

iris

Rakha Hafish Setiawan

2022-08-15

```
skim(iris)
```

Data summary

Name iris
Number of rows 150
Number of columns 5

Column type frequency:

factor 1
numeric 4

Group variables None

Variable type: factor

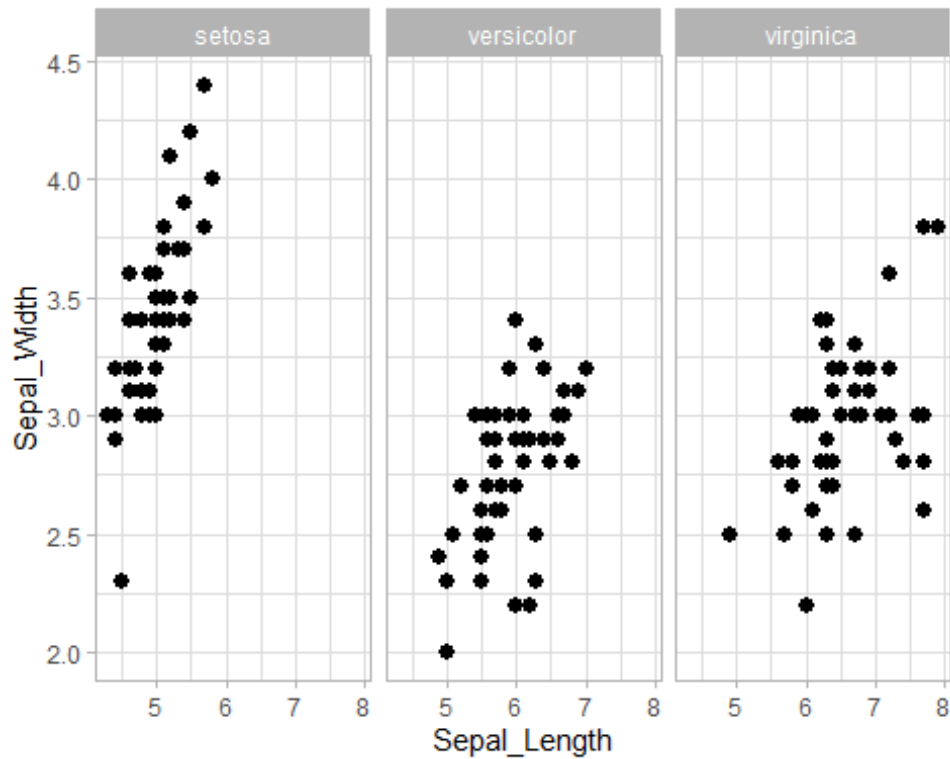
skim_variable	n_missing	complete_rate	ordered	n_unique	top_counts
Species	0	1	FALSE	3	set: 50, ver: 50, vir: 50

Variable type: numeric

skim_variable	n_missing	complete_rate	mean	sd	p0	p2.5	p50	p75	p100	hist
Sepal.Length	0	1	5.84	0.83	4.3	5.1	5.8	6.4	7.9	
Sepal.Width	0	1	3.06	0.44	2.0	2.8	3.0	3.3	4.4	
Petal.Length	0	1	3.76	1.77	1.0	1.6	4.3	5.1	6.9	
Petal.Width	0	1	1.20	0.76	0.1	0.3	1.3	1.8	2.5	

