Pertemuan 7

plirapli

2024-10-31

ggplot2 dan Data Wrangling

```
library(tidyverse)
```

```
## -- Attaching core tidyverse packages ------------------------ tidyverse 2.0.0 --
## v dplyr     1.1.4     v readr     2.1.5
## v forcats   1.0.0     v stringr   1.5.1
## v ggplot2   3.5.1     v tibble    3.2.1
## v lubridate 1.9.3     v tidyr     1.3.1
## v purrr     1.0.2
## -- 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
```

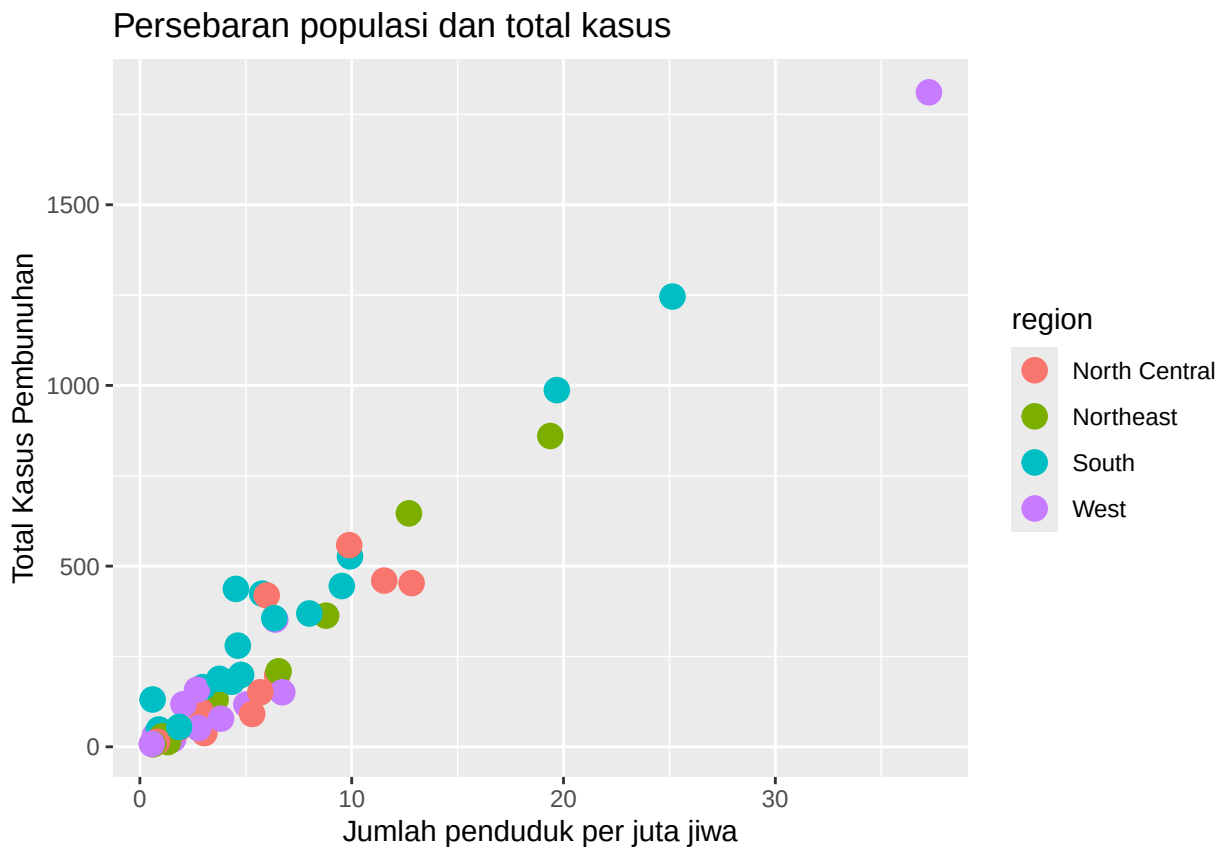
GGPLOT2

```
murders_dataset = read.csv("Dataset/murders_dataset.csv")
vg_dataset = read.csv("Dataset/vg_dataset.csv")
```

Scatter Plot

Membuat scatter plot antara populasi dan total kasus pembunuhan

```
ggplot(
  murders_dataset,
  aes(
    x = population / 10^6,
    y = total_murders,
    color = region
  )
) + geom_point(size = 4) + labs(
  x = "Jumlah penduduk per juta jiwa",
  y = "Total Kasus Pembunuhan",
  title = "Persebaran populasi dan total kasus"
)
```

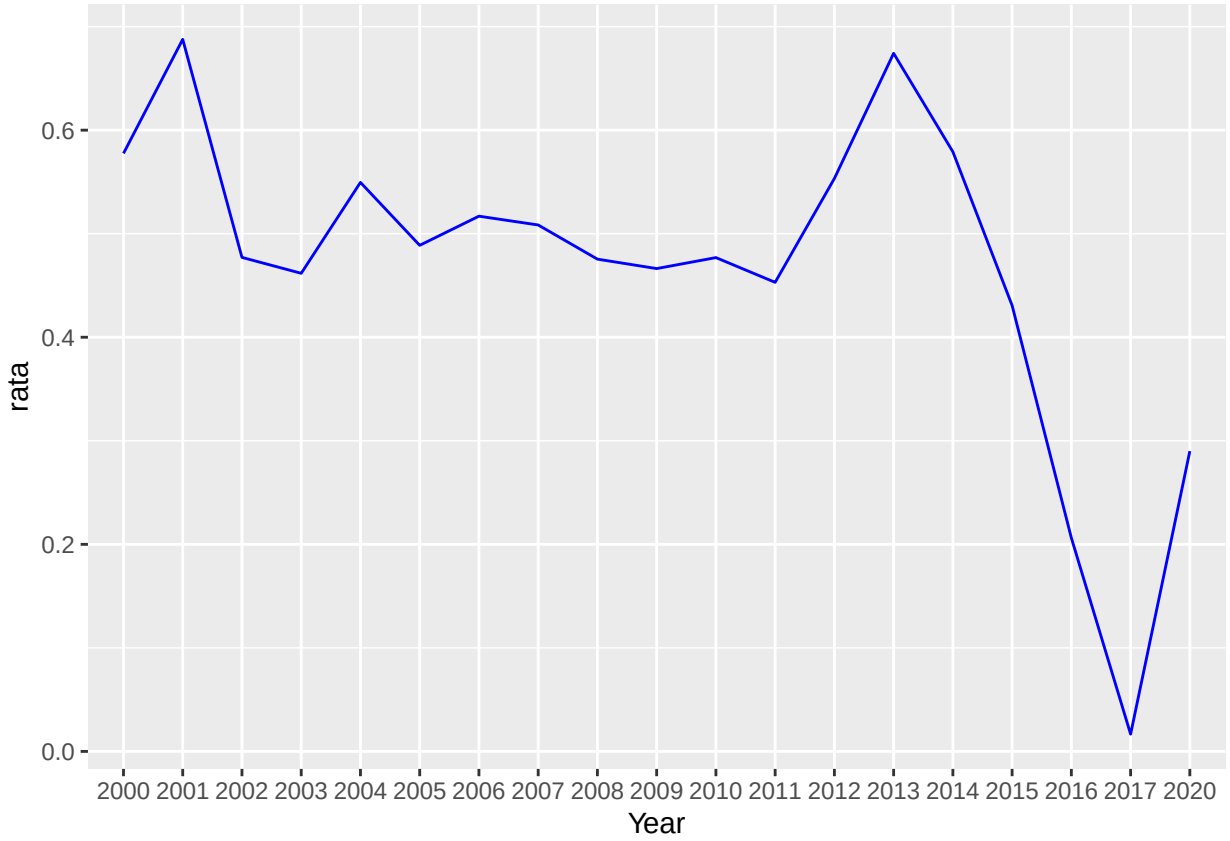


Line Chart dan Bar Chart

Grafik garis rata-rata penjualan video game per tahun

```
vg_dataset = vg_dataset %>% filter(Year >= 2000 & Year != "N/A")

hasil_rerata = vg_dataset %>% group_by(Year) %>% summarise(rata = mean(Global_Sales))

ggplot(
  hasil_rerata,
  aes(
    x = Year,
    y = rata,
    group = 1,
  )
) + geom_line(size = 0.5, color = "blue")
```

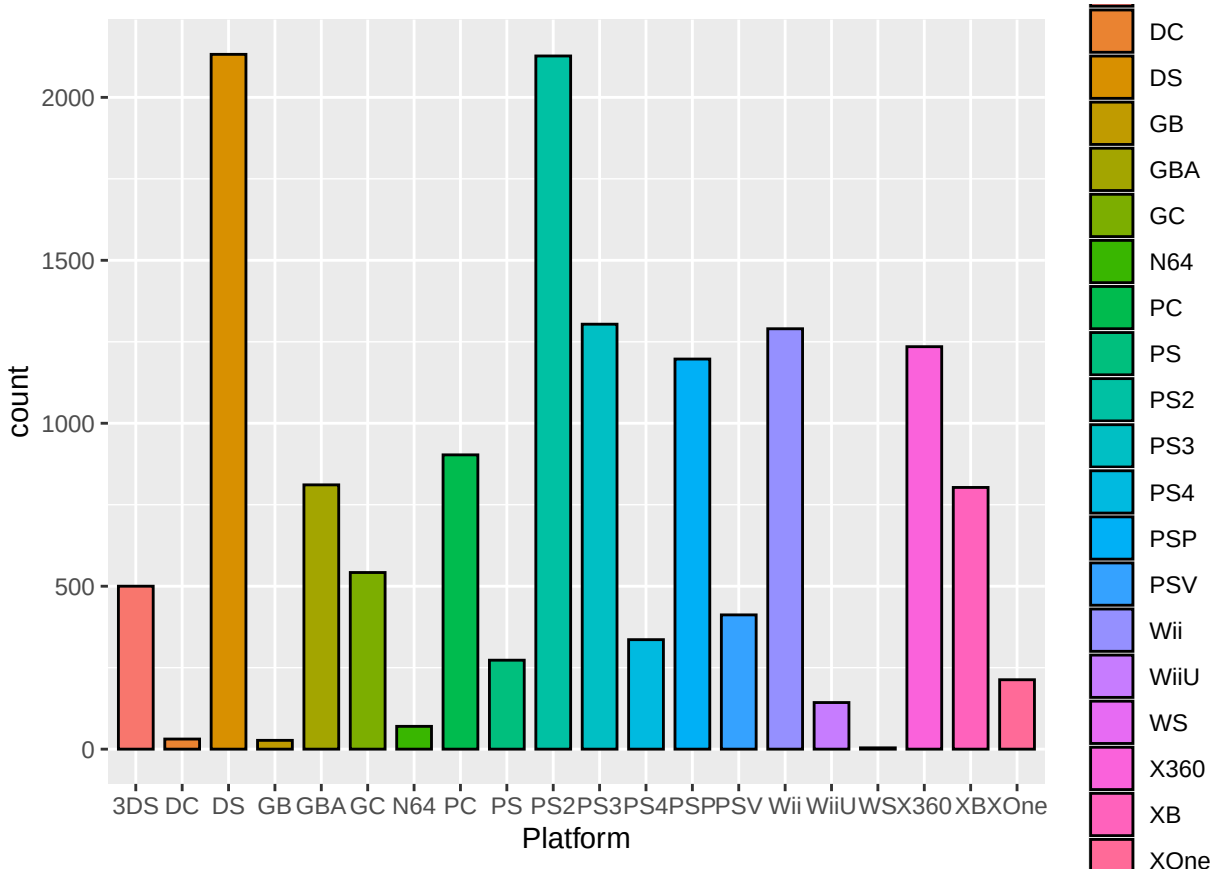
```
## Warning: Using 'size' aesthetic for lines was deprecated in ggplot2 3.4.0.
## i Please use 'linewidth' instead.
## This warning is displayed once every 8 hours.
## Call 'lifecycle::last_lifecycle_warnings()' to see where this warning was
## generated.
```



Histogram

Menghitung jumlah game yang rilis tiap platform

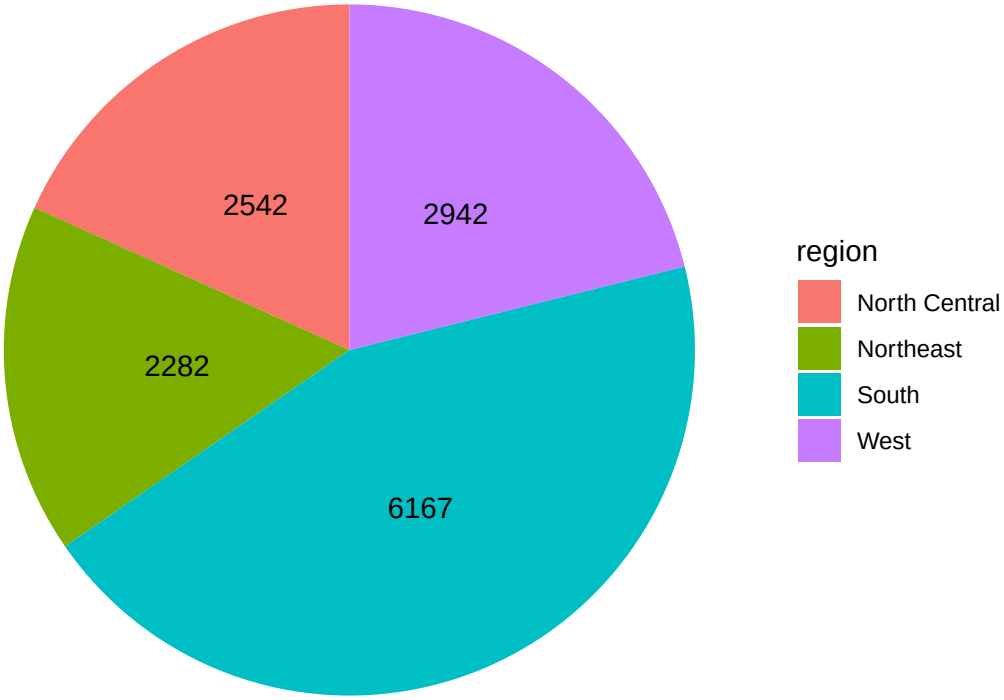
```
ggplot(
  vg_dataset,
  aes(
    x = Platform,
    fill = Platform
  )
) + geom_bar(width = 0.75, color = "black")
```



Pie chart

Perbandingan jumlah kasus pembunuhan tiap region

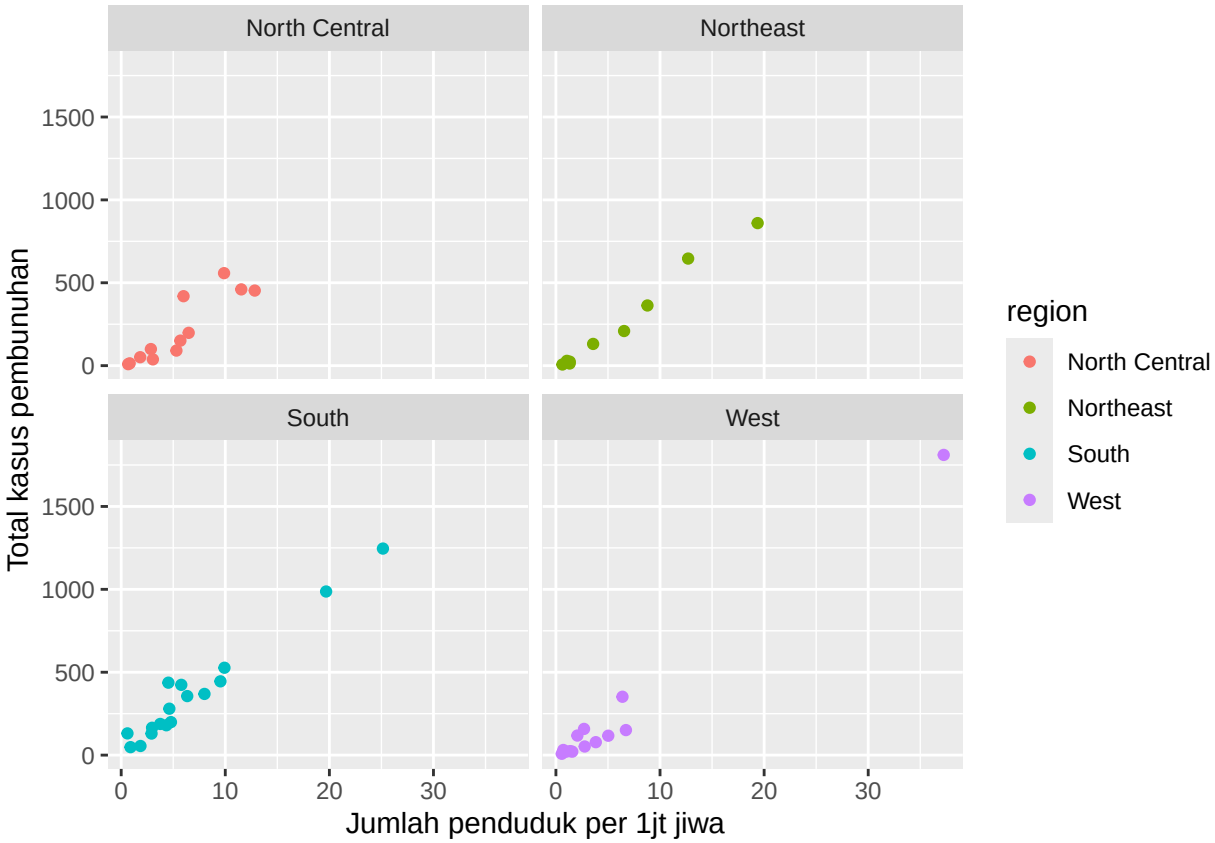
```
total_kasus = murders_dataset %>%
  group_by(region) %>%
  summarise(total = sum(total_murders))

ggplot(
  total_kasus,
  aes(
    x = total,
    y = "",
    fill = region
  )
) + geom_col() + coord_polar() + theme_void() + geom_text(
  aes(label = total),
  position = position_stack(vjust = 0.5),
)
```



Visualisasi berdasarkan Kelompok

Membuat scatter plot antara populasi dan total kasus berdasarkan region

```
ggplot(
  murders_dataset,
  aes(
    x = population / 10^6,
    y = total_murders,
    color = region
  )
) + geom_point() + facet_wrap(~region) + labs(
  x = "Jumlah penduduk per 1jt jiwa",
  y = "Total kasus pembunuhan"
)
```


North Central
Northeast
South
West
Total kasus pembunuhan
0
500
1000
1500
0
10
20
30
Jumlah penduduk per 1jt jiwa
region
North Central
Northeast
South
West