In [1]:
```
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
import matplotlib.ticker as tic
```

In [2]:
```
df=pd.read_csv("rainfall in india 1901-2015.csv")
df
```

Out[2]:

| | index | SUBDIVISION | YEAR | JAN | FEB | MAR | APR | MAY | JUN | JUL | AUG | SEP | OCT |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **0** | 0 | ANDAMAN & NICOBAR ISLANDS | 1901 | 49.2 | 87.1 | 29.2 | 2.3 | 528.8 | 517.5 | 365.1 | 481.1 | 332.6 | 388.5 |
| **1** | 1 | ANDAMAN & NICOBAR ISLANDS | 1902 | 0.0 | 159.8 | 12.2 | 0.0 | 446.1 | 537.1 | 228.9 | 753.7 | 666.2 | 197.2 |
| **2** | 2 | ANDAMAN & NICOBAR ISLANDS | 1903 | 12.7 | 144.0 | 0.0 | 1.0 | 235.1 | 479.9 | 728.4 | 326.7 | 339.0 | 181.2 |
| **3** | 3 | ANDAMAN & NICOBAR ISLANDS | 1904 | 9.4 | 14.7 | 0.0 | 202.4 | 304.5 | 495.1 | 502.0 | 160.1 | 820.4 | 222.2 |
| **4** | 4 | ANDAMAN & NICOBAR ISLANDS | 1905 | 1.3 | 0.0 | 3.3 | 26.9 | 279.5 | 628.7 | 368.7 | 330.5 | 297.0 | 260.7 |
| **...** | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... |
| **4111** | 4111 | LAKSHADWEEP | 2011 | 5.1 | 2.8 | 3.1 | 85.9 | 107.2 | 153.6 | 350.2 | 254.0 | 255.2 | 117.4 |
| **4112** | 4112 | LAKSHADWEEP | 2012 | 19.2 | 0.1 | 1.6 | 76.8 | 21.2 | 327.0 | 231.5 | 381.2 | 179.8 | 145.9 |
| **4113** | 4113 | LAKSHADWEEP | 2013 | 26.2 | 34.4 | 37.5 | 5.3 | 88.3 | 426.2 | 296.4 | 154.4 | 180.0 | 72.8 |
| **4114** | 4114 | LAKSHADWEEP | 2014 | 53.2 | 16.1 | 4.4 | 14.9 | 57.4 | 244.1 | 116.1 | 466.1 | 132.2 | 169.2 |
| **4115** | 4115 | LAKSHADWEEP | 2015 | 2.2 | 0.5 | 3.7 | 87.1 | 133.1 | 296.6 | 257.5 | 146.4 | 160.4 | 165.4 |

4116 rows × 20 columns

In [3]:
```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 4116 entries, 0 to 4115
Data columns (total 20 columns):
 #   Column       Non-Null Count  Dtype  
---  ------       --------------  -----  
 0   index        4116 non-null   int64  
 1   SUBDIVISION  4116 non-null   object 
 2   YEAR         4116 non-null   int64  
 3   JAN          4112 non-null   float64
 4   FEB          4113 non-null   float64
 5   MAR          4110 non-null   float64
 6   APR          4112 non-null   float64
 7   MAY          4113 non-null   float64
 8   JUN          4111 non-null   float64
 9   JUL          4109 non-null   float64
 10  AUG          4112 non-null   float64
 11  SEP          4110 non-null   float64
 12  OCT          4109 non-null   float64
 13  NOV          4105 non-null   float64
 14  DEC          4106 non-null   float64
 15  ANNUAL       4090 non-null   float64
 16  Jan-Feb      4110 non-null   float64
 17  Mar-May      4107 non-null   float64
 18  Jun-Sep      4106 non-null   float64
 19  Oct-Dec      4103 non-null   float64
dtypes: float64(17), int64(2), object(1)
memory usage: 643.2+ KB
```

