

Visualisasi

Muhammad Aswan Syahputra

4/9/2019

Contents

Dalam modul visualisasi ini, Anda akan menggunakan data `evals` dan paket `ggplot2`. Aktifkanlah paket untuk mengimpor data dan paket `ggplot2`!

```
library(tidyverse)
```

```
## Warning: package 'tidyverse' was built under R version 3.5.3
```

```
## -- Attaching packages ----- tidyverse
```

```
## v ggplot2 3.1.1    v purrr  0.3.2
## v tibble  2.1.1    v dplyr  0.8.1
## v tidyr   0.8.3    v stringr 1.4.0
## v readr   1.3.1    v forcats 0.4.0
```

```
## Warning: package 'ggplot2' was built under R version 3.5.3
```

```
## Warning: package 'tibble' was built under R version 3.5.3
```

```
## Warning: package 'tidyr' was built under R version 3.5.3
```

```
## Warning: package 'readr' was built under R version 3.5.3
```

```
## Warning: package 'purrr' was built under R version 3.5.3
```

```
## Warning: package 'dplyr' was built under R version 3.5.3
```

```
## Warning: package 'stringr' was built under R version 3.5.3
```

```
## Warning: package 'forcats' was built under R version 3.5.3
```

```
## -- Conflicts ----- tidyverse
```

```
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
# library(readr)
# library(ggplot2)
```

Sekarang impor dataset `evals` melalui pranala “<https://www.openintro.org/stat/data/evals.csv>”. Simpan obyek data tersebut dengan nama `evals`. (Petunjuk: `?read_csv`)

```
evals <- read_csv("https://www.openintro.org/stat/data/evals.csv")
```

```
## Parsed with column specification:
## cols(
##   .default = col_double(),
##   rank = col_character(),
##   ethnicity = col_character(),
##   gender = col_character(),
##   language = col_character(),
##   cls_level = col_character(),
##   cls_profs = col_character(),
##   cls_credits = col_character(),
##   pic_outfit = col_character(),
##   pic_color = col_character()
## )

## See spec(...) for full column specifications.
```

```
evals
```

