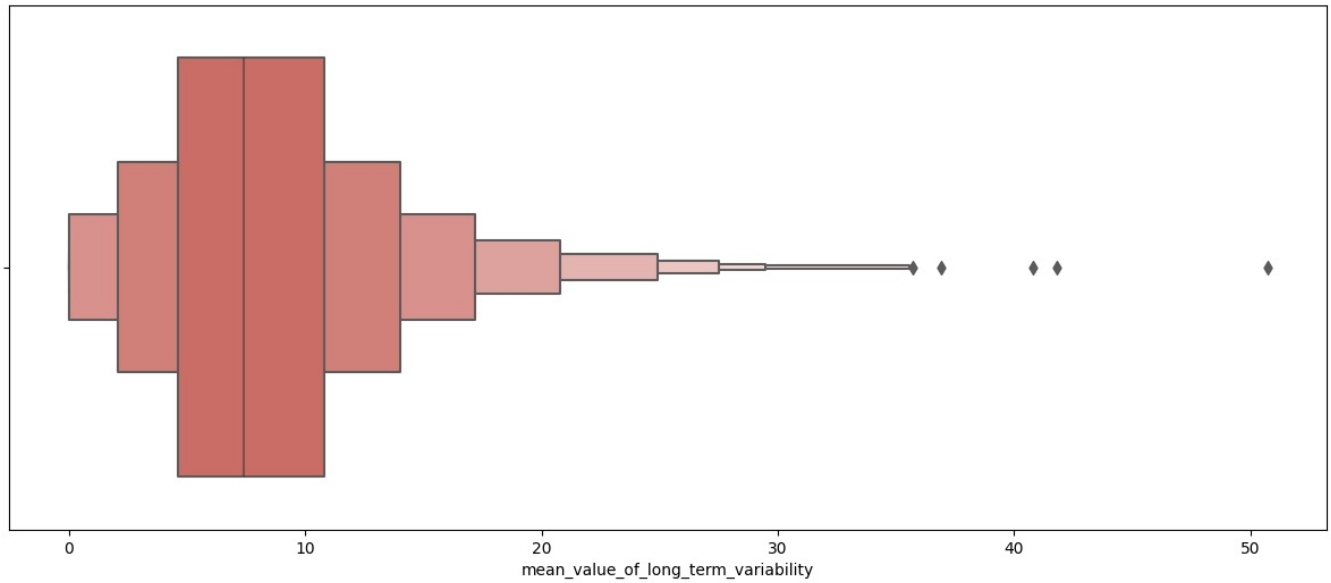

0.0000
0.0002
0.0004
0.0006
0.0008
0.0010
severe_decelerations

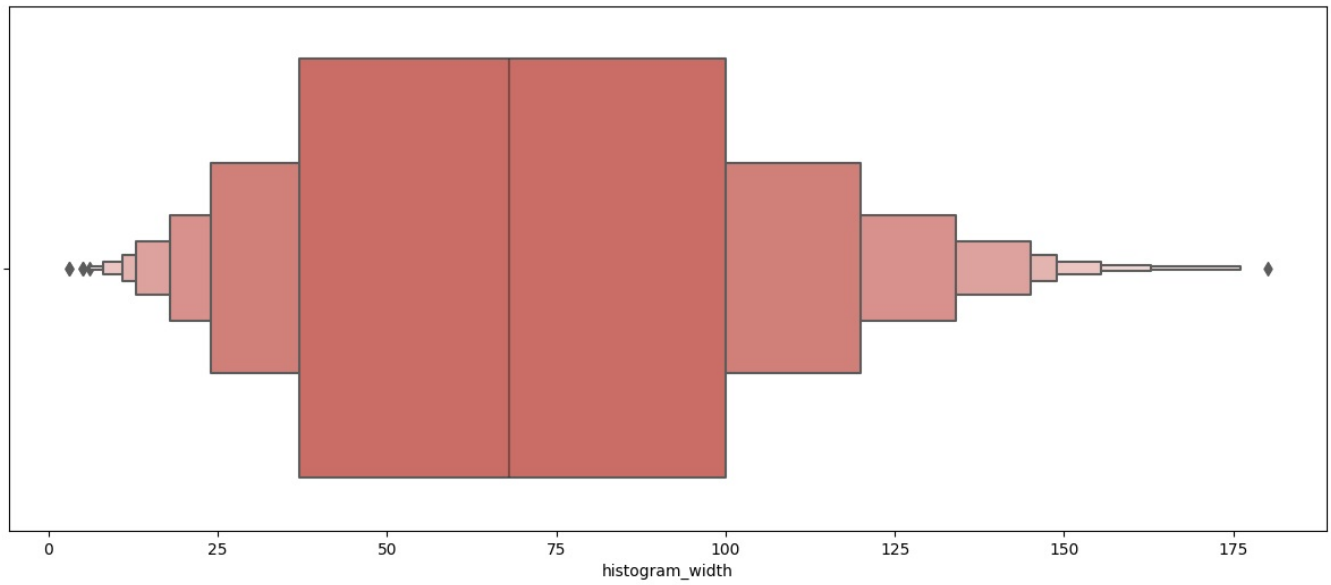

0.000
0.001
0.002
0.003
0.004
0.005
prolongued_decelerations

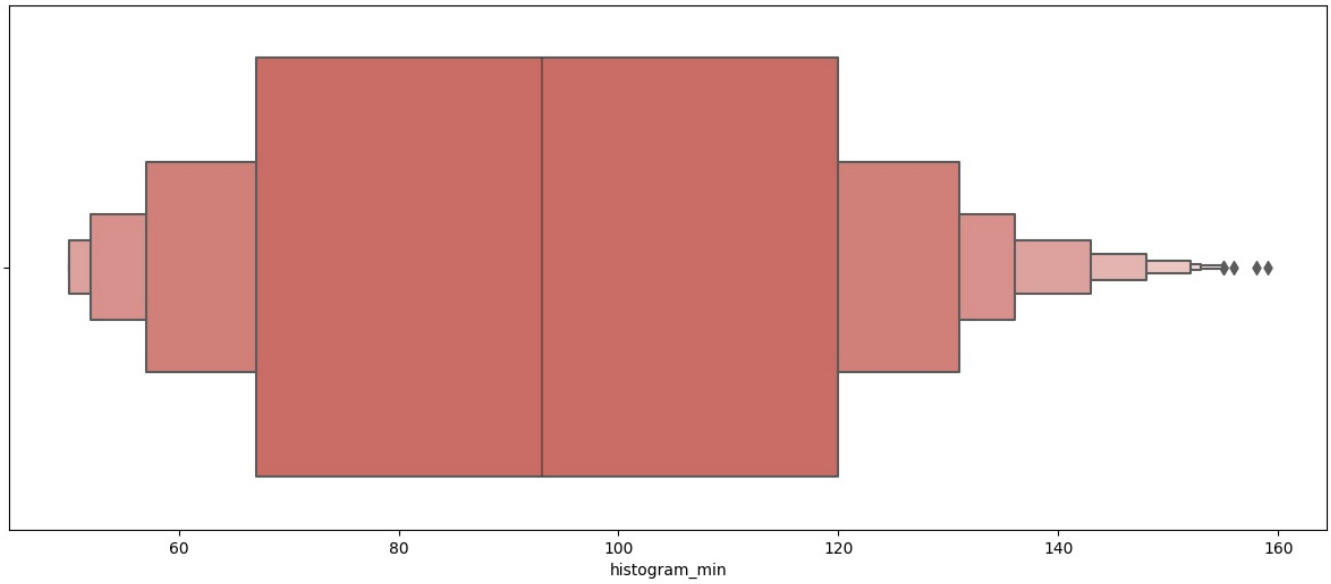

10
20
30
40
50
60
70
80
90
abnormal_short_term_variability

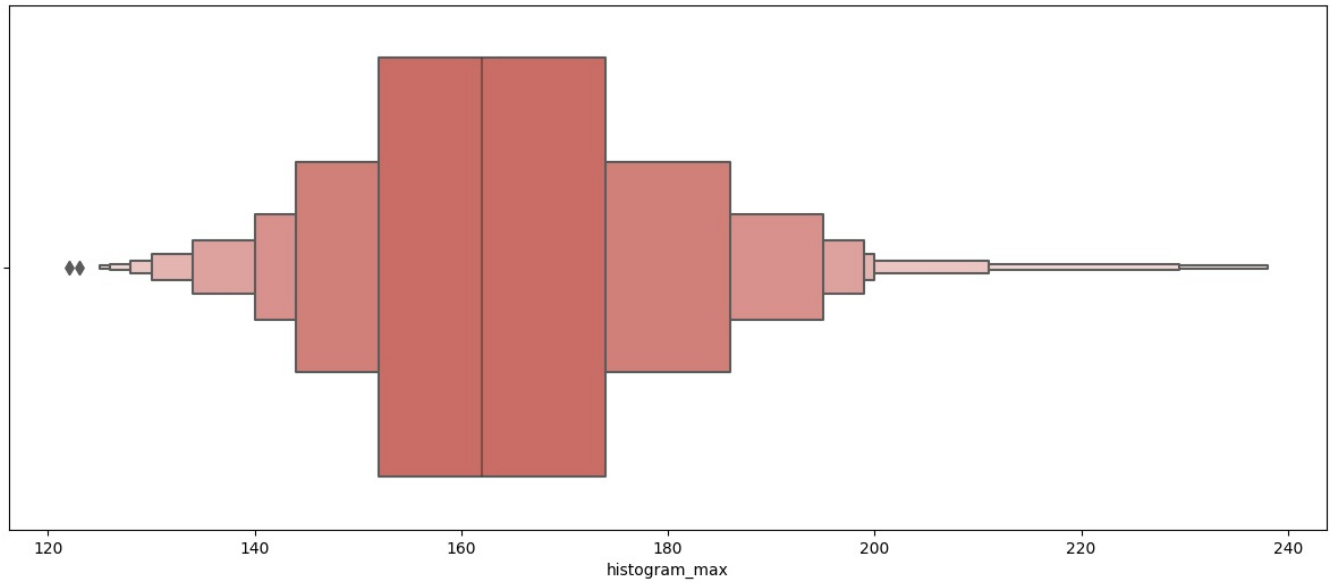

0
1
2
3
4
5
6
7
mean_value_of_short_term_variability

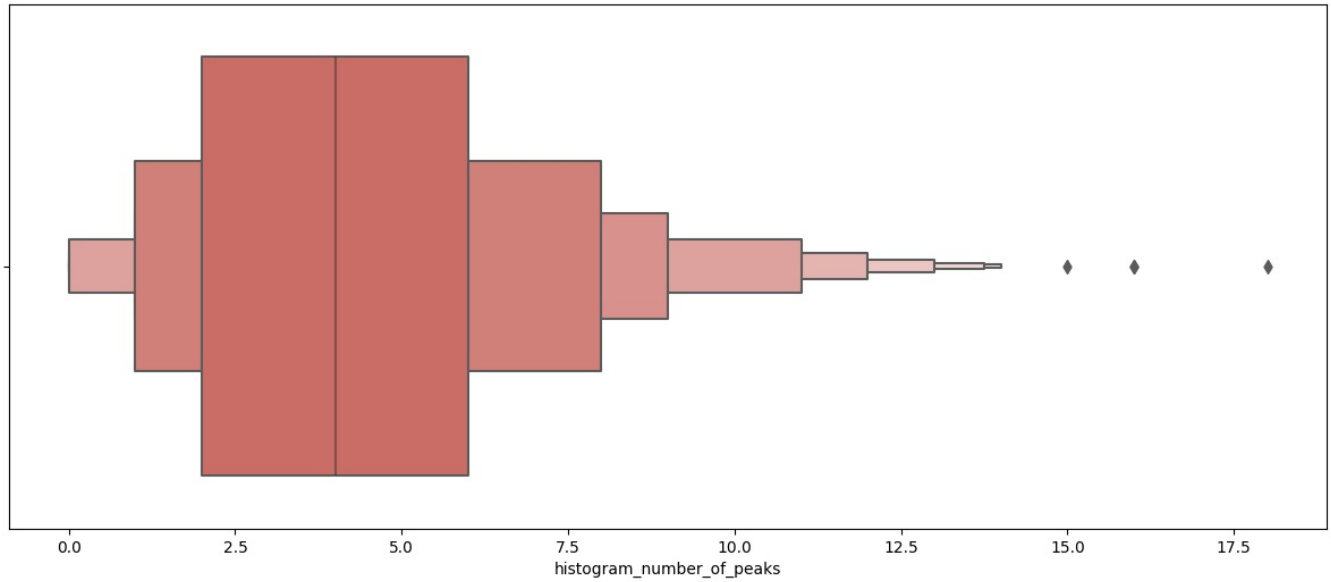

0
20
40
60
80
percentage_of_time_with_abnormal_long_term_variability

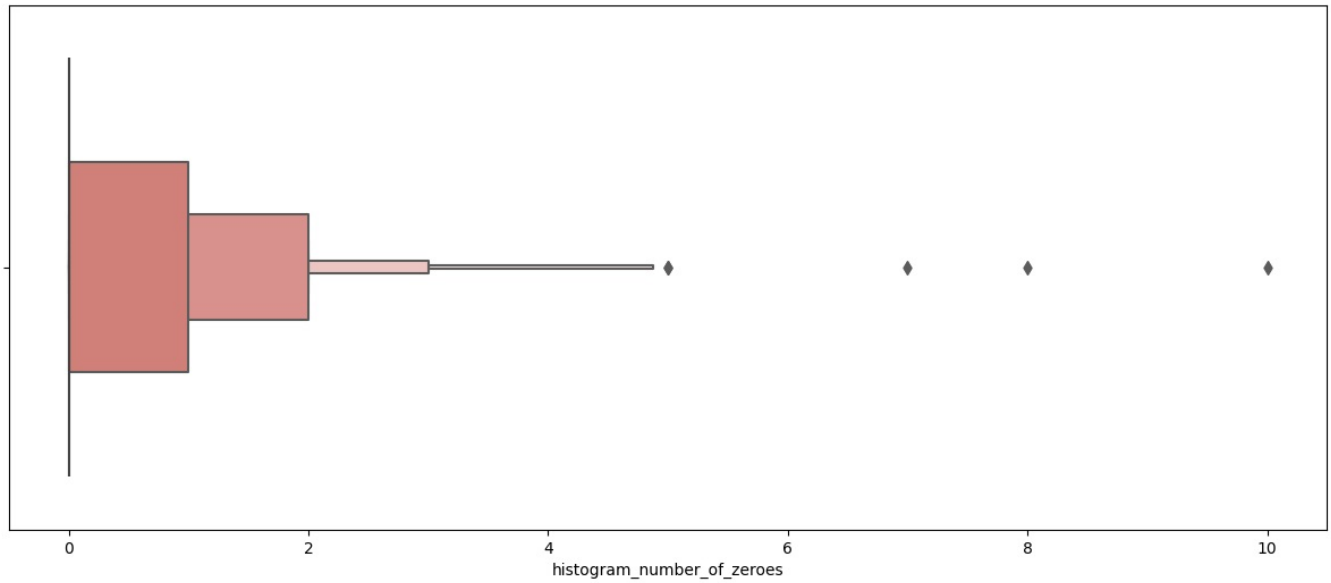

0
10
20
30
40
50
mean_value_of_long_term_variability

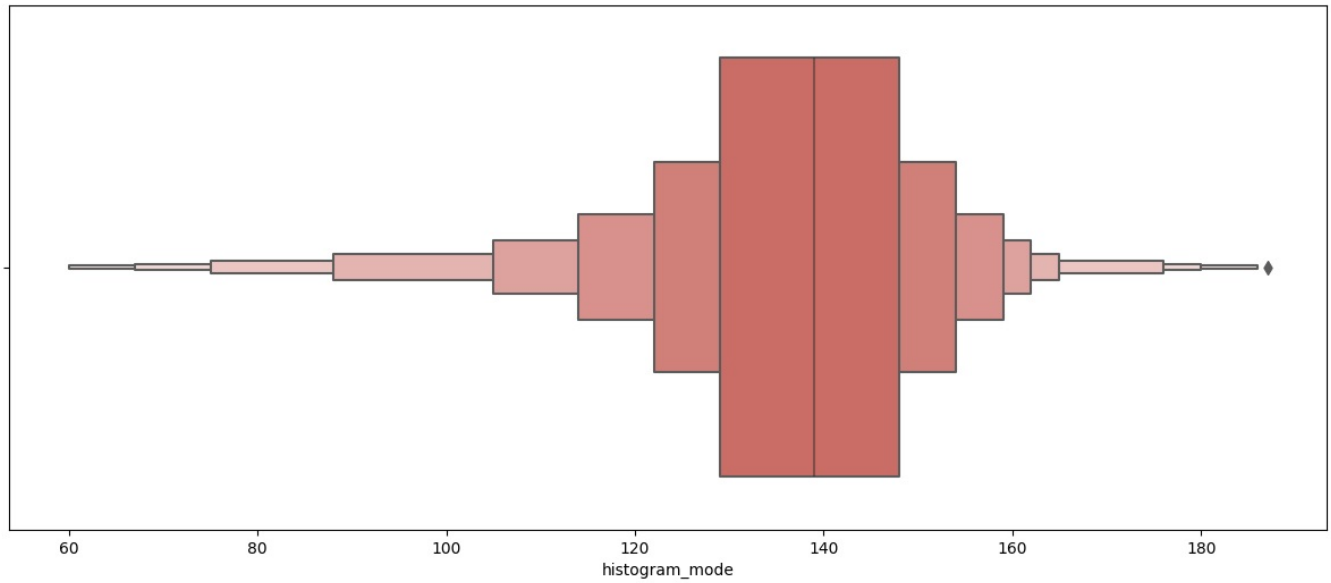

0
25
50
75
100
125
150
175
histogram_width

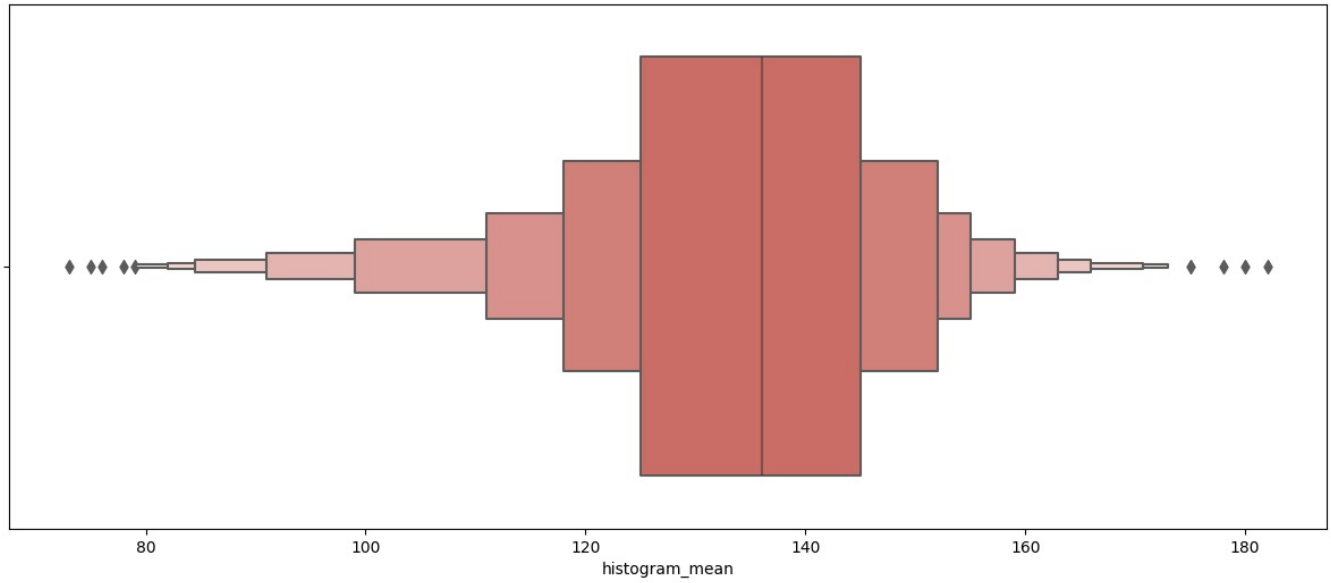

60
80
100
120
140
160
histogram_min

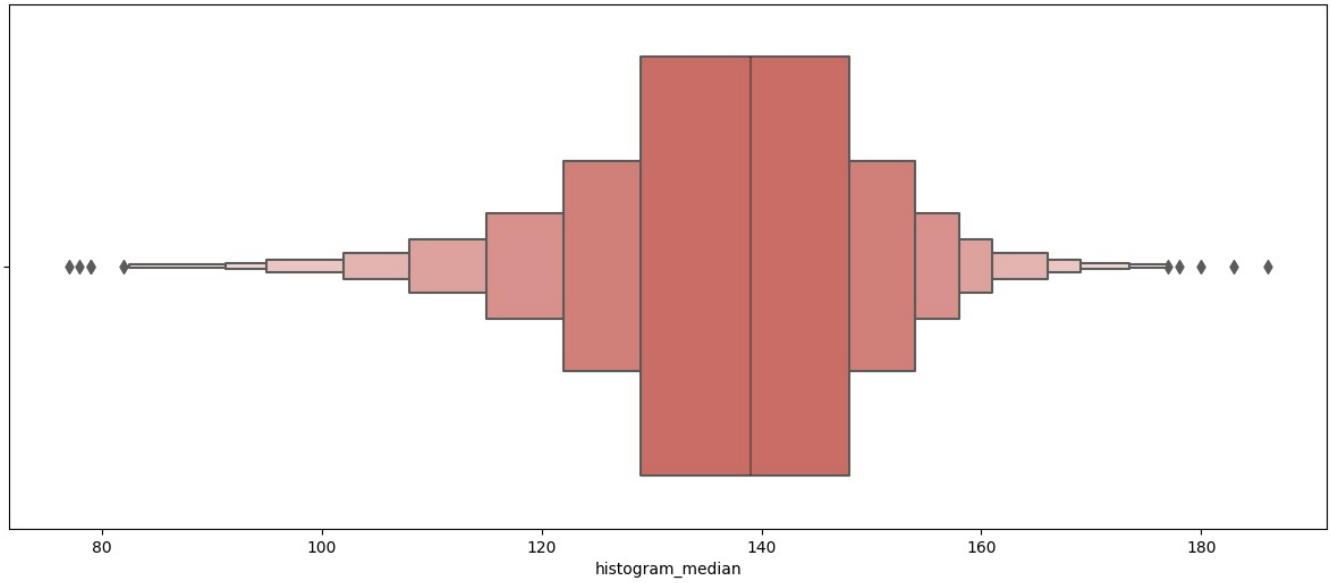

120
140
160
180
200
220
240
histogram_max

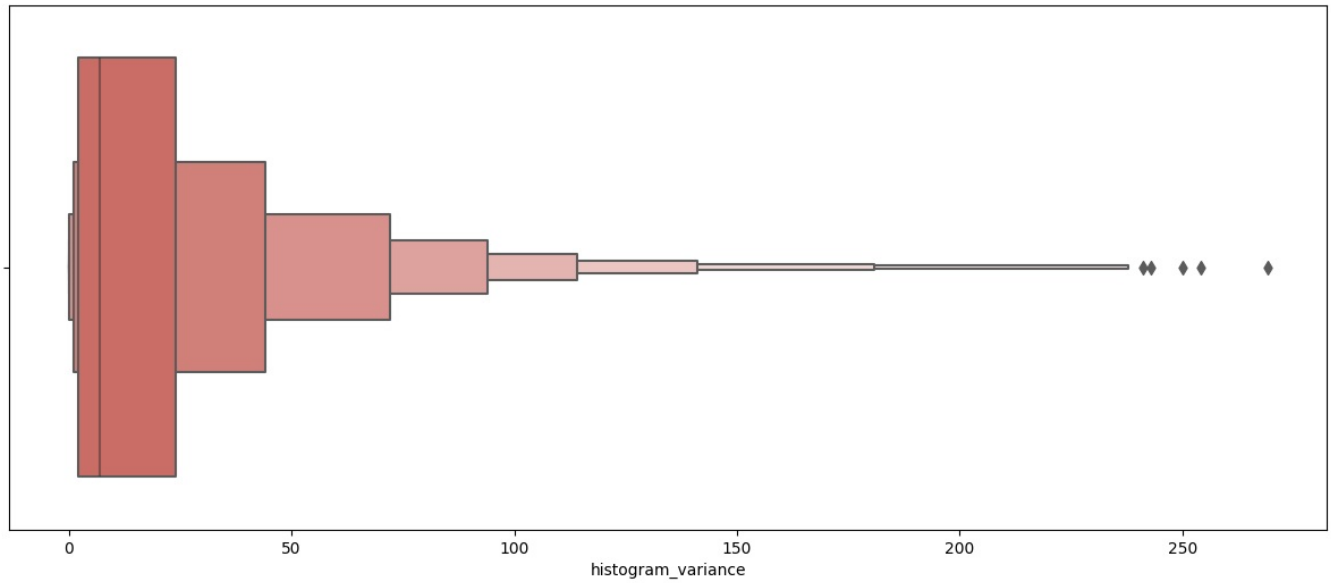

0.0
2.5
5.0
7.5
10.0
12.5
15.0
17.5
histogram_number_of_peaks

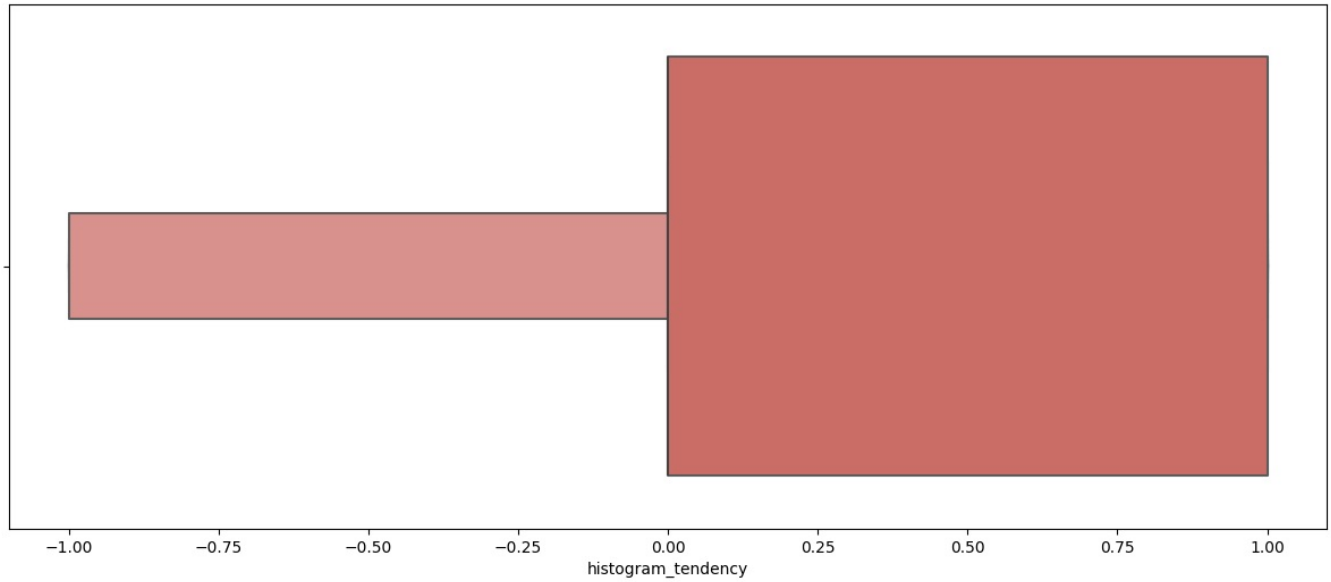

0
2
4
6
8
10
histogram_number_of_zeroes

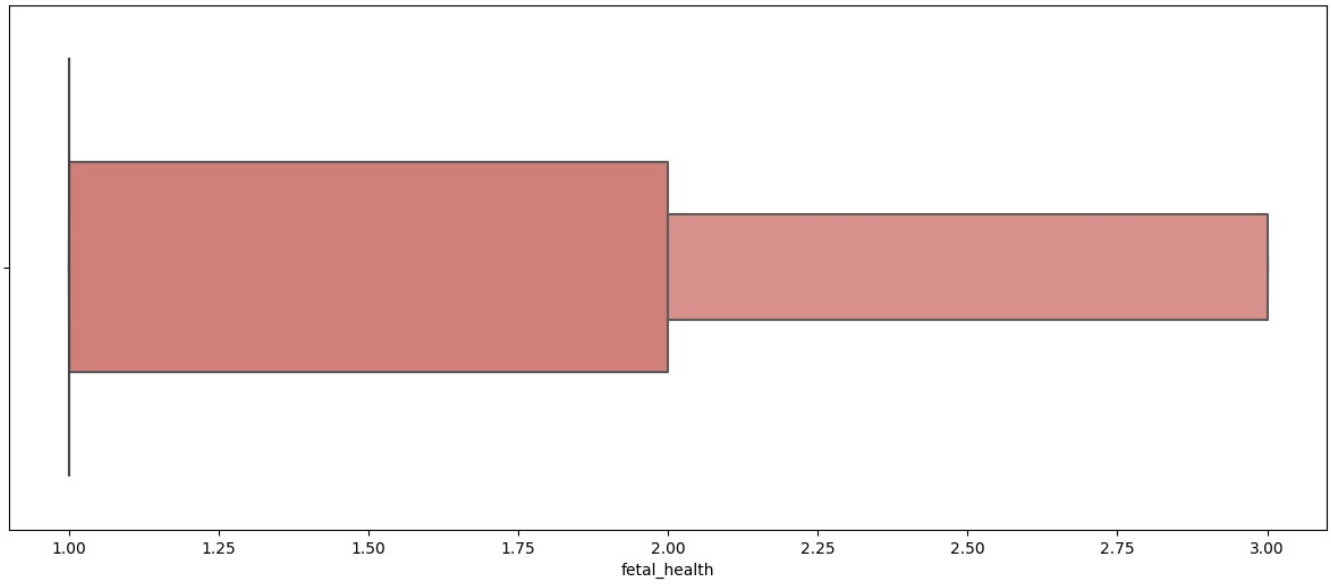

60
80
100
120
140
160
180
histogram_mode


80
100
120
140
160
180
histogram_mean

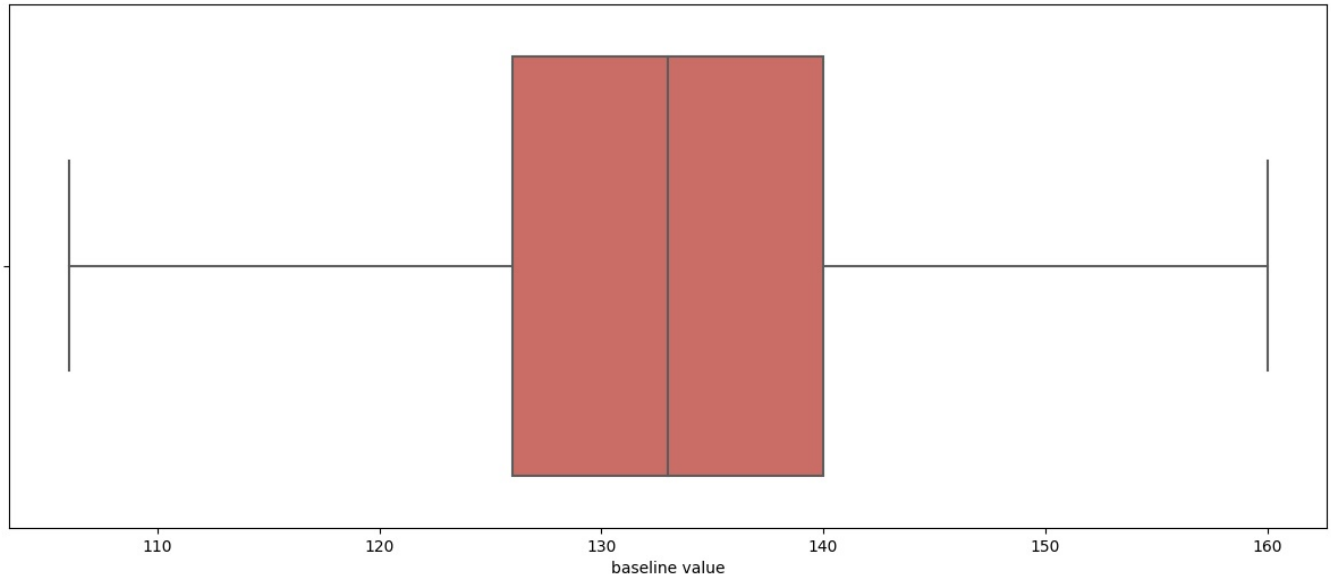

80
100
120
140
160
180
histogram_median

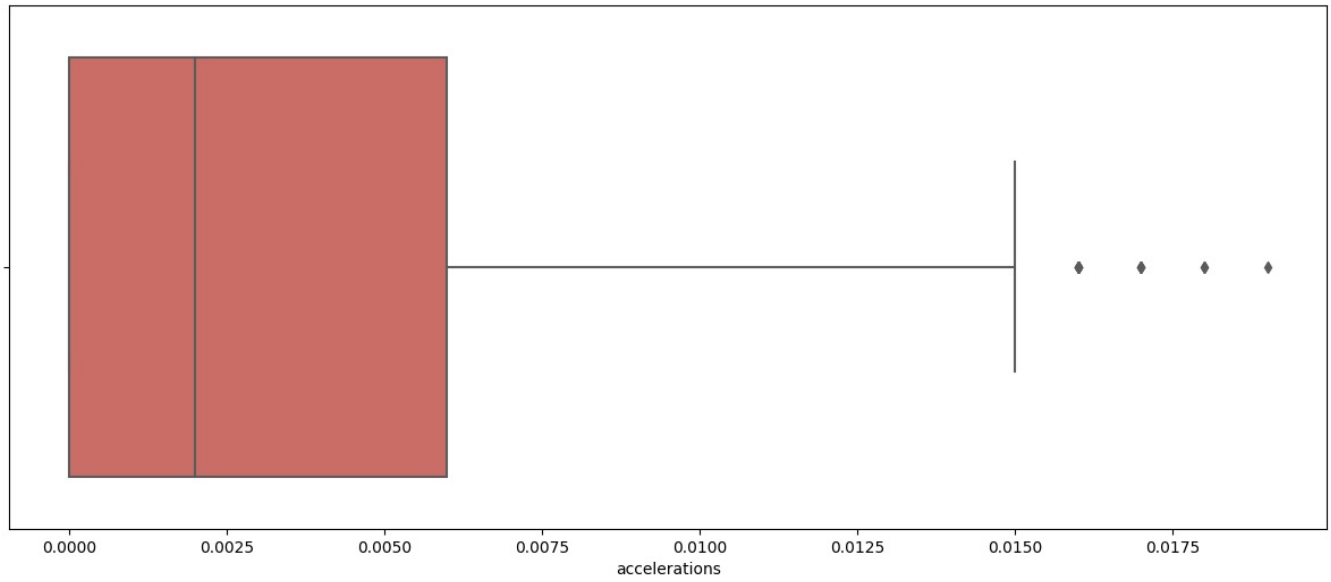


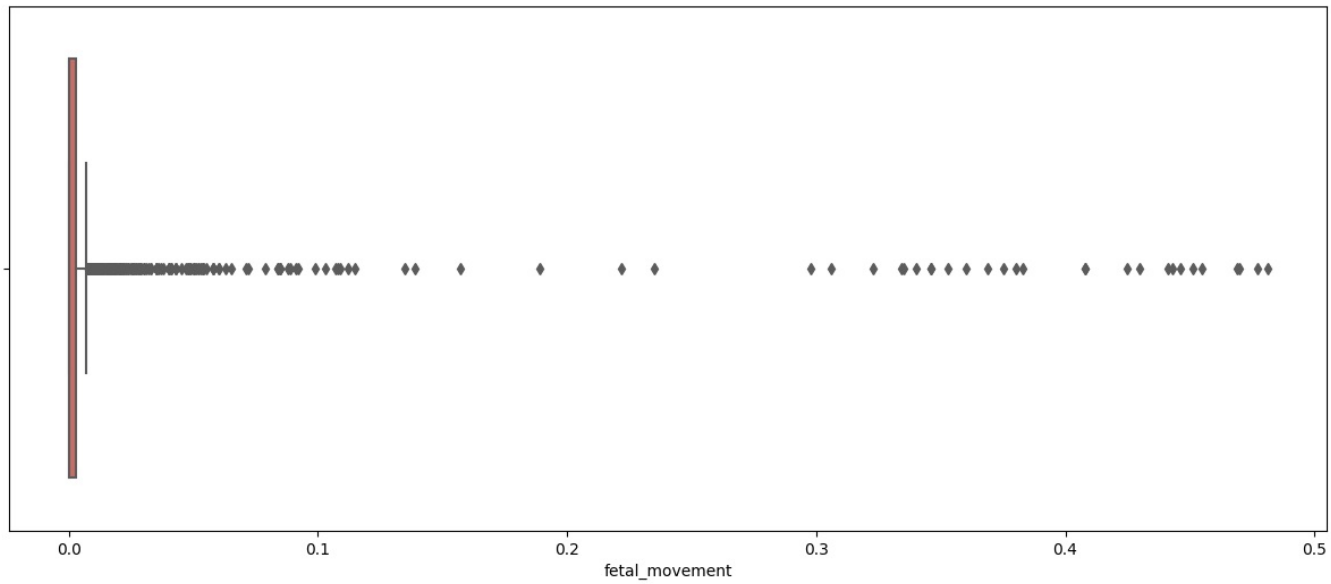


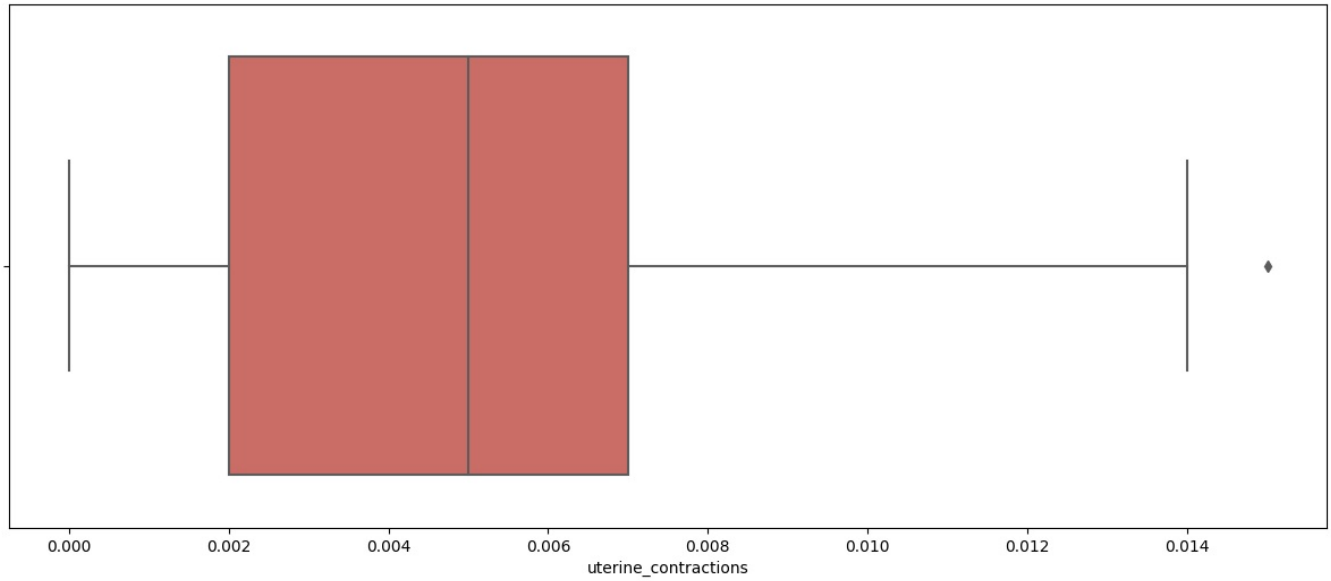


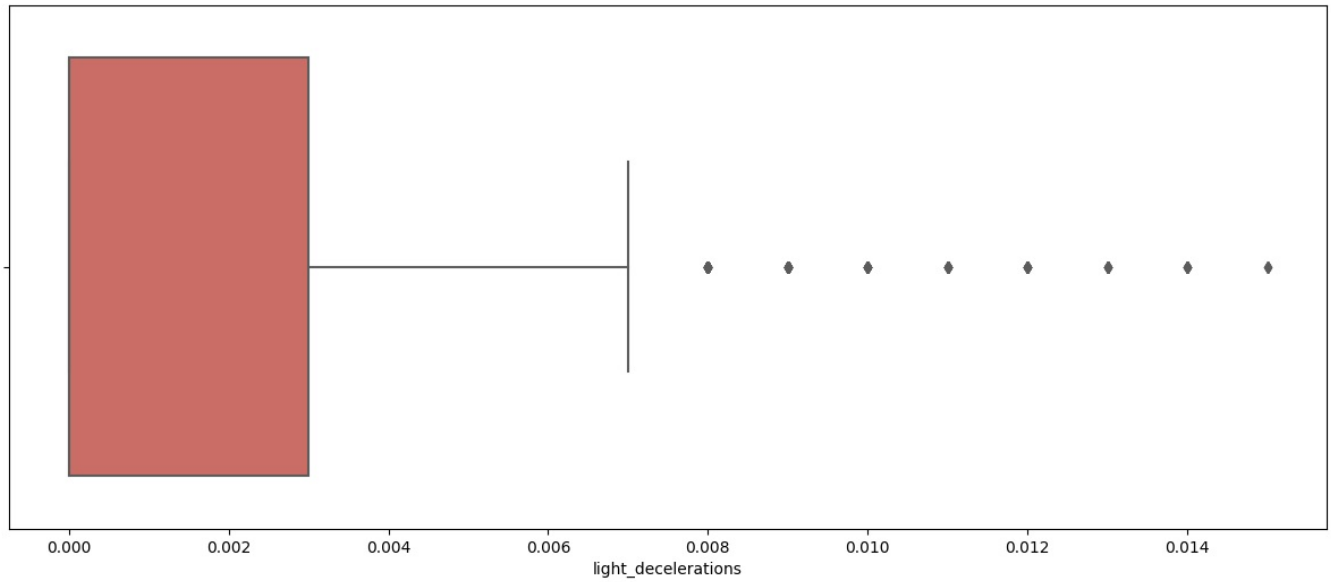
In [19]:
```
for i in df:
    plt.figure(figsize=(15,6))
    sns.boxplot( x=df[i], data = df, palette = 'hls')
    plt.show()
```

