In [1]:

```
from sklearn.datasets import load_breast_cancer
import pandas as pd
import matplotlib.pyplot as plt
dataset=load_breast_cancer()
df=pd.DataFrame(dataset['data'],columns=dataset['feature_names'])
df['target']=dataset['target']
df.head()
```

Out[1]:

	mean radius	mean texture	mean perimeter	mean area	mean smoothness	mean compactness	mean concavity	mean concave points	mear symmetry
0	17.99	10.38	122.80	1001.0	0.11840	0.27760	0.3001	0.14710	0.2419
1	20.57	17.77	132.90	1326.0	0.08474	0.07864	0.0869	0.07017	0.1812
2	19.69	21.25	130.00	1203.0	0.10960	0.15990	0.1974	0.12790	0.2069
3	11.42	20.38	77.58	386.1	0.14250	0.28390	0.2414	0.10520	0.2597
4	20.29	14.34	135.10	1297.0	0.10030	0.13280	0.1980	0.10430	0.1809

5 rows × 31 columns

In [2]:

```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 569 entries, 0 to 568
Data columns (total 31 columns):
 #   Column                   Non-Null Count  Dtype
---  ------                   --------------  -----
 0   mean radius              569 non-null    float64
 1   mean texture             569 non-null    float64
 2   mean perimeter           569 non-null    float64
 3   mean area                569 non-null    float64
 4   mean smoothness          569 non-null    float64
 5   mean compactness         569 non-null    float64
 6   mean concavity           569 non-null    float64
 7   mean concave points      569 non-null    float64
 8   mean symmetry            569 non-null    float64
 9   mean fractal dimension   569 non-null    float64
 10  radius error             569 non-null    float64
 11  texture error            569 non-null    float64
 12  perimeter error          569 non-null    float64
 13  area error               569 non-null    float64
 14  smoothness error         569 non-null    float64
 15  compactness error        569 non-null    float64
 16  concavity error          569 non-null    float64
 17  concave points error     569 non-null    float64
 18  symmetry error           569 non-null    float64
 19  fractal dimension error  569 non-null    float64
 20  worst radius             569 non-null    float64
 21  worst texture            569 non-null    float64
 22  worst perimeter          569 non-null    float64
 23  worst area               569 non-null    float64
 24  worst smoothness         569 non-null    float64
 25  worst compactness        569 non-null    float64
 26  worst concavity          569 non-null    float64
 27  worst concave points     569 non-null    float64
 28  worst symmetry           569 non-null    float64
 29  worst fractal dimension  569 non-null    float64
 30  target                   569 non-null    int32
dtypes: float64(30), int32(1)
memory usage: 135.7 KB
```

In [3]:

```
df.describe()
```

Out[3]:

	mean radius	mean texture	mean perimeter	mean area	mean smoothness	mean compactness	mean concavity
count	569.000000	569.000000	569.000000	569.000000	569.000000	569.000000	569.000000
mean	14.127292	19.289649	91.969033	654.889104	0.096360	0.104341	0.088799
std	3.524049	4.301036	24.298981	351.914129	0.014064	0.052813	0.079720
min	6.981000	9.710000	43.790000	143.500000	0.052630	0.019380	0.000000
25%	11.700000	16.170000	75.170000	420.300000	0.086370	0.064920	0.029560
50%	13.370000	18.840000	86.240000	551.100000	0.095870	0.092630	0.061540
75%	15.780000	21.800000	104.100000	782.700000	0.105300	0.130400	0.130700
max	28.110000	39.280000	188.500000	2501.000000	0.163400	0.345400	0.426800