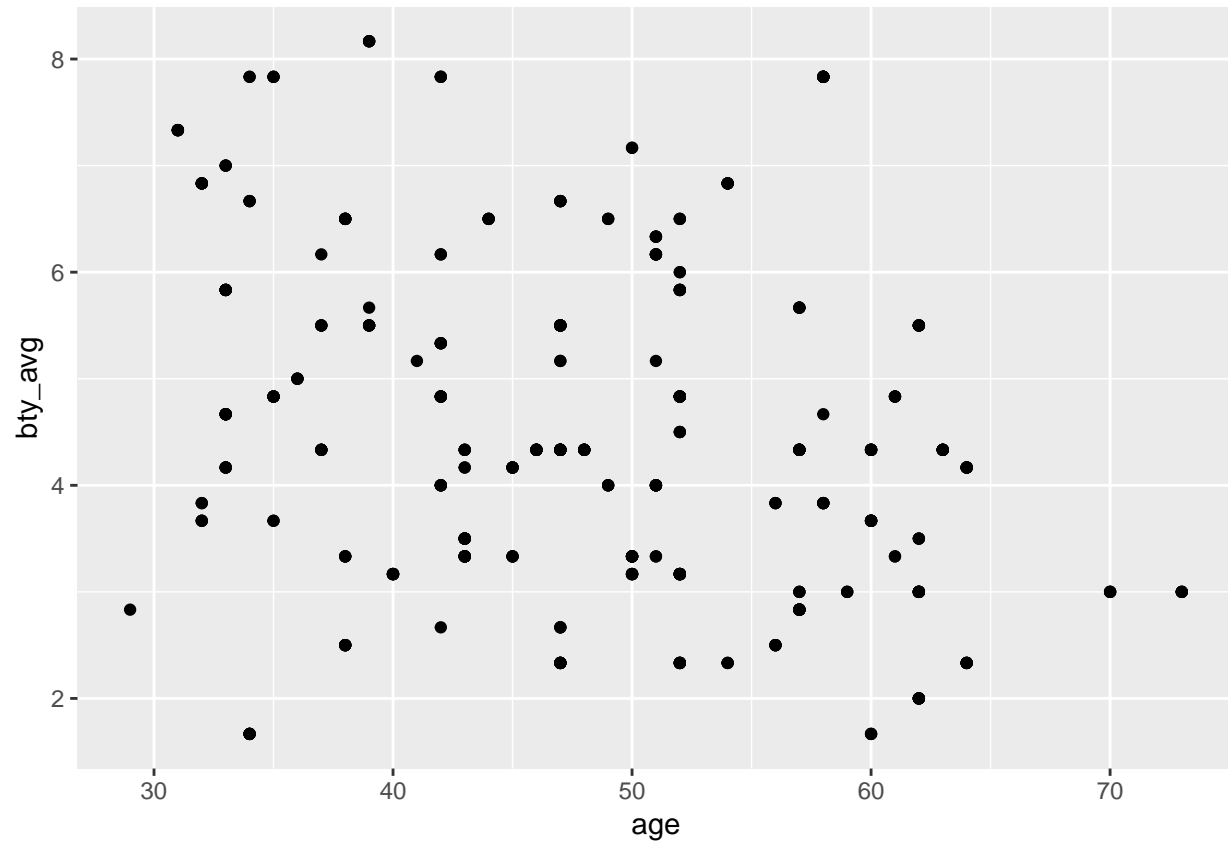
```
## # A tibble: 463 x 21
##   score rank ethnicity gender language age cls_perc_eval cls_did_eval
##   <dbl> <chr> <chr>      <chr> <chr>      <dbl>      <dbl>      <dbl>
## 1  4.7 tenu~ minority female english    36        55.8        24
## 2  4.1 tenu~ minority female english    36        68.8        86
## 3  3.9 tenu~ minority female english    36        60.8        76
## 4  4.8 tenu~ minority female english    36        62.6        77
## 5  4.6 tenu~ not mino~ male   english    59         85         17
## 6  4.3 tenu~ not mino~ male   english    59        87.5         35
## 7  2.8 tenu~ not mino~ male   english    59        88.6         39
## 8  4.1 tenu~ not mino~ male   english    51       100         55
## 9  3.4 tenu~ not mino~ male   english    51        56.9       111
## 10 4.5 tenu~ not mino~ female english    40        87.0         40
## # ... with 453 more rows, and 13 more variables: cls_students <dbl>,
## #   cls_level <chr>, cls_profs <chr>, cls_credits <chr>,
## #   bty_follower <dbl>, bty_follower_upper <dbl>, bty_follower_upper_upper <dbl>,
## #   bty_mollower <dbl>, bty_mollower_upper <dbl>, bty_mollower_upper_upper <dbl>,
## #   bty_avg <dbl>, pic_outfit <chr>, pic_color <chr>
```

Anda dapat membuat grafik melalui `ggplot2` dengan menggunakan kode dengan struktur berikut:

```
ggplot(data, aes(x, y)) +
  geom_X()
```

