PCA in crabs dataset

August 18, 2023

1 Import Libraries

```
[]: import pandas as pd
     pd.set_option('display.precision',3)
     import io
     from google.colab import files
     import matplotlib.pyplot as plt
     import seaborn as sns
[]: uploaded = files.upload()
    <IPython.core.display.HTML object>
    Saving 3 - crabs.csv to 3 - crabs.csv
[]: crabs_data = pd.read_csv("3 - crabs.csv")
     crabs_data
[]:
                          FL
                                RW
                                      CL
                                            CW
                                                  BD
                 index
         sp sex
         В
              М
                     1
                         8.1
                               6.7
                                    16.1
                                          19.0
                                                 7.0
         В
                     2
                         8.8
                                          20.8
     1
             М
                               7.7
                                    18.1
                                                 7.4
     2
         В
                     3
                         9.2
                               7.8
                                    19.0 22.4
                                                 7.7
             М
     3
         В
                     4
                         9.6
                               7.9
                                    20.1 23.1
             М
                                                 8.2
         В
                     5
                         9.8
                               8.0 20.3 23.0
             М
                                                 8.2
                              18.0 41.2 46.2
     195
              F
                        21.4
                                                18.7
         0
                    46
     196 0
              F
                        21.7
                              17.1 41.7 47.2 19.6
                    47
                        21.9
                              17.2 42.6 47.4 19.5
     197 0
              F
                    48
                              17.2 43.0 48.7
                        22.5
     198
              F
                    49
                                                19.8
     199
         0
                    50
                        23.1
                              20.2 46.2 52.5
                                                21.1
     [200 rows x 8 columns]
[]: crabs_data = pd.read_csv("3 - crabs.csv")
     crabs_data.head()
[]:
               index
                       FL
                            RW
                                  CL
                                        CW
                                             BD
       sp sex
                                      19.0
                                            7.0
      В
           Μ
                   1
                      8.1
                           6.7
                                16.1
                   2 8.8 7.7
     1
       В
           Μ
                                18.1
                                      20.8
                                           7.4
```

```
2 B
           Μ
                  3 9.2 7.8 19.0 22.4 7.7
    3 B
                  4 9.6 7.9 20.1
                                     23.1 8.2
           Μ
    4 B
           М
                  5 9.8 8.0 20.3 23.0 8.2
[]: crabs_data = pd.read_csv("3 - crabs.csv")
    crabs_data.tail()
[]:
                                     CL
                                           CW
                index
                         FL
                               RW
                                                BD
        sp sex
    195 0
             F
                   46
                      21.4
                            18.0
                                  41.2 46.2
                                              18.7
                       21.7
                            17.1 41.7 47.2 19.6
    196 0
                   47
    197 0
             F
                   48
                      21.9
                            17.2 42.6 47.4 19.5
    198 0
             F
                   49
                       22.5
                            17.2 43.0 48.7 19.8
    199 0
             F
                   50 23.1 20.2 46.2 52.5
                                              21.1
[]: crabs_data = pd.read_csv("3 - crabs.csv")
     crabs_data.shape
[]: (200, 8)
[]: crabs_data = pd.read_csv("3 - crabs.csv")
    crabs_data.info()
    <class 'pandas.core.frame.DataFrame'>
    RangeIndex: 200 entries, 0 to 199
    Data columns (total 8 columns):
        Column Non-Null Count Dtype
                _____
                                ____
     0
         sp
                 200 non-null
                                object
     1
         sex
                200 non-null
                                object
     2
         index
                200 non-null
                                int64
     3
        FL
                200 non-null
                                float64
     4
        R.W
                200 non-null
                                float64
     5
        CL
                200 non-null
                               float64
        CW
     6
                200 non-null
                                float64
         BD
                200 non-null
                                float64
    dtypes: float64(5), int64(1), object(2)
    memory usage: 12.6+ KB
[]: crabs_data = pd.read_csv("3 - crabs.csv")
    crabs_data.info
[]: <bound method DataFrame.info of
                                                        FL
                                                              RW
                                                                    CL
                                                                          CW
                                                                                BD
                                        sp sex
                                               index
                              6.7 16.1 19.0
                                                7.0
    0
         B
                    1
                        8.1
    1
         В
             Μ
                    2
                        8.8
                              7.7 18.1 20.8
                                                7.4
    2
                    3
                        9.2
                              7.8 19.0 22.4
             M
                                               7.7
    3
         В
                    4
                        9.6
                              7.9 20.1 23.1
                                               8.2
             Μ
                        9.8
                              8.0 20.3 23.0
                    5
    4
         В
             Μ
                                               8.2
```

```
195
         0
               F
                      46
                          21.4
                                 18.0
                                       41.2 46.2
                                                     18.7
     196
                          21.7
                                 17.1
                                      41.7 47.2
                                                     19.6
          0
               F
                      47
                                 17.2
                                       42.6
                                             47.4
     197
               F
                      48
                          21.9
                                                     19.5
               F
                          22.5
                                 17.2
                                       43.0
                                              48.7
     198
                      49
                                                     19.8
     199
           0
                      50
                          23.1
                                 20.2
                                       46.2 52.5
                                                     21.1
     [200 rows x 8 columns]>
[]: crabs_data = pd.read_csv("3 - crabs.csv")
     crabs_data.T