In [4]:
```
df.describe()
```

Out[4]:

| | index | YEAR | JAN | FEB | MAR | APR | MAY | |
|---|---|---|---|---|---|---|---|---|
| **count** | 4116.000000 | 4116.000000 | 4112.000000 | 4113.000000 | 4110.000000 | 4112.000000 | 4113.000000 | 411 |
| **mean** | 2057.500000 | 1958.218659 | 18.957320 | 21.805325 | 27.359197 | 43.127432 | 85.745417 | 23 |
| **std** | 1188.331183 | 33.140898 | 33.585371 | 35.909488 | 46.959424 | 67.831168 | 123.234904 | 23 |
| **min** | 0.000000 | 1901.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | |
| **25%** | 1028.750000 | 1930.000000 | 0.600000 | 0.600000 | 1.000000 | 3.000000 | 8.600000 | 7 |
| **50%** | 2057.500000 | 1958.000000 | 6.000000 | 6.700000 | 7.800000 | 15.700000 | 36.600000 | 13 |
| **75%** | 3086.250000 | 1987.000000 | 22.200000 | 26.800000 | 31.300000 | 49.950000 | 97.200000 | 30 |
| **max** | 4115.000000 | 2015.000000 | 583.700000 | 403.500000 | 605.600000 | 595.100000 | 1168.600000 | 160 |

In [5]:
```
df["JAN"]=df["JAN"].fillna(df["JAN"].median())
df["FEB"]=df["FEB"].fillna(df["FEB"].median())
df["MAR"]=df["MAR"].fillna(df["MAR"].median())
df["APR"]=df["APR"].fillna(df["APR"].median())
df["MAY"]=df["MAY"].fillna(df["MAY"].median())
df["JUN"]=df["JUN"].fillna(df["JUN"].median())
df["JUL"]=df["JUL"].fillna(df["JUL"].median())
df["AUG"]=df["AUG"].fillna(df["AUG"].median())
df["SEP"]=df["SEP"].fillna(df["SEP"].median())
df["OCT"]=df["OCT"].fillna(df["OCT"].median())
df["NOV"]=df["NOV"].fillna(df["NOV"].median())
df["DEC"]=df["DEC"].fillna(df["DEC"].median())
df["ANNUAL"]=df["ANNUAL"].fillna(df["ANNUAL"].mean())
df["Jan-Feb"]=df["Jan-Feb"].fillna(df["Jan-Feb"].mean())
df["Mar-May"]=df["Mar-May"].fillna(df["Mar-May"].mean())
df["Jun-Sep"]=df["Jun-Sep"].fillna(df["Jun-Sep"].mean())
df["Oct-Dec"]=df["Oct-Dec"].fillna(df["Oct-Dec"].mean())
```

In [6]:
```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 4116 entries, 0 to 4115
Data columns (total 20 columns):
 #   Column       Non-Null Count  Dtype  
---  ------       --------------  -----  
 0   index        4116 non-null   int64  
 1   SUBDIVISION  4116 non-null   object 
 2   YEAR         4116 non-null   int64  
 3   JAN          4116 non-null   float64
 4   FEB          4116 non-null   float64
 5   MAR          4116 non-null   float64
 6   APR          4116 non-null   float64
 7   MAY          4116 non-null   float64
 8   JUN          4116 non-null   float64
 9   JUL          4116 non-null   float64
 10  AUG          4116 non-null   float64
 11  SEP          4116 non-null   float64
 12  OCT          4116 non-null   float64
 13  NOV          4116 non-null   float64
 14  DEC          4116 non-null   float64
 15  ANNUAL       4116 non-null   float64
 16  Jan-Feb      4116 non-null   float64
 17  Mar-May      4116 non-null   float64
 18  Jun-Sep      4116 non-null   float64
 19  Oct-Dec      4116 non-null   float64
dtypes: float64(17), int64(2), object(1)
memory usage: 643.2+ KB
```