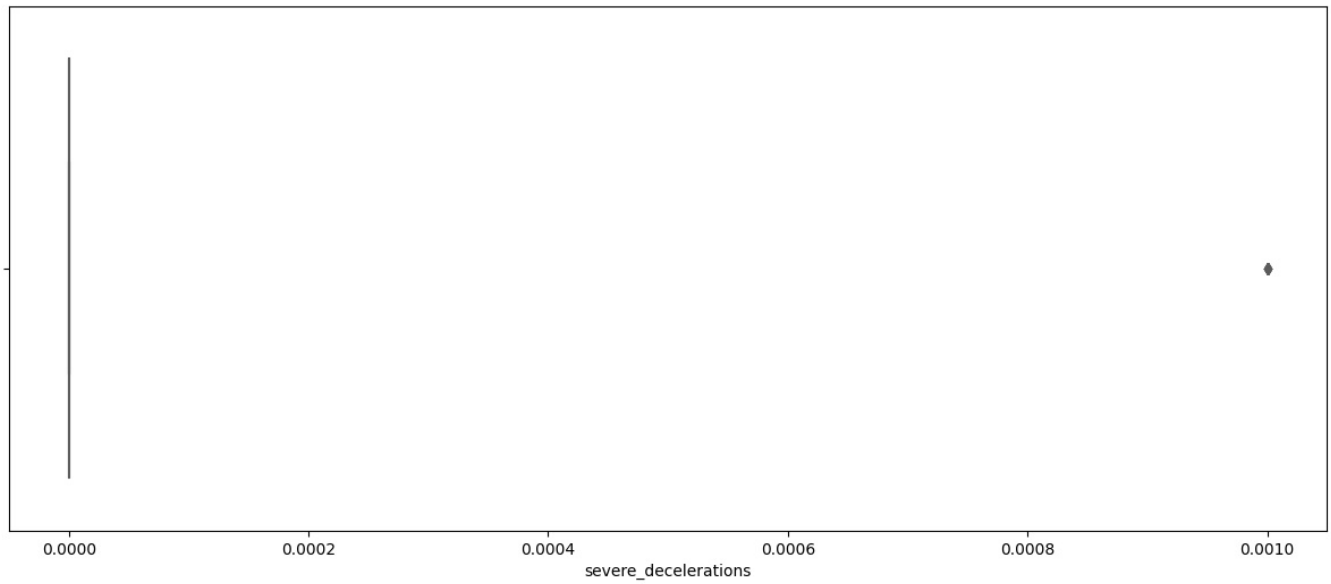

110
120
130
140
150
160
baseline value

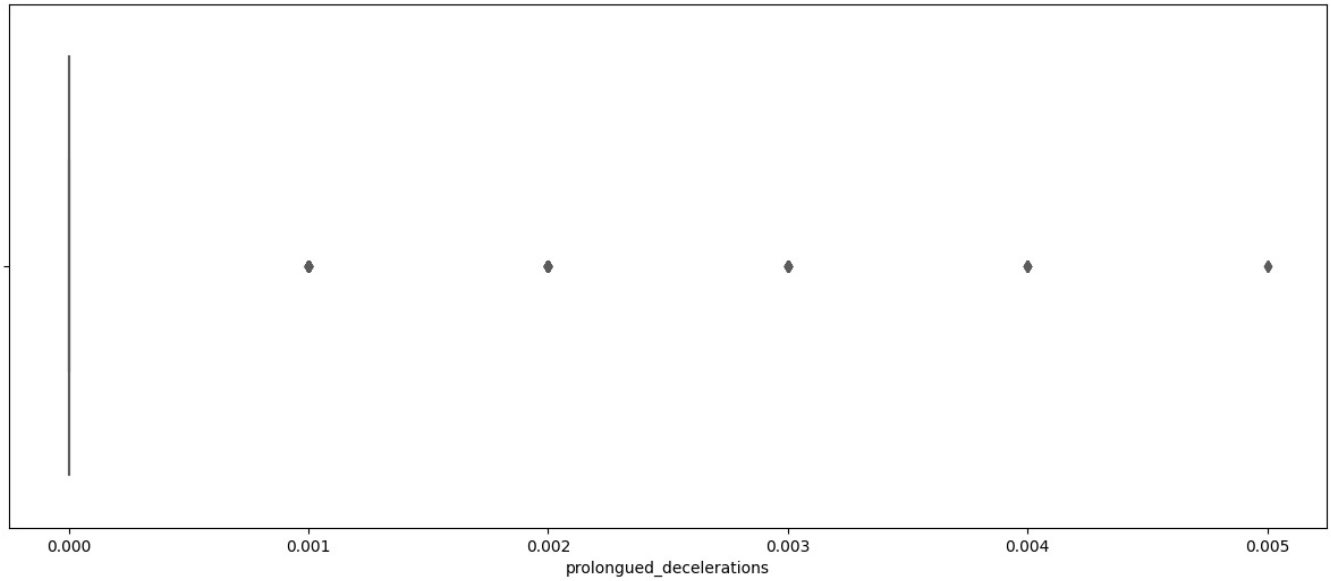

0.0000
0.0025
0.0050
0.0075
0.0100
0.0125
0.0150
0.0175
accelerations

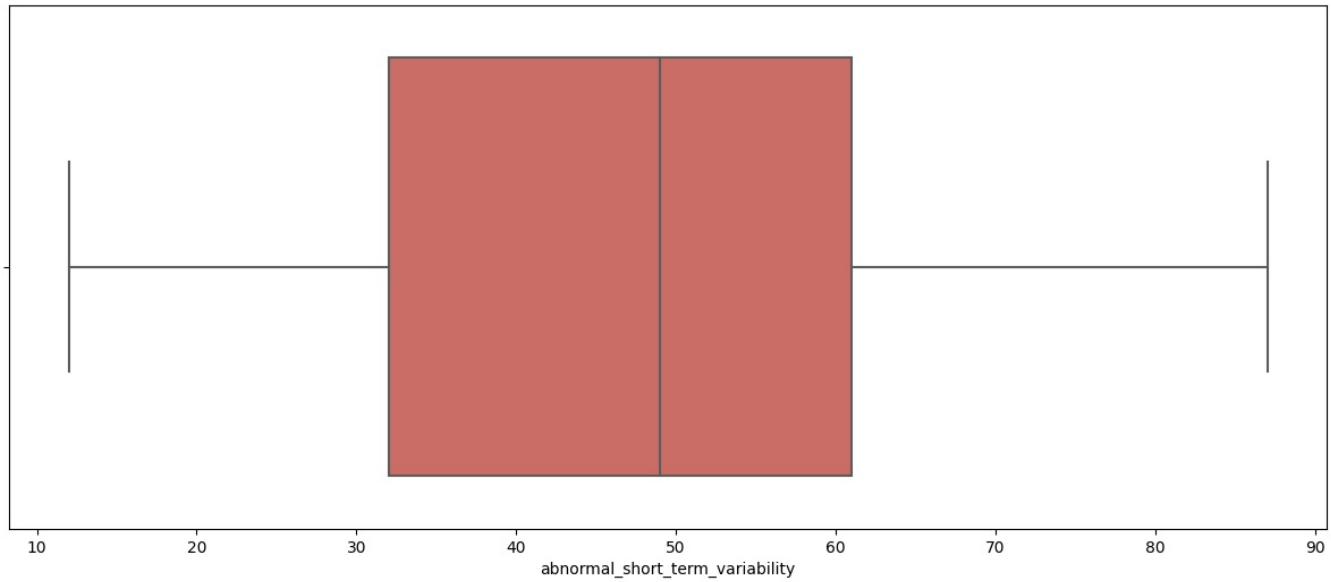

0.0
0.1
0.2
0.3
0.4
0.5
fetal_movement

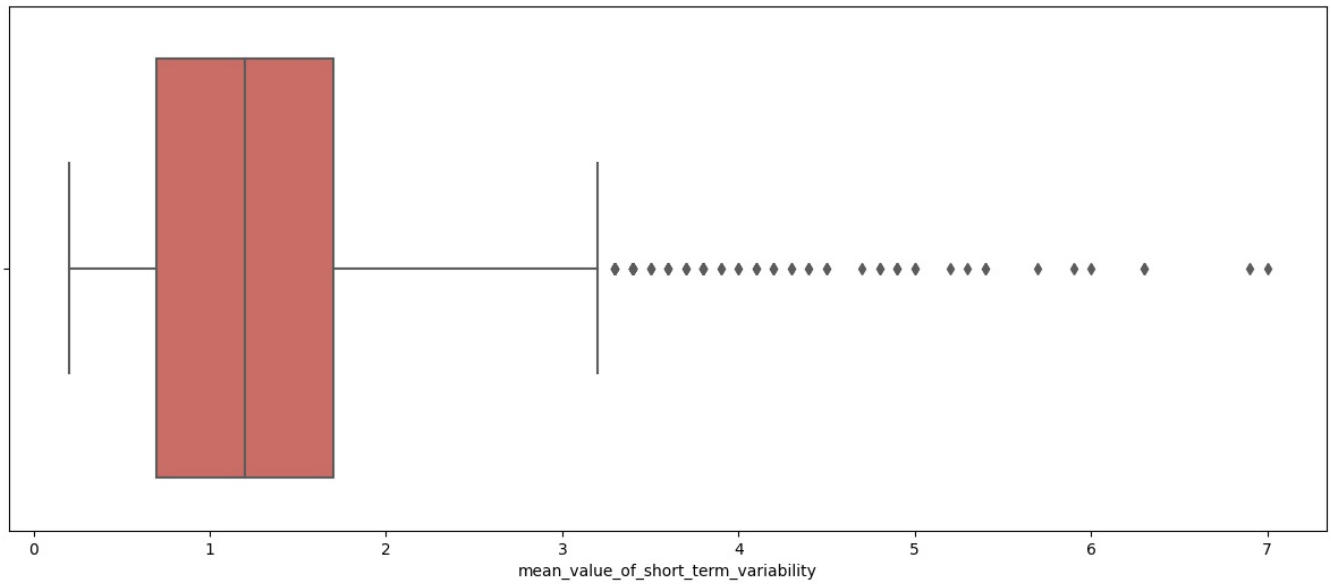

0.000
0.002
0.004
0.006
0.008
0.010
0.012
0.014
uterine_contractions

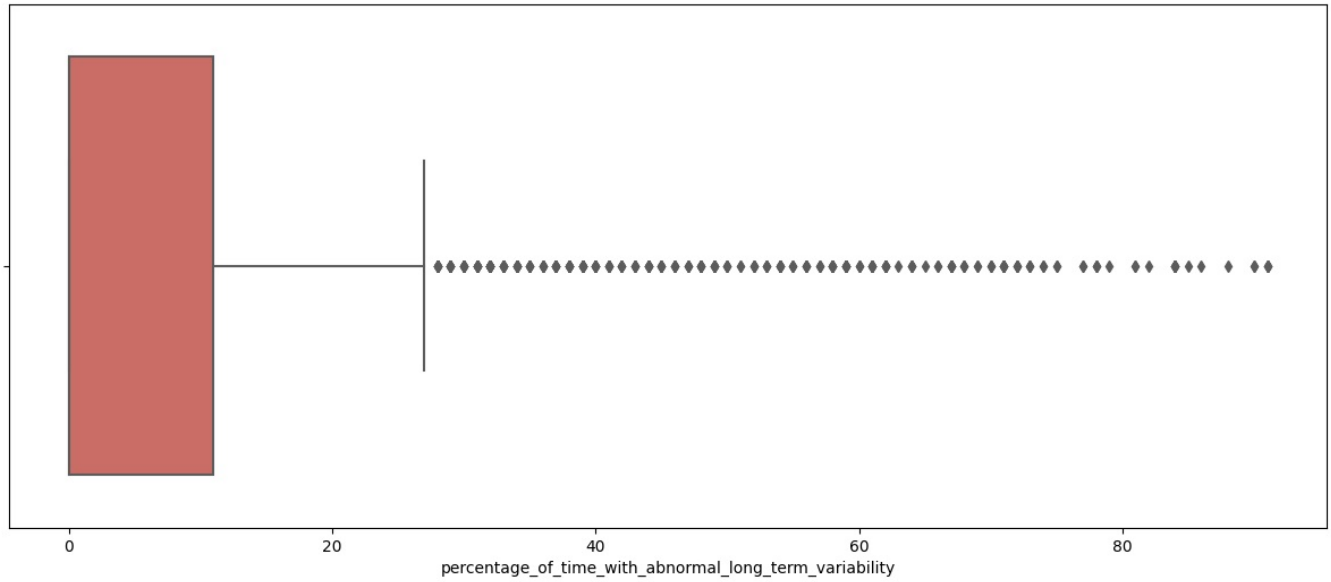

0.000
0.002
0.004
0.006
0.008
0.010
0.012
0.014
light_decelerations

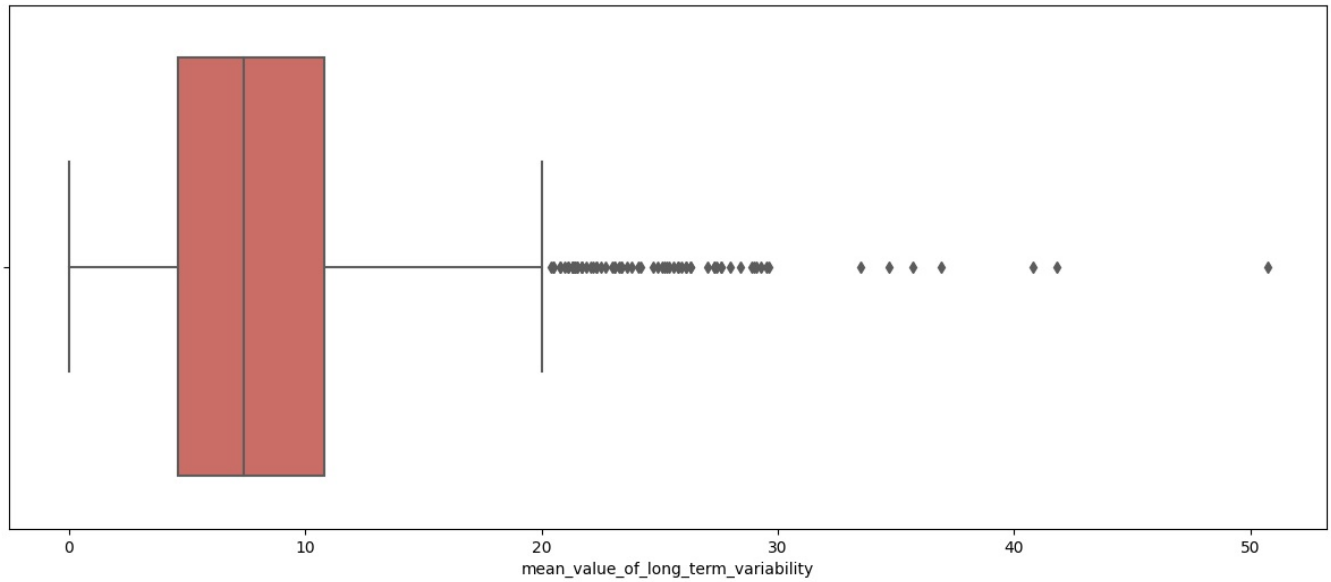

0.0000
0.0002
0.0004
0.0006
0.0008
0.0010
severe_decelerations

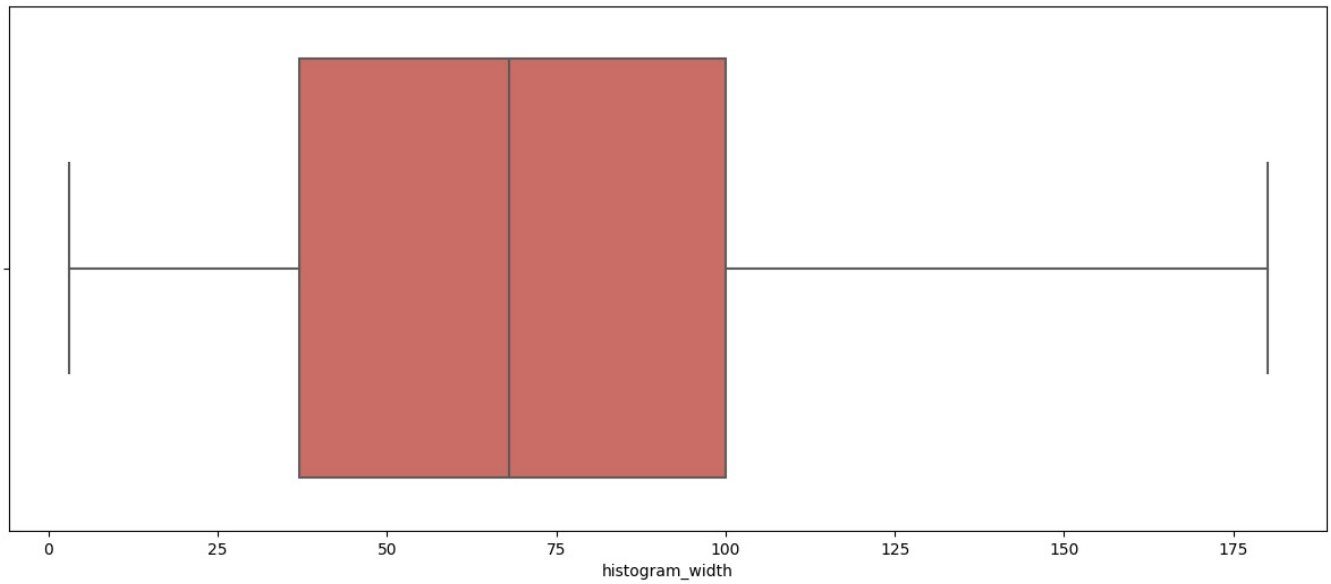

0.000
0.001
0.002
0.003
0.004
0.005
prolongued_decelerations

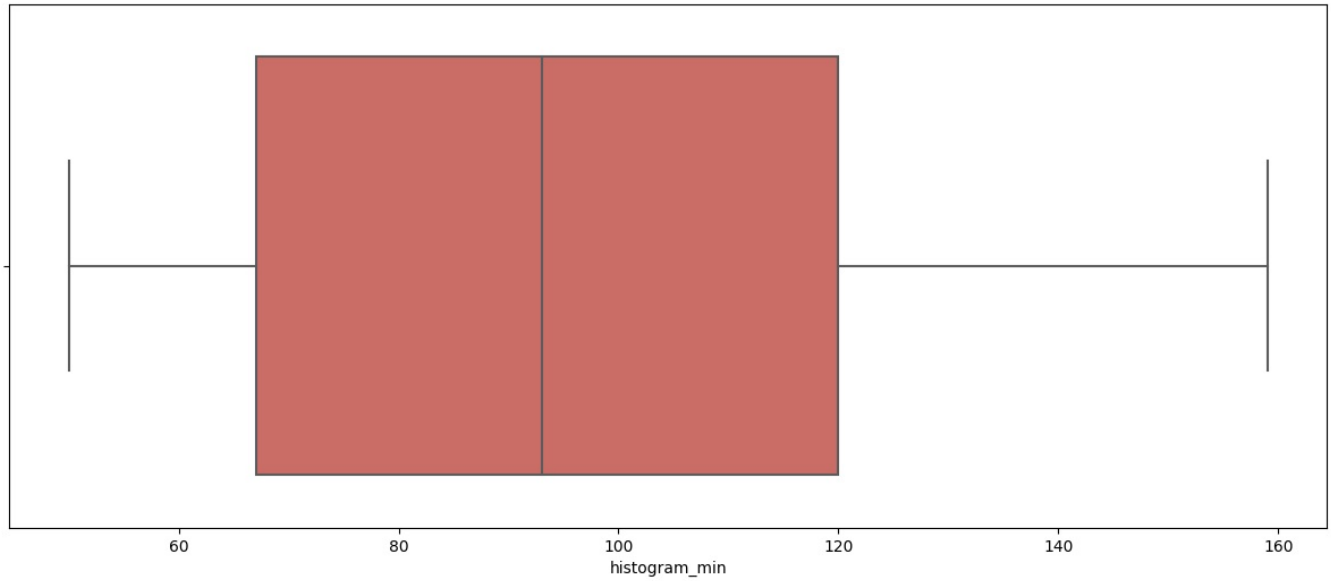

10
20
30
40
50
60
70
80
90
abnormal_short_term_variability

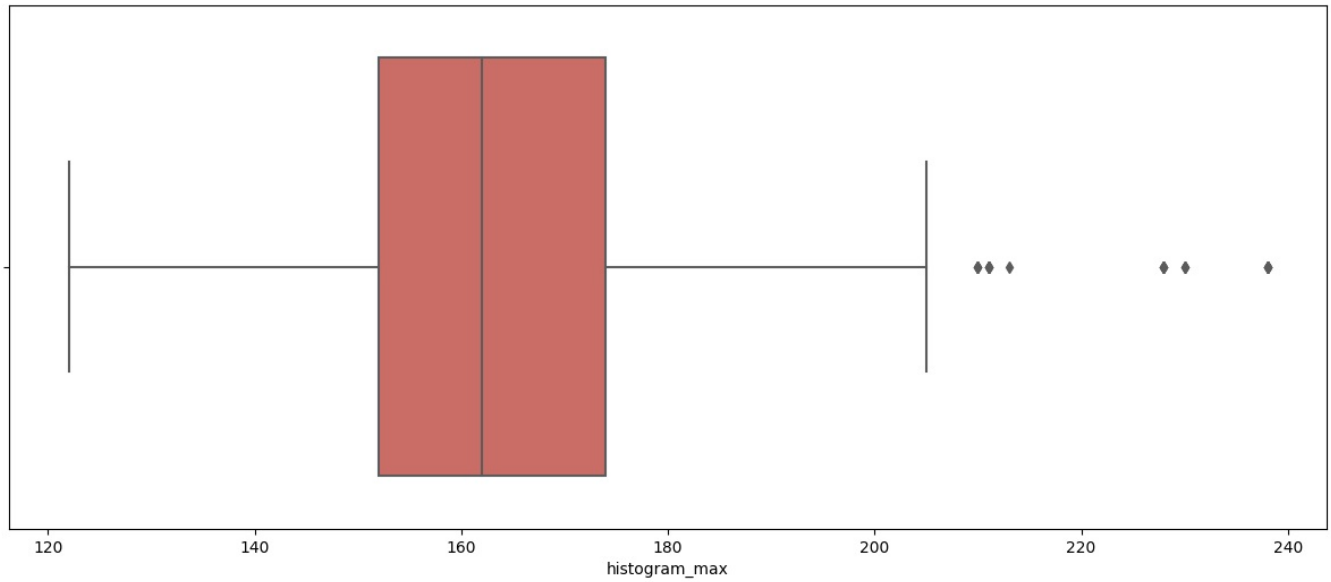

0
1
2
3
4
5
6
7
mean_value_of_short_term_variability

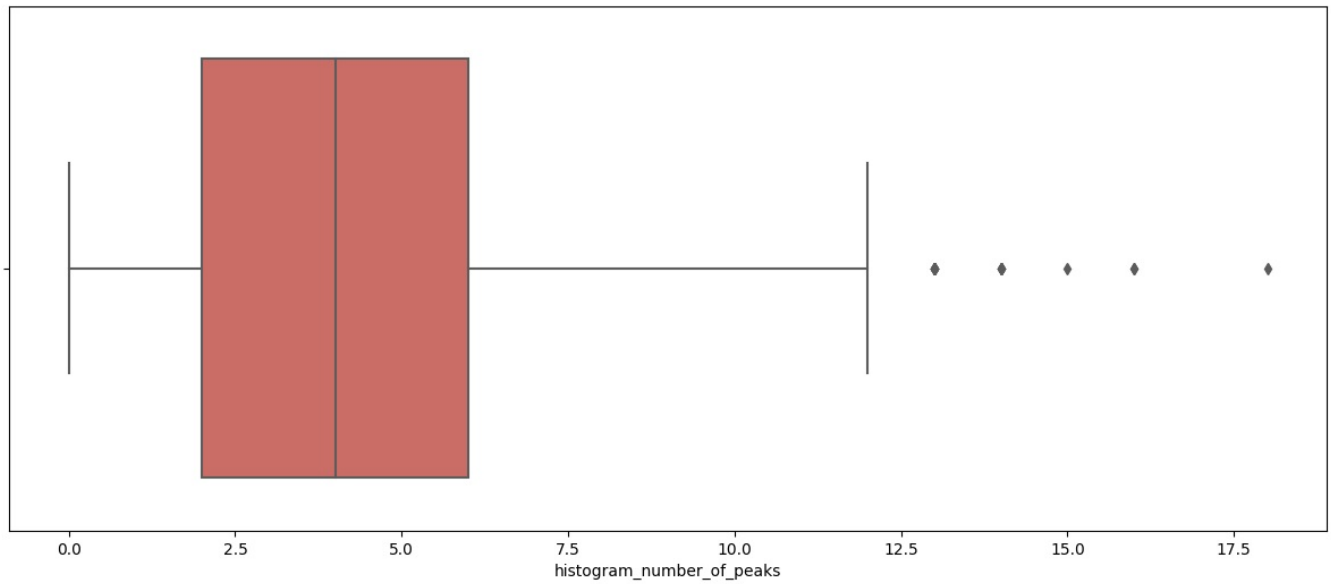

0
20
40
60
80
percentage_of_time_with_abnormal_long_term_variability

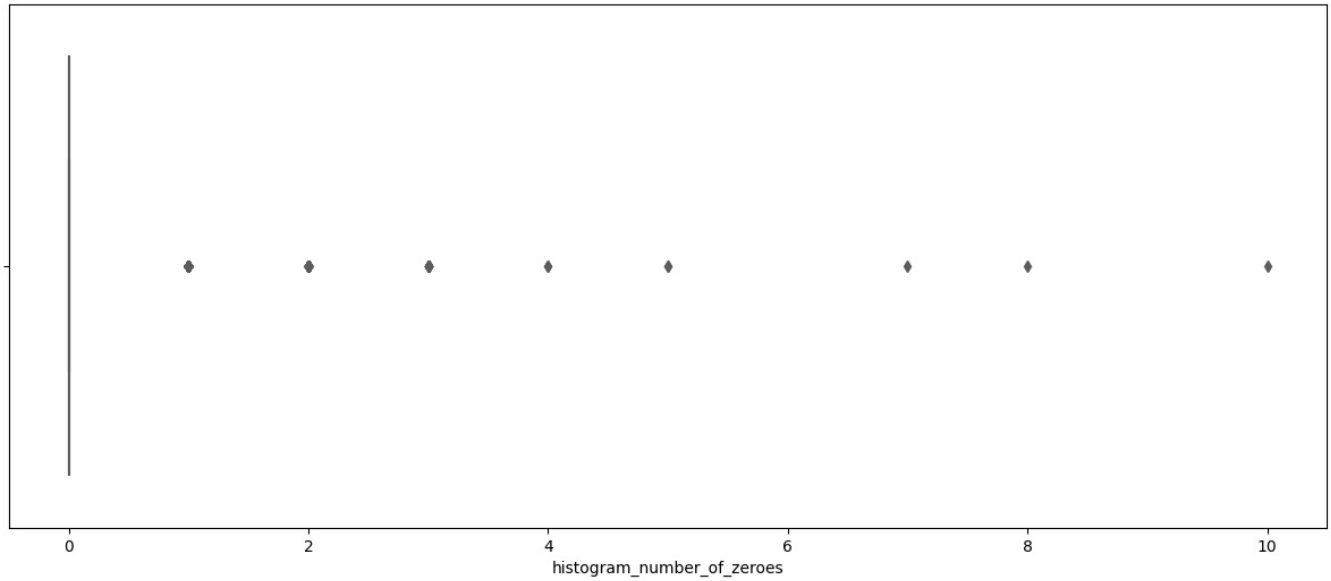

0
10
20
30
40
50
