

Pandas 6

Visualizations

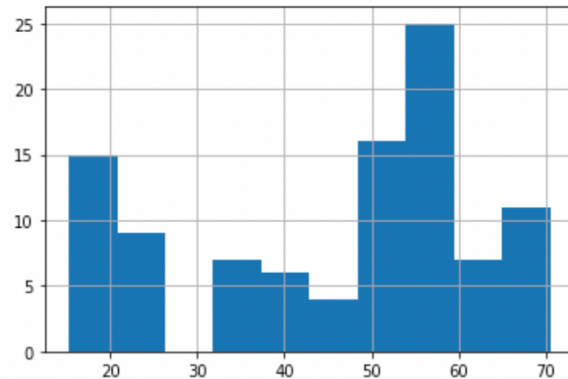
- Making use of Matplotlib to visualize data.
- .hist() to plot hist
- plt.show() - To display the plot
- Here x axis represents the heights of the dogs
- y axis represents the number of dogs of that height

Histograms

```
import matplotlib.pyplot as plt
```

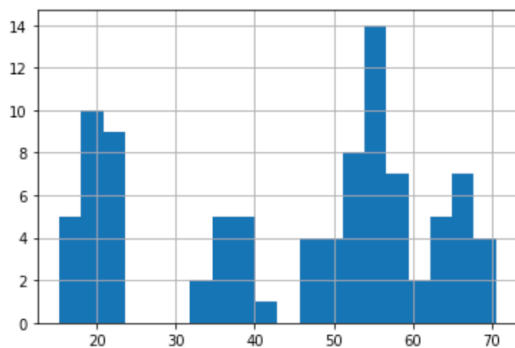
```
dog_pack["height_cm"].hist()
```

```
plt.show()
```

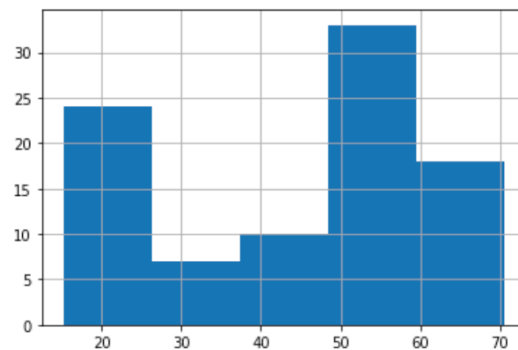


Histograms

```
dog_pack["height_cm"].hist(bins=20)  
plt.show()
```



```
dog_pack["height_cm"].hist(bins=5)  
plt.show()
```



Bar plots

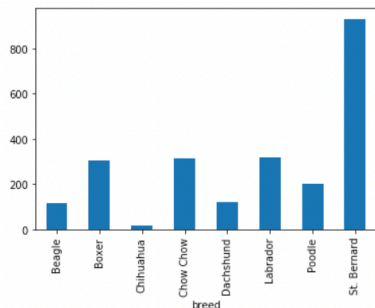
```
avg_weight_by_breed = dog_pack.groupby("breed")["weight_kg"].mean()  
print(avg_weight_by_breed)
```

```
breed  
Beagle      10.636364  
Boxer       30.620000  
Chihuahua   1.491667  
Chow Chow   22.535714  
Dachshund   9.975000  
Labrador    31.850000  
Poodle      20.400000  
St. Bernard 71.576923  
Name: weight_kg, dtype: float64
```

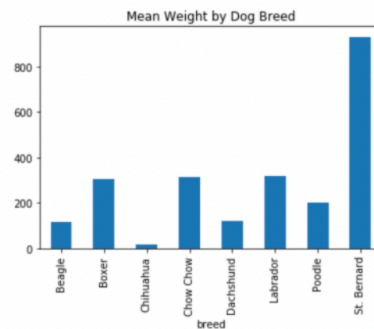
To get the average weight of each breed, we will first group by the breed, then select the weight_kg column and compute its mean
Creating bar plot using mean

