ggplot2-intro

Paul M. Magwene

Data frame example

```
## -- Attaching packages --------------------------------------- tidyverse 1.3.1 --
```

```
## v ggplot2 3.3.5     v purrr   0.3.4
## v tibble  3.1.6     v dplyr   1.0.7
## v tidyr   1.1.4     v stringr 1.4.0
## v readr   2.1.1     v forcats 0.5.1
```

```
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
## Loading required package: viridisLite
```

```
head(iris)
```

```
##   Sepal.Length Sepal.Width Petal.Length Petal.Width Species
## 1          5.1         3.5          1.4         0.2  setosa
## 2          4.9         3.0          1.4         0.2  setosa
## 3          4.7         3.2          1.3         0.2  setosa
## 4          4.6         3.1          1.5         0.2  setosa
## 5          5.0         3.6          1.4         0.2  setosa
## 6          5.4         3.9          1.7         0.4  setosa
```

Data + Geometric Representation + Aesthetic qualities

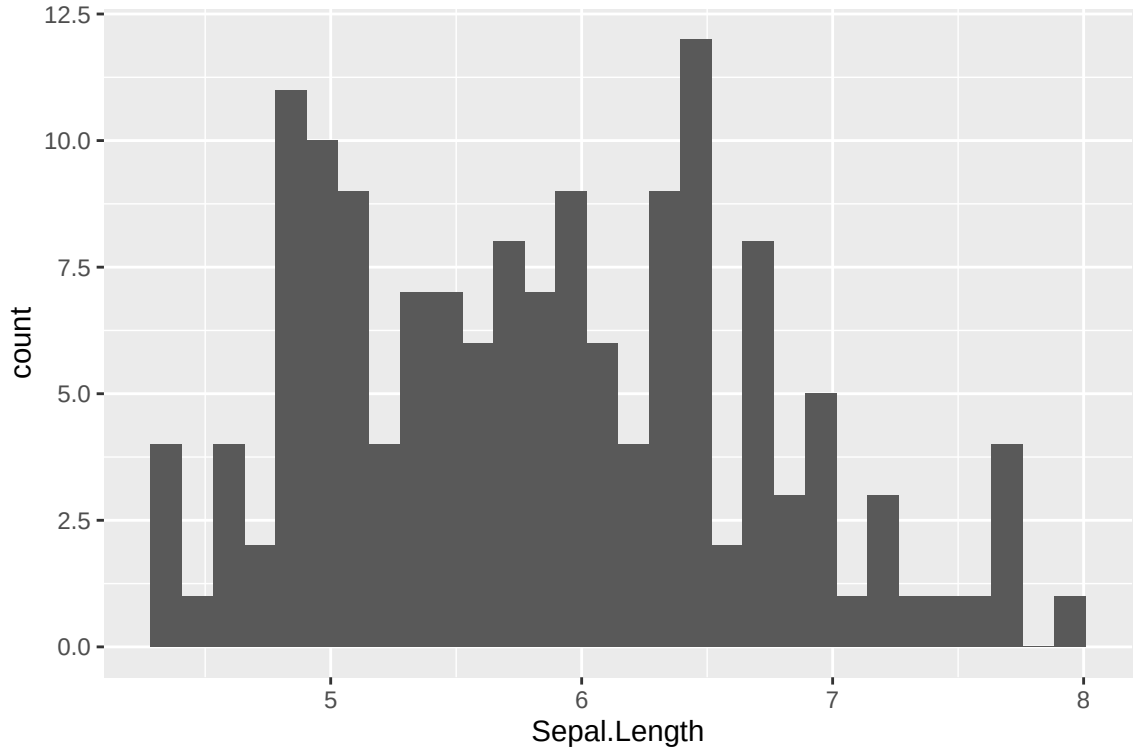
```
library(ggplot2) # import ggplot2 functions

ggplot(data = <DATA>) + <GEOM_FUNCTION>(mapping = aes(<MAPPINGS>))
```

