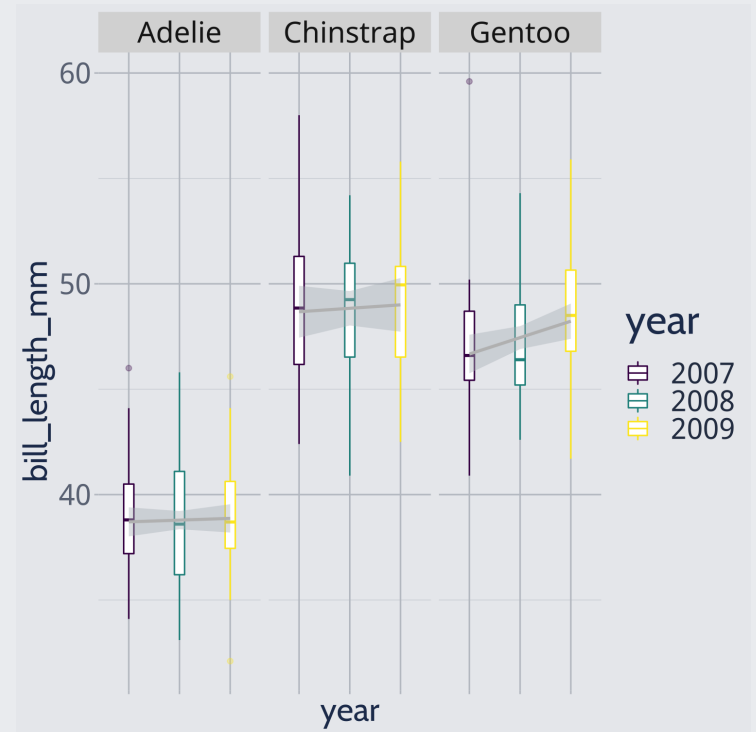
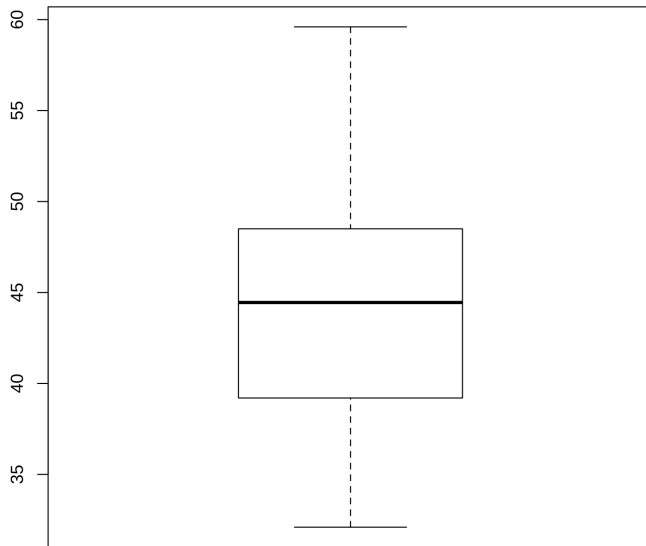


Data Visualisation with ggplot2

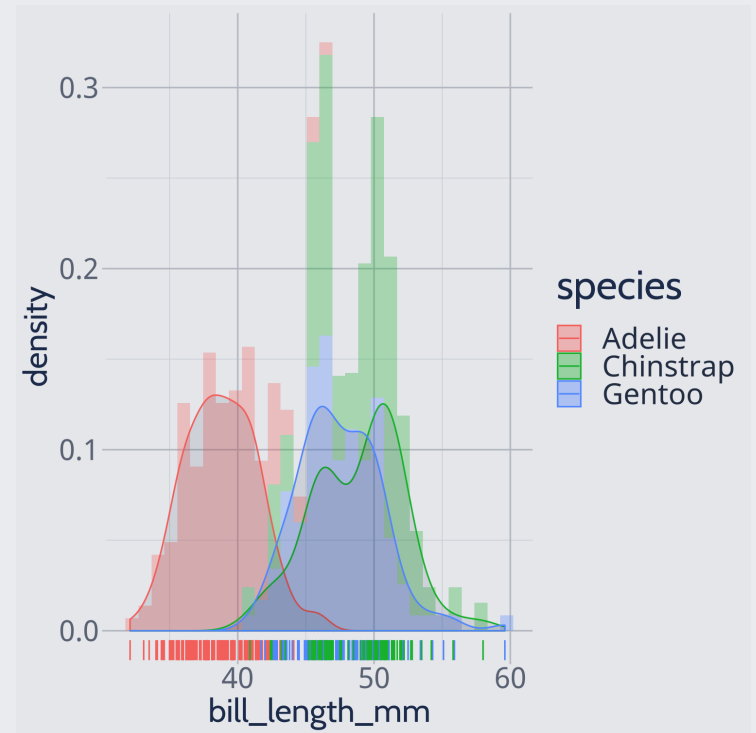
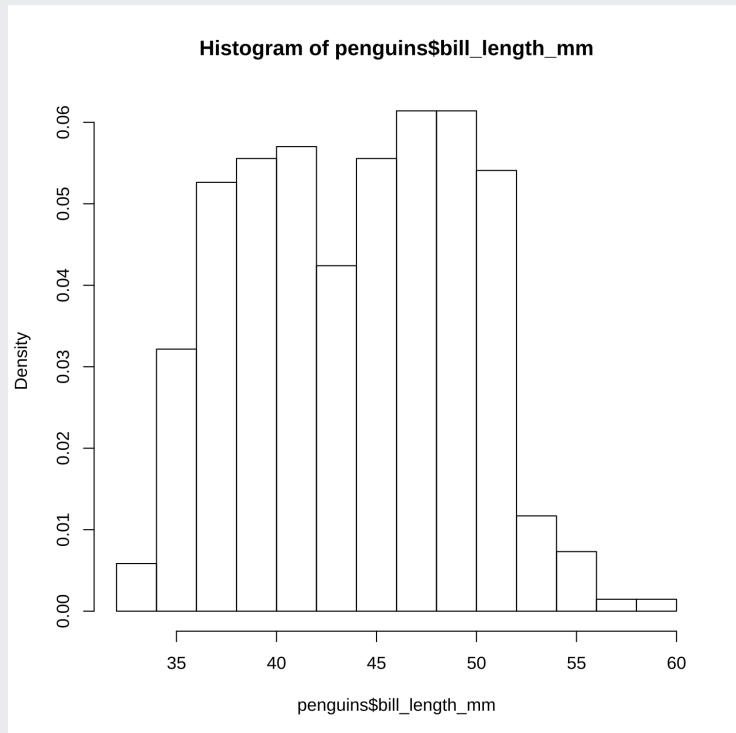
Felix Zaussinger

10.09.2020

Motivation



Motivation



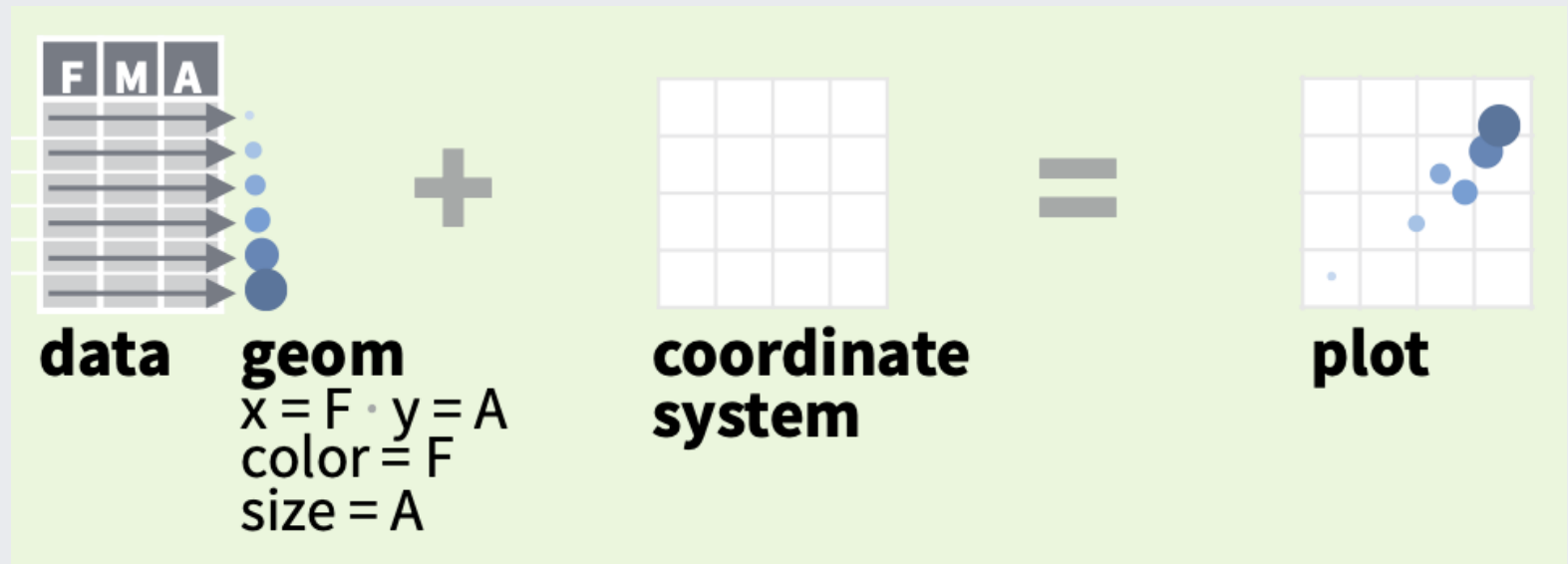
ggplot2

- "The grammar of graphics" -> 3 components make a graph
 - dataset
 - coordinate system
 - geometries ("geoms"): visual marks representing data points

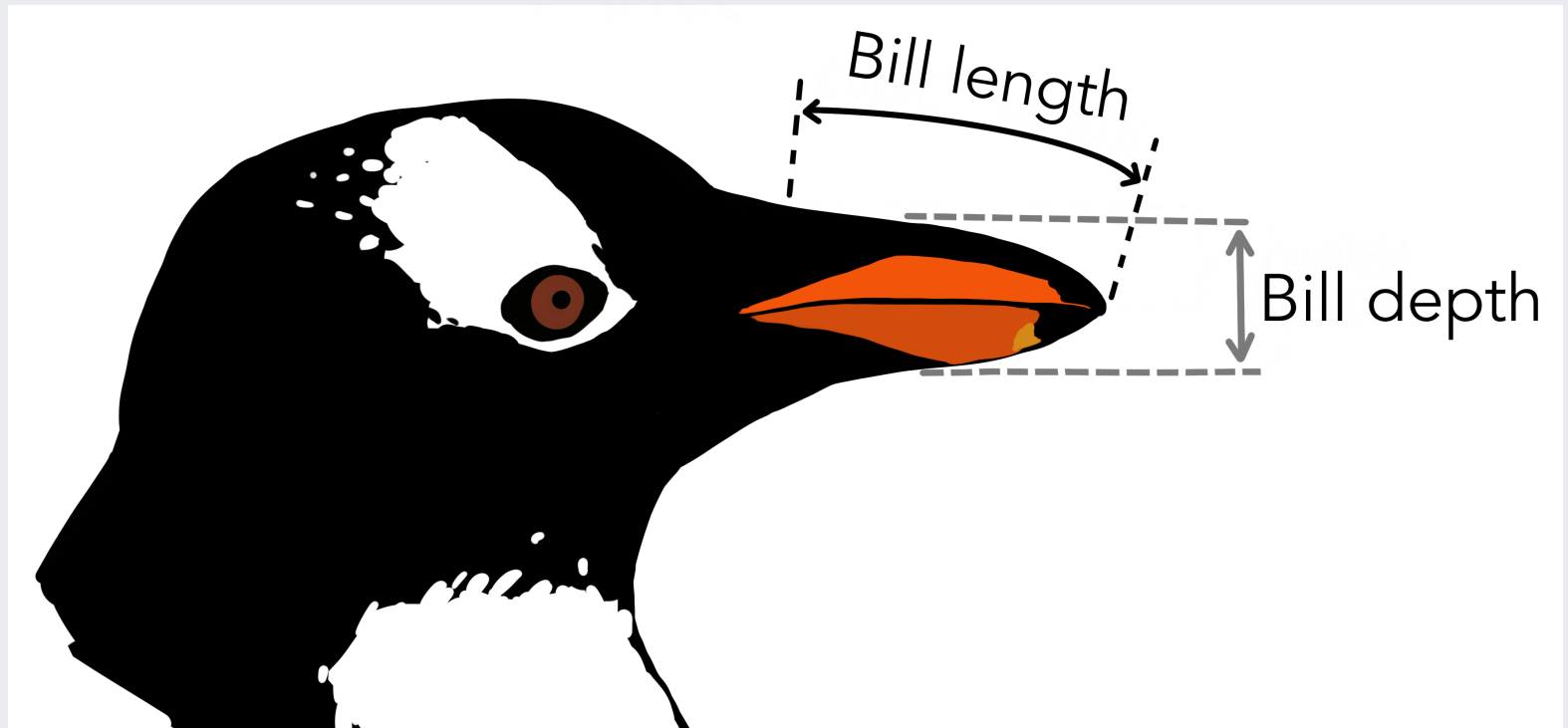


ggplot2

- geom's have properties -> "aesthetics"
 - size
 - color
 - x, y



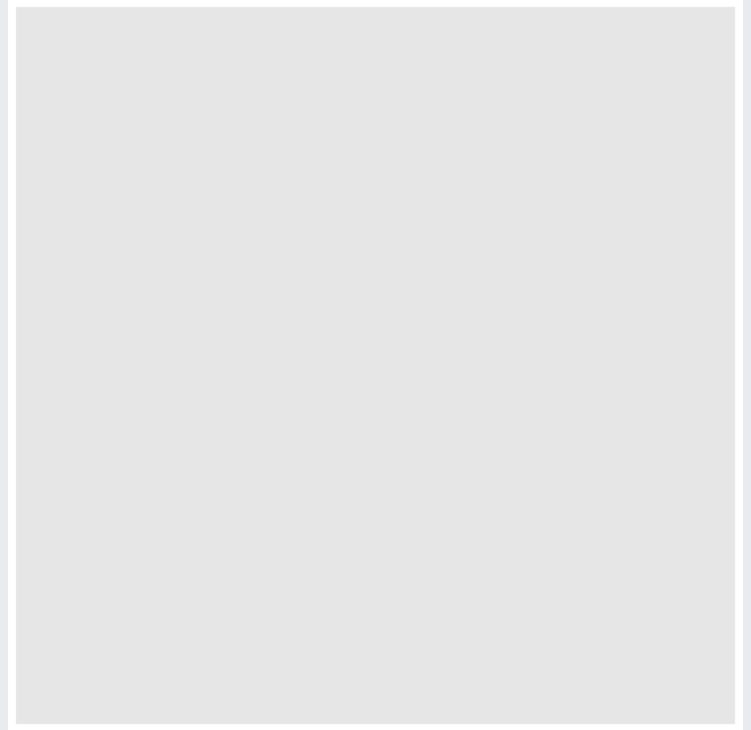
Visualisation practice



(Artwork by @allison_horst, Data from
<https://github.com/allisonhorst/palmerpenguins>)

