ggplot\_1

Daniel\_Kim

2019 8 15

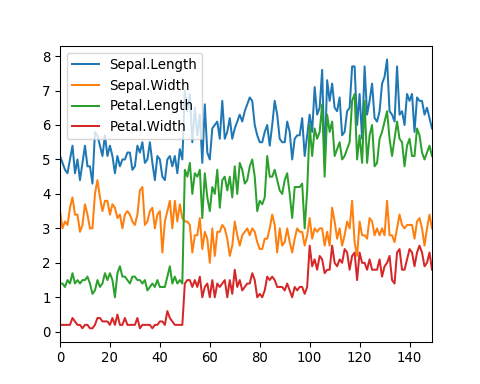
summary(iris)

## Sepal.Length Sepal.Width Petal.Length Petal.Width   
## Min. :4.300 Min. :2.000 Min. :1.000 Min. :0.100   
## 1st Qu.:5.100 1st Qu.:2.800 1st Qu.:1.600 1st Qu.:0.300   
## Median :5.800 Median :3.000 Median :4.350 Median :1.300   
## Mean :5.843 Mean :3.057 Mean :3.758 Mean :1.199   
## 3rd Qu.:6.400 3rd Qu.:3.300 3rd Qu.:5.100 3rd Qu.:1.800   
## Max. :7.900 Max. :4.400 Max. :6.900 Max. :2.500   
## Species   
## setosa :50   
## versicolor:50   
## virginica :50   
##   
##   
##

iris %>% as\_tibble() %>% head()

## # A tibble: 6 x 5  
## Sepal.Length Sepal.Width Petal.Length Petal.Width Species  
## <dbl> <dbl> <dbl> <dbl> <fct>   
## 1 5.1 3.5 1.4 0.2 setosa   
## 2 4.9 3 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 3.6 1.4 0.2 setosa   
## 6 5.4 3.9 1.7 0.4 setosa

import numpy as np  
import pandas as pd  
  
iris = r.iris  
iris.plot()



iris %>%  
 as\_tibble() %>%  
 rowid\_to\_column() %>%  
 gather(Sepal.Length, Sepal.Width, Petal.Length, Petal.Width, key = "Featrures", value = "value") %>%  
 head()

## # A tibble: 6 x 4  
## rowid Species Featrures value  
## <int> <fct> <chr> <dbl>  
## 1 1 setosa Sepal.Length 5.1  
## 2 2 setosa Sepal.Length 4.9  
## 3 3 setosa Sepal.Length 4.7  
## 4 4 setosa Sepal.Length 4.6  
## 5 5 setosa Sepal.Length 5   
## 6 6 setosa Sepal.Length 5.4

iris %>%  
 as\_tibble() %>%  
 rowid\_to\_column() %>%  
 gather(Sepal.Length, Sepal.Width, Petal.Length, Petal.Width, key = "Features", value = "value") %>%  
 separate(Features, sep = "\\.", into = c("Part", "Measure")) %>%  
 spread(Measure, value) %>%  
 select(-rowid) -> iris\_wide  
  
iris\_wide %>% head()

## # A tibble: 6 x 4  
## Species Part Length Width  
## <fct> <chr> <dbl> <dbl>  
## 1 setosa Petal 1.4 0.2  
## 2 setosa Sepal 5.1 3.5  
## 3 setosa Petal 1.4 0.2  
## 4 setosa Sepal 4.9 3   
## 5 setosa Petal 1.3 0.2  
## 6 setosa Sepal 4.7 3.2

iris$Flower <- 1:nrow(iris)  
  
iris %>%  
 gather(key, value, -Species, -Flower) %>%  
 separate(key, c("Part", "Measure"), "\\.") %>%  
 spread(Measure, value) -> iris.wide  
  
iris.wide %>% head()

## Species Flower Part Length Width  
## 1 setosa 1 Petal 1.4 0.2  
## 2 setosa 1 Sepal 5.1 3.5  
## 3 setosa 2 Petal 1.4 0.2  
## 4 setosa 2 Sepal 4.9 3.0  
## 5 setosa 3 Petal 1.3 0.2  
## 6 setosa 3 Sepal 4.7 3.2