                                               5
[]:
              0
                           2
                                  3
                                         4
                                                      6
                                                             7
                                                                   8
                                                                          9
                                                                                    190
                В
                       В
                             В
                                    В
                                           В
                                                 В
                                                        В
                                                               В
                                                                     В
                                                                            В
                                                                                      0
     sp
                       М
                                                                                      F
     sex
                М
                             Μ
                                    М
                                           М
                                                 М
                                                        М
                                                               М
                                                                     М
                                                                            Μ
                       2
                                                        7
                                                                      9
                                                                                     41
     index
                1
                              3
                                    4
                                           5
                                                  6
                                                               8
                                                                           10
     FL
              8.1
                     8.8
                           9.2
                                  9.6
                                         9.8
                                              10.8
                                                     11.1
                                                            11.6
                                                                  11.8
                                                                         11.8
                                                                                   20.3
                                                      9.9
              6.7
                     7.7
                           7.8
                                  7.9
                                               9.0
                                                                                   16.0
     RW
                                         8.0
                                                             9.1
                                                                   9.6
                                                                         10.5
     CL
             16.1
                   18.1
                          19.0
                                 20.1
                                                     23.8
                                                           24.5
                                                                  24.2
                                                                         25.2
                                                                                   39.4
                                       20.3
                                              23.0
                                                                               •••
     CW
             19.0
                    20.8
                          22.4
                                 23.1
                                       23.0
                                              26.5
                                                     27.1
                                                            28.4
                                                                  27.8
                                                                         29.3
                                                                                   44.1
     BD
              7.0
                    7.4
                                  8.2
                                                      9.8
                                                           10.4
                                                                         10.3
                           7.7
                                         8.2
                                               9.8
                                                                   9.7
                                                                                   18.0
              191
                     192
                           193
                                  194
                                         195
                                               196
                                                      197
                                                             198
                                                                   199
                0
                       0
                             0
                                    0
                                           0
                                                 0
                                                        0
                                                               0
                                                                     0
     sp
     sex
                F
                       F
                             F
                                    F
                                           F
                                                 F
                                                        F
                                                               F
                                                                     F
               42
                      43
                            44
                                   45
                                          46
                                                47
                                                              49
                                                                    50
     index
                                                       48
     FL
             20.5
                   20.6
                          20.9
                                 21.3
                                       21.4
                                              21.7
                                                     21.9
                                                           22.5
                                                                  23.1
     RW
             17.5
                   17.5
                          16.5
                                 18.4
                                       18.0
                                              17.1
                                                     17.2
                                                           17.2
                                                                  20.2
     CL
             40.0
                   41.5
                          39.9
                                 43.8
                                       41.2 41.7
                                                     42.6
                                                           43.0
                                                                  46.2
     CW
             45.5
                   46.2
                          44.7
                                 48.4
                                       46.2 47.2
                                                     47.4
                                                           48.7
                                                                  52.5
```