8 rows × 31 columns

In [5]:

```
df.isnull().values.any()
```

Out[5]:

```
False
```

In [6]:

```
df.corr()
```

Out[6]:

	mean radius	mean texture	mean perimeter	mean area	mean smoothness	mean compactness	mea concavi
mean radius	1.000000	0.323782	0.997855	0.987357	0.170581	0.506124	0.67676
mean texture	0.323782	1.000000	0.329533	0.321086	-0.023389	0.236702	0.30241
mean perimeter	0.997855	0.329533	1.000000	0.986507	0.207278	0.556936	0.71613
mean area	0.987357	0.321086	0.986507	1.000000	0.177028	0.498502	0.68598
mean smoothness	0.170581	-0.023389	0.207278	0.177028	1.000000	0.659123	0.52198
mean compactness	0.506124	0.236702	0.556936	0.498502	0.659123	1.000000	0.88312
mean concavity	0.676764	0.302418	0.716136	0.685983	0.521984	0.883121	1.00000
mean concave points	0.822529	0.293464	0.850977	0.823269	0.553695	0.831135	0.92139
mean symmetry	0.147741	0.071401	0.183027	0.151293	0.557775	0.602641	0.50066
mean fractal dimension	-0.311631	-0.076437	-0.261477	-0.283110	0.584792	0.565369	0.33678
radius error	0.679090	0.275869	0.691765	0.732562	0.301467	0.497473	0.63192
texture error	-0.097317	0.386358	-0.086761	-0.066280	0.068406	0.046205	0.07621
perimeter error	0.674172	0.281673	0.693135	0.726628	0.296092	0.548905	0.66039
area error	0.735864	0.259845	0.744983	0.800086	0.246552	0.455653	0.61742
smoothness error	-0.222600	0.006614	-0.202694	-0.166777	0.332375	0.135299	0.09856
compactness error	0.206000	0.191975	0.250744	0.212583	0.318943	0.738722	0.67027
concavity error	0.194204	0.143293	0.228082	0.207660	0.248396	0.570517	0.69127
concave points error	0.376169	0.163851	0.407217	0.372320	0.380676	0.642262	0.68326
symmetry error	-0.104321	0.009127	-0.081629	-0.072497	0.200774	0.229977	0.17800
fractal dimension error	-0.042641	0.054458	-0.005523	-0.019887	0.283607	0.507318	0.44930
worst radius	0.969539	0.352573	0.969476	0.962746	0.213120	0.535315	0.68823
worst texture	0.297008	0.912045	0.303038	0.287489	0.036072	0.248133	0.29987
worst perimeter	0.965137	0.358040	0.970387	0.959120	0.238853	0.590210	0.72956
worst area	0.941082	0.343546	0.941550	0.959213	0.206718	0.509604	0.67598

	mean radius	mean texture	mean perimeter	mean area	mean smoothness	mean compactness	mea concavi
worst smoothness	0.119616	0.077503	0.150549	0.123523	0.805324	0.565541	0.44882
worst compactness	0.413463	0.277830	0.455774	0.390410	0.472468	0.865809	0.75496
worst concavity	0.526911	0.301025	0.563879	0.512606	0.434926	0.816275	0.88410
worst concave points	0.744214	0.295316	0.771241	0.722017	0.503053	0.815573	0.86132
worst symmetry	0.163953	0.105008	0.189115	0.143570	0.394309	0.510223	0.40946
worst fractal dimension	0.007066	0.119205	0.051019	0.003738	0.499316	0.687382	0.51493
target	-0.730029	-0.415185	-0.742636	-0.708984	-0.358560	-0.596534	-0.69636