In [7]:
```
df["SUBDIVISION"].value_counts()
```

Out[7]:
```
EAST UTTAR PRADESH                  115
EAST MADHYA PRADESH                 115
EAST RAJASTHAN                      115
UTTARAKHAND                         115
MATATHWADA                          115
NORTH INTERIOR KARNATAKA            115
RAYALSEEMA                          115
SOUTH INTERIOR KARNATAKA            115
JAMMU & KASHMIR                     115
ASSAM & MEGHALAYA                   115
SUB HIMALAYAN WEST BENGAL & SIKKIM  115
VIDARBHA                            115
GUJARAT REGION                      115
JHARKHAND                           115
COASTAL KARNATAKA                   115
KERALA                              115
WEST MADHYA PRADESH                 115
MADHYA MAHARASHTRA                  115
NAGA MANI MIZO TRIPURA              115
HIMACHAL PRADESH                    115
HARYANA DELHI & CHANDIGARH          115
TELANGANA                           115
KONKAN & GOA                        115
ORISSA                              115
COASTAL ANDHRA PRADESH              115
PUNJAB                              115
WEST RAJASTHAN                      115
WEST UTTAR PRADESH                  115
SAURASHTRA & KUTCH                  115
CHHATTISGARH                        115
TAMIL NADU                          115
BIHAR                               115
GANGETIC WEST BENGAL                115
LAKSHADWEEP                         114
ANDAMAN & NICOBAR ISLANDS           110
ARUNACHAL PRADESH                    97
Name: SUBDIVISION, dtype: int64
```

In [26]:
```
df=df.set_index("index")
```

In [27]:
```
df.to_csv("cleaned_rainfall")
```

# ANDAMAN & NICOBAR ISLANDS

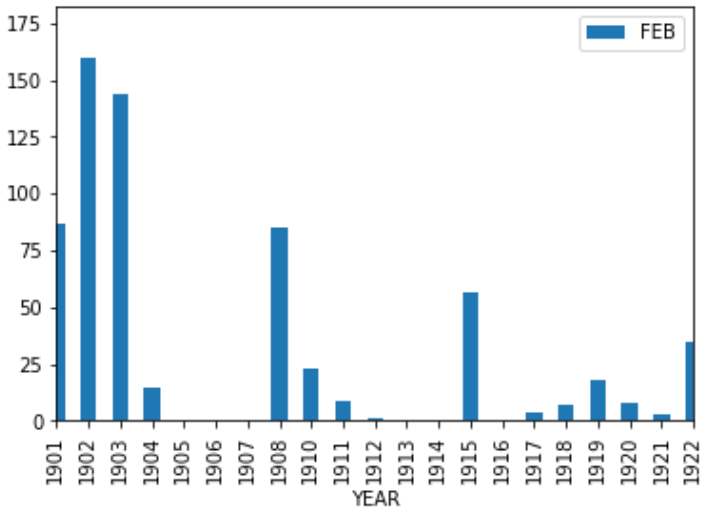
In [11]:
```
dat1=df[df["SUBDIVISION"]=="ANDAMAN & NICOBAR ISLANDS"]
dat1
```

Out[11]:

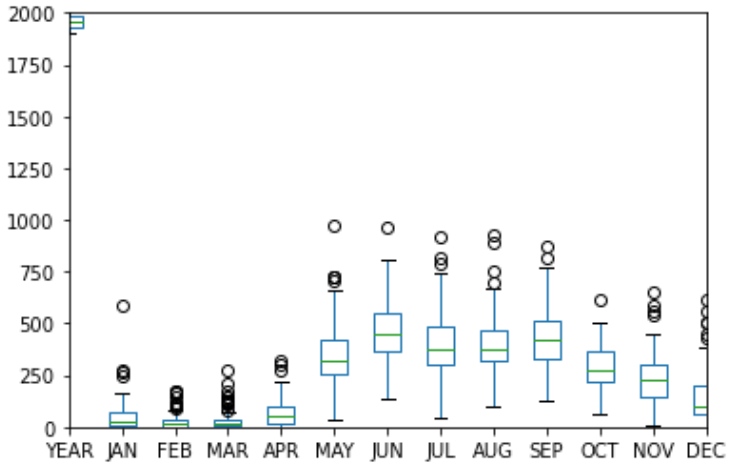
| | index | SUBDIVISION | YEAR | JAN | FEB | MAR | APR | MAY | JUN | JUL | AUG | SEP | OCT | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **0** | 0 | ANDAMAN & NICOBAR ISLANDS | 1901 | 49.2 | 87.1 | 29.2 | 2.3 | 528.8 | 517.5 | 365.1 | 481.1 | 332.6 | 388.5 | 5 |
| **1** | 1 | ANDAMAN & NICOBAR ISLANDS | 1902 | 0.0 | 159.8 | 12.2 | 0.0 | 446.1 | 537.1 | 228.9 | 753.7 | 666.2 | 197.2 | 3 |
| **2** | 2 | ANDAMAN & NICOBAR ISLANDS | 1903 | 12.7 | 144.0 | 0.0 | 1.0 | 235.1 | 479.9 | 728.4 | 326.7 | 339.0 | 181.2 | 2 |
| **3** | 3 | ANDAMAN & NICOBAR ISLANDS | 1904 | 9.4 | 14.7 | 0.0 | 202.4 | 304.5 | 495.1 | 502.0 | 160.1 | 820.4 | 222.2 | 3 |
| **4** | 4 | ANDAMAN & NICOBAR ISLANDS | 1905 | 1.3 | 0.0 | 3.3 | 26.9 | 279.5 | 628.7 | 368.7 | 330.5 | 297.0 | 260.7 | |
| **...** | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | |
| **105** | 105 | ANDAMAN & NICOBAR ISLANDS | 2011 | 265.9 | 84.8 | 272.8 | 111.4 | 326.5 | 383.2 | 583.2 | 441.5 | 757.1 | 212.3 | 1 |
| **106** | 106 | ANDAMAN & NICOBAR ISLANDS | 2012 | 119.9 | 45.6 | 30.9 | 55.8 | 533.9 | 458.2 | 317.3 | 369.6 | 868.9 | 209.7 | 3 |
| **107** | 107 | ANDAMAN & NICOBAR ISLANDS | 2013 | 67.1 | 37.6 | 43.0 | 46.3 | 509.3 | 777.0 | 564.8 | 336.7 | 473.6 | 455.8 | 3 |
| **108** | 108 | ANDAMAN & NICOBAR ISLANDS | 2014 | 41.9 | 8.6 | 0.0 | 11.1 | 238.0 | 416.6 | 467.6 | 321.6 | 412.9 | 402.6 | 2 |
| **109** | 109 | ANDAMAN & NICOBAR ISLANDS | 2015 | 126.8 | 7.6 | 3.1 | 138.2 | 331.9 | 346.4 | 328.9 | 480.0 | 523.3 | 252.1 | 2 |