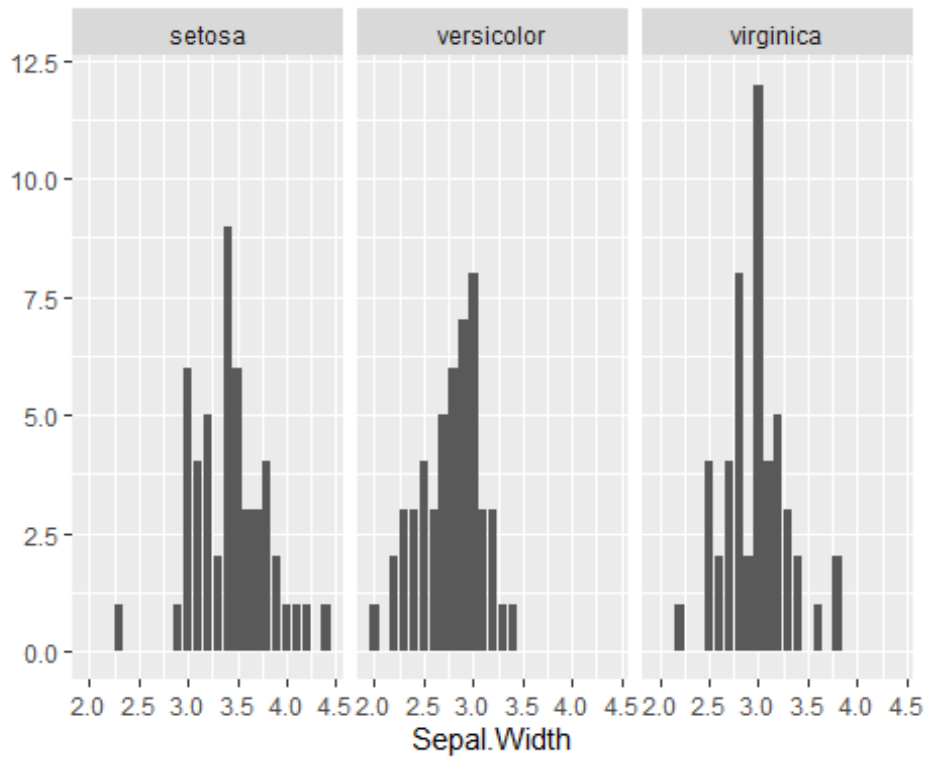
```
iris %>%  
  rename(Petal_Length = Petal.Length) %>%  
  rename(Petal_Width = Petal.Width) %>%  
  rename(Sepal_Width = Sepal.Width) %>%  
  rename(Sepal_Length = Sepal.Length) %>%  
  mutate(ID = row_number()) %>%
```

```
select(ID, Species, Sepal_Width, Sepal_Length) %>%
  ggplot(aes(Sepal_Length, Sepal_Width)) +
  geom_point(size = 2.5) +
  facet_wrap(~Species) +
  theme_light()
```

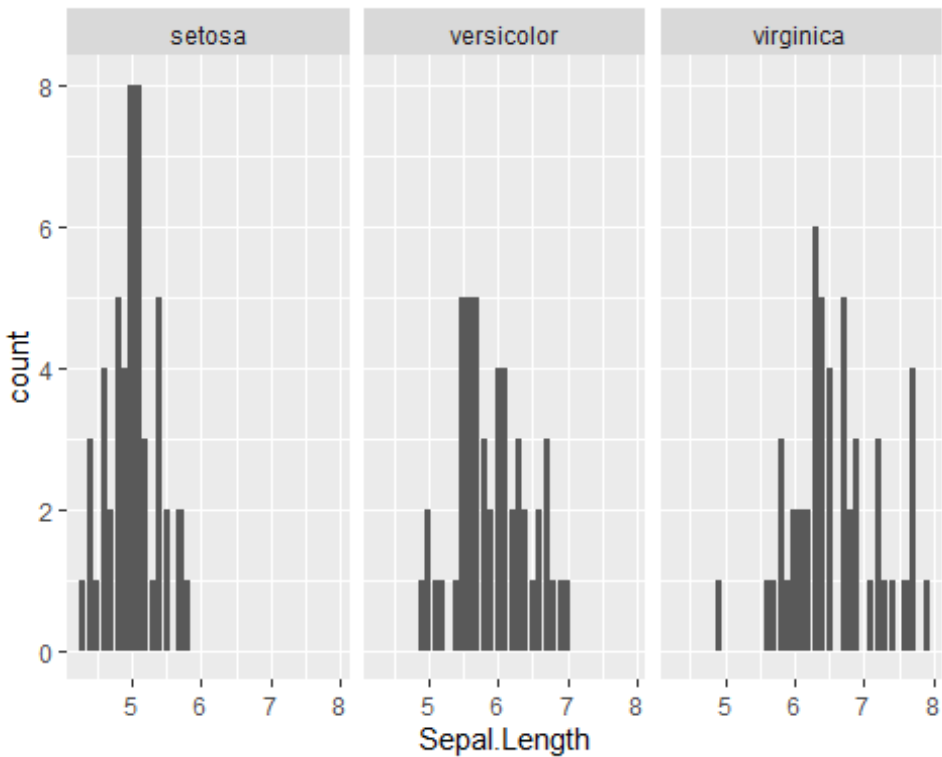


```
iris %>%
  ggplot(aes(Sepal.Width))+
  geom_histogram(stat = "count", binwidth = 5) +
  facet_wrap(~Species) +
  labs(y = "")
```

Warning: Ignoring unknown parameters: binwidth, bins, pad



```
iris %>%  
  ggplot(aes(Sepal.Length))+  
  geom_histogram(stat = "count", binwidth = 5) +  
  facet_wrap(~Species)  
  
## Warning: Ignoring unknown parameters: binwidth, bins, pad
```



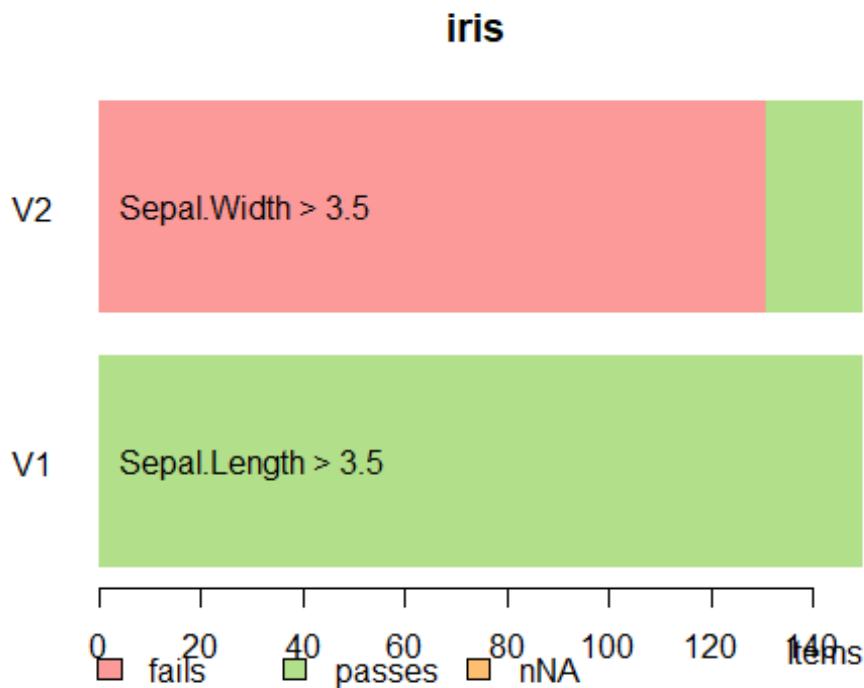
```
iris %>%
  group_by(Species) %>%
  summarise(Count = n())

## # A tibble: 3 × 2
##   Species    Count
##   <fct>      <int>
## 1 setosa         50
## 2 versicolor    50
## 3 virginica     50

test<- validator(
  Sepal.Length > 3.5,
  Sepal.Width > 3.5
)

cf <- confront(iris, test)
barplot(cf, main = "iris")

## Warning: The 'barplot' method for confrontation objects is deprecated. Use
## 'plot' instead
```



```
summary(lm(Sepal.Length~Sepal.Width, data = iris))

##
## Call:
## lm(formula = Sepal.Length ~ Sepal.Width, data = iris)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.5561 -0.6333 -0.1120  0.5579  2.2226
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   6.5262     0.4789   13.63  <2e-16 ***
## Sepal.Width  -0.2234     0.1551   -1.44    0.152
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.8251 on 148 degrees of freedom
## Multiple R-squared:  0.01382,    Adjusted R-squared:  0.007159
## F-statistic: 2.074 on 1 and 148 DF,  p-value: 0.1519

iris %>%
  ggplot(aes(Sepal.Length, Sepal.Width)) +
  geom_point(aes(color = Species), size = 1.85) +
  facet_wrap(~Species) +
  geom_smooth(method = "lm", se = FALSE)
```

```
## `geom_smooth()` using formula 'y ~ x'
```

