

Tugas Modul 9

Nico

1. Jalankan perintah berikut untuk membuat objek baru bernama “co2_wide”:

```
co2_wide <- data.frame(matrix(co2, ncol = 12)) %>%  
  setNames(1:12) %>%  
  mutate(year = as.character(1959:1997))  
head(co2_wide)
```

```
##           1           2           3           4           5           6           7           8           9          10          11  
## 1 315.42 320.42 321.05 320.09 326.60 333.14 336.37 336.68 344.79 353.41 356.06  
## 2 316.31 320.85 318.71 321.16 327.47 333.80 334.51 338.19 345.82 354.04 353.92  
## 3 316.50 320.45 317.66 322.74 327.58 333.43 332.60 339.44 347.25 353.62 352.05  
## 4 317.56 319.45 317.14 323.83 329.56 331.73 332.38 340.57 348.17 352.22 352.11  
## 5 318.13 317.25 318.70 324.26 329.90 329.90 333.75 341.44 348.74 350.27 353.64  
## 6 318.00 316.11 319.25 325.47 328.92 328.40 334.78 342.53 348.07 348.55 354.89  
##           12 year  
## 1 356.00 1959  
## 2 357.59 1960  
## 3 359.05 1961  
## 4 359.98 1962  
## 5 361.03 1963  
## 6 361.66 1964
```

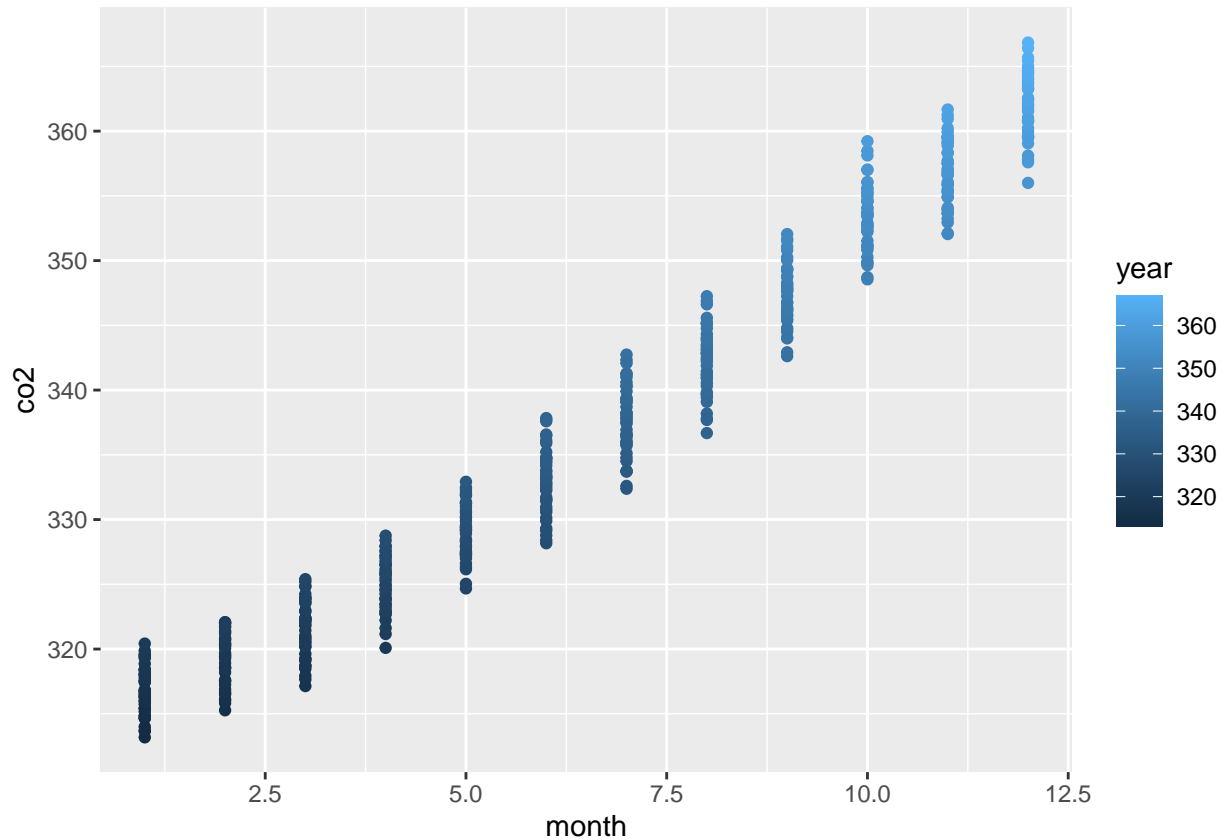
Mengubah dataset menjadi tidy

```
co2_tidy <- gather(co2_wide, month, year, convert = TRUE)  
head(co2_tidy)
```

```
##   month   year  
## 1     1 315.42  
## 2     1 316.31  
## 3     1 316.50  
## 4     1 317.56  
## 5     1 318.13  
## 6     1 318.00
```

2. Plot CO2 versus month dengan plot yang berbeda untuk setiap year menggunakan script ini:

```
co2_tidy %>% ggplot(aes(month, co2, color = year)) + geom_point(mapping = aes(x = month, y = year))
```



```
class(co2_tidy$month)
```

```
## [1] "integer"
```

Hasil dari pembenaran gather

```
co2_tidy <- gather(co2_wide, month, year, convert = TRUE)  
class(co2_tidy$month)
```

```
## [1] "integer"
```

