Tugas Praktikum Komstat

Andrian Agustinus Lumban Gaol

121450090-RA

2023-09-27

# Data Univariat

## Data

library(readr)  
df <- read\_csv("C:/Users/Andrian L Gaol/Downloads/Heart (1).csv")

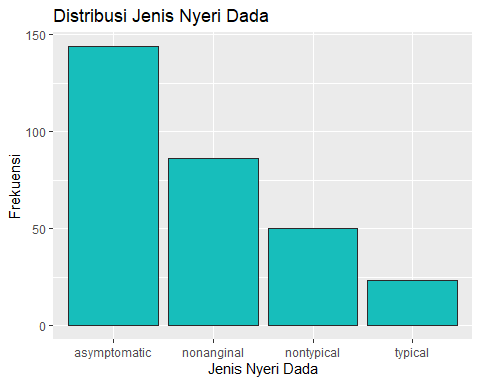
## New names:  
## Rows: 303 Columns: 15  
## ── Column specification  
## ──────────────────────────────────────────────────────── Delimiter: "," chr  
## (3): ChestPain, Thal, AHD dbl (12): ...1, Age, Sex, RestBP, Chol, Fbs, RestECG,  
## MaxHR, ExAng, Oldpeak,...  
## ℹ Use `spec()` to retrieve the full column specification for this data. ℹ  
## Specify the column types or set `show\_col\_types = FALSE` to quiet this message.  
## • `` -> `...1`

## 1. Tampilkan bar chart dari variabel ChestPain dengan menggunakan package ggplot2.

library(ggplot2)

## Warning: package 'ggplot2' was built under R version 4.2.3

ggplot(df, aes(x=ChestPain))+ # menentukan Chespain menjadi variable x  
 geom\_bar(fill="#17BEBB", color="#2E282A")+ # membuat pot dan menambahkan warna pada plot tersebut  
 labs(x= 'Jenis Nyeri Dada',  
 y = 'Frekuensi',  
 title= 'Distribusi Jenis Nyeri Dada') # menambahkan label pada sumbu x dan y serta membetikan judul pada pot



**Analisis**

Dari plot yang dihasilkan dapar dilihat bahwa chespain dibagi menjadi 4 kategori dan dapat dilihat dari frekuensi bahwa asymptomatic memiliki frekuensi terbanyak dan typica menempati frekuensi terrendah

## 2. Tampilkan bar chart seperti pada gambar diatas, ganti skala y dengan persentase

library(dplyr)

## Warning: package 'dplyr' was built under R version 4.2.3

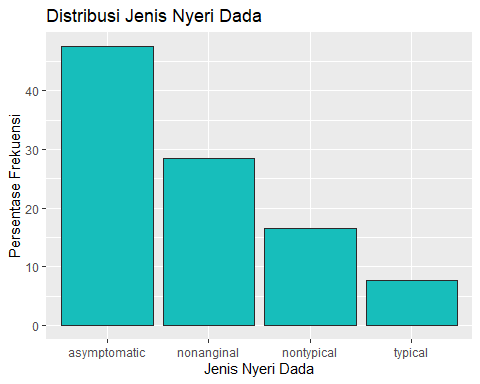
##   
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':  
##   
## filter, lag

## The following objects are masked from 'package:base':  
##   
## intersect, setdiff, setequal, union

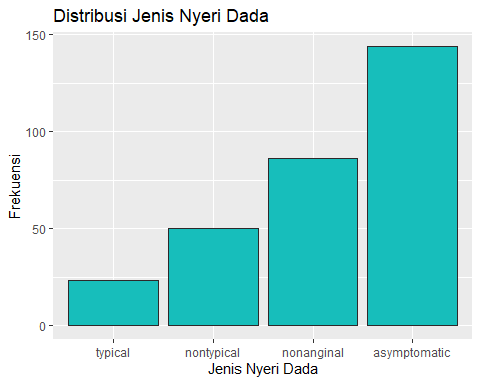
dfa <- df %>%  
 arrange(desc(ChestPain)) # mengurutkan nilai tertinggi di kolom ChestPain hingga nilai terendah.  
  
ggplot(dfa, aes(x = ChestPain)) +  
 # menghitung persentase frekuensi pada setiap kategori chestpain  
 geom\_bar(aes(y = (..count..)/sum(..count..) \* 100), fill="#17BEBB", color="#2E282A") +  
 labs(x= 'Jenis Nyeri Dada',  
 y = 'Persentase Frekuensi',title='Distribusi Jenis Nyeri Dada')

## Warning: The dot-dot notation (`..count..`) was deprecated in ggplot2 3.4.0.  
## ℹ Please use `after\_stat(count)` instead.  
## This warning is displayed once every 8 hours.  
## Call `lifecycle::last\_lifecycle\_warnings()` to see where this warning was  
## generated.



