

p8105_hw1_kw2873

Keyi Wang

9/13/2019

I'm an R Markdown document!

problem 1: dataframe

```
set.seed(12)
s = rnorm(8)
v_logi = s > 0
v_ch = c("yellow", "blue", "green", "red", "black", "white", "gray", "purple")
v_fac = factor(c("AB", "B", "0", "B", "B", "0", "AB", "B"),
               levels = c('AB', 'B', '0'))

example_df = tibble(
  samp = s,
  vec_logical = v_logi,
  vec_char = v_ch,
  vec_factor = v_fac
)
example_df
```

```
## # A tibble: 8 x 4
##   samp vec_logical vec_char vec_factor
##   <dbl> <lgl>      <chr>   <fct>
## 1 -1.48 FALSE      yellow  AB
## 2  1.58 TRUE       blue    B
## 3 -0.957 FALSE     green   0
## 4 -0.920 FALSE     red     B
## 5 -2.00 FALSE     black   B
## 6 -0.272 FALSE     white   0
## 7 -0.315 FALSE     gray    AB
## 8 -0.628 FALSE     purple  B
```

```
### take the mean of each variables
mean(s)
```

```
## [1] -0.6242112
```

```
mean(v_logi)
```

```
## [1] 0.125
```

```
mean(v_ch)
```

```
## Warning in mean.default(v_ch): argument is not numeric or logical:
## returning NA
## [1] NA
```

```
mean(v_fac)
```

```
## Warning in mean.default(v_fac): argument is not numeric or logical:  
## returning NA
```

```
## [1] NA
```

```
### using the as.numeric function
```

```
as.numeric(v_logi)
```

```
## [1] 0 1 0 0 0 0 0 0
```

```
as.numeric(v_ch)
```

```
## Warning: NAs introduced by coercion
```

```
## [1] NA NA NA NA NA NA NA NA
```

```
as.numeric(v_fac)
```

```
## [1] 1 2 3 2 2 3 1 2
```

“ the mean of my sample is -0.6242112

the mean of my logical vector is 0.125

the mean of my character vector is NA

the mean of my factor vector is NA

we noticed that the mean of both character and factor vector doesn't exist. Both logical and factor vector can be converted into numeric vector using as.numeric function, yet, character vector failed to do so.

problem 1 continued

```
new_logi1 = as.numeric(v_logi)
```

```
new_logi2 = as.factor(v_logi)
```

```
result1 = new_logi1*s
```

```
result2 = new_logi2*s
```

```
## Warning in Ops.factor(new_logi2, s): '*' not meaningful for factors
```

```
new_logi3 = as.numeric(new_logi2)
```

```
result3 = new_logi3*s
```

problem 2

```
set.seed(1)
```

```
x = rnorm(500)
```

```
library(tidyverse)
```

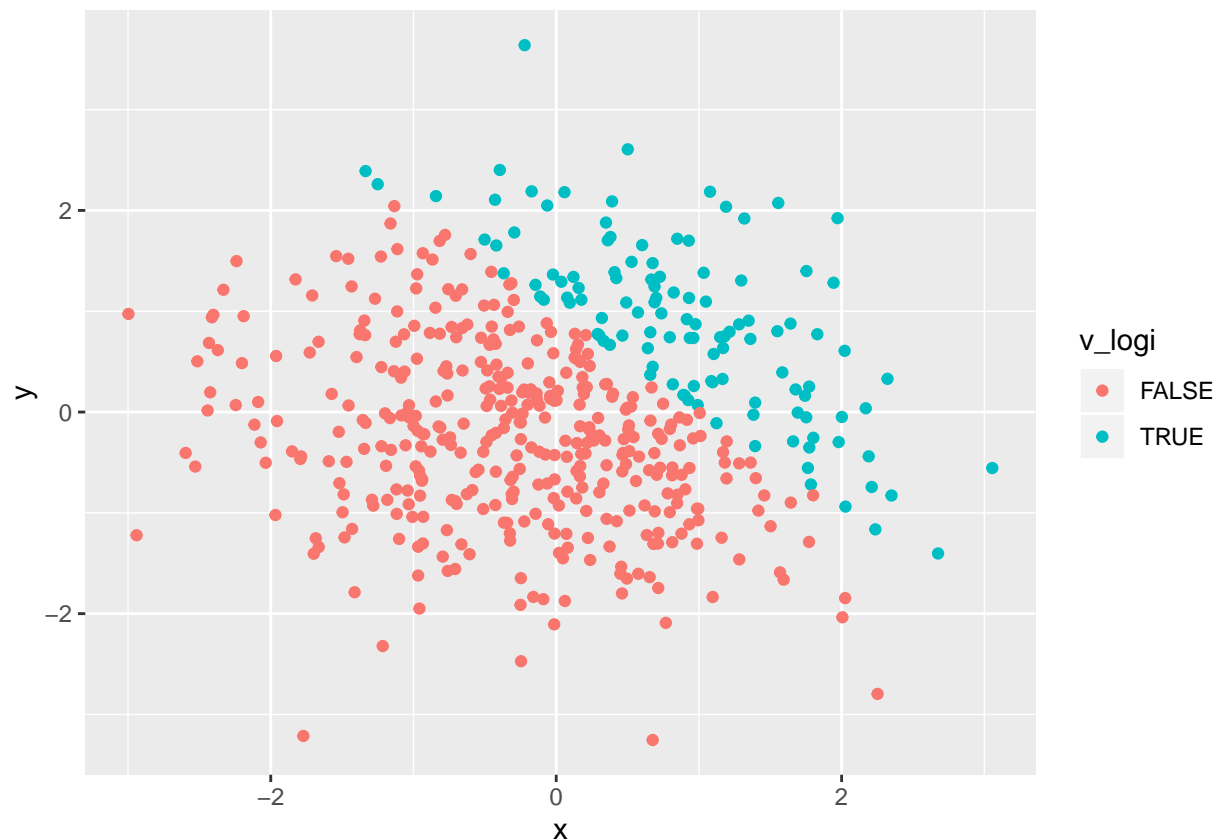
```
plot_df = tibble(  
  x = rnorm(500),  
  y = rnorm(500),  
  v_logi = x + y > 1,  
  v_num = as.numeric(v_logi),
```

```
v_fac = as.factor(v_num),
)
new_logi = pull(plot_df, v_logi)
```

In this dataplot, the number of rows is 500 the number of column is 5 the mean, median, and standard deviation of x are 0.0226441, -0.0367783, 1.0239989 respectively. And the proportion for cases for which $x+y>1$ is 117/ 500

problem 2 continued with making plots

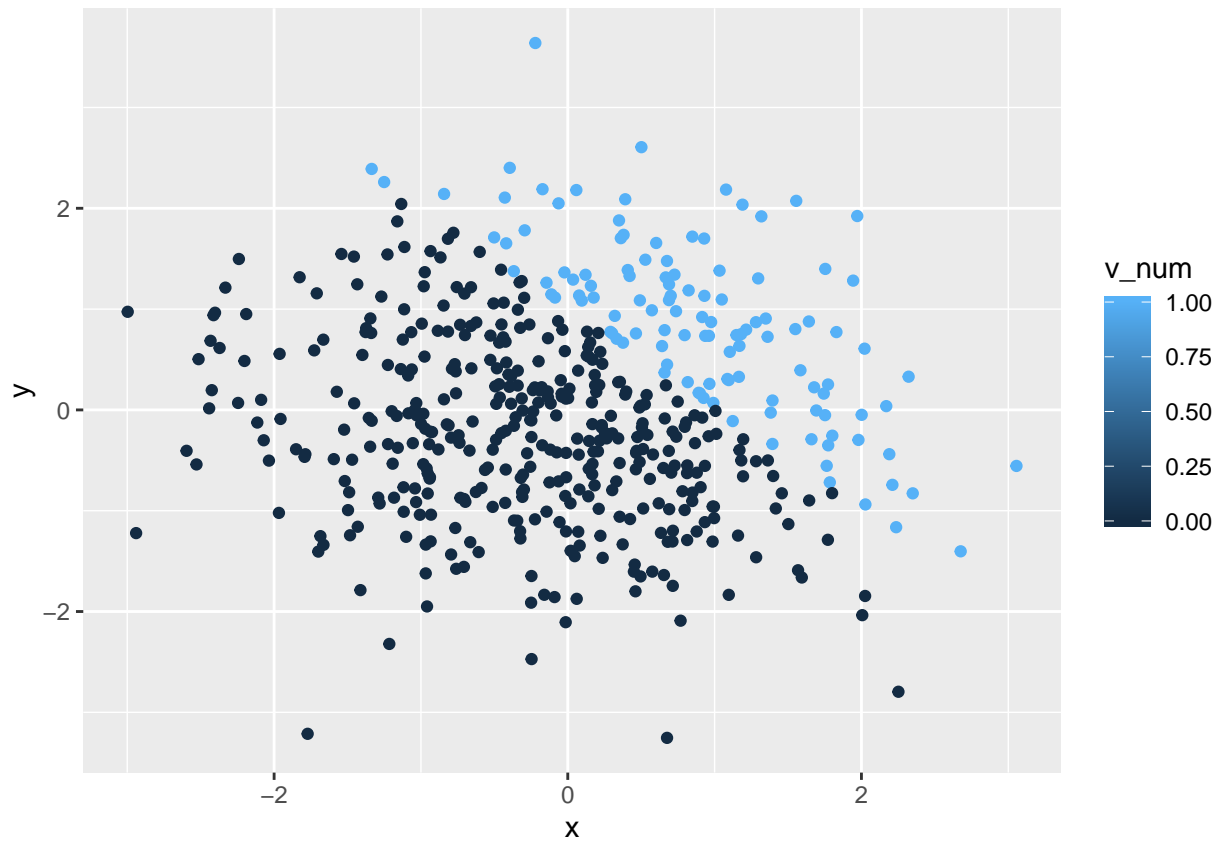
```
ggplot(plot_df, aes(x = x, y = y, color = v_logi)) + geom_point()
```



```
ggsave("scatterplot1.png")
```

```
## Saving 6.5 x 4.5 in image
```

```
ggplot(plot_df, aes(x = x, y = y, color = v_num)) + geom_point()
```



```
ggplot(plot_df, aes(x = x, y = y, color = v_fac)) + geom_point()
```

