Pert 7 - Visualisasi Data dengan R Package : ggplot2

YOHANES FEBRYAN KANA NYOLA_123220198

2024-11-06

Visualisasi Data dengan ggplot2

1. Geom_point

Import Library

```
library(tidyverse)
```

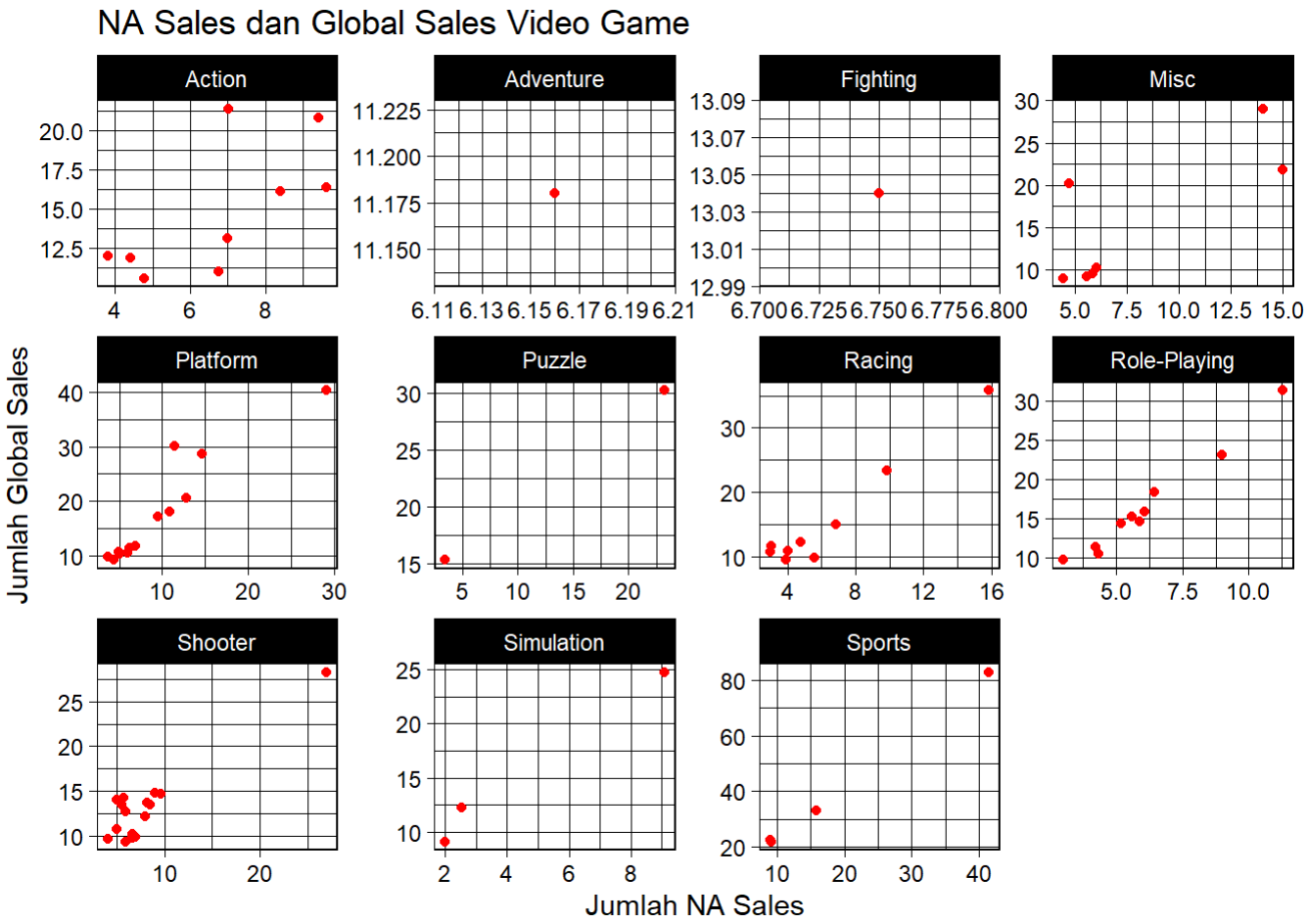
```
## ── Attaching core tidyverse packages ──────────────────────── tidyverse 2.0.0 ──
## ✓ dplyr     1.1.4     ✓ readr     2.1.5
## ✓ forcats   1.0.0     ✓ stringr   1.5.1
## ✓ ggplot2   3.5.1     ✓ tibble    3.2.1
## ✓ lubridate 1.9.3     ✓ tidyr     1.3.1
## ✓ purrr     1.0.2
## ── Conflicts ────────────────────────────────────────── tidyverse_conflicts() ──
## ✖ dplyr::filter() masks stats::filter()
## ✖ dplyr::lag()    masks stats::lag()
## ℹ Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to be
come errors
```

```
library(readxl)
library(dslabs)
```

Importing Data

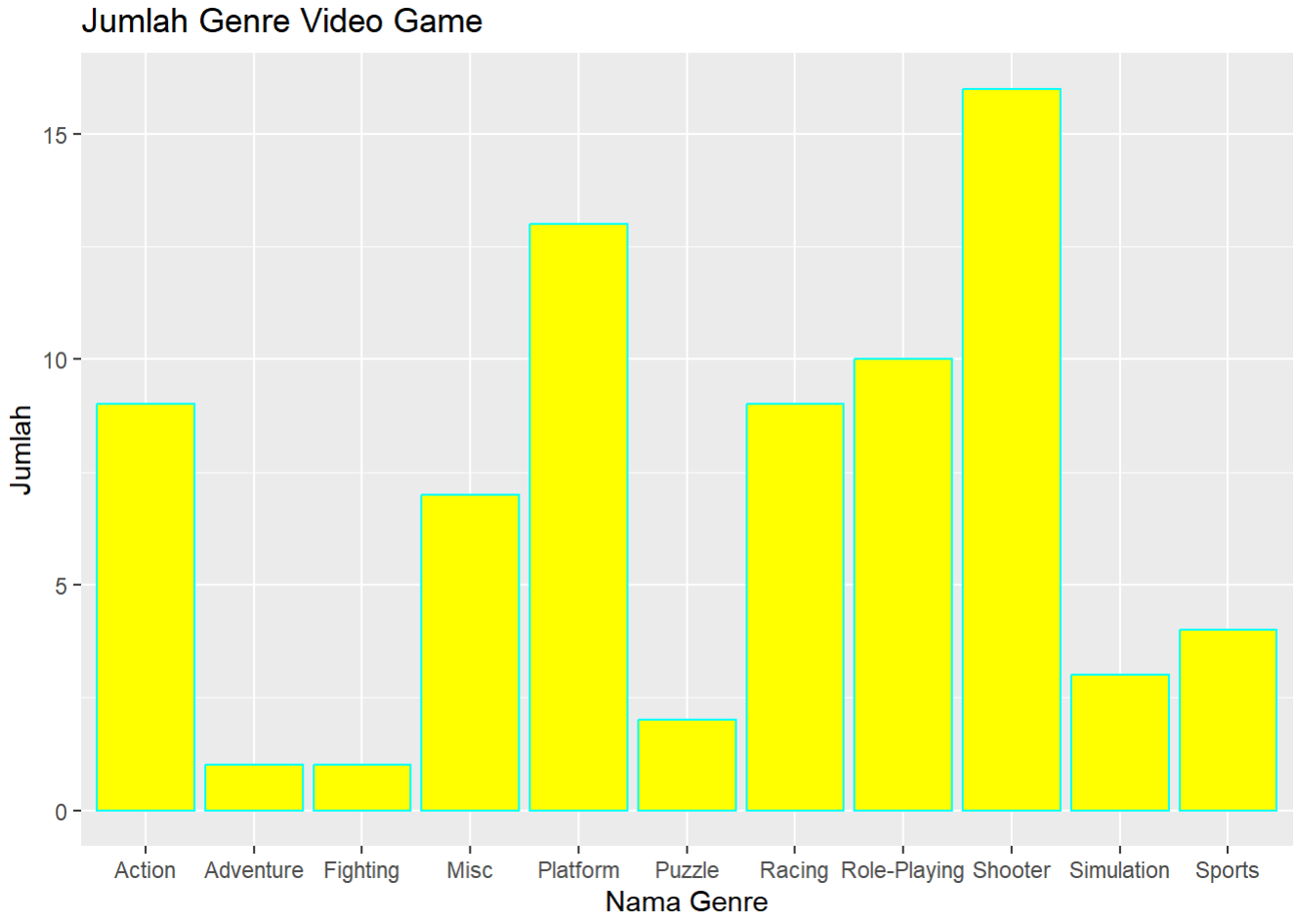
```
data_orang = read.csv("data raw/data_orang.csv")
# View(data_orang)
data_vg = read_excel("data raw/data_video_game.xlsx")
# View(data_vg)
```

```
ggplot(
  data = data_vg,
  aes(
    x = NA_Sales,
    y = Global_Sales,
    color = Genre
  )
) + geom_point(
  color = "red"
) + labs(
  title = "NA Sales dan Global Sales Video Game",
  x = "Jumlah NA Sales",
  y = "Jumlah Global Sales"
) + theme_linedraw() + facet_wrap(~Genre, scales = "free")
```



2. Geom_bar

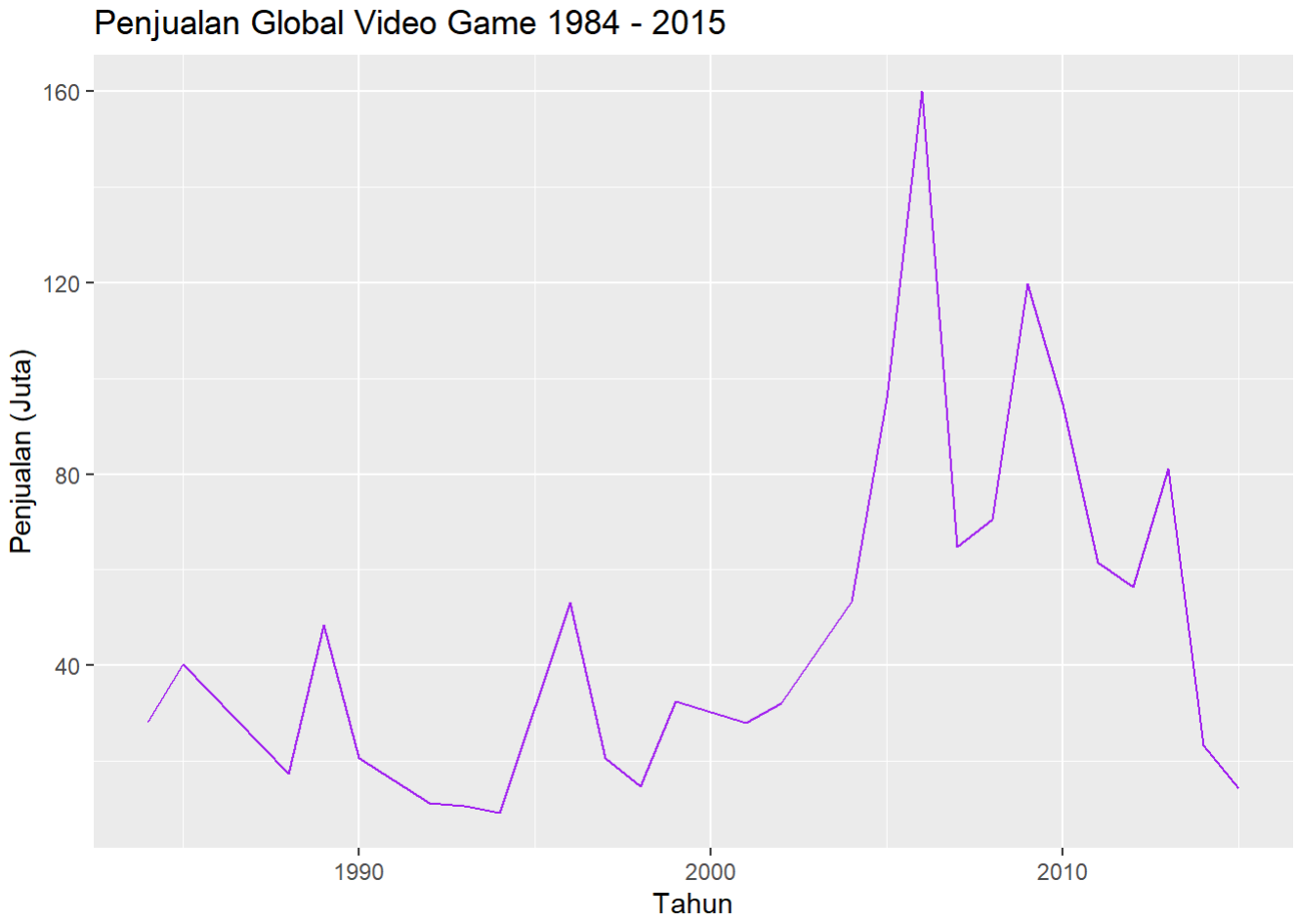
```
ggplot(
  data = data_vg,
  aes(
    x = Genre,
  )
) + geom_bar(
  color = "cyan",
  fill = "yellow"
) + labs(
  title = "Jumlah Genre Video Game",
  x = "Nama Genre",
  y = "Jumlah"
)
```



3. Geom_line

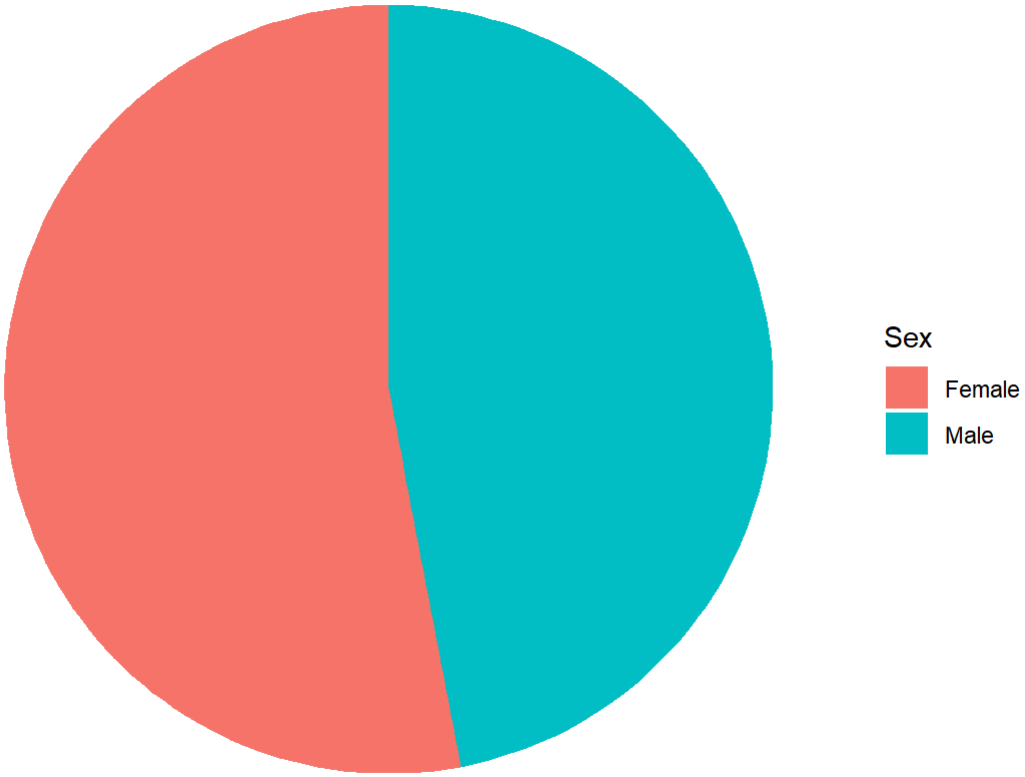
```
new_data_vg = data_vg %>%
  group_by(Year) %>%
  summarize(Total_NA =  sum(NA_Sales),
            Total_Global = sum(Global_Sales))

ggplot(
  data = new_data_vg,
  aes(
    x = Year,
    y = Total_Global
  )
) + geom_line(
  color = "purple"
) + labs(
  title = "Penjualan Global Video Game 1984 - 2015",
  x = "Tahun",
  y = "Penjualan (Juta)"
)
```



4. Pie_Chart

```
new_data_orang = data_orang %>% group_by(Sex) %>% summarize(Total = n())
ggplot(
  new_data_orang,
  aes(
    x = Total,
    y = "",
    fill = Sex
  )
) + geom_col() + coord_polar() + theme_void()
```



Data Wraggling

```
path = system.file("extdata", package = "dslabs")
filename = file.path(path, "fertility-two-countries-example.csv")
wide_data = read.csv(filename)
```

1. Gather

```
gathered_data = wide_data %>% gather(year, fertility_rate, -country, convert = TRUE)

gathered_data
```

2. Spread

```
spread_data = gathered_data %>% spread(year, fertility_rate)

spread_data
```

3. Joining Data

```
data(murders)
data(polls_us_election_2016)

join_data = left_join(murders, polls_us_election_2016, by = "state") %>%
  select(state,samplesize) %>% group_by(state) %>%
  summarise(totalsamplesize = sum(samplesize, na.rm = TRUE))

join_data
```

```
tab1 = slice(murders, 1:6) %>% select(state, population)

tab1
```

```
tab2 = results_us_election_2016 %>%
  filter(state %in% c("Alabama", "Alaska", "Arizona",
                      "California", "Connecticut", "Delaware")) %>%
  select(state, electoral_votes) %>% rename(ev = electoral_votes)

tab2
```

3.1 Left Join

```
left_join = left_join(tab1, tab2, by = "state")
left_join
```

3.2 Right Join

```
right_join = right_join(tab1, tab2, by = "state")
right_join
```

3.3 Inner Join

```
inner_join = inner_join(tab1, tab2, by = "state")
inner_join
```

3.4 Full Join

```
full_join = full_join(tab1, tab2, by = "state")
full_join
```

3.5 Semi Join

```
semi_join = semi_join(tab1, tab2, by = "state")
semi_join
```

3.6 Anti Join

```
anti_join = anti_join(tab2, tab1, by = "state")
anti_join
```