Introduction to Big Data

- Developed by Dr. Keungoui KIM
- https://awekim.github.io/portfolio/

Lecture 5. Data Visualization

```
from google.colab import drive
drive.mount('/content/drive')
```

```
import pandas as pd
import matplotlib as plt
import seaborn as sns

tips = sns.load_dataset("tips")
flights = sns.load_dataset("flights")
titanic = sns.load_dataset("titanic")
```

```
sns.get_dataset_names()
```

Simple Exercise, Flight

```
import pandas as pd
import seaborn as sns

flights = sns.load_dataset("flights")
```

```
flights.head()
```

```
sns.scatterplot(x='year',y='passengers',data=flights)
```

```
sns.lineplot(x='year',y='passengers',data=flights)
```

```
flights_m=flights.groupby('year').mean('passengers')
flights_m.head()
```

```
# Slide
flights_m=flights.groupby('year').mean('passengers')
sns.lineplot(x='year',y='passengers',data=flights_m)
```

```
sns.lineplot(x='year',y='passengers',hue='month',data=flights)
```

```
sns.scatterplot(x='year',y='passengers',data=flights, hue="month")
```

Simple Exercise, Tips

```
import pandas as pd
import seaborn as sns

tips = sns.load_dataset("tips")
```

```
tips.head()
```

```
sns.scatterplot(x='total_bill',y='tip',data=tips)
```

```
sns.scatterplot(x='total_bill',y='tip',data=tips, hue='sex')
```

```
sns.scatterplot(x='total_bill',y='tip',data=tips, hue='time')
```

```
sns.scatterplot(x='total_bill',y='tip',data=tips, hue='smoker')
```

```
sns.scatterplot(x='total_bill',y='tip',data=tips, hue='size')
```

```
sns.histplot(x='tip',data=tips)
```

```
sns.histplot(x='tip',hue='smoker',data=tips)
```

```
sns.histplot(x='total_bill',y='tip',data=tips)
```

```
sns.boxplot(x="day", y="tip", data=tips)
```

```
sns.boxplot(x="day", y="tip", data=tips, hue = "sex")
```

```
sns.boxplot(x="day", y="tip", data=tips, hue = "smoker")
```

```
sns.boxplot(x="day", y="tip", data=tips, hue = "sex")
```

```
sns.pairplot(data=tips)
```

```
sns.pairplot(data=tips, hue='time', diag_kind='hist')
```

```
sns.jointplot(x='total_bill',y='tip', data=tips)
```

```
sns.jointplot(x='total_bill',y='tip', hue='time', data=tips, kind='hist')
```