Lec 11 - seaborn

Statistical Computing and Computation

Sta 663 | Spring 2022

Dr. Colin Rundel

seaborn

> Seaborn is a library for making statistical graphics in Python. It builds on top of **matplotlib** and integrates closely with **pandas** data structures.
>
> Seaborn helps you explore and understand your data. Its plotting functions operate on dataframes and arrays containing whole datasets and internally perform the necessary semantic mapping and statistical aggregation to produce informative plots. Its dataset-oriented, declarative API lets you focus on what the different elements of your plots mean, rather than on the details of how to draw them.

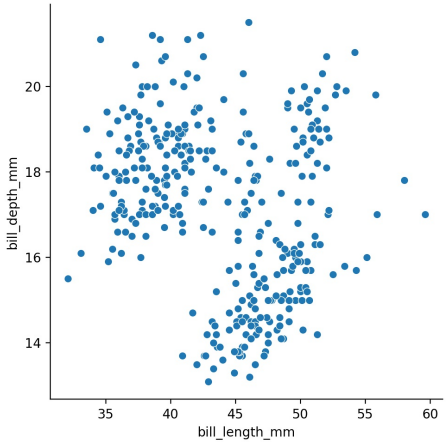
```
import matplotlib.pyplot as plt
import seaborn as sns
```

```
penguins = sns.load_dataset("penguins")
penguins
```

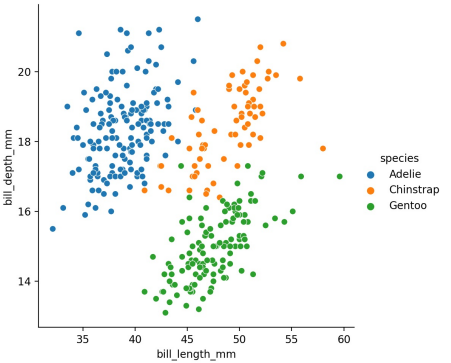
```
##     species     island  bill_length_mm  ...  flipper_length_mm  body_mass_g     sex
## 0    Adelie  Torgersen            39.1  ...              181.0       3750.0    Male
## 1    Adelie  Torgersen            39.5  ...              186.0       3800.0  Female
## 2    Adelie  Torgersen            40.3  ...              195.0       3250.0  Female
## 3    Adelie  Torgersen             NaN  ...                NaN          NaN     NaN
## 4    Adelie  Torgersen            36.7  ...              193.0       3450.0  Female
## ..      ...        ...             ...  ...                ...          ...     ...
## 339  Gentoo     Biscoe             NaN  ...                NaN          NaN     NaN
```

Basic plots

```
sns.relplot(
  data=penguins,
  x="bill_length_mm",
  y="bill_depth_mm"
)
```



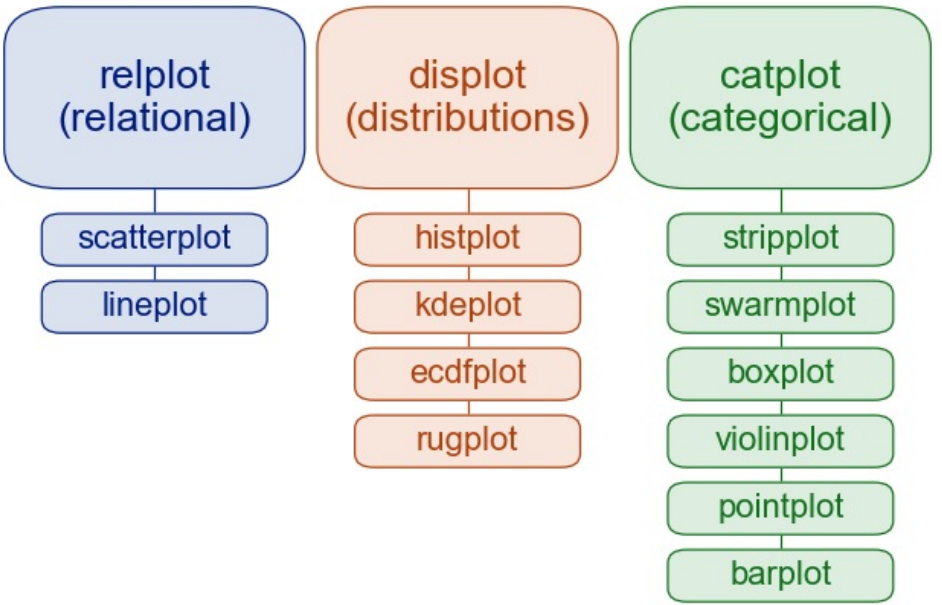

```
sns.relplot(
  data=penguins,
  x="bill_length_mm",
  y="bill_depth_mm",
  hue="species"
)
```



A more complex plot

```
sns.relplot(
  data=penguins,
  x="bill_length_mm", y="bill_depth_mm",
  hue="species",
  col="island", row="species"
)
```