[8 rows x 200 columns]

19.2

17.5

20.0

19.2

BD

19.6

19.5

19.8

21.1

18.7

```
[]:
                                   Frontal Lobe Length Rear Width
                                                                       Carepace Length \
         species
                      sex
                            index
                                                                  6.7
                                                                                    16.1
            BLUE
                     MALE
                                1
                                                     8.1
     0
                     MALE
                                2
                                                     8.8
                                                                  7.7
                                                                                    18.1
     1
            BLUE
            BLUE
                     MALE
                                                     9.2
                                                                  7.8
                                                                                    19.0
     2
                                3
     3
            BLUE
                     MALE
                                4
                                                     9.6
                                                                  7.9
                                                                                    20.1
     4
                     MALE
                                5
            BLUE
                                                     9.8
                                                                  8.0
                                                                                    20.3
```

	195	ORANGE	E FEMA	ALE	46		21.	4	18.0		41.2	
	196	ORANGE	E FEMA	ALE	47		21.	7	17.1		41.7	
	197	ORANGE	E FEMA	ALE	48		21.	9	17.2		42.6	
	198	ORANGE	E FEMA	ALE	49		22.	5	17.2		43.0	
	199	ORANGE	E FEMA	ALE	50		23.	1	20.2		46.2	
		Carepa			Bodyle	_						
	0			9.0		7.0						
	1			8.0		7.4						
	2			2.4		7.7						
	3			3.1		8.2						
	4		23	3.0		8.2						
					•••	40.7						
	195			6.2		18.7						
	196			7.2		19.6						
	197			7.4		19.5						
	198			3.7		19.8						
	199		52	2.5	-	21.1						
	[200	rows 2	r 8 co	lumna	1							
	[200	TOWS 2	. 0 00.	Lumins.	J							
F 7	crab	s_data	.head())								
[]:		_										
				. ,					a			
[]:	sp	ecies	sex			ntal Lo	obe Length R					
	sp 0	ecies BLUE	MALE	:	1	ntal Lo	8.1	6.7	•	16	. 1	
	sp 0 1	ecies BLUE BLUE	MALE MALE	:	1 2	ntal Lo	8.1 8.8	6.7 7.7	,	16 18	. 1 . 1	
	sp 0 1 2	ecies BLUE BLUE BLUE	MALE MALE MALE	:	1 2 3	ntal Lo	8.1 8.8 9.2	6.7 7.7 7.8	, , ,	16 18 19	. 1 . 1 . 0	
	sp 0 1 2 3	BLUE BLUE BLUE BLUE BLUE	MALE MALE MALE MALE	2	1 2 3 4	ntal Lo	8.1 8.8 9.2 9.6	6.7 7.7 7.8	, , , ,	16 18 19 20	.1 .1 .0	
	sp 0 1 2	ecies BLUE BLUE BLUE	MALE MALE MALE	2	1 2 3	ntal Lo	8.1 8.8 9.2	6.7 7.7 7.8	, , , ,	16 18 19	.1 .1 .0	
	sp 0 1 2 3 4	ecies BLUE BLUE BLUE BLUE BLUE	MALE MALE MALE MALE MALE	; ; ;	1 2 3 4 5		8.1 8.8 9.2 9.6	6.7 7.7 7.8	, , , ,	16 18 19 20	.1 .1 .0	
	sp 0 1 2 3 4	BLUE BLUE BLUE BLUE BLUE	MALE MALE MALE MALE MALE	a Boo	1 2 3 4 5 dyleng ^a		8.1 8.8 9.2 9.6	6.7 7.7 7.8	, , , ,	16 18 19 20	.1 .1 .0	
	sp 0 1 2 3 4	ecies BLUE BLUE BLUE BLUE BLUE	MALE MALE MALE MALE MALE MALE	n Boo	1 2 3 4 5 dyleng ^r 7	th	8.1 8.8 9.2 9.6	6.7 7.7 7.8	, , , ,	16 18 19 20	.1 .1 .0	
	sp 0 1 2 3 4 C 0	ecies BLUE BLUE BLUE BLUE BLUE	MALE MALE MALE MALE MALE MALE MALE 19.0	i i i i i i i i i i i i i i i i i i i	1 2 3 4 5 dyleng [*] 7	th .0	8.1 8.8 9.2 9.6	6.7 7.7 7.8	, , , ,	16 18 19 20	.1 .1 .0	
	sp 0 1 2 3 4 C 0 1	ecies BLUE BLUE BLUE BLUE BLUE	MALE MALE MALE MALE MALE MALE MALE 19.0	n Boo 3 4 5 7 8 8	1 2 3 4 5 dyleng [*] 7 7	th .0 .4	8.1 8.8 9.2 9.6	6.7 7.7 7.8	, , , ,	16 18 19 20	.1 .1 .0	
	sp 0 1 2 3 4 C C C C 1 2	ecies BLUE BLUE BLUE BLUE BLUE	MALE MALE MALE MALE MALE MALE MALE 20.8	1 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2	1 2 3 4 5 dyleng ^r 7 7	th .0 .4	8.1 8.8 9.2 9.6	6.7 7.7 7.8	, , , ,	16 18 19 20	.1 .1 .0	
[]:	sp 0 1 2 3 4 C 0 1 2 3 4	ecies BLUE BLUE BLUE BLUE BLUE	MALE MALE MALE MALE MALE MALE 19.0 20.8 22.4 23.0	n Boo 0 3 4 1	1 2 3 4 5 dyleng ^r 7 7	th .0 .4 .7	8.1 8.8 9.2 9.6	6.7 7.7 7.8	, , , ,	16 18 19 20	.1 .1 .0	
	sp 0 1 2 3 4 C 0 1 2 3 4	ecies BLUE BLUE BLUE BLUE BLUE	MALE MALE MALE MALE MALE MALE 19.0 20.8 22.4 23.0	n Boo 0 3 4 1	1 2 3 4 5 dyleng ^r 7 7	th .0 .4 .7	8.1 8.8 9.2 9.6	6.7 7.7 7.8	, , , ,	16 18 19 20	.1 .1 .0	
[]:	sp 0 1 2 3 4 C C C C C C C C C C C C C C C C C C	ecies BLUE BLUE BLUE BLUE BLUE	MALE MALE MALE MALE MALE MALE 20.8 22.4 23.6 23.6	h Boo 0 3 4 1 1	1 2 3 4 5 dyleng ^r 7 7	th .0 .4 .7 .2	8.1 8.8 9.2 9.6	6.7 7.8 7.9 8.0	3333	16 18 19 20	.1 .0 .1 .3	\
[]:	sp 0 1 2 3 4 C C C C C C C C C C C C C C C C C C	ecies BLUE BLUE BLUE BLUE BLUE	MALE MALE MALE MALE MALE MALE Width 19.0 20.8 22.4 23.0 .tail()	h Boo D 33 44 11 D	1 2 3 4 5 dyleng 7 7 7 8 8	th .0 .4 .7 .2	8.1 8.8 9.2 9.6 9.8	6.7 7.8 7.9 8.0	3333	16 18 19 20 20	.1 .0 .1 .3	\
[]:	sp 0 1 2 3 4 C 0 1 2 3 4 C c r a b	ecies BLUE BLUE BLUE BLUE BLUE carepace	MALE MALE MALE MALE MALE MALE Width 19.0 20.8 23.0 .tail()	n Bood	1 2 3 4 5 dyleng 7 7 7 8 8	th .0 .4 .7 .2	8.1 8.8 9.2 9.6 9.8	6.7 7.8 7.9 8.0	vidth	16 18 19 20 20	.1 .0 .1 .3	\
[]:	sp 0 1 2 3 4 0 1 2 3 4 crab	ecies BLUE BLUE BLUE BLUE BLUE sarepace s_data species	MALE MALE MALE MALE MALE MALE Width 19.0 20.8 22.4 23.0 .tail() E FEMA E FEMA	h Boo O B 4 1 O Sex :	1 2 3 4 5 dyleng [*] 7 7 7 8 8	th .0 .4 .7 .2	8.1 8.8 9.2 9.6 9.8	6.7 7.8 7.9 8.0 h Rear W	7idth 18.0	16 18 19 20 20	.1 .0 .1 .3	\
