

Data Science Foundation

Lesson #8 - Storytelling from Seaborn

Ivanovitch Silva
September, 2017



Agenda

- Case study: kaggle competition
- Motivation: seaborn
- Histogram, KDE
- Modifying the appearance of plots
- Conditional distributions

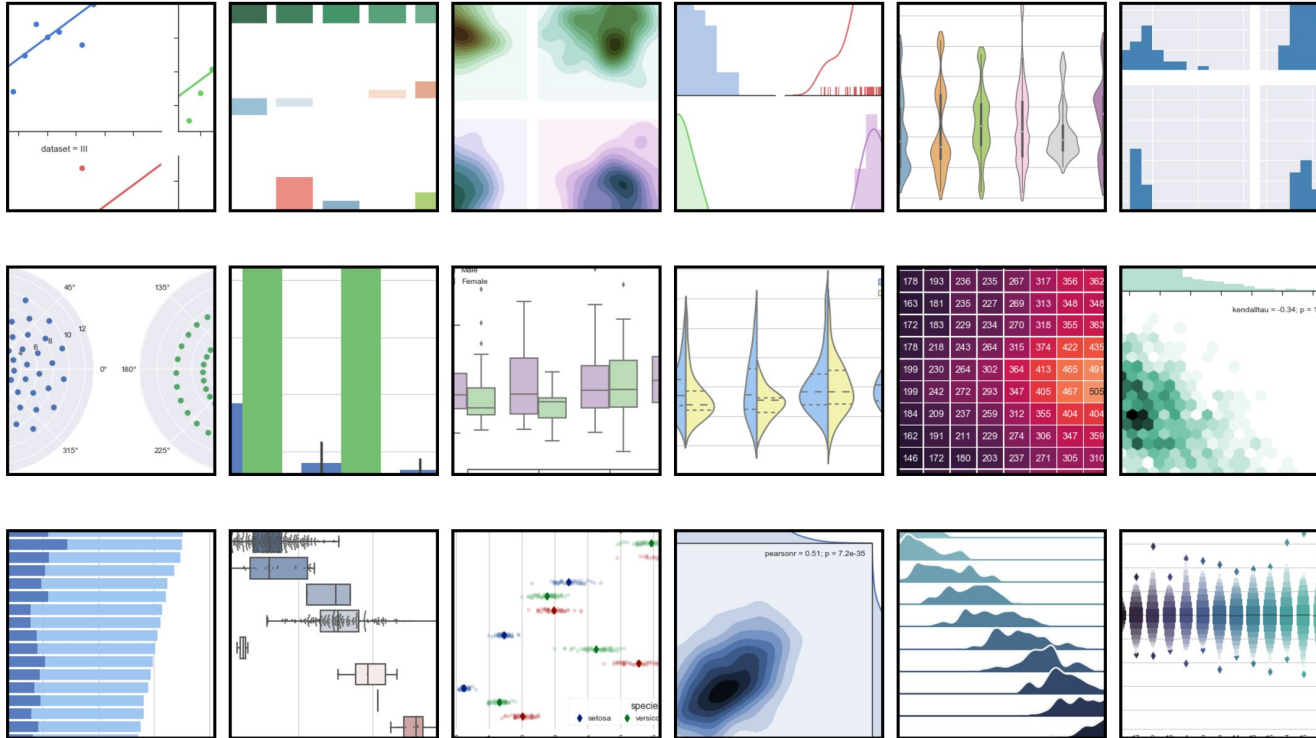
Update the repository

```
git clone https://github.com/ivanovitchm/EEC2006.git
```

Or

```
git pull
```

Motivation - Seaborn



Getting started

```
conda install -c conda-forge seaborn
```

Introduction to dataset



Getting Started Prediction Competition

Titanic: Machine Learning from Disaster

Start here! Predict survival on the Titanic and get familiar with ML basics



Kaggle · 8,380 teams · 3 years to go

<https://www.kaggle.com/c/titanic/data>

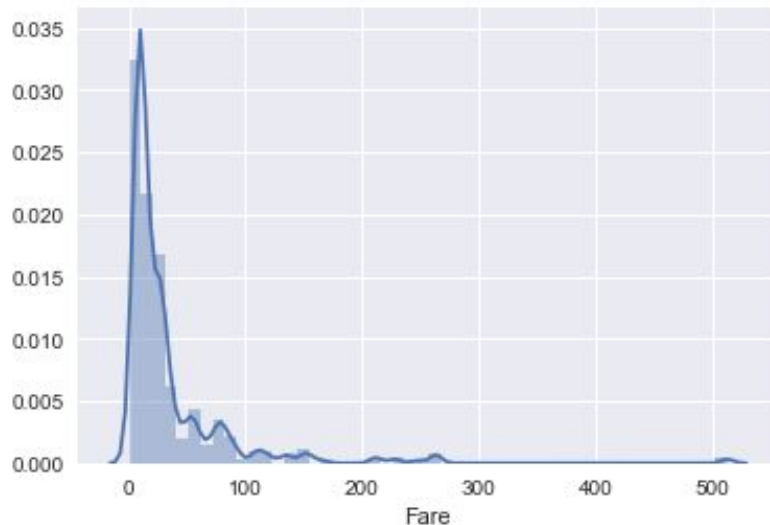
Introduction to dataset

PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin	Embarked
1	0	3	Braund, Mr. Owen Harris	male	22.0	1	0	A/5 21171	7.2500		S
2	1	1	Cumings, Mrs. John Bradley (Florence Briggs Thayer)	female	38.0	1	0	PC 17599	71.2833	C85	C
3	1	3	Heikkinen, Miss. Laina	female	26.0	0	0	STON/O2. 3101282	7.9250		S

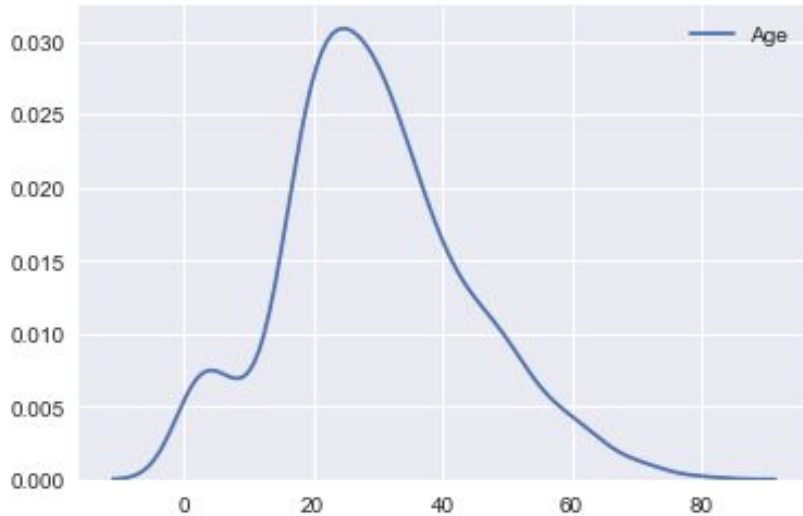
- sibsp: The dataset defines family relations in this way...
 - Sibling = brother, sister, stepbrother, stepsister
 - Spouse = husband, wife (mistresses and fiancés were ignored)
- parch: The dataset defines family relations in this way...
 - Parent = mother, father
 - Child = daughter, son, stepdaughter, stepson
 - Some children travelled only with a nanny, therefore parch=0 for them.

Creating histogram in seaborn

```
# seaborn is commonly imported as `sns`.  
import matplotlib.pyplot as plt  
import seaborn as sns  
  
#to switch to seaborn defaults, simply call the set()  
sns.set()  
  
# The four preset contexts, in order of relative size,  
sns.set_context("notebook")  
  
# plot a univariate distribution of observations.  
sns.distplot(titanic["Fare"])  
plt.show()
```

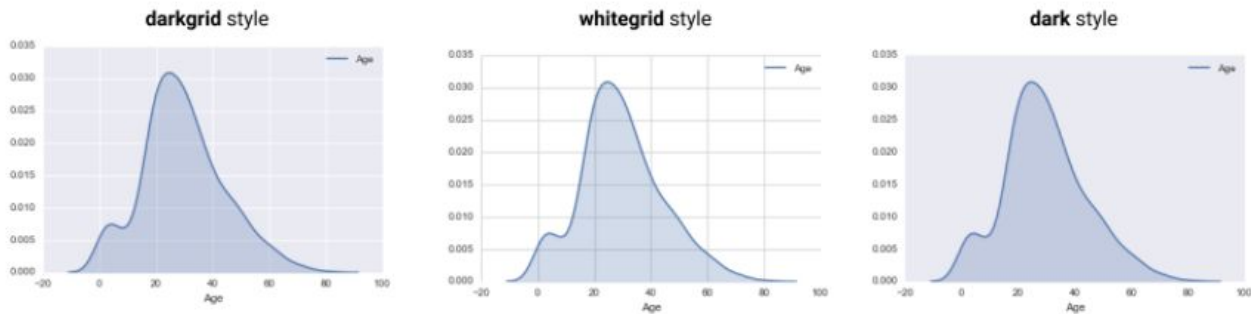


Generating a kernel density plot



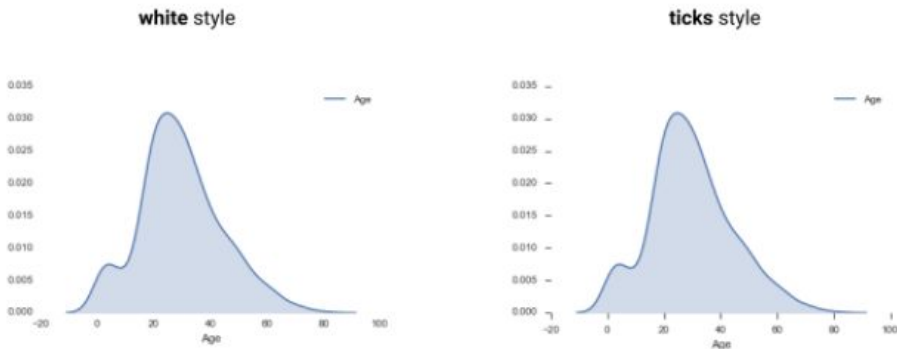
```
sns.kdeplot(titanic["Age"])  
plt.show()
```

Modifying the appearance of plots

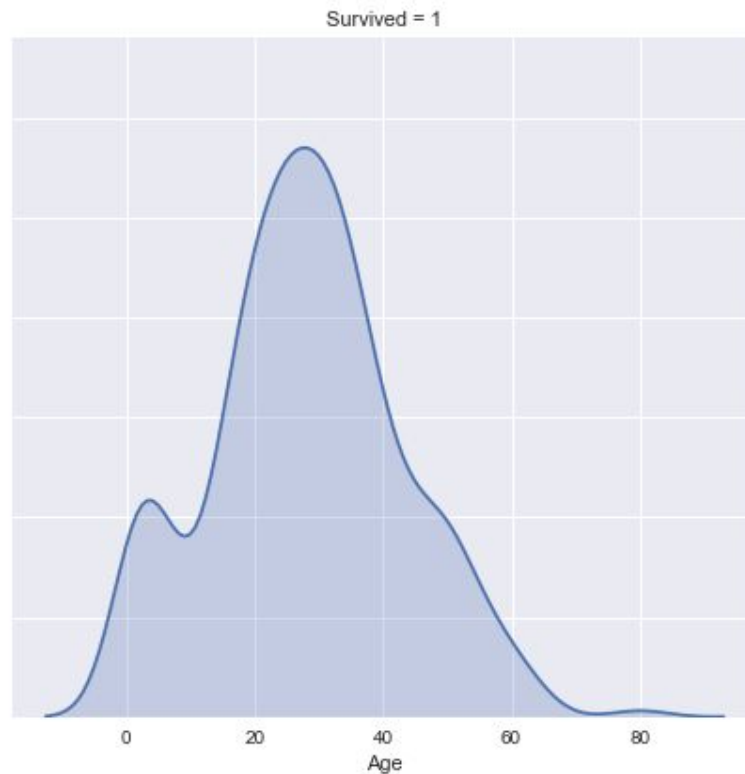
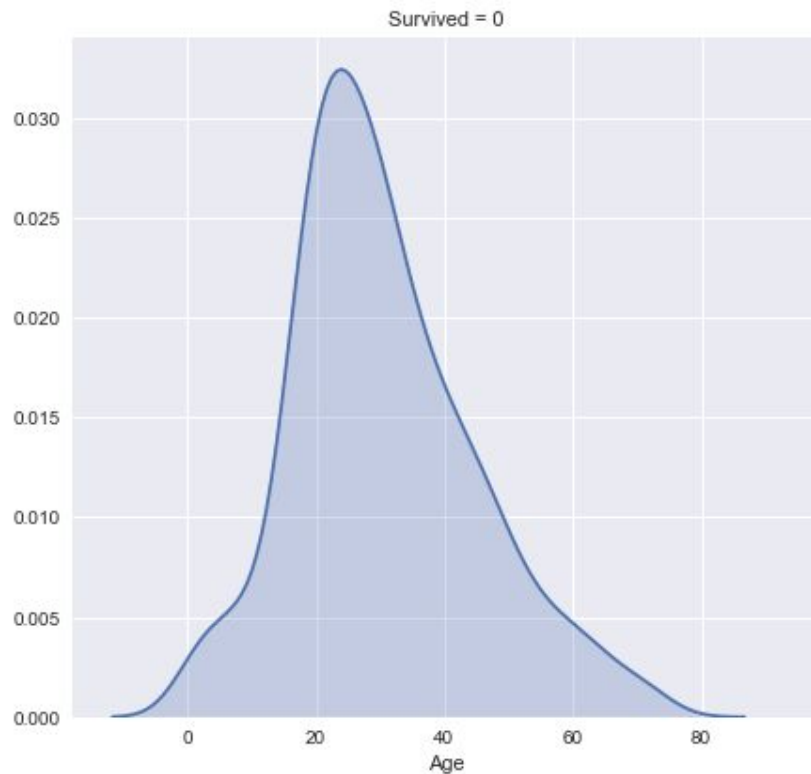


```
sns.set_style("darkgrid")
```

```
sns.despine()
```



Conditional distributions



Conditional distributions

```
# Condition on unique values of the "Survived" column.
```

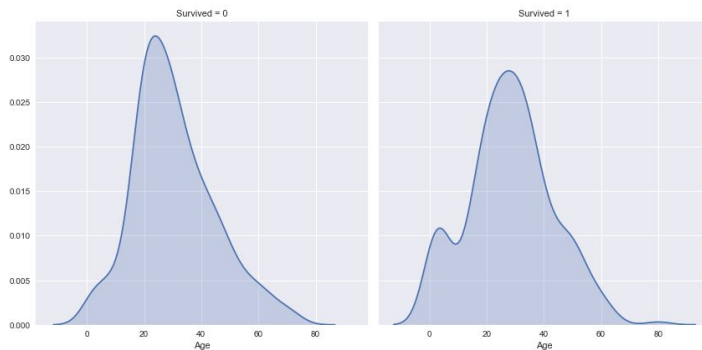
```
g = sns.FacetGrid(titanic, col="Survived", size=6)
```

```
# For each subset of values, generate a kernel density plot of the "Age" columns.
```

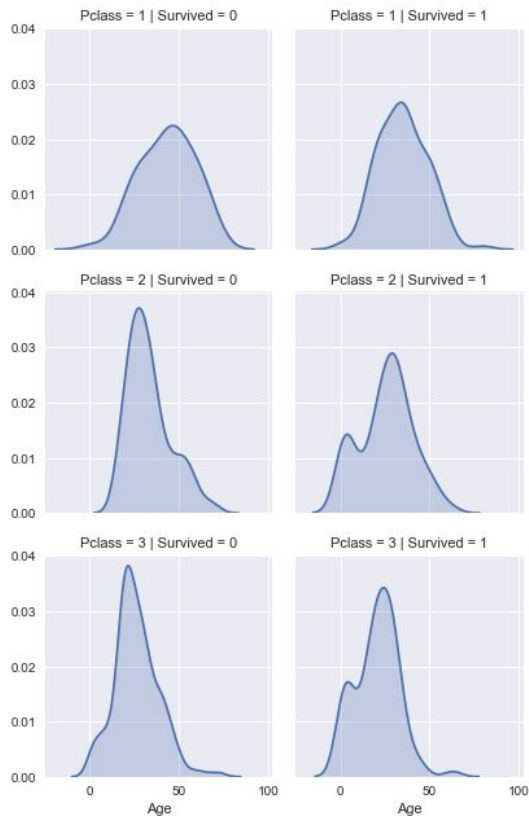
```
g.map(sns.kdeplot, "Age", shade=True)
```

```
# Plot the graph
```

```
plt.show()
```



Creating conditional plots using two conditions



```
g = sns.FacetGrid(titanic, col="Survived", row="Pclass")
g.map(sns.kdeplot, "Age", shade=True)
sns.despine(left=True, bottom=True)
plt.show()
```