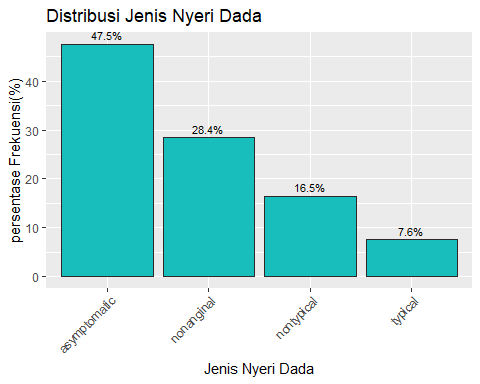
## 3. Urutkan bar chart dari yang terkecil ke terbesar. Gunakan package dplyr dan ggplot

library(dplyr)  
dfb <- df %>% count(ChestPain) # menghitung jumlah frekuensi setiap kategori dalam kolom chestpain dari data frame df  
  
ggplot(dfb,aes(x=reorder(ChestPain,n),y=n))+ # reorder untuk mengurutkan frekuensi chespain dari yang terrendah  
 geom\_bar(stat="Identity",fill="#17BEBB",color ='#2E282A')+  
 labs(x= 'Jenis Nyeri Dada',  
 y = 'Frekuensi',  
 title= 'Distribusi Jenis Nyeri Dada')



## 4. Berikan label dengan nilai numerik pada setiap chart, dengan mengurutkan dari presentase terbesar ke terkecil, pada sumbu X buat label miring dengan 45 derajat

# Menghitung proporsi  
df\_persen <- df %>%  
 group\_by(ChestPain) %>% # Mengelompokkan data berdasarkan kolom ChestPain  
 summarise(Count = n()) %>% # Menghitung jumlah frekuensi untuk setiap kelompok  
 mutate(Percentage = (Count / sum(Count)) \* 100) %>% # Menghitung persentase frekuensi  
 arrange(Percentage) # Mengurutkan data berdasarkan persentase  
  
  
ggplot(df\_persen, aes(x = reorder(ChestPain, -Percentage),   
 y = Percentage)) +  
 geom\_bar(stat="Identity",fill="#17BEBB",color ='#2E282A') +   
 geom\_text(aes(label = sprintf("%.1f%%", Percentage)), vjust = -0.5, size = 3) + # Mengatur label persentase  
 labs(x = 'Jenis Nyeri Dada',   
 y = 'persentase Frekuensi(%)',   
 title = 'Distribusi Jenis Nyeri Dada') +  
 theme(axis.text.x = element\_text(angle = 45, hjust = 1)) # Mengatur label sumbu X miring 45 derajat

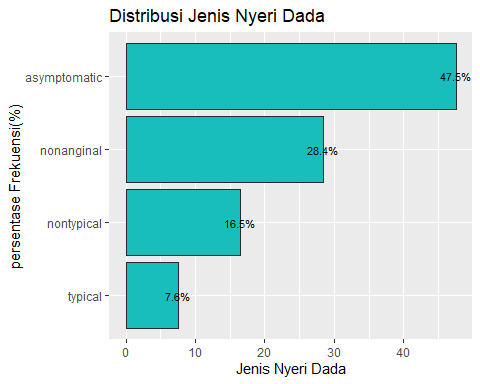


**Analisis**

Dari plot yang dihasilkan dapat di tinjau lebih detail terkait persentase frekuensi tiap kelompok data pada chespain