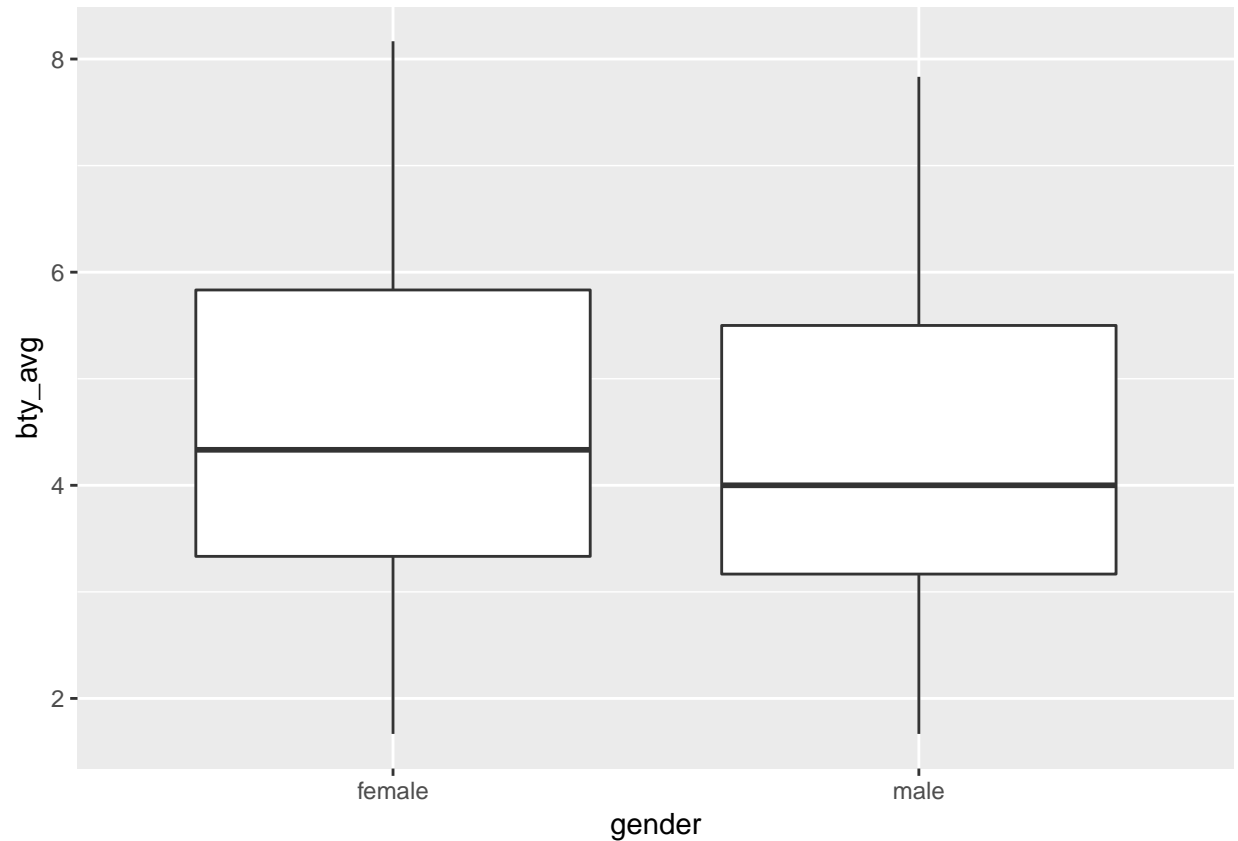
Anda dapat mengganti `X` dengan `geom` yang sesuai. Misalnya `geom_point()` untuk menggambar titik, `geom_line` untuk menggambar garis, `geom_text` untuk memasukan teks dalam gambar. Perhatikan contoh berikut untuk membuat `scatterplot` antara `bty_avg` terhadap `age`.

```
ggplot(data = evals, aes(x = age, y = bty_avg)) +
  geom_point()
```



Sekarang dapatkan Anda membuat grafik dengan `gender` pada sumbu x dan `bty_avg` pada sumbu y? Serta gantilah `geom_point()` menjadi `geom_boxplot()`. Grafik dan informasi apakah yang Anda dapatkan?

```
ggplot(data = evals, aes(x = gender, y = bty_avg)) +  
  geom_boxplot()
```



