HW_Last

GA

2022-11-24

Learning Objectives • Practice stringr, dplyr, and ggplot2. • WARNING: Do not use str_view() or str_view_all() on these data. It will stall your computer. The data aren't that large, but str_view() and str_view_all() are inefficient with medium data. – More stringr options can be found in RDS. # Exercise 1: From RDS:

1. Replace all forward slashes in a string with backslashes. Test it out on the following string:

x <- "hello\////how//are////you////"

```
library(tidyverse)
```

```
## -- Attaching packages --------------------------------------- tidyverse 1.3.2 --
## v ggplot2 3.4.0      v purrr   0.3.5
## v tibble  3.1.8      v dplyr   1.0.10
## v tidyr   1.2.1      v stringr 1.4.1
## v readr   2.1.3      v forcats 0.5.2
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
x <- "hello\\////how//are////you////"
y <- str_replace_all(x, "/", "\\\\")
writeLines(x)
```

```
## hello\////how//are////you////
```

```
writeLines(y)
```

```
## hello\\\\\how\\are\\\\you\\\\
```

2. Construct regular expressions to match words that:

a. Start and end with the same character. A word of length 1 should be matched. Test it out on "A", "AB", and "ABA".

```
str_detect("A", "^((.).*\\2|.)$")
```

```
## [1] TRUE
```

```
str_detect("AB", "^((.).*\\2|.)$")
```

```
## [1] FALSE
```

```
str_detect("ABA", "^((.).*\\2|.)$")
```

```
## [1] TRUE
```

b. Contain a repeated pair of letters (e.g. "church" contains "ch" repeated twice.) Test it out on "AAA", "AAAA", and "AABAA".

```
str_detect("AAA", "(..).*\\1")
```

```
## [1] FALSE
```

```
str_detect("AAAA", "(..).*\\1")
```

```
## [1] TRUE
```

```
str_detect("AABAA", "(..).*\\1")
```

```
## [1] TRUE
```

c. Contain one letter repeated in at least three places (e.g. "eleven" contains three "e"s.) Test it out on "AAA", "AAB", and "AABA".

```
str_detect("AAA", "(.).*\\1.*\\1")
```

```
## [1] TRUE
```

```
str_detect("AAB", "(.).*\\1.*\\1")
```

```
## [1] FALSE
```

```
str_detect("AABA", "(.).*\\1.*\\1")
```

```
## [1] TRUE
```

FACTORS # Exercise 2: Sex, Lies, and Religion The data frame in https://dcgerard.github.io/stat_412_612/data/sexlierel.txt, taken from @clay- ton1971religiosity, contain the following variables:

• gender: The gender of the individual. 1 = Female, 2 = Male. • scale: Different scales of sexual permissiveness. 1 = Ritualistic, 2 = Experiential, 3 = Ideological, 4 = Composite. • perm: The level of sexual permissiveness. 1 = Low, 2 = High. • lie: The propensity to lie. 1 = Lower, 2 = Higher. relig: How religious a person is. 1 = Low, 2 = High • count: The number of individuals satisfying the conditions of the other variables.

1. Read the data into R.

```
library(tidyverse)
library('httr')
sexlierel <- read_table(file = "https://dcgerard.github.io/stat_412_612/data/sexlierel.txt",
                            col_types = cols(
                              gender = col_factor(),
                              scale  = col_factor(),
                              perm   = col_factor(),
                              lie    = col_factor(),
                              relig  = col_factor(),
                              count  = col_double()
                              ))
```

2. Change the level names to something more informative.

```
sexlierel %>%
  mutate(gender = fct_recode(gender,
                             "Female" = "1",
                             "Male"   = "2"),

         scale  = fct_recode(scale,
                             "ritualistic"  = "1",
                             "experiential" = "2",
                             "idiological"  = "3",
                             "composite" = "4"),
```

```
        perm   = fct_recode(perm,
                            "low"  = "1",
                            "high" = "2"),
        lie    = fct_recode(lie,
                            "lower"  = "1",
                            "higher" = "2"),
        relig  = fct_recode(relig,
                            "low"  = "1",
                            "high" = "2")) ->sexlierel

head(sexlierel)
```

```
## # A tibble: 6 x 6
##   gender scale       perm  lie    relig count
##   <fct>  <fct>       <fct> <fct>  <fct> <dbl>
## 1 Female ritualistic low   lower  low      52
## 2 Female ritualistic low   lower  high     74
## 3 Female ritualistic low   higher low      50
## 4 Female ritualistic low   higher high     51
## 5 Female ritualistic high  lower  low      34
## 6 Female ritualistic high  lower  high     13
```

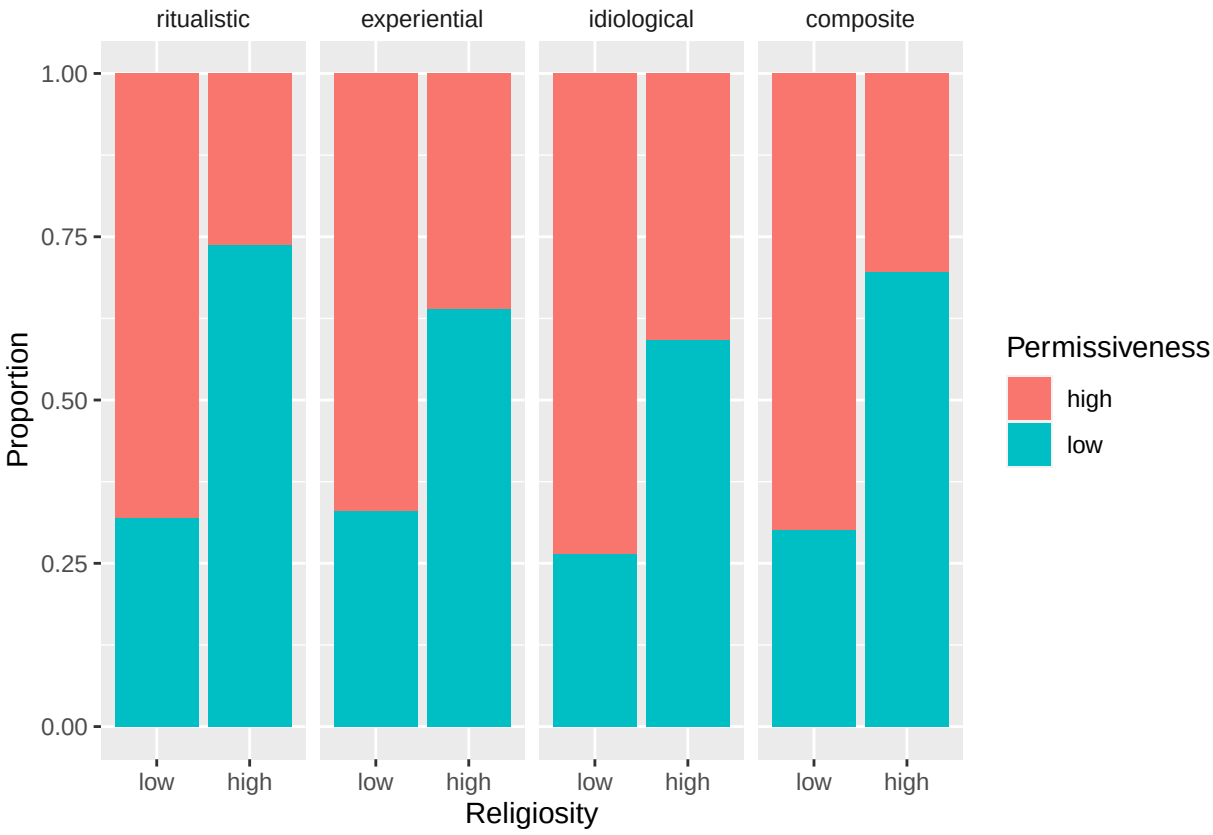
3. Flip the order of the levels in perm.

```
sexlierel %>%
     mutate(perm = fct_rev(perm)) ->
     sexlierel
head(sexlierel)
```

```
## # A tibble: 6 x 6
##   gender scale       perm  lie    relig count
##   <fct>  <fct>       <fct> <fct>  <fct> <dbl>
## 1 Female ritualistic low   lower  low      52
## 2 Female ritualistic low   lower  high     74
## 3 Female ritualistic low   higher low      50
## 4 Female ritualistic low   higher high     51
## 5 Female ritualistic high  lower  low      34
## 6 Female ritualistic high  lower  high     13
```

4. For males, for what scales does there appear to be an association between religiosity and permissiveness? Use one plot to explore.

```
sexlierel %>%
     filter(gender == "Male") %>%
     ggplot(aes(x = relig, y = count, fill = perm)) +
     geom_col(pos = "fill") +
     facet_grid(. ~ scale) +
     ylab("Proportion") +
     xlab("Religiosity") +
     scale_fill_discrete(name = "Permissiveness") +
     theme(strip.background = element_rect(fill = "white"))
```


ritualistic
experiential
idiological
composite
1.00
0.75
0.50
0.25
0.00
Proportion
low
high
Religiosity
Permissiveness
high
low

for all