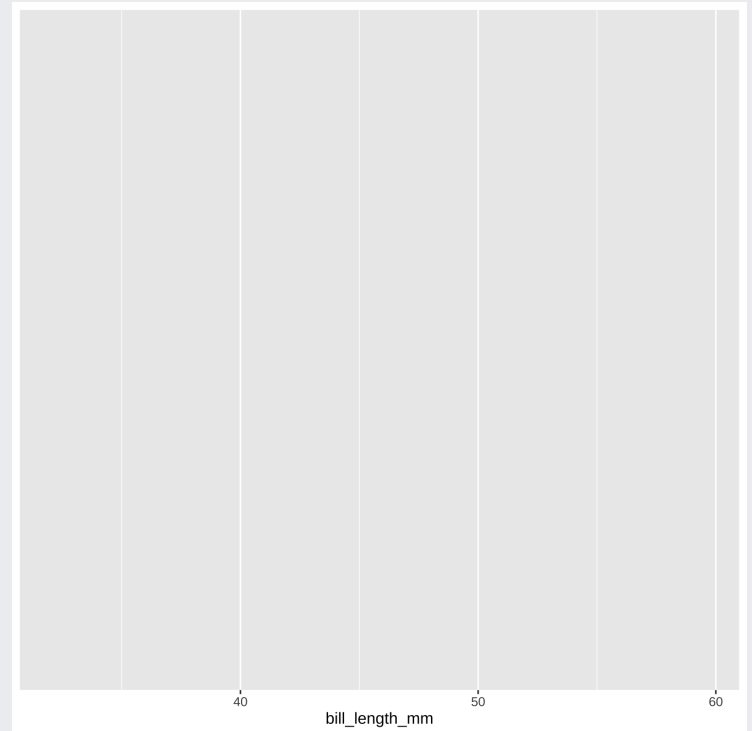
1) Data

```
ggplot(data=penguins)
```



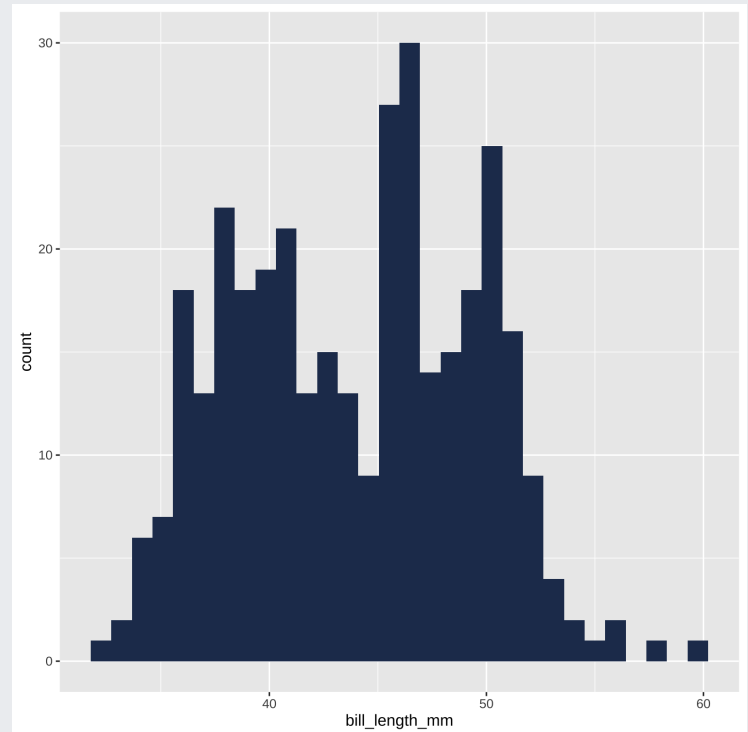
2) Coordinate System

```
ggplot(data=penguins) +  
  aes(x=bill_length_mm)
```



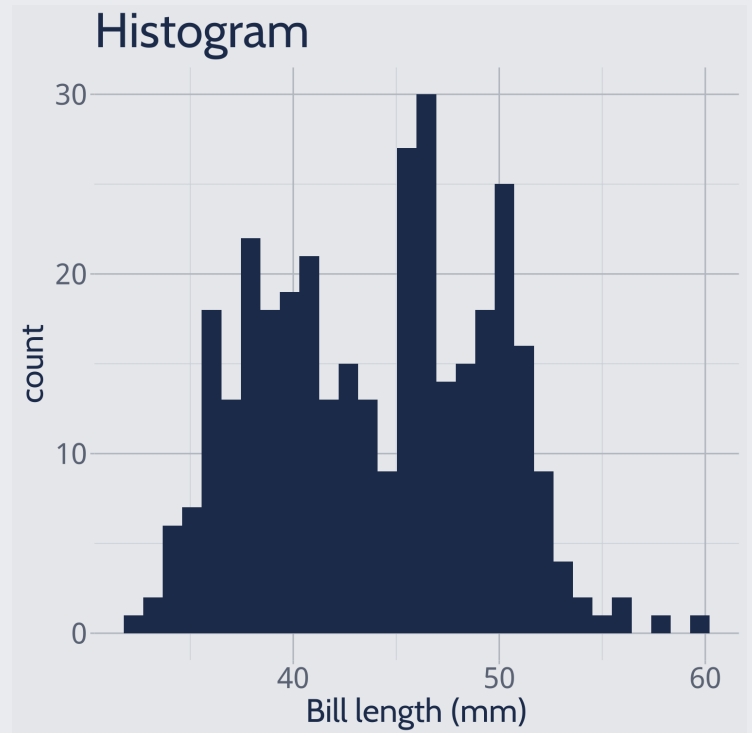
3) Geometry

```
ggplot(data=penguins) +  
  aes(bill_length_mm) +  
  geom_histogram()
```



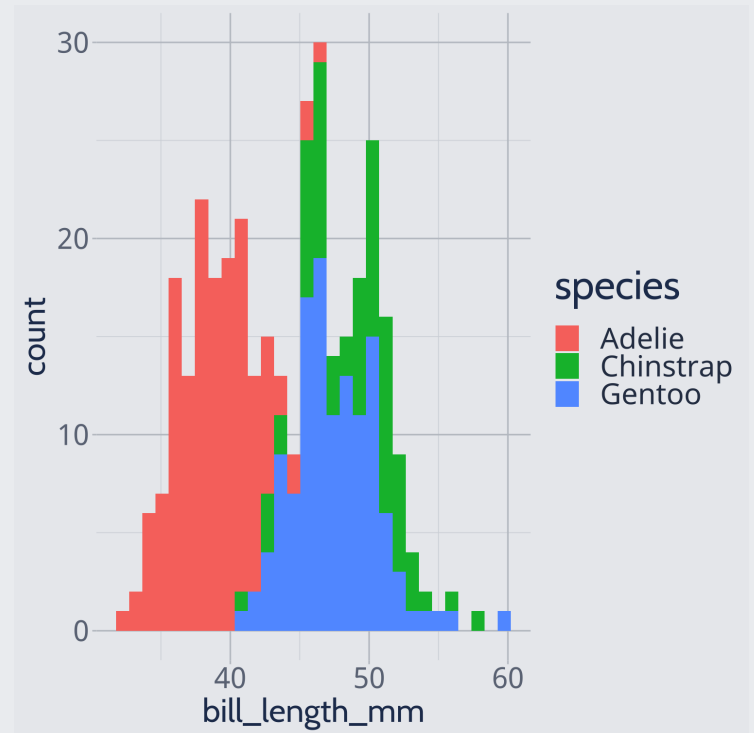
... labeling

```
ggplot(data=penguins) +  
  aes(bill_length_mm) +  
  geom_histogram() +  
  labs(x="Bill length (mm)",  
        title="Histogram") +  
  theme_xaringan()
```



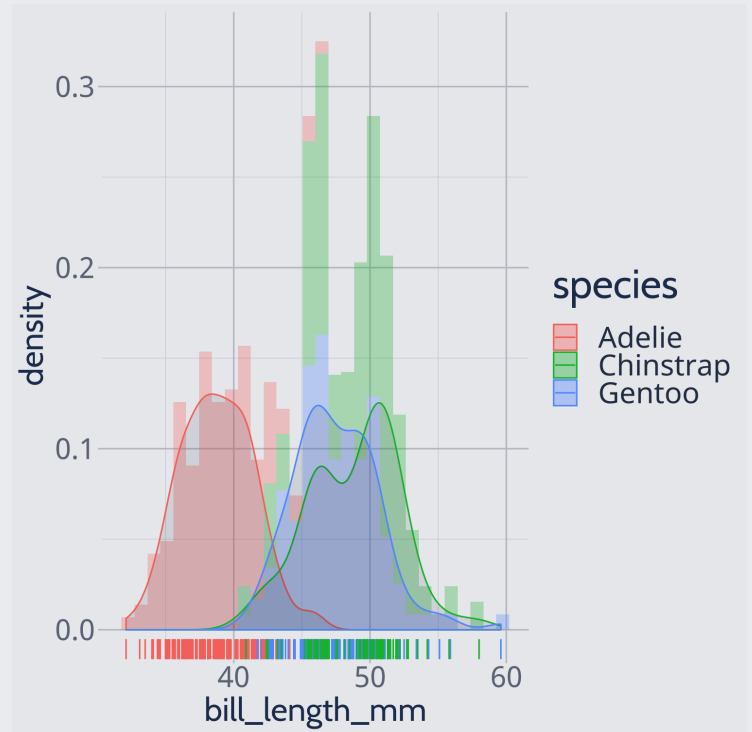
Distinguishing species via colors

```
ggplot(data=penguins) +  
  aes(bill_length_mm) +  
  geom_histogram(  
    aes(fill = species)  
  ) +  
  theme_xaringan()
```



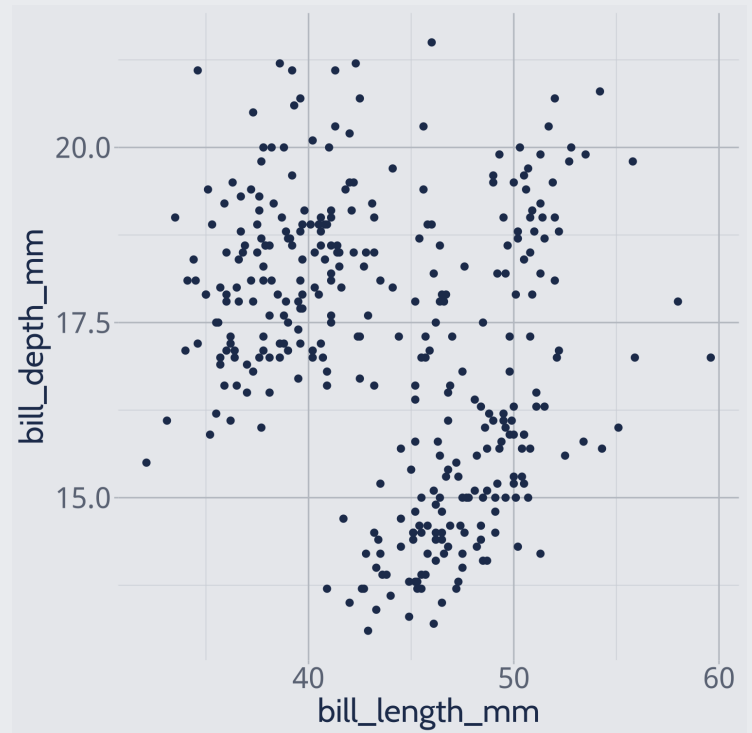
Adding KDE and rug plot

```
ggplot(data=penguins) +  
  aes(bill_length_mm) +  
  geom_histogram(  
    aes(  
      fill = species,  
      y = ..density..  
    ),  
    alpha=0.3  
  ) +  
  geom_density(  
    aes(  
      color = species  
    )  
  ) +  
  geom_rug(aes(color=species)) +  
  theme_xaringan()
```



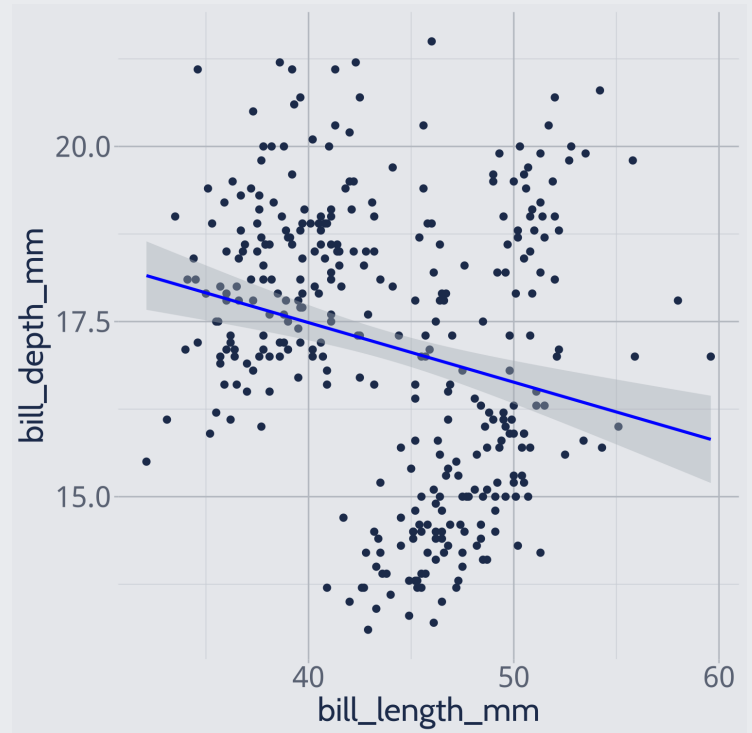
Scatterplot

```
ggplot(data = penguins) +  
  aes(x = bill_length_mm,  
      y = bill_depth_mm) +  
  geom_point(size = 2) +  
  theme_xaringan()
```



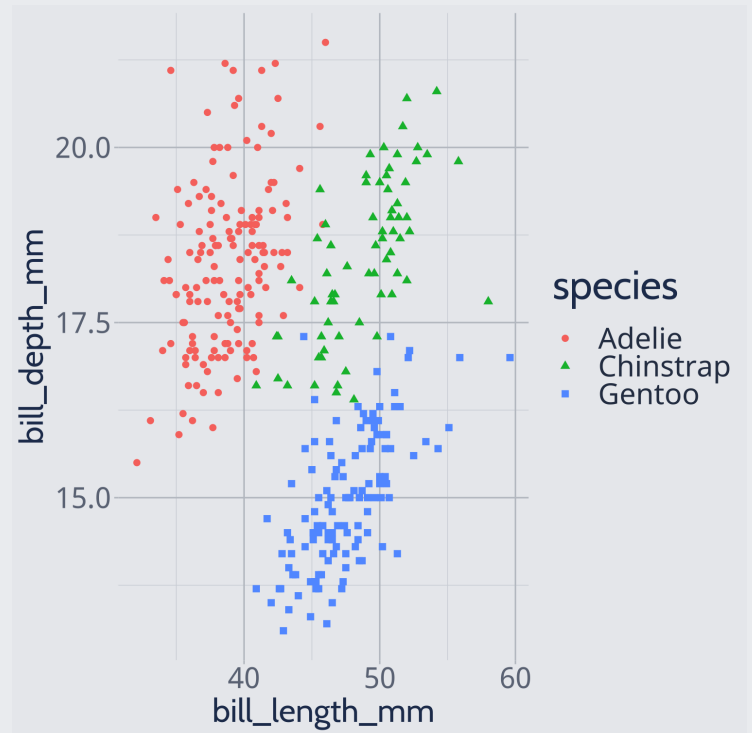
Add regression line

```
ggplot(data = penguins) +  
  aes(x = bill_length_mm,  
      y = bill_depth_mm) +  
  geom_point(size = 2) +  
  geom_smooth(method="lm",  
             color="blue") +  
  theme_xaringan()
```



Distinguish species via colors

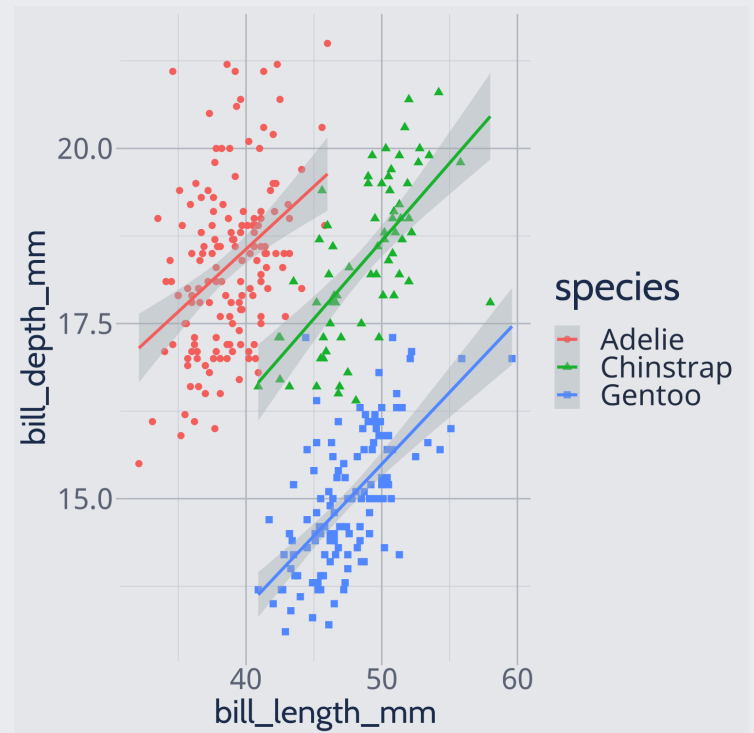
```
ggplot(data = penguins) +  
  aes(x = bill_length_mm,  
      y = bill_depth_mm) +  
  geom_point(  
    aes(color = species,  
        shape = species),  
    size = 2) +  
  theme_xaringan()
```



Add regression lines

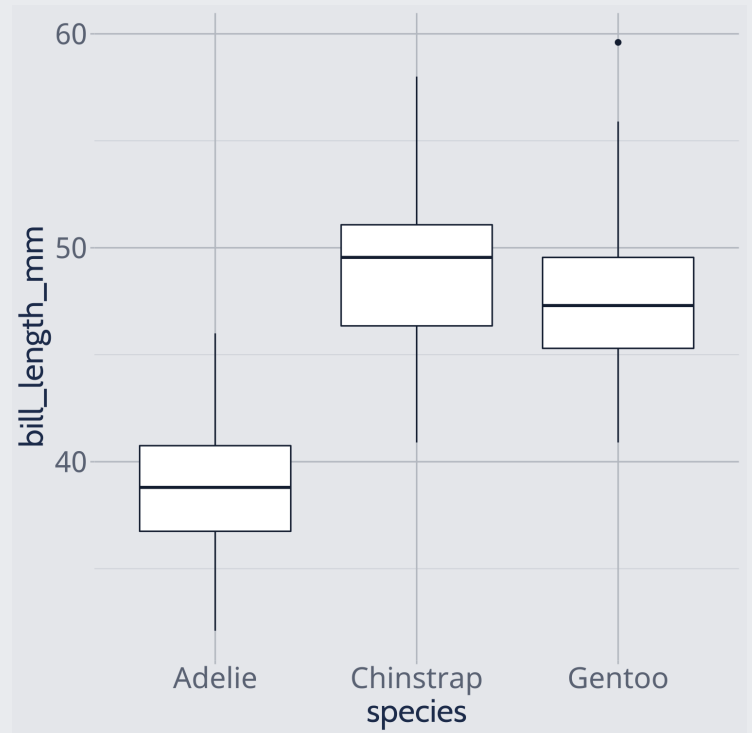
- Simpson's paradox: https://en.wikipedia.org/wiki/Simpson%27s_paradox

```
ggplot(data = penguins) +  
  aes(x = bill_length_mm,  
      y = bill_depth_mm) +  
  geom_point(  
    aes(color = species,  
        shape = species),  
    size = 2  
  ) +  
  geom_smooth(  
    method = "lm",  
    se = TRUE,  
    aes(color = species)  
  ) +  
  theme_xaringan()
```