[]:	sp 0 1 2 3 4 0 1 2 3 4 crab	ecies BLUE BLUE BLUE BLUE BLUE carepace s_data species ORANGE	MALE MALE MALE MALE MALE MALE Width 19.0 20.8 23.0 .tail() E FEMA E FEMA E FEMA	h Boo D B 4 1 D Sex : ALE ALE	1 2 3 4 5 dyleng 7 7 7 8 8 8 index 46 47	th .0 .4 .7 .2	8.1 8.8 9.2 9.6 9.8	6.7 7.8 7.9 8.0 h Rear W	7idth 18.0	16 18 19 20 20	.1 .0 .1 .3 Length 41.2 41.7	\

```
195
                     46.2
                                  18.7
                     47.2
     196
                                  19.6
     197
                     47.4
                                  19.5
     198
                     48.7
                                  19.8
     199
                     52.5
                                  21.1
[]: crabs_data.describe(include='all')
[]:
                                      Frontal Lobe Length
                                                            Rear Width
            species
                       sex
                              index
     count
                 200
                       200
                            200.000
                                                   200.000
                                                                200.000
     unique
                   2
                         2
                                 NaN
                                                       NaN
                                                                    NaN
     top
               BLUE
                     MALE
                                 NaN
                                                       NaN
                                                                    NaN
     freq
                 100
                       100
                                 NaN
                                                       NaN
                                                                    NaN
     mean
                       NaN
                             25.500
                                                                 12.738
                NaN
                                                    15.583
     std
                NaN
                       NaN
                             14.467
                                                     3.495
                                                                  2.573
     min
                NaN
                       NaN
                              1.000
                                                     7.200
                                                                  6.500
     25%
                NaN
                       NaN
                             13.000
                                                    12.900
                                                                 11.000
     50%
                NaN
                       NaN
                             25.500
                                                    15.550
                                                                 12,800
     75%
                             38.000
                                                    18.050
                                                                 14.300
                NaN
                       NaN
     max
                NaN
                       NaN
                             50.000
                                                    23.100
                                                                 20.200
             Carepace Length
                               Carepace Width
                                                 Bodylength
     count
                      200.000
                                       200.000
                                                    200.000
     unique
                          NaN
                                           NaN
                                                        NaN
     top
                          NaN
                                           NaN
                                                        NaN
     freq
                          NaN
                                           NaN
                                                        NaN
     mean
                       32.105
                                        36.415
                                                     14.030
     std
                        7.119
                                         7.872
                                                      3.425
                                                      6.100
                       14.700
     min
                                        17.100
     25%
                       27.275
                                        31.500
                                                     11.400
     50%
                       32.100
                                        36.800
                                                     13.900
     75%
                                        42.000
                                                     16.600
                       37.225
                       47.600
                                        54.600
                                                     21.600
     max
[]: crabs data.columns
[]: Index(['species', 'sex', 'index', 'Frontal Lobe Length', 'Rear Width',
             'Carepace Length', 'Carepace Width', 'Bodylength'],
           dtype='object')
[]: crabs_data.shape
[]: (200, 8)
[]: crabs_data['class'] = crabs_data.species + crabs_data.sex
     crabs_data['class'].value_counts()
```

