

# Multivariate Graph

Bibek Sapkota

2024-08-04

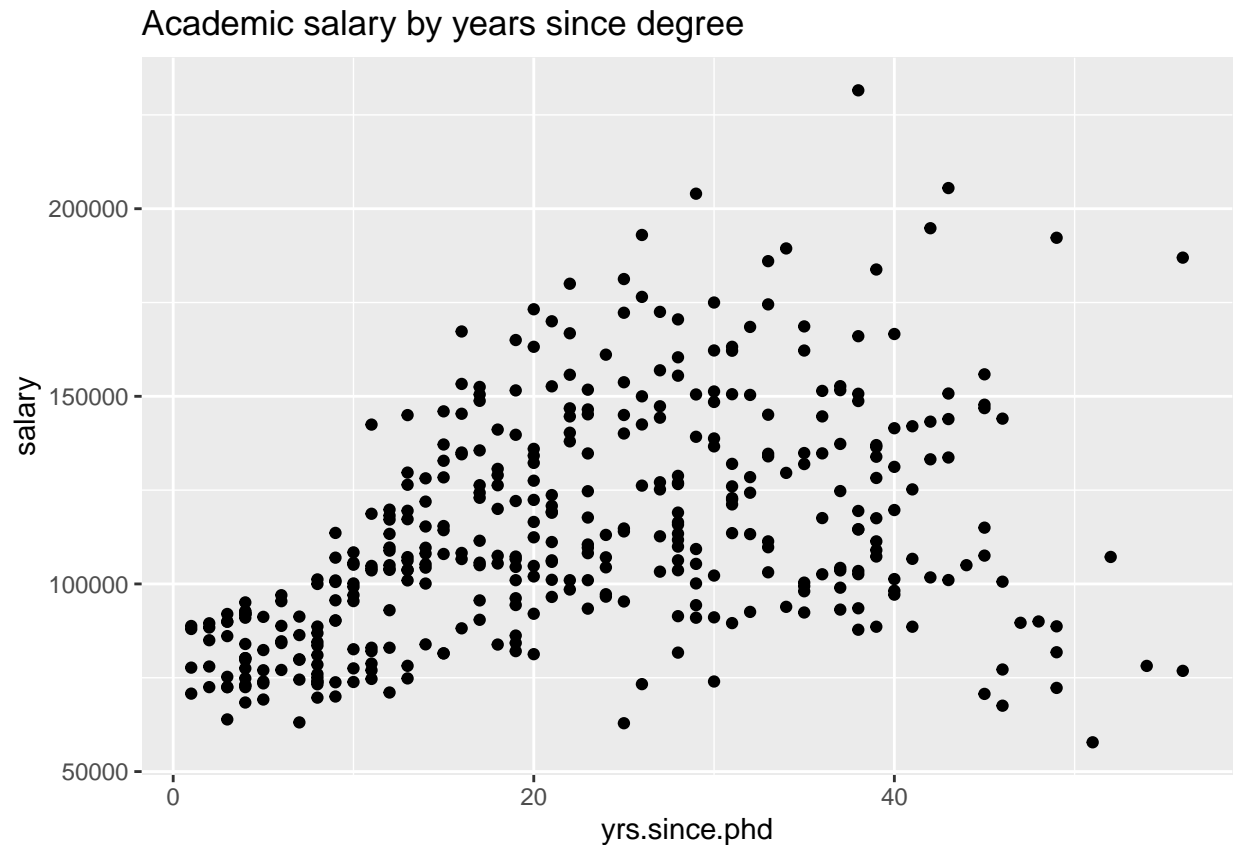
## Importing library and dataset

```
library(ggplot2)
data(Salaries, package="carData")
```

## Grouping

Task:Ploting scatter plot of salary vs year since phd.