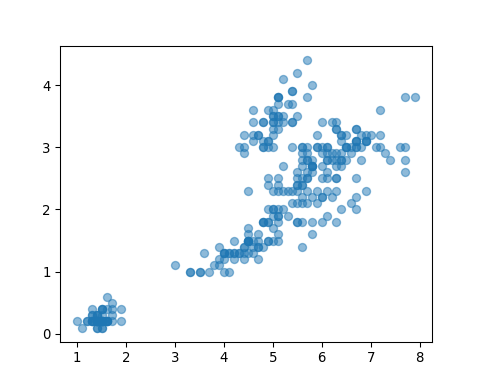
iris\_wide = r.iris\_wide  
iris\_wide['Part'].value\_counts()

## Petal 150  
## Sepal 150  
## Name: Part, dtype: int64

iris\_wide.head()

## Species Part Length Width  
## 0 setosa Petal 1.4 0.2  
## 1 setosa Sepal 5.1 3.5  
## 2 setosa Petal 1.4 0.2  
## 3 setosa Sepal 4.9 3.0  
## 4 setosa Petal 1.3 0.2

import matplotlib.pyplot as plt  
plt.scatter(iris\_wide['Length'], iris\_wide['Width'], alpha=.5)



iris %>%  
 rowid\_to\_column() %>%  
 gather(Sepal.Length, Sepal.Width, Petal.Length, Petal.Width, key = "Features", value = "value") %>%  
 separate(Features, sep = "\\.", into = c("Part", "Measure")) %>%  
 spread(Species, value) -> iris\_tmp  
  
iris\_tmp %>% filter(!is.na(setosa)) %>% select(-versicolor, -virginica) -> iris\_tmp\_1  
iris\_tmp %>% filter(!is.na(versicolor)) %>% select(versicolor) -> iris\_tmp\_2  
iris\_tmp %>% filter(!is.na(virginica)) %>% select(virginica) -> iris\_tmp\_3  
  
iris\_tmp\_1 %>%  
 bind\_cols(iris\_tmp\_2, iris\_tmp\_3) %>%  
 select(-rowid) -> iris\_wide\_2  
  
iris\_wide\_2 %>% head()

## Flower Part Measure setosa versicolor virginica  
## 1 1 Petal Length 1.4 4.7 6.0  
## 2 1 Petal Width 0.2 1.4 2.5  
## 3 1 Sepal Length 5.1 7.0 6.3  
## 4 1 Sepal Width 3.5 3.2 3.3  
## 5 2 Petal Length 1.4 4.5 5.1  
## 6 2 Petal Width 0.2 1.5 1.9

iris %>%  
 rowid\_to\_column() %>%  
 gather(Sepal.Length, Sepal.Width, Petal.Length, Petal.Width, key = "Features", value = "value") %>%  
 separate(Features, sep = "\\.", into = c("Part", "Measure")) %>%  
 select(-rowid) -> iris\_tidy  
  
iris\_tidy %>% head()

## Species Flower Part Measure value  
## 1 setosa 1 Sepal Length 5.1  
## 2 setosa 2 Sepal Length 4.9  
## 3 setosa 3 Sepal Length 4.7  
## 4 setosa 4 Sepal Length 4.6  
## 5 setosa 5 Sepal Length 5.0  
## 6 setosa 6 Sepal Length 5.4

iris %>%  
 gather(key, Value, -Species) %>%  
 separate(key, c("Part", "Measure"), "\\.") -> iris\_tidy; iris\_tidy %>% head()

## Warning: Expected 2 pieces. Missing pieces filled with `NA` in 150 rows  
## [601, 602, 603, 604, 605, 606, 607, 608, 609, 610, 611, 612, 613, 614, 615,  
## 616, 617, 618, 619, 620, ...].

