

**Machine Learning and Pattern Recognition LAB**  
**Chandra Prakash**  
**July - Dec 2018**  
**LAB 1: Pre-processing of DATA using Python**

**VIPASHA DHIMAN**  
**04401032015**  
**B.Tech(IT)**

**1. Identify**

*the variables/features and output/result/classes in the given dataset*

*dait\_dataset.xls*

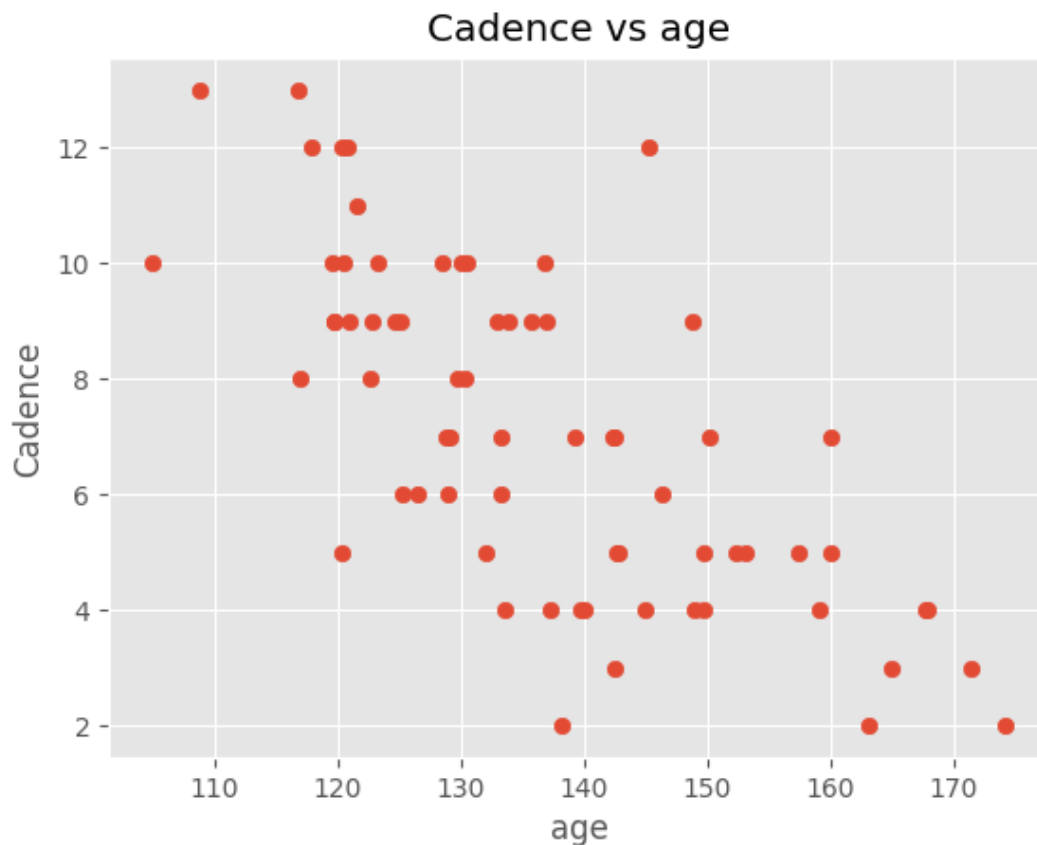
*a) What this dataset is all about?*

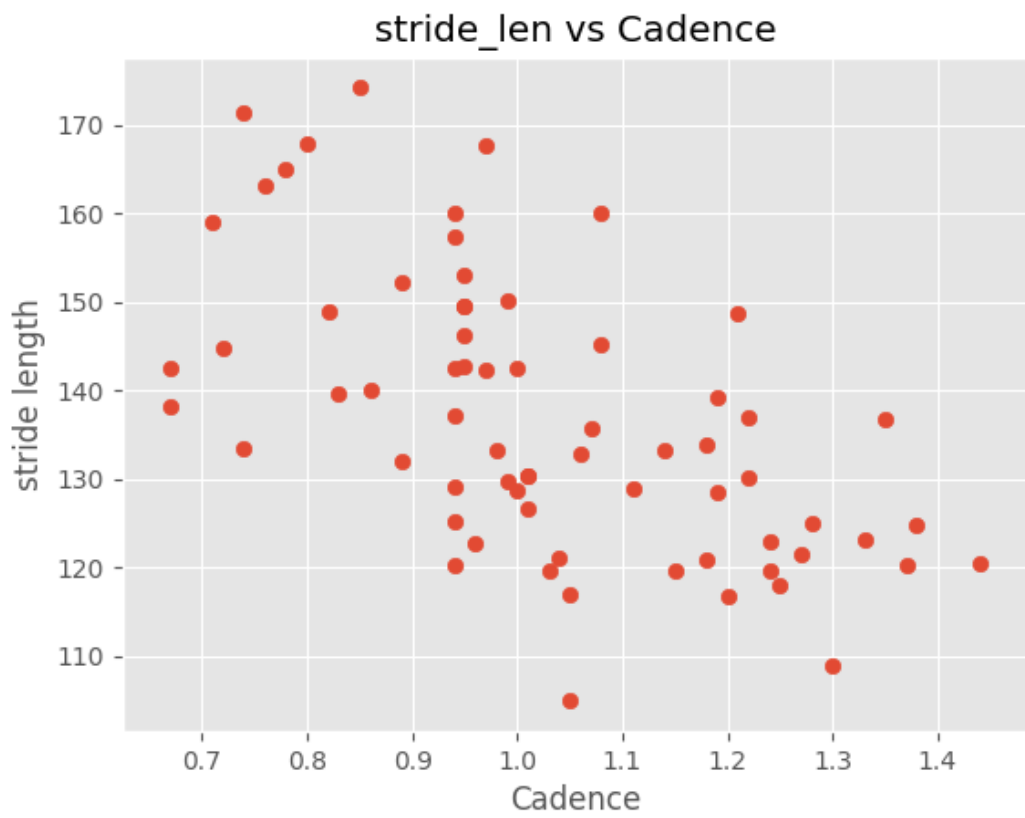
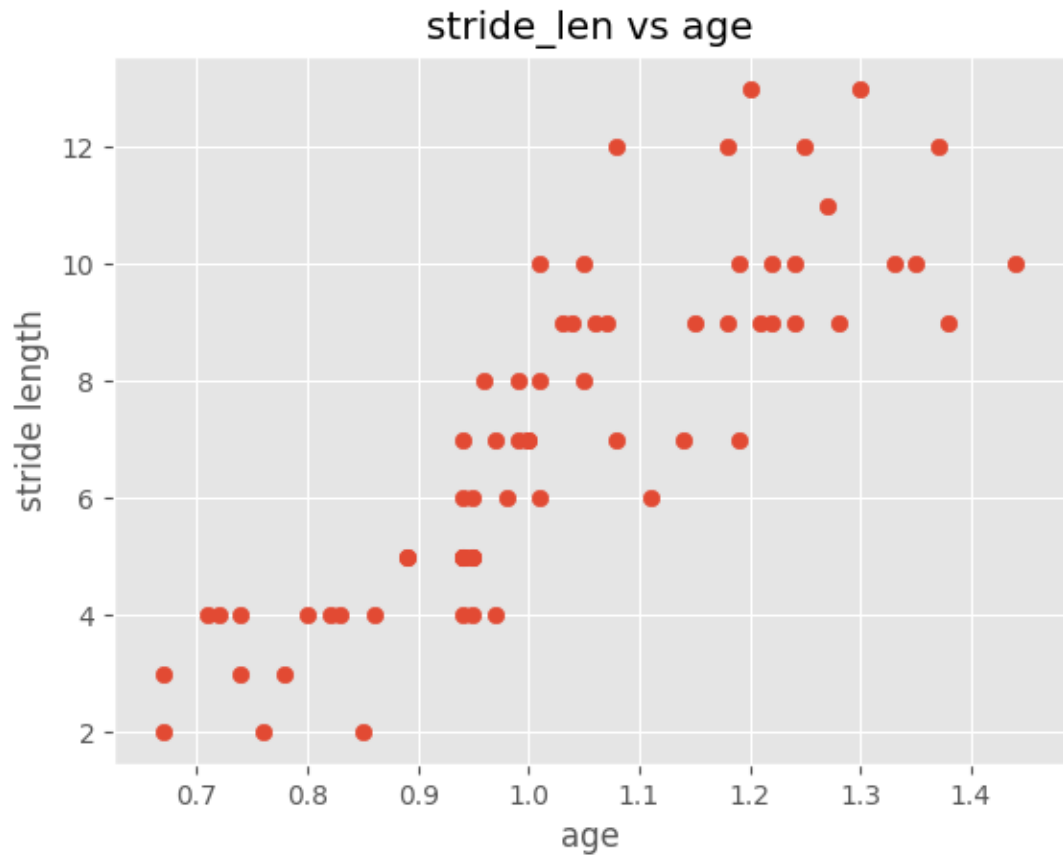
This dataset(Gait\_data.xls) is about gait analysis in human beings with and without cerebral palsy. It has 4 features: stride length (m), cadence (step/min), leg len (m), age (year)