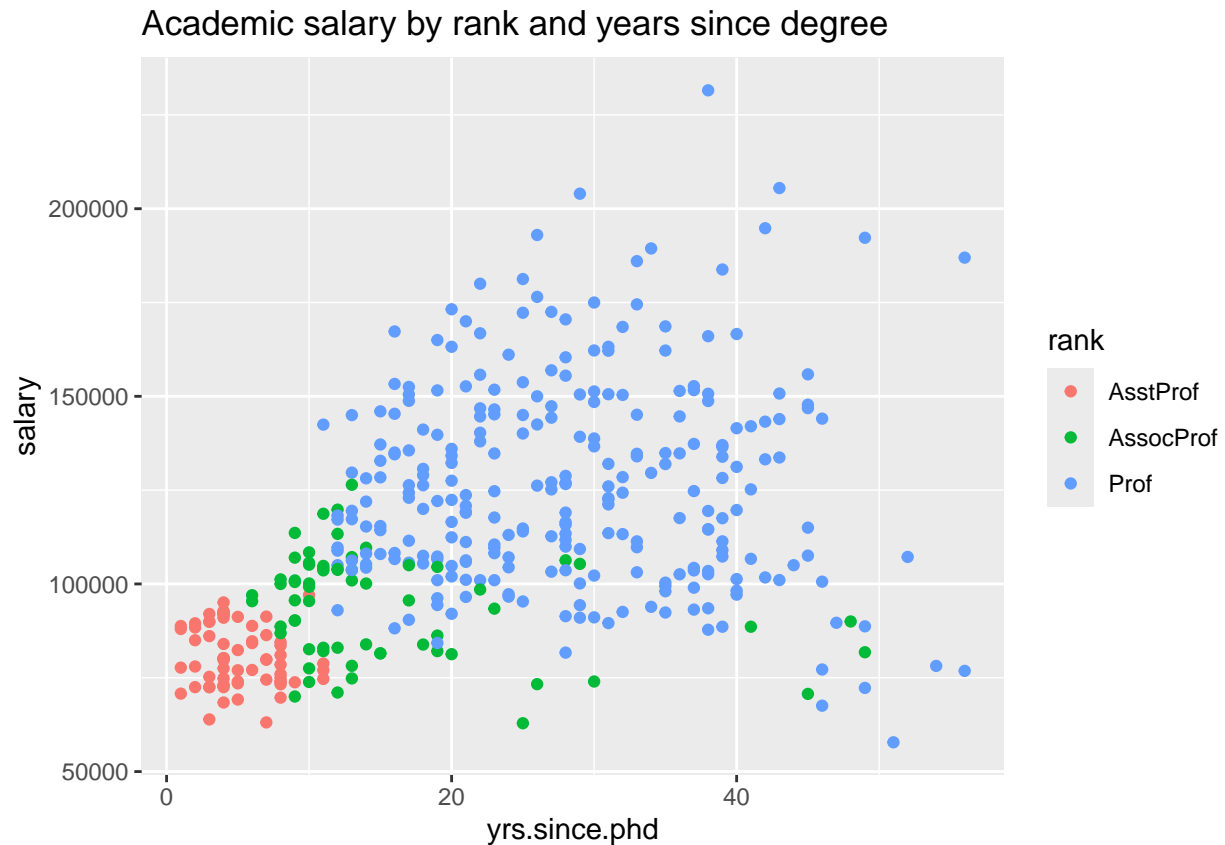
```
ggplot(Salaries,
       aes(x = yrs.since.phd, y = salary)) +
  geom_point() +
  labs(title = "Academic salary by years since degree")
```



Interpretation: This scatter plot depicts the relationship between academic salary and years since obtaining a PhD. It shows a general trend where salaries tend to increase with more years since obtaining a PhD, although there is significant variability, with some high salaries observed even at lower years and some lower salaries at higher years.

Task: Plotting experience vs. salary (color represents rank- i.e AsstProf, AssocProf, Prof)

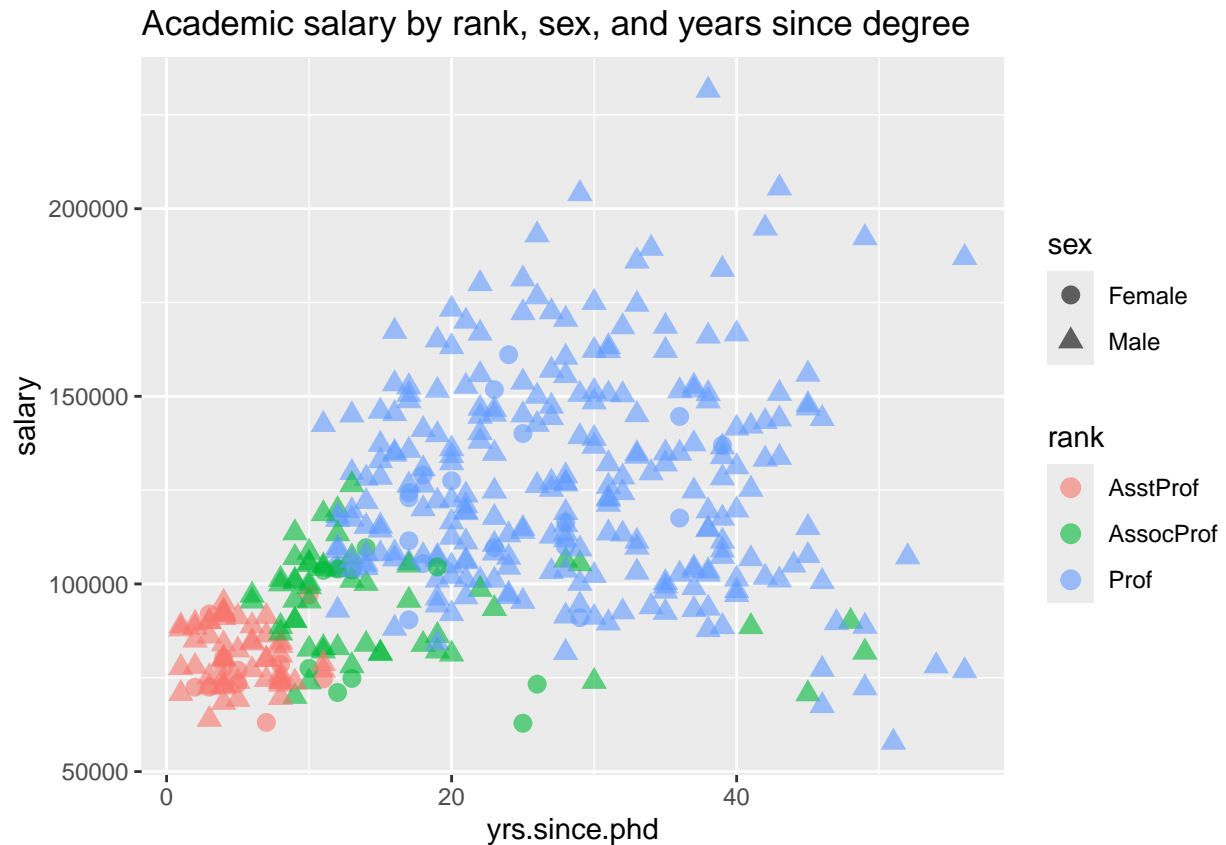
```
ggplot(Salaries, aes(x = yrs.since.phd,
                     y = salary,
                     color=rank)) +
  geom_point() +
  labs(title = "Academic salary by rank and years since degree")
```