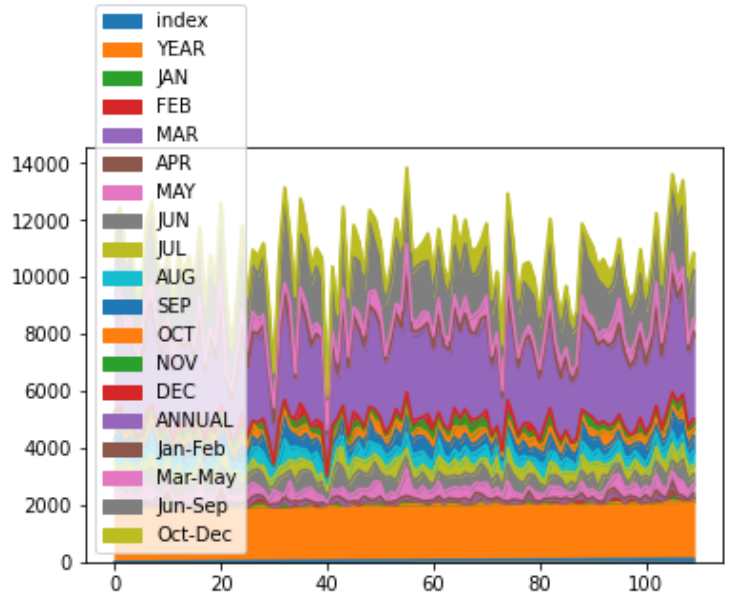
110 rows × 20 columns

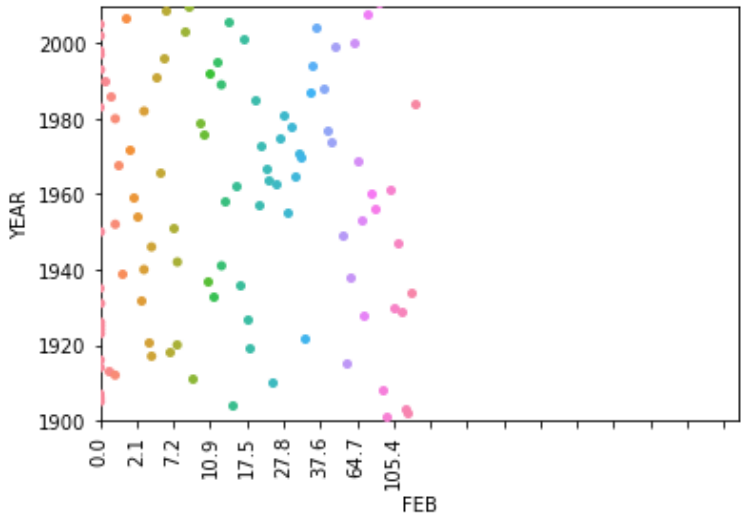
In [23]:
```
dat1.plot.bar("YEAR","FEB")
plt.xlim(0,20)
plt.figure(figsize=(60,30))
plt.show()
dat1.plot.box()
plt.xlim(2,14)
plt.ylim(0,2000)
plt.show()
dat1.plot.area()
dat1.plot.scatter("YEAR","FEB")
sns.stripplot(x=dat1["FEB"],y=dat1["YEAR"],jitter=True)
plt.ylim(1900,2010)
plt.xlim(0,145)
plt.xticks(dat1["FEB"],rotation="vertical")
plt.gca().xaxis.set_major_locator(tic.MultipleLocator(base=10))
plt.show()
dat1.plot.hist()
```



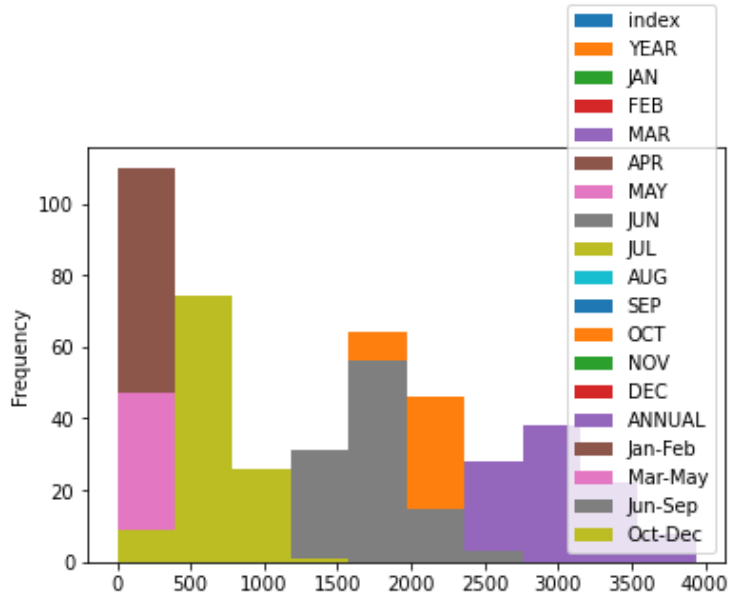

```
<Figure size 4320x2160 with 0 Axes>
```






Out[23]: <AxesSubplot:ylabel='Frequency'>



# ARUNACHAL PRADESH

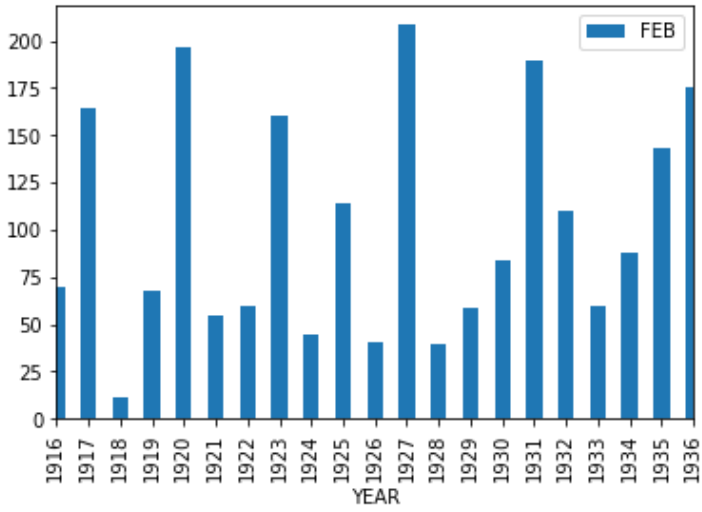
In [13]:
```
dat2=df[df["SUBDIVISION"]=="ARUNACHAL PRADESH"]
dat2
```

Out[13]:

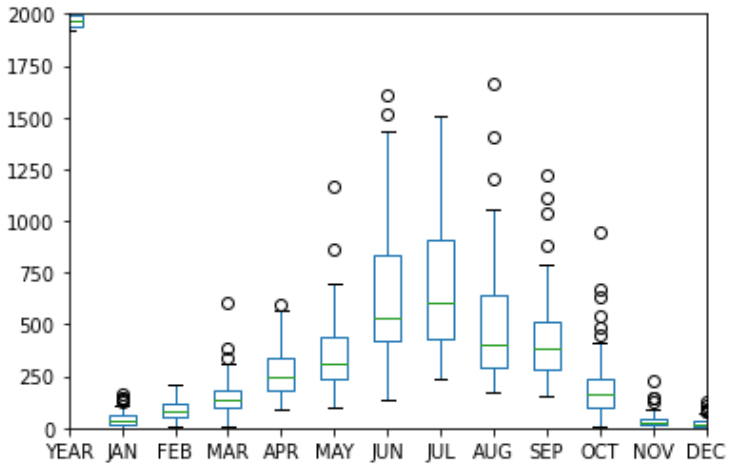
| | index | SUBDIVISION | YEAR | JAN | FEB | MAR | APR | MAY | JUN | JUL | AUG | SEP | OCT |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **110** | 110 | ARUNACHAL PRADESH | 1916 | 48.1 | 69.8 | 71.1 | 316.1 | 424.6 | 1124.9 | 284.8 | 629.7 | 333.9 | 65.2 |
| **111** | 111 | ARUNACHAL PRADESH | 1917 | 21.4 | 164.5 | 7.8 | 269.6 | 107.9 | 823.8 | 909.1 | 628.4 | 411.5 | 199.3 |
| **112** | 112 | ARUNACHAL PRADESH | 1918 | 10.4 | 11.0 | 191.2 | 144.6 | 861.1 | 1609.9 | 1303.0 | 692.6 | 515.8 | 125.2 |
| **113** | 113 | ARUNACHAL PRADESH | 1919 | 34.5 | 67.8 | 28.5 | 256.9 | 420.6 | 973.6 | 999.0 | 286.7 | 628.7 | 948.3 |
| **114** | 114 | ARUNACHAL PRADESH | 1920 | 14.0 | 196.3 | 605.6 | 364.7 | 173.6 | 840.6 | 535.4 | 896.5 | 376.7 | 103.3 |
| **...** | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... |
| **202** | 202 | ARUNACHAL PRADESH | 2011 | 40.0 | 51.3 | 174.5 | 240.8 | 219.6 | 288.4 | 531.4 | 277.6 | 286.7 | 51.9 |
| **203** | 203 | ARUNACHAL PRADESH | 2012 | 57.8 | 35.8 | 134.2 | 403.4 | 187.4 | 645.8 | 638.9 | 316.0 | 724.9 | 248.1 |
| **204** | 204 | ARUNACHAL PRADESH | 2013 | 18.5 | 40.5 | 115.1 | 175.1 | 335.8 | 290.0 | 329.6 | 230.2 | 316.1 | 164.1 |
| **205** | 205 | ARUNACHAL PRADESH | 2014 | 19.0 | 101.9 | 80.3 | 86.7 | 299.0 | 415.8 | 392.4 | 599.6 | 343.0 | 35.1 |
| **206** | 206 | ARUNACHAL PRADESH | 2015 | 30.8 | 47.5 | 97.5 | 287.1 | 238.9 | 637.9 | 329.3 | 595.5 | 374.2 | 65.2 |

