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)
```

```
# Slide
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')

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

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

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

titanic.head()

# Slide
sns.barplot(x='class',y='survived',
            data=titanic)

# Slide
sns.barplot(x='class',y='survived',
            hue='sex',data=titanic,
            dodge=True)

# Slide
sns.barplot(x='survived',y='class', data=titanic)

# Slide
sns.barplot(x='survived',y='class',hue='sex',data=titanic, dodge=True)

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

# Slide
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")

# Slide
sns.pairplot(data=tips)

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

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

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

Review

1. What is the unit of this data set?
2. What is the annual average income of region?

3. Visualize your finding

```
import pandas as pd
import seaborn as sns

korea = pd.read_csv("/content/drive/MyDrive/[Lecture]/IntBigData/BigData_Python/05_Data
korea.head()

# 1. What is the unit of this data set?
korea['id'].nunique()

korea.shape

print(korea.year.nunique(), korea.wave.nunique())

# 2. What is the annual average income of region?
korea_yrreg = korea.groupby(['year','region'])
korea_yrreg

korea_yrreg_income = korea_yrreg['income'].mean().reset_index()
korea_yrreg_income.head()

# 3. Visualize your finding
sns.lineplot(x='year', y='income', hue='region', data=korea_yrreg_income)
```