Interpretation: This scatter plot shows academic salaries by rank and years since obtaining a PhD, indicating that salaries increase with more years since the degree. Assistant Professors have the lowest salaries, Associate Professors have mid-range salaries, and Professors have the highest salaries

Task: plotting scatter plot of experience vs. salary (color represents rank, shape represents sex i.e. male and female)

```
ggplot(Salaries, aes(x = yrs.since.phd,
                    y = salary,
                    color = rank,
                    shape = sex)) +
  geom_point(size = 3, alpha = .6) +
  labs(title = "Academic salary by rank, sex, and years since degree")
```

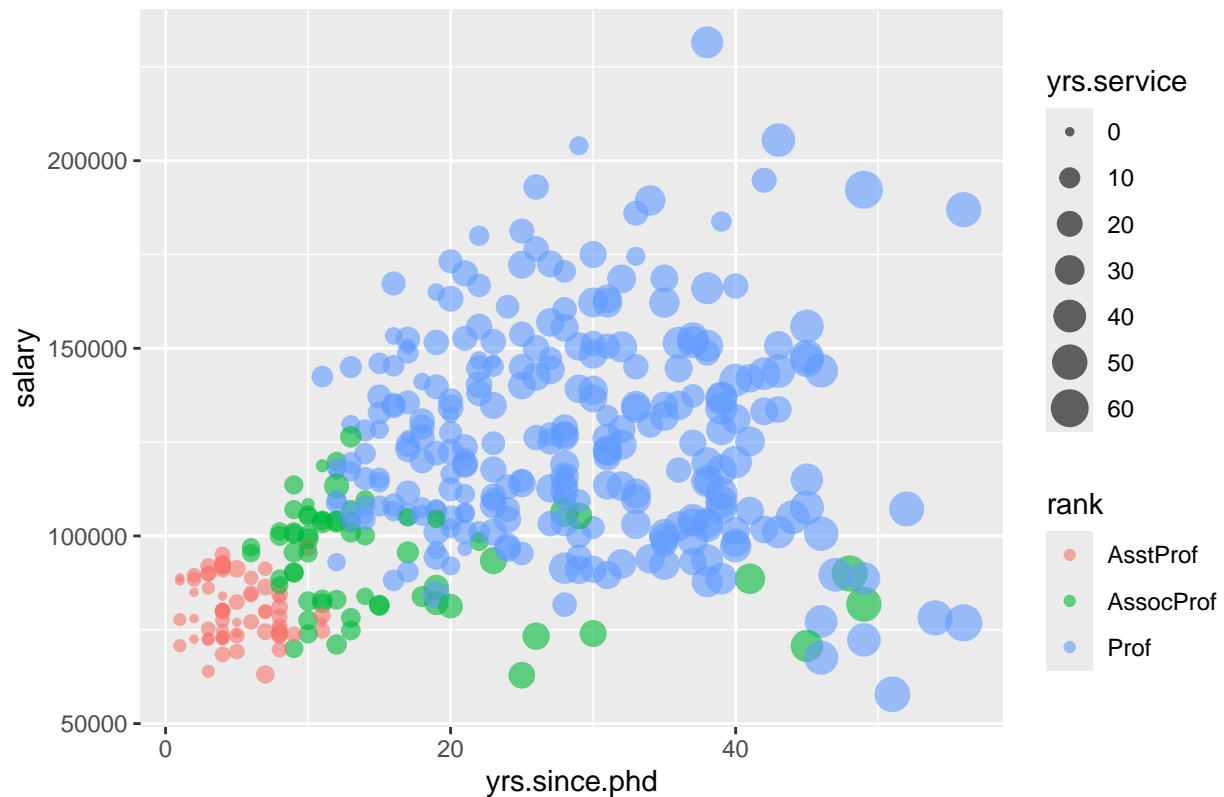


Interpretation: This scatter plot displays academic salaries by rank, sex, and years since obtaining a PhD. It shows that salaries increase with more years since the PhD, with Professors earning the highest salaries. Males (triangles) and females (circles) are distributed across the ranks, but males appear more frequently in higher salary ranges.