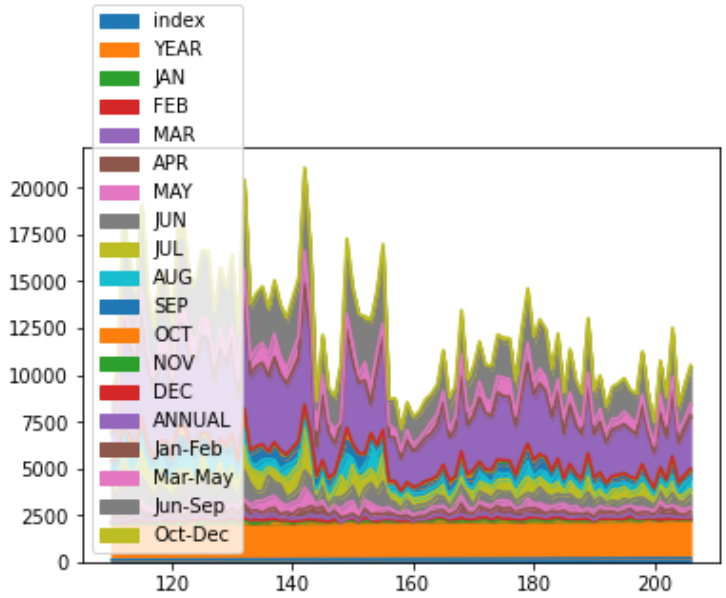
97 rows × 20 columns

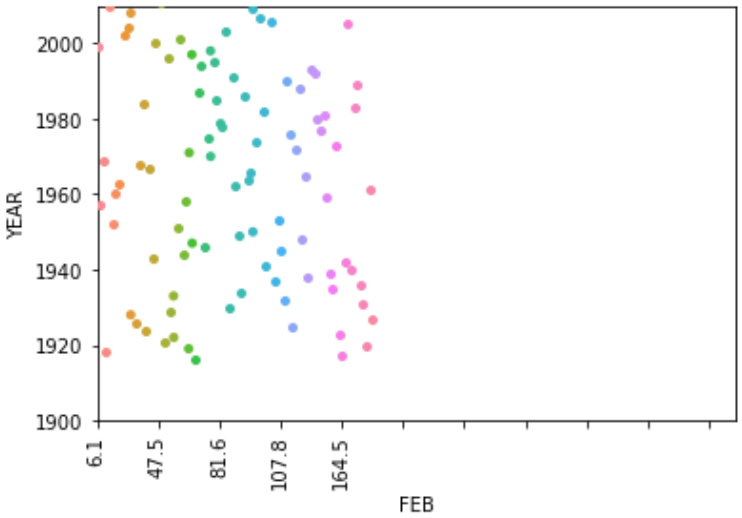
In [24]:
```
dat2.plot.bar("YEAR","FEB")
plt.xlim(0,20)
plt.figure(figsize=(60,30))
plt.show()
dat2.plot.box()
plt.xlim(2,14)
plt.ylim(0,2000)
plt.show()
dat2.plot.area()
dat2.plot.scatter("YEAR","FEB")
sns.stripplot(x=dat2["FEB"],y=dat2["YEAR"],jitter=True)
plt.ylim(1900,2010)
plt.xlim(0,145)
plt.xticks(dat2["FEB"],rotation="vertical")
plt.gca().xaxis.set_major_locator(tic.MultipleLocator(base=20))
plt.show()
dat2.plot.hist()
plt.show()
```