Joining Table

1. Data frame Batting berisi statistik ofensif semua pemain selama beberapa tahun. Lakukan preview data, misalnya, tampilkan 10 batting teratas dengan menjalankan script ini:

```
library(Lahman)
```

```
## Warning: package 'Lahman' was built under R version 3.6.3
```

```
top <- Batting %>%  
  filter(yearID == 2016) %>%  
  arrange(desc(HR)) %>%  
  slice(1:10)  
top %>% as_tibble()
```

```
## # A tibble: 10 x 22  
##   playerID yearID stint teamID lgID      G    AB    R    H   X2B   X3B   HR  
##   <chr>      <int> <int> <fct> <fct> <int> <int> <int> <int> <int> <int> <int>  
## 1 trumbma~  2016     1  BAL   AL    159  613   94  157   27     1   47  
## 2 cruzne02  2016     1  SEA   AL    155  589   96  169   27     1   43  
## 3 daviskh~  2016     1  OAK   AL    150  555   85  137   24     2   42  
## 4 doziebr~  2016     1  MIN   AL    155  615  104  165   35     5   42  
## 5 encared~  2016     1  TOR   AL    160  601   99  158   34     0   42  
## 6 arenano~  2016     1  COL   NL    160  618  116  182   35     6   41  
## 7 cartech~  2016     1  MIL   NL    160  549   84  122   27     1   41  
## 8 frazito~  2016     1  CHA   AL    158  590   89  133   21     0   40  
## 9 bryankr~  2016     1  CHN   NL    155  603  121  176   35     3   39  
## 10 canoro01 2016     1  SEA   AL    161  655  107  195   33     2   39  
## # ... with 10 more variables: RBI <int>, SB <int>, CS <int>, BB <int>,  
## #   SO <int>, IBB <int>, HBP <int>, SH <int>, SF <int>, GIDP <int>
```

Script diatas hanya menampilkan ID, bukan nama pemain. Nama-nama pemain dapat dilihat pada tabel ini :

```
Master %>% as_tibble()
```

```
## # A tibble: 19,617 x 26  
##   playerID birthYear birthMonth birthDay birthCountry birthState birthCity  
##   <chr>      <int>      <int>    <int> <chr>      <chr>      <chr>  
## 1 aardsda~  1981         12      27  USA      CO      Denver  
## 2 aaronha~  1934          2       5  USA      AL      Mobile  
## 3 aaronto~  1939          8       5  USA      AL      Mobile  
## 4 aasedo01  1954          9       8  USA      CA      Orange  
## 5 abadan01  1972          8      25  USA      FL      Palm Bea~  
## 6 abadfe01  1985         12      17  D.R.    La Romana La Romana  
## 7 abadijo~  1850         11       4  USA      PA      Philadel~  
## 8 abbated~  1877          4      15  USA      PA      Latrobe  
## 9 abbeybe~  1869         11      11  USA      VT      Essex  
## 10 abbeych~  1866         10      14  USA      NE      Falls Ci~
```

```
## # ... with 19,607 more rows, and 19 more variables: deathYear <int>,
## #   deathMonth <int>, deathDay <int>, deathCountry <chr>, deathState <chr>,
## #   deathCity <chr>, nameFirst <chr>, nameLast <chr>, nameGiven <chr>,
## #   weight <int>, height <int>, bats <fct>, throws <fct>, debut <chr>,
## #   finalGame <chr>, retroID <chr>, bbrefID <chr>, deathDate <date>,
## #   birthDate <date>
```

Left Join untuk membuat yang berisi ID pemain, nama depan, nama belakang, dan jumlah home run (SDM). Simpan hasil `left_join` pada objek baru

```
tabel_batting <- Batting %>% select(playerID,HR)
tabel_master <- Master %>% select(playerID,nameFirst, nameLast,)
hasil_left_join <- left_join(tabel_master, tabel_batting, by = "playerID" )
# %>% select(Master$playerID, Master$nameFirst, Master$nameLast, Batting$HR)
head(hasil_left_join)
```

```
##   playerID nameFirst nameLast HR
## 1 aardsda01      David  Aardsma  0
## 2 aardsda01      David  Aardsma  0
## 3 aardsda01      David  Aardsma  0
## 4 aardsda01      David  Aardsma  0
## 5 aardsda01      David  Aardsma  0
## 6 aardsda01      David  Aardsma  0
```

2. Dengan menggunakan dataset yang sama, gunakan data frame “Salaries” untuk menambahkan informasi gaji masing-masing pemain ke tabel yang telah dibuat dalam latihan no 1. Perhatikan bahwa besar gaji berbeda setiap tahunnya. Pastikan untuk memfilter gaji pada tahun 2016, lalu gunakan `right_join`. Tampilkan hasil yang berisi: nama depan, nama belakang, tim, SDM, dan gaji.

```
tabel_batting_baru <- Batting %>% select(playerID,teamID,HR)
tabel_salaries <- Salaries %>% filter (yearID == 2016 ) %>% select(playerID,salary)
tabel_master_baru <- Master %>% select(playerID,nameFirst,nameLast)

hasil_right_join_1 <- right_join(tabel_master_baru,tabel_batting_baru, by = "playerID")
view(hasil_right_join_1)
hasil_right_join_2 <- right_join(hasil_right_join_1, tabel_salaries, by = "playerID")

head(hasil_right_join_2)
```

```
##   playerID nameFirst nameLast teamID HR salary
## 1 ahmedni01      Nick    Ahmed     ARI   1 521600
## 2 ahmedni01      Nick    Ahmed     ARI   9 521600
## 3 ahmedni01      Nick    Ahmed     ARI   4 521600
## 4 ahmedni01      Nick    Ahmed     ARI   6 521600
## 5 ahmedni01      Nick    Ahmed     ARI  16 521600
## 6 barreja01      Jake  Barrett     ARI   0 507500
```