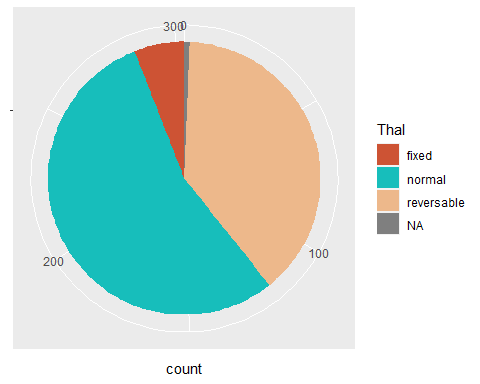
## 5. Transpose sumbu x dan sumbu y

ggplot(df\_persen, aes(x = Percentage, y = reorder(ChestPain, Percentage))) +  
 geom\_bar(stat="Identity",fill="#17BEBB",color ='#2E282A') +  
 geom\_text(aes(label = sprintf("%.1f%%", Percentage)), vjust = 0.5, size = 3) +  
 labs(x = 'Jenis Nyeri Dada',   
 y = 'persentase Frekuensi(%)',   
 title = 'Distribusi Jenis Nyeri Dada') +  
 theme(axis.text.y = element\_text(angle = 0, hjust = 1)) # Memiringkan label sumbu y jika perlu



## 6. Buat pie chart berdasarkan Thal

# Membuat vektor warna kustom  
warna <- c("#CD5334", "#17BEBB", "#EDB88B", "#FAD8D6", "#2E282A")  
  
# Membuat plot pie chart  
ggplot(df, aes(x = "", fill = Thal)) +   
 geom\_bar(width = 1) +   
 coord\_polar(theta = "y") +   
 labs(x = NULL) +   
 scale\_fill\_manual(values = warna)

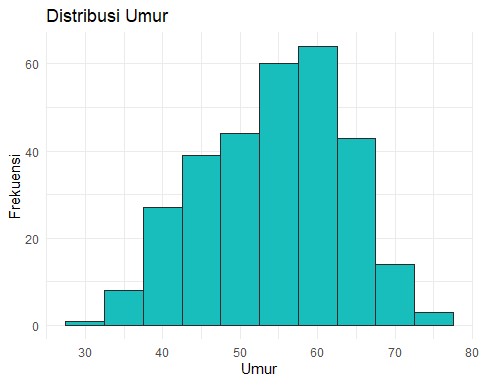


**Analisis**

pada pie chart diatas dapar dilihat bahwa mayoritas pasien kondisi thal(kelainan darah) normal, namund apat dilihat ada pasien yang belom di identifikasi, yang kemungkinan ada data yang kosong atau data pasien yang tidak melampirkan status thal pada bagian yang tidak dapat di identifikasi tersebut

## 7. Buat histogram berdasarkan Age

ggplot(df, aes(x = Age)) +  
 geom\_histogram(binwidth = 5, fill = "#17BEBB", color = "#2E282A") + # membuat histogram dengan lebar interfal adalah 5 (binwidth=5)  
 labs(x = "Umur",   
 y = "Frekuensi",   
 title = "Distribusi Umur") +   
 theme\_minimal()

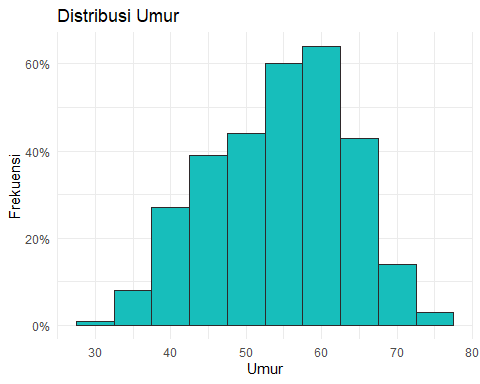


**Analisis**

pada histogram yang di hasilkan dapat dilihat pesebaran data mayoritas ada pada rentang 55-60 tahun

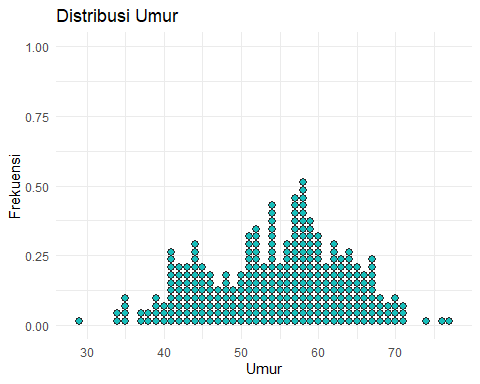
## 8. Buat histogram berdasarkan Age dimana sumbu y dalam persen

ggplot(df, aes(x = Age)) +  
 geom\_histogram(binwidth = 5, fill = "#17BEBB", color = "#2E282A") +  
 labs(x = "Umur",   
 y = "Frekuensi",   
 title = "Distribusi Umur") +   
 scale\_y\_continuous(labels = scales::percent\_format(scale = 1)) + # Mengatur label sumbu y menjadi format persentase  
 theme\_minimal()



## 9. Buatlah diagram titik berdasarkan Age

ggplot(df, aes(x = Age)) +   
 geom\_dotplot(fill = "#17BEBB", color = "#2E282A", binwidth = 1) + # Interfal data yang di gunakan adalah 1, agar diagram lebih akurat/detail  
 theme\_minimal()+  
 labs(x = "Umur",   
 y = "Frekuensi",   
 title = "Distribusi Umur")# Memberi nama label sumbu x dan y serta label subjudul



**Analisis**

pada dotplot yang di hasilkan dapat dilihat lebih detail pesebaran data mayoritas, dan mayoritas umur ada pada rentang 55-60 tahun

# Data Bivariat

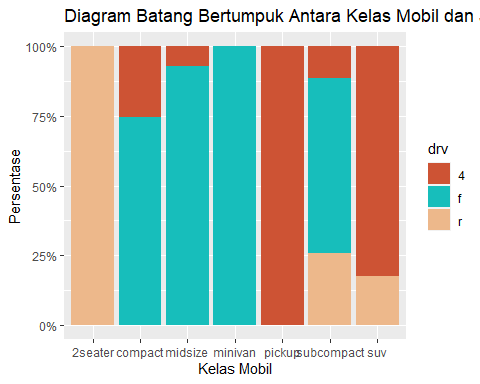
## Plot Diagram Variabel Kategorik Vs Kategorik

df2 = mpg  
str(df2)

## tibble [234 × 11] (S3: tbl\_df/tbl/data.frame)  
## $ manufacturer: chr [1:234] "audi" "audi" "audi" "audi" ...  
## $ model : chr [1:234] "a4" "a4" "a4" "a4" ...  
## $ displ : num [1:234] 1.8 1.8 2 2 2.8 2.8 3.1 1.8 1.8 2 ...  
## $ year : int [1:234] 1999 1999 2008 2008 1999 1999 2008 1999 1999 2008 ...  
## $ cyl : int [1:234] 4 4 4 4 6 6 6 4 4 4 ...  
## $ trans : chr [1:234] "auto(l5)" "manual(m5)" "manual(m6)" "auto(av)" ...  
## $ drv : chr [1:234] "f" "f" "f" "f" ...  
## $ cty : int [1:234] 18 21 20 21 16 18 18 18 16 20 ...  
## $ hwy : int [1:234] 29 29 31 30 26 26 27 26 25 28 ...  
## $ fl : chr [1:234] "p" "p" "p" "p" ...  
## $ class : chr [1:234] "compact" "compact" "compact" "compact" ...

### 1. Plot menggunakan diagram batang bertumpuk, hubungan antara kelas mobil dengan jenis penggerak ( roda depan, roda belakang, atau penggerak 4 roda). (gunakan package “ggplot2”)

ggplot(mpg, aes(x = class, fill = drv)) +  
 geom\_bar(position = "fill") + # Fill digunakan agar menghasilkan diagram batang bertumpuk  
 labs(x = "Kelas Mobil",   
 y = "Persentase",  
 title = "Diagram Batang Bertumpuk Antara Kelas Mobil dan Jenis Penggerak") +  
 scale\_y\_continuous(labels = scales::percent\_format(scale = 100)) + # Mengatur format sumbu y sebagai persentase  
 scale\_fill\_manual(values = warna)



**Analisi**

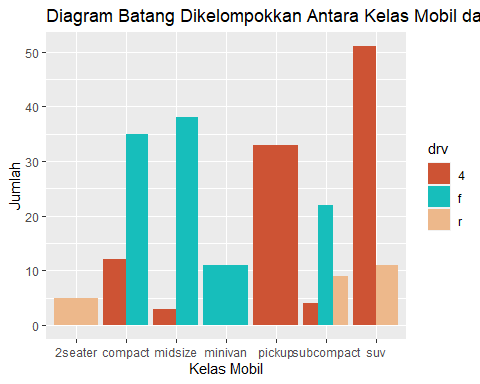
dari plot yang di hasilkan dapat dilihat persentase dari kelas mobil dari jenis penggeraknya. berikut penjelasan klasifikasi pengegrak (drv/drive train) - “4” (Four-Wheel Drive) => penggerak depan belakang atau 4 roda - “f” (Front-Wheel Drive) => penggerak depan - “r” (Rear-Wheel Drive) => penggerrak belakang