mean_value_of_long_term_variability

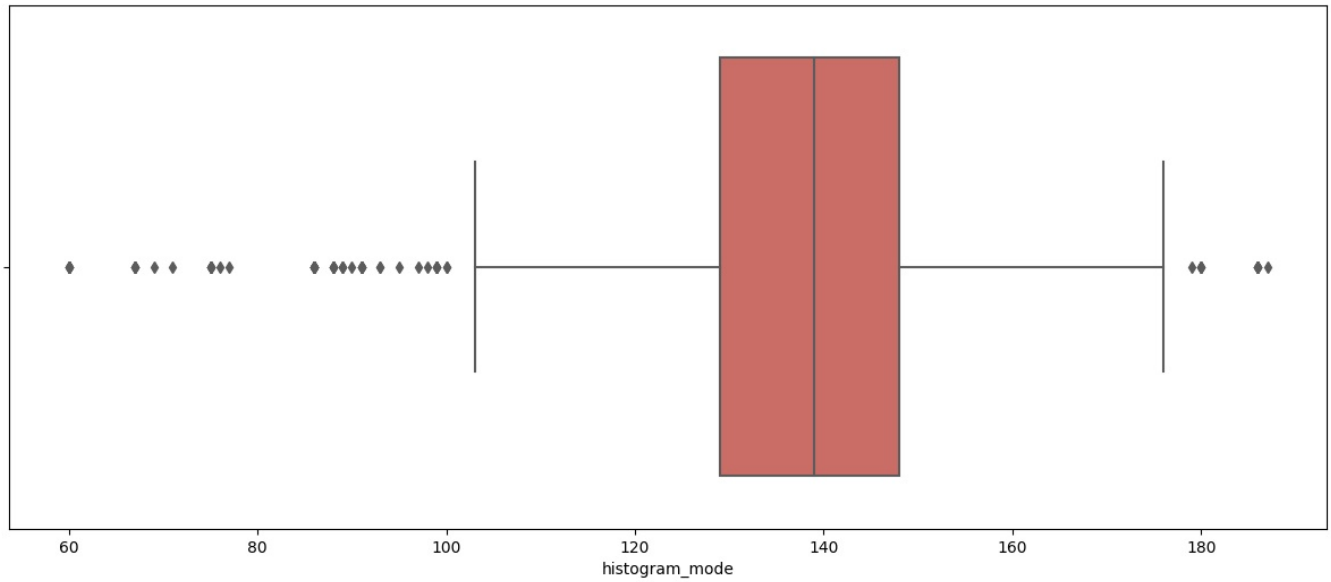

0
25
50
75
100
125
150
175
histogram_width

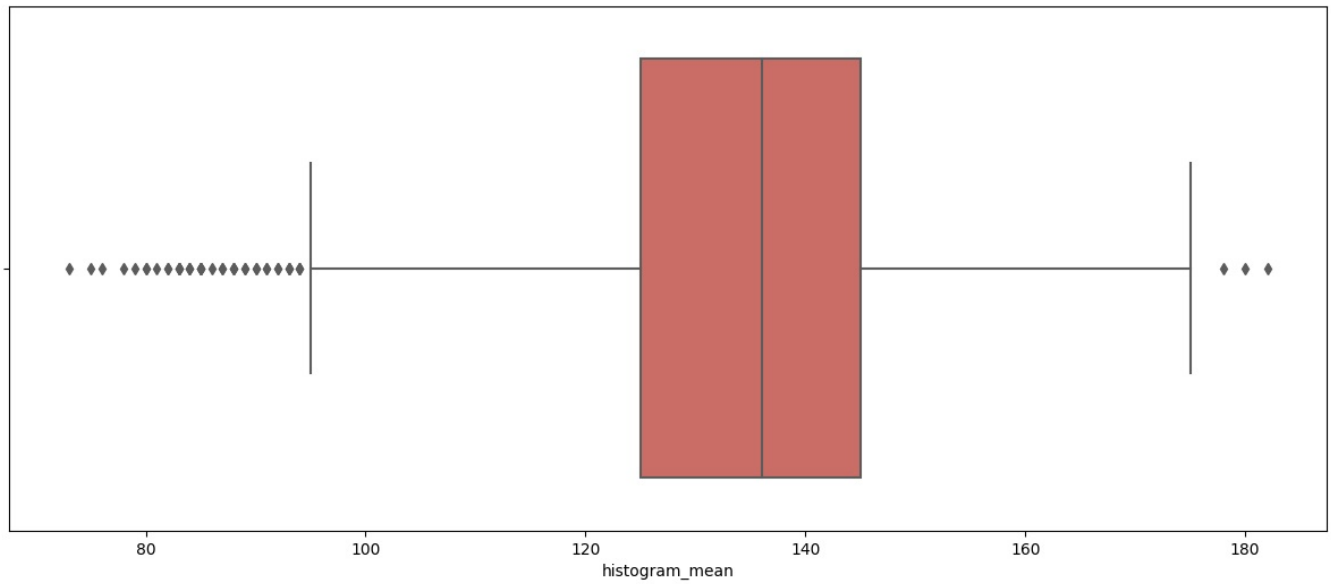

60
80
100
120
140
160
histogram_min

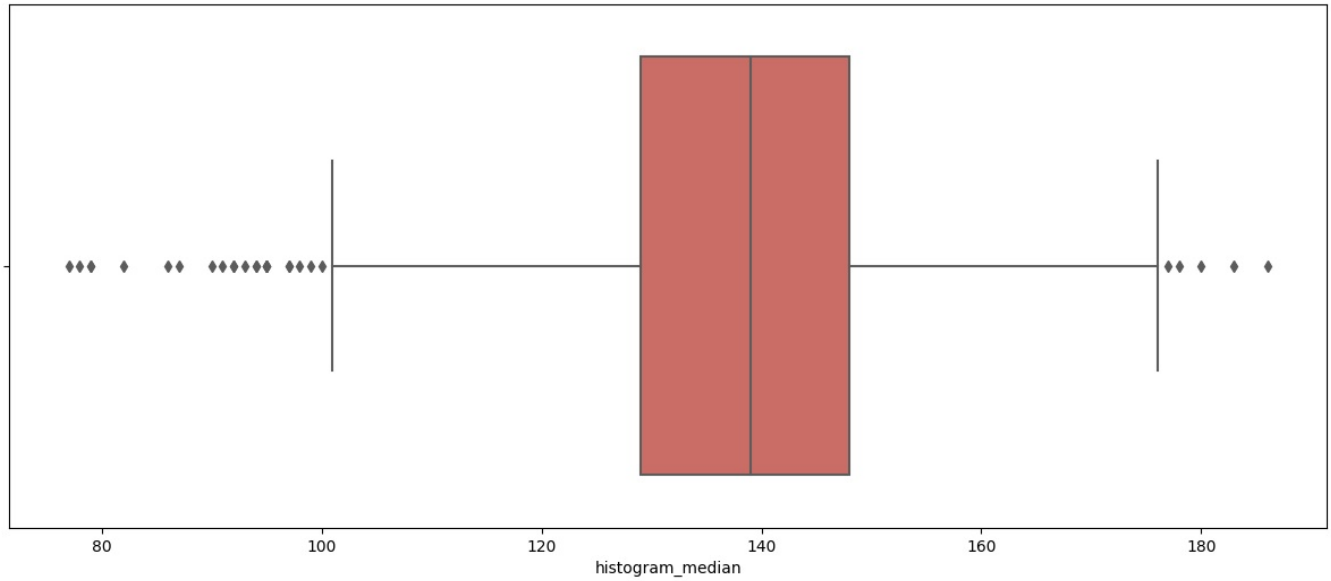

120
140
160
180
200
220
240
histogram_max

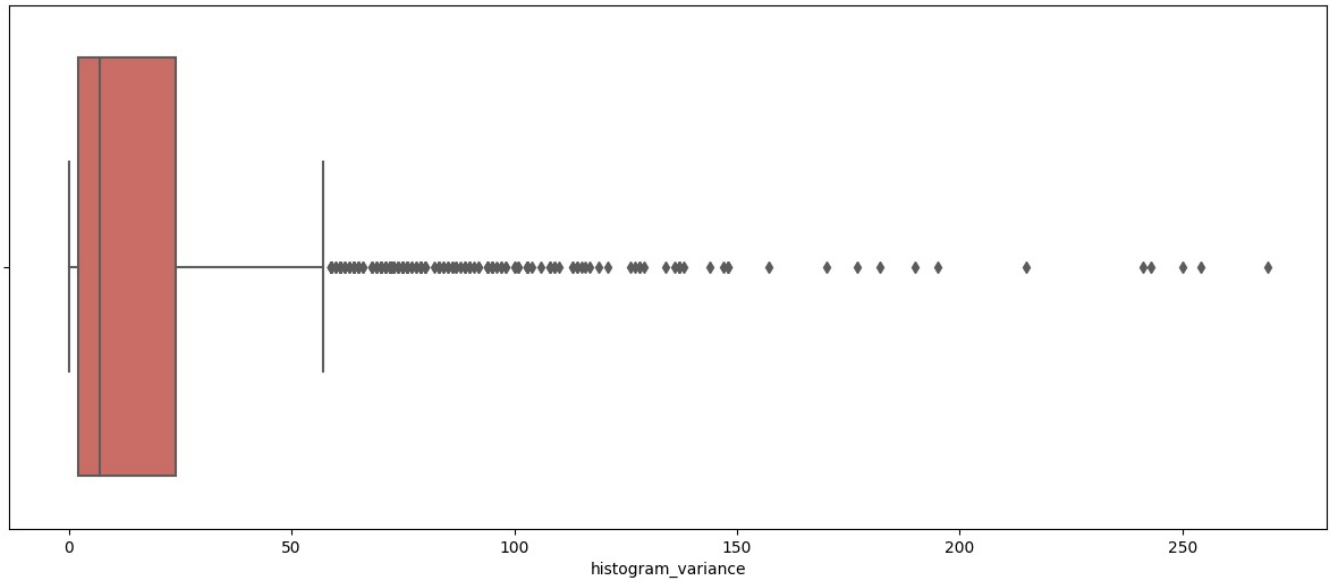

0.0
2.5
5.0
7.5
10.0
12.5
15.0
17.5
histogram_number_of_peaks

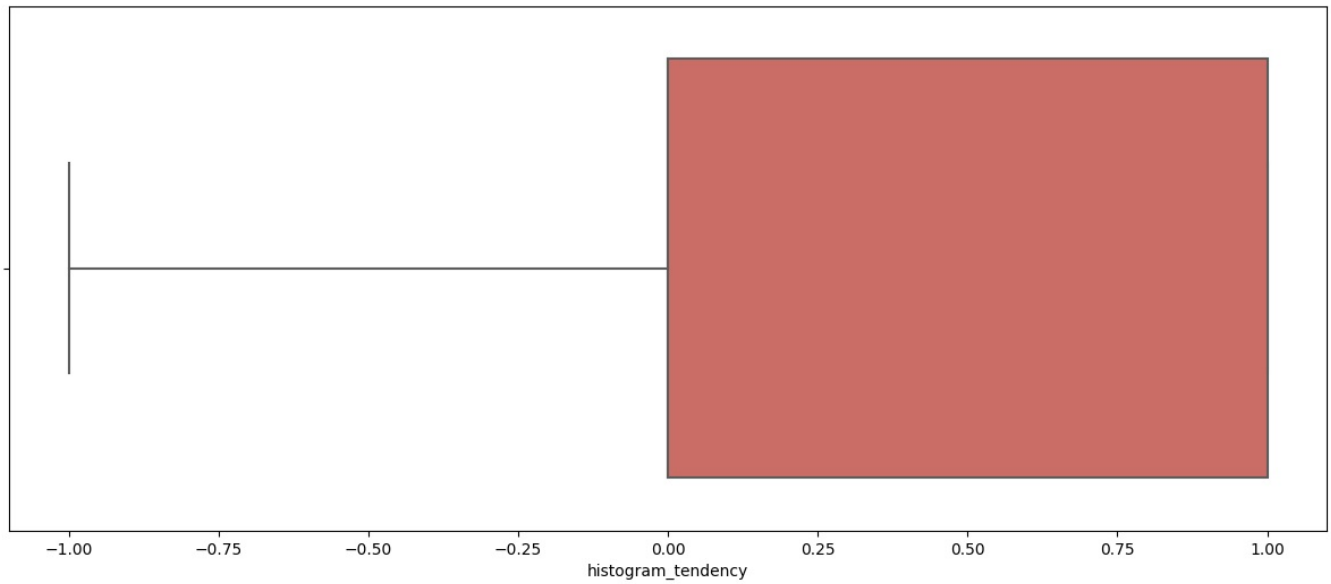

0
2
4
6
8
10
histogram_number_of_zeroes

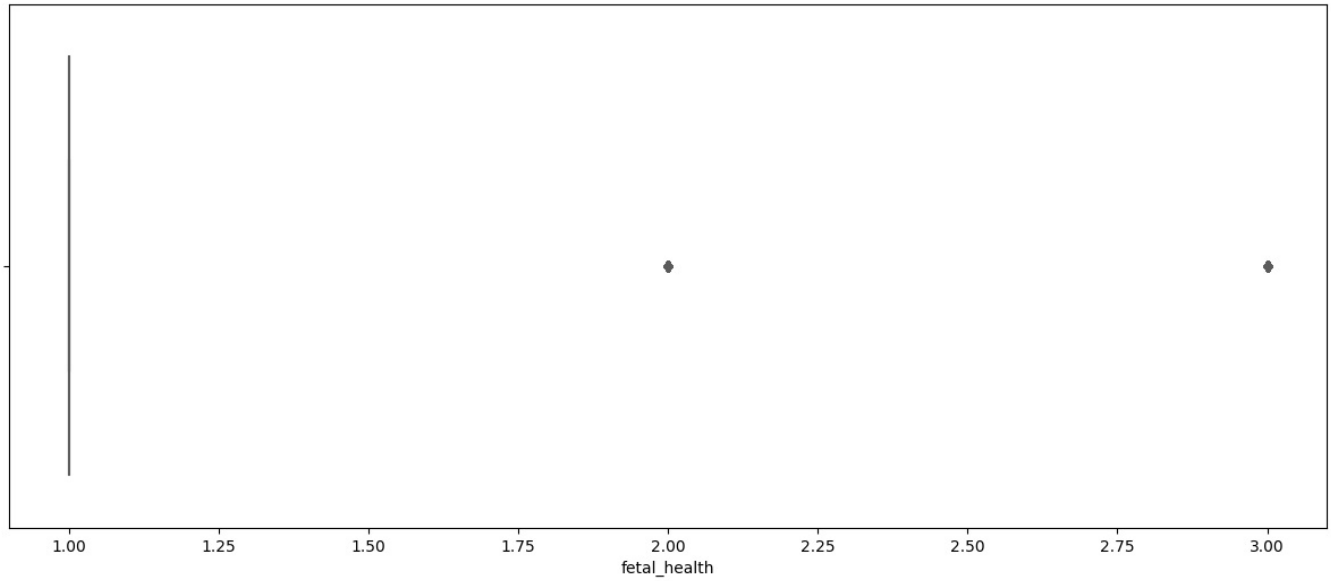

60
80
100
120
140
160
180
histogram_mode


80
100
120
140
160
180
histogram_mean

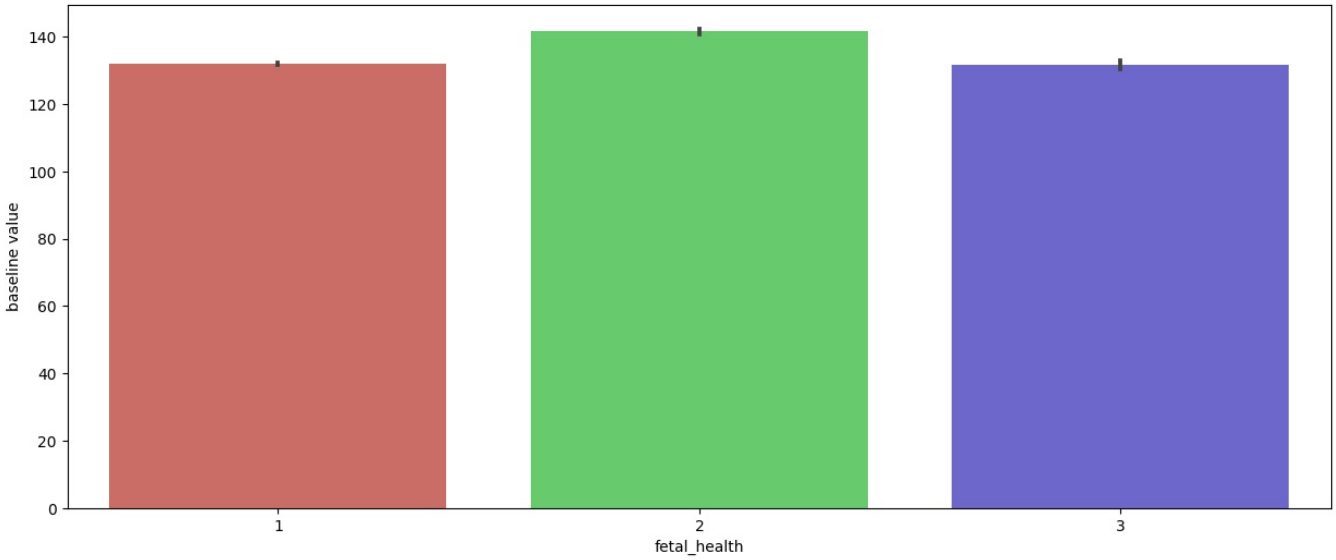

80
100
120
140
160
180
histogram_median

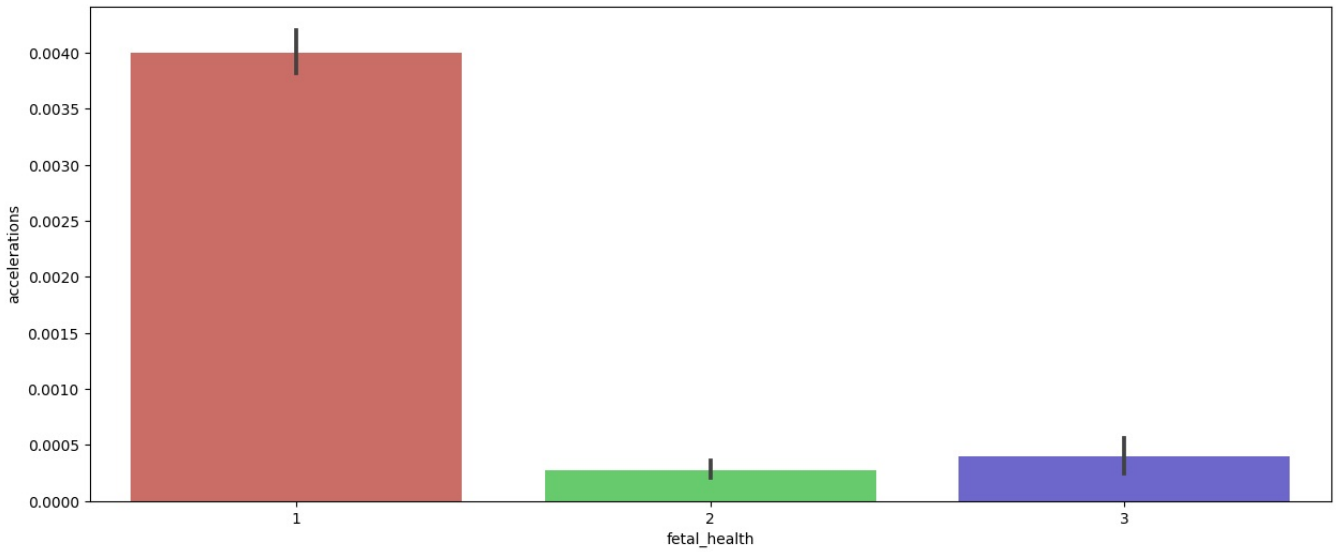

0
50
100
150
200
250
histogram_variance

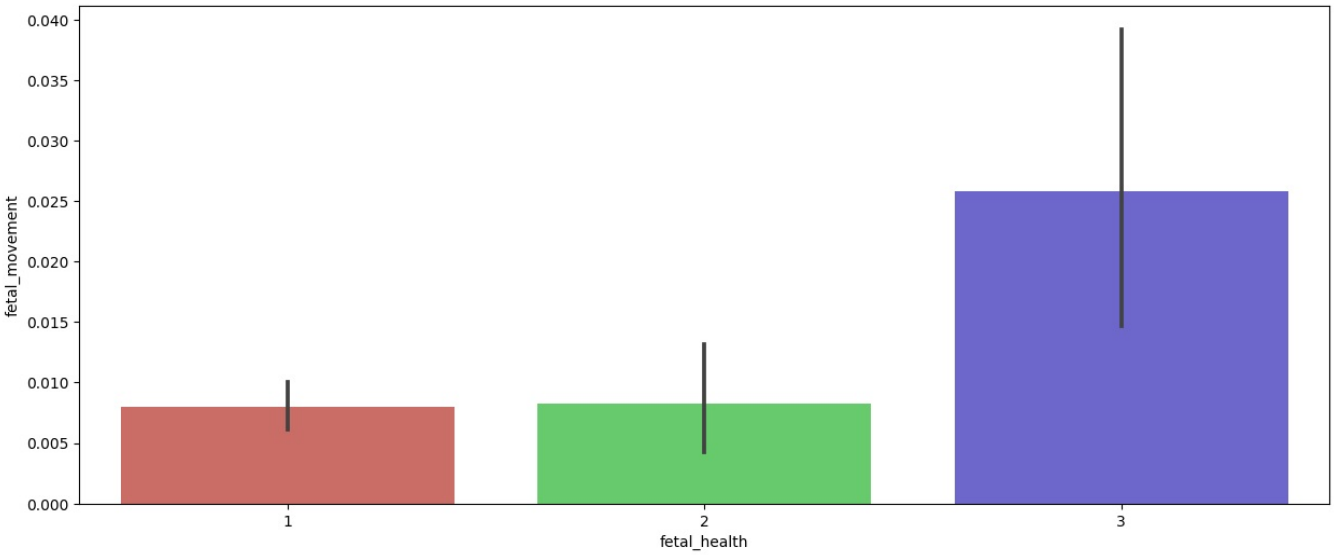

−1.00
−0.75
−0.50
−0.25
0.00
0.25
0.50
0.75
1.00
histogram_tendency

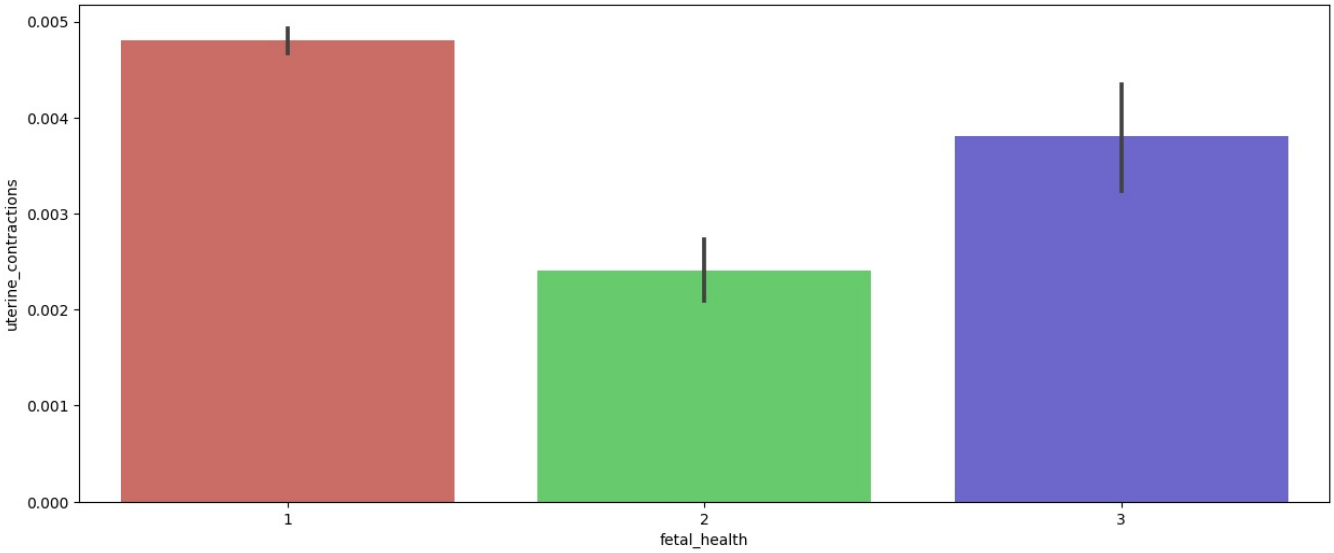


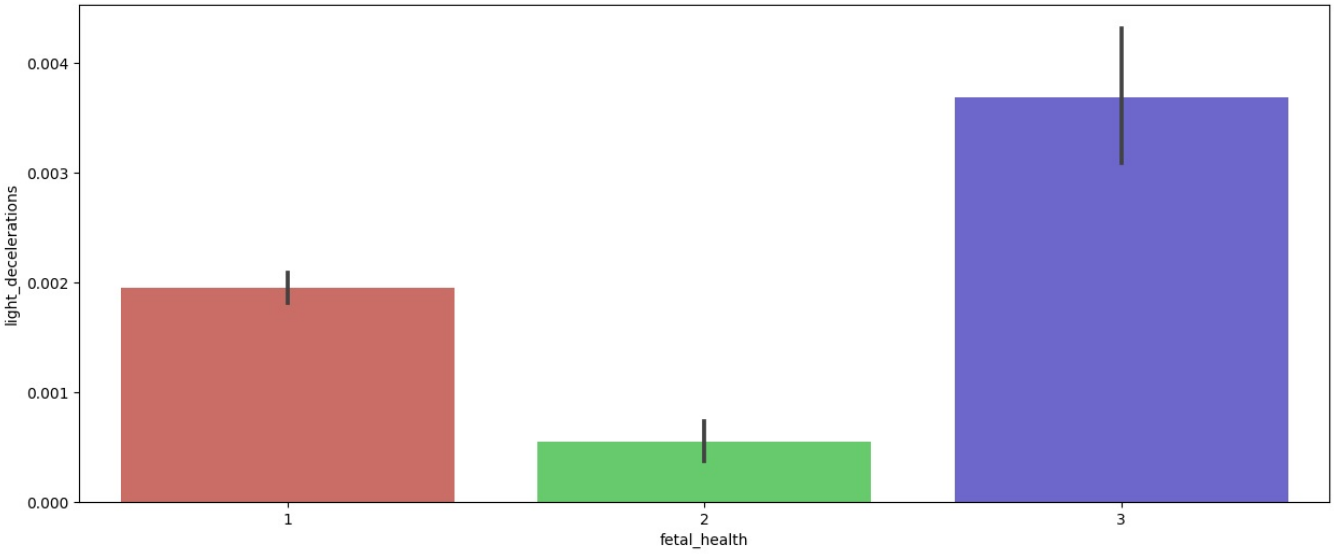
In [20]:
```
for i in df.columns:
    plt.figure(figsize=(15,6))
    sns.barplot(x=df['fetal_health'], y = df[i], palette = 'hls')
    plt.show()
```

