

Nama : Muhammad Rendy Saputra Nim : 121450045 Kelas : RA

- “jenis kelamin” - pria / wanita
- “ras/etnis” - salah satu dari 5 kombinasi ras/etnis
- “tingkat pendidikan orang tua” - tingkat pendidikan tertinggi dari salah satu orang tua
- “makan siang” - apakah siswa menerima makan siang gratis / diskon atau makan siang standar
- “kursus persiapan ujian” - apakah siswa mengambil kursus persiapan ujian
- “matematika” - nilai ujian matematika

```
df <- read.csv("C:/Users/Repets/Downloads/3. Tugas Pertemuan 3.csv")
head(df)
```

```
##   gender race.ethnicity parent_education_level      lunch test_prep_course
## 1 female      group B      bachelor's degree    standard          none
## 2 female      group C          some college    standard      completed
## 3 female      group B      master's degree    standard          none
## 4  male      group A      associate's degree free/reduced          none
## 5  male      group C          some college    standard          none
## 6 female      group B      associate's degree    standard          none
##   math
## 1   72
## 2   69
## 3   90
## 4   47
## 5   76
## 6   71
```

```
library(tidyverse)
```

```
## -- Attaching core tidyverse packages ----- tidyverse 2.0.0 --
## v dplyr      1.1.2      v readr      2.1.4
## v forcats    1.0.0      v stringr   1.5.0
## v ggplot2    3.4.2      v tibble    3.2.1
## v lubridate  1.9.2      v tidyr     1.3.0
## v purrr      1.0.1
## -- Conflicts ----- tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()     masks stats::lag()
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors
```

- Pada data diatas, buat variabel “new” yang berisi nilai dari 1,2,3 dst

```
df <- df |>
  mutate(new = 1:1000)
```

```
head(df)
```

```
##   gender race.ethnicity parent_education_level      lunch test_prep_course
## 1 female      group B      bachelor's degree    standard          none
## 2 female      group C          some college    standard      completed
## 3 female      group B      master's degree    standard          none
```

```
## 4   male      group A      associate's degree free/reduced      none
## 5   male      group C           some college      standard      none
## 6 female      group B      associate's degree      standard      none
##   math new
## 1   72   1
## 2   69   2
## 3   90   3
## 4   47   4
## 5   76   5
## 6   71   6
```

- Ambil amatan pada “lunch”

```
library(janitor)
```

```
##
## Attaching package: 'janitor'

## The following objects are masked from 'package:stats':
##
##   chisq.test, fisher.test
```

```
df |>
  tabyl(lunch)
```

```
##           lunch    n percent
## free/reduced 355    0.355
##      standard 645    0.645
```

terlihat jumlah free/reduced / siswa menerima makan siang gratis memiliki proporsi 0.355, dan siswa yang menerima makan siang standard memiliki proporsi 0.645

- Tentukan variabel mana yang factor kemudian ubah variabel tersebut ke dalam faktor

```
glimpse(df)
```

```
## Rows: 1,000
## Columns: 7
## $ gender      <chr> "female", "female", "female", "male", "male", "~
## $ race.ethnicity <chr> "group B", "group C", "group B", "group A", "gr~
## $ parent_education_level <chr> "bachelor's degree", "some college", "master's ~
## $ lunch       <chr> "standard", "standard", "standard", "free/reduc~
## $ test_prep_course <chr> "none", "completed", "none", "none", "none", "n~
## $ math        <int> 72, 69, 90, 47, 76, 71, 88, 40, 64, 38, 58, 40,~
## $ new         <int> 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, ~
```