Task: plotting experience vs. salary (color represents rank and size represents service)

```
ggplot(Salaries, aes(x = yrs.since.phd,
                    y = salary,
                    color = rank,
                    size = yrs.service)) +
  geom_point(alpha = .6) +
  labs(title = paste0("Academic salary by rank, years of service, ",
                     "and years since degree"))
```

Academic salary by rank, years of service, and years since degree

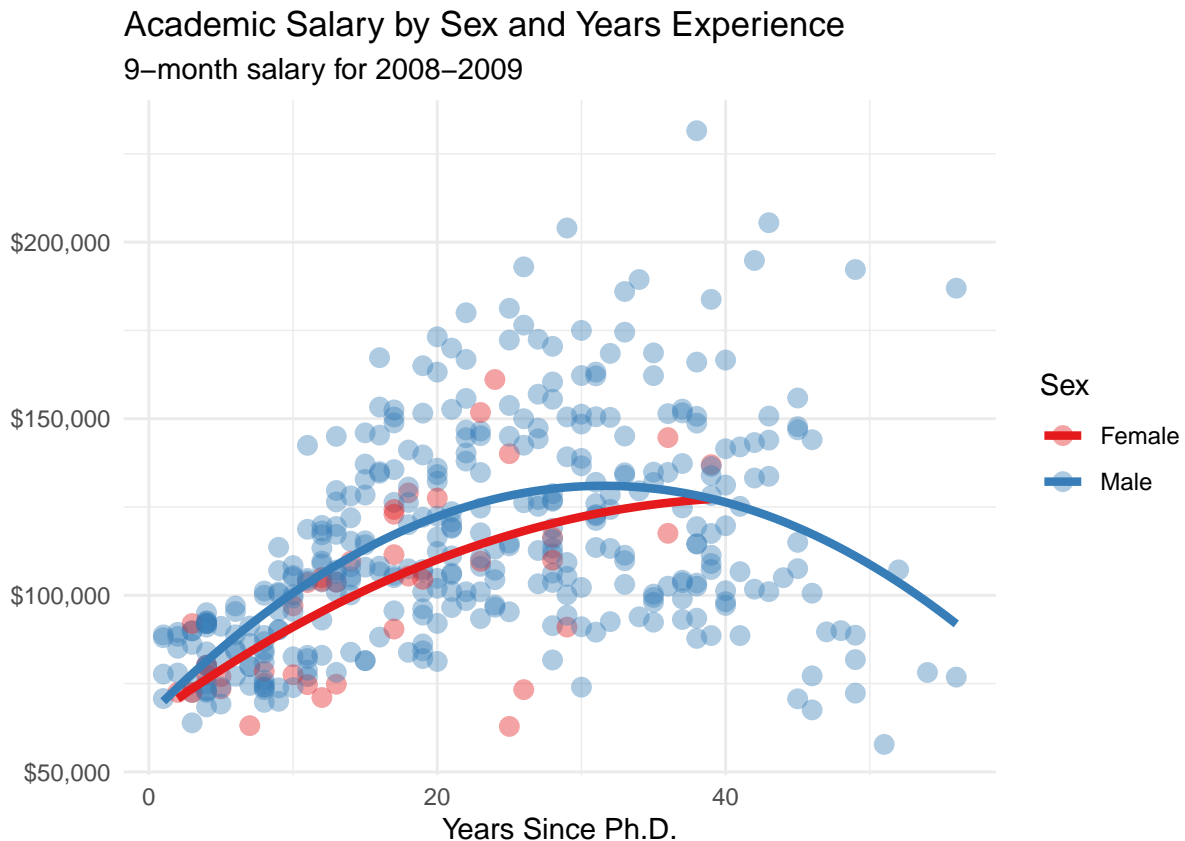


Interpretation: This scatter plot shows academic salaries as a function of years since obtaining a PhD, with different colors representing academic ranks (Assistant Professor, Associate Professor, Professor). The size of the points indicates years of service. Generally, salaries increase with years since PhD, and professors (in blue) tend to have higher salaries and more years of service compared to assistant (red) and associate professors (green).