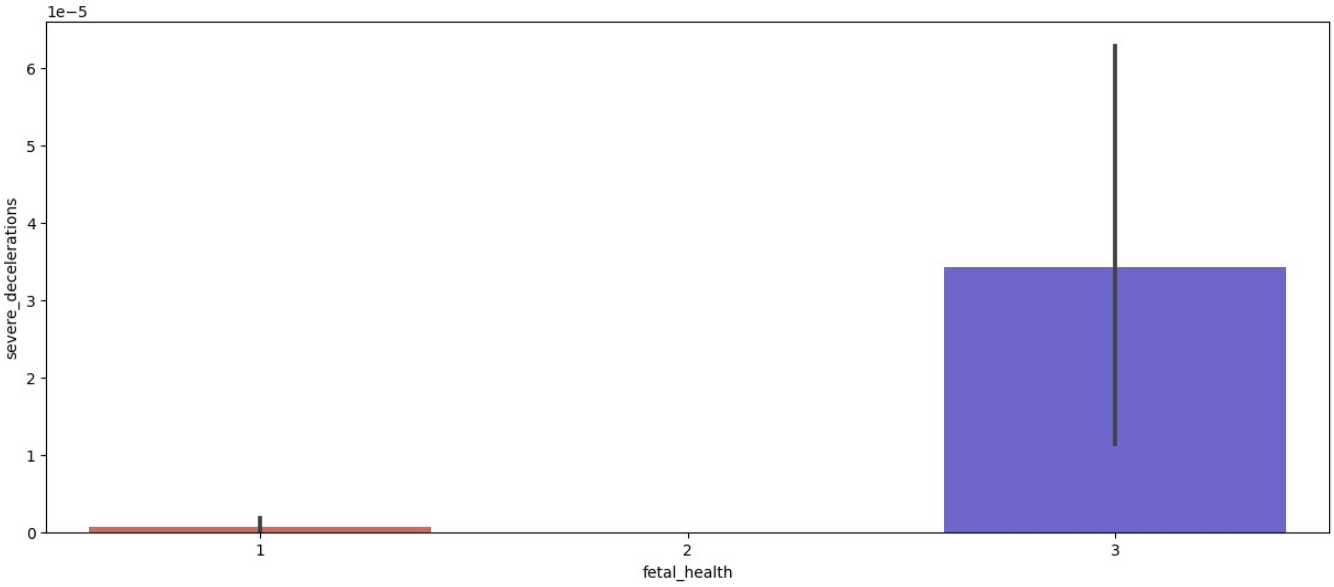


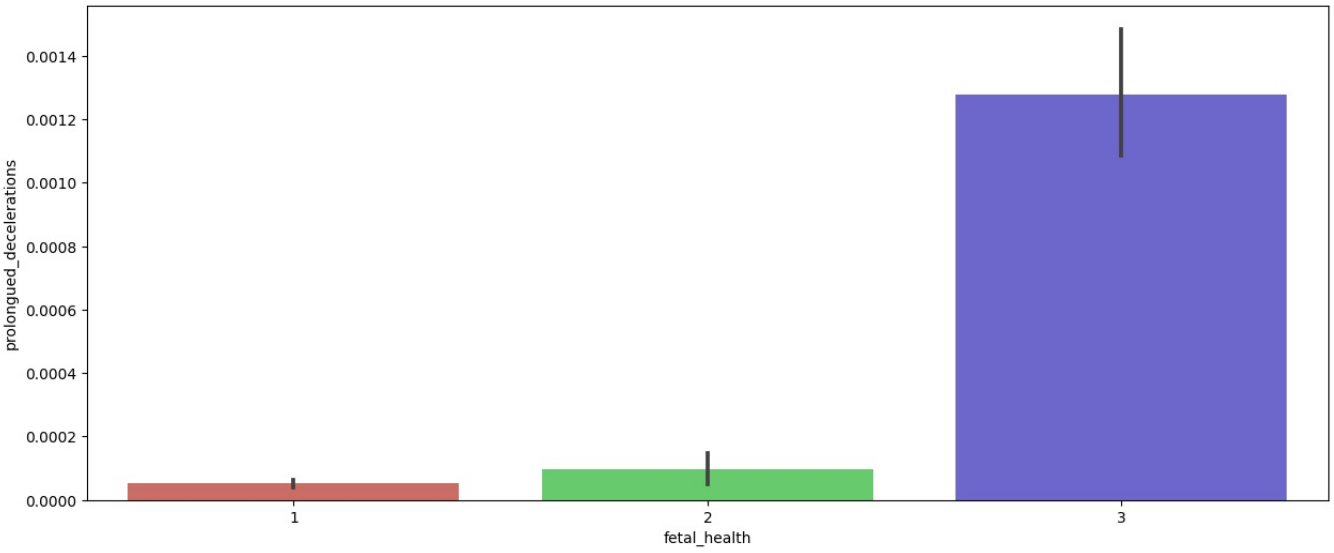


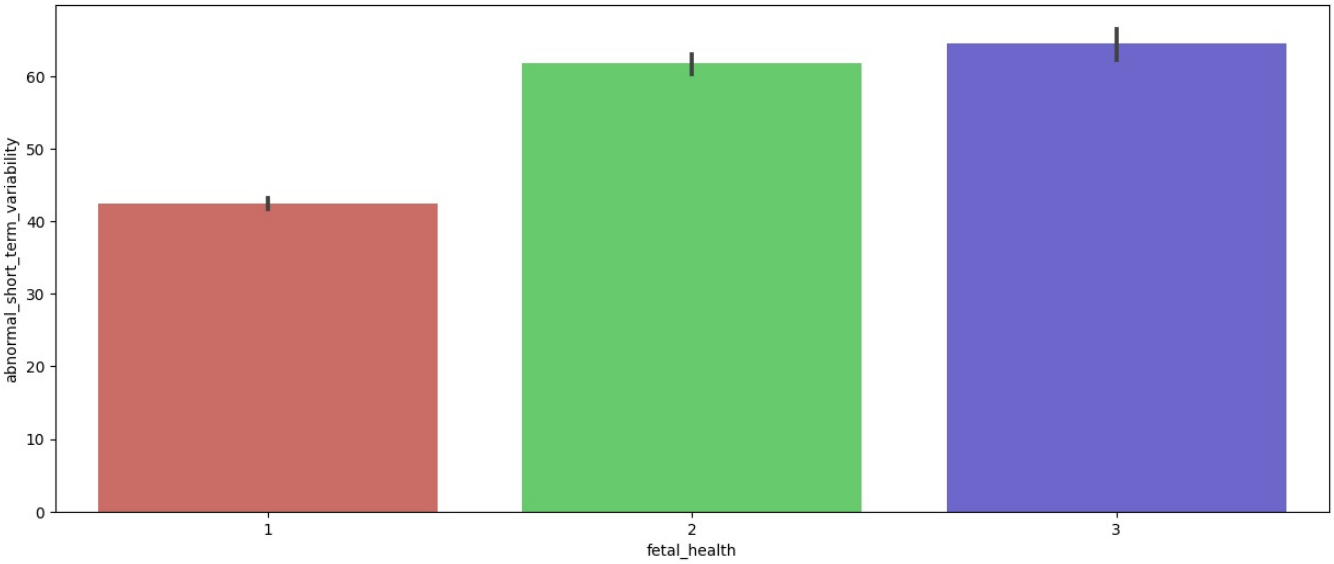

0.040
0.035
0.030
0.025
0.020
0.015
0.010
0.005
0.000
fetal_movement
1
2
3
fetal_health

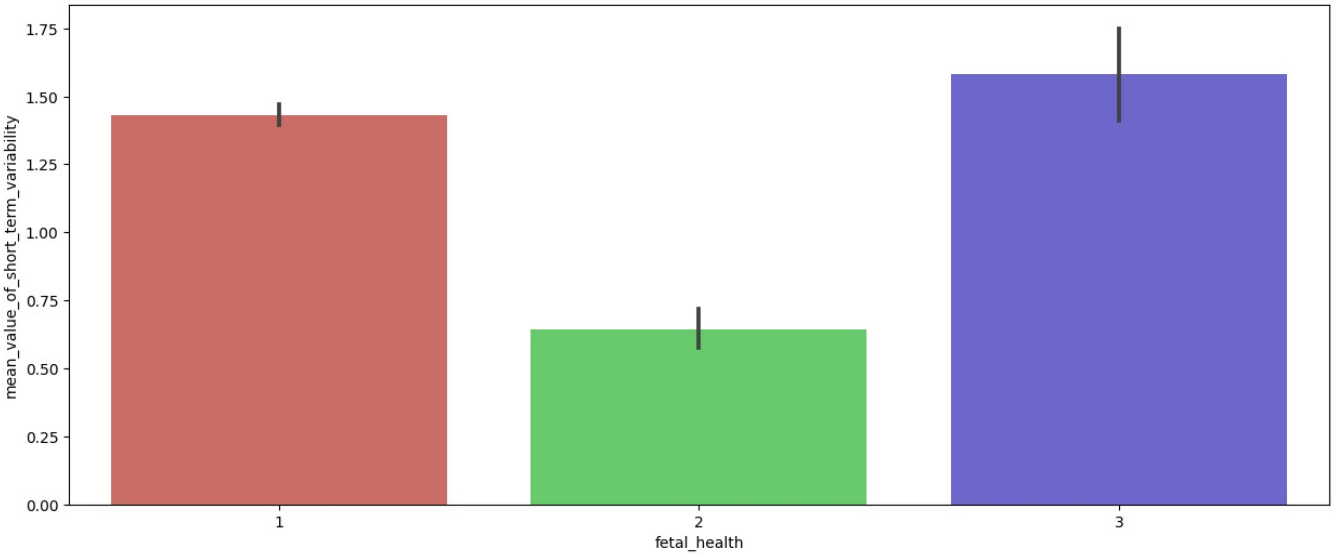

0.005
0.004
0.003
0.002
0.001
0.000
uterine_contractions
1
2
3
fetal_health

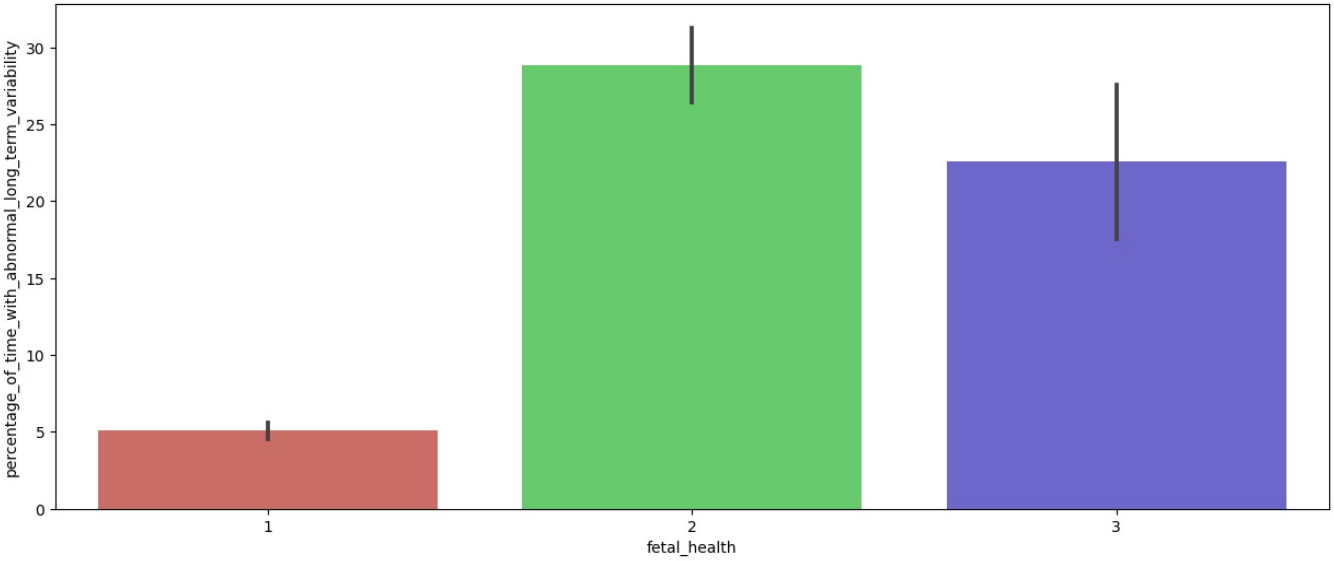

0.004
0.003
0.002
0.001
0.000
light_decelerations
1
2
3
fetal_health

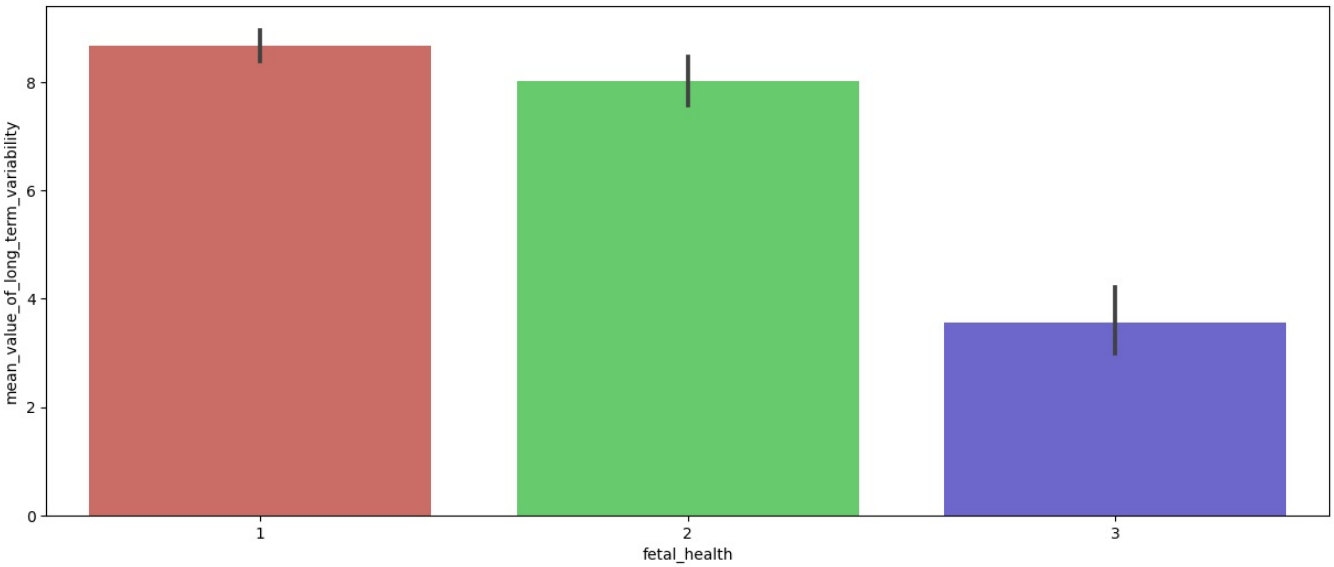

1e−5
severe_decelerations
0
1
2
3
4
5
6
1
2
3
fetal_health

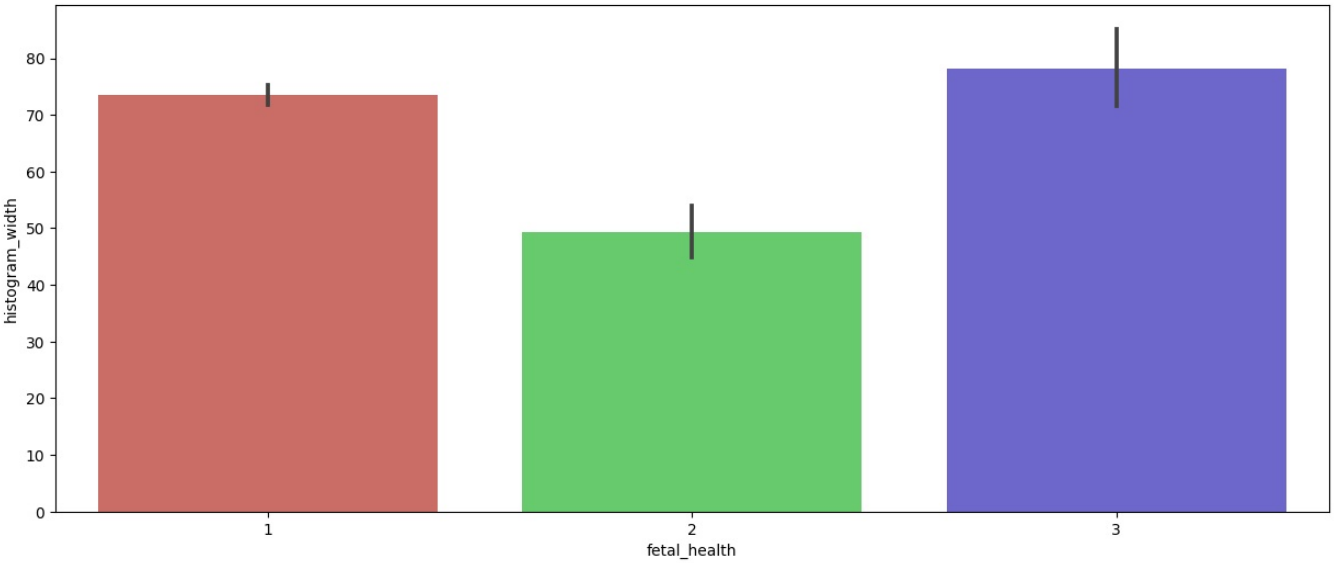

prolongued_decelerations
0.0000
0.0002
0.0004
0.0006
0.0008
0.0010
0.0012
0.0014
1
2
3
fetal_health

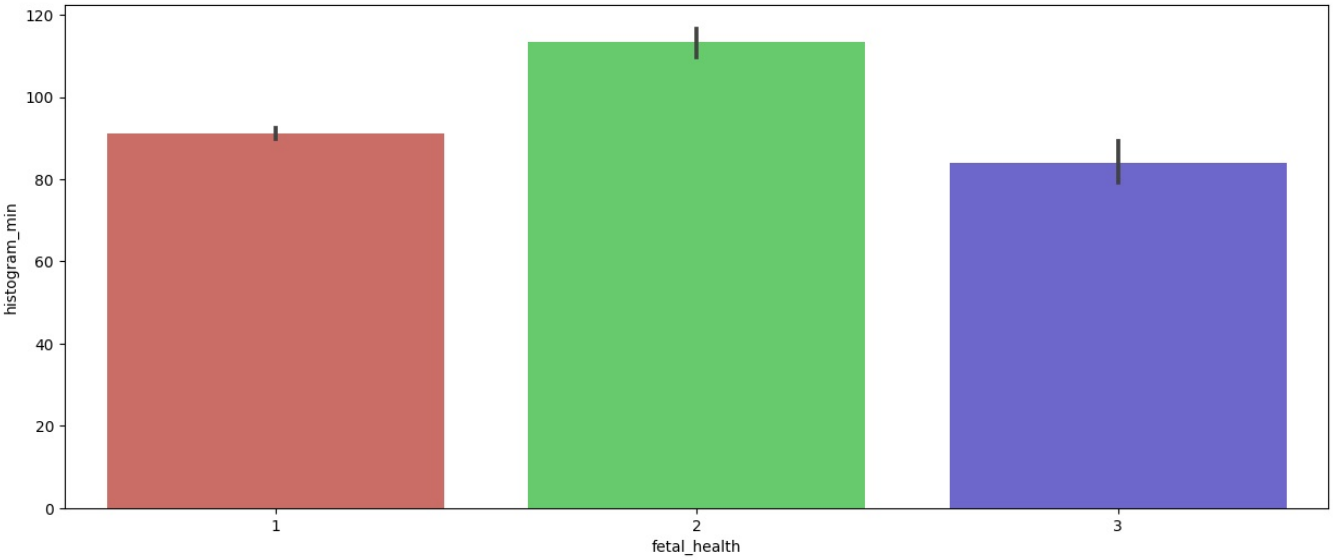

abnormal_short_term_variability
0
10
20
30
40
50
60
1
2
3
fetal_health

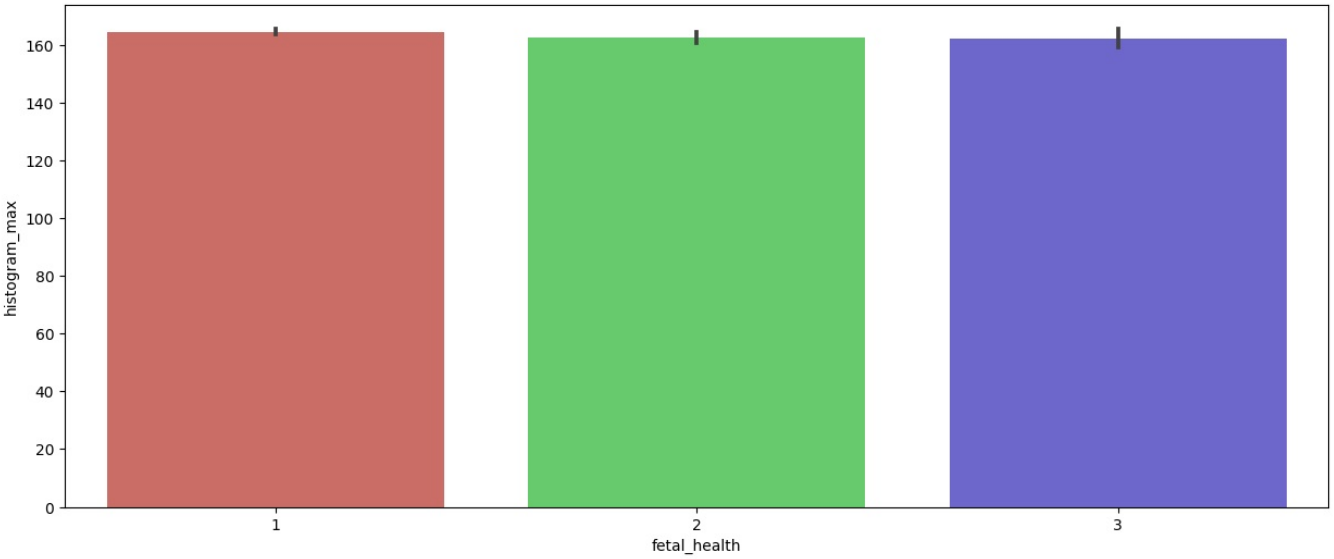

mean_value_of_short_term_variability
1.75
1.50
1.25
1.00
0.75
0.50
0.25
0.00
1
2
3
fetal_health

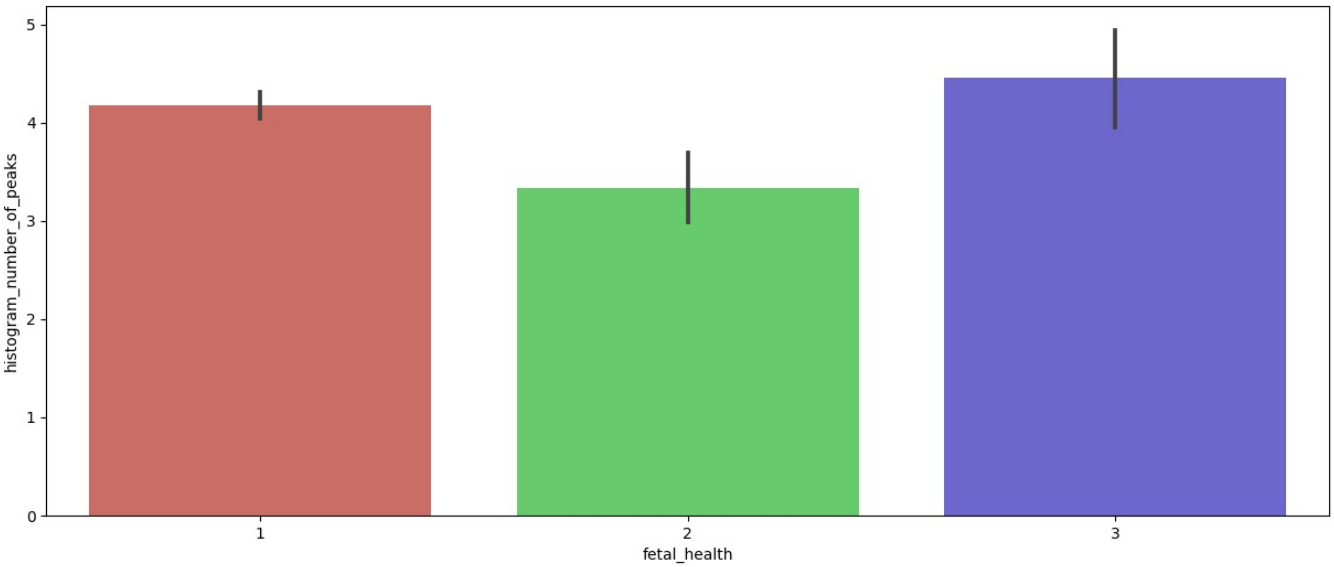

percentage_of_time_with_abnormal_long_term_variability
30
25
20
15
10
5
0
1
2
3
fetal_health