variabel gender, race.ethnicity, parent\_education\_level, lunc, test\_prep\_course bertipe data character

```
df$gender <- as.factor(df$gender)
df$race.ethnicity <- as.factor(df$race.ethnicity)
df$parent_education_level <- as.factor(df$parent_education_level)
df$lunch <- as.factor(df$lunch)
df$test_prep_course <- as.factor(df$test_prep_course)
```

```
glimpse(df)
```

```
## Rows: 1,000
## Columns: 7
## $ gender                <fct> female, female, female, male, male, female, fem~
## $ race.ethnicity        <fct> group B, group C, group B, group A, group C, gr~
## $ parent_education_level <fct> bachelor's degree, some college, master's degree~
## $ lunch                 <fct> standard, standard, standard, free/reduced, sta~
## $ test_prep_course      <fct> none, completed, none, none, none, none, comple~
## $ math                  <int> 72, 69, 90, 47, 76, 71, 88, 40, 64, 38, 58, 40, ~
## $ new                   <int> 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, ~
```

saat ini variabel yang sebelumnya character telah saya ubah menjadi factor

- Lakukan pengkodean terhadap Gender , “parent education level” and “lunch”

```
df$gender <- as.integer(df$gender)
```

```
df |>
  tabyl(parent_education_level)
```

```
## parent_education_level    n percent
##      associate's degree 222    0.222
##      bachelor's degree 118    0.118
##      high school      196    0.196
##      master's degree   59    0.059
##      some college      226    0.226
##      some high school  179    0.179
```

```
df$parent_education_level <- as.integer(df$parent_education_level)
```

```
df$lunch <- as.integer(df$lunch)
```

```
glimpse(df)
```

```
## Rows: 1,000
## Columns: 7
## $ gender                <int> 1, 1, 1, 2, 2, 1, 1, 2, 2, 1, 2, 2, 1, 2, 1, 1, ~
## $ race.ethnicity        <fct> group B, group C, group B, group A, group C, gr~
## $ parent_education_level <int> 2, 5, 4, 1, 5, 1, 5, 5, 3, 3, 1, 1, 3, 5, 4, 6, ~
## $ lunch                 <int> 2, 2, 2, 1, 2, 2, 2, 1, 1, 1, 2, 2, 2, 2, 2, 2, ~
## $ test_prep_course      <fct> none, completed, none, none, none, none, comple~
## $ math                  <int> 72, 69, 90, 47, 76, 71, 88, 40, 64, 38, 58, 40, ~
## $ new                   <int> 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, ~
```

disini saya telah merubah variabel gender, parent\_education\_level, lunch telah saya ubah menjadi integer

- Kemudian Urutkan Data berdasarkan “Race” secara ascending

```
df <- df |>
  arrange(race.ethnicity)
head(df)
```

```
##   gender race.ethnicity parent_education_level lunch test_prep_course math new
## 1      2      group A              1      1      none      47   4
## 2      2      group A              5      2    completed      78  14
## 3      1      group A              4      2      none      50  15
## 4      2      group A              4      1      none      73  26
## 5      1      group A              1      2    completed      55  47
## 6      2      group A              6      1      none      39  62
```

data diatas telah dilakukan pengurutan secara ascending berdasarkan variabel race.ethnicity

- Ubah nilai math menjadi Nilai Huruf dan Tambahkan kolom Status Penilaian sesuai dengan kriteria disamping ini.

```
df <- df |>
  mutate(nilai_huruf = case_when(
    math >= 80 ~ "A",
    math >= 75 ~ "AB",
    math >= 70 ~ "B",
    math >= 65 ~ "BC",
    math >= 60 ~ "C",
    math >= 55 ~ "D",
    TRUE ~ "E"
  ))
head(df)
```

```
##   gender race.ethnicity parent_education_level lunch test_prep_course math new
## 1      2      group A              1      1      none      47   4
## 2      2      group A              5      2    completed      78  14
## 3      1      group A              4      2      none      50  15
## 4      2      group A              4      1      none      73  26
## 5      1      group A              1      2    completed      55  47
## 6      2      group A              6      1      none      39  62
##   nilai_huruf
## 1           E
## 2          AB
## 3           E
## 4           B
## 5           D
## 6           E
```

```
df <- df |>
  mutate(status_penilaian = ifelse(math >= 60, "Lulus", "Tidak Lulus"))
head(df)
```

```
##   gender race.ethnicity parent_education_level lunch test_prep_course math new
## 1      2      group A           1      1           none    47    4
## 2      2      group A           5      2      completed    78   14
## 3      1      group A           4      2           none    50   15
## 4      2      group A           4      1           none    73   26
## 5      1      group A           1      2      completed    55   47
## 6      2      group A           6      1           none    39   62
##   nilai_huruf status_penilaian
## 1           E      Tidak Lulus
## 2          AB           Lulus
## 3           E      Tidak Lulus
## 4           B           Lulus
## 5           D      Tidak Lulus
## 6           E      Tidak Lulus
```