## Species Part Measure Value  
## 1 setosa Sepal Length 5.1  
## 2 setosa Sepal Length 4.9  
## 3 setosa Sepal Length 4.7  
## 4 setosa Sepal Length 4.6  
## 5 setosa Sepal Length 5.0  
## 6 setosa Sepal Length 5.4

str(iris)

## 'data.frame': 150 obs. of 6 variables:  
## $ Sepal.Length: num 5.1 4.9 4.7 4.6 5 5.4 4.6 5 4.4 4.9 ...  
## $ Sepal.Width : num 3.5 3 3.2 3.1 3.6 3.9 3.4 3.4 2.9 3.1 ...  
## $ Petal.Length: num 1.4 1.4 1.3 1.5 1.4 1.7 1.4 1.5 1.4 1.5 ...  
## $ Petal.Width : num 0.2 0.2 0.2 0.2 0.2 0.4 0.3 0.2 0.2 0.1 ...  
## $ Species : Factor w/ 3 levels "setosa","versicolor",..: 1 1 1 1 1 1 1 1 1 1 ...  
## $ Flower : int 1 2 3 4 5 6 7 8 9 10 ...

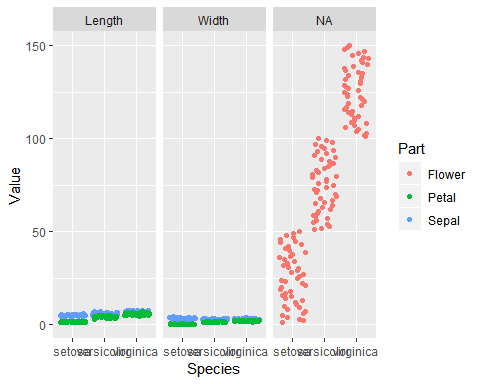
str(iris\_wide)

## Classes 'tbl\_df', 'tbl' and 'data.frame': 300 obs. of 4 variables:  
## $ Species: Factor w/ 3 levels "setosa","versicolor",..: 1 1 1 1 1 1 1 1 1 1 ...  
## $ Part : chr "Petal" "Sepal" "Petal" "Sepal" ...  
## $ Length : num 1.4 5.1 1.4 4.9 1.3 4.7 1.5 4.6 1.4 5 ...  
## $ Width : num 0.2 3.5 0.2 3 0.2 3.2 0.2 3.1 0.2 3.6 ...

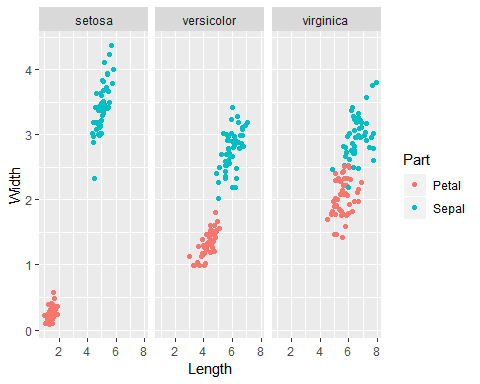
str(iris\_tidy)

## 'data.frame': 750 obs. of 4 variables:  
## $ Species: Factor w/ 3 levels "setosa","versicolor",..: 1 1 1 1 1 1 1 1 1 1 ...  
## $ Part : chr "Sepal" "Sepal" "Sepal" "Sepal" ...  
## $ Measure: chr "Length" "Length" "Length" "Length" ...  
## $ Value : num 5.1 4.9 4.7 4.6 5 5.4 4.6 5 4.4 4.9 ...

ggplot(iris\_tidy, aes(x = Species, y = Value, col = Part)) +   
 geom\_jitter() +   
 facet\_grid(. ~ Measure)



ggplot(iris\_wide, aes(Length, Width, color = Part)) + geom\_jitter() + facet\_grid(. ~ Species)



ggplot(mtcars, aes(x= wt, y = mpg, fill = cyl)) + geom\_point(shape = 21, size = 4, alpha = .6)