Task: Plotting experience vs. salary with fit lines (color represents sex)

```
ggplot(Salaries,
  aes(x = yrs.since.phd,
    y = salary,
    color = sex)) +
  geom_point(alpha = .4,
    size=3) +
  geom_smooth(se=FALSE,
    method="lm",
    formula=y~poly(x,2),
    size = 1.5) +
  labs(x = "Years Since Ph.D.",
    title = "Academic Salary by Sex and Years Experience",
    subtitle = "9-month salary for 2008-2009",
    y = "",
    color = "Sex") +
  scale_y_continuous(label = scales::dollar) +
  scale_color_brewer(palette="Set1") +
  theme_minimal()
```

```
## 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.
```



Interpretation: This scatter plot depicts the relationship between academic salary and years since obtaining a PhD, with data points differentiated by sex (female in red, male in blue). The trend lines indicate that both male and female salaries increase with experience, peaking around 20-25 years post-PhD, before declining slightly. Generally, male academics appear to have higher salaries compared to their female counterparts throughout the observed range of years since PhD.

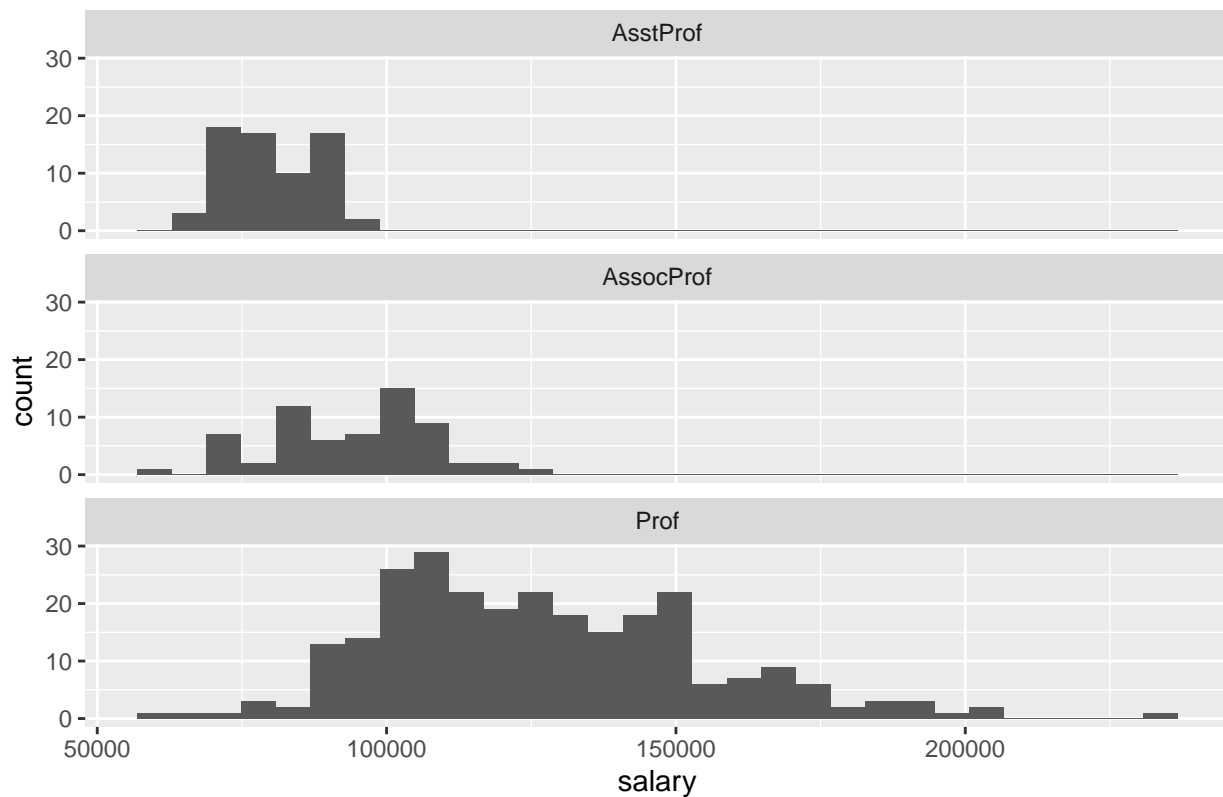
## Faceting

Task: Plotting salary histograms by rank

```
ggplot(Salaries, aes(x = salary)) +
  geom_histogram() +
  facet_wrap(~rank, ncol = 1) +
  labs(title = "Salary histograms by rank")
```

```
## 'stat_bin()' using 'bins = 30'. Pick better value with 'binwidth'.
```

## Salary histograms by rank



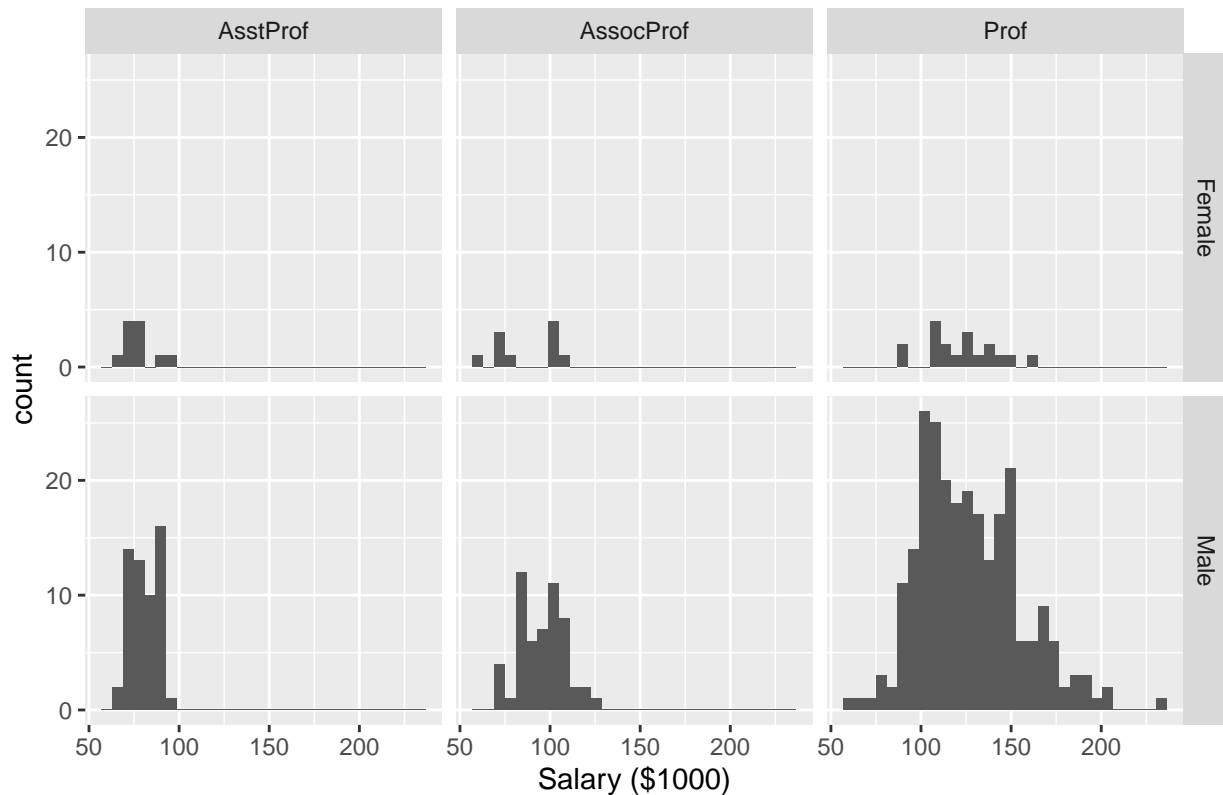
Interpretation: These histograms illustrate salary distributions across academic ranks: Assistant Professors (AsstProf), Associate Professors (AssocProf), and Professors (Prof). Assistant Professors mostly earn between \$70,000 and \$100,000. Associate Professors have salaries mainly between \$80,000 and \$120,000. Professors show the highest and broadest range, with many earning between \$100,000 and \$150,000, and some exceeding \$200,000. Higher ranks correlate with higher salaries.

Task: Plotting salary histograms by rank and sex

```
ggplot(Salaries, aes(x = salary/1000)) +
  geom_histogram() +
  facet_grid(sex ~ rank) +
  labs(title = "Salary histograms by sex and rank",
        x = "Salary ($1000)")
```