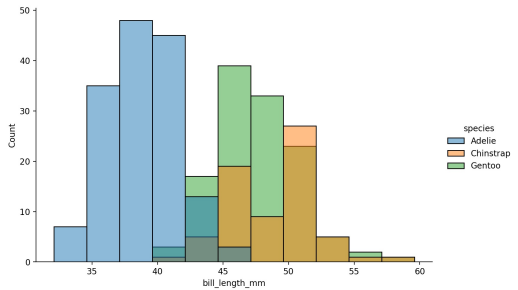
Figure-level vs. axes-level functions



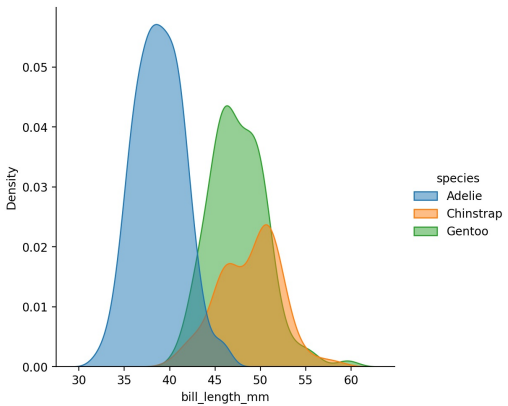
These are not the only axes-level functions - we see additional plotting functions in a bit

displots

```
sns.displot(
  data = penguins,
  x = "bill_length_mm", hue = "species",
  alpha = 0.5, aspect = 1.5
)
```

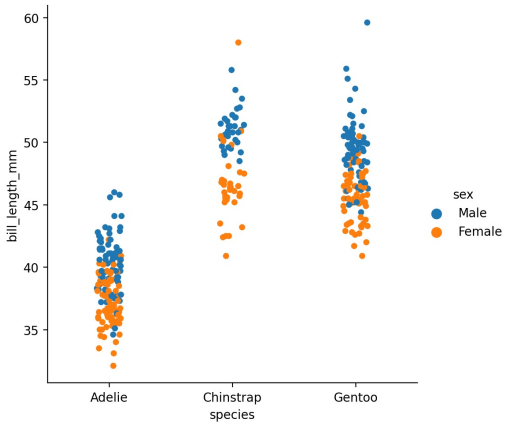



```
sns.displot(
  data = penguins,
  x = "bill_length_mm", hue = "species",
  kind = "kde", fill=True,
  alpha = 0.5, aspect = 1
)
```



catplots

```
sns.catplot(
  data = penguins,
  x = "species", y = "bill_length_mm",
  hue = "sex"
)
```




```
sns.catplot(
  data = penguins,
  x = "species", y = "bill_length_mm",
  hue = "sex",
  kind = "box"
)
```

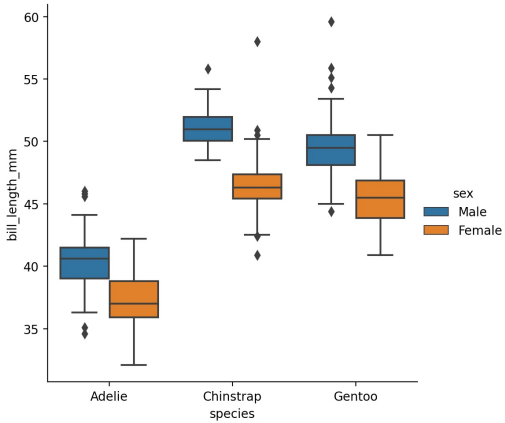
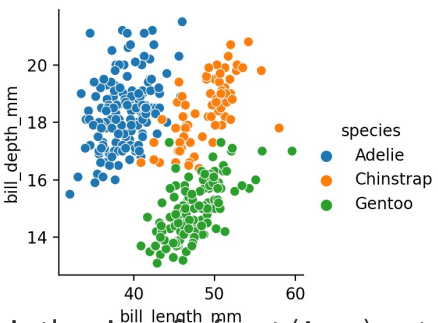


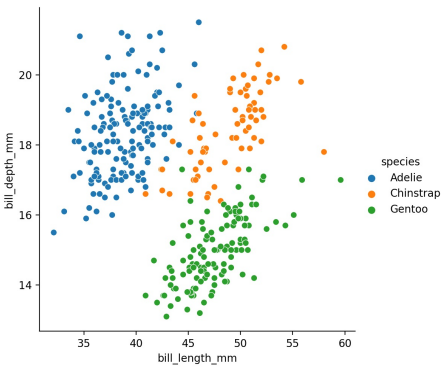
figure-level plot size

To adjust the size of plots generated via a figure-level plotting function adjust the `aspect` and `height` arguments, figure width is `aspect * height`.

```
sns.relplot(
  data=penguins,
  x="bill_length_mm", y="bill_depth_mm",
  hue="species",
  aspect = 1, height = 3
)
```



