Data Science

Intro to Seaborn

CS202 10.05.22

Seaborn

Matrix plots

We will begin by looking at matrix plots in seaboard.

Matrix plots allow you to plot data as color-encoded matrices and can also be used to indicate clusters within the data (later in the machine learning section we will learn how to formally cluster data).

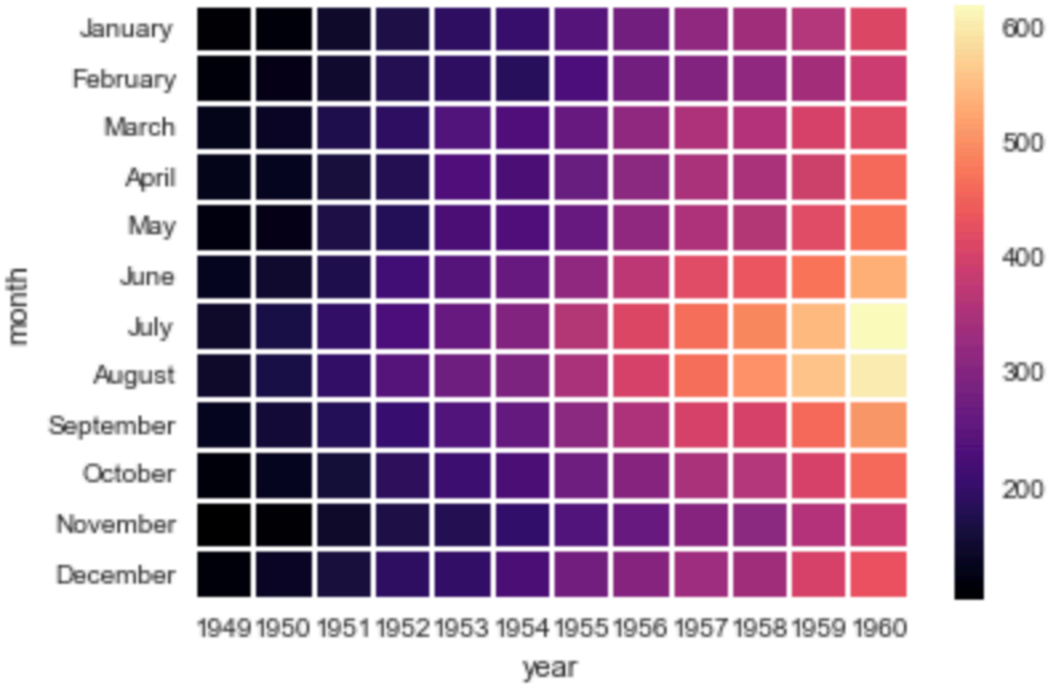
We will be looking at:

- heat maps
-cluster maps

Seaborn

Matrix plots

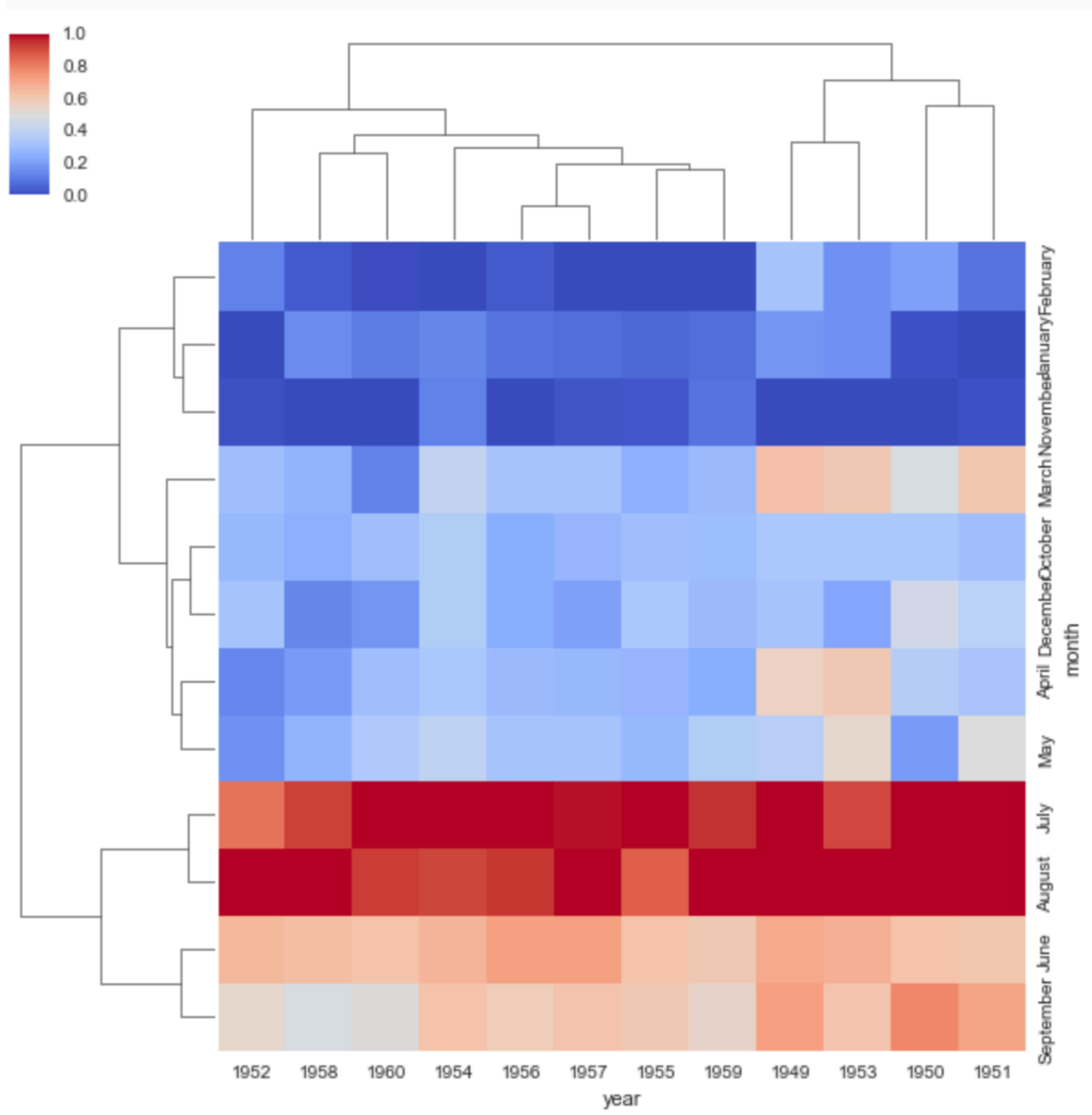
- Heat map
- Heatmap is a way to show some sort of matrix plot.
- In order for a heatmap to work properly, your data should already be in a matrix form.
- The `sns.heatmap()` function basically just colors it in for you.



Seaborn

Matrix plots

- Cluster map
- The cluster map is a clustered heat map.
- The clustermap uses hierarchal clustering to produce a clustered version of the heatmap.
- The heatmap cells are all clustered using a similarity algorithm. Dentograms are drawn for the columns and the rows of the heatmap.
- The `sns.clustermap()` function in seaborn draws a hierarchically clustered heatmap.



Seaborn

Grids

Grids are general types of plots that allow you to map plot types to rows and columns of a grid, this helps you create similar plots separated by features.

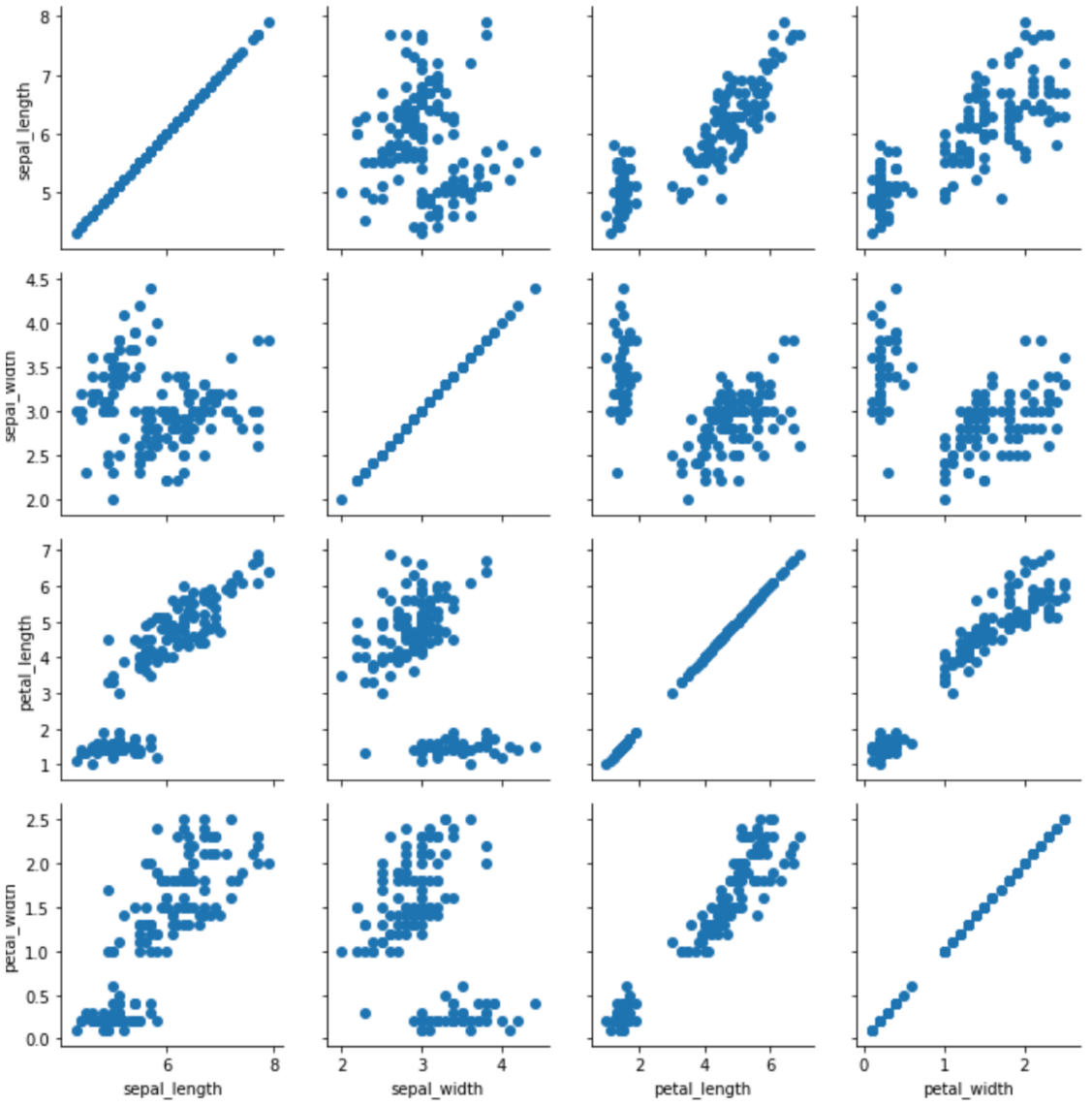
We will be looking at the following grids:

- Pair Grid
- pairplot
- Facet Grid
- Joint Grid

Seaborn

Grids

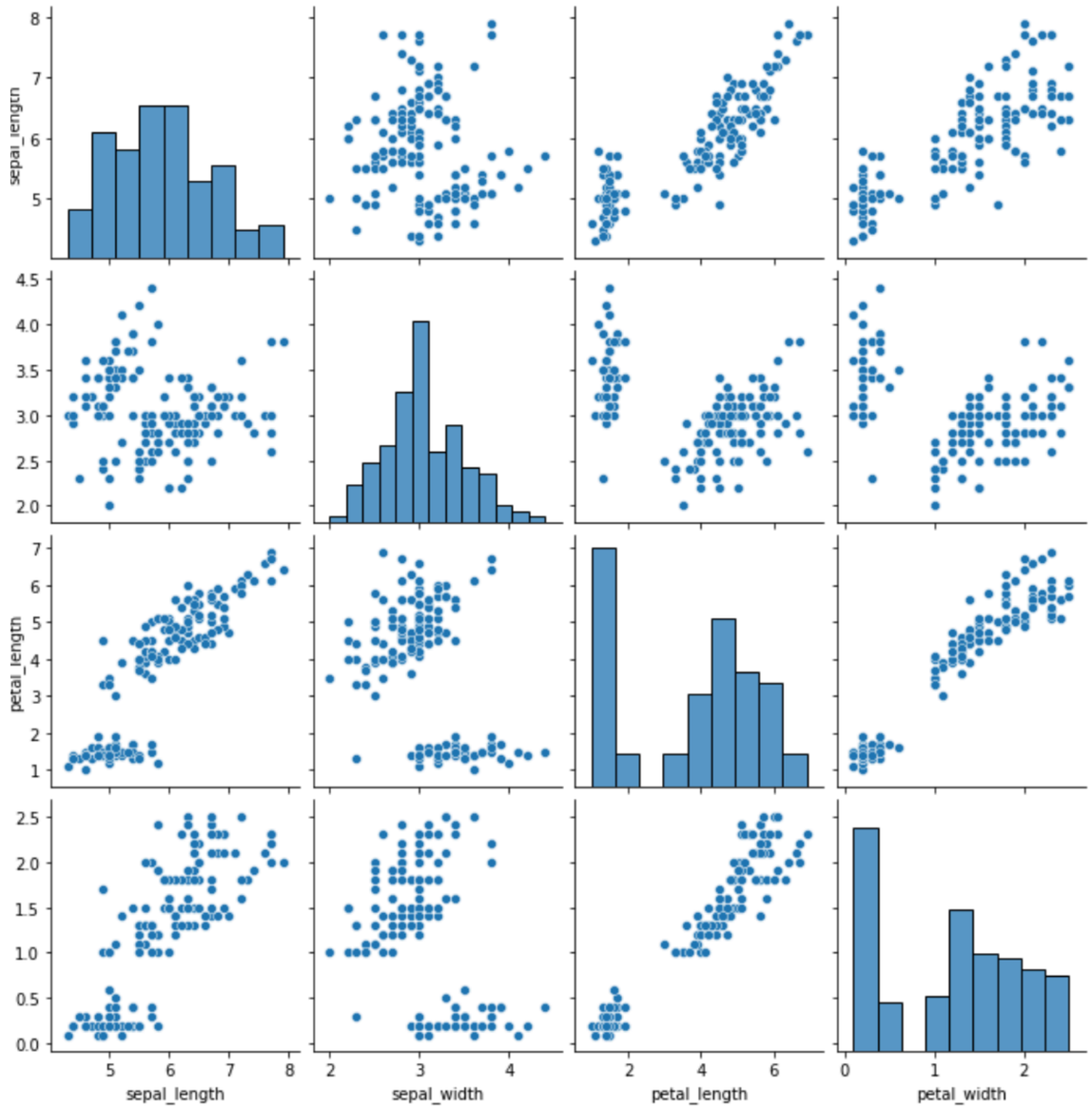
- Pair Grid
- A pair grid is a subplot grid for plotting pairwise relationships in a dataset.
- This object maps each variable in a dataset onto a column and row in a grid of multiple axes. Different axes-level plotting functions can be used to draw bivariate plots in the upper and lower triangles, and the marginal distribution of each variable can be shown on the diagonal.
- The `sns.PairGrid()` function lets you create pair grids, you may use this for more flexibility of graphs you want to plot otherwise you can use `sns.pairplot()`
- Note: the image on the right needs specification with the pair grid, the pair grid alone will just give you a grid of empty graphs.



Seaborn

Grids

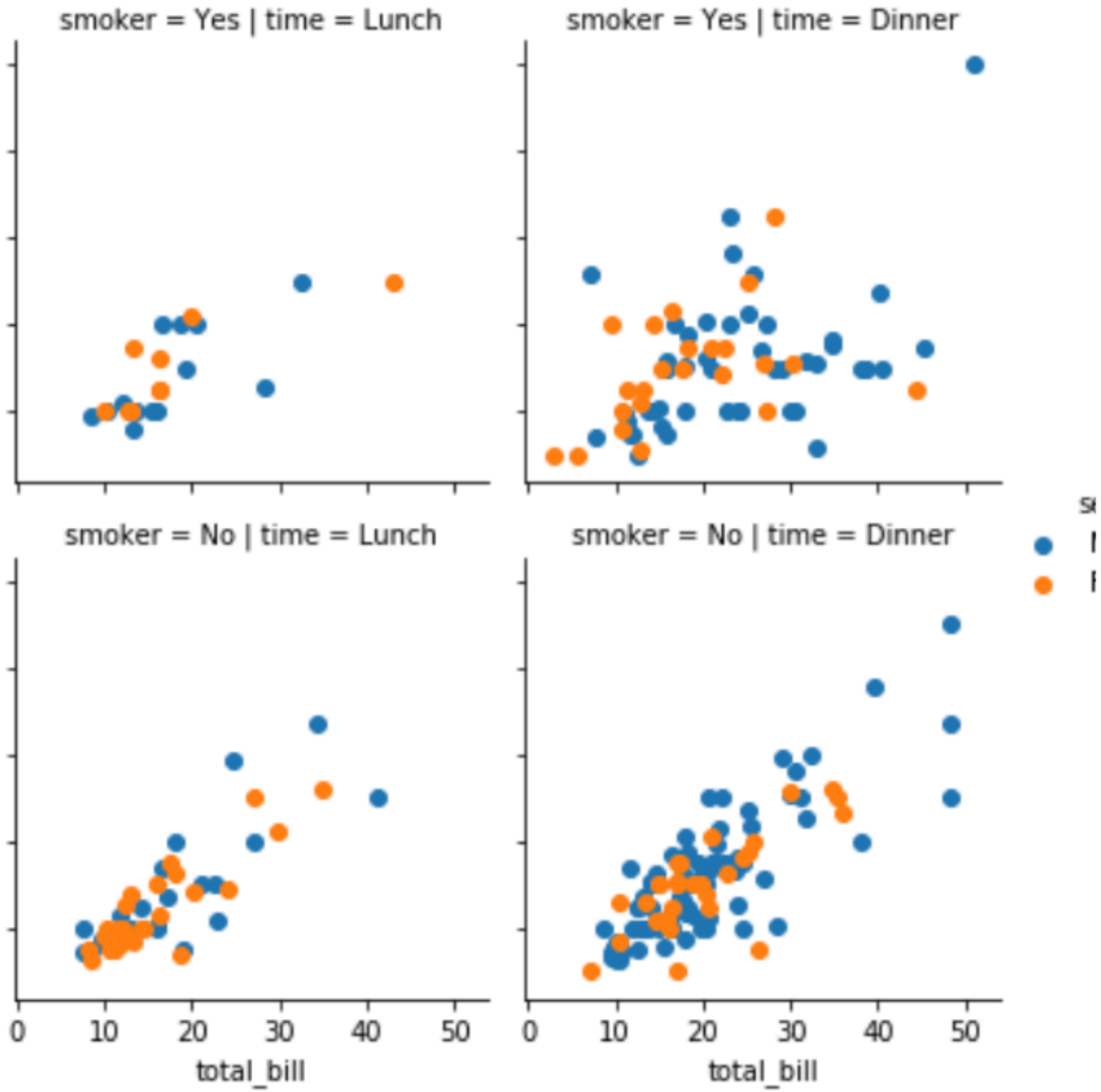
- Pairplot
- A pairplot plots pairwise relationships in a dataset.
- By default, this function will create a grid of Axes such that each numeric variable in data will by shared across the y-axes across a single row and the x-axes across a single column. The diagonal plots are treated differently: a univariate distribution plot is drawn to show the marginal distribution of the data in each column.
- This is intended to make it easy to draw a few common graphs.
- The sns.pairplot() function will generate such a grid of graphs.



Seaborn

Grids

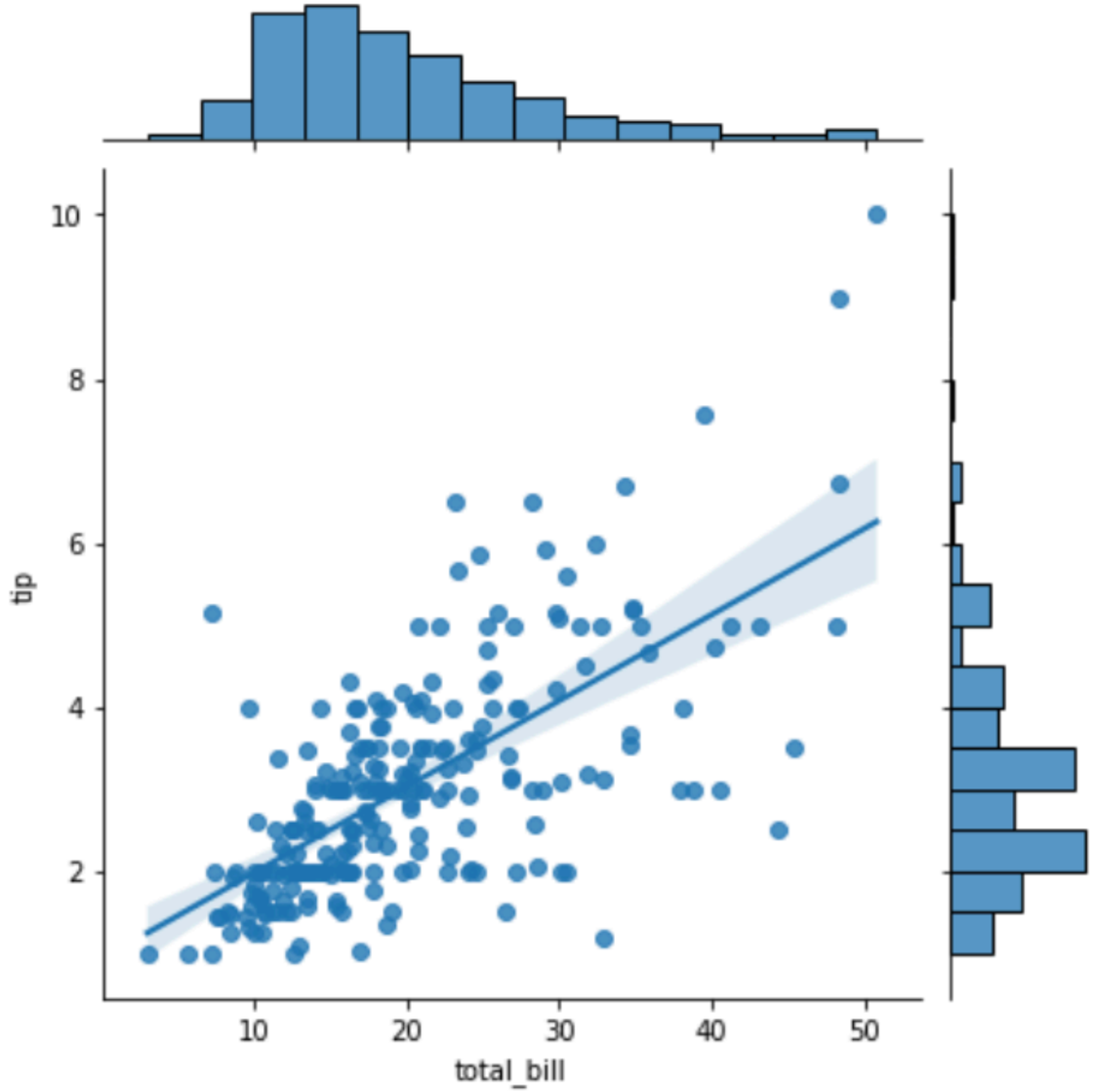
- Facet Grid
- FacetGrid is the general way to create grids of plots based off of a feature.
- Just like the pair grid you must enter the graphs you would like your facet grid to plot
- The `sns.FacetGrid()` function allows you to make a facet grid.
- Parameters: must enter column and row variables: `sns.FacetGrid(df, col = “col1”, row = “col2”)`



Seaborn

Grids

- Joint Grid
- JointGrid is the general version for jointplot() type grids. Joint Grid is a grid for drawing a bivariate plot with marginal univariate plots.
- The `sns.JointGrid()` allows you to plot a joint grid.



Seaborn

Regression Plots

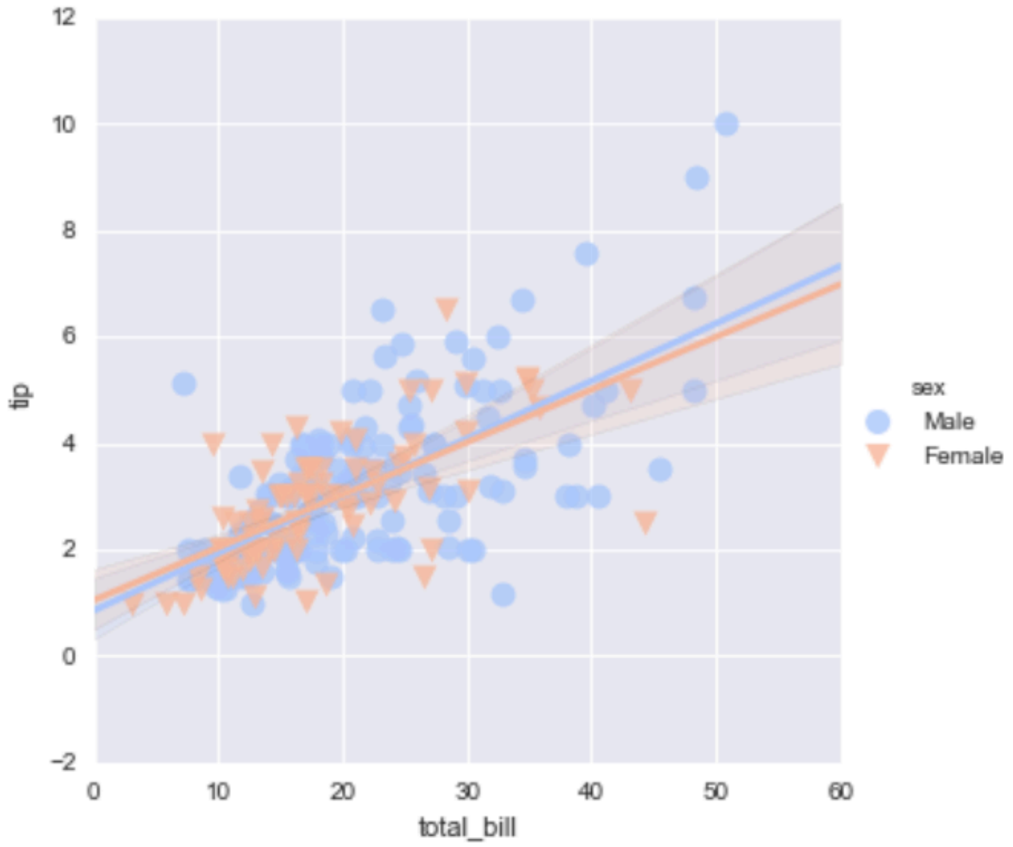
Seaborn has many built-in capabilities for regression plots, however we won't really discuss regression until we begin machine learning in the course, so we will only cover the `lmplot()` function for now.

lmplot allows you to display linear models, but it also conveniently allows you to split up those plots based off of features, as well as coloring the hue based off of features.

Seaborn

Regression plots

- Lmplot
- The lmplot plots data and regression model fits across a FacetGrid.
- This function combines `regplot()` and FacetGrid. It is intended as a convenient interface to fit regression models across conditional subsets of a dataset.
- The sns.lmplot() function is used to plot this regression line.



Seaborn

Styling

There are different ways you can style your plots.

We will see them via demo.

But please read documentations for each plot for practice.

Documentation for each package should act as the textbook that you follow for the course.