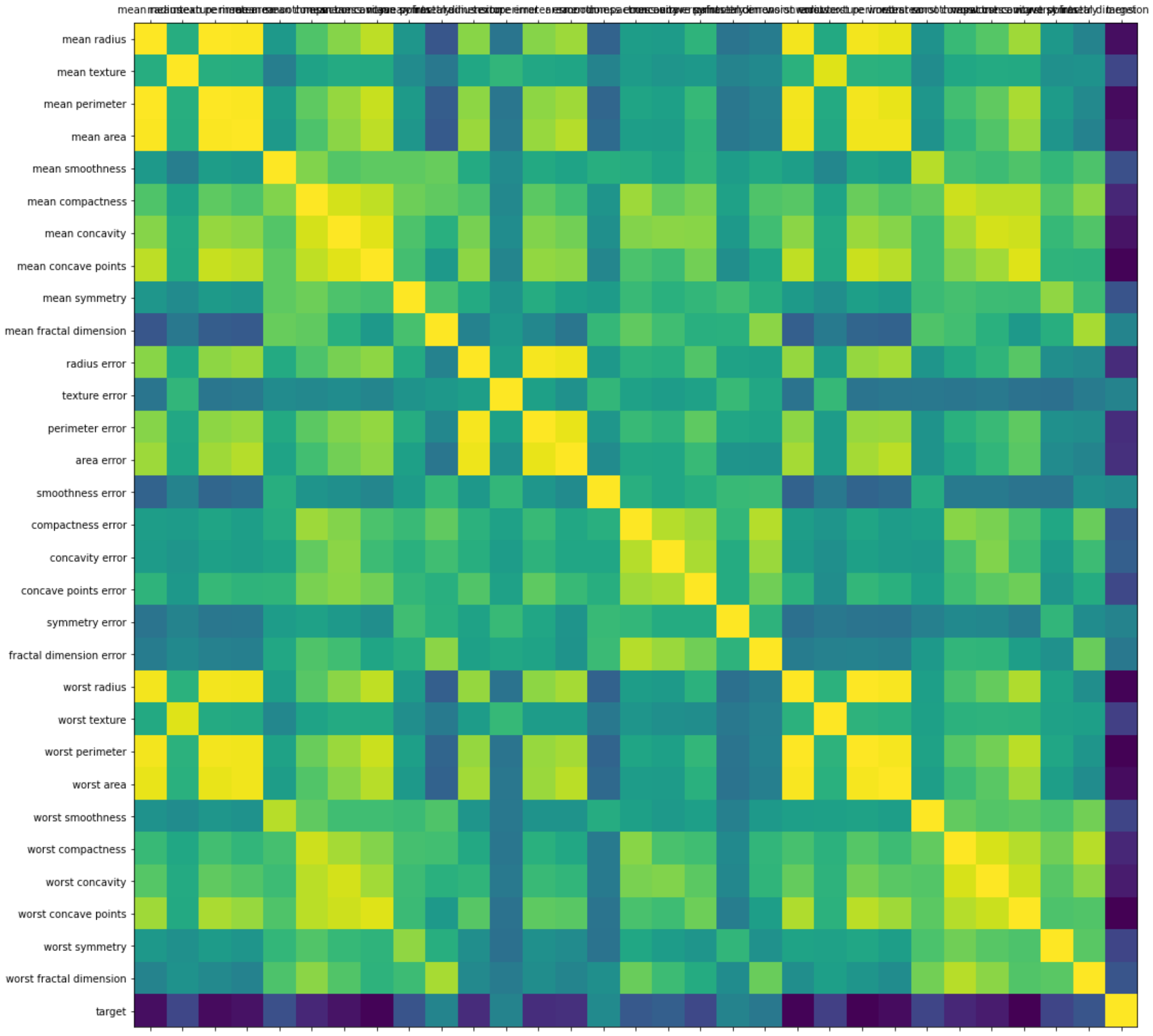
31 rows × 31 columns

In [20]:

```
def plot_corr(df, size=18):
        corr = df.corr()
        fig, ax = plt.subplots(figsize=(size, size))
        ax.matshow(corr)
        plt.xticks(range(len(corr.columns)), corr.columns)
        plt.yticks(range(len(corr.columns)), corr.columns)
```

In [21]:

```
plot_corr(df)
```



In [22]:

```
from sklearn.model_selection import train_test_split
X = df.drop('target',axis=1) # Predictor feature columns (8 X m)
Y = df['target'] # Predicted class (1=True, 0=False) (1 X m)
x_train, x_test, y_train, y_test = train_test_split(X, Y, test_size=0.3, random_state=1)
# 1 is just any random seed number
x_train.head()
```

Out[22]:

```
                                                                                 mean
       mean     mean       mean    mean         mean          mean        mean  concave       me
     radius  texture  perimeter    area   smoothness   compactness   concavity   points    symme
249   11.52    14.93      73.87   406.3      0.10130       0.07808    0.043280 0.029290     0.18
 58   13.05    19.31      82.61   527.2      0.08060       0.03789    0.000692 0.004167     0.18
476   14.20    20.53      92.41   618.4      0.08931       0.11080    0.050630 0.030580     0.15
529   12.07    13.44      77.83   445.2      0.11000       0.09009    0.037810 0.027980     0.16
422   11.61    16.02      75.46   408.2      0.10880       0.11680    0.070970 0.044970     0.18

5 rows × 30 columns
```

