In [1]:
```
import pandas as pd
import numpy as np
import matplotlib as plt
import seaborn as sns
%matplotlib inline
```

In [30]:
```
# Reading the data
titanic_data = pd.read_csv('/Users/apple/Desktop/titanic-data.csv')
```

In [31]:
```
# Displaying header rows
titanic_data.head()
```

Out[31]:

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin	Embarked
0	1	0	3	Braund, Mr. Owen Harris	male	22.0	1	0	A/5 21171	7.2500	NaN	S
1	2	1	1	Cumings, Mrs. John Bradley (Florence Briggs Th...	female	38.0	1	0	PC 17599	71.2833	C85	C
2	3	1	3	Heikkinen, Miss. Laina	female	26.0	0	0	STON/O2. 3101282	7.9250	NaN	S
3	4	1	1	Futrelle, Mrs. Jacques Heath (Lily May Peel)	female	35.0	1	0	113803	53.1000	C123	S
4	5	0	3	Allen, Mr. William Henry	male	35.0	0	0	373450	8.0500	NaN	S

In [32]:
```
# Displaying Description
titanic_data.describe()
```

Out[32]:

	PassengerId	Survived	Pclass	Age	SibSp	Parch	Fare
count	891.000000	891.000000	891.000000	714.000000	891.000000	891.000000	891.000000
mean	446.000000	0.383838	2.308642	29.699118	0.523008	0.381594	32.204208
std	257.353842	0.486592	0.836071	14.526497	1.102743	0.806057	49.693429
min	1.000000	0.000000	1.000000	0.420000	0.000000	0.000000	0.000000
25%	223.500000	0.000000	2.000000	20.125000	0.000000	0.000000	7.910400
50%	446.000000	0.000000	3.000000	28.000000	0.000000	0.000000	14.454200
75%	668.500000	1.000000	3.000000	38.000000	1.000000	0.000000	31.000000
max	891.000000	1.000000	3.000000	80.000000	8.000000	6.000000	512.329200

In [33]:
```
# Dropping unwanted columns
titanic_data.drop(['Cabin', 'SibSp', 'Parch', 'Ticket', 'Embarked'], axis=1).head()
```

Out[33]:

	PassengerId	Survived	Pclass	Name	Sex	Age	Fare
0	1	0	3	Braund, Mr. Owen Harris	male	22.0	7.2500
1	2	1	1	Cumings, Mrs. John Bradley (Florence Briggs Th...	female	38.0	71.2833
2	3	1	3	Heikkinen, Miss. Laina	female	26.0	7.9250
3	4	1	1	Futrelle, Mrs. Jacques Heath (Lily May Peel)	female	35.0	53.1000
4	5	0	3	Allen, Mr. William Henry	male	35.0	8.0500

In [34]:
```
# Replace missing age values with the mean Age
titanic_data['Age'].fillna(titanic_data['Age'].mean()).head()
```

Out[34]:
```
0    22.0
1    38.0
2    26.0
3    35.0
4    35.0
Name: Age, dtype: float64
```

In [41]:
```
# Finding number of entries, listing the dataset, finding datatypes
titanic_data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 891 entries, 0 to 890
Data columns (total 12 columns):
PassengerId    891 non-null int64
Survived       891 non-null int64
Pclass         891 non-null int64
Name           891 non-null object
Sex            891 non-null object
Age            714 non-null float64
SibSp          891 non-null int64
Parch          891 non-null int64
Ticket         891 non-null object
Fare           891 non-null float64
Cabin          204 non-null object
Embarked       889 non-null object
dtypes: float64(2), int64(5), object(5)
memory usage: 83.6+ KB
```

In [46]:
```
# Min and Max Age
print("Min Age:  ", titanic_data['Age'].min(), "(Cute infants were there too!)")
print("Max Age: ", titanic_data['Age'].max())
```

```
Min Age:   0.42 (Cute infants were there too!)
Max Age:  80.0
```

In [47]:
```
# Mean Value of Age attribute
print("Mean Age: ", titanic_data['Age'].mean())
```

```
Mean Age:  29.69911764705882
```

In [53]:
```
# Plotting a Histogram
age_plt = titanic_data['Age'].hist(bins=100)
age_plt.set_xlabel("Age")
age_plt.set_ylabel("Number of people")
age_plt.set_title("Histogram of Age")
```

Out[53]: Text(0.5,1,'Histogram of Age')