Example 1

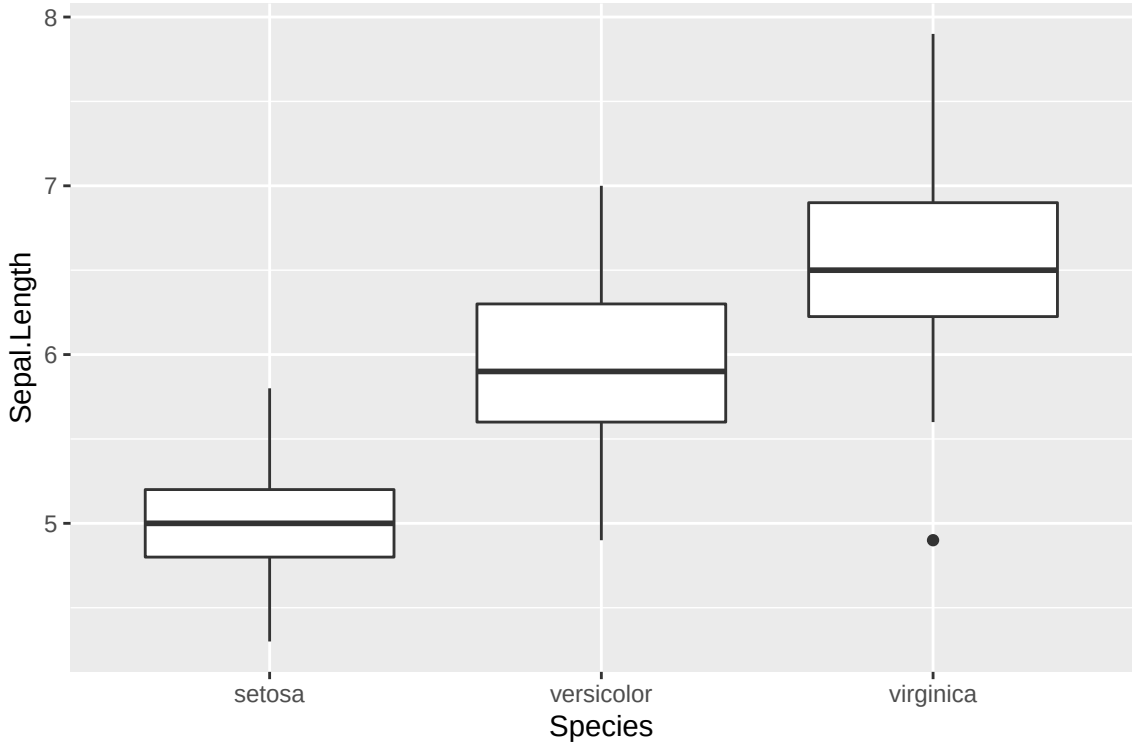
```
ggplot(data = iris) + geom_histogram(mapping = aes(x = Sepal.Length))
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```



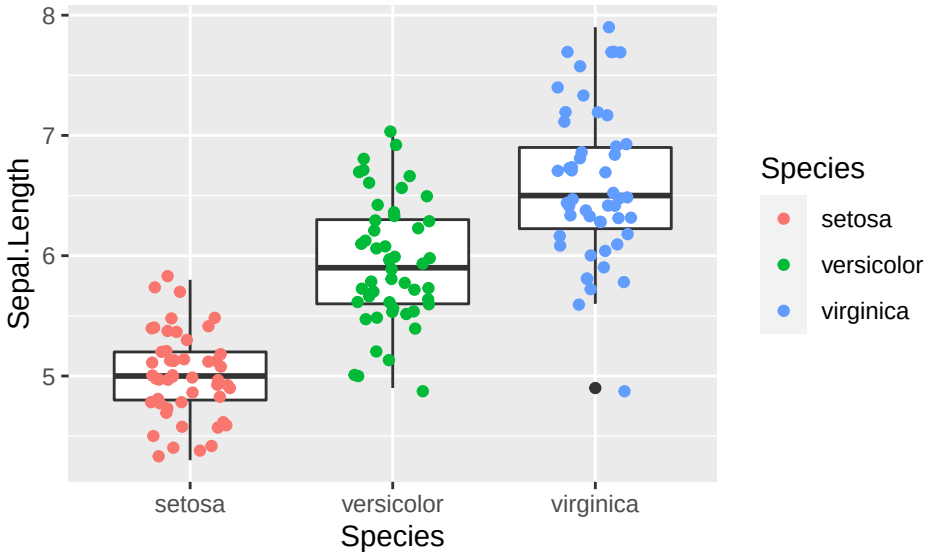
Example 2