In [23]:

```
skew = df.skew()
print("Skew of attribute distributions in the data:\n")
skew
```

```
Skew of attribute distributions in the data:
```

Out[23]:

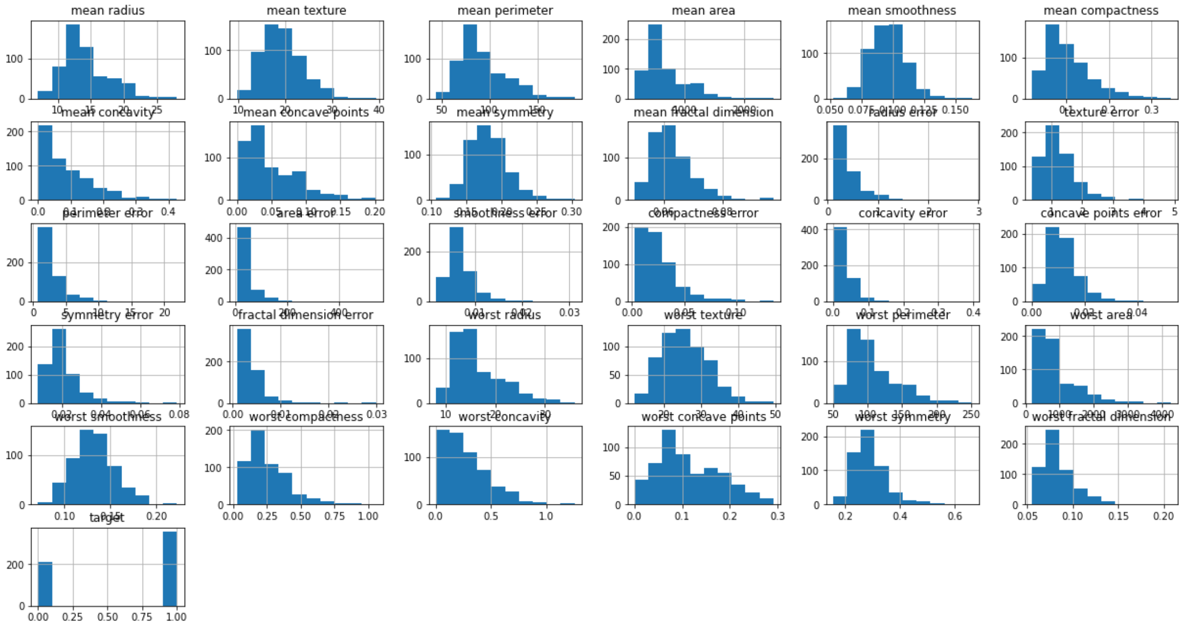
```
mean radius                0.942380
mean texture               0.650450
mean perimeter             0.990650
mean area                  1.645732
mean smoothness            0.456324
mean compactness           1.190123
mean concavity             1.401180
mean concave points        1.171180
mean symmetry              0.725609
mean fractal dimension     1.304489
radius error               3.088612
texture error              1.646444
perimeter error            3.443615
area error                 5.447186
smoothness error           2.314450
compactness error          1.902221
concavity error            5.110463
concave points error       1.444678
symmetry error             2.195133
fractal dimension error    3.923969
worst radius               1.103115
worst texture              0.498321
worst perimeter            1.128164
worst area                 1.859373
worst smoothness           0.415426
worst compactness          1.473555
worst concavity            1.150237
worst concave points       0.492616
worst symmetry             1.433928
worst fractal dimension    1.662579
target                    -0.528461
dtype: float64
```

In [28]:

```
from matplotlib import pyplot
pyplot.rcParams['figure.figsize'] = [21, 11] # set the figure size
```

In [29]:

```
# Draw histograms for all attributes
df.hist()
pyplot.show()
```



In [ ]: