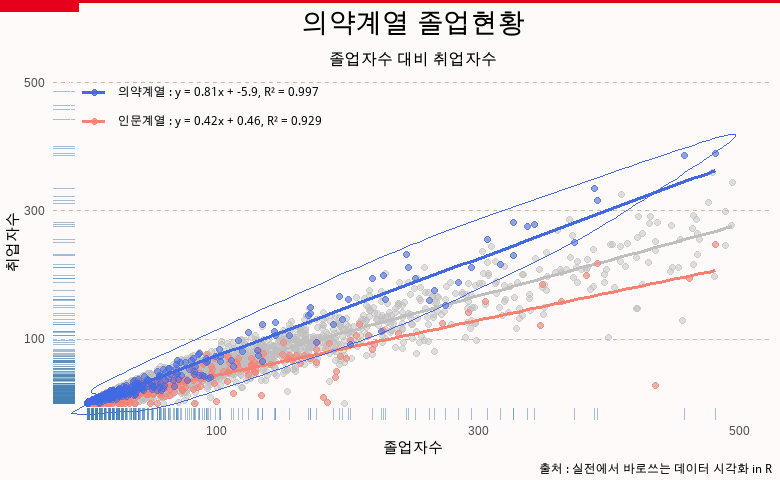
1. 실전 시각화

# 비교 그룹간 분포 비교와 회귀 방정식 표기

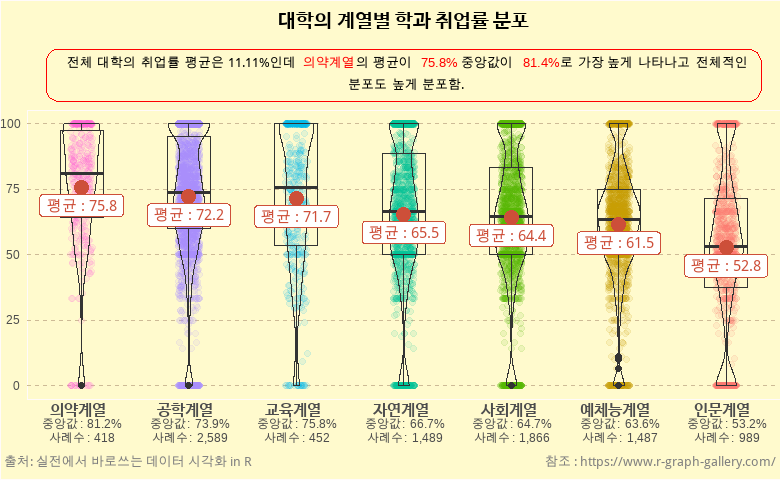
목표 : 비교의 대상이 되는 두 개의 그룹에 대한 산점도와 추세선을 각각 그리고 추세선의 선형회귀 방정식을 표현하여 각각의 그룹의 분포를 서로 비교하고 통계적 유의성을 전달  
활용 시각화 : 산점도, 회귀선, 클러스터 산점도, 러그 플롯, 선형회귀



# 전체 그룹간 다양한 통계 비교 시각화

목표 : 비교의 대상이 되는 두 개의 그룹에 대한 산점도와 추세선을 각각 그리고 추세선의 선형회귀 방정식을 표현하여 각각의 그룹의 분포를 서로 비교하고 통계적 유의성을 전달 활용 시각화 : 박스 플롯, 바이올린 플롯, 산점도

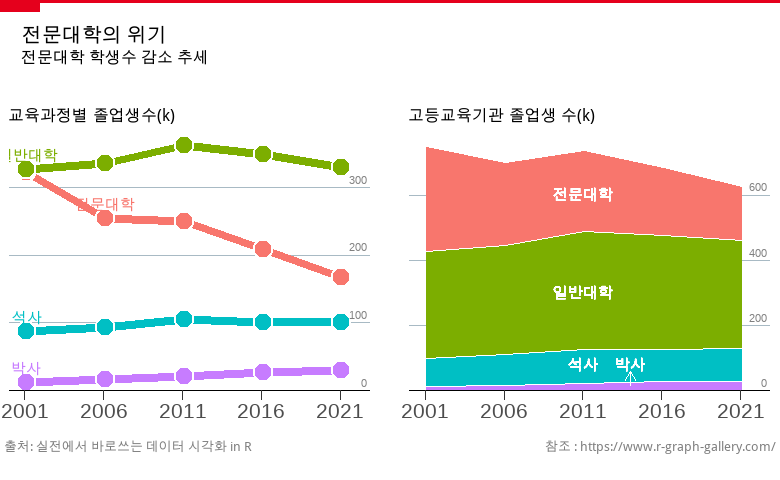
library(ggtext)  
font\_add('NanumBarunGothicBold', 'c:/windows/fonts/NanumBarunGothicBold.ttf')  
  
