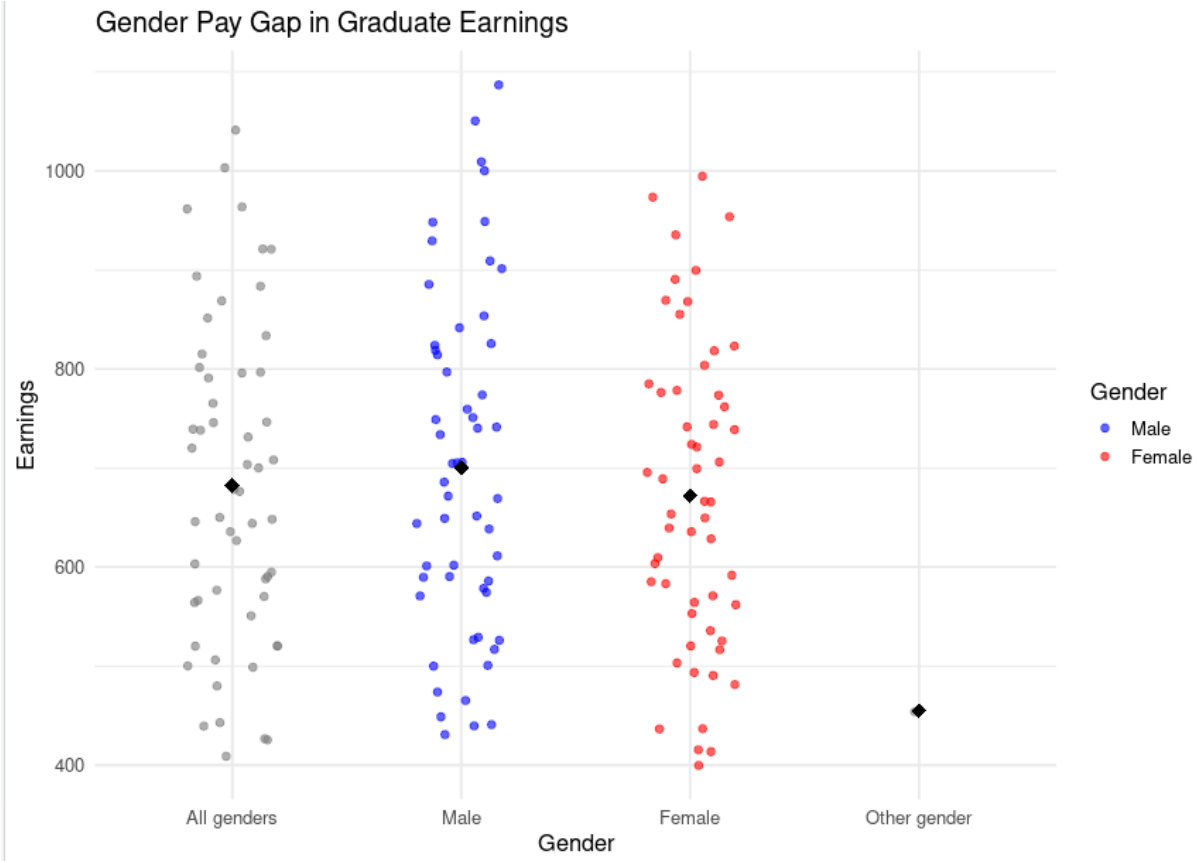
```
library(readr)
library(tidyverse)
library(here)
library(here)
glimpse(gradMS)
dim(gradMS)
gradMS[,1]
gradMS[,2]
gradMS[,3]
gradMS[,4]
gradMS[,5]
gradMS[,6]
gradMS[,7]

gender_pay_gap <- gradMS_gender %>%
pivot_wider(names_from = Gender, values_from = Earnings) %>%
mutate(Gender_Pay_Gap = (Male - Female) / Male * 100)

gradMS %>% filter(Gender == "Other gender")
gradMS %>% filter(Gender != "Other gender")
gradMS_wider <- gradMS|>
pivot_wider(names_from = Statistic, values_from = .)
gradMS_wider_lighter <- gradMS_wider_lighter %>%
gradMS_wider_lighter <- gradMS_wider_lighter %>%
rename(P50 = `P50 Earnings of Graduates`,P25 = `P25 Earnings of Graduates`,P75 = `P75
Earnings of Graduates`)

#Gender pay gap in graduates, comparing male v female
ggplot(gradMS_wider_lighter, aes(x = Gender, y = `P50 Earnings of Graduates`, color =
Gender)) +
geom_jitter(width = 0.2, alpha = 0.6, size = 2) +
stat_summary(fun = mean, geom = "point", shape = 18, size = 4, colour = "black") +
labs(title = "Gender Pay Gap in Graduate Earnings",
x = "Gender",
y = "Earnings") +
scale_color_manual(values = c("Male" = "blue", "Female" = "red")) +
theme_minimal(base_size = 14)

ggplot(gradMS_wider_lighter, aes(x = Gender, y = `P50 Earnings of Graduates`, colour =
Gender)) +
geom_jitter(width = 0.2, alpha = 0.8, size = 2) +
stat_summary(fun = mean, geom = "point", shape = 18, size = 4, color = "black") +
labs(title = "Gender Pay Comparison",x = "Gender",y = "Median Earnings") +
scale_color_manual(values = c("Male" = "blue", "Female" = "red")) +
```

theme_minimal(base_size = 10)