Bar plots

```
avg_weight_by_breed.plot(kind="bar")  
plt.show()
```

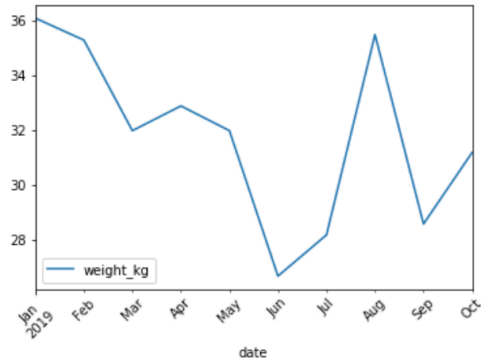


```
avg_weight_by_breed.plot(kind="bar",  
                           title="Mean Weight by Dog Breed")  
plt.show()
```



Rotating axis labels

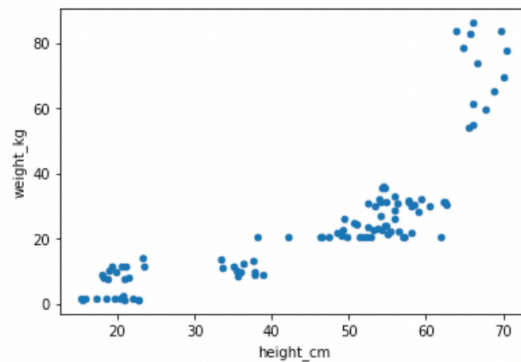
```
sully.plot(x="date", y="weight_kg", kind="line", rot=45)  
plt.show()
```



Rotating axis labels by 45 degrees to make the text easy to read.

Scatter plots

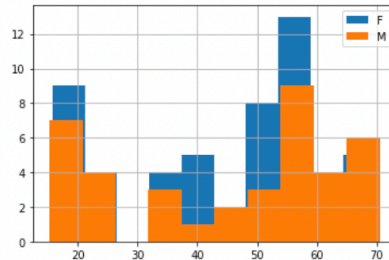
```
dog_pack.plot(x="height_cm", y="weight_kg", kind="scatter")  
plt.show()
```



Layering plots

Add a legend

```
dog_pack[dog_pack["sex"]=="F"]["height_cm"].hist()
dog_pack[dog_pack["sex"]=="M"]["height_cm"].hist()
plt.legend(["F", "M"])
plt.show()
```



to make overlapping histograms translucent, we will specify it in the alpha variable

Code Snippets

```
# Import matplotlib.pyplot with alias plt
import matplotlib.pyplot as plt
```

```
# Get the total number of avocados sold on each date
nb_sold_by_date = avocados.groupby('date')['nb_sold'].sum()
```

```
# Create a line plot of the number of avocados sold by date
nb_sold_by_date.plot(x = 'date', y = 'nb_sold', kind = 'line', rot = 45)
```

```
# Show the plot
plt.show()
```

Missing Values

.isna() - Returns a boolean value indicating whether the value is missing or not.

.isna().sum() - To find the number of missing values, True is indicated as 1 and false as 0. Hence, the sum will give us the number of missing values

.isna.any() - This returns True if there is any missing value in the column for the dataset;

```
dogs.isna().any()
```

```
name           False
breed          False
color          False
height_cm      False
weight_kg       True
date_of_birth  False
dtype: bool
```

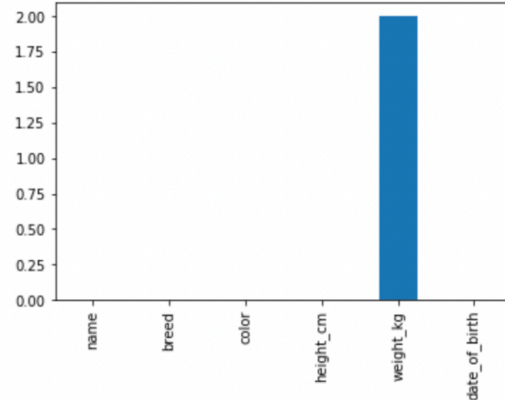
Counting missing values

```
dogs.isna().sum()
```

```
name           0
breed          0
color          0
height_cm      0
weight_kg       2
date_of_birth  0
dtype: int64
```