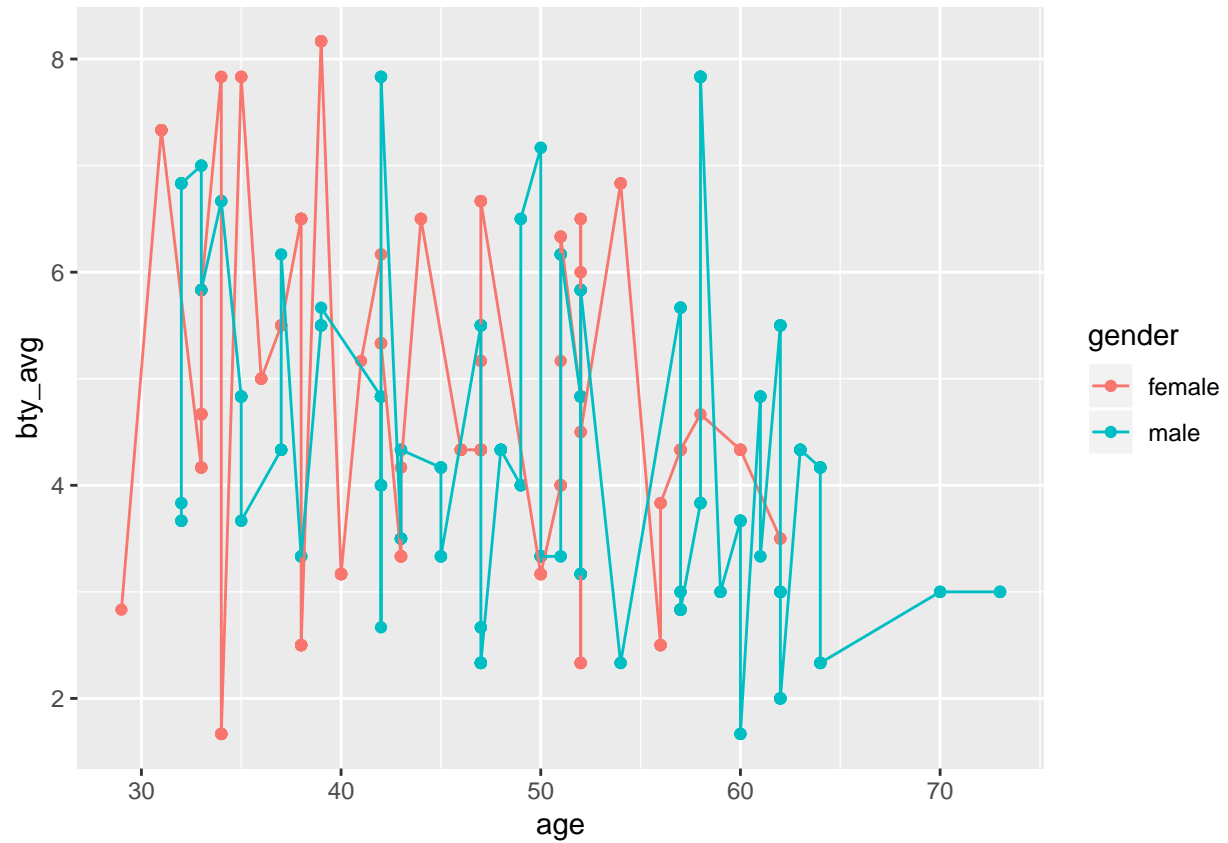
Anda dapat menambahkan informasi mengenai variabel lain kedalam plot dengan menempatkannya pada bagian `aes()` dengan menggunakan `aesthetic` yang sesuai. Apa yang akan terjadi saat anda menambahkan informasi `'colour = gender'` pada baris kode berikut?

```
ggplot(data = evals, aes(x = age, y = bty_avg, colour = gender)) +  
  geom_point()
```



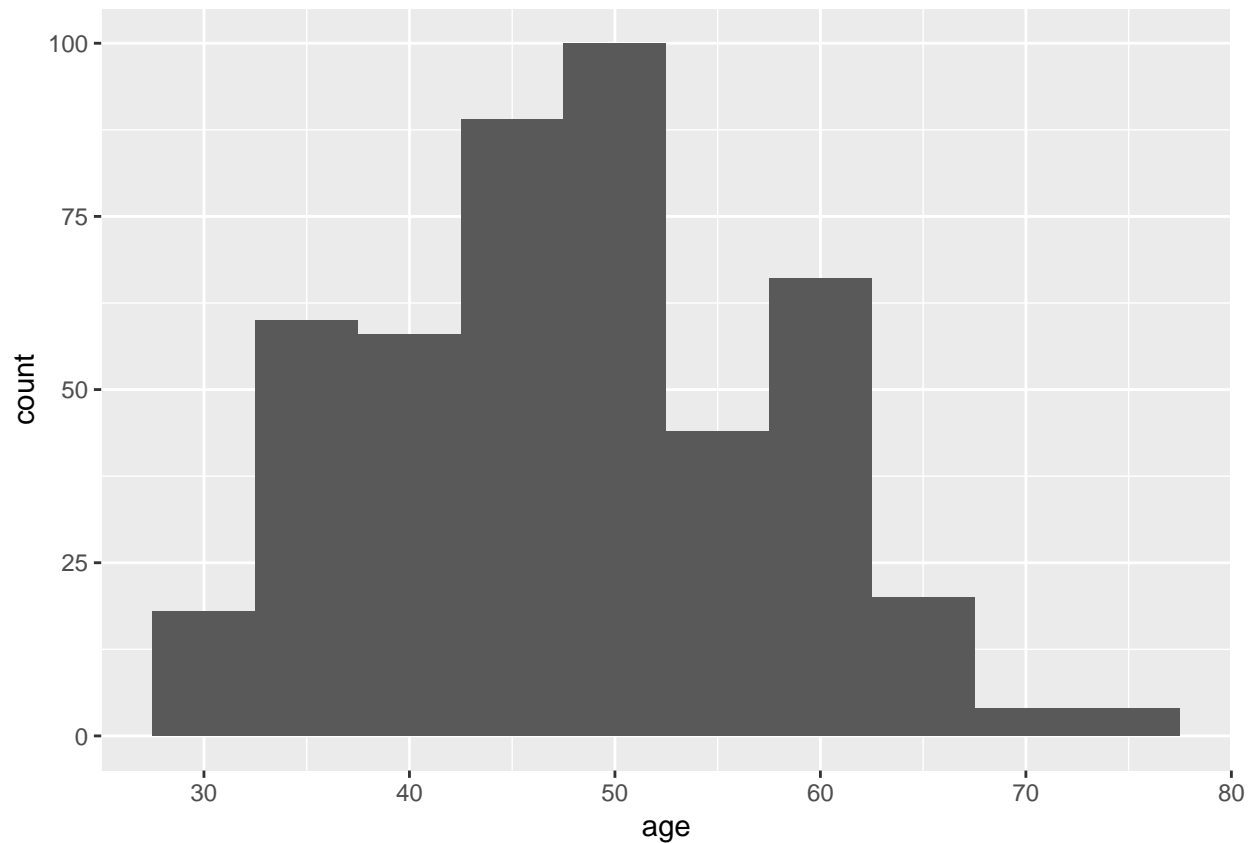
Selain menambahkan `aesthetic`, Anda juga dapat menambahkan `geometry` dengan cara menggunakan tanda `+`. Apa yang akan terjadi jika Anda menambahkan `geom_line` pada baris kode berikut? Menurut Anda data apakah yang cocok digunakan untuk jenis grafik tersebut?

```
ggplot(data = evals, aes(x = age, y = bpy_avg, colour = gender)) +
  geom_point() +
  geom_line()
```



