Clustering Example 2: Iris Data Visualization

Amy Wagaman, Nathan Carter

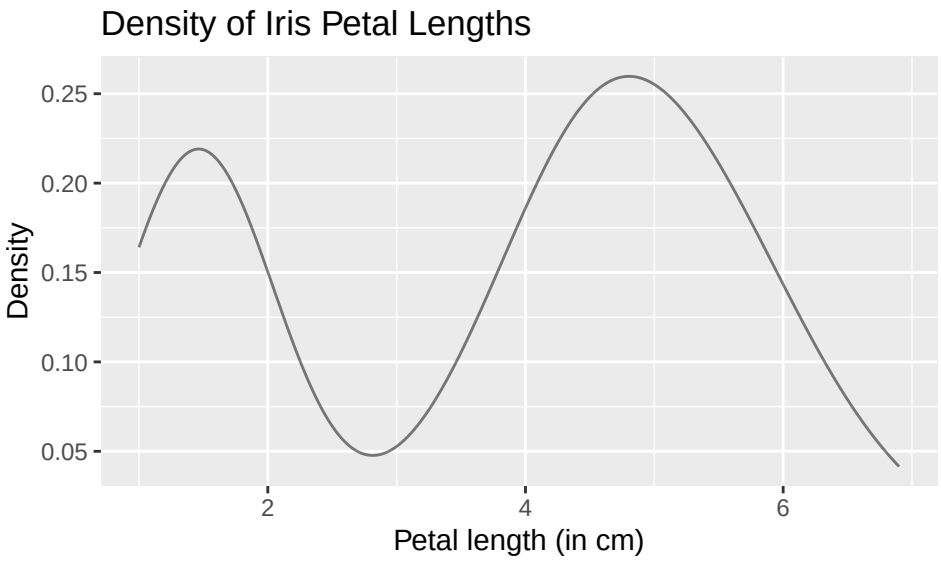
July 2020

Load necessary libraries.

```
library(dplyr)
library(ggformula)
library(GGally)
```

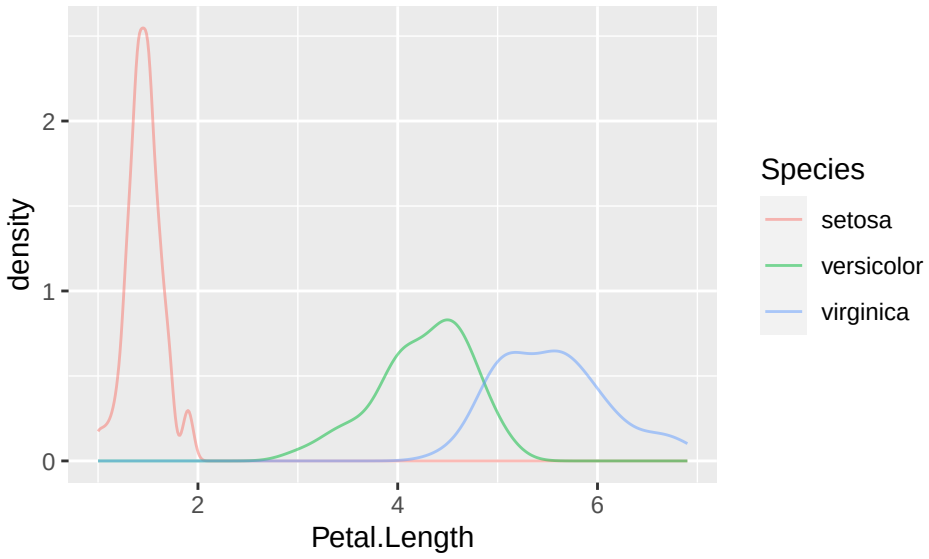
Create a density plot of petal lengths

```
data(iris)

gf_dens(~ Petal.Length, data = iris) %>%
  gf_labs(title = "Density of Iris Petal Lengths",
          x = "Petal length (in cm)", y = "Density")
```



Separate density plot by species

```
gf_dens(~ Petal.Length, color = ~ Species, data = iris)
```

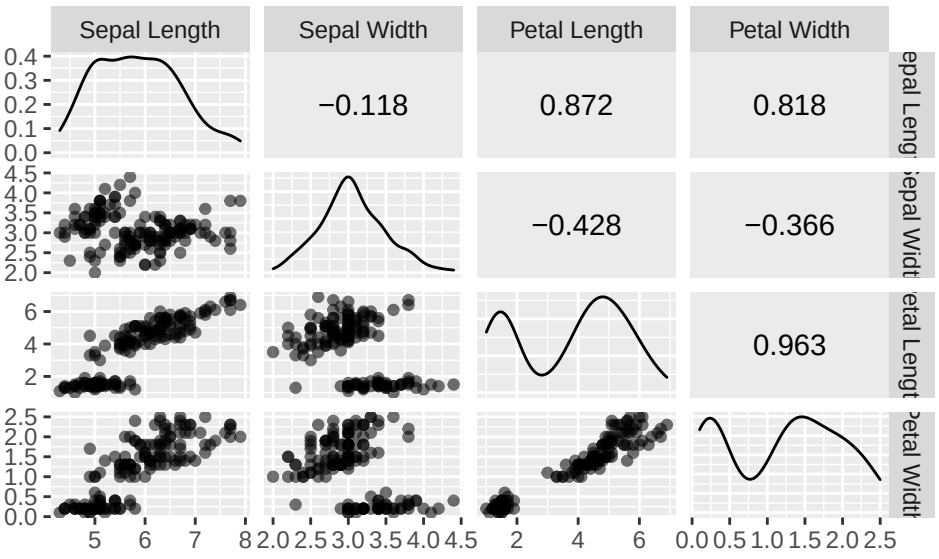


Create a function for annotating the upper triangle of a pair plot with correlations

```
cor_fun <- function (data, mapping, method="pearson", ndp=3, sz=5, ...) {

    x <- eval_data_col(data, mapping$x)
    y <- eval_data_col(data, mapping$y)

    corr <- cor.test(x, y, method=method)
    est <- corr$estimate
    lbl <- round(est, ndp)

    ggplot(data=data, mapping=mapping) +
      annotate("text", x=mean(x, na.rm=TRUE), y=mean(y, na.rm=TRUE),label=lbl, ...)+
      theme(panel.grid = element_blank())
}
```

Create a pair plot (using the function above)

```
ggpairs(iris, columns = 1:4, mapping = ggplot2::aes(alpha = 0.3),
        columnLabels = c("Sepal Length", "Sepal Width", "Petal Length", "Petal Width"),
        upper=list(continuous=cor_fun))
```


Sepal Length
Sepal Width
Petal Length
Petal Width
−0.118
0.872
0.818
−0.428
−0.366
0.963