Carepace Width

Bodylength

```
[]: BLUEMALE 50
BLUEFEMALE 50
ORANGEMALE 50
ORANGEFEMALE 50
```

Name: class, dtype: int64

1.0.1 we will start the basic exploration of dataset

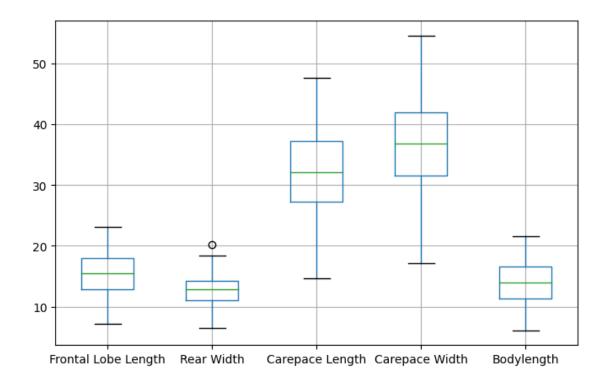
[]:	Frontal Lobe Length	Rear Width	Carepace Length	Carepace Width	١
count	200.000	200.000	200.000	200.000	
mean	15.583	12.738	32.105	36.415	
std	3.495	2.573	7.119	7.872	
min	7.200	6.500	14.700	17.100	
25%	12.900	11.000	27.275	31.500	
50%	15.550	12.800	32.100	36.800	
75%	18.050	14.300	37.225	42.000	
max	23.100	20.200	47.600	54.600	

	Bodylength
count	200.000
mean	14.030
std	3.425
min	6.100
25%	11.400
50%	13.900
75%	16.600
max	21.600

1.0.2 Box plot of the Relevant features

```
[]: fig, ax = plt.subplots(figsize=(8,5))
crabs_data[data_columns].boxplot()
```

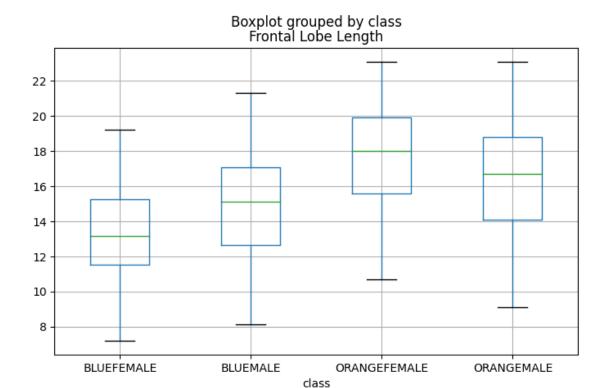
[]: <Axes: >



1.0.3 Initial visualization of classes

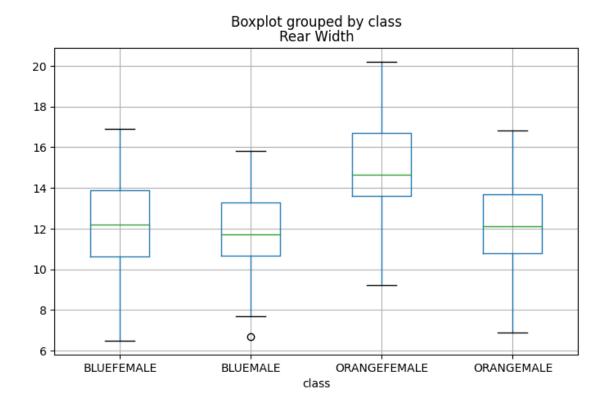
```
[]: crabs_data.boxplot(column = 'Frontal Lobe Length', by = 'class', figsize = ∪ (8,5))
```

[]: <Axes: title={'center': 'Frontal Lobe Length'}, xlabel='class'>



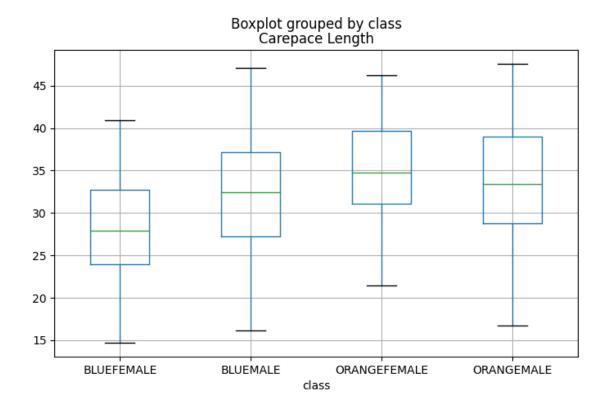
```
[]: crabs_data.boxplot(column = 'Rear Width', by = 'class', figsize = (8,5))
```