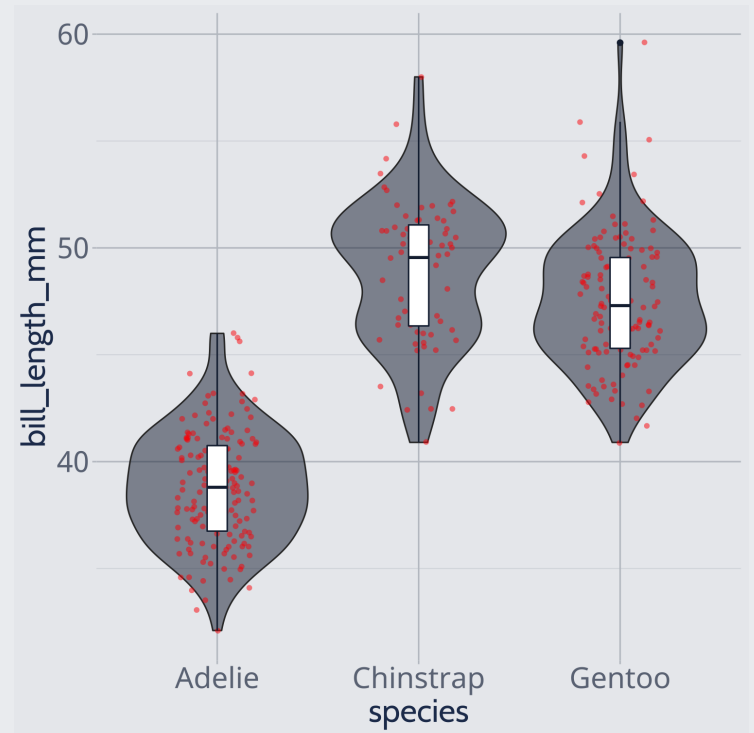
Boxplot

```
ggplot(data = penguins) +  
  aes(x = species,  
      y = bill_length_mm) +  
  geom_boxplot() +  
  theme_xaringan()
```



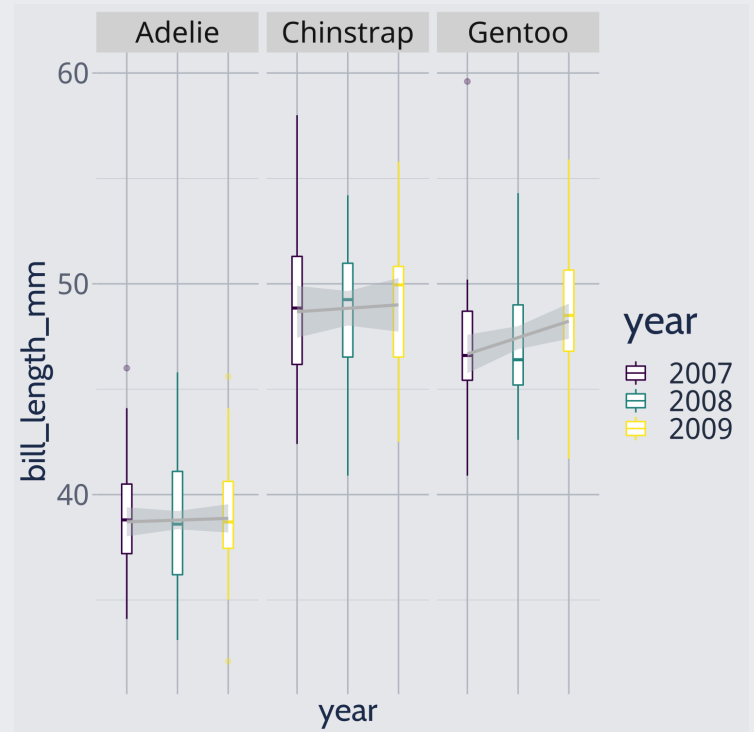
Violinplot

```
ggplot(data = penguins) +  
  aes(x = species,  
      y = bill_length_mm) +  
  geom_violin(alpha=.5) +  
  geom_jitter(  
    shape=16,  
    position=position_jitter(0.2)  
    color="red",  
    alpha=0.5  
  ) +  
  geom_boxplot(width=0.1) +  
  theme_xaringan()
```



Facetting

```
penguins$year <-  
  as.factor(penguins$year)  
ggplot(data = penguins) +  
  aes(x = year,  
      y = bill_length_mm) +  
  geom_boxplot(  
    aes(group=year,  
        color=year),  
    width=0.2,  
    outlier.alpha = 0.3) +  
  geom_smooth(  
    aes(x=as.integer(year),  
        y=bill_length_mm),  
    method="lm",  
    color="grey") +  
  facet_wrap(vars(species)) +  
  theme_xaringan() +  
  scale_colour_viridis_d() +  
  theme(axis.text.x  
        =element_blank())
```



Enough said...