```
ggplot(data = iris) + geom_boxplot(mapping = aes(x = Species,
                                                 y = Sepal.Length))
```



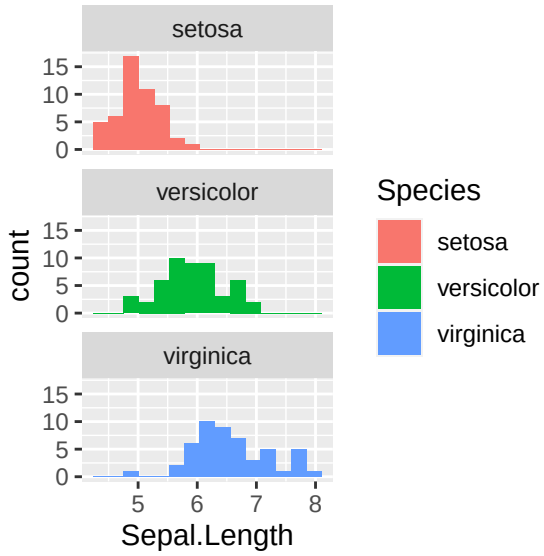
Example 3

```
ggplot(data = iris) + geom_boxplot(mapping = aes(x = Species,
                                                 y = Sepal.Length)) +
                      geom_jitter(mapping = aes(x = Species,
                                                y = Sepal.Length,
                                                color = Species),
                                  width = 0.2)
```



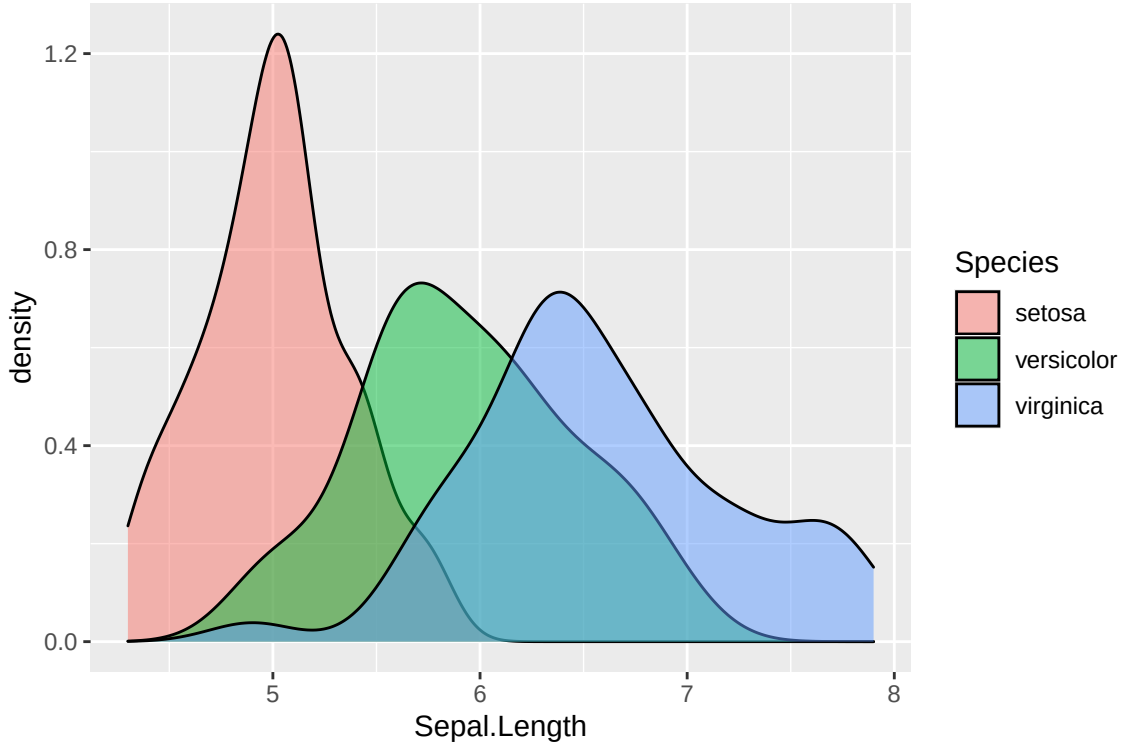
Faceting (conditioning)

```
ggplot(data = iris, aes(x = Sepal.Length, fill=Species)) +
  geom_histogram(bins=15) + facet_wrap(~Species,ncol=1)
```



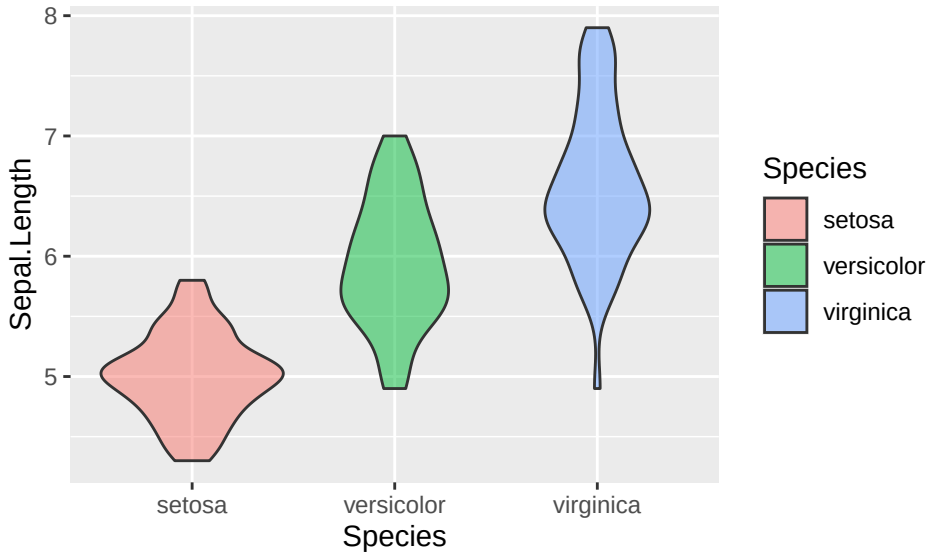
Tour of useful geoms: Density plots

```
ggplot(data = iris, aes(x = Sepal.Length, fill=Species)) +
  geom_density(alpha=0.5)
```



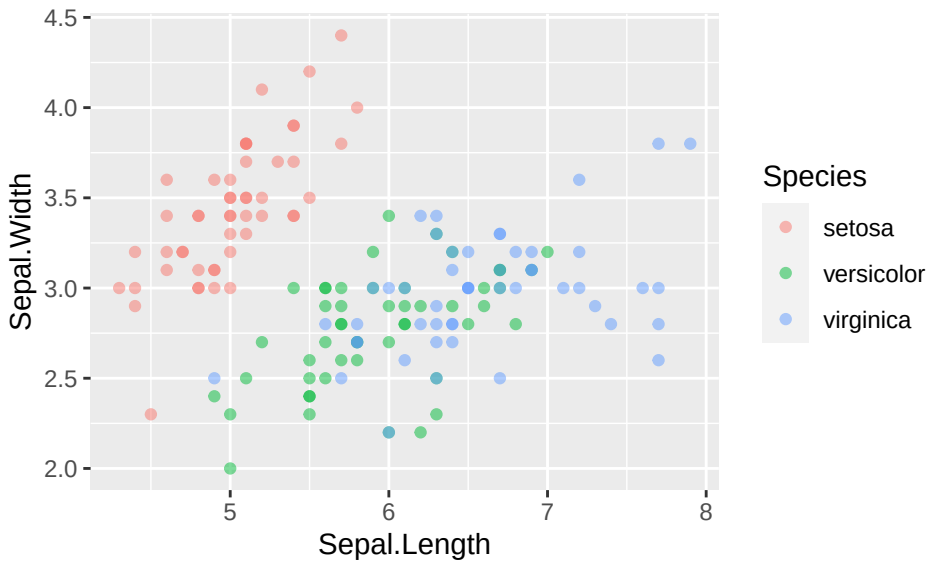
Tour of useful geoms: Violin plots

```
ggplot(data = iris, aes(x = Species,
                        y = Sepal.Length,
                        fill=Species)) +
  geom_violin(alpha=0.5)
```



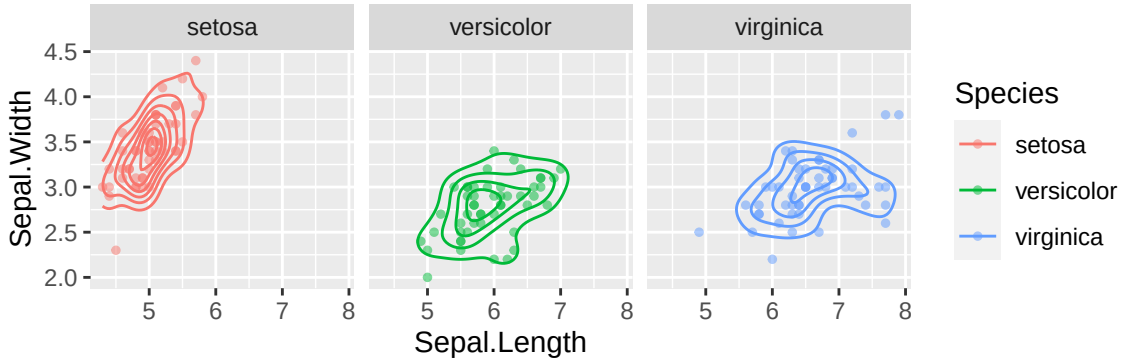
Tour of useful geoms: Scatter plots

```
ggplot(data = iris, aes(x = Sepal.Length,
                        y = Sepal.Width,
                        color=Species)) +
  geom_point(alpha=0.5)
```



Tour of useful geoms: 2D density plots

```
ggplot(data = iris, aes(x = Sepal.Length,
                        y = Sepal.Width,
                        color=Species)) +
  geom_density_2d() +
  geom_point(alpha=0.5,size=1) +
  facet_wrap(~Species)
```



Tour of useful geoms: Line plot

```
## Rows: 756 Columns: 8
```

```
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr  (1): geoid
## dbl  (6): cases, cases_avg, cases_avg_per_100k, deaths, deaths_avg, deaths_a...
```

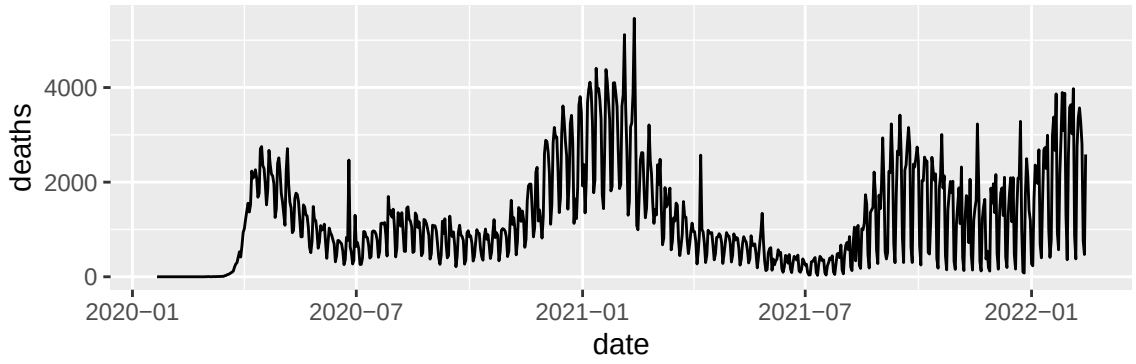
```
## date (1): date
```

```
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```
## Rows: 756
## Columns: 8
## $ date                <date> 2020-01-21, 2020-01-22, 2020-01-23, 2020-01-24, 2~
## $ geoid               <chr> "USA", "USA", "USA", "USA", "USA", "USA", "USA", "~
## $ cases               <dbl> 1, 0, 0, 1, 1, 2, 0, 0, 0, 1, 1, 1, 3, 0, 0, 1, 0,~
## $ cases_avg           <dbl> 0.14, 0.14, 0.14, 0.29, 0.43, 0.71, 0.71, 0.57, 0.~
## $ cases_avg_per_100k  <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,~
## $ deaths              <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,~
## $ deaths_avg          <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,~
## $ deaths_avg_per_100k <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,~
```

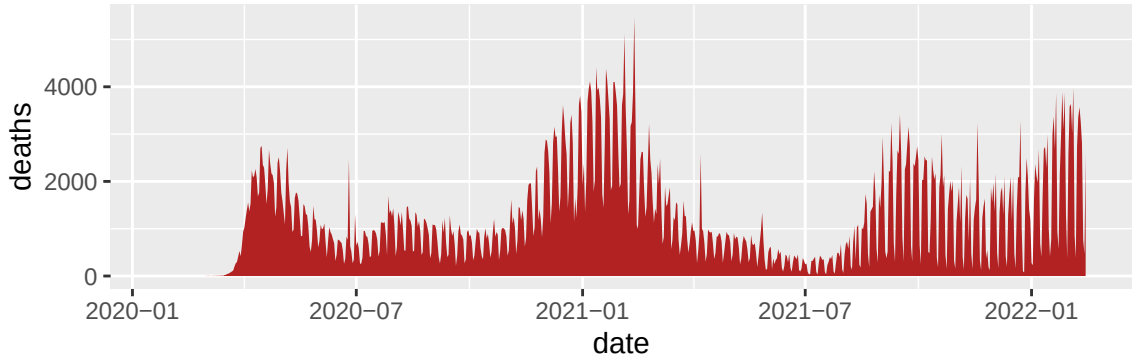
Tour of useful geoms: Line plot

```
ggplot(data = covid, aes(x = date, y = deaths)) +
  geom_line()
```

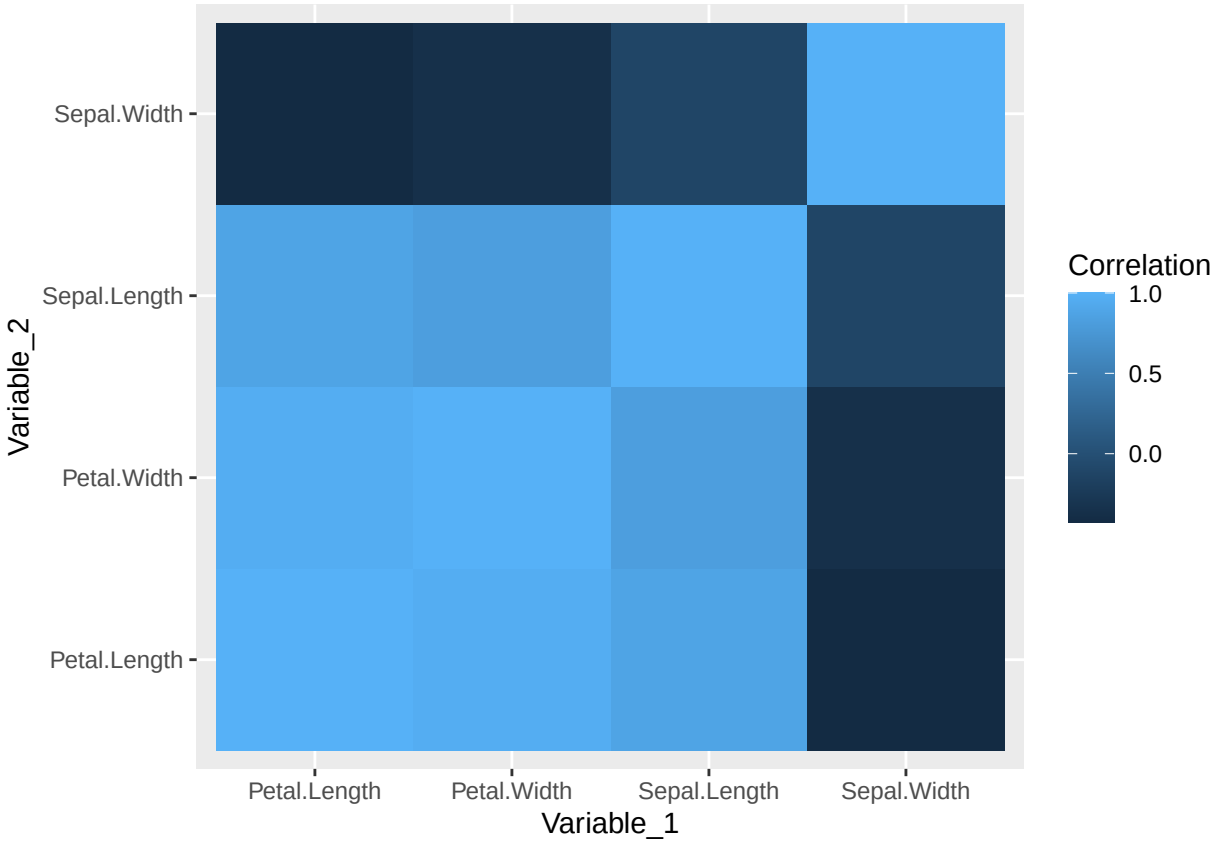


Tour of useful geoms: Area plot

```
ggplot(data = covid, aes(x = date, y = deaths)) +
  geom_area(fill="firebrick")
```



Tour of useful geoms: Heat maps




```
us_states <- read_csv("~/Desktop/us-states.csv")
```

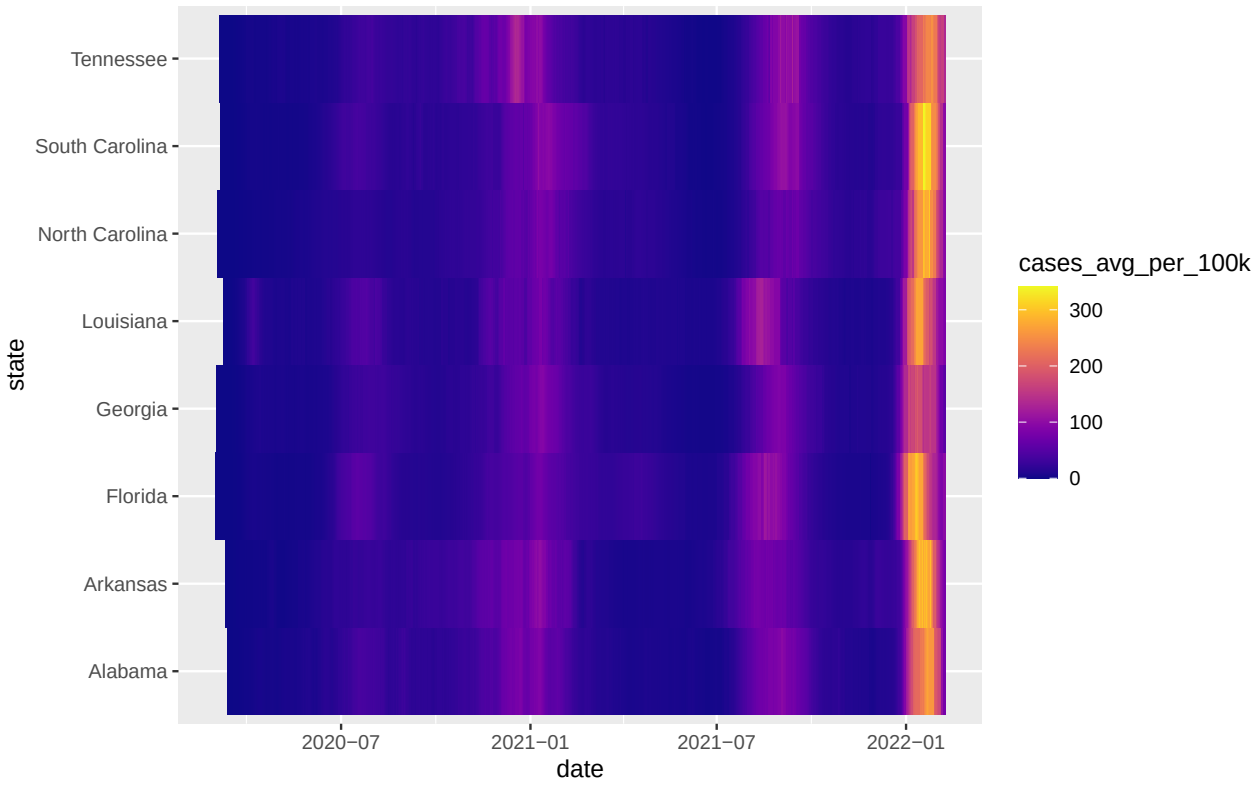
```
## Rows: 39094 Columns: 9
```

```
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr  (2): geoid, state
## dbl  (6): cases, cases_avg, cases_avg_per_100k, deaths, deaths_avg, deaths_a...
## date (1): date
```

```
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```
SE_states <- c("North Carolina","South Carolina",
               "Arkansas","Georgia","Tennessee",
               "Louisiana","Alabama","Florida")

us_states %>%
  filter(state %in% SE_states) %>%
  arrange(state) %>%
  ggplot(aes(x=date, y=state, fill=cases_avg_per_100k)) +
    geom_tile() +
    scale_fill_viridis(option = "C")
```


Tennessee
South Carolina
North Carolina
Louisiana
Georgia
Florida
Arkansas
Alabama
state
2020−07
2021−01
2021−07
2022−01
date
cases_avg_per_100k
300
200
100
0