

# DATA VISUALIZATION WITH PYTHON SESSION - 2

## IN-CLASS NOTLARI

(21 Aralık 2022)

### Fonksiyonel & Object Oriented yöntem:

Birinde doğrudan fonksiyonları kullanıyoruz. Diğerinde object'lere metodlar uyguluyoruz. Aynı işlevi gören fonksiyon ve metod isimlerinde farklılıklar olabiliyor. Bunları notebook'a not alıp geçelim.

\*\*\*\*\*

**Subplots()** fonksiyonu içindeki parametrelerde kaç adet axes tanımlandıysa figure içinde o kadar axes çizdirebiliyorsunuz.

\*\*\*\*\*

Statistics derslerinden iki sayısal değişkenin arasında oluşacak **relationship/association**, **direction** ve **strength** konseptlerini hatırlamakta fayda var. Scatter plot bize ilk iki concept hakkında bilgi verir

\*\*\*\*\*

**Scatter plot** iki farklı sayısal değişkenin değerlerini temsil etmek için noktalar kullanır. Yatay ve dikey eksenlerdeki her noktanın konumu, ayrı bir veri noktası için değerleri gösterir. Dağılım grafikleri, değişkenler arasındaki ilişkileri gözlemlemek için kullanılır

### Scatter plot;

- Biri x ekseninin değerleri için, diğeri de y eksenindeki değerler için olmak üzere aynı uzunlukta iki diziye parametre olarak alır.
- Her gözlem için bir nokta çizer.
- Bize iki değişkenin birbiri arasındaki ilişki, bu ilişkinin kuvveti ve yönü hakkında fikir verir.

\*\*\*\*\*

**colors** parametresi ile noktalar için bir renk belirleyebildiğimiz gibi, her gözlem için ayrı bir renk de verebiliyorsunuz :

```
x = np.array([5,7,8,7,2,17,2,9,4,11,12,9,6])
y = np.array([99,86,87,88,111,86,103,87,94,78,77,85,86])
colors = np.array(["red", "green", "blue", "yellow", "pink", "black",
"orange", "purple", "beige", "brown", "gray", "cyan", "magenta"])

plt.scatter(x, y, c=colors)

plt.show()
```

**Dikkat: Tek renk vermede sorun yok. Ancak her gözlem için ayrı renk verecekseniz gözlem sayısı kadar renk girmeniz gerekiyor!**

\*\*\*\*\*

## Pandas documentasyonundan:

The categorical data type is useful in the following cases:

- A string variable consisting of only a few different values. Converting such a string variable to a categorical variable will save some memory, see [here](#).
- The lexical order of a variable is not the same as the logical order (“one”, “two”, “three”). By converting to a categorical and specifying an order on the categories, sorting and min/max will use the logical order instead of the lexical order, see [here](#).
- As a signal to other Python libraries that this column should be treated as a categorical variable (e.g. to use suitable statistical methods or plot types).

\*\*\*\*\*

### Barplot:

Bir **sayısal** ve **kategorik** değişken arasındaki ilişkiyi görmek için **barplot** kullanırız.

Kategorik değişkeninin her bir entity'si bir bar (çubuk) olarak temsil ediliyor. Barın boyutu ise o kategoriye ait sayısal değeri temsil ediyor

\*\*\*\*\*

### Histogram:

Verinin olasılık dağılımını hızlıca değerlendirmek istediğimizde **histogram** kullanırız arkadaşlar. Histogram barlarının yüksekliği bize o aralığa giren değerlerin kaç kere tekrar ettiğini gösterir. Dolayısıyla **histogram, frekans dağılımlarını gösteren bir grafik** tir.

\*\*\*\*\*

### Histogram plot ile Bar plot farkı nedir.?

**Bar plot** : kategorik verilerin grafiksel gösterimidir.

Her ardışık çubuk çifti arasında eşit boşluk vardır.

Çubukların yüksekliği frekansı gösterir ve çubukların genişliği aynıdır.

**Histogram**: Nicel verilerin grafiksel gösterimidir.

Ardışık çubuklar arasında boşluk yoktur.

Dikdörtgen çubukların alanı verinin sıklığını gösterir ve çubukların genişliğinin aynı olması gerekmez.

\*\*\*\*\*

### Boxplot:

- whisker grafiği olarak da bilinir.
- minimum, 1. quartile, median, 3. quartile ve maksimum özelliklerine sahip datanın özetini görüntülemek için oluşturulur.
- quantitative (sayısal-nicel) verilerin dağılımını, kategorik bir değişkenin level'ları arasında kolayca karşılaştırma yapabilecek şekilde gösterir.
- Birinci çeyrek ile 3. çeyrek arasındaki veriler kutu içinde gösterilir. Bu kutu içinde median değeri ayrı bir çizgi olarak gösterilir.

\*\*\*\*\*

### Pie chart:

Yalnızca tek bir veri dizisini üzerinde bilgi sağlayan, sayısal oranı göstermek için sektörler olarak ayrılmış dairesel istatistiksel grafiklerdir.

Her bir dilimin alanı, datanın o parçasının yüzdesini temsil eder.

Pie chartlar datanın hızlı bir özetini verdiği için satış, operasyon, anket sonuçları, resources (kaynaklar) gibi business sunumlarında yaygın olarak kullanılır.

\*\*\*\*\*

