

Lab 01b: BMI 5/625

Introducing ggplot

Steven Bedrick



R Basics

- R is an interpreter (`>`)
- Name objects in R (`i_like_snake_case ←`) asdf
- Know your object types (`typeof()`)
- Case matters (`my_names ≠ My_names`)
- Use comments! (`# use the hashtag symbol`)
- Functions (`fun!`)
- Use packages (*"install once per machine, load once per R session"*)
- Use the `%>%` (*"dataframe first, dataframe once"*)

Introducing ggplot

- Goal for this session: a very quick `ggplot` refresher
- Meet our dataset:
 - Palmer Penguins

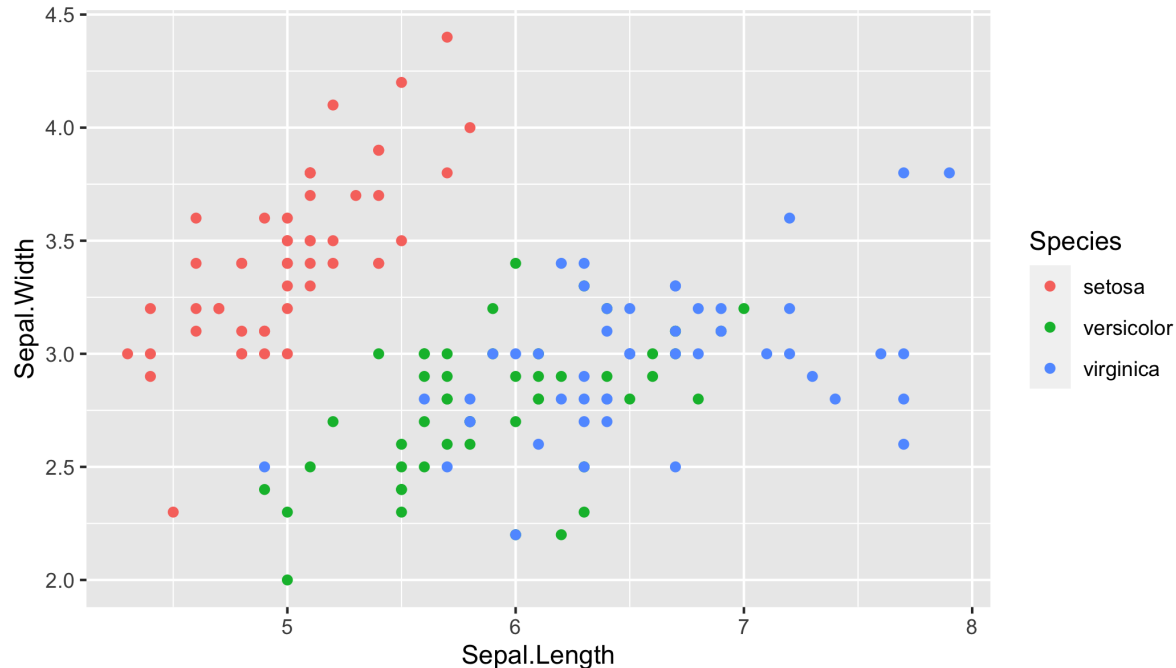
```
glimpse(penguins)
```

```
Rows: 344
```

```
Columns: 8
```

```
$ species      <fct> Adelie, Adelie, Adelie, Adelie, Adelie, Adelie, A  
$ island       <fct> Torgersen, Torgersen, Torgersen, Torgersen, Torge  
$ bill_length_mm <dbl> 39.1, 39.5, 40.3, NA, 36.7, 39.3, 38.9, 39.2, 34.  
$ bill_depth_mm <dbl> 18.7, 17.4, 18.0, NA, 19.3, 20.6, 17.8, 19.6, 18.  
$ flipper_length_mm <int> 181, 186, 195, NA, 193, 190, 181, 195, 193, 190,  
$ body_mass_g   <int> 3750, 3800, 3250, NA, 3450, 3650, 3625, 4675, 347  
$ sex           <fct> male, female, female, NA, female, male, female, m  
$ year          <int> 2007, 2007, 2007, 2007, 2007, 2007, 2007, 2007, 2007, 2
```

Sidebar: Why not Fisher's Irises?



- Hint: The original citation for that data is:
 - R. A. Fisher (1936) "The use of multiple measurements in taxonomic problems." *Annals of Eugenics* 7(2): 179-188

So: Penguins!

```
glimpse(penguins)
```

```
Rows: 344
```

```
Columns: 8
```

```
$ species      <fct> Adelie, Adelie, Adelie, Adelie, Adelie, Adelie, A  
$ island       <fct> Torgersen, Torgersen, Torgersen, Torgersen, Torge  
$ bill_length_mm <dbl> 39.1, 39.5, 40.3, NA, 36.7, 39.3, 38.9, 39.2, 34.  
$ bill_depth_mm <dbl> 18.7, 17.4, 18.0, NA, 19.3, 20.6, 17.8, 19.6, 18.  
$ flipper_length_mm <int> 181, 186, 195, NA, 193, 190, 181, 195, 193, 190,  
$ body_mass_g   <int> 3750, 3800, 3250, NA, 3450, 3650, 3625, 4675, 347  
$ sex           <fct> male, female, female, NA, female, male, female, m  
$ year          <int> 2007, 2007, 2007, 2007, 2007, 2007, 2007, 2007, 2
```

Core ggplot concepts

- *Aesthetics* map data to visual properties of the plot
- *Geometries* ("geoms") actually put "ink on the page"
 - Each type of graph (scatterplot, etc.) has a corresponding geom
 - Different geoms attend to different aesthetics
 - e.g., `geom_point` cares about `x` and `y`
 - Multiple geoms can be combined on the same plot
- *Scales* control axes, fills, etc.
- *Themes* control visual properties (fonts, background colors, etc.)

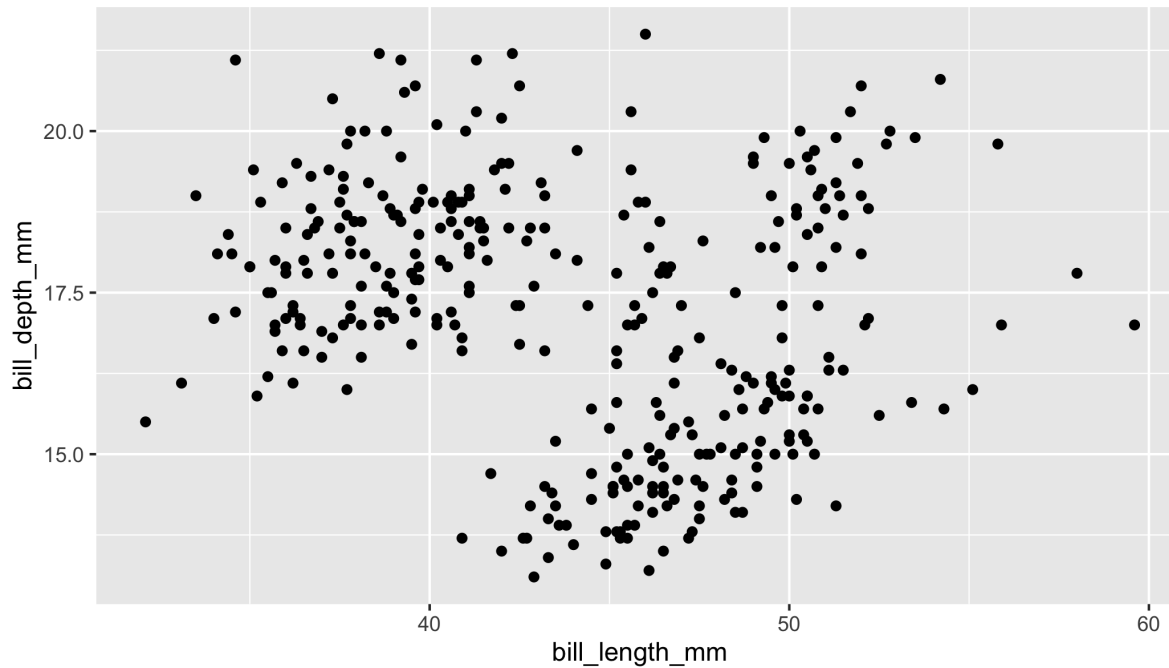
ggplot is an *opinionated* tool!

Data considerations

- By default, `ggplot` assumes "tidy" data
- Think: one "row" per mark on the graph...
- ... and then the various properties for each mark in columns.

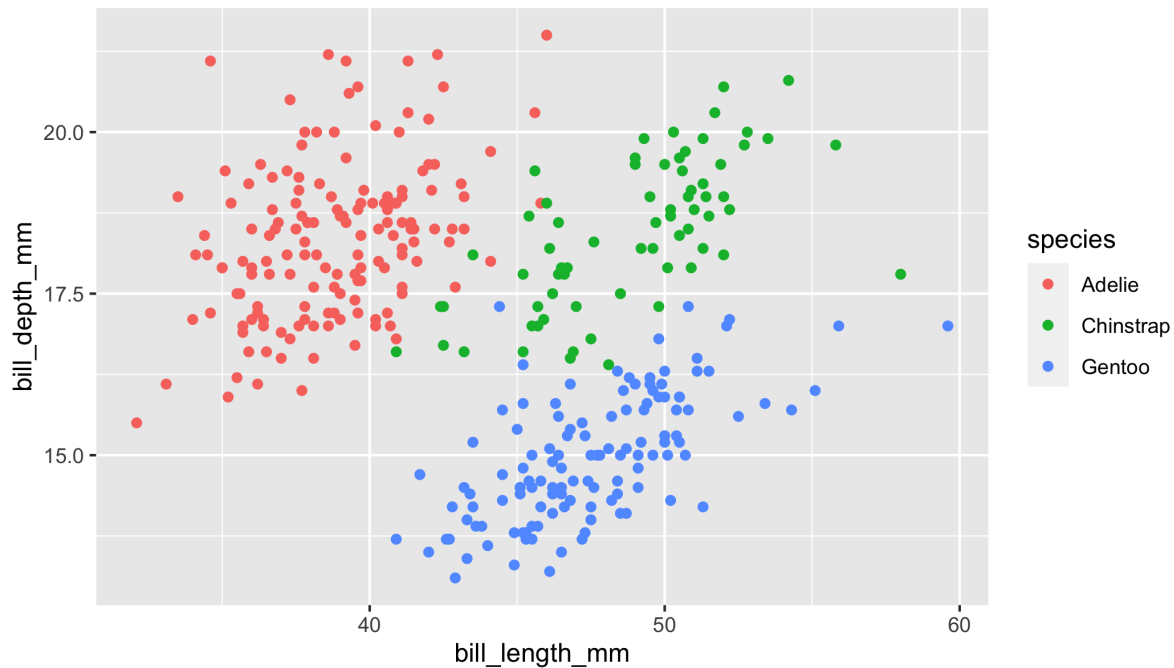
Our first plot

```
penguins %>% ggplot(  
  mapping=aes(x=bill_length_mm, y=bill_depth_mm)  
) + geom_point()
```



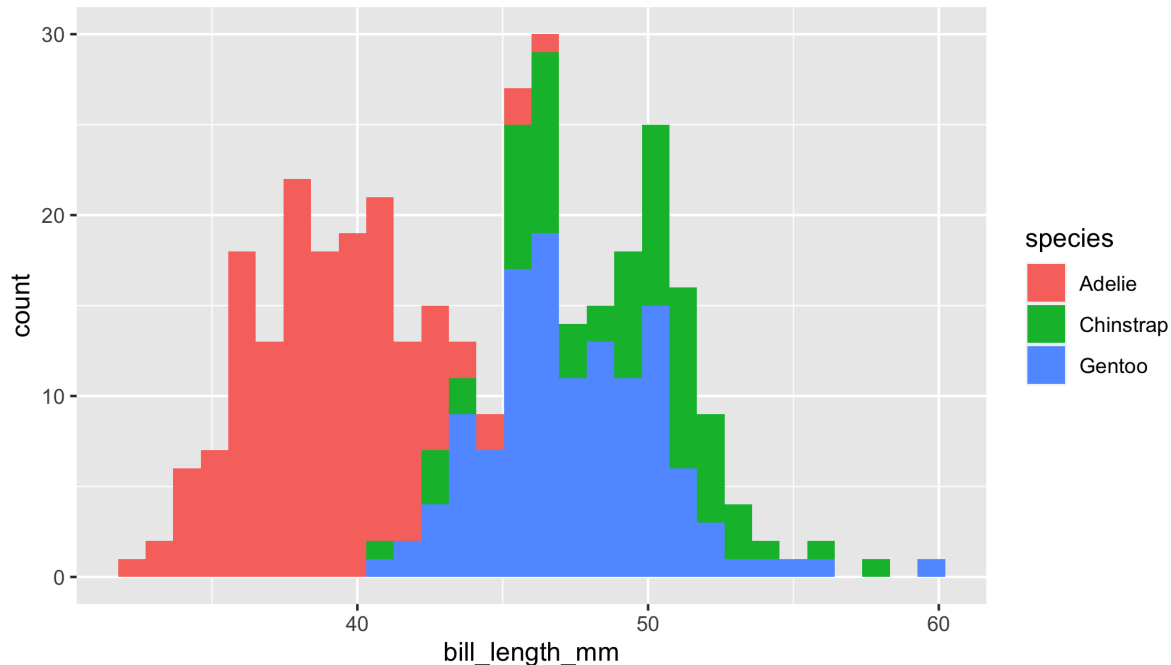
What about color?

```
penguins %>% ggplot(  
  mapping=aes(x=bill_length_mm, y=bill_depth_mm, color=species)  
) + geom_point()
```



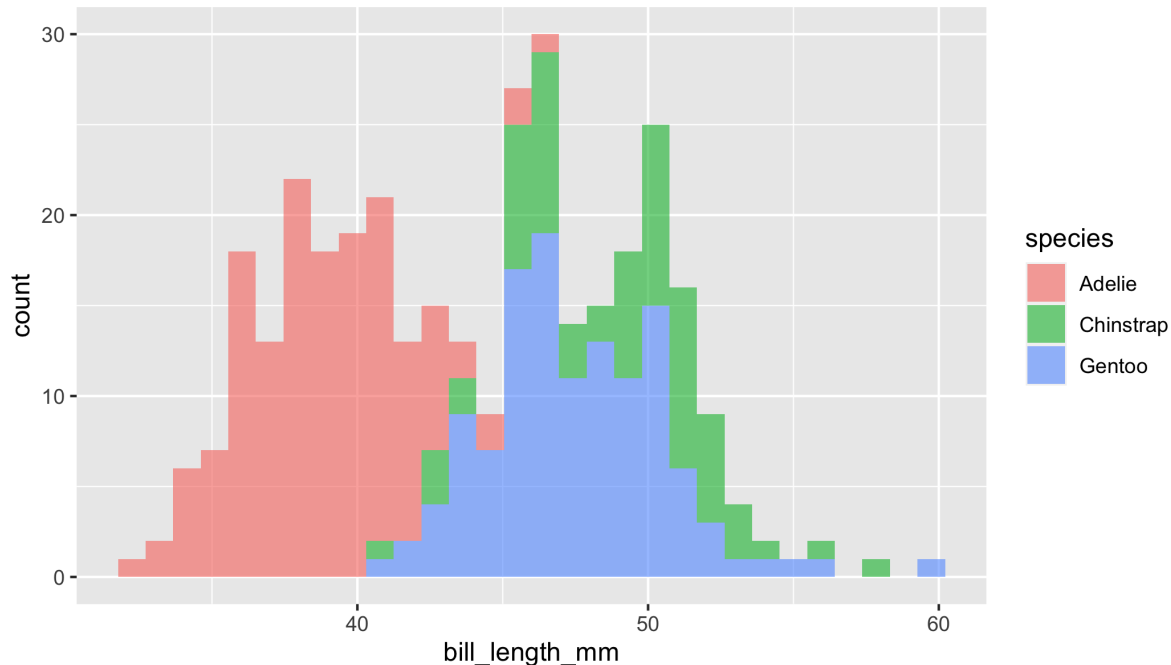
Want a different plot? Try a different geom

```
penguins %>% ggplot(  
  mapping=aes(x=bill_length_mm, fill=species)  
) + geom_histogram()
```



We can specify aesthetic values by hand:

```
penguins %>% ggplot(  
  mapping=aes(x=bill_length_mm, fill=species)  
) + geom_histogram(alpha=0.6)
```



Customizing other aspects of the plot

```
penguins %>% ggplot(  
  mapping=aes(x=bill_length_mm, fill=species)  
) + geom_histogram(alpha=0.6) +  
  labs(x="Bill Length (mm)", y="Frequency") +  
  ggtitle("Bill Length, by species")
```