[]: <Axes: title={'center': 'Rear Width'}, xlabel='class'>



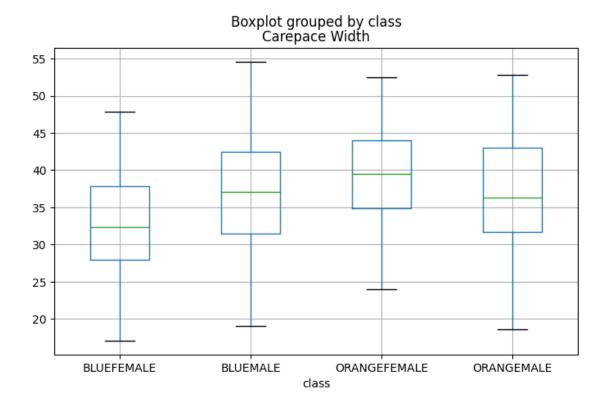
```
[]: crabs_data.boxplot(column = 'Carepace Length', by = 'class', figsize = (8,5))
```

[]: <Axes: title={'center': 'Carepace Length'}, xlabel='class'>



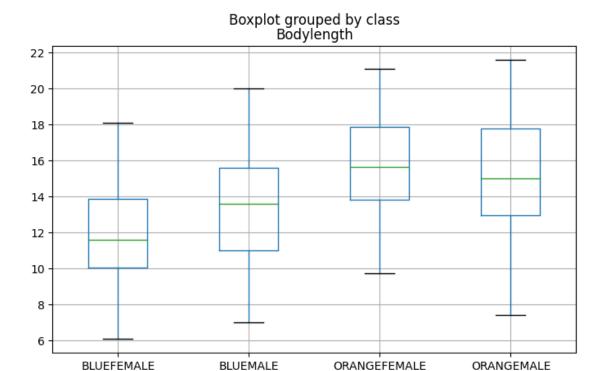
```
[]: crabs_data.boxplot(column = 'Carepace Width', by = 'class', figsize = (8,5))
```

[]: <Axes: title={'center': 'Carepace Width'}, xlabel='class'>



```
[]: crabs_data.boxplot(column = 'Bodylength', by = 'class', figsize = (8,5))
```

[]: <Axes: title={'center': 'Bodylength'}, xlabel='class'>



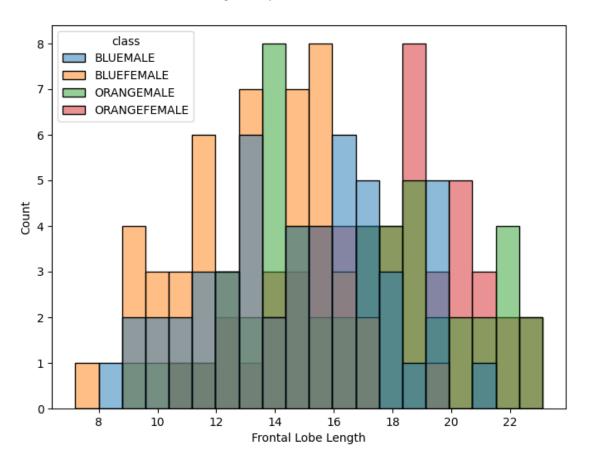
class

1.0.4 Histograms

```
[]: crabs_data[data_columns].hist(figsize=(16,4),layout=(1,6))
[]: array([[<Axes: title={'center': 'Frontal Lobe Length'}>,
              <Axes: title={'center': 'Rear Width'}>,
              <Axes: title={'center': 'Carepace Length'}>,
              <Axes: title={'center': 'Carepace Width'}>,
              <Axes: title={'center': 'Bodylength'}>, <Axes: >]], dtype=object)
            Frontal Lobe Length
                                Rear Width
                                               Carepace Length
                                                                Carepace Width
                                                                                   Bodylength
                                                                              35
                                                             30
          30
                                                                              30
                           40
          25
                                                             25
                                                                              25
          20
                                                             20
                                                                              20
                                            20
          15
                                                             15
                           20
                                            15
                                                             10
          10
                                                                              10
                      20
                                10
                                                20
```

```
[]: plt.figure(figsize=(8,6))
sns.histplot(crabs_data,x="Frontal Lobe Length",hue='class',bins=20)
```

[]: <Axes: xlabel='Frontal Lobe Length', ylabel='Count'>



1.0.5 Pairplots

```
[]: sns.pairplot(crabs_data, hue='class')
```

[]: <seaborn.axisgrid.PairGrid at 0x7da0f5e2e5f0>