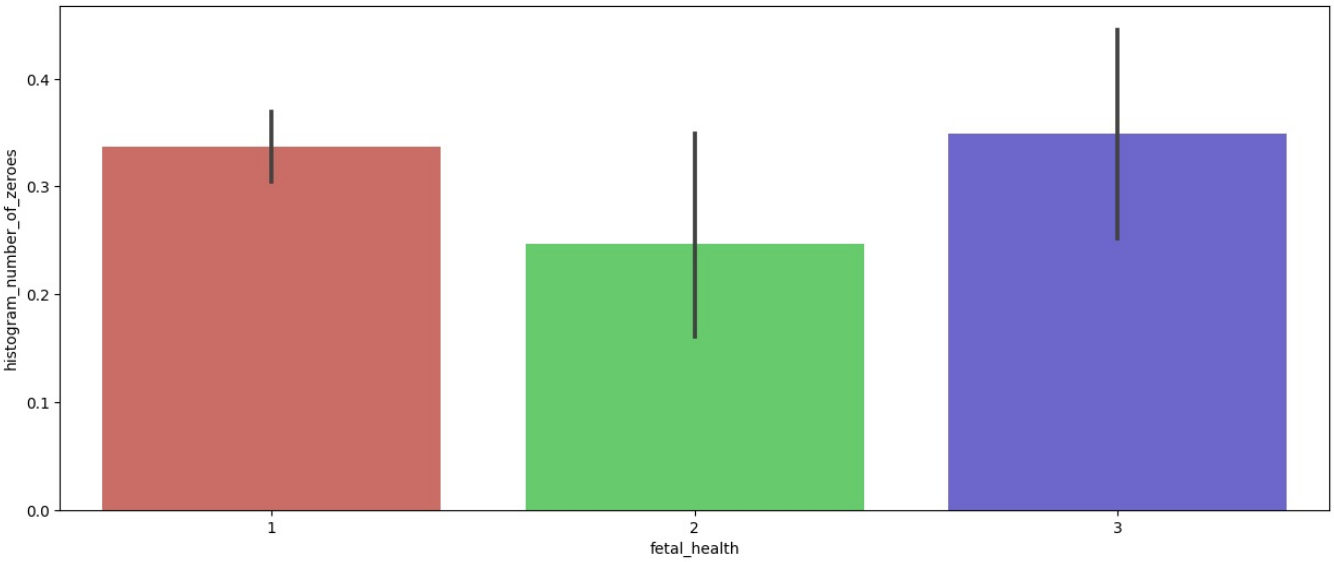

mean_value_of_long_term_variability
8
6
4
2
0
1
2
3
fetal_health

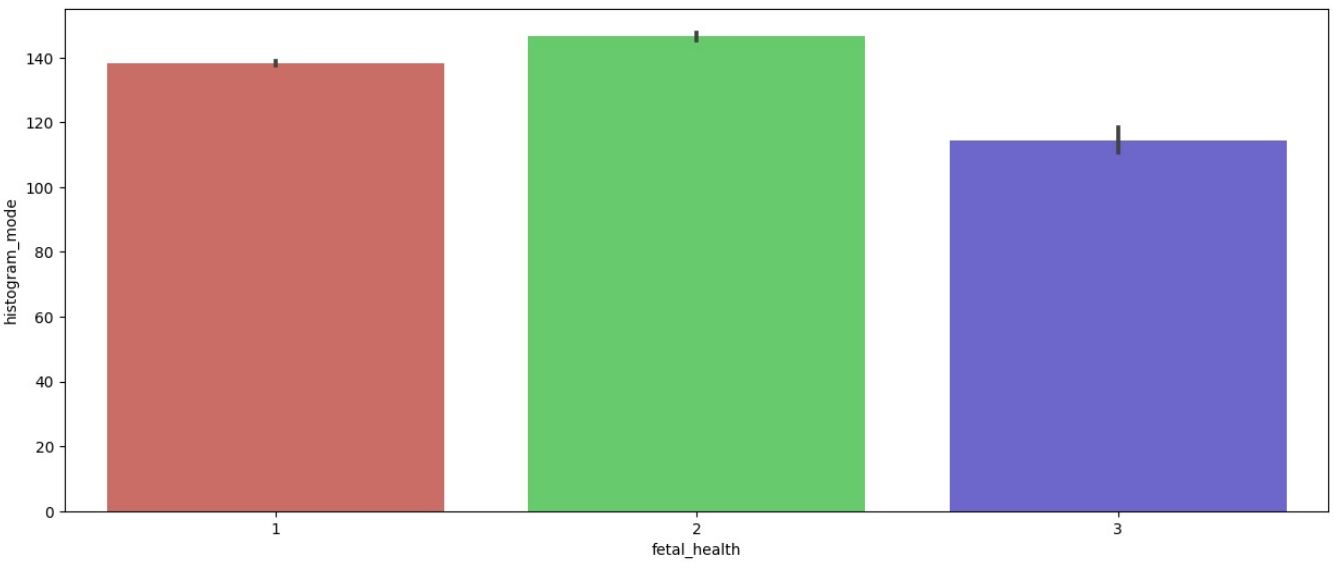

histogram_width
0
10
20
30
40
50
60
70
80
1
2
3
fetal_health

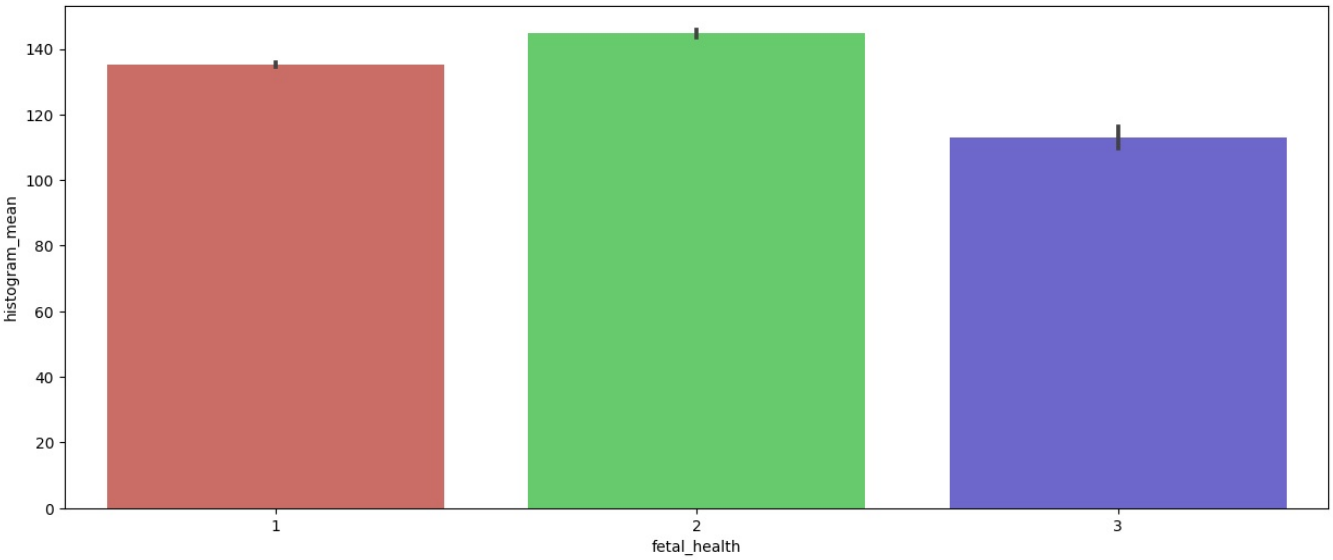

histogram_min
0
20
40
60
80
100
120
1
2
3
fetal_health

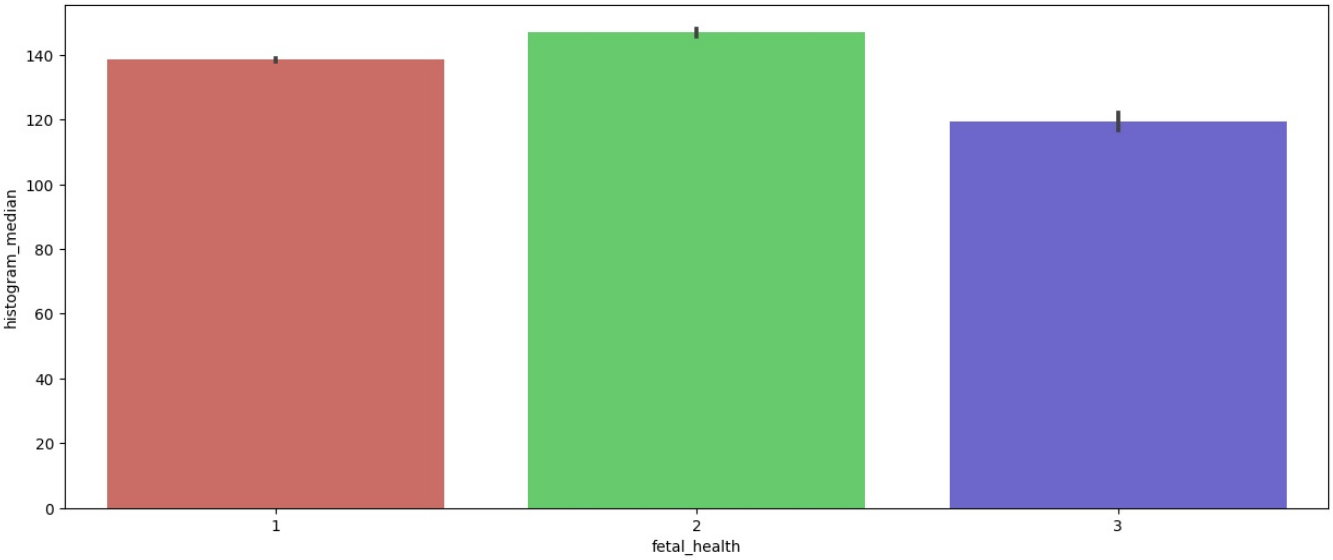

histogram_max
0
20
40
60
80
100
120
140
160
1
2
3
fetal_health

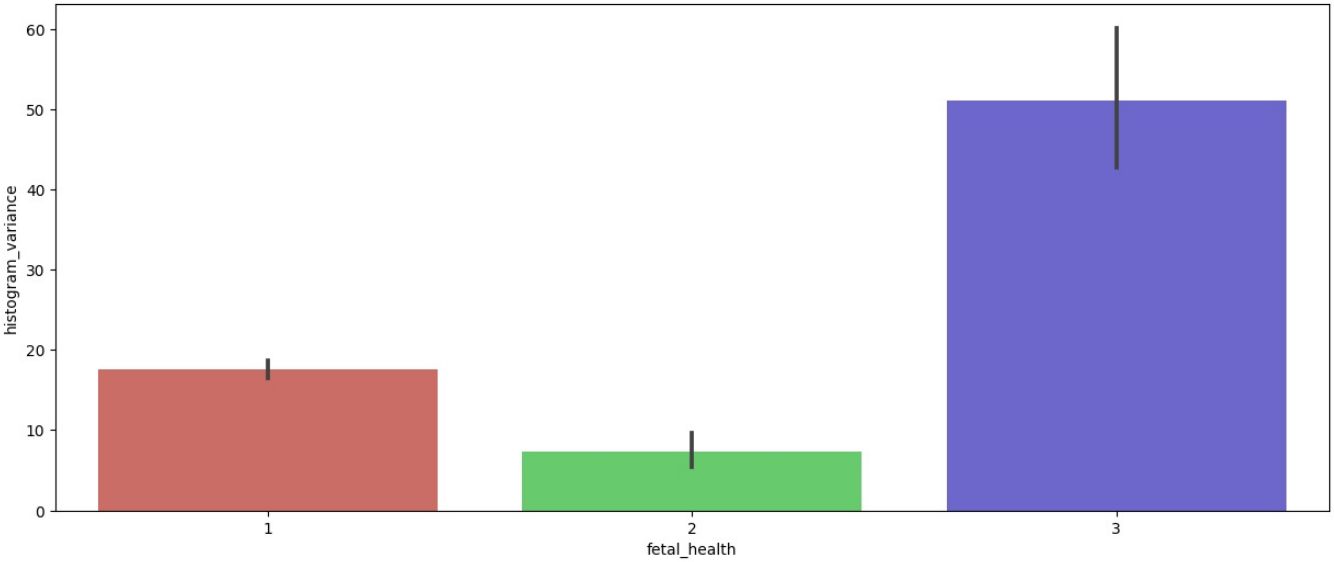

histogram_number_of_peaks
0
1
2
3
4
5
1
2
3
fetal_health

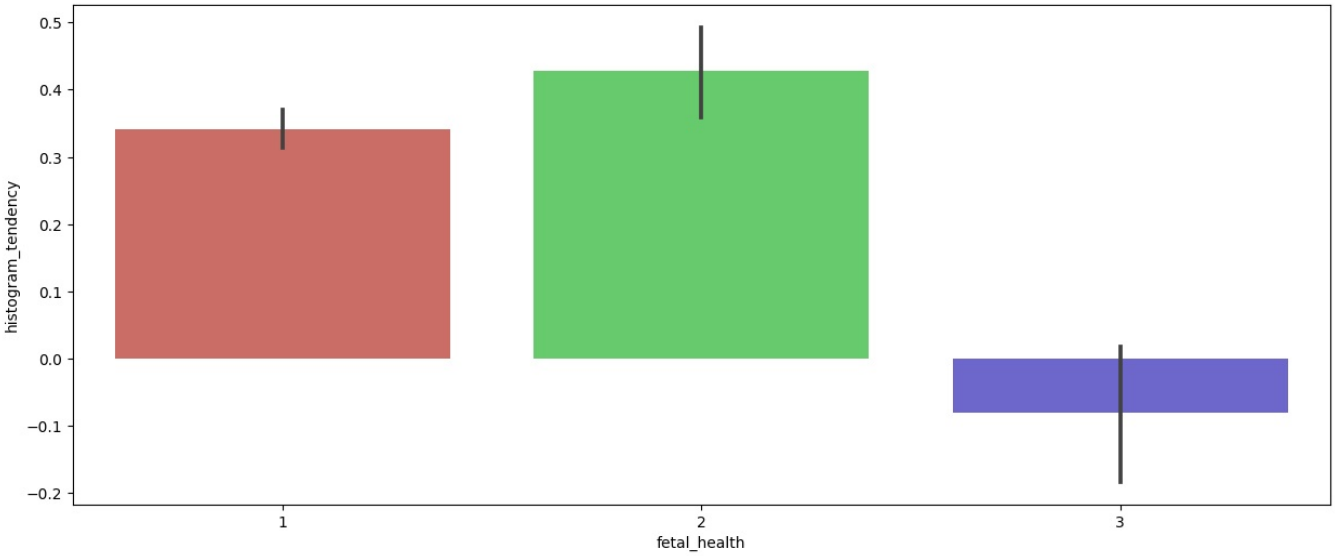

histogram_number_of_zeroes
0.0
0.1
0.2
0.3
0.4
1
2
3
fetal_health

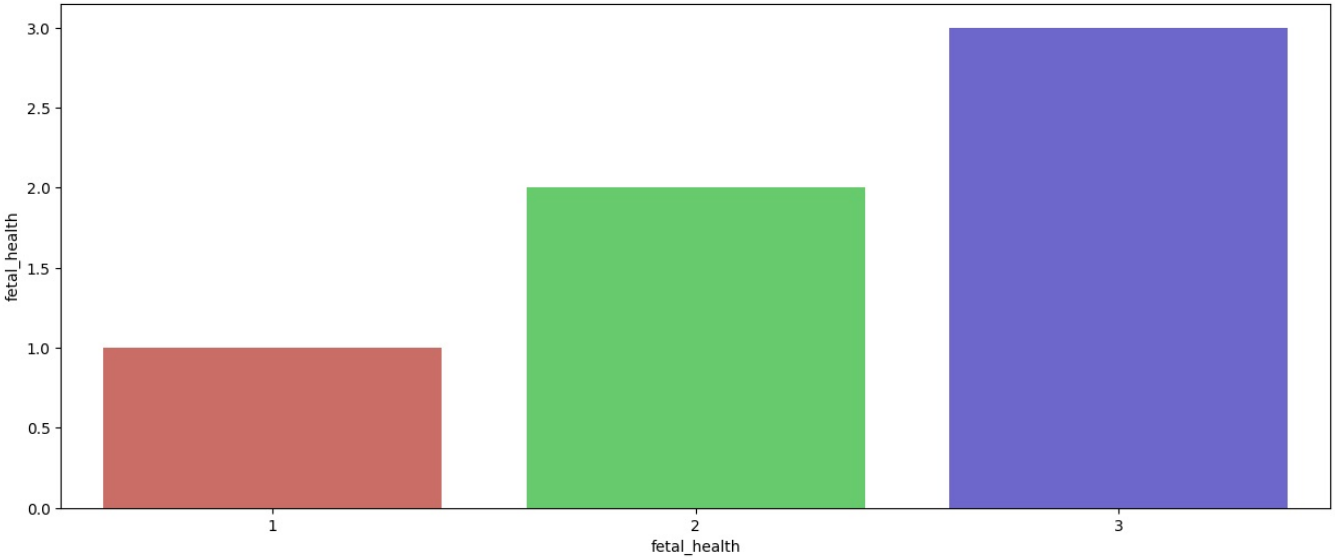

histogram_mode
0
20
40
60
80
100
120
140
1
2
3
fetal_health


histogram_mean
0
20
40
60
80
100
120
140
1
2
3
fetal_health

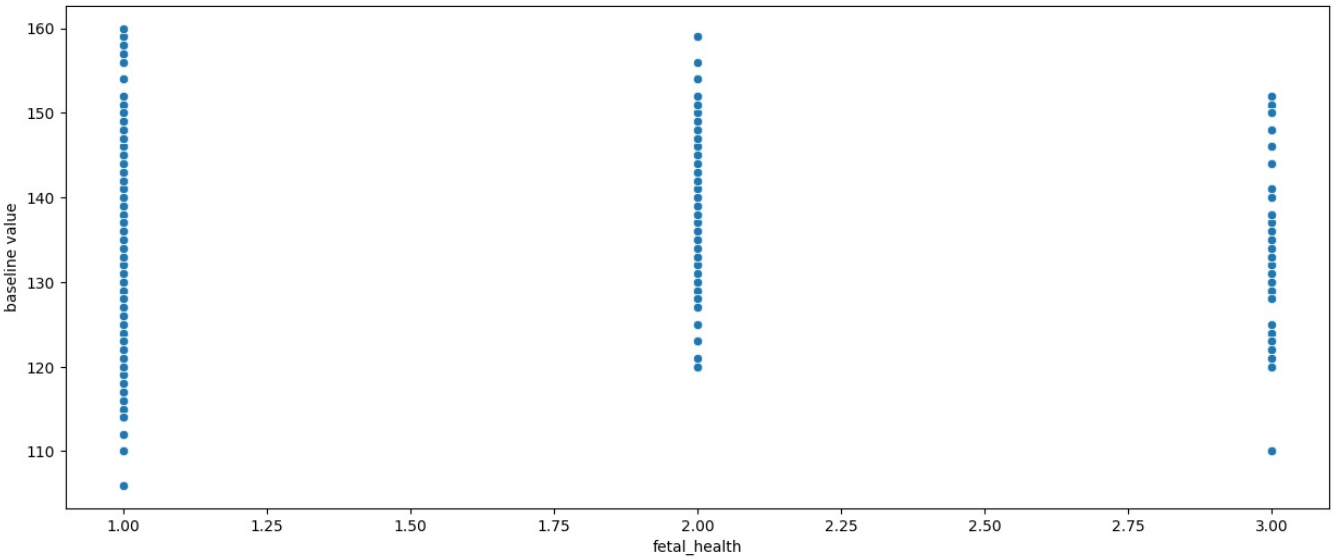

histogram_median
0
20
40
60
80
100
120
140
1
2
3
fetal_health

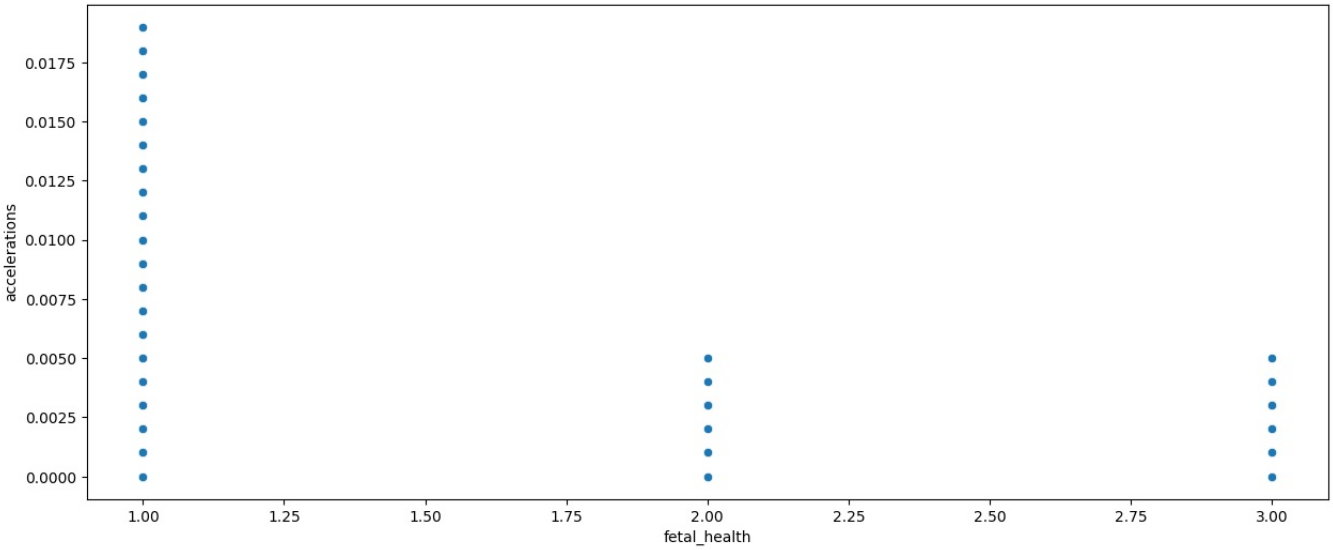

histogram_variance
0
10
20
30
40
50
60
1
2
3
fetal_health

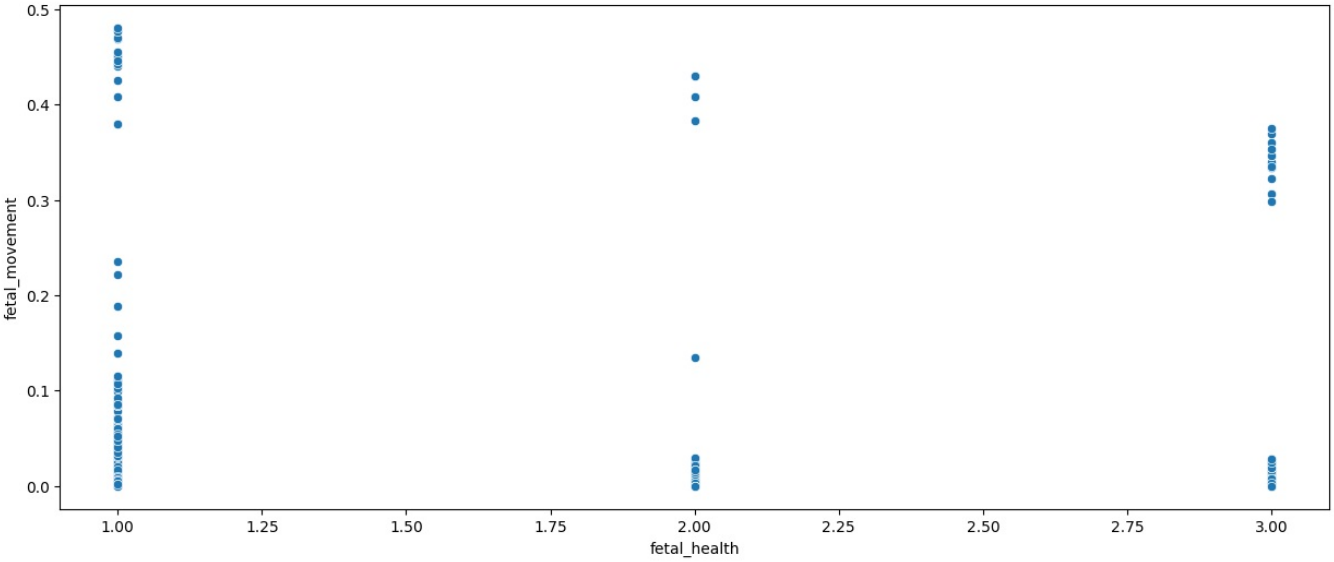


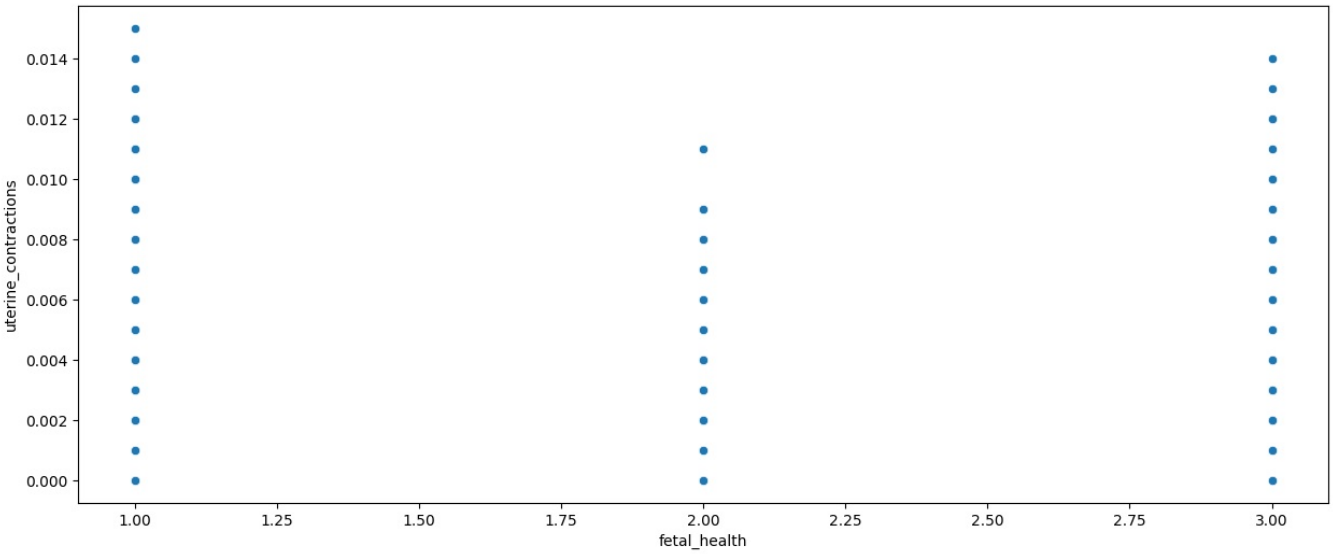


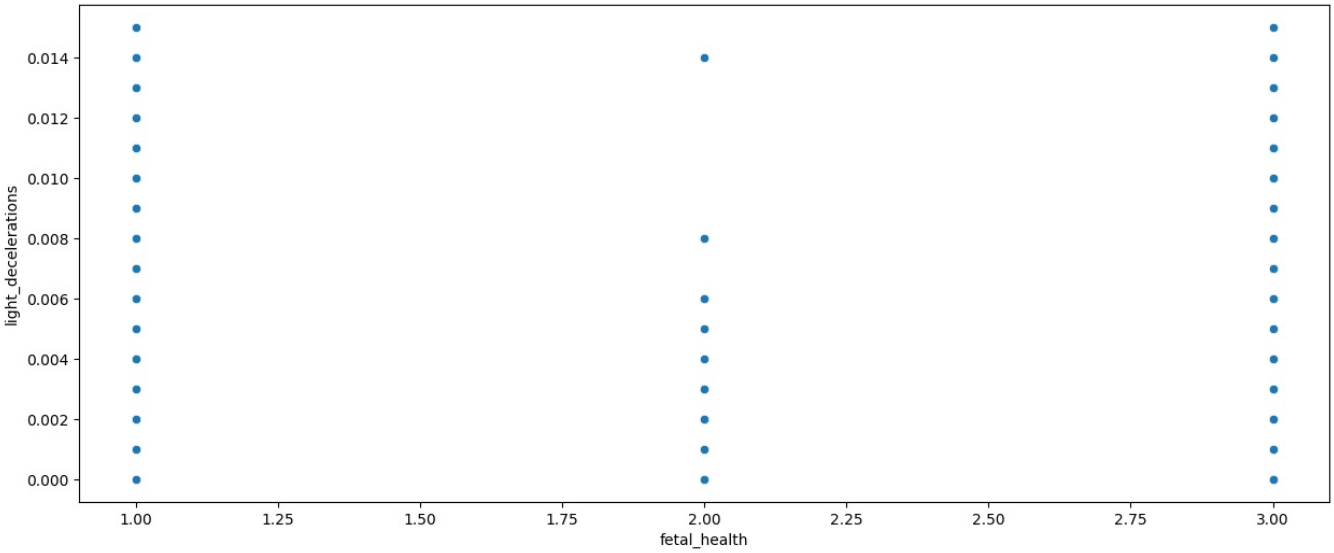
In [21]:
```
for i in df.columns:
    plt.figure(figsize=(15,6))
    sns.scatterplot(x = df['fetal_health'], y = df[i], palette = 'hls')
    plt.show()
```

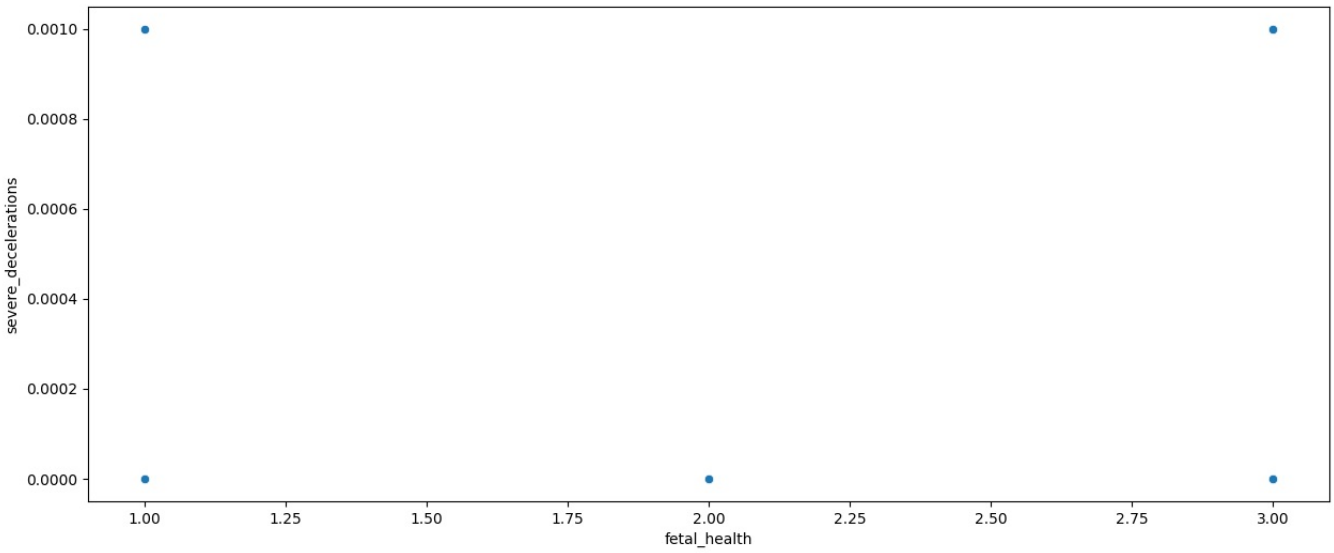


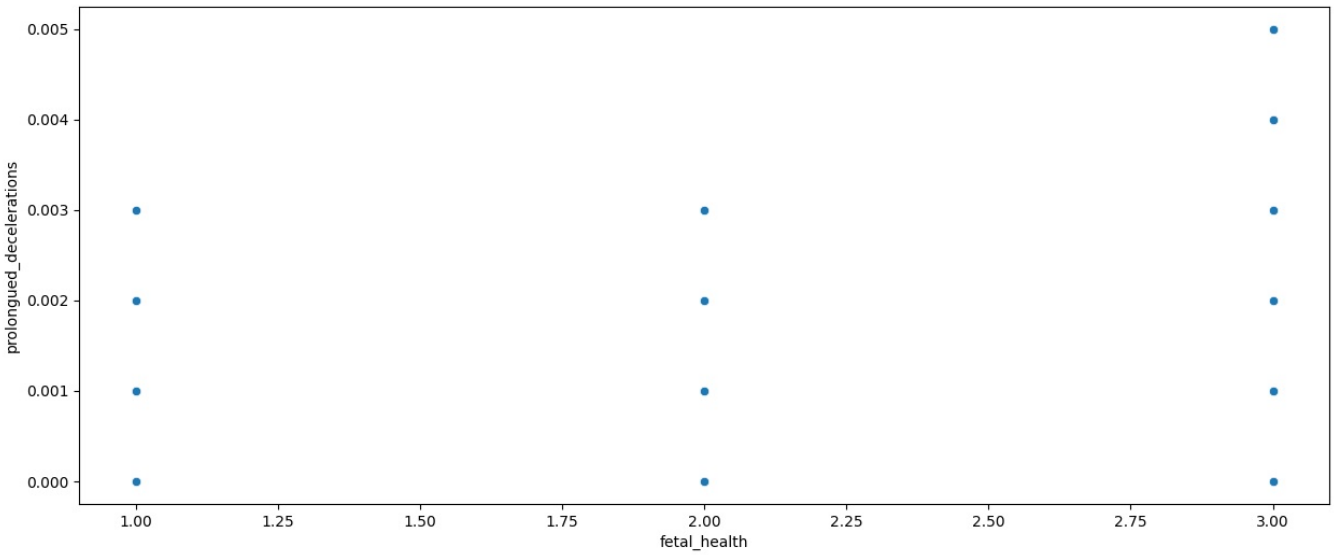

accelerations
0.0175
0.0150
0.0125
0.0100
0.0075
0.0050
0.0025
0.0000
1.00
1.25
1.50
1.75
2.00
2.25
2.50
2.75
3.00
fetal_health

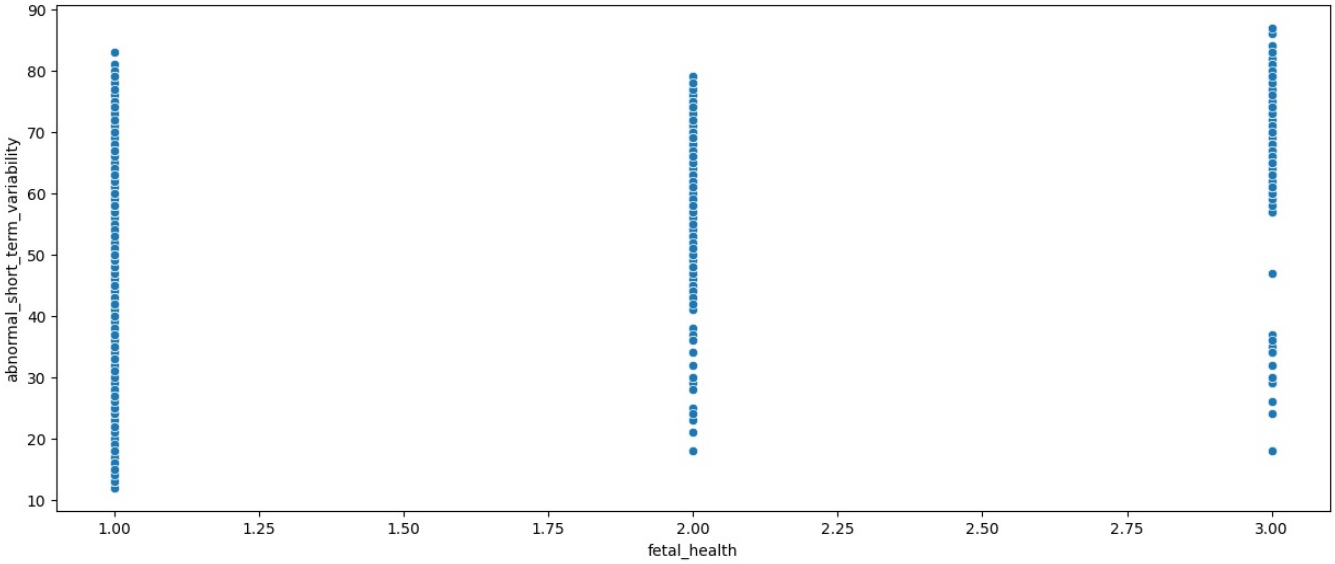

fetal_movement
0.5
0.4
0.3
0.2
0.1
0.0
1.00
1.25
1.50
1.75
2.00
2.25
2.50
2.75
3.00
fetal_health

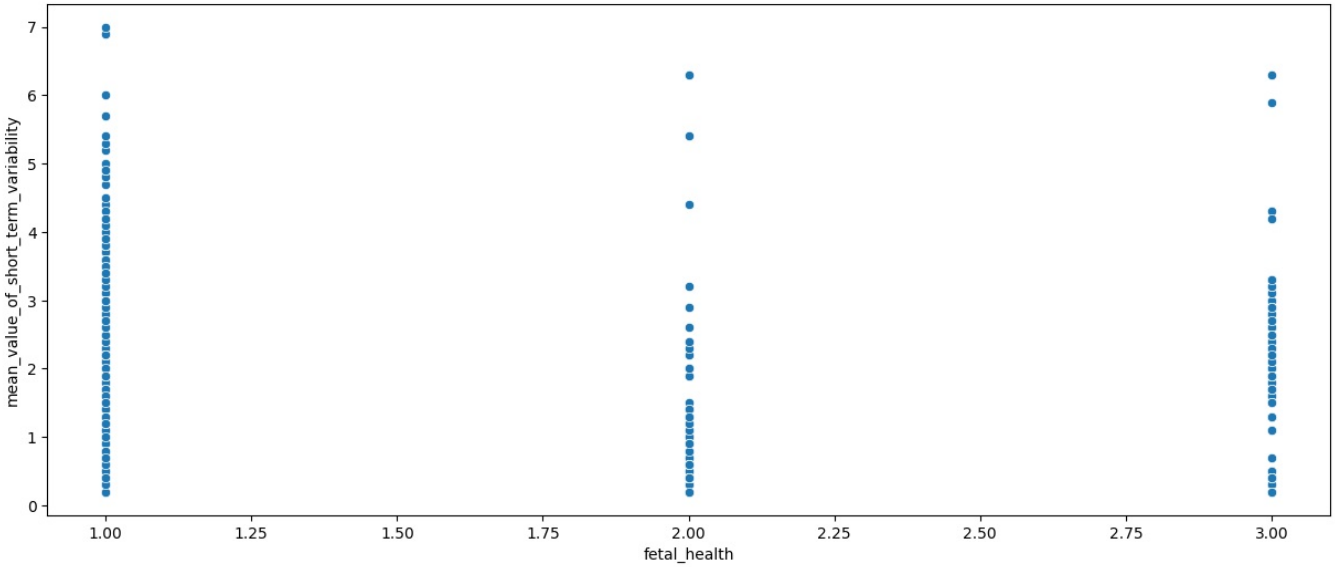

uterine_contractions
0.014
0.012
0.010
0.008
0.006
0.004
0.002
0.000
1.00
1.25
1.50
1.75
2.00
2.25
2.50
2.75
3.00
fetal_health

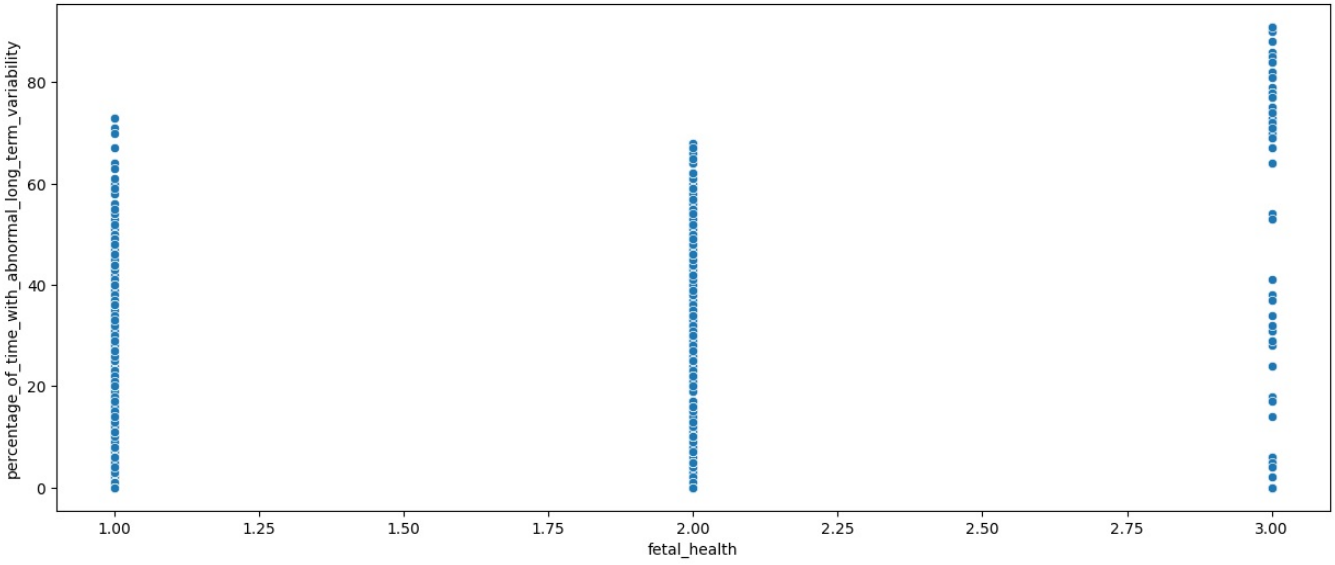

light_decelerations
0.014
0.012
0.010
0.008
0.006
0.004
0.002
0.000
1.00
1.25
1.50
1.75
2.00
2.25
2.50
2.75
3.00
fetal_health

