University of Essex

Online

Plotting with Seaborn

Seaborn is a Python library for making statistical visualizations. It's built to provide eye candy plots and at the same time it makes developers' life easier. We've all been in the situation where we needed to build a simple decent looking histogram and ended up making several function invocations and setting arguments without fully knowing what we're doing.

Seaborn tackles this not by reinventing the wheel but by improving it. It is built on top of Matplotlib and provides a high level API that makes "a well-defined set of hard things easy" (as stated in the docs), amongst other things by making that its methods work greatly by passing a minimal set of arguments.

Import Libraries

```
%matplotlib inline
import pandas as pd
import matplotlib.pyplot as plt
import numpy as np
import seaborn as sns
```

Import Data

Create a dataframe called `titanic` and read the `titanic_modified.csv`

```
titanic = pd.read_csv("path//to// titanic_modified.csv")
```

Investigate `titanic` dataframe

```
titanic.info()

titanic.head()
```

Create few new columns

```
# delete unnecessary columns
titanic = titanic.drop(["name", "ticket", "cabin"], axis=1)
```

Question: how many men and women were in Titanic?

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

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

```
sns.factorplot("class", data=titanic, hue="sex", 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 embark from?

```
sns.factorplot("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: 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: Show correlation within different attributes of the dataset?

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