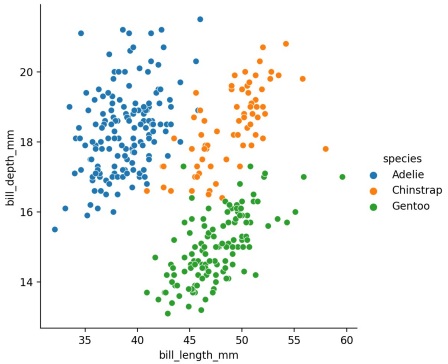

```
sns.relplot(
  data=penguins,
  x="bill_length_mm", y="bill_depth_mm",
  hue="species",
  aspect = 1, height = 5
)
```



Note this is the size of a facet (Axes) not the figure

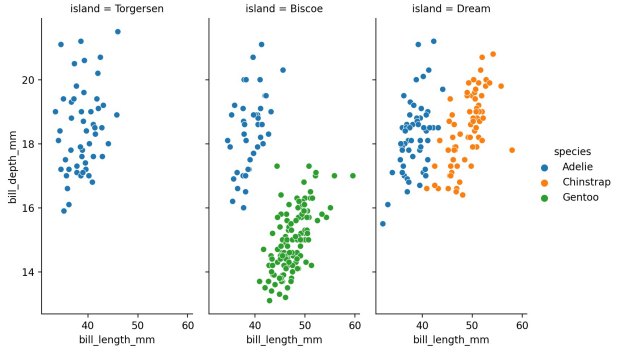
figure-level plot details

```
g = sns.relplot(
  data=penguins,
  x="bill_length_mm", y="bill_depth_mm",
  hue="species",
  aspect = 1
)

g
```




```
print(g)
```

```
h = sns.relplot(
  data=penguins,
  x="bill_length_mm", y="bill_depth_mm",
  hue="species", col="island",
  aspect = 1/2
)

h
```



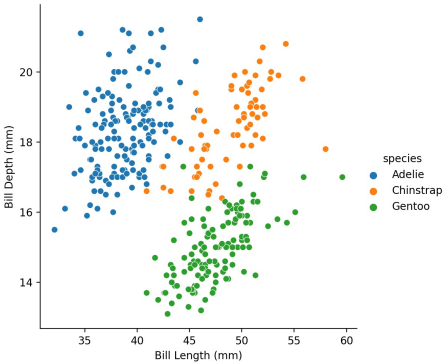

```
print(h)
```

FacetGird methods

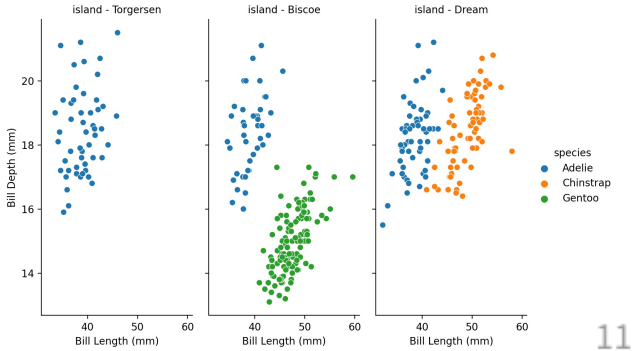
Method	Description
`add_legend()`	Draw a legend, maybe placing it outside axes and resizing the figure
`despine()`	Remove axis spines from the facets.
`facet_axis()`	Make the axis identified by these indices active and return it.
`facet_data()`	Generator for name indices and data subsets for each facet.
`map()`	Apply a plotting function to each facet's subset of the data.
`map_dataframe()`	Like `.map()` but passes args as strings and inserts data in kwargs.
`refline()`	Add a reference line(s) to each facet.
`savefig()`	Save an image of the plot.
`set()`	Set attributes on each subplot Axes.
`set_axis_labels()`	Set axis labels on the left column and bottom row of the grid.
`set_titles()`	Draw titles either above each facet or on the grid margins.
`set_xlabels()`	Label the x axis on the bottom row of the grid.
`set_xticklabels()`	Set x axis tick labels of the grid.