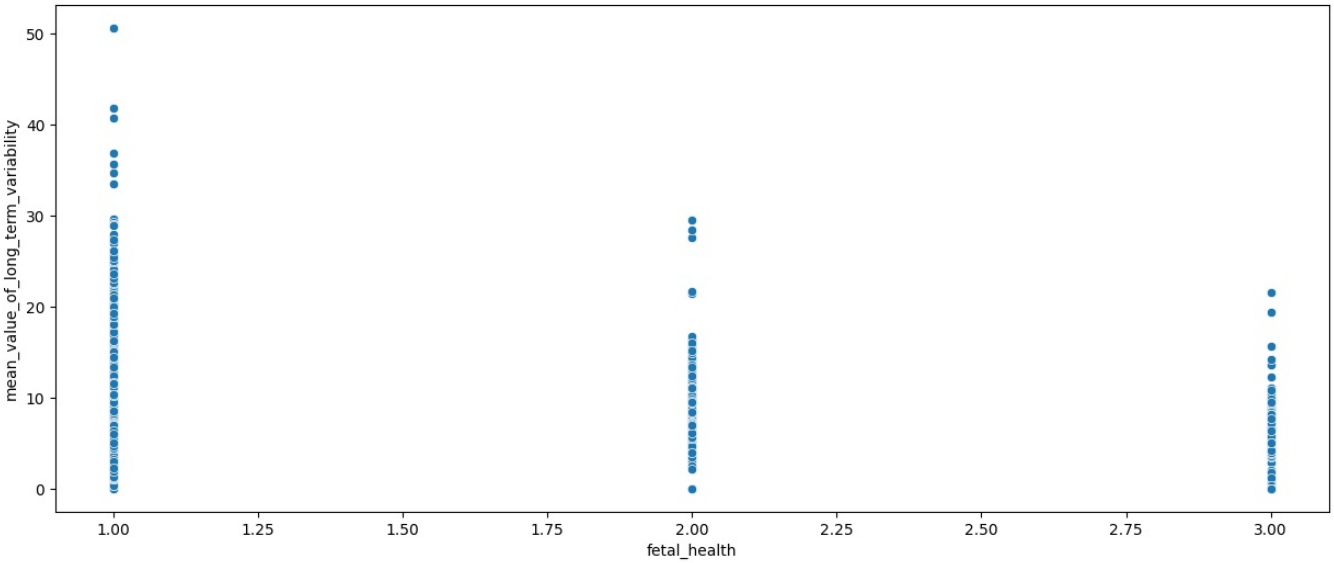

severe_decelerations
0.0010
0.0008
0.0006
0.0004
0.0002
0.0000
1.00
1.25
1.50
1.75
2.00
2.25
2.50
2.75
3.00
fetal_health

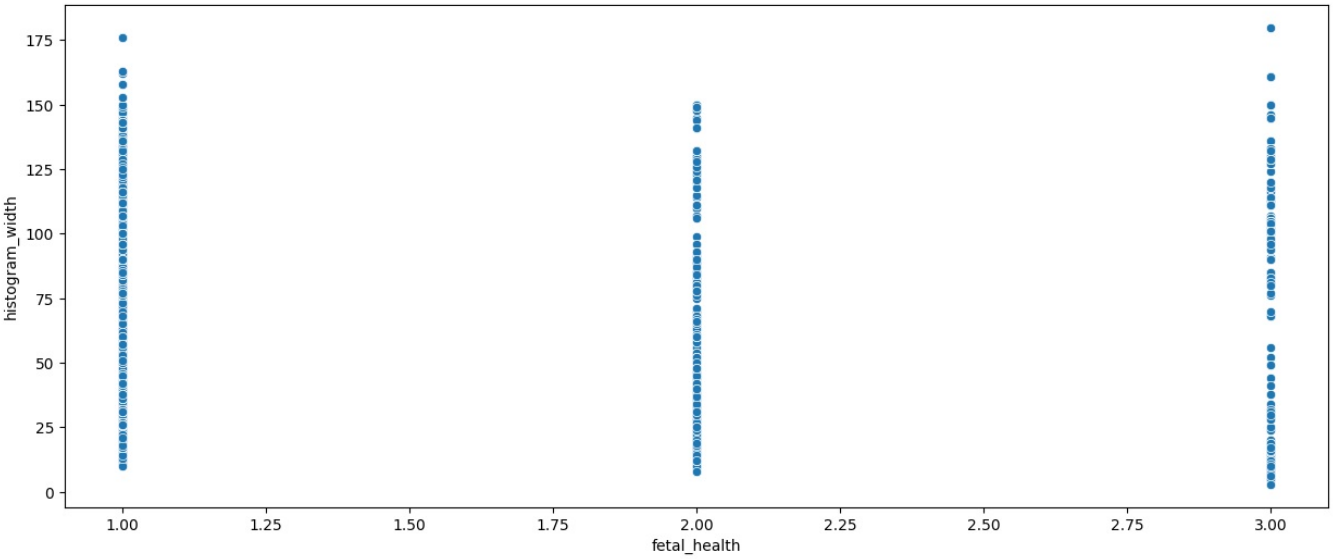

prolongued_decelerations
0.005
0.004
0.003
0.002
0.001
0.000
1.00
1.25
1.50
1.75
2.00
2.25
2.50
2.75
3.00
fetal_health

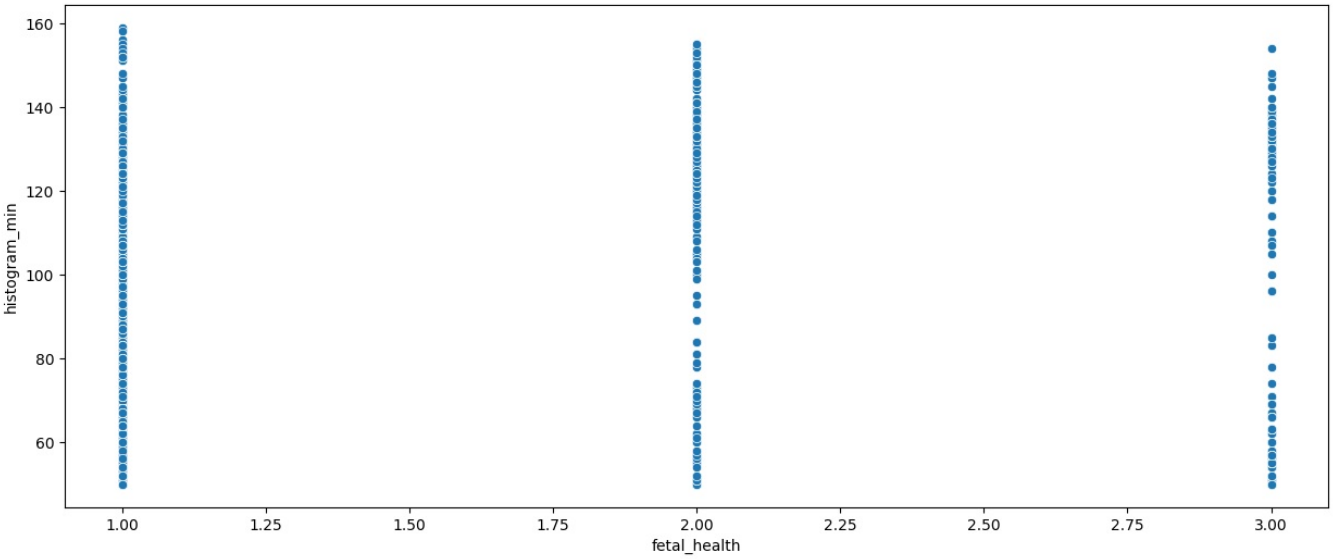

abnormal_short_term_variability
fetal_health

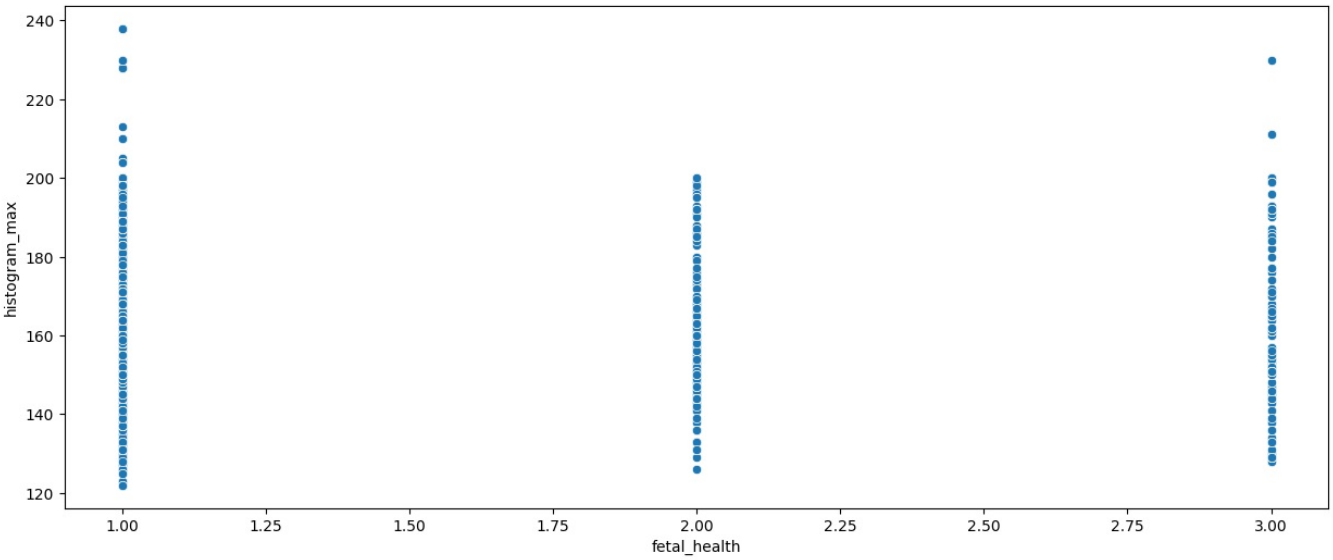

mean_value_of_short_term_variability
fetal_health

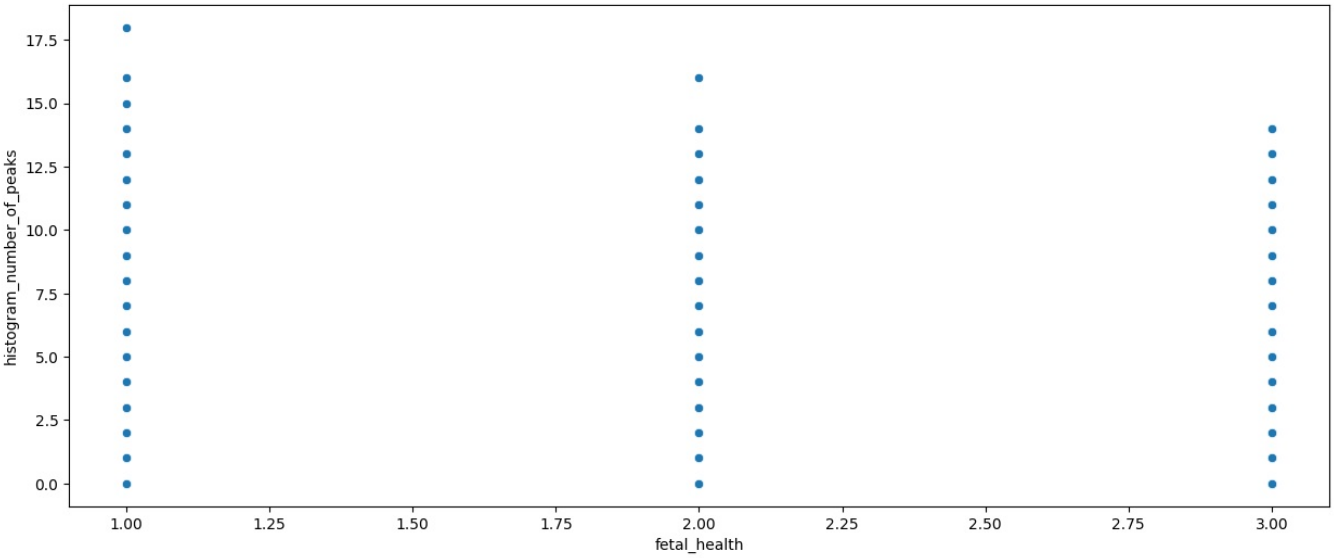

percentage_of_time_with_abnormal_long_term_variability
fetal_health

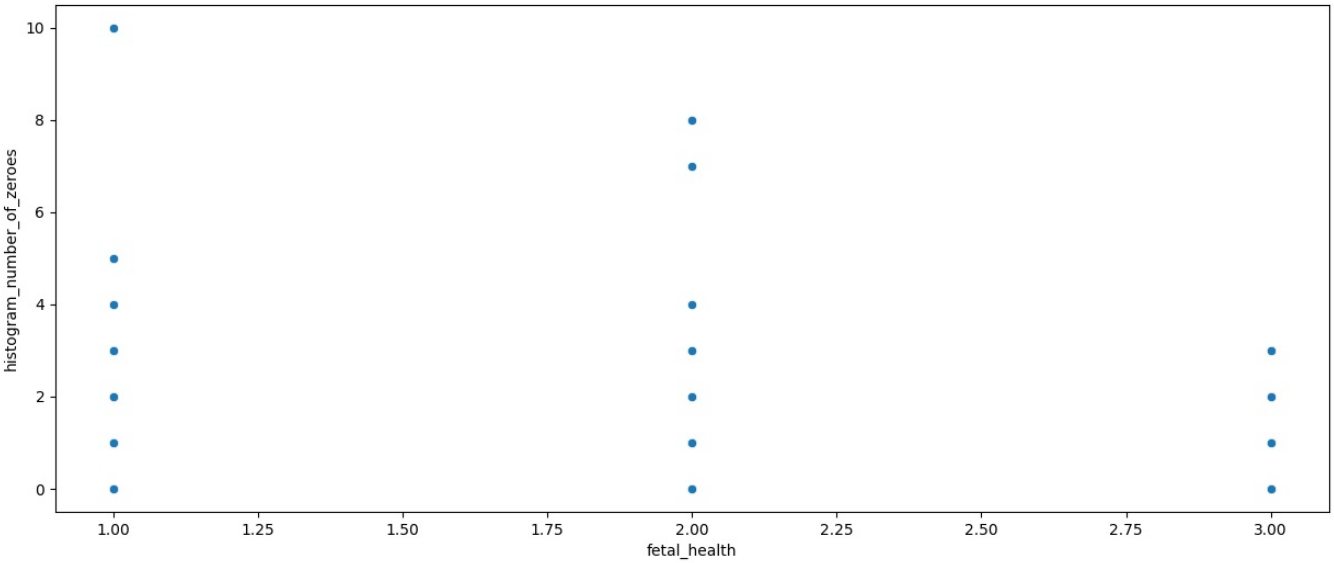

mean_value_of_long_term_variability
fetal_health

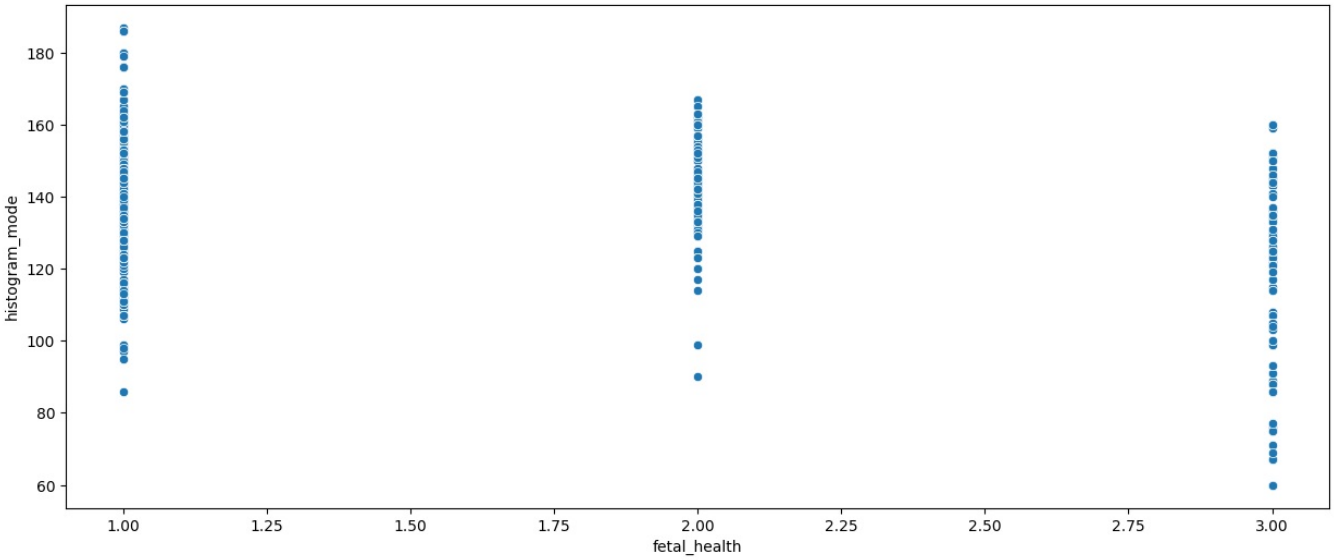

histogram_width
fetal_health

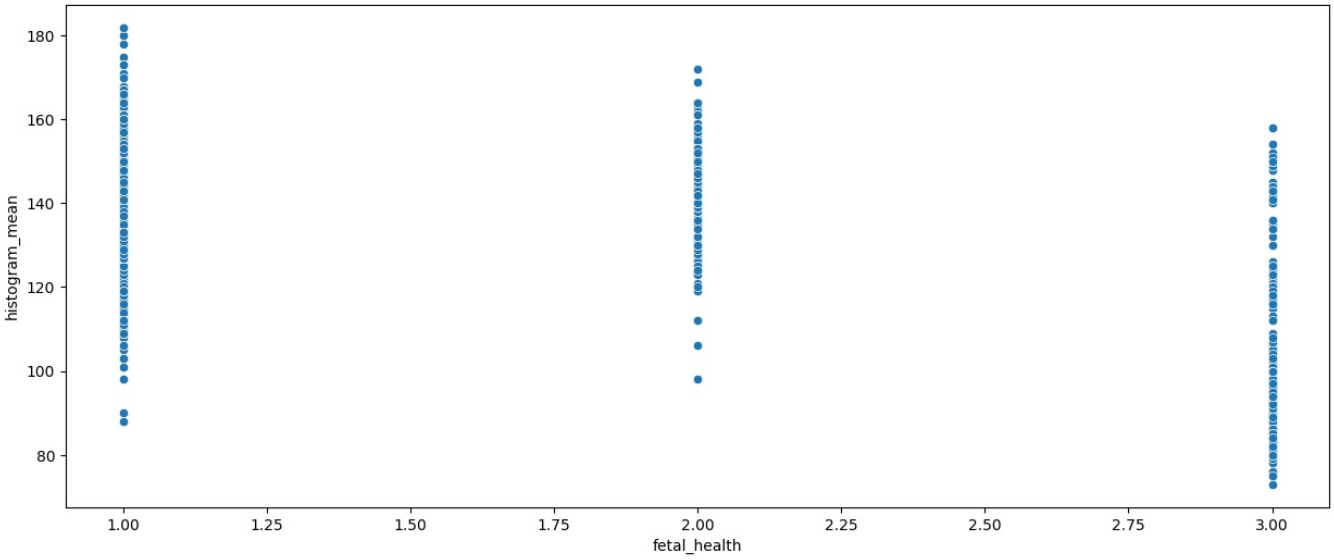

histogram_min
fetal_health

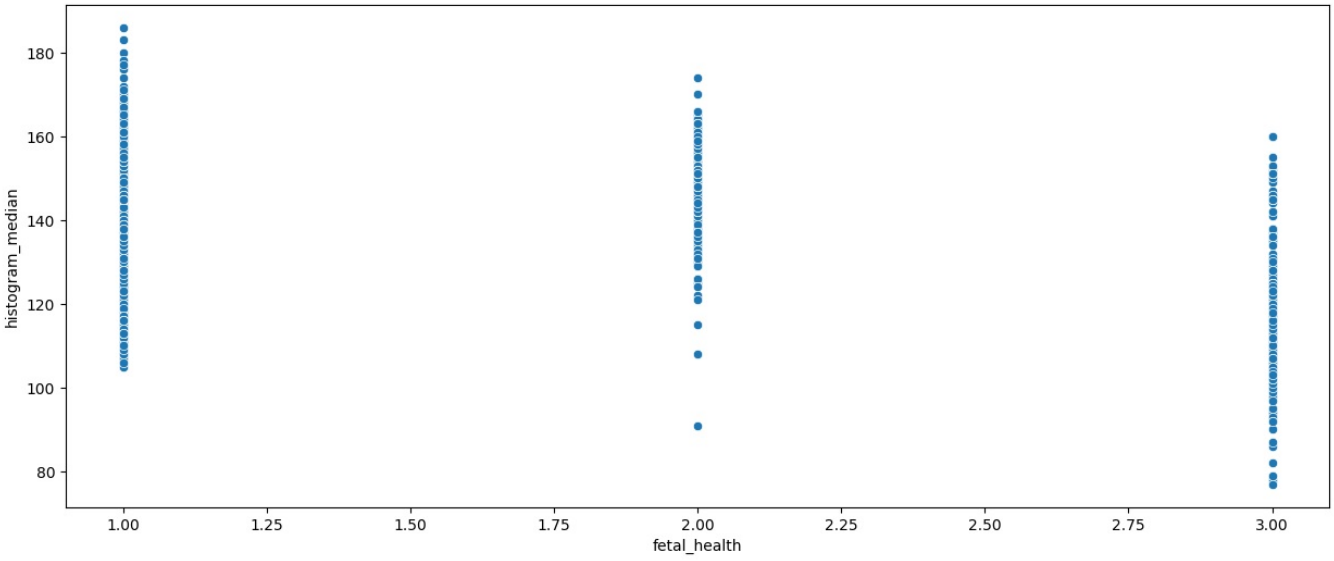

histogram_max
240
220
200
180
160
140
120
1.00
1.25
1.50
1.75
2.00
2.25
2.50
2.75
3.00
fetal_health

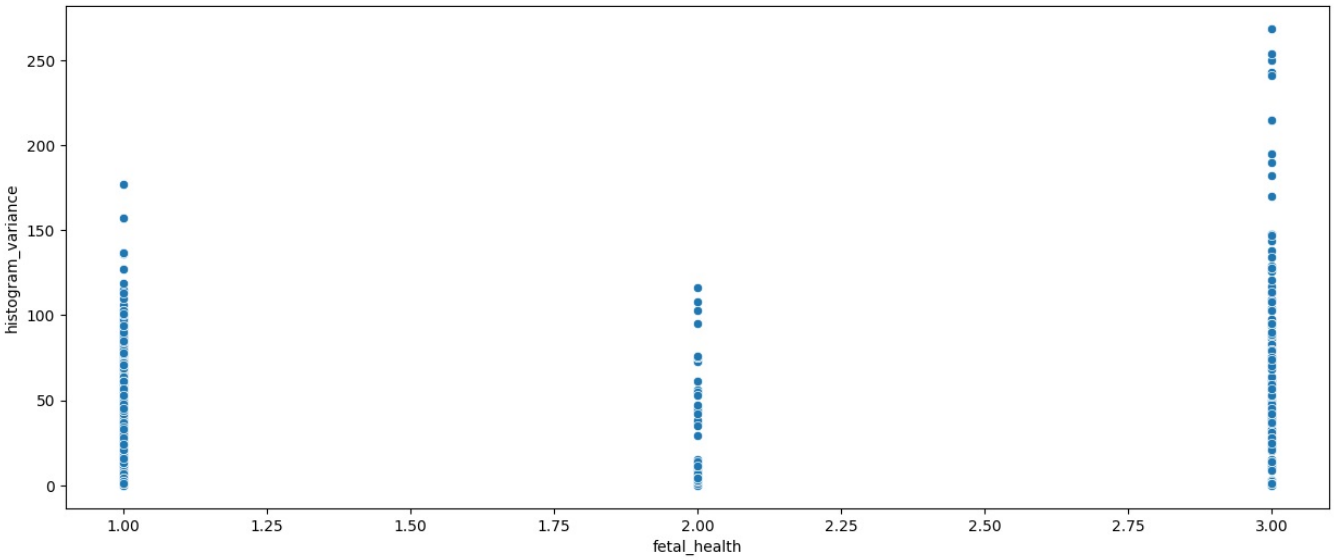

histogram_number_of_peaks
17.5
15.0
12.5
10.0
7.5
5.0
2.5
0.0
1.00
1.25
1.50
1.75
2.00
2.25
2.50
2.75
3.00
fetal_health

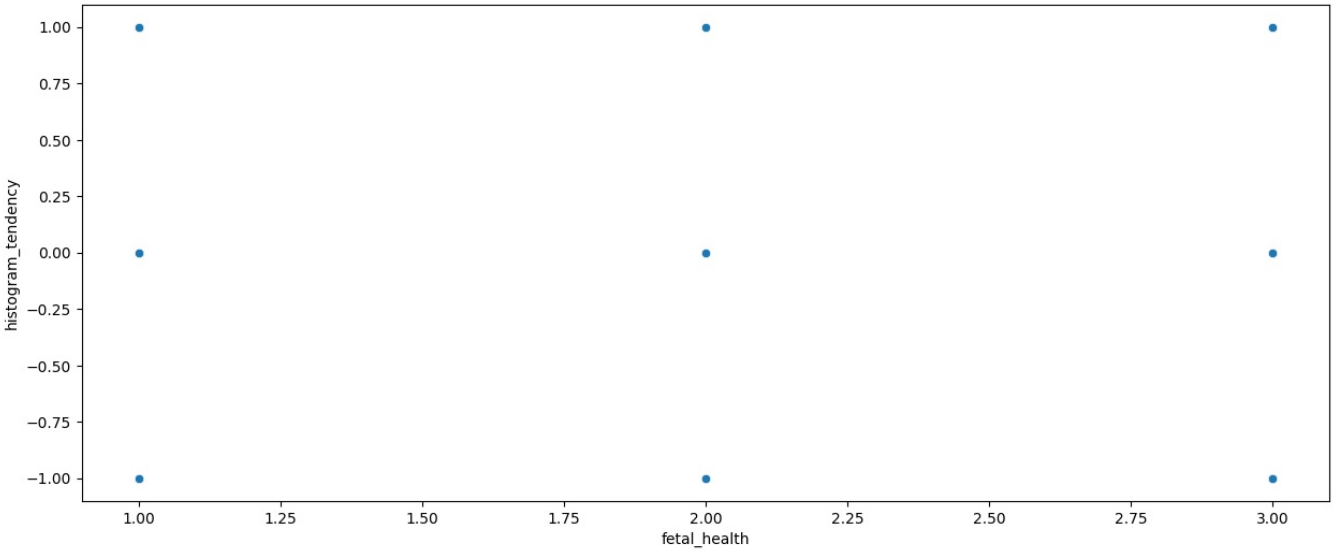

histogram_number_of_zeroes
10
8
6
4
2
0
1.00
1.25
1.50
1.75
2.00
2.25
2.50
2.75
3.00
fetal_health

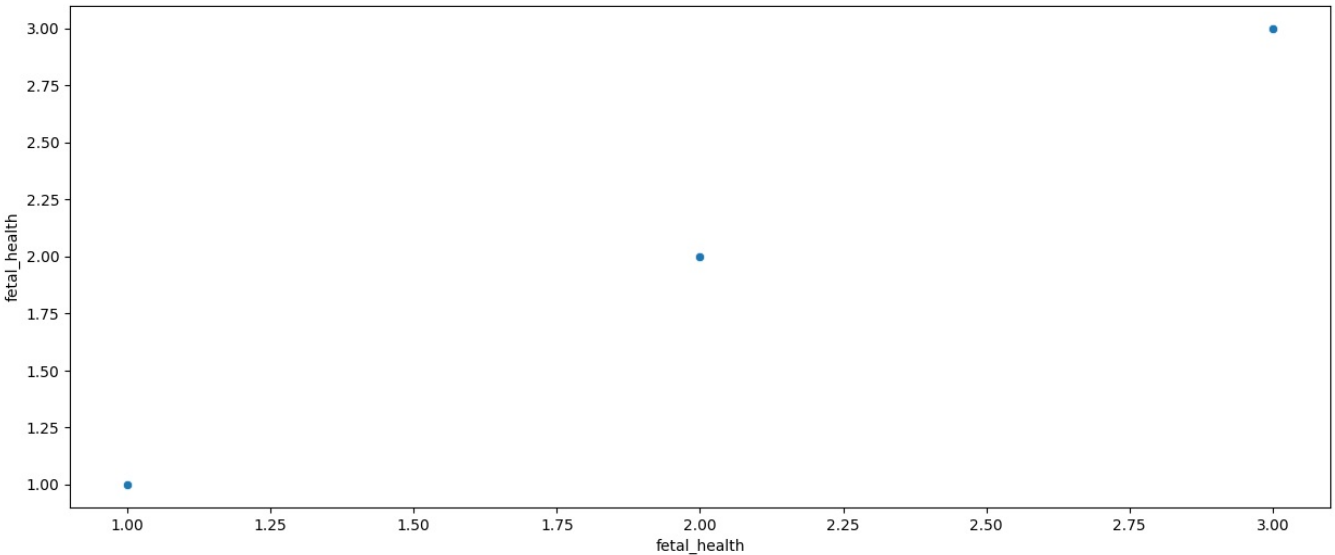

histogram_mode
180
160
140
120
100
80
60
1.00
1.25
1.50
1.75
2.00
2.25
2.50
2.75
3.00
fetal_health


histogram_mean
180
160
140
120
100
80
1.00
1.25
1.50
1.75
2.00
2.25
2.50
2.75
3.00
fetal_health

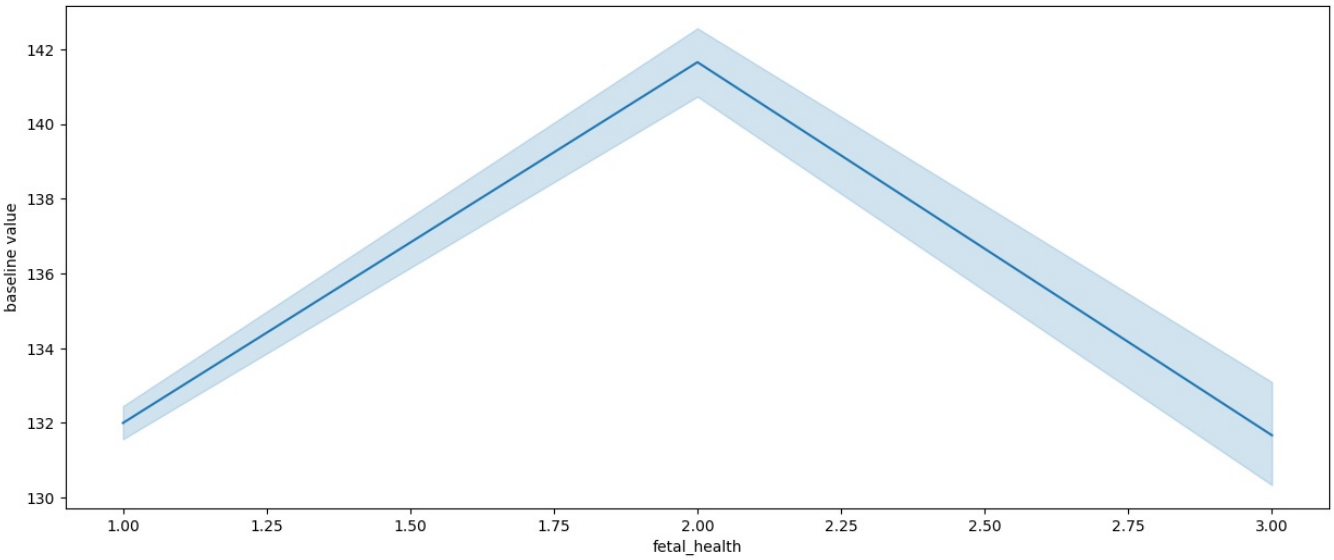

histogram_median
180
160
140
120
100
80
1.00
1.25
1.50
1.75
2.00
2.25
2.50
2.75
3.00
fetal_health

