**Plotting with Seaborn**

**Import Libraries**

%matplotlib inline

import pandas as pd

import matplotlib.pyplot as plt

import numpy as np

import seaborn as sns

import timeit

**Import Data**

Create a dataframe called titanic

url = "https://raw.github.com/mattdelhey/kaggle-titanic/master/Data/train.csv"

titanic = pd.read\_csv(url)

Investigate titanic dataframe

titanic.info()

titanic.head()

Create few new columns

# new column "class" with word description

titanic["class"] = titanic.pclass.map({1: "First", 2: "Second", 3: "Third"})

# transform the embarking town name to actual name

titanic["embark\_town"] = titanic.embarked.map({"C": "Cherbourg", "Q": "Queenstown", "S": "Southampton"})

# transform survived from numeric value to word

titanic["alive"] = titanic.survived.map({0: "no", 1: "yes"})

# delete unnecessary columns

titanic = titanic.drop(["name", "ticket", "cabin", "embarked"], axis=1)

**Task:** show the titanic dataframe

**Task:** show the columns of titanic dataframe

**Task:** save titanic dataframe as csv file

**Question:** how much in total fare each class of passengers paid?

titanic.groupby("class")["fare"].sum()

show the total fare in bar graphs

fig, ax = plt.subplots(figsize=(15,7))

titanic.groupby("class")["fare"].sum().plot(kind='bar')

seaborn: statistical data visualization

[Seaborn](https://seaborn.pydata.org/) is a Python data visualization library based on [matplotlib](https://matplotlib.org/). It provides a high-level interface for drawing attractive and informative statistical graphics.

**Question:** how many men and women were in Titanic?

sns.factorplot(x="sex", data=titanic, kind='count')

**Question:** how many men and women were in different classes?

sns.factorplot(x="class", hue="sex", data=titanic, kind='count')

**Question:** What were the total numbers of passengers in different classes?

**Question:** What were the total numbers of passengers dead or alive?

**Question:** Where did the passengers embarked from which and on which class?

sns.factorplot(x="class", data=titanic, hue="embark\_town", kind="count")

**Question:** What classes had the survivors travelled in?

sns.factorplot("class", "survived", data=titanic)

**Question:** Which gender on which class survived most?

**Question:** Passenger from which port survived most?

**Question:** What is the age distribution of passengers?

sns.distplot(titanic.age.dropna())

sns.plt.show()

**Question:** What is the fare distribution of Titanic?

[FacetGrid](http://seaborn.pydata.org/generated/seaborn.FacetGrid.html#seaborn.FacetGrid): If we wanted to break down a plot (e.g. the last one) by some categories, we needn’t perform boolean queries, nor groupbys, we can use FacetGrid. For example age distribution of survived (0|1) passengers from different classes.

g = sns.FacetGrid(titanic, row='survived', col='class')

g.map(sns.distplot, "age")

plt.show()

**Correlation:** What is the among different attributes of the dataset?

sns.heatmap(titanic.corr(), annot=True, fmt=".2f")

plt.show()