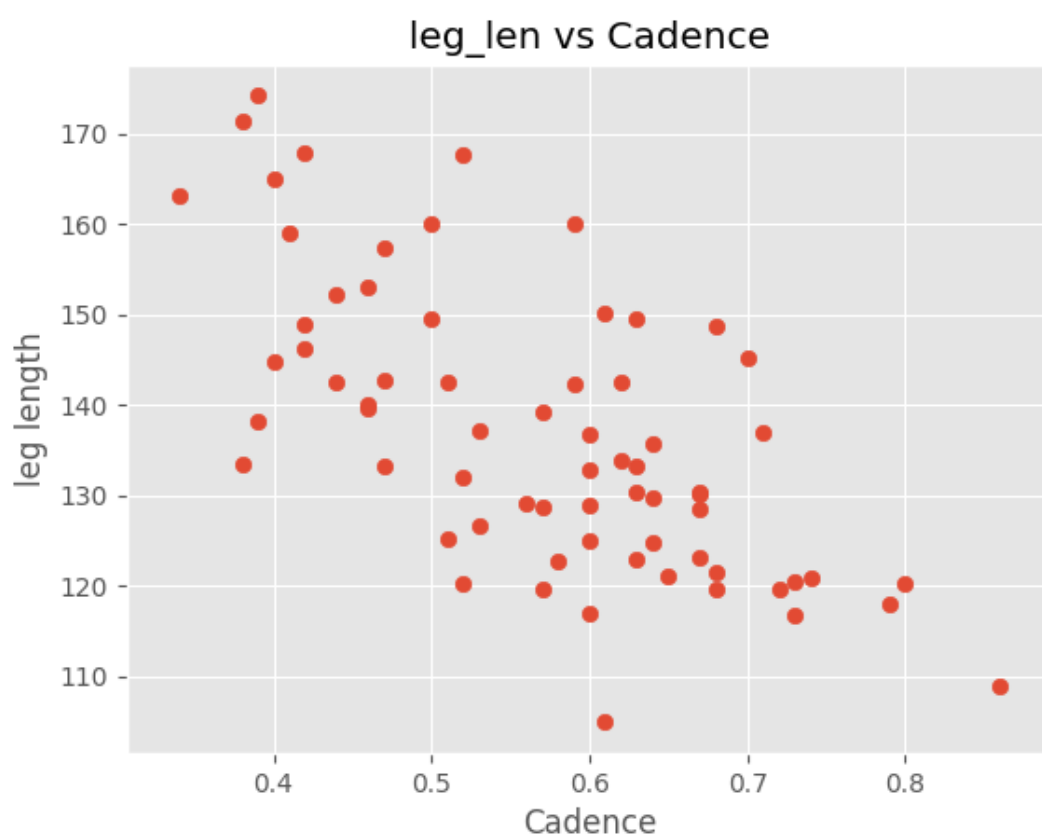
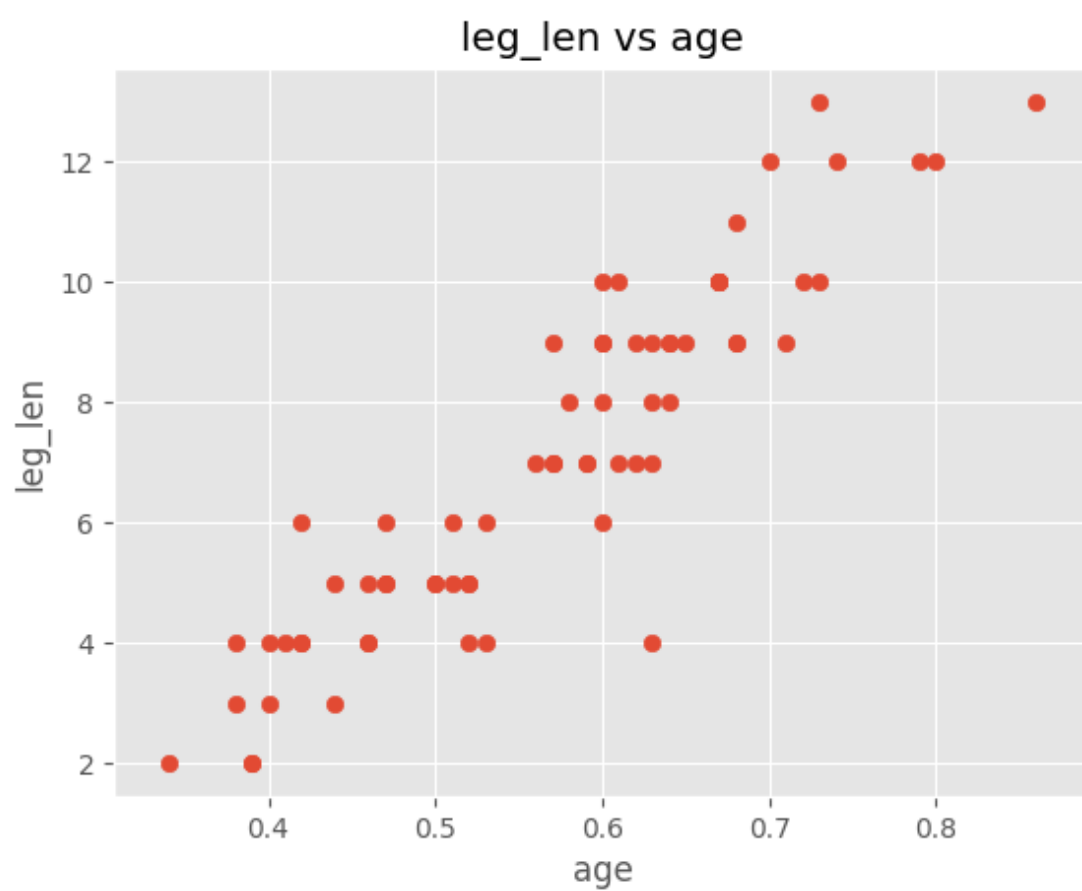
**2. Visual Representation**

