
APSSDC
Andhra Pradesh State Skill Development Corporation
Skill AP
APSSDC
Data Visualization using Seaborn

An Introduction to Seaborn

Seaborn is a library for making statistical graphics in Python. It is built on top of matplotlib and closely integrated with pandas data structures.

Here is some of the functionality that seaborn offers:

- A dataset-oriented API for examining relationships between multiple variables.
- Specialized support for using categorical variables to show observations or aggregate statistics.
- Convenient views onto the overall structure of complex datasets.
- Tools for choosing color palettes that faithfully reveal patterns in your data.

Seaborn Vs Matplotlib

It is summarized that if Matplotlib “tries to make easy things easy and hard things possible”, Seaborn tries to make a well-defined set of hard things easy too.” Seaborn helps resolve the two major problem s faced by Matplotlib; the problems are:

- Default Matplotlib parameters
- Working with data frames As Seaborn compliments and extends Matplotlib, the learning curve is quite gradual. If you know Matplotlib, you are already half way through Seaborn.

Installing and getting started

To install the latest release of seaborn, you can use pip:

```
--- pip install seaborn
```

It’s also possible to install the released version using conda:

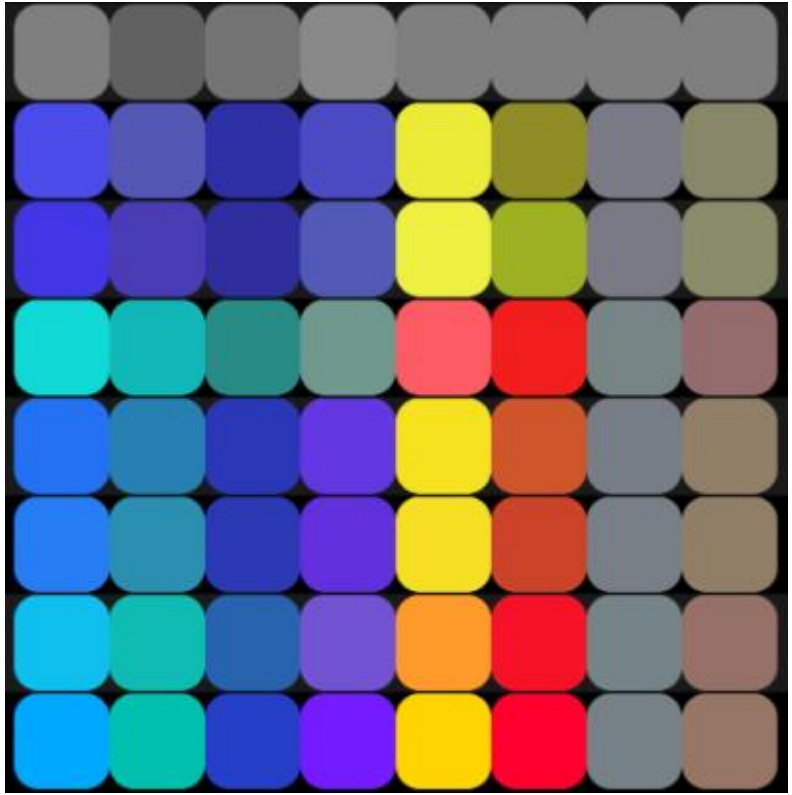
```
--- conda install seaborn
```

To use the seaborn package we need to import it:

```
--- import seaborn
```

Color Palettes

Color is more important than other aspects of figure style because color can reveal patterns in the data if used effectively or hide those patterns if used poorly. There are a number of great resources to learn about good techniques for using color in visualizations



