

Generating graphs: boxplot

Lisa Hopcroft

17/08/2021

```
### Example
### https://rstudio.github.io/reticulate/articles/r\_markdown.html
import pandas
import matplotlib.pyplot as plt
import numpy as np
import seaborn as sns
#
# t = np.arange(0.0, 2.0, 0.01)
# s = 1 + np.sin(2*np.pi*t)
#
# plt.plot( t,s )
# plt.grid(True)
# plt.savefig('test.png')
# plt.show()
```

Generating a boxplot/violin plot etc

Looking at the data

```
head(penguins)
```

```
## # A tibble: 6 x 8
##   species island bill_length_mm bill_depth_mm flipper_length_~ body_mass_g sex
##   <fct>   <fct>         <dbl>         <dbl>         <int>         <int> <fct>
## 1 Adelie  Torge~          39.1          18.7          181          3750 male
## 2 Adelie  Torge~          39.5          17.4          186          3800 fema~
## 3 Adelie  Torge~          40.3           18          195          3250 fema~
## 4 Adelie  Torge~          NA           NA           NA           NA <NA>
## 5 Adelie  Torge~          36.7          19.3          193          3450 fema~
## 6 Adelie  Torge~          39.3          20.6          190          3650 male
## # ... with 1 more variable: year <int>
```

```
print(r.penguins.head(10))
```

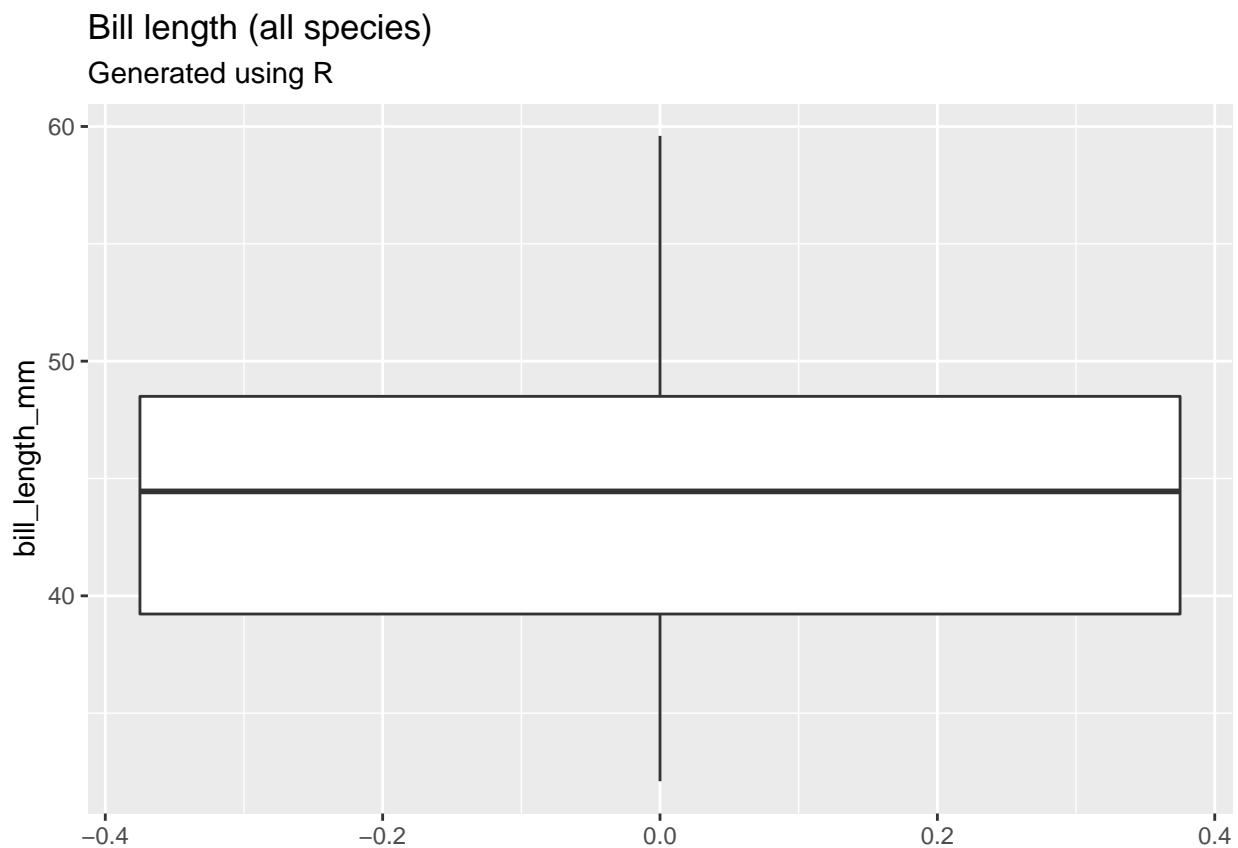
```
##   species   island bill_length_mm ... body_mass_g   sex  year
## 0  Adelie  Torgersen          39.1 ...         3750   male 2007
## 1  Adelie  Torgersen          39.5 ...         3800 female 2007
```

```
## 2 Adelie Torgersen      40.3 ...      3250 female 2007
## 3 Adelie Torgersen      NaN ... -2147483648   NaN 2007
## 4 Adelie Torgersen      36.7 ...      3450 female 2007
## 5 Adelie Torgersen      39.3 ...      3650   male 2007
## 6 Adelie Torgersen      38.9 ...      3625 female 2007
## 7 Adelie Torgersen      39.2 ...      4675   male 2007
## 8 Adelie Torgersen      34.1 ...      3475   NaN 2007
## 9 Adelie Torgersen      42.0 ...      4250   NaN 2007
##
## [10 rows x 8 columns]
```