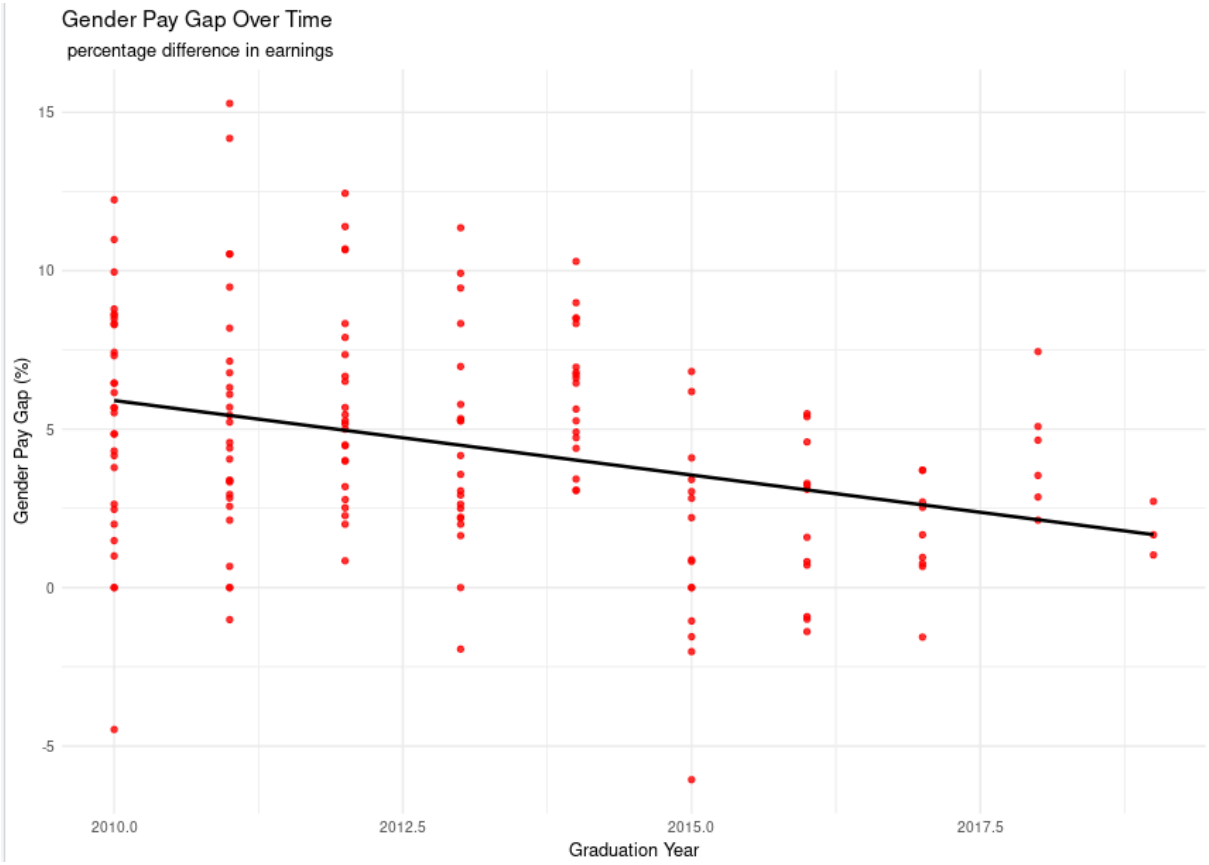


#Median salary over time, comparing male v female salarys

```
ggplot(gradMS_wider_lighter, aes(x = Graduation.Year, y = `P50 Earnings of Graduates`,
colour = Gender)) +
geom_point(alpha = 0.8) +
geom_smooth(method = "lm", se = FALSE) +
labs(title = "Median Monthly Salary by Gender Over Time",
y = "Earnings") +
theme_minimal()
```

Still a lot of cleaning up on this graph to be done. Also not sure why ive labelled it as monthly my mistake.

```
#Pay gap over time as a percent
ggplot(gender_pay_gap, aes(x = Graduation.Year, y = Gender_Pay_Gap)) +
geom_point(color = "red", alpha = 0.8) +
geom_smooth(method = "lm", se = FALSE, colour = "black") +
labs(title = "Gender Pay Gap Over Time",
subtitle = " percentage difference in earnings" ,
x = "Graduation Year",
y = "Gender Pay Gap (%)") +
theme_minimal()
```



This one is a bit unclear but essentially I gathered that if the trend line goes below 0 women are making more than men, this makes sense as from messing around with the data it seems men have always generally earned more than women however the difference has been decreasing.

```
gender_pay_gap <- gradMS_gender %>%
pivot_wider(names_from = Gender, values_from = Earnings) %>%
mutate(Gender_Pay_Gap = (Male - Female) / Male * 100)
```

^thats how I cleaned the data to try see if i could get it as a percent