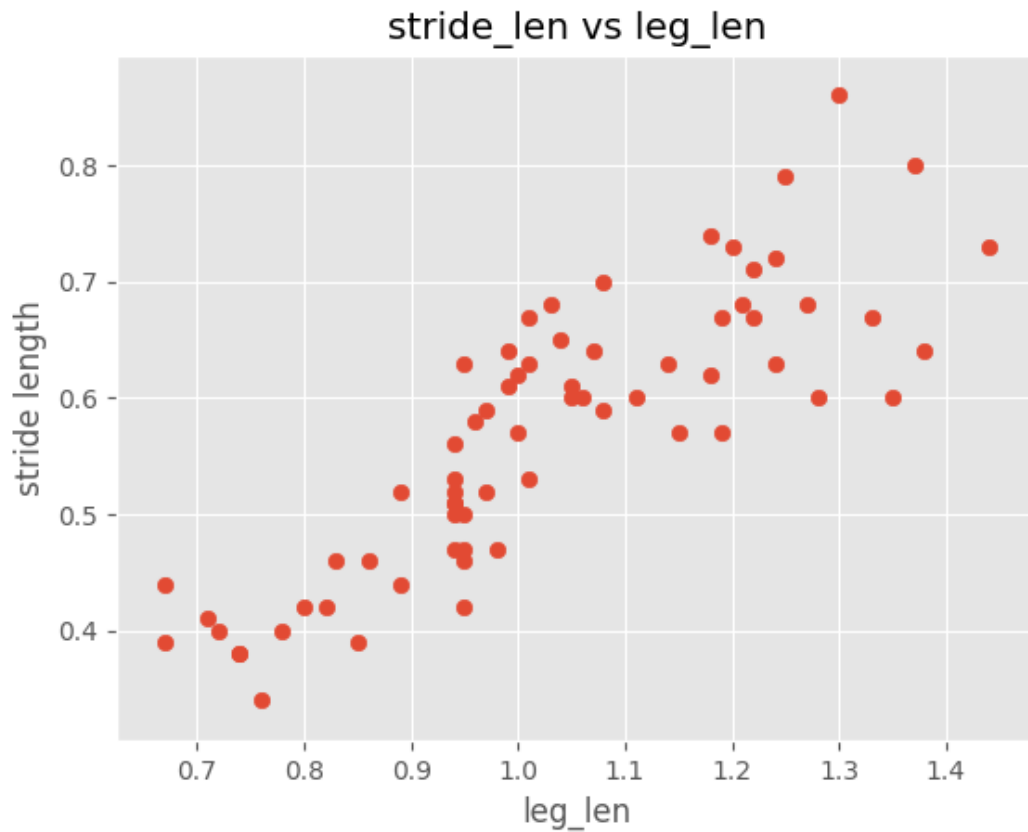
**For Normal Subjects**

- a) Draw scatter plot of two variables at a time
- b) Find the Correlation Coefficient









Correlation coefficients:

Stride length vs Cadence = -0.58877574

Stride length vs leg length = 0.84743651

Stride length vs age = 0.84986344

Cadence vs leg length = -0.66914141

Cadence vs age = -0.70974936

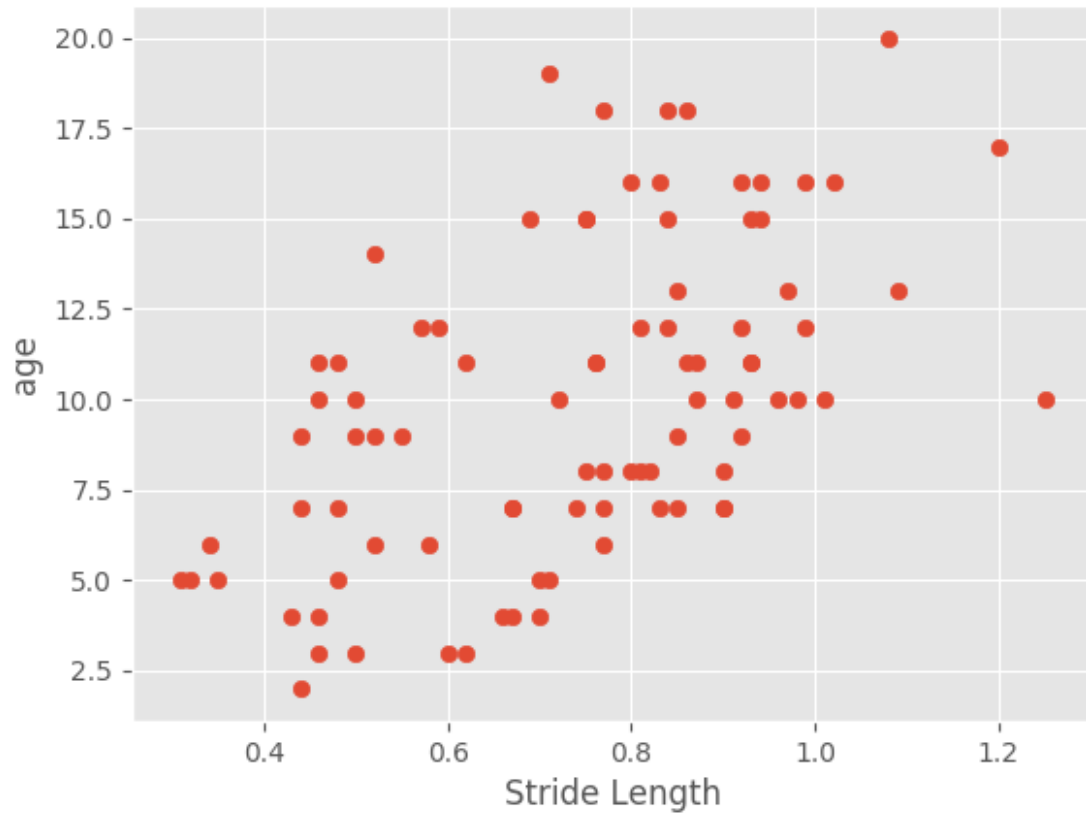
Leg length vs age = 0.91683714

**For Abnormal Subjects (Children with Cerebral Palsy)**

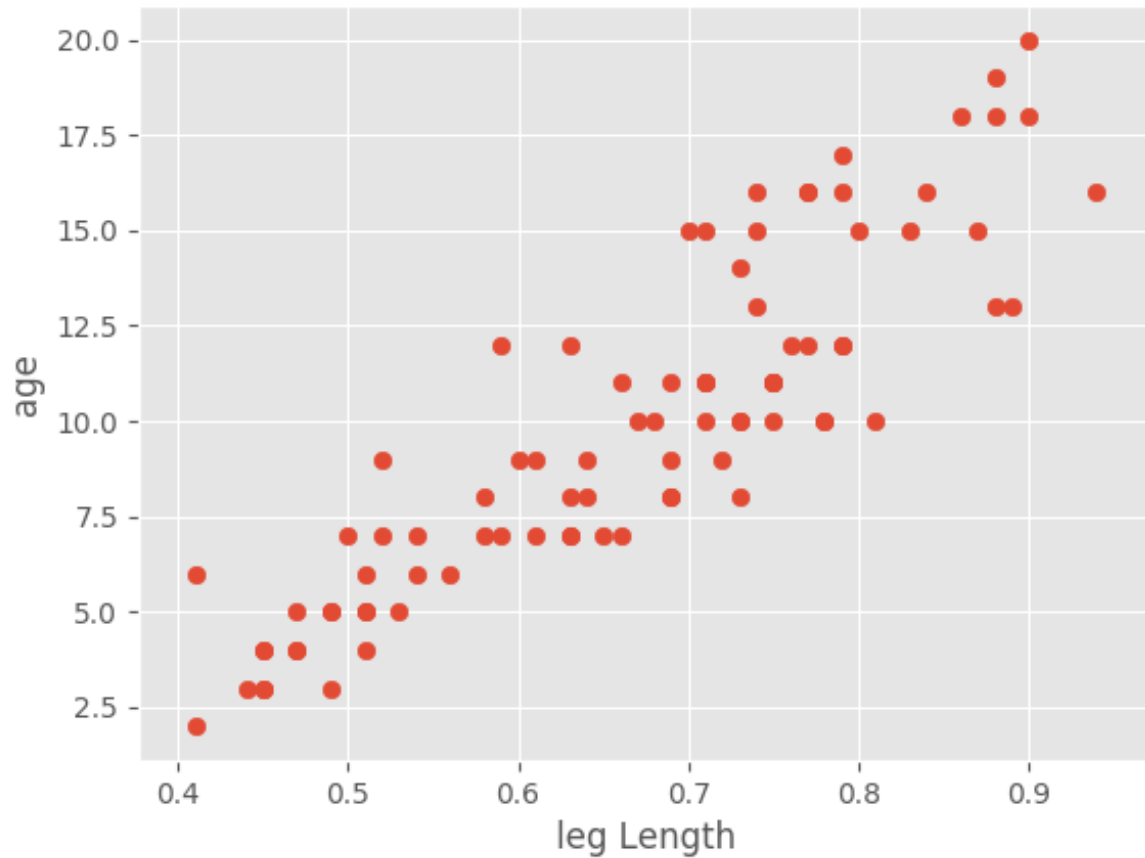
c) Draw scatter plot of two variables at a time

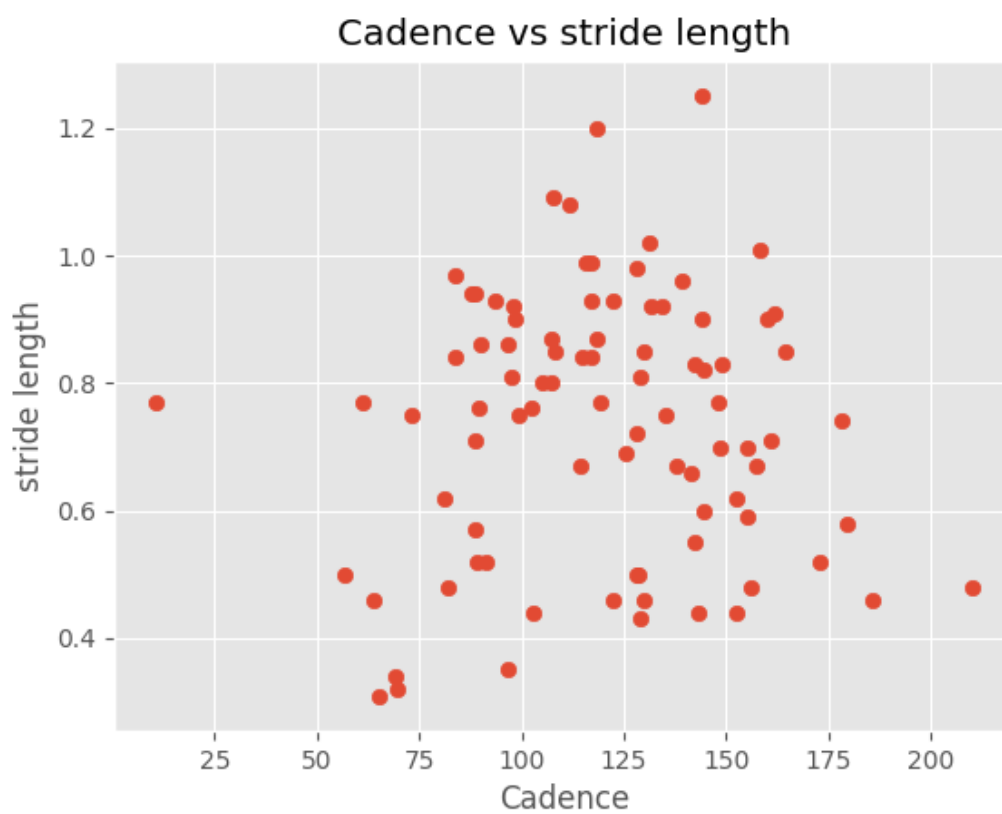
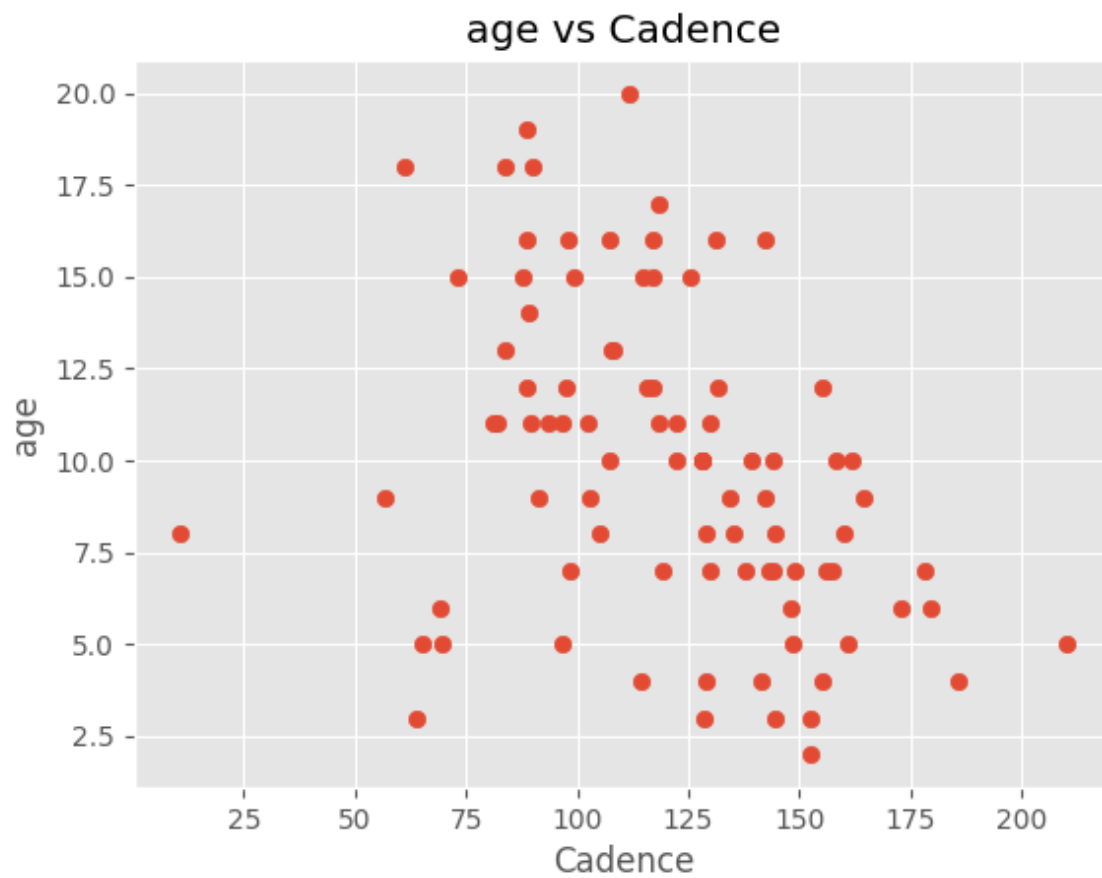
d) Find the Correlation Coefficient