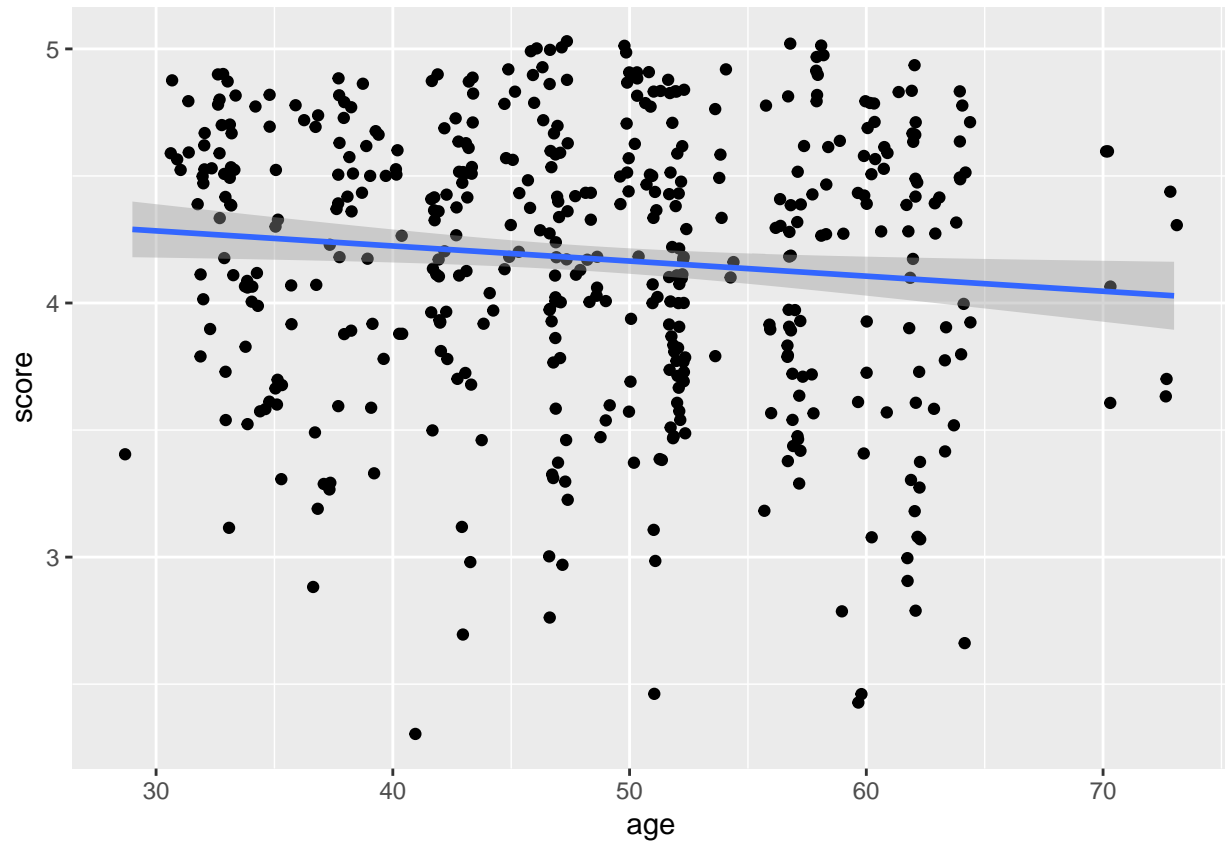
Pada contoh-contoh di atas, Anda menggunakan nilai `x` dan `y` sebagai `aesthetic`. Bagaimana jika Anda hanya menggunakan `x` saja? Jenis `geometry` apakah yang cocok digunakan jika hanya menggunakan satu nilai `x` tersebut? (Petunjuk: distribusi data)

```
ggplot(data = evals, aes(x = age)) +  
  geom_histogram(binwidth = 5)
```



Anda juga dapat membuat grafik yang berisi prediksi model dengan menggunakan `geom_smooth`. Buatlah grafik tersebut dengan menggunakan `geom_point()` dan `geom_smooth()`! Sumbu x adalah 'age' dan sumbu y adalah 'score'. (Pada contoh ini kita akan menggunakan metode regresi linear untuk model, sehingga argumen pada `geom_smooth()` diisi dengan 'method = "lm"').