Dapat dilihat kelas mobil - 2seater hanya memakai penggerak belakang - compact dan midsize di dominasi penggerak belakang dan ada sebagian kecil memakai penggerak depan belakang - minivan hanya memakai penggerak depan - pickup hanya memakai penggerak depan belakang (4roda) - subcompact memakai ketiga jenis penggerak namun paling banyak memakai penggerka depan, kemudian penggerka belakang dan yang paling sedikit adalah penggerak depan belakang - suv i dominasi penggerak depan belakang dan ada sebagian kecil memakai penggerak penggerrak belakang

### 2. Plot menggunakan diagram batang yang dikelompokkan, hubungan antara kelas mobil

dengan jenis penggerak ( roda depan, roda belakang, atau penggerak 4 roda). (gunakan package “ggplot2”)

ggplot(mpg, aes(x = class, fill = drv)) +  
 geom\_bar(position = "dodge") + # Dodge digunakan untuk membuat grafik batang yang mengelompokkan data dalam baris yang berbeda secara berdampingan  
 labs(x = "Kelas Mobil",   
 y = "Jumlah",  
 title = "Diagram Batang Dikelompokkan Antara Kelas Mobil dan Jenis Penggerak") +  
 scale\_fill\_manual(values = warna)



**Analisi**

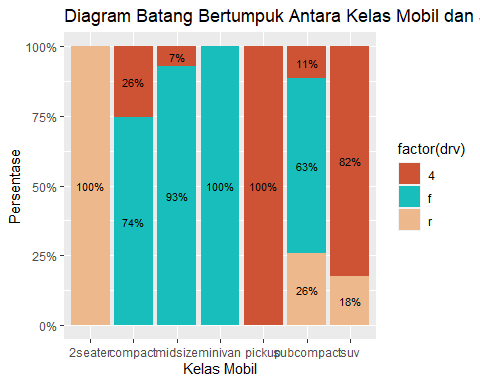
dari plot yang di hasilkan, dapat dilihat dengan lebih jelas tentang jumlah dari kelas mobil dari jenis penggeraknya. dari visualisasi secara tidak langsung menunjukkan jumlah data setiap kelas mobil

### 3. Pada nomor 1, tambahkan persentasenya

# Memproses dan menghitung data untuk plot  
plotdata <- mpg %>%  
 group\_by(class, drv) %>%  
 summarize(n = n()) %>%  
 mutate(pct = n / sum(n), # Menghitung persentase  
 lbl = scales::percent(pct, scale = 100)) # Format pada label persentase

## `summarise()` has grouped output by 'class'. You can override using the  
## `.groups` argument.

# Membuat plot menggunakan ggplot2  
ggplot(plotdata,   
 aes(x = factor(class),  
 y = pct,  
 fill = factor(drv))) +   
  
 geom\_bar(stat = "identity", # Identity digunakan untuk menggambar diagram batang dengan mengikuti nilai yang diberikan dalam dataset  
 position = "fill") + # Fill digunakan agar menghasilkan diagram batang bertumpuk  
  
 geom\_text(aes(label = lbl), # Menambahkan label berdasarkan lbl  
 size = 3,  
 position = position\_stack(vjust = 0.5)) + # Mengatur posisi label  
   
 labs(x = "Kelas Mobil",   
 y = "Persentase",   
 title = "Diagram Batang Bertumpuk Antara Kelas Mobil dan Jenis Penggerak") +   
 scale\_y\_continuous(labels = scales::percent\_format(scale = 100)) + # Mengatur format sumbu y sebagai persentase  
 scale\_fill\_manual(values = warna)



**Anaisis**

dari plot yang di hasilkan, dapat dilihat dengan lebih jelas tentang persentase kelas mobil dari jenis penggeraknya.

## Plot Diagram Variabel Kontinu Vs Kontinu

### Lakukan dot plot terhadap dataset “Salaries” antara pengalaman dengan gaji dari Professor PT

library(carData)   
df3 <- Salaries   
str(df3)

## 'data.frame': 397 obs. of 6 variables:  
## $ rank : Factor w/ 3 levels "AsstProf","AssocProf",..: 3 3 1 3 3 2 3 3 3 3 ...  
## $ discipline : Factor w/ 2 levels "A","B": 2 2 2 2 2 2 2 2 2 2 ...  
## $ yrs.since.phd: int 19 20 4 45 40 6 30 45 21 18 ...  
## $ yrs.service : int 18 16 3 39 41 6 23 45 20 18 ...  
## $ sex : Factor w/ 2 levels "Female","Male": 2 2 2 2 2 2 2 2 2 1 ...  
## $ salary : int 139750 173200 79750 115000 141500 97000 175000 147765 119250 129000 ...

levels(df3$rank)

## [1] "AsstProf" "AssocProf" "Prof"

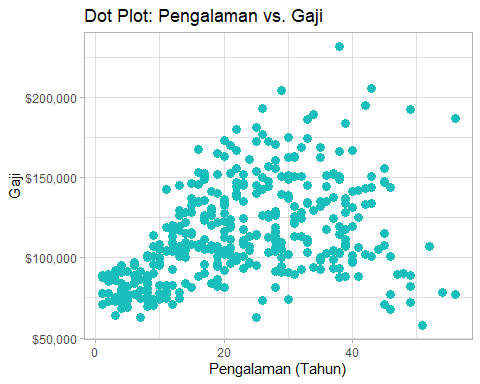
#### Dot Plot: Pengalaman vs. Gaji semua rank

library(scales)

##   
## Attaching package: 'scales'

## The following object is masked from 'package:readr':  
##   
## col\_factor

ggplot(df3, aes(x = yrs.since.phd, y = salary)) + # data yang di ambil meliputi semua kategori rank  
 geom\_point(size = 3, color = "#17BEBB") +   
 scale\_y\_continuous(labels = dollar\_format()) + # Mengubah format sallary menjadi uang dalam dolar  
 labs(x = "Pengalaman (Tahun)",   
 y = "Gaji",  
 title = "Dot Plot: Pengalaman vs. Gaji") +   
 theme\_light()

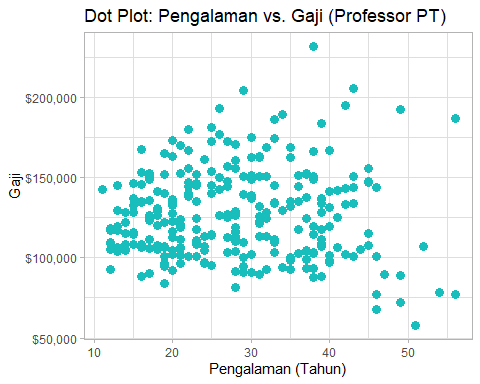


**Analisis**

dari dotplot yang di tunjukkan dapat dilihat bahwa semakin besar pengalaman, gaji cenderung naik, namun tidak selalu naik, dapat dilihat beberapa data pada pengalaman diatas 20 tahun masih memiliki gaji di bawah $ 100,000, namun dapat dilihat gaji tertinggi juga berada pada rentang pengalaman 30-40 tahun

#### Dot Plot: Pengalaman vs. Gaji khusus Professor (prof)

ggplot(data = subset(df3, rank == "Prof"), aes(x = yrs.since.phd, y = salary)) + # data yang di ambil berfokus pada data dengan kategori rank prof  
 geom\_point(size = 3, color = "#17BEBB") +   
 scale\_y\_continuous(labels = dollar\_format()) + # Mengubah format sallary menjadi uang dalam dolar  
 labs(x = "Pengalaman (Tahun)",   
 y = "Gaji",  
 title = "Dot Plot: Pengalaman vs. Gaji (Professor PT)") +   
 theme\_light()

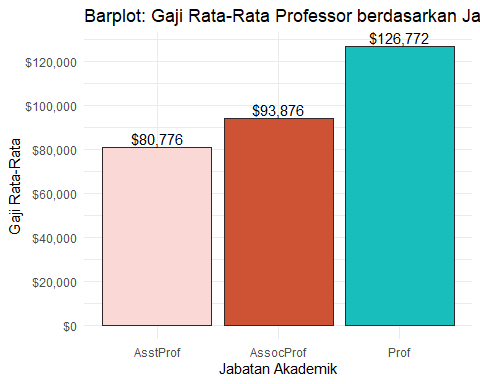


**Analisis**

dari dotplot data prof yang di tunjukkan dapat dilihat bahwa gaji cenderung tidak berubah meski tahun pengalaman bertambah, dan di rentang 45-50 tahun gaji cenderung kecil dibawah $ 100,000, meski rata-rata pesebarna data masih ada di rentang $ 100,000-$150,000

### 1. Lakukan barplot yang menampilkan gaji rata-rata Professor berdasarkan jabatan akademik

warna2 = c('#FAD8D6', '#CD5334', '#17BEBB')  
df3\_gaji <- df3 %>%  
 group\_by(rank) %>% # Mengakses Kolom rank  
 dplyr::summarize(mean\_salary = mean(salary)) # Menghitung rata-rata gaji untuk setiap rank(jabatan akademik)  
  
  
ggplot(df3\_gaji, aes(x = factor(rank), y = mean\_salary)) +  
 geom\_bar(stat = "identity", fill = warna2, color='#2E282A') +   
 geom\_text(aes(label = dollar(mean\_salary)), vjust = -0.25) + # Membuat pelabelan rata-rata gaji  
 scale\_y\_continuous(breaks = seq(0, 130000, 20000), label = dollar\_format()) +   
 theme\_minimal() +   
 labs(x = "Jabatan Akademik",   
 y = "Gaji Rata-Rata",  
 title = "Barplot: Gaji Rata-Rata Professor berdasarkan Jabatan Akademik")

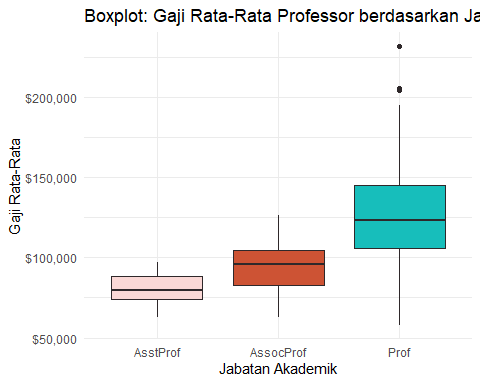


**Analisis**

dari barplot yang di hasilkan dapat dilihat bahwa semakin tinggi jabatan akademik semakin tinggi pula rata-rata gaji nya, dimana prof mempunyai rata-rata gaji sebesar $ 126,772, kemudian assocprof sebesar $ 93,876 dan asstprof sebesar $ 80,776

1. Lakukan boxplot yang menampilkan gaji rata-rata Professor berdasarkan jabatan akademik

ggplot(df3, aes(x = factor(rank), y = salary)) +  
 geom\_boxplot( fill = warna2, color='#2E282A') +   
 scale\_y\_continuous(labels = dollar\_format()) +   
 theme\_minimal() +   
 labs(x = "Jabatan Akademik",   
 y = "Gaji Rata-Rata",  
 title = "Boxplot: Gaji Rata-Rata Professor berdasarkan Jabatan Akademik")



**Analisis**

dari boxplot yang di hasilkan informasi yang di muat hampir sama dengan barpot sebellumnya namun disini dapat dilihat bahwa nilai kuartil bawah jabatan akademik assocprof menyentuh mediansampai kuartil atas dari jabatan akademik asstproff, yang menandakan gaji minimum asscproff hampir sama dengan rata-rata gaji asstproff, namun nilai minimum dari kedua jabatan akademik itu cenderung sama, yang artinya masih ada gaji jabatan assocprof yang sama dengan asstprof yang tergolong rendah

Pada jabatan akademik prof cenderung lebih besar dari yang lain, namun nilai minimum prof cenderung lebih rendah dari kedua jabatan akademik lainnya, yang artinya beberapa data dengan jabatan akademik proff memiliki gaji terrendah dari jabatan akademik lainnya. Pada jabatan akademik ini juga terdapat oulier yang berada di atas nilai maksimum, yang menandakan dari ketiga jabatna akademik ini, jabatan akademik prof memilliki beberapa data yang memiliki gaji tertinggi.