age vs Stride Length

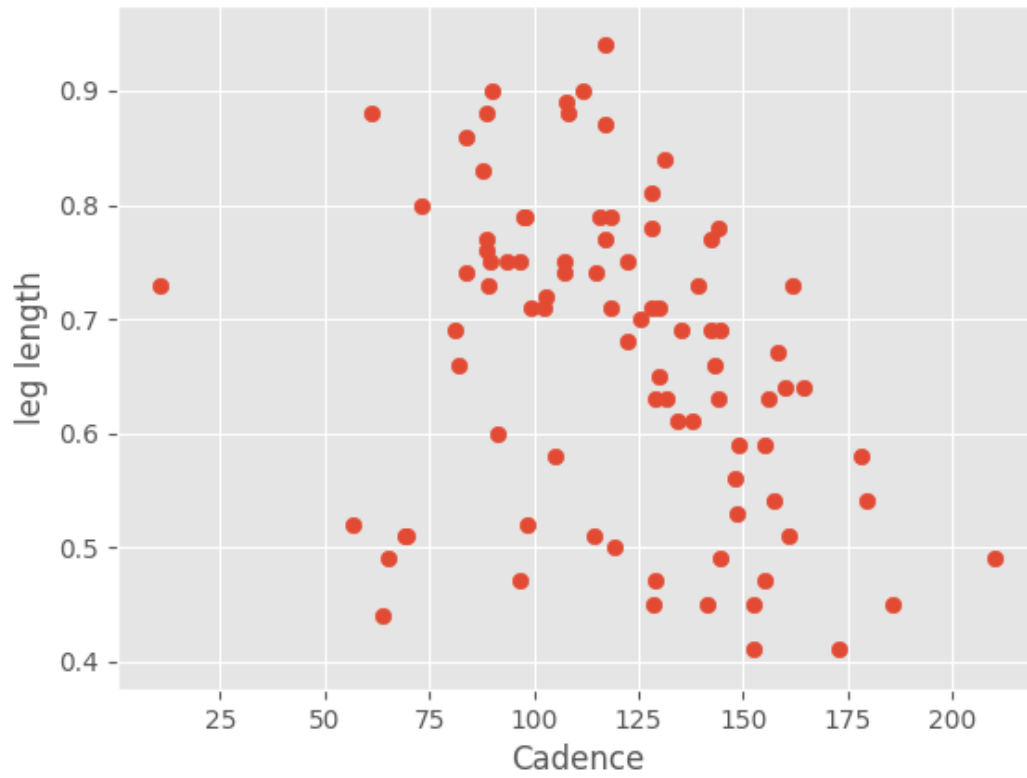


age vs leg Length

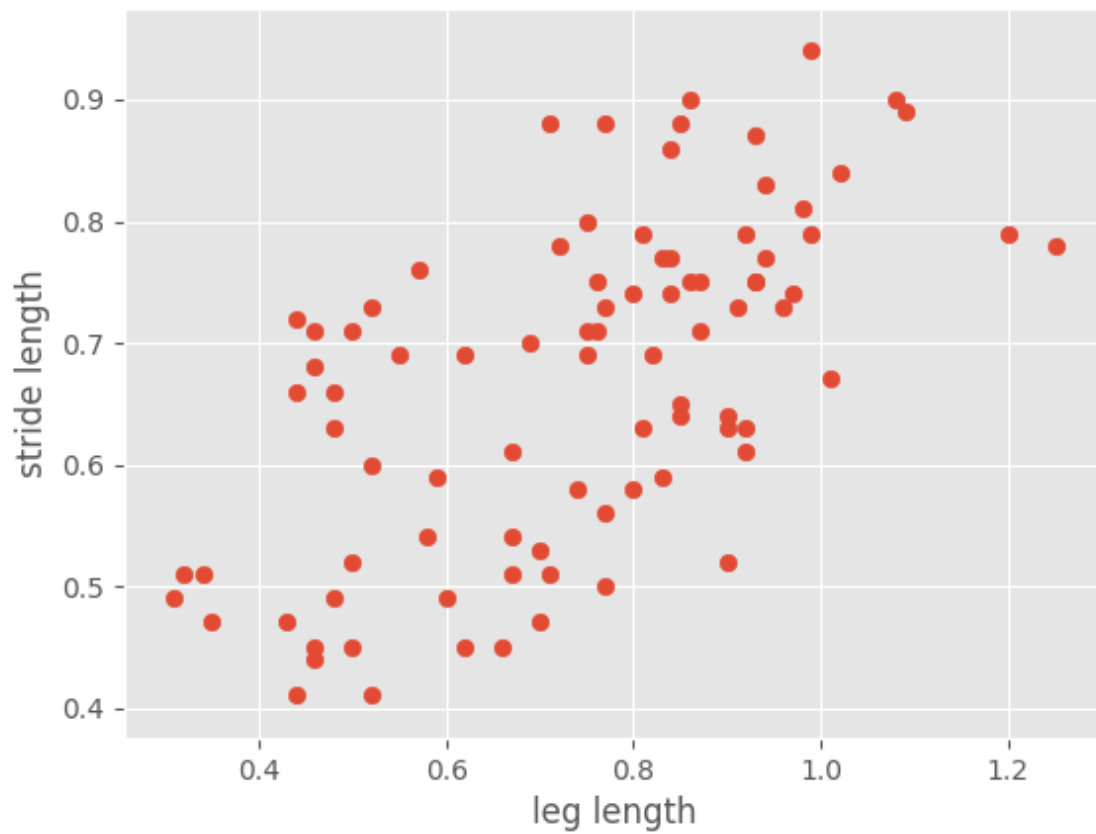




Cadence vs leg length



stride length vs leg length



Correlation coefficients:

Stride length vs Cadence = 0.03199596

Stride length vs leg length = 0.62040285

Stride length vs age = 0.53976861

Cadence vs leg length = -0.3431222

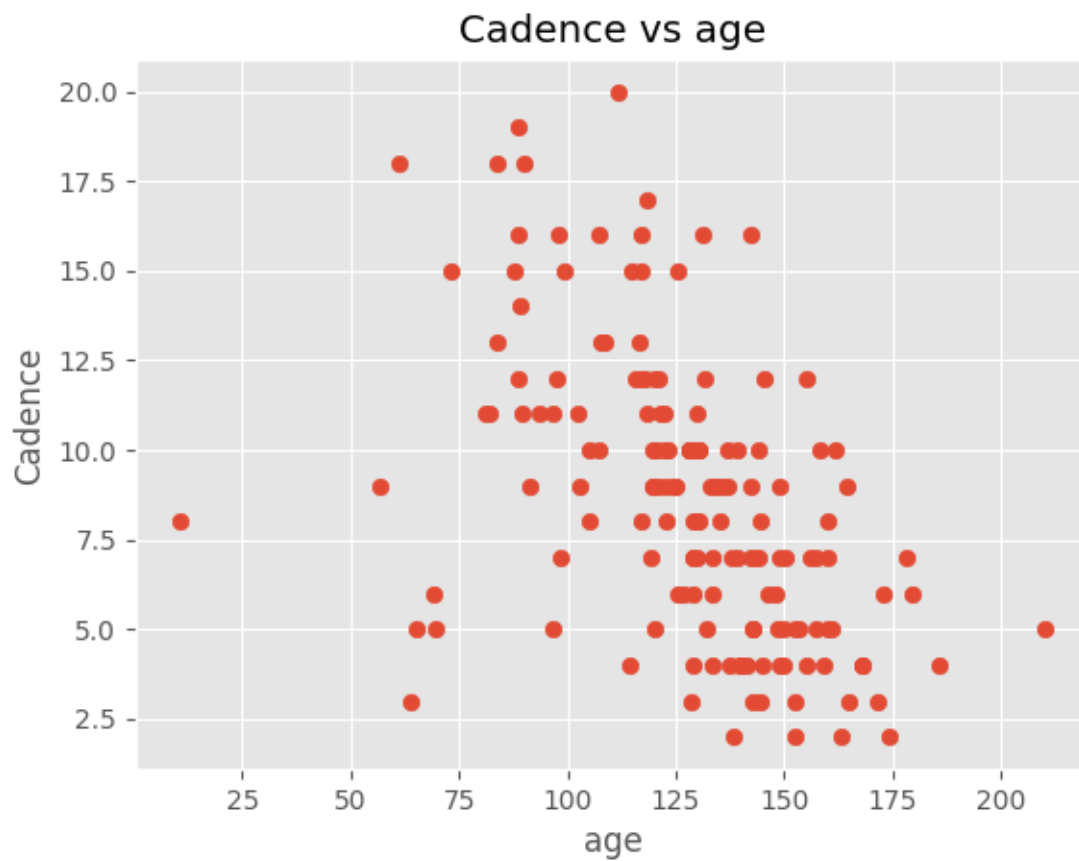
Cadence vs age = -0.37354016

Leg length vs age = 0.88580534

### For ALL Subjects

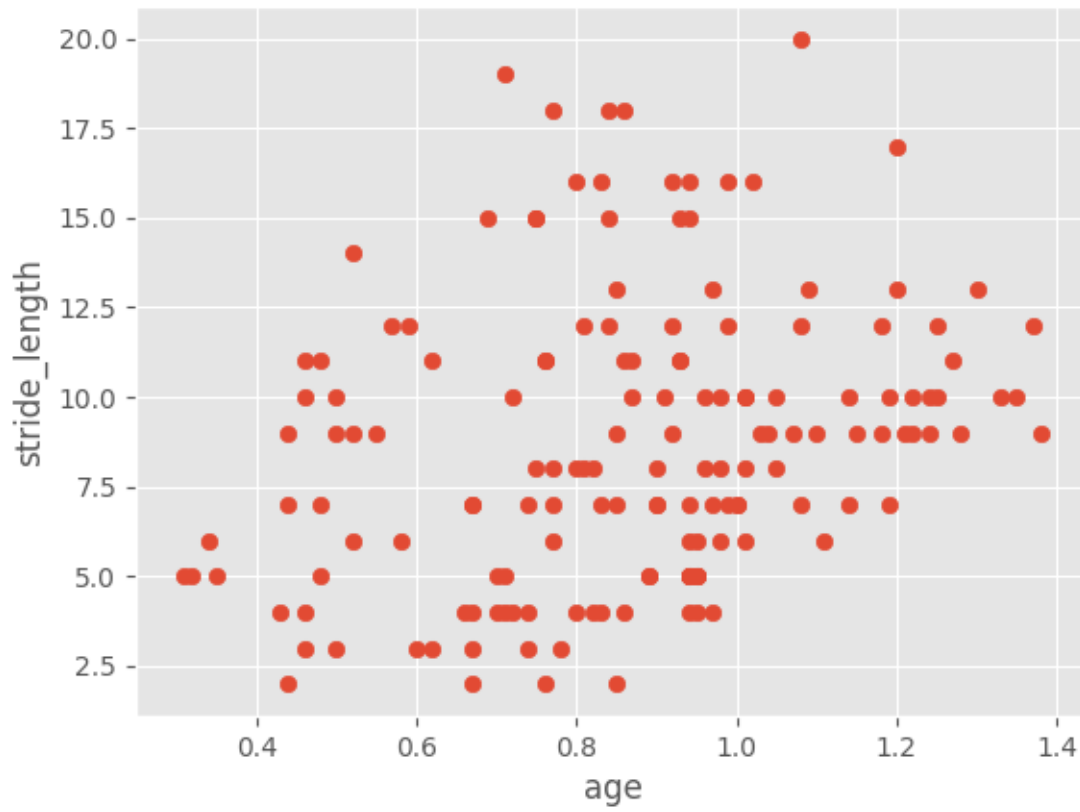
e) Draw scatter plot of two variables at a time