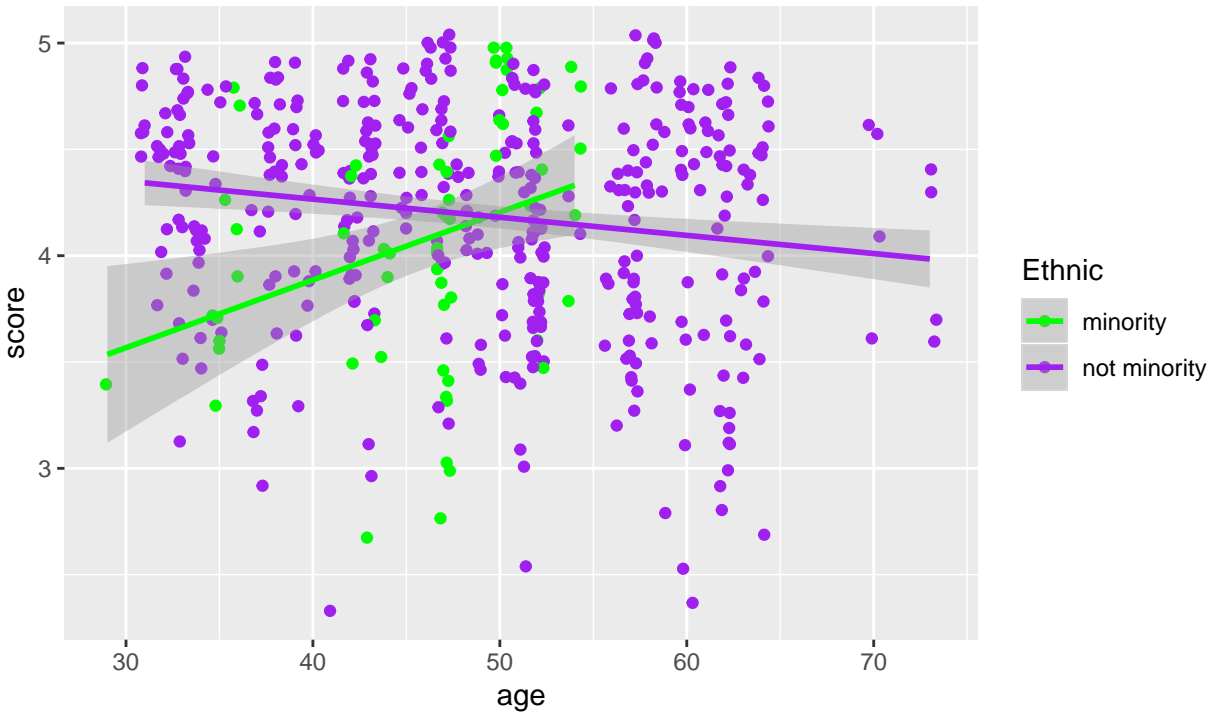
```
ggplot(data = evals, aes(x = age, y = score)) +  
  geom_jitter() +  
  geom_smooth(method = "lm")
```



Bagaimana jika Anda menambahkan argumen 'colour = ethnicity' ke dalam `aes()`? Apakah yang terjadi dan dapatkah Anda menjelaskannya? Anda dipersilahkan untuk mengganti label atau menambahkan judul pada grafik dengan menggunakan `labs`.

```
ggplot(data = evals, aes(x = age, y = score, colour = ethnicity)) +
  geom_jitter() +
  scale_colour_manual(values = c("minority" = "green", "not minority" = "purple")) +
  geom_smooth(method = "lm") +
  labs(
    x = "age",
    y = "score",
    title = "Graph of Score against Age",
    subtitle = "Subtitle",
    caption = "Source of data:...",
    colour = "Ethnic"
  )
```


Graph of Score against Age
Subtitle



Source of data:...