## 'stat\_bin()' using 'bins = 30'. Pick better value with 'binwidth'.

Salary histograms by sex and rank



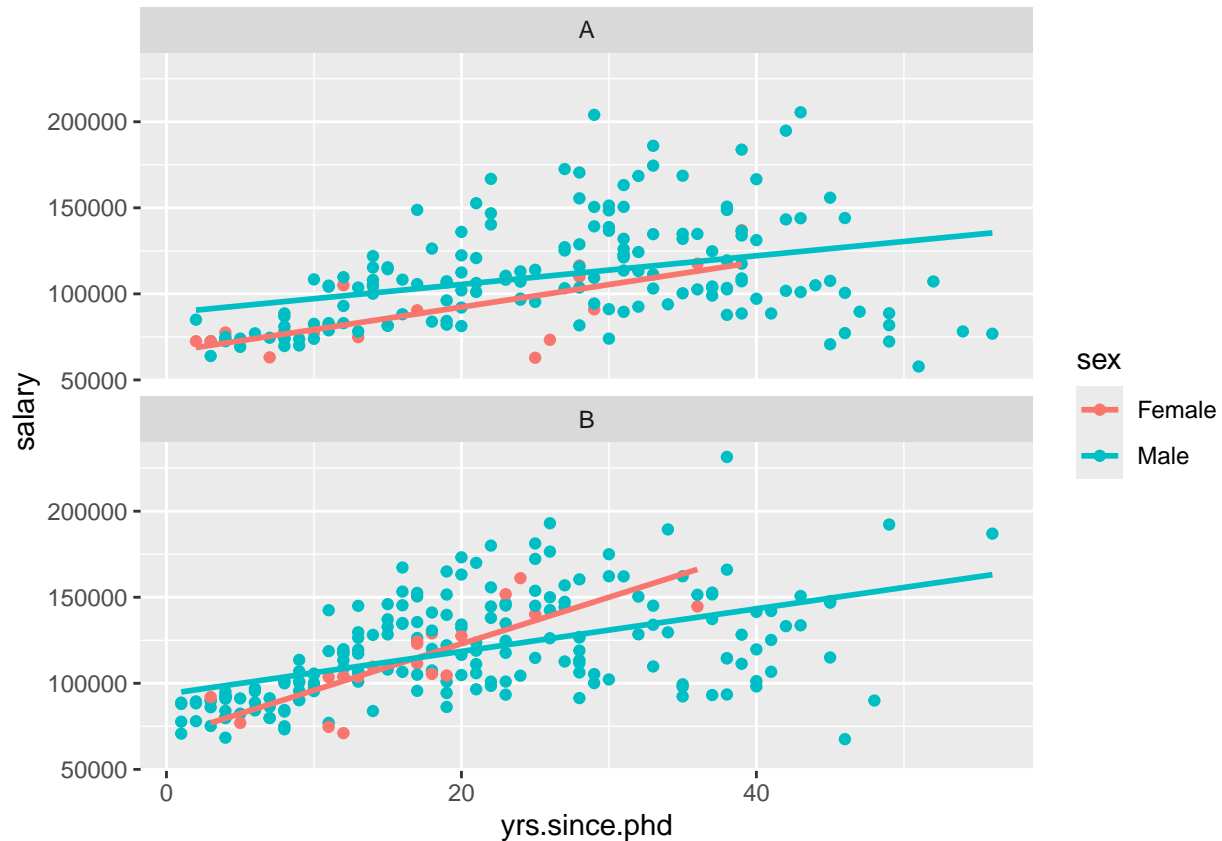
Interpretation: This histograms show salary distributions across academic ranks (Assistant Professor, Associate Professor, and Professor) separated by sex. Male salaries generally have a broader and higher distribution compared to female salaries across all ranks, indicating that men tend to earn more than women within the same rank. This highlights potential salary disparities between male and female faculty members in academia.

Task: Plotting scatter plot of salary by years of experience by sex and discipline with trend line.

```
ggplot(Salaries,
       aes(x=yrs.since.phd, y = salary, color=sex)) +
  geom_point() +
  geom_smooth(method="lm",
             se=FALSE) +
  facet_wrap(~discipline,
            ncol = 1)
```

```
## 'geom_smooth()' using formula = 'y ~ x'
```





Interpretation: This scatter plots depict the relationship between years of experience (years since Ph.D.) and salary, separated by sex and discipline (A and B). The trend lines indicate that, generally, salaries increase with years of experience for both males and females. However, male salaries tend to be higher than female salaries across both disciplines, as shown by the higher positioning of the male trend lines compared to the female trend lines.

Task:plotting salary by years of experience by sex and discipline

```
ggplot(Salaries, aes(x=yrs.since.phd,
                    y = salary,
                    color=sex)) +
  geom_point(size = 2,
            alpha=.5) +
  geom_smooth(method="lm",
            se=FALSE,
            size = 1.5) +
  facet_wrap(~factor(discipline,
                    labels = c("Theoretical", "Applied")),
            ncol = 1) +
  scale_y_continuous(labels = scales::dollar) +
  theme_minimal() +
  scale_color_brewer(palette="Set1") +
  labs(title = paste0("Relationship of salary and years ",
                    "since degree by sex and discipline"),
        subtitle = "9-month salary for 2008-2009",
        color = "Gender",
        x = "Years since Ph.D.",
```

```
y = "Academic Salary")
```

```
## 'geom_smooth()' using formula = 'y ~ x'
```

### Relationship of salary and years since degree by sex and discipline 9-month salary for 2008–2009



Interpretation: This scatter plots show the relationship between academic salary and years since Ph.D., separated by sex and discipline (Theoretical and Applied). Salaries generally increase with experience for both genders, but male salaries are consistently higher than female salaries in both disciplines, indicating a persistent gender pay gap.