f) Find the Correlation Coefficient

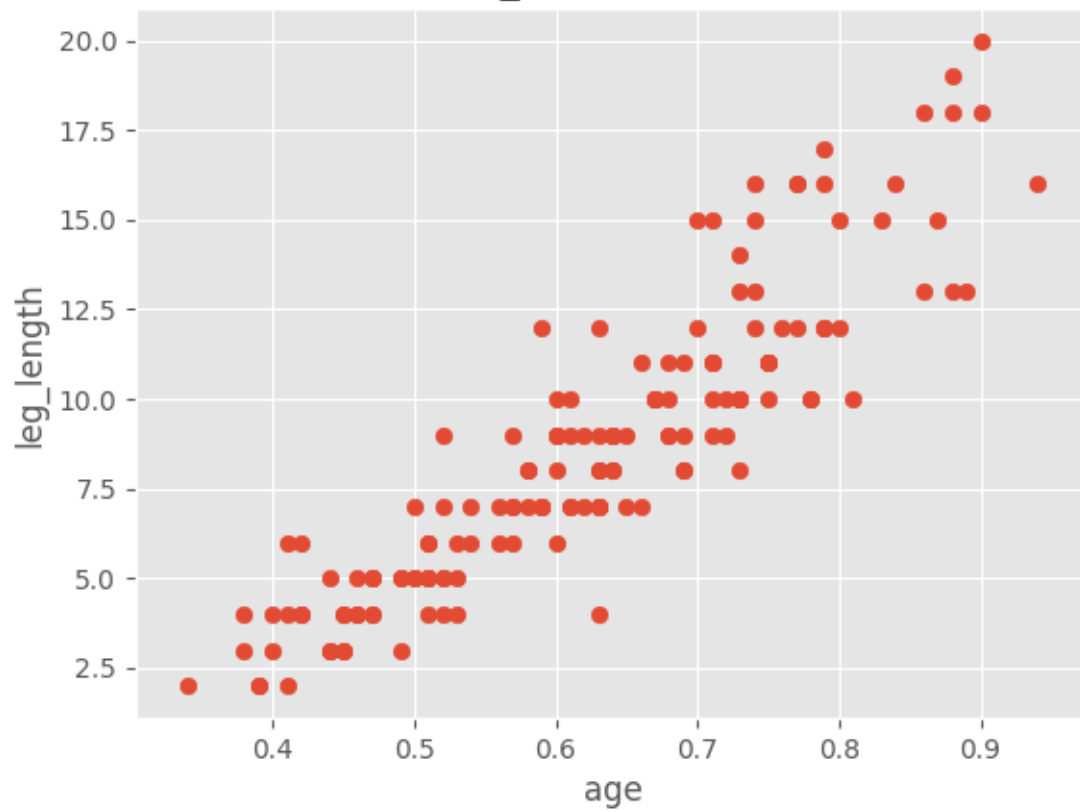




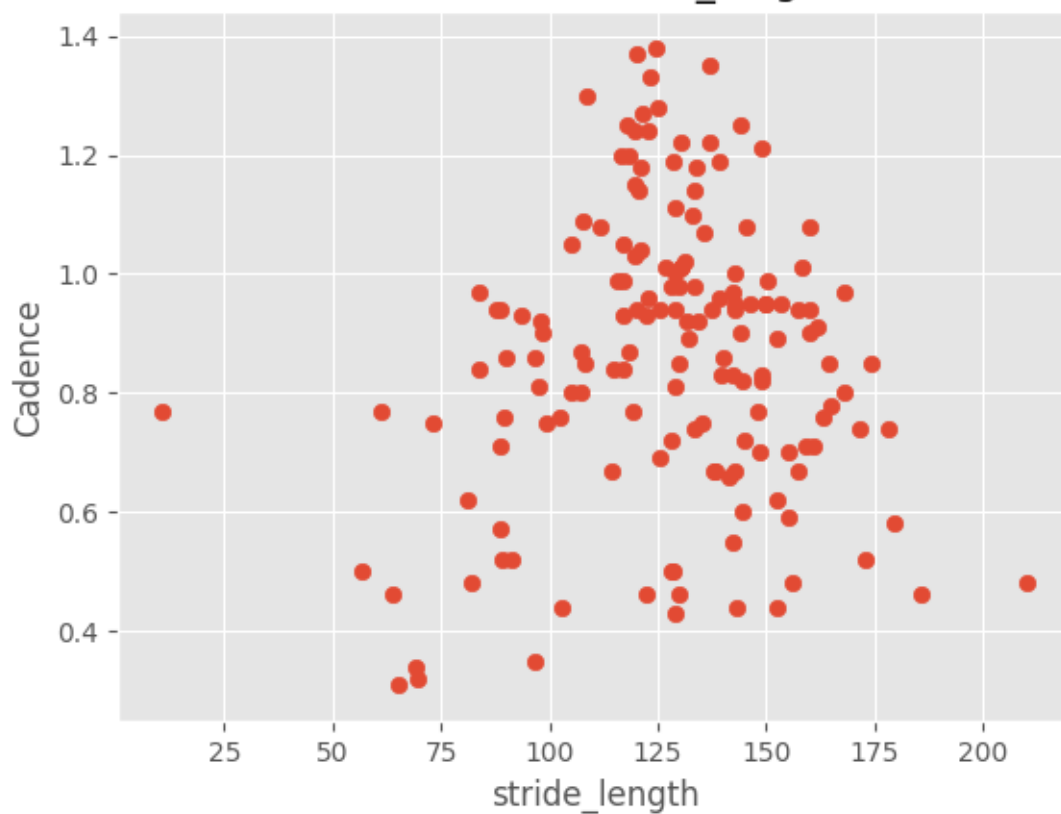
stride\_length vs age



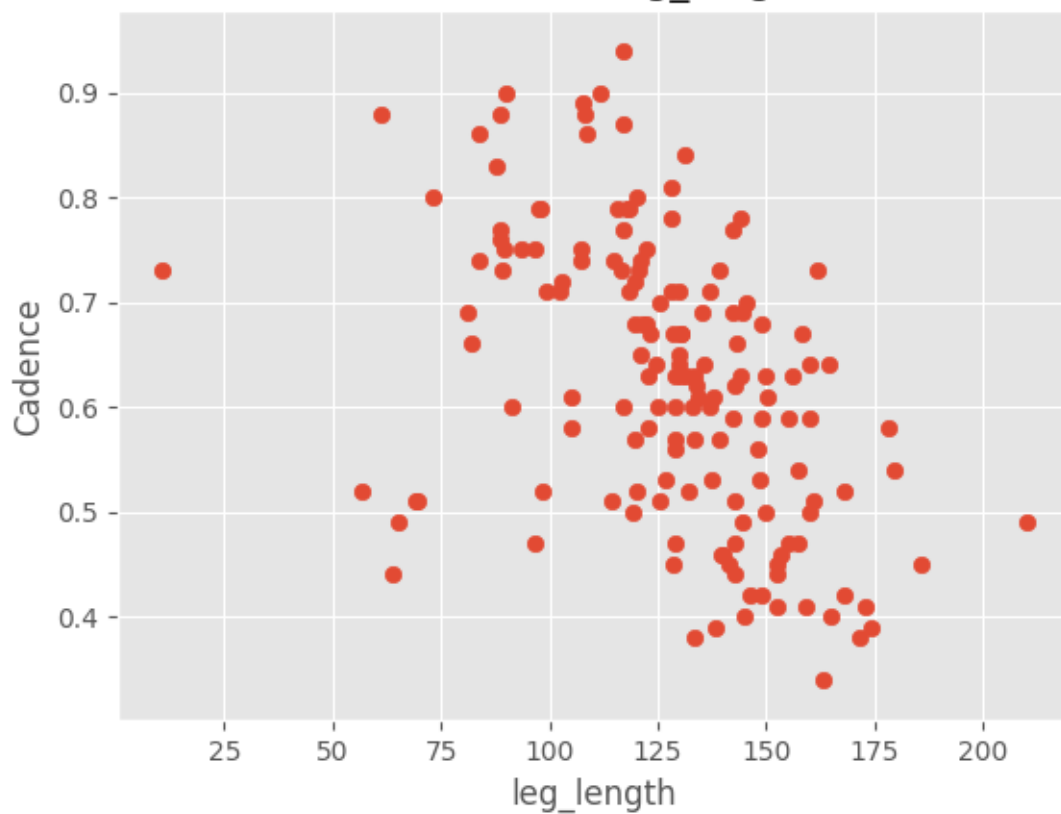
leg\_length vs age

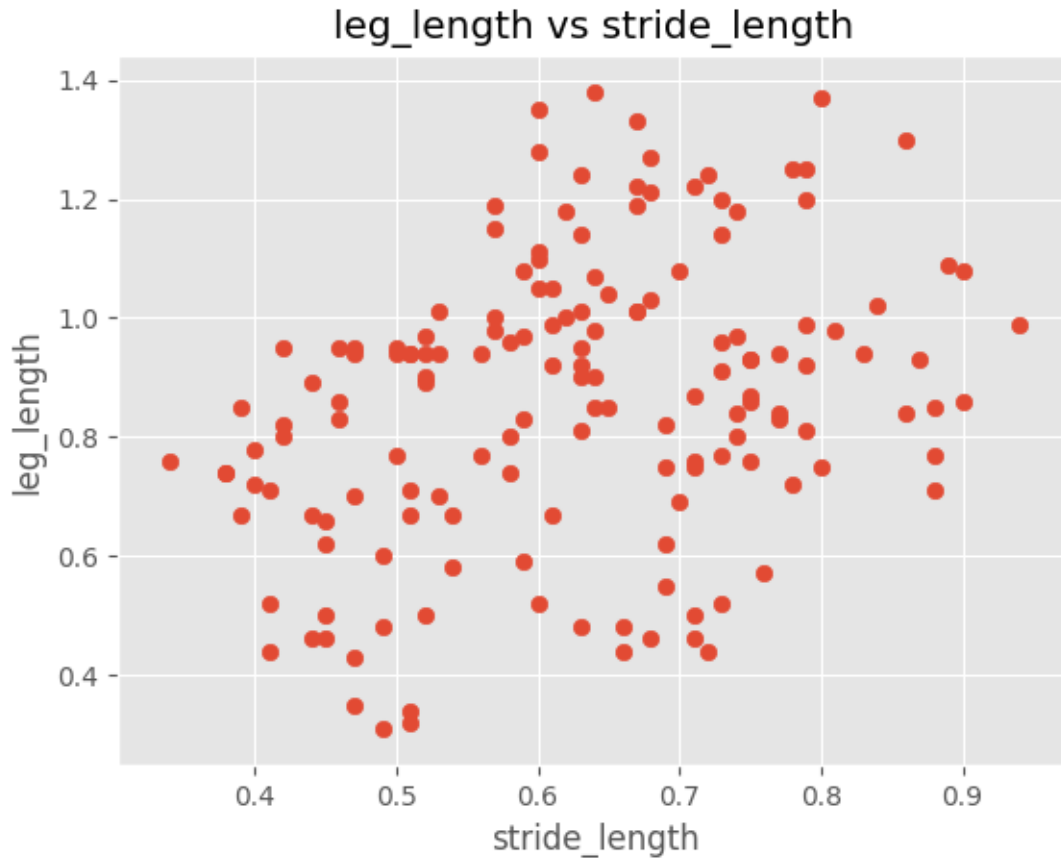


Cadence vs stride\_length



Cadence vs leg\_length





Correlation coefficients:

Cadence vs Stride length = 0.08757209

leg length vs Stride length = 0.32593509

Stride length vs age = 0.27533762

Cadence vs leg length = -0.46857984

Cadence vs age = -0.49176509

Leg length vs age = 0.90287514

### ***3. Statistical Analysis of data set***

***a) Minimum and maximum value***

***e***

***b) mean of each variable/variables/features/input***

***c) Median***

***d) Mode***

***e) Deviation***

***f) Standard deviation***

**Statistical Analysis of data**

### **1. For normal**

Mean of stride\_length is:1.02720588235  
Standard deviation of stride\_length is:0.18555962624  
Median of stride\_length is:1.0  
Mode of stride\_length is:0.94  
Mean of Cadence is:136.839705882  
Standard deviation of Cadence is:15.8129266949  
Median of Cadence is:133.71  
Mode of Cadence is:No unique mode found  
Mean of leg\_length is:0.569705882353  
Standard deviation of leg\_length is:0.11709124632  
Median of leg\_length is:0.59  
Mode of leg\_length is:0.6  
Mean of age is:7.08823529412  
Standard deviation of age is:2.89485026096  
Median of age is:7.0  
Mode of age is:9

### **2. For abnormal**

Mean of stride\_length is:0.736818181818  
Standard deviation of stride\_length is:0.20824689952  
Median of stride\_length is:0.765  
Mode of stride\_length is:No unique mode found  
Mean of Cadence is:120.002727273  
Standard deviation of Cadence is:33.5594737032  
Median of Cadence is:120.72  
Mode of Cadence is:No unique mode found  
Mean of leg\_length is:0.666590909091  
Standard deviation of leg\_length is:0.136043319366  
Median of leg\_length is:0.69  
Mode of leg\_length is:No unique mode found  
Mean of age is:9.88636363636  
Standard deviation of age is:4.33757127387  
Median of age is:10.0  
Mode of age is:7

### **3. For all**

Mean of stride\_length is:127.344487179  
Standard deviation of stride\_length is:28.4638166075  
Median of stride\_length is:129.39  
Mode of stride\_length is:No unique mode found  
Mean of Cadence is:0.625  
Standard deviation of Cadence is:0.136024191397  
Median of Cadence is:0.63  
Mode of Cadence is:0.63  
Mean of leg\_length is:8.66666666667  
Standard deviation of leg\_length is:4.01502554257  
Median of leg\_length is:8.5

Mode of leg\_length is:7

Mean of age is:1.5641025641

Standard deviation of age is:0.497470857366

Median of age is:2.0

Mode of age is:2

### ***Observations***

- Leg length and age has highest correlation for 'normal' and 'abnormal subjects'
- Cadence and leg length has highest correlation for 'all subjects'
- Mean of stride length is maximum for 'all subjects'