Generating a boxplot

```
penguins %>%
  ggplot( aes( y=bill_length_mm) ) +
  geom_boxplot() +
  labs( title="Bill length (all species)",
        subtitle="Generated using R")
```

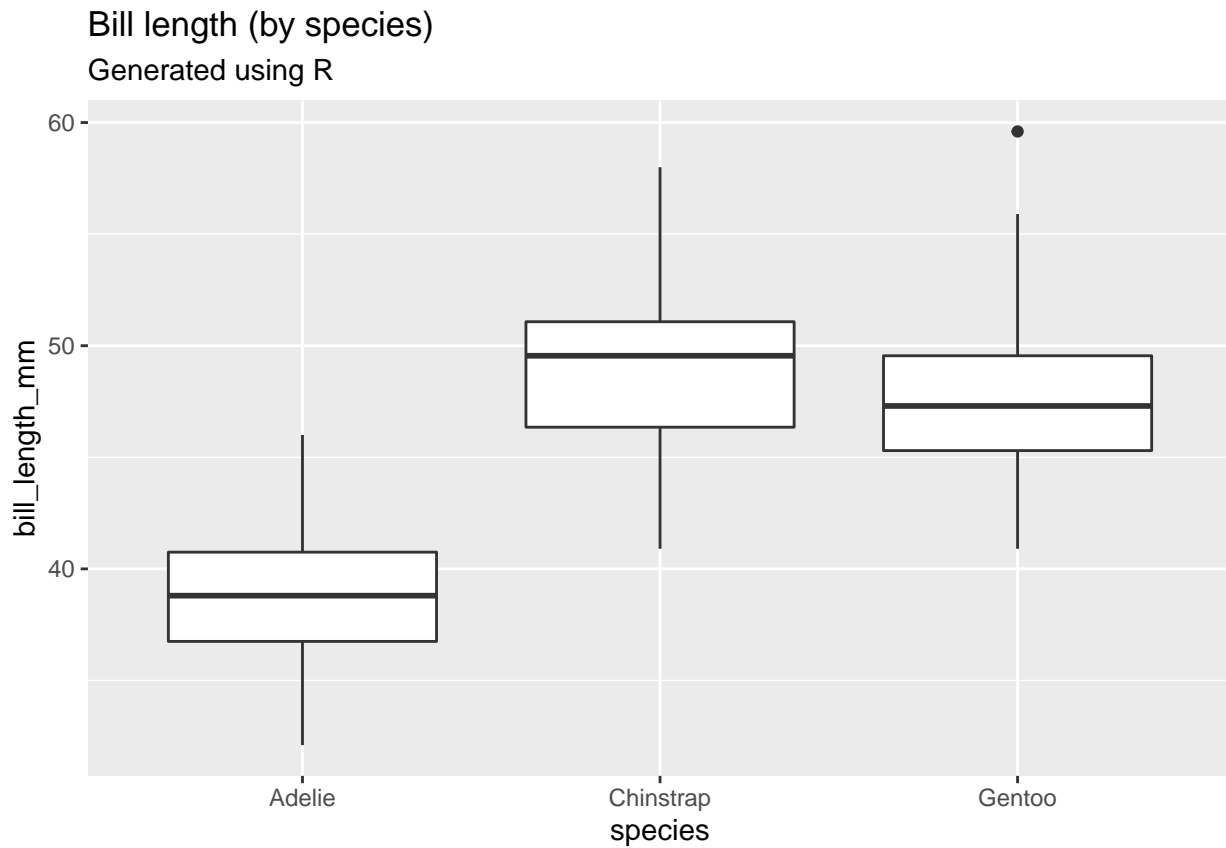
```
## Warning: Removed 2 rows containing non-finite values (stat_boxplot).
```



```
penguins %>%
  ggplot( aes( x=species, y=bill_length_mm) ) +
  geom_boxplot() +
```

```
labs( title="Bill length (by species)",
      subtitle="Generated using R")
```

```
## Warning: Removed 2 rows containing non-finite values (stat_boxplot).
```

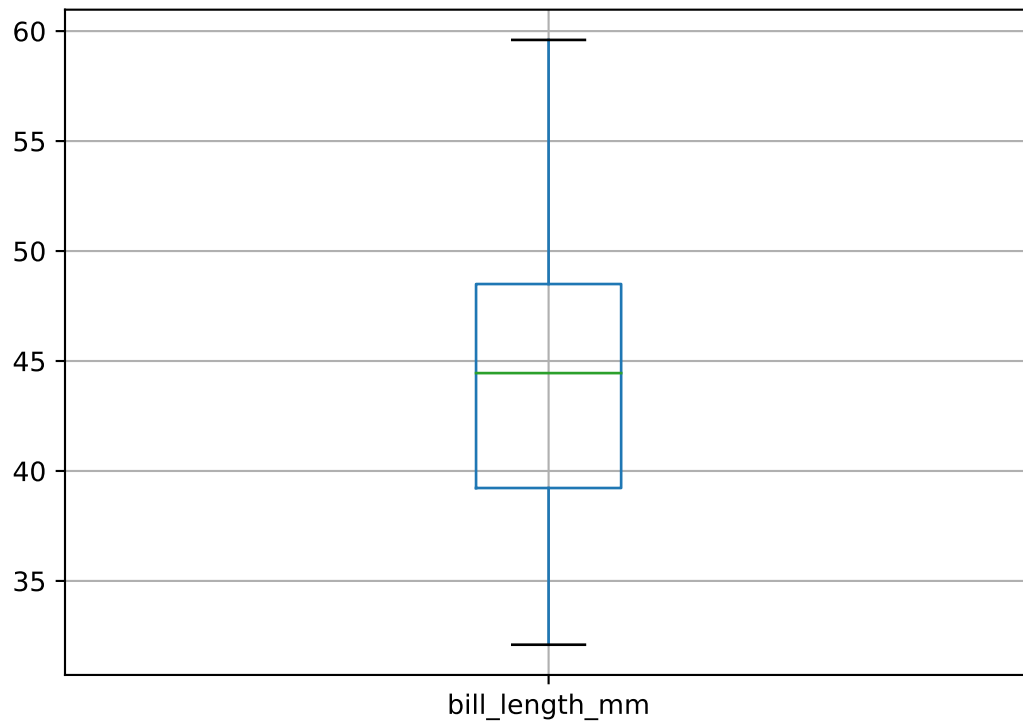


```
### Boxplot using matplotlib
r.penguins.boxplot( column='bill_length_mm' )
```

```
## <AxesSubplot:>
```