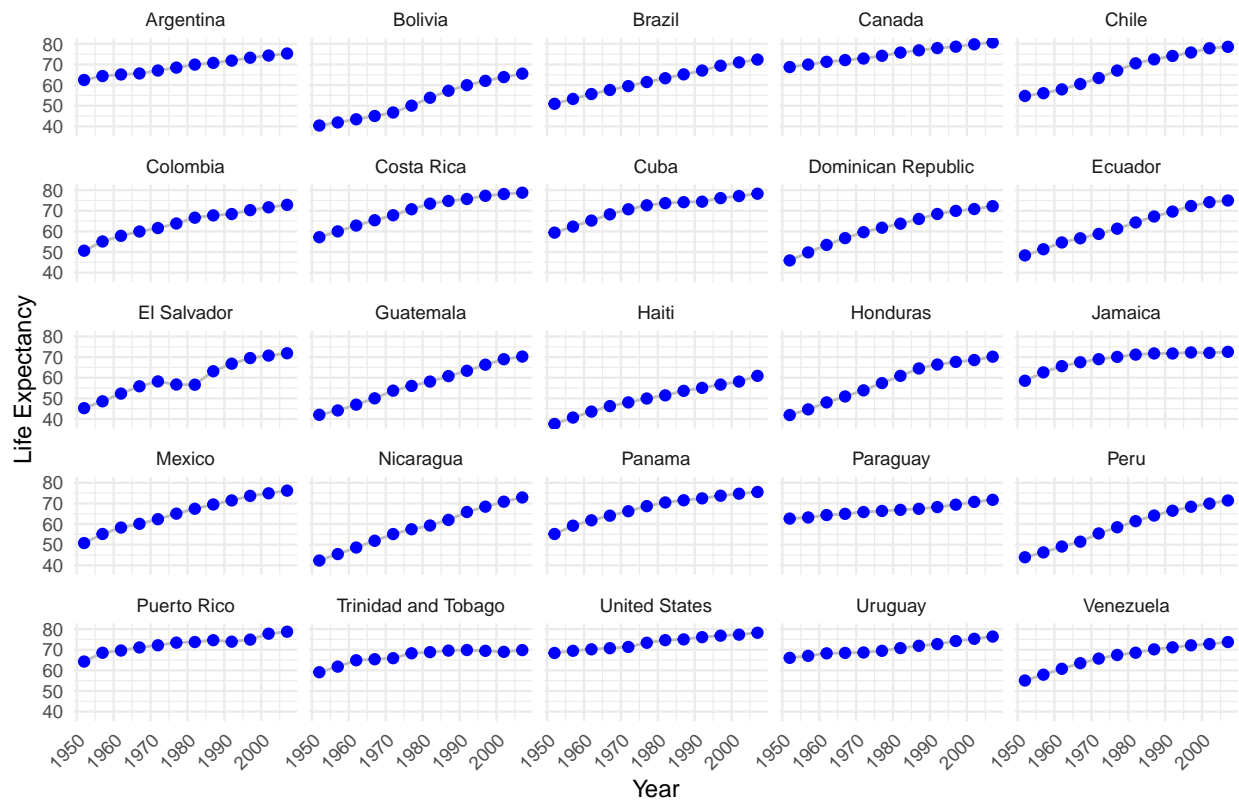
Task: Plotting life expectancy by year separately for each country in the Americas

```
data(gapminder, package = "gapminder")

plotdata <- dplyr::filter(gapminder,
                           continent == "Americas")

ggplot(plotdata, aes(x=year, y = lifeExp)) +
  geom_line(color="grey") +
  geom_point(color="blue") +
  facet_wrap(~country) +
  theme_minimal(base_size = 9) +
  theme(axis.text.x = element_text(angle = 45,
                                     hjust = 1)) +
  labs(title = "Changes in Life Expectancy",
       x = "Year",
       y = "Life Expectancy")
```

## Changes in Life Expectancy



Interpretation: This graph shows the changes in life expectancy from 1950 to 2000 for various countries in the Americas. Each subplot represents a country, tracking life expectancy over time. Generally, life expectancy increased in all countries. Canada, the United States, and Uruguay had consistently higher life expectancies, while Haiti, Bolivia, and Honduras, starting from lower rates, showed notable improvements. The upward trend indicates better healthcare, nutrition, and living conditions across the region, though the rates of increase and initial levels vary, reflecting economic and social differences.