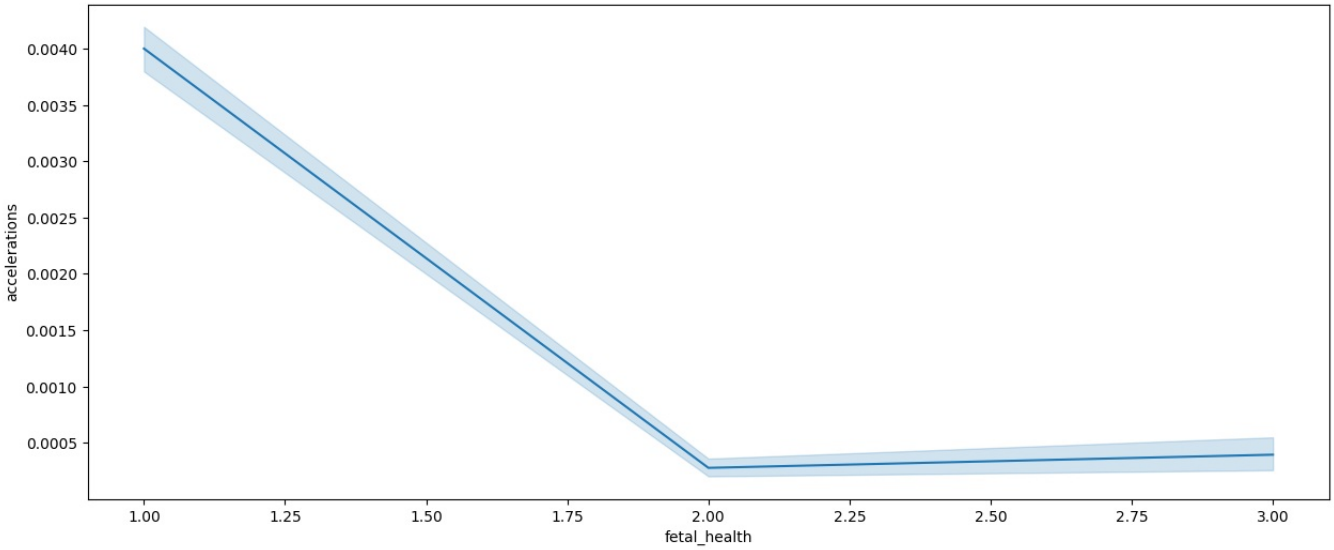


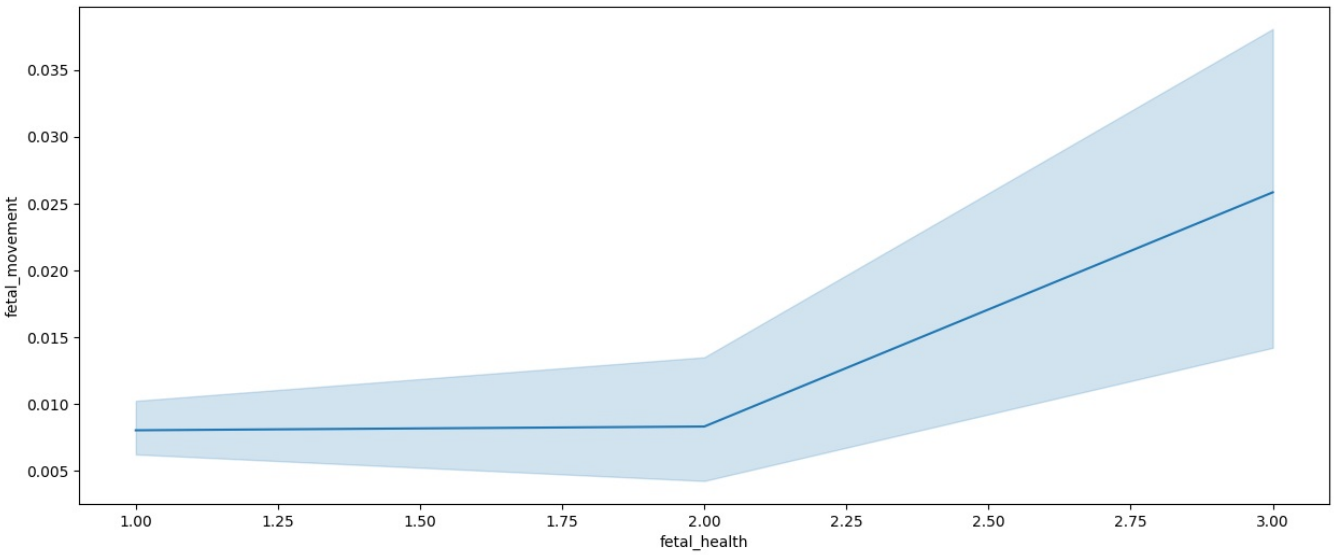


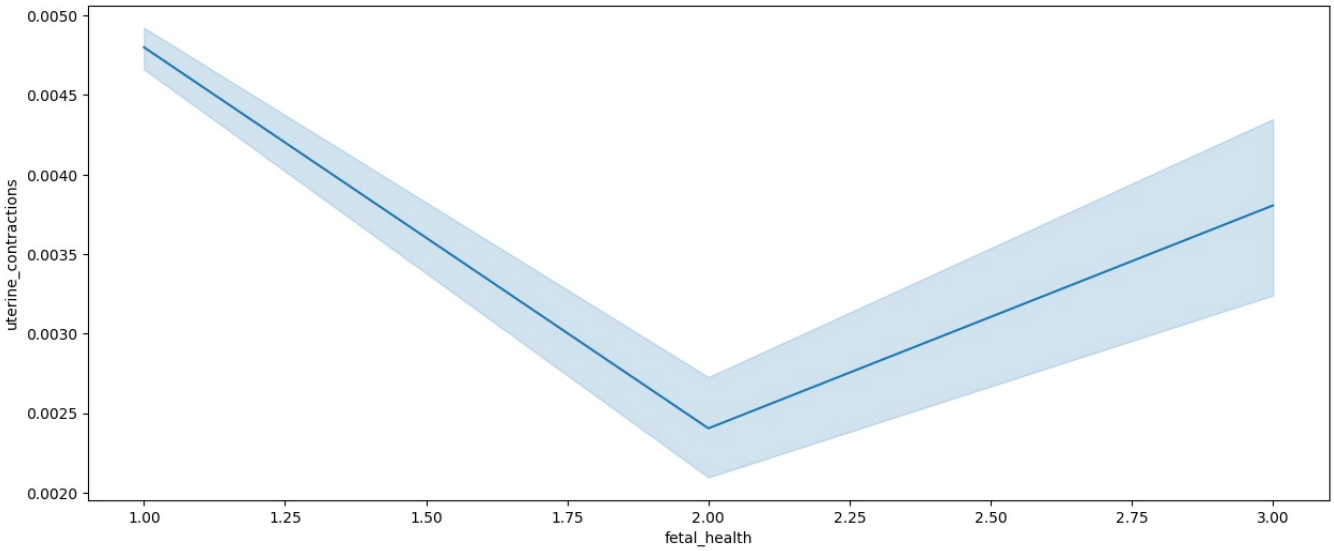


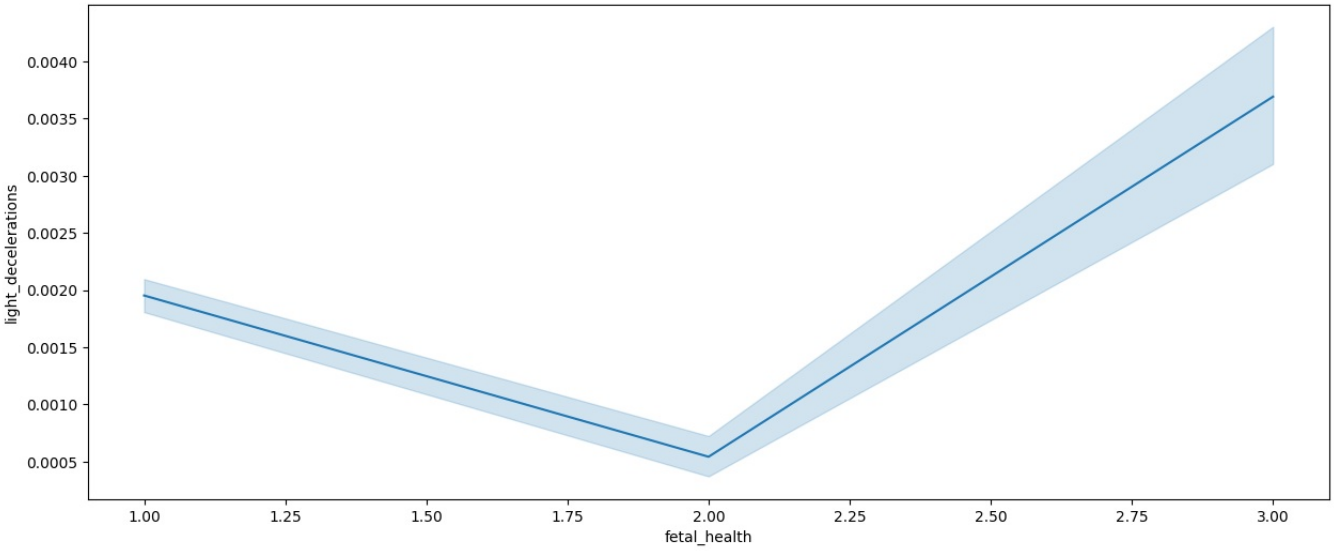
In [22]:
```
for i in df.columns:
    plt.figure(figsize=(15,6))
    sns.lineplot(x = df['fetal_health'], y = df[i], palette = 'hls')
    plt.show()
```

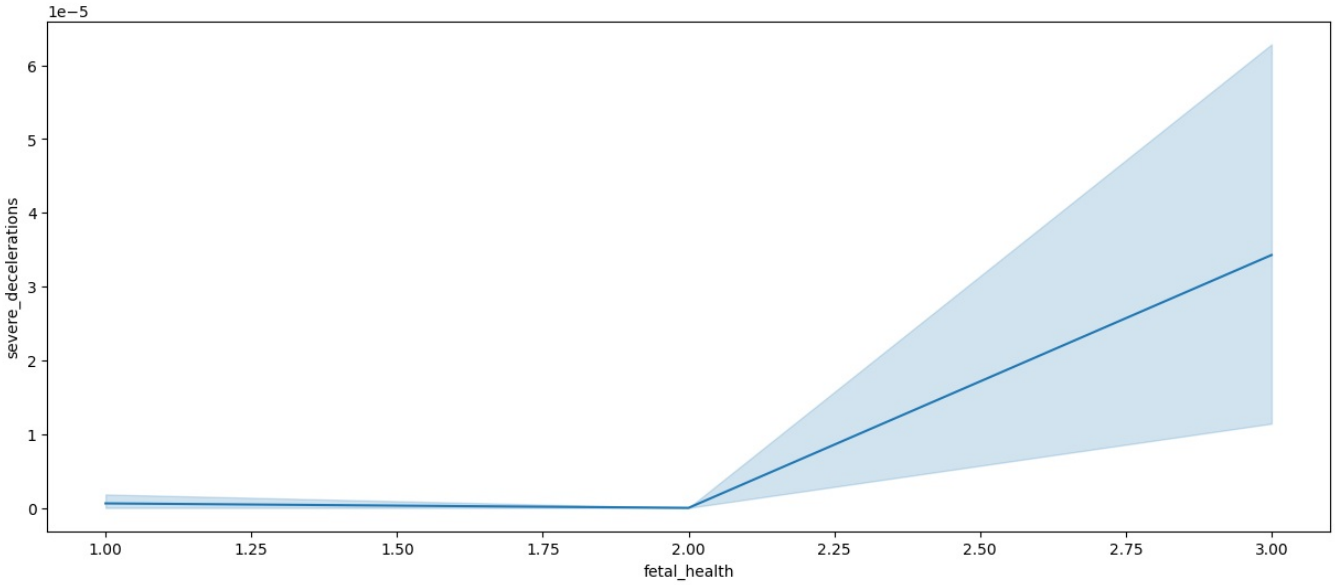

baseline value
142
140
138
136
134
132
130
1.00
1.25
1.50
1.75
2.00
2.25
2.50
2.75
3.00
fetal_health

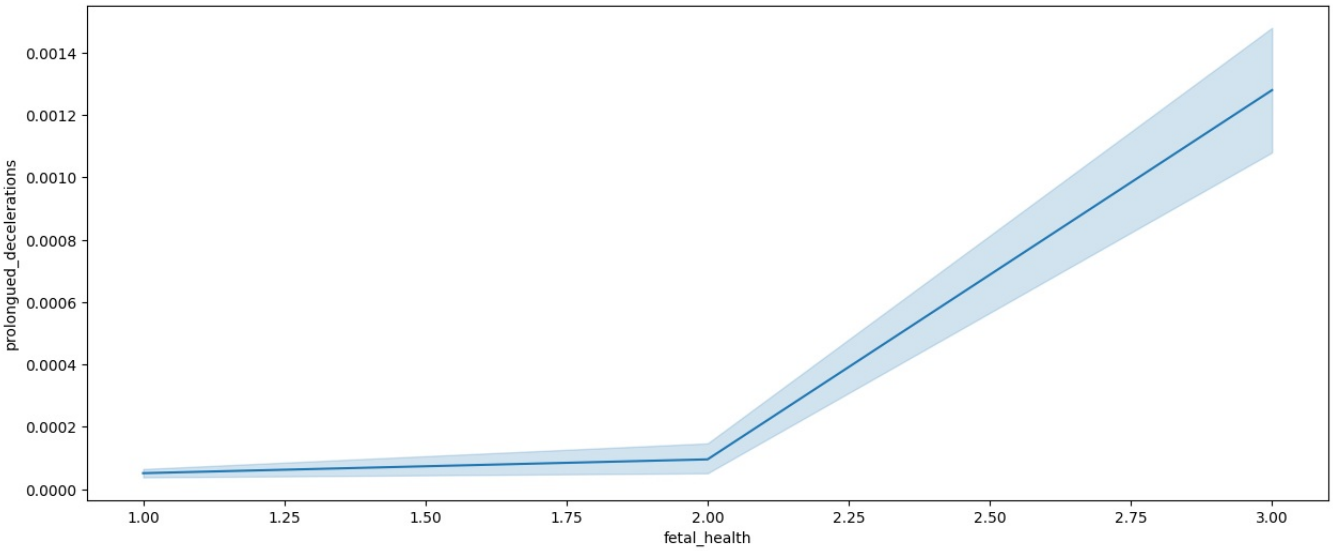

accelerations
0.0040
0.0035
0.0030
0.0025
0.0020
0.0015
0.0010
0.0005
1.00
1.25
1.50
1.75
2.00
2.25
2.50
2.75
3.00
fetal_health

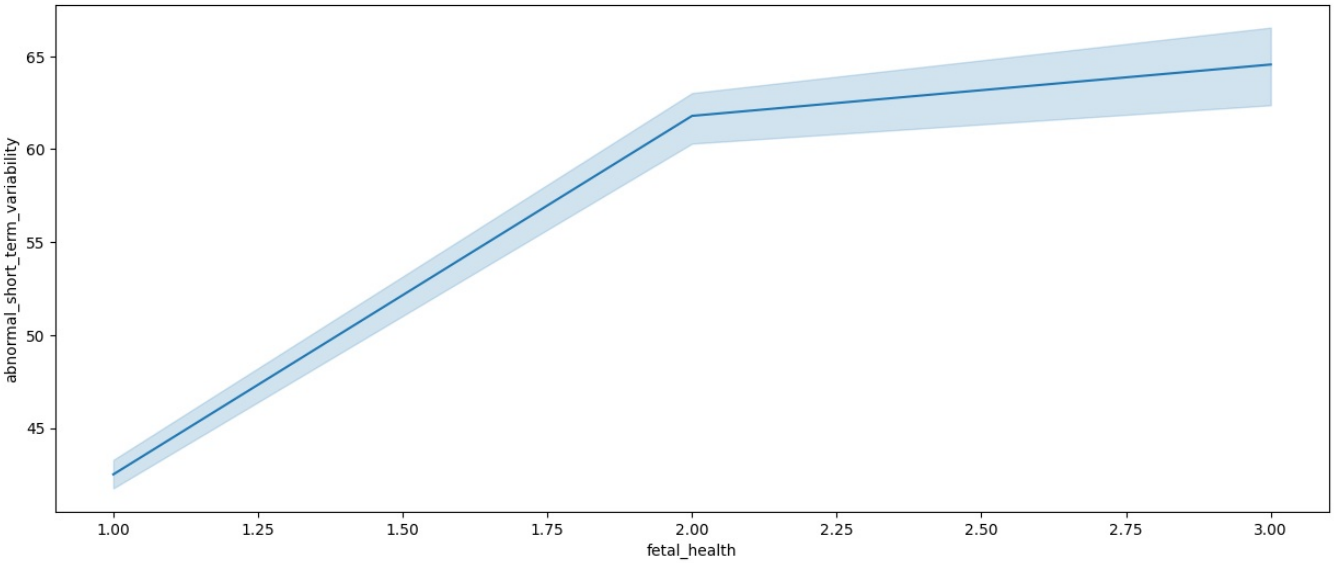

fetal_movement
0.035
0.030
0.025
0.020
0.015
0.010
0.005
1.00
1.25
1.50
1.75
2.00
2.25
2.50
2.75
3.00
fetal_health

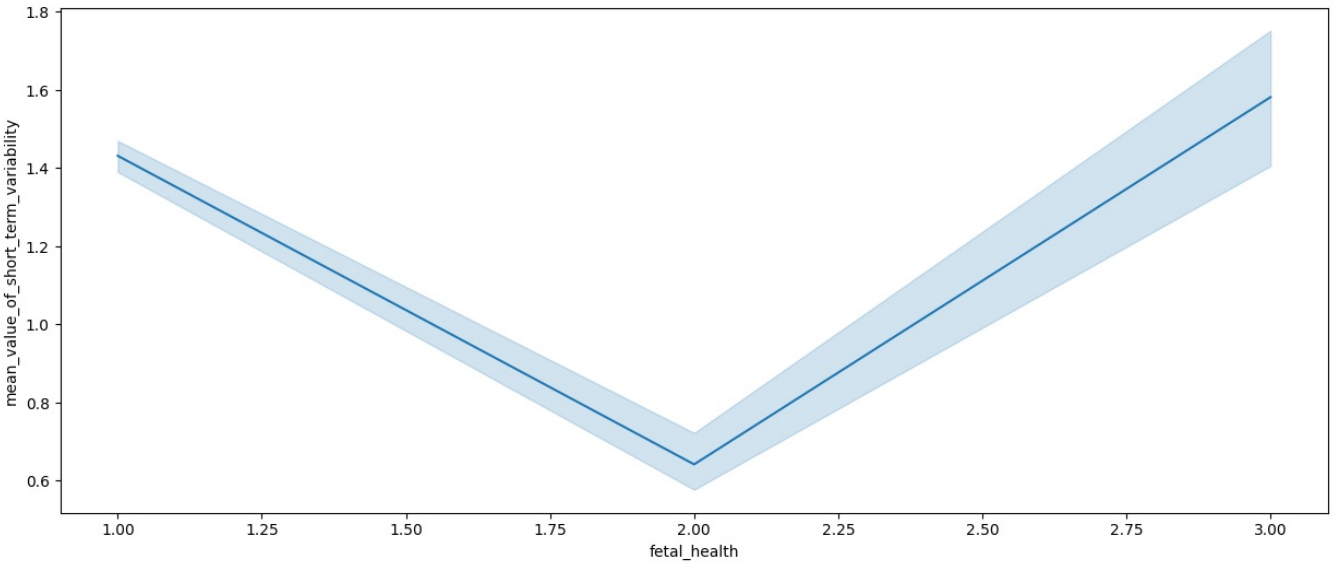

uterine_contractions
0.0050
0.0045
0.0040
0.0035
0.0030
0.0025
0.0020
1.00
1.25
1.50
1.75
2.00
2.25
2.50
2.75
3.00
fetal_health

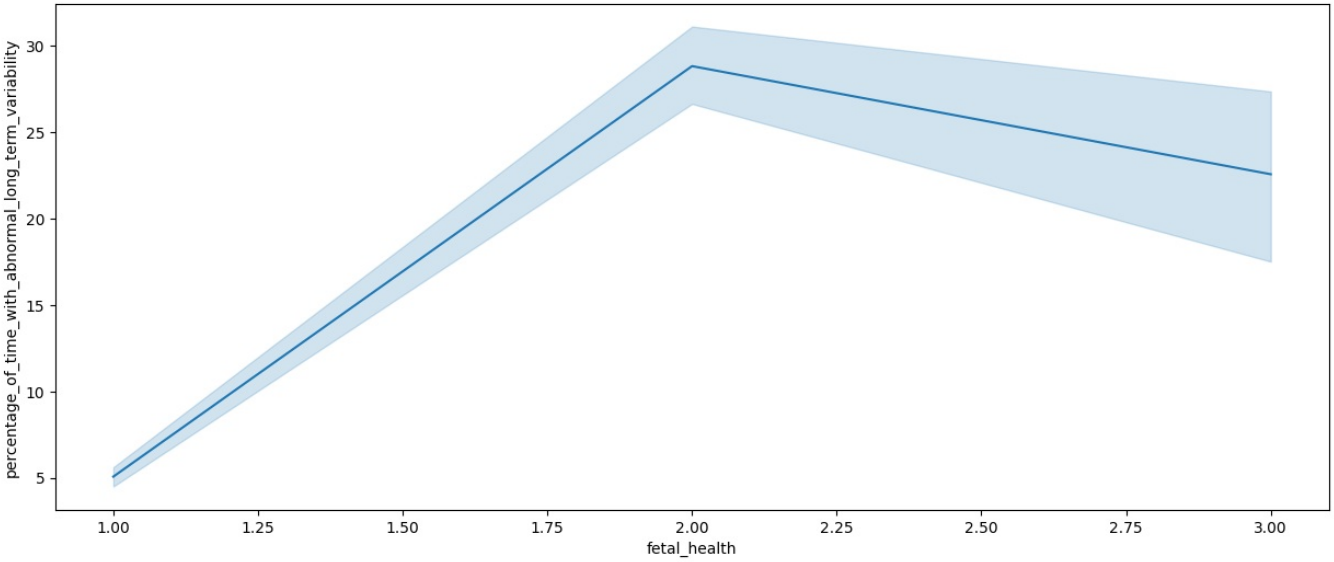

light_decelerations
0.0040
0.0035
0.0030
0.0025
0.0020
0.0015
0.0010
0.0005
1.00
1.25
1.50
1.75
2.00
2.25
2.50
2.75
3.00
fetal_health

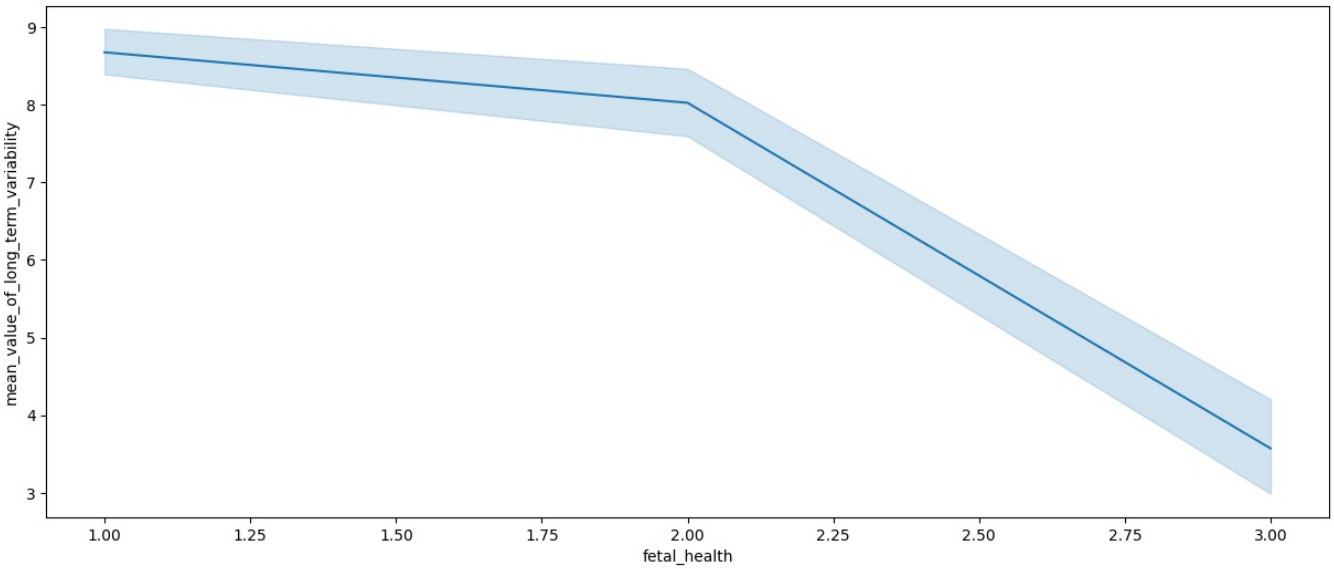

1e−5
severe_decelerations
6
5
4
3
2
1
0
1.00
1.25
1.50
1.75
2.00
2.25
2.50
2.75
3.00
fetal_health

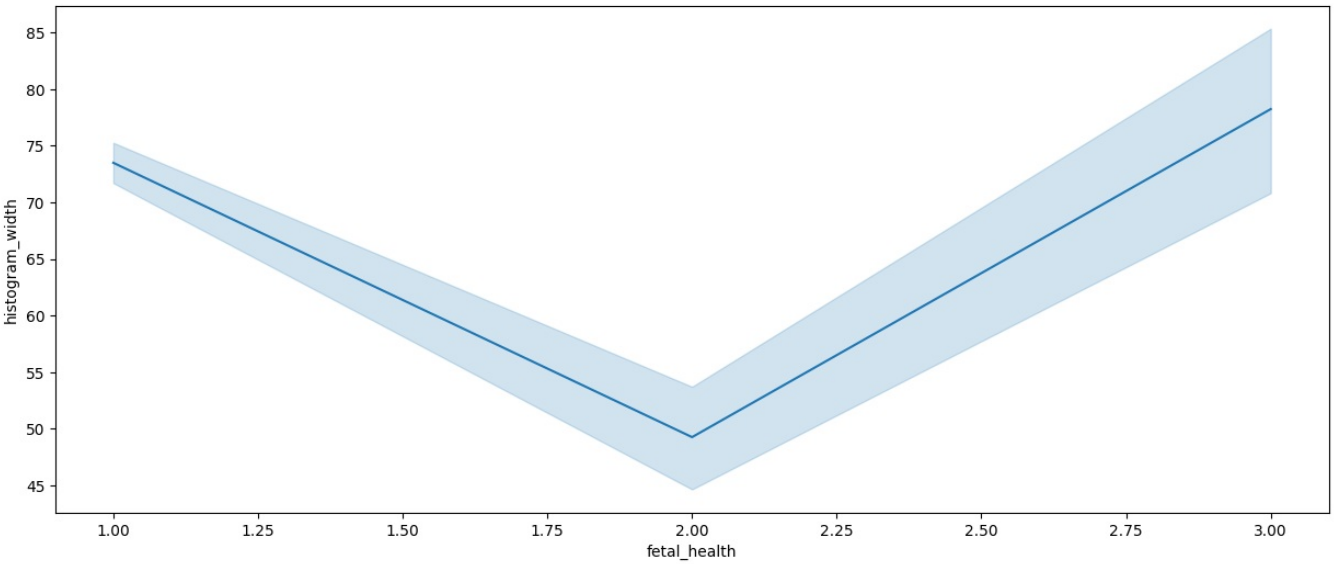

prolongued_decelerations
0.0014
0.0012
0.0010
0.0008
0.0006
0.0004
0.0002
0.0000
1.00
1.25
1.50
1.75
2.00
2.25
2.50
2.75
3.00
fetal_health

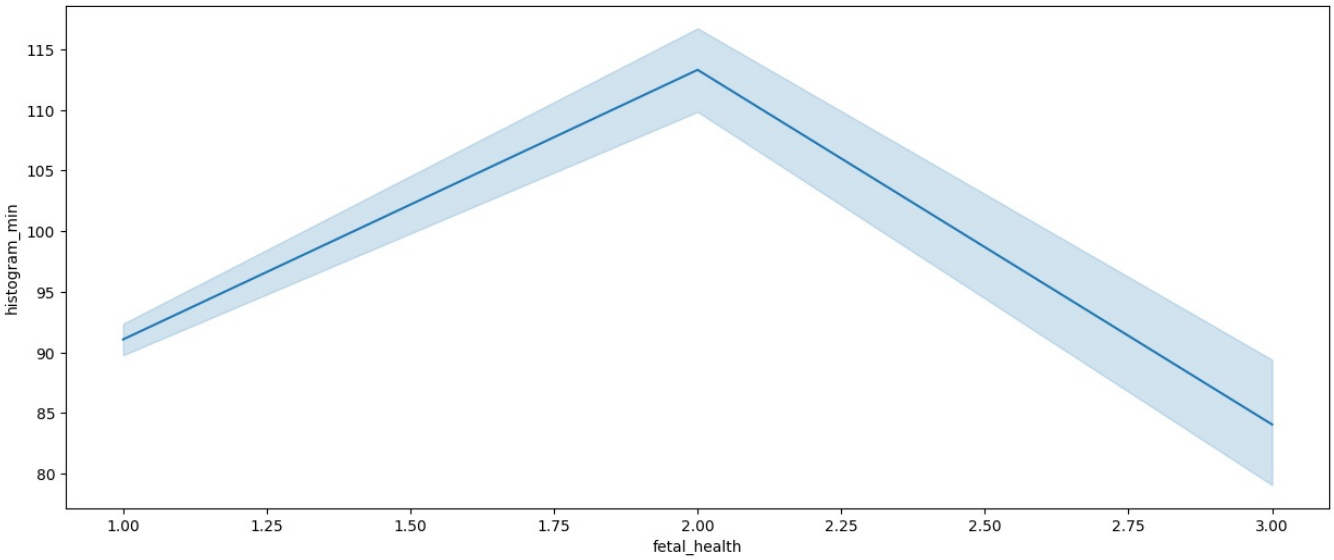

abnormal_short_term_variability
65
60
55
50
45
1.00
1.25
1.50
1.75
2.00
2.25
2.50
2.75
3.00
fetal_health

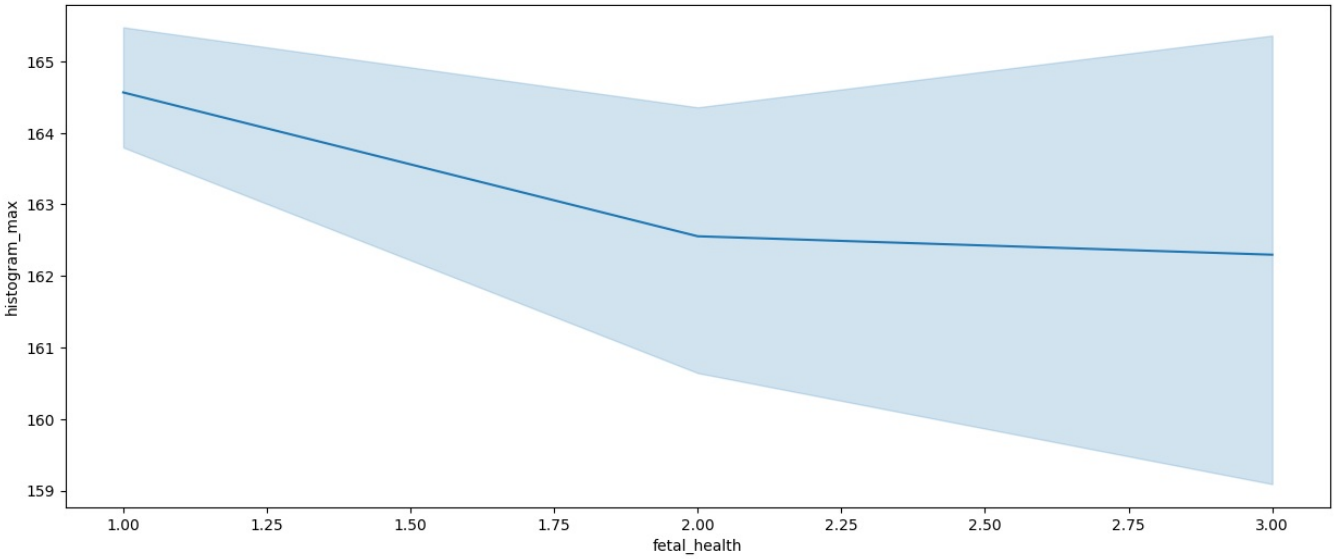

mean_value_of_short_term_variability
1.8
1.6
1.4
1.2
1.0
0.8
0.6
1.00
1.25
1.50
1.75
2.00
2.25
2.50
2.75
3.00
fetal_health

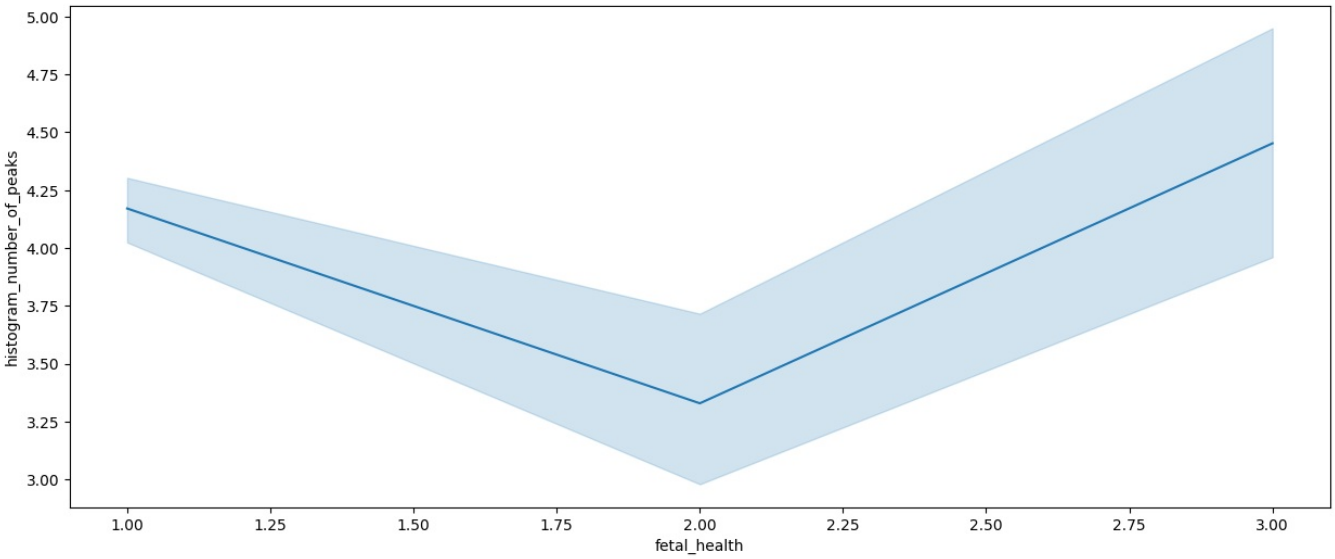

percentage_of_time_with_abnormal_long_term_variability
30
25
20
15
10
5
1.00
1.25
1.50
1.75
2.00
2.25
2.50
2.75
3.00
fetal_health