Adjusting labels

```
sns.relplot(
  data=penguins,
  x="bill_length_mm", y="bill_depth_mm",
  hue="species",
  aspect = 1
).set_axis_labels(
  "Bill Length (mm)", "Bill Depth (mm)"
)
```




```
sns.relplot(
  data=penguins,
  x="bill_length_mm", y="bill_depth_mm",
  hue="species", col="island",
  aspect = 1/2
).set_axis_labels(
  "Bill Length (mm)", "Bill Depth (mm)"
).set_titles(
  "{col_var} - {col_name}"
)
```

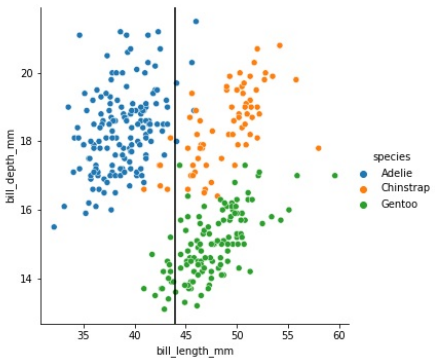


FacetGrid attributes

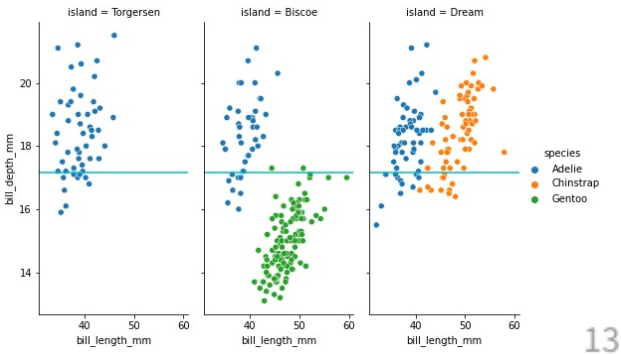
Attribute	Description
`ax`	The `matplotlib.axes.Axes` when no faceting variables are assigned.
`axes`	An array of the `matplotlib.axes.Axes` objects in the grid.
`axes_dict`	A mapping of facet names to corresponding `matplotlib.axes.Axes`.
`figure`	Access the `matplotlib.figure.Figure` object underlying the grid.
`legend`	The `matplotlib.legend.Legend` object, if present.

Using axes to modify plots

```
g = sns.relplot(
  data=penguins,
  x="bill_length_mm", y="bill_depth_mm",
  hue="species",
  aspect = 1
)
g.ax.axvline(
  x = penguins.bill_length_mm.mean(), c="k"
)
```




```
h = sns.relplot(
  data=penguins,
  x="bill_length_mm", y="bill_depth_mm",
  hue="species", col="island",
  aspect = 1/2
)
mean_bill_dep = penguins.bill_depth_mm.mean()

[ ax.axhline(y=mean_bill_dep, c="c")
  for row in h.axes for ax in row ]
```



Why figure-level functions

Advantages:

- Easy faceting by data variables
- Legend outside of plot by default
- Easy figure-level customization
- Different figure size parameterization

Disadvantages:

- Many parameters not in function signature
- Cannot be part of a larger matplotlib figure
- Different API from matplotlib
- Different figure size parameterization

Details based on seaborn docs

lmplots

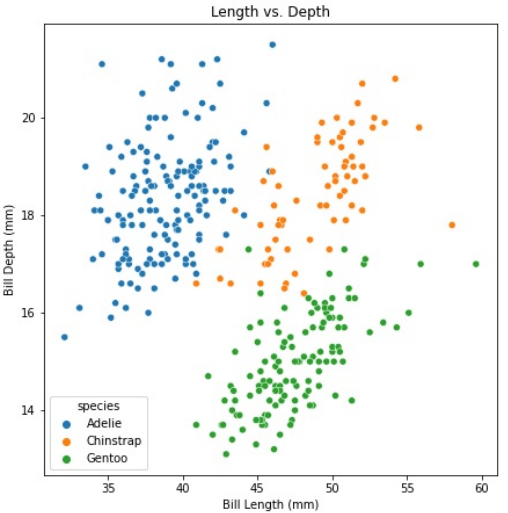
There is one last figure-level plot type - `lmplot()` which is a convenient interface to fitting and ploting regression models across subsets of data,

```
sns.lmplot(
  data=penguins,
  x="bill_length_mm", y="bill_depth_mm",
  hue="species", col="island",
  aspect = 1, truncate=False
)
```

axes-level functions

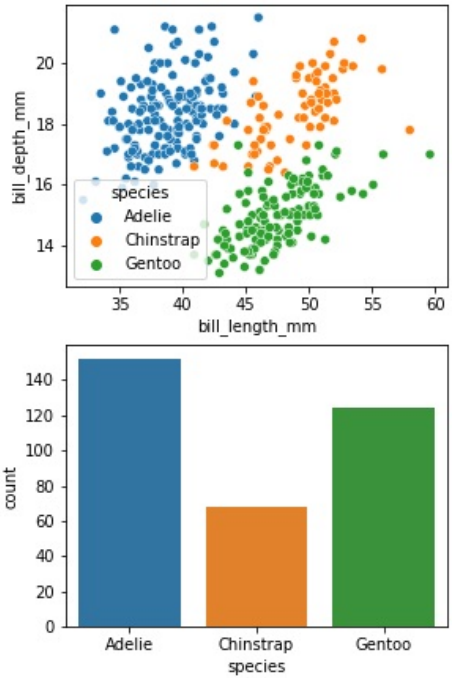
These functions return a `matplotlib.pyplot.Axes` object instead of a `FacetGrid` giving more direct control over the plot using basic matplotlib tools.

```
plt.figure()

sns.scatterplot(
  data=penguins,
  x="bill_length_mm", y="bill_depth_mm",
  hue="species"
)

plt.xlabel("Bill Length (mm)")
plt.ylabel("Bill Depth (mm)")
plt.title("Length vs. Depth")

plt.show()
```



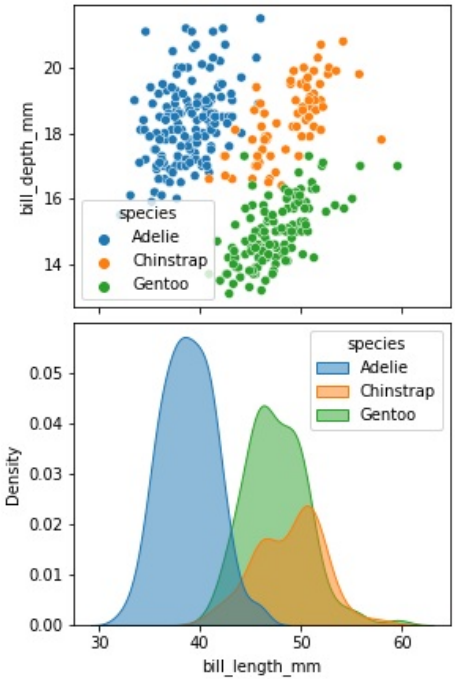
subplots - pyplot style

```
plt.figure(figsize=(4,6), layout="constrained")

plt.subplot(211)
sns.scatterplot(
  data=penguins,
  x="bill_length_mm", y="bill_depth_mm",
  hue="species"
)

plt.subplot(212)
sns.countplot(
  data=penguins,
  x="species"
)

plt.show()
```



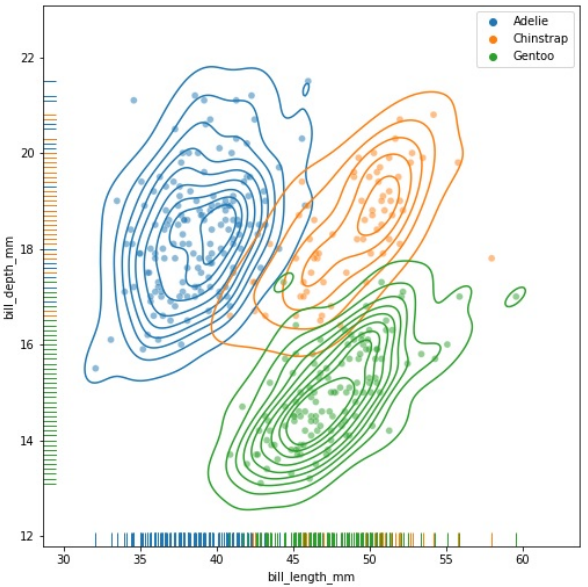
subplots - OO style

```
fig, axs = plt.subplots(
  2, 1, figsize=(4,6),
  layout="constrained",
  sharex=True
)

sns.scatterplot(
  data=penguins,
  x="bill_length_mm", y="bill_depth_mm",
  hue="species",
  ax = axs[0]
)

sns.kdeplot(
  data=penguins,
  x="bill_length_mm", hue="species",
  fill=True, alpha=0.5,
  ax = axs[1]
)

plt.show()
```



layering plots

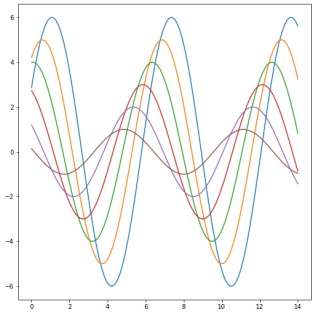
```
plt.figure(layout="constrained")

sns.kdeplot(
  data=penguins,
  x="bill_length_mm", y="bill_depth_mm",
  hue="species"
)

sns.scatterplot(
  data=penguins,
  x="bill_length_mm", y="bill_depth_mm",
  hue="species", alpha=0.5
)

sns.rugplot(
  data=penguins,
  x="bill_length_mm", y="bill_depth_mm",
  hue="species"
)

plt.legend()

plt.show()
```



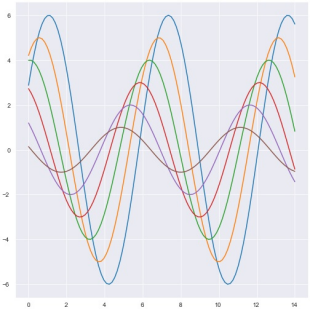
Themes

Seaborn comes with a number of themes (`darkgrid`, `whitegrid`, `dark`, `white`, and `ticks`) which can be enabled by `sns.set_theme()` at the figure level or `sns.axes_style()` at the axes level.

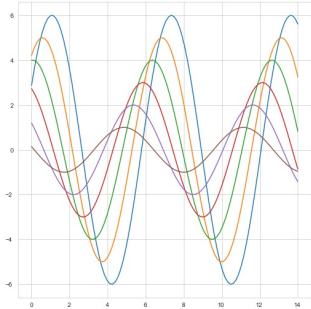
```
def sinplot():
    x = np.linspace(0, 14, 100)
    for i in range(1, 7):
        plt.plot(x, np.sin(x + i * .5) * (7 - i))

sinplot()
plt.show()
```



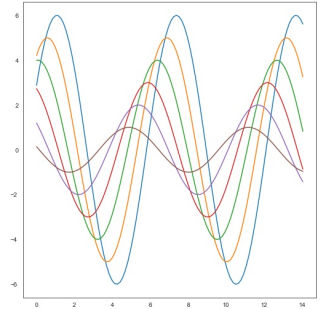
```
with sns.axes_style("darkgrid"):
  sinplot()
  plt.show()
```



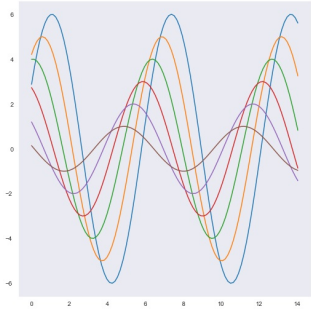
```
with sns.axes_style("whitegrid"):
  sinplot()
  plt.show()
```



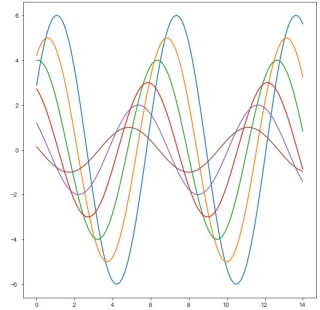
```
with sns.axes_style("white"):
  sinplot()
  plt.show()
```



```
with sns.axes_style("dark"):
  sinplot()
  plt.show()
```

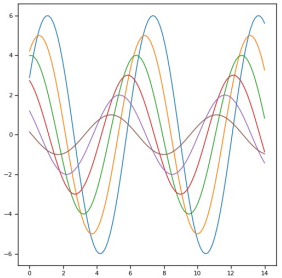


```
with sns.axes_style("ticks"):
  sinplot()
  plt.show()
```

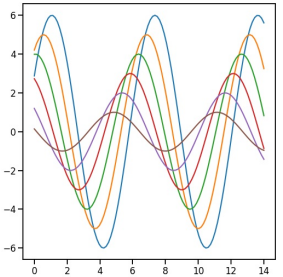


Context

```
sns.set_context("notebook")
sinplot()
plt.show()
```



```
sns.set_context("talk")
sinplot()
plt.show()
```



```
sns.set_context("paper")
sinplot()
plt.show()
```

```
sns.set_context("poster")
sinplot()
plt.show()
```

Color palettes

All of the examples below are the result of calls to `sns.color_palette()` with `as_cmap=True` for the continuous case,

```
show_palette()
```

```
show_palette("tab10")
```

```
show_palette("hls")
```

```
show_palette("husl")
```

```
show_palette("Set2")
```

```
show_cont_palette("cubehelix")
```

```
show_cont_palette("light:b")
```

```
show_cont_palette("dark:salmon_r")
```

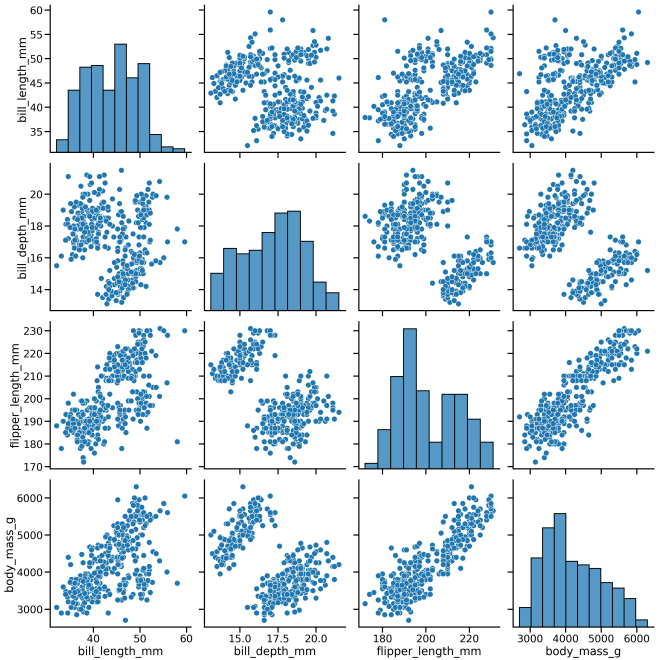
```
show_cont_palette("YlOrBr")
```

```
show_cont_palette("vlag")
```

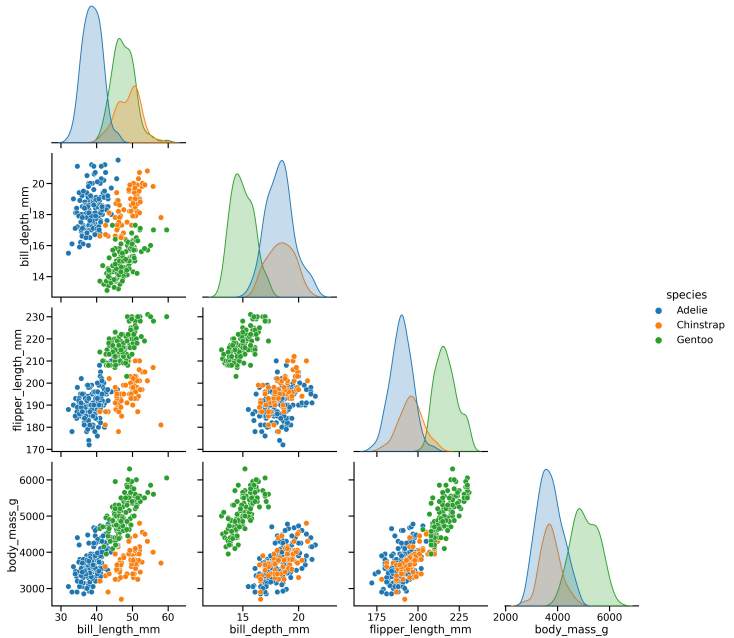
See more examples in the color palettes tutorial

Pair plots

```
sns.pairplot(data = penguins, height=5)
```



```
sns.pairplot(data = penguins, hue="species", height=5, corner=True)
```



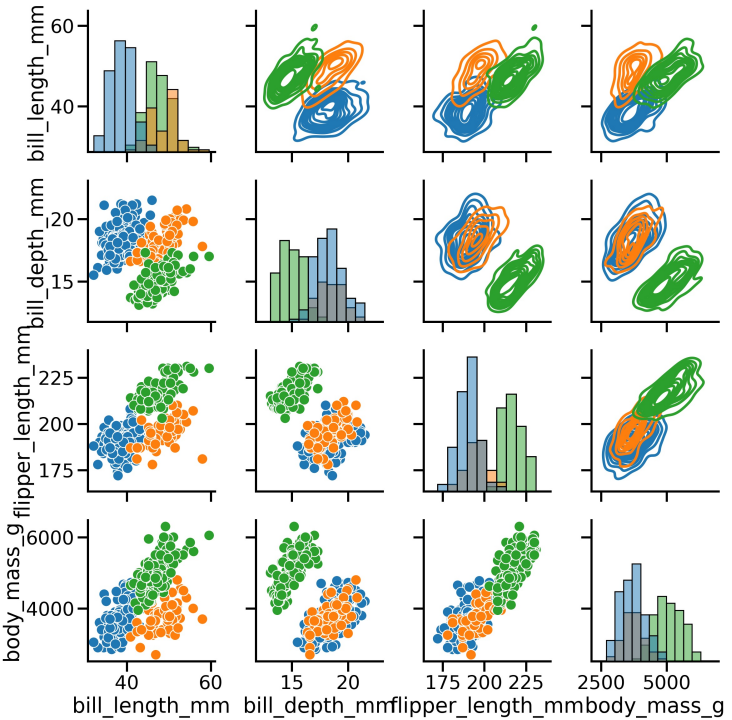
PairGrid

`pairplot()` is a special case of the more general `PairGrid` - once constructed there are methods that allow for mapping plot functions of the different axes,

```
sns.PairGrid(penguins, hue="species", height=5)
```

Mapping

```
g = sns.PairGrid(
  penguins, hue="species",
  height=3
)

g = g.map_diag(
  sns.histplot, alpha=0.5
)

g = g.map_lower(
  sns.scatterplot
)

g = g.map_upper(
  sns.kdeplot
)

g
```