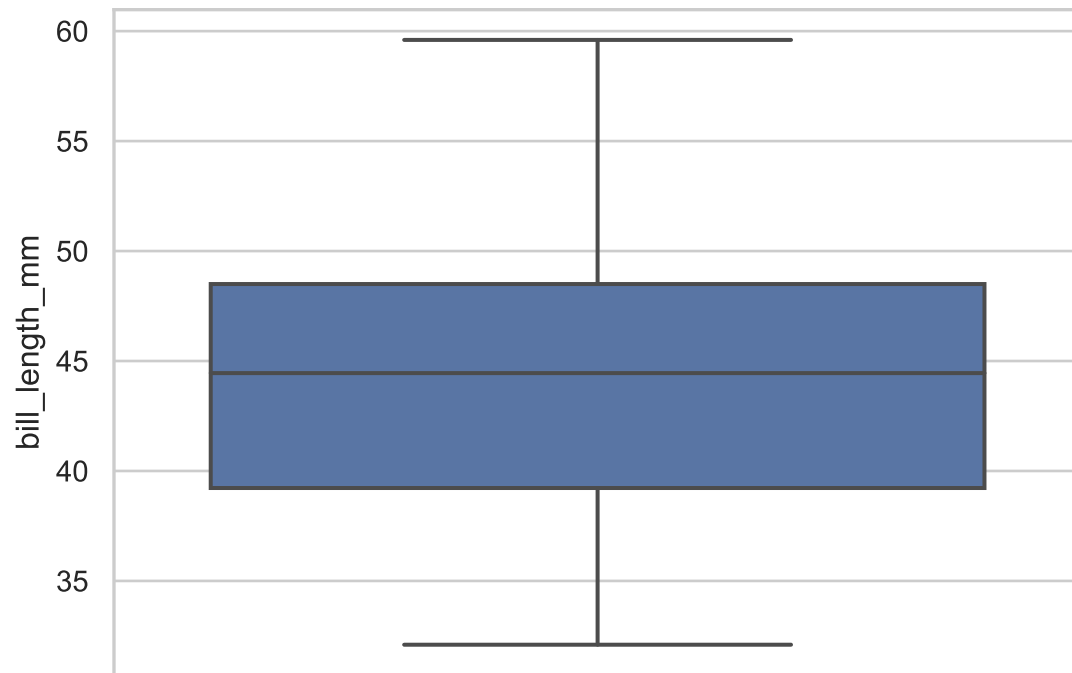
```
plt.show()
```

```
### The same boxplot using Seaborn
### https://seaborn.pydata.org/generated/seaborn.boxplot.html
```

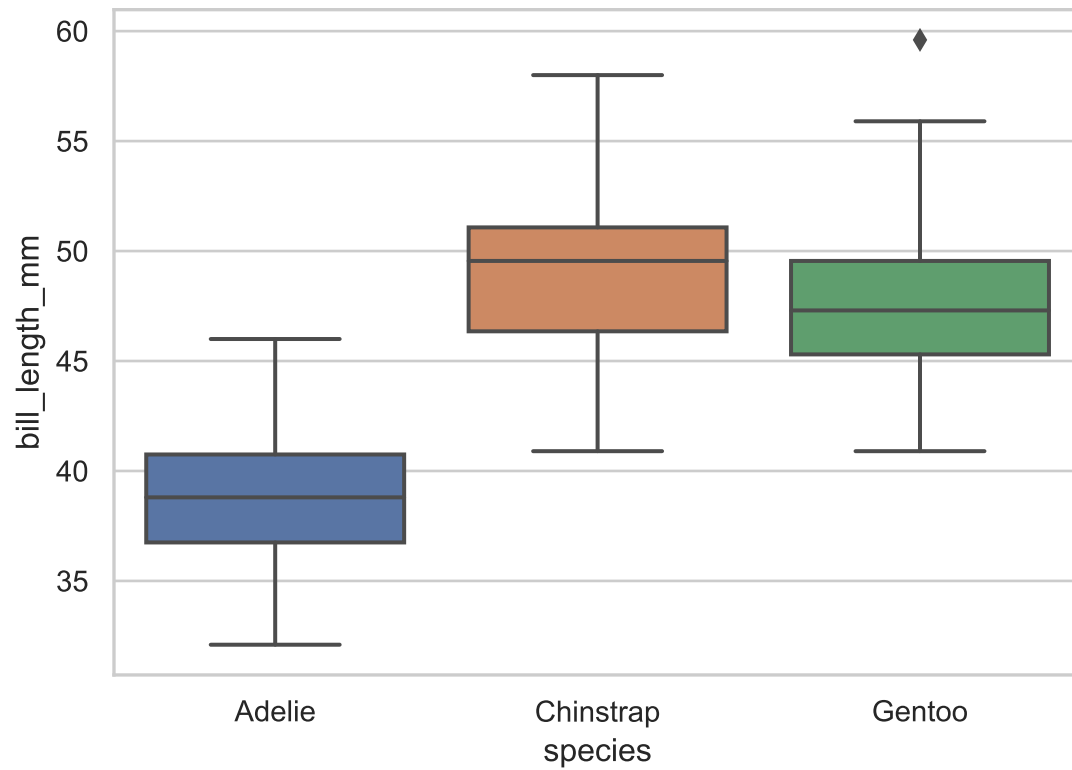


```
sns.set_theme(style="whitegrid")
ax = sns.boxplot( y=r.penguins.bill_length_mm )
plt.show()
```

Splitting the data by species

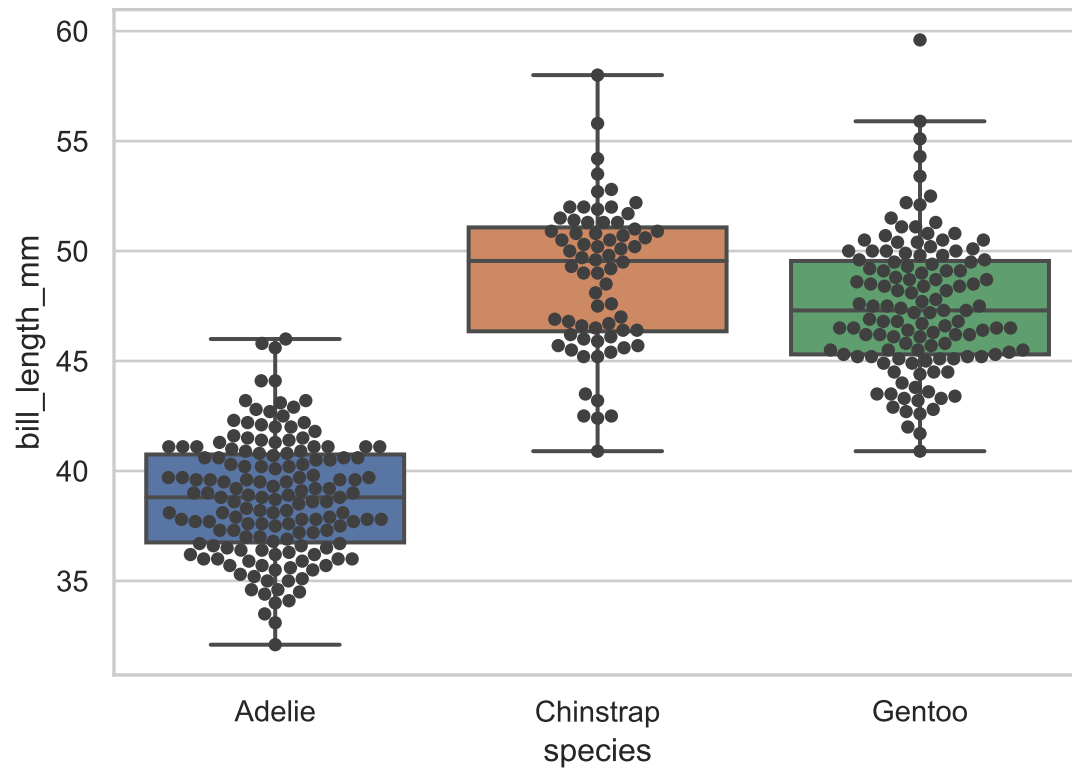


```
ax = sns.boxplot( x="species", y="bill_length_mm", data=r.penguins )  
### Alternatively:  
### ax = sns.boxplot( x=r.penguins.species, y=r.penguins.bill_length_mm )  
plt.show()  
  
### Adding jitter  
### Remove outliers so that they are not plotted twice
```



```
ax = sns.boxplot( x="species", y="bill_length_mm", data=r.penguins, showfliers = False )
ax = sns.swarmplot(x="species", y="bill_length_mm", data=r.penguins, color=".25" )
plt.show()
```

```
### Colouring by another categorical variable
### Remove outliers so that they are not plotted twice
```

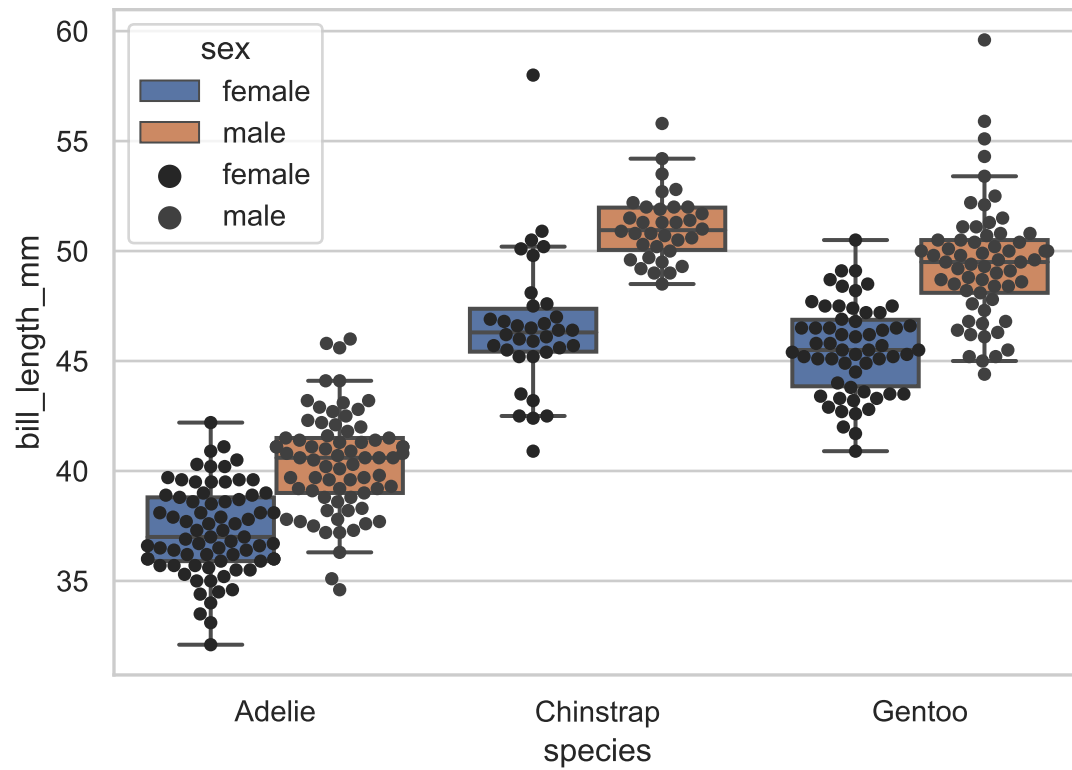


```
ax = sns.boxplot( x="species", y="bill_length_mm", hue="sex", data=r.penguins, showfliers = False )
ax = sns.swarmplot(x="species", y="bill_length_mm", hue="sex", color=".25", data=r.penguins, dodge=True)
```

```
## /Users/lisahopcroft/opt/anaconda3/lib/python3.8/site-packages/seaborn/categorical.py:1296: UserWarning
##   warnings.warn(msg, UserWarning)
## /Users/lisahopcroft/opt/anaconda3/lib/python3.8/site-packages/seaborn/categorical.py:1296: UserWarning
##   warnings.warn(msg, UserWarning)
```

```
plt.show()
```