pada data di atas dibuat variable baru dengan membuat pengkategorian dengan nama nilai\_huruf dan status\_penilaian berdasarkan nilai math

- Dari nilai math ambil, nilai yang bersala dari bilangan prima

```
is_prime <- function(n) {
  if (n <= 1) return(FALSE)
  if (n <= 3) return(TRUE)
  if (n %% 2 == 0 || n %% 3 == 0) return(FALSE)
  i <- 5
  while (i * i <= n) {
    if (n %% i == 0 || n %% (i + 2) == 0) return(FALSE)
    i <- i + 6
  }
  return(TRUE)
}
```

```
df |>
  filter(sapply(math, is_prime)) |>
  head()
```

```
##   gender race.ethnicity parent_education_level lunch test_prep_course math new
## 1      2      group A           1      1           none    47    4
## 2      2      group A           4      1           none    73   26
## 3      1      group A           1      1           none    41   73
## 4      2      group A           6      2      completed    47  251
## 5      1      group A           6      2           none    71  379
## 6      2      group A           5      2           none    53  402
##   nilai_huruf status_penilaian
## 1           E      Tidak Lulus
## 2           B           Lulus
## 3           E      Tidak Lulus
## 4           E      Tidak Lulus
## 5           B           Lulus
## 6           E      Tidak Lulus
```

- Tambahkan 5 data baru untuk melakukan reshaping, reshaping dengan Long to wide data format beri nama datawide, reshaping dengan datawide menjadi long

```
country_long <- data.frame(
  expand.grid(country = c("Sweden", "Denmark", "Norway"), year = 1994:1996),
  avgtemp = round(runif(9, 3, 12), 0)
)
country_long
```

```
##   country year avgtemp
## 1  Sweden 1994      7
## 2 Denmark 1994      4
## 3  Norway 1994      5
## 4  Sweden 1995      9
## 5 Denmark 1995      4
## 6  Norway 1995      3
## 7  Sweden 1996      6
## 8 Denmark 1996     11
## 9  Norway 1996     11
```

pada proses di atas saya membuat variabel baru berbentuk long yang saya beri nama country long

```
country_wide <- country_long |>
  pivot_wider(names_from = "country", values_from = "avgtemp")
```

```
country_wide
```

```
## # A tibble: 3 x 4
##   year Sweden Denmark Norway
##   <int>   <dbl>   <dbl>   <dbl>
## 1  1994      7      4      5
## 2  1995      9      4      3
## 3  1996      6     11     11
```

pada tahap di atas saya merubah data country long yang sebelumnya berbentuk long menjadi wide berdasarkan variabel country sebagai kolom, dan variabel avgtemp sebagai nilai dari kolom

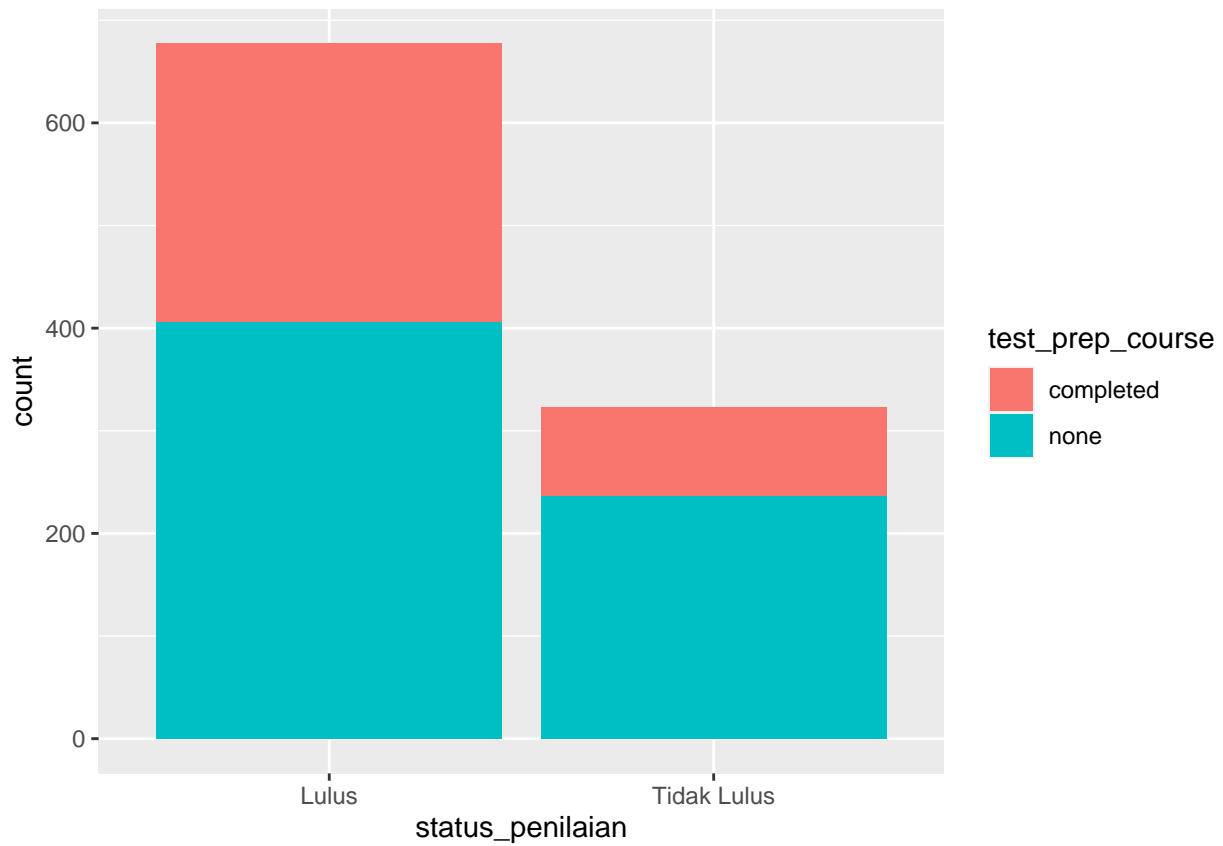
- Menurut Anda, Apakah dengan mengambil Kursus persiapan ujian Dapat menentukan seseorang lulus Atau tidak pada kasus disamping.

```
df1 <- df |>
  select(test_prep_course, status_penilaian) |>
  group_by(status_penilaian, test_prep_course) |>
  count()
```

```
df1
```

```
## # A tibble: 4 x 3
## # Groups:   status_penilaian, test_prep_course [4]
##   status_penilaian test_prep_course    n
##   <chr>           <fct>         <int>
## 1 Lulus           completed      271
## 2 Lulus           none          406
## 3 Tidak Lulus     completed      87
## 4 Tidak Lulus     none          236
```

```
ggplot(data = df, mapping = aes(x = status_penilaian, fill = test_prep_course)) +  
  geom_bar()
```



berdasarkan data dan grafik siswa tanpa persiapan ujian lebih tinggi mengalami kelulusan jika dibandingkan dengan siswa yang mempersiapkan ujian terlihat proporsi lebih tinggi, tetapi lebih tinggi pula tingkat ketidaklulusannya jika dibandingkan dengan siswa yang mempersiapkan ujian