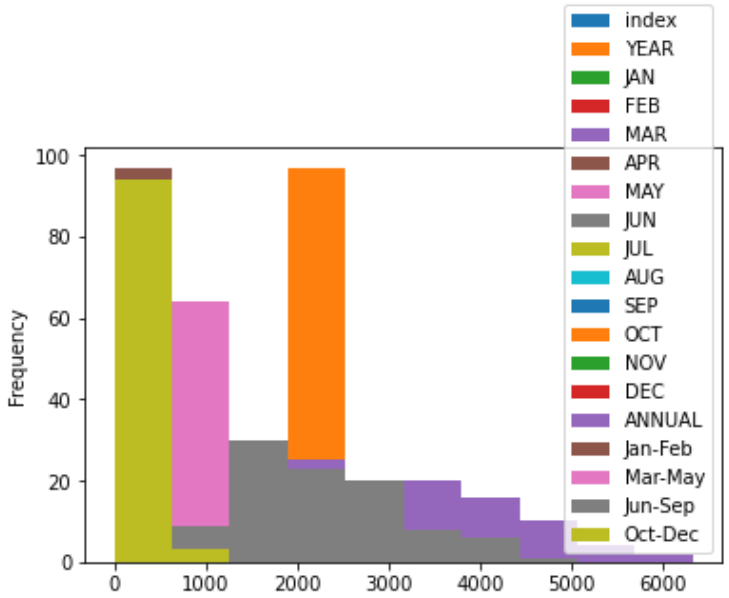



```
<Figure size 4320x2160 with 0 Axes>
```

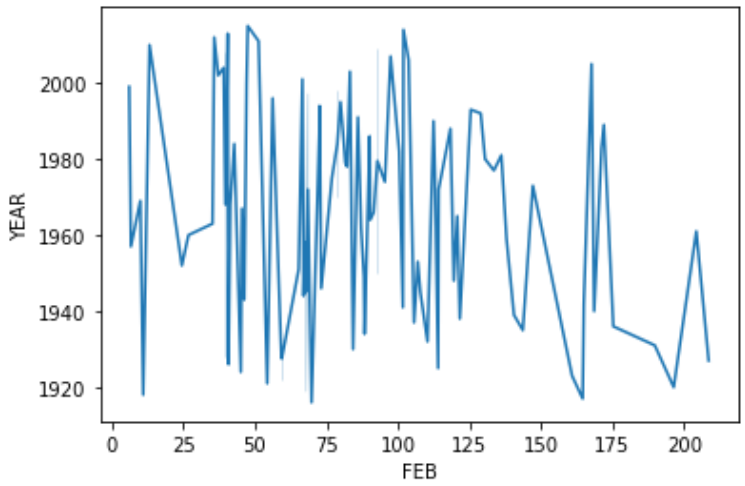



index
YEAR
JAN
FEB
MAR
APR
MAY
JUN
JUL
AUG
SEP
OCT
NOV
DEC
ANNUAL
Jan-Feb
Mar-May
Jun-Sep
Oct-Dec
20000
17500
15000
12500
10000
7500
5000
2500
0
120
140
160
180
200


2000
1980
1960
1940
1920
1900
YEAR
6.1
47.5
81.6
107.8
164.5
FEB


index
YEAR
JAN
FEB
MAR
APR
MAY
JUN
JUL
AUG
SEP
OCT
NOV
DEC
ANNUAL
Jan-Feb
Mar-May
Jun-Sep
Oct-Dec
100
80
60
40
20
0
Frequency
0
1000
2000
3000
4000
5000
6000

In [25]:
```
sns.lineplot(x=dat2["FEB"],y=dat2["YEAR"])
plt.show()
```




In [ ]:

In [ ]: