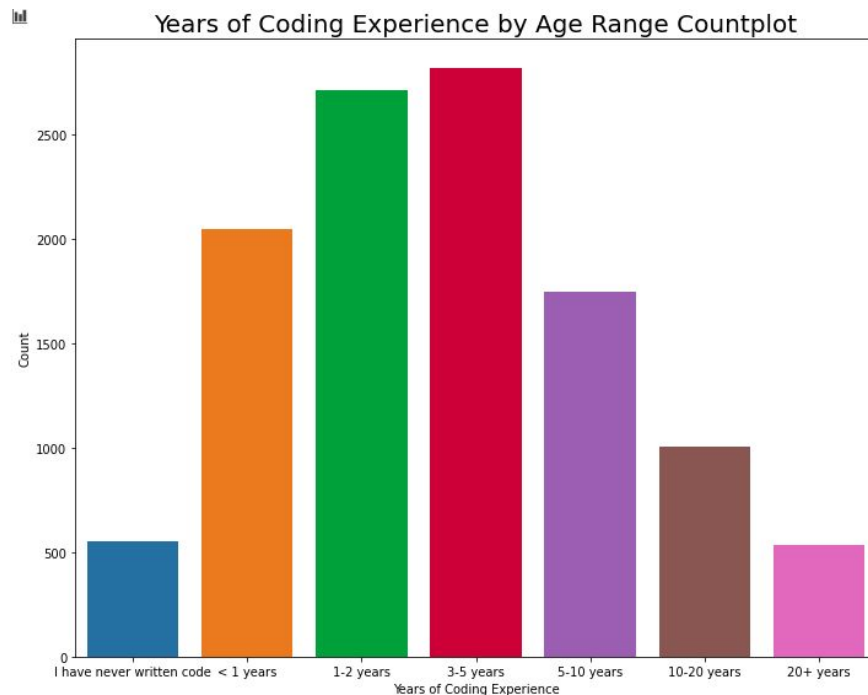# Impact of Gender & Education on Income in DS & ML

"tell a data story about a subset of the data science community represented in this survey, through a combination of both narrative text and data exploration"

Samuel Atkins - October 2020
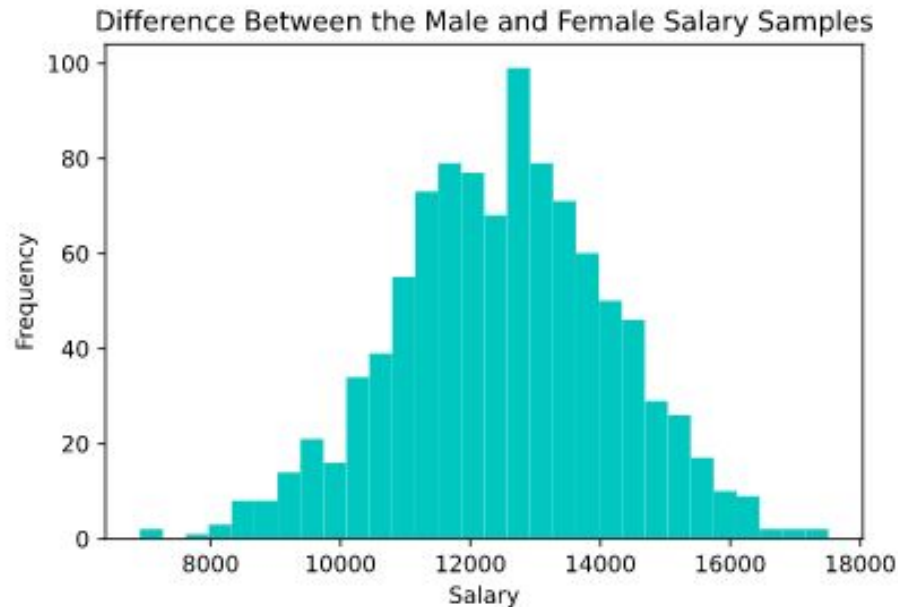
Exploratory Data Analysis

- Data Visualization
- Salary vs. Gender & Salary vs. Education Box Plots
- Descriptive Statistics

	sex	country	age	educ	coding_exp_yrs	salary
0	Male	France	22-24	Master's degree	1-2 years	40000
1	Male	India	40-44	Professional degree	I have never written code	7500
2	Male	Australia	40-44	Master's degree	1-2 years	300000
3	Male	India	22-24	Bachelor's degree	< 1 years	5000
4	Male	France	50-54	Master's degree	20+ years	70000

Years of Coding Experience by Age Range Countplot

Comparing Male and Female Salaries in DS & ML

Two-Sample t-Test & Bootstrapping

Difference Between the Male and Female Salary Samples

```
tc, pc = stats.ttest_ind(males_df["salary"], females_df["salary"])
print ("t-test: t = %g  p = %g" % (tc, pc))
```
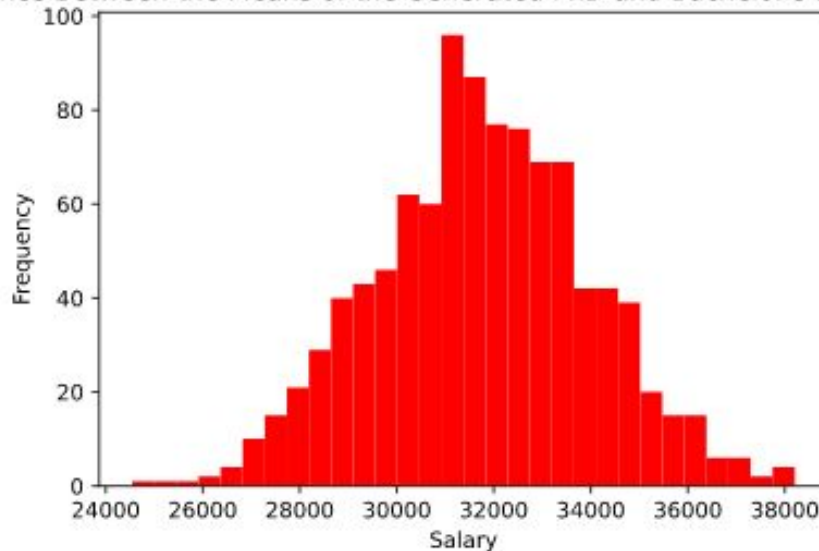t-test: t = 6.44284 p = 1.22048e-10

```
tc, pc = stats.ttest_ind(male_means, female_means)
print ("t-test: t = %g  p = %g" % (tc, pc))
```
t-test: t = 241.718 p = 0

Impact of Education on Income

ANOVA, Bootstrapping, & Estimating the Difference Between the Means

Difference Between the Means of the Generated PhD and Bachelor's Salary Samples

```
f_stat, p_val = stats.f_oneway(phd_df.salary, masters_df.salary, ba_df.salary)
print ("F-stat: F = %g   p = %g" % (f_stat, p_val))

F-stat: F = 119.618  p = 4.36862e-52
```

```
f_stat, p_val = stats.f_oneway(phd_means, master_means, ba_means)
print ("F-stat: F = %g   p = %g" % (f_stat, p_val))

F-stat: F = 130362  p = 0
```

Conclusion

Difference Between Salaries of Males and Females in DS & ML:

- *Males earn approximately $12,514.74 more on average*

Average Salary Increases for Education Tiers Surpassing a Bachelor's Degree:

- *Bachelor's Degree to Master's degree: $14,118.39 approximate pay increase*
- *Bachelor's Degree to PhD: $31,765.27 approximate pay increase*