

Seaborn Workshop

Plotting on data-aware grids

- ▶ When exploring medium-dimensional data, a useful approach is to draw multiple instances of the same plot on different subsets of your dataset.
- ▶ This technique is sometimes called either lattice, or trellis plotting, and it is related to the idea of small multiples. It allows a viewer to quickly extract a large amount of information about complex data.
- ▶ Matplotlib offers good support for making figures with multiple axes; seaborn builds on top of this to directly link the structure of the plot to the structure of your dataset.

Seaborn Workshop

To use these features, your data has to be in a Pandas DataFrame and it must take the form of what Hadley Wickam calls tidy data. In brief, that means your dataframe should be structured such that each column is a variable and each row is an observation.

Seaborn Workshop

- ▶ For advanced use, you can use the objects discussed in this part of the tutorial directly, which will provide maximum flexibility. Some seaborn functions (such as `Implot()`, `factorplot()`, and `pairplot()`) also use them behind the scenes.
- ▶ Unlike other seaborn functions that are Axes-level and draw onto specific (possibly already-existing) matplotlib Axes without otherwise manipulating the figure, these higher-level functions create a figure when called and are generally more strict about how it gets set up.

Seaborn Workshop

- ▶ In some cases, arguments either to those functions or to the constructor of the class they rely on will provide a different interface attributes like the figure size, as in the case of `Implot()` where you can set the height and aspect ratio for each facet rather than the overall size of the figure.
- ▶ Any function that uses one of these objects will always return it after plotting, though, and most of these objects have convenience methods for changing how the plot, often in a more abstract and easy way.

Seaborn Workshop

```
%matplotlib inline
import numpy as np
import pandas as pd
import seaborn as sns
from scipy import stats
import matplotlib as mpl
import matplotlib.pyplot as plt
sns.set(style="ticks")
np.random.seed(sum(map(ord, "axis_grids")))
```

Subsetting data with FacetGrid

- ▶ The FacetGrid class is useful when you want to visualize the distribution of a variable or the relationship between multiple variables separately within subsets of your dataset.
- ▶ A FacetGrid can be drawn with up to three dimensions: row, col, and hue.
- ▶ The first two have obvious correspondence with the resulting array of axes; think of the hue variable as a third dimension along a depth axis, where different levels are plotted with different colors.

Seaborn Workshop

The class is used by initializing a `FacetGrid` object with a dataframe and the names of the variables that will form the row, column, or hue dimensions of the grid. These variables should be categorical or discrete, and then the data at each level of the variable will be used for a facet along that axis. For example, say we wanted to examine differences between lunch and dinner in the tips dataset.

Seaborn Workshop

Additionally, both `lmplot()` and `factorplot()` use this object internally, and they return the object when they are finished so that it can be used for further tweaking.

```
tips = sns.load_dataset("tips")  
g = sns.FacetGrid(tips, col="time")  
../_images/axis_grids_8_0.png
```

Initializing the grid like this sets up the matplotlib figure and axes, but doesn't draw anything on them.

Seaborn Workshop

The main approach for visualizing data on this grid is with the `FacetGrid.map()` method. Provide it with a plotting function and the name(s) of variable(s) in the dataframe to plot. Lets look at the distribution of tips in each of these subsets, using a histogram.

```
g = sns.FacetGrid(tips, col="time")  
g.map(plt.hist, "tip");  
../_images/axis_grids_10_0.png
```

Seaborn Workshop

This function will draw the figure and annotate the axes, hopefully producing a finished plot in one step. To make a relational plot, just pass multiple variable names. You can also provide keyword arguments, which will be passed to the plotting function:

```
g = sns.FacetGrid(tips, col="sex", hue="smoker")
g.map(plt.scatter, "total_bill", "tip", alpha=.7)
g.add_legend();
../_images/axis_grids_12_0.png
```

Seaborn Workshop

There are several options for controlling the look of the grid that can be passed to the class constructor.

```
g = sns.FacetGrid(tips, row="smoker", col="time", margin_titles=True)
g.map(sns.regplot, "size", "total_bill", color=".3", fill=False)
sns.savefig("../_images/axis_grids_14_0.png")
```

Note that `margin_titles` isn't formally supported by the matplotlib API, and may not work well in all cases. In particular, it currently can't be used with a legend that lies outside of the plot.

