

# Assignment-1

Aravind

2024-02-01

*#Reference:*

*#Data is taken from <https://opendataphilly.org/dataset/philadelphia-universities-and-colleges/resource/baa34ebf-0352-4c33-a9f6-8ae061a50939>*

```
dataset1 = read.csv("C:/Users/aravi/OneDrive/Documents/Universities List - 1.csv")
dataset1
```

##	University_names	PARCEL_ID	GROSS_AREA	Shape__Area	Shape__Length
## 1	Temple University	155059	231347	231347.865	2519.9679
## 2	Temple University	392760	47425	47426.015	1281.7251
## 3	Temple University Medical	13161	29301	29348.387	707.2776
## 4	Temple University Medical	324604	34030	33840.215	746.5503
## 5	Temple University Medical	18474	22500	22697.679	602.6341
## 6	Temple University Medical	272372	506603	498503.438	3085.6383
## 7	University of the Arts	208422	13350	13349.988	477.9994
## 8	University of the Arts	542656	38640	39003.837	931.2586
## 9	University of the Arts	88046	16648	16634.061	708.5103
## 10	University of the Arts	139329	13567	13747.561	469.1568
## 11	University of the Arts	195308	5193	5213.345	340.0758
## 12	University of the Arts	259264	19294	19378.185	615.5444
## 13	University of the Arts	288419	4920	5144.847	327.9308
## 14	University of the Arts	487380	21150	21081.861	581.2803
## 15	La Salle University	16858	1383	1369.704	208.3377
##	Code				

```
## 1   Green
## 2   Green
## 3     Red
## 4     Red
## 5     Red
## 6     Red
## 7  Orange
## 8  Orange
## 9  Orange
## 10 Orange
## 11 Orange
## 12 Orange
## 13 Orange
## 14 Orange
## 15   Blue
```

*#Quantitative Descriptive Statistics*

```
mean(dataset1$Shape__Area)
```

```
## [1] 66539.13
```

```
sd(dataset1$Shape__Area)
```

```
## [1] 131862.3
```

*#Categorical Variables*

```
table(dataset1$Code)
```

```
##
```

```
##   Blue   Green Orange    Red
```

```
##     1     2     8     4
```

```
dataset1$Shape__Area = mean(dataset1$Shape__Area) - sd(dataset1$Shape__Area)
dataset1$Shape__Area
```

```
## [1] -65323.15 -65323.15 -65323.15 -65323.15 -65323.15 -65323.15 -65323.15
```

```
## [8] -65323.15 -65323.15 -65323.15 -65323.15 -65323.15 -65323.15 -65323.15
```

```
## [15] -65323.15
```

*#Below is the example for Histogram of Quantitative variables*

*#Scatter Plot*

```
hist(dataset1$Shape__Length)
```

A histogram showing the frequency distribution of the variable 'Shape\_Length' from the 'dataset1' dataset. The x-axis is labeled 'dataset1\$Shape\_Length' and has major tick marks at 0, 500, 1000, 2000, and 3000. The y-axis is labeled 'Frequency' and has major tick marks from 0 to 7. The histogram consists of several bars: a bar from 0 to 500 with a frequency of 5, a bar from 500 to 1000 with a frequency of 7, a bar from 1000 to 1500 with a frequency of 1, a bar from 2500 to 3000 with a frequency of 1, and a bar from 3000 to 3500 with a frequency of 1. There are no bars between 1500 and 2500.

Shape_Length Range	Frequency
0 - 500	5
500 - 1000	7
1000 - 1500	1
1500 - 2000	0
2000 - 2500	0
2500 - 3000	1
3000 - 3500	1

A scatter plot titled "Area and Length" showing the relationship between Area (x-axis) and Length (y-axis). The x-axis ranges from -90,000 to -40,000, and the y-axis ranges from 0 to 3,000. The data points are clustered around Area = -65,000, with Length values ranging from approximately 200 to 3,000.