font\_add('NanumBarunGothic', 'c:/windows/fonts/NanumBarunGothic.ttf')  
  
df\_전체\_요약 <- df\_취업통계 |>  
 summarise(mean = mean(취업률\_계), median = median(취업률\_계), n = n())  
  
subtitle <- paste0('전체 사례수 : ', scales::comma(df\_전체\_요약$n, accuracy = 1))  
  
df\_계열\_요약 <- df\_취업통계 |>  
 group\_by(대계열) |>  
 summarise(median = median(취업률\_계), n = n())  
  
labels <- paste0('<span style= "font-family: NanumBarunGothicBold;font-size: 12pt" >', df\_계열\_요약$대계열, '</span> <br>', '중앙값 : ', scales::percent(df\_계열\_요약$median, scale = 1), '<br>', '사례수 : ', scales::comma(df\_계열\_요약$n, accuracy = 1))  
  
labels\_name <- pull(df\_계열\_요약, 대계열)  
  
title <- "<b><span style= 'font-family: NanumBarunGothicBold;font-size: 15pt'>대학의 계열별 학과 취업률 분포</b>"  
  
subtitle <- "<span style = 'font-size:10pt'>전체 대학의 취업률 평균은 11.11%인데 <span style = 'color:red;'>의약계열</span>의 평균이 <span style = 'color:red;'> 75.8%</span> 중앙값이 <span style = 'color:red;'> 81.4%</span>로 가장 높게 나타나고 전체적인 분포도 높게 분포함.</span>"  
  
df\_취업통계$대계열 = reorder(df\_취업통계$대계열, df\_취업통계$취업률\_계, mean)  
  
df\_취업통계 |>  
 ggplot(aes(x = 대계열, y = 취업률\_계)) +  
 geom\_jitter(aes(color = 대계열), width = 0.1, alpha = 0.1, show.legend = F) +  
 geom\_boxplot(fill = NA, width = 0.4) +  
 geom\_violin(fill = NA, width = 0.4) +  
 geom\_point(aes(x = 대계열, y = 취업률\_계), stat = 'summary', fun.y = 'mean', color = 'tomato3', size = 4) +  
 geom\_label(aes(x = 대계열, y = 취업률\_계, label = paste0('평균 : ', round(..y.., 1))), stat = 'summary', fun.y = 'mean', color = 'tomato3', size = 4, nudge\_y = -7) +   
 labs(title = title, x = NULL, y = NULL, subtitle = subtitle) +  
 scale\_x\_discrete(expand = expansion(add = c(0.5, 0.5)), labels = setNames(labels, labels\_name), limits = rev) +  
 theme(strip.text.x = element\_blank(),  
 strip.background = element\_rect(colour="white", fill="white"),  
 legend.position=c(.2,.9),   
 legend.background = element\_rect(fill = NA),   
 legend.key = element\_rect(fill = NA),   
 plot.title = element\_textbox\_simple(halign = 0.5),  
 plot.subtitle = element\_textbox\_simple(width = unit(0.95, "npc"), size = 13,  
 lineheight = 1,  
 padding = margin(5.5, 0.5, 5.5, 5.5),  
 margin = margin(10, 0, 5.5, 0),  
 halign = 0.5,   
 box.color = 'red',  
 r = grid::unit(8, "pt"),  
 linetype = 1),  
 axis.line = element\_blank(),   
 plot.background = element\_rect(fill = 'lemonchiffon'),   
 panel.background = element\_rect(fill = 'lemonchiffon'),   
 axis.text.x = element\_markdown(),   
 plot.margin = margin(0.025, 0, 0.075, 0, "npc"),  
 )  
  
  
grid.text(  
 '출처: 실전에서 바로쓰는 데이터 시각화 in R',   
 x = 0.005,   
 y = 0.03,   
 just = c("left", "bottom"),  
 gp = gpar(  
 col = "grey50",  
 fontsize = 10,  
 fontfamily = "Econ Sans Cnd"  
 )  
)  
  
# Add third caption  
grid.text(  
 "참조 : https://www.r-graph-gallery.com/",   
 x = 0.995,   
 y = 0.03,   
 just = c("right", "bottom"),  
 gp = gpar(  
 col = "grey50",  
 fontsize = 10,  
 fontfamily = "Econ Sans Cnd"  
 )  
)



# 종합

목표 : 비교의 대상이 되는 두 개의 그룹에 대한 산점도와 추세선을 각각 그리고 추세선의 선형회귀 방정식을 표현하여 각각의 그룹의 분포를 서로 비교하고 통계적 유의성을 전달 활용 시각화 : 박스 플롯, 바이올린 플롯, 산점도

library(ggtext)  
library(grid)  
library(tidyverse)  
  
theme\_set(  
 theme\_grey()  
)  
  
df\_total\_line <- df\_입학자\_long |> filter(지역 == '전체', lubridate::year(연도) %in% c(seq(from = 2001, to = 2021, by = 5)), 학교종류 %in% c('전문대학', '일반대학', '석사', '박사'))   
  
data\_labels <- bind\_rows(df\_total\_line |> filter(lubridate::year(연도) == 2001, 학교종류 %in% c('일반대학', '석사', '박사')), df\_total\_line |> filter(lubridate::year(연도) == 2006, 학교종류 %in% c('전문대학')))  
  
df\_total\_line |>  
 ggplot(aes(x = 연도, y = (입학생수)/1000)) +   
 ## geom\_line 레이어를 생성  
 geom\_line(aes(group = 학교종류, color = 학교종류), size = 2.4) +  
 geom\_point(aes(fill = 학교종류), size = 5,   
 shape = 21, # Type of point that allows us to have both color (border) and fill.  
 color = "white",   
 stroke = 1 # The width of the border, i.e. stroke.  
 ) +   
 scale\_x\_date(  
 expand = c(0, 0), # The horizontal axis does not extend to either side  
 limits = c(as.Date('2000-01-01'), as.Date('2022-12-01')),   
 breaks = seq(from = as.Date("2001-01-01"), to = as.Date("2021-01-01"),  
 by = "5 years"),   
 labels = lubridate::year(seq(from = as.Date("2001-01-01"), to = as.Date("2021-01-01"),  
 by = "5 years")) # Set custom break locations  
 # Set custom break locations  
# labels = c("2008", "12", "16", "20") # And custom labels on those breaks!  
 ) +   
 scale\_y\_continuous(  
 limits = c(0, 380),  
 expand = c(0, 0)  
 ) -> line\_ch7\_1  
  
line\_ch7\_1 +  
 theme(  
 # Set background color to white  
 panel.background = element\_rect(fill = "white"),  
 # Remove all grid lines  
 panel.grid = element\_blank(),  
 # But add grid lines for the vertical axis, customizing color and size   
 panel.grid.major.y = element\_line(color = "#A8BAC4", size = 0.3),  
 # Remove tick marks on the vertical axis by setting their length to 0  
 axis.ticks.length.y = unit(0, "mm"),   
 # But keep tick marks on horizontal axis  
 axis.ticks.length.x = unit(2, "mm"),  
 # Remove the title for both axes  
 axis.title = element\_blank(),  
 # Only the bottom line of the vertical axis is painted in black  
 axis.line.x.bottom = element\_line(color = "black"),  
 # Remove labels from the vertical axis  
 axis.text.y = element\_blank(),  
 # But customize labels for the horizontal axis  
 axis.text.x = element\_text(family = "Econ Sans Cnd", size = 16)  
 ) -> line\_ch7\_2  
  
  
  
line\_ch7\_2 +   
 geom\_text(data = data\_labels, aes(x = 연도, y = (입학생수/1000) + 20, label = 학교종류, color = 학교종류), show.legend = F) +   
 geom\_text(  
 data = data.frame(x = as.Date('2022-09-01'), y = seq(0, 300, by = 100)),  
 aes(x, y, label = y),  
 hjust = 1, # Align to the right  
 vjust = -0.5, # Align to the bottom  
 family = "Econ Sans Cnd",  
 size = 3,   
 color = 'grey50'  
 ) +   
 theme(legend.position = 'none') +  
 labs(  
 title = "\*\*교육과정별 졸업생수(k)\*\*",  
 ) +   
 theme(  
 # theme\_markdown() is provided by ggtext and means the title contains   
 # Markdown that should be parsed as such (the '\*\*' symbols)  
 plot.title = element\_markdown(  
 family = "Econ Sans Cnd",   
 size = 12  
 )  
 ) -> line\_ch7\_3  
  
  
ggplot(df\_total\_line) +  
 # color = "white" indicates the color of the lines between the areas  
 geom\_area(aes(x = 연도, y = (입학생수)/1000, group = 학교종류, fill = 학교종류), color = "white") +  
# scale\_fill\_manual(values = c('grey', 'brown', 'green', 'blue')) +  
 theme(legend.position = "None") + # no legend +  
 scale\_x\_date(  
 expand = c(0, 0), # The horizontal axis does not extend to either side  
 limits = c(as.Date('2000-01-01'), as.Date('2022-12-01')),   
 breaks = seq(from = as.Date("2001-01-01"), to = as.Date("2021-01-01"),  
 by = "5 years"),   
 labels = lubridate::year(seq(from = as.Date("2001-01-01"), to = as.Date("2021-01-01"),  
 by = "5 years")) # Set custom break locations  
 # Set custom break locations  
 # labels = c("2008", "12", "16", "20") # And custom labels on those breaks!  
 ) +   
 scale\_y\_continuous(  
 limits = c(0, 790),  
 expand = c(0, 0)  
 ) -> area\_ch7\_1  
  
area\_ch7\_1 +   
 theme(  
 # Set background color to white  
 panel.background = element\_rect(fill = "white"),  
 # Remove all grid lines  
 panel.grid = element\_blank(),  
 # But add grid lines for the vertical axis, customizing color and size   
 panel.grid.major.y = element\_line(color = "#A8BAC4", size = 0.3),  
 # Remove tick marks on the vertical axis by setting their length to 0  
 axis.ticks.length.y = unit(0, "mm"),   
 # But keep tick marks on horizontal axis  
 axis.ticks.length.x = unit(2, "mm"),  
 # Remove the title for both axes  
 axis.title = element\_blank(),  
 # Only the bottom line of the vertical axis is painted in black  
 axis.line.x.bottom = element\_line(color = "black"),  
 # Remove labels from the vertical axis  
 axis.text.y = element\_blank(),  
 # But customize labels for the horizontal axis  
 axis.text.x = element\_text(family = "Econ Sans Cnd", size = 16)  
 ) -> area\_ch7\_2  
  
area\_ch7\_2 +   
 geom\_text(aes(x = as.Date('2011-01-01'), y = 600), label = '전문대학', color = 'white') +   
 geom\_text(aes(x = as.Date('2011-01-01'), y = 300), label = '일반대학', color = 'white') +   
 geom\_text(aes(x = as.Date('2011-01-01'), y = 80), label = '석사', color = 'white') +   
 geom\_text(aes(x = as.Date('2014-01-01'), y = 80), label = '박사', color = 'white') +   
 geom\_segment(aes(x = as.Date('2014-01-01'), xend = as.Date('2014-01-01'), y = 12, yend = 60), color = 'white', arrow = arrow(angle = 30, length = unit(0.1, "inches"))) +   
 geom\_text(  
 data = data.frame(x = as.Date('2022-09-01'), y = seq(0, 800, by = 200)),  
 aes(x, y, label = y),  
 hjust = 1, # Align to the right  
 vjust = -0.5, # Align to the bottom  
 family = "Econ Sans Cnd",  
 size = 3,   
 color = 'grey50'  
 ) +   
 theme(legend.position = 'none') +   
 labs(  
 title = "고등교육기관 졸업생 수(k)",  
 ) +   
 theme(  
 plot.title = element\_markdown(  
 family = "Econ Sans Cnd",   
 size = 12  
 )  
 ) -> area\_ch7\_3  
  