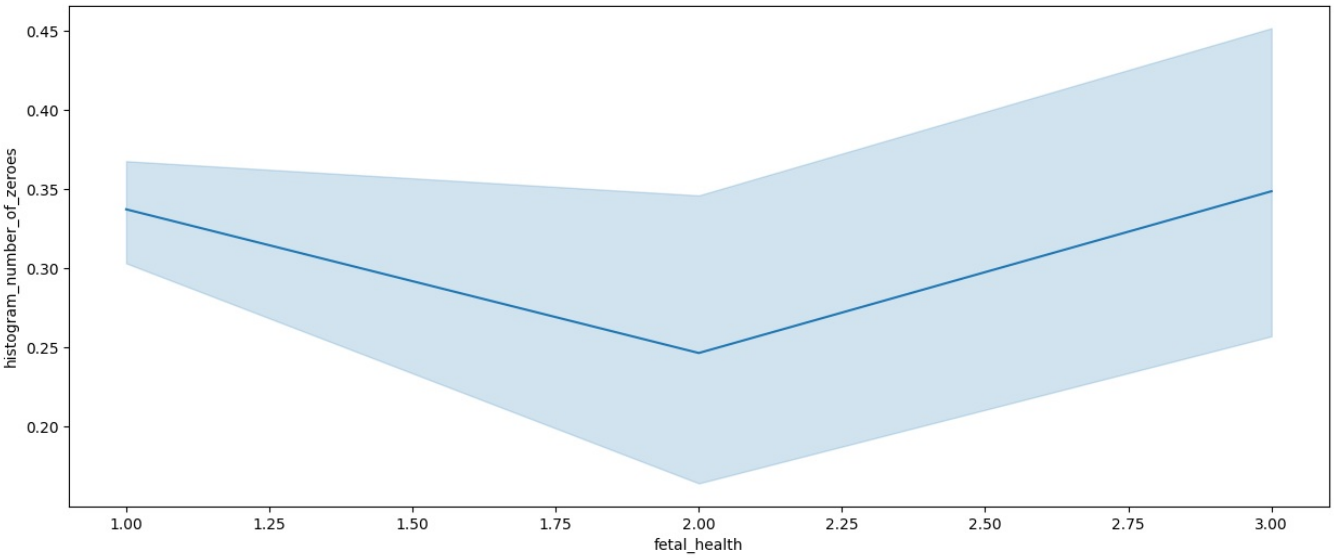

mean_value_of_long_term_variability
9
8
7
6
5
4
3
1.00
1.25
1.50
1.75
2.00
2.25
2.50
2.75
3.00
fetal_health

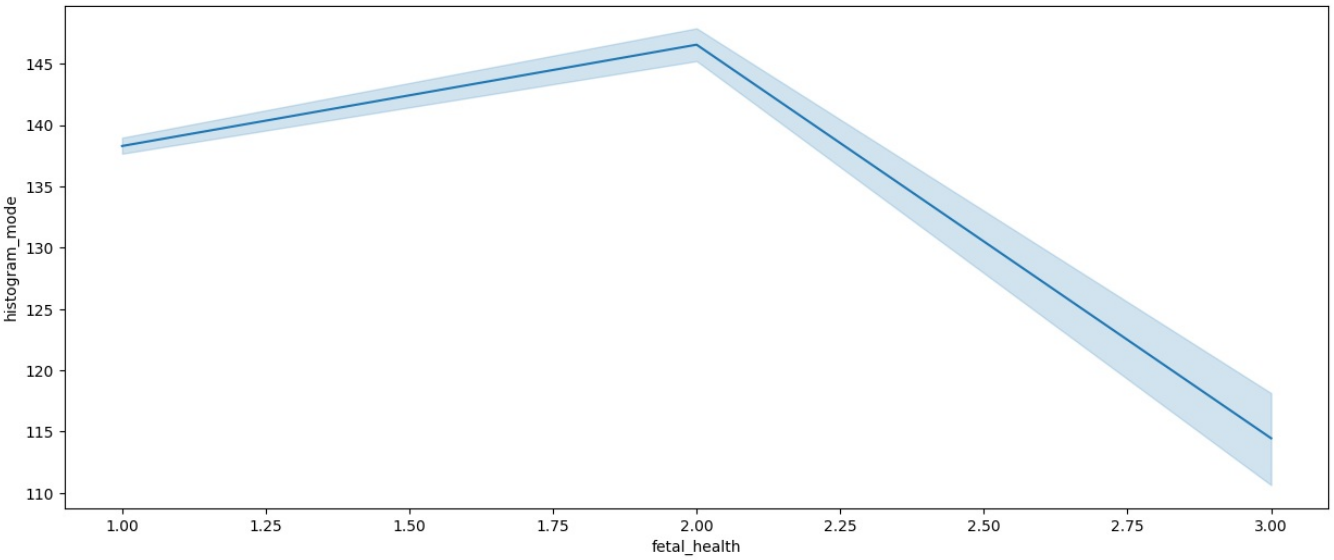

histogram_width
85
80
75
70
65
60
55
50
45
1.00
1.25
1.50
1.75
2.00
2.25
2.50
2.75
3.00
fetal_health

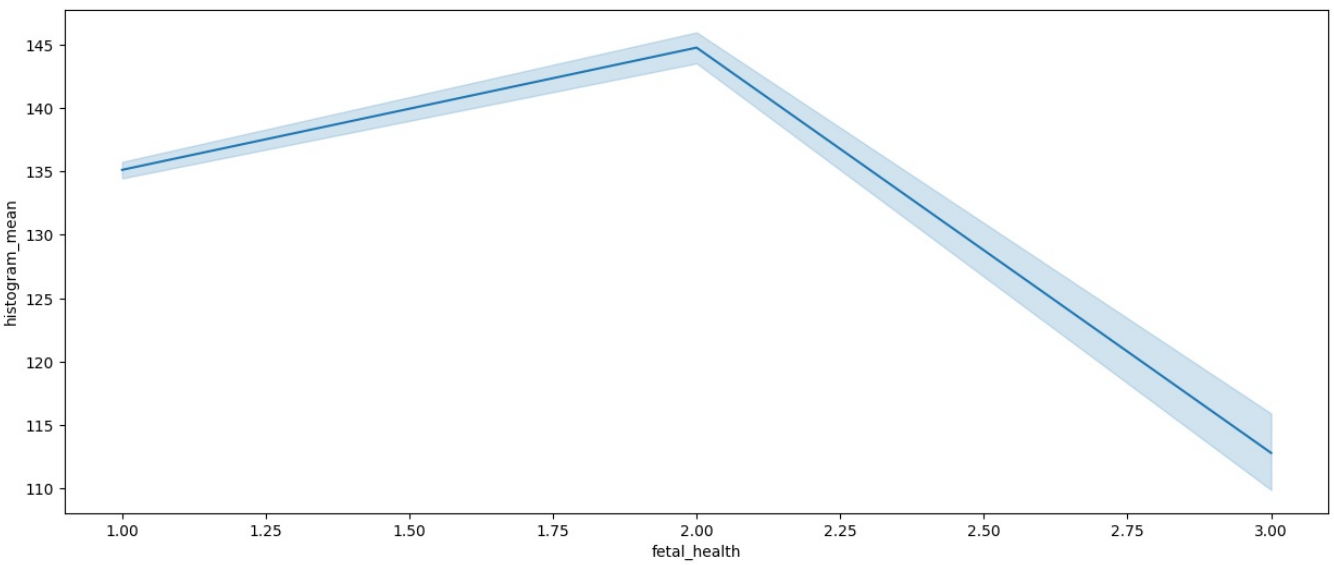

histogram_min
115
110
105
100
95
90
85
80
1.00
1.25
1.50
1.75
2.00
2.25
2.50
2.75
3.00
fetal_health

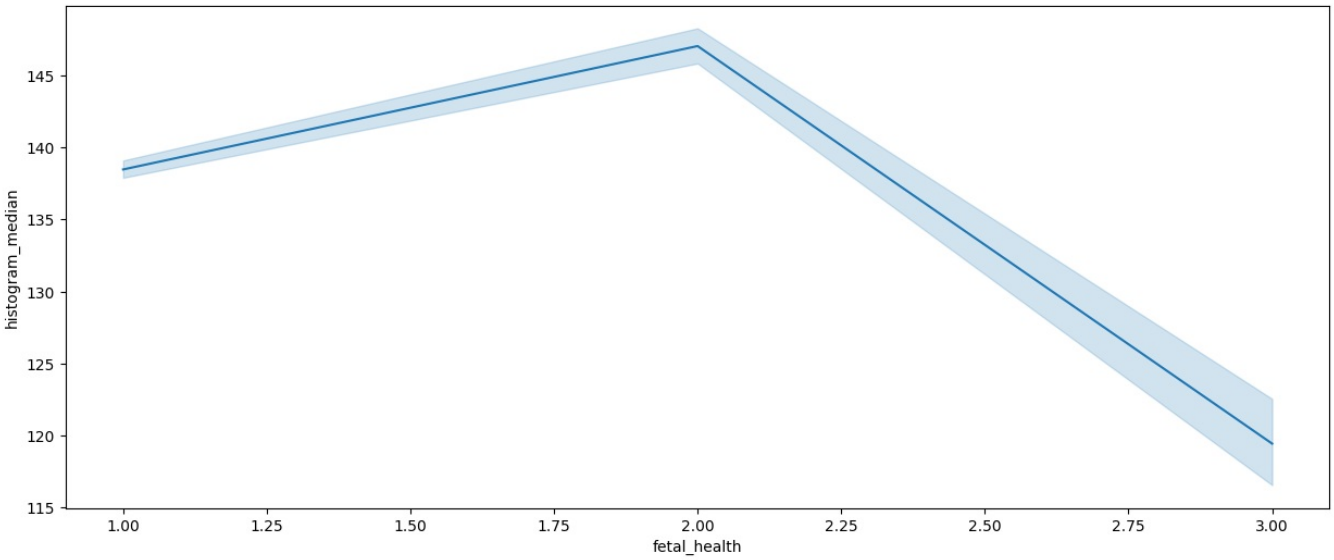

histogram_max
165
164
163
162
161
160
159
1.00
1.25
1.50
1.75
2.00
2.25
2.50
2.75
3.00
fetal_health

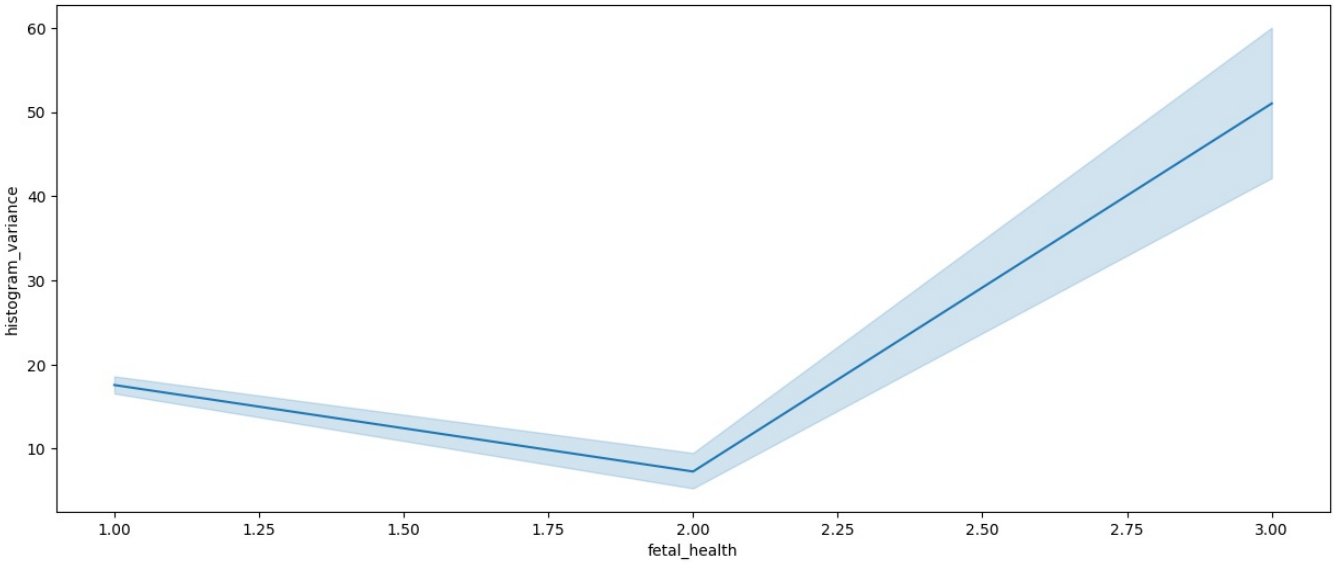

histogram_number_of_peaks
5.00
4.75
4.50
4.25
4.00
3.75
3.50
3.25
3.00
1.00
1.25
1.50
1.75
2.00
2.25
2.50
2.75
3.00
fetal_health

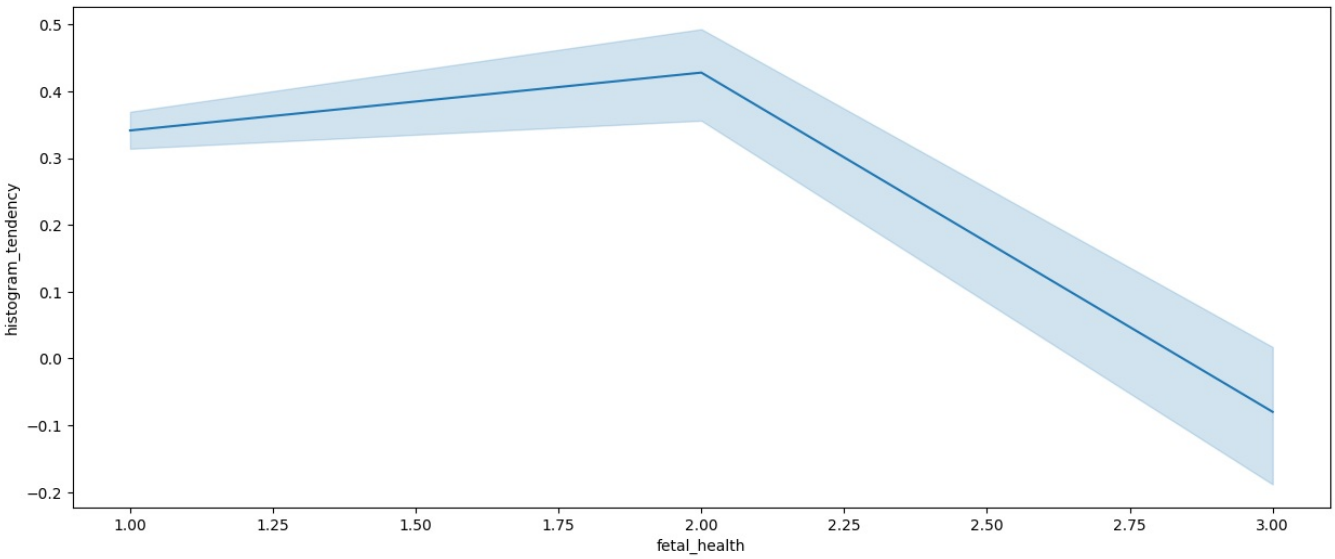

histogram_number_of_zeroes
0.45
0.40
0.35
0.30
0.25
0.20
1.00
1.25
1.50
1.75
2.00
2.25
2.50
2.75
3.00
fetal_health

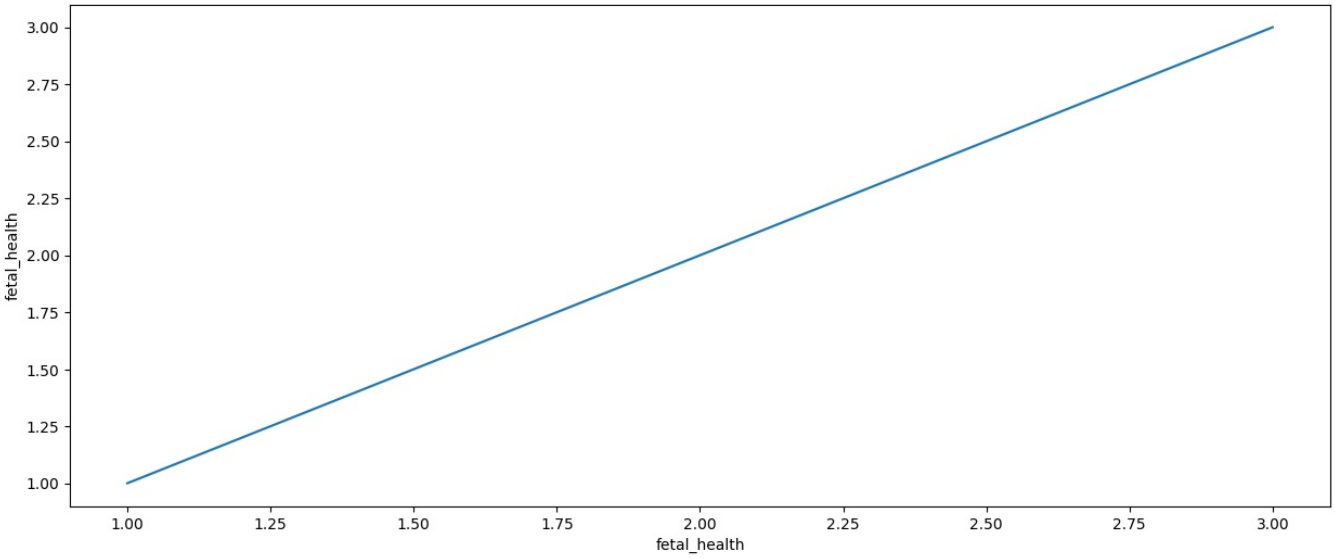

histogram_mode
145
140
135
130
125
120
115
110
1.00
1.25
1.50
1.75
2.00
2.25
2.50
2.75
3.00
fetal_health


histogram_mean
145
140
135
130
125
120
115
110
1.00
1.25
1.50
1.75
2.00
2.25
2.50
2.75
3.00
fetal_health


histogram_median
145
140
135
130
125
120
115
1.00
1.25
1.50
1.75
2.00
2.25
2.50
2.75
3.00
fetal_health


histogram_variance
60
50
40
30
20
10
1.00
1.25
1.50
1.75
2.00
2.25
2.50
2.75
3.00
fetal_health


histogram_tendency
0.5
0.4
0.3
0.2
0.1
0.0
−0.1
−0.2
1.00
1.25
1.50
1.75
2.00
2.25
2.50
2.75
3.00
fetal_health



In [23]:
```
Q1 = df.quantile(0.25)
Q3 = df.quantile(0.75)
IQR = Q3 - Q1
print(IQR)
```

```
baseline value                                          14.000
accelerations                                            0.006
fetal_movement                                           0.003
uterine_contractions                                     0.005
light_decelerations                                      0.003
severe_decelerations                                     0.000
prolongued_decelerations                                 0.000
abnormal_short_term_variability                         29.000
mean_value_of_short_term_variability                     1.000
percentage_of_time_with_abnormal_long_term_variability  11.000
mean_value_of_long_term_variability                      6.200
histogram_width                                         63.000
histogram_min                                           53.000
histogram_max                                           22.000
histogram_number_of_peaks                                4.000
histogram_number_of_zeroes                               0.000
histogram_mode                                          19.000
histogram_mean                                          20.000
histogram_median                                        19.000
histogram_variance                                      22.000
histogram_tendency                                       1.000
fetal_health                                             0.000
dtype: float64
```

In [24]:
```
df_new = df[~((df < (Q1 - 1.5 * IQR)) | (df > (Q3 + 1.5 * IQR))).any(axis = 1)]
```

In [25]:
```
df_new.shape
```

Out[25]:
```
(824, 22)
```

In [26]:
```
df.shape
```

Out[26]:
```
(2113, 22)
```

In [27]:
```
import numpy as np

def detect_outliers_zscore(data, threshold=3):
    """
    Detect outliers using the Z-score method.

    Parameters:
    -----------
    data : array-like
        Input data.
    threshold : float, optional (default=3)
        The threshold value for the Z-score.

    Returns:
    --------
    A boolean array with True for the positions of outliers.
    """
    z_scores = np.abs((data - np.mean(data)) / np.std(data))
    return z_scores > threshold
```

In [28]:
```
import numpy as np

# Generate some random data
data = np.random.normal(loc=0, scale=1, size=100)
```

```
# Add an outlier
data[0] = 10

# Detect outliers using the Z-score method
outliers = detect_outliers_zscore(data)

# Print the indices of the outliers
print(np.where(outliers))
```

```
(array([0], dtype=int64),)
```

In [29]:
```
df_corr= df.corr()
```

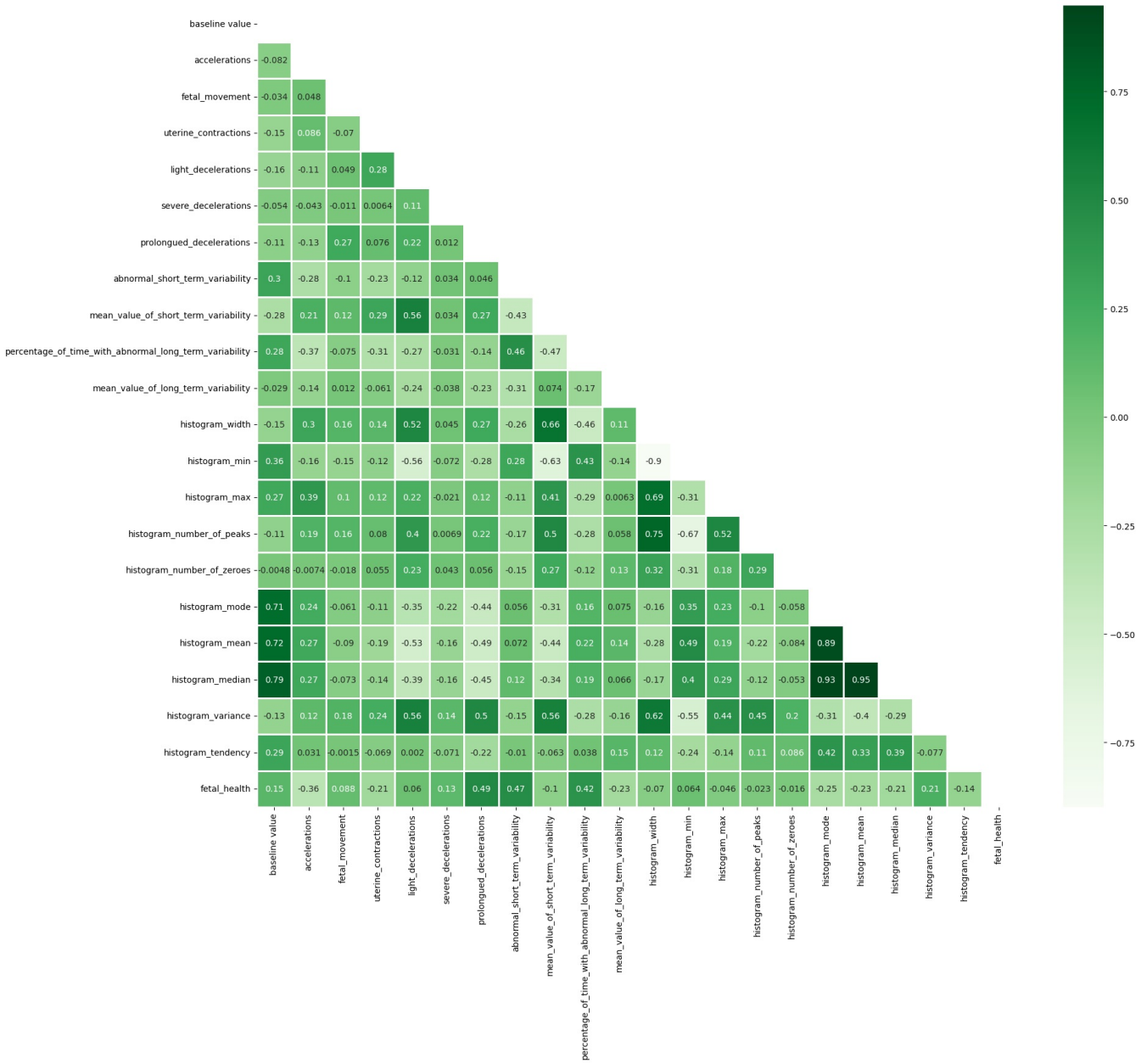
In [30]:
```
df_corr
```

Out[30]:

	baseline value	accelerations	fetal_movement	uterine_contractions	light_decelerations	sev
baseline value	1.000000	-0.081885	-0.033949	-0.149587	-0.159836	
accelerations	-0.081885	1.000000	0.048114	0.086174	-0.110595	
fetal_movement	-0.033949	0.048114	1.000000	-0.069867	0.048795	
uterine_contractions	-0.149587	0.086174	-0.069867	1.000000	0.282325	
light_decelerations	-0.159836	-0.110595	0.048795	0.282325	1.000000	
severe_decelerations	-0.053706	-0.043237	-0.011022	0.006416	0.107483	
prolongued_decelerations	-0.105003	-0.128682	0.265802	0.075533	0.224888	
abnormal_short_term_variability	0.303502	-0.280495	-0.104876	-0.234868	-0.119912	
mean_value_of_short_term_variability	-0.278344	0.206762	0.121532	0.289004	0.562265	
percentage_of_time_with_abnormal_long_term_variability	0.283918	-0.373507	-0.074900	-0.306057	-0.271542	
mean_value_of_long_term_variability	-0.028901	-0.141413	0.011749	-0.061488	-0.241392	
histogram_width	-0.147150	0.298350	0.162803	0.141496	0.520556	
histogram_min	0.360129	-0.155306	-0.154297	-0.115445	-0.555570	
histogram_max	0.273402	0.392684	0.099703	0.117391	0.216314	
histogram_number_of_peaks	-0.113242	0.189209	0.164645	0.080074	0.397093	
histogram_number_of_zeroes	-0.004807	-0.007360	-0.018122	0.054975	0.233940	
histogram_mode	0.708074	0.243083	-0.061496	-0.107352	-0.348386	
histogram_mean	0.722152	0.270266	-0.089938	-0.189689	-0.528620	
histogram_median	0.788487	0.272507	-0.072676	-0.143001	-0.389898	
histogram_variance	-0.134458	0.124433	0.179115	0.236030	0.563352	
histogram_tendency	0.294412	0.030670	-0.001459	-0.069177	0.001977	
fetal_health	0.146077	-0.363947	0.088057	-0.205117	0.059651	

22 rows × 22 columns

In [31]:
```
plt.figure(figsize=(20, 17))
matrix = np.triu(df_corr)
sns.heatmap(df_corr, annot=True, linewidth=.8, mask=matrix, cmap="Greens");
plt.show()
```



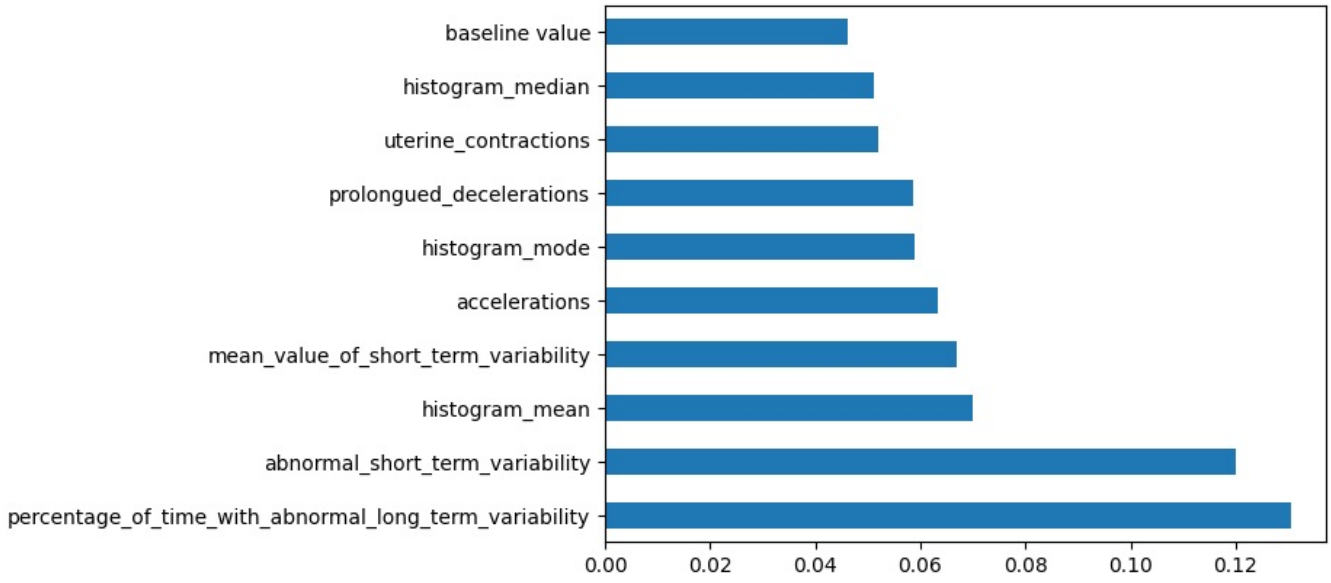
In [32]:
```
X = df.drop('fetal_health',axis=1)
y =df['fetal_health']
```

In [33]:
```
from sklearn.ensemble import ExtraTreesClassifier
model = ExtraTreesClassifier()
model.fit(X,y)
print(model.feature_importances_)
```

```
[0.04620224 0.06334155 0.02809479 0.05190644 0.01516886 0.00220232
 0.05861738 0.12006312 0.06705327 0.13065495 0.03626855 0.0393669
 0.0390858  0.02980682 0.02387896 0.01094209 0.05888911 0.06996851
 0.05107105 0.0339248  0.0234925 ]
```

In [34]:
```
X = df.iloc[:,:-1]
```

In [35]:
```
feat_importances = pd.Series(model.feature_importances_,index=X.columns)
feat_importances.nlargest(10).plot(kind='barh')
plt.show()
```



In [36]:
```
top_10 = pd.DataFrame({'Feature Importance': feat_importances.nlargest(10)})
```

In [37]:
```
top_10
```

Out[37]:

	Feature Importance
percentage_of_time_with_abnormal_long_term_variability	0.130655
abnormal_short_term_variability	0.120063
histogram_mean	0.069969
mean_value_of_short_term_variability	0.067053
accelerations	0.063342
histogram_mode	0.058889
prolongued_decelerations	0.058617
uterine_contractions	0.051906
histogram_median	0.051071
baseline value	0.046202

In [38]:
```
from sklearn.preprocessing import MinMaxScaler
scaler = MinMaxScaler()
```

In [39]:
```
x = scaler.fit_transform(X)
```

In [40]:
```
from sklearn.model_selection import train_test_split
```

In [41]:
```
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size= 0.25, random_state= 42, stratify = y)
```

In [42]:
```
from sklearn.metrics import accuracy_score, precision_score, recall_score, f1_score, confusion_matrix, classifi
```

In [43]:
```
from sklearn.linear_model import LogisticRegression
lr = LogisticRegression()
lr.fit(X_train, y_train)
```

Out[43]:
```
▾ LogisticRegression
LogisticRegression()
```

In [44]:
```
y_pred = lr.predict(X_test)
```

In [45]:
```
print("Accuracy:", accuracy_score(y_test, y_pred))
```

```
Accuracy: 0.8412098298676749
```

In [46]:
```
from sklearn.tree import DecisionTreeClassifier
dt = DecisionTreeClassifier()
dt.fit(X_train, y_train)
```

Out[46]:
```
▾ DecisionTreeClassifier
DecisionTreeClassifier()
```

In [47]:
```
y_pred = dt.predict(X_test)
```

In [48]:
```
print("Accuracy:", accuracy_score(y_test,y_pred))
```

```
Accuracy: 0.9395085066162571
```

In [49]:
```
from sklearn.ensemble import RandomForestClassifier
rfc = RandomForestClassifier()
rfc.fit(X_train, y_train)
```

Out[49]:
```
▾ RandomForestClassifier
RandomForestClassifier()
```

In [52]:
```
y_pred = rfc.predict(X_test)
```

In [53]:
```
print("Accuracy:", accuracy_score(y_test,y_pred))
```

```
Accuracy: 0.947069943289225
```

In [ ]: