earthquake-prediction-model

October 17, 2023

0.1 EARTHQUAKE PREDICTION MODEL USING PYTHON

---

1 *Uploading_dataset*

In data analysis and machine learning, working with datasets is a fundamental task. To get started, we need to upload our dataset into our Python environment. we have uploaded a dataset using the popular pandas library. Pandas simplifies the process of working with structured data, making it an ideal choice for handling datasets in Python.

[15]:
```
!pip install pandas
import pandas as pd
dataset = pd.read_csv('bronze.csv')
print(dataset.head())    # Display the first few rows
print(dataset.info())    # Display information about columns and data types
print(dataset.describe()) # Display summary statistics
print(dataset)
```

```
Requirement already satisfied: pandas in /usr/local/lib/python3.10/dist-packages
(1.5.3)
Requirement already satisfied: python-dateutil>=2.8.1 in
/usr/local/lib/python3.10/dist-packages (from pandas) (2.8.2)
Requirement already satisfied: pytz>=2020.1 in /usr/local/lib/python3.10/dist-
packages (from pandas) (2023.3.post1)
Requirement already satisfied: numpy>=1.21.0 in /usr/local/lib/python3.10/dist-
packages (from pandas) (1.23.5)
Requirement already satisfied: six>=1.5 in /usr/local/lib/python3.10/dist-
packages (from python-dateutil>=2.8.1->pandas) (1.16.0)
                      time  latitude  longitude  depth  mag magType  nst  \
0  1930-12-08T08:01:02.000Z    23.261    120.277   15.0  6.3      mw  NaN
1  1930-12-03T18:51:47.000Z    18.233     96.298   10.0  7.4      mw  NaN
2  1930-12-02T07:01:30.000Z    25.854     98.356   35.0  6.2      mw  NaN
3  1930-11-28T07:32:56.000Z    18.779   -106.767   15.0  6.3      mw  NaN
4  1930-11-25T19:02:53.000Z    35.050    139.129   15.0  6.9      mw  NaN

   gap  dmin  rms  …                  updated place type  horizontalError  \
0  NaN   NaN  NaN  …  2015-05-13T18:52:43.000Z   NaN  NaN              NaN
```

```
1  NaN   NaN  NaN  …  2015-05-13T18:52:43.000Z   NaN  NaN             NaN
2  NaN   NaN  NaN  …  2015-05-13T18:52:43.000Z   NaN  NaN             NaN
3  NaN   NaN  NaN  …  2015-05-13T18:52:43.000Z   NaN  NaN             NaN
4  NaN   NaN  NaN  …  2015-05-13T18:52:43.000Z   NaN  NaN             NaN

   depthError  magError  magNst  status  locationSource  magSource
0         NaN       NaN     NaN     NaN             NaN        NaN
1         NaN       NaN     NaN     NaN             NaN        NaN
2         NaN       NaN     NaN     NaN             NaN        NaN
3         NaN       NaN     NaN     NaN             NaN        NaN
4         NaN       NaN     NaN     NaN             NaN        NaN

[5 rows x 22 columns]
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 797046 entries, 0 to 797045
Data columns (total 22 columns):
 #   Column           Non-Null Count   Dtype
---  ------           --------------   -----
 0   time             797046 non-null  object
 1   latitude         797046 non-null  float64
 2   longitude        797046 non-null  float64
 3   depth            797041 non-null  float64
 4   mag              797046 non-null  float64
 5   magType          796940 non-null  object
 6   nst              421658 non-null  float64
 7   gap              470629 non-null  float64
 8   dmin             202838 non-null  float64
 9   rms              614095 non-null  float64
 10  net              797046 non-null  object
 11  id               797046 non-null  object
 12  updated          797046 non-null  object
 13  place            0 non-null       float64
 14  type             0 non-null       float64
 15  horizontalError  0 non-null       float64
 16  depthError       0 non-null       float64
 17  magError         0 non-null       float64
 18  magNst           0 non-null       float64
 19  status           0 non-null       float64
 20  locationSource   0 non-null       float64
 21  magSource        0 non-null       float64
dtypes: float64(17), object(5)
memory usage: 133.8+ MB
None
            latitude      longitude          depth            mag  \
count  797046.000000  797046.000000  797041.000000  797046.000000
mean       19.038314     -12.911692      53.726903       3.884296
std        29.219884     118.010192      99.510254       0.911611
min       -84.422000    -179.999000      -4.900000       2.500000
```

```
25%        -4.727000    -118.086000       10.000000        3.000000
50%        29.826000     -52.385000       23.000000        4.000000
75%        39.405000     120.892000       45.620000        4.600000
max        87.221000     180.000000      735.800000        9.500000

                nst            gap           dmin            rms  place  \
count  421658.000000  470629.000000  202838.000000  614095.000000    0.0
mean       33.170883     141.002356       1.757801       0.763249    NaN
std        55.796692      81.447357       3.734481       0.477974    NaN
min         0.000000       0.000000       0.000000       0.000000    NaN
25%         8.000000      75.000000       0.082880       0.420000    NaN
50%        16.000000     125.000000       0.583000       0.810000    NaN
75%        35.000000     195.900000       1.973000       1.070000    NaN
max       934.000000     360.000000     127.420000      69.320000    NaN

       type  horizontalError  depthError  magError  magNst  status  \
count   0.0              0.0         0.0       0.0     0.0     0.0
mean    NaN              NaN         NaN       NaN     NaN     NaN
std     NaN              NaN         NaN       NaN     NaN     NaN
min     NaN              NaN         NaN       NaN     NaN     NaN
25%     NaN              NaN         NaN       NaN     NaN     NaN
50%     NaN              NaN         NaN       NaN     NaN     NaN
75%     NaN              NaN         NaN       NaN     NaN     NaN
max     NaN              NaN         NaN       NaN     NaN     NaN

       locationSource  magSource
count             0.0        0.0
mean              NaN        NaN
std               NaN        NaN
min               NaN        NaN
25%               NaN        NaN
50%               NaN        NaN
75%               NaN        NaN
max               NaN        NaN
                          time  latitude  longitude   depth  mag magType  \
0       1930-12-08T08:01:02.000Z   23.2610   120.2770   15.00  6.3      mw
1       1930-12-03T18:51:47.000Z   18.2330    96.2980   10.00  7.4      mw
2       1930-12-02T07:01:30.000Z   25.8540    98.3560   35.00  6.2      mw
3       1930-11-28T07:32:56.000Z   18.7790  -106.7670   15.00  6.3      mw
4       1930-11-25T19:02:53.000Z   35.0500   139.1290   15.00  6.9      mw
…                            …         …          …      …   …     …
797041  2018-09-01T01:14:38.230Z  -30.4830  -177.9279   43.90  4.3      mb
797042  2018-09-01T01:07:59.120Z  -10.7558   124.3621   10.00  4.0      mb
797043  2018-09-01T01:00:13.810Z   -5.5167   147.1735  217.56  4.6      mb
797044  2018-09-01T00:27:11.440Z   46.8819   155.6566   10.00  4.3      mb
797045  2018-09-01T00:00:47.980Z  -55.7508   -28.3561   10.00  4.8      mb

        nst    gap   dmin   rms  …                  updated place type  \
```

```
0       NaN    NaN    NaN   NaN  …  2015-05-13T18:52:43.000Z   NaN  NaN
1       NaN    NaN    NaN   NaN  …  2015-05-13T18:52:43.000Z   NaN  NaN
2       NaN    NaN    NaN   NaN  …  2015-05-13T18:52:43.000Z   NaN  NaN
3       NaN    NaN    NaN   NaN  …  2015-05-13T18:52:43.000Z   NaN  NaN
4       NaN    NaN    NaN   NaN  …  2015-05-13T18:52:43.000Z   NaN  NaN
…     …    …    …   …  …                          …   …  …
797041  NaN  165.0  1.233  0.87  …  2018-11-07T18:37:12.040Z   NaN  NaN
797042  NaN  112.0  0.998  1.23  …  2018-11-07T18:37:12.040Z   NaN  NaN
797043  NaN  119.0  3.455  0.71  …  2018-11-07T18:37:07.040Z   NaN  NaN
797044  NaN   94.0  6.370  1.21  …  2018-11-07T18:37:12.040Z   NaN  NaN
797045  NaN   78.0  4.905  1.31  …  2018-11-07T18:37:07.040Z   NaN  NaN

        horizontalError  depthError  magError  magNst  status  locationSource  \
0                   NaN         NaN       NaN     NaN     NaN             NaN
1                   NaN         NaN       NaN     NaN     NaN             NaN
2                   NaN         NaN       NaN     NaN     NaN             NaN
3                   NaN         NaN       NaN     NaN     NaN             NaN
4                   NaN         NaN       NaN     NaN     NaN             NaN
…                 …         …       …     …     …               …
797041              NaN         NaN       NaN     NaN     NaN             NaN
797042              NaN         NaN       NaN     NaN     NaN             NaN
797043              NaN         NaN       NaN     NaN     NaN             NaN
797044              NaN         NaN       NaN     NaN     NaN             NaN
797045              NaN         NaN       NaN     NaN     NaN             NaN

        magSource
0             NaN
1             NaN
2             NaN
3             NaN
4             NaN
…           …
797041        NaN
797042        NaN
797043        NaN
797044        NaN
797045        NaN

[797046 rows x 22 columns]
```

2 *Preprocessing*

Before extracting meaningful insights or building accurate machine learning models, we must preprocess our dataset. Data preprocessing involves a series of steps to clean, transform, and structure your data for analysis.

[27]:
```
print(dataset.isnull().sum())




# Specify the columns you want to delete (e.g., 'column1', 'column2')
columns_to_delete =␣
 ↪['place','type','horizontalError','depthError','magError','magNst','status','locationSource

# Use the drop method to delete the specified columns
dataset.drop(columns=columns_to_delete, inplace=True, errors='ignore')

# The specified columns are deleted from the dataset.
print(dataset)
```

```
time               0
latitude           0
longitude          0
depth              5
mag                0
magType          106
nst           375388
gap           326417
dmin          594208
rms           182951
net                0
id                 0
updated            0
type          797046
depthError    797046
magSource     797046
dtype: int64
                          time  latitude  longitude   depth  mag magType  \
0       1930-12-08T08:01:02.000Z   23.2610   120.2770   15.00  6.3      mw
1       1930-12-03T18:51:47.000Z   18.2330    96.2980   10.00  7.4      mw
2       1930-12-02T07:01:30.000Z   25.8540    98.3560   35.00  6.2      mw
3       1930-11-28T07:32:56.000Z   18.7790  -106.7670   15.00  6.3      mw
4       1930-11-25T19:02:53.000Z   35.0500   139.1290   15.00  6.9      mw
…                            …         …          …      …   …      …
797041  2018-09-01T01:14:38.230Z  -30.4830  -177.9279   43.90  4.3      mb
797042  2018-09-01T01:07:59.120Z  -10.7558   124.3621   10.00  4.0      mb
797043  2018-09-01T01:00:13.810Z   -5.5167   147.1735  217.56  4.6      mb
797044  2018-09-01T00:27:11.440Z   46.8819   155.6566   10.00  4.3      mb
797045  2018-09-01T00:00:47.980Z  -55.7508   -28.3561   10.00  4.8      mb
```

```
        nst    gap   dmin   rms     net            id  \
0       NaN    NaN    NaN   NaN  iscgem  iscgem907791
1       NaN    NaN    NaN   NaN  iscgem  iscgem907777
2       NaN    NaN    NaN   NaN  iscgem  iscgem907773
3       NaN    NaN    NaN   NaN  iscgem  iscgem907769
4       NaN    NaN    NaN   NaN  iscgem  iscgem907761
…       …    …    …    …   …      …           …
797041  NaN  165.0  1.233  0.87      us    us2000hafq
797042  NaN  112.0  0.998  1.23      us    us2000hafw
797043  NaN  119.0  3.455  0.71      us    us2000h6as
797044  NaN   94.0  6.370  1.21      us    us2000hafn
797045  NaN   78.0  4.905  1.31      us    us2000h6aq

                        updated
0       2015-05-13T18:52:43.000Z
1       2015-05-13T18:52:43.000Z
2       2015-05-13T18:52:43.000Z
3       2015-05-13T18:52:43.000Z
4       2015-05-13T18:52:43.000Z
…                           …
797041  2018-11-07T18:37:12.040Z
797042  2018-11-07T18:37:12.040Z
797043  2018-11-07T18:37:07.040Z
797044  2018-11-07T18:37:12.040Z
797045  2018-11-07T18:37:07.040Z

[797046 rows x 13 columns]
```

3 *Data Visualization*

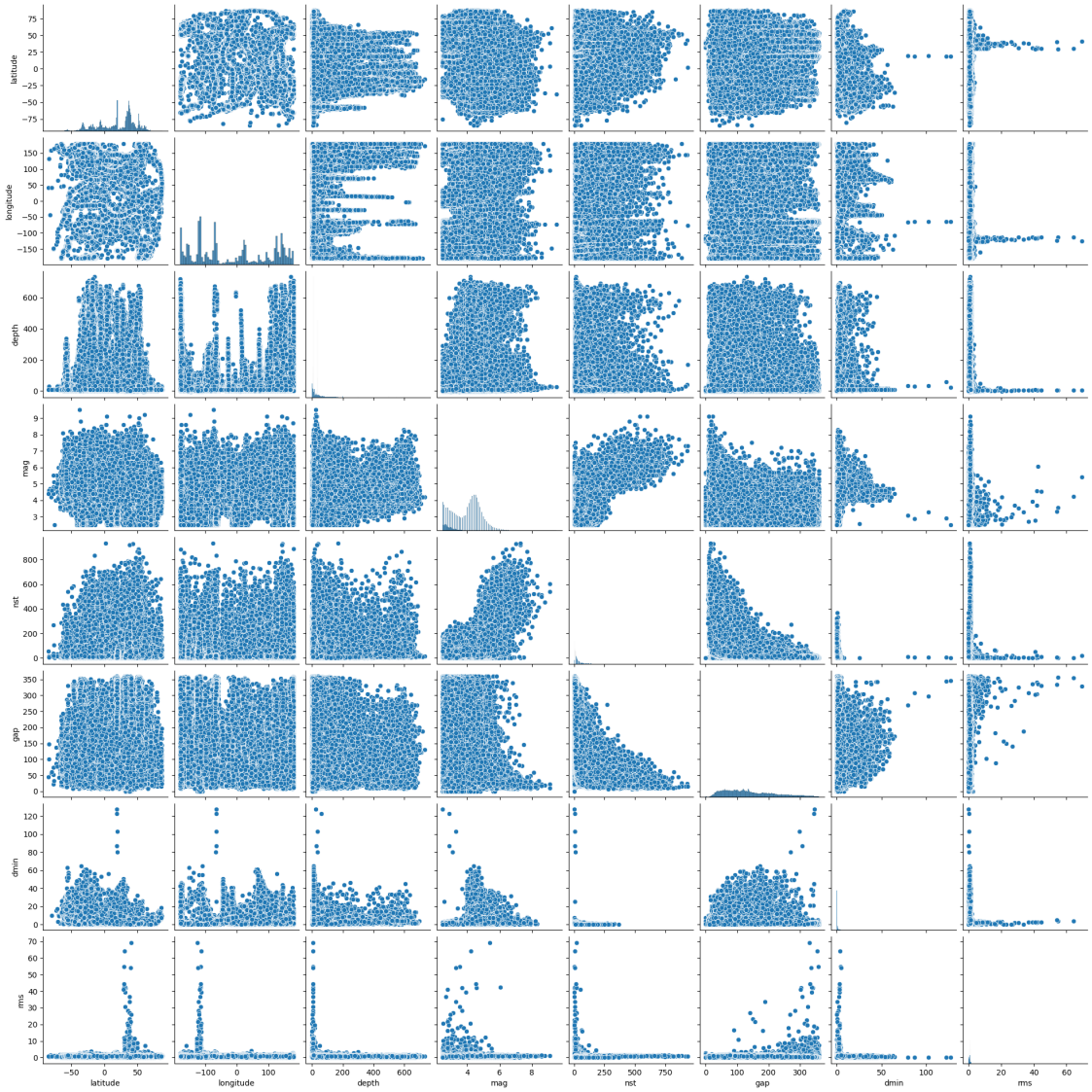
Data visualization is a fundamental part of data analysis. It helps us to understand our dataset, identify patterns, and communicate your findings to others. In this guide, we will explore how to create informative and visually appealing data visualizations using Python, with a focus on the Matplotlib and Seaborn libraries.

[28]:
```
!pip install matplotlib seaborn
import matplotlib.pyplot as plt
import seaborn as sns

print(dataset.head())
print(dataset.describe())
print(dataset.dtypes)
sns.pairplot(dataset)
plt.show()
```

```
Requirement already satisfied: matplotlib in /usr/local/lib/python3.10/dist-
packages (3.7.1)
Requirement already satisfied: seaborn in /usr/local/lib/python3.10/dist-
```

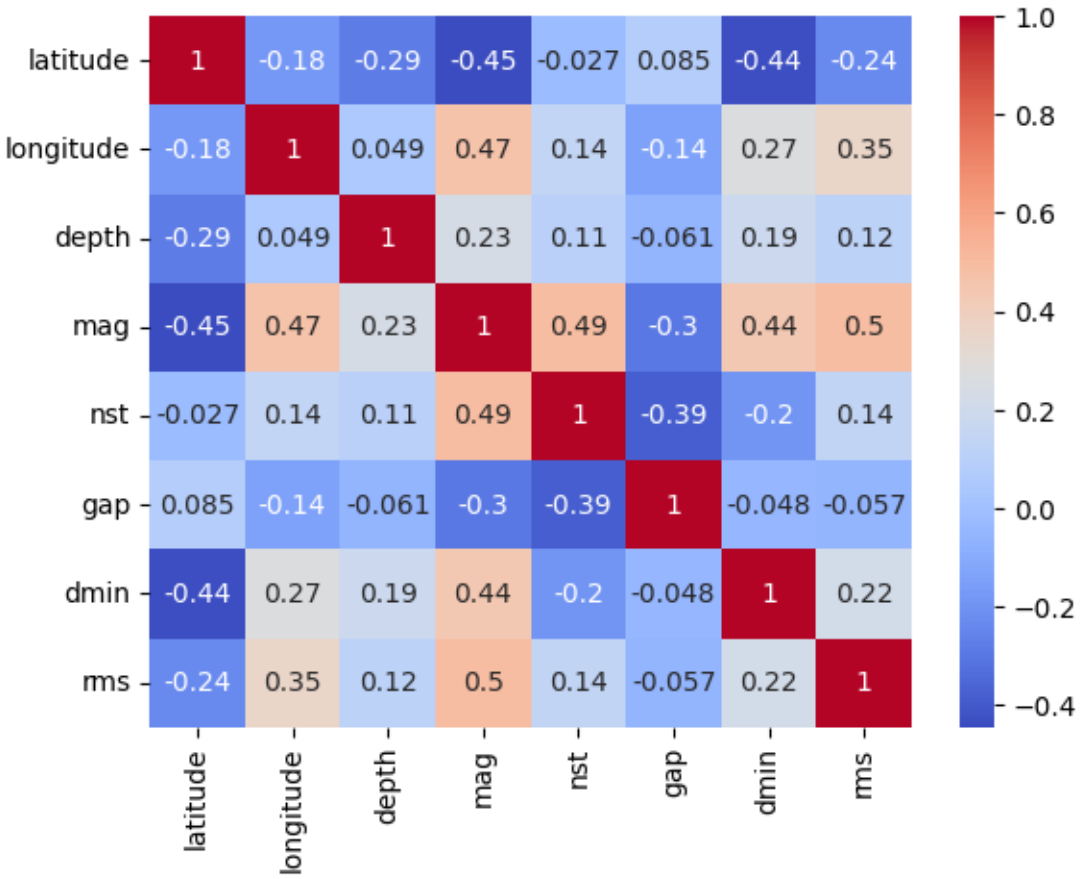
```
packages (0.12.2)
Requirement already satisfied: contourpy>=1.0.1 in
/usr/local/lib/python3.10/dist-packages (from matplotlib) (1.1.1)
Requirement already satisfied: cycler>=0.10 in /usr/local/lib/python3.10/dist-
packages (from matplotlib) (0.12.1)
Requirement already satisfied: fonttools>=4.22.0 in
/usr/local/lib/python3.10/dist-packages (from matplotlib) (4.43.1)
Requirement already satisfied: kiwisolver>=1.0.1 in
/usr/local/lib/python3.10/dist-packages (from matplotlib) (1.4.5)
Requirement already satisfied: numpy>=1.20 in /usr/local/lib/python3.10/dist-
packages (from matplotlib) (1.23.5)
Requirement already satisfied: packaging>=20.0 in
/usr/local/lib/python3.10/dist-packages (from matplotlib) (23.2)
Requirement already satisfied: pillow>=6.2.0 in /usr/local/lib/python3.10/dist-
packages (from matplotlib) (9.4.0)
Requirement already satisfied: pyparsing>=2.3.1 in
/usr/local/lib/python3.10/dist-packages (from matplotlib) (3.1.1)
Requirement already satisfied: python-dateutil>=2.7 in
/usr/local/lib/python3.10/dist-packages (from matplotlib) (2.8.2)
Requirement already satisfied: pandas>=0.25 in /usr/local/lib/python3.10/dist-
packages (from seaborn) (1.5.3)
Requirement already satisfied: pytz>=2020.1 in /usr/local/lib/python3.10/dist-
packages (from pandas>=0.25->seaborn) (2023.3.post1)
Requirement already satisfied: six>=1.5 in /usr/local/lib/python3.10/dist-
packages (from python-dateutil>=2.7->matplotlib) (1.16.0)
                     time  latitude  longitude  depth  mag magType  nst  \
0  1930-12-08T08:01:02.000Z    23.261    120.277   15.0  6.3      mw  NaN
1  1930-12-03T18:51:47.000Z    18.233     96.298   10.0  7.4      mw  NaN
2  1930-12-02T07:01:30.000Z    25.854     98.356   35.0  6.2      mw  NaN
3  1930-11-28T07:32:56.000Z    18.779   -106.767   15.0  6.3      mw  NaN
4  1930-11-25T19:02:53.000Z    35.050    139.129   15.0  6.9      mw  NaN

   gap  dmin  rms     net            id                    updated
0  NaN   NaN  NaN  iscgem  iscgem907791  2015-05-13T18:52:43.000Z
1  NaN   NaN  NaN  iscgem  iscgem907777  2015-05-13T18:52:43.000Z
2  NaN   NaN  NaN  iscgem  iscgem907773  2015-05-13T18:52:43.000Z
3  NaN   NaN  NaN  iscgem  iscgem907769  2015-05-13T18:52:43.000Z
4  NaN   NaN  NaN  iscgem  iscgem907761  2015-05-13T18:52:43.000Z
            latitude      longitude          depth            mag  \
count  797046.000000  797046.000000  797041.000000  797046.000000
mean       19.038314     -12.911692      53.726903       3.884296
std        29.219884     118.010192      99.510254       0.911611
min       -84.422000    -179.999000      -4.900000       2.500000
25%        -4.727000    -118.086000      10.000000       3.000000
50%        29.826000     -52.385000      23.000000       4.000000
75%        39.405000     120.892000      45.620000       4.600000
max        87.221000     180.000000     735.800000       9.500000
```

```
                nst            gap           dmin            rms
count  421658.000000  470629.000000  202838.000000  614095.000000
mean       33.170883     141.002356       1.757801       0.763249
std        55.796692      81.447357       3.734481       0.477974
min         0.000000       0.000000       0.000000       0.000000
25%         8.000000      75.000000       0.082880       0.420000
50%        16.000000     125.000000       0.583000       0.810000
75%        35.000000     195.900000       1.973000       1.070000
max       934.000000     360.000000     127.420000      69.320000
time          object
latitude     float64
longitude    float64
depth        float64
mag          float64
magType       object
nst          float64
gap          float64
dmin         float64
rms          float64
net           object
id            object
updated       object
dtype: object
```

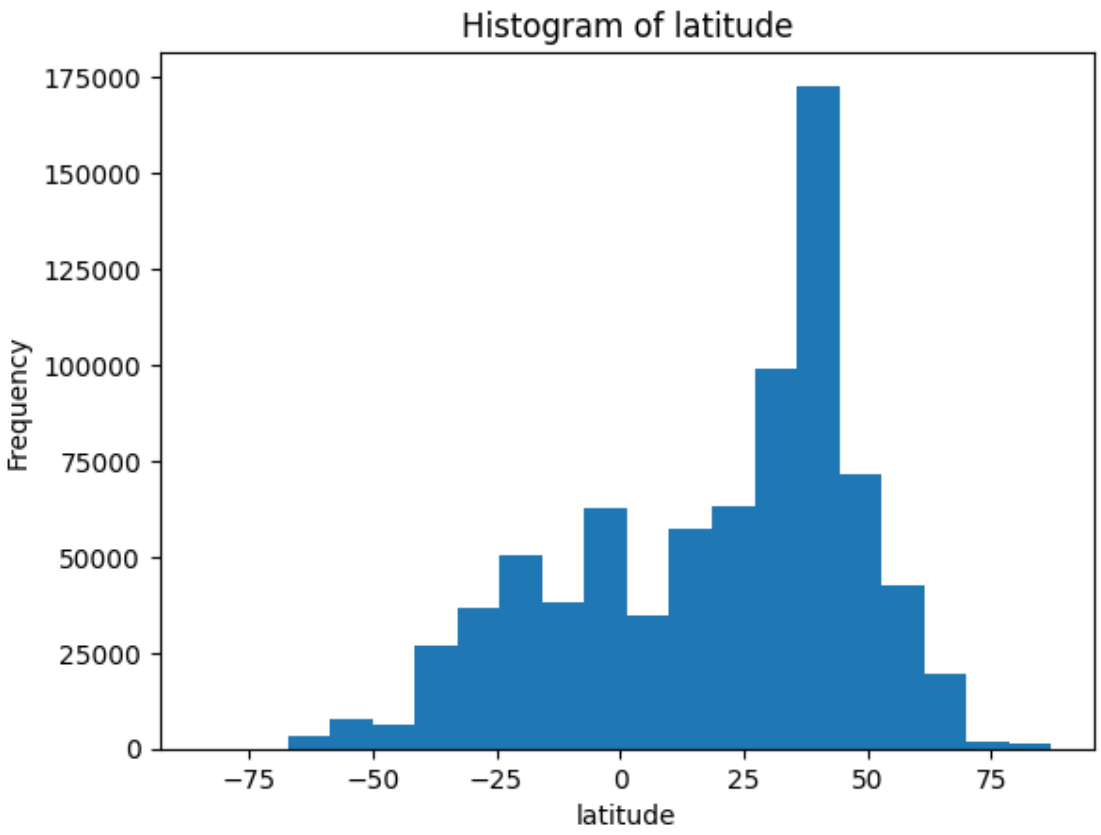


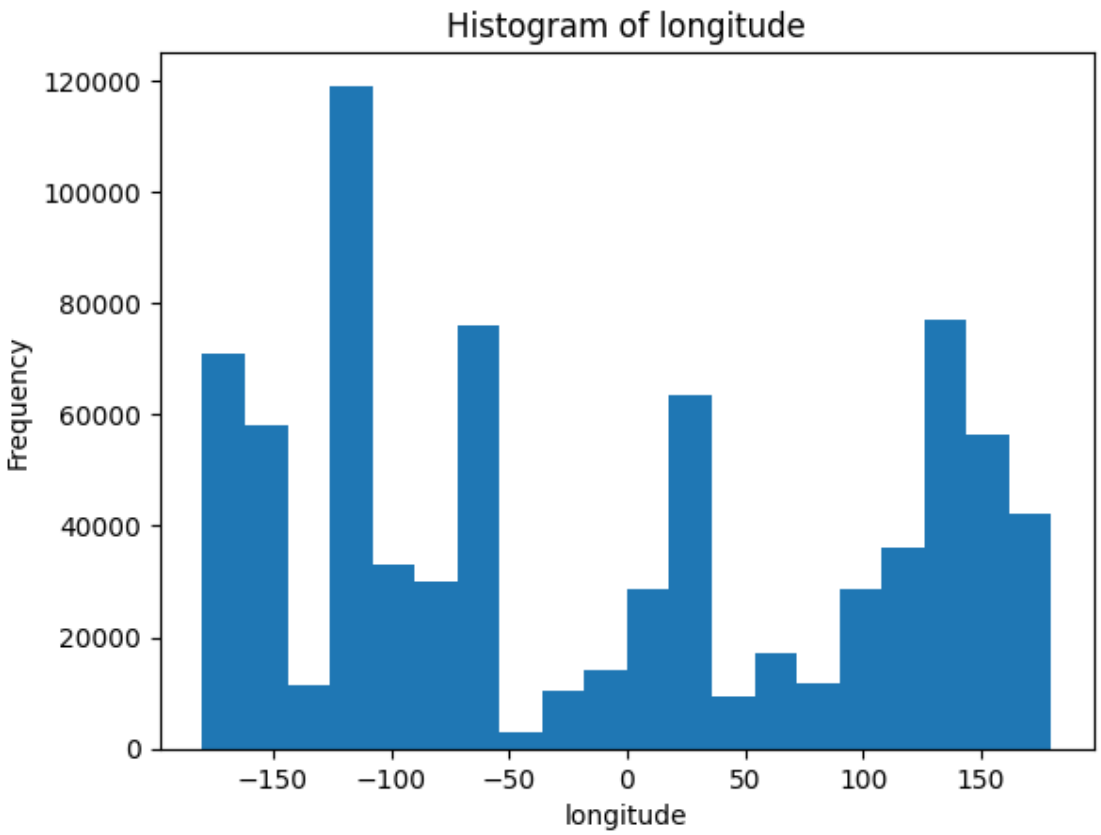
[29]:
```
sns.heatmap(dataset.corr(), annot=True, cmap="coolwarm")
plt.show()
```

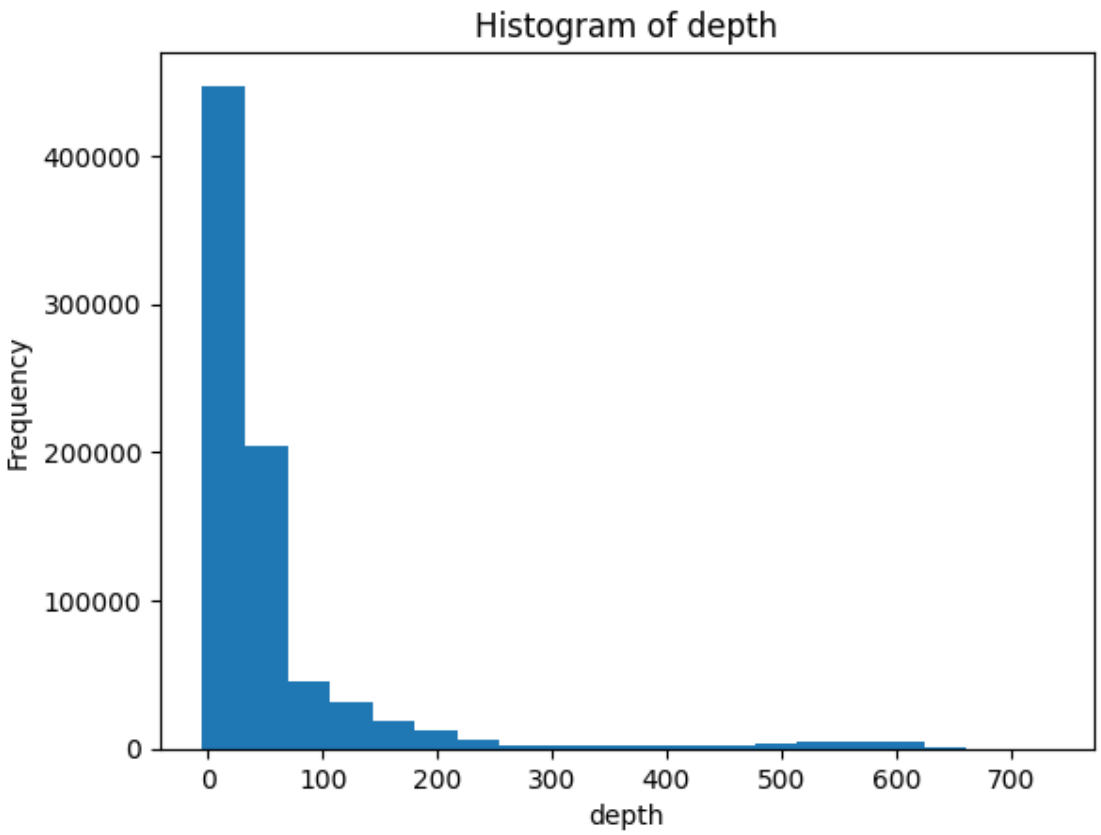
```
<ipython-input-29-26dfcbc1272d>:1: FutureWarning: The default value of
numeric_only in DataFrame.corr is deprecated. In a future version, it will
default to False. Select only valid columns or specify the value of numeric_only
to silence this warning.
  sns.heatmap(dataset.corr(), annot=True, cmap="coolwarm")
```

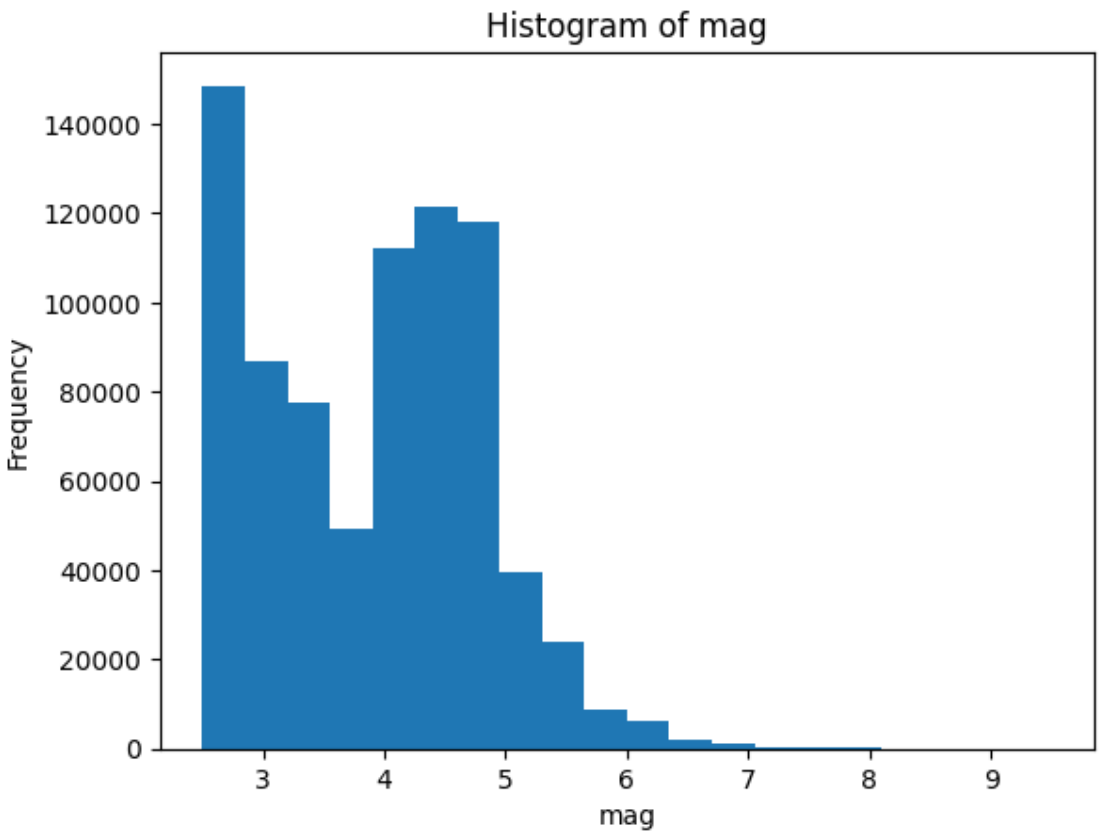


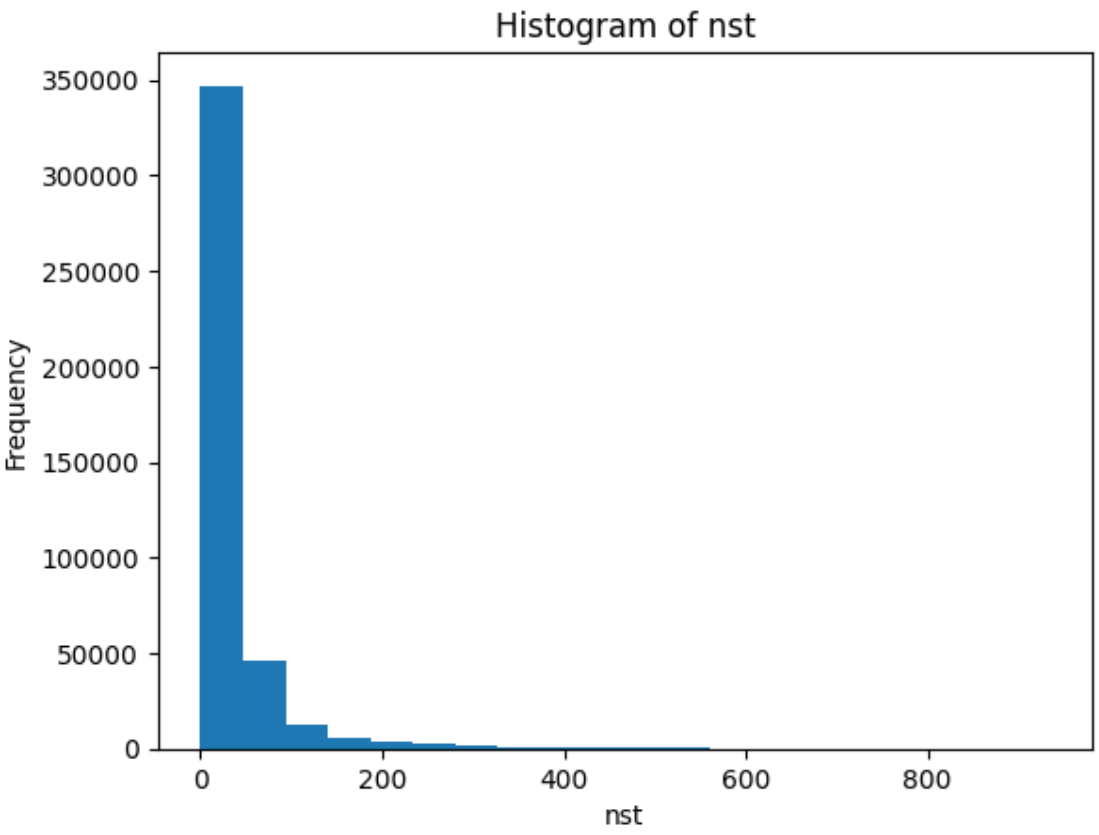
[30]:
```
for column in dataset.select_dtypes(include=['int', 'float']):
    plt.hist(dataset[column], bins=20)
    plt.xlabel(column)
    plt.ylabel("Frequency")
    plt.title(f"Histogram of {column}")
    plt.show()
```

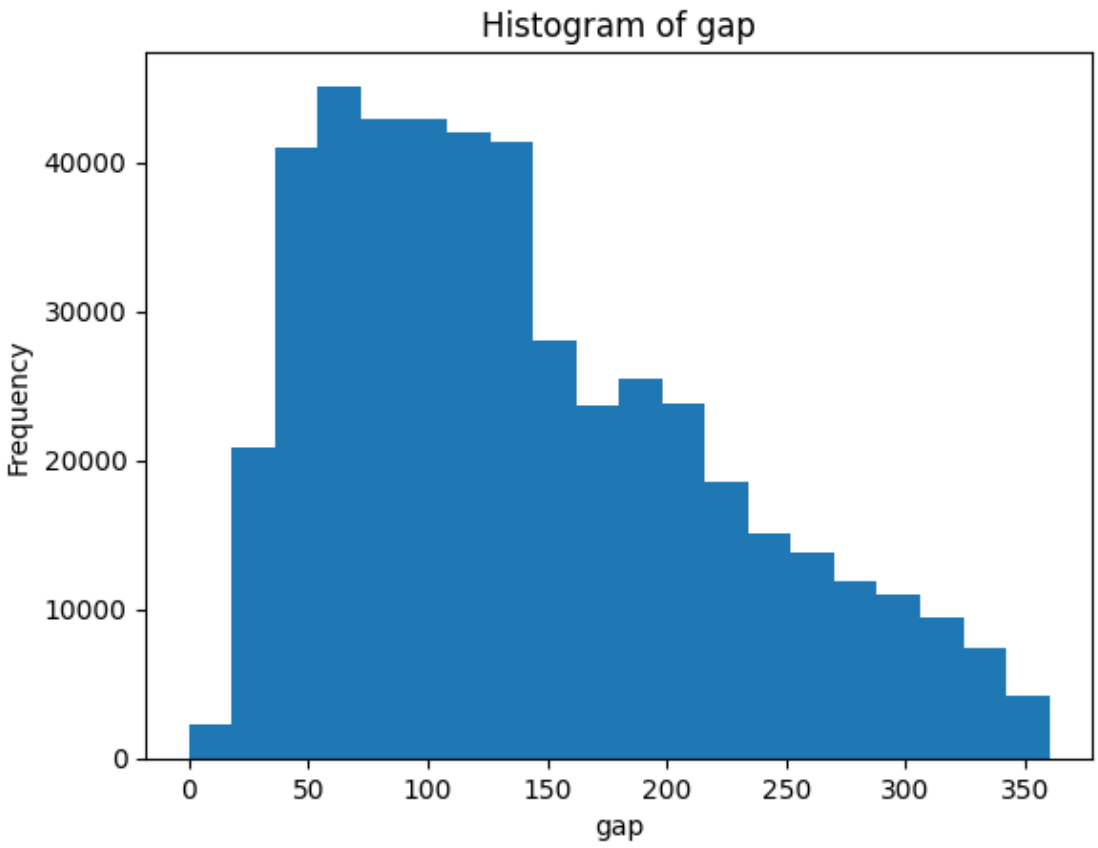

Histogram of latitude
Frequency
175000
150000
125000
100000
75000
50000
25000
0
−75
−50
−25
0
25
50
75
latitude

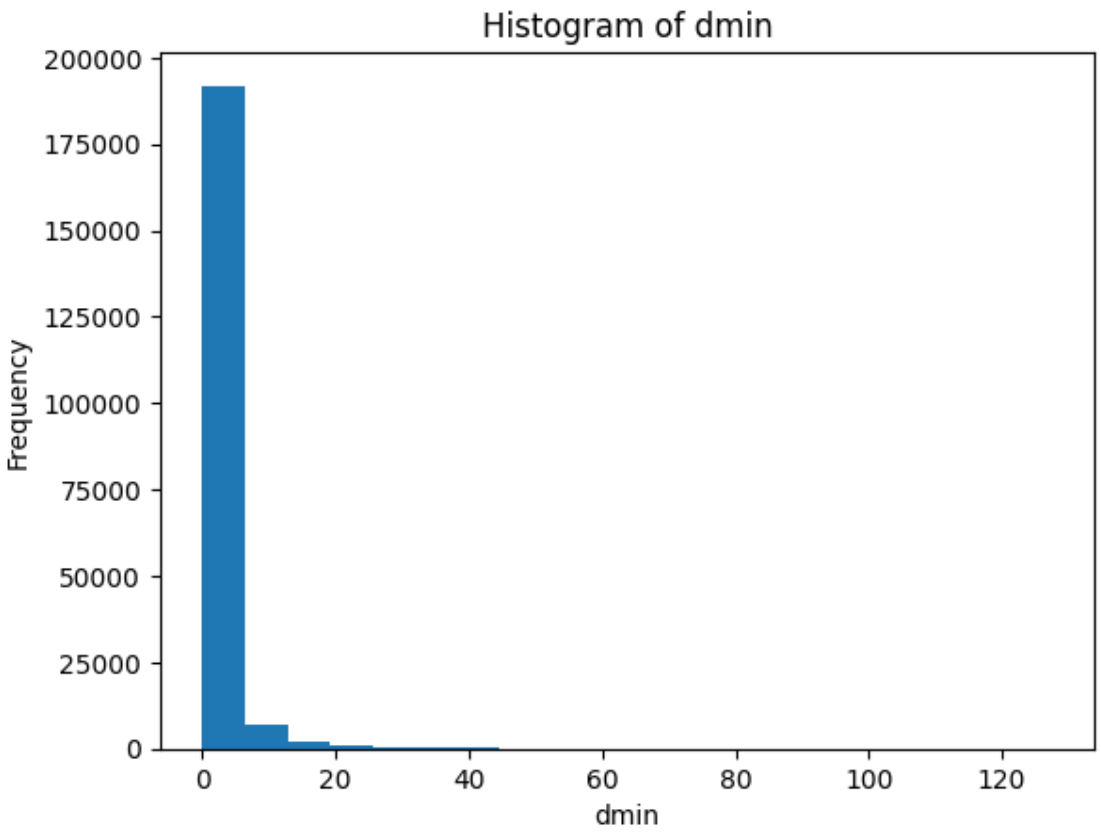

Histogram of longitude
Frequency
120000
100000
80000
60000
40000
20000
0
−150
−100
−50
0
50
100
150
longitude

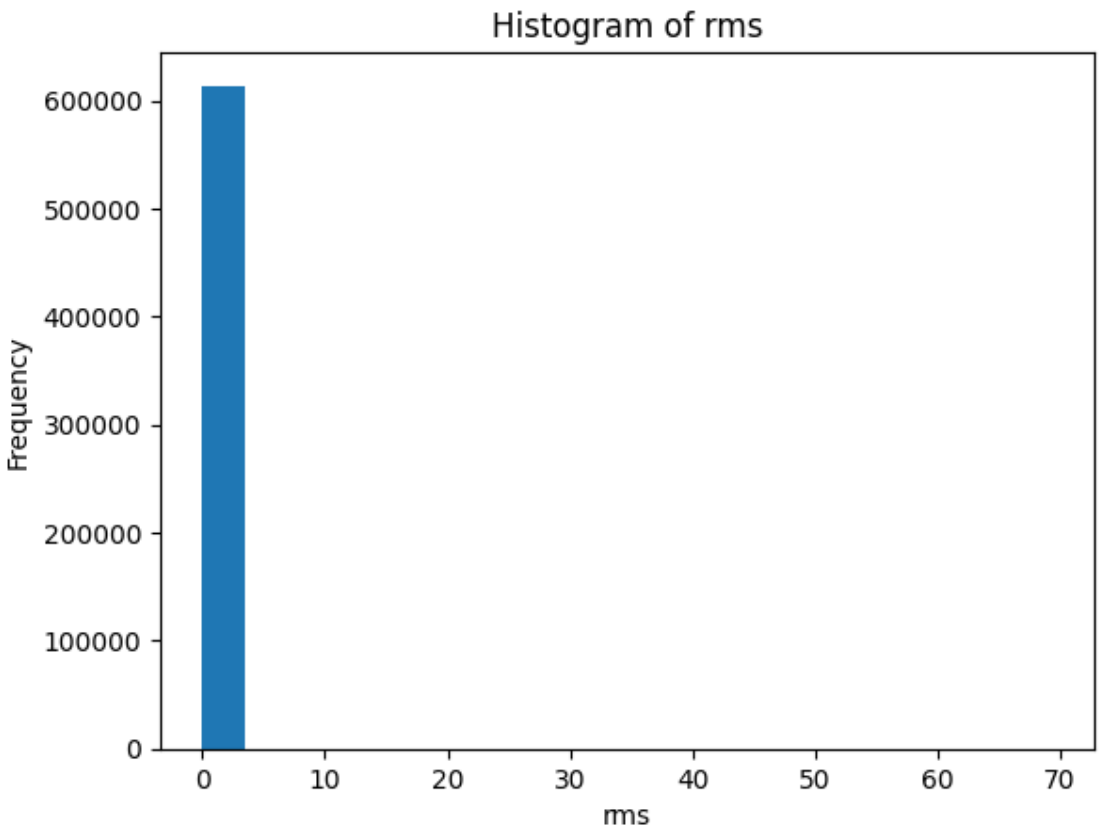

Histogram of depth
Frequency
400000
300000
200000
100000
0
0
100
200
300
400
500
600
700
depth


Histogram of mag
Frequency
140000
120000
100000
80000
60000
40000
20000
0
3
4
5
6
7
8
9
mag


Histogram of nst
Frequency
350000
300000
250000
200000
150000
100000
50000
0
0
200
400
600
800
nst


Histogram of gap
Frequency
40000
30000
20000
10000
0
0
50
100
150
200
250
300
350
gap


Histogram of dmin
Frequency
dmin
200000
175000
150000
125000
100000
75000
50000
25000
0
0
20
40
60
80
100
120


Histogram of rms
Frequency
600000
500000
400000
300000
200000
100000
0
0
10
20
30
40
50
60
70
rms