Plotting missing values

```
import matplotlib.pyplot as plt
dogs.isna().sum().plot(kind="bar")
plt.show()
```



Removing missing values

`df.dropna()` - This method will drop all the rows having missing values in any of the columns

Removing missing values

```
dogs.dropna()
```

	name	breed	color	height_cm	weight_kg	date_of_birth
1	Charlie	Poodle	Black	43	24.0	2016-09-16
2	Lucy	Chow Chow	Brown	46	24.0	2014-08-25
4	Max	Labrador	Black	59	29.0	2017-01-20
5	Stella	Chihuahua	Tan	18	2.0	2015-04-20
6	Bernie	St. Bernard	White	77	74.0	2018-02-27

Replacing missing values

`.fillna(0)`

Replacing missing values

```
dogs.fillna(0)
```

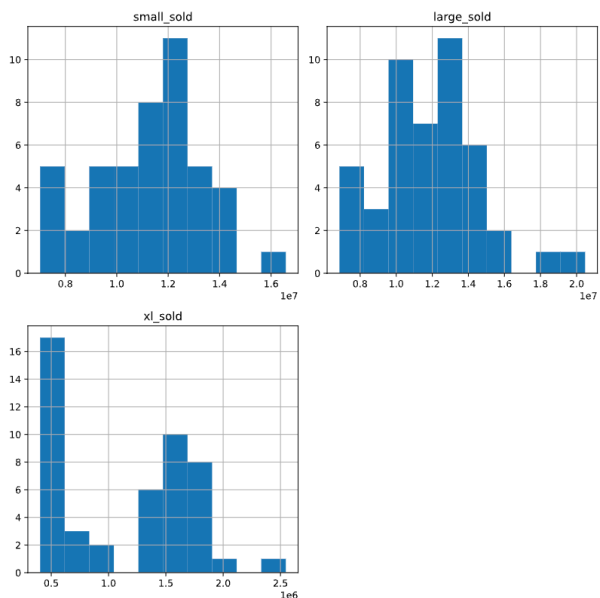
	name	breed	color	height_cm	weight_kg	date_of_birth
0	Bella	Labrador	Brown	56	0.0	2013-07-01
1	Charlie	Poodle	Black	43	24.0	2016-09-16
2	Lucy	Chow Chow	Brown	46	24.0	2014-08-25
3	Cooper	Schnauzer	Gray	49	0.0	2011-12-11
4	Max	Labrador	Black	59	29.0	2017-01-20
5	Stella	Chihuahua	Tan	18	2.0	2015-04-20
6	Bernie	St. Bernard	White	77	74.0	2018-02-27

script.py

Light Mode

```
1 # List the columns with
  missing values
2 cols_with_missing =
  ["small_sold",
  "large_sold", "xl_sold"]
3
4 # Create histograms
  showing the distributions
  cols_with_missing
5 avocados_2016
  [cols_with_missing].hist()
6
7 # Show the plot
8 plt.show()
```

Plots



Creating Dataframes

List of dictionaries - by row

name	breed	height (cm)	weight (kg)	date of birth
Ginger	Dachshund	22	10	2019-03-14
Scout	Dalmatian	59	25	2019-05-09

```
new_dogs = pd.DataFrame(list_of_dicts)
print(new_dogs)
```

```
   name    breed  height_cm  weight_kg  date_of_birth
0  Ginger  Dachshund      22         10   2019-03-14
1  Scout   Dalmatian      59         25   2019-05-09
```

Dictionary of lists - by column

name	breed	height	weight	date of birth
Ginger	Dachshund	22	10	2019-03-14
Scout	Dalmatian	59	25	2019-05-09

- **Key** = column name
- **Value** = list of column values

```
dict_of_lists = {
    "name": ["Ginger", "Scout"],
    "breed": ["Dachshund", "Dalmatian"],
    "height_cm": [22, 59],
    "weight_kg": [10, 25],
    "date_of_birth": ["2019-03-14",
                     "2019-05-09"]
}
new_dogs = pd.DataFrame(dict_of_lists)
```


DataFrame to CSV

```
new_dogs.to_csv("new_dogs_with_bmi.csv")
```