Project 2: Data Visualization

For this project, I compared wages across genders to see if there was a wage gap between males and females. Additionally, I looked to see if other factors besides gender, such as age or education levels, impacted the wage gap between males and females.

In [3]:
```
import numpy as np
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt
```

In [4]:
```
md = pd.read_csv("datasets/wageGenderEduAge.csv")
```

Wage Distribution by Gender

In [5]:
```
#Setting the style of my plots
sns.set_style('whitegrid') #adding gridlines to the background of my plots
sns.set_context('notebook', font_scale=1.5) #changing the font size of my plots
```

In [6]:
```
sns.displot(md, x="wage", hue="gender", kind="hist", kde=True,
            palette='Set1', height=5, aspect=2, element='step')
plt.xlabel("Wage")
plt.title("Wage Distributions by Gender")
```

Out[6]:
```
Text(0.5, 1.0, 'Wage Distributions by Gender')
```

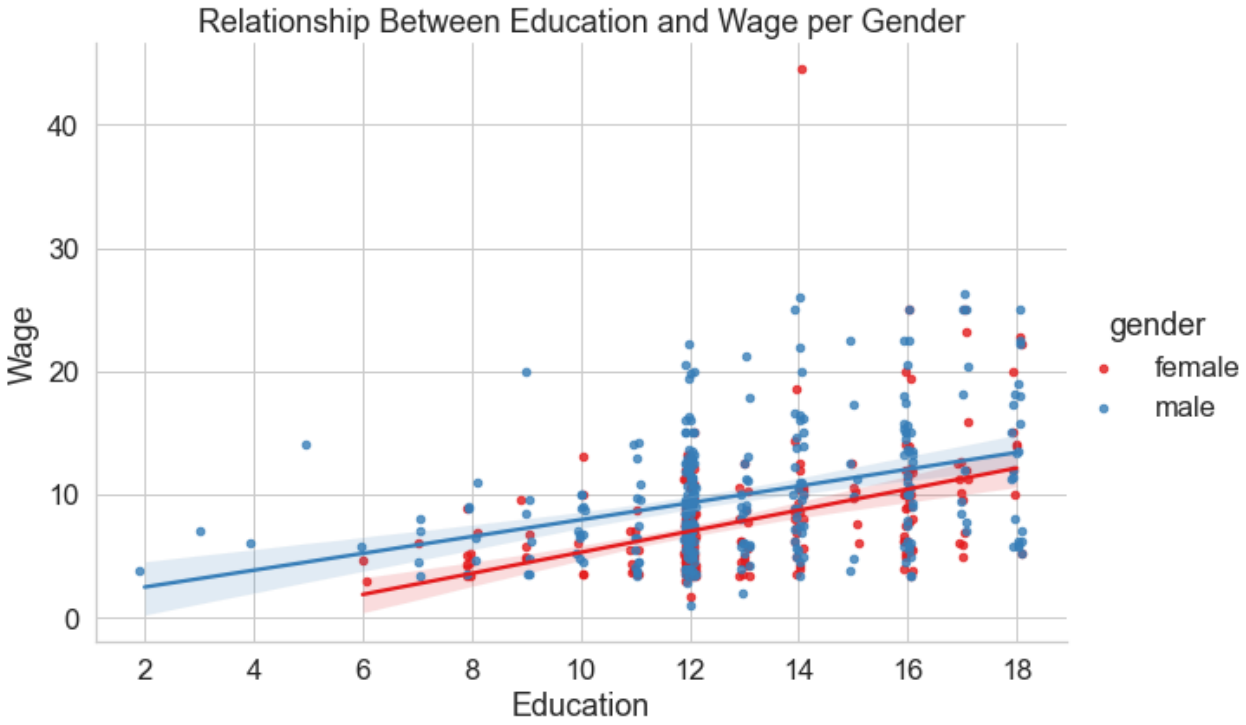
Interpretation of the plot above: The plot above displays the distribution of wages for each gender. The female distribution of wages is depicted in red, while the male distribution is depicted in blue. The first thing to note about the wage distribution for both genders is that they peak at relatively low wages, and the distributions are positvely skewed. This means that the median (the wage at which 50% of people earn less and 50% of people earn more) is less than the mean (the

average wage of this sample). More people earn low wages, regardless of their gender, and only a few lucky people earn wages as high as 20. However, there is a concerning difference between the female and male wage distributions. The female wage distribution is shifted to the left of the male wage distribution, signifying a lower average wage for females. More females earn low wages than males in this sample, evidenced by the consistently higher female wage distribution at lower wages. Furthermore, the female distribution peaks at a lower wage than the male distribution. After the female wage distribution peaks, it is consistently lower than the male wage distrubtion at higher wages. So, females in this sample earn less than the males on average. There could be other factors besides gender influencing the wages though, so I investigated how age and education are related to wages.

Wage's Relationship with Education for each Gender

In [6]:
```
sns.lmplot(data=md, x='education', y='wage', hue='gender',
           palette='Set1', height=6, aspect=1.5, scatter_kws = {"s" : 20}, x_jitter=0.1
plt.title("Relationship Between Education and Wage per Gender")
plt.xlabel("Education")
plt.ylabel("Wage")
```

Out[6]:
```
Text(49.094457561728404, 0.5, 'Wage')
```



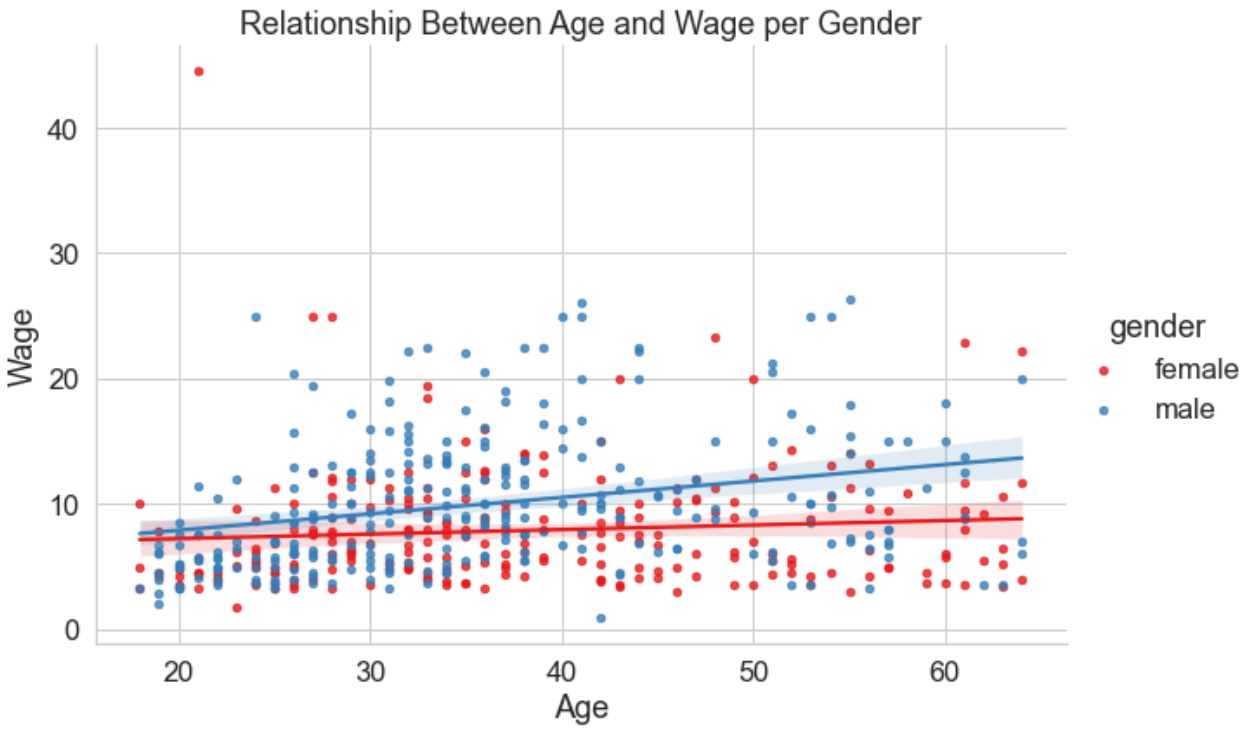
Interpretation of the plot above: Some people may assert that males earn higher wages because they attain more education than females. However, this graph clearly shows that this is not the case. This graph displays the relationship between education level and wage. Each point represents an individual's education level and their wage. The lines overlaying the points represent the relationship between education and wage. The data is also separated by gender, so females are represented in red and males are represented in blue, like the graph above. First, the overall trend that is visible, regardless of gender, is that more education leads to higher wages. I can tell this

because of the positively sloped regression lines, and even when considering the the potential error (represented by the shaded area around the lines), the relationship is always positive. However, the relationship between education and wage is slightly different for each gender. Females in this sample earn less than males at every level of education. So even though a female may have the same level of education as a male, she will probably earn less than that male for some inexplicable reason. Luckily, the wage gap decreases as females get more education, evidenced by the steeper slope of the female regression line, but females still earn less than males even with the highest education level possible. Females attaining less education than males doesn't cause the wage gap because the gap is present regardless of education level. There may be another underlying factor for the wage gap, so I also looked at the relationship between wage and age for each gender below.

Wage's Relationship with Age for each Gender

In [7]:
```
sns.lmplot(data=md, x='age', y='wage', hue='gender',
           palette='Set1', height=6, aspect=1.5, scatter_kws = {"s" : 20})
plt.title("Relationship Between Age and Wage per Gender")
plt.xlabel("Age")
plt.ylabel("Wage")
```

Out[7]:
```
Text(49.094457561728404, 0.5, 'Wage')
```



Interpretation of the plot above: The graph above displays the relationship between age and wage for each gender. Each data point represents an individual's age and their corresponding wage. The lines overlaid on the data represent the general relationship between age and wage based on this data, and they are surrounded by a confidence interval. The general trend in the data, when not considering gender, is that wage has either a slight positive or no relationship with age. However, the female and male data display slightly different trends once again. Female wages are consistently lower than male wages across all ages. Looking at the red female regression line, there doesn't seem

to be a relationship between wage and age because the line is relatively flat and the error bounds contain both positive and negative slopes. In contrast, there might be a positive relationship between wage and age for males because the blue regression line has a positve slope, and only positively sloped lines can be drawn within the error bounds. Males earn more than females on average when they are younger, though not by much. However, as males get older, there wage also increases. Since female wages don't increase with wage, the gap between male and female wages grows as they get older.

Conclusion

According to the graphs above, females earn less than males on average. Although the wage gap decreases as females get more education, more education does not solve the wage gap because females earn less than males at all education levels. Additionally, the wage gap grows as people get older because males earn more as they age while females do not. There is certainly a wage gap between males and females based on this data, and society must find a way to correct this problem.