Seaborn Workshop

The size of the figure is set by providing the height of each facet, along with the aspect ratio:

```
g = sns.FacetGrid(tips, col="day", size=4, aspect=.5)
g.map(sns.barplot, "sex", "total_bill");
../_images/axis_grids_16_0.png
```

Seaborn Workshop

With versions of matplotlib \geq 1.4, you can pass parameters to be used in the gridspec module. This can be used to draw attention to a particular facet by increasing its size. It's particularly useful when visualizing distributions of datasets with unequal numbers of groups in each facet.

```
titanic = sns.load_dataset("titanic")
titanic = titanic.assign(deck=titanic.deck.astype(object))
g = sns.FacetGrid(titanic, col="class", sharex=False,
                  gridspec_kws={"width_ratios": [5, 3, 1]})
g.map(sns.boxplot, "deck", "age");
```

../_images/axis_grids_18_0.png

Seaborn Workshop

The default ordering of the facets is derived from the information in the DataFrame. If the variable used to define facets has a categorical type, then the order of the categories is used. Otherwise, the facets will be in the order of appearance of the category levels. It is possible, however, to specify an ordering of any facet dimension with the appropriate `*_order` parameter:

```
ordered_days = tips.day.value_counts().index
g = sns.FacetGrid(tips, row="day", row_order=ordered_days,
                  size=1.7, aspect=4,)
g.map(sns.distplot, "total_bill", hist=False, rug=True)
../_images/axis_grids_20_0.png
```

Seaborn Workshop

- ▶ Any seaborn color palette (i.e., something that can be passed to `color_palette()`) can be provided.
- ▶ You can also use a dictionary that maps the names of values in the hue variable to valid matplotlib colors:

```
pal = dict(Lunch="seagreen", Dinner="gray")
g = sns.FacetGrid(tips, hue="time", palette=pal, size=5)
g.map(plt.scatter, "total_bill", "tip", s=50, alpha=.7)
g.add_legend();
../_images/axis_grids_22_0.png
```

Seaborn Workshop

- ▶ You can also let other aspects of the plot vary across levels of the hue variable, which can be helpful for making plots that will be more comprehensible when printed in black-and-white.
- ▶ To do this, pass a dictionary to `hue_kws` where keys are the names of plotting function keyword arguments and values are lists of keyword values, one for each level of the hue variable.

Seaborn Workshop

```
g = sns.FacetGrid(tips, hue="sex", palette="Set1", size=4, aspect=1.5)  
g.map(plt.scatter, "total_bill", "tip", s=100, linewidth=1)  
g.add_legend()  
../_images/axis_grids_24_0.png
```

Seaborn Workshop

If you have many levels of one variable, you can plot it along the columns but wrap them so that they span multiple rows. When doing this, you cannot use a row variable.

```
attend = sns.load_dataset("attention").query("subject < 10")
g = sns.FacetGrid(attend, col="subject", col_wrap=4, sharey=True)
g.map(sns.pointplot, "solutions", "score", color=".3",
      ../_images/axis_grids_26_0.png
```

Seaborn Workshop

- ▶ Once you've drawn a plot using `FacetGrid.map()` (which can be called multiple times), you may want to adjust some aspects of the plot.
- ▶ There are also a number of methods on the `FacetGrid` object for manipulating the figure at a higher level of abstraction.
- ▶ The most general is `FacetGrid.set()`, and there are other more specialized methods like `FacetGrid.set_axis_labels()`, which respects the fact that interior facets do not have axis labels. For example:

Seaborn Workshop

```
with sns.axes_style("white"):

    g = sns.FacetGrid(tips, row="sex", col="smoker", ma
    g.map(plt.scatter, "total_bill", "tip", color="#334488")
    g.set_axis_labels("Total bill (US Dollars)", "Tip");
    g.set(xticks=[10, 30, 50], yticks=[2, 6, 10]);
    g.fig.subplots_adjust(wspace=.02, hspace=.02);
    ../_images/axis_grids_28_0.png
```

Seaborn Workshop

- ▶ For even more customization, you can work directly with the underlying matplotlib Figure and Axes objects, which are stored as member attributes at `fig` and `axes` (a two-dimensional array), respectively.
- ▶ When making a figure without row or column faceting, you can also use the `ax` attribute to directly access the single axes.