  
plt1 <- line\_ch7\_3 + theme(plot.margin = margin(0, 0.05, 0, 0, "npc"))  
plt2 <- area\_ch7\_3 + theme(plot.margin = margin(0, 0, 0.05, 0, "npc"))  
plt <- plt1 | plt2  
  
title\_theme <- theme(  
 plot.title = element\_text(  
 hjust = 0.02,  
 size = 15,  
 margin = margin(0.8, 0, 0.3, 0, "npc")  
 ),  
 plot.subtitle = element\_text(  
 hjust = 0.02,  
 size = 12,  
 margin = margin(0.4, 0, 0.5, 0, "npc")  
 )  
)  
  
plt <- plt + plot\_annotation(  
 title = "전문대학의 위기",  
 subtitle = "전문대학 학생수 감소 추세",  
 theme = title\_theme  
) +  
 theme(  
 plot.margin = margin(0.075, 0, 0.1, 0, "npc"),  
 )  
   
plt  
  
# Add line on top of the chart  
grid.lines(  
 x = c(0, 1),  
 y = 1,  
 gp = gpar(col = "#e5001c", lwd = 4)  
)  
  
# Add rectangle on top-left  
# lwd = 0 means the rectangle does not have an outer line  
# 'just' gives the horizontal and vertical justification  
grid.rect(  
 x = 0,  
 y = 1,  
 width = 0.05,  
 height = 0.025,  
 just = c("left", "top"),  
 gp = gpar(fill = "#e5001c", col = "#e5001c", lwd = 0)  
)  
# Add first caption  
grid.text(  
 '출처: 실전에서 바로쓰는 데이터 시각화 in R',   
 x = 0.005,   
 y = 0.06,   
 just = c("left", "bottom"),  
 gp = gpar(  
 col = "grey50",  
 fontsize = 10,  
 fontfamily = "Econ Sans Cnd"  
 )  
)  
  
# Add third caption  
grid.text(  
 "참조 : https://www.r-graph-gallery.com/",   
 x = 0.995,   
 y = 0.06,   
 just = c("right", "bottom"),  
 gp = gpar(  
 col = "grey50",  
 fontsize = 10,  
 fontfamily = "Econ Sans Cnd"  
 )  
)



## Lists

Amet nunc eros curabitur tellus massa, eros maximus porttitor sociosqu, pellentesque.

* Erat mauris egestas finibus tincidunt sed in rhoncus a tellus etiam.
  + A adipiscing per ultricies justo tellus lorem.
    - Imperdiet ut dui primis, sed gravida, at sed nulla.
    - Sem posuere lacus consequat inceptos dapibus duis malesuada finibus.
  + Urna sed dui, ornare, eu turpis mus pellentesque amet amet bibendum.
* Himenaeos tincidunt, auctor dapibus scelerisque, montes nunc faucibus sodales malesuada ridiculus sed cubilia ligula.

1. Erat mauris egestas finibus tincidunt sed in rhoncus a tellus etiam.
   1. A adipiscing per ultricies justo tellus lorem.
      1. Imperdiet ut dui primis, sed gravida, at sed nulla.
      2. Sem posuere lacus consequat inceptos dapibus duis malesuada finibus.
   2. Urna sed dui, ornare, eu turpis mus pellentesque amet amet bibendum.
2. Himenaeos tincidunt, auctor dapibus scelerisque, montes nunc faucibus sodales malesuada ridiculus sed cubilia ligula.

Lorem dolor interdum orci eros pellentesque semper tristique, sodales, et sed ut! Porta mattis natoque et. Ac facilisi ipsum viverra elementum vestibulum ligula amet enim magnis luctus ullamcorper. Rhoncus rhoncus elit in at nisl. Tincidunt habitant sit.

* Aptent conubia quam montes id sagittis.
  + Mattis nisi nascetur, aliquam duis ex, tristique.
    - Imperdiet ut dui primis, sed gravida, at sed nulla.
    - Donec ligula nulla ac. Nisl ac at accumsan sagittis eros felis lobortis amet nec phasellus urna bibendum sapien.
  + Eu dui ac id, dictum proin consectetur convallis.
* Facilisi eu lectus mauris lorem. Et sed sapien pellentesque sed etiam vehicula.
* In porttitor id lorem eu efficitur, nisl dis!

## Reference

see figure and table !