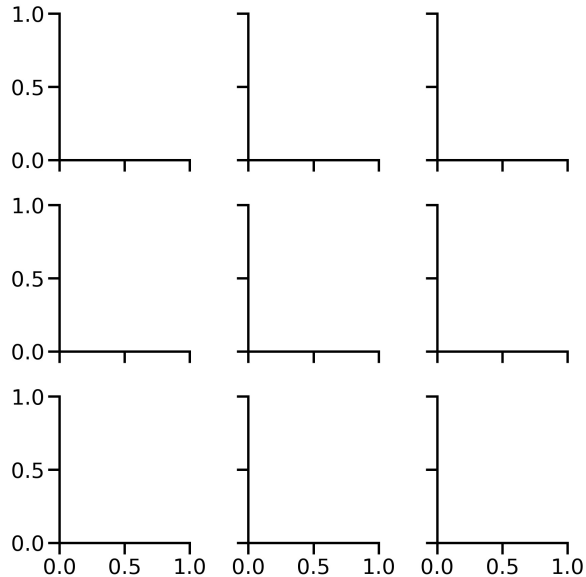
Pair subsets

```
x_vars = ["body_mass_g", "bill_length_mm", "bill_depth_mm", "flipper_length_mm"]
y_vars = ["body_mass_g"]

g = sns.PairGrid(penguins, hue="species", x_vars=x_vars, y_vars=y_vars, height=3)

g = g.map_diag(sns.kdeplot, fill=True)
g = g.map_offdiag(sns.scatterplot, size=penguins["body_mass_g"])
g = g.add_legend()

g
```

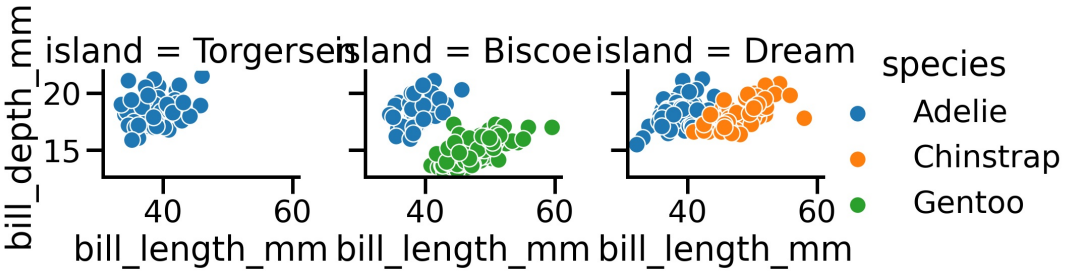
Custom FacetGrids

Just like `PairGrid`s it is possible to construct `FacetGrid`s from scratch,

```
sns.FacetGrid(penguins, col="island", row="species")
```

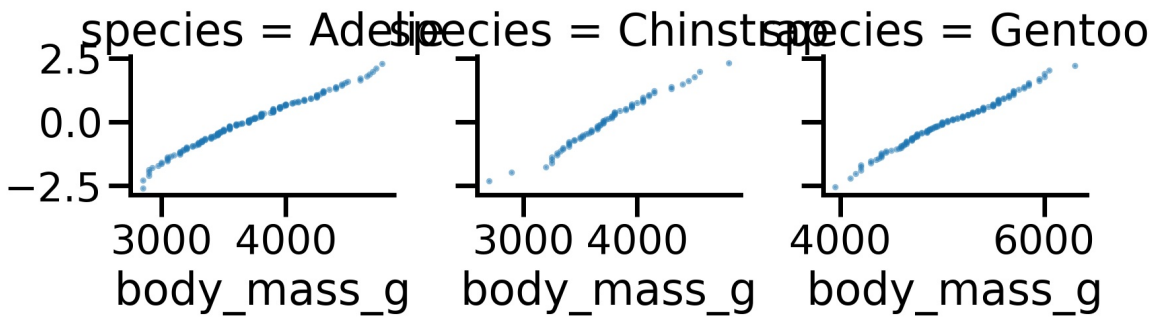


```
g = sns.FacetGrid(penguins, col="island", hue="species")

g = g.map(sns.scatterplot, "bill_length_mm", "bill_depth_mm")
g = g.add_legend()

g
```



Custom plots / functions

```
from scipy import stats
def quantile_plot(x, **kwargs):
    quantiles, xr = stats.probplot(x, fit=False)
    plt.scatter(xr, quantiles, **kwargs)

g = sns.FacetGrid(penguins, col="species", height=3, sharex=False)
g.map(quantile_plot, "body_mass_g", s=2, alpha=0.5)
```



Example from axis grid tutorial

jointplot

One final figure-level plot, is a joint plot which includes marginal distributions along the x and y-axis.

```
g = sns.jointplot(data=penguins, x="bill_length_mm", y="bill_depth_mm", hue="species")
```

```
## /opt/homebrew/lib/python3.9/site-packages/seaborn/axisgrid.py:1740: UserWarning: Tight layout not applied. t
##   f.tight_layout()
```

```
plt.show()
```

Adjusting

The main plot (joint) and the margins (marginal) can be modified by keywords or via layering (use `plot_joint()` and `plot_marginals()` methods).

```
g = sns.jointplot(data=penguins, x="bill_length_mm", y="bill_depth_mm", hue="species", marginal_kws=dict(
```

```
## /opt/homebrew/lib/python3.9/site-packages/seaborn/axisgrid.py:1740: UserWarning: Tight layout not applied. t
##   f.tight_layout()
```

```
g = g.plot_joint(sns.kdeplot, alpha=0.5, levels=5)
plt.show()
```