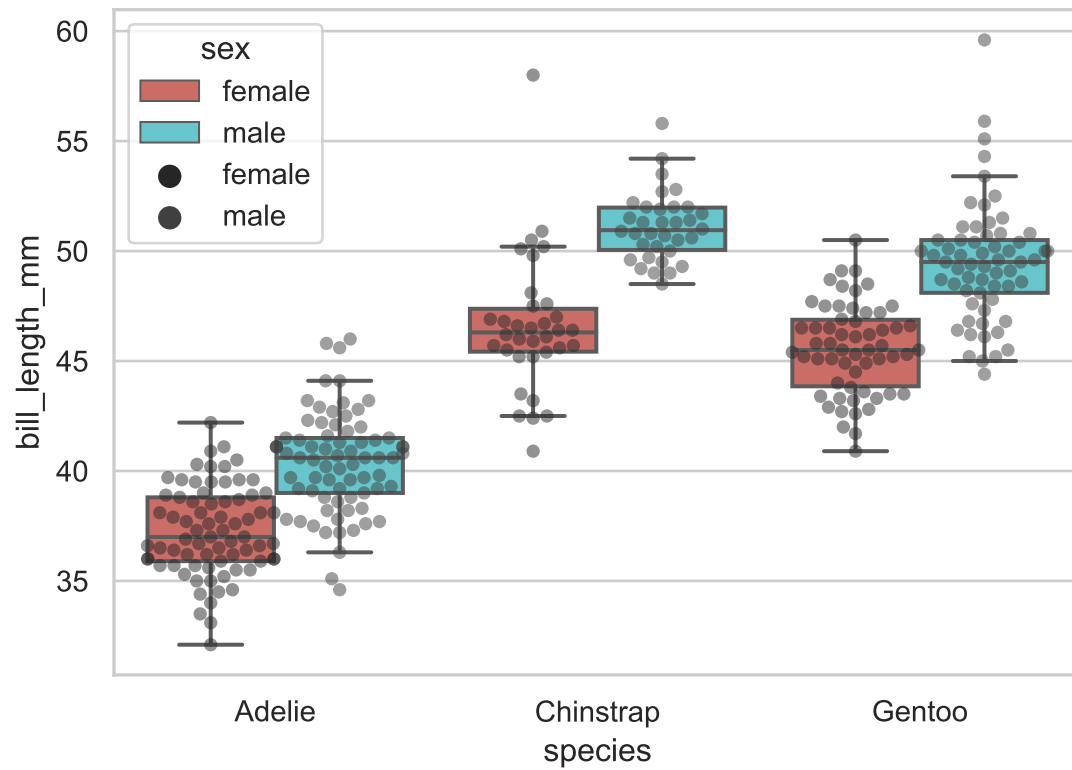
```
### Applying a colour palette and changing alpha
### https://seaborn.pydata.org/tutorial/color\_palettes.html
```



```
ax = sns.boxplot( x="species", y="bill_length_mm", hue="sex", data=r.penguins, palette="hls", showfliers=True)
ax = sns.swarmplot(x="species", y="bill_length_mm", hue="sex", color=".25", alpha=0.5, data=r.penguins,
```

```
## /Users/lisahopcroft/opt/anaconda3/lib/python3.8/site-packages/seaborn/categorical.py:1296: UserWarning:
##   warnings.warn(msg, UserWarning)
## /Users/lisahopcroft/opt/anaconda3/lib/python3.8/site-packages/seaborn/categorical.py:1296: UserWarning:
##   warnings.warn(msg, UserWarning)
```

```
plt.show()
```

Using seaborn

```
citation("palmerpenguins")
```

```
##
## To cite palmerpenguins in publications use:
##
## Horst AM, Hill AP, Gorman KB (2020). palmerpenguins: Palmer
## Archipelago (Antarctica) penguin data. R package version 0.1.0.
## https://allisonhorst.github.io/palmerpenguins/
##
## A BibTeX entry for LaTeX users is
##
## @Manual{,
##   title = {palmerpenguins: Palmer Archipelago (Antarctica) penguin data},
##   author = {Allison Marie Horst and Alison Presmanes Hill and Kristen B Gorman},
##   year = {2020},
##   note = {R package version 0.1.0},
##   url = {https://allisonhorst.github.io/palmerpenguins/},
## }
```