Seaborn Workshop

```
g = sns.FacetGrid(tips, col="smoker", margin_titles=True)
g.map(plt.scatter, "total_bill", "tip", color="#338844")
for ax in g.axes.flat:
    ax.plot((0, 50), (0, .2 * 50), c=".2", ls="--")
g.set(xlim=(0, 60), ylim=(0, 14));
```

Mapping custom functions onto the grid

- ▶ You're not limited to existing matplotlib and seaborn functions when using FacetGrid. However, to work properly, any function you use must follow a few rules:
- ▶ It must plot onto the currently active matplotlib Axes. This will be true of functions in the matplotlib.pyplot namespace, and you can call `plt.gca` to get a reference to the current Axes if you want to work directly with its methods.
- ▶ It must accept the data that it plots in positional arguments. Internally, FacetGrid will pass a Series of data for each of the named positional arguments passed to `FacetGrid.map()`.

Seaborn Workshop

- ▶ It must be able to accept color and label keyword arguments, and, ideally, it will do something useful with them.
- ▶ In most cases, its easiest to catch a generic dictionary of `**kwargs` and pass it along to the underlying plotting function.
- ▶ Lets look at minimal example of a function you can plot with. This function will just take a single vector of data for each facet:

Seaborn Workshop

```
def quantile_plot(x, **kwargs):  
    qntls, xr = stats.probplot(x, fit=False)  
    plt.scatter(xr, qntls, **kwargs)  
  
g = sns.FacetGrid(tips, col="sex", size=4)  
g.map(quantile_plot, "total_bill");  
../_images/axis_grids_32_0.png
```

Seaborn Workshop

If we want to make a bivariate plot, you should write the function so that it accepts the x-axis variable first and the y-axis variable second:

```
def qqplot(x, y, **kwargs):  
    _, xr = stats.probplot(x, fit=False)  
    _, yr = stats.probplot(y, fit=False)  
    plt.scatter(xr, yr, **kwargs)  
  
g = sns.FacetGrid(tips, col="smoker", size=4)  
g.map(qqplot, "total_bill", "tip");  
../_images/axis_grids_34_0.png
```

Seaborn Workshop

Because `plt.scatter` accepts color and label keyword arguments and does the right thing with them, we can add a hue facet without any difficulty:

```
g = sns.FacetGrid(tips, hue="time", col="sex", size=4)
g.map(qqplot, "total_bill", "tip")
g.add_legend();
../_images/axis_grids_36_0.png
```

Seaborn Workshop

This approach also lets us use additional aesthetics to distinguish the levels of the hue variable, along with keyword arguments that won't be dependent on the faceting variables:

```
g = sns.FacetGrid(tips, hue="time", col="sex", size=4,  
hue_kws={"marker": ["s", "D"]})  
g.map(qqplot, "total_bill", "tip", s=40, edgecolor="w")  
g.add_legend();  
../_images/axis_grids_38_0.png
```

Seaborn Workshop

- ▶ Sometimes, though, you'll want to map a function that doesn't work the way you expect with the `color` and `label` keyword arguments.
- ▶ In this case, you'll want to explicitly catch them and handle them in the logic of your custom function.
- ▶ For example, this approach will allow use to map `plt.hexbin`, which otherwise does not play well with the FacetGrid API:

Seaborn Workshop

```
def hexbin(x, y, color, **kwargs):  
    cmap = sns.light_palette(color, as_cmap=True)  
    plt.hexbin(x, y, gridsize=15, cmap=cmap, **kwargs)  
  
with sns.axes_style("dark"):  
    g = sns.FacetGrid(tips, hue="time", col="time", size=(8, 4))  
    g.map(hexbin, "total_bill", "tip", extent=[0, 50, 0, 10])  
    plt.savefig("../_images/axis_grids_40_0.png")
```