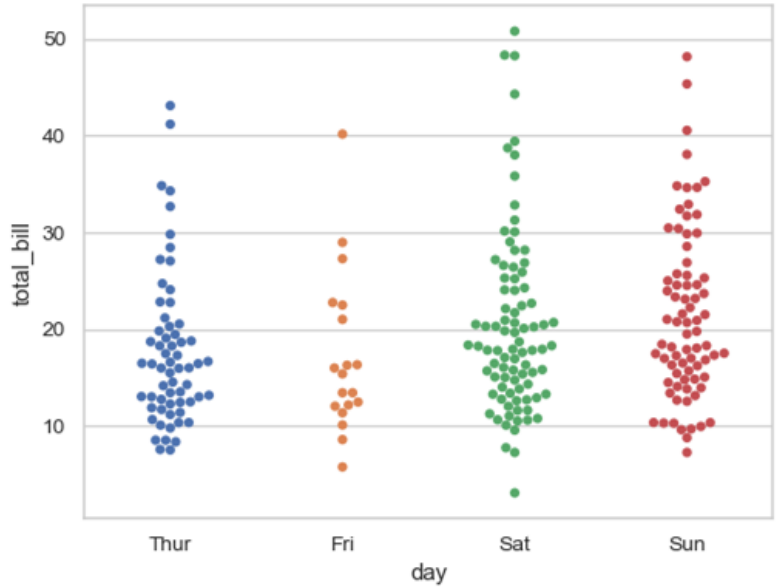
Types of Plots:

The various types of plots available in seaborn are:

1. Swarm plot
2. Box plot
3. Violin plot
4. Bar plot
5. Count plot
6. Regression plot
7. Heat map

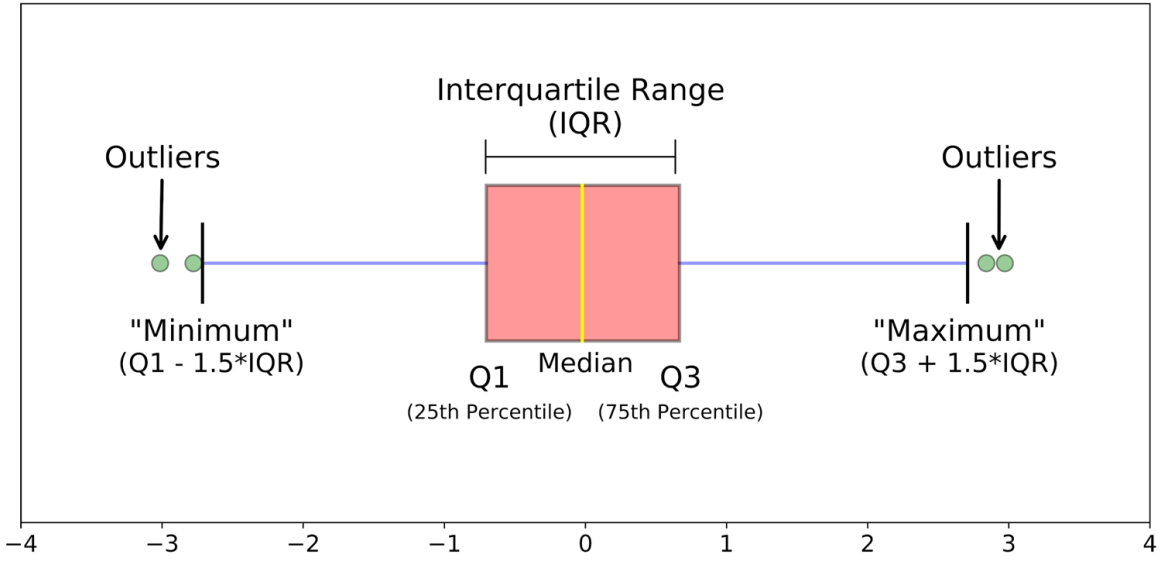
Swarm plot

Seaborn’s strip plot and swarn plot is virtually identical except that it prevents data points from overlapping.



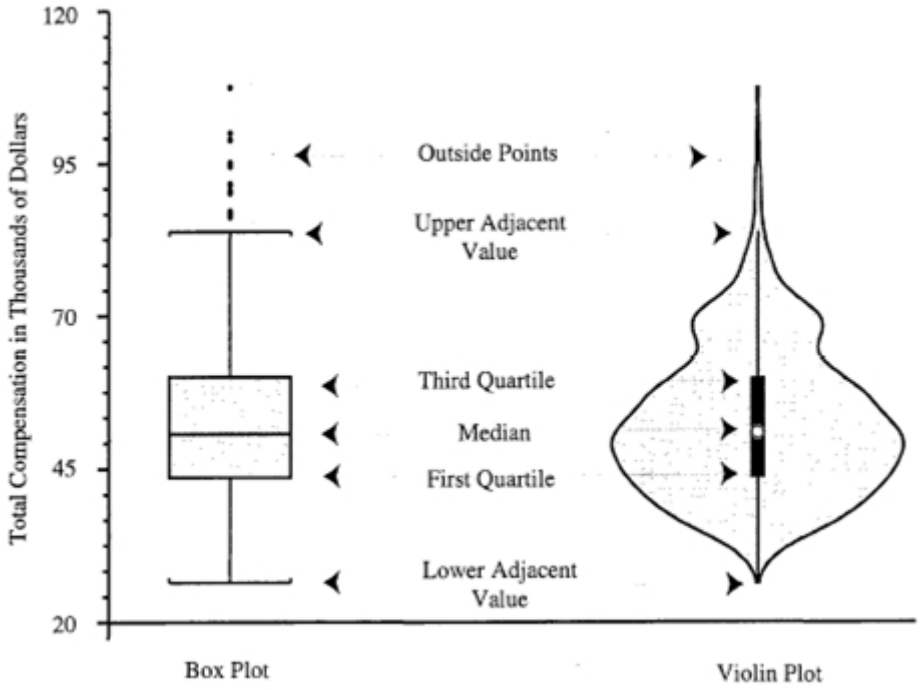
Box plot

Boxplots summarize numeric data over a set of categories. The data is divided into four groups called quartiles.



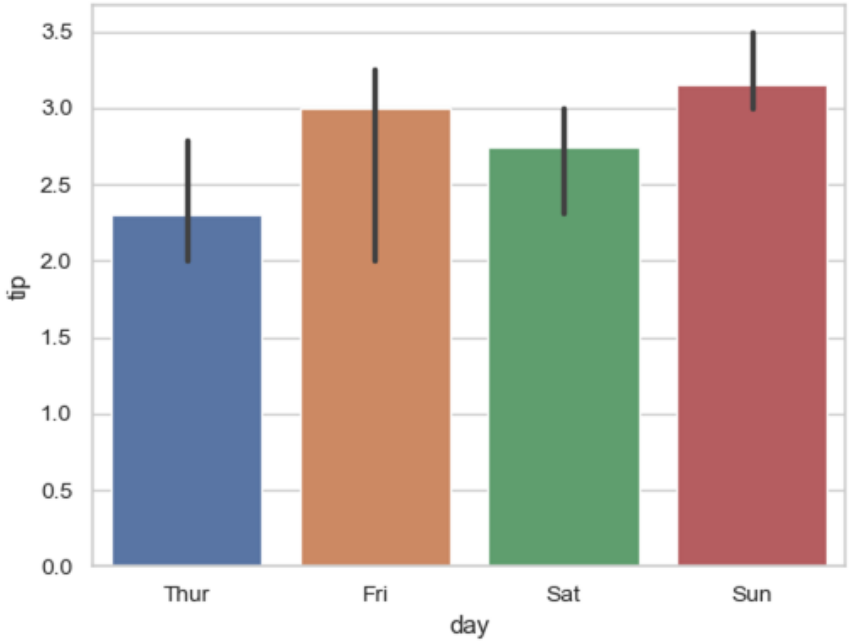
Violin plot

Violin plots summarize numeric data over a set of categories. They are essentially a box plot with a kernel density estimate (KDE) overlaid along the range of the box and reflected to make it look nice. They provide more information than a box plot because they also include information about how the data is distributed within the inner quartiles.



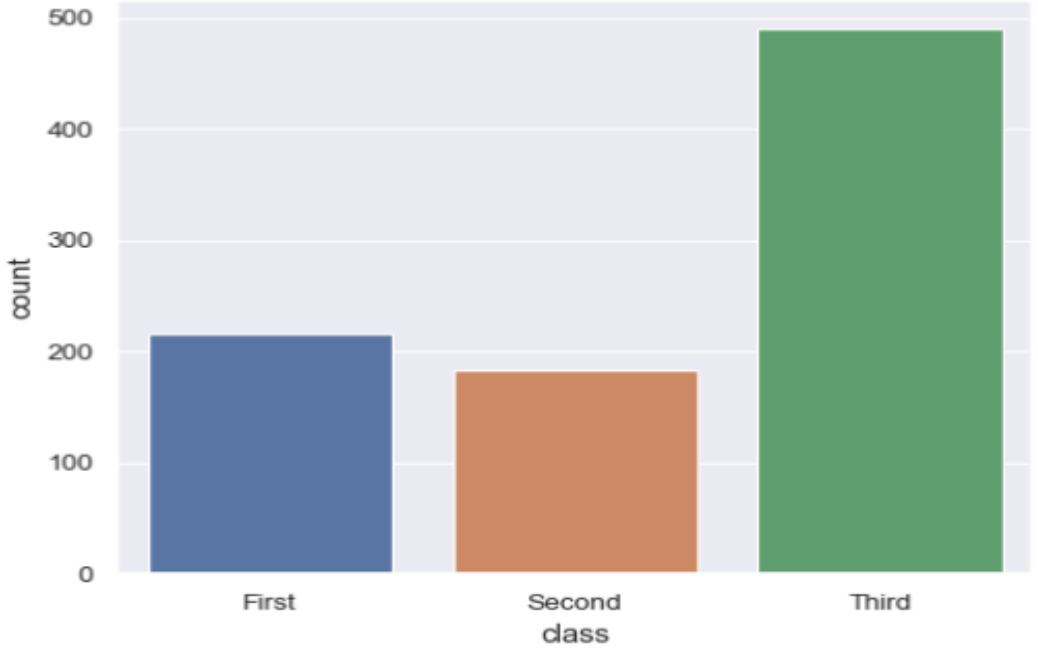
Bar plot

Bar graphs are useful for displaying relationships between categorical data and at least one numerical variable.



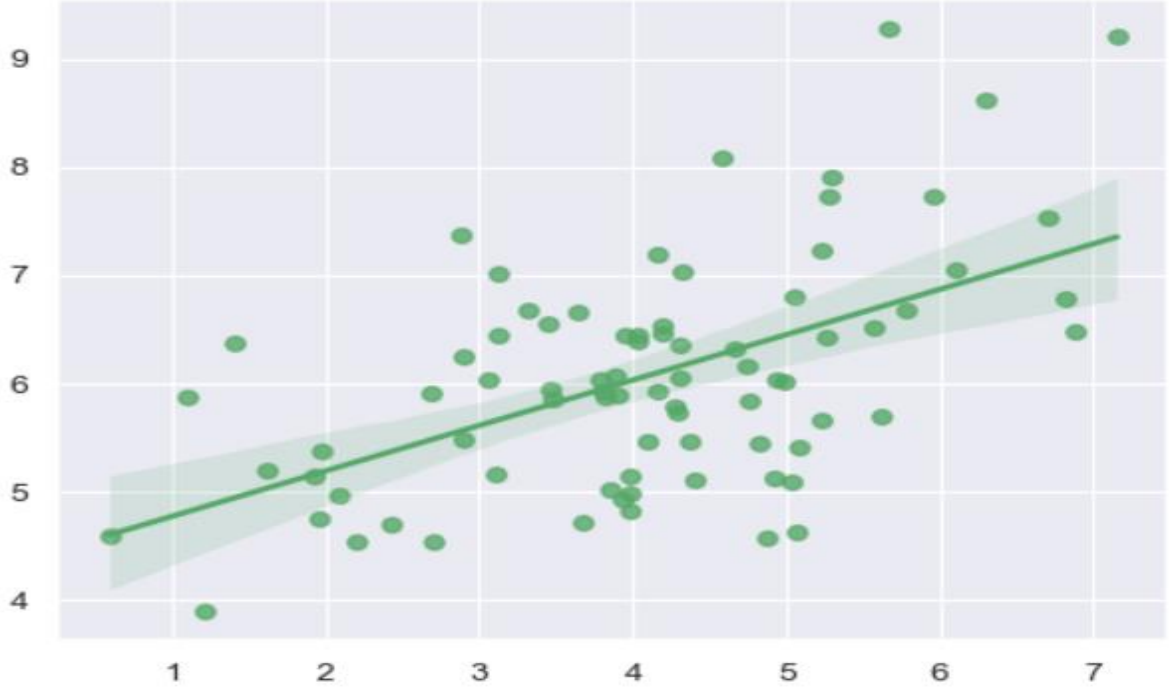
Count plot

A count plot can be thought of as a histogram across a categorical, instead of quantitative, variable. The basic API and options are identical to those for barplot(), so you can compare counts across nested variables.



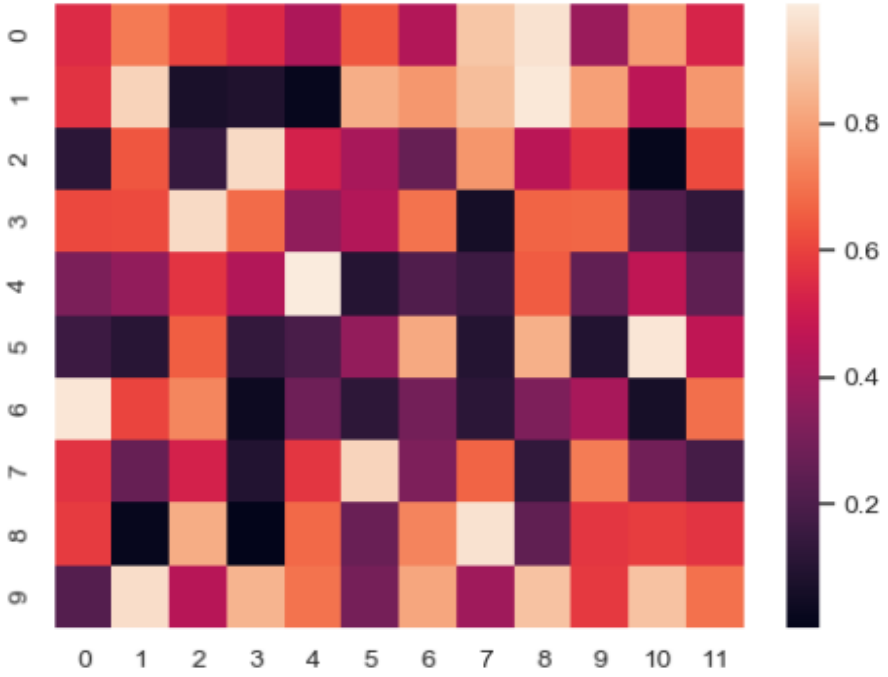
Regression plot

Regplot performs a simple linear regression model fit and plot. lmplot() combines regplot() and FacetGrid.



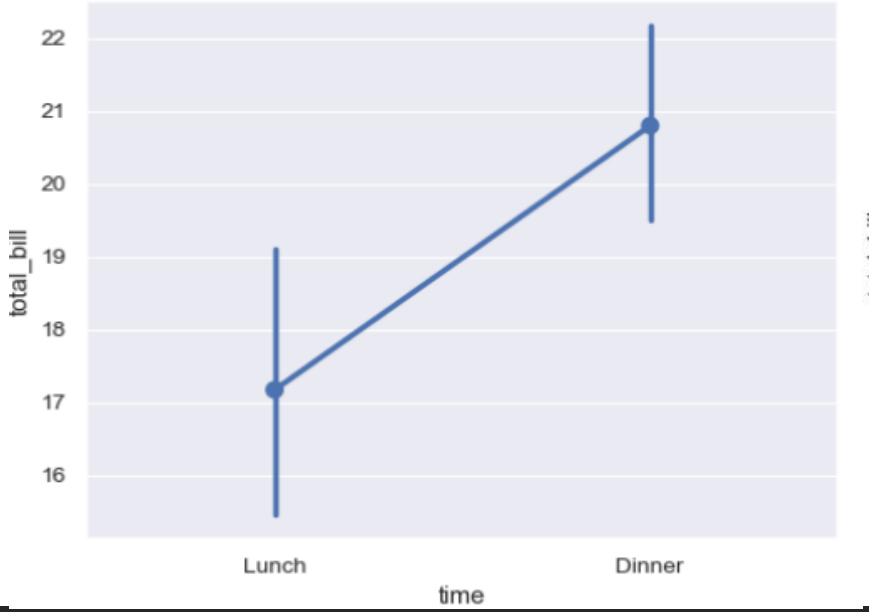
Heat map

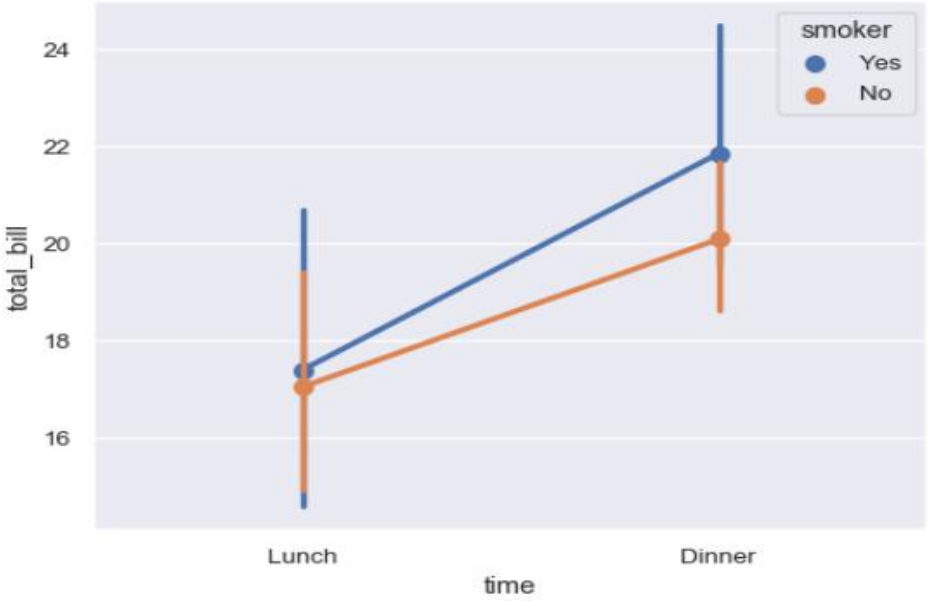
Plot rectangular data as a color-encoded matrix.



Point plot

Point plots can be more useful than bar plots for focusing comparisons between different levels of one or more categorical variables.





APSSDC

Andhra Pradesh State Skill Development Corporation

Print("THANK YOU")

Slides created, maintained and distributed by

Andhra Pradesh State Skill Development Corporation